Supplementary Figures 1-13

Neprosin, a selective prolyl endoprotease for bottom-up proteomics and histone mapping

Christoph U. Schräder[1], Linda Lee[1], Martial Rey[1], Vladimir Sarpe[1], Petr Man[2,3], Seema Sharma[4], Vlad Zabrouskov[4], Brett Larsen[5], David C. Schriemer[1,6]*

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada

[2]BioCev - Institute of Microbiology, Czech Academy of Sciences, Vestec, Czech Republic

[3]Department of Biochemistry, Faculty of Science, Charles University in Prague, Prague, Czech Republic

[4]Thermo Fisher Scientific, San Jose, California, United States

[5]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada

[6]Department of Chemistry, University of Calgary, Calgary, Alberta, Canada

*Corresponding author:

David C. Schriemer, Ph.D.

Department of Biochemistry and Molecular Biology

The University of Calgary, Room 300 Heritage Medical Research Building

3330 Hospital Drive NW

Calgary, Alberta, Canada

T2N4N1

Email: dschriem@ucalgary.ca

Running title: A Post-Pro-cleaving enzyme for extended bottom-up proteomics

Fig. S1. Overview of Pro-cleaving protease families together with their domain structure, size, optimal pH and cleavage specificity.

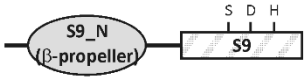
Protease Family	Name	Domain Organization	Species	MW (kDa)	Optimal pH	Substrate specificity
DUF239	Neprosin (Npr1)		*Nepenthes ventrata*	29	2-4	Pro-X
S9	Prolyl oligopeptidase (POP)		*Pyrococcus furiosus, Myxococcus xanthus, Flavobacterium meningosepticum*	70-80	6-8	C-terminal Pro-X, oligopeptide
S28	Propyl endoprotease (PEP)		*Aspergillus niger*	50-60	4-5	Pro-X

Fig. S2. Autolytic processing and activation of recombinant neprosin under acidic condition. The purified recombinant hisMBP-neprosin (rMBP-Npr1) fusion protein was incubated in 100mM Gly-HCl pH2.5 at 37°C for the indicated number of days (lanes 1-4). Digests were conducted with enzyme incubated for a minimum of 7 days. Samples were analyzed by 8% SDS-PAGE followed by silver staining. The mature recombinant neprosin (rNpr1) appeared to be slightly smaller than the endogenous neprosin (Npr1, lane 5) isolated from *Nepenthes* digestive fluid. The difference was confirmed to arise from glycosylation.

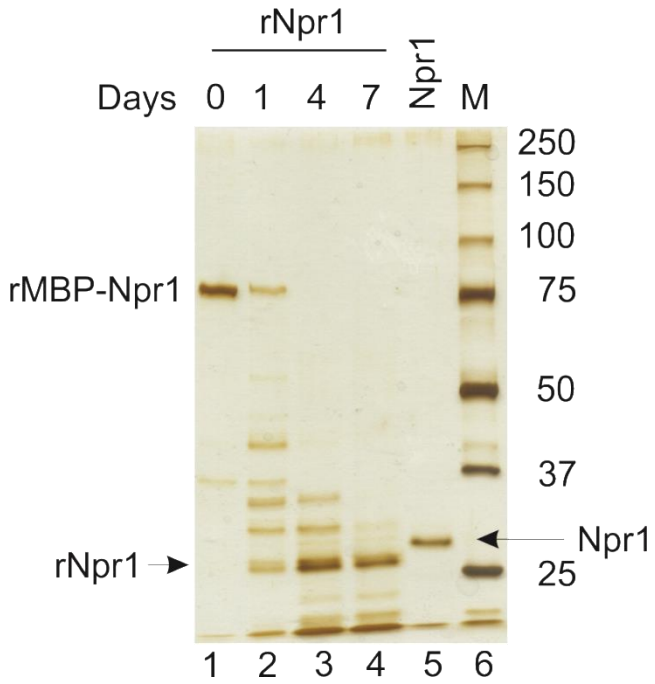


Fig. S3. SDS-PAGE analysis of BSA digested with neprosin in the presence of urea in different concentrations up to 8 M. Reduced BSA (0.05 mg mL^{-1}) was incubated with recombinant neprosin in 100 mM Gly-HCl pH 2.5 and the indicated urea concentration at 37 °C for 1 hour. The reaction was analyzed by 8% SDS-PAGE and followed by silver staining.

