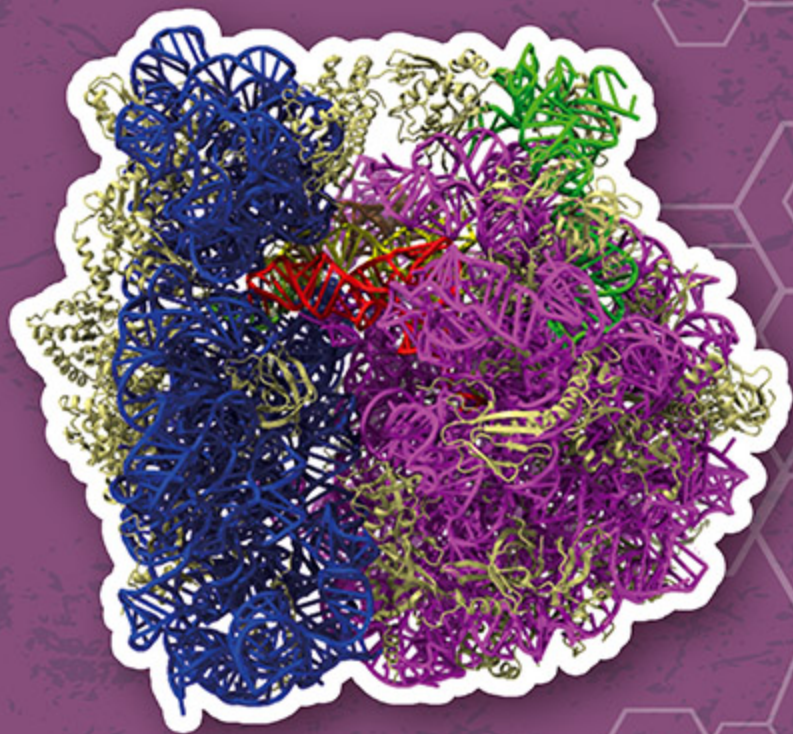


Андрій СИВОЛОБ

# МОЛЕКУЛЯРНА БІОЛОГІЯ



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

**Андрій Сиволоб**

# **МОЛЕКУЛЯРНА БІОЛОГІЯ**

**Підручник**

*Друге видання, перероблене та доповнене*

УДК 577.2(075.8)  
С34

**Рецензенти:**

д-р біол. наук, проф., акад. НАН України Г. В. Єльська,  
д-р біол. наук, проф., чл.-кор. НАН України О. І. Корнелюк  
(Інститут молекулярної біології і генетики НАН України);  
д-р біол. наук, проф., акад. НАН України С. О. Костерін  
(Інститут біохімії імені О. В. Палладіна НАН України)

*Рекомендовано до друку вченою радою ННЦ "Інститут біології та медицини"  
(протокол № 9 від 18 квітня 2022 року)*

*Ухвалено науково-методичною радою  
Київського національного університету імені Тараса Шевченка  
(протокол № 9-22 від 1 грудня 2022 року)*

**Сиволоб Андрій**

С34 Молекулярна біологія : підручник / Андрій Сиволоб. – 2-ге вид., перероб. і доп. – К. : ВПЦ "Київський університет", 2023. – 511 с. : кольор. іл.

**ISBN**

Оновлений підручник, у якому подано сучасні уявлення про молекулярні механізми загальних біологічних процесів. Викладено принципи структурної організації білків і нуклеїнових кислот, а також взаємодій біологічних макромолекул. Описано молекулярні механізми транскрипції, процесингу РНК, білкового синтезу, реплікації, репарації і рекомбінації ДНК, а також принципи основних методів молекулярної біології. Найбільшу увагу приділено структурним і динамічним властивостям макромолекул і макромолекулярних комплексів, які забезпечують здійснення біологічних функцій.

Для студентів біологічних факультетів університетів, медичних і аграрних закладів вищої освіти, а також аспірантів і науковців.

**УДК 577.2(075.8)**

ISBN

© Сиволоб А. В., 2008  
© Київський національний університет імені Тараса Шевченка,  
ВПЦ "Київський університет", 2008  
© Сиволоб Андрій, 2023, зі змінами  
© Київський національний університет імені Тараса Шевченка,  
ВПЦ "Київський університет", 2023

## ВІД АВТОРА

Молекулярну біологію можна визначити як розділ біологічних наук, присвячений вивченню молекулярних механізмів біологічних процесів. Цілком очевидно, що сьогодні майже вся біологія є молекулярною або принаймні прагне бути такою. У пропонованому підручнику основну увагу зосереджено на молекулярних механізмах лише найзагальніших біологічних процесів – передавання і реалізації спадкової інформації. Проте, оскільки загальні принципи функціонування біологічних макромолекул є універсальними, то підручник можна розглядати як базовий для подальшого вивчення також інших біологічних явищ на молекулярному рівні.

Перше видання підручника, присвяченого ключовим принципам молекулярної біології, вийшло у 2008 році. Друге видання, принципово не відрізняючись за структурою від першого, має на меті суттєво оновити зміст підручника відповідно до результатів, що з'явилися за цей час у сфері, яка бурхливо розвивається.

Сподіваюся, що ця книга, в основу якої покладено курс молекулярної біології, що автор викладає у ННЦ "Інститут біології та медицини" Київського національного університету імені Тараса Шевченка, буде корисною не тільки для біологів, а й для медиків, аграріїв, фізиків і хіміків – усіх, хто цікавиться молекулярною біологією.

Щиро дякую за обговорення підручника й корисні зауваження рецензентам книги академікам Ганні Єльській і Сергію Костеріну та члену-кореспонденту НАН України, професору Олександрю Корнелюку. Особлива подяка моїй дружині Катерині Афанасьєвій.

*Андрій Сиволоб*

## ВСТУП

Спільними за головними рисами для всіх живих систем процесами, що лежать в основі функціонування організмів, є процеси збереження та реалізації спадкової інформації.

Магістральний шлях передавання інформації в біологічних системах відображає схема, запропонована свого часу Френсісом Кріком під назвою *центральна догма молекулярної біології*:



Формулювання центральної догми можна вважати початком молекулярної біології як окремої науки. Саме за такою схемою, без порушення певної традиції, побудовано цю книгу. Слід зауважити, що, крім відображених на схемі, у живих системах реалізуються також шляхи передавання інформації від РНК на РНК і на ДНК.

Вузловими точками схеми – джерелом і об'єктами передавання інформації – є біологічні *макромолекули* двох основних типів: *нуклеїнові кислоти* та *білки*. Префікс "макро-" відображає той простий факт, що молекула має великий розмір, нехарактерний для "звичайних" молекул неорганічного світу. Великий розмір біологічних макромолекул є принциповою відмінністю живих систем: тільки молекула, яка складається з великої кількості атомів, здатна мінімізувати перешкоджаючий вплив випадкових теплових флуктуацій і здійснювати більш-менш точні операції на основі численних взаємодій з іншими молекулами. У першому розділі книги розглянуто деякі важливі фізико-хімічні принципи, які є основою молекулярної біології.

Макромолекули – це полімерні ланцюги, що складаються з мономерних хімічних груп певного класу. Кінцевою точкою передання біологічної інформації є білки (або протеїни) – біополімери, побудовані переважно із 20 типів *амінокислот*. Велике розмаїття амінокислотних послідовностей створює можливості для реалізації різноманітних шляхів укладання ланцюгів у просторі – утворення специфічних просторових структур білків для виконання певних специфічних завдань. Головним із цих завдань є каталіз численних біохімічних реакцій (зокрема й тих, що забезпечують передавання біологічної інформації). Здійснення реакцій, які мають відбуватися у вузькому діапазоні фізіологічних умов, було б практично неможливим без такого каталізу. Крім того, білки є основним будівним матеріалом будь-якої біологічної системи й забезпечують виконання всіх інших біологічних функцій: транспортування речовин, передавання регуляторних сигналів, спрямованих рухів частин системи, захисту від чужорідних молекул тощо. Принципи структурної організації білків та їхнього функціонування розглянуто в розділі 2.

Мономерами інших важливих біополімерів – нуклеїнових кислот – є *нуклеотиди*. Нуклеотиди чотирьох типів утворюють полінуклеотидний ланцюг, а два такі ланцюги, утримуючись разом за рахунок певних взаємодій, формують дволанцюгову спіральну молекулу *дезоксирибонуклеїнової кислоти* (ДНК). Нуклеотиди, які трохи відрізняються за хімічною будовою, утворюють (зазвичай одноланцюгову) полімерну молекулу *рибонуклеїнової кислоти* (РНК). Молекула ДНК відрізняється високою стабільністю, що робить її надійним сховищем спадкової інформації. Сама інформація записана у вигляді послідовності нуклеотидів чотирьох типів, а структура дволанцюгової молекули забезпечує функціонування ДНК як носія інформації. Ключовим у цьому сенсі є *принцип комплементарності*, згідно з яким чотири нуклеотиди здатні утворити дві комплементарні пари і ДНК є стабільною тільки за умови, що у двох ланцюгах один проти одного розташовані нуклеотиди однієї такої пари. Отже, кожен із ланцюгів є своєрідним дзеркальним відображенням іншого. Структурну організацію ДНК і принципи її взаємодії з білками, що забезпечують укладання ДНК у клітині й регуляцію її актив-

ності, описано в розділі 3. Особливості різних типів РНК, які виконують різноманітні важливі допоміжні функції у процесах реалізації спадкової інформації, розглянуто окремо в розділах, присвячених механізмам відповідних процесів.

Спадкова інформація, записана в послідовності нуклеотидів на ділянці ДНК, є інформацією про послідовність амінокислот у складі білка. Окрема змістовна ділянка ДНК, у послідовності якої закодована амінокислотна послідовність білка, називається *геном*, а сукупна нуклеотидна послідовність ДНК у клітинах певного організму (яка включає не тільки гени) – *геномом*. У розділі 4 викладено загальні принципи організації геномів, а також упаковки ДНК у клітинах – у складі ДНК-білкового комплексу, який називають *хроматином*. У цьому самому розділі описано систему співвідношень між елементами послідовності нуклеотидів і амінокислотними послідовностями білків – *генетичний код*, який використовується живими системами для перекладання нуклеотидного тексту на послідовність амінокислот у складі білка.

Наведене визначення гена насправді не зовсім точне. Крім білкових генів, геном кожного організму містить також гени, кінцевими продуктами яких є різноманітні молекули РНК. Однак, незалежно від типу гена, первинним продуктом його активності (проміжним для білкових генів) є молекула РНК. Під час *транскрипції* нуклеотидна послідовність одного з ланцюгів ДНК за принципом комплементарності переписується в нуклеотидну послідовність РНК – ДНК використовується як матриця, на якій будується комплементарна РНК-репліка. Молекула РНК, що синтезується на білковому гені, стає далі матрицею для білкового синтезу (матрична або інформаційна РНК – мРНК). Молекулярні механізми транскрипції та її регуляції у бактерій і еукаріотів описано в розділах 5, 6. Тісно пов'язаний із транскрипцією *процесинг* еукаріотичних мРНК (остаточне "дозрівання" функціонально активної молекули) розглянуто в розділі 7.

Розділ 8 присвячено молекулярним механізмам кінцевого етапу експресії білкових генів – синтезу білків. Матрицею для білкового синтезу є молекула мРНК, нуклеотидна послідовність якої за правилами генетичного коду перекладається на амінокислотну послідовність білка під час *трансляції*. При цьому знов

використовується принцип комплементарності: елемент послідовності мРНК упізнається комплементарною ділянкою у складі особливої молекули РНК – транспортної РНК (тРНК), яка несе на собі амінокислоту. Специфічне зв'язування певної тРНК із відповідною амінокислотою є важливою передумовою трансляції. Процес сканування матриці та синтезу амінокислотного ланцюга забезпечується *рибосомою* – мультимолекулярним комплексом, основою якого є рибосомні РНК (рРНК), зв'язані з рибосомними білками. Після закінчення синтезу на рибосомі амінокислотного ланцюга (або іноді під час синтезу) відбувається його просторове укладання у специфічну білкову структуру. Закономірності цього процесу також викладено в розділі 8.

Суттєвим моментом функціонування живої системи є не тільки реалізація (експресія) генетичної інформації, а й її збереження і подвоєння з метою передання наступному поколінню. Подвоєння інформації – *реплікація* – це відтворення молекули ДНК у двох ідентичних дочірніх копіях. Головним механізмом реплікації є вже згадуваний кілька разів принцип комплементарності: кожен із ланцюгів ДНК використовується як матриця для синтезу комплементарної ДНК-репліки. Спеціальні молекулярні системи забезпечують високу точність синтезу ДНК під час реплікації. Виправлення помилок, які все ж таки виникають у ході цього процесу, а також пошкоджень у молекулі ДНК, що з'являються під дією зовнішніх впливів, називається *репарацією* ДНК. Молекулярні механізми реплікації і репарації описано в розділі 9.

Розділ 10 присвячено молекулярним процесам, що зумовлюють надзвичайно біологічно важливе "перетасування" генетичного матеріалу, – *рекомбінації* ДНК.

В останньому, 11 розділі подано огляд деяких найважливіших методів дослідження у молекулярній біології.

З метою полегшення сприйняття інформації в тексті відсутні посилання. Перелік рекомендованої літератури наведено наприкінці кожного розділу; нижче представлено літературу, яку можна вважати загальною для цього підручника.

Зображення структур макромолекул, якими проілюстровано книгу, створено на основі атомних координат із Банку даних біл-

кових структур (PDB – Protein Data Bank, <http://www.rcsb.org/pdb/>, відповідні коди PDB наведено в підписах до рисунків) за допомогою програмних пакетів PyMOL (<http://www.pymol.org/>) та UCSF Chimera (<http://www.cgl.ucsf.edu/chimera>).

## Рекомендована література

Allison, L.A. (2021) *Fundamental molecular biology*. 3d ed. Oxford UK : Wiley-Blackwell.

Alberts, B., Johnson, A., Lewis, J., et al. (2015) *Molecular biology of the cell*. 6<sup>th</sup> ed. New York : Garland Science.

Karp, G., Iwasa, J. and Marshall, W. (2019) *Karp's cell and molecular biology*. 9<sup>th</sup> ed. New Jersey NY : John Wiley and Sons.

Krebs, J.E., Goldstein, E.S. and Kilpatric, S.T. (2018) *Lewin's Genes XII*. Burlington MA : Jones and Bartlett Learning.

Lodish, H., Berk, A., Kaiser, C.A., et al. (2021) *Molecular cell biology*. 9<sup>th</sup> ed. New York : W.H. Freeman and Company.

Pal, S. (2020) *Fundamentals of molecular structural biology*. London : Academic Press.

Watson, J., Baker, T., Bell, S., et al. (2013) *Molecular biology of the gene*. 7<sup>th</sup> ed. London : Pearson.

Weaver, R.F. (2011) *Molecular biology*. 5<sup>th</sup> ed. New York : McGraw-Hill Companies.

# Розділ 1

## ФІЗИКО-ХІМІЧНІ ОСНОВИ МОЛЕКУЛЯРНОЇ БІОЛОГІЇ

All science is either physics or stamp collection.  
E. Rutherford

Незважаючи на надзвичайну складність живих систем, в основі їхнього функціонування лежать ті самі закони фізики, що й для неживої природи, – і жодних інших. Сьогодні таке твердження може здаватися банальним, однак важливо пам'ятати, що саме воно є базовим для сучасної парадигми молекулярної біології. Біологічні явища і процеси залежать від функціонування молекул, що складаються з великої кількості атомів, які вступають у різноманітні взаємодії між собою – як у межах однієї такої макромолекули, так і при формуванні макромолекулярних комплексів. Цей розділ має на меті нагадати деякі найважливіші фізичні закономірності, без розуміння яких неможливо розглядати принципи структурної організації біологічних макромолекул і взаємодій між ними.

### **Енергетика взаємодії між атомами й молекулами**

#### **Вільна енергія**

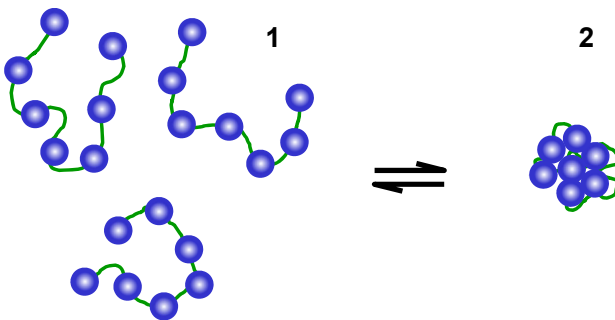
Напрямок і ймовірність перебігу будь-яких молекулярних процесів визначаються змінами *вільної енергії*. Для введення цього важливого поняття розглянемо ідеалізовану систему – спрощену модель полімерного ланцюга, зображену на рис. 1.1: мономерні елементи (кульки), з'єднані гнучкими перемичками. Нехай мономери при цьому є інертними, тобто не здатними до взаємодій між собою (немає ані взаємного притягання, ані від-

штовхування). Просторова форма (*конформація*) такого ланцюга буде визначатися лише конфігурацією перемичок. Унаслідок теплового руху ланцюг постійно змінюватиме конформацію, тобто буде реалізована *форма неупорядкованого клубка*.

На рис. 1.1 цей стан неупорядкованого клубка зображено як стан 1 ліворуч (представлено кілька з великої кількості  $W_1$  можливих конформацій ланцюга). Певний компактний стан 2 (праворуч) може бути реалізований унаслідок випадкового контакту між мономерами. Зрозуміло, що цьому стану відповідає значно менша кількість  $W_2$  конформацій перемичок. Відношення імовірності  $f_2$  знайти ланцюг у стані 2 до ймовірності  $f_1$  стану 1 (константа рівноваги  $K$  між станами 2 і 1) визначається відношенням кількостей відповідних конформацій:

$$K = \frac{f_2}{f_1} = \frac{W_2}{W_1}.$$

Оскільки  $W_2 \ll W_1$ , то рівновага зсунута ліворуч – у бік клубка. Отже, за інших однакових умов імовірність знайти будь-яку систему в певному стані визначається кількістю мікростанів (конфігурацій системи), якими цей стан реалізується. Чим більша кількість мікростанів (у нашому прикладі  $W_1$  та  $W_2$ ), тим менш упорядкованим є зазначений стан системи (зрозуміло, що ступінь упорядкованості значно зростає для стану 2).



**Рис. 1.1.** Клубок (1) і компактний стан (2) полімерного ланцюга

Нехай тепер мономери здатні взаємодіяти таким чином, що при утворенні контакту між ними в навколишнє середовище виділяється тепло (знижується внутрішня енергія нашої системи). Як відомо з термодинаміки, усі процеси за інших однакових умов відбуваються в напрямку виділення тепла. Імовірність певного стану системи визначається відношенням *ентальпії*  $H$  (величина, яка за умов постійного тиску й об'єму є еквівалентною внутрішній енергії системи) до середньої кінетичної енергії теплового руху  $k_B T$ , де  $k_B$  – константа Больцмана (Ludwig Boltzmann),  $T$  – абсолютна температура. Отже, імовірність пропорційна до величини  $\exp\left(-\frac{H}{k_B T}\right)$ : менше значення ентальпії відповідає зростанню імовірності (принцип Больцмана). Якщо зниження ентальпії при компактизації ланцюга є значним, то це може змінити конформаційну рівновагу на користь стану 2. Таким чином, у загальному вигляді тепер запишемо

$$K = \frac{f_2}{f_1} = \frac{W_2 \exp(-H_2 / k_B T)}{W_1 \exp(-H_1 / k_B T)},$$

або, позначивши  $S = k_B \ln W$  і  $G = H - TS$ ,

$$K = \frac{f_2}{f_1} = \frac{\exp\left(-\frac{H_2 - TS_2}{k_B T}\right)}{\exp\left(-\frac{H_1 - TS_1}{k_B T}\right)} = \frac{\exp\left(-\frac{G_2}{k_B T}\right)}{\exp\left(-\frac{G_1}{k_B T}\right)} = \exp\left(-\frac{\Delta G}{k_B T}\right), \quad (1.1)$$

де  $\Delta G = G_2 - G_1$ .

Останнє дуже важливе співвідношення має універсальне значення. У рівнянні (1.1) введена Больцманом величина  $S$ , що залежить від кількості мікростанів, є мірою неупорядкованості системи й називається *ентропією*. Величина

$$G = H - TS, \quad (1.2)$$

яка була введена фундатором термодинаміки Гіббсом (Josiah Willard Gibbs), називається *вільною енергією*. Саме вона в загальному випадку визначає імовірність перебування системи в

певному стані й, відповідно, напрямком, у якому відбувається той чи інший процес. Наведені нижче твердження, що базуються на рівняннях (1.1), (1.2), виконуються для будь-яких систем, зокрема (і повною мірою) молекулярно-біологічних:

- зміна стану системи відбувається в напрямку зниження вільної енергії, тобто будь-який процес, що спостерігається, не може супроводжуватись підвищенням цієї величини;
- рівноважний стан, що може бути стабільним скільки завгодно довго, відповідає мінімуму вільної енергії;
- зміна вільної енергії залежить від двох складових: зміни ентальпії (внутрішньої енергії), яка супроводжується виділенням / поглинанням тепла; зміни ентропії, що характеризує зростання впорядкованості / неупорядкованості в системі.

При цьому величина ентропійного внеску зростає з температурою (рівняння (1.2)), тобто за високих температур переважними завжди стають менш упорядковані стани.

Слід зробити також кілька інших важливих зауважень:

- вільна енергія є величиною відносною: має сенс тільки *різниця* між вільними енергіями різних станів *тієї самої* системи;
- різниця між вільними енергіями двох станів системи визначає тільки напрямок перетворення одного стану на інший і *нічого* не дозволяє сказати про те, наскільки швидко та яким шляхом буде відбуватися цей процес. Інакше кажучи, хоча стан із меншою величиною вільної енергії завжди є більш імовірним, практично він може не спостерігатись, якщо на реалізацію цього стану потрібно дуже багато часу (оскільки на шляху перетворення є високоенергетичні проміжні стани);
- енергія взаємодії між елементами системи (атомами, молекулами) не обов'язково залежить тільки від ентальпійної складової, як у простому прикладі, що його ілюструє рис. 1.1. Насправді, як буде показано нижче, вільна енергія взаємодій може мати суттєву ентропійну складову або навіть залежати тільки від ентропійних ефектів.

### **Рівновага утворення комплексів молекул у розчині**

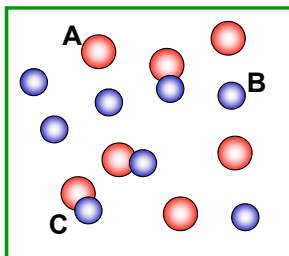
Перш ніж перейти до розгляду типів взаємодій між атомами й молекулами, розглянемо ще один приклад рівноваги у простій, але важливій для молекулярної біології системі. У розчині певного об'єму містяться молекули двох типів (А і В), які здатні утворити між собою комплекс С (рис. 1.2):



(літерою  $K$  тут знову позначено константу рівноваги, у цьому випадку – для процесу утворення / дисоціації комплексів). Ентропійна частина вільної енергії кожного компонента містить складову, яка залежить від концентрації. Наприклад, вільну енергію  $G_A$  компонента А (її в цьому випадку називають також *хімічним потенціалом*) можна записати як

$$G_A = G_A^0 + RT \ln[A], \quad (1.4)$$

де  $G_A^0$  – стандартна вільна енергія,  $R$  – універсальна газова стала, яка є константою Больцмана в перерахунку на моль речовини:  $R = k_B N_{AV}$ ,  $N_{AV}$  – число Авогадро (Amedeo Avogadro); квадратними дужками позначено молярну концентрацію А. Зміст рівняння (1.4) дуже простий:  $G_A^0$  є вільною енергією компонента А за його концентрації 1 моль/л і залежить від природи цього компонента; зростання концентрації спричинює зниження ентропії (чим більше молекул, тим менше доступних місць для розташування кожної з них, тобто тим меншою є кількість мікростанів, див. наведене вище визначення ентропії) і, відповідно, зростання вільної енергії.



**Рис. 1.2.** Взаємодія між молекулами А та В з утворенням комплексу С

Записавши аналогічні вирази для компонентів В і С, знайдемо різницю між правою і лівою частинами рівняння (1.3)  $\Delta G = G_C - G_A - G_B$ :

$$\Delta G = \Delta G^0 + RT \ln \frac{[C]}{[A][B]}, \quad (1.5)$$

де  $\Delta G^0$  – різниця між відповідними стандартними вільними енергіями,  $[A]$  і  $[B]$  – концентрації відповідних *вільних* компонентів у дисоційованому стані. Процес зміни концентрацій у системі буде тривати до моменту досягнення рівноваги – такого співвідношення між концентраціями, коли вільна енергія комплексу дорівнюватиме сумарній енергії вільних компонентів А та В, а  $\Delta G = 0$  (можна показати, що *загальна вільна енергія розчину* буде мінімальною, коли різниця  $\Delta G$  між правою і лівою частинами рівняння (1.3) дорівнюватиме нулю). Оскільки  $\Delta G^0$  – константа, яка залежить від природи компонентів, то співвідношення концентрацій під логарифмом у рівнянні (1.5) також є константним за рівноважних умов. Саме це співвідношення і є в даному прикладі константою рівноваги, яка залежить тільки від різниці стандартних вільних енергій:

$$K = \frac{[C]}{[A][B]} = \exp\left(-\frac{\Delta G^0}{RT}\right). \quad (1.6)$$

Отже, якщо при утворенні комплексу вільна енергія знижується ( $\Delta G^0 < 0$ ), то це сприяє взаємодії між компонентами. У ви-

падку, що розглядається, константу  $K$  називають *константою асоціації*, або *константою зв'язування* (зворотна величина  $K^{-1}$  – *константа дисоціації*), а величину  $\Delta G^0$  – *стандартною вільною енергією зв'язування*. Обидві константні величини є мірою *спорідненості* компонентів, що утворюють комплекс: чим негативнішою є енергія зв'язування (чим більше енергетично вигідних взаємодій реалізується між компонентами), тим вище значення  $K$  і тим стабільнішим є комплекс. Слід також зауважити, що ефективність утворення комплексу, під якою можна розуміти, наприклад, частку компонента А у складі комплексу  $[C]/([C]+[A])$ , залежить від концентрації компонентів. Рівняння (1.6) є однією з форм запису відомого *закону діючих мас*, або *принципу Ле Шательє* (Henry Le Chatelier), згідно з яким додавання в систему одного з компонентів зсуває рівновагу в протилежний бік.

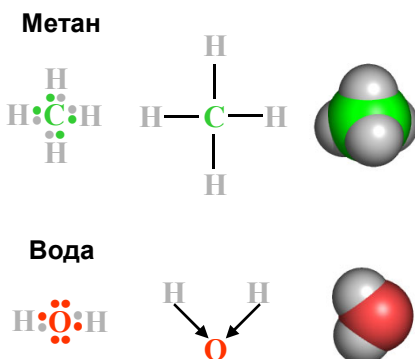
Рівняння (1.3) може описувати деяку *хімічну реакцію*, коли внаслідок перебудови *ковалентних зв'язків* (див. нижче) у молекулах А та В утворюється нова молекула С. Тоді рівняння (1.4)–(1.6) і висновки з них також можуть бути застосовані. У цьому випадку знак зміни стандартної вільної енергії вказує на напрямок, у якому відбувається хімічна реакція. Варто підкреслити ще раз: *усі* хімічні реакції в живих системах відбуваються зі зниженням вільної енергії. Тут може виникнути запитання: якщо, наприклад, вільна енергія знижується при розпаданні молекули С на А та В, але клітина має потребу саме в молекулі С, чи можливий синтез С? Відповідь є однозначною: синтез С із молекул А та В неможливий. Однак можна використати "обхідний" шлях: спочатку хімічно модифікувати одну з молекул (напр., змінити молекулу А на А'), після чого утворення С із А' і В стане можливим, якщо така реакція супроводжуватиметься зниженням вільної енергії. Повернемося до обговорення цього питання в розділі 2.

## Ковалентний зв'язок між атомами в молекулах

### Властивості ковалентного зв'язку

Ковалентний (або хімічний) зв'язок – це єдина сила, що підтримує цілісність певної системи атомів, яка називається молекулою. Основою утворення ковалентного зв'язку є узагальнення електронів двох атомів – ефект, що описується у квантовій механіці. Результатом узагальнення є формування *молекулярної орбіталі*, яка належить двом атомам одночасно й на якій може міститися не більше двох електронів (принцип заборони Паулі (Wolfgang Pauli)).

Наприклад, на зовнішньому електронному шарі атома карбону є чотири електрони, а максимально в цьому шарі їх може бути вісім. Він доповнюється електронами чотирьох атомів гідрогену (по одному електрону на кожному атомі), у результаті чого утворюється молекула метану. Аналогічно доповнення зовнішнього шару оксигену за рахунок двох атомів гідрогену зумовлює утворення молекули води (рис. 1.3). Між двома атомами, залежно від їхньої природи, можливо також формування двох (іноді й трьох) молекулярних орбіталей (по два електрони на кожній), тобто утворення подвійного (чи потрійного) ковалентного зв'язку.



**Рис. 1.3.** Структура молекул метану й води.

Показано пари електронів, які відповідають ковалентним зв'язкам. Стрілки символізують зсув електронної щільності до атома оксигену

Головні властивості ковалентного зв'язку:

- ковалентний зв'язок дуже міцний. Інакше кажучи, кожна молекула (про це свідчить сам факт її існування) перебуває в досить глибокій *потенційній ямі* – мінімумі вільної енергії;
- кожен атом містить певну кількість електронів (або "вакансій" для електронів на зовнішньому шарі), які можуть бути узагальнені з іншими атомами, тобто кожен атом може утворити лише певну кількість ковалентних зв'язків (має певну *валентність*). Наприклад, валентність карбону дорівнює 4, кисню – 2, водню – 1 (див. рис. 1.3);
- ковалентний зв'язок характеризується певною геометрією. По-перше, кожен зв'язок має конкретну довжину (звичай 1–2 Å). По-друге, електронні щільності різних орбіталей у молекулі взаємодіють між собою, у результаті чого кожен зв'язок певним чином орієнтований у просторі відносно суміжних зв'язків. Типові значення валентних кутів (кутів між зв'язками, що йдуть від одного атома) становлять близько 109 або 120°.

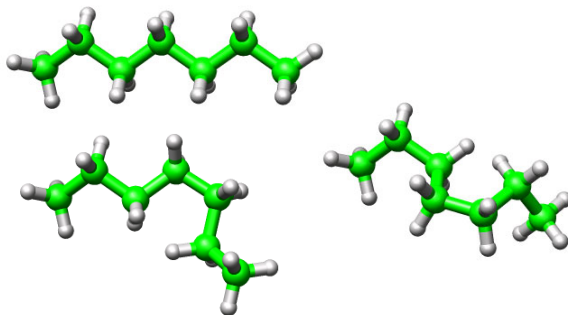
Слід також зауважити, що ступінь узагальнення електронної пари між двома атомами залежить від природи цих атомів. Наприклад, у випадку зв'язку С–Н електронна щільність розподілена приблизно рівномірно між двома атомами. Такий зв'язок називають *неполяризованим*. У випадку зв'язку О–Н, унаслідок високої *електронегативності* кисню (здатності відтягувати на себе електронну щільність), електронна пара зсунута до цього атома (*поляризований* зв'язок показано стрілочками на рис. 1.3). У результаті в молекулі води на кисні виникає парціальний негативний (приблизно  $-2/3$  від заряду електрона), а на кожному водні – позитивний ( $+1/3$ ) заряд. Іншим важливим для біополімерів прикладом електронегативного атома є азот.

## Механізм гнучкості полімерного ланцюга

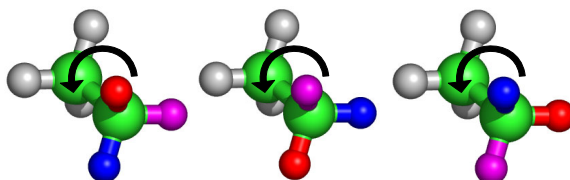
Полімерний ланцюг містить велику кількість атомів, у його складі певні мономерні одиниці з'єднані між собою ковалентними зв'язками. У прикладі, наведеному на рис. 1.1, було припущено (і це справді так для реальних полімерів), що такий ланцюг дуже гнучкий і здатен існувати у вигляді великої кількості конформацій. Однак зрозуміло, що реальний полімерний ланцюг не може бути *абсолютно* гнучким, як у моделі на рис. 1.1. У чому ж полягає механізм переходу від однієї конформації до іншої? Априорі можна уявити три можливості: 1) зміну довжини ковалентного зв'язку; 2) зміну валентних кутів; 3) обертання навкруг ковалентного зв'язку. За умов кімнатної (чи фізіологічної) температури (20–37 °С) ковалентний зв'язок є дуже жорстким щодо повздовжніх деформацій – практично такі температури не здатні збудити навіть коливання довжини. Коливання валентних кутів можливі за фізіологічної температури, але їхня амплітуда доволі невелика (до 5 °С). Отже, головну роль у гнучкості полімерного ланцюга відіграє обертання навкруг ковалентних зв'язків (рис. 1.4). Власне, конформація часто й описується в термінах кутів обертання.

Два важливі зауваження:

- обертання є можливим лише навкруг одинарного зв'язку. Подвійний зв'язок, унаслідок особливостей будови задіяних у його утворенні молекулярних орбіталей, допускає лише дві обертальні конформації: з кутами обертання 0 і 180°. Хімічні групи, присьдані до атомів, з'єднаних таким зв'язком, містяться тоді відповідно у *цис*- і *транс*-орієнтаціях;
- обертання навкруг одинарного зв'язку також не є абсолютно вільним. Взаємодія між суміжними хімічними групами зумовлює те, що певні кути обертання є переважними (вигіднішими енергетично). Інакше кажучи, практично реалізуються лише певні вузькі діапазони значень кута обертання – різні конформації є *обертальними ізомерами* (рис. 1.5).



**Рис. 1.4.** Кілька конформацій вуглеводневого ланцюжка (зелений – С, сірий – Н), отримані шляхом зміни кутів обертання навкруг зв'язків С–С



**Рис. 1.5.** Три обертальні ізомери, що є переважними для молекули етану. Атоми гідрогену при одному з карбонів забарвлені різними кольорами

### **Перебудови системи ковалентних зв'язків у хімічних реакціях**

Хімічну реакцію можна визначити як процес перебудови системи ковалентних зв'язків у межах однієї молекули чи групи молекул (субстрати реакції) з утворенням нових молекул-продуктів. Варто зауважити, що ковалентний зв'язок практично ніколи не руйнується в хімічних реакціях (це дуже не вигідно енергетично) – він просто замінюється на зв'язок з іншим атомом. Перший принцип, який уже було сформульовано в цьому розділі, полягає в тому, що реакція відбувається тоді й тільки тоді, коли вільна енергія продуктів реакції є нижчою, ніж вільна енергія субстратів (рис. 1.6).

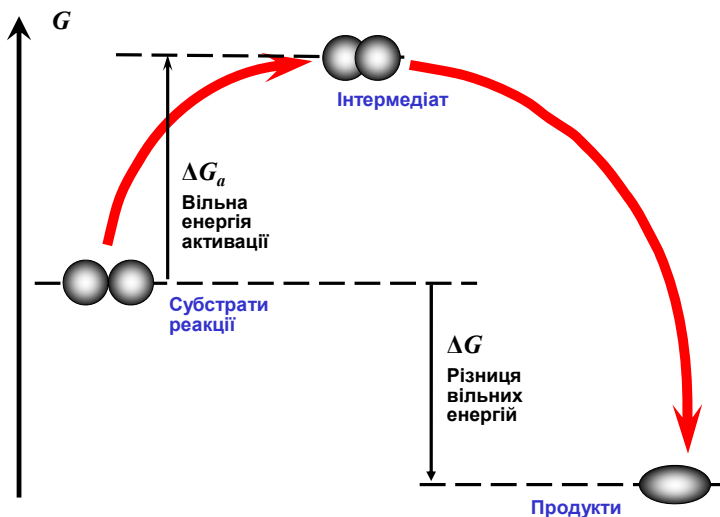


Рис. 1.6. Енергетична схема процесу хімічної реакції

Проте ця умова є лише необхідною, але не достатньою. Оскільки кожна молекула вже існує в мінімумі вільної енергії (система ковалентних зв'язків дуже стабільна), то на шляху перебування електронної щільності неодмінно має існувати перехідний стан (інтермедіат) з вищою енергією. Це означає, що електронну систему субстратів необхідно спочатку "зламати" – збудити, відхиливши її від рівноважного стабільного стану, після чого вона з більшою імовірністю "впаде" до стану продуктів, оскільки цей стан перебуває нижче на шкалі вільної енергії.

Отже, енергія інтермедіатного стану неодмінно є вищою за енергію субстратів: наявність інтермедіату створює на шляху кожної хімічної реакції енергетичний бар'єр, величина якого називається *енергією активації* (рис. 1.6). При цьому вільна енергія активації містить ентропійну складову: реакція відбувається тільки тоді, коли субстрати стикаються за певної взаємної орієнтації – лише однієї із великої кількості орієнтацій при випадкових стиканнях. Отже, малоімовірна подія є передумовою утворення інтермедіату – різниця ентропії між інтермедіатом і субстратами (ентропія активації) є негативною величиною, тобто робить позитивний внесок у вільну енергію активації.

Енергія активації, згідно із принципом Больцмана (рівняння (1.1)), визначає імовірність переходу в інтермедіатний стан, тобто швидкість реакції. Великий (порівняно з тепловою енергією  $k_B T$ ) енергетичний бар'єр означає, що реакція уповільнюється настільки, що практично не відбувається. Найпростіший шлях прискорити реакцію – підвищити температуру (збільшити енергію теплового руху). Зрозуміло, що в живих системах таке рішення неможливе. Кожна хімічна реакція, яка відбувається у клітині, потребує (для того, щоб вона могла відбуватися за розумний час) спеціального молекулярного пристрою – каталізатора, який прискорює реакцію за рахунок зниження енергії активації. Біокаталізаторами переважно (але не завжди) стають білкові ферменти (ензими), механізми роботи яких розглядатимуться в розділі 2.

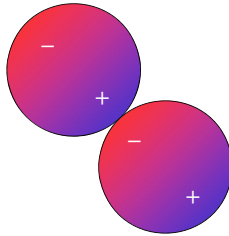
## **Нековалентні взаємодії між атомами й молекулами**

Крім ковалентних зв'язків, між атомами однієї чи різних молекул реалізуються також слабкіші, але не менш важливі, нековалентні взаємодії кількох типів, що будуть розглянуті нижче. Саме від таких взаємодій значною мірою залежить структурна організація і фізико-хімічна поведінка макромолекул, що містять велику кількість атомів.

### **Вандерваальсові взаємодії**

Навіть у межах неполяризованої молекули розподіл електронної щільності є рівномірним лише в разі його усереднення за часом. У кожний момент часу внаслідок флуктуацій щільності виникають локальні неоднорідності, тобто парціальні заряди різного знака, і молекула стає тимчасовим *диполем*. При наближенні іншої молекули такий тимчасовий диполь першої молекули індукує диполь у межах другої молекули, унаслідок чого виникає притягання електричної природи (рис. 1.7). Квантова механіка свідчить, що енергія пари молекул, електронні системи яких поляризовані, нижча за енергію пари неполяризованих молекул. У результаті всі молекули й атоми на близьких відстанях притягуються один до одного. Взаємодії такого

типу були описані Ван дер Ваальсом (Johannes Van der Waals) і отримали його ім'я.



**Рис. 1.7.** Взаємодія між тимчасовими диполями двох молекул

Вандерваальсові взаємодії між неполяризованими молекулами називають *дисперсійними*, між двома поляризованими молекулами (постійними диполями) – *орієнтаційними* (або *диполь-дипольними*), між неполяризованою молекулою і постійним диполем – *індукційними* (або взаємодіями типу *диполь-індукований диполь*).

Основні властивості вандерваальсових взаємодій:

- вандерваальсові взаємодії, які зумовлюють притягання на малих відстанях, реалізуються між *будь-якими* молекулами й атомами незалежно від їхньої природи;
- вандерваальсові взаємодії є короткодіючими: енергія взаємодій швидко знижується з відстанню  $r$  між молекулами чи атомами (для дисперсійних та індукційних взаємодій пропорційно до  $1/r^6$ , для орієнтаційних – до  $1/r^3$ ), тобто вони суттєві при безпосередньому контакті між молекулами (атомами, хімічними групами);
- слова "безпосередній контакт" означають, що існує певна оптимальна величина  $r_0$  – відстань, на якій енергія взаємодій є максимальною (тобто енергія системи – мінімальною). При зниженні  $r$  відносно  $r_0$  притягання між молекулами змінюється на відштовхування: електронні оболонки різних молекул (уже насичені валентними електронами) не можуть проникати одна в одну (відповідно до принципу заборони Паулі). Величина  $r_0$  характеризує розмір кожного атома (вандерваальсів радіус), який використовується-

ся при побудові молекулярних моделей (як, напр., на рис. 1.3). Непроникність атомів один для одного (стеричні обмеження) має велике значення для структури біополімерів: серед великої кількості конформацій можливими є лише ті, для яких не спостерігається проникнення однієї частини ланцюга в іншу (два атоми не можуть займати одну область простору);

- вандерваальсові взаємодії є найслабшими серед усіх нековалентних взаємодій. Для порівняння: величина середньої кінетичної енергії теплового руху  $RT$  при  $27\text{ }^\circ\text{C}$  становить  $0,6$  ккал/моль; типові значення енергії вандерваальсових взаємодій (на відстані  $r_0$ ) варіюють у межах  $0,1\text{--}0,5$  ккал/моль. Наприклад, такі слабкі взаємодії між молекулами метану здатні забезпечити його рідкий стан тільки за дуже низьких температур (нижче  $-161,6\text{ }^\circ\text{C}$ ).

Останнє твердження потребує уточнення стосовно макромолекул. Коли взаємодіють доволі великі поверхні, між якими реалізується велика кількість міжатомних контактів, вандерваальсові взаємодії стають суттєвим фактором стабілізації структури макромолекул і макромолекулярних комплексів. Для того, щоб таких контактів було справді багато (щоб не було порожнеч), дві поверхні мають добре відповідати одна одній (як ключ із замком), тобто бути *просторово комплементарними*. Просторова комплементарність між поверхнями макромолекул є одним із важливих механізмів специфічної взаємодії між ними.

### **Електростатичні (іонні) взаємодії**

Певні хімічні групи в біологічних макромолекулах здатні іонізуватися, тобто набувати сталого негативного чи позитивного заряду, втрачаючи чи набуваючи протон. Взаємодія між такими зарядами підпорядковується добре відомому закону Кулона (Charles de Coulomb), а саме: вільна енергія взаємодії двох зарядів  $q_1$  і  $q_2$  залежить від відстані між зарядами  $r$  як

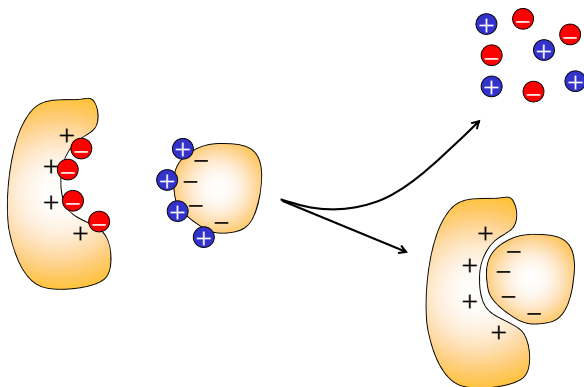
$$U = \frac{q_1 q_2}{\epsilon r},$$

де  $\epsilon$  – діелектрична проникність середовища. При цьому слід звернути увагу на кілька важливих моментів.

Величина  $\epsilon$  залежить від полярності середовища. Поляризованість молекул води зумовлює доволі велике значення проникності ( $\epsilon \approx 80$  за кімнатної температури). У неполярному середовищі проникність значно нижча (напр., усередині молекули білка  $\epsilon \approx 3$ ) і, відповідно, електростатична енергія зростає. Занурення зарядженої групи або навіть пари наближених різнойменних зарядів у неполярне середовище є дуже не вигідним енергетично: заряджені групи частіше залишаються на поверхні макромолекул, не занурюючись усередину.

Висока діелектрична проникність води свідчить про те, що електростатичні взаємодії у ній пов'язані передусім з упорядкованістю її диполів навкруг зарядів і зміною цієї впорядкованості при наближенні чи віддаленні зарядів, тобто електростатичні взаємодії мають ентропійну природу – при наближенні (взаємній нейтралізації) різнойменних зарядів ступінь упорядкованості диполів води знижується (відповідно зростає ентропія і знижується вільна енергія).

Найвагомішими електростатичні взаємодії є для макромолекул, на поверхні яких сконцентрована велика кількість заряджених груп (яскравим прикладом таких макромолекул є нуклеїнові кислоти, див. розділ 3). На перший погляд здається очевидним, що дві макромолекули з позитивно й негативно зарядженими поверхнями мають взаємодіяти за законом Кулона. Однак крім води в живих системах завжди є неорганічні іони, які, за тим самим законом Кулона, зв'язуються із зарядженими поверхнями макромолекул (рис. 1.8). Права частина рис. 1.8 енергетично краща за ліву лише звільненням неорганічних іонів у зовнішній розчин, тобто більш низькою впорядкованістю (вищою ентропією і нижчою вільною енергією). Таким чином, взаємодія двох заряджених макромолекул – цілком ентропійний ефект. Величина цього ефекту знижується зі зростанням концентрації неорганічних іонів (іонною силою): якщо іонів багато в зовнішньому розчині, то ефект від їхнього звільнення дуже незначний. Тоді взаємна спорідненість макромолекул знижується і комплекс дисоціює.



**Рис. 1.8.** Ентропійна природа електростатичних взаємодій між високочарядженими макромолекулами

### **Водневий зв'язок**

Деякі молекули й хімічні групи здатні на взаємодії особливо-го типу – водневі зв'язки. Прикладом можуть бути молекули води – саме водневі зв'язки між ними зумовлюють рідкий стан води за температур до 100 °С. Розглянемо дві молекули води (рис. 1.9). Як було зазначено вище, унаслідок високої електронегативності оксигену молекула води поляризована, тобто на її атомах є парціальні електричні заряди. Відповідно дві молекули будуть взаємодіяти одна з одною електростатично. Однак тут є ще дві важливі обставини: 1) атом гідрогену має лише один електрон, який відтягнутий на оксиген у молекулі води; 2) атом оксигену містить на зовнішньому шарі дві неподілені пари електронів (рис. 1.3, 1.9), які не залучені до утворення ковалентного зв'язку. У результаті слабка електронна оболонка гідрогену "продавлюється" атомом оксигену другої молекули; його неподілена пара електронів частково узагальнюється з гідрогеном, електрон якого відтягнутий власним оксигеном. Взаємодія такого типу називається *водневим зв'язком*, який є чимось середнім між електростатичною взаємодією (притяганням парціальних

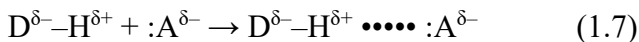
зарядів) і ковалентним зв'язком (частковим узагальненням пари електронів між двома атомами).

Кожна молекула води містить два атоми гідрогену й дві неподілені електронні пари, отже, максимальна кількість водневих зв'язків, які може утворити одна молекула води з іншими, дорівнює чотирьом. У середньому саме чотири водневі зв'язки припадають на одну молекулу в рідкій воді. Однак слід розуміти, що середній час життя одного зв'язку становить приблизно 1 пс – система водневих зв'язків у рідкій воді характеризується ультрашвидкою динамікою (саме тому спостерігається рідкий стан). Відповідно будь-які вигадки про "пам'ять" чи "структуру" рідкої води не мають жодних підстав.

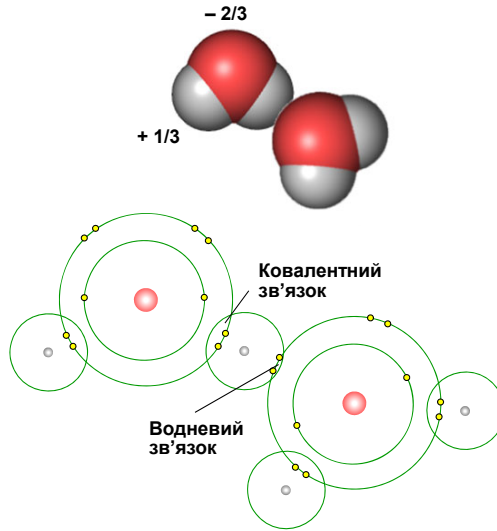
Не тільки молекули води містять хімічні групи, здатні до утворення водневих зв'язків. Наведемо типові пари хімічних груп, що можуть з'єднуватись водневими зв'язками в макромолекулах:



Отже, водневий зв'язок зазвичай формується між атомом гідрогену, приєднаним до електронегативного атома із частковим негативним зарядом  $\delta^-$  (ця група віддає протон і називається *донором* водневого зв'язку), і електронегативним атомом іншої хімічної групи чи молекули (*акцептором*), який містить неподілену пару електронів:



Зокрема, такі групи завжди розміщені на поверхні водорозчинних макромолекул. Саме із цієї причини такі молекули розчиняються у воді, взаємодіючи з нею за рахунок водневих зв'язків. Слід зазначити, що шари гідратації біля поверхні макромолекул швидко обмінюються з молекулами основного водного розчину – середній час затримки молекули води у шарі гідратації оцінюють у 10–20 пс.



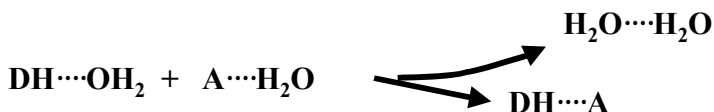
**Рис. 1.9.** Дві молекули води, зв'язані водневим зв'язком (внизу – схема електронної структури двох молекул)

Дві властивості водневого зв'язку є наслідком його природи:

- 1) водневий зв'язок є дуже вигідним енергетично: його енергію (утворення водневого зв'язку супроводжується виділенням тепла, тобто він має ентальпійну природу) можна оцінити приблизно у 5 ккал/моль (порівняйте з енергією теплового руху  $RT = 0,6$  ккал/моль);
- 2) на відміну від електростатичних взаємодій, але так само, як ковалентний зв'язок, водневий зв'язок є чутливим до взаємної орієнтації донора й акцептора – напрямки водневого й ковалентного зв'язків D–H мають збігатися (їхнє припущиме відхилення становить  $20\text{--}30^\circ$ ).

Друга властивість дуже важлива. Вона означає, що водневі зв'язки між макромолекулами утворюються тільки тоді, коли між донорними й акцепторними групами є чітка просторова відповідність. Це робить водневий зв'язок найсуттєвішим фактором у визначенні *специфічності* міжмолекулярних взаємодій – міжмолекулярному впізнанні.

Перша властивість також потребує уточнень. Водневий зв'язок, якщо він може утворитись, дуже не вигідно втрачати. Однак можна легко його замінити на водневий зв'язок із молекулою води, яка зазвичай є у великій кількості в оточенні макромолекул. Якщо взаємодія між донором і акцептором відбувається у водному середовищі, то рівняння (1.7) має бути переписане таким чином:

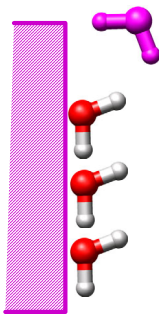


Як бачимо, кількість водневих зв'язків однакова ліворуч і праворуч (ентальпія системи не змінюється), різною може бути лише ентропія лівого й правого станів. Якщо донор і акцептор – маленькі молекули, то переважним буде лівий стан (велика кількість молекул води робить його менш упорядкованим): маленькі полярні молекули розчиняються у воді, тому що добре з нею взаємодіють. Якщо донор і акцептор належать макромолекулі, то переважним може стати правий стан: пара молекул води може вільніше рухатися, що збільшує ентропію системи. Однак у будь-якому разі внаслідок конкуренції з водою у водному оточенні водневий зв'язок між донорною й акцепторною групами на поверхні макромолекул не може бути занадто стабільним. З іншого боку, зрозуміло, що в неполярному середовищі (де немає конкуренції з водою) водневий зв'язок буде значно стабільнішим.

### **Гідрофобні взаємодії**

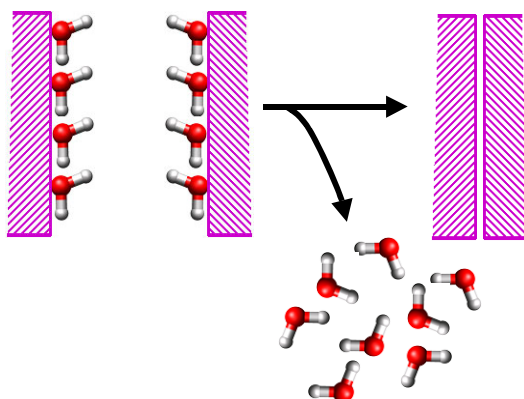
Сильна взаємодія між молекулами води за рахунок водневих зв'язків і те, що ці зв'язки можливі лише за певної взаємної орієнтації молекул, зумовлює *гідрофобний ефект* при зануренні у воду *неполярних* молекул. неполярність молекули означає, що вона не здатна утворювати водневі зв'язки. Однак молекули води не тільки здатні, а й прагнуть обов'язково їх утворити, оскільки енергія зв'язку є високою і втрачати його

дуже не вигідно. Єдиний вихід для молекул води – утворити зв'язки між собою. Проте неполярна поверхня створює перешкоди: як видно з рис. 1.10, група ОН верхньої молекули спрямована на неполярну (гідрофобну) поверхню і не може бути задіяною в утворенні водневого зв'язку. За збереження водневих зв'язків молекулам води доводиться "платити" зростанням упорядкованості – вони формують поблизу гідрофобної поверхні, як показано на рис. 1.10, упорядкований кригоподібний кластер. Отже, занурення гідрофобної молекули у водне оточення за рахунок зниження ентропії є дуже не вигідним енергетично.



**Рис. 1.10.** Кластер молекул води біля гідрофобної поверхні. Верхня молекула має не вигідну орієнтацію відносно поверхні

Якщо додати ще одну таку гідрофобну поверхню, то в системі буде дуже простий вихід (рис. 1.11): злипання поверхонь супроводжується визволенням молекул води, тобто зростанням ентропії. Отже, гідрофобні взаємодії мають цілком ентропійну природу. З рисунків 1.10 і 1.11 зрозуміло, що величина гідрофобного ефекту (енергетичний вигаиш від визволення води) пропорційна площі гідрофобної поверхні.



**Рис. 1.11.** Гідрофобна взаємодія між двома неполярними поверхнями

Гідрофобний ефект і гідрофобні взаємодії залежать від температури. Далі наведено дещо спрощене пояснення такої залежності (без занурення в детальний термодинамічний аналіз). Оскільки гідрофобний ефект має ентропійну природу, а ентропійний внесок у вільну енергію зростає з температурою (рівняння (1.2)), то ефективність гідрофобних взаємодій збільшується при підвищенні температури (зростає ефект визволення води в середовище, де рух її молекул прискорюється). Саме цим зростанням гідрофобного ефекту пояснюється добре відоме зниження розчинності у воді неполярних молекул газів при підвищенні температури. Для біологічних систем це також має велике значення: підвищення температури, навіть не дуже значне (до 35–37 °С), змінює енергетичний баланс міжмолекулярних взаємодій і призводить у пойкилотермних організмів до теплового шоку (див. розділ 8).

Гідрофобні взаємодії надзвичайно важливі для живих систем. Достатньо зазначити, що саме вони лежать в основі утворення мембран, без яких неможливим було б існування клітини. У наступних розділах буде показано, що гідрофобні взаємодії відіграють також провідну роль у підтриманні структури біологічних макромолекул.

## Контрольні запитання

1. Що таке вільна енергія? Як вона залежить від ентальпії та ентропії системи? У якому напрямку змінюється вільна енергія при переході системи до стану рівноваги?

2. Чи може при переході до стану рівноваги зрости ентальпія системи? За якої умови? Чи може зрости ступінь упорядкованості при переході до рівноважного стану?

3. Скориставшись рівнянням 1.6, розрахуйте відносну частку компонента А у складі комплексу С при концентраціях вільного компонента В  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$  моль/л. Вважайте, що  $K = 10^5$  л/моль.

4. Що таке ковалентний зв'язок, які основні властивості йому притаманні? За рахунок чого полімерний ланцюг може змінити конформацію?

5. Назвіть дві умови, за яких буде відбуватися хімічна реакція. Якою величиною визначається швидкість реакції?

6. У чому полягає фізичний механізм вандерваальсових взаємодій?

7. Поясніть ентропійну природу іонних взаємодій між двома високозарядженими макромолекулами.

8. Які властивості мають бути притаманні двом хімічним групам, щоб між ними міг утворитися водневий зв'язок? Чи можливий водневий зв'язок між двома метильними групами?

9. Чому водневий зв'язок часто буває важливим для специфічного міжмолекулярного впізнання?

10. Поясніть фізичну природу гідрофобного ефекту.

## Рекомендована література

Сиволоб, А.В. (2011) *Фізика ДНК*. Київ : ВПЦ "Київський університет".

Allen, J.P. (2008) *Biophysical chemistry*. Oxford UK : Wiley-Blackwell.

Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical chemistry*. New York : W.H. Freeman and Company.

Dill, K.A., Truskett, T.M., Vlachy, V. and Hribar-Lee, B. (2005) Modeling water, the hydrophobic effect, and ion solvation. *Annu. Rev. Biophys. Biomol. Struct.* 34, 173–199.

Nelson, D.L. and Cox, M.M. (2021) *Lehninger principles of biochemistry*. 8<sup>th</sup> ed. New York : W.H. Freeman and Company.

Pal, S. (2020) *Fundamentals of molecular structural biology*. London : Academic Press.

Pal, S.K., Zhao, L., Xia, T. and Zewail, A.H. (2003) Site- and sequence-selective ultrafast hydration of DNA. *Proc. Natl. Acad. Sci. USA* 100, 13746–13751.

Schrödinger, E. (1992) *What is life?: with Mind and matter and Autobiographical sketches*. Cambridge UK : Cambridge University Press.

Tanford, C. (1980) *The hydrophobic effect: formation of micelles and biological membranes*. New York : John Wiley & Sons.

## Розділ 2

# СТРУКТУРА І ПРИНЦИПИ ФУНКЦІОНУВАННЯ БІЛКІВ

Thus we have come to the conclusion that an organism and all the biologically relevant processes that it experiences must have an extremely 'many-atomic' structure and must be safeguarded against haphazard, 'single-atomic' events ...  
*E. Schrödinger. What is life?*

Переважає більшість процесів у живій клітині залежить від білків – біополімерів, мономерними одиницями яких є залишки амінокислот. За всієї великої різноманітності білків, їхня структурна організація підпорядкована достатньо простим загальним фізичним принципам, які визначають також можливість виконання білками численних функцій.

У структурному аспекті прийнято розрізняти кілька основних типів білків. Найважливішими з них є глобулярні водорозчинні білки – такі, що формують компактну – *глобулярну* – структуру й добре розчиняються у воді. Саме про такі білки йтиметься у першу чергу. Окремо будуть розглянуті також деякі інші важливі типи білків.

# Хімічна будова білків

## Амінокислоти

Узагальнену хімічну формулу амінокислот, що входять до складу білків, зображено на рис. 2.1.

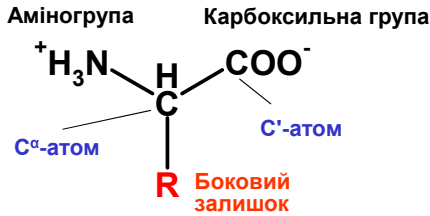
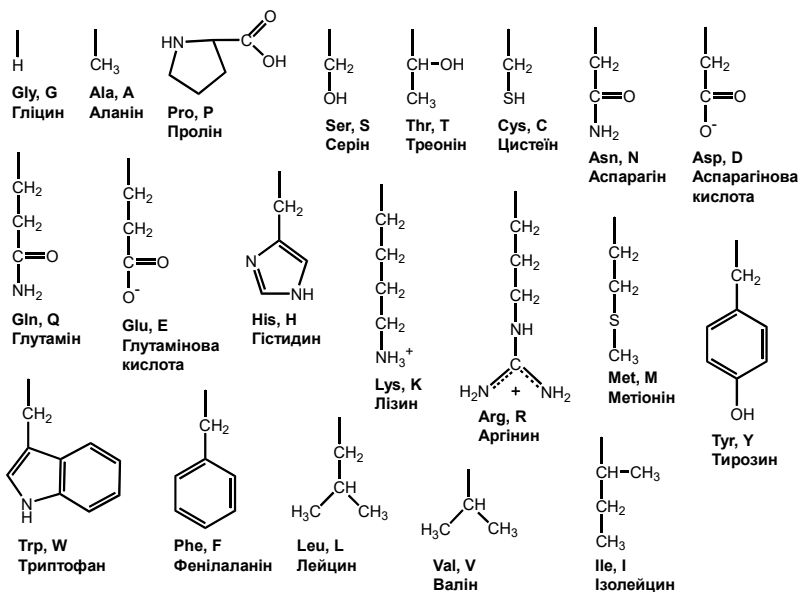


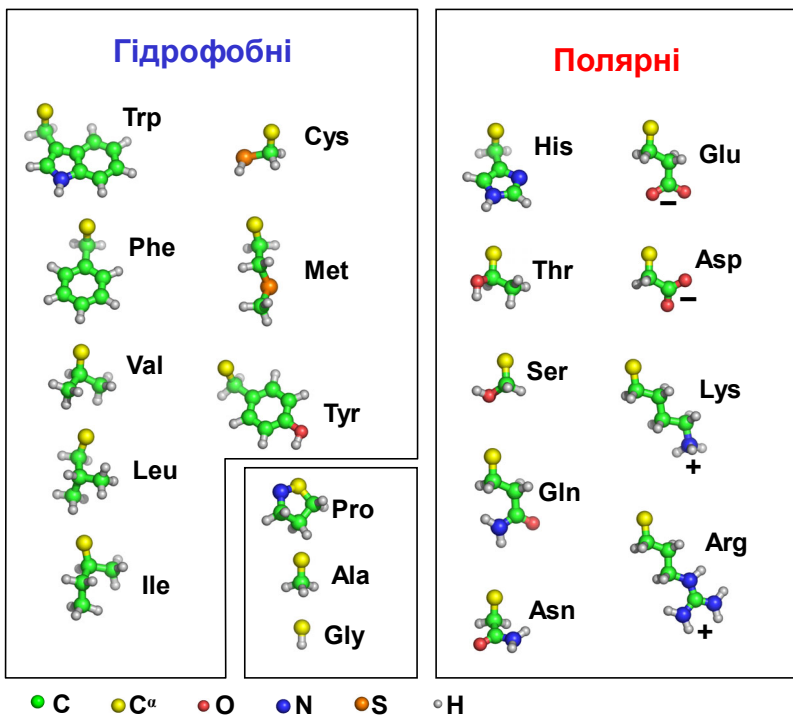
Рис. 2.1. Структурна формула амінокислоти

Як видно з рисунка, за нейтральних рН аміно- й карбоксильна групи (NH<sub>2</sub> і COOH), приєднані до спільного атома карбону (C<sup>α</sup>), перебувають в іонізованому стані. Саме ці, стандартні для всіх амінокислот, групи (наявність яких і зумовила назву сполук) забезпечують утворення полімерної молекули білка при її синтезі. Розрізняють амінокислоти за типом бокового (амінокислотного) залишку. Загалом у складі білків зустрічаються 20 амінокислот. Деякі білки містять також кілька додаткових мінорних амінокислот, але саме 20 визначаються генетичним кодом (розділи 4, 8). Назви, загальноприйняті скорочені позначення (три- й однолітерні), хімічні структурні формули та структури амінокислотних залишків зображено на рис. 2.2, 2.3. Слід зауважити, що, строго кажучи, серед наведених 20 сполук амінокислотами є 19: пролін є імінокислотою, оскільки у його складі боковий залишок утворює кільце з азотом, унаслідок чого замість аміногрупи NH<sub>2</sub> залишається іміногрупа NH (рис. 2.2).



**Рис. 2.2.** Структурні формули 20 амінокислотних залишків. Імінокислота пролін зображена повністю, для усіх інших амінокислот наведено їхні бокові залишки, приєднані до  $C^\alpha$ -атома (у стані іонізації, який вони мають при pH 7,0)

Два параметри амінокислотних залишків мають особливе значення для структури білка: розмір і здатність взаємодіяти з водою. За останньою ознакою залишки можна поділити на дві групи: полярні й неполярні (гідрофобні). Три гідрофобні залишки (окрема група на рис. 2.3) – маленького розміру, тому їх можна вважати нейтральними щодо спорідненості до води (гідрофобний ефект залежить від розміру поверхні, див. розділ 1).



**Рис. 2.3.** Структури амінокислотних залишків, приєднаних до C<sup>α</sup>-атома. Кольоровий код (унизу) позначення хімічних елементів зберігається на всіх рисунках далі

У глобулярних водорозчинних білках співвідношення між гідрофобними й полярними залишками становить у середньому 50 : 50. Унаслідок гідрофобного ефекту неполярні залишки прагнуть опинитися всередині структури білка, а полярні залишаються на поверхні, зберігаючи контакт із водою. Це і є головною рушійною силою, яка змушує білковий ланцюг укладатися певним чином у просторі, формуючи компактну структуру – *глобулу*.

Полярні амінокислоти тому так називаються, що прагнуть взаємодіяти з водою за рахунок водневих зв'язків (містять донорні й акцепторні групи). Серед них є залишки, які мають постійний заряд при pH 7,0: негативний у Asp і Glu (за рахунок карбоксильних

груп у складі бокового залишку), позитивний у Lys (аміногрупа) та Arg (гуанідинова група). Ніс також може набувати позитивного заряду в структурі білка в певному мікрооточенні (у розчині він іонізується при  $pH < 6,5$ ).

### Пептидний зв'язок і поліпептидний ланцюг

При білковому синтезі карбоксильна група однієї амінокислоти й аміногрупа іншої, втрачаючи молекулу води, утворюють ковалентний зв'язок, який називається *пептидним* (рис. 2.4). Група атомів  $O-C-N-H$  називається *пептидною групою*. Оскільки в утворенні зв'язку між мономерами беруть участь різні хімічні групи, то поліпептидний ланцюг має напрямок: його кінець, де залишається аміногрупа, називається N-кінцем, протилежний (карбоксильна група) – C-кінцем. Синтезуються поліпептидні ланцюги (розділ 8) у напрямку  $N \rightarrow C$ , у цьому ж напрямку завжди записують за умовчанням амінокислотні послідовності білків (*первинні структури*).

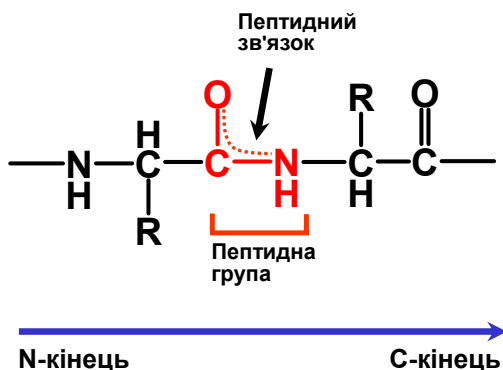
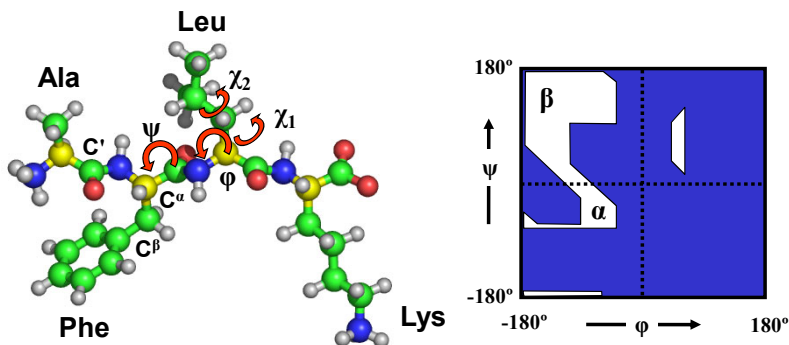


Рис. 2.4. Фрагмент поліпептидного ланцюга

Дві властивості пептидної групи відіграють ключову роль в утворенні просторової структури білка:

- 1) *Пептидна група полярна* – вона одночасно є донором і акцептором водневих зв'язків. Отже, універсальний остов поліпептидного ланцюга (пептидні групи та  $C^\alpha$ -атоми, що чергуються), від якого відгалужуються полярні й неполярні бокові залишки (див. рис. 2.5), є полярним.

2) Одна з пар електронів, що відповідає зв'язку C'-O (зазвичай позначається як подвійний, див. праворуч на рис. 2.4), насправді делокалізована також між атомами C'-N: *пептидний зв'язок є напівподвійним*, що унеможливорює обертання навкруг нього. Для більшості амінокислот реалізується вигідніша транс-конформація пептидного зв'язку: приєднані до пептидної групи C<sup>α</sup>-атоми спрямовані в різні боки (рис. 2.5). Єдиний виняток – пептидний зв'язок перед Pro, приблизно 10 % таких зв'язків перебувають у *цис*-конформації.



**Рис. 2.5.** Тетрапептид і приблизна діаграма Рамачандрана: синім кольором позначено заборонені, білим – дозволені для більшості амінокислотних залишків зони. Літери  $\alpha$  і  $\beta$  позначають зони, які відповідають певним регулярним конформаціям

Оскільки обертання навкруг пептидного зв'язку є неможливим, то конформація поліпептидного ланцюга може змінюватись тільки за рахунок обертання навкруг суміжних із пептидною групою одинарних зв'язків (обертальні кути  $\phi$  і  $\psi$ , рис. 2.5), а також обертання навкруг зв'язків у межах бокових залишків ( $\chi_1, \chi_2 \dots$ , рис. 2.5). Унаслідок стеричних обмежень не всі комбінації кутів обертання  $\phi$  і  $\psi$  є можливими – для поліпептидного ланцюга реалізується обертально-ізомерний механізм гнучкості. Дозволені для більшості залишків зони конформаційного простору в координатах  $\phi - \psi$  (діаграма Рамачандрана (Gopalasamudram Narayana Ramachandran)) показано на рис. 2.5.

Діаграми Рамачандрана мають специфічні ознаки для окремих залишків. Особливо це виявляється, якщо ввести третю координату – вільну енергію для кожної пари обертальних кутів. Інакше кажучи, переважна (найбільш імовірна) *локальна конформація ланцюга залежить від амінокислотної послідовності*.

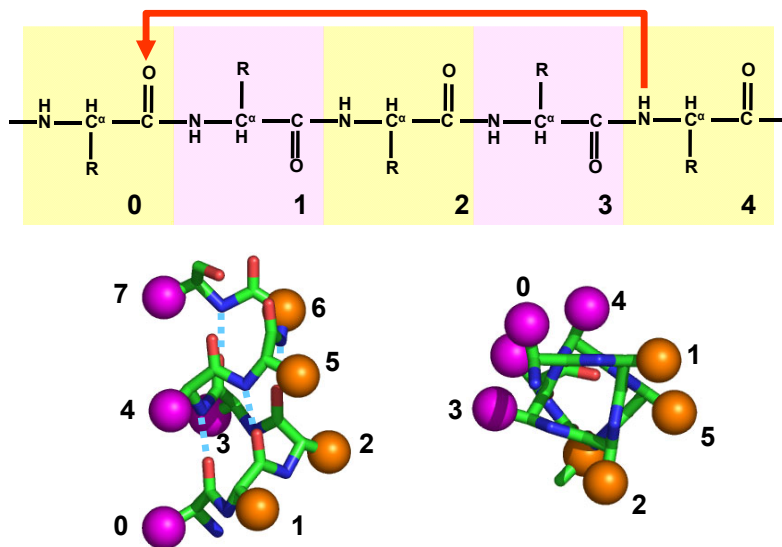
## Вторинна структура

Локальну конформацію амінокислотного ланцюга називають *вторинною структурою* білка. Частина ланцюга (іноді значна) може залишатись *невпорядкованою* – існувати у вигляді великого набору конформацій (кути  $\phi$  і  $\psi$  змінюються у часі в допустимих межах). Частина ланцюга перебуває у впорядкованій, але *нерегулярній* конформації (кути  $\phi$  і  $\psi$  мають стабільні значення, але варіюють від одного залишку до іншого). Проте у глобулярних білках завжди є також ділянки у *регулярних конформаціях, стабілізованих водневими зв'язками між пептидними групами*, тому термін "вторинна структура" іноді у вузькому сенсі відносять саме до них. Переважно у глобулярних білках представлено всього два типи таких конформацій –  $\alpha$ -спіраль ( $\alpha$ -helix) і  $\beta$ -структура ( $\beta$ -structure).

### $\alpha$ -Спіраль

Одна з допустимих конформацій на карті Рамачандрана ( $\phi \approx -60^\circ$ ,  $\psi \approx -45^\circ$ , рис. 2.5) передбачає наближення кисню пептидної групи  $i$ -го залишку до NH-групи  $i + 4$ . Перша група є акцептором водневого зв'язку, друга – донором. Система таких водневих зв'язків між кожним  $i$ -м та  $i + 4$ -м залишком стабілізує регулярну конформацію, яка й називається  *$\alpha$ -спіраллю* (рис. 2.6). Ця спіраль є правою (точка, що рухається вздовж ланцюга, обертається навкруг осі спіралі *за годинниковою стрілкою* й одночасно пересувається вздовж осі *вперед*), на один виток припадає 3,6 амінокислотних залишків, довжина одного витка вздовж осі (крок спіралі) становить 5,4 Å. Водневий зв'язок між пептидними групами замикає в кільце *13 атомів чотирьох залишків*: відповідно  $\alpha$ -спіраль позначається також як *спіраль  $4_{13}$* .  $\alpha$ -Спіраль можна приблизно уявити як циліндр діаметром  $\approx 10$  Å, на поверхні якого розташовані амінокислотні залишки.

