
Two-Dimensional Gel Electrophoresis as an Information Base for Human Proteome

Stanislav Naryzhny

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75125>

Abstract

The main intricacy in the human proteome is that it is tremendously complex and composed of 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 should be performed. The approaches that allow disclosing and keeping the information about human proteome using two-dimensional gel electrophoresis (2DE) are described. Experimental identification methods such as mass spectrometry of high resolution and sensitivity (MALDI-TOF MS and ESI LC-MS/MS) or immunodetection in combination with bioinformatics and 2DE can be used for the development of a comprehensive knowledge base of the human proteome.

Keywords: two-dimensional gel electrophoresis, proteome, proteoforms, mass spectrometry, bioinformatics

1. Introduction

The first aim in human proteomics, as it was proposed by the Human Proteome Organization (HUPO), is a complete catalog of all human proteins. Due to collaborative efforts inside the Chromosome-based Human Proteome Project (C-HPP), this task is close to completion now [1, 2]. According to NextProt release from Aug 1, 2017, 17,168 from 20,199 predicted proteins have already been found. Still, ~3000 proteins are in the list of so-called “missing proteins.” But even after completion of this task, only representative proteins will be identified in most cases [3, 4]. The situation here is much more complicated, as proteins can exist as different proteoforms (protein species) [5–7]. Proteoforms, as the smallest units of the proteome, are molecules (polypeptides) arising from all combinatorial sources of variation after

expression of a single gene. Each proteoform is a chemically clearly defined molecule. These molecules are different due to genetic variation, alternatively spliced RNA, transcripts and posttranslational modifications [5, 7]. Accordingly, the term protein refers to its coding gene and, therefore, becomes as the umbrella term for all developing proteoforms/protein species [6]. Sometimes the term “proteoform” is used for the description of structural variants of proteins as well [8]. But it will make the issue of terminology in proteomics more complicated and confused, as even inside the abovementioned definition, a proteomics field of proteoforms is very broad and could encompass many billions of components [9–12]. For instance, all combinations of 30 known modifications of histone H3 alone can theoretically produce more than 1 billion of proteoforms [10, 13]. Because of such a variety, huge range of concentration (7–8 orders of magnitude in blood plasma), and dynamic changes during life cycle, their identification, quantitation and database organization is a serious challenge. Nevertheless, there is evident progress in this area. So far, the main workhorse in proteomics was bottom-up mass spectrometry, but the top-down approach is becoming pre-eminent today [14]. 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 [14, 15]. But top-down proteomics cannot be just a one-step procedure. There are also several approaches based on protein separation that are involved in proteoform analysis. Among these methods, two-dimensional gel electrophoresis (2DE) occupies the special place. Accordingly, different schemes could be used to establish a basis for a comprehensive knowledge base for protein/proteoform inventory.

2. The principles of protein separation by 2DE

Application of different electrophoretic separation methods in two-dimensional combinations has a long history [16–23]. Continuous development on improvement of electrophoretic techniques and working out new support media for separation of proteins allowed to achieve better and better resolution. Finally, the best combination was chosen and optimized by O’Farrell in 1975 [24]. The method is based on separations in completely denaturing conditions according to two independent parameters (pI and Mw). Isoelectric focusing (IEF) that separates proteins due to difference in their isoelectric points (pI) is used as a first direction. High resolution according to protein size or weight (Mw) in the second direction is achieved by SDS polyacrylamide gel electrophoresis (SDS-PAGE) [23]. Quickly, the method was accepted by scientific community as the most powerful approach for separation of complex protein samples [25, 26]. Besides the high resolution, an important part of this method is that all stages of separation are performed in denaturing conditions, so in case of complex mixtures like cell extracts, all proteins or proteoforms are separated according to their basic parameters (pI and Mw) of their primary chemical structure (**Figure 1**). This is very important, as it allows not only to separate proteoforms but **also to** experimentally determine their pI and Mw. Also, these parameters can be calculated based on information about the chemical structure of the

