Journal of Chromatography A, 1337 (2014) 95–105

Contents lists available at ScienceDirect

Journal of Chromatography A

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

A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins

Steffen Frey*, Dirk Görlich*

Max-Planck-Institut für biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 20 December 2013
Received in revised form 4 February 2014
Accepted 7 February 2014
Available online 19 February 2014

Keywords:
Affinity tag
Tag-removing protease
Salt tolerance
P_1' promiscuity
SUMO
NEDD8

ABSTRACT

Engineered protein tags that confer specific binding to standardized affinity resins have revolutionized recombinant protein purification. Ideally, these tags should, however, be removed during or following purification to restore an authentic N-terminus. We introduce here a new set of proteases and corresponding protease recognition modules that are optimally suited for this purpose: a SUMO-specific and a NEDD8-specific protease from *Brachypodium distachyon* (bdSENP1 and bdNEDP1), the NEDP1 protease from *Salmo salar* (ssNEDP1), *Saccharomyces cerevisiae* Atg4p (scAtg4) and *Xenopus laevis* Usp2 (xlUsp2). These new proteases are highly specific and cleave tags from a 50-fold (xlUsp2) to 10,000-fold (bdSENP1) molar excess of substrate per hour at 0 °C. They are thus up to 1000-fold more active than TEV protease. The most efficient protease, bdSENP1, is even more active and far more salt tolerant than its yeast ortholog scUlp1, allowing efficient tag removal also in high salt buffers containing, e.g. 1 M NaCl. ssNEDP1 is distinguished by an exceptional salt tolerance, and a considerable tolerance toward charged and bulky residues in the P_1' position. xlUsp2 is unique in that it can restore, with low efficiency though, an N-terminal proline. As shown in the accompanying paper (S. Frey, D. Görlich, J. Chromatogr. A (2014), http://dx.doi.org/10.1016/j.chroma.2014.02.029), the orthogonality between bdSENP1, NEDP1, scAtg4 and xlUsp2 can be exploited for purifying multi-subunit protein complexes of defined stoichiometry.

1. Introduction

Purified proteins have gained considerable impact in modern biomedical research. As the traditional chromatographic methods for purification of proteins from their natural source are tedious and often lead to limited yield and purity, most of the proteins are nowadays produced as "recombinant" proteins in suitable host organisms [1]. In such systems, the DNA encoding the target protein is fused to foreign DNA elements. It can, e.g. be put under the control of a strong inducible promoter in order to over-express the target protein and hence allow for higher product yields [2,3]. Most importantly, however, recombinant expression systems can be used to modify the target protein, i.e. to introduce mutations, deletions or to genetically fuse the target protein with engineered "tags" [2,4–7]. Such tags often promote protein expression and solubility. Typically, they also mediate high-affinity binding to standardized affinity matrices and therefore allow for highly efficient and streamlined purification schemes [5,7,8].

In the ideal case, such tags can be removed from the target protein during the purification process and thereby allow production of target proteins lacking any unwanted extensions at their termini. This step is often accomplished by site-specific proteases recognizing a unique, short recognition motif that has been artificially introduced between the tag and the target protein. For this purpose, commercial suppliers offer various proteases, e.g. Thrombin, Factor Xa, enterokinase, or the 3C proteases from *Tobacco etch virus* (TEV) or human rhinovirus [7,9].

In practice, the application of these proteases is often hampered by inefficient cleavage, a requirement for elevated temperature during cleavage, pronounced preferences for certain amino acids in the P_1' position (the position after the scissile bond) or a narrow optimum for buffer and/or salt conditions. Also, most of these proteases leave unwanted residues at the N-terminus of the target protein [9]. In addition, the limited specificity of some commonly used proteases (e.g. thrombin) might lead to the degradation of sensitive target proteins.

Recently, an alternative system utilizing the *Saccharomyces cerevisiae* SUMO-specific protease Ulp1p (scUlp1) has been introduced [10–12]. This protease can be used to specifically

* Corresponding authors at: Max-Planck-Institut für biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany. Tel.: +49 5512012460.
(S. Frey)/ +49 5512012401 (D. Görlich)
E-mail addresses: sfrey@gwdg.de (S. Frey), goerlich@mpibpc.mpg.de (D. Görlich).

http://dx.doi.org/10.1016/j.chroma.2014.02.029
0021-9673

cleave off target proteins fused to the C-terminus of yeast Smt3p (scSUMO).

The scUlp1 protease cleaves scSUMO-containing substrates also in its cellular context. SUMO (small ubiquitin-related modifier) can be covalently attached to numerous acceptor proteins, whereby an isopeptide bond is formed between SUMO's C-terminus and a lysine ε-amino group on the acceptor protein [13]. The SUMO pathway relies on two scUlp1-mediated proteolytic events: The enzyme removes a C-terminal extension from the scSUMO precursor protein, thereby creating the characteristic C-terminal Gly–Gly motif that is required for SUMO-conjugation. In addition, scUlp1 cleaves isopeptide bonds between scSUMO and acceptor proteins, and thereby reverses scSUMO modifications.

The standard nomenclature [14] describes the context of a protease cleavage site by listing the residues preceding (P_N ... P_2, P_1) and following (P_1', P_2'...P_N') the scissile (P_1–P_1') bond. scUlp1 is more specific than a conventional protease, because it recognizes not just a short peptide motif, but also the folded SUMO domain including the C-terminal Gly–Gly motif [15]. It cleaves C-terminally of the invariant Gly–Gly motif, i.e. scUlp1 accepts only Gly as P_1 and P_2 residues. Regarding P_1', however, scUlp1 is remarkably promiscuous: It tolerates 19 of the possible 20 P_1' residues [12]; it can thus restore an authentic N-terminus in many cases, but not an N-terminal proline or an acetylated N-terminus.

SUMO is just one representative of a larger group of paralogous eukaryotic modifiers that also includes ubiquitin (Ub), NEDD8, and Atg8 [16,17]. While SUMO, ubiquitin and NEDD8 possess a characteristic double-glycine (GG) motif at their mature C-termini, Atg8 proteins feature the sequence Phe–Gly (FG) at the corresponding position [17]. Despite these differences, these modifiers share a common fold, a similar conjugation mechanism, and they are similarly processed and deconjugated by dedicated proteases [17]. We therefore speculated that further proteases of this type could potentially be used for the removal of engineered tags from recombinant proteins.

To minimize the loss of activity of a given target protein during its purification because of thermal denaturation or buffer incompatibility, purifications are preferably done at low temperature (i.e. 0–4 °C) and in buffers optimal for the target protein. Therefore, proteases used for tag removal should be able to efficiently cleave their target proteins also at low temperature and within a wide range of ionic conditions.

Here, we characterize new proteases ideally matching these criteria: bdSENP1 and bdNEDP1 from *Brachypodium distachyon* (bd) as well as ssNEDP1 from salmon (*Salmo salar*, ss) (Table 1). We further show that the yeast Atg8-specific protease Atg4 and the *Xenopus laevis* ubiquitin-specific protease Usp2 can also be used for efficient removal of tags from recombinant proteins. Comparing the new proteases to both, yeast scUlp1 and a stabilized variant of TEV protease, bdSENP1 was found to be possibly one of the most efficient and versatile proteases characterized for tag removal so far.

Table 1
Nomenclature of substrates and proteases used in this study.

Organism	Substrate	Protease
Saccharomyces cerevisiae	scSUMO (Smt3p)	scUlp1 (Ulp1p)
Brachypodium distachyon	bdSUMO	bdSENP1
Brachypodium distachyon	bdNEDD8	bdNEDP1
Salmo salar	ssNEDD8	ssNEDP1
Saccharomyces cerevisiae	scAtg8 (Atg8p)	scAtg4 (Atg4p)
Xenopus laevis	xlUb	xlUsp2
Tobacco etch virus	TEV site	TEV protease

Abbreviations: sc: *Saccharomyces cerevisiae*; bd: *Brachypodium distachyon*; ss: *Salmo salar*; xl: *Xenopus laevis*.

2. Materials and methods

2.1. Identification and assembly of proteases and substrates

Iterative BLAST searches were performed on EST and genomic databases using the known human and yeast SUMO- and SENP1 orthologs as templates. Full-length sequences of the primary hits were assembled from multiple overlapping clones and used as input sequences for further BLAST searches. Similarly, NEDD8- and NEDP1 orthologs were identified.

2.2. Structure-based alignment and structure evaluation

Folded protein domains were subjected to structure-based sequence alignments using Expresso [18] and the following templates: hsSUMO1: 2G4D_B [19]; hsSUMO2/bdSUMO: 2CKH_B [20]; scSUMO: 1EUV_B [15]; hsSENP1/bdSENP1: 2CKH_A [20]; scUlp1: 1EUV_A [15]; all NEDD8 orthologs: 2BKR_B [21]; all NEDP1 orthologs: 2BKR_A [21].