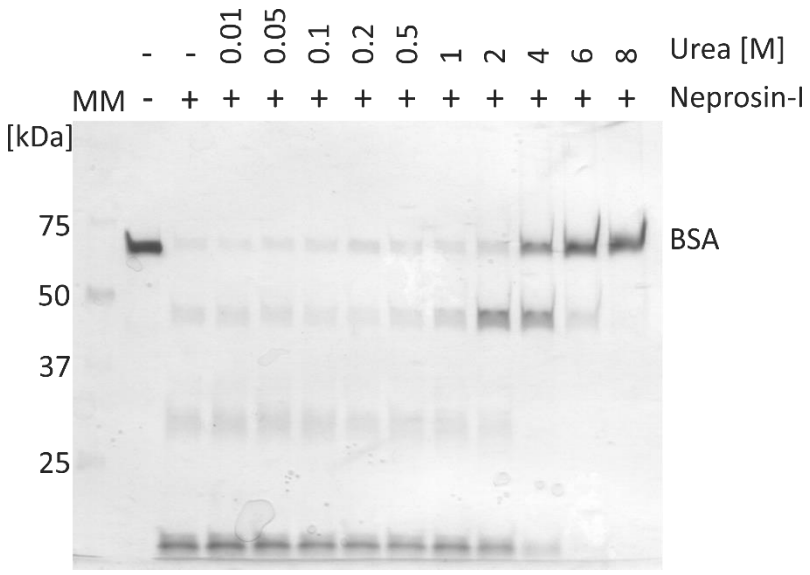


Fig. S4. Cleavage analysis of trypsin expressed as iceLogo and heat map visualizations (from up to down). Cleavage sites were calculated based on HeLa digests, which yielded 17,058 and 16,209 unique cleavage sites with trypsin. Differences in iceLogo and heat map visualizations are displayed as the amino acid occurrence in a certain position normalized to the natural occurrence in *Homo sapiens*.