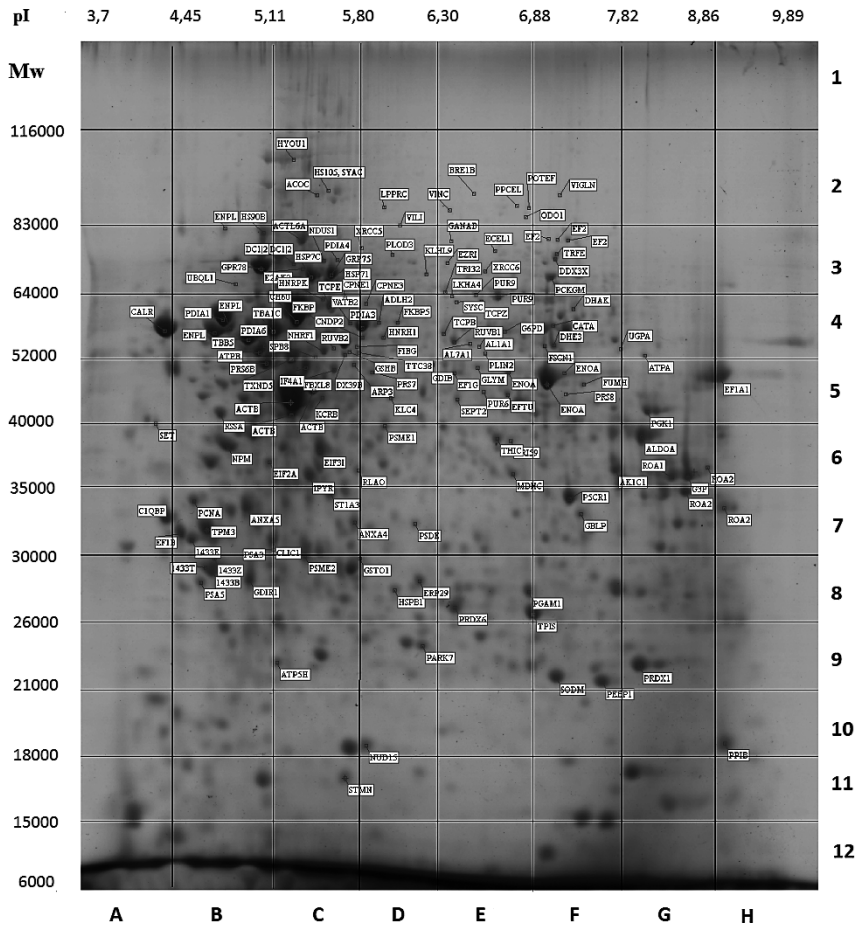


Figure 1. 2DE map of HepG2 proteins. Spots with proteins identified by MALDI TOF-MS are annotated. Sections in 2D gel selected for following LC ESI-MS/MS analysis are shown. Adapted from Naryzhny et al. [28].

proteoforms. Accordingly, two-dimensional separation of proteoforms can be performed not only experimentally but virtually as well [27–29] (**Figure 2**).

3. Information produced by 2DE gels

Separation with high resolution of proteoforms is the main function of 2DE, but extra steps should be done toward obtaining necessary information. Protein identification is the next very important step. Immunological methods as most specific approaches were mainly used for this

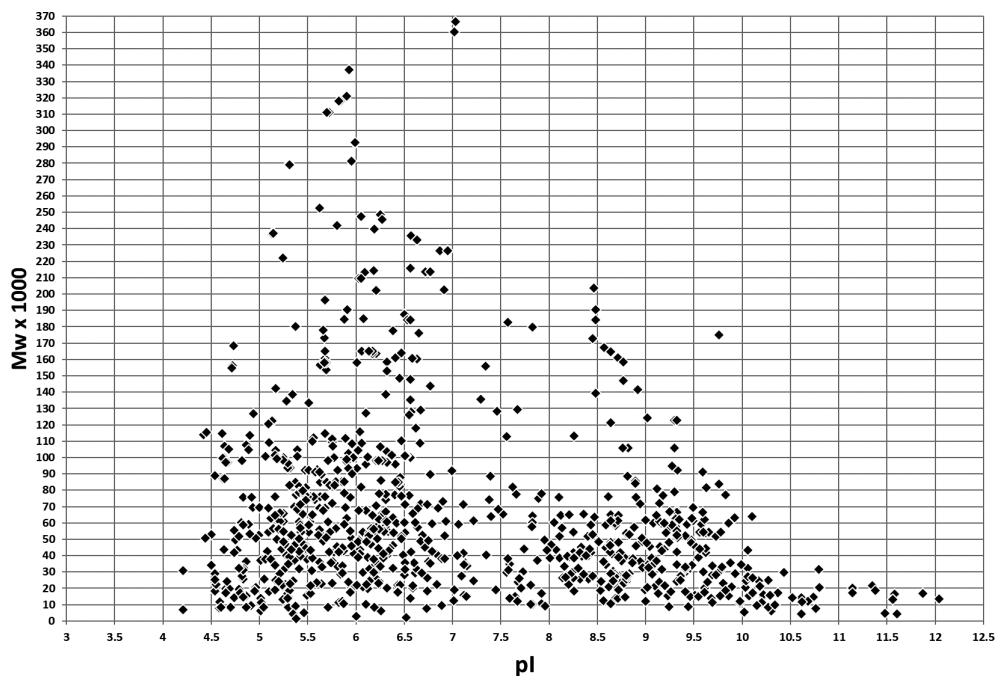


Figure 2. A virtual 2DE map of proteins coded by human chromosome 18.

purpose. They are still successfully used in 2DE-based proteomics for protein identification (Western blot) (Figure 3).

To study protein-protein interactions, so-called Far-Western blot is used (Figure 4) [32, 33]. In this case, the proteins after 2DE separation are transferred to a membrane and then treated with a set of buffers, which apply washes that allow “prey” proteins to denature and renature [32–34]. The membrane is then blocked and probed with a purified “bait” protein (the protein, which binding partners (“preys”) need to be detected). The bait protein is detected on spots where the prey protein is located if the bait proteins and the prey protein bind together. For detection of bait proteins, they can be labeled, or antibodies can be used [32, 33]. This technique is not only very informative but also laborious and not convenient for a large-scale analysis. The situation was radically improved when mass spectrometry became a central element for proteomics analysis [35, 36].

The MS approach for protein identification by searching for the best match between peptide masses produced by specific hydrolysis and peptide masses calculated from theoretical cleavage of proteins was developed simultaneously by different groups [37–41]. In this case, the peptide mass sets are acquired by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). This approach was named peptide mass fingerprinting (PMF). For a long time, PMF has been the most powerful technique for high-throughput

