Journal of Proteomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

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

Inventory of proteoforms as a current challenge of proteomics: Some technical aspects

Stanislav Naryzhny[a,b,*]

[a] Orekhovich Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Pogodinskaya 10, Moscow 119121, Russia
[b] B.P. Konstantinov Petersburg Nuclear Physics Institute, National Research Center "Kurchatov Institute", Gatchina, Leningrad Region 188300, Russia

ARTICLE INFO

Keywords:
Protein species
Proteoform
Inventory
2-DE
Proteome
Mass spectrometry

ABSTRACT

The main intricacy in the human proteome is that it is tremendously complex and composed from diverse and heterogeneous gene products. These products are called protein species or proteoforms and are the smallest units of the proteome. In pursuit of the comprehensive profiling of the human proteome, significant advances in the technology of so called "Top-Down" mass spectrometry based proteomics, have been made. However, the scale of performance of this approach is still far behind the "Bottom-Up", peptide-centric techniques. The classical two-dimensional electrophoresis (2-DE) as the most powerful and convenient method for separation of proteoforms remains as a superior method in "Top-Down" proteomics. Here, some aspects of approaches for establishing an inventory of proteoforms based on 2-DE and mass spectrometry are discussed.

Biological significance: The systematic efforts in the Human Proteome project to map the entire human proteome greatly depend on currently available and emerging techniques and approaches. Here, the possibilities of a visual representation of the human proteome by combination of virtual/experimental 2-DE with protein identification by mass spectrometry or immunologically is discussed. By application of this approach on several profiles of gene products we show its convenience in informative representation of the whole proteome and single gene products, proteoforms (protein species). This approach could be very helpful in the emerging global inventory of all human proteoforms.

1. Introduction

The main aim in human proteomics as it was proposed by the Human Proteome Organization (HUPO) is a complete catalogue of all human proteins. Until now, this survey of proteins was performed mainly using the bottom-up approach that includes protein digestion and subsequent mass spectrometric analysis of the peptides produced by shot-gun on target mass spectrometry [1]. As the human genome is deciphered and the number and the sequences of protein coding genes, or even mRNA are known, the task is clear – find the corresponding protein(s). Due to collaborative efforts inside the Chromosome-based Human Proteome Project (C-HPP) this task is close to completion now. From the list of 20,199 predicted proteins, 17,168 have already been found (NextProt release 2017-08-01). Still ~3000 proteins remain as so-called "missing proteins" and attract the special attention of the proteomics community involved in C-HPP. But even after complete identification of the human proteome by a high-throughput bottom-up approach, only representative proteins will be identified in most cases [2]. The situation in proteomics is much more complicated as proteins can exist as different forms (protein species/proteoform) [3–5]. The term "protein species" was introduced for the description of the smallest unit in the proteome – "chemically clearly-defined molecule" [3]. Lately, the another term basically having the same meaning was coined. This term 'proteoform' is "to be used to designate all of the different molecular forms in which the protein product of a single gene can be found, including changes due to genetic variations, alternatively spliced RNA transcripts and post-translational modifications" [5,6]. It looks like this term is more compatible with a gene-centric approach for referring to a protein/gene, from which each proteoform originates, as citing Jungblut et al. "the term protein refers to its coding gene and, therefore, is the umbrella term for all of the developing protein species" [7]. It is of note to mention that the term "proteoform" actually sometimes is used for conformational variants of the same protein as well [8]. But in this review, this word is used only as a synonym for protein species with above mentioned definition made by Smith, and Kelleher [5]. Even inside this definition frame, a proteomics avenue of

Abbreviations: ABC, ammonium bicarbonate; FA, formic acid; C-HPP, Chromosome-based Human Proteome Project; CTDP, Consortium for Top-Down Proteomics; IF, isoelectric focusing; MRM, Multiple Reaction Monitoring; emPAI, exponentially modified form of protein abundance index

* Corresponding author at: Orekhovich Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Pogodinskaya 10, Moscow 119121, Russia.
E-mail address: naryzhnyy_sn@pnpi.nrcki.ru.

https://doi.org/10.1016/j.jprot.2018.05.008
Received 25 October 2017; Received in revised form 11 May 2018; Accepted 12 May 2018

proteoforms is very wide and long, and could encompass many billions of components [9–12]. For instance, all combinations of modifications (~30) of just histone H3 alone can theoretically produce > 1 billion of protein species [10,13]. Therefore, it seems that refined versions of high-throughput bottom-up approach and a clear strategy in proteoform profiling is needed [14–16]. Because of such a variety of proteoforms, range of concentration (7–8 orders of magnitude), and dynamic changes during life cycle, their identification and quantitation is a challenge difficult to attain. Nevertheless, there is evident progress in this area. If, to date, the main workhorse in proteomics was Bottom-Up mass spectrometry, it is the Top-Down approach that is becoming preeminent today [17]. Top-Down proteomics implies that mass spectrometry is applied at the proteoform level, allowing the acquisition of information about all intramolecular complexity preserved during analysis, that might be overlooked in Bottom-Up shotgun workflows [17,18]. Here, several approaches will be discussed, where a combination of top-down and bottom-up approaches is utilized to optimize the proteomics analysis. Though MS is a central technique and a driving force in proteomics, large scale Top-Down proteomics cannot be just a one-step procedure. There are also several protein fractionation or separation steps involved in protein species/proteoform analysis. Accordingly, different schemes could be used to establish a base for a comprehensive protein species/proteoform inventory.

2. Top-Down mass spectrometry based proteomics

Bottom-Up mass-spectrometry, especially targeted mass-spectrometry (MRM), offers unrivaled utility for sensitive and accurate detection of target proteins. However, the protein complexity the typical MRM assays cannot overpower and may fail because a same signature peptide may actually be present in different proteoforms.

