

©Рыбина

## 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.gybina@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.

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

**DOI:** 10.18097/PBMC20247006475

### ВВЕДЕНИЕ

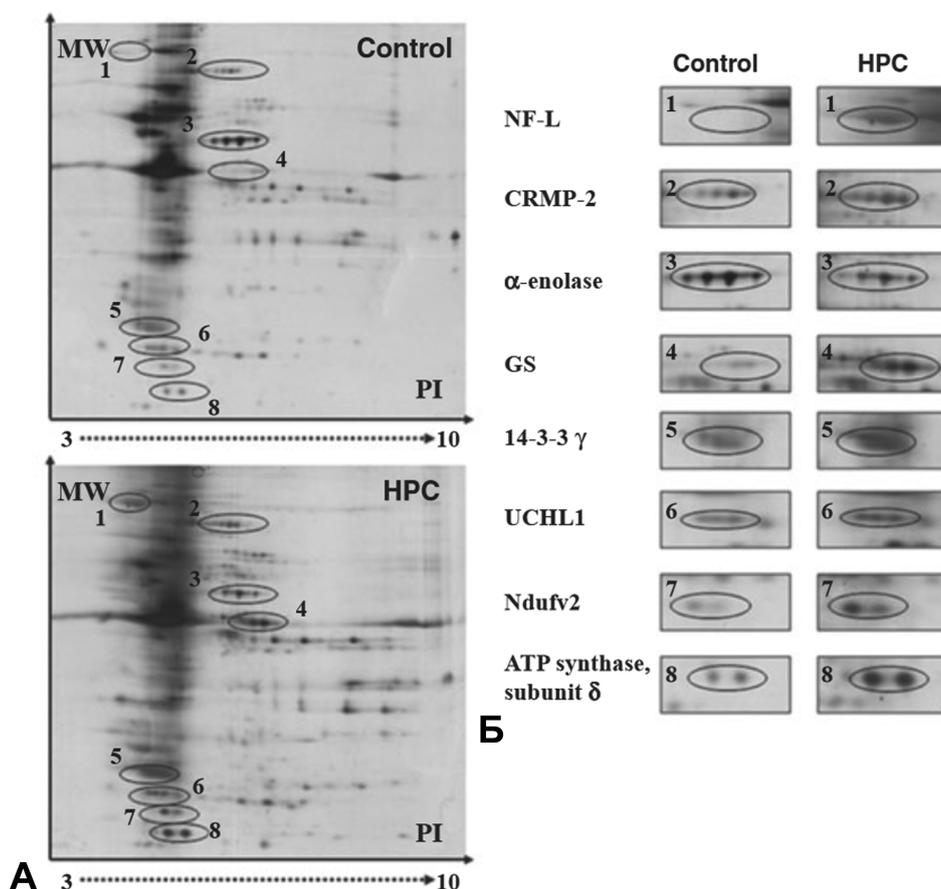
Посттрансляционные модификации (ПТМ) белков играют важную роль в регуляции многих биохимических процессов в клетке в норме и при патологии [1]. В экспериментах с использованием двумерного (2D) гель-электрофореза часто регистрируют белки, представленные несколькими протеоформами [2]. Если при этом электрофоретический сдвиг наблюдается в основном по значению изоэлектрической точки белка (pI), а не по массе белка, то с наибольшей вероятностью он связан с наличием ПТМ, а не наличием протеоформ, существующих в результате альтернативного сплайсинга или деградации последовательности белка. Анализ данных 2D гель-электрофореза далеко не всегда позволяет выявить ПТМ. Например, при детекции белков методом иммуноферментного анализа не удаётся различить модифицированные и немодифицированные формы белков [2]. Даже если провести для выделенных участков анализ методом тандемной масс-спектрометрии (MS/MS), то в большинстве случаев исследователю нет необходимости получить полное покрытие белка или искать модифицированные участки; как правило, достаточно идентифицировать 2–3 характеристических пептида. В связи с этим исследователи редко настраивают программы идентификации на поиск широкого спектра ПТМ, так как это существенно увеличивает время расчётов и может быть источником ошибок. В то же время с высокой долей вероятности масса первичных ионов (а иногда и вторичные спектры) для пептидов, содержащих модификации, при проведении MS/MS анализа измерена и записана. Ранее был предложен метод, дающий возможность генерации гипотез о конкретных ПТМ белков

на карте 2D гель-электрофореза на основании предсказания величины pI [3]. В данной работе сделана попытка показать, что эти предсказания действительно работают.