Interacting residues (<4.6 Å center–center distance between the nearest atoms of neighboring residues) were computed using MacPyMol (DeLano Scientific). To annotate SUMO-interacting surface of hsSENP1, the structure of its complex with hsSUMO2 (accession number 2CKH [20]) was analyzed.

2.3. Expression of proteases and substrates

All recombinant proteins were expressed in *Escherichia coli* from appropriate expression vectors (see Tables S1 and S2) and purified via an engineered polyHis-tag using standard Ni^{2+} chelate chromatography with imidazole elution. If required, proteases were further purified by gel filtration on a Superdex200 column (GE Healthcare). Untagged proteases were obtained by cleaving the polyHis-tag with an appropriate His-tagged protease. After a gel filtration step, remaining traces of cleaved polyHis-tag and polyHis-tagged protease were removed by "reverse Ni^{2+} chelate" chromatography. This guaranteed the final enzyme preparation to be free of any contaminating proteolytic activity. A specific purification example is given in Fig. 2. All proteins were quantified via their absorption at 280 nm and computed extinction coefficients. Accuracy of quantification and purity of the proteins were validated by SDS–PAGE followed by Commassie-staining. Protein and plasmid sequences are available on request.

2.4. Protease and substrate truncations

In order to narrow down the minimal domains required for proteolytic cleavage, a detailed truncation analysis was performed for the bdSUMO/bdSENP1 and bdNEDD8/bdNEDP1 substrate/protease pairs (Figs. S1–S3). The presumably unstructured N-terminal extensions of both bdSENP1 (amino acids 1–247) and its substrate bdSUMO (amino acids 1–20) could be deleted without affecting cleavage efficiency (Fig. S1 and S2). As in the case of bdSUMO this extension may contribute to its solubility-enhancing effect (S.F. and D.G., manuscript in preparation), our standard substrates contain $bdSUMO^{2–97}$ (including the N-terminal extension). In contrast, if not stated otherwise, all experiments were performed with the catalytically active $bdSENP1^{242–481}$ fragment. As the bdNEDP1 protease and its substrate bdNEDD8 do not possess comparable extensions, already minor N-terminal truncations within both proteins significantly reduced the cleavage efficiency (Fig. S3). Deletion of more than 23 amino acids from the C-terminus of the bdNEDP1 protease resulted in a drastic loss of protease activity.

For all direct comparisons to other proteases, we used a solubility-enhanced and autocleavage-resistant variant of TEV protease (TEV(SH) [22]) lacking the C-terminal autoinhibitory peptide

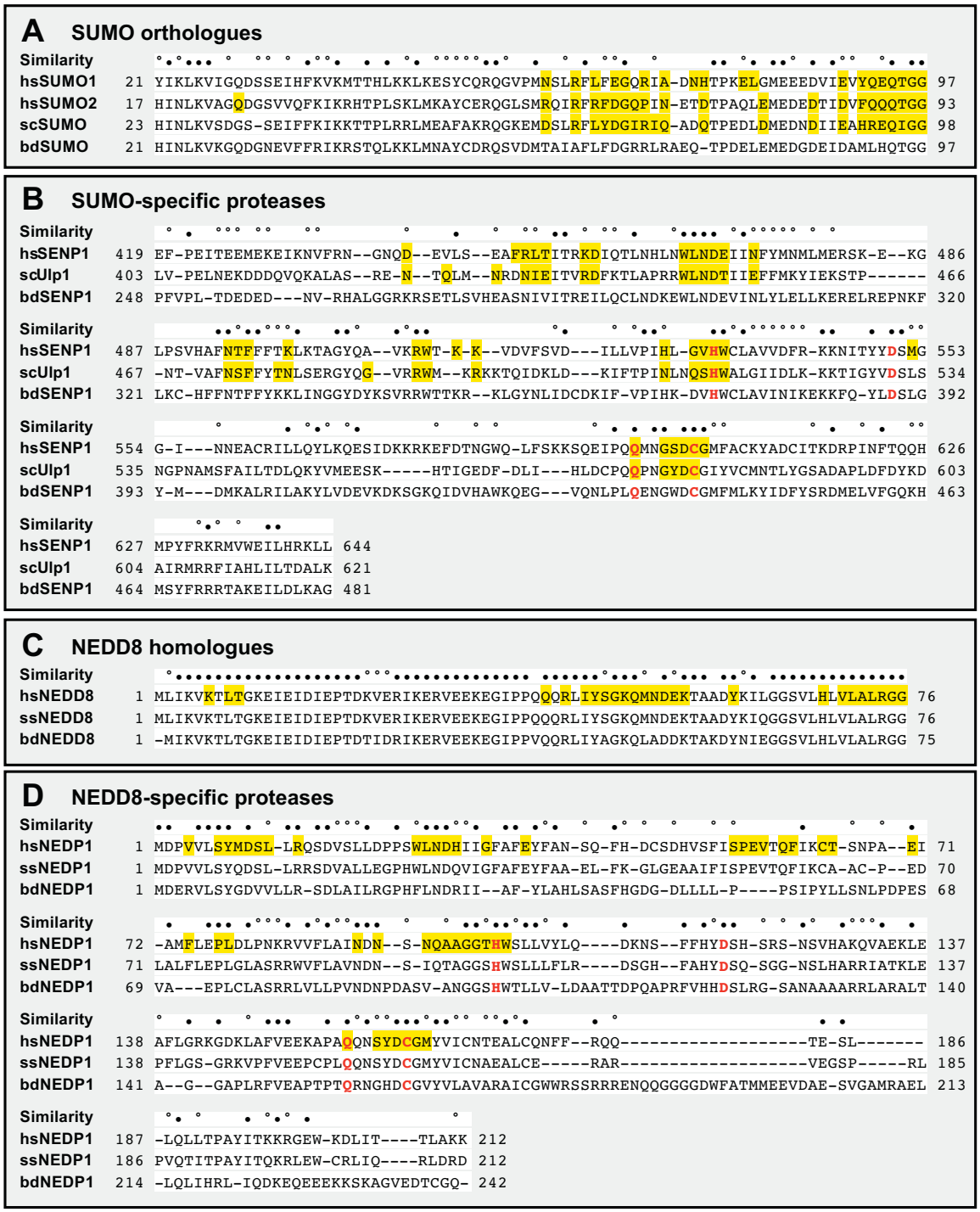

Fig. 1. Structure-based sequence alignment of SUMO- and NEDD8-orthologs and SUMO/NEDD8-specific proteases with their human orthologs. Relevant protein sequences were assembled from available EST and genomic sequence and aligned based on the results obtained from the Expresso server [18]. Residue conservation at each position was classified as similar (◦) or identical (•). Amino acids near the interface to the respective binding partner are highlighted in yellow boxes. Residues directly involved in peptide bond hydrolysis are marked in bold. No structures were available for bdSUMO•bdSENP1, bdNEDD8•bdNEDP1 and ssNEDD8•ssNEDP1.

ELVYSQ [23]. The catalytic activity of this protease is indistinguishable from the parental full-length enzyme (Fig. S4).

2.5. Solution cleavage assay

If not stated otherwise, cleavage reactions were performed in LS-S buffer (250 mM NaCl, 40 mM Tris/HCl pH7.5, 2 mM $MgCl_2$, 250 mM sucrose, 2 mM DTT). Generally, substrates and proteases were pre-diluted in LS-S buffer to twice the aspired end-concentration. Cleavage was initiated by mixing identical volumes of substrate and protease pre-dilutions and stopped by adding 9 volumes of hot SDS sample buffer. A fraction corresponding to 2.5 μg of substrate was separated by SDS-PAGE on 7–15% gradient gels. Gels were stained with Coomassie G250 and scanned.

3. Results

3.1. New SENP1-like and NEDP1 proteases intended for selective tag removal

The primary motivation of our study was to introduce new proteases for the specific cleavage of tags from recombinant proteins. We aimed at proteases with an extraordinary specificity and therefore restricted our search to proteases involved in processing and deconjugation of ubiquitin-like modifiers. We wanted these proteases also to be stable, denaturation-resistant and yet highly active at low temperature. We therefore focused pragmatically on species that tolerate a wide temperature range, namely poikilothermic vertebrates (fish and amphibians) as well as temperate grasses, like wheat, which survives temperatures from −17 °C to +47 °C, before protein denaturation causes lethal damage [24].

By iterative homology searches and assembling cDNA as well as genomic sequences, we identified the so-far non-annotated putative SENP1 and NEDP1 proteases from the grass *B. distachyon* (purple false brome, a relative of wheat) and the putative NEDP1 ortholog from *S. salar* (salmon). Fig. 1 shows a structure-based sequence alignment of the respective protease sequences with their putative human orthologs.

