proteomes

MDPI

Article

The "2DE-Pattern" Database for Inventory of Proteoform Profiles: 2026 Upgrade and Update on Outcomes

Stanislav Naryzhny [1,*], Nikolay Klopov [1], Natalia Ronzhina [1], Elena Zorina [2] and Olga Legina [1]

[1] Petersburg Nuclear Physics Institute Named by B. P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina 188300, Russia; klopov_nv@pnpi.nrcki.ru (N.K.); ronzhina_nl@pnpi.nrcki.ru (N.R.); olgaleg@mail.ru (O.L.)

[2] Institute of Biomedical Chemistry, Pogodinskaya, 10, Moscow 119121, Russia; el.zorina@ibmc.msk.ru

* Correspondence: snaryzhny@mail.ru; Tel.: +7-9111764453

Abstract

Background: Modern proteomics faces a critical bottleneck: the vast discrepancy between the number of genes in the human genome and the exponentially greater variety of functional proteoforms that actually drive biological processes. Methods: Our paper addresses the urgent need for high-resolution systematic mapping of these proteoforms, arguing that the true frontier of molecular biology lies in the precise identification and categorization of protein variants. It centers on the development and expansion of the "2DE-pattern" database, a specialized platform designed to bridge the gap between theoretical protein sequences and the physical reality of proteins as captured through two-dimensional electrophoresis (2DE). The "2DE-pattern" database is based on information obtained by separation of proteoforms using 2DE followed by shotgun ESI LC-MS/MS. It was launched in 2020, contains multiple isoform-centric patterns of proteoforms, and can be freely used. Results: Here, we report the additional data and all updates that were added into this database. Also, the database was upgraded to be more research-oriented. Tools were incorporated into the database to allow convenient comparative analysis of the data. Conclusions: New additions and enhancements now allow us to consider our database a knowledge base.

Keywords: proteoforms; database; bioinformatics; pattern; 2DE

1. Introduction

As proteomics delves deeper into the protein heterogeneity associated with a huge number of proteoforms, the amount of information grows very rapidly, so there is high demand for convenient ways to store and use this information [1,2]. The main users of proteomic databases are researchers in the fields of biology, bioinformatics, and clinical science analyzing protein levels, their sequences, and molecular networks. Therefore, databases are a crucial part of proteomics. The most well-known open-access databases are located at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) and on the SIB ExPASy bioinformatics resource portal, https://www.expasy.org/, including, for example, NeXtProt and UniProt (https://www.nextprot.org/, http://www.uniprot.org/) [3,4]. The Human Protein Atlas is another prominent research portal dedicated to mapping all human proteins across cells, tissues, organs, and blood [5]. The Consortium for Top-Down Proteomics (CTDP) is involved in a project aiming to define the complete set of human proteoforms [6,7]. Accordingly, a proteoform database called

Check for updates

Academic Editor: Xavier Roucou

Received: 9 June 2026
Revised: 22 August 2026
Accepted: 31 August 2026
Published: 7 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Proteoform Atlas was organized (http://atlas.topdownproteomics.org) [8,9]. In addition, archive databases such as PRIDE (the PRoteomics IDEntifications database) or PeptideAtlas, where information about proteoforms can be found, are available [10,11]. Typically, the data collected in the databases are based on the specific method by which they were obtained. As for proteoforms, two-dimensional gel electrophoresis (2DE) is ideal for their separation and subsequent identification [12–15]. Moreover, 2DE allows one to measure the basic parameters of a protein molecule—molecular weight (MW) and pI—and allows for the storage of the obtained information in a highly visually accessible format [16–20]. Although the subsequent analysis is performed using bottom-up mass spectrometry, the overall approach is called top-down integrative proteomics [21–25]. To enable a complete analysis of samples, another approach was developed whereby the entire gel was divided into individual sections, which were then processed and subjected to mass spectrometry. Based on this approach, we generated proteoform profiles for several cell types [26–28]. Moreover, these data were used to create a web database called "2DE-pattern" [29]. Although this approach is not as accurate as classical top-down mass spectrometry, it allows one to obtain a general idea of proteoform families, referred to as patterns. Here, we report on more data as well as more features that were incorporated into this database recently. These updates have significantly expanded the database and enhanced its discovery power.

2. Materials and Methods

2.1. Sample Preparation

Human hepatocellular carcinoma cells (HepG2, collection of Institute of Biomedical Chemistry, Moscow, Russia), embryonic lung fibroblasts (FLEH, Russian Collection of Vertebrate Cell Cultures in Institute of Cytology of the Russian Academy of Sciences, St.-Petersburg, Russia), and glioblastoma (primary line L, generated in the Laboratory of Cell Biology (NRC "Kurchatov Institute"-PNPI, Gatchina, Russia) cells were cultured, collected, and extracted in the same way as was described in detail before [27,30–32]. Hepatocellular carcinoma and control tissues (n = 5) were extracted according to [28]. The samples contained 2 mg of protein in 100 µL of lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 1% dithiothreitol (DTT), 2% ampholytes, pH 3–10, protease inhibitor mixture). The protein concentration in the sample was determined by the method of Bradford [33].

