

Міністерство освіти і науки України  
ДВНЗ «Ужгородський національний університет»  
Медичний факультет  
Кафедра мікробіології, вірусології, епідеміології з курсом інфекційних  
хвороб

# Робочий конспект-зошит для практичних занять з курсу «Молекулярна біологія»

*Навчальний посібник для студентів медичного факультету*



**ДВНЗ «Ужгородський національний університет»  
Медичний факультет  
Кафедра мікробіології, імунології, епідеміології з курсом  
інфекційних хвороб**

**Мірутенко В.В., Коваль Г.М., Карбованець О.І.,  
Попович О.В., Лушнікова О.В., Голомб Л.А.**

**Робочий конспект-зошит  
для практичних занять з курсу  
«Молекулярна біологія»**

*СТУДЕНТА (КИ)*      *КУРС* \_\_\_\_\_  
*група / підгрупа* \_\_\_\_\_  
*факультет* \_\_\_\_\_  
*П.І.П.* \_\_\_\_\_  
*Викладач* \_\_\_\_\_

**Ужгород - 2019**

**УДК 576.851**

Робочий конспект-зошит для практичних занять з курсу «Молекулярна біологія»  
Мірутенко В.В., Коваль Г.М., Карбованець О.І., Попович О.В., Лушнікова О.В., Голомб Л.А.  
«Інвазор», Ужгород: Ужгородський національний університет. 2019.

Методичні вказівки складено у відповідності до вимог типової програми викладання молекулярної біології для студентів медичних спеціальностей, призначені для покращення організації і виконання самостійного вивчення даної теми, згідно вимог Болонського процесу. Наявність контрольних питань, тестів та ситуаційних задач дозволяє в значній мірі активізувати роботу студентів і звести до мінімуму можливість її механічного виконання, оскільки для відповідей на ці запитання необхідне попереднє опрацювання теоретичного матеріалу, а також сприяє більш глибокому осмисленню пройденого матеріалу. Також описано ряд методів, які не виконуються студентами на практичних заняттях, проте з якими їм необхідно ознайомитись для успішного засвоєння курсу молекулярної біології. Видання включає також перелік довідкової вітчизняної, періодичної та зарубіжної літератури, використання якої рекомендується студентам для підготовки до практичних занять по даній дисципліні.

Рецензент : д.м.н., проф. Болдіжар О.О.

д.м.н., проф. Рогач І.М.

## Перелік лекційних тем

1. **Молекулярна біологія, її завдання та методи.** Геноміка, протеоміка і біоінформатика – основні напрямки сучасної молекулярної біології. Методи секвенування. Нанотехнології та їх застосування для секвенування ДНК. **Рівні структурної організації біополімерів.** Білки і нуклеїнові кислоти. Надмолекулярні структури.
2. **Характеристика нуклеїнових кислот** та їх компонентів про- і еукаріотичних організмів. Макромолекулярна структура ДНК. Структура хроматину. Нуклеосоми та їх будова. Гістони. **Генетична функція хромосом.** Картування геномів. Гени, що перекриваються, мозаїчні гени. Інтрони в генах еукаріотів.
3. **Первинна структура білків.** Визначення амінокислотних послідовностей, методи. Просторова структура білків. Основні положення стереохімічної теорії вторинної структури білків.
4. **Експресія генів, транскрипція.** РНК-полімерази: будова, каталітичний механізм. Регуляція транскрипції у бактерій. і еукаріотів.
5. **Біосинтез білків. Трансляція.** Транспортні РНК і аміноацил-тРНК-синтетази. Транспептидація, трисайтова модель функціонування рибосоми. Ініціація і термінація трансляції. Фолдинг білків. Патологічний фолдинг білків, пріони.

## ПРЕДМЕТ І ІСТОРІЯ РОЗВИТКУ МОЛЕКУЛЯРНОЇ БІОЛОГІЇ

Молекулярна біологія є однією з наймолодших галузей біології і вивчає організацію всього живого на найнижчому рівні – рівні молекул. Початок її розвитку припадає на 1950-ті рр., коли вченим вдалося розшифрувати ДНК людини: саме ця нуклеїнова кислота забезпечує зберігання, передачу та реалізацію генетичної програми розвитку живого організму, тобто, простіше кажучи, в ДНК закодовано всі його ознаки. Особливості досліджень в молекулярній біології дають змогу цілеспрямовано, точково аналізувати найрізноманітніші процеси реалізації генетичної програми, виявляти їх порушення, які в майбутньому здатні спричинити появу тих чи інших захворювань.

Молекулярна біологія – комплекс біологічних наук, які вивчають механізми зберігання, передачі та реалізації генетичної (спадкової) інформації, будову, властивості і функції нерегулярних біополімерів – білків і нуклеїнових кислот, а також молекулярних комплексів (молекулярних машин) на їх основі. Виникнувши як біохімія нуклеїнових кислот, молекулярна біологія пережила період бурхливого розвитку власних методів дослідження, якими тепер відрізняється від біохімії.

Досягнення молекулярної біології в пізнанні живої природи важко переоцінити. Великих успіхів вдалося досягти завдяки вдалій концепції досліджень: складні біологічні процеси розглядаються з позиції окремих молекулярних систем, що дозволяє застосовувати точні фізико-хімічні методи дослідження. Це також залучило в цю галузь науки багато великих умів із суміжних напрямів, що також благотворно вплинуло на масштаби і швидкість розвитку наукових знань у цій галузі. Настільки значимі відкриття, як визначення структури ДНК, розшифровка генетичного коду, штучна спрямована модифікація геному,

дозволили значно глибше зрозуміти специфіку процесів розвитку організмів і успішно вирішувати численні найважливіші фундаментальні та прикладні наукові, медичні та соціальні завдання, які ще не так давно вважалися нерозв'язними.

Молекулярна біологія як окремих напрямок біохімії почала формуватися в 30-х роках минулого століття. Саме тоді для поглибленого розуміння феномена життя виникла необхідність у цілеспрямованих дослідженнях на молекулярному рівні процесів збереження і передачі спадкової інформації в живих організмах. Тоді й визначилися завдання молекулярної біології у вивченні структури, властивостей і взаємодії нуклеїнових кислот і білків. Термін «молекулярна біологія» був вперше використаний британським фізиком і молекулярним біологом Вільямом Астбері в контексті досліджень, що стосувалися з'ясування залежностей між молекулярною структурою і фізичними та біологічними властивостями фібрилярних білків, таких, як колаген, фібрин крові або скоротливі білки м'язів.

На зорі виникнення молекулярної біології РНК вважалася компонентом рослин та грибів, а ДНК розглядалася як типовий компонент клітин тварин. Першим дослідником, що довів, що ДНК міститься в рослинах, був радянський біолог, біохімік та основоположник молекулярної біології в СРСР А.М. Белозерський, який виділив у 1935 р. ДНК гороху. Це відкриття встановило той факт, що ДНК є універсальною нуклеїною кислотою, яка присутня у клітинах рослин і тварин.

Серйозним досягненням стало встановлення американськими генетиками Джорджем Бідлом і Едуардом Тейтемом прямого причинно-наслідкового зв'язку між генами і білками. У своїх експериментах вони піддавали клітини нейроспорів (*Neurospora crassa*) ретгеновському опроміненню, що викликало мутації. Отримані результати показали, що це призводило до зміни властивостей специфічних ферментів. У 1940 р. американський біохімік Альбер Клод виділив з цитоплазми тваринних клітин цитоплазматичні РНК, що містили в собі гранули, які були менше мітохондрій. Він назвав їх мікросомами. Згодом при дослідженні структури і властивостей виділених часток була встановлена їх основна роль в процесі біосинтезу білка. У 1958 р. на першому симпозіумі, присвяченому цим частинкам, було прийнято рішення називати ці частинки рибосомами. Ще одним важливим кроком у розвитку молекулярної біології стали опубліковані в 1944 р. дані експерименту американських молекулярного біолога і імунолога Освальда Евері і молекулярних біологів Коліна Маклауда і Макліна МакКарті, які показали, що причиною трансформації бактерій є ДНК. Це був перший експериментальний доказ ролі ДНК у передачі спадкової інформації, розвінчавший існуюче раніше уявлення про білкову природу генів. На початку 50-х рр. британський біохімік Фредерік Сенгер показав, що білковий ланцюг є унікальною послідовністю амінокислотних залишків. В кінці 50-х рр. британські біохіміки Макс Перуц і Джон Кендрю розшифрували просторову будову перших білків. Вже в 2000 р. були відомі сотні тисяч природних амінокислотних послідовностей і тисячі просторових структур білків.

Приблизно в той же час дослідження американського біохіміка Ервіна Чаргаффа дозволили йому сформулювати правила, що описують співвідношення а.п. в ДНК, що допомогло надалі зробити найбільший прорив у молекулярній біології і одне з найбільших відкриттів у біології взагалі. Ця визначна подія відбулася в 1953 р., коли американський і британський молекулярні біологи Джеймс Уотсон і Френсіс Крик, ґрунтуючись на роботах британських біофізика і рентгенографа Розалін Франклін і біофізика Моріса Вілкінса по рентгено-структурному аналізу ДНК, встановили двохспіральну структуру молекули ДНК. Це відкриття дозволило відповісти на принципове питання про здатність носія спадкової

інформації до самовідтворення і зрозуміти механізм передачі такої інформації. Цими ж вченими був сформульований принцип комплементарності а.п., що має ключове значення для розуміння механізму утворення надмолекулярних структур. Це принцип, який застосовується тепер для опису всіх молекулярних комплексів, дозволяє описувати і передбачати умови виникнення слабких (невалентних) міжмолекулярних взаємодій, що обумовлюють можливість формування вторинної, третинної і т.д. структури макромолекул, протікання самозбірки надмолекулярних біологічних систем, що визначають таку велику різноманітність молекулярних структур та їх функціональних наборів. Тоді ж, в 1953 р. був заснований науковий журнал «Journal of Molecular Biology». Його очолив Джон Кендрю, сферою наукових інтересів якого було дослідження структури глобулярних білків. Аналогічний російськомовний науковий журнал під назвою «Молекулярная биология» був заснований в СРСР радянським біохіміком і молекулярним біологом В.О. Енгельгардтом в 1966 р. У 1958 р. Френсіс Крик сформулював центральну догму молекулярної біології – уявлення про необоротність потоку генетичної інформації від ДНК через РНК до білків. У 1970 р. ця догма була кілька поправлена з врахуванням накопичених знань, оскільки незалежно американським генетиком Хоуардом Теміном і американським біохіміком, молекулярним біологом і вірусологом Дейвідом Балтімором було відкрито явище зворотної транскрипції – процес утворення двуланцюгової ДНК на матриці одноланцюгової РНК, який відбувається у онкогенних вірусів. Їми був виявлений фермент зворотна транскриптаза або ревертаза, відповідальний за здійснення цього процесу.

Слід зазначити, що сувора необхідність потоку генетичної інформації від нуклеїнових кислот до білків до цих пір залишається основою молекулярної біології. У 1957 р. радянський біохімік і молекулярний біолог О.С. Спірін спільно з А.М. Белозерським показали, що, при істотних відмінностях у нуклеотидному складі ДНК з різних організмів, склад сумарних РНК подібний. На підставі цих даних вони дійшли сенсаційного висновку про те, що сумарна РНК клітини не може виступати в якості переносника генетичної інформації від ДНК до білків, оскільки не відповідає їй за своїм складом. Разом з тим вони помітили, що існує мінорна фракція РНК, яка повністю відповідає за своїм нуклеотидним складом ДНК і яка може бути істинним переносником генетичної інформації від ДНК до білків. В результаті вони передбачили існування відносно невеликих молекул РНК, які є за будовою аналогами окремих ділянок ДНК і виконують роль посередників при передачі генетичної інформації, що міститься в ДНК, в рибосому, де з використанням цієї інформації здійснюється синтез білкових молекул. У 1961 р. південно-африканським біологом Сіднеєм Бреннером, французьким генетиком і мікробіологом Франсуа Жакобом, американським мікробіологом Метью Месельсоном з одного боку і Ф. Гро, Франсуа Жакобом і французьким біохіміком і мікробіологом Жаком Моно з іншого, вперше отримано дослідне підтвердження існування таких молекул – матричної або інформаційної РНК. Тоді ж вони розробили концепцію і модель функціональної одиниці ДНК – оперона, яка дозволила пояснити, як саме здійснюється регуляція експресії генів у прокариот. Дослідження механізмів біосинтезу білка і принципів структурної організації і роботи молекулярних машин – рибосом – дозволило сформулювати постулат, що описує рух генетичної інформації, званий центральною догмою молекулярної біології.

У 1961 р. і протягом наступних декількох років американськими біохіміками і генетиками Хайнріхом Маттеєм і Маршаллом Ніренбергом, а потім індійським і американськими молекулярним біологом Харом Кораной і біохіміком Робертом Холлі були проведені кілька робіт з розшифровки генетичного коду, в результаті яких було встановлено

безпосередній взаємозв'язок між структурою ДНК і синтезом білків і визначена послідовність нуклеотидів, яка визначає набір амінокислот у білку. Також були отримані дані про універсальність генетичного коду. У 1962 р. генетичний код був розшифрований.

Для розвитку сучасних уявлень про функції РНК вирішальним було відкриття некодуєчих РНК, зроблене за результатами робіт О.С. Спіріна спільно з А.М. Белозерським у 1958 р. і Чарльзом Бреннером із співавторами і американським молекулярним біологом Солом Шпігельманом у 1961 р. відповідно.

Серйозний розвиток отримали способи культивування та гібридизації тваринних клітин. У 1963 р. Франсуа Жакобом і Сіднеєм Бреннером були сформульовані уявлення про реплікон – послідовність невід'ємно реплікуючихся генів, що пояснює важливі аспекти регуляції реплікації генів. У 1967 р. в лабораторії О.С. Спіріна було вперше продемонстровано, що форма компактно згорнутої РНК визначає морфологію рибосомної частинки. У 1967 р. американським біохіміком Артуром Корнбергом, був здійснений синтез *in vitro* біологічно активної ДНК. У 1968 р. було зроблено значне фундаментальне відкриття. Японський біохімік Рейджі Оказакі, виявивши фрагменти ДНК відстаючого ланцюга при дослідженні процесу реплікації, названими на честь неї фрагментами Оказакі, уточнила механізм реплікації ДНК. У 1970 р. незалежно Хоуардом Теміном і Дейвідом Балтімором було зроблено значне відкриття: був виявлений фермент ревертаза, відповідальний за здійснення зворотної транскрипції – утворення дволанцюгової ДНК на матриці одноланцюгової РНК, яке відбувається у онкогенних вірусів, що містять РНК. У цьому ж році, Харой Кораною був проведений хімічний синтез гена.

Ще одним важливим досягненням молекулярної біології стало пояснення механізму мутацій на молекулярному рівні. У результаті серії досліджень були встановлені основні типи мутацій: дуплікації, інверсії, делеції, транслокації та транспозиції. Це дало можливість розглядати еволюційні зміни з точки зору генних процесів, дозволило розробити теорію молекулярних годин, яка застосовується в філогенії.

До початку 70-х рр. були сформульовані основні принципи функціонування нуклеїнових кислот і білків в живому організмі. Було встановлено, що білки і нуклеїнові кислоти в організмі синтезуються за матричним механізмом, молекула-матриця несе в собі зашифровану інформацію про послідовність амінокислот (у білку) або нуклеотидів (в нуклеїновій кислоті). При реплікації або транскрипції такою матрицею служить ДНК, при трансляції або зворотній транскрипції – мРНК. Таким чином, були створені теоретичні передумови для розвитку прикладних напрямів молекулярної біології, зокрема, генної інженерії. У 1972 р. американські біохіміки Пол Берг, Герберт Боер і Стенлі Коен розробили технологію молекулярного клонування. Тоді ними вперше була отримана в пробірці рекомбінантна ДНК. Ці видатні експерименти заклали основи генної інженерії, а цей рік вважається датою народження цього наукового напрямку.

У 1974 р. американськими мікробіологами Гамільтоном Смітом і Даніелем Натансом, та швейцарським мікробіологом і генетиком Вернером Арбером, були відкриті рестриктази. 1978 р. став роком відкриття сплайсингу. Його відкрив американський молекулярний біолог і генетик Філіп Шарп. У 1977 р. Фредерік Сенгер, і незалежно американські біохіміки і молекулярні біологи Аллан Максам і Уолтер Гілберт розробили різні методи визначення первинної структури (секвенування) ДНК. Метод Сенгера, так званий метод обриву ланцюга, є основою сучасного методу секвенування. У 1976 р. Фредерік Сенгер розшифрував нуклеотидну послідовність ДНК фага φ<sub>X174</sub> довжиною 5375 нуклеотидних пар.

1981 р. – серповидноклітинна анемія стає першою генетичною хворобою, що діагностується за допомогою аналізу ДНК. 1982-1983 рр. – зроблено відкриття каталітичної функції РНК в американських лабораторіях канадського молекулярного біолога Сіднея Олтмена і американського молекулярного біолога Томаса Чека. Це змінило існуюче уявлення про виняткову роль білків. За аналогією з каталітичними білками – ензимами, каталітичні РНК були названі рибозимами.

У 1982 р. Томасом Чеком був відкритий автосплайсінг. У 1983 р. американський біохімік і молекулярний біолог Кері Муліс відкрив полімеразну ланцюгову реакцію – ПЛР, завдяки якій стало можливо штучно значно збільшити кількість молекул ДНК в розчині для подальшої роботи. На сьогоднішній день це один з найбільш важливих методів молекулярної біології, що застосовується при вивченні генів, генетичному встановленні особи і встановленні спорідненості, дослідженні спадкових і вірусних захворювань і т.п. У 1990 р. одночасно трьома групами вчених був опублікований метод, що дозволяв швидко одержувати в лабораторії синтетичні функціонально активні РНК (штучні рибозими або молекули, які взаємодіють з різними лігандами – аптамери). Цей метод отримав назву «еволюція в пробірці». А незабаром після цього, в 1991-1993 рр. в лабораторії російського молекулярного біолога А.Б. Четверіна була експериментально показана можливість існування, розвитку і ампліфікації молекул РНК у формі колоній на твердих середовищах.

У 1996 р. британськими генетиками і ембріологами Яном Вілмутом і Кейтом Кемпбелом був успішно проведений експеримент по генетичному клонуванню тварини. В результаті даного експерименту на світ з'явилась вівця Долі – перша тварина, отримана з генетичного коду іншої дорослої істоти шляхом генетичного клонування. Ця визначна подія стала революцією у генній інженерії.

У 1998 р. практично одночасно американські генетики Крейг Мелло і Ендрю Файер описали спостерігаючийся раніше при генних експериментах з бактеріями і квітами механізм РНК-інтерференції, при якому невелика дволанцюгова молекула РНК призводить до специфічного придушення експресії гена. Відкриття механізму РНК-інтерференції має дуже важливе практичне значення для сучасної молекулярної біології. Це явище широко використовується в наукових експериментах в якості інструменту для «виключення», тобто, придушення експресії окремих генів. Особливий інтерес викликаний тим, що цей спосіб дозволяє здійснювати зворотне (тимчасове) придушення активності досліджуваних генів. Ведуться дослідження можливості застосування цього явища для лікування вірусних, пухлинних, дегенеративних і метаболічних захворювань. Слід зазначити, що в 2002 р. були відкриті мутанти вірусу поліомієліту, здатні уникати РНК-інтерференції, тому потрібна ще кропітка робота для розробки дійсно ефективних методів лікування на основі цього явища.

У 1999-2001 рр. декількома групами дослідників визначена з дозволом від 5,5 до 2,4 Å структура бактеріальної рибосоми.

У 2001 р. вченими з міжнародного консорціуму і групи американського генетика і біолога Крейга Вентера у найавторитетніших наукових журналах світу «Nature» та «Science» були опубліковані перші дані про геном людини. Повний геном був опублікован у 2003 р. Публікування було закінчено у 2006 р. До 2006 р. «Protein Data Bank» (PDB) містив дані про просторову структуру 40000 білків.

Протеоміка (англ. *proteomics*, протеомікс) — наука про протеом, що вивчає протеїни в цілому та їх структуру і функції зокрема. Протеїни є однією з важливих складових живих організмів, оскільки вони є головними складовими фізіологічних шляхів метаболізму клітини. Термін протеоміка (англ. *proteomics*) був вперше застосований в 1997 по аналогії



з геномікою (англ. *genomics*), наукою, що вивчає гени. Слово протеом (англ. *proteome*) походить від «протеїн» та «геном», та було запропоновано генетиком Марком Уїлкінсом в 1994 році, під час його роботи над протеїнами та геномом, яку він виконував як PhD-студент (докторант). Основне завдання протеоміки полягає в ідентифікації нових білків і їх кількісному аналізі. Протеом — це сукупність всіх протеїнів та їх модифікацій, що складають протеїновий набір, утворений в результаті діяльності певної біологічної системи. Протеом біологічних систем варіює в часі та під впливом різноманітних факторів, наприклад, стресу, впливу якого може зазнавати як окремі клітини так і весь організм. Протеоміка — міждисциплінарна наука, що виникла як результат науково-дослідного пошуку в рамках Проекту Геному Людини (англ. *Human Genome Project*). Крім того, протеоміка — це галузь наукових досліджень протеому, що стрімко розвивається, та важливий компонент функціональної геноміки.

Незважаючи на те, що протеоміка часто асоціюється з масштабним експериментальним аналізом протеїнів, ця галузь також застосовується для очищення протеїнів та мас-спектрометрії.

Контрольні питання:

1. Що вивчає молекулярна біологія?
2. Коли молекулярна біологія почала формуватися, як окремий напрямок біохімії?
3. Коли і ким була встановлена двоспіральна структура молекули ДНК?
4. Коли і ким була сформульована центральна догма молекулярної біології?
5. Перерахуйте основні відкриття у сфері молекулярної біології.

## ПРАКТИЧНА РОБОТА №1

### Загальна характеристика нуклеїнових кислот та їх компонентів.

Порівняно з білками, вуглеводами та ліпідами нуклеїнових кислот у клітині дуже мало, але їхнє значення настільки велике, що можна сказати: без нуклеїнових кислот життя не можливе. Нуклеїнові кислоти — це інформаційний банк, у якому знаходяться усі відомості про склад, розвиток і функціонування живих систем. У них не тільки зберігається спадковий матеріал, він активно реалізується під час процесів синтезу, розвитку організму, поділу клітин та розмноження.

Уперше нуклеїнові кислоти виділив і описав швейцарський біохімік Фрідріх Мішер, який у 1869 р. вилучив із ядер клітин специфічну речовину і назвав її «нуклеїном» (від лат. *nucleus* — ядро). Дослідження нуклеїнових кислот триває й досі.

Нуклеїнові кислоти — складні високомолекулярні біополімери, мономерами яких є нуклеотиди. Число нуклеотидів у складі однієї молекули нуклеїнової кислоти може сягати 200 млн. Вони можуть існувати вільно і ускладні полімерів — молекул ДНК (дезоксирибонуклеїнової кислоти) і РНК (рибонуклеїнової кислоти).

