PRACTICAL APPLICATIONS OF PROTEOMICS-A TECHNIQUE FOR LARGE- SCALE STUDY OF PROTEINS: AN OVERVIEW

Satyanand Tyagi[1]*, Raghvendra[2], Usha Singh [3], Taruna Kalra[4], Kavita Munjal[4], Vikas[4]
[1] Department of Pharmaceutical Chemistry, K.N.G.D Modi Institute of Pharmaceutical Education & Research, Modinagar, Ghaziabad (dt), U.P, India-201204.
[2] Department of Pharmaceutics, Aligarh College of Pharmacy, Aligarh (dt), U.P, India-202001.
[3] Pharmacist, Max Neeman International Ltd, Okhla Phase-3, New Delhi, India-110020.
[4] Department of Pharmacy, B.S.Anangpuria Educational Institutes, Faridabad (dt), Haryana, India-121004.
*Email: sntyagi9@yahoo.com

ABSRTACT

The dream of having genomes completely sequenced is now a reality. The complete sequence of many genomes including the human one is known. However, the understanding of probably half a million human proteins encoded by less than 30'000 genes is still a long way away and the hard work to unravel the complexity of biological systems is yet to come. A new fundamental concept called proteome (protein complement to a genome) has recently emerged that should drastically help to unravel biochemical and physiological mechanisms of complex multivariate diseases at the functional molecular level. The discipline of proteomics has been initiated to complement physical genomic research. Proteomics can be defined as the qualitative and quantitative comparison of proteomes under different conditions to further unravel biological processes. Proteomics is the premier international source for information on all aspects of applications and technologies in proteomics. It provides the most comprehensive coverage in the field, spanning significant technical developments in all major gel and non-gel based proteomic platforms and their application in every area of life sciences. In the present article, we have concentrated on various practical applications of Proteomics. This article presents a brief review of Proteomics with an emphasis on its various practical applications; the article reveals that how this technique is helpful and useful in study of large scale study of proteins.

Keywords: genomes, proteomics, proteome, proteins, expression proteomics, interaction proteomics, structural proteomics

INTRODUCTON

The focus of proteomics is a biological group called the proteome. The proteome is dynamic, defined as the set of proteins expressed in a specific cell, given a particular set of conditions. Within a given human proteome, the number of proteins can be as large as 2 million. Proteins themselves are macromolecules: long chains of amino acids. This amino acid chain is constructed when the cellular machinery of the ribosome translates RNA transcripts from DNA in the cell's nucleus. The transfer of information within cells commonly follows this path, from DNA to RNA to protein.

Proteins can be organized in four structural levels:

- Primary (1°): The amino acid sequence, containing members of a (usually) twenty-unit alphabet

- Secondary (2°): Local folding of the amino acid sequence into α helices and β sheets

- Tertiary (3°): 3D conformation of the entire amino acid sequence

- Quaternary (4°): Interaction between multiple small peptides or protein subunits to create a large unit

Each level of protein structure is essential to the finished molecule's function. The primary sequence of the amino acid chain determines where secondary structures will form, as well as the overall shape of the final 3D conformation. The 3D conformation of each small peptide or subunit determines the final structure and function of a protein conglomerate.

There are many different subdivisions of proteomics, including:

- Structural proteomics ---- in-depth analysis of protein structure

- Expression proteomics ---- analysis of expression and differential expression of proteins

- Interaction proteomics --- analysis of interactions between proteins to characterize complexes and determine function.

Proteomics has both a physical laboratory component and a computational component. These two parts are often linked together; at times data derived from laboratory work can be fed directly into sequence and structure prediction algorithms. Mass spectrometry of multiple types is used most frequently for this purpose[1].

Proteomics is the large-scale study of proteins, particularly their structures and functions[2, 3]. Proteins are vital parts of living organisms, as they are the main components of the physiological metabolic pathways of cells. The term "proteomics" was first coined in 1997 to make an analogy with genomics, the study of the genes[4]. The word "proteome" is a blend of "protein" and "genome", and was coined by Marc Wilkins in 1994 while working on the concept as a PhD student[5]. The proteome is the entire complement of proteins, including the modifications made to a particular set of proteins, produced by an organism or

system. This will vary with time and distinct requirements, or stresses, that a cell or organism undergoes.

Complexity of the problem

After genomics, proteomics is considered the next step in the study of biological systems. It is much more complicated than genomics mostly because while an organism's genome is more or less constant, the proteome differs from cell to cell and from time to time. This is because distinct genes are expressed in distinct cell types. This means that even the basic set of proteins which are produced in a cell needs to be determined. In the past this was done by mRNA analysis, but this was found not to correlate with protein content[6]. It is now known that mRNA is not always translated into protein[7], and the amount of protein produced for a given amount of mRNA depends on the gene it is transcribed from and on the current physiological state of the cell. Proteomics confirms the presence of the protein and provides a direct measure of the quantity present.

