Sebastian W. Meister, Natalie M. Hendrikse and John Löfblom*

Directed evolution of the 3C protease from coxsackievirus using a novel fluorescence-assisted intracellular method

<https://doi.org/10.1515/hsz-2018-0362>

Received August 31, 2018; accepted November 21, 2018; previously published online December 6, 2018

Abstract: Proteases are crucial for regulating biological processes in organisms through hydrolysis of peptide bonds. Recombinant proteases have moreover become important tools in biotechnological, and biomedical research and as therapeutics. We have developed a label-free high-throughput method for quantitative assessment of proteolytic activity in *Escherichia coli*. The screening method is based on co-expression of a protease of interest and a reporter complex. This reporter consists of an aggregation-prone peptide fused to a fluorescent protein via a linker that contains the corresponding substrate sequence. Cleavage of the substrate rescues the fluorescent protein from aggregation, resulting in increased fluorescence that correlates to proteolytic activity, which can be monitored using flow cytometry. In one round of flow-cytometric cell sorting, we isolated an efficiently cleaved tobacco etch virus (TEV) substrate from a 1:100 000 background of non-cleavable sequences, with around 6000-fold enrichment. We then engineered the 3C protease from coxsackievirus B3 (CVB3 3C^{pro}) towards improved proteolytic activity on the substrate LEVLFQ↓GP. We isolated highly proteolytic active variants from a randomly mutated CVB3 3C^{pro} library with up to 4-fold increase in activity. The method enables simultaneous measurement of proteolytic activity and protease expression levels and can therefore be applied for protease substrate profiling, as well as directed evolution of proteases.

Keywords: Coxsackievirus B3; FACS; GFP-fusion; intracellular assay; protease engineering; protease substrate profiling.

Introduction

Proteases regulate biological processes in all domains of life through their ability to catalyze the hydrolysis of peptide bonds. In multicellular organisms, activation or inactivation of biomolecules through proteolysis is implicated in the regulation of, for example, immunity, development and tissue repair (Doucet and Overall, 2008). About 2% of genes in all organisms encode for proteases, highlighting the crucial role this protein family plays in maintaining homeostasis (Rawlings et al., 2014).

Due to the catalytic nature of proteolysis, recombinantly produced proteases are used in a wide range of applications, e.g. as therapeutics, industrial enzymes and research reagents (Craik et al., 2011; Waugh, 2011; Vojcic et al., 2015). As of 2011, the Food and Drug Administration (FDA) has approved 12 proteases for therapeutic applications (Craik et al., 2011). For example, recombinant human factor VIII is a therapeutic protease, which is indicated for the treatment and prophylaxis of bleeding episodes in patients suffering from factor VIII deficiency (hemophilia A). Alkaline proteases from various *Bacillus* sp. are used as industrial enzymes in laundry detergent formulations (Banik and Prakash, 2004). Viral proteases like tobacco etch virus (TEV) protease and human rhinovirus (HRV) 3C protease are characterized by a particularly high substrate specificity and are therefore used as research reagents, e.g. for the processing of fusion proteins or activation of proenzymes (Phan et al., 2002; Gray et al., 2010; Chen et al., 2017). HRV 3C protease is attractive among commercially available proteases because of its robust activity at low temperatures (Ullah et al., 2016).

Aiming to expand the scope of practical applications of this enzyme class, researchers have engineered or evolved proteases with tailor-made properties such as increased stability or solubility, prolonged activity, activity at low temperatures, altered substrate specificity and catalytic activity (Dumont et al., 2012; Vojcic et al., 2015; Packer et al., 2017;

*Corresponding author: John Löfblom, Department of Protein Science, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden, e-mail: lofblom@kth.se.
<https://orcid.org/0000-0001-9423-0541>

Sebastian W. Meister: Department of Protein Science, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

Natalie M. Hendrikse: Department of Fibre and Polymer Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

Zhao and Feng, 2018). To this end, both rational design and directed evolution strategies have been employed (Guerrero et al., 2017). For instance, substrate specificity of proteases for which biochemical and structural data are available has been successfully altered using rational design approaches (Hedstrom et al., 1992), whereas directed evolution of proteases via selection or screening of combinatorial protein libraries is beneficial in cases where structural data is scarce. Another benefit of the latter approach is that a relatively extensively mutated protein library increases the possibility of identifying epistatic interactions between mutations (Longwell et al., 2017). In recent years, several screening and selection methods have been reported for directed evolution of protease enzymes (Guerrero et al., 2017). These methods generally rely on the coupling of the proteolytic activity to a measurable biochemical property, such as antibiotic resistance (Carrico et al., 2016), propagation (Dickinson et al., 2015), fluorescence (Kostallas et al., 2011) or surface expression of reporter tag sequences (Yi et al., 2013). Microbial cells are commonly used as reaction vessels for the enzymatic reaction and enable phenotype-genotype coupling.

Here, we have developed a label-free method for directed evolution of protease activity and protease substrate profiling in the cytoplasm of *Escherichia coli*. To this end, we designed a two-plasmid system that links protease activity to the solubility of a fluorescent reporter protein.

It has previously been shown that fusion of the aggregation-prone amyloid beta 1-42 ($A\beta_{1-42}$) peptide, which is associated with Alzheimer's disease, to the N-terminus of green fluorescent protein (GFP) prevents proper folding of GFP in *E. coli* and thereby decreases its fluorescence signal (Wurth et al., 2002). Such fusions have been applied in previous studies, e.g. to assess aggregation tendencies of different $A\beta$ peptides (Kim and Hecht, 2008), and for the screening of $A\beta$ aggregation-inhibitors (Kim et al., 2006; Chakrabortee et al., 2012; Lindberg et al., 2017). Based on this principle, we designed a reporter protein consisting of N-terminal $A\beta_{1-42}$, followed by a protease cleavage site, and enhanced green fluorescent protein (EGFP) at the C-terminal. The hypothesis was that co-expressing the plasmid encoding the reporter protein and a plasmid encoding the protease of interest would result in proteolysis of the substrate and separation of $A\beta_{1-42}$ from EGFP and restored fluorescence signal. We fused the protease to mCherry, which would allow for monitoring of expression levels of protease variants, while also potentially improving solubility of proteases expressed in the cytoplasm (Van der Henst et al., 2010). In order to verify our hypothesis, we tested the method with respect to substrate discrimination and directed evolution of a protease.

