We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Introductory Chapter: Proteoforms

Xianquan Zhan

1. Introduction

The completion of human genome sequence has driven the research focusing from structural genomics to functional genomics. Transcriptomics and proteomics are two main contents in the era of functional genomics. Human genome contains about 20,300 genes [1]. However, RNA splicing and other factors in the transcriptional process from a gene to RNA result in multiple transcripts that are derived from the same single-one gene. Thus, human transcriptome is estimated to contain at least 100,000 transcripts, much more than the number of human genes. Each transcript guides the ribosome to synthesize an amino acid sequence of a protein. The synthesized protein in the ribosome must be translocated and redistributed to the appropriate locations to form a special conformation and interact with surrounding molecules, namely, a complex, to exert its biological functions. Also, protein is modified by many posttranslational modifications (PTMs) and even unknown factors in the process of translocation and redistribution. An estimated 400–600 PTMs in human body are the main factors to cause the complexity and diversity of proteins, namely, protein species [2, 3] or proteoforms [4, 5]. Thus, multiple proteoforms are often derived from one same transcript, and it is estimated that human proteome contains at least 1,000,000 proteoforms [6]. A proteoform is the basic unit in a proteome, and it is defined as its amino acid sequence + PTMs + spatial conformation + localization + cofactors + binding partners + a function (**Figure 1**), which is the final functional performer of a gene [6]. A protein is an umbrella term for all proteoforms coded by the same gene. Moreover, the different proteoforms derived from one same gene might have different conformation and functions. Each proteoform has its own copy number or abundance, which can be quantified between given conditions [4]. Studies on proteoforms will offer much more in-depth insights into a proteome, which will directly lead to the discovery of reliable biomarkers to understand accurate molecular mechanisms, the discovery of effective therapeutic targets, and for effective prediction, diagnosis, and prognostic assessment.

