Journal of proteome research

Perspective

pubs.acs.org/jpr

Next Steps on in Silico 2DE Analyses of Chromosome 18 Proteoforms

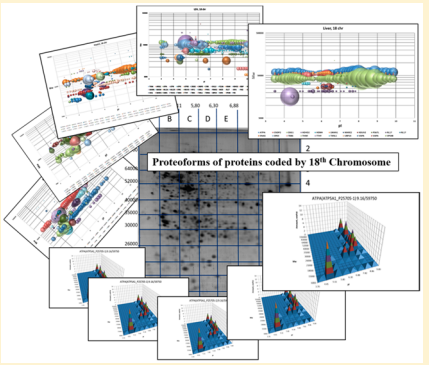
Stanislav N. Naryzhny,*[,†,‡] Elena S. Zorina,[†] Arthur T. Kopylov,[†] Victor G. Zgoda,[†] Olga A. Kleyst,[‡] and Alexander I. Archakov[†]

[†]Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Pogodinskaya 10, Moscow 119121, Russia

[‡]Petersburg Nuclear Physics Institute, National Research Center "Kurchatov Institute", Leningrad Region, Gatchina 188300, Russia

Supporting Information

ABSTRACT: In the boundaries of the chromosome-centric Human Proteome Project (c-HPP) to obtain information about proteoforms coded by chromosome 18, several cell lines (HepG2, glioblastoma, LEH), normal liver, and plasma were analyzed. In our study, we have been using proteoform separation by two-dimensional electrophoresis (2DE) (a sectional analysis) and a semivirtual 2DE with following shotgun mass spectrometry using LC-ESI-MS/MS. Previously, we published a first draft of this research, where only HepG2 cells were tested. Here, we present the next step using more detailed analysis and more samples. Altogether, confident (2 significant sequences minimum) information about proteoforms of 117 isoforms coded by 104 genes of chromosome 18 was obtained. The 3D-graphs showing distribution of different proteoforms from the same gene in the 2D map were generated. Additionally, a semivirtual 2DE approach has allowed for detecting more proteoforms and estimating their pI more precisely. Data are available via ProteomeXchange with identifier PXD010142.



KEYWORDS: *proteome, inventory, proteoforms, chromosome 18, two-dimensional electrophoresis, mass spectrometry, ESI LC−MS/MS, chromosome-centric*

INTRODUCTION

The final goal of the "Human Proteome Project" (HPP) is complete knowledge about all human proteins. Keeping in mind the huge volume of information that needs to be disclosed and efforts to be done, a special project was launched. This project is chromosome-centric (c-HPP) to promote more effective collaborations[1−3] and to create a comprehensive knowledge base of all human proteins for use as a tool in different areas such as education, science, or medicine that need this information. Russia is responsible for chromosome 18, which harbors 275 protein coding genes. Since these proteins are presented as multiple variants/proteoforms/protein species, we expect to detect at least 3000 proteoforms generated from chromosome 18. Usage of an interactive virtual 2DE map of proteins coded by genes of chromosome 18 in combination with experimental 2DE data will allow more effective execution of the Russian part of c-HPP. As protein extracts of normal as well as cancer cells are used in our study, information about proteoform distribution (proteome) will allow better understanding of protein changes connected to carcinogenesis. It will allow the identification of characteristic quantitative and qualitative differences between normal and cancer cells for use as biomarkers or targets for therapy. Here, we represent our data extracted from analysis of fibroblasts (LEH), HepG2, glioblastoma, human liver, and plasma.

EXPERIMENTAL SECTION

Chemicals and Materials

All reagents used were obtained from Sigma-Aldrich (St. Louis, MO, USA), unless another manufacturer is specified. The remaining reagents were obtained from the following companies: Thermo Fisher Scientific (Waltham, MA, USA): dithiothreitol (DTT), protease inhibitor cocktail, penicillin, streptomycin; GE Healthcare (Pittsburgh, PA, USA): IPG DryStrip (gel strips), IPG-buffers, DryStrip-coating liquid, Coomassie R350; Promega (Madison, WI, USA): Trypsin Gold; Bio-Rad (Hercules, CA, USA) molecular weight markers for protein electrophoresis; Biolot (Moscow, Russia): DMEM/F12 for cell growth, fetal calf serum, Trypsin-EDTA solution; Orange Scientific (Braine-l'Alleud, Belgium): Carrel culture flasks.

Cells

Human hepatocellular carcinoma (HepG2), lung embryonic fibroblasts (LEH), and glioblastoma cells were cultured in medium (DMEM/F12 supplemented with 10% fetal bovine serum (FBS) and 100 U/mL penicillin/streptomycin) under

Special Issue: Human Proteome Project 2018

Received: May 29, 2018
Published: September 21, 2018

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

standard conditions (5% CO_2, 37 °C). To prepare cell samples for protein extraction, the cells were detached with 0.25% Trypsin-EDTA solution, washed 3 times with PBS, and treated by lysis buffer.[4,5] Liver tissue samples of people who died as a result of an accident were purchased from the ILSBioBiobank within the framework of collaboration on the Chromosome-Centric Human Proteome Project (c-HPP). All ethical requirements have been met. Extraction was performed by lysis buffer after grinding the tissue in liquid nitrogen according to 2-DE protocol described in the literature.[6,7]

Sample Preparation and Two-Dimensional Electrophoresis (2DE)

Samples were prepared as described previously.[8,9] Cells (~10^7) containing 2 mg of protein were treated by 100 μL of lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 1% dithiothreitol (DTT), 2% (v/v) ampholytes, pH 3−10, protease inhibitor cocktail). The protein concentration in the sample was determined by the method of Bradford.[10] 2DE was performed using immobilized pH gradient (IPG)[11] for isoelectric focusing (IEF). Proteins were separated by IEF using Immobiline DryStrip 3−11 NL, 7 cm (GE Healthcare) following the manufacturer's protocol. The samples in the lysis buffer were mixed with rehydrating buffer (7 M urea, 2 M thiourea, 2% CHAPS, 0.3% DTT, 0.5% IPG (v/v) buffer, pH 3−11 NL, 0.001% bromophenol blue) in final volume of 125 μL (250−300 μg of protein)/strip. Strips (4 per experiment) were passively rehydrated for 4 h at 4 °C. IEF was performed on an IPGphor (GE Healthcare), which was programmed as follows: first step, 300 V, 1 h; second step, gradient to 1000 V, 1 h, third step; gradient to 5000 V, 1.5 h; fourth step, 5000 V, 1 h, temperature 20 °C and maintained at the voltage 500 V. After IEF, strips were soaked 10 min in the equilibration solution (50 mM Tris, pH 8.8, 6 M urea, 2% sodium dodecyl sulfate (SDS) and 30% (v/v) glycerol, 1% DTT). This process was followed by 10 min incubation in the equilibration solution containing iodoacetamide instead of DTT (50 mM Tris, pH 8.8, 6 M urea, 2% sodium dodecyl sulfate (SDS) and 30% (v/v) glycerol, 5% iodoacetamide). The strips were placed on the top of the 12% polyacrylamide gel of the second direction and sealed with a hot solution of 1 mL of 0.5% agarose prepared in electrode buffer (25 mM Tris, pH 8.3, 200 mM glycine, and 0.1% SDS) and electrophoresed to second direction under denaturing conditions using the system Ettan DALTsix (240 × 200 × 1 mm, GE Healthcare). Electrophoresis was carried out at room temperature at constant power 3.5 W per gel.[12,13] Gels were stained with Coomassie Blue R350, scanned by ImageScanner III (GE Healthcare) or Typhoon FLA 9500 (GE Healthcare), and analyzed using ImageMaster 2D Platinum 7.0 (GE Healthcare).