We used TEV protease and a reporter protein harboring the canonical substrate sequence for optimization of the system. Using optimized culturing conditions, we isolated cells expressing reporter plasmids with the canonical sequence from a 1:100 000 background of cells expressing a non-processed sequence.

Finally, we engineered a 3C protease from coxsackievirus B3 towards improved proteolytic activity on the peptide substrate LEVLFQ↓GP by directed evolution. Protease variants were simultaneously selected for high expression of soluble protein in the flow cytometer. 3C protease variants enriched after three consecutive rounds of FACS exhibited increased proteolytic activity (1.8–3.7-fold).

Results

Design and construction of expression vectors and proof of concept experiment

We designed a two-plasmid method for co-expression of the reporter construct and the protease in *E. coli*, with the aim of simultaneous flow-cytometric screening of protease activity and soluble protease expression levels (Figure 1).

The gene for the amyloid beta peptide 1-42 was subcloned into the reporter plasmid as N-terminal fusion to EGFP via a peptide linker that contained the canonical TEVp cleavage sequence ENLYFQ↓G (subG) under control of a rhamnose promoter. The gene encoding for TEVp was subcloned into the protease plasmid as N-terminal fusion to the fluorescent protein mCherry under control of an arabinose promoter.

We first tested the principle of the method in a proof of concept experiment. Both plasmids were co-transformed into *E. coli* and cells were analyzed by flow cytometry upon expression of the fusion proteins. The low whole-cell fluorescence of cells in which only the reporter complex was expressed indicated that introducing a peptide linker with the protease substrate did not interfere with aggregation of the fusion protein. Moreover, a 2-fold increase in whole-cell EGFP fluorescence was observed upon expression of TEVp, showing that the principle of the method worked as intended (Figure 2A).

Investigation of cultivation conditions and aggregation-prone peptides

In order to increase the sensitivity of the method, we aimed to expand its dynamic range, reflected in the shift of

A

B

Figure 1: Schematic overview of the intracellular method for flow-cytometric screening of protease activity and protease substrate profiling. (A) Genetic fusion of $A\beta_{1-42}$ to the N-terminus of GFP prevents the GFP portion of the fusion protein from forming its native fluorescent structure, resulting in a low fluorescence signal. Protease substrate sequences are introduced into the linker between $A\beta_{1-42}$ and GFP. Co-expression of a corresponding protease results in proteolysis of the substrate and separation of $A\beta_{1-42}$ from GFP and the fluorescence signal is restored. (B) In combinatorial applications, different expression levels of protease candidates can lead to biases in the selection procedure. Fusion of the protease/library to fluorescent mCherry allows normalization of proteolysis data against differences in expression levels on single cell basis.

fluorescence, by exploring different experimental conditions and different aggregation prone peptides. We investigated the effect of inducer concentrations, respective time points of induction of reporter and protease plasmid and cultivation temperatures. Briefly, TEVp-mCherry was induced with 0.2%, 0.4%, 0.6% or 0.8% L-arabinose. The cultures were incubated for 0 h, 2 h, 4 h or 8 h before induction of the reporter proteins with 6 mM, 8 mM, 10 mM or 12 mM L-rhamnose. The experiments were carried out at 30°C and 37°C. At 16 h after induction, GFP fluorescence signals of the cells were measured using flow cytometry. The mean fluorescence signals of cells expressing TEVp-mCherry and the reporter protein were compared to cells expressing only the reporter plasmid. The largest dynamic range was observed after cultivation at 37°C and induction of the protease plasmid with 0.2% L-arabinose, followed by induction of the reporter plasmid with 12 mM

L-rhamnose 8 h later (Figure S1), corresponding to around 6.5-fold increase in EGFP fluorescence in cells expressing both reporter construct and protease, compared to cells expressing reporter protein alone. Next, we explored the possibility of improving the dynamic range of the method by utilizing different $A\beta_{1-42}$ variants in the reporter construct. The single amino acid substitutions Q15L and E22G have been reported to enhance the aggregation propensity of the $A\beta_{1-42}$ peptide (Dahlgren et al., 2002; Kim and Hecht, 2008). Reporter plasmids encoding $A\beta_{1-42}$ (E22G) or $A\beta_{1-42}$ (Q15L/E22G) were constructed and co-transformed with the previously described TEVp-mCherry plasmid. The largest dynamic range was observed for cells containing the $A\beta_{1-42}$ (Q15L/E22G) reporter, corresponding to a 9-fold increase in EGFP fluorescence for cells expressing protease and reporter protein, compared to cells expressing reporter protein alone (Figure 2B). As a control, we

Figure 2: Flow-cytometric analysis of intracellular protein expression in *E. coli*.

(A) Representative histograms from flow-cytometric analysis of cells over-expressing only $A\beta_{1-42}$ -EGFP or $A\beta_{1-42}$ -ENLYFQG-EGFP are shown in purple and red, respectively. Cells co-expressing $A\beta_{1-42}$ -ENLYFQG-EGFP and TEVp-mCherry are represented by the green histogram. (B) Flow-cytometric analysis of intracellular protein expression after optimization of the dynamic range. A representative histogram of cells over-expressing only $A\beta_{1-42}$ (Q15L/E22G)-ENLYFQG-EGFP (negative control) is shown in red. Cells co-expressing $A\beta_{1-42}$ (Q15L/E22G)-ENLYFQG-EGFP and TEVp-mCherry are represented by the green histogram and cells co-expressing $A\beta_{1-42}$ (Q15L/E22G)-ENLYFQ-EGFP and TEVp-mCherry are represented by the blue histogram.

generated a reporter plasmid containing the peptide sequence ENLYFQ↓P (subP) in the linker, which has been reported to be non-cleavable (Kapust et al., 2002). Indeed, flow-cytometric analysis showed no shift in whole-cell fluorescence when comparing cells expressing TEVp and subP reporter construct to cells expressing a reporter construct only (Figure 2B).