The two NEDP1 enzymes and the catalytic domain of the bdSENP1 (residues 242–481) were then cloned as polyHis-tag expression constructs, over-expressed in *E. coli* and purified on a Ni^{2+} chelate resin. For some applications, the polyHis-tag was proteolytically removed before a polishing step on a Superdex200 column. The catalytic domain (amino acids 403–621) of the scSUMO-cleaving protease scUlp1 [12] and a stabilized version of TEV protease (TEV(SH) [22]) lacking the C-terminal auto-inhibitory peptide [23] were chosen as reference proteases and purified analogously. All these proteases were perfectly soluble and could be purified with excellent yields. As an example, the preparation of tag-free bdSENP1 is detailed in Fig. 2.

3.2. scAtg4 and xlUsp2: Candidates for cleaving Atg8 and ubiquitin tags, respectively

As further examples of proteases involved in processing of ubiquitin-like modifiers, we expressed and purified the *S. cerevisiae* Atg8-specific protease Atg4 (scAtg4; [25]) as well as the catalytic domain (amino acids 43–383) of the *X. laevis* (xl) ubiquitin-specific protease Usp2 (xlUsp2). These two enzymes were not as well behaved as the SENP1 and NEDP1 protease mentioned above, i.e. their large-scale production requires further optimization. Nevertheless, we could obtain them in sufficient scale and quality to carry out comprehensive test series.

3.3. Design of protease substrates for cleavage assays in solution

We wanted to compare the new proteases with scUlp1 and the TEV protease directly in terms of substrate processing. We therefore designed SUMO- and NEDD8-, Atg8- and Ub-containing substrates according to a common scheme that ensured an identical sequence context (and thus similar accessibility) of the scissile bond (Fig. 3A): The scissile bond was followed by the tri-peptide Ala–Gly–Thr and the target protein (*E. coli* maltose binding protein, MBP). Within the TEV protease substrate, the sequence after the scissile bond was Gly–Thr, in agreement with the natural and preferred TEV recognition sequence [26,27]. Here, a 14 kDa ZZ-tag [28] was fused to the N-terminus of the protease recognition module to allow a clear electrophoretic discrimination between full length and cleaved substrate. All substrate proteins contained an N-terminal polyHis-tag for efficient purification on a Ni^{2+} chelate resin.

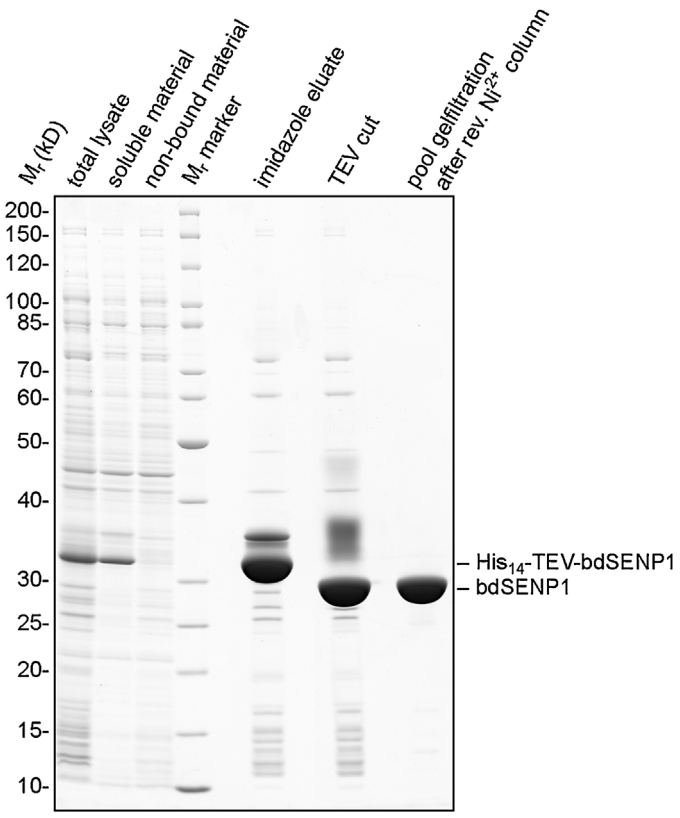

Fig. 2. Purification of bdSENP1. *E. coli* strain NEB Express (MiniF LacIq) transformed with pSF1389 encoding His_{14}-TEV-bdSENP1 was grown at 37 °C in TB-medium. At OD_{600} = 2.0 protein expression was induced with 0.2 mM IPTG and the temperature reduced to 18 °C. Cultures were further shaken over night at 18 °C. The culture was treated with 5 mM EDTA. Cells were harvested by centrifugation, resuspended in LS-ID buffer (290 mM NaCl, 45 mM Tris–HCl pH7.5, 4.5 mM $MgCl_2$, 10 mM DTT, 15 mM imidazole) and lysed by sonication. The total lysate was cleared by centrifugation (1 h, 200,000*g*). The supernatant (soluble material) was incubated in batch mode with a Ni^{2+} chelate resin for 1 h at 4 °C. After washing off non-bound proteins with LS-ID buffer, the target protein was eluted with LS-ID buffer containing 400 mM imidazole (imidazole eluate). The His_{14}-TEV tag was cleaved off using a 1/30 molar ratio of polyHis-tagged TEV protease over night at room temperature (TEV cut). The complete reaction was subjected to gel filtration on a Superdex200 26/60 column (GE healthcare) equilibrated with degassed LS buffer (290 mM NaCl, 45 mM Tris–HCl pH7.5, 4.5 mM $MgCl_2$) supplemented with 5 mM DTT. TEV protease co-eluting with bdSENP1 was removed by passing the pooled peak fractions twice over a Ni^{2+} chelate resin. Without further optimization, this procedure yielded 63 mg of bdSENP1 per liter of *E. coli* culture, enough to cleave >1 kg of a 50 kD target protein within 1 h at 0 °C.

3.4. All new proteases are highly efficient tools for tag removal

To compare the cleavage efficiencies of the new proteases with scUlp1 and TEV protease, we incubated a fixed concentration (100 μM) of each substrate protein with a wide concentration range of the corresponding proteases (Fig. 3B and C). These initial tests revealed that already ≈15–50 nM of scUlp1 or bdSENP1 efficiently (≥95%) cleaved the corresponding substrate proteins within 1 h at 0 °C. At 25 °C, efficient cleavage required no more than 10 nM of either protease (Fig. 3C). At either temperature, the two NEDP1 enzymes and scAtg4 required an ≈10- to 15-fold, xlUsp2 even an ≈30- to 50-fold higher concentration for a similarly efficient cleavage of their substrates. Strikingly–but consistent with the reported low specific activities of commercial "TEV protease" (Sigma-Aldrich No. T4455) and "TurboTEV" (Nacalai USA No. NU0102) preparations–10 μM TEV protease was insufficient to cleave more than ≈40% of substrate protein within 1 h at 0 °C. At 25 °C, ≈2–3 μM of TEV protease was required for efficient cleavage.