In the case of a semivirtual 2DE, the 24 cm IPG strip was cut into 48 equal sections using scissors, and each section was transferred to Eppendorf tube. For complete reduction, 300 μL of 3 mM DTT, 100 mM ammonium bicarbonate was added to each sample and incubated at 50 °C for 15 min. For alkylation, 20 μL of 100 mM iodoacetamide (IAM) were added to the same tube and incubated in the dark at room temperature for 15 min. For digestion, stock solution of trypsin (0.1 mg/mL) was diluted 1:10 by 25 mM ammonium bicarbonate, and 100 μL of diluted trypsin was added into each tube. Samples were incubated overnight for 4−24 h at 37 °C. Supernatants that may contain peptides that have diffused out of the gel slices were collected to new labeled 0.5 mL tubes. Peptides were extracted by adding 150 μL of 60% acetonitrile, 0.1% trifluoroacetic acid (TFA) to each tube containing gel slices. Extracts were dried in Speed Vac, reconstituted in 20 μL of 0.1% TFA, and analyzed by Orbitrap Q-Exactive Plus mass spectrometer. Protein identification and relative quantification were performed using Mascot 2.4.1 (Matrix Science) and emPAI. A table with information about all detected protein proteoforms was built. All proteins detected in the same section were given the pI of this section. Accordingly, the same proteins detected in different sections were considered as different proteoforms.

ESI LC−MS/MS Analysis

All procedures were performed according to the protocol described previously.[13] The gel was divided into 96 sections with determined coordinates. Each section (~0.7 cm^2) was cut, shredded, and treated by trypsin according to the protocol for single spots identification with proportionally increased volumes of solutions. Tryptic peptides were eluted from the gel with extraction solution (5% (v/v) ACN, 5% (v/v) formic acid) and dried in a vacuum centrifuge. Peptides were dissolved in 5% (v/v) formic acid. Tandem mass spectrometry analysis was carried out in duplicate on an Orbitrap Q-Exactive mass spectrometer (Thermo Scientific, USA). Mass spectra were acquired in positive ion mode. High resolution data was acquired in the Orbitrap analyzer with a resolution of 30 000 (m/z 400) for MS and 7500 (m/z 400) for MS/MS scans.

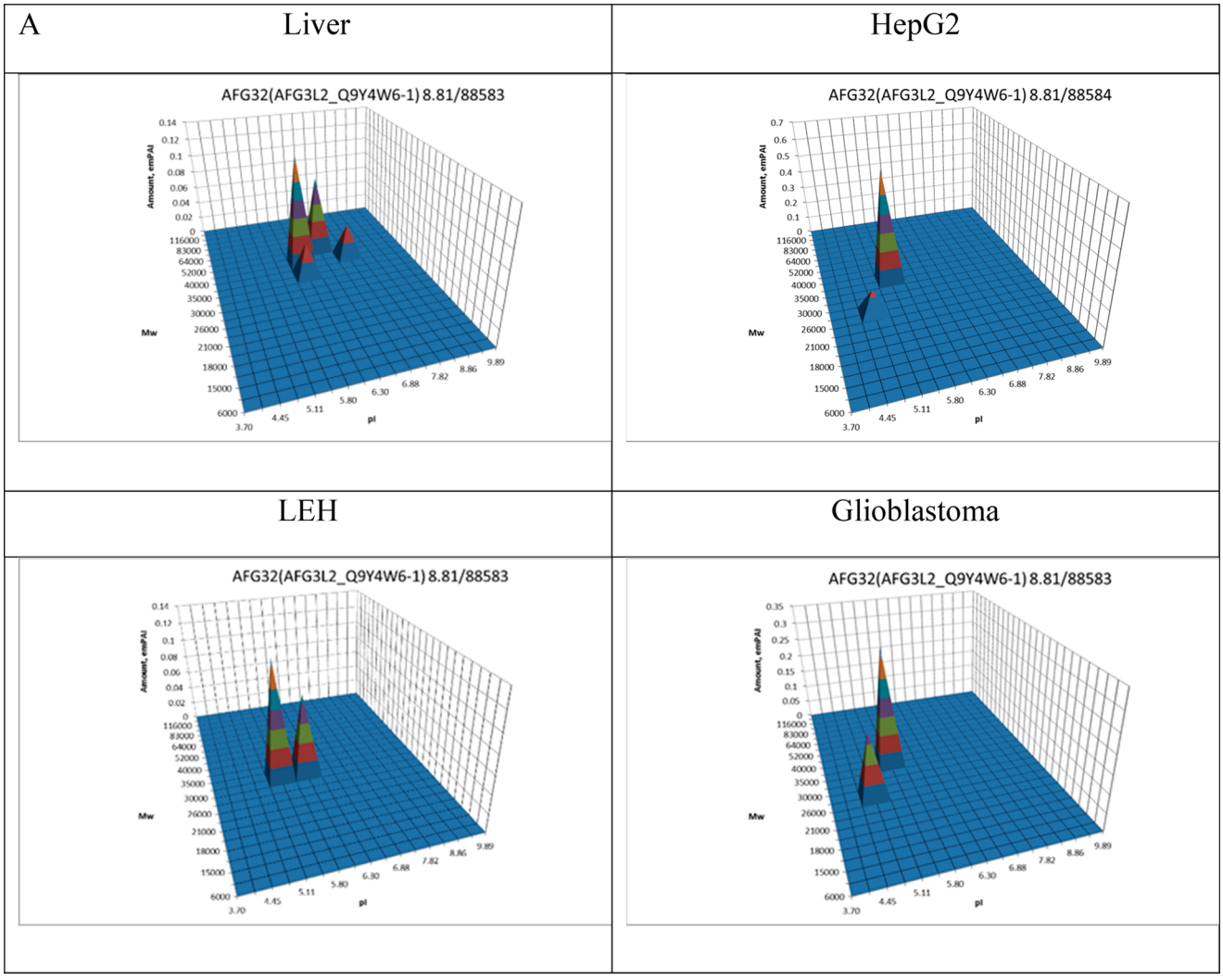
Identification of proteins was performed using Mascot 2.4.1 (Matrix Science, London, UK) by searching UniProt/Swiss-Protein sequence database (October 2014, 20 196 total sequences). The following search parameters were used: trypsin as the cutting enzyme, mass tolerance for the monoisotopic peptide window was set to ±50 ppm, missed cleavages 1. Cysteine carbamidomethyl was chosen as a fixed and oxidized methionine as a variable modification. NeXtProt database was used as a protein sequence database. For FDR assessment, a separate decoy database was generated from the protein sequence database. False positive rate of 1% was allowed for protein identification. A minimum Mascot ion score of 30 was used for accepting peptide MS/MS spectra. Data were also searched, using the SearchGUI, an open-source graphical user interface.[14] Two unique peptides per protein were required for all protein identifications. Exponentially modified PAI (emPAI), the exponential form of protein abundance index (PAI) defined as the number of identified peptides divided by the number of theoretically observable tryptic peptides for each protein, was used to estimate protein abundance.[15]

RESULTS AND DISCUSSION

High Yield Proteome Analysis

We ran the same type of 2DE (7 cm × 8 cm) using different human samples (HepG2, LEH, liver, glioblastoma, plasma) (Figure S1). Each gel was stained and scanned, and the produced experimental 2DE maps were calibrated according to the position of the several known protein spots. Next, the gels were identically divided into 96 sections, identified as 1−12 along the M_w dimension and A−H along the pI dimension. All these gel sections were chopped and treated with trypsin according to protocol for mass spectrometry by LC-ESI-MS/MS. The tryptic peptides obtained from each 2DE section were analyzed using an Orbitrap Q-Exactive mass spectrometer. Finally, protein identification and relative quantification were performed using Mascot 2.4.1 or SearchGUI and Peptide Shaker. If the same protein (product of the same gene) was

