Biomeditsinskaya Khimiya, 2024 vol. 70, issue 5, pp. 342–348.

EXPERIMENTAL STUDIES

DETECTION OF LOW-COPY PROTEINS IN PROTEOMIC STUDIES: ISSUES AND SOLUTIONS

A.I. Archakov, N.E. Vavilov*, V.G. Zgoda

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: n.vavilov95@gmail.com

Detection of low-copy proteins in complex biological samples is one of the most important issues of modern proteomics. The main reason for inefficient detection of low protein concentrations is the insufficient sensitivity of mass spectrometric detectors and the high dynamic range of protein concentrations. In this study we have investigated the possibilities and limitations of a targeted mass spectrometric analysis using the reconstructed system of standard proteins UPS1 (Universal Proteomic Standard 1) as an example. The study has shown that the sensitivity of the method is affected by the concentration of target proteins of the UPS1 system, as well as by a high level of biological noise modelled by proteins of whole *E. coli* cell lysate. The limitations of the method have been overcome by concentrating and pre-fractionating the sample peptides in a reversed phase chromatographic system under alkaline elution conditions. Proteomic analysis of the biological sample (proteins of the human hepatocellular carcinoma cell line HepG2 encoded by genes of human chromosome 18) showed an increase in the sensitivity of the method as compared to the standard targeted mass spectrometric analysis. This culminated in registration of 94 proteins encoded by genes located on human chromosome18.

Key words: proteomics; mass spectrometry; HepG2; human chromosome 18 gene products

DOI: 10.18097/PBMC20247005342

INTRODUCTION

The Human Proteome Project was initiated in 2010 at the International Congress in Sydney [1]. The main goal of the project was the search and description of proteins and their corresponding mRNAs, encoded by the human genome. The Chromosome-Centric Human Proteome Project, C-HPP, was also proposed, in which participating countries were invited to study the gene products of a certain human chromosome [2]. The main goal of the project is to detect all proteins encoded by human genes. The successful realization of the project will expand existing knowledge of human biology at the cellular level, which in turn will serve as the basis for future prognostic, diagnostic, therapeutic, and preventive medical applications based on protein data obtained during the project. A group of researchers from Russia led by A.I. Archakov has chosen human chromosome 18 to search for and describe the products of protein-coding genes. During the project, a mass spectrometric method for recording and measuring the absolute concentration of proteins has been developed; it is characterized by high sensitivity and selectivity. The developed method was tested on various types of human biomaterial: the human hepatocellular carcinoma cell line HepG2, human liver cells, and human blood serum [3]. The registered proteins encoded by the genes located on human chromosome 18 cover 12 orders of concentration from 10^{-6} M to 10^{-18} M. Detection of ultra-low concentrations of proteins was achieved by the method of irreversible binding of analytes in solution [4].

However, during the work on the project it became clear that modern proteomic methods for detecting proteins did not have sufficient sensitivity required for analysis of human biological samples [5]. For example, using methods of shotgun mass spectrometry for analysis of a complex sample of human hepatocellular carcinoma cells it is possible to register about 20,000 unique peptides. However, proteomic software also detects and counts unregistered precursor ions in the studied sample. Typically, the number of unregistered peptides is much higher, amounting to approximately 120,000 precursors. Thus, the shotgun technique provides a high number of identifications, but the minimum protein concentration usually does not exceed 10^{-9} M [6]. Methods of targeted mass spectrometry are characterized by increased sensitivity. However, to maintain a high level of sensitivity, it is necessary to limit the number of detected proteins to 200–300 per method. Targeted mass spectrometry employing internal standards in combination with prefractionation, which was used to obtain a deep proteome of HepG2 human hepatocellular carcinoma cells, resulted in detection of 264 proteins encoded by genes of human chromosome 18. In order to determine the sensitivity of the technique at different signal-to-noise ratios, the simplified UPS1 protein system has been used.

MATERIALS AND METHODS

Reagents

The following reagents were used in the study: formic acid (Sigma-Aldrich, Switzerland),

trifluoroacetic acid, 2-chloroacetamide, tris-2-carboxyethylphosphine, sodium deoxycholate, DMEM, Universal Proteomic Standard 1 (UPS1), phosphate buffer (Sigma-Aldrich, USA), urea, sodium chloride (Acros Organics, USA), acetonitrile (Fisher Chemical, China), triethylammonium bicarbonate (TAB) (Fluka, Switzerland), modified pig trypsin (Promega, USA), fetal bovine serum (FBS) (Biowest, France), penicillin/streptomycin (Gibco, USA), trypsin/EDTA sterile solution (Paneco, Russia).