Flow-cytometric cell sorting of a mock protease substrate library

To investigate the potential of the method to enrich preferred protease substrates from combinatorial protein libraries, we carried out a mock selection experiment using fluorescence-activated cell sorting (FACS). To this end, *E. coli* cells were transformed with TEVp-mCherry and a reporter plasmid encoding either the non-cleavable sequence ENLYFQ↓P (subP) or the consensus TEVp substrate ENLYFQ↓G (subG). For the mock selection, cells expressing TEVp-mCherry/ $A\beta_{1-42}$ (Q15L/E22G)-subG-EGFP were spiked into cells expressing TEVp-mCherry/ $A\beta_{1-42}$ (Q15L/E22G)-subP-EGFP at a 1:100 000 ratio. One round of cell sorting was performed in which cells exhibiting high EGFP fluorescence signals were collected. Sequencing of the reporter plasmids of the sorted cells revealed

that 6.25% contained the consensus TEVp substrate ENLYFQ↓G, corresponding to a 6250-fold enrichment in only one round of cell sorting (Figure S2).

Selection of coxsackievirus 3C protease variants with improved catalytic activity

Having demonstrated the effectiveness of the method in a mock selection experiment, we next investigated whether it could be used to engineer protease variants with desired properties. We aimed to engineer the 3C protease from coxsackievirus B3 (CVB3 3C^{pro}) towards improved catalytic activity on the substrate LEVLFQ↓GP. LEVLFQ↓GP is a commonly used sequence for HRV 3C^{pro}-catalyzed cleavage of fusion proteins, however, it is not the preferred substrate of CVB3 3C^{pro} (Lee et al., 2009). Via fusion of the 3C^{pro} candidates to mCherry, we simultaneously aimed to enrich variants that might exhibit higher expression of soluble protein. We generated a random mutagenesis protease library by carrying out three consecutive rounds of error-prone polymerase chain reaction (PCR) on the CVB3 3C^{pro} gene. *Escherichia coli* cells containing the LEVLFQ↓GP reporter plasmid were transformed with the protease library, resulting in a diversity of approximately 2.3×10^7 unique clones. The diversity of amino

Figure 3: Flow-cytometric sorting of CVB3 3C^{pro} library.

(A) Representative dot plots of the three rounds of flow-cytometric sorting are shown. The cells were cultured using the optimal conditions (37°C, 220 rpm, 0.2% L-arabinose, 8 h incubation, 12 mM L-rhamnose, 16 h incubation) and analyzed by flow-cytometry. The leftmost plot shows cells before sorting. The middle and rightmost plots show cells after one and two rounds of sorting, respectively. The sorting gates with percentages of collected cells of the three sorting rounds are indicated in the plots. (B) Sequence analysis of CVB3 3C^{pro} library. The frequency of amino acid mutations compared to wild-type CVB3 3C^{pro} in the naïve library (left) and after three rounds of FACS (right) is shown (56 sequences were analyzed for the naïve library and 76 sequences were analyzed after three rounds of FACS).

acid mutations in the protease library was verified by sequencing. Among the sequenced library candidates, the most prevalent number of amino acid mutations was seven, indicating that on average approximately 3.8% of the residues of the 183 aa CVB3 3C^{pro} sequence were mutated (Figure 3B).

Three consecutive rounds of FACS with increasing stringency were carried out to collect cells exhibiting both high EGFP and mCherry fluorescence signals (Figure 3A). Sequencing of 76 of the selected clones identified 39 unique CVB3 3C^{pro} variants. The most enriched variant (A05) had a prevalence of nine out of 76 and contained five mutations compared to wild-type CVB3 3C^{pro}. In general, we observed a tendency towards fewer mutations (two to five) in the sort output (Figure 3B), compared to the unsorted library. Interestingly, 16 out of 39 enriched variants contained the amino acid

substitution T130I, while three out of 39 variants contained a T130S mutation (Figure 4).

We chose to analyze all variants with a prevalence >1 (14 variants) for proteolytic activity on the LEVLFQ↓GP substrate. Single clones were cultured according to the optimized protocol and whole-cell EGFP fluorescence signals were analyzed using flow cytometry. All analyzed clones showed larger shifts in EGFP fluorescence signals upon expression of the protease variant, as compared to the shift observed for wild-type CVB3 3C^{pro} (Figure S3). Strikingly, CVB3 3C^{pro} variant A04 showed a 5-fold larger shift using the intracellular method (Figure 5A, Table 1).

We chose five potentially interesting protease variants based on either enrichment (sequence prevalence in sort output) or observed shift in whole-cell fluorescence. The sequences of the selected protease variants are shown in Figure S4. These five variants as well as

Figure 4: Crystal structure of 3C protease of coxsackievirus B3 (PDB: 3ZYD).

Close-up of the active site with amino acids forming the catalytic triad (H40, E71, C147) depicted as orange sticks. The purple stick represents amino acid T130, which is a mutational hotspot in enriched protease variants.

wild-type CVB3 3C^{pro} were expressed as soluble proteins with an N-terminal hexahistidine tag using the pET-45b(+) vector in *E. coli* BL21 Star (DE3). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the soluble fractions of lysed *E. coli* cells indicated higher expression levels of soluble protein for at least a majority of the selected protease variants compared to wild-type (Figure S5). Following purification via IMAC, only protein bands corresponding to the proteases were visible after SDS-PAGE (Figure S6). Proteolytic activity of the protease candidates was characterized in an *in vitro* assay (Abcam plc; ab211088). The engineered protease variants showed 1.8–3.7-fold increase in protease activity compared to wild-type CVB3 3C^{pro} (Figure 5B, Table 1).