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX



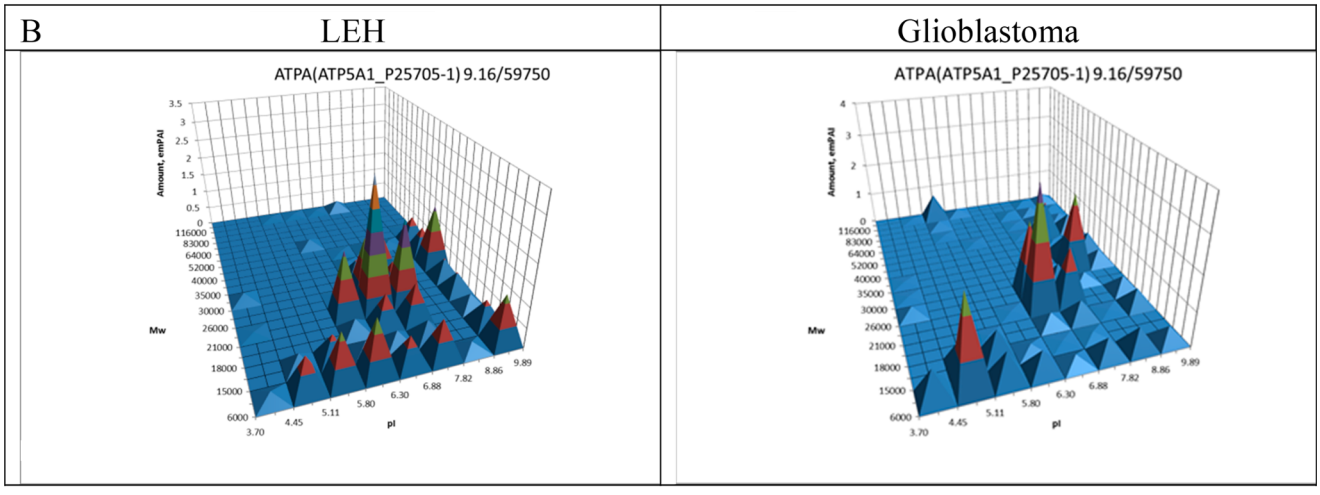


Figure 1. continued

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

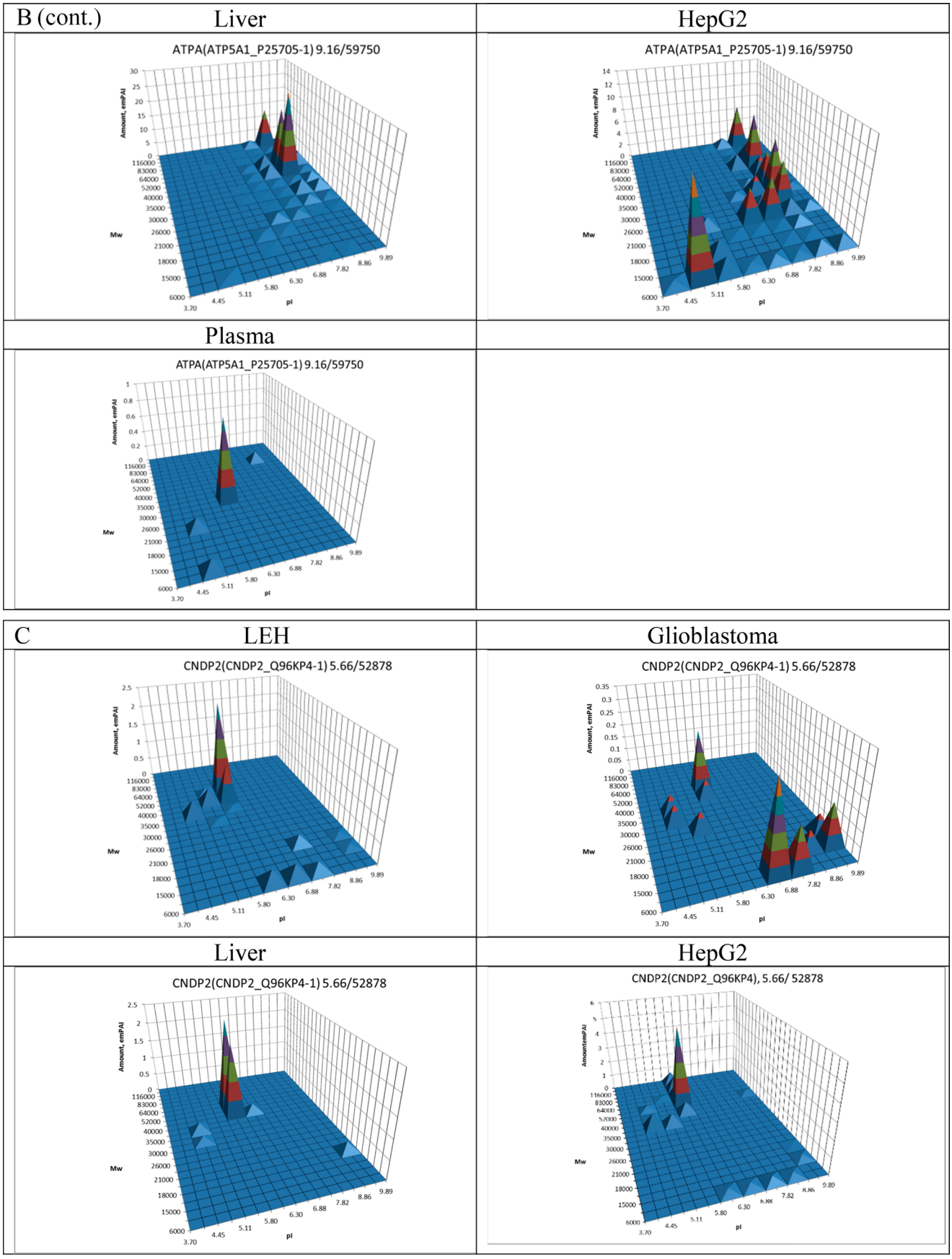


Figure 1. continued

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

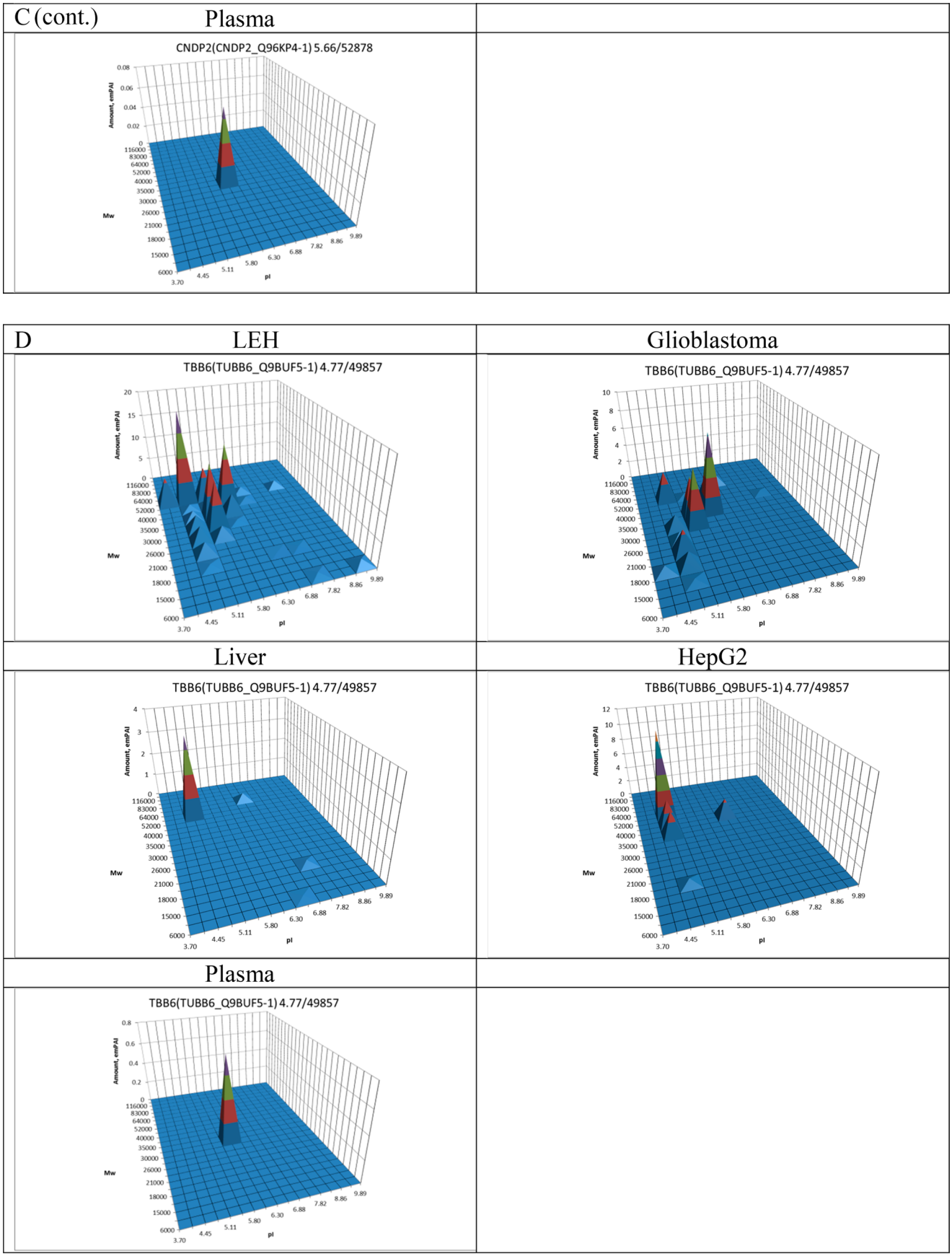
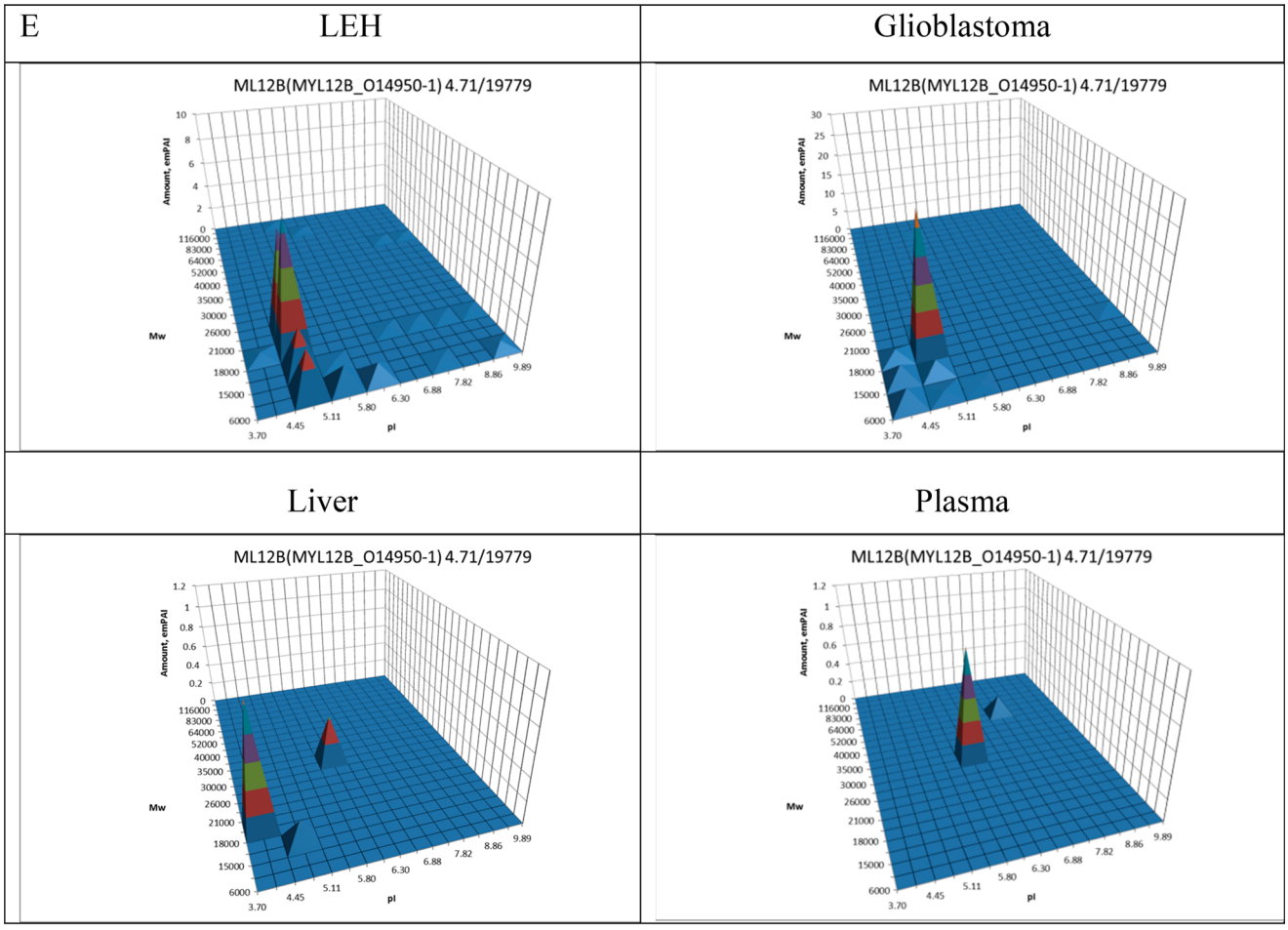


