

**Навчально-науковий центр «Інститут біології та медицини»**

**Кафедра біофізики та медичної інформатики**

**О.В. Жолос**

НАВЧАЛЬНИЙ ПОСІБНИК

## **СУЧАСНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ У БІОЛОГІЇ**

Затверджено на засіданні кафедри біофізики та  
медичної інформатики,  
протокол №15 від 26 квітня 2022 року  
Зав. кафедри, проф. О.В. Жолос

---

Затверджено на засіданні Вченої Ради ННЦ  
«Інститут біології та медицини»,  
протокол №4 від 15 листопада 2022 року  
Голова Вченої Ради, проф. Л.І. Остапченко

---

### **Рецензенти:**

**Цимбалюк О.В.**, доктор біол. наук, доцент, професор кафедри молекулярної біотехнології та біоінформатики КНУ імені Тараса Шевченка

**Берест В.П.**, доктор фіз.-мат. наук, доцент, завідувач кафедри молекулярної і медичної біофізики ХНУ імені В.Н. Каразіна

**Київ – 2022**

## ЗМІСТ

ВСТУП .....	4
РОЗДІЛ 1. Основи інформатики та роботи з науковою літературою (PubMed, Mendeley), графічне представлення даних .....	11
1.1. Поняття «інформація», її види, властивості і кількісна характеристика. Роль комп'ютерних технологій в сучасній біології. ....	11
1.2. Робота з числовою та текстовою інформацією за допомогою Microsoft Office, Libre Office та інструментів Google. Пошукова система PubMed та програма Mendeley. ....	33
1.3. Графічне представлення даних векторної графіки за допомогою програм CorelDraw та Inkscape. Графічне представлення даних растрової графіки за допомогою програми ImageJ.....	73
РОЗДІЛ 2. Основи роботи у програмі Origin та програмному середовищі R. ....	83
2.1. Основи роботи з науковими даними у програмі Origin: імпорт даних, їх аналіз, перетворення, апроксимація, створення сценаріїв обробки даних. ....	83
2.2. Можливості Origin для побудови 2D і 3D графіків та кількісного аналізу зображень. ....	90
2.3. Статистичний та автоматизований аналіз даних в програмі Origin. ....	96
2.4. Вступ до мов програмування R та Wolfram Mathematica для біологів. Огляд наявних в них інструментів для візуалізації біологічних даних та ....	102
їх статистичної обробки. Приклади комп'ютерного моделювання з використанням цих програм .....	102
РОЗДІЛ 3. Вступ до біоінформатики та галузі її застосування.....	104
3.1. Основні розділи і галузі застосування біоінформатики. Основні бази даних послідовностей нуклеїнових кислот, білків, сигнальних та метаболічних шляхів.....	104

3.2. Основні підходи та алгоритми біоінформатичного аналізу: кластерний аналіз, ланцюги Маркова, штучні нейронні мережі, штучний інтелект, DeepLearning та інші.....	122
3.3. Основні програми біоінформатики, їх алгоритми і галузі застосування: вирівнювання послідовностей, філогенетичний аналіз, візуалізація і аналіз молекулярних структур.....	128
РОЗДІЛ 4. Практичні роботи.....	134
4.1. Пошук біологічної інформації в Інтернеті. Робота з пошуковою системою PubMed та менеджером цитувань Mendeley.....	134
4.2. Термінологія і функції програми Excel.....	144
4.3. Графічне представлення даних за допомогою програм CorelDraw, Inkscape і BioDraw.....	152
4.4. Аналіз наукових даних за допомогою програми Origin.....	153
4.5. Побудова графіків та кількісний аналіз зображень в Origin та ImageJ. Статистичний і автоматизований аналіз даних.....	155
4.6. Дослідження амінокислотних послідовностей, еволюційних стосунків та видоспецифічних змін в структурі білків.....	164
4.7. Аналіз біологічних даних з використанням мови програмування R.....	170
КОНТРОЛЬНІ ЗАПИТАННЯ.....	175
РЕКОМЕНДОВАНІ ДЖЕРЕЛА.....	177
ДОДАТКИ.....	182

## ВСТУП

“Хто володіє інформацією, той володіє світом”

Натан Ротшильд, Уїнстон Черчилль

Ми живимо в інформаційному суспільстві, де потоки інформації, комп’ютерні та телекомунікаційні технології пронизують буквально всі сфери діяльності як окремої людини, так і суспільства у цілому. Це особливо стосується таких сфер як освіта і наука. Навіть у таких традиційних галузях як медицина, історія розвитку якої така ж давня, як і історія людства, зараз широко запроваджуються комп’ютерні та інформаційні технології, створюються великі бази даних, які акумулюють інформацію щодо поширеності тих чи інших захворювань та їх профілактики (наприклад Центри з контролю та профілактики захворювань в США<sup>1</sup>), електронні медичні картки пацієнтів (Electronic Health Record<sup>2</sup>), широко запроваджується електронна система охорони здоров'я (eHealth<sup>3</sup>) як нова форма діагностики захворювань, їх лікування та менеджменту із застосуванням комп’ютерних технологій та телекомунікаційних засобів, таких як телемедицина.

Глобальні тренди цифровізації суспільства<sup>4</sup> авторитетно ілюструє дослідження вчених Університету Південної Каліфорнії Гілберта і Лопес, яке

---

<sup>1</sup> Centers for Disease Control and Prevention <https://www.cdc.gov/>

<sup>2</sup> Menachemi N., Collum T.H. (2011). Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy* 4, 47–55.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3270933/>

<sup>3</sup> Асоціація eHealth в Україні. <https://ehealth.org.ua/>

<sup>4</sup> Musik C., Bogner A. (2019). Digitalization & society. *Österreich Z Soziol* 44, 1–14.

було опубліковано у журналі *Science* у 2011 р.<sup>5</sup> Автори дослідили світові можливості зберігати, передавати і аналізувати інформацію за допомогою 60 аналогових і цифрових технологій за період з 1986 по 2007 р. і дійшли таких висновків. У 2007 р. людство змогло зберігати  $2,9 \times 10^{20}$  байт інформації, обмінюватись  $2 \times 10^{21}$  байт інформації у цифровому форматі і виконувати  $6,4 \times 10^{18}$  інструкцій в секунду за допомогою персональних комп'ютерів (ПК). За цей період (1986-2007 рр.) зростання комп'ютерної потужності становило у середньому 58% щороку. Паралельно з цим зростали можливості телекомунікаційних технологій (28% щороку) і зберігання даних у цифровому форматі (23% щороку). У своєму коментарі для BBC World Service Гілберт для наочності цих даних відзначив, що якби всю наявну у 2007 р. інформацію, а це 295 ексабайтів ( $295 \times 10^{18}$  байтів), надрукувати у книжках, то їх би вистачило, щоб вкрити територію США 13 разів, а якщо її записати на компакт дисках, то їх стопка була б вищою, ніж відстань від Землі до Місяця, тобто понад 400 тис. км! І якщо у 2000 р. 75% усієї інформації зберігалось в аналоговому форматі, то вже у 2007 р. 94% інформації зберігалась у цифровому форматі, і таким чином вона стала доступною для використання із застосуванням комп'ютерних технологій. Цей феномен дістав назву інформаційної революції. Автори цього дослідження підкреслюють, що кожні 40-60 років у суспільстві щось починає розвиватись прискореними темпами, і якщо раніше електрика чи автомобіль повністю змінили суспільство, то зараз це інформація. Прочитуємо їх слушну думку дослівно: «Те, що ви можете робити з інформацією - це передавати її

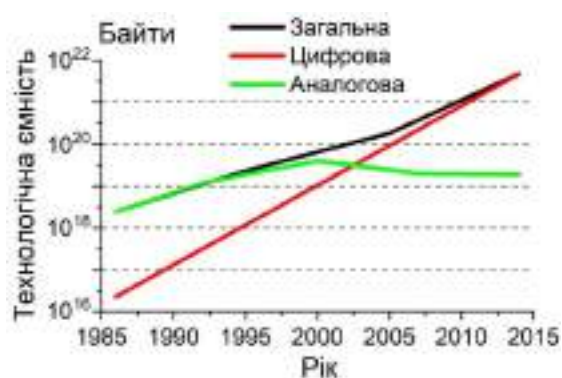
---

<sup>5</sup> Hilbert M., López P. (2011). The World's technological capacity to store, communicate, and compute information. *Science* 332, 60-65.

<https://www.science.org/doi/10.1126/science.1200970>

через простір, і ми називаємо це комунікацією. Ви можете передати її через час, і ми називаємо це зберіганням. Або ви можете трансформувати її, маніпулювати нею, змінити її сенс, і ми називаємо це обчисленням». Датою настання цифрового віку можна вважати 2002 р. – саме тоді обсяги інформації в аналоговій і цифровій формах зрівнялись між собою, а далі відбулось стрімке, експоненційне зростання останньої, як показав більш недавній аналіз цих трендів<sup>6</sup> (зверніть увагу, що на цій ілюстрації трендів кількість інформації по висі ординат представлена у логарифмічному масштабі).

З точки зору біології цікаво відмітити, що за оцінками авторів цього дослідження кількість інформації, яка міститься у ДНК однієї людини, у 300 разів перебільшує її кількість в усіх технологічних пристроях усього світу (так звана технологічна ємність).



Свого часу найбільш потужні імпульси для розвитку комп'ютерних технологій надали дві галузі. Одна з них стосувалась космічних досліджень і керування космічними апаратами, а інша – складного аналізу рентген-структурних даних будови білків та інших біологічних макромолекул. Сучасні біологічні дослідження продукують величезний обсяг даних, зокрема це пов'язано з геномною революцією, а також із широким запровадженням технологій, які здатні швидко генерувати велику кількість даних: ДНК-мікрочип (DNA microarray), мас-спектроскопія, секвенування наступного

<sup>6</sup> The data revolution. World development report 2016.

[https://documents1.worldbank.org/curated/en/896971468194972881/310436360\\_201602630\\_200201/additional/102725-PUB-Replacement-PUBLIC.pdf](https://documents1.worldbank.org/curated/en/896971468194972881/310436360_201602630_200201/additional/102725-PUB-Replacement-PUBLIC.pdf)

покоління (next-generation sequencing, скорочено NGS), двогібридний скринінг (two-hybrid screening), роботизований планарний петч-клемп та інші високопродуктивні (так звані high throughput) технології. Останнім часом сформувалась і швидко прогресує так звана «біологія великих масивів даних» (Big Data Biology), яка створює не тільки безпрецедентні наукові можливості, але і багато викликів для вчених-біологів<sup>7</sup>. Біологічні лабораторії вже поділились на «мокрі» (wet labs проводять традиційні дослідження *in vitro*, *in situ*, *in vivo*) і «сухі» (dry labs, які працюють *in silico*). Багато вчених (у тому числі і автор цих рядків) все більше переконуються у тому, що оптимум лежить десь посередині, і найкраще, якщо в одній і тій самій лабораторії поєднуються експериментальні дослідження з комп'ютерними технологіями. Ця ситуація дещо нагадує історію становлення біофізики як окремої біологічної дисципліни. На початку здавалось, що для створення біофізичної лабораторії біологам достатньо буде залучити до співпраці фізиків, але все виявилось значно складнішим... Тим не менше, є багато яскравих прикладів, коли відкриття були зроблені виключно *in silico*. Один з таких прикладів наводиться у статті, яка була опублікована у журналі Nature у 2013 р.<sup>8</sup> Група вчених з університету штату Монтана відкрила 11 нових ретровірусів риб жодного разу не доторкнувшись до води в озері чи лабораторії! Це було зроблено шляхом комп'ютерного аналізу геномів. Більш того, дослідники розробили власне програмне забезпечення для пошуку вірусів людини. В процесі дослідження ними були згенеровані тисячі гігабайт інформації, яку вони надали у вільний

---

<sup>7</sup> Leonelli S. (2019). The challenges of big data biology. *eLife* 8, e47381.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6450665/>

<sup>8</sup> Marx V. (2013). The big challenges of big data. *Nature* 498, 255-260.

<https://www.nature.com/articles/498255a.pdf>

доступ іншим вченим. Це яскравий приклад того, як працює сучасна біологія. Прочитуємо звернення до студентів одного з авторів цього дослідження: «До початку занять мої студенти не уявляли, що вони можуть отримати нові, значущі знання і сформулювати нові гіпотези на основі вже наявних, не своїх даних. Великі дані в біології створюють нові можливості для вчених, оскільки в базах даних вони «недоаналізовані» по всьому світу».

Кожне наукове дослідження, від курсової роботи до статті у рейтинговому міжнародному журналі, починається з аналізу літератури стосовно конкретної проблеми, формулювання робочих гіпотез і плану досліджень. Зберігання отриманих даних (а більшість з них у сучасній біології вже у процесі реєстрації є у цифровому форматі), їх аналіз і графічне представлення є інтегральними складовими дослідження. І останній етап – обговорення отриманих результатів і формулювання висновків – знову таки вимагає звернення до літературних джерел з тим, щоб визначити новизну і значущість результатів, порівняти їх з відомими фактами та коректно інтерпретувати. Кожен з цих етапів зараз базується на застосуванні сучасних комп'ютерних технологій. Тому у цьому курсі ми не будемо обмежуватись питаннями біоінформатики, саме з якою насамперед і виникає асоціація при згадуванні сучасних комп'ютерних технологій в біології, хоча питанням біоінформатики теж буде приділена увага.

Нашою загальною метою буде формування та розвиток компетентностей у галузі сучасних інформаційно-комунікаційних технологій (ІКТ) для забезпечення ефективного, раціонального використання сучасного програмного забезпечення загального та спеціального призначення при опрацюванні біологічних даних, і зокрема вміння знаходити у глобальних мережах інформацію професійного характеру, застосовувати програмне забезпечення для отримання, зберігання, аналізу та обміну інформацією у своїй професійній діяльності. Не менш важливо і те, щоб студенти відчували перспективи у цьому

напрямку, адже бази біологічних даних постійно зростають за обсягом і якістю представлених у них даних, і так само постійно розвивається арсенал наших методичних можливостей для їх опрацювання – програми, алгоритми, підходи, автоматизація всіх ланок цієї складної роботи.

Сфера застосування таких знань є надзвичайно багатопрофільною. При вивченні *першого розділу* студенти навчаються професійно проводити пошук потрібних їм по конкретній темі публікацій в базах даних PubMed, Scopus та Web of Science, а також формуванню списків першоджерел за допомогою бібліографічного менеджера Mendeley. Вони також отримають навички аналізу даних за допомогою MS Excel і роботи з графічними зображеннями в CorelDraw, BioDraw, Inkscape та ImageJ, познайомляться з питаннями, як у сучасному світі науки оцінюються досягнення, тобто різними наукометричними показниками.

У *другому розділі* основна увага буде приділена роботі з програмою Origin, яка зараз є однією з найбільш широко вживаних для наукових і інженерних розрахунків, фітування та графічного представлення даних. У цьому ж розділі ми розглянемо і питання як за допомогою нескладних програм (швидше сценаріїв, або користувацьких скриптів, Додаток 2) можна автоматизувати цю складну роботу і тим самим значно зекономити час для проведення експериментів. Більше того, автоматизація аналізу даних дозволяє наочно побачити результат у процесі експерименту у реальному часі, і за потреби ввести потрібні корективи у протокол дослідження. Цей розділ завершується вступом до мови програмування R для біологів.

*Третій розділ* є більш «біоінформатичним». Розглядаються основні бази даних біологічної інформації та типові онлайн-інструменти для біоінформатичного аналізу. При його опрацюванні студенти дізнаються про ресурси Національного центру біотехнологічної інформації (National Center for

Biotechnology Information, або NCBI), Європейського інституту біоінформатики (EMBL's European Bioinformatics Institute, або EMBL-EBI), познайомляться з базами даних, алгоритмами аналізу даних та програмним забезпеченням, зокрема для вирівнювання нуклеотидних і амінокислотних послідовностей, філогенетичного аналізу, візуалізації та аналізу молекулярних структур. Цей розділ проілюстровано прикладами успіхів використання сучасних інформаційних технологій в біології. Ці приклади підібрані за двома критеріями. По-перше, такі дослідження у принципі не могли бути б проведені без застосування інформаційних комп'ютерних технологій. По-друге, вони ілюструють широкий спектр і міждисциплінарний характер проблем з тим, щоб незалежно від обраної спеціалізації студент чи студентка побачили б власну перспективу опанування сучасними інформаційними технологіями в галузі біології. Як працює серце людини на молекулярному рівні, як плаває медуза чи літає комар, скільки дерев є на нашій планеті, як шляхом молекулярних моделювання, динаміки і докінгу створюються нові лікарські препарати, як можна прогнозувати розвиток ракової пухлини – відповіді на ці та інші питання нам дали саме сучасні комп'ютерні технології у поєднанні з експериментом.

Метою останнього, *четвертого розділу* є перехід «від теорії до практики». У ньому наведені інструкції до практичних робіт у цьому курсі, які студенти виконують протягом семестру. Ці інструкції деталізовані і проілюстровані з тим, щоб роботи можна було виконати цілком самостійно. Крім того, у цьому та інших розділах наводяться гіперпосилання на додаткові ресурси, які важливо використовувати під час вивчення цієї дисципліни.

Даний посібник не варто розглядати як вичерпний для опанування цим предметом, швидше метою автора було допомогти студентам визначити власну «дорожню карту» у широкій мережі сучасних інформаційних технологій в біології.

## РОЗДІЛ 1. ОСНОВИ ІНФОРМАТИКИ ТА РОБОТИ З НАУКОВОЮ ЛІТЕРАТУРОЮ (PUBMED, MENDELEY), ГРАФІЧНЕ ПРЕДСТАВЛЕННЯ ДАНИХ

### 1.1. Поняття «інформація», її види, властивості і кількісна характеристика. Роль комп'ютерних технологій в сучасній біології.

Термін «інформація» має латинське походження від слова «*informatio*», який розуміється як представлення фактів чи подій, їх виклад, роз'яснення. Наприклад, ще Цицерон використовував цей термін у контексті теорії ідей або форм давньогрецького мислителя Платона. У першій половині ХХ ст. «інформація» стала предметом математичних досліджень та теорій, особливо у галузях статистики та теорії ймовірності. Насамперед це стосується робіт Клода Шеннона (Shannon, 1948), про які ми згадаємо детальніше далі. Зараз теорія інформації широко використовується у багатьох галузях, таких як математика, фізика, комп'ютерні науки, логіка та лінгвістика.

Кембридзький словник англійської мови<sup>9</sup> визначає інформацію як «facts about someone or something». Britannica Dictionary<sup>10</sup> додає до цього визначення «знання, які ви отримуєте про когось або щось: факти або деталі стосовно предмету» («knowledge that you get about someone or something: facts or details about a subject»). Таке тлумачення близьке до оригінального значення лат. «*informatio*», що означає ознайомлення, знання, навчання. Синонімами цього слова є дані, факти. Приклади, які наводяться в якості ілюстрації цього терміну, добре ілюструють те, що інформація – це дуже широке поняття, яке може мати досить різне значення в залежності від контексту. Наведемо деякі з таких прикладів<sup>9,10</sup>:

---

<sup>9</sup> Cambridge Dictionary <https://dictionary.cambridge.org/>

<sup>10</sup> Britannica Dictionary <https://www.britannica.com/dictionary>

- Чи є у вас будь-яка *інформація* про розклад поїздів?
- Я прочитав цікаву *інформацію* в газеті.
- Я відвідав декілька бібліотек, щоб зібрати *інформацію* по цій темі.
- Сучасні комп'ютери можуть зберігати велику кількість *інформації*.
- Цей документ містить конфіденційну *інформацію*.
- Ми не маємо більш детальної *інформації* стосовно цієї події.

Merriam-Webster словник для студентів деталізує це поняття у залежності від контексту наступним чином:

1: надання чи отримання знань;

2а: знання, які були отримані в результаті дослідження або навчання;

2б: знання про конкретну подію чи ситуацію - новини;

2в: характеристика або якість, наприклад ДНК чи комп'ютерна програма;

3: якість чогось (наприклад, ДНК або комп'ютерної програми), що представляє один конкретний вибір з низки можливих інших варіантів.

Зупинимось на останньому визначенні, адже воно є ключовим для розуміння більш технічних аспектів терміну «інформація». Дійсно, отримання інформації є тим засобом, що дозволяє принести більше впорядкованості і визначеності для нашого розуміння тих чи інших подій, фактів чи даних. Якщо ентропія характеризує міру невизначеності певної системи, то отримання інформації від системи, навпаки, є мірою зменшення ентропії. Іншими словами, якщо невизначеність події (тобто ймовірність її реалізації) є високою, то відповідно потрібно отримати велику кількість інформації, щоб вирішити цю невизначеність. Згадаємо визначення українського вченого-кібернетика, піонера комп'ютерних технологій і засновника Інституту кібернетики НАН України, який зараз носить його ім'я, академіка В.М. Глушкова: «Інформація у самому загальному її розумінні являє собою міру неоднорідності розподілу матерії та

енергії у просторі та у часі, міру змін, якими супроводжується всі процеси, що протікають у світі...».

Проілюструємо це положення простим прикладом. Якщо підкинути монету і гральну кісточку, то доти, поки вони обертаються у повітрі їх кінцевий стан є невизначеним, але коли вони впадуть на землю, то будуть реалізовані відповідно 1 з 2 або 1 з 6 можливих їх станів. Отже, неважко зрозуміти, що міра невизначеності, а адже і кількість інформації у гральній кісточці є більшою, ніж у монети. Чи можна якимось чином кількісно схарактеризувати цю різницю? Одиницею інформації у цифровому вигляді є 1 біт (скорочення від англійського binary digit). Це мінімальна кількість інформації, один двійковий розряд, який може приймати лише одне з двох значень - 0 або 1. В контексті інформації це така змінна, значення якої буде відомим, коли відбудеться одна з двох однаково ймовірних подій. У нашому прикладі це падіння монети на одну з її сторін. Отже, повідомлення про реалізацію такої події містить 1 біт інформації. Щодо гральної кісточки, то у цьому прикладі буде реалізована 1 з 6 незалежних і однаково ймовірних подій, а повідомлення про це буде містити кількість інформації, яку можна розрахувати за формулою, яку запропонував у 1928 р. американський вчений Ральф Хартлі:

$$I(N) = \log_2 N$$

відповідно до якої інформація  $I$ , яка міститься в одному повідомленні, розраховується як логарифм за основою 2 загального числа можливих повідомлень  $N$ . У нашому прикладі з гральною кісточкою  $I = \log_2 6 = 2,585$  біт. З цього рівняння витікає, що якщо подія є повністю визначеною, тобто вона гарантовано відбудеться ( $N=1$ ), то вона не несе ніякої нової інформації, адже  $\log_2 1=0$ .

Якщо в загальній кількості подій  $N$  деякі з них  $N_i$  виникають більш ймовірно, а інші мають меншу ймовірність реалізації  $P_i$  (наприклад, білок має

певні більш стабільні, тобто більш ймовірні конформаційні стани), то для обчислення кількості інформації в такій системі використовується формула, яку запропонував американський математик Клод Шеннон у 1948 р.:

$$I(N) = P_1 \log_2 N_1 + P_2 \log_2 N_2 + \dots + P_i \log_2 N_i$$

Можна побачити, що за умови, що всі події однаково ймовірні ( $P_1=P_2=\dots=P_i$ ) ми отримуємо формулу Хартлі, яка таким чином є частковим варіантом формули Шеннона.

Клод Шеннон є засновником теорії інформації, і зокрема саме він почав використовувати термін «біт». У цифрових технологіях і телекомунікаціях використовується інша одиниця: 1 байт=8 біт. 1 байт – це та мінімальна кількість інформації, яка необхідна для представлення певного знаку (літера, цифра тощо) у цифровому виді за допомогою бінарного коду ASCII (American Standard Code for Information Interchange). Наприклад, дерево можна представити як рисунок або записати як слово на клавіатурі комп'ютера, де кожна літера буде представлена в коді ASCII наступним чином:

0101 0100 - t

0111 0010 - r

0110 0101 - e

0110 0101 - e

Для бінарної репрезентації більш широкого спектру унікальних символів (інші мови тощо) було розроблено новий стандарт Unicode (Universal Coded Character Set), в якому кожний унікальний символ представлений числом, що може містити до 4 байт. Для кодування літер та інших символів широко використовується також HTML (Hypertext Markup Language). В Інтернет технологіях домінує стандарт кодування UTF-8 (Unicode Transformation Format – 8-bit). У ньому для кодування кожного унікального символу використовується від 1 до 4 однобайтних одиниць, що дозволяє кодування понад 1 млн. символів.

Час, коли ємність жорстких дисків вимірювалась у мегабайтах ( $10^6$ ), вже минув. Зараз об'єми інформації вимірюються у гіга ( $10^9$ )-, тера ( $10^{12}$ )- та пета ( $10^{15}$ )-байтах, а масивний обмін інформацією через Інтернет вже дістав назву екса ( $10^{18}$ ) повню (exaflood). Наведемо одиниці кількості інформації:

$$1 \text{ байт} = 2^3 \text{ біт} = 8 \text{ біт}$$

$$1 \text{ кбайт} = 2^{10} \text{ байт} = 1024 \text{ байт}$$

$$1 \text{ мбайт} = 2^{10} \text{ кбайт} = 1024 \text{ кбайт (або (або 1 048 576 байт))}$$

$$1 \text{ гбайт} = 2^{10} \text{ мбайт} = 1024 \text{ мбайт}$$

$$1 \text{ терабайт} = 2^{10} \text{ гбайт} = 1024 \text{ гбайт}$$

$$1 \text{ петабайт} = 2^{10} \text{ терабайт} = 1024 \text{ терабайт}$$

У сучасних цифрових пристроях інформація зберігається у вигляді біт, хоча передаватись між ними вона може і значно більшими масивами. Інформація може передаватись як у часі шляхом її зберігання, наприклад на комп'ютерних жорстких дисках, так і у просторі із застосуванням телекомунікаційних технологій.

До основних властивостей інформації відносяться наступні її характеристики (Табл. 1.1.1):

**Таблиця 1.1.1.** Основні властивості інформації.

Характеристика	Її тлумачення
Точність, коректність	Наскільки інформація є деталізованою і коректною в усіх її деталях. Є однією з ключових властивостей інформації, адже помилкові або неточні вихідні дані унеможливають їх правильну інтерпретацію. Також важливо, щоб всі користувачі сприймали ту чи іншу інформацію однозначним чином.

Повнота	Наскільки вона є вичерпною. Неповна інформація може зробити її непридатною для використання.
Надійність, достовірність, безпомилковість	Наскільки інформація є достовірною і чи не суперечить вона іншим надійним джерелам інформації. Це ще одна з ключових характеристик інформації, адже якщо у даних з різних джерел виявляються розбіжності, то таким даним не можна довіряти.
Актуальність, своєчасність	Характеризує наскільки інформація є дійсно потрібною її користувачеві, а також наскільки інформація є сучасною, наприклад чи відповідає вона вимогам для її опрацювання у режимі реального часу. Застаріла інформація значно втрачає свою цінність.
Цінність	Наскільки інформація відповідає певній конкретній меті користувача.
Доступність	Характеризує можливість отримання певної інформації користувачем.
Захищеність	Наскільки можливим є отримання недозволеного доступу тим користувачем, для кого ця інформація не призначена.

Дещо парадоксальним є те, що хоча інформація є абстрактним поняттям, вона існує лише на певних матеріальних носіях – книги, компакт диски, жорсткі диски комп'ютерів тощо, які є носіями інформації. У живих системах інформація зберігається і передається молекулами (ДНК, РНК, рецептори, сигнальні мережі, іонні канали, які відповідають за генерацію та розповсюдження потенціалів дії, тощо).

Біологам буде добре зрозумілою класифікація інформації за способом її сприйняття, адже вона стосується наших органів чуттів: візуальна, звукова,

нюхова, смакова, тактильна. В залежності від її матеріального носія чи засобу передачі можна розрізнити різні види інформації за формою її представлення: текстова, числова, звукова, у вигляді зображень (графіки, діаграми тощо) та ін. Нарешті, інформація може відрізнятись в залежності від її призначення, з усіма градаціями від масової до суто особистої інформації. Так, зокрема, спеціальна інформація призначена для відносно вузького кола фахівців. Кожна галузь знань також має значну специфіку, у тому числі термінологічного характеру, і відповідно можна розрізнити інформацію медичного, біологічного, технічного, економічного, соціального тощо характеру.

Хоча ці питання виходять за межі цього посібника, варто зазначити, що інформація тісно пов'язана з такими науковими дисциплінами як математика (зокрема теорія ймовірності) і кібернетика, які є теоретичною основою науки про інформацію – інформатики (інколи використовують термін «комп'ютерні науки», але у даному контексті це менш вдалий термін), яка досліджує структуру представлення та трансформації/передачі інформації у технічних пристроях.

Як вже відмічалось, інформатизація пронизує всі галузі сучасного суспільства, і у більшості країн вона регулюється на законодавчому рівні. У 1992 р. був прийнятий закон України «Про інформацію»<sup>11</sup>, який визначає терміни (зокрема сам термін інформація як «будь-які відомості та/або дані, які можуть бути збережені на матеріальних носіях або відображені в електронному вигляді» і регулює «відносини щодо створення, збирання, одержання, зберігання, використання, поширення, охорони, захисту інформації».

Ми вже наводили дані досліджень щодо сумарного обсягу інформації, яка зберігається на різноманітних носіях у сучасному суспільстві і оцінюється у

---

<sup>11</sup> Закон України «Про інформацію». <https://zakon.rada.gov.ua/laws/show/2657-12#Text>

декілька сотень ексабайтів, тобто порядку  $10^{20}$ - $10^{21}$  байтів. Не менш вражають і оцінки кількості інформації у живих системах. Так, для яйцеклітини ссавця це приблизно  $6 \times 10^9$  бітів, а для організму людини -  $3 \times 10^{25}$  бітів в молекулярному представленні, або  $2 \times 10^{28}$  в атомарному представленні. За деякими оцінками на свідомому рівні людина може обробляти 100 бітів інформації за секунду, а на підсвідомому –  $10^9$  бітів за секунду. Дійсно, з математичними обчисленнями комп'ютери справляються набагато більш ефективно, ніж люди. Можна навести такий приклад. У Німеччині кожні 2 роки проходить Mental Calculation World Cup. На останньому з них (2018 р.) були встановлені декілька рекордів:

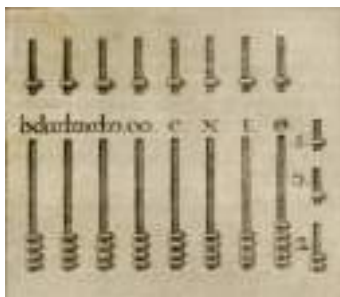
- додавання десяти 10-значних чисел - 10 завдань за 242 сек.
- множення двох 8-значних чисел - 10 завдань за 295 сек.
- корінь квадратний з 6-значних чисел – 10 правильних розв'язків за 135 сек.

Звісно, комп'ютер вирішив би такі завдання за долі секунди. Сучасні комп'ютерні технології створюють безпрецедентні можливості для прискорення обробки та аналізу інформації, представлення її у найбільш доступному форматі, надійного зберігання у великих базах даних, а також для комунікації даних. Далі ми познайомимось з тим, як саме відбувається прогрес цих технологій і яку роль вони відіграють у сучасній біології.

### **Роль комп'ютерних технологій в сучасній біології.**

Прогрес людства у створенні обчислювальних пристроїв можна візуалізувати на Рис. 1.1.1. В Україні перші електронні обчислювальні машини були створені на базі Інституту кібернетики під керівництвом академіка В.М. Глушкова у 1965 р.

Римська рахівниця  
(відома до н.е.)



Арифметична машина  
(Франція, 1673 р.)



Механічний калькулятор  
(1900 р.)



Один з перших  
електронних “універ-  
сальних” комп’ютерів  
(Pilot Ace, 1950 р.)



Перший  
“мікрокомп’ютер”  
(Altair 8800, 1974 р.)



Суперкомп’ютер  
«Blue Gene»  
(1999 - )



**Рис. 1.1.1.** Ілюстрація механічних та електронних пристроїв для обчислень, від рахівниці до сучасного суперкомп’ютера «Blue Gene», який застосовується для обчислень, цифрової реконструкції і симуляції функції мозку миші у проєкті Blue Brain у Швейцарії<sup>12</sup> (модифіковано з публічних доменів Інтернету).

Кожного разу, коли ми купуємо новий ПК, ноутбук чи смартфон ми на власному досвіді пересвідчуємось, що відбувається неспинний, стрімкий прогрес апаратного (hardware) та програмного (software) забезпечення цих пристроїв. Ще більш вражальним він є у галузі створення суперкомп’ютерів, сам факт розробки і постійного вдосконалення яких, незважаючи на

<sup>12</sup> Blue Brain Project. <https://www.epfl.ch/research/domains/bluebrain/>

надзвичайно високу вартість виробництва та експлуатації, можливо якнайкраще ілюструє високу затребуваність комп'ютерної обробки даних у сучасному суспільстві.

Але чи можна якимось чином кількісно схарактеризувати цей прогрес і чи існують якісь закономірності його поступу? Продуктивність комп'ютерів, особливо у галузі наукових та інженерних розрахунків, вимірюється в одиницях, які відомі як FLOPS (Floating Point Operations Per Second — “операцій з плаваючою комою на секунду”). Відповідна формула для її розрахунку:

$$\text{FLOPS} = \text{кількість ядер процесора} \times \text{частота процесора} \times \text{FLOPS/такт}$$

Наприклад, для процесора, що виконує 4 FLOPS за цикл, має 4 ядра (Core 2 Quad) і працює на частоті 3,5 ГГц, теоретична межа його продуктивності становить:

$$4 \times 3,5 \times 10^9 \times 4 = 56 \text{ гігафлопс } (56 \times 10^9 \text{ флопс})$$

З огляду на значне зростання продуктивності сучасних процесорів один з її виробників, Intel, останнім часом в якості базової одиниці почав використовувати GFLOPS (1 гігафлопс), або APP (Adjusted Peak Performance). Наведемо деякі дані з посиланням на Intel<sup>13</sup>, які наочно ілюструють динаміку збільшення продуктивності процесорів (Табл. 1.1.2):

---

<sup>13</sup> Where can I find information about FLOPS per cycle for Intel® processors?

<https://www.intel.com/content/www/us/en/support/articles/000057415/processors.html>

**Таблиця 1.1.2.** Продуктивність деяких мікропроцесорів Intel® Core™.

<b>Тип</b>	<b>GFLOPS</b>	<b>APP</b>
<b>4 покоління</b>		
i7-4550U	48	0,0144
<b>5 покоління</b>		
i7-5550U	64	0,0192
<b>6 покоління</b>		
i7-6500U	80	0,0240
<b>7 покоління</b>		
i7-7500U	86,4	0,02592
<b>8 покоління</b>		
i7-8557U	108,8	0,03264
<b>9 покоління</b>		
i7-9700TE	230,4	0,06912
<b>10 покоління</b>		
i7-10510U	115,2	0,03456
<b>11 покоління</b>		
i7-11375H	211,2	0,06336

Дані наведені у Табл. 1.1.3 далі характеризують основні тренди у розвитку комп'ютерних технологій за понад 70 років.

**Таблиця 1.1.3.** Продуктивність мікропроцесорів ПК та суперкомп'ютерів.

<b>Персональні комп'ютери</b>	<b>Суперкомп'ютери</b>
IBM PC/XT 4,77 МГц (1983 р.) — 6,9 Кфлопс	ENIAC (1946 р., маса 27 т) — 300 флопс
Intel 80386 40 МГц (1985 р.) — 0,6 Мфлопс	IBM 709 (1957 р.) — 5 Кфлопс
<b>Мега 10<sup>6</sup></b>	<b>Мега 10<sup>6</sup></b>
Intel Pentium 75 МГц (1993 р.) — 7,5 Мфлопс	CDC 6600 (1964 р.) — 3 Мфлопс
Intel Pentium II 300 МГц (1997 р.) — 50 Мфлопс	БЭСМ-6 (1968 р.) — 1 Мфлопс
<b>Гіга 10<sup>9</sup></b>	Cray-1 (1974 р.) — 133 Мфлопс
Intel Pentium III 1 ГГц (1999 р.) — 2 Гфлопс	Ельбрус-2 (1984 р.) — 125 Мфлопс
AMD Athlon XP1800+ 1533 МГц (2002 р.) — 3 Гфлопс	Електроніка СС БІС (1991 р.) — 500 Мфлопс
AMD Athlon 64 2,211 ГГц (2003 р.) — 8 Гфлопс	<b>Гіга 10<sup>9</sup></b>
AMD Athlon 64 X2 4200 2,2 ГГц (2006 р.) — 13,2 Гфлопс	Cray-2 (1985 р.) — 1,9 Гфлопс
Intel Core 2 Duo 2,4 ГГц (2006 р.) — 19,2 Гфлопс	Cray Y-MP (1988 р.) — 2,3 Гфлопс
Intel Core i7-975 XE 3,33 ГГц (2009 р.) — 53,28 Гфлопс	<b>Тера 10<sup>12</sup></b>
Intel Core i7-3930K 3,8 ГГц, 6-cores (2011 р.) — 182 Гфлопс	ASCI Red (1993 р.) — 1 Тфлопс
<b>Тера 10<sup>12</sup></b>	IBM Blue Gene/L (2006 р.) — 478,2 Тфлопс
Intel's Core i9, 4.5GHz, 18 core (2017 р.)	<b>Пета 10<sup>15</sup></b>
	Cray Jaguar (2008 р.) — 1,06 Пфлопс
	IBM Roadrunner (2008 р.) — 1,04 Пфлопс
	Jaguar Cray XT5-NE (2009 р.) — 1,76 Пфлопс
	Ломоносов (2011 р.) — 1,3 Пфлопс
	IBM Sequoia (2012 р.) — 20 Пфлопс
	Tianhe-2 (2013 р.) — 33,86 Пфлопс

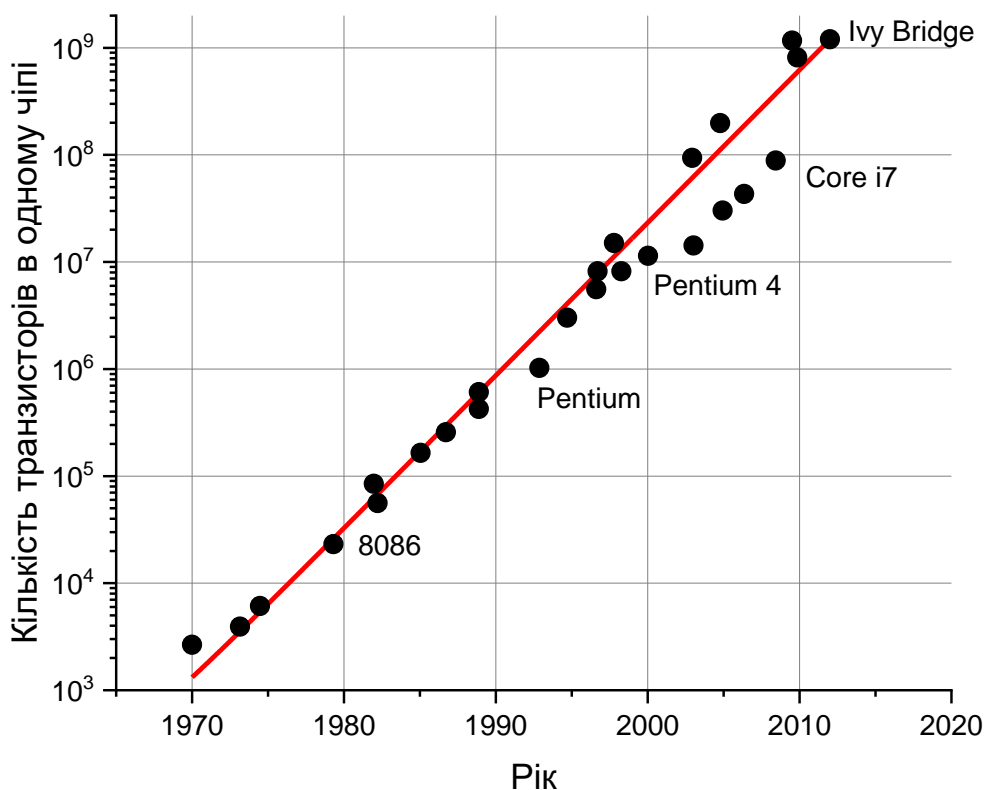
Як бачимо, потужність суперкомп'ютерів випереджає ПК на 15-25 років. Звісно, при цьому відбувалось і продовжує відбуватись різке зменшення вартості застосування комп'ютерних технологій, у результаті чого вони стали загальнодоступними. Так, якщо у 1984 р. найдешевшою системою з продуктивність 1 Гфлопс був би суперкомп'ютер Cray X-MP вартістю понад 8 млн. доларів, то у 2015 р. вартість 1 Гфлопса зменшилась до 8 центів (процесор графічної карти AMD Radeon R9 295X2).

Такі тренди передбачив ще у 1965 р. американський інженер і співзасновник корпорації Intel Г. Мур (Gordon E. Moore) у статті, яка була опублікована у журналі Electronics Magazine. На основі достатньо обмеженої на той час кількості даних він сформував загальний постулат, який пройшов випробування часом дотепер. Емпіричний закон Мура говорить про те, що кількість транзисторів у силіконовому чіпі подвоюється щороку. У 1975 р. Мур уточнив цей період і збільшив його до 2 років, тоді як за 40 років починаючи з 1961 р. таке подвоєння відбувалось кожні 18 місяців, тобто в межах прогнозованого Г. Муром (Рис. 1.1.2).

У 2007 р. Г. Мур відмітив, що таке зростання не може бути нескінченним, адже природа накладає певні обмеження, зокрема це стосується швидкості розповсюдження електромагнітних імпульсів. Але зараз швидко розвиваються новітні технології, такі як нанокомп'ютери (на базі напівпровідникових транзисторів розміром менше 100 нм), які можна побачити лише за допомогою мікроскопу, квантові комп'ютери, штучний інтелект.

Вчені-біологи також спробували оцінити наскільки швидко працює мозок людини. Такі оцінки базувались як на швидкості обробки візуальної інформації (Н. Moravec), так і на кількості нейронів головного мозку разом з тим, наскільки швидко відбувається передача сигналів між окремими нейронами (С. Westbury).

Вони дали відповідно  $100 \times 10^{12}$  флопс (100 терафлопс) та  $20 \times 10^{15}$  (20 петафлопс) операцій в секунду.



**Рис. 1.1.2.** Дані для різних мікропроцесорів накопичені за останні приблизно 50 років підтверджують передбачення Г. Мура<sup>14</sup>. Зростання відбувається по експоненті (зверніть увагу на логарифмічний масштаб по вісі ординат).

Аналогічним чином паралельно з процесорами відбувається експоненційне зростання ємності твердих дисків разом із зменшенням вартості зберігання одиниці інформації. Ще у 2005 р. М. Крайдер (Mark Kryder, 2005) прогнозував, що щільність запису інформації на магнітних носіях буде

---

<sup>14</sup> König K, Ostendorf A, editors. *Optically Induced Nanostructures: Biomedical and Technical Applications*. Berlin: De Gruyter; 2015.

подвоюватись кожні 14-18 місяців. Для порівняння, у 1956 р. ця щільність становила 3,1 біт/мм<sup>2</sup> поверхні диску, вона збільшилась до 155 Кбіт/мм<sup>2</sup> в 1990-х роках і до 170 Мбіт/мм<sup>2</sup> в 2005 р. Щоправда, останнім часом вартість зберігання інформації не зменшується так швидко, як передбачає закон М. Крайдера, і навіть стабілізується. Надійне і довготривале зберігання даних має певні технічні обмеження. Цікаво, що вже була продемонстрована принципова можливість запису цифрової інформації в ДНК як одного з найбільш надійних носіїв інформації (згадаємо успішне секвенування геному вимерлих мамонтів). Ці результати були опубліковані у журналі *Nature* у 2013 р. і показали можливість запису, а потім зчитування файлів у форматах тексту (ASCII коди), PDF, JPEG та MP3<sup>15</sup>.

Для ефективного використання баз даних, Інтернет ресурсів і взагалі обміну інформацією між її споживачами ключове значення також має швидкість телекомунікаційних мереж. Стосовно цього показника ефективності комп'ютерних технологій також існують емпіричні закони, які відомі як закон Батера (Gerry Butcher, Law of Photonics) та закон Нільсена (Jakob Nielsen). Перший з них передбачає, що кількість даних, яку можна передати через одне оптичне волокно, подвоюється кожні 9 місяців, а другий свідчить про те, що швидкість домашніх комунікаційних мереж зростає на 50% кожного року і подвоюється кожні 21 місяців. Це значно менше, ніж відповідно до закону Мура, і тому Нільсен вважав, що швидкість передачі даних стане для користувачів комп'ютерами одним з важливих обмежувальних факторів.

---

<sup>15</sup> Goldman et al. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494, 77–80.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3672958/>

Для використання комп'ютерних технологій в галузі аналізу зображень, про що буде мова у наступних розділах, також важливим є удосконалення комп'ютерної цифрової графіки. Ціна моніторів у розрахунку на 1 піксель зображення також постійно зменшується. Важливим параметром є щільність пікселів, яка позначається PPI (Pixels Per Inch, тобто кількість пікселів на 1 дюйм) – чим цей параметр вищий, тим більш деталізоване зображення можна побачити на моніторі. Цей параметр розраховується як відношення кількості пікселів по діагоналі монітора до довжини діагоналі. Наприклад, для стандарту Full HD (монітор 24 дюймів з роздільною здатністю 1920 x 1080 пікселів) можна розрахувати, що PPI=95,78.

Вже відмічалось, що свого часу, ще на початку розвитку комп'ютерних технологій, одним з потужних імпульсів разом з космічними дослідженнями стала біологічна галузь, яка стосувалась складних розрахунків для інтерпретації рентген-структурних даних будови білків та інших біологічних макромолекул. В наші дні біологія продовжує бути одним з лідерів серед наукових дисциплін, в яких генеруються величезні обсяги даних (так звана Big Data Biology). Особливо помітним це стало після запровадження платформ NGS для секвенування геномів починаючи приблизно з 2005 р. Лінкольн Стейн в роботі, яка була опублікована ще у 2010 р.<sup>16</sup>, проаналізувавши ці тренди дійшов висновку, що вартість секвенування ДНК зменшувалась у 2 рази кожні 19 місяців до 2004 р., а із широким застосуванням платформ NGS з 2005 р. ця вартість почала зменшуватись у 2 рази кожні 5 місяців. Якщо порівняти з вищезазначеними законами, які існують у розвитку комп'ютерної техніки і телекомунікаційних мереж, то ми бачимо експоненційне і значно випереджаюче

---

<sup>16</sup> Stein L.D. The case for cloud computing in genome informatics. *Genome Biology* 2010, 11:207. <https://genomebiology.biomedcentral.com/track/pdf/10.1186/gb-2010-11-5-207.pdf>

зростання кількості даних лише в одній біологічній галузі – геноміці. Автор статті пропонує для вирішення цієї дилеми більш широке використання хмарних технологій («cloud computing» - computation-as-a-service), хоча і застерігає, що для цього потрібно буде вирішити багато проблем.

На веб-сторінках AAAS (American Association for the Advancement of Science) теж можна побачити графіки, які ілюструють цю дещо тривожну ситуацію, коли «data volume increases faster than computational power»<sup>17</sup>. Це особливо стосується таких біологічних галузей, як нейроіміджінг та геноміка, у яких починаючи приблизно з 2010 р. генерація масивів даних значно стала перевищувати зростання комп'ютерних можливостей відповідно до закону Мура. На допомогу приходять насамперед суперкомп'ютери, і деякі приклади їх застосування у біологічних дослідженнях ми розглянемо у наступних розділах. З посиланням на сайт Top500<sup>18</sup> проілюструємо значний прогрес у цій галузі за останні 5 років (Табл. 1.1.4).

**Таблиця 1.1.4.** Характеристики Top-10 суперкомп'ютерів у 2017 р. та 2021 р.

№	Країна	Система	Ядра	Rmax (TFlop/s)	Потужність (кВт)
<b>2017 р.</b>					
1	США	Summit	2 282 544	122 300	8 806
2	Китай	Sunway TaihuLight	10 649 600	93 015	15 371
3	США	Sierra	1 572 480	71 610	-

<sup>17</sup> AAAS - The Big Data Blog, Part V: Interview with Dr. Ivo Dinov.

<https://www.aaas.org/news/big-data-blog-part-v-interview-dr-ivo-dinov>

<sup>18</sup> Top500. <https://www.top500.org/>

4	Китай	Tianhe-2A	4 981 760	61 444,5	18 482
5	Японія	AI Bridging Cloud Infrastructure (ABCI)	391 680	19 880	1 649
6	Швейцарія	Piz Daint	361 760	19 590	2 272
7	США	Titan	560 640	17 590	8 209
8	США	Sequoia	1 572 864	17 173	7 890
9	США	Trinity	979 968	14 137	3 844
10	США	Cori	622 336	14 015	3 939
<b>Червень 2021 р.</b>					
1	Японія	Fugaku, A64FX 48C	7 630 848	442 010	29 899
2	США	Summit	2 414 592	148 600	10 096
3	США	Sierra	1 572 480	94 640	7 438
4	Китай	Sunway TaihuLight	10 649 600	93 015	15 371
5	США	Perlmutter	706 304	64 590	2 528
6	США	Selene	555 520	63 460	2 646
7	Китай	Tianhe-2A	4 981 760	61 444,5	18 482
8	Німеччина	JUWELS	449 280	44 120	1 764
9	Італія	HPC5	669 760	35 450	2 252
10	США	Frontera	448 448	23 516,4	38 745,9

Крім того, варто відмітити значні обчислювальні можливості з використанням ґрід-технології, що працює подібно до Інтернету, але з використанням інших інструментів, таких як системне програмне забезпечення ґрід (Middleware) – це проміжне програмне забезпечення для організації спільної роботи багатьох окремих обчислювальних ресурсів. Тобто ґрід-технології дозволяють не тільки передавати інформацію, як це відбувається в Інтернеті, а і інтегрувати обчислювальні можливості, програмне забезпечення і

бази даних. Перші грід-сервери в Україні з'явилися ще у 2004 р., в Інституті теоретичної фізики ім. М.М. Боголюбова НАН України і Київському національному університеті імені Тараса Шевченка, а з часом було створено Український національний грід (УНГ)<sup>19</sup>. Інститут теоретичної фізики ім. М.М. Боголюбова НАН України залишається основною базою для розвитку проєкту УНГ. На сайті УНГ зазначається, що основними його ресурсами є наступні:

- комп'ютерні кластери, робочі станції, сервери і персональні комп'ютери, що підключені до національної грід-інфраструктури;
- комунікаційні мережі, що об'єднують ці комп'ютери;
- системи збереження даних, підключені до національної грід-інфраструктури;
- інші активні компоненти і обладнання мережі, підключене до національної грід-інфраструктури;
- служби та сервіси, бібліотеки програм, прикладне і системне програмне забезпечення, бази даних і документів, інформаційні сховища і файлові архіви, встановлені на комп'ютерах і мережевому обладнанні національної грід-інфраструктури.

УНГ включає понад 40 кластерів, які існують як в академічних інститутах, так і у закладах вищої освіти України. Більше того, УНГ приєднався до низки міжнародних грід-проєктів та організацій, у тому числі це WLCG (Worldwide LHC Computing Grid, ЦЕРН), EGEE (Enabling Grids for E-sciencE), EGI (European Grid Initiative). Для користувачів грід є дуже значним ресурсом для складних обчислень та зберігання великих масивів даних.

