Journal of proteome research

Cite This: J. Proteome Res. XXXX, XXX, XXX−XXX

Article

pubs.acs.org/jpr

Host Cell Proteome of *Physcomitrella patens* Harbors Proteases and Protease Inhibitors under Bioproduction Conditions

Sebastian N. W. Hoernstein,[†] Benjamin Fode,[‡] Gertrud Wiedemann,[†] Daniel Lang,[†,§] Holger Niederkrüger,[‡] Birgit Berg,[‡] Andreas Schaaf,[‡] Thomas Frischmuth,[‡] Andreas Schlosser,[∥] Eva L. Decker,[†] and Ralf Reski*[,†,⊥]

[†]Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaenzlestrasse 1, D-79104 Freiburg, Germany

[‡]Greenovation Biotech GmbH, Hans-Bunte-Strasse 19, D-79108 Freiburg, Germany

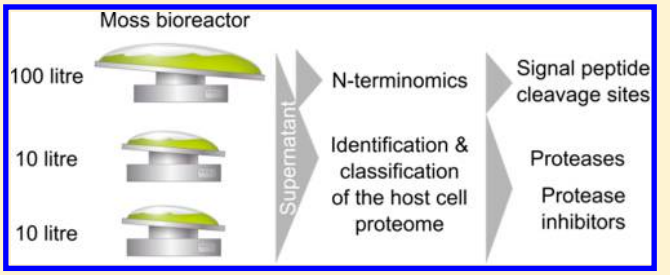
[§]Plant Genome and System Biology, Helmholtz Center Munich, D-85764 Neuherberg, Germany

[∥]Rudolf-Virchow-Center for Experimental Biomedicine, University of Wuerzburg, D-97080 Wuerzburg, Germany

[⊥]BIOSS - Centre for Biological Signalling Studies, University of Freiburg, D-79104 Freiburg, Germany

Supporting Information

ABSTRACT: Host cell proteins are inevitable contaminants of biopharmaceuticals. Here, we performed detailed analyses of the host cell proteome of moss (*Physcomitrella patens*) bioreactor supernatants using mass spectrometry and subsequent bioinformatics analysis. Distinguishing between the apparent secretome and intracellular contaminants, a complex extracellular proteolytic network including subtilisin-like proteases, metallo-proteases, and aspartic proteases was identified. Knockout of a subtilisin-like protease affected the overall extracellular proteolytic activity. Besides proteases, also secreted protease-inhibiting proteins such as serpins were identified. Further, we confirmed predicted cleavage sites of 40 endogenous signal peptides employing an N-terminomics approach. The present data provide novel aspects to optimize both product stability of recombinant biopharmaceuticals as well as their maturation along the secretory pathway. Data are available via ProteomeXchange with identifier PXD009517.



KEYWORDS: *bioreactor, biopharmaceutical, moss, proteases, signal peptides*

INTRODUCTION

The demand for recombinant biopharmaceuticals is continuously increasing,[1] and their overall complexity requires suitable cost-efficient production platforms with amenability for genetic engineering. The moss *Physcomitrella patens* provides an alternative production system for complex glycoproteins to the sector.[2,3] Several recombinant human glycoproteins have been successfully produced in *P. patens* such as vascular endothelian growth factor (VEGF), erythropoietin (EPO), or complement factor H.[4−8] Further, the human keratinocyte growth factor FGF7 (Bryokine FGF7) produced in *P. patens* is already commercially available for research purposes, and moss-aGal to treat Fabry disease successfully passed clinical phase I (www.greenovation.com). Common obstacles in the production of biopharmaceuticals are residual host cell proteins (HCP) that need to be minimized during downstream processing. Possible safety risks of such inevitable production residuals need to be addressed in obligatory toxicological studies.[9] Frequently, immunoassays are used to quantify the HCP content of products.[10]

Additionally, host cell proteases can significantly decrease the yield and quality of the product.[11−14] As a consequence, an RNA antisense approach in *Nicotiana tabacum* targeting several host proteases decreased proteolytic activity concomitant with an increased stability of recombinant proteins.[15] Several proteases were recently identified in the secretome of *P. patens* gametophores[16] and an even more complex network including proteolytically active serine and cysteine proteases was uncovered in agroinfiltrated *Nicotiana benthamiana* leaves.[17] In terms of bioproduction, filamentous protonemata, and not gametophores, are the dominating tissue in moss bioreactors.[18,19] Therefore, we analyzed the HCP of *P. patens* protonema filaments in axenic moss bioreactors under bioproduction conditions to gain a deeper understanding of the production process and to find novel targets for the enhancement of product quality and yield. Particularly we applied in-depth bioinformatics to characterize the HCP and to identify several secreted proteases that may interfere with the stability of recombinant proteins. Among the HCP we identified different classes of serine, cysteine, aspartic, and metallo-proteases as well as several proteins with known protease-

Received: June 5, 2018
Published: September 18, 2018

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

inhibiting domains. We established a knockout line of one of the most abundant proteases and observed a significant decrease in extracellular proteolytic activity.

EXPERIMENTAL PROCEDURES

Cultivation of *Physcomitrella patens* Protonema

For analysis of the host cell proteome the *P. patens* basic strain Pp_A_Doko#069 of Greenovation Biotech GmbH (Freiburg, Germany; IMSC acc. no. 40828) was used, in which the FT3 and XT-genes have been knocked out.[4]

Wave 100 Batch. The strain Pp_A_Doko#069 was cultivated as described.[20] Wave 10 batches: The cultivation process was performed as described.[21] In short, the strain Pp_A_Doko#069 was cultivated for 4 weeks in a 20 L disposable bag (Cellbag 20, GE Healthcare, Germany) placed in a Wave Reactor Rocker (BioWave 20 SPS, Wave Biotech AG, Switzerland). The cultivation parameters were according to Greenovation's production process using secretion-enhancing medium, fumigation of pressurized air supplemented with CO_2, and 24 h of light per day, delivered from light panels equipped with Osram FQ 24W 840 HO, Lumilux Cool White. The medium was supplemented with 1000x Nitsch vitamin mixture (Duchefa, Netherlands) according to the manufacturer's instructions. The pH of the fermentation was controlled automatically at pH 5−6 with the help of WAVEPOD I (GE Healthcare) in combination with Pump20 (GE Healthcare).

At the end of cultivation, the culture broth was clarified through a 3-step filtration cascade: (1) removal of moss through cake filtration in customized polypropylene filtration housing,[22] equipped with Zetaplus (01SP B3002, 3M, Germany), (2) depth filtration through a double layer Scale-Up Capsule (E0340FSA60SP03A, 3M, Germany), and (3) a sterile filtration (Millipore ExpressTM Plus, 0.22 μm, Millipore, Germany).

Cultiflask 20 mL Cultures. To quickly assess productivity and extracellular protease activity in producer moss cell lines, 20 mL-small-scale cultivations of each cell line were carried out in cultiflasks (50 mL tubes, lid with membrane for gas exchange, Greiner Bio-One, Kremsmünster, Austria). The cultures were incubated in secretory medium (SM) for up to 7 days in a shaker (Adolf Kühner AG, Birsfelden, Switzerland) with 2% CO_2, 25 °C and 24 h of light (85 μE, LED-modules from Infors AG, Bottmingen, Switzerland). For inhibition of protease activity, serine protease inhibitor AEBSF (4-(2-aminoethyl)-benzenesulfonyl fluoride, AppliChem GmbH, Darmstadt, Germany) with 0.25 mM final concentration was added to cultures prior incubation and on days 3 and 5. Supernatant was harvested at indicated time points to measure the product via ELISA or protease activity.

Sample Preparation

Supernatants of one 100 L reactor (Wave_100) and two samples from 10 L reactors (Wave_10A, Wave_10B) were provided as ultraconcentrate. The supernatant of the 100 L reactor was concentrated to 40 mL, and buffer was exchanged to 100 mM Tris, 2.5 M NaCl, pH 7.5 using a Cross-Flow (Centramate 500s, Pall, Crailsheim, Germany) with a 10 kDa cut off (Centramate Delta cassette, Pall). The ultrafiltrate was divided into aliquots of 5 mL, and each was precipitated with acetone. The supernatants from the two 10 L reactors were concentrated each to 170 mL, and aliquots of 35 mL were precipitated overnight using 10% trichloroacetic acid (TCA). The aliquots were centrifuged at 20,000*g* for 15 min at 4 °C. The supernatant was discarded, and the remaining pellet was washed with acetone containing 0.2% DTT. After centrifugation an additional washing step using acetone without DTT was performed. The remaining protein pellets were air-dried and stored at −20 °C for further use.

Protein pellets were dissolved in 100 mM HEPES-NaOH, pH 7.5, 0.2% SDS, and the protein concentration was measured using the BCA assay.[23] 50 μg dissolved proteins were mixed with an equal volume of Laemmli buffer (Bio Rad, Munich, Germany). Reduction of cysteine residues was carried out using Reducing Agent (Life Technologies, Carlsbad, USA) 1:10 at 95 °C for 10 min. Alkylation was performed at a final concentration of 100 mM iodacetamide for 20 min at RT. SDS-PAGE was performed with 12% or 7.5% Mini-Protean TGX gels (Bio Rad). After electrophoresis, gels were stained with PageBlue (Thermo Scientific, Waltham, USA) as recommended by the manufacturer. Each lane was cut into 15 gel slices. Preparation of the slices and trypsin digestion was done as described.[24] Gel slices were chopped to pieces using a scalpel and destained with 30% acetonitrile (ACN) in 100 mM NH_4HCO_3 with constant shaking for 10 min. The supernatant was discarded, and the destained gel pieces were equilibrated with 100 mM NH_4HCO_3. The supernatant was discarded, and gel pieces were shrunk with 100% ACN and dried in a vacuum concentrator. Tryptic digestion was performed with 0.1 μg trypsin per gel slice in 50 mM NH_4HCO_3 overnight at 37 °C. The supernatant was recovered. Additional peptides were extracted with 5% formic acid (FA) and pooled with the first supernatant. STAGE-Tip purification was performed as described below (enrichment of N-terminal peptides).