Cultivation of HepG2 Human Hepatocellular Carcinoma Cells

The HepG2 cell line was obtained from the cell bank of the Institute of Biomedical Chemistry (IBMC, Moscow, Russia). The cell culture was grown in a medium supplemented with DMEM, 10% FBS, and antibiotics (penicillin/streptomycin 100 units/ml). Cells were cultivated in an incubator at 37°C and 5% CO_2. During cultivation, the medium was renewed every 2–3 days. When the cells were more than 80% confluent they were collected using 0.25% trypsin/EDTA to obtain a cell suspension.

Lysis and Hydrolysis of HepG2 Cell Samples

The cell sediment was washed to remove serum in 100 mM phosphate buffered saline pH 7.4. After centrifugation (10 min, 500 *g*), the supernatant was removed and 10 μl of lysis buffer (4 M urea, 1% sodium deoxycholate, 15% acetonitrile, 5 mM triscarboxyethylphosphine, 100 mM phosphate buffer pH 7.4, 100 mM chloride sodium) was added to the cell sediment. The solution was sonicated to reduce the viscosity (5 cycles of on the Q125 Sonicator (Qsonica Sonicators, USA): 10 s — sonication, 30 s — pause with power of 30%). The homogenate was then heated to 60°C for 30 min. After cooling to room temperature, 2-chloroacetamide was added at a final concentration of 50 mM. Incubation was carried out at room temperature in the dark. After this, 90 μl of 100 mM TAB and 1 μg of trypsin were added and incubated at 38°C; after 3 h, another aliquot of trypsin (1 μg) was added. The total hydrolysis time was 18 h. The reaction was stopped by adding formic acid to a final concentration of 2%, and then samples were centrifuged (10 min, 10,000 g) was performed to obtain supernatant. The supernatant was transferred into a glass vial (Agilent, Germany), evaporated in a vacuum concentrator (Concentrator plus, Eppendorf, Germany) and redissolved in 0.1% formic acid.

The preparation of UPS1 system proteins for mass spectrometric analysis was carried out as described previously [7].

Mass Spectrometry Analysis

The samples were analyzed using a 6495 Triple Quad LC/MS instrument (Agilent, USA). Its settings and the detection method were the same as described previously [8].

Visualization and processing of the results of the targeted analysis were carried out using Skyline software (version 23.1) as described previously [8].

Transcriptomic Analysis

Total RNA was isolated from HepG2 cells using a commercial Extract RNA kit (Evrogen, Russia). The quality of isolation was controlled spectrophotometrically using a Bioanalyzer 2100 System (Agilent Technologies, USA). Transcriptomic analysis was performed on an Illumina HiSeq 2500 system instrument (Illumina, USA) according to the manufacturer's protocol. For each replicate, between 32 and 59 million reads were obtained [9].

Cultivation of E. coli K-12 Cells

Cells of the bacterial culture were grown according to the standard protocol. Protein isolation and preparation for mass spectrometry analysis were performed as described previously [7].

RESULTS AND DISCUSSION

Determination of the Sensitivity and Limitations of Targeted Mass Spectrometry (SRM) at Different Protein Concentrations of the Simplified Protein System UPS1

The low depth of coverage of proteins encoded by the genes of human chromosome 18 is explained by the insufficient sensitivity of modern mass spectrometry methods. The reason for the low sensitivity may be due to the simultaneous elution and ionization of a huge number of peptides, among which only a small part can be detected [10, 11]. In order to determine the operating range of the method and ways to overcome the lack of sufficient sensitivity, a model system simulating a biological sample was constructed.

The simplified protein system UPS1 is an equimolar mixture of 47 recombinant, highly purified human proteins; it is used for the development and validation of new methods for protein analysis [12, 13]. UPS1 is used as a model to simulate the target peptide signal in a biological sample. In order to exclude signal interference between human proteins and UPS1, another biological object, *E. coli* strain K-12, was used as noise.

Thus, using these components, 3 different versions of the model system with different concentrations of target UPS1 proteins were generated: 1) UPS1 solution without addition of *E. coli* preparation; 2) UPS1 solution with addition of *E. coli* preparation, in which concentrations of UPS1 and *E. coli* preparation decreased proportionally; 3) UPS1 solution with addition of *E. coli* preparation in the same maximum permissible (high) concentration.

