

УДК 51-76:577.29

Олевська Ю.Б.^а, Олевський В.І.^б, Олевський О.В.^в**РОЗВИТОК МЕТОДІВ КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ В ПРОТЕОМИЦІ**^а ДВНЗ Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна^б ДВНЗ «Український державний хіміко-технологічний університет», м. Дніпро, Україна^в ДВНЗ «Дніпровський національний університет імені Олеся Гончара», м. Дніпро, Україна

Актуальною потребою сучасних досліджень в області медико-біологічних наук є вирішення проблеми підвищення точності результатів електрофореграфії в їх загальновживаному вигляді та обґрунтування заснованих на таких експериментах висновків, зокрема, підвищення точності ідентифікації білків. Проблеми розрахунків у даному напрямі вирішуються як універсальними, так і спеціалізованими комп’ютерними системами, різноманітність яких обумовлена, по-перше, конкретними потребами природничих наук, і, по-друге, математичним апаратом, що застосовується для досить об’ємних розрахунків. Зазвичай в такому разі використовуються методи математичної статистики, але в даному випадку застосування стохастичного аналізу даних не могло бути досить коректним через відсутність репрезентативної вибірки або її неоднорідності. У зв’язку з цими потребами було створено комп’ютерний пакет FANSPREL – «Fuzzy ANalysis System for Protein Electrophoresis» («Система нечіткого аналізу результатів електрофорезу білків») для обробки електрофорограм та ідентифікації білків при застосуванні методів нечіткої математики на основі алгоритмів, які було розроблено авторами в попередніх дослідженнях. Пропонується підхід, заснований на визначені нечіткої шкали за результатами обробки нечітких даних відомих білків, розпізнаванні білків, одержаних за результатами дослідження, та порівнянні між собою білків, отриманих за результатами подальших експериментів. Математична частина інформаційної системи являє собою побудову наближення функції належності на основі отриманих даних кривою з чотирма параметрами та побудову нечіткої моделі шкали, визначення нечітких мас білків та їх порівняння шляхом кон’юнкції нечітких значень, та наступну дефазифікацію результатів. Система дає можливість більш точної ідентифікації білків, зокрема розмежування тих, маси яких при інших формах обробки при певній точності ідентифікувалися як однакові або був отриманий невизначений результат. Програма має зручний інтерфейс, передбачає стандартний вигляд вхідних і вихідних даних і апробована на вирішенні задач експериментів, що відбувалися для безпосередніх потреб сільського господарства, зокрема для покращення різноманітних видів ґрунту, який використовується для вирощування кукурудзи.

Ключові слова: нечітке моделювання, електрофореграми, комп’ютерна система.

DOI: 10.32434/2521-6406-2020-8-2-49-58

Постановка проблеми в загальному вигляді та її зв’язок з важливими практичними завданнями

В сучасному світі зростаючі потреби в кількісному і якісному аспектах обробки експериментальних даних в області молекулярних досліджень в області біології і медицини стимулюють здійснення переходу на новий рівень от-

римання результатів експериментів та їх аналізу математичними методами. Останні досягнення і розробки дозволяють бачити і впливати на причину явища на молекулярному рівні в діагностичному та лікувальному аспектах медицини, популяційної генетики, питаннях підвищення врожайності та інших проблемах, пов’язаних з найважливішими аспектами життєдіяльності

людини. Описаний комплекс завдань обумовлює застосування методу математичного моделювання для використання в ширину – опису біологічних макропроцесів і в глибину – дослідженю будови і функціонування клітин і органів, що складають тваринний або рослинний організм. Обидва ці підходи можуть слугувати і основою, і наслідком один для іншого. В силу специфіки будови клітин згаданих організмів при здійсненні детальних досліджень, пов’язаних з отриманням експериментальних даних, аналізом їх коректності, математичною обробкою та оцінкою отриманих результатів, використовуються різні складові, що несуть певний набір властивостей. При цьому, як показує досвід досліджень на молекулярному рівні, найбільш інформативними в сенсі точності результатів, сфери охоплених питань і, як наслідок, найбільш розвиненої експериментальної бази, є дослідження білків. Моделювання процесів і експериментів на макро- і мікрорівні в біології та медицині виконує дві функції – дослідницьку і прогнозуючу. Різноманітність завдань обумовлює різноманітність відповідних математичних моделей і застосування розрахункових методів, які залежать від цілей дослідження. В умовах сьогодення інтенсивно розвивається застосування методів хемометрики [1–4] в біології та медицині [3–5].

Аналіз публікацій, виділення невирішених частин загальної проблеми

Сучасні дослідження білків базуються на біологічному та хімічному фундаменті як класичному, так і тому, що знаходиться на піку розвитку в силу актуальності поставлених питань [3,6]. Білки – це лінійні полімерні молекули, що складаються з послідовності амінокислотних залишків, сполучених пептидними зв’язками. Термін «білковий» був вперше застосований французьким фізіологом Ф. Кене по відношенню до всіх рідин тваринного походження. Хімічний склад білків був вперше описаний Йенсом Якобом Берцеліусом – шведським хіміком, мінералогом, президентом Шведської академії наук. Їм же був запропонований і термін «протеїн». Перша модель хімічної будови білків, яка була запропонована Геррітом Яном Мульдером в 1836 році, з усіма своїми уточненнями і розвитком лежить в основі сучасних досліджень. При розгляді питань, так чи інакше пов’язаних з функціонуванням білкових структур, враховується, що білкам властива здатність до різноманітних внутрішніх і міжмолекулярних взаємодій. У зв’язку з актуальністю дослідження

білків в біології, медицині та випливаючих з них напрямами, а також завдяки розвитку нових можливостей у виконанні експериментів, постійному удосконаленню математичного апарату і комп’ютерних технологій, провідним вектором досліджень в даному напрямі з кінця ХХ століття стає протеоміка [2,6].

Протеоміка – це кількісний і якісний аналіз білкових структур живих організмів і властивостей білкових молекул.

При дослідженнях білків найбільш часто, в силу своєї інформативності, в експериментах визначаються рухливість білка, маса, склад молекули (зокрема, молекули ДНК). Отже, при моделюванні найбільш важливим є визначення маси білка, характеру його руху, складу генома.

