### A Practical Beginner's Guide to Proteomics

This manuscript (<u>permalink</u>) was automatically generated from <u>jessegmeyerlab/proteomics-tutorial@2eccdd4</u> on February 1. 2022.

#### **Authors**

#### • Jesse G. Meyer

© 0000-0003-2753-3926 · ♥ jessegmeyerlab · ♥ j my sci

Department of Biochemistry, Medical College of Wisconsin · Funded by Grant R21 AG074234; Grant R35 GM142502

#### Dina Schuster

Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich 8093, Switzerland; Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich 8093, Switzerland; Laboratory of Biomolecular Research, Division of Biology and Chemistry, Paul Scherrer Institute, Villigen 5232, Switzerland

#### Muralidharan Vanuopadath

© 0000-0002-9364-917X · • vanuopadathmurali · V MuraleeDhar School of Biotechnology, Amrita Vishwa Vidyapeetham, Kollam-690 525, Kerala, India

#### Devasahayam Arokia Balaya Rex

-Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, Yenepoya (Deemed to be University), Mangalore 575018, India

#### **Abstract**

Proteomics is the large scale study of protein structure and function from biological systems. "Shotgun proteomics" or "bottom-up proteomics" is the prevailing strategy, in which proteins are hydrolyzed into peptide that are analyzed by mass spectrometry. Proteomics studies can be applied to diverse studies ranging from simple protein identification to studies of protein-protein interactions, absolute and relative protein quantification, post-translational modifications, and protein stability. To enable this range of different experiments, there are diverse strategies for proteome analysis. The nuances of how proteomic workflows differ may be difficult to understand for new practitioners. Here, we provide a comprehensive tutorial of different proteomics methods. Our tutorial covers all necessary steps starting from protein extraction and ending with biological interpretation. We expect that this work will serve as a basic resource for new practitioners of the field of shotgun or bottom-up proteomics.

### Introduction

Proteomics is the large scale study of protein structure and function. Proteins are translated from mRNAs that are transcribed from the genome. Although the genome encodes potential cellular functions and states, the study of proteins is necessary to truly understand biology. Currently, proteomic studies are facilitated by mass spectrometry, although alternative methods are being developed.

Modern proteomics started around the year 1990 with the introduction of soft ionization methods that enabled, for the first time, transfer of large biomolecules into the gas phase without destroying them [1,2]. Shortly afterward, the first computer algorithm for matching peptides to a database was introduced [3]. Another major milestone that allowed identification of over 1000 proteins were actually improvements to chromatography [4]. As the volume of data exploded, methods for statistical analysis transitioned use from the wild west to modern informatics based on statistical models [5] and the false discovery rate [6].

Two strategies of mass spectrometry-based proteomics differ fundamentally by whether proteins are cleaved into peptides before analysis: "top-down" and "bottom-up". Bottom-up proteomics (also refered to as shotgun proteomics) is defined by the hydrolysis of proteins into peptide pieces [7]. Therefore, bottom-up proteomics does not actually measure proteins, but must infer their presence [8]. Sometimes proteins are infered from only one peptide sequence representing a small fraction of the total protein sequence predicted from the genome. In contrast, top-down proteomics attempts to measure all proteins intact [9]. The potential benefit of top-down proteomics is the ability to measure proteoforms [10]. However, due to analytical challenges, the depth of protein coverage that is achievable by top-down proteomics is less than the depth that is achievable by bottom-up proteomics.

In this tutorial we focus on the bottom-up proteomics workflow. The most common version of this workflow is generally comprised of the following steps. First, proteins in a biological sample must be extracted. Usually this is done by denaturing and solubilizing the proteins while disrupting DNA and tissue. Next, proteins are hydrolyzed into peptides, usually using a protease like trypsin. Peptides from proteome hydrolysis must be purified. Most often this is done with reversed phase chromatography cartridges or tips. The peptides are then almost always separated by liquid chromatography before they are ionized and introduced into a mass spectrometer. The mass spectrometer then collects precursor and fragment ion data from those peptides. The data analysis is usually the rate limiting step. Peptides must be identified, and proteins are inferred and quantities are assigned. Changes in proteins across conditions are determined with statistical tests, and results must be interpreted in the context of the relevant biology.

There are many variations on this workflow. The wide variety of experimental goals that are achievable with proteomics technology leads to a wide variety of potential proteomics workflows. Even choice is important and every choice will affect the results. In this tutorial, we cover all of the required steps in detail to serve as a tutorial for new proteomics practioners.

- 1. Types of experiments enabled by proteomics
- 2. Protein extraction
- 3. proteolysis
- 4. Isotopic Labeling
- 5. Enrichments
- 6. Peptide purification
- 7. Mass Spectrometry
- 8. Peptide Ionization

- 9. Data Acquisition10. Basic Data Analysis11. Biological Interpretation12. Experimental considerations and design

### **Types of Experiments**

A wide range of questions are addressable with proteomics technology, which translates to a wide range of variations of proteomics workflows. Sometimes identifying what proteins are present is desired, and sometimes the quantities of as many proteins as possible are desired. Proteomics experiments can be both qualitative and quantitative.

### **Qualitative experiments**

- Identifying proteins
- Identifying post translational modifications
- Identifying protein isoforms

### **Quantitative experiments**

- Protein abundance changes
- Phosphoproteomics
- Glycoproteomics
- Structural techniques (XL-MS, HDX-MS, FPOP, protein-painting, LiP-MS, radical footprinting, ion mobility)
- Protein stability and small molecule binding (Thermal proteome profiling, TPP, or cellular thermal shift assay, CETSA)
- Protein-protein interactions (PPIs): AP-MS, APEX, BioID

#### **Protein Extraction**

First, proteins must be isolated from the sample matrix. Because some proteins alter other proteins, the goal is to simultaneously solubilize and denature proteins. This is achieved with a combination of salt and chaotropic agent.