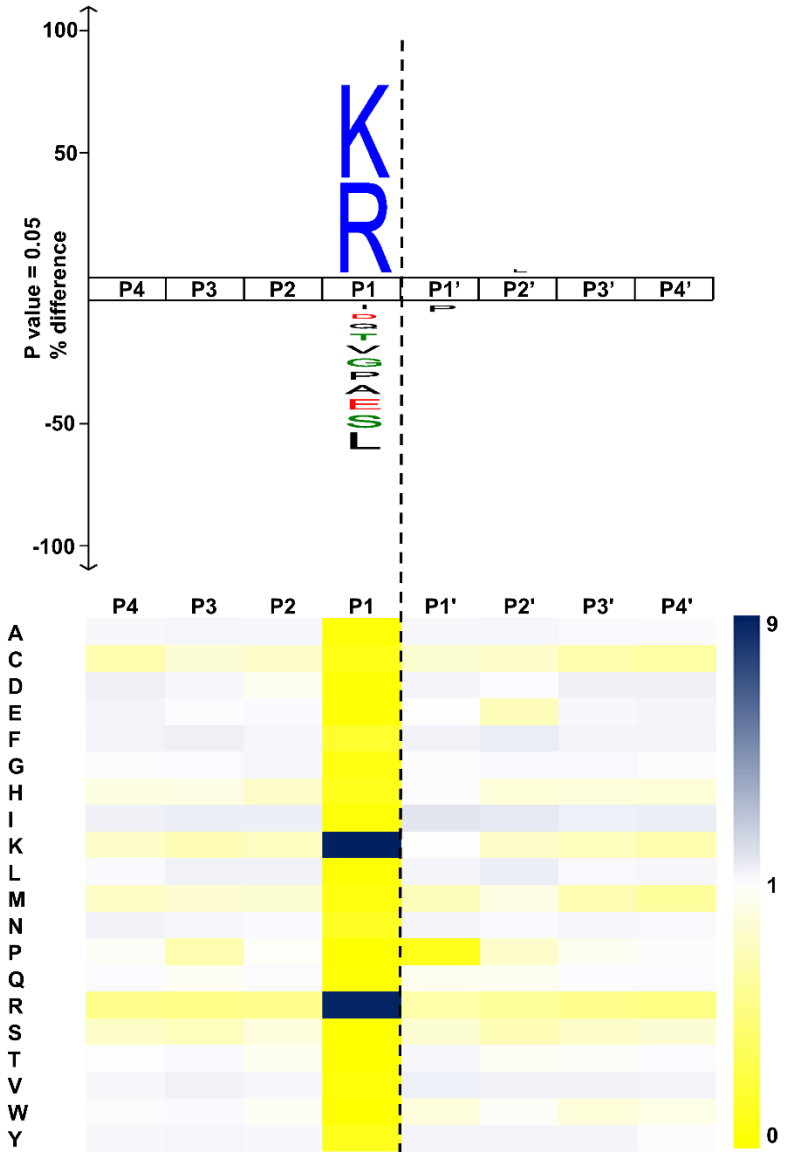


Fig. S5. Cleavage analysis of endogenous neprosin expressed as iceLogo and heat map visualizations (from up to down). Cleavage sites were calculated based on HeLa digests, which yielded 3,524 and 3,668 unique cleavage sites with endogenous neprosin. Differences in iceLogo and heat map visualizations are displayed as the amino acid occurrence in a certain position normalized to the natural occurrence in *Homo sapiens.*

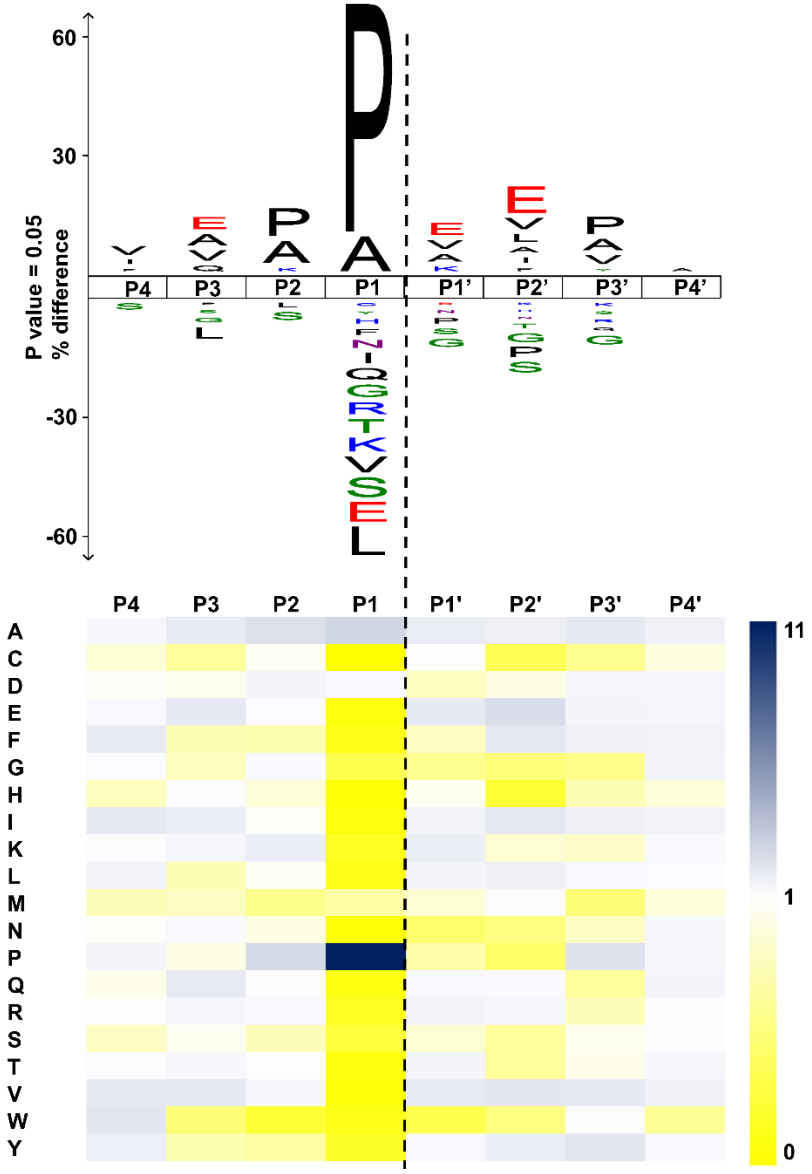


Fig S6. Comparative cleavage analysis of AN-PEP and recombinant neprosin. (A) Cleavage analysis of AN-PEP as an iceLogo visualization. Cleavage sites were calculated based on HeLa digests, which yielded 8,515 and 8,199 unique cleavage sites. Differences are displayed as the amino acid occurrence in a certain position normalized to the natural occurrence in *Homo sapiens*. (B) Cleavage analysis of AN-PEP compared to recombinant neprosin. Experiments were carried out in two biological replicates and unique cleavage sites are displayed relative to all identified cleavages. Observed cleavages C-terminal to the corresponding amino acid are shown as red bars for neprosin and for AN-PEP as blue bars, respectively.