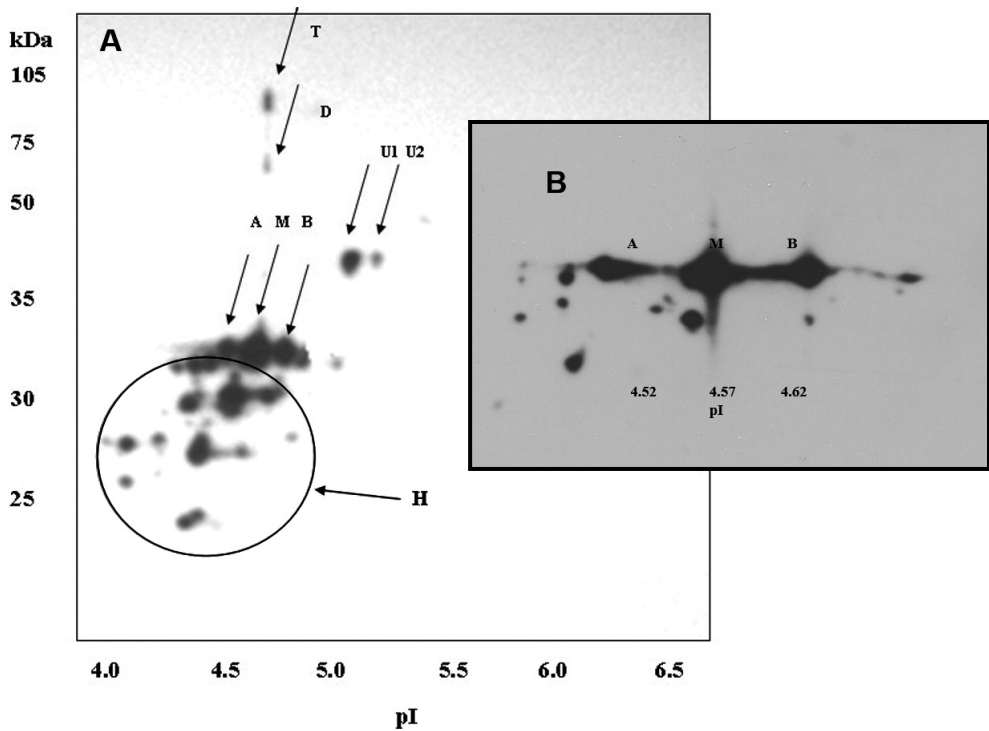


Figure 3. Multiple spots of PCNA (proteoforms) can be detected by 2DE. After 2DE separation and transfer to Immobilon-P membrane, immunostaining was performed using PCNA-specific antibody PC10. A, M, B, T and D denote acidic, main, basic, probable trimer and probable dimer forms of PCNA, respectively. U1 and U2 are ubiquitinated forms of PCNA monomer. H is a cluster of PCNA hydrolysis products, which is produced by proteasome action [30]. (A) 2DE (pH 4–7, 13 cm – first dimension, 10% SDS-PAGE, 13 cm – second dimension). (B) 2DE (pH 4–5, 18 cm – first dimension, 10% SDS-PAGE, 13 cm – second dimension). Adapted from Naryzhny [31].

protein identification. Using more sophisticated MS instruments (especially ESI LC–MS/MS), it was revealed that, depending on the gel resolution, the 2DE spots often contain more than a single protein, especially in the case of mammalian cells [29, 42]. Accordingly, the quantitation of proteins became an ambiguous task, as the densitometry of spots cannot be used for accurate proteoform quantitation. Fortunately, the special MS-based quantitative approaches can be applied in this case. In addition, as only the proteins detected as stained spots are analyzed, a lot of information is missing. This problem can be solved only by analyzing all parts of the gel [28, 43]. The general view of this method is shown in **Figure 1**. The main steps are as follows:

1. 2DE separation.
2. Staining the gel with Coomassie R350.
3. Scanning the image produced (2DE map).

- Analyzes and calibration of the 2DE map by Image Master 2DE Platinum (GE Healthcare, Pittsburgh, PA, USA). Calibration was performed according to the position of several major protein spots that had been previously identified: actin cytoplasmic (ACTB_HUMAN, pI 5.29/Mw 42,052), 78 kDa glucose-regulated protein (GRP78_HUMAN, pI 5.07/Mw 72,333), tropomyosin alpha-3 chain (TPM3_HUMAN, pI 4.68/Mw 32,950), stathmin 1 (STMN_HUMAN, pI 5.76/Mw 17,292) and alpha-enolase (ENOA_HUMAN, pI 7.01/Mw 47,481).
- The separation of the gel into 96 sections (identified as 1–12 along the Mw dimension and A–H along the pI dimension). According to the calibration, each section is given pI/Mw coordinates.