- 1. Choice of Lysis buffer
- Urea can cause chemical modifications
- 2. Sample type and homogenisation methods
- specialised sample preparation protocols for non-denaturing protein isolation (i.e. for LiP-MS, HDMX-MS etc)
- 4. chemicals to avoid: PEGs, detergents etc
- 5. removal of contaminations, Protein Precipitation
- detergent reomoval resins, S-TRAP (Protifi) etc
- 7. protein alkylation
- choices of reduction and alkylation reagents, TCEP/DTT/2BME, Chloroacetamide/iodoacetamide, nethyl maleimide

### **Proteolysis**

Proteolysis is the defining step that differentiates bottom-up or shotgun proteomics from top-down proteomics. Hydrolysis of proteins is extremely important because it defines the population of potentially identifyable peptides. Generally peptides between a length of 7-35 amino acid are considered useful for mass spectrometry analysis. Peptides that are too long are difficult to identify by tandem mass spectrometry, or may be lost during sample preparation due to Peptides that are too short are less likely to uniquely match to a single protein. There are many choices of enzymes and chemicals that hydrolyze proteins into peptides. This section summarizes potential choices and their strengths and weaknesses.

Trypsin is the most common choice of protease for proteome hydrolysis [11]. Trypsin is favorable because of its specificity, availability, efficiency and low cost. Trypsin cleaves at the C-terminus of basic amino acids, arginine and lysine. Many of the peptides generated from trypsin are short in length (less than  $\sim$  20 amino acids), which is ideal for chromatographic separation, MS-based peptide fragmentation and identification by database search. The main drawback of trypsin is that majority (56%) of the tryptic peptides are  $\leq$  6 amino acids, and hence using trypsin alone limits the observable proteome [12,13,14]. This limits the number of identifiable protein isoforms and post translational modifications.

- 3. theoretical studies of proteolysis enzymes [15]
- 4. Challenges associated with alternative enzyme choices (non-specific and semi-specific enzymes)

Many alternative proteases are available with different specificities that complement trypsin to reveal different proteomic sequences [12,16], which can help distinguish protein isoforms [17]. The enzyme choice mostly depends on the application. In general, for a mere protein identification mostly trypsin is the choice due to the reasons aforementioned. However, alternative enzymes can facilitate *de novo* assembly when the genomic data information is limited in the public database repositories [18,19,20,21,22]. Use of multiple proteases for proteome digestion also can improve the sensitivity and accuracy of protein quantification [23].

Lysyl endopeptidase (Lys-C) obtained from *Lysobacter enzymogenesis* is a serine protease invloved in cleaving carboxyl terminus of lysine [13]. Like trypsin, the optimum pH required for its activity is from pH 7-9. One among the major advantage of Lys-C is that it is resistant to denaturing agents, including 8 M urea [17]. Compared to Arginine, trypsin is less efficient in cleaving Lysine, which could limit the quality of quantitation from tryptic peptides. Hence to achieve complete protein digestion with minimal missed cleavages, Lys-C is often used in parallel to complement tryspin digestion [25].

Alpha-lytic protease (aLP) is also secreted by the soil bacterial *Lysobacter enzymogenesis* [26]. Wildtype aLP (WaLP) and an active site mutant of aLP, M190A (MaLP), have been used to expand proteome coverage [27]. Based on observed peptide sequences from yeast proteome digestion, WaLP showed a specificity for small aliphatic amino acids like alanine, valine, and glycine, but also threonine and serine. MaLP showed a specificity for slightly larger amino acids like methionine, phenylalanine, and surprisingly, a preference for leucine over isoleucine. The specificity of WaLP for threonine enabled the first method for mapping endogenous human SUMO sites [28].

Glutamyl peptidase I, commonly known as Glu-C or V8 protease, is a serine protease obtained from *Staphyloccous aureus* [29]. Glu-C cleaves at the C-terimus of glutamate, but also after aspartate [29,30].

Peptidyl-Asp metallopeptidase, commonly known as Asp-N, is a metalloprotease obtained from *Pseudomonas fragi* [31]. Asp-N catalyzes the hydrolysis of peptide bonds at the N-terminal of aspartate residues. The optimum activity of this enzyme is between pH 4 and 9. As with any metalloprotease, chelators like EDTA should be avoided for digestion buffers when using Asp-N. Studies also suggests that Asp-N cleaves at the amino terminus of glutamate when a detergent is present in the proteolysis buffer [31]. Asp-N often leaves many missed cleavages [17].

Chymotrypsin or Chymotrypsinogen A is a serine protease obtained from porcine or bovine pancreas with an optimum pH range from 7.8 to 8.0 [32]. It cleaves at the C-terminus of hydrphobic amino acids Phe, Trp, Tyr and barely Met and Leu residues. Since the transmembrane region of membrance proteins lacks tryptic cleavage sites, this enzyme works well with membrane proteins having more hydrophobic residues [17,33,34]. Also, the chymotryptic peptides generated after proteolysis will cover the proteome space orthogonal to that of tryptic peptides both in a quantitative and qualitative manner [34,35,36]

Clostripain, commonly known as Arg-C, is a cysteine protease obtained from *Clostridium histolyticum* [37]. It hydrolyse mostly the C-terminal Arg residues and sometimes Lys residues, but with less efficiency. The peptides generated are generally longer than that of tryptic peptides. Arg-C is often used with other proteases for improving qualitative proteome data and also for investigating PTMs [13].