[19,20]. To be unambiguously identified, proteoforms should be analysed with 100% sequence coverage [21]. So far this task has appeared surreal in proteomics [21]. A strategy of top down mass spectrometry combines separation of proteins with individual analysis. Not long ago, Top-Down mass spectrometry allowed the confident identification of only the most abundant proteins [22]. However, now much deeper large-scale analysis is possible due to vast improvements [23–25]. Now Top-Down proteomics can be performed not only by Fourier-transform ion cyclotron resonance mass spectrometry but also on Orbitrap MS instruments [17,26–28]. A number of reviews have already been published about technical aspects and application of the field [16,17,25,29], but some points can be mentioned here. Various fragmentation modes (ETD, HCD, UVPD, ETDhcD) have been exploited to maximize sequence coverage and increase the chance of detecting PTMs inside intact proteins [30]. Among them, higher energy collisional dissociation (HCD) is a well-characterized method that provides the best compromise between speed and the number of identified proteoforms [26]. Alternative fragmentation methods have been developed to increase sequence coverage and number of identified PTMs: ultraviolet photo dissociation (UVPD), electron transfer dissociation (ETD), hybrid electron transfer dissociation / higher energy collisional dissociation (EThcD) [26,31–34]. The complementary nature of HCD and UVPD fragmentation for more extensive protein profiling and proteoform characterization was shown using an Orbitrap Fusion Lumos platform [30]. Usage of different fragmentation modes allows direct quantification of proteoforms [18], and determination of the order of multiple modifications [35,36]. For instance, electron transfer dissociation (ETD) fragmentation allows the detection of twice as many ubiquitination sites compared to a traditional higher-energy collision dissociation (HCD) method [37]. It is interesting that the EThcD method can give almost the same information as HCD and ETD together. It appears that EThcD is especially advantageous for *de novo* sequencing and detection of PTMs, phosphorylation, and N-glycosylation, for instance [38,39]. Another challenge for top-down mass-spectrometry is data analysis and the application of tools for protein database searching. The data generated by mass spectrometers should be adequately processed to completely characterize proteoforms. Now, there is a panel of software tools available for top-down proteomics (PIITA, MASH Suite, MS-Align+, MS-Deconv, ProSight PTM 2.0, ProteinGoggle and others). ProSight PTM was the first one designed for the identification of intact proteins [40,41]. Also BIG Mascot or MascotTD that utilizes the popular Bottom Up software platform, Mascot, was applied for top-down analysis by doing some adjustments (extending the precursor mass cutoff to 110 kDa, for instance, as a precursor ion limit with this cutoff is much higher than in case of bottom-up analysis) [42]. The size of the precursor ion is only one of many issues that need to be considered in Top-Down software. But as top-down proteomics continues to develop, undoubtedly better software will be designed [43]. What is more, a Consortium for Top-Down Proteomics (CTDP) has been formed http://www.topdownproteomics.org, with the mission 'to promote innovative research, collaboration and education accelerating the comprehensive analysis of intact proteins' [43,44]. CTDP launched a proteoform database "Proteoform Atlas" and is working now to develop rules for the method (nomenclature), to use for describing fully characterized protein species/proteoforms. This nomenclature would be a valuable part of a proteoform inventory https://topdownproteomics.github.io/ProteoformNomenclatureStandard/

2.1. Separation techniques for sample preparation

Successful implementation of a Top Down approach requires the use of one or more steps of separation prior to mass spectrometric analysis [16,45,46]. Some separations can be used on-line with a mass spectrometer, but many can be applied off-line only. Examples of on-line separations are chromatography (HPLC) and capillary electrophoresis. HPLC is a robust and routine technique used for bottom-up and top-down mass spectrometry. Capillary electrophoresis (CE) followed by top-down MS is also a powerful technique to fully characterize protein isoforms and combinatorial post-translational modifications (PTMs) [47,48]. All these on-line separations allow increased throughput and substantially reduce sample handling but have limitations to sample loading, data acquisition, and separation conditions. Accordingly, off-line separation(s) look very attractive, even considering increased time needs to collect and treat fractions. But this approach is more flexible and allows the use of more diverse and multiple separations. The workflow for high-throughput Top-Down proteomics used in the Kelleher group is an example of such an approach [18,28]. Here, a whole cell extract goes through two different off-line separation steps and one on-line prior to analysis by mass spectrometry. The off-line separation steps are solution isoelectric focusing and a size-based separation (Gel Elution Liquid Fraction Entrapment Electrophoresis, GELFREE). The fractions are cleaned to remove SDS and go to reversed phase liquid chromatography prior to MS analysis http://www.kelleher.northwestern.edu/research/top-down-proteomics/. Finally, protein identification and characterization are completed with ProSight [49]. What is important, as you may notice, this workflow actually is a remake of the well-known two dimensional gel electrophoresis (2-DE) [50].

3. Two dimensional gel electrophoresis (2-DE)

2-DE was a driving force in the appearance of the term "proteome" [51]. For a long time, 2-DE was a workhorse of proteomics [50,52–56]. It is interesting that despite the domination of mass spectrometry, this approach still remains in high demand to-day. Among protein separation methods, a high resolution 2-DE is the best and can be considered not just as a separation technique but as an analytical method, as it is the technology with the highest separative power for proteins being made with engineered pH gradients. What is more, the protein parameters that can be measured by 2-DE are isoelectric point (pI) and molecular weight (Mw). These parameters are specific to every

