Lecture 8. Proteomics and methods of proteome investigation.

Learning outcomes:

1. Give the definition to the terms “olygopeptides”, “polypeptides”, “proteins”, “proteome”, “proteomics”.
2. Explain and analyze the experimental methods of proteomics: MALDI-mass-spectrometry, ESI-mass-spectrometry, different types of chromatography, 2D- and 3D-PAGE, ELISA, nuclear magnetic resonance (NMR), X-ray diffraction and etc. What are the reasons for selecting any of these methods for specific proteins?
3. Describe the computational (bioinformatical) methods of protein research: amino acid sequence alignment, protein structure prediction, analysis of X-ray diffraction pattern and 3D-modelling of protein structure.
4. Characterize the Protein Data Bank (PDB), Uni-ProtKB/Swiss-Prot and other bioinformatical databases of protein information.

Proteomics is the large-scale study of proteins. **Proteins** are vital parts of living organisms, with many functions. The **proteome** is the **entire set of proteins that is produced or modified by an organism or system**. Proteomics has enabled the identification of ever increasing numbers of protein. This varies with time and distinct requirements, or stresses, that a cell or organism undergoes. Proteomics is an interdisciplinary domain that has benefitted greatly from the genetic information of various genome projects, including the **Human Genome Project**. It covers the exploration of proteomes from the overall level of protein composition, structure, and activity. It is an important component of functional genomics.

Proteomics generally refers to the **large-scale experimental analysis** of proteins and proteomes, but often is used specifically to refer to **protein purification and mass spectrometry**.

There are several specific techniques and protocols that use **antibodies** for **protein detection**. **The enzyme-linked immunosorbent assay (ELISA)** has been used for decades to detect and quantitatively measure proteins in samples. The **western blot** may be used for detection and quantification of individual proteins, where in an initial step, a complex protein mixture is separated using **SDS-PAGE** and then the protein of interest is identified using an antibody.

Mass spectrometry (MS) is an analytical technique that is used to measure the **mass-to-charge ratio of ions**. The results are typically presented as a mass spectrum, a plot of intensity as a function of the mass-to-charge ratio. Mass spectrometry is used in many different fields and is applied to pure samples as well as complex mixtures. **Two techniques** often used with liquid and solid biological samples include **electrospray ionization** (invented by John Fenn) and **matrix-assisted laser desorption/ionization (MALDI**, initially developed as a similar technique "Soft Laser Desorption (SLD)" by K. Tanaka for which a Nobel Prize was awarded and as MALDI by M. Karas and F. Hillenkamp). An important enhancement to the mass resolving and mass determining capabilities of mass spectrometry is using it in tandem with **chromatographic** and other **separation** techniques.

Chromatography is a laboratory technique for the separation of a mixture. The mixture is dissolved in a fluid (gas, solvent, water, ...) called the **mobile phase**, which carries it through a system (a column, a capillary tube, a plate, or a sheet) on which is fixed a material called the **stationary phase**. The different constituents of the mixture have different affinities for the stationary phase. The different molecules stay longer or shorter on the stationary phase, depending on their interactions with its surface sites. So, they travel at different apparent velocities in the mobile fluid, causing them to separate. The separation is based on the differential partitioning between the mobile and the stationary phases. Subtle differences in a compound's partition coefficient result in differential retention on the stationary phase and thus affect the separation.

Much proteomics data is collected with the help of high throughput technologies such as mass spectrometry and microarray. It would often take weeks or months to analyze the data and

perform comparisons by hand. For this reason, biologists and chemists are collaborating with computer scientists and mathematicians to create programs and pipeline to computationally analyze the protein data. Using **bioinformatics** techniques, researchers are capable of faster analysis and data storage. A good place to find lists of current programs and databases is on the **ExPASy** bioinformatics resource portal. **The applications of bioinformatics-based proteomics includes medicine, disease diagnosis, biomarker identification, and many more.**

Mass spectrometry and microarray produce peptide fragmentation information but do not give identification of specific proteins present in the original sample. Due to the lack of specific protein identification, past researchers were forced to decipher the peptide fragments themselves. However, there are currently programs available for protein identification. These programs take the peptide sequences output from mass spectrometry and microarray and return information about matching or similar proteins. This is done through algorithms implemented by the program which perform alignments with proteins from known databases such as **UniProt** and **PROSITE** to predict what proteins are in the sample with a degree of certainty.

The biomolecular structure forms the **3D configuration** of the protein. Understanding the protein's structure aids in identification of the protein's interactions and function. It used to be that the 3D structure of proteins could only be determined using **X-ray crystallography** and **NMR spectroscopy**. As of 2017, Cryo-electron microscopy is a leading technique, solving difficulties with crystallization (in X-ray crystallography) and conformational ambiguity (in NMR); resolution was 2.2Å as of 2015. Now, through bioinformatics, there are computer programs that can in some cases predict and model the structure of proteins. These programs use the chemical properties of amino acids and structural properties of known proteins to predict the 3D model of sample proteins. This also allows scientists to model protein interactions on a larger scale. In addition, biomedical engineers are developing methods to factor in the flexibility of protein structures to make comparisons and predictions.