### МЕТОДИКА

Источником данных для данной работы были результаты двух экспериментальных работ разных групп исследователей, полученные на мышах. В одной животных подвергали гипоксическому прекодиционированию (autohypoxia-induced hypoxic preconditioning, HPC) [4]. После процедуры 2D-электрофореза белки мозга идентифицировали с использованием метода MALDI-TOF MS. В другой моделировали ишемию головного мозга с использованием окклюзии средней мозговой артерии (middle cerebral artery occlusion (MCAO)-induced cerebral ischemia) и препараты мозга анализировали методом LC-MS/MS [5]. Исходные данные работы [5] доступны в базе данных ProteomeXchange [6] (accession code PXD032141).

Анализ карт 2D-электрофореза был проведён по методике, описанной в работе [3], предсказание величины изоэлектрической точки выполнено программой pIPredict 3.0 [7] (вариант с учётом влияния соседних аминокислотных остатков). Были использованы изображения и аннотации двух карт цитозольной фракции коры головного мозга мышей контрольной группы и животных, подвергнутых HPC. Авторы экспериментальной работы [4] выделили 8 групп точек на картах 2D-электрофореза (рис. 1), соответствующих 8 белкам с не менее чем двумя неконкретизированными протеоформами. Эти данные были дополнены



**Рисунок 1.** Положение на картах 2D-электрофореза (А) белков с протеоформами, идентифицированных в [4]. (Б) увеличенное изображение фрагментов карты. Модифицировано из [4].

(табл. 1) полученной из UniProt информацией о всех подтверждённых экспериментально или идентифицированных на основе гомологов ПТМ, встречающихся в этих белках. Затем было сформировано пространство всех возможных пептидов, которые могут быть получены при обработке трипсином. Допускались пептиды с одним “недорезом” (пропущенным сайтом рестрикции при гидролизе белка). Для каждого из этих пептидов было предсказано время удержания (RT) с использованием программы RTP [8] и сформирован виртуальный спектр фрагментации (CID), которые были использованы для сравнения с данными из второй работы.

В настоящей работе идентификация пептидов по исходным данным [5] была проведена заново по следующему алгоритму:

1. Средствами программы PEAKS-X Pro [9] была выполнена идентификация пептидов (погрешность идентификации для первичных ионов 5 ppm и 0,1 Да для вторичных ионов, уровень False Discovery Rate (FDR) 1%). Условием идентификации белка было наличие трёх уникальных пептидов. В данной работе был использован только один набор данных из 4 (“young male mice”, 24 биологические пробы).

2. С использованием программы Progenesis LC-MS (“Nonlinear Dynamics”, Великобритания) проведено выравнивание данных всех 24 проб по RT.

Затем с использованием полученных на этапе 1 данных идентификации пептидов, не имеющих ПТМ, значения RT были выравнены с значениями RT, предсказываемыми программой RTP.

3. Для каждого из виртуальных пептидов был проведён поиск совпадений по всей совокупности вторичных спектров, полученных в работе [5], со следующими параметрами: погрешность идентификации для первичных ионов 2 ppm, отклонение RT от предсказанного в пределах окна, равного 3 мин, погрешность идентификации для фрагментов вторичных ионов 0,05 Да.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты поиска совпадений приведены в таблице 2. На первом этапе (данные анализа с использованием программы PEAKS-X Pro) было идентифицировано только 6 из 8 белков, причём ни одного пептида с ожидаемыми ПТМ зафиксировано не было. Можно предположить, что проблема с идентификацией состоит в том, что алгоритм поиска динамически настраивается на идентификацию пептида с наилучшими показателями в случае, если спектр смешанный, т.е. содержит более одного пептида. В работе [5] заявлено окно в 5 ppm. Косвенным свидетельством того, что вторичные спектры преимущественно

Таблица 1. Белки с протеоформами, идентифицированные в работе [4]