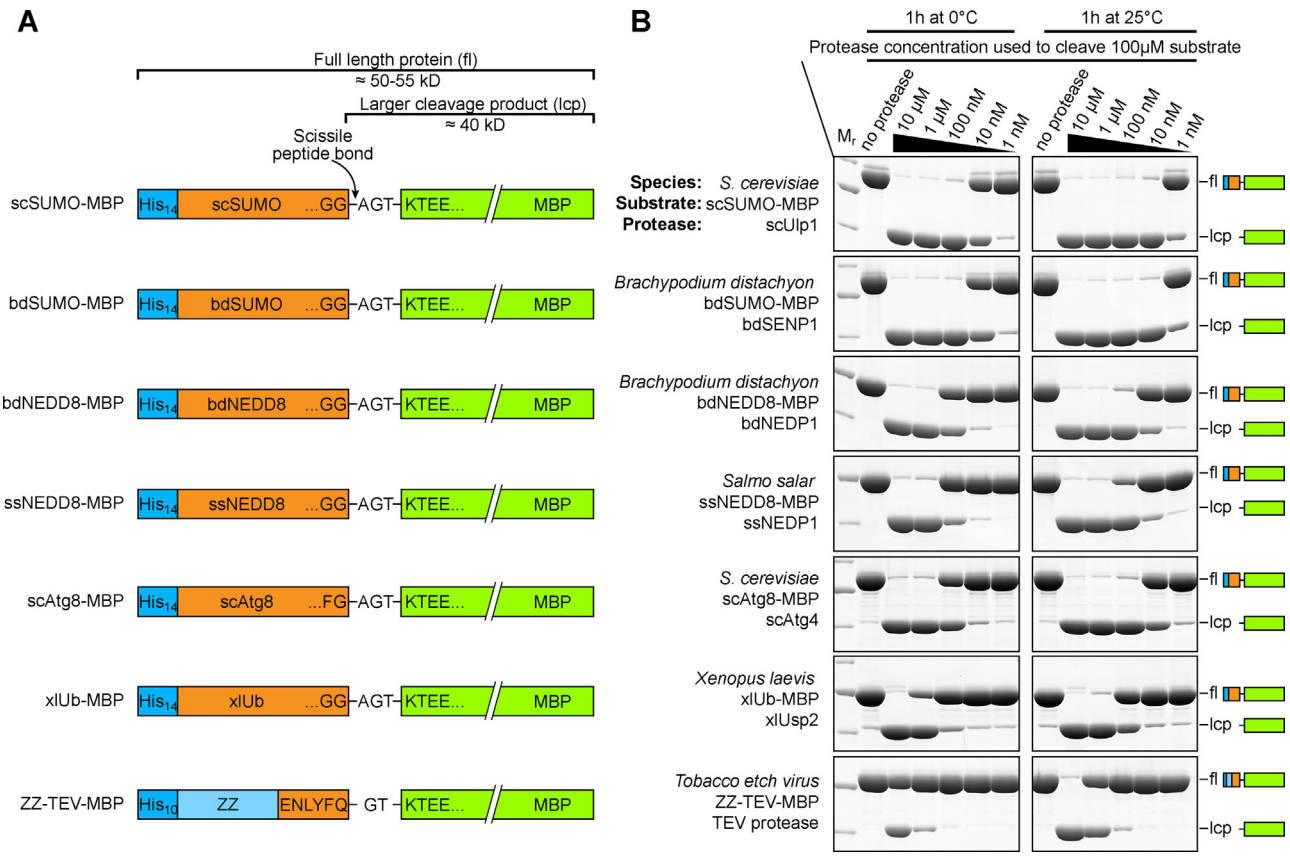

Fig. 3. Activity of tag-cleaving proteases. (A) General design of protease substrates. All substrates contain an N-terminal polyHis-tag (His_{14} or His_{10}), a protease recognition module (box left of the scissile bond) and the target protein MBP (maltose-binding protein; MBP). To ensure equivalent cleavage conditions, identical sequences had been placed behind the scissile bond of SUMO-, NEDD8-, scAtg8 and xlUb-containing substrates. (B) Protease titration. Protease substrates (100 μM) sketched in (A) were incubated for 1 h at 0 °C (left) or at 25 °C (right) in the presence of the corresponding proteases. Proteases were titrated down from 10 μM to 1 nM. Reactions were stopped by dilution in hot SDS sample buffer. Cleavage products were separated by SDS–PAGE and stained with Coomassie G250. For the sake of clarity and space economy, only a section of the gel is depicted, showing non-cut (full length) proteins (fl) and the larger cleavage products (lcp). Bands of the molecular weight marker (M_r) correspond to 40 kD, 50 kD (more intense band) and 60 kD (not always visible). Complete gels for representative experiments are shown as Fig. S5.

3.5. *The Brachypodium SENP1 enzyme is the most efficient tag-cleaving protease tested*

For a more detailed comparison of cleavage efficiencies, we further analyzed the cleavage kinetics at 0 °C for each cognate substrate•protease pair (Fig. 4A). To facilitate a qualitative comparison, orthologous protease pairs were tested at identical concentrations (30 nM for SUMO-specific proteases, 300 nM for NEDP1 enzymes). This setup revealed even subtle differences between similar proteases: While digestion was ≈90% complete after 30 min with 30 nM of bdSENP1, a comparably complete digestion with scUlp1 took more than twice as long. Similarly, at 300 nM concentration, bdNEDP1 cleaved its substrate ≈2-times faster than the corresponding salmon enzyme. In this assay, the activity of scAtg4 was comparable to bdNEDP1. The xlUsp2 enzyme required about 3-fold higher concentration to process its substrate at the same rate as ssNEDP1. At 0 °C, even 20 μM TEV protease was insufficient to process 100 μM of substrate (i.e. a 5-fold molar excess) within 4 h.

These results classified bdSENP1 as the most efficient tag-cleaving protease tested. Taking into account the different protease concentrations and cleavage kinetics, this direct comparison indicated that, at 0 °C, the two NEDP1 proteases and scAtg4 are ≈10–15 times less active than the SUMO-specific enzymes, but still, they outperform TEV protease by a factor of ≈150–300. The least active of the newly characterized proteases, xlUsp2, has still a >20-fold higher turnover than TEV protease.

3.6. *The SUMO-, NEDD8- and Atg8-specific proteases are highly active between 0 °C and 37 °C*

The temperature dependence of protease activity was analyzed by incubating a fixed concentration of substrate proteins at various temperatures with a limiting amount of the respective proteases (Fig. 4B). As expected, the cleavage efficiency increased from 0 °C to 37 °C for all substrate/protease pairs. Also in this assay, bdSENP1 performed better than its yeast ortholog and consistently showed a more efficient cleavage of its substrate at all temperatures. In a direct comparison of the two NEDP1 enzymes, the *Brachypodium* enzyme was more active than its salmon counterpart between 0 °C and 25 °C, and overall showed a similar temperature dependence as scAtg4. The strongest preference for higher temperatures was observed for xlUsp2. Although it cleaved its substrate also at 0 °C, this enzyme was the only one that performed significantly better at 37 °C than at 25 °C. In this assay, TEV protease was at all temperatures at least 10-fold less efficient than any of the other proteases tested. Thus, while the SUMO-, NEDD8- and Atg8-specific proteases tested here can be used for efficient tag removal at 0 °C, TEV protease needs higher temperatures, more enzyme and/or significantly longer incubation times for similar results. The xlUsp2 enzyme is

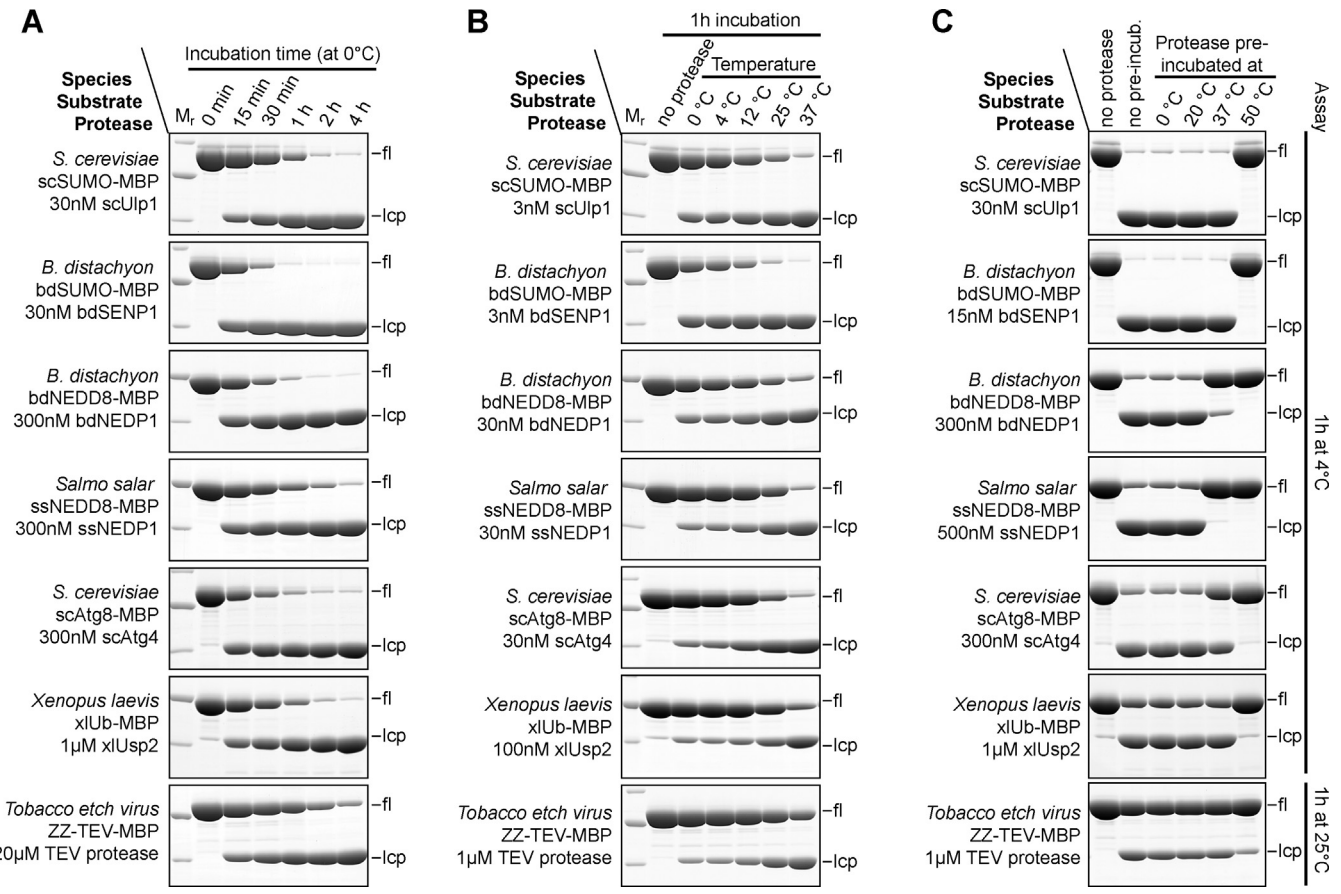

Fig. 4. Cleavage kinetics, temperature dependence and temperature stability of tag-cleaving proteases. (A) Cleavage kinetics. 100 μM protease substrates were incubated at 0 °C with corresponding proteases at indicated concentrations. Samples were taken at indicated time points. (B) Temperature dependence. Proteases were allowed to cleave their respective substrates for 1 h at various temperatures. Note that in comparison to (A), the protease concentration had been significantly reduced. Also note that bdSENP1 is consistently more active than scUlp1 while bdNEDP1 outperforms ssNEDP1 at lower temperatures. (C) Long-term temperature stability. All proteases were pre-incubated for 16 h at indicated temperatures under argon to exclude oxygen that could potentially damage the active site cysteines. Thereafter, the remaining activity was assayed by treating 100 μM of substrate with indicated concentrations of the corresponding protease for 1 h at 0 °C (SENP1 and NEDP1 enzymes, scAtg4 and xlUsp2) or 25 °C (TEV protease).

also a good alternative to TEV protease, but exploits its full potential only at higher temperatures.