2.2. 2DE

All procedures (at least 3 technical replicates) were performed according to the protocols described previously [27,30,31,34]. Briefly, for classical 2DE, 1000 µg of protein was loaded onto a 24 cm long nonlinear IPG strip. After isoelectric focusing (IEF), the separation was produced on a 12% polyacrylamide gel (1 mm × 25 cm × 20 cm) under denaturing conditions. The gel was stained with Coomassie Blue R350, scanned with an ImageScanner III (GE Healthcare, Pittsburgh, PA, USA), and analyzed with an ImageMaster 2D Platinum 7.0 (GE Healthcare).

For sectional 2DE, 500 µg of protein was loaded onto a 7 cm long linear IPG strip. After the second dimension, the gel (1 mm × 8 cm × 8 cm) was stained, scanned, analyzed using ImageMaster 2D Platinum 7.0, and divided into 96 sections with defined coordinates designated as 1–12 along the MW dimension and A–H along the pI dimension. Each section was digested with trypsin. Tryptic peptides were eluted from the gel with extraction solution (5% (v/v) acetonitrile, 5% (v/v) formic acid) and dried in a Speed Vac vacuum centrifuge (Thermo Scientific, Waltham, MA, USA). The peptides were dissolved in 5% (v/v) formic acid.

In the case of semi-virtual 2DE [26,35], the first step was the same as in classical 2DE. After IEF, the strip was cut into 48 equal sections and processed according to the trypsin

digestion protocol. Samples were digested overnight at 37 °C. Peptides were extracted with 150 µL of 60% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA). The extracts were dried in a Speed Vac vacuum dryer and dissolved in 20 µL of 0.1% TFA before analysis by mass spectrometry (MS). Protein identification and relative quantification were performed using SearchGUI v 4.0.12. A semi-virtual 2D map was constructed according to the theoretical MW and experimental pI of the detected proteins.

The virtual 2DE maps were constructed using data obtained by the so-called Filter-Aided Sample Preparation (FASP) method [28,36]. In this case, the detected proteins were plotted on the map according to their theoretical pI/MW and experimentally detected abundance (emPAI).

2.3. ESI LC-MS/MS Analysis

The analysis was completed by the "shotgun" approach as described previously [26,35]. Panoramic proteomic analysis of the obtained extracts was performed using filter processing and subsequent mass spectrometry (FASP). Centrifuge concentrators (Microcon YM–30, Merck, Rahway, NJ, USA) were used for this purpose. In short, extracts containing the required amount of protein (300 µg) were placed in concentrators and sequentially treated with solutions: (a) for the reduction of disulfide bonds (100 mM DTT in 100 mM Tris-HCl, pH 8.5), then (b) for the alkylation of sulfhydryl groups (50 mM iodoacetamide, 8 M urea, 100 mM Tris-HCl, pH 8.5), and (c) for hydrolysis with trypsin (Trypsin Gold, Promega, Madison, WI, USA). Tandem mass-spectrometry analysis was carried out on an Orbitrap Q-Exactive mass spectrometer (Thermo Scientific, USA) according to the protocols described previously, using an Agilent HPLC system 1100 Series ("Agilent Technologies", Santa Clara, CA, USA) [37].

2.4. Protein Identification

Identification of proteins was performed using SearchGUI, an open-source graphical user interface [38]. Two unique peptides per protein were required for all protein identifications. The exponentially modified PAI (emPAI), the exponential form of the protein abundance index (PAI), defined as the number of detected peptides divided by the number of theoretically observable tryptic peptides for each protein, was used to estimate protein abundance [39]. All additional information about the methods can also be reached through the front page of the database by clicking the corresponding links for protocols or articles.

2.5. Software Used for Database Construction

All data were stored in a Mysql database (version 15.1) on a Linux server. The scripts that process requests from the user interface were developed using Perl 5 (version 30). The website is based on webserver Apache/2.4.39. The interactive user interface for the database is implemented using HTML, CSS, JavaScript, and JQuery library. The queries to server programs are executed using AJAX technology.

3. Results and Discussion

3.1. Overview of the Database

Our database "2DE-pattern" was developed to provide a simple and comprehensive tool to store and analyze information about proteoforms separated according to 2DE principles (Figure 1). These principles are based on the fundamental physicochemical parameters of polypeptides, namely isoelectric point (pI) and molecular weight (MW). Each proteoform (polypeptide) has a specific set of these parameters. Unfortunately, the very useful database SWISS-2DPAGE is not available anymore, and the interactive website https://world-2dpage.expasy.org/swiss-2dpage/ has been discontinued (accessed on 1 January

2026). Accordingly, we are trying to enlarge and upgrade our database. We are using three different approaches based on experimental and virtual 2DE separation and identification using mass spectrometry. The first approach is based on classical 2DE. In addition to the experimental 2DE, we performed virtual 2DE for HCC samples. In this case, the information about protein abundances was obtained by the direct ESI LC-MS/MS analysis of extracts. The virtual 2DE was conducted based on this experimental information and the theoretical parameters (pI, MW) available for each polypeptide. The second one is sectional 2DE, in which a whole gel, not only selected spots, is analyzed by ESI LC-MS/MS section by section. The third one is semi-virtual 2DE, in which proteoforms are separated only by IEF according to their pI. Each approach allows for the production of a specific proteoform pattern for every isoform. All three approaches have limitations, but they are complementary to each other and allow for obtaining a better view of the combined proteoform profiles of isoforms (Figure 2). To generate the complete image of the proteoform landscape, detailed information (sequence and PTMs) about each proteoform should be obtained. But this is a task for the next step in proteoform identification. Currently, the database contains information on proteoform patterns for 8014 isoforms of 7053 proteins. This represents a twofold increase compared with the first version of the database.