Post-translational modifications

Not only does the translation from mRNA cause differences, many proteins are also subjected to a wide variety of chemical modifications after translation. A lot of these post-translational modifications are critical to the protein's function.

Phosphorylation

One such modification is phosphorylation, which happens to many enzymes and structural proteins in the process of cell signaling. The addition of a phosphate to particular amino acids—most commonly serine and threonine[8] mediated by serine/threonine kinases, or more rarely tyrosine mediated by tyrosine kinases—causes a protein to become a target for binding or interacting with a distinct set of other proteins that recognize the phosphorylated domain.

Because protein phosphorylation is one of the most-studied protein modifications many "proteomic" efforts are geared to determining the set of phosphorylated proteins in a particular cell or tissue-type under particular circumstances. This alerts the scientist to the signaling pathways that may be active in that instance.

Ubiquitination

Ubiquitin is a small protein that can be affixed to certain protein substrates by enzymes called E3 ubiquitin ligases. Determining which proteins are poly-ubiquitinated can be helpful in understanding how protein pathways are regulated. This is therefore an additional legitimate "proteomic" study. Similarly, once it is determined what substrates are ubiquitinated by each ligase, determining the set of ligases expressed in a particular cell type will be helpful.

Distinct proteins are made under distinct settings

Even if one is studying a particular cell type, that cell may make different sets of proteins at different times, or under different conditions. Furthermore, as mentioned, any one protein can undergo a wide range of post-translational modifications. Therefore a "proteomics" study can become

quite complex very quickly, even if the object of the study is very restricted. In more ambitious settings, such as when a biomarker for a tumor is sought - when the proteomics scientist is obliged to study sera samples from multiple cancer patients - the amount of complexity that must be dealt with is as great as in any modern biological project.

Limitations to genomic study

Scientists are very interested in proteomics because it gives a much better understanding of an organism than genomics. First, the level of transcription of a gene gives only a rough estimate of its level of expression into a protein. An mRNA produced in abundance may be degraded rapidly or translated inefficiently, resulting in a small amount of protein. Second, as mentioned above many proteins experience post-translational modifications that profoundly affect their activities; for example some proteins are not active until they become phosphorylated. Methods such as phosphoproteomics and glycoproteomics are used to study post-translational modifications. Third, many transcripts give rise to more than one protein, through alternative splicing or alternative post-translational modifications. Fourth, many proteins form complexes with other proteins or RNA molecules, and only function in the presence of these other molecules. Finally, protein degradation rate plays an important role in protein content[9].

METHODS OF STUDYING PROTEINS

Determining proteins which are post-translationally modified

One way in which a particular protein can be studied is to develop an antibody which is specific to that modification. For example, there are antibodies which only recognize certain proteins when they are tyrosine-phosphorylated; also, there are antibodies specific to other modifications. These can be used to determine the set of proteins that have undergone the modification of interest. For sugar modifications, such as glycosylation of proteins, certain lectins have been discovered which bind sugars. These too can be used. A more common way to determine post-translational modification of interest is to subject a complex mixture of proteins to electrophoresis in "two-dimensions", which simply means that the proteins are electrophoresed first in one direction, and then in another... this allows small differences in a protein to be visualized by separating a modified protein from its unmodified form. This methodology is known as "two-dimensional gel electrophoresis".

Recently, another approach has been developed called PROTOMAP which combines SDS-PAGE with shotgun proteomics to enable detection of changes in gel-migration such as those caused by proteolysis or post translational modification.

Determining the existence of proteins in complex mixtures

Classically, antibodies to particular proteins or to their modified forms have been used in biochemistry and cell biology studies. These are among the most common tools used by practicing biologists today. For more quantitative determinations of protein amounts, techniques such as

ELISAs can be used. For proteomic study, more recent techniques such as Matrix-assisted laser desorption/ionization have been employed for rapid determination of proteins in particular mixtures.

Establishing protein-protein interactions

Most proteins function in collaboration with other proteins, and one goal of proteomics is to identify which proteins interact. This is especially useful in determining potential partners in cell signaling cascades.

Several methods are available to probe protein-protein interactions. The traditional method is yeast two-hybrid analysis. New methods include protein microarrays, immunoaffinity chromatography followed by mass spectrometry, dual polarisation interferometry and experimental methods such as phage display and computational methods.

