SUPPORTING INFORMATION

Native structural and functional proteoform characterization of the prolyl-alanyl-specific endoprotease EndoPro from *Aspergillus niger*

Guusje van Schaick[1], Elena Domínguez-Vega[1], Christoph Gstöttner[1], Johanna H. van den Berg-Verleg[2], Olaf Schouten[2], Michiel Akeroyd[2], Maurien M.A. Olsthoorn[2], Manfred Wuhrer[1], Albert J. R. Heck[3,4], Nicolas Abello[2], Vojtech Franc[3,4 *]

[1] Leiden University Medical Center, Center for Proteomics and Metabolomics, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands

[2] DSM Biotechnology Center, Center for Enabling Innovation, Delft, Alexander Fleminglaan 1, 2613 AX, the Netherlands

[3] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH, Utrecht, the Netherlands.

[4] Netherlands Proteomics Center, Padualaan 8, 3584 CH, Utrecht, the Netherlands

* corresponding author

E-mail: vfranc@proqr.com

Table of contents

Figure S1.	Sequence of EndoPro	S-3
Figure S2.	Analysis of EndoPro using LC-MS and gel-based approaches	S-4
Figure S3.	AEX-UV chromatograms of mobile phase optimization	S-5
Figure S4.	Charge variant profile of EndoPro by AEX-UV and AEX-MS	S-6
Figure S5.	Zero charge deconvoluted native mass spectra of AEX peaks	S-7
Figure S6.	Assigned MALDI-FTICR mass spectra of released N-glycans	S-8
Figure S7.	EICs and MS/MS spectra of phospho- and neutral glycans	S-9
Figure S8	Glycoproteomic analysis of AEX fractions	S-10
Figure S9.	Gel of CytC digestion by EndoPro	S-11
Figure S10.	BSA peptide characteristics after digestion with AEX fractions	S-12
Figure S11.	C-terminal ragging of six BSA peptides	S-13

M R A F S A V A A A A L A L S W A S L A Q A A R P R L V P K P V S R P A S S K S
A A T T G E A Y F E Q L L D H H N P E K G T F S Q R Y W W S T E Y W G G P G S P
V V L F T P G E V S A D G Y E G Y L T N E T L T G V Y A Q E I Q G A V I L I E H
R Y W G D S S P Y E V L N A E T L Q Y L T L D Q A I L D M T Y F A E T V K L Q F
D N S T R S N A Q N A P W V M V G G S Y S G A L T A W T E S V A P G T F W A Y H
A T S A P V E A I Y D Y W Q Y F Y P I Q Q G M A Q N C S K D V S L V A E Y V D K
I G K N G T A K E Q Q A L K E L F G L G A V E H F D D F A A V L P N G P Y L W Q
D N D F A T G Y S S F F Q F C D A V E G V E A G A A V T P G P E G V G L E K A L
A N Y A N W F N S T I L P D Y C A S Y G Y W T D E W S V A C F D S Y N A S S P I
Y T D T S V G N A V D R Q W E W F L C N E P F F Y W Q D G A P E G T S T I V P R
L V S A S Y W Q R Q C P L Y F P E T N G Y T Y G S A K G K N A A T V N S W T G G
W D M T R N T T R L I W T N G Q Y D P W R D S G V S S T F R P G G P L A S T A N
E P V Q I I P G G F H C S D L Y M A D Y Y A N E G V K K V V D N E V K Q I K E W
V E E Y Y A

Figure S1. EndoPro sequence. Amino acid composition of EndoPro. Color code shows in purple the signal peptide, in green the known N-terminal truncation, in red the N-glycosylation sequons, and in blue the cysteines involved in disulfide bridges.