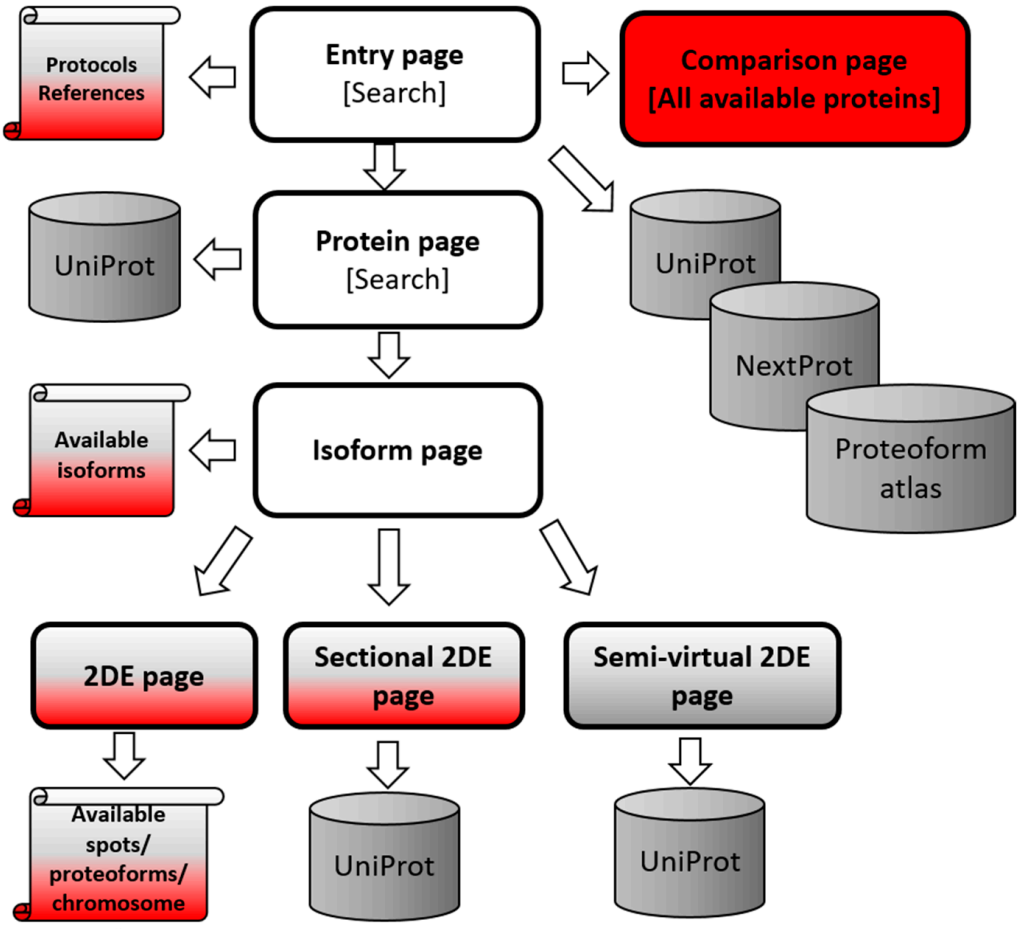


Figure 1. A flow chart of the proteoform database "2DE-pattern". A user starts from the top, at the "Entry page", where a choice of search parameters can be found. The next steps can be done by following the arrows. Blocks colored red indicate content that differs from the first edition of the database. The graph is a modified version of that presented in [29].

NN	Protein	Uniprot	Obj	N	sum	max	%m/s
0	1433B	P31946-1	Glio	27	94.61	25.90	27.4
1	1433B	P31946-1	HCC	7	3.26	0.81	24.8
2	1433B	P31946-1	HCCcontrol	8	5.34	2.53	47.4
3	1433B	P31946-1	LEH	22	71.16	22.42	31.5
4	1433B	P31946-1	LEH	3	21.56	20.13	93.4
5	1433E	P62258-1	Glio	30	787.36	443.61	56.3
6	1433E	P62258-1	HCC	15	7.48	1.66	22.2
7	1433E	P62258-1	HCCcontrol	20	22.09	7.11	32.2
8	1433E	P62258-1	LEH	23	261.05	184.72	70.8
9	1433F	Q04917-1	Glio	11	33.13	15.04	45.4
10	1433F	Q04917-1	HCC	4	1.02	0.32	31.4
11	1433F	Q04917-1	HCCcontrol	3	2.31	1.66	71.9
12	1433F	Q04917-1	LEH	11	39.30	14.71	37.4
13	1433G	P61981-1	Glio	24	78.81	28.75	36.5
14	1433G	P61981-1	HCC	10	3.24	0.56	17.3
15	1433G	P61981-1	HCCcontrol	11	5.78	3.10	53.6
16	1433G	P61981-1	HepG2	15	102.75	47.46	46.2
17	1433G	P61981-1	LEH	21	89.16	32.29	36.2
18	1433S	P13947-1	Glio	2	6.99	5.83	83.4
19	1433S	P13947-1	HepG2	10	46.14	20.38	44.2
20	1433T	P27348-1	Glio	31	180.94	40.06	22.1
21	1433T	P27348-1	HCC	3	1.15	0.45	39.1
22	1433T	P27348-1	HCCcontrol	5	3.62	1.26	34.8
23	1433T	P27348-1	HepG2	28	271.84	73.09	26.9
24	1433T	P27348-1	LEH	28	102.30	29.54	28.9