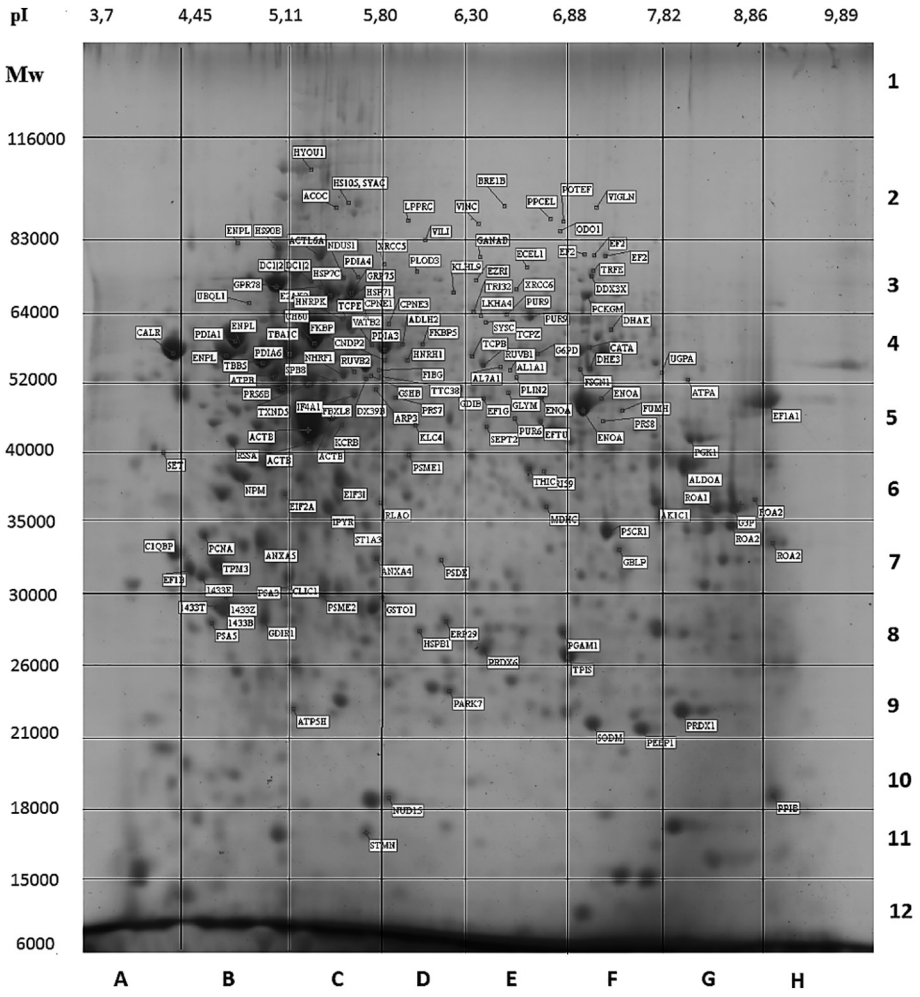

Fig. 1. 2DE map of HepG2 proteins. Spots with proteins identified by MALDI TOF-MS are annotated. Gel image analysis was performed by the program ImageMaster 2D Platinum 7.0 (GE Healthcare). Sections in the 2D gel selected for following LC ESI-MS/MS analysis are shown. Adapted from [69].

chemical signature representing a protein molecule. No wonder that not only the term “proteome” but also “protein species” originated from 2-DE [4,51,57]. Many comprehensive reviews describing strong as well as weak points of 2-DE were published [58–61]. Though not complete proteome can be analysed by 2-DE (some hydrophobic proteoforms and proteoforms with extremely acidic or basic pI can be missing), a database for proteins based on 2-DE (World-2DPAGE Repository) http://world-2dpage.expasy.org/repository was developed. The data used in this database were derived mostly from its combination with mass spectrometry and immunodetection. The data representation is based on visual properties of 2-DE gels. An example of such a protein inventory is shown in Fig. 1. What is more, when using 2-DE with subsequent immunostaining (2-DE Western-blot), it is possible to obtain profiles of specific proteoforms [62–64]. Yet, it should be noted that Western-blotting with a single antibody may not detect all proteoforms of a protein. It depends on antigen specificity of the antibody, as not all proteoforms of the same protein can contain this antigen. Additional application of a panel of more specific antibodies (phosphorylation, acetylation, glycosylation…) or *in vivo* isotope labelling can give more detailed information about these proteoforms [65–68]. Such a profile of PCNA proteoforms is shown in Fig. 2. It should be also kept in mind that these profiles can be ambiguous, as pI/Mw parameters for some proteoforms can coincide. For instance, a same single modification (phosphorylation, acetylation…) but in different sites, will produce different proteoforms having the same pI/Mw. In a classical 2-DE-based approach in proteomics, after 2-DE separation and protein staining, the spots are cut and analysed by mass spectrometry (usually by MALDI TOF MS using peptide mass fingerprinting, PMF). This method can be called “a spot-based approach”.

PMF algorithms assume that there is only one protein in a spot and the peptides are derived from this protein only [70]. Accordingly, by PMF usually only one protein in a spot can be identified. If only a single protein presents in a single spot, this technique is very efficient. This opinion has been the prevailing one since the invention of 2-DE [50]. In accordance with this opinion, comparative spot analysis of 2-DE gels is performed, when it is desired to compare protein levels between different samples. But recently, using MS/MS techniques, it was revealed that a single 2-DE spot can contain several proteins [72,73]. In this case, ESI LC-MS/MS has evident advantages over MALDI TOF MS and even MALDI TOF MS/MS, as it can perform a qualitative and quantitative analysis of each spot in a single run. An example of this approach is shown in Fig. 1. Here, spots are marked according to protein identification by MALDI TOF MS. In addition to these major proteins, subsequent analysis by ESI LC-MS/MS revealed the presence of less abundant proteins in many spots. Also, the same proteins were found in different spots (different proteoforms). Several examples of such proteins (ENOA, GRP75, and G3P) are shown in Fig. 3 (left). There are several weak points in this “spot-based” approach. One of them is that it completely depends on staining sensitivity and only stained spots can be analysed. In the case of sophisticated samples containing thousands of proteoforms, such as mammalian cell extracts, a lot of information is missing even when sensitive fluorescent dyes are used. Even further improvements cannot change the situation drastically here [74,75]. To solve this problem, a sectional analysis of 2-DE gels, or so called “pixel-