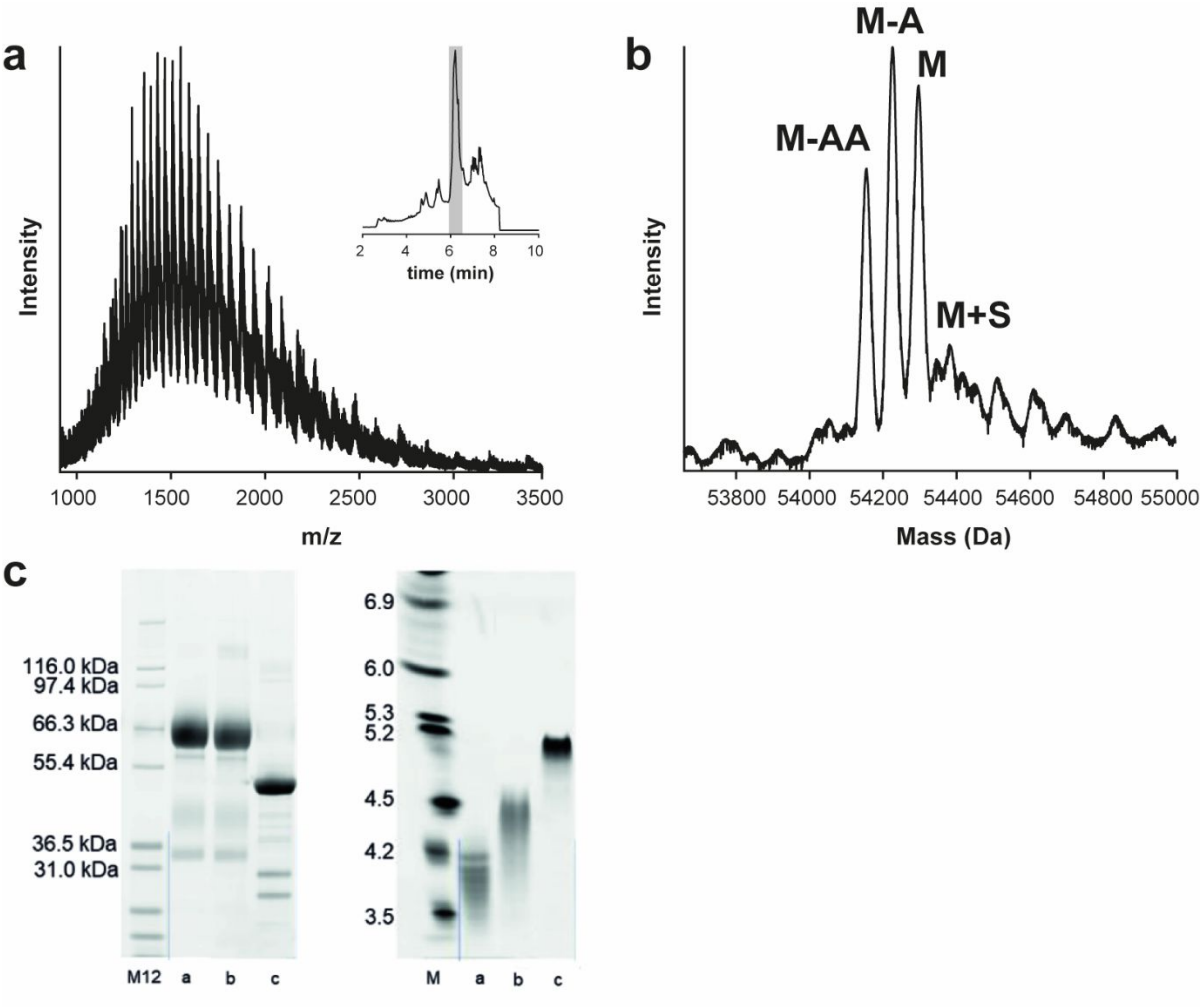

Figure S2. Characterization of EndoPro using LC-MS and gel-based separation approaches. (**a**) Mass spectrum of deglycosylated EndoPro recorded using RPLC-MS. The inset shows the total ion chromatogram of the RPLC measurement. The indicated peak is deglycosylated EndoPro. (**b**) Zero-charge deconvoluted mass spectrum of the spectrum shown in "a". The presence of the variants with different N-terminal processing is indicated. (**c**) SDS-PAGE (left) and IEF gel (right) of EndoPro reference sample (lane a), blank incubation of EndoPro without PNGase F (enzyme blank, lane b), and incubation of EndoPro with PNGase F (lane c). In the SDS-PAGE gel, the main band is the mature EndoPro protein, while the other bands correspond to either background proteins or autolysis products.