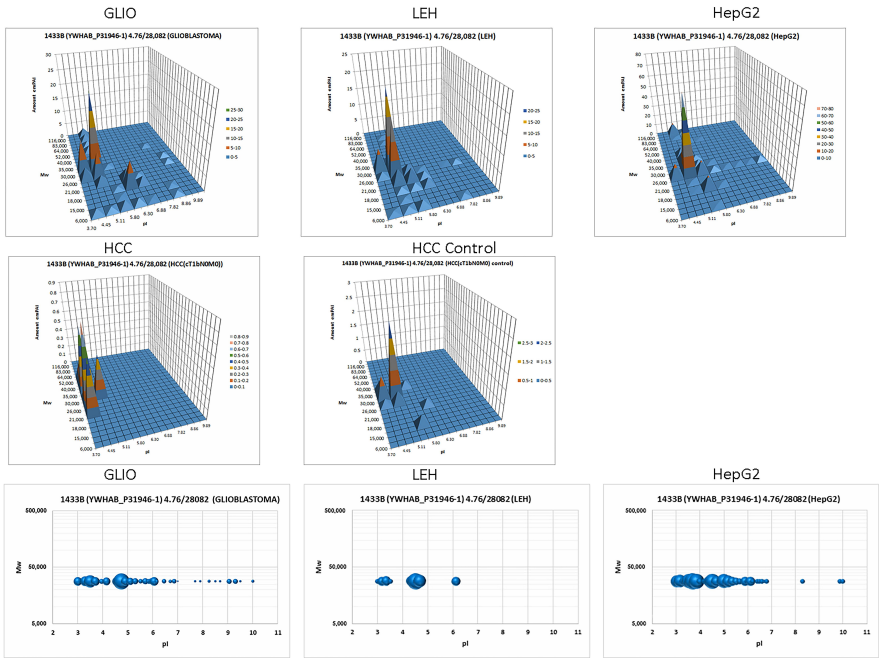


Figure 2. An example of the "Comparison Page" (a partial view). The comparison of the available images can be performed here after selection of the isoform number.

3.2. Additions to the Database Content and Search Features

Users can search the database using three distinct identifiers from the front (entry) page:

- The UniProt accession number of the target protein.
- The official name of the protein.
- The gene symbol or name.

The home page provides direct links to the underlying data sources:

- Protocols: Detailed laboratory methods used to generate the data.
- References: Published papers validating the database content.

We also added the ability to check the general information about all isoforms included in the database. This can be done by clicking "All available proteins (isoforms)—Click HERE". This will transfer users to the Comparison Page, where they can evaluate and compare protein expressions:

- emPAI Values: View quantitative data for detected proteoforms.
- Bulk Graph Loading: Render all available charts simultaneously.
- Pattern Comparison: Contrast and compare 2DE maps across different samples.

The data currently originate from five specific sources:

- Glia-L: Glioblastoma cell line.
- LEH: Human embryonic lung fibroblasts.
- HepG2: Hepatocellular carcinoma cell line.
- HCC: Hepatocellular carcinoma tissue samples.
- HCCCONTROL: Healthy control tissue for hepatocellular carcinoma.

The original version of the database contained only Glia-L and LEH.

3.3. Navigation Workflow

The database utilizes a hierarchical transition model to navigate from general protein data to specific experimental proteoform patterns:

[Protein Page] ---> [Isoform Page] ---> [Sample Pages (3 Types)]

Protein Page: Displays basic information, SWISS-PROT ID/accession number, description line, and all available isoforms.

Isoform Page: Links to different sample pages and includes a master table mapping isoform data (Figure 1).

Sample Pages: Delivers specialized proteoform patterns based on the chosen analytical approach.

3.4. Sample Page Types and Methods

The database offers three types of sample pages determined by the experimental workflow used to analyze the proteome. The first type is classical or virtual 2DE maps:

Classical 2DE Map: This displays physical, experimental gel spots where proteins were detected. In the 2DE map, spots where different proteoforms of the same isoform were detected are highlighted. Accordingly, a proteoform pattern of this isoform is produced. Additionally, basic experimental information about the spot abundance and the isoform (isoform name, protein name, gene name, chromosome, theoretical pI, theoretical Mw, experimental pI, theoretical MW, emPAI, modifications) is shown. Minimal information obtained by mass spectrometric analysis is shown in the table. Also, by clicking on a spot a user can extract information about all the isoforms that were detected in this spot. Additionally, there is an option to check each chromosome for the detected proteoforms. Classical 2DE is used for the samples Glia-L, LEH, and HepG2.

Virtual 2DE Map: This was used for HCC and HCCCONTROL samples (Figure 3). It constructs a digital map using quantitative FASP analysis of extracts combined with theoretical physical–chemical parameters (pI and MW) of proteins. Virtual 2DE map data provide additional visual representation of the proteome state. Interestingly, the virtual 2DE map of the master protein forms (277 proteins) encoded by human chromosome 18 also resembles a liver or tumor protein map. This provides additional grounds for extrapolating many of the data obtained on chromosome 18 proteins to the entire proteome.