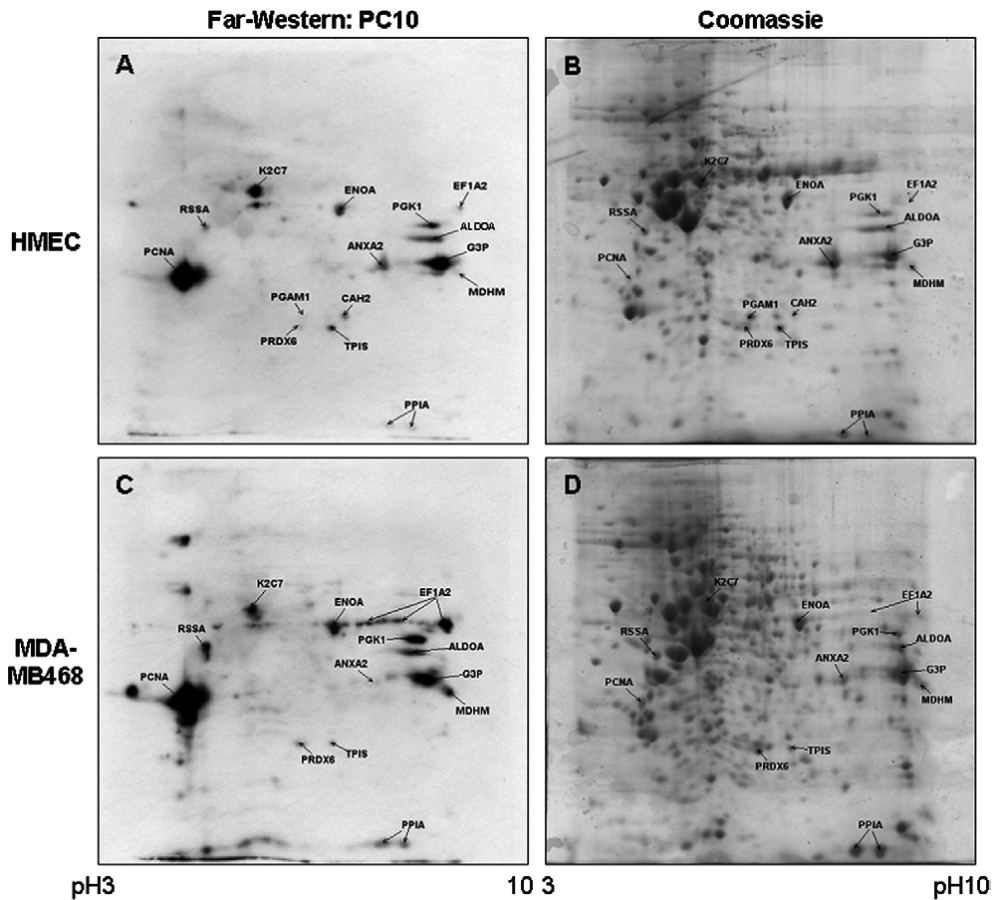


Figure 4. Identification of PCNA-binding proteins in normal and cancer cells by 2DE-Far-Western blotting. Protein extracts from HMEC (A and B) or MDA-MB468 (C and D) were separated by 2DE, transferred onto a membrane, renatured and incubated with PCNA. PCNA-binding proteins were then detected with the PC10 monoclonal anti-PCNA antibody. (B and D) Proteins separated by 2DE were stained with Coomassie R350, protein spots corresponding to signals on the membrane (A and C) were identified, cut out, and the identity of the proteins was determined by mass spectrometry. Adapted from Naryzhny et al. [32].

6. The treatment of each section with trypsin according to the protocol for mass-spectrometry by ESI LC–MS/MS.
7. Analysis of the tryptic peptides obtained from each 2DE section by Orbitrap Q-Exactive mass spectrometer.
8. Protein identification and relative quantification are performed using Mascot “2.4.1” and exponentially modified form of protein abundance index (emPAI).

In total, up to 500 unique proteins were identified in each section of 2DE gel after separation of proteins from glioblastoma cell extract (**Figure 1**). All proteins detected in the same section were given the pI/Mw parameters of this section.

Respectively, the same proteins detected in different sections were considered as different proteoforms. About 20,000 proteoforms coded by ~4000 genes were identified using this approach [28, 43]. Additionally, 3D graphs represented gene centric expression of proteoforms can be generated (**Figure 5**). Here, the proteoform profiles for each gene can

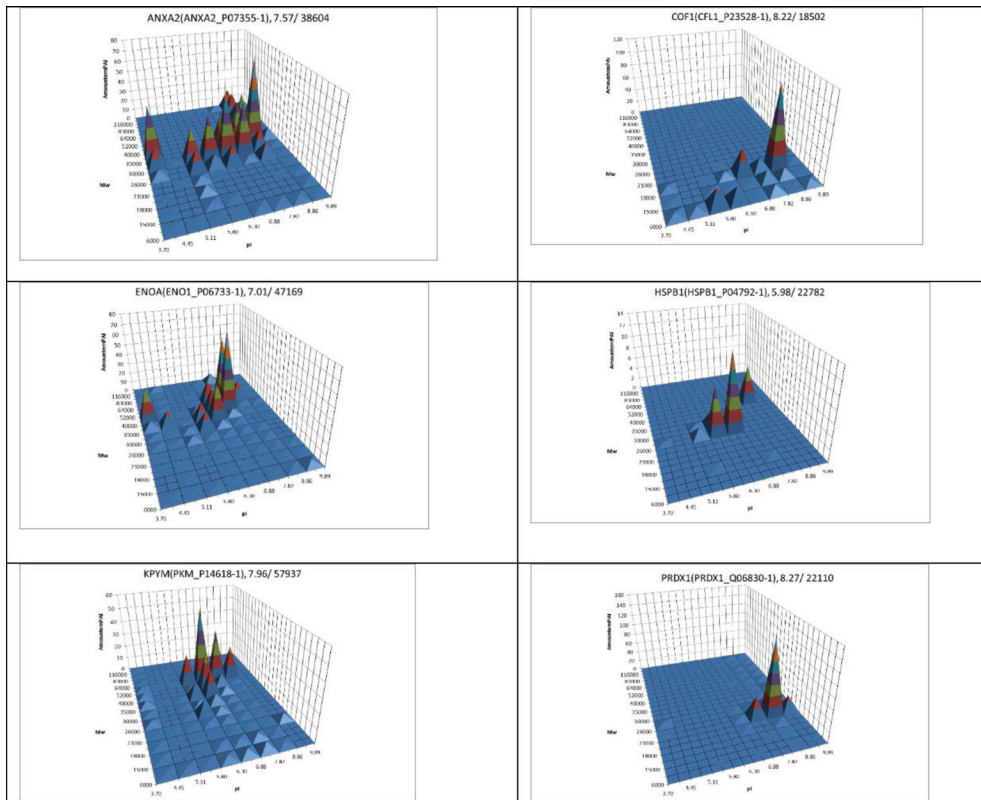


Figure 5. The examples of 3D graphs showing distribution of proteoforms between different sections of the 2DE map. A semiquantitative (estimated by emPAI) distribution of the same protein (gene) around the different gel sections was plotted. Proteins, the potential biomarkers, are shown. Reproduced with permission from Naryzhny et al. [43].

