IDENTIFICATION OF MOUSE BRAIN PROTEOFORMS: COMPARISON OF 2D-ELECTROPHORESIS DATA AND INDEPENDENT EXPERIMENT WITH MASS SPECTROMETRIC IDENTIFICATION

A.V. Rybina

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; e-mail: aleona.rybina@ibmc.msk.ru

A previously developed algorithm for the preliminary identification of protein proteoforms associated with post-translational modifications (PTMs) based on 2D electrophoresis data (DOI: 10.18097/BMCRM00191) has been used in this study for analysis of experimental data obtained using mice and reported in two papers by different authors. The authors of the first paper identified 8 groups of spots on 2D electrophoretic maps corresponding to 8 proteins with at least two unconcretised proteoforms. The authors of the second paper analyzed brain samples by means of the LC-MS/MS. In this study identification of peptides with PTMs was repeated using the raw data from the second paper. Among the 8 target proteins, 7 were identified in most of the biological samples. For 4 of them, 17 possible peptides with modifications were found. The 5 proteoform variants with identified PTMs matched the spots on the 2D electrophoresis maps. Thus, the prediction of pI values for proteins with hypothetical PTMs allows to form a set of hypotheses about the presence of particular proteoforms on the 2D electrophoretic maps.

Key words: bioinformatics; 2D electrophoresis; mass spectrometry; ischemic stroke

DOI: 10.18097/PBMC20247006475

INTRODUCTION

Post-translational modifications (PTMs) of proteins play an important role in the regulation of many biochemical processes in the cell in normal and pathological conditions [1]. Experiments performed using two-dimensional (2D) gel electrophoresis often reveal proteins existing as multiple proteoforms [2]. In the case of the change in the pI value rather than the protein mass, the electrophoretic shift is most likely attributed to the presence of PTMs rather than the presence of proteoforms originated from alternative splicing or degradation of the protein sequence. However, analysis of 2D gel electrophoresis data does not always culminate in detection of PTMs. For example, enzyme immunoassay-based protein detection cannot distinguish between modified and unmodified forms of proteins [2]. Even when tandem mass spectrometry (MS/MS) is performed for analysis of selected 2D gel spots, a researcher does not often need to achieve full coverage of the protein sequence or search for modified sites; in many cases it is usually sufficient to identify 2–3 characteristic peptides. For this reason, researchers rarely configure identification programs to search for a large number of PTMs, as this significantly increases computational time and can be an additional source of errors. At the same time, it is very likely that primary ion masses (and sometimes secondary spectra) for peptides containing modifications are measured and recorded during MS/MS analysis. Previously, a method has been proposed for generation of hypotheses about specific PTMs of proteins identified in a 2D gel electrophoresis map; it is based on the prediction of protein pI values [3]. This paper attempts to evaluate how these predictions work.

METHODS

Results of two experimental studies in mice obtained by different groups of researchers were used as the data for analysis in this study. In one of them, animals were subjected to autohypoxia-induced hypoxic preconditioning (HPC) [4]. After a 2D electrophoresis procedure, brain proteins were identified using the MALDI-TOF MS method. In the other one, brain ischemia was simulated by middle cerebral artery occlusion (MCAO) and brain preparations were analyzed by the LC-MS/MS method [5]. The raw data from [5] are available in the ProteomeXchange database [6] (accession number PXD032141).

2D electrophoresis maps were analyzed using the method described in [3], and isoelectric point value prediction was performed using pIPredict 3.0 software [7] (a variant taking into account the influence of neighboring amino acid residues). Images and annotations of two maps of the cytosolic fraction of the cerebral cortex of control and HPC-treated mice were used. The authors of the experimental work [4] identified 8 groups of points on the 2D electrophoretic maps (Fig. 1), corresponding to 8 proteins with at least two unconcretised proteoforms.

These data were supplemented (Table 1) with information obtained from UniProt on all experimentally confirmed or identified by homology PTM present