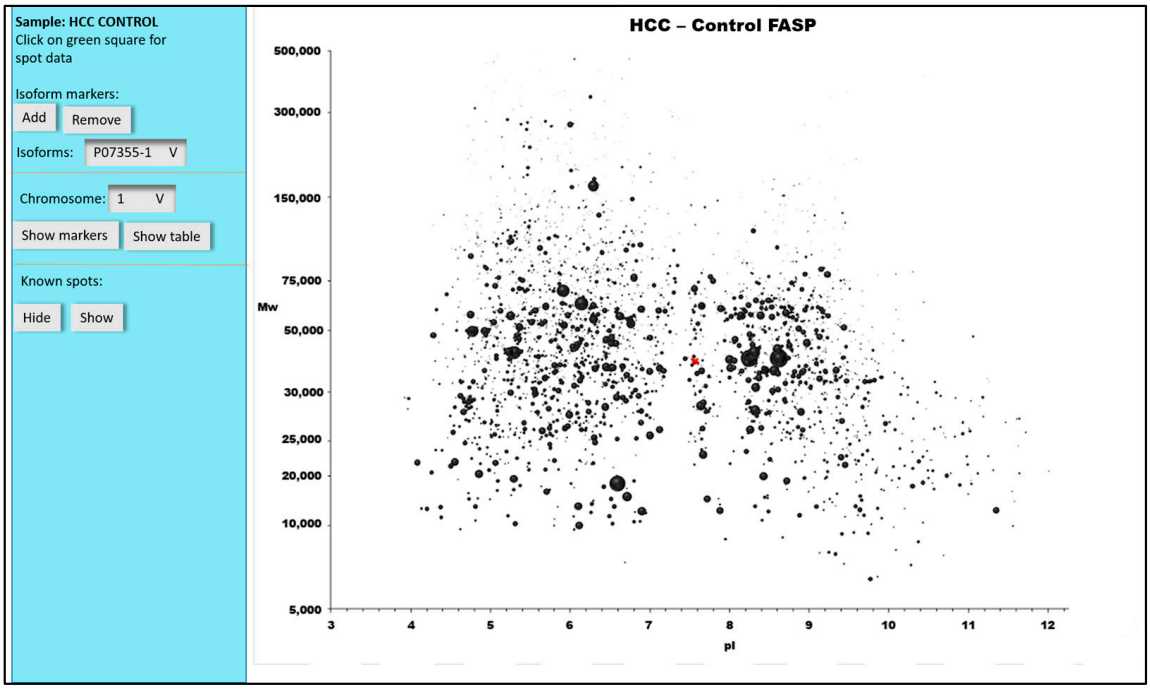
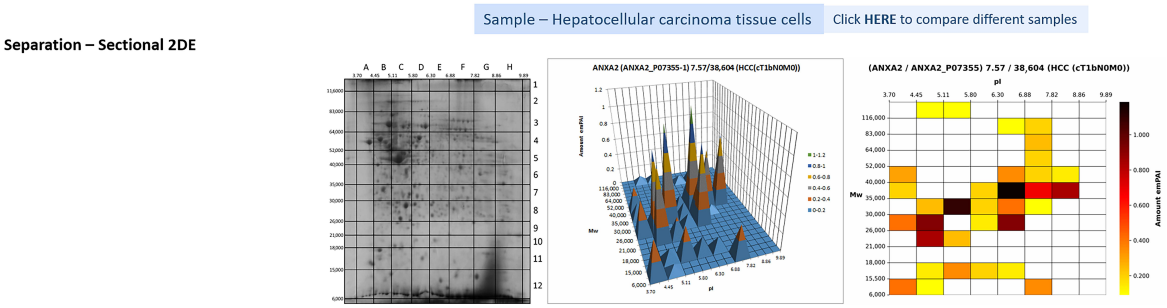


Figure 3. An example of a "Sample Page" on the human proteoform database "2DE-pattern", where virtual 2DE was used. Information about each spot is presented in the table. A spot of the master form of isoform 1 of ANXA2 (P07355-1) is marked by the red cross.

A "Sample Page" with sectional 2DE was also upgraded (Figure 4). To improve perception and analysis, two types of patterns were constructed: volumetric (three-dimensional), in the form of pyramids, where the height of the pyramid of each section corresponded to the level of the proteoform located in this section (the emPAI parameter); and in the form of heat maps, where the color corresponded to the level of the proteoform. A "Sample Page" with semi-virtual 2DE looks the same as it did when the database was launched in 2020.



