

Towards the analysis of protein species: an overview about strategies and methods

Peter R. Jungblut · Hartmut Schlüter

Received: 30 November 2010 / Accepted: 21 December 2010 / Published online: 18 January 2011
© Springer-Verlag 2011

Abstract The deciphering of the relationship between function and exact chemical composition of a defined protein species in the context of the proteome is one of the major challenges in proteomics and molecular cell physiology. In the Special Issue of Amino Acids about the analysis of protein species current approaches are reviewed and new methods described focusing on the investigation of protein species. On the basis of the articles in this Special Issue it can be summarized that first important and promising steps towards the comprehensive analysis of protein species have been done. It is already possible to obtain full (100%) sequence coverage of proteins by mass spectrometry, if the amount of proteins available for their analysis allows their proteolytic degradation by more than one protease and the subsequent mass spectrometric analysis of the resulting peptides. Employing affinity chromatography helps to analyse proteins with defined post-translational modifications thus opening a targeted view on e.g. the phosphoproteome. In the future the aim to identify the exact chemical composition including not one but every posttranslational modification and complete sequence coverage on the protein species level should be achievable with further progress in sample preparation techniques, especially concerning separation techniques on the protein level, mass spectrometry and algorithms for

mass spectrometric data processing. For determining the function of defined protein species a closer cooperation between cell biologists and proteomics experts is desirable.

The term protein species was introduced by Jungblut et al. (1996) for the description of the smallest unit in the proteome (Jungblut et al. 2008; Schlüter et al. 2009) and for avoiding an ambiguity concerning the term isoforms, which is already claimed for the description of genetic variations such as allelic forms according to the IUPAC rules (Joint Commission on Biochemical Nomenclature IUPAC-IUBMB 1992). Why is the term protein species so important for the unravelling of the molecular mechanisms in the life sciences? This Special Issue is intended to give some answers and also to raise new questions to improve the access to the molecules, which decide about so many different processes in life: the protein species.

Whereas the genome predestines the structure and the function of an organism, the molecular phenotype at a certain time in the lifecycle of an organism is better described by the proteome and the metabolome. Pursuing the deciphering of the genomes of hundreds of microorganisms, plants and animals snapshots of proteomes of many biological situations were taken in the last 20 years. Most of these investigations were focused on the finding of proteins, which show changes in abundance between two or more biological situations such as disease and control. The importance of post-translational modifications was getting obvious long before proteomics raised as a new discipline of biochemical analysis. By classical biochemical approaches meanwhile numerous results demonstrate the significance of posttranslational modifications towards the determination of the function of a protein. An excellent example is given in the review of Deribe et al. (2010),

P. R. Jungblut (✉)
Core Facility Protein Analysis, Max-Planck-Institute for
Infection Biology, Charitéplatz 1, 10117 Berlin, Germany
e-mail: jungblut@mpiib-berlin.mpg.de

H. Schlüter (✉)
Institute of Clinical Chemistry - Mass Spectrometric Proteomics,
University Medicine Hamburg-Eppendorf, Martinistr. 52,
20246 Hamburg, Germany
e-mail: hschluet@uke.de

which focuses on the impact of polyubiquitination on cell signalling. Since many years the importance of post-translational modifications was stressed by biochemists and the detection of the modification sites such as phosphorylation was improved to end up in the detection of 10,000 phosphorylation sites in the phosphoproteome of *Drosophila Schneider* cells (Hilger et al. 2009). The editors of this Special Issue formulated the protein species concept in 2008 and 2009 (Jungblut et al. 2008; Schlüter et al. 2009) and worked out that the combination of posttranslational modifications on a certain protein distinguishes about the function within an organism. If functional aspects have to be elucidated, it is necessary to investigate the proteome at the protein species level. Now 2 years later the Editors invited 15 authors to write a review or to present their original investigations under the aspect of the protein species concept.

One of the major prerequisites to reach the protein species level, 100% sequence coverage, was picked out as a central theme by Björn Meyer et al., who agreed that it is necessary to reach the protein species level, but because of the complexity this seems to be rather “surreal”. Nevertheless, they review solutions such as top-down or middle-down approaches, which separate in the first step the protein species and in the second step the more or less purified protein species are analysed. It will be interesting to read this review 10 years later and if the present surrealism will change to realism. In the review of Wolfgang Hoehenwarter et al. the surrealism seems to become reality already today: Protein species are functionally analysed by shotgun proteomics combined with linear mathematics. Is it really possible to get information about a protein species from a peptide (protein sequence coverage down to 1%)? Membrane proteins are especially difficult to analyse. Jacek Wisniewski reviewed the state of the art of phospho- and glycoproteomics of plasma membranes. Today with LC-MS methods it is possible to gain access to many thousand membrane proteins and even their post-translational modifications, but not at the protein species level. The modified peptides can be accessed but not the protein species with their distinct combination of post-translational modifications.