LysargiNase, also known as Ulilysin, is a recently discovered protease belonging to the metalloprotease family. It is a thermophilic protease derived from *Methanosarcina acetivorans* that specifically cleaves at the N-terminus of Lys and Arg residues [38]. Hence, it enabled discovery of C-terminal peptides that were not observed using trypsin. In addition, it also can cleave modified amino acids such as methylated or dimethylated arginine and lysine [38].

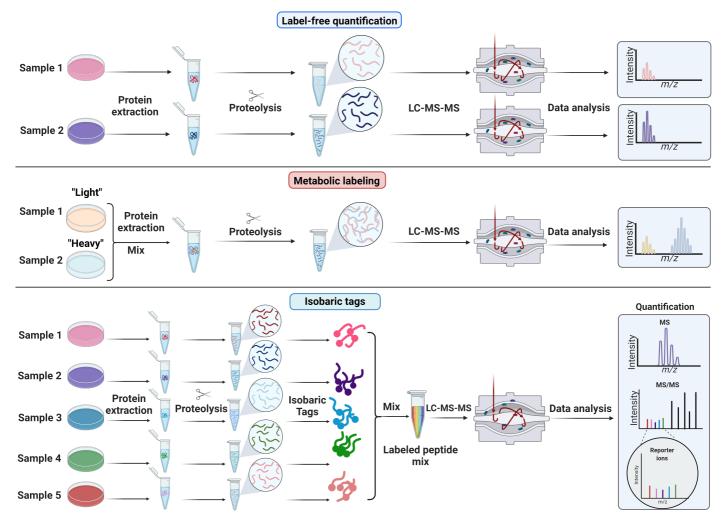
Peptidyl-Lys metalloendopeptidase, or Lys-N, is an metalloprotease obtained from *Grifola frondosa* [39]. It cleaves at N-terminal of Lys and has an optimal activity at pH 9.0. Unlike trypsin, Lys-N is more resistant to denaturing agents and can be heated up to 70 °C [13]. Reports suggests that the peptides generated after Lys-N digestion produces more of c-type ions in a ETD-based mass spectrometer [40]. Hence this can be used for analysing PTMs, identification of C-terminal peptides and also for de novo sequencing strategies [40,41].

Pepsin A, commonly known as pepsin, is an aspartic protease obtained from bovine or porcrine pancreas [42]. This was the first protease that was identified and crystallized and was awarded the Nobel prize [43,44,45]. Pepsin works at an optimum pH from 1-4 and, in a pH depended manner it specifically cleaves Trp, Phe, Tyr and Leu [13]. Since it possess high enzyme activity and broad specificity at lower pH, it is preferred over other proteases for MS-based disulphide mapping [46,47]. Pepsin is also used extensively for structural mass spectrometry studies with hydrogen-deuterium exchange (HDX) because the rate of back exchange of the amide deuteron is minimized at low pH [48,49].

### **Peptide and Protein Labeling**

Discussion of methods to isotopically label peptides or proteins that enable quantification

- 1. SILAC/SILAM
- 2. iTRAQ
- 3. TMT
- 4. dimethyl labeling



**Figure 1: Quantitative strategies commonly used in proteomics.** A) Label-free quantitation. Proteins are extracted from samples, enzymatically hydrolyzed into peptides and analyzed by mass spectrometry. Chromatographic peak areas from peptides are compared across samples that are analyzed sequentially. B) Metabolic labelling. Stable isotope labeling with amino acids in cell culture (SILAC) is based on feeding cells stable isotope labeled amino acids ("light" or "heavy"). Samples grown with heavy or light amino acids are mixed before cell lysis. The relative intensities of the heavy and light peptide are used to compute protein changes between samples. C) Isobaric or chemical labelling. Proteins are isolated separately from samples, enzymatically hydrolyzed into peptides, and then chemically tagged with isobaric stable isotope labels. These isobaric tags produce unique reporter mass-to-charge (m/z) signals that are produced upon fragmentation with MS/MS. Peptide fragment ions are used to identify peptides, and the relative reporter ion signals are used for quantification.

### **Peptide or Protein Enrichment**

### Protein enrichment (e.g. for protein protein interactions)

- APEX
- bioID
- bioplex

### **Peptide enrichment**

- antibody enrichments of modifications, e.g. lysine acetylation [50].
- TiO2 and Fe enrichment of phosphorylation
- Glycosylation
- SISCAPA

### **Methods for Peptide Purification**

- 1. Reverse phase including tips and cartridges
- 2. stage tips
- 3. in stage tip (iST)
- 4. SP2, SP3
- 5. s traps

### **Types of Mass Spectrometers used for Proteomics**

- 1. QQQ
- 2. Q-TOF
- 3. Q-Orbitrap
- 4. LTQ-Orbitrap
- 5. TOF/TOF
- 6. FT-ICR
- 7. types of ion mobility
- SLIM
- FAIMS
- traveling wave
- tims

### **Peptide Ionization**

The 2002 Nobel Prize in Chemistry was awarded to partially to John Fenn and Koichi Tanaka "for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules" [51/].

#### MALDI

### **Electrospray Ionization**

### **Data Acquisition**

Data acquisition strategies for proteomics fall generally within targeted or untargeted, and they can depend on the data (data dependent acquisition or DDA) or be data independent (data-independent acquisition or DIA).