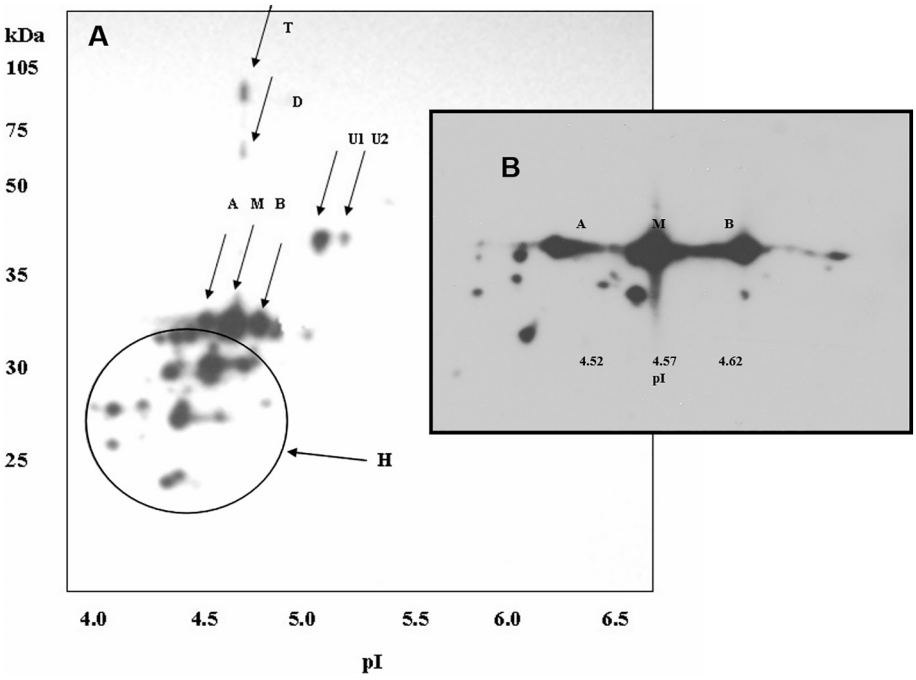

Fig. 2. Multiple spots of PCNA (proteoforms) can be detected by 2-DE. Immunostaining using PCNA specific antibody PC10 after 2-DE separation and transfer to Immobilon-P membrane. A, M, B, T and D denote acidic, main, basic, probable trimer and probable dimer forms of PCNA respectively. U1 and U2 are ubiquitinated forms of PCNA monomer. H is a cluster of PCNA hydrolysis products, produced by proteasome action [71]. (A) 2-DE (pH 4–7, 13 cm – first dimension, 10% SDS-PAGE, 13 cm – second dimension). (B) 2-DE (pH 4–5, 18 cm – first dimension, 10% SDS-PAGE, 13 cm – second dimension). Adapted from [63].

based approach" was introduced [76,77]. As you can see in Fig. 1, a whole gel (like the one used for "spot-based analysis") can be cut into sections (marked from A to H in pI direction and from 1 to 12 in Mw direction). Each section is analysed by ESI LC-MS/MS and the distribution of proteoforms around the gel can be revealed (Fig. 1). All proteoforms that are separated and located inside the 2-DE gel should be identified using this approach. The final detection of these proteoforms depends ultimately on the sensitivity of mass spectrometry. For instance, in the case of analysis of HepG2 cells, 20,462 proteoforms were identified, encoded by 3774 genes [69]. For glioblastoma cells – 16,012 proteoforms, encoded by 4050 genes were found [77]. What is more, this approach allows the generation of a "panoramic" view (3D-graphs or pictures) of proteoforms corresponding to each of these genes (so-called "one gene proteomes"). Examples of these pictures are shown in Fig. 3 (right). Note that even with decreased resolution, many more proteoforms (peaks) compared to Fig. 3 (left) are revealed for the same protein. Such a visualization of proteoforms is just a first step, the next most important step is identification of PTMs that are responsible for each proteoform. This information can be extracted from MS/MS spectra, considering application of fragmentation modes that are compatible with PTM detection and database search including as many PTMs as possible [35,36,38,39,78]. As this workflow is very similar to the workflow used by the Kelleher Research Group, it makes sense to compare them (Table 1). According to this comparison we can make a conclusion that these approaches are actually complementary.

If we can find an approach that allows us to combine the data produced by these approaches it will give us a chance to organize all this information in a convenient way.

Another approach for data representation is called a semi-virtual two dimensional gel electrophoresis or "strip-based approach" [79]. In this case, only one direction of 2-DE is used as a separation step (isoelectric focusing, IF). Mass-spectrometry works here as a second dimension of 2-DE generating information about the Mw of proteins after their identification using database searching. The final virtual 2D picture can be produced using Excel (Fig. 4).

In summary, 2-DE methodology (experimental, virtual and semi-virtual) can be a good complement to other separative techniques to implement the proteoforms atlas. This is possible because the 2-DE methodology is based on two basic proteoform parameters, pI and Mw, that can be measured (experimental approach) and calculated (virtual approach).

4. Perspectives

It seems obvious that information about intact proteoforms (Top-Down proteomics) should be obtained to identify variation in the human proteome. However, the complete and comprehensive Top-Down profiling of the human proteome is still beyond our reach. The continuing evolution of mass spectrometry technology and usage of it in combination with optimum protein separation techniques will finally allow us to get an image of the whole human proteome. A combination of methods for separation of proteins (2-DE) with bottom-up mass spectrometry (shot-gun analysis of peptides by ESI LC-MS/MS) is an efficient approach for increasing the productivity of tandem mass-spectrometry. Additionally, this combination of Top-Down and Bottom-up approaches together with immunological identification after 2-DE allows very convenient visual representation (profiling) of information about diverse proteoforms. This way will permit the development of a knowledge base for an inventory of all human protein species/proteoforms that is visually attractive, clear, easy to search and perceptive. The development of such an inventory will build on existing databases like http://world-2dpage.expasy.org/, https://www.nextprot.org/, http://www.uniprot.org/, http://atlas.topdownproteomics.org, which are of crucial importance in the establishment of an inventory of all human protein species/proteoforms.

Conflict of interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The work was performed in the framework of the Program for Basic Research of State Academies of Sciences for 2013–2020. Mass-spectrometry measurements were performed using the equipment of "Human Proteome" Core Facilities of the Institute of Biomedical Chemistry (Russia) which is supported by Ministry of Education and Science of the Russian Federation (agreement 14.621.21.0017, unique project ID RFMEFI62117X0017). Carita Lanner is acknowledged for

