Міністерство освіти і науки України
Вінницький національний технічний університет

# Методичні вказівки
# до виконання практичних робіт
# з дисципліни
# «Біофізика»
# для здобувачів вищої освіти за спеціальністю
# 163 – Біомедична інженерія







**Developed in the frame of project "Erasmus+ (CBHE) BioArt "Innovative Multidisci-plinary Curriculum in Artificial Implants for Bio-Engineering BSc / MSc Degrees" (586114-EPP-1-2017-1-ES-EPPKA2-CBHE-JP).**

Вінниця
ВНТУ
2020

## *ПРАКТИЧНА РОБОТА № 1*
## ВИРІВНЮВАННЯ БІЛКОВИХ ПОСЛІДОВНОСТЕЙ

*Мета роботи* – навчитись здійснювати пошук та порівняння відомих та нових білкових послідовностей для аналізу їх властивостей, філогенетичної близькості організмів або функціональної подібності пептидних структур.

### Короткі теоретичні відомості

З розвитком методів біохімії та молекулярної біології, а також обчислювальної техніки, в біологічній науці сформувався окремий підхід до вивчення живих систем – *in silico*. Даний підхід передбачає дослідження біологічних систем із застосуванням обчислювальних машин, тобто розглядає живі системи як набір інформації про них. Однією з біологічних наук, що реалізують підхід *in silico*, є біоінформатика.

Предметом біоінформатики є алгоритми обробки даних про структуру біологічних макромолекул. Об'єкт біоінформатики – нуклеїнові кислоти і білки, біологічні макромолекули, структура яких принципово визначає властивості живих організмів і їх частин. З розвитком технологій секвенування, що дозволяють отримувати інформацію про первинну структуру біологічних макромолекул, і методів структурного аналізу, які досліджують їх просторову будову, виникла необхідність в обробці великої кількості відповідних даних. У сучасній біоінформатиці є значна кількість спеціальних методів роботи з біологічної інформацією; в той же час безперервно йде пошук нових методів, підходів і алгоритмів математичного (модельного) аналізу біомолекул.

Робота з первинною структурою біологічних макромолекул – тобто з нуклеотидною або амінокислотною послідовністю – мабуть, найбільш поширена завдача біоінформатики. Цьому сприяє насамперед відносна легкість отримання інформації про первинну структуру нуклеїнових кислот та білків. На порівнянні послідовностей нуклеїнових кислот засновані сучасні методи класифікації живих організмів, а аналіз множини гомологічних білкових послідовностей дозволяє отримати інформацію про структурні і функціональні особливості молекул.

Послідовності білків і нуклеїнових кислот дуже легко можуть бути записані у вигляді тексту, оскільки вони є лінійними та не містять розгалужень. Первинна структура описується як послідовність мономерів: у білків – в напрямку від N-кінця до C-кінця; у нуклеїнових кислот – від 5'-кінця

до 3'-кінця. Послідовності записуються буквами англійського алфавіту в однобуквеному кодуванні (табл. 1).

Таблиця 1 – Кодування та властивості амінокислот

| Однобуквенний код | Трибуквенний код | Повна назва | Властивості (заряд радикала при pH=6–7, хімічні групи) |
|---|---|---|---|
| A | Ala | Аналін | Гідрофобна; незаряджена; аліфатична, метил |
| R | Arg | Аргінін | Основна; несе позитивний заряд; гуанідо |
| N | Asn | Аспарагін | Гідрофільна; незаряджена; амід |
| D | Asp | Аспарагінова кислота | Кислотна; несе негативний заряд |
| V | Val | Валін | Гідрофобна; незаряджена; аліфатичний ізопропіл |
| H | His | Гістидин | Основна; несе позитивний заряд; імідазол |
| G | Gly | Гліцин | Гідрофобна; незаряджена |
| Q | Gln | Глутамін | Гідрофільна; незаряджена; амід |
| E | Glu | Глутамінова кислота | Кислотна; несе негативний заряд |
| I | Ile | Ізолейцин | Гідрофобна; незаряджена; аліфатична |
| L | Leu | Лейцин | Гідрофобна; незаряджена; аліфатична |
| K | Lys | Лізин | Основна; несе позитивний заряд; амін |
| M | Met | Метіонін | Гідрофобна; незаряджена; тіоетерна |
| P | Pro | Пролін | Імінокислота; гідрофобна; незаряджена; іміно, піролідин |
| S | Ser | Серин | Гідрофільна; незаряджена; гідроксил |
| Y | Tyr | Тирозин | Гідрофільна; незаряджена; фенольна (ароматична) |
| T | Thr | Треонін | Гідрофільна; незаряджена; гідроксил |
| W | Trp | Трептофан | Гідрофобна; незаряджена; ароматичний індол |
| F | Phe | Фенілаланін | Гідрофобна; незаряджена; ароматичний феніл |
| C | Cys | Цистеїн | Гідрофільна; незаряджена; тіольна |

***Самостійно повторіть, чим зумовлені гідрофільні та гідрофобні властивості амінокислот.***

Для запису біологічних послідовностей використовується формат FASTA – текст, розмічений особливим чином. Кожна послідовність може мати заголовок, який розпочинається з символу ">". Після переходу на новий рядок записується сама послідовність. Текстовий файл (або текстове поле), що містить кілька послідовностей (з різними заголовками), позначають терміном "multi-FASTA" (рис. 1). Якщо у файлі або текстовому полі міститься тільки одна послідовність, вона може не мати заголовка.

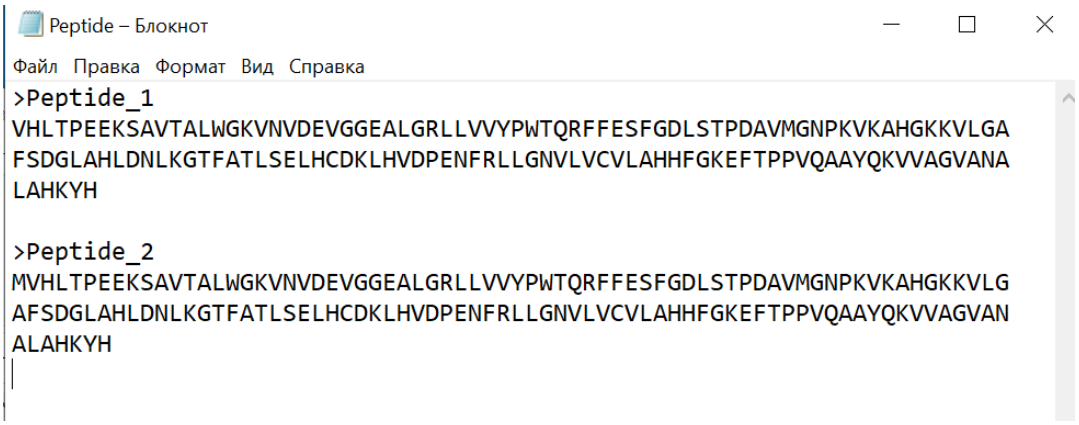



Рисунок 1 – Запис пептидних послідовностей у форматі multi-FASTA

Використовуючи відмінності у молекулярних послідовностях, їх можна застосовувати як джерело даних для класифікації організмів шляхом кількісної оцінки їх схожості. Для кількісної оцінки подібності послідовностей використовують **вирівнювання – розташування послідовностей одна під іншою зі збереженням порядку символів таким чином, щоб досягти збігу або схожості мономерів в найбільшій кількості позицій**. Вирівнювання проводиться шляхом зміщення частин послідовностей одна відносно іншої та додавання розривів в послідовності у разі невідповідності.

При кількісному порівнянні послідовностей біомолекул необхідно враховувати фізико-хімічні властивості мономерів.

Для диференційованої оцінки збігів або розбіжностей різних мономерів використовуються **матриці замін амінокислотних залишків**, побудовані на основі статистики амінокислотних або нуклеотидних замін в гомологічних молекулах з відомою структурою. Числа на перетинах рядків та стовпців показують, наскільки часто в молекулах відбувається заміна одного мономера на інший, а значить, наскільки відповідність цих мономерів у вирівнюванні свідчить про подібність молекул, які порівнюються (рис. 2, 3).

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ala (A) | 4 | | | | | | | | | | | | | | | | | | |
| Arg (R) | −1 | 5 | | | | | | | | | | | | | | | | | |
| Asn (N) | −2 | 0 | 6 | | | | | | | | | | | | | | | | |
| Asp (D) | −2 | −2 | 1 | 6 | | | | | | | | | | | | | | | |
| Cys (C) | 0 | −3 | −3 | −3 | 9 | | | | | | | | | | | | | | |
| Gln (Q) | −1 | 1 | 0 | 0 | −3 | 5 | | | | | | | | | | | | | |
| Glu (E) | −1 | 0 | 0 | 2 | −4 | 2 | 5 | | | | | | | | | | | | |
| Gly (G) | 0 | −2 | 0 | −1 | −3 | −2 | −2 | 6 | | | | | | | | | | | |
| His (H) | −2 | 0 | 1 | −1 | −3 | 0 | 0 | −2 | 8 | | | | | | | | | | |
| Ile (I) | −1 | −3 | −3 | −3 | −1 | −3 | −3 | −4 | −3 | 4 | | | | | | | | | |
| Leu (L) | −1 | −2 | −3 | −4 | −1 | −2 | −3 | −4 | −3 | 2 | 4 | | | | | | | | |
| Lys (K) | −1 | 2 | 0 | −1 | −3 | 1 | 1 | −2 | −1 | −3 | −2 | 5 | | | | | | | |
| Met (M) | −1 | −1 | −2 | −3 | −1 | 0 | −2 | −3 | −2 | 1 | 2 | −1 | 5 | | | | | | |
| Phe (F) | −2 | −3 | −3 | −3 | −2 | −3 | −3 | −3 | −1 | 0 | 0 | −3 | 0 | 6 | | | | | |
| Pro (P) | −1 | −2 | −2 | −1 | −3 | −1 | −1 | −2 | −2 | −3 | −3 | −1 | −2 | −4 | 7 | | | | |
| Ser (S) | 1 | −1 | 1 | 0 | −1 | 0 | 0 | 0 | −1 | −2 | −2 | 0 | −1 | −2 | −1 | 4 | | | |
| Thr (T) | 0 | −1 | 0 | −1 | −1 | −1 | −1 | −2 | −2 | −1 | −1 | −1 | −1 | −2 | −1 | 1 | 5 | | |
| Trp (W) | −3 | −3 | −4 | −4 | −2 | −2 | −3 | −2 | −2 | −3 | −2 | −3 | −1 | 1 | −4 | −3 | −2 | 11 | |
| Tyr (Y) | −2 | −2 | −2 | −3 | −2 | −1 | −2 | −3 | 2 | −1 | −1 | −2 | −1 | 3 | −3 | −2 | −2 | 2 | 7 |
| Val (V) | 0 | −3 | −3 | −3 | −1 | −2 | −2 | −3 | −3 | 3 | 1 | −2 | 1 | −1 | −2 | −2 | 0 | −3 | −1 4 |
| | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y V |

Рисунок 2 – Приклад матриці амінокислотних замін. Чим менше число, тим більші відмінності у властивостях амінокислот

| | A | T | G | C |
|---|---|---|---|---|
| A | 1 | | | |
| T | -10000 | 1 | | |
| G | -10000 | -10000 | 1 | |
| C | -10000 | -10000 | -10000 | 1 |

Рисунок 3 – Приклад матриці нуклеотидних замін у послідовності ДНК

Кількісний показник подібності біологічних послідовностей – рахунок вирівнювання (англ. score) – обчислюється шляхом додавання чисел з матриці замін для кожної позиції. Рахунок – це міра подібності, тобто чим він вищий, тим більш схожі послідовності (рис. 4).