**Дезоксирибонуклеїнова кислота (ДНК)** — один із двох типів природних нуклеїнових кислот, що забезпечує зберігання, передачу з покоління в покоління і реалізацію генетичної програми розвитку й функціонування живих організмів. Основна роль ДНК в клітинах — довготривале зберігання інформації про структуру РНК і білків.

У клітинах еукаріотів (наприклад, тварин, рослин або грибів) ДНК міститься в ядрі клітини в складі хромосом, а також в деяких клітинних органелах (мітохондріях і пластидах). У клітинах прокаріотів (бактерій і архей) кільцева або лінійна молекула ДНК, так званий нуклеоїд, міститься в цитоплазмі й прикріплена зсередини до клітинної мембрани. У них і у нижчих еукаріотів (наприклад дріжджів) зустрічаються також невеликі автономні кільцеві молекули ДНК, так звані плазміди. Крім того, одно- або дволанцюгові молекули ДНК можуть утворювати геном ДНК-вірусів.

З хімічної точки зору ДНК — це довга полімерна молекула, що складається з послідовності блоків — нуклеотидів. Кожний нуклеотид складається з азотистої основи, цукру (дезоксирибози) і фосфатної групи. Зв'язки між нуклеотидами в ланцюгу утворюються дезоксирибозою й фосфатною групою. У переважній більшості випадків (окрім деяких вірусів, що містять одноланцюгові ДНК) макромолекула ДНК складається з двох ланцюгів, орієнтованих азотистими основами один проти одного. Ця дволанцюгова молекула утворює спіраль. У цілому структура молекули ДНК отримала назву «подвійної спіралі». У ДНК зустрічається чотири види азотистих основ (аденін, гуанін, тимін і цитозин). Азотисті основи одного з ланцюгів сполучені з азотистими основами іншого ланцюга водневими зв'язками згідно з принципом комплементарності: аденін з'єднується тільки з тиміном, гуанін — тільки з цитозином.

Послідовність нуклеотидів дозволяє «кодувати» інформацію про різні типи РНК, найважливішими з яких є матричні (мРНК), рибосомні (рРНК) і транспортні (тРНК) та інші некодуючі РНК. Крім кодуючих послідовностей, ДНК клітини містить некодуючі послідовності, що виконують регуляторні та структурні функції, або не виконують ніяких функцій. Ділянки кодуючих послідовностей разом із регуляторними ділянками називаються генами. Сукупність всіх генів, регуляторних послідовностей, некодуючих послідовностей, тобто вся нуклеотидна послідовність ДНК незалежно від її функцій, формує геном організму.

У геномах еукаріотів містяться також довгі послідовності без очевидної функції (некодуючі послідовності). Також у складі геному досить поширені генетичні паразити — транспозони та вірусні або схожі на них послідовності. Проте організм може використовувати транспозони для виконання певних функцій, також транспозони можуть впливати на еволюцію генів.

Розшифровка структури ДНК, виконана в 1953 році, стала одним з поворотних моментів в історії біології. За видатний внесок у це відкриття Френсісу Кріку, Джеймсу Ватсону і Морісу Вілкінсу була присуджена Нобелівська премія з фізіології або медицини 1962 року.

**Рибонуклеїнові кислоти (РНК)** — органічні біополімери, які входять до складу всіх живих клітин і складаються з великої кількості рибонуклеотидів, з'єднаних між собою фосфодієфірними зв'язками. РНК містять азотисті основи пуринового і піримідинового ряду (аденін, гуанін, цитозин, урацил), пентозу (рибозу) та залишок фосфорної кислоти. У деяких РНК є, крім постійних азотистих основ, і їх похідні, або «мінорні» основи, наявність яких зумовлює специфічність функцій РНК. Молекули РНК, як правило, побудовані з одного полінуклеотидного ланцюга; дволанцюгову структуру мають лише РНК деяких вірусів. Окремі ділянки молекули РНК утворюють спіралізовані петлі — «шпильки» — за рахунок водневих зв'язків між комплементарними азотистими основами А-У та Г-Ц. Наявність спіралізованих ділянок характерна для всіх типів РНК. РНК клітини відрізняються нуклеотидним складом, розмірами, локалізацією в клітині та функціями.

Існують три головні типи РНК: інформаційні, або матричні (мРНК), рибосомні (рРНК) і транспортні (тРНК). У еукаріотів є ще клас малих ядерних РНК (мяРНК), які впливають на процесинг РНК, та ядерні попередники мРНК — гетерогенні ядерні РНК (гяРНК).

РНК є основним генетичним матеріалом у деяких вірусів тварин та рослин. Синтез усіх видів РНК клітини здійснюється на ДНК під час транскрипції. Реакції синтезу каталізує фермент ДНК-залежна-РНК-полімераза. Синтезована молекула РНК комплементарна фрагменту ДНК-матриці, оскільки порядок включення нуклеотидів у ланцюг РНК визначається послідовністю нуклеотидів у матриці ДНК. Молекули РНК синтезуються, як правило, у вигляді попередників, які внаслідок процесингу перетворюються на зрілі молекули РНК.

- Завдання:**
1. Користуючись муляжами і таблицями, вивчити будову молекул ДНК.
  2. Замалювати загальну схему будови ДНК (фрагмент молекули ДНК з трьох нуклеотидів), відмітити комплементарні азотисті основи в молекулі ДНК.
  3. Розглянути при малому та великому збільшенні мікроскопа мікропрепарати печінки пацюків. Визначити локалізацію ДНК в клітині.
  4. По таблицях вивчити будову різних типів РНК.
  5. Замалювати схему будови м-РНК і т-РНК.

**Питання для самопідготовки:** Нуклеїнові кислоти, їх значення. Історія вивчення нуклеїнових кислот. Типи нуклеїнових кислот, їх класифікація. Локалізація нуклеїнових кислот в клітині.

## ПРАКТИЧНА РОБОТА №2

### Характеристика ДНК та РНК про- та еукаріотичних організмів.

Нуклеїнові кислоти — полімерні молекули, мономерами яких є **нуклеотиди**. Нуклеотид, в свою чергу, складається з трьох компонентів: залишків нітратної основи, вуглеводу (пентози) та залишку ортофосфатної кислоти.

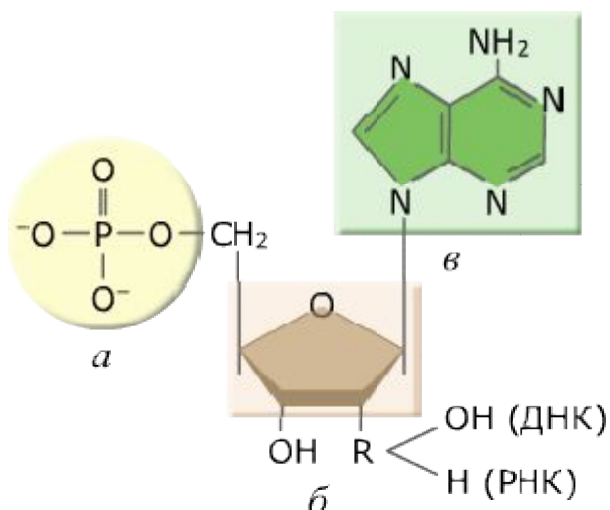


Рис. 2.1. Загальна структурна формула нуклеотиду:

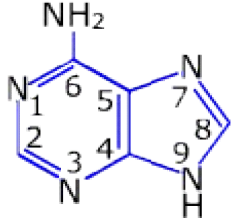
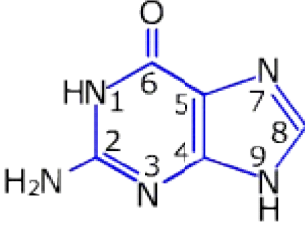

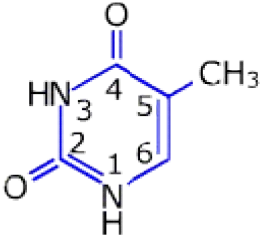
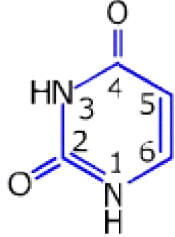
*a* — ортофосфатна кислота, *б* — п'ятиуглецевий моносахарид (пентоза), *в* — нітратна основа.

Залежно від виду пентози, що входить до складу нуклеотиду, розрізняють два типи нуклеотидів: рибонуклеотиди, що мають пентозу **рибозу**, і дезоксирибонуклеотиди, що мають пентозу **дезоксирибозу**. Рибонуклеотиди утворюють полімерні нуклеїнові кислоти РНК, а дезоксирибонуклеотиди — ДНК.



Нуклеотиди можуть відрізнятися не тільки за видом пентози, а й за залишками нітратних основ (Табл. 2.1). Нуклеотиди з аденіном, гуаніном і цитозіном входять до складу як ДНК, так і РНК. Проте тимін міститься лише в нуклеотидах ДНК, а урацил — лише в РНК (Табл. 2.2).

Таблиця 2.1

Назва залишку нїтратної основи	Позначення для відповідних нуклеотидів	Структурна формула нїтратної (азотистої) основи
<b>Пуринові</b>		
Аденін	А	
Гуанін	Г	
<b>Піримідинові</b>		
Цитозин	Ц	
Тимін	Т	
Урацил	У	

## Відмінності у складі ДНК та РНК за азотистими основами

ДНК	РНК
А	А
Г	Г
Ц	Ц
Т	У

Молекула РНК — це нерозгалужений ланцюг залишків нуклеотидів, які з'єднані один з одним послідовно ковалентним полярним зв'язком. Цей зв'язок утворюється між залишком пентози одного нуклеотиду і залишком ортофосфатної кислоти іншого нуклеотиду (Рис. 2.2).

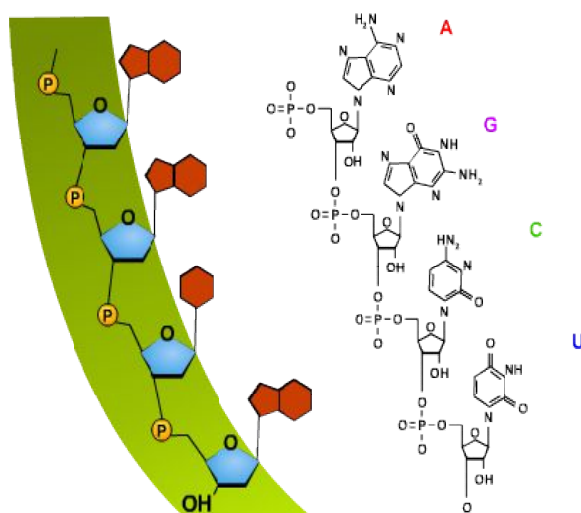


Рис. 2.2. Схематична будова фрагменту РНК

РНК — нерегулярний полімер, тобто нуклеотиди різного виду чергуються нерегулярно: ...АГЦГУАГЦЦАУ... РНК зустрічається в усіх живих організмах: вірусах, прокаріотах та еукаріотах. Існують декілька видів РНК, розглянемо їхню будову, функції та місцезнаходження у клітині.

**Інформаційна (або матрична) РНК (іРНК, або мРНК)** може знаходитись у цитоплазмі, ядрі, мітохондріях, хлоропластах, утворювати з рибосомами комплекс — полісому (Рис. 2.3). Основні функції іРНК — перенесення генетичної інформації від основного джерела (як правило, від ДНК) до місця синтезу білка та безпосередня участь у синтезі білкових молекул.

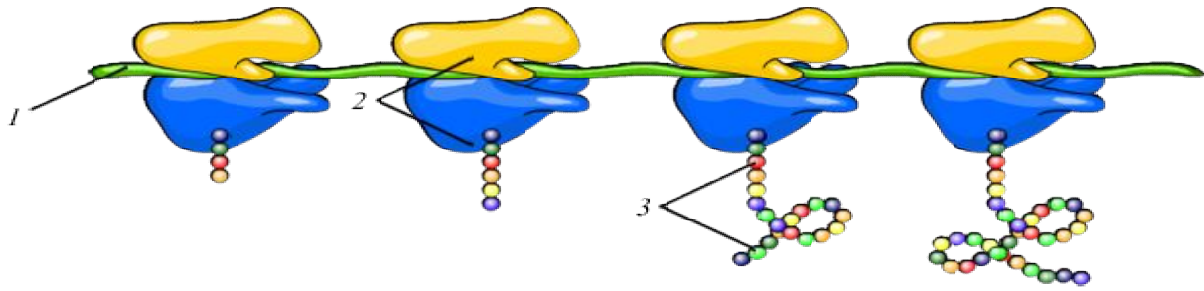


Рис. 2.3. Полісома:  
1 — іРНК, 2 — рибосома, 3 — поліпептид

**Транспортна РНК (тРНК)** міститься в цитоплазмі, мітохондріях і пластидах, має найменші розміри серед усіх молекул РНК (складається з 70-90 нуклеотидів). Основна функція тРНК — переносити амінокислоти до рибосом, на яких відбувається синтез білкових молекул. Кожен вид тРНК високо специфічний, тобто переносить тільки конкретну амінокислоту. Хоча до складу білка входить 20 амінокислот, існує 60 видів тРНК (декілька видів тРНК можуть переносити один вид амінокислот).

Транспортна РНК має *вторинну структуру*, що підтримується водневими зв'язками і формою нагадує листок конюшини (Рис. 2.4).

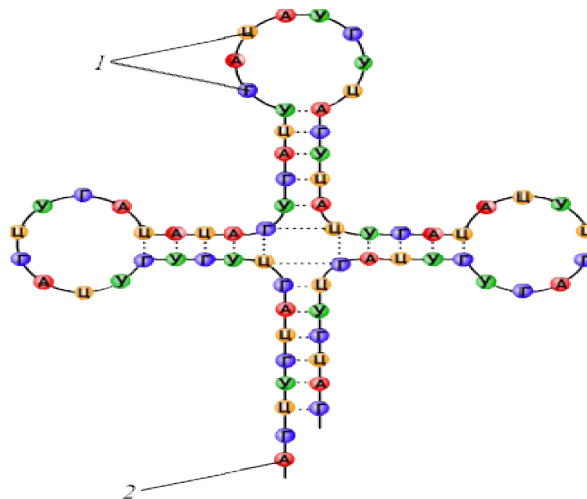


Рис. 2.4. Вторинна структура тРНК:  
1 — антикодон, 2 — місце приєднання амінокислот

Біля верхівки «листка» містяться три нуклеотида, які відповідають певній амінокислоті за генетичним кодом. Вони називаються *антикодоном*. А з протилежного боку, біля основи молекули тРНК, є ділянка, до якої приєднується амінокислота. Молекула тРНК може утворювати *третинну структуру*, що нагадує латинську літеру «L» (Рис. 2.5).

**Рибосомна РНК (рРНК)** є найбільшою (3-5 тис. нуклеотидів), вона є складовою рибосом і разом з білками забезпечує певне розташування іРНК і тРНК під час синтезу білкової молекули. рРНК становлять близько 80% усієї РНК клітини; вони локалізовані в основному в цитоплазмі, а також у мітохондріях. рРНК разом зі специфічними білками становлять основу структури рибосом — нуклеопротейінових структур, призначених для синтезу білка по РНК-матриці. Просторова структура рРНК повністю не вивчена. Кожна з рРНК має специфічну тривимірну структуру, зумовлену характером внутрішньо-

молекулярної взаємодії азотистих основ. Установлено, що в одноланцюгових молекулах рРНК є багато спіралізованих ділянок («шпильок»).

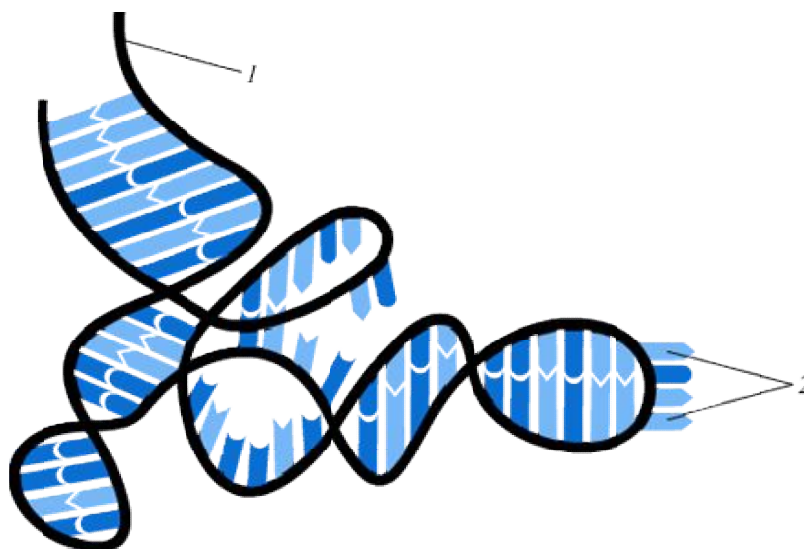


Рис. 2.5. Третинна структура тРНК:  
1 — антикодон, 2 — місце приєднання амінокислоти

**Будова ДНК.** ДНК у еукаріотів міститься в ядрі, пластидах і мітохондріях клітин, а у прокаріотів, які не мають оформленого ядра, — входить до складу *нуклеоїду*.

Розшифрування структури нуклеїнових кислот та розуміння їхнього значення для спадкової передачі ознак та властивостей організмів стало значною подією, яка відкрила нові горизонти у розумінні сутності життя і була відзначена Нобелівською премією 1962 року. Її отримали англійський фізик Ф. Крік і американський біохімік Дж. Уотсон.

ДНК подібна до РНК, але має більш складну будову. Вона складається з двох ланцюгів, кожен з яких принципово схожий на ланцюг РНК, з'єднаних між собою водневими зв'язками. Кожен нуклеотид одного ланцюга має водневі зв'язки з нуклеотидом другого ланцюга. Ці зв'язки виникають між азотистими основами нуклеотидів точно за правилом — залишок аденіну нуклеотиду одного ланцюга сполучається із залишком тиміну нуклеотиду іншого ланцюга, утворюючи три зв'язки, а залишок гуаніну — із залишком цитозину, утворюючи два зв'язки. Така закономірність називається *принципом, або правилом, комплементарності* (Рис. 2.6).

**Принцип комплементарності** — це закономірність розташування нуклеотидів у ДНК або під час синтезу РНК, згідно з якою гуаніну одного ланцюга завжди відповідає цитозин другого ланцюга, а аденіну одного ланцюга — тимін другого ланцюга ДНК або урацил РНК. Завдяки здатності утворювати водневі зв'язки аденіну з тиміном (урацилом), а гуаніну — з цитозином, нуклеотиди наче «доповнюють» один одного.

Отже, знаючи послідовність нуклеотидів одного ланцюга ДНК, можна абсолютно точно визначити послідовність для другого ланцюга.



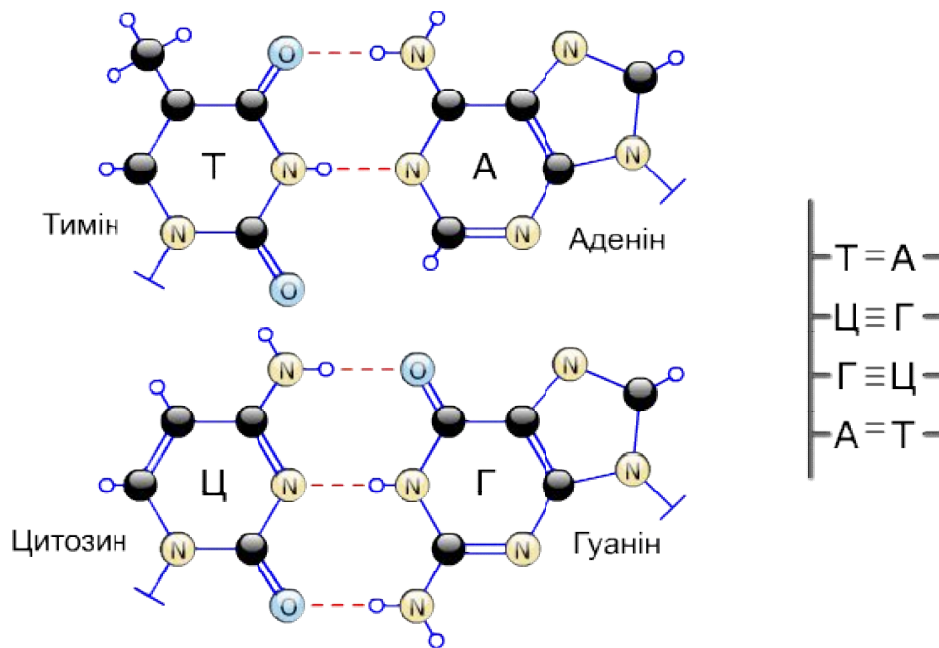


Рис. 2.6. Комплементарне з'єднання нуклеотидів в молекулі ДНК.

**Правила Чаргаффа.** Ервін Чаргафф в 1949-51 рр. визначив кількісне відношення азотистих основ, що входять до складу нуклеїнових кислот. Своє відкриття вчений виклав у так званих «Правилах Чаргаффа»:

1. Вміст аденіну рівний вмісту тиміну, а вміст гуаніну — кількості цитозину:  $A=T$ ,  $G=C$ .
2. Кількість пуринів дорівнює кількості піримідинів:  $A+G=T+C$ .
3. Кількість основ з 6 аміногруп дорівнює кількості основ з 6 кетогруп:  $A+C=G+T$  (Це правило слідує з першого). Разом з тим, співвідношення частка  $G+C$  (вміст  $G+C$ ) може бути різним у ДНК різних видів. У одних переважають пари  $AT$ , в інших —  $G$ .

- Завдання:**
1. Ознайомитись з принципом компліментарності і правилами Чаргаффа.
  2. Навчитися визначати послідовність нуклеотидів у молекулах ДНК та РНК.
  3. Розв'язувати задачі по визначенню складу нуклеїнових кислот, їх довжині та масі.

**Питання для самопідготовки:** Нуклеїнові кислоти: РНК та ДНК, склад, типи зв'язків між структурними компонентами. Типи РНК та їх біологічна роль. Подвійна спіраль ДНК, її особливості. Комплементарні основи. Нуклеотиди як складні мономерні одиниці нуклеїнових кислот: склад, будова, типи зв'язків.

- Задачі:**
1. На ділянці ДНК, яка має склад Ц-А-Т-Г-Г-Ц-Т-А-Т, синтезований фрагмент іРНК, вкажіть його склад.
  2. Ділянка ланцюга іРНК містить послідовність нуклеотидів: А-У-Г-У-У-У-А-А-Ц-Ц-Г-Ц-У-А-А-Г-У-У-Ц-У-У-У-А-Ц. Визначити будову ділянки ДНК, на якій синтезувалася ця копія.
  3. Один з ланцюгів ДНК має склад: Г-Ц-Т-Т-А-Г-Г-Т-А-А-Ц-Ц-Г-Т-Г-А-Ц-Ц-Ц-Т-Г. Скільки водневих зв'язків утворюють два ланцюги ДНК на цьому фрагменті?
  4. У нуклеїновій кислоті аденін складає 17% (що відповідає 500 нуклеотидам), гуанін та цитозин – по 25%. Визначити склад нуклеїнової кислоти, її довжину та масу.