Isoform – P07355-1 (Annexin A2)

UniProt	Theor. pI	Theor. Mw	Gene	Protein	Chromosome	Protein name
P07355-1	7.57	38,604	ANXA2	ANXA2	15	Annexin A2

Proteoforms

Position	Exp. pI	Exp. Mw	Score	#Validated Peptides	Amount emPAI	Modification sites
#A12	3.70-4.45	6,000-15,000	Infinity	8	0.42	
#A5	3.70-4.45	40,000-52,000	Infinity	6	0.30	Oxidation of M(M240)
#A6	3.70-4.45	35,000-40,000	Infinity	4	0.19	
#A8	3.70-4.45	26,000-30,000	Infinity	8	0.42	Oxidation of M(M240)
#B1	4.45-5.11	11,6000-up	Infinity	2	0.09	
#B11	4.45-5.11	15,000-18,000	Infinity	3	0.14	
#B12	4.45-5.11	6,000-15,000	Infinity	2	0.09	
#B7	4.45-5.11	30,000-35,000	Infinity	5	0.24	
#B8	4.45-5.11	26,000-30,000	Infinity	15	0.92	
#B9	4.45-5.11	21,000-26,000	Infinity	14	0.84	Oxidation of M(M217, M240)
#C1	5.11-5.80	11,6000-up	Infinity	2	0.09	

Figure 4. An example of a "Sample Page" with sectional 2DE. A 2DE gel (HCC) divided into sections with determined coordinates is shown on the left. Pyramid-type and heat-map-type distributions of proteoforms among the sections (proteoform pattern) of isoform 1 of ANXA2 (P07355-1) are shown on the right. A table contains basic theoretical information about this isoform (pI/MW, isoform, protein, gene, chromosome). Another table contains experimental data about the proteoforms (pI/MW coordinates of the sections, where the isoform was detected, and MS information—score, number of detected sequences, emPAI, modification sites).

All the different approaches have advantages and disadvantages. The data obtained by classical experimental 2DE have high resolution but are limited in terms of the number of proteoforms, as they were detected only in the stained spots. Virtual 2DE allows the user to get unbiased images of the proteomes, which are convenient to compare with each other. But they contain only the master forms of proteins and are missing information about other proteoforms. While sectional 2DE suffers from low resolution and overlapping proteoform signals, it maps the entire gel to detect invisible, unstained proteoforms. The pattern obtained by semi-virtual 2DE has high enough resolution in the pH direction but is missing information about the real MW of the proteoforms. Integrating information from these four approaches, a user can reveal a comprehensive, "ideal" proteoform profile. The "Sample page" contains direct links to external databases for deeper analysis. Users can access UniProt to view known post-translational modifications (PTMs) for the isoform. Alternatively, they can use the Proteoform Atlas to check if specific proteoforms have been previously detected via top-down mass spectrometry.

4. Conclusions

This paper describes an upgraded version of the database "2DE-pattern", which was developed for storing and analyzing information about patterns or profiles of human proteoforms generated according to 2DE principles [29]. The database can be freely used at

http://2de-pattern.pnpi.nrcki.ru. A central theme here is the necessity of 2DE in an era dominated by mass spectrometry [23,40]. Critics often argue that high-resolution mass spectrometers can now allow identifying thousands of proteins in a single run, seemingly rendering the "visual" approach of the 2DE-pattern database obsolete [41]. However, this argument overlooks the difference between identification and characterization [42]. Our approach allows generating an "integral image" of proteoform distribution according to their physicochemical parameters. This allows performing a direct comparison between normal and cancer samples to find changes that are involved in malignant transformation. This comparison can enable discovering the presence of specific proteoforms suitable for use as biomarkers or as targets for therapy. A panel of such proteins for hepatocarcinoma was revealed recently [28]. The complete characterization of these proteoforms is a more difficult problem to solve. It seems that top-down MS can provide a solution to this task. Indeed, there has been major progress in this area, as the Proteoform Atlas now has data about 66496 proteoforms detected in human cells (http://human-proteoform-atlas.org/ accessed on 22 May 2026). What is important is that the data from the Proteoform Atlas and integrative top-down proteomics (2DE-pattern) can be complementary to each other [25,28]. The primary challenge lies in unifying these experimental results. Implementing data standardization and unification protocols would provide significant benefits. Specifically, integrating key metadata—such as the average molecular weight (MW) and calculated isoelectric point (pI) for each entry—into the Proteoform Atlas would enable direct cross-database comparisons, particularly with the "2DE-pattern" database.

The future of proteoform-based proteomics will depend on the systematic integration of proteoform mapping, functional characterization, and predictive modeling. This vision requires progress in several areas that will enhance the translational value of proteoform research [43–45]. Maintaining and replenishing the proteoform databases should surely be a basis of this process [46].

Author Contributions: Conceptualization, supervision, writing—review and editing, S.N.; software, data curation, resources, N.K.; formal analysis, data curation, N.R.; formal analysis, data curation, E.Z.; formal analysis, data curation, O.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of the Russian Scientific Center of Surgery named after B.V. Petrovsky (protocol № 14 from 16 December 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: The data that support the findings of this study are openly available at http://2de-pattern.pnpi.nrcki.ru/index.html (accessed on 19 June 2026) and https://proteomecentral.proteomexchange.org/ui?pxid=PXD010142 (accessed on 19 June 2026).