Рисунок 4 – Приклад рахунку вирівнювання для пар олігопептидів. У пари праворуч більше точних збігів амінокислот, однак рахунок вирівнювання більший для пари ліворуч, оскільки амінокислоти замінені подібними їм за властивостями

Якщо вирівнювання містить розриви, вони також повинні бути враховані. Очевидно, що розриви (англ. gaps) – це ознака відмінності в молекулах, тому наявність розривів має знижувати рахунок вирівнювання. Цифра, яку віднімають від рахунку вирівнювання для врахування розривів, називається штрафом за делецію – gap penalty. Довгі розриви в послідовності, що охоплюють кілька позицій, оцінюються як ціле, тому що розриви інтерпретуються як мутації в одній з гомологічних молекул, при чому довгий розрив розглядається як одна мутація, яка зачепила відразу кілька мономерів. Для цього використовуються окремі величини штрафів за «відкриття» (gap open) і за «продовження» (gap extend) розривів (рис. 5). Значення штрафів вибирають, виходячи з характеристик послідовностей, що вирівнюється, особливостей конкретного алгоритму і сформованих практик.

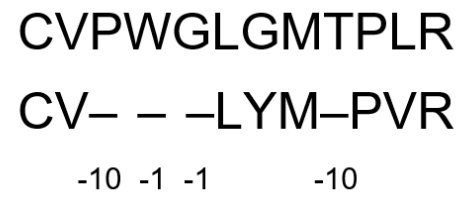



Рисунок 5 – Розриви при порівнянні послідовностей.
Штраф за відкриття розриву –10, за продовження –1

Для здійснення парного вирівнювання, тобто оцінки подібності двох послідовностей, використовуються різні модифікації алгоритму динамічного програмування (алгоритм Нідлмана–Вунша, алгоритм Сміта–Вотермана та ін.), які видають вирівнювання з найбільшим рахунком з усіх можливих (для заданих послідовностей, матриці замін і величин штрафів за розриви). Зі збільшенням числа послідовностей, що вирівнюються, обчислювальна складність (і необхідний час) алгоритму динамічного програмування зростає до неприйнятних величин, тому для множинного вирівнювання – тобто одночасного вирівнювання трьох або більшої кількості послідовностей – використовуються інші, так звані евристичні алгоритми (Clustal, Muscle тощо).

Зовсім особливу групу алгоритмів для роботи з послідовностями складають алгоритми пошуку по банкам даних, серед яких найбільш відомий алгоритм **BLAST (Basic Local Alignment Search Tool) – класичний алгоритм, який дозволяє знайти за заданою послідовністю найбільш подібні до неї серед тих, що наявні у базі даних**.

## Виконання роботи

### 1.1 Пошук і порівняння білкових послідовностей

Одним з центральних ресурсів пошуку інформації в біології є портал Національного центру біотехнологічної інформації США:

https://www.ncbi.nlm.nih.gov/

Перейдіть у веб-браузері за вказаною адресою.

Цей ресурс є величезною збіркою різноманітних інструментів аналізу біологічної інформації, часом здатних замінити десятки спеціалізованих програм. Проте, на головній сторінці ви бачите єдиний рядок пошуку – оскільки **першим кроком до роботи з біологічними даними є пошук інформації в спеціальних базах і банках даних**. В нашій роботі ви будете використовувати пошук за ключовими словами: система буде шукати записи у базах даних, що містять введені вами слова.

Випадний список вгорі сторінки, в якому за замовчуванням обрано “All Databases”, дозволяє вибрати необхідну базу даних (наприклад, Protein – для пошуку послідовностей білків, Nucleotide – нуклеїнових кислот, Structure – просторових структур біомолекул, Pubmed – для пошуку наукових публікацій).

**Знайдіть записи, які стосуються білка гемоглобіну**. Для цього в випадному списку виберіть розділ Protein, впишіть в поле введення Hemoglobin і натисніть Search або Enter на клавіатурі. В результаті на екран будуть виведені посилання на всі записи, що містять вказане слово. **Зробіть скріншот. Поясніть, що означають позначені стрілками записи** (рис. 6).

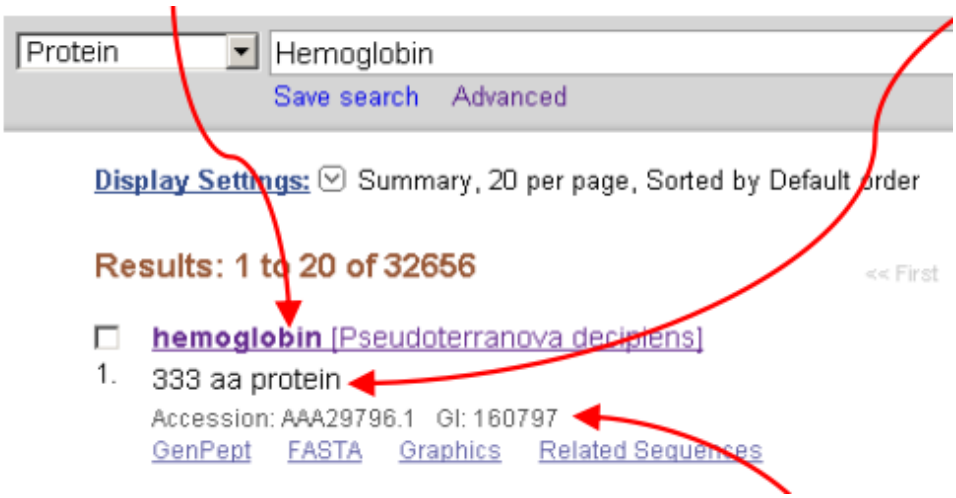



Рисунок 6 – Опишіть, що означають записи, позначені стрілками

Перейдіть за посиланням до першого запису зі списку результатів. Запис складається з полів, позначених великими літерами, які містять інформацію про послідовність. Ознайомтесь зі вмістом полів.

Також можна скористатись розширеним режимом пошуку.

Поверніться на попередню сторінку – до результатів пошуку. Системою були знайдені записи, в яких слово Hemoglobin зустрічається в будь-якому з полів. При такому підході часто буває складно знайти потрібну послідовність серед тисяч нерелевантних результатів. Більш гнучкі можливості пошуку надає розширений режим, в який можна перейти за допомогою посилання Advanced під полем введення.

У розширеному режимі можна вказувати, в яких полях повинні зустрічатися шукані слова. Нехай нам потрібно знайти записи, що містять слово Hemoglobin виключно в заголовку. У першому випадному списку виберіть Title, в поле введення поряд з ним впишіть Hemoglobin. Натисніть Search або Enter, щоб побачити результати пошуку. Порівняйте кількість знайдених записів в цьому випадку і в попередньому.

**В розширеному режимі виконайте пошук записів, які відповідають гемоглобіну людини**. Використовуйте додатковий рядок пошуку і логічний оператор AND.

Перед кожним додатковим рядком можна встановити один з трьох логічних операторів:

- AND – шукати відповідно до вмісту обох рядків – цього і попереднього;
- OR – шукати записи, які відповідають хоча б одному рядку – цьому та попередньому;
- NOT – шукати всі записи, які не містять вказаної в цьому рядку інформації.

**Знайдіть записи, які відповідають гемоглобіну всіх еукаріот, крім людини. Зробіть скріншот**. Коректно виконаний пошук видасть близько 15000 результатів.

**У розширеному режимі складіть запит для пошуку білка глобіну** (title = globin) **людини** (organism = Human) **і корови** (organism = Bos). Зверніть увагу на правильне використання логічних операторів. Коректно виконаний пошук видасть близько 240 результатів. **Зробіть відповідний скріншот**. За необхідності можете скористатись можливістю редагувати пошуковий запит (рис. 7).

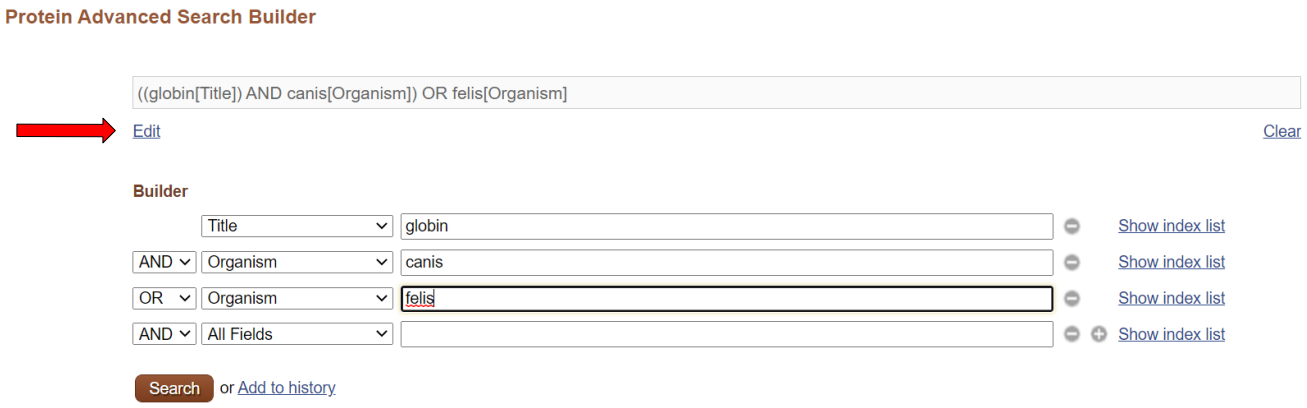



Рисунок 7 – Розширений режим пошуку NCBI

## 1.2 Парне вирівнювання білкових послідовностей

У цій роботі ми використаємо серію парних вирівнювань для вивчення білків глобінів – на прикладі людського і тваринного білків. Виберіть один з видів для дослідження: курка, корова, коза, або інший вид тварин, для яких у базі даних є інформація про структуру α- і β-глобінів (зважайте на те, що далеко не для всіх біологічних видів з'ясована структура глобінів).

Глобіни наявні у багатьох організмів і здатні до утворення четвертинної структури, при цьому існує два типи субодиниць глобінів – α і β. Вважають, що всі глобіни походять від єдиної початкової молекули. В результаті виконання цієї роботи ми з'ясуємо, які глобіни більш подібні – субодиниці різного типу, але з одного біологічного виду, або ж одного типу, хоча і з різних організмів. Відповідно, ми зможемо сказати, яка подія відбулася раніше – поділ предкової молекули на α- і β-субодиницю, або ж поява досліджуваних біологічних видів. Вважається, хоча і з застереженнями, що чим більше відрізняються дві молекули (тобто послідовності), тим більше часу пройшло з їх еволюційного відгалуження – з тих часів, коли вони ще були однією молекулою.

Для виконання вирівнювання білків нам потрібні їх первинні структури, тобто послідовності. Вам необхідно **створити текстовий файл («Блокнот») та скопіювати в нього всі чотири послідовності в форматі Fasta** і використовувати в міру необхідності.

Використовуючи звичайний пошук (розширений тут не знадобиться), знайдіть 4 білкових послідовності – альфа- і бета-глобіни людини і обраної тварини. Приклад запиту для пошуку: «*alpha globin Sus*», де Sus – латинська родова назва тварини.