5. Відомий склад і-РНК: аденіну – 41%, гуаніну – 7%, урацилу – 39%. Визначити нуклеотидний склад комплементарної ділянки ДНК.
6. Фрагмент одного ланцюга молекули ДНК має послідовність нуклеотидів: Т-Т-Г-А-Г-Ц-А-Ц-Г-Г-Т-А-А-А-Т-Ц-Г-А. Побудуйте схему дволанцюгової ДНК. Яку довжину має цей фрагмент ДНК? Яку масу має дволанцюговий фрагмент ДНК?
7. В складі фрагменту молекули і-РНК 60 нуклеотидів. Визначте довжину і масу молекули ДНК. На якій було синтезовано даний фрагмент.
8. Довжина фрагмента молекули ДНК – 221 нм. Скільки нуклеотидів в даному фрагменті?
9. У молекулі ДНК 120 гуанілових нуклеотидів, що становить 15% від загальної кількості. Визначіть кількість інших нуклеотидів, довжину і масу цієї молекули ДНК.
10. В молекулі і-РНК людини в результаті біохімічних досліджень виявлено 440 гуанілових нуклеотидів, 325 аденілових, 128 цитидилових та 348 уридилових нуклеотидів. Скільки цитидилових нуклеотидів міститься у фрагменті молекули ДНК, транскрипційною копією якої є дана і-РНК?
11. Ферменти, які здійснюють реплікацію ДНК, рухаються зі швидкістю 0,6 мкм за 1 хв. Скільки часу необхідно для подвоєння ДНК у хромосомі, яка має 500 репліконів, якщо довжина кожного реплікону 60 мкм?

## ПРАКТИЧНА РОБОТА №3

### Експресія генів. Транскрипція. Трансляція. Біосинтез білків

Експресія генів — процес, при якому спадкова інформація генів (нуклеотидна послідовність) використовується для синтезу функціонального продукту: білка або РНК. Якщо кінцевим продуктом є білок, процес експресії генів називається біосинтезом білків а ген — білок-кодуєчим (англ. protein-encoding gene). Процес складається із кроків транскрипції мРНК та її процесингу (кепування, сплайсингу, постраскрипційних модифікацій мРНК), трансляції та посттрансляційної модифікації. В інших випадках, для генів, що не кодують білки, а кодують так звані некодуєчі РНК (наприклад, тРНК чи рРНК), набір кроків дещо відрізняється. Для експресії генів може використовуватися як генетична інформація, так і епігенетична. Застарілий термін «реалізація генетичної інформації» посилається тільки на інформацію першого типу, якої, проте, може бути недостатньо для отримання функціонального продукту.

Експресія генів в багатьох випадках активно регулюється, змінюючи час та кількість синтезованого генетичного продукту. Кілька кроків у процесі експресії генів можуть модулюватися, зокрема транскрипція і посттрансляційна модифікація. Регуляція експресії генів надає клітині контроль за кількістю та структурою синтезованих біополімерів і є основою диференціації клітин, морфогенезу і адаптації організму до умов навколишнього середовища. Регулювання експресії генів також може приводити до еволюційних змін.

Процес експресії генів відбувається в організмах усіх живих істот: еукаріотів (у тому числі багатоклітинних організмах), прокаріотів (у бактерій і архей), а також вірусів — для створення макромолекулярних основ для їх життєдіяльності. Деякі процеси, які відбуваються під час експресії генів можуть модулюватися певними чинниками, наприклад транскрипція, сплайсинг РНК, трансляція і посттрансляційна модифікація білка.

Експресія генів забезпечує підтримання структури та функції клітини, що є основою для диференціації клітин, морфогенезу, а також універсальної адаптованості будь-якого організму до умов існування. Регуляція генів може також служити як субстрат для еволюційних змін, оскільки контроль за часом, місцем і інтенсивністю експресії генів може мати величезний вплив на функції (дію) генів у клітині або у багатоклітинному організмі.

У генетиці, вплив експресії генів розглядається на фундаментальному рівні, адже під час цього процесу під дією генотипу формується фенотип. Генетичний код зберігається у ДНК у вигляді нуклеотидної послідовності «яка інтерпретується» під час експресії генів, а властивості продуктів експресії генів призводить до формування фенотипу організму.

**Транскрипцією** називається процес зчитування генетичного коду з молекули ДНК. При цьому на одному з ланцюжків ДНК синтезується одноланцюжкова молекула інформаційної або матричної РНК (мРНК) згідно з принципом комплементарності. Послідовність з трьох нуклеотидів в мРНК, відповідна послідовності ДНК, що кодує визначену амінокислоту, називається кодоном. Основну роль в транскрипції грає фермент РНК-полімераза. У випадку синтезу з частини ДНК, що кодує білок — з так званих білок-кодуєчих генів, транскрипція є першим кроком біосинтезу білків, процесу, який кінець кінцем приводить до перекладу генетичного коду, через мРНК як проміжної ланки, у поліпептидну послідовність білка. Транскрипція каталізується ферментом ДНК-залежною РНК-полімеразою. Процес синтезу РНК протікає в напрямку від 5'- до 3'- кінця<sup>[en]</sup>,

тобто РНК-полімераза рухається матричним (також антизмістовним) ланцюжком ДНК у напрямку 3'→5'. Транскрипція різних видів РНК здійснюється різними РНК-полімеразами.

Рівень транскрипції більшості генів чітко регулюється за допомогою факторів транскрипції. Саме на цьому етапі відбувається більша частина регуляції експресії генів. Проте і якість синтезованих РНК регулюється на різних етапах транскрипції, абортівне завершення транскрипції (термінація) призводить до неповної молекули РНК, яка часто деградується за допомогою спеціальних механізмів, таких як робота ядерної екзосоми чи інших РНКаз.

Зазвичай процес транскрипції поділяється на 4 стадії — пре-ініціацію, ініціацію, елонгацію і термінацію.

**Трансляція** полягає в синтезі поліпептидного ланцюжка відповідно до інформації, закодованої в і-РНК. Амінокислотна послідовність шифрується за допомогою транспортних РНК, які утворюють з амінокислотами комплекси — аміноацил-тРНК. Кожній амінокислоті відповідає своя тРНК, що має відповідний антикодон, «відповідний» до кодону і-РНК (Рис. 3.1.). Під час трансляції рибосома рухається упродовж і-РНК, у міру цього нарощується поліпептидний ланцюжок. Енергією біосинтезу білка забезпечується за рахунок молекул АТФ.

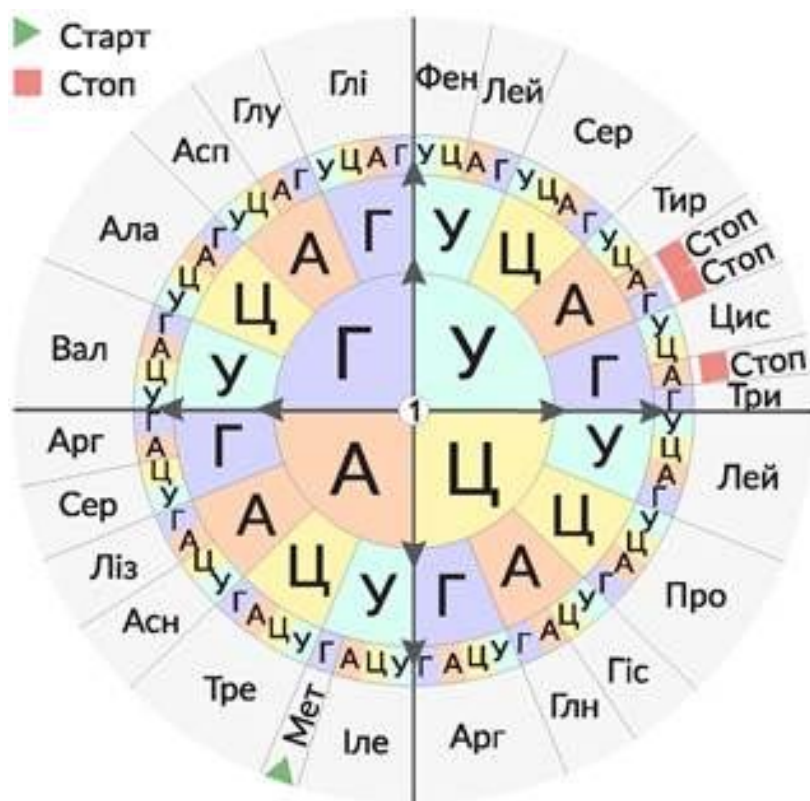


Рис. 3.1. Таблиця генетичного коду.

Готова білкова молекула потім відщеплюється від рибосоми і транспортується в потрібне місце клітини. Тоді як цитоплазматичні білки рухаються за допомогою дифузії та молекулярних моторів, мембранні білки, білки органел та білки позначені для секреції синтезуються на мембранах клітини (у випадку еукаріотів на мембранах ендоплазматичного ретикулума), одразу проходять встроюються мембрану та направляються до відповідної

органели або секретуються відповідно до сигнальної послідовності у складі білка (яка зазвичай видаляється після цього за допомогою протеолітичних ферментів).

Трансляція відбувається в цитоплазмі, де знаходяться рибосоми клітини. Під час трансляції інформація, що міститься в і-РНК, розшифровується згідно з правилами, відомими як генетичний код, та використовується для синтезу закодованої поліпептидної послідовності. Процес трансляції можна поділити на чотири фази: активацію, ініціацію, елонгацію та термінацію (Рис. 3.2).

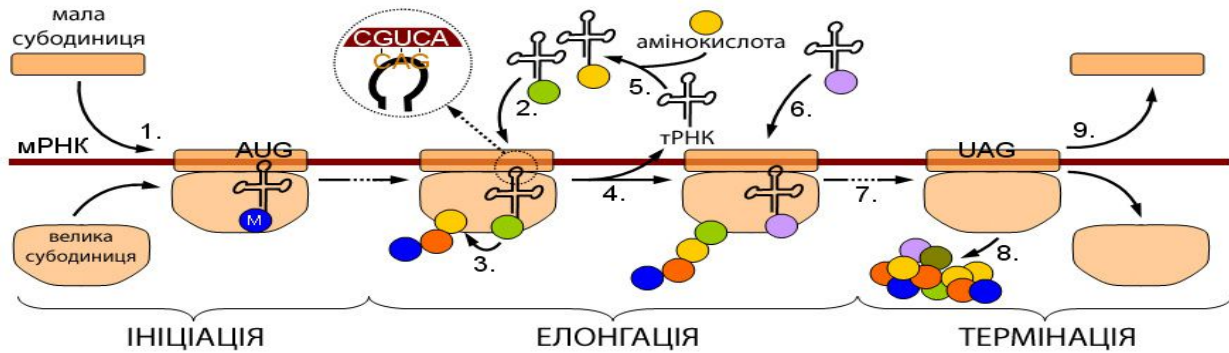


Рис. 3.2. Етапи трансляції.

Існує багато комп'ютерних програм, здібних до моделювання трансляції послідовності ДНК/РНК у білкову послідовність. Проте небагато програм здатні видавати правильну послідовність у всіх «особливих» випадках, таких як використання альтернативних ініціаторних кодонів. Наприклад рідкісний код альтернативного кодону ініціації TTG використовується для метіоніну (Рис. 3.1.), коли використовується як ініціаторний кодон, і для лейцину в решті випадків.

**Завдання:** 1. Ознайомитись з таблицею генетичних кодів.

2. Навчитися визначати послідовність нуклеотидів у молекулах РНК.
3. Навчитися визначати послідовність амінокислот у молекулах білків.
4. Розв'язувати задачі по визначенню складу нуклеїнових кислот та білків.

**Питання для самопідготовки:** Основні етапи біосинтезу білка в клітині. Транскрипція: процесинг, сплайсинг. Трансляція: ініціація, елонгація, термінація. Активація амінокислот. Генетичний код. Сучасний стан теорії гена. Властивості генетичного коду Посттрансляційні перетворення білків. Регуляція експресії генів. Екзонно-інтронна організація генома еукаріотів.

**Задачі:** 1. Некодуєчий ланцюг молекули ДНК має наступну структуру: Г-А-Г-А-Г-Г-Ц-Г-Т-Т-Г-А-Ц-Г-Г. Визначте будову відповідної частини молекули білка, що синтезується за участю кодуєчого ланцюга ДНК.

2. Визначте молекулярну масу та довжину гена, якщо він несе інформацію про білок, молекулярна маса якого дорівнює 280000 а.о.м.
3. Молекулярна маса білкової молекули становить 50000, а молекулярна маса однієї амінокислоти – 100 а.о.м. Визначте масу відповідного гена, що кодує цю білкову молекулу, якщо маса одного нуклеотиду складає 345 а.о.м.
4. В молекулі і-РНК людини в результаті біохімічних досліджень виявлено 440 гуанілових нуклеотидів, 325 аденілових, 128 цитидилових та 348 уридилових нуклеотидів. Скільки

цитидилових нуклеотидів міститься у фрагменті молекули ДНК, транскрипційною копією якої є дана і-РНК?

5. В молекулі і-РНК людини в результаті біохімічних досліджень виявлено 440 гуанілових нуклеотидів, 325 аденілових, 128 цитидилових та 348 уридилових нуклеотидів. Скільки аденілових нуклеотидів міститься у фрагменті молекули ДНК, транскрипційною копією якої є дана і-РНК?
6. Скільки нуклеотидів містить ген (обидва ланцюги ДНК), у якому закована первинна структура білка, що складається із 145 амінокислотних залишків? Яка молекулярна маса та довжина цього гена?
7. Відомий склад і-РНК: аденіну – 41%, гуаніну – 7%, урацилу – 39%. Визначити нуклеотидний склад відповідної ділянки ДНК.
8. Фрагмент ланцюга А білка нормального гемоглобіну складається із 7 амінокислот, розміщених у такій послідовності: вал – лей – лей – тре – про – глн – ліз.  
Яка будова фрагмента і-РНК, що є матрицею для синтезу цього фрагмента молекули гемоглобіну? Яка будова фрагмента ДНК, що кодує дану і-РНК?
9. До складу білка входить 800 амінокислот. Визначте довжину та молекулярну масу гена, який кодує синтез цього білка.
10. Довжина фрагмента ДНК становить 1530 нм. Скільки в ньому заковано білкових молекул, які складаються в середньому із 300 амінокислотних залишків?
11. Білок А – мономер, який складається з 560 амінокислотних залишків. Ген цього білка має 2 інтрони (по 10 000 пар нуклеотидів) і 3 екзони, кожен з яких містить однакову кількість пар нуклеотидів. Скільки пар нуклеотидів у складі даного гена? Скільки пар нуклеотидів у складі кожного з екзонів? Скільки нуклеотидів містить кодуюча зона іРНК даного білка?
12. У молекулі про-іРНК на інтронні ділянки припадає 800 нуклеотидів. Визначте відносну молекулярну масу й довжину структурного гена, якщо в ньому заковано поліпептид, відносна молекулярна маса якого становить 20 000.
13. Оперон (сукупність структурних генів і гена-оператора) містить 9300 нуклеотидів. У ньому заковано три поліпептидні ланцюги, кожен з яких складається із 250-ти амінокислотних залишків. Маса інтронних фрагментів дорівнює масі екзонних. Визначте лінійні розміри гена-оператора.

## ПРАКТИЧНА РОБОТА №4

### Сучасні методи дослідження в молекулярній біології та генетиці.

Основною проблемою в галузі охорони здоров'я є вчасна діагностика, адже виявлення захворювання на ранній стадії значно підвищує шанси пацієнта на одужання, оскільки в такому разі медицина має більше можливостей, аби запобігти подальшому розвитку патології. Ця проблема є однаково актуальною для всіх країн світу, але передусім для України, де й досі немає реальної програми для ранньої діагностики населення на предмет наявності поширених важких захворювань, зокрема онкологічних і серцево-судинних хвороб, цукрового діабету, які сукупно спричиняють найбільшу кількість летальних випадків. У розвинених країнах із високим рівнем фінансування наукових досліджень виникають особливі проблеми: найсучасніші та відносно надійні методи діагностики (на зразок томографічного обстеження) є або досить шкідливими для живого організму і нерідко виступають додатковим чинником ризику, або надто вартісними для масового впровадження та, попри точність, не здатні зафіксувати наявність патології на рівні клітини чи груп клітин. Над розробленням нових методів раннього розпізнавання захворювань сучасні вчені працюють у межах таких суміжних і порівняно молодих наук, як молекулярна біологія та генетика. Доведено, клітина має набір білкових компонентів, організованих у певні мережі, структури.

На даний момент використовують три основні методи діагностики – на рівні ДНК, на рівні РНК та на рівні білків. На обох зазначених рівнях дослідження при здійсненні аналізу крові пацієнта за допомогою розроблених науковцями тестсистем – за умови наявності моноклональних антитіл та специфічних праймерів для полімеразної ланцюгової реакції – цілком можливо виявити патології навіть за присутності лише однієї пошкодженої клітини. В сучасній медицині вже чітко прослідковується тенденція до персоналізованої терапії. Що стосується генетики, то вона вже дає змогу за короткий час здійснити сиквенс (простіше кажучи, дешифрування) геному людини й отримати персональну карту, яка містить інформацію про пошкодження ДНК (якщо такі виявлено) або брак її складових та потенційну схильність особи до певних хвороб.

Відомо, наприклад, що при серповидно-клітинній анемії відбувається мутація в одній з амінокислот, і в більшості випадків це не залишає шансу на виживання новонародженим. Для деяких метаболічних розладів вже розроблено комплекс лікувальних заходів, які можна здійснювати впродовж усього життя людини. Ясна річ, змінити ДНК в усьому трильйоні клітин, з яких складається людський організм, неможливо. Наразі йдеться про те, щоб підтримувати життя, штучно вводячи, зокрема, відсутній білок, який через генні порушення не синтезується в організмі, або стовбурові клітини із правильною копією гена, частково модифікуючи таким чином ДНК. До цього методу зазвичай вдаються, якщо виникають проблеми з кістковим мозком: тоді за допомогою хіміо- та радіотерапії вбивають аномальні клони, що містять пошкоджені, онкологічні, гени, а потім вводять стовбурові клітини – носії нормального гена. Розмножуючись, ці клітини мають забезпечувати людині імунний захист. Однак, ідеального результату науці й медицині поки що досягти не вдалося. З іншого боку, встановити причину захворювання нині можна у переважній більшості випадків.

## Молекулярно-генетичні методи

Це різноманітна група методів, що застосовуються для виявлення варіацій у структурі досліджуваної ділянки ДНК, а також для розшифрування первинної послідовності основ. Ці методи ґрунтуються на "маніпуляціях" з ДНК і РНК.

Широкого застосування в молекулярній біології набули різноманітні фізичні методи, важливість яких важко переоцінити. Нижче наведено перелік основних таких методів із зазначенням особливостей і характеру інформації, яку можна отримати за їх допомоги.

### Методи безпосереднього спостереження

Історично першим методом спостереження великих макромолекулярних комплексів із роздільною здатністю у нанометровому діапазоні стала **електронна мікроскопія**. При застосуванні цього методу зразок розміщують у вакуумній камері мікроскопа, опромінують променем електронів, який далі проектується на флуоресцентний екран. Зразок при цьому має бути «зафарбований» барвником, непрозорим для електронів. Одна з поширених технік полягає в тому, що макромолекули наносять на карбонову підкладку, практично прозору для електронного променя (як і макромолекули), і фіксують на ній. Потім підкладку промивають розчином солі важкого металу, наприклад ураніацетатом. Після висушування така підкладка стає непрозорою для електронів скрізь, крім ділянки, де розміщена макромолекула, й ураніацетат не зв'язався з підкладкою. Така техніка негативного контрасту дозволяє спостерігати під електронним мікроскопом молекули ДНК, комплекси ДНК із білками (такі як нуклеосоми), мультибілкові комплекси тощо. Розділення при негативному контрасті лімітоване розміром частинок важкого металу.

Важливим варіантом електронної мікроскопії, який дозволяє уникнути фіксації та контрастування зразка, є **кріоелектронна мікроскопія**. Тонкий шар (~100 нм) розчину макромолекул на підкладці швидко заморожується до  $-160\text{ }^{\circ}\text{C}$  і при цій температурі розміщується в охолодженій вакуумній камері мікроскопа. Виявляється, що такі зразки дають достатньо високий рівень контрасту без додаткової обробки. Кріоелектронна мікроскопія відіграла важливу роль у дослідженнях структури та механізмів функціонування рибосоми.

Сучасною альтернативою електронної мікроскопії є **мікроскопія атомних сил** (AFM – Atomic Force Microscopy). Макромолекули адсорбуються на підкладці, після чого зразок може бути дослідженим як після висушування, так і в разі, коли залишається шар рідини.

Підкладка сканується голкою з надзвичайно тонким вістряем, яке за рахунок міжатомних (вандерваальсових) сил взаємодіє з поверхнею, відстежуючи її «ландшафт» (Рис. 4.1).

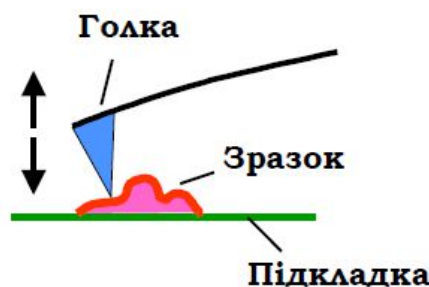


Рис. 4.1. Принцип мікроскопії атомних сил (AFM).

Рух голки реєструється оптичною системою (лазерний промінь, що віддзеркалюється від голки на фотодіод), і перетворюється комп'ютером у зображення поверхні. Роздільна



здатність є порівнянною з такою електронного мікроскопа, однак, на відміну від електронної мікроскопії, AFM не потребує фіксації та контрастування зразка, дозволяє працювати з макромолекулами, що перебувають в умовах, близьких до фізіологічних. Крім того, існує можливість кількісної оцінки висоти зміщення голки відносно поверхні підкладки.

### **Рентгеноструктурний аналіз**

Головною технікою, яка дозволяє встановити просторову структуру макромолекули з роздільною здатністю до кількох ангстремів, є рентгеноструктурний аналіз. Об'єктом аналізу є кристал, у складі якого однакові макромолекули чи макромолекулярні комплекси утворюють впорядковану тривимірну решітку. Кристали, придатні для рентгеноструктурного аналізу (розміром щонайменше 0,3 мм), готують із перенасичених розчинів макромолекул.

Вузкий рентгенівський промінь буде частково розсіюватись на атомах кристала. Оскільки групи атомів періодично повторюються у кристалі, розсіяні промені будуть інтерферувати між собою, підсилюючи один одного в певних точках на детекторі, що реєструє рентгенівські промені після проходження їх крізь зразок. У результаті на детекторі формується впорядкований розподіл так званих рефлексів – точок високої інтенсивності.