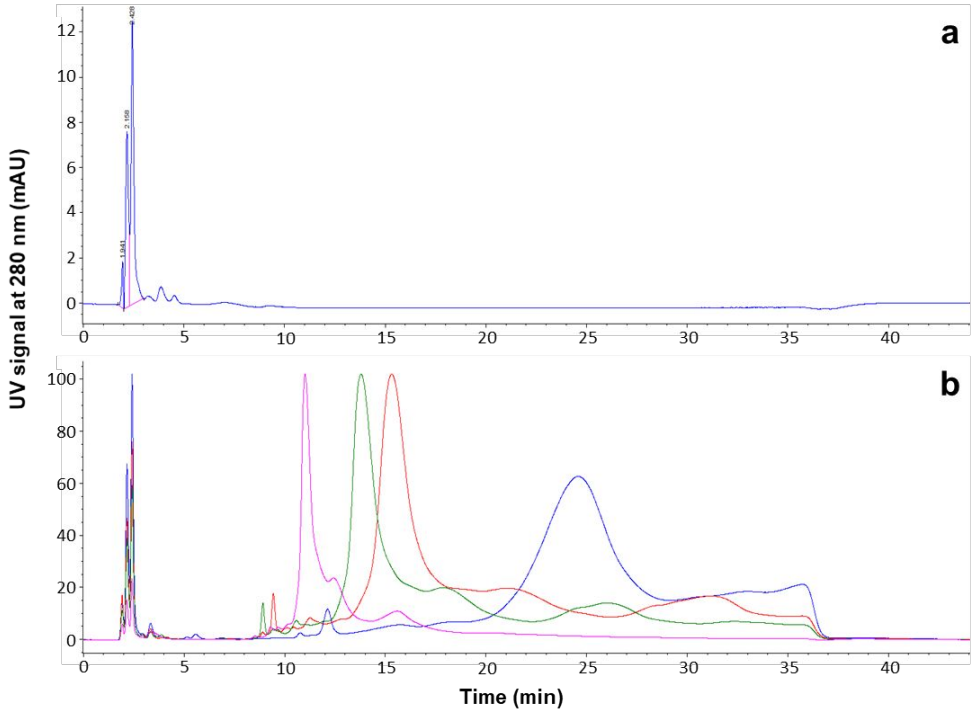

Figure S3. AEX method optimization. (**a**) UV chromatogram obtained using a pH gradient from 50 mM ammonium acetate pH 5.5 to pH 3.0. (**b**) Overlay of UV chromatograms using different concentrations of ammonium formate. All chromatograms use a pH gradient from pH 5.5 to pH 2.5 for elution. Blue is 25 mM, red is 50 mM, green is 60 mM and pink is 75 mM.

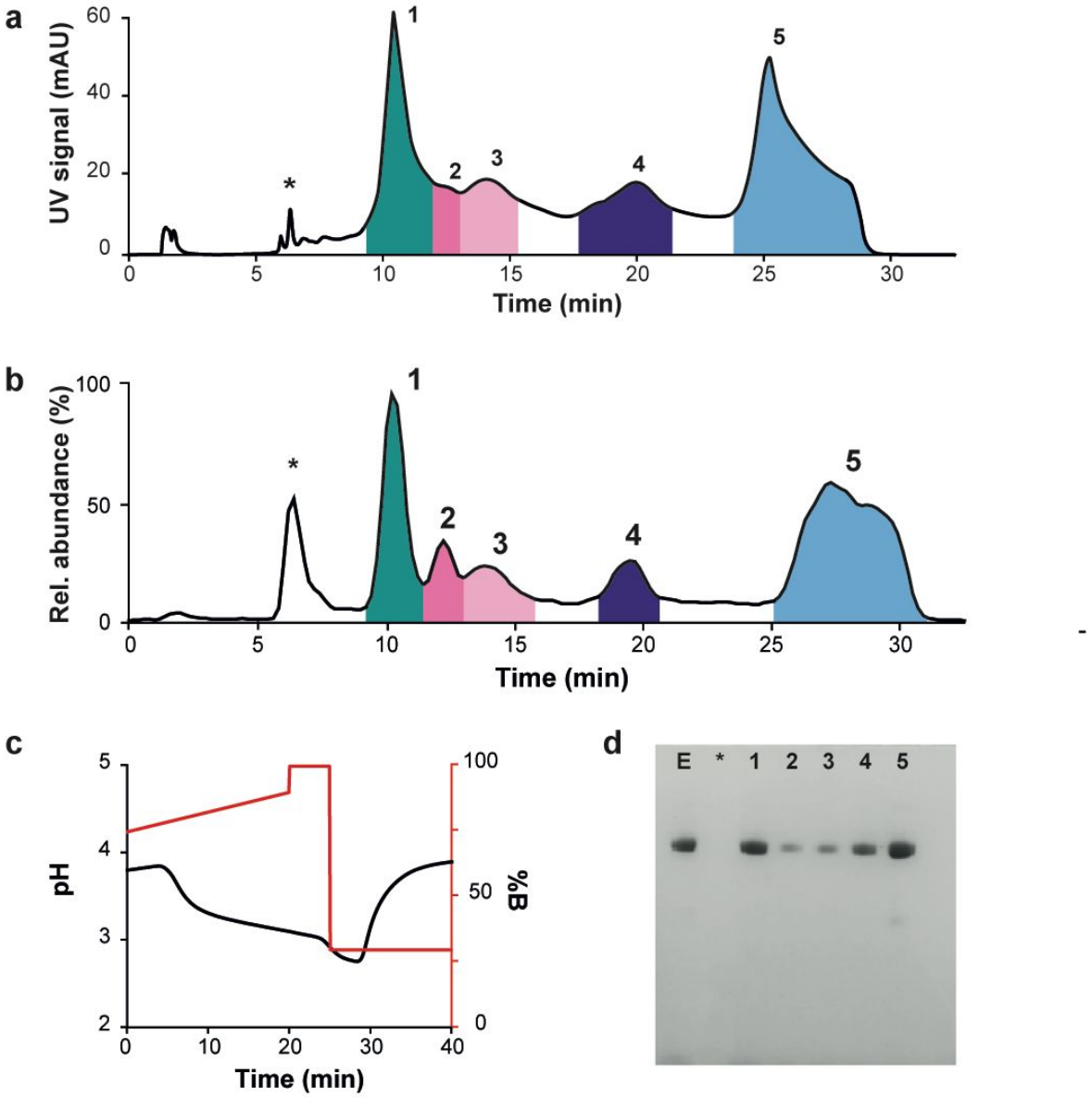

Figure S4. Chromatographic separation of EndoPro. (**a**) UV chromatogram of anion-exchange chromatography AEX method. The peak marked with * is an unknown (most likely a non-proteinaceous) species. (**b**) Base Peak Chromatogram of the AEX-MS method. (**c**) Online measured pH gradient of the AEX method (black trace) compared with the programmed gradient (red trace). (**d**) Reducing gel, where lanes represent EndoPro or one of the collected AEX peaks indicated in the UV chromatogram.