A series of dilutions with a low signal-to-noise value most closely simulated the biological effect of the matrix, realized during analysis of complex biological samples [14].

The results of targeted mass spectrometric analysis demonstrate that in the high-concentration region of 10^{-9} M, the method had 100% sensitivity regardless of the presence of biological noise in the form of *E. coli* preparation in the sample (Fig. 1). The most pronounced decrease in the number of identifications of target proteins of the UPS1 system was observed when their concentration decreased to 10^{-10} M; in the sample with the highest noise level there were 26 identifications, in a pure UPS1 solution and in a sample with a low noise level there were 45 and 44 identifications, respectively. This trend continued during the further decrease in the concentration of target proteins. In the presence of a high level of biological noise, even at a concentration of 10^{-12} M, it was not possible to register a single protein of the UPS1 system; at the same time, in the absence or low level of noise, it was still possible to register 10 and 8 proteins, respectively. However, at a concentration of 10^{-13} M, even in the absence of biological noise or at its low levels, not a single target protein was detected. Therefore, the results obtained using the UPS1 model object can be extrapolated to proteins encoded by genes of chromosome 18, and the number of proteins that can be detected using targeted mass spectrometric analysis can be predicted. Thus, we can conclude that using this method it is possible to detect 100% of proteins at 10^{-9} M concentration or higher, 56% of proteins at 10^{-10} M concentration, and only 19% of proteins at a concentration of 10^{-11} M. It appears that this method is not applicable for detecting lower concentrations of proteins in in complex biological samples. Thus, the influence of two factors that affect the registration of proteins has been demonstrated: the concentration of target proteins and the presence of a high noise level.

In order to demonstrate the effect of the concentration of target proteins on the number of detected proteins, an experiment with reverse concentrating of the diluted protein mixture was also performed. UPS1 samples with and without added *E. coli* preparations at a concentration of 10^{-12} M were concentrated 100-fold to 10^{-10} M and analyzed by targeted mass spectrometry. This approach restored the number of detected proteins in the case of a pure UPS1 solution and a solution with a low noise level to 45 and 44, respectively. At the same time, concentrating a sample with a high noise level did not improve the result, because the matrix proteins were concentrated along with the target proteins. To reduce the influence of the matrix, it is necessary to apply a sample fractionation technique. The use of prefractionation by reverse-phase high-performance liquid chromatography under alkaline elution conditions restored the number of identifications of proteins of the UPS1 system to 26 (Table 1).

Thus, the combination of sample concentration and fractionation methods solves the problem of low concentration of target proteins and also reduces the noise level; this increases the number of identifications in the reconstructed UPS1/*E. coli* protein system. Therefore, the use of this method on a biological sample, the human hepatocellular carcinoma cell line HepG2, can also increase the coverage of proteins encoded by the genes located on human chromosome 18.