Most programs available for protein analysis are not written for proteins that have undergone **post-translational modifications**. Some programs will accept post-translational modifications to aid in protein identification but then ignore the modification during further protein analysis. It is important to account for these modifications since they can affect the protein's structure. In turn, computational analysis of post-translational modifications has gained the attention of the scientific community. The current post-translational modification programs are only predictive. Chemists, biologists and computer scientists are working together to create and introduce new pipelines that allow for analysis of post-translational modifications that have been experimentally identified for their effect on the protein's structure and function.

The questions for self - control:

1. What are the “olygopeptides”, “polypeptides”, “proteins”, “proteome”, “proteomics”?
2. Experimental methods of proteomics.
3. Bioinformatical methods of protein research.
4. Bioinformatical databases of protein information.

Recommended readings:

1. Anderson NL, Anderson NG; Anderson (1998). "Proteome and proteomics: new technologies, new concepts, and new words". Electrophoresis. 19 (11): 1853–61. doi:10.1002/elps.1150191103. PMID 9740045. S2CID 28933890.
2. Blackstock WP, Weir MP; Weir (1999). "Proteomics: quantitative and physical mapping of cellular proteins". Trends Biotechnol. 17 (3): 121–7. doi:10.1016/S0167-7799(98)01245-1. PMID 10189717.
3. Anderson, Johnathon D.; Johansson, Henrik J.; Graham, Calvin S.; Vesterlund, Mattias; Pham, Missy T.; Bramlett, Charles S.; Montgomery, Elizabeth N.; Mellema, Matt S.; Bardini, Renee L. (2016-03-01). "Comprehensive Proteomic Analysis of Mesenchymal Stem Cell Exosomes Reveals Modulation of Angiogenesis via Nuclear Factor-KappaB

Signaling". Stem Cells. 34 (3): 601–613. doi:10.1002/stem.2298. ISSN 1549-4918. PMC 5785927. PMID 26782178.
4. Hood, Leroy; Rowen, Lee (2013-09-13). "The human genome project: big science transforms biology and medicine". Genome Medicine. 5 (9): 79. doi:10.1186/gm483. PMC 4066586. PMID 24040834.
5. Wilson, DH; Rissin, DM; Kan, CW; Fournier, DR; Piech, T; Campbell, TG; Meyer, RE; Fishburn, MW; Cabrera, C; Patel, PP; Frew, E; Chen, Y; Chang, L; Ferrell, EP; von Einem, V; McGuigan, W; Reinhardt, M; Sayer, H; Vielsack, C; Duffy, DC (2016). "The Simoa HD-1 Analyzer: A Novel Fully Automated Digital Immunoassay Analyzer with Single-Molecule Sensitivity and Multiplexing". J Lab Autom. 21 (4): 533–47. doi:10.1177/2211068215589580. PMID 26077162.
6. Sparkman, O. David (2000). Mass spectrometry desk reference. Pittsburgh: Global View Pub. ISBN 978-0-9660813-2-9.
7. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (October 1989). "Electrospray ionization for mass spectrometry of large biomolecules". Science. 246 (4926): 64–71. Bibcode:1989Sci...246...64F. CiteSeerX 10.1.1.522.9458. doi:10.1126/science.2675315. PMID 2675315.
8. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T (1988). "Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of flight Mass Spectrometry". Rapid Commun Mass Spectrom. 2 (20): 151–3. Bibcode:1988RCMS....2..151T. doi:10.1002/rcm.1290020802.
9. Karas M, Bachman D, Bahr U, Hillenkamp F (1987). "Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds". Int J Mass Spectrom Ion Proc. 78: 53–68. Bibcode:1987IJMSI..78...53K. doi:10.1016/0168-1176(87)87041-6.
10. McMurry J (2011). Organic chemistry: with biological applications (2nd ed.). Belmont, CA: Brooks/Cole. pp. 395. ISBN 9780495391470.
11. "UniProt". www.uniprot.org.
12. "ExPASy - PROSITE". prosite.expasy.org.
13. Wang H, Chu C, Wang W, Pai T; Chu; Wang; Pai (April 2014). "A local average distance descriptor for flexible protein structure comparison". BMC Bioinformatics. 15 (95): 1471–2105. doi:10.1186/1471-2105-15-95. PMC 3992163. PMID 24694083.
14. Petrov D, Margreitter C, Gandits M, Ostenbrink C, Zagrovic B; Margreitter; Grandits; Oostenbrink; Zagrovic (July 2013). "A systematic framework for molecular dynamics simulations of protein post-translational modifications". PLOS Computational Biology. 9 (7): e1003154. Bibcode:2013PLSCB...9E3154P. doi:10.1371/journal.pcbi.1003154. PMC 3715417. PMID 23874192.
15. Margreitter C, Petro D, Zagrovic B; Petrov; Zagrovic (May 2013). "Vienna-PTM web server: a toolkit for MD simulations of portein post-translational modifications". Nucleic Acids Res. 41 (Web Server issue): W422–6. doi:10.1093/nar/gkt416. PMC 3692090. PMID 23703210.