

Commentary

Open Access

## Finding one's way in proteomics: a protein species nomenclature

Hartmut Schlüter\*<sup>1</sup>, Rolf Apweiler<sup>2</sup>, Hermann-Georg Holzhütter<sup>3</sup> and Peter R Jungblut\*<sup>4</sup>

Address: <sup>1</sup>University Medical Centre Hamburg-Eppendorf, Hamburg, Germany, <sup>2</sup>The EMBL Outstation-The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, Great Britain, <sup>3</sup>Charité Berlin, Institut für Biochemie, Berlin, Germany and <sup>4</sup>Max Planck Institute for Infection Biology, Berlin, Germany

Email: Hartmut Schlüter\* - [hschluet@uke.de](mailto:hschluet@uke.de); Rolf Apweiler - [apweiler@ebi.ac.uk](mailto:apweiler@ebi.ac.uk); Hermann-Georg Holzhütter - [hergo@charite.de](mailto:hergo@charite.de); Peter R Jungblut\* - [jungblut@mpiib-berlin.mpg.de](mailto:jungblut@mpiib-berlin.mpg.de)

\* Corresponding authors

Published: 9 September 2009

Received: 8 June 2009

*Chemistry Central Journal* 2009, 3:11 doi:10.1186/1752-153X-3-11

Accepted: 9 September 2009

This article is available from: <http://journal.chemistrycentral.com/content/3/1/11>

© 2009 Schlüter et al

### Abstract

Our knowledge of proteins has greatly improved in recent years, driven by new technologies in the fields of molecular biology and proteome research. It has become clear that from a single gene not only one single gene product but many different ones - termed protein species - are generated, all of which may be associated with different functions. Nonetheless, an unambiguous nomenclature for describing individual protein species is still lacking. With the present paper we therefore propose a systematic nomenclature for the comprehensive description of protein species. The protein species nomenclature is flexible and adaptable to every level of knowledge and of experimental data in accordance with the exact chemical composition of individual protein species. As a minimum description the entry name (gene name + species according to the UniProt knowledgebase) can be used, if no analytical data about the target protein species are available.

### Introduction

The number of publications in the field of proteomics has increased dramatically over the last decade. The driving force behind this development has been the hope of gaining additional insights into the functioning of a cell or of a complete organism by identification and quantification of proteins in different biological states such as disease and health, wild type and mutant, baseline and perturbed state, among others. The dynamics and the influence of post-translational protein modifications were mostly ignored in the course of development of the basic technologies during these years. The focus on technology has nonetheless resulted in dramatic improvements in mass spectrometry and in coupling MS with two-dimensional electrophoresis and liquid chromatography. The improvements and results gained over time in proteomics research have shown that the behaviour and variability of proteins are more complex than had ever been imagined. The fact

that the same protein was found at several different spots on 2-dimensional electrophoresis gels made it necessary to define a new term for these different forms of a single protein: protein species [1,2]. Each additional modification and each new combination of modifications represents an additional protein species of that single protein. Though the term protein species had been used earlier in the literature [3,4], it was not clearly defined and used more in the sense of a protein complex consisting of several subunits [3] or to differentiate between different proteins (e.g. catalase and actin were two different protein species) [4]. According to the IUPAC rules [5] the term "isoform" is to be used for genetic variations such as allelic forms. Therefore, it was necessary to find a term for any chemical modification and any combination of chemical modifications. The term "protein species" has been defined by Jungblut et al. at the chemical, molecular level [1,2]. According to this definition, isoforms repre-

sent different protein species, because they are also chemically different. In contrast, two proteins with different post-translational modifications represent different protein species but not different isoforms.

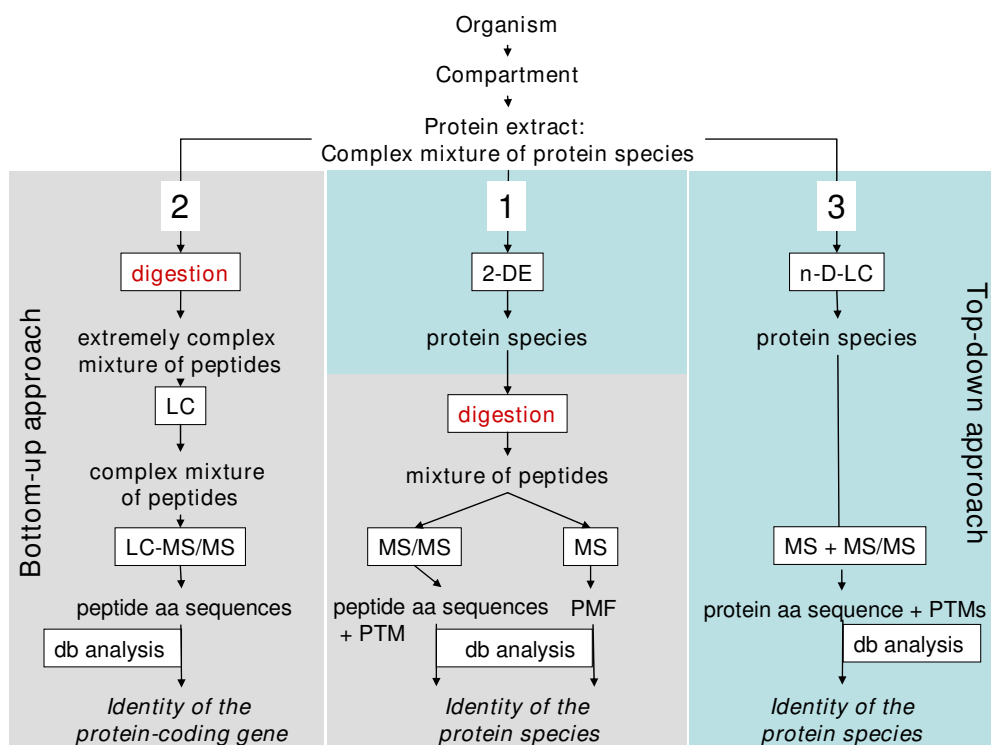
About 600 different post-translational modifications (PTMs) are included in the database UNIMOD (August, 2009, [6]), which was developed by Creasy and Cottrell [7]. Based on the results they obtained in a proteome analytical study employing high-resolution Fourier-Transform-Ion-Cyclotron mass spectrometry (FT-ICR-MS) Nielsen et al. concluded that the estimated level of 8-12 modified peptides per unmodified tryptic peptide present at >1% level approaches one modification per amino acid on average [8]. An example for the important relationship between the exact chemical composition of proteins including PTMs and their function is the polyubiquitylation of proteins. Lysin-48-linked poly-ubiquitin chains target proteins for proteasome-mediated proteolysis, whereas lysine-63-linked ubiquitin chains mediate various non-degradative functions, including the activation of signalling factors and protein trafficking [9]. To move the field forward, it has become necessary, therefore, to take the speciation of proteins and the kinetics of their protein species into account [2].