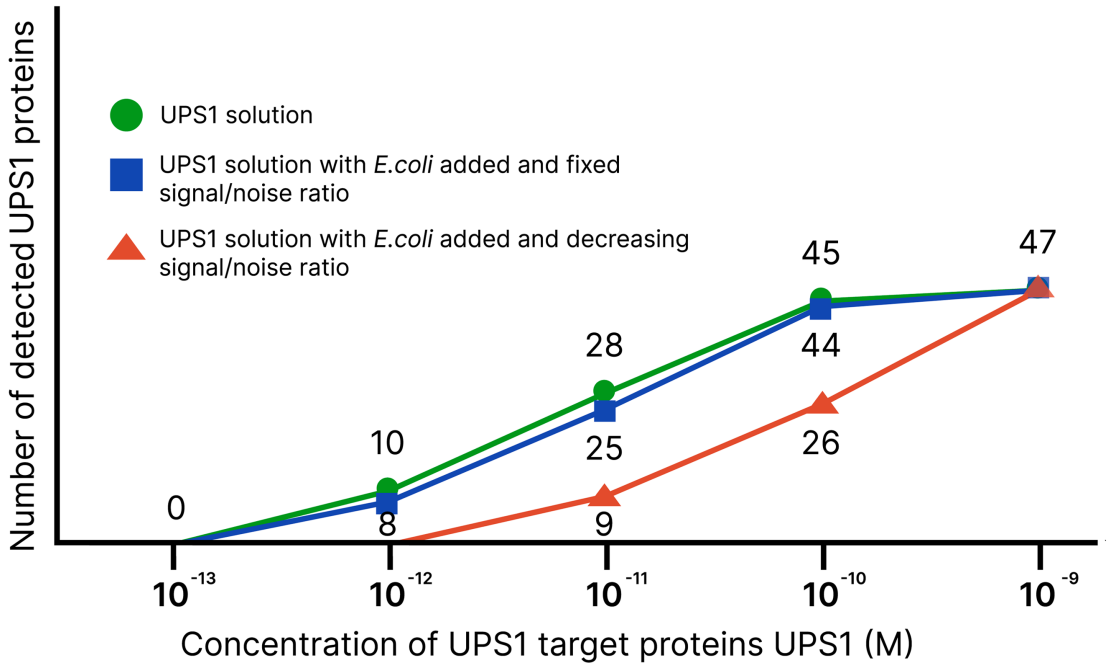

Figure 1. The number of detected proteins of the UPS1 system in dependence of the target protein concentration in the model samples.

Table 1. Results of the experiment on concentrating UPS1 samples

Sample	The number of proteins registered before sample concentration	The number of proteins registered after 100-fold sample concentration
UPS1 solution	10	45
UPS1 solution supplemented with *E. coli* with fixed signal-to-noise ratio	8	44
UPS1 solution supplemented with *E. coli* with decreasing signal-to-noise ratio	0	26*

*Biochemical fractionation of a sample into 12 fractions was carried out.

Targeted Mass Spectrometry Analysis (SRM SIS) of Proteins Encoded by Genes Located on Human Chromosome 18: Focus on HepG2 Human Hepatocellular Carcinoma Cells

For detection of proteins encoded by the genes located on human chromosome 18, the method of targeted mass spectrometric analysis was used in combination with prefractionation under alkaline elution conditions [7]. It was shown that the use of an additional fractionation step resulted in detection of 26 unique proteins, which were not detected in the unfractionated sample (Fig. 2). Based on the results of using both methods, a total of 94 proteins encoded by the genes located on human chromosome 18 were registered. The majority of proteins (66%) were detected in both 1D and 2D analysis, but a small proportion of proteins (6.4%) was identified exclusively by the SRM analysis. This phenomenon, as well as the significant difference in the measured concentrations of some proteins recorded in both 1D and 2D analysis is explained by the scheme for collecting fractions during prefractionation. Fractions are collected continuously and during switching to the next fraction, part of the chromatographic peak of the peptide can be collected in one fraction, and another part in the next. Thus, 6 peptides could appear in two different fractions, and their concentration could be below the sensitivity limit of the device.

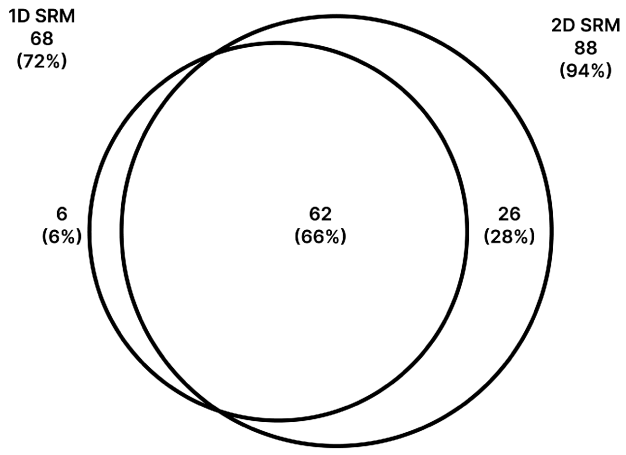

Figure 2. The Euler-Venn diagram of comparison of sets of unique proteins found during 1D SRM of an unfractionated sample and 2D SRM of a fractionated sample.

We could not observe this effect during analysis of proteins of the UPS1 system. This may have been hindered by the limited sample of 47 proteins; in this case, we observed 94 proteins.

The range of measured concentrations varies from 10^{-8} M to 10^{-12} M or from 300,000 to 20 copies of protein per cell. Among the proteins detected only by using targeted analysis with prefractionation, there were transcription factors (Q92908 transcription factor GATA-6) and protein kinases (Q13464-Rho associated protein kinase 1, P31152 mitogen-activated protein kinase 4), responsible for signal transduction in the cell and regulation of gene expression [15–17]. Consequently, using this technique it is possible to obtain additional biologically significant information about the studied object, which is not available in the case of using standard methods of analysis.