Enrichment of N-Terminal Peptides

Enrichment of N-terminal peptides was performed using HCP precipitate from the 100 L wave reactor. Employment of different proteases can be beneficial for the identification of N-termini by providing alternative evidence for an N-terminus or to identify N-termini with unfavorable peptide lengths after tryptic digest.[25] Hence we employed three different proteases (trypsin, GluC, chymotrypsin).

The dimethylation was performed as described.[24] In short, protein pellets (200 μg each) were dissolved in 100 mM HEPES-NaOH, 0.2%SDS and reduced with Reducing Agent (Life Technologies) at 95 °C for 5 min. Reduced proteins were alkylated in 100 mM iodacetamide for 20 min in the dark. For dimethylation 2 μL of a 4% formaldehyde solution (isotope labeled $^{13}C,d_2$, Sigma-Aldrich) and 2 μL of $NaCNBH_3$ (500 mM) per 100 mL sample was added. After incubation for 4 h at 37 °C the same amounts of formaldehyde and $NaCNBH_3$ were added and the reaction was continued overnight. Quenching was achieved by adding 2 μL of a 4% NH_4OH solution and incubation for 1 h at 37 °C. Dimethylated proteins were precipitated with acetone for 3 h at −20 °C. Dried protein pellets were dissolved in binding buffer[26] containing 20 mM NaH_2PO_4 and 150 mM NaCl pH 7.5 with 0.2% SDS, and in-solution digest using either trypsin (Promega, Madison, USA), GluC (Thermo Scientific), or chymotrypsin (Promega) was performed at an enzyme-to-substrate ratio of 1:25 for 4 h at 37 °C (trypsin, GluC) or 25 °C (chymotrypsin). Then the ratio was increased to 1:20 and the reaction was carried out overnight. Enrichment of amino-terminal labeled peptides was carried out using 200 μL NHS-sepharose slurry (GE Healthcare). The slurry was centrifuged for 30 s at 200*g*. The supernatant was discarded, and 400 μL ice-cold 1 mM HCl was added. The slurry was centrifuged again, and the supernatant was discarded. Afterward the sepharose was washed with 1 mL binding buffer without

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

SDS. The samples were applied to the prepared sepharose and incubated for 4 h at RT. The sepharose was again centrifuged, and the supernatant was transferred to a new tube containing freshly prepared sepharose. The used sepharose was washed with 20 μL binding buffer, and the supernatant was also added to the freshly prepared sepharose. The enrichment reaction was carried out overnight at 4−8 °C. The enriched peptides were desalted using 200 μL C18 StageTips (Thermo Scientific) that were supplemented with an additional layer of Empore SPE Disk C18 material (Sigma-Aldrich). The tips were washed prior to use with 100 μL 0.1% trifluoroacetic acid (TFA) and subsequently with 100 μL 80% ACN, 0.1% TFA. The tips were again equilibrated with 100 μL 0.1% TFA and the samples were loaded afterward. The remaining sepharose was washed with 50 μL binding buffer, and the supernatant was also transferred to the tip. The tips were washed with 100 μL binding buffer, and the retained peptides were eluted with 100 μL of 80% ACN, 0.1% TFA. The eluate was vacuum-dried and the dried peptides were directly used for MS/MS analysis.

LC-MS/MS Analysis

NanoLC-MS/MS analyses were performed on an LTQ Orbitrap Velos Pro (Thermo Scientific) equipped with an EASY-Spray ion source and coupled to an EASY nLC 1000 (Thermo Scientific). Peptides were loaded on a trapping column (2 cm × 75 μm ID, PepMap C18, 3 μm particles, 100 Å pore size, Dionex, Thermo Scientific) and separated either on a 25 cm EASY-Spray column (25 cm × 75 μm ID, PepMap C18, 2 μm particles, 100 Å pore size) with a 30 min linear gradient from 3% to 30% ACN (Promochem) and 0.1% FA (Thermo Scientific) at 400 nL/min in the case of in-gel digested samples or with a 240 min linear gradient from 3% to 30% ACN and 0.1% FA at 200 nL/min in the case of in solution digested proteins (enriched N-terminal peptides). MS scans were acquired in the Orbitrap analyzer with a resolution of 30000 at m/z 400; MS/MS scans were acquired in the Orbitrap analyzer with a resolution of 7500 at m/z 400 using HCD fragmentation with 30% normalized collision energy. A TOP5 data-dependent MS/MS method was used. Dynamic exclusion was applied with a repeat count of 1 and an exclusion duration of 30 s or 2 min in the case of long gradients. Singly charged precursors were excluded from selection. Minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with a target value of 10^6 for MS scans and 5×10^4 for MS/MS scans. Lock mass option was applied for internal calibration using background ions from protonated decamethylcyclopentasiloxane (m/z 371.10124).

Raw Data Processing and Database Search

Raw data processing was performed using Mascot Distiller V2.5.1.0 (Matrix Science, Boston, Massachusetts, USA). Database searches on the processed raw data were submitted to the Mascot search engine using Mascot Daemon V2.4 (Matrix Science) against the *Physcomitrella patens* database containing all version 1.6 protein models[27] as well as their reversed sequences used as decoys and simultaneously against an in-house database containing all sequences of known typical contaminants (e.g., human Keratins, Trypsin, 267 total entries, available on request). For database searches of dimethylated samples the fixed modifications were carbamidomethyl (C) + 57.021464 Da and $^{13}C,d_2$ dimethyl (K) + 34.063117 Da. Variable modifications used for database search were Gln > pyro Glu (N term Q) − 17.026549 Da, oxidation (M) + 15.994915 Da, acetyl (N-term) + 42.010565 Da, $^{13}C,d_2$ dimethyl (N-term) + 34.063117 Da, hybrid-methylation (N-term) + 31.047208 Da, formyl (N-term) + 27.994915 Da.

For database searches of the in-gel digested samples the fixed modifications were carbamidomethyl (C) + 57.021464 Da. Variable modifications used for database search were Gln → pyro Glu (N term Q) − 17.026549 Da, oxidation (M) + 15.994915 Da, acetyl (N-term) + 42.010565 Da, hydroxyproline (P) + 15.994915 Da, and phospho (ST) + 79.966331 Da. For all searches the peptide mass tolerance was ±8 ppm and the fragment mass tolerance was set to ±0.02 Da.

For database searches on trypsin-digested samples the enzyme specificity was set to semitryptic with a total of 2 in the case of in-gel digested samples or 5 missed cleavage sites in the case of in-solution digested samples. For samples digested with GluC or chymotrypsin the specificity was set to "none".

All Mascot searches were loaded into Scaffold 4 (Version 4.2.1, Proteome Software, Portland, Oregon, USA). Database searches from amino-terminal labeled samples were additionally searched using X!Tandem[28] implemented in Scaffold 4 against the *P. patens* protein models V1.6 with the same search parameters as used for the Mascot searches. Samples were loaded and analyzed using the legacy *PeptideProphet* scoring (high mass accuracy) with standard experiment wide protein grouping. Results were filtered either using the *Protein-* and *PeptideProphet*[29,30] or directly with the FDR filter implemented in Scaffold 4 software. The filter settings for the SDS-PAGE samples were as follows:

Protein threshold: 99% ProteinProphet, peptide threshold 95% PeptideProphet, minimum number of peptides: 2 and for the measurements of enriched N-terminal peptides:

Protein threshold: 1% FDR, peptide threshold: 0.5 FDR, minimum numbers of peptides: 1.

For further analysis of the identified N-terminal peptides, all peptides were considered that showed either a dimethylated N-terminus, an acetylated N-terminus or a pyro-Glutamate. Peptides with N-terminal pyro-Glutamate were only considered if the preceding P1 amino acid did not match the specificity of the experimentally used protease:

tryspin	C-Term	K|R
GluC	C-Term	D|E
chymotrypsin	C-Term	F|L|W|Y

Data Analysis

All data analysis was performed using custom Perl scripts (www.perl.org) and Bioperl modules (www.bioperl.org). If not stated otherwise, assignment of homologous genes was done according to the OrthoMCL[31] cluster analysis published in ref 27. If not stated otherwise, the gene identifiers used for proteins of other organisms are either TAIR accession numbers (http://www.arabidopsis.org/) in the case of *A. thaliana* proteins or UniProt accession numbers (http://www.uniprot.org/). All plots and charts presented here as well as PCA and subsequent clustering using Mclust[32] were created using R (http://www.r-project.org/). Data imputing of missing values for the PCA has been performed using the R package *impute*.[33] GO enrichment analysis of the clusters obtained from the Mclust clustering was performed using the R package *topGO* V1.0.[34] To test for significantly enriched GO Terms, a Fisher's exact test in combination with the weight01 method implemented in topGO was performed. Locus-wise annotations from pred2goa[27] were used as reference annotations.