Figure 1. continued

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX



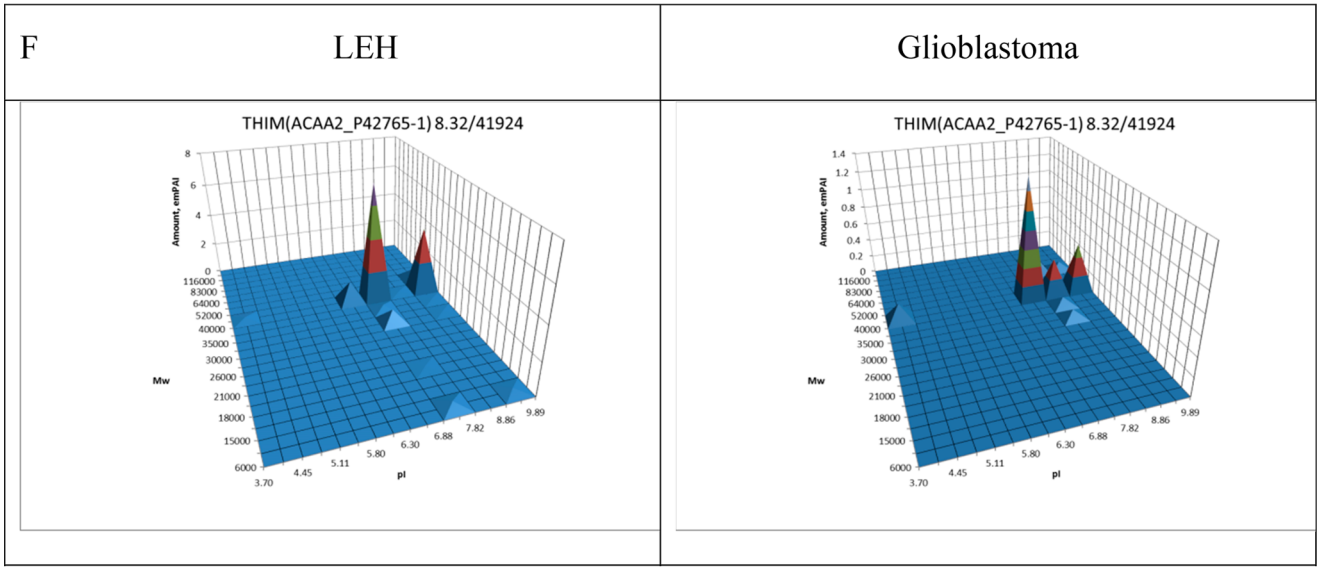


Figure 1. continued

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

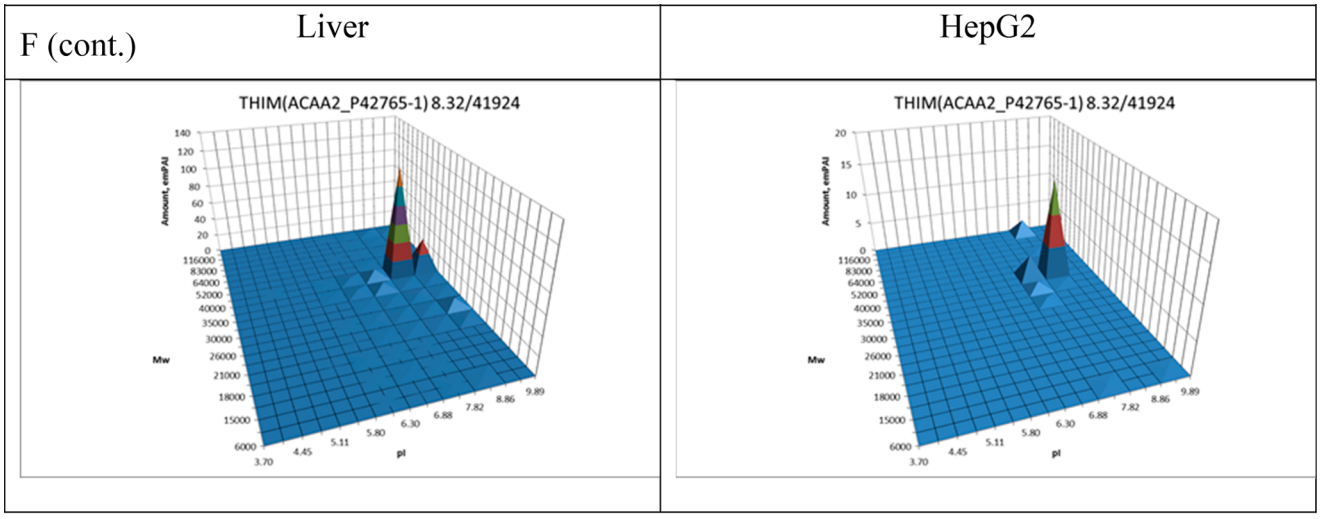


Figure 1. 3D graphs showing profiles of proteoforms of the same gene (isoform) in liver, HepG2, LEH, glioblastoma cells, or plasma. Proteins coded by chromosome 18 are shown. (A) AFG3L2 (Q9Y4W6−1), AFG3-like protein 2. (B) ATP5A1 (P25705−1), ATP synthase subunit alpha, mitochondrial. (C) CNDP2 (Q96KP4−1), cytosolic nonspecific dipeptidase. (D) TUBB6 (Q9BUF5−1), tubulin beta-6 chain. (E) MYL12B (O14950−1), myosin regulatory light chain 12B. (F) ACAA2 (P42765−1), 3-ketoacyl-CoA thiolase, mitochondrial isoform.

identified in different sections, it was considered to exist as different proteoforms. These forms have different pI/M_w parameters (coordinates) and are in different areas (spots) of the gel. After analysis of all sections and protein identification, a table of proteins coded by chromosome 18 and detected in extracts from several origins was generated (Table S1). Finally, the 3D graphs showing patterns of proteoforms (profiles) of the same isoform (unique UniProt number) coded by the same gene of 18th chromosome were generated for all samples analyzed (Figure S2). Some of these profiles are shown in Figure 1A−F. These graphical images give us a general representation of these proteoform profiles in different cells or plasma. The examples of genes presented are AFG3L2, ATP5A1, CNDP2, TUBB6, MYL12B, ACAA2. Using virtual-experimental approach we can actually try to explain these profiles and proteoforms.[13] In these graphs, a master protein (product of a nonmodified canonical sequence) is presented usually as a most abundant proteoform. There are some exceptions of the case, when polypeptide is processed into mature form by cleavage of the signal sequence, for instance. It happens with mitochondrial proteins that have a mitochondrial targeting signal (1−43 AA), which should be usually cleaved once a targeting is complete. In the case of chromosome 18, there are several such proteins: AFG3-like protein 2 (AFG32_HUMAN), ATP synthase subunit alpha (ATPA_HUMAN), ferrochelatase (HEMH_HUMAN), (NADH dehydrogenase [ubiquinone] flavoprotein 2 (NDUV2_HUMAN), 3-ketoacyl-CoA thiolase (THIM_HUMAN). Accordingly, in the case of mitochondrial proteins, main proteoforms observed are not master but mature forms. We often do not see a master protein in their profiles at all (AFG32, HEMH, NDUV2) (Figure 1A) or see it at a very low abundance (ATPA_HUMAN) (Figure 1B). There are some exceptions as well. For instance, mitochondrial transit peptide (1−16) in 3-ketoacyl-CoA thiolase (THIM_HUMAN) is supposed to be not cleaved. Accordingly, in liver and HepG2 cells, we see that its main form has the same parameters (pI/M_w, 8.32/42 000) as the master protein (Figure 1F). But in LEH and glioblastoma cells, the most abundant peak of THIM has more acidic pI, may be because of the signal peptide cleavage (Figure 1F). An alternative reason for this shift could be heavy acetylation or phosphorylation. In concordance with this assumption, we can find in Proteoform Atlas (http://atlas.topdownproteomics.org/) that it has already detected 3 sites of THIM acetylation[16]

The sectional analysis, which we are performing, actually has a low resolution (0.5−0.8 pH range per section), and single PTMs (like acetylation, phosphorylation...) that produce shift in the range of 0.05−0.5 pH unit/per PTM usually cannot be differentiated here.[17−19] But in the case of multiple PTMs, the corresponding proteoforms should be easily observed. Also, some single PTMs like ubiquitination (∼8.5 kDa, a single shift), SUMOylation (∼12 kDa, a single shift), glycosylation (up to several kDa shift)[20,21] should be easily come out. These could be reasons for the observation of proteoforms with M_w bigger than theoretically predicted ones, like for serpin B3 (SPB3_HUMAN) or serpin B4 (SPB4_HUMAN) in the case of liver and HepG2 (Figure S2).

Additionally, smaller proteoforms that are most likely proteolytic products are often observed (ATPA, CNDP2, DSC1) (Figure S2). One interesting example is AFG3-like protein 2 (AFG32_HUMAN) (Figure 1A). The theoretical parameters pI/M_w of its mature mitochondrial form are 8.01/81184. But in all observed samples (liver, HepG2, LEH, glioblastoma), we see only acidic half-size proteoforms (pH 5−7, M_w ∼40 000) (Figure 1A). The half-size proteoform with acidic pI ∼ 5.5 (C-terminus part) could be produced by cleavage of AFG32 by endopeptidase somewhere in the middle of the chain. But in this case, we should also see a N-terminus part with a basic pI as well (pI ∼ 9.4). As we do not see this polypeptide, it could be completely hydrolyzed by another proteinase. What is interesting, a very similar situation is also observed with mitochondrial ferrochelatase (HEMH_HUMAN). A mature form of this protein should have pI/M_w: 8.61/42186.67. This proteoform is observed only in liver cells, but in HepG2, LEH, and glioblastoma cells, we see only half-size proteoform with pI 5−6/M_w ∼ 21000. Taking together, we can assume that this could be a result of an unknown type of processing for these proteins. Complementing the theme of processing, very interesting and unexpected results were obtained with human plasma samples. Majority of proteins coded by chromosome 18