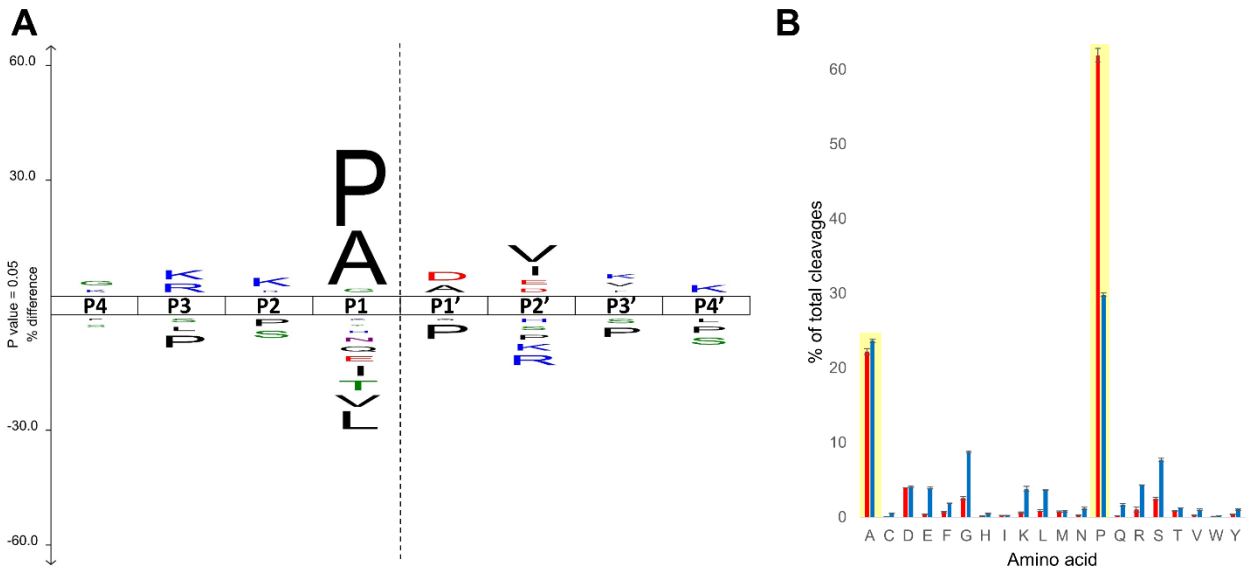


Fig. S7. Cleavage specificity of neprosin towards dipeptide motifs. (A) Cleavage probability of dipeptide motifs in HeLa normalized to the natural occurrence in *Homo sapiens*. The 40 most often cleaved motifs are displayed on the horizontal axis, from highest (left) to lowest (right) normalized activity. Cleavage events, which are C-terminal to Pro and Ala are displayed in red and blue, respectively. Cleavages after Asp are displayed in yellow and other cleavages are displayed in grey. (B) Subsite analysis of Asp fixed to P1 position and (C) Pro fixed to P1' position as indicated by black arrows using iceLogo visualization. Differences are displayed as the amino acid occurrence in a certain position normalized to the natural occurrence in *Homo sapiens.*