Картина рентгенівського розсіювання містить інформацію щодо просторових позицій атомів у кристалі. Вилучення цієї інформації є досить складним математичним завданням, хоча сьогодні комп'ютерна техніка дозволяє аналізувати рентгенограми досить швидко: лімітуючим кроком рентгеноструктурного аналізу є практичне вирошування кристалів.

Безпосереднім результатом аналізу рентгенограми є складна тривимірна карта електронної щільності. Інтерпретація цієї карти є можливою за умови, якщо послідовність біополімеру в складі кристала є відомою. Шляхом численних підгонок на комп'ютері ця послідовність (чи послідовності кількох елементів досліджуваного комплексу) узгоджується з електронною картою, результатом чого є встановлення просторової координати кожного атома в макромолекулі. Саме таким шляхом отримано інформацію про структуру численних білків і міжмолекулярних комплексів.

Проте, як правило, великі макромолекули виявляються надто складними для того, щоб рентгенівське розсіювання на їхніх кристалах дозволило отримати тривимірну структуру. Для визначення структури більшості білків, нуклеїнових кислот і комплексів застосовують метод ізоморфних похідних, який потребує набору кристалів макромолекули: вихідний кристал і принаймні 2-3 кристали, які були б ідентичними, але містили б атоми важких металів, що значно інтенсивніше розсіюють рентгенівське світло. Важкі атоми вводять у кристал або шляхом дифузії важких металів у нього, або використовуючи мутантні форми білка. Таким чином, розвиток методів рекомбінантної ДНК та експресії рекомбінантних білків став важливою передумовою структурних досліджень: завдяки рекомбінантним технологіям стало можливим не тільки отримувати достатню кількість високоочищених білків, придатних для подальшої кристалізації, а й готувати їхні ізоморфні похідні. Порівняння рентгенограм, отриманих від немодифікованого кристала та його похідних, дозволяє врешті решт розрахувати структуру макромолекули.

Численні експериментальні результати свідчать, що структура макромолекули в кристалі, не тільки у своїх головних загальних рисах, а й у багатьох деталях, збігається зі структурою в розчині. Тому важливим доповненням до рентгеноструктурного аналізу є різноманітні методи дослідження структури макромолекул у розчині.

### **Дослідження структури макромолекул у розчині**

Вивчити особливості конформаційної рухливості та взаємодії макромолекул у розчині дають змогу різноманітні спектральні методи.

Загальний принцип будь-якої спектроскопії полягає в наступному: розчин опромінюється тим чи іншим типом електромагнітного випромінювання, яке мінімально взаємодіє зі зразком. Відповідно, жорстке іонізуюче випромінювання, яке здатне руйнувати ковалентні зв'язки, не є придатним для спектральних досліджень. Різні типи молекулярної спектроскопії використовують діапазони випромінювання, які характеризуються не дуже високою енергією – від ближнього ультрафіолету до високих радіочастот.

Найчастіше для досліджень макромолекул застосовують **спектрофотометрію**, круговий дихроїзм, флуоресцентну спектроскопію та ядерний магнітний резонанс.

Спектрофотометр – прилад для вимірювання ефективності поглинання світла в ультрафіолетовій і видимій областях – є неодмінним атрибутом будь-якої молекулярно-біологічної лабораторії. У першу чергу, спектрофотометрія використовується з метою визначення концентрацій молекул: ефективність поглинання є пропорційною до концентрації. Але крім того, ефективність поглинання може залежати від мікрооточення хімічних груп, які відповідають за поглинання – хромофорів.

Найяскравіше така залежність виявляється для нуклеїнових кислот: поглинання азотистих основ в ультрафіолеті значно знижується (так званий гіпохромний ефект) за умови реалізації між ними стекінг-взаємодій. Дослідження денатурації нуклеїнових кислот (плавлення подвійної спіралі) за допомогою протилежного, гіперхромного, ефекту (зростання в поглинанні при руйнуванні стекінг-взаємодій) дає інформацію про стабільність подвійних спіралей залежно від зовнішніх умов, нуклеотидного складу, іноді – особливостей послідовності тощо.

**Спектроскопія ядерного магнітного резонансу (ЯМР)** діє на основі того принципу, що деякі атомні ядра (найчастіше йдеться про ядра гідрогену) мають власний магнітний момент (спін), який орієнтується вздовж зовнішнього постійного магнітного поля. Електромагнітне випромінювання певної частоти в радіочастотному діапазоні здатне змінювати цю орієнтацію на протилежну: відбувається резонансне поглинання енергії, що приводить до зміни магнітних характеристик зразка, яку можна зафіксувати. Значення резонансної частоти залежить від того, до складу якої хімічної групи належить даний гідроген, і якого типу хімічні групи містяться поруч. Тобто спектр ЯМР – залежність ефективності поглинання від частоти – несе інформацію про типи хімічних груп, їхнє взаємне розташування та ступінь їх рухливості у складі макромолекули.

Особливо цінною є техніка **двовимірного ЯМР**, коли сигнали резонансу детектують після складної комбінації кількох радіочастотних імпульсів і відображають як функцію двох координат (наприклад, у найпростішому варіанті – це частота та інтервал між двома імпульсами).

На спектрі такого типу можна розрізнити сигнали гідрогенів різних амінокислот і виміряти зсуви цих сигналів, зумовлені іншими наближеними амінокислотами. Величина зсуву є мірою відстані між амінокислотами: порівняння цієї інформації з амінокислотою послідовністю дозволяє з'ясувати просторову структуру білка. Сьогодні двовимірний ЯМР – єдиний метод, за допомогою якого можна встановити структуру білка в розчині. На жаль, це можливо тільки для порівняно невеликих білків (до 200 амінокислот). Та оскільки результатом двовимірного ЯМР є набір дуже подібних одна до одної структур, що відображає конформаційну рухливість білка, і оскільки метод у принципі позбавлений можливих артефактів, пов'язаних із кристалічною упаковкою, двовимірний ЯМР є важливим доповненням до рентгеноструктурного аналізу.

Нині банк даних білкових структур містить понад 36 тис. структур, отриманих методом

рентгеноструктурного аналізу, і понад 6 тис. – методом двовимірного (чи останнім часом мультимірного) ЯМР.

Ще одним важливим доповненням до експериментальних методів вивчення структури макромолекул є комп'ютерні розрахунки в рамках підходу **моделювання молекулярної динаміки**.

Метод розрахунку базується на використанні другого закону Ньютона:

$$F = ma,$$

де  $F$  – сила, що діє на частинку,  $m$  – її маса,  $a$  – прискорення руху частинки.

У програму розрахунку вводяться дані про координати атомів макромолекули, отримані методами рентгеноструктурного аналізу чи ЯМР. Знаючи силу, яка діє на кожен атом (на основі між атомних відстаней та енергії нековалентних взаємодій різних типів), можна обчислити прискорення кожного атома в системі: отримати систему диференціальних рівнянь руху. Наступне інтегрування цих рівнянь дозволяє обчислити залежності траєкторій і швидкостей руху атомів від часу. Таким чином, розрахунки молекулярної динаміки можуть дати детальну інформацію про флуктуації та конфірмаційні перетворення білків і нуклеїнових кислот.

#### Методи дослідження одиничних макромолекул

Більшість методів дослідження (за винятком мікроскопії різних типів) передбачає роботу з досить великою кількістю – ансамблем – однакових молекул. Останнім часом розвиваються методи, що дозволяють слідкувати за поведінкою одиничних (звичайно, досить великих) макромолекул і макромолекулярних комплексів: ДНК та її комплексів з білками, хроматинової фібрили, актоміозинової фібрили тощо.

Одним із прикладів таких підходів є **застосування магнітного пінцета** (magnetic tweezers). У випадку ДНК (Рис. 4.2), молекули (довжиною ~5 тис. пар основ чи більше) пришивають за один кінець до покривного скельця мікроскопа, за інший – до парамагнітної кульки ~2 мкм у діаметрі. Розмір кульки дозволяє слідкувати за нею (тобто за індивідуальною молекулою ДНК) за допомогою звичайного оптичного мікроскопа. Кулька знаходиться в полі електромагніта, який можна переміщувати або обертати, як показано на рисунку 4.2, прикладаючи тим самим зовнішню силу до кінця молекули. Відповідь на зміну зовнішньої сили, що проявляється у відстані кульки від скельця (дефектується відеокамерою), несе інформацію про еластичні властивості молекули, тобто про сили, що цю молекулу стабілізують.

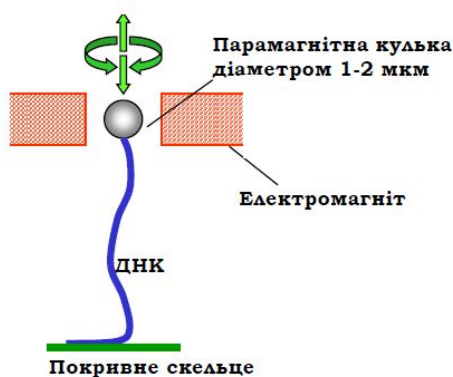


Рис. 4.2. Принципова схема експерименту з ДНК за допомогою магнітного пінцета.

За допомогою підходів такого типу отримано цінну інформацію, зокрема про жорсткість ДНК щодо розтягування та кручення, структурні особливості деяких білково-нуклеїнових комплексів, конформаційні перебудови нуклеосом, тонкі механізми роботи РНК-полімерази в реальному часі.

- Завдання:**
1. Ознайомитися із методами сучасної генетики та молекулярної біології.
  2. Ознайомитися із експериментальними методами вивчення просторової структури білків.
  3. Запропонувати галузі та напрямки застосування досягнень молекулярної біології (у вигляді реферату, пошуку наукових статей, презентацій).

**Питання для самопідготовки:** Клонування, ампліфікація і секвенування ДНК. Методи дослідження ДНК-білкових взаємодій. Методи аналізу структури й експресії генів. Фізичні методи дослідження структури й активності біомакромолекул. Динаміка білків. Функціонування ферментів.

## ПРАКТИЧНА РОБОТА №5

### Методика аналізу ДНК. Полімеразна ланцюгова реакція (ПЛР) як метод ампліфікації ДНК. Схема ПЛР.

#### Основні етапи аналізу нуклеотидних послідовностей:

1. Отримання зразків ДНК або РНК – вихідний етап всіх методів. Джерелом геномної ДНК є будь-які клітини, що мають ядро. Частіше використовують периферійну кров (лейкоцити), хоріон, амніотичні клітини, культури фібробластів. Основне завдання – накопичення необхідної кількості певних фрагментів ДНК.

2. Полімеразна ланцюгова реакція (ПЛР) – це метод ампліфікації ДНК за умов *in vitro*. Відповідно до нуклеотидної послідовності кінців досліджуваної ділянки застосовують два олігонуклеотидних праймери (приманки). Довжина праймерів 20-30 нуклеотидів. Процес ампліфікації полягає у здійсненні повторюваних циклів (Рис. 5.1, 5.2).

3. Рестрикція ДНК на фрагменти за допомогою рестриктаз. Основна їх властивість – розривати дволанцюгову ДНК у визначених послідовностях нуклеотидів. Рестриктази – ферменти, виділені з бактеріальних клітин, розрізають молекулу ДНК на фрагменти у визначених місцях. Застосування цих ферментів дає можливість одержати досить короткі фрагменти ДНК, в яких легко можна визначити послідовність нуклеотидів. Розробка методу зворотної транскрипції ДНК на молекулах мРНК визначених білків з наступним клонуванням цих ДНК призвела до появи ДНК-зондів. Використання таких зондів для гібридизації з ДНК-клітин пацієнта дає можливість точно локалізувати генну мутацію.

4. Електрофорез фрагментів ДНК. Кожен фрагмент ДНК займає певне місце у вигляді дискретної смуги в конкретному місці геля.

5. Візуалізація та ідентифікація фрагментів ДНК у гелі.

#### Полімеразна ланцюгова реакція.

Полімеразна ланцюгова реакція (ПЛР / PCR) — експериментальний метод молекулярної біології, спосіб значного збільшення малих концентрацій бажаних фрагментів ДНК в біологічному матеріалі (пробі). Крім простого збільшення числа копій ДНК (цей процес називається *ампліфікацією*), ПЛР дозволяє проводити безліч інших маніпуляцій з генетичним матеріалом (введення мутацій, зрощення фрагментів ДНК), і широко використовується в біологічній і медичній практиці, наприклад для клонування генів, введення мутацій, виділення нових генів, секвенування, для створення і визначення генетично модифікованих організмів, діагностики захворювань (спадкових, інфекційних), ідентифікації малих кількостей ДНК, встановлення батьківства.

Полімеразна ланцюгова реакція була відкрита Кері Маллісом. Він став лауреатом Нобелівської премії з хімії 1993 року за це відкриття, через сім років після того, як він і його колеги з корпорації Cetus запропонували його для практичного використання. Вперше метод був винайдений у 1983 році, під час роботи Малліса в компанії Cetus в місті Емерівіль, США, в одній з перших біотехнологічних компаній. В журналі *Scientific American* К. Малліс підвів підсумок методу: «Починаючи з єдиної молекули ДНК, носія генетичної інформації, ПЛР може надати 100 мільярдів подібних молекул за кілька годин. Реакцію дуже легко провести, вона вимагає однієї пробірки, незначної кількості реагентів, та джерела тепла».

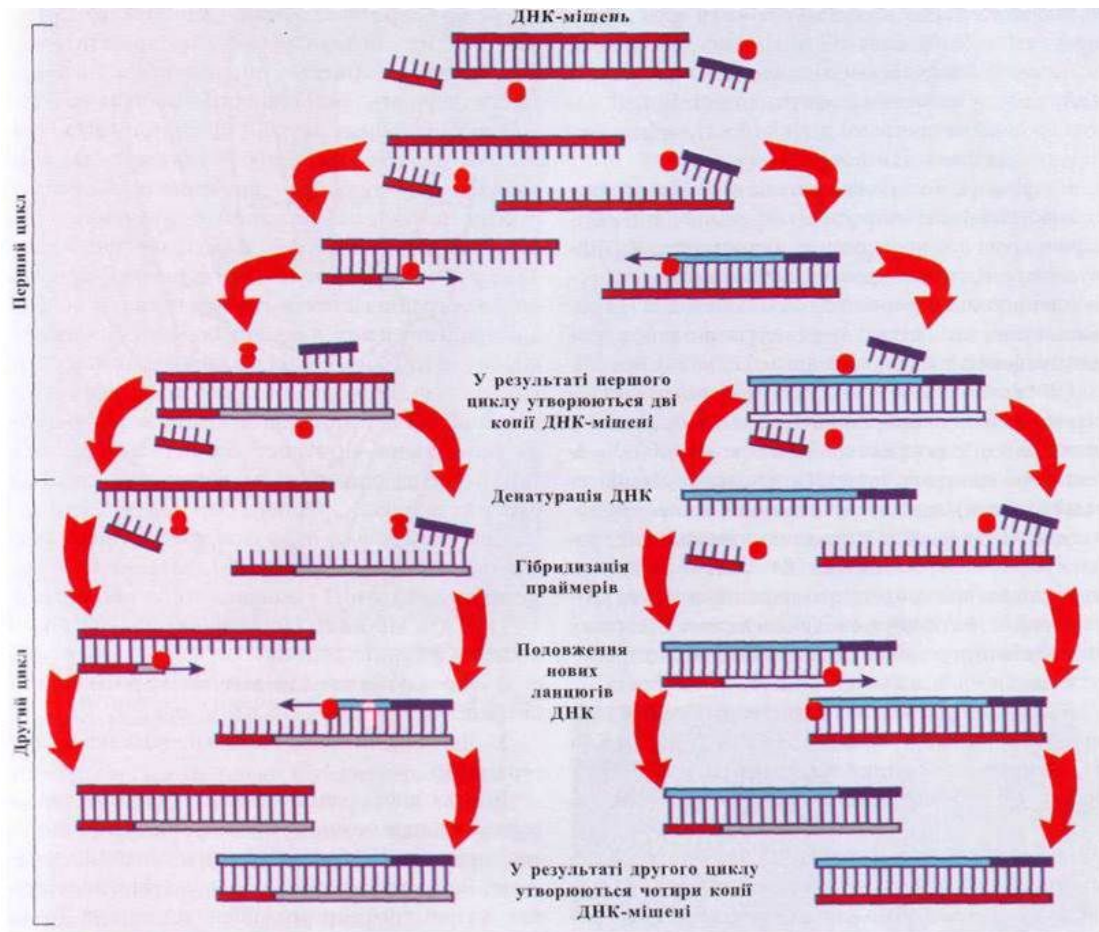


Рис. 5.1. Схема полімеразної ланцюгової реакції.

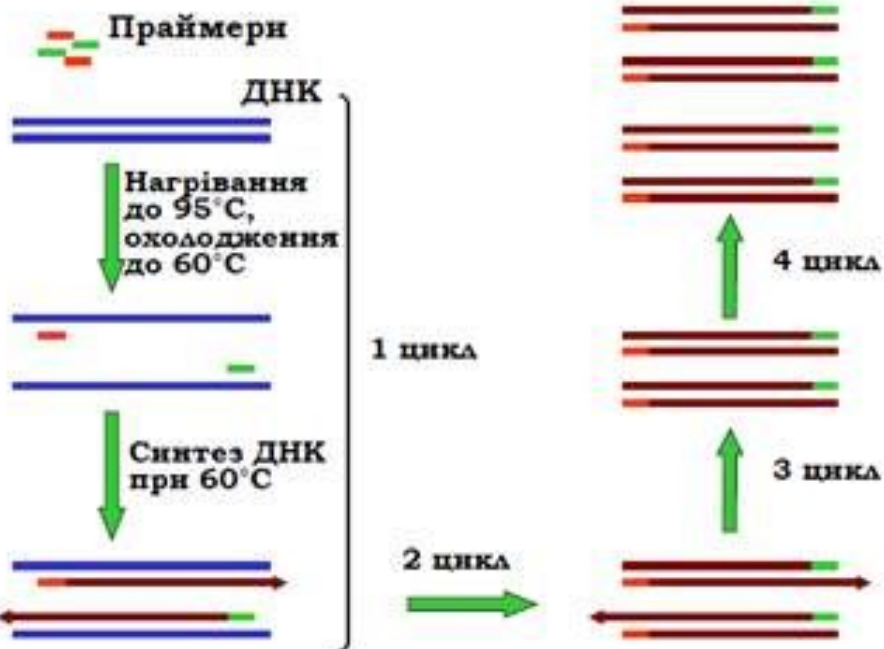


Рис. 5.2. Принцип полімеразної ланцюгової реакції.

Метод заснований на багаторазовому вибіркового копіюванні певної ділянки ДНК за допомогою ферментів в штучних умовах (*in vitro*). При цьому відбувається копіювання

тільки тієї ділянки, яка задовольняє заданим умовам, і тільки в тому випадку, якщо вона присутня в досліджуваному зразку.

На відміну від ампліфікації ДНК в живих організмах (реплікації), за допомогою ПЛР ампліфікують відносно короткі ділянки ДНК. У звичайному ПЛР-процесі довжина копійованих ДНК-ділянок становить не більше 3000 пар основ.

Для проведення ПЛР в найпростішому випадку потрібні такі компоненти:

- ДНК-матриця, що містить ту ділянку ДНК, яку потрібно ампліфікувати.
- Два праймери – короткі синтетичні олігонуклеотиди довжиною 18-30 базіоснов, що комплементарні кінцям різних ланцюгів необхідного фрагмента ДНК.
- Термостабільна ДНК-полімераза – фермент, який каталізує реакцію полімеризації ДНК. Полімераза для використання в ПЛР повинна зберігати активність при високій температурі тривалий час, тому використовують ферменти, виділені з термофілів – *Thermus aquaticus* (Taq-полімераза), *Pyrococcus furiosus* (Pfu-полімераза), *Pyrococcus woesei* (Pwo-полімераза) та інші.
- Дезоксирибонуклеозидтрифосфати (dATP, dGTP, dCTP, dTTP).
- Буферний розчин, що забезпечує необхідні умови реакції.

ПЛР проводять в ампліфікаторі (рис. 5.3) – приладі, що забезпечує періодичну та швидко зміну температури (охолодження і нагрівання) тестових пробірок із розчином, зазвичай з точністю не менше 0,1 °С.



Рис. 5.3. Ампліфікатор для ПЛР.

Реакція проходить в 30-50 циклів, кожен з яких складається з трьох стадій: денатурація, віджиг, елонгація (Рис. 5.4):

1. Дволанцюгову ДНК-матрицю нагрівають до 94-96 °С (або до 98 °С, якщо використовується особливо термостабільна полімераза) на 0,5-10 хв., щоб ланцюги ДНК розділилися. Ця стадія називається денатурацією — руйнуються водневі зв'язки між двома ланцюгами. Іноді перед першим циклом проводять попереднє прогрівання реакційної суміші протягом 2-5 хв. для повної денатурації матриці і праймерів.



2. Коли ланцюги розійшлися, температуру знижують, щоби праймери могли зв'язатися з одноланцюговою матрицею; відбувається гібридизація праймерів та матриці. Ця стадія називається відпалом (англ. *annealing*). Температура відпалу залежить від послідовності праймерів і зазвичай вибирається на 4-5°C нижче за їх температуру плавлення. Тривалість стадії — 0,5-2 хв.

3. ДНК-полімераза реплікує матричний ланцюжок, використовуючи праймер як затравку. Це так звана стадія елонгації. Температура елонгації залежить від полімерази. Полімерази Taq і Pfu, що найчастіше використовуються, найактивніші за 72 °C. Час елонгації залежить як від типу ДНК-полімерази, так і від довжини фрагмента, який ампліфікують. Середня швидкість елонгації — 1000 пар основ за 1 хв. Після закінчення всіх циклів зазвичай проводять додаткову стадію фінальної елонгації, щоб добудувати всі одноланцюжкові фрагменти. Ця стадія триває 5-15 хв.