Acknowledgments: The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (theme №1024011100004-4-1.6.8;1.6.4;1.6.7;1.6.1. Functional and structural organization of complex, multicomponent biological systems and their dynamics. Molecular and cellular components of the pathogenesis of socially significant diseases). Mass-spectrometry measurements were performed using the equipment of the "Human Proteome" Core Facilities of the Institute of Biomedical Chemistry (Moscow, Russia). No AI or AI-assisted technologies were used in the preparation of this work. The authors take full responsibility for the content and originality of the publication.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

2DE	two-dimensional electrophoresis
ESI LC-MS/MS	liquid chromatography–electrospray ionization–tandem mass spectrometry
DTT	dithiothreitol
MS	mass spectrometry
PTM	post-translational modification
emPAI	exponential modified form of protein abundance index
SIB	Swiss Institute of Bioinformatics

References

1. Toby, T.K.; Fornelli, L.; Kelleher, N.L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev. Anal. Chem.* **2016**, *9*, 499–519. [CrossRef] [PubMed] [PubMed Central]
2. Schaffer, L.V.; Millikin, R.J.; Shortreed, M.R.; Scalf, M.; Smith, L.M. Improving Proteoform Identifications in Complex Systems Through Integration of Bottom-Up and Top-Down Data. *J. Proteome Res.* **2020**, *19*, 3510–3517. [CrossRef] [PubMed]
3. Bairoch, A. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45–48. [CrossRef] [PubMed]
4. Lane, L.; Argoud-Puy, G.; Britan, A.; Cusin, I.; Duek, P.D.; Evalet, O.; Gateau, A.; Gaudet, P.; Gleizes, A.; Masselot, A.; et al. NeXtProt: A knowledge platform for human proteins. *Nucleic Acids Res.* **2012**, *40*, D76–D83. [CrossRef] [PubMed]
5. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef] [PubMed]
6. Smith, L.M.; Agar, J.N.; Chamot-Rooke, J.; Danis, P.O.; Ge, Y.; Loo, J.A.; Paša-Tolić, L.; Tsybin, Y.O.; Kelleher, N.L. The Human Proteoform Project: Defining the human proteome. *Sci. Adv.* **2021**, *7*, eabk0734. [CrossRef]
7. Dang, X.; Scotcher, J.; Wu, S.; Chu, R.K.; Tolić, N.; Ntai, I.; Thomas, P.M.; Fellers, R.T.; Early, B.P.; Zheng, Y.; et al. The first pilot project of the consortium for top-down proteomics: A status report. *Proteomics* **2014**, *14*, 1130–1140. [CrossRef] [PubMed]
8. Hollas, M.A.R.; Robey, M.T.; Fellers, R.T.; LeDuc, R.D.; Thomas, P.M.; Kelleher, N.L. The Human Proteoform Atlas: A FAIR community resource for experimentally derived proteoforms. *Nucleic Acids Res.* **2021**, *50*, D526–D533. [CrossRef] [PubMed]
9. Fornelli, L.; Toby, T.K.; Schachner, L.F.; Doubleday, P.F.; Srzentić, K.; DeHart, C.J.; Kelleher, N.L. Top-down proteomics: Where we are, where we are going? *J. Proteom.* **2018**, *175*, 3–4. [CrossRef] [PubMed]
10. Perez-Riverol, Y.; Bandla, C.; Kundu, D.J.; Kamatchinathan, S.; Bai, J.; Hewapathirana, S.; John, N.S.; Prakash, A.; Walzer, M.; Wang, S.; et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **2025**, *53*, D543–D553. [CrossRef] [PubMed]
11. Desiere, F.; Deutsch, E.W.; King, N.L.; Nesvizhskii, A.I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S.N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, D655-8. [CrossRef] [PubMed]
12. Farrell, P.H.O. The pre-omics era: The early days of two-dimensional gels. *Electrophoresis* **2009**, *8*, 4842–4852. [CrossRef] [PubMed]
13. Klose, J. From 2-D electrophoresis to proteomics. *Electrophoresis* **2009**, *30*, 142–149. [CrossRef] [PubMed]
14. Zhan, X.; Jungblut, P.R. The comparison between 2DE-MS and bottom-up LC-MS demands high-end techniques for both technologies. *Electrophoresis* **2022**, *43*, 1242–1245. [CrossRef] [PubMed]
15. Simula, M.P.; Notarpietro, A.; Toffoli, G.; De Re, V. 2-D gel electrophoresis: Constructing 2D-gel proteome reference maps. *Methods Mol. Biol.* **2012**, *815*, 163–173. [CrossRef] [PubMed]
16. Anderson, L.; Anderson, N.G. High resolution two-dimensional electrophoresis of human plasma proteins. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5421–5425. [CrossRef] [PubMed]
17. Linn, N.H.; Doustjalali, S.R.; Bhuiyan, M.; Al-Jashamy, K.; Kadir, S.Y.A.; Appalanaidu, V.; Arzuman, H.; Krishnan, A.; Lin, W.M.; Saraji, A.; et al. Two dimensional gel electrophoresis: An overview of proteomic technique in cancer research. *J. Proteom. Bioinform.* **2014**, *7*, 3. [CrossRef]
18. Westermeier, R.; Görg, A. Two-Dimensional Electrophoresis in Proteomics. In *Protein Purification: Principles, High Resolution Methods, and Applications: Third Edition*; HitGen Inc.: Chengdu, China, 2011; pp. 411–439.
19. Lee, P.Y.; Saraygord-Afshari, N.; Low, T.Y. The evolution of two-dimensional gel electrophoresis—From proteomics to emerging alternative applications. *J. Chromatogr. A* **2020**, *1615*, 460763. [CrossRef] [PubMed]
20. Rabilloud, T.; Chevallet, M.; Luche, S.; Lelong, C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J. Proteom.* **2010**, *73*, 2064–2077. [CrossRef] [PubMed]
21. Naryzhny, S. Puzzle of proteoforms—Where is a key? *Proteomes* **2024**, *12*, 15. [CrossRef] [PubMed]
22. Woodland, B.; Necakov, A.; Coorssen, J.R. Optimized Proteome Reduction for Integrative Top-Down Proteomics. *Proteomes* **2023**, *11*, 10. [CrossRef] [PubMed]