## Discussion

### Three main proteomic approaches

The classic strategy, the 2-DE/MS approach, starts with the separation of the proteins by two-dimensional electrophoresis [4] followed by enzymatic digestion and identification of the proteins by mass spectrometric analysis of the peptide digest (Figure 1, path 1). The comparison of the mass spectrometric data with sequence databases results in the identification of the protein. This approach has the advantage of separation of the proteins at the protein species level with a high resolution of up to 10,000 spots [10] and application of the identification procedure at the peptide level, where MS is very sensitive, fast and accurate [11].

The second strategy (path 2 in Figure 1) starts directly with the digestion of the proteins of a complex mixture. Low femtomole protein identifications in mixtures by online LC/MS/MS were first reported in 1997 [12]. The digestion yields a huge number of peptides, which are separated in the next step, typically via one-dimensional or multidimensional chromatographic methods. This procedure is a bottom-up approach because it starts on the level of separation with peptides. Peptides eluting from a reversed-phase column are then identified by mass spectrometry



**Figure 1**

**The three main strategies in proteomics.** 2-DE: two-dimensional electrophoresis; LC: liquid chromatography; n-D: n-dimensional; db: database; MS: mass spectrometry; PMF: peptide mass fingerprint; PTM: post-translational modification.

(MS), complementing MS with MS/MS experiments. The latter analysis yields amino acid sequences by which the original proteins can be identified, again by sequence database comparison. The bottom-up approach is fast and sensitive, but does not allow the differentiation between protein species.

A third strategy, a top-down approach (path 3 in Figure 1) starts with liquid chromatography for separation of the protein species followed by identification of the protein species by mass spectrometry [13]. This approach, however, has been largely limited to low mass proteins up to 30 kDa, although there has been a report of identification of a protein with a mass of more than 200 kDa [14]. With the present technology the technique does afford a high sample amount, protein fractions with a largely reduced complexity of composition, and high mass accuracies in the low ppm range. A top-down study in HeLa cells resulted in the identification of 45 protein species, containing polymorphisms, alternative splicing products and modifications [15].

In the bottom-up, top-down terminology the traditional 2-DE/MS approach represents a top-down separation with bottom-up identification. The critical step in the first two strategies is the protein digestion step, since usually not all peptides of the digest are detected by mass spectrometric analysis. During the separation steps peptides may be lost due to unspecific interactions with surfaces and chromatographic materials. Other peptides may also not be identified or cannot be used for the peptide mass fingerprinting, if they contain uncommon or complex post-translational modifications, or if they lie outside the optimal mass range between 500 and 3000 Da. As a result, the protein identification is in most cases based on a subset of peptides and does not cover 100% of the amino acid sequence of the analysed protein. A consequence of this is that RNA and protein splice variants, proteolytically processed protein species, and protein polymorphisms cannot be distinguished from each other. As a further consequence, protein species containing post-translational modifications may not be identified. This problem is clearly reduced with the 2-DE strategy [16], since the protein species are separated and detected in spots. All the peptides of one protein are within one mass spectrum, permitting analysis of the modifications of the predicted primary structure. Additionally, the sequence coverage can be increased by using different digestion procedures. In the pure bottom-up approach (strategy 2 in Figure 1) the peptides derived from one protein species are distributed over all fractions of the LC, which makes it impossible to assign them to their respective protein species. Since peptides with an identical amino acid sequence stemming from different protein species elute within a single peak,

quantification of the individual protein species is not possible using this method.

Recently, an integrated top-down and bottom-up strategy for broadly characterizing protein species has been developed to overcome the limitations of pure LC-MS strategies [17]. Quantification and resolution of low-abundance protein species still remain difficult problems for all of the proteomic approaches used today.