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

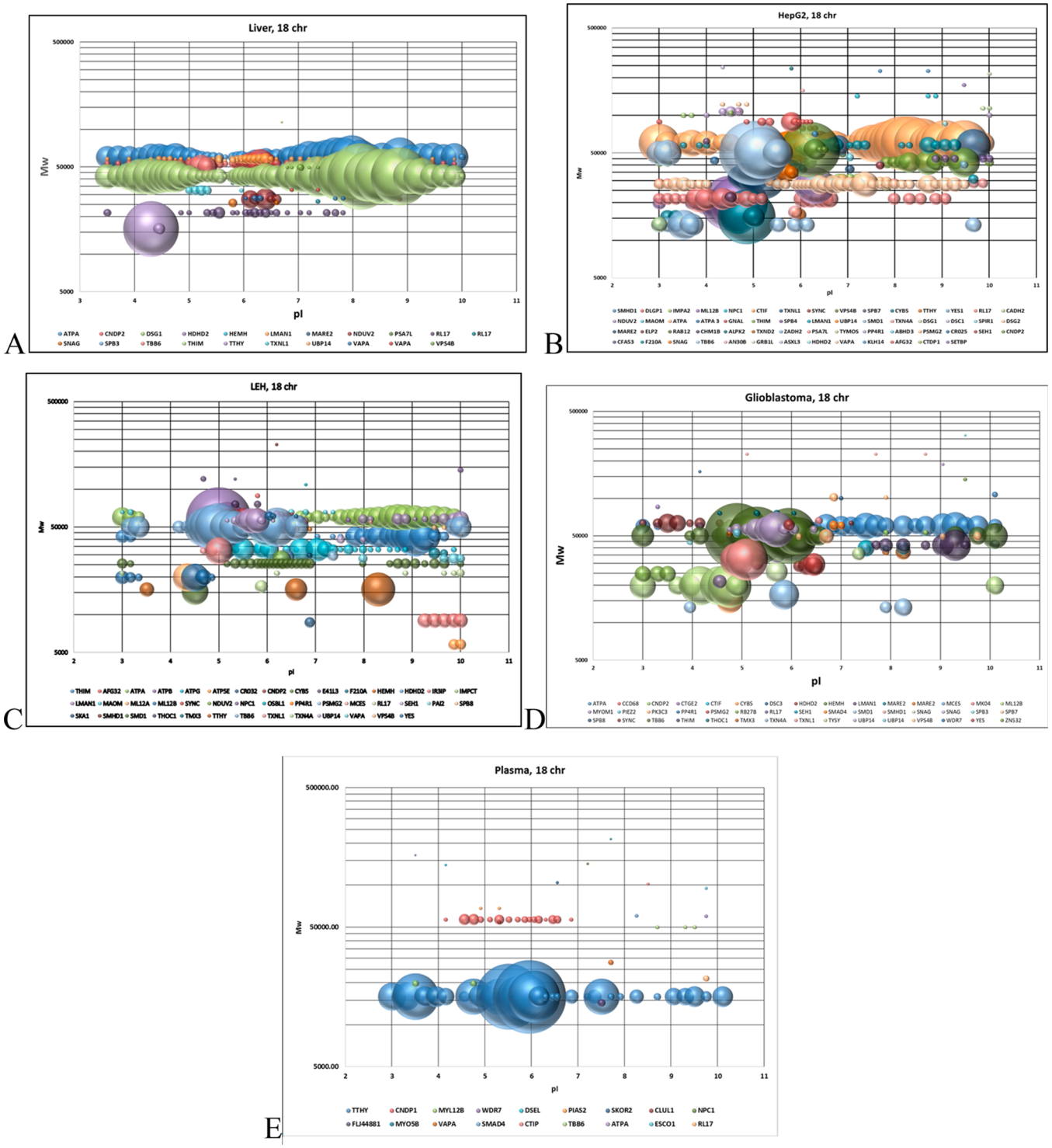


Figure 2. A semivirtual 2DE of proteins coded by chromosome 18. (A) Liver. (B) HepG2, adapted from ref 31. (C) LEH. (D) Glioblastoma. (E) Plasma.

are presented in plasma as proteoforms that are not master or mature forms (Figure S2). Their profiles in plasma are very different from the profiles in the cellular extracts (Figure 1B–E). Most likely, these proteins are not bearing any function in plasma and coming to the bloodstream as a byproduct from different cells. Some of them could be a product of proteolysis (smaller forms) but some (bigger forms) possibly are heavily modified (Figure S2). An example of a protein coded by chromosome 18 that is functional in plasma is beta-Al-His dipeptidase (CNDP1_HUMAN), also known as serum carnosinase. This enzyme is abundant in blood serum, where it plays an important role in hydrolysis of dietary carnosine.[22,23] Another protein, transthyretin (TTHY_HUMAN) is a serum and cerebrospinal fluid protein that transports holo-retinol-binding protein and thyroxine. Its serum concentration has been widely used to assess clinical nutritional status.[24]

We should keep in mind that there is a possibility that some of the observed truncated products can be produced during 2DE. 2DE is a long multistage procedure, and there is a chance of extra modification of proteins during it. Previously, we had