#### **DDA**

**Targeted DDA** 

**Untargeted DIA** 

DIA

**Targeted DIA** 

**Untargeted DIA** 

### **Analysis of Raw Data**

The goal of basic data analysis is to convert raw spectral data into identities and quantities of peptides and proteins that can be used for biologically-focused analysis. This step may often include measures of quality control, cross-run data normalization, quantification on different levels (precursor, peptide, protein), protein inference, PTM (post translational modification) localization and also first steps of data analysis, such as statistical hypothesis tests.

In typical bottom-up proteomics experiments, proteins are digested into peptides and further analyzed with LC-MS/MS systems. Peptides can have different PTMs and ionize differently depending on their length and amino acid distributions. Therefore, mass spectrometers often record different charge and modification states of one single peptide. The entity that is recorded on a mass spectrometer is usually referred to as a precursor ion (peptide with its modification and charge state). This precursor ion is fragmented and the precursor or peptide sequences are obtained though spectral matching. The quantity of a precursor is estimated with various methods. The measured precursor quantities are combined to generate a peptide quantity. Peptides are also often combined into a protein group through protein inference, which combines multiple peptide identifications into a single protein identification [52] [53]. Protein inference is still a challenge in bottom-up proteomics.

Due to the inherent differences in the data structures of DDA and DIA measurements, there exist different types of software that can facilitate the steps mentioned above. The existing software for DDA and DIA analysis can be further divided into freeware and non-freeware:

#### **DDA** freeware

Name	Publication	Website
MaxQuant	Cox and Mann, 2008[ <u>54</u> ]	<u>MaxQuant</u>
MSFragger	Kong et al., 2017[ <u>55</u> ]	<u>MSFragger</u>
Mascot	Perkins et al., 1999[ <u>56</u> ]	<u>Mascot</u>
MS-GF+	Kim et al., [ <u>57</u> ]	MS-GF+

#### **DIA freeware:**

Name	Publication	Website
MaxDIA	Cox and Mann, 2008[ <u>54</u> ]	<u>MaxQuant</u>
Skyline	MacLean et al., 2010[ <u>58</u> ]	<u>Skyline</u>
DIA-NN	Demichev et al., 2019[ <u>59</u> ]	<u>DIA-NN</u>

### **Targeted proteomics freeware:**

Name	Publication	Website
Skyline	MacLean et al., 2010[ <u>58</u> ]	<u>Skyline</u>

#### **DDA** non-freeware:

Name	Publication	Website
ProteomeDis coverer		<u>ProteomeDiscoverer</u>
Mascot	Perkins et al., 1999[ <u>56</u> ]	<u>Mascot</u>
Spectromine		<u>Spectromine</u>
PEAKS	Tran et al., 2018[ <u>60</u> ]	<u>PEAKS</u>

### **DIA non-freeware:**

Name	Publication	Website
Spectronaut	Bruderer et al., 2015[ <u>61</u> ]	<u>ProteomeDiscoverer</u>
PEAKS	Tran et al., 2018[ <u>60</u> ]	<u>PEAKS</u>

### **Analysis of DDA data**

Strategies for analysis of DIA data

**Targeted proteomics data analysis** 

**Quality control** 

**Statistical hypothesis testing** 

### **Biological Interpretation**

- 1. term enrichment analysis (KEGG, GO)
- 2. network analysis methods
- 3. structure analysis
- 4. isoform analysis
- 5. follow-up experiments

### **Experiment Design**

This section should discuss trade offs and balancing them to design an experiment. 1. constraints: Each experiment will have different constraints, which may include the number of samples needed for analysis, or desire to quantify a specific subset of proteins within a sample. 2. sample size 3. statistics 4. costs

### References

#### 1. Electrospray Ionization for Mass Spectrometry of Large Biomolecules

John B Fenn, Matthias Mann, Chin Kai Meng, Shek Fu Wong, Craig M Whitehouse *Science* (1989-10-06) https://doi.org/cq2q43

DOI: 10.1126/science.2675315 · PMID: 2675315

# 2. Protein and polymer analyses up tom/z 100 000 by laser ionization time-of-flight mass spectrometry

Koichi Tanaka, Hiroaki Waki, Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida, Tamio Yoshida, T Matsuo

Rapid Communications in Mass Spectrometry (1988-08) https://doi.org/ffbwwr

DOI: <u>10.1002/rcm.1290020802</u>

# 3. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.

JK Eng, AL McCormack, JR Yates

*Journal of the American Society for Mass Spectrometry* (1994-11)

https://www.ncbi.nlm.nih.gov/pubmed/24226387

DOI: <u>10.1016/1044-0305(94)80016-2</u> · PMID: <u>24226387</u>

### 4. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics

Dirk A Wolters, Michael P Washburn, John R Yates

Analytical Chemistry (2001-10-25) https://doi.org/bn4kq6

DOI: 10.1021/ac010617e · PMID: 11774908

#### 5. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry

Alexey I Nesvizhskii, Andrew Keller, Eugene Kolker, Ruedi Aebersold

Analytical Chemistry (2003-07-15) https://doi.org/b2xv45

DOI: 10.1021/ac0341261 · PMID: 14632076

# 6. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry

Joshua E Elias, Steven P Gygi

Nature Methods (2007-03) https://doi.org/djz7fz

DOI: <a href="https://doi.org/10.1038/nmeth1019">https://doi.org/10.1038/nmeth1019</a>