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. Primers Used for Cloning and Identification of Subtilase Knockout Lines

Name	Sequence 5′→3′	Tm [°C]	Product [bp]	Function
S8KO_fwd1	CGTCGAGCTCCGAAGTGTATGCTTCTG	70	1367	KO-construct fragment 1
S8KO_rev1	GTCCCGGACTACAGGTCGGAAATTCCCATGATCTG	74		
S8KO_fwd2	CCGACCTGTAGTCCGGGACGATGAACAAC	73	1154	KO-construct fragment 2
S8KO_rev2	AGGTGTGCCAACTGCCAAAGATCC	69		
S8KO_g_fw	ATGATGACATGGCCTTGGAG	61	1483 (Ctrl)	Analysis of subtilase locus
S8KO_g_rv	GAGATAGTCGGATCCAAACC	59	------ (KO)	
S8_RT_fw	GGGTTTGGATCCGACTATCTC	61	354	Subtilase expression
S8_RT_rv	CGAATCGGTCCCACATTTG	60		
Act_RT_fw	GGATGGGCAGGTGATAAC	59	321	Actin expression
Act_RT_rv	CTGAGGGATGCGAGAATG	59		

Prediction of Subcellular Localization

Cleavage sites for plastid, mitochondrial, and secretory signal peptides were predicted using TargetP 1.1[35,36] for all V1.6 protein models. Additional cleavage site predictions for the selected secretome candidates were performed using Signal-3L[37−40] using the plant reference set (http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/). The whole genome prediction for leaderless secretion was performed using SecretomeP 2.0.[41] The prediction of transmembrane domains was performed using MEMSAT,[42] TMHMM,[43] and HMMTOP.[44]

Annotation of PFAM-Domains

The sequences of all identified isoforms from the SDS-PAGE samples of the HCP as well as from the N-terminal peptide enrichment experiments were searched against the PFAM database version 27.0 (http://pfam.sanger.ac.uk/) by the HMMER3 algorithm using the "gathering threshold" settings.[45,46]

Microarray Analysis

Microarray experiments were performed using a Nimblegen v1.6 *P. patens* 135k x 12 array (A-MTAB-528 on ArrayExpress, design OID40596/110407_Ppat_SR_exp_HX12.[47] For microarray analysis, Pp_A_Doko#69 protonema was grown in flasks with a dry weight of 120 mg/L and harvested 6 days after the last subculture. Total RNA was extracted as described.[48] 200 ng of total RNA was reverse transcribed and amplified with the WTA Kit (Sigma-Aldrich). The design of the Nimblegen *Physcomitrella patens* v1.6 transcript microarray (Roche, Switzerland) is based on the V1.6 gene models[27] and represents 32,275 transcripts with an average of four 60mer probes per gene. One microgram of cDNA was labeled with Cy3 according to the NimbleGen One-Color DNA Labeling Kit (Roche); microarray hybridization using 4 μg of labeled cDNA, washing, imaging, and data processing were performed as described[47] and on ArrayExpress (www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5374). Microarray expression values were analyzed as described[49,50] using the Expressionist Analyst 7.5.7 (Genedata, Basel, Switzerland). Microarray data used in this study are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-5374.

Generation of the Protease S8 Knockout

The moss strain Pp_P_aGal-1#001 used for the generation of the protease KO was based on Greenovation's glycan-engineered basic strain Pp_A_Doko#069. Additionally, this strain is capable of producing recombinant human alpha-Galactosidase A (aGal). In this strain, the gene for the most abundant extracellular S8 peptidase (Pp1s48_63V6.1) was targeted by homologous recombination using a construct with genomic sequences surrounding exons 2 and 3 and the connecting intron. PCR-amplification on moss genomic DNA with primers S8KO fwd1 and S8KO rev1 (Table 1) gave rise to a 5′ homologous sequence fragment upstream of exon. The 3′ homologous sequence fragment was amplified and connected to the 5′ fragment by overlapping PCR with primers S8KO fwd1 and S8KO rev2.

Integration of this construct into the moss genome lead to simultaneous KO of the exon 2−3 parts of S8 peptidase Pp1s48_63V6.1 and insertion of an in-frame stop codon in the remaining part of exon 2 (following Leu58). Success of KO was confirmed by PCR on the genomic and transcript level.

Preparation and regeneration of moss protoplasts were performed as described[51] with modifications. 200 mL moss protonema culture grown in shake flasks was used for generation of protoplasts by digestion with Driselase (Sigma-Aldrich/Merck, Munich, Germany) for 2 h. PEG-mediated transformation of moss protoplasts was carried out as described[52] with modifications. DNA was prepared and purified using Qiagen's Plasmid Plus Mega or Giga Kits (QIAGEN, Venlo, The Netherlands). Transformation was carried out in 50 mL plastic tubes for 12 min at RT, with gentle mixing every 2 min. After transformation, regenerated protoplasts were diluted prior to spreading on agar plates due to lack of a selection marker.

Alpha-Galactosidase ELISA

For detection of alpha-Galactosidase A (aGal) product levels in moss cultiflasks, a customized aGal-ELISA (BioGenes, Berlin, Germany) was used, based on standard sandwich-technology with two different polyclonal antibodies against recombinant aGal raised in rabbits.

Protease Activity Assay

The protease activity in moss supernatants from cultiflasks was measured using the Pierce Fluorescent Protease Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States). The assay detects decreases in fluorescence quenching that occur as the fluorescein isothiocyanate (FITC) labeled native casein substrate is digested into smaller fluorescein-labeled fragments. Therefore, the trypsin standard, 100 μL of each sample and blanks was mixed with 100 μL FITC-Casein in a black 96-well plate and incubated for 60 min in the dark at RT. In addition, all samples were diluted in a 1:4 ratio with 25 mM Tris, 150 mM NaCl pH 7.2 prior measurements. For detection on a ClarioStar fluorescence reader (BMG Labtech, Ortenberg, Germany), standard fluorescein excitation/emission filters (485/538 nm) were used.

Other Tools

Venn diagrams were created using BioVenn.[53] Sequence logos were created using iceLogo software.[54]

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

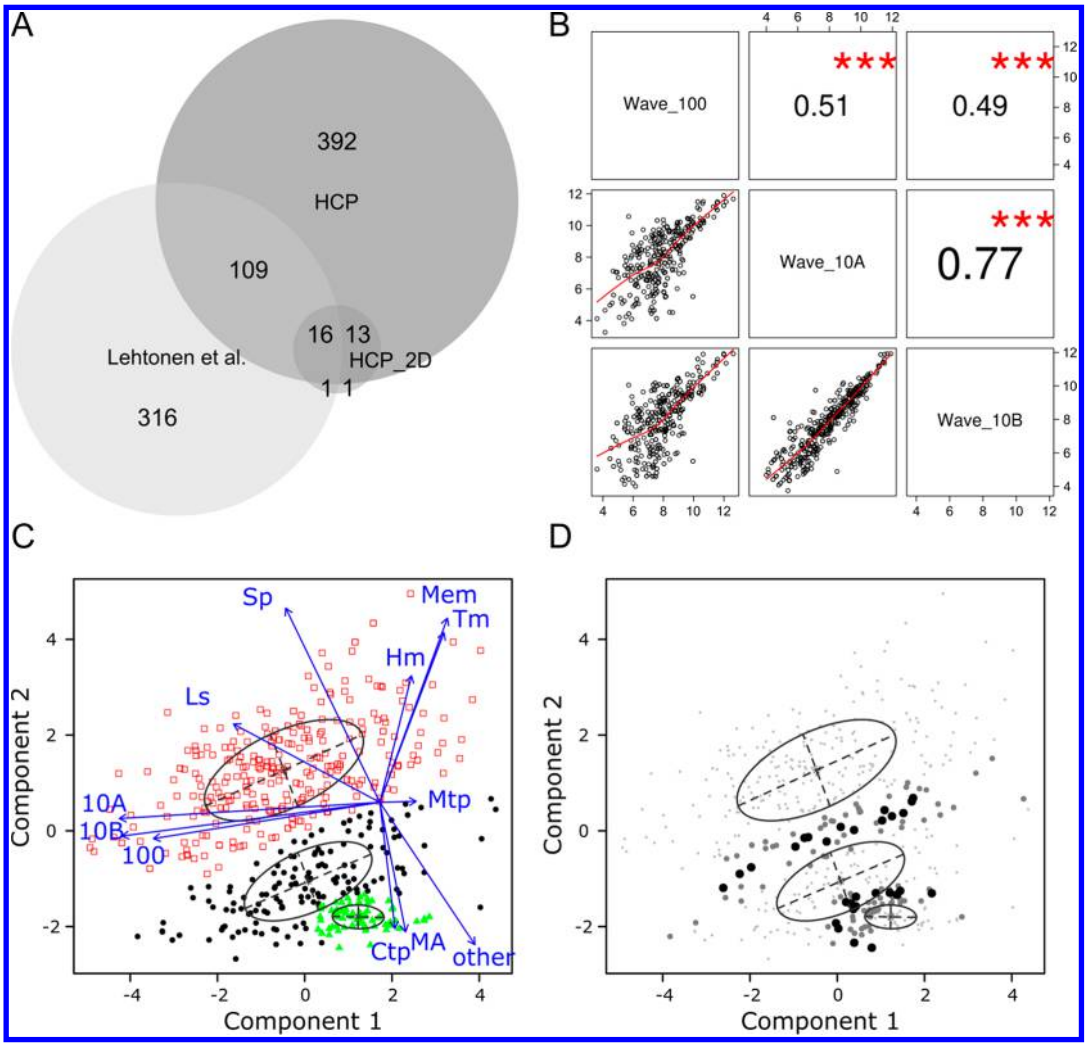


Figure 1. Overview of proteins identified in the HCP and their characterization. (A) Overlap of proteins identified in the present study by SDS-PAGE and enrichment of N-terminal peptides (HCP), proteins identified from the HCP of the 100 l wave reactor by 2D-PAGE (HCP_2D[20]) and proteins identified in the secretome of *P. patens* gametophores.[16] Diagram created using BioVenn.[53] (B) Relative abundance levels (SPKM-values, see Supplemental Experimental Procedures) of the identified proteins correlate statistically significantly ($p < 0.001$) and strong (Kendalls $\tau > 0.7$) between the 10 L wave reactors whereas the correlation of the abundance levels between the 10 and 100 L reactors is moderate (Kendalls $\tau < 0.7$) but statistically significant ($p < 0.001$). (C) Principal component analysis of all protein isoforms identified in the in-gel digested samples. Values used for the PCA are SPKM values for every protein isoform identified in the three reactor supernatants (10A, 10B, 100), prediction scores for leaderless prediction (Ls) by SecretomeP, predictions scores for cleavable signal peptides (Sp), plastid transit peptides (Ctp), mitochondrial presequences (Mtp), or other localizations than the aforementioned (other) predicted by TargetP. Additionally, the expression level of the corresponding transcripts (MA) from the microarray as well as the number of potential transmembrane domains predicted by HMMTOP (Hm), MEMSAT (Mem), and TMHMM (Tm) were included. The length of the arrows is proportional to the strength of the vector. Colors indicate the clusters defined using Mclust. The ellipses indicate the center of each cluster. (D) Depiction of the cluster uncertainty for every data point (protein isoform) from the PCA. The size of each data point is proportional to the level of uncertainty.