Discussion

Herein, we describe a label-free, high-throughput and whole-cell method for directed evolution of protease activity and protease substrate profiling in *E. coli*. As a proof of concept, the consensus substrate of TEVp was enriched

Figure 5: Analysis of the proteolytic activity of evolved CVB3 3C^{pro} variants.

(A) Representative histograms of cells co-expressing Aβ₁₋₄₂(Q15L/E22G)-LEVLFQGP-EGFP and any of the protease variants wt CVB3 3C^{pro} (green), A03 (gray), A05 (yellow), B06 (purple), A09 (orange), A04 (blue). The cells were cultured using the optimal conditions (37°C, 220 rpm, 0.2% L-arabinose, 8 h incubation, 12 mM L-rhamnose, 16 h incubation) and analyzed by flow-cytometry. The red histogram represents a negative control of cells over-expressing Aβ₁₋₄₂(Q15L/E22G)-LEVLFQGP-EGFP only. (B) 3C protease activities of the evolved protease variants normalized to the activity of wt CVB3 3C^{pro} determined using an *in vitro* protease activity assay. Statistical evaluation was performed using an unpaired, two-tailed Student's *t*-test.

Table 1: Overview of the properties of evolved CVB3 3C^{pro} variants.

	Number of mutations	Prevalence	Relative EGFP shift ^{a,b}	Relative 3C ^{pro} activity ^{a,c,d} (U/mg)
Wild-type	0	3/76	1.0	1.0 ± 0.10
A03	5	3/76	3.0	2.1 ± 0.16
A04	3	5/76	5.1	1.8 ± 0.32
A05	5	9/76	3.1	3.3 ± 0.37
A09	3	5/76	3.7	1.9 ± 0.28
B06	4	3/76	3.4	3.7 ± 0.32

^aNormalized to wild-type.^bBased on intracellular GFP signals as analyzed by flow cytometry.^cBased on results from 3C protease activity assay on soluble proteases.^dOne unit (U) is defined as the amount of enzyme which can cleave 1 nmol of substrate per minute under assay conditions.

from a pool of non-cleavable substrates by 6500-fold in one single round of sorting. The potential of the method was further demonstrated by engineering the 3C protease from coxsackievirus B3 towards improved catalytic activity on the substrate LEVLFQ↓GP.

The proposed method relies on linking protease activity to the solubility and correct folding of a fluorescent reporter protein. To verify the method, we first evaluated the possibility of restoring fluorescence of the reporter protein upon cleavage by a respective protease. Whole-cell fluorescence of *E. coli* cells expressing a reporter protein containing the TEVp consensus substrate was monitored. The EGFP signal of cells expressing Aβ₄₂-subG-EGFP was low (Figure 2A), showing that introduction of a protease substrate linker in between Aβ₄₂ and EGFP does not abolish aggregation of the reporter protein. Upon co-expression of TEVp we observed an increase in whole-cell EGFP fluorescence, supporting the strategy for screening protease activity. Moreover, changing one amino acid from the consensus TEVp substrate to a non-cleavable sequence abolished the observed effect, demonstrating that the increase in whole-cell fluorescence relies on specific cleavage of the substrate sequence within the linker of the reporter construct. By optimizing the cultivation conditions and exploring different Aβ₄₂ mutations within the reporter plasmid a 9-fold shift in EGFP fluorescence was observed for cells co-expressing TEVp reporter protein compared to cells expressing only the reporter protein.

The protease and reporter protein are expressed from two separate plasmids, where the promoter within the reporter plasmid allows for titratable protein expression (Giacalone et al., 2006). We believe that the two-plasmid system allows for better control over selection

experiments. It should be possible to fine-tune selection stringency by utilizing different induction times and reporter protein concentrations.

In order to show that the method can be applied to select preferred protease or substrate variants, we performed a mock selection experiment with the aim to isolate cells containing the subG reporter protein from a large background of cells containing the subP reporter protein. Sequencing of reporter plasmids after only one single round of flow-cytometric sorting revealed an enrichment of subG by 6250-fold.

Next, we aimed to isolate variants of the 3C protease from coxsackievirus B3 with improved proteolytic activity on the peptide substrate LEVLFQ↓GP from a randomly-mutated protease library. After three consecutive rounds of sorting, 39 unique protease sequences were enriched. Interestingly, 16 of the enriched variants after sorting harbored a T130I mutation, which thus indicates a significant effect on the protease-substrate interaction. This is also supported by the notion that T130 is located within or in close proximity to a groove that forms the substrate binding pocket of the CVB3 3C^{pro} (Figure 4). We speculate that the hydrophobic nature of isoleucine could be advantageous in accommodating the LEVLFQ↓GP substrate in comparison to the polar side chain of threonine. Alternatively, the mutation could induce a conformational shift in the active site of the protease that helps to process the substrate. However, more future studies are required to elucidate the importance of T130I in detail.

The observed intracellular proteolytic activity could be influenced by various factors, such as the amino acid sequences flanking the substrate in the reporter construct and the environment within the bacterial cells. To confirm the observed proteolytic activities of the engineered CVB3 3C^{pro} variants using an independent method, an *in vitro* activity assay using the substrate sequence capped with the chromophore pNA was carried out. All protease variants exhibited increased activity compared to wild type, ranging from 1.8 to 3.7-fold increased proteolytic activity. In addition, analysis of the soluble fraction from lysed bacteria by SDS-PAGE indicated higher amount of soluble protease for four out of five of the evolved variants as compared to wild-type CVB3 3C^{pro} (Figure S5).