Оскільки найімовірніша локальна конформація ланцюга залежить від послідовності, то різні амінокислоти та їхні комбінації мають різний потенціал щодо включення в  $\alpha$ -спіраль (як і в будь-яку іншу вторинну структуру). Наприклад, Pro не може бути включеним до  $\alpha$ -спіралі (немає донора водневого зв'язку). Gly також уникає регулярних структур, оскільки у їхньому складі цей дуже маленький залишок втратив би свою високу конформаційну свободу (ентропійна ціна була б надто високою). При цьому в глобулярних білках спостерігається закономірність: гідрофобні амінокислотні залишки зазвичай згруповані на поверхні спіралі в більш-менш суцільний гідрофобний кластер (рис. 2.6). Отже,  $\alpha$ -спіраль переважно утворюється на таких ділянках послідовності з високим спіральним потенціалом, які здатні сформувати гідрофобний кластер на поверхні спіралі.



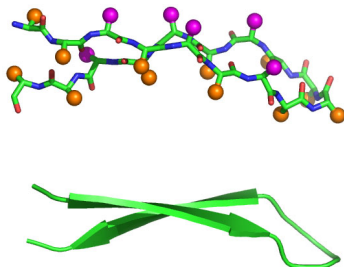
**Рис. 2.6.**  $\alpha$ -Спіраль. Червона стрілочка – водневий зв'язок.

Унизу: дві проєкції  $\alpha$ -спіралі, пунктирні лінії – водневі зв'язки між пептидними групами. Бокові залишки схематично показані у вигляді кульок, різне забарвлення символізує два типи залишків – полярні й неполярні

Така закономірність пояснюється просто. Унаслідок конкуренції з молекулами води водневі зв'язки між пептидними групами (а отже, і спіраль) у водному оточенні не дуже стабільні, адже утворення спіралі вимагає енергетичних витрат – зниження ентропії за рахунок обмеження конформаційної свободи. А водневі зв'язки у присутності води (тобто заміна зв'язку між пептидною групою і водою на зв'язок між пептидними групами) не здатні суттєво компенсувати ці витрати (див. розділ 1). Якщо ж на поверхні спіралі є гідрофобний кластер, то він прагне адсорбувати іншу гідрофобну поверхню (спіралі чи інших елементів регулярної вторинної структури). Тоді частина пептидних груп і водневих зв'язків між ними опиняється в неполярному оточенні, що суттєво стабілізує і ці зв'язки, і  $\alpha$ -спіраль. У результаті у глобулі залишаються лише  $\alpha$ -спіралі з гідрофобним кластером на поверхні. Це свідчить, що формування регулярної вторинної структури й укладання в глобулу – єдиний процес. Нижче при обговоренні глобулярної структури з'ясуємо, чому взагалі утворюється регулярна вторинна структура у глобулярних білках.

### $\beta$ -Структура

Інший тип регулярної вторинної структури, який часто зустрічається в білках, – це  $\beta$ -структура. Вона складається із витягнутих фрагментів поліпептидного ланцюга: кілька таких фрагментів –  $\beta$ -тяжів ( $\beta$ -strands) – утворюють  $\beta$ -структурний шар, у площині якого розташовані водневі зв'язки між пептидними групами тяжів, а від обох поверхонь відходять амінокислотні залишки (рис. 2.7).



**Рис. 2.7.** Антипаралельна  $\beta$ -структурна шпилька (амінокислотні залишки зображено кульками, як на рис. 2.6).  
Унизу: та сама шпилька в загальноприйнятому схематичному зображенні

Половина пептидних груп  $\beta$ -тяжів, які містяться на границях шару, залишаються ненасиченими водневими зв'язками, тому чим більше тяжів входять до складу шару, тим більш стабільним він є.  $\beta$ -Шар може бути антипаралельним (сусідні тяжі спрямовані у протилежні один щодо одного напрямки, як на рис. 2.7) або паралельним (це можливо, якщо тяжі не розташовані один за одним по ланцюгу). Як видно з рис. 2.7,  $\beta$ -шар завжди є дещо скрученим через те, що окремі тяжі не є абсолютно прямими.

Сусідні по ланцюгу амінокислотні залишки розташовані по різні боки  $\beta$ -структурного шару. І знову спостерігається вже зрозуміла закономірність: принаймні одна поверхня шару є переважно гідрофобною.

### **Спіраль $3_{10}$**

Права спіраль  $3_{10}$  (водневий зв'язок між  $i$ -м та  $i + 3$ -м залишками) зустрічається не дуже часто у вигляді коротких (три – чотири залишки) фрагментів.

### **$\beta$ -Поворот**

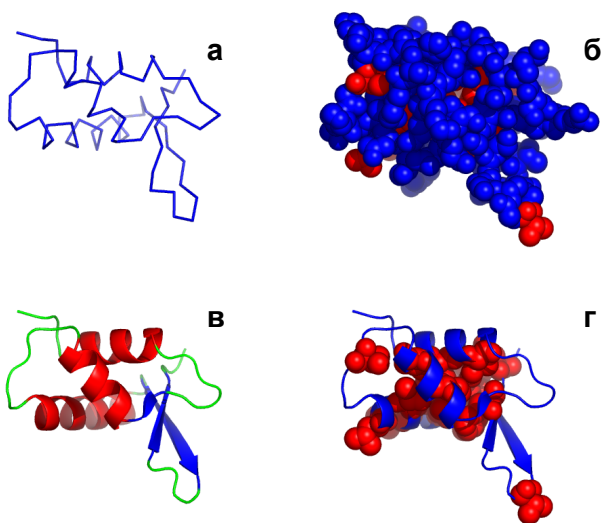
$\beta$ -Поворот, або  $\beta$ -вигин ( $\beta$ -turn), – невеликий елемент ланцюга із чотирьох залишків, стабілізований водневим зв'язком між пептидними групами першого й четвертого.  $\beta$ -Поворот забезпечує поворот ланцюга на  $180^\circ$  і часто міститься між двома сусідніми  $\beta$ -тяжами в антипаралельних  $\beta$ -шпильках.

### **Глобулярна структура**

Поліпептидний ланцюг глобулярного білка завдовжки 50–150 (максимально 250) амінокислотних залишків укладається в єдину щільну глобулу розміром 25–40 Å. Усередині такої глобули утворюється *гідрофобне ядро* з неполярних залишків, закрановане від води полярною поверхнею глобули. Ланцюги більшого розміру при формуванні такої структури займали б занадто великий об'єм – так, що полярні залишки змушені були б опинитися всередині, втративши вигідні контакти з водою.

Відповідно довші ланцюги утворюють кілька субглобул із власним гідрофобним ядром кожна – структурних *доменів*. Просторову структуру будь-якого поліпептидного ланцюга називають *третинною структурою*. Часто молекула білка складається з кількох поліпептидних ланцюгів, які можуть укладатися в єдину щільну глобулу або формувати окремі глобули, що взаємодіють між собою. В обох випадках кажуть про *четвертинну (субодичинну) структуру* білка.

Різні варіанти зображення просторової структури білка показані на рис. 2.8. На частинах (б) і (г) добре видно утворення всередині глобули щільного гідрофобного ядра.



**Рис. 2.8.** Різні варіанти зображення структури глобулярного білка (глобулярний домен гістону H5, код PDB 1HST):  
 а – остов ланцюга (з'єднані між собою  $C^\alpha$ -атоми);  
 б – атомна модель, червоні – атоми неполярних, сині – полярних залишків; в – схема будови, де показані сегменти вторинної структури; г – до попередньої схеми додано атоми неполярних залишків. Усі зображення в однакових проєкції та масштабі

## **Роль вторинної структури в утворенні глобули**

Розглядаючи структури глобулярних білків (на сьогодні їх установлено понад 160 тис.), можна сформулювати кілька загальних правил, яким вони підпорядковані.

**Правило 1. Каркас глобули завжди складається з ділянок вторинної структури ( $\alpha$ -спіралей і/або  $\beta$ -структурних шарів), гідрофобні поверхні яких, взаємодіючи між собою, утворюють гідрофобне ядро.**

Чому саме з ділянок регулярної вторинної структури? Це стає зрозумілим, якщо брати до уваги полярність / неполярність елементів поліпептидного ланцюга. Неполярні амінокислотні залишки прагнуть опинитися всередині глобули внаслідок гідрофобного ефекту – їхній контакт із водою є дуже не вигідним енергетично. Однак разом з ними в неполярному оточенні мають опинитися *полярні* пептидні групи остова. Втрата їхніх водневих зв'язків із молекулами води є енергетично не вигідною (енергія водневого зв'язку дуже висока). Єдиний вихід для пептидних груп – утворити водневі зв'язки між собою, а це можливо тільки у складі ділянок регулярної вторинної структури. Отже, *насичення пептидних груп водневими зв'язками у складі вторинної структури є необхідною передумовою утворення всередині глобули гідрофобного ядра*. Петлі й перемички між ділянками вторинної структури перебувають у нерегулярній конформації і залишаються на поверхні глобули, їхні пептидні групи при цьому зберігають контакт із водою. Звідси випливає

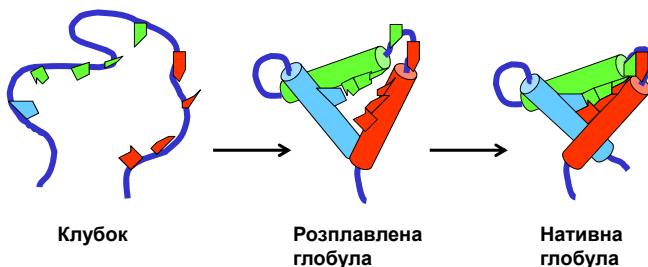
**Правило 2. Сегменти вторинної структури протягнуті від одного краю глобули до іншого. Перемички між ділянками вторинної структури залишаються на поверхні й ніколи не занурюються всередину глобули.**

Обидва правила мають ясний фізичний зміст і могли б бути сформульованими, виходячи із загальних фізичних міркувань.

## **Стабілізація глобули**

Як зрозуміло з викладеного вище, *головною рушійною силою утворення глобулярної структури є гідрофобні взаємодії*. Однак, щоб здійснювати певні функції, білок повинен мати ду-

же стабільну просторову структуру, яку гідрофобні взаємодії не можуть забезпечити: для їхньої реалізації достатньо просто витіснити воду з інтерфейсу між гідрофобними поверхнями. При цьому поверхні, тобто сегменти вторинної структури, яким вони належать, будуть зберігати певну свободу руху. Саме така ситуація спостерігається для проміжного стану, який реалізується при укладанні білка, – *розплавленої глобули* (рис. 2.9). Першим етапом укладання є одночасне й доволі швидке утворення сегментів вторинної структури та їхнє "злипання" за рахунок гідрофобних взаємодій. Загальна архітектура розплавленої глобули приблизно така сама, як у нативного функціонального білка, але спостерігаються значні рухи ділянок вторинної структури одна відносно одної – розплавлена глобула нагадує краплю рідини, де елементами, що рухаються, є ділянки вторинної структури. На другому етапі відбувається щільне пакування глобули за рахунок вандерваальсових взаємодій – точна взаємна підгонка амінокислотних залишків з утворенням численних контактів між ними, без порожнин усередині глобули. Результатом стає *нативна глобула*, яка є вже не краплею рідини, а *твердим тілом*.



**Рис. 2.9.** Схема укладання глобулярного білка

Твердість – найбільш характерна й важлива ознака білка, вона забезпечує наявність стабільної специфічної робочої поверхні, яка й виконує певну функцію. Решта молекули – лише каркас, який утримує цю робочу поверхню. Твердість забезпечує також надійність роботи білка, дозволяє уникнути розмивання специфічності. Адже молекула білка поводить себе як тверде тіло при зміні умов: витримує зовнішні зміни до певних меж, не змі-

нюючись сама, а за цими межами руйнується як одне ціле – за принципом "усе або нічого" (у процесі так званої *денатурації*).

Головні взаємодії, що стабілізують білкову глобулу:

- гідрофобні взаємодії між неполярними амінокислотними залишками, розташованими в межах сегментів вторинної структури;
- водневі зв'язки між пептидними групами в межах тих самих сегментів, які підтримують їхню структуру й, головне, забезпечують можливість занурення пептидних груп у гідрофобне ядро;
- вандерваальсові контакти всередині глобули, які забезпечують остаточне "огранювання" структури й надають їй жорсткості.

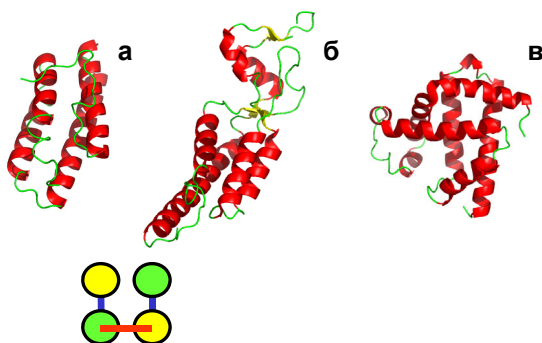
Додаткову роль у стабілізації можуть відігравати: водневі зв'язки всередині глобули за участю певних амінокислотних залишків; електростатичні взаємодії на поверхні глобули; ковалентні дисульфідні зв'язки між залишками Cys, наближеними один до одного внаслідок реалізації головних факторів укладання (дисульфідні зв'язки найбільше характерні для білків маленького розміру, що екскретуються клітиною); взаємодії з певними невеликими молекулами небілкової природи – лігандами, простетичними групами, іонами металів тощо.

### **Структурна класифікація глобулярних білків**

Оскільки глобула будується з елементів регулярної вторинної структури, яких лише два, то цілком природно розділити структурні домени глобулярних білків на три групи:  $\alpha$ -спіральні,  $\beta$ -структурні та змішаного типу. Виявляється, що кількість типів взаємного розташування у просторі сегментів вторинної структури (*структурних мотивів*), або типів *укладки* (fold), які забезпечують утворення гідрофобного ядра, обмежена.

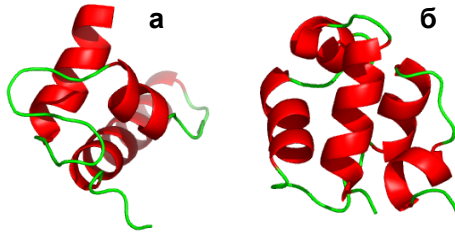
**$\alpha$ -Спіральні білки.** Серед структурних мотивів, які складаються тільки з  $\alpha$ -спіралей, переважно зустрічаються лише два типи. Довгі спіралі (завдовжки до 5–6 витків) формують  $\alpha$ -спіральні пучки ( $\alpha$ -helix bundles). Дві (рідше три) спіралі у складі пучка утворюють шар із гідрофобною поверхнею; два

такі шари, укладаючись паралельно (під кутом  $\sim 20^\circ$ ) або ортогонально ( $\sim 90^\circ$ ), формують гідрофобне ядро (рис. 2.10). Слід зробити такі важливі зауваження щодо зображених на цьому рисунку структур: усі вони не мають гомології між собою щодо амінокислотної послідовності; білки подібної структури (гемеритрин і білок вірусу тютюнової мозаїки) виконують різні функції; білки різної структури (гемеритрин і міоглобін) виконують подібну функцію зв'язування кисню.



**Рис. 2.10.** Спіральні пучки з паралельною (а, б) чи ортогональною (в) упаковкою в гемеритрині (а, 1HMD), у білку оболонки вірусу тютюнової мозаїки (б, 1E17), у міоглобіні (в, 1A6M). Внизу – схема взаємного розташування  $\alpha$ -спіралей у паралельних пучках, N- і C-кінцеві торці забарвлені в різні кольори, червона перемичка міститься ближче до глядача, сині – на протилежному боці

Другим, більш розповсюдженим, структурним мотивом є укладання трьох або більше  $\alpha$ -спіралей у приблизно сферичну глобулу. Така упаковка реалізується для коротких спіралей, які значно частіше зустрічаються в білках (близько 10–12 залишків, тобто 3–3,5 витка). Приклади наведено на рис. 2.11 (подібна упаковка реалізується також для трьох  $\alpha$ -спіралей, зображених на рис. 2.8). Короткі спіралі в цьому випадку розташовані по ребрах квазісферичного багатогранника, оточуючи приблизно сферичне гідрофобне ядро.



**Рис. 2.11.** Домени актинідину (а, 1AEC) і білка Cro бактеріофага 434 (б, 3CRO)

Указані структурні мотиви зустрічаються найчастіше, оскільки вони найкраще відповідають сформульованим вище двом правилам структурної організації білків. Крім того, розглядаючи ці структури, можна сформулювати ще одне правило.

**Правило 3. Перемички між сегментами вторинної структури уникають взаємного перехрещування – суміжні по ланцюгу перемички розташовані по різні боки глобули.**

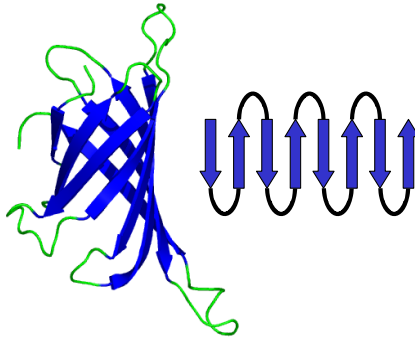
Таке перехрещування призвело б або до контакту між перемичками (тобто до їхньої дегідратації), або до значного вигину однієї з перемичок. Обидва наслідки є енергетично не вигідними. Інакше кажучи, у край мало можна уявити фрагментів амінокислотної послідовності, які б могли перетнутися без суттєвих енергетичних витрат.

Крім того, описані  $\alpha$ -спіральні мотиви найкраще забезпечують вимогу щільної упаковки глобули – саме за цих типів взаємної орієнтації спіралей їхні поверхні можуть здійснити найбільшу кількість вандерваальсових контактів між собою.

**$\beta$ -Структурні білки.** Основою  $\beta$ -структурних доменів є антипаралельні  $\beta$ -шари. Два шари  $\beta$ -структури, взаємодіючи своїми гідрофобними поверхнями, укладаються один на одного у своєрідний штабель,  $\beta$ -тяжі двох шарів орієнтуються при цьому паралельно або перпендикулярно (ортогональна упаковка) один до одного.

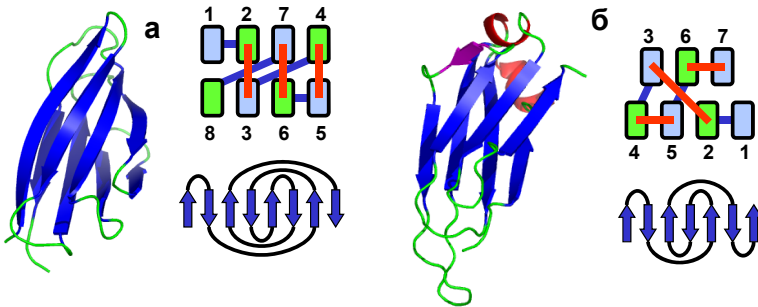
За ортогональної упаковки, унаслідок скрученості  $\beta$ -шарів, форма штабеля нагадує циліндр зі значним нахилом тяжів до осі (рис. 2.12). Ця упаковка так і називається –  $\beta$ -циліндр, або  $\beta$ -діжка ( $\beta$ -barrel). Проекція ходу ланцюга на площину для такого структурного мотиву має дуже простий вигляд: сусідні по ланцюгу ді-

лянки є також сусідами у складі шару, а проєкція еквівалентна візерунку "меандр" (часто зустрічається на античних амфорах).



**Рис. 2.12.**  $\beta$ -Циліндр у стрептовідині (1HQQ); праворуч – проєкція ходу ланцюга на площину, яка відповідає візерунку "меандр"

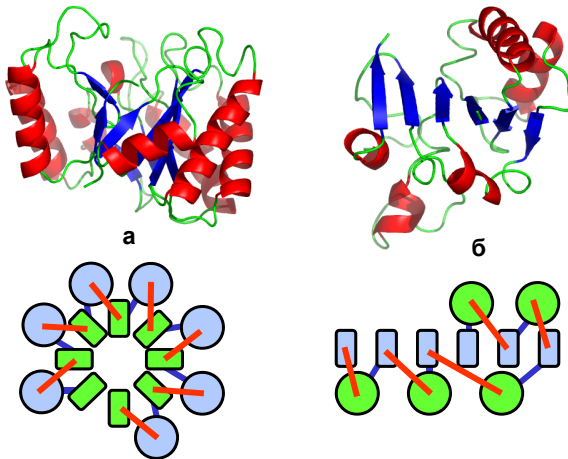
Інший тип ходу ланцюга нагадує ще один античний візерунок – "грецький ключ" (рис. 2.13), відповідна просторова упаковка має вигляд двох накладених один на одного шарів  $\beta$ -структури ( $\beta$ -сандвіч).



**Рис. 2.13.** а – ксиланзв'язувальний домен ксиланози (1E5B); б – імуноглобулінова упаковка в константному домені важкого ланцюга імуноглобуліну (15C8). Праворуч показано схеми упаковки двох шарів із торця ( $\beta$ -тяжі пронумеровані, блакитний – N-, зелений – С-кінець фрагмента) і проєкції ходу ланцюга на площину (для структури (б) короткі  $\alpha$ -спіралі та один  $\beta$ -тяж, які прикривають торець, на схемах не показано)

Наведені приклади просторової організації  $\beta$ -структурних доменів із певними варіаціями зустрічаються дуже часто у глобулярних білках, оскільки найкраще відповідають сформульованим вище правилам структурної організації білків і вимозі щільної упаковки глобули.

**$\alpha/\beta$ -Білки.** Так позначають білки, у яких  $\alpha$ -спіралі та  $\beta$ -тяжі регулярно чергуються по ланцюгу (відповідно  $\beta$ -структура є паралельною,  $\alpha$ -спіралі також паралельні одна одній). Для них найхарактернішими є два типи упаковок (рис. 2.14): 1)  $\alpha/\beta$ -циліндр – закручений у циліндр паралельний  $\beta$ -шар, оточений циліндром із  $\alpha$ -спіралей; відповідно утворюються два гідрофобні ядра – усередині  $\beta$ -циліндра та на інтерфейсі між  $\beta$ -структурою і  $\alpha$ -спіралями; 2) упаковка Россманна (Michael Rossmann) – приблизно планарний шар паралельної  $\beta$ -структури, прикритий двома шарами  $\alpha$ -спіралей; тут також утворюються два гідрофобні ядра на інтерфейсах між  $\beta$ -структурою та  $\alpha$ -спіралями.



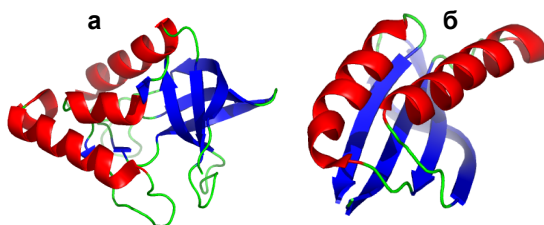
**Рис. 2.14.** а –  $\alpha/\beta$ -циліндр ендоглюканазі (1A3N);

б – упаковка Россманна в трансгідрогеназі (1D4O).

Унизу – відповідні схеми упаковок із торця  $\alpha$ -спіралей і  $\beta$ -тяжів

**$(\alpha+\beta)$ -Білки** – білки з нерегулярним чергуванням  $\alpha$ -спіралей і  $\beta$ -структурних тяжів. У їхньому складі два типи вторинної структури можуть бути рознесені: наприклад,  $\beta$ -сендвіч, торці якого

прикриті  $\alpha$ -спіралями чи  $\alpha$ -спіральною субдоменом (рис. 2.15, а) – ОВ-упаковка (oligonucleotide binding) із розташуванням елементів вторинної структури по ланцюгу  $\beta\beta\beta\alpha\beta\beta\alpha\beta\alpha$ . Іншим варіантом є  $\alpha\beta$ -складка ( $\alpha\beta$ -plait), зображена на рис. 2.15, б: шар антипаралельної  $\beta$ -структури, гідрофобна поверхня якого закрита  $\alpha$ -спіралями, розташування елементів вторинної структури по ланцюгу  $\beta\alpha\beta\beta\alpha\beta$ . Отже, як і для  $\alpha/\beta$ -білків, у складі єдиної глобули, утвореної шарами сегментів вторинної структури,  $\alpha$ -спіралі та  $\beta$ -структура завжди розділяються по різних шарах. Причиною є те, що занурення  $\alpha$ -спіралі в шар  $\beta$ -структури зруйнувало б систему водневих зв'язків між  $\beta$ -тяжами.



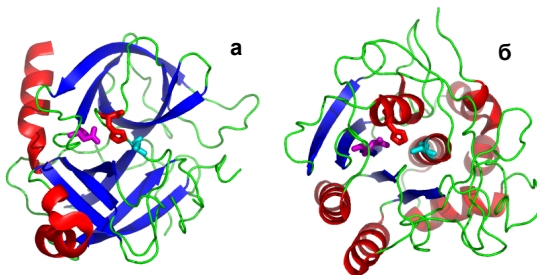
**Рис. 2.15.** а – стафілококова нуклеаза (1EY0);  
б – глобулярний домен рибосомного білка S6 (1J5E)

Вище було розглянуто деякі найтипівіші структурні мотиви, що зустрічаються у глобулярних білках. Насправді таких мотивів (упаковок), беручи до уваги деякі деталі, нараховують дещо більше 100. Зрозуміло, що білки з дуже подібною первинною структурою мають схожу просторову організацію. Однак десятки тисяч білків, які не мають нічого спільного в амінокислотних послідовностях, також вписуються в доволі обмежену кількість стандартних структурних мотивів. Причиною цього нетривіального спостереження є те, що стандартні мотиви найкраще відповідають певним правилам (частина яких сформульована вище) структурної організації, правилам, які впливають просто із загальних фізичних міркувань. Загалом можна уявити порушення будь-якого із цих правил. Наприклад, можна так "спроєкувати" амінокислотну послідовність, щоб гідрофобна перемичка між сегментами вторинної структури занурювалась у гідрофобне

ядро з утворенням там водневих зв'язків її пептидними групами. Однак зрозуміло, що для цього потрібен дуже обізнаний "дизайнер". Сам факт наявності стандартних упаковок свідчить про те, що *за походженням молекули білків є статистичними (випадковими) кополімерами 20 амінокислот*. І чим більше послідовностей може бути вписано в певну стандартну архітектуру глобули без порушення її стабільності, тим частіше така архітектура спостерігається у природі.

Це не означає, що *будь-який* статистичний кополімер амінокислот може сформувати стабільну глобулу. У результаті еволюційного відбору залишились лише ті кополімери (дуже невелика кількість), які допускають утворення стабільного, щільно упакованого (з великою кількістю вандерваальсових контактів) гідрофобного ядра. Слід зауважити, що вимога утворення гідрофобного ядра не означає, що *всі* неполярні залишки мають опинитися всередині (див., напр., рис. 2.8, (б), (г)): у середньому для всіх білків тільки 80–85 % гідрофобних залишків занурюються у гідрофобне ядро, решта, які залишаються на поверхні, не впливають суттєво на стабільність глобул. І це є ще одним підтвердженням випадковості білкових послідовностей.

Отже, сучасні білки є випадковими полімерами амінокислот, спочатку відібраними за критерієм стабільності їхньої глобулярної структури, а потім "відредагованими" добором шляхом невеликої кількості амінокислотних замін у двох напрямках: підвищення стабільності просторової структури й утворення активного центра, який має виконувати певну функцію. У результаті білки з різними функціями часто мають подібну просторову будову. І навпаки, білки, що виконують подібні функції, можуть бути зовсім різними за структурою. На рис. 2.16 зображено дві серинові протеази (однакова функція), які належать до різних структурних типів, але містять однаково організований активний центр. Отже, для функціонування часто важливою є не стільки просторова структура білка, скільки сам факт її наявності: жорсткий каркас глобули утримує активний центр у певній просторовій конфігурації, не дозволяючи йому рухатися під дією теплових флуктуацій.



**Рис. 2.16.** а –  $\alpha$ -хімотрипсин (4CHA); б – субтилізін (1SBC).  
Для кожного білка показано три залишки (Asp, His, Ser), які формують активний центр цих протеаз

## Типи неглобулярних білків

У цьому розділі досі йшлося про глобулярні водорозчинні білки – найважливіший тип білків, які й далі будуть у центрі нашої уваги. Крім таких білків, розрізняють також фібрилярні, мембранні й водорозчинні неструктуровані білки.

### Фібрилярні білки

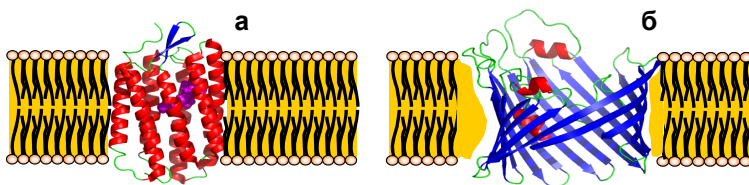
Під фібрилярними білками часто традиційно розуміють усі білки, що беруть участь у формуванні різноманітних фібрил як усередині клітини, так і у позаклітинному просторі. Внутрішньоклітинні білкові фібрили (мікротрубочки, актинові фібрили, проміжні філаменти) утворюються з білкових субодиниць, що мають глобулярну структуру. Для взаємодії між субодиницями проміжних філаментів використовується також особлива супервторинна структура – дві закручені одна навкруг одної  $\alpha$ -спіралі (*coiled coil*). Така структура достатньо часто зустрічається і в інших білках, деякі приклади стосовно ДНК-зв'язувальних білків наведені у наступному розділі. Позаклітинні фібрилярні білки (колаген, кератин, фіброїн шовку тощо) не мають глобулярної структури й формують фібрили за рахунок утворення різноманітних міжланцюгових супервторинних структур – на основі  $\alpha$ -спіралей,  $\beta$ -структури, особливої потрійної спіралі у складі колагену.

## Мембранні білки

Особливості структурної організації мембранних білків або доменів, що занурюються у клітинну мембрану, пов'язані з тим, що вони перебувають у неполярному оточенні. Основа мембрани – це подвійний шар, який утворюється за рахунок гідрофобних взаємодій між неполярними вуглеводневими хвостами молекул ліпідів (здебільшого фосфоліпідів), полярні головки ліпідів при цьому залишаються зовні подвійного шару, взаємодіючи з водою. Близько половини (або більше) ваги мембрани становлять мембранні білки, що переважно функціонують як канали для пасивного чи активного транспорту крізь мембрану, і рецептори, які сприймають певні хімічні сигнали.

Уже описані закономірності структурної організації водорозчинних білків дозволяють легко зрозуміти основні вимоги до білків мембранних. Оскільки білок або його частина має зануритися в гідрофобне оточення, то:

- 1) пептидні групи мають бути насичені водневими зв'язками – у мембрану занурюються або довгі (по товщині мембрани)  $\alpha$ -спіралі, або  $\beta$ -циліндр, оскільки він з'єднує водневі зв'язки границі  $\beta$ -структурного шару (рис. 2.17);
- 2) зовнішній (обернений до ліпідів) бік  $\alpha$ -спіралі чи  $\beta$ -структурного тяжа має містити переважно гідрофобні залишки.



**Рис. 2.17.** Структури двох мембранних білків: бактеріородопсину (а, 1R2N) і порину (б, 2POR)

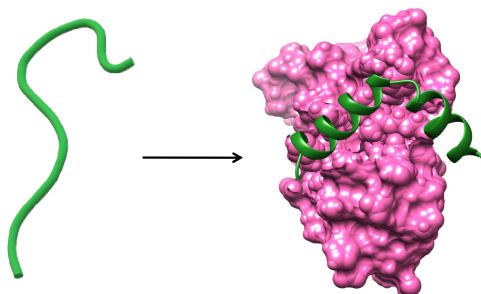
На рис. 2.17 зображено дві такі типові структури. Молекула бактеріородопсину містить пучок із семи довгих  $\alpha$ -спіралей, перемички між якими виходять із мембрани в полярне оточення. Усередині пучка утворюється вузький канал, на внутрішній по-

верхні якого міститься невелика кількість полярних залишків. У каналі розміщується також кофактор небілкової природи – ретинол, за допомогою якого крізь канал проводиться протон. Молекула порину утворює широкий  $\beta$ -циліндр, внутрішня поверхня якого (внутрішній діаметр циліндра  $\sim 15 \text{ \AA}$ ) полярна, тобто канал є мембранною порою низької селективності.

### **Неструктуровані білки**

Далеко не всі водорозчинні білки мають упорядковану просторову структуру. Доволі велика кількість білків (до 40 % у клітинах еукаріотів) не утворюють за фізіологічних умов жорсткої третинної структури, залишаючись неупорядкованими (неструктурованими), до 70 % білків мають у своєму складі окремі неупорядковані ділянки. Такі неструктуровані білки (*intrinsically disordered proteins*) або окремі ділянки глобулярних білків, часто розташовані на кінцях поліпептидного ланцюга, зазвичай збірнені на гідрофобні амінокислоти, що робить неупорядкований стан енергетично вигіднішим.

Доволі загальною рисою неструктурованих білків є їхня здатність взаємодіяти з багатьма партнерами, що робить їх своєрідними платформами для збирання мультимолекулярних комплексів (у наступних розділах наведено кілька прикладів такого типу), – тим самим вони виконують важливі функції з регуляції багатьох клітинних процесів. Часто зв'язування неструктурованого білка з іншими білками чи нуклеїновими кислотами зумовлює утворення певної просторової структури, яку стабілізують ці нові взаємодії (рис. 2.18). Утворення такої структури, своєю чергою, зумовлює формування робочої поверхні, що має спорідненість до наступної макромолекули: неструктурований білок (який стає структурованим унаслідок міжмолекулярних взаємодій) спрацьовує як своєрідний перемикач, необхідний для формування мультимолекулярного комплексу, задіяного до певного регуляторного акту.



**Рис. 2.18.** Неструктурований білок утворює впорядковану структуру при взаємодії з іншим білком (1KDX)

## Конформаційна рухливість білків

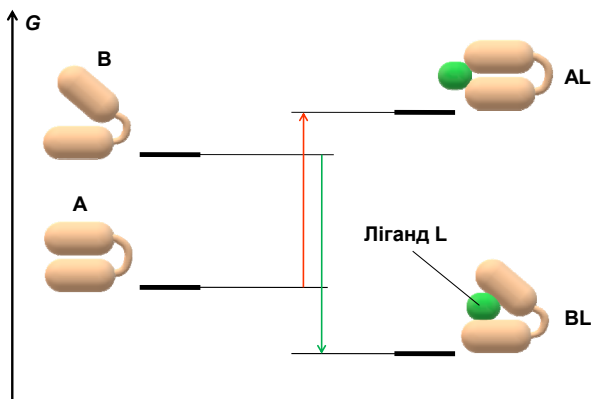
У відповідь на екстремальні зовнішні впливи *in vitro* (напр. підвищення температури) глобулярний білок може втратити впорядковану структуру – стати *денатурованим*. Дослідження цього процесу є дуже важливими для з'ясування механізмів стабілізації структури білка. Однак ще цікавішою є можливість зміни структурної форми за фізіологічних умов – саме в цьому сенсі будемо далі розуміти конформаційну рухливість білка.

Хоча нативна молекула білка – тверде тіло, вона доволі маленька, тобто, на відміну від макроскопічних твердих тіл, молекула білка перебуває під дією теплового руху як власних частин, так і оточуючих молекул. Невеличкі конформаційні зміни, передусім кутів обертання основного ланцюга й бокових залишків на перемичках між елементами вторинної структури, можуть зумовити тимчасові взаємні зсуви ділянок вторинної структури. Більша рухливість притаманна мультидоменним і субодиничним білкам: окремі домени (субглобули) можуть деформуватися і, що важливіше, рухатися відносно один одного. Отже, в основі конформаційної рухливості білка лежить взаємний зсув великих структурних блоків (субодиниць, доменів або елементів вторинної структури) за рахунок невеликих локальних деформацій. Іноді відбуваються локальні перебудови вторинної структури в

різних структурних станах білка – руйнування або утворення спіралей тощо. Зрозуміло, що структурні перебудови змінюють форму робочої поверхні молекули, тобто впливають на виконання білком його функцій.

Однак якщо структурні блоки (сегменти вторинної структури, домени, субодиниці) випадково зсуваються один відносно одного, то взаємодія між ними швидко відновлює вихідну конфігурацію – у випадку, коли немає додаткових факторів, що можуть порушити баланс вільної енергії. Відповідно для реалізації радикальних структурних перебудов білка чи білкового комплексу в системі має з'явитись деякий додатковий елемент.

Припустимо, що амінокислотна послідовність білка дозволяє принципове існування двох структурних станів – А і В. Зазвичай вільна енергія одного зі станів (напр. А) є нижчою, тоді конформаційна рівновага зсунута саме до нього (рис. 2.19, ліворуч). Як можна перемкнути цю рівновагу на користь стану В (якщо, напр., саме він має виконувати якусь функцію)? Типовий шлях такого перемикання полягає у використанні певної сполуки (ліганду). Лігандами можуть бути органічні низькомолекулярні сполуки, іони, інші білки – залежно від конкретної системи (одним із прикладів радикальної структурної зміни є утворення просторової структури у згаданих вище неструктурованих білках за умов їхньої взаємодії з іншими макромолекулами). На рис. 2.19 показано ситуацію, коли ліганд має спорідненість саме до форми В і не має такої до форми А. Це означає, що вільна енергія знижується при зв'язуванні ліганду з формою В, але зростає при зв'язуванні з формою А, тобто зв'язування з А не відбувається. Головним є те, що у присутності ліганду здійснюється зміна балансу вільної енергії: стан В (точніше комплекс ВL) стає переважним – його вільна енергія нижча, ніж вільна енергія стану А.



**Рис. 2.19.** Схема структурних переходів між станами А і В молекули білка, опосередкованих лігандом. G – шкала вільної енергії

Результатом такої фіксації одного зі структурних станів за рахунок зв'язування з лігандом є зміна робочої поверхні білка. При цьому робоча поверхня часто розташована зовсім в іншому місці відносно ділянки зв'язування ліганду. Ліганд (або його відсутність) є лише сигналом, регулятором, який перемикає структурні, а відповідно, і функціональні стани. Регуляція такого типу отримала назву *алостеричної регуляції* (від *άλλος* – інший).

### **Принципи функціонування білків**

Усі різноманітні операції, які виконують білки, можна описати трьома словами:

ЗВ'ЯЗАТИ – ТРАНСФОРМУВАТИ – ВІДПУСТИТИ.

Звичайно, не обов'язково кожний білок здійснює всі три операції. Може виконуватися лише одна з них – зв'язування. Так працюють імуноглобуліни: з надзвичайно високою селективністю зв'язують антиген (точніше, антигенний детермінант – невеличку ділянку на поверхні чужорідної для організму молекули). Константа зв'язування настільки висока, що практично взаємодія є

нерівноважною (неможлива дисоціація комплексу). Далі комплекс просто руйнується певними клітинами імунної системи.

Операцію зв'язування з іншими білками й молекулами небілкової природи виконують білки, що залучені до утворення різноманітних клітинних структур. У цьому випадку зв'язування може змінюватись дисоціацією (напр., при утворенні / руйнуванні філаментів цитоскелета), що часто супроводжується структурними змінами в молекулі білка.

Операції зв'язати – відпустити виконують численні регуляторні білки (транспортні білки, рецептори, фактори транскрипції тощо). Для їхнього функціонування часто велике значення мають структурні перебудови за механізмом, показаним схематично на рис. 2.19: певний ліганд, виконуючи роль хімічного сигналу, індукує або блокує спорідненість білка-регулятора до певної молекули-мішені. Конкретні приклади такого типу будуть досить часто зустрічатися в наступних розділах.

Найважливіший для живих систем функціональний клас білків – ферменти, або ензими (enzymes), які здійснюють каталіз численних біохімічних реакцій, виконуючи три операції: зв'язують субстрати, сприяють їхній трансформації, відпускають продукти реакції.

### **Механізм ферментативного каталізу**

Енергетичну схему хімічної реакції (рис. 1.6) було розглянуто в розділі 1. На рис. 2.20 ту саму схему доповнено ефектами, що є наслідком присутності ферменту. Уповільнення реакції зумовлено високим значенням вільної енергії активації – за відсутності ферменту реакція практично не відбувається (червоний шлях на рис. 2.20). Відповідно каталітична дія ферменту полягає у зниженні цього енергетичного бар'єра. Оскільки бар'єр, як будь-яка вільна енергія, має дві складові – ентальпійну й ентропійну (див. розділ 1), то фермент здійснює зниження енергії активації двома шляхами (зазвичай обидва мають значення).

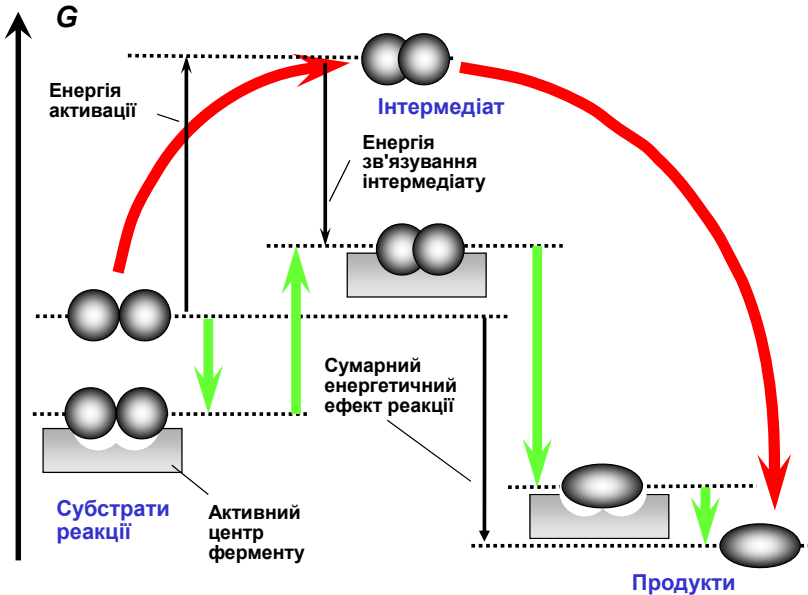


Рис. 2.20. Енергетична схема ферментативної реакції;  
 $G$  – шкала вільної енергії

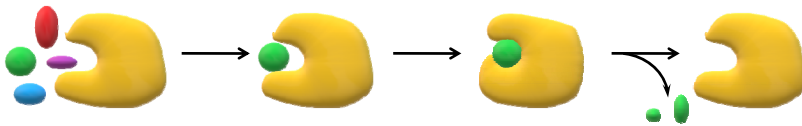
- *Ентропійний каталіз.* Перше, що робить фермент, – зв'язує субстрати у своєму активному центрі за рахунок спорідненості до них (перша зелена стрілка вниз). Однак головне полягає в тому, що субстрати зв'язуються не абияк – реалізується певна, найбільш сприятлива для реакції, взаємна орієнтація і певна орієнтація щодо хімічних груп активного центра. Це означає, що з великої кількості орієнтацій обирається лише одна: взаємодії субстратів з активним центром компенсують ентропійні витрати на впорядкування субстратів.
- *Ентальпійний каталіз.* На шляху перебудови системи ковалентних зв'язків субстратів обов'язково існує проміжна високоенергетична сполука – інтермедіат (див. розділ 1). Саме до нього активний центр ферменту має високу спорідненість (вищу, ніж до субстратів), реалізуючи з ним певні взаємодії. Слова "висока спорідненість" означають, що

зв'язаному в активному центрі інтермедіату відповідає значно нижча енергія, ніж вільному (рис. 2.20). Саме за рахунок цієї високої спорідненості до інтермедіату активний центр і знижує ентальпійну складову енергії активації: бар'єр (зелена стрілочка вгору) залишається, але стає значно нижчим – таким, що його можна швидко подолати за рахунок теплових флуктуацій.

Наступні дві зелені стрілочки вниз на рис. 2.20 показують закінчення процесу: з бар'єра система спускається униз по градієнту вільної енергії – у стан продуктів реакції; далі (і це також важлива подія) продукти, які не мають спорідненості до активного центра (вільна енергія дисоційованого продукту нижча, ніж зв'язаного), звільняють його для інших субстратів.

Для успішного виконання завдання каталізу обома шляхами *активний центр ферменту повинен бути дуже жорстким* – мати чітко зафіксовану у просторі й часі структуру. Для цього, власне, і потрібна твердість глобули, на якій розташований активний центр, і тут зовсім не потрібна (навпаки – шкідлива) конформаційна рухливість ферменту. Білковим ферментам деяких простих реакцій (напр. протеазам на рис. 2.16) висока конформаційна рухливість не притаманна.

Однак ферменти часто мають виконувати складніше завдання: дискримінувати субстрати, тобто перетворити лише один субстрат із певного набору аналогів. У типовому сценарії у цьому випадку задіяні принаймні дві структурні форми ферменту й, відповідно, його активного центра (рис. 2.21). У *відкритій* формі активний центр має невисоку спорідненість до субстратів певного типу (завдяки чому можливе швидке перебирання варіантів) і не здатен каталізувати хімічне перетворення. Якщо нарешті зв'язується потрібний субстрат, то він спрацьовує як ліганд на рис. 2.19 – індукує перетворення ферменту на *закриту* форму. Такий процес було названо Кошландом (Daniel E. Koshland) *індукованою відповідністю* (induced fit). Закрита форма жорстко фіксує субстрат, каталітичний центр спрацьовує і здійснюється хімічна реакція. У результаті ліганд змінюється і відбувається зворотне перетворення білка на відкриту форму зі звільненням продуктів.

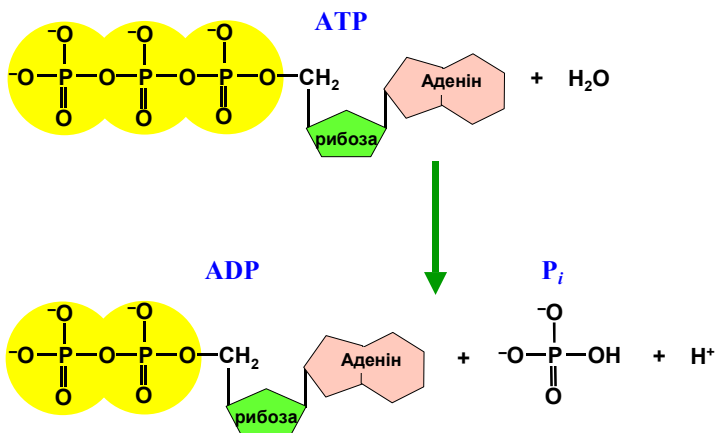


**Рис. 2.21.** Індукована відповідність: структурні перебудови ферменту при дискримінації субстратів

Повертаючись до загальної схеми реакції, зображеної на рис. 2.20, варто ще раз зауважити: з ферментом чи без, реакція може відбуватися тільки в напрямку зниження вільної енергії. Протилежний напрямок може здійснюватися лише за умови використання "обхідних шляхів" і особливих механізмів (також за участю білків).

### Механізми використання енергії гідролізу АТФ

Джерелом енергії для всіх процесів, що її потребують, у живих системах є зазвичай нуклеозидтрифосфати (див. розділ 3). Найчастіше використовується аденозинтрифосфат (adenosine triphosphate, АТФ). Ця сполука, яка утворюється при перенесенні протонів крізь мембрану хлоропластів (фотосинтез) чи мітохондрій (окиснювальне фосфорилування), схематично показана на рис. 2.22.



**Рис. 2.22.** Гідроліз АТФ

Вона складається з азотистої основи (аденіну), пентозного цукру (рибози) і трьох залишків фосфорної кислоти. Гідроліз АТР спричиняє відщеплення одного з них – залишаються аденозиндифосфат (adenosine diphosphate, ADP) і неорганічний фосфат  $P_i$  (рис. 2.22). Як будь-яка інша реакція, гідроліз АТР потребує каталізу (за участю широкого класу ферментів – АТРаз). Роль АТР як джерела енергії зумовлена тим, що ця реакція супроводжується доволі великим зниженням вільної енергії.

Зв'язок між фосфатними залишками, який руйнується при перетворенні АТР на ADP, іноді так і називають – *макроергичним*. Слід розуміти, що це метафора, яка зовсім не означає, що цей зв'язок заощаджує надзвичайно велику енергію, яка виділяється при його руйнуванні. По-перше, у цьому зв'язку немає нічого особливого. По-друге, загальна кількість ковалентних зв'язків не змінюється в ході реакції – фосфатний залишок переноситься на молекулу води (рис. 2.22). Змінюється тільки *вільна* енергія набору хімічних сполук, які зображені на рис. 2.22, причому головний внесок у цю зміну роблять концентраційні (тобто ентропійні) ефекти.

Скориставшись рівнянням (1.5), різницю вільних енергій продуктів і субстратів реакції, зображеної на рис. 2.22, можна записати як

$$\Delta G = \Delta G^* + RT \ln \frac{[\text{ADP}][P_i][\text{H}^+]}{[\text{ATP}][\text{H}_2\text{O}]},$$

де у квадратних дужках – молярні концентрації відповідних компонентів,  $\Delta G^*$  – стандартна зміна вільної енергії,  $RT = 0,6$  ккал/моль (див. розділ 1). Величина  $\Delta G^*$  є *позитивною* – стандартна вільна енергія продуктів вища, ніж у субстратів, тому, здавалося б, реакція не має відбуватись. Проте ця величина у розглядуваному випадку не має великого сенсу: енергетичний ефект реакції  $\Delta G$  дорівнює  $\Delta G^*$  за концентрацій усіх компонентів 1 моль/л. Однак концентрації води та протонів зберігаються постійними як у буферному розчині *in vitro*, так і у клітині, причому вони значно відрізняються від 1 моль/л: у розведеному розчині  $[\text{H}_2\text{O}] \approx 55,5$  моль/л, при рН 7,0  $[\text{H}^+] = 10^{-7}$  моль/л. Отже, відно-

шення цих концентрацій – також постійна величина, і можна ввести іншу, більш змістовну, стандартну вільну енергію реакції

$$\Delta G^0 = \Delta G^* + RT \ln \frac{[H^+]}{[H_2O]}$$

і записати

$$\Delta G = \Delta G^0 + RT \ln \frac{[ADP][P_i]}{[ATP]}.$$

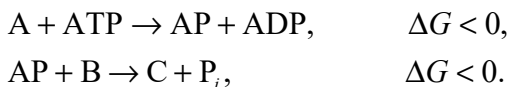
Тепер величина  $\Delta G^0 \approx -7,3$  ккал/моль, що вказує на високу енергетичну вигідність реакції гідролізу. Головний внесок дає та обставина, що в реакції народжується вільний протон, який виділяється у середовище з надзвичайно низькою концентрацією протонів, що зумовлює підвищення ентропії системи. Інакше кажучи, за законом діючих мас, підтримання на дуже низькому рівні загальної концентрації протонів має зсувати рівновагу в бік їх утворення. При цьому в клітинах концентрації усіх компонентів, що залишилися в останньому рівнянні, також підтримуються на певних рівнях:  $3 \div 8$  ммоль/л АТР;  $0,2 \div 1$  ммоль/л АDP;  $2 \div 8$  ммоль/л неорганічного фосфату. Підстановка цих значень дає реальний енергетичний ефект від гідролізу АТР:  $\Delta G = -12 \div -16$  ккал/моль.

Ця енергія використовується двома шляхами: 1) для хімічної модифікації субстратів, які самі не можуть вступити в певну реакцію (переважно маються на увазі реакції синтезу), оскільки вона супроводжувалась би зростанням вільної енергії; 2) для виконання механічної роботи молекулярними машинами.

**Модифікація субстратів.** Розглянемо хімічну реакцію



Оскільки утворення С супроводжується зростанням вільної енергії, то такий процес є неможливим. Наведемо типовий спосіб, за яким жива система діє у цьому випадку:



Тут, власне, відбувається розділення процесу гідролізу АТР на два кроки: спочатку здійснюється перенесення фосфатного залишку на молекулу А (фосфорилування), потім фосфорильована молекула А вступає в реакцію з В з одночасним відщепленням фосфату. Обидві реакції відбуваються, оскільки супроводжуються зниженням вільної енергії. Сумарне зниження вільної енергії дорівнює (або скоріше є трохи меншим за абсолютною величиною) енергії гідролізу АТР. З певними модифікаціями (напр., на першому кроці на одну зі сполук переноситься не фосфат, а два фосфатні залишки або аденозинмонофосфат АМР) саме за такою схемою відбуваються хімічні реакції синтезу, зокрема біополімерів.

**Молекулярні машини** – білкові молекули чи, частіше, макромолекулярні комплекси, утворені за участю білків, які здатні виконувати механічну роботу. Робота таких машин забезпечує переміщення клітинних структур і органел (напр. розходження хромосом при клітинному поділі), зміну форми клітин, їхнє активне переміщення і транспортування крізь мембрану проти градієнтів концентрації, м'язові скорочення, переміщення полімераз уздовж матриці при синтезі біополімерів тощо. Усе це потребує енергії, джерелом якої також є АТР (або інші нуклеозидтрифосфати).

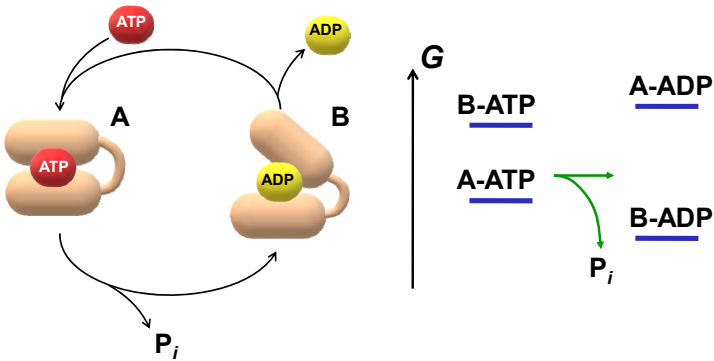
Узагалі машина – це пристрій для перетворення тієї чи іншої форми енергії на рух. Звичайна механічна машина перетворює будь-яку енергію на механічну, умовою заощадження механічної енергії при цьому є висока інерція (маса) частин машини й певна їхня жорстка взаємна орієнтація. Точніше, конструкція машини зумовлює рухи її частин тільки певними заданими траєкторіями.

Молекулярна машина, на відміну від макроскопічної механічної машини, по-перше, маленька, тобто піддається тепловому руху своїх власних частин і зовнішніх молекул, що робить неможливим заощадження механічної енергії. По-друге, вона побудована з полімерів, а це означає, що її частинам притаманна конформаційна рухливість. Проте внаслідок того, що молекула білка має певну доволі жорстку просторову організацію, конструкція молекулярної машини допускає не будь-які, а певним чином спрямовані у просторі рухи її частин. Саме на цих особ-

ливостях молекулярної машини побудована її робота, основними принципами якої є такі:

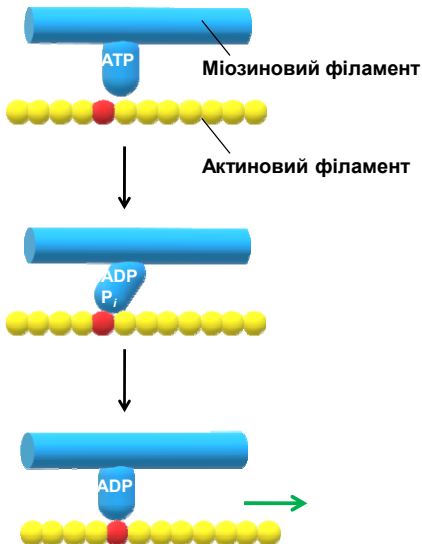
- *Конформаційна рухливість* молекулярної машини забезпечує їй можливість існувати в кількох структурних станах (двох-трьох), які розрізняються переважно на рівні просторового розташування великих структурних блоків – доменів або субодиниць.
- Структурні стани мають різну спорідненість до певних лігандів. *Взаємодії з лігандами* (факторами) фіксують певні стани.
- *Хімічні реакції*, які каталізуються машиною, спричиняють заміну лігандів, а відповідно – і перехід в інший структурний стан.
- Рушійною силою для переміщення блоків є *тепловий рух*: блоки рухаються хаотично (хоча й відповідно до конструкції машини); зв'язування лігандів і заміна їх унаслідок реакцій каналізують ці рухи в певних напрямках.
- Результатом структурних перебудов є *переміщення* структурних блоків у просторі та/або зміна характеру взаємодії машини з її оточенням – рух або всієї машини, або відносний рух її частин.

Загальний принцип перемикання структурних станів такий самий, як на рис. 2.19 – для фіксації станів використовується енергія зв'язування з лігандом. Лігандами часто стають АТФ і продукти його гідролізу. Реакція гідролізу (тут гідроліз відбувається за один крок, як зображено на рис. 2.22) є просто засобом змінити ліганд – і перемкнути структурні стани. Висока вільна енергія гідролізу визначає високу ймовірність проходження реакції (див. рівняння (1.6)), а відповідно й ймовірність перемикання конформацій білка. Типовий елементарний "такт" роботи молекулярної машини зображено на рис. 2.23: структурна форма А зв'язує АТФ із високою спорідненістю; спрацьовує АТФ-зв'язаний каталітичний центр; АДФ, що виникає внаслідок гідролізу, має підвищену спорідненість до форми В, наслідком чого є структурна перебудова. Потім АДФ витісняється новою молекулою АТФ і цикл повторюється.



**Рис. 2.23.** Схема роботи молекулярної машини.  
Праворуч: рівні вільної енергії  $G$  структурних станів у комплексах із різними лігандами

Приблизно за такою схемою працює, наприклад, актоміозинний комплекс у м'язових (і не тільки) клітинах (рис. 2.24).



**Рис. 2.24.** Спрощена схема роботи актоміозинового комплексу

Міозиновий філамент складається з молекул міозину – великого білка складної структури. Від основи цього філаменту відходять глобулярні домени міозину – головки (показано лише одну). Головка має спорідненість до АТР і містить АТРазний каталітичний центр. Зв'язування АТР фіксує головку в структурному стані, який не має спорідненості до філаменту, що складається з молекул іншого білка – актину. Після гідролізу АТР нові ліганди – ADP і неорганічний фосфат – фіксують інший стан: головка зв'язується з актином, відхиляючись ліворуч. Третій стан головки індукується окремою молекулою ADP після дисоціації фосфату: головка зміщується праворуч без втрати зв'язку з актином. У результаті актиновий філамент зсувається разом із головою. Далі ADP витісняється новою молекулою АТР і все повторюється. Актиновий філамент при дисоціації головки не повертається назад, оскільки фіксується іншими головками, що перебувають на інших стадіях свого робочого циклу. Інші приклади роботи молекулярних машин розглянуто в наступних розділах.

### **Контрольні запитання**

1. Напишіть стандартні скорочення для 20 амінокислот. На які дві групи їх можна поділити за фізико-хімічними властивостями?
2. Які властивості має пептидна група? Чи взаємодіє вона з водою?
3. Що таке вторинна структура білка? Укажіть типи й основні риси вторинних структур, що зустрічаються у глобулярних білках. Які взаємодії стабілізують регулярну вторинну структуру?
4. Які взаємодії є основною рушійною силою, що зумовлює укладання поліпептидного ланцюга в компактну глобулу?
5. Укажіть основні загальні риси структури глобулярних водорозчинних білків. Чому глобула завжди формується елементами регулярної вторинної структури?
6. Які взаємодії зумовлюють твердість білкової глобули? Яке значення має твердість білка для його функціонування?
7. Що таке структурний мотив? Укажіть основні структурні мотиви, що часто зустрічаються у глобулярних білках.

8. Як структура мембранних білків відрізняється від такої водорозчинних?

9. За рахунок чого стає можливим перемикання структури білка між двома конформаційними станами?

10. У чому полягає фізична природа ферментативного каталізу?

11. За яким принципом використовується вільна енергія гідролізу АТФ для реакцій синтезу?

12. Як використовується вільна енергія гідролізу АТФ для виконання роботи молекулярними машинами? Сформулюйте основні принципи роботи молекулярних машин.

## Рекомендована література

Branden, C.-I. and Tooze, J. (1999) *Introduction to protein structure*. New York : Garland Science.

Creighton, T. (1992) *Proteins: structure and molecular properties*. – New York : W.H. Freeman and Company.

Dill, K.A. and MacCallum, J.L. (2012) The protein-folding problem, 50 years on. *Science* 338, 1042–1046.

Elber, R. and Kirmizialtin, S. (2013) Molecular machines. *Curr. Opin. Struct. Biol.* 23, 206–211.

Kessel, A. and Ben-Tal, N. (2018) *Introduction to proteins: structure, function, and motion*. Boca Raton FL : CRC Press.

Lesk, A.M. (2001) *Introduction to protein architecture: the structural biology of proteins*. Oxford : Oxford University Press.

Mavroidis, C., Dubey, A. and Yarmush, M.L. (2004) Molecular machines. *Annu. Rev. Biomed. Eng.* 6, 363–395.

Nelson, D.L. and Cox, M.M. (2021) *Lehninger principles of biochemistry*. 8<sup>th</sup> ed. New York : W.H. Freeman and Company.

Spirin, A.A. (2002) Ribosome as a molecular machine. *FEBS Letters* 514, 2–10.

Wright, P.E. and Dyson, H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell. Biol.* 16, 18–29.

## Розділ 3

# ДНК

They are in you and me; they created us, body and mind;  
and their preservation is the ultimate rationale for our existence.  
*R. Dawkins. The selfish gene*

Другий після білків тип біополімерів, що буде в центрі уваги в цьому підручнику, – нуклеїнові кислоти, що складаються з нуклеотидів як мономерних одиниць. Серед двох типів нуклеїнових кислот ДНК, якій переважно присвячено цей розділ, займає особливе місце як носій спадкової інформації у більшості біологічних систем. Саме з ДНК зазвичай "усе розпочинається" у клітині. Особливості різноманітних молекул РНК будуть розглянуті в наступних розділах.

### Хімічна будова нуклеїнових кислот

#### Нуклеотиди

Мономерна одиниця нуклеїнової кислоти – *нуклеотид* – складається із трьох елементів: азотистої основи, пентозного цукру й залишку фосфорної кислоти.

*Азотисті основи* – гетероциклічні сполуки (рис. 3.1), у кільцях яких містяться карбон і азот, а всі зв'язки частково подвійні. До складу нуклеїнових кислот входять два типи азотистих основ: пурини (purines, загальноприйняте позначення R) – аденін (A) і гуанін (G); піримідини (pyrimidines, Y) – урацил (U), тимін (T), цитозин (C). Загальне позначення для всіх основ – N, часто використовують також позначення S (strong) для G або C і W (weak) для A або T.

Найсуттєвішими для структури нуклеїнових кислот та їхніх взаємодій з іншими молекулами є такі властивості азотистих основ:

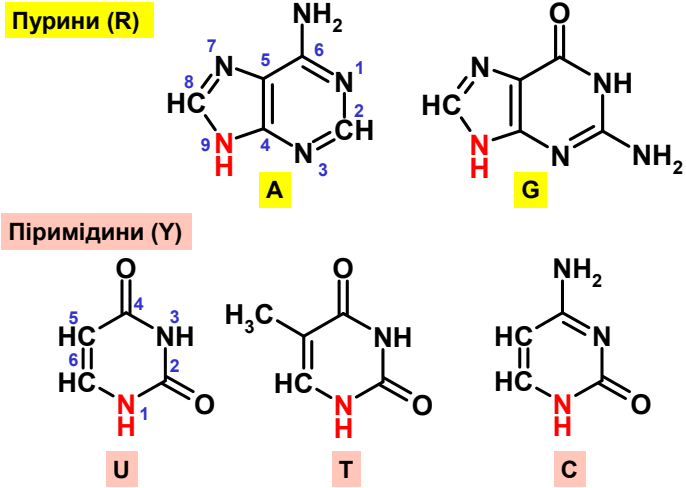
- Оскільки всі зв'язки в кільці частково подвійні, то азотисті основи є конформаційно жорсткими і *планарними* – усі атоми кільця лежать практично в одній площині.
- Атоми азоту кільця і приєднані до нього атоми оксигену є акцепторами, а NH і NH<sub>2</sub>-групи – донорами водневого зв'язку. Однак водночас площина кільця є гідрофобною, тому азотисті основи погано розчиняються у воді.

Один із атомів азоту кільця (виділено червоним на рис. 3.1) приєднується у складі нуклеотиду до карбону пентозного цукру *глікозидним* зв'язком (рис. 3.2). Цей атом пентози позначається як C1' (символ ' прийнято додавати, щоб відрізнити атоми фуранозного кільця пентози від атомів азотистої основи). Інші C'-атомі пентози нумеруються далі за порядком їхнього розташування (рис. 3.2). *До 3'-атома завжди приєднана ОН-група*. Пентоза, у складі якої ОН-група знаходиться також при 2'-атомі, називається *рибозою* (ribose). Пентоза іншого типу – *дезоксирибоза* (deoxyribose) – відрізняється лише заміною цієї ОН-групи на атом гідрогену.

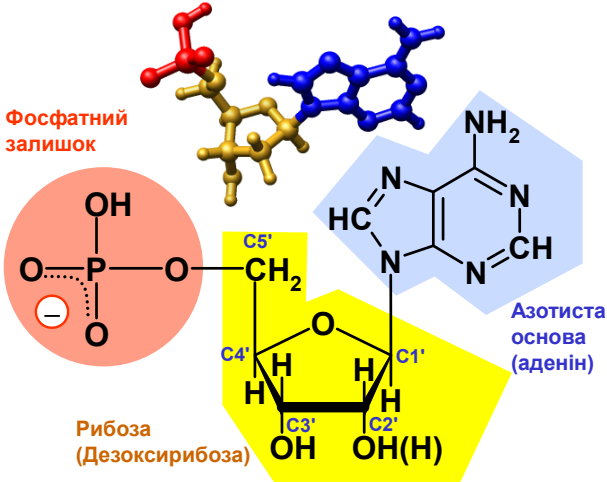
Сполука азотистої основи й пентози називається *нуклеозидом* (залежно від типу азотистої основи: аденозин, гуанозин, цитидин, тимідин, уридин) або ж, залежно від типу пентози, *рибо-* чи *дезоксирибонуклеозидом*. Рибонуклеозиди входять до складу РНК – рибонуклеїнової кислоти (ribonucleic acid, RNA), дезоксирибонуклеозиди – до складу ДНК – дезоксирибонуклеїнової кислоти (deoxyribonucleic acid, DNA). Інша хімічна різниця між ДНК і РНК стосується набору піримідинових азотистих основ: Т у ДНК замість U в РНК.

Фосфорилування ОН-групи при 5'-атомі пентози у складі нуклеозиду зумовлює утворення нуклеотиду (рис. 3.2). Найважливішою для нуклеїнових кислот властивістю фосфатного залишку є *негативний заряд*, який він має за нейтральних рН. Отже, нуклеотид є нуклеозидфосфатом або нуклеозидмонофосфатом (NMP). Зокрема нуклеотид, зображений на рис. 3.2, є

аденозинмонофосфатом (АМР); відповідні ди- і трифосфат (АДР і АТР) було розглянуто в розділі 2 (див. рис. 2.22).

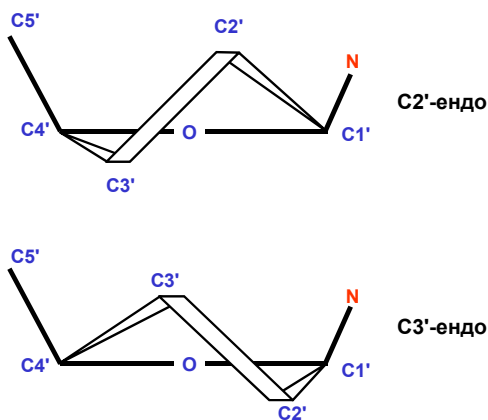


**Рис. 3.1.** Азотисті основи. Червоним позначено атом N, з'єднаний у нуклеотиді з атомом С1' пентози (див. рис. 3.2). Наведено стандартну нумерацію атомів у складі пуринових і піримідинових кілець



**Рис. 3.2.** Структура нуклеотиду

На відміну від азотистих основ, пентози й фосфатні залишки є полярними й добре взаємодіють із водою. Крім того, пентоза не містить подвійних зв'язків. Тому, на відміну від азотистої основи, у фуранозному кільці пентози можливе обертання навкруг зв'язків і воно не є планарним. У складі нуклеїнових кислот здебільшого зустрічаються дві конформації, схематично показані на рис. 3.3. Конформація *C2'-ендо* ( $C2'$ -атом виходить у бік  $C5'$ -атома із площини, заданої атомами  $C4'-O-C1'$ ) є переважною для дезоксирибози. Поява  $OH$ -групи замість  $H$  при  $C2'$ -атомі робить цю конформацію неможливою, тому для рибози переважною стає конформація *C3'-ендо*. Перемикання цих двох конформацій суттєво змінює геометрію інших зв'язків у складі нуклеотиду й тому лежить в основі перетворень між структурними формами нуклеїнових кислот (див. нижче).



**Рис. 3.3.** Дві переважні конформації пентозного кільця цукру у складі нуклеїнових кислот

Отже, нуклеотид складається із двох елементів, які розрізняються за полярністю і конформаційною рухливістю:

- 1) полярна й гнучка частина – пентозний цукор із фосфатним залишком, причому останній має негативний заряд;
- 2) жорстка планарна азотиста основа, яка загалом є неполярною, хоча й містить донорні й акцепторні групи.

## Полінуклеотидний ланцюг

Дві хімічні групи – ОН-група при 3'-атомі пентози одного нуклеотиду та фосфат при 5'-атомі іншого – використовуються для утворення *фосфодієфірного зв'язку* між нуклеотидами (рис. 3.4).

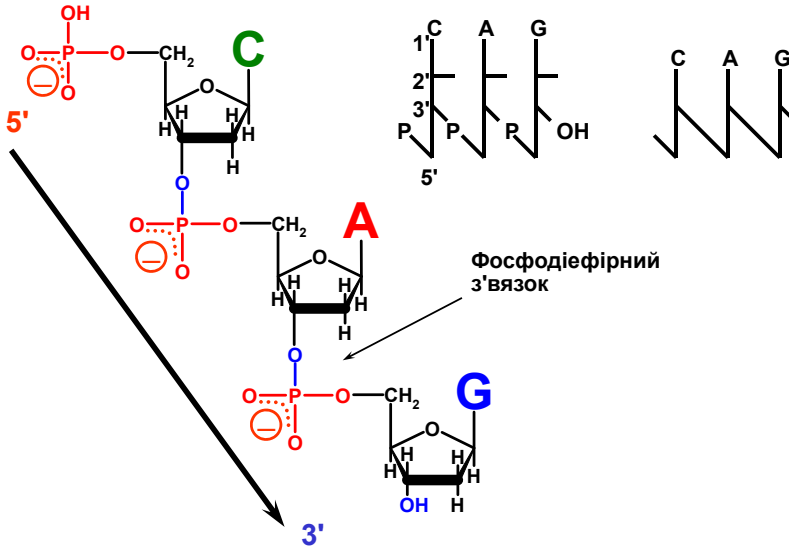


Рис. 3.4. Полінуклеотидний ланцюг.

Угорі праворуч – два варіанти схематичного зображення ланцюга: вертикальна лінія позначає пентозу, P – фосфатний залишок

Відповідно, як і поліпептидний (розділ 2), полінуклеотидний ланцюг має напрямок. На одному його кінці залишається 5'-фосфат (5'-кінець), на іншому – 3'ОН-група (3'-кінець). За умовчанням послідовності нуклеотидів записують у напрямку 5' → 3', у тому самому напрямку відбувається синтез усіх нуклеїнових кислот. Слід також розуміти, що конформаційні особливості ланцюга залежать від напрямку, у якому розташовані сусідні нуклеотиди. Наприклад, послідовність СА має інші особливості, ніж послідовність АС. Загалом у складі ланцюга може бути 16 типів таких динуклеотидних контактів (16 можливих комбінацій із чотирьох по два).

Отже, полярний остов полінуклеотидного ланцюга складається із фосфатних залишків і пентоз, що чергуються (*цукрофосфатний остов*). Від нього відходять азотисті основи як бокові залишки.

## Нуклеази

Деградація нуклеїнових кислот відбувається при різноманітних функціональних процесах шляхом гідролізу фосфодієфірного зв'язку в реакції, яка каталізується спеціальними ферментами – *нуклеазами*. Нуклеази поділяють на різні типи за певними ознаками:

- РНКази та ДНКази є специфічними до відповідних нуклеїнових кислот.
- *Екзонуклеази* відщеплюють нуклеотиди один за одним з кінців ланцюга. Відповідно розрізняють 5'- та 3'-екзонуклеази. *Ендонуклеази* гідролізують фосфодієфірний зв'язок усередині ланцюга.
- Серед ендонуклеаз, які мають за свій субстрат дволанцюгову молекулу ДНК, є такі, що роблять односторонній розріз – *нік* (*nick*), залишаючи інший ланцюг інтактним. Інші здійснюють дволанцюгові розрізи.
- Серед ендонуклеаз є неспецифічні щодо послідовності нуклеотидів. Однак є також велика група бактеріальних ферментів – *рестриктаз*, які діють лише на певні невеличкі (частіше чотири або шість нуклеотидів) елементи послідовності. Рестриктази захищають бактеріальну клітину від чужорідної ДНК бактеріофагів і широко використовуються як інструменти дослідження в молекулярній біології (див. розділ 11).