***How do gene polymorphisms, alternatively spliced transcripts, proteolytically processed protein species, and post-translational modifications affect the number of protein species derived from a single gene?***

Nucleotide polymorphisms, alternative splicing, and proteolytic cleavage, as well as post-translational modifications, are widespread and have an obvious effect on the number of protein species that can derive from a single gene. There are no systematic data available on the distribution of protein species derived from all genes in a single organism, but evidence for the widespread occurrence of multiple protein species per protein can be found in many publications that present the results of 2-dimensional electrophoresis (2-DE). One such report is by Scheler et al. [18], where the authors presented a 2-DE pattern showing 59 spots of the Hsp27 protein. In another article, Klose et al. [19] identified 24 protein spots encoded by the  $\gamma$ -enolase-2 gene and 52 protein spots of the HSP-70 gene in mouse brain tissue. Even in microorganisms, many different protein species, particularly of the heat-shock proteins such as HspX and GroES, have been found [20]. Better known is the speciation in the case of histones and a special histone code [21] has been postulated connecting certain histone species with defined functions. Twenty modifications and all of their combinations result in a total of more than 3 million protein species alone for histone H4 [22]. The biological function of most of these species is completely uncertain at present, but an extension of the histone code to other proteins has already been discussed, and designated as the "protein code" [23].

***How do polymorphisms, alternatively spliced transcripts, proteolytically processed protein species, and posttranslational modifications affect the protein function?***

Many examples of different protein species derived from an initial synthesis product having different molecular roles are now known. Reviews focusing on HSP-70 - the example mentioned above - illustrate the involvement of HSP-70 in many different cellular processes [24,25]. It can be assumed that different protein species are responsible for different cellular tasks. For the last twenty to thirty years it has been well known that, for example, enzymatic activities can be switched on and off by phosphorylation/dephosphorylation processes. Protease-activated recep-

tors (PAR), as well as many proteases, are activated by the cleavage of a peptide from the protein by specific proteases. The ESAT-6 gene product of *Mycobacterium tuberculosis* differentiates into at least 8 protein species [26]. In this investigation, four of the ESAT-6 protein species were acetylated at the N-terminus and were not able to interact with CFP10, an interaction necessary for the transfer of both proteins out of the bacterial cell. This observation of CFP 10 binding and CFP 10 non-binding protein species in the extra-cellular protein fraction raises new functional questions.

The biological significance of the protein species concept [1,2], which defines protein species chemically [2], is clearer if we take a look at our current knowledge of gene expression and the flow of information from genes to proteins (Figure 2). In recent years it has become obvious that the original paradigm of one gene encoding one protein is not correct and must be replaced by a new one based on the following facts:

1. At the end, one gene encodes many proteins - more exactly, protein species - which are usually strongly inter-related based on their amino acid sequence, but which can, nonetheless, differ sometimes quite dramatically in chemical structure as a result of alternative splicing at the

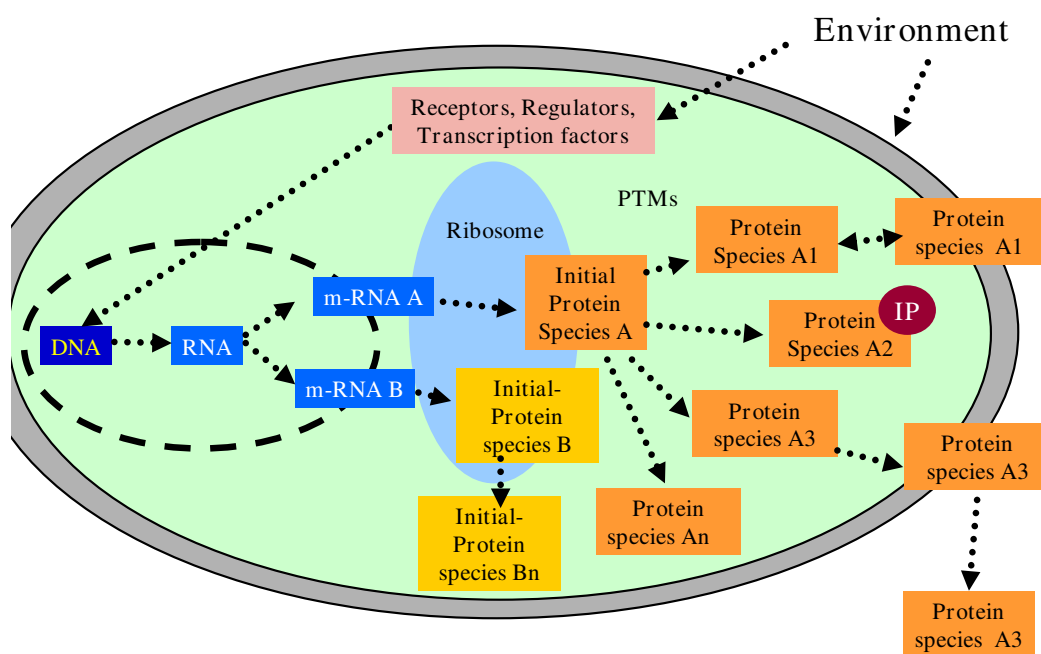
RNA and protein level [27] and/or due to post-translational modifications.

2. A protein species with a defined function is not only the product of one gene but of many genes, since many other products of completely different genes are involved in processing the chemical structure of the mature protein species.

3. Protein species are also modified by their environment, e.g. by physical factors such as temperature or light, or chemically by interactions with other molecules.

4. Furthermore, protein species interact with other proteins encoded by their own genes, as well as with small molecules. Figure 2 summarizes the new paradigm and gives an overview of the flow of information.