Three reviews comprise information on defined posttranslational modifications: chondroitin and dermatan sulphate glycosaminoglycans, ADP-ribosylation and metalloproteases. The first and the second review are typical examples for the topic of different classes of building blocks, which are used for the classification of proteins. Eugen Sisu et al. focus in depth on the mass spectrometry based analysis of chondroitin sulphate and dermatan sulphate, which originally were bound to proteins. The methods, which are described in the review, are especially useful for analysing the composition of glycosaminoglycan after removal of the oligosaccharides from the core

proteins. Therefore, an assignment of the oligosaccharide composition towards a single protein species is possible only if the protein species has been purified prior to the analysis of the composition of its glycosaminoglycan. High-through-put analysis applying the described strategies on protein mixtures will yield data about the glycosaminoglycan status of the protein mixture, which may be characteristic for the development of diseases. The review of Sabrina Laing et al. about arginine-specific ADP-ribosylation gives a holistic and impressive overview about its physiological impact including the enzymes responsible for this modification and its reversal as well as consequences in protein functions. Furthermore the authors discuss methods for the detection of arginine-specific ADP-ribosylation. The topic ADP-ribosylation of proteins displays the current discrepancies between protein biochemistry research focusing on the functions of proteins and the field of proteomics, which claims the aim to investigate the complement of the genome and thereby giving the genes functions. Although the importance of ADP-ribosylation of proteins for living organisms is fully established today as described in the review of Laing et al. this topic is more or less ignored by the proteomics research community, which becomes obvious with a view on PubMed. A search with the term “ADP-ribosylation” combined with “proteomics” yielded only 31 hits. In comparison, the term “phosphorylation” combined with “proteomics” resulted in 1,611 hits. This also clearly shows that proteomics still is far from the aim to give a comprehensive functional status of an investigated proteome. This problem also becomes obvious by reading the review of Theo Klein and Rainer Bischoff. The authors gave insight into a facet of the world of proteases—the matrix metalloproteases. They clearly and detailed summarize the physiological and pathophysiological importance of these proteases thus emphasizing the importance of the protein species. The numerous different proteases, which are present e.g. in humans, have an enormous impact on the regulation of the functions of proteins only by shortening their target proteins. Again many of the current proteomic approaches are neglecting the question of the exact amino acid sequence of proteins thereby not being able to give information about the functional status of individual proteins.

Jörg Seidler et al. address in their study the problem of poor detection efficiency of phosphopeptides in LC-MS. Although this field has already been investigated by many other colleagues, the results of their study show that improvement of the analytical methods for analysing phosphopeptides is still possible and necessary. Therefore, this review should encourage everyone continuing method development.

Alessandro Cuomo et al. report about the investigation of PTMs on histone H3 and H4 from four human breast

cancer cell lines in comparison with normal epithelial breast cells. By applying the stable isotope labelling with amino acids in cell culture (SILAC) to quantify the modification changes in cancer cells the authors observed a “breast cancer-specific epigenetic signature”. This study convincingly shows that today it is possible to perform an “omics” approach not only on the level of tryptic peptides but also on the level of defined posttranslational modifications, here lysine acetylation and methylation on fourteen distinct sites in H3 and H4. Of course this approach is much focused and does not include all aspects of the protein species concept especially according to the total exact chemical composition. However, it is the right step forward towards the analysis of protein species.

The manuscript of Zsuzsanna Darula et al. describes the re-interrogation of electron-transfer dissociation (ETD) MS/MS data of a complex mixture of O-linked glycopeptides with newly developed analysis tools. The re-interrogation approach—processing and interpretation of “old” mass spectrometric data, with new algorithms—caused some important discussions between the Editors since one reviewer suggested to reject this manuscript with the argument that the old data (processed with an “older” software) were already published elsewhere. In contrast to the reviewer our opinion is that re-interrogation of “old” mass spectrometric data should be supported, since it opens the chance to yield better and more results from the original mass spectrometric experiment. Especially with respect to the protein species topic a re-interrogation of mass spectrometric data is recommendable, because the actual sequence databases are incomplete concerning posttranslational modifications and the currently available software is not yet really able to account every possible posttranslational modification, but in the future the software will improve in this direction as the manuscript from Zsuzsanna Darula et al. shows. Since proteomics investigations are expensive, the re-interrogation of older mass spectrometric data is additionally cost saving and allows the sustainable use of these data. Furthermore, because the complexity of proteomes is extremely huge, as much as possible information should be collected in www-accessible databases from each individual experiment.

Stefanie Henning et al. describe a method for the direct analysis of alpha- and beta-chains of haemoglobins from diluted blood samples by simultaneous in-capillary proteolytic digestion and nano-ESI-MS and MS/MS. With their newly developed method the authors achieved high sequence coverage for both haemoglobin chains and therefore this remarkable study is also a step forward towards the comprehensive analysis of protein species.