Реакція гідролізу фосфодієфірного зв'язку енергетично вигідна. Із цього випливає, що нуклеотид (нуклеозидмонофосфат) не може бути використаним для синтезу нуклеїнової кислоти, натомість до реакцій синтезу залучаються відповідні нуклеозидтрифосфати (розділи 5, 9).

## Подвійна спіраль

### Стабілізація подвійної спіралі

Два полінуклеотидні ланцюги (ДНК, РНК або гібридні) можуть об'єднуватись у єдину дволанцюгову структуру (*дуплекс*), схему якої зображено на рис. 3.5. Таке об'єднання відбувається за жорсткої умови: певні азотисті основи мають розташовуватися одна проти одної – А проти Т (чи U), G проти C. Цей *принцип комплементарності*, сформульований Уотсоном і Кріком (James D. Watson, Francis H. C. Crick), зумовлений утворенням специфічних водневих зв'язків між названими основами: два зв'язки в парі А-Т, три в парі G-С. Два ланцюги при цьому спрямовані в різні боки (*є антипаралельними*), цукрофосфатні остови розташовані зовні, пари основ – усередині структури.

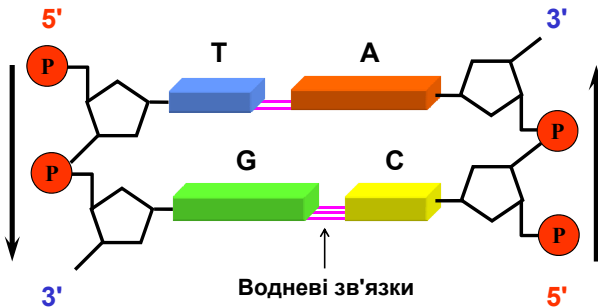
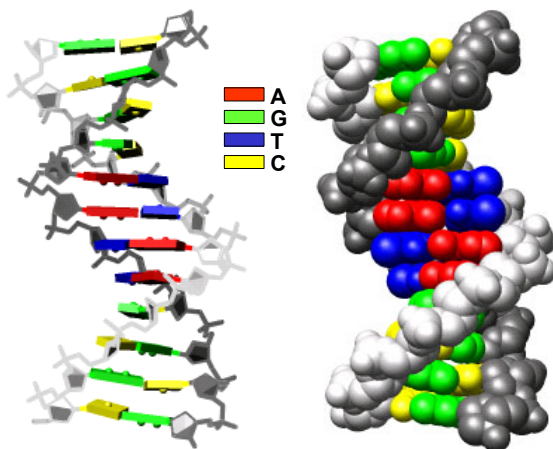


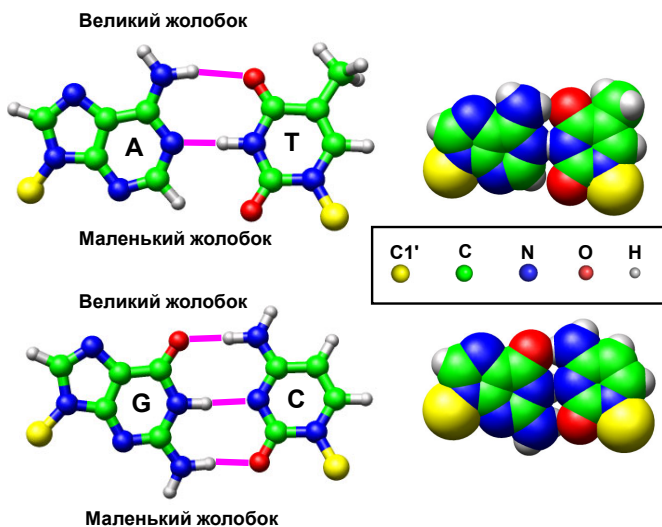
Рис. 3.5. Схема об'єднання двох динуклеотидів у дволанцюгову структуру

У складі такого дволанцюгового комплексу неполярні площини сусідніх пар основ мають наблизитися одна до одної, відносна гнучкість остовів дозволяє це зробити простим шляхом – закрутити два ланцюги один навкруг одного в подвійну спіраль (рис. 3.6). Усередині формується *стос* (stack) щільно укладених одна на одну пар основ, захищений на поверхні спіралі закрученими один навкруг одного полярними цукрофосфатними остовами. Між остовами на поверхні спіралі утворюються два *жолобки* різного розміру – великий і маленький (*major / minor groove*), у яких опиняються донорні й акцепторні групи азотистих основ (рис. 3.7). У великому жолобку містяться по два акцептори (O, N) і одному донору (NH<sub>2</sub>) водневого

зв'язку кожної пари, у маленькому – два акцептори пари А-Т, два акцептори й один донор пари G-C.



**Рис. 3.6.** Два варіанти зображення подвійної спіралі ДНК (кристалічна структура додекамеру, 355D)



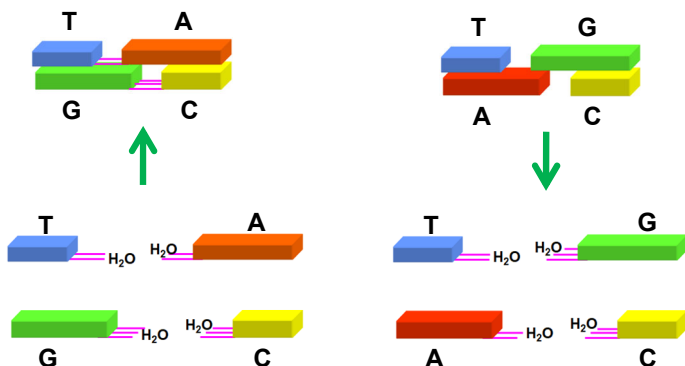
**Рис. 3.7.** Комплементарні пари основ у складі ДНК з водневими зв'язками між основами. Праворуч атоми тих самих пар зображені відповідно до їхніх ван-дер-ваальсових радіусів

Отже, подібно до білків (розділ 2), *гідрофобні взаємодії* (розділ 1) між площинами пар азотистих основ зумовлюють формування подвійної спіралі. При цьому реалізується щільна упаковка – утворюються *вандерваальсові контакти* між сусідніми парами. Велика кількість таких контактів уздовж осі спіралі робить їх суттєвим фактором стабілізації дуплекса. Вандерваальсові взаємодії між парами основ є за характером орієнтаційними та, унаслідок нерівномірності розподілу електронної щільності по кільцях основ, диполь-дипольними (див. розділ 1). Такі взаємодії складного характеру (гідрофобний ефект + вандерваальсові взаємодії) у складі стосу мають спеціальну назву – *стекінг-взаємодії*.

***Стекінг-взаємодії є головним, а точніше, практично єдиним, фактором стабілізації подвійної спіралі.*** Функція комплементарних водневих зв'язків між основами, які самі по собі енергетично не дуже важливі (водневі зв'язки основ із водою при розходженні ланцюгів практично такі самі в енергетичному сенсі), полягає в тому, що вони є необхідною передумовою утворення стабільного стосу. Це знову цілком аналогічно ролі водневих зв'язків між пептидними групами у складі регулярної вторинної структури глобулярних білків (розділ 2). Донорні й акцепторні групи азотистих основ прагнуть утворити водневі зв'язки (енергія водневого зв'язку надто висока, щоб його втрачати). Опиняючись усередині неполярного стосу, вони стають недоступними для води, тому єдиний шанс зберегти водневі зв'язки – утворити їх із відповідними групами іншої основи. А це можливо лише за умови комплементарності. Відповідно за відсутності комплементарності азотисті основи зберігають водневі зв'язки з водою, що унеможливорює формування подвійної спіралі (рис. 3.8).

Ефективність стекінг-взаємодій залежить від типу контактуючих пар основ, а також їхньої взаємної орієнтації та порядку розташування в ланцюгах дуплекса – типу динуклеотидного контакту вздовж осі подвійної спіралі. Замість 16 типів динуклеотидних контактів у одноланцюговій нуклеїновій кислоті у подвійній спіралі залишається 10 – унаслідок комплементарності ланцюгів певні контакти еквівалентні. Наприклад, контакт AG в одному ланцюзі автоматично означає, що в комплементарному ланцюзі реалізується контакт СТ (слід пам'ятати, що ланцюги є

антипаралельними і послідовність нуклеотидів читається в напрямку 5' → 3'; усі динуклеотидні контакти подвійної спіралі наведено в табл. 3.2). Для певних контактів, наприклад пурин-піримідинових порівняно з піримідин-пуриновими або між сусідніми GC-парами порівняно з AT-парами, стекінг-взаємодії більш ефективні. Відповідно *стабільність подвійної спіралі залежить від послідовності пар основ*.



**Рис. 3.8.** Порівняння подвійної спіралі (зверху) і одноланцюгового (денатурованого) стану (знизу) за комплементарності між азотистими основами (ліворуч) та її відсутності (праворуч). Стрілки вказують на стан, що є більш вигідним енергетично за фізіологічних умов

Утворення подвійної спіралі діє проти конформаційної ентропії: розходження двох ланцюгів супроводжується значним зростанням їхньої конформаційної свободи (невпорядкованості). Отже, *стабільність подвійної спіралі залежить від температури*: при зростанні температури ентропійна компонента вільної енергії збільшується (рівняння (1.2)) і відбувається розходження ланцюгів – *плавлення* подвійної спіралі. Зрозуміло, що температура плавлення, тобто температура, за якої 50 % пар основ стають розплавленими, є мірою стабільності й залежить від послідовності пар основ. Оскільки контакти між сусідніми GC-парами більш стабільні, то температура плавлення прямо пропорційна вмісту GC-пар у складі ДНК.

Суттєвий дестабілізуючий внесок у вільну енергію подвійної спіралі роблять також електростатичні взаємодії. Оскільки кожен фосфатний залишок несе на собі негативний заряд (рис. 3.4), то молекула ДНК є гігантським поліаніоном із високою поверхневою щільністю зарядів. У розчині ці заряди суттєво нейтралізовані протиіонами – неорганічними катіонами: навколо ДНК утворюється іонна атмосфера з доволі високою локальною концентрацією катіонів, тобто в системі виникає нерівномірний їх розподіл, унаслідок чого зростає відповідна ентропійна компонента вільної енергії. Цей невідгідний ефект буде тим меншим, чим вищою буде концентрація солі в розчині, тобто різниця між концентрацією катіонів поблизу ДНК і на віддаленні від неї. Отже, *стабільність подвійної спіралі* (і температура плавлення) *залежить від іонної сили розчину* – зростає при підвищенні концентрації солі.

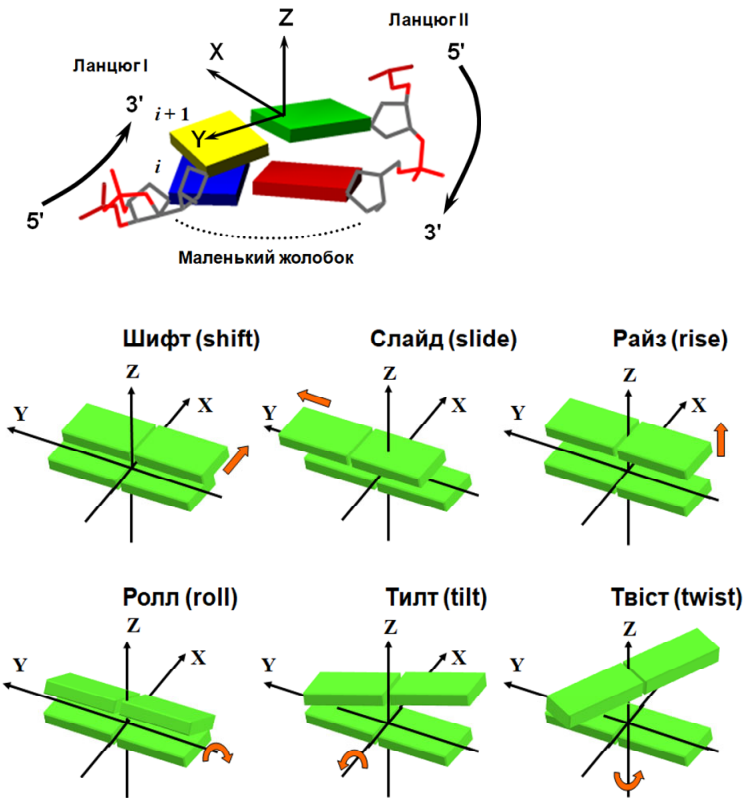
За фізіологічних умов подвійна спіраль є цілком стабільною структурою, і саме в цій формі ДНК (і значна частина комплементарних ділянок РНК) існує в живих системах. Розходження ланцюгів, яке має тимчасово відбуватися при різноманітних функціональних процесах, потребує виконання роботи молекулярними машинами з використанням енергії гідролізу АТР.

### **Конформаційні параметри подвійної спіралі**

Як для будь-якого іншого полімеру, конформаційна рухливість нуклеїнових кислот реалізується завдяки зміні кутів обертання навкруг ковалентних зв'язків. У складі подвійної спіралі ці обертальні рухи обмежені міцними стекінг-взаємодіями та комплементарними водневими зв'язками. Дозволені обертання навкруг різних зв'язків зумовлюють урешті-решт зміну взаємної орієнтацій основ і пар основ. Саме набором параметрів, що описують такі відносні орієнтації, прийнято характеризувати конформації подвійних спіралей.

В основі номенклатури конформаційних параметрів подвійних спіралей нуклеїнових кислот лежить локальна декартова система координат, що ставиться у відповідність до кожної пари основ (рис. 3.9): початок координат знаходиться в геометричному центрі пари; площина  $X$ – $Y$  збігається із площиною пари, вісь

У паралельна лінії, яка з'єднує С1'-атоми, і спрямована на ланцюг, який береться за перший; вісь Z спрямована на площину наступної (у напрямку 5'-3' по першому ланцюгу) пари основ; напрямок осі X обирається за правилами утворення правої декартової системи координат – жолобок, на який вона спрямована, позначається як "великий". Насправді терміни "великий / маленький жолобок" не обов'язково мають буквального значення (див. нижче).



**Рис. 3.9.** Дві сусідні пари основ у молекулі ДНК, локальна декартова система координат, пов'язана з однією з них, і шість параметрів, що описують взаємне розташування двох сусідніх пар

Для опису відносного положення двох сусідніх пар необхідно й достатньо шести параметрів: три зсуви однієї пари відносно іншої по трьох осях (будь-який зсув можна розкласти на три компоненти) і три кути обертання пари відносно сусідньої навколо трьох осей (рис. 3.9):

- шифт (shift) характеризує зсув у напрямку одного із жолобків;
- слайд (slide) – зсув у напрямку цукрофосфатного остова одного з ланцюгів;
- райз (rise) – відстань між парами основ уздовж осі Z;
- кут ролла (roll) характеризує локальний вигин подвійної спіралі в напрямку одного із жолобків;
- кут тилта (tilt) – вигин у напрямку цукрофосфатного остова;
- твіст (twist) є кутом обертання навкруг осі Z, який характеризує ступінь спірального закручення (для правої спіралі прийнято використовувати позитивні значення твіста).

Аналогічно шість параметрів характеризують відносне положення основ, що належать до однієї пари: зсув уздовж осі X і обертання навколо неї (*shear* і *buckle*), зсув і обертання відносно осі Y (*stretch* і *propeller*) і відносно осі Z (*stagger* і *opening*). Як само відбуваються рухи однієї основи відносно іншої, легко зрозуміти з рис. 3.9.

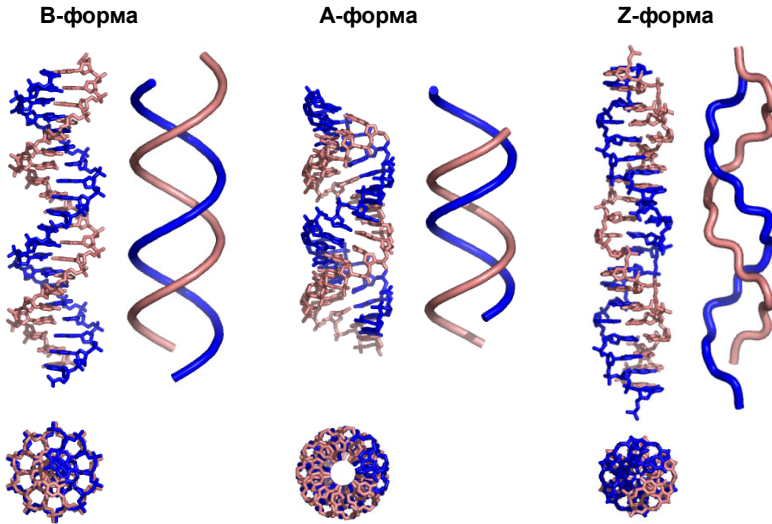
## Структурні форми ДНК

ДНК із майже будь-якою послідовністю пар основ може існувати у двох структурних формах, що позначаються як А і В. Деякі усереднені параметри двох форм у кристалах ДНК наведено в табл. 3.1, структури показано на рис. 3.10.

Обидві *спіралі є правими*. Основні відмінності між двома формами ДНК зумовлені різною конформацією цукру – С2'- та С3'-ендо в В- і А-формах, відповідно. Саме внаслідок цього в А-формі порівняно з В-формою:

- скорочується відстань між фосфатними залишками;
- зменшується ступінь правого спірального закручення;
- кожна пара основ суттєво нахилиється до осі спіралі (у В-формі пари майже перпендикулярні осі);

- унаслідок нахилу скорочується відстань між сусідніми парами основ уздовж осі спіралі (проекція райза на вісь спіралі);
- кожна пара суттєво зсунута до периферії спіралі в А-формі (вісь спіралі проходить майже крізь центр пари у В-формі);
- "великий" і "маленький" жолобки є справді великим і маленьким для В-форми, в А-формі їхні відносні розміри міняються місцями.



**Рис. 3.10.** Структури трьох форм ДНК у двох проєкціях, зображено також хід цукрофосфатних ланцюгів. Усі молекули містять однакову кількість пар основ. Зображено ідеалізовані структури (на основі усереднених структурних параметрів), створені за допомогою програмного пакету 3DNA (<http://x3dna.org/>)

Крім того, є різниця у конформаційній рухливості: А-форма однорідніша за структурою, а В-форма допускає ширші варіації параметрів залежно від послідовності. Узагалі частіше кажуть не про форми ДНК, а про А- і В-родини структурних форм.

За фізіологічних умов у розчині ДНК існує тільки у В-формі, перехід в А-форму здійснюється за екстремальних умов *in vitro* – за певних концентрацій солі та спиртів у розчині, при знижен-

ні вологості у фібрилах. Однак це не означає, що А-форма не є фізіологічною. Через те, що рибоза не може набувати С2'-ендо конформації, усі *подвійні спіралі РНК існують у А-формі* за фізіологічних умов. З тієї самої причини набувають А-форми гібридні подвійні спіралі РНК-ДНК, які, зокрема, виникають тимчасово під час транскрипції. Крім того, ДНК може переходити в А-форму або наближену до неї в комплексах із деякими білками.

**Таблиця 3.1. Порівняння між А- і В-формами подвійних спіралей ДНК**

	<b>А-форма</b>	<b>В-форма</b>
Твіст $\Omega$ (°)	+31,1	+36,0
Кількість пар основ на виток спіралі (360°/ $\Omega$ )	11,6	10,0
Слайд (Å)	-1,5	0,2
Ролл (°)	8,0	0,6
Райз (Å)	3,3	3,3
Відстань між сусідніми парами вздовж осі спіралі (Å)	2,8	3,3
Конформація дезоксирибози	С3'-ендо	С2'-ендо
Відстань між сусідніми фосфатами по ланцюгу (Å)	5,9	7,0
Великий жолобок: глибина (Å), ширина (Å)	13,5 2,7	8,5 11,7
Маленький жолобок: глибина (Å), ширина (Å)	2,8 11,0	7,5 5,7

Ще одна регулярна форма подвійної спіралі реалізується тільки для альтернувальних пурин-піримідинових послідовностей, передусім для  $\text{poly}(\text{GC})$ , коли G і C чергуються по ланцюгу. Це так звана *Z-форма* (рис. 3.10). Її назва пов'язана з тим, що лінія, яка з'єднує фосфатні залишки у складі Z-форми, нагадує зигзаг. Найбільш вражаюча особливість цієї форми ДНК полягає в тому, що це – *ліва спіраль* із приблизно 12 парами основ на виток. Альтернувальні послідовності G та C є у природних ДНК, але перехід у Z-форму відбувається *in vitro* за умов, які є дуже далекими від фізіологічних – при концентрації NaCl 2,5 моль/л. Проте у циркулярних ДНК (див. нижче) Z-форма може бути

реалізована за фізіологічних умов. Біологічне значення Z-форми залишається не зовсім зрозумілим, хоча знайдені білки, які мають високу спорідненість саме до неї, тобто можуть індукувати B → Z перехід *in vivo*.

### Конформація В-ДНК

Основною структурною формою ДНК, у якій вона існує у живих системах, є В-форма. Структура цієї форми не є абсолютно регулярною. Конформаційні особливості й особливості стекингу в різних контактах між парами основ зумовлюють суттєву залежність локальної конформації подвійної спіралі від послідовності, тобто можна сказати, що послідовність несе інформацію про структурні особливості ДНК (подібно до того, як амінокислотна послідовність зумовлює просторову структуру білка).

Щоб проілюструвати залежність конформації ДНК від послідовності, у табл. 3.2 наведено значення кута твіста (див. рис. 3.9) для 10 типів контактів між парами основ у кристалах вільної ДНК і білково-нуклеїнових комплексів.

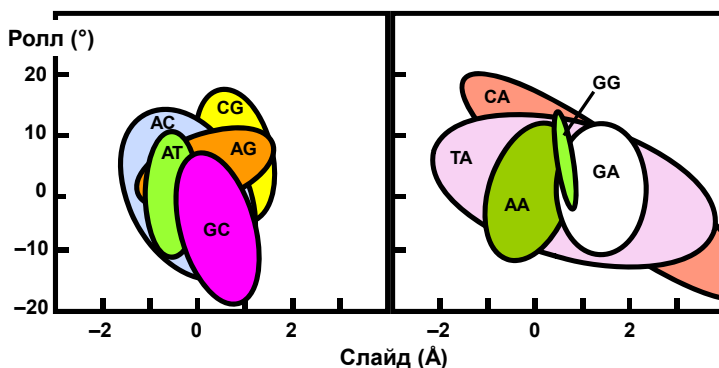
**Таблиця 3.2.** Твіст ДНК для 10 типів контактів між парами основ (за даними Olson et al., 1998)

Тип контакту	Середній кут твіста ± стандартне відхилення (°)	
	Кристали вільної ДНК	Білково-нуклеїнові комплекси
AA/TT	35,5 ± 3,6	35,1 ± 3,9
AG/CT	30,6 ± 4,7	31,9 ± 4,5
GA/TC	39,6 ± 3,0	36,3 ± 4,4
GG/CC	35,3 ± 4,9	32,9 ± 5,2
AC/GT	33,1 ± 4,6	31,5 ± 4,2
AT	31,6 ± 3,4	29,3 ± 4,5
GC	38,4 ± 3,0	33,6 ± 4,7
CA/TG	37,7 ± 9,3	37,3 ± 6,5
CG	31,3 ± 4,7	36,1 ± 5,5
TA	43,2 ± 5,5	37,8 ± 5,5

Твіст у комплексах є дуже близьким до значення твіста вільної ДНК у розчині за фізіологічних умов – середнє значення для

всіх контактів відповідає 10 парам основ на виток спіралі у кристалах і 10,5 парам основ у комплексах із білком. Як видно з таблиці, твіст варіює в доволі широких межах – ступінь локального спірального закручення ДНК суттєво змінюється залежно від послідовності пар основ. Крім того, варіюють стандартні відхилення від середнього значення твіста. Останній параметр показує, наскільки легко значення твіста може відхилитися від переважного середнього значення, тобто характеризує конформаційну рухливість контакту.

Інші конформаційні параметри подвійної спіралі (як середні значення, так і ширина розподілу) також залежать від послідовності пар основ. При цьому 6 конформаційних параметрів, наведених на рис. 3.9, не є незалежними один від одного – їхні зміни взаємно пов'язані. Для прикладу на рис. 3.11 показано співвідношення між роллом і слайдом (доступний сегмент конформаційного простору) для 10 контактів. Загалом конформаційна рухливість негативно корелює з ефективністю стекингу: вона вища для піримідин-пуринових і АТ-збагачених контактів.



**Рис. 3.11.** Усередині зафарбованих зон – доступні для кожного з 10 контактів значення двох конформаційних параметрів для кристалів ДНК (за даними Olson et al., 1998)

Характер залежності конформації ДНК від послідовності насправді ще складніший: конформаційні особливості визначаються не тільки найближчими сусідами – типом контакту між парами основ, але й контекстом послідовності, у якій він міститься.

Отже, хоча конструкція молекули ДНК здається простішою за білкову, на сьогодні неможливо точно передбачити деталі структури ДНК, виходячи з її послідовності. Залежні від послідовності особливості структури подвійної спіралі та потенціал щодо конформаційних змін – основа механізму специфічного впізнання послідовностей ДНК білками.

## **Білково-нуклеїнові взаємодії**

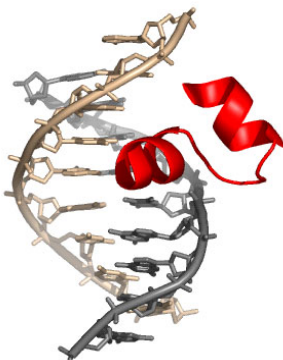
Нуклеїнові кислоти, зокрема й ДНК, практично ніколи не існують *in vivo* у вільному вигляді – вони постійно взаємодіють із численними білками. Білки, що взаємодіють із ДНК, поділяють на три функціональні класи: структурні білки визначають структурну організацію ДНК у клітинах; регуляторні білки залучені до регуляції різноманітних процесів (найбільша група серед них – транскрипційні фактори); ферменти каталізують різноманітні хімічні реакції, що здійснюються на ДНК і зв'язаних із нею білках.

### **Структурна класифікація білків, які взаємодіють із ДНК**

Так само, як більшість глобулярних білків можна вписати в обмежений набір стандартних укладок глобули (розділ 2), структурні елементи білків, які безпосередньо взаємодіють із ДНК, можна поділити на обмежену кількість стандартних типів – ДНК-зв'язувальних структурних мотивів.

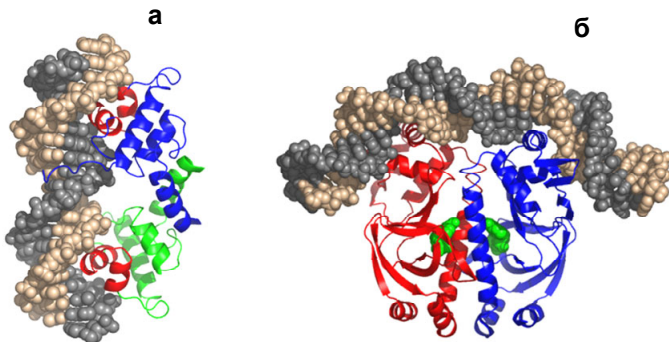
**Мотив НТН** (*helix-turn-helix*, спіраль – поворот – спіраль), приклад якого зображено на рис. 3.12, є одним із найбільш розповсюджених. Мотив складається із двох  $\alpha$ -спіралей, часто майже перпендикулярних одна до одної, з'єднаних короткою перемичкою. Одна із цих спіралей (яка впізнає нуклеотидну послідовність) виконує функцію *головки, що читає*, взаємодіючи з екзоциклічними групами азотистих основ у великому жолобку подвійної спіралі. Орієнтація цієї спіралі щодо ДНК може варіювати для різних білків від приблизно перпендикулярної відносно осі подвійної спіралі до паралельної цукрофосфатному остову. Друга  $\alpha$ -спіраль часто утворює допоміжні контакти із

цукрофосфатним остовом. Взаємодія  $\alpha$ -спіралі з великим жолобком часто зустрічається і в інших ДНК-зв'язувальних мотивах. Причина полягає в майже точній просторовій відповідності між цими двома елементами – можна сказати, що  $\alpha$ -спіраль є просторово комплементарною великому жолобку.



**Рис. 3.12.** НТН-мотив репресора бактеріофага  $\lambda$  у комплексі з ДНК (1LMB)

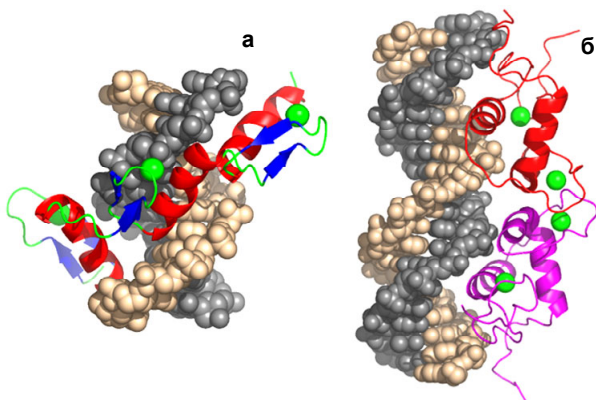
Двоспіральний НТН-мотив утримується у структурі того чи іншого структурного домену, наприклад у випадку репресора фага  $\lambda$ , у структурі  $\alpha$ -спірального пучка (рис. 3.13, *a*). Часто білки, що містять НТН-мотив, взаємодіють із ДНК у вигляді гомодимеру (дві однакові субодиниці), це особливо характерно для прокаріотичних білків. Тоді дві однакові "читаючі головки" впізнають дві однакові послідовності, що симетрично розташовані в сайті взаємодії. Інший приклад такого самого типу – білок активації катаболізму (CAP – Catabolite Activator Protein) *Escherichia coli* (рис. 3.13, *b*). При взаємодії гомодимеру CAP зі специфічним сайтом відбувається доволі значне вигинання ДНК у бік білка (подвійна спіраль "огортає" білок). Спорідненість білка до цього сайту залежить від ліганду – циклічного аденозин-монофосфату (сАМР), у складі якого 5'-фосфат з'єднаний внутрішнім ковалентним зв'язком із 3'-ОН-групою рибози. Зв'язування сАМР індукує структурні перебудови в молекулі білка (див. розділ 2), у результаті чого змінюється робоча поверхня і білок набуває здатності впізнавати специфічну послідовність.



**Рис. 3.13.** Комплекси з ДНК гомодимерів репресора бактеріофага  $\lambda$  (а, 1LMB) і білка CAP *E. coli* (б, 1CGP, зеленим кольором позначено молекули cAMP)

Розповсюджений клас білкових мотивів, що взаємодіють із ДНК, – **Zn-координуючі мотиви**. Це дуже невеличкі структурні елементи, які складаються з обмеженої, недостатньої для формування жорсткої глобули кількості сегментів вторинної структури. Іон цинку надає жорсткості такому мотиву, утворюючи координаційні зв'язки із чотирма амінокислотними залишками, частіше Cys та His. Координаційний зв'язок формується неподіленою парою електронів атомів O, N або S, яка узагальнюється з атомом металу шляхом часткового "перетікання" на його низьку незаповнену орбіталь. Чотири такі зв'язки створюють жорсткий каркас, який утримує білкову поверхню, що має взаємодіяти з ДНК.

Один зі структурних мотивів такого типу, що часто зустрічається, – **цинковий палець** (Zn finger, рис. 3.14, а). Він має дуже просту будову: одна  $\alpha$ -спіраль і маленький  $\beta$ -шар із двох  $\beta$ -ділянок; одна з них і  $\alpha$ -спіраль містять залишки, що утворюють координаційні зв'язки із Zn. Зазвичай кілька таких пальців (три на рис. 3.14, а), з'єднаних перемичкою, спіралью обгортають великий жолобок.



**Рис. 3.14.** Три цинкові пальці у складі ДНК-зв'язувального білка (а, 1ZAA) і ДНК-зв'язувальний домен (димер) глюкокортикоїдного рецептора (б, 1GLU). Зелені сфери – іони Zn

Інший приклад – ДНК-зв'язувальний домен гормонового рецептора (рис. 3.14, б). Гормонові рецептори – це еукаріотичні фактори транскрипції, здебільшого гомодимери, які набувають спорідненості до певних елементів послідовності ДНК після зв'язування з білком стероїдного гормону (розділ 6). ДНК-зв'язувальний домен має дві пари структурних елементів, а саме  $\alpha$ -спіраль – суміжна нерегулярна петля; кожна пара координує іон Zn. Одна пара елементів бере участь у димеризації, інша – у взаємодії з ДНК. Зустрічаються також інші Zn-координуючі мотиви, більшість яких, як і розглянуті, реалізують взаємодію  $\alpha$ -спіралі з великим жолобком.

Ще один мотив, який взаємодіє з великим жолобком ДНК за рахунок  $\alpha$ -спіралей – *лейциновий zipper* (Leu zipper, рис. 3.15). Він складається із двох довгих  $\alpha$ -спіралей (належать різним поліпептидним ланцюгам), кожна з яких можна розділити на дві частини. Одна частина обох спіралей здійснює димеризацію – за рахунок гідрофобних взаємодій між неполярними амінокислотними залишками (часто Leu, звідки й назва мотиву) утворюється трохи закручена подвійна спіраль (coiled coil). Друга частина кожної  $\alpha$ -спіралі взаємодіє з великим жолобком ДНК.

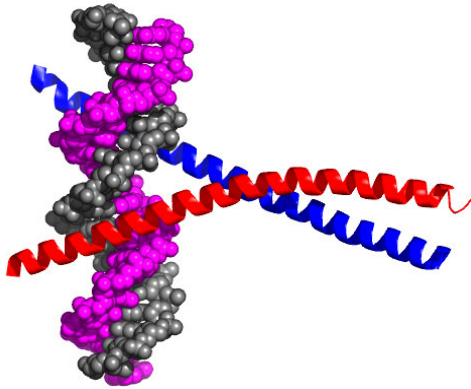


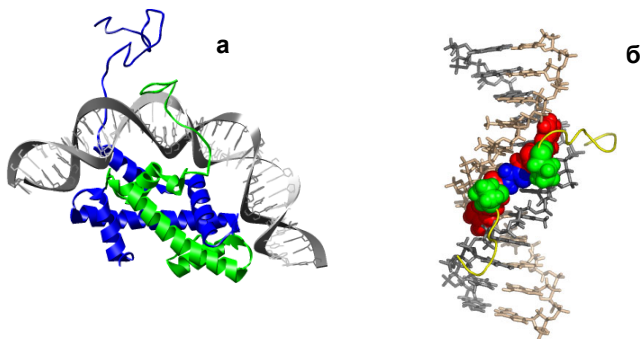
Рис. 3.15. Лейциновий зіпер (1NWQ)

Варіацією лейцинового зіпера є мотив *спіраль – петля – спіраль* (helix-loop-helix), різниця полягає лише в тому, що кожна довга  $\alpha$ -спіраль розділена на дві коротші, з'єднані петлею: одна пара спіралей здійснює димеризацію, інша – взаємодіє з ДНК.

Наведені приклади не означають, що тільки  $\alpha$ -спіраль використовується для взаємодії білків із великим жолобком ДНК. У великий жолобок добре вкладається також  $\beta$ -шар із двох  $\beta$ -ділянок, саме він є взаємодіючим елементом для білків типу метіонінового репресора. Численний клас білків становить родину імуноглобуліноподібних транскрипційних факторів (глобула має укладку імуноглобулінового типу, див. рис. 2.13, б), які реалізують взаємодії перемичок між ділянками  $\beta$ -структури з великим жолобком. Отже, хоча  $\alpha$ -спіраль використовується частіше, реалізуються і всі інші можливості.

Усе сказане не означає, що білки взаємодіють із ДНК тільки через великий жолобок (хоча через великий – частіше). На рис. 3.16 зображено приклади взаємодії перемичок між  $\alpha$ -спіралями (а) і витягнутого поліпептидного ланцюга (б) з маленьким жолобком. У першому випадку – димер гістонів у складі нуклеосоми (див. розділ 4) – маємо ще один структурний мотив, який зустрічається також в інших ДНК-зв'язувальних білках, а саме гістонову укладку: одна довга  $\alpha$ -спіраль фланкована двома

короткими. Димер двох таких укладок взаємодіє із  $\sim 2,5$  витками подвійної спіралі. У випадку гістонів ця взаємодія неспецифічна: утворюються електростатичні контакти позитивно заряджених амінокислотних залишків із фосфатами ДНК з боку маленького жолобка.

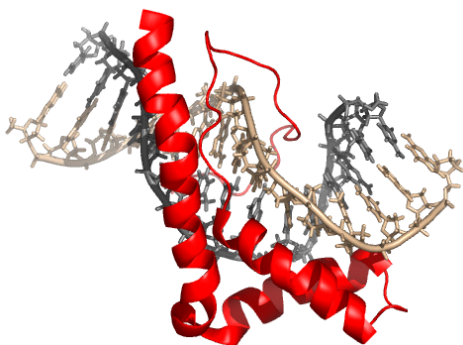


**Рис. 3.16.** Димер гістонів H3-H4 (а, 1A0I) і АТ-гак у складі білка HMG-A (б, 2EZD)

Другий приклад (рис. 3.16, б) стосується білків HMG-A – білків групи високої рухливості (High Mobility Group) типу А (HMG – історична назва для білків різних структурних класів, мається на увазі рухливість при електрофорезі). Білки HMG-A взагалі не мають глобулярної структури та є непорядкованими. Вони містять 3 пентапептидні елементи (Pro-Arg-Gly-Arg-Pro), що називаються *АТ-гаками* (AT-hook). Такий гак добре укладається у звужений маленький жолобок невеликої ділянки подвійної спіралі, збагаченої АТ-парами (тобто АТ-гак має специфічність до невеликих АТ-збагачених сайтів, завдяки чому він і отримав свою назву).

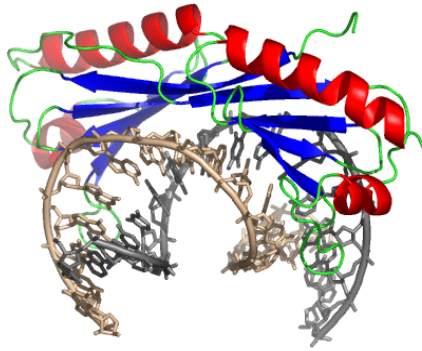
Наступні приклади стосуються дуже важливих випадків взаємодії елементів регулярної вторинної структури білків із маленьким жолобком. Оскільки  $\alpha$ -спіраль і дволанцюговий  $\beta$ -шар добре укладаються у великий жолобок, то це автоматично означає, що в маленькому жолобку їм не вистачає місця. Отже, така взаємодія буде можливою тільки за умови суттєвої деформації подвійної спіралі ДНК (розширення жолобка). Цей ефект має

важливі функціональні наслідки. На рис. 3.17 зображено доволі розповсюджений мотив – *HMG-бокс*, який, зокрема, входить до складу білків HMG-B (ще один клас білків HMG). Мотив складається із трьох  $\alpha$ -спіралей, дві з них (коротші) вбудовуються у маленький жолобок. Це супроводжується розкрученням подвійної спіралі й вигинанням на  $\sim 80^\circ$  у протилежний щодо білка бік. Серед HMG-боксів є як специфічні до певних послідовностей пар основ, так і неспецифічні.



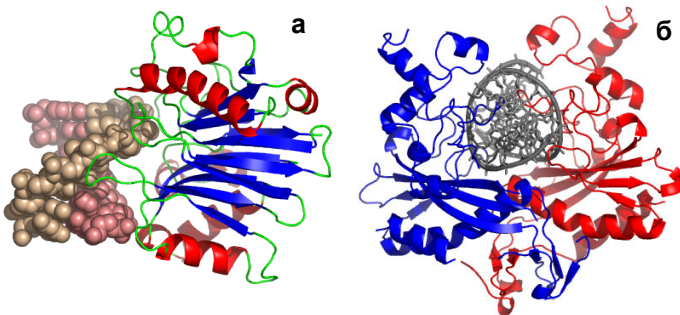
**Рис. 3.17.** HMG-бокс у складі білка HMG-B (1J5N)

У складі білка TBP (TATA-Box Binding Protein) – важливого елемента ініціації транскрипції в еукаріотів – доволі широкий  $\beta$ -шар взаємодіє з маленьким жолобком подвійної спіралі в зоні TATA-боксу – регуляторного елемента послідовності ДНК (розділ 5). Наслідком цієї взаємодії також є значна деформація подвійної спіралі з її розкрученням і значним вигинанням у протилежний від білка бік (рис. 3.18). Вигин, як і у випадку HMG-боксу, підсилюється інтеркаляцією двох гідрофобних амінокислотних залишків між парами основ. У місці інтеркаляції порушуються стекінг-взаємодії та утворюється кінк (kink) – різкий злам подвійної спіралі.



**Рис. 3.18.** ДНК-зв'язувальний домен TBP у комплексі з TATA-боксом (1QNE)

На відміну від розглянутих факторів транскрипції та структурних білків, ферменти, які працюють на ДНК, доволі важко описати в термінах простих структурних мотивів. Ферменти використовують складні комбінації різноманітних елементів білкової структури для впізнання ДНК і зв'язування з нею. Два приклади – мономерна ДНКаза I (неспецифічна ендонуклеаза, яка здійснює одноланцюговий розріз у ділянці маленького жолобка) і гомодимерна рестриктаза *EcoRV* – зображені на рис. 3.19.



**Рис. 3.19.** Комплекси з ДНК ДНКазі I (а, 1DNK) і рестриктази *EcoRV* (б, 1AZ0)

## Принципи білково-нуклеїнового впізнання

Усі випадки взаємодії численних білків із ДНК можна поділити на дві категорії: *неспецифічне зв'язування* білка з ДНК будь-якої послідовності (характерна константа зв'язування  $K \sim 10^5\text{--}10^6$  л/моль, визначення константи зв'язування див. у розділі 1) і *специфічне впізнання* білком певної послідовності пар основ (із константою  $\sim 10^9\text{--}10^{10}$  л/моль). Білки, що здійснюють таке впізнання, зазвичай неспецифічно взаємодіють і з будь-якою іншою ДНК. Розглянемо на простому прикладі, що практично означають наведені цифри.

Нехай у бактеріальній клітині радіусом 1 мкм міститься 10 молекул певного білка та 4 млн 600 тис. пар основ ДНК, де наявна одна специфічна ділянка (оператор). Виходячи з об'єму клітини, можна легко розрахувати загальні молярні концентрації білка й оператора (скориставшись числом Авогадро). Концентрація комплексу з оператором визначається рівнянням (1.6), де вільні концентрації білка й оператора дорівнюють їхнім загальним концентраціям мінус концентрація комплексу. Звідси, якщо відома  $K$ , можна розрахувати концентрацію комплексу.

Якщо білок зв'язується неспецифічно ( $K = 10^5$  л/моль), то відношення концентрації комплексу до загальної концентрації оператора дорівнює  $4 \cdot 10^{-4}$  – оператор практично вільний від білка. Якщо ж замість концентрації оператора використати загальну концентрацію потенційних сайтів зв'язування на ДНК (білок зв'язується будь-де), яка у 4 млн 600 тис. разів вища (кожна пара основ потенційно може бути початком сайта зв'язування), то відношення концентрації комплексу до загальної концентрації білка дорівнюватиме 0,995, тобто практично білок увесь час зв'язаний із ДНК.

При специфічному зв'язуванні ( $K = 10^{10}$  л/моль) відношення концентрації комплексу до загальної концентрації оператора дорівнює 0,97, тобто десяти молекул білка достатньо, щоб більшу частину часу оператор був зв'язаним. При цьому з оператором у нашому прикладі в певний момент часу може бути зв'язана лише одна молекула білка, решта перебуватиме не у вільному стані, а на ДНК, взаємодіючи з нею неспецифічно.

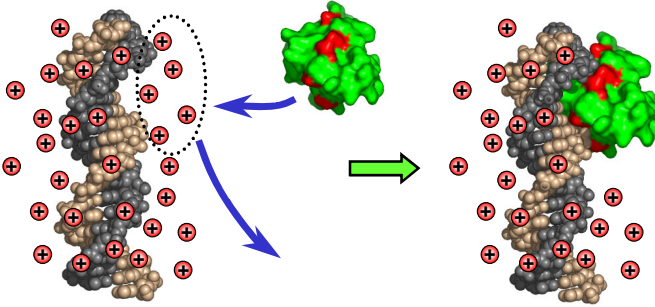
Чим визначається висока специфічність зв'язування? **Головне правило білково-нуклеїнового впізнання – відсутність жорстких правил.** Розглянуті різні структурні мотиви білків є різними еволюційними рішеннями для специфічної взаємодії з тією чи іншою послідовністю: існує багато шляхів для того, щоб сформувати білкову поверхню для впізнання послідовності пар основ. Не існує і будь-якого коду впізнання – чіткої відповідності між амінокислотними залишками й парами основ. Однак є певні загальні закономірності, частина яких уже має бути зрозумілою з розгляду ДНК-зв'язувальних структурних мотивів. Зокрема, *білково-нуклеїновий інтерфейс* частіше представлений парою  $\alpha$ -спіраль – великий жолобок завдяки хорошій просторовій відповідності між цими двома елементами. Проте використовуються також і  $\beta$ -структура, і перемички між елементами вторинної структури для впізнання пар основ у великому жолобку. Маленький жолобок також використовується для впізнання послідовностей пар основ, причому у випадку взаємодії елементів вторинної структури білка ( $\alpha$ -спіраль чи  $\beta$ -структура) із маленьким жолобком така взаємодія супроводжується розкриттям останнього. Це розкриття стає можливим унаслідок значної деформації подвійної спіралі (розкручування, вигинання, кінк у результаті порушення стекінг-взаємодій).

На поверхні ДНК розташовані фосфатні залишки й донорні / акцепторні групи азотистих основ у жолобках. Саме ці групи залучаються до контактів з амінокислотними боковими залишками й пептидними групами на поверхні білка.

#### ***Типи контактів (взаємодій) між ДНК і білками***

***Електростатичні взаємодії*** між позитивно зарядженими амінокислотними залишками й негативно зарядженими фосфатами наявні завжди в білково-нуклеїнових комплексах. Зрозуміло, що електростатичні взаємодії відповідають за неспецифічне зв'язування. Зазвичай вони додатково стабілізують і специфічні комплекси. Електростатичне зв'язування білка з нуклеїною кислотою має цілком ентропійну природу (див. розділ 1). Висока концентрація негативних зарядів (фосфатів) на поверхні ДНК зумовлює формування навкруг ДНК іонної атмосфери з доволі

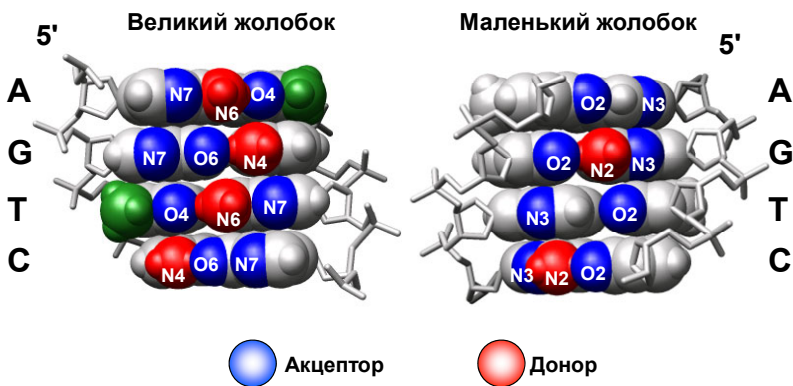
високою локальною концентрацією катіонів (рис. 3.20). Зв'язування позитивно зарядженого білка спричиняє визволення частки катіонів у зовнішній розчин, тобто зростання неупорядкованості (ентропії) у системі. Енергетичний вигравш від зв'язування білка (його спорідненість до ДНК) є тим більшим, чим меншою є концентрація солі в розчині, тобто різниця між концентрацією катіонів поблизу ДНК і на віддаленні від неї.



**Рис. 3.20.** Електростатичне зв'язування білка (червоним позначені позитивно заряджені амінокислотні залишки) з ДНК за рахунок визволення неорганічних катіонів

**Водневі зв'язки між донорними й акцепторними групами білка та фосфатами й екзоциклічними групами азотистих основ.** Водневі зв'язки, оскільки вони потребують чіткої взаємної орієнтації донора й акцептора, відіграють роль головного фактора специфічного впізнання. Кожна послідовність пар основ утворює в жолобках подвійної спіралі власний патерн донорних і акцепторних груп (рис. 3.21, див. також рис. 3.7), який і може упізнаватися поверхнею білка (білок здійснює *безпосереднє зчитування* – direct readout – послідовності). Слід зазначити, що цей патерн більш варіабельний у великому жолобку, де легше розрізнити пари основ і динуклеотидні контакти. Це ще одна причина, чому саме великий жолобок частіше використовується для впізнання. Деякі амінокислотні залишки здатні утворювати два водневі зв'язки з азотистою основою, наприклад Arg із гуаніном. Проте Arg контактує також з усіма іншими основами. Чи буде певний залишок залучений до утворення водневого зв'язку,

і якого саме, залежить від його орієнтації на поверхні. Певні закономірності (своєрідний код, коли залишок певного типу в певному місці утворює зв'язок із певною основою) спостерігаються іноді тільки в межах однієї родини структурних мотивів, але й тоді вони не мають абсолютного характеру.



**Рис. 3.21.** Патерни донорів і акцепторів водневого зв'язку в жолобках подвійної спіралі для зазначеної послідовності чотирьох пар основ

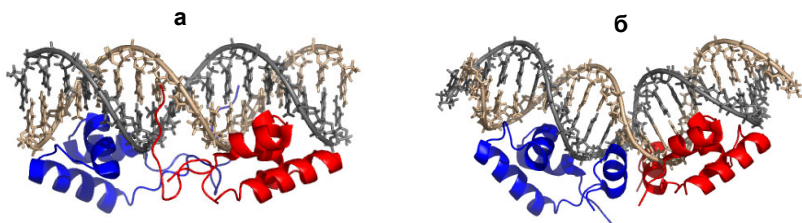
**Водневі зв'язки, опосередковані молекулами води.** Вода взаємодіє з поверхнями як білків, так і ДНК. Визволення води з білково-нуклеїнового інтерфейсу вносить додаткову ентропійну складову у стабілізацію комплексу. Однак часто окремі молекули води можуть залишатися в інтерфейсі, виконуючи роль бівалентної шивки, тобто утворюючи водневі зв'язки з білковими групами й фосфатами або групами азотистих основ.

**Гідрофобні контакти** можуть здійснюватись за участю метильної групи тиміну у великому жолобку (рис. 3.7, 3.21), але суттєвішими є гідрофобні взаємодії при інтеркаляції неполярних амінокислотних залишків між парами основ при деформації подвійної спіралі внаслідок занурення елементів вторинної структури білка в маленький жолобок (рис. 3.17, 3.18).

**Головною умовою реалізації розглянутих контактів між ДНК і білком є взаємна підгонка структури взаємодіючих елементів унаслідок відповідних конформаційних перетворень.**

Шляхом певних конформаційних змін подвійної спіралі (вигини, зміни твіста тощо) змінюється розмір жолобків, хімічні групи "підводяться" під водневі зв'язки й міцні електростатичні контакти, збільшується загальна поверхня, що може взаємодіяти з білком. Це означає, що для ефективного впізнання необхідні певна конформація подвійної спіралі та/або певні зміни цієї конформації. І те й інше визначається послідовністю пар основ – білок, таким чином, зчитує не стільки нуклеотидну послідовність як таку, скільки її структурно-динамічні особливості, тобто відбувається *опосередковане зчитування* (indirect readout). Отже, основою для білково-нуклеїнового впізнання є структурно-динамічний поліморфізм нуклеотидних послідовностей – залежні від послідовності структурні особливості й конформаційна рухливість подвійної спіралі.

При цьому білок часто потребує певної конформаційної підгонки під подвійну спіраль. Один із прикладів такої, доволі значної, конформаційної зміни як у молекулі білка (формування додаткових  $\alpha$ -спіралей), так і у молекулі ДНК (вигин подвійної спіралі) при утворенні специфічного комплексу показано на рис. 3.22. Саме в процесі взаємної конформаційної підгонки білка та ДНК реалізується специфічний патерн контактів у інтерфейсі між двома молекулами. Від того, наскільки легко потрібні конформаційні зміни відбуваються при зустрічі конкретного білка з конкретною послідовністю пар основ, залежить ефективність упізнання. Отже, головним механізмом упізнання послідовності ДНК білком є структурно-динамічна комплементарність двох молекул.



**Рис. 3.22.** Структура ДНК-зв'язувальних доменів димеру репресора лактозного оперона *E. coli* у комплексі з неспецифічною ДНК (а, 1OSL) та своїм специфічним оператором (б, 1L1M)

## Циркулярна ДНК

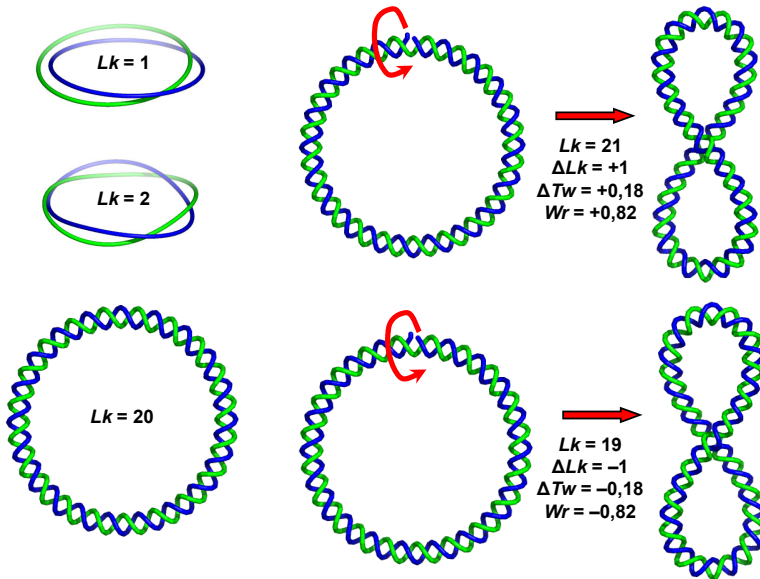
ДНК майже ніколи не існує *in vivo* у вигляді лінійної молекули із двома вільними кінцями. Прокаріотична ДНК є циркулярною ковалентно замкненою, в еукаріотів у складі хромосом утворюються еквівалентні циркулярній ДНК петлі, закріплені своїми кінцями так, що ці кінці не можуть обертатись один відносно одного (розділ 4). У молекулі, два кінці якої жорстко зафіксовані або з'єднані один з одним, виникають *топологічні* обмеження, і це має важливі біологічні наслідки.

Топологія – розділ математики, присвячений вивченню таких властивостей просторових об'єктів, які не залежать від деформацій цих об'єктів. Центральним поняттям топології циркулярної ДНК є *число зчеплень*  $Lk$  (linking number) двох полінуклеотидних ланцюгів – двох кілець, закручених одне навкруг одного в подвійну спіраль. Число зчеплень визначається як кількість перетинів одним циркулярним контуром поверхні, що натягнута на інший контур. У найпростішому випадку два кільця можна зчепити одне з одним один раз ( $Lk = 1$ , рис. 3.23), а будь-які деформації кілець (за умови збереження цілісності кожного кільця) не змінять цієї величини: кільця не можна ні розвести, ні збільшити ступінь зчеплень. Отже, і це одна із двох важливих властивостей числа зчеплень:  ***$Lk$  системи двох кілець є сталою величиною*** – топологічним інваріантом, поки обидва кільця інтактні (не містять розривів). Легко уявити два кільця, зчеплені одне з одним два (рис. 3.23) або сто разів, але не може бути двох із половиною зчеплень, тобто  ***$Lk$  є цілим числом***. Саме ці дві властивості числа зчеплень накладають топологічні обмеження на молекулу циркулярної ДНК.

На рис 3.22 показано планарну (вісь подвійної спіралі лежить в одній площині) циркулярну молекулу ДНК з  $Lk = 20$ . Зрозуміло, що для такої системи  $Lk$  дорівнює кількості витків подвійної спіралі.

Кількість витків подвійної спіралі позначають як твіст  $Tw$  (twist, дорівнює сумі кутів твіста для всіх пар основ, поділений на  $360^\circ$ , із позначкою "+" для правої спіралі та "-" для лівої), для планарної циркулярної ДНК  $Lk = Tw$ . Найвигідніший твіст ДНК

$T_{w_0}$  (часто записують також  $Lk_0$ ) визначається зовнішніми умовами й послідовністю пар основ (типове середнє значення за фізіологічних умов відповідає приблизно 10,5 парам основ на виток подвійної спіралі), тобто зовсім не має бути цілим числом. Лише тоді, коли  $T_w = T_{w_0}$  (припустимо, що ця умова виконується для прикладу на рис. 3.23 і  $T_{w_0} = 20$ ), при замиканні в кільце ДНК набуде найбільш енергетично вигідної планарної форми, у складі якої реалізується мінімальний вигин молекули.



**Рис. 3.23.** Ліворуч: два кільця з різним числом зчеплень  $Lk$ , у тому числі планарна форма циркулярної ДНК із  $Lk = 20$ . Праворуч: збільшення чи зменшення ступеня закручення подвійної спіралі на один оберт зумовлює відповідно позитивну чи негативну надспіралізацію

Якщо ж (як це частіше буває)  $T_{w_0}$  не є цілим числом, то замикання в кільце (точне зведення кінців обох ланцюгів) можливе або за рахунок зміни твіста до цілого значення (треба підкрутити подвійну спіраль, щоб виставити один кінець точно навпроти одного), або шляхом додаткових вигинань молекули з відхиленням

від планарної форми. У будь-якому разі замикання в кільце буде можливим лише за умови деформацій молекули ДНК.

Отже, у загальному випадку, коли  $Lk$  (яке має бути цілим числом) не збігається із  $T_{w_0}$ , число зчеплень має дві складові згідно з рівнянням Уайта – Фуллера (James H. White, F. Brock Fuller):

$$Lk = T_w + W_r,$$

де *райзинг*  $W_r$  (writhing) – параметр, який є мірою відхилення осі подвійної спіралі від планарної конфігурації (для планарного кільця  $W_r = 0$ ). Розподіл величини  $Lk$  по двох складових залежить від механічних властивостей молекули ДНК: той факт, що  $Lk \neq T_{w_0}$ , означає наявність деформації (відхилення конформації від найвигіднішої), і ця деформація мінімізується, розподіляючись певним чином між зміною твіста  $\Delta T_w = T_w - T_{w_0}$  і ненульовим райзингом унаслідок зростання вигину молекули. Отже, загальною мірою деформації є величина

$$\Delta Lk = Lk - T_{w_0} = \Delta T_w + W_r.$$

Оскільки зростання райзингу (при зростанні загальної деформації) пов'язане зі спіралізацією осі подвійної спіралі – *надспіралізацією* (supercoiling), то величина  $\Delta Lk$  використовується як міра надспіралізації в циркулярній ДНК. Надспіралізація є тим більшою, чим більше  $Lk$  відрізняється від  $T_{w_0}$  і чим більша напруга, пов'язана з деформацією, накопичується в ДНК.

Описане ілюструє рис. 3.23 на прикладі ДНК із  $T_{w_0} = 20$ . Якщо зробити односторонній розрив, збільшити кількість витків подвійної спіралі на 1 (закрутити подвійну спіраль торсійно) і зашити розрив, відновивши ковалентний зв'язок, то тим самим кількість зчеплень зросте на 1:  $\Delta Lk$  набуде значення +1, яке розподілиться між зміною твіста й райзингом. У цьому випадку кажуть про *позитивну надспіралізацію*, яка топологічно еквівалентна зростанню закручення подвійної спіралі. Відповідно *розкручення* подвійної спіралі після розриву й подальше відновлення цілісності ланцюга спричинить *негативну надспіралізацію* з  $\Delta Lk = -1$ .

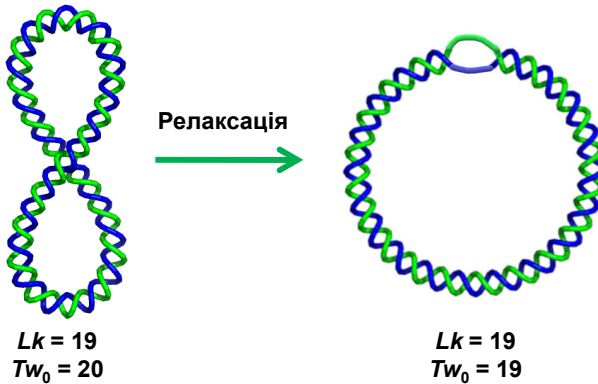
Однакові циркулярні молекули ДНК, які різняться лише числом зчеплень (як три молекули із  $Lk = 19, 20$  та  $21$  на рис. 3.23), називаються *топоізомерами*. Зрозуміло, що практично застосувати описану процедуру для отримання різних топоізомерів доволі

важко. Однак це можна зробити, здійснивши ензиматичне зшивання ДНК у кільце в різних умовах, що відповідають різним значенням  $T_{w_0}$ . Отримані топоізмери будуть відповідати найменш енергетичним формам – таким, для яких значення  $Lk$  є максимально наближеними до значень найвигіднішого твіста. Щойно зациклення відбулося, значення  $Lk$  залишається незмінним. А зміна умов і нове значення  $T_{w_0}$  спричинить зростання величини  $\Delta Lk$  у негативний чи позитивний бік і, відповідно, надспіралізацію.

З іншого боку, значення  $Lk$  залишається незмінним лише за умови цілісності обох полінуклеотидних ланцюгів. Якщо внести хоча б один розрив у хоча б один із ланцюгів, то два кінці ланцюга в місці розриву отримають свободу обертатися навкруг інтактного ланцюга. У результаті будь-яка надспіралізація і пов'язана з нею еластична напруга зникнуть – відбудеться *релаксація* циркулярної ДНК

Для негативно надспіралізованої ДНК існує ще один шлях релаксації (принаймні часткової) без розривів – локальне розкручування подвійної спіралі (рис. 3.24). Припустимо, що лінійна молекула ДНК із  $T_{w_0}=20$  (як на рис. 3.23) може існувати також у формі зі зруйнованим (розплавленим) одним витком подвійної спіралі, тобто з величиною  $T_{w_0}=19$ . За фізіологічних умов розведення ланцюгів потребує зростання вільної енергії, тобто така форма неможлива. Однак локальне плавлення подвійної спіралі у складі негативного топоізомеру буде супроводжуватись також зниженням вільної енергії за рахунок релаксації, оскільки  $\Delta Lk=19-20=-1$  зміниться на  $\Delta Lk=19-19=0$ . У результаті, якщо загальна вільна енергія знизиться, то локальна дестабілізація подвійної спіралі стане ймовірною – негативна надспіралізація суттєво полегшує руйнування подвійної спіралі. За таким самим механізмом енергетично не вигідне конформаційне перетворення стає вигідним за рахунок того, що знімає негативну надспіралізацію в кільці. У негативно надспіралізованій ДНК (за фізіологічних умов!) відбуваються також інші конформаційні перетворення (напр. перехід у Z-форму, яка характеризується розкрученням подвійної спіралі порівняно з B-ДНК). Оскільки більшість функціональних процесів, що відбуваються у ДНК, потребують локального руйнування подвійної спіралі, то зрозу-

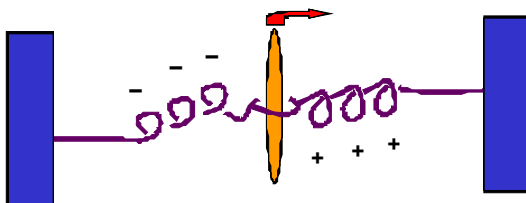
міло, що негативна надспіралізація активує ці процеси, полегшуючи руйнування на найменш стабільних ділянках.



**Рис. 3.24.** Релаксація негативно надспіралізованої ДНК за рахунок локального розкручування подвійної спіралі

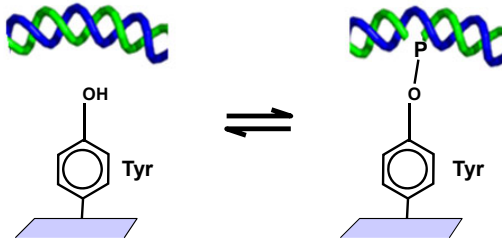
З іншого боку, функціональні процеси (передусім транскрипція і реплікація) самі зумовлюють надспіралізацію. Проходження такого процесу пов'язане з пересуванням (*транслокацією*) уздовж ДНК того чи іншого ферменту (*транслокази*), який локально руйнує подвійну спіраль. Транслоказами є РНК- і ДНК-полімерази, які розглядатимуться в наступних розділах, а також деякі інші ферменти. Оскільки молекула ДНК – спіраль, то пересування транслокази має супроводжуватись або її обертанням навкруг осі подвійної спіралі (на кшталт обертання гайки навкруг гвинта), або прокручуванням самої подвійної спіралі (обертанням гвинта в гайці). Саме друга можливість реалізується, оскільки транслоказа працює завжди у складі величезного мультибілкового комплексу – "гайка" є надто масивною. Крім того, транслоказа може бути закореною на певних структурах, що робить її обертання узагалі неможливим. Якщо (як це має місце у клітині) кінці ДНК жорстко зафіксовані й не можуть обертатися, то транслокація буде створювати топологічні проблеми (рис. 3.25). У процесі пересування вздовж ДНК транслоказа руйнує подвійну спіраль попереду від себе й від-

новлює її позаду. Локальне розкручування подвійної спіралі попереду транслокази має бути компенсованим позитивною надспіралізацією, відновлення спіралі (закручування) – надспіралізацією негативною. У результаті попереду й позаду від транслокази виникають дві хвилі надспіралізації протилежного знака (рис. 3.25).



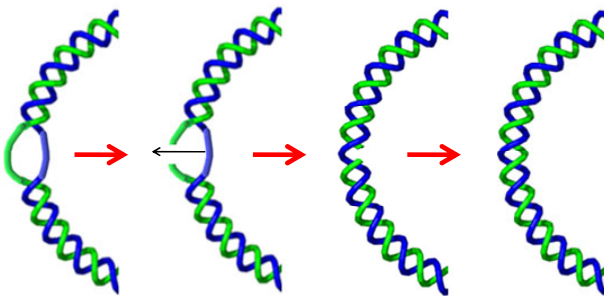
**Рис. 3.25.** Дві хвилі надспіралізації у процесі роботи транслокази

Зрозуміло, що накопичення надспіралізації (еластичної напруги) не може тривати нескінченно: урешті-решт напруга буде блокувати процес транслокації. Отже, має бути спосіб розв'язувати цю проблему – знімати надспіралізацію, релаксуючи ДНК. Інструментами, що для цього використовуються, є спеціальні ферменти – *ДНК-топоізомерази* (topoisomerases). Оскільки зміна кількості зчеплень можлива лише за умови порушення цілісності полінуклеотидних ланцюгів, то загальний механізм роботи топоізомераз полягає у внесенні тимчасового розриву. При цьому один кінець у місці розриву полінуклеотидного ланцюга залишається вільним, а фосфат на іншому кінці тимчасово ковалентно пришивається до ОН-групи залишку Туг в активному центрі ферменту (хімічна реакція заміни одного фосфодієфірного зв'язку на інший – *трансестерифікація*). У зворотній реакції цей зв'язок руйнується і відновлюється цілісність полінуклеотидного ланцюга (рис. 3.26). Зміна  $Lk$  відбувається у проміжку між цими двома подіями – поки існує розрив.



**Рис. 3.26.** Тимчасовий одноланцюговий розрив у ДНК із пришиванням фосфату до залишку Tyr в активному центрі топоізомерази

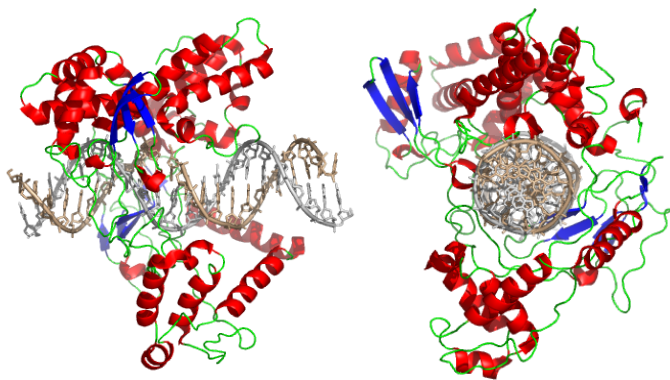
Топоізомерази поділяють на два класи. **Топоізомерази I** (мономерні білки), які поділяють на два підкласи Ia і Ib, індукують одноланцюговий розрив ДНК. **Топоізомерази Ia** (наявні як у про-, так і в еукаріотів) здатні релаксувати тільки негативно надспіралізовану ДНК. Фермент упізнає дестабілізовану ділянку подвійної спіралі (яка має бути в негативно надспіралізованій молекулі, див. вище), робить одноланцюговий розрив і протягує крізь нього інтактний ланцюг (рис. 3.27). У результаті число зчеплень змінюється на одиницю в напрямку зниження рівня негативно надспіралізації. Оскільки впізнається дестабілізована ділянка, то топоізомераза Ia спрацьовує тільки за наявності високої негативно надспіралізації і не здатна релаксувати ДНК повністю (знизити  $\Delta Lk$  до 0). Основна функція топоіомераз цього типу – підтримувати певний оптимальний рівень надспіралізації.



**Рис. 3.27.** Схема дії топоіомераз Ia

*Топоізомерази Ib* (наявні тільки в еукаріотів) переводять будь-яку ДНК у максимально релаксований стан. Механізм їхньої дії дуже простий: мультидоменний мономерний білок зв'язується з ДНК, оточуючи подвійну спіраль з усіх боків (рис. 3.28); індукується одноланцюговий розрив; один із кінців, які утворилися, отримує змогу вільно обертатися навколо інтактного ланцюга; після трьох-чотирьох випадкових обертів розрив зшивається і фермент дисоціює. Оскільки оберти є вільними, то вони здійснюються в напрямку зниження напруги – максимальної релаксації. Зрозуміло, що *Lk* у результаті роботи топоізомерази змінюється на величину, кратну одиниці, – на кількість обертів.

Отже, топоізомерази I обох типів змінюють число зчеплень шляхом зміни твіста подвійної спіралі. Іншою спільною ознакою топоізомераз I є незалежність їхньої активності від АТР: відновлення цілісності полінуклеотидного ланцюга відбувається після тимчасового ковалентного пришивання одного з кінців до активного центра ферменту й не потребує зовнішніх джерел вільної енергії.



**Рис. 3.28.** Еукаріотична топоізомераза I у комплексі з ДНК у двох проєкціях (1K4S)

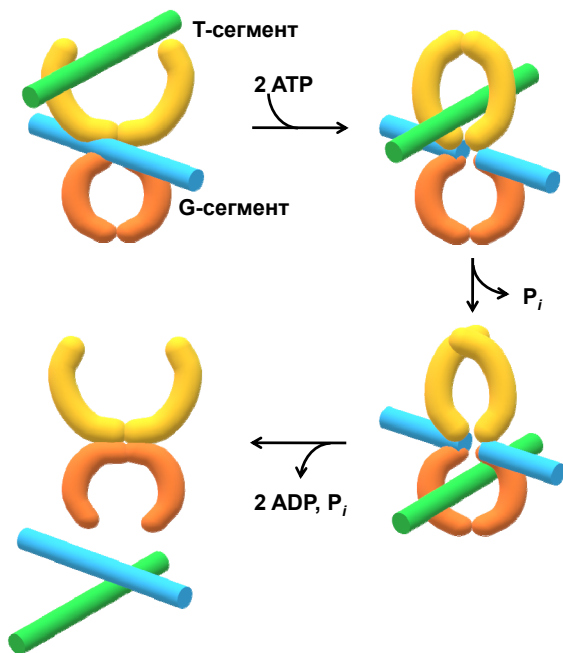
*Топоізомерази II*, на відміну від топоізомераз першого класу, змінюють число зчеплень шляхом зміни райзингу циркулярної ДНК. Є також дві інші відмінності: ці субодиничні білки мають два активні центри, у яких відбувається розрізання обох поліну-

клеотидних ланцюгів; фермент є активним тільки у присутності АТР, гідроліз якої використовується для здійснення конформаційних змін білка. Структурні домени двох субодиниць еукаріотичної топоізомерази II (рис. 3.29) або чотири субодиниці прокаріотичних топоізомераз цього типу утворюють своєрідні верхні й нижні "ворота", які здатні розкриватися і закриватися при структурних перебудовах.



**Рис. 3.29.** Еукаріотична топоізомераза II у комплексі із G-сегментом ДНК, у якому індукується дволанцюговий розрив (2RGR)

Схему роботи топоізомерази II зображено на рис. 3.30: 1) ділянка ДНК, яка називається G-сегментом, зв'язується з ферментом, що викликає спорідненість ферменту до АТР; 2) у відповідь на зв'язування АТР відбувається конформаційна зміна, яка супроводжується замиканням іншої зв'язаної ділянки ДНК – Т-сегмента – і дволанцюговим розривом у складі G-сегмента; 3) відбувається гідроліз АТР, що зумовлює проштовхування Т-сегмента крізь розрив; 4) здійснюється зшивання розриву та звільнення обох сегментів. Шляхом описаної операції можна перетворити, наприклад, одну надспіралізовану молекулу з рис. 3.23 на іншу, звідки зрозуміло, що число зчеплень при роботі топоізомераз II змінюється на величину, кратну двом.