Transcriptoproteomic Analysis of HepG2 Cells

The data used in this section represent results from high-throughput RNA sequencing performed on the Illumina platform. For qualitative assessment of the convergence of the proteome and transcriptome of the HepG2 cells, different cutoffs were selected based on the RPKM value (Reads Per Kilobase Million) of the transcriptomic data, which were compared with the corresponding proteins detected by targeted mass spectrometry methods [18]. The highest convergence of proteomic and transcriptomic data is shown by the cutoff at the level of RPKM≤1 (48%). At this cutoff, the number of registered transcripts was 145, while the number of corresponding proteins was 94 (Table 2). At the same time, correlation analysis of quantitative proteomic and transcriptomic data at appropriate RPKM cutoff levels showed a tendency for the resulting correlation coefficient to decrease as the RPKM cutoff values increased from 0.6 to 0.4. Data on protein concentration measurements and RPKM of the corresponding transcripts are presented in the Supplementary Materials.

In this study, high-throughput RNA sequencing data have been used to guide the expression of the corresponding protein in the HepG2 cells. By analogy with the simplified UPS1 protein system, where the manufacturer guaranteed the presence of 47 human proteins, in this case, transcript detection was considered as a guaranteed sign

Table 2. Percent ratio of detected unique proteins, transcripts and jointly registered products encoded by genes of human chromosome 18 at various RPKM cutoff levels and correlation coefficient for quantitative proteomic and transcriptomic data at corresponding cutoff level

RPKM cutoff level (≥)	Simultaneously registered proteins and transcripts [%]	Registered unique transcripts [%]	Registered unique proteins [%]	Spearman correlation coefficient (rs) for quantitative proteomic and transcriptomic data	*p*-value significance level
0.01	42	54	3	0.6	6.3×10^{-9}
0.1	43	51	6	0.6	7.0×10^{-8}
0.5	47	44	9	0.6	6.0×10^{-8}
1	48	42	10	0.6	2.0×10^{-7}
5	40	28	31	0.5	0.1×10^{-3}
10	31	20	50	0.4	0.2×10^{-1}
50	8	2	90	0.4	0.4

("gold standard") of the presence of the corresponding protein. For 76 proteins, both the transcript and the corresponding protein were reported. However, no corresponding protein products were detected for 67 transcripts (Fig. 3). In contrast to the UPS1 set, we cannot know protein concentrations in the cells. Even if they are present in the cell, their concentration may be lower than the working range of the SRM assay. However, 16 proteins were reported for which no corresponding transcripts were found. The most probable explanation for this phenomenon is the longer half-life and stability of the protein molecule relative to corresponding mRNA [19, 20]. During mRNA isolation some molecules may degrade despite all precautions, while proteins that are more stable in their chemical structure are preserved. It should be also noted that in a living cell, the half-life of proteins is many times higher than the half-life of the corresponding mRNA molecules [21, 22]. Measurements of the half-life of proteins show that, on average, proteins perform their functions for tens of hours before their degradation, while half-life of mRNA molecules is much shorter (seconds or minutes) [23, 24]. Thus, detection of the protein corresponding to the particular mRNA is determined both by the sensitivity of the proteomic analysis and by the half-life of the mRNA molecule in the cell.

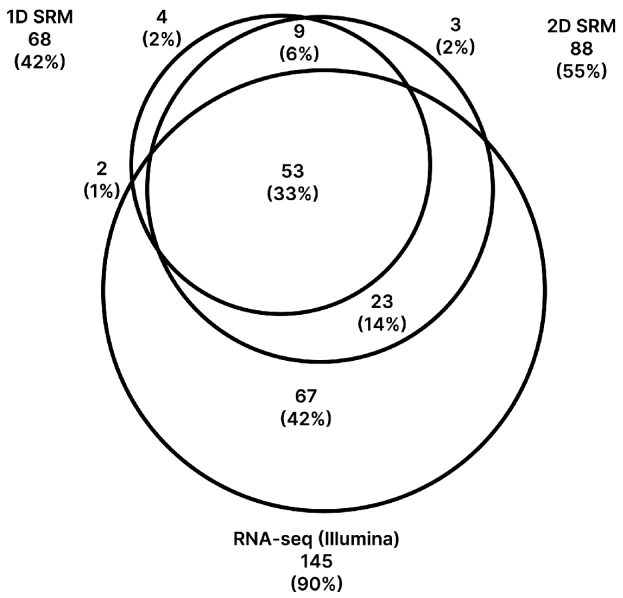

Figure 3. The Euler-Venn diagram representing comparisons of three sets, transcripts registered in the HepG2 sample by RNA-seq, proteins encoded by genes of chromosome 18, registered by 1D SRM, proteins registered by 2D SRM.