Дані білки містять 140–150 амінокислотних залишків. Відкрийте знайдені послідовності у форматі Fasta за допомогою відповідного посилання (рис. 8).

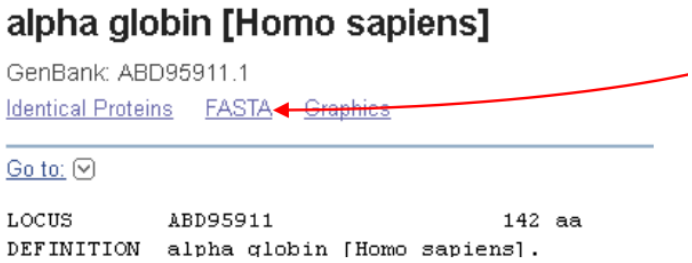



Рисунок 8 – Тут знаходиться інформація про первинну структуру білка

Скопіюйте послідовності в текстовий редактор і перейменуйте, наприклад, так: «beta_globin_Homo» (див. рис. 1). Перейменування (тобто видалення символів пробілів з назви) необхідно для коректної роботи сервісу вирівнювання.

За допомогою серії парних вирівнювань **з'ясуйте, які білки більш подібні**: однакові субодиниці різних організмів, або різні субодиниці одного організму.

Один з найбільш зручних сервісів парного вирівнювання розташовується на сайті Європейського інституту біоінформатики: https://www.ebi.ac.uk/Tools/psa/

Знайдіть на сторінці посилання на сервіс глобального вирівнювання білків з використанням алгоритму Нідлмана–Вунша та перейдіть по ній (рис. 9).

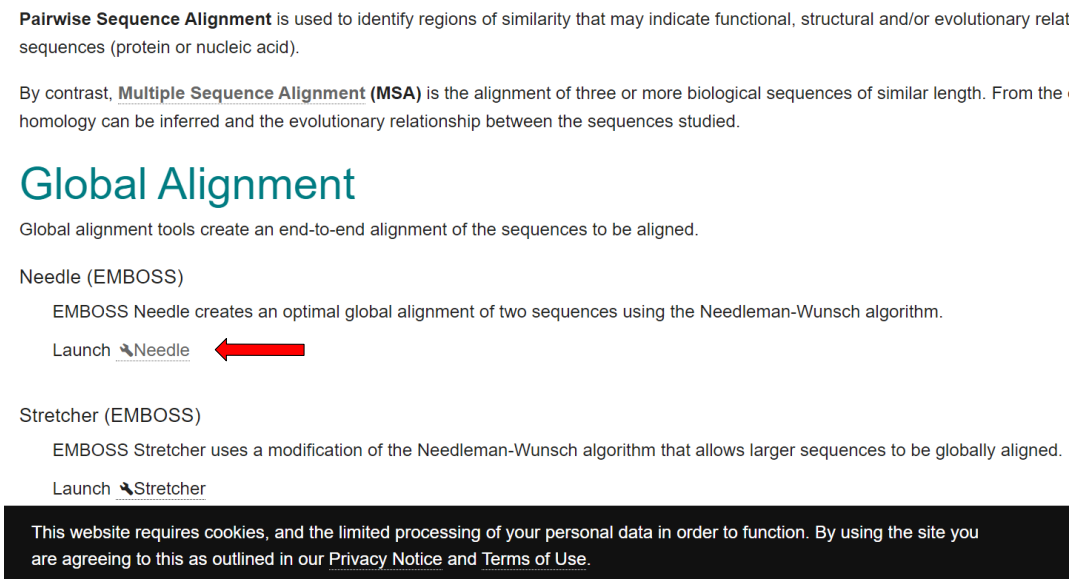



Рисунок 9 – Сервіс парного вирівнювання EBI

Скопіюйте у два текстових поля досліджувані послідовності. Натисніть кнопку «More options», яка дозволяє налаштувати параметри вирівнювання. Параметри вказані у таблиці 2.

Встановіть величину штрафу за відкриття розриву рівною 10, за продовження розриву – 1.

Таблиця 2 – Параметри парного вирівнювання

| Назва параметра | Опис параметра |
|---|---|
| MATRIX | Матриця амінокислотних замін |
| GAP OPEN | Штраф за відкриття розриву |
| GAP EXTEND | Штраф за продовження розриву |
| END GAP PENALTY | Врахування кінцевих розривів, тобто розривів на кінцях молекули. Зазвичай ці розриви не враховуються, значення параметра «false». У цьому випадку величини наступних параметрів не мають значення. |
| END GAP OPEN | Штраф за відкриття кінцевого розриву |
| END GAP EXTEND | Штраф за продовження кінцевого розриву |

Натисніть Submit. За умови коректних вхідних параметрів через деякий час будуть відображені результати. Основні результати показані в таблиці 3.

Таблиця 2 – Параметри парного вирівнювання

| Назва параметра | Опис параметра |
|---|---|
| Length | Довжина вирівнювання |
| Identity | Рівень ідентичності – абсолютна та відносна кількість позицій вирівнювання з однаковими амінокислотними залишками |
| Similarity | Рівень подібності – абсолютна та відносна кількість позицій вирівнювання з подібними амінокислотними залишками |
| Gaps | Кількість розривів в обох послідовностях |
| Score | Рахунок вирівнювання |

Унизу сторінки відображається саме вирівнювання – паралельне розташування послідовностей, між амінокислотними залишками яких відображаються три символи:

| – точний збіг (ідентичність);

: – високий рівень подібності;

. – низький рівень подібності;

(відсутність символу) – амінокислотні залишки не подібні (володіють кардинально різними фізико-хімічними властивостями).

**Виконайте вирівнювання альфа-глобінів людини і тварини з використанням матриць BLOSUM з різними номерами (із випадного списку)**. Заповніть таблицю 3.

Таблиця 3 – Залежність рахунку вирівнювання від матриці замін

| Матриця амінокислотних замін (Matrix) | Ідентичність (Identity) | Рахунок вирівнювання (Score) |
|---|---|---|
| BLOSUM30 | | |
| BLOSUM40 | | |
| BLOSUM50 | | |
| BLOSUM60 | | |
| BLOSUM70 | | |
| BLOSUM80 | | |
| BLOSUM90 | | |

**Зробіть висновок, як номер матриці впливає на рахунок вирівнювання**. **Спробуйте сформулювати принцип вибору коректної матриці для парного вирівнювання**. Обговоріть свої висновки з одногрупниками.

Виконайте вирівнювання інших трьох пар білків, заповнивши таблицю 4. Використовуйте матриці замін BLOSUM з правильними номерами. **Зробіть відповідні висновки**.

Таблиця 4 – Результати парних вирівнювань

| Білкові послідовності, що порівнюються | Рівень ідентичності | Матриця амінокислотних замін | Рахунок вирівнювання |
|---|---|---|---|
| α-глобіни людини і тварини | | | |
| β-глобіни людини і тварини | | | |
| α- і β-глобіни людини | | | |
| α- і β-глобіни тварини | | | |

### 1.3 Множинне вирівнювання білкових послідовностей

Множинне вирівнювання аналогічне парному, адже дозволяє досліджувати одразу кілька послідовностей, однак використовує принципово інші алгоритми. Строго кажучи, аналіз чотирьох послідовностей з попередньої роботи простіше і більш правильно було б здійснити саме за допомогою множинного вирівнювання (хоча результати, звісно, повинні бути однакові).

Множинне вирівнювання є першим кроком філогенетичного аналізу. Крім того, воно дозволяє виявити особливі ділянки в послідовності споріднених білків – ділянки, важливі для формування структури або функціонування молекули (які називають консервативними ділянками).

**Виконайте множинне вирівнювання з α- і β-глобінами людини і тварини**. Відкрийте сторінку множинного вирівнювання на сайті Європейського інституту біоінформатики: https://www.ebi.ac.uk/Tools/msa/

Запустіть метод Clustal Omega за допомогою відповідного посилання («Launch Clustal Omega»).

Скопіюйте в текстове поле (Step 1) порівнювані послідовності у форматі multi-Fasta. Натисніть Submit. За коректних вхідних параметрів через деякий час будуть відображені результати, згруповані в кілька вкладок. **Зробіть скріншот**.

На основній вкладці «Alignmentп» під вирівнюванням розташовується рядок, що показує ступінь подібності залишків у кожній позиції. Символ «зірочка» (*) позначає збіг символів у всіх досліджуваних послідовностях.

Увімкніть відображення кольорів символів за допомогою кнопки «Show colors». Проаналізуйте вирівнювання відповідно до властивостей амінокислот (див. табл. 1) і **заповніть таблицю 5**.

Таблиця 5 – Кольорове позначення амінокислотних залишків

| Колір | Амінокислотні залишки | Ознака, що позначається кольором |
|---|---|---|
| Червоний | | |
| Синій | | |
| Пурпуровий | | |
| Зелений | | |
| Сірий | | |

За допомогою вкладок «Guide tree» та «Phylogenetic tree» можна здійснювати аналіз філогенетичного дерева, побудованого за даними чотирьох послідовностей. **Зробіть відповідні скріншоти**.

**Сформулюйте висновок, чи підтвердились результати попередніх парних вирівнювань? Дайте відповідь, чим відрізняється вміст вкладок «Guide tree» та «Phylogenetic tree»?**

## *ПРАКТИЧНА РОБОТА № 2*
## ВИВЧЕННЯ ПРОСТОРОВОЇ СТРУКТУРИ МАКРОМОЛЕКУЛ

*Мета роботи* – навчитись здійснювати пошук та аналіз просторової структури білкових молекул для дослідження та передбачення їх вторинної та третинної структури, що визначає конформацію та ферментативну і реакційну активність білків.

### Короткі теоретичні відомості

У розвитку біології XX століття важко переоцінити роль фізичних методів, що дозволили встановлювати просторову структуру біологічних макромолекул з точністю до атомів. Завдяки різним методам – від рентген-структурного аналізу і ЯМР до атомно-силової мікроскопії – накопичено велику кількість даних про просторову структуру білків і нуклеїнових кислот та їх комплексів з різними сполуками, що відіграють важливу біологічну роль. Дослідження просторових структур біомолекул дозволяють не тільки краще зрозуміти молекулярну механіку життєвих процесів, але також створювати нові сполуки з заданими біологічними властивостями. Паралельно з розвитком фізичних методів - оскільки отримання просторових структур для деяких молекул досі є вкрай трудомістким процесом - розвиваються і комп'ютерні, математичні методи аналізу одержуваних даних. На відміну від методів роботи з послідовностями, робота з просторовими структурами є більш складною як з обчислювальної точки зору, так і з позиції алгоритмів дії і інтерпретації результатів.