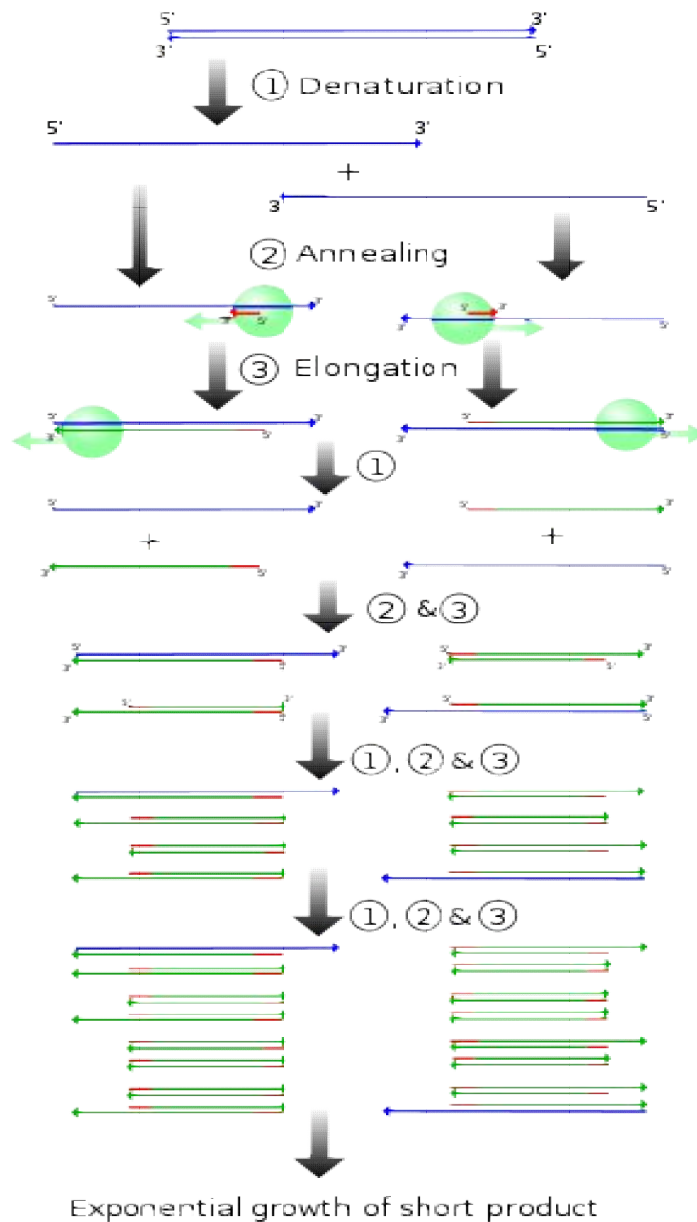


Рис. 5.4. Схематичне зображення процесу ПЛР: 1 - денатурація при 94-96 °C; 2 - відпал при ~65 °C; 3 - елонгація при 72 °C.

На малюнку зображені чотири цикли реакції.

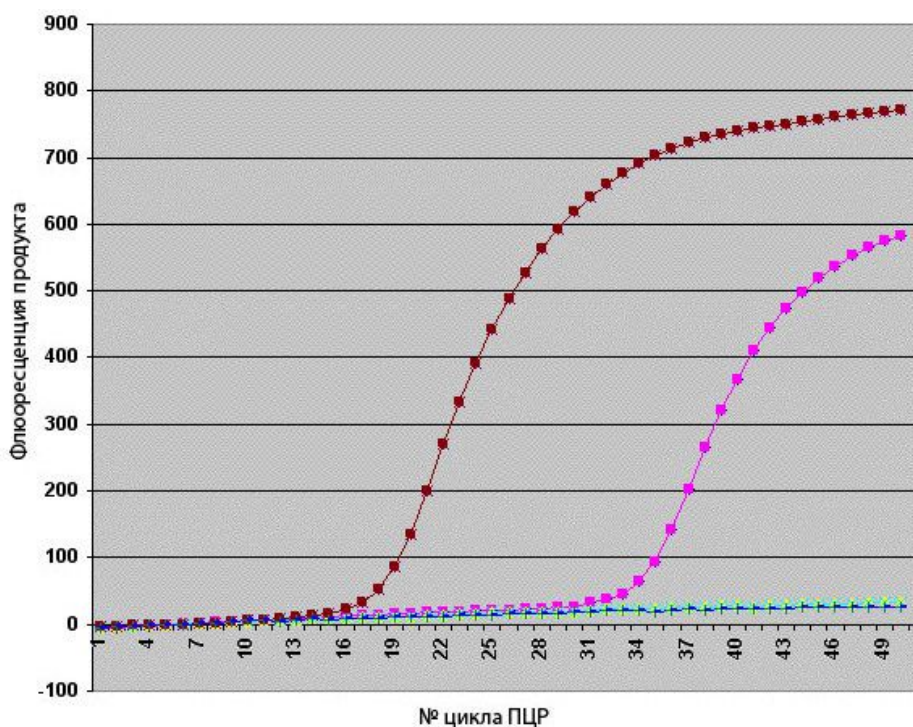


### ПЛР в режимі реального часу (real-time PCR):

Real-time PCR це група методик, що дають можливість аналізу продуктів ПЛР (PCR) в режимі реального часу по ходу проведення реакції.

Основними етапами даного методу є:

- Визначення виходу продукту реакції після кожного циклу ампліфікації.
- Побудова за цими даними кінетичної кривої ПЛР.



Визначення відносної концентрації субстрату на підставі аналізу цієї кривої:

Для детекції PCR-продукту використовуються флуоресцентні барвники, що забезпечують флуоресценцію, прямо пропорційну кількості ПЛР-продукту. Механізми її генерації розрізняються залежно від конкретного типу real-time PCR.

Кінетична крива PCR має S-подібну форму.

В ній можна виділити три стадії:

- Стадію ініціації (коли PCR-продукти ще не детектуються флуоресцентною міткою).
- Експонентну стадію (в якій спостерігається експоненціальна залежність кількості флуоресценції від циклу PCR).
- Плато (стадію насичення).

Початкова кількість субстрату визначається порогом циклу (threshold cycle), такий цикл  $n$ , на якому досягається певний заданий рівень флуоресценції – порогова флуоресценція за умови, що ефективність реакції і заданий рівень порогової флуоресценції однакові для кожної з порівнюваних реакцій.

Варіанти real-time PCR розрізняються за способами генерації флуоресценції. Два основних використовуваних в даний час способи – це:

- Застосування інтерколюючих флуоресцентних агентів, флуоресценція яких значно зростає при зв'язуванні з дволанцюговою ДНК,
- Використання мічених флуоресцентними агентами олігонуклеотидних проб, комплементарних області PCR-продукту.

Переваги методу ПЛР, як методу діагностики інфекційних захворювань:

- Виявлення ДНК збудника методом ПЛР дає пряму вказівку на присутність збудника інфекції.

- Висока специфічність методу ПЛР обумовлена тим, що в досліджуваному матеріалі виявляється унікальний, характерний тільки для даного збудника фрагмент ДНК, виключає можливість отримання помилкових результатів, на відміну від методу імуноферментного аналізу, де часті помилки в зв'язку з перехресно-реагуючими антигенами.

- Метод ПЛР, завдяки високій чутливості, дозволяє виявляти навіть поодинокі клітини бактерій або вірусів. Чутливість ПЛР-аналізу складає 10-1000 клітин в пробі (чутливість імунологічних і мікроскопічних тестів – 10<sup>3</sup>-10<sup>5</sup> клітин).

- Універсальність процедури виявлення різних збудників полягає в тому, що матеріалом для дослідження методом ПЛР слугує ДНК або РНК збудника. Подібність хімічного складу всіх нуклеїнових кислот дозволяє застосовувати уніфіковані методи проведення лабораторних досліджень. Це дає можливість діагностувати кілька збудників з однієї біопроби.

- Висока швидкість отримання результату аналізу обумовлена тим, що для проведення ПЛР-аналізу не потрібно виділення і вирощування культури збудника, що займає велику кількість часу.

### Секвенування

Секвенування, секвенс (від англ. Sequence – послідовність) – визначення первинної амінокислотної або нуклеотидної послідовності біополімерів (білків і нуклеїнових кислот – ДНК і РНК). В результаті виходить лінійний символічний опис, який коротко пояснює структуру молекули (Рис. 5.5). Для секвенування в даний час зазвичай застосовується метод Сенгера з дідезоксинуклеозидтрифосфатами (ddNTP). Зазвичай, до початку секвенування за допомогою ПЛР проводять ампліфікації ділянки ДНК, послідовність якої потрібно визначити.

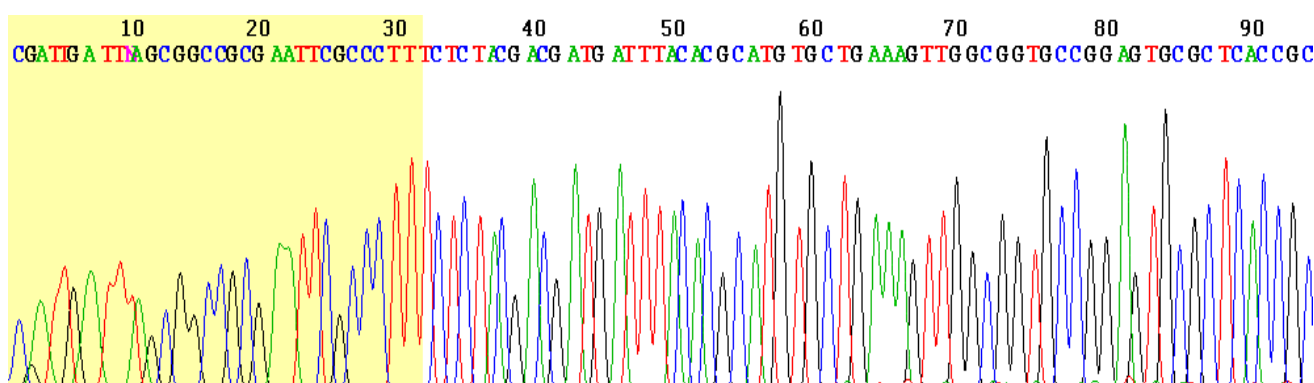


Рис. 5.5. Графічне зображення послідовності нуклеотидів. Метод Сенгера із застосуванням барвників-термінаторів.

Інший сучасний підхід у секвенуванні цілих геномів (так зване піросеквенування), який реалізується на автоматизованих секвенаторах, дозволяє визначити послідовність значно швидше, дешевше і при цьому не потребує ані клонування ДНК, ані електрофорезу. Одноланцюгові фрагменти, отримані з невеликої кількості геномної ДНК, пришивають 5'-кінцями до мікрокульок (один фрагмент на кульку) та ампліфікують за допомогою ПЛР. Кожну кульку із пришитими до неї ампліфікованими ідентичними фрагментами розміщують у мікрореакторі, де здійснюється ДНК-полімеразна реакція. Нуклеозидтрифосфати

подаються в реакційну суміш імпульсно один за одним. Якщо нуклеотид певного типу виявляється комплементарним матриці та включається у зростаючий ланцюг, пірофосфат, що при цьому звільняється, залучається до низки хімічних реакцій, де остання реакція супроводжується випромінюванням світла (хемілюмінесценція). Світловий сигнал фіксується оптичною системою, і послідовність таких сигналів читається як нуклеотидна послідовність. Реакція здійснюється паралельно у 200 тисячах мікрореакторів (для 200 тис. фрагментів, які перекриваються), що дозволяє встановити послідовність приблизно 200 млн. пар основ протягом 4,5 години.

Однак, існуючі технології ампліфікації ДНК обмежують розміри аналізованої ДНК тисячою парою основ. В зв'язку з цим зараз розробляються нові методики аналізу послідовностей ДНК, які дозволять не лише знизити вартість, часові затрати і рівень помилок, що характерні для аналізу ДНК попередніми методами, але й вивчати нитки ДНК, значно довші, ніж це можливо внаслідок обмежень, характерних для існуючих методів.

**Завдання:** 1. Ознайомитися з методиками ампліфікації ДНК.

2. Ознайомитися з різними методами секвенування нуклеїнових кислот.

3. Ознайомитися з принципом роботи гель-електрофорезу та різними його методами.

**Питання для самопідготовки:** Фрагментування ДНК. Стадії, які включає клонування фрагмента ДНК. Мета ампліфікації ДНК. Метод ПЛР. Обладнання для ПЛР. ПЛР-маркери. Методи гель-електрофорезу. Візуалізації результатів гель-електрофорезу. Секвенування ДНК. Методи секвенування. Генна терапія, її різновиди.

## ПРАКТИЧНА РОБОТА №6

### Аналіз білкових послідовностей. Розрахунок молекулярної маси, амінокислотного складу білків.

Білки — складні високомолекулярні природні органічні речовини, що складаються з амінокислот, сполучених пептидними зв'язками (Рис. 6.1).



Рис. 6.1. Стрічкова молекулярна модель білка — ядерного антигену проліферуючих клітин (PCNA) людини.

Зазвичай білки є лінійними полімерами — поліпептидами, хоча інколи мають складнішу структуру. Невеликі білкові молекули, тобто олігомери поліпептидів, називаються пептидами. Послідовність амінокислот у конкретному білку визначається відповідним геном і зашифрована генетичним кодом. Хоча генетичний код більшості організмів визначає лише 20 «стандартних» амінокислот, їхнє комбінування уможливорює створення великого різноманіття білків із різними властивостями. Крім того, амінокислоти у складі білка часто піддаються посттрансляційним модифікаціям, які можуть виникати і до того, як білок починає виконувати свою функцію, і під час його «роботи» в клітині. Для досягнення певної функції білки можуть діяти спільно, і часто зв'язуються, формуючи великі стабілізовані комплекси (наприклад, фотосинтетичний комплекс).

Функції білків в клітині різноманітніші, ніж функції інших біополімерів — полісахаридів і нуклеїнових кислот. Так, білки-ферменти каталізують протікання біохімічних реакцій і грають важливу роль в обміні речовин. Деякі білки виконують структурну або механічну функцію, утворюючи цитоскелет, що є важливим засобом підтримки форми клітин. Також білки грають важливу роль в сигнальних системах клітин, клітинній адгезії, імунній відповіді і клітинному циклі.

Білки — важлива частина харчування тварин і людини, оскільки ці організми не можуть синтезувати повний набір амінокислот і повинні отримувати частину з них із білковою їжею. У процесі травлення протеолітичні ферменти руйнують спожиті білки, розкладаючи їх до рівня амінокислот, які використовуються при біосинтезі білків організму або піддаються подальшому розпаду для отримання енергії.

Білки були вперше описані шведським хіміком Єнсом Якобом Берцеліусом в 1838 році, який і дав їм назву протеїни, від грец. *πρότα* — «першорядної важливості». Проте їхня центральна роль в життєдіяльності всіх живих організмів була виявлена лише у 1926 році, коли Джеймс Самнер показав, що фермент уреаза також є білком. Секвенування першого білка — інсуліну, тобто визначення його амінокислотної послідовності, принесло Фредеріку Сенгеру Нобелівську премію з хімії 1958 року. Перші тривимірні структури білків гемоглобіну і міоглобіну були отримані за допомогою рентгеноструктурного аналізу, за що автори методу, Макс Перуц і Джон Кендрю, отримали Нобелівську премію з хімії 1962 року.

### Протеоміка і біоінформатика

Повний набір білків, які в конкретний момент присутні в клітині, певному типі клітин або в організмі, називається протеомом, а дослідження великих наборів даних про ці білки та зв'язки між ними називається протеомікою (Рис. 6.2), що була названа за аналогією з геномікою. Головні експериментальні методи протеоміки включають: мас-спектроскопію, що дозволяє швидко ототожнення високих кількостей білків і встановлення пептидної послідовності, білкові мікрочипи, що дозволяють одночасне визначення відносних кількостей великого числа білків в клітині, двогібридний скринінг, що дозволяє систематичне дослідження білок-білкової взаємодії (встановлюючи їхню повну систему — інтерактом) та інші. Систематичні спроби визначення структури білків та всіх можливих станів кожного білка називаються структурною геномікою.

### Modern quantitative proteomics is central to understanding biology

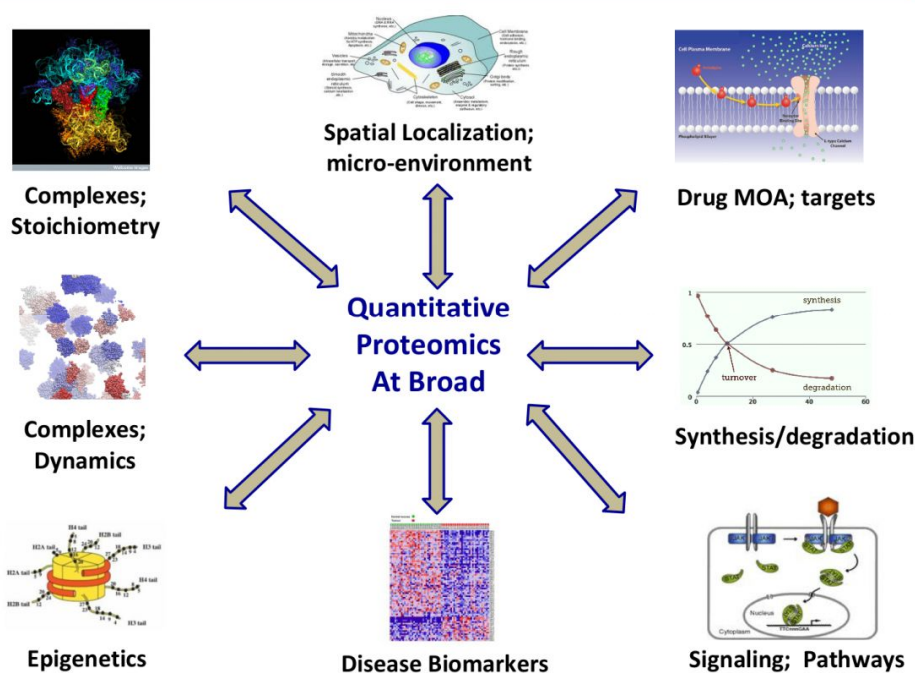


Рис. 6.2. «Кількісна протеоміка в цілому» – галузі застосування протеоміки.

Визначення просторової (тривимірної, 3D-) структури білків є необхідним етапом для встановлення взаємозв'язку між структурою та функцією білків.

Історично першим підходом для аналізу складної білкової суміші став **двовимірний електрофорез** у поліакриламідному гелі. У цьому методі білкова суміш послідовно розганяється у двох взаємно перпендикулярних напрямках у різних електрофоретичних

буферах (з різним рН, іонним складом тощо). Часто в першому напрямку відбувається розділення білків за їхнім зарядом, у другому – за молекулярною масою, що дозволяє ефективно розділити до 1 тис. білків одночасно.

Зручним і потужним методом розділення білкових сумішей є **високоєфективна рідинна хроматографія** (HPLC – High Performance Liquid Chromatography). Найчастіше розчин денатурованих білків наноситься на колонку, заповнену твердими пористими частинками, які адсорбують білки за рахунок гідрофобних взаємодій. Змивання (елюція) розчинником, який пропускається через колонку під високим тиском і полярність якого варіюється, дозволяє розділити окремі білки, що незначною мірою розрізняються один від одного за своєю гідрофобністю. Метод використовують як для аналізу, так і з метою виділення та очистки білків.

Найпотужнішим засобом аналізу складних білкових сумішей є **мас-спектрометрія**, принцип якої полягає в розділенні пучка заряджених частинок в електричному та магнітному полі на окремі фракції з однаковим відношенням маси до заряду. Сучасні методи іонізації зразків дозволяють іонізувати й переводити в газову фазу великі органічні сполуки, у тому числі олігопептиди та білки. Наявність 10–15 моль/л білка з молекулярною вагою до 200 тис. може бути детектована за допомогою сучасних мас-спектрометрів.

Проте визначення маси білка ще не дозволяє точно ідентифікувати його. З цією метою білок (після його виділення за допомогою двовимірного електрофорезу або рідинної хроматографії) піддають обмеженій протеолітичній деградації. Отримана комбінація олігопептидів, маса яких визначається за допомогою масспектрометрії, є своєрідним відбитком даного білка. Наступний аналіз геномних послідовностей дозволяє знайти відповідну амінокислотну послідовність білка, що практично позбавляє від необхідності хімічного визначення такої послідовності.

Поєднання масспектрометра з рідинним хроматографом – протеоліз білкової суміші, розділення олігопептидів методами рідинної хроматографії, аналіз їх за допомогою масспектрометрії та наступний пошуку баз даних з метою їхньої ідентифікації – це найперспективніший шлях розвитку кількісної протеоміки.

Наступним кроком у вивченні протеому є з'ясування складної картини білок-білкових взаємодій. Одним із найефективніших методів дослідження таких взаємодій є так званий **pull-down**: молекула певного білка, що використовується як пастка (bait), адсорбується на колонці (завдяки пришитому до білка специфічному «ярлику» – tag), через колонку пропускається білкова суміш, утворені білкові комплекси аналізуються за допомогою гелелектрофорезу. Ідентифікацію білків, вилучених із комплексів і розділених шляхом електрофорезу, можна також здійснити за допомогою мас-спектрометрії. Цілком аналогічно до ДНК-мікроареїв, у дослідженнях білок-білкових взаємодій використовують також білкові мікроареї – скельця з пришитими до них білками (понад 5 тис. білків), які дозволяють здійснювати скринінг протеому.

Переважає більшість відомих структур білків (близько 90% в Банку даних білків) були визначені за допомогою рентгенівської кристалографії. Близько 9% відомих структур визначені за допомогою ЯМР-спектроскопії, що також дозволяє отримання структур високої роздільної здатності. Останнім часом також набирає популярності кріоелектронна мікроскопія, як дуже швидкий, хоча і низькоточний, метод визначення структур білків. Хоча цей метод і не здатний визначити розташування окремих амінокислот або деталі вторинної структури, він широко застосовується для дослідження великих білкових комплексів.



**Банки даних послідовностей.** Інформація про визначені послідовності (первинні структури) біополімерів зберігається у спеціалізованих банках даних, які безперервно і автоматизовано поповнюються. Доступ до цих послідовностей можливий через пошук за ключовими словами або методом пошуку за гомологією. В записах банків даних наводяться додаткові відомості та різні перехресні посилання (наукові статті, інформація про вторинну структуру, різні аналізи, їх похибки тощо). Для генів і геномів це банки даних *GenBank*, *EMBL*, *DDBJ* та інші, а для білків — *GenPept*, *TrEMBL*, *Swiss-Prot*, *OWL*, *NBRF-PIR* та інші. Рекомендується використовувати автоматизовані банки даних послідовностей *GenPept* та *TrEMBL*, у яких зберігаються "трансляції" всіх нуклеотидних послідовностей з *GenBank* (*EMBL*). Інший банк даних — *Swiss-Prot* є "класичним" і містить розвинену систему анотацій та посилань на інші чисельні банки даних.

Важливим етапом комп'ютерного аналізу білка є пошук гомологічних амінокислотних (АК) послідовностей серед усіх відомих послідовностей. Існують численні програми пошуку за гомологією, які можуть використовувати різні матриці та алгоритми вирівнювання, і враховують не тільки ідентичні, а й схожі за властивостями АК-залишки. При використанні удосконалених методів вирівнювання багатьох послідовностей враховується не тільки формальна схожість двох послідовностей, а й консервативність окремих залишків.

Найбільш вживаною є програма BLAST та її чисельні спеціалізовані клони. Офіційний сервер — BLAST, існують варіанти цієї програми (*blastp*, *blastn*, *tblastn*, *blastx*, *tblastx*) для різних типів пошуку залежно від типу надісланої послідовності і банку даних, в якому іде пошук. Інтегровані бази даних NR (*non redundant*) — це ті, в яких видалені всі дублюючі послідовності. Подібними за призначенням є також програми *FastA*, *Blitz*. Варіант пошуку за гомологією, коли процес вирівнювання повторюється ітеративно з уточненням ролі найбільш консервативних залишків, дозволяє знаходити значно віддалені гомологи (програми *PSIBLAST* *PHI-BLAST*).

