

A personal view on proteomics

Editorial

When Marc Wilkins coined the word “Proteomics” in 1994,¹ describing a global approach to identify and quantify all proteins in a cell, a tissue, an organism at a certain state, this idea was not really new. Already in the 1970s Norman Anderson used 2D gel electrophoresis to visualize the proteins in human plasma.^{2,3} In this time Edman sequencing was the only method to identify proteins, comprehensive databases of any genome were not available, each sequenced gene could be published. Also in 1994, the technique to identify proteins still immature and only the genome of the yeast was available. In the same time the human genome sequencing became a major program of the US government, described under <http://www.genome.gov/12011239>. The sequence of the first individual human genome led to a big surprise, from originally over 1M expected proteins, only around 22,000 coding sequences were found in the human genome when finally finished around 2005. The rest of the genome was named “junk DNA”, but nature does not generate energy wasting junk, the functions have yet to be discovered.⁴ In the meanwhile many more genomes from different species have been sequenced, also as many as 1,000s of individual human genomes.

Both Genomics and Proteomics gave big promises to revolutionise the development of drugs, to understand diseases, and to identify new biomarkers.

Proteomics wanted to answer the major question Genomics cannot answer: how biology works. But Proteomics is only one part of the bigger picture, to understand the biology all components have to be known, their interactions, and their role at a certain state of a biological system. This includes the genes, the RNA, the metabolites, the glycans, the lipids, the proteins together with their specific modifications and interactions, the special distribution of these in the system; an almost endless list.

With the techniques in the early 1990s these questions could never be answered, but these questions led to a rapid development of new methods allowing comprehensive analysis on the biological relevant level. One major example is the still ongoing development of mass spectrometry, from the invention of electrospray,⁵ over static nano-electrospray⁶ to state-of-the-art nano-liquid chromatography coupled to high resolution mass spectrometers^{7,8} with the fully automatic fragmentation of peptides to identify the protein in databases. These developments allowed getting the first draft of the human Proteome, published in 2014.^{9,10} But this draft Proteome does not give comprehensive answers to biological questions. Interestingly it revealed some proteins which are not predicted from the genome and some predicted proteins have not been found yet, but it did not tell anything about the actual active form of the protein itself as the post-translational modifications¹¹ were not covered. Neil Kelleher predicted over 1 Billion active protein forms, called “Proteform”¹² in the human organism. For example only the Histone H3 modifications allow over 50,000 different Proteoforms; few of these have been seen by today; many may not exist at all. But the published draft Proteome opens new avenues to look into biology and to draw knowledge which leads not only to deep understanding but may also fulfil promises given at the publication of the human Genome and at the rise of Proteomics. Big Pharma companies jumped on the bandwagon of Genomics and

Volume 1 Issue 2 - 2014

Manfred Raida

Life Sciences Institute, National University of Singapore, Singapore

Correspondence: Manfred Raida, Life Sciences Institute, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077, Tel +6596387692, Email Manfred_raida@proteomics.asia

Received: June 22, 2014 | **Published:** June 23, 2014

later Proteomics, soon recognizing that it does not result in new blockbuster drugs and does not explain the development of diseases in a way allowing to fight them. End 1990s many big Pharma companies invested in Proteomics facilities, closing most of them down in early 2000s. Old school drug development was continued, with a single protein as a target and 100,000s to millions of small molecules to test them in-vitro as mainly inhibitors. For a while the drug companies complained to have too many targets to follow, suddenly they run out of targets and only very few new drugs got approved. Many drug developments also failed due to safety reasons as toxic side effects or even because of inefficiency. The approach “one target/one disease/one drug” has failed in most cases. Now big Pharma industry is rethinking and looking into the phenotypic approach, the effect of small molecules, peptides, antibodies, natural compounds, proteins, to name only a few, on a whole biological system, a cell, an animal model, an organ. This is the point where Proteomics comes back into the game. To understand the biological effect of a developmental drug the targets have to be identified in an unbiased manner and their role in the biological system has to be well understood. Chemical Proteomics^{13,14} allows this unbiased identification of the protein targets from complex biological systems, and also the off-targets and the reasons of toxicity. Interesting developments have been published in the last years and drawing attention from big Pharma to be included into their drug development pipelines. The dramatic teratogenic side effect of Thalidomide¹⁵ could be explained using Chemical Proteomics approaches; 50years after the drug was taken off the market.