У сучасній хімії існує надзвичайно велика кількість форматів опису просторової структури (конформації) молекул. Проте для зберігання інформації про структуру біологічних молекул в переважній більшості випадків використовується формат PDB (рис. 1). Він передбачає використання текстових файлів (з розширенням «pdb»), розмічених особливим чином. Як правило, більша частина PDB-файлу – а саме рядки, що починаються зі слів «ATOM» і «HETATM» – містить інформацію про позиції атомів в системі трьох координат. Рядки «ATOM» використовуються для опису атомів макромолекул – білків і нуклеїнових кислот; «HETATM» – це атоми інших з'єднань, зазвичай низькомолекулярних, що знаходяться в комплексі з макромолекулою або просто присутніх в складі структури (як молекули води).

```
ATOM   4893  O   ALA B 360      49.951  41.917  16.965  1.00  9.36           O
ATOM   4894  C   SER B 361      49.279  43.271  20.854  1.00 11.33           C
ATOM   4895  CA  SER B 361      49.327  43.061  19.367  1.00 11.24           C
ATOM   4896  CB  SER B 361      48.148  43.776  18.700  1.00 13.43           C
ATOM   4897  N   SER B 361      49.313  41.640  19.097  1.00 10.14           N
ATOM   4898  O   SER B 361      48.840  42.398  21.604  1.00 11.00           O
ATOM   4899  OG  SER B 361      46.909  43.243  19.118  1.00 16.54           O
ATOM   4900  C   CYS B 362      48.605  45.712  22.906  1.00 14.14           C
ATOM   4901  CA  CYS B 362      49.797  44.791  22.660  1.00 12.69           C
ATOM   4902  CB  CYS B 362      51.123  45.451  23.031  1.00 11.63           C
ATOM   4903  N   CYS B 362      49.805  44.414  21.267  1.00 11.41           N
ATOM   4904  O   CYS B 362      47.830  45.963  21.954  1.00 15.40           O
ATOM   4905  OXT CYS B 362      48.417  46.113  24.065  1.00 16.83           O
ATOM   4906  SG  CYS B 362      52.450  44.205  23.117  1.00  9.98           S
TER
HETATM 4907  C1  NAG 1   3      22.652  22.504   7.778  1.00 10.60           C
HETATM 4908  C2  NAG 1   3      22.211  21.391   8.711  1.00 11.16           C
HETATM 4909  C3  NAG 1   3      23.236  20.270   8.687  1.00 13.07           C
HETATM 4910  C4  NAG 1   3      23.586  19.850   7.247  1.00 12.90           C
HETATM 4911  C5  NAG 1   3      23.882  21.062   6.362  1.00 13.35           C
```

Рисунок 1 – Атом № 4899 є атомом кисню (див. символ у кінці рядка) в гамма-позиції серину - 361-го амінокислотного залишку B-ланцюга білка. Цей атом знаходиться в точці з координатами по осі X - 47 Å, Y - 43 Å, Z - 19 Å

Для роботи з просторовими структурами біомолекул використовуються спеціальні програми-переглядачі, або «в'юери» (viewer), що дозволяють представити молекулу у вигляді моделей різного вигляду (рис. 2). Також зазвичай присутній стереоскопічний режим, який полегшує сприйняття конформації молекули (але вимагає деякого тренування очей).

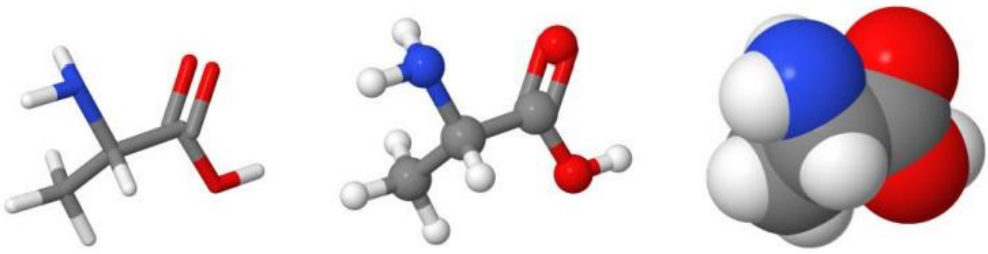


Рисунок 2 – Різні моделі молекули L-аланіну: а – каркасна (скелетна або дротяна – stick), б – кулько-стрижнева (ball and stick), в – модель Ван-дер-Ваальсових радіусів атомів (CPK)

Кожному хімічному елементу можна призначити свій колір. І хоча цю опцію можна змінювати, загальноприйнятим є таке кольорове позначення (рис. 3).




Рисунок 3 – Загальноприйняті кольорові позначення хімічних елементів

Білки та нуклеїнові кислоти мають додаткові модельні позначення для відображення їх вторинної та третинної структури (рис. 4).

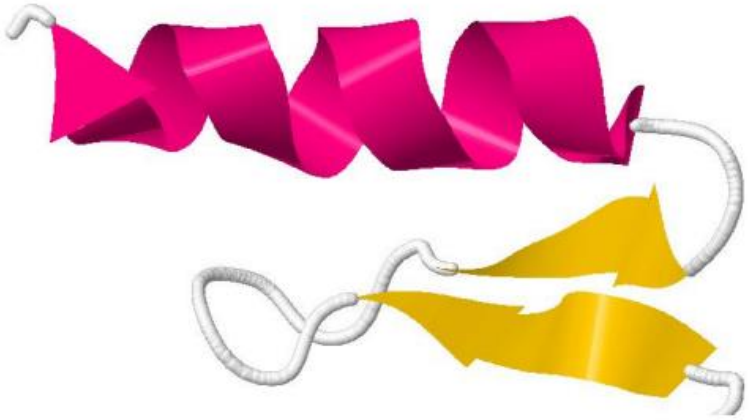


Рисунок 4 – Схематичне зображення елементів вторинної структури білків: угорі – α-спіраль, унизу – β-лист (складчастий шар), сірим кольором позначені петлі та шпильки

## Виконання роботи

### 2.1 Пошук структур макромолекул і файли PDB

Як уже зазначалось, найпоширенішим форматом файлу опису просторової структури біологічних молекул є формат PDB. Файл у форматі PDB – це звичайний текстовий файл з розширенням «pdb», хоча можуть використовуватися і інші розширення. Його можна відкрити будь-яким текстовим редактором.

Центральним сховищем даних про структуру біологічних макромолекул є Protein Data Bank, доступний за адресою: https://www.rcsb.org

Крім того, отримати доступ до цих даних і здійснювати пошук за ними можна на знайомому вам сайті NCBI (база даних Structure).

Унікальним ідентифікатором кожної структури в Protein Data Bank є PDB ID. Даний ідентифікатор складається з чотирьох символів, перший з яких, як правило, цифра; літери в PDB ID прийнято писати великими, хоча більшість програм, які з ними працюють, не чутливі до регістру символів. Приклади PDB ID: 4L9K, 1BTI, 2R33.

Знайдіть за допомогою ресурсу NCBI структуру гемоглобіну (hemoglobin, рис. 5). Список результатів пошуку в розділі (базі даних) Structure містить зображення структур. Відкрийте перший результат пошуку.

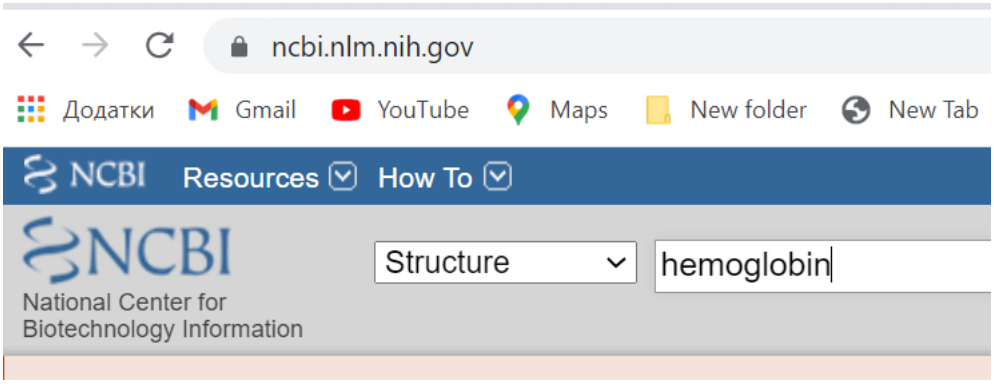



Рисунок 5 – Пошук структури на сервісі NCBI

Розгляньте сторінку, на якій представлена структура молекули.

Розгляньте об'ємну структуру на темному полі (рис. 6). **Що буде, якщо натиснути кнопки "3D view" і "full featured 3D viewer"?**

Розгляньте структуру взаємодій в молекулі (біле поле праворуч від чорного, див. рис. 6). Спробуйте порухати елементи мишкою. Зробіть подвійний клік на одному з квадратиків. **Що відбувається?**

Трохи нижче розміщено блок "Download Stucture Data" (саме так, з помилкою). **Запишіть, у яких форматах можна завантажити структуру?**

Розгляньте розділ "Molecule", розташований нижче на сторінці (див. рис. 6). **Які у ньому є підрозділи?**

**Куди ведуть посилання у табличці?**

Для завантаження структури у форматі PDB з цієї сторінки можна вибрати з випадного списку Format - PDB і натиснути кнопку "Download".

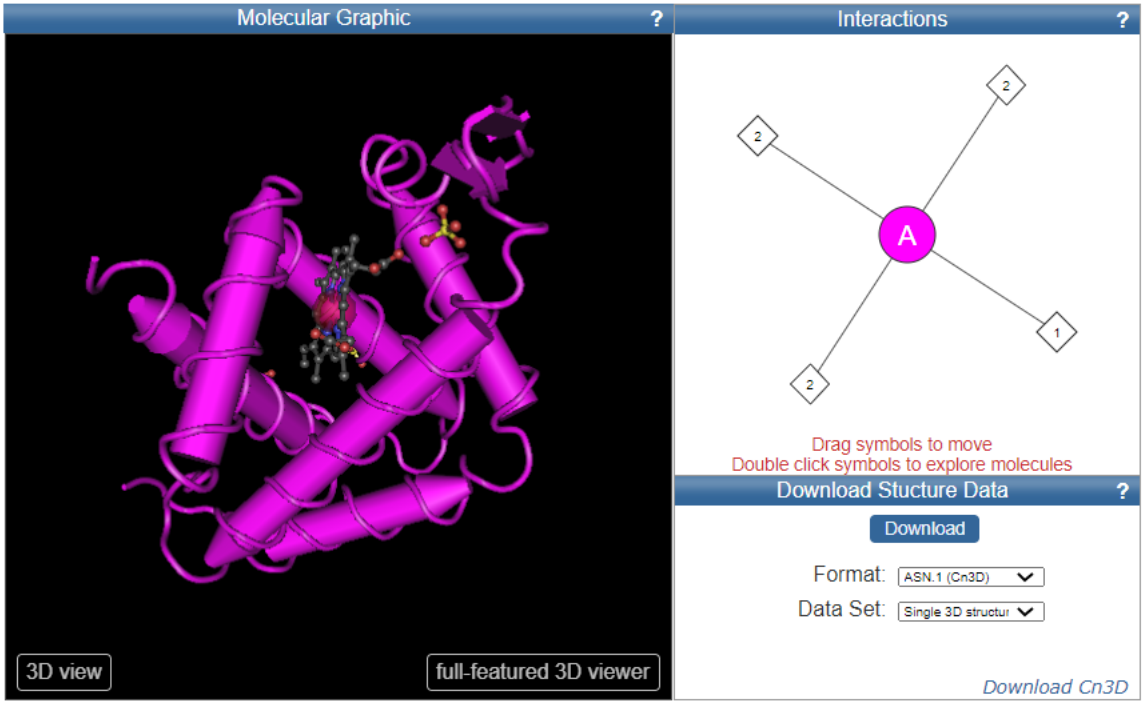
Однак зручніше отримати файл структури з сайту PDB. Перейдіть за посиланням:

https://www.rcsb.org/

На головній сторінці сайту PDB також є рядок пошуку, що дозволяє шукати структури за ключовими словами. **Виконайте пошук за запитом «hemoglobin»**. Ви можете помітити, що список знайдених структур відрізняється від такого на сайті NCBI. Самостійно вивчіть призначення окремих написів, посилань і елементів інтерфейсу на сторінці результатів пошуку.