Термін «протеом» вперше був запропонований в 1994 році Марком Вілкінсом. Це поняття виникло в процесі розвитку методу поділу білків двовимірним електрофорезом в поліакриlamідному гелі, який виявився найбільш вдалим за своїми параметрами для дослідження білків. І саме метод електрофорезу є найбільш затребуваним, розвиваючись з моменту відкриття явища в 1807 році.

Існують різні методи знаходження маси білка, які обумовлені відповідними властивостями молекул [3,4]. Оптичні методи засновані на тому, що маса молекули білка строго визначає його вид, а також поверхня молекули має заряд, змінюваний за допомогою зміни pH розчину. Для дослідження поведінки білкових молекул в розчинах найбільш ефективно використовуються методи статичного і динамічного розсіювання світла, а також атомна силова мікроскопія. Перевага оптичних методів визначається, зокрема, їх неінвазивністю і відсутністю руйнівної дії.

Метод мас-спектрометрії білків передбачає іонізацію білків і подальше дослідження їх в електричному або магнітному полі.

Метод ультрацентрифугування розчинів полімерів застосовується, зокрема для визначення молекулярних мас білків на основі аналізу осідання білків під дією високошвидкісної ультрацентрифуги, яка була сконструйована Т. Свебергом в 1926 році.

Метод гель-хроматографії заснований на аналізі співвідношення розмірів молекули і розмірів пор використовуваних гелів, через які відбувається фільтрація.

Одним з центральних методів визначення молекулярних мас білків і, отже, їх розпізнавання, є електрофоретичний метод [7,8]. Електро-

кінетичні явища були відкриті Ф.Ф. Рейссом в 1807 році при дослідженні електролізу води [9]. У двох поставлених експериментах були відкриті два явища: електроосмос – переміщення рідини в пористих тілах під дією електричного поля і електрофорез – переміщення частинок дисперсної фази в електричному полі. Вперше застосований А. Тізеліусом в 1930 роки, метод, заснований на направленому русі заряджених мікрочастинок в рідкому середовищі під дією електричного поля, в даний час займає фундаментальне місце по частоті використання в відповідних теоретичних і прикладних розробках.

Метод полягає в пересуванні колоїдних частинок до полюсів, або до позитивного – анода, або до негативного – катода під дією електричного поля. Можливість застосування даного методу і отримання електрофорограм для наступних розшифрувань і аналізу детально розглянуті в [7]. Можливість дослідження колоїдів протоплазми, і, зокрема, білків, що є одним із першочергових завдань біології і медицини, обумовлена їх специфічними хімічними властивостями. Зокрема, дослідження базуються на електричних зарядах білкових колоїдних частинок. Важлива сфера застосування електрофорезу – поділ органічних і високомолекулярних компонентів розчину, тому метод має надзвичайно широке застосування в медицині та біології для розділення і аналізу білків. Молекули білків є з'єднанням багатьох амінокислот, які, в свою чергу, є амфотерними електролітами, тобто слабкими електролітами, які здатні виявляти властивості як слабких кислот, так і слабких основ в залежності від природи речовини, з яким вони вступають в реакцію. Водні розчини білків зазвичай мають негативний заряд і частинки білків пересуваються до анода. У крові і органах організму, наприклад, середа злегка лужна. Тому багато білків організму і поверхні клітин мають негативний заряд. Цей факт і покладено в основу електрофоретичного аналізу білкового складу біологічних рідин.

Електрофоретична трубка електрофоретичного апарату, створеного Тізеліусом [10], являла собою U-образну посудину, канал якої в поперечному перерізі має форму прямокутника. Дослідження велося при температурі близько 0°C, щоб уникнути розігрівання, обумовленого різницею потенціалів, і для усунення конвекційних струмів. Білкові фракції, що утворилися в силу електрофоретичної рухливості, мали різні коефіцієнти заломлення. Промінь, проходячи

через зони білкових фракцій, давав різне відхилення. За допомогою особливого пристрою оптичної реєстрації цей промінь автоматично викреслював на фотопластинці криві відхилення. Кожному індивідуальному білку відповідав характерний для нього пік. За результатами аналізу можна було не тільки судити про кількість окремих білкових фракцій, а й визначати концентрацію того чи іншого білка в системі. В апараті Тізеліуса також було можливо перевірити однорідність того чи іншого білка, і на підставі цього робити висновок про відсутність інших білкових домішок.

Електрофорез [7] виконується в однорідному електричному полі. Електричний струм пропускається через буферний розчин, налітий в канал з ізоляючого матеріалу або яким просочено середу-носій. Концентрація в буферному розчині вільних іонів та їх електрофоретична рухливість задають опір R буферного розчину. Під електрофоретичною рухливістю мається на увазі швидкість руху зарядженої молекули в одиничному полі:

$$u = \frac{V \text{ (см/г)}}{E \text{ (В/см)}},$$

де V – швидкість міграції білкової молекули, E – напруженість електричного поля.

З моменту свого виникнення і до теперішнього часу метод, зберігаючи концептуальну платформу, змінювався і вдосконалювався відповідно до нових теоретичних розробок і модифікації технічної бази, що робить його менш громіздким і дорогим, і сприяє підвищенню точності розшифрування результатів досліджень. У сучасній інтерпретації електрофоретичні методи включають в себе зональний (капілярний) електрофорез, ізоелектричного фокусування, ізотахофорез, іммуноелектрофорез [7,11]. При зональному електрофорезі досліджуваний матеріал розміщується в буферному розчині і піддається поділу до стабілізації електрофоретичних зон. За типом носія зміни, що відбулися з моменту першого застосування методу і до сьогоднішнього часу, включають в себе електрофорез з рухомою границею (відсутність носія), на фільтрувальному або хроматографічному папері, на плівках з ацетату целюлози, в крохмальній гелі, і два, які найбільш відповідають вимогам сучасності, – електрофорез в агаровом і агарозном гелях, електрофорез в поліакриламідному гелі (ПААГ).