#### 7. Mass-spectrometric exploration of proteome structure and function

Ruedi Aebersold, Matthias Mann

Nature (2016-09) <a href="https://doi.org/f83zqm">https://doi.org/f83zqm</a>

DOI: 10.1038/nature19949 · PMID: 27629641

#### 8. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry

\* Alexey I. Nesvizhskii, \* Andrew Keller, ‡ and Eugene Kolker, Ruedi Aebersold *ACS Publications* (2003-07-15) <a href="https://pubs.acs.org/doi/abs/10.1021/ac0341261">https://pubs.acs.org/doi/abs/10.1021/ac0341261</a>

#### 9. High-throughput quantitative top-down proteomics

Kellye A Cupp-Sutton, Si Wu

Molecular Omics (2020) https://doi.org/gnx98p

DOI: 10.1039/c9mo00154a · PMID: 31932818 · PMCID: PMC7529119

#### 10. Proteoforms as the next proteomics currency

Lloyd M Smith, Neil L Kelleher

Science (2018-03-09) https://doi.org/gn6p4x

DOI: 10.1126/science.aat1884 · PMID: 29590032 · PMCID: PMC5944612

#### 11. Getting intimate with trypsin, the leading protease in proteomics

Elien Vandermarliere, Michael Mueller, Lennart Martens

Mass Spectrometry Reviews (2013-06-15) https://doi.org/gn64qb

DOI: <u>10.1002/mas.21376</u> · PMID: <u>23775586</u>

#### 12. Value of using multiple proteases for large-scale mass spectrometry-based proteomics.

Danielle L Swaney, Craig D Wenger, Joshua J Coon

Journal of proteome research (2010-03-05) https://www.ncbi.nlm.nih.gov/pubmed/20113005

DOI: 10.1021/pr900863u · PMID: 20113005 · PMCID: PMC2833215

#### 13. **Proteomics beyond trypsin.**

Liana Tsiatsiani, Albert JR Heck

The FEBS journal (2015-04-14) <a href="https://www.ncbi.nlm.nih.gov/pubmed/25823410">https://www.ncbi.nlm.nih.gov/pubmed/25823410</a>

DOI: <u>10.1111/febs.13287</u> · PMID: <u>25823410</u>

#### 14. Jesse G Meyer

ISRN computational biology (2014-04-22) https://www.ncbi.nlm.nih.gov/pubmed/30687733

DOI: 10.1155/2014/960902 · PMID: 30687733 · PMCID: PMC6347401

# 15. In Silico Proteome Cleavage Reveals Iterative Digestion Strategy for High Sequence Coverage

Jesse G Meyer

ISRN Computational Biology (2014-04-22) https://doi.org/gb6s2r

DOI: 10.1155/2014/960902 · PMID: 30687733 · PMCID: PMC6347401

# 16. Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS.

Gargi Choudhary, Shiaw-Lin Wu, Paul Shieh, William S Hancock

Journal of proteome research <a href="https://www.ncbi.nlm.nih.gov/pubmed/12643544">https://www.ncbi.nlm.nih.gov/pubmed/12643544</a>

DOI: 10.1021/pr025557n · PMID: 12643544

#### 17. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin.

Piero Giansanti, Liana Tsiatsiani, Teck Yew Low, Albert JR Heck

Nature protocols (2016-04-28) https://www.ncbi.nlm.nih.gov/pubmed/27123950

DOI: <u>10.1038/nprot.2016.057</u> · PMID: <u>27123950</u>

#### 18. Combination of Proteogenomics with Peptide

B Blank-Landeshammer, I Teichert, R Märker, M Nowrousian, U Kück, A Sickmann

mBio (2019-10-15) https://www.ncbi.nlm.nih.gov/pubmed/31615963

DOI: 10.1128/mbio.02367-19 · PMID: 31615963 · PMCID: PMC6794485

#### 19. **Precision**

Hao Yang, Yan-Chang Li, Ming-Zhi Zhao, Fei-Lin Wu, Xi Wang, Wei-Di Xiao, Yi-Hao Wang, Jun-Ling

Zhang, Fu-Qiang Wang, Feng Xu, ... Ping Xu

*Molecular & cellular proteomics : MCP* (2019-01-08)

https://www.ncbi.nlm.nih.gov/pubmed/30622160

DOI: <u>10.1074/mcp.tir118.000918</u> · PMID: <u>30622160</u> · PMCID: <u>PMC6442358</u>

# 20. Mass spectrometry-assisted venom profiling of Hypnale hypnale found in the Western Ghats of India incorporating de novo sequencing approaches.

Muralidharan Vanuopadath, Nithin Sajeev, Athira Radhamony Murali, Nayana Sudish, Nithya Kangosseri, Ivy Rose Sebastian, Nidhi Dalpatraj Jain, Amit Pal, Dileepkumar Raveendran, Bipin Gopalakrishnan Nair, Sudarslal Sadasivan Nair

*International journal of biological macromolecules* (2018-07-07)

https://www.ncbi.nlm.nih.gov/pubmed/29990557

DOI: 10.1016/j.ijbiomac.2018.07.016 · PMID: 29990557

### 21. Venomics and antivenomics of Indian spectacled cobra (Naja naja) from the Western Ghats

Muralidharan Vanuopadath, Dileepkumar Raveendran, Bipin Gopalakrishnan Nair, Sudarslal Sadasivan Nair

Acta Tropica (2022-04) https://doi.org/gpbzf7

DOI: <u>10.1016/j.actatropica.2022.106324</u> · PMID: <u>35093326</u>

# 22. Sequencing-Grade De novo Analysis of MS/MS Triplets (CID/HCD/ETD) From Overlapping Peptides

Adrian Guthals, Karl R Clauser, Ari M Frank, Nuno Bandeira

Journal of Proteome Research (2013-05-30) https://doi.org/f47kqd

DOI: <u>10.1021/pr400173d</u> · PMID: <u>23679345</u> · PMCID: <u>PMC4591044</u>

# 23. Multiple-Enzyme-Digestion Strategy Improves Accuracy and Sensitivity of Label- and Standard-Free Absolute Quantification to a Level That Is Achievable by Analysis with Stable Isotope-Labeled Standard Spiking.