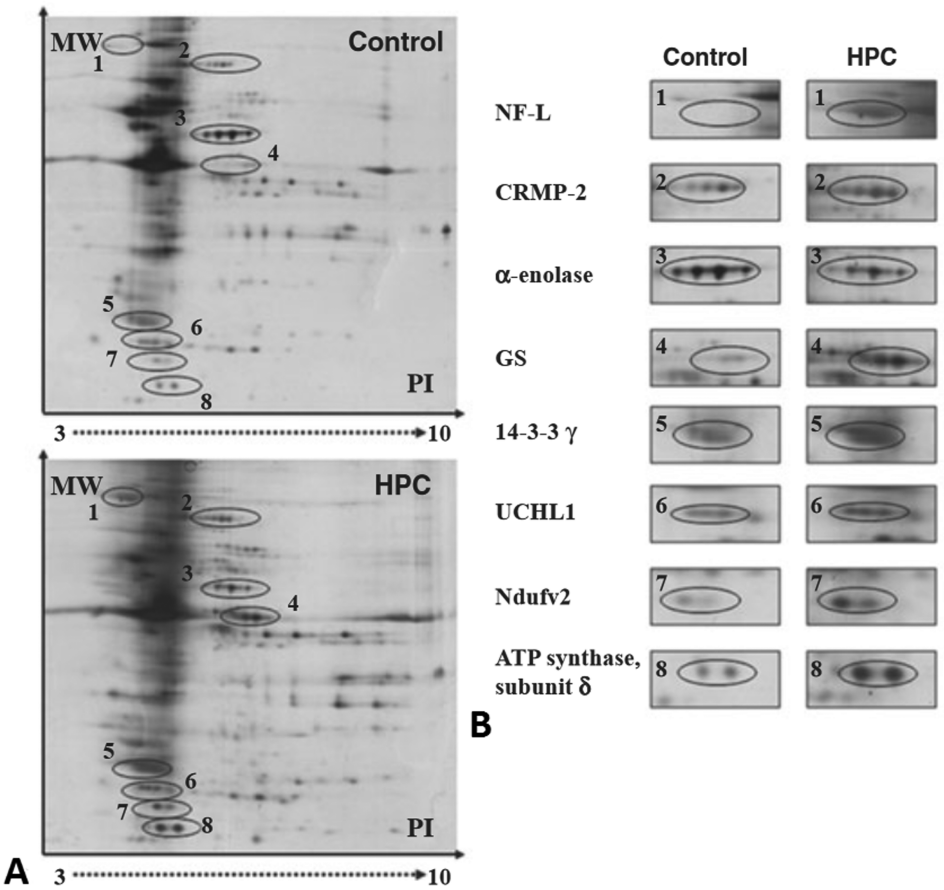

Figure 1. Position on 2D electrophoresis maps (**A**) of proteins with proteoforms identified in [4]. (**B**) is an enlarged image of map fragments. Modified from [4].

in these proteins. The space of all possible peptides that could be produced by trypsin treatment was then generated. Peptides with a single missing cleavage were allowed. For each of these peptides, the retention time (RT) was predicted using the RTP program [8] and a virtual fragmentation spectrum (CID) was generated, which was used for comparison with the data from the second study.

In the present work the following algorithm was used to re-identify peptides from the original data [5]:

1. Peptide identification was performed using PEAKS-X Pro [9] (identification error for primary ions 5 ppm and 0.1 Da for secondary ions, false discovery rate (FDR) 1%). The condition for protein identification was the presence of three unique peptides. Only one data set of 4 (young male mice, 24 biological samples) was used in this work.

2. The Progenesis LC-MS program (Nonlinear Dynamics, UK) was used to align the data from all 24 samples to RT. The RT values were then aligned with the RT values predicted by the RTP program using the non-PTM peptide identification data obtained in step 1.

3. For each of the virtual peptides, the entire set of secondary spectra obtained in [5] was searched for matches using the following parameters: primary ion identification error 2 ppm, RT deviation from the predicted value within a window of 3 min, secondary ion fragment identification error 0.05 Da.

RESULTS AND DISCUSSION

The results of the match search are summarized in Table 2. In the first stage (data analysis using PEAKS-X Pro) only 6 out of 8 proteins were identified and no peptide with the expected PTMs was found. It can be assumed that the problem with identification is that the search algorithm is dynamically tuned to identify the peptide with the best performance when the spectrum is mixed, i.e. contains more than one peptide. A window of 5 ppm is claimed in [5]. Indirect evidence that secondary spectra are predominantly mixed is provided by the fact that a large number of secondary spectra (even with a small total mass of the primary ion) contain several thousand peaks. The data obtained in this step were therefore only used for the subsequent RT alignment of individual samples and it was necessary to use the identification procedure described for step 3.