It is a big challenge in the methodology to study the over millions of human proteoforms [1, 6]. The common bottom-up mass spectrometry (MS)-based strategies cannot identify proteoforms, which in fact only identify protein-coded genes, a protein group. This type of method includes stable isotope-labeled two-dimensional liquid chromatography-tandem mass spectrometry (2DLC-MS/MS) and stable isotope-free 2DLC-MS/MS, which only identify peptides and PTMs (**Figure 2**) [6]. Top-down MS-based strategies have been developed to identify proteoforms [7–9]. This type of method can identify proteoforms, which obtains the proteoform message including the amino acid sequence and PTMs. However, the obtained message of proteoform is only partial information of the above defined proteoform. Also, the protein must be purified prior to MS analysis, with different types of protein isolation techniques such as capillary zone electrophoresis (CZE) and liquid
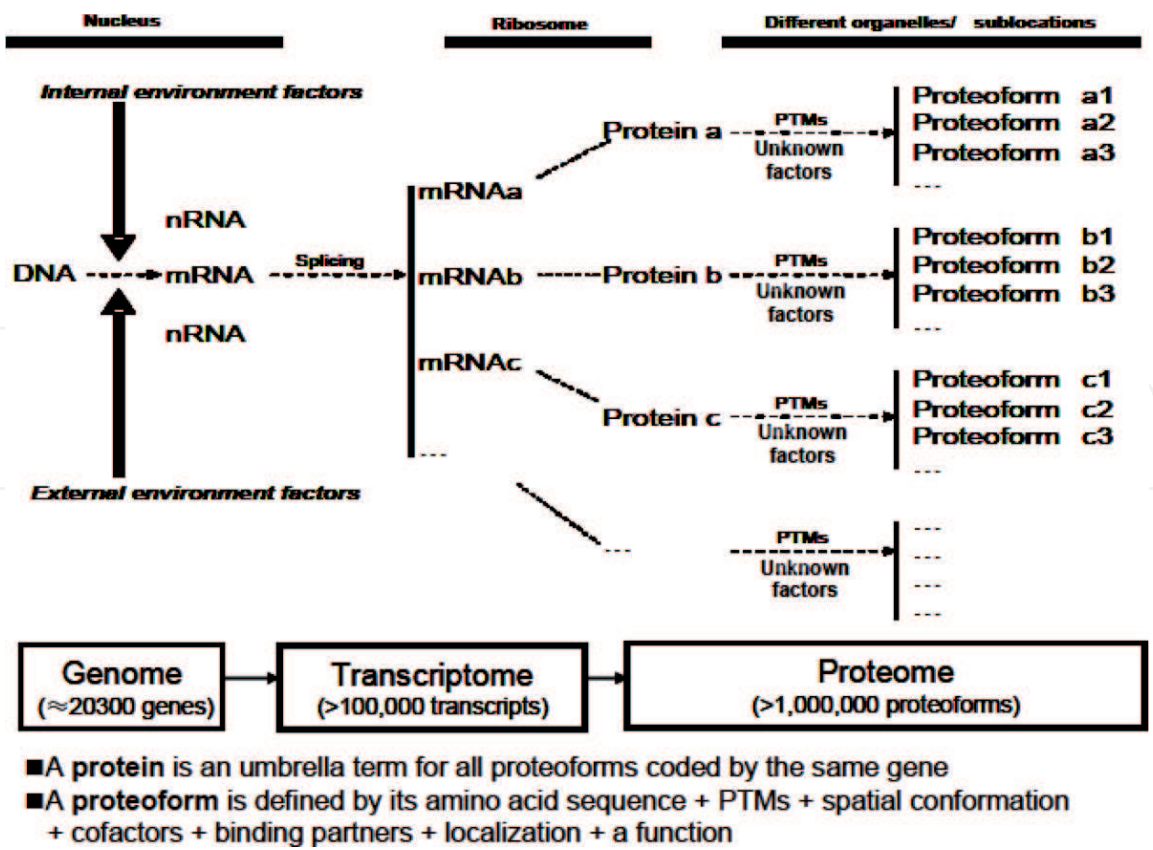
Figure 1.
The concept and formation model of proteoform. (Reproduced from Zhan et al. [1, 6], copyright permission with open access policy.)
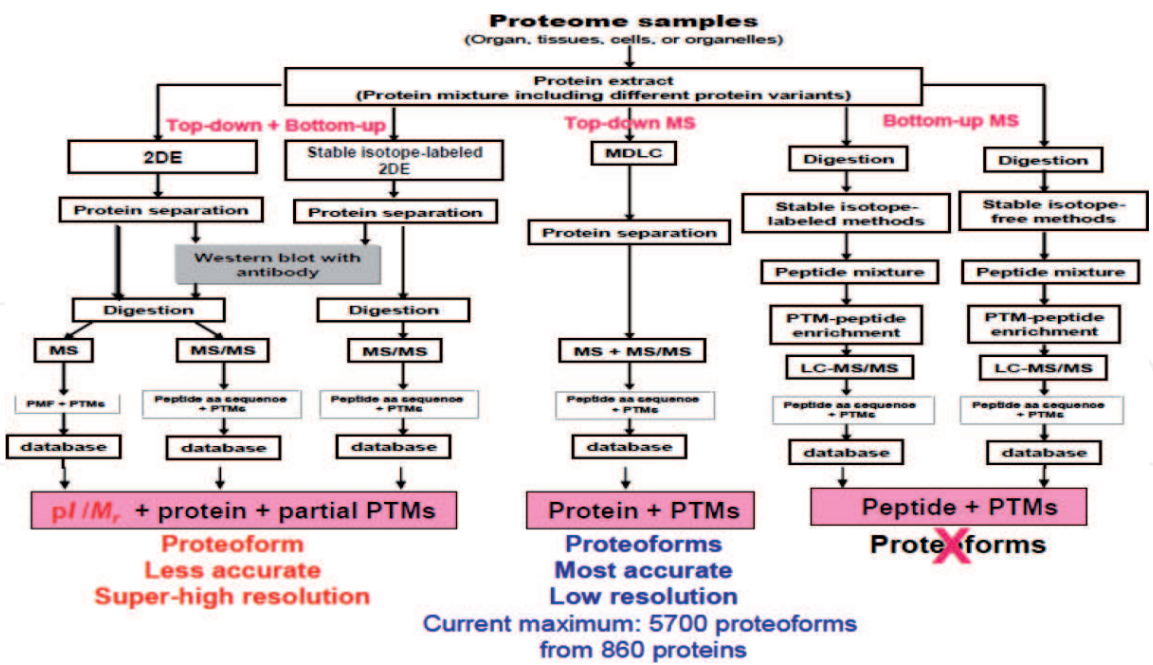