Although the method proved useful for engineering of CVB3 3C^{pro} variants with improved proteolytic activity on the LEVLFQ↓GP substrate, we cannot rule out the possibility of an overall relaxed specificity of the evolved variants. In future studies, it would thus be interesting to profile the substrate specificity of the engineered 3C^{pro} variants using the same method and a combinatorial substrate library or other previously reported substrate profiling methods,

like substrate phage display (Matthews and Wells, 1993; Ratnikov et al., 2009) or substrate cell surface display (Sandersjöo et al., 2017). Moreover, it should be possible to introduce counter-selection substrates into the system by utilizing other fluorescent proteins fused to A β_{42} or antibiotic resistance in upcoming experiments. In this study, we used bacterial cells as enzyme reaction vessels for the quantification of protease activity and for genotype-phenotype coupling. However, the cytoplasm of *E. coli* is not suited for production of many mammalian proteases that, e.g. depend on post-translational modifications in their active form. Hussein et al. recently reported the use of A β -GFP to study the effect of DNAJB6 overexpression on A β -aggregation in HEK293 cells (Hussein et al., 2015). Hence, it should be interesting to explore whether the GFP solubility reporter system for the screening of protease activity would work in mammalian cells as well.

In summary, we have developed a new method that enables directed evolution of proteases and protease substrate profiling that links protease activity to the solubility of a fluorescent reporter protein in the cytoplasm of *E. coli*. Using this method, we successfully evolved 3C protease variants with improved proteolytic activity for the substrate peptide LEVLFQ↓GP. As opposed to other protease engineering strategies, there is no need for production and purification of the investigated protease and/or substrate variants, which otherwise can be laborious and expensive. Moreover, our method enables direct measurement of relative protease activities with single cell resolution using flow cytometry, and without the need of labeling procedures. We believe that this method holds promise as an inexpensive and versatile system for the selection of proteases with desired properties from large mutant libraries.

Materials and methods

Bacterial strains

The *E. coli* strain TOP10 [F-*mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *araD139* Δ (*araleu*)7697 *galU* *galK* *rpsL* (StrR) *endA1* *nupG*] (Thermo Fisher Scientific, Waltham, MA, USA) was used for subcloning work and intracellular protein expression. *Escherichia coli* strain BL21 Star (DE3) [F-*ompT* *hsdS_B* (*r_B⁻*, *m_B⁻*) *galdcmrne131* (DE3)] was used for soluble expression of protease proteins.

Construction of expression vectors

For the assembly of the reporter plasmid the gene encoding A β_{1-42} -G4S-substrate-G4S-EGFP was designed with *SaII* and *XhoI*-sites upstream and downstream. A *NcoI*-site between A β_{1-42} and

G4S-substrate, and a *BamHI*-site between substrate-G4S and EGFP were included to be able to change the protease substrate in the reporter plasmid. The gene was ordered from Thermo Fisher Scientific. Insert and pRha67K vector (Xbrane Bioscience, Stockholm, Sweden) were digested using *SaII*-HF and *XhoI* restriction enzymes (New England BioLabs). Digested insert and vector were purified using QIAquick PCR Purification Kit (Qiagen GmbH, Hilden, Germany) and QIAquick Gel Extraction Kit (Qiagen), respectively. Purified insert and vector were ligated using T4 DNA Ligase (NEB). Genes encoding different protease substrates were designed with a *NcoI*-G4S sequence upstream and a G4S-*BamHI* sequence downstream and purchased from Integrated DNA Technologies. Protease substrate inserts and reporter plasmid were digested using *NcoI*-HF and *BamHI*-HF restriction enzymes (NEB), purified, and ligated as described already.

For the assembly of the protease plasmid the gene encoding TEVp-G4S-mCherry was designed with *KpnI* and *SphI*-sites upstream and downstream. A TEV protease S219V mutant (Kapust et al., 2001) was used in all experiments. A *NotI*-site between TEVp-G4S and mCherry was included to be able to change the protease gene. The gene was ordered from ThermoFisher Scientific. Insert and pBAD33 vector (Guzman et al., 1995) were digested using *KpnI*-HF and *SphI*-HF restriction enzymes (NEB), purified and ligated as described already. The gene encoding 3C protease of coxsackievirus B3 (PDB ID: 3ZYD) was amplified by PCR using specific primers (IDT) to introduce a *KpnI* site upstream and a G4S-*NotI* sequence downstream. The amplified 3C protease gene was purified using QIAquick PCR Purification Kit (Qiagen). 3C insert and protease plasmid were digested using *KpnI*-HF and *NotI*-HF restriction enzymes (NEB), purified, and ligated as described already.

All ligated plasmids were transformed to *E. coli* TOP10 (Thermo Fisher Scientific) by heat shock. Plasmids were prepared using QIAprep Spin Miniprep Kit (Qiagen) and the sequences were verified by DNA sequencing (Microsynth AG, Balgach, Switzerland).

Proof-of-concept for intracellular cleavage of A β_{1-42} -substrate-EGFP reporter protein

Protease plasmid encoding either TEVp-mCherry or 3Cpro-mCherry was co-transformed with the reporter plasmid encoding the respective protease substrate into TOP10 by heat shock. Individual colonies containing both plasmids were inoculated to Luria Bertani (LB) medium (Merck KGaA, SV Darmstadt, Germany) containing appropriate antibiotics (10 μ g/ml chloramphenicol and 50 μ g/ml kanamycin) and grown for 16 h at 37°C and 150 rpm. An aliquot of the culture was diluted 1:200 in LB containing antibiotics and grown at 37°C and 150 rpm until OD₆₀₀ of 0.6 was reached. Protease-mCherry expression was induced by addition of 0.6% L-arabinose (Merck KGaA). The cultures were incubated for 2 h at 37°C and 150 rpm. Expression of the reporter protein was induced by addition of 6 mM L-rhamnose (Merck KGaA). Induced cultures were cultivated for 16 h at 37°C and 150 rpm. Approximately 10⁷ cells were pelleted (1500 g, 4°C, 4 min) and washed in phosphate-buffered saline (PBS) before resuspension in 600 μ l PBS. The cells were analyzed using a Gallios™ flow cytometer (Beckman Coulter, Inc., Indianapolis, IN, USA). Cleavage of the A β_{1-42} -substrate-EGFP reporter protein was monitored by measuring mean fluorescence intensities (MFI) of EGFP.