RESULTS AND DISCUSSION

Protein Identification and Characterization of the Host Cell Proteome

For the analysis of the HCP under production conditions, we used *Physcomitrella patens* wave bag reactor supernatants and analyzed proteins by SDS-PAGE and subsequent mass spectrometry. Here we reanalyzed the supernatant of a 100 L wave reactor (Wave_100) which has already been investigated via 2d-PAGE[20] and further employed enrichment of N-terminal peptides of proteins to identify cleavage sites of secretory signal peptides. In addition, we analyzed the supernatants of two parallel-cultivated 10 L wave reactors and thus generated three independent biological data sets.

We identified peptides corresponding to 530 different protein isoforms across all performed measurements (Table S1, Table S2) including all except two proteins identified in the previous study[20] by 2D-PAGE from the same supernatant of the 100 l reactor (HCP_2D, Figure 1A) used in the present study. We identified more than 80% of all proteins in at least two of the supernatants (Figure S1A). Proteins identified only in a single supernatant were on average identified with less assigned spectra (Figure S1B), indicating their low abundance in the HCP. The study of Lehtonen et al.[16] analyzed the *P. patens* gametophore secretome in response to a fungal elicitor (Chitosan). Remarkably, the overlap between our study and the gametophore secretome (treated and untreated) was comparably low (~15% of all identified proteins), revealing significant

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

differences in the HCP composition of *P. patens* gametophores and protonema despite a "core-HCP" of 125 proteins present in both (Figure 1a).

The relative abundance (SPKM values, see Supplemental Experimental Procedures) of the HCP identified via SDS-PAGE correlated strongly (Kendall's τ = 0.77) and significantly (p < 0.001) when comparing the two 10 l wave reactors (Figure 1B), indicating very low batch-to-batch variations of the HCP. The correlation of protein abundances between the 100 l and the 10 l wave reactors was moderate (Kendall's τ = 0.51 and 0.49, respectively), but nevertheless statistically significant (p < 0.001). These slight differences may originate from different sample preparations prior to MS analysis (see Experimental Procedures). Different protein precipitation methods can lead to marginal changes in proteome compositions,[55] which would explain the comparably lower but still significant correlation of protein abundances. However, regarding the fact that the samples analyzed here are not replicates in a strict sense, the observed overlap and significantly correlating abundances of proteins clearly demonstrate the highly reproducible production conditions of wave reactors, where minor differences only occur on the level of low abundant proteins.

Next, we aimed to dissect the identified HCP into proteins of the true secretome, namely proteins that are transported into the extracellular space either via the secretory pathway or by leaderless secretion, and contaminating proteins due to cell leakage. To achieve this, we performed a principle component analysis (PCA) utilizing the following criteria: (1) normalized protein abundance values (SPKM), (2) prediction scores for cleavable targeting peptides (plastid, mitochondrion, secretory),[35,36] leaderless secretion[41] and numbers of transmembrane domains,[42−44] and (3) transcript abundance levels derived from a transcriptomics approach based on microarray analysis (Table S3). Subsequent clustering was done by finite Gaussian mixture modeling employing the Mclust algorithm.[32] All values used for the PCA as well as the result of the clustering are depicted in Table S3. We obtained three main clusters of proteins (Figure 1C, black, red, green) even though the clusters were not separated clearly and showed interfaces with high cluster uncertainties (Figure 1D). Further, several protein isoforms derived from the same genomic locus were split across different clusters as a result of alternative splicing of transcripts (see Supplemental Discussion). To validate the performance of our clustering, we performed GO enrichment analysis for each cluster using locus-wise annotations based on manual annotation (www.cosmoss.org) or the pred2goa algorithm.[27] This algorithm includes predictions for subcellular localization employing several independent prediction algorithms and decides on a localization consensus. The enrichment analysis (Table S4) revealed that cluster 2 (Figure 2, red) was strongly enriched in proteins targeted to the extracellular space as well as proteins associated with membranes. The strong overrepresentation of membrane-associated proteins is most likely a consequence of the application of poly alkylalcohols during the cultivation process to enhance the recovery of the secreted product.[56] In contrast, cluster 1 (Figure 2, black) and cluster 3 (Figure 2, green) were enriched in proteins targeted to organelles. We performed extensive manual validation of protein annotation to evaluate the performance of our cluster analysis (see Supplemental Discussion and Tables S5−S8). In consequence, our PCA, incorporating experimentally observed abundance values and computation predictions, robustly dissects the HCP into the apparent secretome and intracellular contaminants. According to this analysis, the HCP of *P. patens* protonema in wave reactors contains at least 55% secreted proteins and maximum 45% intracellular contaminants.

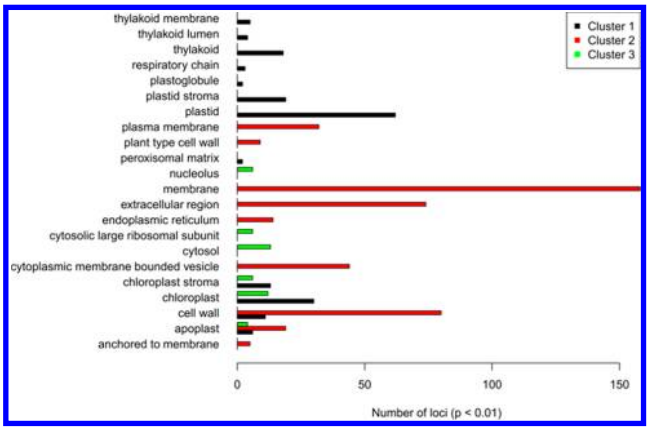


Figure 2. Significantly (p < 0.01) overrepresented GO-terms for each of the PCA clusters. Cluster 2 (red) is mainly composed of proteins associated with membranes as well as proteins localized to the extracellular space including the cell wall. Cluster 1 (black) and cluster 3 (green) are enriched in proteins localized to organelles. A detailed list is available in Table S4.

Protease Content of the Host Cell Proteome

Proteases generally may negatively influence the yield and quality of recombinant biopharmaceuticals.[57] To investigate the protease content of the *P. patens* HCP, we searched the sequences of all identified proteins against the PFAM database[46] and inspected all proteins for the presence of protease domains (Table S9 and S10). In total, we identified 10 different types of proteases (according to the MEROPS nomenclature[58]) including serine-type proteases (S8, S10, S28), cysteine proteases (C1, C13), aspartic proteases (ASP), and metalloproteases (M1, M6, M10, M49) represented by proteins encoded from 25 different genomic loci. The abundance range of these proteases covers 3 orders of magnitude (Figure 3a) with serine protease type S8 being the most abundant.

The number of proteases identified here from protonema cultures is remarkably higher than in the *P. patens* gametophore secretome where six different types of proteases encoded by 12 genomic loci were identified.[16] Additionally, we identified several proteins harboring TAXi domains. Proteins of this domain family are supposed to have evolved from pepsin-type aspartic proteases and are inhibitors of cell wall degrading xylanases but lost their proteolytic function.[59,60] Interestingly, four out of the five proteins (Pp1s165_128V6.1, Pp1s387_51V6.1, Pp1s38_350V6.1, Pp1s38_390V6.1) are homologous to aspartic proteases in *Arabidopsis thaliana*. A multiple sequence alignment (Figure S2) comprising all of the identified TAXi domain-containing proteins and their homologues from *A. thaliana* revealed that all of them harbor the two conserved active site motives of aspartic proteases Asp-Thr/Ser-Gly.[61,62] In contrast, the representative TAXi domain protein from wheat (Q8H0K8)[63] does not harbor the conserved active side residues as described.[59] In fact, all of the TAXi domain proteins have the same conserved active site domain as another annotated aspartic protease that we identified in the HCP (Pp1s93_73V6.2). Hence, we suggest that the annotation of the TAXi domain by PFAM may be wrong and that all of these proteins are in fact aspartic proteases (ASP, PF00026). In consequence, the aspartic proteases represent the second most

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

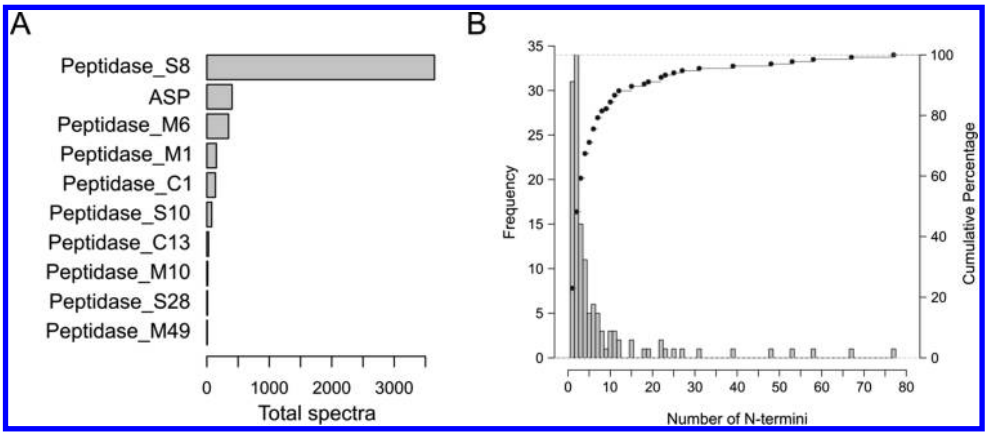


Figure 3. Abundance of protease families and frequency of identified N-termini. (A) All spectra assigned to proteins corresponding to a family of proteases (according to the MEROPS nomenclature)[58] were counted. Spectral counts for the aspartic protease family (ASP) include the spectral counts for the falsely annotated TAXi-domain proteins. Subtilisin-like proteases (SBTs, S8) are among the most abundant proteases. Aspartic proteases (ASP) and metalloproteases of the M6 class were also identified with high spectral counts. Other serine proteases (S10, S28), cysteine proteases (C1, C13), as well as other metalloproteases (M1, M10, M49) were only detected with a few spectral counts. (B) Frequency of the number of identified N-termini per protein is depicted as histogram. Dots indicate the cumulative frequency for the distribution. More than 70% of all proteins were identified with more than one N-terminus.

abundant group of proteases in the HCP identified here and they were all assigned to cluster 2, representing the apparent secretome. Aspartic proteases represent a family with 51 members in *A. thaliana*[64] and 32 encoding loci in *P. patens*, and they contribute to development[65] and stress response signaling.[66] Further, Pp1s93_73V6.2 harbors two saposin-like domains (SapB_1, SapB_2, Table S10). This domain is necessary for vacuolar targeting.[67−69] In fact, this protease localizes to the ER-Golgi network and the vacuole in transfected *P. patens* protoplasts.[70] In this respect, this protease might represent an intracellular contamination, although we cannot exclude additional targeting to the extracellular space.