The results of this study show that the use of targeted analysis together with the method of prefractionation of the sample increases the depth of coverage of proteins encoded by genes located on chromosome 18. Using this approach we have registered 94 proteins, which constitute 48% of the registered transcriptome and 34% of all protein-coding genes localized on human chromosome 18.

CONCLUSIONS

The main problem of targeted analysis of complex biological samples is the presence of a high dynamic range of protein concentrations in the studied system. Experiments with the reconstituted UPS1 system have shown that the sensitivity of the SRM assay depends on the concentration of target proteins, as well as the presence of high levels of biological noise. The use of 2D fractionation has partially solved the sensitivity problem and reduced the dynamic range of proteins.

During analysis of proteins encoded by genes located on human chromosome 18 in the HepG2 human hepatocellular carcinoma cells 264 proteins are targeted, while the remaining proteins, the number of which can exceed 6 million, are the matrix or, in other words, "biological noise" [25]. Thus, the use of sample prefractionation helps to enrich each fraction with target proteins and increase the signal-to-noise ratio in each fraction. Low concentration proteins (10^{-10} M – 10^{-12} M) represent the majority (76 of 94) of the human chromosome 18 proteome in the HepG2 cells. Consequently, during analysis of proteins at a concentration of 10^{-10} M and lower, we cannot be sure that all proteins that are expressed in the cell have been recorded, as the losses in the reconstructed UPS1 protein system at a similar

concentration were 50%. At the same time, there were 67 transcripts of gene products for which the corresponding proteins were not registered in the studied cells. The findings suggest that prefractionation can detect additional proteins, but still does not reach the level of sensitivity of transcriptomic methods.

ACKNOWLEDGMENTS

The study was carried out using the equipment of the Center of Collective use at the Institute of Biomedical Chemistry.

FUNDING

The work was carried out within the framework of the Program of Fundamental Scientific Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100168-2).

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any research involving humans or the use of animals as objects.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

Supplementary materials are available in the electronic version at the journal site (pbmc.ibmc.msk.ru).