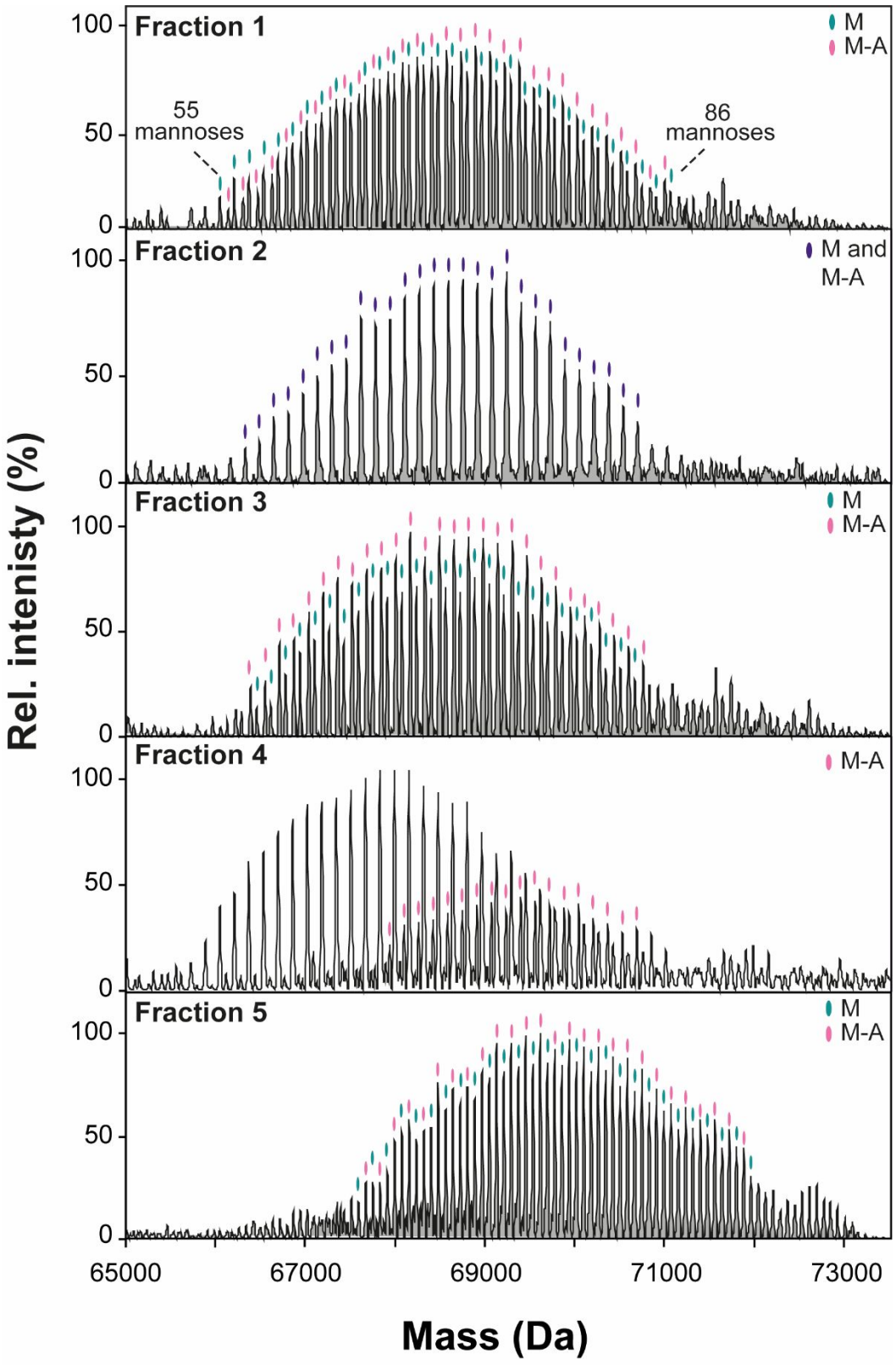

Figure S5. Zero charge deconvoluted native mass spectra of the peaks of the AEX-MS measurement. Where possible, the different sequence variants are indicated (*i.e.,* M and M-A). For the UV chromatogram and the BPC of the AEX separation, see **Figure S4**.