Proteolytic activity on recombinant proteins has been repeatedly observed in plant host systems[57] although susceptibility to various classes of proteases differs between different recombinant proteins. Moreover, S8 serine protease activity in the culture medium of plant production systems is measurable[71] and negatively affects recombinant proteins.[15] In particular, degradation of the recombinant heavy chain of the human anti-HIV antibody 2F5 was dependent on serine protease activity as well as cysteine protease activity[72] whereas degradation of a recombinant plasminogen activator (DSPAα1) was dependent on metalloprotease activity.[11,73,57]

Proteases of the S8 family are subtilisin-like proteases (SBTs) and possess Asp, Ser, and His as catalytic triad.[74] They comprise 56 members in *A. thaliana*[75] and 22 protein-encoding loci in *P. patens*. Interestingly, five (Pp1s48_63V6.1, Pp1s232_23V6.2, Pp1s170_22V6.1, Pp1s78_186V6.1, Pp1s7_237V6.1) out of the six identified peptidases of the S8 family (Table S10) were assigned to cluster 1 which is mainly composed of intracellular contaminations. Only one SBT identified in the HCP (Pp1s19_223V6.1) was predicted to harbor a secretory signal peptide (Table S3). This is in contrast to the large majority of the *A. thaliana* SBTs.[75] Two of the SBTs identified here (Pp1s48_63V6.1, Pp1s7_237V6.1) were predicted by TargetP to harbor mitochondrial presequences which is presumably causing assignment to cluster 1. SBTs were identified in several plant secretomes (*A. thaliana*,[76,77] rice,[78] *P. patens*,[16] maize[79]) and are involved in a variety of processes including defense against pathogens,[80] programmed cell death,[81] conversion of pro-proteins,[74] or maturation of plant peptide hormones from precursors.[82] Here, the lack of closely related homologues from other species hampers unambiguous assignment of the localization of the identified *P. patens* SBTs. However, regarding their overall high abundance and the fact that most of them were also identified in the gametophore secretome[16] we conclude that they are all present in the protonema HCP upon targeted secretion.

Members of the M6 metalloproteases were also among the more highly abundant proteases but were mostly assigned to cluster 1 and are thus most probably intracellular contaminations (Table S10). One of these proteases (Pp1s37_124V6.1) was identified as another splice variant (Pp1s37_124V6.3) in the gametophore secretome[16] although splicing does not affect the N-terminus and neither TargetP nor SecretomeP predicts secretion (Table S3). Therefore, this protease may represent an intracellular contamination present in both secretomes. Additionally, we identified a single protease of the M10 class, which is supposed to be secreted (Pp1s184_7V6.1, Table S10).

Here, we provide an estimation of abundance for each class of proteases as well as localization annotation for each protease. Further, extensive proteolytic activity in the HCP is indicated as for approximately 80% of all proteins identified via enrichment of N-terminal peptides two or more different N-termini were identified (Figure 3b). However, our data do not allow tracing back proteolytic effects to a certain class of proteases. Taken together, our PCA reveals the presence of secreted proteases including the subtilisin class, the metalloprotease class, as well as several members of cysteine proteases. All of these protease classes affect recombinant proteins in other plant systems.[72,73]

Knockout of a Subtilisin-like Protease

One way to stabilize recombinant proteins is to reduce their proteolytic degradation by targeting host-specific proteases either by genome engineering or by microRNA-based knockdown.[15] *P. patens* provides an efficient toolbox for genetic engineering via homologous recombination.[83] Here we knocked out one of the most abundant proteases identified in the secretome, a subtilisin family protease (Pp1s48_63V6.1), in the background of the glyco-engineered strain used for our

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

secretome analysis. This strain is also capable of producing and secreting human α-Galactosidase A (aGal). Positive knockout lines were confirmed by RT-PCR, and two lines were identified (Figure S3a). Both, knockout (S8KO) and aGal line accumulated the same amount of dry weight during 7 days under bioproduction conditions (Figure S3b) and no obvious phenotypical deviations were observed (Figure S3c). Although the difference in the aGal titer between the parental (aGal) and the subtilase knockout line (S8KO) during the first 3 days of the production process was significant ($p < 0.05$), this difference was compensated after 7 days (Figure 4a). However, the aGal levels after 3 days are quite low. Interestingly, the same observations were made when the proteolytic activities of serine proteases were inhibited (Figure 4a, AEBSF (4-(2-aminoethyl)-benzenesulfonyl fluoride)) during the production process and the titers were comparable with titers of untreated plants. Conversely, we observe a significantly lower ($p < 0.05$) extracellular protease activity in the supernatant of the knockout line (Figure 4b). Hence, we conclude that the knockout of the subtilisin-family protein is suitable to decrease extracellular proteolytic activity significantly and that α-Galactosidase is most likely not targeted by serine proteases effectively. Otherwise, an increase of the aGal level after broad range inhibition of serine proteases with AEBSF should have been observable. Thus, knockout of this protease may represent a promising approach to stabilize proteins prone to degradation by serine proteases.

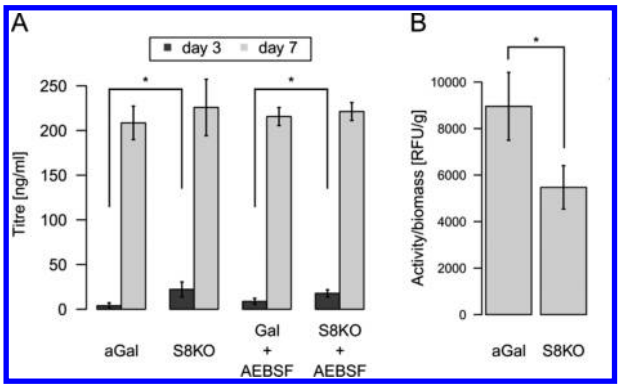


Figure 4. α-Galactosidase titer and extracellular protease activity in the subtilase knockout line (S8KO) and the parental line (aGal). (A) Extracellular aGal titer (mean of three biological replicates and standard deviation) determined via ELISA in the S8KO line and the parental line with and without serine protease inhibitor treatment (AEBSF) on day 3 and 7 of the production process. After 3 days the difference in the aGal level is significant ($p < 0.05$) although the level itself is low. After 7 days, the aGal levels are equalized between the parental and the knockout line. Treatment with AEBSF showed no further effect on the aGal levels. (B) Extracellular protease activity after 7 days of the production process in the parental line and the S8KO line (mean of three biological replicates and standard deviation). The activity is significantly reduced ($p < 0.05$) in the S8KO line.

Secreted Protease Inhibitors

Currently, different strategies are being employed to minimize degradation of recombinant proteins by proteases.[57] One of these strategies, currently available for *P. patens*, is the addition of stabilizing additives or the coexpression of human serum albumin.[84] However, the coexpression of different protease inhibiting proteins is especially promising.[13,71,85,86] Here, we identified four serpins (Pp1s22_263V6.1, Pp1s414_23V6.1, Pp1s61_265V6.1, Pp1s86_63V6.1, Table S10) which comprise all known proteins of this class in *P. patens*. Serpins are effective irreversible inhibitors of proteases as they mimic cleavage targets for proteases but capture them in a kinetic trap via covalent bonds.[87] One of the serpins identified here is a homologue of AtSerpin1, which inhibits cysteine protease RD21 (AT1G47128.1) in *A. thaliana* and is targeted to the ER and the apoplast.[88,89] In turn, RD21 is a homologue of four cysteine proteases of the C1 class in *P. patens* (Pp1s285_10V6.1, Pp1s292_39V6.1, Pp1s369_30V6.1, Pp1s49_32V6.1), which we identified in the HCP and which are targeted to the extracellular space (Table S10). Proteases of the C1 family are involved in several processes such as pathogen defense, disease resistance signaling, and senescence.[90,91] Interestingly, all serpins except one (Pp1s86_83V6.1) identified here are predicted to be extracellular according to SecretomeP and only one splice variant (Pp1s61_265V6.4) has a signal peptide predicted by TargetP (Table S3 and S8). None of them was identified in the gametophore secretome,[16] which raises the question of the biological functions of serpins in *P. patens*. Although various functions are discussed for plant serpins, their exact biological role remains elusive.[92] However, in view of the function of AtSerpin1 they may act at least partially as inhibitors of secreted moss proteases. Thus, these serpins may be another target to further optimize yield and quality of biopharmaceuticals from moss bioreactors.

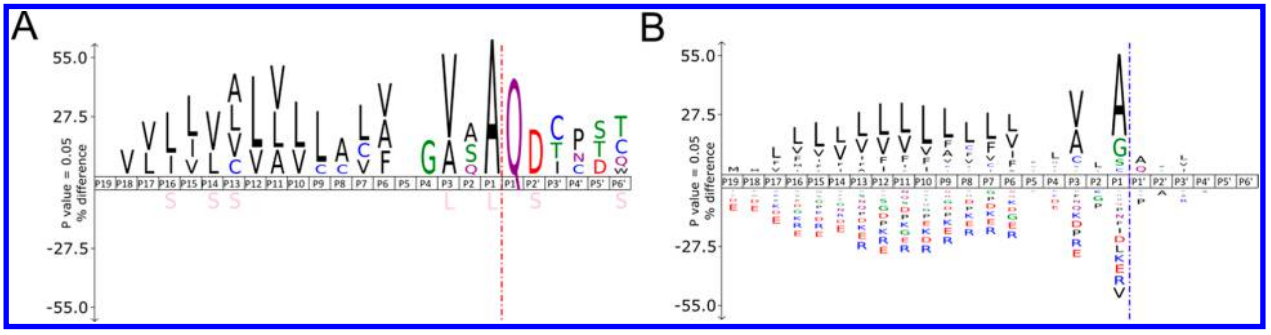


Figure 5. Sequence logos of identified and predicted signal peptide cleavage sites. (A) Sequence logo of the identified cleavage sites in the HCP ($N = 40$ unique sequences of 57 protein isoforms) using iceLogo.[54] Sequence window: 25 amino acids in total. All sequences were aligned at the identified cleavage site (P1′ position, red line). Reference: *P. patens* V1.6 gene models (cosmoss.org). The dashed line (red) indicates the cleavage site. (B) Sequence logo for all *P. patens* signal peptides predicted by TargetP[35,36] aligned at the predicted cleavage site (blue line, $N = 3987$ different sequences). All sequences are available in Table S12.