3.7. Long-term stability of tag-cleaving proteases

If desired, increasing the reaction temperature or the incubation times will generally reduce the amount of protease necessary for tag removal. This, however, requires the protease to be stable under these conditions. To test for long-term stability, the activities of all five proteases were assayed after over-night incubation at 0 °C, 20 °C, 37 °C or 50 °C in the absence of oxygen (Fig. 4C). The two SUMO proteases, xlUsp2 as well as TEV protease sustained temperatures up to 37 °C for 16 h without visibly loosing activity. In contrast, the two NEDP1 enzymes and scAtg4 significantly lost activity upon pre-incubation at 37 °C, but remained fully active after incubation at 20 °C. These results indicate that all tested enzymes are fully compatible with over-night incubations at room temperature or below. In the case of sufficiently stable target proteins, all tested enzymes except for TEV protease can be heat inactivated at 50 °C.

3.8. ssNEDP1 and bdSENP1 are far more tolerant to high salt than scUlp1

In order to test for salt tolerance, protease activities were assayed at NaCl concentrations ranging from 200 mM up to 1 M (Fig. 5A). While TEV protease and ssNEDP1 remained fully active in 1 M NaCl, the bdNEDP1 activity decreased by ≈30% and the xlUsp2 activity by ≈50%. bdSENP1 cleaved its substrate efficiently between 200 mM and 750 mM NaCl, but 1 M NaCl was already slightly inhibitory. The most striking salt sensitivity was evident for scUlp1 and scAtg4: the two enzymes were strongly inhibited by 1 M NaCl. A detailed time course indicated that scUlp1 cleaved its substrate at 1 M NaCl ≈50 times slower than at 250 mM NaCl (Fig. 5B). In comparison, the salt-induced kinetic inhibition was only ≈3-fold for bdSENP1 (Fig. 5C).

3.9. P_1' sensitivity

A number of proteases show sensitivity toward the P_1' position of the substrate's (P_1–P_1') scissile bond (see e.g. [9,12,26]). In such cases, the P_1' residue may be regarded to be part of the recognition sequence.

In order to probe the sensitivities of the here analyzed SUMO-, NEDD8-, scAtg8- and xlUb-specific proteases toward variations in the P_1' position, the standard substrates ($P_1'_{Ala}$; Fig. 3A) were compared to substrates containing the non-preferred P_1' residues Met, Tyr, Glu, Arg or Pro (Fig. 6A). Methionine was included, because it is the P_1' residue when the authentic N-terminus of a target protein, whose start methionine is neither cleaved nor acetylated, needs to be restored. Tyrosine represents a non-related bulky hydrophobic residue while glutamic acid and arginine are examples for charged

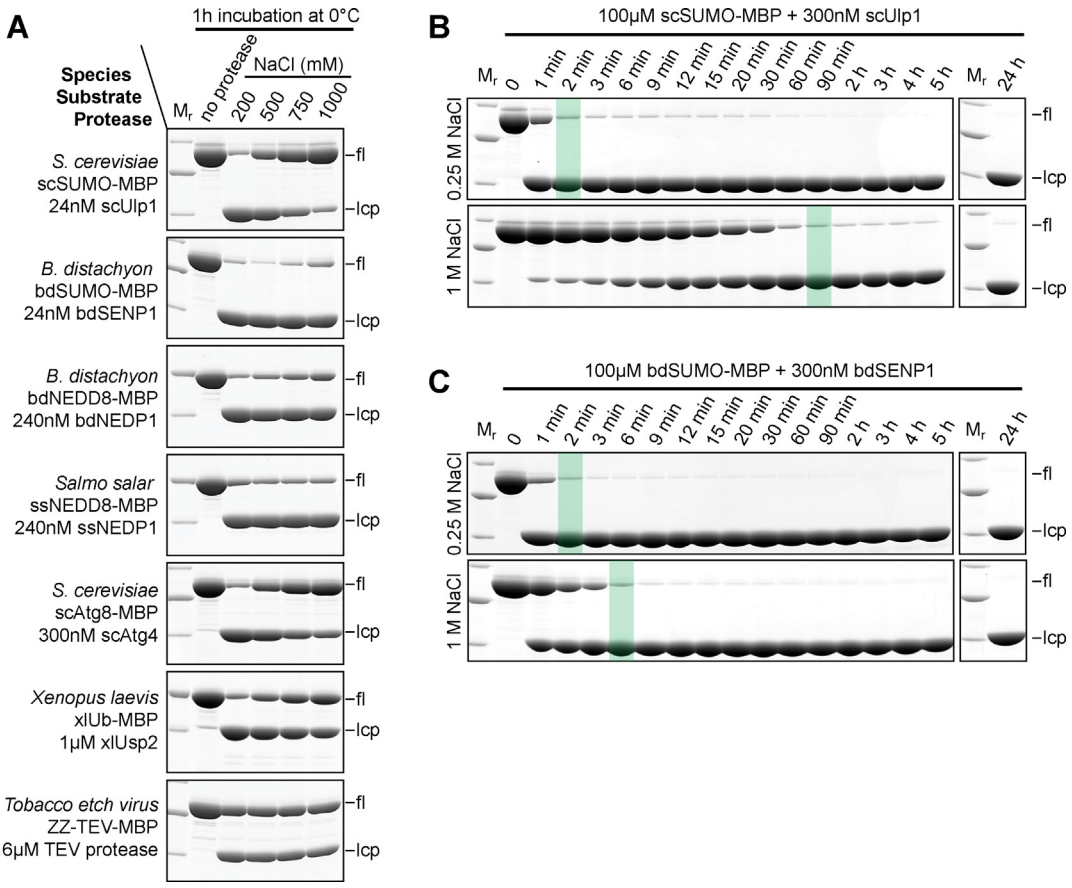

Fig. 5. Salt sensitivity of tag-cleaving proteases. (A) Substrates were incubated for 1 h at 0 °C with the corresponding proteases at various NaCl concentrations. Strikingly, scUlp1 and scAtg4 show a pronounced sensitivity to high NaCl concentrations. (B) and (C) scSUMO- (B) or bdSUMO- (C) containing substrates were incubated at 0 °C in the presence of 250 mM (upper panels) or 1 M NaCl (lower panels) with 300 nM of their corresponding protease. Samples were taken after various time points and analyzed by SDS-PAGE. Green bars mark lanes with efficient digestion of cognate protease/substrate pairs. Note that scUlp1 needed ≈50-times longer to digest 95% of its cognate substrate in the presence of 1 M NaCl as compared to 250 mM NaCl. In contrast, 1 M NaCl lowered the activity of bdSENP1 only by a factor of ≈3.

residues. Lastly, we also tested the performance of all proteases on substrates with a P_1' proline, a residue known to be problematic for most proteases (see e.g. [12,26]).

As expected, most SUMO- and NEDD8-containing substrates with other residues than Ala in the P_1' position required higher protease concentrations for efficient cleavage (Fig. 6B and C). These effects were generally moderate (≈3–10-fold) for substrates with Met, Tyr and Arg in the P_1' position. An up to 30-fold higher protease concentration was, however, required for $P_1'_{Glu}$ substrates. A remarkable exception is ssNEDP1, which cleaved all tested substrates (except for the $P_1'_{Pro}$ substrate) with comparable efficiency (Fig. 6C, right panel).

In contrast to published data [26], substrate cleavage by TEV protease was virtually unaffected by the $P_1'_{Met}$ or $P_1'_{Tyr}$ mutations (Fig. 6F). The $P_1'_{Arg}$ substrate, however, required a ≈3-fold higher protease concentration. The $P_1'_{Glu}$ mutant required an even 10-fold higher concentration, i.e. a 1:1 molar substrate: protease ratio was needed to cleave the substrate efficiently within 1 h at 25 °C.