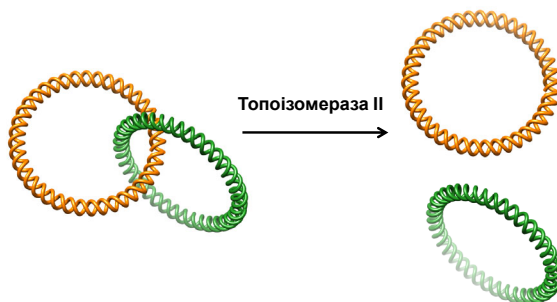


**Рис. 3.30.** Схема дії топоізомерази II

Напрямок зміни  $Lk$  залежить від взаємної орієнтації G- і T-сегментів. Еукаріотичні топоізомерази II дозволяють обидві орієнтації: у результаті  $Lk$  змінюється в напрямку релаксації ДНК. Однак до класу топоізомераз II належить також бактеріальний фермент – *гіраза* (*gyrase*), який допускає тільки одну орієнтацію (див. рис. 3.30). Результатом роботи гірази є зміна  $Lk$  тільки в одному напрямку – накопичення негативної надспіралізації. Це має важливі функціональні наслідки, оскільки сприяє дестабілізації подвійної спіралі ДНК на певних важливих ділянках у складі бактеріальної ДНК.

Оскільки топоізомерази II продукують тимчасовий дволанцюговий розрив, то вони виконують ще одну важливу функцію – розділюють переплутані (зав'язані у вузол) або зчеплені молекули ДНК. Дві дволанцюгові циркулярні молекули ДНК, зчеплені одна за одну, називають *катенаном* (рис. 3.31). Зрозу-

міло, що розвести дві такі молекули (здійснити декатенацію) можна, тільки зробивши дволанцюговий розріз в одній із них, і саме на це здатні топоізомерази класу II. Така або аналогічна операція є абсолютно необхідною для завершення процесу реплікації ДНК (див. розділ 9).



**Рис. 3.31.** Розділення катенану за допомогою топоізомерази II

### Контрольні запитання

1. З яких трьох елементів складається нуклеотид? Назвіть основні фізико-хімічні властивості цих елементів.
2. Чим хімічно відрізняються між собою рибо- і дезоксирибонуклеїнові кислоти?
3. Між якими хімічними групами утворюється полінуклеотидний ланцюг за рахунок ковалентного зв'язку? Як позначаються кінці ланцюга й чому саме так?
4. Опишіть основні риси структури подвійної спіралі ДНК.
5. Які взаємодії стабілізують подвійну спіраль? Що лежить в основі комплементарності нуклеотидів і яке значення має комплементарність для стабілізації спіралі?
6. Які структурні форми подвійних спіралей нуклеїнових кислот ви знаєте? Чим вони відрізняються? Які із цих форм є основними формами існування подвійних спіралей ДНК і РНК *in vivo*?
7. Які конформаційні параметри характеризують вигин і ступінь спірального закручення молекули ДНК?

8. Запишіть усі типи динуклеотидних контактів у складі полінуклеотидного ланцюга й подвійної спіралі. Як і чому різняться ці два набори контактів?

9. Чим визначаються локальна конформація і динамічні властивості подвійної спіралі?

10. Назвіть основні структурні мотиви білків, які взаємодіють із ДНК. Які елементи структури білків залучаються до взаємодії, з якими структурними елементами подвійної спіралі?

11. Взаємодії якого типу реалізуються між ДНК і білками?

12. Що відбувається при взаємодії елементів регулярної вторинної структури білків із маленьким жолобком ДНК?

13. Сформулюйте головний принцип, за яким здійснюється специфічне впізнання білком певної послідовності пар основ ДНК.

14. Що таке число зчеплень, твіст і райзинг циркулярної ДНК? Як співвідносяться ці величини? Що таке топоізомер?

15. Як виникає надспіралізація в циркулярних ДНК? Яка різниця між негативною і позитивною надспіралізацією?

16. У чому полягають основні механізми роботи й функціональне значення ДНК-топоізомераз?

## Рекомендована література

Сиволоб, А.В. (2011) *Фізика ДНК*. Київ : ВПЦ "Київський університет".

Anderson, C.F. and Record, M.T. (1995) Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.* 46, 657–700.

Bates, A.D. and Maxwell, A. (2005) *DNA topology*. New York : Oxford University Press.

Bloomfield, V.A., Crothers, D.M. and Tinoco, I. (2000) *Nucleic acids: structures, properties and functions*. Sausalito : University Science Books.

Calladine, C.R. and Drew, H.R. (1986) Principles of sequence-dependent flexure of DNA. *J. Mol. Boil.* 192, 907–918.

Dickerson, R.E. (1992) DNA structure from A to Z. *Methods Enzymol.* 211, 67–111.

Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell.* 8, 937–946.

Kouzine, F., Gupta, A., Baranello, L., et al. (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* 20, 396–403.

Neidle, S. (2008) *Principles of nucleic acid structure*. London : Academic Press.

Olson, W.K., Gorin, A.A., Lu, X.-J., et al. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* 95, 11163–11168.

Olson, W.K. and Zhurkin, V.B. (2000) Modeling DNA deformations. *Curr. Op. Struct. Biol.* 10, 286–297.

Rohs, R., Jin, X., West, S.M., J., et al. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* 79, 233–269.

Sarai, A. and Kono, H. (2005) Protein-DNA recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Struct.* 34, 379–398.

Stormo, G.D. (2013) *Introduction to protein-DNA interactions. Structure, thermodynamics, and bioinformatics*. Cold Spring Harbor : Cold Spring Harbor Laboratory Press.

Vos, S.M., Tretter, E.M., Schmidt, B.H. and Berger, J.M. (2011) All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat. Rev. Mol. Cell. Biol.* 12, 827–841.

Vologodskii, A.V. (2015) *Biophysics of DNA*. Cambridge : Cambridge University Press.

Watson, J.D. (2012) *Double helix*. Annotated and illustrated. New York : Simon and Schuster.

Yakovchuk, P., Protozanova, E. and Frank-Kamenetskii, M.D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucl. Acids Res.* 34, 564–574.

## Розділ 4

# ОРГАНІЗАЦІЯ ДНК У КЛІТИНАХ: ГЕНОМИ ТА СТРУКТУРА ХРОМАТИНУ

Il n'existe rien hors contexte.  
J. Derrida

"I am a fair length – a fair length," said Kaa with a little pride.  
R. Kipling. The Jungle Book

У цьому розділі розглянуто два важливі аспекти організації ДНК у клітинах. Перший – інформаційний: який вигляд має загальна сукупність спадкової інформації, записаної у нуклеотидних послідовностях ДНК. Другий аспект – структурний: як упакована ДНК у клітинах. Цей аспект особливо важливий для еукаріотичних клітин, ядра яких містять надзвичайно велику кількість ДНК, і має безпосереднє відношення до механізмів реалізації спадкової інформації – щільна упаковка повинна дозволити знімати спадкову інформацію з визначених ділянок у визначені моменти життя клітини.

## Організація геномів

### Генетичний код

У послідовності нуклеотидів ДНК записана інформація про послідовність амінокислот у складі білків. Відповідність між цими двома типами текстів – комбінаціями нуклеотидів і амінокислотами – називається *генетичним кодом*. Загальна кількість комбінацій по два нуклеотиди із чотирьох дорівнює 16, а по три – 64, тобто мінімальна кількість нуклеотидів у одному слові нуклеотидного тексту, оскільки такими словами потрібно зако-

дувати 20 амінокислот, має дорівнювати трьом. Саме це й спостерігається: одне слово – *кодон* – є триплетом нуклеотидів.

Послідовності всіх 64 кодонів (у напрямку від 5'- до 3'-кінця у складі мРНК) наведено на рис. 4.1. Серед 64 кодонів три є сигналами зупинки синтезу білка (*стоп-кодони*, або *нонсенс-кодони*), решта, 61 змістовний кодон, відповідають двадцяти амінокислотам. Співвідношення між кодонами й амінокислотами є однозначним: певний триплет кодує одну й тільки одну амінокислоту. Зворотнє співвідношення не є однозначним: більшість амінокислот кодуються кількома триплетами – код є *виродженим* (*degenerate*). Дві амінокислоти, Trp і Met, невироджені й кодуються лише одним кодоном кожна. Усі інші 18 амінокислот кодуються серіями кодонів-синонімів – від двох до шести кодонів на серію.

Якщо розглянути триплети синонімічних серій, то стає очевидним, що три позиції кодону нерівнозначні.

- Найважливішою для визначення змісту кодону є комбінація нуклеотидів у першій і другій позиції, при цьому друга позиція визначає зміст кодону найбільш жорстко – нуклеотидні заміни у ній завжди зумовлюють зміну змісту (амінокислотну заміну).
- Найменш змістовною є третя позиція – нуклеотидні заміни в ній у 70 % випадків не зумовлюють зміну змісту кодону.

Залежно від того, який нуклеотид стоїть у другій позиції, усі кодони й відповідні амінокислоти можна розділити на чотири групи (чотири стовпчики в таблиці коду на рис. 4.1). Виявляється, що до групи U (урацил у другій позиції) належать гідрофобні амінокислоти великого розміру з повністю аліфатичними боковими залишками (див. також рис. 2.2, 2.3), тобто нуклеотидні заміни в першій і третій позиціях у межах цієї групи не викликають зміни властивостей амінокислоти, мінімізуючи вплив таких амінокислотних замін на просторову структуру білка (див. розділ 2). До групи A належать великі полярні амінокислоти (за виключенням Туг, але й він, хоча загалом гідрофобний, також здатен утворювати водневий зв'язок). Група С сформована амінокислотами маленького розміру. До групи G потрапля-

ють амінокислоти з екстремальними властивостями: найбільша гідрофобна – Trp, найбільша полярна – Arg, найбільш реакційно-здатна – Cys, найменша – Gly.

		2й нуклеотид			
		U	C	A	G
1й нуклеотид	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }
	A	AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys	AGU } Ser AGC } AGA } Arg AGG }
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }

Рис. 4.1. Таблиця генетичного коду

Генетичний код є доволі універсальним – наведена таблиця коду реалізується як для бактерій, так і для ссавців. Однак у деяких випадках (окремі прокаріоти, гриби, водорості, а також автономна генетична система мітохондрій) спостерігаються невеличкі відхилення від цієї універсальної таблиці.

При синтезі білка (розділ 8) триплети зчитуються з нуклеотидного тексту один за одним: сусідні триплети не перекриваються, між ними відсутні проміжки. Відповідно будь-яка послідовність нуклеотидів може бути прочитана трьома різними способами, тобто містить три рамки зчитування (рис. 4.2). Рамка, що міститься між стартовим (зазвичай як стартовий використовується метіоніновий кодон AUG) і стоп-кодомом, називається *відкритою рамкою зчитування* (ORF, open reading frame).



**Рис. 4.2.** Три можливі рамки зчитування, одна з яких (позначена червоними дужками) є відкритою

## Гени

Відкрита рамка зчитування (кодувальна послідовність), яка містить інформацію про амінокислотну послідовність білка, є найважливішою змістовною частиною *гена* (gene). Однак для того, щоб відбулась *експресія* генетичної інформації (через синтез РНК і потім білка), не менш важливими є регуляторні послідовності ДНК, які за рахунок спорідненості до специфічних білків використовуються для увімкнення / вимкнення транскрипції (див. розділи 5, 6) як першої стадії експресії гена. Отже, визначення гена можна сформулювати так: ***ген – це ділянка ДНК, яка є необхідною і достатньою для повноцінного синтезу функціональної молекули РНК.*** Ділянка ДНК, яка може вважатися геном, має містити кодувальну послідовність, де записана інформація про продукт, а також певний набір регуляторних елементів послідовності, від яких залежить запуск / блокування процесу транскрипції, шлях зчитування інформації тощо. Згідно з визначенням міжнародного консорціуму онтології послідовностей (Sequence Ontology Consortium), ***ген – це певна визначена зона послідовності ДНК, яка відповідає одиниці спадковості й асоційована з регуляторними ділянками, ділянками, що транскрибуються, та/або іншими функціональними ділянками послідовності.***

Одним із найважливіших типів продуктів транскрипції генів є *mРНК* – матричні РНК (messenger RNA, mRNA), які використовуються далі як матриці для синтезу білків (піддаються трансляції, див. розділ 8). У цьому випадку білок є кінцевим продуктом гена, який, відповідно, називається *білковим*. Крім того, доволі велика кількість генів – *гени РНК* – кодують різноманітні молекули РНК, які не піддаються трансляції (є кінцевими продуктами), серед яких: *pРНК* – рибосомні РНК (ribosome RNA, rRNA, розділ 8); *mРНК* – транспортні РНК (transfer RNA, tRNA, розділ 8); *snRNA* – маленькі ядерні РНК (small nuclear RNA, розділ 7);

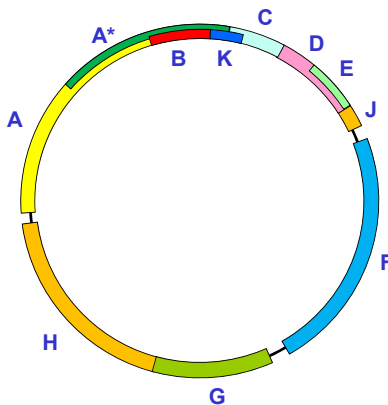
*snoRNA* – маленькі ядерцеві РНК (small nucleolar RNA, розділ 8); *miRNA* – мікроРНК (microRNA) і некодувальні регуляторні РНК (розділ 6), молекули РНК, які є компонентами деяких ферментів; інші види РНК, не для всіх із яких з'ясовано їхні функції.

## Геноми

ДНК, що міститься у клітині, – це не тільки гени: принаймні мають бути з'єднувальні міжгенні ділянки. **Сукупність нуклеотидних послідовностей ДНК у гаплоїдному наборі називається геномом** (мається на увазі геном *виду* – під геномом клітини або окремої особини розуміють повний (диплоїдний) набір послідовностей ДНК). На сьогодні повністю розшифровані послідовності більше 41 тис. вірусних, 23 тис. прокаріотичних і 150 еукаріотичних геномів (більше 17 тис. останніх установлені частково). Інформація про геномні послідовності розміщена у спеціальних загальнодоступних базах даних (див. список рекомендованої літератури). Головна відмінність між геномами різних груп організмів полягає в тому, що у прокаріотичних і вірусних геномах кодувальні послідовності становлять до 95 %, тоді як їхня частка у геномах еукаріотів значно менша – у більшості багатоклітинних еукаріотів на неї припадає мінорна частина геному (у дріжджів *Saccharomyces cerevisiae* – 70 %, у нематоди *Caenorhabditis elegans* – 45 %, у плодової мушки *Drosophila melanogaster* – 20 %, у ссавців – 1–2 %). Розміри деяких геномів і оцінка кількості генів у їхньому складі наведені в табл. 4.1. Ця оцінка залежить від способу анотації генів і тому може трохи розрізнятися у різних базах даних.

**Прокаріотичні й вірусні геноми.** Схему організації одного з найпростіших (і першого з вивчених) геномів – геному бактеріофага фХ-174 (слід читати "фі-десять") – показано на рис. 4.3. Геном побудований надзвичайно економно: десять генів (один із них дає два різні транскрипти А та А\*) займають практично всю циркулярну ДНК бактеріофага загальним розміром 5 386 нуклеотидів (усередині фагової частинки міститься одностанцюгова циркулярна молекула ДНК, яка перетворюється на двостанцюгову всередині бактеріальної клітини). Більше того, кілька генів перекриваються за рахунок використання різних рамок зчитування: гени А і С та С і D перекриваються своїми кінцями,

гени В, К і Е повністю знаходяться у межах інших генів; три гени – А, С і К – використовують усі три можливі рамки зчитування на одній ділянці ДНК (у цьому випадку всі три рамки є відкритими). Явище перекриття генів за рахунок використання різних рамок зчитування спостерігається також для кількох інших бактеріофагів, проте іноді зустрічається в еукаріотів. Загалом геноми вірусів і бактеріофагів побудовані за подібним економним принципом: порівняно невелика кількість генів, мінімальна кількість міжгенної ДНК.



**Рис. 4.3.** Геном бактеріофага φX-174.  
Літерами позначено гени у його складі

У геномі прокаріотичної клітини кількість ДНК і генів значно зростає, але зберігається принцип економності щодо використання більшості послідовностей для кодування генетичної інформації. Наприклад, геном *E. coli* представлений однією циркулярною молекулою ДНК (бактеріальною хромосомою) завдовжки приблизно 4,6 млн пар основ (це достатньо великий геном – у більшості вивчених бактерій розмір геному не перевищує 2 млн пар основ). Близько 90 % цієї ДНК припадає на кодувальні послідовності ~4 тис. білкових генів і ~200 генів РНК, що не транскрибується. Середній розмір гена становить 950 пар основ, середня довжина міжгенної ділянки – 118 пар основ. Проте міжгенні зони мають доволі нерівномірний розподіл за довжиною, яка варіює від 0 до 1 тис. 730 пар основ.

**Таблиця 4.1. Загальні характеристики деяких геномів**

Організм	Розмір геному (пари основ)	Кількість хромосом (молекул ДНК)*	Кількість білкових генів	Кількість генів РНК, що не транскрибуються
Бактеріофаг φX-174	5 386	1	10	–
Бактерія <i>Escherichia coli</i>	$4,6 \cdot 10^6$	1	4 316	207
Аскоміцет <i>Saccharomyces cerevisiae</i>	$1,2 \cdot 10^7$	16	6 600	424
Нематода <i>Caenorhabditis elegans</i>	$10^8$	6	20 191	24 791
Плодова мушка <i>Drosophila melanogaster</i>	$1,4 \cdot 10^8$	5	13 968	4 044
Курка <i>Gallus gallus</i>	$10^9$	41	16 878	7 166
Миша <i>Mus musculus</i>	$2,7 \cdot 10^9$	21	22 468	16 060
Людина <i>Homo sapiens</i>	$3,2 \cdot 10^9$	24	20 442	23 982
Мітохондріальна ДНК людини	16 569	1	13	24

\* Для еукаріотів наведено кількість хромосом у гаплоїдному наборі, для організмів із диморфізмом статевих хромосом (починаючи з *D. melanogaster*) – кількість аутосом плюс дві статеві хромосоми

Приблизно  $\frac{3}{4}$  транскрипційних одиниць *E. coli* містять один ген, решта реалізує характерний для прокаріотів (і тільки для них) оперонний принцип організації генетичного матеріалу. *Оперон* – це кластер *структурних генів*, на яких синтезується одна молекула мРНК, що має кілька (на кожен структурний ген) послідовних (які не перекриваються) відкритих рамок зчитування для трансляції відповідних білків (див. приклади організації та регуляції оперонів у розділі 6). У межах оперона зазвичай згруповані структурні гени, які відповідають за синтез білків,

залучених до одного ланцюжка біохімічних перетворень (ферменти синтезу або деградації певної сполуки). Крім структурних генів, оперон має регуляторні ділянки, за рахунок яких здійснюється регуляція його транскрипції як цілого. У геномі *E. coli* міститься ~650 таких одиниць транскрипції.

Характерною особливістю бактерій є наявність у клітині, поряд із бактеріальною хромосоною, невеликих автономних елементів геному – *плазмід*. Плазміда є циркулярною молекулою ДНК (типовий розмір ~3 тис. пар основ), яка містить кілька генів і реплікується незалежно від бактеріальної хромосоми. Плазміди широко використовуються як зручний інструмент молекулярно-біологічних досліджень (див. розділ 11).

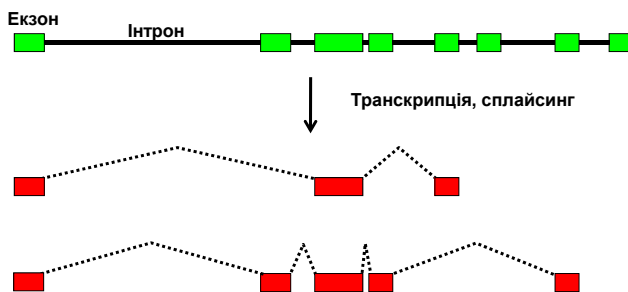
Циркулярна ДНК бактеріальної хромосоми існує у клітині у вигляді комплексу з білками. Ця взаємодія доволі динамічна, і практично весь бактеріальний геном (що принципово відрізняє його від еукаріотичного геному, див. нижче) перебуває в потенційно транскрипційно активному стані: гени й оперони є об'єктами швидкої оперативної регуляції у відповідь на зміну зовнішніх умов.

**Еукаріотичні геноми** містять значно більшу кількість ДНК порівняно з геномами прокаріотів (табл. 4.1), причому переважна її частина представлена послідовностями, що не є кодувальними. Зокрема, приблизно половина геному багатоклітинних еукаріотів – це послідовності, представлені багатьма копіями (послідовності, що повторюються). Еукаріотична ДНК існує у клітинному ядрі у складі хромосом, кожна хромосома містить одну гігантську лінійну молекулу ДНК. Послідовності, що повторюються, зосереджені, зокрема, на кінцях хромосом (теломери) і в зонах прикріплення хромосом до веретена поділу при мітозі (центромери).

Характерною ознакою білкових генів еукаріотів (на відміну від прокаріотів) є мозаїчний принцип будови кодувальної частини. Власне кодувальна частина – це послідовність окремих змістовних ділянок – *екзонів* (exon), розділених беззмістовними *інтронами* (intron). У середньому один ген містить 0,04 інтрони у *S. cerevisiae* (більшість генів є однокзонними), 3 інтрони у *D. melanogaster* і 9 – у людини. Часто екзони відповідають ок-

ремим структурним доменам мультидоменних білків: еволюційне збирання білка з "кубиків"-доменів може здійснюватись шляхом перетасування екзонів на рівні ДНК. Беззмистовними інтронами є в тому сенсі, що не містять інформації про кінцевий продукт, але в межах інтронів часто розташовані важливі регуляторні ділянки. При транскрипції молекула РНК синтезується суцільно (первинний продукт транскрипції – первинний транскрипт – містить екзони та інтрони). Отже, необхідним етапом експресії гена є процес *сплайсингу* (розділ 7) – вирізання інтронів і зшивання екзонів у кінцевий транскрипт, який може бути використаний як матриця для білкового синтезу.

Сплайсинг може відбуватися різними шляхами (рис. 4.4, *альтернативний сплайсинг*), унаслідок чого утворюються дещо різні білки як кінцеві продукти. Середня кількість таких альтернативних продуктів на один білковий ген у геномі, наприклад, людини, дорівнює приблизно п'яти. Отже, альтернативний сплайсинг можна інтерпретувати як механізм перекриття еукаріотичних генів. Інший випадок перекриття генів у геномах еукаріотів пов'язаний із можливістю розміщення гена (із своїми власними інтронами й екзонами) у межах інтрона іншого гена. В окремих випадках спостерігається також перекриття відкритих рамок зчитування.



**Рис. 4.4.** Мозаїчна будова кодувальної частини гена та схема утворення різних кінцевих продуктів унаслідок альтернативного сплайсингу

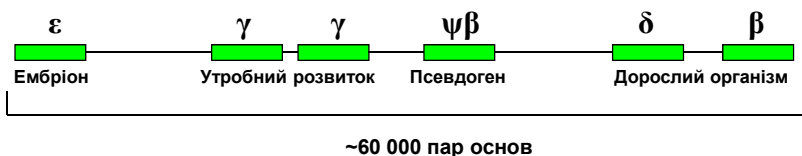
Усі гени багатоклітинного організму можна поділити на дві групи: 1) гени, від яких залежать певні універсальні функції та які активні в усіх клітинах, – *гени домашнього господарства* (housekeeping genes); 2) гени, що специфічно активуються у клітинах певного типу, – *гени розкоші* (luxury genes). Загальною ознакою генів першої групи є розташування в їхніх регуляторних зонах *CpG-острівців* (CpG-islands) – ділянок із підвищеним вмістом динуклеотидів CpG (контактів типу CG між парами основ, див. розділ 3). Вміст цих динуклеотидів у еукаріотичних геномах приблизно у п'ять разів менший за очікуваний унаслідок метилювання цитозину в складі CpG-контакту: 5mC (5-метилцитозин) із часом спонтанно перетворюється на тимін. Метилювання цитозину в регуляторних ділянках є одним із механізмів репресії генів (див. розділ 6). Відповідно гени, що зберігають активність у більшості клітин, містять неметилювані динуклеотиди CpG, вміст яких зберігається на високому рівні (для ДНК випадкової послідовності із 40 % GC-пар очікуваний вміст динуклеотидів CpG становить 4 %).

Типове значення вмісту GC-пар в еукаріотичних геномах становить близько 40 % (у геномі людини – 41 %). Проте спостерігаються варіації локального GC-вмісту в широких межах – від 30 до 60 %; зустрічаються й доволі довгі ділянки (до кількох мільйонів пар основ) з аномальним вмістом GC-пар. Загальна тенденція при цьому така: високий вміст GC-пар корелює з високою щільністю генів, зменшеною довжиною інтронів і низькою щільністю послідовностей, що повторюються.

Серед еукаріотичних генів 25–50 % є унікальними (представлені в геномі єдиною копією), решта належать до родин генів, що складаються з кількох копій, зазвичай неідентичних. Відповідні (гомологічні, але неідентичні) білки становлять родину білків. Кілька родин (протеїнази, транскрипційні фактори певного типу, імуноглобуліни) містять сотні білків, але більшість складаються із кількох (до 30) білків. Гени такої родини часто об'єднані в геномі в кластери, тобто розташовані поряд у певній хромосомі (кластери генів теплового шоку, глобінові гени). Слід зауважити, що такий кластер не є опероном – кожен

ген піддається регуляції як окрема одиниця транскрипції (хоча у кластера може бути й спільна регуляторна зона, яка спроможна переводити весь кластер у потенційно активний / неактивний стан). Наприклад, кластер генів  $\beta$ -субодиниці гемоглобіну містить гомологічні гени, які активуються на різних стадіях індивідуального розвитку (рис. 4.5).

$\beta$ -Глобіновий кластер містить також неактивний *псевдоген*. Псевдогени – це послідовності, які гомологічні певним генам, але не експресуються. Таку інактивацію гена може спричинити мутація, яка порушує ініціацію транскрипції, сплайсинг, викликає невчасну термінацію тощо. Після інактивації псевдоген перестає бути об'єктом добору, у ньому накопичується велика кількість мутацій. Зрозуміло, що передусім псевдогени виникають саме в кластерах, коли є кілька копій гена і пошкодження одного з них не призводить до фатальних наслідків.

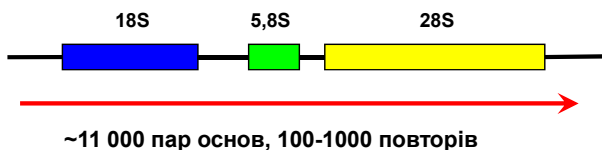


**Рис. 4.5.** Кластер генів  $\beta$ -глобіну у 16-й хромосомі людини (кожен ген містить інтрони).

Указано стадії розвитку, на яких відповідні гени активні

Типовою ознакою еукаріотичних геномів є наявність кількох типів генних кластерів, що повторюються багато разів. Серед білкових генів це стосується генів гістонів – структурних білків хроматину (див. наступний підрозділ). Гени п'яти молекул гістонів завжди згруповані у кластер (кожен ген – окрема одиниця транскрипції), який повторюється до 100 разів. Іншим прикладом кластерів, що повторюються, є гени рибосомної РНК (рис. 4.6, тип рибосомних РНК позначають за їхньою константою седиментації, див. розділ 8). У цьому випадку кластер є одиницею транскрипції, первинний транскрипт піддається далі "дозріванню" – частковій деградації з утворенням окремих молекул рРНК. Такий тандемний повтор рибосомного кластера є ядер-

цевим організатором – зоною, де утворюється ядерце (місце збирання рибосом). Гени рРНК ще одного типу – 5S – також тандемно повторюються в іншому місці геному. Також зібрані в тандемні кластери частина генів тРНК і маленьких ядерних РНК.



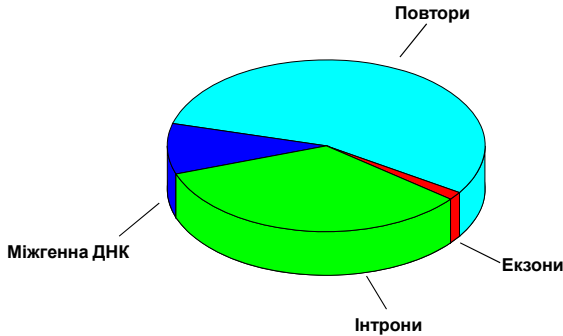
**Рис. 4.6.** Кластер генів рибосомних РНК, який повторюється в зоні ядерцевого організатора. Червона стрілка позначає первинний транскрипт

Загальна кількість білкових генів у геномах вищих еукаріотів становить приблизно 20 тис. (табл. 4.1). Наприклад, у геномі людини їхня кількість трохи більше 20 тис., ще приблизно стільки ж генів кодують РНК, які не піддаються трансляції.

Наведемо як приклад кілька загальних характеристик білкових генів людини:

- середній розмір гена (ураховуючи інтрони) – 100 тис. пар основ (максимум 2,4 млн);
- середній розмір кодувальної частини – 1,4 тис. пар основ (максимум 80 тис.);
- середній розмір екзона – 150 пар основ;
- середня кількість екзонів – 9 (максимум 178);
- середній розмір інтрона – 5 тис. пар основ (максимум 820 тис.).

Як видно з рис. 4.7, кодувальні послідовності цих генів займають лише ~1,5 % геному. Решта геному припадає на міжгенну ДНК (де містяться також регуляторні ділянки й гени РНК), інтрони (~34 %), а більше половини – на послідовності, що повторюються.



**Рис. 4.7.** Відносний вміст нуклеотидних послідовностей різних типів у геномі людини

### **Основні типи повторів у геномах вищих еукаріотів**

1. *Псевдогени* – близько 15 тис. у геномі людини.
2. *Тандемні повтори*. До цього класу можна віднести тандемні повтори генних кластерів, які вже згадувалися вище. Також до них відносять багатократні повтори коротких послідовностей по шість-вісім пар основ у теломерах (TTAGGG у хребетних) і повтори  $\alpha$ -сателітної ДНК у центромерах (довжина повтору варіює від 7 пар основ у дрозофіли до 200 пар у ссавців, у людини – 171 пара основ). По всьому геному розподілені також *прості повтори* (SSR, simple sequence repeats, ~3 % у геномі людини). Зазвичай виділяють *мікросателіти* – 1–15 пар основ, що повторюються від 10 до кількох тисяч разів (до цієї категорії належить і теломерний повтор), і *мінісателіти* – 15–500 пар основ, які повторюються до 100 разів. У геномі людини є принаймні 30 тис. міні- та 200 тис. мікросателітних локусів.
3. *Сегментні дуплікації* – великі блоки завдовжки 1–200 тис. пар основ (~5 % у геномі людини), які є гомологічними. Імовірно, сегментні дуплікації є продуктом порушення хромосом. Частіше зустрічаються в перичентромерних і субтеломерних зонах.
4. *Інтерсперсні (мобільні) елементи* (~44 % у геномі людини), здатні до переміщення і розмноження в межах геному.

Значна частина таких послідовностей є результатом колишньої активності мобільних елементів (таких, що втратили здатність до переміщення), але деякі понині зберігають активність.

**Основні типи мобільних елементів у геномі людини** (механізми їхнього переміщення детальніше розглядатимуться в розділі 10):

- ДНК-транспозони (3 %) – переміщення здійснюється шляхом вирізання ділянки ДНК із подальшим вбудовуванням її в інше місце;
- LTR-ретротранспозони (8 %) – як і для інших двох типів мобільних елементів, переміщення відбувається через проміжну молекулу РНК: на елементі послідовності (який містить довгі кінцеві повтори – Long Terminal Repeats – і кілька генів, зокрема ген зворотної транскриптази) здійснюється транскрипція; молекула РНК прямує до цитоплазми, де відбувається *зворотна транскрипція* – синтез ДНК на РНК-матриці за допомогою зворотної транскриптази (РНК-залежна ДНК-полімераза); ДНК повертається до ядра, де вбудовується в геном;
- мобільні елементи LINE (Long Interspersed Elements, 20 %) містять кілька генів, включаючи ген зворотної транскриптази. На відміну від LTR-ретропозонів, зворотна транскрипція здійснюється у ядрі. У геномі людини є три родини таких елементів, одна з них активна і служить основним джерелом зворотної транскриптази у клітинах;
- мобільні елементи SINE (Short Interspersed Elements, 13 %) – короткі (100–400 пар основ) беззмістовні елементи, які використовують для переміщення ферменти системи LINE. До цього класу належить, зокрема, *Alu*-повтор (від назви відповідної рестриктази, яка здатна специфічно гідролізувати цей елемент послідовності).

Мобільні елементи розподілені в геномі нерівномірно: є довгі ділянки, що на 90 % складаються із них, і такі, де інтерсперсні елементи відсутні. Загалом спостерігається негативна кореляція між щільністю генів і мобільних елементів. Винятком із цієї закономірності є позитивна кореляція між щільністю генів і елементів типу SINE.

**Геном мітохондрій і хлоропластів** – автономний елемент еукаріотичного геному. Циркулярна (зазвичай) молекула мітохондріальної ДНК містить від 6 тис. до 2 млн пар основ і певний набір генів (рРНК, тРНК, деяких білків). Мітохондріальні гени містять інтрони (але не у ссавців). Розмір мітохондріального геному зменшується в ході еволюції. Циркулярна ДНК хлоропластів більша за розміром і містить до 200 тис. пар основ, хлоропластний геном кодує до ~100 білків.

## Молекулярна організація хроматину

Загальна довжина ДНК у ядрі клітини ссавців становить близько 2 м. Така кількість ДНК вимагає її щільної упаковки, яка зумовлює тотальне пригнічення функціональних активностей у більшій частині геному. Однак при цьому упаковка ДНК у клітинному ядрі має дозволяти вибірккову активацію певних ділянок у певні моменти часу. Ці альтернативні завдання розв'язуються завдяки тому, що ДНК існує у клітинному ядрі у вигляді складного нуклеопротейнового комплексу – *хроматину*. Під хроматином можна розуміти й сукупність спадкової речовини всередині ядра, і матеріал, з якого побудовані хромосоми. Тут і далі переважно матимемо на увазі хромосоми під час інтерфази, тобто тоді, коли гени можуть піддаватись транскрипції. Хромосому можна визначити як нуклеопротейновий комплекс, до складу якого входить одна лінійна молекула ДНК. Середня довжина такої молекули для хромосом, наприклад, людини – близько 5 см. Приблизно третину маси хроматину становить ДНК, третину – основні структурні білки хроматину *гістони*, третину – інші хроматинові білки.

На першому рівні організації хроматину ДНК кожної хромосоми практично по всій своїй довжині формує за рахунок взаємодії з білками елементарні утворення – *нуклеосоми* – із середньою щільністю одна нуклеосома на 200 пар основ. Білковий компонент нуклеосоми (*кор*) складається із восьми молекул *корових гістонів* H2A, H2B, H3 і H4 – по дві молекули кожного типу. На вищих рівнях структурної організації така полінуклеосомна

фібрила піддається додатковій компактизації за участю корових гістонів і ще одного (*лінкерного*) гістону H1, а також інших білків. Важливим структурним елементом вищих рівнів організації хроматину є *петельні домени* – петлі, кінці яких жорстко закріплені на особливих білках. Ступінь наднуклеосомної компактизації варіює у різних зонах хроматину: розрізняють менш компактний (потенційно активний) *еухроматин* і більш компактний (репресований) *гетерохроматин*. Утворення гетерохроматину здійснюється у репресованих ділянках геному, а також завжди в ділянках, що містять повтори, – центромерах, теломерах, зонах концентрації мобільних елементів.

### **Нуклеосома**

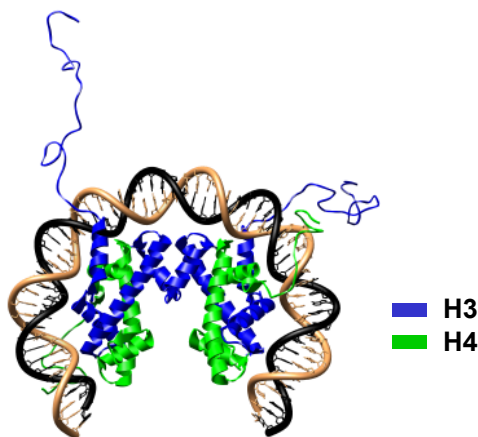
Білковий компонент нуклеосоми становлять гістони, які є одним із найбільш еволюційно консервативних класів білків. Усі корові гістони (містять від 102 до 135 амінокислотних залишків) мають спільну схему будови. У первинній структурі виділяють дві частини: глобулярну та N-кінцеву невпорядковану (хвіст) завдовжки від 20 (H2A) до 40 (H3) амінокислотних залишків (згадайте невпорядковані білки, що розглядалися у розділі 2). Гістон H2A має також помітний C-кінцевий хвіст завдовжки близько 15 залишків. Невпорядковані хвости практично не містять гідрофобних залишків, збагачені позитивно зарядженими амінокислотами і є мішенями для численних посттрансляційних модифікацій (див. нижче).

Глобулярна частина всіх корових гістонів, своєю чергою, також має спільну структуру. Вона має вигляд характерного триспірального гістонового мотиву (*histone fold*), у якому одна довга  $\alpha$ -спіраль фланкована двома короткими (див. рис. 3.16, *a*). Гістон H3 містить також додаткову  $\alpha$ -спіраль із боку N-кінця мотиву ( $\alpha$ N-спіраль), а гістон H2B – додаткову  $\alpha$ C-спіраль. Гістоновий мотив не формує гідрофобного ядра, заекранованого від розчинника, – значна кількість гідрофобних залишків опиняється на поверхні. Унаслідок цього одна молекула корового гістону не може існувати як окремий глобулярний білок у водному середовищі. Мінімальними стабільними структурними одиницями є гетеродимери H2A-H2B та H3-H4 (мають подібну струк-

туру, зображену на рис. 3.16, *a*). Два гістонові мотиви формують у складі димерів щільне гідрофобне ядро, а специфічність формування димерів залежить від наявності додаткових  $\alpha$ N- та  $\alpha$ C-спіралей у гістонах H3 та H2B, відповідно.

На поверхні димеру розташовані три зони скупчення позитивно заряджених амінокислотних залишків, які, відповідно до електростатичного механізму, здатні взаємодіяти з ДНК. Разом ці три сайти створюють платформу для зв'язування ділянки ДНК завдовжки 27–28 пар основ (~ 2,5 витка подвійної спіралі, рис. 3.16, *a*).

Два гетеродимери H3-H4 взаємодіють між собою за рахунок утворення чотириспірального пучка між гістоновими мотивами двох молекул H3. У результаті формується тетрамер  $(H3-H4)_2$  – центральний комплекс у структурі нуклеосоми, який іноді тимчасово може існувати у хроматині (рис. 4.8). Структура тетрамеру нагадує підкову, яка характеризується хіральністю – утворює елемент лівої спіралі. Вісь симетрії тетрамерного комплексу (яка водночас є і віссю симетрії всієї нуклеосоми) проходить через інтерфейс між двома молекулами H3.



**Рис. 4.8.** Тетрамер гістонів  $(H3-H4)_2$  у комплексі з ДНК у складі нуклеосоми (1KX5)

За рахунок взаємодії гістонів H4 і H2B "підкова" тетрамеру симетрично продовжується двома димерами H2A-H2B в обидва боки, результатом чого є утворення октамеру  $(H2A-H2B-H3-H4)_2$ . Отже, глобулярні частини гістонів утворюють октамерний комплекс, який і служить білковим кором нуклеосоми (рис. 4.9). На глобулярній поверхні октамеру існує своєрідний трек позитивно заряджених амінокислотних залишків, який використовується для взаємодії з нуклеосомною ДНК завдовжки 145–147 пар основ.

Окрема нуклеосома – октамер гістонів + 145 пар основ ДНК – може бути вилучена із хроматину за допомогою мікрококової нуклеази: вона робить дволанцюговий розріз у ДНК, а оскільки один із ланцюгів нуклеосомної ДНК завжди взаємодіє з гістонами (рис. 4.9), то доступною для нуклеази є лише ДНК за межами нуклеосоми. У хроматині *вся* ДНК формує нуклеосоми із середньою щільністю одна нуклеосома на 200 пар основ, сусідні нуклеосоми з'єднані міжнуклеосомними *лінкерними* (linker) ділянками. Нуклеосомна ДНК разом із лінкерною ділянкою становлять *нуклеосомний повтор*, довжина якого (середнє значення 200 пар основ) варіює як уздовж полінуклеосомного ланцюга, так і залежно від функціонального стану, типу клітин тощо.

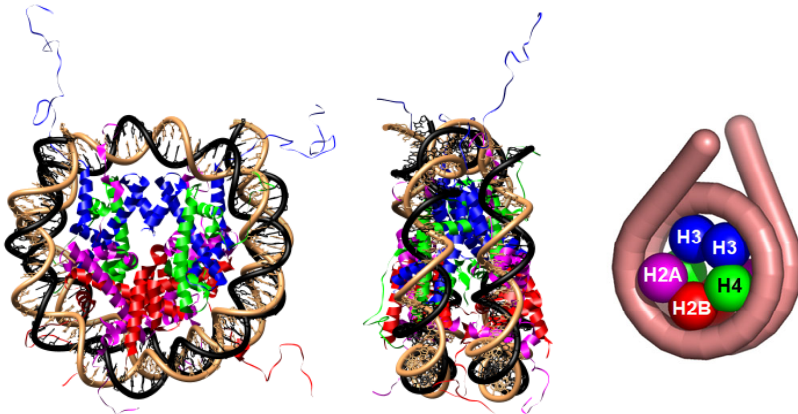


Рис. 4.9. Структура нуклеосоми у двох проєкціях (1KX5) і схема цієї структури

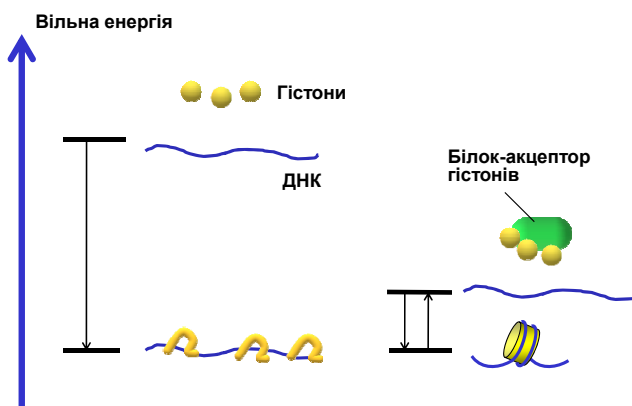
Вісь симетрії нуклеосоми проходить через центральну точку нуклеосомної ДНК, у якій великий жолобок подвійної спіралі повернутий до поверхні октамеру гістонів (рис. 4.9). ДНК-гістонові взаємодії здійснюються в позиціях, де маленький жолобок контактує з поверхнею октамеру: саме тут реалізуються взаємодії ДНК із позитивно зарядженими сайтами на поверхні гістонових мотивів (порівн. рис. 3.16, а; 4.8; 4.9). Як видно з рис. 4.9, нуклеосомна ДНК значно вигнута на поверхні октамеру гістонів і формує майже ідеальну ліву суперспіраль із радіусом  $\sim 4,2$  нм і кроком  $\sim 2,6$  нм. Зрозуміло, що значний вигин нуклеосомної ДНК потребує енергетичних витрат, які компенсуються ДНК-гістоновими взаємодіями.

Структура нуклеосомної ДНК зумовлена її взаємодією із глобулярною частиною октамеру гістонів. N-кінцеві хвости гістонів H3 та H2B виходять за межі нуклеосоми через канали, сформовані маленькими жолобками двох дуплексів сусідніх витків нуклеосомної суперспіралі (рис. 4.9). Ділянки хвостів, безпосередньо розташованих у каналах, позитивно заряджені й, отже, додатково скріплюють сусідні витки суперспіралі. N-кінцеві хвости гістонів H4 і H2A також взаємодіють із маленьким жолобком зовні нуклеосомної суперспіралі. Особливе місце входу / виходу нуклеосомної ДНК (рис. 4.9) займає найдовший N-кінцевий хвіст гістону H3. Позитивно заряджений хвіст H3 стабілізує структуру нуклеосоми в цій зоні, де спостерігається найвища щільність негативних зарядів ДНК.

Значна частина гістонових хвостів виходить за межі нуклеосоми. Завдяки структурній лабільності вони беруть участь в організації хроматину на наднуклеосомному рівні, а також служать платформою для зв'язування різноманітних білків, спорідненість яких залежить від посттрансляційних модифікацій хвостів (див. нижче).

Електростатичні взаємодії між ДНК і гістонами за фізіологічних умов дуже міцні – остаточне руйнування нуклеосоми *in vitro* відбувається при концентрації солі 2 моль/л. Отже, за фізіологічної іонної сили (приблизно 150 ммоль/л) зв'язування гістонів з ДНК є необоротним (неможлива рівновага між зв'язаними й дисоційованими гістонами). Хоча структура нуклеосоми

відповідає мінімуму вільної енергії, цього мінімуму неможливо досягти за розумний проміжок часу: при змішуванні *in vitro* гістони швидко й безладно зв'язуються з ДНК без подальшої дисоціації. Для реалізації рівноважних умов ДНК-гістонових взаємодій *in vivo* у хроматині існують *проміжні акцептори* гістонів, що виконують роль факторів збирання / руйнування нуклеосом. Ці негативно заряджені білки мають спорідненість до гістонів дещо меншу, ніж спорідненість гістонів до ДНК (рис. 4.10), що й забезпечує можливість рівноважного обміну гістонами між ДНК і проміжними акцепторами зі зсувом цієї рівноваги в бік ДНК-гістонових комплексів. Наявність проміжних акцепторів гістонів робить можливим обмін димерами H2A-H2B між різними нуклеосомами: тимчасове видалення димерів є важливим шляхом структурної динаміки хроматину.



**Рис. 4.10.** Роль проміжних акцепторів гістонів у збиранні нуклеосоми: дисоціація гістонів практично заборонена за фізіологічної іонної сили; у присутності акцептора взаємодія гістонів з ДНК стає рівноважною

Незважаючи на те, що ДНК-гістонові взаємодії в нуклеосомі є неспецифічними щодо послідовності пар основ ДНК (електростатичні взаємодії між фосфатами ДНК і позитивно зарядженими амінокислотними залишками), у хроматині спостерігається феномен *переважного позиціонування* нуклеосом щодо послідов-

ності. Основною причиною такого позиціонування є те, що здатність до деформацій, які супроводжують формування нуклеосоми, суттєво залежить від нуклеотидної послідовності ДНК (див. розділ 3). У результаті нуклеосома більш імовірно обирає таку ділянку, де вигин ДНК потребує менших енергетичних витрат. Це зумовлює важливі функціональні наслідки – диференційне експонування ділянок ДНК до дії регуляторних факторів (розділ 6). При функціонуванні хроматину виникає необхідність у змінах такого експонування / екранування певних ділянок ДНК – репозиціонуванні нуклеосом. Міцність ДНК-гістонових взаємодій не дозволяє спонтанного зсуву нуклеосом уздовж ДНК, що зумовлює потребу в особливих АТР-залежних молекулярних пристроях, які здійснюють таке репозиціонування, – факторах ремоделювання хроматину (розділ 6), що також можуть виступати факторами збирання / руйнування нуклеосом.

### **Посттрансляційні модифікації гістонових хвостів**

Невпорядковані хвости корових гістонів є субстратом для ковалентних посттрансляційних модифікацій, яким піддаються конкретні (такі, що мають конкретну позицію у складі поліпептидного ланцюга) амінокислотні залишки певного типу.

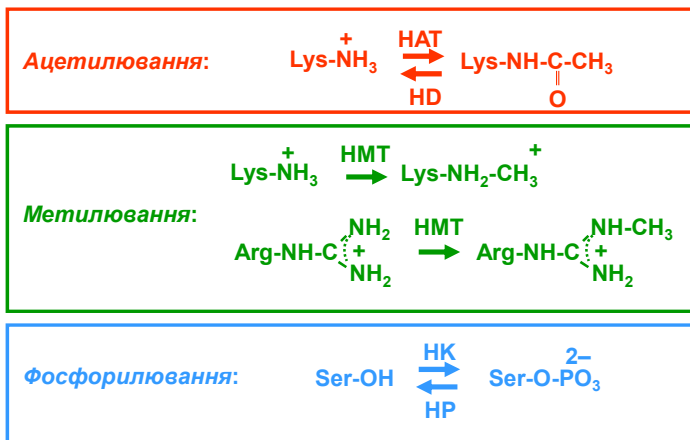
Найбільш важливі типи цих модифікацій (рис. 4.11, 4.12) наведено нижче.

**Ацетилювання** залишків Lys – перенесення залишку оцтової кислоти (ацетату) на аміногрупу Lys. Реакція каталізується *гістон-ацетилтрансферазами* (НАТ, Histone Acetyl Transferase), різні їхні типи мають певну специфічність щодо залишків-мішеней. Результатом реакції є зникнення позитивного заряду на залишку Lys (рис. 4.11). Ацетилювання є динамічною модифікацією: гістон-деацетилази (HD, Histone Deacetylase, позначаються також HDAC) здійснюють відщеплення ацетатних залишків. Два типи ферментів-антагоністів підтримують певний динамічний гомеостаз ацетилювання / деацетилювання, зсунутий у той чи інший бік у певних ділянках хроматину.

**Метилювання** залишків Lys та Arg – перенесення метильної групи на аміногрупу Lys або гуанідинову групу Arg (одна, дві або три метильні групи можуть переноситись на Lys, одна

або дві – на Arg), позитивний заряд при цьому залишається. Реакція каталізується гістон-метилтрансферазами (НМТ, Histone Methyl Transferase), кожна з яких є специфічною щодо конкретного залишку-мішені. Існують також ферменти-антагоністи – гістонові деметилази, але їхня активація відбувається внаслідок достатньо складних регуляторних процесів. Відповідно, на відміну від ацетилювання, метилювання є дуже стабільною у часі модифікацією.

**Фосфорилування** залишків Ser – короткострокова модифікація, що полягає у перенесенні фосфатного залишку на ОН-групу Ser. Реакція каталізується гістон-кіназою (НК, Histone Kinase), відщеплення фосфату – фосфатазою. Внесення негативного заряду внаслідок фосфорилування (рис. 4.11) не дуже суттєве, оскільки сайтів фосфорилування небагато (рис. 4.12).



**Рис. 4.11.** Основні посттрансляційні модифікації амінокислотних залишків у гістонах (показано лише хімічні групи у складі залишків, які безпосередньо модифікуються).

НАТ – гістон-ацетилтрансфераза, HD – гістон-деацетилаза,

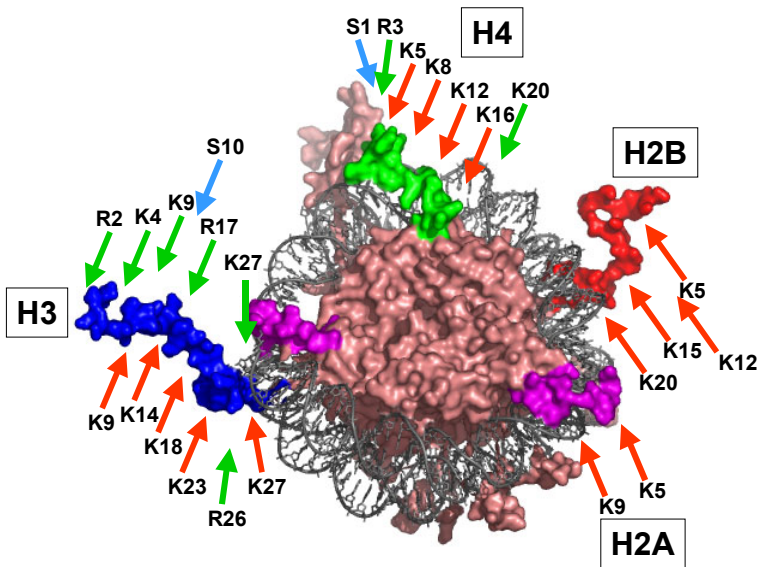
НМТ – гістон-метилтрансфераза, НК – гістон-кіназа,

HP – гістон-фосфатаза

Серед усіх модифікацій лише про ацетилювання можна сказати, що воно чітко корелює із транскрипційною активністю: гіперацетилювані гістони наявні в активних ділянках хромати-

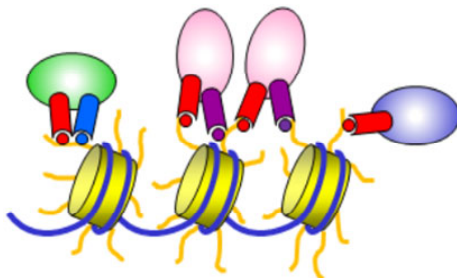
ну, у репресованих підтримується деацетилований статус. Інші модифікації впливають на функціональний стан складніше: метилювання одного залишку може викликати репресію транскрипції, іншого (через кілька амінокислот від першого в тому самому хвості) – супроводжувати активацію; фосфорилування H3-Ser10 корелює з активацією транскрипції, але також супроводжує гіперконденсацію хроматину при переході до мітозу.

Оскільки ацетилювання численних залишків Lys (рис. 4.12) зумовлює суттєве пониження позитивного заряду хвостів, то лише воно може безпосередньо впливати на взаємодію хвостів із ДНК. Це не викликає змін структури нуклеосоми (яка практично не залежить від хвостів), але знижує її стабільність і сприяє декомпактизації полінуклеосомного ланцюга.



**Рис. 4.12.** Кінцеві хвости гістонів у структурі нуклеосоми (1KX5) та їхні модифікації. Кольором виділено по одному N-кінцевому хвосту кожного гістону, а також С-кінцевий хвіст гістону H2A (поряд із N-кінцем H3). Стрілочками вказано ацетилювання (червоні), метилювання (зелені) і фосфорилування (блакитні) залишків певного типу, які займають певні позиції в поліпептидних ланцюгах

Однак головним механізмом впливу модифікованих гістонових хвостів на функціональний стан хроматину є **специфічне впізнання модифікованих хвостів іншими білками**: регуляторами, факторами транскрипції, ферментами тощо (рис. 4.13). Наприклад, ацетилювані залишки Lys упізнаються особливим структурним блоком таких білків – *бромодоменом*. Інший блок – *хромодомен* – упізнає метильований Lys. Гістонові хвости служать своєрідними платформами для збирання різноманітних білкових комплексів, склад яких залежить від патерну модифікацій – розподілу певних модифікованих груп по хвостах. Співвідношення між патерном модифікацій і набором білків, які впізнають такий патерн, що, своєю чергою, має певні функціональні наслідки, називають *гістоновим кодом*.



**Рис. 4.13.** Принцип реалізації гістонового коду – упізнання модифікованих залишків гістонових хвостів структурними елементами певних білків

Наприклад, фосфорилювання H3-Ser10 із одночасним ацетилюванням H3-Lys14 і метилюванням H3-Lys4 супроводжується активацією транскрипції, деацетилюванням H3-Lys14 із одночасним деметилюванням H3-Lys4 і метилюванням H3-Lys9 – репресією. Це лише невеличкий елемент такого коду: зрозуміло, що різноманітних комбінацій модифікованих залишків на восьми гістонових хвостах (рис. 4.12) може бути велика кількість. Специфічні комбінації модифікацій (і білків, що їх впізнають) характерні для ділянок гетеро- й еухроматину, певні особливості спостерігаються для інтронів і екзонів активних генів, стартових точок транскрипції тощо. Деякі за-

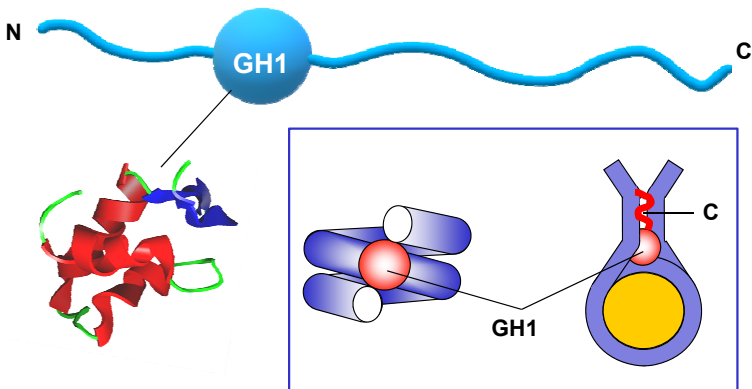
кономічності та приклади залучення гістонового коду до регуляції транскрипції наведені у розділі 6.

Слід також зауважити, що геномні профілі гістонових модифікацій передаються у дочірні клітини при клітинному поділі (розділ 9), що є одним із головних механізмів *епігенетичної спадковості* – успадкування у клітинних поколіннях не тільки генетичної інформації у вигляді нуклеотидних послідовностей ДНК, а й програми реалізації цієї інформації. Власне, модифікації гістонів (разом із метилюванням ДНК, про яке йтиметься у розділі 6) так і називають – *епігенетичними маркерами*.

### **Гістон H1 і компактизація хроматинової фібрили**

Загальний вміст у хроматині п'ятого гістону (лінкерного гістону H1) становить не більше однієї молекули на одну нуклеосому (співвідношення H1/нуклеосома дорівнює 1 чи трохи більше у термінально диференційованих клітинах; у стовбурових та інших більш транскрипційно активних клітинах воно близьке до 0,5). На відміну від корових гістонів, взаємодія H1 із ДНК є динамічною: спостерігається швидкий обмін лінкерних гістонів між хроматином та їхнім пулом у ядрі – можна сказати, що молекули H1 *узагальнені* всіма ділянками хроматину.

За структурою гістон H1 (мономерний білок) суттєво відрізняється від корових гістонів (рис. 4.14). Молекула H1 містить N-кінцевий неупорядкований хвіст, глобулярний домен GH1 і довгий, збагачений позитивно зарядженими залишками (насамперед Lys) C-кінцевий хвіст завдовжки приблизно 100 амінокислотних залишків (приблизно половина молекули). Глобулярний домен взаємодіє з нуклеосомною ДНК на виході з нуклеосоми, додатково стабілізуючи нуклеосомну суперспіраль (рис. 4.14). Позитивно заряджений C-кінцевий хвіст H1 сприяє об'єднанню двох лінкерів (завдовжки по 10–30 пар основ), що виходять із нуклеосоми, у стеблоподібну структуру (рис. 4.14), яка зумовлює наближення одна до одної сусідніх нуклеосом.

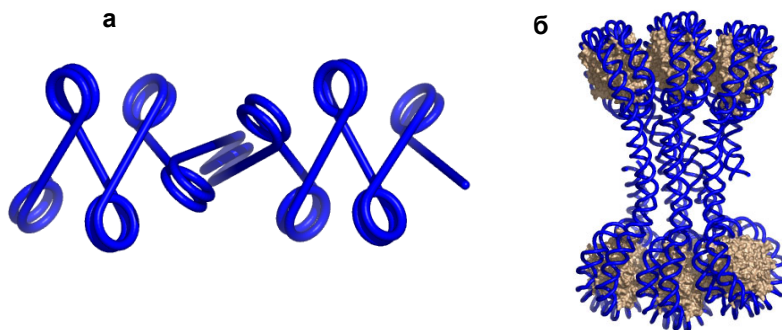


**Рис. 4.14.** Схема будови гістону H1 і структура його глобулярного домену GH1 (1GHC).  
На вставці – схема взаємодії гістону H1 з нуклеосою

У хроматині нуклеосоми з'єднані лінкерами завдовжки  $\sim 50$  пар основ. Якщо лінкер (у полінуклеосомному ланцюзі без гістону H1) просто продовжує хід нуклеосомної ДНК по прямій, то нуклеосоми у складі полінуклеосомного ланцюга мають бути розташовані зигзагом (рис. 4.15, *a*). Зрозуміло, що достатня гнучкість лінкерів робить такий зигзаг конформаційно рухливим, а середній діаметр фібрили дорівнює приблизно 10 нм (розмір нуклеосоми). Саме такий зигзагоподібний вигляд і має декомпактизована (якщо відсутній H1 і за низької іонної сили) полінуклеосомна фібрила під мікроскопом (електронним чи атомно-силовим).

Полінуклеосомний зигзаг може конденсуватися з утворенням фібрили діаметром 30 нм (рис. 4.15, *б*). **Умова конденсації** (необхідна, але недостатня) – фізіологічна іонна сила (крім одновалентних мають бути наявні двовалентні катіони) для зниження електростатичного відштовхування між нуклеосомами. **Фактор конденсації** (її рушійна сила) – неупорядковані хвости корових гістонів: лабільні позитивно заряджені хвости ефективно "зшивають" фібрилу, взаємодіючи з ДНК сусідніх нуклеосом і лінкерів. При цьому за наявності гістону H1 компактна хроматинова фібрила завтовшки 30 нм стає значно стабільнішою за рахунок як

додакових позитивно заряджених хвостів Н1, так і стебла із двох лінкерних ділянок, яке сприяє наближенню сусідніх нуклеосом.

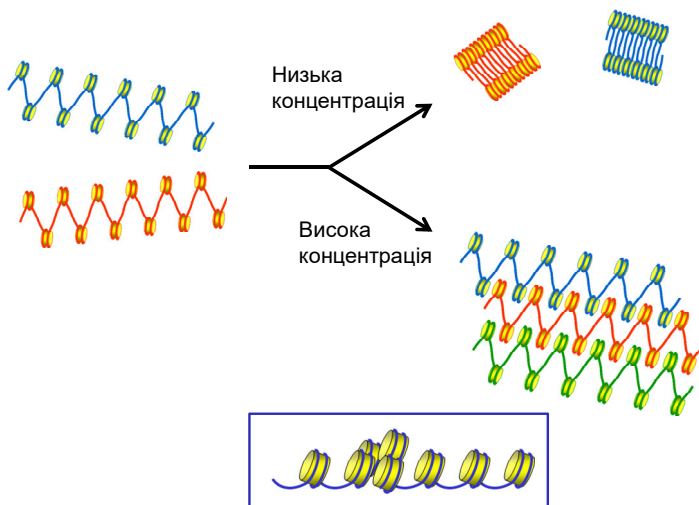


**Рис. 4.15.** Ідеалізована зигзагоподібна конфігурація полінуклеосомного ланцюга (а) і структура однієї з можливих регулярних компактних форм такого ланцюга, що складається із шести нуклеосом (б, 6НКТ)

Отже, суперструктура конденсованої фібрили діаметром 30 нм – це тривимірний зигзаг нуклеосом, з'єднаних лінкерами, які спрямовані всередину фібрили. Один із можливих варіантів такої регулярної структури зображено на рис. 4.15, б. Проте отримати таку структуру можна тільки *in vitro*. По-перше, для того щоб структура конденсованої фібрили була регулярною, усі лінкери повинні мати однакову довжину. Це можна зробити шляхом реконструкції нуклеосом на ДНК, що складається з однакових тандемних повторів деякої нуклеотидної послідовності, таких, що в межах кожного повтору розташовується одна нуклеосома у визначеній позиції. Однак у хроматині лінійне розташування нуклеосом зазвичай не є строго регулярним.

По-друге, і це більш важлива обставина, фібрилу діаметром 30 нм можна отримати *in vitro* за низької концентрації не надто довгих полінуклеосомних ланцюгів, інакше відбувається конденсація не всередині фібрили, а між різними фібрилами (або між віддаленими ділянками однієї довгої фібрили) за рахунок тих самих електростатичних взаємодій гістонових хвостів із ДНК (рис. 4.16). Саме така ситуація має спостерігатись у клітинному ядрі, де містяться дуже довгі фібрили у дуже високій концентрації (за даними електронномікроскопічних досліджень основ-

ним матеріалом усередині ядра є фібрили діаметром  $\sim 10$  нм). Найвні також невеличкі скупчення по  $\sim 5$ – $6$  сусідніх по ланцюгу нуклеосом (рис. 4.16). Кількість таких скупчень підвищується у термінально диференційованих клітинах, що корелює зі зростанням там кількості молекул гістону H1.



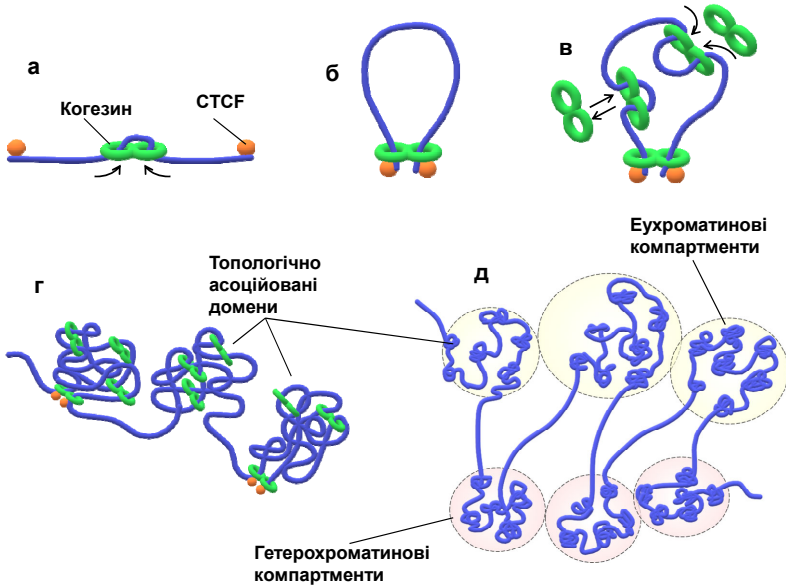
**Рис. 4.16.** Внутрішньоланцюгова або міжланцюгова конденсація хроматину залежно від його концентрації.  
На вставці – скупчення сусідніх по ланцюгу нуклеосом

Отже, основний матеріал інтерфазного хроматину – полінуклеосомний ланцюг, який, власне, і формує хроматинову фібрилу. Ця фібрила піддається компактизації різного ступеня за рахунок віддалених по ланцюгу взаємодій у межах кожної хромосоми. Зниження ступеня конденсації в еухроматинових зонах забезпечується ацетилюванням (зниженням позитивного заряду) гістонових хвостів і збідненням на гістон H1. Підвищення ступеня компактизації у гетерохроматинових зонах забезпечується додатковими факторами компактизації – особливими білками. При цьому в зонах обох типів реалізується додатковий механізм компактизації – утворення петельних доменів, яке, зокрема, дозволяє запобігти переплутуванню довгих полінуклеосомних фібрил.

### **Петельні й топологічно асоційовані домени хроматину**

Границі петель хроматину фіксуються двома білками: ДНК-зв'язувальним білком CTCF (CCCTC-binding factor) і білковим комплексом когезином (cohesin). Білок CTCF містить цинкові пальці (розділ 3), що впізнають певний мотив нуклеотидної послідовності (у геномі людини ~40 тис. таких мотивів), який визначає границю петлі. Спорідненість білка до мотиву знижується за умови метилювання останнього, яке варіює у клітинах різних типів (детальніше про метилювання ДНК йтиметься у розділі 6). Когезин, який здатен взаємодіяти із CTCF, є комплексом певних білків, що формує замкнене кільце із внутрішнім діаметром ~35 нм (цілком достатнім для того, щоб там могла розміститись полінуклеосомна фібрила). Кільце когезину може відкриватись у АТР-залежний спосіб за допомогою кофакторів, після його замикання хроматинова фібрила опиняється всередині. При цьому сам когезиновий комплекс має АТРазну активність, тобто є молекулярною машиною, що використовує енергію гідролізу АТР для транслокації уздовж фібрили.

Найімовірніше, два когезинових кільця, що взаємодіють між собою, завантажуються на хроматинову фібрилу, після чого починається процес *екструзії* – унаслідок транслокації когезинів у протилежних напрямках петля збільшується (рис. 4.17, *a*). Цей процес зупиняється, коли когезиновий комплекс зустрічається з білками CTCF, які служать своєрідними стоп-сигналами (рис. 4.17, *b*). Взаємодія когезинів із CTCF робить петлю достатньо стабільною – у геномі людини таких петель (медіанний розмір 185 тис. пар основ) налічується до 10 тис. Штучне видалення із клітини або АТР, або когезину, або CTCF спричиняє зникнення петельних доменів. Процес екструзії петлі автоматично виключає утворення вузлів усередині неї, можливі вузли поза межами петлі видаляються завдяки активності ДНК-топоізомераз II (розділ 3), що концентруються на границях петель.



**Рис. 4.17.** Структурна організація хроматину на високих рівнях: а – завантаження комплексу двох кілець когезину на хроматинову фібрилу і початок екструзії петлі (стрілками показано напрямок руху фібрили крізь когезин); б – зустріч когезину з білками СТСФ зупиняє процес екструзії, унаслідок чого формується стабільна петля; в – динамічна екструзія тимчасових петель зумовлює підвищення частоти контактів між будь-якими точками; г – унаслідок динамічної екструзії як у межах петель, так і поза ними утворюються топологічно асоційовані домени, збагачені контактами всередині себе; д – сусідні топологічно асоційовані домени об'єднуються в еу- й гетерохроматиніві компартменти

Процес екструзії є насправді динамічним – когезиновий комплекс може час від часу завантажуватись на хроматинову фібрилу, здійснювати транслокацію і зніматись, якщо не зустріне СТСФ (рис. 4.17, в). У результаті різні точки в основах тимчасових петель доволі часто входять у контакт між собою. Це відбувається як усередині петельних доменів, обмежених білками СТСФ, так і поза їхніми межами (рис. 4.17, г), унаслідок чого утворюються *топологічно асоційовані* (або *контактні*) домени (Topologically Associating (Contact) Domains, TADs). Отже, топологічно асоційований домен – це зона хроматину (типовий

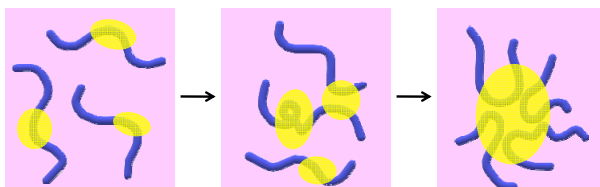
розмір до ~200 тис. пар основ), усередині якої рееструються часті контакти між будь-якими двома точками, що належать такому домену. Функціональні наслідки існування топологічно асоційованих доменів розглядатимуться у розділі 6.

### Компартменти хроматину

На наступному рівні структурної організації сусідні по ланцюгу топологічно асоційовані домени об'єднуються у *хроматинові компартменти* – зони з типовим розміром ~5 млн пар основ (рис. 4.17, д). Слід зауважити, що у цьому випадку під компартментом мають на увазі частину простору клітинного ядра, не обмежену мембранами. Є два основні типи таких компартментів: менш компактні еухроматинові й більш компактні гетерохроматинові. При цьому однойменні компартменти схильні контактувати між собою, різнойменні – уникають таких контактів.

Механізм виникнення компартментів не пов'язаний із петельними або топологічно асоційованими доменами. У його основі лежить зчитування різними наборами білків різних (еу- й гетерохроматинових) епігенетичних маркерів. Частина цих білків визначає менш компактний стан еухроматину, частина – є факторами додаткової компактизації гетерохроматину (приклад таких білків будуть розглянуті у розділі 6).

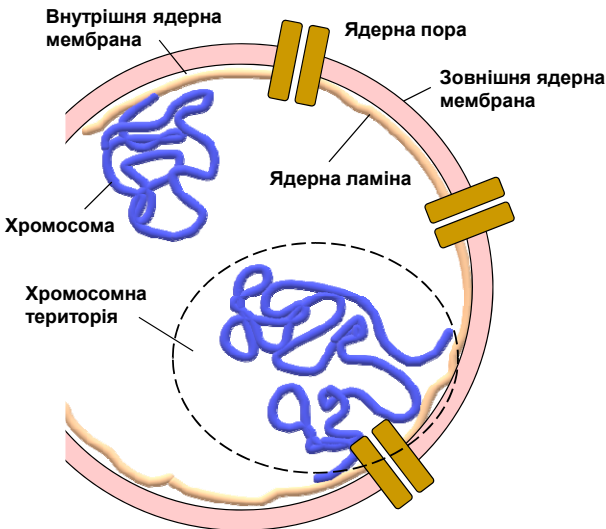
Взаємодія між однотипними наборами білків зумовлює збільшення частоти контактів між однойменними компартментами й *розділення рідких фаз* (liquid phase separation) у просторі ядра. Наприклад, взаємодія між факторами додаткової компактизації гетерохроматину спричинює об'єднання відповідних компартментів у гетерохроматинову мікрокраплю (рис. 4.18). Така крапля є рідкою, усередині відбувається рух усіх її елементів, але при цьому вона відділена від зовнішнього оточення.



**Рис. 4.18.** Зростання мікрокраплі гетерохроматину за рахунок взаємодії між факторами додаткової компактизації (зони скупчення таких факторів показано жовтим кольором)

## Організація хроматину в масштабі клітинного ядра

Усередині клітинного ядра (на відміну від цитоплазми) практично відсутня розгалужена мережа білкових скелетних структур, однак на його периферії розташована скелетна структура, асоційована із внутрішньою ядерною мембраною – *ядерна ламіна*. Білкові філаменти ламіни (належать до категорії проміжних філаментів цитоскелета) протягнуті від однієї ядерної пори до іншої і позиціонують порові комплекси у площині мембрани (рис. 4.19). Від ламіни всередину ядра можуть відгалужуватись білкові філаменти, але доволі короткі (зокрема актинові). Подвійна ядерна мембрана, ядерні пори й ламіна утворюють єдину систему – *ядерний конверт*.