| № | Код на рисунке 1               | Идентификатор UniProt | Количество точек на 2D-карте (particulate control) | Количество точек на 2D-карте (HPC) | Название белка по UniProt   | Количество гипотетических сайтов с ПТМ (по данным UniProt) | Возможные ПТМ (по данным UniProt)  | Количество биологических образцов, в которых идентифицирован белок (*) |
|---|--------------------------------|-----------------------|--|------------------------------------|---|--|--|--|
| 1 | NF-L                           | P08551                | 0  | 2                                  | Neurofilament light polypeptide   | 10   | N-концевое ацетилирование, фосфорилирование  | 24   |
| 2 | CRMP-2                         | O08853/<br>P97821     | 5  | 5                                  | Dipeptidyl peptidase 1  | 0  | —  | 5  |
| 3 | $\alpha$ -enolase              | P17182                | 4  | 3                                  | Alpha-enolase   | 22   | N-концевое ацетилирование, ацетилирование лизина, фосфорилирование, сукцинирование | 24   |
| 4 | GS                             | P15105                | 2  | 2                                  | Glutamine synthetase  | 3  | N-концевое ацетилирование, фосфорилирование  | 24   |
| 5 | 14-3-3 $\gamma$                | P61982                | 3  | 2                                  | 14-3-3 protein gamma  | 7  | N-концевое ацетилирование, фосфорилирование  | 24   |
| 6 | UCHL1                          | Q9R0P9                | 3  | 3                                  | Ubiquitin carboxyl-terminal hydrolase isozyme L1                        | 1  | фосфорилирование   | 24   |
| 7 | Ndufv2                         | Q9D6J6                | 2  | 2                                  | NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial precursor | 2  | ацетилирование лизина, фосфорилирование  | 14   |
| 8 | ATP synthase, subunit $\delta$ | Q9DCX2                | 2  | 2                                  | ATP synthase subunit $\delta$ , mitochondrial                           | 11   | N-концевое ацетилирование, ацетилирование лизина, сукцинирование                   | 24   |

Примечание: \* – идентификация выполнена в данной работе.

смешанные, служит тот факт, что большое число вторичных спектров (даже при небольшой общей массе первичного иона) содержит нескольких тысяч пиков. Таким образом, данные, полученные на этом этапе, были использованы только для последующего выравнивания отдельных проб по RT, и было необходимо использовать процедуру идентификации, описанную для этапа 3.

Из 8 целевых белков во всех пробах было идентифицировано также 6 белков, ещё один в более чем половине проб и один в 5 пробах (табл. 1). Для 4 из них найдены 17 пептидов с искомыми ПТМ (табл. 2). При таком варианте поиска сложно сказать, какого числа фрагментов вторичного спектра достаточно для идентификации. В таблицу вошли варианты с как минимум 4 ( $y$ - или  $b$ -ионы). Также некоторые из пептидов имеют совпадающий фрагмент, что связано с наличием “недорезов” или подтверждённым различным положением ПТМ. Как минимум для 2 белков можно говорить о том,

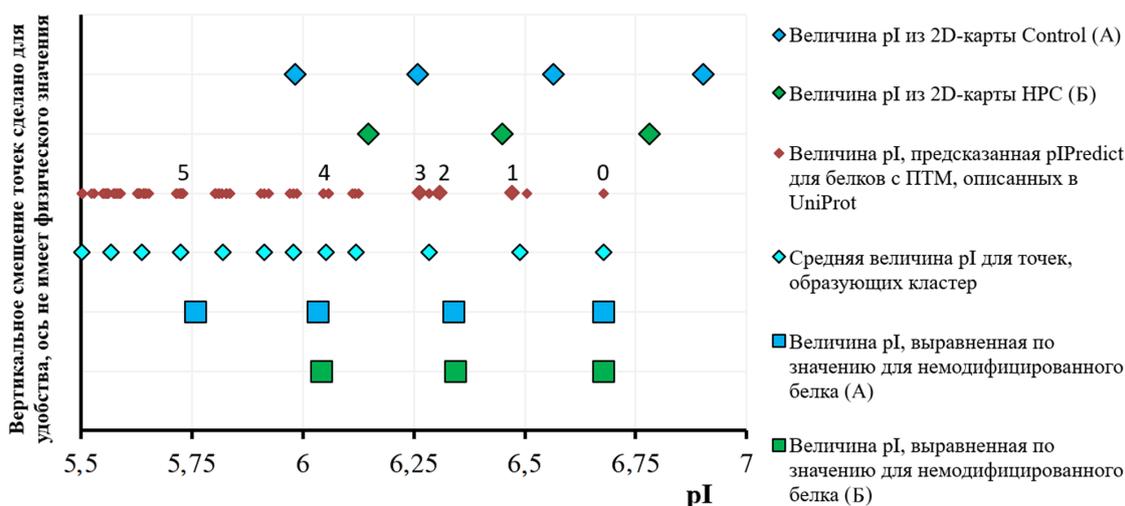
что найденные пептиды с ПТМ могут соответствовать протеоформам, представленным на 2D-картах (табл. 2). Ещё в одном случае (для белка глутаминсинтетаза, UniProt ID P15105) возможная протеоформа соответствует той же области затемнения, в которой выделены точки на 2D-карте, и так как она соответствует N-концевому ацетилированию, то можно предположить неправильное определение “нулевой” точки (положение на карте электрофореза немодифицированного белка), которая не выражена ярко и должна быть смещена в основную область. Для белка альфа-енолаза (P17182) было найдено 7 вариантов ПТМ в 11 пептидах (табл. 2 и рис. 2), для части из которых можно найти соответствие протеоформам на 2D-картах. При этом следует иметь в виду, что их комбинация может дать протеоформы, также представленные на 2D-картах. Белок альфа-енолаза, судя по всему, подвержен множественному ацетилированию, так как есть точки, которым можно поставить в соответствие