Незважаючи на те, що електрофорез є од-

ним з класичних методів сучасної лабораторії молекулярної біології, який зберігає ідейну ідентичність, постійно відбувається процес втілення нових розробок, заснованих на розвитку математичного апарату і лабораторного оснащення [3,4,7], багато в чому завдяки потребам біомедичної спільноти. Розробляються нові механізми поділу і нові види матриць просіювання. Гель-електрофорез в імпульсному полі (PFGE) – метод генотипування, який використовується для поділу великих молекул ДНК (цільної геномної ДНК) після їх обробки рестрикційними ферментами і нанесення на гелевий матрикс під дією електричного поля, яке періодично змінює напрямок. PFGE – це варіант електрофорезу в агарозному гелі, який дозволяє аналізувати фрагменти бактеріальної ДНК на порядок більше, ніж при звичайному аналізі рестрикційних ферментів. Він забезпечує чітко розрізнювані і добре розділені фрагменти ДНК. Двовимірний електрофорез [7,12] – процес поділу складних зв'язків білків, в якому поєднуються електрофорез білків та ізоелектричне фокусування. Він передбачає поділ білків в одному напрямку, а потім в напрямку, йому перпендикулярному, що дозволяє збільшувати роздільну здатність при поділі суміші великої кількості білків.

Властивості білків як хімічних сполук лежать в основі кількісних і якісних обробок отриманих експериментальних і статистичних даних. Можливості аналітичної обробки експериментальних даних розширилися з розвитком сучасних комп’ютерних технологій. Крім безпосереднього використання універсальних і спеціалізованих пакетів комп’ютерних програм з’явилася можливість використання постійно оновлюваних загальносвітових банків даних характеристик різних структур, зокрема білків. Один з найбільших і найвідоміших – Protein Data Bank (PDB), що містить інформацію про білки і нуклеїнові структури. Тривимірні дані отримані експериментальним шляхом за допомогою апаратури високої точності і можуть бути використані для математичної обробки як універсальними комп’ютерними програмами, так і в процесі проміжних розрахунків. Також завдяки еталонним даним є можливість тестування правильності роботи розроблюваних програмних продуктів.

Специфіка даних, оброблюваних при розгляді задач протеоміки, передбачає потужний математичний апарат, підтримуваний сучасними засобами обчислювальної техніки. Обчислювальний потенціал і графічні можливості необ-

хідного комп’ютера визначаються на підставі обсягу інформації (великого чи, навпаки, надзвичайно малого), використання стандартних структур із зовнішніх баз даних, складності розрахунків і необхідної точності. Взаємозв’язок хмарних і локальних обчислень оптимально відповідає інтересам поставлених завдань.

У процесі досліджень спочатку відбувається обробка даних, отриманих в результаті експериментів, а потім обробка та інтерпретація результатів розрахунків і отримання нових даних, остаточних або проміжних. Для розрахунків використовуються універсальні та спеціалізовані математичні пакети. До універсальних відносяться такі загальновідомі програмні продукти, як MATLAB, MathCad, GNU Octave, Conda, Maple, SPSS. Спеціалізовані комп’ютерні системи розроблені під конкретні цілі розрахунків та подання даних. До них відносяться GROMACS – універсальний пакет для моделювання процесів в молекулярній динаміці за допомогою моделювання ньютоновських рівнянь руху для систем з сотнями і мільйонами частинок, ESPResSo – універсальний програмний пакет для виконання та аналізу багаточастинкових симуляцій крупномолекулярних атомистических моделей молекулярної динаміки, що використовуються в дослідженнях і моделюванні м’яких структур у фізиці, хімії і молекулярній біології. Його можна використовувати для моделювання таких систем, як полімери, рідкі кристали, колоїди, поліелектроліти, феррорідини і біологічні системи, наприклад, ДНК і ліпідні мембрани. PyMol – система візуалізації молекул, Unscrambler – інструмент для моделювання, прогнозування та оптимізації багавимірних даних, Umetrics – пакет комплексних рішень для аналізу даних.

Формулювання цілей статті та постановка завдання

Завданням представленої роботи є розробка комп’ютерного програмного пакета для обробки та аналізу електрофореграм білків на основі нечіткого аналізу. Математичний апарат системи базується на раніше розробленому авторами методі нечіткого визначення маси білків при гель-електрофорезі [13].

Виклад основного матеріалу дослідження

У зв’язку з сучасними цілями підвищення точності розшифровки електрофореграм білків було розроблено комп’ютерний пакет FANSPREL [14] – «Fuzzy ANalysis System for Protein Electrophoresis» («Система нечіткого аналізу результатів електрофорезу білків»). Базуючись на методах нечіткого аналізу, програмний

продукт являє собою науково-програмний комплекс, призначення якого має дві складові – практичну (стосовно застосування даних біологічних експериментів на практиці з відповідною економічною доцільністю) і теоретичну (стосовно застосування математичного методу нечіткого аналізу даних в біологічній та інших науках щодо опису та висновків експериментів з об'єктивно малою кількістю дослідів).

Необхідність створення та практичного застосування інформаційної системи виникла при спробах коректно описати експериментальні дані, спрогнозувати дані наступних експериментів та, в якості висновків, дати відповідь на питання доцільності подальших досліджень і впровадження їх результатів в практичну галузь аграрної промисловості. Біологічні дослідження призначенні для покращення різноманітних видів ґрунту, який використовується для вирощування кукурудзи. Відомо, що характеристики ґрунту можуть змінюватись після його переробки дощовими черв'яками роду *Eisenia*. Була здійснена низка експериментів по опроміненню лазером представників черв'яків, отриманню їх потомства і переробки ґрунту цим потомством. Вибірковий якісний аналіз показав зростання врожайності кукурудзи на перероблених ділянках, але для повного обґрунтuvання дослідження і подальшого впровадження його результатів у відповідну галузь сільського господарства була потрібна аналітична обробка матеріалу і побудова математичної моделі з урахуванням усіх факторів. Зазвичай в такому разі використовуються методи математичної статистики, але у даному випадку застосування стохастичного аналізу даних не могло бути досить коректним через відсутність репрезентативної вибірки або її неоднорідності.