Відкрийте перший запис із списку результатів. Інформація про структуру розділена на декілька вкладок. Вивчіть вміст вкладок "Structure

summary” та “Sequence”. **Опишіть, яка інформація наведена у цих вкладках.**




**Molecular Components in 3ZHW**

| Label | Count | Molecule |
|---|---|---|
| **Protein (1 molecule)** | | |
| A | 1 | **Non-symbiotic Hemoglobin 1 (Gene symbol: HB1)** 1 Protein; Domain Families; Specific Hits; Super Families; class1_nsHb-like; Globin-like superfamily |
| **Chemicals and Non-standard biopolymers (4 molecules)** | | |
| 1 | 1 | Protoporphyrin Ix Containing Fe |
| 2 | 3 | Sulfate Ion |

Рисунок 6 – Опис структури на ресурсі NCBI

Для завантаження файла структури на локальний комп’ютер натисніть “Download files – PDB Format”. **Збережіть PDB-файл.**

PDB-файл – це звичайний текстовий файл з розширенням «pdb». Як правило, робота з такими файлами здійснюється за допомогою спеціальних програм – переглядачів, або «в’юерів». Однак в деяких випадках корисним буває вміння відредагувати файл структури в текстовому режимі.

Відкрийте завантажений PDB-файл в текстовому редакторі (наприклад, у «Блокноті»). Наша структура містить дві молекули (ланцюга) білка, кожна з

них сполучена з молекулою гема. Розгляньте розділи файлу. Опишіть, яка інформація міститься в розділах PDB-файла «HELIX», «SHEET», «ATOM», «HETATM».

Знайдіть рядки, що описують структуру однієї з молекул гема. Скільки в ній атомів? **Створіть новий PDB-файл, який містить молекули гема.** Найпростіший спосіб зробити це – видалити всі рядки, крім тих, що починаються з HETATM, і зберегти файл (розширення залишається PDB).

### 2.2 Перегляд структур у середовищі Discovery Studio Visualizer

Однією з найбільш зручних якісних програм для аналізу структури молекул є програма «BIOVIA Discovery Studio Visualizer» компанії Dassault Systèmes. Ця програма є безкоштовною і має версії під Windows і Linux. У наших роботах ми розглянемо лише малу частину широкого набору функцій цієї програми.

Завантажити її можна за посиланням:

https://discover.3ds.com/discovery-studio-visualizer-download

При цьому потрібно зареєструватись через електронну пошту та ввести деякі дані про себе.

Для виконання подальших завдань вам знадобиться PDB-файл структури гемоглобіну. За відсутності у вас такого файлу завантажте його з ресурса rcsb.org (PDB ID 4YU3). **Відкрийте файл**. За правильних системних налаштувань він відкриється одразу в програмі Discovery Studio. У разі виникнення проблем з відкриттям файлу ви можете відкрити програму, а потім перетягнути файл на її вікно або вибрати його, обравши “File – Open…” в меню програми.

Вивчіть основні принципи роботи мишкою в цій програмі. Заповніть таблицю 2.1.

Таблиця 2.1 – Керування мишкою в програмі Discovery Studio

| Дія мишкою | Результат |
|---|---|
| Клік по структурі | |
| Клік на порожньому полі | |
| Подвійник кліп по структурі | |
| Рух з натиснутою лівою клавішою | |
| Рух з натиснутою правою клавішою | |
| Рух з натиснутою середньою клавішою (колесом) | |

| | |
|---|---|
| Рух з натиснутою правою клавішою, утримуючи клавішу Shift | |
| Рух з натиснутою правою клавішою, утримуючи клавішу Ctrl | |
| Виділення з утриманням клавіші Shift | |
| Обертання колеса мишки | |

Програма Discovery Studio забезпечує відображення всіх основних моделей візуалізації молекул. За вибір режиму відображення відповідає діалогове вікно Display Style. Його можна відкрити, обравши відповідний пункт контекстного меню (після кліку правою клавішою мишки) або у меню програми “View – Display Style…”

Зверніть увагу: якщо попередньо виділити будь-яку частину структури (виділена частина відображається жовтим кольором), то зміни, зроблені у вікні Display Style, будуть застосовані тільки до виділеної частини молекули. Якщо ніяка частина структури не виділена, зміни застосуються до всієї структури.

Часто в структуру біологічних макромолекул включаються молекули води, які взаємодіють з її гідрофільними частинами. Для більш комфортного перегляду їх можна приховати за допомогою меню “Scripts – Visualization - Show / Hide Waters”.

Відкрийте збережений файл, що містить структуру гемоглобіну. **Відобразіть молекулу гему в режимі Ван-дер-Ваальсових радіусів атомів**: виділіть гем, виберіть у контекстному меню Display Style, в діалозі на вкладці Atom виберіть CPK; для правильного відображення кольорів атомів параметр “Color by” повинен мати значення “Element”. Закрийте діалог, натиснувши OK. **Зробіть скріншот**.

**Відобразіть глобін у вигляді вторинних структур**: для цього виділіть обидві молекули глобіну, виберіть у контекстному меню Display Style, на вкладці Atom виберіть Off; на вкладці Protein – Schematic.

Колір фону можна змінити, обравши у контекстному меню пункт “Color – Background…”

Зніміть всі виділення та **зробіть скріншот отриманої структури**.

### 2.3 Робота з послідовностями у Discovery Studio

У цій частині роботи ви проаналізуєте структуру дисульфідних зв'язків в інгібіторі трипсину (PDB ID 2R33). Вам необхідно завантажити відповідний файл структури. **Відкрийте цей файл за допомогою Discovery Studio**.

Задайте відображення білка у вигляді вторинних структур (“Display Style – Proteine – Solid ribbon”, хоча зазвичай такий режим увімкнений за замовчуванням).

**Відобразіть амінокислотну послідовність** (тобто первинну структуру білка) за допомогою пункту контекстного меню Show Sequence або через меню програми “Sequence – Show Sequence”.

Первинна структура відображається в нижній частині вікна. Розмір цієї частини за замовчуванням, як правило, надлишковий; ви можете зменшити її, потягнувши за верхню межу.

Якщо в програмі Discovery Studio відкрито одразу кілька файлів, вони відображаються в загальному вікні, але в окремих вкладках (ви можете побачити їх у верхній частині вікна). Послідовності також відображаються у вкладках у нижній частині вікна; при цьому перемикання вкладок структур і вкладок послідовностей відбувається незалежно, що може призвести до плутанини. Тому, якщо ви працюєте з кількома послідовностями – уважно стежте, чи належать структура і послідовність до одного і того ж файлу.

Послідовність відображається в однобуквенному форматі. Іноді у файлах структур відсутня «початкова» частина молекули (тобто N-кінець поліпептидного ланцюга або 5'-кінець полінуклеотида); в цьому випадку номери на лінійці не відповідатимуть «справжнім» номерами залишків. Перевірте цей факт на вашій структурі, порівнявши початок послідовності у Discovery Studio та на ресурсах NCBI або PDB. **Наведіть кількість амінокислотних залишків, що відсутні на N-кінці вашої структури. Як ви думаєте, чому вони відсутні?**

Вкладка Sequence дозволяє виділяти окремі літери в послідовності практично так само, як і в звичайному текстовому редакторі. При цьому відповідні ділянки візуалізованої структури також виділяються. І навпаки.

Як у звичайному текстовому редакторі, ви можете шукати окремі літери або їх поєднання за допомогою комбінації клавіш “Ctrl + F”. Ця команда працює, тільки якщо послідовність знаходиться у фокусі введення. Тобто якщо при натисканні Ctrl + F нічого не відбувається, клацніть на послідовність і повторіть спробу.

Як відомо, дисульфідні зв'язки утворюються між залишками цистеїну. **Знайдіть залишки цистеїну в структурі 2R33** за допомогою функції пошуку, описаної вище.

Відобразіть знайдені залишки в режимі кулько-стрижневої моделі (за допомогою діалогового вікна Display Style). **Чи всі залишки цистеїну залучені до формування дисульфідних мостиків?**

**Зробіть скріншот одного з дисульфідних зв’язків між сусідніми β-листами.**

### 2.4 Карти Рамачандрана і вторинна структура білків

Карти Рамачандрана – це спеціальні діаграми для опису значень торсіонних кутів φ і ψ поліпептидного ланцюга. Як правило, карти Рамачандрана описують конформацію цілих молекул білка; кожна точка на карті позначає один амінокислотний залишок. При цьому положення точки по горизонталі визначається кутом φ, а по вертикалі – ψ (рис. 7).

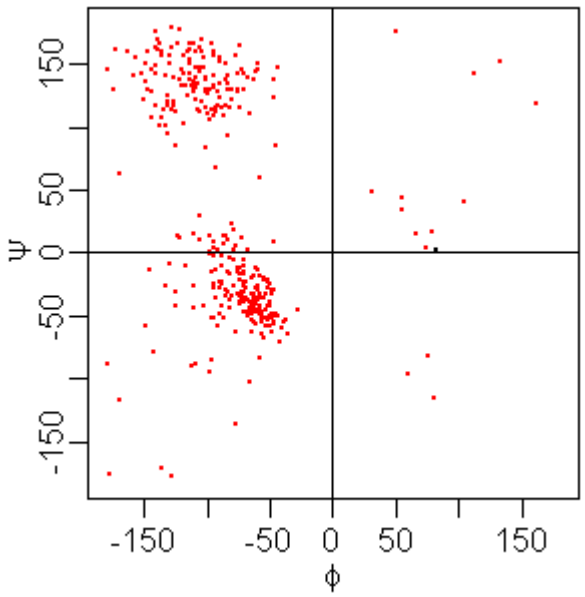



Рисунок 7 – Приклад карти Рамачандрана

Торсіонний кут (навколо хімічного зв'язку) – це двогранний кут, що визначає взаємний поворот частин молекули, які знаходяться по різні сторони від цього зв'язку. Зручною моделлю для демонстрації торсіонних кутів є умовна система з чотирьох атомів (A-B-C-D), як на малюнку нижче.

Торсіонний кут навколо зв'язку B–C – це двогранний кут між площинами (A, B, C) і (B, C, D), що показано на рис. 8.

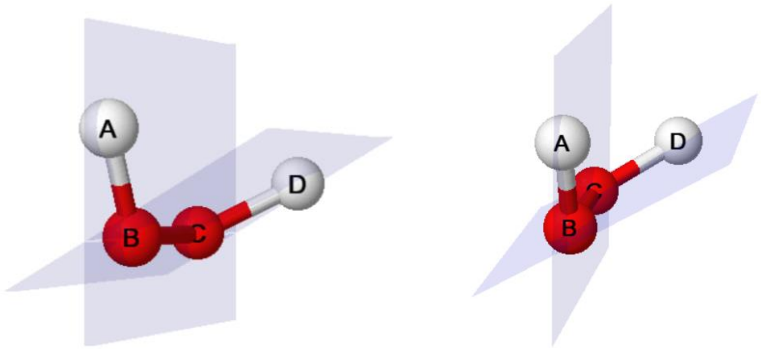



Рисунок 8 – Визначення торсіонного кута в системі чотирьох атомів

Відлік величини кута ведуть від зв'язку, ближчого до спостерігача, до зв'язку, віддаленого від спостерігача. Якщо уявний рух найкоротшим шляхом від ближнього зв'язку до дальнього відбувається за годинниковою стрілкою, кут вважається додатнім; якщо проти годинникової стрілки – від'ємним. Торсіонний кут не залежить від того, з якого боку знаходиться спостерігач - з боку атома «B» чи з боку атома «C». Для структури на рис. 8 він рівний приблизно 40°.