Eight original articles analysed proteomes of cells, blood and tissues. All of the contributions using 2-DE recognized the occurrence of protein species. The 100%

sequence coverage was not reached in any of these investigations. Only Frank Schmidt et al. describe discriminative post-translational modifications between two protein species. One of two spots of the proteasomal protein alpha-7 was phosphorylated and one of its cysteines was oxidized to cysteine sulfonic acid, whereas the other protein species was unmodified. These two proteins were differentially regulated in apoptosis. A clear differentiation into protein species was also recognized for superoxide dismutase type 1 by Alexandra Scholze et al. One of the protein species was down-regulated in the controls in the comparison with haemodialysis and chronic kidney disease patients. All of the other protein species were down-regulated in the disease samples, in spite of upregulation of the superoxide dismutase type 1 gene at the transcriptomics level, again showing that a correlation between transcriptomics and proteomics cannot be expected per se and the necessity to analyse the proteome at the protein species level. Evelyn Zeindl-Eberhart et al. started with a proteomic investigation on papillary thyroid carcinomas and found protein speciation for several cytoskeletal proteins and differential regulation of different protein species of the same protein during carcinoma development. Karima Schwab et al. elucidated several proteins occurring with several protein species in a 2-DE proteomic analysis on the influence of the phytohormone genistein on the normal murine heart proteome. By multiprotease digestion and FTICR-MS identification it was possible to find some candidates for posttranslational modifications, but 100% sequence coverage was not reached. The study of Frank Kozielski et al. shows the high resolution power of LC-MS. Microtubule-associated proteins and their interaction partners from mammalian brain were resolved into 573 proteins. They were able to reveal many post-translational modifications in the analysed peptides, predicting many different protein species not accessible by this approach.

Even after 6 years, the investigation of Okkels et al. (2004) remains one of the rare examples where with the help of proteomics one protein (ESAT-6) was separated into eight protein species and four of them were analysed with 100% sequence coverage, allowing a real protein species differentiation and a correlation to different functions. Only the non-acetylated protein species were able to interact with CFP-10. This interaction is important for the transfer of ESAT-6 out of the bacterial cell. It may be speculated that in the secretome, which was analysed, the rebinding of ESAT-6 to CFP-10 should be avoided and, therefore, ESAT-6 is acetylated to fulfil its task without competition with the interaction to CFP-10.

The articles of this Special Issue give a representative overview about the state of the art of protein science focusing on the topic of protein species and show clearly the speciation for many proteins and in some cases the

differential behaviour of the protein species of one protein concerning different biological situations. The consequences of these observations should be: (1) To be careful with the results from shotgun proteomic investigations concerning correlation of identified proteins from peptides with their function. (2) To develop new strategies and tools in the next years for the investigation of the proteome on the protein species level as well as to establish a higher degree of collaboration between cell biologists and proteomics specialists, although proteomics has reached high-throughput for the analysis of proteins and the complete access to proteomes at the protein level was already claimed (Nilsson et al. 2010). The analysis at the protein species level, which is necessary to reach the functional level of proteomics, is still in its infancy. (3) To add entries for each protein species in protein sequence databases allowing to identify proteins on the protein species level with search engines. Up to now identification of proteins by mass spectrometry is based on searches in protein sequence databases, which have entries for each protein. In these entries posttranslational modifications are annotated. After accepting the importance of the combination of posttranslational modifications for the function of a polypeptide, it is necessary to identify the proteins at the protein species level and this means that it cannot be avoided to adapt the protein sequence database to the requirements of an unambiguous description of the protein species. (4) To come to an agreement concerning the unambiguous description of protein species as suggested in Schlüter et al. (2009).

The Editors thank the authors for their excellent contributions, the reviewers for their engagement, and

Springer Wien New York for their support to produce this Special Issue on Protein Species. Hopefully this Special Issue will contribute to improve the understanding of health and disease at the molecular level and will contribute to find more effective diagnostically and therapeutically relevant biomarkers in the future.

References

- Deribe YL, Pawson T, Dikic I (2010) Post-translational modifications in signal integration. *Nat Struct Mol Biol* 17(6):666–672
- Hilger M, Bonaldi T, Gnad F, Mann M (2009) Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Mol Cell Proteomics* 8:1908–1920
- Joint Commission on Biochemical Nomenclature IUPAC-IUBMB (1992) Nomenclature of multiple forms of enzymes. In: Colchester LC (ed) *Biochemical nomenclature and related documents*, 2nd edn. Portland Press, London
- Jungblut P, Thiede B, Zimny-Armdt U, Muller EC, Scheler C, Wittmann-Liebold B, Otto A (1996) Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* 17:839–847
- Jungblut PR, Holzhütter HG, Apweiler R, Schlüter H (2008) The speciation of the proteome. *Chem Cent J* 18(2):16
- Nilsson T, Mann M, Aebersold R, Yates J III, Bairoch A, Bergeron JJM (2010) Mass spectrometry in high-throughput proteomics: ready for the big time. *Nature Methods* 7:681–685
- Okkels LM, Müller EC, Schmid M, Rosenkrands I, Kaufmann SHE, Andersen P, Jungblut PR (2004) CFP10 discriminates between nonacetylated and acetylated ESAT-6 of *Mycobacterium tuberculosis* by differential interaction. *Proteomics* 4:2954–2960
- Schlüter H, Apweiler R, Holzhütter HG, Jungblut PR (2009) Finding one's way in proteomics: a protein species nomenclature. *Chem Cent J* 9(3):11