A low sensitivity for residues in the P_1' position was observed for scAtg4 (Fig. 6D). At 25 °C, 1 μM protease was generally sufficient to process an 80–100-fold excess of all tested P_1' variants, except for $P_1'_{Pro}$.

With regard to the P_1' position, xlUsp2 was remarkably promiscuous and processed substrates with Ala, Met, Tyr, Arg or Glu in the P_1' position equally well (Fig. 6E). Surprisingly, and as the only protease tested here, xlUsp2 could even process a $P_1'_{Pro}$ substrate—although with significantly reduced efficiency.

3.10. SUMO-specific proteases and NEDP1 enzymes show very high turnover rates also at limiting substrate concentrations

A comprehensive kinetic description of the here described substrate/protease systems would require knowledge of three parameters for each substrate/protease pair and incubation condition, namely k_{cat} (the maximum turnover rate of the enzyme), K_M (the Michaelis–Menten constant describing enzyme saturation by the substrate [29]) as well $K_{Product}$ (accounting for end product inhibition). Such level of detail is beyond the scope of this study, nevertheless, we wanted to develop guidelines as to what concentration of protease is needed for quantitative tag cleavage when substrate type, substrate concentration, incubation time and temperature as well as ionic strength of the buffer are given.

For that, we analyzed the effect of substrate concentration on the protease activity in two slightly different setups (Fig. 7). In the first one, the initial concentrations of both, substrate and protease were varied proportionally, keeping the initial substrate/protease ratio constant (Fig. 7A). In a second assay, the substrate concentration was varied at a constant absolute protease concentration (Fig. 7B). For these experiments, the standard substrates (see Fig. 3A) were used.

These assays showed that the two SUMO-specific proteases can cleave a >10,000-fold excess of substrate within 1 h at 0 °C. Remarkably, the number of substrate molecules cleaved per molecule of protease remained largely constant when the concentrations of both, substrate and protease, were reduced proportionally to each

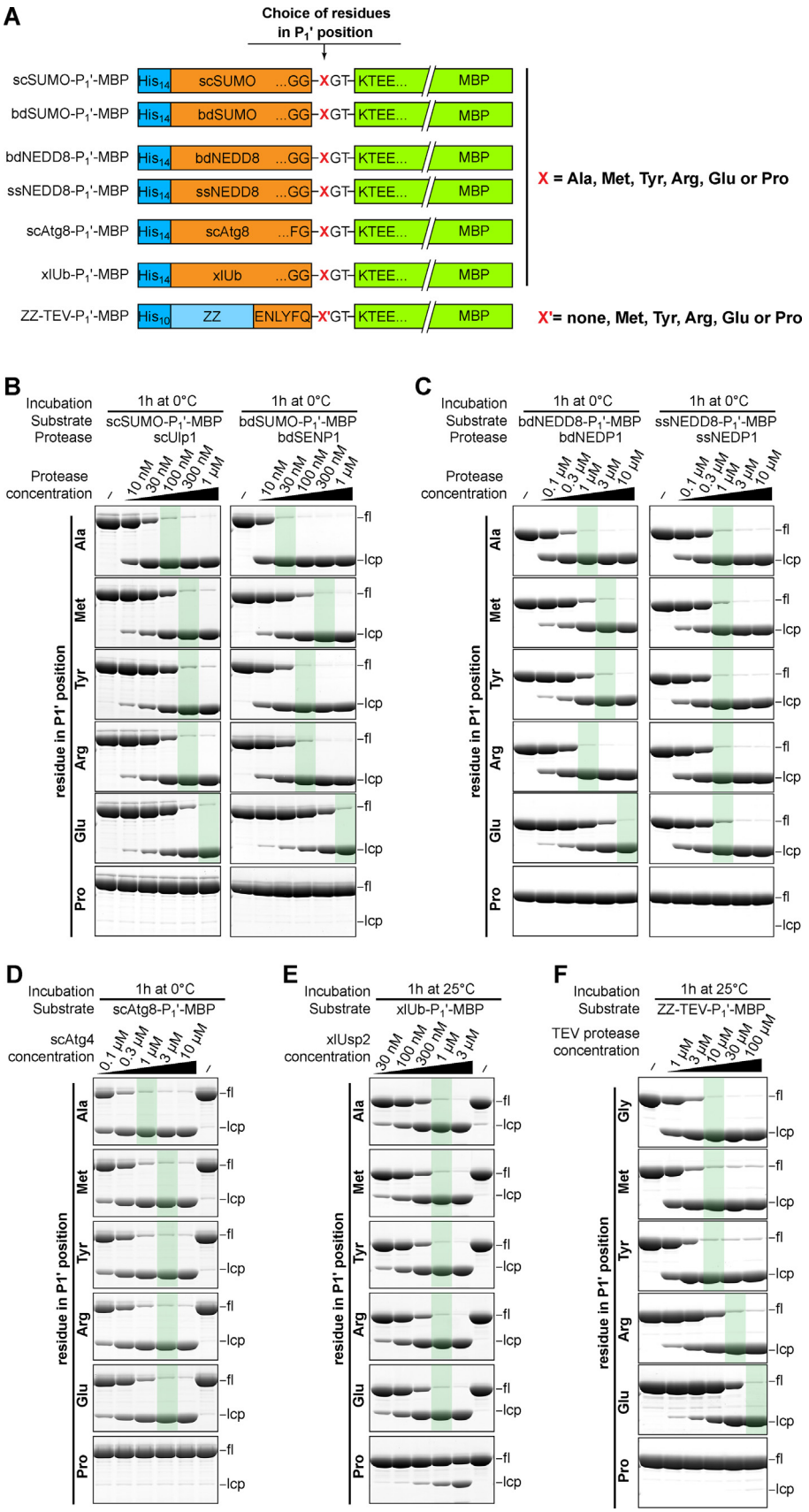

Fig. 6. P_1' preference of tag-cleaving proteases. (A) Scheme of protease substrates used. They follow the general outline shown in Fig. 3A, however, the P_1' position of the P_1–P_1' scissile bond had been mutated to the potentially non-preferred residues methionine (Met), tyrosine (Tyr), arginine (Arg), glutamic acid (Glu), or proline (Pro). For TEV substrates, the respective residues had been inserted in front of the original glycine residue. (B)–(F) Protease substrates with different P_1' residues were incubated for 1 h with various concentrations of their dedicated proteases. Cleavage reactions were performed at 0 °C for the SUMO-, NEDD8- and Atg8-specific proteases. Incubations with TEV protease or xlUsp2 were done at 25 °C. To facilitate a direct comparison between orthologous pairs of SUMO- and NEDD8-specific proteases, identical

