A Practical Beginner's Guide to Proteomics

This manuscript (permalink) was automatically generated from jessegmeyerlab/proteomics-tutorial@21ec587 on January 16, 2022.

Authors

- **Dina Schuster**
  0000-0001-6611-8237 · dschust-r · dina_sch
  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

- **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

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, 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

[paragraph about what proteomics means today] 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.

[history of proteomics? how we got here] How did we get here? Modern proteomics really started around 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 machine 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].

The wide variety of experimental goals leads to equal diversity in 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 Acquisition 10. Basic Data Analysis 11. Biological Interpretation 12. Experimental considerations and design

[paragraph about what proteomics can do - leads into the next section] A wide range of questions are addressable with proteomics experiments, 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.

Types of Experiments

[List of common types of experiments and brief description]

- Protein abundance changes
- Phosphoproteomics
- Glycoproteomics
- Structural techniques (XL-MS, HDX-MS, FPOP, protein-painting, LiP-MS, ...)
- Protein stability measurements (Thermal denaturation)
- PPIs: AP-MS, APEX, BioID
- ...

Protein Extraction

Discussion of methods for protein extraction and solubilizaition.

1. Choice of Lysis buffer

- Urea can cause chemical modifications

2. chemicals to avoid
3. removal of contaminations, Protein Precipitation
4. protein alkylation

- choices of reduction and alkylation reagents, TCEP/DTT/2BME, Chloroacetamide/iodoacetamide, n-ethyl maleimide

Proteolysis

1. discussion of protein sequence coverage is determined by the choice of proteolysis
2. why trypsin is the most common choice (charge and length character)
3. theoretical studies of proteolysis and enzyme [7]
4. Challenges associated with alternative enzyme choices (non-specific and semi-specific enzymes)
5. Alternative enzyme choices (one paragraph each?) - LysC
6. GluC
7. AspN
8. Alpha-lytic protease [8] and how it enables mapping human SUMO sites [9].
9. others?

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

Peptide or Protein Enrichment

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

- coIP
- APEX
- bioID
- bioplex

Peptide enrichment

- antibody enrichments of modifications, e.g. lysine acetylation [10].
- 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” [11/].

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 [12] [13]. 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: - MaxQuant [14/] - MSFragger [15/] - Mascot (for smaller data sets) [16/] [17] - MS-GF+ [18]

DIA freeware: - MaxDIA (within MaxQuant) [14/] - Skyline [19] - DIA-NN [20] Targeted proteomics freeware: - Skyline [19]

DDA non-freeware: - ProteomeDiscoverer[21] - Mascot (for larger data sets) [16/] - Spectromine [22/?gclid=Cj0KCQiAoY-PBhCNARIsABcz770mjUz6iavBr9Ql7RPUdMvaHu9RYgPNrEfZco1wExEeoFwnQXuCHscaAlgBEALw_wcB] - PEAKS [23/]

DIA non-freeware: - Spectronaut [24/?gclid=Cj0KCQiAoY-PBhCNARIsABcz770nuaU2SgIriS-ZJJGsC6CtzXc9AC8b9K3w5FlFDsDfGtnuUjIhankaAvegEALw_wcB] - PEAKS [23/]

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: https://doi.org/10.1002/rcm.1290020802

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: 10.1016/1044-0305(94)80016-2 · PMID: 24226387

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 (2007-02-27) https://doi.org/djz7fz
DOI: https://doi.org/10.1038/nmeth1019

7. **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: https://doi.org/10.1155/2014/960902

8. **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: https://doi.org/10.1074/mcp.m113.034710

9. **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) https://www.ncbi.nlm.nih.gov/pubmed/29079793
DOI: 10.1038/s41467-017-01271-3 · PMID: 29079793 · PMCID: PMC5660086

10. **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

11. **The Nobel Prize in Chemistry 2002**
NobelPrize.org
https://www.nobelprize.org/prizes/chemistry/2002/summary/

12. **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

13. **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: 10.1016/j.jprot.2016.08.002 · PMID: 27498275

14. **MaxQuant** https://www.maxquant.org/

15. **MSFragger**
MSFragger
https://msfragger.nesvilab.org/

16. **Mascot search engine | Protein identification software for mass spec data**
https://www.matrixscience.com/

17. **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

18. **MS-GF+ makes progress towards a universal database search tool for proteomics**
Sangtae Kim, Pavel A Pevzner
Nature Communications (2014-10-31) https://doi.org/ggkdq8
DOI: https://doi.org/10.1038/ncomms6277

19. **Start Page: /home/software/Skyline**
https://skyline.ms/project/home/software/Skyline/begin.view

20. **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) https://doi.org/gj9xgj
DOI: https://doi.org/10.1038/s41592-019-0638-x

21. **Proteome Discoverer Software - US** //www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html

22. **Proteomics Analysis Software | Shotgun Proteomics | DDA Proteomics**
Biognosys
https://biognosys.com/software/spectromine/

23. **Protein Identification & Quantification Software, PTM & Variant Search**
Bioinformatics Solutions Inc.
https://www.bioinfor.com/peaks-studio/

24. **Proteomics Software | DIA Proteomics | Discovery Proteomics**
Biognosys
https://biognosys.com/software/spectronaut/