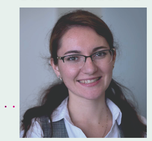
A Proline- and Alanine-specific protease is complementary to trypsin in proteomics applications

Diana Samodova*, Chris Hosfield, Christian N. Cramer, Maria V. Giuli, Giulia Franciosa, Enrico Cappellini, Michael Rosenblatt, Christian D. Kelstrup, Jesper V. Olsen

Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
diana.samodova@cpr.ku.dk



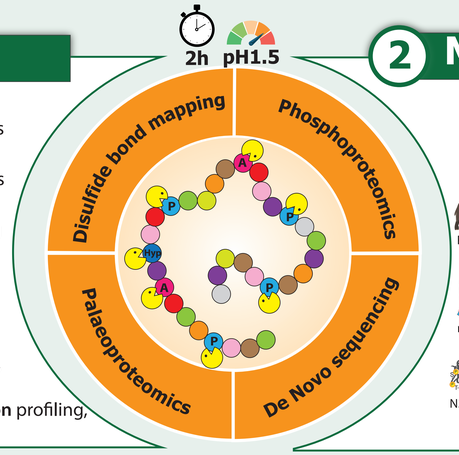
The Novo Nordisk Foundation Center for Protein Research

OLSEN GROUP

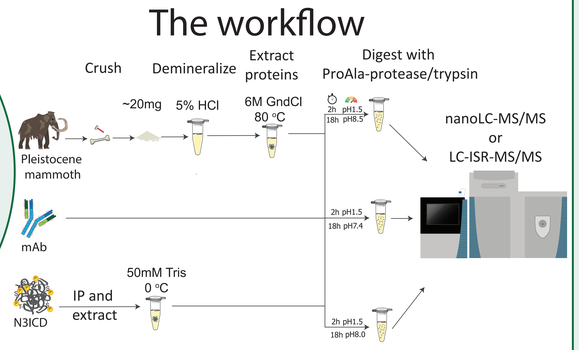
1 Introduction and goals

- Trypsin is the protease of choice in bottom-up proteomics[1];
- Too few or too many tryptic cleavage sites (R and K) in some of the proteins (e.g. collagen)[2];
- Trypsin is mainly active at pH 7-9, while in some cases lower digestion pH is required (e.g. disulfide bond mapping)[3];
- Ambiguos phosphosite localization in tryptic phosphopeptides containing basophylic phosphorylation sites[4];
- Orthogonal peptides are required for *de novo* protein sequencing[5];
- Proteases alternative to trypsin are desired for specific proteomics applications.

The Goal: to test the application of a **proline- and alanine-specific protease**, which is active at **pH1.5** in **2h** of protein digestion, to a series of proteomics investigations comprising digestion of **proline-rich** proteins, **phosphorylation** profiling, **disulfide bond mapping** and ***de novo*** protein sequencing.



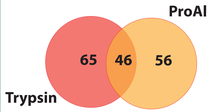
2 Methods



Results

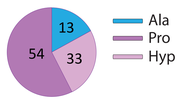
3 Palaeoproteomics

a) Pleistocene mammoth bone proteins



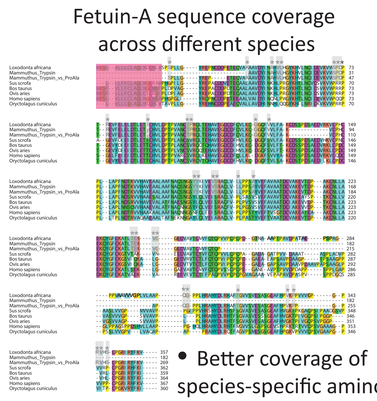
- Complementary non-collagenous proteins in tryptic and ProAla-digests.

b) Cleavage sites of ProAla-protease, %



- Cleavage after alanine, proline and hydroxy-proline.

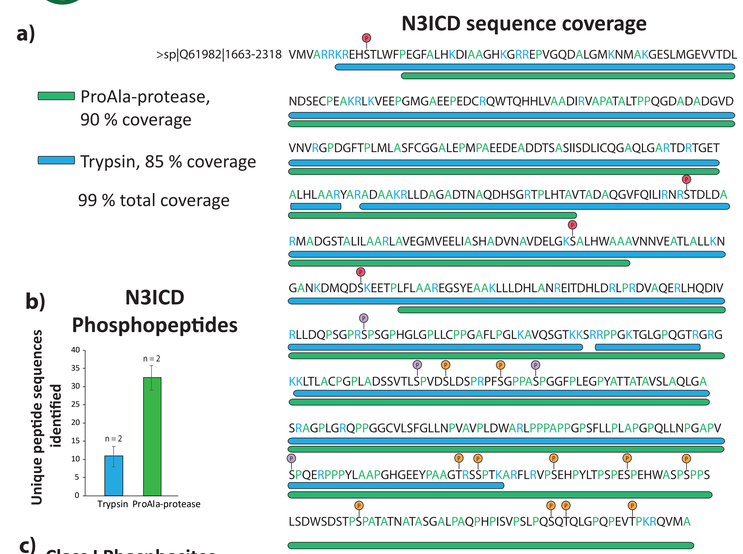
c) Fetuin-A sequence coverage across different species



- Better coverage of species-specific amino acid substitutions.