Окремо зупинимось також на деяких важливих питаннях, які стосуються методів захисту та зберігання інформації, тобто антивірусних програмах та

---

<sup>19</sup> Український національний грід. <http://ung.in.ua/ua/>

резервних копіях файлів. На жаль, ані hardware, ані software не дивлячись на такий значний технічний прогрес у галузі комп'ютерних технологій, не є абсолютно надійними. Багатьох проблем можна було б уникнути, якби користувачі цих технологій знали і дотримувались декількох простих правил. Це включає насамперед регулярне оновлення програмного забезпечення (software update) та використання антивірусних програм. Операційні системи Windows 10 і 11 автоматично надають ці можливості і містять вбудований антивірусний захист – Microsoft Defender Antivirus як інтегральна складова Windows Security<sup>20</sup>. Windows Security постійно, у реальному часі перевіряє комп'ютер з моменту його увімкнення на предмет появи malware (malicious software), комп'ютерних вірусів (програми, що можуть відтворюватись прихованим чином, знищувати чи ушкоджувати дані, надавати до даних незгодуваний доступ, уповільнювати або навіть унеможливити роботу операційної системи ПК) та інших загроз безпеці ПК. Цей захист систематично оновлюється в автоматичному режимі, якщо комп'ютер приєднаний до мережі Інтернету (Рис. 1.1.3).



**Рис. 1.1.3.** Можливості Windows Security в операційній системі Windows 10.

<sup>20</sup> Stay protected with Windows Security. <https://support.microsoft.com/en-us/windows/stay-protected-with-windows-security-2ae0363d-0ada-c064-8b56-6a39afb6a963>

Є також і спеціалізовані окремі антивірусні програми, наприклад McAfee AntiVirus, Norton AntiVirus, Bitdefender Antivirus, Webroot SecureAnywhere AntiVirus, Malwarebytes, ESET NOD32 Antivirus, Cylance Smart Antivirus та інші. В разі їх інсталяції Microsoft Defender Antivirus автоматично деактивується. Microsoft надає наступні рекомендації щодо ефективного захисту комп'ютеру:

- окремий файл або папку можна швидко перевірити, якщо на них клацнути правою кнопкою миші та вибрати опцію «Scan with Microsoft Defender»;
- якщо виникає підозра, що в комп'ютері з'явився вірус або malware, то набагато швидше буде це перевірити, якщо виконати швидке сканування, тому що повне сканування може зайняти багато часу;
- якщо загрози не виявлені, варто провести розширену (advanced) перевірку – це може бути повна перевірка (включає всі файли і ті програми, які зараз виконуються на комп'ютері), скан окремих файлів чи папок, а також Microsoft Defender Offline scan. Останній потребує перезавантаження комп'ютера і виконується автоматично до завантаження операційної системи.

Для кращого захисту даних також рекомендовано створювати та регулярно оновлювати резервні (backup) копії важливих файлів, або навіть усього вмісту жорсткого диску. Для спрощення та прискорення виконання цієї задачі є спеціальне програмне забезпечення, таке як Nova Backup, PC Backup, PC Backup & Recovery, PoerBackup та інше. Останнім часом все більшої популярності набуває «хмарне» зберігання (Рис. 1.1.4).



**Рис. 1.1.4.** Ілюстрація сучасних можливостей для створення резервних копій даних (авторська вибірка з веб-ресурсів у публічному домені).

Варто підкреслити, що багато наукових журналів рекомендують зберігати результати експериментів принаймні 5 років після їх публікації. Для отримання наукових грантів також є жорсткі критерії щодо архівації та надання даних іншим вченим (data management policies). Наприклад, BBSRC (Великобританія) пропонує вченим розробити план менеджменту даних як окрему складову частину наукового проєкту<sup>21</sup>. Такий план є обов'язковим додатком до гранту, який проходить рецензування і містить інформацію щодо типу даних, які будуть отримані під час виконання проєкту (реєстрації, експериментальні моделі, зображення тощо), стандартів даних і метаданих, відповідності іншим даним, яким саме чином ці дані будуть доступними для інших дослідників, формат остаточних наборів даних. При цьому потрібно пояснити як саме дані будуть зберігатись (включно з наявними технічними можливостями), у тому числі із застосуванням резервних архівів, і як буде гарантована безпека даних.

---

<sup>21</sup> BBSRC: Data management plan. <https://www.ukri.org/councils/bbsrc/guidance-for-applicants/what-to-include-in-your-application/data-management-plan/>

## ***1.2. Робота з числовою та текстовою інформацією за допомогою Microsoft Office, Libre Office та інструментів Google. Пошукова система PubMed та програма Mendeley.***

### **Обробка числової і текстової інформації за допомогою програми Excel.**

MS Excel - це інтегральна частина Microsoft Office. І якщо основним призначенням програми є зберігання даних в електронних таблицях, то у нашій дисципліні нас більше цікавлять можливості MS Excel для аналізу даних, розрахунків та графічного представлення отриманих результатів. Студенти навчаються працювати з електронними листами Excel, проводити статистичний аналіз результатів, робити інші розрахунки за допомогою електронних таблиць, а потім візуалізувати як оригінальні дані, так і результати їх аналізу. Для виконання таких більш складних задач рекомендовано додати до MS Excel пакет аналізу (Solver)<sup>22</sup> за допомогою меню «Файл > Параметри > Надбудови» і додати «Пакет аналізу», «Пакет аналізу VBA» та «Пошук розв'язання». Після успішного додавання цього пакету у меню «Дані» додаються опції для «Data analysis» та «Розв'язувач» (Рис. 1.2.1; всі ілюстрації зроблені з використанням ліцензованого MS Office 2016).



**Рис. 1.2.1.** Пакет аналізу даних MS Excel.

<sup>22</sup> Solver Add-in in Excel. <https://support.microsoft.com/en-us/office/load-the-solver-add-in-in-excel-612926fc-d53b-46b4-872c-e24772f078ca>

Практика викладання дисципліни показує, що студенти добре обізнані з роботою в MS Excel, а також аналогічною програмою LibreOffice – Calc (spreadsheets)<sup>23</sup>. LibreOffice сумісний із форматами документів MS Office, таких як MS Word (.doc, .docx), Excel (.xls, .xlsx) та PowerPoint (.ppt, .pptx). Програма MS Excel на заголовній сторінці містить посібник з формул, посібник зі зведеної таблиці та посібник із секторних діаграм (Рис. 1.2.2). Наприклад, посібник з формул – це інтерактивний ресурс, що зліва містить детальні інструкції, а справа користувач може власноруч їх виконувати у запропонованій послідовності (Рис. 1.2.3).



**Рис. 1.2.2.** Посібники MS Excel.

---

<sup>23</sup> LibreOffice <https://www.libreoffice.org/discover/libreoffice/>

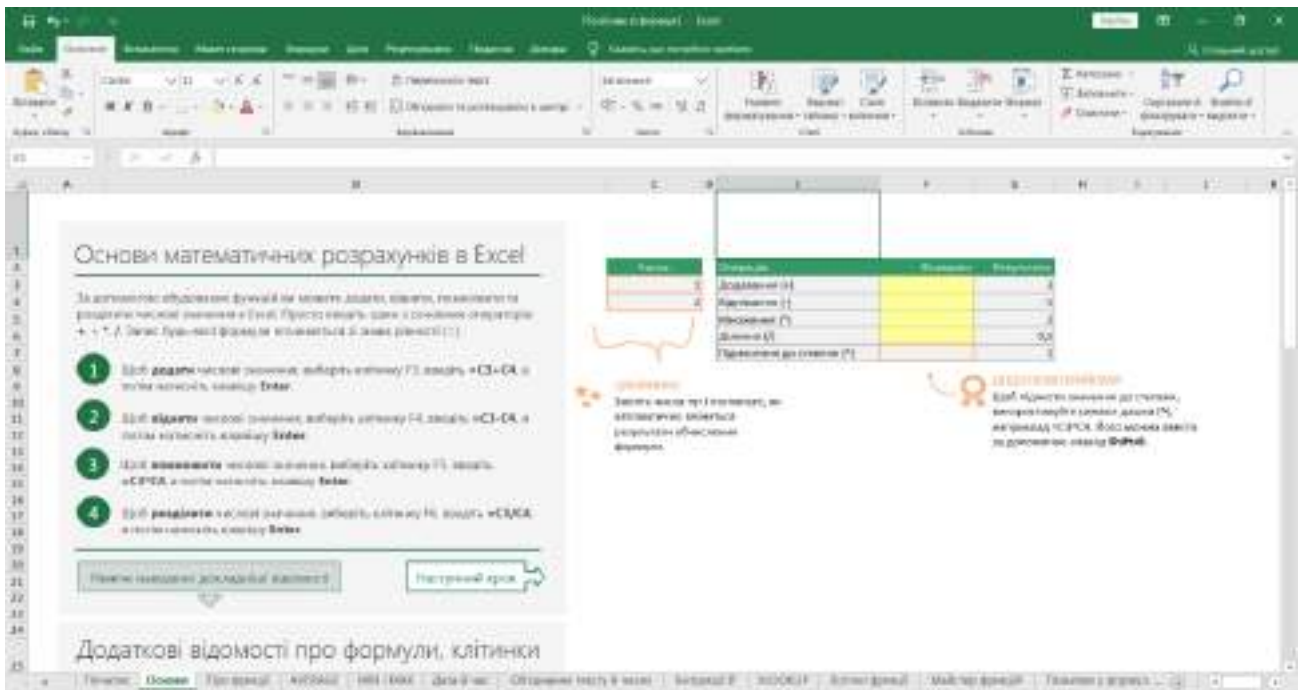


Рис. 1.2.3. Ілюстрація посібника з формул.

У кінці сторінки MS Excel пропонує скористатись ресурсами Інтернету для отримання більш детальної інформації (Рис. 1.2.4). Зокрема це і безкоштовні онлайн відеокурси<sup>24</sup>.

Дізнатись більше про можливості цієї програми та знайти відповіді на свої запитання студенти можуть також за допомогою «Допомога та навчання з Excel» від Microsoft<sup>25</sup>.

<sup>24</sup> [Відео курси](#) з Excel.

<sup>25</sup> Допомога та навчання з Excel <https://support.microsoft.com/uk-ua/excel>

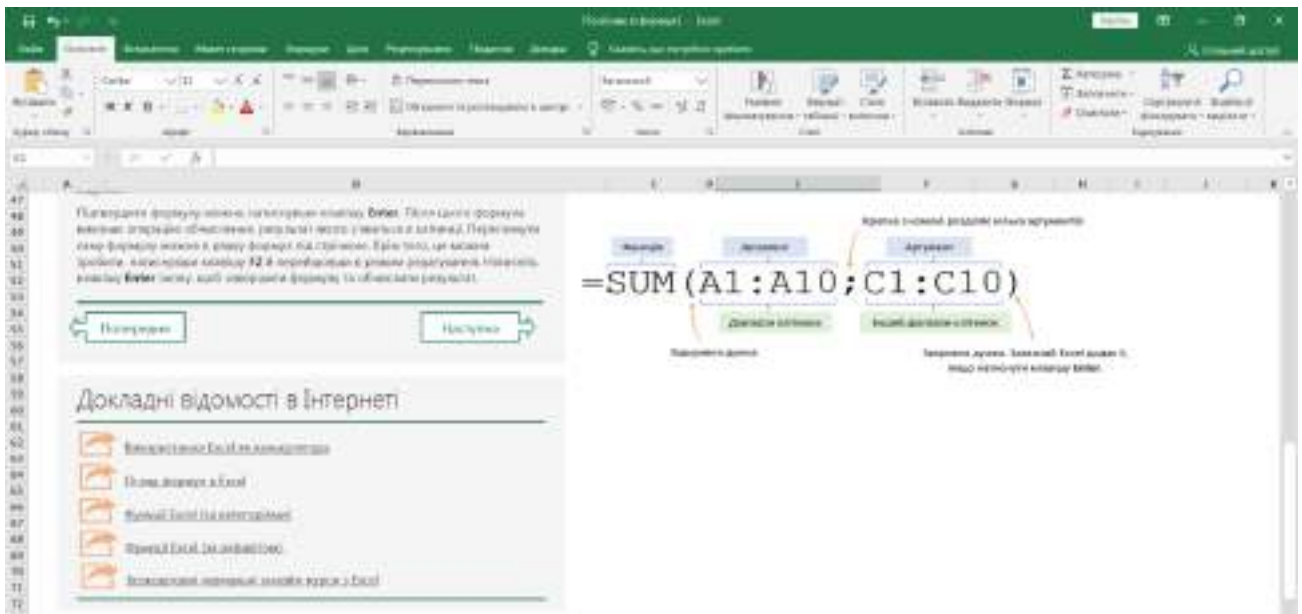


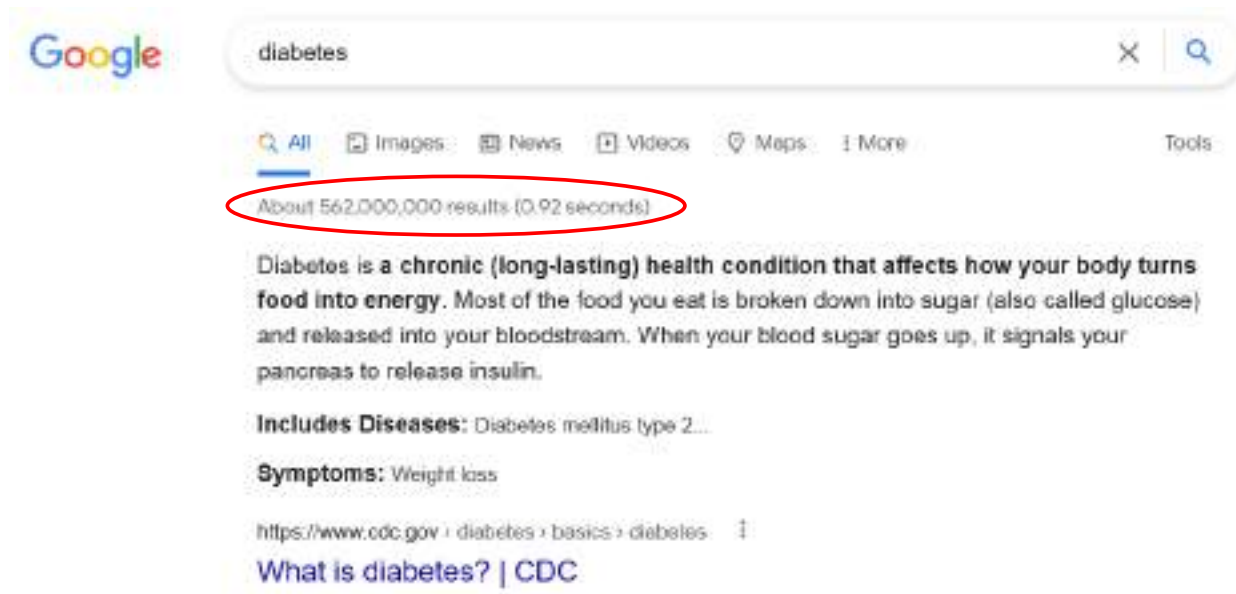
Рис. 1.2.4. Докладніше про роботу з формулами в MS Excel.

## Інструменти Google

Для коректного планування любого наукового дослідження, будь то дипломна студентська робота, дисертаційна робота чи науковий проєкт, першим і найбільш важливим етапом є ознайомлення із сучасним станом проблеми, що дозволяє сформулювати загальну мету і цілі дослідження. У самому процесі дослідження часто виникає потреба скористатись доступними базами даних, розробленими і запрограмованими алгоритмами їхнього аналізу. На завершальному етапі під час обговорення результатів знову виникає потреба порівняти отримані результати з відомими даними, і таким чином визначити і довести їх новизну, визначити подальші перспективи роботи у цьому науковому напрямку, зробити висновки. Відповідно, найбільша кількість цитувань першоджерел як правило припадає на такі розділи роботи як «Вступ» та «Обговорення результатів».

Стандартний пошук через Google продукує від сотень тисяч до сотень мільйонів результатів. Наприклад, якщо ввести у пошуковий рядок Google

термін “diabetes”, то менш ніж за секунду отримаємо близько 560 млн. результатів (Рис. 1.2.5).



**Рис. 1.2.5.** Ілюстрація результатів пошуку інформації стосовно діабету за допомогою Google.

Пошуковий алгоритм Google<sup>26</sup> - PageRank алгоритм - є надзвичайно ефективним, адже ми з його допомогою майже миттєво отримали велику кількість посилань на веб-ресурси стосовно діабету. Якщо не заглиблюватись у деталі, то процес пошуку включає 3 основні стадії:

- повзання (crawling): відбувається скан Інтернету на предмет нових або оновлених веб-сторінок, при цьому чим більше було звернень на конкретний ресурс, тим легше Google його знаходить;
- індексація: аналізується контент кожного ресурсу, і після цього він заноситься у базу даних Google index;

---

<sup>26</sup> Google Search - How Search algorithms work.

<https://www.google.com/search/howsearchworks/algorithms/>

- обслуговування (serving): на цьому етапі визначається, які з цих веб-сторінок є найбільш доречними для конкретного пошукового запиту (відбувається їх ранжирування – ranking). Враховується велика кількість факторів – від якості веб-ресурсу і до місця знаходження користувача. «Ваговий коефіцієнт» також може бути різним – наприклад для пошуку стосовно новин найбільшу вагу буде мати така характеристика інформації як її актуальність, своєчасність (Табл. 1.1.1), а для пошуку в онлайн словниках чи енциклопедіях ця характеристика буде менш вагомою.

З наведеного прикладу ми бачимо, що першим результатом пошуку є CDC – це авторитетний Centers for Disease Control and Prevention, державний ресурс США<sup>27</sup>. Але у результати пошуку потраплять також і статті з Вікіпедії, дисертації, книги, матеріали з сайтів професійних організацій, навчальних закладів, новини, блоги, препринти та багато іншого.

Нас зараз цікавить питання – як у цьому океані інформації знайти найбільш доречну для мого конкретного наукового питання? Для студентів і науковців цінним інструментом у цій ситуації може стати інший інструмент - Google Академія (Google Scholar)<sup>28</sup> (Рис. 1.2.6).

Зверніть увагу, що вверху зліва Google Scholar надає можливості для створення власного профілю вченого, в який можна імпортувати свої публікації та побачити їх наукометричні показники (кількість цитувань тощо), а також створити власну бібліотеку онлайн літературних джерел.

---

<sup>27</sup> Centers for Disease Control and Prevention. <https://www.cdc.gov/>

<sup>28</sup> Google Scholar. <http://scholar.google.com>

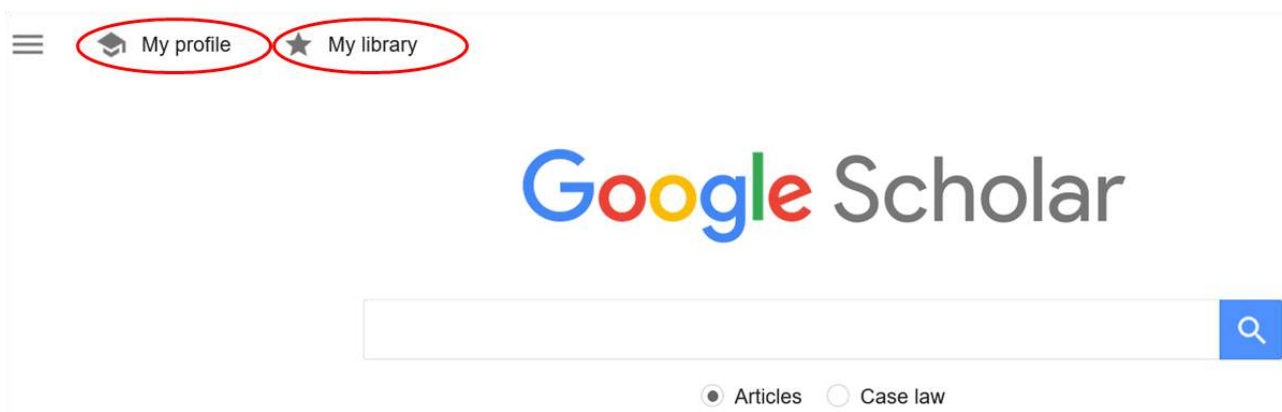


Рис. 1.2.6. Заголовна сторінка Google Академія.

Приклад My Library наведено на Рис. 1.2.7. В ній є можливості, хоча і досить обмежені у порівнянні з іншими ресурсами, які ми будемо розглядати далі, для сортування і фільтрування джерел (це опції зліва сторінки), проведення пошуку за ключовими словами (у цьому прикладі TRPC4 – рецептор-керований катіонний канал), а також показані варіанти можливості для отримання доступу до тексту публікації (у тому числі безкоштовного – Free from Publisher тощо), якщо такі існують.

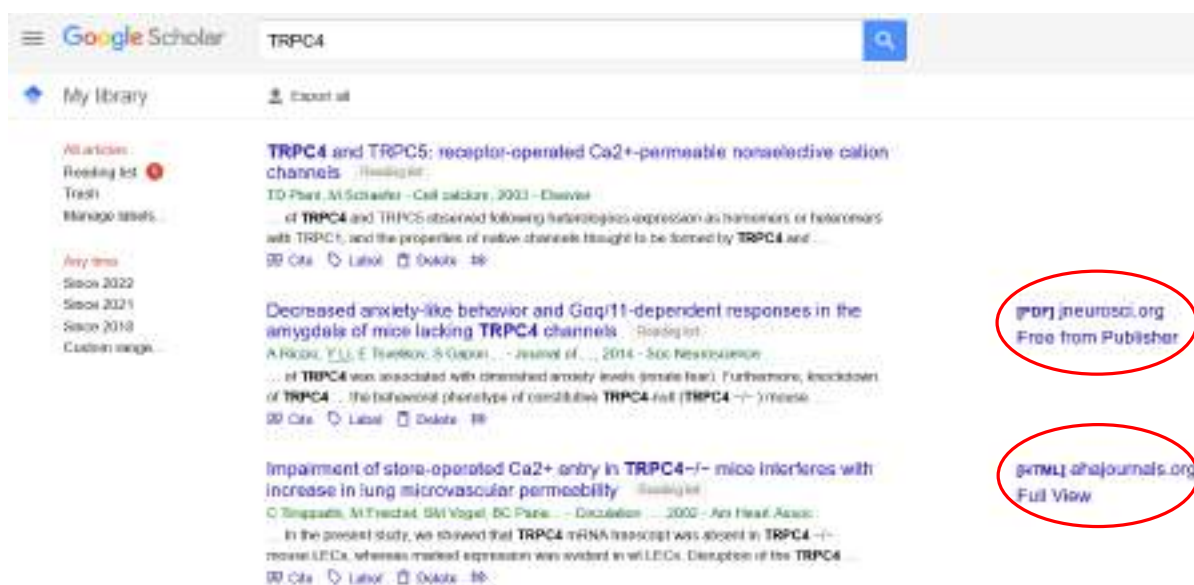


Рис. 1.2.7. Власна бібліотека Google Академія.

Якщо ми тепер проведемо пошук публікацій щодо діабету, але за допомогою ресурсу Google Академія, а не просто Google, то отримаємо значно менше результатів – саме академічного (наукового чи освітнього) контенту, як показано на Рис. 1.2.8 (порівняйте з Рис. 1.2.5).

Для кожного джерела наводиться важлива інформація та додаткові опції – можна додати публікацію до власної бібліотеки (Save), процитувати цю публікацію (після натискання на Cite з’являється вікно, яке містить різні формати цитування), джерела, які цитують дану публікацію (Cited by) та подібні публікації (Related articles). Останні дві допомагають продовжити і розширити пошук літературних джерел по даній темі. Неважко уявити, що якщо користувачеві вдалось знайти статтю, яка найкраще відповідає конкретному запиту (пригадайте таку важливу характеристику інформації як її цінність для користувача – Табл. 1.1.1), то далі варто буде переглянути як подібні публікації (Related articles), так і більш недавні публікації, які цитують цю роботу (Cited by).

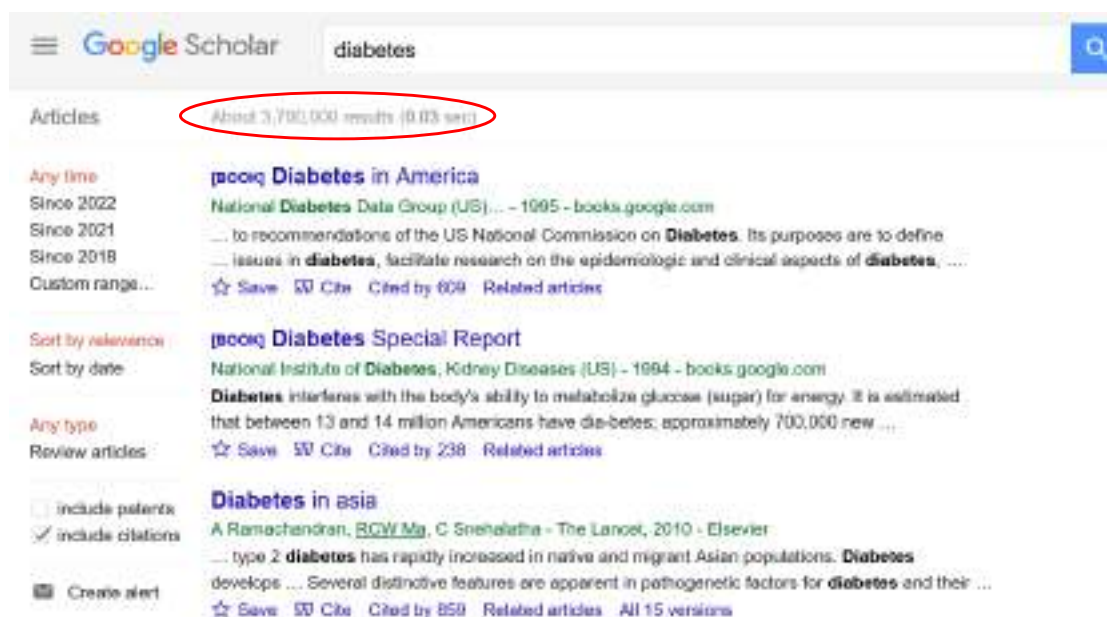


Рис. 1.2.8. Ілюстрація результатів пошуку інформації стосовно діабету за допомогою Google Академія.

За своїм наповненням Google Академія є найбільшим онлайн ресурсом, який охоплює понад 160 млн. джерел. Для Google Академія є посібник користувача, який містить багато цінних порад, зокрема для ефективного пошуку інформаційних джерел академічного контенту<sup>29</sup>. Наведемо з посиланням на цей ресурс основні переваги Google Академія:

- шукайте всю наукову літературу в одному зручному місці;
- досліджуйте пов'язані роботи, цитати, авторів та публікації;
- знайдіть повний документ у своїй бібліотеці або в Інтернеті;
- будьте в курсі останніх подій у будь-якій галузі досліджень;
- перевірте, хто цитує ваші публікації, створіть загальнодоступний профіль автора.

Наведемо також основні поради від Google Академія стосовно організації пошуку літератури<sup>25</sup>:

*Як шукати за автором?*

Використовуйте оператор "author:", наприклад, author:"a sanders" або author:"alexander v sanders".

*Як шукати за назвою?*

Візьміть назву статті у лапки: «Історія Китайського моря».

*Як сортувати за датою?*

Ви будете отримуватимете кращі результати, якщо шукатимете лише останні статті, але сортуватиме їх за релевантністю, а не за датою. Наприклад, натисніть «З 2012 року» на лівій бічній панелі сторінки результатів пошуку.

Щоб спочатку переглянути абсолютно нові статті, натисніть «Сортувати за датою» на бічній панелі. Якщо ви часто використовуєте цю функцію, вам

---

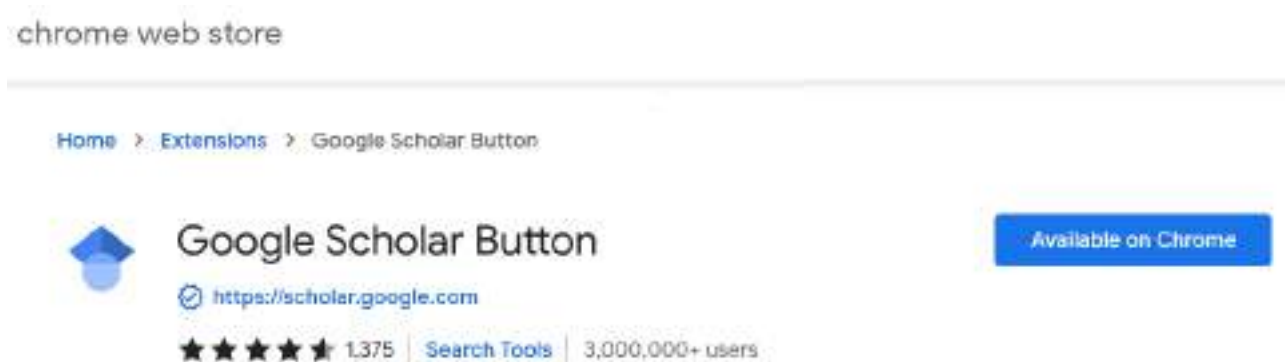
<sup>29</sup> Google Академія - посібник користувача.

<https://scholar.google.com.ua/intl/uk/scholar/help.html>

також може бути корисно налаштувати автоматичне сповіщення електронною поштою.

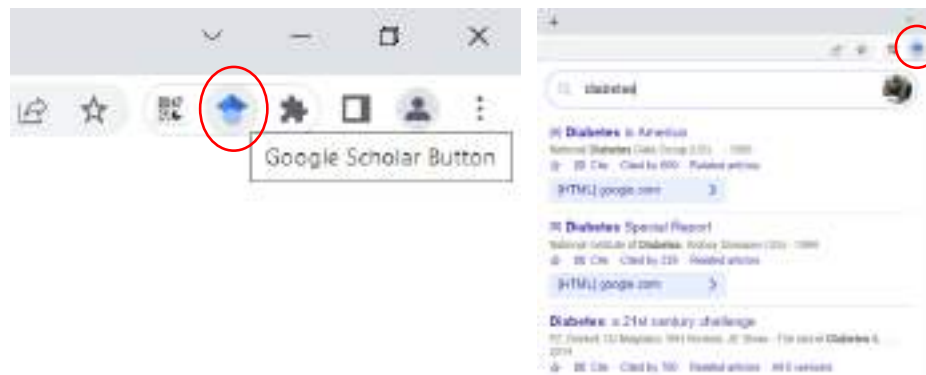
Ще більше пошукових можливостей надає розширений пошук (Advanced search). Для його застосування потрібно клацнути на значок з 3 горизонтальними рисками у верхньому лівому кутку сторінки (Рис. 1.2.8).

Для зручності роботи з Google Академія у браузері Google Chrome є додаток, який дозволяє швидко перейти в цю пошукову систему з будь-якої веб-сторінки. Його можна інстальовати за допомогою Веб-магазину Chrome як показано на Рис. 1.2.9.



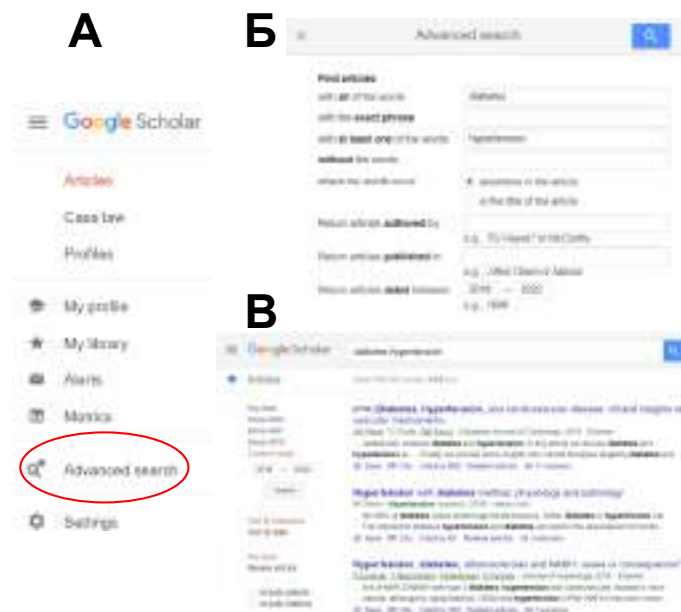
**Рис. 1.2.9.** Інсталяція кнопки Google Академія.

Натисканням на цю кнопку можна миттєво активувати Google Академія, зробити пошук за ключовим словом і отримати 3 найбільш релевантні результати (Рис. 1.2.10).



**Рис. 1.2.10.** Використання кнопки Google Академія для швидкого пошуку з будь-якої сторінки Google Chrome.

Далі потрібно задати параметри пошуку. На Рис. 1.2.11Б пошук літератури стосується одного з поширених ускладнень діабету – гіпертензії, публікації протягом останніх 5 років. В результаті пошуку отримуємо вже не 3,7 млн. (Рис. 1.2.8) , а 362 тис. джерел (Рис. 1.2.11В).



**Рис. 1.2.11.** Ілюстрація результатів пошуку інформації стосовно діабету та одного з його поширених ускладнень, гіпертензії, за допомогою розширеного пошуку в Google Академія. А ілюструє як саме обрати Advanced search, Б ілюструє параметри розширеного пошуку, В ілюструє результати пошуку.

## PubMed

Google Академія, як вже зазначалось, є найбільшим у світі онлайн ресурсом, який охоплює понад 160 млн. джерел. Але це створює і певні проблеми у роботі з ним. Згадаємо таку надзвичайну властивість інформації як її надійність, достовірність, безпомилковість (Табл. 1.1.1). У галузі біомедицини існує такий надійний ресурс, функціонування якого забезпечують такі організації як Національні Інститути Здоров'я США, Національна медична бібліотека і Національний центр біотехнологічної інформації США – PubMed<sup>30</sup>. Він зараз містить понад 33 млн. літературних джерел біомедичного профілю з баз даних MEDLINE, наукових журналів та книг (Рис. 1.2.12).



Рис. 1.2.12. Заголовна сторінка PubMed.

Перевагами цього ресурсу порівняно з Google Академія є те, що, по-перше, інформація, яка потрапляє в цю базу даних проходить ретельний фаховий відбір (так, далеко не всі журнали чи книги в ній індексуються), а по-друге, PubMed пропонує цілий набір інструментів для ефективної роботи з

---

<sup>30</sup> PubMed. <https://pubmed.ncbi.nlm.nih.gov/>

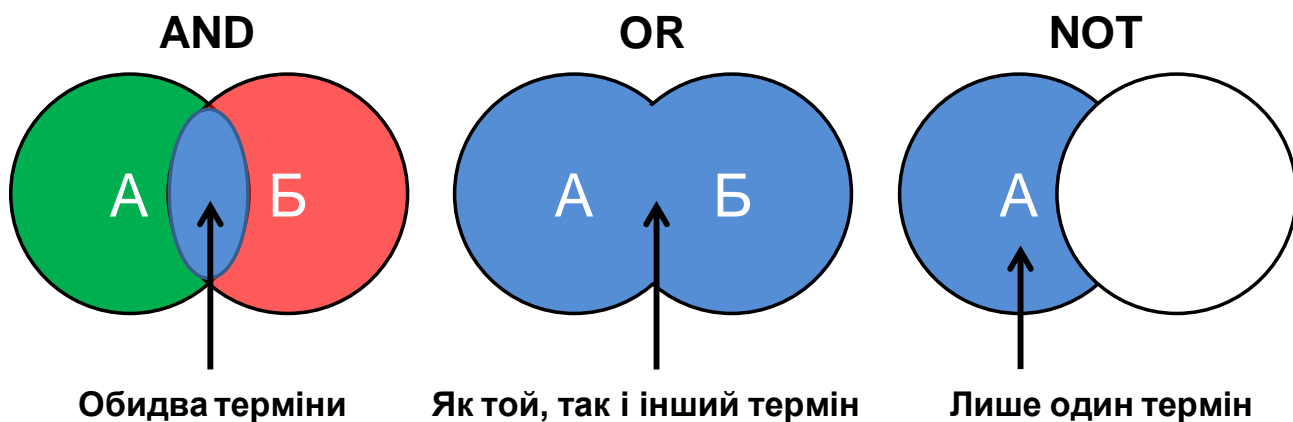
літературою, серед яких є «власний кабінет» (My NCBI), який дозволяє зберігати результати пошуку і потім продовжувати пошук з того місця, на якому ви закрили сторінку веб-браузера.

Познайомимось з основними функціями і можливостями PubMed. Для ефективної роботи в PubMed нам насамперед потрібно пригадати, що таке логічні оператори і як вони працюють. Ці оператори також називаються булевими. Наведемо таблицю логічних операторів, які дозволяють об'єднувати результати пошуку за певними правилами (Табл. 1.2.1). Важливо – ці оператори потрібно записувати великими літерами - AND, OR, NOT.

**Табл. 1.2.1.** Булеві оператори, які використовуються для пошуку за ключовими словами у PubMed.

Логічний оператор	Результат застосування
AND (також символ &)	Використовується між концепціями. Результати включають всі терміни.
OR	Результати включають принаймні один з термінів.
NOT	Результати виключають термін(и) після NOT.

Пошук проводиться зліва направо, а для більшої деталізації порядку їх виконання, подібно до алгебри, застосовуються дужки. Наприклад, *(diabetes AND hypertension) NOT (stroke OR heart)* означає, що спочатку буде виконано пошук *(diabetes AND hypertension)*, потім - *(stroke OR heart)*, і в останню чергу ці два набори результатів будуть об'єднані, щоб виключити з першого набору другий. Графічно результати об'єднання результатів пошуку представлені на Рис. 1.2.13.



**Рис. 1.2.13.** Графічне представлення результатів об'єднання двох наборів даних, А і Б, з використанням логічних операторів AND, OR, NOT.

При застосуванні логічних операторів важливо використовувати теги (tags) після терміну для конкретизації його значення. Повний список тег можна знайти за посиланням Search Field descriptions and tags<sup>31</sup>. Наведемо лише деякі, які вживаються найбільш часто (Табл. 1.2.2).

**Табл. 1.2.2.** Деякі дескриптори термінів в PubMed.

[au]	автор	[ti]	назва	[ad]	приналежність (організація)
[all]	всі поля	[tiab]	назва/абстракт	[book]	книга
[ta]	журнал	[tw]	слово у тексті	[crdt]	дата публікації

Виконаємо, як і раніше, пошук для терміну «diabetes» (Рис. 1.2.14). Насамперед звернемо увагу на таке: (1) кількість отриманих результатів є значно меншою, ніж в пошуку Google Академія – 868 647; (2) зліва сторінки наведено графік кількості публікацій по роках, який наочно демонструє

<sup>31</sup> PubMed: Search Field descriptions and tags.

<https://pubmed.ncbi.nlm.nih.gov/help/#search-tags>

динаміку обсягу досліджень по даній темі; (3) справа вверху є можливість задати параметри відображення результатів, зокрема формату їх представлення (Summary, Abstract, PubMed, PMID). PMID – це унікальний номер, який є ідентифікатором публікації в системі PubMed.

The image shows the PubMed search results page for the query 'diabetes'. At the top, the NIH logo and 'National Library of Medicine' are visible. The search bar contains 'diabetes' and the search button is labeled 'Search'. Below the search bar, there are options for 'Advanced', 'Create alert', and 'Create RSS'. A red box highlights the 'Sorted by: Best match' and 'Display options' buttons. The results section shows '668,647 results' and 'Page 1 of 66,865'. A 'RESULTS BY YEAR' chart is highlighted with a red box, showing a sharp increase in results starting around 2015, peaking in 2022. Two search results are visible: 'Diabetes insipidus' by Tizard JP (1948) and 'Gestational diabetes insipidus: a review of an underdiagnosed condition' by Aleksandrov N et al. (2010). A 'TEXT AVAILABILITY' section allows filtering by Abstract, Free full text, or Full text.

Рис. 1.2.14. Результати пошуку публікацій у PubMed стосовно діабету.

Далі є можливості сортування результатів, а саме – найбільш релевантні публікації, найбільш недавні статті, по даті їх публікації, по прізвищу першого автора, по назві журналу. Також можна змінити кількість посилань, які будуть відображатись на одній сторінці. Останнє важливе, якщо відмічати джерела галочками у квадратах біля них і потім перейти на наступну сторінку, то ці відмітки втрачаються, отже варто вибирати максимальну кількість – це 200 на сторінку.

Уявимо, що ми вперше знайомимось з цим напрямком досліджень і скористаємось фільтрами, які є зліва сторінки. Напевно нас будуть цікавити насамперед недавні оглядові статті, які є у вільному доступі. Отже відмічаємо ці опції і повторюємо пошук (Рис. 1.2.15).

The screenshot shows a PubMed search interface. On the left, there are filter sections: 'TEXT AVAILABILITY' with 'Free full text' checked; 'ARTICLE ATTRIBUTE' with 'Associated data' unchecked; 'ARTICLE TYPE' with 'Review' checked; and 'PUBLICATION DATE' with '5 years' selected. The search bar contains 'diabetes' and shows '21,543 results'. Below the search bar, there are buttons for 'Save', 'Email', and 'Send to', and a 'Sorted by: Best match' dropdown. The results list shows two articles with their titles, authors, and publication details.

Рис. 1.2.15. Пошук у PubMed із застосуванням фільтрів.

Нам вдалось звузити коло пошуків до 21 543 результатів, але все одно це досить багато, навіть для простого перегляду назв і абстрактів цих статей. На цьому етапі пошуку варто задатись питанням – а що саме мене цікавить у цій проблематиці? Можливо це гіпертонічна хвороба як одне з типових ускладнень діабету. Скористаємось логічними операторами і у пошуковому рядку введемо diabetes [tiab] AND hypertension [tiab]. Табл. 1.2.2 показує, що тега [tiab] вимагає,

щоб ці терміни були у назві та/або абстракті статті. Ці прості дії дозволили нам звузити пошук приблизно на порядок – ми отримали 2 289 результатів, які тепер можна сортувати по «Best match» та «Most recent» і починати опрацьовувати.

PubMed містить дуже деталізовані інструкції для користувачів, які включають багато практичних порад<sup>32</sup>. Розглянемо як приклад пошук за автором. Ми знаємо, що прізвище вченого Brown, він нейрофізіолог, який досліджує мускаринові рецептори і М-струми і працює в Університетському коледжі Лондону. Кількість результатів нашого пошуку по мірі його деталізації будуть наступними:

Brown – 282 190

Brown [au] – 936.

Якщо відоме ім'я або ініціали автора, то можемо їх додати наступним чином

Brown, David [au] або Brown DA [au] – отримуємо 2259 і 936 результатів, відповідно. Це все рівно надто багато результатів, тому нам потрібно звужувати пошук далі. Додаймо мускаринові рецептори, які досліджує цей вчений:

Brown DA [au] AND "M current" – 50 результатів.

Додаймо організацію:

Brown DA [au] AND "M current" AND "University College London" [ad] – 29 результатів.

Таким чином, хоча прізвище Браун є одним з найпоширеніших прізвищ, нам вдалось знайти публікації саме того автора, який нам був потрібен. Студенти можуть самостійно повторити ці пошукові кроки для набуття практичних навичок. У пошуку авторів можна також використати додаткові теги - [1au] для першого автора і [lastau] для останнього автора у списку авторів

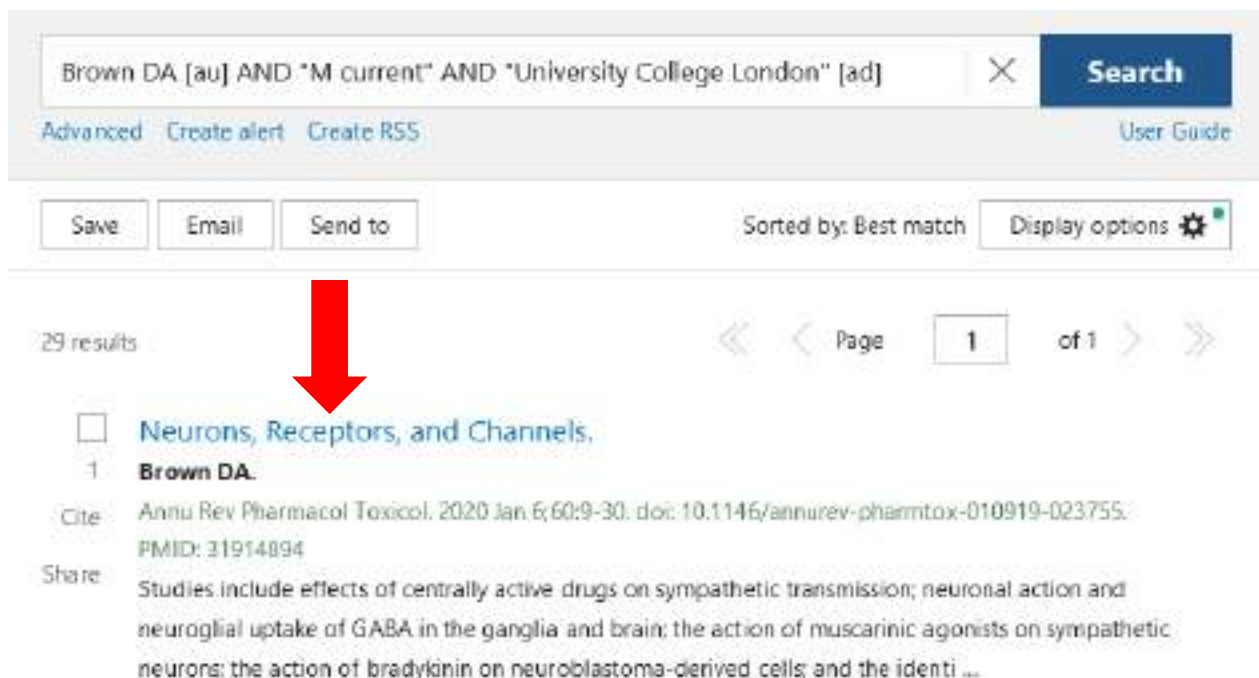
---

<sup>32</sup> PubMed User Guide.

[https://pubmed.ncbi.nlm.nih.gov/help/#pubmedhelp.How\\_do\\_I\\_search\\_PubM](https://pubmed.ncbi.nlm.nih.gov/help/#pubmedhelp.How_do_I_search_PubM)

публікації. Порядку авторів зараз приділяють увагу, адже загальноприйняте, що перший автор виконав більшу частину досліджень або найбільш важливу їх частину, а останній є науковим керівником проєкту.

Але виникає питання – чи всі публікації конкретного автора нам вдалось знайти таким чином? Очевидно, що ні. В деяких з них він міг використати лише перший ініціал (Brown D), в інших - David A. Brown, і звісно не всі публікації автора стосувались досліджень М-струмів. Клацнемо курсором на назву першої статті у списку для того, щоб отримати більше інформації (Рис. 1.2.16).



**Рис. 1.2.16.** Отримання додаткової інформації про статтю.

Відкривається нова сторінка, як показано на Рис. 1.2.17. По-перше, у нас тепер з'являється можливість знайти у PubMed всі публікації цього автора. Для цього потрібно лише клацнути по прізвищу автора (Рис. 1.2.17 - 1), і їх виявляється є 941. По-друге, можна ознайомитись з абстрактм статті (2), а також відкрити повний текст статті (іконка справа вверху, крок 3). Крім того,

можна отримати список подібних, релевантних статей (Рис. 1.2.17 - 4). І нарешті, є можливість отримати доступ до значно більшої кількості баз даних і ресурсів, які не входять у систему NCBI. Цей сервіс називається LinkOut (Рис. 1.2.17 - 5)<sup>33</sup>.

> Annu Rev Pharmacol Toxicol. 2020 Jan 6;60:9-30. doi: 10.1146/annurev-pharmtox-010919-023755.

## Neurons, Receptors, and Channels

David A. Brown <sup>1</sup>

Affiliations — collapse

### Affiliation

<sup>1</sup> Departments of Neuroscience, Physiology, and Pharmacology, University College London, London WC1E 6BT, United Kingdom; email: d.a.brown@ucl.ac.uk.

PMID: 31914894 · DOI: 10.1146/annurev-pharmtox-010919-023755

### Abstract

Here, I recount some adventures that I and my colleagues have had over some 60 years since 1957 studying the effects of drugs and neurotransmitters on neuronal excitability and ion channel function, largely, but not exclusively, using sympathetic neurons as test objects. Studies include effects of centrally active drugs on sympathetic transmission; neuronal action and neuroglial uptake of GABA in the ganglia and brain; the action of muscarinic agonists on sympathetic neurons; the action of bradykinin on neuroblastoma-derived cells; and the identification of M-current as a target for muscarinic action, including experiments to determine its distribution, molecular composition, neurotransmitter sensitivity, and intracellular regulation by phospholipids and their hydrolysis products. Techniques used include electrophysiological recording (extracellular, intracellular microelectrode, whole-cell, and single-channel patch-clamp), autoradiography, messenger RNA and complementary DNA expression, antibody injection, antisense knockdown, and membrane-targeted lipidated peptides. I finish with some recollections about my scientific career, funding, and changes in laboratory life and pharmacology research over the past 60 years.

**Keywords:** GABA; M-current; acetylcholine; autobiography; brain slices; muscarinic receptors; sympathetic ganglion.

FULL TEXT LINKS

ABSTRACT REVIEWED FULL-TEXT ARTICLE

ACTIONS

Cite

Favorites

SHARE

Twitter Facebook LinkedIn

PAGE NAVIGATION

< Title & authors

Abstract

Similar articles

Publication types

MeSH terms

Substances

Grant support

LinkOut - more resources

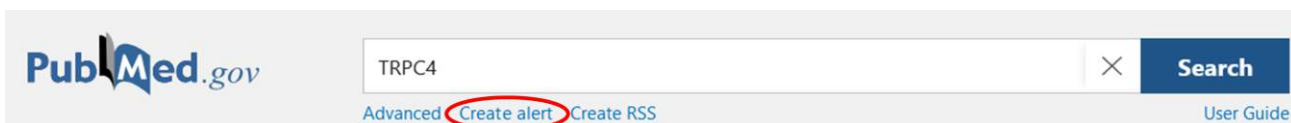
Рис. 1.2.17. Повна інформація про джерело.

<sup>33</sup> NCBI LinkOut. <https://www.ncbi.nlm.nih.gov/projects/linkout/doc/linkout.html>

Для ще більш ефективної, а головне персоналізованої роботи з ресурсами PubMed потрібно створити власний безкоштовний кабінет, який називається My NCBI. Є декілька варіантів – від використання Google акаунту до реєстрації користувача (login).



В My NCBI можна зберігати результати пошуків – для цього просто натисніть на Create alert під пошуковим рядком:



Власний кабінет має багато інших функцій, як показано на Рис. 1.2.18.