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

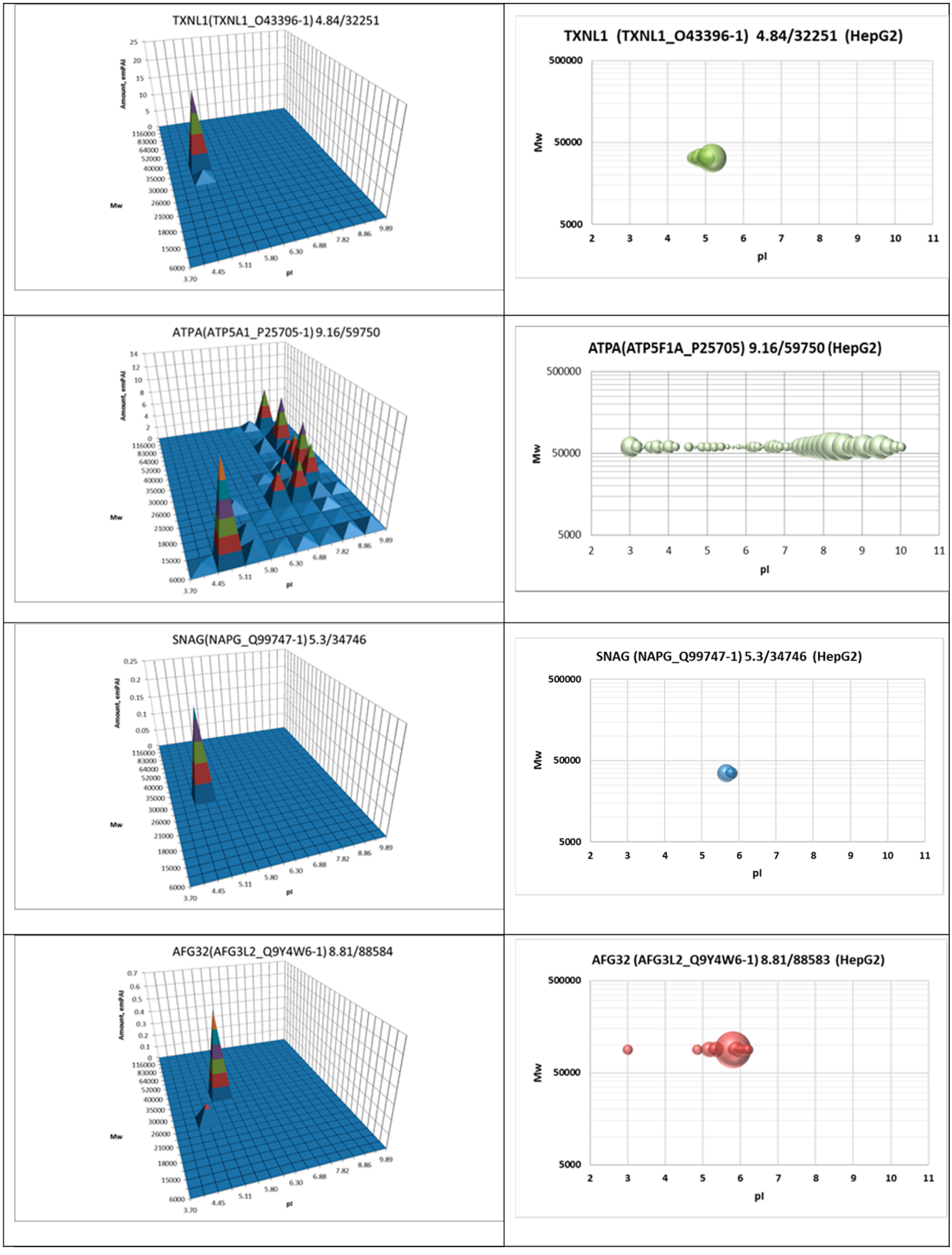


Figure 3. continued

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

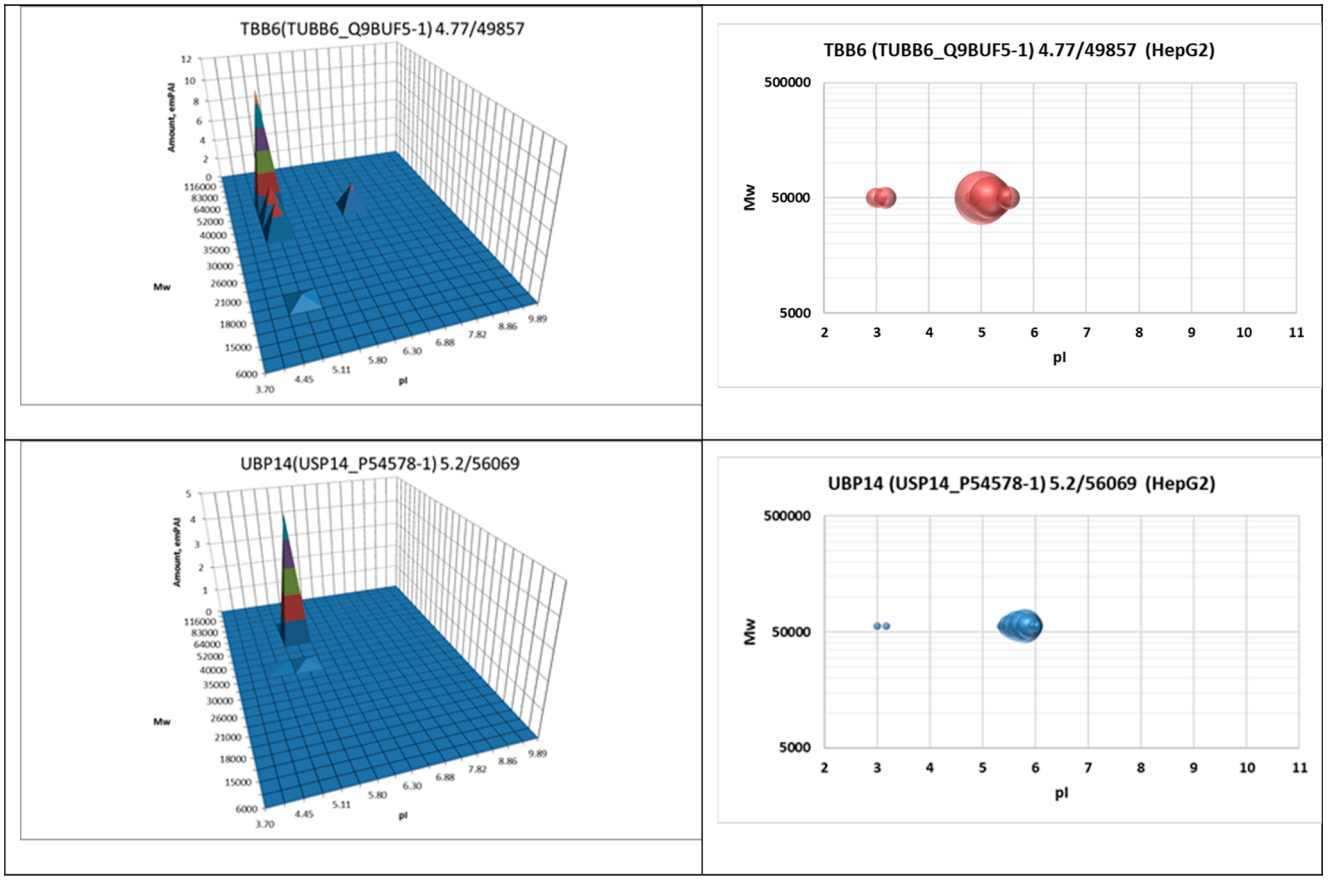


Figure 3. Comparison of a sectional analysis and a semivirtual 2DE for some proteins coded by chromosome 18. TXNL1 (thioredoxin-like protein), AFG32 (AFG3-like protein 2), SNAG (gamma-soluble NSF attachment protein), UBP14 (ubiquitin carboxyl-terminal hydrolase 14), TBB6 (tubulin beta-6 chain), ATPA (ATP synthase subunit alpha, mitochondrian).

been analyzing this situation on an example of PCNA (proliferation cell nuclear antigen).[25] We found that its proteasomal degradation can take place during 2DE (IEF) if inhibitors (thiourea or proteasomal inhibitor) are absent.[25,26] A positive side of this situation is that this proteasomal degradation happens in vivo as well and may be relevant to the physiological functions of proteins. Increased in vitro proteolysis that is detected using 2DE may be a kind of amplification of the proteolysis that happens in vivo.[25] Though we make every effort to exclude this situation using thiourea and protease inhibitors cocktail, we cannot guarantee that it is not completely happening in our experiments.[26] It seems that this is a general situation with 2DE, which has manifested itself after using sensitive large-scale mass spectrometry (ESI-LC-MS/MS).[27] As like with PCNA, a similar situation is observed with ATPA (Figure 1B). It looks like the observed truncated proteoforms of ATPA are produced by endopeptidase (proteasome). According to detected peptides, we can even reconstruct ATPA proteoforms located in different sections. For instance, ATPA proteoform located in section B12 highly likely is AA59−168 polypeptide (theoretical pI/M_w: 4.57/11309), in section C12, AA2−168 (pI/M_w: 5.64/17253), in section G7, N-terminus peptide AA24−317 (pI/M_w: 8.43/31495) and C-terminus peptide AA253−554 (pI/M_w: 7.93/32826), in section F9, N-terminus peptide AA46−260 (pI/M_w: 7.73/22999) and C-terminus peptide AA329- 554 (pI/M_w 7.05/24468).

To obtain more information about proteoforms, we also performed a semivirtual 2DE, where high resolution of IEF was implemented in a better degree than in the 2DE sectional analysis.[28] In the case of 24 cm strips with pH 3−11 (48 sections), a pH range of each section taken for analysis is ∼0.166 and ∼0.21 in the case of 18 cm strips (36 sections). Though semivirtual 2DE does not give the value of experimental M_w of detected proteoforms, it allows to detect more of them and measure their pI more precisely (Figure 2A−F, Table S2, Figure S3A−F). As a complement, it is improving a resolution of the previously performed sectional analysis. Let's see some examples. For instance, in the case of AFG3-like protein 2 (AFG32_HUMAN), by 2DE sectional analysis of HepG2 proteins only one major peak in the section with coordinates pI 5.11−5.80 and M_w 35 000−40 000 was detected. But using semivirtual approach, seven proteoforms with pI around 5.8 were detected (Figure 3). For thioredoxin-like protein 1 (TXNL1_HUMAN), four proteoforms instead of one peak around pI 5.0 were detected. In the case of tubulin beta-6 chain (TBB6_HUMAN), 4 proteoforms around pI 5.0 were detected. In the case of myosin regulatory light chain 12B (ML12B_HUMAN), 6 proteoforms around pI 4.7 were detected. For gamma-soluble NSF attachment protein isoform (SNAG_HUMAN), two proteoforms around pI 5.5 were detected, and for ubiquitin carboxyl-terminal hydrolase 14 isoform (UBP14_HUMAN), six proteoforms around pI 5.8 were detected (Figure 3). In the case of mitochondrial ATP synthase subunit alpha (ATPA_HUMAN), the situation is more complicated. A resolution of the semivirtual 2DE in pH direction is much higher (its limit is 46 fractions against only 8 fractions in the