Two examples of proteins present in the cell as different protein species are the well-known proteins angiotensin-converting-enzyme (ACE) (Figure 3) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Figure 4). ACE is a component of the renin-angiotensin system, which regulates blood pressure, fluid homeostasis, and electrolyte balance. ACE contributes to blood pressure regulation by generating the vasoactive octapeptide angiotensin II from inactive decapeptide angiotensin I and by cleaving the



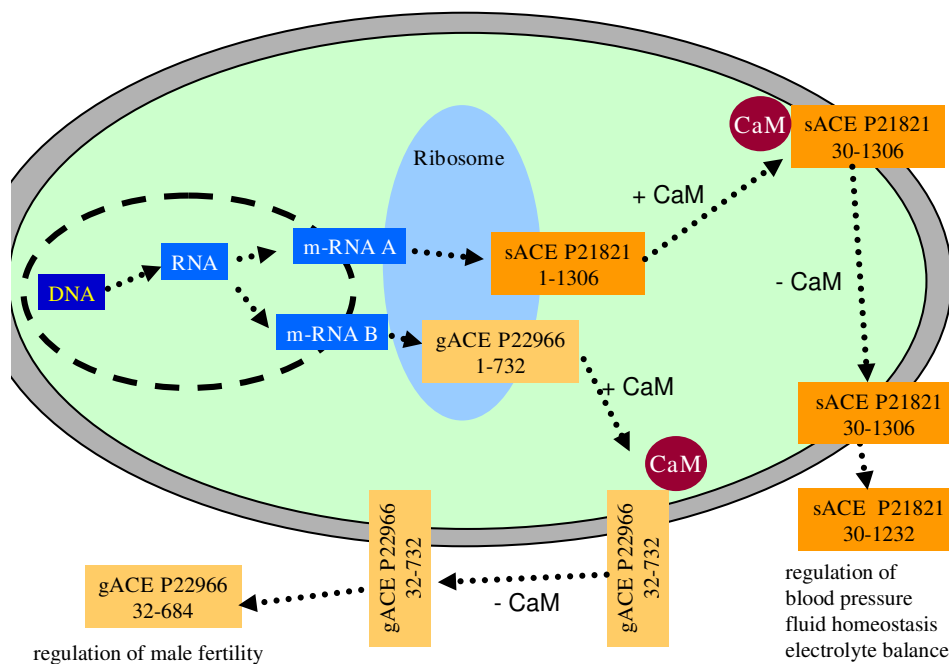
**Figure 2**  
Scheme of the flow of information starting with extracellular events via receptors, transcription factors and other proteins regulating gene expression and determining synthesis and processing of protein species. IP: interacting partner biomolecules.

vasodilator bradykinin. ACE exists as at least two protein species, germinal ACE (gACE) and somatic ACE (sACE), which are coded by the same gene but differently transcribed from two alternative tissue-specific promoters [28]. The larger glycosylated protein species (150 - 170 kDa [29]), sACE, is synthesized in neuronal cells, macrophages, renal epithelial cells and vascular endothelial cells and is involved in the regulation of blood pressure and renal function. The smaller glycosylated form (100 - 110 kDa [29]), gACE, is synthesized in maturing sperm cells and is involved in male fertility. Both protein species of the ACE gene are located in the membrane of the cell with a short cytoplasmic domain, a transmembrane domain, and a long extracellular domain containing the active sites. They are cleaved at the extracellular domain to release the soluble form of ACE into the extracellular fluid, including the blood [29-31]. ACE shedding is negatively regulated by the cytoplasmic domain of ACE, a domain that is not required for recognition by the ACE-secretase. Shedding of ACE is regulated by calmodulin (CaM), which binds to the cytoplasmic domain of ACE [32]. Dissociation of CaM from the cytoplasmic domain of ACE stimulates the cleavage secretion.

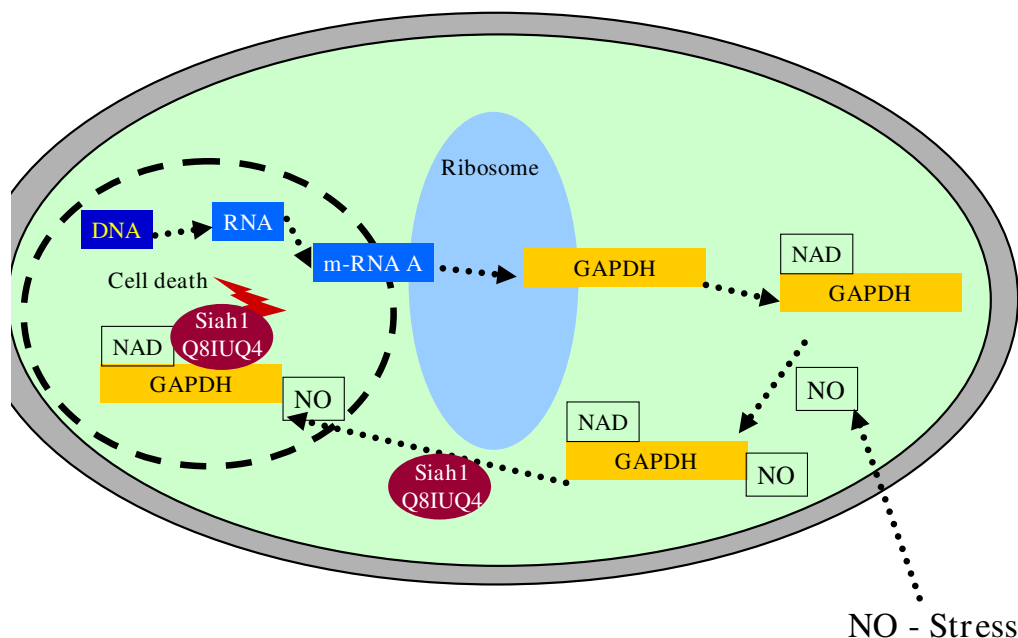
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was originally considered to be a glycolytic protein involved in energy generation. Recent results have shown,