Таким чином, торсіонний кут може набувати значень від –180° до 180°. Природньо, що ці крайні значення відповідають одній і тій самій конформації.

В ситуації, коли від атома «B» або «C» відходять кілька зв'язків, для розрахунку кута вибирають ті, що пов'язані з більш масианим атомом або хімічною групою.

При описі конформації молекули білка прийнято виділяти т. зв. головний ланцюг (атоми, що утворюють «кістяк» поліпептидного ланцюга – атоми N, $C_\alpha$ і C амінокислотних залишків) і бічні ланцюга (радикали амінокислотних залишків).

Кути (як «звичайні», так і торсіонні) прийнято позначати грецькими літерами. Кути визначаються навколо зв'язків, позначених на рис. 9 відповідними літерами. Кути навколо зв'язків, що утворюють головний ланцюг, позначаються літерами φ (фі), ψ (псі) та ω (омега); кути навколо зв'язків в радикалах позначаються літерою χ (хі) з цифровими індексами, відраховуючи від атома $C_\alpha$.

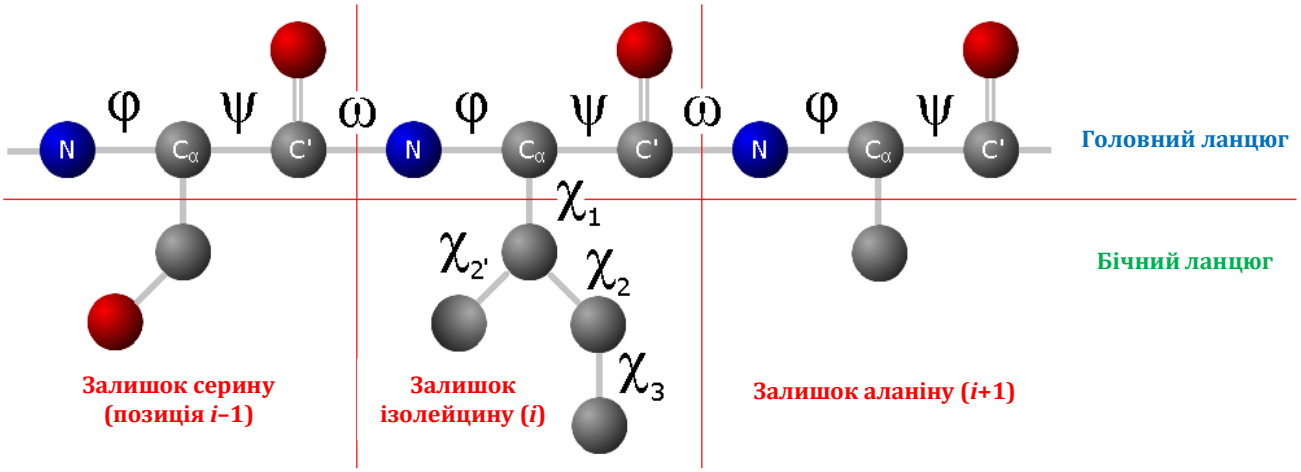



Рисунок 9 – Позначення торсіонних кутів у поліпептиді

Розглянемо докладніше головний ланцюг. Для *i*-го амінокислотного залишку кут φ (навколо зв'язку N–$C_\alpha$) розраховується як кут між зв'язками $C'_{i-1}$–$N_i$ і $C_{\alpha i}$–$C'_i$. Кут ψ (навколо $C_\alpha$–C') – між $N_i$–$C_{\alpha i}$ і $C'_i$–$N_{i+1}$. Іншими словами, всі чотири атома, що визначають торсіонний кут φ або ψ, належать кістяку поліпептидного ланцюга. Такий принцип вірний не тільки «інтуїтивно», а й формально: адже для розрахунку кута ми вибираємо найбільш великі групи або атоми; і в даному випадку дві частини поліпептидного ланцюга, що тягнуться по обидві сторони від нашого кута, можна розглядати як гігантські «хімічні групи».

Кут ω також підкоряється цим правилом, однак його зазвичай не беруть до уваги, оскільки ω – практично константа. У зв'язку з таутомерією пептидної групи обертання навколо пептидного зв'язку ускладнене; атоми пептидної групи (C=O, N–H) лежать в одній площині, в тій же площині лежать і пов'язані з ними $C_\alpha$ з одного боку і N з іншого, а значить, ω = 180°. Отже, кут ω не впливає на конформацію поліпептиду.

Досить очевидно, що просторова структура білкової молекули в цілому може бути визначена як конформація головного ланцюга. Інакше кажучи, вся можлива гнучкість поліпептиду забезпечується обертанням хімічних груп навколо зв'язків N–$C_\alpha$ і $C_\alpha$–C'. А значить, конформацію поліпептиду можна описати як сукупність торсіонних кутів φ і ψ для всіх амінокислотних залишків ланцюга.

Отже запам'ятаємо, що третинна структура білка – тобто конформація глобули – може бути визначена як сукупність торсіонних кутів φ і ψ всіх складових амінокислотних залишків, що входять до його складу.

Слід зауважити, що карти Рамачандрана не визначають третинну структуру білка однозначно, оскільки не враховують порядок з'єднання

амінокислот. Вони лише дозволяють побачити переважну конформацію пептидного ланцюга у білку. На практиці карти Рамачандрана використовуються для демонстрації присутності в білку тих чи інших вторинних і надвторинних структур (структурних мотивів); інакше кажучи, карти Рамачандрана дозволяють виявити так звані регулярні структури.

Регулярні структури в білках – це ділянки поліпептидного ланцюга, в яких амінокислотні залишки мають подібну конформацію (в тому сенсі, що у них приблизно рівні кути φ і ψ; а конформація радикалів не розглядається). Регулярні структури – це різновид вторинних структур. Прикладами регулярних структур є α-спіралі і β-листи.

Значна кількість точок в області карти, що характерна для певної вторинної структури, дозволяє судити про ступінь присутності цієї структури в білку.

**У звіті з лабораторної роботи коротко опишіть карти Рамачандрана і їх застосування в молекулярній біології.**

Для побудови карт Рамачандрана та вивчення принципів структурної організації білків в цій роботі необхідно дослідити такі структури:

- міоглобін (1WLA);
- тріозофосфатізомераза (3UWZ);
- фактор росту фібробластів (1RML);
- пластоціанін (6PCY).

За допомогою онлайн-сервісу http://eds.bmc.uu.se/ramachan.html **побудуйте і занесіть у звіт карти Рамачандрана для знайдених білків**.

Розглянемо схему можливих розташувань точок на карті Рамачандрана (рис. 10). На цій спрощеній діаграмі білі області відповідають конформаціям, в яких відстань між атомами поліпептидного ланцюга більша, ніж їх Ван-дер-Ваальсова взаємодія. Ці області стерично неможливі для всіх амінокислот, крім гліцину, який не має бічного ланцюга. Червоні області відповідають конформаціям, в яких немає стеричних зіткнень, тобто це допустимі області, а саме альфа-спіральні і бета-листові конформації. Жовті області охоплюють допустимі області, якщо в розрахунку використовується трохи коротший радіус дії Ван-дер-Ваальсових сил, тобто якщо атоми можуть наближатися один до одного. Це виявляє додаткову область, яка відповідає лівозакрученій альфа-спіралі.

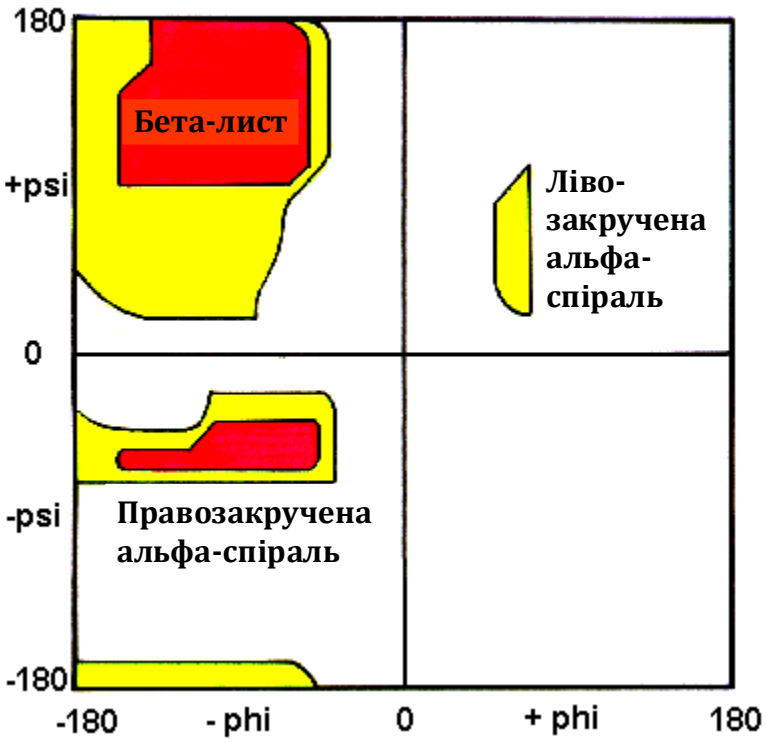



Рисунок 10 – Області карти Рамачандрана

L-амінокислоти не можуть утворювати протяжні області лівої спіралі, але окремі особливі залишки можуть приймати таку конформацію. Ці залишки зазвичай представлені гліцином, але також можуть бути аспарагін або аспартат, де бічний ланцюг утворює водневий зв'язок з основним ланцюгом і таким чином стабілізує цю несприятливу конформацію.

**Проаналізуйте отримані вами діаграми**. Які значення кутів переважають в кожному білку? Зробіть припущення про переважний тип вторинної структури (α-спіралі або β-структури) в кожному з білків. **Заповніть другу колонку табл. 2.2**.

Таблиця 2.2 – Аналіз структури білків

| Білок | Переважний тип вторинної структури | Клас |
|---|---|---|
| Міоглобін | | |
| Тріозофосфатізомераза | | |
| Фактор росту фібробластів | | |
| Пластоціанін | | |

Дослідіть просторову структуру вказаних білків за допомогою програми Discovery Studio. Для цього необхідно завантажити PDB-файли цих структур.

Визначте, до якого класу належить кожен білок (табл. 3.3). **Заповніть третю колонку табл. 2.3**.

Таблиця 2.3 – Класифікація білків за регулярними структурами

| Клас | Регулярні вторинні структури |
|---|---|
| α-білки | Присутні лише альфа-спіралі |
| β-білки | Наявні лише бета-листи |
| α/β-білки | Строге чергування альфа-спіралей і бета-листів |
| α+β-білки | Альфа-спіралі та бета-листи в довільному порядку |
| Нерегулярні білки | Регулярні структури відсутні |

## *ПРАКТИЧНА РОБОТА № 3*
## АНАЛІЗ БІЛКОВИХ ТА НУКЛЕОТИДНИХ ПОСЛІДОВНОСТЕЙ

*Мета роботи* – навчитись визначати закономірності формування просторової структури біологічних макромолекул в залежності від їх послідовностей, освоїти підходи до аналізу просторових структур білків та нуклеїнових кислот; визначати особливі ділянки в молекулі білка та встановлювати зв'язок між їх амінокислотними послідовностями і структурно-функціональною роллю цих ділянок.

### Виконання роботи

### 3.1 Дослідження формування третинної структури білка

Знайдіть за допомогою ресурсу NCBI структуру 1V2W (рис. 1).