З огляду на ці умови авторами було запропоновано здійснити обчислення, аналіз і прогнозування даних методами нечіткої математики. Програмний комплекс містить в собі декілька етапів обробки даних. По-перше, для коректної ідентифікації білків будеться нечітка шкала за результатами обробки нечітких даних відомих білків. Далі за допомогою цієї шкали надається можливість розпізнавання білків, отриманих за результатами дослідження. Окремо вирішується питання порівняння між собою білків, отриманих за результатами подальших експериментів. Результати роботи програми надаються в зручному для користування вигляді, в аналітичній, табличній і графічній формах.

Теоретична цінність комп'ютерної програми актуальна для таких галузей науки, як біо-

гія та математика. З одного боку, використання програмного комплексу надає можливість перевірки і, в разі позитивної відповіді, обґрунтування біологічних гіпотез щодо генотипу нащадків опромінених черв'яків. З іншого боку, розробка і застосування методу в математичному сенсі дає можливість розповсюдити його на експерименти подібного роду, з малою кількістю і достатньою різноманітністю даних, на інші галузі природничих та технічних наук.

Програму розроблено у відкритому середовищі GNU Octave. Вона має зручний інтерфейс, який відповідає всім потребам користувача та ергономічним вимогам (рис. 1,а,б,в).

Вихідними даними є: експериментальні параметри (щільність гелю, температура, напруга електричного поля); результати електрофорезу у вигляді графічного зображення з його фізичними розмірами; параметри дискретизації електрофореграм (лінійні розміри, кількість точок дискретизації у вертикальному та горизонтальному напрямках). Використовуються дві категорії електрофореграм – відомих білків (маркерів) для побудови шкали та білків, які отримані у конкретному експерименті, для їх ідентифікації та порівняння. Особливістю введення даних є пряме використання стандартних фотографій електрофореграм відповідного біологічного експерименту (рис. 2), подальше їх представлення в графічному вигляді (рис. 3) та автоматизоване отримання даних для розрахунків.

Вихідними даними є: звіти з параметрами, які отримані при побудові математичної моделі; звіти щодо шкали та ідентифікації білків в табличному і графічному вигляді. Висновки надаються в графічному (рис. 4,а,б) та табличному (рис. 5) вигляді для зручності аналізу і подальшого користування.

Створена інформаційна система розрахована на спеціалістів з мета-аналізу з точки зору застосування математичного апарату і біологів з метою коректного обчислення результатів експериментів. Програмна розробка надає можливість математичної обробки даних біологічного експерименту у разі відсутності репрезентативної вибірки чи її неоднорідності, що є необхідним для статистичного аналізу. Пропонується підхід, заснований на визначені нечіткої шкали за результатами обробки нечітких даних відомих білків, розпізнаванні білків, одержаних за результатами дослідження, та порівнянні між собою білків, одержаних за результатами подальших експериментів. Математична частина інформаційної системи являє собою побудову набли-