Investigation of cultivation conditions using flow cytometry

Protease plasmid encoding TEVp-mCherry was co-transformed with either a reporter plasmid encoding the consensus TEVp substrate ENLYFQ↓G (subG) or a reporter plasmid encoding the non-cleavable sequence ENLYFQ↓P (subP) into TOP10 by heat shock. Individual colonies containing both plasmids were inoculated to LB medium (Merck KGaA) containing appropriate antibiotics and grown for 16 h at 37°C and 150 rpm. An aliquot of the culture was diluted 1:200 in LB containing antibiotics at 37°C and 220 rpm until OD₆₀₀ of 0.6 was reached. TEVp-mCherry expression was induced by addition of 0.2%, 0.3%, 0.4%, 0.6% or 0.8% L-arabinose (Merck KGaA). The cultures were incubated for 0 h, 2 h, 4 h or 8 h at 30°C or 37°C and 220 rpm. Expression of the reporter protein was induced by addition of 6 mM, 8 mM, 10 mM or 12 mM L-rhamnose (Merck KGaA). Induced cultures were cultivated for 16 h at 30°C or 37°C and 220 rpm. Approximately 10⁷ cells were pelleted (1500 g, 4°C, 4 min) and washed in phosphate-buffered saline (PBS) before resuspension in 600 µl PBS. The cells were analyzed using a Gallios flow cytometer (Beckman Coulter) or a MoFlo Astrio EQ cell sorter (Beckman Coulter). Cleavage of the Aβ₁₋₄₂-substrate-EGFP reporter protein was monitored by measuring mean fluorescence intensities (MFI) of EGFP and expression of TEVp was monitored by measuring MFI of mCherry.

For all following experiments the cells were cultured using the optimal conditions (37°C, 220 rpm, 0.2% L-arabinose, 8 h incubation, 12 mM L-rhamnose, 16 h incubation).

Construction of reporter plasmids expressing Aβ₁₋₄₂ variants

The quickchange method was used to construct reporter plasmids encoding Aβ₁₋₄₂(E22G)-G4S-subG-G4S-EGFP and Aβ₁₋₄₂(Q15L/E22G)-G4S-subG-G4S-EGFP. Complementary primers containing the respective mutants were used to PCR amplify the entire plasmid. Template plasmid was eliminated by digestion with *DpnI* and the PCR products were transformed to *E. coli* TOP10 (Thermo Fisher Scientific) by heat shock. Plasmids were prepared using QIAprep Spin Miniprep Kit (QIAGEN) and the sequences were verified by DNA sequencing (Microsynth).

Cell sorting of mock protease substrate library

Cells co-expressing pBAD TEVp-mCherry/pRha Aβ₁₋₄₂(Q15L/E22G)-subG-EGFP were mixed at a ratio of 1:100 000 with cells co-expressing pBAD TEVp-mCherry/pRha Aβ₁₋₄₂(Q15L/E22G)-subP-EGFP. Approximately 10⁷ cells were pelleted (1500 g, 4°C, 4 min) and washed in PBS before resuspension in 600 µl PBS. One round of sorting was performed using a MoFlo Astrios EQ cell sorter. The sorting gate was drawn based on the MFI signal of cells co-expressing pBAD TEVp-mCherry/pRha Aβ₁₋₄₂(Q15L/E22G)-subG-EGFP. Twice 100 cells were sorted directly onto two LB agar plates containing appropriate antibiotics (100 cells per plate), and incubated at 37°C for 16 h. Ninety-six single colonies were picked and the sequences of the reporter plasmids were analyzed by DNA sequencing (Microsynth).

Generation of random mutagenesis 3C protease library

Error prone PCR was done using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA), following the provided protocol's instructions. The 3C protease of coxsackievirus B3 gene was used as template DNA with the already mentioned 3C primers. Three consecutive rounds of error prone PCR were carried out. The PCR product was purified using QIAquick PCR Purification Kit (QIAGEN) after each round. After three rounds of error prone PCR, the PCR product and protease plasmid were digested using KpnI-HF and NotI-HF restriction enzymes (NEB), purified and ligated as described already.

The reporter plasmid encoding the 3C substrate LEVLFQ↓GP was transformed to *E. coli* TOP10 (Thermo Fisher Scientific) by heat shock. An individual colony containing the reporter plasmid was picked, grown and electrocompetent cells were prepared as described elsewhere (*Molecular Cloning: A Laboratory Manual*, Third Edition). The ligated random mutagenesis 3C protease plasmid was purified using QIAquick PCR Purification Kit (Qiagen) and transformed to TOP10 cells containing the reporter plasmid by electroporation. Ten electroporation reactions were carried out in total and the cells were transferred to LB medium (Merck KGaA) and incubated for 30 min at 37°C and 150 rpm. A sample of the cell culture was diluted and spread on agar plates containing appropriate antibiotics to determine the total number of transformants. The cell culture was thereafter inoculated into 500 ml LB medium (Merck KGaA) containing appropriate antibiotics and incubated for 16 h at 37°C and 150 rpm. The 500 ml culture was centrifuged at 5000 g, 4°C for 5 min and the cell pellet resuspended in 15% glycerol. The *E. coli* library was thereafter stored in aliquots at -80°C.