Among the 8 target proteins, 6 were also identified in all samples, one in more than half of the samples and one in 5 samples (Table 1). For 4 of these proteins,

Table 1. Proteins with proteoforms identified in [4]

#	Code in Figure 1	UniProt ID	Number of spots on the control gel	Number of spots on the HPC gel	UniProt protein name	Number of hypothetical PTM sites (by UniProt)	Putative PTMs (by UniProt)	Number of biological samples with protein identification (*)
1	NF-L	P08551	0	2	Neurofilament light polypeptide	10	N-terminal-acetylation, phosphorylation	24
2	CRMP-2	O08853/P97821	5	5	Dipeptidyl peptidase 1	0	—	5
3	α-enolase	P17182	4	3	Alpha-enolase	22	N-terminal-acetylation, acetylation of K, phosphorylation, succinylation	24
4	GS	P15105	2	2	Glutamine synthetase	3	N-terminal-acetylation, phosphorylation	24
5	14-3-3 γ	P61982	3	2	14-3-3 protein gamma	7	N-terminal-acetylation, phosphorylation	24
6	UCHL1	Q9R0P9	3	3	Ubiquitin carboxyl-terminal hydrolase isozyme L1	1	phosphorylation	24
7	Ndufv2	Q9D6J6	2	2	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial precursor	2	acetylation of K, phosphorylation	14
8	ATP synthase, subunit δ	Q9DCX2	2	2	ATP synthase subunit δ, mitochondrial	11	N-terminal-acetylation, acetylation of K, succinylation	24

(*) Identification was performed in that paper.

17 peptides with the expected PTMs were found (Table 2). With this search variant it is difficult to say how many fragments of the secondary spectrum are sufficient for identification. The table includes variants with at least 4 (y or b ions). Some of the peptides also have a matching fragment due to the presence of "missing cleavage" or confirmed different PTM positions. For at least 2 proteins it can be said that the peptides found with PTMs may correspond to proteoforms represented in 2D maps (Table 2). In another case (for glutamine synthetase, UniProt ID P15105), the possible proteoform corresponds to the same darkened region where the spots on the 2D map are highlighted. Since it corresponds to the N-terminal acetylation, it can be assumed that the "zero" point (position on the electrophoretic map of the unmodified protein) is misidentified and is not pronounced and should be shifted to the basic region. For alpha-enolase (P17182), 7 PTM variants were found in 11 peptides (Table 2 and Fig. 2); some of them corresponded to proteoforms on 2D maps. It should be noted that their combination can result in proteoforms that are also represented on 2D maps. Alpha-enolase seems to be subjected to multiple acetylation, as there are spots that correspond to proteoforms with triple acetylation and even with triple acetylation and phosphorylation, while there is no point in the single acetylation site on the map.

In some of the identified peptides acetylation of the C-terminal lysine was detected. Although this conflicts trypsin specificity rules, it is nevertheless known that trypsin can (albeit with much lower efficiency) hydrolyze not only after positively charged residues [10]. For some of these peptides, we also found forms where this modified amino acid residue is located in the internal part of the amino acid sequence, which may serve as additional evidence for the existence of this PTM.

Table 2. Peptides with PTMs identified in this work. Comparison with the experimental position of the spot on the gel was made with the calculated values for the whole protein with the detected modification

#	Protein	Peptide	Sequence position of PTM	Maximum of coincident MS/MS fragments	Number of biological samples with the identified peptide	Matching the proteoform variant with these PTMs to the spots on the gel
1	P08551	S(+80)YSSSSGSLMPSLENLDLSQVAAISNDLK	56	12	11	Not present.
2	P17182	AAVPSGASTGIY(+80)EALELR	44	16	5	Present (most likely).
		GLFRAAVPSGASTGIY(+80)EALELR		4	2	
3	P17182	EALELLK(+100)TAIAK	228	10	8	Present (less likely than phosphorylation). Completely identical in pI to double acetylation.
4	P17182	KVNVVEQEK(+100)	89	5	2	Present (less likely than phosphorylation). Completely identical in pI to double acetylation.
5	P17182	FMGK(+42)GVSQAVEHINK(+42)	60; 71	6	3	Present (less likely than phosphorylation).
6	P17182	EALELLK(+42)TAIAK	228	6	8	Not present (as the only modification, but in combination with other PTMs possible).
7	P17182	EALELLKTAIAK(+42)	233	7	8	Not present (as the only modification, but in combination with other PTMs possible).
		TAIAK(+42)AGYTDQVVIGMDVAASEFYR		6	1	
8	P17182	VNVVEQEK(+42)IDK(+42)	89; 92	9	9	Present (less likely than phosphorylation).
9	P17182	YITPDQLADLYK(+42)	281	14	8	Not present (as the only modification, but in combination with other PTMs possible).
		YITPDQLADLYK(+42)SFVQNYPVVSIEDPFDQDDWGAWQK		5	1	
10	Q9DCX2	ANVAK(+42)PGLVDDFEK(+42)	63; 72	14	9	Not present in control, possibly present in HPC (equal probability with single acetylation).
		ANVAK(+42)PGLVDDFEK(+42)K		9	7	
11	Q9DCX2	MRNIIPFDQMTIDDLNEIFPETK(+42)	144	8	7	Present.
		NIIPFDQMTIDDLNEIFPETK(+42)		5	1	
12	Q9DCX2	IPVPEDK(+42)YTALVDQEEK(+100)	85; 95	17	8	Not present.
13	Q9DCX2	IPVPEDK(+42)YTALVDQEEK(+42)	85; 95	4	5	Not present in control, possibly present in HPC (equal probability with single acetylation).
14	Q9DCX2	K(+100)YPYWPHQPIENL	149	13	3	Not present in control, possibly present in HPC (equal probability with single acetylation).
15	Q9DCX2	YNALK(+100)IPVPEDK(+100)	78; 85	5	2	Not present.
16	Q9DCX2	IPVPEDK(+100)YTALVDQEEK(+42)	78; 85	15	9	Not present.
17	P15105	A(+42)TSASSHLNKGIK	1	4	4	Not present (but is within the blurred area, spot undefined).