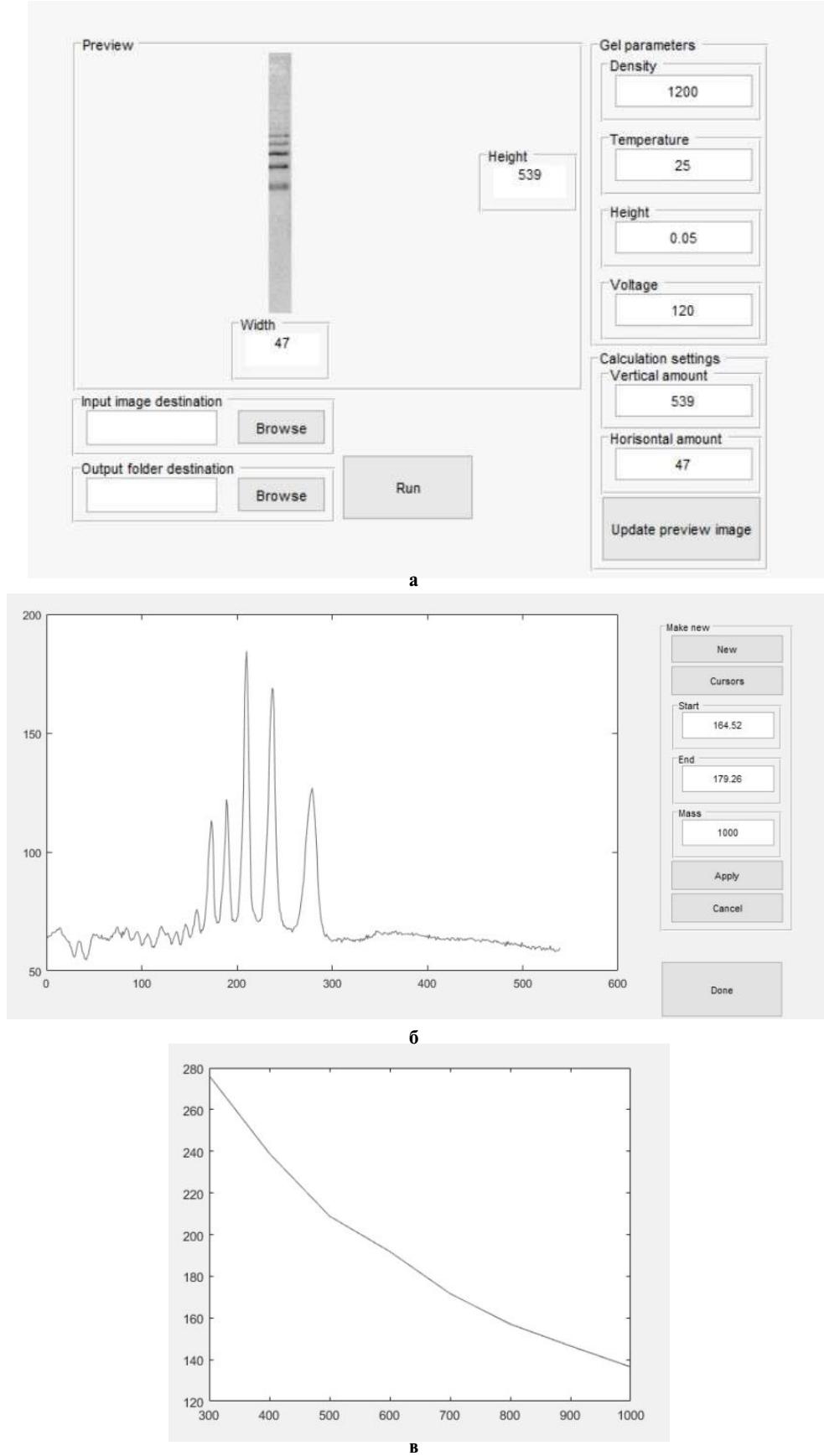


Рис. 1. Інтерфейс програми

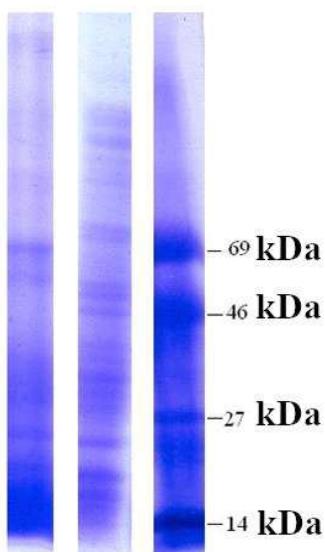


Рис. 2. Електрофореграми біологічного експерименту

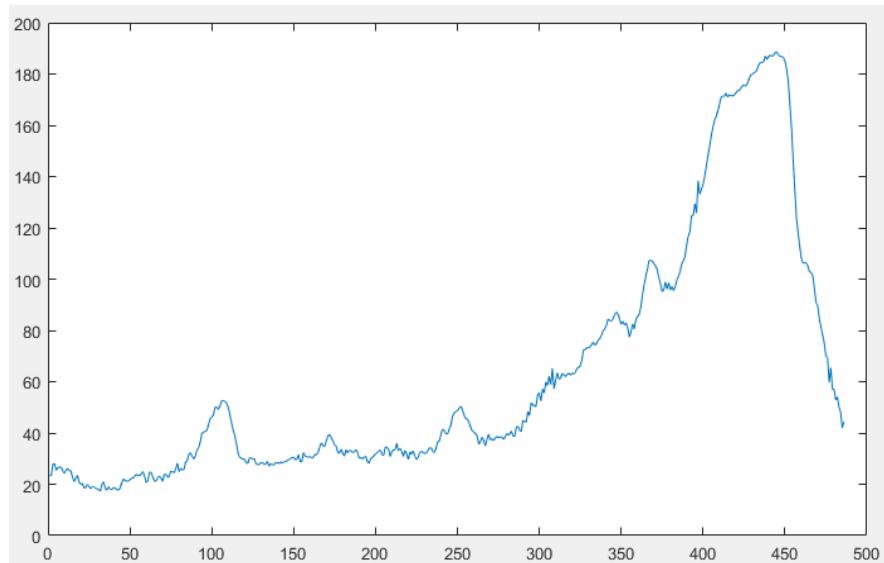
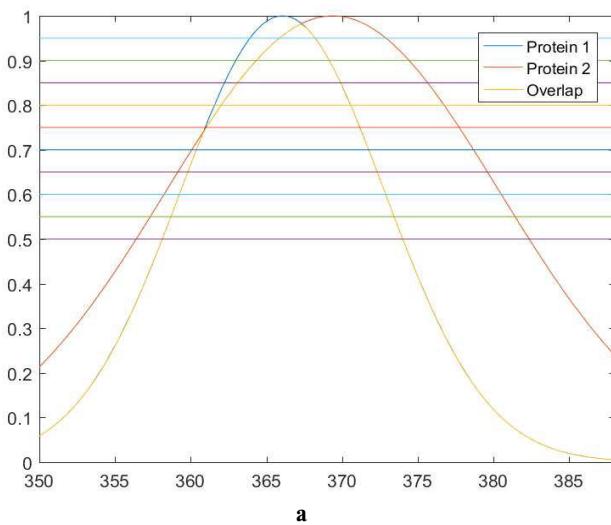
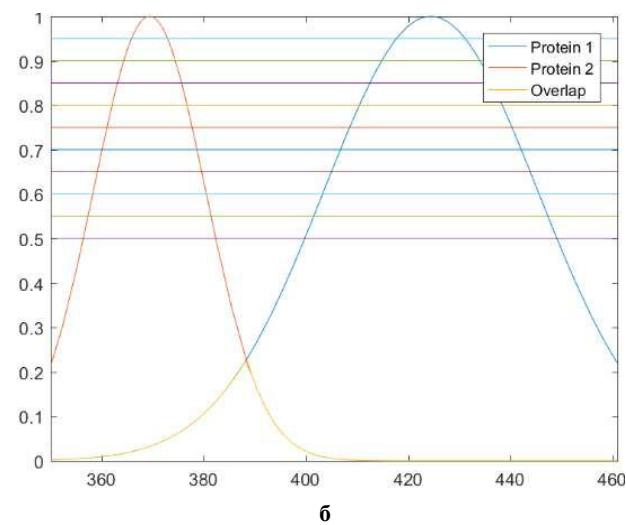


Рис. 3. Графічне представлення



а

Рис. 4. Графічне представлення результатів



б

Alpha-level	Min	Max
0.50	409.990	449.394
0.55	411.667	447.717
0.60	413.343	446.040
0.65	415.020	443.525
0.70	416.697	441.848
0.75	418.374	440.172
0.80	420.051	438.495
0.85	422.566	436.818
0.90	424.242	434.303
0.95	427.596	430.949

Рис. 5. Табличне представлення результатів

ження функції належності на основі одержаних даних кривою з чотирма параметрами і побудову нечіткої моделі шкали, визначення нечітких мас білків та їх порівняння шляхом кон'юнкції нечітких значень, а також дефазифікацію результатів.

Функціональні особливості створюваної

інформаційної системи наступні:

- обробка введених даних;
- побудова математичної моделі;
- порівняння білків щодо їх ідентичності;
- формування звітів.

Структурна схема інформаційної системи представлена на рис. 6.

За функціональним призначенням можна виділити наступні головні модулі, які пов'язані між собою та містять окремі блоки.

