Future Challenges in Proteomics

Hartmut Schlüter

[1] Core Facility Protein Analysis, Charité – University Medicine Berlin
Tucholskystr. 2, 10115 Berlin, Germany
Hartmut.Schlueter@charite.de

Abstract. The benefits of the present proteomic approaches for the life science community are limited by 3 major problems. 1. The identification of proteins, regardless whether the peptide mass fingerprint or the shot-gun approach is used, usually is based on an incomplete set of peptides leaving large parts of the full amino acid sequence and other structural details of the original protein species in the dark; 2. Missing validation of many of the identified proteins by orthogonal biological experiments thus risking false positive results; 3. Missing standardization in nomenclature. In sum these problems may hinder progress in life science projects employing proteomic strategies and may be especially risky for system biology approaches since the ambiguities resulting from the above mentioned problems may cause wrong models. It is recommended to guide future proteome analytical investigations by a hypothesis and to focus to a smaller number of proteins which should ideally be analyzed in detail covering 100 % sequence coverage as well as all posttranslational modifications and will allow validation by additional biological experiments.

1 Introduction

In the past 10 years an impressive increase in publications about proteome analytical investigations can be recognized. However there are many analytical problems which are yet not really well solved such as the quantification problem or the resolution of low abundant proteins. Beside the individual restrictions, which accompany every currently available proteomic method, there are 3 major obstacles, which will be discussed here.

2 Discussion

2.1 The identification problem

The first problem concerns the protein identification. Independently from the approach, either peptide mass fingerprint (PMF) or sequencing of peptides of a digest

Dagstuhl Seminar Proceedings 05471
Computational Proteomics
http://drops.dagstuhl.de/opus/volltexte/2006/544

by MS/MS experiments, the identification of each individual protein is based on the mass spectrometric data derived from only a few of its peptides, which cover only a part of the total sequence of the protein. As a result splicing variants, truncations or mutations can not be identified. In addition in many cases no or only minor information about posttranslational modifications are yielded. Are these protein species relevant in number? Answers can be found in many publications presenting the results of 2-dimensional electrophoresis (2DE) such as the article from Klose et al. [1]. In this publication 24 protein species of the γ-enolase-2 and 52 species of HSP-70 were identified in mouse brain tissue. Are these protein species biologically relevant? Reviews about HSP-70 [2, 3] show, that this protein is involved in many different cell physiological processes. E.g. the extent of its in vitro phosphorylation was found to be significantly reduced by heat stress [4], thus demonstrating, that the phosphorylation state is associated with environmental conditions. Not only phosphorylation or other posttranslational modifications (PTMs) but also protein truncations play an important role in the regulation of protein function, as we know from protease-activated receptors (PAR) [5]. However, none of the common proteomic approaches clarifies the question, by which structural details individual protein species are characterized. Furthermore, with those proteomic approaches, which start with enzymatic digestions (shot-gun approach), followed by chromatographic separation prior to MS/MS analysis, it is nearly impossible to answer the question, which protein species of a protein were present in a sample. Quantification of proteins using the shot-gun approach may even yield results which are completely wrong since a defined peptide fragment detected by MS may originate from different protein species with identical amino acid sequence, which most probably differ in their concentration as well as in their functions. For example the activity of many enzymes can be switched on and off by phosphorylation and depshosphorylation. Both protein species in this case differ only by one phosphate group, which is bound covalently to an amino acid of the phosphorylated species. After the digestion of both protein species, the origin of the peptides can not be distinguished anymore with the exception of the phosphorylated peptide. As a result it is impossible to answer the biologically important question about the ratio of the active and the inactive enzyme.

Thus, protein analytical methods are urgently needed, which allow a comprehensive analysis of the total structure of protein species and their quantification. A new promising strategy towards these requirements is the top-down mass spectrometric approach, using the Fourier transform ion cyclotron resonance (FT-ICR)-MS technology, although this approach for the analysis of intact proteins is still in its infancy. In the future top-down proteomics may enable the study of intact proteins including their PTMs. Nevertheless, further progress in the FT-ICR technology is necessary to be able to analyze larger proteins. Furthermore advances in sample preparation and separation of intact proteins and data processing are needed to realize the potential of FT-ICR-MS for proteome analyses.

2.2 The validation problem

The second problem results from the large number of identified proteins, especially in protein mapping and other high-throughput studies. In these cases experimental verification of the identity of each protein usually is not possible, since the necessary biological experiments are much more time consuming than the automated protein identification procedure. Here, it may be beneficial to link the proteomic study to a hypothesis, which will reduce the number of proteins of interest from hundreds to a few. These proteins of interest after their identification can then be investigated by other techniques in more detail to validate their biological relevance in correspondence to their function. This procedure will also help to reduce the number of wrongly identified proteins.