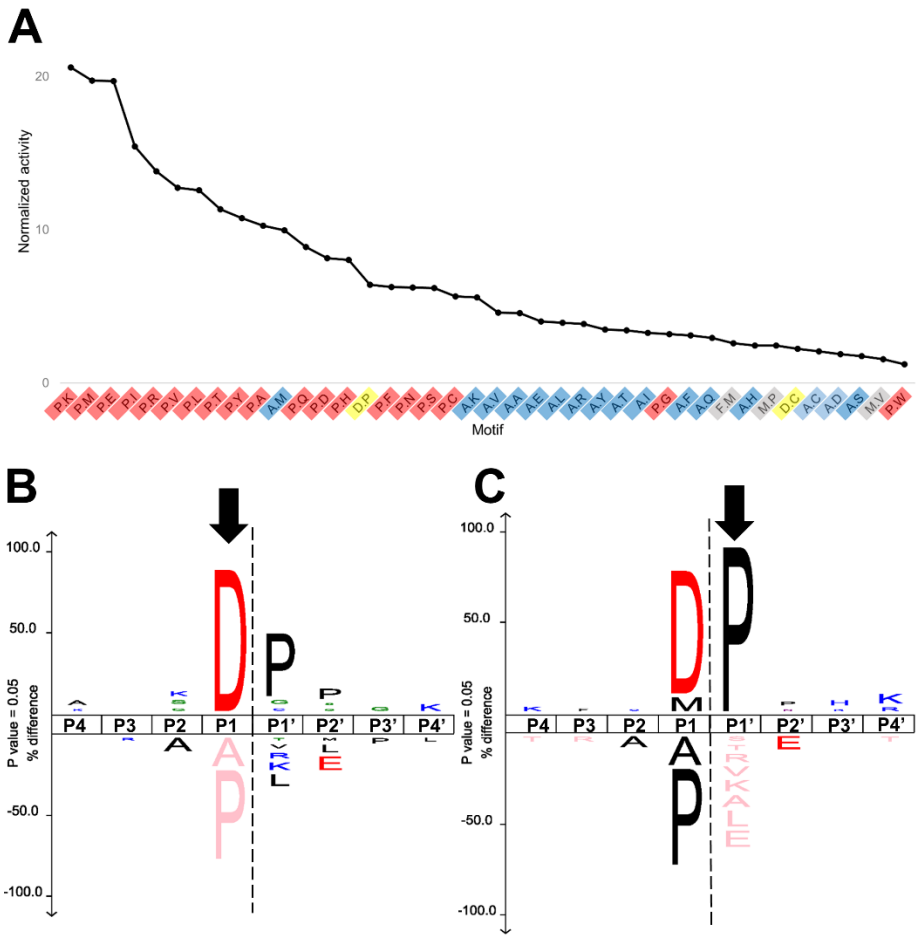


Fig. S8. Heat map visualization of observed differences in cleavage sites of a tryptic HeLa digest with and without incubation overnight at pH 2.5 and 37 °C. Differences are normalized to the natural occurrence of the corresponding dipeptide motifs in *Homo sapiens*.

Fig. S9. Missed cleavage site analysis of neprosin. (A) Observed missed cleavage sites, based on missed Pro and Ala residues, for HeLa cell lysate digested with neprosin. Zero missed cleavages are shown in red, one missed cleavage is shown in orange, two missed cleavages in light grey, three missed cleavage events in dark grey and missed cleavages equal or superior to a number of four are shown in black. (B) IceLogo visualization of amino acids C-terminal to Pro and (C) Ala in case of a missed cleavage event. Differences are displayed as the amino acid occurrence in a certain position normalized to the natural occurrence in *Homo sapiens*.