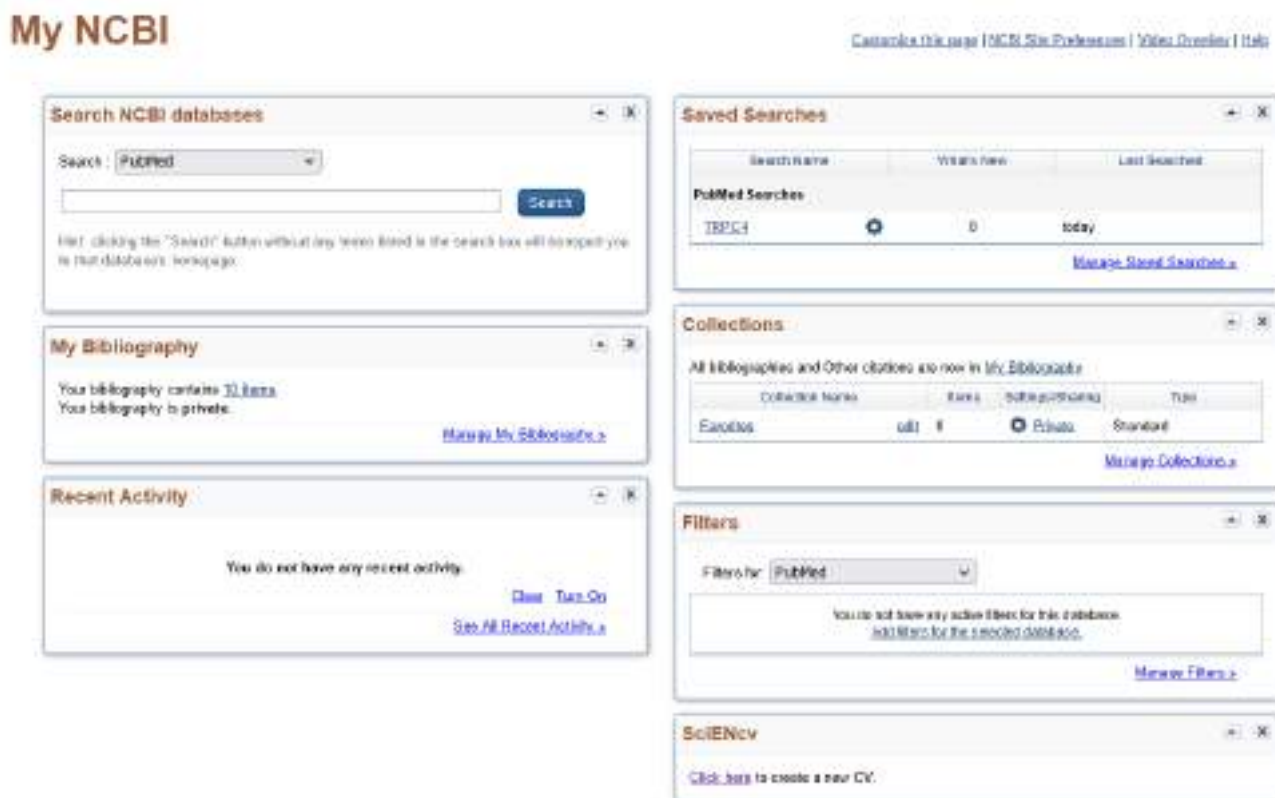


Рис. 1.2.18. Функції My NCBI.

Крім вже відмічених можливостей пошуку джерел (Search NCBI databases) та зберігання результатів пошуку (Saved Searches) є можливість створення власної онлайн картотеки (My Bibliography), колекцій статей (Collections) використовуючи при цьому власні фільтри (Filters), і навіть є можливість створити своє власне професійне резюме (SciENcv)!

Для пошуку джерел крім базових процедур, з більшістю яких ви вже познайомились, є також розширені можливості:



Вони зокрема значно спрощують використання логічних операторів і тег. Для активації розширеного пошуку потрібно натиснути на «Advanced» під пошуковим рядком.

Так, наведений вище приклад для пошуку diabetes [tiab] AND hypertension [tiab] можна реалізувати наступним чином, якщо скористатись веб-сторінкою для розширеного пошуку (Рис. 1.2.19).

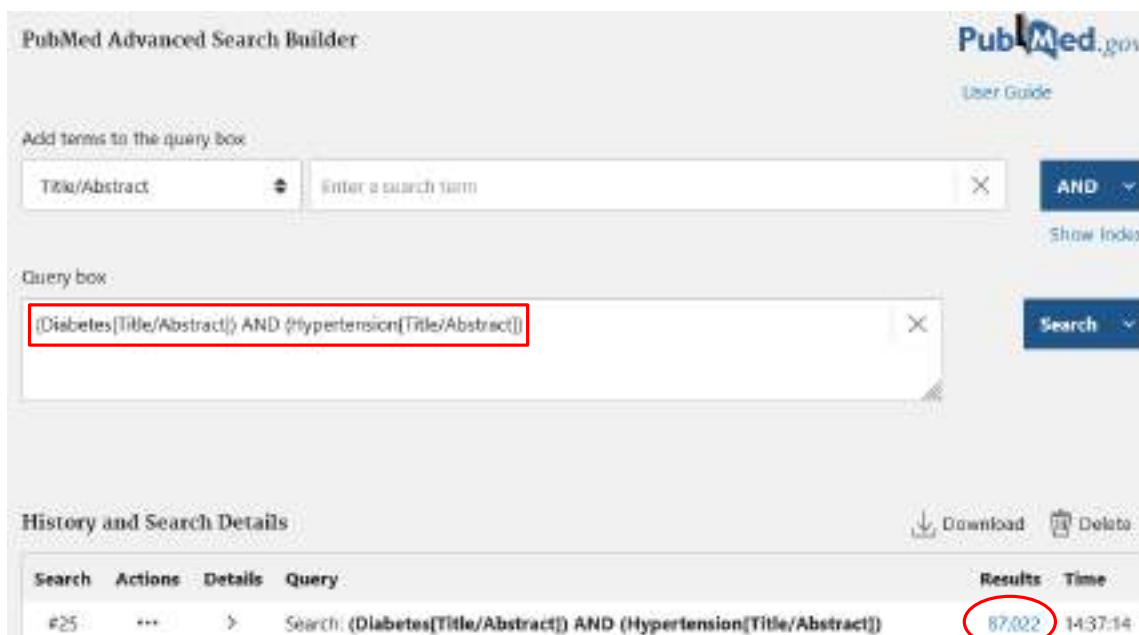


Рис. 1.2.19. Розширений пошук у PubMed.

На сторінці розширеного пошуку, який дуже влучно називається «PubMed Advanced Search Builder», можна додавати нові і нові терміни, обираючи при цьому потрібну тегу та логічний оператор, і одночасно спостерігати за кількістю отриманих результатів, яка показується у реальному часі внизу справа сторінки (Рис. 1.2.19).

Відповіді на питання, які часто виникають у процесі роботи з PubMed, можна знайти у посібнику користувача за посиланням <https://pubmed.ncbi.nlm.nih.gov/help/>. У другій частині він також містить і детальні інструкції щодо алгоритмів пошуку (Search PubMed). Цей ресурс дає можливість обрати інструкції для дослідників, видавництв, бібліотекарів, викладачів, медиків і для тих, хто не є фахівцями (вкладка Resources for you). Наприклад, для дослідників, які працюють у галузі обчислювальної молекулярної біології, пропонуються наступні ресурси NCBI:

- BLAST
- GenBank
- Molecular Structure and Function
- Taxonomy Database
- Conserved Domain Database (CDD)
- Reference Sequences (RefSeq)
- Vector Alignment Search Tool (VAST)

Починати працювати з посібником користувача варто з проходження швидких турів (Quick Tours for Everyone). Це серія коротких відео, які ілюструють наступні теми:

What is in PubMed?

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/whatsin/>

PubMed: Find articles by author.

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/author/>

PubMed: Find articles by journal.

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/journal/>

PubMed: Find articles from a citation.

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/citation/>

PubMed: Find articles on a topic

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/citation/>

PubMed: Find the latest treatments for a disease or disorder

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/treatment/>

PubMed: Get the full text for an article

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/fulltext/index.html>

PubMed: Save searches and set e-mail alerts

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/alerts/index.html>

PubMed subject search: How it works

[https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/topic\\_how\\_it\\_works/index.html](https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/topic_how_it_works/index.html)

PubMed and MedlinePlus: Consumer health

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/consumer/index.html>

Під час перегляду цих відео варто робити паузи і повторювати ті чи інші продемонстровані кроки для кращого засвоєння цих матеріалів.

В розділі швидких турів для експертних пошуків (Quick Tours for Expert Searches) є відео стосовно:

Using PubMed to Find Human-Related Studies

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/humans/index.html>

PubMed: Using the Advanced Search Builder

<https://www.nlm.nih.gov/oet/ed/pubmed/quicktours/advanced/index.html>

Можливості розширеного пошуку ми вже достатньо детально розглянули. Посібник користувача також містить посилання на онлайн-класи, які тривають

значно довше, до 90 хвилин, а деякі з них складаються з окремих модулів, за проходження яких можна отримати CE (continuing education) кредити.

Для кращого ознайомлення з інформаційними ресурсами PubMed та отримання практичних навичок з їх використання студентам пропонується виконати практичну роботу 4.1. «Пошук біологічної інформації в Інтернеті. Робота з пошуковою системою PubMed та менеджером цитувань Mendeley», а також підготувати і провести в інтерактивному з іншими студентами режимі презентацію на одну з пропонованих тем:

1. Вступ до PubMed.
2. Як мені знайти публікації автора?
3. Як мені знайти публікації конкретного журналу?
4. Як мені знайти статтю, якщо у мене є цитата?
5. Як мені знайти наукову статтю по моїй темі, яка є у вільному доступі?
6. Як мені організувати власний кабінет в NCBI?
7. Пошук біомедичної літератури за допомогою MeSH<sup>34</sup>.
8. Як мені зберегти результати мого пошуку і встановити автоматичні сповіщення, які будуть надсилатись на мою електронну пошту?
9. Пошук інформації щодо лікарських препаратів<sup>35</sup>.
10. Як мені створити власну бібліографічну базу статей?

---

<sup>34</sup> Medical Subject Heading (MeSH). <https://nmlm.gov/training/class/medical-subject-heading-mesh-changes-and-pubmed-searching-demand>

<sup>35</sup> Drug information Portal. <https://druginfo.nlm.nih.gov/drugportal/>

## Пошук та аналіз літератури в базах даних Scopus та Web of Science. Оцінка наукової роботи і наукометричні показники.

Іншим не менш важливим ресурсом для роботи з науковою літературою є база даних (БД) Scopus. Деякі функції Scopus подібні до таких у PubMed, але є цілий ряд суттєвих відмінностей. Отже для найбільш ефективного опрацювання наукової літератури це не питання – PubMed чи Scopus? – а їх поєднання – PubMed та Scopus. На відміну від PubMed користування БД Scopus не є безкоштовним. Студенти з університетських аудиторій мають доступ до цієї БД, адже Київський національний університет імені Тараса Шевченка має на неї підписку.

Однією з унікальних особливостей (БД) Scopus – це наявність наукометричних показників, таких як індекс Хірша (h index), тому насамперед нам потрібно познайомитись із сучасною наукометриєю. Чи можна якось кількісно «виміряти» внесок окремого науковця, організації чи наукового видання – це дуже дискусійне питання<sup>36 37 38</sup>. Так, вчені навіть дискутують, чи дійсно найкращою статтею автора є та, яка отримала найбільшу кількість цитувань<sup>39</sup>. У 2005 р. американський фізик Джордж Хірш (Jorge E. Hirsch)

---

<sup>36</sup> Комісаренко С.В., Романюк С.І. (2015). Нотатки з етики в науці, або чи можуть відкриття бути запланованими і як здобутки другої наукової революції допоможуть вижити українській науці. *Вісник НАН України*, №4, с. 13-23.

<sup>37</sup> Van Noorden R. (2010). Metrics: A profusion of measures. *Nature* 465, 864–866. <https://www.nature.com/articles/465864a.pdf>

<sup>38</sup> Colquhoun D. (2003). Challenging the tyranny of impact factors. *Nature* 423: 479. <http://www.dscience.net/colquhoun-nature-impact-2003.pdf>

<sup>39</sup> Ioannidis J., Boyack K., Small, H. et al. (2014). Bibliometrics: Is your most cited work your best? *Nature* 514, 561–562. <https://www.nature.com/articles/514561a.pdf>

опублікував статтю у журналі *Proceedings of the National Academy of Sciences*, в якій він запропонував як можна розрахувати індекс, який би характеризував індивідуальний внесок вченого: «A scientist has index  $h$  if  $h$  of his or her  $N_p$  papers have at least  $h$  citations each and the other  $(N_p - h)$  papers have  $\leq h$  citations each»<sup>40</sup>.

Цікаво, що автор вже у цій піонерській роботі відмітив той факт, що «...high  $h$  indices in the life sciences are much higher than in physics». Зокрема це пов'язано з тим, що у різних наукових галузях працює дуже різна кількість вчених, видається різна кількість журналів тощо, і тому кількість цитувань робіт теж є дуже різною. Отже, однією з проблемою є те, що з використанням цього індексу не можна один до одного порівнювати успіхи окремих вчених, які працюють у різних наукових галузях. І якщо кількість публікацій (чим їх більше тим краще) не викликає значних суперечок, то оцінка якості публікацій за цим формальним показником досі залишається тим питанням, навколо якого точиться багато дискусій. Наприклад, якщо у дослідженні сталась якась помилка (дизайн експериментів, інтерпретація даних тощо), то не дивно, що така стаття спричинить багато дискусій, а отже і її цитувань буде багато, але у цьому прикладі у негативному контексті. Відмітимо, що ця піонерська робота Хірша отримала велику кількість цитувань – 6287 на цей час (цікаво, що наступна найбільш цитована його робота у галузі фізики має 2186 цитувань), що само по собі свідчить про актуальність цього питання.

Значно кращим і більш об'єктивним наукометричним показником є так званий імпакт фактор журналу – це один з узагальнювальних показників стосовно цитувань робіт, які публікує той чи інший журнал. Він розраховується

---

<sup>40</sup> Hirsch J.E. (2005). An index to quantify an individual's scientific research output.

*Proceedings of the National Academy of Sciences* 102 (46), 16569-16572.

<https://www.pnas.org/doi/epdf/10.1073/pnas.0507655102>

як відношення кількості цитувань у поточному році (А) статей, що були опубліковані у журналі протягом двох останніх років (Б), тобто  $IF=A/B$ . Таким чином, імпакт фактор журналу – це кількісна характеристика частоти цитування «середньої» статті за певний період часу (замість 2-х років також може бути період у 5 років).

Цей показник теж не є однозначним, адже на нього впливають такі фактори, як відносна кількість оглядових статей, які як правило отримують більшу кількість цитувань, самоцитування авторів статей, знову таки наукова галузь досліджень та багато інших<sup>41</sup>. Clarivate аналізує і регулярно публікує оновлені дані щодо імпакт факторів журналів в Journal Citation Reports (JCR)<sup>42</sup>.

Останнім часом все більше уваги приділяється так званим квартилям журналів, які дозволяють порівняти журнали з різних наукових галузей. Для визначення квартилю журнали в одній певній галузі розташовуються по мірі зменшення їх імпакт-факторів, і тоді до першого квартилю (позначається Q1) потрапляють верхні зі списку журнали (>75%), до Q2 потрапляє група 51-75%, до Q3 – група 26-50% і до Q4 – група 25% і менше. Таким чином, до Q1 у різних галузях науки можуть відноситись журнали з дуже різними імпакт факторами.

Важливо також зазначити, що один і той самий журнал може відноситись до різних квартилей в залежності від дисципліни. Інформацію щодо рейтингу журналу можна отримати за допомогою ресурсу SJR Scimago Journal & Country

---

<sup>41</sup> The Clarivate Analytics Impact Factor.

<https://clarivate.com/webofsciencegroup/essays/impact-factor/>

<sup>42</sup> SCI® Journal Citation Reports®: a bibliometric analysis of science journals in the ISI® database. Philadelphia: Institute for Scientific Information, Inc.®, 1993.

Rank<sup>43</sup>. Наведемо ілюстрацію з цього ресурсу для журналу Cell Calcium, яка надається як у табличній (Табл. 1.2.3), так і у графічній формах (Рис. 1.2.20).

**Таблиця 1.2.3.** Квартилі журналу Cell Calcium, який стосується трьох дисциплін, за період з 1999 р. по 2020 р.

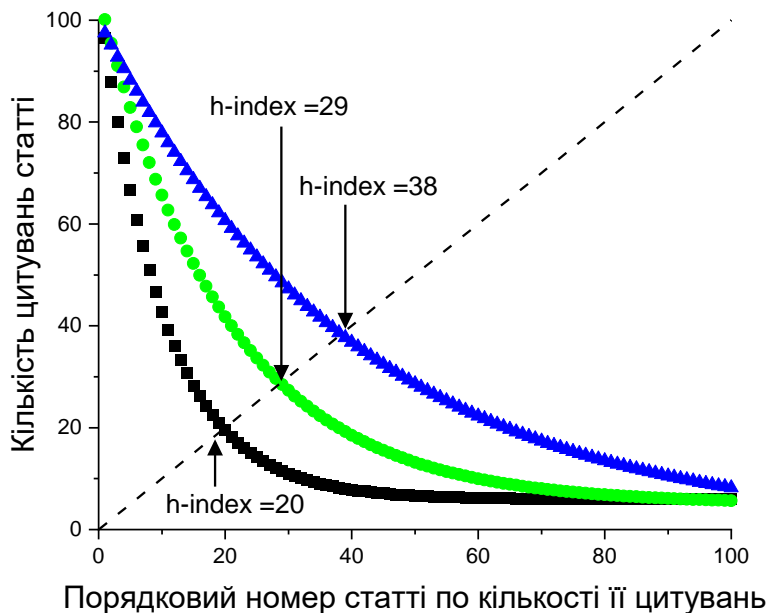
Рік	Категорія	Квартиль	Категорія	Квартиль	Категорія	Квартиль
1999	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2000	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2001	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2002	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2003	Cell Biology	Q2	Molecular Biology	Q2	Physiology	Q1
2004	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2005	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2006	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2007	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2008	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2009	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2010	Cell Biology	Q2	Molecular Biology	Q1	Physiology	Q1
2011	Cell Biology	Q2	Molecular Biology	Q2	Physiology	Q1
2012	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2013	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2014	Cell Biology	Q2	Molecular Biology	Q1	Physiology	Q1
2015	Cell Biology	Q2	Molecular Biology	Q2	Physiology	Q1
2016	Cell Biology	Q2	Molecular Biology	Q2	Physiology	Q1
2017	Cell Biology	Q2	Molecular Biology	Q1	Physiology	Q1
2018	Cell Biology	Q2	Molecular Biology	Q2	Physiology	Q1
2019	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1
2020	Cell Biology	Q1	Molecular Biology	Q1	Physiology	Q1



**Рис. 1.2.20.** Графічне представлення цієї інформації на веб-сторінці SJR Scimago Journal & Country Rank. Q1 - зелений колір, Q2 - жовтий колір.

<sup>43</sup> Scimago Journal & Country Rank. <https://www.scimagojr.com/>

Графічно визначення індексу Хірша проілюстровано на Рис. 1.2.21.



**Рис. 1.2.21.** Графічна ілюстрація розрахунку індексу Хірша.

З часом у намаганні врахувати певні обмеження h-індексу були запропоновані інші його варіанти. Вони добре представлені і пояснені розробниками програми Publish or Perish, яка є безкоштовною і надає зручні можливості для пошуку літературних джерел<sup>44</sup>. Наведемо та пояснимо деякі з цих індексів з посиланням на цей ресурс<sup>45</sup>:

- g-індекс надає більше ваги статтям, що отримали велику кількість цитувань (розраховується кількість цитувань у квадраті);

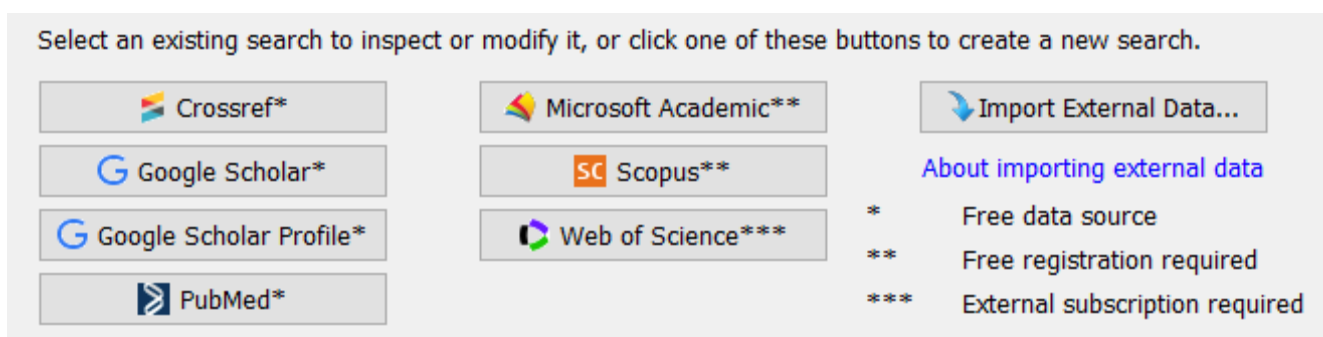
---

<sup>44</sup> Publish or Perish: A citation analysis software program which helps academics to present their case for research impact to its best advantage. <https://harzing.com/resources>

<sup>45</sup> Publish or Perish: Citation metrics. <https://harzing.com/resources/publish-or-perish/manual/using/query-results/metrics>

- сучасний h-індекс (hc-index) надає більшу вагу статтям, які були опубліковані більш недавно;
- індивідуальний h-індекс (є три його варіанти - hI-index, hI,norm, hm-index) робить спробу врахувати кількість співавторів статті.

Програма Publish or Perish є досить універсальним ресурсом, адже вона надає можливість проводити пошук першоджерел у декількох базах даних (Рис. 1.2.22).



**Рис. 1.2.22.** Скріншот сторінки Publish or Perish, що ілюструє з якими саме базами даних працює ця програма.

Розглянемо більше детально дві з них – вже згадану вище БД Scopus, а також Web of Science (скорочено WoS) як два авторитетні ресурси для роботи з науковою літературою. Scopus – це ресурс Elsevier, база даних публікацій і їх цитувань, яка формується та перевіряється експертами<sup>46</sup>. Далеко не всі журнали, книги, матеріали конференцій та інші публікації проходять такий ретельний відбір. Більш того, Scopus надає набір аналітичних інструментів для розрахунку наукометричних показників. На цей час Scopus включає приблизно 27 млн.

<sup>46</sup> About Scopus - Abstract and citation database Elsevier.

<https://www.elsevier.com/solutions/scopus>

документів. Пошук в цій базі даних можна проводити як для окремих публікацій, так і за авторами або організаціями (Рис. 1.2.23).



Рис. 1.2.23. Варіанти пошуку в Scopus (скріншот заголовної сторінки Scopus).

Наведемо приклад пошуку для автора. У результаті такого пошуку ми отримуємо не тільки список публікацій даного автора, а і розгорнутий аналіз публікаційної активності. На першій сторінці, над списком публікацій, є метричні показники – загальна кількість публікацій автора у БД Scopus, а також загальна кількість цитувань усіх публікацій (Рис. 1.2.24). Графічно представлена динаміка цих показників по роках, а також перелічені основні галузі досліджень автора.

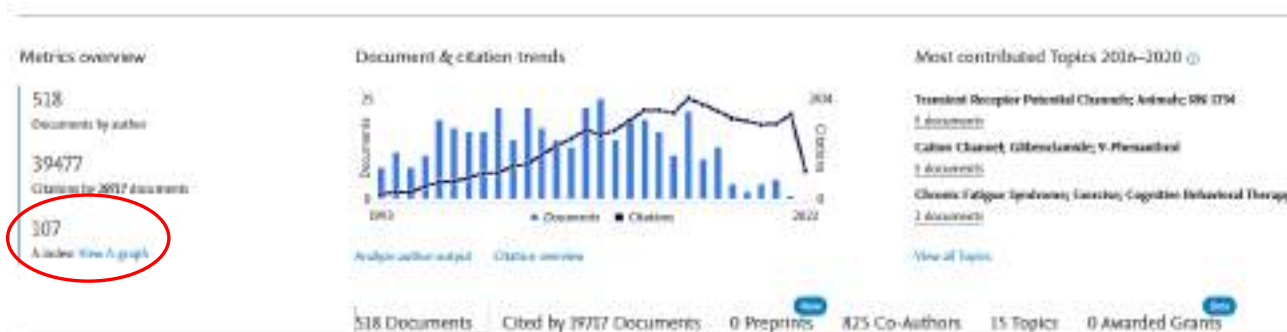
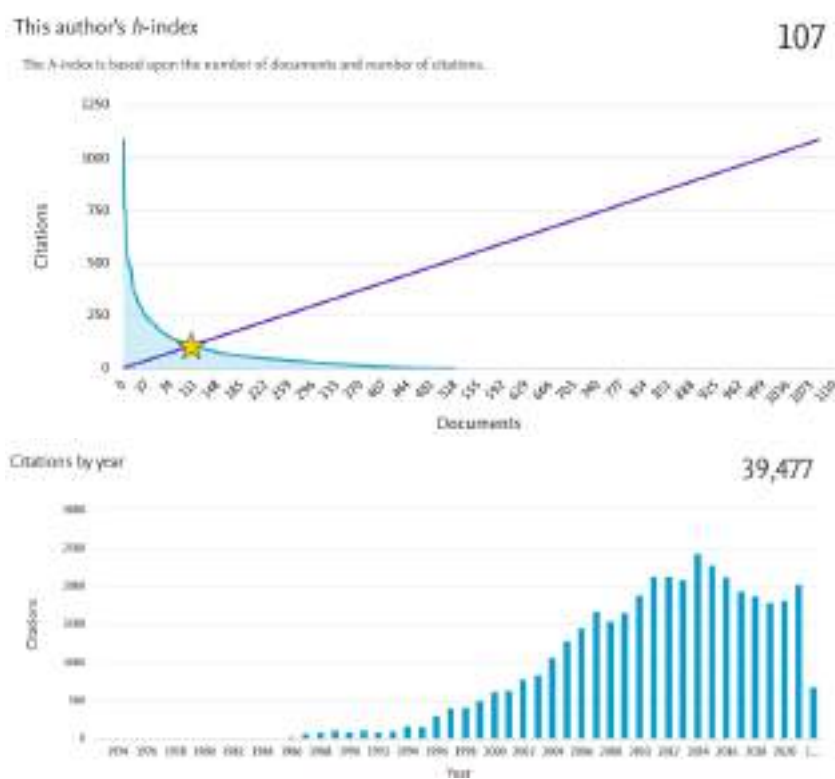


Рис. 1.2.24. Метричні показники публікацій науковця у БД Scopus.

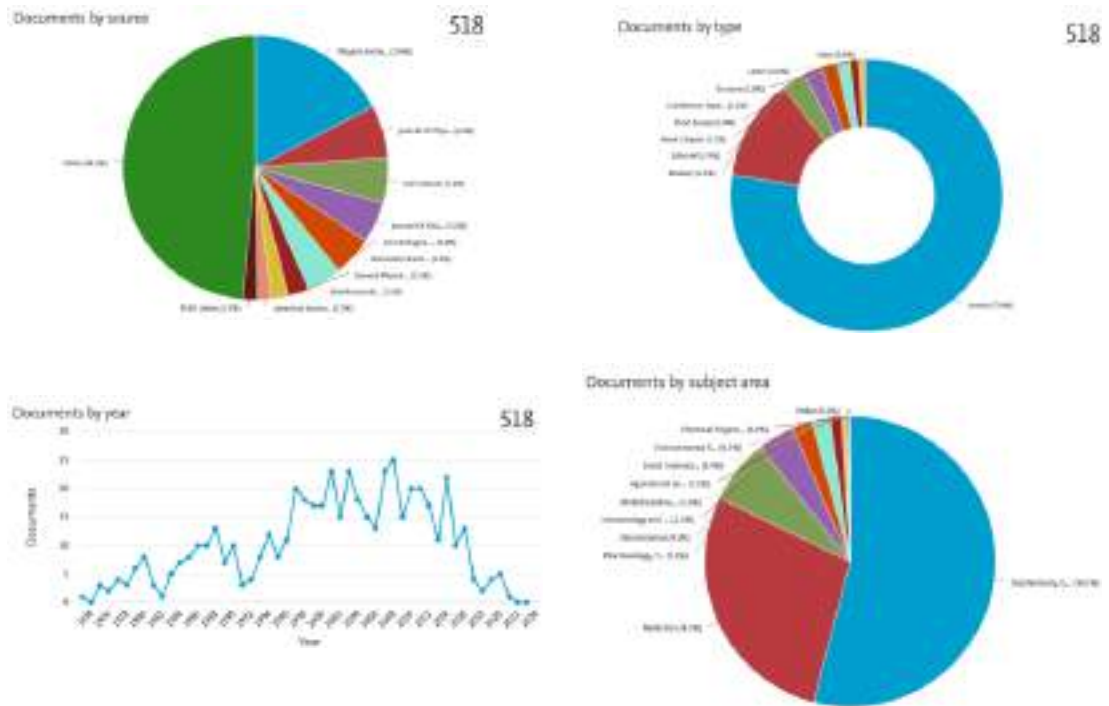
Якщо клацнути курсором мишки по «View h-graph», то ми отримаємо ще більш деталізований аналіз (Рис. 1.2.25). По перше, це графік, який ілюструє зв'язок між кількістю цитувань і кількістю публікацій (порівняйте Рис. 1.2.25 і 1.2.21). Зірочка на перетині кривих позначає точку на графіку, де порядковий номер публікації співпадає з кількістю її цитувань, тобто це і є h-індекс. У цьому прикладі індекс Хірша становить 107. Нижче наведено графік кількості цитувань усіх статей автора по роках.



**Рис. 1.2.25.** Аналіз кількості цитувань автора у БД Scopus.

Інструменти БД Scopus надають додаткові можливості для отримання інформації щодо журналів, в яких були опубліковані ці статті, кількості публікацій кожного року, типу публікації (оригінальна стаття, оглядова стаття, монографія тощо), а також до яких наукових галузей відносяться ці публікації

(Рис. 1.2.26). Дані представлені у вигляді кругової діаграми, що дозволяє візуалізувати відносний внесок тих чи інших кількісних показників.



**Рис. 1.2.26.** Кількісний та якісний аналіз публікацій науковця у БД Scopus.

БД Web of Science (WoS) – це ресурс компанії Clarivate Analytics, яка до 2016 р. була частиною відомої компанії Thomson Reuters. БД Web of Science складається з основної бази, або колекції (WoS Core Collection) та більш широкої платформи WoS. Якщо в питаннях вмісту БД Scopus співпрацює з міжнародною групою експертів (Content Selection and Advisory Board), то WoS має власну групу експертів, які відповідають за підбір контенту цієї бази даних. Особливо високим вимогам відповідають WoS Core Collection, і зокрема Science Citation Index Expanded, Social Sciences Citation Index та Arts & Humanities Citation Index. Тому не дивно, що розмір і наповнення цих баз даних дещо різний, загалом даних у WoS дещо менше, ніж у Scopus, і вони перекриваються між собою приблизно на 25%. Отже, робимо висновок, що варто

використовувати обидві бази даних, адже кожна з них має як унікальні інструменти, так і значною мірою унікальний контент.

Фахівець з аналітичних ресурсів компанії Clarivate Analytics к.б.н. І.О. Тихонкова на каналі YouTube пропонує [цикл вебінарів](#) українською мовою. Теми цих вебінарів стосуються широкого кола питань, зокрема тих, які розглядаються на практичних заняттях з дисципліни «Сучасні інформаційні технології в біології». Студентам можна порадити прослухати наступні:

[Розширені можливості нового інтерфейсу Web of Science](#)

[Профіль установи у Web of Science](#)

[Як опублікуватися у виданні, що індексується Web of Science](#)

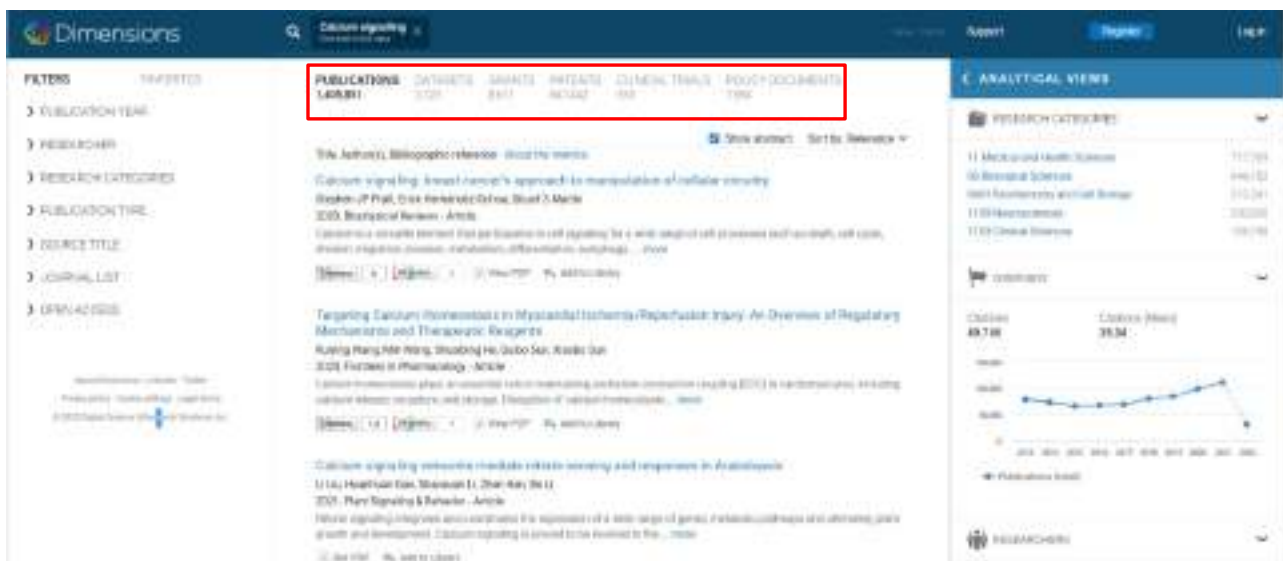
Є також вебіари англійською мовою від інших фахівців, наприклад [Literature searching with Web of Science](#).

Крім розглянутих нами детально БД Scopus та Web of Science є інші міждисциплінарні бібліографічні бази даних, такі як Dimensions, Crossref та Microsoft Academic. Dimensions – це ресурс Digital Science, який включає не тільки літературні джерела, а і гранти, патенти, клінічні випробування, оригінальні результати досліджень<sup>47</sup> (Рис. 1.2.27). Також є можливість застосувати різні типи фільтрів для звуження пошуку. Наприклад, пошук за ключовими словами «кальцієва сигналізація» продукує понад 1,4 млн. літературних джерел, і крім того 2123 наборів даних, 8617 грантів, 319 клінічних випробувань і 1554 регуляторних документів (Рис. 1.2.27).

Під посиланнями знаходимо іконки, що показують кількість цитувань, можливість отримання повного тексту статті, а також додати публікацію до власної бібліографічної бази - ReadCube Papers Cloud Library.

---

<sup>47</sup> Dimensions data and solutions for discovery and analytics. <https://www.dimensions.ai/>



**Рис. 1.2.27.** Скріншот веб сторінки Dimensions, який ілюструє результати пошуку інформації у галузі кальцієвої сигналізації.

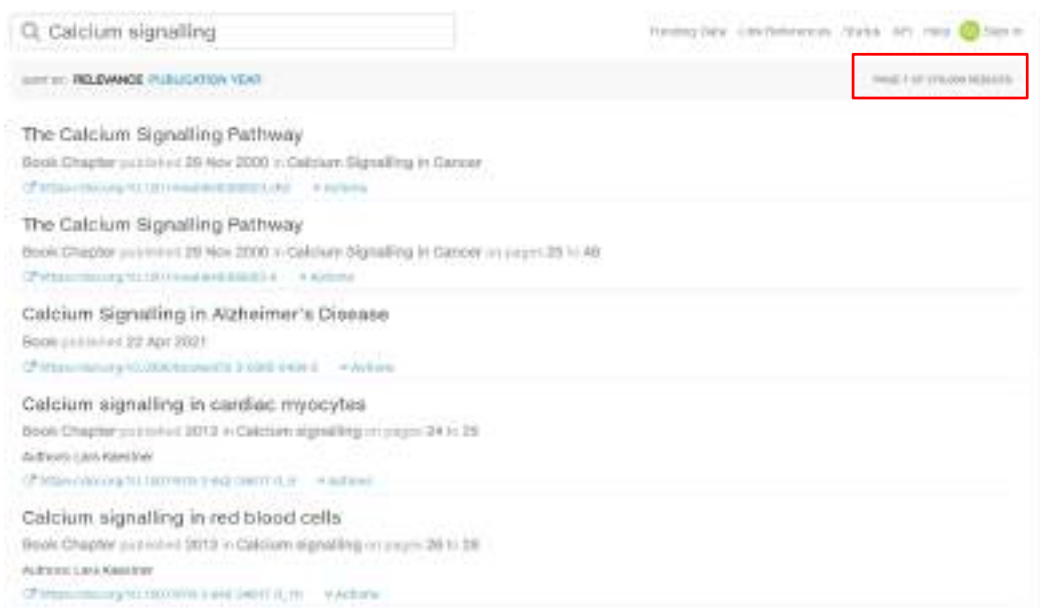
Crossref – це неприбуткова організація, метою якої є сприяння обміну інформацією академічного характеру<sup>48</sup>. Рис. 1.2.28 ілюструє результати пошуку літератури у цій базі даних – понад 270 тис. джерел у даному прикладі розташованих за їх релевантністю.

Microsoft Academic – це ресурс від компанії Microsoft (Microsoft Research)<sup>49</sup>. Великі набори даних є у вільному доступі<sup>50</sup>.

<sup>48</sup> Crossref. <https://www.crossref.org/>

<sup>49</sup> Microsoft Academic. <https://www.microsoft.com/en-us/research/about-microsoft-research/>

<sup>50</sup> Microsoft Academic Graph. <https://zenodo.org/record/2628216>



**Рис. 1.2.28.** Скріншот веб сторінки Crossref, який ілюструє результати пошуку інформації у галузі кальцієвої сигналізації.

Visser et al. у статті, яка була опублікована у журналі *Quantitative Science Studies*<sup>51</sup>, надали розгорнутий порівняльний аналіз цих баз даних, у тому числі наскільки їх контент перекривається чи є унікальним і яка між ними є різниця в залежності від часу публікації, типу публікації, наукової галузі, а також повноти і точності посилань на цитування робіт (пригадайте основні властивості інформації, Табл. 1.1.1). Зокрема автори відмітили той факт, що 1,9% лінків, що містить БД WoS, вони не змогли знайти у БД Scopus, а якщо перевірити навпаки, то це 5,8%. Розбіжності з іншими базами даних є навіть більш суттєвими, наприклад до 12,7% у БД Scopus відсутні у Microsoft Academic, а якщо порівняти БД Scopus та Crossref, то ця розбіжність досягає 57,9%.

---

<sup>51</sup> Visser M., van Eck N.J., Waltman L. (2020). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies* 2 (1): 20–41.

## **Ознайомлення з функціями програми для роботи з науковою літературою Mendeley.**

Після знаходження потрібних літературних джерел у пошуковій системі PubMed виникає питання – як можна їх зберегти, а потім вставити ці цитування у документ MS Word при написанні курсової роботи чи наукової статті, а також сформувати список цитованих першоджерел. Всі ці операції дозволяє виконати програмний ресурс від Elsevier – Mendeley. Ця програма є у вільному доступі, але її використання потребує створення акаунту (електронна адреса та пароль користувача)<sup>52</sup>. По суті Mendeley – це бібліографічний менеджер для роботи з літературою, у базі даних якого є понад 100 млн. джерел, і їх кількість постійно зростає. Пошук можна починати безпосередньо з домашньої сторінки цього ресурсу (Рис. 1.2.29).

З посиланням на цей ресурс зазначимо, що він дозволяє:

- зберігати, упорядковувати та шукати всі посилання у своїй бібліотеці;
- легко вставляти посилання у свої документи MS Word за допомогою Mendeley Cite;
- читати, виділяти та коментувати PDF файли статей, а також додавати свої коментарі до опису статті;
- співпрацюйте з іншими користувачами Mendeley, обмінюватись з ними бібліотеками та думками.

Mendeley також дозволяє експортувати та імпортувати бібліографічні дані у форматах BibTeX, EndNote XML та RIS для використання у інших бібліографічних менеджерах, наприклад Reference Manager або EndNote<sup>53</sup>.

---

<sup>52</sup> Mendeley. <https://www.mendeley.com>

<sup>53</sup> EndNote. <https://endnote.com/>



**Рис. 1.2.29.** Скріншот заголовної веб-сторінки Mendeley.

Вищезгаданий Mendeley Cite для MS Word дозволяє швидко вставляти посилання з бібліотеки Mendeley у документ Word, автоматично змінювати формат цитувань відповідно до конкретних вимог того чи іншого журналу та генерувати список використаних літературних джерел. Це безкоштовний додаток, який потрібно завантажити та інсталиювати у Word (Рис. 1.2.30).

Mendeley Cite надає такі можливості:

- знаходити посилання у своїй бібліотеці Mendeley та вставляти їх у документ Word, при цьому одночасно можна вставляти одне або декілька посилань;
- генерувати список літератури для всіх цитувань у цьому документі;
- довільно міняти формат цитувань, при цьому відбувається автоматичне оновлення як формату цитат у тексті, так і у списку літератури;
- ваша бібліотека Mendeley зберігається та завантажується за допомогою хмарних технологій (це власні сервера Mendeley), отже її неможливо втратити (згадайте таку властивість інформації як її захищеність та вимоги до резервних копій файлів у Табл. 1.1.1);

- Mendeley Cite відкривається як окрема панель і не заважає працювати з текстом Word документа.



**Рис. 1.2.30.** Скріншот веб-сторінки, де можна отримати Mendeley Cite <https://appsource.microsoft.com/en-US/product/office/wa104382081>

Mendeley є як в онлайн версії (Mendeley Web), що не потребує інсталяції на комп'ютері, але працює лише за наявності зв'язку з Інтернетом, так і off-line версії, яка потребує інсталяції і працює автономно – це програма Mendeley Desktop<sup>54</sup>. Для імпорту посилань з PubMed до інтернет-браузера потрібно додати Web Importer<sup>55</sup> – це App для браузера, яке дозволяє легко додавати посилання з будь-якої веб-сторінки.

У самій програмі джерела можна сортувати за такими категоріями:

- Recently Added – це ті, які ви додали останніми.
- Recently Read – показує список нещодавно прочитаних PDF файлів.

---

<sup>54</sup> Mendeley Desktop. <https://www.mendeley.com/download-reference-manager/windows>

<sup>55</sup> Mendeley Web Importer для Mozilla Firefox. <https://addons.mozilla.org/en-US/firefox/addon/mendeley-web-importer/>

- Favorites – ті, що відмічені користувачем.
- Needs Review – ті, що потребують уваги, наприклад уточнення інформації про джерело, неправильній рік чи номери сторінок тощо. Після внесених коректив статтю можна видалити з цього списку шляхом підтвердження коректності інформації.
- My Publications – містить список публікацій автора - користувача Mendeley.
- Unsorted – посилання, які не були додані до жодної з наведених вищій категорій.

У разі виникнення питань можна скористатись центром підтримки<sup>56</sup>, а також довідником для користувачів<sup>57</sup>.

Студенти отримують практичні навички користування PubMed та Mendeley при виконанні практичної роботи 4.1.

---

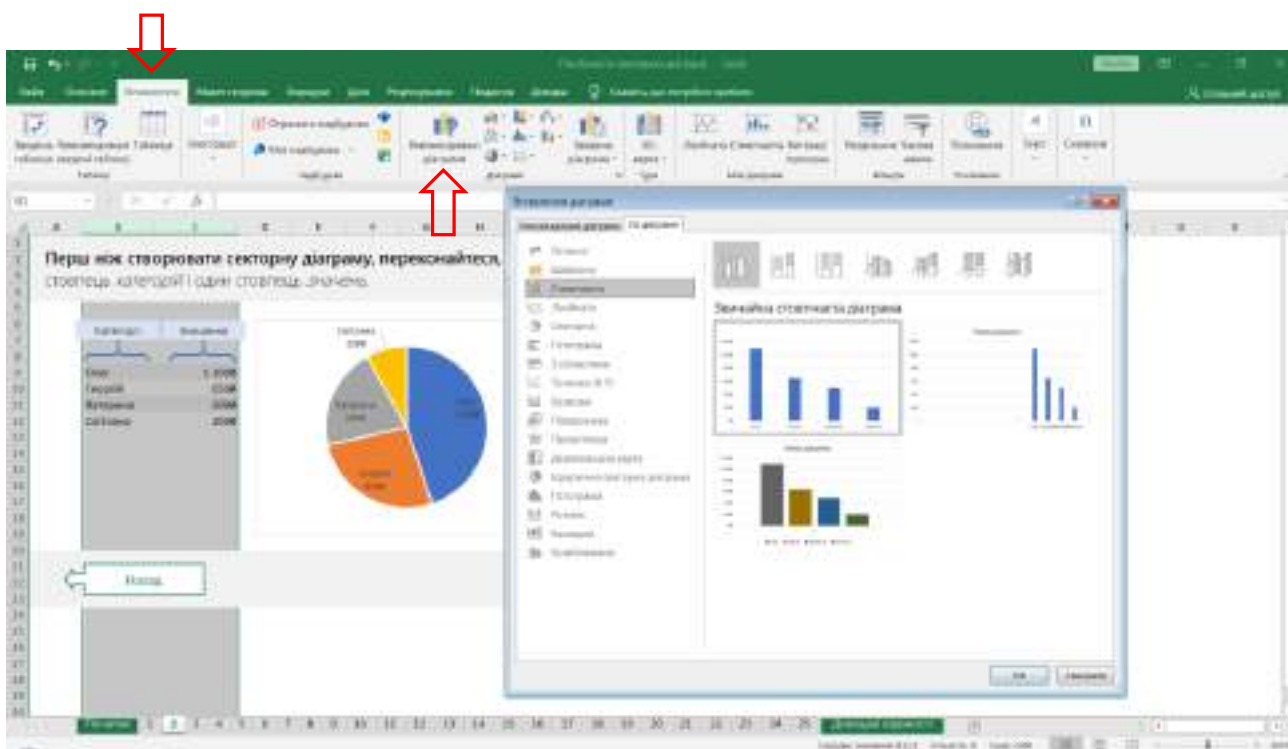
<sup>56</sup> Mendeley Support Center. <https://service.elsevier.com/app/home/supporthub/mendeley/>

<sup>57</sup> Help guides. <https://www.mendeley.com/guides>

### *1.3. Графічне представлення даних векторної графіки за допомогою програм CorelDraw та Inkscape. Графічне представлення даних растрової графіки за допомогою програми ImageJ.*

#### **MS Excel**

Для кращого порівняння з програмами векторної графіки спочатку розглянемо графічні можливості MS Excel, з якими студенти, як правило, вже добре знайомі. Їх скориставшись вищезгаданим посібником з секторних діаграм. Для візуалізації даних у вигляді діаграми потрібно вибрати стовпчики з даними, а потім через меню «Вставлення» обрати один з рекомендованих форматів для діаграми (Рис. 1.3.1).



**Рис. 1.3.1.** Графічне представлення даних в MS Excel у вигляді діаграм.

Як саме створювати графіки різних типів в MS Excel буде детально продемонстровано у лабораторній роботі 4.2.

## CorelDRAW

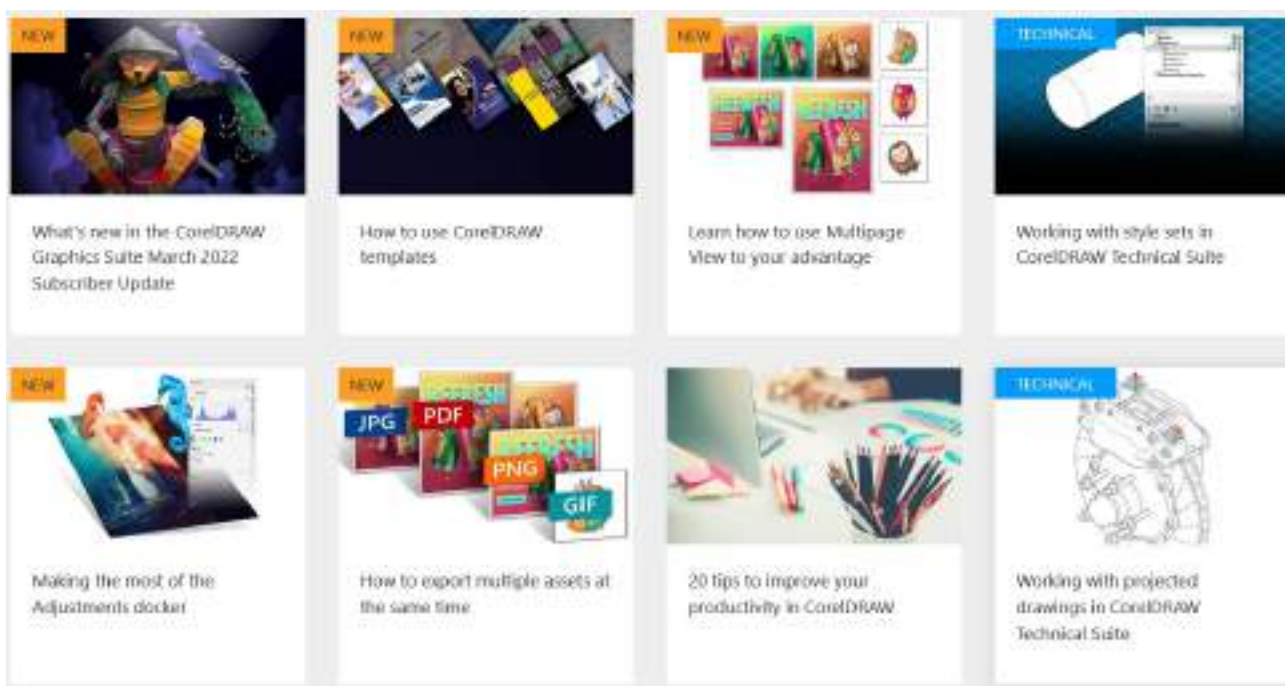
Для якісної ілюстрації результатів наукових досліджень корисною може бути векторна графіка, яка застосовується в CorelDRAW<sup>58</sup>. Векторні зображення складаються з елементів (лінії, форми, тощо), які зберігаються разом з геометричними формулами. Іншими словами, всі точки і лінії на зображенні математично визначені, отже їх можна редагувати з високим рівнем деталізації і при збільшенні масштабу зображення його якість не втрачається. Найпоширенішими типами файлів векторної графіки є такі: \*.CDR, SVG, \*PDF, \*EPS, \*AI (з позначкою \* можуть бути і растровими зображеннями).

Растрові (або бітові) зображення у цифровому форматі представлені масивом пікселей у прямокутній сітці. При збільшенні такого зображення в залежності від кількості та розмірів пікселей чіткість зображення може досить суттєво втрачатись. Різними типами такого роду зображень є BMP, JPG, JPEG, GIF, PICT, PNG, \*PSD, TIFF (\*PSD може бути також і векторним зображенням). Для вивчення CorelDRAW можна порадити скористатись навчальним ресурсом від розробника цієї програми (Рис. 1.3.2)<sup>59</sup>. Ця програма є платною, але файли \*.CDR можна переглядати у безкоштовному Libre Office Draw.

---

<sup>58</sup> CorelDRAW. <https://www.coreldraw.com/en/>

<sup>59</sup> CorelDRAW Tutorials. <https://www.coreldraw.com/en/learn/tutorials/?topNav=en#ctg-education>



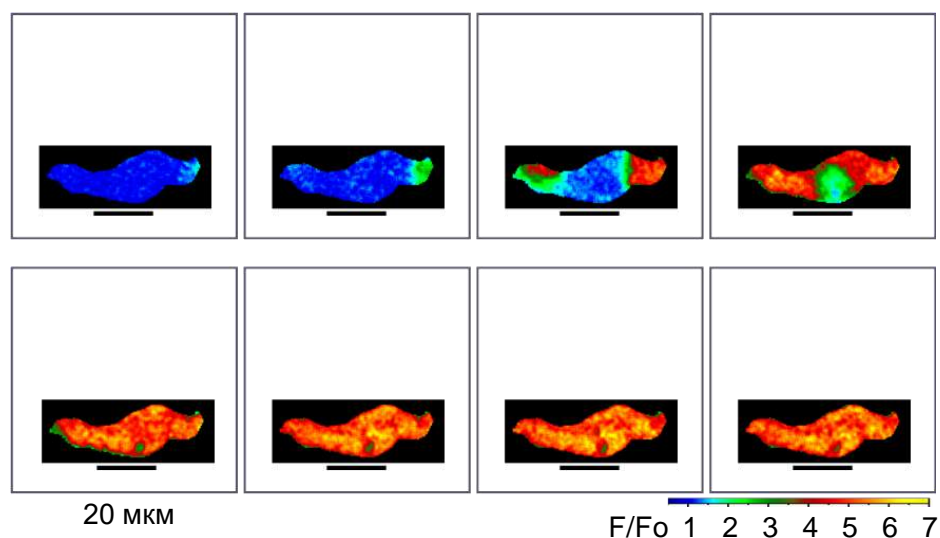
**Рис. 1.3.2.** Скріншот сторінки навчальних ресурсів CorelDRAW.