however, that it is a multifunctional protein with cytosolic, nuclear and perinuclear localizations [33] (Table 1). Briefly, GAPDH becomes nitrosylated in the presence of high NO concentrations in the cytosol. After its nitrosylation, the protein interaction of GAPDH and Siah1 is significantly augmented. The GAPDH-Siah1 complex then is translocated to the nucleus. Siah1 is a component of a multiprotein E3 ubiquitin ligase complex that targets nuclear proteins for destruction via the proteasome. The GAPDH-Siah1 complex in the nucleus has been shown to facilitate the degradation of nuclear substrates, which results in cell death [34]. Figure 4 summarizes the role of GAPDH in apoptosis. GAPDH is a further representative example for the importance of the structure-function relationship of protein species. This example, among others, demonstrates the urgent need for a systematic nomenclature taking the structure-function relationship into account.

As a consequence of the new paradigm of the gene-protein relationship, we propose that the identity of each protein should be described more precisely in the future. Already in 1996 Jungblut et al. introduced and defined [1] the term protein species for proteins which are derived from a single gene but differ in protein chain length and/or post-translational modification. The term "protein species" [2], in this sense, is central to the new paradigm and the appli-



**Figure 3**  
**Protein species (sACE P21821; gACE P22966) derived from the angiotensin-converting enzyme gene. CaM: calmodulin.**



**Figure 4**  
**Protein species (GAPDH P04406) derived from the GAPDH gene.** Siah I: protein interacting with GAPDH.

cation of the protein species concept enables the development of a systematic nomenclature required for a comprehensive understanding of different functions of a protein.

#### **Protein species nomenclature based on the protein species concept**

For a protein species nomenclature we need to distinguish various levels: At the gene level a gene may have one or more than one transcript (transcript level); each transcript may lead to one or more than one translation. For the pro-

tein species nomenclature the following levels must be considered:

1. The initial protein species is the primary translation product. One would expect that, usually, one distinct transcript would lead to one distinct primary translation product. However, some of these initial protein species may be indistinguishable although derived from different transcripts. In *Caenorhabditis elegans*, for example, there are four genes coding for actins. The protein sequences encoded by genes 1 and 3 are identical (UniProt Knowledgebase AC P10983) [35]. On

**Table 1: Some examples of different functions and localizations of GAPDH.**

Function/Cellular role	Localization	PTM*	Interaction partner	Reference
Apoptosis (Figure 4)	Nucleus	Nitrosylation	Siah	[36]
Plasma membrane fusion	Membrane	-	Anion exchanger protein	[37]
Microtubule structure and function	Cytoplasm	-	Tubulin	[38]
Translational regulation	-	-	5V and 3VUTR mRNA sequence	[39]
Cell transport properties	Secretory vesicles	Phosphorylation	PKCL/E	[40]
Transcription control	Nucleus	-	Oct-1 transcription factor + promyelocytic leukemia nuclear bodies	[41] [42]
Intranuclear transport	Nucleus	-	Telomeres	[43]
Cell proliferation	-	-	tRNA precursors	[44]
Nuclear membrane fusion	Nuclear envelope	-	Ap <sub>4</sub> A	[45,46]
			-	[47]

\*post-translational modification

the other hand, some unique transcripts may lead - due to different transcription start sites - to different initial protein species, and there are even a few cases where a human gene gives rise to an mRNA that seems to be bistrionic and encodes for two different products. Examples are TREX2 (UniProt Knowledgebase AC Q9BQ50) [35] and UCHL5IP (UniProt Knowledgebase AC Q99871) [35]. Most mRNAs that encode UCHL5IP also include the N-terminal part of TREX2. The initial protein species may be the final functional protein species, or may undergo further processing such as proteolytic cleavage and/or chemical modification.

2. Proteolytically processed protein species. The protein processing procedure may lead to the final functional protein species or the protein species may still be subject to additional chemical modification.

3. A protein may undergo additional chemical modification such as acetylation, phosphorylation, etc.

On a species scale, on all these levels, we also have to consider nucleotide polymorphisms (SNPs, insertions, deletions, etc) that may change the amino acid composition of the derived protein species. Because there is currently no systematic nomenclature available for the exact description of protein species, we present such a nomenclature. The following combination of terms is suggested for a complete description of an individual protein species. Each describing parameter is contained in square brackets. An overview of the individual terms of the protein species nomenclature is given in Table 2. Below the detailed description of the terms is given.

#### *Gene level*

The name of a protein species starts with a descriptor which is identical with its entry name in the knowledgebase UniProt [35], containing the term for the gene name and the term for the species. E.g., the descriptor for the human gene coding an angiotensin-converting-enzyme is G\_ACE\_human. If no analytical data are available the descriptor for the gene gives far more reliable information about the identity of the protein species for the description of biological experiments than synonyms which are still in wide use.

#### *Nucleotide polymorphism level*

Describe the exact protein isoform adding the accession number of the record describing the polymorphism. E.g., SNP\_rs10853044 describes an SNP of the human angiotensin-converting-enzyme which is responsible for the replacement of Leu by Pro at the amino acid position 132 of P12821-1\_1.