Cell sorting of random mutagenesis 3C protease library

Approximately 3.0×10⁹ cells co-expressing the plasmid encoding random mutagenesis 3C proteases and the reporter plasmid encoding the 3C substrate LEVLFQ↓GP were pelleted (1500 g, 4°C, 4 min). The pellet was washed in phosphate buffered saline (PBS) before resuspension in 1 ml PBS. The cells were sorted based on GFP and mCherry signals. Three rounds of cell sorting were carried out with increasing sorting stringency for GFP and mCherry signals (1.5%, 0.5% and 0.1% of cells within the sort gate, respectively). A total of 5.0×10⁸ cells were sorted in the first round to oversample the number of transformants 20-fold. In subsequent rounds a 10-fold excess of previously sorted cells was sampled. The cells were sorted using a MoFlo Astrio EQ cell sorter (Beckman Coulter). In the last round, twice 100 cells were sorted directly onto two LB agar plates containing appropriate antibiotics (100 cells per plate), and incubated at 37°C for 16 h. Ninety-six single colonies were picked and the sequences of the protease plasmids were analyzed by DNA sequencing (Microsynth).

Ranking of engineered 3C protease candidates based on intracellular GFP signals

Cells containing enriched protease sequences as determined by DNA sequencing (incidence >1) were picked and cultured as described already. GFP fluorescence intensities of the cells were analyzed using

a Gallios™ flow cytometer (Beckman Coulter). Cells containing the top five protease candidates based on GFP fluorescence signals were cultured as described and GFP and mCherry fluorescence intensities were analyzed using a MoFlo Astrio EQ cell sorter (Beckman Coulter).

Expression and purification of engineered 3C protease candidates

Sequences encoding for engineered variants of 3C protease of coxsackievirus B3 were cloned into pET-45b(+) (Merck KGaA) vector to introduce a N-terminal hexahistidine tag. Cells containing engineered protease candidates were cultured and plasmids were prepared using QIAprep Spin Miniprep Kit (Qiagen). The genes encoding engineered protease variants were amplified by PCR using specific primers (IDT) to introduce a *KpnI* site upstream and a *XhoI* sequence downstream. Protease variants inserts, the wild-type CVB3 3C^{pro} insert, and pET-45b(+) were digested using *NcoI*-HF and *BamHI*-HF restriction enzymes (NEB), purified, and ligated as described above. All ligated plasmids were transformed to *E. coli* TOP10 (Thermo Fisher Scientific) by heat shock. Plasmids were prepared using QIAprep Spin Miniprep Kit (Qiagen) and the sequences were verified by DNA sequencing (Microsynth).

Plasmids were transformed to *E. coli* BL21 Star (DE3). Individual colonies were inoculated to LB medium (Merck KGaA) containing antibiotics (50 µg/ml ampicillin) and grown for 16 h at 37°C and 150 rpm. An aliquot of the culture was diluted 1:200 in LB containing antibiotics and grown at 37°C and 150 rpm until OD₆₀₀ of 0.8 was reached. Protein expression was induced by addition of 0.1 mM IPTG (Chemtronics AB, Sollentuna, Sweden). The cultures were incubated for 4 h at 37°C and 150 rpm. The bacteria were lysed using sonication and the recombinant proteins were purified using HisPur™ Cobalt Resin (Thermo Fisher Scientific) according to the manufacturer's instructions. SDS-PAGE analysis of recombinant proteins before and after purification was carried out using NuPAGE™ 4–12% Bis-Tris 1.0 mm × 10 well gels (Thermo Fisher Scientific). For the comparison of expression levels of soluble enzymes, equal numbers of *E. coli* cells expressing the different protease variants were lysed and the soluble fractions were loaded on the gel. For the assessment of purity, 2 µg of each of the purified proteins was loaded on the gel. Electrophoresis was carried out at 200 V and 0.3 A for 45 min. The gels were stained using GelCode™ Blue Safe Protein Stain (Thermo Fisher Scientific) according to the manufacturer's protocol.

3C protease activity assay

The activity of the engineered 3C protease candidates was determined using an HRV 3C protease activity kit (Abcam plc, Cambridge, UK) according to the manufacturer's protocol. Engineered 3C protease proteins and wild-type coxsackievirus 3C protease were purified as described and used in the assay. Briefly, the proteins were diluted in assay buffer and 3C protease substrate was added. The release of the chromophore pNA was immediately measured at an absorbance of OD 405 nm on a microplate reader (CLARIOstar, BMG Labtech GmbH, Ortenberg, Germany) for 1 h. The assay was carried out in triplicates for CVB 3C^{pro}, A09, A05, A04, A03 and in duplicate for variant B06. The amount of released pNA was determined by comparison to a freshly prepared pNA standard curve and 3C protease activity was calculated.

Acknowledgments: This work was supported by the ProNova VINN Excellence Center for Protein Technology (VINNOVA, Funder Id: 10.13039/501100001858, 2015-04037).