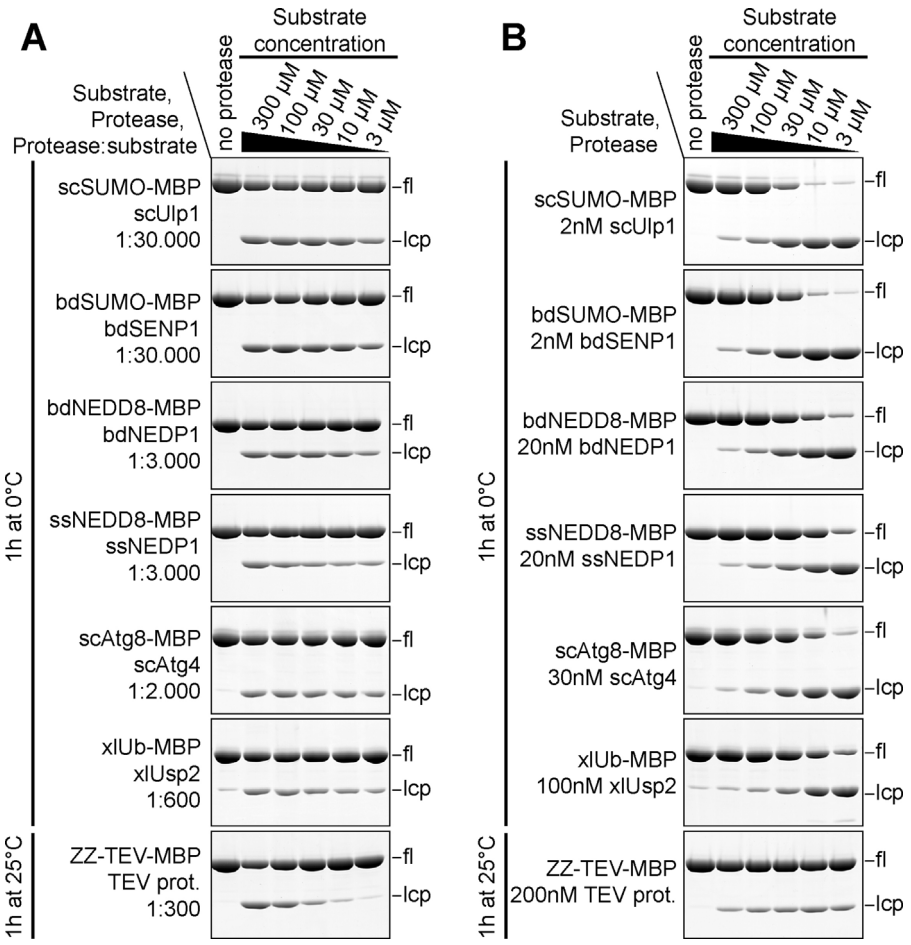

Fig. 7. Digestion efficiency at different substrate concentrations. In two different setups, indicated concentrations of substrate were incubated with their corresponding protease. A fraction of each reaction corresponding to 1.2 μg (≈20 pmol) of substrate protein was analyzed by SDS-PAGE and Coomassie staining. Due to the variable substrate concentration, the absolute volume of the digestion reaction analyzed by SDS-PAGE was inversely proportional to the substrate concentration in the digestion reaction. (A) The concentrations of both substrate and protease were titrated at constant protease: substrate ratio. Note that the cleavage efficiency remained similar for SUMO- and NEDD8-, scAtg8- and xlUb- specific proteases also at low protein concentrations, while TEV protease showed very inefficient turn over at low substrate concentrations. (B) The substrate concentration was varied from 300 μM to 3 μM while keeping the protease concentration constant. In this setup, the fraction of substrate cleaved by the TEV protease remains rather constant. For all other proteases assayed, the share of cleaved substrate increases at lower substrate concentrations (i.e. at higher protease: substrate ratio).

other, down to a substrate concentration of 3 μM (Fig. 7A). These results indicated that the K_M for the reaction is low, probably in the lower single-digit micromolar range. In line with this interpretation, the completeness of digestion drastically increased if the concentration of substrate was reduced at a constant protease concentration (Fig. 7B). The two NEDD8-specific proteases as well as scAtg4 and xlUsp2 showed a similar general behavior with low K_M, although the maximal turnover rate of these enzymes was significantly lower. The substrate excess that could be successfully cleaved within one1 h at 0 °C was therefore limited to ≈500–1000-fold (NEDP1 enzymes), ≈400-fold (scAtg4) and ≈100-fold (xlUsp2), respectively. According to these results, all analyzed SUMO-, NEDD8-, scAtg8- and xlUb-specific proteases cleave highly efficient also at low substrate concentrations. At substrate concentrations typical for preparative applications (>10 μM substrate) these enzymes can therefore operate near their maximal turnover rates.

Similar titration experiments performed with TEV protease and a TEV protease substrate showed that, here, even at 25 °C and at an exceedingly high substrate concentration (300 μM), the number of substrate molecules cleaved per molecule of TEV protease within 1 h was limited to ≈100–150. Moreover, the substrate turnover per protease is further reduced at low substrate concentration: When titrating down the substrate at constant protease concentration the fraction of cleaved substrate increased only marginally (Fig. 7B). Along the same lines, reducing the concentration of both, the substrate and the protease, significantly impaired cleavage (Fig. 7A). These results are consistent with the rather high K_M of the reaction that is reported in the literature (50–90 μM [26,30,31]). Thus, complete substrate cleavage by TEV protease generally requires a high concentration of TEV protease, independently of how low the substrate concentration is. At low substrate concentration, the required ratio is even higher.

4. Discussion

The focus of this study has been the identification, recombinant expression and characterization of new protease/substrate pairs that can be used for removal of affinity tags from recombinant proteins. Our results show that the new proteases bdSENP1, bdNEDP1, ssNEDP1, scAtg4 and xlUsp2 can efficiently cleave appropriately

used within the respective pairs. Green bars mark lanes with the lowest protease concentration that cleaved its substrate efficiently. Note the remarkably low P_1' sensitivity of ssNEDP1, xlUsp2 and scAtg4. The xlUsp2 enzyme is the only enzyme tested that is able to cleave a P_1'–Pro substrate. For all P_1' variants tested, the bdSENP1 protease shows the highest absolute activity.

tagged target proteins within a wide range of temperatures, including 0 °C. These proteases therefore have great potential as general tools for purification of recombinant proteins and protein complexes. Important for preparative and biotechnological applications, all SUMO- and NEDD8-specific proteases could easily be produced in large amounts, excellent purity and activity. scAtg4 and xlUsp2 show also very promising catalytic properties, however, their expression and purification have not yet been fully optimized.

The tagging system with bdSUMO-, NEDD8-, Atg8- or ubiquitin-modules appears ideal for protein production in prokaryotic hosts. With respect to eukaryotic expression, it needs to be considered that cognate host proteases residing in the nucleus or cytoplasm might rapidly cleave fusions with ubiquitin-like modifiers. This restriction should, however, not apply to secreted recombinant proteins, because the cognate proteases should not reside in the lumen of the endoplasmic reticulum and the subsequent compartments of the secretory pathway.

4.1. bdSUMO and NEDD8 as expression and solubility enhancers

Frequently encountered problems in recombinant protein expression in *E. coli* are low levels of expression and/or poor solubility of the expressed proteins. A fusion of e.g. MBP, Protein A, or indeed SUMO to the N-terminus of the problematic protein can often relieve these problems. We found that bdSUMO, bdNEDD8 and ssNEDD8 have similar expression- and solubility-enhancing effects as already described for scSUMO [32]. In quite a few cases, NEDD8 was even superior to SUMO fusions (S.F. and D.G., manuscript in preparation). Of course, SUMO and NEDD8 can be combined with any other tag, such as polyHis, ZZ, MBP, or GST tags, be it for further enhancing expression and solubility or for conferring binding to dedicated affinity matrices.

4.2. Species specificity and sequence conservation

With scUlp1/bdSENP1 and bdNEDP1/ssNEDP1, we included two pairs of orthologous proteases in our analysis. Interestingly, both NEDP1 enzymes behave similar in most assays using the standard $P_1'_{Ala}$ substrates. This is surprising, especially when considering the moderate degree of conservation between the corresponding NEDP1 enzymes (33% identity; see also Fig. 1). According to the available structure of the human NEDD8•NEDP1 complex [21], the significant differences seen with respect to their P_1' preferences (Fig. 6) can probably be attributed to significant exchanges in protease residues contacting the substrate C-terminally of the scissile bond.

Compared to the *Brachypodium* and salmon NEDD8 orthologs, which share 84% identity on the amino acid level, the SUMO orthologs analyzed here are less conserved (46% identity; also compare Fig. 1A and C). Moreover, while the yeast SUMO (Smt3p) has a high similarity to the human SUMO1 isoform, the bdSUMO is more related to hsSUMO2. Similar to their substrates, also the SUMO proteases from yeast, *Brachypodium* and human show a low degree of sequence conservation. Structural alignments (including structure predictions for the *Brachypodium* enzyme) [18], however, indicate that all theses enzymes adopt a similar three-dimensional structure. According to published structure of the yeast scSUMO•scUlp1 complex [15], the substrate•enzyme interfaces of the respective yeast and *Brachypodium* complexes differ in a significant number of residues that may easily account for their species-specific properties.

4.3. Salt tolerance

More specifically, bdSENP1 outperforms its yeast ortholog in several aspects: already at standard salt concentrations (250 mM NaCl; see e.g. Figs. 3 and 8), bdSENP1 has a 2–3-fold higher specific activity as compared to scUlp1 (see e.g. Fig. 4). In addition, bdSENP1 can efficiently cleave substrates in a wide range of salt conditions while the yeast counterpart significant looses activity at NaCl concentrations above 250 mM (Fig. 5). This finding contrasts the relatively mild salt sensitivity (30% remaining activity at 1 M NaCl) reported for scUlp1 in the literature [12]. In this earlier report, however, the protease concentration used was significantly (presumably ≈10-fold) higher than needed for complete substrate cleavage under low-salt conditions. Consequently, a salt-induced decrease in protease activity by ≈90% would have escaped detection completely.

Similar to bdSENP1, also the two NEDP1 enzymes and xlUsp2 show an excellent tolerance to high salt conditions. These enzymes can therefore conveniently be used as tag-cleaving proteases in a variety of different buffers.

4.4. Protease/substrate ratio

A variety of commercial vectors include the TEV protease recognition site ("TEV site") e.g. after the GST tag. TEV protease is thus often considered as the first choice for removing affinity tags from target proteins. While comparing the catalytic properties of TEV protease to proteases of the SENP1 and NEDP1 enzyme families, it turned out that TEV protease has major limitations that should be considered in practice.

First, the effective turnover rate of TEV protease is poor. Even at 25 °C and at very high substrate concentrations, each molecule of TEV protease can cleave only ≈150 substrate molecules per hour (Fig. 7). In addition, because of the high K_M of the reaction (50–90 μM) [26,30,31], this turnover rate can only be reached at very high substrate concentration (>100–200 μM). At lower substrate concentrations, the number of substrate molecules cleaved per protease molecule drops significantly. Consequently, regardless of the concentration of substrate to be cleaved, roughly the same (high) concentration of protease is required. In practice, these properties have two major consequences. First, a complete cleavage by TEV protease is hard to achieve and generally requires long incubation times (e.g. 3–16 h), elevated temperature (generally 16–30 °C, as recommended by the commercial suppliers) or high enzyme concentrations. Second, any cleavage product will be contaminated with a rather high fraction of protease unless the substrate can be supplied in unreasonably high concentrations (>200 μM). For applications in an analytical or semi-preparative scale, the potential of TEV protease is therefore limited.