#### *Initial Protein Species level*

The identity of a protein species giving access to its initial amino acid sequence, which is observed directly after its protein synthesis (initial protein species level), is defined by the protein database accession number and sequence version number. In the nomenclature suggested here, the name of a protein starts with AC\_ followed by the accession number and the sequence version number. For example, the name for the somatic species of the human angiotensin-converting enzyme is [AC\_ P12821-1\_1]. In many cases accession numbers for different splicing variants are already available. For example, the two splicing variants of the angiotensin-converting enzyme are designated P12821 (representing the somatic species) and P22966 (representing the testis species).

The human serine/threonine-protein phosphatase also exists in several splicing variants. Here the name of the regulatory subunit beta species is [AC\_ Q15173\_2]. The name for the species beta-2 is [AC\_ Q15173-2\_2]. The splicing variant is indicated here by -2. The sequence version number is designated by \_2.

If no accession number exists for the exact transcript translation (in the case of an alternative spliced protein species without a splice isoform number), describe the exact transcript translation starting with S\_ (S for splicing). S should be followed by D for deletion (SD\_), if amino acids are missing compared to the canonical protein record described by the protein accession and sequence version number. If, say, the unspliced protein species consists of 100 amino acids and the spliced species is missing the amino acids 70-79 the term is [SD\_70-79].

If a peptide chain is inserted as compared to the canonical protein record, the term starts with SI\_ followed by the number of the amino acid preceding the inserted peptide chain. This number is followed by the amino acid sequence written in the one-letter code. E.g., if the peptide chain LLELFVMFL is inserted at the position of amino acid number 43 the term then is [SI\_43\_ LLELFVMFL].

An exchange (E) of one amino acid or several amino acids can be designated [SE\_38\_RQELWQG\_SKEHWNQ]. Where 38 indicates the number of the first exchanged amino acid of the peptide. SKEHWNQ is the peptide present in the protein species, which was substituted for RQELWQG.

#### *Proteolytically processed protein species level*

Describe the protein sequence after proteolytic processing, starting with T\_: e.g., [T\_1-17] indicates that the first 17 amino acids were removed.

**Table 2: Overview of the individual terms (descriptors) of the protein species nomenclature**

Level	Comment	Descriptor
A: Gene	Entry name (gene name + species)	G_ACE_human <sup>a</sup>
B: Nucleotide polymorphism	SNP Accession number	SNP_rs10853044
C: Initial amino acid sequence of the protein synthesized at the ribosome	Accession number (AC) + sequence version (_1)	AC_P12821-1_1
	Splicing variant (S) with deletion _ localisation of the deletion	SD_70-79
	Splicing variant with insertion _ localisation of the insertion	SI_43_LLELFVMFL
	Splicing variant with an exchange of amino acids	SE_38_RQELWQG_SKEHWNQ <sup>b</sup>
D: Proteolysis	Truncated amino acids: 1-17: the first 17 amino acids were removed	T_1-17
E: Posttranslational Modification (P)	Position of the modified amino acid	P_33_21 <sup>c</sup>
F: Cofactor (C)	Identity of non-covalently bound cofactors	C_Zn(2x) <sup>d</sup>
G: Further descriptors	Can be added, if necessary	X, Y

<sup>a</sup>according to the term in the UniProt knowledgebase [35]

<sup>b</sup>38 indicates the number of the first exchanged amino acid of the peptide. SKEHWNQ is the peptide present in the protein species, which was substituted for RQELWQG.

<sup>c</sup> the first number (here: 33) gives the position of the modified amino acid; the second number gives the type of the post-translational modification according to the UniMod [6] accession number (here: 21 indicating a phosphorylation)

<sup>d</sup> here the term indicates the presence of 2 cations of Zn

#### Chemically modified protein level

Describe the post-translational modification(s), starting with P\_ followed by the number of the amino acid that has been modified and the UniMod accession number [6]. As an example take: [P\_33\_21]. Here 33 indicates the amino acid that is phosphorylated and 21 is the accession number for phosphorylation from UniMod.

#### Cofactor level

Describe the identity of non-covalently bound cofactors starting with C\_.

Further descriptors [X,Y] can be added, if necessary. As an example we give the combination of descriptors for the full description of the human somatic angiotensin-converting enzyme, including an SNP located in the membrane of vascular smooth muscle cells:

[G\_ACE\_human]+ [SNP\_rs10853044]+ [AC\_P12821-1\_1]+ [T\_1-29/1233-1306]+ [P\_N-linked-carbohydrate chains at 38, 54, 74, 111, 146, 160, 318, 445, 509, 677, 695, 714, 760, 942, 1191]+[P\_21]+ [C\_Zn(2x)].

The data for the carbohydrate chains were taken from the UniProt Knowledgebase [35].

This nomenclature takes the structure-function relationship into account. Furthermore, it is well suited for database searches and will provide a more reliable foundation for systems biology approaches. We recommend that in future publications on protein species authors use this nomenclature, presenting the complete protein species term at least in the Abstract and in the Introduction of

their manuscript. Since the complete protein species descriptor is long, the author should define a short form substitute of the protein species for use in the rest of the paper.

If not every detail of the exact chemical composition can be determined, the author should include those terms which are accessible. As a minimum, the entry number according to UniProt Knowledgebase [35] should be given.

Protein databases such as the UniProt Knowledgebase [35] will ideally create unique identifiers for each protein species identified. This will take time to implement, however, and even then the assignment of such unique identifiers will lag behind the detection of new protein species. The creation and use of protein species descriptors will therefore be of permanent value - unless it in its turn is replaced someday by an improved nomenclature.