В якості приклада застосування CorelDRAW у наукових дослідженнях проілюструємо конфокальні зображення розповсюдження внутрішньоклітинних кальцієвих хвиль у гладеньком'язових клітинах тонкого кишечника, для візуалізації яких ми застосовували флуоресцентний кальцієвий індикатор fluo-3, а отримані за допомогою лазерної конфокальної мікроскопії x-y скани окремих ізольованих клітин представляли у програмі CorelDRAW<sup>60</sup> (Рис. 1.3.3).

---

<sup>60</sup> Gordienko D.V, Zholos A.V. (2004). Regulation of muscarinic cationic current in myocytes from guinea-pig ileum by intracellular Ca<sup>2+</sup> release: a central role of inositol 1,4,5-trisphosphate receptors. *Cell Calcium* 36(5):367-386.

<https://www.sciencedirect.com/science/article/abs/pii/S0143416004000740>



**Рис. 1.3.3.** Виникнення та розповсюдження кальцієвої хвилі у міоциті тонкого кишечника реєструвалось за допомогою методу лазерної конфокальної мікроскопії. Внутрішньоклітинна концентрація кальцію показана як нормована інтенсивність флуоресцентного сигналу кальцій-чутливого барвника fluo-3 у кольоровій шкалі від низьких (синій колір) до високих (жовтий колір) із усіма проміжними градаціями кольорів (модифіковано зі статті Gordienko, Zholos, Cell Calcium 2004).

## Inkscape

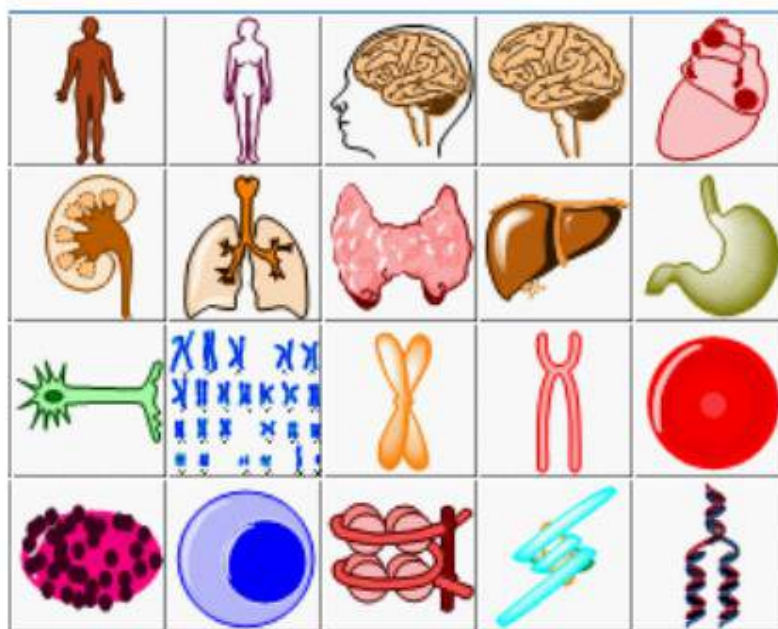
Inkscape<sup>61</sup> – це також програма для векторної графіки. Вона є безкоштовною. Програма включає гнучкі інструменти для малювання, забезпечує сумісність форматів файлів з іншими програмами (зокрема це SVG, редактор XML, PNG, OpenDocument Drawing, DXF, sk1, PDF, EPS та PostScript формати), а також надає має текстові інструменти. Розробники Inkscape

<sup>61</sup> Inkscape <https://inkscape.org/>

підготували посібники та відео презентації, з якими можна ознайомитись на веб сторінках Learning Inkscape<sup>62</sup>.

## BioDraw

BioDraw – це корисний ресурс для біологічних ілюстрацій, презентацій або розробки веб-сторінок з біологічним контентом. Є багато шаблонів BioDraw, які можна далі редагувати і вільно використовувати для власних цілей<sup>63</sup>. За допомогою меню File > Open > Templates користувач може завантажити один з потрібних шаблонів (Рис. 1.3.4).



**Рис. 1.3.4.** Ілюстрація шаблонів для анатомії у ChemBioOffice<sup>64</sup>.

<sup>62</sup> Learning Inkscape <https://inkscape.org/learn/>

<sup>63</sup> Gordon J. BioDraw Templates: A Biological Illustrator's Resource.  
<http://insideinformatics.cambridgesoft.com/articles/675/biodraw-templates>

<sup>64</sup> ChemBioOffice Ultra 14.0 Suite.

<https://www.adeptscience.de/products/lab/chembiooffice/chembiooffice-ultra-suite.html>

## ImageJ

NIH Image був започаткований у Національних Інститутах Здоров'я США (NIH) ще на самому початку запровадження комп'ютерних технологій у біомедичній галузі – у 1987 р. З часом він набув значної популярності і швидко прогресував, зокрема завдяки зусиллям багатьох дослідників і програмістів усього світу, адже це відкрите програмне забезпечення (open source software) на GitHub<sup>65</sup>. Стаття Schneider et al. (2012) отримала понад 40 тис. цитувань! Завдяки такій популярності серед вчених платформа містить багато «know how» додатків для вирішення тих чи інших конкретних наукових задач. Деякі з них інсталиються разом з інсталяцією програми, деякі потрібно додавати після інсталяції основної програми.

У 2012 р. був розроблений інтегрований пакет Fiji ImageJ для візуалізації та аналізу біологічних зображень, який включає більшість з таких додатків<sup>66</sup> (Рис. 1.3.5). Для роботи ImageJ на комп'ютері має бути також встановлене програмне середовище Java<sup>67</sup>.

При виконанні програми відкривається одне вікно, яке містить меню і панель інструментів під ним, і інше – консоль (Рис. 1.3.6). В меню «Аналіз» можна знайти багато інструментів для різного роду операцій із зображеннями. Вже згадані додатки (plugins) дозволяють виконати макроси і велику кількість різноманітних типів аналізу, зокрема кількісного аналізу зображень. Проілюструємо ці можливості наступним прикладом.

---

<sup>65</sup> Schneider CA, Rasband WS, Eliceiri KW. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 9(7):671-675.

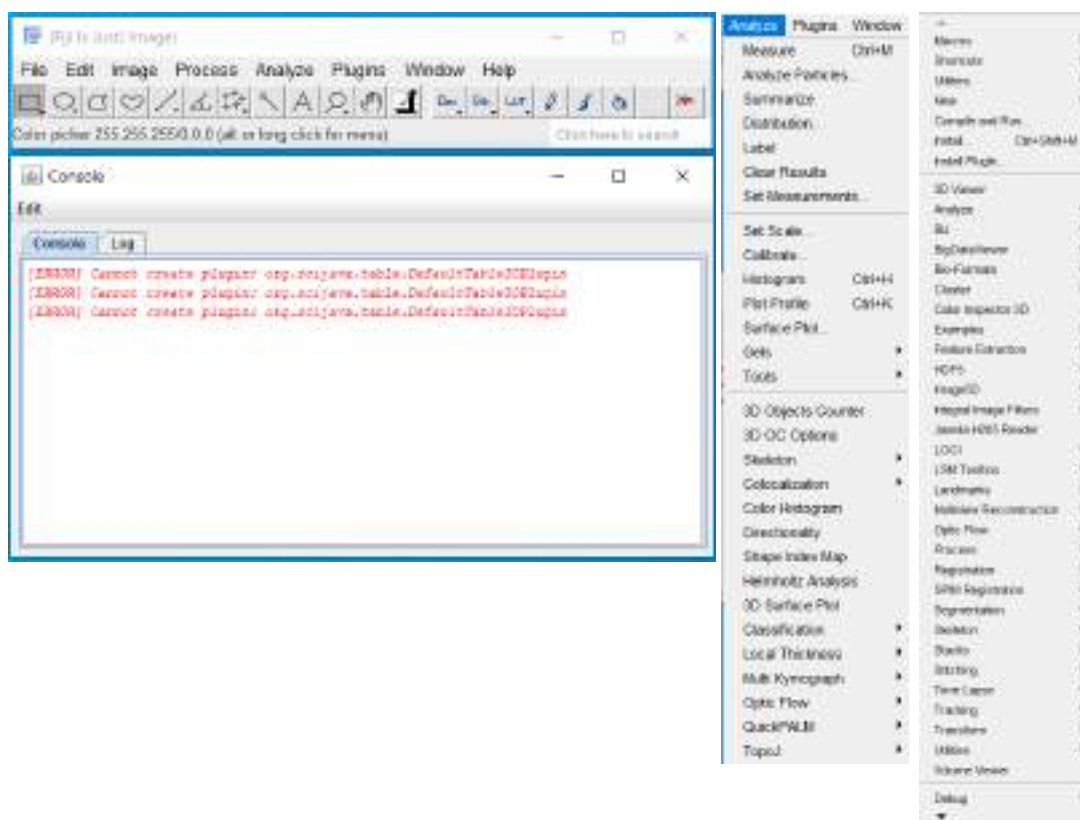
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5554542/>

<sup>66</sup> Schindelin et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9(7), 676–682.

<sup>67</sup> Java. <https://www.java.com/en/>



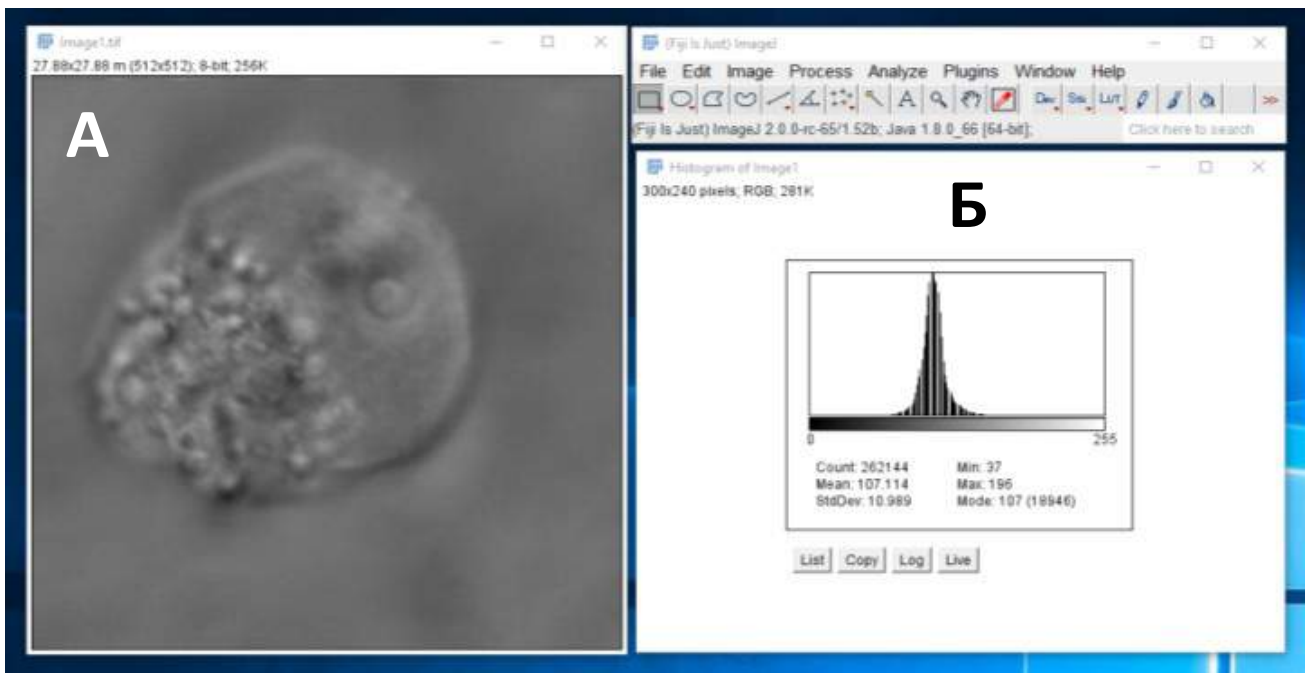
**Рис. 1.3.5.** Заголовна сторінка Fiji ImageJ, яка містить веб-лінк для завантаження Windows версії цієї безкоштовної програми<sup>68</sup> (для інших версій - Fiji Downloads <https://imagej.net/software/fiji/downloads>).



**Рис. 1.3.6.** Меню та можливості ImageJ для аналізу зображень.

<sup>68</sup> Fiji Image J image processing package. <https://fiji.sc/>

Оригінальна мікрофотографія клітини (Рис. 1.3.7А) не є дуже якісною з точки зору її контрасту та яскравості зображення. Нагадаємо, що контраст зображення залежить від різниці в яскравості між його світлими і темними частинами. Іншими словами, контраст визначає кількість відтінків – у даному прикладі відтінків сірого кольору (grayscale), тобто зображення містить тільки білий і чорний кольори і проміжні відтінки сірого кольору.

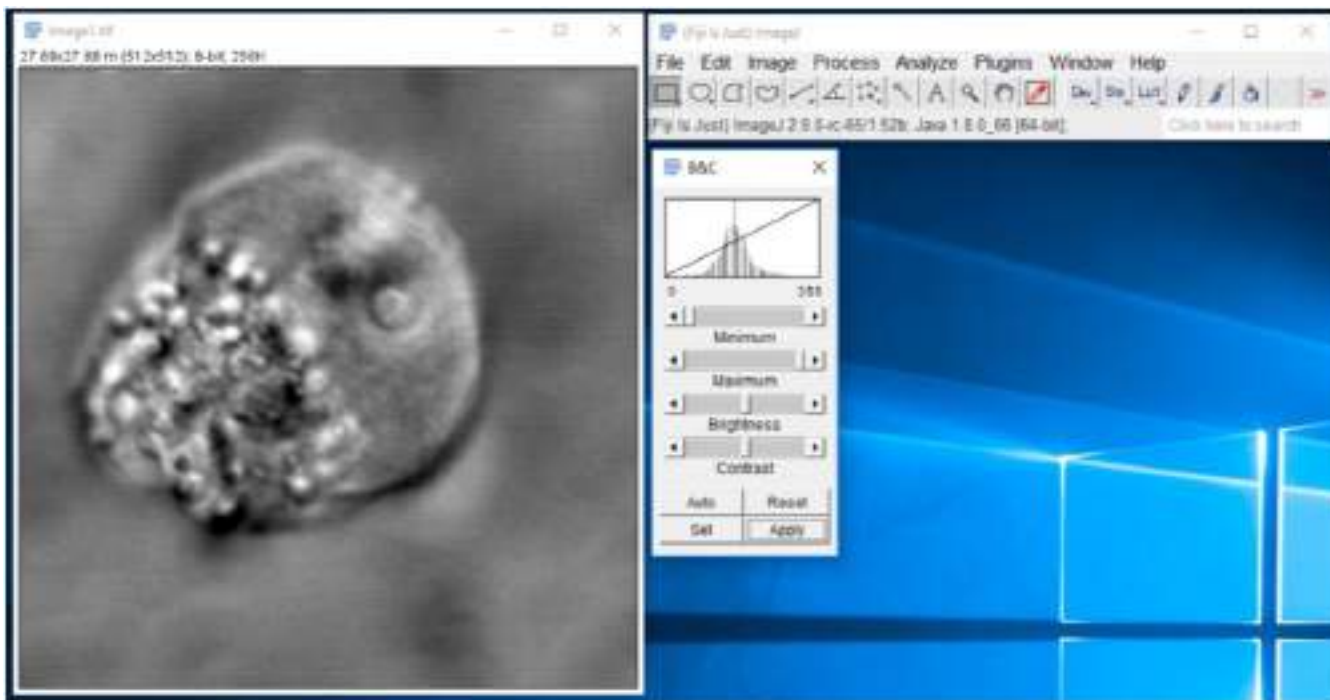


**Рис. 1.3.7.** Мікрофотографія клітини (А) і гістограма цього зображення (Б).

Кожен піксель grayscale записується у двійковому коді і складається з 8 біт. Відповідно є  $2^n$  різних комбінацій коду для кожного пікселя, де  $n$  – це бітність. Отже в grayscale є  $2^8=256$  відтінків сірого кольору. Яскравість зображення стосується його загальної світлості чи темності.

Чи можна покращити ці характеристики зображення не «на око», а шляхом комп'ютерного аналізу? Для цього за допомогою ImageJ спочатку побудуємо гістограму зображення (Рис. 1.3.7Б). Її аналіз показує, що є велика

кількість невикористаних значень відтінків сірого кольору. Наступним кроком за допомогою повзунків встановимо мінімальні і максимальні значення прямо на гістограмі, після чого отримуємо оптимізоване зображення (Рис. 1.3.8).



**Рис. 1.3.8.** Оптимізоване за допомогою ImageJ зображення клітини.

У цифрове зображення були внесені наступні корективи. Зміна яскравості відбулась шляхом додавання або віднімання певної сталої величини від значення для кожного пікселя, це зсуває гістограму уздовж осі абцис, але розподіл залишається таким же. Для покращення контрастності найменше значення (88 в оригінальному зображенні) приймається як 0, а найвище значення (166) прирівнюється до 255. Відповідно до цієї нової школи відбувається лінійно пропорційне масштабування значення кожного з пікселів зображення.

Студенти отримують практичні навички роботи з ImageJ під час виконання лабораторної роботи 4.5. Підкреслимо, що всі процедури щодо

внесення змін в оригінальні файли зображень мають бути деталізовані у частині «Методи дослідження». При підготовці наукових публікацій даних, як це зазначено в інструкціях для авторів у багатьох журналах, неприпустимо робити зміни вибірково в одній із частин зображення і не робити такі ж зміни в інших його частинах.

## РОЗДІЛ 2. ОСНОВИ РОБОТИ У ПРОГРАМІ ORIGIN ТА ПРОГРАМНОМУ СЕРЕДОВИЩІ R.

### *2.1. Основи роботи з науковими даними у програмі Origin: імпорт даних, їх аналіз, перетворення, апроксимація, створення сценаріїв обробки даних.*

#### **Вступ до Origin**

Наукова робота, дисертаційне дослідження чи курсова робота – всі вони вимагають обчислень за допомогою комп'ютерних технологій, статистичного аналізу даних, графічного представлення даних, їхнього аналізу та порівняння (наприклад, шляхом фітування даних тією чи іншою функцією). Всі ці задачі можна виконати в єдиному програмному продукті Origin, який був розроблений заснованою у 1992 р. компанією - OriginLab Corporation (США) і працює в операційній системі Microsoft Windows<sup>69</sup>. Зараз в усьому світі є понад 500 тис. користувачів цього програмного продукту, які працюють у приватних компаніях, університетах, дослідницьких лабораторіях, державних організаціях, на підприємствах тощо.

У цьому модулі студенти знайомляться з основними можливостями програми, яка зараз вважається світовим ресурсом №1 для інтерактивного аналізу і представлення результатів наукових досліджень. Про універсальність цієї програми свідчить те, що вона підтримує імпорт даних всіх існуючих форматів, від простого тексту в ASCII кодах і до зображень. Зважаючи на те, що більшість сучасних приладів у наукових чи комерційних лабораторіях генерують цифрові дані у вигляді окремих файлів, не дивно, що програма Origin дістала таке визнання та широке застосування.

---

<sup>69</sup> OriginLab. <https://www.originlab.com/>

Варто підкреслити і те, що з OriginLab Corporation співпрацюють науковці, інженери, програмісти всього світу шляхом створення програмних додатків (Origin Apps) для вирішення тих чи інших конкретних задач. Компанія розміщує ці додатки на своїх веб-сторінках. Автором були розроблені такі додатки для імпорту результатів петч-клемп експериментів з подальшим їхнім аналізом у автоматизованому режимі<sup>70 71</sup>, а також додаток для перегляду файлів<sup>72</sup>. На сайті OriginLab можна зареєструватися і завантажити безкоштовну демонстраційну версію програми, яка є повністю функційною, але термін її використання є обмеженим (21 день). Всі ілюстрації у цьому розділі зроблені автором у ліцензованій версії програми OriginPro 2021b (64-bit) SR2.

Якщо виникає питання, на яке користувач не може самотійно знайти відповідь, можна звернутись до міжнародної спільноти через форум<sup>73</sup>. Також є можливість обміну файлами між користувачами<sup>74</sup>. З цього сайту можна завантажити потрібні Apps, а для їх перегляду передбачене сортування Apps по даті завантаження, популярності, коментарям тощо. Крім того, є можливість скористатись фільтрами і обрати Apps з категорій: для графічного представлення даних, аналізу даних, роботи з даними, імпорту і експорту даних та інших, які не потрапляють у ці категорії.

---

<sup>70</sup> Zholos A.V. Simple pCLAMP Analyzer.

<http://www.originlab.com/fileExchange/details.aspx?fid=315>

<sup>71</sup> Zholos A.V. Advanced pCLAMP Analyzer.

<http://www.originlab.com/fileExchange/details.aspx?fid=357>

<sup>72</sup> Zholos A.V. Origin File Browser.

<https://www.originlab.com/fileExchange/details.aspx?mode=sandbox&fid=338>

<sup>73</sup> Origin User Forum. <https://my.originlab.com/forum/default.asp>

<sup>74</sup> Origin File Echange. <https://www.originlab.com/fileexchange/index.aspx>

З різними варіантами графічного представлення даних можна ознайомитись на відповідній веб сторінці<sup>75</sup>. Програмний пакет містить понад 175 графічних шаблонів, у тому числі для створення 3D графіків, якими легко скористатись. OriginLab також надає детальний опис аналітичних процедур та статистичного аналізу<sup>76</sup>. В Origin є понад 200 вбудованих функцій для фітування даних, і також можна записувати власні математичні рівняння для апроксимації результатів. Дуже корисним є ресурс, який містить відео презентації з детальними поясненнями всіх технічних моментів<sup>77</sup>. Для перегляду файлів, зроблених у програмі Origin, є безкоштовний Origin Viewer<sup>78</sup>.

Допомога від розробника програми включає вже згадані відео ілюстрації, блог, відповіді на типові запитання, форум користувачів, Apps, галерею графіків, ілюстровані приклади наукових досліджень та вебінари (Рис. 2.1.1).



**Рис. 2.1.1.** Скріншот веб сторінки, що ілюструє різні форми допомоги для користувача<sup>79</sup>.

<sup>75</sup> Origin Graphing. <https://www.originlab.com/index.aspx?go=Products/Origin/Graphing>

<sup>76</sup> Origin Data Analysis and Statistics.

[https://www.originlab.com/index.aspx?go=Products/Origin#Data\\_Analysis\\_and\\_Statistics](https://www.originlab.com/index.aspx?go=Products/Origin#Data_Analysis_and_Statistics)

<sup>77</sup> Origin/OriginPro Videos. <https://www.originlab.com/videos/index.aspx>

<sup>78</sup> Origin Viewer. <https://www.originlab.com/viewer/>

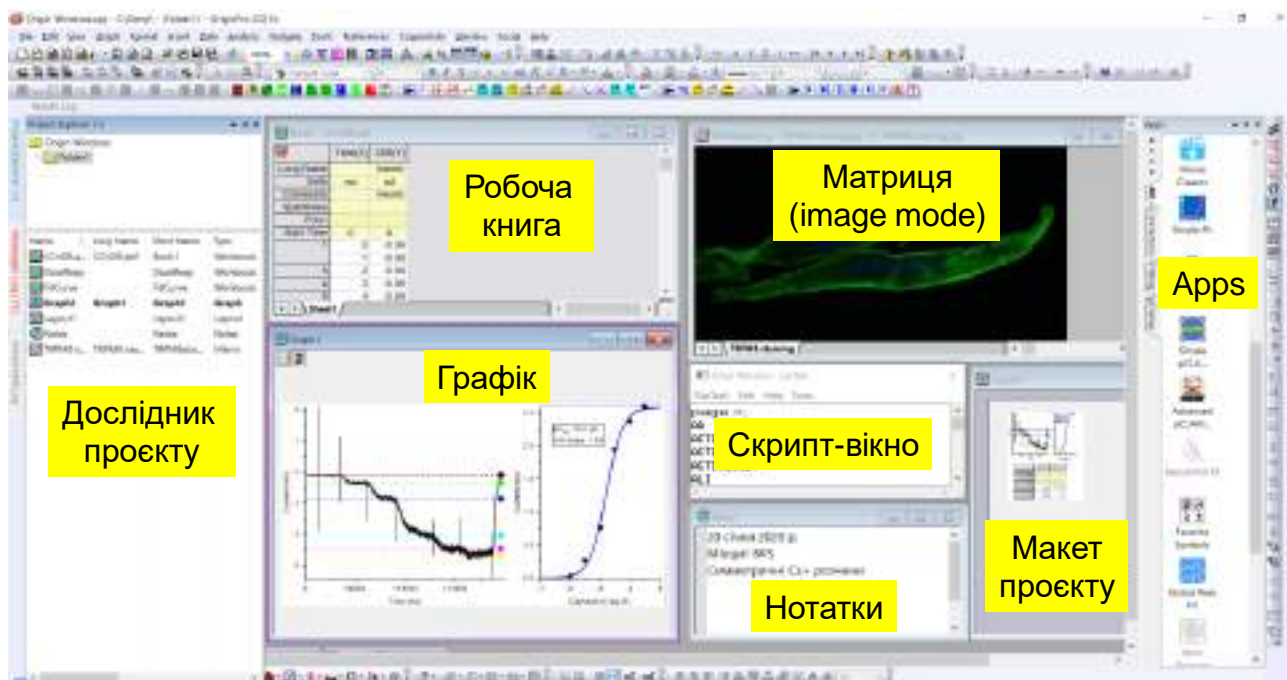
<sup>79</sup> Origin Help Center.

<https://www.originlab.com/index.aspx?go=Support/DocumentationAndHelpCenter>

Для першого знайомства з можливостями Origin варто переглянути брошуру<sup>80</sup>. Для подальшої роботи з програмою є детальний посібник користувача<sup>81</sup>.

**Основи роботи з науковими даними у програмі Origin: імпорт даних, їх аналіз, перетворення, апроксимація, створення сценаріїв обробки даних.**

Новий проєкт Origin відкривається з вікном “робоча книга” (Workbook). Іншими вікнами Origin є: матриця (Matrix), графік (Graph), скрипт-вікно (Script window), нотатки (Notes) і макет (Layout) (Рис. 2.1.2).



**Рис. 2.1.2.** Основні вікна програми Origin.

Зліва сторінки відкривається дослідник проєкту (Project Explorer), який дає можливість переглянути список всіх вікон даного відкритого проєкту.

<sup>80</sup> Origin Booklet. <https://www.originlab.com/brochure/>

<sup>81</sup> Origin User Guide. <https://www.originlab.com/doc/user-guide>

Робоча книга подібна до книги MS Excel, в ній також можна додавати нові аркуші (Sheet1, Sheet2 – до максимальної кількості 1024). Так само як і в MS Excel основним призначенням робочої книги є зберігання даних у цифровому форматі. Формат колонки може бути величини X, Y, Z (останній для 3D графіків), а також стандартна похибка або стандартне відхилення середньої величини. Звісно, у комірки можна також записувати і текстові дані, зокрема ярлички (label). Для робочих книг є певні обмеження щодо кількості колонок і рядків в одній книзі<sup>82</sup>. Наприклад, для книги з однією колонкою максимальна кількість рядків становить 200 млн. На практиці ці обмеження як правило не відчуються.

Вікно матриці – це насправді свого роду контейнер для однієї або більше матриць, який може мати одну або більше вкладок, і кожна з яких містить один або більше об'єктів матриць. Матриця (Matrix) виконує функцію встановлення взаємозв'язків між X, Y і Z величинами. Це вектор величин Z, які співвідносяться між собою в координатах X і Y відповідно до їх позиції в колонках і рядках. Вони є основою для створення 3D графіків. Зокрема, вони призначені для імпорту і оцифровки зображень, про що мова буде йти далі.

Графік (Graph) разом з робочою книгою є двома найбільш використовуваними вікнами Origin. Це основна структура для графічного представлення експериментальних даних і результатів їх аналізу. Графік може містити один або більше шарів (Layer), в які вносяться ті чи інші дані з робочих книг для їх візуалізації. Вверху зліва на графіку відображаються іконки, які позначають, з яким саме шаром працює у даний момент часу користувач (active layer). Кожен елемент шару (вісі, дані, підписи тощо) можна форматувати потрібним чином (змінювати ширину лінії, її колір, тип тощо).

---

<sup>82</sup> Origin Limitations. <https://www.originlab.com/doc/Origin-Help/Origin-Limitation>

Скрипт-вікно (Script window) – це командний інтерпретатор Origin. У ньому можна виконувати математичні операції (наприклад записавши  $\log(20)=$ ; отримуємо результат  $\log(20)=1.301029995664$ ), але більш важливе його призначення – це виконання команд вбудованої мови програмування Origin, яка називається LabTalk. Для написання скриптів за допомогою LabTalk є відповідний посібник користувача<sup>83</sup> (приклади скриптів також наведені у Додатку 2).

Нотатки (Notes) містять довільний текст. В нотатках можна наприклад записати дату експерименту, номер препарату чи клітини з протоколу дослідження, експериментальні умови такі як склад розчинів з тим, щоб ця інформація зберігалась в одному файлі разом з даними, результатами їх аналізу та графічного представлення.

Макет (Layout) призначений для візуалізації всіх компонентів проекту на одній сторінці. Сюди можна додавати любі інші вікна даного проекту.

Всі ці типи вікон можна відкрити з меню “New” або “Window”. Різні типи вікон дозволяють представити дані у різних форматах, наприклад, у вигляді робочої книги або графічному, що значно полегшує обробку результатів і їх графічне представлення.

Основи роботи з науковими даними у програмі Origin включають імпорт даних, їх аналіз, обробку, перетворення, фітування і графічне представлення, а також автоматизований аналіз даних.

Імпорт та експорт даних можливий у різних форматах: ASCII та Binary дані, файли MATLAB, pClamp, LabVIEW, SigmaPlot, Lotus, Image та багато інших. Origin інтерактивно допомагає у процесі імпорту даних через import wizard. Origin також дозволяє відкривати файли Microsoft Excel і працювати з

---

<sup>83</sup> LabTalk Scripting Guide. <https://www.originlab.com/doc/LabTalk/guide>

ними в середовищі Origin, а також експортувати робочі книги у форматі MS Excel.

Робота з даними та їх модифікація включає видалення зайвої кількості даних (reduce dataset), фільтрування даних, як це зокрема буває потрібно для «згладжування» кривих, екстрагування потрібних даних з великого їх набору, розрахунок та трансформація величин, сортування даних, їх нормалізацію, математичні операції з даними та ін.

Аналіз даних включає статистику (тест на нормальний розподіл, *t*-тест, ANOVA та ін.), апроксимацію даних функціями (як вбудовані функції Origin, так і власні функції), обробку сигналів, пошук локальних мінімумів і максимумів, аналіз зображень тощо.

Представлення наукових результатів можливе у формі 2D & 3D графіків (є більше 175 шаблонів графіків), статистичних графіків, зображень тощо. Для ілюстрації цих можливостей наведемо деякі з них (Рис. 2.1.3).



**Рис. 2.1.3.** Графічні можливості програми Origin<sup>84</sup>.

<sup>84</sup> Origin Graph Gallery. <https://www.originlab.com/www/products/graphgallery.aspx>

Для спілкування з користувачем (user interface) існують можливості організації великих проєктів з використанням Project Explorer, меню, інструментів та діалогових вікон.

Для програмування і автоматизація операцій в останній версії Origin є Code Builder, LabTalk, Origin C, Pyton, X-функції, Automation Server та LabVIEW VI.

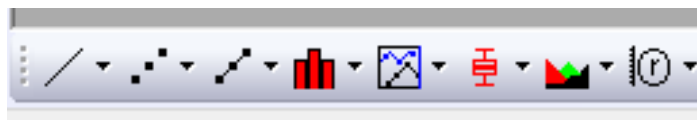
Меню містить набір команд для:

- операцій з активним робочим зошитом;
- ряду загальних операцій, таких як виклик файлу допомоги.

Важливо пам'ятати, що набір операцій в меню залежить від типу активного вікна. Origin також має панелі інструментів (toolbar buttons) для тих команд, які використовуються найбільш часто. Як вже зазначалось, команди можна також виконати з використанням Script Window. Більш того, за допомогою команд можна також звертатись до тих чи інших компонентів меню за допомогою скриптів.

## ***2.2. Можливості Origin для побудови 2D і 3D графіків та кількісного аналізу зображень.***

Наукові публікації містять переважно 2D графіки, які ілюструють залежність між x та y величинами. Такого типу графіки легко створити за допомогою пре-програмованих кнопок на панелі інструментів (Рис. 2.2.1).



**Рис. 2.2.1.** Панель інструментів Origin містить кнопки для швидкого створювання різних типів 2D графіків – зліва направо: лінія, символи, символи з'єднані прямими лініями, колонки, графік, що має дві вісі ординат, коробковий графік або боксплот, площа, графік у полярних координатах.

Крім цих основних типів 2D графіків за допомогою меню Plot-Basic 2D можна обрати і багато інших варіантів графіків цього типу (Рис. 2.2.2).

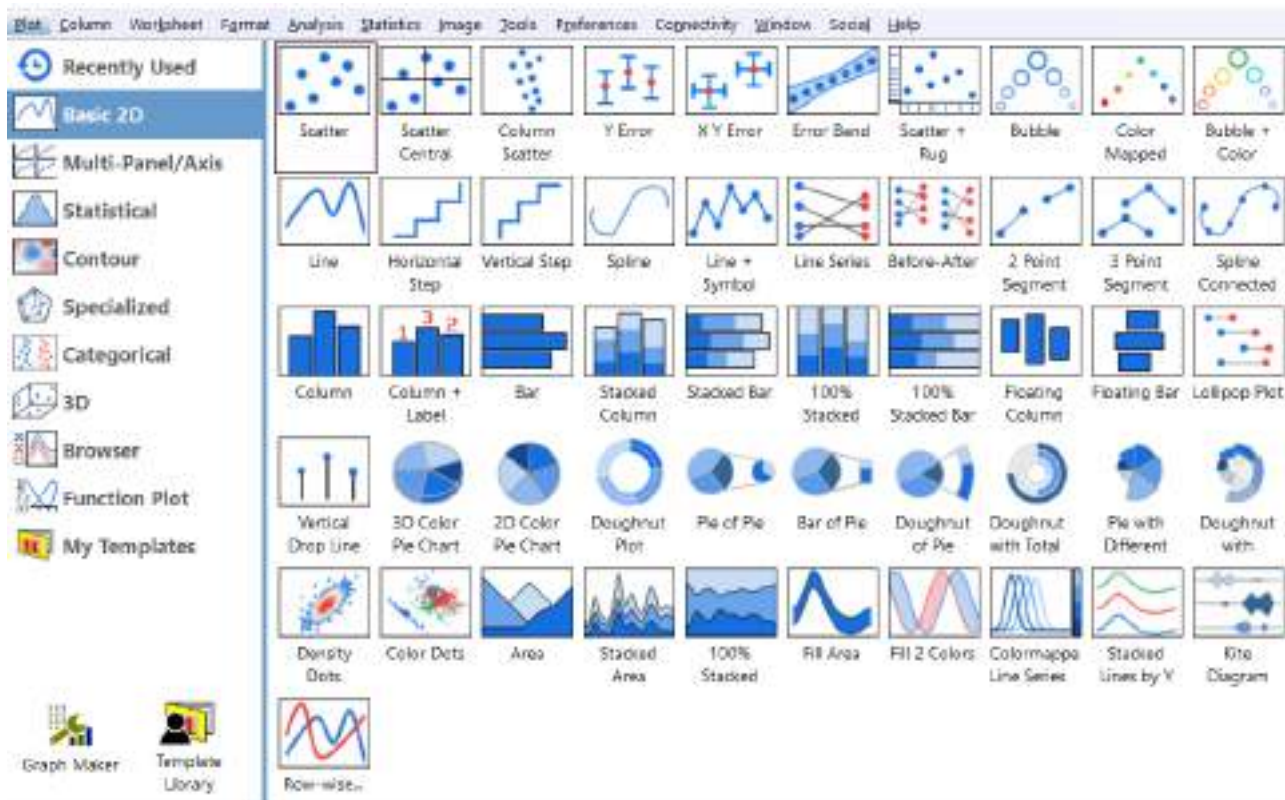


Рис. 2.2.2. Ілюстрація вибору типу 2D графіка.

Так само є багато варіантів представлення даних у 3D вигляді (Рис. 2.2.3). Але при підготовці наукової публікації не варто надто захоплюватись 3D форматом даних, адже «виглядає гарним» - це не дуже переконливий аргумент. 3D графіку дійсно важливо використовувати у тих ситуаціях, коли вона несе додаткову інформацію, яку не можна візуалізувати за допомогою звичайних 2D графіків. Проілюструємо це наступним прикладом. За допомогою методу лазерної конфокальної мікроскопії і з використанням специфічних до

рецепторів холоду і ментолу антитіл ми досліджували субклітинну локалізацію цього білку (TRPM8) у міоцитах судин<sup>85</sup>.



Рис. 2.2.3. Ілюстрація вибору типу 3D графіка.

Оригінальне конфокальне зображення ізольованого міоциту судин наведено на Рис. 2.2.4.

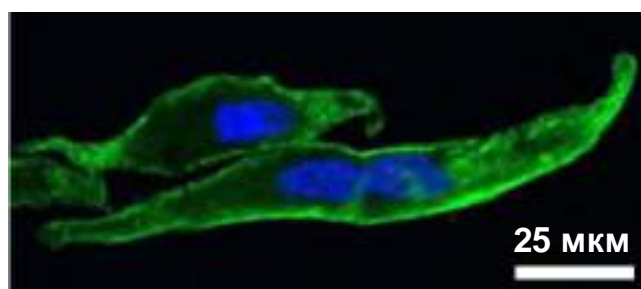


Рис. 2.2.4. Мічені антитілами TRPM8 рецептори показані зеленим кольором.

<sup>85</sup> Johnson C.D., Melanaphy D., Purse A., Stokesberry S.A., Dickson P., Zholos A.V. (2009). Transient receptor potential melastatin 8 channel involvement in the regulation of vascular tone. *Am J Physiol Heart Circ Physiol*. 296(6):H1868-1877.

<https://pubmed.ncbi.nlm.nih.gov/19363131/>

Рецептор холоду – це трансмембранний (інтегральний) білок. Він є одним з трьох клітинних «сенсорів», за відкриття яких була присуджена Нобелівська премія у галузі фізіології та медицини за 2021 р.<sup>86</sup> Видно, що у міоцитах судин цей білок переважно локалізований по периметру клітин, тобто у плазматичній мембрані. Але забарвлення меншої інтенсивності також спостерігається і у цитозолі клітин. Чи можемо ми якимось чином отримати кількісну інформацію щодо їхнього співвідношення? Для цього скористуємось можливостями кількісного аналізу зображень в Origin. Спочатку відкриємо нове вікно матриці (Matrix) та імпортуємо наше зображення за допомогою меню Data > Import from file > Image to matrix. На Рис. 2.2.5 наведені результати імпорту (2 робочі книги справа сторінки) у режимах їх перегляду View > Image mode та View > Data mode.

Далі через меню Plot > Image profiles побудуємо потрібний нам графік (Рис. 2.2.6). На цьому графіку за визначенням є вертикальні і горизонтальні лінії, але також можна провести і довільну лінію натисканням на кнопку «Add Arbitrary Line» (Рис. 2.2.6., панель інструментів вверху). Ця довільна лінія показана зеленим кольором. На основному графіку наведено зображення, а на панелях вверху і справа відповідними кольорами показані результати кількісного аналізу локалізації протеїну, де піки відповідають його мембранній локалізації.

---

<sup>86</sup> Discoveries of receptors for temperature and touch.

<https://www.nobelprize.org/prizes/medicine/2021/advanced-information/>

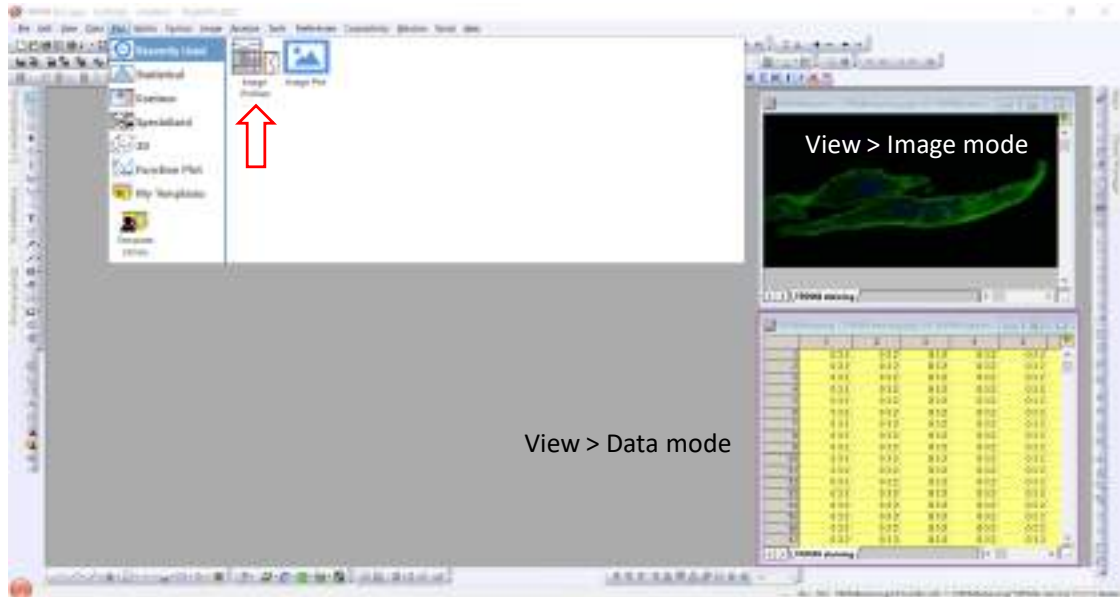


Рис. 2.2.5. Етапи імпорту зображення міоциту.

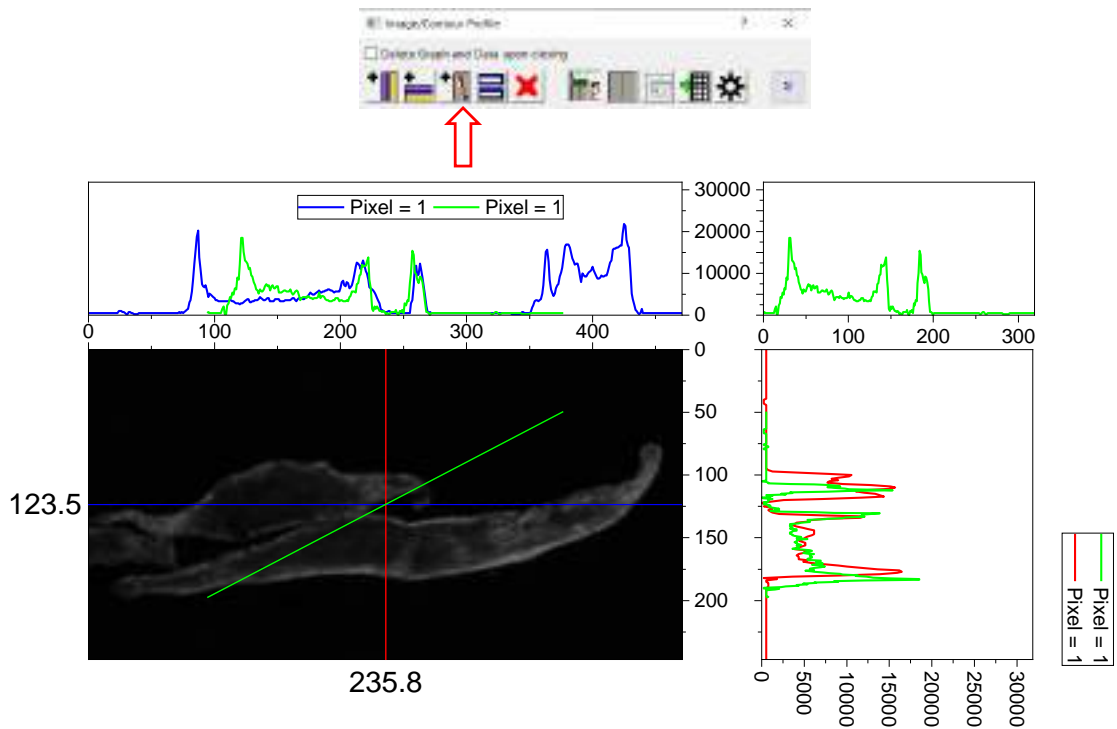
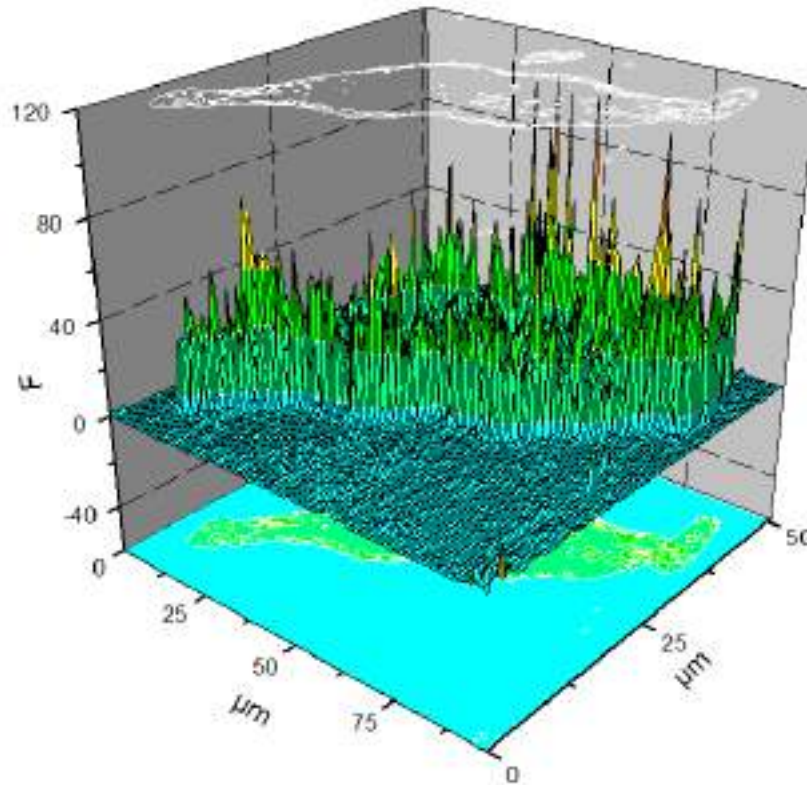


Рис. 2.2.6. Профілі розподілу рецепторів холоду у клітині, які були розраховані уздовж ліній з відповідним кольором.

Проілюструємо на цьому прикладі можливості отримання додаткової інформації за допомогою 3D графіки, як показано на Рис. 2.2.7.



**Рис. 2.2.7.** 3D репрезентація субклітинної локалізації рецепторів холоду у міоциті судин.

Якщо тепер порівняти рисунки 2.2.7 і 2.2.6, то легко переконатись, що у даному випадку 3D графіка є більш доцільною для візуалізації потрібної інформації.

Студенти отримують практичні навички роботи з Origin під час виконання практичної роботи 4.5.

### ***2.3. Статистичний та автоматизований аналіз даних в програмі Origin.***

Аналіз великих масивів даних (згадайте Big Data Biology) може зайняти дуже багато часу, коли доводиться рутинним чином повторювати ті самі операції. В такій ситуації варто задатися питанням – а чи не можна автоматизувати ці процедури і так позбутись рутини, з однієї сторони, а з другої сторони отримати більше часу для творчої наукової діяльності. З точки зору проведення експерименту є і додаткові переваги автоматизації аналізу, адже він дозволяє швидко проаналізувати дані і за потреби ввести корективи у протокол експерименту. Уявіть, якщо цього не робити, проблеми протоколу стали б очевидними лише після завершення експерименту, коли вже пізно щось міняти, а це втрачені ресурси – час дослідника, матеріали, реактиви.

Автоматизація процесів обробки результатів дозволяє збільшити продуктивність, зекономити час, звести до мінімуму втручання користувача (більше об'єктивності, менше тривіальних помилок, які починають виникати по мірі втоми людини).

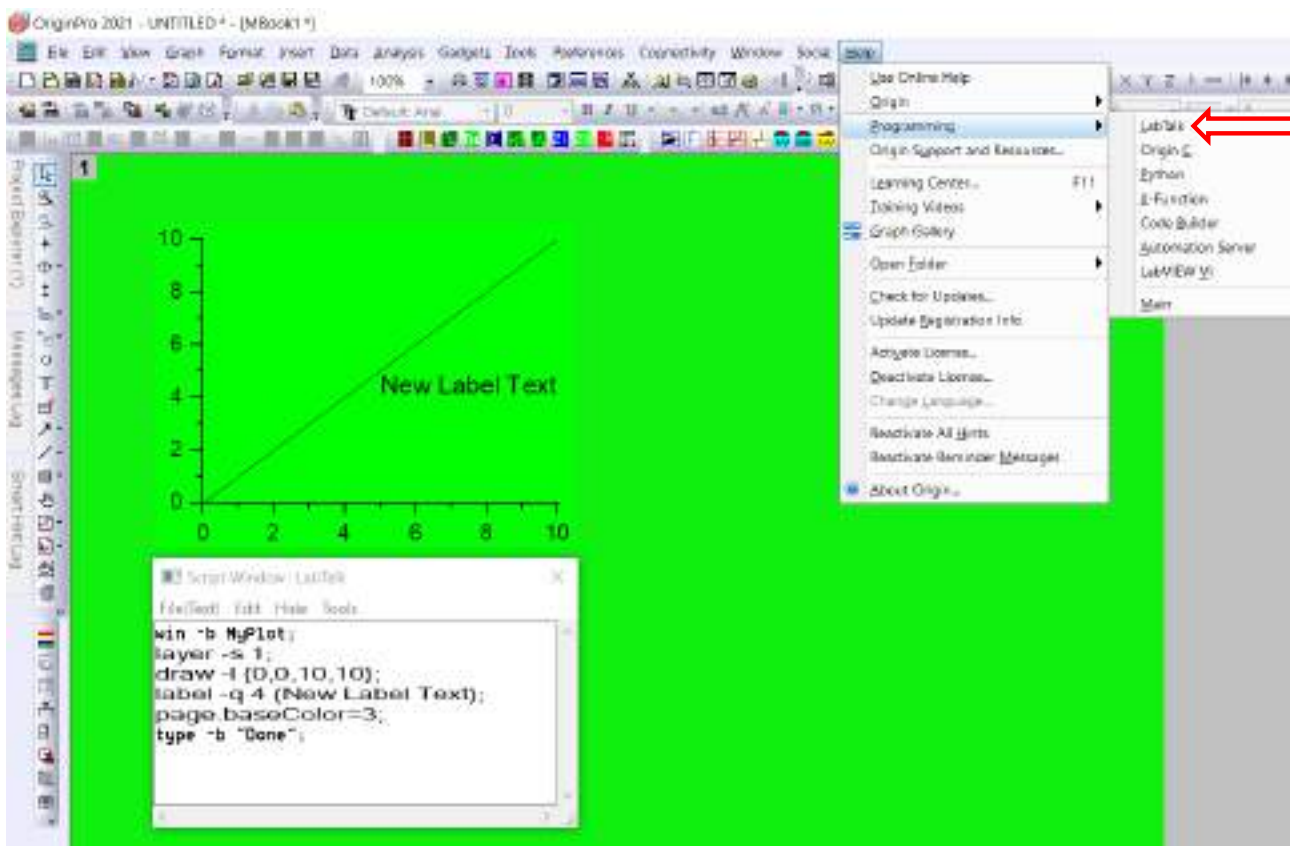
Для вирішення цих задач ми зокрема зосередимося на вбудованій мові сценаріїв Origin, LabTalk™. Це особливо корисно, якщо аналіз даних та/або побудова графіків вимагають великої кількості повторювальних операцій.