Figure 2.
The methods to study proteoforms. (Reproduced from Zhan et al. [1, 4, 6], copyright permission with open access policy.)

chromatography (LC) [10, 11]. Another drawback is the low ratio of signal to noise (S/N) in the MS analysis. All of those factors result in a relative low throughput in identification of human proteoforms. Currently the maximum throughput of top-down MS is up to 5700 proteoforms corresponding to 860 proteins (**Figure 2**) [6]. Two-dimensional gel electrophoresis (2DE)-liquid chromatography-MS

(2DE-LC/MS) strategy combines the top-down technique (2DE) and bottom-up technique (LC/MS), which is currently superhigh-throughput method to identify the large-scale proteoforms [1, 6, 12, 13]. With the innovating concept and practice of 2DE, 2DE is a real prefractionation method, which can effectively recognize isoelectric point (pI) and the relative mass (Mr)—two essential parameter of a proteoform; each 2D gel spot contains over 50 to several hundred proteoforms, and most of proteoforms are low-abundance. Currently, the largest 2D gel is 30 cm x 40 cm, which can separate 10,000 2D gel spots; thus at least 500,000 or 1,000,000 proteoforms can be identified. LC/MS can identify protein sequences and partial PTMs (**Figure 2**) [1, 13]. 2DE-LC/MS has great potential in analysis of large-scale proteoforms. 2DE-LC/MS and top-down MS are complementary in the achievement of maximum coverage of human proteoforms in a proteome.

Proteoform is the final functional format of a protein coded by a gene, which has important scientific merits in the fields of life sciences and medical sciences, and it is the research hot spot and international scientific frontiers. In the past 1–2 years, one has gradually paid more attention to the proteoform study. A total of 532 publications can be obtained through searching in the PubMed dataset with the keyword "proteoform or proteoforms." For example, 24 growth hormone (GH) proteoforms were identified with 2DE-LC/MS in human pituitary tissues [14], and 20 and 22 kDa GH proteoforms functioned in different signaling profiles. Six prolactin (PRL) proteoforms were identified with 2DE-LC/MS and 2DE-Western blot in human pituitary tissues, and the proportional ratio of six PRL proteoforms were significantly different among different subtype nonfunctional pituitary adenoma relative to control pituitary tissues [15]. The six PRL proteoforms bind to different long or short PRL receptors to exert their functions. A total of 3090 proteoforms were identified with liquid chromatography-MS (LC/MS), and 417 proteoforms were identified with sheathless CZE-MS, in seminal plasma [10]. A total of 3028 proteoforms corresponding to 387 proteins from *E. coli* cells were identified with coupling size exclusion chromatography (SEC) to CZE-activated ion electron transfer dissociation (CZE-AI-ETD) [16]. Human sperm protamine proteoforms were identified with a combination of top-down and bottom-up MS [17]. The glioblastoma [12, 13] and pituitary adenoma [13–15] tissue proteoforms were investigated with 2DE-LC-MS/MS. Proteoforms were identified from several cell lines (HepG2, glioblastoma, LEH) with 2DE-LC/MS [18]. Also, proteoform dynamics is also investigated underlying the senescence associated secretory phenotype [19].

In summary, development of proteoforms or protein species significantly enriches the concept of proteome, which is the next-generation research direction in the field of proteomics. 2DE-LC/MS and top-down MS are the complementary method to study the large-scale proteoforms. In-depth investigating proteoforms in a proteome with different pathophysiological conditions will directly cause to deeply understand disease molecular mechanisms, discover the reliable and effective therapeutic targets, and identify effective predictive, diagnostic, and prognostic biomarkers. Further, each proteoform is involved in a molecular network system and has multiple PTMs. It is the research hot spot how different PTMs competitively or synergistically affect proteoform structure and functions and their involved molecular network system [20–24]. Molecular network-based proteoform pattern biomarkers will have more important scientific merits.