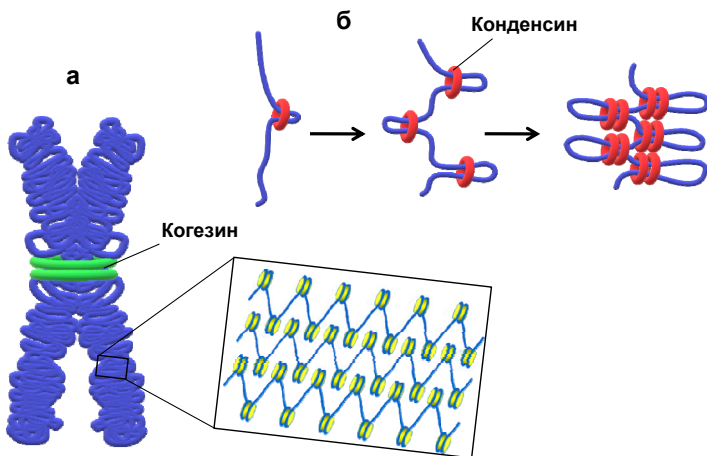
**Рис. 4.19.** Загальна схема структурної організації клітинного ядра

З ламіною взаємодіє значна частина гетерохроматину, зокрема центромери й теломери хромосом, що робить внесок у згадане вище розділення фаз еу- й гетерохроматину. Отже, зазвичай еухроматинові компартменти мають тенденцію розта-

шовуватися ближче до центра ядра, гетерохроматинові – ближче до периферії. У результаті фіксації окремих зон на ламіні хромосома займає певну зону в об'ємі ядра – *хромосомну територію* (рис. 4.19). Наявність хромосомних територій має важливі функціональні наслідки. З одного боку, хромосомна територія не є чітко визначеною – сусідні території частково перекриваються. Це створює можливість для розміщених там хромосом взаємодіяти, а отже, впливати одна на одну при здійсненні функціональних процесів, що на цих хромосомах відбуваються. І навпаки, розміщення хромосом у віддалених одна від одної територіях виключає можливість такого взаємного впливу. Розподіл хромосом по хромосомних територіях відрізняється у клітинах різних типів.

### **Мітотична хромосома**

Усе, про що йшлося вище, стосується організації хроматину та хромосом у інтерфазному ядрі. Однак варто сказати кілька слів і про структурну організацію надкомпактної (і функціонально інертної) мітотичної хромосоми, відкритої ще у XIX столітті, численні зображення якої всім добре відомі. У процесі реплікації ДНК (розділ 9) у S-фазі клітинного циклу утворюються дві дочірні копії кожної хромосоми. Ці дві копії (що їх за традицією, яка йде із XIX століття, прийнято називати хроматидами) утримуються поряд одна з одною завдяки знайомому вже когезину – кільця когезинових комплексів оточують обидві копії. Коли клітина вступає до мітозу, починається надкомпактизація обох дочірніх копій хромосоми. При цьому когезини видаляються скрізь, крім центромерної зони, і саме завдяки цьому по завершенню надкомпактизації можна бачити перетяжку в ділянці двох центромер (рис. 4.20, *a*). Пізніше у певний момент мітозу ця когезинова перетяжка руйнується, що дає змогу розвести дві дочірні хромосоми до полюсів клітини.



**Рис. 4.20.** Схема організації мітотичної хромосоми (двох дочірніх хромосом у надкомпактному стані, а) і процесу надкомпактизації завдяки екструзії петель крізь конденсини (б)

Основний механізм надкомпактизації мітотичної хромосоми дуже схожий на розглянутий вище процес утворення петель у інтерфазному хроматині. Тут також відбувається екструзія петель, але крізь дещо інші білкові комплекси – *конденсини*, які складаються з білків, споріднених до тих, що входять до складу когезинів. Так само як когезини, конденсини здійснюють транслокацію в АТР-залежний спосіб. Велика кількість конденсинових комплексів завантажується на хроматинову фібрилу, і починається екструзія, яка гальмується, коли один конденсин зустрічає інший (рис. 4.20, б). У результаті суттєво скорочується лінійний розмір хромосоми, а конденсини накопичуються уздовж осі цієї паличкоподібної структури. Як і у випадку інтерфазного хроматину, матеріалом, з якого "зроблена" мітотична хромосома, залишається полінуклеосомний ланцюг (фібрила діаметром ~10 нм), що упакується завдяки віддаленим по ланцюгу взаємодіям гістонових хвостів із ДНК, а також за участю деяких додаткових факторів компактизації.

## Контрольні запитання

1. Дайте визначення кодону і відкритої рамки зчитування. Скільки існує кодонів? Які позиції нуклеотидів у складі кодону є найбільш і найменш суттєвими для визначення його змісту?

2. Що таке ген?

3. Що таке геном? У чому полягає найсуттєвіша відмінність між геномами про- й еукаріотів?

4. У чому полягає перекриття генів у вірусів і еукаріотів?

5. Дайте визначення інтрона й екзона.

6. Яка різниця між кластером генів і опероном?

7. Назвіть основні типи послідовностей ДНК, що повторюються.

8. З яких елементів складається нуклеосома? Яке місце в її структурі займають білки? Як організована у просторі ДНК у нуклеосомі?

9. Які взаємодії стабілізують нуклеосому? За яким принципом відбувається збирання нуклеосом?

10. У чому полягає механізм позиціонування нуклеосом щодо послідовності ДНК?

11. Назвіть основні типи посттрансляційних модифікацій гістонів. Які амінокислотні залишки піддаються модифікаціям? У яких частинах молекул гістонів ці залишки розташовані?

12. Як організована хроматинова фібрила у просторі? За яким основним механізмом фібрила піддається компактизації? У чому полягає функція гістону H1?

13. Як утворюються петельні й топологічно асоційовані домени хроматину?

14. Що таке компартменти хроматину? Як вони утворюються?

15. Опишіть загальну схему структурної організації клітинного ядра і хромосом у його складі.

16. Чим відрізняється мітотична хромосома від інтерфазної? Як здійснюється надкомпактизація мітотичної хромосоми?

## Рекомендована література

1000 Genomes Project Consortium. (2015) A global reference for human genetic variation. *Nature* 526, 68–74.

Akey, C.W. and Luger, K. (2003) Histone chaperones and nucleosome assembly. *Curr. Op. Struct. Biol.* 13, 6–14.

Bednar, J., Garcia-Saez, I., Boopathi, R., et al. (2017) Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. *Mol. Cell* 66, 384–397.

Brown, T.A. (2017) *Genomes 4*. 4<sup>th</sup> ed. New York : Garland Science.

Dekker, J. and Mirny, L. (2016) The 3D genome as moderator of chromosomal communication. *Cell* 164, 1110–1121.

Dekker, J. and Misteli, T. (2015) Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.* 7, a019356.

Fritz, A.J., Sehgal, N., Pliss, A., et al. (2019) Chromosome territories and the global regulation of the genome. *Genes Chromosomes Cancer* 58, 407–426.

Garcia-Saez, I., Menoni, H., Boopathi, R., et al. (2018) Structure of an H1-bound 6-nucleosome array reveals an untwisted two-start chromatin fiber conformation. *Mol. Cell* 72, 902–915.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., et al. (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17, 669–681.

Goloborodko, A., Marko, J.F. and Mirny, L.A. (2016) Chromosome compaction by active loop extrusion. *Biophys. J.* 110, 2162–2168.

Gruenbaum, Y., Margalit, A., Goldman, R.D., et al. (2005) The nuclear lamina comes of age. *Nat. Rev.* 6, 21–31.

International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

Koonin, E.V. (2011) *The logic of chance: the nature and origin of biological evolution*. New Jersey : Pearson Education.

Kornberg, R.D. and Lorch, Y. (2020) Primary role of the nucleosome. *Mol. Cell* 79, 371–375.

Krietenstein, N., Abraham, S., Venev, S.V., et al. (2020) Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* 78, 554–565.

Luger, K., Dechassa, M.L. and Tremethick, D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell. Biol.* 13, 436–447.

Luger, K. and Richmond, T. J. (1998) DNA binding within the nucleosome core. *Curr. Op. Struct. Biol.* 8, 33–40.

Luger, K. and Richmond, T.J. (1998) The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* 8, 140–146.

Naumova, N., Imakaev, M., Fudenberg, G., et al. (2013) Organization of the mitotic chromosome. *Science*. 342, 948–953.

Ohno, M., Priest, D.G. and Taniguchi, Y. (2018) Nucleosome-level 3D organization of the genome. *Biochem. Soc. Trans.* 46, 491–501.

Parmar, J.J. and Padinhateeri, R. (2020) Nucleosome positioning and chromatin organization. *Curr. Opin. Struct. Biol.* 64, 111–118.

Pearson, H. (2006) Genetics: what is a gene? *Nature* 441, 398–401.

Racko, D., Benedetti, F., Goundaroulis, D. and Stasiak, A. (2018) Chromatin loop extrusion and chromatin unknotting. *Polymers* 10, 1126.

Rao, S.S., Huntley, M.H., Durand, N.C., et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.

Rothbart, S.B. and Strahl, B.D. (2014) Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta.* 1839, 627–643.

Saier, M.H. Jr. (2019) Understanding the genetic code. *J. Bacteriol.* 201, e00091–19.

Sherman, R.M. and Salzberg, S.L. (2020) Pan-genomics in the human genome era. *Nat. Rev. Genet.* 21, 243–254.

Sivolob, A. and Prunell, A. (2004) Nucleosome conformational flexibility and implications for chromatin dynamics. *Phil. Trans. Roy. Soc. Lond. A* 362, 1519–1547.

Song, F., Chen, P., Sun, D., et al. (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 344, 376–380.

Strom, A.R., Emelyanov, A.V., Mir, M., et al. (2017) Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245.

The ENCODE Project Consortium. (2007) Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

Vian, L., Pękowska, A., Rao, S.S.P., et al. (2018) The energetics and physiological impact of cohesin extrusion. *Cell* 173, 1165–1178.

Wolffe, A.P. (2012) *Chromatin: structure and function*. 3d ed. London : Academic Press.

Wu, H.-L., Bagby, S. and van den Elsen J.M.H. (2005) Evolution of the genetic triplet code via two types of doublet codons. *J. Mol. Evol.* 61, 54–64.

Zhang, T., Cooper, S. and Brockdorff, N. (2015) The interplay of histone modifications – writers that read. *EMBO Rep.* 16, 1467–1481.

## Розділ 5

# ТРАНСКРИПЦІЯ

Down, down, down. There was nothing else to do.  
*Lewis Carroll. Alice's adventures in Wonderland*

Транскрипція (transcription) – процес синтезу РНК із використанням одного з ланцюгів ДНК як матриці, тобто "переписування" послідовності нуклеотидів ДНК у послідовність нуклеотидів РНК. Зростання ланцюга РНК відбувається в напрямку від 5'- до 3'-кінця. Субстратами реакції є 3'-кінцева ОН-група рибози зростаючого *транскрипту* (ланцюга РНК, що синтезується) і рибонуклеозидтрифосфати (rNTP). Ферментом, що каталізує цю реакцію, є *ДНК-залежна РНК-полімераза* (DNA dependent RNA Polymerase, RNAP). Молекулярні механізми синтезу РНК, за основними рисами спільні для всіх живих організмів, на початку цього розділу розглянуто на прикладі бактеріальної системи транскрипції. Особливостям транскрипції в еукаріотів присвячено другу частину розділу, а в наступному розділі описуватимуться принципи її регуляції.

### Загальний сценарій процесу транскрипції

РНК-полімераза має порівняно невисоку спорідненість до ДНК, що дозволяє їй зв'язуватись на деякий короткий час, дисоціювати і зв'язуватись знову. У такий спосіб здійснюється пошук особливої ділянки, з якої транскрипція має розпочинатись – *промотору*. Для ефективного впізнання промотору (підвищення спорідненості саме до нього), а також забезпечення початкових стадій синтезу РНК-полімераза потребує деякого загального фактора (факторів) ініціації транскрипції.

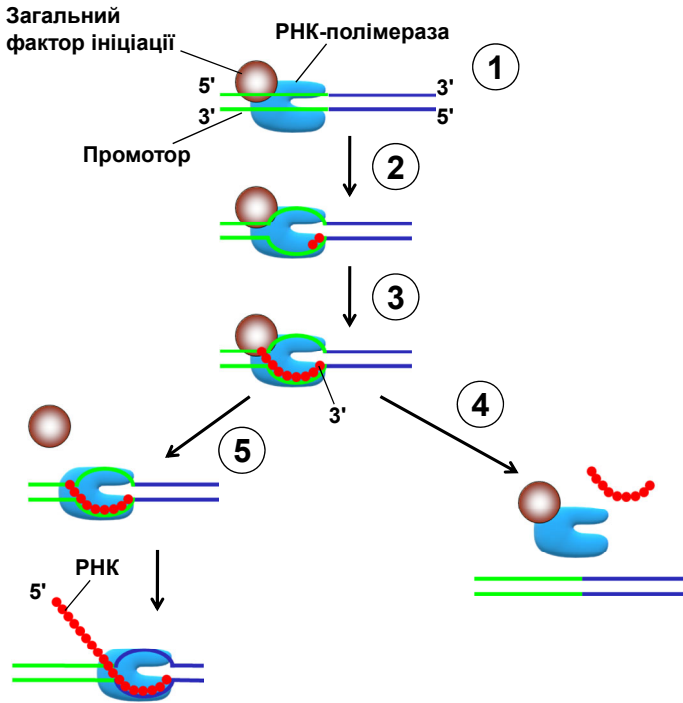
Розглянемо стадії робочого циклу РНК-полімерази.

**1. Ініціація** транскрипції, яка, своєю чергою, є достатньо складним багатоетапним процесом (рис. 5.1), включає такі кроки:

- зв'язування РНК-полімерази і загального фактора ініціації транскрипції з промотором, у результаті чого формується *закритий комплекс*, у складі якого ДНК зберігає форму подвійної спіралі;
- локальне плавлення подвійної спіралі з утворенням *відкритого комплексу* – розходження ланцюгів ДНК, яке дозволяє використовувати один із них як матрицю;
- включення перших двох нуклеотидів до молекули РНК (синтезу першого фосфодієфірного зв'язку в активному центрі полімерази); це найповільніша стадія процесу;
- зростання первинного короткого транскрипту – приєднання 8–9 нуклеотидів, після чого (або в процесі) найчастіше відбувається *абортивна ініціація* (визволення нефункціонального короткого транскрипту й руйнування усього комплексу) – невдала спроба ініціації;
- альтернативою є *очищення промотору* – дисоціація від РНК-полімерази загального фактора ініціації, що маркує перехід до елонгації транскрипції.

**2. Елонгація** транскрипції, у кожному елементарному акті якої (елонгаційному циклі) відбувається приєднання чергового нуклеотиду до 3'-кінця РНК з одночасним відщепленням пірофосфату (енергетичне джерело реакції) і пересування РНК-полімерази на один нуклеотид уздовж матриці (*транслокація*). У процесі елонгації полімераза особливо міцно зв'язана з ДНК – настільки, що потрібні спеціальні зусилля для зупинки процесу.

**3. Термінація** транскрипції при впізнанні полімеразою спеціального сигналу (особливого елемента послідовності), коли відбувається визволення транскрипту й дисоціація РНК-полімерази. Далі фермент може вступати у новий раунд транскрипції на тому чи іншому промоторі.



**Рис. 5.1.** Етапи ініціації транскрипції:

- 1 – зв'язування РНК-полімерази у комплексі із загальним фактором ініціації з промотором (формування закритого комплексу);
- 2 – локальне плавлення ДНК (утворення відкритого комплексу), синтез першого фосфодієфірного зв'язку;
- 3 – зростання первинного транскрипту; 4 – абортівна ініціація;
- 5 – альтернативне очищення промотору і початок елонгації транскрипції

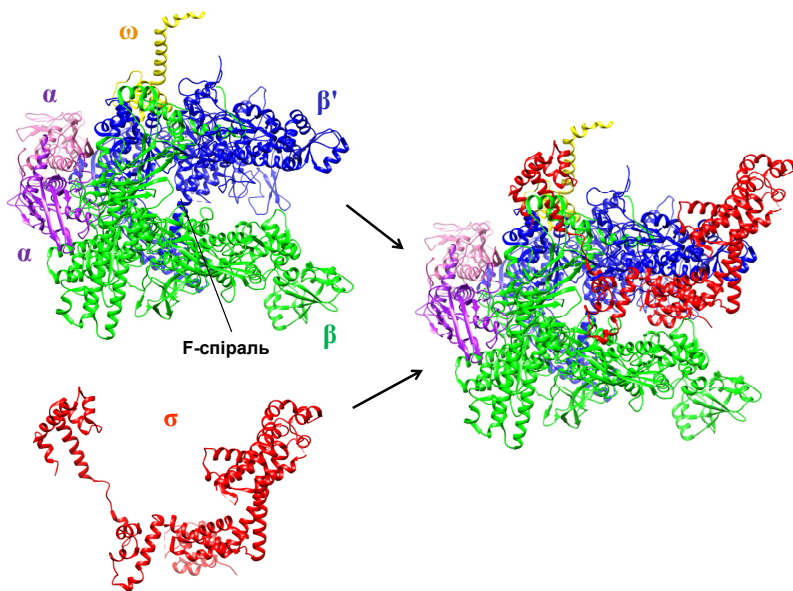
Розглядаючи цей загальний сценарій, потрібно відповісти на кілька запитань: чому процес ініціації транскрипції не є ефективним, а його частим результатом є відсутність результату – абортівна ініціація; навіщо потрібен загальний фактор ініціації; чому, якщо спроба ініціації виявилась вдалою і РНК-полімераза здійснює елонгацію, її важко зупинити? Для відповідей на ці запитання необхідно зрозуміти структурну організацію РНК-полімерази й механізми її взаємодії з нуклеїновими кислотами.

## Бактеріальна РНК-полімераза

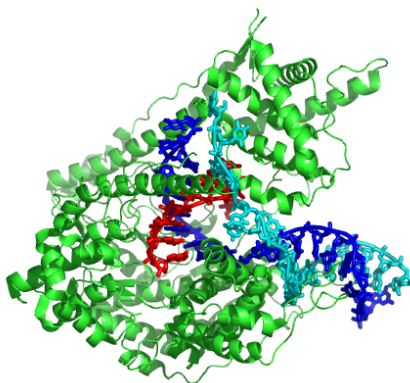
Бактеріальна РНК-полімераза складається з кількох білкових субодиниць та існує, залежно від стадії транскрипції, у двох формах: 1) *кор-фермент* (core enzyme) у складі субодиниць, що позначаються як  $\alpha$  (дві копії),  $\beta$ ,  $\beta'$  та  $\omega$ ; 2) *голофермент* (holoenzyme) – комплекс кор-ферменту із субодиницею  $\sigma$ . Кор-фермент працює під час елонгації транскрипції, голофермент здійснює ініціацію. Отже, у випадку бактеріальної системи транскрипції РНК-полімераза на рис. 5.1 означає кор-фермент, а субодиниця  $\sigma$  виконує функцію загального фактора ініціації.

Дві субодиниці  $\alpha$  взаємодіють між собою у складі кор-ферменту бактеріальної РНК-полімерази, виконуючи структурну функцію: вони сприяють утриманню разом інших частин ферменту (рис. 5.2). Стабілізує структуру мультибілкового комплексу й найменша із субодиниць –  $\omega$ , яка сприяє також зв'язуванню  $\sigma$ -фактора з кор-ферментом. Дві великі субодиниці –  $\beta$  і  $\beta'$  – формують характерні *щелепи* (форма нагадує також клішню краба), у щілині між якими з ферментом взаємодіє ДНК, розташована нижче (downstream) у напрямку руху полімерази.

Глибоко між щелепами, ліворуч від F-спіралі (F bridge helix, див. рис. 5.2), яка належить до найбільшої субодиниці  $\beta'$ , міститься каталітичний активний центр – 3 залишки Asp (також належать до  $\beta'$ ), що утримують ключовий для каталізу іон  $Mg^{2+}$  (див. нижче). Структурна організація бактеріальної РНК-полімерази (рухливі щелепи, активний центр з іоном  $Mg^{2+}$  та інші описані нижче особливості) є загальною для всіх РНК-полімераз – як для гомологічних еукаріотичних полімераз (див. нижче), так і для багатьох інших. Зокрема, деякі бактеріофаги (зазвичай фаги використовують полімеразу клітини-хазяїна) мають власні РНК-полімерази, що є мономерними білками без гомології в амінокислотній послідовності до субодиниць бактеріальної полімерази (рис. 5.3). Аналогічна мономерна РНК-полімераза працює і в мітохондріях. Простота мономерних полімераз (яка робить їх зручним об'єктом досліджень) зумовлена тим, що вони мають упізнавати лише кілька промоторів фагових генів, однак указані вище загальні риси структурної організації РНК-полімераз і базові принципи функціонування зберігаються і для них.

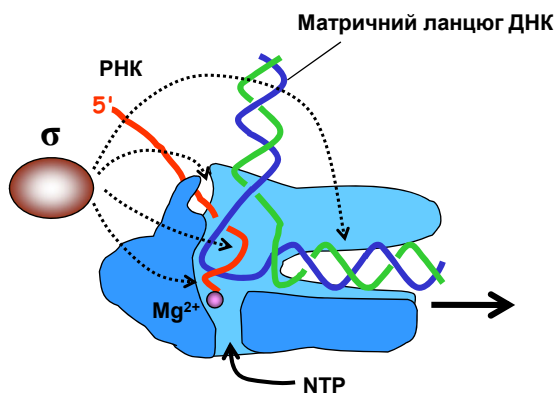


**Рис. 5.2.** Кор-фермент,  $\sigma$ -фактор і їхнє об'єднання у складі голоферменту РНК-полімерази *E. coli* (4LK1)



**Рис. 5.3.** Структура елонгаційного комплексу мономерної РНК-полімерази бактеріофага T7 (1MSW). Матричний ланцюг ДНК позначено темно-синім, РНК – червоним

Схему будови бактеріальної РНК-полімерази в комплексі з ДНК і РНК під час елонгації транскрипції наведено на рис. 5.4 (за основними рисами ця схема зберігається і для еукаріотичних РНК-полімераз). Структурні елементи, які оточують активний центр, під час елонгації транскрипції забезпечують взаємодію з нематричним ланцюгом ДНК (підтримуючи його розплавлений стан), дволанцюговим гібридом ДНК-РНК, сприяють руйнуванню подвійної спіралі ДНК нижче активного центра, руйнуванню подвійної спіралі гібрида ДНК-РНК і відновленню подвійної спіралі ДНК вище (upstream) активного центра. Від активного центра, де міститься 3'-кінець РНК, що зростає, РНК виходить за межі полімеразного комплексу через спеціальний канал. Інший, *вторинний канал*, дозволяє нуклеозидтрифосфатам потрапляти до точки зростання РНК.



**Рис. 5.4.** Схема організації полімеразного комплексу під час елонгації транскрипції і вплив  $\sigma$ -фактора під час ініціації.  $Mg^{2+}$  – активний центр, стрілками показано напрямок руху полімерази і шлях входу нуклеозидтрифосфатів (NTP), пунктирними стрілками – основні точки взаємодії  $\sigma$ -фактора із кор-ферментом під час ініціації транскрипції

Взаємодія структурних елементів полімерази з нуклеїновими кислотами забезпечує надзвичайно високу спорідненість до гібридної подвійної спіралі ДНК-РНК. Саме це спричинює високу *процесивність* РНК-полімераз – здатність працювати на матриці

без дисоціації. Така зумовлена високою спорідненістю до гібрида процесивність визначає необхідність особливих механізмів термінації транскрипції. Однак гібрид ДНК-РНК потрібно спочатку створити – він існує під час елонгації і його ще немає (або він дуже короткий) на початкових стадіях ініціації. Отже, на початкових стадіях виникає необхідність розв'язання кількох завдань, яке не може забезпечити кор-фермент сам по собі: додатково стабілізувати полімераза на ДНК, жорстко зафіксувати субстрати (перші два нуклеотиди) в активному центрі, стабілізувати короткий гібрид ДНК-РНК на початкових стадіях синтезу. Усе це бере на себе фактор ініціації  $\sigma$ , який є абсолютно необхідним для початку синтезу РНК.

Цей білок із мультидоменною структурою здійснює численні взаємодії з кор-ферментом (рис. 5.2, 5.4): фіксує певне положення щелеп; взаємодіє з активним центром і його оточенням, сприяючи жорсткій фіксації перших двох нуклеотидів і стабілізації короткого гібрида ДНК-РНК; взаємодіє з ДНК у ділянці downstream, сприяючи утворенню відкритого комплексу при ініціації транскрипції. Крім того,  $\sigma$ -фактор блокує канал виходу 5'-кінця РНК, що іммобілізує РНК-полімераза, не дозволяючи їй рухатись (під час ініціації невелика частина ДНК з ділянки downstream втягується всередину РНК-полімерази, що спричиняє зростання розміру розплавленої зони). Відповідно по завершенню процесу ініціації дисоціація  $\sigma$  (очищення промотору) є необхідною для виходу синтезованої РНК, тобто для вільного руху полімерази вздовж матриці під час елонгації (при цьому відновлюється вихідний розмір розплавленої зони). Отже, під час ініціації між 5'-кінцем транскрипту і  $\sigma$ -фактором виникає конкуренція за канал виходу. Доволі часто "перемагає"  $\sigma$ -фактор, що є однією із причин абортивної ініціації. Другою причиною є короткий (а отже, нестабільний) гібрид ДНК-РНК на початкових стадіях синтезу первинного транскрипту – зусилля  $\sigma$ -фактора з його стабілізації іноді виявляються недостатніми і короткий транскрипт дисоціює. Загалом абортивну ініціацію можна сприймати як своєрідну плату за можливість розпочати синтез РНК, яку надає  $\sigma$ -фактор.

## Бактеріальний промотор та ініціація транскрипції

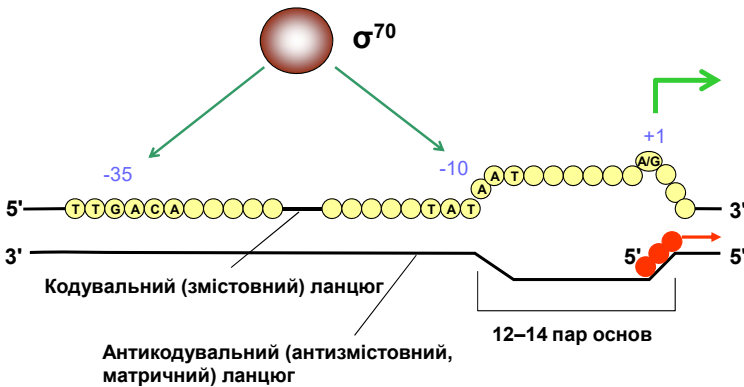
Крім зазначених вище функціональних ролей фактора  $\sigma$ , у нього є ще одна – упізнання промотору. Бактеріальні промотори розрізняються за спорідненістю до голоферменту (*силою промотору*). Упізнання слабких промоторів залежить від додаткових, специфічних для гена чи групи генів, факторів ініціації (див. розділ 6).

Зрозуміло, що РНК-полімераза може бути орієнтована відносно ДНК у два способи й кожній із цих орієнтацій відповідатимуть два можливі напрямки руху ферменту при транскрипції. Відповідно один із двох ланцюгів ДНК буде обиратися як *матричний* – той, з якого зчитується інформація в напрямку 3'–5' і на якому синтезується комплементарний ланцюг РНК у протилежному напрямку 5'–3' (рис. 5.5).

Послідовність ланцюга ДНК, який є комплементарним матричному, збігається з послідовністю РНК, що синтезується. Тому саме цей нематричний ланцюг називають *кодуювальним*, або *змістовним* (coding, sense), і саме його послідовність прийнято наводити як послідовність гена (матричний ланцюг називають, відповідно, *антикодуювальним*, або *антизмістовним*). Нуклеотид кодуювального ланцюга, який відповідає стартовому нуклеотиду РНК (частіше пуриновий), позначають +1 (рідше 0), наступні нуклеотиди (чи пари основ) у напрямку руху полімерази нумерують із позначкою "+": +2, +3 тощо. У зворотному напрямку (ліворуч на рис. 5.5) нуклеотиди нумерують із позначкою "-" (починаючи з -1).

Типовий бактеріальний промотор (рис. 5.5) складається із двох елементів послідовності, які знаходяться на відстані приблизно -10 і -35 від стартової точки. Консенсусна (узагальнена для різних промоторів) послідовність елемента -35 має вигляд  $T_{82}T_{84}G_{78}A_{65}C_{54}A_{45}$  (числами позначено частоту зазначеного нуклеотиду у відсотках). Послідовність елемента -10 (його називають також *боксом Прибноу* (David Pribnow) – Pribnow box) така:  $T_{80}A_{95}T_{45}A_{60}A_{50}T_{96}$ . Варіабельність послідовностей визначає різну ефективність ініціації – силу промоторів. Саме еле-

менти  $-35$  і  $-10$  безпосередньо впізнає  $\sigma$ -фактор при ініціації транскрипції. Наведені послідовності упізнаються варіантом  $\sigma$ -фактора з молекулярною вагою  $70$  кДа ( $\sigma^{70}$ ), який є доволі широко вживаним, але не єдиним. Кільком іншим варіантам  $\sigma$ , під контролем яких перебувають певні групи промоторів, відповідають інші консенсусні послідовності  $-35$  та  $-10$ . Взаємодія  $\sigma$ -фактора із промотором чітко орієнтує РНК-полімеразу відносно ДНК, тобто визначає і стартову точку транскрипції, і її напрямок (і, відповідно, вибір одного із двох ланцюгів як матричного).

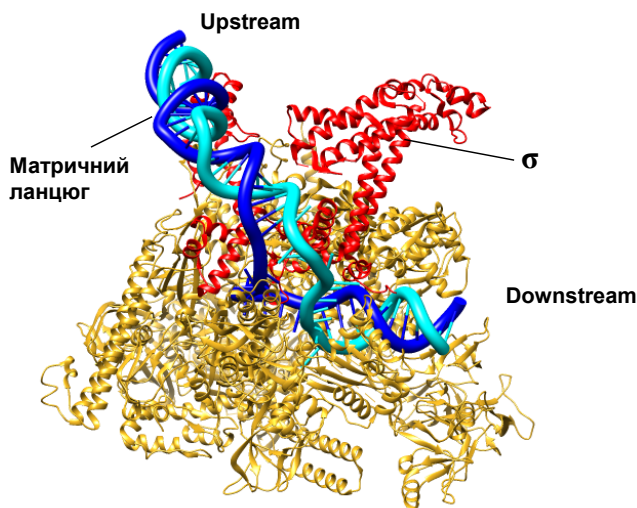


**Рис. 5.5.** Типовий промотор *E. coli*. Зелена стрілка позначає старт транскрипції всередині розплавленої зони

У складі голоферменту РНК-полімерази субодинаця  $\sigma$  утримує щелепи у відкритому положенні (рис. 5.2). Це сприяє підвищенню швидкості дисоціації білка від непромоторної ДНК і гарантує низьку ефективність транскрипції з випадкових сайтів на ДНК. За умови впізнання  $\sigma$ -фактором двох елементів промотору відбувається структурна перебудова полімеразного комплексу із замиканням щелеп. У результаті ДНК у зоні стартової точки потрапляє всередину щілини між щелепами – утворюється закритий комплекс.

Дуже швидко закритий комплекс перетворюється на відкритий: замикання ДНК у щілині зумовлює те, що подвійна спіраль не може розміститися там через брак простору. У результаті від-

бувається локальне плавлення 12–14 пар основ у зоні від  $-7$  до  $+5$  відносно стартової точки (рис. 5.5, 5.6). Це плавлення, на відміну від еукаріотичної системи транскрипції (див. нижче), не потребує енергетичних витрат у вигляді гідролізу АТР – воно відбувається за рахунок взаємодії двох ланцюгів із внутрішньою поверхнею щілини: нематричний ланцюг захоплюється певними структурними елементами полімерази, матричний – опиняється в активному центрі. Вагомим фактором, що сприяє локальному плавленню подвійної спіралі у складі відкритого комплексу, є наявність у циркулярній бактеріальній ДНК певного рівня негативної надспіралізації – напружень, що полегшують розкручування дуплекса (див. розділ 3). Важливою властивістю, що додатково (крім нуклеотидної послідовності стандартних елементів промотору) визначає силу бактеріальних промоторів, є стабільність подвійної спіралі: сильні промотори відрізняються зниженою стабільністю, тобто підвищеною здатністю до плавлення у відповідь на негативну надспіралізацію.



**Рис. 5.6.** Відкритий комплекс РНК-полімерази *E. coli* (6XL5)

Наступна, найповільніша стадія ініціації – синтез першого фосфодієфірного зв'язку, тобто ініціація власне синтезу РНК.

Здатність ініціювати синтез нуклеїнової кислоти – унікальна властивість РНК-полімерази (ДНК-полімераза здатна лише продовжувати синтез уже існуючого ланцюга, див. розділ 9). Ініціація синтезу дуже важка, оскільки при цьому дві порівняно невеликі молекули – перші два нуклеотиди – мають бути жорстко зафіксовані в активному центрі в певній взаємній орієнтації. Як обговорювалось вище, така фіксація залежить від  $\sigma$ -фактора.

Далі процес нарощування транскрипту продовжується ще до 8–9 нуклеотидів, які можуть розміститися в межах РНК-полімеразного комплексу до виходу в спеціальний канал. Після зростання транскрипту до 9–10 нуклеотидів  $\sigma$ -фактор більше не потрібен. У випадку, коли конкуренцію за канал "виграє" РНК і не відбулось руйнування комплексу при абортивній ініціації (див. вище),  $\sigma$ -фактор після синтезу  $\sim 12$  нуклеотидів витісняється з комплексу і розпочинається елонгація транскрипції.

## Елонгація транскрипції

При елонгації транскрипції РНК-полімераза рухається вздовж матриці разом із транскрипційним міхуром – розплавленими 12–14 парами основ (рис. 5.7). При цьому на кожному кроці полімерази одна пара основ руйнується попереду розплавленої зони й одна відновлюється позаду. Розділення подвійної спіралі ДНК перед міхуром полегшує G-петля РНК-полімерази (G-trigger loop, рис. 5.7). У зоні міхура з полімеразою завжди зв'язаний гібрид ДНК-РНК завдовжки вісім-дев'ять пар основ: вісім-дев'ять нуклеотидів на 3'-кінці РНК залучені до утворення подвійної спіралі з матричним ланцюгом, наступні п'ять-шість – локалізовані в каналі виходу РНК, решта нуклеотидів на 5'-кінці виходять за межі полімеразного комплексу. Руйнуванню гібрида, а отже, і відновленню ДНК-дуплекса сприяє L-петля (lid loop, рис. 5.7). Як обговорювалось вище, висока спорідненість певних структурних елементів полімерази до гібрида ДНК-РНК, а також взаємодія РНК з каналом виходу, зумовлюють високу процесивність: дисоціація полімерази, яка була можливою під

час синтезу первинного транскрипту, під час елонгації стає дуже малоймовірною подією.

На рис. 5.7 схематично зображено вихідну конфігурацію системи на початку кожного елонгаційного циклу перед зв'язуванням чергового NTP, що потрапляє до свого сайту зв'язування через вторинний канал. Сайт зв'язування формується 3'-кінцевим нуклеотидом РНК, неспареним нуклеотидом матричного ланцюга ДНК (позначеним як  $i + 1$ ), активним центром полімерази та його оточенням, зокрема F-спіраллю.

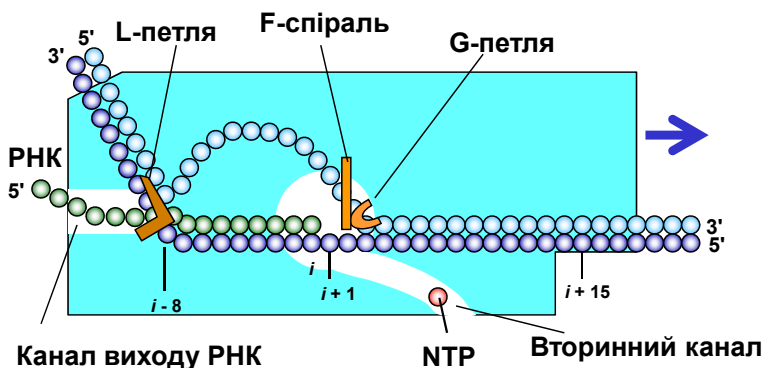


Рис. 5.7. Вихідна конфігурація системи транскрипції на початку елонгаційного циклу

**Перший етап** елонгаційного циклу – *відбір і зв'язування* нуклеозидтрифосфату, яке контролюється матрицею (нуклеотидом  $i + 1$ ). Спорідненість усіх чотирьох типів NTP до РНК-полімерази невисока, що дозволяє їм здійснювати швидке зв'язування / дисоціацію. Реалізація комплементарних взаємодій NTP із нуклеотидом матриці (поява нового ліганду в сайті зв'язування – комплементарної нуклеотидної пари) викликає конформаційну зміну із замиканням NTP і жорсткою фіксацією субстратів у активному центрі.

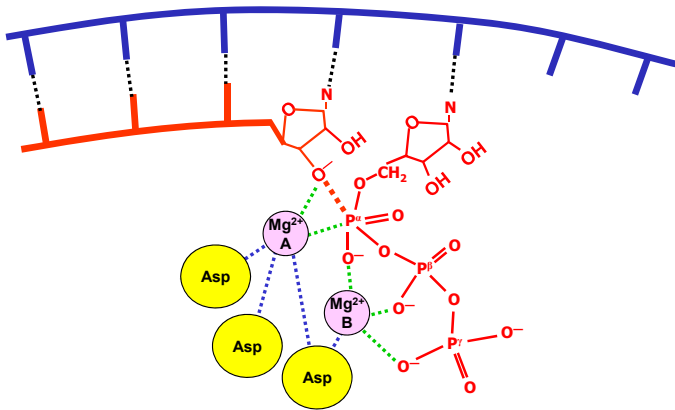
**Другий етап** – хімічна реакція *приседнання чергового нуклеотиду* до 3'-кінця транскрипту (рис. 5.8). За умови реалізації комплементарних взаємодій між NTP і нуклеотидом у складі матриці NTP жорстко фіксується в активному центрі в певній

напруженій реакційноздатній конформації, яка може розглядатися як інтермедіат реакції. Активний центр формується трьома залишками Asp, які належать субодиноці  $\beta'$ . Найважливішу роль у фіксації інтермедіату відіграють два іони  $Mg^{2+}$ , що утримуються негативно зарядженими залишками Asp: один (іон А) наявний у активному центрі завжди, інший (іон В), який потрапляє туди разом із NTP, – тимчасово. Перший іон взаємодіє з  $\alpha$ -фосфатом NTP і 3'-ОН-групою, забезпечуючи, зокрема, її іонізацію (рис. 5.8). Другий іон стабілізує певну напружену конформацію всіх трьох фосфатних залишків. Спорідненість активного центра до інтермедіату й жорстка фіксація субстратів забезпечує зниження активаційного бар'єра реакції, тобто її каталіз (див. розділ 2). Такий механізм каталізу за участю іонів  $Mg^{2+}$ , що утримуються трьома залишками Asp, є спільним для всіх полімераз (РНК- і ДНК-полімераз), що здійснюють синтез нуклеїнових кислот. Результатом реакції є подовжений на один нуклеотид 3'-кінець РНК і відщеплений пірофосфат. Крім того, унаслідок реакції змінюється розташування ліганду (3'-кінця) відносно активного центра (рис. 5.7, 5.8).

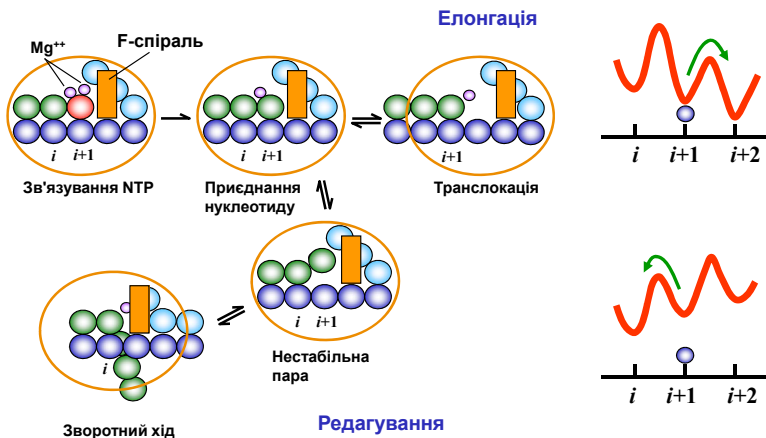
Зміна ліганду й дисоціація пірофосфату спричиняє "розмикання" структури в зоні активного центра, що дозволяє рухатися полімеразі. Відбувається *третій етап – транслокація* – у напрямку оптимального розташування подовженого 3'-кінця відносно активного центра (рис. 5.9).

Рух полімерази при транслокації відбувається як одновимір-на дифузія вздовж матриці за рахунок енергії теплових флуктуацій, тобто можливим є рух як уперед, так і назад. Пересування полімерази вперед супроводжується руйнуванням однієї нуклеотидної пари у складі гібрида РНК-ДНК і відновленням однієї пари ДНК позаду від розплавленої зони; одночасно руйнується одна пара ДНК попереду. Якщо взяти до уваги, що на попередньому етапі внаслідок хімічної реакції відбулося утворення пари в гібриді, то маємо практично ізоенергетичний процес – дві нуклеотидні пари утворюються замість двох зруйнованих. Аналогічно пересування полімерази назад буде супроводжуватись руйнуванням пари в гібриді на 3'-кінці РНК (3'-кінцевий нуклеотид при цьому опиняється у вторинному

каналі) і пари ДНК позаду від міхура з одночасним утворенням пари ДНК попереду й гібрида позаду.



**Рис. 5.8.** Схема інтермедіатного стану в активному центрі РНК-полімерази у процесі хімічної реакції приєднання нуклеотиду до 3'-кінця транскрипту (червоний)



**Рис. 5.9.** Послідовність подій в елонгаційному циклі (збільшений фрагмент схеми на рис. 5.7). Показано також альтернативний шлях при редагуванні помилок. Праворуч: енергетичні профілі, що відповідають руху полімерази вперед відносно матриці при елонгації (угорі) і назад при редагуванні (унизу)

Різниця між кінцевими результатами двох рухів – в енергії взаємодій РНК-полімерази з нуклеїновими кислотами. У вихідному претранслокаційному стані (положення  $i + 1$ , рис. 5.9) реалізуються певні нуклеопротейнові взаємодії, що зумовлює перебування системи в локальному мінімумі вільної енергії. Із двох інших локальних мінімумів, що відповідають положенням  $i$  та  $i + 2$ , другий є більш глибоким, оскільки для нього реалізуються найсприятливіші взаємодії між активним центром і 3'-кінцем РНК. Максимуми енергії (бар'єри) між мінімумами на рис. 5.9 відповідають проміжним станам при пересуванні полімерази (часткове порушення взаємодій, конформаційні зміни полімерази). Ці максимуми невисокі, тобто легко долаються за рахунок теплової енергії. Отже, полімераза з більш високою імовірністю спонтанно переміщується вперед, система опиняється в більш глибокому мінімумі вільної енергії (рис. 5.9), звідки починається наступний елонгаційний цикл.

Якщо *напрямок* переміщення полімерази визначається положенням мінімумів на шкалі вільної енергії, то *швидкість* переміщення залежить від висоти максимумів вільної енергії: чим нижчим є максимум, тим легше він долається. Висота максимумів залежить від послідовності пар основ у складі гібрида, а також може змінюватись під дією додаткових факторів (білкові фактори елонгації, шпильки в мРНК тощо).

Проте у випадку приєднання помилкового нуклеотиду профіль вільної енергії змінюється, сприяючи зворотному руху полімерази, що дає можливість *редагування помилок*. Забезпечити стовідсоткову точність операцій з такими невеликими молекулами, як нуклеотиди, практично неможливо. Зокрема, джерелом (одним з основних) помилкового приєднання нуклеотидів під час транскрипції (так само як і під час синтезу ДНК, розділ 9) є *таутомерія* азотистих основ. Спонтанні переміщення протона з однієї хімічної групи на іншу й перебудови електронних систем гетероциклів зумовлюють те, що кожна основа існує у вигляді двох основних таутомерних форм: аміно- чи іміноформи для А, С, енольної чи кетоформи для G, U, T (рис. 5.10). Рівновага зсунута в бік аміно- й кетоформ, які наявні у складі подвійних спіралей (див. рис. 3.1, 3.7) і для яких реалізуються правила ком-

плементарності А-Т, G-C. Однак спарювання основ підпорядковується іншим правилам для мінорних таутомерних форм: наприклад, іміноформа А й аміноформа С утворюють між собою два водневі зв'язки, що може відбутися під час упізнання матриці нуклеозидтрифосфатом. Тоді після хімічної реакції, у результаті швидкого повернення до мажорної таутомерної форми, у складі гібрида ДНК-РНК на 3'-кінці транскрипту залишиться некомплементарна пара нуклеотидів.

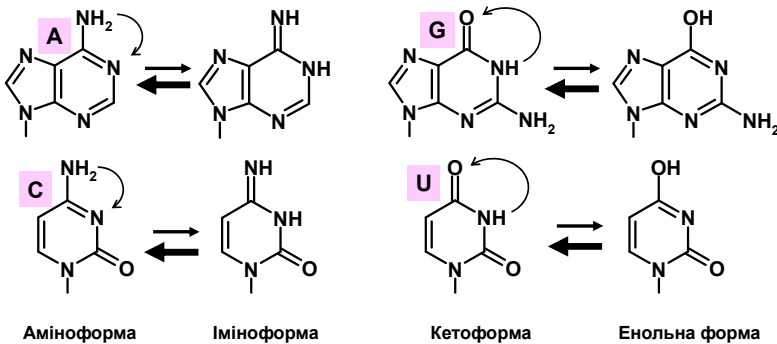


Рис. 5.10. Таутомерні форми азотистих основ

Нестабільність гібрида на 3'-кінці у випадку, коли щойно приєднаний нуклеотид є помилковим, зумовлює зміни енергетичного балансу між різними позиціями РНК-полімерази відносно матриці (рис. 5.9): тепер рух уперед буде означати збереження нестабільної (енергетично невигідної) пари, а рух назад відновить стабільність гібрида по всій його довжині. Відповідно найбільш імовірною (такою, що відповідає найнижчому мінімуму енергії) тепер буде позиція *i*. Отже, відбувається зворотний рух полімерази з виходом 3'-кінця транскрипту у вторинний канал. При цьому ніщо більше не фіксує 3'-кінець транскрипту, і всі позиції позаду від *i* (*i*, *i* - 1, *i* - 2, ...) є майже ізоенергетичними (варіації залежать від стабільності гібрида, тобто послідовності нуклеотидів) – зворотний рух відбувається на невизначену випадкову відстань.

Коли 3'-кінцевий фрагмент транскрипту опиняється у вторинному каналі, з полімеразою взаємодіє особливий фактор елонгації транскрипції, який можна назвати *фактором редагування* і який існує у двох варіантах: GreA та GreB (еукаріотичним аналогом цих факторів є фактор TFIIIS, див. нижче). Gre-фактори, через конформаційні зміни РНК-полімерази, індують нуклеазну активність її активного центра. У результаті відбувається відщеплення 3'-кінцевого фрагмента транскрипту і в активному центрі залишається новий 3'-кінець (зі стабільною нуклеотидною парою гібрида), який використовується для відновлення елонгаційного процесу. Загальна точність процесу транскрипції дорівнює приблизно  $10^{-6}$  (частота помилково включених нуклеотидів, що залишаються у складі РНК).

З викладеного видно, що РНК-полімераза є одним із прикладів молекулярної машини (див. розділ 2). *Конформаційна рухливість* полімерази забезпечує їй можливість існувати в кількох структурних станах. Ці стани мають різну спорідненість до певних лігандів (NTP, пари основ у складі гібрида РНК-ДНК, фактори елонгації). *Взаємодії з лігандами* фіксують певні стани. *Хімічні реакції*, які каталізуються машиною, спричиняють заміну лігандів, а отже, і перехід в інший структурний стан. Рушійною силою для переміщення полімерази є *тепловий рух*: структурні блоки полімеразного комплексу рухаються відповідно до конструкції машини; зв'язування лігандів і їхня заміна внаслідок реакцій каналізують ці рухи в певних напрямках. Результатом структурних перебудов є переміщення полімерази вздовж матриці з одночасним синтезом транскрипту зі швидкістю не менше 40 нуклеотидів/с.

## Термінація транскрипції у бактерій

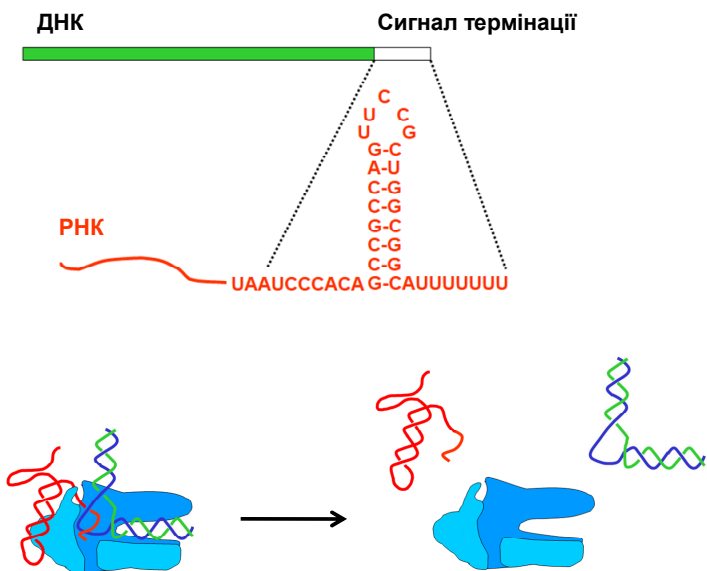
Як було зазначено вище, швидкість пересування РНК-полімерази під час елонгації визначається висотою максимумів вільної енергії (рис. 5.9). Якщо ці максимуми (по обидва боки від мінімуму, що відповідає позиції  $i + 1$  на рис. 5.9) за певних при-

чин виявляться високими, то полімераза опиниться в енергетичній пастці й зупиниться на тривалий час.

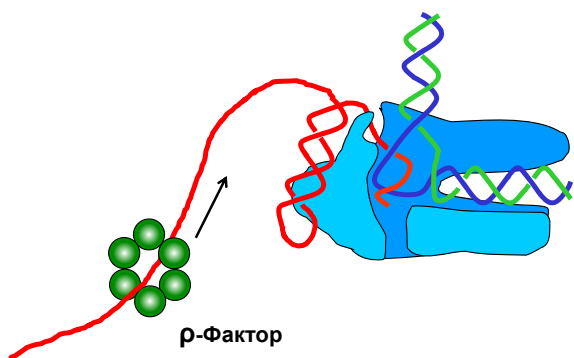
Саме це й відбувається під час термінації транскрипції на певних елементах послідовності. Бактеріальний *сигнал термінації* – це інвертований повтор (паліндром – послідовність ДНК, що читається однаково в обох ланцюгах у напрямках 5'–3'), безпосередньо фланкований послідовністю oligoT (близько 8 нуклеотидів). Відповідно у складі транскрипту утворюється дволанцюгова шпилька, фланкована послідовністю oligoU (рис. 5.11). Гібрид ДНК-РНК за шпилькою складається з менш стабільних А-У пар.

Загальний сценарій термінації може бути описаний таким чином. Порівняно нестабільний гібрид ускладнює елонгацію транскрипції, а шпилька – зворотний рух полімерази. Більше того, під час свого формування шпилька здійснює механічний тиск на полімеразу, сприяючи витягуванню oligoU-кінця із гібридного дуплекса. Відбувається руйнування гібрида, відновлення подвійної спіралі ДНК і визволення транскрипту (рис. 5.11).

У кількох бактеріальних оперонах термінація залежить від фактора  $\rho$  (рис. 5.12). Цей фактор (моногексамерний білок) зв'язується з РНК-транскриптом у певних С-збагачених ділянках. Після цього починається АТР-залежне пересування  $\rho$ -фактора вздовж транскрипту в напрямку 5'–3' зі швидкістю, яка є нижчою за швидкість руху полімерази. Сигнал термінації в цьому випадку нічим не відрізняється від описаних вище термінаторів і зумовлює зупинку полімерази. У результаті  $\rho$ -фактор доганяє полімеразу і, продовжуючи рух уздовж РНК, руйнує гібридну подвійну спіраль (працює як *геліказа* (helicase) – так називаються ферменти, що АТР-залежно руйнують подвійні спіралі нуклеїнових кислот). Результатом є відновлення подвійної спіралі ДНК і визволення транскрипту.



**Рис. 5.11.** Типовий бактеріальний сигнал термінації і витіснення РНК-полімерази із комплексу з гібридом РНК-ДНК під дією дволанцюгової шпильки у складі транскрипту



**Рис. 5.12.** ρ-Залежна термінація

Суттєвою особливістю прокариотичної системи транскрипції білкових генів є те, що молекула мРНК зв'язується з рибосомами безпосередньо під час транскрипції – транскрипція мРНК і білковий синтез (розділ 8) є єдиним процесом. Власне, за відсутності рибосом мРНК буде швидко піддаватись нуклеазній деградації. Поєднання транскрипції і трансляції зумовлює їхній взаємний вплив і використовується бактеріальною клітиною для регуляції експресії генів (розділ 6).

Зрозуміло, що в еукаріотів транскрипція і трансляція розділені у просторі й часі. Розглянуті вище базові механізми транскрипції зберігаються для еукаріотів, хоча, звісно, еукаріотична транскрипційна система має певні особливості.

## Еукаріотичні РНК-полімерази

В еукаріотичних клітинах функціонують РНК-полімерази трьох типів:

- *РНК-полімераза I* працює на кластерах генів рибосомної РНК (розділ 4) і здійснює синтез рРНК 18S, 28S та 5,8S;
- *РНК-полімераза II* транскрибує білкові гени (синтезує всі мРНК), а також гени багатьох РНК, що не транслюються;
- *РНК-полімераза III* здійснює синтез тРНК, рибосомної РНК 5S і кількох інших низькомолекулярних РНК.

У рослин наявні також додаткові РНК-полімерази IV і V – похідні полімерази II, залучені до синтезу певних РНК, що не транслюються.

Кожна із трьох стандартних еукаріотичних полімераз містить чотири гомологічні корові субодиниці, які є водночас гомологами субодиниць  $\alpha$ ,  $\beta$  і  $\beta'$  прокариотичної полімерази (див. вище). Крім того, до складу полімераз входять п'ять спільних для всіх трьох ферментів субодиниць, а також певний набір специфічних субодиниць (кількістю 5, 3 і 7 для РНК-полімераз I, II і III, відповідно). Загальна архітектура еукаріотичних полімераз (наявність щелеп, між якими зв'язується ДНК, каналу виходу РНК, вторинного каналу для входу нуклеозидтрифосфатів) дуже схо-

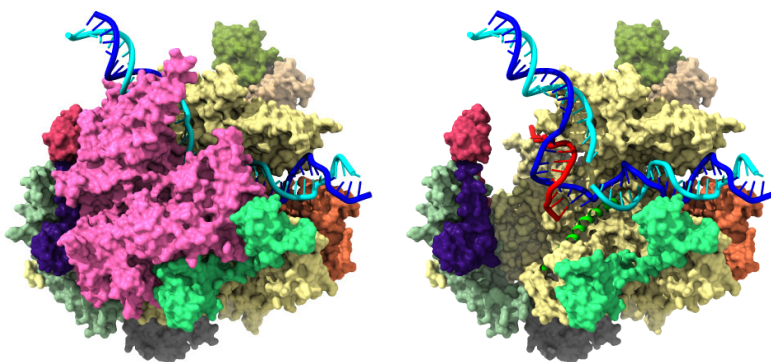
жа на таку прокаріотичної полімерази. Між про- та еукаріотичними полімеразами спостерігається також висока подібність внутрішньої поверхні щілини щелеп і активного центра; спільними є й основні механізми роботи полімераз, розглянуті вище у цьому розділі. Проте відсутня гомологія зовні: додаткові до чотирьох корових субодиниці еукаріотичних полімераз створюють специфічну поверхню для взаємодії з елементами еукаріотичної системи транскрипції. Із переліку функціональних ролей трьох полімераз зрозуміло, що полімерази I і III є високоспеціалізованими, а полімераза II виконує основну роботу із синтезу РНК в еукаріотичній клітині.

## РНК-полімераза II

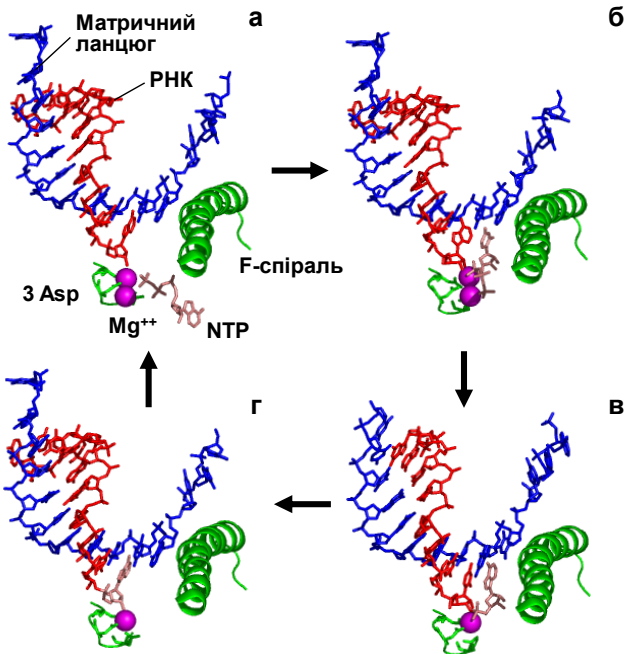
РНК-полімераза II складається із 12 субодиниць (рис. 5.13). Найбільша субодиниця (гомолог субодиниці  $\beta'$  прокаріотичної полімерази) формує, зокрема, активний центр, структура й найближче оточення якого практично ідентичні таким бактеріальній полімерази. Відповідно ідентичними є і механізми елонгації синтезу РНК. На рис. 5.14 показано структуру оточення активного центра на різних стадіях елонгаційного циклу (порівн. з рис. 5.7, 5.8): вхід NTP через вторинний канал; зв'язування NTP із матрицею і його фіксація в активному центрі завдяки двом іонам  $Mg^{2+}$ , які утримуються трьома залишками Asp; каталіз реакції приєднання нуклеотиду до 3'-кінця транскрипту; транслокація полімерази на один нуклеотид уздовж матриці, унаслідок чого система здатна вступити в новий цикл елонгації.

Під час елонгації, так само як і для бактеріальній полімерази, за умови нестабільності гібрида РНК-ДНК на 3'-кінці відбувається зворотний рух полімерази з виходом 3'-кінця транскрипту у вторинний канал. У цьому випадку взаємодія з фактором транскрипції **TFIIIS** (Transcription Factor S of RNA-polymerase II, аналог бактеріальних факторів GreA/B) індукує нуклеазну активність, яка забезпечує редагування помилок при транскрипції.

Особливістю РНК-полімерази II (характерною тільки для цієї полімерази) є наявність у найбільшій субодиниці С-кінцевого домену – **CTD** (**C**-**T**ermin**a**l **D**omain). CTD – це довгий невпорядкований хвіст (не показаний на рис. 5.13), амінокислотна послідовність якого є гептапептидом Tyr-Ser-Pro-Thr-Ser-Pro-Ser, що тандемно повторюється 52 рази. Три залишки Ser у складі гептапептиду є субстратами фосфорилування / дефосфорилування для специфічних кіназ і фосфатаз. CTD, який відходить від полімерази поблизу від каналу виходу мРНК, є платформою для зв'язування численних білків, спорідненість яких залежить від патерну фосфорилування. Зокрема, CTD відіграє ключову роль у перемиканні між ініціацією та елонгацією транскрипції (див. нижче) і збиранні елементів системи процесингу мРНК (розділ 7) під час елонгації.



**Рис. 5.13.** РНК-полімераза II під час елонгації транскрипції (5XON). Праворуч: одна із субодиниць видалена для візуалізації нуклеїнових кислот (матричний ланцюг – темно-синій, транскрипт – червоний); праворуч від 3'-кінця транскрипту показано F-спіраль, подібну до такої, що наявна у складі бактеріальної полімерази (порівн. із рис. 5.2)



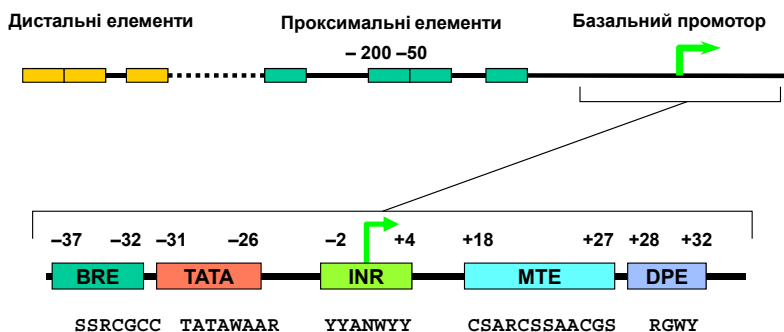
**Рис. 5.14.** Стадії елонгаційного циклу РНК-полімерази II:  
 а – вхід нуклеозидтрифосфату (NTP) до активного центра (1R9T);  
 б – впізнання матриці та зв'язування NTP (1R9S);  
 в – претранслокаційний стан після приєднання нуклеотиду до транскрипту і видалення пірофосфату разом з одним іоном Mg<sup>2+</sup> (116H); г – посттранслокаційний стан (1SFO)

## Ініціація транскрипції РНК-полімеразою II

### Промотор РНК-полімерази II

Розглянемо елементи нуклеотидної послідовності, з яких може складатися промотор РНК-полімерази II, узагальнену схему якого зображено на рис. 5.15 (для конкретних промоторів спостерігаються її численні варіації).

Проксимальні регуляторні елементи (типове розташування приблизно в зоні  $-50 \dots -200$  пар основ відносно старту транскрипції) мають спорідненість до специфічних факторів транскрипції, взаємодія з якими активує / блокує ініціацію (див. розділ 6). Характерною особливістю промоторів РНК-полімерази II є те, що вони часто перебувають також під контролем *дистальних* (розташованих будь-де відносно старту) елементів послідовності (розділ 6). Якщо зона, де розташовані дистальні елементи, підвищує ефективність ініціації, то її називають *енхансером* (enhancer), якщо навпаки – *сайленсером* (silencer). Сукупність проксимальних елементів, що активують транскрипцію, іноді також називають енансером. У зоні старту міститься *базальний промотор*, на який, власне, і завантажується РНК-полімераза. Іноді базальний промотор називають просто промотором (у вузькому сенсі).



**Рис. 5.15.** Приблизна узагальнена схема організації промотору РНК-полімерази II і консенсусні послідовності деяких основних елементів базального промотору. BRE – B Recognition Element, TATA – TATA-box, INR – Initiator, MTE – Motif Ten Element, DPE – Downstream Promoter Element. Стрілкою позначено стартову точку транскрипції

Базальний промотор може включати (не обов'язково) кілька стандартних елементів послідовності (рис. 5.15): TATA-бокс на відстані приблизно  $-30$  пар основ від стартової точки; ініціаторний елемент (Initiator, Inr) безпосередньо в зоні старту; DPE (Downstream Promoter Element) на відстані приблизно  $+30$  пар

основ від старту; розташований поряд з ТАТА-боксом BRE (B Recognition Element), який упізнається базальним фактором транскрипції TFIIВ (див. нижче); розміщений між Inr та DPE MTE (Motif Ten Element). Приблизно половина промоторів у геномі людини містять Inr, трохи більше 10 % – DPE, до 20 % – ТАТА-бокс. Inr та DPE часто наявні (якщо наявні) одночасно, DPE і ТАТА-бокс, також як ТАТА-бокс і BRE, навпаки, мають тенденцію уникати один одного. При цьому достатньо велика частина промоторів не містять жодного із зазначених стандартних елементів – тоді промотор визначається або іншими, більш специфічними, елементами послідовності у проксимальній зоні, або особливостями структури хроматину (див. розділ 6).

### **Загальні фактори транскрипції РНК-полімерази II і преініціаторний комплекс**

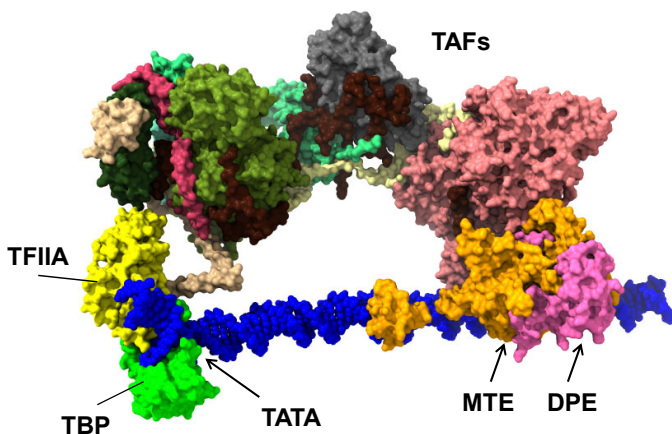
РНК-полімеразу II на рис. 5.13 можна вважати корферментом, що здійснює елонгацію транскрипції. Для забезпечення ініціації вона, так само як і бактеріальна РНК-полімераза, потребує загальних факторів ініціації. Однак замість одного  $\sigma$ -фактора тут їх шість – для ініціації транскрипції необхідне збирання на базальному промоторі *преініціаторного комплексу* (PIC – Pre-Initiation Complex) у складі РНК-полімерази II та загальних (базальних) факторів транскрипції TFIIA, TFIIВ, TFIIД, TFIIЕ, TFIIF, TFIIH.

Послідовність збирання преініціаторного комплексу може бути різною, але всі базальні фактори мають бути у складі PIC для подальшого запуску транскрипції (незалежно від наявності в базальному промоторі відповідних елементів, до яких деякі з факторів мають підвищену спорідненість).

**TFIIД** – мультибілковий комплекс, один із найважливіших базальних факторів, який визначає впізнання стандартних елементів базального промотору (якщо вони наявні). Найважливішою складовою TFIIД є **TBP** (TATA-box Binding Protein) – білок з підвищеною спорідненістю до ТАТА-боксу. Незважаючи на цю підвищену спорідненість, TBP чудово зв'язується і з іншими послідовностями ДНК. Доволі широкий  $\beta$ -шар у складі TBP взаємодіє з маленьким жолобком ДНК, наслідком чого є суттєва

деформація подвійної спіралі з її розкручуванням і значним вигинанням у протилежний від білка бік (див. рис. 3.18). Саме із цього протилежного боку (усередині вигину) розташовуються інші білкові компоненти преініціаторного комплексу. Отже, вигин ДНК, який індукується ТВР, організує промотор і зв'язані з ним елементи в єдиний компактний комплекс і підтримує його стабільність. Часто (але не обов'язково) зв'язування ТВР у складі TFIIID ініціює збирання PIC.

Крім ТВР до складу TFIIID входять ТВР-асоційовані фактори (TAFs – TBP Associated Factors) – до 13 білків, деякі з яких наявні у двох копіях (конкретний склад факторів може варіювати для різних промоторів). Комплекс TFIIID доволі гнучкий і здатний до структурних перебудов (одну зі структурних форм, що реалізується у комплексі з базальним промотором, зображено на рис. 5.16). TAFs взаємодіють між собою, деякі – з ТВР та іншими базальними факторами транскрипції, деякі – із ДНК. Зокрема, певні ТВР-асоційовані фактори здатні впізнавати стандартні елементи промотору нижче стартової точки (MTE, DPE, рис. 5.16), тобто забезпечувати зв'язування TFIIID із базальним промотором, у складі якого відсутній ТАТА-бокс.



**Рис. 5.16.** Комплекс TFIIID із базальним промотором і загальним фактором транскрипції TFIIA (6M2M)

Крім того, деякі TAFs містять структурні блоки, що мають спорідненість до характерних маркерів активації транскрипції у гістонах (розділ 4) – ацетилованих лізінів, триметильованого Lys4 в гістоні H3. У такий спосіб TFIID впізнає розміщені у промоторі нуклеосоми, що несуть зазначені маркери (більш детально цей аспект розглядатиметься у розділі 6). Нарешті, один із TBP-асоційованих факторів (гістон-ацетилтрансфераза) може здійснювати ацетилювання гістонів у промоторній зоні.

Загалом TFIID можна розглядати як засіб завантаження TBP – білка, абсолютно необхідного для ефективного збирання пре-ініціаторного комплексу, – у ділянку приблизно –30 пар основ від стартової точки транскрипції. Не обов'язково всі TAFs залишаються у цьому комплексі після його остаточної збірки. TFIID є не єдиним комплексом, що використовується для завантаження TBP. Інший варіант такого комплексу – SAGA (PCAF у людини), який містить гістон-ацетилтрансферазу і набір TAFs, частина яких входять також до складу TFIID.

**TFIIA** – гетеродимерний білок, який можна було б розглядати як TBP-асоційований фактор. Він додатково стабілізує комплекс ДНК із TBP (рис. 5.16), а також із загальним фактором транскрипції TFIIB.

**TFIIB** – фактор, найважливіший для ініціації транскрипції – мономерний білок із мультидоменною структурою. Два С-кінцевих домени взаємодіють із ДНК по обидва боки від сайту зв'язування TBP, із внутрішнього боку вигину, який індукується цим білком (рис. 5.17). Два білки, які взаємодіють також між собою, зв'язуються з ДНК *кооперативно*, взаємно підсилюючи спорідненість до неї. TFIIB бере участь у впізнанні базального промотору, взаємодіючи із підвищеною спорідненістю з елементами BRE (рис. 5.15), якщо вони наявні. Однак цим функція TFIIB далеко не вичерпується. N-кінцева частина білка реалізує численні взаємодії з РНК-полімеразою (рис. 5.17): у процесі ініціації транскрипції TFIIB стабілізує перші два нуклеотиди на початку синтезу й короткий первинний гібрид ДНК-РНК, а також блокує канал виходу РНК, тобто TFIIB практично є аналогом бактеріального  $\sigma$ -фактора (див. вище). Відповідно TFIIB, з одного

боку, часто спричинює абортівну ініціацію, з іншого – має бути звільненим при переході від ініціації до елонгації транскрипції.

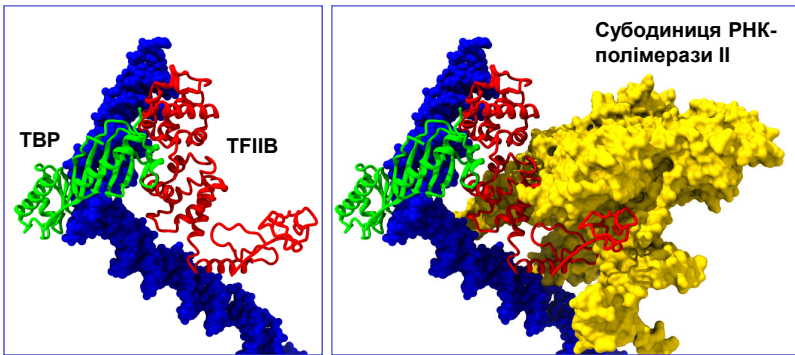


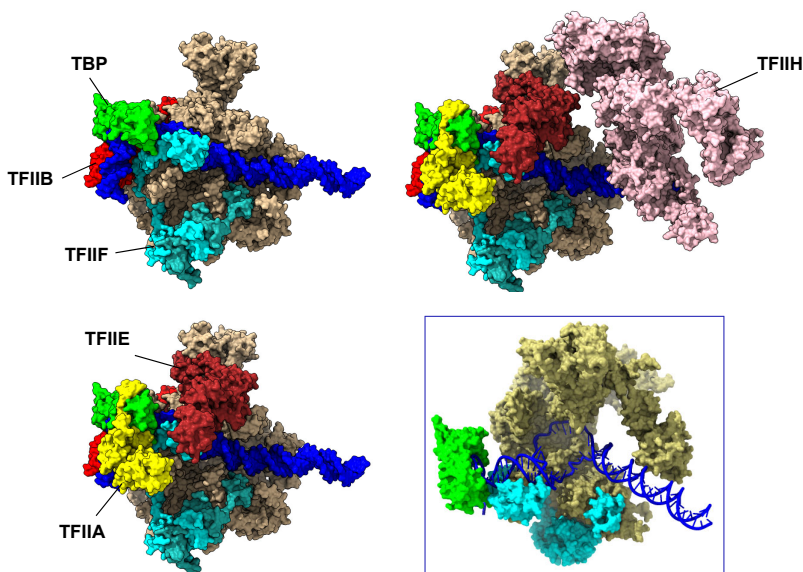
Рис. 5.17. Комплекс ДНК із TBP і TFIIIB та РНК-полімеразою II (представлено одну її субодинацю, 5FMF)

**TFIIF** – гетеродимер, взаємодіє з РНК-полімеразою II, ДНК, базальними факторами TFIIE та TFIIH (рис. 5.18). Після первинного плавлення ДНК при ініціації транскрипції TFIIF здійснює взаємодії з нематричним ланцюгом. Після завершення ініціації TFIIF разом з іншими базальними факторами дисоціює від РНК-полімерази, але може реасоціювати з нею під час елонгації у випадку зупинки ферменту. Його функція як фактора елонгації транскрипції полягає у стимуляції продовження руху полімерази.

**TFIIE** – гетеродимер (рис. 5.18), сприяє зв'язуванню TFIIH і стимулює його роботу, взаємодіє з розплавленою зоною ДНК після утворення відкритого комплексу.

