Development of the Human Proteome Peptide Catalog – A comprehensive Repository of Reference Peptides for the Human Proteome

Karsten Schnatbaum[1], Daniel P. Zolg[2], Mathias Wilhelm[2], Tobias Knaute[1], Johannes Zerweck[1], Holger Wenschuh[1], Bernhard Kuster[2,3,4], Ulf Reimer[1]*

[1] JPT Peptide Technologies GmbH, Berlin, Germany, [2] Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany, [3] Center for Integrated Protein Science Munich, Freising, Germany, [4] Bavarian Center for Biomolecular Mass Spectrometry, Freising, Germany.

Introduction

Synthetic reference peptides are an integral part of current MS-based proteomics. Our project ProteomeTools[1] aims to synthesize and provide data for reference peptides covering all accessible proteins of the human proteome. Using the data generated in the project we describe here a comprehensive peptide catalog that allows to order customized sets of peptides. The catalog contains >400,000 reference peptides that can be searched with multiple filter options and provides links to reference spectra.

Methods and Results

Peptides were selected based on known proteotypicity, or - when this was not known - by synthesizing all suitable peptides of a protein. The peptides were prepared by SPOT synthesis, combined in pools of 1000 peptides per pool, and analyzed by LC-MS/MS using five different fragmentation methods.

Peptides that were reliably detected with an Andromeda score >50 were selected for inclusion into the catalog. This resulted in a list of 428,681 peptides for which a total of >50 Mio reference spectra is available (Figure 1a). The peptides represent 19,764 human proteins (98.6% of the human proteome, Figure 1b) and 41,636 isoforms of these proteins. For >90% of the proteins at least five peptides per protein are available (Figure 1c).

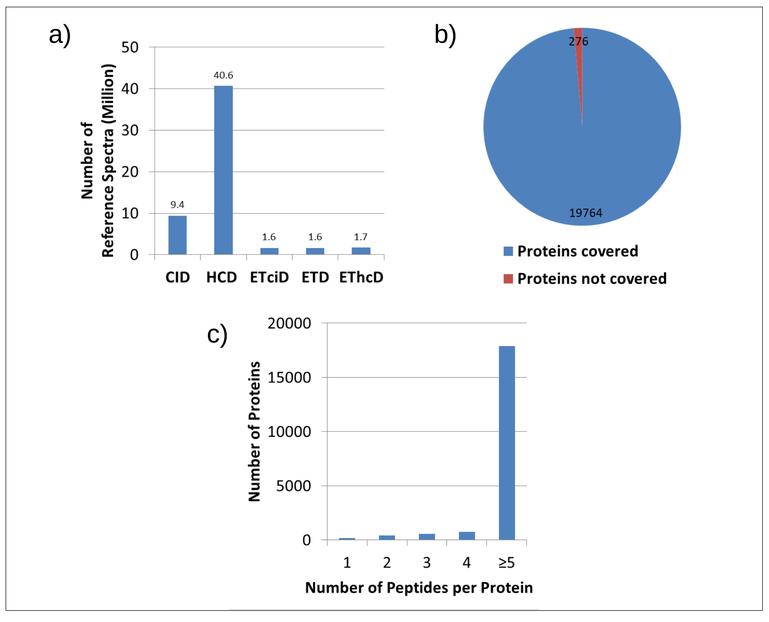

Figure 1: a) Number of reference spectra by ionization method, b) Human proteins covered (blue) or not covered (red) by the catalog, c) Number of peptides per protein.

To enable fast access to the data and reagents, a publicly available database was set up[2] that applies fast and easy-to-use filter functions to select suitable peptides (Figure 2). Efficient peptide selection is supported by links to Uniprot and reference spectra in ProteomicsDB[3]. Finally, an interface for ordering user-defined subsets of peptides is provided as well.