Наведемо приклад простої програми в Origin LabTalk. Для її виконання потрібно скопіювати цей код в Script Window (відкривається через меню Window > Script Window), відмітити весь текст і натиснути Enter.

```
win -b MyPlot; // ця команда створює графічне вікно «MyPlot»  
layer -s 1; // ця команда дозволяє вибрати потрібний шар  
draw -l {0,0,10,10}; // ця команда проводить лінію із заданими x-y координатами  
label -q 4 (New Label Text); // ця команда створює ярлик з текстом «New Label  
Text» і розміщує його у четвертому квадранті графіка  
page.baseColor=3; // ця команда задає колір графіка (=3 – зелений)
```

type -b "Done";// ця команда виводить на екран текст «Done» після завершення всіх цих операцій.

В результаті створюється наведений нижче графік (Рис. 2.3.1).

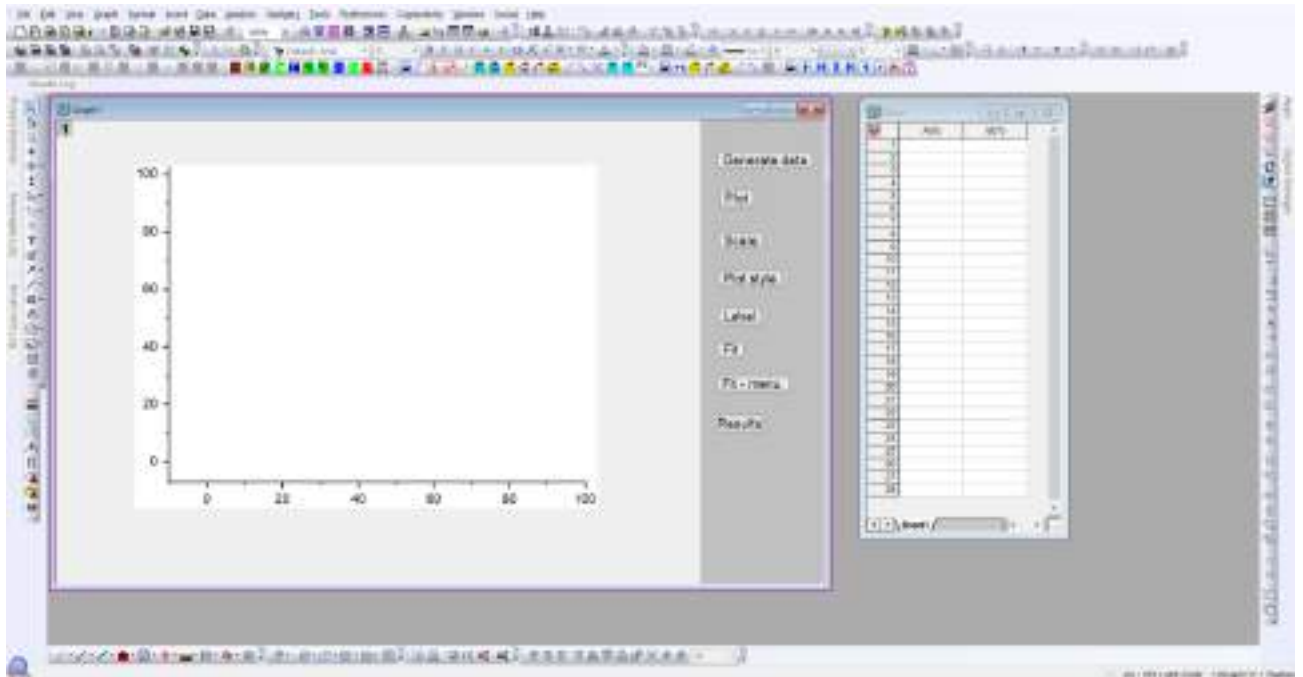


**Рис. 2.3.1.** Скріншот вікна програми Origin, що ілюструє графік, який створюється після виконання скрипту.

Зверніть увагу, що текст після символів // програмою ігнорується. Це зручно для нотаток у процесі написання скрипту.

За допомогою скриптів можна генерувати і фітувати дані, будувати графіки, модифікувати їх елементи, наприклад колір та розмір символів, масштаб вісі, шрифт тексту тощо (Додаток 3). Проілюструємо на простому прикладі ці можливості автоматизації тих чи інших процедур. На Рис. 2.3.2 наведені вікна проєкту Origin, в якому графічне вікно містить ряд кнопок.

Створення скрипту, який виконується при натисканні на кнопку – це є ще одним зручним способом як можна виконати програму.



**Рис. 2.3.2.** Вікна проєкту до виконання скриптів.

Нижче наведені самі скрипти:

*Згенерувати дані:*

```
for (i=1; i<101; i+1) {Data1_A[i]=$i;}  
Data1_B=100*exp(-Data1_A/20);
```

*Побудувати їх графік*

```
win -a Graph1;  
layer -i201 Data1_B;  
layer -at;
```

*Змінити початок шкали X*

```
x1=-10;
```

*Змінити кольори символів і кривих*

```
set data1_B -c 2;  
set data1_B -k 3;  
set data1_B -kf 3;
```

*Додати текст*

```
label -xb (Time (s));  
label -p (-10) (-10) (Signal);
```

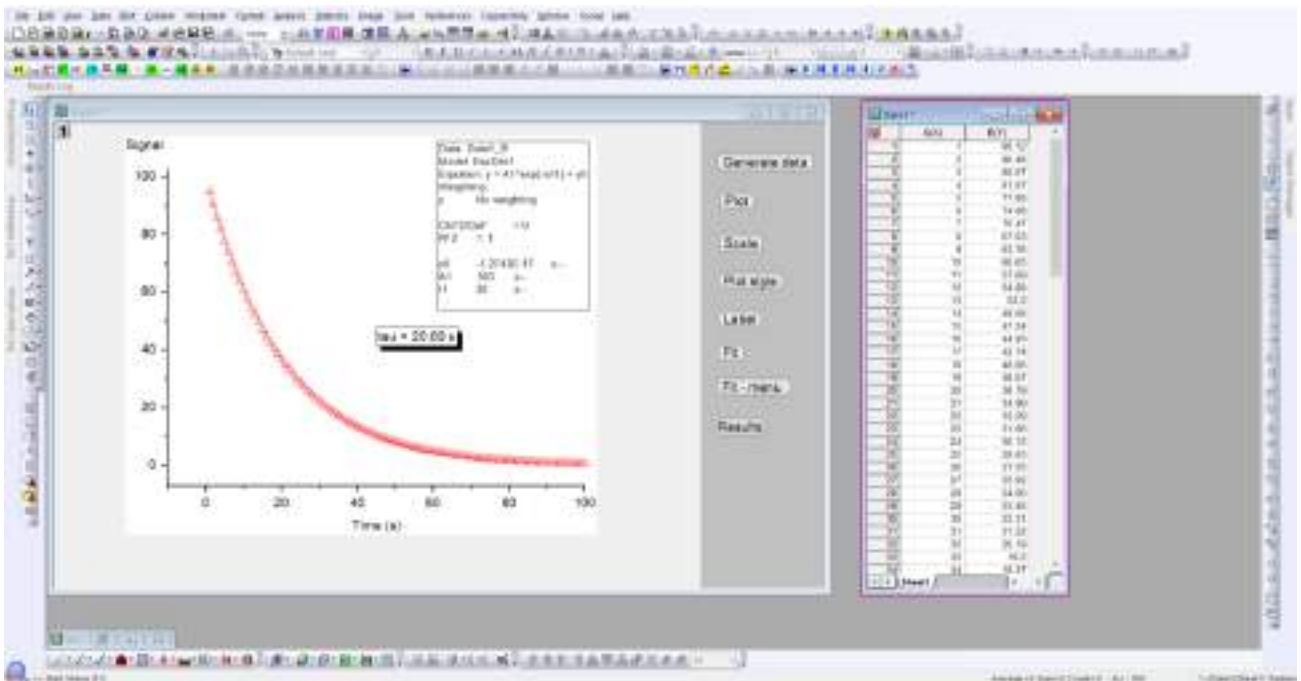
*Апроксимувати дані експоненційною функцією*

```
lsf.func$ = ExpDec1;           //вибір функції ExpDecay  
nlsf.cleanupfitdata();  
nlsf.fitdata1$ = Data1_B;     //вибір даних для фітування  
nlsf.dataBegin=1;           //почати фітування з 1-го рядку  
nlsf.dataEnd=100;          //фітувати дані до 100-го рядку  
nlsf.p1 = 0;                // y0  
nlsf.p2 = 100;              // A1  
nlsf.p3 = 50;                // t1  
nlsf.fit(100);              //виконати 100 ітерацій  
label -sa -q 4 -b 2 -s "tau = $(t1,.2) s";
```

*Виконати команду меню «Fit»*

```
menu -e 34165;
```

В результаті їх виконання отримуємо графік наведений на Рис. 2.3.3. Отже, цей простий приклад ілюструє, як за допомогою декількох рядків скриптів можна ефективно працювати з робочими книгами, графіками, створювати чи модифікувати їх різні елементи, аналізувати і фітувати дані.



**Рис. 2.3.3.** Вікна проєкту після виконання скриптів

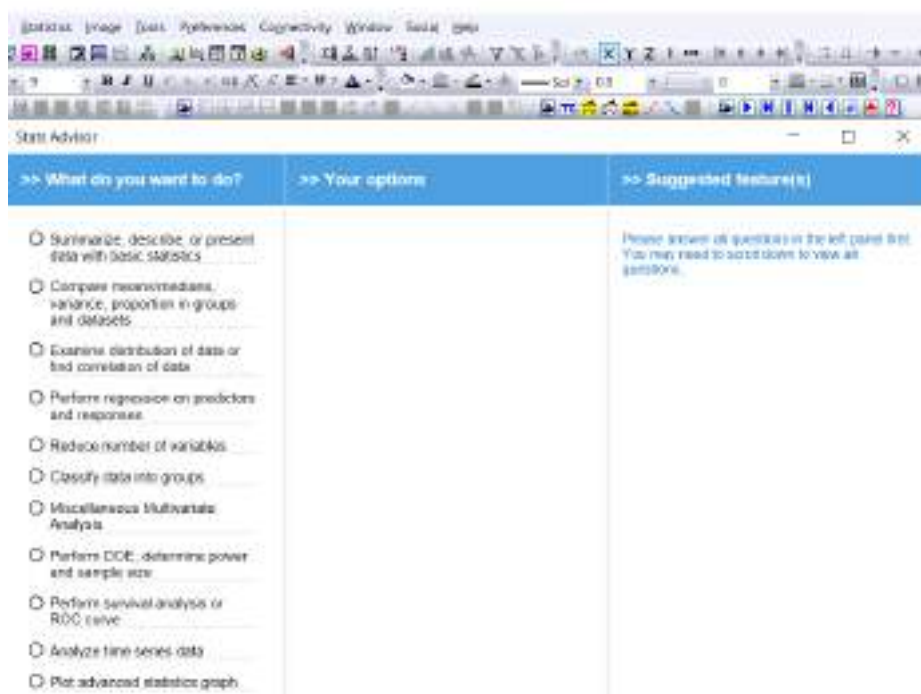
Для автоматизації аналізу даних без застосування програмування є ще дві можливості. Перша – це створення шаблонів аналізу (Analysis Templates), які можна зберегти і використовувати за потреби повторення того чи іншого типу аналізу<sup>87</sup>. Для створення шаблону аналізу потрібно зберегти робочу книгу скориставшись меню «Save Workbook as Analysis Template». У цьому разі зберігається інформація щодо всіх аналітичних процедур, і якщо у таку робочу книгу внести нові дані, то вони будуть автоматично проаналізовані відповідно до заданого шаблону. Друга можливість – це використання обробки даних у пакеті (Batch Processing)<sup>88</sup>. Інструмент пакетної обробки даних дозволяє

<sup>87</sup> Origin Analysis Templates. <https://www.originlab.com/doc/en/Origin-Help/Analysis-Templates>

<sup>88</sup> Origin Batch Processing. <https://www.originlab.com/doc/en/Origin-Help/Batch-Processing>

виконувати повторний аналіз кількох наборів даних за допомогою шаблону аналізу.

Програма Origin має низку можливостей для статистичного аналізу даних, які включають описову статистику, тести гіпотез з однією та двома вибірками, а також дисперсійний аналіз даних (ANOVA). Крім того, Origin містить шаблони різних типів статистичних діаграм, наприклад декілька варіантів гістограм<sup>89</sup>. Відмітимо також, що Origin має вбудований помічник з питань статистики (Stats Advisor), який в залежності від типу даних інтерактивно допомагає вибрати оптимальний підхід до їх статистичної обробки (Рис. 2.3.4).



**Рис. 2.3.4.** Скріншот додатку Stats Advisor (меню Statistics > Stats Advisor).

Студенти знайомляться з питаннями статистичного аналізу даних в Origin під час виконання практичних робіт 4.4 та 4.5.

---

<sup>89</sup> Origin Statistics. <https://www.originlab.com/doc/Origin-Help/Statistics>

## ***2.4. Вступ до мов програмування R та Wolfram Mathematica для біологів. Огляд наявних в них інструментів для візуалізації біологічних даних та їх статистичної обробки. Приклади комп'ютерного моделювання з використанням цих програм.***

Мова програмування R була спочатку створена для статистичних розрахунків (приблизно починаючи з 1995 р.), але з часом вона набула широкого застосування у галузі біоінформатики, зокрема у галузі геноміки, де широко використовуються різні статистичні інструменти аналізу.

Популярності мови програмування R сприяють такі чинники:

- R є відкритим, тобто безкоштовним, кодом. Це зокрема важливо у тих країнах, де коштовне програмне забезпечення малонедоступне. Крім того, це означає, що R активно удосконалюється міжнародною спільнотою, а саме програмне оновлення регулярно оновлюються.
- Широке використання R означає, що в Інтернеті є багато якісних ресурсів, які можна використовувати для вивчення і використання цієї мови програмування.
- R працює у різних операційних системах (Windows, MacOS, Linux) і може опрацьовувати значно ширші формати даних, ніж, наприклад, електронні таблиці MS Excel, з якими ми знайомились у попередніх розділах. Крім того, так само як і у програмі Origin в R є можливість створення сценаріїв аналізу і графічного представлення даних.
- У вільному доступі є велика кількість додаткових програмних пакетів R для наукових досліджень, у тому числі у галузі біології (наприклад, SeqinR).

Відмітимо деякі з основних можливостей і функцій R. Інструмент для аналізу даних (Data Analysis Tool) використовується для статистичного аналізу даних, їх візуалізації та моделювання. Мова програмування R використовується для написання сценаріїв і задання функцій. Об'єкти,

функції та оператори використовуються для обробки та обчислення даних. Варто підкреслити, що навіть складні обчислення базуються на лише декількох рядків коду. Аналізатор статистики (Statistics Analyzer) застосовується для аналізу, моделювання і графічного представлення даних.

Більш детальне ознайомлення з мовою програмування R виходить за межі цього посібника, тим більше, що для цього існує велика кількість доступних і детальних ресурсів. Студентам можна порекомендувати наступні:

- R Crash Course<sup>90</sup>
- Coursera's Computing for Data Analysis course on R<sup>91</sup>
- Solving Biological Problems with R<sup>92</sup>

Студенти також отримають необхідні практичні навички роботи з R під час виконання практичної роботи 4.7.

**Wolfram Mathematica** (інколи спрощено – Mathematica) була розроблена компанією Wolfram Research<sup>93</sup> та використовує Wolfram Language code для чисельних розрахунків (наприклад, рішення систем рівнянь, інтегрування і диференціювання, перетворення Фур'є і Лапласа) та різних аналітичних процедур. Програма має можливості для роботи з графікою. Так само, як і у програмі Origin, є можливості для створення як 2D, так і 3D графіків, зокрема графіків математичних функцій, геометричних фігур тощо.

---

<sup>90</sup> <https://bioinformatics-core-shared-training.github.io/r-crash-course/crash-course.nb.html>

<sup>91</sup> <https://blog.revolutionanalytics.com/2012/12/coursera-videos.html>

<sup>92</sup> <http://cambiotraining.github.io/r-intro/>

<sup>93</sup> <https://www.wolfram.com/mathematica/>

## **РОЗДІЛ 3. ВСТУП ДО БІОІНФОРМАТИКИ ТА ГАЛУЗІ ЇЇ ЗАСТОСУВАННЯ.**

У Вступі ми вже зазначали, що біологічні лабораторії зараз вже поділились на «мокрі» і «сухі». В останніх вчені проводять дослідження переважно з використанням комп'ютерних інформаційних технологій. Необхідність таких досліджень зокрема пов'язана зі стрімким накопиченням великих об'ємів інформації у різних базах даних (пригадайте Big Data Biology), а отже виникає потреба в аналізі і систематизації цих даних.

У цьому модулі студенти знайомляться з такими базами даних (зокрема базами даних геномів, білків, ліпідів, вуглеводів, клітинної сигналізації), а також ресурсами для роботи з ними. Багато з цих питань є складними, наприклад це стосується деталей різних алгоритмів для проведення філогенетичного аналізу чи молекулярного моделювання і докінгу, і виходять за межі нашої дисципліни, але ознайомлення з основними базами біологічної інформації є важливим, адже вони є не тільки науковим, а і значним навчальним ресурсом.

### ***3.1. Основні розділи і галузі застосування біоінформатики. Основні бази даних послідовностей нуклеїнових кислот, білків, сигнальних та метаболічних шляхів.***

#### **Історія розвитку біоінформатики. Роботи Хогевег.**

Біоінформатика загалом поєднує комп'ютерні науки з науками про живу природу (біологія, медицина, фармакологія). Дехто може думати, що біоінформатика як самостійна наукова дисципліна виникла зовсім нещодавно, як результат накопичення великої кількості даних разом з паралельним розвитком комп'ютерних інформаційних технологій для їх інтерпретації. Але сам термін «біоінформатика» вперше був використаний ще у 1977 р., і це було

пов'язано з дослідженнями П. Хогевег та її колег біологів-теоретиків, дослідників складних біологічних систем як динамічних систем, в яких постійно відбувається обробка інформації на різних взаємопов'язаних рівнях<sup>94 95</sup>.

Потужний запит для розвитку комп'ютерного аналізу біологічних даних виник ще раніше у зв'язку з дослідженнями будови білків та інших біологічних макромолекул за допомогою рентгеноструктурного аналізу. Перших значних успіхів у цій галузі досягла британська вчена Дороті Ходжкин. У 1956 р. вона визначила 3D будову молекули вітаміну В12 (Нобелівська премія з хімії 1964 р.), а згодом - будову інсуліну (1967 р.). Для ілюстрації успіхів запровадження комп'ютерних інформаційних технологій у біологічну галузь можна розглянути будову міоглобіну як класичний приклад, який часто згадується у підручниках з біоінформатики. На Рис. 3.1.1 наведена ілюстрація моделі гемоглобіну зробленої з пластиліну, яка зберігається в експозиції музею науки у Лондоні, а також деякі варіанти сучасного графічного представлення цієї молекули за допомогою комп'ютерних технологій. У базі даних білків PDB зараз можливо візуалізувати молекулу на рівні атомів, побачити її з різних кутів зору шляхом обертання, переглянути зображення у різних форматах, таких як cartoon, backbone, space fill, ball and stick, trace, ribbon.

Більше того, зараз розвиваються комп'ютерні технології, які дозволяють передбачити будову білків<sup>96</sup>. Вже було виконано 14 таких проєктів (The Critical Assessment of Protein Structure Prediction - CASP), і у 2022 р. починається 15-й

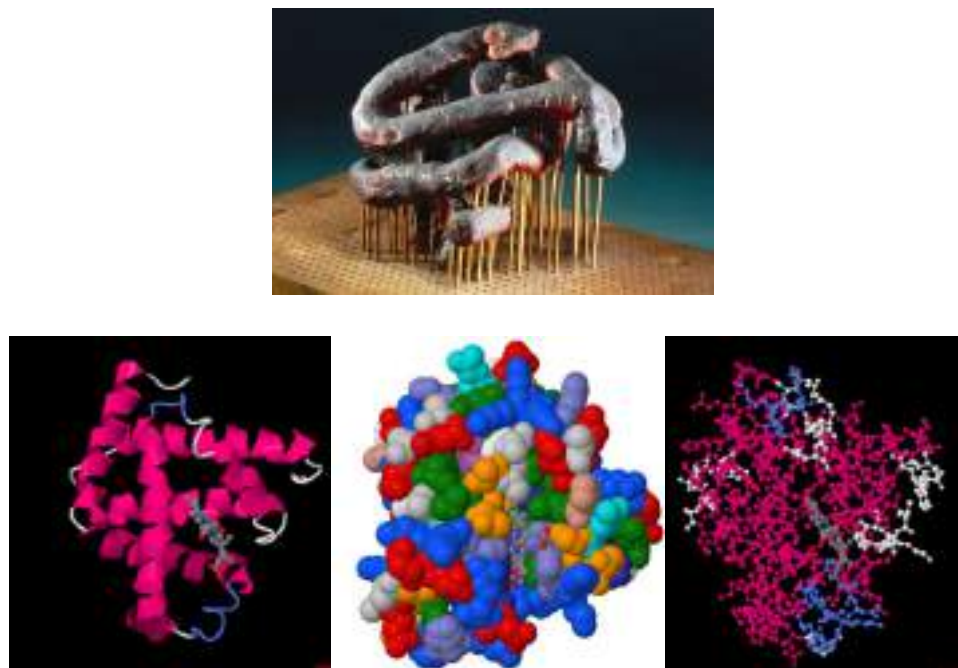
---

<sup>94</sup> Hogeweg P. (1978). Simulating the growth of cellular forms. *Simulation* 31, 90-96.

<sup>95</sup> Hogeweg P., Hesper B. (1978). Interactive instruction on population interactions. *Computational Biology & Medicine* 8:319-327.

<sup>96</sup> Protein Structure Prediction Center. <https://predictioncenter.org/>

проект<sup>97</sup>. У вирішенні цих проблем застосовуються комп'ютерні алгоритми DeepMind. Дослідники вже зробили загальнодоступною систему для аналізу будови білків на основі штучного інтелекту - AlphaFold.



**Рис. 3.1.1.** Вверху - перша 3D модель протеїну (пластилін): міоглобін (John Kendrew, 1957 р.; Нобелівська премія з хімії 1962 р.), яка можна побачити в онлайн колекції Музею науки у Лондоні<sup>98</sup>. Внизу – декілька варіантів представлення структури міоглобіну у базі даних PDB (код молекули 1MBN<sup>99</sup>).

---

<sup>97</sup> AlphaFold: a solution to a 50-year-old grand challenge in biology.

<https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

<sup>98</sup> <https://collection.sciencemuseumgroup.org.uk/objects/co13543/kendrews-original-model-of-the-myoglobin-molecule-molecular-models-proteins>

<sup>99</sup> The stereochemistry of the protein myoglobin. <https://www.rcsb.org/structure/1MBN>

Значним стимулом для запровадження комп'ютерних технологій у галузь біологічних досліджень стали вже згадані у вступі геномна революція, а також розвиток експериментальних методів, які швидко продукують величезний обсяг даних - так звані high throughput технології. Без комп'ютерних технологій обробка таких обсягів інформації була б напевно просто неможливою.

### **Основні розділи і галузі застосування біоінформатики.**

Обговорені вище методи біоінформатичного аналізу вже дають певне уявлення щодо галузей застосування біоінформатичного аналізу. Вони дійсно є дуже широкими. Наведемо деякі з них:

- еволюційні дослідження
- молекулярне моделювання і розробка нових ліків
- біотехнології
- молекулярна, персоналізована, профілактична медицина
- резистентність до антибіотиків
- аналіз геномів і генна терапія
- ветеринарія
- покращення врожайності та багато інших.

Один з важливих аспектів біоінформатики стосується питання – яким чином можна проаналізувати та представити дані з різних інформаційних джерел? П'ять найбільш застосовуваних доменів і форматів даних є такі:

- послідовність
- структура
- взаємодія
- експресія
- функція.

Зараз розвиваються алгоритми для автоматизованого передбачення функції на основі аналізу перших чотирьох з них. Типовою послідовністю проведення біоінформатичного аналізу є наступна. Починаючи з послідовності гену або білку в доступних базах проводиться пошук даних для ідентичної або подібної послідовності. Бази даних містять багато інформації щодо цієї послідовності, структури молекули, а також цитування відповідних наукових публікацій. Крім того, амінокислотну послідовність білку можна також аналізувати щодо наявності у ній певних законсервованих доменів, функція яких добре відома в інших білках. Це дає більш достовірну інформацію щодо функції білку, ніж простий аналіз подібності амінокислотної послідовності шляхом її множинного вирівнювання з іншими послідовностями<sup>100</sup>.

До безпосередніх сфер інтересів біоінформатики відносяться також і чисельні «-оміки» - наукові галузі, в яких вчені зараз намагаються зрозуміти окремі молекули як компоненти складних мереж, їх структуру та взаємодії. Це такі галузі як:

- геноміка (орієнтована на аналіз ДНК)
- протеоміка (орієнтована на аналіз білків)
- транскриптоміка (орієнтована на аналіз РНК)
- ліпідоміка (дослідження ліпідів клітин)
- глікоміка (дослідження вуглеводів)
- систеоміка (системні дослідження)
- інтерактоміка (дослідження взаємодій молекул)
- метаболоміка (макромолекули і шляхи їх перетворень)

---

<sup>100</sup> Kersey P., Arweiler R. (2006). Linking publication, gene and protein data. *Nat Cell Biol.* 8(11):1183-1189. <https://pubmed.ncbi.nlm.nih.gov/17060904/>

- фармакогеноміка (пошук нових молекулярних мішеней для лікування захворювань)

- фізіоміка (дослідження функцій в нормі та їх зміни при патологіях).

Рис. 3.1.2 ілюструє наскільки швидко зараз прогресує секвенування і анотація геномів різних організмів. Після секвенування геному його необхідно анотувати, щоб отримати більш детальну інформацію щодо його структурних особливостей та функціональних елементів, а це неможливо без використання комп'ютерних технологій. Пріоритетність анотацій геномів базується на таких критеріях:

- пріоритети Національних Інститутів охорони здоров'я (НІН, США), і зокрема організми ссавців

- біологічне чи економічне значення

- інтереси чи запити науковців.



Рис. 3.1.2. Кількість анотованих еукаріотичних геномів за даними NCBI<sup>101</sup>.

<sup>101</sup> Eukaryotic Genome Annotation at NCBI.

[https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/)

Згадаємо декілька масштабних біоінформатичних проєктів у цій галузі. Проєкт «1000 геномів», що виконувався протягом 2008-2015 рр., мав на меті секвенування геномів великої кількості людей для розуміння генетичних варіацій у всьому світі. Були знайдені всі основні варіації в геномі людини, які зустрічаються в досліджених популяціях з частотою принаймні 1%. Ці дані стали доступними для вчених шляхом створення відкритих баз даних<sup>102</sup>. Протягом виконання проєкту було створено багато новітніх комп'ютерних інструментів<sup>103</sup>. Проєкт «1000 геномів» зокрема з'ясував властивості та розподіл поширених та рідкісних генетичних змін, процеси, що формують генетичне різноманіття, та сприяв більш глибокому розумінню причин виникнення патофізіологічних станів людини. Ще більш масштабний проєкт «100 тис. геномів» стосувався виникнення ракових та деяких рідкісних захворювань<sup>104</sup>.

Метою проєкту “Геном 10000” було дослідження складної еволюції тваринного світу на основі змін в ДНК та яким чином можна використати цю інформацію для збереження видів<sup>105</sup>. Надзвичайно амбітним є теперішній “Earth BioGenome” та ряд пов'язаних з ним проєктів<sup>106</sup>. В галузі біології рослин зараз виконується масштабний проєкт “10,000 Plant Genomes”.

Не менш вражаючі успіхи є і у галузі розробки лікарських препаратів з використанням методів молекулярного моделювання, молекулярної динаміки і

---

<sup>102</sup> IGSR: The International Genome Sample Resource.

<https://www.internationalgenome.org/>

<sup>103</sup> Software Tools. <https://www.internationalgenome.org/tools>

<sup>104</sup> 100,000 Genomes Project. <https://www.genomicsengland.co.uk/initiatives/100000-genomes-project>

<sup>105</sup> Genome 10K. <http://www.genome10k.org>

<sup>106</sup> Earth BioGenome. <https://www.earthbiogenome.org/>

докінгу. Протягом 2001-2005 р. виконувався проєкт “Геном-5000”, який демонструє роль біоінформатики у розробці ліків. У центрі уваги дослідників були гени, які кодують білки з найбільш високим потенціалом для фармацевтичної індустрії (іонні канали, рецептори, кінази та ін.). Вони дослідили функцію 4650 таких генів (нокаутні моделі, мутації тощо) та визначили понад 100 нових потенційних мішеней для різних терапевтичних застосувань. RCSB PDB наводить ілюстрації основних груп лікарських препаратів разом з їх молекулярними мішенями<sup>107</sup>. Наш прогрес у розумінні молекулярної будови білків ілюструє портал PDB-101 “Молекула місяця”<sup>108</sup>, який можна рекомендувати як актуальний навчальний ресурс, де регулярно з’являється нова інформація щодо наукових успіхів у цій галузі. PDB-101 як навчальний портал для студентів надає повну інформацію про кожен молекулу – вступ стосовно її будови та функцій, пояснення чому молекула важлива як потенційна мішень для лікування захворювань людини, поради де і як користувачі можуть отримати більш детальну інформацію.

Значну роль сучасна біоінформатика і комп’ютерні технології відіграють і у дослідженні модельних організмів. Наприклад, у проєкті OpenWorm<sup>109</sup> була створена повна комп’ютерна симуляція процесів життєдіяльності, таких як рух, на основі симуляції кожної з 959 клітин *C. elegans*, від молекулярного рівня до функцій (Рис. 3.1.3). До виконання цього міжнародного проєкту можна долучатись як вченим, так і студентам.

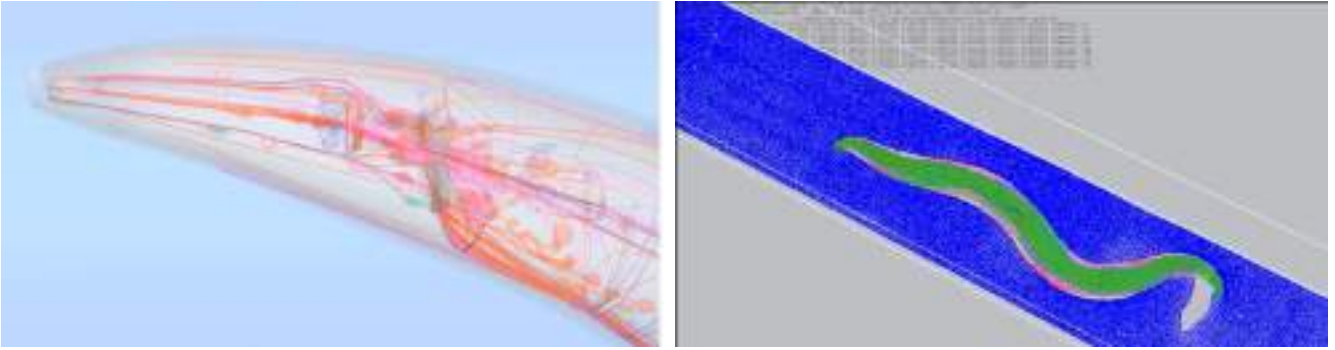
---

<sup>107</sup> How do drugs work: Examples from the PDB archive.

<https://cdn.rcsb.org/pdb101/learn/resources/how-do-drugs-work-flyer.pdf>

<sup>108</sup> PDB-101 Molecule of the Month.. <https://pdb101.rcsb.org/motm/motm-by-date>

<sup>109</sup> OpenWorm. <http://www.openworm.org/>



WormBase – Нервова система

OpenWorm локомоція

**Рис. 3.1.3.** Будова нервової системи, яка показана на рівні окремих нейрональних клітин, та симуляція рухів нематоди в OpenWorm<sup>102</sup>.

В Інституті Сенгера (The Wellcome Trust Sanger Institute, UK) на моделі даніо-реріо з використанням комп'ютерних технологій вчені досліджують мутації, сигнальні мережі клітин і таким чином моделюють захворювання людини. Триває виконання міждисциплінарного проєкту «Фізіом» (Physiom Project<sup>110</sup>). Вже розроблені моделі і симуляції стосуються внутрішньоклітинної кальцієвої динаміки, функції серцево-судинної системи, клітинного циклу, міграції клітин, електрофізіології, гепатології, імунології, транспорту іонів метаболізму, нейробиології, сигнальних мереж клітини та інших питань .

Проєкт Е-клітина (E-Cell project) має на меті дуже амбітну задачу – розробити технології та програмне забезпечення для комп'ютерної симуляції цілої клітини на молекулярному рівні<sup>111</sup>. Це знову таки відкритий міжнародний проєкт, до якого можна за бажанням приєднатись.

---

<sup>110</sup> <https://physiomeproject.org/>

<sup>111</sup> E-Cell project. <https://www.e-cell.org/>

Підкреслимо, що сама можливість взяти участь у виконанні спільного міжнародного проєкту не залишаючи свою лабораторію якнайкраще ілюструє широкі можливості сучасних інформаційних технологій у біології!

### **Основні бази даних послідовностей нуклеїнових кислот, білків, сигнальних та метаболічних шляхів.**

#### ***Знайомство з ресурсами NCBI.***

Національний центр біотехнологічної інформації (NCBI) є частиною Національної медичної бібліотеки США - підрозділу Національних Інститутів охорони здоров'я США. За оцінками Центру, він має приблизно 500 млн. користувачів щодня. Таке визнання зокрема стосується пошукової системи PubMed для роботи з біомедичною літературою, з чим ми вже познайомились у розділі 1.2. Центр надає вільний доступ до великої кількості інформаційних ресурсів у різних галузях (Рис. 3.1.4).



**Рис. 3.1.4.** Перелік основних розділів ресурсів на заголовній сторінці NCBI<sup>112</sup>.

---

<sup>112</sup> NCBI. <https://www.ncbi.nlm.nih.gov/>

Серед них для навчання можна порекомендувати такий ресурс як «Книги» (Books) (Рис. 3.1.5).



**Рис. 3.1.5.** Каталог книг NCBI у результаті пошуку книг з молекулярної біології.

Розділ «Дані і програми» містить бази даних (NCBI C++ Toolkit Manual), різні інструменти (наприклад BLAST Microbial Genomes, CDTree, Cn3D, Conserved Domain Search Service та багато інших). Тут можна завантажити повні геноми організмів<sup>113</sup>, а також програми для аналізу даних, наприклад: BLAST (вирівнювання нуклеотидних та амінокислотних послідовностей) NCBI ToolBox (C/C++ вихідний код для програм NCBI)

<sup>113</sup> <https://www.ncbi.nlm.nih.gov/guide/howto/dwn-genome/>

CD-Tree (перегляд і редагування результатів вирівнювання послідовностей в CD)

Cn3D (для перегляду 3D структур і редагування вирівнювання)

Genome Workbench (програма для перегляду та аналізу даних послідовностей).

Окремо є ресурси стосовно:

- ДНК і РНК (<https://www.ncbi.nlm.nih.gov/guide/dna-rna/>),
- білків (<https://www.ncbi.nlm.nih.gov/guide/proteins/>),
- доменів і структур (<https://www.ncbi.nlm.nih.gov/guide/domains-structures/>),
- генів і експресії (<https://www.ncbi.nlm.nih.gov/guide/genes-expression/>),
- таксономії (<https://www.ncbi.nlm.nih.gov/guide/taxonomy/>) та ін.

Для набуття навичок як фахово користуватись цими ресурсами NCBI пропонує навчальні курси, які включають вебінари, конференції і презентації та онлайн практичні заняття<sup>114</sup>.

Бази даних (БД), з якими знайомляться студенти у цьому розділі – це далеко не всі, які зараз є у загальному доступі. Постійно розробляються нові БД, тому нашою метою є проілюструвати інформаційні ресурси у різних галузях біології з тим, щоб студенти, незалежно від подальшої спеціалізації, знали, що практично у кожній галузі сучасної біології є той чи інший доступний інформаційний ресурс. Ресурси NCBI були вже розглянуті як одні з найбільших у світі. На Рис. 3.1.6 проілюстровано результат пошуку для “antisense RNA” в NCBI. NCBI - геноми<sup>115</sup> містить послідовності і карти даних для цілих геномів більш ніж 1000 організмів, як повністю, так і частково секвенованих. Тут представлені всі домени життя (бактерії, археї і еукаріоти), а також багато вірусів, фагів, плазмід і органел (Рис. 3.1.7).

---

<sup>114</sup> NCBI Learn. <https://www.ncbi.nlm.nih.gov/home/learn/>

<sup>115</sup> NCBI Genome. <http://www.ncbi.nlm.nih.gov/genome/>

## GQuery

NCBI Global/Cross-database Search

Search NCBI databases

antense RNA

Search

### Literature

- 72710 PubMed: scientific & medical abstracts/databases
- 85731 PubMed Central: full-text journal articles
- 113 NLM Catalog: books, journals and more in the NLM Collections

- 10 MeSH: ontology used for PubMed indexing
- 435 Books: books and reports
- 19 Site Search: NCBI web and FTP site index

### Health

- 10000 PubMed Health: clinical effectiveness, adverse and drug reports
- 1 MedGen: medical genetic literature and links
- 10000 GTR: genetic testing registry
- 1 i2b2: genotypic/phenotype interaction studies

- 200 ClinVar: human variations of clinical significance
- 153 OMM: online mendelian inheritance in man
- 14 OMA: online mendelian inheritance in animals

### Organisms

- 10000 Taxonomy: taxonomic classification and nomenclature catalog

### Nucleotide Sequences

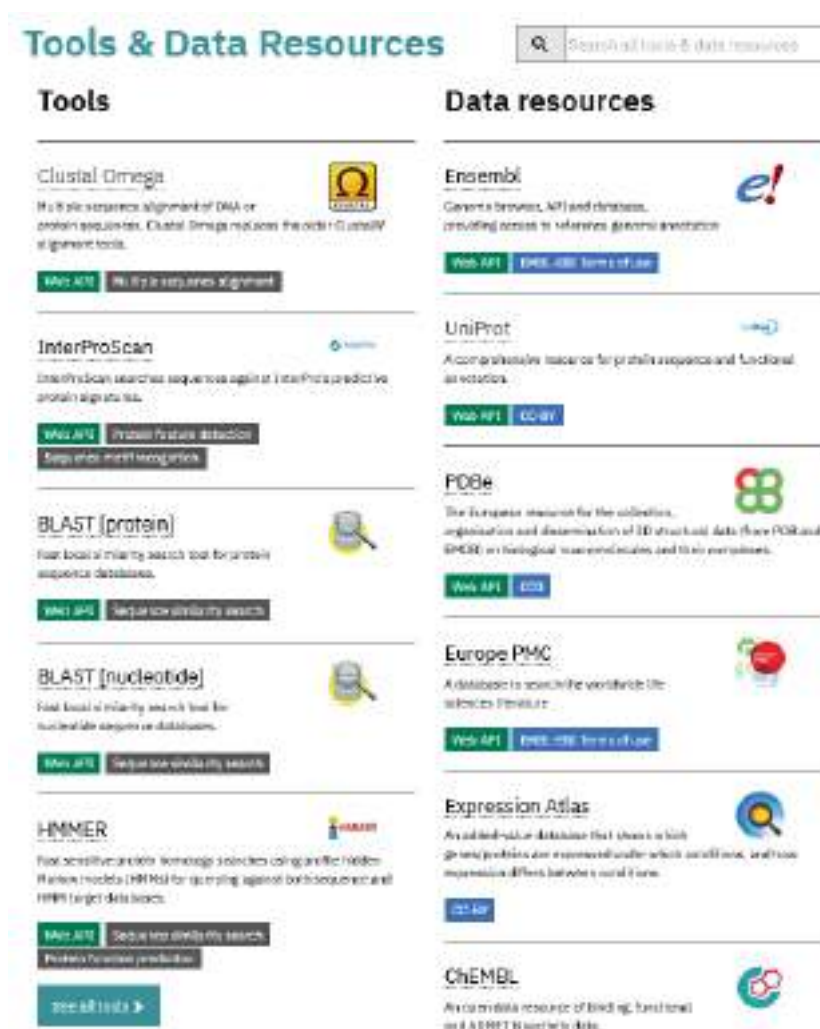
- 3042 Nucleotide: DNA and RNA sequences
- 10000 GSS: genome survey sequences
- 49787 EST: expressed sequence tag sequences

- 11 SRA: high-throughput DNA and RNA sequence read archive
- 4 PopSet: sequence sets from phylogenetic and population studies
- 4010 Probe: sequence-based probes and primers

Рис. 3.1.6. Пошук інформації у БД NCBI.

Рис. 3.1.7. Скріншот заголовної сторінки БД NCBI – геноми.

Сайт Європейського Інституту біоінформатики (EBI<sup>116</sup>) також є дуже значним інформаційним порталом, який містить дані для роботи у багатьох галузях сучасної біології. Це і інформація про нуклеотидні і амінокислотні послідовності разом з інструментами для її множинного вирівнювання, і молекулярні структури разом з інструментами для їх аналізу (Рис. 3.1.8). Вони представляють інтерес особливо для молекулярних біологів.



**Рис. 3.1.8.** Скріншот веб-сторінки, що ілюструє різноманітні інформаційні та аналітичні ресурси EMBL-EBI.

<sup>116</sup> European Bioinformatics Institute. <http://www.ebi.ac.uk/>

Існує велика кількість більш спеціалізованих онлайн БД, зокрема для ДНК, РНК і білків. Одними з найбільш популярних є GENBANK, PDB і UNIPROT. Зараз також збільшується кількість баз даних для окремих організмів. Метою UniProt є надання науковій спільноті найбільш повної, високоякісної та безкоштовної інформації стосовно структури і функції білків<sup>117</sup>. Отриманню практичних навичок роботи з БД UniProt присвячена практична робота 4.6.

ExpASY – це ресурс на порталі SIB (Swiss Institute of Bioinformatics)<sup>118</sup>, який надає доступ до наукових баз даних і програм для їх аналізу в різних галузях біомедичних наук, у тому числі це протеоміка, геноміка, філогенез, системна біологія, популяційна генетика, транскриптоміка тощо. Основними категоріями цих ресурсів є наступні:

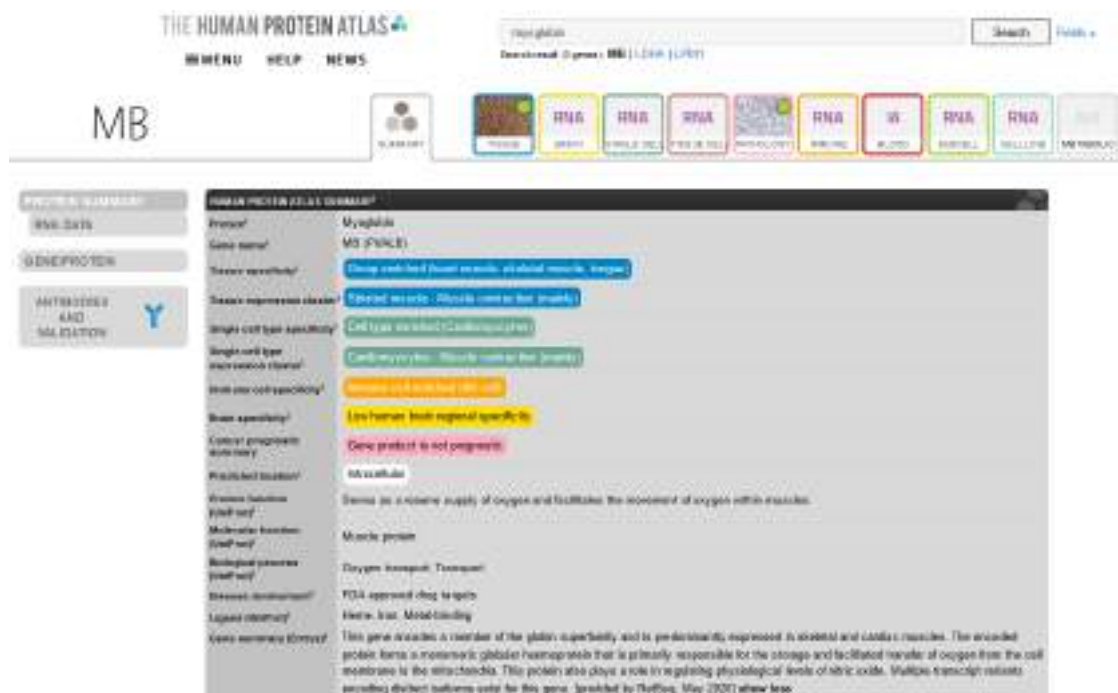
- протеоміка
- геноміка
- структурна біоінформатика
- системна біологія
- філогенез і еволюція
- популяційна генетика
- транскриптоміка
- біофізика
- іміджінг
- інфраструктура інформаційних технологій
- розробка лікарських засобів.

---

<sup>117</sup> UniProt. <http://www.uniprot.org/>

<sup>118</sup> SIB Bioinformatics Resource Portal. <http://expasy.org/>

Банк даних білків (Protein Data Bank, PDB)<sup>119</sup> – це ще один достатньо універсальний інформаційний ресурс, особливо для 3D структур білків, нуклеїнових кислот та інших біологічних макромолекул. Згадаємо також більш спеціалізований Атлас білків людини (The Human Protein Atlas)<sup>120</sup>. У практичній роботі 4.6 досліджуються будова, еволюційні стосунки та видоспецифічні зміни в амінокислотній послідовності міоглобіну. Яку саме інформацію про цей білок можна знайти у цій БД проілюстровано на Рис. 3.1.9.



**Рис. 3.1.9.** Знаходження інформації про міоглобін на порталі «Атлас білків людини». Зверніть увагу, що цей скріншот ілюструє лише загальну інформацію, а більш спеціальну можна отримати натисканням на відповідну вкладку вверху – це експресія, метаболізм тощо.

<sup>119</sup> PDB. <https://www.rcsb.org/>

<sup>120</sup> The Human Protein Atlas. <https://www.proteinatlas.org>

Банк даних нуклеїнових кислот (Nucleic Acid Database, NDB<sup>121</sup>) – це спеціалізований ресурс, особливістю якого на відміну від інших БД є те, що він містить інформацію щодо експериментально підтвердженої просторової будови молекул ДНК та РНК. БД Genomes OnLine Database<sup>122</sup> є онлайн інформаційним ресурсом для проектів з секвенування геномів і метагеномів та пов'язаних з ними даними в усьому світі.

БД KEGG (Kyoto Encyclopedia of Genes and Genomes)<sup>123</sup> створювалась для розуміння біологічних функцій на рівнях вищого порядку, таких як клітина, організм і екосистема. Вона включає інформацію починаючи з молекулярного рівня, зокрема дані, що були отримані в результаті секвенування геномів та за допомогою інших високопродуктивних технологій. Її особливістю є надання детальної інформації щодо сигнальних метаболічних шляхів клітини (Рис. 3.1.10). Цю інформацію доповнює БД Pathguide<sup>124</sup>, яка на сьогодні містить інформацію про понад 700 біологічних процесів та молекулярних взаємодій, зокрема це білок-білкові взаємодії, атлас сигналізації у ракових клітинах, БД гормонів, біохімічних реакцій та багато інших цікавих і корисних ресурсів. Сигнальним мережам на молекулярному рівні присвячені також БД CellCircuits<sup>125</sup>, а питанням молекулярної динаміки - Database of Macromolecular

---

<sup>121</sup> Nucleic Acid Database. <http://ndbserver.rutgers.edu/>

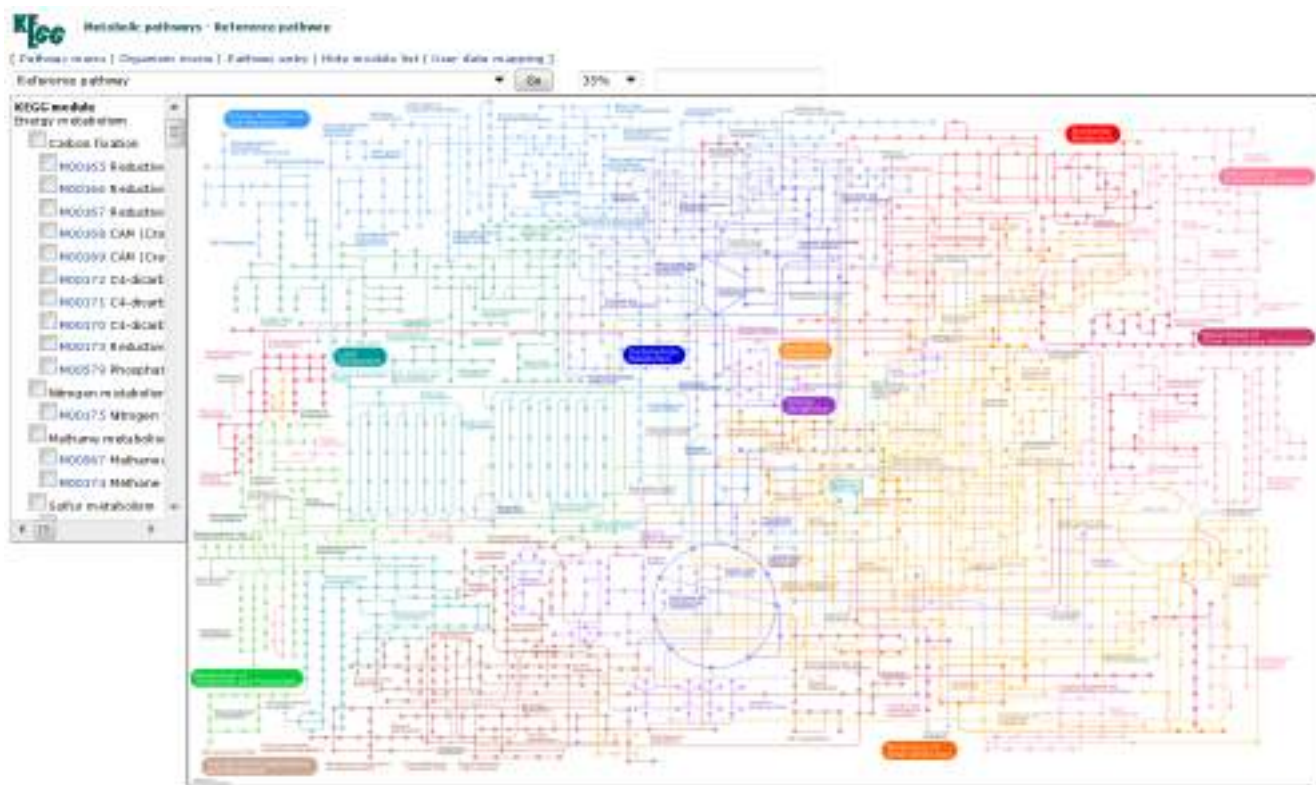
<sup>122</sup> Genomes OnLine Database. <https://gold.jgi.doe.gov/>

<sup>123</sup> Kyoto Encyclopedia of Genes and Genomes. <http://www.genome.jp/kegg/>

<sup>124</sup> Pathguide. <http://www.pathguide.org/>

<sup>125</sup> CellCircuits. [http://www.cellcircuits.org/search/about\\_cell\\_circuits.html](http://www.cellcircuits.org/search/about_cell_circuits.html)

Movements<sup>126</sup>. Цікавою її особливістю є застосування анімацій для більш доступного представлення цієї інформації.