**Завдання:** 1. Ознайомитись з будовою поліпептидів.

2. Ознайомитись з методами визначення якісного та кількісного складу білків.
3. Навчитися проводити пошук амінокислотних послідовностей для різних білків в базах даних.
4. Розв'язувати задачі по визначенню маси та довжини білків.

**Питання для самопідготовки:** Хімічна будова поліпептидів. Властивості амінокислот. Функції білків. Молекулярна маса білків. Методи аналізу амінокислотних послідовностей та молекулярної маси пептидів. Біоінформатика: завдання і перспективи. Особливості роботи спеціалізованого програмного забезпечення по визначенню амінокислотних послідовностей.

**Задачі:** 1. Білок вазопресин складається з 9 амінокислот. Яка його молекулярна маса, якщо молекулярна маса однієї амінокислоти в середньому 100 а.о.м.?

2. Молекулярна маса каталази 22400 а.о.м. Скільки амінокислотних залишків в цій молекулі?
3. Один з ланцюгів ДНК має молекулярну масу 119025 а.о.м. Визначте кількість мономерів білка, закодованого у цьому ланцюгу ДНК.
4. Молекулярна маса кодуючої ділянки гена становить 172500 а.о.м. Визначте масу білкової молекули, яка закодована даним геном.
5. Білок складається з 400 амінокислот. Визначте довжину білка, якщо половина білкової молекули скручена в  $\alpha$ -спіраль.

6. Молекулярна маса пепсину – 35 000 а.о.м. Скільки амінокислотних залишків в складі молекули? Яка довжина первинної структури білка?
7. Ген еукаріотичного організму складається із 14100 нуклеотидів, причому на інтронні ділянки припадає 11400 нуклеотидів. За який час відбувається транскрипція та трансляція, якщо швидкість транскрипції у еукаріот складає 30 нуклеотидів за 1 секунду, а швидкість трансляції – 2 амінокислоти за 1 секунду.
8. Молекула білка гемоглобіну містить 0,34% Феруму (Fe). Обчисліть молекулярну масу молекули гемоглобіну.
9. Білок містить 0,5% гліцину. Чому дорівнює молекулярна маса цього білка, якщо  $M_{\text{гліцину}} = 75,1$  а.о.м.? Скільки амінокислотних залишків в даному білку?



## ПРАКТИЧНА РОБОТА №7

### Методи моделювання вторинної, третинної та четвертинної структур білків.

Встановлення амінокислотних (АК) послідовностей та просторових структур великої кількості білків викликало підвищену зацікавленість до вивчення природи сил, що стабілізують їх нативну конформацію. Було сформульовано спрощені вихідні принципи організації структури білків:

- 1) АК-послідовність однозначно визначає нативну конформацію білка;
- 2) згортання поліпептидного ланцюга в нативну структуру здійснюється спонтанно, прямо і неопосередковано;
- 3) стабільна конформація білка відповідає мінімальній вільній енергії Гіббса.

Зрозуміло, що жоден з цих принципів не є абсолютно вичерпним. Для багатьох встановлених структур нативна конформація відповідає тільки одному з локальних мінімумів вільної енергії. Але доповнення скоріше уточнюють, а не скасовують ці принципи.

Природна (нативна) АК послідовність набуває за фізіологічних умов досить визначеної, унікальної конформації, яка детермінує біологічну функцію даного білка. Білкову глобулу формують як ковалентні, так і слабкі нековалентні взаємодії — водневі зв'язки, гідрофобні, електростатичні та Ван-дер-Ваальсові взаємодії. Бічні радикали АК у молекулах білків визначають унікальні властивості нативної конформації білка.

Окрім структурної протеоміки, існує область досліджень, яка займається передбаченням структури білків і прагне розвинути ефективні методи створення правдоподібних моделей білків, структури яких не були визначені експериментально. Найуспішніший вид передбачення структури, відомий як моделювання гомології, використовує відому структуру «матриці» зі схожою послідовністю до модельованого білка; завданням є лише знайти відмінності та передбачити, яким чином вони впливатимуть на структуру. Хоча створення точних моделей залишається дуже складним, якщо доступні тільки приблизно подібні «матриці», вважається, що вузьким місцем методу є вирівнювання послідовностей. Це підтверджується тим, що у випадку «досконалого» вирівнювання можуть бути отримані дуже точні моделі. Багато методів передбачення структури призначені для використання в новій галузі проектування білків, що вже дозволила спроектувати кілька нових типів білкових структур. Складніша обчислювальна проблема — передбачення міжмолекулярних взаємодій, таких як молекулярний докінг і передбачення білок-білкової взаємодії.

Основним методом передбачення структури білків та білок-білкової взаємодії є симуляція процесів згортання та зв'язування білків із використанням методів молекулярної динаміки. Загалом ці методи вимагають величезних обчислювальних ресурсів. Для задоволення цієї потреби створюються сімейства найшвидших сучасних суперкомп'ютерів Blue Gene. Як альтернативу дослідники дедалі більше використовують розподілені обчислення, такі як проект Folding@Home («згортання вдома»). Згортання маленьких спіральних областей білків, наприклад «головки» віліну і допоміжного білка ВІЛ вже вдалося успішно просимулювати з атомною точністю *in silico*, а гібридні методи, що комбінують методи молекулярної динаміки із методами квантової хімії, дозволили просимулювати електронні стани родопсину.

На основі експериментально визначених структур білків розроблені емпіричні правила та алгоритми передбачення їх 3D-структури, які покладені в основу молекулярного моделювання просторової організації білків. Ці методи розвивалися спочатку як допоміжні при уточненні РСА- та ЯМР-структур, що дало змогу встановити загальні правила упаковки поліпептидних ланцюгів, а згодом розвинути новий підхід на основі конформаційного аналізу та мінімізації вільної енергії молекул білків.

Для більшості білків характерна модульна (мультидоменна) будова з окремих доменів, яким притаманні незалежні походження і структура. Аналіз доменної будови білків (кількості і границь доменів) проводиться з банками даних доменів *ProDom*, *Pfam*, *HSSP*, *SBASE*, *ProtFam*, *BLOCKS*.

Існують потужні інтернет-ресурси, які є світовими центрами з біоінформатики. Наприклад, *NCBI*, *ExPASy*, *BioPlanet*, *Frontier*.

Важливу інформацію щодо структури білка дає передбачення його вторинної структури та рівня експонованості АК залишків. Таке передбачення допомагає ідентифікувати окремі елементи структури ( $\alpha$ -спіралі,  $\beta$ -тяжі, повороти поліпептидного ланцюга, ділянки петель тощо), а потім і тип згортання окремих доменів, а також границі доменів.

Якість передбачення значно залежить від кількості і рівня гомології різних послідовностей у множинному вирівнюванні, з яким працюють програми передбачення вторинної структури. Наприклад, програма *PHD*, точність якої становить понад 72% для розчинних глобулярних білків. Подібними за призначенням програми *PSSP*, *PROF*, *SSpro*, *Predator*, *JPRED* та деякі інші.

Дані про експериментально встановлені 3D-структури білків методами РСА та ЯМР депонуються і зберігаються у банках даних просторових структур. Найважливішим з них є **PDB (Protein Data Bank)** — головний всесвітній депозитарій просторових структур. На сьогодні *PDB* містить понад 15 тис. встановлених структур різних білків. Кожній структурі відповідає окремий файл з просторовими (декартівськими) координатами окремих атомів. В анотації до *PDB*-файлів описуються важливі структурні дані. Існує багато форматів таких координатних файлів, які розпізнаються різними програмами. Стандартом *de facto* для білкових структур є формат банку даних *PDB* (файли типів \*.pdb та \*.ent). Існують зручні програми з конвертації таких файлів у потрібний формат, наприклад, *BabelWin*.

До споріднених банків даних належать *MMDB*, *3Dee*, *FSSP* та інші, які є загальнодоступними в інтернет-мережі. Дані про структури окремих білкових доменів систематизуються у спеціалізованих банках даних у вигляді певних ієрархічних систем (наприклад, Class - Fold - Superfamily - Family - Domain - Organism Variant). До найбільш цитованих банків даних належать *SCOP* (Structural Classification of Proteins), *CATH*, *CE*, які мають власні розвинені пошукові системи та програми.

Для візуалізації та зміни типу зображення (рендерінгу) просторових структур використовують програми-вьювери (переглядачі). Такі програми дозволяють проводити виділення за вибраними параметрами (електростатичний потенціал, тип атомів, елементи вторинної структури, окремі АК залишки, доступна для розчинника поверхня, кути та довжини зв'язків, водневі зв'язки, Ван-дер-Ваальсові радіуси тощо). Програма *WebLab* — це простий, але потужний вьювер для перегляду та редагування координатних файлів різних форматів. Програма *Swiss-PDB viewer* містить також функції різних необхідних підпрограм-аналізаторів, а також є інтегрованою з сервером *Swiss-Model*, що полегшує роботу з цим сервером. До поширених вживаних вьюверів можна також віднести *MolMol*, *Protein Explorer*

і *RasMol*. Також для перегляду зображень (слайдів), записаних у різних форматах, які легко генеруються програмами-вьюверами, можна використовувати Windows-програму загального призначення *ACDSee*.

### **Класифікація методів передбачення структури білків**

Ці методи спрощено можна розділити на три класи за повнотою фізичного підходу до їх опису:

1) квантово-механічні методи розрахунку *ab initio*, які базуються на вихідних фізичних принципах та враховують усі електрони атомів (дедуктивні методи); 2) напівемпіричні, які враховують взаємодію валентних електронів атомів; 3) методи класичної молекулярної механіки (емпіричні індуктивні методи), які не враховують електрони, але використовують статистично встановлені правила у вигляді певних силових полей. Теоретично можливо, виходячи з базових фізичних принципів, розраховувати структуру макромолекул без додаткових емпіричних даних, що становить напрямок "строного" моделювання конформації білків на потужних суперкомп'ютерах. Напівемпіричні методи характеризуються різними підходами (MNDO, CNDO тощо), що базуються на різних типах врахування взаємодії валентних електронів.

Моделювання за гомологією включає наступні етапи: 1) пошук у банках даних білків-матриць з експериментально визначеною просторовою структурою, які мають високий ступінь гомології з АК послідовністю білка-мішені, для якого будується структурна модель; 2) вирівнювання послідовності, яку необхідно змодельовати, з однією чи декількома послідовностями-матрицями. Якщо ідентичність досліджуваної послідовності порівняно з матрицями понад 40%, то якість вирівнювання визначає якість моделі; 3) корекція вирівнювання; 4) генерація ковалентно-неперервного ланцюга (каркаса) моделі на основі вирівнювання. Це може бути каркас найбільш гомологічної матриці або гібридний з кількох матриць; 5) генерація "канонічних" поверхневих петель, отриманих з банків даних. Петлі визначають переважно конформаційні відмінності гомологічних білків; 6) "вбудова" бічних радикалів у каркас та їх оптимізація. Консервативні залишки мають однакові конформації; 7) добудова петель *ab initio*, якщо в них є інсерції або делеції; 8) мінімізація вільної енергії всієї моделі, іноді з використанням молекулярної динаміки; 9) перевірка моделі вибірково повторенням попередніх етапів.

Моделювання за гомологією можна безпосередньо проводити в он-лайн за допомогою моделюючих систем *Swiss-Model*, *Modeller*, *CPHmodels*, *SDSCI*, *FAMS*, *3D-JIGSAW*. За допомогою цих серверів можна автоматично провести пошук білків-матриць, саме моделювання, оптимізацію геометрії і перевірку якості отриманої моделі. Деякі сервери надають повний контроль за процесом побудови моделі (крок за кроком).

### **Проектування білків**

Важливою областю досліджень сучасних молекулярної біології та генної інженерії стало не тільки вивчення білків, створених природою, або комбінування їх у штучних білках, але й проектування принципово нових білків із потрібними властивостями. Методи проектування білків можна розбити на дві головні групи: раціональне проектування та направлену еволюцію.

В методі проектування білків робота найчастіше починається із знаходження природного білка із відомою структурою, найближчого за властивостями до потрібного, після серії таких мутацій можна отримати новий білок. Хоча невеликі зміни активно вносяться у білки починаючи з середини 1980-х років, зараз стало можливим конструювати

відносно складні білки, наприклад, рецептори нових сполук, та конструювати білки de novo, без використання природного шаблону.

В альтернативному методі направленої еволюції використовується подібний до необхідного природний білок, до якого додаються випадкові мутації, а в результаті сконструйованого випробування відбираються найкращі примірники. Цей процес може бути повторений багаторазово, часто з додаванням рекомбінації частин різних успішних білків (аналог гомологічної рекомбінації). Перевагою методу є непотрібність будь-яких знань про структуру і методи роботи білка, проте його недоліком є неможливість легкого отримання деяких білків у великих кількостях та рекомбінантні маніпуляції з ними.

- Завдання:**
1. Ознайомитися із сучасними методами моделювання структурних рівнів білків.
  2. Ознайомитися із програмним забезпеченням по моделюванню структури білків.
  3. Ознайомитися із сучасними напрямками генної інженерії.

**Питання для самопідготовки:** Будова та структурні рівні організації білків. Вторинна структура білків. Глобулярна структура. Неструктуровані білки. Принципи функціонування білків. Класифікація методів моделювання білків. Проблематика сучасної біоінформатики. Білкова інженерія.

## ПРАКТИЧНА РОБОТА №8

### Геном людини. Генетичні захворювання.

ДНК однієї клітини людини може зберігати близько 1,5 Гб інформації. Водночас ДНК усіх клітин людського організму займають 60 мільярдів Тб, що зберігаються лише на 150-160 грамах ДНК. Таким чином, молекулу ДНК можна вважати найкращим у світі «жорстким диском».

**Геном** (нім. *Genom* від гр. *Genos* – походження) – сукупність генів гаплоїдного набору хромосом певного виду; основний гаплоїдний набір хромосом. Термін запровадив німецький ботанік Г. Вінклер у 1920 році.

Геном людини — геном *Homo sapiens* – складається з 23 пар хромосом: 22 аутосомних + X + Y. Розмір генома людини в розрахунку пар нуклеотидів на гаплоїдний набір хромосом має в сумі 3,2 мільярдів пар основ – bp (від англ. basic pair, пара основ ДНК) та містить приблизно 20 000-25 000 кодуючих генів. Повний каріотип, тобто диплоїдний набір хромосом, містить вдвічі більше ДНК. Проект геному людини привів до отримання послідовності еухроматину геному людини, яка використовується у всьому світі в біомедичних науках. Геном людини має менше генів, ніж очікувалося раніше, і тільки приблизно 1,5% геному кодує білки, решту складають РНК-гени, регуляторні послідовності, інтрони та інша некодуюча ДНК.

Дані проекту ENCODE у 2012 р. свідчать про те, що більшість геному людини (80,4%) бере участь у РНК-хроматин взаємодіях, тоді як 95% ДНК послідовності лежить на відстані 8 тис. пар основ від ДНК-білкової взаємодії, а 99% геному лежить на відстані 1,7 тисяч пар основ від принаймні однієї біохімічної взаємодії, що зафіксована в межах проекту ENCODE. В межах проекту ENCODE є частина GENCODE, що відповідає за регіони ДНК, що кодують гени (в які входить як і білок-кодуючі РНК, так і некодуючі РНК, і псевдогени). GENCODE у 2012 р. показав, що у людини існує 20 687 генів що кодують білки, а альтернативний сплайсинг в середньому дає 6,3 варіантів мРНК. Також ідентифіковано 11 224 псевдогенів, 863 з яких транскрибуються і асоціюються з хроматином.

**Загальна характеристика геному людини.** Завдяки реалізації програми „Геном людини”, що завершена у 2001 р у США, повністю розшифрований весь геном людини. Встановлені всі 24 теоретично можливі групи зчеплення генів, з яких 22 містяться в аутосомах. У кожній з них міститься по кілька сотень генів. Більше 100 генів локалізовано у статевих хромосомах (92 – в X-хромосомі; в Y-хромосомі – 9 генів, гомологічних з X-хромосою та 5 – не гомологічних з X-хромосою).

Весь набір конденсованих мітотичних хромосом має назву **каріотип** (термін ввів російський цитолог Левітський у 1924 р. Точну кількість хромосом у людини встановили шведські вчені Тийо та Леван у 1956 р.

#### Молекулярна організація хромосом

Типова метафазна хромосома складається з 2-х сестринських молекул ДНК, які укладені окремо одна від одної у вигляді хроматид і сполучені між собою за допомогою центромери. Такі мітотичні хромосоми звичайно несуть на своїй поверхні багато інших молекул, зокрема, рибонуклеопротейнів.

Речовина хромосом еукаріотів – хроматин являє собою нуклеопротейд, головними складовими частинами якого є ДНК та 2 види білків: основні (**гістони**) та кислі (**негістонові білки**).

У людини генотип складається з приблизно 3 млрд. пар нуклеотидів. Загальна довжина всіх нуклеотидних послідовностей – біля 2 м. Оскільки вони розподілені по 46 хромосомам, то в розтягнутому стані довжина подвійної спіралі ДНК, що міститься в кожній хромосомі людини, складає в середньому біля 4-5 см. Сумарна товщина молекул ДНК у диплоїдній клітині складає біля 180 см. Загальна молекулярна маса геному людини –  $10^{12}$  а.о.м.

Кожна еукаріотична хромосома містить лише одну молекулу ДНК. Виходячи з цього, відповідні хромосомам молекули ДНК людини в середньому повинні містити по  $(3 \times 10^9) : 23 = 1,3 \times 10^8$  нуклеотидів.

За допомогою гістонів така довга молекула укладається у клітинному ядрі, діаметр якого дорівнює лише декілька мкм.

**Укладання ДНК в хромосомі.** На теперішній час відомо декілька рівнів укладання ДНК:

1. Основною структурною одиницею **1-го рівня** укладання хроматину є **нуклеосома** – комплекс ДНК та гістонів, що виглядає на електронних мікрофотографіях як частинка дископодібної форми з діаметром біля 11 мкм. Кожна нуклеосома містить набір з 8 молекул висококонсервативних гістонів – по 2 кожного виду (H2A, H2B, H3 та H4). Це її білкова серцевина, або **гістоновий кор**, на який накручується фрагмент 2-ланцюгової ДНК довжиною у 146-250 пар нуклеотидів. ДНК тягнеться у вигляді неперервної нитки від нуклеосоми до нуклеосоми. Ділянки ДНК між нуклеосомами (**лінкерні** ділянки) мають різну довжину (в середньому біля 6-60 пар нуклеотидів). Нуклеосома та відповідна лінкерна ДНК складають повторювану структурну одиницю хроматину, що містить в цілому приблизно 200 пар нуклеотидів. Таким чином, весь гаплоїдний геном людини ( $3 \times 10^9$  пар нуклеотидів) може бути укладений у  $1,5 \times 10^7$  нуклеосом. Нуклеосоми та ДНК, що їх огортає, утворюють спіраль з діаметром біля 13 нм і відстанню між витками (крок спіралі) – біля 50 нм.

2. На **2-му рівні** укладання нуклеосомна нитка укладається у **30 нм хроматинову фібрилу**. Хроматин одної хромосоми людини, що містить молекулу ДНК довжиною 5 см, у вигляді нуклеосомної нитки мав би довжину біля 2 см, а у вигляді 30нм-хроматинової фібрили – довжину 1,2 мм. За укладання нуклеосомної нитки у фібрилу відповідає гістон H1. Амінокислотна послідовність H1 зазнала в ході еволюції значно більших змін, ніж структура нуклеосомних гістонів. Молекула H1 має видовжені кінцеві N- і C-ділянки, якими він зв'язується з лінкерною ділянкою ДНК однієї нуклеосоми та гістоновим кором сусідньої нуклеосоми, притягаючи сусідні нуклеосоми одну до одної.

Можуть також утворюватися цілі довгі послідовності молекул H1 вздовж молекули ДНК. Отже, хроматин набуває блочної структури, причому кожний з блоків може незалежно від інших конденсуватися і деконденсуватися згідно рівня активності генів, що містяться в ньому. Це супроводжується перебудовою або видаленням цілої групи молекул H1. Крім того, існують різні варіації молекул H1 і зв'язування саме певної з них з даним блоком хроматину є ще одною причиною диференціальної активності хроматину.

З окремими нуклеосомами зв'язуються також негістонові білки (1 млн молекул на клітину або 1 молекула на 10 нуклеосом), що беруть участь у процесах транскрипції генів. Найважливішими з них є два білки: HMG-14 та HMG-17, що знайдені у всіх клітинах ссавців. Вони завжди пов'язані з нуклеосомами, що містять активні гени, і приєднуються безпосередньо до гістонового кору.

Всі названі білки є лише окремими представниками тисяч інших ще не відкритих білків, що здатні зв'язувати ДНК і брати участь в процесах транскрипції. В багатоклітинному організмі вони забезпечують тканинноспецифічний характер білкового синтезу, ініціацію синтезу ДНК, формування окремих блоків (доменів) у довгому полінуклеотидному ланцюгу

ДНК. Білки можуть розпізнавати специфічні ділянки ДНК завдяки тому, що утворюють водневі зв'язки з певними азотистими основами, а також „відчувають” особливості геометрично-просторової конформації різних ділянок ДНК.

Існує також механізм регуляції генної активності, що може пояснити природу клітинної пам'яті. Він пов'язаний з метилюванням (або деметилюванням) специфічними ферментами залишків цитозину у ДНК: функціонально важливі ділянки ДНК можуть стабільно знаходитися у метилюваній (неактивній) або неметилюваній (активній) формі протягом багатьох клітинних генерацій. При активації генів відбувається деметилювання ДНК.

3. На **3-му рівні** укладання ДНК відбувається укладання ДНК у вигляді петель. Такі петлі формуються і підтримуються за допомогою білків, що здатні розпізнати віддалені специфічні нуклеотидні послідовності та зблизити їх, утворюючи петлю. Середня довжина петель – 20 000 – 80 000 пар нуклеотидів. Таким чином, типова хромосома людини може містити  $(1,3 \times 10^8) : 50\ 000 = 2600$  петель, кожна з яких утворена ділянкою хроматин-нової фібрили із середньою довжиною біля 400 нм (0,4 мкм). Така зібрана у петлі ДНК людини матиме у довжину вже біля 100 мкм. Але для того, аби вона могла уміститися всередині клітинного ядра, вона повинна згорнутися у ще більш щільну структуру.

4. На **4-му рівні** укладання ДНК, яке відбувається під час мітозу у вигляді суперспіралізації або конденсації ДНК, лінійні розміри ДНК людини зменшуються врешті-решт до 5 мкм. Це супроводжується фосфорилуванням в клітині всіх молекул гістону H1.