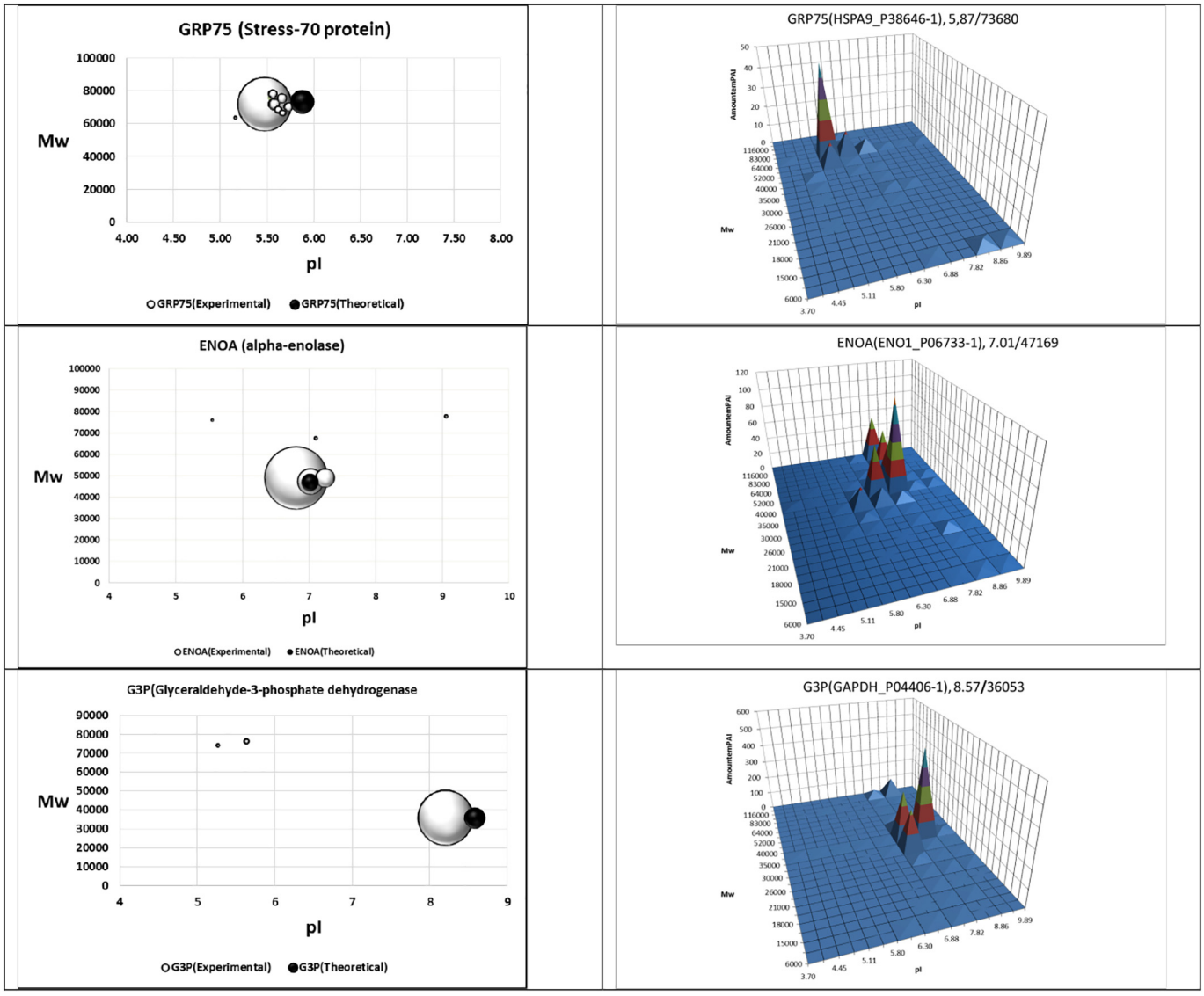

Fig. 3. Proteoforms identified after 2-DE separation and following ESI LC-MS/MS analysis. Detection was performed in spots (left) or sections (right). Proteoform abundance (emPAI) is expressed as a ball size or a peak height.

Table 1
Comparison of panoramic proteoform analysis by the Kelleher Research Group and by 2-DE sectional approach.

Item	Kelleher Research Group workflow http://www.kelleher.northwestern.edu/research/top-down-proteomics/	2-DE sectional analysis [67,68]
Number of fractions in pI direction (3 – 11)	8	8 and up
Number of fractions in Mw direction	12	12 and up
Mass range	10–100 kDa 35–150 kDa 75–500 kDa (different cartridge kits)	5–500 kDa single run
Number of all samples	96	96 and up
Amount of protein in a run	Up to 5 mg	Up to 2 mg
Gel staining and handling	No	Yes
Specific digestion	No	Yes
Recovery	> 80%	~100%
Collection of fractions	Manual	Manual
Sample treatment before ESI LC MS/MS	Yes	Yes
Hardware	LC MS/MS 12 T LTQ-FTMS, Orbitrap Q Exactive HF	Orbitrap Q Exactive
Software	ProSight PTM 2.0	2D Platinum, Mascot, SearchGui
Volume of information	~2000 proteoforms, ~700 genes	~20,000 proteoforms, ~4000 genes
Type of information	Type of PTMs	pI/Mw parameters of proteoforms, partial information about PTMs

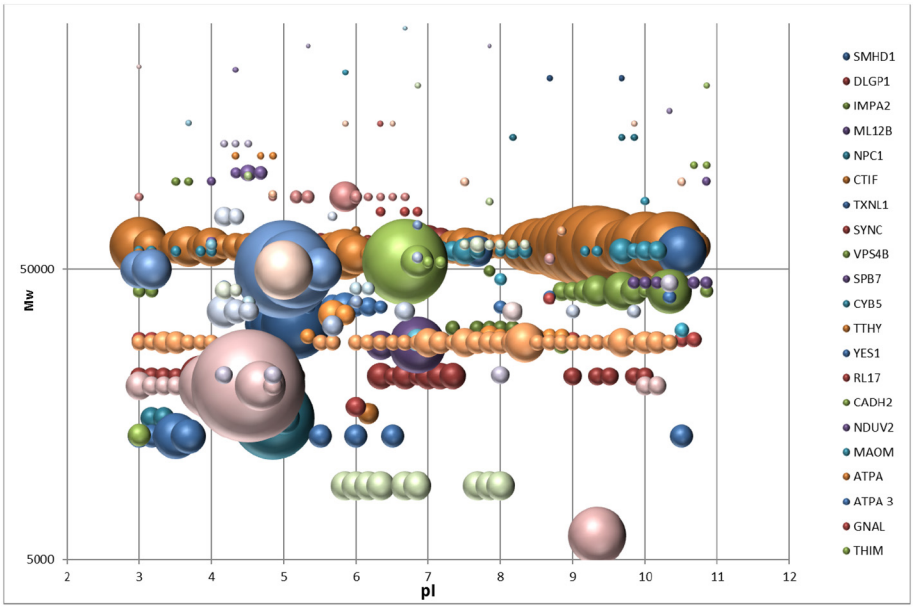

Fig. 4. 2-DE virtual representation of detected proteoforms of 18th chromosome in HepG2 cellular extract after IF and following ESI LC-MS/MS. A ball size is proportional to the proteoform abundance (emPAI).

editing assistance.