#### Conclusion

We have presented a scheme for the accurate and detailed description of protein species. We propose that future investigations of protein function focusing on defined protein species use this protein species terminology which includes the complete terms necessary for the comprehensive description of a protein species. It is desirable not only to characterize the protein function, but to identify the target protein species in question with a sequence coverage as high as possible - ideally 100% - and the identification of all post-translational modifications. These requirements should also be addressed in proteome analytical investigations, by digesting protein mixtures not



only with trypsin but by several other proteases in parallel, thus achieving higher sequence coverage. In the short term in many cases it may not be possible to achieve 100% sequence coverage and/or the identification of every PTM of individual protein species. Nonetheless, even in these cases use of the protein species nomenclature will be helpful, indicating the current level of knowledge of the exact chemical structure and giving hints as to the direction further investigations should take.

In summary, we propose a tool for the storage of information on protein species which makes accessible a clear chemical description of protein molecules.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

HS developed the nomenclature, wrote the text, and contributed Figures 2 to 4. Parts of the text and Figure 1 were prepared by PRJ, who contributed - together with HGH and RA - to the concept of the nomenclature and critically revised the text.

### Acknowledgements

Anna Walduck and Peter Germain are acknowledged for editorial support.

### References

- Jungblut P, Thiede B, Zimny-Arndt U, Muller EC, Scheler C, Wittmann-Liebold B, Otto A: **Resolution power of two-dimensional electrophoresis and identification of proteins from gels.** *Electrophoresis* 1996, **17**:839-47.
- Jungblut PR, Holzhuetter HG, Apweiler R, Schlüter H: **The Speciation of the Proteome.** *Chem Cent J* 2008, **2**:16.
- Blanchard JM, Brissac C, Jeanteur P: **Characterization of a protein species isolated from HeLa cell cytoplasm by affinity chromatography on polyadenylate-sepharose.** *Proc Natl Acad Sci USA* 1974, **71**:1882-6.
- O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins.** *J Biol Chem* 1975, **250**:4007-21.
- Joint Commission on Biochemical Nomenclature IUPAC-IUBMB: **Nomenclature of multiple forms of enzymes.** In *Biochemical Nomenclature and Related Documents* 2nd edition. Edited by: Liébecq C. Colchester: Portland Press; 1992.
- UNIMOD - Protein modifications for mass spectrometry: [[http://www.unimod.org/modifications\\_list.php](http://www.unimod.org/modifications_list.php)].
- Creasy DM, Cottrell JS: **Protein modifications for mass spectrometry.** *Proteomics* 2004, **4**:1534-6.
- Nielsen ML, Savitski MM, Zubarev RA: **Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics.** *Mol Cell Proteomics* 2006, **5**:2384-91.
- Sun SC: **Deubiquitylation and regulation of the immune response.** *Nat Rev Immunol* 2008, **8**:501-11.
- Klose J, Kobalz U: **Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome.** *Electrophoresis* 1995, **16**:1034-59.
- Jungblut P, Thiede B: **Protein identification from 2-DE gels by MALDI mass spectrometry.** *Mass Spectrom Rev* 1997, **16**:145-62.
- McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR 3rd: **Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level.** *Anal Chem* 1997, **69**:767-76.
- Kelleher N: **Top-down proteomics.** *Anal Chem* 2004, **76**:197A-203A.
- Han X, Jin M, Breuker K, McLafferty FW: **Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons.** *Science* 2006, **314**(5796):109-12.
- Roth M, Forbes A, Boyne M, Kim Y, Robinson D, Kelleher N: **Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry.** *Mol Cell Proteomics* 2005, **4**:1002-8.
- Schmidt F, Donahoe S, Hagens K, Mattow J, Schaible U, Kaufmann S, Aebersold R, Jungblut PR: **Complementary analysis of the Mycobacterium tuberculosis proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology.** *Mol Cell Proteomics* 2004, **3**(1):24-42.
- Wu S, Lourette NM, Tolic N, Zhao R, Robinson EW, Tolmachev AV, Smith RD, Pasa-Tolic L: **An integrated top-down and bottom-up strategy for broadly characterizing protein isoforms and modifications.** *J Proteome Res* 2009, **8**(3):1347-57.
- Scheler C, Muller E, Stahl J, Muller-Verdan U, Sainikow J, Jungblut P: **Identification and characterization of heat shock protein 27 protein species in human myocardial two-dimensional electrophoresis patterns.** *Electrophoresis* 1997, **18**:2823-31.
- Klose J, Nock C, Herrmann M, Stuhler K, Marcus K, Bluggel M, Krause E, Schalkwyk LC, Rastan S, Brown SD, Bussow K, Himmelbauer H, Lehrach H: **Genetic analysis of the mouse brain proteome.** *Nat Genet* 2002, **30**(4):385-93.
- Mattow J, Jungblut PR, Muller E, Kaufmann S: **Identification of acidic, low molecular mass proteins of Mycobacterium tuberculosis strain H37Rv by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry.** *Proteomics* 2001, **1**:494-507.
- Strahl B, Allis C: **The language of covalent histone modifications.** *Nature* 2000, **403**:41-5.
- Brumbaugh J, Phanstiel D, Coon JJ: **Unraveling the histone's potential: a proteomics perspective.** *Epigenetics* 2008, **3**:254-7.
- Sims RJ 3rd, Reinberg D: **Is there a code embedded in proteins that is based on post-translational modifications?** *Nat Rev Mol Cell Biol* 2008, **9**:815-20.
- Mayer MP, Bukau B: **Hsp70 chaperones: cellular functions and molecular mechanism.** *Cell Mol Life Sci* 2005, **62**(6):670-84.
- Morishima N: **Control of cell fate by Hsp70: more than an evanescent meeting.** *J Biochem (Tokyo)* 2005, **137**(4):449-53.
- Okkels L, Muller E, Schmid M, Rosenkrands I, Kaufmann S, Andersen P, Jungblut PR: **CFPI10 discriminates between nonacetylated and acetylated ESAT-6 of Mycobacterium tuberculosis by differential interaction.** *Proteomics* 2004, **4**:2954-60.
- Warren EH, Vigneron NJ, Gavin MA, Coulie PG, Stroobant V, Dalet A, Tykodi SS, Xuereb SM, Mito JK, Riddell SR, Eynde BJ Van den: **An antigen produced by splicing of noncontiguous peptides in the reverse order.** *Science* 2006, **313**:1444-7.
- Kumar RS, Thekkumkara TJ, Sen GC: **The mRNAs encoding the two angiotensin-converting isozymes are transcribed from the same gene by a tissue-specific choice of alternative transcription initiation sites.** *J Biol Chem* 1991, **266**:3854-62.
- Ramchandran R, Sen I: **Cleavage processing of angiotensin-converting enzyme by a membrane-associated metalloprotease.** *Biochemistry* 1995, **34**:12645-52.
- Ramchandran R, Kasturi S, Douglas JG, Sen I: **Metalloprotease-mediated cleavage secretion of pulmonary ACE by vascular endothelial and kidney epithelial cells.** *Am J Physiol* 1996, **271**(2 Pt 2):H744-51.
- Sadhukhan R, Santhamma KR, Reddy P, Peschon JJ, Black RA, Sen I: **Unaltered cleavage and secretion of angiotensin-converting enzyme in tumor necrosis factor-alpha-converting enzyme-deficient mice.** *J Biol Chem* 1999, **274**:10511-6.
- Chattopadhyay S, Santhamma KR, Sengupta S, McCue B, Kinter M, Sen GC, Sen I: **Calmodulin binds to the cytoplasmic domain of angiotensin-converting enzyme and regulates its phosphorylation and cleavage secretion.** *J Biol Chem* 2005, **280**(40):33847-55.
- Sirover MA: **New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase.** *Biochim Biophys Acta* 1999, **1432**:159-184.
- Hara MR, Snyder SH: **Nitric Oxide-GAPDH-Siah: A Novel Cell Death Cascade.** *Cell Mol Neurobiol* 2006, **26**:527-38.
- The UniProt Consortium: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2009, **37**:D169-D174.