**Рис. 3.1.10.** Скріншот схеми метаболічних шляхів у БД KEGG.

Цей список сучасних інформаційних БД у галузі біології можна було б продовжувати і далі, але як вже зазначалось, надати їх вичерпний перелік було б неможливо. На завершення зробимо висновок, що навіть з цієї обмеженої кількості розглянутих БД стає очевидним, що у кожній галузі сучасної біології зараз знайдеться загальний або спеціалізований інформаційний ресурс.

---

<sup>126</sup> Database of Macromolecular Movements. <http://www.molmovdb.org/>

### **3.2. Основні підходи та алгоритми біоінформатичного аналізу: кластерний аналіз, ланцюги Маркова, штучні нейронні мережі, штучний інтелект, DeepLearning та інші.**

Біотехнологічна корпорація Genscript прокоментувала 100 найбільш цитованих публікацій. Серед них є 6 статей у галузі біоінформатики – роботи, які чи не найкраще ілюструють прогрес цієї галузі. Серед них:

- розробка Basic Local Alignment Search Tool (BLAST)<sup>127 128</sup> для пошуку у базах даних найбільш подібних за будовою молекул (ДНК, РНК та білки) серед тисяч організмів;
- програма Clustal (її сучасна версія відома як ClustalX)<sup>129 130</sup> для вирівнювання множинних послідовностей різних організмів, що дозволяє виявити еволюційні стосунки на молекулярному рівні і навіть

---

<sup>127</sup> Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. J. Basic local alignment search tool . *Mol. Biol.* 215, 403–410 (1990).

<https://pubmed.ncbi.nlm.nih.gov/2231712/>

<sup>128</sup> Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402 (1997).

<https://academic.oup.com/nar/article/25/17/3389/1061651>

<sup>129</sup> Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680 (1994).

<https://pubmed.ncbi.nlm.nih.gov/7984417/>

<sup>130</sup> Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools . *Nucleic Acids Res.* 25, 4876–4882 (1997).

<https://pubmed.ncbi.nlm.nih.gov/9396791/>

передбачити яким чином зміна у певній частині гену чи білку може вплинути на функцію;

- метод для створення філогенетичних дерев відомий у біоінформатиці як метод приєднання сусідів (neighbor-joining method)<sup>131</sup>;
- метод бутстреп (bootstrap)<sup>132</sup> – інструмент статистичного аналізу для оцінки точності різних частин філогенетичного дерева. Бутстреп використовує багаторазову повторну вибірку даних з одного і того самого набору, а потім аналізує варіації в отриманих оцінках для визначення достовірності для окремих філогенетичних гілок. І якщо певна група з'являється в 95% випадків або більше, то це вважається статистично значущим результатом.

Розглянемо принципи роботи цих та інших алгоритмів біоінформатичного аналізу. Пошук у базах даних нуклеотидних чи амінокислотних послідовностей за допомогою звичайних динамічних програмних методів, таких як алгоритм Сміта-Ватермана (Smith–Waterman, 1981 р.), який використовує матриці замін і штрафні бали за пропуски (gaps, тобто вставки або делеції), є надто повільним. Так, наприклад, пошук у базах даних усього 300 тис. послідовностей, якщо запит складає 100 амінокислотних залишків, може зайняти до 3 годин часу. Тому для прискорення цього аналізу були запропоновані евристичні методи, які на відміну від динамічних методів перевіряють лише невеличку частку можливих співпадінь. Такими евристичними методами зараз є BLAST та

---

<sup>131</sup> Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees . *Mol. Biol. Evol.* 4, 406–425 (1987).

<https://pubmed.ncbi.nlm.nih.gov/3447015/>

<sup>132</sup> Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap evolution 39, 783–791 (1985). <https://www.jstor.org/stable/2408678?origin=crossref>

FASTA. Обидва алгоритми знаходять подібні послідовності шляхом ідентифікації їх подібних сегментів (тобто коротких ідентичних або майже ідентичних послідовностей). Якщо порівняти амінокислотні послідовності з реченнями, то ці алгоритми базуються на припущенні, що принаймні одне слово у двох реченнях має бути тим самим. Знайшовши такі співпадіння в окремих словах, довше вирівнювання можна отримати шляхом розширення областей подібності слів. Після виявлення областей із високою подібністю послідовності суміжні області можуть бути об'єднані в повне вирівнювання. Програма BLAST була розроблена в NCBI у 1990 році, і зараз вона є однією з найбільш застосованих для аналізу послідовностей нуклеотидів чи амінокислот.

Для кількісної оцінки подібності сегментів послідовностей використовується скорингова матриця (scoring matrix) BLOSUM62 (BLOCKS SUBstitution Matrix), яка була розроблена шляхом аналізу частот амінокислотних замін у кластерах споріднених білків. Вчені обчислювальним шляхом визначили частоти всіх амінокислотних замін в цих кластерах білків і на основі отриманих результатів побудували матрицю BLOSUM62 (ідентичність амінокислотних залишків у цих блоках становила принаймні 62%). Оцінка BLOSUM62 для кожної конкретної заміни — це логарифм ймовірності біологічно значимої заміни у порівнянні з ймовірністю випадкової заміни.

Для проведення такого аналізу був створений спеціалізований веб портал Національного центру біотехнологічної інформації – NCBI<sup>133</sup>. Пошукові інструменти дозволяють зробити пошук як по базах даних нуклеотидних чи амінокислотних послідовностей (Рис. 3.2.1), так і низку спеціальних пошуків (Рис. 3.2.2).

---

<sup>133</sup> NCBI Basic Local Alignment Search Tool <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

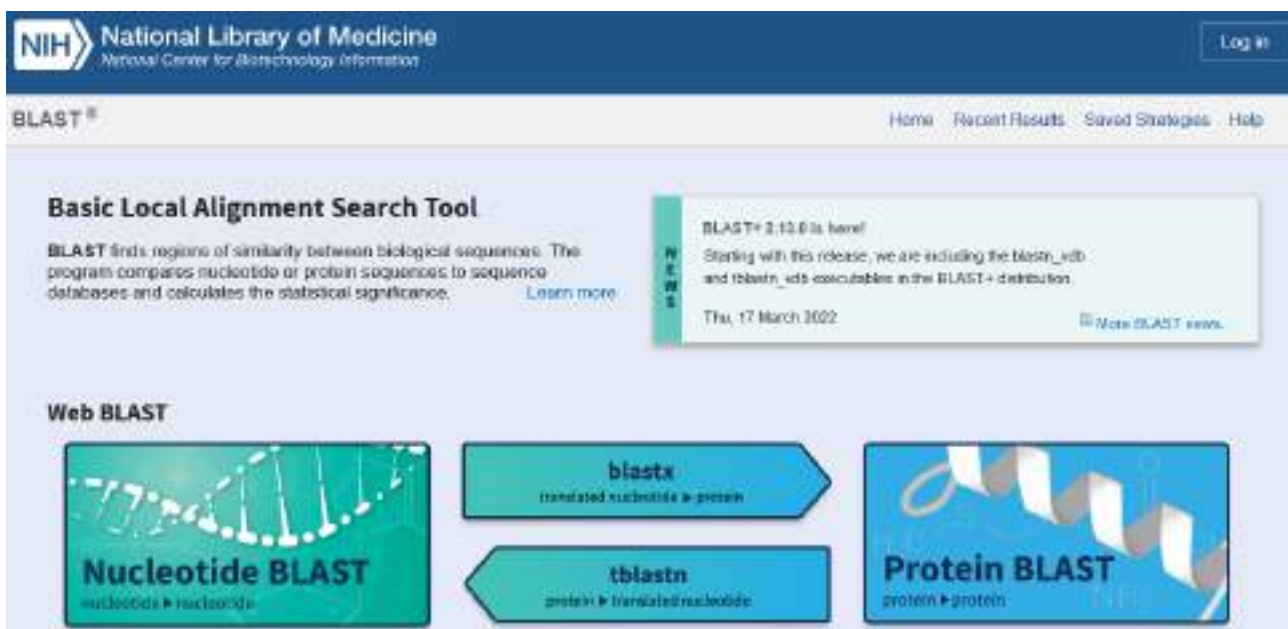


Рис. 3.2.1. Загальні інструменти BLAST на веб сервері NCBI.



Рис. 3.2.2. Спеціалізовані пошукові біоінформатичні ресурси на веб сервері NCBI.

Результатами такого пошуку є список збігів попарної послідовності, які упорядковані за їх статистичною значимістю, що дозволяє розрізняти

еволюційно пов'язані послідовності з непов'язаними. Як правило, відображаються лише ті дані, що перевищують певний поріг – так зване *E*-значення (*E*-value). Це значення надає інформацію щодо ймовірності випадкових збігів.

Інструмент FASTA (FAST All) представлений на веб порталі Європейського інституту біоінформатики<sup>134</sup>, і він також базується на евристичному методі пошуку (Рис. 3.2.3). Це був перший інструмент такого типу, що був розроблений ще до BLAST. Так само як і BLAST, FASTA має декілька підпрограм, і цей інструмент також використовує *E*-значення для статистичних порівнянь. Якщо порівнювати ці інструменти, то BLAST працює дещо швидше і, у залежності від баз даних, може мати кращу специфічність порівняно з FASTA.



**Рис. 3.2.3.** Інструмент FASTA Європейського інституту біоінформатики.

Ці інструменти біоінформатичного аналізу настільки ефективні і прості у використанні, що їх користувачам потрібно лише відкрити відповідну сторінку у веб-браузері, ввести послідовність ДНК, РНК або білка, і майже миттєво

<sup>134</sup> EMBL-EBI Protein Similarity Search <https://www.ebi.ac.uk/Tools/sss/fasta/>

будуть знайдені не тільки подібні послідовності з тисяч організмів, але і надана інформація про функції цих послідовностей разом з посиланнями на відповідні публікації.

Зараз у галузі біоінформатики все більшого значення та популярності набуває кластерний аналіз. Кластерні алгоритми дозволяють ідентифікувати у великих обсягах даних певні дискретні підмножини даних. Крім того, ці методи стають доступними для програмування і вже існуючих комп'ютерних програм (Python Matlab, SPSS та інші). Кластерний аналіз тісно пов'язаний з алгоритмами машинного навчання, метою яких є виявлення, без втручання користувача, підмножин даних, що характеризуються дискретними схожими відмінностями. Кластерний аналіз – це складна статистична процедура, і одна з проблем його використання – це застосування алгоритмів, які були розроблені для одного типу даних, для зовсім іншого типу даних. Серед багатьох різних підходів і методів кластеризації даних відмітимо штучні нейронні мережі, серед яких однією з найбільш поширених є модель Кохонена (self-organizing maps – SOM; Kohonen 1982, 2001). Наприклад, алгоритми SOM були ефективно використані разом з прихованою Марківською моделлю (Hidden Markov Model – НММ) для з'ясування локальних філогенетичних стосунків між вирівняними послідовностями, в результаті чого було доведено, що досить значна частка геному приматів (приблизно 3%) пов'язана з імунною системою<sup>135</sup>. SOM також використовувався для аналізу експресії генів та вирішення інших складних задач.

---

<sup>135</sup> Zamani et al. Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics* 14: 347 (2013). <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-347>

Зараз у цій галузі застосовуються і інші алгоритми аналізу, зокрема це штучний інтелект та DeepLearning. Яскравим прикладом прогресу у цьому напрямку є створення бази даних структури білків AlphaFold, яка була нещодавно розроблена у Європейському інституті біоінформатики EMBL-EBI<sup>136</sup>. Ця база даних надає інформацію про понад 200 млн. білкових структур і є системою, що базується на алгоритмах штучного інтелекту розроблених DeepMind, що дозволяє передбачення 3D структури на основі амінокислотної послідовності.

### ***3.3. Основні програми біоінформатики, їх алгоритми і галузі застосування: вирівнювання послідовностей, філогенетичний аналіз, візуалізація і аналіз молекулярних структур.***

Щодо доступних програм для аналізу даних, то ця галузь так само рясніє їх розмаїттям. Ми знову лише проілюструємо на деяких прикладах основні підходи та галузі застосування алгоритмів цих ресурсів. Слід підкреслити, що алгоритми аналізу можуть бути дуже складними (наприклад аналіз молекулярної будови, взаємодій і динаміки базується на законах термодинаміки, квантової фізики, теорії ймовірності, стохастичної поведінки окремих макромолекул, тощо), але для фахівців-біологів ці програми є цілком доступними у роботі, адже вони мають інструкції для користувача, сайти розробників містять вебінари, онлайн допомогу тощо. Так, SWISS-MODEL<sup>137</sup> – це повністю автоматизований сервер для моделювання білків на основі їх структурної гомології. Цей ресурс доступний через ExPASy веб-сервер або за

---

<sup>136</sup> AlphaFold Protein Structure Database developed by DeepMind and EMBL-EBI.

<https://alphafold.ebi.ac.uk/>

<sup>137</sup> SWISS-MODEL. <http://swissmodel.expasy.org/>

допомогою програми DeepView (Swiss Pdb-Viewer). Мета цього сервера - зробити моделювання білків доступним для біохіміків і молекулярних біологів усього світу.

Крім загальних програм, так само як із базами даних, є дуже спеціалізовані програми. Наприклад, для дослідників іонних каналів є програма для аналізу і візуалізації іонних каналів - ChExVis<sup>138</sup>. Так, нами була проаналізована будова селективного фільтру (це найбільш вузька частина пори іонного каналу, що забезпечує селективний відбір певного типу іона) рецептору холоду і ментолу TRPM8<sup>139</sup>. Дані молекулярного моделювання були нами підтвержені експериментально шляхом точкового мутагенезу білка, що є важливим для молекулярного моделювання – поєднання теоретичних передбачень та експериментальних спостережень.

Значні ресурси для молекулярного моделювання і візуалізації молекулярних структур надає Центр теоретичної і комп'ютерної біофізики Університету Іллінойса (NIH Center for Macromolecular Modeling & Bioinformatics<sup>140</sup>). Він включає багато ресурсів як для наукових досліджень, так і програм для аналізу даних (Рис. 3.3.1). Серед них назвемо такий популярний ресурс як Visual Molecular Dynamics (VMD)<sup>141</sup>. Програма має відкритий код.

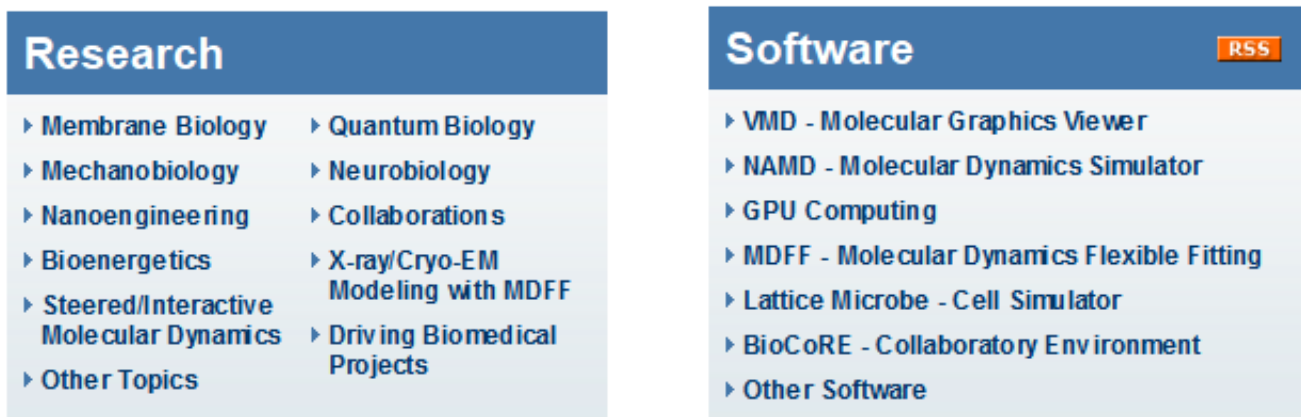
---

<sup>138</sup> ChExVis. <http://vgl.serc.iisc.ernet.in/chexvis/>

<sup>139</sup> Bidaux G., Sgobba M., Lemonnier L., Borowiec A.S. Noyer L., Jovanovic S., Zholos A.V., Haider S. (2015). Functional and modeling studies of the transmembrane region of the TRPM8 channel. *Biophys J.* 109(9):1840-1851.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4643257/>

<sup>140</sup> NIH Center for Macromolecular Modeling & Bioinformatics. <http://www.ks.uiuc.edu/>

<sup>141</sup> Visual Molecular Dynamics. <https://www.ks.uiuc.edu/Research/vmd/>



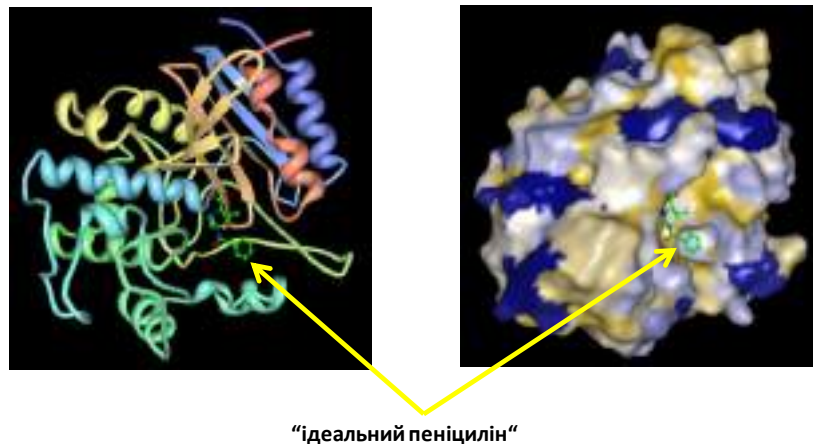
**Рис. 3.3.1.** Скріншот веб-сторінки NIH Center for Macromolecular Modeling & Bioinformatics.

VMD призначений для моделювання, візуалізації та аналізу біологічних макромолекул (білки, нуклеїнові кислоти, ліпідний бішар мембрани). У програмі VMD можна також відкривати стандартні файли БД білків (PDB), вона дозволяє візуалізувати, змінювати кольори, робити анімації молекул, проводити моделювання молекулярної динаміки. Більш спеціалізованим для розрахунків молекулярної динаміки ресурсом є Molecular Dynamics Simulator NAMD<sup>142</sup>. Gromacs<sup>143</sup> теж часто використовується для вирішення таких задач. Автоматизований докінг молекул часто проводиться за допомогою AutoDock<sup>144</sup>. Ці методи набувають все більшого застосування у процесі розробки лікарських препаратів, вони використовуються R&D підрозділами фармацевтичної індустрії. На Рис. 3.3.2 наведено приклад «ідеального» пеніциліну розробленого таким чином.

<sup>142</sup> NAMD - Molecular Dynamics Simulator. <https://www.ks.uiuc.edu/Research/namd/>

<sup>143</sup> Gromacs. <http://www.gromacs.org/>

<sup>144</sup> AutoDockTools. <https://autodock.scripps.edu/>



**Рис. 3.3.2.** Ілюстрація комплексу DD-пептідази з “ідеальним пеніциліном” (зліва) та гідрофобний профіль молекули з БД PDB (1pwc)<sup>145</sup>.

Ми також використовували ці методи для дослідження механізмів взаємодії наноструктурованих матеріалів з різними типами іонних каналів<sup>146 147</sup>.

Комп’ютерні технології відіграють ключову роль у встановленні філогенетичних стосунків від молекулярного рівня<sup>148</sup> і до видів. В якості

---

<sup>145</sup> PDB: penicilloyl acyl enzyme complex of the Streptomyces R61 DD-peptidase with penicillin G. <https://www.rcsb.org/3d-view/jsmol/1pwc/1>

<sup>146</sup> Melnyk M.I., Ivanova I.V., Dryn D.O., Prylutsky Y.I., Hurmach V.V., Platonov M., Al Kury L.T., Ritter U., Soloviev A.I., Zholos A.V. (2019). C<sub>60</sub> fullerenes selectively inhibit BK<sub>Ca</sub> but not K<sub>v</sub> channels in pulmonary artery smooth muscle cells. *Nanomedicine: Nanotechnology, Biology and Medicine* 19, 1-11. <https://pubmed.ncbi.nlm.nih.gov/30981819/>

<sup>147</sup> Al Kury L.T., Papandreou D., Hurmach V.H., Dryn D.O., Melnyk M.I., Platonov M.O., Prylutsky Y.I., Ritter U., Scharff P., Zholos A.V. (2021). Single-walled carbon nanotubes inhibit TRPC4-mediated muscarinic cation current in mouse ileal myocytes. *Nanomaterials* 11(12), 3410. <https://pubmed.ncbi.nlm.nih.gov/34947764/>

ілюстрації наведемо інтерактивну карту генетичної історії людства<sup>149</sup>, створення якої стало можливим завдяки розробці складних статистичних алгоритмів для аналізу ДНК 1490 людей з різних куточків світу<sup>150</sup>. Студенти отримують практичні навички користування методами вирівнювання послідовностей і філогенетичного аналізу при виконанні практичних робіт 4.5-4.7.

На початку нашого знайомства з технічними можливостями сучасних інформаційних технологій ми розглядали як ґрід-технології можуть допомогти інтегрувати обчислювальні можливості, програмне забезпечення і бази даних. У галузі молекулярного моделювання зараз з'явилися нові, безпрецедентні можливості для виконання обчислювальних задач за допомогою сервера Galaxy<sup>151</sup>, на базі якого був створений проєкт Galaxy<sup>152</sup>. Тепер у науковій спільноті з'явилися нові можливості для проведення великих обсягів обчислювань онлайн. Для молекулярного докінгу Galaxy має такі інструменти

---

<sup>148</sup> Dryn D.O., Gryshchenko A.V., Bolton T.B., Zhu M.X., Zholos A.V. (2016). Species-related differences in the properties of receptor-operated TRPC4 channels in intestinal myocytes of rodents. *Neurophysiology* 48 (4), 220-229.

<https://link.springer.com/article/10.1007/s11062-016-9592-8>

<sup>149</sup> Interactive map of human genetic history revealed.

<https://www.ucl.ac.uk/news/2014/feb/interactive-map-human-genetic-history-revealed>

<sup>150</sup> Hellenthal G. et al. (2014). A Genetic Atlas of Human Admixture History. *Science* 343, 747-751. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4209567/>

<sup>151</sup> Galaxy. <https://usegalaxy.eu/>

<sup>152</sup> Проєкт Galaxy. <https://galaxyproject.eu/index-cheminformatics.html>

як AutoDock Vina і rDock, а для пошуку потенційних місць зв'язування лігандів з білковою молекулою – focket. Сервер також надає низку інструментів для молекулярної динаміки з використанням GROMACS. З посиланням на вебсторінку, з якої рекомендовано починати більш детальне ознайомлення з цими ресурсами<sup>153</sup> Galaxy — це наукова платформа для інтеграції даних, збереження, аналізу та публікації даних, яка має на меті *зробити обчислювальну біологію доступною для дослідників, які не мають досвіду програмування.*

---

<sup>153</sup> Introduction to Galaxy Analyses. <https://training.galaxyproject.org/training-material/topics/introduction/>

## РОЗДІЛ 4. ПРАКТИЧНІ РОБОТИ.

### *4.1. Пошук біологічної інформації в Інтернеті. Робота з пошуковою системою PubMed та менеджером цитувань Mendeley.*

**Тема заняття:** Вивчення літератури щодо гладеньких м'язів судин за допомогою PubMed та формування списку цитованих джерел за допомогою програми Mendeley.

**Мета заняття:** засвоєння практичних навичок роботи з PubMed та Mendeley.

**Завдання:** Ви знаєте, що скоротлива активність гладеньком'язового шару судин регулює діаметр судин, а отже тонус судин, динаміку кровопостачання органів, тиск крові та інші фізіологічні процеси, а порушення регуляції тонусу судин призводить до гіпертонічної хвороби. Таким чином, Ви хочете дізнатися більше про гладенькі м'язи судин, механізми регуляції їх скоротливої активності та пов'язані з її порушеннями захворювання. Для цього Вам буде потрібно знайти публікації (до 10, бажано за останні 5-10 років) на тему, яка Вас цікавить, зробити деякі замітки для подальшого їх опрацювання та зробити список літератури.

Перш за все варто буде ознайомитись з усіма категоріями, які проіндексовані в NLM MeSH за посиланням:

<https://www.ncbi.nlm.nih.gov/mesh/?term=vascular+smooth+muscles>

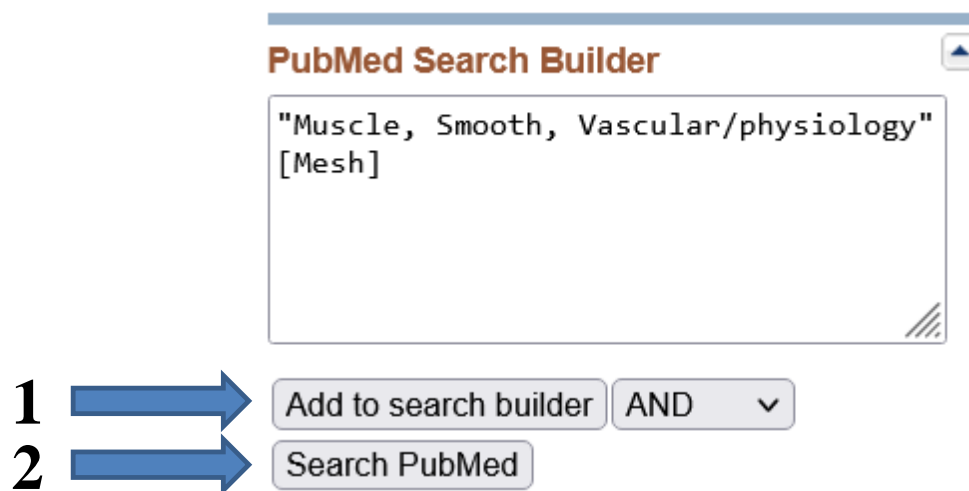


Ці категорії потім за потреби можна використати для розширеного пошуку (PubMed search builder options):

Subheadings:

abnormalities	enzymology	pathology
analysis	etiology	physiology
anatomy and histology	growth and development	physiopathology
blood supply	immunology	radiation effects
chemistry	injuries	surgery
cytology	innervation	therapy
diagnosis	metabolism	transplantation
diagnostic imaging	microbiology	ultrastructure
drug effects	parasitology	virology
embryology		

Наприклад, якщо обрати категорію «фізіологія» і натиснути на кнопку “Add to search builder” (стрілка 1), то буде згенеровано наступний запис для подальшого пошуку у PubMed:



У результаті такого пошуку ми отримаємо наступну інформацію (стрілка 2):

PubMed.gov

"Muscle, Smooth, Vascular/physiology"[Mesh] Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Most recent ↓ Display options ⚙

MP MEDLINE FILTERS

All (35,032)

published last 5 years (2,629)

Hypertension (5,634)

TRP (233)

Diabetes mellitus (1,039)

Protein kinase C (1,662)

35,032 results

Page 1 of 3,504

1 Dexmedetomidine-Induced Aortic Contraction Involves Transactivation of the Epidermal Growth Factor Receptor in Rats.

Cite Lee SH, Kwon SC, Ok SH, Ahn SH, Bae SI, Kim JY, Hwang Y, Park KE, Kim M, Sohn JT. Int J Mol Sci. 2022 Apr 13;23(8):4320. doi: 10.3390/ijms23084320.

Share PMID: 35457136 Free PMC article.

Внизу зліва червоним прямокутником відмічені деякі важливі терміни для подальшого фокусування нашого пошуку наукових публікацій стосовно фізіології гладеньких м'язів судин. Серед них ми знаходимо гіпертонічну хворобу, Са-проникні канали транз'єнтного рецепторного потенціалу (TRP), цукровий діабет (адже гіпертонічна хвороба є поширеним ускладненням діабету) та протеїн кіназу С - ключовий фермент, який регулює багато процесів і зокрема чутливість скоротливого апарату клітини до кальцію. Так, нами було показано, що підвищена активність протеїн кінази С викликає кальцієву сенситизацію гладеньких м'язів судин при діабеті<sup>154</sup>. Отже, ми бачимо, що система пошуку публікацій у PubMed інтер- та проактивно допомагає користувачеві сфокусувати та оптимізувати пошук потрібної інформації.

---

<sup>154</sup> Kizub I.V., Pavlova O.O., Johnson C.D., Soloviev A.I., Zholos A.V. (2010). Rho kinase and protein kinase C involvement in vascular smooth muscle myofilament calcium sensitization in arteries from diabetic rats. *Br J Pharmacol.* 159(8):1724-1731. <https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/j.1476-5381.2010.00666.x>

### Хід роботи:

1. Індивідуальні завдання - оберіть для опрацювання одну з запропонованих тем у наступних розділах з таблиці та проаналізуйте її у поєднанні з «vascular smooth muscle»:

Розділ	Тема
Скоротлива активність	Actin
	Myosin
	Vascular tone
	Calcium
	Protein kinase C
	Sarcoplasmic reticulum
Іонні канали	Calcium channels
	Potassium channels
	Smooth muscle
Захворювання	Diabetes
	Hypertension
	Stroke
	Myocardial infarction
	Chronic kidney disease

2. Проведіть пошук джерел у PubMed у наступній послідовності (проілюстровано на прикладі ролі TRP каналів у регуляції скоротливої

активності гладеньких м'язів судин, як було показано у наших дослідженнях<sup>155</sup>  
156 157):

2. Відкрийте веб-сторінку [PubMed](#) і скористайтеся можливостями розширеного пошуку (PubMed Advanced Search Builder), а саме Query box - (Vascular smooth muscle\*[Title/Abstract]) AND (TRP channel\*[Title/Abstract]). Зверніть увагу, що застосування символу \* (тобто wildcard) дозволяє пошуковій системі шукати не тільки muscle та channel, а також muscles та channels<sup>158</sup>.



---

<sup>155</sup> Zholos A., Johnson C., Burdyga T., Melanaphy D. (2011). TRPM channels in the vasculature. *Adv Exp Med Biol.* 704:707-729. [https://link.springer.com/chapter/10.1007/978-94-007-0265-3\\_37](https://link.springer.com/chapter/10.1007/978-94-007-0265-3_37)

<sup>156</sup> Zholos A.V., Curtis T.M. (2013). TRP channels in vascular disorders. *Curr Top Med Chem.* 13(3):295-309. <https://www.eurekaselect.com/article/50279>

<sup>157</sup> Zholos AV. Studying endogenous TRP channels in visceral and vascular smooth muscles. In: Zhu MX, editor. *TRP Channels*. Boca Raton (FL): CRC Press/Taylor & Francis; 2011. Chapter 9. <https://www.ncbi.nlm.nih.gov/books/NBK92812/>

<sup>158</sup> Skill Kit: Truncation in PubMed. [https://www.nlm.nih.gov/pubs/techbull/jf06/jf06\\_skillkit.html](https://www.nlm.nih.gov/pubs/techbull/jf06/jf06_skillkit.html)

2. Для отриманих посилань застосуємо фільтри, як це було проілюстровано на Рис. 1.2.15. Таким чином ми обираємо оглядові статті, які є у вільному доступі і які були опубліковані протягом останніх 5 років. Ви можете також клацнути курсором миші на “Create custom filter” і додати власний фільтр. Цей скріншот ілюструє результати пошуку:

The screenshot shows a search interface with the following elements:

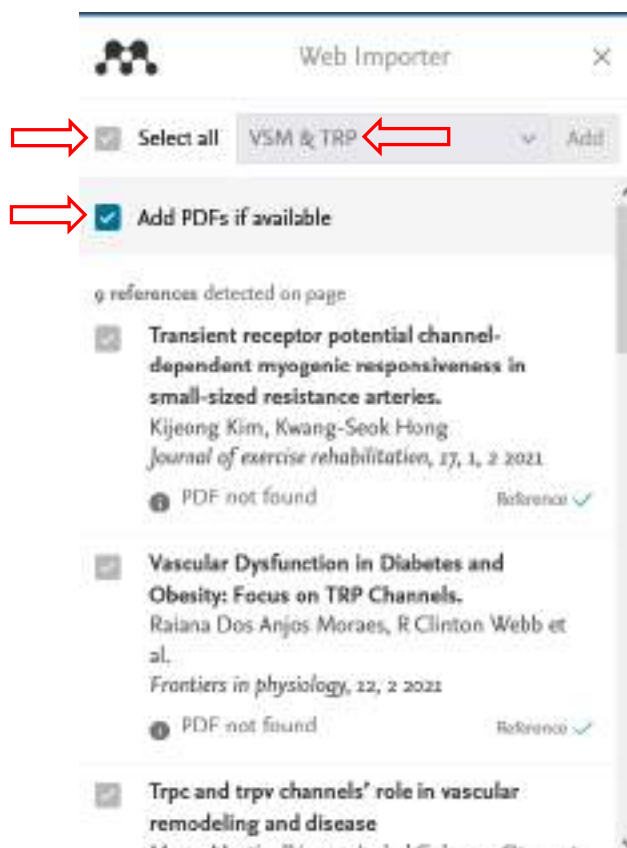
- Search Bar:** Contains the query `(Vascular smooth muscle*[Title/Abstract]) AND (TRP channel*[Title/Abstract])`. A blue **Search** button is to the right.
- Advanced Search:** Links for [Advanced](#), [Create alert](#), and [Create RSS](#). A [User Guide](#) link is also present.
- Actions:** Buttons for [Save](#), [Email](#), and [Send to](#).
- Sorting:** A dropdown menu set to **Sorted by: Most recent**.
- Display Options:** A **Display options** button with a gear icon.
- Results Summary:** Shows **8 results** and a pagination control for **Page 1 of 1**.
- Filters:** A yellow box indicates *Filters applied: Full text, Review, in the last 5 years. Clear all*.
- Result 1:**
  - Transient receptor potential channel-dependent myogenic responsiveness in small-sized resistance arteries.**
  - Cite:** Kim K, Hong KS.
  - Share:** J Exerc Rehabil. 2021 Feb 23;17(1):4-10. doi: 10.12965/jer.2040836.418. eCollection 2021 Feb. PMID: 33728282 [Free PMC article](#). [Review](#).
  - Abstract:** Although the roles of transient receptor potential (TRP) channels in pressure-induced vasoconstriction are not fully understood to date, TRP channels that are directly activated by mechanical stimuli (e.g., stretch of VSMCs) or indirectly evoked by int ...

Зараз у пошук включено 8 публікацій. Тепер можна переглянути їх і вибрати ті, які Ви хотіли б прочитати першими. Виберіть опцію, щоб відображати 50 записів на сторінці, відсортованих за найкращим збігом (Best Match). Поставте прапорці, щоб зробити Ваш вибір статей (до 10 публікацій):

Filters applied: Full text, Review, in the last 5 years. Clear all

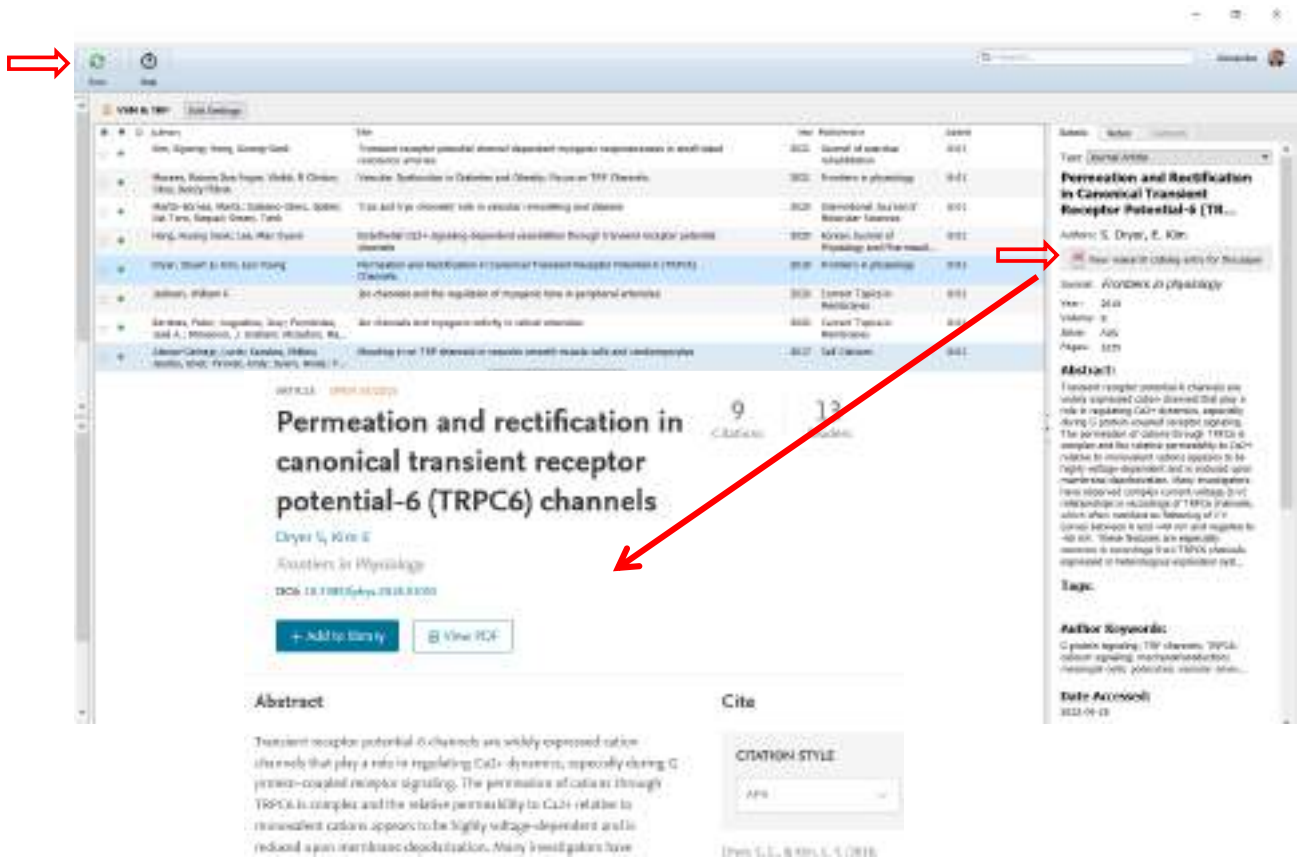
- Transient receptor potential **channel**-dependent myogenic responsiveness in small-sized resistance arteries.**  
Cite Kim K, Hong KS.  
J Exerc Rehabil. 2021 Feb 23;17(1):4-10. doi: 10.12965/jer.2040836.418. eCollection 2021 Feb.  
Share PMID: 33728282 [Free PMC article](#), [Review](#).  
Although the roles of transient receptor potential (**TRP channels**) in pressure-induced vasoconstriction are not fully understood to date, **TRP channels** that are directly activated by mechanical stimuli (e.g., stretch of VSMCs) or indirectly evoked by int ...
- Vascular Dysfunction in Diabetes and Obesity: Focus on **TRP Channels**.**  
Cite Moraes RDA, Webb RC, Silva DF.  
Front Physiol. 2021 Feb 26;12:645109. doi: 10.3389/fphys.2021.645109. eCollection 2021.  
Share PMID: 33716794 [Free PMC article](#), [Review](#).  
In the blood vessels, **TRP channels** are present in endothelial cells, **vascular smooth muscle** cells, perivascular adipose tissue (PVAT) and perivascular sensory nerves, and these **channels** have been implicated in the regulation of **vascula** ...
- TRPC and TRPV **Channels**' Role in **Vascular Remodeling and Disease**.**  
Cite Martin-Bórmex M, Galeano-Otero I, Del Toro R, Smani T.  
Int J Mol Sci. 2020 Aug 25;21(17):6125. doi: 10.3390/ijms21176125.  
Share PMID: 32854408 [Free PMC article](#), [Review](#).  
Transient receptor potentials (TRPs) are non-selective cation **channels** that are widely expressed in **vascular** beds. They contribute to the Ca(2+) influx evoked by a wide spectrum of chemical and physical stimuli, both in endothelial and **vascular smooth** ...
- [Anti-VEGF Therapy for Diabetic Eye Diseases.](#)  
1. Bahrami B, Hong T, Gilles MC, Chang A.  
Asia Pac J Ophthalmol (Phila). 2017 Nov-Dec;6(6):535-545. doi: 10.22608/APO.2017350. Epub 2017 Oct 26. Review.  
PMID: 29076303 [Free Article](#)

3. Імпортуйте ці джерела в Mendeley за допомогою Web Importer plug-in:



Оберіть опції **Select all** та **Add PDFs if available**, а також із розкритого списку Вашу бібліотеку для цих джерел, яку Ви перед цим створили (у наведеному прикладі це (VSM & TRP)).

4. Завантажте програму Mendeley, ввійдіть з Вашим user account email та паролем і оберіть бібліотеку VSM & TRP. Як проілюстровано на рисунку внизу, отримуємо список імпортованих публікацій (рекомендовано його оновлення за допомогою кнопки Sync), а справа наведена повна інформація про вибрану зі списку статтю. За потреби можна відкрити повний текст статті безпосередньо з програми Mendeley, як показано червоними стрілками.

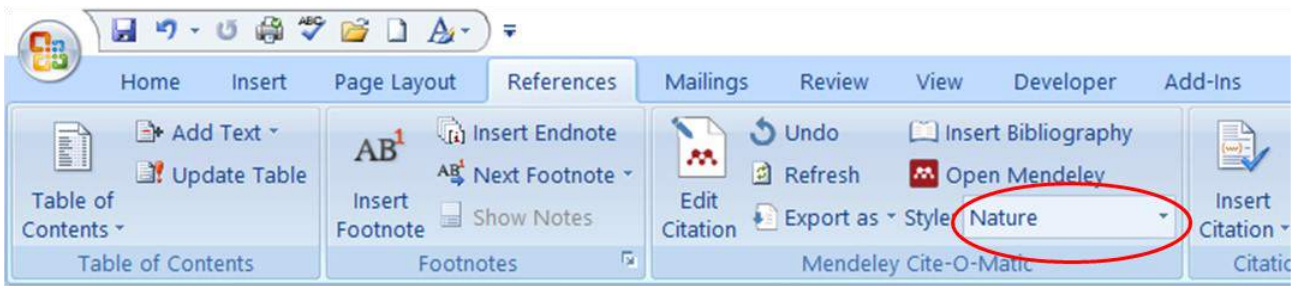


6. Завантажте MS Word, зробіть деякі замітки на цю тему використовуючи реферати оглядів, які Ви щойно знайшли в PubMed, і вставте цитати, де це доречно.

7. Зробіть список використаних джерел за допомогою Mendeley (меню References → Insert Bibliography in MS Word).

8. Спочатку оберіть стиль цитування «Nature», потім «Journal of Physiology» та «Cell Calcium» і переконайтесь, що:

- як формат цитат у тексті так і
- формування списку літератури за допомогою програми Mendeley відбувається автоматично у відповідності до вимог того чи іншого журналу.



9. Збережіть файл для перевірки викладачем.

Приклад оформленої роботи для наведеного вище пошуку літератури наведено у Додатку 1.

## 4.2. Термінологія і функції програми Excel.

**Тема заняття:** Статистичний аналіз та графічне представлення даних за допомогою програми MS Excel.

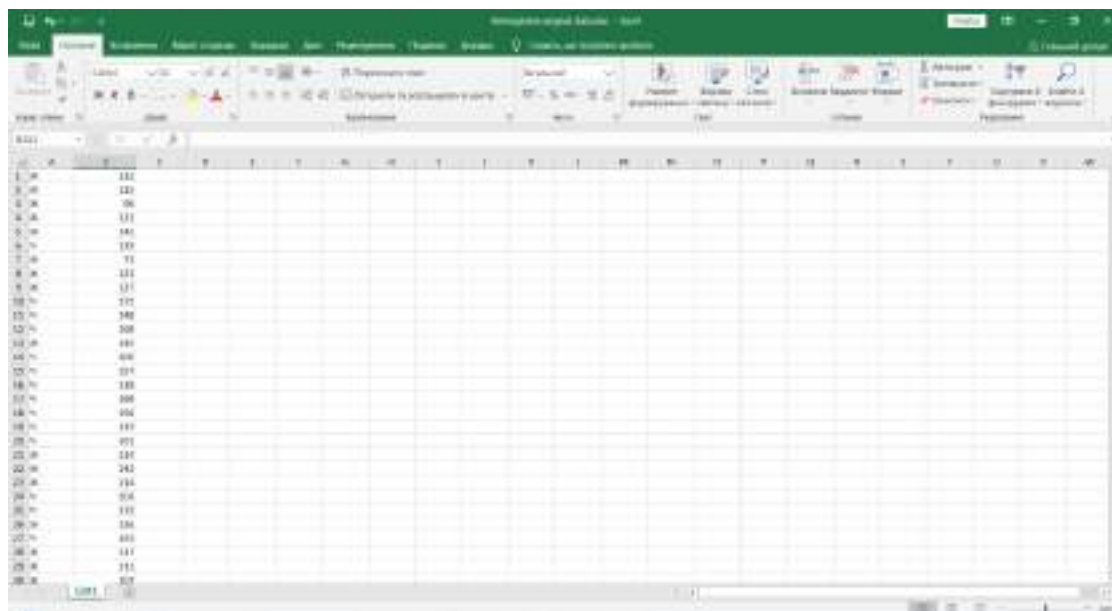
**Мета заняття:** засвоєння практичних навичок роботи з науковими даними за допомогою MS Excel.

**Завдання:** проаналізувати дані лабораторних аналізів концентрації гемоглобіну крові.

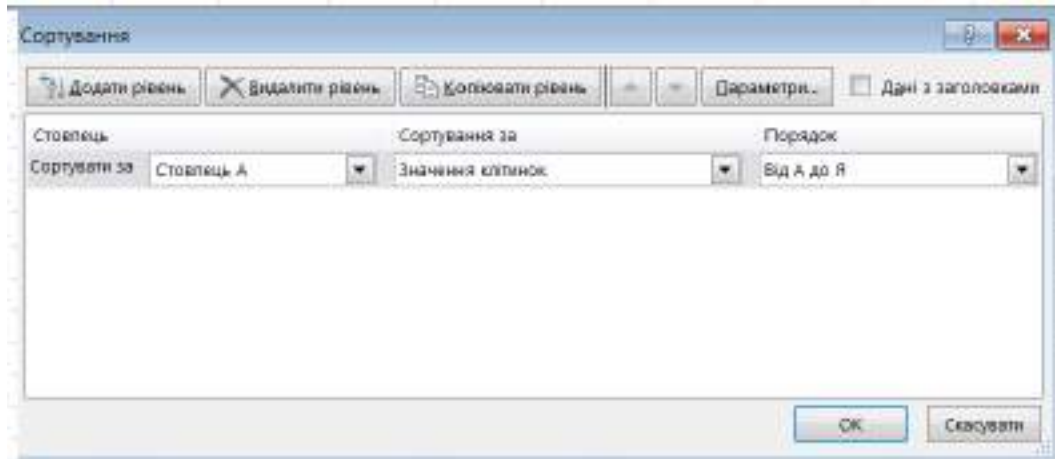
Студенти використовують Excel файл з даними щодо концентрації гемоглобіну у крові 150 пацієнтів, з них 75 – жінки і 75 – чоловіки, який надається їм викладачем. У першій частині роботи студенти виконують аналіз цих даних за допомогою формул MS Excel і будують точкові графіки окремо для цих двох груп. У другій частині вони будують гістограми розподілу даних та виконують статистичний аналіз даних за допомогою додатку «Data analysis» (розділ 1.2).

### Хід роботи:

1. Відкрийте файл з даними

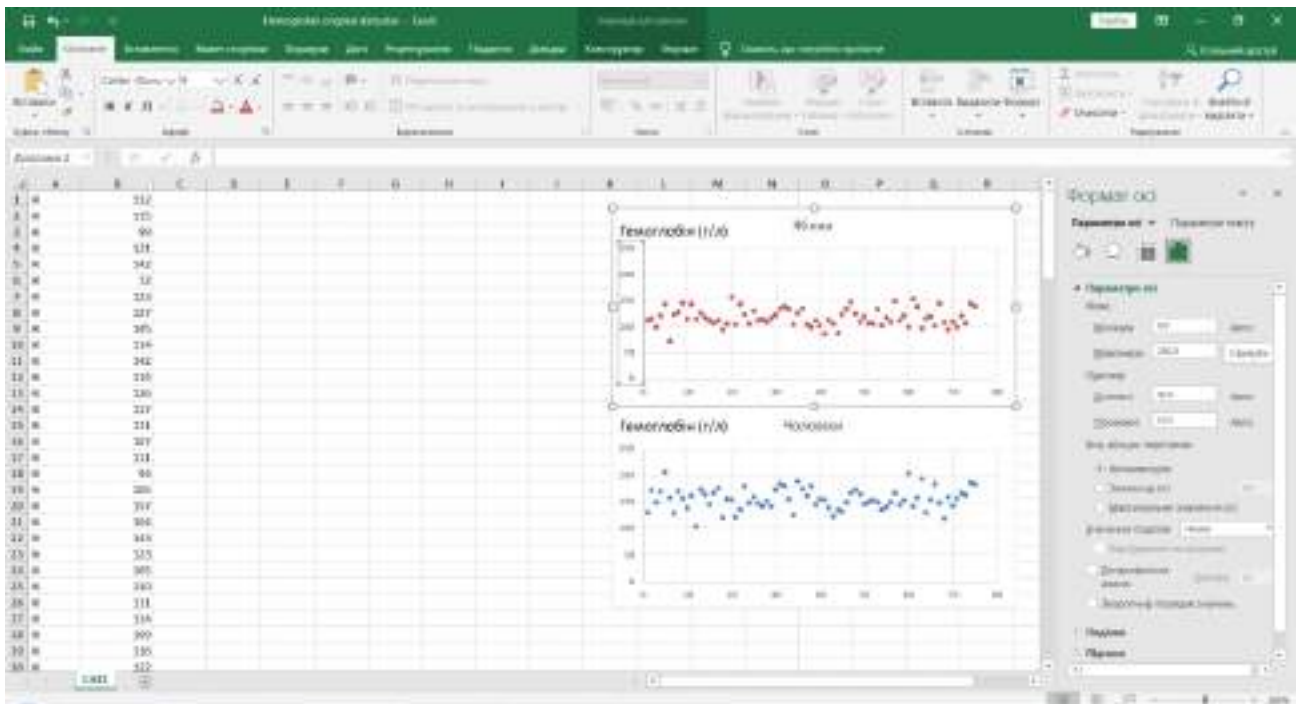


та проведіть їх сортування як показано на наступному рисунку.



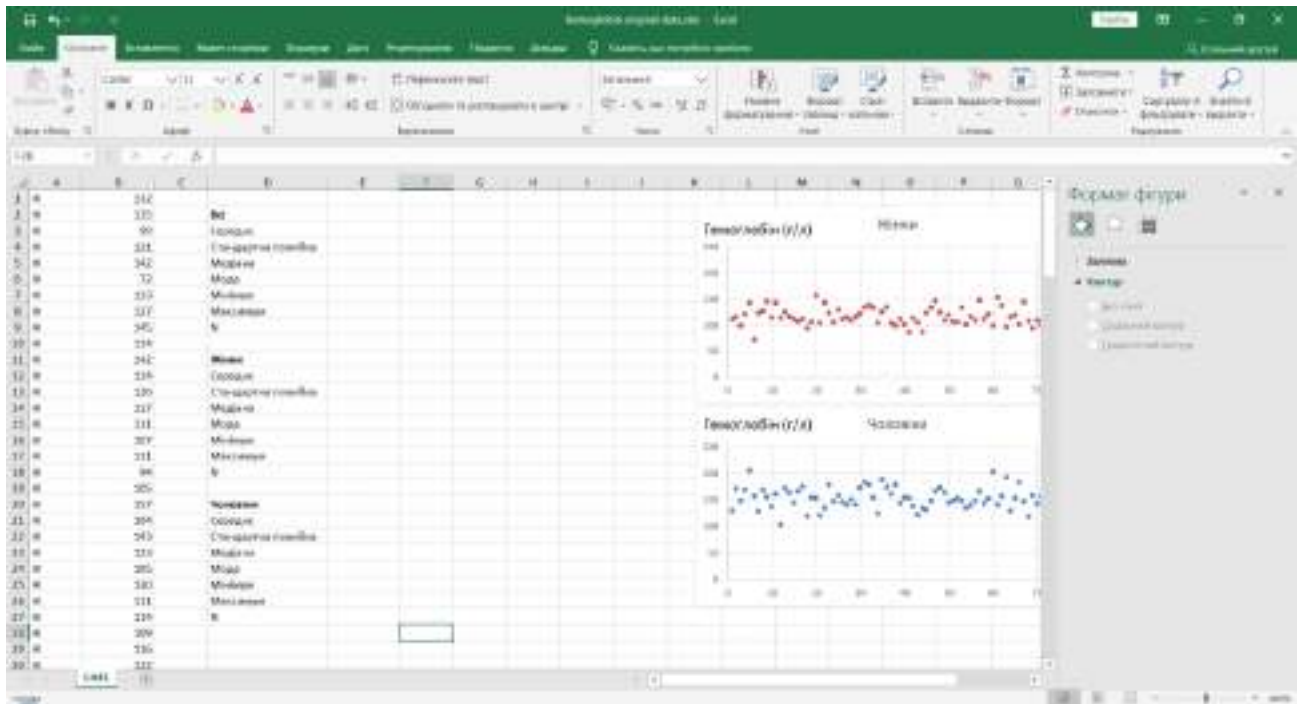
Перевірте, що перші 75 рядків тепер містять дані для жінок, а рядки з 76 по 150 – для чоловіків.