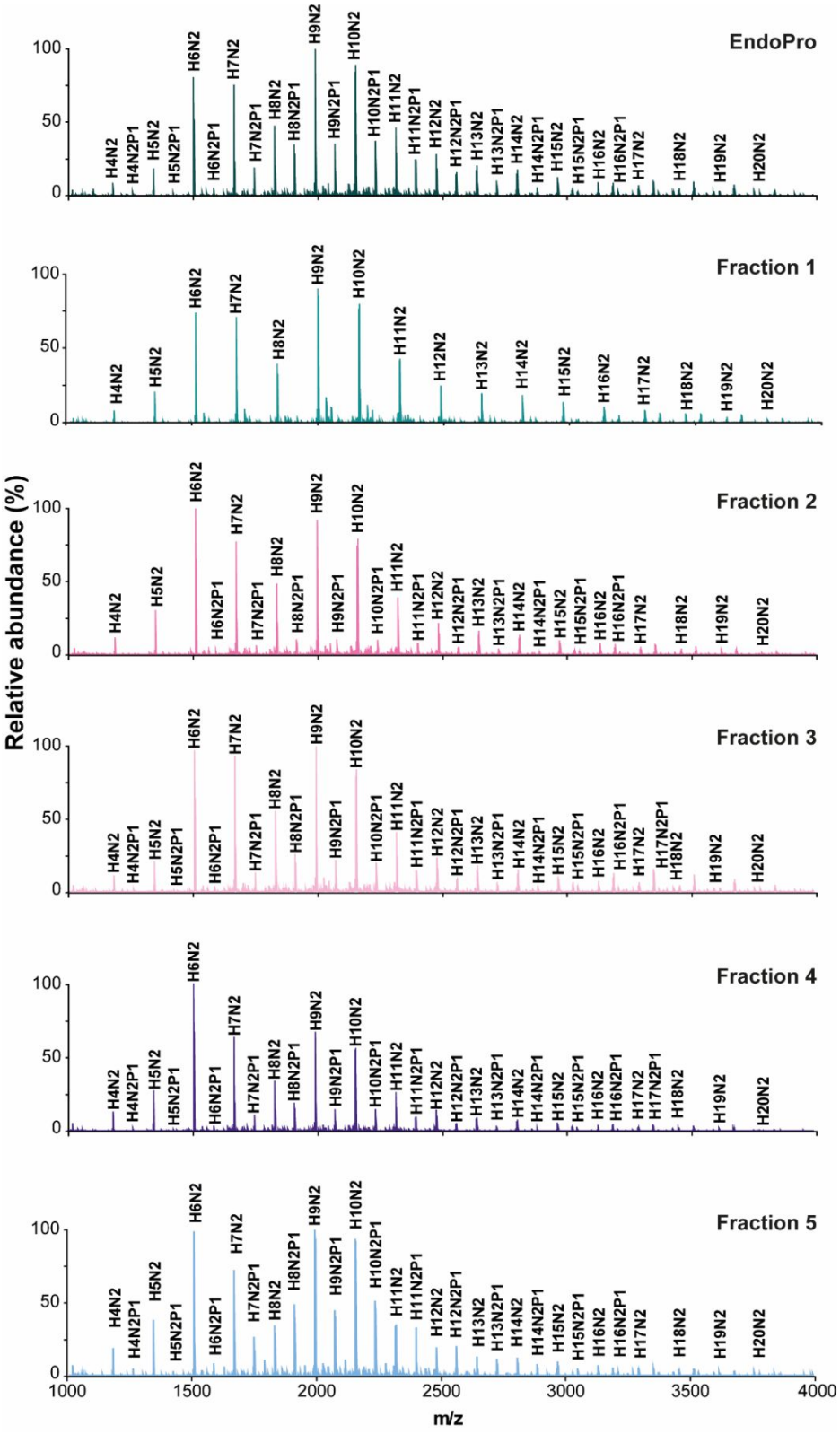

Figure S6. EndoPro released N-glycan analysis. Assigned MALDI-FTICR mass spectra of released N-glycans of nonseparated EndoPro and the AEX fractions (Fraction 1 in green, Fraction 2 in pink, Fraction 3 in pale pink, Fraction 4 in purple, and Fraction 5 in blue). Only

high mannose type N-glycans were detected for both fractions. Fraction 1 contains no phosphoglycans, whereas the other fractions contain these glycoforms.