Jacek R Wiśniewski, Christine Wegler, Per Artursson

Journal of proteome research (2018-10-30) <a href="https://www.ncbi.nlm.nih.gov/pubmed/30336047">https://www.ncbi.nlm.nih.gov/pubmed/30336047</a>

DOI: 10.1021/acs.jproteome.8b00549 · PMID: 30336047

### 24. Use of endoproteinase Lys-C from Lysobacter enzymogenes in protein sequence analysis.

PA Jekel, WJ Weijer, JJ Beintema

Analytical biochemistry (1983-10-15) https://www.ncbi.nlm.nih.gov/pubmed/6359954

DOI: <u>10.1016/0003-2697(83)90308-1</u> · PMID: <u>6359954</u>

# 25. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion.

Timo Glatter, Christina Ludwig, Erik Ahrné, Ruedi Aebersold, Albert JR Heck, Alexander Schmidt Journal of proteome research (2012-10-16) <a href="https://www.ncbi.nlm.nih.gov/pubmed/23017020">https://www.ncbi.nlm.nih.gov/pubmed/23017020</a> DOI: <a href="https://www.ncbi.nlm.nih.gov/pubmed/23017020">10.1021/pr300273g</a> · PMID: <a href="https://www.ncbi.nlm.nih.gov/pubmed/23017020">23017020</a>

# 26. The alpha-lytic protease gene of Lysobacter enzymogenes. The nucleotide sequence predicts a large prepro-peptide with homology to pro-peptides of other chymotrypsin-like enzymes.

DM Epstein, PC Wensink

*The Journal of biological chemistry* (1988-11-15)

https://www.ncbi.nlm.nih.gov/pubmed/3053694

PMID: 3053694

#### 27. Expanding Proteome Coverage with Orthogonal-specificity α-Lytic Proteases

Jesse G Meyer, Sangtae Kim, David A Maltby, Majid Ghassemian, Nuno Bandeira, Elizabeth A Komives

Molecular & Cellular Proteomics (2014-03) https://doi.org/f5vgcg

DOI: 10.1074/mcp.m113.034710 · PMID: 24425750 · PMCID: PMC3945911

# 28. Site-specific identification and quantitation of endogenous SUMO modifications under native conditions.

Ryan J Lumpkin, Hongbo Gu, Yiying Zhu, Marilyn Leonard, Alla S Ahmad, Karl R Clauser, Jesse G Meyer, Eric J Bennett, Elizabeth A Komives

Nature communications (2017-10-27) <a href="https://www.ncbi.nlm.nih.gov/pubmed/29079793">https://www.ncbi.nlm.nih.gov/pubmed/29079793</a>

DOI: <u>10.1038/s41467-017-01271-3</u> · PMID: <u>29079793</u> · PMCID: <u>PMC5660086</u>

#### 29. Purification and properties of an extracellular protease of Staphylococcus aureus.

GR Drapeau, Y Boily, J Houmard

*The Journal of biological chemistry* (1972-10-25)

https://www.ncbi.nlm.nih.gov/pubmed/4627743

PMID: 4627743

# 30. Mildly acidic conditions eliminate deamidation artifact during proteolysis: digestion with endoprotease Glu-C at pH 4.5.

Shanshan Liu, Kevin Ryan Moulton, Jared Robert Auclair, Zhaohui Sunny Zhou

Amino acids (2016-01-09) https://www.ncbi.nlm.nih.gov/pubmed/26748652

DOI: <u>10.1007/s00726-015-2166-z</u> · PMID: <u>26748652</u> · PMCID: <u>PMC4795971</u>

# 31. Specificity of endoproteinase Asp-N (Pseudomonas fragi): cleavage at glutamyl residues in two proteins.

D Ingrosso, AV Fowler, J Bleibaum, S Clarke

Biochemical and biophysical research communications (1989-08-15)

https://www.ncbi.nlm.nih.gov/pubmed/2669754

DOI: 10.1016/0006-291x(89)90848-6 · PMID: 2669754

#### 32. Chymotrypsin: molecular and catalytic properties.

W Appel

Clinical biochemistry (1986-12) https://www.ncbi.nlm.nih.gov/pubmed/3555886

DOI: 10.1016/s0009-9120(86)80002-9 · PMID: 3555886

#### 33. Mass-spectrometry-based draft of the human proteome.

Mathias Wilhelm, Judith Schlegl, Hannes Hahne, Amin Moghaddas Gholami, Marcus Lieberenz, Mikhail M Savitski, Emanuel Ziegler, Lars Butzmann, Siegfried Gessulat, Harald Marx, ... Bernhard Kuster