У блоках побудови шкали проводиться створення моделі шкали для подальшого нечіткого аналізу білків шляхом знаходження залежності параметрів носія від його маси.

Модуль створення файлів білків дозволяє зберігати інформацію про досліджуваний носій

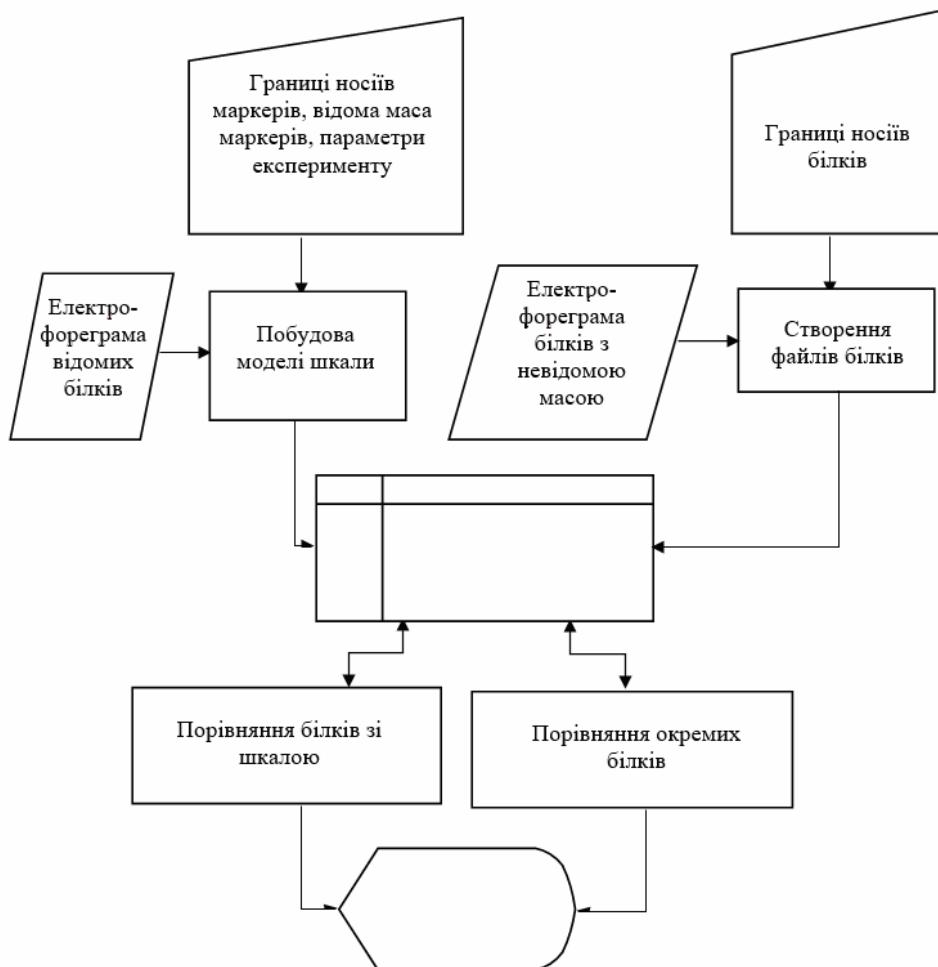


Рис. 6. Структурна схема інформаційної системи

у пам'ять комп'ютера для подальшої обробки.

Модуль порівняння білків зі шкалою є основним в програмі і виконує алгоритм визначення маси білка методами нечіткого аналізу.

Модуль порівняння білків дозволяє провести розрізнення окремих білків та визначити можливість того, що вони справді є різними білками.

Висновки

Програмний продукт FANSPREL, який було розроблено на основі нечіткого аналізу, вирішує задачу досягнення певної точності визначення маси білків. Система дає змогу розрізняти між собою білки за стандартними формами вихідних даних електрофореграм, які до цього моменту ідентифікувалися як один і той же або мали сумнівну ідентифікацію.

СПИСОК ЛІТЕРАТУРИ

1. Geladi P., Esbensen K. Chemometrics, a growing and maturing discipline (Editorial) // Chemom. Intell. Lab. Syst., 7. – 1990. – 197 p. [https://doi.org/10.1016/0169-7439\(90\)80100-K](https://doi.org/10.1016/0169-7439(90)80100-K)
2. Rehm H. Protein biochemistry and proteomics. 1st ed. – London: Academic Press, 2006. – 236 p. <https://www.elsevier.com/books/protein-biochemistry-and-proteomics/rehm/978-0-12-088545-9>
3. Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing / D. Granato, P. Putnik, D.B. Kovacevic, J.S.V. Calado, R.S. Rocha, A.G. Da Cruz, B. Jarvis, O.Ye. Rodionova, A. Pomerantsev // Comp Rev Food Sc Food Saf, 17: – 2018. – P.663-677. <https://doi.org/10.1111/1541-4337.12341>
4. Chemometrics in analytical chemistry – part II: modeling, validation, and applications / R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J.M. Roger, B. Walczak, R. Tauler // Anal Bioanal Chem, 410, 26: – 2018. – P.6691-6704. <https://doi.org/10.1007/s00216-018-1283-4>

5. Application of self-assembly techniques in the design of biocompatible protein microarray surfaces / M. Schaeferling, S. Schiller, H. Paul, M. Kruschina, P. Pavlickova, M. Meerkamp, C. Giammasi, D. Kambhampati // Electrophoresis, 23, – 2002. – 3097 p.

[https://doi.org/10.1002/1522-2683\(200209\)23:18<3097::AID-ELPS3097>3.0.CO;2-G](https://doi.org/10.1002/1522-2683(200209)23:18<3097::AID-ELPS3097>3.0.CO;2-G)

6. Нарыжный С.Н. Введение в протеомику. – Гатчина Ленинградской обл.: Изд-во НИЦ «Курчатовский институт» – ПИЯФ, 2020. – 66 с.