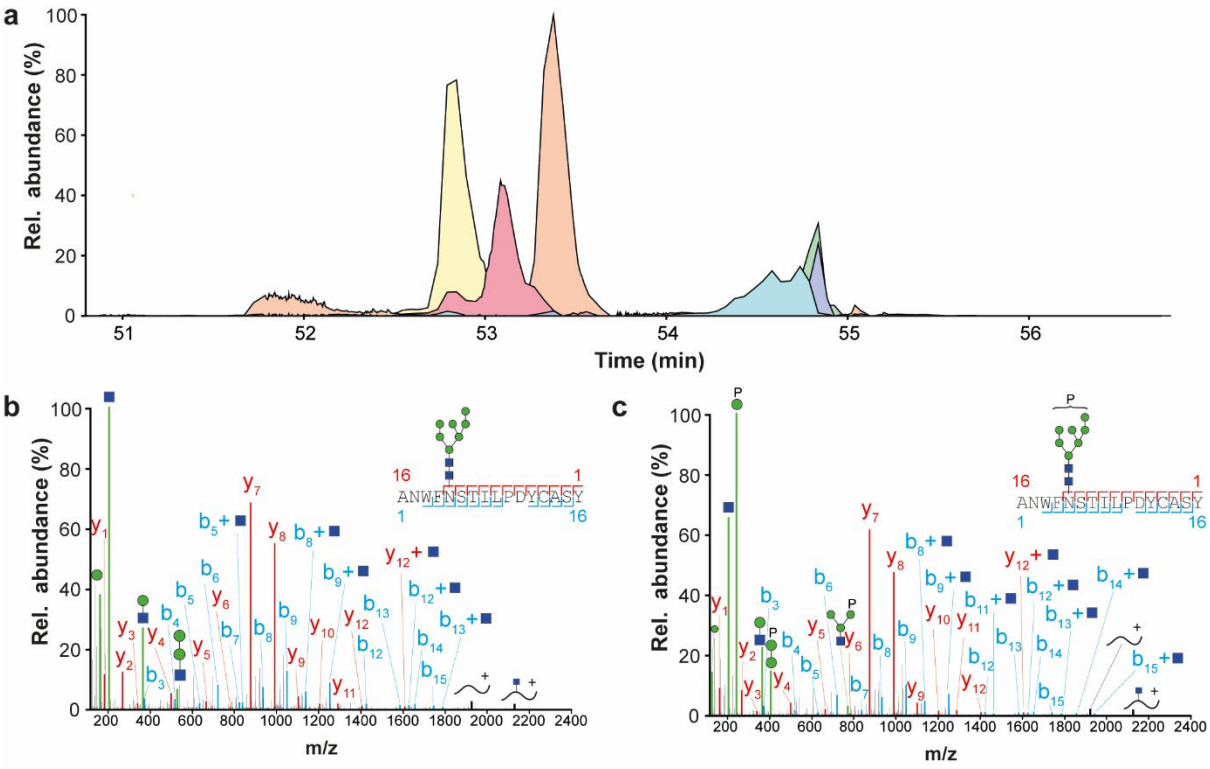

Figure S7. Illustrative glycopeptide LC-MS/MS data. (**a**) Extracted Ion chromatograms (EICs) of peptide ANWFNSTILPDYCASY (site N288) containing different glycoforms (*i.e.*, orange is H7N2, red is H8N2, yellow is H9N2, purple is H7N2P1, green is H8N2P1 and blue is H9N2P1). (**b**) HCD fragmentation spectrum of glycoform H7N2 from peptide displayed in EIC. (**c**) HCD fragmentation spectrum of glycoform H7N2P1 from peptide displayed in EIC.