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

Identification of Potential Signal Peptide Cleavage Sites

Improving the yield of recombinant protein production implies a range of tasks, starting from engineering protein folding after protein synthesis, proceeding with engineering the intracellular trafficking of the protein, and ending up with improving the stability of the secreted protein in the supernatant by blocking proteolytic activity.[93,94] One approach in engineering the intracellular trafficking of the target protein along the secretory pathway is the selection of an appropriate signal peptide. Signal peptide sequences are conserved among eukaryotes,[95] and recombinant proteins bearing their native signal sequences are correctly processed in foreign host production systems.[96] However, more recent publications report differences in protein secretion efficiency by different signal peptides or by genetically engineered signal peptides.[97,98]

Here, we used the enrichment of N-terminal peptides to identify endogenous signal sequences in *P. patens* by comparing experimentally observed N-termini with predicted signal peptide cleavage sites (Table S11). Several endogenous signal sequences of *P. patens* have already been compared regarding their secretion efficiency.[99] Proteins corresponding to five signal peptides tested by[99] were identified in the present study (Pp1s93_72V6.2, Pp1s545_15V6.1, Pp1s139_88V6.1, Pp1s236_87V6.1, Pp1s81_246V6.1). This confirms that one of the sequences showing the highest secretion efficiency in a previous study[99] was that of Xyloglucan endotransglucosylase/hydrolase (Pp1s81_246V6.1), which was among the most abundant proteins in the HCP, as identified in the present study.

In total, we identified N-termini corresponding to predicted signal peptide cleavage sites for 57 protein isoforms (Table S12, 40 unique sequences). Sequence analysis using iceLogo[54] revealed a significant overrepresentation of glutamine as resulting N-terminal amino acid at the cleavage site (Figure 5a). The fraction of N-terminal glutamine was much higher in our experimentally observed cleavage sites than for all predicted signal peptide cleavage sites in the whole genome (Figure 5b). Nearly all of the N-terminal glutamine residues identified here were deamidated and cyclized to pyro-glutamate, a reaction that can either occur artificially after enzymatic digestion during sample preparation for MS[100] or enzymatically.[101] Here, the deamidation obviously occurred after cleavage of the signal peptide; otherwise, the N-terminus would have been dimethylated in most cases. This difference might be related to a potentially more stable N-terminus of extracellular proteins bearing glutamine as an N-terminal amino acid. In fact, human proteins are protected from exopeptidase degradation when bearing N-terminal pyro-glutamate.[102] Apart from its protective role, the modification did not affect the function of the tested proteins. Consequently, this might provide an interesting aspect for molecular engineering of signal peptides for the production of biopharmaceuticals in order to enhance their stability in the culture medium.

CONCLUSION

Here we present a detailed qualitative and quantitative analysis of the host cell proteome of moss bioreactors under production conditions. We find the secreted proteolytic network in the HCP to be far more complex than previous secretome studies in *P. patens* suggest. We also identified several potential protease-inhibitors, which were secreted by the production host. Targeting a single highly abundant protease made it possible to minimize the overall proteolytic activity in moss bioreactor supernatants. Based on our findings, several aspects for an improved production of recombinant biopharmaceuticals in moss can be addressed in the future, including the down-regulation of proteolytic activity in the supernatant and the optimization of product maturation along the secretory pathway.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[103] via the PRIDE partner repository with the data set identifier PXD009517.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00423.

1. Supplemental experimental procedures; 1.1. Calculation of SPKM-values: A detailed description how relative quantitative values for each observed protein isoform were calculated; 2. Supplemental Discussion; 2.1. Manual validation of the PCA performance: A detailed discussion using selected proteins about the clustering performance for classification ambiguities for each PCA-based cluster; 2.2. Protein isoforms originating from the same genomic locus appearing in different clusters: A detailed discussion of protein isoforms resulting from alternative splicing of the same initial transcript that were differentially classified in the PCA-based clustering; Figure S1: Overlap of proteins identified via SDS-PAGE and numbers of spectra assigned for proteins identified either in a single experiment or identified in more than one experiment; Figure S2: Section of a multiple sequence alignment for proteins with an assigned TAXI-domain and one aspartic protease; Figure S3: Identification and characterization of a subtilase knockout line (PDF)

Supplemental Tables: Table S1: This table contains all protein identification data exported from Scaffold 4 software for proteins identified from SDS-PAGE; Table S2: This table contains all protein identification data exported from Scaffold 4 software for proteins identified via enrichment of N-terminal peptides; Table S3: This table contains results of the data imputing and the principal component analysis with subsequent Mclust-clustering; Table S4: This table contains the results of the GO-enrichment analysis from the different cluster obtained from the PCA; Table S5: This table contains the results of the manual annotation of selected proteins of cluster 2; Table S6: This table contains the results of the manual annotation of selected proteins of cluster 1; Table S7: This table contains the results of the manual annotation of selected proteins of cluster 3; Table S8: This table contains all proteins where different isoforms were assigned to different clusters in the PCA; Table S9: This table contains the results of the PFAM-domain annotation of all proteins identified in the HCP; Table S10: This table contains all proteins with protease or protease-inhibiting domains and their cluster classification from the PCA; Table S11: This table contains only proteins with an identified N-terminus after enrichment of N-terminal peptides, the prediction values of TargetP for every protein and the calculated difference of an identified N-terminus to a predicted targeting peptide length; Table S12: This table contains all sequence used

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

for the sequence logos made with iceLogo software (XLSX)

AUTHOR INFORMATION

Corresponding Author

*Phone: +49 761 203 6969; fax: +49 761 203 6967; e-mail: ralf.reski@biologie.uni-freiburg.de.

ORCID

Ralf Reski: 0000-0002-5496-6711

Notes

The authors declare the following competing financial interest(s): R.R. is an inventor of the moss bioreactor and a founder of Greenovation Biotech. He currently serves as advisory board member of this company. T.F. is CEO and A.S. is CSO at Greenovation Biotech. B.F., B.B., and H.N. are staff members of this company. Greenovation Biotech develops and markets moss-based biopharmaceuticals.

ACKNOWLEDGMENTS

R.R. acknowledges financial support from the Excellence Initiative of the German Federal and State Governments (EXC 294). We thank Anne Katrin Prowse for proof reading the manuscript.

REFERENCES

(1) Chauhan, R.; Sood, N. Biopharmaceuticals: New yet Natural. *Br. Biotechnol. J.* **2016**, *14*, 1−19.

(2) Decker, E. L.; Reski, R. Moss bioreactors producing improved biopharmaceuticals. *Curr. Opin. Biotechnol.* **2007**, *18*, 393−398.

(3) Decker, E. L.; Reski, R. Glycoprotein production in moss bioreactors. *Plant Cell Rep.* **2012**, *31*, 453−460.

(4) Koprivova, A.; Stemmer, C.; Altmann, F.; Hoffmann, A.; Kopriva, S.; Gorr, G.; Reski, R.; Decker, E. L. Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic N-glycans. *Plant Biotechnol. J.* **2004**, *2*, 517−523.

(5) Weise, A.; Altmann, F.; Rodriguez-Franco, M.; Sjoberg, E. R.; Baumer, W.; Launhardt, H.; Kietzmann, M.; Gorr, G. High-level expression of secreted complex glycosylated recombinant human erythropoietin in the *Physcomitrella* Delta-fuc-t Delta-xyl-t mutant. *Plant Biotechnol. J.* **2007**, *5*, 389−401.

(6) Buettner-Mainik, A.; Parsons, J.; Jerome, H.; Hartmann, A.; Lamer, S.; Schaaf, A.; Schlosser, A.; Zipfel, P. F.; Reski, R.; Decker, E. L. Production of biologically active recombinant human factor H in. *Physcomitrella. Plant Biotechnol. J.* **2011**, *9*, 373−383.

(7) Parsons, J.; Altmann, F.; Arrenberg, C. K.; Koprivova, A.; Beike, A. K.; Stemmer, C.; Gorr, G.; Reski, R.; Decker, E. L. Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol. J.* **2012**, *10*, 851−861.

(8) Michelfelder, S.; Parsons, J.; Bohlender, L. L.; Hoernstein, S. N. W.; Niederkrüger, H.; Busch, A.; Krieghoff, N.; Koch, J.; Fode, B.; Frischmuth, T.; Schaaf, A.; Pohl, M.; Zipfel, P. F.; Reski, R.; Decker, E. L.; Häffner, K. Moss-based production of glyco-optimized and biologically active human complement factor H for therapeutic applications. *J. Am. Soc. Nephrol.* **2017**, *28*, 1462−1474.

(9) Wang, X.; Hunter, A. K.; Mozier, N. M. Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnol. Bioeng.* **2009**, *103*, 446−458.

(10) Tscheliessnig, A. L.; Konrath, J.; Bates, R.; Jungbauer, A. Host cell protein analysis in therapeutic protein bioprocessing - Methods and applications. *Biotechnol. J.* **2013**, *8*, 655−670.

(11) Schiermeyer, A.; Schinkel, H.; Apel, S.; Fischer, R.; Schillberg, S. Production of *Desmodus rotundus* salivary plasminogen activator α1 (DSPAα1) in tobacco is hampered by proteolysis. *Biotechnol. Bioeng.* **2005**, *89*, 848−858.

(12) Doran, P. M. Foreign protein degradation and instability in plants and plant tissue cultures. *Trends Biotechnol.* **2006**, *24*, 426−432.