### **Організація ДНК в геномі людини**

Більша частина геномної ДНК вищих організмів і людини не кодує білків. Це доводиться такими фактами:

- Відносна кількість ДНК в геномі різних організмів не пов'язана з їхнім положенням на еволюційному дереві. Так, клітини людини містять у 800 разів більше ДНК, ніж клітини бактерії кишкової палички (*Escherichia coli*), але кількість ДНК в клітинах деяких амфібій та рослин у 30 разів вища, ніж у людини.

- У близько споріднених видів вміст ДНК в клітині може сильно варіювати (наприклад, у різних видів амфібій – в 100 разів).

- Оскільки зміни ДНК (мутації) шкідливі для організму, якщо вони зачіпають життєво важливі послідовності ДНК, то повинно існувати якесь визначене число таких послідовностей, які забезпечують еволюційну стабільність геному. Виходячи з відомої частоти мутацій, було визначено, що у ссавців не більше 1% їх геному припадає на долю ДНК, що кодує та регулює синтез життєво важливих білків.

Отже, хоча геном ссавців має біля  $3 \times 10^9$  нуклеотидів, чого достатньо для кодування майже 3 млн. білків, жоден організм не здатний мати більше ніж 30 000 реально кодуючих індивідуальних послідовностей ДНК. Інша ДНК, що складає основну масу геному вищих організмів, виконує структурну роль, беручи участь в укладанні хроматину.

Нуклеотидні послідовності ДНК вищих організмів і людини можна розділити на 3 групи за ступенем їхньої повторюваності у геномі: **унікальна, сателітна та помірно-повторювана ДНК.**

Більшість послідовностей ДНК геному унікальні, тобто представлені в одиничному екземплярі. В клітинах ссавців біля 70% нуклеотидних послідовностей є такими неповторними унікальними генами. Саме вони визначають видову специфічність ДНК організмів. Решта 30% сумарної ДНК представлена послідовностями, що багаторазово повторюються в геномі: сателітна ДНК та помірно повторювана ДНК.

**Сателітна ДНК** складає близько 10% клітинної ДНК. Вона являє собою короткі послідовності нуклеотидів, що повторюються цілими серіями (мільйони і десятки мільйонів копій на геном). Повторювальна одиниця у таких послідовностях може складатися усього з 1-2 нуклеотидів, але найчастіше її довжина складає 170-250 нуклеотидів. Такі послідовності мають незвичайний нуклеотидний склад, тому можуть бути відділені від основної маси клітинної ДНК як окремий невеликий компонент – „сателіт” („супутник”). Послідовності сателітної ДНК, ймовірно, не транскрибуються і в більшості випадків локалізовані в гетерохроматині центромерних зон хромосом.

Послідовності сателітної ДНК в процесі еволюції зазнавали надзвичайно швидких змін і навіть змінювали свою локалізацію в хромосомах. Наприклад, геном людини містить, як мінімум, три домінуючі послідовності сателітної ДНК, котрі у різних пропорціях присутні в кожній центромері. Якщо порівняти дві гомологічні митотичні хромосоми будь-якого індивідуума, то практично завжди з'ясується, що деякі перебудови у сателітних послідовностях, що успадковані від батька, сильно відрізняються від таких у послідовностях, що успадковані від матері.

Еволюційну нестабільність послідовностей сателітної ДНК можна пояснити її здатністю до багаторазових рекомбінацій: в цих ділянках відбувається парування гомологічних ділянок двох сестринських молекул ДНК та обмін фрагментами між ними. Крім того, повторювані послідовності у сателітній ДНК провокують дуплікації та делеції великих ділянок між двома новоутвореними сестринськими молекулами ДНК.

В принципі будь-яка послідовність ДНК може дуплікуватися, а утворена копія – підлягати подальшій модифікації. Отже, утворяться додаткові білки, що чимось відмінні від вихідного. Диплоїдні організми завжди мають „у запасі” другу, „зайву” копію кожного гена, що може модифікуватися без ризику для організму втратити вихідний білок. Таким чином, наприклад, могла виникнути серія відмінних між собою у різних організмів глобінів, які є основою молекули гемоглобіну. У людини гени, що кодують різні глобіни, складають групу гомологічних послідовностей ДНК (**кластер**), що розташована в одній з хромосом на ділянці довжиною у 50 000 нуклеотидів. Аналогічним шляхом, ймовірно, виникли різні типи імуноглобулінів, альбумінів та колагенів.

Наявність в еукаріотичній ДНК некодуючих вставок – **інтронів** – значно прискорює процес нагромадження дуплікацій, бо довгі інтрони значно примножують кількість потенційних точок генної рекомбінації та дуплікації. При відсутності інтронів змістовна послідовність ДНК могла б містити лише невелику кількість ділянок, у яких рекомбінаційний обмін ділянками між сестринськими молекулами ДНК міг призвести б до дуплікації цілого функціонального активного гена. Отже, інтрони в цілому підвищують частоту виникнення нових генів, таким чином, прискорюють еволюцію.

Далеко не кожна дуплікація ДНК призводить до появи нового функціонального гена. Так, деякі дупліковані послідовності глобінової ДНК, що локалізовані у тій же ділянці хромосоми, не є справжніми генами. Ці послідовності називають **псевдогени**. Вони гомологічні функціонально активним генам, але не здатні до експресії внаслідок мутаційних змін, що відбулися з ними.

Отже, еволюційну історію людини можна прослідкувати по нашим хромосомам, якщо детально проаналізувати подібність і відмінності гомологічних генів на різних ступенях філогенетичного дерева.

**Помірно повторювана ДНК** – це короткі розсіяні по геному повторення, що складають біля 20% клітинної ДНК. До таких помірних повторень відноситься ряд



функціонально важливих ділянок геному, наприклад, гени, що кодують р-РНК, 5S-РНК, т-РНК та гістони. Але повторювані послідовності, функції котрих відомі, складають менше 1% геному еукаріот. Природа ж більшості помірних повторень лишається нез'ясованою.

Так, в клітинах людини більша частина несателітної повторюваної ДНК представлена єдиною домінуючою родиною генів. Дана група ділянок ДНК відрізняється від сателітної ДНК тим, що:

- ці послідовності являють собою не **тандемні** (блочні) повторення, що локалізовані в певних ділянках хромосом, а повторення, що розсіяні по всьому геному у вигляді сотень тисяч індивідуальних копій, кожна з яких має довжину біля 300 нуклеотидів;
- вони ефективно транскрибуються як у складі довгих тандемних РНК, так і у вигляді самостійних коротких молекул РНК;
- ці послідовності в ході еволюції ссавців змінювалися по своїй первинній структурі і положенню в геномі набагато повільніше, ніж елементи сателітної ДНК.

Ця родина розсіяних повторень, що нараховує  $10^5$ - $10^6$  копій на гаплоїдний геном, виникла дуже давно, навіть ще до того моменту, коли виокремилася еволюційна лінія приматів. Всі члени цієї родини повторень, напевно, пішли від однієї предкової послідовності, котра розмножилася і як своєрідний паразит розповсюдилася по всьому геному наших далеких предків. Ця первинна послідовність ДНК могла дуже нагадувати вірус або транспозон (мобільний генетичний елемент). Ці рухливі елементи геному ампліфікуються (багаторазово реплікуються) і дають початок множині ідентичних послідовностей-копій, які здатні вбудовуватися у нові ділянки основної клітинної ДНК та реплікуватися у складі її хромосоми, отримуючи можливість пасивно зберігатися в еволюції.

Конкретні функції цих послідовностей поки що невідомі, хоча теоретично вони можуть бути точками початку реплікації ДНК, сигналами процесингу та сплайсингу РНК, контрольними ділянками, що регулюють транскрипцію генів, або ж структурними елементами хромосом. Можливо, що вони взагалі не виконують в клітині ніяких функцій.

**„Егоїстична” ДНК.** Можливо, деякі високо повторювані послідовності ДНК, що присутні у геномі сучасних вищих організмів, являють собою фракцію „егоїстичної” ДНК, яка збереглася в процесі еволюції тільки завдяки своїй властивості реплікуватися і розповсюджуватися в межах геному, не завдаючи шкоди організму.

Егоїстична ДНК складає частину властивого вищим організмам величезного надлишку ДНК, що не кодує білки та, ймовірно, не має життєво важливого значення. Але з часом характер її взаємостосунків з клітинною ДНК набуває форми симбіозу. Для пом'якшення впливу егоїстичної ДНК на основну ДНК (і для збільшення шансів на своє власне збереження) на кінцях мобільних послідовностей можуть формуватися прикордонні послідовності нуклеотидів, що відповідають інтронним зонам основних генів. Отже, ці зони можуть слугувати контрольними точками сплайсингу, тобто надавати організму додаткові можливості в удосконаленні механізмів генного контролю та утворення нових білків. Тоді організм, що має таку егоїстичну ДНК, отримує певні переваги перед іншими організмами.

**Патологічні зміни геному** пов'язані з виникненням шкідливих мутацій. За розрахунками Штерна, у кожному поколінні виникає 2% людей з патологічними мутаціями у гетерозиготному стані (генетичний вантаж).

У людини досліджено характер успадкування декількох тисяч генів. Серед них є домінантні і рецесивні. Деяка частина генів кодує нейтральні ознаки (форма вух, колір очей і волосся, наявність ластовиння тощо). Але більшість генів формують життєво важливі ознаки

організму (наприклад, домінуючий ген резус-фактору, гени аглютиногенів А і В). Найменша зміна подібних генів спричиняє розвиток вад і спадкових хвороб, наприклад, домінуючі аутосомні гени полідактилії, карликовості; рецесивні аутосомні гени фенілкетонурії, тирозинкетонурії, алькаптонурії, альбінізма; рецесивні зчеплені з Х-хромосомою гени гемофілії, дальтонізму; зчеплені з У-хромосомою гени іхтіозу, гіпертрихозу. У 1% новонароджених наявні хромосомні вади (значні хромосомні аномалії спричиняють загибель ще на стадії зародка або плоду).

Розшифрування будови геному людини – відкриття, яке дорівнює за силою впливу на життя людини у всіх її проявах можна поставити в один ряд з відкриттями будови Сонячної системи, Теорії відносності та Періодичної системи Менделєєва. Сто років знадобилося людству, щоб від моменту народження генетики як науки розкрити таємницю будови геному і підійти до вивчення його функції, що призвело, в свою чергу, до відкриття молекулярної медицини. Якщо протягом минулих 100 років з моменту народження генетики як науки накопичувалася інформація про види і варіанти спадкових хвороб, про будову геному людини, його нормальної і патологічної анатомії, то кінець ХХ століття ознаменувався повною розшифровкою його будови. А результатом стало народження молекулярної медицини, яка народилася від злиття фундаментальної науки та практичної медицини і ознаменувала початок в ній епохи генетики, в яку нам доводиться жити.

Вивчення будови генома дозволило встановити локалізацію генів, зміни в яких приводять до декількох тисяч спадкових хвороб, визначити гени частих і рідкісних спадкових хвороб, застосувати методи діагностики в генетичній практиці.

Все частіше глибоке клінічне та молекулярне дослідження дозволяє знаходити мішені тяжких хвороб, які розташовані на нитці ДНК і нерідко діагностуються вперше (Рис. 8.1). Така форма результатів обстеження поки що незвична для більшості лікарів, але недалеко той час, коли вона витіснить багатослівні описи діагнозу.

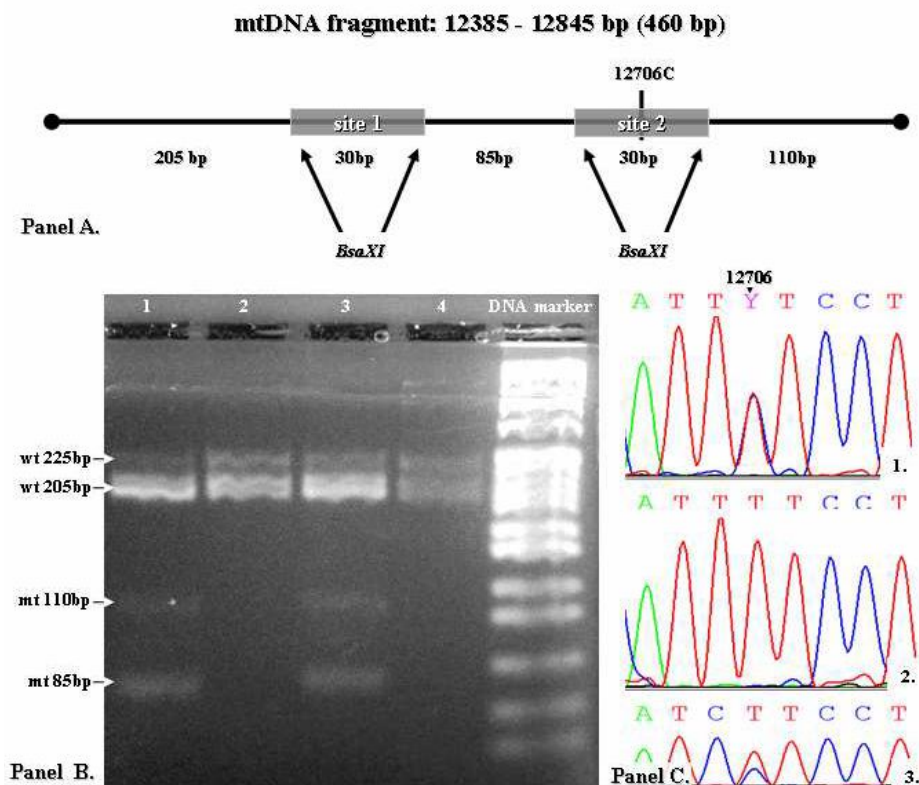


Рис. 8.1. Схематичний огляд методу ПЛР-ПДРФ з виявлення мутації 12706С.

Панель А. Діаграма ампліфікованого фрагмента м-ДНК із зазначенням положення сайту рестрикції *BsaXI*. У дикому типі м-ДНК відсутній другий рестрикційний сайт *BsaXI*, що призводить до поділу на 3 фрагменти по 225, 205, і 30 п.н. При наявності мутації 12706С в рестрикційному сайті *BsaXI* відбувається поділ на п'ять фрагментів по 205, 110, 85, 30 і 30 п.н.

Панель В. Зображення ампліфікованого фрагмента гена ND5 з рестрикцій *BsaXI* в гелі Пробанда К. та Л. (лінії 1 і 3) містять комбінацію молекул мутантного та дикого типів, у той час як, у матері пробанда К. (лінія 2), як і в негативному контролі (лінія 4), спостерігаються тільки смуги дикого типу з відсутністю мутації.

Панель С. Представлені електрофореграми пробанда К (1), його матері (2), і пробанда Л (3), з нумерацією ліній, яка відповідає Панелі В.

Сьогодні стали доступними молекулярній діагностиці багато спадкових хвороб: агамаглобулінемія; синдром адреногенітальний; альбінізм типу ОСА 1; синдром Альпорта; недостатність альфа-1-антитрипсину; аміотрофія спінальна Вердніга-Гоффмана; аміотрофія Кугельберга-Веландер; аміотрофія невральна Шарко-Марі-Тута; аміотрофія спінальна Х-зчеплена; синдром Анжельмана; синдром Апера; атаксія Фрідрейха; ахопроплазія; синдром Беквита-Відемана; бета-таласемія; хвороба Віллебранда; вроджена контрактурна арахнодактілія; вроджена м'язова дистрофія тип Фукуяма; гемофілія А; гемофілія В; гепатолентикулярна дегенерація (хвороба Вільсона-Коновалова); сімейна гіперхолестеринемія; гіпофізарний нанізм (дефіцит гормону росту); глікогенози; глухота нейросенсорна несиндромальна; синдром Грейга; дефіцит ацил-КоА дегідрогенази; синдром подовженого інтервалу QT; синдром Жильбера; зонулярна катаракта; синдром Коффіна-Лоурі; синдром Криглера-Найара; синдром лімфопроліферативний; Х-зчеплені (хвороба Дункана; синдром Пуртільо); лімфідема Мінроя; синдром Луї-Бар (телеангіоектазія); синдром Леш-Ніхана; синдром Мартіна-Белл (ламкої Х-хромосоми); синдром Марфана; синдром Міллера-Дікер; синдром Мілроя; мідистрофія Дюшенна/Беккера; мідистрофія Емері-Дрейфуса; міотонічна дистрофія; муковісцидоз; мукополісахаридози; несиндромальна нейросенсорна приглухуватість; синдром Ніймеген; хвороба Норрі; мідистрофія окулофарингеальна; періодична хвороба; псевдоахондропластична дисплазія; синдром Прадера-Віллі; рецесивний полікістоз нирок; ретинобластома; синдром Сміта-Лемлі-Опітц; спастична параплегія Штрюмпеля; спино-бульбарна м'язова атрофія (хвороба Кеннеді); синдром тестикулярної фемінізації; хвороба Унферріхта-Лундберга; фенілкетонурія; хвороба Хантера; синдром Хольт-Орана; Хорея Гентингтона; ектодермальна ангідротична дисплазія; синдром Елерса-Данло (класичний тип).

Молекулярна медицина поділена на 3 практичні складові: молекулярна діагностика, профілактична медицина і генна терапія (введення нуклеїнових кислот в клітину з метою лікування). Найбільшого успіху в розвинених країнах сьогодні досягли молекулярна діагностика та профілактична медицина. Що ж до генної терапії спадкових і неспадкових хвороб, то вона, стрімко розвиваючись, все ще має масу невирішених проблем, тому що при генній терапії ген набуває функції лікувального препарату і забезпечує індивідуальність лікування хворого. Стало можливим зупинити зростання злоякісних пухлин, позбавивши їх клітини «безсмертя», яким ракові клітини відрізняються від звичайних, повернути їм властивість апоптозу (програмованої загибелі). Якщо таке лікування вдається, то ракові клітини гинуть, а людина одужує. Така генна корекція вже врятувала деяких хворих і інтенсивно розробляється в світі генної терапії. Як тільки розробники досягнуть стійких позитивних результатів, такий вид лікування стане доступним.

Профілактична медицина є найбільш перспективним напрямком молекулярної медицини, а її розвиток визначається розширенням числа генетичних тестів, які дозволяють застосовувати досімptomатичну діагностику спадкових хвороб, для яких розроблені методи ефективної терапії, що позитивно впливають на показники дитячої смертності. Реалізувалася можливість виявлення носійства патологічних генів в родинах високого ризику спадкової патології. Реальною стала допологова діагностика, як метод однозначного прогнозування та вторинної профілактики. Розроблено підходи до генетичного тестування генів схильності до спадкових захворювань, які проявляються в дорослому періоді, онкогенетичних захворювань і поширених мультифакторіальних хвороб.

Проте без точного діагнозу використання профілактичної медицини нереально. Чим вище рівень діагностики спадкової патології, тим ефективніше розвиток всіх складових молекулярної медицини. Рівень діагностики забезпечується:

- Високим рівнем медико-генетичного консультування (оцінкою історії сім'ї, хвороби; зовнішніх даних людини і його внутрішніх органів; аналізом родоводів);
- Оцінкою біохімічних характеристик людини, яка відображає правильність роботи генів;
- Адекватною оцінкою виявленої мутації як причини фенотипових проявів.

- Завдання:**
1. Ознайомитися із будовою геному.
  2. Ознайомитися із молекулярною будовою хромосом.
  3. Навчитися класифікувати та визначити основні типи генетичних хвороб людини.

**Питання для самопідготовки:** Поняття геному. Молекулярна організація хромосом. Групи повторюваностей ДНК. Історія дослідження геному людини. Класифікація генетичних хвороб людини. Молекулярна діагностика спадкових хвороб. Прикладні напрямки застосування білкової інженерії в біотехнології та медицині.

**Задачі:** 1. У п'ятирічної дитини порушений тирозиновий обмін. Це приводить до ураження нервової системи і недоумкуватості, але легко лікується спеціальною дієтою, призначеною у ранньому віці. Яке це захворювання?

- А гемофілія;
- Б цистинурія;
- В фенілкетонурія;
- Г брахідактилія;
- Д таласемія.

2. Під час аналізу сечі тримісячної дитини виявлено підвищену кількість гомогентизинової кислоти, сеча при триманні її на повітрі набуває темного забарвлення. Для якого із перелічених нижче захворювань характерні описані зміни?

- А алькаптонурії;
- Б альбінізму;
- В аміноацидурії;
- Г цистинурії;
- Д фенілкетонурії.

3. Чотирирічна дівчинка має вивих кришталіків, довгі і тонкі пальці, спадкову ваду серця і високий рівень оксипроліну (амінокислота) у сечі. Всі ці дефекти викликані аномалією сполучної тканини. Для якого захворювання характерні ці клінічні симптоми?
- А синдрому Марфана;
  - Б фенілкетонурії;
  - В гіпофосфатемії;
  - Г фруктозурії;
  - Д галактоземії.
4. У разі хвороби Вілсона-Коновалова у тканинах мозку і печінки накопичується і викликає їх дегенерацію:
- А фосфор;
  - Б тирозин;
  - В феніланін;
  - Г ліпіди;
  - Д мідь.
5. Позитивна реакція проби Феллінга, затхлий специфічний запах сечі і поту, уповільнений моторний і психічний розвиток із шестимісячного віку, висвітлення волосся характерні для:
- А синдрому Шерешевського-Тернера;
  - Б галактоземії;
  - В фруктозурії;
  - Г фенілкетонурії;
  - Д синдрому Патау.
6. У разі амавротичної ідіотії Тея-Сакса розвиваються необернені тяжкі порушення центральної нервової системи, які призводять до смерті. Визначте причину захворювання:
- А розлад вуглеводного обміну;
  - Б розлад амінокислотного обміну;
  - В розлад мінерального обміну;
  - Г розлад ліпідного обміну;
  - Д розлад обміну нуклеїнових кислот.
7. Альбіноси, перебуваючи на сонці, часто отримують опіки. Порушення метаболізму якої амінокислоти лежить в основі цього явища?
- А глютамінової кислоти;
  - Б гістидину;
  - В тирозину;
  - Г метіоніну;
  - Д триптофану.
8. У шестимісячної дитини уповільнений моторний і психічний розвиток, блідість шкірного покриву, волосся і райдужної оболонки очей, позитивна проба із 5% розчином трихлоруксусного заліза. Яке із указаних спадкових захворювань виявлено у дитини?
- А галактоземія;
  - Б алькаптонурія;
  - В хвороба Дауна;
  - Г альбінізм;
  - Д фенілкетонурія.
9. Які хвороби виникають у людини при порушенні обміну ліпідів?

- А галактоземія;
- Б хвороба Німанні-Піка;
- В хвороба Тея-Сакса;
- Г хвороба Жильбера;
- Д хвороба Вільсона-Коновалова.

10. В результаті порушення якого обміну у людини розвивається подагра?

- А вуглеводного;
- Б мінерального;
- В ліпідного;
- Г пуринового і піримідинового.

11. Порушення обміну вуглеводів є причиною виникнення:

- А цукрового діабету;
- Б подагри;
- В альбінізму;
- Г гепатоцеребральної дегенерації;
- Д фенілкетонурії.