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

semivirtual 2DE). But many proteoforms detected by the sectional analysis are truncated products of ATPA and have the similar pI (especially in the range 7.82−8.86) (Figure 3). Accordingly, they are not separated by the semivirtual 2DE and detected in the same fractions. This is the reason why the amounts of detected by both methods proteoforms (44 and 45) are similar. But the actual amount is higher. For instance, instead of only 2 peaks (M_w 6000−15 000) with acidic pH (3.7−5.11) detected by the sectional analysis, the semivirtual 2DE gives 13 proteoforms (Figure 3). It will rise the number of proteoforms at least to 56. This more detailed set of proteoforms makes it possible to more accurately attribute to them the information obtained about posttranslational modifications (PTMs) (Table S2).

In summary, if the sectional analysis and the semivirtual 2DE are used together, more detailed proteoform patterns can be obtained. First, information about proteoforms (emPAI and pI) is extracted from the semivirtual 2DE. Next, additional information about the experimental M_w of proteoforms can be added from the sectional analysis.

CONCLUSION

This is our next step in the creation of a 2DE-based knowledge database of proteins coded by chromosome 18 that was carried out using small 2DE gels (7 cm × 8 cm). Such a scale and division of the gels into 96 sections allowed to generate proteoform profiles coded not only by the 18th chromosome but by other chromosomes as well. A representation of proteomes based on theoretical and experimental parameters (pI/M_w) allowed to get a clearer view of the general state at the scale of complete proteomes and products of single genes. In particular, using this way of representation, we have shown a striking difference between specific profiles of proteoforms in cellular and plasma proteomes in terms of PTMs. Addition of a semivirtual 2DE allowed to produce more detailed profiles of proteoforms coded by each gene, and a graphical representation of the profiles gives a chance to get information about profiles in a convenient way. Also, parameters (pI/M_w) of detected proteoforms give us as a hint what kind of PTMs can produce these proteoforms. But the next steps should be done to disclose the fine details of proteoforms, including their PTMs. At this moment, we can give only some information about PTMs of proteins coded by chromosome 18 in HepG2 (Table S2). A deeper database search of our raw data for other samples is in progress. Available in Proteoform Atlas, information about already detected proteoforms also will be very helpful.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00386.

Figures S1 (adapted from ref 29) and S2 (PDF)
Table S1 (XLSX)
Table S2 (XLSX)

AUTHOR INFORMATION

Corresponding Author

*Tel: (+7) 9111764453. E-mail: snaryzhny@mail.ru.