PTM: (+80) – phosphorylation, (+42) – acetylation, (+100) – succinylation.

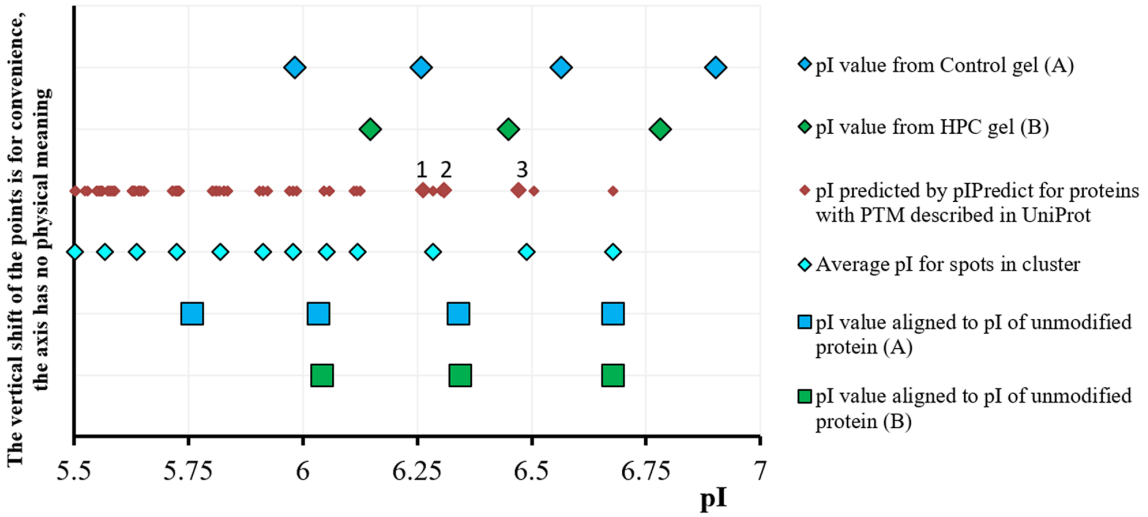

Figure 2. Analysis of control (A) and HPC (B) electrophoretic maps for protein P17182. The numbered zones correspond to the proteoform variants: 0 – no PTM, 1 – single acetylation, 2 – single phosphorylation, 3 – single succinylation or double acetylation, 4 – triple acetylation, 5 – triple acetylation and single phosphorylation.

CONCLUSIONS

Thus, using pI value prediction for proteins with hypothetical PTMs it is possible to form a set of hypotheses about particular proteoforms observed in the 2D electrophoresis maps. In this work, for 3 of the 8 proteins, 5 variants of proteoforms with identified PTMs that coincided with spots on the 2D electrophoresis maps could be hypothesized. In addition, using combined techniques (2D electrophoresis and LC-MS/MS), it is possible to specify a given region on the gel to be analyzed to identify specific proteoforms.

FUNDING

The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100170-5).

COMPLIANCE WITH ETHICAL STANDARDS

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

CONFLICT OF INTEREST

The author declare no conflicts of interest.