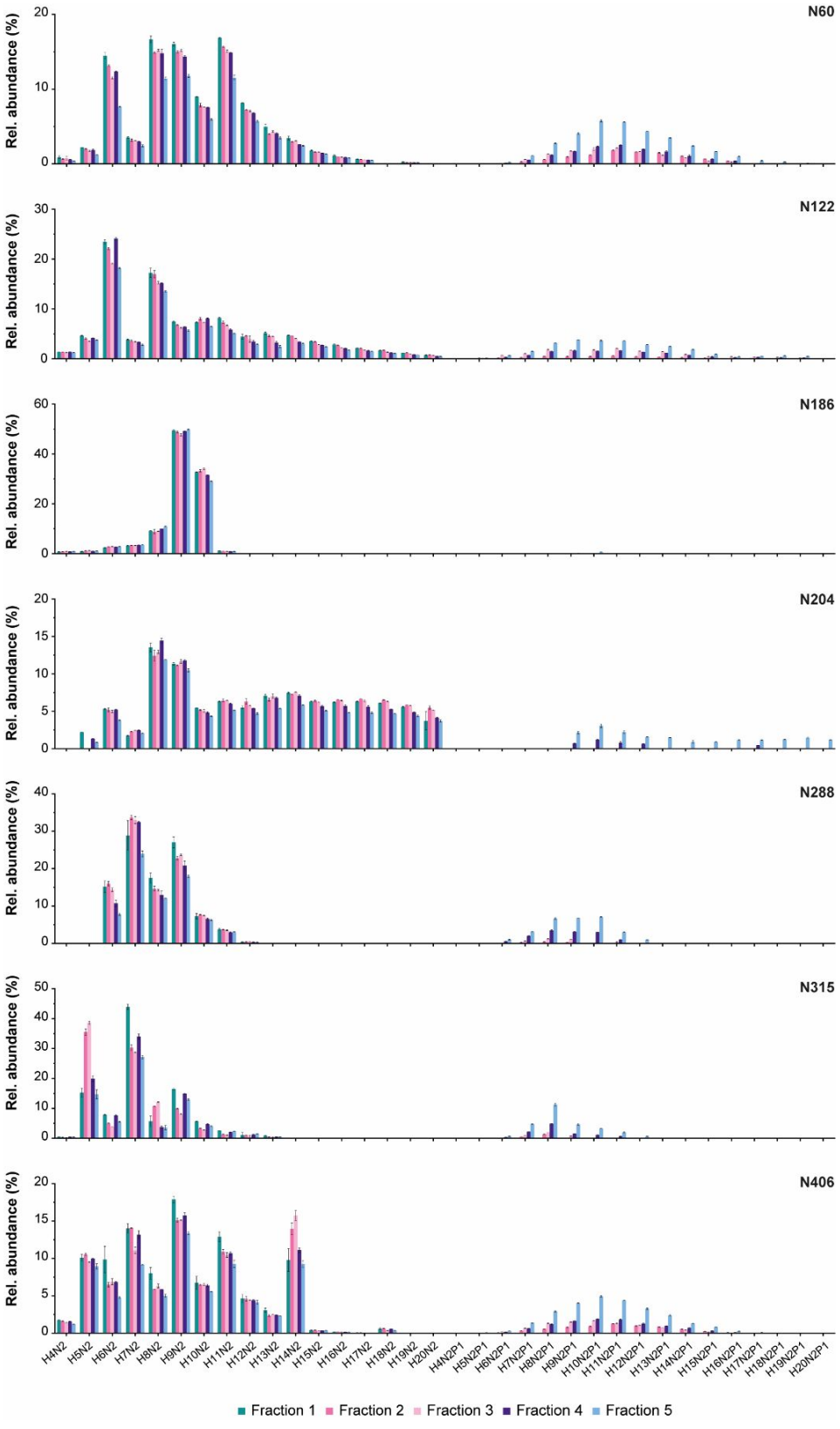

Figure S8. Peptide-centric glycoproteomic analysis of the AEX fractions for profiling of the site-specific glycosylation. For each glycosylation site (*i.e.,* N60, N122, N186, N204, N288, N315, and N406), the glycopeptide profile is shown. Samples were measured in duplicate and

the error bars show standard deviation. The relative abundance of each glycoform can be found in **Table S4.**

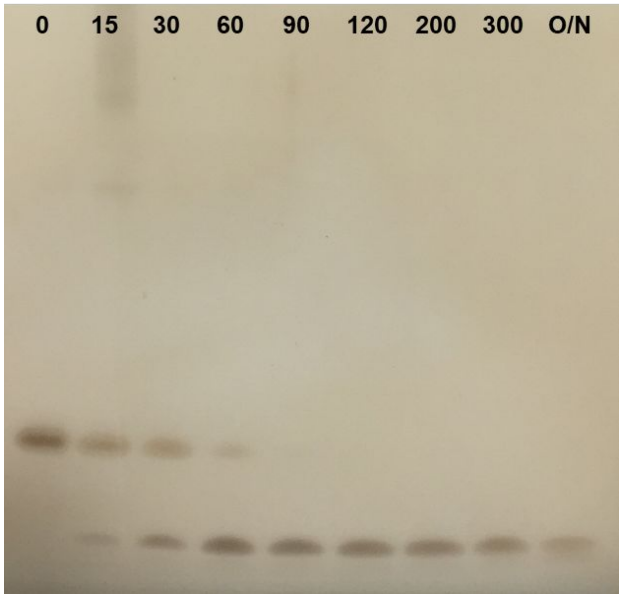

Figure S9. Time course of CytC digestion by EndoPro. Digestion of CytC by EndoPro was performed at different time points (*i.e.,* 0, 15, 30, 60, 90, 120, 200, 300, and 480 min) measured by SDS-PAGE analysis.