ORCID

Stanislav N. Naryzhny: 0000-0002-4102-3423
Elena S. Zorina: 0000-0003-4456-6850

Notes

The authors declare no competing financial interest.
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[30] partner repository with the data set identifier PXD010142 and 10.6019/PXD010142.

ACKNOWLEDGMENTS

This work was funded by a grant of RSF (Russian Science Foundation) #14-25-00132. 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 (unique project ID RFMEFI62117X0017).

ABBREVIATIONS

2DE, two-dimensional electrophoresis; ESI LC−MS/MS, liquid chromatography-electrospray ionization-tandem mass spectrometry; HCD, higher energy collisional dissociation; ABC, ammonium bicarbonate; ACN, acetonitrile; PTM, post-translation modification; emPAI, exponential modified form of protein abundance index.

REFERENCES

(1) Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; et al. The Human Proteome Project : Current State and Future Direction. *Mol. Cell. Proteomics* **2011**, *10*, 1−5.

(2) Paik, Y. K.; Jeong, S. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H. J.; Na, K.; Choi, E. Y.; Yan, F.; et al. The Chromosome-Centric Human Proteome Project for Cataloging Proteins Encoded in the Genome. *Nat. Biotechnol.* **2012**, *30*, 221−223.

(3) Paik, Y. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H. J.; et al. Standard Guidelines for the Chromosome-Centric Human Proteome Project. *Journal of Proteome Research.* **2012**, *11*, 2005−2013.

(4) Naryzhny, S. N.; Maynskova, M. A.; Zgoda, V. G.; Ronzhina, N. L.; Novikova, S. E.; Belyakova, N. V.; Kleyst, O. A.; Legina, O. K.; Pantina, R. A. F. M. Proteomic Profiling of High-Grade Glioblastoma Using Virtual-Experimental 2DE. *J. Proteomics Bioinf.* **2016**, *9* (6), 158−165.

(5) Naryzhny, S. N.; Lisitsa, A. V.; Zgoda, V. G.; Ponomarenko, E. A.; Archakov, A. I. 2DE-Based Approach for Estimation of Number of Protein Species in a Cell. *Electrophoresis* **2014**, *35* (6), 895−900.

(6) Zabel, C.; Klose, J. Protein Extraction for 2DE. *Methods Mol. Biol.* **2009**, *519*, 171−196.

(7) Naryzhny, S.; Maynskova, M.; Zgoda, V.; Archakov, A. Dataset of Protein Species from Human Liver. *Data Br.* **2017**, *12*, 584−588.

(8) Naryzhny, S.; Zgoda, V.; Kopylov, A.; Petrenko, E.; Archakov, A. A Semi-Virtual Two Dimensional Gel Electrophoresis: IF−ESI LC−MS/MS. *MethodsX* **2017**, *4*, 260.

(9) Naryzhny, S. N.; Maynskova, M. A. Proteomic Profiling of High-Grade Glioblastoma Using Virtual Experimental 2DE. *J. Proteomics Bioinf.* **2016**, DOI: 10.4172/jpb.1000402.

(10) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities Utilizing the Principle of. *Anal. Biochem.* **1976**, *72*, 248−254.

(11) Gorg, A.; Postel, W.; Domscheit, A.; Gunther, S. Two-Dimensional Electrophoresis with Immobilized pH Gradients of Leaf Proteins from Barley (Hordeum Vulgare): Method, Reproducibility and Genetic Aspects. *Electrophoresis* **1988**, *9* (11), 681−692.

(12) Naryzhny, S. N. Blue Dry Western: Simple, Economic, Informative, and Fast Way of Immunodetection. *Anal. Biochem.* **2009**, *392* (1), 90.

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX

(13) Naryzhny, S. N.; Zgoda, V. G.; Maynskova, M. A.; Novikova, S. E.; Ronzhina, N. L.; Vakhrushev, I. V.; Khryapova, E. V.; Lisitsa, A. V.; Tikhonova, O. V.; Ponomarenko, E. A.; et al. Combination of Virtual and Experimental 2DE Together with ESI LC−MS/MS Gives a Clearer View about Proteomes of Human Cells and Plasma. *Electrophoresis* **2016**, *37* (2), 302−309.

(14) Vaudel, M.; Barsnes, H.; Berven, F. S.; Sickmann, A.; Martens, L. SearchGUI: An Open-Source Graphical User Interface for Simultaneous OMSSA and X!Tandem Searches. *Proteomics* **2011**, *11* (5), 996−999.

(15) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell. Proteomics* **2005**, *4* (9), 1265−1272.

(16) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; et al. Mapping Intact Protein Isoforms in Discovery Mode Using Top-down Proteomics. *Nature* **2011**, *480* (7376), 254−258.

(17) Naryzhny, S. N.; Lee, H. 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.* **2004**, *279* (19), 20194−20199.

(18) Halligan, B. D.; Ruotti, V.; Jin, W.; Laffoon, S.; Twigger, S. N.; Dratz, E. A. ProMoST (Protein Modification Screening Tool): A Web-Based Tool for Mapping Protein Modifications on Two-Dimensional Gels. *Nucleic Acids Res.* **2004**, *32*, W638.

(19) Halligan, B. D. ProMoST: A Tool for Calculating the pI and Molecular Mass of Phosphorylated and Modified Proteins on Two-Dimensional Gels. *Methods Mol. Biol.* **2009**, *527*, 283.

(20) Xu, G.; Jaffrey, S. R. Proteomic Identification of Protein Ubiquitination Events. *Biotechnol. Genet. Eng. Rev.* **2013**, *29* (1), 73−109.

(21) Hendriks, I. A.; Vertegaal, A. C. O. A Comprehensive Compilation of SUMO Proteomics. *Nat. Rev. Mol. Cell Biol.* **2016**, *17* (9), 581−595.

(22) Bellia, F.; Vecchio, G.; Rizzarelli, E. Carnosinases, Their Substrates and Diseases. *Molecules* **2014**, *19*, 2299−2329.

(23) Peters, V.; Zschocke, J.; Schmitt, C. P. Carnosinase, Diabetes Mellitus and the Potential Relevance of Carnosinase Deficiency. *J. Inherited Metab. Dis.* **2018**, *41*, 39−47.

(24) Buxbaum, J. N.; Reixach, N. Transthyretin: The Servant of Many Masters. *Cell. Mol. Life Sci.* **2009**, *66*, 3095−3101.

(25) Naryzhny, S. N.; Lee, H. Observation of Multiple Isoforms and Specific Proteolysis Patterns of Proliferating Cell Nuclear Antigen in the Context of Cell Cycle Compartments and Sample Preparations. *Proteomics* **2003**, *3*, 930−936.

(26) Castellanos-Serra, L.; Paz-Lago, D. Inhibition of Unwanted Proteolysis during Sample Preparation: Evaluation of Its Efficiency in Challenge Experiments. *Electrophoresis* **2002**, *23* (11), 1745−1753.

(27) Thiede, B.; Koehler, C. J.; Strozynski, M.; Treumann, a.; Stein, R.; Zimny-Arndt, U.; Schmid, M.; Jungblut, P. R. Protein Species High Resolution Quantitative Proteomics of HeLa Cells Using SILAC-2-DE-nanoLC/LTQ-Orbitrap Mass Spectrometry. *Mol. Cell. Proteomics* **2013**, *12* (2), 529−538.

(28) Naryzhny, S.; Zgoda, V.; Kopylov, A.; Petrenko, E.; Archakov, A. A Semi-Virtual Two Dimensional Gel Electrophoresis: IF−ESI LC−MS/MS. *MethodsX* **2017**, *4*, 260−264.

(29) Naryzhny, S. N.; Maynskova, M. A.; Zgoda, V. G.; Ronzhina, N. L.; Kleyst, O. A.; Vakhrushev, I. V.; Archakov, A. I. Virtual-Experimental 2DE Approach in Chromosome-Centric Human Proteome Project. *J. Proteome Res.* **2016**, *15*, 525.

(30) Vizcaíno, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 Update of the PRIDE Database and Its Related Tools. *Nucleic Acids Res.* **2016**, *44* (D1), D447−D456.

(31) Naryzhny, S. Inventory of Proteoforms as a Current Challenge of Proteomics: Some Technical Aspects. *J. Proteomics* **2018**, DOI: 10.1016/j.jprot.2018.05.008.

DOI: 10.1021/acs.jproteome.8b00386
J. Proteome Res. XXXX, XXX, XXX−XXX