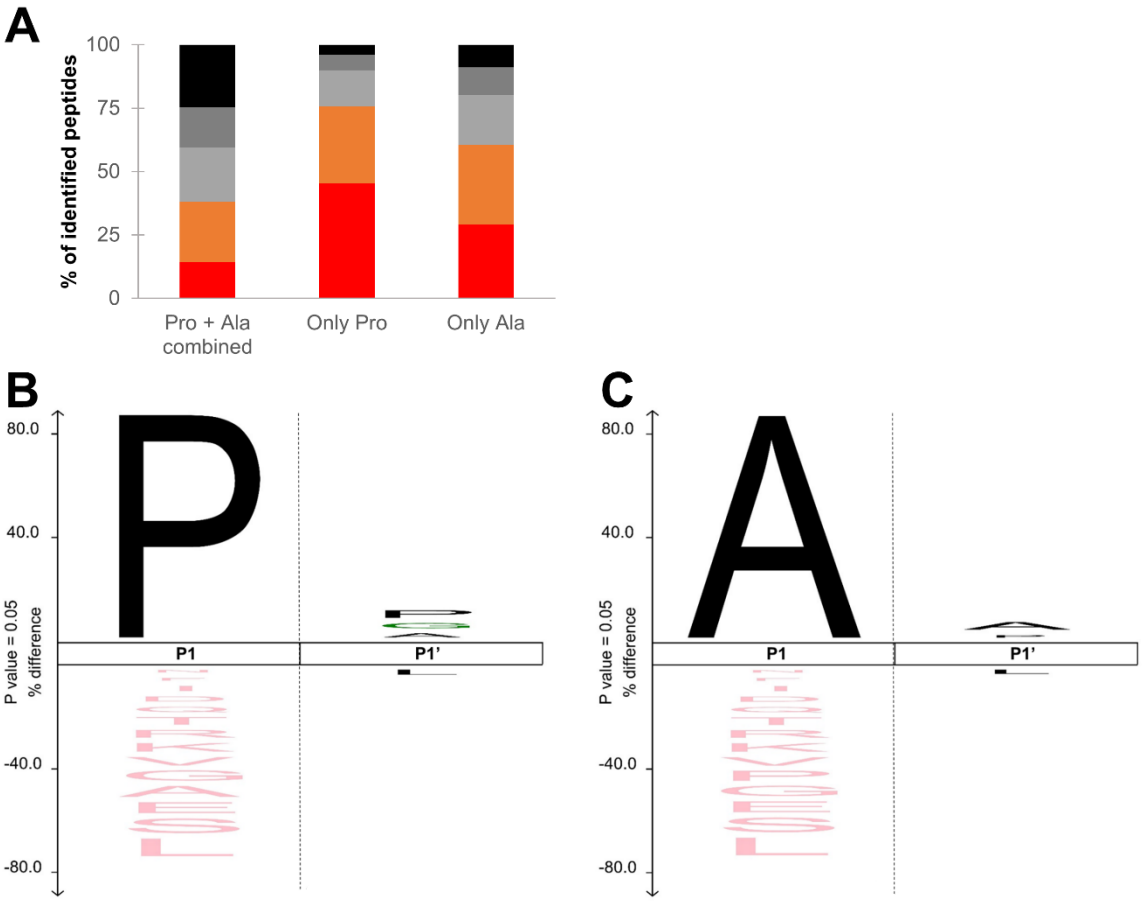


Fig. S10. Venn diagram visualization of the corresponding chosen fragmentation mode for identified peptide spectrum matches in HeLa digests generated by (A) trypsin, (B) LysC and (C) neprosin using the decision tree-based LC-MS method. The fragmentation mode was chosen depending on the nature of the selected ions: Doubly charged ions, triply charged ions with *m/z* > 650, quadruply charged ions with *m/z* > 900 and quintuply charged ions with *m/z* > 950 were triggered with CID. Triply charged ions with *m/z* ≤ 650, quadruply charged ions with *m/z* ≤ 900, quintuply charged ions with *m/z* ≤ 950 and ions with a charge state of six and higher were fragmented using ETD.