2.3 The problem of nomenclature

The third problem is given by the protein-nomenclature, which is not standardized yet. Still, many different synonyms are used for individual proteins, which are transcripts of one single gene. The protein-disulfid-isomerase A3 is a representative example. The SwissProt accession number is P30101. SwissProt lists the synonyms EC 5.3.4.1, disulfide isomerase ER-60, ERp60, 58 kDa microsomal protein, p58, ERp57 and 58 kDa glucose regulated protein. A SwissProt search with "synonym" EC 5.3.4.1 finds a whole family of protein-disulfide-isomerases with the accession numbers P07237, Q13087, P30101, P13667, Q14554, Q15084 and Q96JJ7. A search with P30101 in the protein data base of the NCBI [6] results in 1 hit. In contrast a search with p58 results in 330 hits, thus demonstrating that this name is used for many different proteins with completely different amino acid sequences. Searching PubMed for "proteome" and ER60 yields 3 hits [7 - 9], which all of them are different from those 3 hits using "proteome" and "ERp57" [10 - 12]. Some authors obviously recognized this problem and therefore introduced a term consisting of 2 synonyms (ER60/ERp57) [13] or even a name formed by 4 synonyms: "glucose-regulated protein 58, GRP58/ERp57/ER-60" [14]. The missing standardization of the nomenclature of proteins may result in inappropriate conclusions concerning the relationship of function and identity of proteins and in the worst case may generate incorrect models of biochemical mechanisms, especially if data-base oriented modeling is performed.

This problem will be solved only if all data bases and all journals will follow the same standardized rules for the nomenclature of proteins, which allows an unambiguous identification of the individual protein species including and addressing its structural features. It is suggested that additionally to the already existing names the accession numbers should be used, with terms added, which characterize the structural properties of the protein species.

References

1. Klose, J., Nock, C., Herrmann, M., Stuhler, K., Marcus, K., Bluggel, M., Krause, E., Schalkwyk, L.C., Rastan, S., Brown, S.D., Bussow, K., Himmelbauer, H., Lehrach, H.: Genetic analysis of the mouse brain proteome. Nat. Genet. 30 (2002) 385-93
2. Mayer, M.P., Bukau, B.: Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol. Life Sci. 62 (2005) 670-84
3. Morishima, N.: Control of cell fate by Hsp70: more than an evanescent meeting. J. Biochem (Tokyo). 137 (2005) 449-53
4. Cvoro, A., Dundjerski, J., Trajkovic, D., Matic, G.: The level and phosphorylation of Hsp70 in the rat liver cytosol after adrenalectomy and hyperthermia. Cell Biol. Int. 23 (1999) 313-20
5. Coughlin, S.R.: Protease-activated receptors in hemostasis, thrombosis and vascular biology. J. Thromb. Haemost. 3 (2005) 1800-14
6. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=protein
7. Piec, I., Listrat, A., Alliot, J., Chambon, C., Taylor, R.G., Bechet, D.: Differential proteome analysis of aging in rat skeletal muscle. FASEB J. 19 (2005) 1143-5
8. Stierum, R., Gaspari, M., Dommels, Y., Ouatas, T., Pluk, H., Jespersen, S., Vogels, J., Verhoeckx, K., Groten, J., van Ommen B.: Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2. Biochim Biophys Acta. 1650 (2003) 73-91
9. Boraldi, F., Bini, L., Liberatori, S., Armini, A., Pallini, V., Tiozzo, R., Pasquali-Ronchetti, I., Quaglino, D.: Proteome analysis of dermal fibroblasts cultured in vitro from human healthy subjects of different ages. Proteomics. 3 (2003) 917-29
10. Blake, C.A., Brown, L.M., Duncan, M.W., Hunsucker, S.W., Helmke, S.M.: Estrogen regulation of the rat anterior pituitary gland proteome. Exp Biol Med (Maywood). 230 (2005) 800-7
11. Baik, J.Y., Lee, M.S., An, S.R., Yoon, S.K., Joo, E.J., Kim, Y.H., Park, H.W., Lee, G.M.: Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. Biotechnol. Bioeng. 93 (2006) 361-71
12. Giorda, E., Sibilio, L., Martayan, A., Moretti, S., Venturo, I., Mottolese, M., Ferrara, G.B., Cappellacci, S., Eibenschutz, L., Catricala, C., Grammatico, P., Giacomini, P.: The antigen processing machinery of class I human leukocyte antigens: linked patterns of gene expression in neoplastic cells. Cancer Res. 63 (2003) 4119-27
13. Lindquist, J.A., Hammerling, G.J., Trowsdale, J.: ER60/ERp57 forms disulfide-bonded intermediates with MHC class I heavy chain. FASEB J. 15 (2001) 1448-50
14. Kita, K., Okumura, N., Takao, T., Watanabe, M., Matsubara, T., Nishimura, O., Nagai, K.: Evidence for phosphorylation of rat liver glucose-regulated protein 58, GRP58/ERp57/ER-60, induced by fasting and leptin. FEBS Lett. 580 (2006) 199-205

Springer Auslieferungsgesellschaft
Haberstrasse 7
69126 Heidelberg
Germany
Fax: +49 6221 345-229
Phone: +49 6221 345-0

Springer-Verlag New York Inc.
P.O. Box 2485
Secaucus, NJ 07096-2485
USA
Fax: +1 201 348 4505
Phone: +1-800-SPRINGER
(+1 800 777 4643), toll-free in USA

Preferential orders can also be placed by sending an email to
orders@springer.de or orders@springer-ny.com.
For information about shipping, please contact one of the above mentioned orders departments. Sales tax is required for residents of: CA, IL, MA, MO, NJ, NY, PA, TX, VA, and VT. Canadian residents please add 7% GST. Payment for the book(s) plus shipping charges can be made by giving a credit card number together with the expiration date (American Express, Eurocard/Mastercard, Discover, and Visa are accepted) or by enclosing a check (mail orders only).