2. Зробіть графіки цих даних у вигляді точкових (X, Y) діаграм. Позначте символи різними кольорами та зробіть однаковий масштаб по вісі ординат для обох графіків:



Зверніть увагу, що значення концентрації гемоглобіну крові є дещо вищими у чоловіків, ніж у жінок. Але чи статистично достовірною є ця різниця? Адже видно, що у деяких жінок ця концентрація є вищою, ніж у деяких чоловіків, і навпаки. Проаналізуємо це питання.

3. Запишіть текст у комірці колонки D як показано на рисунку:



4. У комірках стовпчика C введіть відповідні формули для розрахунків цих параметрів, а саме:

Average

Standard Deviation

Median

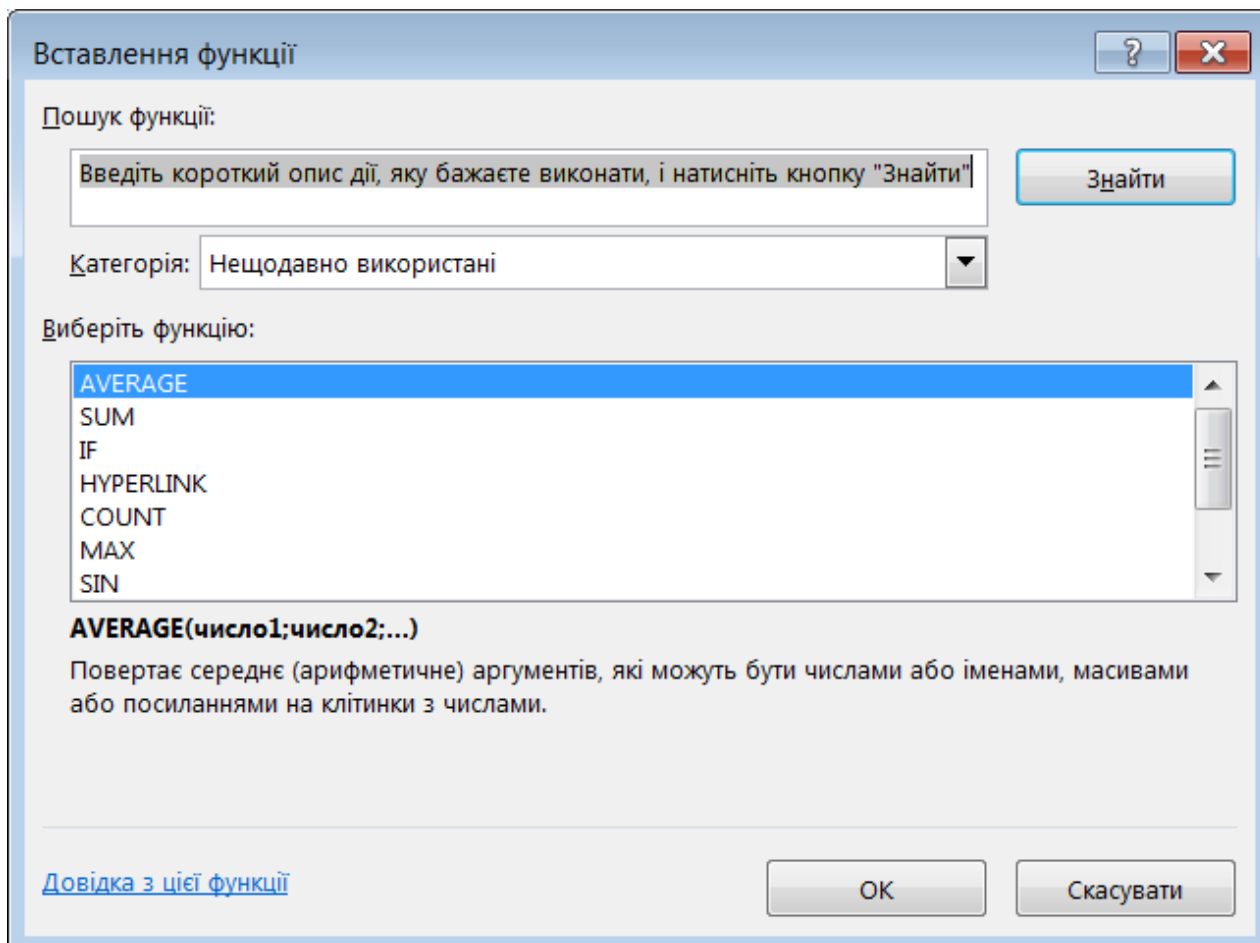
Mode

Minimum

Maximum

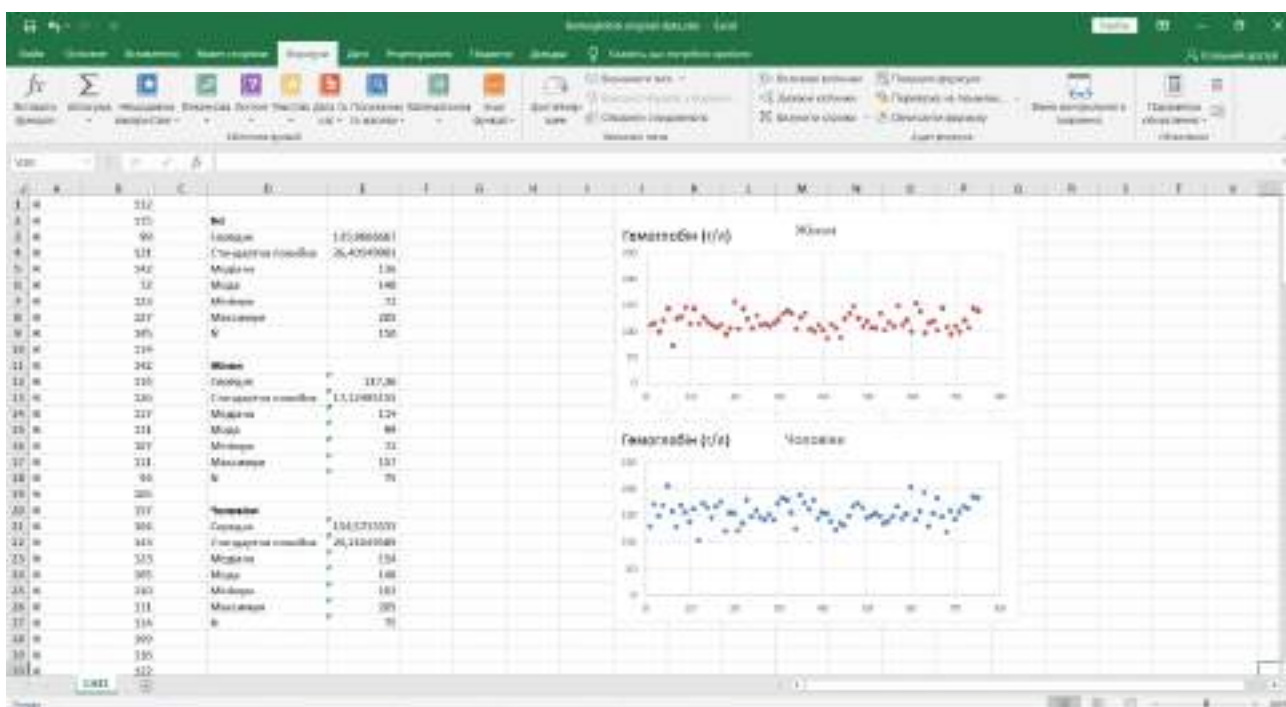
Count

Наприклад, для розрахунку середнього значення концентрації гемоглобіну у крові для всіх пацієнтів розмістіть курсор у комірці E3 та виконайте меню «Формули – Вставити функцію - Average».



Діапазоном для розрахунків у цьому прикладі буде B1:B150. Далі розмістіть курсор у комірці E13 і повторіть розрахунки середнього для жінок (діапазон B1:B75), потім у комірці E20 – для чоловіків (діапазон B76:B150).

Аналогічним чином зробіть розрахунки для всіх вищезазначених параметрів з використанням функцій STDEV, MEDIAN, MODE.SNGL, MIN, MAX, COUNT, відповідно.



Отже, концентарція гемоглобіну крові у середньому становила  $117,4 \pm 17,1$  г/л у жінок ( $n=75$ ) та  $154,6 \pm 20,2$  г/л у чоловіків ( $n=75$ ) ( $\text{Mean} \pm \text{SD}$ ).

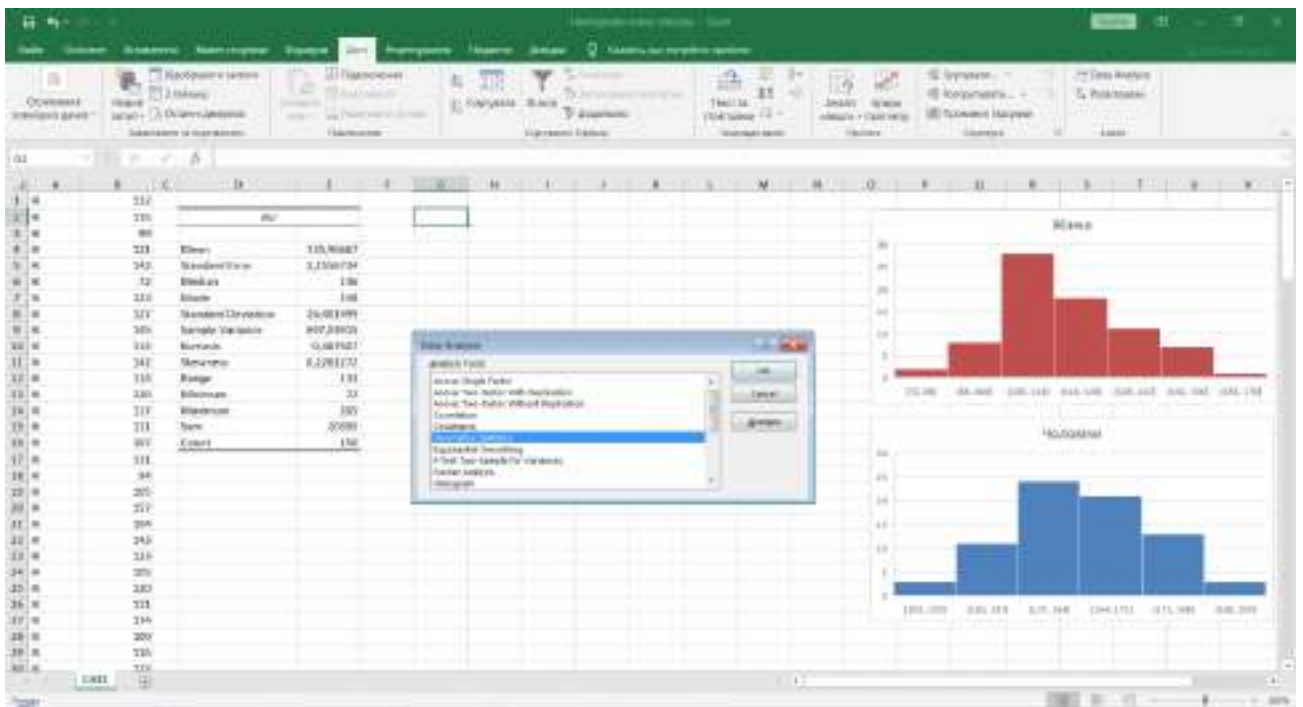
5. Тепер ми можемо відповісти на поставлене запитання щодо статистичної значущості цієї різниці у середніх величинах<sup>159</sup>. Для цього у комірці D29 запишіть  $P=$  , потім поставте курсор у E29, клацніть курсором мишки по іконці функцій і оберіть TTEST() у категорії Statistical. Далі введіть:

- Array 1 – B1:B75 (для жінок)
- Array 2 – B76:B150 (для чоловіків)
- Боки – 2 (двобічний розподіл)
- Type – 2 (двопарний  $t$ -тест із рівною дисперсією)

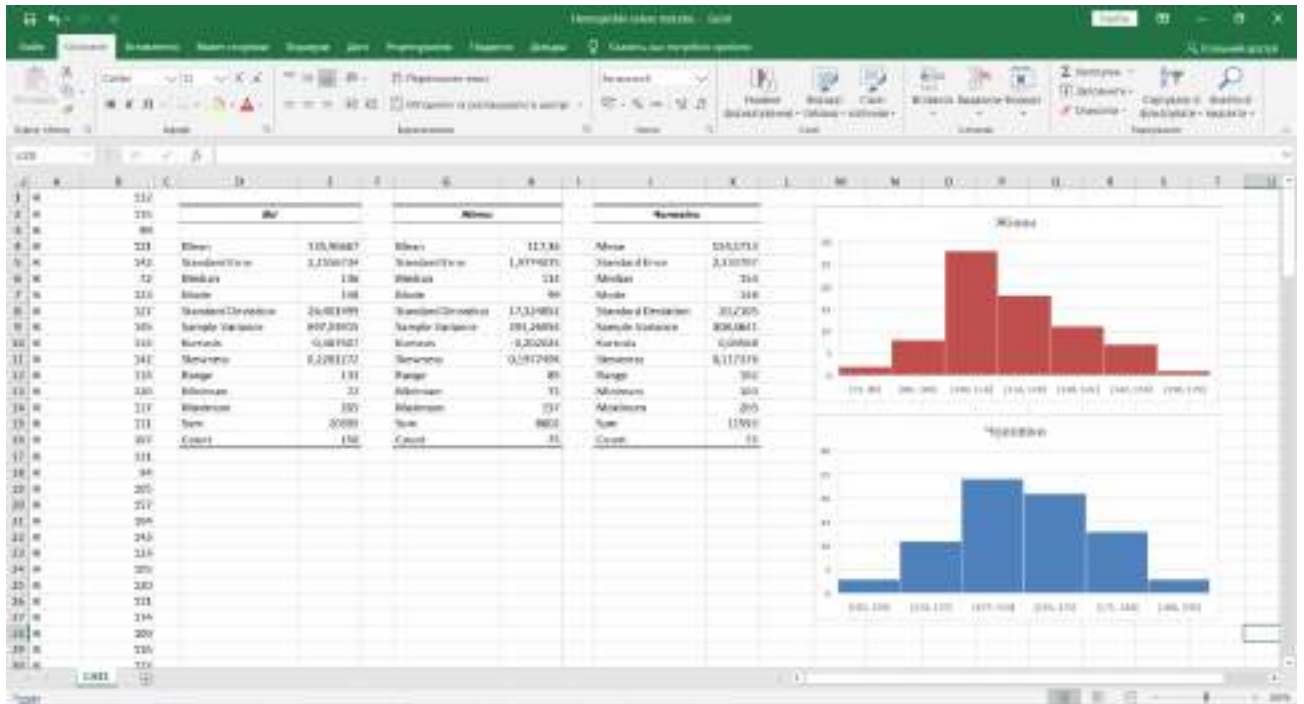
і натисніть кнопку Ok для виконання цих розрахунків.

<sup>159</sup> Microsoft Support: T.TEST function. <https://support.microsoft.com/en-us/office/t-test-function-d4e08ec3-c545-485f-962e-276f7cbcd055>





\$B\$1:\$B\$75) та окремо для чоловіків (діапазон \$B\$76:\$B\$150), отримуємо остаточний результат статистичного аналізу:



Зверніть увагу, що статистичні дані включають не тільки стандартне відхилення, а і стандартну похибку середнього (у науковій літературі SEM використовується дещо частіше), а також Range – це різниця між Max та Min величинами.

### 4.3. Графічне представлення даних за допомогою програм CorelDraw, Inkscape і BioDraw.

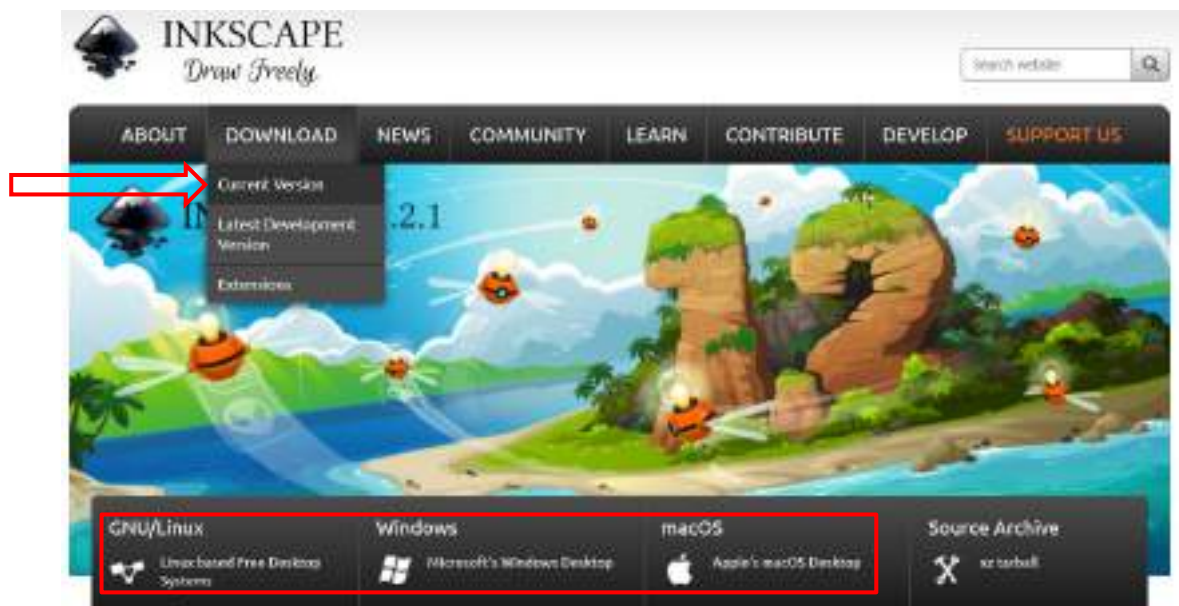
**Тема роботи:** застосування інструментів векторної графіки.

**Мета роботи:** ознайомлення студентів з інструментами векторної графіки за допомогою програми Inkscape, яка знаходиться у вільному доступі.

#### Хід роботи

1. Завантажити та інсталиувати Inkscape для потрібної Вам операційної системи

<https://inkscape.org/release/inkscape-1.2.1/>

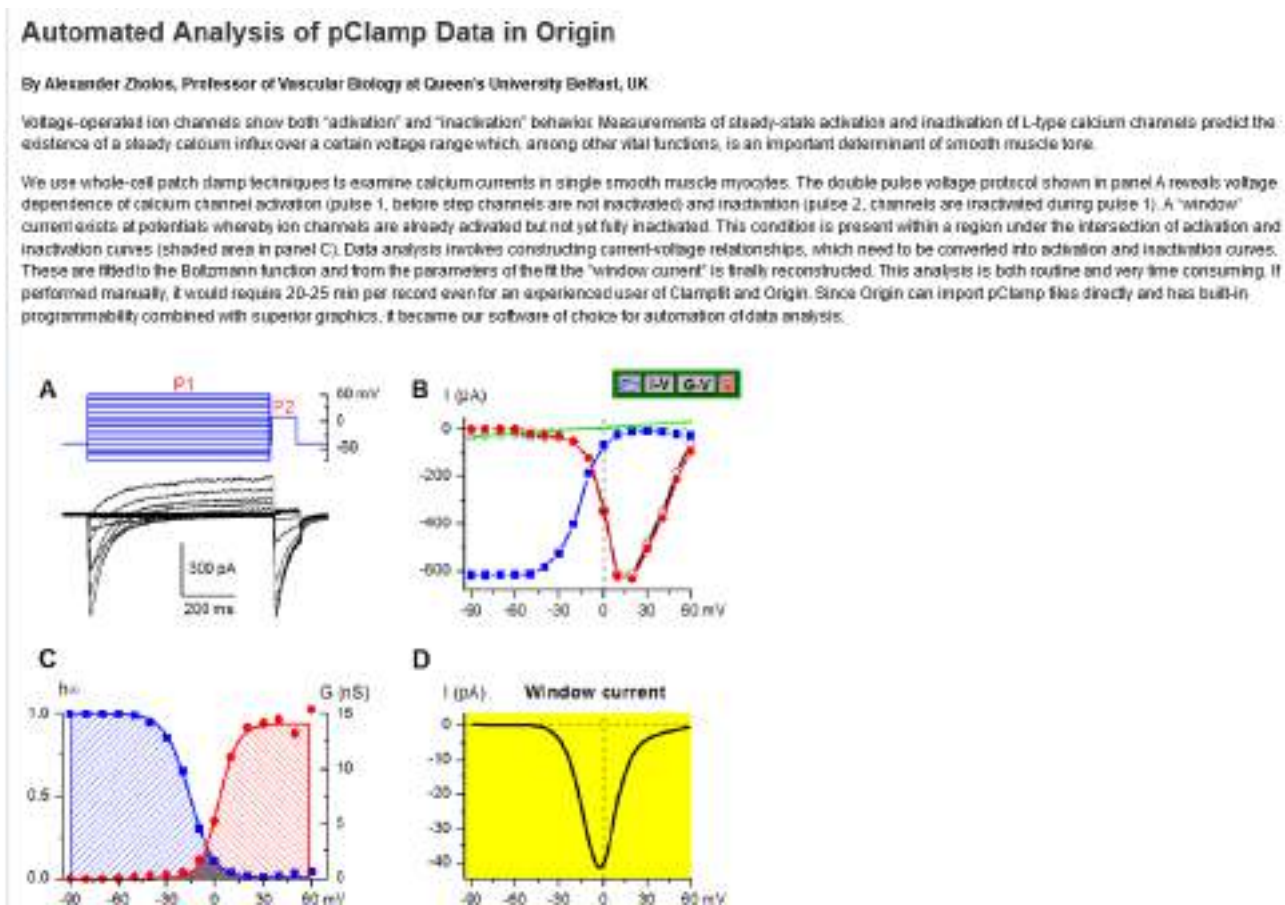


2. Самостійно обрати біологічний об'єкт та схематично зобразити його з використанням інструментів векторної графіки Inkscape.

3. Збережіть файл для перевірки викладачем.

#### 4.4. Аналіз наукових даних за допомогою програми Origin.

Ця практична робота складається з двох частин. Спочатку студенти знайомляться з різними застосуваннями Origin на прикладах наукових досліджень - Origin Case Studies <https://www.originlab.com/casestudies>. Ці матеріали містять описи експериментальних методів, наукових задач і як вони вирішувались засобами Origin, наприклад<sup>160</sup>.



У другій частині студенти використовують оригінальні дані імуноцитохімічного дослідження субклітинної локалізації рецепторів холоду і

<sup>160</sup> Zholos A.V. Automated Analysis of pClamp Data in Origin.

<https://www.originlab.com/index.aspx?go=Solutions/CaseStudies&pid=1540>

ментолу TRPM8 у міоцитах хвостової артерії щура в Origin, які були отримані за допомогою методу лазерної конфокальної мікроскопії<sup>161</sup>, для їх подальшого кількісного аналізу в Origin. Один з таких прикладів був пояснений у розділі 2.3.

---

<sup>161</sup> Johnson C.D., Melanaphy D., Purse A., Stokesberry S.A., Dickson P., Zholos A.V. (2009). Transient receptor potential melastatin 8 channel involvement in the regulation of vascular tone. *Am J Physiol Heart Circ Physiol*. 296(6):H1868-1877.

<https://pubmed.ncbi.nlm.nih.gov/19363131/>

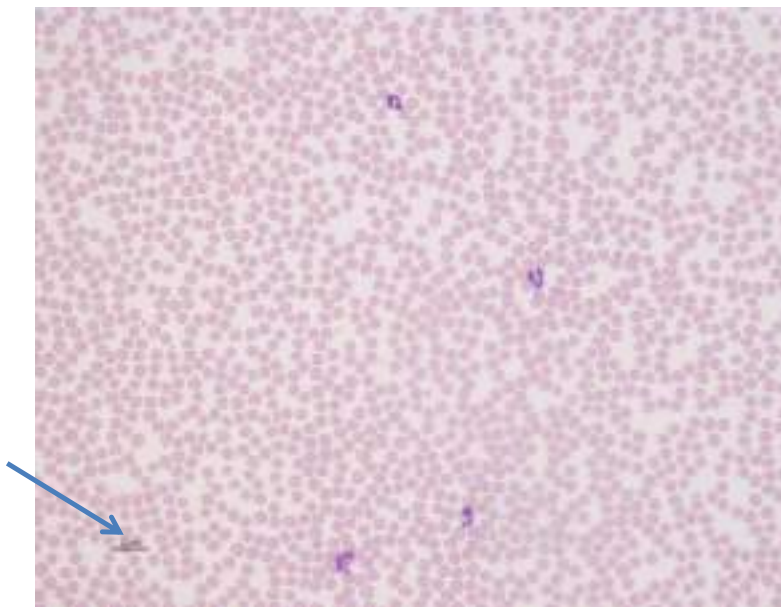
#### ***4.5. Побудова графіків та кількісний аналіз зображень в Origin та ImageJ. Статистичний і автоматизований аналіз даних.***

##### **А. Автоматичний підрахунок кількості клітин у полі зору мікроскопу та аналіз розподілу їх розмірів за допомогою ImageJ.**

У першій частині цієї лабораторної роботи студенти отримують практичні навички роботи з ImageJ. Програма вже інстальована на ПК у комп'ютерному класі.

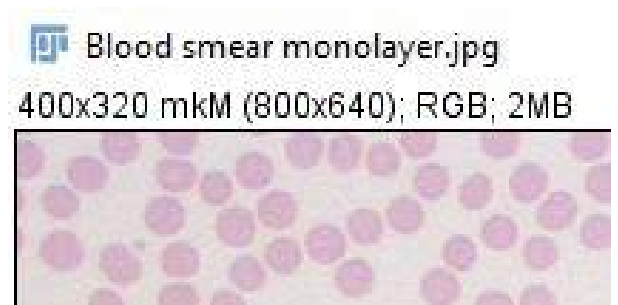
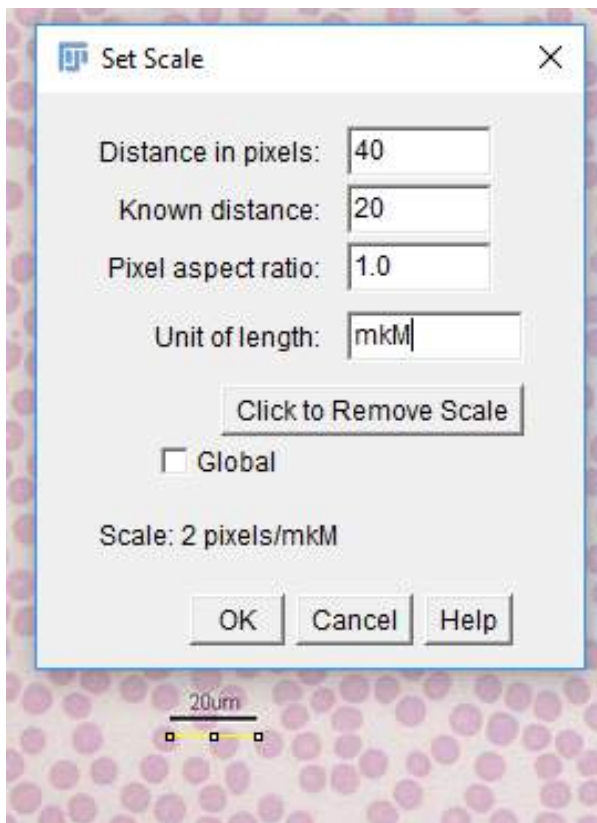
##### **Хід роботи:**

1. Відкрити файл із зображенням еритроцитів.



На ньому є калібрівка 20 мкм, яка буде використовуватись для розрахунків розміру (площі) клітин.

2. Провести лінію скориставшись відповідною кнопкою на панелі інструментів і, з врахуванням розмірів зображення у пікселях - 800×640, за допомогою функції меню «“Analyze-Set scale”» задати шкалу зображення як показано нижче:

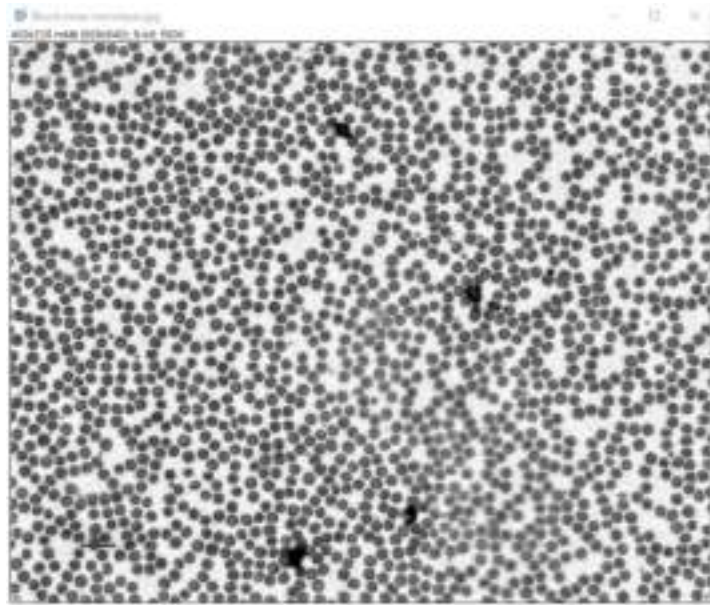


3. Виконайте процесінг зображення для його подальшого кількісного аналізу наступним чином:

Image – Type – 8 bit

Image-Adjust-Brightness/Contrast

Process-Subtract Background



4. Виконайте подальший процесинг зображення шляхом «Image-Adjust-Threshold»:

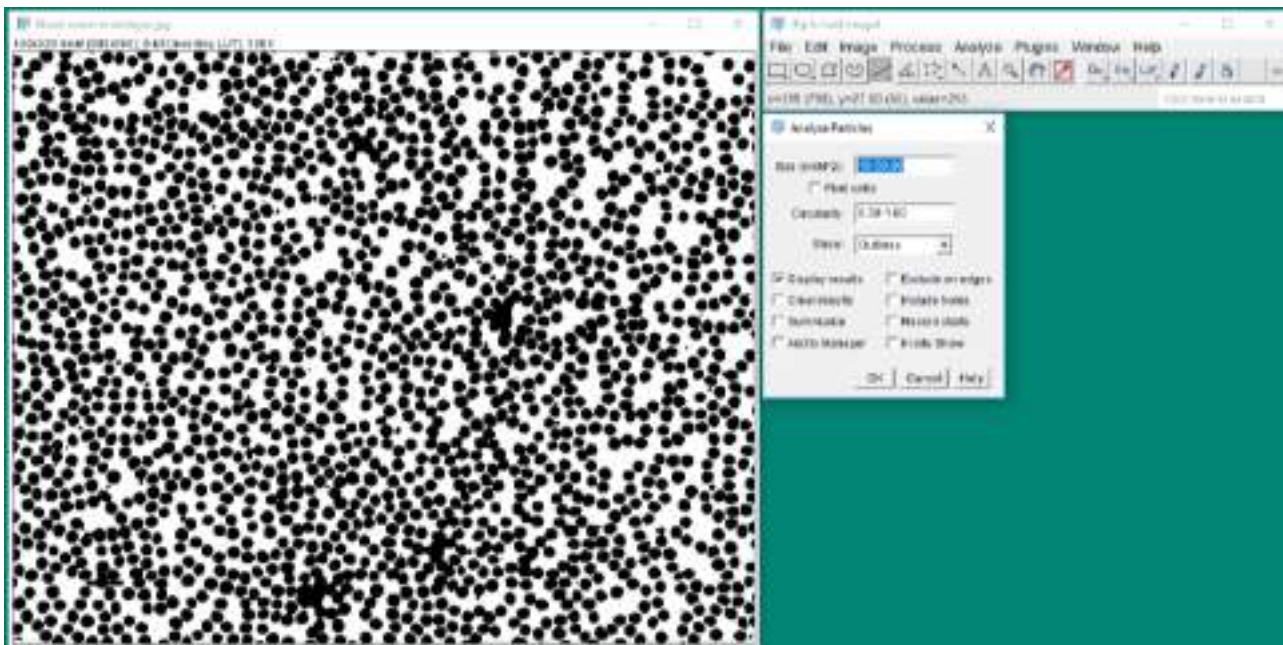


5. В меню «Analyze» оберіть «Analyze Particles...»<sup>162</sup>. Далі за допомогою діалогового вікна налаштуйте аналізатор частинок. Частинки менші ніж зазначено в полі «Розмір» ігноруються. Округлість =1 означає ідеальне коло,

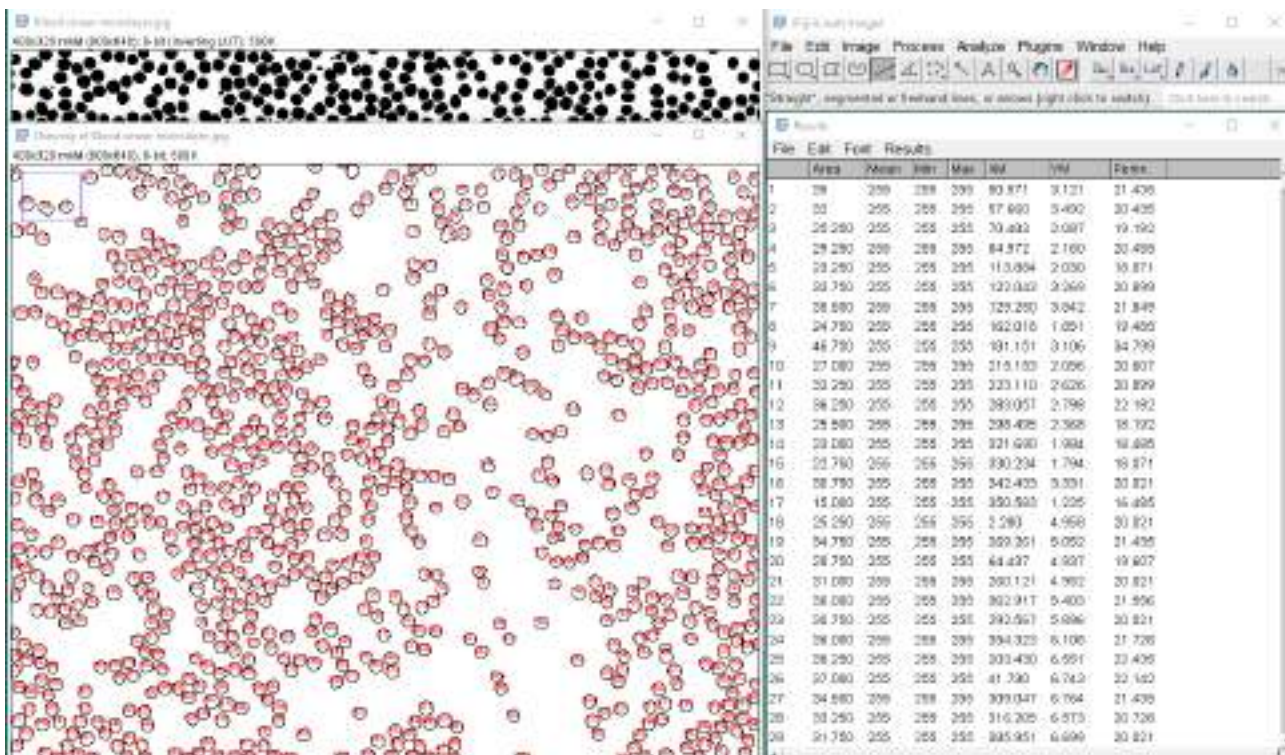
---

<sup>162</sup> ImageJ Analyze Particles. <https://imagej.nih.gov/ij/docs/menus/analyze.html>

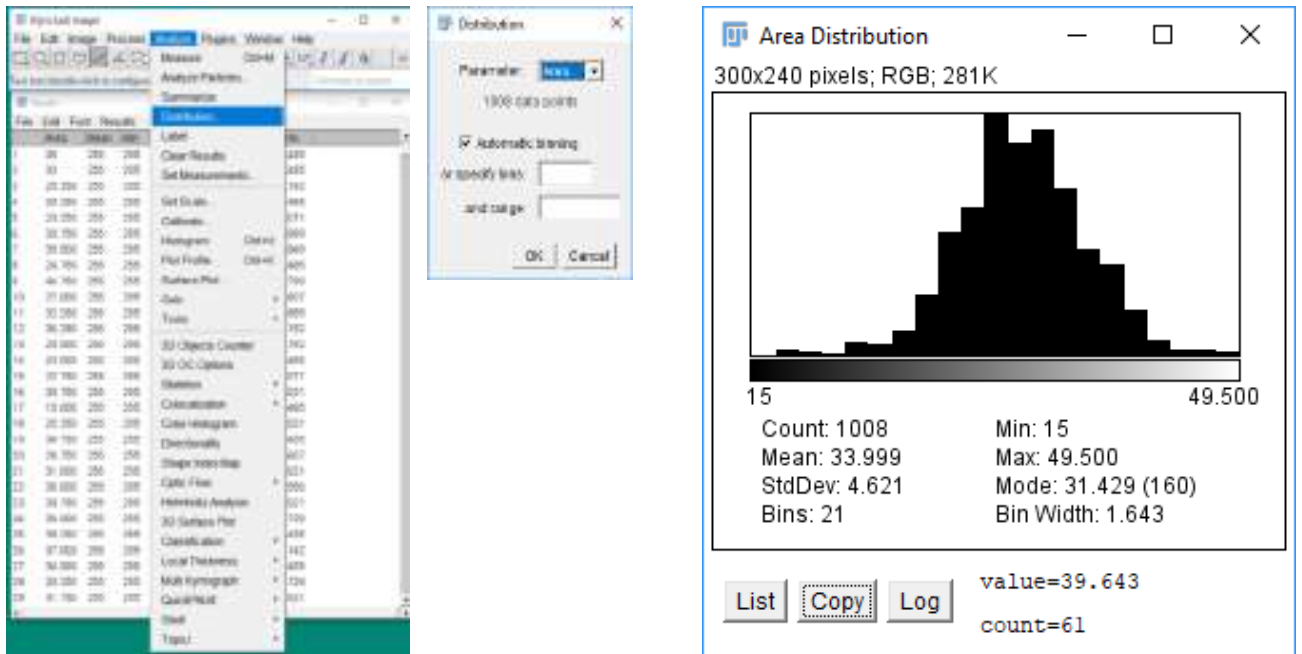
якщо ж вони деформовані більше певної межі (у цьому прикладі 0,3), то такі частинки теж ігноруються.



6. У результаті цього аналізу отримуємо кількісну інформацію про всі знайдені таким чином еритроцити:



7.3 використанням меню «Analyze Particles > Distribution» проаналізуємо цю інформацію, при цьому нас цікавить площа клітин, розмір бінів гистограми обираємо автоматичним:



8. Отже, можемо зробити висновок про те, що розмір еритроцитів (площа) має нормальний розподіл і у середньому становить  $34,0 \pm 4,6 \mu\text{m}^2$  (середнє  $\pm$  S.D.;  $n=1008$ ).

### Рекомендовані ресурси для кращого вивчення цього матеріалу:

ImageJ: Tutorials and Examples

<https://imagej.nih.gov/ij/docs/examples/index.html>

ImageJ Basics

<https://imagej.nih.gov/ij/docs/pdfs/ImageJ.pdf>

Image Processing with ImageJ

[https://imagej.nih.gov/ij/docs/pdfs/Image\\_Processing\\_with\\_ImageJ.pdf](https://imagej.nih.gov/ij/docs/pdfs/Image_Processing_with_ImageJ.pdf)

ImageJ Tutorial

<https://imagej.nih.gov/ij/docs/examples/IJ-M&M08.ppt>

## **Б. Кількісний аналіз біологічних зображень за допомогою програми Origin.**

### **Оформлення роботи:**

Робота має складатись з таких частин:

Вступ

Теоретична частина

Методи (інструменти Origin для аналізу зображень)

Результати дослідження

Висновки

Список літератури

### **Вимоги:**

Вступ містить стисле пояснення біологічної суті проблеми (до 1 сторінки), що дозволяє сформулювати мету дослідження.

Теоретична частина – містить детальне пояснення зображення, яке буде аналізуватися засобами Origin (1-2 сторінки).

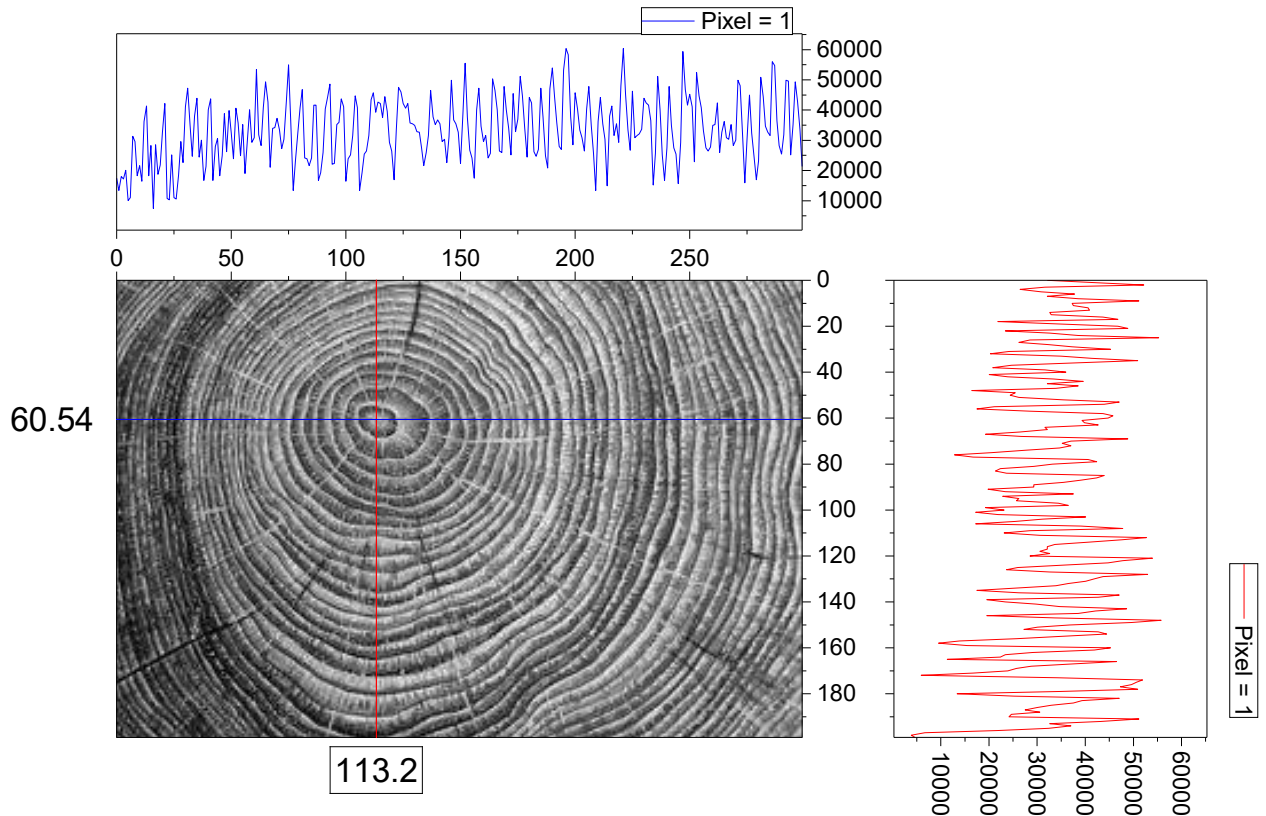
Методи – зазначити можливості та інструменти Origin для аналізу зображень, послідовність виконаних операцій.

### **Хід роботи:**

В якості приклада наведено аналіз зріза стовбура дерева. Для аналізу зображення потрібно:

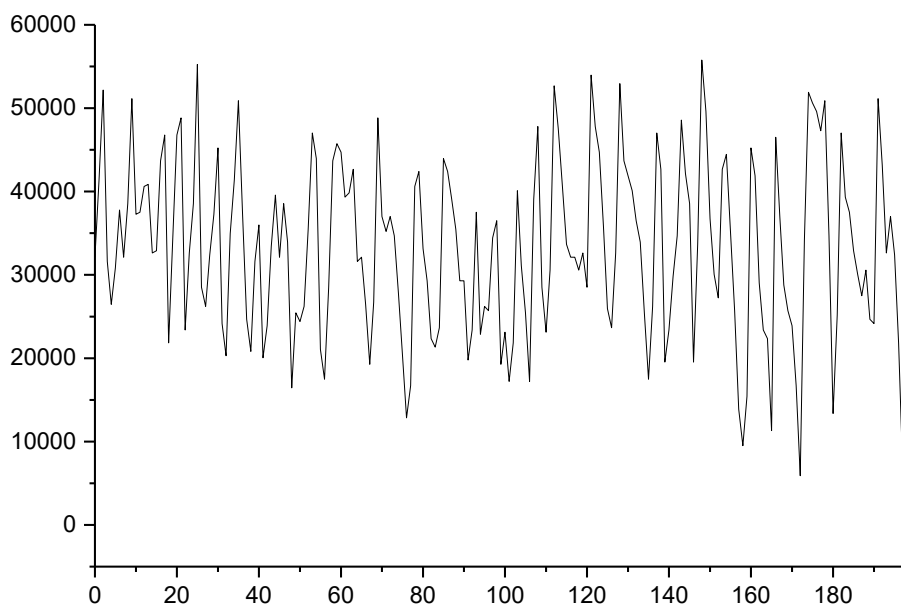
1. Зберегти обране зображення в одному з графічних форматів (JPG, TIFF, PNG тощо).
2. В програмі Origin створити нове вікно Matrix (File>New>Matrix).

3. Імпортувати зображення (File>Import>Image to Matrix).
4. Створити графік зображення (Plot>Profile>Image profiles).
5. Розташувати вертикальний та горизонтальний курсори, або провести довільну лінію (в залежності від задачі) і вставити отриманий графік в звіт (меню Edit>Copy page>Paste in MS Word) (Рис. 1).



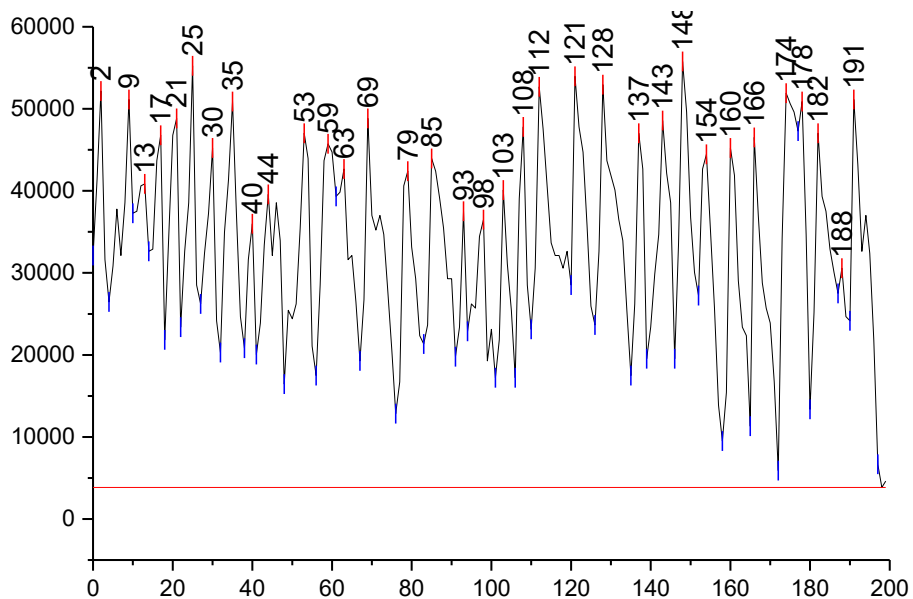
**Рис. 1.** Горизонтальний та вертикальні профілі зображення зрізу стовбура дерева.

6. Відкрити робочу книгу (Workbook), що містить профіль зображення, і побудувати його окремий графік (вибрати колонку Y →Plot>Line) (Рис. 2):



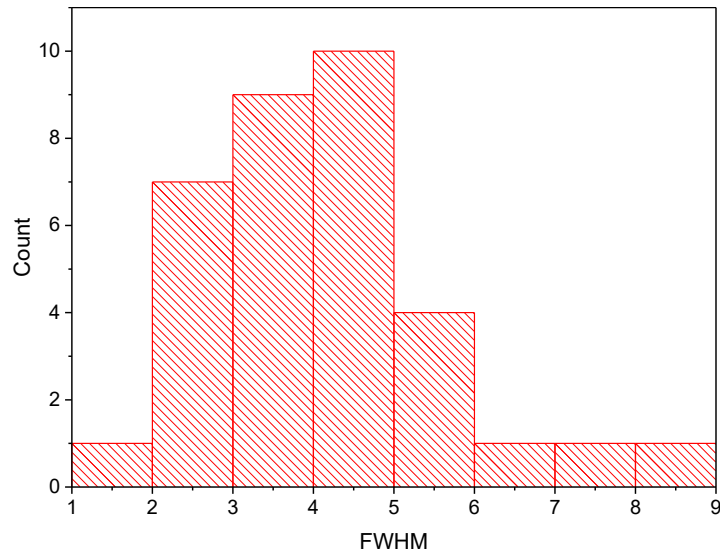
**Рис. 2.** У наведеному прикладі цей графік показує профіль зображення по лінії вертикального курсора.

7. Провести кількісний аналіз зображення. Для цього використайте функцію Origin: Analysis>Peak Analyzer>Find Peaks (Рис. 3).



**Рис. 3.** Аналіз піків вертикального профілю зображення.

8. Відкрити робочу книгу (Workbook), що містить профіль зображення, проаналізувати отриману кількісну інформацію щодо профілю зображення. Наприклад, на Рис. 4 наведена гістограма ширини піків:



**Рис. 4.** Гістограма ширини річних кілець у стовбурі дерева.

**Рекомендовані ресурси для кращого вивчення цього матеріалу:**

OriginLab Webinars: Working with Image

<https://www.originlab.com/index.aspx?go=SUPPORT/ONLINETRAINING&pid=31>

[69](#)

Під час вебінару демонструється Origin Apps - Image Profile та Object Counting.


#### 4.6. Дослідження амінокислотних послідовностей, еволюційних стосунків та видоспецифічних змін в структурі білків.

**Тема роботи:** Дослідження філогенезу міоглобіну засобами біоінформатичного аналізу.

**Мета роботи:** проаналізувати еволюційні стосунки у молекулі кисень-зв'язувального білку міоглобіну (позначається Mb або MB) ссавців<sup>163</sup>, який транспортує кисень, зокрема у скелетних м'язах.