(13) Kim, T. G.; Lee, H. J.; Jang, Y. S.; Shin, Y. J.; Kwon, T. H. M.; Yang, M. S. Co-expression of proteinase inhibitor enhances recombinant human granulocyte−macrophage colony stimulating factor production in transgenic rice cell suspension culture. *Protein Expression Purif.* **2008**, *61*, 117−121.

(14) Navarre, C.; De Muynck, B.; Alves, G.; Vertommen, D.; Magy, B.; Boutry, M. Identification, gene cloning and expression of serine proteases in the extracellular medium of *Nicotiana tabacum* cells. *Plant Cell Rep.* **2012**, *31*, 1959−1968.

(15) Mandal, M. K.; Fischer, R.; Schillberg, S.; Schiermeyer, A. Inhibition of protease activity by antisense RNA improves recombinant protein production in *Nicotiana tabacum* cv. Bright Yellow 2 (BY-2) suspension cells. *Biotechnol. J.* **2014**, *9*, 1065−1073.

(16) Lehtonen, M. T.; Takikawa, Y.; Rönnholm, G.; Akita, M.; Kalkkinen, N.; Ahola-Iivarinen, E.; Somervuo, P.; Varjosalo, M.; Valkonen, J. P. T. Protein secretome of moss plants (*Physcomitrella patens*) with emphasis on changes induced by a fungal elicitor. *J. Proteome Res.* **2014**, *13*, 447−459.

(17) Grosse-Holz, F.; Kelly, S.; Blaskowski, S.; Kaschani, F.; Kaiser, M.; van der Hoorn, R. A. L. The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire. *Plant Biotechnol. J.* **2018**, *16*, 1068−1084.

(18) Hohe, A.; Reski, R. Optimisation of a bioreactor culture of the moss *Physcomitrella patens* for mass production of protoplasts. *Plant Sci.* **2002**, *163*, 69−74.

(19) Hohe, A.; Decker, E. L.; Gorr, G.; Schween, G.; Reski, R. Tight control of growth and cell differentiation in photoautotrophically growing moss (*Physcomitrella patens*) bioreactor cultures. *Plant Cell Rep.* **2002**, *20*, 1135−1140.

(20) Dabrowska-Schlepp, P.; Sommerschuh, S.; Knappenberger, M.; Niederkruger, H.; Gliese, N.; Schaaf, A. Development of a novel host-cell protein assay: Supporting the *Physcomitrella patens* expression system. *BioProcess Int.* **2016**, *14*, 38−49.

(21) Niederkrüger, H.; Dabrowska-Schlepp, P.; Schaaf, A. Suspension culture of plant cells under phototrophic conditions. In *Industrial Scale Suspension Culture of Living Cells*; Meyer, H.-P., Schmidhalter, D. R., Eds.; Wiley-VCH: Weinheim, 2014; pp 259−292.

(22) Grosse, T., Niederkrüger, H.; Schaaf, A. Filtration of cell culture supernatants. *Patent* 2014, EP2687592A1.

(23) Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **1985**, *150*, 76−85.

(24) Hoernstein, S. N. W.; Mueller, S. J.; Fiedler, K.; Schuelke, M.; Vanselow, J. T.; Schuessele, C.; Lang, D.; Nitschke, R.; Igloi, G. L.; Schlosser, A.; Reski, R. Identification of targets and interaction partners of arginyl-tRNA protein transferase in the moss *Physcomitrella patens*. *Mol. Cell. Proteomics* **2016**, *15*, 1808−1822.

(25) Lange, P. F.; Huesgen, P. F.; Nguyen, K.; Overall, C. M. Annotating N-termini for the human proteome project: N-termini and N-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome. *J. Proteome Res.* **2014**, *13*, 2028−2044.

(26) McDonald, L.; Beynon, R. J. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat. Protoc.* **2006**, *1*, 1790−1798.

(27) Zimmer, A. D.; Lang, D.; Buchta, K.; Rombauts, S.; Nishiyama, T.; Hasebe, M.; Van de Peer, Y.; Rensing, S. A.; Reski, R. Reannotation and extended community resources for the genome of the non-seed plant *Physcomitrella patens* provide insights into the evolution of plant gene structures and functions. *BMC Genomics* **2013**, *14*, 498.

(28) Craig, R.; Beavis, R. C. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2310−2316.

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

(29) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383−5392.
(30) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646−4658.
(31) Chen, F.; Mackey, A. J.; Stoeckert, C. J., Jr.; Roos, D. S. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* **2006**, *34*, D363−D368.
(32) Fraley, C.; Raftery, A. E. Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *J. Classif.* **2003**, *20*, 263−286.
(33) Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R. B. Missing value estimation methods for DNA microarrays. *Bioinformatics* **2001**, *17*, 520−525.
(34) Alexa, A.; Rahnenführer, J.; Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **2006**, *22*, 1600−1607.
(35) Emanuelsson, O.; Brunak, S.; von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953−971.
(36) Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng., Des. Sel.* **1997**, *10*, 1−6.
(37) Chou, K. C. Using subsite coupling to predict signal peptides. *Protein Eng., Des. Sel.* **2001**, *14*, 75−79.
(38) Chou, K. C.; Shen, H. B. Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.* **2007**, *357*, 633−640.
(39) Shen, H. B.; Chou, K. C. Ensemble classifier for protein fold pattern recognition. *Bioinformatics* **2006**, *22*, 1717−1722.
(40) Shen, H. B.; Chou, K. C. Signal-3L: A 3-layer approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.* **2007**, *363*, 297−303.
(41) Bendtsen, J. D.; Jensen, L. J.; Blom, N.; von Heijne, G.; Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng., Des. Sel.* **2004**, *17*, 349−356.
(42) Jones, D. T.; Taylor, W. R.; Thornton, J. M. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **1994**, *33*, 3038−3049.
(43) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567−580.
(44) Tusnády, G. E.; Simon, I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **1998**, *283*, 489−506.
(45) Finn, R. D.; Clements, J.; Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29−W37.
(46) Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G. A.; Tate, J.; Bateman, A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279−D285.
(47) Yaari, R.; Noy-Malka, C.; Wiedemann, G.; Gershovitz, N. A.; Reski, R.; Katz, A.; Ohad, N. (2015) DNA Methyltransferase 1 is involved in mCG and mCCG DNA methylation and is essential for sporophyte development in *Physcomitrella patens*. *Plant Mol. Biol.* **2015**, *88*, 387−400.
(48) Richardt, S.; Timmerhaus, G.; Lang, D.; Qudeimat, E.; Correa, L. G. G.; Reski, R.; Rensing, S. A.; Frank, W. Microarray analysis of the moss *Physcomitrella patens* reveals conserved transcriptional regulation of salt stress and abscisic acid signalling. *Plant Mol. Biol.* **2010**, *72*, 27−45.
(49) Hiss, M.; Laule, O.; Meskauskiene, R. M.; Arif, M. A.; Decker, E. L.; Erxleben, A.; Frank, W.; Hanke, S. T.; Lang, D.; Martin, A.; Neu, C.; Reski, R.; Richard, S.; Schallenberg-Rüdinger, M.; Szövényi, P.; Theodhor, T.; Wiedemann, G.; Wolf, L.; Zimmermann, P.; Rensing, S. A. Large-scale gene expression profiling data for the model moss *Physcomitrella patens* aid understanding of developmental progression, culture and stress conditions. *Plant J.* **2014**, *79*, 530−539.
(50) Beike, A. K.; Lang, D.; Zimmer, A. D.; Wüst, F.; Trautmann, D.; Wiedemann, G.; Beyer, P.; Decker, E. L.; Reski, R. Insights from the cold transcriptome of *Physcomitrella patens*: global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation. *New Phytol.* **2015**, *205*, 869−881.
(51) Frank, W.; Decker, E. L.; Reski, R. Molecular tools to study *Physcomitrella patens*. *Plant Biol.* **2005**, *7*, 220−227.
(52) Strepp, R.; Scholz, S.; Kruse, S.; Speth, V.; Reski, R. Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4368−4373.
(53) Hulsen, T.; de Vlieg, J.; Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **2008**, *9*, 488.
(54) Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **2009**, *6*, 786−787.
(55) Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* **2010**, *31*, 3573−3579.
(56) Jost, W.; Knappenberger, M.; Claußnitzer, D.; Schaaf, A. Method for increasing the secretion of recombinant proteins. *Patent* EP 2653548A1, 2013.
(57) Mandal, M. K.; Ahvari, H.; Schillberg, S.; Schiermeyer, A. Tackling unwanted proteolysis in plant production hosts used for molecular farming. *Front. Plant Sci.* **2016**, *7*, 267.
(58) Rawlings, N. D.; Barrett, A. J.; Finn, R. D. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2016**, *44*, D343−D350.
(59) Sansen, S.; Ranter, C. J. D.; Gebruers, K.; Brijs, K.; Courtin, C. M.; Delcour, J. A.; Rabijns, A. Structural basis for inhibition of *Aspergillus niger* xylanase by *Triticum aestivum* xylanase inhibitor-I. *J. Biol. Chem.* **2004**, *279*, 36022−36028.
(60) Walton, J. D. Deconstructing the cell wall. *Plant Physiol.* **1994**, *104*, 1113−1118.
(61) Guo, R.; Xu, X.; Carole, B.; Li, X.; Gao, M.; Zheng, Y.; Wang, X. Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. *BMC Genomics* **2013**, *14*, 554.
(62) Kay, J. Aspartic proteinases and their inhibitors. *Biochem. Soc. Trans.* **1985**, *13*, 1027−1029.
(63) Fierens, K.; Brijs, K.; Courtin, C. M.; Gebruers, K.; Goesaert, H.; Raedschelders, G.; Robben, J.; Van Campenhout, S.; Volckaert, G.; Delcour, J. A. Molecular identification of wheat endoxylanase inhibitor TAXI-I1, member of a new class of plant proteins. *FEBS Lett.* **2003**, *540*, 259−263.
(64) Faro, C.; Gal, S. Aspartic proteinase content of the *Arabidopsis* genome. *Curr. Protein Pept. Sci.* **2005**, *6*, 493−500.
(65) Ge, X.; Dietrich, C.; Matsuno, M.; Li, G.; Berg, H.; Xia, Y. An *Arabidopsis* aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. *EMBO Rep.* **2005**, *6*, 282−288.
(66) Yao, X.; Xiong, W.; Ye, T.; Wu, Y. Overexpression of the aspartic protease ASPG1 gene confers drought avoidance in *Arabidopsis*. *J. Exp. Bot.* **2012**, *63*, 2579−2593.
(67) Kervinen, J.; Tobin, G. J.; Costa, J.; Waugh, D. S.; Wlodawer, A.; Zdanov, A. Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO J.* **1999**, *18*, 3947−3955.
(68) Törmäkangas, K.; Hadlington, J. L.; Pimpl, P.; Hillmer, S.; Brandizzi, F.; Teeri, T. H.; Denecke, J. A vacuolar sorting domain may also influence the way in which proteins leave the endoplasmic reticulum. *Plant Cell* **2001**, *13*, 2021−2032.
(69) Bryksa, B. C.; Grahame, D. A.; Yada, R. Y. Comparative structure-function characterization of the saposin-like domains from potato,