36. Hara MR, Cascio MB, Sawa A: **GAPDH as a sensor of NO stress.** *Biochim Biophys Acta* 2006, **1762**:502-509.
37. Hessler RJ, Blackwood RA, Brock TG, Francis JW, Harsh DM, Smolen JE: **Identification of glyceraldehyde-3-phosphate dehydrogenase as a Ca<sup>2+</sup>-dependent fusogen in human neutrophil cytosol.** *J Leukoc Biol* 1998, **63(3)**:331-6.
38. Muronetz VI, Wang ZX, Keith TJ, Knull HR, Srivastava DK: **Binding constants and stoichiometries of glyceraldehyde 3-phosphate dehydrogenase-tubulin complexes.** *Arch Biochem Biophys* 1994, **313**:253-260.
39. Schultz DE, Hardin CC, Lemon SM: **Specific interaction of glyceraldehyde-3-phosphate dehydrogenase with the 5V-nontranslated RNA of hepatitis A virus.** *J Biol Chem* 1996, **271**:14134-14142.
40. Tisdale EJ: **Glyceraldehyde-3-phosphate dehydrogenase is required for vesicular transport in the early secretory pathway.** *J Biol Chem* 2001, **276**:2480-2486.
41. Zheng L, Roeder RG, Luo Y: **S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component.** *Cell Mol Life Sci* 2003, **114**:255-266.
42. Carlisle GW, Tatton WG, Borden DLB: **Demonstration of a RNA-dependent nuclear interaction between the promyelocytic leukaemia protein and glyceraldehyde-3-phosphate dehydrogenase.** *Biochem J* 1998, **335**:691-696.
43. Sundararaj KP, Wood RE, Ponnusamy S, Salas AM, Szulc Z, Bielawska A, Obeid LM, Hannun YA, Ogretmen B: **Rapid shortening of telomere length in response to ceramide involves the inhibition of telomere binding activity of nuclear glyceraldehyde-3-phosphate dehydrogenase.** *J Biol Chem* 2004, **279**:6152-6162.
44. Singh R, Green MR: **Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase.** *Science* 1993, **259**:365-368.
45. Baxi MD, Vishwanatha JK: **Uracil DNA lysoylase/glyceraldehyde-3-phosphate dehydrogenase is an Ap4A binding protein.** *Biochemistry* 1995, **34**:9700-9707.
46. Xing C, LaPorte JR, Barbay JK, Myers AG: **Identification of GAPDH as a protein target of the saframycin antiproliferative agents.** *Proc Natl Acad Sci USA* 2004, **101**:5862-5866.
47. Nakagawa T, Hirano Y, Inomata A, Yokota S, Miyachi K, Kaneda M, Umeda M, Furukawa K, Omata S, Hoarigome T: **Participation of a fusogenic protein, glyceraldehyde-3-phosphate dehydrogenase, in nuclear membrane assembly.** *J Biol Chem* 2003, **278**:20395-20404.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

*“Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge.”*

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
<http://www.chemistrycentral.com/manuscript/>



**ChemistryCentral**