In comparison, the new proteases characterized here are highly efficient tag-removing enzymes. Remarkably, when using these enzymes, the substrate/protease ratio required for efficient cleavage remains rather constant even at low substrate concentrations. Therefore, especially bdSENP1, bdNEDP1 and ssNEDP1 are ideally suited for driving tag removal to completion. Importantly, when using these enzymes, the amount of protease used for cleavage

Table 2
Suggested protease concentrations for a near quantitative (>95%) substrate cleavage.

Protease	Cleavage at 0 °C[a]	Cleavage at 25 °C[a]
scUlp1	50–100 nM	10 nM
bdSENP1	20–50 nM	5 nM
bdNEDP1	0.5 μM	0.1 μM
ssNEDP1	1 μM	0.2 μM
scAtg4	0.5 μM	0.1 μM
xlUsp2	2–3 μM	1 μM
TEV protease	30 μM	5 μM

[a] Given protease concentrations refer to cleavage at standard conditions: 100 μM $P_{1'Ala}$ or $P_{1'Gly}$ substrates, 1 h, LS-S buffer (250 mM NaCl, 40 mM Tris/HCl pH7.5, 2 mM $MgCl_2$, 250 mM sucrose, 2 mM DTT).

Table 3
Correction factors for suggested protease concentrations at conditions deviating from standard conditions.

Protease	Parameter deviating from standard conditions[a]								
	0.5 M NaCl	1 M NaCl	16 h	Substrate ≤10 μM	$P_1'_{Met}$	$P_1'_{Tyr}$	$P_1'_{Arg}$	$P_1'_{Glu}$	$P_1'_{Pro}$
scUlp1	3	20	0.1	0.1	3	3	3	10	n.c.[b]
bdSENP1	1	2	0.1	0.1	10	3	3	30	n.c.[b]
bdNEDP1	2	3	0.1	0.1	3	3	1	10	n.c.[b]
ssNEDP1	1	1	0.1	0.1	1	1	1	1	n.c.[b]
scAtg4	3	10	0.1	0.1	3	3	3	3	n.c.[b]
xlUsp2	2	3	0.1	0.1	1	1	1	1	10
TEV protease	1	1	0.1	1	1	1	3	10	n.c.[b]

[a] Standard conditions as defined in Table 2.
[b] n.c.: No cleavage.

can be lowered according to the substrate concentration. As a rule of thumb, at 0 °C one molecule of bdSENP1 will cleave roughly 2–4 substrate molecules per second, i.e. in practice, a 5000–15,000-fold molar excess of substrate can easily be cleaved within 1 h at 0 °C. The NEDD8-specific enzymes have an approximately 10-fold lower turnover rate. Nevertheless, the two NEDP1 proteases can still digest an up to 1000-fold excess of substrate within 1 h at 0 °C.

At such enzyme/substrate ratio, the remaining "contaminant" protease that is used for cleavage can be neglected for most laboratory purposes. If desired, the protease concentration used for cleavage can, however, be further drastically decreased if the cleavage reaction is performed at higher temperature or for a longer time (for recommended protease concentrations and correction factors see Tables 2 and 3). This is easily possible as the characterized SUMO-and NEDD8-specific proteases remain fully active even after over-night incubation at 37 °C or 20 °C, respectively. A complete removal of the protease is possible using a protease variant harboring an engineered affinity tag. Together, these measures should allow for the removal of even trace amounts of protease.

5. Conclusions

In summary, we are confident that the new proteases described here will become widely applied tools for the specific tag removal from recombinant proteins. In the accompanying paper [33], we describe the application of these proteases for on-column cleavage of recombinant proteins and in a general approach allowing for the purification of recombinant protein complexes with defined stoichiometry.

Acknowledgments

We wish to thank Heike Behr for excellent technical support, Matthias Samwer and Bastian Hülsmann for comments on the manuscript, and the Max-Planck-Gesellschaft and the Deutsche Forschungsgemeinschaft (SFB860) for funding.

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.chroma.2014.02.029

References

[1] A.L. Demain, P. Vaishnav, Biotechnol. Adv. 27 (2009) 297.
[2] S.C. Makrides, Microbiol. Rev. 60 (1996) 512.
[3] H.P. Sorensen, K.K. Mortensen, J. Biotechnol. 115 (2005) 113.
[4] M. Uhlen, G. Forsberg, T. Moks, M. Hartmanis, B. Nilsson, Curr. Opin. Biotechnol. 3 (1992) 363.
[5] J.J. Lichty, J.L. Malecki, H.D. Agnew, D.J. Michelson-Horowitz, S. Tan, Protein Expression Purif. 41 (2005) 98.
[6] D.S. Waugh, Trends Biotechnol. 23 (2005) 316.
[7] C.L. Young, Z.T. Britton, A.S. Robinson, Biotechnol. J. 7 (2012) 620.
[8] J. Nilsson, S. Stahl, J. Lundeberg, M. Uhlen, P.A. Nygren, Protein Expression Purif. 11 (1997) 1.
[9] J. Arnau, C. Lauritzen, G.E. Petersen, J. Pedersen, Protein Expression Purif. 48 (2006) 1.
[10] S.J. Li, M. Hochstrasser, Nature 398 (1999) 246.
[11] T.R. Butt, S.C. Edavettal, J.P. Hall, M.R. Mattern, Protein Expression Purif. 43 (2005) 1.
[12] M.P. Malakhov, M.R. Mattern, O.A. Malakhova, M. Drinker, S.D. Weeks, T.R. Butt, J. Struct. Funct. Genomics 5 (2004) 75.
[13] S. Muller, C. Hoege, G. Pyrowolakis, S. Jentsch, Nat. Rev. Mol. Cell Biol. 2 (2001) 202.
[14] I. Schecter, A. Berger, Biochem. Biophys. Res. Commun. 27 (1967) 157.
[15] E. Mossessova, C.D. Lima, Mol. Cell 5 (2000) 865.
[16] E.T. Yeh, L. Gong, T. Kamitani, Gene 248 (2000) 1.
[17] A.G. van der Veen, H.L. Ploegh, Annu. Rev. Biochem. 81 (2012) 323.
[18] F. Armougom, S. Moretti, O. Poirot, S. Audic, P. Dumas, B. Schaeli, V. Keduas, C. Notredame, Nucleic Acids Res. 34 (2006) W604.
[19] Z. Xu, S.F. Chau, K.H. Lam, H.Y. Chan, T.B. Ng, S.W. Au, Biochem. J. 398 (2006) 345.
[20] L.N. Shen, C. Dong, H. Liu, J.H. Naismith, R.T. Hay, Biochem. J. 397 (2006) 279.
[21] L.N. Shen, H. Liu, C. Dong, D. Xirodimas, J.H. Naismith, R.T. Hay, EMBO J. 24 (2005) 1341.
[22] S. van den Berg, P.A. Lofdahl, T. Hard, H. Berglund, J. Biotechnol. 121 (2006) 291.
[23] C.M. Nunn, M. Jeeves, M.J. Cliff, G.T. Urquhart, R.R. George, L.H. Chao, Y. Tscuchia, S. Djordjevic, J. Mol. Biol. 350 (2005) 145.
[24] J.R. Porter, M. Gawith, Eur. J. Agron. (1999).
[25] T. Kirisako, Y. Ichimura, H. Okada, Y. Kabeya, N. Mizushima, T. Yoshimori, M. Ohsumi, T. Takao, T. Noda, Y. Ohsumi, J. Cell Biol. 151 (2000) 263.
[26] R.B. Kapust, J. Tozser, T.D. Copeland, D.S. Waugh, Biochem. Biophys. Res. Commun. 294 (2002) 949.
[27] G. Kostallas, P.A. Lofdahl, P. Samuelson, PLoS One 6 (2011) e16136.
[28] B. Nilsson, T. Moks, B. Jansson, L. Abrahmsen, A. Elmblad, E. Holmgren, C. Henrichson, T.A. Jones, M. Uhlen, Protein Eng. 1 (1987) 107.
[29] L. Michaelis, M.L. Menten, Biochem. Z. 49 (1913) 333.
[30] R.B. Kapust, J. Tozser, J.D. Fox, D.E. Anderson, S. Cherry, T.D. Copeland, D.S. Waugh, Protein Eng. 14 (2001) 993.
[32] J.G. Marblestone, S.C. Edavettal, Y. Lim, P. Lim, X. Zuo, T.R. Butt, Protein Sci. 15 (2006) 182.
[33] S. Frey, D. Gorlich, J. Chromatogr. A 1337 (2014) 106.