12. Алькаптонурія – це захворювання яке належить до:

- А патології обміну тирозину, зумовлене нестачею ферменту гомогентизинази;
- Б дефіциту у печінці і шкірі ферменту гістидази;
- В відсутності ферменту цистатіонінсинтетази;
- Г дефіциту ферменту фенілаланінгідроксилази;
- Д патології обміну цистину.

13. Хвороба “кленового сиропу” розвивається за відсутності або дефіциту:

- А коферменту тіамініпрофосфату;
- Б галактокінази;
- В цистину;
- Г ферменту гістидази;
- Д незамінної амінокислоти фенілаланіну.

14. Хвороба Помпе викликана:

- А відсутністю лізосомної альфа-глюкозидази;
- Б відсутністю гамма-амілази;
- В зниженою активністю фруктозо-1-фосфатальдолази;
- Г відсутністю активності глюкозо-6-фосфатази;
- Д відсутністю амільо-1,6-глюкозидази.

15. Хвороба Німана-Піка зумовлена порушенням обміну:

- А білків;
- Б ліпідів;
- В вуглеводів;
- Г амінокислот;
- Д вітамінів.

16. Хвороба Тея-Сакса (амавротична ідіотія) належить до групи хвороб:

- А обміну білків;
- Б спадкових хвороб нуклеїнових кислот;
- В спадкових хвороб обміну гормонів;
- Г спадкових хвороб обміну ліпідів;
- Д порушення метаболізму амінокислот.

17. До спадкових хвороб обміну мінеральних речовин належить:

- А синдром Леша-Нааяна;
- Б хвороба Вільсона-Коновалова;
- В хвороба Тея-Сакса; 4) фруктоземія;
- Г хвороба Німана-Піка.

18. Генотерапія пов'язана з:

- А зміною спадкового апарату людини;
- Б заміною органів;
- В заміною тканин;
- Г заміною хромосоми;
- Д заміною дефектних генів нормальними.

## ПРАКТИЧНА РОБОТА №9

### Інформаційно-аналітичні методи в молекулярній біології.

В сучасні медицині та біології генетичні методики застосовують для вирішення широкого кола завдань. Інтерпретація первинних даних з машинного секвенування починається з автоматизованої обробки даних. Результати генетичного аналізу та секвенування підлягають подальшому «розшифруванню» генетичної інформації з метою візуалізації, а саме – вирівнювання та картування зчитаних послідовностей нуклеотидів.

Перше секвенування геному людини тривало 13 років і коштувало близько 2,7 млрд. доларів, але завдяки новим методам вартість аналізу ДНК стрімко зменшується, і по всьому світу створюють усе більше масштабних «біобанків» генетичних профілів. Мета таких проєктів — не лише дізнатись більше про хвороби у загальній популяції, але й створити персоналізоване лікування на основі конкретних генетичних особливостей людей.

Аналіз молекулярних даних є одним з підходів до теоретичного вивчення структури і функції генетичних макромолекул (РНК, ДНК, білків) і їх еволюційного перетворення в популяціях живих організмів. Основна мета **філогенетичного аналізу** – вивчення еволюційного порядку дивергенції послідовностей генів і білків або їх частин, а також відновлення списків еволюційних подій (замін нуклеотидів, делецій і вставок) в предкових лініях цих макромолекул (Рис. 9.1).

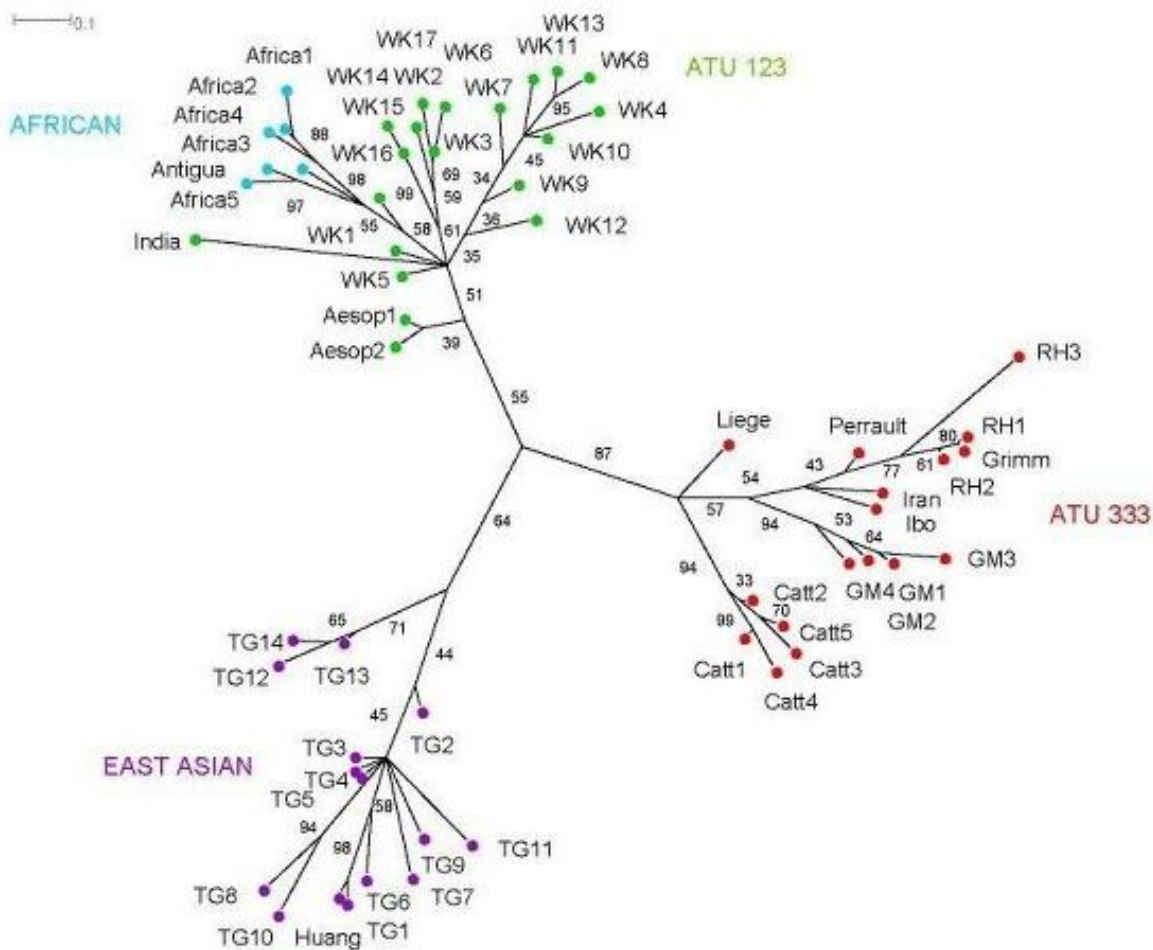


Рис. 9.1. Приклад філогенетичного дерева.



Основним інструментом філогенетичного аналізу є порівняння близьких за структурою або по функції генів або білків, і перш за все, порівняння їх первинних послідовностей. Найважливішою властивістю функціонально значущих структур макромолекул є їх еволюційний консерватизм. Чим менше функціональна важливість окремих ділянок генів, тим більше вони мають тенденцію до еволюційної мінливості.

### **Біоінформатика**

Велика кількість послідовностей (у тому числі повних послідовностей геномів), які вже встановлені та продовжують накопичуватись, створює проблему їхнього збереження та аналізу. Зрозуміло, що таке питання, яке постає вже при порівнянні послідовностей окремих клонованих фрагментів з метою з'ясування нуклеотидної послідовності геному, може бути вирішеним лише за допомогою комп'ютерів. Біоінформатика – галузь, пов'язана з розв'язанням подібних проблем, є сьогодні невід'ємною складовою частиною молекулярної біології.

Найбільші загальнодоступні через Інтернет бази даних нуклеотидних і амінокислотних послідовностей створено в Європейській Лабораторії Молекулярної Біології (EMBL Sequence Data Base) і Національних Інститутах Здоров'я США (Gen Bank). На відповідних порталах розміщено також програмні засоби порівняння послідовностей, пошуку промоторів, стартових точок транскрипції, інтронів та екзонів, визначення білкових послідовностей із послідовностей нуклеотидів тощо. Серед баз даних слід згадати також бази даних просторових структур білків (Protein Data Bank) і нуклеїнових кислот (Nucleic Acid Database).

На основі біоінформатичного аналізу можна встановити, наприклад, функціональне значення невідомого щойно клонованого гена за його гомологією з відомим геном іншого організму, структуру гена (знайти промотор, екзони, точки термінації транскрипції), структурно-функціональні особливості нових білків, функціональні зв'язки між різними білками та генами, філогенетичні стосунки між різними таксонами. Завдяки розвитку біоінформатики існує можливість (у доповнення до результатів, що отримують *in vivo* та *in vitro*) розв'язувати різноманітні проблеми за допомогою комп'ютерного забезпечення – *in silico*.

- Завдання:**
1. Ознайомитися з роботою програмних продуктів по розшифровці результатів секвенування.
  2. Провести вирівнювання (alignment) для ряду зразків, отриманих з генетичних банків даних.
  3. Провести порівняльний аналіз для декількох видів/особин послідовностей ДНК окремих генів.

**Питання для самопідготовки:** Аналіз результатів секвенування. Філогенетичний аналіз. Біоінформатика. Галузі застосування результатів генетичного аналізу. Перспективи дослідження геномів окремих популяцій та індивідів.

## ПРАКТИЧНА РОБОТА №10

### Генетичні банки даних нуклеїнових кислот та білків. GenBank та Swiss-Prot.

Основою успіху біоінформатики є не тільки великий інтерес до неї науковців, медиків та промисловців, але й використання нею нетрадиційних для математики, фізики або хімії підходів. По-перше, це створення і широке використання загальнодоступних молекулярно-біологічних баз даних (БД), які є одним з основних її інструментів. По-друге, це надання певним емпіричним правилам, що одержані шляхом статистичного аналізу інформації з відповідних БД, статусу «законів», наприклад, «золоті правила», які дозволяють робити висновки про фізико-хімічні та функціональні характеристики складних молекул на основі їх подібності до інших молекул з уже відомими властивостями.

Призначення та класифікація баз даних З 1982 р. по молекулярній біології, зокрема, нуклеотидним та амінокислотним послідовностям, створено декілька тисяч БД. Серед них є електронні банки даних як загального призначення, так і вузькоспеціалізовані, а також БД наукових робіт з біології та медицини (Medline, PubMed та інші).

Взагалі, база даних – це сукупність пов'язаної інформації, що об'єднана за певними ознаками. Більшість БД для збереження даних використовують таблиці. Кожна таблиця складається з рядків та стовпчиків, які називаються записами та полями, відповідно. Один запис може містити багато однакових полів, в які заноситься різна інформація.

Біоінформаційні БД забезпечують зручне та ефективно зберігання великої кількості інформації, систематизацію амінокислотних і нуклеотидних послідовностей, спрямовану на їх порівняльний аналіз, зокрема для:

- трансліювання амінокислотних послідовностей білків;
- ідентифікації організмів, їх таксономічної приналежності і рівнів еволюційного розвитку, побудови філогенетичних дерев;
- задач генної інженерії; • виявлення у неперервній послідовності символів окремих структурних одиниць та визначення їх функціонального навантаження;
- розшифровки просторової структури білків;
- виявлення структурно-функціональних взаємозв'язків груп білків;
- виявлення генів, які кодують макромолекули – потенційні мішені дії нових ліків та їх синтез (drug-designing).

Основним призначенням БД, крім збереження інформації, є швидкий пошук та цілеспрямоване структурування останньої. Електронні БД також повинні забезпечувати користувача засобами для управління всіма їх даними та інструментами для аналізу відповідної інформації. Існує багато різних типів баз даних, які відрізняються за джерелом надходження інформації (архівні БД, БД, що куруються, похідні БД та інші) або за тематикою (нуклеотидні послідовності або цілі геноми, амінокислотні послідовності або просторова структура білків, конкретні організми або наукова література та інше). Архівні, або первинні БД містять необроблені дані у тій формі, у якій вони були отримані із джерела. Існують також і вторинні БД (БД, що куруються), які містять відібрану після аналізу з архівних БД інформацію.

Так БД GenBank є найбільшою архівною БД нуклеотидних послідовностей, а БД Swissprot, що курується, є найбільш достовірною БД по білках. Ще в один тип БД можна виділити інтегровані БД, у яких зібрана уся інформація з певної тематики як з БД, що

куруються, так і з БД, які не куруються. Якщо у таку БД ввести назву гену можна знайти усе, що про нього відомо – у яких організмах він зустрічається, у якому місці він локалізований, які функції виконує, яку просторову структуру він має і таке 60 інше. Такою БД є NCBI Entrez – доступ до інформації про нуклеотидні і амінокислотні послідовності і їх структурах.

За тематикою БД, як уже зазначалося, часто поділяють на два типи, а саме на бази даних загального та спеціального призначення. БД ДНК, білків, вуглеводів та інші є базами загального призначення. У якості прикладів спеціалізованих БД можна навести БД поліморфізмів окремих нуклеотидів, послідовностей, що характеризують геном, резус даних про імуногенні білки і таке інше.

БД загального призначення, у свою чергу, часто поділяють на бази даних послідовностей і бази даних структур. Бази даних структур містять записи окремих послідовностей – нуклеотидів, амінокислот, білків. Бази даних структур містять записи окремих послідовностей, структури макромолекул яких визначені біохімічними методами (наприклад, БД Protein 3D structure). Записи БД містять нові експериментальні результати і додаткові відомості у формі анотацій. Анотації дають інформацію про джерела даних і методи отримання цих даних.

Також надається інформація про дослідників і перелік публікацій по даному питанню. Не менш важливим є те, що записи забезпечують посилання на відповідні записи інших БД. Кожна база даних має свій формат представлення інформації, крім того, деякі з них виникли досить давно та зараз рідко використовуються. Для переводу з формату в формат існують різні програми-конвертори. Крім форматів EMBL, GeneBank, Swiss-Prot та інших БД, орієнтованих на сприйняття інформації людиною, необхідно знати один з найбільш поширених комп'ютерно-орієнтованих форматів – FASTA-формат, який слугує для представлення нуклеотидних та амінокислотних послідовностей у вигляді, зручному для обробки на обчислювальних машинах. Важливо також вміти переводити в нього дані з більш високорівневих форматів.

### **Методики пошуку інформації у базах даних**

Усі існуючі бази даних надають можливість роботи з ними через Internet та практично усі вони використовують стандартні методики пошуку, наприклад, можливість роботи з пошуковими системами: Entrez (пошук по назві, номеру, організму, автору і т. ін.). Забезпечує доступ до амінокислотних і нуклеотидних послідовностей, їх тривимірних структур, а також до повних секвенованих геномів, надає графічне відображення генів.

Практично для кожної послідовності можна підібрати подібні послідовності та вже розраховані і визначені дво- та тривимірні структури, що відносяться до даної послідовності. BLAST (basic local alignment search tool, пошук за подібністю) – порівнює надану інформацію з послідовностями, що вже є в базі для пошуку подібних послідовностей. Є різні модифікації програми BLAST: BLASTp (вирівнювання амінокислотних послідовностей), BLASTn (вирівнювання нуклеотидних послідовностей), BLASTx (вирівнювання всіх можливих транслятів нашої нуклеотидної послідовності проти банка амінокислотних послідовностей), TBLASTx (вирівнювання всіх можливих транслятів нашої нуклеотидної послідовності проти всіх транслятів банка нуклеотидних послідовностей). Якщо при пошуку за допомогою BLAST було віднайдено декілька подібних до вихідної послідовностей (для кожної з яких побудовано тільки парне вирівнювання з досліджуваною послідовністю), то виникає задача написати всі ці послідовності одна під одною, щоб визначити, в якій мірі вони співпадають, що в них консервативно (стійко) повторюється, а що ні. Ця задача називається множинним

вирівнюванням. Offline-інтерфейс – спочатку з мережі Інтернет на локальний комп'ютер скачується частина бази даних, потім з цією частиною проводиться подальша робота.

Режим клієнт-сервер – на локальному комп'ютері встановлюється програма математичної обробки нуклеотидних послідовностей або послідовностей амінокислот, далі дана програма з'єднується з сервером бази даних і обробляє інформацію без скачування останньої на локальний комп'ютер. В той же час інтенсивно розвиваються системи обробки інформації та пошукові системи, що збирають і обробляють інформацію відповідно до запитів користувачів.

Програмне забезпечення баз даних повинно задовільняти наступним функціональним вимогам:

– Об'єм баз даних повинний бути практично не обмеженим (тобто обмежений лише параметрами апаратних засобів).

– БД повинна бути достатньо гнучкою для забезпечення проходження процесу перебудови по мірі її заповнення, так як попереднє проектування детальної структури бази даних є неможливим.

– БД повинні бути інтегровані з іншими БД та підтримувати не лише стандартні мультимедійні формати, але й ряд спеціальних гіпермедіасередовищ (просторові структури молекул, хімічні структурні формули та ін.).

– Експлуатація та поповнення баз даних через комп'ютерні мережі має бути легко доступним та зрозумілим для користувачів, які не мають комп'ютерної підготовки (біологи, медики).

### **Характеристики біоінформаційних ресурсів**

Однією з найвідоміших міжнародних організацій з тих, що створюють інструменти для аналізу інформації та здійснюють нагляд за наповненням баз даних біоінформаційного напрямку, є Міжнародна система баз даних нуклеотидних послідовностей (International Nucleotide Sequence Database Collaboration (INSDC) – <http://www.ncbi.nlm.nih.gov/collab>), яка об'єднує три установи, нуклеотидні БД яких мають різний набір сервісів:

NCBI – National Center for Biotechnology Information, USA (БД GenBank містить більше  $8 \times 10^7$  записів): <http://www.ncbi.nlm.nih.gov/>

EBI – European Bioinformatics Institute, United Kingdom (БД EMBL містить біля  $8 \times 10^7$  нуклеотидних послідовностей)

TrEMBL – більше  $5 \times 10^6$  автоматично трансльованих з даних EMBL амінокислотних послідовностей): <http://www.ebi.ac.uk/>

NIG – National Institute of Genetics, Japan (БД DDBJ містить біля  $8 \times 10^7$  записів): <http://www.ddbj.nig.ac.jp/>

До найбільш відомих організацій також відносяться:

RCSB – Research Collaboratory for Structural Bioinformatics

USA (БД PDB – Protein data bank – біля  $5 \times 10^4$  просторових структур білків), <http://www.rcsb.org/>

SIB – Swiss Institute of Bioinformatics, Switzerland (БД Swissprot містить більше  $3 \times 10^5$  записів білків): <http://www.isb-sib.ch/>

Розглянемо більш докладно деякі з БД:

– БД нуклеотидних послідовностей: EMBL (European Molecular Biology Laboratory, Geyselberg, Germany: <https://www.embl.de/>). Заснована у 1982 р. БД нуклеотидних послідовностей. Поповнюється безпосередньо авторами, що визначили первинну структуру фрагмента ДНК чи РНК. В форматі БД EMBL представлена інформація про нуклеотидну

послідовність, її функціональну розмітку, посилання на відповідні експериментальні дані та статті. Картка EMBL представляє з себе текстовий файл з обмеженою довжиною строки. Ключові слова, що визначають характер представленої в різних областях картки інформації є жорстко заданими. Для картки EMBL ключові слова створюють двосимвольні послідовності на початку строки та служать для розширення можливостей двовимірної БД.

GenBank (National Centre for Biotechnology Information, Los-Alamos, USA: <http://www.ncbi.nlm.nih.gov/>). Заснована у 1982 р. БД генетичних послідовностей, містить анотовану колекцію всіх загальнодоступних послідовностей ДНК, РНК та білків разом з літературними посиланнями. Поновлюється раз на 2 місяці (Рис. 1).

Center for Information Biology (CIB, DNA Data Bank of Japan: <http://www.ddbj.nig.ac.jp/>). Заснована у 1984 р. БД генетичних послідовностей. Збір інформації проводиться в першу чергу у японських вчених та з літератури. 75% зібраних послідовностей являють собою частково секвеновані фрагменти ДНК з декількох сотень експресованих генів, так званих EST (Expressed Sequence Tags).

Department of Medical Biochemistry of the University, Switzerland, Geneva: <http://www.expasy.ch/sprot/sprot-top.html>). База даних, що містить анотовані амінокислотні послідовності, трансльовані з нуклеотидних послідовностей EMBL; адаптовані послідовності з PIR; а також послідовності опубліковані в літературі і надіслані безпосередньо авторами. Містить високоякісні анотації без збиткової інформації, посилання на споріднені бази даних (EMBL, Prosite, PDB). Кожна анотація містить опис функції білка, його доменної структури, особливостей пост-трансляційної модифікації. Оновлюється щотижня. Для академічних користувачів є безкоштовною.

PIR (Protein Information Resource, National Biomedical Research Foundation, USA: <http://www-nbrf.georgetown.edu/pirwww/dbinfo/pirpsd.html>). База даних, що містить інформацію щодо білків для яких відомі нуклеотидні послідовності. Пошук організовано як по таксономії так і гомології. Має низький рівень зайвої інформації. Поновлюється щотижня.

MMDB (Molecular Modelling Database, Georgetown University, USA: <http://www.ncbi.nlm.nih.gov/Structure/>). База, яка містить просторові структури білків, визначені дослідним шляхом (рентгеноструктурною кристало-графією та ЯМР-спектроскопією), надає інформацію про біологічну функцію та 67 механізми, що з нею пов'язані; еволюційну історію і взаємозв'язок між макромолекулами. Входить до складу PDB, що містить також теоретичні моделі. Всі структури цієї бази даних мають первинні структури в NCBI. Оновлюється щоденно.

ENZYME (<http://www.expasy.ch/enzyme>). База даних, що містить інформацію щодо номенклатури ферментів, і описує всі типи білків, яким присвоєно номер ЕС (Enzyme Commission). Пошук реалізовано по ЕС-номеру, класам ферментів, хімічним компонентам, по кофакторам, по назвам хвороб, пов'язаних з ферментом.

Таким чином, окрім високого професійного рівня в області біології, ключовим моментом в оволодінні всіма можливостями сучасних біоінформаційних технологій є наявність у фахівця навичок швидкого та ефективного пошуку інформації в спеціалізованих БД через мережу Інтернет. Оволодіння навичками користування інтернет-ресурсами молекулярної біології відкриває широкі можливості у використанні обчислювальної молекулярної біології не лише для пошуку і аналізу вже існуючої інформації, але й для отримання нових знань з меншими затратами матеріальних та часових ресурсів порівняно з фізико-хімічними дослідженнями.

- Завдання:**
1. Ознайомитися із генетичними базами та банками даних.
  2. Навчитися проводити пошук нуклеотидних послідовностей по базах для окремих видів організмів.
  3. Навчитися проводити пошук нуклеотидних послідовностей для різних генів людини.

**Питання для самопідготовки:** Генетичні бази даних та їх типи. Методи пошуку інформації в базах даних. Значення та перспективи використання біологічних та медичних інформаційних ресурсів.