Another promising, but generally failed field is the identification of relevant biomarkers for early diagnosis of diseases. While over 25,000 papers on protein biomarkers have been published, only very few made the way into the clinics. There are many reasons for this failure, as study design, cohort selection, number of individuals, the definition of the disease, the ethnicity, and the finally validation of the proposed biomarkers. Here a long used method made its way into Proteomics, the targeted mass spectrometry, also known as multiple or single reaction monitoring (MRM or SRM), or Western-Blot of mass spectrometry.¹⁶ MRM/SRM allows the detection and quantification of proteins in complex biological mixtures, as serum or urine, based on proteotypic tryptic peptides, representative for the targeted protein. With the now public available data from the human draft Proteome and from many other projects, these peptides can be selected for almost all human proteins.¹⁷ With state-of-the-art instruments 100s and even 1,000s of proteins can be detected and quantified in a single run in complex samples, overcoming the limitations of antibodies.

Now proposed biomarkers can be validated in larger cohorts, first commercial companies offer these services already.

But there are still many open questions that require both, technical and computational bioinformatics developments.

From the over 600 known protein modifications only a few have been tackled in a larger scale, as phosphorylation.¹⁸ The challenge to analyse highly complex modifications, as glycosylation is not solved yet, many other modifications are not even tackled or showed to be inaccessible to current techniques. Here lies the big challenges for the future, understanding biology require understanding how the active proteins are modified and what the modification affects.

The other big challenges are the overwhelming amount of already available information. A lot of hidden knowledge lies here, difficult to excavate and turn it into actions. The data and information are spread over many databases, from literature to sequences. To combine these into an easy searchable manner will help researchers to plan meaningful experiments leading to new knowledge in understanding biology and finally to new insights into diseases.

Also the identification of proteins from mass spectrometric data, especially MS/MS data, requires new approaches. The commonly used MOWSE¹⁹ has shown many drawbacks; mutations, multiple modifications may get not detected using this approach. New approaches, as de-novo sequencing with consecutive database searches may overcome the limitations.²⁰

After a downtime of Proteomics in the early 2000s, due to unfulfilled promises, it is back to a new and promising rise. Many mistakes have been done, researches should learn out of them and avoid their repetition.

Marvin Vestal once stated (literally): “*Genomics took 20years, Proteomics will take 200years*” a very optimistic view.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Wilkins MR, Sanchez JC, Gooley AA, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev.* 1996;13:19–50.
2. Anderson NG. Adventures in clinical chemistry and proteomics: a personal account. *Clin chem.* 2010;56(2):154–160.
3. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 2002;1(11):845–867.
4. Pennisi E. Genomics. ENCODE project writes eulogy for junk DNA. *Science.* 2012;337(6099):1159.
5. Fenn JB, Mann M, Meng CK, et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989;246(4926):64–71.
6. Wilm M, Shevchenko A, Houthaeve T, et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature.* 1996;379(6564):466–469.
7. Oosterkamp AJ, Gelpi E, Abian J. Quantitative peptide bioanalysis using column-switching nano liquid chromatography/mass spectrometry. *J Mass Spectrom.* 1998;33(10):976–983.
8. Chervet JP, Ursem M, Salzman JP. Instrumental requirements for nanoscale liquid chromatography. *Anal Chem.* 1996;68(9):1507–1512.
9. Wilhelm M, Schlegel J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. *Nature.* 2014;509(7502):582–587.
10. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. *Nature.* 2014;509(7502):575–581.
11. Larsen MR, Roepstorff P. Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. *Fresenius J Anal Chem.* 2000;366(6–7):677–690.
12. Smith LM, Kelleher NL; Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nat Methods.* 2013;10(3):186–187.
13. Raida M. Drug target deconvolution by chemical proteomics. *Curr Opin Chem Biol.* 2011;15(4):570–575.
14. Beroza P, Villar HO, Wick MM, et al. Chemoproteomics as a basis for post-genomic drug discovery. *Drug Discov Today.* 2002;7(15):807–814.
15. Ito T, Ando H, Suzuki T, et al. Identification of a primary target of thalidomide teratogenicity. *Science.* 2010;327(5971):1345–1350.
16. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics.* 2006;5(4):573–588.
17. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26(7):966–968.
18. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics.* 2013;12(12):3444–3452.
19. Pappin DJ, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol.* 1993;3(6):327–332.
20. Zhang J, Xin L, Shan B, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics.* 2012;11(4):M111.010587.