be observed. Considering that only 96 sections from a small gel (8 cm × 8 cm) were taken for analysis, the situation can be improved significantly by increasing the gel size, the sample loading and the number of sections. This approach also solves the problem with a sensitivity of the staining. However, there are still some issues that need to be tackled. One of them is a resolution, which drops largely if we cut the gel into big sections. Ideally, the size of these sections should be close to the size of the smallest spots. But in this case, we are facing a dramatically increased amount of samples (around 1000). As there is a limitation in the processing time of each sample by ESI LC-MS/MS (usually at least 0.5 h per sample), we will need about 1 month of continuous work on the Orbitrap to analyze these samples from a single 2DE gel. Another issue is proteoform quantitation. There are several options available. The exponentially modified form of protein abundance index (emPAI) [28, 43, 44] is not ideal, since it gives only a relative and not very accurate estimation of the protein content. Another option is MaxQuant that recently became possibly the most frequently used platforms for mass-spectrometry (MS)-based proteomics data analysis [45]. Since its release in 2008, it has been improved substantially [45, 46]. Selected reaction monitoring (SRM) and isotope-coded affinity tag (ICAT) are other excellent examples of the power of MS technology in protein quantitation [47–49]. There are also more approaches that are described and reviewed [50, 51].

4. Databases based on 2DE gels

The overall approach has been used to generate annotated 2DE gel databases for many cell types. The first 2DE database, SWISS-2DPAGE database has been launched in 1993 and is maintained by the Central Clinical Chemistry Laboratory of the Geneva University Hospital and the Swiss Institute of Bioinformatics (SIB). Now this database is a part of the World-2DPAGE (<http://world-2dpage.expasy.org/list/>), a dynamic portal to query simultaneously world-wide gel-based proteomics databases. These databases put together over 250 maps for 23 species, totalizing nearly 40,000 identified spots, making it the biggest gel-based proteomics dataset accessible from a single interface. Here, we can select a 2DE map which will be displayed for inspection. The database can be queried by keywords (protein description, protein name, gene name, species, author, full text, protein spot serial number) or graphically by clicking on a spot. Each spot is linked to a page containing the corresponding gene (protein) information and identification details. Also, information is displayed about other spots in different maps, where product of the same gene is detected. All these spots are highlighted in the maps and the calculated parameters [isoelectric point (pI) and molecular weight (Mw)] are displayed. There is a possibility for cross-references and obtaining more information from different 2DE databases and from UniProtKB. UniProtKB, a comprehensive protein sequence knowledge base has two sections: UniProtKB/Swiss-Prot, which is manually curated and UniProtKB/TrEMBL that contains computer-annotated entries. UniProtKB/Swiss-Prot entries provide users with cross-links to about 100 external databases and with access to additional information or tools [52].

5. Collection of experimental and theoretical data on the platform of 2DE gel

The proteomes can be retrieved from experimental data or generated by available programs, which calculate theoretical protein parameters [27, 53, 54]. According to their basic parameters (pI/Mw), the unique principles of separation of polypeptides allow to organize the crosstalk between experimental and theoretical data. In a simple way, this approach was realized in abovementioned 2DE databases. The main idea behind the approach, where each spot is considered as containing only one protein, works well only with simple proteomes (for instance, mycoplasma), where the number of proteoforms are not so big as in mammalian cells [11, 12, 55, 56]. A European pathogenic microorganism proteome database focused on pathogenic microorganisms was launched. Now, under the name “The Proteome 2D-PAGE Database,” this database currently contains 13,893 identified spots and 3245 mass peak lists in 57 reference maps representing experiments from 26 different organisms and strains. The database provides protein information such as ORF name, predicted isoelectric point (pI) and molecular weight (Mw), several protein identifiers, identification method, sequence coverage, and so on. [57]. This database contains information about mammalian cells as well. Therefore, multiple proteins can be attributed to the same spot in case of these cells.

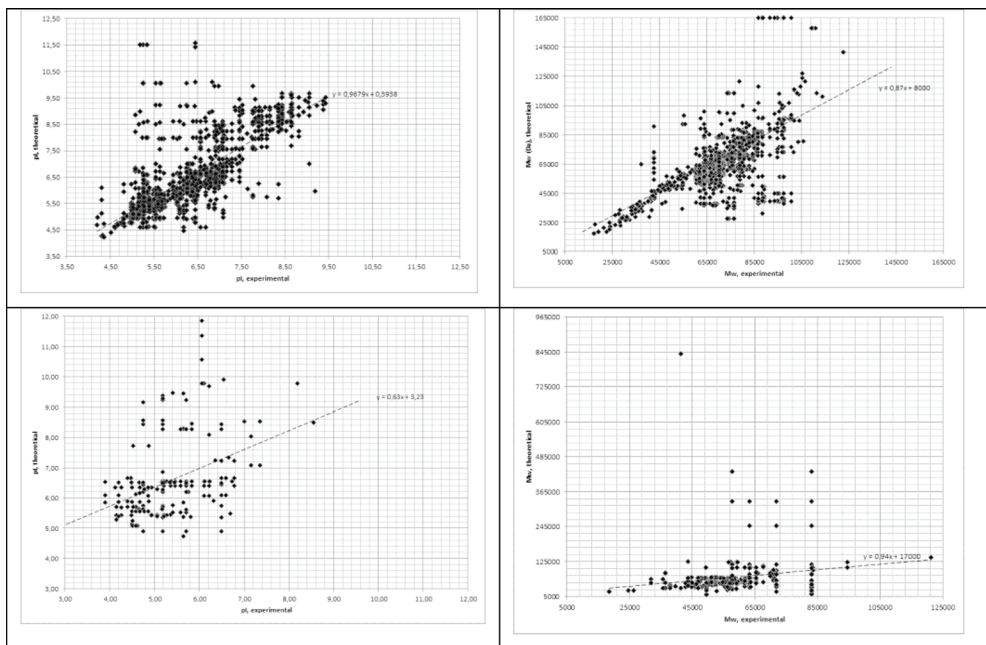


Figure 6. Relationship between theoretically (in silico) determined pI (left) or mw (right) of proteins (canonical forms) and experimentally detected pI/mw of proteoforms. Top – HepG2 cells, bottom – human depleted plasma. Adapted from Naryzhny et al. [29].