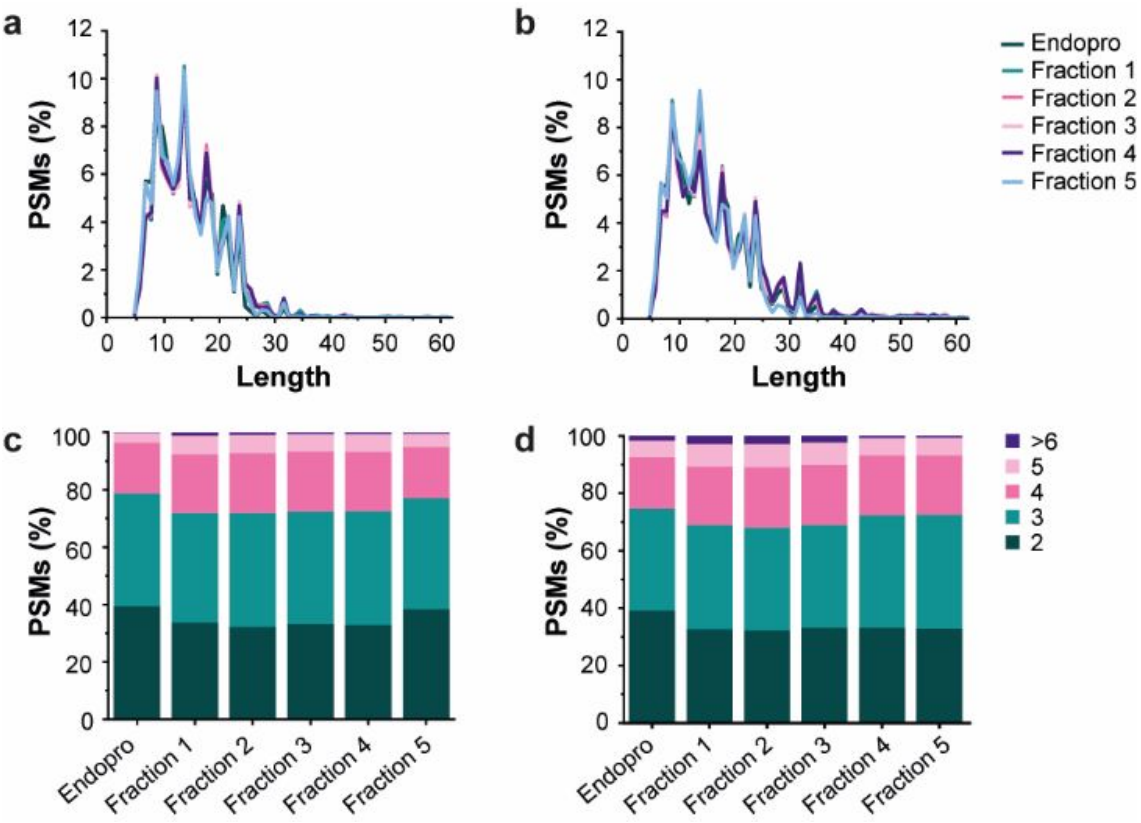

Figure S10. Digestion of BSA with EndoPro. The peptide length distribution after digestion of BSA with EndoPro or the AEX fractions at pH 2 (**a**) and pH 4 (**b**). The peptide charge distribution after digestion with EndoPro or the AEX fractions at pH 2 (**c**) and pH 4 (**d**).

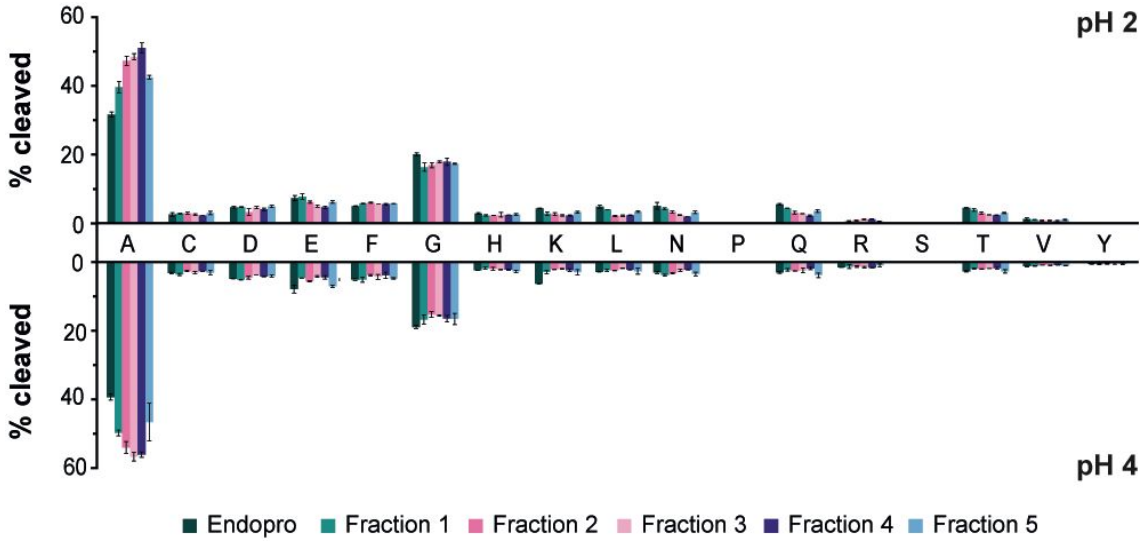

Figure S11. EndoPro cleavage pattern. The C-terminal amino acids from six BSA peptides generated by digestion EndoPro or the AEX fractions (EndoPro proteoforms). These peptides showed C-terminal ragging. The data is a mean percentage of two measurements and the error bars show standard deviations. The peptides: "FDEHVKLVNELTEFAKTCVADESHAG", "ELLYYANKYNGVFQECCQAEDKGA", "EVTKLVTDLTKVHKECCHGDLLECADDRA", "EDKDVCKNYQEAKDAFLG", "QNLIKQNCDQFEKLGEYGFQNALIVRY", "TEEQLKTVMENFVAFVDKCCAA".