References

[1] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (2003) 198–207.
[2] L.M.F. de Godoy, J.V. Olsen, J. Cox, M.L. Nielsen, N.C. Hubner, F. Fröhlich, T.C. Walther, M. Mann, Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast, Nature 455 (2008) 1251–1254.
[3] H. Schlüter, R. Apweiler, H.-G. Holzhütter, P.R. Jungblut, Finding one's way in proteomics: a protein species nomenclature, Chem. Cent. J. 3 (2009) 11.
[4] P. Jungblut, B. Thiede, U. Zimny-Arndt, E.C. Müller, C. Scheler, B. Wittmann-Liebold, A. Otto, Resolution power of two-dimensional electrophoresis and identification of proteins from gels, Electrophoresis (1996) 839–847.
[5] L.M. Smith, N.L. Kelleher, Proteoform: a single term describing protein complexity, Nat. Methods 10 (2013) 186–187.
[6] L.M. Smith, N.L. Kelleher, Proteoforms as the next proteomics currency, Science 359 (2018) 1106–1107.
[7] P.R. Jungblut, H.G. Holzhütter, R. Apweiler, H. Schlüter, The speciation of the proteome, Chem. Cent. J. 2 (2008) 16.
[8] V.N. Uversky, p53 proteoforms and intrinsic disorder: an illustration of the protein structure–function continuum concept, Int. J. Mol. Sci. 17 (2016).
[9] N.L. Kelleher, A cell-based approach to the human proteome project, J. Am. Soc. Mass Spectrom. 23 (2012) 1617–1624.
[10] P.R. Jungblut, The proteomics quantification dilemma, J. Proteome 107 (2014) 98–102.
[11] S.N. Naryzhny, A.V. Lisitsa, V.G. Zgoda, E.A. Ponomarenko, A.I. Archakov, 2DE-based approach for estimation of number of protein species in a cell, Electrophoresis 35 (2014) 895–900.
[12] S.N. Naryzhny, V.G. Zgoda, M.A. Maynskova, N.L. Ronzhina, N.V. Belyakova, O.K. Legina, A.I. Archakov, Experimental estimation of proteome size for cells and human plasma, Biomed. Khim. 61 (2015) 279–285.
[13] B.A. Garcia, J.J. Pesavento, C.A. Mizzen, N.L. Kelleher, Pervasive combinatorial modification of histone H3 in human cells, Nat. Methods 4 (2007) 487–489.
[14] J. Cox, M. Mann, Is proteomics the new genomics? Cell 130 (2007) 395–398.
[15] A. Lisitsa, S. Moshkovskii, A. Chernobrovkin, E. Ponomarenko, A. Archakov, Profiling proteoforms: promising follow-up of proteomics for biomarker discovery, Expert Rev. Proteomics. 11 (2014) 121–129.
[16] A.D. Catherman, O.S. Skinner, N.L. Kelleher, Top down proteomics: facts and perspectives, Biochem. Biophys. Res. Commun. 445 (2014) 683–693.
[17] T.K. Toby, L. Fornelli, N.L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms, Annu. Rev. Anal. Chem. 9 (2016) 499–519.
[18] K.R. Durbin, L. Fornelli, R.T. Fellers, P.F. Doubleday, M. Narita, N.L. Kelleher, Quantitation and identification of thousands of human proteoforms below 30 kDa, J. Proteome Res. 15 (2016) 976–982.
[19] J.P. Savaryn, A.D. Catherman, P.M. Thomas, M.M. Abecassis, N.L. Kelleher, The emergence of top-down proteomics in clinical research, Genome Med. 5 (2013) 53.
[20] R. Aebersold, J.N. Agar, I.J. Amster, M.S. Baker, C.R. Bertozzi, E.S. Boja, C.E. Costello, B.F. Cravatt, C. Fenselau, B.A. Garcia, Y. Ge, J. Gunawardena, R.C. Hendrickson, P.J. Hergenrother, C.G. Huber, A.R. Ivanov, O.N. Jensen, M.C. Jewett, N.L. Kelleher, L.L. Kiessling, N.J. Krogan, M.R. Larsen, J.A. Loo, R.R. Ogorzalek Loo, E. Lundberg, M.J. Maccoss, P. Mallick, V.K. Mootha, M. Mrksich, T.W. Muir, S.M. Patrie, J.J. Pesavento, S.J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S.A. Slavoff, L.M. Smith, M.P. Snyder, P.M. Thomas, M. Uhlén, J.E. Van Eyk, M. Vidal, D.R. Walt, F.M. White, E.R. Williams, T. Wohlschlager, V.H. Wysocki, N.A. Yates, N.L. Young, B. Zhang, How many human proteoforms are there? Nat. Chem. Biol. 14 (2018) 206–214.
[21] B. Meyer, D.G. Papasotiriou, M. Karas, 100% protein sequence coverage: a modern form of surrealism in proteomics, Amino Acids 41 (2011) 291–310.
[22] B.A. Parks, L. Jiang, P.M. Thomas, C.D. Wenger, M.J. Roth, M.T. Boyne, P.V. Burke, K.E. Kwast, N.L. Kelleher, Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers, Anal. Chem. 79 (2007) 7984–7991.
[23] A.D. Catherman, K.R. Durbin, D.R. Ahlf, B.P. Early, R.T. Fellers, J.C. Tran, P.M. Thomas, N.L. Kelleher, Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence, Mol. Cell. Proteomics 12 (2013) 3465–3473.
[24] J.C. Tran, L. Zamdborg, D.R. Ahlf, J.E. Lee, A.D. Catherman, K.R. Durbin, J.D. Tipton, A. Vellaichamy, J.F. Kellie, M. Li, C. Wu, S.M.M. Sweet, B.P. Early, N. Siuti, R.D. LeDuc, P.D. Compton, P.M. Thomas, N.L. Kelleher, Mapping intact protein isoforms in discovery mode using top-down proteomics, Nature 480 (2011) 254–258.
[25] B. Chen, K.A. Brown, Z. Lin, Y. Ge, Top-down proteomics: ready for prime time? Anal. Chem. 90 (2018) 110–127.
[26] D.R. Ahlf, P.D. Compton, J.C. Tran, B.P. Early, P.M. Thomas, N.L. Kelleher, Evaluation of the compact high-field orbitrap for top-down proteomics of human cells, J. Proteome Res. 11 (2012) 4308–4314.
[27] Z. Tian, N. Tolić, R. Zhao, R.J. Moore, S.M. Hengel, E.W. Robinson, D.L. Stenoien, S. Wu, R.D. Smith, L. Paša-Tolić, Enhanced top-down characterization of histone post-translational modifications, Genome Biol. 13 (2012) R86.
[28] L. Fornelli, K.R. Durbin, R.T. Fellers, B.P. Early, J.B. Greer, R.D. LeDuc, P.D. Compton, N.L. Kelleher, Advancing top-down analysis of the human proteome using a benchtop quadrupole-orbitrap mass spectrometer, J. Proteome Res. 16 (2017) 609–618.
[29] L. Fornelli, T.K. Toby, L.F. Schachner, P.F. Doubleday, K. Srzentic, C.J. DeHart, N.L. Kelleher, Top-down proteomics: where we are, where we are going? J. Proteome 175 (2016) 3–4.
[30] T.P. Cleland, C.J. Dehart, R.T. Fellers, A.J. Vannispen, J.B. Greer, R.D. LeDuc, W.R. Parker, P.M. Thomas, N.L. Kelleher, J.S. Brodbelt, High-throughput analysis of intact human proteins using UVPD and HCD on an orbitrap mass spectrometer, J. Proteome Res. 16 (2017) 2072–2079.
[31] A. Michalski, N. Neuhauser, J. Cox, M. Mann, A systematic investigation into the nature of tryptic HCD spectra, J. Proteome Res. 11 (2012) 5479–5491.
[32] J.B. Shaw, W. Li, D.D. Holden, Y. Zhang, J. Griep-Raming, R.T. Fellers, B.P. Early, P.M. Thomas, N.L. Kelleher, J.S. Brodbelt, Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation, J. Am. Chem. Soc. 135 (2013) 12646–12651.
[33] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 9528–9533.
[34] A.M. Brunner, P. Lössl, F. Liu, R. Huguet, C. Mullen, M. Yamashita, V. Zabrouskov, A. Makarov, A.F.M. Altelaar, A.J.R. Heck, Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization,