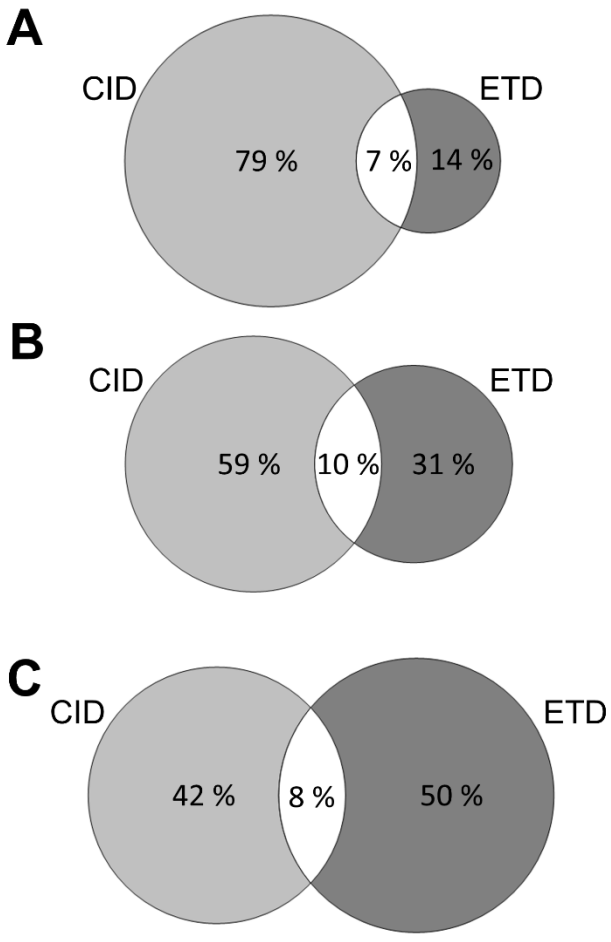


Fig. S11. HCD fragment ion spectra of peptides containing a C-terminal Pro-residue identified in HeLa lysate digested with neprosin. Three fragment spectra are displayed as representing species with different positions of charged residues within the peptide sequence. Basic residues are shown in blue and bold. (A) HCD fragment ion spectrum of the peptide MEVEKNSTP (precursor ion was at *m/z* 517.2725; z=2). (B) HCD fragment ion spectrum of the peptide TMSGVTTCLRFP (precursor ion was at *m/z* 685.3405; z=2). (C) HCD fragment ion spectrum of the peptide KIAILTCP (precursor ion was at *m/z* 458.2725; z=2).

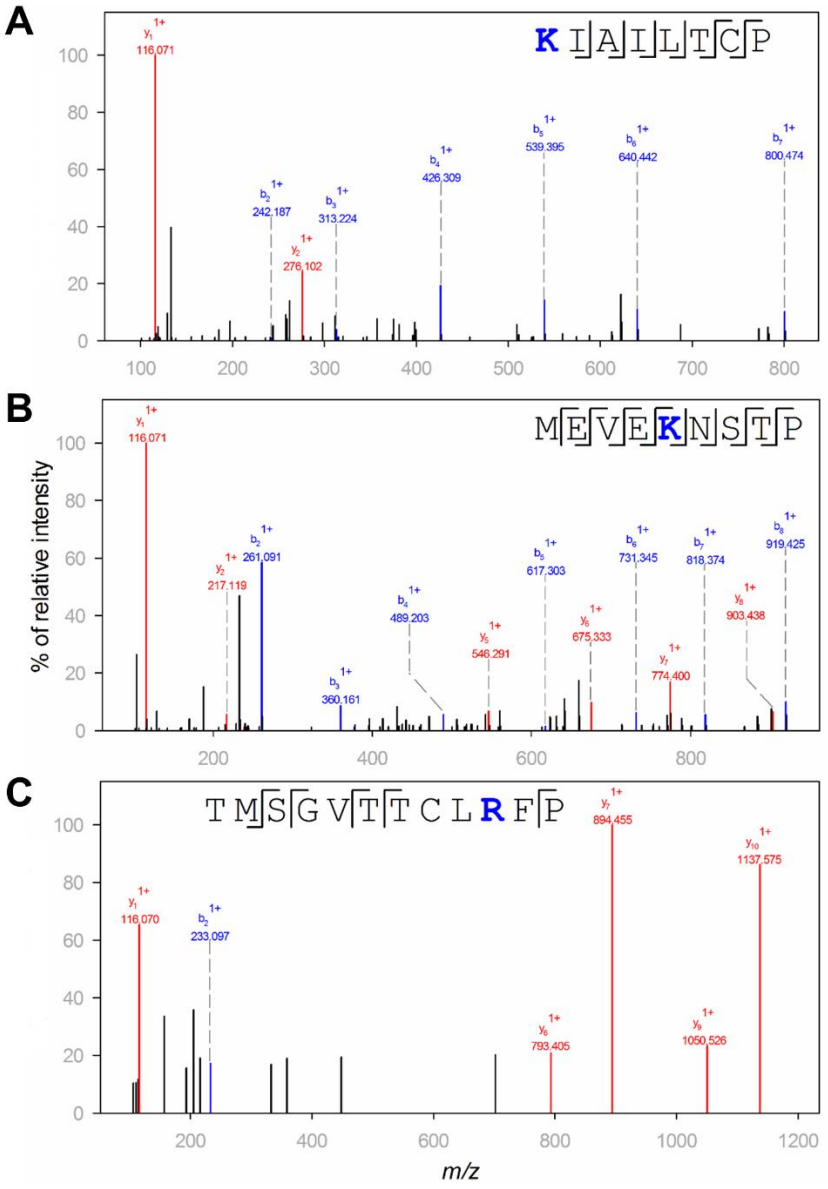


Fig. S12. Whole histone digestion using neprosin. RP-HPLC-MS chromatogram (first 20 min, displayed as the TIC) of histone tails from calf thymus after digestion with neprosin. Eluting peptides are labelled in the chromatogram. The precursor ion mass spectrum at 5.37 min is displayed as an insert showing the abundant co-elution of highly charged and differently modified peptides from histone H3 tail (peptide 1-38) under these conditions.