Nature (2014-05-29) https://www.ncbi.nlm.nih.gov/pubmed/24870543

DOI: 10.1038/nature13319 · PMID: 24870543

#### 34. Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics.

Xiaofeng Guo, David C Trudgian, Andrew Lemoff, Sivaramakrishna Yadavalli, Hamid Mirzaei *Molecular & cellular proteomics : MCP* (2014-04-02)

https://www.ncbi.nlm.nih.gov/pubmed/24696503

DOI: 10.1074/mcp.m113.035170 · PMID: 24696503 · PMCID: PMC4047476

# 35. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis.

Teck Yew Low, Sebastiaan van Heesch, Henk van den Toorn, Piero Giansanti, Alba Cristobal, Pim Toonen, Sebastian Schafer, Norbert Hübner, Bas van Breukelen, Shabaz Mohammed, ... Victor Guryev

Cell reports (2013-11-27) https://www.ncbi.nlm.nih.gov/pubmed/24290761

DOI: 10.1016/j.celrep.2013.10.041 · PMID: 24290761

#### 36. Protease bias in absolute protein quantitation.

Mao Peng, Nadia Taouatas, Salvatore Cappadona, Bas van Breukelen, Shabaz Mohammed, Arjen Scholten, Albert JR Heck

Nature methods (2012-05-30) https://www.ncbi.nlm.nih.gov/pubmed/22669647

DOI: 10.1038/nmeth.2031 · PMID: 22669647

# 37. Studies on the active site of clostripain. The specific inactivation by the chloromethyl ketone derived from -N-tosyl-L-lysine.

WH Porter, LW Cunningham, WM Mitchell

The Journal of biological chemistry (1971-12-25)

https://www.ncbi.nlm.nih.gov/pubmed/4332560

PMID: <u>4332560</u>

#### 38. LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification.

Pitter F Huesgen, Philipp F Lange, Lindsay D Rogers, Nestor Solis, Ulrich Eckhard, Oded Kleifeld, Theodoros Goulas, FXavier Gomis-Rüth, Christopher M Overall

Nature methods (2014-11-24) https://www.ncbi.nlm.nih.gov/pubmed/25419962

DOI: 10.1038/nmeth.3177 · PMID: 25419962

#### 39. Proteomic analyses using Grifola frondosa metalloendoprotease Lys-N.

Laura Hohmann, Carly Sherwood, Ashley Eastham, Amelia Peterson, Jimmy K Eng, James S Eddes, David Shteynberg, Daniel B Martin

Journal of proteome research (2009-03) https://www.ncbi.nlm.nih.gov/pubmed/19195997

DOI: <u>10.1021/pr800774h</u> · PMID: <u>19195997</u> · PMCID: <u>PMC2798736</u>

#### 40. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase.

Nadia Taouatas, Madalina M Drugan, Albert JR Heck, Shabaz Mohammed *Nature methods* (2008-04-20) <a href="https://www.ncbi.nlm.nih.gov/pubmed/18425140">https://www.ncbi.nlm.nih.gov/pubmed/18425140</a>
DOI: <a href="https://www.ncbi.nlm.nih.gov/pubmed/18425140">10.1038/nmeth.1204</a> · PMID: <a href="https://www.ncbi.nlm.nih.gov/pubmed/18425140">18425140</a>

#### 41. Cleavage specificities of the brother and sister proteases Lys-C and Lys-N.

Reinout Raijmakers, Pieter Neerincx, Shabaz Mohammed, Albert JR Heck *Chemical communications (Cambridge, England)* (2010-10-18)

https://www.ncbi.nlm.nih.gov/pubmed/20953479

DOI: 10.1039/c0cc02523b · PMID: 20953479

#### 42. A history of pepsin and related enzymes.

Joseph S Fruton

The Quarterly review of biology (2002-06) <a href="https://www.ncbi.nlm.nih.gov/pubmed/12089768">https://www.ncbi.nlm.nih.gov/pubmed/12089768</a>

DOI: <u>10.1086/340729</u> · PMID: <u>12089768</u>

#### 43. CRYSTALLINE PEPSIN: I. ISOLATION AND TESTS OF PURITY.

IH Northrop

The Journal of general physiology (1930-07-20)

https://www.ncbi.nlm.nih.gov/pubmed/19872561

DOI: 10.1085/jgp.13.6.739 · PMID: 19872561 · PMCID: PMC2141071

#### 44. CRYSTALLINE PEPSIN: II. GENERAL PROPERTIES AND EXPERIMENTAL METHODS.

JH Northrop

*The Journal of general physiology* (1930-07-20)

https://www.ncbi.nlm.nih.gov/pubmed/19872562

DOI: <u>10.1085/jgp.13.6.767</u> · PMID: <u>19872562</u> · PMCID: <u>PMC2141088</u>

#### 45. **CRYSTALLINE PEPSIN.**

JH Northrop

Science (New York, N.Y.) (1929-05-31) <a href="https://www.ncbi.nlm.nih.gov/pubmed/17758437">https://www.ncbi.nlm.nih.gov/pubmed/17758437</a>

DOI: 10.1126/science.69.1796.580 · PMID: 17758437

#### 46. Protein disulfide bond determination by mass spectrometry.

Jeffrey J Gorman, Tristan P Wallis, James J Pitt

Mass spectrometry reviews <a href="https://www.ncbi.nlm.nih.gov/pubmed/12476442">https://www.ncbi.nlm.nih.gov/pubmed/12476442</a>

DOI: 10.1002/mas.10025 · PMID: 12476442

# 47. Facilitating protein disulfide mapping by a combination of pepsin digestion, electron transfer higher energy dissociation (EThcD), and a dedicated search algorithm SlinkS.