7. Стручкова И.В., Кальясова Е.А. Теоретические и практические основы проведения электрофореза белков в поликариламидном геле: Электронное учебно-методическое пособие. – Нижегородский госуниверситет, 2012. – 60 с. http://www.unn.ru/pages/e-library/methodmaterial/files/Struchkova_Kalyasova.pdf

8. Остерман Л.А. Методы исследования белков и нуклеиновых кислот: Электрофорез и ультрацентрифугирование (практическое пособие). – М.: Наука, 1981. – 288 с. <http://www.enzyme.chem.msu.ru/eduproc/prac/task7/task7book2.pdf>

9. Zaitseva E.A. Deutsche an der Moskauer Universität im 19. Jahrhundert: Ferdinand Friedrich v.Reuss (1778–1852) // In: Deutsch-russische Beziehungen in Medizin und Naturwissenschaften / D.v.Engelhardt u. I.Kdstner (Hgg), Bd.4. Aachen: Shaker Verlag. – 2001. – P.209-226.

10. Kyle R.A., Shampo M.A. Arne Tiselius – father of electrophoresis // Mayo Clinic Proceedings. – 2005. – 80(3). – 302 p. PMID 15757008

11. Езерская А.А., Пивовар М.Л. Капиллярный электрофорез: основные принципы, применение в фармацевтическом анализе // Вестник фармации. – № 1 (83). – 2019. – С.35-44.

https://elib.vsmu.by/bitstream/123/20856/1/vf_2019_1_35-44.pdf

12. Применение 2D-электрофореза для получения «белковых портретов» лизатов бактериальных культур возбудителей особо опасных инфекций / Т.А. Полунина, Ю.С. Варшавская, С.П. Заднова, Я.М. Краснов // Проблемы особо опасных инфекций. – 2016(1). – С.97-101.

<https://doi.org/10.21055/0370-1069-2016-1-97-101>

13. Modified method of fuzzy recognition of proteins in electrophoresis in populationgenetics / Yu.B. Olevska, V.I. Olevskyi, N.M. Ausheva, O.V. Olevskyi // AIP Conference Proceedings 2302, 080006 (2020).

<https://doi.org/10.1063/5.0033556>

14. А.с. № 96330 Україна. Комп'ютерна програма «Система нечіткого аналізу результатів електрофорезу білків (FANSPREL)» / Ю.Б. Олевська, В.І. Олевський, О.В. Олевський (Україна). – 1 с. Опубл. 25.02.2020, Бюл. № 57.

<https://www.m.e.gov.ua/Documents/Detail?lang=uk-UA&id=11c1271e-eb9e-470a-ba8e-43b4e5301bee&title=OfitsiiniiBuletenevtorskePravoISumizhniPrava&isSpecial=true>

РАЗВИТИЕ МЕТОДОВ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ В ПРОТЕОМИКЕ

Олевская Ю.Б., Олевский В.И., Олевский А.В.

Потребностью современных исследований в области медико-биологических наук является решение проблемы повышения точности результатов электрофорографии в их общеупотребительном виде и обоснования основанных на таких экспериментах выводов, в частности, повышение точности идентификации белков. Проблема расчетов в данном направлении решается как универсальными, так и специализированными компьютерными системами, разнообразие которых обусловлено, во-первых, конкретными потребностями естественных наук, и, во-вторых, математическим аппаратом, который применяется для достаточно объемных расчетов. Обычно в таком случае используются методы математической статистики, но в данном случае применение стохастического анализа данных не было достаточно корректным из-за отсутствия репрезентативной выборки или же ее неоднородности. В связи с этими потребностями был создан компьютерный пакет FANSPREL – «Fuzzy AAnalysis System for Protein Electrophoresis» («Система нечеткого анализа результатов электрофореза белков») для обработки электрофорограмм и идентификации белков при применении методов нечеткой математики на основе алгоритмов, разработанных авторами в предыдущих исследованиях. Предлагается подход, основанный на определении нечеткой шкалы по результатам обработки нечетких данных известных белков, распознавании белков, полученных по результатам исследования, и сравнении между собой белков по результатам экспериментов. Математическая часть информационной системы представляет собой построение аппроксимации функции принадлежности на основе полученных данных кривой с четырьмя параметрами и построение нечеткой модели шкалы, определения нечетких масс белков и их сравнение путем конъюнкции нечетких значений, и последующую дефазификацию результатов. Система дает возможность более точной идентификации белков, в частности разграничение тех, массы которых при других формах обработки при определенной точности идентифицировались как одинаковые или был получен неопределенный результат. Программа имеет удобный интерфейс, предусматривает стандартный вид входных и выходных данных и апробирована на решении задач экспериментов, которые происходили для непосредственных нужд сельского хозяйства, в частности для улучшения различных видов почвы, которая используется для выращивания кукурузы.

Ключевые слова: нечеткое моделирование, электрофорограммы, компьютерная система.

Надійшла до редакції 30.11.2020

DEVELOPMENT OF COMPUTER MODELING METHODS IN PROTEOMICS

Olevska Yu.B.^a, Olevskyi V.I.^b, Olevskyi O.V.^c

^a National Technical University «Dnipro Polytechnic», Dnipro, Ukraine

^b Ukrainian State University of Chemical Technology, Dnipro, Ukraine

^c Oles Honchar Dnipro National University, Dnipro, Ukraine

The need for modern research in the field of biomedical sciences is to solve the problem of improving the accuracy of the results of electrophoretography in their common form and substantiating the conclusions based on such experiments, in particular, increasing the accuracy of protein identification. Calculation problems in this direction are solved by both universal and specialized computer systems, the variety of which is due, firstly, to the specific needs of the natural sciences, and, secondly, by the mathematical apparatus that is used for fairly voluminous calculations. Usually, in this case, methods of mathematical statistics are used, but in this case the use of stochastic data analysis was not sufficiently correct due to the lack of a representative sample or its heterogeneity. In connection with these needs, a computer package FANSPREL – «Fuzzy ANalysis System for Protein Electrophoresis» was created for processing electrophoregrams and identifying proteins using fuzzy mathematics methods based on algorithms developed by the authors in previous studies. An approach is proposed based on the determination of a fuzzy scale based on the results of processing fuzzy data of known proteins, recognition of proteins obtained from the results of research, and comparison of proteins among themselves based on the results of experiments. The mathematical part of the information system is the construction of an approximation of the membership function based on the obtained data of a curve with four parameters and the construction of a fuzzy scale model, determination of fuzzy masses of proteins and their comparison by conjunction of fuzzy values, and subsequent defuzzification of the results. The system makes it possible to more accurately identify proteins, in particular to distinguish between those whose masses were identified as the same with other forms of processing with a certain accuracy or an undefined result was obtained. The program has a user-friendly interface, provides a standard form of input and output data, and is tested for solving problems of experiments that took place for the immediate needs of agriculture, in particular to improve various types of soil that is used to grow corn.