#### Хід роботи

1. В базі даних UniProt <http://www.uniprot.org/> провести пошук 10-15 секвенованих ізоформ міоглобіну (Myoglobin) різних видів, що кодується геном *Mb*, бажано з різних таксономічних груп.

Особливу увагу потрібно звернути на ановані (Reviewed) білки, що позначені символом  а також на кількість амінокислотних залишків (близько 154). Інформацію щодо цих молекул занести в таблицю. Проаналізовані ізоформи міоглобіну різних видів з бази даних Uniprot занести у таблицю подібним до першого прикладу чином.

UniProt Entry <a href="http://www.uniprot.org">www.uniprot.org</a>	Назва виду латинською мовою	Назва виду українською мовою	Кількість амінокислотних залишків
P04247	<i>Mus musculus</i>	Миша хатня, або звичайна	154

<sup>163</sup> Khoshouei M, Danev R, Plitzko JM, Baumeister W. (2017). Revisiting the structure of hemoglobin and myoglobin with cryo-electron microscopy. *J Mol Biol.* 429(17):2611-2618.


У додатку до роботи навести ці послідовності наступним чином:

Міоглобін миші (UniProtKB - P04247 (MYG\_MOUSE)):

>sp|P04247|MYG\_MOUSE Myoglobin OS=Mus musculus OX=10090 GN=Mb PE=1 SV=3

MGLSDGEWQLVLNVWGKVEADLAGHGQEVLIQLFKTHPETLDKFDKFKNL  
 KSEEDMKGSEDLKKHGCTVLTALGTILKKKGQHAAEIQLAQSHTKHKIPV  
 KYLEFISEPIIEVLKKRHSDFGADAQGAMSKALELFRNDIAAKYKELGFQG

... (продовжити для інших видів)

2. Відмітити потрібні білки  та зберегти ці послідовності в текстовому форматі FASTA (canonical).

Entry	Entry name	Protein name	Gene name	Organism
<input type="checkbox"/> P02144	MYG_HUMAN	Myoglobin	MB	Homo sapiens (Human)
<input type="checkbox"/> P04247	MYG_MOUSE	Myoglobin	Mb	Mus musculus (Mouse)
<input type="checkbox"/> P02192	MYG_BOVIN	Myoglobin	MB	Bos taurus (Bovine)
<input type="checkbox"/> Q9QZ76	MYG_RAT	Myoglobin	Mb	Rattus norvegicus (Rat)

Download selected (8)  Download all (919)

**Format:** FASTA (canonical) ▾

Compressed  Uncompressed

[Preview first 10<sup>i</sup>](#)

3. Провести множинне вирівнювання амінокислотних послідовностей з використанням веб-серверів ClustalW та/або ClustalW2.

<http://www.genome.jp/tools/clustalw/>

<http://www.ebi.ac.uk/Tools/msa/clustalo/>

Після цього отримати філогенетичне дерево у формі Rooted phylogenetic tree with branch length (UPGMA) – ця функція знаходиться в кінці веб-сторінки.

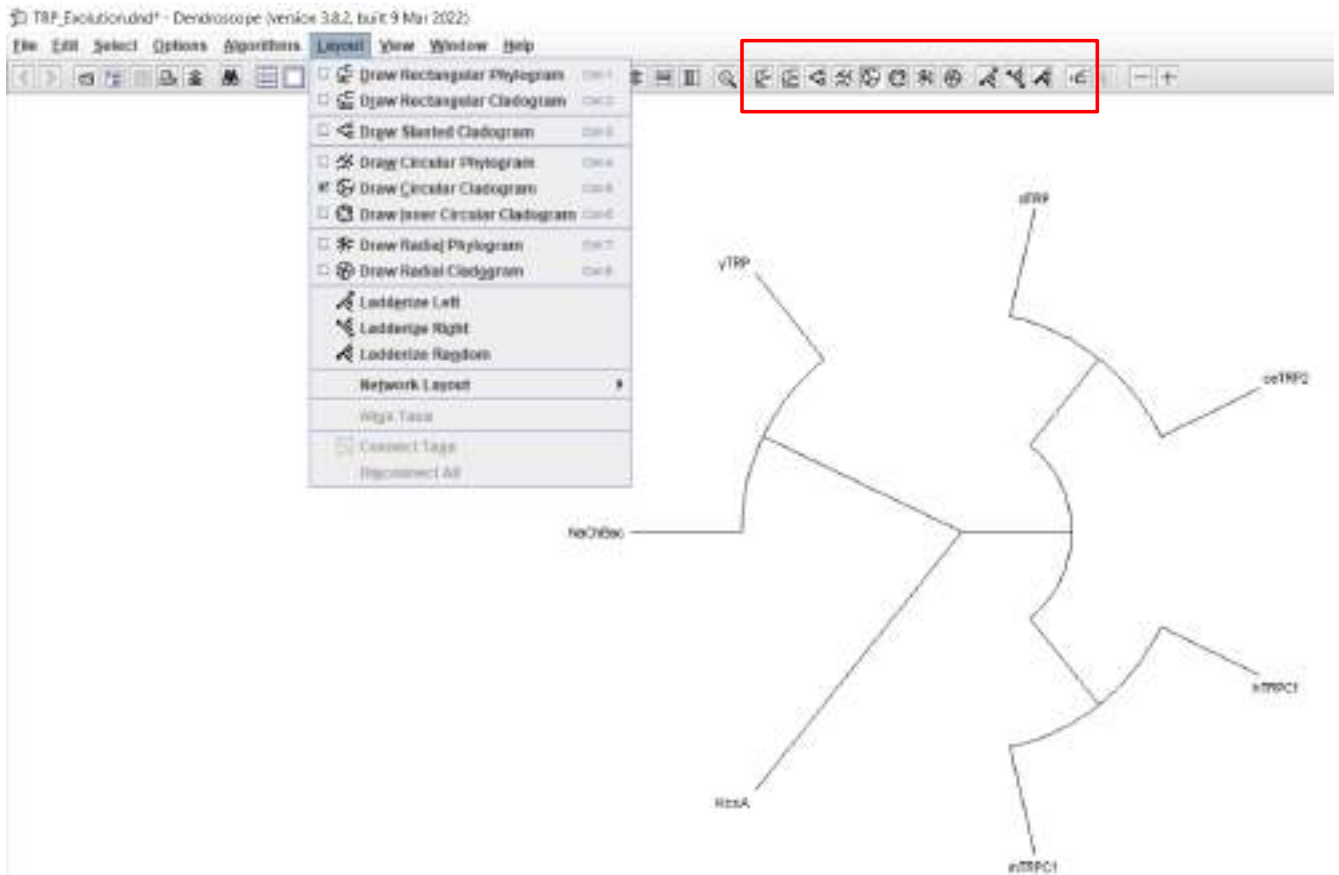
4. Зберегти результати вирівнювання послідовностей у форматі clustalw.dnd та відкрити цей файл у програмі Dendroscope<sup>164</sup>.

5. Побудувати філогенетичне дерево міоглобіну у програмі Dendroscope. Можливі декілька варіантів представлення даних, як проілюстровано на Рис. 1:

---

<sup>164</sup> Dendroscope3 - An interactive viewer for rooted phylogenetic trees and networks.

<http://ab.inf.uni-tuebingen.de/software/dendroscope/>



**Рис. 1.** Для графічного представлення даних скористайтесь меню “Layout” або кнопками на панелі інструментів, що відмічені червоним прямокутником.

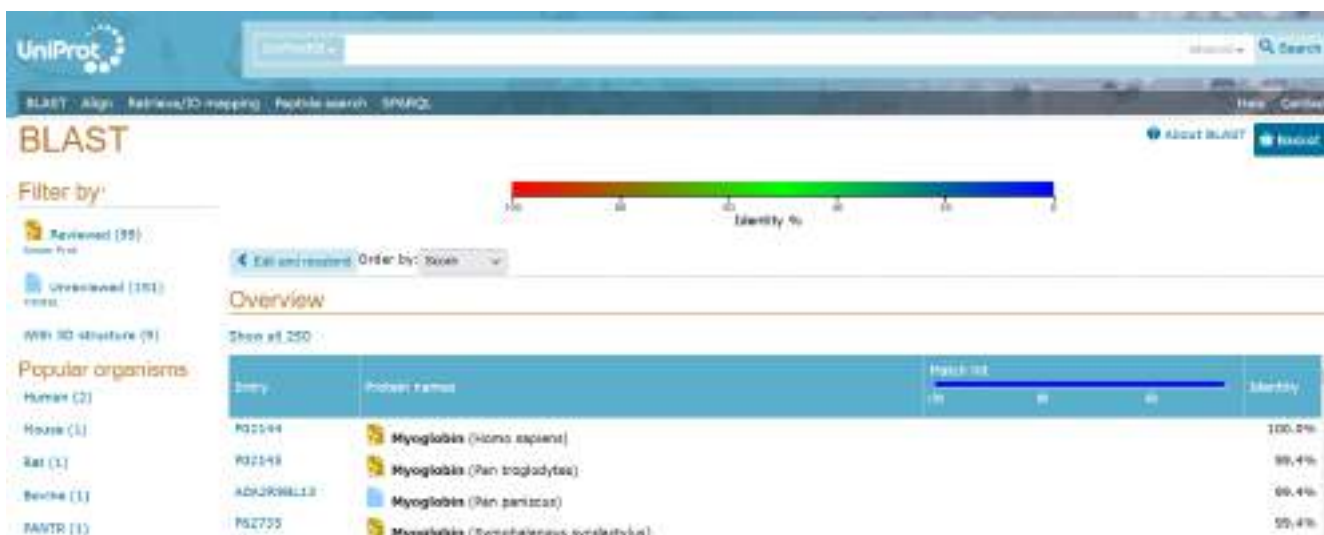
6. Побудувати дерево еволюції відповідних видів з використанням NCBI “Taxonomy - Common tree” <http://www.ncbi.nlm.nih.gov/taxonomy>

Для цього потрібно зберегти дерево еволюції у форматі phylog tree і потім відкрити цей файл у програмі Dendroscope 3. Приклад для TRP каналів показано на Рис. 1.

7. Порівняти філогенетичне дерево міоглобіну з деревом еволюції для цих же видів і зробити висновки.

8. База даних UniProt також дозволяє виконати множинне вирівнювання амінокислотних послідовностей, детальні інструкції і відео ілюстрація знаходяться за посиланням: <https://www.uniprot.org/help/sequence-alignments>

Крім того, ресурс Basic Local Alignment Search Tool (BLAST) дозволяє провести пошук подібних амінокислотних послідовностей у базі даних UniProt. Детальні інструкції і відео ілюстрація знаходяться за посиланням: <https://www.uniprot.org/help/sequence-searches>. Приклад такого аналізу проілюстровано на Рис. 2.



**Рис. 2.** Аналіз амінокислотних послідовностей з використанням BLAST.

9. Оформити роботу відповідно до наступних інструкцій.

### Оформлення роботи

1. Вступ – будова і функція міоглобіну (0,5-1 сторінки). Ви можете скористатися інформацією, що містить база даних UniProt, а також Protein Data Bank – PDB: <https://www.rcsb.org>.

2. Методи дослідження - навести характеристику використаних програм, їх призначення та можливостей (1-2 сторінки).
3. Результати дослідження та їх обговорення (навести отримані рисунки, діаграми з підписами і короткими поясненнями).
4. Обговорення отриманих результатів і висновки.

#### ***4.7. Аналіз біологічних даних з використанням мови програмування R.***

**Тема заняття:** Статистичний аналіз даних за допомогою мови програмування R.

**Мета заняття:** подальше ознайомлення студентів з можливостями мови програмування R для аналізу і візуалізації біологічних даних (розділ 2.4).

**Завдання:** проаналізувати дані лабораторних аналізів концентрації гемоглобіну крові з використанням мови програмування R.

Студенти використовують Excel файл з даними щодо концентрації гемоглобіну у крові 150 пацієнтів, з них 75 – жінки і 75 – чоловіки. Ці дані вже були використані у практичній роботі 4.2, отже у студентів буде можливість порівняти аналіз тих самих даних в MS Excel та за допомогою R. У першій частині роботи студенти інсталиують R та Rstudio і виконують аналіз цих даних за допомогою наведеного нижче R коду. У другій частині вони виконують індивідуальне завдання і аналізують дані, які є в R для навчання.

#### **Хід роботи:**

1. Інсталиувати R та RStudio з веб-порталів <https://cran.rstudio.com/> або <https://cloud.r-project.org/> та <https://rstudio.com/products/rstudio/download/#download>.
2. Скопіювати Excel файл з даними «Hemoglobin data.xlsx» (файл надається викладачем) у папку C:\Temp, як передбачено у наведеному нижче коді (відмічено жовтим кольором).
3. Відкрити R Console і вставити наступний код:

```
install.packages("readxl")  
install.packages("moments")
```

```
###sknewness and kurtosis come from moments
```

```
library("readxl")
```

```
library("moments")
```

```
Sys.setlocale("LC_CTYPE", "ukrainian")
```

```
datah <- read_excel("C:\\Temp\\Hemoglobin data.xlsx", col_names = c("Gender",  
"Grams"))
```

```
datah <- unstack(datah, form = Grams ~ Gender)
```

```
sapply(datah, length)
```

```
summary(datah)
```

```
Mode <- function(x) {
```

```
  ux <- unique(x)
```

```
  mode <- ux[which.max(tabulate(match(x, ux)))]
```

```
  return(mode)
```

```
}
```

```
sapply(datah, Mode)
```

```
t.test(datah$Ч, datah$Ж, alternative = "two.sided", var.equal = FALSE)
```

```
hist(datah$Ч)
```

```
###hist(datah$Ч,breaks=20)
```

```
hist(datah$Ж)
```

```
sapply(datah, shapiro.test)
```

```
sapply(datah, skewness)
```

```
sapply(datah, kurtosis)
```

4. Після виконання цього коду миттєво отримуємо наступні результати:

```
> sapply(datah, length)
```

```
Ч Ж
```

```
75 75
```

```
>
```

```
> summary(datah)
```

```
Ч Ж
```

```
Min. :103.0 Min. : 72.0
```

```
1st Qu.:141.5 1st Qu.:105.0
```

```
Median :154.0 Median :114.0
```

```
Mean :154.6 Mean :117.4
```

```
3rd Qu.:167.0 3rd Qu.:128.5
```

```
Max. :205.0 Max. :157.0
```

```
>
```

```
> Mode <- function(x) {
```

```
+ ux <- unique(x)
```

```
+ mode <- ux[which.max(tabulate(match(x, ux)))]
```

```
+ return(mode)
```

```
+ }
```

```
>
```

```
> sapply(datah, Mode)
```

```

У Ж
148 99
>
> t.test(datah$У, datah$Ж, alternative = "two.sided", var.equal = FALSE)

```

### Welch Two Sample t-test

```

data: datah$У and datah$Ж
t = 12.166, df = 144.12, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 31.16740 43.25926
sample estimates:
mean of x mean of y
154.5733 117.3600

```

```

>
> hist(datah$У)
> ###hist(datah$У,breaks=20)
> hist(datah$Ж)
>
> sapply(datah, shapiro.test)
      У          Ж
statistic 0.995023      0.9797328
p.value  0.993056      0.2724655
method   "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"      "X[[i]]"

```

```
> sapply(datah, skewness)
```

```
Ч Ж
```

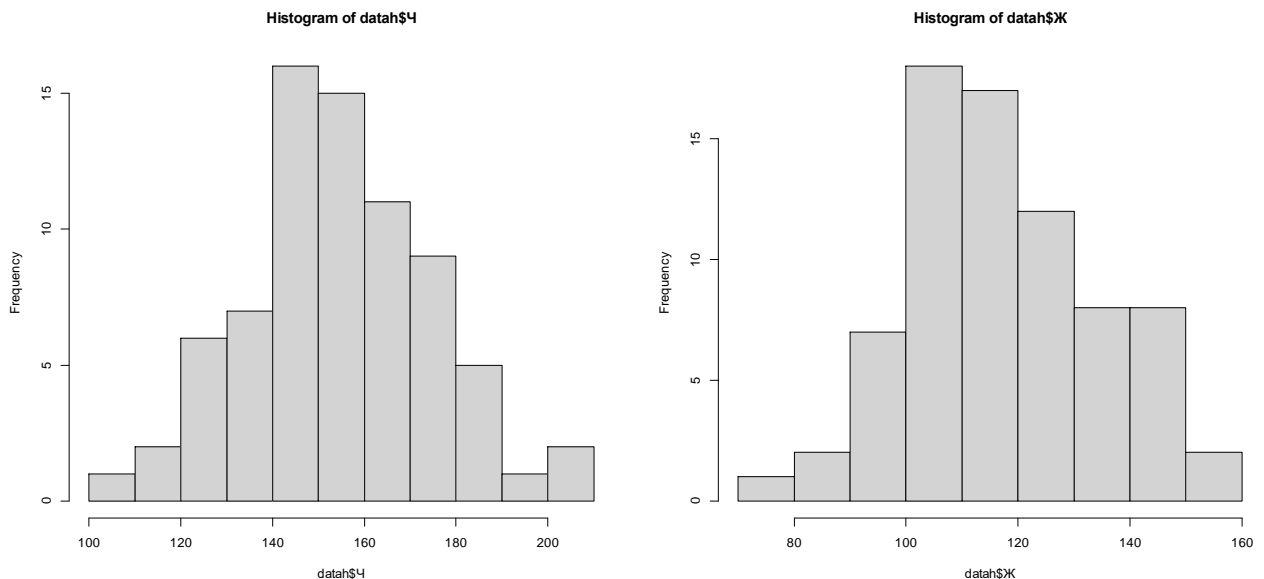
```
0.1150151 0.1937722
```

```
> sapply(datah, kurtosis)
```

```
Ч Ж
```

```
2.963743 2.732248
```

```
>
```



5. Разом з викладачем розібрати використані у цьому коді функції R<sup>165</sup> та отримані результати аналізу даних, що включають:

1. Середню ( $\pm$ SD) концентрацію гемоглобіну крові у чоловіків ( $n=75$ ) і жінок ( $n=75$ ), медіани, мінімальні та максимальні величини.
2. Тест на нормальність розподілу даних.
3. Визначено, що різниця у концентарціях гемоглобіну крові чоловіків і жінок є статистично достовірною ( $p\text{-value} < 2.2e-16$ ).
4. Гістограми, що ілюструють розподіл цих величин окремо у жінок і чоловіків.

---

<sup>165</sup> List of R Commands & Functions. <https://statisticsglobe.com/r-functions-list/>

## КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть основні властивості інформації з їх короткими поясненнями.
2. Як класифікується інформація за способом її сприйняття? Наведіть деякі приклади.
3. Як класифікується інформація за способом її подання? Наведіть деякі приклади.
4. Наведіть і поясніть методи збереження і захисту інформації.
5. У чому полягає суть закону Мура з точки зору розвитку комп'ютерних технологій?
6. Одиниця вимірювання швидкодії обчислювальних приладів, формула для її розрахунку.
7. Якими електронними базами публікацій можна скористатись для роботи з науковою літературою? Поясніть особливості та переваги деяких з них.
8. Які наукометричні показники Ви знаєте? Поясніть на конкретному прикладі як можна розрахувати один з цих показників.
9. Застосування бібліографічного менеджера Mendeley.
10. Вкажіть, якими є головні застосування програми MS Excel.
11. Наукові задачі, які можна вирішувати із застосуванням MS Excel.
12. Основні "вікна" програми Origin та їх призначення.
13. Яким чином краще організувати свій проєкт в Origin?
14. Поясніть послідовність аналізу біологічних зображень з використанням програми Origin.
15. Можливості Origin для побудови 2D і 3D графіків.
16. Можливості Origin для кількісного аналізу зображень.
17. Основні категорії команд LabTalk, наведіть деяких приклади таких команд.
18. Як можна автоматизувати процес аналізу даних та побудови графіків в Origin?

19. Призначення програми ImageJ. Поясніть в якості прикладу одну з її функцій.
20. Наведіть відомі Вам галузі застосування біоінформатики.
21. Наведіть 5 найбільш застосовуваних доменів і форматів біоінформатичного аналізу.
22. Які завдання вирішував проєкт «Геном-5000»?
23. Наведіть декілька прикладів відомих Вам модельних організмів.
24. Які бази даних генів Ви знаєте? Деталізуйте яку саме інформацію можна отримати з цих баз та наведіть деякі приклади.
25. Для вирішення яких задач використовується програма VMD?

## РЕКОМЕНДОВАНІ ДЖЕРЕЛА

### *Основні: (Базові)*

1. Медична інформатика в модулях: практикум / І.Є. Булах, Л.П. Войтенко, М.Р. Мруга та ін.; за ред. І.Є.Булах. – К.: Медицина, 2012. – 208 с.
2. Прилуцький Ю.І., Богуцька К.І., Залоіло І.А., Оглобля О.В., Цимбалюк О.В., Мірошниченко М.С. Методичні вказівки до курсу “Інформатика” для студентів біологічного факультету. - К: ВПЦ КУ, 2002. - 67 с.
3. Baker M. Big biology: The 'omes puzzle // Nature. – 2013. – 494. - P. 416–419.
4. Bidaux G., Sgobba M., Lemonnier L., Borowiec A.S., Noyer L., Jovonovic S., Zholos A.V., Haider S. (2015). Functional and modelling studies of the transmembrane region of the TRPM8 channel. Biophysical Journal 109 (9), 1840-1851.
5. Etchings J.A. Strategies in Biomedical data science: driving force for innovation. – Wiley, 2017. – 464 pp.
6. Informatics in Medical Imaging: Imaging in Medical Diagnosis and Therapy / Editors G.C. Kagadis, S.G. Langer. - Boca Raton: CRC Press, 2011. - 367 pp.
7. Kalet I.J. Principles of Biomedical Informatics, 2nd edition / Academic Press, 2013 – 708 pp.
8. Medical Imaging Informatics / Editors A.T. Bui, R.K. Taira. - New York Dordrecht Heidelberg London: Springer, 2010. - 551 pp.
9. Microsoft Excel. GSFGlobal [Електронний ресурс]. - Режим доступу: <https://edu.gcfglobal.org/en/topics/excel/>
10. Origin User Guide [Електронний ресурс]. - Режим доступу: <http://www.originlab.com/doc/User-Guide>
11. Pevsner J. Bioinformatics and Functional Genomics. - Wiley-Blackwell, 2009. – 992 p. [Електронний ресурс]. - Режим доступу: <http://www.bioinfbook.org>.
12. The NCBI Handbook, 2<sup>nd</sup> edition / Bethesda (MD): National Center for

Biotechnology Information, 2018. – 513 pp. [Електронний ресурс]. – Режим доступу: [https://www.ncbi.nlm.nih.gov/books/NBK143764/pdf/Bookshelf\\_NBK143764.pdf](https://www.ncbi.nlm.nih.gov/books/NBK143764/pdf/Bookshelf_NBK143764.pdf).

*Додаткові:*

- 13.Одинець К.О., Івахно С.С., Ковальський Д.Б., Токовенко Б.Т., Корнелюк О.І. Структурна біоінформатика в постгеномну еру // Біополімери і клітина. - 2004. - 20, №1-2. - С. 78-91.
- 14.Річард Резнік: Вітаємо вас у геномній революції (TEDxBoston 2011) [Електронний ресурс]. - Режим доступу: <https://www.youtube.com/watch?v=nAljnpoKVLw>.
- 15.AlphaFold: Using AI for scientific discovery. [Електронний ресурс]. - Режим доступу: <https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery>.
- 16.Handbook of the Philosophy of Science. Volume 8: Philosophy of Information / Editors P. Adriaans, J. van Benthem. – Elsevier, 2008. – 808 pp.
- 17.Meyer E.F., Swanson S.M., Williams J.A. Molecular modeling and drug design // Pharmacology & Therapeutics. - 2000. - 85. - P. 113-121.
- 18.Origin Tutorials [Електронний ресурс]. - Режим доступу: <http://www.originlab.com/doc/Tutorials>.
- 19.Origin LabTalk Scripting Guide [Електронний ресурс]. - Режим доступу: <http://www.originlab.com/doc/LabTalk/guide>.
- 20.Systems Biology - Bioinformatic Data Analysis. [Електронний ресурс]. - Режим доступу: <http://tbb.bio.uu.nl/BDA/>.
- 21.Xiong J. Essential bioinformatics. – Cambridge, UK: Cambridge University Press, 2006.

### *Інтернет ресурси:*

22. Аналіз білків, мутацій, кутів та відстаней між атомами, інтерактивна графіка Swiss-PdbViewer. [Електронний ресурс]. - Режим доступу: <http://spdbv.vital-it.ch>.
23. Банк даних білків (Protein Data Bank, PDB). [Електронний ресурс]. - Режим доступу: <http://www.pdb.org/pdb/home/home.do>, <http://www.rcsb.org/pdb/home/home.do>.
24. База даних генів і геномів Kyoto Encyclopedia of Genes and Genomes (KEGG). [Електронний ресурс]. - Режим доступу: <http://www.genome.jp/kegg/>.
25. Банк даних геномів Genomes OnLine Database. [Електронний ресурс]. - Режим доступу: <http://www.genomesonline.org/cgi-bin/GOLD/index.cgi>.
26. Банк даних нуклеїнових кислот. [Електронний ресурс]ю - Режим доступу: <http://ndbserver.rutgers.edu/>.
27. Візуалізація і 3D моделювання біомолекул VMD. [Електронний ресурс]. - Режим доступу: <http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD>.
28. Візуалізація і аналіз молекулярних структур MGLTools. [Електронний ресурс]. - Режим доступу: <http://mgltools.scripps.edu>.
29. Віртуальна лабораторія MolDynGrid. [Електронний ресурс]. - Режим доступу: <http://moldyngrid.org/main.php>.
30. Вирівнювання послідовностей ClustalW2. [Електронний ресурс]. - Режим доступу: <http://www.ebi.ac.uk/Tools/msa/clustalw2>.
31. Молекулярна динаміка Gromacs. [Електронний ресурс]. - Режим доступу: <http://www.gromacs.org/>.
32. Молекулярний докінг AutoDockTools. [Електронний ресурс]. - Режим доступу: <http://autodock.scripps.edu/resources/adt>.

33. Пошукова система PubMed [Електронний ресурс]. - Режим доступу: <http://www.ncbi.nlm.nih.gov/pubmed/citmatch>.
34. Ресурси NCBI: BLAST, Bookshelf, Gene, Genome, Nucleotide, OMIM, Protein, PubChem, PubMed, PubMed Central, SNP. [Електронний ресурс]. - Режим доступу: <http://www.ncbi.nlm.nih.gov>.
35. Ресурси SIB Bioinformatics Resource Portal: Proteomics, Genomics, Structural bioinformatics, Systems biology, Phylogeny/evolution, Population genetics, Transcriptomics, Biophysics, Imaging, IT infrastructure, Drug design. [Електронний ресурс]. - Режим доступу: <http://expasy.org>.
36. Ресурси Європейського Інституту Біоінформатики (EBI): Nucleotide and Protein sequence searching, Multiple Sequence Alignment, Pairwise Sequence Alignments, Protein Functional Analysis, Functional Genomics Tools, Molecular structural analysis, Scientific literature text mining, Sequence Translation. [Електронний ресурс]. - Режим доступу: <http://www.ebi.ac.uk/>.
37. Сервер для порівняльного моделювання білків SWISS-MODEL. [Електронний ресурс]. - Режим доступу: [http://swissmodel.expasy.org/workspace/index.php?func=show\\_workspace](http://swissmodel.expasy.org/workspace/index.php?func=show_workspace).
38. Симуляція молекулярної динаміки Molecular Dynamics Simulator (NAMD). [Електронний ресурс]. - Режим доступу: <http://www.ks.uiuc.edu/Research/namd>.
39. Створення та використання бібліографічних баз даних Mendeley [Електронний ресурс]. - Режим доступу: <http://www.mendeley.com>.
40. Структура і функція білків UniProt. [Електронний ресурс]. - Режим доступу: <http://www.uniprot.org>.
41. Brain Imaging Data Structure (BIDS) [Електронний ресурс]. - Режим доступу: <http://bids.neuroimaging.io>.
42. Excel help & learning. Microsoft Corporation [Електронний ресурс]. - Режим

- доступу: <https://support.microsoft.com/en-us/excel>
- 43.Excel video training. Microsoft Corporation [Электронный ресурс]. - Режим доступа: [https://support.microsoft.com/en-us/office/excel-video-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb?wt.mc\\_id=otc\\_home&ui=en-us&rs=en-us&ad=us](https://support.microsoft.com/en-us/office/excel-video-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb?wt.mc_id=otc_home&ui=en-us&rs=en-us&ad=us)
- 44.ImageJ (National Institutes of Health, USA) [Электронный ресурс]. - Режим доступа: <https://imagej.nih.gov/ij>.
- 45.Origin video tutorials [Электронный ресурс]. - Режим доступа: <http://www.originlab.com/index.aspx?go=Support/VideoTutorials>
- 46.CorelDRAW graphics suite tutorials [Электронный ресурс]. - Режим доступа: <http://www.corel.com/corel>
- 47.PubMed tutorials [Электронный ресурс]. - Режим доступа: <http://www.nlm.nih.gov/bsd/disted/pubmed.html>
- 48.PubMed quick start guide [Электронный ресурс]. - Режим доступа: [http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.PubMed\\_Quick\\_Start](http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.PubMed_Quick_Start)

## ДОДАТКИ

### Додаток 1.

#### Формат цитувань відповідно до вимог журналу Nature

Катіон-селективні  $\text{Ca}^{2+}$ -проникні канали транзйентного рецепторного потенціалу широко представлені у різних типах клітин, і зокрема у гладеньком'язових клітинах судин. Вони відіграють центральну роль у регуляції динаміки змін концентрації внутрішньоклітинного кальцію<sup>1</sup>, зокрема при активації спряжених з G білками рецепторів<sup>2</sup>. Особливо важливу роль у цих процесах мають канали підродин TRPC та TRPV<sup>3</sup>.

Було показано, що TRP канали регулюють міогенний тонус судин малого діаметру<sup>4 5 6</sup>. Крім того, TRP канали відіграють важливу роль в ендотелій-залежній регуляції скоротливої активності гладеньких м'язів судин<sup>7</sup>. Дисфункція цих каналів при захворюванні на діабет спричиняє порушення регуляції тонуусу судин та гіпертонічну хворобу<sup>8</sup>.

#### Список літератури:

1. Alonso-Carbajo, L. *et al.* Muscling in on TRP channels in vascular smooth muscle cells and cardiomyocytes. *Cell Calcium* vol. 66 48–61 (2017).
2. Dryer, S. E. & Kim, E. Y. Permeation and Rectification in Canonical Transient Receptor Potential-6 (TRPC6) Channels. *Front. Physiol.* **9**, 1055 (2018).
3. Martín-Bórnez, M., Galeano-Otero, I., Del Toro, R. & Smani, T. Trpc and trpv channels' role in vascular remodeling and disease. *International Journal of Molecular Sciences* vol. 21 1–17 (2020).
4. Kim, K. & Hong, K.-S. Transient receptor potential channel-dependent myogenic responsiveness in small-sized resistance arteries. *J. Exerc. Rehabil.* **17**, 4–10 (2021).
5. Jackson, W. F. Ion channels and the regulation of myogenic tone in peripheral

- arterioles. in *Current Topics in Membranes* vol. 85 19–58 (Academic Press Inc., 2020).
6. Barabas, P. *et al.* Ion channels and myogenic activity in retinal arterioles. in *Current Topics in Membranes* vol. 85 187–226 (Academic Press Inc., 2020).
  7. Hong, K.-S. & Lee, M.-G. Endothelial Ca<sup>2+</sup> signaling-dependent vasodilation through transient receptor potential channels. *Korean J. Physiol. Pharmacol.* **24**, 287–298 (2020).
  8. Moraes, R. D. A., Webb, R. C. & Silva, D. F. Vascular Dysfunction in Diabetes and Obesity: Focus on TRP Channels. *Front. Physiol.* **12**, 645109 (2021).

### **Формат цитувань відповідно до вимог журналу *The Journal of Physiology***

Катіон-селективні Ca<sup>2+</sup>-проникні канали транз'єтного рецепторного потенціалу широко представлені у різних типах клітин, і зокрема у гладеньком'язових клітинах судин. Вони відіграють центральну роль у регуляції динаміки змін концентрації внутрішньоклітинного кальцію (Alonso-Carbajo *et al.*, 2017), зокрема при активації спряжених з G білками рецепторів (Dryer & Kim, 2018). Особливо важливу роль у цих процесах мають канали підродин TRPC та TRPV (Martín-Bórnez *et al.*, 2020).

Було показано, що TRP канали регулюють міогенний тонус судин малого діаметру (Kim & Hong, 2021) (Jackson, 2020) (Barabas *et al.*, 2020). Крім того, TRP канали відіграють важливу роль в ендотелій-залежній регуляції скоротливої активності гладеньких м'язів судин (Hong & Lee, 2020). Дисфункція цих каналів при захворюванні на діабет спричиняє порушення регуляції тонусу судин та гіпертонічну хворобу (Moraes *et al.*, 2021).

### Список літератури:

Alonso-Carbajo L, Kecskes M, Jacobs G, Pironet A, Syam N, Talavera K &

- Vennekens R (2017). Muscling in on TRP channels in vascular smooth muscle cells and cardiomyocytes. *Cell Calcium* **66**, 48–61.
- Barabas P, Augustine J, Fernández JA, McGeown JG, McGahon MK & Curtis TM (2020). Ion channels and myogenic activity in retinal arterioles. In *Current Topics in Membranes*, pp. 187–226. Academic Press Inc.
- Dryer SE & Kim EY (2018). Permeation and Rectification in Canonical Transient Receptor Potential-6 (TRPC6) Channels. *Front Physiol* **9**, 1055.
- Hong K-S & Lee M-G (2020). Endothelial Ca<sup>2+</sup> signaling-dependent vasodilation through transient receptor potential channels. *Korean J Physiol Pharmacol* **24**, 287–298.
- Jackson WF (2020). Ion channels and the regulation of myogenic tone in peripheral arterioles. In *Current Topics in Membranes*, pp. 19–58. Academic Press Inc.
- Kim K & Hong K-S (2021). Transient receptor potential channel-dependent myogenic responsiveness in small-sized resistance arteries. *J Exerc Rehabil* **17**, 4–10.
- Martín-Bórnez M, Galeano-Otero I, Del Toro R & Smani T (2020). Trpc and trpv channels' role in vascular remodeling and disease. *Int J Mol Sci* **21**, 1–17.
- Moraes RDA, Webb RC & Silva DF (2021). Vascular Dysfunction in Diabetes and Obesity: Focus on TRP Channels. *Front Physiol* **12**, 645109.

### **Формат цитувань відповідно до вимог журналу Cell Calcium**

Катіон-селективні Ca<sup>2+</sup>-проникні канали транз'єтного рецепторного потенціалу широко представлені у різних типах клітин, і зокрема у гладеньком'язових клітинах судин. Вони відіграють центральну роль у регуляції динаміки змін концентрації внутрішньоклітинного кальцію [1], зокрема при активації спряжених з G білками рецепторів [2]. Особливо важливу роль у цих процесах мають канали підродин TRPC та TRPV [3].

Було показано, що TRP канали регулюють міогенний тонус судин малого діаметру [4][5][6]. Крім того, TRP канали відіграють важливу роль в ендотелій-залежній регуляції скоротливої активності гладеньких м'язів судин [7]. Дисфункція цих каналів при захворюванні на діабет спричиняє порушення регуляції тону судин та гіпертонічну хворобу [8].

#### Список літератури:

- [1] L. Alonso-Carbajo, M. Kecskes, G. Jacobs, A. Pironet, N. Syam, K. Talavera, R. Vennekens, Muscling in on TRP channels in vascular smooth muscle cells and cardiomyocytes, *Cell Calcium*. 66 (2017) 48–61. <https://doi.org/10.1016/j.ceca.2017.06.004>.
- [2] S.E. Dryer, E.Y. Kim, Permeation and Rectification in Canonical Transient Receptor Potential-6 (TRPC6) Channels., *Front. Physiol.* 9 (2018) 1055. <https://doi.org/10.3389/fphys.2018.01055>.
- [3] M. Martín-Bórnez, I. Galeano-Otero, R. Del Toro, T. Smani, Trpc and trpv channels' role in vascular remodeling and disease, *Int. J. Mol. Sci.* 21 (2020) 1–17. <https://doi.org/10.3390/ijms21176125>.
- [4] K. Kim, K.-S. Hong, Transient receptor potential channel-dependent myogenic responsiveness in small-sized resistance arteries., *J. Exerc. Rehabil.* 17 (2021) 4–10. <https://doi.org/10.12965/jer.2040836.418>.
- [5] W.F. Jackson, Ion channels and the regulation of myogenic tone in peripheral arterioles, in: *Curr. Top. Membr.*, Academic Press Inc., 2020: pp. 19–58. <https://doi.org/10.1016/bs.ctm.2020.01.002>.
- [6] P. Barabas, J. Augustine, J.A. Fernández, J.G. McGeown, M.K. McGahon, T.M. Curtis, Ion channels and myogenic activity in retinal arterioles, in: *Curr. Top. Membr.*, Academic Press Inc., 2020: pp. 187–226. <https://doi.org/10.1016/bs.ctm.2020.01.008>.

- [7] K.-S. Hong, M.-G. Lee, Endothelial  $\text{Ca}^{2+}$  signaling-dependent vasodilation through transient receptor potential channels., *Korean J. Physiol. Pharmacol.* 24 (2020) 287–298. <https://doi.org/10.4196/kjpp.2020.24.4.287>.
- [8] R.D.A. Moraes, R.C. Webb, D.F. Silva, Vascular Dysfunction in Diabetes and Obesity: Focus on TRP Channels., *Front. Physiol.* 12 (2021) 645109. <https://doi.org/10.3389/fphys.2021.645109>.

**Додаток 2.** Скрипти Origin на основі LabTalk, які ілюструють найбільш часто застосовані алгоритми аналізу електрофізіологічних даних.

Наведений приклад ілюструє одне із застосувань скриптів автора<sup>166</sup> розміщених на сайті Origin, а саме напівавтоматизований аналіз кривих доза-ефект. Такий аналіз проводиться як у академічних лабораторіях, так і у фармацевтичній індустрії для визначення кількісних параметрів взаємодії активаторів рецепторів (агоністів) з таргетними молекулами, таких як концентрація агоністу, що викликає напівмаксимальний ефект ( $EC_{50}$ ) і коефіцієнт Хілла (h). У наведеному прикладі аналізується амплітуда мускаринового струму у міоциті тонкого кишечника в залежності від концентрації агоніста мускаринових ацетилхолінових рецепторів карбахоліну<sup>167</sup>. Для аналізу потрібно натиснути 2 кнопки на панелі інструментів цього додатку:



Перша кнопка активує скрипт імпорту pClamp файлу, що містить дані експерименту. Друга кнопка активує аналітичну частку, що складається з таких основних кроків:

А – з'являється діалогове вікно, в якому потрібно зазначити назву агоністу та його концентрації – максимум до 10. Це приклад інтерфейсу користувача (user

---

<sup>166</sup> Zholos A.V. Advanced pCLAMP Analyzer.

<http://www.originlab.com/fileExchange/details.aspx?fid=357>

<sup>167</sup> Zholos A.V., Bolton T.B. (1997). Muscarinic receptor subtypes controlling the cationic current in guinea-pig ileal smooth muscle. *Br J Pharmacol.* 122(5):885-893.

<https://pubmed.ncbi.nlm.nih.gov/9384504/>

interface), який зручно використовувати у LabTalk для інтерактивного аналізу даних.

Б – з’являється графік з даними, який містить горизонтальні курсори, один для базової лінії сигналу, інші – по одному для кожної концентрації агоністу. Для прикладу на графіку внизу таких курсорів 6.

Agonist	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Carbachol	1	3	10	30	100	300	0	0	0	0

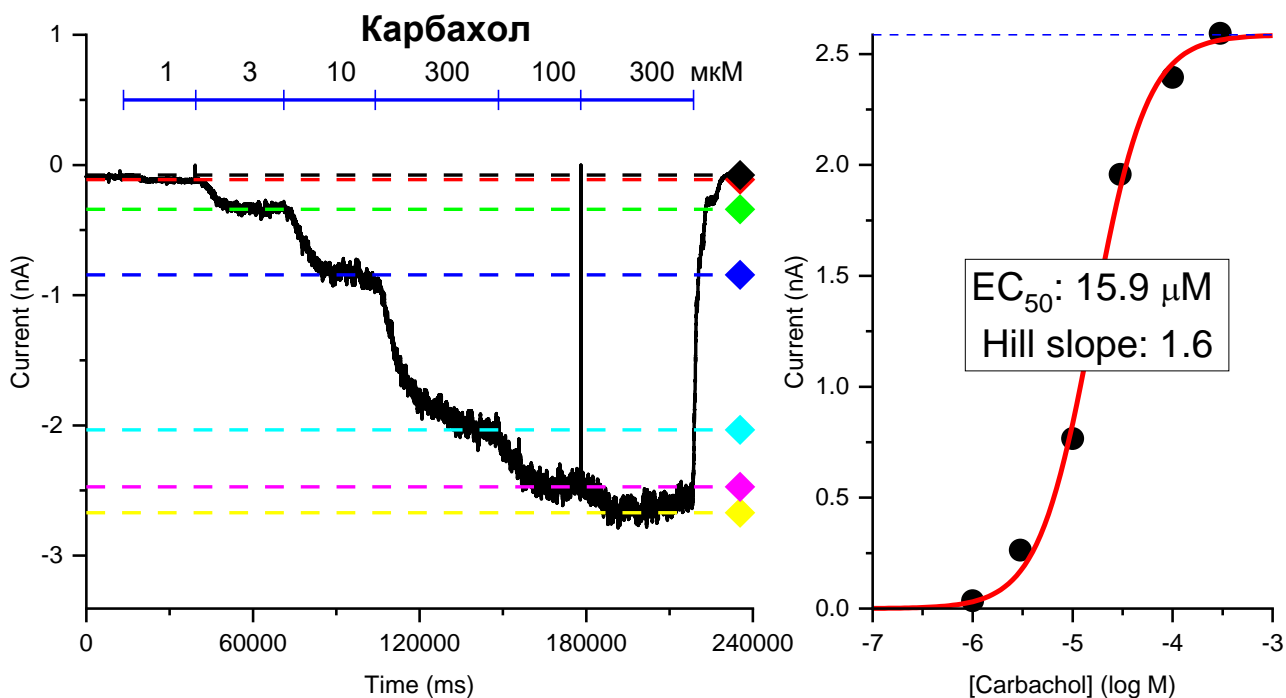
В – з’являється інше діалогове вікно, яке пропонує позначити курсорами потрібні сегменти кривої, а потім натиснути Ok.

Move Cursors to Position

Done Position horizontal cursors in the left graph.  
Click on the Done button when done.

Г – після клацання Ok виконання скрипту продовжується вже у повністю автоматичному режимі, адже користувач вже ввів всю потрібну програмі інформацію. Проводяться потрібні виміри, дані додаються до графіка і апроксимуються рівнянням Хілла.

Д – скрипт завершується додаванням до графіка текстового ярличка (label), де зазначені величини  $EC_{50}$  і коефіцієнт Хілла  $h$ :



Такий аналіз в напівавтоматичному режимі займає не більше 1 хвилини, якщо ж його робити вручну, то це зайняло б десятки хвилин часу.

Зверніть увагу, що текст після символів // програмою ігнорується. Це зручно для нотаток у процесі написання скрипту. Для найпоширеніших команд LabTalk надаються гіперлінки, щоб можна було з ними швидко ознайомитись.

У скрипті для імпорту даних використовується X-функція Origin `imppClamp`. X-функції (<https://www.originlab.com/doc/en/X-Function/ref>) - це свого роду невеликі блоки кодів для поширених задач, які значно прискорюють і полегшують написання власних скриптів. Вони є у багатьох категоріях, наприклад фітування, аналіз зображень, імпорт та експорт даних, графіка, процесинг сигналів, спектроскопія і статистичний аналіз.

// Початок роботи. Текст у квадратних дужках – це назва нової секції скрипту.

[ImportCurrent]

```
if (fdlog.multiopen.count==0 || fdlog.multiopen.count==1) doc -N;  
//створює новий проєкт  
if (num1==num2 && fdlog.multiopen.count>1) {doc -N;}  
doc -e p {win -cd;} //видаляє всі попередні графіки  
doc -e w {win -cd;} //видаляє всі попередні робочі книги  
if (exist(num1)==4) {run.section(,OpenFile);}  
else { fdlog.usegroup(pClamp);  
fdlog.useType("ABF DAT");  
fdlog.dlgName$ = "Import Multiple pCLAMP Files";  
fdlog.multiopen();  
num2=fdlog.multiopen.count;  
if( num2==0 )  
return;  
num1=0;  
run.section(,OpenFile);} //далі буде виконуватись наступна секція скрипту
```

[OpenFile]

```
if (fdlog.multiopen.count>1) {  
doc -e p {win -cd;}  
doc -e w {win -cd;}  
}  
run.section(,Import); //далі буде виконуватись наступна секція скрипту
```

[Import]

```
newbook; //створює нову книгу
```

```

num1=$(num1+1);
fdlog.get(Q, num1); //imppClamp fname:="%Q" trheader:=trheader
trfiles.Stimulus:=0 trfiles.ShowTimeCol:=1 trFile.bHoldingDuration:=1
impFileSel xfname:=imppClamp fname:="%Q" trFiles:=tt;
tt.Channels.ChannelName2 = 0; //імпортує лише один канал, якщо дані
містять багато каналів реєстрації
tt.Stimulus=0;
tt.ShowTimeCol=1;
tt.bHoldingDuration=1;
imppClamp fname:="%Q" trheader:=trheader trfiles:=tt
Options.sparklines:=0
Options.XFBar:=0;
if(trheader.File1.HeaderInfo.nOperationMode==3) { //тип даних (гар-
free)
dataset=wks.maxrows; //Dataset = кількість рядків даних
if (fdlog.multiopen.count==1)
getnumber -s (Number of data points?) dataset (Enter number);
else dataset=5000;
if (dataset<wks.maxrows) {
%M=%H;
%V=%H_%( %H,@C,2);
newbook Reduced_%( %H, @PL);
wks.col1.name$=Time;
wks.col2.name$=Ch0;
%J=%H_%( %H,@C,2);
copy -s $(dataset) %V %J;
win -a %M;

```

```
win -cd;}} //завершення імпорту даних
%L=%H; //книга даних - %L
run.section(,Plot1); //далі буде виконуватись наступна секція скрипту
```

[Plot1]

```
win -b "Graph$(num1)" "% @AAAdvanced pCLAMP Analyzer\APCA";
%K=%H; //Plot - %K
%H!page.active=1; //обрати шар (layer) 1 графіка
layer -w %N 200; //додати криві потенцілів
layer -at;
layer.y.inc=int((y2-y1)/50)*10;
layer.y.firstTick=y1;

%H!page.active=2; // обрати шар (layer) 2 графіка
layer -w %L 200; //додати криві струмів
@MP=1;
mark -u;
layer -at;
run.section(,AgonistDR); //далі буде виконуватись наступна секція скрипту
```

[AgonistDR]

```
//Ввести назва агоністу та концентрації – до 10
string Agonist$ = "Carbachol";
CCh1 = 1; CCh2 = 3; CCh3 = 10; CCh4 = 30; CCh5 = 100; CCh6 = 300;
CCh7=0; CCh8=0; CCh9=0; CCh10=0;
getn (Agonist) Agonist$
(1st) CCh1
```

(2nd) CCh2

(3rd) CCh3

(4th) CCh4

(5th) CCh5

(6th) CCh6

(7th) CCh7

(8th) CCh8

(9th) CCh9

(10th) CCh10

(Agonist concentrations used (in mkM));

```
Loop (jj,1,10) {
```

```
if (CCh$(jj)==0) break;
```

```
}
```

```
k1=jj; //k1: baseline + number of concentrations used
```

```
win -a %K;
```

```
if (page.nLayers>2) { //цей графік має 2 шари
```

```
%H!page.active=1;
```

```
layer -d;
```

```
}
```

```
%H!page.active=1;
```

```
layer 50 60 8 20;
```

```
layer -at;
```

```
layer.y.inc=nint((y1-y2)/500)*100;
```

```
loop (jj,1,k1) {
```

```
draw -n Line$(jj) -L -h [y2-(y2-y1)*(jj-1)/k1];
```

//додати k1 горизонтальних курсорів для вимірювання амплітуд

```
Line$(jj).color=$(jj);  
Line$(jj).linewidth=1;  
Line$(jj).linetype=2;  
Line$(jj).arrowEndLength=15;  
Line$(jj).arrowEndWidth=15;  
Line$(jj).arrowEndShape=12;  
Line$(jj).vmove=1;  
Line$(jj).keepinside=1;  
Line$(jj).realtime=1;  
}
```

```
work - b DoseResp;
```

```
k2=k1-1;
```

```
loop (jj,1,k2) {
```

```
DoseResp_A[jj+1]=$(CCh$(jj));
```

```
}
```

```
DoseResp_A*=1E-6;
```

```
DoseResp_A=log(DoseResp_A);
```

```
win -a %K;
```

```
%H!page.active=2;
```

```
layer 30 60 67 20;
```

```
%H!page.active=1;
```

//Пауза скрипта для встановлення курсорів

```

/*break -be Position cursors and click on the End button when done;
break -r 1 60;
for (ii=1;ii<61;ii++) {
break -p ii;
sec -p 1;           //пауза 1 с
};*/
run.LoadOC("% @AAdvanced pCLAMP Analyzer\DoneButton.c", 16);
string strLT, strText;
strLT$ = "run.section(% @AAdvanced pCLAMP Analyzer\APCA.ogs,
EventAfterDoneButton, %(Agonist$));";
strText$ = "Position horizontal cursors in the left graph.(LF)Click on the
Done button when done.";
TestDoneButtonDialog( strLT$, "Move Cursors to Position", strText$ );

```

[EventAfterDoneButton]

//Збереження отриманих результатів вимірів

```

loop (jj,1,k1) {
DoseResp_B[jj]=line$(jj).y;
}
DoseResp_B-=DoseResp_B[1];
mark -d DoseResp_B -b 1 -e 1;
DoseResp_B=abs(DoseResp_B);

```

//Сортування результатів по мірі збільшення концентрації агоністу

```

win -a DoseResp;
col(1)=sort(col(1));
col(2)=sort(col(2));

```

//Фітування даних

```
work -b FitCurve;
for (i=1;i<102;i++) {
  FitCurve_A[i]=DoseResp_A[1]-1+(i-1)*(int(DoseResp_A[6])+1-
DoseResp_A[1])/100;}
range fitx=1, fity=2;
win -a DoseResp;
nlbegin (1,2) DoseResp tt;
tt.f_A1=1;
tt.A1=0;    //фіксує A1=0 (за визначенням I=0 якщо [agonist]=0)
nlfит;
fity=fit(fitx);
nlend;
```

//Графічне представлення отриманих результатів

```
win -a %K;
%N!page.active=2;
layer -i201 DoseResp_B;
set DoseResp_B -k 2;
set DoseResp_B -c 1;
set DoseResp_B -z 9;
layer -i200 FitCurve_B;
layer -at;
draw -n MaxValue -L -h $(tt.A2);
Maxvalue.color=4;
maxvalue.linewidth=1;
maxvalue.linetype=2;
```

```

k3=100*DoseResp_B[k2]/$(tt.A2);
label -s -xt (Imax is  $$(k3,.1)$  % of plateau level);
xt.fsize=16;
label -sa -p 10 10 -b 1 -s -n Params (EC $\backslash$ -(50):  $$(10^$(tt.Logx0)*10^6),.1)$ 
\g(m)M \n Hill slope:  $$(tt.p,.2)$ );
Params.fsize=16;
label -s -xb ([% 1] (log M));
y1=0;
y2=$(tt.A2);
layer.y.inc=nint((y1-y2)/500)*100;

```