PRACTICAL APPLICATIONS OF PROTEOMICS[10-13]

One of the most promising developments to come from the study of human genes and proteins has been the identification of potential new drugs for the treatment of disease. This relies on genome and proteome information to identify proteins associated with a disease, which computer software can then use as targets for new drugs. For example, if a certain protein is implicated in a disease, its 3D structure provides the information to design drugs to interfere with the action of the protein. A molecule that fits the active site of an enzyme, but cannot be released by the enzyme, will inactivate the enzyme. This is the basis of new drug-discovery tools, which aim to find new drugs to inactivate proteins involved in disease. As genetic differences among individuals are found, researchers expect to use these techniques to develop personalized drugs that are more effective for the individual.

A computer technique which attempts to fit millions of small molecules to the three-dimensional structure of a protein is called "virtual ligand screening". The computer rates the quality of the fit to various sites in the protein, with the goal of either enhancing or disabling the function of the protein, depending on its function in the cell. A good example of this is the identification of new drugs to target and inactivate the HIV-1 protease. The HIV-1 protease is an enzyme that cleaves a very large HIV protein into smaller, functional proteins. The virus cannot survive without this enzyme; therefore, it is one of the most effective protein targets for killing HIV.

Biomarkers

Understanding the proteome, the structure and function of each protein and the complexities of protein-protein interactions will be critical for developing the most effective diagnostic techniques and disease treatments in the future. An interesting use of proteomics is using specific protein biomarkers to diagnose disease. A number of techniques allow to test for proteins produced during a particular disease, which helps to diagnose the disease quickly. Techniques include western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA) or mass spectrometry. The following are some of the diseases that have characteristic biomarkers that physicians can use for diagnosis-:

Alzheimer's disease

In Alzheimer's disease, elevations in beta secretase create amyloid/beta-protein, which causes plaque to build up in the patient's brain, which is thought to play a role in dementia. Targeting this enzyme decreases the amyloid/beta-protein and so slows the progression of the disease. A procedure to test for the increase in amyloid/beta-protein is immunohistochemical staining, in which antibodies bind to specific antigens or biological tissue of amyloid/beta-protein.

Heart disease

Heart disease is commonly assessed using several key protein based biomarkers. Standard protein biomarkers for cardiovascular diseases include interleukin-6, interleukin-8, serum amyloid A protein, fibrinogen, and troponins. Cardiac troponin I (cTnI) increases in concentration within 3 to 12 hours of initial cardiac injury and remains elevated days after an acute myocardial infarction. A number of commercial antibody based assays as well as other methods are used in hospitals as primary tests for acute heart infarctions.

REFERENCES

1. Colinge, Jacques, Keiryn L. Bennett, "Introduction to Computational Proteomics", PLoS Comput Biol, 3(7), 2007, 114.

2. Anderson NL, Anderson NG, "Proteome and proteomics: new technologies, new concepts, and new words", Electrophoresis, 19 (11), 1998, 1853–1861.

3. Blackstock WP, Weir MP, "Proteomics: quantitative and physical mapping of cellular proteins", Trends Biotechnol, 17 (3), 1999, 121–127.

4. P. James, "Protein identification in the post-genome era: the rapid rise of proteomics", Quarterly reviews of biophysics, 30 (4), 1997, 279–331.

5. Marc R. Wilkins, Christian Pasquali, Ron D. Appel, et al, From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis", Nature Biotechnology, 14 (1), 1996, 61–65.

6. Simon Rogers, Mark Girolami, Walter Kolch, et al, "Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models", Bioinformatics 24 (24), 2008, 2894–2900.

7. Vikas Dhingraa, Mukta Gupta, Tracy Andacht, Zhen F. Fu, "New frontiers in proteomics research: A

perspective", International Journal of Pharmaceutics, 299 (1–2), 2005, 1–18.

8. Olsen JV, Blagoev B, Gnad F, et al, "Global, in vivo and site-specific phosphorylation dynamics in signaling networks", Cell, 127 (3), 2006, 635–648.

9. Archana Belle, Amos Tanay, Ledion Bitincka, Ron Shamir, "Quantification of protein half-lives in the budding yeast proteome", PNAS 103 (35), 2006, 13004–13009.

10. Hye A, Lynham S, Thambisetty M, et al, "Proteome-based plasma biomarkers for Alzheimer's disease", Brain 129 (Pt 11), 2006, 3042–3050.

11. Perroud B, Lee J, Valkova N, et al, "Pathway analysis of kidney cancer using proteomics and metabolic profiling", Mol Cancer, 5, 2006, 64.

12. 12.Yohannes E, Chang J, Christ GJ, Davies KP, Chance MR, "Proteomics analysis identifies molecular targets related to diabetes mellitus-associated bladder dysfunction", Mol. Cell Proteomics, **7** (7), 2008, 1270–1285.

13. Vasan RS, "Biomarkers of cardiovascular disease: molecular basis and practical considerations", Circulation, 113 (19), 2006, 2335–62.

****