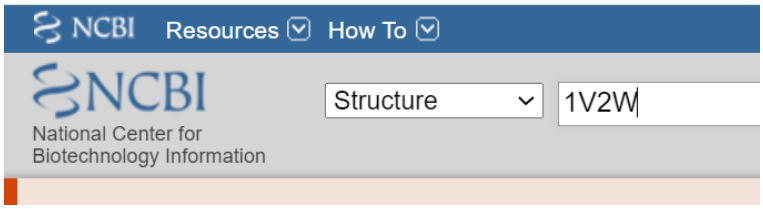



Рисунок 1 – Пошук структури на сервісі NCBI

Відкрийте перший результат пошуку. Одразу завантажте PDB-файл цієї структури – кнопка Download (рис. 2) знаходиться у правій верхній частині сторінки (цей файл знадобиться нам пізніше).

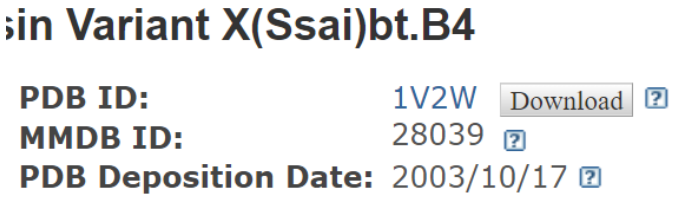



Рисунок 2 – Посилання на PDB-файл структури

Розгляньте відкриту сторінку з об'ємною будовою цього молекулярного комплексу. **Що він собою представляє? Опишіть, що входить до його складу**.

Відкрийте білкову послідовність. Для цього натисніть «1_protein» трохи нижче на сторінці (рис. 3).

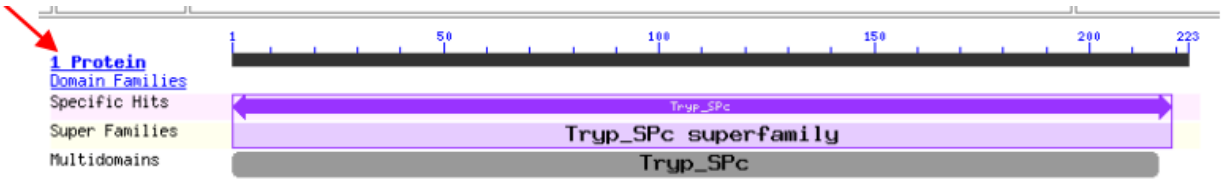



Рисунок 3 – Перехід до перегляду білкової послідовності

Прогляньте сторінку і перейдіть у розділ Graphics (рис. 4).

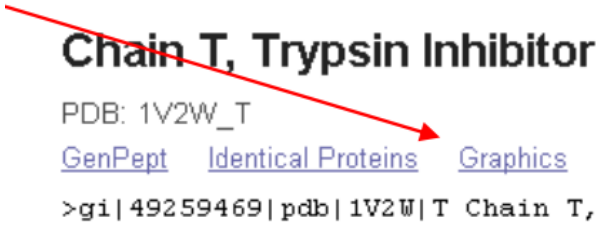



Рисунок 4 – Розділ розширених даних щодо просторової структури білка

Дослідіть представлену на сторінці схему і способи роботи з нею, знайдіть елементи керування масштабуванням і прокручуванням (рис. 5). У верхній частині вказані номери амінокислотних залишків. Збільшуючи масштаб, можна досягти буквенних позначень амінокислот.

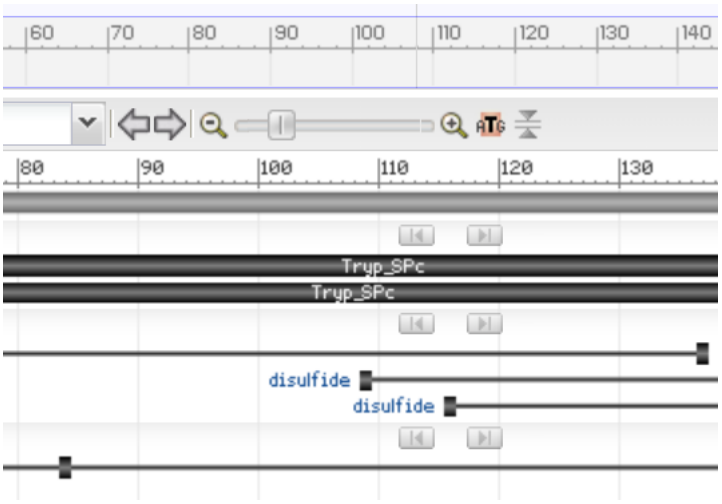



Рисунок 5 – Фрагмент розділу Graphics з елементами керування, номерами амінокислотних залишків та дисульфідними зв’язками

**Визначте, між якими амінокислотами утворені дисульфідні зв’язки.**

На відкритій вами схемі показані в тому числі й амінокислотні залишки, що формують активний центр (active) і центр зв'язування з субстратом (substrate binding site). Визначте номери крайніх залишків, які формують їх (наприклад, перший залишок активного центру має номер 40). **Внесіть**

**порядкові номери та буквенне позначення цих амінокислот у звіт** (всього їх повинно бути 6).

Запустіть виконання алгоритму BLAST для цієї послідовності – для цього у правій верхній частині сторінки передбачена кнопка “Run BLAST”.

На вкладці Graphic Summary поставте галочку біля опції «Show Conserved Domains» (рис. 6).

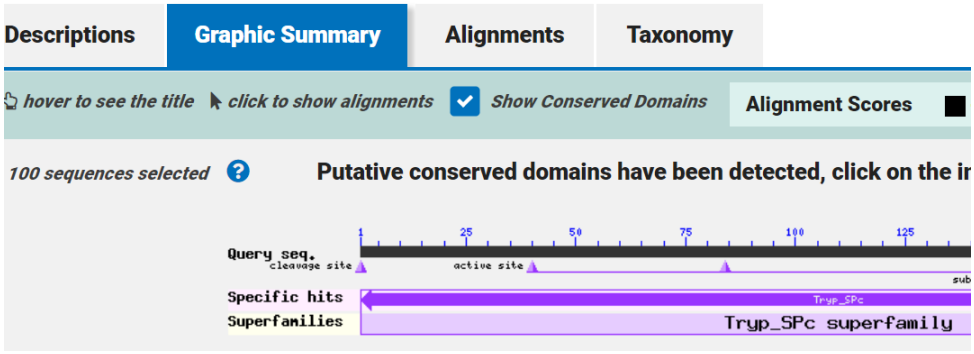



Рисунок 6 – Показ консервативних ділянок в молекулі білка

Відобразиться схема, на якій зокрема будуть показані активний центр і центр зв’язування з субстратом. Натиснувши на фіолетову схему, ви перейдете на сторінку аналізу консервативних доменів. Поставте галочку біля опції «Zoom to residue level» та **перевірте, чи відповідають номери та літери амінокислот** тим, що були визначені раніше.

Для того, щоб перевірити, чи є ці амінокислотні залишки консервативними, **проведіть множинне вирівнювання 100 найближчих гомологів** цього білка.

Для цього зручно скористатись сервісом множинного вирівнювання NCBI, який може бути запущений зі сторінки результатів пошуку BLAST. Натисніть посилання «Multiple alignment» на сторінці результатів пошуку BLAST (а не на тій, де ви тільки що були – рис. 7).

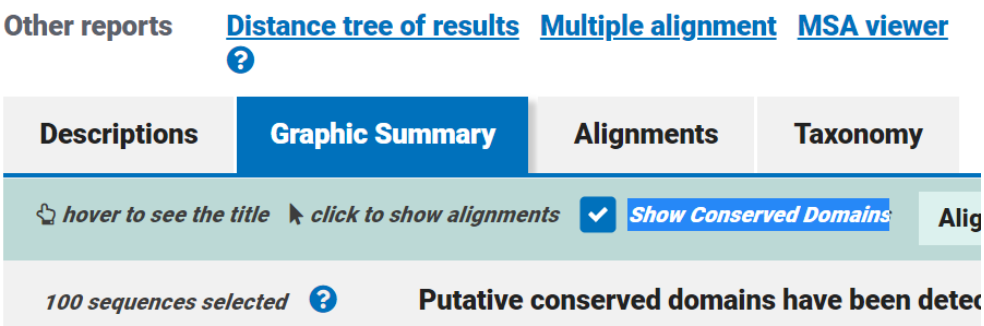



Рисунок 7 – Запуск множинного вирівнювання результатів пошуку BLAST

На сторінці результатів вирівнювання згорніть розділи Graphical Overview та Descriptions, перейдіть до розділу Alignments. Змініть параметр Conservation Setting (встановіть Identity), щоб червоним кольором були виділені тільки консервативні позиції амінокислот (рис. 8), які є однаковими для всіх 100 білків.

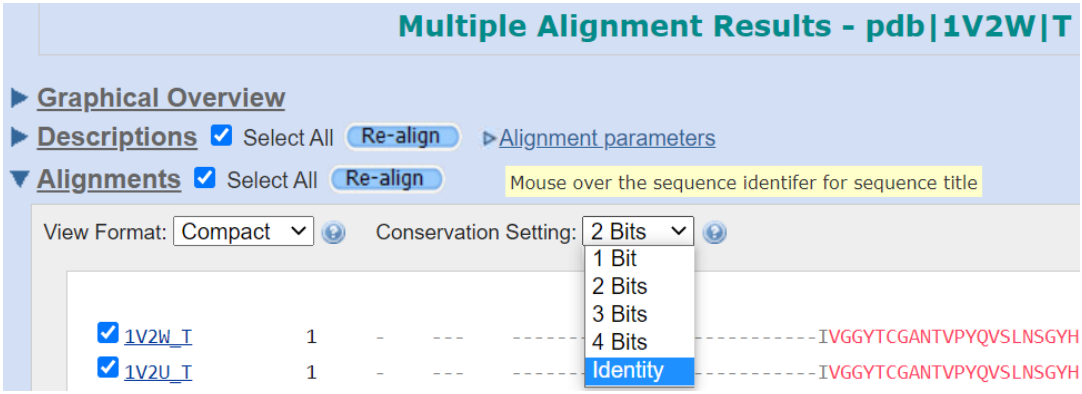



Рисунок 8 – Встановлення відображення консервативних позицій амінокислот

**Визначте, чи є раніше виписані вами амінокислотні залишки консервативними?**

Як можна побачити, досить багато позицій вирівнювання є ідентичними для усіх знайдених білків. Прогляньте послідовність нашого білка (вона показана першим рядком у кожній групі вирівнювання і позначена кодом 1V2W_T).

**Знайдіть і випишіть у ліву колонку табл. 3.1 всі консервативні ділянки, які містять мінімум 5 ідентичних амінокислот поспіль**.

Таблиця 3.1 – Консервативні ділянки в структурі білка

| Консервативні послідовності | Роль і місце у структурі білка |
|---|---|
| YQVSLN | |
| | |
| … | |

Поверніться до завантаженого раніше файла структури 1V2W у форматі PDB і відкрийте його у програмі Discovery Studio. Задайте відображення білка у вигляді Solid Ribbon, а інгібітора – у вигляді Ball & Stick, якщо таке не задано автоматично. Знайдіть в цій структурі ділянки послідовності, визначені вами як консервативні. **Заповніть другу колонку таблиці**. Потрібно вказати, до якої вторинної структури належить ділянка, чи входить вона до складу активних центрів, визначити її положення в глобулі (в глибині або на

поверхні), чи утворює вона дисульфідні мостики, чи контактує ця ділянка з небілковими елементами структури.