Keywords: fuzzy modeling, electrophoregrams, computer system.

REFERENCES

1. Geladi P., Esbensen K. Chemometrics, a growing and maturing discipline (Editorial). Chemom. Intell. Lab. Syst., 7, 1990, 197 p. [https://doi.org/10.1016/0169-7439\(90\)80100-K](https://doi.org/10.1016/0169-7439(90)80100-K)
2. Rehm H. Protein biochemistry and proteomics. 1st ed, London: Academic Press, 2006, 236 p. <https://www.elsevier.com/books/protein-biochemistry-and-proteomics/rehm/978-0-12-088545-9>
3. Granato D., Putnik P., Kovacevic D.B., Calado J.S.V., Rocha R.S., Cruz A.G.Da, Jarvis B., Rodionova O.Ye., Pomerantsev A. Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing. Comp Rev Food Sc Food Saf, 17, 2018, pp.663-677. <https://doi.org/10.1111/1541-4337.12341>
4. Brereton R.G., Jansen J., Lopes J., Marini F., Pomerantsev A., Rodionova O., Roger J.M., Walczak B., Tauler R. Chemometrics in analytical chemistry – part II: modeling, validation, and applications. Anal Bioanal Chem, 410 (26), 2018, pp.6691-6704. <https://doi.org/10.1007/s00216-018-1283-4>
5. Schaeferling M., Schiller S., Paul H., Kruschina M., Pavlickova P., Meerkamp M., Giammasi C., Kambhampati D. Application of self-assembly techniques in the design of biocompatible protein microarray surfaces, Electrophoresis, 23, 2002, pp. 3097. [https://doi.org/10.1002/1522-2683\(200209\)23:18<3097::AID-ELPS3097>3.0.CO;2-G](https://doi.org/10.1002/1522-2683(200209)23:18<3097::AID-ELPS3097>3.0.CO;2-G)
6. Naryzhnyj S.N. *Vvedenie v proteomiku* [Introduction to proteomics]. Gatchina Leningradskoj obl.: Izd-vo NIC «Kurchatovskij institut», PIJaF, 2020, 66 p. (in Russian).
7. Struchkova I.V., Kal'jasova E.A. *Teoreticheskie i prakticheskie osnovy provedenija elektroforeza belkov v poliakrilamidnom gele: Elektronnoe uchebno-metodicheskoe posobie* [Theoretical and practical foundations of carrying out electrophoresis of proteins in polyacrylamide gel: Electronic teaching aid.]. Nizhegorodskij gosuniversitet, 2012, 60 p. (in Russian). http://www.unn.ru/pages/e-library/methodmaterial/files/Struchkova_Kalyasova.pdf
8. Osterman L.A. *Metody issledovanija belkov i nukleinovyh kislot: Elektroforez i ul'tracentrifugirovanie (prakticheskoe posobie)* [Research methods of proteins and nucleic acids: Electrophoresis and ultracentrifugation (practical guide)]. Moscow, Nauka, 1981, 288 p. (in Russian) <http://www.enzyme.chem.msu.ru/eduproc/prac/task7/task7book2.pdf>
9. Zaitseva E.A. Deutsche an der Moskauer Universität im 19. Jahrhundert: Ferdinand Friedrich v.Reuss (1778–1852). In: Deutsch-russische Beziehungen in Medizin und Naturwissenschaften, D.v Engelhardt u. I.Kdstner (Hgg), Bd.4. Aachen: Shaker Verlag, 2001, pp.209-226.
10. Kyle R.A., Shampo M.A. Arne Tiselius – father of electrophoresis. Mayo Clinic Proceedings, 2005, 80(3), 302 p., PMID 15757008
11. Ezerskaja A.A., Pivovar M.L. *Kapilljarnyj elektroforez: osnovnye principy, primenie v farmacevticheskem analize* [Capillary electrophoresis: basic principles, application in pharmaceutical analysis], Vestnik farmacii, no. 1, 2019, pp.35-44. (in Russian) https://elib.vsmu.by/bitstream/123/20856/1/vf_2019_1_35-44.pdf
12. Polunina T.A., Varshavskaja Ju.S., Zadnova S.P., Krasnov Ja.M. *Primenenie 2D-elektroforeza dlja poluchenija «belkovykh portretov» lizatov bakterial'nyh kul'tur vozбудitelej osobu opasnyh infekcij* [Application of 2D-electrophoresis for obtaining «protein portraits» of lysates of bacterial cultures of pathogens of especially dangerous infections], Problemy osobu opasnyh infekcij, 2016, pp.97-101. (in Russian). <https://doi.org/10.21055/0370-1069-2016-1-97-101>
13. Olevska Yu.B., Olevskyi V.I., Ausheva N.M., Olevskyi O.V. Modified method of fuzzy recognition of proteins in electrophoresis in population genetics, AIP Conference Proceedings 2302, 080006 (2020). <https://doi.org/10.1063/5.0033556>
14. A.s. № 96330 Україна. Komp'juterna programma «Sistema nechitkogo analizu rezul'tativ elektroforezu bilkv (FANSPREL)» [Computer program «System of fuzzy analysis of protein electrophoresis results (FANSPREL)】 / Ju.B. Olev's'ka, V.I. Olev's'kij, O.V. Olev's'kij (Ukraina). – 1 s. Opubl. 25.02.2020, Bjul. № 57. (in Ukrainian) <https://www.me.gov.ua/Documents/Detail?lang=uk-UA&id=11c1271e-eb9e-470a-ba8e-43b4e5301bee&title=OfitsiiniiBiuletenvtorskePravoISumizhniPrava&isSpecial=true>