23. Zhan, X.; Li, B.; Zhan, X.; Schlüter, H.; Jungblut, P.R.; Coorssen, J.R. Innovating the concept and practice of two-dimensional gel electrophoresis in the analysis of proteomes at the proteoform level. *Proteomes* **2019**, *7*, 36. [CrossRef] [PubMed]
24. Carbonara, K.; Padula, M.P.; Coorssen, J.R. Quantitative assessment confirms deep proteome analysis by integrative top-down proteomics. *Electrophoresis* **2023**, *44*, 472–480. [CrossRef] [PubMed]
25. Kaulich, P.T.; Tholey, A. Top-Down Proteomics: Why and When? *Proteomics* **2025**, *25*, 6–12. [CrossRef] [PubMed]
26. 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*, 525–530. [CrossRef] [PubMed]
27. 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*, 302–309. [CrossRef] [PubMed]
28. Zorina, E.; Ronzhina, N.; Legina, O.; Klopov, N.; Zgoda, V.; Naryzhny, S. Proteoform Patterns in Hepatocellular Carcinoma Tissues: Aspects of Oncomarkers. *Proteomes* **2025**, *13*, 27. [CrossRef] [PubMed]
29. Naryzhny, S.; Klopov, N.; Ronzhina, N.; Zorina, E.; Zgoda, V.; Kleyst, O.; Belyakova, N.; Legina, O. 2DE-pattern: A database for inventory of human proteoforms. *Electrophoresis* **2020**, *41*, 1118–1124. [PubMed]
30. Naryzhny, S.; Maynskova, M.; Zgoda, V.; Archakov, A. Dataset of protein species from human liver. *Data Brief* **2017**, *12*, 584–588. [CrossRef] [PubMed]
31. Naryzhny, S.; Maynskova, M.; Zgoda, V.; Archakov, A. Zipf's Law in Proteomics. *J. Proteom. Bioinform.* **2017**, *10*, 79–84.
32. Naryzhny, S.; Ronzhina, N.; Zorina, E.; Kabachenko, F.; Klopov, N.; Zgoda, V. Construction of 2DE Patterns of Plasma Proteins: Aspect of Potential Tumor Markers. *Int. J. Mol. Sci.* **2022**, *23*, 11113. [CrossRef] [PubMed]
33. Bradford, M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254. [CrossRef] [PubMed]
34. Gorg, A.; Obermaier, C.; Boguth, G.; Harder, A.; Scheibe, B.; Wildgruber, R.; Weiss, W. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **2000**, *21*, 1037–1053. [CrossRef]
35. 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. [CrossRef] [PubMed]
36. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [CrossRef] [PubMed]
37. Zgoda, V.G.; Moshkovskii, S.A.; Ponomarenko, E.A.; Andreewski, T.V.; Kopylov, A.T.; Tikhonova, O.V.; Melnik, S.A.; Lisitsa, A.V.; Archakov, A.I. Proteomics of mouse liver microsomes: Performance of different protein separation workflows for LC-MS/MS. *Proteomics* **2009**, *9*, 4102–4105. [CrossRef] [PubMed]
38. 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*, 996–999. [CrossRef] [PubMed]
39. 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. Proteom.* **2005**, *4*, 1265–1272. [CrossRef] [PubMed]
40. Kim, Y.I.; Cho, J.Y. Gel-based proteomics in disease research: Is it still valuable? *Biochim Biophys. Acta Proteins Proteom.* **2019**, *1867*, 9–16. [CrossRef] [PubMed]
41. Mann, M.; Kelleher, N.L. Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18132–18138. [CrossRef] [PubMed]
42. Carbonara, K.; Andonovski, M.; Coorssen, J.R. Proteomes Are of Proteoforms: Embracing the Complexity. *Proteomes* **2021**, *9*, 38. [CrossRef] [PubMed]
43. Korchak, J.A.; Stephen Yi, S.; Kelleher, N.L.; Sahni, N.; Sheynkman, G.M. Proteoform medicine: Characterizing and targeting protein forms in human disease. *Nat. Rev. Genet.* **2026**, *27*, 271–291. [CrossRef] [PubMed]
44. Hoàng, T.; Hu, Y.; Zhang, H. ProteoformDB: A Built-In Application to Generate Proteoform Database. *J. Proteome Res.* **2026**, *25*, 145–156. [CrossRef] [PubMed]
45. Po, A.; Eyers, C.E. Top-Down Proteomics and the Challenges of True Proteoform Characterization. *J. Proteome Res.* **2023**, *22*, 3663–3675. [CrossRef] [PubMed]
46. Fang, Z.; Zhang, Y.; Feng, X.; Li, N.; Chen, L.; Zhan, X. Proteoformics: Current status and future perspectives. *J. Proteom.* **2025**, *321*, 105524. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.