**TFIIH** – найбільший із базальних факторів, складається із 10 субодинаць і володіє двома важливими ферментативними активностями. Перша з них – АТР-залежна **ДНК-геліказна** (гелікази – helicases – ферменти, що розкручують подвійну спіраль). Частина TFIIH взаємодіє з РНК-полімеразою та іншими базальними факторами, частина – з ДНК нижче по ходу транскрипції від активного центра полімерази (рис. 5.18). Конформаційні зміни білка, зумовлені гідролізом АТР, штовхають ДНК усередину полімеразного комплексу, де для подвійної спіралі немає місця – два ланцюги розводяться і захоплюються певними

структурними елементами РНК-полімерази, тобто утворюється відкритий комплекс (рис. 5.18). РНК-полімераза II є єдиною з РНК-полімераз, яка потребує такої геліказної активності при ініціації транскрипції. Інші РНК-полімерази, еукаріотичні та бактеріальні, здійснюють локальне плавлення ДНК завдяки певним ДНК-білковим взаємодіям у специфічних промоторах. ТFIID надає РНК-полімеразі II принципову можливість завантажитись на будь-яку ділянку ДНК, але наявність в еукаріотичному хроматині нуклеосом заважає такому безладному зв'язуванню – структура хроматину є одним із найважливіших факторів регуляції транскрипції (розділ 6).

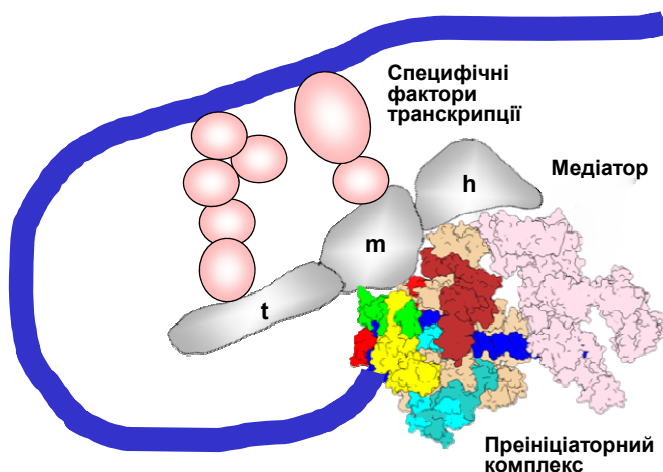


**Рис. 5.18.** Преініціаторний закритий комплекс РНК-полімерази II з ДНК і різними наборами базальних факторів транскрипції (5IY6). Праворуч унизу – відкритий комплекс, що утворюється внаслідок активності TFIID (більшість базальних факторів і кілька субодиниць полімерази не показано для візуалізації розплавленої зони, 5IYB)

Друга ферментативна активність TFIID – *кіназна*. Частину комплексу TFIID, що відповідає за цю активність, позначають також як *TFIIK*. За рахунок кіназної активності здійснюється фосфорилування С-кінцевого домену РНК-полімерази при пе-

реході від ініціації до елонгації транскрипції. У дефосфорильованій формі, на стадії збирання преініціаторного комплексу, CTD взаємодіє з базальними факторами транскрипції; після фосфорилування певних залишків Ser ця взаємодія порушується.

**Медіатор.** Ефективне збирання преініціаторного комплексу можливе лише за участю ще одного структурного модуля – медіатора (mediator), який містить понад 20 субодиниць (склад варіює для різних промоторів – певний мінімальний набір субодиниць доповнюється більш специфічними за рахунок взаємодії із транскрипційними факторами). Цей мультибілковий комплекс має видовжену структуру, яку можна розділити на три частини: голова (h, head), середня частина (m, middle) і хвіст (t, tail). Медіатор здійснює лише білок-білкові взаємодії: голова й середня частина взаємодіють з РНК-полімеразою та іншими компонентами преініціаторного комплексу, середня частина та хвіст – зі специфічними факторами транскрипції, що зв'язані на проксимальних і дистальних елементах промотору (рис. 5.19). Отже, медіатор є засобом передавання активаційних сигналів із регуляторних елементів послідовності на РНК-полімеразу: збільшення кількості взаємодій підсилює ефективність збирання РІС.



**Рис. 5.19.** Схема преініціаторного комплексу РНК-полімерази II, що взаємодіє зі специфічними факторами транскрипції за допомогою медіатора

## Процес ініціації транскрипції

Збирання преініціаторного комплексу є першою подією процесу ініціації. Безпосередній результат цього процесу – *закритий* преініціаторний комплекс (рис. 5.18, 5.20).

Наступна подія – утворення *відкритого* комплексу – АТР-залежне локальне плавлення ДНК за рахунок геліказної активності ТFIІН. Нематричний ланцюг після цього захоплюється фактором ТFIІF, матричний – занурюється в активний центр, де взаємодіє також із N-кінцевою частиною ТFIІВ. Починається синтез короткого первинного транскрипту, коротка гібридна подвійна спіраль стабілізується тим самим N-кінцевим доменом ТFIІВ, який при цьому блокує канал виходу РНК.

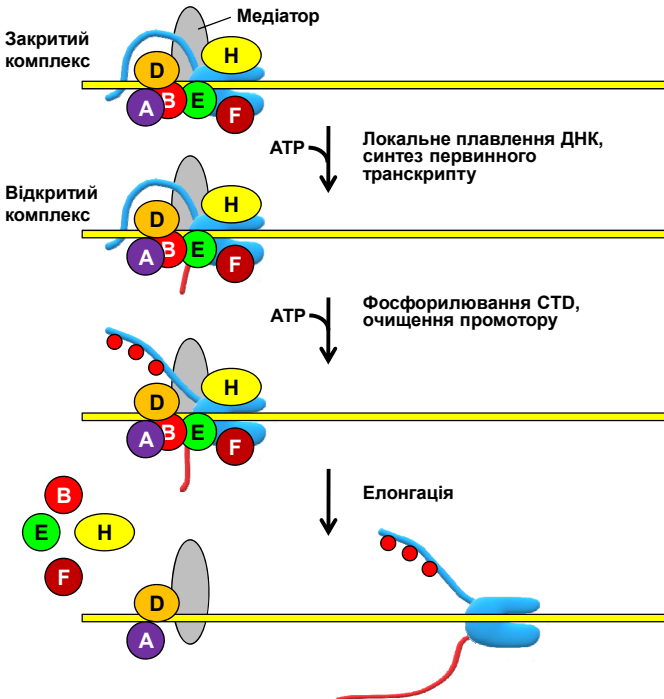


Рис. 5.20. Процес ініціації транскрипції РНК-полімеразою II

На наступному кроці спрацьовує кіназна активність ТФІІН: відбувається фосфорилування СТД, яке є точкою перемикавання ініціації на елонгацію. СТД втрачає зв'язок із базальними факторами транскрипції, які дисоціюють від полімерази – відбувається очищення промотору. РНК-полімераза продовжує синтез РНК, а частина базальних факторів може залишатися на базальному промоторі й ініціювати зв'язування іншої полімерази.

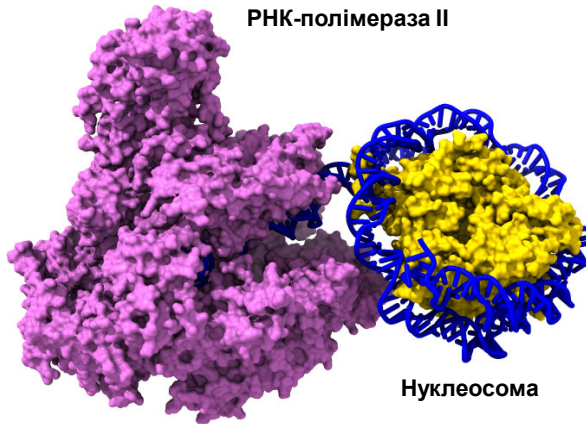
Після термінації транскрипції (цей процес в еукаріотів тісно пов'язаний із процесингом мРНК і буде розглядатись у розділі 7) РНК-полімераза зв'язується із ТФІІФ, який рекрутує специфічну фосфатазу, – відбувається дефосфорилування СТД, і полімераза вступає в новий робочий цикл.

Дослідження кінетики роботи РНК-полімерази II на певному гені у живих клітинах дають такі оцінки проміжків часу, який полімераза перебуває у тому чи іншому стані, та ймовірностей переходів від одного стану до іншого (напевно, на різних генах тривалість часових інтервалів та інші параметри можуть варіювати залежно від ефективності роботи системи регуляції): середній час імобілізації РНК-полімерази на промоторі після зв'язування з ним – 6 с; більшість молекул полімерази після цього дисоціюють, лише 13 % зв'язаних полімераз переходять до ініціації транскрипції – тоді цей етап триває у середньому 54 с; лише 9 % таких полімераз переходять до стадії елонгації транскрипції, решта здійснюють абортивну ініціацію. Отже, тільки одна молекула полімерази з ~90 таких, що зв'язались із промотором, продукує молекулу мРНК. Така низька ефективність може мати біологічне значення, знижуючи ймовірність випадкової транскрипції будь-де в геномі (*транскрипційний шум*). Швидкість безупинної елонгації транскрипції становить 30–70 нуклеотидів за секунду. Однак частина полімераз зупиняються у паузах до кількох хвилин.

Основною причиною таких пауз є наявність нуклеосом – фундаментальна особливість процесу транскрипції в еукаріотів полягає в тому, що РНК-полімераза змушена працювати на хроматиновій фібрилі.

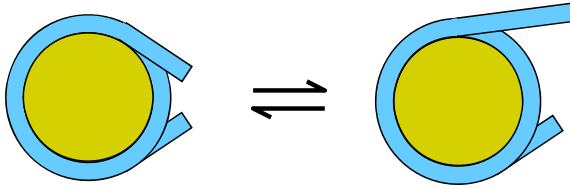
## Елонгація РНК-полімерази через хроматин

Зрозуміло, що на стадії елонгації транскрипції нуклеосома має створювати суттєві перешкоди для проходження РНК-полімерази (рис. 5.21). При цьому хроматин, навіть у межах активних генів, загалом зберігає нуклеосомну структуру.



**Рис. 5.21.** Нуклеосомний бар'єр на шляху РНК-полімерази (6184)

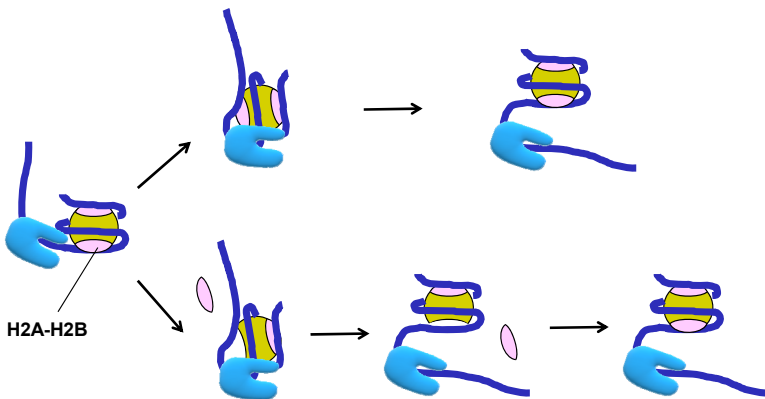
Для проникнення всередину нуклеосомної ДНК РНК-полімераза використовує структурну динаміку нуклеосоми. Електростатичні взаємодії між ДНК і гістонами реалізуються кожного разу, коли маленький жолобок подвійної спіралі контактує з поверхнею октамера гістонів (загалом здійснюються 14 таких контактів, див. розділ 4). Як обговорювалось у розділі 4, за фізіологічних значень іонної сили сумарна вільна енергія цих взаємодій є такою, що нуклеосома жорстко фіксується на ДНК. Однак енергетична цінність одного контакту порівняно невисока. Відповідно спостерігається спонтанне руйнування контактів на виході з нуклеосоми із частковим розкручуванням нуклеосомної ДНК – приблизно 10 % часу нуклеосома існує у такій відкритій формі (рис. 5.22).



**Рис. 5.22.** Структурна динаміка нуклеосоми з тимчасовим руйнуванням контакту ДНК з гістонами на виході з нуклеосоми

Зустрічаючи на своєму шляху нуклеосому, РНК-полімераза чекає на таке розкручування, просувається вперед, чекає на руйнування наступного контакту і т. д. При цьому частина гістонової поверхні, яка звільнилась від взаємодій із нуклеосомною ДНК, починає взаємодіяти з вільною ДНК на сусідній лінкерній ділянці. Результатом стає перенесення нуклеосоми на ділянку позаду від полімерази (рис. 5.23). Залежно від швидкості руху полімерази може відбуватись або перенесення октамеру гістонів як цілого (за порівняно низької швидкості), або (при підвищенні швидкості руху) частинами (у вигляді гістонових комплексів). Зокрема, часто спостерігаються видалення одного димеру H2A-H2B, перенесення гексамеру гістонів і реасоціація димеру на ділянці позаду від РНК-полімерази (рис. 5.23).

Описані процеси роблять нуклеосому практично "прозорою" для РНК-полімерази, але за однієї простої умови: усе це спостерігається в експериментах *in vitro* за дещо підвищеної (до  $\sim 300$  ммоль/л) іонної сили, яка послаблює електростатичні ДНК-гістонові взаємодії. Відповідно у хроматині мають існувати та існують фактори елонгації транскрипції, які, беручи на себе роль підвищеної іонної сили, полегшують проходження РНК-полімерази.



**Рис. 5.23.** Проходження РНК-полімерази через нуклеосому

До таких факторів елонгації транскрипції відносять:

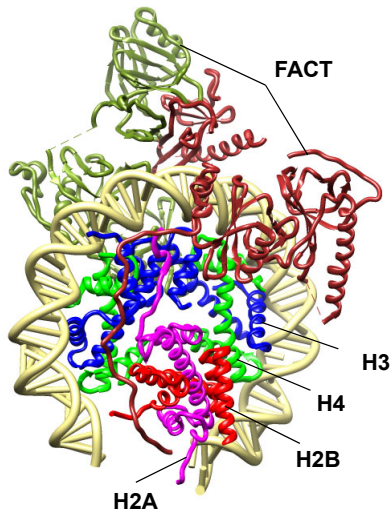
- ***elongator*** – мультибілковий комплекс, одна із субодиниць якого є гістон-ацетилювальний фермент. Комплекс взаємодіє з РНК-полімеразою II і здійснює ацетилювання гістонів, яке знижує їхню спорідненість до ДНК, під час елонгації транскрипції;
- ***Spt 4-5*** – димер, зв'язаний з РНК-полімеразою II, який полегшує її проходження через нуклеосому, конкуруючи з ДНК за гістони;
- ***FACT* (FAcilitates Chromatin Transcription)** – складається із двох субодиниць, здатен взаємодіяти з нуклеосомою (рис. 5.24) і має підвищену спорідненість до димерів H2A-H2B; виконує функцію проміжного переносника димерів, забезпечуючи їх видалення з нуклеосом попереду полімерази й повернення до хроматину позаду;
- ***Spt6*** – проміжний акцептор зі спорідненістю до гістонів H3-H4.

Отже, переважно роль факторів, що полегшують проходження полімерази, виконують проміжні акцептори гістонових комплексів (див. рис. 4.10). Крім зазначених вище, у хроматині наявні також інші проміжні акцептори гістонів.

При здійсненні транскрипції у петельних доменах хроматину, кінці яких жорстко зафіксовані (розділ 4), виникає ще один фактор, що впливає на нуклеосоми. Проходження РНК-полімерази (як і будь-якого іншого ферменту, що здійснює транслокацію вздовж ДНК) за умов закріплення кінців матриці (основ хроматинової петлі) має створювати позитивну надспіралізацію попереду й негативну – позаду полімеразного комплексу (див. рис. 3.25). Надспіралізація у процесі транскрипції виникає у зв'язку з тим, що обертальний рух РНК-полімеразного комплексу навкруг подвійної спіралі значно загальмований унаслідок масивності комплексу, у тому числі наявності в його складі довгого ланцюга РНК, що синтезується. Крім цього, значна кількість полімераз залучені до великих мультимолекулярних комплексів – транскрипційних фабрик (розділ 6). Позитивна надспіралізація дестабілізує *ліву* суперспіраль нуклеосомної ДНК попереду полімерази, що сприяє дії проміжних переносників гістонів. І навпаки – негативна надспіралізація позаду полімерази сприяє ефективному відновленню нуклеосом.

Завдяки описаним механізмам в еукаріотичних клітинах здійснюється ефективна елонгація транскрипції при збереженні нуклеосомної структури навіть у транскрипційно активних ділянках. Проте іноді спостерігається повне видалення нуклеосом по всій довжині петельного домену на початку транскрипції. Такий ефект описаний для випадку, коли транскрипція відбувається із двох близько розташованих промоторів, спрямованих у різні боки: дві хвилі позитивної надспіралізації дестабілізують нуклеосоми і сприяють перенесенню гістонових комплексів на проміжні акцептори особливого типу, "розчищаючи дорогу" для двох РНК-полімераз, що рухаються у протилежних напрямках.

Крім того, наявність нуклеосом залежить від рівня транскрипційної активності – кількості РНК-полімераз, що одночасно працюють на певному гені. Якщо ген гіперактивний, то ДНК може бути недоступною для гістонових комплексів, які залишаються на проміжних акцепторах. Зокрема, така ситуація спостерігається у кластерах генів рибосомних РНК, що транскрибуються РНК-полімеразою I.



**Рис. 5.24.** Комплекс двох субодиниць фактора елонгації транскрипції FACT із субнуклеосомною частинкою, що містить гексамер гістонів – тетрамер  $(H3-H4)_2$  і один димер H2A-H2B (6УРК)

## РНК-полімерази I і III

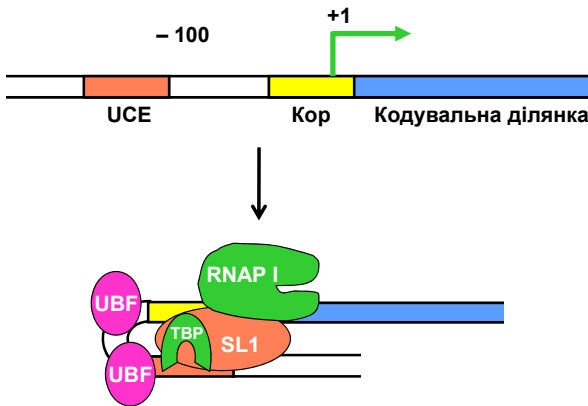
Оскільки РНК-полімерази I і III мають обмежену кількість мішеней, то системи ініціації транскрипції для них значно простіші порівняно з РНК-полімеразою II.

### **Ініціація транскрипції генів рибосомної РНК**

Промотор кожного кластера генів рибосомної РНК (рис. 5.25, див. також рис. 4.6) містить два елементи послідовності: корелемент у зоні стартової точки та UCE (Upstream Control Element) на відстані приблизно 100 пар основ від старту. Ініціація залежить від двох основних транскрипційних факторів: SL1 (іноді позначається TFIB) і UBF (Upstream Binding Factor).

Зв'язування UBF є передумовою ініціації, але цей фактор не має високої специфічності щодо елементів промотору: він зв'я-

зується будь-де в межах рибосомного кластера. Димери UBF містять у своїй структурі HMG-бокси (див. рис. 3.17), які сприяють вигинам ДНК. У результаті зростає імовірність наближення у просторі двох елементів промотору, з якими специфічно взаємодіє SL1. Цей фактор є аналогом фактора TFIID РНК-полімерази II, він містить TBP і певний специфічний набір асоційованих із ним факторів, від яких залежить спорідненість SL1 до промотору рибосомного кластера. SL1 рекрутує до промотору РНК-полімеразу I, яка починає синтез попередника рРНК.



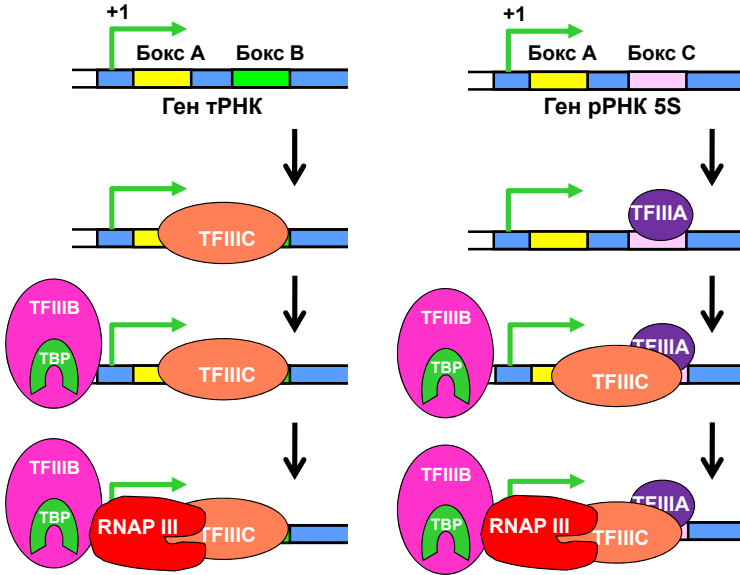
**Рис. 5.25.** Ініціація транскрипції РНК-полімеразою I

Отже, для РНК-полімерази I виконується та сама закономірність, що й для РНК-полімерази II: незалежно від наявності чи відсутності ТАТА-боксу (який відсутній у промоторі RNAPI), TBP має бути обов'язковим компонентом преініціаторного комплексу, сприяючи організації промоторної ДНК навкруг полімерази. Ця закономірність зберігається також для РНК-полімерази III.

### **Ініціація транскрипції РНК-полімеразою III**

Промотори генів тРНК і рРНК 5S, які транскрибуються РНК-полімеразою III, містяться всередині кодувальної частини гена (рис. 5.26). Два елементи промотору гена тРНК, бокси А і В,

упізнаються фактором TFIIIC, що рекрутує мультибілковий комплекс TFIIIB, який, своєю чергою, викликає зв'язування полімерази. Фактор TFIIIB містить TBP і кілька специфічних для РНК-полімерази III TAFs.



**Рис. 5.26.** Ініціація транскрипції РНК-полімеразою III

Усередині гена рРНК 5S замість боксу В міститься інший елемент – бокс С, що упізнається фактором TFIIIA (рис. 5.26). Далі цей білок рекрутує TFIIIC і формування преініціаторного комплексу продовжується так само, як і на гені тРНК.

Описані події відбуваються однократно. Коли починається транскрипція, полімераза витісняє TFIIIC (і TFIIIA), що знаходиться попереду від неї. TFIIIB при цьому залишається на місці, рекрутуючи нову полімеразу для нового раунду транскрипції.

## Контрольні запитання

1. Як розрізняються кор- і голофермент бактеріальної РНК-полімерази? Яке значення має ця різниця для ініціації транскрипції?

2. Навіщо для ініціації транскрипції потрібен особливий білковий фактор? Опишіть функціональну роль фактора  $\sigma$ .

3. Яку будову має стандартний бактеріальний промотор?

4. Чим розрізняються закритий і відкритий комплекси РНК-полімерази із промотором?

5. У якому напрямку здійснюється зчитування інформації з ДНК під час транскрипції? Що таке змістовний і антизмістовний ланцюг? Який із них є матричним?

6. Як забезпечується висока процесивність РНК-полімерази?

7. Як організовано активний центр РНК-полімерази? Які сполуки є субстратами (будівним матеріалом) синтезу РНК?

8. Опишіть основні етапи елонгаційного циклу РНК-полімерази. Чому фермент рухається вздовж ДНК під час транскрипції?

9. У який спосіб здійснюється при транскрипції редагування помилок приєднання нуклеотидів?

10. Як здійснюється термінація транскрипції у бактерій?

11. Порівняйте структуру РНК-полімерази II з такою бактеріальною полімерази. Що є спільного між ними й чим вони відрізняються?

12. Яку будову й функціональне значення має С-кінцевий домен (CTD) РНК-полімерази II?

13. Яку будову має промотор РНК-полімерази II?

14. Охарактеризуйте базальні фактори транскрипції РНК-полімерази II й основні етапи ініціації транскрипції.

15. Яке значення має білок ТВР? Яку роль відіграють фактори ТФІВ та ТФІН у ініціації транскрипції РНК-полімеразою II?

16. Що таке медіатор? Яку функцію він виконує?

17. Яку спеціалізацію мають еукаріотичні РНК-полімерази різних типів? Порівняйте системи ініціації транскрипції РНК-полімераз I та III. Що спільного між ними й системою ініціації РНК-полімерази II?

18. Опишіть механізми, за допомогою яких здійснюється рух еукаріотичної РНК-полімерази через нуклеосому.

## Рекомендована література

- Bai, L., Santangelo, T.J. and Wang, M.D. (2006) Single-molecule analysis of RNA polymerase transcription. *Annu. Rev. Biophys. Biomol. Struct.* 35, 343–360.
- Ben-Shem, A., Papai, G. and Schultz, P. (2021) Architecture of the multi-functional SAGA complex and the molecular mechanism of holding TBP. *FEBS J.* 288, 3135–3147.
- Bintu, L., Kopaczynska, M., Hodges, C., et al. (2011) The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nat. Struct. Mol. Biol.* 18, 1394–1399.
- Boeger, H., Bushnell, D.A., Davis, R., et al. (2005) Structural basis of eukaryotic gene transcription. *FEBS Letters* 579, 899–903.
- Ciampi, M.S. (2007) Rho-dependent terminators and transcription termination. *Microbiology* 152, 2515–2528.
- Cramer, P., Armache, K.J., Baumli, S., et al. (2008) Structure of eukaryotic RNA polymerases. *Annu. Rev. Biophys.* 37, 337–352.
- Darzacq, X., Shav-Tal, Y., de Turris, V., et al. (2007) In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806.
- Fazal, F.M., Meng, C.A., Murakami, K., et al. (2015) Real-time observation of the initiation of RNA polymerase II transcription. *Nature* 525, 274–277.
- Geiduschek, E.P. and Kassavetis, G.A. (2001) The RNA polymerase III transcription apparatus. *J. Mol. Biol.* 310, 1–26.
- Gouge, J., Satia, K., Guthertz, N., et al. (2015) Redox signaling by the RNA polymerase III TFIIB-related factor Brf2. *Cell* 163, 1375–1387.
- Greive, S.J. and von Hippel, P.H. (2005) Thinking quantitatively about transcriptional regulation. *Nat. Rev.* 6, 221–232.
- Griesenbeck, J., Tschochner, H. and Grohmann, D. (2017) Structure and function of RNA polymerases and the transcription machineries. *Subcell. Biochem.* 83, 225–270.

Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Struct. Mol. Biol.* 11, 394–403.

He, Y., Fang, J., Taatjes, D.J. and Nogales, E. (2013) Structural visualization of key steps in human transcription initiation. *Nature* 495, 481–486.

Heintzman, N.D. and Ren, B. (2007) The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell. Mol. Life Sci.* 64, 386–400.

Herbert, K.M., Greenleaf, W.J. and Block, S.M. (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* 77, 149–76.

Hodges, C., Bintu, L., Lubkowska, L., et al. (2009) Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* 325, 626–628.

Kujirai, T. and Kurumizaka, H. (2020) Transcription through the nucleosome. *Curr Opin Struct Biol.* 61, 42–49.

Lai, W.K.M. and Pugh, B.F. (2017) Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell. Biol.* 18, 548–562.

Lane, W.J. and Darst, S.A. (2006) The structural basis for promoter –35 element recognition by the group IV  $\sigma$  factors. *PLoS Biol.* 4, e269.

Larson, M.H., Greenleaf, W.J., Landick, R. and Block, S.M. (2008) Applied force reveals mechanistic and energetic details of transcription termination. *Cell* 132, 971–982.

Lee, J. and Borukhov, S. (2016) Bacterial RNA polymerase-DNA interaction – the driving force of gene expression and the target for drug action. *Front Mol Biosci.* 3, 73.

Lenhard, B., Sandelin, A. and Carninci, P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* 13, 233–245.

Liu, X., Bushnell, D.A., Wang, D., et al. (2010) Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* 327, 206–209.

Liu, Y., Zhou, K., Zhang, N., Wei H, et al. (2020) FACT caught in the act of manipulating the nucleosome. *Nature* 577, 426–431.

Louder, R.K., He, Y., López-Blanco, J.R., et al. (2016) Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* 531, 604–609.

Murakami, K., Elmlund, H., Kalisman, N., et al. (2013) Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342, 1238724.

Murakami, K.S. (2015) Structural biology of bacterial RNA polymerase. *Biomolecules* 5, 848–864.

Nudler, E. (2009) RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* 78, 335–361.

Patel, A.B., Louder, R.K., Greber, B.J., et al. (2018) Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* 362, eaau8872.

Plaschka, C., Hantsche, M., Dienemann, C., et al. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* 533, 353–358.

Roberts, J.W. (2019) Mechanisms of bacterial transcription termination. *J. Mol. Biol.* 431, 4030–4039.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., et al. (2016) Structure of a complete mediator-RNA polymerase II pre-initiation complex. *Cell* 166, 1411–1422.

Russell, J. and Zomerdijk, J.C. (2006) The RNA polymerase I transcription machinery. *Biochem. Soc. Symp.* 73, 203–216.

Schier, A.C. and Taatjes, D.J. (2020) Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* 34, 465–488.

Sims, R.J., Belotserkovskaya, R. and Reinberg, D. (2004) Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* 18, 2437–2468.

Steitz, T.A. (2006) Visualizing polynucleotide polymerase machines at work. *EMBO J.* 25, 3458–3468.

Studitsky, V.M., Walter, W., Kireeva, M., et al. (2004) Chromatin remodeling by RNA polymerases. *Trends Biochem. Sci.* 29, 127–135.

Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119, 481–489.

Young, B.A., Gruber, T.M. and Gross, C.A. (2002) Views of transcription initiation. *Cell* 109, 417–420.

## Розділ 6

# РЕГУЛЯЦІЯ ТРАНСКРИПЦІЇ

According to the strictly structural concept, the genome is considered as a mosaic of independent molecular blue-prints for the building of individual cellular constituents. In the execution of these plans, however, co-ordination is evidently of absolute survival value.

*F. Jacob, J. Monod*

Only when a system behaves in a sufficiently random way may the difference between past and future, and therefore irreversibility, enter its description.

*I. Prigogine, I. Stengers. Order Out of Chaos*

У переважній більшості гени не транскрибуються постійно (за винятком згаданих у розділі 4 генів домашнього господарства, активність яких завжди підтримується на більш-менш високому рівні), а вмикаються / вимикаються в певні моменти залежно від типу клітини, зовнішніх умов, стадій клітинного циклу тощо. Головним, хоча й не єдиним моментом, у який здійснюється регуляторний вплив на експресію генів, є ініціація транскрипції. У складній системі регуляції транскрипції можна виділити три основні шляхи, якими відбувається вплив на промотори: білкові фактори транскрипції, регуляторні РНК різних типів і (для еукаріотів) зміни у структурі хроматину. Хоча всі три шляхи тісно пов'язані між собою, транскрипційні фактори є основним інструментом регуляції. Мішенями, специфічна взаємодія з якими транскрипційних факторів (*транс*-факторів), що вільно дифундують (*транспортуються*) у просторі клітини, зумовлює активацію чи репресію транскрипції, є *цис-елементи* – регуляторні елементи послідовності ДНК, які фізично зв'язані з певним геном. В еукаріотів *цис*-елементи розміщені у проксимальних і дистальних регуляторних зонах (див. рис. 5.15), у прокаріотів

вони часто називаються *операторами* й зазвичай розташовані в безпосередній близькості до промоторів. Якщо зв'язування *транс*-фактора із *цис*-елементом зумовлює активацію транскрипції, то кажуть, що фактор є *активатором* і здійснює *позитивну* регуляцію. Якщо фактор блокує зв'язування РНК-полімерази, то його називають *репресором* і кажуть про *негативну* регуляцію.

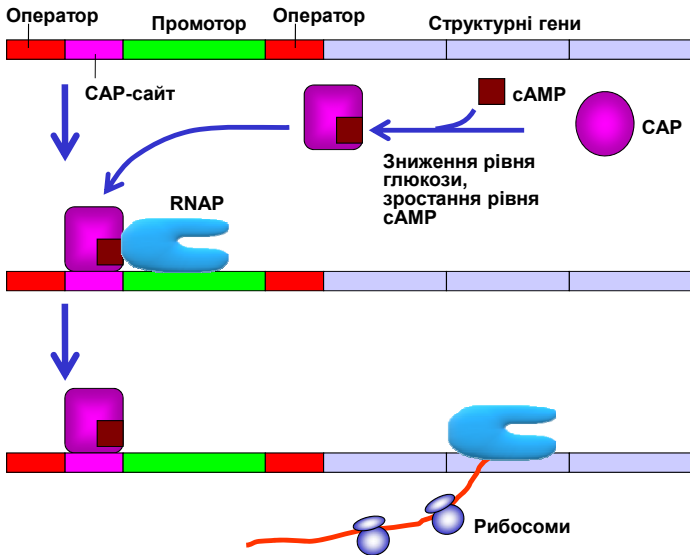
На початку цього розділу деякі загальні принципи регуляції проілюстровано на типових прикладах для бактеріальної системи транскрипції. У наступних підрозділах розглядатимуться механізми регуляції транскрипції в еукаріотів.

## Регуляція транскрипції у бактерій

Особливостями транскрипційної системи бактерій є порівняно невелика кількість генів, наявність оперонів (одиниць транскрипції, де об'єднано по кілька структурних генів, див. розділ 4), відсутність складних хроматинових структур і тісна координація з білковим синтезом, який відбувається під час транскрипції.

### Лактозний оперон

Лактозний оперон (*lac*-оперон) *E. coli* став у свій час, завдяки дослідженням Жакоба і Моно (François Jacob, Jacques Monod), першою детально вивченою системою регуляції транскрипції. До складу оперона (рис. 6.1) входять три структурні гени, що кодують ферменти, залучені до катаболізму лактози. Транскрипція усіх трьох генів відбувається з одного промотору (синтезується єдина *поліцистронна* молекула мРНК, яка має три послідовні відкриті рамки зчитування). Промотор оточують дві однакові операторні ділянки (*lac*-оператори), що мають спорідненість до *lac*-репресора, і сайт зв'язування білка-активатора катаболізму CAP (Catabolite Activator Protein).

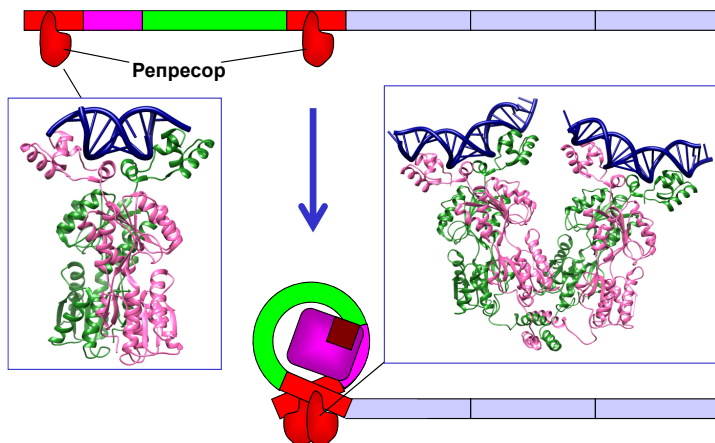


**Рис. 6.1.** Позитивна регуляція *lac*-оперона білком-активатором катаболізму CAP

Промотор *lac*-оперона є слабким і має доволі низьку спорідненість до РНК-полімерази. Навіть якщо в середовищі є лактоза, але наявна також глюкоза (з якої починається магістральний шлях катаболізму в усіх організмів), то транскрипція *lac*-оперона майже не відбувається. Зниження рівня глюкози зумовлює підвищення внутрішньоклітинної концентрації сAMP (циклічного аденозинмонофосфату), зв'язування якого із CAP індукуює конформаційну перебудову білка та появу його специфічної спорідненості до відповідного сайту на ДНК (див. структуру комплексу на рис. 3.13, б). Взаємодія CAP із РНК-полімеразою підсилює її спорідненість до промотору – CAP рекрутує полімеразу, яка далі розпочинає синтез мРНК (рис. 6.1).

Описаний сценарій позитивної регуляції реалізується лише за умови, що *lac*-оператори не взаємодіють із *lac*-репресором. За відсутності лактози (коли ферменти її утилізації не потрібні) гомодимери репресора (незалежно від можливої наявності CAP) зв'язуються з обома операторами (рис. 6.2, див. також рис. 3.22)

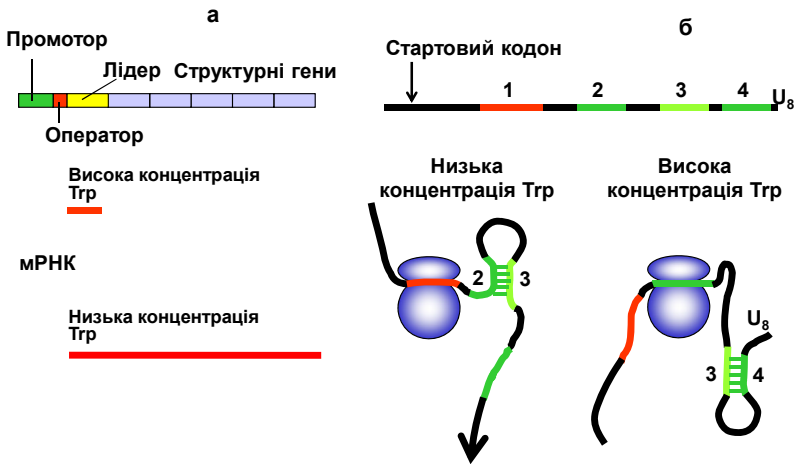
і при цьому взаємодіють між собою: утворюється тетрамерний комплекс, що утримує петлю ДНК розміром ~90 пар основ (рис. 6.2). Усередині петлі розташований промотор, що запобігає зв'язуванню з ним РНК-полімерази. Коли з'являється лактоза, її невелика кількість перетворюється на алолактозу, яка спрацьовує як *індуктор* *lac*-операона: зв'язування алолактози з репресором індукує втрату його спорідненості до оператора. Унаслідок руйнування петлі РНК-полімераза зв'язується із промотором і оперон починає працювати.



**Рис. 6.2.** Негативна регуляція *lac*-операона *lac*-репресором. На вставках: структура димеру репресора в комплексі з оператором (1JWL) і тетрамерний комплекс із двома операторами (1LBG)

### Триптофановий оперон: атенюація

Та обставина, що транскрипція і трансляція у бактерій об'єднані у єдиний процес, використовується для регуляції активності генів шляхом *атенюації* (attenuation). Така система працює, зокрема, у триптофановому опероні (*trp*-операоні) *E. coli*. Оперон містить 5 структурних генів, продукти яких відповідають за синтез амінокислоти Трп, перед ними розташовані промотор, оператор і *лідерна послідовність* (рис. 6.3, а).



**Рис. 6.3.** Схема будови *trp*-оперона, на якому синтезуються два РНК-продукти залежно від внутрішньоклітинної концентрації *Trp* (а), і схема організації лідерної послідовності та два варіанти спарювання основ у її складі залежно від розташування рибосоми (б)

Лідруюча частина РНК містить стартовий кодон, що упізнається рибосомою, і чотири елементи послідовності: ділянка 1 містить два сусідні триптофанові кодони, ділянки 2–3 і 3–4 є попарно комплементарними, за ділянкою 4 розташована оліго-U-послідовність (рис. 6.3). Отже, шпилька 3–4, фланкована оліго-U, є сигналом термінації транскрипції (див. рис. 5.11). Коли концентрація *Trp* низька (існує потреба в цій амінокислоті й оперон має бути активним), рибосома зупиняється на триптофанових кодонах ділянки 1 (просто тому, що *Trp* у клітині немає). У цьому випадку утворюється шпилька 2–3 (ділянка 3 не залучається до утворення термінувальної шпильки) і РНК полімераза продовжує синтез повноцінної мРНК. Рибосоми зв'язуються зі стартовими кодонами, що відповідають структурним генам, і синтезуються відповідні білки.

За високого рівня *Trp* рибосома швидко проходить через ділянку 1 на ділянку 2 і зупиняється на стоп-кодоні. У результаті утворюється шпилька 3–4 – формується сигнал термінації,

РНК-полімераза зупиняє транскрипцію після синтезу короткої нефункціональної лідерної РНК.

*trp*-Оперон перебуває також під контролем *trp*-репресора. Алостеричним регулятором репресора є сам Trp, у комплексі з ним репресор набуває конформаційної форми, що має високу спорідненість до оператора. При зниженні концентрації Trp репресор дисоціює і ефективність ініціації транскрипції підвищується в ~70 разів. Атенуація є додатковим, менш ефективним механізмом регуляції: за її рахунок ефективність транскрипції підвищується у ~10 разів за відсутності Trp (за наявності Trp ~10 % РНК-полімераз долають сигнал термінації і продовжують працювати, а за його відсутності – практично всі). Сумісна дія атенуації і негативного контролю за рахунок репресора дозволяє змінювати активність оперона в ~700 разів залежно від внутрішньоклітинної концентрації Trp.

Приклад *trp*-оперона демонструє можливість, залежно від розташування рибосоми, експонувати або приховувати сигнал термінації. Такий самий принцип використовується в системах регуляції за механізмом *антитермінації*. У цьому випадку все залежить від наявності / відсутності у клітині спеціального білкового фактора транскрипції: саме він запобігає упізнанню РНК-полімеразою сигналів термінації, що знаходяться всередині кодувальної частини деяких бактеріальних генів. Такі білки-антитермінатори взаємодіють або безпосередньо з РНК-полімеразою, або зі спеціальною ділянкою послідовності поряд із сигналом термінації і одночасно з РНК-полімеразою. В обох випадках ця взаємодія допомагає полімеразі ефективно долати термінаційний сигнал.

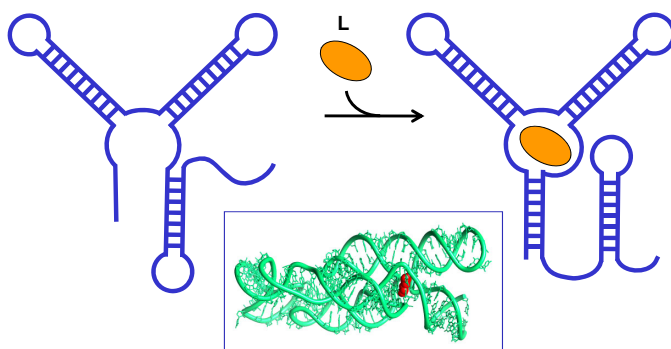
Крім того, приклад *trp*-оперона демонструє використання для регуляції транскрипції структурних перебудов мРНК, які у *trp*-опероні залежать від розташування рибосоми, але в інших випадках можуть індукуватись невеликими лігандами.

### **РНК-перемикачі**

У складі молекул РНК доволі часто виникають дволанцюгові шпильки, які, взаємодіючи між собою, продукують складні просторові структури. Подібно до білків (див. рис. 2.19) такі струк-

тури можуть існувати у різних формах залежно від взаємодії з тим чи іншим лігандом. Часто структурна форма за відсутності ліганду є конформаційно рухливою, тоді як зв'язування ліганду фіксує певну жорстку конформацію. Такі молекули або частини молекул РНК, що здатні "відповідати" на наявність ліганду, називають *РНК-перемикачами* (*рибоперемикачами*, *riboswitches*). Молекули РНК зі складною структурою, здатні з високою специфічністю зв'язувати різноманітні ліганди, називають також *аптамерами* (*aptamers*). Синтетичні аптамери широко використовують у сучасних біотехнологіях.

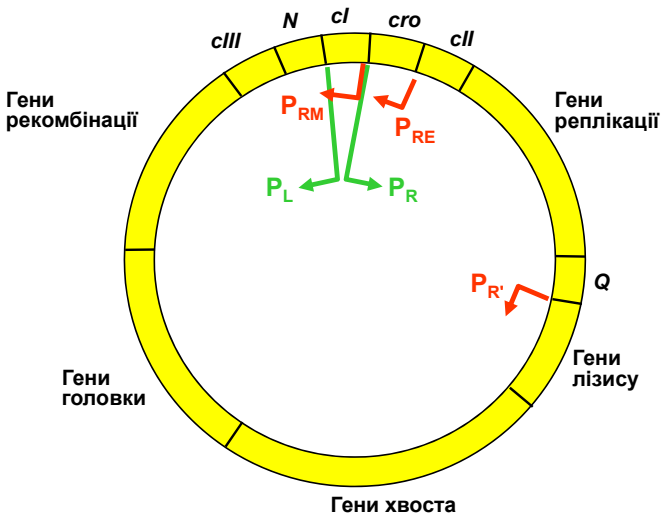
РНК-перемикачі часто беруть участь у регуляції експресії генів, зокрема для експонування чи маскуванню термінаційних сигналів у бактерій (рис. 6.4). Функцію ліганду зазвичай виконує той чи інший метаболіт, що є кінцевим продуктом активності гена / оперона. За відсутності ліганда термінаційний сигнал маскується у структурі перемикача і, відповідно, не сприймається РНК-полімеразою, що продовжує синтезувати функціональний транскрипт. У присутності ліганду (коли виникає потреба вимкнути ген) реалізується інша структурна форма перемикача, яка сприяє формуванню термінаційного сигналу. РНК-полімераза зупиняється на ньому, створивши тільки короткий нефункціональний транскрипт.



**Рис. 6.4.** Схема залежного від ліганду L структурного перетворення РНК-перемикача. На вставці: структура одного з таких перемикачів у комплексі з лігандом (червоний), роль якого у цьому випадку відіграє лізин (3DIL)

## Регуляція транскрипції бактеріофага $\lambda$

Система регуляції транскрипції генів бактеріофага  $\lambda$  може служити відносно простим прикладом регуляції у масштабі нехай і маленького, але цілого геному. Усередині фагових частинок міститься лінійна молекула ДНК завдовжки  $\sim 50$  тис. пар основ, на кінцях якої розміщені взаємно комплементарні одноланцюгові вирости – *cos*-сайти. Після проникнення в бактеріальну клітину ця лінійна ДНК замикається в кільце бактеріальною лігазою (ligase – фермент, що каталізує утворення фосфодієфірного зв'язку, див. розділ 9). Геном фага містить гени, що відповідають за синтез білків головки й хвоста фагової частинки, реплікацію фагової ДНК, лізис бактерії, рекомбінацію (вбудовування фагової ДНК у бактеріальний геном), і кілька регуляторних генів, що кодують фактори транскрипції (рис. 6.5).

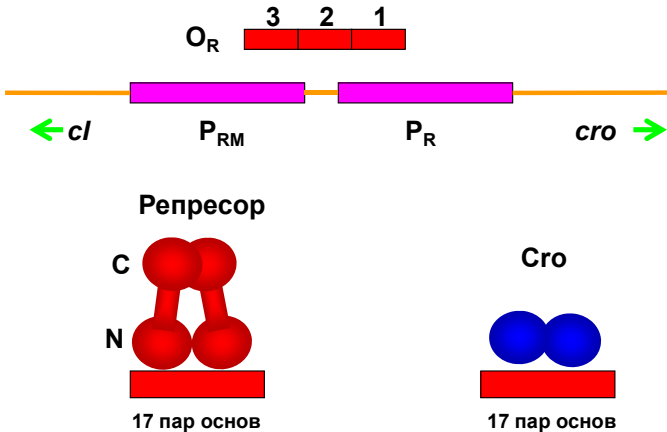


**Рис. 6.5.** Схема розташування генів у циркулярній ДНК бактеріофага  $\lambda$ . Стрілочками позначено промотори й напрямок транскрипції: зелені – сильні промотори, червоні – такі, що потребують активації

Існують два альтернативні шляхи розвитку фага: *лізогенія* – лінеаризація фагової ДНК, вбудовування її в бактеріальний

геном і блокування більшості фагових генів; *лізис* (одразу після інфікування або шляхом вирізання фагової ДНК із бактеріального геному та її циркуляризації після лізогенії) – активація реплікації фагової ДНК і синтезу білків оболонки, збирання фагових частинок і руйнування клітини.

Одразу після інфікування бактеріальна РНК-полімераза зв'язується із двома сильними промоторами  $P_R$  і  $P_L$  (рис. 6.5), з яких відбувається транскрипція у протилежних напрямках на генах *cro* і *N*, відповідно. Обидва гени закінчуються термінаторами, але щойно з'являється білок *N*, він зв'язується із транскриптами, рекрутує кілька бактеріальних білків, і цей комплекс запобігає впізнанню термінаторів за механізмом антитермінації. РНК-полімераза в цьому випадку продовжує синтезувати поліцистронну мРНК на *cIII* і генах рекомбінації (із промотору  $P_L$ ) і *cII*, генах реплікації та *Q* (із промотору  $P_R$ ). Після цього можливе розгалуження на два альтернативні шляхи, ключову роль у виборі одного з яких виконують два фактори транскрипції: білок *Cro* і  $\lambda$ -репресор – продукти відповідно генів *cro* і *cI*, що містяться поруч у фаговому геномі (рис. 6.5, 6.6).



**Рис. 6.6.** Зона контакту між генами *cI* і *cro* фага  $\lambda$ : промотори генів частково перекривають три оператори  $O_R$  1, 2, 3, кожен з яких (завдовжки 17 пар основ) має спорідненість до репресора (продукту гена *cI*) і білка *Cro* (продукту гена *cro*)

Розглянемо спочатку процес лізогенії (рис. 6.7). Продукт гена *cII* – активатор транскрипції, який забезпечує зв'язування РНК-полімерази зі слабким промотором  $P_{RE}$ . Із цього промотору (у напрямку, протилежному до напрямку транскрипції *cro*, рис. 6.5) транскрибується ген *cI*, унаслідок чого з'являється  $\lambda$ -репресор. Крім того, білок *cII* активує ген інтегрази (міститься серед генів рекомбінації) – ферменту, що забезпечує вбудовування фагової ДНК у геном клітини-хазяїна (див. розділ 10). Продукт гена *cIII* захищає білок *cII* від бактеріальних протеаз, тобто підвищує час життя активатора. Отже, за умови високої концентрації *cII* виникає певна кількість  $\lambda$ -репресорів, а фагова ДНК вбудовується інтегразою в бактеріальний геном.

Репресор (гомодимер, N-кінцеві частини обох субодиниць взаємодіють із ДНК, див. рис. 3.12, 3.13, *a*) має спорідненість до двох наборів операторів. Один із цих наборів – оператори  $O_R 1, 2, 3$ , кожен завдовжки 17 пар основ, – частково перекриває промотори  $P_R$  і  $P_{RM}$  (рис. 6.6). Промотор  $P_{RM}$  є слабким і може використовуватись для транскрипції гена *cI* тільки за умови активації. Послідовності операторів схожі між собою, але розрізняються за спорідненістю до репресора (найвищу спорідненість має ділянка  $O_R 1$ ). При заповненні цього оператора репресором промотор  $P_R$  гена *cro* й усіх генів праворуч від нього заблокований. С-кінцеві частини двох димерів репресора здатні до взаємодії: за наявності репресора в сайті  $O_R 1$  швидко заповнюється також сайт  $O_R 2$ . Інакше кажучи, репресор *кооперативно* взаємодіє з ділянками  $O_R 1, 2$ : його зв'язування з ДНК підсилюється білок-білковими взаємодіями димерів. При заповненні оператора  $O_R 2$  репресор (який є негативним регулятором гена *cro*) спрацьовує як позитивний регулятор свого власного гена *cI*, тобто індукує зв'язування РНК-полімерази із промотором  $P_{RM}$ . Спорідненість репресора до оператора  $O_R 3$  є найнижчою: ця ділянка некооперативно заповнюється при зростанні концентрації репресора, унаслідок чого відбувається блокування гена *cI*. У такий спосіб активність гена *cI* і концентрація репресора підтримуються на оптимальному рівні за принципом зворотного зв'язку. На стадії лізогенії тільки ген *cI* є активним –  $\lambda$ -репресор

блокує також промотор  $P_L$  (через відповідні оператори  $O_L$ ), тобто транскрипцію всіх генів ліворуч від  $N$ .

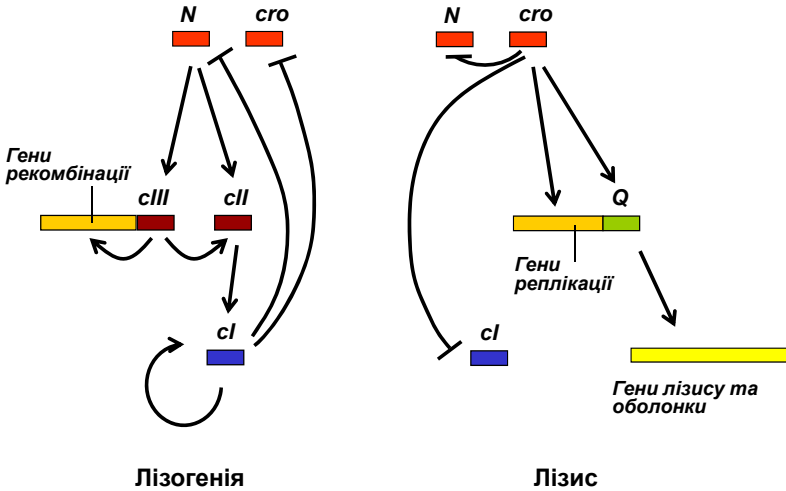


Рис. 6.7. Мережа регуляторних взаємодій між генами фага  $\lambda$

Якщо бактерія піддається дії мутагенів, коли виникає ризик загибелі клітини (разом із вбудованою фаговою ДНК), то відбувається активація певної бактеріальної протеази, яка розрізає молекулу репресора між її N- та С-кінцевими доменами. Результатом стають порушення димеризації мономерів репресора і втрата спорідненості до ДНК. Унаслідок звільнення операторів  $O_R$  та  $O_L$  РНК-полімераза зв'язується із промоторами  $P_R$  та  $P_L$  і починає транскрибувати гени *cro* та *N*, тобто знову виникає ситуація, що спостерігається одразу після інфікування, – спрацьовують гени реплікації, *cII* і *Q*. Однак при цьому білок cII швидко деградується протеазами, не встигаючи активувати синтез репресора. Така сама ситуація (висока активність протеаз) може реалізуватися одразу після інфікування за умови збагачення середовища на харчові ресурси.

Білок Cro (також гомодимер) некооперативно взаємодіє з операторами  $O_R$  1, 2, 3 (рис. 6.6), спорідненість до яких знижується в порядку 3 – 2 – 1. Заповнення  $O_R$  3 білком Cro остаточно вимикає синтез репресора із промотору  $P_{RM}$ . Подальше заповнення опера-

торів при зростанні концентрації *cro* вимикає транскрипцію *cro* і генів праворуч від нього, так само як і генів ліворуч від *N*, за рахунок зв'язування *cro* з операторами *O<sub>L</sub>*. Однак головний наслідок активності *cro* полягає в появі білка *Q* (рис. 6.5, 6.7).

Ще один промотор  $P_{R'}$  є насправді сильним, але одразу за ним розташований термінатор – РНК-полімераза за відсутності *Q* синтезує із цього промотору короткий нефункціональний транскрипт. Білок *Q* спрацьовує як антитермінатор, забезпечуючи додання цього бар'єра. У результаті відбуваються синтез поліцистронної мРНК на генах лізису та білків оболонки (праворуч від *Q*) і лізис клітини з виходом фагових частинок.

Розглянута доволі складна система регуляції транскрипції порівняно із простим геномом бактеріофага  $\lambda$  дає уявлення про те, наскільки ускладнюється загальна система регуляції транскрипції у бактерій, а тим більше в еукаріотів. Хоча складність загальних систем внутрішньоклітинної регуляції, які залежать від тонкого балансу великої кількості різноманітних впливів, не дозволяє стверджувати про остаточне з'ясування усіх деталей, ми сьогодні непогано розуміємо базові принципи регуляції в еукаріотичній системі транскрипції.

## Еукаріотичний транскриптом

Крім розділення процесів транскрипції і трансляції у часі та просторі, особливістю еукаріотичної системи експресії генів є упаковка ДНК у хроматині, що, за умовчанням, заважає завантаженню РНК-полімерази. Відповідно існують механізми перебудови структури хроматину з метою або активувати транскрипцію на певних ділянках, або, навпаки, гарантовано вимкнути її.

Не менш важливою особливістю еукаріотів є відсутність оперонів і суттєве зростання, порівняно із прокаріотами, кількості генів. Кілька десятків тисяч еукаріотичних генів (зрозуміло, що йдеться переважно про гени, які транскрибуються РНК-полімеразою II) потребують диференційної активації/репресії в певні моменти залежно від типу клітин, стадії розвитку, зовнішніх умов тощо.

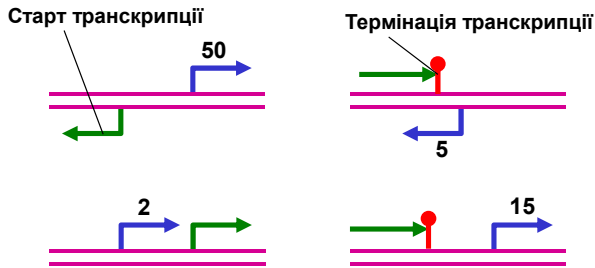
Безпосереднім загальним результатом диференційної активності еукаріотичних генів є сумарна сукупність усіх транскрип-

тів клітини – *транскриптом*, аналіз якого не тільки дає відповідь на запитання, що у клітині транскрибується, а й дозволяє виявити певні закономірності загальної системи регуляції транскрипційної активності геному.

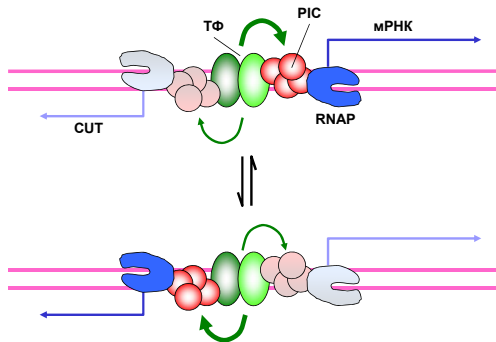
У доволі компактному геномі найпростіших еукаріотів – дріжджів – по обох полінуклеотидних ланцюгах транскрибується ~85 % ДНК (у фазі експоненційного росту на багатому середовищі). При цьому значна частина транскриптому складається з різноманітних РНК, що не піддаються трансляції. За винятком молекул РНК з відомими функціями, ці транскрипти, що не транслюються, можна розділити на дві групи: 1) стабільні неанотовані транскрипти (SUT, Stable Unannotated Transcripts) – стабільні молекули РНК із поки що невідомими функціями; 2) нестабільні "загадкові" транскрипти (CUT, Cryptic Unstable Transcripts) – молекули РНК завдовжки в середньому близько 260 нуклеотидів, які швидко деградуються нуклеазами у спеціальних комплексах – екзосомах. Іноді такі нестабільні транскрипти збігаються за напрямком транскрипції з відомими мРНК (є *змістовними* щодо цих мРНК), але частіше вони є *антизмістовними* щодо відомих генів і транскрибуються у протилежному напрямку (рис. 6.8).

Загалом, початок транскрипції ~70 % всіх транскриптів дріжджів збігається або із промотором іншого транскрипта (частіше при цьому інформація зчитується з тієї самої стартової ділянки, але в іншому напрямку та з іншого ланцюга ДНК), або із зоною в ділянці термінації транскрипції іншого гена (рис. 6.8) – у цьому випадку обидва напрямки транскрипції зазвичай збігаються. Отже, переважна більшість промоторів дріжджів є *двоспрямованими* – одна зона (зазвичай збіднена на нуклеосоми, див. нижче) використовується як промотор для транскрипції в обох напрямках по обох ланцюгах (рис. 6.9). При цьому в деяких випадках таке узагальнення промоторної зони двома генами (або геном і нестабільним транскриптом) може забезпечувати координовану регуляцію – дві частини промотору взаємно підсилюють ініціацію транскрипції. У деяких двоспрямованих промоторах, навпаки (як на схемі рис. 6.9), відбувається конкуренція між двома напрямками за базальні фактори транскрипції та РНК-полімерази. Рівновага може бути зсунута в той чи інший бік залежно

від контексту послідовності. Наприклад, наявність ТАТА-боксу чи інших стандартних елементів послідовності в одному із двох базальних промоторів сприятиме зв'язуванню РНК-полімерази у напрямку транскрипції мРНК, а не нестабільного транскрипту, якщо присутні відповідні активуючі впливи. Якщо ж таких впливів немає, то рівновага зсувається в бік синтезу нестабільного нефункціонального транскрипту.



**Рис. 6.8.** Розподіл стартових точок для транскриптів дріжджів (синя стрілка), асоційованих з іншими транскриптами (зелена стрілка). Цифри – приблизна частка у відсотках транскриптів зазначеного типу



**Рис. 6.9.** Схема організації двоспрямованого промотору з конкуренцією між мРНК і нестабільним транскриптом (CUT).  
ТФ – специфічні транскрипційні фактори,  
PIC – преініціаторний комплекс

Аналізу транскриптому присвячено, зокрема, міжнародний проєкт ENCODE (*Encyclopedia of DNA Elements*), зосереджений передусім на дослідженні транскриптому людини. Не враховуючи генів РНК, на ~20 тис. білкових генів (розділ 4) припадає 39,5 % геному (від стартової точки до зони термінації транскрипції, включаючи інтрони). У рамках проєкту було досліджено 15 ліній зляжисних клітин (напевно, у нормальних клітинах людини наведені нижче цифри дещо інакші). Виявилось, що частка геному, яка загалом транскрибується у всіх клітинних лініях, становить 74 %, у середньому в кожній лінії 39 % (не більше 57 %). Для ініціації усіх цих транскриптів використовується ~113 тис. промоторів. Напевно, частина великої кількості первинних транскриптів є просто наслідком неспецифічної хаотичної активності РНК-полімераз. Проте значна частка транскриптів містить консервативні для ссавців елементи послідовності. Імовірно, у багатьох випадках ці транскрипти є невідомими раніше РНК, які не піддаються трансляції (для більшості з них їхнє функціональне значення ще має бути з'ясовано).

Часто піддаються транскрипції енхансерні ділянки. Можливо, іноді така транскрипція є просто одним із засобів підтримувати регуляторну ділянку в доступному стані деконденсованого хроматину. Важливу роль дистальних елементів у системі регуляції транскрипції в еукаріотів, а також складність взаємовідносин між дистальними елементами й базальними промоторами, демонструють такі цифри: у середньому одна стартова точка транскрипції перебуває під контролем 3,9 дистальних елементів, тоді як один дистальний елемент контролює у середньому 2,5 стартових точок.

Для білкових генів була продемонстрована наявність великої кількості невідомих раніше стартових точок транскрипції, зокрема й таких, що віддалені на 100 тис. пар основ від анованих раніше стартів транскрипції цих генів. Деякі зі стартових точок використовують промотори інших генів: одна стартова точка може бути спільною для двох чи трьох генів, а первинний транскрипт іноді містить кілька генних локусів (груп екзонів). Відповідно спостерігається високий рівень альтернативного сплайсингу, зокрема й на рівні різних генів (розділ 7). Резуль-

тати проєкту ENCODE вказують також на диспергований характер розподілу регуляторних елементів по всьому геному: багато регуляторних елементів розташовані всередині екзонів і інтронів, при цьому вони можуть бути елементами системи регуляції зовсім іншого гена.

У геномі людини виявлено біля 3 млн регуляторних *цис*-елементів, що упізнаються специфічними факторами транскрипції. Серед них багато однакових або подібних за нуклеотидною послідовністю. Загалом типів регуляторних елементів (спільних мотивів послідовності) налічується дещо менше 700, тобто одно-типні *цис*-елементи зустрічаються у багатьох регуляторних зонах.

## Транскрипційні фактори й регуляторні РНК в еукаріотів

### Транскрипційні фактори

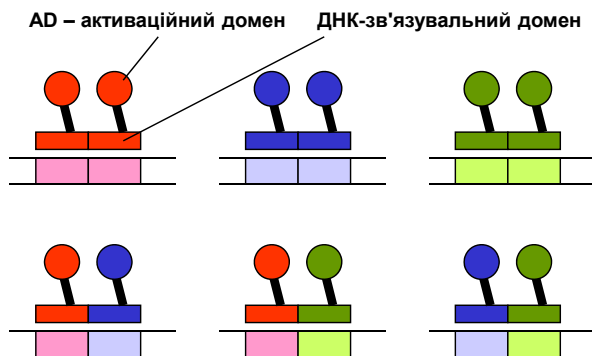
Як і у прокаріотів, ключовими елементами еукаріотичної системи регуляції транскрипції є білкові транскрипційні фактори, що упізнають *цис*-елементи у проксимальних і дистальних регуляторних зонах. Більшість факторів транскрипції мають у структурі принаймні два домени: той, що зв'язується з ДНК, і *активаційний* (AD, Activation Domain), який використовується для взаємодії з іншими білками. Переважна більшість активаційних доменів належать до категорії невпорядкованих (розділ 2), що передбачає широкі можливості взаємодій із різноманітними партнерами.

Велика кількість еукаріотичних генів і той факт, що вони не об'єднуються в оперони (коли кілька генів перебували б під контролем одного транскрипційного фактора), роблять неможливим однозначне співвідношення між генами і факторами транскрипції: кожен ген не може контролюватися власним фактором транскрипції, оскільки фактор також є продуктом певного гена, який також має контролюватися певним фактором, і т. д.

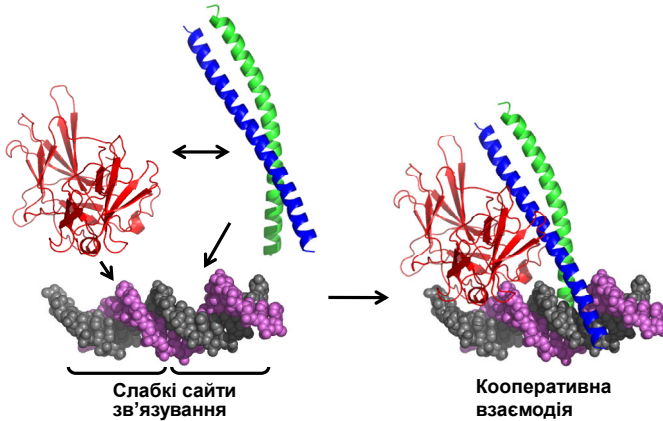
Відповіддю на необхідність регулювати активність великої кількості генів великим, але лімітованим набором факторів транскрипції, є *принцип модульності* будови регуляторних зон еукаріо-

тичних промоторів. Цей принцип ілюструє схема на рис. 6.10: три регуляторні елементи послідовності мають спорідненість до трьох транскрипційних факторів; із трьох пар такої взаємодії можна скласти шість комбінацій. Для згаданих вище ~700 типів регуляторних елементів кількість можливих комбінацій по 5–10 в одній регуляторній зоні практично нескінченно велика. Отже, кожен промотор може мати власний набір модулів, який відрізняє його від інших промоторів, і, відповідно, власний набір транскрипційних факторів, потрібних для впливу на цей промотор.

Із принципом модульності тісно пов'язаний **принцип кооперативності** взаємодії транскрипційних факторів із *цис*-елементами, що містяться поряд. Кожен із транскрипційних факторів зазвичай має порівняно невисоку спорідненість до відповідних елементів послідовності. Однак якщо два *цис*-елементи розташовані поруч і два білкові фактори здатні взаємодіяти між собою, то спорідненість кожного з них до промотору підвищується (рис. 6.11): стабільність комплексу значно зростає (експоненційно, див. рівняння (1.6)) завдяки сумі невеликих знижень вільної енергії при взаємодії кожного окремого білка. Приклад на рис. 6.11 показує лише один елементарний акт кооперативної взаємодії кількох транскрипційних факторів із промотором.



**Рис. 6.10.** Модульний принцип взаємодії транскрипційних факторів із промоторами



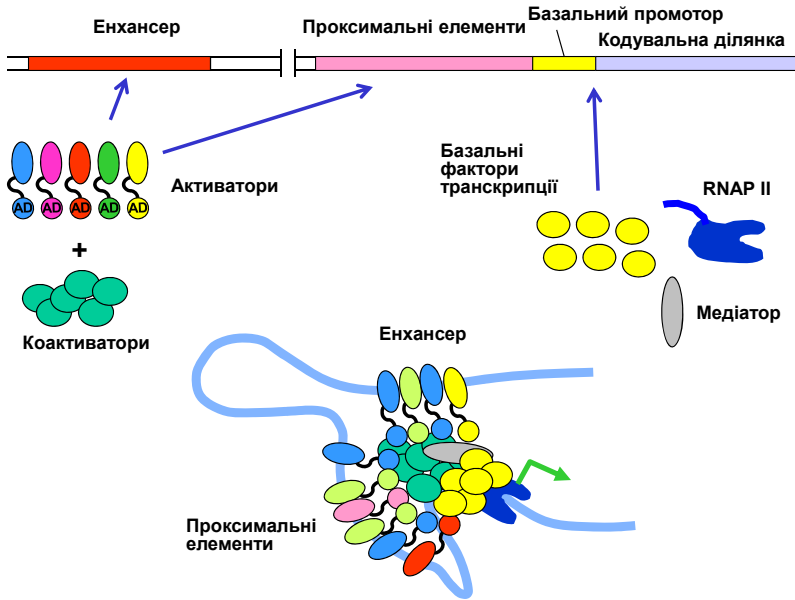
**Рис. 6.11.** Приклад кооперативної взаємодії двох транскрипційних факторів із ДНК (1A02)

Крім того, як зазначено вище, більшість факторів транскрипції мають активаційні домени, що зв'язують білкові кофактори (коактиватори). У результаті на дистальних і проксимальних регуляторних зонах формується складний мультибілковий комплекс – *енхансосома* (enhanceosome). До складу енхансосоми часто входять також ферменти, що вносять хімічні модифікації в гістони, білки, які упізнають такі модифікації (розділ 4), і фактори ремоделювання хроматину (див. нижче). Отже, транскрипційні фактори є також засобом рекрутування у регуляторні й промоторні зони машинерії, від якої залежать хімічні модифікації і перебудови структури хроматину.

Різні елементи енхансосоми діють синергічно, підвищуючи загальну стабільність комплексу. З іншого боку, відсутність кількох елементів може зумовити дестабілізацію та розпад енхансосоми, де спорідненість до ДНК кожного окремого елемента порівняно невисока. Це забезпечує *динамізм* активації: енхансосома не є фіксованою, а збирається / розбирається в певні моменти.

Активаційні домени транскрипційних факторів і коактиватори мають, своєю чергою, спорідненість до медіатора й базальних факторів транскрипції (розділ 5). Результатом такої взаємодії є ефективне збирання преініціаторного комплексу на базальному промоторі (рис. 6.12). Слід зауважити, що компоненти

мультибілкових комплексів, які збираються на промоторах, можуть, навпаки, блокувати ініціацію транскрипції – тоді їх називають репресорами та корепресорами.



**Рис. 6.12.** Схема збирання комплексу активації транскрипції

Гени транскрипційних факторів самі є мішенями регуляції з боку інших транскрипційних факторів. У клітині працює складна регуляторна мережа, організована за ієрархічним принципом: фактори вищого рівня, що контролюють найбільшу кількість генів, впливають на експресію генів інших факторів, під контролем яких перебуває набір факторів нижчого рівня тощо – до різноманітних генів-мішеней. Наприклад, частина транскрипційних факторів дрозофіли (76 білків) здійснюють 158 588 регуляторних впливів на гени-мішені. Загальна кількість мішеней цих факторів становить 12 482, тобто один ген-мішень регулюється в середньому ~12 факторами транскрипції (максимально до 54), що відображає великі можливості комбінаторного контролю.

До цієї регуляторної мережі разом із генами білкових транскрипційних факторів залучені також гени, продуктами яких є мікроРНК.

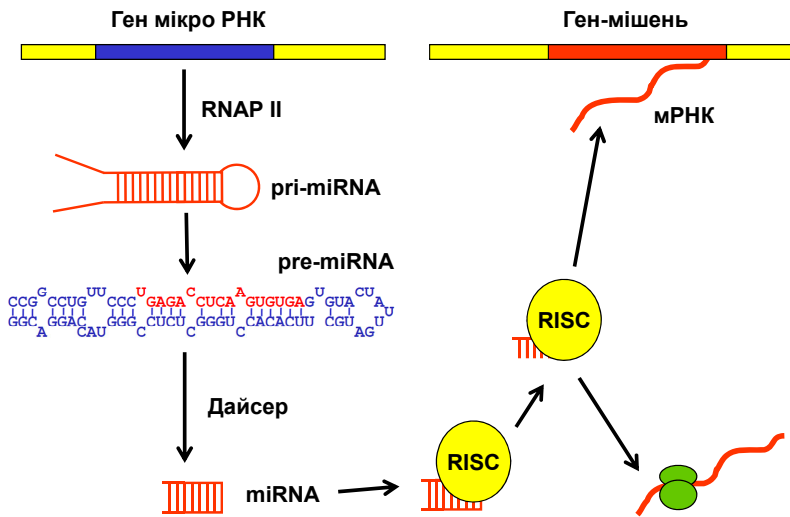
## Регуляторні РНК

Велика кількість транскриптів, що продукуються РНК-полімеразою II, є різноманітними молекулами РНК, що не піддаються трансляції і виконують регуляторну функцію. Серед них зупинимось на двох найкраще вивчених категоріях – мікроРНК (miRNA) і довгих некодувальних (регуляторних) РНК (large non-coding RNA, lncRNA, або large intergenic non-coding RNA, lincRNA).