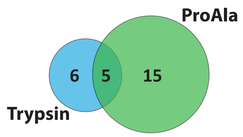
4 Phosphorylation profiling

a)



b) N3ICD Phosphopeptides

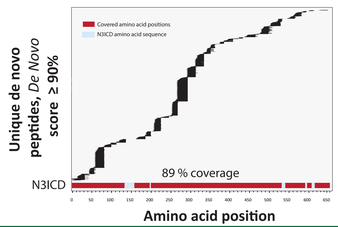
c) Class I Phosphosites in N3ICD protein



- 3x more unique phosphopeptides and 2x more class I localized phosphosites in ProAla-digests, compared to trypsin.

5 De Novo sequencing

De Novo sequencing of N3ICD protein using ProAla-protease and trypsin

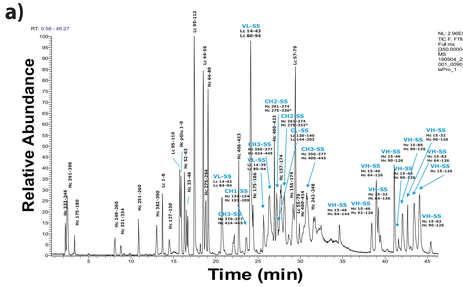


- 89% sequence coverage of N3ICD protein by overlapping high-score tryptic and ProAla-*De Novo* peptides.

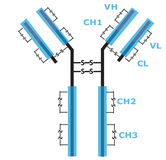
6 Disulfide bond mapping

ProAla-digested mAb chromatogram

a)

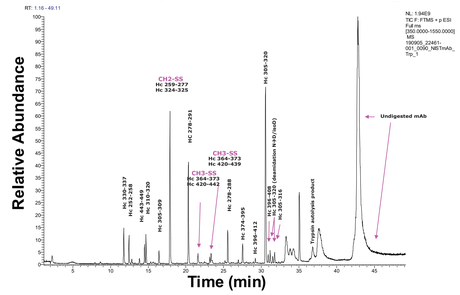


Antibody coverage

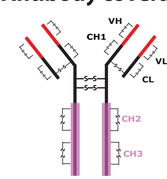


- Constant and variable regions (except for the hinge region) of the antibody are covered by ProAla-peptides.

b) **Trypsin-digestion mAb chromatogram**



Antibody coverage



- Constant region of the antibody is covered by tryptic peptides.

7 Conclusions and Future perspective

ProAla-digestion showed an improved phosphorylation profiling in purified proline-rich single protein N3ICD, compared to trypsin, as well as allowed to increase total sequence coverage of the protein by combining peptides generated by both proteases (4). A similar increase in total sequence coverage was observed for non-collagenous proteins in Pleistocene mammoth bone sample, allowing to cover more species-specific amino acid substitutions relevant for phylogenetic placement. Notably, cleavage also occurs at the C-terminus of hydroxyproline, facilitating efficient digestion of bone collagen and improving the identification of non-collagenous bone proteins (3). Using ProAla-protease at pH 1.5 in 2h of protein digestion potentially allows to decrease scrumbling of the disulfide bonds. We observed a higher digestion efficiency of a non-reduced and non-denatured NIST mAb and almost complete coverage of its sequence and disulfide-containing fragments, compared to trypsin. This demonstrates ProAla-protease as a powerful tool for efficient disulfide bond mapping (6). Finally, we performed a near-complete *de novo* sequencing of N3ICD protein, using a combination of ProAla- and tryptic peptides (5). Taken together, this demonstrates the broad utility of ProAla-protease for numerous proteomics applications. A potential future application of the protease, not covered in this study, is the reduction of artificially-introduced deamidation at low pH of proteolytic digestion[6].

8 References and Acknowledgements

1. Tabb, D. L., Huang, Y., Wysocki, V. H. & Yates, J. R., 3rd. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal. Chem.* **2004**, *76*, 1243–1248.
2. Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a004978.
3. Lakbub, J. C., Shipman, J. T. & Desaire, H. Recent mass spectrometry-based techniques and considerations for disulfide bond characterization in proteins. *Anal. Bioanal. Chem.* **2018**, *410*, 2467–2484.
4. Huesgen, P. F. et al. LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification. *Nat. Methods* **2015**, *12*, 55–58.
5. Guthals, A., Clauser, K. R., Frank, A. M. & Bandeira, N. Sequencing-grade de novo analysis of MS/MS triplets (CID/HCD/ETD) from overlapping peptides. *J. Proteome Res.* **2013**, *12*, 2846–2857.
6. Cao, M. et al. An Automated and Qualified Platform Method for Site-Specific Succinimide and Deamidation Quantitation Using Low-pH Peptide Mapping. *J. Pharm. Sci.* **2019**, *108*, 3540–3549.

Work at The Novo Nordisk Foundation Center for Protein Research (CPR) is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001). D.S., J.V.O and E.C are supported by the Marie Skłodowska-Curie European Training Network (ETN) TEMPERA, a project funded by the European Union's EU Framework Program for Research and Innovation Horizon 2020 under Grant Agreement No. 722606.

TEMPERA

Promega