Anal. Chem. 87 (2015) 4152–4158.
[35] H. Zhang, Y. Ge, Comprehensive analysis of protein modifications by top-down mass spectrometry, Circ. Cardiovasc. Genet. 4 (2011) 711.
[36] X. Han, A. Aslanian, J.R. Yates, Mass spectrometry for proteomics, Curr. Opin. Chem. Biol. 12 (2008) 483–490.
[37] T.R. Porras-Yakushi, M.J. Sweredoski, S. Hess, ETD outperforms CID and HCD in the analysis of the Ubiquitylated proteome, J. Am. Soc. Mass Spectrom. 26 (2015) 1580–1587.
[38] C.K. Frese, H. Zhou, T. Taus, A.F.M. Altelaar, K. Mechtler, A.J.R. Heck, S. Mohammed, Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD), J. Proteome Res. 12 (2013) 1520–1525.
[39] Y. Mechref, Use of CID/ETD mass spectrometry to analyze glycopeptides, Curr. Protoc. Protein Sci. 11 (2012) 1–11.
[40] G.K. Taylor, Y. Bin Kim, A.J. Forbes, F. Meng, R. McCarthy, N.L. Kelleher, Web and database software for identification of intact proteins using “top down” mass spectrometry, Anal. Chem. 75 (2003) 4081–4086.
[41] R.D. LeDuc, G.K. Taylor, Y. Bin Kim, T.E. Januszyk, L.H. Bynum, J.V. Sola, J.S. Garavelli, N.L. Kelleher, ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry, Nucleic Acids Res. 32 (2004).
[42] N.M. Karabacak, L. Li, A. Tiwari, L.J. Hayward, P. Hong, M.L. Easterling, J.N. Agar, Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry, Mol. Cell. Proteomics 8 (2009) 846–856.
[43] Z.R. Gregorich, Y. Ge, Top-down proteomics in health and disease: challenges and opportunities, Proteomics 14 (2014) 1195–1210.
[44] X. Dang, J. Scotcher, S. Wu, R.K. Chu, N. Tolić, I. Ntai, P.M. Thomas, R.T. Fellers, B.P. Early, Y. Zheng, K.R. Durbin, R.D. Leduc, J.J. Wolff, C.J. Thompson, J. Pan, J. Han, J.B. Shaw, J.P. Salisbury, M. Easterling, C.H. Borchers, J.S. Brodbelt, J.N. Agar, L. Paša-Tolić, N.L. Kelleher, N.L. Young, The first pilot project of the consortium for top-down proteomics: a status report, Proteomics 14 (2014) 1130–1140.
[45] J.F. Kellie, J.C. Tran, J.E. Lee, D.R. Ahlf, H.M. Thomas, I. Ntai, A.D. Catherman, K.R. Durbin, L. Zamdborg, A. Vellaichamy, P.M. Thomas, N.L. Kelleher, The emerging process of top down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput, Mol. BioSyst. 6 (2010) 1532.
[46] A.L. Capriotti, C. Cavaliere, P. Foglia, R. Samperi, A. Laganà, Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics, J. Chromatogr. A 1218 (2011) 8760–8776.
[47] X. Han, Y. Wang, A. Aslanian, M. Bern, M. Lavallée-Adam, J.R. Yates, Sheathless capillary electrophoresis-tandem mass spectrometry for top-down characterization of pyrococcus furiosus proteins on a proteome scale, Anal. Chem. 86 (2014) 11006–11012.
[48] K.N. Toby, Fornelli L. TK, Progress in top-down proteomics and the analysis of proteoforms, Annu Rev Anal Chem (Palo Alto, Calif) 9 (2016) 499–519.
[49] J.E. Lee, J.F. Kellie, J.C. Tran, J.D. Tipton, A.D. Catherman, H.M. Thomas, D.R. Ahlf, K.R. Durbin, A. Vellaichamy, I. Ntai, A.G. Marshall, N.L. Kelleher, A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome, J. Am. Soc. Mass Spectrom. 20 (2009) 2183–2191.
[50] P.H. O'Farrell, High resolution two-dimensional electrophoresis of proteins, J. Biol. Chem. 250 (1975) 4007–4021.
[51] M. Wilkins, Proteomics data mining, Expert Rev. Proteomics. 6 (2009) 599–603.
[52] J. Klose, Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals, Humangenetik 26 (1975) 231–243.
[53] J. Klose, From 2-D electrophoresis to proteomics, Electrophoresis 30 (2009) 142–149.
[54] G.A. Scheele, Two-dimensional gel analysis of soluble proteins. Charaterization of guinea pig exocrine pancreatic proteins, J. Biol. Chem. 250 (1975) 5375–5385.
[55] N.G. Anderson, N.L. Anderson, The human protein index, Clin. Chem. 28 (1982) 739–748.
[56] N.G. Anderson, A. Matheson, N.L. Anderson, Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology, Proteomics 1 (2001) 3–12.
[57] V.C. Wasinger, S.J. Cordwell, A. Cerpa-Poljak, J.X. Yan, A.A. Gooley, M.R. Wilkins, M.W. Duncan, R. Harris, K.L. Williams, I. Humphery-Smith, Progress with gene-product mapping of the Mollicutes:Mycoplasma genitalium, Electrophoresis 16 (1995) 1090–1094.