**ИДЕНТИФИКАЦИЯ ПРОТЕОФОРМ ПО ДАННЫМ 2D-ЭЛЕКТРОФОРЕЗА**

Таблица 2. Пептиды с ПТМ, идентифицированные в работе. Сравнение с экспериментальным положением точки на геле проводили с расчётными значениями для полного белка с обнаруженной модификацией

| №  | Белок  | Пептид                                     | Позиция ПТМ в последовательности | Максимум совпадающих фрагментов MS/MS | Количество биологических образцов, в которых идентифицирован пептид | Соответствие варианта протеоформы с этими ПТМ точкам на 2D карте   |
|----|--------|--|----------------------------------|---------------------------------------|---|--|
| 1  | P08551 | S(+80)YSSSSGLMPSLENLDSQVAAISNDLK           | 56                               | 12                                    | 11  | Не представлен.  |
| 2  | P17182 | AAVPSGASTGIY(+80)EALELR                    | 44                               | 16                                    | 5   | Представлен (наиболее вероятно).   |
|    |        | GLFRAAVPSGASTGIY(+80)EALELR                |                                  | 4                                     | 2   |  |
| 3  | P17182 | EALELLK(+100)TAIAK                         | 228                              | 10                                    | 8   | Представлен (менее вероятно, чем фосфорилирование). Полностью совпадает по pI с двойным ацетилированием. |
| 4  | P17182 | KVNVVVEQEK(+100)                           | 89                               | 5                                     | 2   | Представлен (менее вероятно, чем фосфорилирование). Полностью совпадает по pI с двойным ацетилированием. |
| 5  | P17182 | FMGK(+42)GVSVQAVENHNK(+42)                 | 60; 71                           | 6                                     | 3   | Представлен (менее вероятно, чем фосфорилирование).  |
| 6  | P17182 | EALELLK(+42)TAIAK                          | 228                              | 6                                     | 8   | Не представлен (как единственная модификация, но в комбинации с другими ПТМ возможно).                   |
| 7  | P17182 | EALELLKTAIAK(+42)                          | 233                              | 7                                     | 8   | Не представлен (как единственная модификация, но в комбинации с другими ПТМ возможно).                   |
|    |        | TAIAK(+42)AGYTDQVVGMDVAASEFYR              |                                  | 6                                     | 1   |  |
| 8  | P17182 | VNVVVEQEK(+42)IDK(+42)                     | 89; 92                           | 9                                     | 9   | Представлен (менее вероятно, чем фосфорилирование).  |
| 9  | P17182 | YITPDQLADLYK(+42)                          | 281                              | 14                                    | 8   | Не представлен (как единственная модификация, но в комбинации с другими ПТМ возможно).                   |
|    |        | YITPDQLADLYK(+42)SFVQNYPVVSIEDPFDQDDWGAWQK |                                  | 5                                     | 1   |  |
| 10 | Q9DCX2 | ANVAK(+42)PGLVDDFEK(+42)                   | 63; 72                           | 14                                    | 9   | Не представлен в Control, возможно есть в НРС (равная вероятность с одиночным ацетилированием).          |
|    |        | ANVAK(+42)PGLVDDFEK(+42)K                  |                                  | 9                                     | 7   |  |
| 11 | Q9DCX2 | MRNIPFDQMTIDDLNEIFPETK(+42)                | 144                              | 8                                     | 7   | Представлен.   |
|    |        | NIPFDQMTIDDLNEIFPETK(+42)                  |                                  | 5                                     | 1   |  |
| 12 | Q9DCX2 | IPVPEDK(+42)YALVDQEEK(+100)                | 85; 95                           | 17                                    | 8   | Не представлен.  |
| 13 | Q9DCX2 | IPVPEDK(+42)YALVDQEEK(+42)                 | 85; 95                           | 4                                     | 5   | Не представлен в Control, возможно есть в НРС (равная вероятность с одиночным ацетилированием).          |
| 14 | Q9DCX2 | K(+100)YRYWRHQRIENL                        | 149                              | 13                                    | 3   | Не представлен в Control, возможно есть в НРС (равная вероятность с одиночным ацетилированием).          |
| 15 | Q9DCX2 | YNALK(+100)IPVPEDK(+100)                   | 78; 85                           | 5                                     | 2   | Не представлен.  |
| 16 | Q9DCX2 | IPVPEDK(+100)YALVDQEEK(+42)                | 78; 85                           | 15                                    | 9   | Не представлен.  |
| 17 | P15105 | A(+42)TSASSHLNKGIK                         | 1                                | 4                                     | 4   | Не представлен (но находится в пределах неоформленного затемнения).                                      |