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX

barley, cardoon and *Arabidopsis* aspartic proteases. *Biochim. Biophys. Acta, Biomembr.* **2017**, *1859*, 1008−1018.
(70) Schaaf, A.; Reski, R.; Decker, E. L. A novel aspartic proteinase is targeted to the secretory pathway and to the vacuole in the moss *Physcomitrella patens*. *Eur. J. Cell Biol.* **2004**, *83*, 145−152.
(71) Goulet, C.; Khalf, M.; Sainsbury, F.; D'Aoust, M. A.; Michaud, D. A protease activity−depleted environment for heterologous proteins migrating towards the leaf cell apoplast. *Plant Biotechnol. J.* **2012**, *10*, 83−94.
(72) Niemer, M.; Mehofer, U.; Torres Acosta, J. A.; Verdianz, M.; Henkel, T.; Loos, A.; Strasser, R.; Maresch, D.; Rademacher, T.; Steinkellner, H.; Mach, L. The human anti-HIV antibodies 2F5, 2G12, and PG9 differ in their susceptibility to proteolytic degradation: Down-regulation of endogenous serine and cysteine proteinase activities could improve antibody production in plant-based expression platforms. *Biotechnol. J.* **2014**, *9*, 493−500.
(73) Mandal, M. K.; Fischer, R.; Schillberg, S.; Schiermeyer, A. Biochemical properties of the matrix metalloproteinase NtMMP1 from *Nicotiana tabacum* cv. BY-2 suspension cells. *Planta* **2010**, *232*, 899−910.
(74) Schaller, A.; Stintzi, A.; Graff, L. Subtilases − versatile tools for protein turnover, plant development, and interactions with the environment. *Physiol. Plant.* **2012**, *145*, 52−66.
(75) Rautengarten, C.; Steinhauser, D.; Büssis, D.; Stintzi, A.; Schaller, A.; Kopka, J.; Altmann, T. Inferring hypotheses on functional relationships of genes: Analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Comput. Biol.* **2005**, *1*, e40.
(76) Bayer, E. M.; Bottrill, A. R.; Walshaw, J.; Vigouroux, M.; Naldrett, M. J.; Thomas, C. L.; Maule, A. J. *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. *Proteomics* **2006**, *6*, 301−311.
(77) Cheng, F.; Blackburn, K.; Lin, Y.; Goshe, M. B.; Williamson, J. D. Absolute protein quantification by LC/MSE for global analysis of salicylic acid-induced plant protein secretion responses. *J. Proteome Res.* **2009**, *8*, 82−93.
(78) Kim, S. G.; Wang, Y.; Lee, K. H.; Park, Z. Y.; Park, J.; Wu, J.; Kwon, S. J.; Lee, Y. H.; Agrawal, G. K.; Rakwal, R.; Kim, S. T.; Kang, K. Y. In-depth insight into *in vivo* apoplastic secretome of rice-*Magnaporthe oryzae* interaction. *J. Proteomics* **2013**, *78*, 58−71.
(79) Ma, W.; Muthreich, N.; Liao, C.; Franz-Wachtel, M.; Schütz, W.; Zhang, F.; Hochholdinger, F.; Li, C. The mucilage proteome of maize (*Zea mays L.*) primary roots. *J. Proteome Res.* **2010**, *9*, 2968−2976.
(80) Chichkova, N. V.; Kim, S. H.; Titova, E. S.; Kalkum, M.; Morozov, V. S.; Rubtsov, Y. P.; Kalinina, N. O.; Taliansky, M. E.; Vartapetian, A. B. A plant caspase-like protease activated during the hypersensitive response. *Plant Cell* **2004**, *16*, 157−171.
(81) Vartapetian, A. B.; Tuzhikov, A. I.; Chichkova, N. V.; Taliansky, M.; Wolpert, T. J. A plant alternative to animal caspases: subtilisin-like proteases. *Cell Death Differ.* **2011**, *18*, 1289−1297.
(82) Schardon, K.; Hohl, M.; Graff, L.; Pfannstiel, J.; Schulze, W.; Stintzi, A.; Schaller, A. Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science* **2016**, *354*, 1594−1597.
(83) Reski, R.; Parsons, J.; Decker, E. L. Moss-made pharmaceuticals: from bench to bedside. *Plant Biotechnol. J.* **2015**, *13*, 1191−1198.
(84) Baur, A.; Reski, R.; Gorr, G. Enhanced recovery of a secreted recombinant human growth factor using stabilizing additives and by co-expression of human serum albumin in the moss *Physcomitrella patens*. *Plant Biotechnol. J.* **2005**, *3*, 331−340.
(85) Komarnytsky, S.; Borisjuk, N.; Yakoby, N.; Garvey, A.; Raskin, I. Cosecretion of protease inhibitor stabilizes antibodies produced by plant roots. *Plant Physiol.* **2006**, *141*, 1185−1193.
(86) Goulet, C.; Benchabane, M.; Anguenot, R.; Brunelle, F.; Khalf, M.; Michaud, D. A companion protease inhibitor for the protection of cytosol-targeted recombinant proteins in plants. *Plant Biotechnol. J.* **2010**, *8*, 142−154.
(87) Roberts, T. H.; Hejgaard, J. Serpins in plants and green algae. *Funct. Integr. Genomics* **2008**, *8*, 1−27.
(88) Vercammen, D.; Belenghi, B.; Cotte, B.; van de Beunens, T.; Gavigan, J.-A.; De Rycke, R.; Brackenier, A.; Inzé, D.; Harris, J. L.; Van Breusegem, F. Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J. Mol. Biol.* **2006**, *364*, 625−636.
(89) Lampl, N.; Alkan, N.; Davydov, O.; Fluhr, R. Set-point control of RD21 protease activity by AtSerpin1 controls cell death in *Arabidopsis*. *Plant J.* **2013**, *74*, 498−510.
(90) van der Hoorn, R. A. L. Plant proteases: From phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* **2008**, *59*, 191−223.
(91) Misas-Villamil, J. C.; van der Hoorn, R. A. L.; Doehlemann, G. Papain-like cysteine proteases as hubs in plant immunity. *New Phytol.* **2016**, *212*, 902−907.
(92) Fluhr, R.; Lampl, N.; Roberts, T. H. Serpin protease inhibitors in plant biology. *Physiol. Plant.* **2012**, *145*, 95−102.
(93) Faye, L.; Boulaflous, A.; Benchabane, M.; Gomord, V.; Michaud, D. Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* **2005**, *23*, 1770−1778.
(94) Idiris, A.; Tohda, H.; Kumagai, H.; Takegawa, K. Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 403−417.
(95) Gitzinger, M.; Parsons, J.; Reski, R.; Fussenegger, M. Functional cross-kingdom conservation of mammalian and moss (*Physcomitrella patens*) transcription, translation and secretion machineries. *Plant Biotechnol. J.* **2009**, *7*, 73−86.
(96) Matsumoto, S.; Ikura, K.; Ueda, M.; Sasaki, R. Characterization of a human glycoprotein (erythropoietin) produced in cultured tobacco cells. *Plant Mol. Biol.* **1995**, *27*, 1163−1172.
(97) Vadhana, A. K. P.; Samuel, P.; Berin, R. M.; Krishna, J.; Kamatchi, K.; Meenakshisundaram, S. Improved secretion of *Candida antarctica* lipase B with its native signal peptide in *Pichia pastoris*. *Enzyme Microb. Technol.* **2013**, *52*, 177−183.
(98) Ng, D. T. W.; Sarkar, C. A. Engineering signal peptides for enhanced protein secretion from *Lactococcus lactis*. *Appl. Environ. Microbiol.* **2013**, *79*, 347−356.
(99) Schaaf, A.; Tintelnot, S.; Baur, A.; Reski, R.; Gorr, G.; Decker, E. L. Use of endogenous signal sequences for transient production and efficient secretion by moss (*Physcomitrella patens*) cells. *BMC Biotechnol.* **2005**, *5*, 30.
(100) Khandke, K. M.; Fairwell, T.; Chait, B. T.; Manjula, B. N. Influence of ions on cyclization of the amino terminal glutamine residues of tryptic peptides of streptococcal PepM49 protein. Resolution of cyclized peptides by HPLC and characterization by mass spectrometry. *Int. J. Pept. Protein Res.* **1989**, *34*, 118−123.
(101) Schilling, S.; Wasternack, C.; Demuth, H.-U. Glutaminyl cyclases from animals and plants: a case of functionally convergent protein evolution. *Biol. Chem.* **2008**, *389*, 983−991.
(102) Coillie, E.; Van; Proost, P.; Van Aelst, I.; Struyf, S.; Polfliet, M.; De Meester, I.; Harvey, D. J.; Van Damme, J.; Opdenakker, G. Functional comparison of two human monocyte chemotactic protein-2 isoforms, role of the amino-terminal pyroglutamic acid and processing by CD26/dipeptidyl peptidase IV. *Biochemistry* **1998**, *37*, 12672−12680.
(103) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P. A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H. J.; Albar, J. P.; Martinez-Bartolomé, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223−226.

DOI: 10.1021/acs.jproteome.8b00423
J. Proteome Res. XXXX, XXX, XXX−XXX