### 3.2 Аналіз нуклеотидних послідовностей і структур

Ще одним типом біологічної послідовності є нуклеїнові кислоти.

Розглянемо деякі аспекти роботи з інформацією, яка описує структуру ДНК і РНК, а також особливості формування вторинної та третинної структури РНК.

Вам задана послідовність ДНК, наведена нижче.

gaaccattatgcactcttcaatagttttggccaccgtgctctttgtagcgattgcttcagcatcaaaaacgcgagagctatgcatgaaatcgctcgagca
tgccaaggttggcaccagcaaggaggcgaagcaggacggcatcgacctctacaaacatatgttcgagcactatccagcaatgaagaaatacttc
aagcatcgtgaaaattatacaccggccgatgtccaaaaggatcccttctttattaaacaaggtcaaaatatcttgctcgcctgtcacgttttgtgcgcca
catacgacgatcgtgagacattcgacgcgtacgttggtgagctgatggcacgacacgagcgggaccatgttaaagtaccgaatgatgtttggaatc
acttctgggaacatttcatcgagtttctgggaagtaagaccacgttggacgagccaaccaagcacgcatggcaagagatcggtaaagaattctcac
atgaaatcagccaccacggtcgacattcggttcgcgaccattgcatgaactcgttggagtatatcgcgatcggcgataaggaacatcaaaagcag
aatggcattgacctttacaagcatatgttcgagcattatccacatatgagaaaggcattcaagggacgcgaaaacttcacgaaagaagacgttcaa
aaggacgcattcttcgttaacaaggacacaagattctgttggcccttcgtatgctgtgactcctcatacgatgacgagccaacattcgactattttgttgat
gccctaatggatcgtcatatcaaagatgatattcatctacctcaggaacaatggcatgagttctggaaattgtttgccgaatatttgaacgaaaagagt
caccaacatttgacagaagccgagaaacatgcatggagtacaataggtgaggacttcgcgcatgaggccgataagcatgcaaaggccgaaaa
agaccatcatgaaggagagcacaaagaggaacaccactgaaccaacccgtcgtcgttcaacttaagccttcagcttaagctcgagctaaagcct
cagcttgagctcaatcttatgtcctcaggcctaaacttgaattttaaaagcattttgttgaagcagtgctagccaatctcttatcttatcggtgctattatcaat
ttactctatgccaccccccccccccctctctctgttctctatttgatattctgttcttttagtgccagatgttagtaccagatgttatttctgcataattttcttctcttta
cttcgttatttttcgttcttctatttttatggccaatttttgtgatgtcgaagtcaataaaaccatttttt

**Визначте, чи кодує наведена послідовність який-небудь білок**. Ділянка послідовності ДНК, яка несе інформацію про амінокислотну послідовність білка, називається відкритою рамкою зчитування (Open Reading Frame – ORF).

Для цього скористайтесь сервісом пошуку рамок зчитування NCBI ORF Finder: https://www.ncbi.nlm.nih.gov/orffinder

Опишіть у звіті, **які параметри пошуку можна при цьому задавати?**

Виконайте пошук рамок зчитування в наведеній послідовності з параметрами за замовчуванням.

Розгляньте сторінку результатів пошуку. У верхній частині сторінки знаходиться вже знайома вам «лінійка» послідовності, на яку спроектовані знайдені ORF. Нижче знаходиться таблиця, в якій перераховані знайдені ORF та їх параметри.

Виділіть найдовшу рамку зчитування в одержаному переліку, щоб побачити амінокислотну послідовність, яку вона кодує (ліворуч від списку). Використовуючи навички, набуті під час виконання попередніх робіт, **з'ясуйте, якому білку належить ця послідовність**. Відобразіть це у звіті.

Дослідіть отриманий поліпептид за допомогою сервісу Sequence Manipulation Suite: http://www.bioinformatics.org/sms2/

**Визначте молекулярну масу та ізоелектричну точку**. Посилання на відповідні сервіси розміщені в розділі Sequence Analysis у лівій частині сторінки. **Вкажіть, чому дорівнює середня молекулярна маса одного амінокислотного залишку? Самостійно знайдіть та наведіть визначення ізоелектричної точки білка та опишіть одиницю виміру молекулярної маси протеїнів.**

**Знайдіть в банку PDB структуру 1VTQ**, яка є моделлю молекули транспортної РНК. Збережіть pdb-файл на комп'ютер і відкрийте його у програмі Discovery Studio.

Порівняйте вторинну і просторову структури т-РНК. Знайдіть у просторовій структурі акцепторне стебло (або черенок), D-петлю, TψC-петлю, антикодонову петлю (рис. 9).

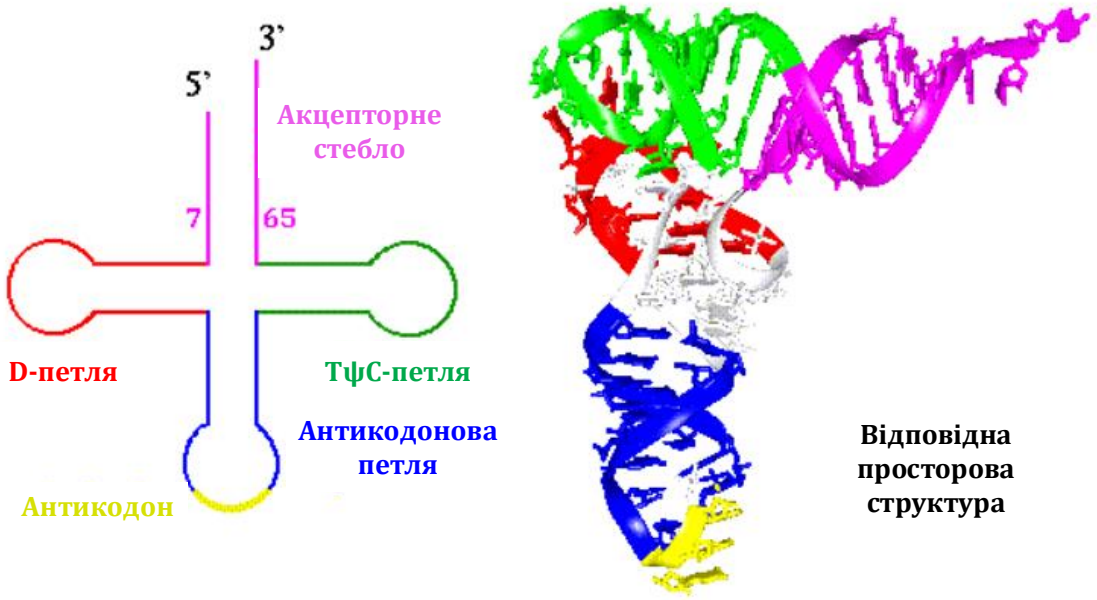



Рисунок 9 – Схематична та просторова будова транспортної РНК

**Визначте номери пар нуклеїнових основ, з яких починаються відповідні спіралі**, тобто перші пари основ цих спіралей (номери для черенка вже показані на рисунку, визначте номери для інших трьох спіралей).

Відобразіть послідовність нуклеотидну послідовність в програмі Discovery Studio, застосувавши команду “Show Sequence”. Зверніть увагу, що послідовність РНК містить літери «N», що позначають модифіковані нуклеотиди. Для коректної роботи сервісу передбачення вторинної структури, який ви будете використовувати в наступному завданні, досить замінити N на

X (тобто позначити їх як невідомі нуклеотиди). Однак біологічно більш правильно буде позначити їх літерами вихідних (не модифікованих) основ, взявши їх зі сторінки структури на сайті PDB.

Поверніться на сторінку, з якої ви завантажили структуру 1VTQ. У вкладці Structure summary, праворуч від назви структури оберіть “Display Files – FASTA Sequence” (рис. 10). Ви побачите послідовність без урахування модифікацій.

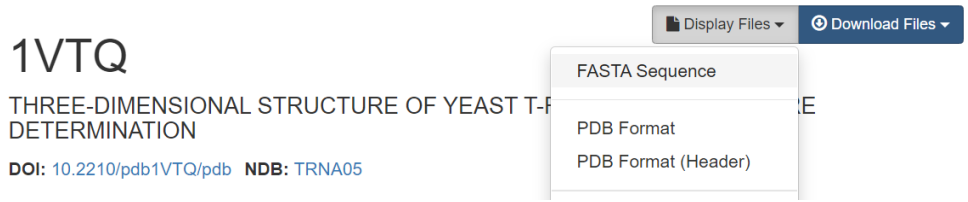



Рисунок 10 – Відкриття нуклеотидної послідовності на сервісі PDB

Щоб побачити, які саме модифікації були виявлені у цій структурі, натисніть на кнопку Structure під графічним зображенням РНК (рис. 11). Кожен модифікований нуклеотид має своє позначення, наприклад, PSU – псевдоуридин.




Рисунок 11 – Посилання для вивчення модифікацій у нуклеотидній послідовності

Для передбачення вторинної структури РНК використовують спеціальні обчислювальні алгоритми, які шукають можливі просторові розташування

нуклеотидного ланцюга та варіанти комплементарного з'єднання пар нуклеотидів

Скористаємось одним з таких сервісів, який розміщений за посиланням:

http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html

Перегляньте вхідні параметри для виконання алгоритму. Скопіюйте вашу послідовність (без урахування модифікацій) у відповідне поле введення і запустіть виконання програми.

Через певний час ви побачите результати роботи сервісу, який запропонує кілька варіантів вторинної структури РНК. **Зробіть скріншот** найбільш відповідної структури. **Порівняйте передбачені початки спіралей (пар основ) з реальними, знайденими вами раніше**.

Прогляньте сторінку результатів. **Коротко опишіть, що на ній відображено. Зробіть відповідні висновки.**

# Список літератури


1. Momand J., McCurdy A., Heubach S. , Warter-Perez N. Concepts in Bioinformatics and Genomics. New York : Oxford University Press, 2017. 504 p.
2. Азнакаєв Е. Г. Моделювання біомедичних процесів. Геноміка : навчальний посібник / МОН України, НАУ. К. : Освіта України, 2011. 428 с.
3. Шмараков І. О., Марченко М. М. Біоінформатика : навчально-методичний посібник. Чернівці : Рута, 2008. 76 с.
4. Огурцов А. Н. Основы биоинформатики : учеб. пособие. Харьков : НТУ «ХПИ», 2013. 400 с.
5. Васильева Н. Ю. Биоинформатика. Множественное выравнивание. Филогенетические деревья : методическое пособие. Одесса : Одесский национальный университет имени И. И. Мечникова, 2014. 70 с.
6. Sung W.-K. Algorithms in Bioinformatics: A Practical Introduction. – Boca Raton : CRC Press, 2009. 407 p.
7. Огурцов А. Н. Выравнивание белковых последовательностей : учеб.-метод. пособие. Харьков : НТУ «ХПИ», 2015. 80 с.
8. Біофізика. Фізичні методи аналізу та метрологія : підручник / Е. І. Личковський, В. О. Тіманюк, О. В. Чалий [та ін.] ; за ред. Е. І. Личковського, В. О. Тіманюка. – Вінниця : Нова Книга, 2014. – 464 с.



***This project has been funded with support from the European Commission. This publication / communication reflects the views only of the author, and the Commission cannot be held responsibility for any use which may be made of the information contained therein***