Proteoforms are involved in the entire life science and medical sciences. This book contains only a fraction of the important frontier "proteoforms," which serve as a spur to stimulate and encourage researchers who study proteoforms to come forward with its scientific merits to research and clinical practice. This book will focus on the concept of proteoform, technologies to study proteoforms, and applications of proteoforms.

Acronyms and abbreviations

CZE	capillary zone electrophoresis
GH	growth hormone
LC	liquid chromatography
Mr.	relative mass
MS	mass spectrometry
MS/MS	tandem mass spectrometry
pI	isoelectric point
PRL	prolactin
PTM	posttranslational modification
S/N	ratio of signal to noise
2DE	two-dimensional gel electrophoresis
2DLC	two-dimensional liquid chromatography

Author details

Xianquan Zhan[1,2,3]

1 University Creative Research Initiatives Center, Shandong First Medical University, Shandong, China

2 Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha, China

3 State Local Joint Engineering Laboratory for Anticancer Drugs, Xiangya Hospital, Central South University, Changsha, China

*Address all correspondence to: yjzhan2011@gmail.com

IntechOpen

References

[1] Zhan X, Li N, Zhan X, Qian S. Revival of 2DE-LC/MS in proteomics and its potential for large-scale study of human proteoforms. Med One. 2018;**3**:e180008. DOI: 10.20900/mo.20180008

[2] Jungblut PR, Holzhütter HG, Apweiler R, Schlüter H. The speciation of the proteome. Chemistry Central Journal. 2008;**2**:16. DOI: 10.1186/1752-153X-2-16

[3] Schlüter H, Apweiler R, Holzhütter HG, Jungblut PR. Finding one's way in proteomics: A protein species nomenclature. Chemistry Central Journal. 2009;**3**:11. DOI: 10.1186/1752-153X-3-11

[4] Zhan X, Long Y, Lu M. Exploration of variations in proteome and metabolome for predictive diagnostics and personalized treatment algorithms: Innovative approach and examples for potential clinical application. Journal of Proteomics. 2018;**188**:30-40. DOI: 10.1016/j.jprot.2017.08.020

[5] Smith LM, Kelleher NL. Consortium for top down proteomics. Proteoform: A single term describing protein complexity. Nat. Methods. 2013;**10**(3):186-187. DOI: 10.1038/nmeth.2369

[6] Zhan X, Li B, Zhan X, Schlüter H, Jungblut PR, Coorssen JR. Innovating the concept and practice of two-dimensional gel electrophoresis in the analysis of proteomes at the proteoform level. Proteomes. 2019;**7**(4):36. DOI: 10.3390/proteomes7040036

[7] Chaffer LV, Millikin RJ, Miller RM, Anderson LC, Fellers RT, Ge Y, et al. Identification and quantification of proteoforms by mass spectrometry. Proteomics. 2019;**19**(10):e1800361. DOI: 10.1002/pmic.201800361

[8] Cupp-Sutton KA, Wu S. High-throughput quantitative top-down proteomics. Molecular Omics. 2020. DOI: 10.1039/c9mo00154a

[9] Shaw JB, Liu W, Vasil Ev YV, Bracken CC, Malhan N, Guthals A, et al. Direct determination of antibody chain pairing by top-down and middle-down mass spectrometry using electron capture dissociation and ultraviolet photodissociation. Analytical Chemistry. 2020;**92**(1):766-773. DOI: 10.1021/acs.analchem.9b03129

[10] Gomes FP, Diedrich JK, Saviola AJ, Memili E, Moura AA, Yates JR. EThcD and 213 nm UVPD for top-down analysis of bovine seminal plasma proteoforms on electrophoretic and chromatographic time frames. Analytical Chemistry. 2020;**92**(4):2979-2987. DOI: 10.1021/acs.analchem.9b03856