Fan Liu, Bas van Breukelen, Albert JR Heck

*Molecular & cellular proteomics : MCP* (2014-06-30)

https://www.ncbi.nlm.nih.gov/pubmed/24980484

DOI: <u>10.1074/mcp.o114.039057</u> · PMID: <u>24980484</u> · PMCID: <u>PMC4189002</u>

# 48. Online, High-Pressure Digestion System for Protein Characterization by Hydrogen/Deuterium Exchange and Mass Spectrometry

Lisa M Jones, Hao Zhang, Ilan Vidavsky, Michael L Gross

Analytical Chemistry (2010-01-22) https://doi.org/b993rm

DOI: 10.1021/ac902477u · PMID: 20095571 · PMCID: PMC2826105

#### 49. Hydrogen/deuterium exchange in mass spectrometry

Yury Kostyukevich, Thamina Acter, Alexander Zherebker, Arif Ahmed, Sunghwan Kim, Eugene Nikolaev

Mass Spectrometry Reviews (2018-03-30) https://doi.org/gffzx8

DOI: 10.1002/mas.21565 · PMID: 29603316

## 50. Simultaneous Quantification of the Acetylome and Succinylome by 'One-Pot' Affinity Enrichment

Nathan Basisty, Jesse G Meyer, Lei Wei, Bradford W Gibson, Birgit Schilling

PROTEOMICS (2018-08-19) https://doi.org/gn4cmb

DOI: 10.1002/pmic.201800123 · PMID: 30035354 · PMCID: PMC6175148

#### 51. The Nobel Prize in Chemistry 2002

NobelPrize.org

https://www.nobelprize.org/prizes/chemistry/2002/summary/

#### 52. Interpretation of Shotgun Proteomic Data

Alexey I Nesvizhskii, Ruedi Aebersold

Molecular & Cellular Proteomics (2005-10) https://doi.org/cm99cj

DOI: https://doi.org/10.1074/mcp.r500012-mcp200

# 53. In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics

Enrique Audain, Julian Uszkoreit, Timo Sachsenberg, Julianus Pfeuffer, Xiao Liang, Henning Hermjakob, Aniel Sanchez, Martin Eisenacher, Knut Reinert, David L Tabb, ... Yasset Perez-Riverol

Journal of Proteomics (2017-01) https://doi.org/f9r8r6

DOI: <u>10.1016/j.jprot.2016.08.002</u> · PMID: <u>27498275</u>

# 54. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification

Jürgen Cox, Matthias Mann

Nature Biotechnology (2008-11-30) https://doi.org/crn24x

DOI: 10.1038/nbt.1511 · PMID: 19029910

#### 55. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometrybased proteomics

Andy T Kong, Felipe V Leprevost, Dmitry M Avtonomov, Dattatreya Mellacheruvu, Alexey I Nesvizhskii

Nature Methods (2017-04-10) https://doi.org/f9z6p7

DOI: 10.1038/nmeth.4256 · PMID: 28394336 · PMCID: PMC5409104

# 56. Probability-based protein identification by searching sequence databases using mass spectrometry data.

DN Perkins, DJ Pappin, DM Creasy, JS Cottrell

Electrophoresis (1999-12) https://www.ncbi.nlm.nih.gov/pubmed/10612281

DOI: 10.1002/(sici)1522-2683(19991201)20:18<3551::aid-elps3551>3.0.co;2-2 · PMID: 10612281

### 57. MS-GF+ makes progress towards a universal database search tool for proteomics

Sangtae Kim, Pavel A Pevzner

Nature Communications (2014-10-31) <a href="https://doi.org/ggkdq8">https://doi.org/ggkdq8</a>

DOI: <u>10.1038/ncomms6277</u> · PMID: <u>25358478</u> · PMCID: <u>PMC5036525</u>

# 58. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments

Brendan MacLean, Daniela M Tomazela, Nicholas Shulman, Matthew Chambers, Gregory L Finney, Barbara Frewen, Randall Kern, David L Tabb, Daniel C Liebler, Michael J MacCoss *Bioinformatics* (2010-02-09) https://doi.org/bqx9rg

DOI: <u>10.1093/bioinformatics/btg054</u> · PMID: <u>20147306</u> · PMCID: <u>PMC2844992</u>

# 59. **DIA-NN:** neural networks and interference correction enable deep proteome coverage in high throughput

Vadim Demichev, Christoph B Messner, Spyros I Vernardis, Kathryn S Lilley, Markus Ralser *Nature Methods* (2019-11-25) <a href="https://doi.org/gj9xg">https://doi.org/gj9xg</a>j

DOI: <u>10.1038/s41592-019-0638-x</u> · PMID: <u>31768060</u> · PMCID: <u>PMC</u>6949130

# 60. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry

Ngoc Hieu Tran, Rui Qiao, Lei Xin, Xin Chen, Chuyi Liu, Xianglilan Zhang, Baozhen Shan, Ali Ghodsi, Ming Li

Nature Methods (2018-12-20) <a href="https://doi.org/gftvmn">https://doi.org/gftvmn</a> DOI: <a href="https://doi.org/gftvmn">10.1038/s41592-018-0260-3</a> PMID: <a href="https://doi.org/gftvmn">30573815</a>

# 61. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues

Roland Bruderer, Oliver M Bernhardt, Tejas Gandhi, Saša M Miladinović, Lin-Yang Cheng, Simon Messner, Tobias Ehrenberger, Vito Zanotelli, Yulia Butscheid, Claudia Escher, ... Lukas Reiter *Molecular & Cellular Proteomics* (2015-05) https://doi.org/f7b76h

DOI: 10.1074/mcp.m114.044305 · PMID: 25724911 · PMCID: PMC4424408