REFERENCES

1. *Anonymous* (2010) The call of the human proteome. Nat. Methods, **7**(9), 661. DOI: 10.1038/nmeth0910-661
2. *Hancock W., Omenn G., Legrain P., Paik Y.K.* (2011) Proteomics, human proteome project, and chromosomes. J. Proteome Res., **10**(1), 210. DOI: 10.1021/pr101099h
3. *Zgoda V.G., Kopylov A.T., Tikhonova O.V., Moisa A.A., Pyndyk N.V., Farafonova T.E., Novikova S.E., Lisitsa A.V., Ponomarenko E.A., Poverennaya E.V., Radko S.P., Khmeleva S.A., Kurbatov L.K., Filimonov A.D., Bogolyubova N.A., Ilgisonis E.V., Chernobrovkin A.L., Ivanov A.S., Medvedev A.E., Mezentsev Y.V., Moshkovskii S.A., Naryzhny S.N., Ilina E.N., Kostrjukova E.S., Alexeev D.G., Tyakht A.V., Govorun V.M., Archakov A.I.* (2013) Chromosome 18 transcriptome profiling and targeted proteome mapping in depleted plasma, liver tissue and HepG2 cells. J. Proteome Res., **12**(1), 123. DOI: 10.1021/pr300821n
4. *Kopylov A.T., Zgoda V.G., Lisitsa A.V., Archakov A.I.* (2013) Combined use of irreversible binding and MRM technology for low- and ultralow copy-number protein detection and quantitation. Proteomics, **13**(5), 727. DOI: 10.1002/pmic.201100460
5. *Anderson N.L., Polanski M., Pieper R., Gatlin T., Tirumalai R.S., Conrads T.P., Veenstra T.D., Adkins J.N., Pounds J.G., Fagan R., Lobley A.* (2004) The human plasma proteome: A nonredundant list developed by combination of four separate sources. Mol. Cell. Proteomics, **3**(4), 311. DOI: 10.1074/mcp.M300127-MCP200
6. *Michalski A., Cox J., Mann M.* (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. J. Proteome Res., **10**(4), 1785. DOI: 10.1021/pr101060v
7. *Vavilov N., Ilgisonis E., Lisitsa A., Ponomarenko E., Farafonova T., Tikhonova O., Zgoda V., Archakov A.* (2022) Number of detected proteins as the function of the sensitivity of proteomic technology in human liver cells. Curr. Protein Pept. Sci., **23**(4), 290. DOI: 10.2174/1389203723666220526092941
8. *Vavilov N.E., Zgoda V.G., Tikhonova O.V., Farafonova T.E., Shushkova N.A., Novikova S.E., Yarygin K.N., Radko S.P., Ilgisonis E.V., Ponomarenko E.A., Lisitsa A.V., Archakov A.I.* (2020) Proteomic analysis of Chr 18 proteins using 2D fractionation. J. Proteome Res., **19**(12), 4901. DOI: 10.1021/acs.jproteome.0c00856
9. *Deinichenko K., Krasnov G., Radko S., Ptitsyn K., Shapovalova V., Timoshenko O., Khmeleva S., Kurbatov L., Kiseleva Y., Ilgisonis E., Pyatnitskiy M., Poverennaya E., Kiseleva O., Vakhrushev I., Tsvetkova A., Buromski I., Markin S., Zgoda V., Archakov A., Lisitsa A., Ponomarenko E.* (2021) Human CHR18: "Stakhanovite" genes, missing and uPE1 proteins in liver tissue and HepG2 cells. Biomedical Chemistry: Research and Methods, **4**(1), e00144. DOI: 10.18097/BMCRM00144
10. *Furey A., Moriarty M., Bane V., Kinsella B., Lehane M.* (2013) Ion suppression: A critical review on causes, evaluation, prevention and applications. Talanta, **115**, 104. DOI: 10.1016/j.talanta.2013.03.048
11. *Mei H., Hsieh Y., Nardo C., Xu X., Wang S., Ng K., Korfmacher W.A.* (2003) Investigation of matrix effects in bioanalytical high-performance liquid chromatography/ tandem mass spectrometric assays: Application to drug discovery. Rapid Commun. Mass Spectrom., **17**(1), 97. DOI: 10.1002/rcm.876
12. *Muller L., Fornecker L., van Dorsselaer A., Cianférani S., Carapito C.* (2016) Benchmarking sample preparation/ digestion protocols reveals tube-gel being a fast and repeatable method for quantitative proteomics. Proteomics, **16**(23), 2953. DOI: 10.1002/pmic.201600288
13. *Ramus C., Hovasse A., Marcellin M., Hesse A.M., Mouton-Barbosa E., Bouyssié D., Vaca S., Carapito C., Chaoui K., Bruley C., Garin J., Cianférani S., Ferro M., van Dorssaeler A., Burlet-Schiltz O., Schaeffer C., Couté Y., Gonzalez de Peredo A.* (2016) Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. J. Proteomics, **132**, 51. DOI: 10.1016/j.jprot.2015.11.011
14. *Matuszewski B.K., Constanzer M.L., Chavez-Eng C.M.* (2003) Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal. Chem., **75**(13), 3019. DOI:10.1021/ac020361s
15. *Kodo K., Nishizawa T., Furutani M., Arai S., Yamamura E., Joo K., Takahashi T., Matsuoka R., Yamagishi H.* (2009) GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. Proc. Natl. Acad. Sci. USA, **106**(33), 13933. DOI: 10.1073/pnas.0904744106