References

- Banik, R.M. and Prakash, M. (2004). Laundry detergent compatibility of the alkaline protease from *Bacillus cereus*. *Microbiol. Res.* 159, 135–140.
- Carrico, Z.M., Strobel, K.L., Atreya, M.E., Clark, D.S., and Francis, M.B. (2016). Simultaneous selection and counter-selection for the directed evolution of proteases in *E. coli* using a cytoplasmic anchoring strategy. *Biotechnol. Bioeng.* 113, 1187–1193.
- Chakrabortee, S., Liu, Y., Zhang, L., Matthews, H.R., Zhang, H., Pan, N., Cheng, C.R., Guan, S.H., Guo, D.A., Huang, Z., et al. (2012). Macromolecular and small-molecule modulation of intracellular Aβ 42 aggregation and associated toxicity. *Biochem. J.* 442, 507–515.
- Chen, Y., Li, Q., Yang, J., and Xie, H. (2017). Promoting tag removal of a MBP-fused integral membrane protein by TEV protease. *Appl. Biochem. Biotechnol.* 181, 939–947.
- Craik, C.S., Page, M.J., and Madison, E.L. (2011). Proteases as therapeutics. *Biochem. J.* 435, 1–16.
- Dahlgren, K.N., Manelli, A.M., Stine, W.B. Jr., Baker, L.K., Krafft, G.A., and LaDu, M.J. (2002). Oligomeric and fibrillar species of amyloid-β peptides differentially affect neuronal viability. *J. Biol. Chem.* 277, 32046–32053.
- Dickinson, B.C., Packer, M.S., Badran, A.H., and Liu, D.R. (2015). A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* 5, 1–8.
- Doucet, A. and Overall, C.M. (2008). Protease proteomics: revealing protease in vivo functions using systems biology approaches. *Mol. Aspects Med.* 29, 339–358.
- Dumont, J.A., Liu, T., Low, S.C., Zhang, X., Kamphaus, G., Sakorafas, P., Fraley, C., Drager, D., Reidy, T., McCue, J., et al. (2012). Prolonged activity of a recombinant factor VIII-Fc fusion protein in hemophilia A mice and dogs. *Blood* 119, 3024–3030.
- Giacalone, M.J., Gentile, A.M., Lovitt, B.T., Berkley, N.L., Gunderson, C.W., and Surber, M.W. (2006). Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *BioTechniques* 40, 355–364.
- Gray, D.C., Mahrus, S., and Wells, J.A. (2010). Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* 142, 637–646.
- Guerrero, J.L., Daugherty, P.S., and O'Malley, M.A. (2017). Emerging technologies for protease engineering: new tools to clear out disease. *Biotechnol. Bioeng.* 114, 33–38.
- Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121–4130.
- Hedstrom, L., Szilagyi, L., and Rutner, W.J. (1992). Converting trypsin to chymotrypsin: the role of surface loops. *Science* 255, 1249–1253.
- Hussein, R.M., Hashem, R.M., and Rashed, L.A. (2015). Evaluation of the amyloid beta-GFP fusion protein as a model of amyloid beta

- peptides-mediated aggregation: a study of DNAJB6 chaperone. *Front Mol Neurosci.* 8, 1–9.
- Kapust, R.B., Tözsér, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* 14, 993–1000.
- Kapust, R.B., Tözsér, J., Copeland, T.D., and Waugh, D.S. (2002). The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* 294, 949–955.
- Kim, W. and Hecht, M.H. (2008). Mutations enhance the aggregation propensity of the Alzheimer's A β peptide. *J. Mol. Biol.* 377, 565–574.
- Kim, W., Kim, Y., Min, J., Kim, D.J., Chang, Y.T., and Hecht, M.H. (2006). A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *ACS Chem. Biol.* 1, 461–469.
- Kostallas, G., Löfdahl, P.A., and Samuelson, P. (2011). Substrate profiling of tobacco etch virus protease using a novel fluorescence-assisted whole-cell assay. *PLoS One* 6, 1–10.
- Lee, C.C., Kuo, C.J., Ko, T.P., Hsu, M.F., Tsui, Y.C., Chang, S.C., Yang, S., Chen, S.J., Chen, H.C., Hsu, M.C., et al. (2009). Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. *J. Biol. Chem.* 284, 7646–7655.
- Lindberg, H., Sandersjö, L., Meister, S.W., Uhlén, M., Löfdahl, J., and Ståhl, S. (2017). Flow-cytometric screening of aggregation-inhibitors using a fluorescence-assisted intracellular method. *Biotechnol. J.* 12, 1–11.
- Longwell, C.K., Labanieh, L., and Cochran, J.R. (2017). High-throughput screening technologies for enzyme engineering. *Curr. Opin. Biotechnol.* 48, 196–202.
- Matthews, D.J. and Wells, J.A. (1993). Substrate phage: selection of protease substrates by monovalent phage display. *Science* 260, 1113–1117.
- Packer, M.S., Rees, H.A., and Liu, D.R. (2017). Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* 8, 1–11.
- Phan, J., Zdanov, A., Evdokimov, A.G., Tropea, J.E., Peters, H.K. 3rd, Kapust, R.B., Li, M., Wlodawer, A., and Waugh, D.S. (2002). Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* 277, 50564–50572.
- Ratnikov, B., Cieplak, P., and Smith, J.W. (2009). High throughput substrate phage display for protease profiling. *Methods Mol Biol.* 539, 1–21.
- Rawlings, N.D., Barrett, A.J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, D503–D509.
- Sandersjö, L., Jonsson, A., and Löfdahl, J. (2017). Protease substrate profiling using bacterial display of self-blocking affinity proteins and flow-cytometric sorting. *Biotechnol. J.* 12, 1–11.
- Ullah, R., Shah, M.A., Tufail, S., Ismat, F., Imran, M., Iqbal, M., Mirza, O., and Rahman, M. (2016). Correction: Activity of the human rhinovirus 3C protease studied in various buffers, additives and detergents solutions for recombinant protein production. *PLoS One* 11, 1–13.
- Van der Henst, C., Charlier, C., Deghelt, M., Wouters, J., Matroule, J.Y., Letesson, J.J., and De Bolle, X. (2010). Overproduced brucella abortus PdhS-mCherry forms soluble aggregates in *Escherichia coli*, partially associating with mobile foci of IbpA-YFP. *BMC Microbiol.* 10, 1–9.
- Vojcic, L., Pitzler, C., Körfer, G., Jakob, F., Martinez, R., Maurer, K.H., and Schwaneberg, U. (2015). Advances in protease engineering for laundry detergents. *N. Biotechnol.* 32, 629–634.
- Waugh, D.S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* 80, 283–293.
- Wurth, C., Guimard, N.K., and Hecht, M.H. (2002). Mutations that reduce aggregation of the Alzheimer's A β 42 peptide: an unbiased search for the sequence determinants of A β amyloidogenesis. *J. Mol. Biol.* 319, 1279–1290.
- Yi, L., Gebhard, M.C., Li, Q., Taft, J.M., Georgiou, G., and Iverson, B.L. (2013). Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc. Natl. Acad. Sci. USA* 110, 7229–7234.
- Zhao, H. and Feng, H. (2018). Engineering *Bacillus pumilus* alkaline serine protease to increase its low-temperature proteolytic activity by directed evolution. *BMC Biotechnol.* 18, 1–12.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/hsz-2018-0362>).