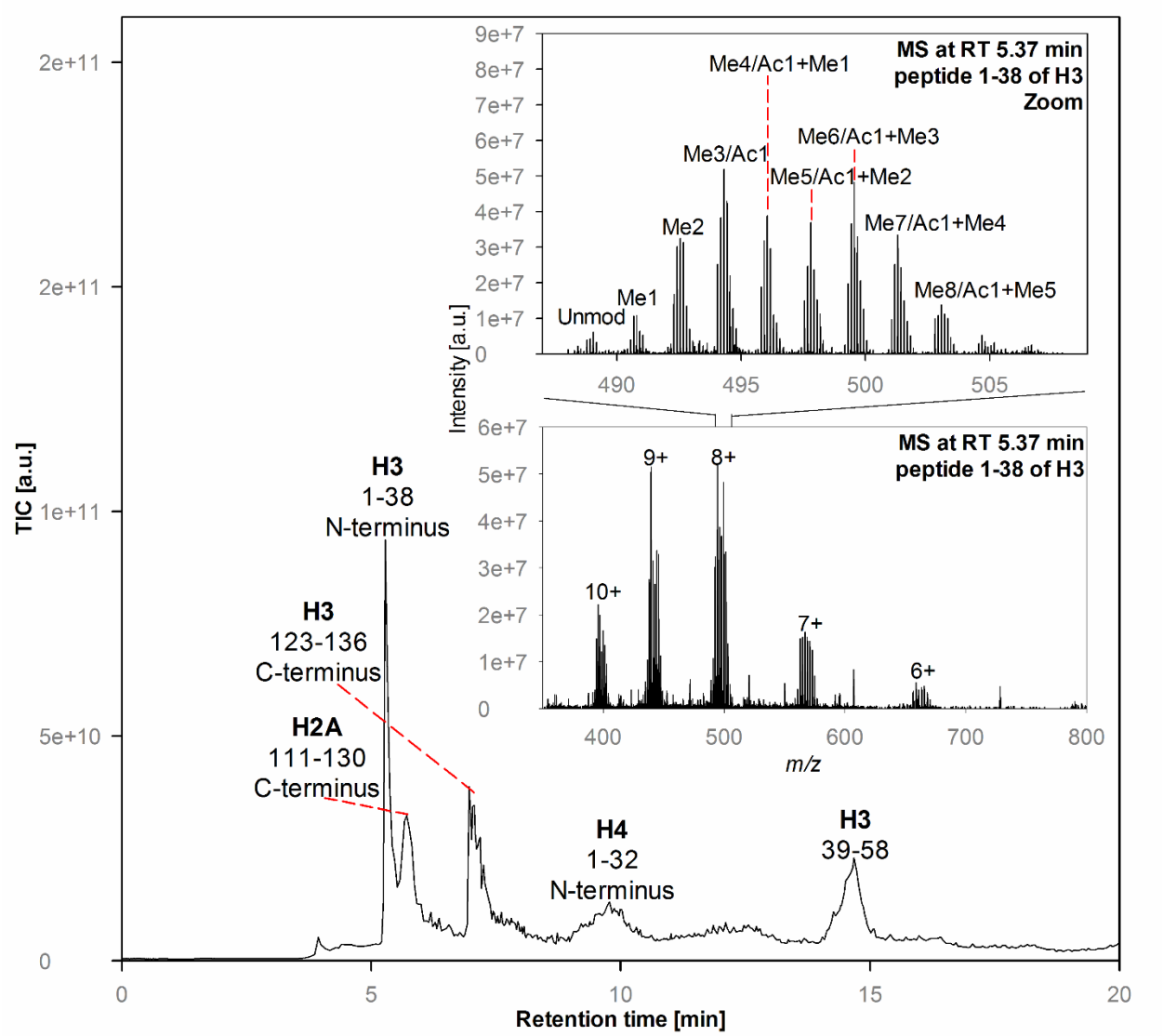


Fig. S13. Histone H2A mapping using neprosin as identified within a digest of whole histones from calf thymus. Underlying peptides are shown as red bars in a sequence coverage plot. Pro residues are highlighted in red and bold, while Arg and Lys residues are shown in black and bold. Legend for identified PTMs is provided in the figure.