**МікроРНК** – один із найважливіших типів регуляторних РНК. Механізм регуляторної дії мікроРНК (miRNAs) базується на **РНК-інтерференції**. Цей процес (RNA interference, RNAi) може запускатись будь-якою дволанцюговою РНК (у т. ч. штучно введеною до клітини). Дволанцюгова РНК стає субстратом для спеціалізованої РНКазиди, яка отримала назву *дайсер* (Dicer). Дайсер розрізає молекулу на дволанцюгові фрагменти завдовжки близько 21 пара основ, які позначають siRNA (short interfering RNA). Вони зв'язуються з кількома білками, утворюючи комплекс RISC (**R**NA **I**nduced **S**ilencing **C**omplex). Найважливішим компонентом цього комплексу є білок, що належить до родини *аргонавт* (Argonaute), один із доменів білка має нуклеазну активність. Один із ланцюгів siRNA руйнується, а той, що залишився, здійснює комплементарне спарювання з ділянкою мРНК, що синтезується у процесі транскрипції (вихідна дволанцюгова РНК має збігатися за послідовністю з кодувальною частиною зазначеного гена), спрямовуючи туди RISC. За рахунок нуклеазної активності аргонавта здійснюється деградація транскрипту. Отже, хоча транскрипція відбувається, реалізується посттранскрипційне вимкнення гена (PTGS – posttranscriptional gene silencing). При штучному введенні у клітину дволанцюгової РНК у такий спосіб можна на деякий час припинити експресію бажаного гена – здійснити *генний нокаун*. Такий підхід широко використовують як дослідницький засіб, а також як інструмент генної терапії.

У клітині РНК-інтерференція запускається молекулами мікроРНК. Ці молекули є продуктами численних генів мікроРНК (транскрибуються РНК-полімеразою II, частина мікроРНК походить також з інтронів деяких білкових генів). Безпосередній продукт транскрипції (pri-miRNA) піддається частковій деградації у ядрі й виходить у цитоплазму у вигляді пре-мікроРНК (pre-miRNA) – молекули РНК, яка практично є дволанцюговою шпилькою (рис. 6.13).

Із цієї шпильки дайсер вирізає дволанцюговий фрагмент завдовжки ~24 пари основ, який є аналогом siRNA. Ця РНК утворює комплекс RISC, у складі якого нарешті залишається одноланцюговий фрагмент – мікроРНК, що спрямовує комплекс до мРНК-мішені (рис. 6.13). Результатом взаємодії з мРНК може бути її посттранскрипційна деградація або блокування білкового синтезу (у цьому разі взаємодія з мРНК відбувається в цитоплазмі). Крім того (і це ще одна важлива функція мікроРНК), зв'язування комплексу RISC із мРНК під час транскрипції викликає рекрутування до хроматину ферментів, що здійснюють певні модифікації гістонів (більш детально цей аспект буде обговорено нижче у зв'язку з механізмами забезпечення гарантованої репресії).

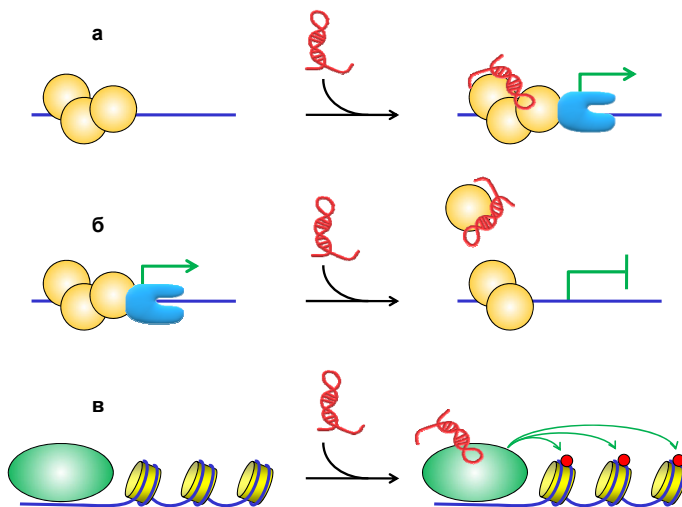


**Рис. 6.13.** Участь мікроРНК у регуляції експресії генів.

У складі наведеного прикладу пре-мікроРНК червоним кольором позначено одноланцюгову мікроРНК, що залишається у комплексі RISC

З одного боку, гени мікроРНК контролюються транскрипційними факторами, а з іншого – гени транскрипційних факторів (як і інші гени-мішені) самі часто перебувають під контролем мікроРНК. Обидві обставини створюють єдину регуляторну мережу, що формується генами обох типів.

**Довгі регуляторні РНК**, що містять більше 200 нуклеотидів і формують складні просторові структури, є продуктами генів, які містяться або в ділянках між білковими генами, або всередині інтронів білкових генів, або таких, що транскрибуються у зворотному напрямку щодо напрямку транскрипції якогось білкового гена (із двоспрямованого промотору, див. вище). Довгі регуляторні РНК залучені до регуляції багатьох процесів, зокрема процесингу мРНК і трансляції. Їхня участь у регуляції транскрипції може бути різноманітною (деякі приклади наведено на рис. 6.14). Регуляторні РНК можуть бути компонентами енхансосою, необхідними для ефективного запуску транскрипції (рис. 6.14, *a*), або, навпаки, взаємодіяти з факторами чи кофакторами транскрипції, що сприяє розбиранню енхансосою (рис. 6.14, *б*). Крім того, регуляторні РНК залучені до активації машинерії, від якої залежить внесення хімічних модифікацій до компонентів хроматину (рис. 6.14, *в*).



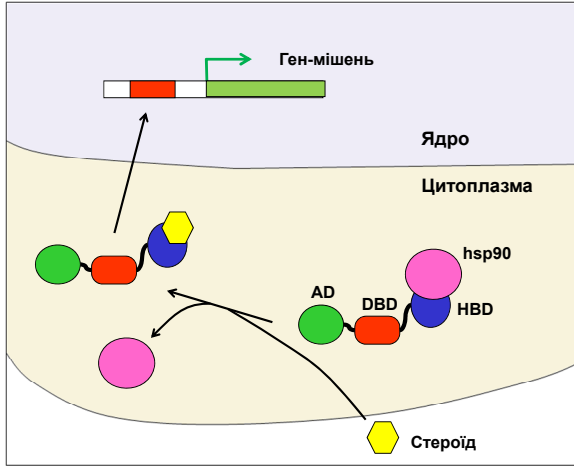
**Рис. 6.14.** Участь довгих регуляторних РНК у регуляції транскрипції: зв'язування РНК з енхансосою, що зумовлює активацію транскрипції (*a*); зв'язування РНК із транскрипційним фактором, що зумовлює інактивацію транскрипції (*б*); зв'язування РНК із ферментом, який вносить модифікації у гістони, що зумовлює активацію такого ферменту (*в*)

## **Зовнішня регуляція активності транскрипційних факторів**

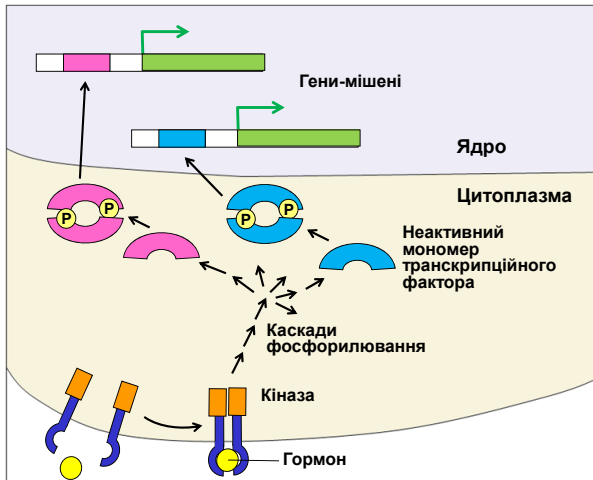
Білкові транскрипційні фактори й кофактори, а також регуляторні РНК, інкорпоровані в єдину складну систему регуляції транскрипції, яка забезпечує специфічний для певного типу клітин залежний від концентрації регуляторів профіль експресії геному – у клітині певного типу відбувається синтез специфічного набору транскрипційних факторів, що зумовлює активацію специфічного набору генів. У той самий час експресія деяких генів має оперативну контролюватися у відповідь на зовнішні сигнали шляхом зміни активності вже синтезованих транскрипційних факторів. Два найважливіші механізми такої регуляції – взаємодія транскрипційних факторів певного типу зі стероїдними гормонами й каскади посттрансляційних модифікацій у відповідь на дію хімічних сигналів (сигнальна трансдукція).

**Гормонові рецептори** – транскрипційні фактори, активність яких залежить від стероїдних гормонів. Гормоновий рецептор – зазвичай гомодимер, кожна субодиноця якого складається із трьох структурних доменів: 1) ДНК-зв'язувальний домен, що має специфічну спорідненість до певних елементів послідовності ДНК (див. рис. 3.14, б); 2) активаційний домен; 3) гормонзв'язувальний домен. За відсутності гормону білок міститься в цитоплазмі, де гормонзв'язувальний домен взаємодіє з білком теплового шоку hsp90 (детальніше білки теплового шоку буде розглянуто в розділі 8), який підтримує недоструктурований стан домену. У результаті гормоновий рецептор є інактивованим (рис. 6.15).

Коли гормон проникає в цитоплазму, він взаємодіє з гормонзв'язувальним доменом, витісняючи hsp90; відбувається остаточне структурування гормонного рецептора, і той стає активним – прямує до ядра, де зв'язується зі специфічним елементом послідовності. Взаємодія гормонного рецептора з ДНК часто запускає каскад збирання енхансосоми: рецептор рекрутує інші транскрипційні фактори за рахунок свого активаційного домену.



**Рис. 6.15.** Активация гормонального рецептора стероидным гормоном. AD – активационный, DBD – ДНК-связывающий, HBD – гормонсвязывающий домены



**Рис. 6.16.** Активация транскрипционных факторов у відповідь на дію білкового гормону

**Сигнальна трансдукція.** Загальну схему одного з типових варіантів сигнальної трансдукції наведено на рис. 6.16. Білковий гормон не проникає у клітину, а зв'язується рецептором на зовнішньому боці плазматичної мембрани. Рецептор є гомодимером, і за відсутності гормону дві субодиниці вільно дифундують у площині мембрани. Зв'язування гормону викликає димеризацію, унаслідок чого в цитоплазматичній частині рецептора виникає кіназна активність. Примембранна кіназа запускає каскад фосфорилування – фосфорилує білок, який набуває внаслідок цього кіназної активності. Ця нова кіназа фосфорилує інший білок (або кілька різних білків, завдяки чому здійснюється підсилення сигналу та/або його розгалуження по кількох шляхах, спрямованих до кількох кінцевих мішеней), перетворюючи його на кіназу, і так далі – до фосфорилування субодиниць транскрипційних факторів, що знаходяться в цитоплазмі. Фосфорильовані субодиниці об'єднуються в димер, унаслідок чого транскрипційний фактор стає активним: він проникає в ядро, де знаходить специфічну послідовність ДНК.

## **Структура хроматину й регуляція транскрипції**

Зміни у структурі хроматину є ще одним дуже важливим механізмом регуляції транскрипції в еукаріотів – таким, що інтегрує інші впливи в єдину систему. Достатньо сказати, що для регуляції транскрипції за допомогою транскрипційних факторів регуляторні елементи нуклеотидної послідовності мають бути доступними й для цих факторів, і для компонентів преініціаторного комплексу РНК-полімерази. Загалом нуклеосоми виступають як загальний репресор генної активності й тим самим допомагають забезпечити пригнічення активності більшості генів в еукаріотичній клітині, за винятком тих, чия активація потрібна. Отже, активація транскрипції потребує структурних перебудов у напрямку деконденсації хроматину й визволення *цис*-елементів від нуклеосом. Для реалізації таких перебудов є два основні інструменти, які діють у тісній координації один з одним: система посттрансляційних модифікацій гістонів і АТР-залежне перемі-

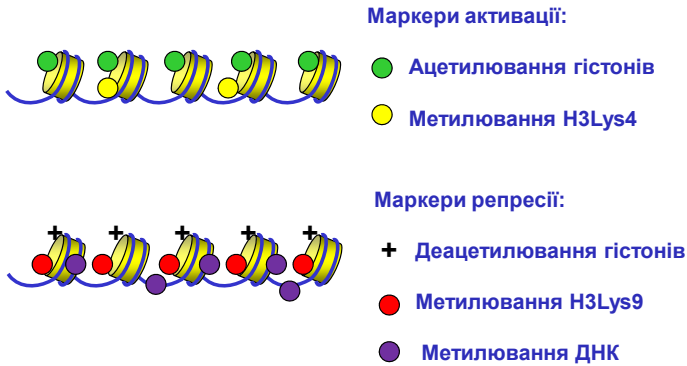
щення (репозиціонування) і тимчасове видалення нуклеосом. Хімічні модифікації компонентів хроматину відіграють також важливу роль у його структурній організації на вищих рівнях, що не менш важливо для регуляції транскрипції.

### **Хімічні модифікації хроматину**

Як описано у розділі 4, патерни хімічних модифікацій компонентів хроматину визначають розділення хроматинової фібрили на еу- і гетерохроматинові компартменти, що розрізняються ступенем компактності, а отже, і загальної доступності ДНК для різноманітних впливів. Зокрема, характерними маркерами репресованих зон є метильований Lys9 у гістоні H3 і метильований цитозин у ДНК (нижче гетерохроматинові маркери будуть розглянуті детальніше). Потенційно активні й активні зони, збагачені на метильований Lys4 в тому самому гістоні H3, комбінації певних метильованих і ацетильованих залишків маркують зони елонгації транскрипції (рис. 6.17): гістоновий код дає можливість тонкої регуляції та модуляції транскрипції за рахунок упізнання гістонових модифікацій численними білками – факторами й кофакторами транскрипції, ферментами, що здійснюють модифікації гістонів і ДНК тощо.

Серед інших модифікацій ацетилювання залишків Lys (у певних консервативних позиціях, див. рис. 4.12) завжди корелює з активацією транскрипції – ацетильовані гістон-ацетилтрансферазами гістони акумулюються в активних промоторах, і навпаки – дія гістон-деацетилаз зумовлює інактивацію. Гістон-ацетилтрансферази (НАТ) входять до складу мультибілкових комплексів, які часто є компонентами енхансосом. НАТ-комплекси можуть рекрутуватися до промоторів транскрипційними факторами чи кофакторами (напр. активаційним доменом гормонового рецептора). Зв'язування НАТ може індукуватися іншими гістоновими модифікаціями, зокрема через упізнання фосфорильованого Ser10 гістону H3. Крім того, часто до складу НАТ входять бромодомени – структурні модулі, які мають специфічну спорідненість до ацетильованих лізинів, тобто НАТ упізнають Lys, уже ацетильовані іншими НАТ, і здійснюють

ацетилювання сусідніх нуклеосом, підтримуючи у такий спосіб ацетилюваний статус певної ділянки хроматину.



**Рис. 6.17.** Характерні хімічні маркери, що визначають активний і репресований стани хроматину

Механізми головних впливів ацетилювання гістонів на структуру хроматину й активацію транскрипції можна згрупувати у три категорії.

- 1) Ацетилювання сприяє *деконденсації хроматину* за рахунок зниження позитивного заряду головних факторів конденсації, якими є гістонові хвости. У результаті знижується і спорідненість ацетилюваних нуклеосом до гістону H1, що далі сприяє деконденсації. Деконденсація еухроматинових компартментів створює "вікно можливості" для зв'язування регуляторних факторів із "відкритою" ДНК.
- 2) Хоча ацетилювання гістонів не змінює структуру нуклеосоми, зниження позитивного заряду гістонових хвостів спричиняє *дестабілізацію нуклеосоми* на виході з неї нуклеосомної ДНК за рахунок підвищення електростатичного розштовхування між сусідніми витками нуклеосомної суперспіралі (див. розділ 4). У результаті полегшується тимчасове руйнування нуклеосом за рахунок перенесення гістонів на проміжні акцептори й роботи комплексів ремоделювання хроматину (див. нижче).

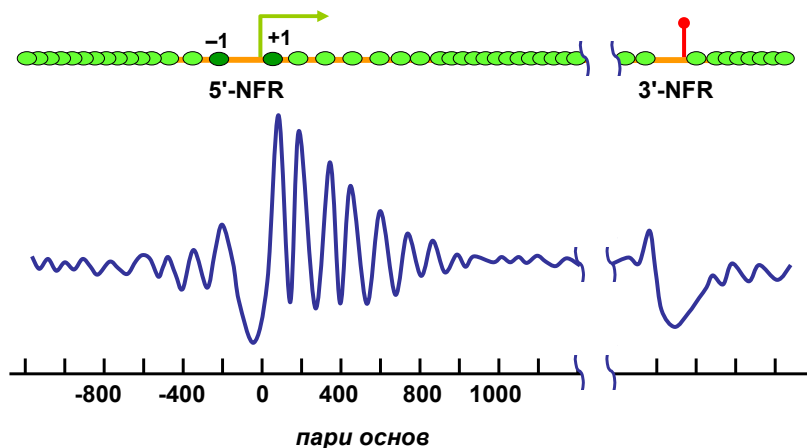
- 3) Ацетильовані лізинові залишки гістонів можуть безпосередньо упізнаватися факторами й кофакторами транскрипції. Наприклад, наявність бромодому у складі одного із ТВР-асоційованих факторів (розділ 5) сприяє підвищенню локальної концентрації TFIIID у ацетильованих ділянках хроматину.

### **Розташування нуклеосом у промоторах**

Залежне від послідовності ДНК переважне позиціонування нуклеосом (розділ 4) зумовлює диференційне експонування ділянок ДНК до дії регуляторних факторів. З одного боку, будь-яка послідовність пар основ диктує певний "передустановлений" розподіл нуклеосом, з іншого – активація промоторів потребує змін у їхньому розташуванні.

В еукаріотичних геномах існують загальні закономірності щодо розташування нуклеосом відносно стартових точок транскрипції (рис. 6.18). Вище від старту транскрипції на відстані  $\sim -220$  пар основ із високою імовірністю розташована нуклеосома, позначена як  $-1$  на рис. 6.18. Одразу за нею у позитивному напрямку часто розташована зона, вільна від нуклеосом (NFR – Nucleosome-Free Region), завдовжки  $\sim 150$  пар основ – прогалина, де могла б бути розміщена одна нуклеосома. За цією зоною, яка позначається як 5'-NFR, на стартовій точці або одразу за нею з високою імовірністю розміщена нуклеосома  $+1$ . Нуклеосоми  $+1$  належать до найсильніше позиційованих у геномі. Нуклеосома  $+1$  задає рамку фазованих нуклеосом у кодуючій частині гена: нуклеосоми, розташовані за сильно позиційованою, займають не переважні, а просто доступні позиції, що не перекриваються. У результаті на ділянці певної довжини спостерігається впорядковане розташування нуклеосом, коли на багатьох фрагментах ДНК, що аналізуються, у тій самій ділянці з високою імовірністю знаходиться нуклеосома. На віддаленні від сильно позиційованої нуклеосоми накопичуються випадкові зсуви нуклеосом у той чи інший бік, і регулярність їхнього розташування порушується – на багатьох фрагментах

ДНК нуклеосоми хаотично займають місця, що перекриваються. У ділянці термінації транскрипції є інша характерна зона, вільна від нуклеосом, – 3'-NFR (рис. 6.18).



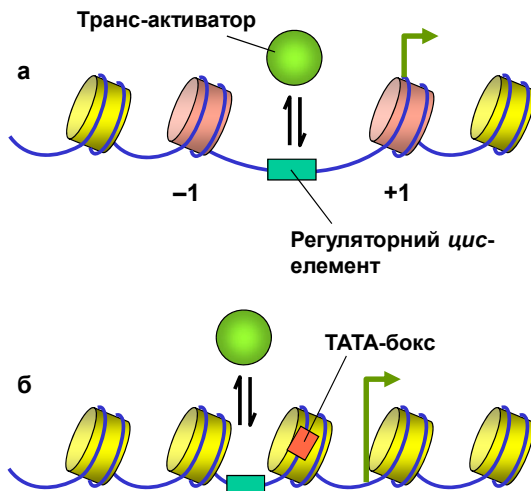
**Рис. 6.18.** Приблизний усереднений по всіх стартових точках транскрипції розподіл нуклеосом відносно старту (нульова точка). Крива відображає профіль частоти зустрічальності пар основ у складі нуклеосом. На верхній панелі нуклеосоми зображено як овали; зелена стрілка – старт транскрипції; червоне коло – сайт термінації транскрипції; NFR – зони, вільні від нуклеосом; -1 та +1 – дві нуклеосоми, що фланкують старт транскрипції

Зрозуміло, що 5'-NFR є зоною, доступною для збирання преініціаторного комплексу РНК-полімерази, а 3'-NFR – для збирання мультибілкового комплексу, від якого залежить термінація транскрипції (розділ 7). Цілком можливо, що саме особливості розташування нуклеосом (які зумовлені передусім нуклеотидною послідовністю) у зоні стартової точки є первинним механізмом визначення еукаріотичного промотору, а стандартні елементи базального промотору виникли у процесі еволюції пізніше як адаптація, що підсилює спорідненість до загальних факторів транскрипції.

Нуклеосоми +/-1 збагачені на особливий варіант гістону H2A – H2A.Z, який зумовлює зниження їхньої стабільності. Як

зазначалося вище, нуклеосоми в активованому промоторі зазвичай піддаються ацетилюванню. Крім того, дві позиційовані нуклеосоми по обидва боки від 5'-NFR часто містять гістони H3, у складі яких метильовано Lys4. Обидві модифікації упізнаються TBP-асоційованими факторами, що входять до складу преініціаторного комплексу, тобто сильно позиційована нуклеосома +1 і вільна від нуклеосом ділянка ДНК поряд із нею формують зручний сайт для початкових етапів збирання преініціаторного комплексу. Водночас після збирання преініціаторного комплексу нуклеосома +1 стає бар'єром для РНК-полімерази, яка має розпочати рух для продовження транскрипції. На початку елонгації виникає необхідність у дестабілізації або видаленні цієї нуклеосоми.

Значимо, що розподіл нуклеосом на рис. 6.18 є *усередненим*: далеко не всі промотори характеризуються наявністю зони, збідненої на нуклеосоми. Серед промоторів можна виділити дві характерні групи, між якими спостерігаються численні проміжні варіанти: *відкриті (конститутивні)* промотори, для яких яскраво виявляється загальна закономірність відсутності нуклеосоми перед стартовою точкою транскрипції (5'-NFR), і *закриті*, тобто такі, що не містять зони, вільної від нуклеосоми (рис. 6.19). Промотори першого типу, характерні для генів домашнього господарства, збіднені на стандартні елементи базального промотору (такі як ТАТА-бокс). Їхня вільна від нуклеосоми ділянка містить регуляторний *цис*-елемент, доступний для специфічного активатора транскрипції, зв'язування якого запускає збирання преініціаторного комплексу. Промотори другого типу збагачені на ТАТА-бокс та інші стандартні елементи, недоступні для транскрипційної машинерії, оскільки знаходяться в межах нуклеосом. Зрозуміло, що такі промотори потребують спеціальних механізмів для активації, які запускаються зазвичай зв'язуванням певного "піонерського" специфічного транс-активатора у лінкерній міжнуклеосомній ділянці.



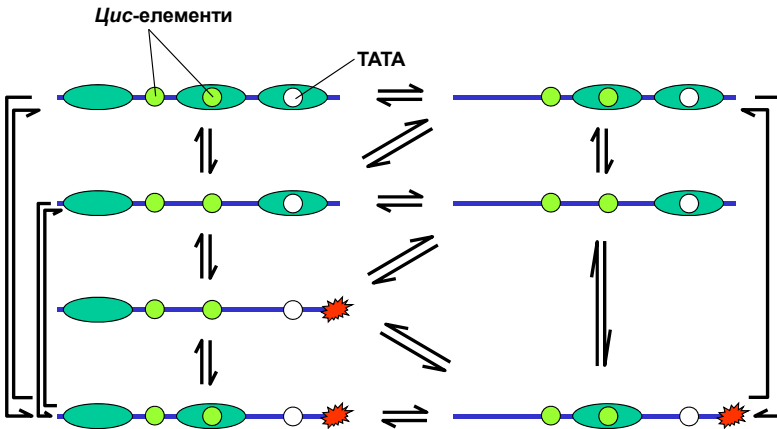
**Рис. 6.19.** Два типи промоторів: відкриті із позиційованими нуклеосомами  $-1$  та  $+1$  і зоною, вільною від нуклеосоми (а, див також рис. 6.17), і закриті (б)

### Ремоделювання промоторів

Ефективна ініціація транскрипції у закритих промоторах часто залежить від змін характеру розташування нуклеосом у промоторних зонах – *ремоделювання хроматину*.

Рисунок 6.20 ілюструє детально досліджений приклад: наявність нуклеосом у закритому промоторі дріжджового гена *RHO5*, вивчену Корнбергом зі співавторами (Roger D. Kornberg). У неактивному стані промотор містить 3 нуклеосоми у специфічних позиціях, у межах однієї знаходиться ТАТА-бокс, іншої – специфічна регуляторна послідовність. При активації промотору (цьому передуює гіперацетилювання гістонів) спостерігається втрата нуклеосом у всіх трьох сайтах, але ця втрата не є повною в жодному з них. У середньому втрачається 1,9 нуклеосоми із трьох; кожна з нуклеосом від 0,18 до 0,6 частини часу зберігається в активному промоторі. Це означає, що при активації підтримується певна рівновага між видаленням і реформуванням нуклеосом.

Тимчасове видалення нуклеосом з активних промоторів є загальним правилом. Тотальний аналіз розподілу нуклеосом по ділянках усього геному дріжджів та інших еукаріотів укажує, що існує зворотна кореляція між щільністю нуклеосом у промоторах і рівнем транскрипційної активності відповідних генів, причому за умови активації / репресії гена часто спостерігається перебудова нуклеосомної організації: деякі нуклеосоми зникають / виникають на певних позиціях або змінюють свої позиції.



**Рис. 6.20.** Динаміка нуклеосом (овали) у промоторі гена *PHO5* дріжджів *S. cerevisiae* після його активації: "визволення" ТАТА-боксу запускає транскрипцію

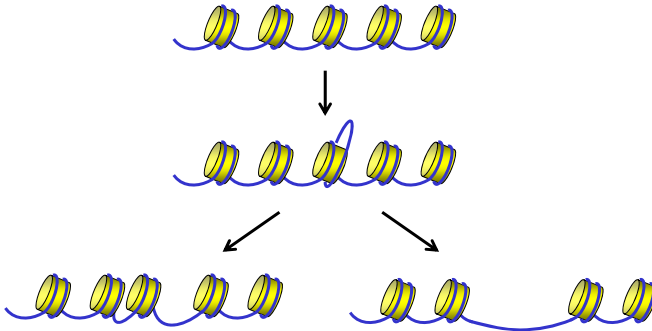
Підвищення доступності промоторів потребує спеціальних механізмів. Адже за фізіологічної іонної сили електростатичні взаємодії ДНК і гістонів дуже міцні й нуклеосома зберігає високу стабільність. Ця стабільність практично виключає навіть переміщення нуклеосоми вздовж ДНК: таке переміщення потребує появи високоенергетичних інтермедіатів із частковим порушенням взаємодії гістонів із ДНК. Оскільки переміщення нуклеосом необхідне для експонування регуляторних сайтів на ДНК до дії транскрипційних факторів, то у клітині існує спеціальна систе-

ма, що сприяє репозиціюванню нуклеосом за рахунок індукування проміжних структурних станів: АТР-залежні *комплекси ремоделювання хроматину*.

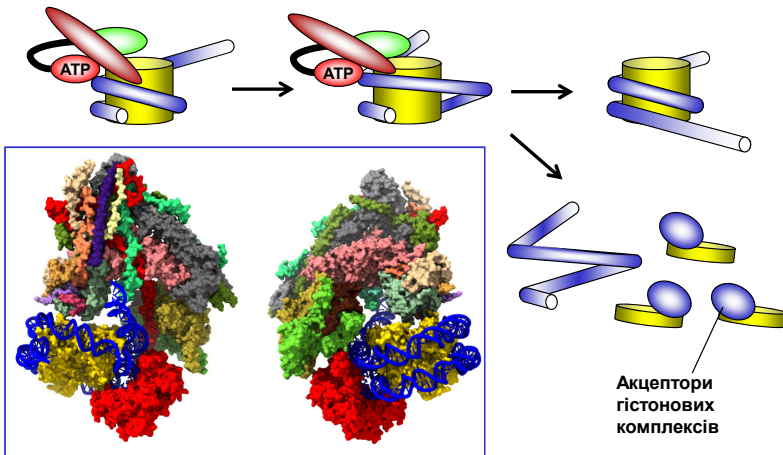
Принцип роботи комплексів ремоделювання полягає в тому, що вони використовують енергію гідролізу АТР для порушення ДНК-гістонових взаємодій – продукують високоенергетичний інтермедіатний стан нуклеосоми із частково розгорнутою нуклеосомною ДНК (рис. 6.21). Виходом із такого проміжного стану є відновлення стабільної нуклеосоми на іншій позиції або тимчасове переміщення гістонових комплексів на проміжні акцептори гістонів, що обговорювались у розділі 4, тобто можна сказати, що комплекси ремоделювання хроматину здійснюють каталіз структурних перетворень нуклеосоми.

Мультибілкові комплекси ремоделювання мають велику молекулярну вагу (у деяких випадках вона значно перевищує таку нуклеосоми) і містять від 4 до 15 субодиниць. Одна з них володіє АТРазною активністю. Власне АТРазний домен, гомологічний ДНК-геліказам певного класу, в усіх комплексах дуже схожий. Залежно від наявності додаткових структурних доменів різного типу у складі каталітичної субодиниці комплекси ремоделювання поділяють на чотири родини, що характеризуються певними особливостями.

Комплекси ремоделювання здатні здійснювати численні взаємодії з нуклеосомною ДНК, гістоновими хвостами й певними хімічними модифікаціями в їхньому складі, глобулярними частинами гістонів, а також специфічними й загальними факторами транскрипції, гістон-ацетилтрансферазами тощо. При взаємодії з нуклеосомною комплекс жорстко фіксується на ній, а АТРазний домен каталітичної субодиниці взаємодіє з нуклеосомною ДНК приблизно за два витки подвійної спіралі від центра (рис. 6.22).



**Рис. 6.21.** Каталіз структурних перетворень комплексами ремоделювання хроматину: реалізація проміжних інтермедіатних станів нуклеосоми із частково розгорнутою ДНК, які сприяють або репозиціюванню, або видаленню нуклеосом



**Рис. 6.22.** Механізм дії комплексу ремоделювання.

АТРазний домен позначено червоним, ДНК-зв'язувальний – зеленим, інші субодиниці, що взаємодіють із гістонами – коричневим. АТП-залежна транслокація АТРазного домену, закореного на нуклеосомі, викликає формування петлі ДНК на поверхні октамеру гістонів. Переміщення петлі зумовлює зміни позиції нуклеосоми, а проміжні акцептори здатні видалити гістонові комплекси із дестабілізованої нуклеосоми.

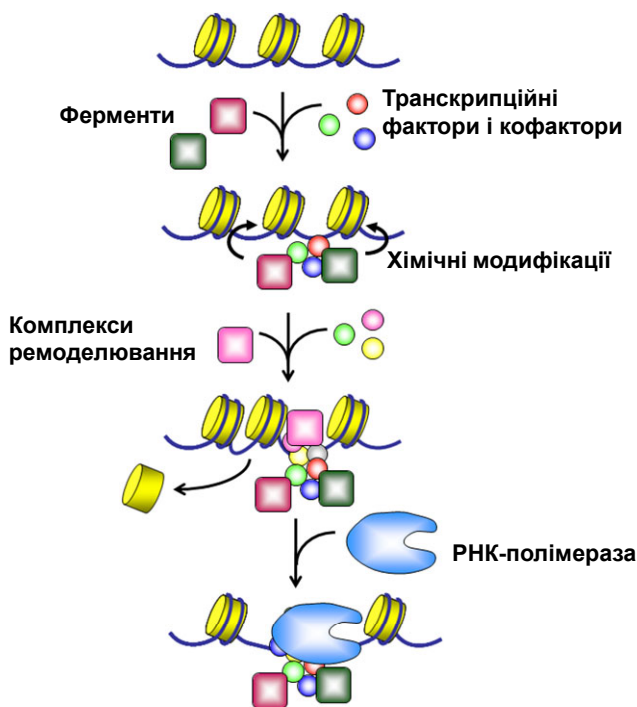
На вставці: структура комплексу ремоделювання SWI/SNF, зв'язаного з нуклеосомою (6UXW), у двох проєкціях (АТРазна субодиниця зафарбована червоним)

Механізм дії комплексів ремоделювання базується на гомології із ДНК-геліказами. Хоча АТРазна субодиниця не здатна діяти як геліказа (розводити два ланцюги), вона, подібно до геліказ, може пересуватись уздовж ДНК (але без її руйнування), використовуючи енергію гідролізу АТР. Загальний механізм схематично зображено на рис. 6.22. АТРазний домен взаємодіє з нуклеосомною ДНК поблизу від центра, інший ДНК-зв'язувальний домен АТРазної субодиниці – на вході в нуклеосому. Цей останній домен порушує взаємодії кінцевої ділянки ДНК із гістонами, "віддираючи" ДНК з поверхні октамеру. Зв'язування і гідроліз АТР індують конформаційні зміни й, відповідно, транслокацію АТРазного домену вздовж ДНК у напрямку до ДНК-зв'язувального домену. Оскільки положення АТРазного домену при цьому фіксоване, то така транслокація викликає "шттовання" нуклеосомної ДНК із периферії усередину нуклеосоми – на поверхні гістонового октамеру утворюється петля ДНК, яка потім мігрує від центра у другу половину нуклеосоми. Результатом міграції петлі та, зрештою, виходу її назовні є трансляційне переміщення нуклеосоми.

Різні комплекси ремоделювання можуть створювати петлі ДНК різного розміру на поверхні нуклеосоми: від мікропетель, що містять до 12 пар основ, до великих петель розміром до ~100 пар основ (як це характерно, зокрема, для комплексу SWI/SNF). У першому випадку нуклеосома під час ремоделювання несуттєво відрізняється за структурою від "звичайної", у другому – генеруються інтермедіатні структури нуклеосоми, у складі яких великі ділянки нуклеосомної ДНК втрачають зв'язок із гістонами, що створює умови для ефективного перенесення гістонових комплексів на проміжні гістонові акцептори, тобто тимчасового видалення нуклеосоми.

Функціональні наслідки ремоделювання хроматину певним комплексом не обов'язково пов'язані з активацією транскрипції, іноді зміни в розташуванні нуклеосом спричиняють репресію – залежно від контексту інших функціонально важливих впливів, у кооперації з якими працює комплекс. У разі активації система АТР-залежного ремоделювання хроматину завжди працює коопе-

ративно із системою ацетилювання гістонів. Реалізуються різноманітні стратегії рекрутування комплексів ремоделювання до промоторів: за рахунок взаємодії комплексів із транскрипційними факторами і кофакторами, упізнання ацетильованих та інших модифікованих залишків у гістонах тощо (один із типових сценаріїв зображено на рис. 6.23). Отже, комплекси ремоделювання хроматину інкорпоровані в загальну систему регуляції транскрипції і працюють у тісній кооперації із транскрипційними факторами й системою посттрансляційних модифікацій гістонів.



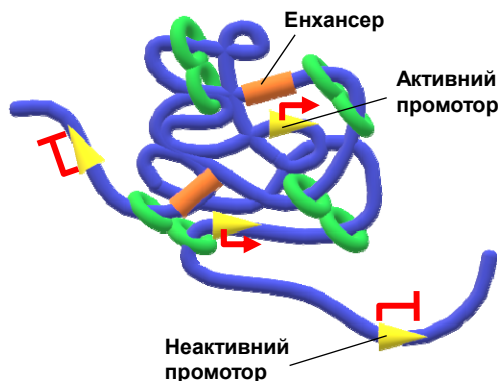
**Рис. 6.23.** Один із можливих сценаріїв послідовності подій при активації промотору

## Активация транскрипції на відстані: енхансери

У попередньому підрозділі йшлося про перебудови хроматину на нуклеосомному рівні. Дія *енхансерів* – *цис*-елементів, які розташовані на будь-якій відстані від стартової точки транскрипції гена, – зазвичай пов'язана зі структурою хроматину на вищих рівнях його організації.

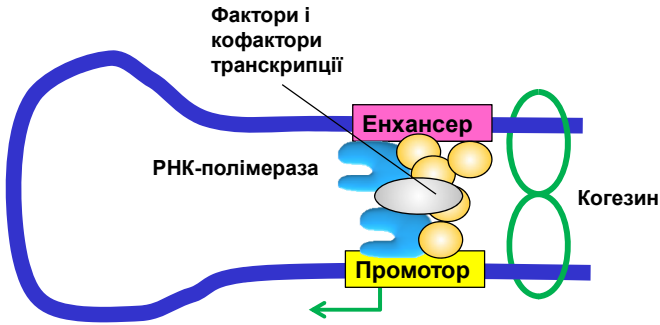
Як зазначалося вище, енхансери часто є також промоторами, куди завантажуються РНК-полімераза і звідки розпочинається транскрипція. Іноді транскрипт, що продукується енхансером, є регуляторною РНК, яка діє як *транс*-фактор, тобто включається в енхансому, що формується на деякому промоторі (рис. 6.14, *а*). Проте частіше механізм впливу енхансера полягає у його прямій фізичній взаємодії із промотором через транскрипційні фактори й кофактори, зв'язані з обома цими зонами (рис. 6.12). Зрозуміло, що між промотором і енхансером у такому випадку формується хроматинова петля, часто стабілізована когезином і білками CTCF (розділ 4). Отже, "передумовлені" хроматинові петлі, розподіл яких по геному варіює у клітинах різних типів, визначають можливість взаємодій певного енхансера з певним промотором. При цьому не тільки промотор і енхансер, що локалізовані в основах петлі, можуть взаємодіяти один з одним. Процес тимчасової екструзії (розділ 4) суттєво підвищує частоту контактів між будь-якими точками всередині петельного (або топологічно асоційованого) домену, що і створює можливість взаємодії. І навпаки, промотори (або енхансери), що перебувають поза межами топологічно асоційованого домену, втрачають таку можливість (рис. 6.24).

Якщо енхансер і промотор містяться в основах петлі, причому енхансер є одночасно промотором, то в основі петлі опиняються дві РНК-полімерази у складі фактично єдиного мультибілкового комплексу різноманітних факторів і кофакторів транскрипції (рис. 6.25). На рис. 6.24, таким чином, зображено мінімальну *транскрипційну фабрику*, у складі якої можуть бути об'єднані кілька РНК-полімераз і, відповідно, кілька промоторів і/або енхансерів.

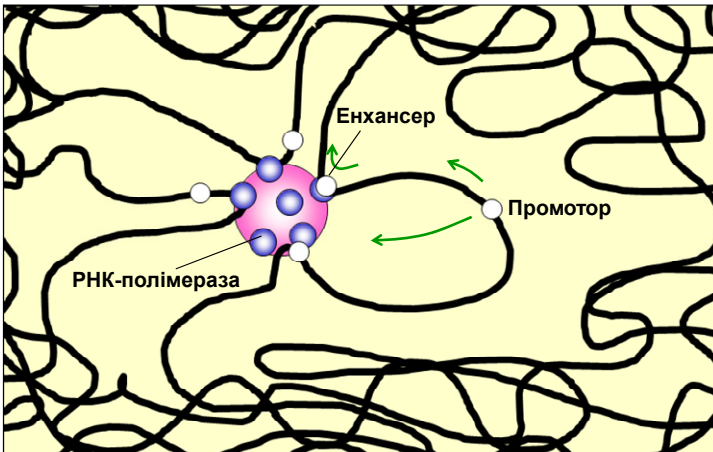


**Рис. 6.24.** Енхансери впливають на промотори в межах петельного (топологічно асоційованого) домену й не впливають на них поза його межами

**Транскрипційні фабрики** – великі мультимолекулярні комплекси, які містять віддалені по ланцюгу регуляторні *цис*-елементи, зв'язані з ними білки й молекули РНК-полімерази (рис. 6.26). У різних клітинах містяться від кількох сотень до кількох тисяч таких фабрик, на яких об'єднуються від двох до кількох десятків транскрипційних одиниць (одна з таких спеціалізованих фабрик – ядерце, де здійснюється синтез рибосомної РНК РНК-полімеразою I на об'єднаних у єдину структуру ядерцевих організаторах різних хромосом, – давно й добре відома). Різноманітні елементи системи активації транскрипції можуть бути об'єднані в межах транскрипційної фабрики і спільно використовуватись кількома промоторами, які зв'язуються з фабрикою при ініціації транскрипції. При цьому певні елементи системи регуляції транскрипції (включаючи РНК-полімеразу) можуть зв'язуватись із промоторами заздалегідь, переводячи їх у стан потенційної здатності взаємодіяти із транскрипційною фабрикою. Така фабрика може служити місцем для узгодженої транскрипції групи генів, що активуються спільним набором активаторів і, завдяки об'єднанню у єдиний комплекс, взаємно підтримують активований стан одне одного.



**Рис. 6.25.** Об'єднання двох РНК-полімераз, зв'язаних із енхансером і промотором, у єдиний комплекс



**Рис. 6.26.** Транскрипційна фабрика й один із механізмів активуючого впливу енхансера на промотор

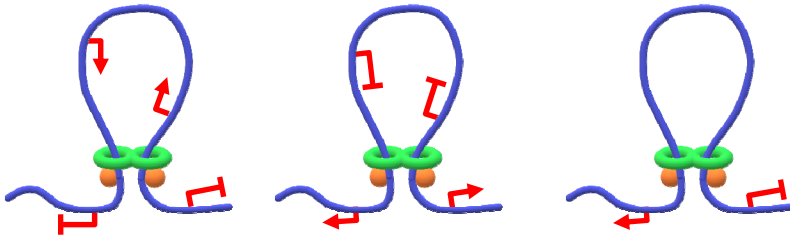
Якщо енхансер є промотором, з якого розпочинається транскрипція у напрямку до промотору гена, що активується енхансером (рис. 6.26), то, оскільки РНК-полімераза заякорена у транскрипційній фабриці, цей останній промотор згодом "втягується"

до фабрики внаслідок транскрипційної активності енхансера (ще один механізм дії енхансерів).

Поряд із транскрипційними фабриками, у межах яких об'єднуються кілька промоторів, у клітинному ядрі має місце ще один феномен – об'єднання кількох особливих енхансерів у *суперенхансер*. Суперенхансер формує мікрокраплю (подібну до таких, що обговорювались у розділі 4, див. також рис. 4.18), де зосереджені транскрипційні фактори й кофактори у високій локальній концентрації. На промоторах, які опиняються поряд із такою мікрокраплею (переважно це стосується промоторів генів, що визначають клітинний тип і активуються на ранніх стадіях диференціації клітин), надзвичайно ефективно запускається транскрипція.

Важливою умовою того, що енхансер у клітинах певного типу буде впливати на промотори певного гена, є не тільки наявність певного набору транскрипційних факторів, що взаємодіють із енхансером, а й відсутність перешкод на шляху передавання регуляторного сигналу з енхансера на промотор. Такі перешкоди створюють *інсулятори* – спеціальні зв'язані з інсуляторними білками ділянки, які встановлюють границі між активними й неактивними доменами хроматину. Основний механізм інсуляції було проілюстровано на рис. 6.24: енхансери ефективно впливають на промотори, що містяться у спільному із цими енхансерами петельному або топологічно асоційованому домені, і не впливають на промотори з іншого такого домену. Отже, інсуляторами часто стають границі петельного домену, де знаходяться білки CTCF, які є основним типом інсуляторних білків.

Іноді розрізняють щойно розглянуті енхансерблокувальні інсулятори та бар'єрні, тобто такі, що розділяють гетеро- й еухроматинові компартменти. Бар'єрні інсулятори часто також пов'язані з петельною організацією хроматину (рис. 6.27). Деякі інші механізми створення границь між еу- й гетерохроматином розглянуто нижче.



**Рис. 6.27.** Петельний домен як границя між різними функціональними зонами хроматину (зліва направо): у межах домену містяться активні промотори, тоді як неактивні – поза його межами; неактивні промотори всередині домену, активні зовні; петельний домен відмежовує активні зони хроматину від неактивних

### Конститутивна репресія транскрипції: гетерохроматин

Великий розмір еукаріотичного геному, майже половина якого припадає на послідовності, що повторюються, вимагає існування систем, які б визначали гарантовану інактивацію значної частини генетичного матеріалу. Прикладами *конститутивного гетерохроматину* (такого, що утворюється в усіх клітинах) є субтеломерні й перичентромерні зони хромосом, одна із X-хромосом самок ссавців, ділянки скупчення мобільних елементів. Інші гетерохроматинові зони є специфічними для клітин певного типу (*факультативний гетерохроматин*). При цьому інактивованій стан тих чи інших ділянок хроматину відтворюється у дочірніх клітинах після мітозу. Як уже згадувалось у розділі 4, успадкування такої тканиноспецифічної системи репресії (при збереженні повної генетичної програми організму) називають *епігенетичною спадковістю*. У її основі лежить передання клітинам-нащадкам не просто батьківської ДНК, а хроматину разом із різноманітними хімічними маркерами. Ці маркери упізнаються відповідними білками, які вступають у різноманітні взаємодії з репресорами й корепресорами транскрипції,

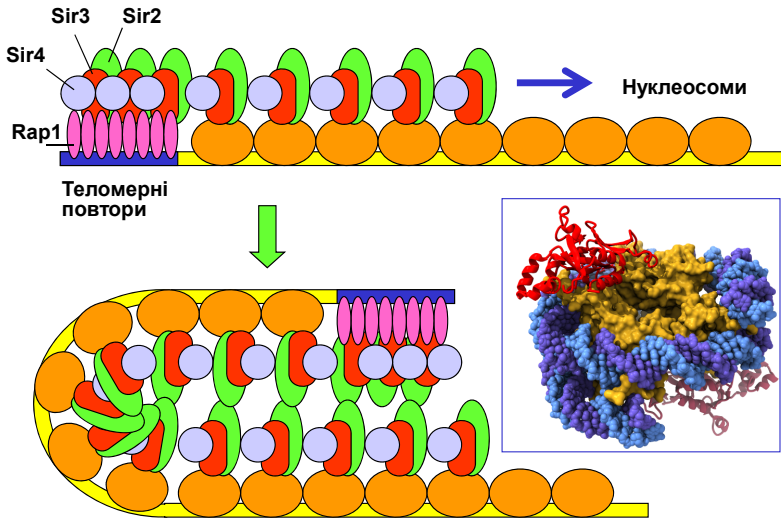
ферментами, що здійснюють посттрансляційні модифікації, і білками, які додатково компактизують хроматинову фібрилу. Розглянемо принципи виникнення та підтримання гетерохроматинового стану на кількох прикладах.

### **Субтеломерні зони у *S. cerevisiae***

Якщо ацетилювання гістонів завжди корелює з підвищеною транскрипційною активністю, то деацетилювання, яке здійснюється гістон-деацетилазними комплексами (HD), завжди пов'язане з репресією. Як і ацетилтрансферази, деацетилази постійно безадресно працюють у хроматині, підтримуючи певний базовий баланс ацетилювання / деацетилювання гістонів. При активації певного промотору ацетилтрансферази здійснюють адресне гіперацетилювання, а після зникнення активуючого сигналу HD повертають промотор до базового неактивного стану. Деацетилази також можуть бути адресно рекрутовані до промоторів репресорами транскрипції для підтримання гарантованого деацетилюваного статусу.

Поряд із такою динамічною регуляцією активності, деацетилювання гістонів завжди здійснюється в гетерохроматинових ділянках. Прикладом, який демонструє визначальну роль деацетилювання у підтриманні гетерохроматинового стану, є теломери дріжджів *S. cerevisiae* (рис. 6.28).

Теломерні повтори на кінці хромосоми упізнаються ДНК-зв'язувальним білком Rap1 (Repressor activator protein), який рекрутує білок Sir4 (Silencing information regulator). Цей останній зв'язує ще два білки – Sir3 та Sir2, другий із них є гістон-деацетилазою. Sir2 здійснює деацетилювання кількох специфічних лізинових залишків гістонів H4 та H3 найближчої нуклеосоми. Ці деацетилювані залишки упізнаються новим Sir3, унаслідок чого утворюється новий Sir-комплекс і здійснюється деацетилювання нової нуклеосоми. Процес деацетилювання і збирання Sir-комплексу розповсюджується вздовж фібрили, а за рахунок взаємодії між Sir-комплексами на кінцях хромосоми забезпечується надкомпактизація хроматину (рис. 6.28).



**Рис. 6.28.** Розповсюдження і самопідтримання гетерохроматинового стану в субтеломерах дріжджів. На вставці: комплекс нуклеосоми із двома білками Sir3 (3TU4)

Виникнення гетерохроматину в субтеломерних зонах демонструє кілька загальних принципів: визначальну роль відіграють хімічні маркери на гістонах; білки, що упізнають такі маркери, служать засобом додаткової компактизації; гетерохроматин здійснює експансію на сусідні ділянки на невизначену відстань.

Останній принцип зумовлює необхідність у певній точці зупинити таку експансію. Цю функцію виконують бар'єрні інсулятори. Інсуляторами, що відокремлюють гетерохроматинові компартменти від еухроматинових, можуть бути границі петельних доменів (рис. 6.27). Крім того, бар'єрний інсулятор може бути зоною, збідненою на нуклеосоми, тобто прогалиною, на якій відбувається обривання ланцюга реакцій, що передають гетерохроматинові маркери з нуклеосоми на нуклеосому. Наявність такої прогалини може залежати від нуклеотидної послідовності, зв'язування специфічних білків, активності ремодельовальних комплексів. Ще одна можливість утворення границі (яка не виключає попередні) – формування зони, де зв'язуються специфіч-

ні білки, які рекрутують ферментативні активності, що вносять у хроматин еухроматинові маркери, – гістон-ацетилтрансферази та гістон-метилтрансферази, від яких залежить метилювання Lys4 у гістоні H3. Гіперацетилювання гістонів і метилювання Lys4 блокує роботу ферментів, що вносять гетерохроматинові маркери, перешкоджає зв'язуванню гетерохроматинових білків. Інсулятором може бути також гіперактивний ген (зокрема, гетерохроматинову зону в субтеломерах дріжджів відокремлює ген tРНК) або сильний енхансер. Узагалі, бар'єрні інсулятори мають багато спільних рис із енхансерами й промоторами: усі ці ділянки рекрутують подібні активності й часто збіднені на нуклеосомі.

### **HP1-залежна система репресії**

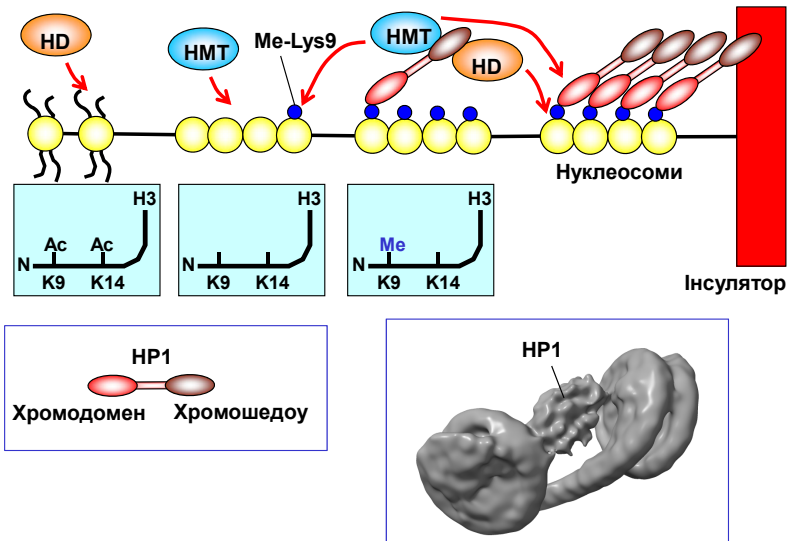
Система репресії, опосередкована білком HP1 (Heterochromatin Protein 1), має доволі універсальний характер і реалізується у багатьох репресованих ділянках. Зокрема, такий механізм працює у перицентромерних зонах хромосом.

Білок HP1 містить два структурні домени: *хромодомен* (модуль, що впізнає метильовані лізинові залишки) і *хромощедоу* (chromoshadow) – домен, який має спорідненість до певних специфічних HD і специфічної гістон-метилтрансферази (HMT), а також здатен взаємодіяти з таким самим доменом іншої молекули HP1 (рис. 6.29).

Як і в субтеломерних зонах, процес формування гетерохроматину в цьому випадку може розпочинатись із деацетилювання гістонів. Деацетилювання Lys9 та Lys14 гістону H3 сприяє зв'язуванню специфічної гістон-метилтрансферази (HMT), яка здійснює метилювання Lys9 гістону H3 (рис. 6.29). Цей метильований Lys (Me-Lys9) специфічно упізнається хромодоменом HP1. Завдяки хромощедоу-домону HP1 рекрутує HD, яка підтримує деацетилюваний статус суміжних нуклеосом, і ту саму HMT, що здійснює метилювання Lys9 на суміжних нуклеосомах. Знову виникає лавиноподібний процес, що самопідтримується і розповсюджується на сусідні ділянки. Взаємодія між білками HP1 (зв'язаними із сусідніми або віддаленими по ланцюгу нуклеосомами) забезпечує додаткову компактизацію фібрили. Підвищення локальної концентрації HP1 запускає процес форму-

вання відокремленої мікрокраплі гетерохроматину (див. рис. 4.18). Крім того, HP1 має спорідненість до білків ламіни (розділ 4), що зумовлює розташування гетерохроматину в периферичних зонах клітинного ядра.

Отже, додаткова компактизація забезпечується особливим білком, що впізнає певний набір гетерохроматинових маркерів. При реплікації ДНК цей набір відтворюється у дочірніх хромосомах (розділ 9) – гетерохроматиновий стан спадкується дочірніми клітинами.

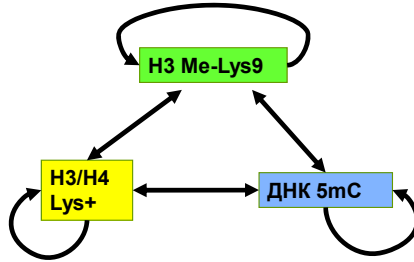


**Рис. 6.29.** Розповсюдження і самопідтримання гетерохроматинового стану в перичентромерних зонах.

На вставках: стан модифікацій двох ключових залишків у N-кінцевому хвості гістону H3, схема структури HP1 і комплекс димеру HP1 із двома сусідніми нуклеосомами (зображення отримане за допомогою криоелектронної мікроскопії, The Electron Microscopy Data Bank, код 6738)

Метилування Lys9 гістону H3 і деацетильований стан лізинів гістонів H3/H4 є найхарактернішою ознакою репресованих ділянок. Обидві модифікації самопідтримуються і підтримують

одна одну через опосередковану дію HP1. При цьому важливу роль у забезпеченні репресованого стану відіграє метилювання цитозинів ДНК – інша ковалентна модифікація, яка також відновлюється при клітинному поділі й замикає своєрідне коло репресії (рис. 6.30).

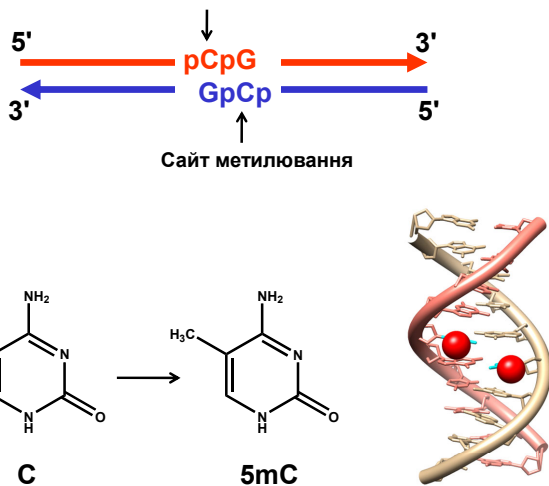


**Рис. 6.30.** Взаємний вплив деацетилювання гістонів, метилювання Lys9 гістону H3 і метилювання ДНК у гетерохроматині

### Метилювання ДНК

Метилюванню в ДНК еукаріотів переважно піддаються цитозини (метильна група приєднується до п'ятого атома кільця з утворенням 5mC – 5-метилцитозину) у складі динуклеотидів CpG. Дві симетрично розташовані метильні групи опиняються у результаті у великому жолобку подвійної спіралі (рис. 6.31). Іноді це може впливати на взаємодію ДНК-зв'язувальних білків із ДНК. Зокрема, метилювання CTCF-мотивів спричиняє зниження їхньої спорідненості до білків CTCF і знищує границю петельного домену.

Загалом 70–80 % динуклеотидних контактів CpG є метильованими в геномі диференційованих клітин хребетних. Зони, де підтримується деметильований стан CpG (CpG-острівці, див. розділ 4), часто розташовані у промоторах генів домашнього господарства – таких, що є активними незалежно від спеціалізації клітин. Геномний профіль тканиноспецифічного метилювання ДНК є результатом двох процесів, яким відповідають два типи ДНК-метилтрансфераз (DNA methyltransferase, Dnmt): підтримання метильованого статусу після реплікації і метилювання *de novo*.



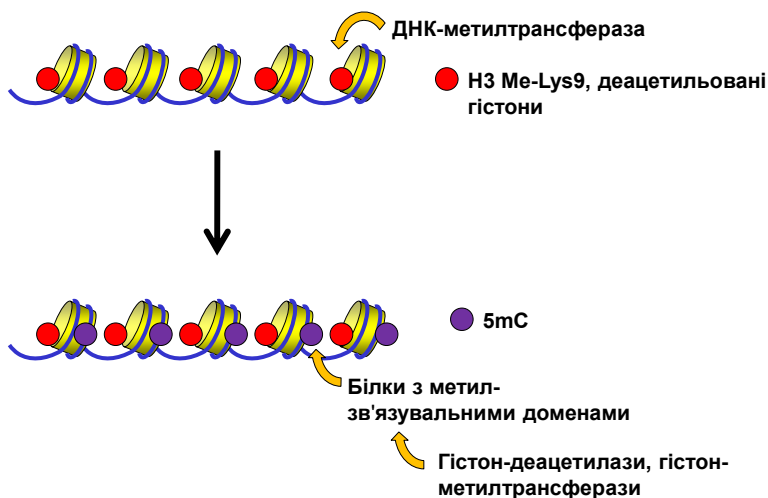
**Рис. 6.31.** Сайти метилювання в ДНК – динуклеотиди CpG, структурні формули цитозину (C) і 5-метил-цитозину (5mC) і положення метильних груп (червоні кульки) у великому жолобку подвійної спіралі (6JV5)

Поява метильних груп у ДНК відбувається на ранніх стадіях ембріонального розвитку хребетних, коли ДНК тотально деметильована. У процесі диференціації реалізується принцип "спочатку гістони, потім ДНК": визначаються репресовані зони хроматину, де накопичуються деацетильовані гістони й відбувається метилювання Lys9 у гістоні H3. Обидві модифікації сприяють рекрутуванню ДНК-метилтрансферази, що забезпечує метилювання *de novo* (метилювання ДНК надійніше закріплює репресований статус).

Залучення 5mC до репресії пов'язане з наявністю у складі певних білків особливих структурних модулів – метилзв'язувальних доменів (Methyl Binding Domain, MBD), які мають специфічну спорідненість до метильованих динуклеотидів CpG. Білки, що містять MBD, є компонентами різноманітних репресуючих комплексів. Зокрема, такі білки рекрутують до метильованих ділянок хроматину гістон-деацетилази й /гістон-метилтрансферазу, яка здійснює метилювання Lys9

гістону H3 (рис. 6.32). Виникає зв'язок між метилюванням ДНК та Lys9 і деацетилюванням гістонів, відображений на рис. 6.30. Якщо деацетилюваний стан гістонових хвостів блокує деметилювальні активності, то ацетилювання хвостів, навпаки, може викликати деметилювання ДНК і порушення репресованого статусу відповідної ділянки.

Коли певний профіль метилювання ДНК уже сформований, він має спадкуватись дочірніми клітинами. За це відповідає підтримувальна ДНК-метилтрансфераза, що працює на дочірніх ланцюгах ДНК під час реплікації. Одразу після синтезу дві дочірні молекули ДНК містять по одному батьківському ланцюгу ДНК (із 5mC у складі CpG) і по одному синтезованому, де С не є метильованим. Підтримувальна метилтрансфераза упізнає такі напівметилювані динуклеотидні контакти й відновлює симетрію щодо метилювання цитозину (рис. 6.33).



**Рис. 6.32.** Метилювання ДНК *de novo* і його зв'язок із хімічними модифікаціями гістонів

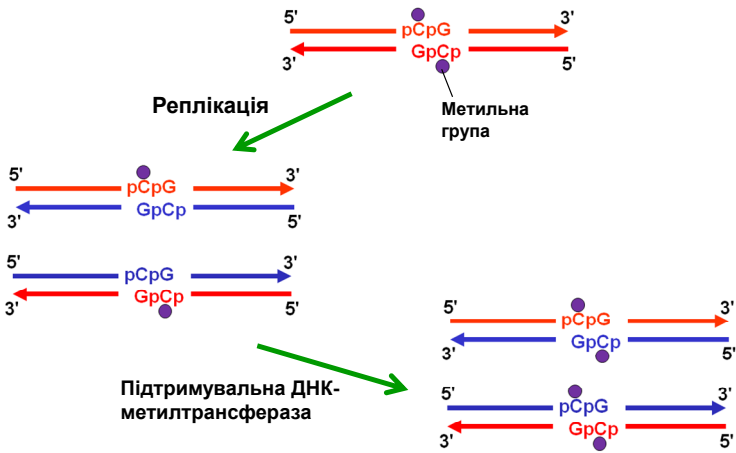


Рис. 6.33. Відновлення метилювання динуклеотидів CpG у процесі реплікації ДНК

### Білки групи полікомб і некодувальні РНК

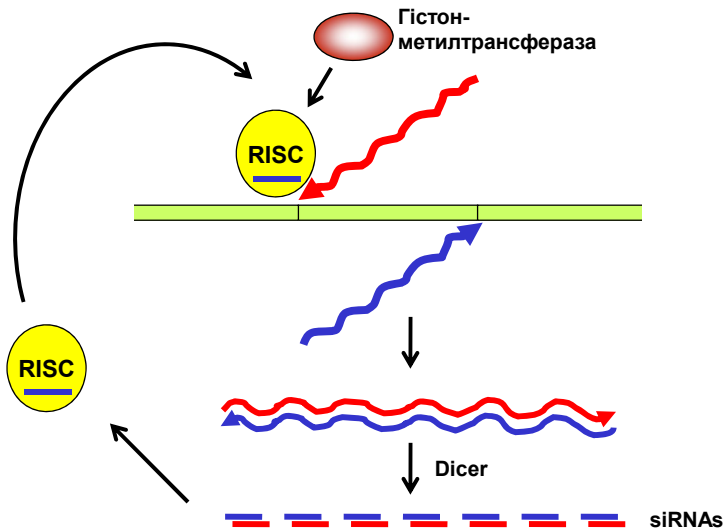
Хоча 5mC і метильовані Lys9 та деякі інші лізинові залишки є загальними маркерами репресованих ділянок хроматину, не завжди репресія і додаткова компактизація залежать від HP1 – реалізуються також інші системи. Одним із прикладів таких систем є консервативні для еукаріотів *репресорні комплекси полікомб* (PRC, Polycomb Repressive Complex), що складаються із кількох субодиниць (білків групи полікомб, PcG) і забезпечують додаткову компактизацію певних гетерохроматинових зон. До репресорних комплексів полікомб можуть залучатись також різноманітні регуляторні РНК. Ключова роль у забезпеченні гетерохроматинізації тут належить ще одному характерному маркеру репресії – триметильованому Lys27 у гістоні H3. Окремі субодиниці PRC можуть як упізнавати цей метильований залишок, так і забезпечувати метилювання.

Інший приклад системи гетерохроматинізації у масштабі цілої хромосоми – інактивація однієї із Х-хромосом у клітинах самок ссавців. Така інактивована хромосома залишається у надкомпактному стані під час інтерфази, її можна побачити в оптич-

ний мікроскоп у вигляді тільця Барра (Barr body). Ключову роль тут відіграє велика некодувальна РНК *Xist* разом із деякими іншими некодувальними РНК. У складі Х-хромосоми, яка буде інактивована (обирається випадково на ранніх стадіях розвитку), спрацьовує ген *Xist*, що продукує *Xist* як транскрипт. РНК *Xist* укривають собою хромосому і взаємодіють із особливими білками, серед яких є варіант гістону H2A *macroH2A* і білки групи полікомб. Ці білки, забезпечуючи надкомпактизацію, рекрутують гістон-метилтрансферази, що здійснюють метилювання Lys9 та Lys27 гістону H3, і гістон-деацетилазу. Метилювання Lys9, своєю чергою, зумовлює метилювання ДНК, що забезпечує надійну епігенетичну спадковість.

До виникнення гетерохроматинових зон (у т. ч. опосередкованих HP1) має відношення розглянутий вище процес РНК-інтерференції. По-перше, зв'язування певних мікроРНК у складі комплексу RISC із мРНК під час транскрипції (див. рис. 6.13) не тільки зумовлює її деградацію, а й викликає рекрутування до хроматину гістон-метилтрансферази, що здійснює метилювання Lys9 у гістоні H3 із подальшою репресією гена (за схемою на рис. 6.29), у результаті чого здійснюється входження в коло репресії, зображене на рис. 6.30.

По-друге, гетерохроматин утворюється передусім на послідовностях ДНК, які повторюються, причому має значення не послідовність як така, а саме наявність повторів. Випадкова транскрипція на таких повторах у різних напрямках (зокрема, це може відбуватись на центромерних повторах при порушенні компактизації) з певною імовірністю продукує взаємно комплементарні транскрипти: утворюється дволанцюгова РНК, яка запускає процес РНК-інтерференції (рис. 6.34). Знову, крім деградації транскриптів, RISC, який опиняється в зоні повторів, рекрутує до хроматину гістон-метилтрансферазу, що здійснює метилювання Lys9 гістону H3. За вже відомою схемою (рис. 6.29) відбувається зв'язування HP1 і компактизація гетерохроматину.

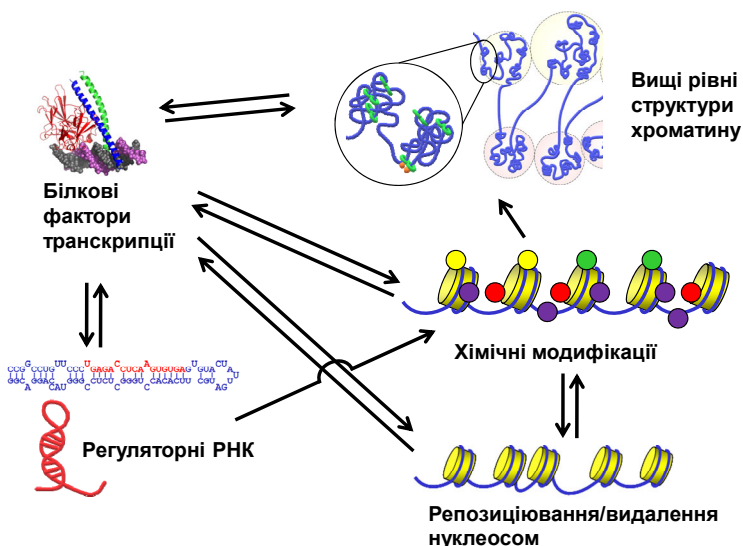


**Рис. 6.34.** РНК-інтерференція як механізм підтримання гетерохроматинового стану послідовностей ДНК, що повторюються

## Загальна схема регуляції транскрипції в еукаріотів

Підсумовуючи обговорені вище механізми регуляції транскрипції в еукаріотів, можна зобразити схему, у якій об'єднані три основні шляхи впливу на експресію генів на рівні транскрипції через транскрипційні фактори, регуляторні РНК і зміни патерну хімічних модифікацій і структури хроматину (рис. 6.35).

Усі три шляхи взаємопов'язані та інтегровані в єдину систему. Власне, усі процеси, спрямовані на регуляцію транскрипції, базуються на різноманітних випадкових міжмолекулярних взаємодіях, баланс яких зсувається у той чи інший бік (активації чи репресії) залежно від локальних концентрацій різноманітних компонентів у різних ділянках хроматину і, відповідно, взаємного підсилення впливів різних факторів.



**Рис. 6.35.** Взаємозв'язок між елементами системи регуляції транскрипції в еукаріотів

Зокрема, ефективність взаємодії транскрипційних факторів і кофакторів визначається доступністю їхніх сайтів зв'язування у ДНК і патерном модифікацій ДНК і гістонів. І навпаки, транскрипційні фактори здатні рекрутувати ферментативні активності, що вносять хімічні модифікації, які упізнаються новими кофакторами, і комплекси ремоделювання хроматину, унаслідок роботи яких з'являються нові доступні сайти зв'язування для транскрипційних факторів і загальної транскрипційної машинерії. Під контролем транскрипційних факторів перебувають гени мікроРНК та інших регуляторних РНК, і навпаки, гени транскрипційних факторів самі є мішенями контролю з боку мікроРНК. Крім того, регуляторні РНК залучаються до активаційних чи репресорних мультимолекулярних комплексів і здатні стимулювати певні хімічні модифікації у хроматині. Ці останні відіграють визначальну роль у формуванні еу- й гетерохроматинових компартмен-

тів, у межах яких створюється відповідно або потенційна можливість транскрипції, або тотальне блокування доступу для транскрипційної машинерії.

### Контрольні запитання

1. Яка різниця між *цис*-елементами і *транс*-факторами системи регуляції транскрипції?

2. Опишіть систему регуляції лактозного оперона. Яку роль у регуляції відіграють *lac*-репресор і білок CAP?

3. Як здійснюється атенуація у триптофановому опероні?

4. Що таке РНК-перемикач?

5. Охарактеризуйте стадії життєвого циклу бактеріофага  $\lambda$ . Як відбувається перемикання між різними шляхами розвитку?

6. Репресор бактеріофага  $\lambda$  є позитивним чи негативним регулятором транскрипції гена *cro*; гена *cI*?

7. Що таке транскриптом? Які типи РНК входять до його складу?

8. Що таке енхансосома? Поясніть принципи модульності та кооперативності при взаємодії факторів транскрипції з еукаріотичними промоторами.

9. Що таке гормональний рецептор?

10. Дайте визначення сигнальної трансдукції.

11. Що таке РНК-інтерференція?

12. Дайте визначення мікроРНК і опишіть їхню роль у регуляції транскрипції.

13. Яка посттрансляційна модифікація гістонів позитивно корелює із транскрипційною активністю? У чому полягають механізми такої кореляції?

14. Назвіть модифікації гістонів, характерні для активного й репресованого станів.

15. Які особливості структури хроматину спостерігаються в еукаріотичних промоторах?

16. Охарактеризуйте комплекси ремоделювання хроматину. Яким є головний результат їхньої дії? У чому полягають механізми такої дії?

17. Опишіть механізми дії енхансерів на промотори.

18. Що таке інсулятор? Опишіть механізми дії інсуляторів.
19. Опишіть функціональну роль білка HP1 у підтриманні гетерохроматинового стану.
20. Яку роль відіграє метилювання ДНК у репресії транскрипції?
21. У чому полягає роль РНК-інтерференції в підтриманні гетерохроматинового стану?

## Рекомендована література

Ali, T., Renkawitz, R. and Bartkuhn, M. (2016) Insulators and domains of gene expression. *Curr. Opin. Genet. Dev.* 37, 17–26.

Allshire, R.C. and Madhani, H.D. (2018) Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* 19, 229–244.

Browning, D.F. and Busby, S.J. (2016) Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* 14, 638–650.

Browning, D.F., Butala, M. and Busby, S.J.W. (2019) Bacterial transcription factors: regulation by pick "N" mix. *J. Mol. Biol.* 431, 4067–4077.

Cairns, B.R. (2009) The logic of chromatin architecture and remodelling at promoters. *Nature* 461, 193–198.

Carter, D.R.F., Eskiw, C. and Cook, P.R. (2008) Transcription factories. *Biochem. Soc. Trans.* 36, 585–589.

Chen, D. and Lei, E.P. (2019) Function and regulation of chromatin insulators in dynamic genome organization. *Curr. Opin. Cell Biol.* 58, 61–68.

Clapier, C.R. and Cairns, B.R. (2009) The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* 78, 273–304.

Cramer, P. (2019) Organization and regulation of gene transcription. *Nature* 573, 45–54.

David, L., Huber, W., Granovskaia, M., et al. (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* 103, 5320–5325.

Dykes, I.M. and Emanuelli, C. (2017) Transcriptional and post-transcriptional gene regulation by long non-coding RNA. *Genomics Proteomics Bioinformatics* 15, 177–186.

Flynt, A.S. and Lai, E.C. (2008) Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat. Rev. Genet.* 9, 831–842.

Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.

Guttman, M. and Rinn, J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339–346.

Han, Y., Reyes, A.A., Malik, S. and He, Y. (2020) Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature* 579, 452–455.

Hnisz, D., Shrinivas, K., Young, R.A., et al. (2017) A phase separation model for transcriptional control. *Cell* 169, 13–23.

Holoch, D. and Margueron, R. (2017) Mechanisms regulating PRC2 recruitment and enzymatic activity. *Trends Biochem. Sci.* 42, 531–542.

Janssen, A., Colmenares, S.U. and Karpen, G.H. (2018) Heterochromatin: guardian of the genome. *Annu. Rev. Cell. Dev. Biol.* 34, 265–288.

Jiang, C. and Pugh, B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10, 161–172.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., et al. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458, 362–366.

Klemm, S.L., Shipony, Z. and Greenleaf, W.J. (2019) Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220.

Kouzine, F., Gupta, A., Baranello, L., et al. (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* 20, 396–403.