[11] Melby JA, Jin Y, Lin Z, Tucholski T, Wu Z, Gregorich ZR, et al. Top-down proteomics reveals myofilament proteoform heterogeneity among various rat skeletal muscle tissues. Journal of Proteome Research. 2020;**19**(1):446-454. DOI: 10.1021/acs.jproteome.9b00623

[12] Peng F, Li J, Guo T, Yang H, Li M, Sang S, et al. Nitroproteins in human astrocytomas discovered by gel electrophoresis and tandem mass spectrometry. Journal of the American Society for Mass Spectrometry. 2015;**26**(12):2062-2076. DOI: 10.1007/s13361-015-1270-3

[13] Zhan X, Yang H, Peng F, Li J, Mu Y, Long Y, et al. How many proteins can be identified in a 2-DE gel spot within an analysis of a complex human cancer tissue proteome? Electrophoresis. 2018;**39**:965-980. DOI: 10.1002/elps.201700330

[14] Zhan X, Giorgianni F, Desiderio DM. Proteomics analysis of growth hormone isoforms in the human pituitary. Proteomics. 2005;**5**(5): 1228-1241. DOI: 10.1002/pmic. 200400987

[15] Qian S, Yang Y, Li N, Cheng T, Wang X, Liu J, et al. Prolactin variants in human pituitaries and pituitary adenomas identified with two-dimensional gel electrophoresis and mass spectrometry. Frontiers in Endocrinology. 2018;**9**:468. DOI: 10.3389/fendo.2018.00468

[16] McCool EN, Lodge JM, Basharat AR, Liu X, Coon JJ, Sun L. Capillary zone electrophoresis-tandem mass spectrometry with activated ion electron transfer dissociation for large-scale top-down proteomics. Journal of the American Society for Mass Spectrometry. 2019;**30**(12):2470-2479. DOI: 10.1007/s13361-019-02206-6

[17] Soler-Ventura A, Gay M, Jodar M, Vilanova M, Castillo J, Arauz-Garofalo G, et al. Characterization of human sperm protamine proteoforms through a combination of top-down and bottom-up mass spectrometry approaches. Journal of Proteome Research. 2020;**19**(1):221-237. DOI: 10.1021/acs.jproteome.9b00499

[18] Naryzhny SN, Zorina ES, Kopylov AT, Zgoda VG, Kleyst OA, Archakov AI. Next steps on in silico 2DE analyses of chromosome 18 proteoforms. Journal of Proteome Research. 2018;**17**(12):4085-4096. DOI: 10.1021/acs.jproteome.8b00386

[19] Doubleday PF, Fornelli L, Kelleher NL. Elucidating proteoform dynamics underlying the senescence associated secretory phenotype. Journal of Proteome Research. 2020;**19**(2):938-948. DOI: 10.1021/acs. jproteome.9b00739

[20] Long Y, Lu M, Cheng T, Zhan X, Zhan X. Multiomics-based signaling pathway network alterations in human non-functional pituitary adenomas. Frontiers in Endocrinology. 2019;**10**:835. DOI: 10.3389/fendo.2019.00835

[21] Lu M, Zhan X. The crucial role of multiomic approach in cancer research and clinically relevant outcomes. The EPMA Journal. 2018;**9**(1):77-102. DOI: 10.1007/s13167-018-0128-8

[22] Zhan X, Long Y. Exploration of molecular network variations in different subtypes of human non-functional pituitary adenomas. Frontiers in Endocrinology. 2016;**7**:13. DOI: 10.3389/fendo.2016.00013

[23] Cheng T, Zhan X. Pattern recognition for predictive, preventive, and personalized medicine in cancer. The EPMA Journal. 2017;**8**(1):51-60. DOI: 10.1007/s13167-017-0083-9

[24] Zhan X, Desiderio DM. Editorial: Molecular network study of pituitary adenomas. Frontiers in Endocrinology. 2020;**11**:26. DOI: 10.3389/ fendo.2020.00026