16. *Maekawa M., Ishizaki T., Boku S., Watanabe N., Fujita A., Iwamatsu A., Obinata T., Ohashi K., Mizuno K., Narumiya S.* (1999) Signaling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase. Science, **285**(5429), 895. DOI: 10.1126/science.285.5429.895
17. *Déléris P., Trost M., Topisirovic I., Tanguay P.L., Borden K.L., Thibault P., Meloche S.* (2011) Activation loop phosphorylation of ERK3/ERK4 by group I p21-activated kinases (PAKs) defines a novel PAK-ERK3/4-MAPK-activated protein kinase 5 signaling pathway. J. Biol. Chem., **286**(8), 6470. DOI: 10.1074/jbc.M110.181529
18. *Koch C.M., Chiu S.F., Akbarpour M., Bharat A., Ridge K.M., Bartom E.T., Winter D.R.* (2018) A beginner's guide to analysis of RNA sequencing data. Am. J. Respir. Cell Mol. Biol., **59**(2), 145. DOI: 10.1165/rcmb.2017-0430TR
19. *Reimegård J., Tarbier M., Danielsson M., Schuster J., Baskaran S., Panagiotou S., Dahl N., Friedlander M.R., Gallant C.J.* (2021) A combined approach for single-cell mRNA and intracellular protein expression analysis. Commun. Biol., **4**(1), 624. DOI: 10.1038/s42003-021-02142-w
20. *Shyu A.B., Greenberg M.E., Belasco J.G.* (1989) The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev., **3**(1), 60. DOI: 10.1101/gad.3.1.60
21. *Treisman R.* (1985) Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5′ element and c-fos 3′ sequences. Cell, **42**(3), 889. DOI: 10.1016/0092-8674(85)90285-5
22. *Mathieson T., Franken H., Kosinski J., Kurzawa N., Zinn N., Sweetman G., Poeckel D., Ratnu V.S., Schramm M., Becher I., Steidel M., Noh K.M, Bergamini G., Beck M., Bantscheff M., Savitski M.M.* (2018) Systematic analysis of protein turnover in primary cells. Nat. Commun., **9**(1), 689. DOI: 10.1038/s41467-018-03106-1
23. *Bevilacqua A., Ceriani M.C., Canti G., Asnaghi L., Gherzi R., Brewer G., Papucci L., Schiavone N., Capaccioli S., Nicolin A.* (2003) Bcl-2 protein is required for the adenine/uridine-rich element (ARE)-dependent degradation of its own messenger. J. Biol. Chem., **278**(26), 23451. DOI: 10.1074/jbc.M210620200
24. *Yang E., van Nimwegen E., Zavolan M., Rajewsky N., Schroeder M., Magnasco M., Darnell J.E. Jr.* (2003) Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. Genome Res., **13**(8), 1863. DOI: 10.1101/gr.1272403
25. *Ponomarenko E.A., Poverennaya E.V., Ilgisonis E.V., Pyatnitskiy M.A., Kopylov A.T., Zgoda V.G., Lisitsa A.V., Archakov A.I.* (2016) The size of the human proteome: The width and depth. Int. J. Anal. Chem., **2016**, 7436849. DOI: 10.1155/2016/7436849

Received: 27. 04. 2024.
Revised: 10. 06. 2024.
Accepted: 11. 06. 2024.

ДЕТЕКЦИЯ НИЗКОКОПИЙНЫХ БЕЛКОВ В ПРОТЕОМНЫХ ИССЛЕДОВАНИЯХ: ПРОБЛЕМЫ И ПУТИ РЕШЕНИЯ

А.И. Арчаков, Н.Э. Вавилов*, В.Г. Згода

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича, 119121, Москва, ул. Погодинская, 10; *эл. почта: n.vavilov95@gmail.com

Одна из наиболее актуальных проблем современной протеомики — детекция низкокопийных белков в комплексных биологических образцах. Главная причина малоэффективной детекции низких концентраций белков заключается в недостаточной чувствительности масс-спектрометрического детектора и высоком динамическом диапазоне концентраций белков. В данной работе были исследованы возможности и ограничения метода таргетного масс-спектрометрического анализа на примере реконструированной системы стандартных белков UPS1 (Универсальный Протеомный Стандарт 1, Universal Proteomic Standard 1). Показано, что на чувствительность метода влияет концентрация целевых белков системы UPS1, а также высокий уровень биологического шума в виде белков цельного лизата *E. coli*. Ограничения метода удалось преодолеть с помощью концентрирования и предварительного фракционирования пептидов образца в хроматографической системе на обращённой фазе в щелочных условиях элюции. Протеомный анализ биологического образца — белков клеточной линии гепатоцеллюлярной карциномы человека HepG2, кодируемых генами 18 хромосомы человека, — показал повышение чувствительности метода по сравнению со стандартным таргетным масс-спектрометрическим анализом. Это позволило зарегистрировать 94 белка, кодируемых генами 18 хромосомы человека.

Полный текст статьи на русском языке доступен на сайте журнала (http://pbmc.ibmc.msk.ru).

Ключевые слова: протеомика; масс-спектрометрия; HepG2; продукты генов 18 хромосомы человека

Финансирование. Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы) (№ 122030100168-2).

Поступила в редакцию: 27.04.2024; после доработки: 10.06.2024; принята к печати: 11.06.2024.