The idea to manage the theoretical proteomes and link them with experimental, 2DE-based data has been tried to realize in another online database, DynaProt 2D [58]. It was developed for dynamic access to proteomes and 2DE gels. Here, a 2DE gel could serve as a reference map and as a tool for navigation of the database [58]. Integrated into 2DE database a complete theoretical proteome could provide a powerful tool allowing simply linking newly identified spots to the already available appropriate theoretical data [58]. But this idea was tried for one organism only, *Lactococcus lactis*, and stopped in realization.

If we plot experimentally measured physicochemical parameters of proteoforms (pI or Mw) against the theoretical ones, the general view of the proteome according to the diversity of proteoforms is revealed (**Figure 6**). Here, the dots are distributed along the diagonal in the graph, if experimentally detected parameters match or are close to the theoretical values. Otherwise, dots are distributed above or below the diagonal, and the bigger the difference between theoretical and experimental parameters, the bigger the deviation of the dot position from the diagonal. In particular, in case of pI, the location of proteoform dots above the diagonal shows that the experimental pIs of this proteoform is smaller than the theoretical

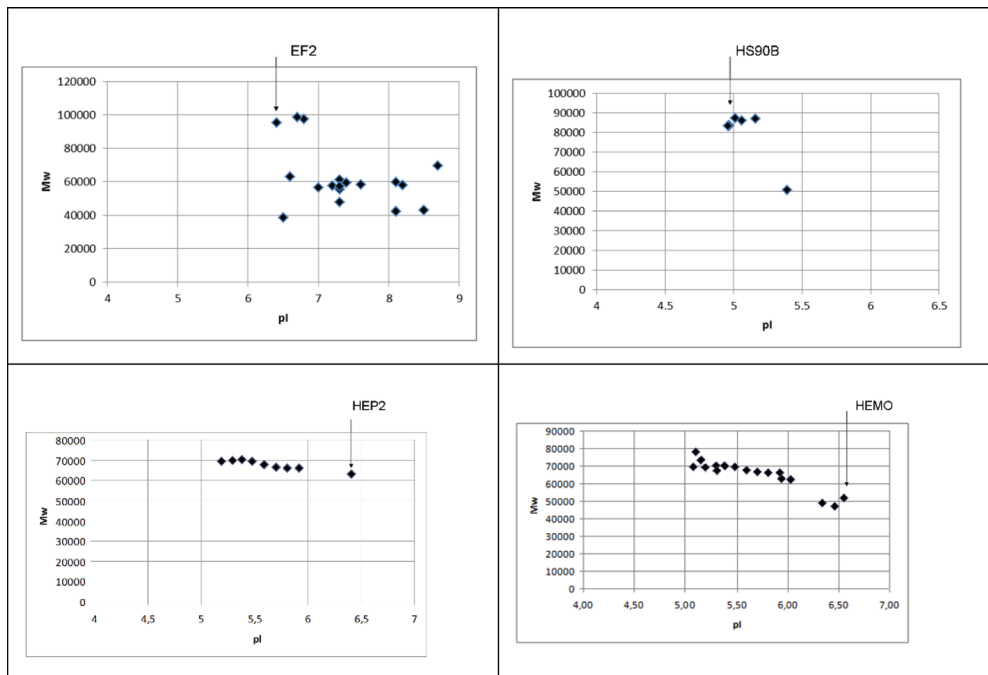


Figure 7. The examples of variety of proteoforms in HepG2 cells (top) and human plasma (bottom). The arrow shows the predicted location of the master polypeptide (polypeptide coded by the canonical sequence). Adapted from Naryzhny et al. [29].

one. It can be usually because of adding negative charges (phosphorylation) or removing positive charges (acetylation). An opposite situation with increased pI (the dots are under the diagonal) is observed; then, PTMs are removing negative charges from proteins (esterification). In case of Mw, location of a dot above the diagonal means that theoretical mass (weight) of this polypeptide is bigger than the experimentally observed one. It could be if the polypeptide is truncated or proteolytically processed. If the dots are observed below the diagonal, it can be because of several reasons. It can happen because of technical issues of 2DE procedure (protein polymerization, aggregation or precipitation). Glycosylation also can strongly increase a polypeptide mass. This situation can be analyzed in more detail by representation of all proteoforms corresponding to the same gene on the separate chart (Figure 7).

6. Conclusion

The continuing evolution of the detection technique (mostly mass spectrometry) and usage of it in combination with optimum protein separation techniques will finally allow us to reach the main aim of the HUPO—image of the whole human proteome. A union of such a classic proteomics method for separation of proteins as 2DE 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 union of top-down and bottom-up approaches allows very convenient visual representation (profiling) of information about diverse proteoforms. As 2DE maps are a convenient and effective way to represent information about proteomes and navigate around all its proteoforms, it will allow the construction of a knowledge base for an inventory of all human protein species/proteoforms, that is, visually attractive, clear, easy to search and perceptive.

Particularly, the development of chromosome-centric interactive virtual 2DE maps of proteins coded by specific genes in combination with experimental 2DE protein maps will allow executing more effectively C-HPP, to estimate more accurately the number of proteoforms and could be a basis for the knowledge base of human proteins. The development of such inventory will be based on existing databases like <http://world-2dpage.expasy.org/>, <https://www.nextprot.org/>, <http://www.uniprot.org/> and <http://atlas.topdownproteomics.org>.

Acknowledgements

The study was performed in frames of Program of Fundamental 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).