REFERENCES

1. *Knorre D.G., Kudryashova N.V., Godovikova T.S.* (2009) Chemical and functional aspects of posttranslational modification of proteins. Acta Naturae, **1**(3), 29–51. DOI: 10.32607/20758251-2009-1-3-29-51
2. *Naryzhny S.N., Legina O.K.* (2019) Structural-functional diversity of p53 proteoforms. Biomeditsinskaya Khimiya, **65**(4), 263–276. DOI: 10.18097/PBMC20196504263
3. *Rybina A.V.* (2023) Proteoform identification in 2D electrophoresis maps by using isoelectric point prediction. Biomedical Chemistry: Research and Methods, **6**(1), e00191.DOI: 10.18097/BMCRM00191
4. *Bu X., Zhang N., Yang X., Liu Y., Du J., Liang J., Xu Q., Li J.* (2011) Proteomic analysis of cPKCβII-interacting proteins involved in HPC-induced neuroprotection against cerebral ischemia of mice. J. Neurochemistry, **117**(2), 346–356. DOI: 10.1111/j.1471-4159.2011.07209.x
5. *Ramiro L., Faura J., Simats A., García-Rodríguez P., Ma F., Martín L., Canals F., Rosell A., Montaner J.* (2023) Influence of sex, age and diabetes on brain transcriptome and proteome modifications following cerebral ischemia. BMC Neuroscience, **24**(1), 7. DOI: 10.1186/s12868-023-00775-7
6. ProteomeXchange, project PXD051750. DOI: 10.6019/PXD051750.
7. *Skvortsov V.S., Voronina A.I., Ivanova Y.O., Rybina A.V.* (2021) The prediction of the isoelectric point value of peptides and proteins with a wide range of chemical modifications. Biomedical Chemistry: Research and Methods, **4**(4), e00161. DOI: 10.18097/BMCRM00161
8. *Voronina A.I., Rybina A.V.* (2023) A program for predicting the retention time of peptides with post-translational modifications. Biomedical Chemistry: Research and Methods, **6**(3), e00196. DOI: 10.18097/BMCRM00196
9. *Ma B., Zhang K., Hendrie C., Liang C., Li M., Doherty-Kirby A., Lajoie G.* (2003) PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom., **17**(20), 2337–2342. DOI: 10.1002/rcm.1196
10. *Dixon M., Webb E.C.* (1979) Enzymes. New York: Academic Press. 1116 p.

Received: 10. 10. 2024.
Revised: 03. 12. 2024.
Accepted: 05. 12. 2024.

ИДЕНТИФИКАЦИЯ ПРОТЕОФОРМ МОЗГА МЫШЕЙ: СРАВНЕНИЕ ДАННЫХ 2D-ЭЛЕКТРОФОРЕЗА И НЕЗАВИСИМОГО ЭКСПЕРИМЕНТА С МАСС-СПЕКТРОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИЕЙ

А.В. Рыбина

Научно-исследовательский институт биомедицинской химии им. В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; эл. почта: aleona.rybina@ibmc.msk.ru

Ранее был предложен алгоритм предварительной идентификации белковых протеоформ, ассоциированных с посттрансляционными модификациями (ПТМ), на основе данных 2D-электрофореза (DOI: 10.18097/BMCRM00191). В данной работе предпринята попытка оценить его работоспособность, используя экспериментальные данные из двух работ разных авторов, полученных на мышах. Авторы первой работы выделили 8 групп пятен на 2D-электрофоретических картах, соответствующих 8 белкам, имеющим не менее двух неконкретизированных протеоформ. Авторы второй анализировали образцы мозга, используя метод LC-MS/MS. В данной работе идентификация пептидов с ПТМ была выполнена заново с использованием исходных данных из второй работы. Из 8 целевых белков 7 были идентифицированы в большинстве биологических проб. Для 4 из них было найдено 17 возможных модификаций. 5 вариантов протеоформ с идентифицированными ПТМ совпали с точками на картах 2D-электрофореза. Таким образом, использование предсказания величины pI для белков с гипотетическими ПТМ позволяет сформировать набор гипотез, какие конкретно протеоформы наблюдаются на картах 2D-электрофореза.

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

Ключевые слова: биоинформатика; 2D-электрофорез; масс-спектрометрия; ишемический инсульт

Финансирование. Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы) (№ 122030100170-5).

Поступила в редакцию: 10.10.2024; после доработки: 03.12.2024; принята к печати: 05.12.2024.