Journal of Proteomics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Review

# Protein species as diagnostic markers

Pascal Steffen [a], Marcel Kwiatkowski [a], Wesley D. Robertson [b], Arash Zarrine-Afsar [c,d], Diana Deterra [a], Verena Richter [a], Hartmut Schlüter [a,*]

[a] University Medical Centre Hamburg-Eppendorf, Institute for Clinical Chemistry, Department for Mass Spectrometric Proteomics, Martinistraße 52, 20246 Hamburg, Germany
[b] Max Planck Institute for the Structure and Dynamics of Matter, Atomically Resolved Dynamics Division, Luruper Chaussee 149, 22761, Hamburg, Germany
[c] Techna Institute for the Advancement of Technology for Health, University Health Network, Toronto, ON M5G-1P5, Canada
[d] Department of Medical Biophysics, University of Toronto, 101 College Street Suite 15-701, Toronto, ON M5G 1L7, Canada

ARTICLE INFO

*Article history:*
Received 4 August 2015
Received in revised form 28 November 2015
Accepted 9 December 2015
Available online xxxx

*Keywords:*
Mass spectrometry
Protein species
Clinical chemistry
Protein diagnostics

ABSTRACT

Many diseases are associated with protein species perturbations. A prominent example of an established diagnostic marker is the glycated protein species of hemoglobin, termed HbA1c. HbA1c concentration is increased in the blood of diabetes mellitus patients due to their poor control of blood glucose levels resulting in an increased non-enzymatic glycosylation of hemoglobin producing HbA1c. This important diagnostic marker is routinely measured in the blood of diabetes patients. As in the case of HbA1c, protein species can mirror pathophysiological events. Shifts in the levels of protein species can be associated with or even be responsible for disease making them well suited as diagnostic markers. However, only a few protein species are currently used as diagnostic markers in routine clinical chemistry laboratories, despite being widely established in clinical proteomics research. This review provides an overview of the biochemical characteristics associated with protein species as well as examples of pathophysiological mechanisms, which cause modifications in the protein species composition, thereby emphasizing the importance of screening for protein markers at the species level. Further, we highlight techniques, which are currently utilized for investigating protein species markers in clinical research.
*Biological significance:* The success rate of FDA approved diagnostic protein markers until today is very low compared to the number of published candidate disease markers. It is hypothesized that one important reason is the gene-centric view which is still followed in clinical proteomics: In many investigations proteins are still digested in small peptides thus making it nearly impossible to discriminate between healthy proteins and pathologic proteins causing diseases. Thus this review is focusing on the biochemistry and patho-biochemistry of proteins, is highlighting the need for screening for disease markers on the protein species level and is giving an overview about available techniques.



**Contents**



* Corresponding author.
*E-mail addresses:* p.steffen@uke.de (P. Steffen), m.kwiatkowksi@uke.de (M. Kwiatkowski), wesley.robertson@mpsd.mpg.de (W.D. Robertson), arash.zarrine.afsar@utoronto.ca (A. Zarrine-Afsar), Diana.Deterra@gmx.de (D. Deterra), vrichter-bs@web.de (V. Richter), hschluet@uke.de (H. Schlüter).

http://dx.doi.org/10.1016/j.jprot.2015.12.015
1874-3919/

## 1. Introduction

Proteins have been used as diagnostic disease markers for more than 150 years [1]. Consequently, following the beginning of the "proteomics age" in the mid-90s, techniques for proteome analysis were also applied to screening for diagnostic markers. Although the success of the mass spectrometry based proteomics tools over the past 20 years is overwhelming, there is surprisingly a discrepancy with respect to the US food and drug administration (FDA) approval of methods for identifying diagnostic protein markers. According to a study from 2008, performed by Anderson et al., 109 protein target assays in plasma were approved by the FDA. In total, 205 clinical protein assays existed at that time including those which are not FDA approved but found in the test menus of major reference laboratories or referred to in DORA (declaration on research assessment) [2]. It is striking that 80% of these 109 FDA approved tests were developed before 1993 and since then only 22 additional tests have been introduced, giving a rate of 1.5 tests per year. This rate is especially surprising in the age of proteomics, which was "born" in 1995 following the creation of the term proteome [3] and which marked a steep increase in the number of publications since that time. The low success rate in the discovery of diagnostic protein markers emphasizes the need to scrutinize and improve the strategies for screening for protein markers and to push the development of new protein assay technologies with applicability in diagnostic routine. Many factors have been discussed which may be responsible for the low success rate of bringing new diagnostic protein markers into the clinic. Critical points include study design, patient specimen sampling, processing, storage, preparation, as well as the correct classification of the disease and the total number of patients within a collective [2,4,5].