Author details

Stanislav Naryzhny^{1,2}

Address all correspondence to: snaryzhny@mail.ru

1 Institute of Biomedical Chemistry, Moscow, Russia

2 Petersburg Nuclear Physics Institute, National Research Center "Kurchatov Institute", Leningrad Region, Russia

References

- [1] Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. *Nature* [Internet]. 2014;**509**(7502):575-581. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24870542>
- [2] Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014;**509**(7502):582-587
- [3] Aebersold R, Mann M. Mass spectrometry based proteomics. *Nature* [Internet]. 2003;**422**:198-207. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12634793>
- [4] de Godoy LMF, Olsen J V., Cox J, Nielsen ML, Hubner NC, Fröhlich F, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* [Internet]. 2008;**455**(7217):1251-1254. Available from: <http://www.nature.com/doi/10.1038/nature07341>
- [5] Schlüter H, Apweiler R, Holzhütter H-G, Jungblut PR. Finding one's way in proteomics: a protein species nomenclature. *Chemistry Central Journal* [Internet]. 2009;**3**:11. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2758878&tool=pmcentrez&rendertype=abstract>
- [6] Jungblut P, Thiede B, Zimny-Arndt U, Müller EC, Scheler C, Wittmann-Liebold B, et al. Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis*. 1996;**17**(5):839-847
- [7] Smith LM, Kelleher NL. Proteoform: A single term describing protein complexity. *Nature Methods*. 2013;**10**:186-187
- [8] Uversky VN. p53 Proteoforms and intrinsic disorder: An illustration of the protein structure-function continuum concept. *International Journal of Molecular Sciences*. 2016;**17**(11):1874
- [9] Kelleher NL. A cell-based approach to the human proteome project. *Journal of the American Society for Mass Spectrometry*. 2012;**23**(10):1617-1624
- [10] Jungblut PR. The proteomics quantification dilemma. *Journal of Proteomics* [Internet]. 2014;**107**:98-102 Available from: <http://dx.doi.org/10.1016/j.jprot.2014.03.015>

- [11] Naryzhny SN, Lisitsa AV, Zgoda VG, Ponomarenko EA, Archakov AI. 2DE-based approach for estimation of number of protein species in a cell. *Electrophoresis*. 2014; **35**(6):895-900
- [12] Naryzhny SN, Zgoda VG, Maynskova MA, Ronzhina NL, Belyakova N V, Legina OK, et al. Experimental estimation of proteome size for cells and human plasma. *Biomeditsinskaia Khimiia* [Internet]. 2015;**61**(2):279-285. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25978394>
- [13] Garcia BA, Pesavento JJ, Mizzen CA, Kelleher NL. Pervasive combinatorial modification of histone H3 in human cells. *Nature Methods* [Internet]. 2007;**4**(6):487-489 Available from: <http://www.nature.com/doi/10.1038/nmeth1052>
- [14] Toby TK, Fornelli L, Kelleher NL. Progress in top-down proteomics and the analysis of Proteoforms. *Annual Review of Analytical Chemistry* [Internet]. 2016;**9**(1):499-519 Available from: <http://www.annualreviews.org/doi/10.1146/annurev-anchem-071015-041550>
- [15] Durbin KR, Fornelli L, Fellers RT, Doubleday PF, Narita M, Kelleher NL. Quantitation and identification of thousands of human proteoforms below 30 kDa. *Journal of Proteome Research*. 2016;**15**(3):976-982
- [16] Bussard A, Huet JM. Description of a technic simultaneously combining electrophoresis and immunological precipitation in gel: Electrosyneresis. *Biochimica et Biophysica Acta*. 1959;**34**:258-260
- [17] Laurell CB. Electroimmuno assay. *Scandinavian Journal of Clinical and Laboratory Investigation* [Internet]. 1972;**29**(s124):21-37 Available from: <http://www.tandfonline.com/doi/full/10.3109/00365517209102748>
- [18] Laurell C-B. Quantitative estimation of proteins by electrophoresis in agarose gel containing antibodies. *Analytical Biochemistry* [Internet]. 1966;**15**(1):45-52 Available from: <http://www.sciencedirect.com/science/article/pii/0003269766902466>
- [19] Awdeh ZL, Williamson AR, Askonas BA. Isoelectric focusing in polyacrylamide gels. *Nature*. 1968;**219**:66-67
- [20] Righetti PG, Drysdale JW. Isoelectric focusing in gels. *Journal of Chromatography. A*. 1974;**98**(2):271-321
- [21] Dale G, Latner AL. Isoelectric focusing in polyacrylamide gels. *Lancet*. 1968;**1**:847-848
- [22] Fawcett JS. Isoelectric fractionation of proteins on polyacrylamide gels. *FEBS Letters*. 1968;**1**:81-82
- [23] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;**227**(5259):680-685
- [24] O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* [Internet]. 1975;**250**(10):4007-4021 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2874754&tool=pmcentrez&rendertype=abstract>