Примечание. ПТМ: (+80) – фосфорилирование, (+42) – ацетилирование, (+100) – суццинирование.



**Рисунок 2.** Анализ карт электрофореза контрольной (А) и HPC (Б) для белка P17182. Отмеченные цифрами зоны соответствуют вариантам протеоформ: 0 – без ПТМ, 1 – одиночное ацетилирование, 2 – одиночное фосфорилирование, 3 – одиночное сукцинирование, либо двойное ацетилирование, 4 – тройное ацетилирование, 5 – тройное ацетилирование и одиночное фосфорилирование.

протеоформы с тройным ацетилированием и даже с тройным ацетилированием и фосфорилированием, в то время как в зоне одиночного ацетилирования точки на карте нет.

В части идентифицированных пептидов есть ацетилирование С-концевого лизина, что, возможно, противоречит общеизвестным представлениям о специфичности трипсина. Однако известно, что трипсин может (хотя и со значительно меньшей эффективностью) осуществлять гидролиз не только после положительно заряженных остатков [10]. Для части таких пептидов найдены также и формы, где данный модифицированный аминокислотный остаток находится во внутренней части аминокислотной последовательности, что может служить дополнительным доказательством существования данной ПТМ.

## ЗАКЛЮЧЕНИЕ

Таким образом, использование предсказания величины pI для белков с гипотетическими ПТМ позволяет сформировать набор гипотез, какие конкретно протеоформы наблюдаются на картах 2D-электрофореза. В данной работе для 3 из 8 белков можно предположить 5 вариантов протеоформ с идентифицированными ПТМ, которые совпали с точками на картах 2D-электрофореза. Кроме того, при использовании комбинированных методик (2D электрофорез и LC-MS/MS) можно указать заданную область на геле, которую нужно анализировать для идентификации конкретных протеоформ.

## ФИНАНСИРОВАНИЕ

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

## СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Данная работа не содержит каких-либо исследований с использованием людей и животных в качестве объектов исследования.

## КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

## ЛИТЕРАТУРА

1. Кнорре Д.Г., Кудряшова Н.В., Годовикова Т.С. (2009) Химические и функциональные аспекты посттрансляционной модификации белков. *Acta Naturae*, **1**(3), 32–56. [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. Нарыжный С.Н., Легина О.К. (2019) Структурно-функциональное многообразие протеоформ белка p53. *Биомедицинская химия*, **65**(4), 263–276. [Naryzhny S.N., Legina O.K. (2019) Structural-functional diversity of p53 proteoforms. *Biomeditsinskaya Khimiya*, **65**(4), 263–276.] DOI: 10.18097/PBMC20196504263
3. Рыбина А.В. (2023) Идентификация протеоформ в результатах 2D электрофореза по предсказанным значениям величины изоэлектрической точки. *Biomedical Chemistry: Research and Methods*, **6**(1), e00191. [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 cPKC $\beta$ II-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. *Скворцов В.С., Воронина А.И., Иванова Я.О., Рыбина А.В.* (2021) Предсказание значения изоэлектрической точки пептидов и белков с широким спектром химических модификаций. *Biomedical Chemistry: Research and Methods*, **4**(4), e00161. [*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. *Воронина А.И., Рыбина А.В.* (2023) Программа предсказания времени удержания пептидов с учётом посттрансляционных модификаций. *Biomedical Chemistry: Research and Methods*, **6**(3), e00196. [*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.

Поступила в редакцию: 10. 10. 2024.  
После доработки: 03. 12. 2024.  
Принята к печати: 05. 12. 2024.

**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.

*The whole English version is available at <http://pbmc.ibmc.msk.ru>.*

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

**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).

Received: 10.10.2024, revised: 03.12.2024, accepted: 05.12.2024.