[58] T. Rabilloud, M. Chevallet, S. Luche, C. Lelong, Two-dimensional gel electrophoresis in proteomics: past, present and future, J. Proteome 73 (2010) 2064–2077.
[59] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, Proc. Natl. Acad. Sci. 97 (2000) 9390–9395.
[60] P. Pomastowski, B. Buszewski, Two-dimensional gel electrophoresis in the light of new developments, Trends Anal. Chem. 53 (2014) 167–177.
[61] R. Westermeier, A. Görg, Two-dimensional electrophoresis in proteomics, Protein Purif. Princ. High Resolut. Methods, Appl, Third ed., 2011, pp. 411–439.
[62] J.C. Sanchez, P. Wirth, S. Jaccoud, R.D. Appel, C. Sarto, M.R. Wilkins, D.F. Hochstrasser, Simultaneous analysis of cyclin and oncogene expression using multiple monoclonal antibody immunoblots, Electrophoresis 18 (1997) 638–641.
[63] S.N. Naryzhny, Proliferating cell nuclear antigen: a proteomics view, Cell. Mol. Life Sci. C 65 (2008) 3789–3808.
[64] S.N. Naryzhny, Blue dry western: simple, economic, informative, and fast way of immunodetection, Anal. Biochem. 392 (2009) 90–95.
[65] K. Mayer, S. Albrecht, A. Schaller, Targeted analysis of protein phosphorylation by 2D electrophoresis, Plant Phosphoproteomics Methods Protoc, 2015, pp. 167–176.
[66] S.N. Naryzhny, H. Lee, The post-translational modifications of proliferating cell nuclear antigen: acetylation, not phosphorylation, plays an important role in the regulation of its function, J. Biol. Chem. 279 (2004) 20194–20199.
[67] D. Montes-Sánchez, J.L. Ventura, I. Mitre, S. Frías, L. Michán, A. Espejel-Nuñez, F. Vadillo-Ortega, A. Zentella, Glycosylated VCAM-1 isoforms revealed in 2D western blots of HUVECs treated with tumoral soluble factors of breast cancer cells, BMC Chem. Biol. 9 (2009) 7.
[68] K. Kusch, M. Uecker, T. Liepold, W. Möbius, C. Hoffmann, H. Neumann, H. Werner, O. Jahn, Partial immunoblotting of 2D-gels: a novel method to identify post-translationally modified proteins exemplified for the myelin Acetylome, Proteomes. 5 (2017) 3.
[69] S.N. Naryzhny, M.A. Maynskova, V.G. Zgoda, N.L. Ronzhina, O.A. Kleyst, I.V. Vakhrushev, A.I. Archakov, Virtual-experimental 2DE approach in chromosome-centric human proteome project, J. Proteome Res. 15 (2) (2016) 525–530.
[70] A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, H. Boucherie, M. Mann, Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 14440–14445.
[71] S.N. Naryzhny, H. Lee, Observation of multiple isoforms and specific proteolysis patterns of proliferating cell nuclear antigen in the context of cell cycle compartments and sample preparations, Proteomics 3 (2003) 930–936.
[72] H. Lim, J. Eng, J.R. Yates, S.L. Tollaksen, C.S. Giometti, J.F. Holden, M.W.W. Adams, C.I. Reich, G.J. Olsen, L.G. Hays, Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide mass mapping to ?? LC-ESI tandem mass spectrometry, J. Am. Soc. Mass Spectrom. 14 (2003) 957–970.
[73] B. Thiede, C.J. Koehler, M. Strozynski, a. Treumann, R. Stein, U. Zimny-Arndt, M. Schmid, P.R. Jungblut, Protein species high resolution quantitative proteomics of HeLa cells using SILAC-2-DE-nanoLC/LTQ-Orbitrap mass spectrometry, Mol. Cell. Proteomics 12 (2012) 529–538.
[74] V.J. Gauci, M.P. Padula, J.R. Coorssen, Coomassie blue staining for high sensitivity gel-based proteomics, J. Proteome 90 (2013) 96–106.
[75] M. Padula, I. Berry, M. O'Rourke, B. Raymond, J. Santos, S.P. Djordjevic, A comprehensive guide for performing sample preparation and top-down protein analysis, Proteomes 5 (2017) 11.
[76] S. Naryzhny, Towards the full realization of 2DE power, Proteomes 4 (2016) 33.
[77] S.N. Naryzhny, M.A. Maynskova, V.G. Zgoda, N.L. Ronzhina, S.E. Novikova, N.V. Belyakova, O.A. Kleyst, O.K. Legina, R.A. Pantina, Proteomic profiling of high-grade glioblastoma using virtual-experimental 2DE, J Proteomics Bioinform. 9 (2016) 158–165.
[78] A.M. Silva, R. Vitorino, M.R. Domingues, C.M. Spickett, P. Domingues, Post-translational modifications and mass spectrometry detection, Free Radic. Biol. Med. 65 (2013) 925–941.
[79] S. Naryzhny, V. Zgoda, A. Kopylov, E. Petrenko, A. Archakov, A semi-virtual two dimensional gel electrophoresis: IF–ESI LC-MS/MS, MethodsX 4 (2017) 260–264.