- [25] Anderson NGAN. The human protein index. *Clinical Chemistry* [Internet]. 1982;**28**(4): 739-748 Available from: <http://clinchem.aaccjnls.org/content/28/4/739.long>
- [26] Anderson NG, Matheson A, Anderson NL. Back to the future: The human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* [Internet]. 2001;**1**(1):3-12 Available from: [http://onlinelibrary.wiley.com/doi/10.1002/1615-9861\(200101\)1:1%3C3::AID-PROT3%3E3.0.CO;2-T/abstract](http://onlinelibrary.wiley.com/doi/10.1002/1615-9861(200101)1:1%3C3::AID-PROT3%3E3.0.CO;2-T/abstract)
- [27] Hiller K, Schobert M, Hundertmark C, Jahn D, Münch R. JVirGel: Calculation of virtual two-dimensional protein gels. *Nucleic Acids Research*. 2003;**31**(13):3862-3865
- [28] Naryzhny SN, Maynskova MA, Zgoda VG, Ronzhina NL, Kleyst OA, Vakhrushev IV, et al. Virtual-experimental 2DE approach in chromosome-centric human proteome project. *Journal of Proteome Research*. 2016
- [29] Naryzhny SN, Zgoda VG, Maynskova MA, Novikova SE, Ronzhina NL, Vakhrushev IV, 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
- [30] Naryzhny SN, 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;**6**:930
- [31] Naryzhny SN. Proliferating cell nuclear antigen: A proteomics view. *Cellular and Molecular Life Sciences C* [Internet]. 2008;**65**(23):3789-3808. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18726183>
- [32] Naryzhny SN, Lee H. Proliferating cell nuclear antigen in the cytoplasm interacts with components of glycolysis and cancer. *FEBS Letters*. 2010;**584**(20):4292-4298
- [33] Wu Y, Li Q, Chen X-Z. Detecting protein-protein interactions by far western blotting. *Nature Protocols*. 2007;**2**(12):3278-3284
- [34] Guichet A, Copeland JW, Erdélyi M, Hlousek D, Závorszky P, Ho J, et al. The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. *Nature* [Internet]. 1997;**385**(6616):548-552. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9020363>
- [35] Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* [Internet]. 1999;**20**(18):3551-3567. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16638152>
- [36] Webster J, Oxley D. Protein identification by MALDI-TOF mass spectrometry. *Methods in Molecular Biology* [Internet]. 2012;**800**:227-240. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21964792>
- [37] Pappin DJC, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*. 1993;**3**(6):327-332
- [38] Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein

- sequence databases. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 1993;**90**(11):5011-5015 Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=46643&tool=pmcentrez&rendertype=abstract>
- [39] Mann M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biological Mass Spectrometry*. 1993;**22**(6):338-345
- [40] James P, Quadroni M, Carafoli E, Gonnet G. Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications* [Internet]. 1993;**195**:58-64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8363627>
- [41] Yates JR, Speicher S, Griffin PR, Hunkapiller T. Peptide mass maps: A highly informative approach to protein identification. *Analytical Biochemistry* [Internet]. 1993;**214**(2):397-408 Available from: <http://www.sciencedirect.com/science/article/pii/S0003269783715149>
- [42] Thiede B, Koehler CJ, Strozynski M, Treumann A, Stein R, Zimny-Arndt U, et al. Protein species high resolution quantitative proteomics of HeLa cells using SILAC-2-DE-nanoLC/LTQ-Orbitrap mass spectrometry. *Molecular & Cellular Proteomics* [Internet]. 2012;**12**(2):529-538. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3567871&tool=pmcentrez&rendertype=abstract>
- [43] Naryzhny SN, Maynskova MA, Zgoda VG, Ronzhina NL, Novikova SE, Belyakova NV, Kleyst OA, Legina OK, Pantina RAFM. Proteomic profiling of high-grade glioblastoma using virtual-experimental 2DE. *Journal of Proteomics and Bioinformatics*. 2016;**9**(6):158-165
- [44] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially Modified Protein Abundance Index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics* [Internet]. 2005;**4**(9):1265-1272. Available from: <http://www.mcponline.org/content/4/9/1265.abstract>
- [45] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*. 2016;**11**(12):2301-2319
- [46] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*. 2008;**26**(12):1367-1372
- [47] Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular Systems Biology*. 2008;**4**(222):222
- [48] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* [Internet]. 1999;**17**(10):994-999 Available from: <http://dx.doi.org/10.1038/13690>
- [49] Kusebauch U, Campbell DS, Deutsch EW, Chu CS, Spicer DA, Brusniak MY, Slagel J, Sun Z, Stevens J, Grimes B, Shteynberg D, Hoopmann MR, Blattmann P, Ratushny AV, Rinner O, Picotti P, Carapito C, Huang CY, Kapousouz M, Lam H, Tran T, Demir E, Aitchison JD, Human MR. SRMAtlas: A resource of targeted assays to quantify the complete human proteome. *Cell*. 2016;**166**(3):766-778

- [50] Cox J, Hein MY, Lubner CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics* [Internet]. 2014;**13**(9):2513-2526 Available from: <http://www.mcponline.org/lookup/doi/10.1074/mcp.M113.031591>
- [51] Blein-Nicolas M, Zivy M. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochimica et Biophysica Acta – Proteins and Proteomics*. 2015
- [52] Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. *Methods in Molecular Biology* [Internet]. 2007;**406**:89-112. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18287689>
- [53] Perez-Riverol Y, Vizcaíno JA. Synthetic human proteomes for accelerating protein research. *Nature Methods*. 2017;**14**:240-242
- [54] Pruess M, Fleischmann W, Kanapin A, Karavidopoulou Y, Kersey P, Kriventseva E, et al. The proteome analysis database: A tool for the in silico analysis of whole proteomes. *Nucleic Acids Research*. 2003;**31**:414-417
- [55] Ueberle B, Frank R, Herrmann R. The proteome of the bacterium *Mycoplasma pneumoniae*: Comparing predicted open reading frames to identified gene products. *Proteomics*. 2002;**2**:754-764
- [56] Jungblut PR, Hecker M, editors. *Proteomics of microbial pathogens*. Wiley-VCH Verlag Weinheim; 2007. 326 p. DOI: 10.1002/9783527610099.ch6
- [57] Pleißner KP, Eifert T, Jungblut PR. A European pathogenic microorganism proteome database: Construction and maintenance. *Comparative and Functional Genomics*. 2002;**3**(2):97-100
- [58] Drews O, Görg A. DynaProt 2D: An advanced proteomic database for dynamic online access to proteomes and two-dimensional electrophoresis gels. *Nucleic Acids Research*. 2005;**33**(Database Issue):D583-D587