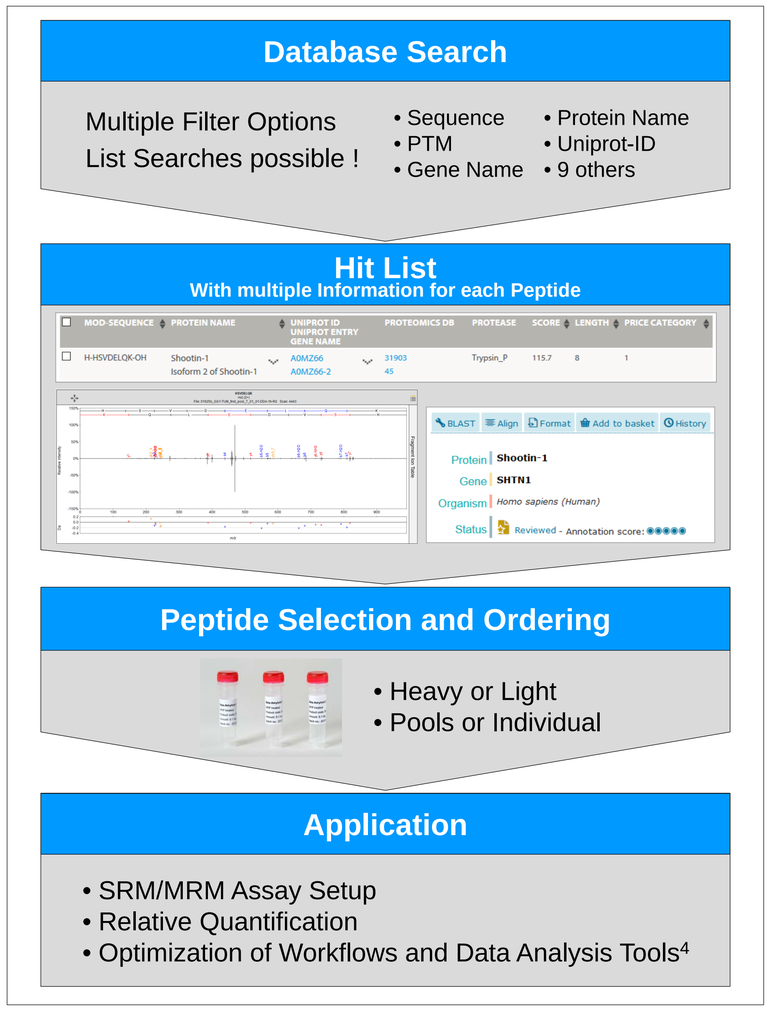

Figure 2: Workflow for usage of the catalog.

Several exemplary searches were performed to show the usability and speed of the catalog. A large number of peptides was obtained for selected protein (groups) and searching was very fast (Table 1).

Table 1: Examples of search performance and protein coverage.

Search Term	Search Time	# Reference Peptides
Cytokine	3 sec	951
Interleukin	3 sec	1911
Cytochrome	2 sec	1369

Search Term	Search Time	# Reference Peptides
Antigen	5 sec	2880
Receptor	20 sec	26161
Kinase	22 sec	17360

Figure 3 shows an exemplary use of the catalog data. Plotted are two experimental spectra from an external dataset[5] (top), identified as AQGLVTFR (left) and LSGVEDHVK (right) by Andromeda/Maxquant, in comparison to the reference spectra available in the catalog (bottom). The reference peptide measurements strongly suggest that AQGLVTFR is a false positive identification, while the low scoring LSGVEDHVK is confirmed.

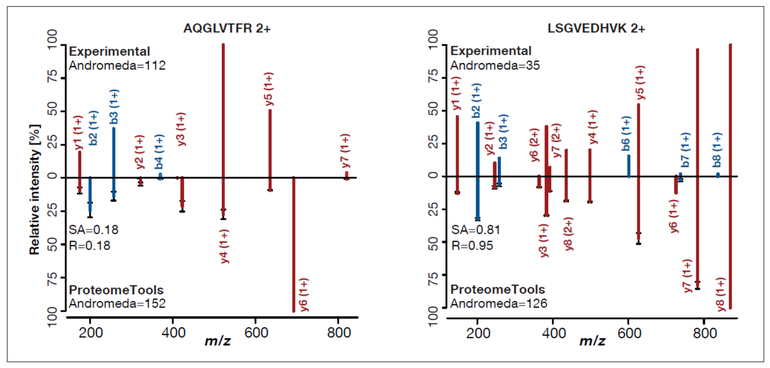

Figure 3: Example for the use of reference peptides for correct sequence assignment.

Conclusion

A repository of validated reference peptides for proteomics is presented which includes >400,000 peptides, covering essentially all human proteins. For each peptide CID, HCD, ETciD, ETD and EThcD spectra are available. Ongoing efforts focus on the expansion of the peptide list, including peptides with important PTMs.

The Human Proteome Peptide Catalog
- **Is the most comprehensive repository of standards for MS-based proteomics**
- **Currently holds > 400,000 peptides**
- **Has peptides with favorable LC-MS characteristics**
- **Provides links to high confidence spectra for each peptide.**

References

(1) (a) D. P. Zolg et al., *Nat. Meth.* **2017**, *14(3)*, 259-262. (b) http://www.proteometools.org
(2) www.jpt.com
(3) T. Schmidt et al., *Nucl. Acids Res.* **2018**, *46*, D1271-D1281.
(4) S. Gessulat et al., *Nat. Meth.* **2019**, in press.
(5) D. B. Bekker-Jensen et al., *Cell Syst.* **2017**, *4(6)*, 587-599.

* Correspondence should be addressed to Karsten Schnatbaum (schnatbaum@jpt.com)

jpt
Innovative Peptide Solutions