## 2. Patho/biochemistry of protein species — enzyme catalyzed formation of protein species

Presumably the biochemistry of proteins itself and especially the occurrence of a multitude of protein species originating from a single gene is a major reason for the low success rate of translating marker candidates into approved markers, as described above. The term protein species was introduced in 1996 by Peter Jungblut [6] to avoid the ambiguity problem with the term isoform with respect to the IUPAC nomenclature. A protein species is the smallest unit of the proteome [7] and is defined by its structural formula comprising full amino acid sequence coverage and every posttranslational modification [8]. Smith et al. recently suggested the term "proteoform" as a synonym for the term "protein species" [9]. The authors have invented the term proteoform from a gene-centric viewpoint. In contrast the term protein species is associated with a chemo-centric viewpoint. Since in this review the non-enzymatic formation of new species by the reaction of the proceeding species with exogenous molecules is also included, we here use the term protein species.

From one single gene a large number of protein species can arise. E.g. histones occur in huge quantities from an individual protein species [10]. Since the discovery of the significance of phosphorylation with respect to the control of protein functions [11] more than 30 years ago, it has become more and more evident that the function of a protein critically depends on its structural formula [8]. Recently it was estimated that the human proteome comprises more than a billion protein species, also termed proteoforms [12]. The large number of different specialized cell types, which have their own repertoire of protein species adapted to their special tasks can explain this enormous number. A further reason is the long journey a protein perambulates, starting from the *1. gene*, *2. expression of its gene*, via *3. RNA processing* including possible *splicing*, *4. translation of messenger RNA* (*mRNA*), *5. protein synthesis*, *6. proteolytic processing*, which can occur several times within a lifespan of a gene product, *7. addition of* diverse *posttranslational modification*, *8. formation of complexes* with other protein species, until its *9. degradation*. During every conversion of a protein species a new protein species is formed. All of these steps in the lifetime of a gene product are strictly controlled, since enzymes catalyze almost all conversions, and are present with defined activities and concentrations. At every step, from gene expression to protein degradation, during the lifetime of a gene product and its many protein species, a perturbation is possible, which can change the quantity or quality of a converting enzyme and result in an inappropriate, in the worst case "pathologic" protein species, which is characterized by a dysfunctional protein.

Well known for many years are perturbations on the *1. gene* level, which are caused by mutation of a gene affecting a functionally important epitope in an enzyme or receptor and result in a monogenic disease ([13,14]). These mutations are usually associated with an exchange of a single amino acid – a very small change of the chemical composition in comparison to the total chemical composition of a protein – and a detrimental outcome for the affected individual. Dysregulation of *2. gene expression* is also often observed in cancer [15] and causes shifts in the quantitative composition of protein species resulting in significantly different protein patterns as commonly reported in cancer proteomics studies [16,17].

Beside mutations, which cause diseases and can today be easily detected by genetic testing, every individual has many mutations in his genome, which do not have a significant effect on protein function but can cause a problem for mass spectrometric analysis of diagnostic proteins since the exchange of an amino acid can shift the species molecular weight [18]. Nedlkov et al. investigated the protein diversity in human populations – termed population proteomics [19]. The authors pointed out that proteomic approaches require a paradigm shift to include subtle variations among individuals. These individual variations must also be considered regarding the development of new protein quantification assays for the clinical laboratory. The research institute for population proteomics investigated the extent of protein diversity in and across human populations. In a large study analyzing 1000 plasma samples from different individuals by an MS-immunoassay, different species of cystatin c, retinol binding protein, transferrin and transthyretin were analyzed by monitoring unique modifications such as oxidation, truncations, sulfonations and point mutations [18]. These studies provide the first impressions of the extent of protein species diversity that exists in human populations and it is critical that all of these species are mapped and their wild-type protein profiles determined.

On the *3. path from gene expression to the mature messenger RNA* (mRNA), failures in the expression and splicing machinery can result in new protein species with different amino acid sequences. Epigenetic information can also have an impact on protein species by affecting gene expression and pre-mRNA splicing. Transcriptional activity is

Please cite this article as: P. Steffen, et al., Protein species as diagnostic markers, J Prot (2015), http://dx.doi.org/10.1016/j.jprot.2015.12.015