Krauss, G. (2003) *Biochemistry of signal transduction and regulation*. Weinheim : Wiley-VCH Verlag.

Lai, W.K.M. and Pugh, B.F. (2017) Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell. Biol.* 18, 548–562.

Latchman, D. (2015) *Gene control*. 2<sup>nd</sup> ed. New York : Garland Science.

Li, B., Carey, M. and Workman, J.L. (2007) The role of chromatin during transcription. *Cell* 128, 707–719.

Maison, C. and Almouzni, G. (2004) HP1 and the dynamics of heterochromatin maintenance. *Nature Rev.* 5, 296–304.

Matzke, M.A. and Birchler, J.A. (2004) RNAi-mediated pathways in the nucleus. *Nature Rev.* 6, 24–35.

Moritz, L.E., Trievel, R.C. (2018) Structure, mechanism, and regulation of polycomb-repressive complex 2. *J. Biol. Chem.* 293, 13805–13814.

Neil H., Malabat, C., d'Aubenton-Carafa, Y., et al. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042.

Ong, C.T. and Corces, V.G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283–293.

Panne, D. (2008) The enhanceosome. *Curr. Opin. Struct. Biol.* 18, 236–242.

Ptashne, M. (2004) *A genetic switch, third edition: phage lambda revisited*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Rando, O.J., Chang, H.Y. (2009) Genome-wide views of chromatin structure. *Annu. Rev. Biochem.* 78, 245–271.

Reyes, A.A., Marcum, R.D. and He, Y. (2021) Structure and function of chromatin remodelers. *J. Mol. Biol.* 433, 166929.

Rothbart, S.B. and Strahl, B.D. (2014) Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* 1839, 627–643.

Ruthenburg, A.J., Li, H., Patel, D.J. and Allis, C.D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 8, 983–994.

Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell* 152, 17–24.

Shlyueva, D., Stampfel and G., Stark, A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286.

The ENCODE Project Consortium. (2007) Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group. (2005) The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.

The modENCODE Consortium. (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330, 1787–1797.

Wilson, C.J., Zhan, H., Swint-Kruse, L. and Matthews, K.S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol. Life Sci.* 64, 3–16.

Xu, Z., Wei, W., Gagneur, J., et al. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037.

Zhang, T., Cooper, S. and Brockdorff, N. (2015) The interplay of histone modifications – writers that read. *EMBO Rep.* 16, 1467–1481.

Zhou, C.Y., Johnson, S.L., Gamarra, N.I. and Narlikar G.J. (2016) Mechanisms of ATP-dependent chromatin remodeling motors. *Annu. Rev. Biophys.* 45, 153–181.

## Розділ 7

# ПРОЦЕСИНГ ЕУКАРІОТИЧНИХ мРНК

The ability to simplify means to eliminate the unnecessary  
so that the necessary may speak.

*Hans Hofmann*

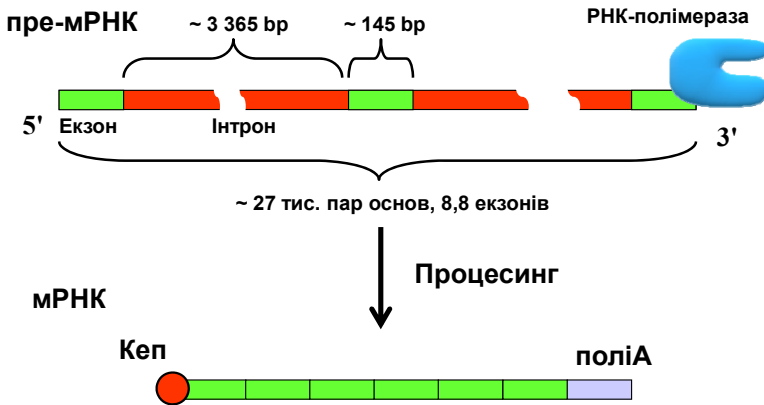
Безпосереднім продуктом транскрипції еукаріотичних білкових генів РНК-полімеразою II є пре-мРНК – попередник функціональної мРНК. Пре-мРНК не може бути використана як матриця для білкового синтезу, принаймні тому, що вона містить інтрони. Дозрівання пре-мРНК з утворенням функціональної матриці називають *процесингом* (processing). Він складається із трьох операцій (рис. 7.1).

*Кепування* – модифікація 5'-кінця з утворенням *кепа* (cap).

*Сплайсинг* (splicing) – вирізання інтронів і зшивання екзонів – процес, у результаті якого мРНК стає копією лише кодувальної частини гена або її фрагментів. Сплайсинг часто може відбуватися кількома альтернативними шляхами (альтернативний сплайсинг). У результаті сплайсингу із пре-мРНК видаляється до 90 % нуклеотидів.

*Поліаденілювання* – приєднання до 3'-кінця polyA-послідовності. Ця операція тісно пов'язана з термінацією транскрипції.

***Усі операції процесингу відбуваються під час транскрипції на РНК-полімеразному комплексі***, тобто процесинг є невід'ємною частиною транскрипції. Місцем збирання машинерії процесингу служить С-кінцевий домен (CTD) РНК-полімерази II, який відходить від полімерази поряд із каналом виходу РНК (розділ 5). Інакше кажучи, РНК-полімераза "тягне на собі" фактори процесингу, одночасно синтезуючи пре-мРНК і забезпечуючи її дозрівання.



**Рис. 7.1.** Процесинг мРНК.  
Показано середні значення довжини  
й кількості елементів пре-мРНК  
у первинних транскриптах людини

## Кепування

Будову кепа й хімічний процес його утворення показано на рис. 7.2. У складі пре-мРНК до 5'-атома рибози першого нуклеозидтрифосфату, який був включений до ланцюга РНК, залишаються приєднаними три фосфатні залишки.

*Перша стадія* кепування: відщеплення одного фосфату ( $\gamma$ -фосфату) із 5'-кінця, що здійснюється спеціальною фосфатазою.

*Друга стадія:* перенесення гуанозинмонофосфату (GMP) з гуанозинтрифосфату (GTP) на два фосфатні залишки, що залишились на 5'-кінці (відповідний фермент – гуанілілтрансфераза). Результатом перенесення є утворення ковалентного зв'язку між 5'-фосфатом GMP і 5'-кінцевим  $\beta$ -фосфатом пре-мРНК, тобто формування нехарактерного для нуклеїнових кислот зв'язку між нуклеотидами.

Третя стадія: метилювання метилтрансферазою кінцевого G з утворенням 7-метилгуаніну (7mG).

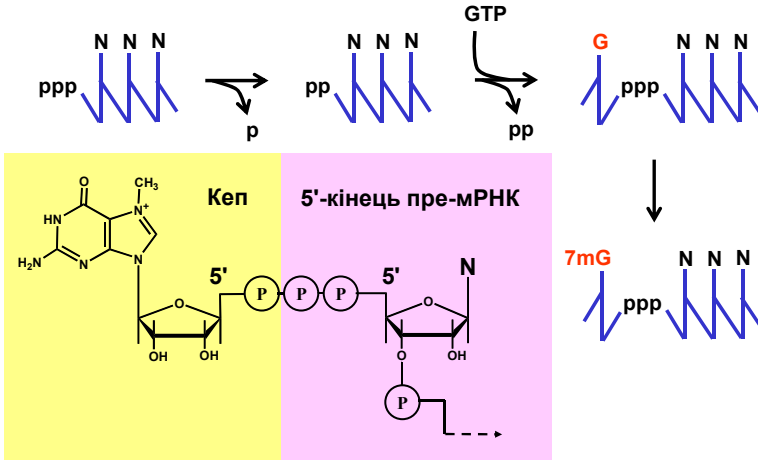


Рис. 7.2. Стадії утворення і хімічна структура кепа

Три ферментативні активності, які здійснюють кепування, містяться у трьох (дріжджі) або двох (ссавці) білках. Ці ферменти рекрутуються на CTD і 5'-кінець пре-мРНК після синтезу перших 20–40 нуклеотидів транскрипту (рис. 7.3).

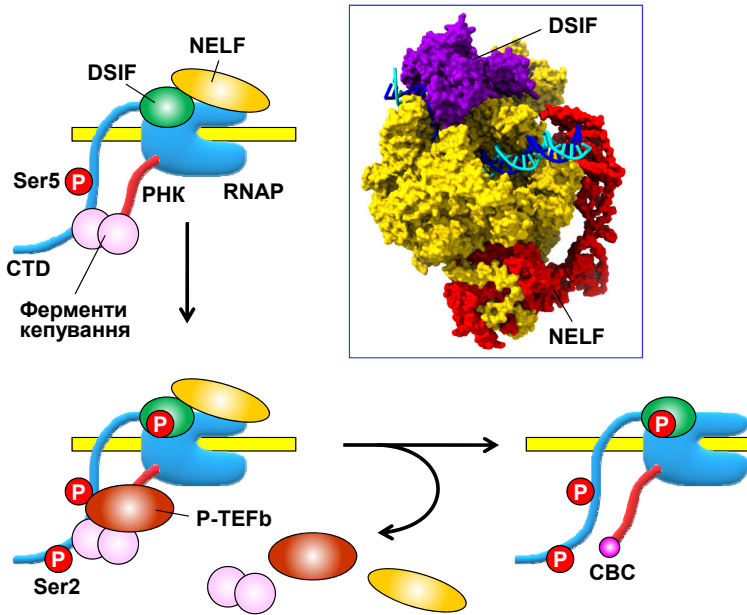
Функціонування CTD як платформи для збирання машинерії процесингу мРНК залежить від патерну фосфорилування залишків Ser, які входять до складу гептапептиду, що повторюється. Наприкінці ініціації транскрипції за рахунок кіназної активності TFIIK здійснюється масоване фосфорилування Ser5 у складі повторів, що зумовлює втрату зв'язку CTD із базальними факторами транскрипції та початок руху полімерази (розділ 5).

Під час руху, що почався, із полімеразою зв'язується білковий фактор DSIF (DRB-Sensitivity Inducing Factor, DRB – синтетичний інгібітор транскрипції, що спрацьовує тільки за наявності цього фактора). Цей двохсубодиничний білок є ні чим іншим, як фактором елонгації транскрипції Spt4-5, який згадувався у розділі 5 у зв'язку з механізмами проходження РНК-полімерази через нуклеосоми. На початку транскрипції він рекрутує

NELF (Negative Elongation Factor). Останній складається з п'яти субодиниць і взаємодіє з РНК-полімеразним комплексом, зумовлюючи зупинку його руху (рис. 7.3).

Під час цієї паузи (верхній лівий кут на рис. 7.3) із CTD і 5'-кінцем РНК зв'язуються ферменти кепування, які рекрутують Р-TEFb (Positive Transcription Elongation Factor b) – гетеродимер, одна із субодиниць якого має кіназну активність. Відбувається фосфорилування Ser2 у складі CTD, а також фосфорилування Spt5 у складі DSIF, що спричиняє дисоціацію NELF. Після цього РНК-полімераза здатна рухатись далі разом із фосфорильованим DSIF (Spt4-5), який саме у такій формі полегшує подолання нуклеосомних бар'єрів.

Пауза в роботі РНК-полімерази одразу за стартом транскрипції є важливим моментом регуляції експресії гена. Часто полімераза зупиняється під час створення кепа на невизначений термін. Особливо це властиво генам, що мають дуже швидко відповідати на зміну зовнішніх умов (напр. генам теплового шоку, продукти яких детально обговорюватимуться у розділі 8). Оскільки процес ініціації транскрипції дуже неефективний, то ініціація на промоторах таких генів здійснюється заздалегідь – РНК-полімераза починає працювати, але майже одразу виявляється "заарештованою" під час створення кепа, чекаючи на відповідні зовнішні сигнали (для генів теплового шоку – підвищення температури). Отже, зв'язування Р-TEFb, який знімає блокаду РНК-полімерази, залежить від наявності певних транскрипційних факторів і змін у характері модифікацій гістонів у складі найближчої нуклеосоми (ідеться про позиційовану нуклеосому +1, яка згадувалась у попередньому розділі, – саме вона міститься одразу за стартовою точкою транскрипції).



**Рис. 7.3.** Механізм утворення кепа в координації із транскрипцією.  
На вставці: комплекс РНК-полімерази II з факторами DSIF і NELF (6GML)

Після зняття блокади РНК-полімераза продовжує рух, а сформований кеп одразу зв'язується із двома білками, що формують **CBC** (**C**ap **B**inding **C**omplex). Білки CBC залишаються на кепах до моменту транспортування мРНК до цитоплазми, де замінюються на цитоплазматичні фактори.

Функціональне значення кепа є багатоплановим:

- *захист 5'-кінця від деградації*; особливість зв'язку між першими двома нуклеотидами робить цей зв'язок непомітним для екзонуклеаз;
- *участь у інших реакціях процесингу* – CBC стимулює сплайсинг першого інтрона й поліаденілювання 3'-кінця (див. нижче);

- *транспортування мРНК у цитоплазму* здійснюється завдяки взаємодії СВС із ядерною порою: саме своїм 5'-кінцем мРНК виштовхується в цитоплазматичний простір;
- *ініціація трансляції*: саме кеп є первинною точкою збирання рибосоми й інших елементів системи ініціації білкового синтезу (розділ 8).

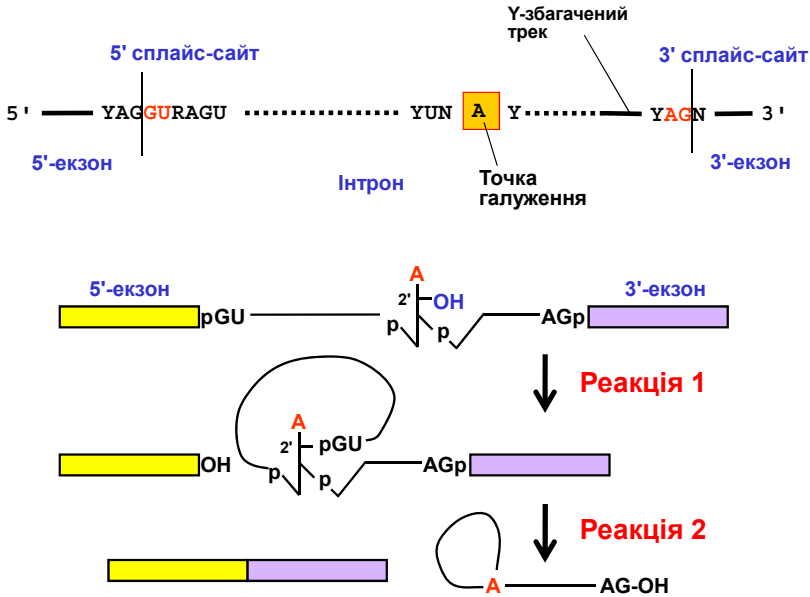
## Сплайсинг

Продовжуючи елонгацію транскрипції, рано чи пізно РНК-полімераза дістається границі першого інтрона. Узагальнену схему інтрона зображено на рис. 7.4. На кінцях переважної більшості інтронів розташовані стандартні динуклеотиди GU та AG, що містяться у складі певних консенсусних послідовностей. Усередині інтрона, ближче до його 3'-кінця, існує ще один консенсус, до складу якого обов'язково входить аденозин. Цей А є *точкою галуження* (branch point), яка виконує особливу функцію.

Сплайсинг інтрона полягає у двох послідовних хімічних реакціях *трансестерифікації* (заміни одного фосфодієфірного зв'язку на інший), які також показано на рис. 7.4.

1. Аденозин у точці галуження має підвищену реакційну здатність (унаслідок особливостей просторової структури інтрона в цій ділянці, див. нижче). 2'-ОН-група його рибози здійснює нуклеофільну атаку на фосфат у 5'-сплайс-сайті на межі між 5'-екзоном та інтроном. У результаті зв'язок між цим фосфатом і 3'-кінцевим нуклеотидом екзона замінюється на зв'язок між фосфатом і 2'-ОН-групою аденінового нуклеотиду – у 5'-кінцевій частині інтрона утворюється *ласо* (lariat).

2. ОН-група, що залишилася на 3'-кінці першого екзона, атакує фосфодієфірний зв'язок у 3'-сплайс-сайті. Цей зв'язок розривається, замінюючись на зв'язок між двома екзонами.



**Рис. 7.4.** Узагальнена схема нуклеотидної послідовності інтрона і двох стадій сплайсингу

Оскільки кількість фосфодієфірних зв'язків не змінюється, то реакція трансестерифікації є ізоенергетичною і не потребує АТР як джерела енергії. Однак, безумовно, обидві реакції сплайсингу потребують каталізу. На відміну від інших біохімічних реакцій, що ми розглядали до сих пір, **каталізаторами сплайсингу не є білкові ферменти.**

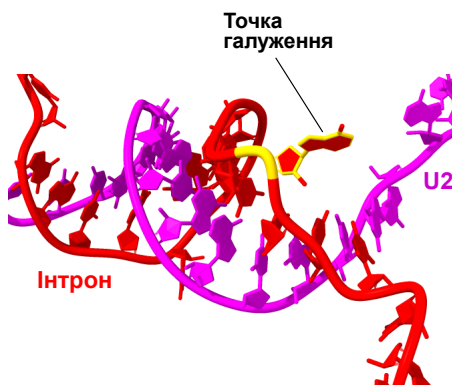
### Сплайсосома: механізм сплайсингу

Зрозуміло, що для проходження обох реакцій сплайсингу відповідні хімічні групи на кінцях інтрона й у точці галуження мають бути наближені одна до одної. Реакції сплайсингу здійснюються у складі спеціальної мультимолекулярної структури – **сплайсосоми (spliceosome)**, яка зв'язана із СТД і утворюється

на кожному інтроні за участю самої пре-мРНК, білків і особливих молекул мяРНК – *маленьких ядерних РНК* (small nuclear RNA, snRNA).

Збагачені на урацил маленькі ядерні РНК синтезуються РНК-полімеразою II (деякі – РНК-полімеразою III) на відповідних генах, згрупованих у кластери. Саме ці молекули **відіграють ключову роль у визначенні просторової структури, формуванні й функціонуванні сплайсосоми**. У сплайсингу беруть участь п'ять основних типів маленьких ядерних РНК завдовжки 100–200 нуклеотидів: U1, U2, U4, U5 та U6.

Маленькі ядерні РНК, які існують у клітині у вигляді комплексів зі специфічними білками, мають певну просторову структуру за рахунок утворення комплементарних подвійних спіралей – шпильок. Завдяки комплементарному спарюванню з консенсусними послідовностями інтрона відбувається їхня взаємодія із пре-мРНК, також вони взаємодіють між собою. Зокрема, мяРНК U2 впізнає консенсус у зоні галуження, причому згаданий **аденозин залишається неспареним** і випетлюється між сусідніми парами основ (рис. 7.5). Саме ця обставина визначає його особливий конформаційно-напружений стан і, як наслідок, підвищену реакційну здатність.

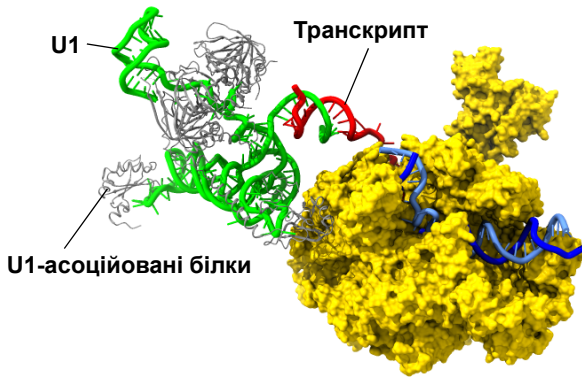


**Рис. 7.5.** Фрагмент структури комплексу маленької ядерної РНК U2 і зони галуження інтрона (5258)

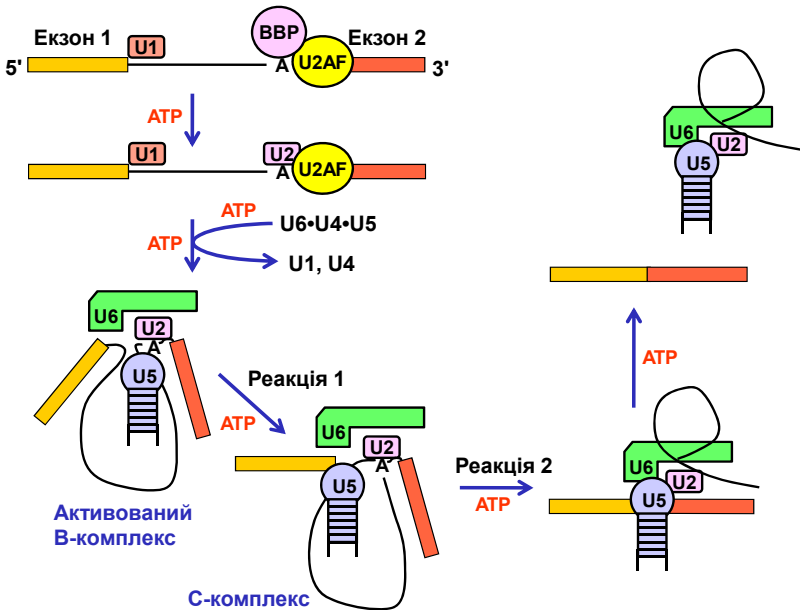
Оскільки подвійні спіралі РНК у сплайсосомі короткі, то вони потребують додаткової стабілізації. Крім того, під час збирання і функціонування сплайсосоми певні спіралі мають бути зруйновані й замінені на інші. Обидві операції забезпечуються численними білками сплайсосоми, які виконують важливі, але допоміжні функції. Ці функції переважно зводяться до:

- *розкручування подвійних спіралей РНК* АТР-залежними РНК-геліказами;
- *стабілізації* подвійних спіралей і загальної просторової структури сплайсосоми білками зі специфічною спорідненістю до РНК;
- *регуляції сплайсингу* – блокування чи підсилення ефективності збирання сплайсосоми на інтроні білками-регуляторами сплайсингу (див. нижче).

При появі у складі транскрипту 5'-сплайс-сайта границя інтрона упізнається маленькою ядерною РНК U1 (рис. 7.6). Загальний сценарій наступних подій схематично зображено на рис. 7.7. Після зв'язування з U1 у ході синтезу пре-мРНК зона галуження упізнається специфічним білком ВВР (Branch point Binding Protein, позначається також як SF1 у ссавців). 3'-Кінцева зона інтрона та Y-збагачений трек перед нею упізнаються допоміжним білком U2AF (U2 Auxiliary Factor), який далі сприяє заміні ВВР на мРНК U2. Зв'язування U2 (яке, власне, і визначає точку галуження, див. рис. 7.5) потребує АТР-залежного руйнування певних подвійних спіралей у складі цієї молекули. На подальших стадіях збирання і перебудови сплайсосоми (рис. 7.7) також відбувається АТР-залежне руйнування частини подвійних спіралей і заміна їх іншими.



**Рис. 7.6.** Комплекс РНК-полімерази II і маленької ядерної РНК U1, що взаємодіє з певними білками і 5'-сплайс-сайтом у складі транскрипту (7B0Y)



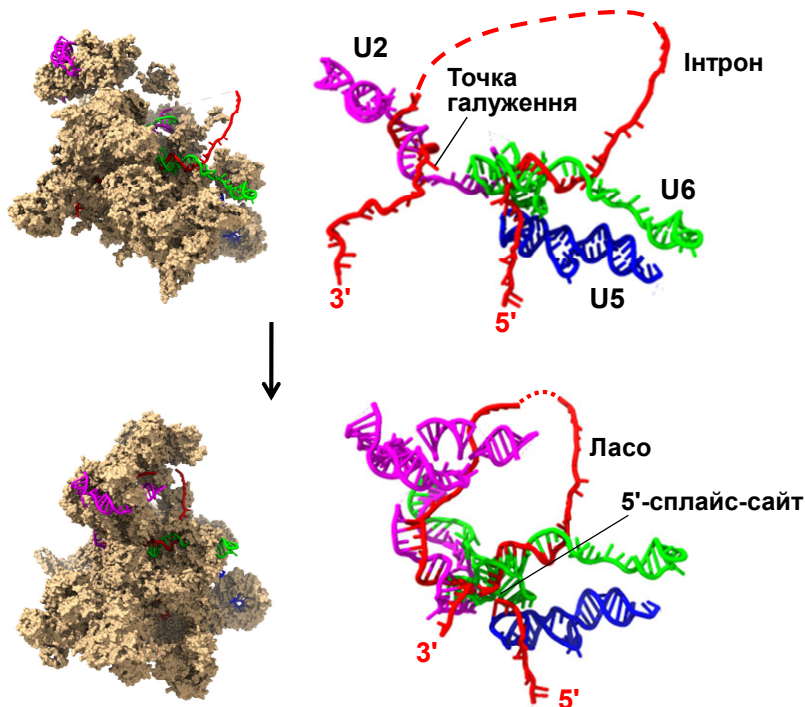
**Рис. 7.7.** Схема формування і функціонування сплайсосоми

Після зв'язування U2 з інтроном взаємодіє потрійний комплекс маленьких ядерних РНК U4-U5-U6, унаслідок чого утворюється сплайсосома у вигляді *B-комплексу* (невідображеного на рис. 7.7) маленьких ядерних РНК, інтрона і білків. На завершальному етапі із цього комплексу звільняється U4, а U6 витісняє U1, взаємодіючи із 5'-кінцевою зоною інтрона – формується *активований B-комплекс* (рис. 7.7, 7.8), готовий для здійснення першої реакції сплайсингу. У його складі U6 взаємодіє з U2, сприяючи наближенню 5'-сплайс-сайта до точки галуження. Додатково структура стабілізується маленькою ядерною РНК U5, яка взаємодіє із 3'-кінцем першого (5'-кінцевого) екзона та іншими елементами сплайсосоми.

У складі активованого B-комплексу здійснюється структурна перебудова сплайсосоми, результатом якої є безпосереднє наближення 5'-сплайс-сайта до точки галуження. Виникають умови для першої реакції трансестерифікації з утворенням ласо (див. також рис. 7.4). Хімічна перебудова інтрона викликає нову перебудову просторової структури сплайсосоми – утворення C-комплексу (рис. 7.7, 7.8). Далі за рахунок одночасної взаємодії з U5 наближуються один до одного кінці екзонів, що створює умови для другої трансестерифікації. На останньому етапі від двох уже з'єднаних екзонів АТР-залежним шляхом (порушення комплементарних взаємодій) видаляється комплекс інтрона з маленькими ядерними РНК.

Отже, сплайсосома працює як АТР-залежна молекулярна машина. Структурні перебудови машини забезпечують перегрупування елементів – субстратів реакцій сплайсингу. ***Каталіз обох реакцій сплайсингу здійснюється молекулами РНК.*** Механізм каталізу є таким самим, як для білкових ферментів (див. розділ 2). Наприклад, для першої реакції активний центр формується зоною галуження інтрона й маленькими ядерними РНК U2 та U6. Просторова структура активного центра жорстко утримує субстрати – 5'-сплайс-сайт і точку галуження – у певній взаємній орієнтації та забезпечує підвищену реакційну здатність аденінового нуклеотиду, який можна розглядати одночасно і як субстрат у збудженому проміжному стані, і як компонент активного центра. У хлоропластах сплайсинг відбувається за механі-

змом *самосплайсингу* (self-splicing), тобто без участі білків і маленьких РНК: пре-мРНК сама набуває просторової структури, яка має каталітичну активність щодо власного сплайсингу.



**Рис. 7.8.** Загальний вигляд сплайсосоми на стадіях (зверху донизу) активованого В-комплексу (6Z58) і С-комплексу (5XJC). Праворуч – ті самі структури (у такій самій орієнтації, але у збільшеному вигляді), з яких видалено білки для візуалізації нуклеїнових кислот

За аналогією з білковими *ензимами*, молекули РНК, які мають каталітичну активність, називають *рибозимами*. Одним із таких рибозимів є сплайсома. Зазвичай каталітична активність рибозиму (як і у випадку сплайсосоми) залежить від білків, які виконують допоміжні функції, стабілізуючи структуру активного центра.

Загальноприйнятий погляд полягає у тому, що на ранніх етапах добіологічної еволюції головними (або єдиними) біологічними макромолекулами були молекули РНК (так званий *РНК-світ*), оскільки це єдиний тип макромолекул, що можуть одночасно виконувати функції носіїв спадкової інформації і каталізаторів. Згодом більш стабільні молекули ДНК перебрали на себе функцію носіїв інформації, а різноманітніші за просторовою структурою білки – функцію каталізаторів. Однак у деяких важливих випадках (і сплайсинг – не єдиний приклад) рибозими виявились еволюційно консервативними.

Описаний механізм сплайсингу реалізується на ~99 % інтронів, які відповідають загальній схемі на рис. 7.4 (*мажорні інтрони*, або *інтрони U2-типу*). Решта інтронів вирізаються у подібний спосіб, але мають трохи інші консенсусні послідовності на границях і в точці галуження, хоча там також наявний особливий аденозин. Це *мінорні інтрони*, або *інтрони U12-типу*. Замість маленької ядерної РНК U2 у формуванні сплайсосоми на таких інтронах беруть участь маленька ядерна РНК U12 (визначає точку галуження) і маленька ядерна РНК U11.

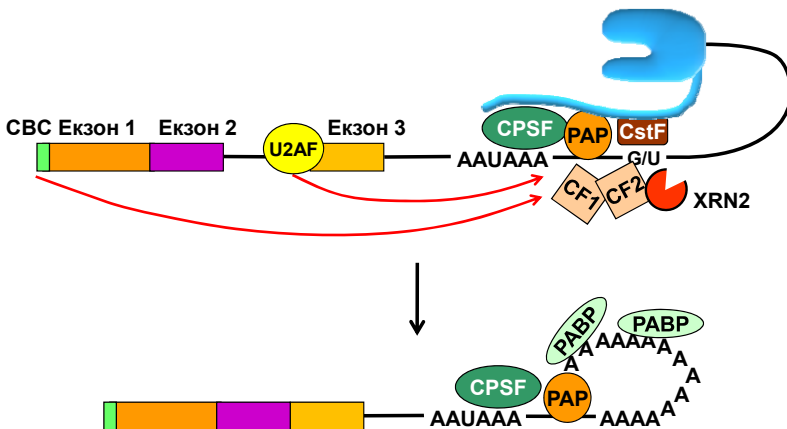
Сплайсосоми утворюються на інтронах пре-мРНК (і на STD РНК-полімерази) послідовно під час транскрипції. Майже одразу після синтезу сплайс-сайтів вони упізнаються відповідними елементами системи сплайсингу. При цьому СВС, що зв'язаний із кепом, підсилює ефективність зв'язування U1 у першому 5'-сплайс-сайті, імовірно шляхом прямої взаємодії; вплив СВС на сплайсинг наступних інтронів послаблюється.

Швидкість збирання сплайсосом визначається швидкістю транскрипції. З іншого боку, зі сплайсосомою взаємодіють фактори елонгації транскрипції – наявність сплайс-сайта сприяє прискоренню руху полімерази. Такий процес синтезу РНК та її сплайсингу триває до того моменту, поки у складі пре-РНК не з'являється специфічна нуклеотидна послідовність – сигнал термінації.

## Поліаденілювання мРНК і термінація транскрипції

Сигнал термінації транскрипції та поліаденілювання 3'-кінця мРНК – *polyA-сигнал* (PAS, polyadenylation signal) – складається з консенсусної послідовності AAUAAA (рис. 7.9), часто асоційованої із U-збагаченою ділянкою вище і U- або G/U-збагаченою ділянкою нижче (на ~20 нуклеотидів) від неї. Консенсус AAUAAA упізнається гетеротетрамерним білком CPSF (Cleavage-Polyadenylation Specificity Factor). CPSF є одним із компонентів комплексу, що забезпечує розрізання пре-мРНК і поліаденілювання 3'-кінця, який виникає внаслідок цього (комплекс CPA, Cleavage and PolyAdenylation). Інші важливі компоненти цього комплексу – фактор, що стимулює розрізання, CstF (Cleavage stimulation Factor), і два фактори розрізання (Cleavage Factors) CF 1 і 2. До комплексу CPA долучаються також інші білки, серед яких polyA-полімераза (PAP) і 5'-екзонуклеаза XRN2 (Rat1 у дріжджів). Збирання CPA-комплексу стимулюється зв'язаним із кепом CBC, а також сплайсосою на останньому інтроні. Сплайсинг останнього інтрона й розрізання / поліаденілювання РНК здійснюються одночасно й стимулюють одне одного. Зокрема, білок U2AF, який міститься на останньому інтроні, взаємодіє з PAP.

У процесі впізнання polyA-сигналу і збирання на ньому CPA-комплексу РНК-полімераза зазвичай продовжує синтез РНК за сигналом (до 1000 нуклеотидів). У складі сформованого CPA-комплексу одна із субодиниць CPSF, що має ендонуклеазну активність, здійснює розрізання пре-мРНК у ділянці між консенсусом AAUAAA та U-(G/U)-збагаченою послідовністю. До 3'-кінця, що виник унаслідок розрізання, поліА-полімераза PAP (за стимулюючої дії CPSF) у безматричний спосіб приєднує один за одним від ~50 до 200 аденінових нуклеотидів. З polyA-хвостом, що зростає, одразу зв'язується специфічний білок PABP (PolyA Binding Protein), який підвищує процесивність PAP (рис. 7.9). Слід зауважити, що деякі мРНК у зоні перед поліА-сигналом (3'-зоні, що не транлюється, див. нижче) мають сигнальні послідовності, які зумовлюють додаткове поліаденілювання уже в цитоплазмі, що підвищує час життя мРНК і ефективність ініціації трансляції (розділ 8).

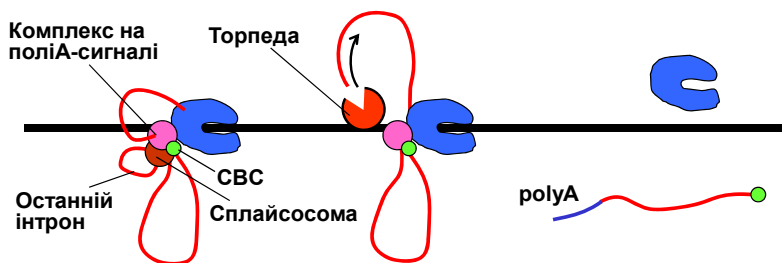


**Рис. 7.9.** Упізнання polyA-сигналу і поліаденілювання мРНК.  
 Фактори упізнання поліА-сигналу й розрізання мРНК:  
 CPSF – Cleavage-Polyadenylation Specificity Factor,  
 CstF – Cleavage stimulation Factor, CF1/2 – Cleavage Factors,  
 PAP – поліА-полімераза (PolyA Polymerase),  
 XRN2 – 5'-екзонуклеаза (торпеда)

Розрізання / поліаденілювання мРНК є тригером термінації транскрипції (рис. 7.10). Хоча детальні молекулярні механізми цього процесу в еукаріотів залишаються недостатньо вивченими, але ясно, що важливу роль фактора термінації відіграє згадана вище 5'-екзонуклеаза XRN2 (яку називають *торпедою*). У місці розрізу, крім 3'-кінця, що піддається поліаденілюванню, вочевидь виникає і 5'-кінець, з якого розпочинається ділянка пре-мРНК, що закінчується на РНК-полімеразному комплексі. Саме із цього 5'-кінця XRN2 розпочинає деградацію РНК (*котранскрипційну деградацію*). Урешті-решт ця деградація доходить до 3'-кінцевої ділянки РНК у складі гібрида ДНК-РНК – XRN2 "торпедує" РНК-полімеразу, що сприяє руйнуванню гібрида та її дисоціації.

Для того, щоб торпеда ефективно спрацювала, вона повинна мати можливість наздогнати полімеразу, тобто необхідно зупинити або загальмувати останню. Одним із факторів, від якого залежить таке гальмування, є зв'язаний із РНК-полімеразою Spt4-5 (DSIF). Нагадаємо, що у дефосфорильованій формі цей

фактор зумовлює зупинку РНК-полімерази, необхідну для утворення кепа. Після фосфорилування (рис. 7.3) він сприяє елонгації транскрипції, допомагаючи долати нуклеосомні бар'єри (розділ 5). У складі СРА-комплексу наявна фосфатаза, яка знову забезпечує дефосфорилування Spt4-5, що ускладнює рух полімерази після проходження поліА-сигналу. Термінація транскрипції іноді потребує також інших регуляторних елементів послідовності. Одні з них можуть знаходитись у межах останнього інтрона та стимулювати / блокувати впізнання polyA-сигналу, інші – нижче від polyA-сигналу й зумовлювати паузи в роботі полімерази, які сприяють термінації.



**Рис. 7.10.** Синхронізація поліаденілювання і термінації транскрипції

Іноді, незважаючи на високу процесивність РНК-полімерази, термінація може здійснюватись незалежно від поліА-сигналу. Особливо часто це відбувається перед або під час паузи, що необхідна для створення кепа. Іноді полімераза зупиняється на тривалий час за різних причин (напр., зустрічаючи надкомпактизовану гетерохроматинову ділянку), тоді вона піддається протеолітичній деградації і знімається з ДНК лише у такий радикальний спосіб.

Майже єдиним винятком серед еукаріотичних мРНК є гістонові мРНК, які не піддаються сплайсингу (гістонові гени не містять інтронів) і поліаденілюванню 3'-кінця. Термінація транскрипції гістонових генів залежить від двох сигнальних елементів послідовності РНК: один утворює шпильку (як у прокариотів, див. розділ 5), інший спарюється з маленькою ядерною РНК U7;

між цими елементами індукується розріз, що здійснюється тим самим компонентом CPSF, що й на інших білкових генах. Аналогічно відбувається термінація транскрипції маленьких ядерних РНК, які також не піддаються поліаденілюванню.

Узагальнену схему будови зрілої еукаріотичної мРНК, яка звільняється з полімеразного комплексу, транспортується в цитоплазму й використовується як матриця для білкового синтезу, зображено на рис. 7.11. Між кепом і початком кодувальної ділянки (стартовим кодоном, найчастіше AUG) розташована 5'-кінцева зона, що не транслюється (5' UTR – UnTranslated Region). За кодувальною ділянкою, що закінчується одним зі стоп-кодонів, і перед polyA-послідовністю розташована 3'-кінцева зона, що не піддається трансляції. Обидві зони, що не транслюються, містять важливі елементи послідовності, які використовуються для регуляції білкового синтезу (розділ 8).

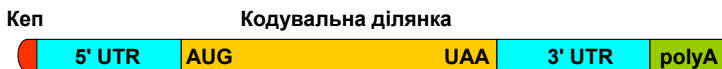


Рис. 7.11. Схема будови мРНК

Час півжиття молекул мРНК (нижче наведено цифри для фібробластів миші) варіює від 30 хв до 20 год (медіанне значення – 9 год). Для порівняння: час півжиття білків (для приблизно 5 тис. білків) становить від 30 хв до 200 год (медіана – 46 год). Цікаво, що обидва часові показники практично не корелюють між собою – час життя мРНК і білків регулюється окремими незалежними механізмами. Натомість спостерігається кореляція між кількістю копій мРНК і відповідних білків у клітині. Кількість копій мРНК варіює від 1 до 1 тис. (медіана – 17), білків – від 10 до  $10^7$  (медіана – 16 тис). Отже, перша стадія – транскрипція й асоційований із нею процесинг – є хоча й не єдиним, але визначальним етапом експресії генетичної інформації.

## Редагування мРНК

Процес дозрівання мРНК може включати також різноманітні модифікації, які об'єднують під назвою *редагування мРНК* (mRNA editing). Редагування, яке найчастіше відбувається під час сплайсингу або одразу після нього, не є обов'язковим для всіх мРНК процесом, хоча й зустрічається практично в усіх таксонах.

У мітохондріях простіших більшість мРНК піддається редагуванню за РНК-залежним механізмом. *Гідова РНК* (guide RNA) спарюється з комплементарною ділянкою мРНК та ініціює збирання мультибілкової едитосоми (editosome), компоненти якої забезпечують видалення чи інсерцію уридину в певне місце ділянки мРНК.

У мітохондріях і хлоропластах рослин часто відбувається дезамінування цитидину із заміною його на уридин, що змінює кодувальні властивості багатьох мРНК.

Дезамінування аденіну в певному місці ланцюга з перетворенням його на інозин (неканонічна пуринова основа, див. розділ 8) відбувається в мРНК різних субодиноць Glu-залежного  $\text{Ca}^{2+}$ -каналу нервових клітин ссавців. Наслідком нуклеотидної заміни є амінокислотні заміни у складі білка (усередині трансмембранної пори); комбінація різних (редагованих і нередагованих) субодиноць створює розмаїття каналів щодо їхньої чутливості до хімічного сигналу (амінокислоти Glu) і швидкості потоку  $\text{Ca}^{2+}$  крізь мембрану. Редагування такого типу залежить від активності спеціальних ферментів (аденозиндеаміназ). Вважають, що вони зв'язуються із дволанцюговими шпильками, які тимчасово формуються в мРНК під час синтезу, де впізнають певні елементи послідовності пар основ.

Іншим прикладом є перетворення шляхом дезамінування певного цитидину на уридин у мРНК аполіпобілка В (ApoB, apolipoprotein B) печінки ссавців. Заміна викликає перетворення глутамінового кодону на стоп-кодон. У результаті білок може існувати у двох формах із різною довжиною поліпептидного ланцюга. Обидві форми використовуються для транспортування тригліцеридів і холестерину через кров, причому довша форма пов'язана з підвищенням ризику атеросклерозу. Редагування

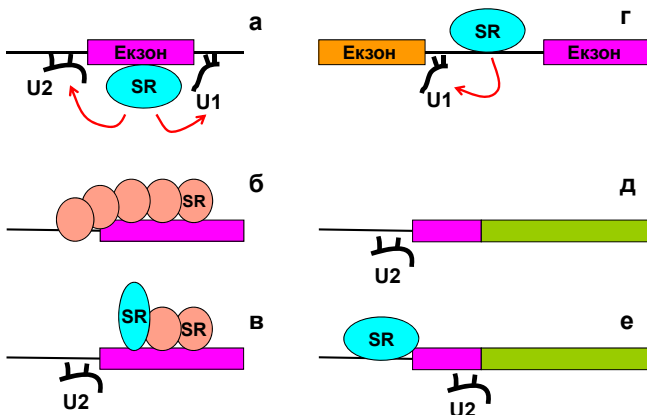
мРНК АроВ, яке відбувається паралельно з поліаденілюванням, залежить від певної послідовності мРНК, що упізнається комплексом білків, серед яких – специфічна деаміназа.

## Альтернативний процесинг

Пре-мРНК, що синтезується під час транскрипції, може піддаватися сплайсингу й поліаденілюванню різними альтернативними шляхами (зазвичай тканиноспецифічними): кілька екзонів можуть вирізатися із транскрипту; транскрипт може обрізатися і піддаватися поліаденілюванню за рахунок використання polyA-сигналу всередині одного з інтронів. У результаті у клітинах різних типів утворюються різні молекули мРНК, що містять різні набори екзонів і, відповідно, кодують різні білки. Багатоконпонентність (і необхідність кооперації між компонентами) сплайсосоми й системи розрізання / поліаденілювання дозволяє здійснювати тонку регуляцію утворення мРНК певного типу за рахунок зміни транскрипційної активності генів, що кодують ті чи інші компоненти машинерії процесингу, зміни концентрацій компонентів, їхньої хімічної модифікації.

Ключова роль у визначенні шляхів сплайсингу належить білкам, що відносять до групи SR (серін-аргінин-збагачені, SR-rich). Доволі часто білок-регулятор специфічно зв'язується з певною послідовністю нуклеотидів усередині ексона та стимулює впізнання сплайс-сайтів по обидва боки від нього (рис. 7.12, *a*). Зрозуміло, що за відсутності регулятора такий екзон буде вирізаний разом з інтронами, що його фланкують.

У випадку негативної регуляції зв'язування регулятора утруднює впізнання сплайс-сайтів. Наприклад, взаємодія з регулятором ініціюється всередині ексона, молекули білка кооперативно зв'язуються поруч одна з одною і в решті-решт блокують 3'-сплайс-сайт (рис. 7.12, *b*). Замість заблокованого сприймається наступний такий сайт, і екзон вирізається з молекули разом із двома інтронами. За наявності іншого регулятора, який також упізнає специфічну послідовність у екзоні, процес зростання білкового комплексу зупиняється і 3'-сплайс-сайт стає вільним для впізнання елементами сплайсосоми (рис. 7.12, *в*).



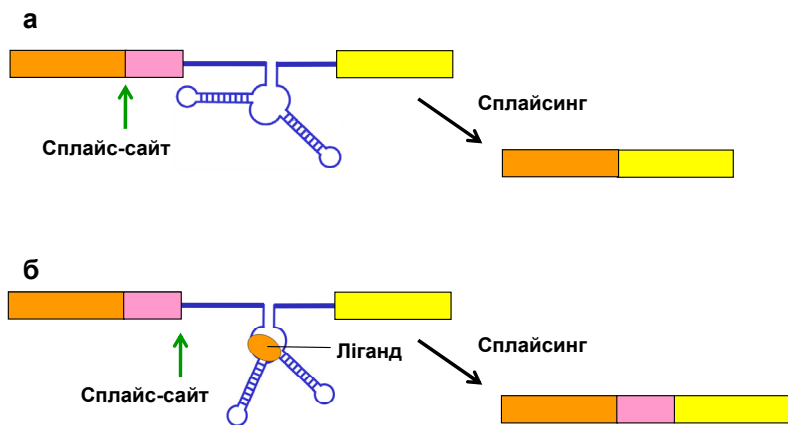
**Рис. 7.12.** Кілька варіантів участі сплайсинг-регуляторів, що зв'язуються всередині екзона (а, б, в) або інтрона (г, д, е), у визначенні шляху сплайсингу

Сплайсинг-регулятори можуть також мати сайти зв'язування в інтронах, впливаючи на ефективність збирання сплайсосоми (рис. 7.12, з). Регулятор може блокувати, наприклад, 3'-сплайс-сайт. Якщо така блокада відсутня, то сайт розпізнається (рис. 7.12, д), за наявності регулятора сприймається інший 3'-сплайс-сайт усередині екзона, а у складі мРНК залишається скорочений екзон (рис. 7.12, е).

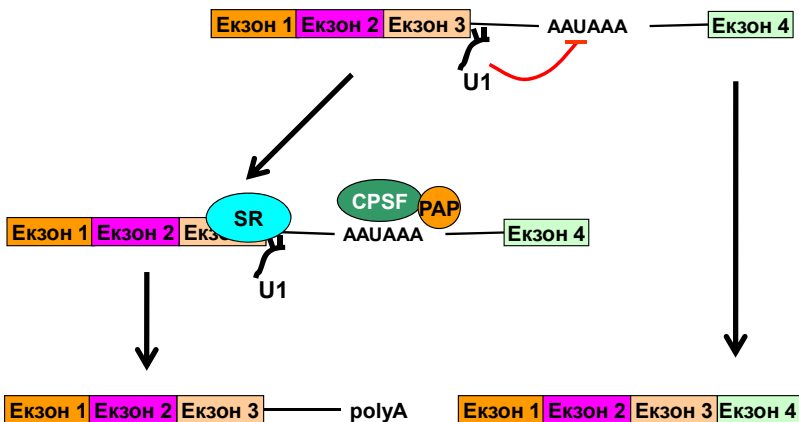
Іноді на впізнання сплайс-сайтів впливає просторова структура пре-мРНК: дволанцюгові шпильки можуть або маскувати сплайс-сайт (тоді його впізнання можливе за наявності білка, що руйнує шпильку), або навпаки – сприяти його впізнанню. Крім цього, до регуляції сплайсингу залучаються РНК-перемикачі (розділ 6) усередині інтронів. У такому випадку блокада сплайс-сайта, що створює елемент просторової структури РНК, знімається при структурній перебудові, що індукується певним лігандом (рис. 7.13).

Вибір шляху утворення кінцевого мРНК-продукту часто залежить також від альтернативного вибору роуА-сигналів, які можуть існувати не лише нижче останнього екзона, а й усередині інтронів. Приклад такого типу зображено на рис. 7.14. Узага-

лі, якщо polyA-сигнал міститься всередині інтрона, то його впізнання блокується 5'-сплайс-сайтом – білками, асоційованими з маленькою ядерною РНК U1 (polyA-сигнал, що розташований після термінального екзона, не блокується, оскільки такий 5'-сплайс-сайт відсутній). У результаті такої блокади polyA-сигнал не упізнається і вирізається разом з інтроном. Утім, якщо діє специфічний сплайсинг-регулятор, який заважає впізнанню 5'-сплайс-сайта або взаємодії U1-білків з елементами системи впізнання polyA-сигналу, то сигнал упізнається, а мРНК втрачає всі екзони, розташовані нижче від сигналу. Наприклад, сплайсинг мРНК важкого ланцюга імуноглобуліну IgM спрямовується першим шляхом (внутрішній polyA-сигнал ігнорується) у В-лімфоцитах, що розвиваються. Тоді синтезується мембранна форма білка, яка містить С-кінцевий домен, зв'язаний із мембраною. У плазматичних клітинах спрацьовує внутрішній polyA-сигнал (упізнання сигналу підсилюється також завдяки зростанню концентрації CstF), унаслідок чого синтезується секреторна форма білка.



**Рис. 7.13.** Участь РНК-перемикача у складі інтрона у виборі шляху сплайсингу. В одній структурній формі перемикач блокує один із сплайс-сайтів (а), в іншій, залежній від певного ліганду, робить цей сайт доступним (б)



**Рис. 7.14.** Приклад двох шляхів процесингу, вибір яких залежить від упізнання polyA-сигналу всередині інтрона

Поряд із концентрацією елементів машинерії процесингу та сплайсинг-регуляторів, доволі важливим фактором регуляції сплайсингу є швидкість транскрипції. У випадку довгого інтрона 3'-сплайс-сайт синтезується суттєво пізніше 5'-сплайс-сайта. За цей час упізнання 5'-сплайс-сайта вже завершилося, і перший 3'-сплайс-сайт упізнається як такий. Якщо за короткий час синтезуються кілька 3'-сплайс-сайтів, то виникає конкуренція між ними за відповідні елементи сплайсосоми. При цьому перший із сайтів буде вигравати за низької швидкості руху полімерази.

Ефективність упізнання сплайс-сайтів залежить від контексту нуклеотидної послідовності пре-мРНК: при швидкому русі полімерази "слабкі" сайти не встигають упізнаватись компонентами машинерії сплайсингу (РНК-полімераза "проскакує" через них), але упізнаються при уповільненні елонгації транскрипції. Швидкість елонгації залежить від тривалості пауз РНК-полімерази, які регулюються факторами елонгації транскрипції, а також від статусу фосфорилування STD й активності комплексів ремоделювання хроматину. Значною мірою шлях процесингу визначається вже на етапі ініціації транскрипції. При цьому до промотору (відповідно і до РНК-полімеразного комплексу) можуть рекрутуватися також регулятори сплайсингу.

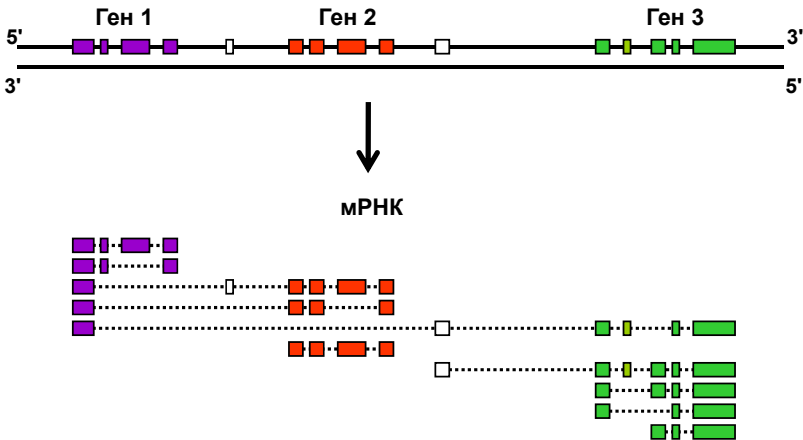
Тому на химерних генах із чужорідним промотором часто виявляються дефекти процесингу.

Нарешті, щільність і характер розміщення нуклеосом усередині частини гена, що транскрибується, може служити додатковим регуляторним фактором, який впливає на швидкість транскрипції і на тісно пов'язаний із транскрипцією процес сплайсингу мРНК. Нуклеосоми розподілені всередині еукаріотичних генів неоднорідно: екзони (незалежно від транскрипційної активності гена) характеризуються підвищеною імовірністю наявності нуклеосом порівняно з інтронами. Особливо збагачені на нуклеосомі екзони, що оточені довгими інтронами (середня довжина екзона ссавців передбачає, що всередині його міститься тільки одна позиційована нуклеосома). Підвищеною є присутність нуклеосом усередині екзонів, які мають слабкі сплайс-сайти (порівняно з екзонами, що відокремлені сильними сплайс-сайтами, які більш ефективно упізнаються машинерією сплайсингу). Псевдоекзони (ділянки всередині інтронів, які оточені сильними сплайс-сайтами, але не включаються до зрілих мРНК), навпаки, збіднені на нуклеосомі. Крім того, нуклеосомі в екзонах характеризуються специфічним набором хімічних модифікацій гістонів.

Отже, особливості позиціонування нуклеосом по обидва боки від сплас-сайта можуть служити додатковим маркером екзона, а зміни цього позиціонування під впливом комплексів ремоделювання хроматину – впливати на альтернативний сплайсинг. По-перше, нуклеосома всередині екзона може полегшувати впізнання сплайс-сайтів, які знаходяться в наближених у просторі лінкерах (зв'язування факторів сплайсингу у двох сусідніх сплайс-сайтах характеризується позитивною кооперативністю). По-друге, нуклеосомі можуть відігравати певну роль у рекрутуванні факторів сплайсингу, безпосередньо взаємодіючи з ними. Можливо, така взаємодія залежить від гістонових модифікацій. По-третє, розміщення нуклеосом має впливати на кінетику транскрипції, яка, своєю чергою, є важливим фактором альтернативного сплайсингу.

## Транс-сплайсинг

Для еукаріотичних генів характерне явище *транс-сплайсингу* – об'єднання у складі мРНК екзонів різних генів. Основою транс-сплайсингу є та обставина, що для багатьох генів існують не одна, а кілька альтернативних стартових точок транскрипції, зокрема й такі, що розташовані доволі далеко від гена й водночас є стартовими точками інших генів (рис. 7.15). У результаті транскрипція іноді здійснюється через кілька генів (ситуація дещо нагадує прокаріотичні оперони, розділ 5) і сплайсинг відбувається на рівні таких "об'єднаних" первинних транскриптів.



**Рис. 7.15.** Зона еукаріотичного геному: показано три гени й відповідні екзони (прямокутники) у складі змістовних ланцюгів. Прямокутники білого кольору важко віднести до конкретного гена. Унизу: кінцеві транскрипти, синтезовані у цій зоні

Описано також випадки транс-сплайсингу між пре-мРНК, які є продуктами різних одиниць транскрипції, іноді такими, що розташовані на різних хромосомах; тоді на 5'- і 3'-кінцях двох пре-мРНК міститься ніби розірваний інтрон. Об'єднання двох кінців у сплайсосомі (зрозуміло, що у такому випадку вона формується окремо від процесу транскрипції) спричиняє видалення цього інтрона та зшивання кінцевих екзонів. Найкраще це явище вивчено в нематод, але, імовірно, воно чимало поширене в еукаріотів узагалі.

## Контрольні запитання

1. З яких операцій складається процесинг мРНК? Де й коли він відбувається?
2. Опишіть загальну будову зрілої еукаріотичної мРНК.
3. Опишіть хімічну будову кепа. У чому полягає його функціональне значення?
4. Як і коли відбувається кепування 5'-кінця мРНК?
5. Опишіть послідовність етапів сплайсингу. Що таке точка галуження?
6. Що таке сплайсосома? З яких елементів вона складається, де й коли формується? Опишіть основні етапи утворення сплайсосоми.
7. Як здійснюється каталіз хімічних реакцій спласингу? Що таке рибозим?
8. З якою метою при збиранні й роботі сплайсосоми використовується енергія гідролізу АТР?
9. Як відбувається термінація транскрипції РНК-полімеразою II?
10. Охарактеризуйте основні механізми альтернативного сплайсингу.

## Рекомендована література

Akinyi, M.V. and Frilander, M.J. (2021) At the intersection of major and minor spliceosomes: crosstalk mechanisms and their impact on gene expression. *Front. Genet.* 12, 700744.

Baralle, F.E. and Giudice, J. (2017) Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451.

Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754.

Eaton, J.D. and West, S. (2020) Termination of transcription by RNA polymerase II: BOOM! *Trends Genet.* 36, 664–675.

Hariharan, A., Sun, S., Wipplinger, M. and Felley-Boscom E. (2020) RNA editing in mesothelioma: a look forward. *Open Biol.* 10, 200112.

Kachaev, Z.M., Lebedeva, L.A., Kozlov, E.N. and Shidlovskii, Y.V. (2020) Interplay of mRNA capping and transcription machineries. *Biosci. Rep.* 40, BSR20192825.

Kuehner, J.N., Pearson, E.L. and Moore, C. (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell. Biol.* 12, 283–294.

Lee, Y. and Rio, D.C. (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323.

Margaritis, T. and Holstege, F.C.P. (2008) Poised RNA polymerase II gives pause for thought. *Cell* 133, 581–584.

Matlin, A.J., Clark, F. and Smith, C.W.J. (2005) Understanding alternative splicing: towards a cellular code. *Nature Rev.* 6, 386–398.

Proudfoot, N.J. (2016) Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* 352, aad9926.

Ramanathan, A., Robb, G.B. and Chan, S.H. (2016) mRNA capping: biological functions and applications. *Nucleic Acids Res.* 44, 7511–7526.

Schwanhäusser, B., Busse, D., Li, N., et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342.

Shi, Y. (2017) Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell. Biol.* 18, 655–670.

Shi, Y. and Manley, J.L. (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29, 889–897.

Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326.

Teichert, I. (2020) Fungal RNA editing: who, when, and why? *Appl. Microbiol. Biotechnol.* 104, 5689–5695.

Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell. Biol.* 18, 18–30.

Tilgner, H. and Guigó, R. (2010) From chromatin to splicing. RNA-processing as a total artwork. *Epigenetics* 5, 180–184.

Ule, J. and Blencowe, B.J. (2019) Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* 76, 329–345.

Vos, S.M., Farnung, L., Urlaub, H. and Cramer, P. (2018) Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* 560, 601–606.

Zhang, X., Yan, C., Hang, J., et al. (2017) An atomic structure of the human spliceosome. *Cell* 169, 918–929.

Zhang, X., Yan, C., Zhan, X., et al. (2018) Structure of the human activated spliceosome in three conformational states. *Cell Res.* 28, 307–322.

Zorio, D.A.R. and Bentley, D.L. (2004) The link between mRNA processing and transcription: communication works both ways. *Exp. Cell Res.* 296, 91–97.

## Розділ 8

# СИНТЕЗ БІЛКІВ

So he started to climb out of the hole...  
and in a little while his nose was out... and then his ears...  
and then his front paws... and then his shoulders...  
A. Miln. Winnie the Pooh and all, all, all

Інформація щодо амінокислотної послідовності білка записана у вигляді послідовності нуклеотидів мРНК відповідно до генетичного коду (рис. 4.1). Зчитування цієї інформації та її перекладання на амінокислотний текст (*трансляція*, translation) має розпочинатися зі стартового кодону, де при *ініціації трансляції* відбувається остаточне збирання із двох субодиниць головного пристрою трансляції – *рибосоми* (ribosome). Обидві субодиниці є комплексами рибосомних РНК і білків. Рибосома сканує нуклеотидну послідовність мРНК, рухаючись уздовж неї кроками по три нуклеотида від 5'- до 3'-кінця під час *елонгації трансляції* (рис. 8.1) до стоп-кодону, де відбувається *термінація* процесу. Під час сканування рибосома працює і як каталізатор синтезу пептидного зв'язку між амінокислотами, і як декодувальний пристрій, забезпечуючи впізнання кодонів комплементарними триплетами (*антикодонами*) у складі *тРНК* (транспортна РНК, tRNA – transfer RNA). Певний антикодон відповідає амінокислоті певного типу, яку несе на собі тРНК. Саме тРНК є ключовою ланкою реалізації генетичного коду, оскільки вони доставляють амінокислоти до рибосоми в порядку, який відповідає послідовності кодонів. Отже, найважливішою підготовчою для білкового синтезу операцією є приєднання до тРНК амінокислоти, яка точно відповідає антикодону.

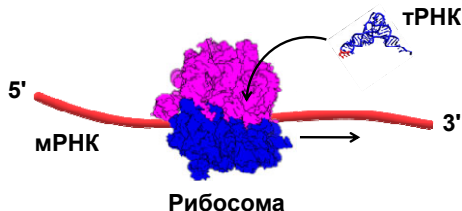


Рис. 8.1. Схема білкового синтезу

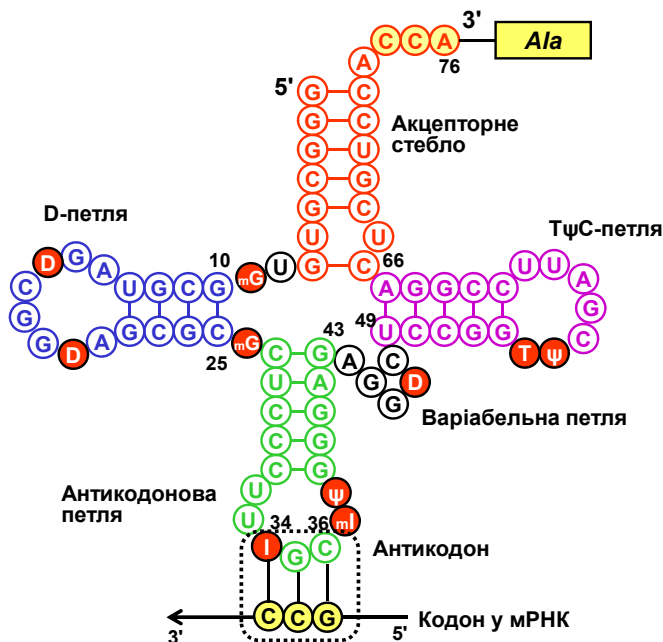
## Транспортні РНК

### Структура тРНК

Молекули тРНК містять 74–95 (найчастіше 76) нуклеотидів, серед яких зустрічаються неканонічні (утворюються внаслідок посттранскрипційних модифікацій) – тимідин, інозин (I), дигідроуридин (D), псевдоуридин ( $\psi$ ) тощо.

За рахунок взаємної комплементарності між ділянками ланцюга у складі молекули формуються дволанцюгові стебла та шпильки за єдиною для всіх тРНК схемою, що нагадує лист конюшини (рис. 8.2). Кінцеві фрагменти молекули об'єднуються у дволанцюгове стебло, причому чотири нуклеотиди на 3'-кінці залишаються неспареними. 3'-Кінцевий триплет ССА є стандартним для всіх тРНК, до рибози кінцевого аденозину ковалентно приєднується (акцептується) амінокислота, отже, стебло називають **акцепторним**. 5'-Кінцева частина акцепторного стебла переходить у шпильку з петлею, яка часто містить дигідроуридин (D-стебло і D-петля). У деяких тРНК D-петля може містити на один нуклеотид менше або на один-три нуклеотиди більше відносно типового розміру петлі, зображеного на рис. 8.2. D-стебло переходить у стебло з петлею, у складі якої розташований антикодонавий триплет (у позиції 34–36), – **антикодонаве стебло / антикодонава петля**. За антикодонавим стеблом розташована **варіабельна петля (V-петля)**. За розміром V-петлі всі тРНК можна поділити на два структурні класи: до першого належить більшість тРНК із короткою (3–5 нуклеотидів), до другого – кілька тРНК із довгою (до 16 додаткових нуклеотидів)

петлею. Нумерація нуклеотидів на рис. 8.2 є стандартною, додаткові нуклеотиди D- та V-петлі, якщо вони наявні, нумерують окремо. За варіабельною петлею міститься ТψС-стебло з петлею (у складі петлі часто зустрічається консенсус ТψС), яке переходить у 3'-кінцеву частину акцепторного стебла.



**Рис. 8.2.** Схема спарювання нуклеотидів у складі аланінової тРНК

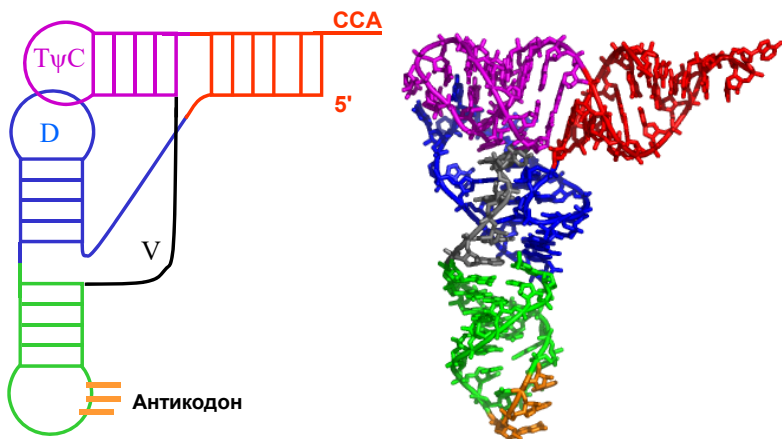
тРНК конкретного типу, яка відповідає певній амінокислоті, позначають верхнім індексом, наприклад тРНК<sup>Ala</sup>. Якщо до молекули тРНК приєднана амінокислота, то таку аміноацильовану тРНК позначають Ala-тРНК<sup>Ala</sup>. Загальне позначення для аміноацильованих тРНК – aa-тРНК (aa-tRNA – aminoacyl-tRNA).

Еукаріотичні гени тРНК (близько 500 активних генів тРНК у геномі людини, частина яких зібрана у кластери) транскрибуються РНК-полімеразою III. Продуктом транскрипції генів є довші молекули-попередники, деякі з них містять інтрон у межах майбутньої антикодонової петлі. Процесинг цих поперед-

ників з утворенням зрілих тРНК полягає у відщепленні певними нуклеазами зайвих фрагментів на кінцях, сплайсингу інтрона (здійснюється специфічними ендонуклеазами й лігазою), приєднанні до 3'-кінця стандартного триплету ССА (у складі бактеріальних тРНК-попередників цей триплет уже є), хімічній модифікації певних азотистих основ. Бактеріальні гени тРНК (87 у геномі *E. coli*, не містять інтронів) або транскрибуються окремо (як у еукаріотів), або є частинами оперонів, і в цьому випадку первинний транскрипт містить кілька майбутніх молекул тРНК. До того ж деякі бактеріальні гени тРНК містяться у складі оперона генів рибосомної РНК. В усіх випадках тРНК вирізаються з попередників нуклеазами, після чого піддаються хімічним модифікаціям. Одна з нуклеаз, яка бере участь у процесингу тРНК у про- та еукаріотів, – РНКаза Р – варта особливої уваги. Фермент складається із двох субодиниць, одна з яких є білковою, а інша – молекулою РНК зі складною просторовою структурою. Саме ця РНК-субодиниця має каталітичну активність, тобто РНКаза Р є рибозимом.

Схема спарювання "лист конюшини" (рис. 8.2) не дає уявлення щодо просторової структури тРНК. Насправді акцепторне та ТψС-стебло, переходячи одне в одне, утворюють єдину майже пряму подвійну спіраль (рис. 8.3), під приблизно прямим кутом до якої розташована інша подвійна спіраль, сформована D- і антикодоновим стеблом. ТψС- і D-петлі при цьому наближаються одна до одної, між ними реалізується комплементарне спарювання основ. У результаті молекула тРНК має Г-подібну (або L-подібну) форму із двома плечима різної довжини. На кінці коротшого плеча акцептується амінокислота (акцепторне плече), на кінці довшого розташований антикодон (антикодонове плече).

У складі дволанцюгових спіральних зон тРНК міститься ~55 % нуклеотидів, але ~90 % основ залучено до стекінг-взаємодій. Отже, молекула характеризується високою упорядкованістю, жорсткістю структури. Це стосується також петель, серед яких лише антикодонова не залучена до взаємодій з іншими елементами. Однак п'ять основ петлі (включаючи антикодон) утворюють доволі жорсткий стос. Додатково структура тРНК стабілізується іонами  $Mg^{2+}$ , лише у присутності яких молекула функціонально активна.



**Рис. 8.3.** Схема і просторова структура тРНК (тРНК<sup>Phe</sup>, 6TNA). Елементи структури пофарбовані однаково ліворуч і праворуч

Загальна кількість типів тРНК, які обслуговують процес білкового синтезу, близька до 40 (напр., усі гени тРНК людини можна поділити на 49 родин за властивостями антикодонів). Оскільки типів тРНК *більше, ніж амінокислот*, то одній амінокислоті може відповідати кілька тРНК (*ізоакцепторні тРНК*). Серед них є такі, що містять різні (але, зрозуміло, синонімічні) антикодони, – гетерокодонові. Є також гомокодонові тРНК, які можуть бути або продуктами різних генів і розрізнятися за послідовністю нуклеотидів, або продуктами одного гена, розрізняючись модифікаціями основ. Оскільки типів тРНК *менше, ніж кодонів*, то одна тРНК здатна впізнавати кілька синонімічних кодонів, що забезпечується неоднозначністю спарювання між першою позицією антикодону і третьою (за якою переважно розрізняються синонімічні кодони, див. розділ 4) – кодону. Наприклад, U та G здатні впізнавати по два нуклеотиди у третій позиції кодону, I (який доволі часто зустрічається в першій позиції антикодону) упізнає три нуклеотиди (табл. 8.1).

**Таблиця 8.1.** Відповідність між нуклеотидами в першій позиції антикодону і третій позиції кодону

Антикодон	Кодон
C	G
A	U
U	A, G
G	U, C
I	U, C, A

### **Аміноацилювання тРНК**

Порядок залучення амінокислот до поліпептидного ланцюга, що утворюється при білковому синтезі, залежить лише від взаємодій між нуклеїновими кислотами – кодоном і антикодоном; амінокислота, яку несе тРНК, жодним чином не розпізнається рибосомою. Отже, акцептування певної амінокислоти молекулою тРНК відповідного типу (і тільки *відповідного*) є одним із найважливіших підготовчих моментів білкового синтезу: від точності процесу акцептування залежить точність синтезу білка загалом.

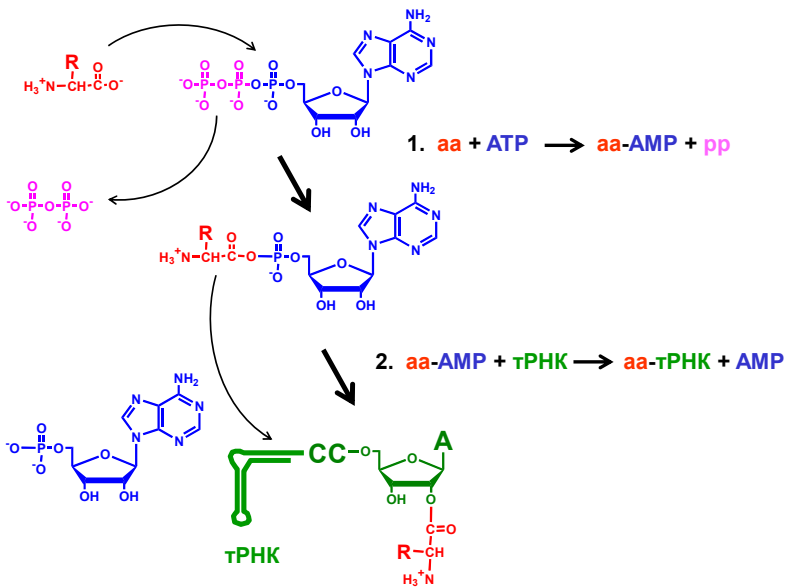
Процес приєднання амінокислот до тРНК каталізується *аміноацил-тРНК-синтетазами* (АРСаза, aaRS – aminoacyl-tRNA-Synthetase). Кожен із 20 типів (за кількістю амінокислот) цих ферментів каталізує дві хімічні реакції (рис. 8.4).

На першій стадії відбувається *активування амінокислоти* – її приєднання до аденозинмонофосфату АМР з утворенням *аміноацил-аденілату*, коли пірофосфат (pp) у складі АТР замінюється на амінокислоту (aa). Активування амінокислоти супроводжується зниженням вільної енергії, але при цьому значна частина вільної енергії, що вивільняється при відщепленні пірофосфату від АТР, заощаджується у формі аміноацил-аденілату – молекули, гідроліз якої також супроводжується великим енергетичним ефектом. Молекула АТР, що використовується на етапі активування амінокислоти, є єдиним джерелом енергії для майбутнього синтезу пептидного зв'язку на рибосомі.

Далі аміноацил-аденілат утворює проміжний комплекс з активним центром ферменту й ефективно атакує ОН-групу рибози

3'-кінцевого аденозину тРНК (3'- або 2'-ОН-групу, залежно від класу АРСази): відбувається *трансфер* (transfer) – перенесення амінокислоти на тРНК. Після дисоціації від АРСази у складі aa-тРНК має місце спонтанний обмін амінокислоти між двома ОН-групами рибози; при зв'язуванні з рибосомою амінокислота фіксується на 3'-ОН-групі.

Aa-тРНК – також макроергічна сполука: руйнування зв'язку між амінокислотою і тРНК енергетично вигідне й забезпечує потім утворення пептидного зв'язку на рибосомі. Отже, АРСази не тільки залучають амінокислоти до білкового синтезу, приєднуючи їх до тРНК, а й забезпечують заощадження вільної енергії, необхідної для приєднання амінокислоти до поліпептидного ланцюга.



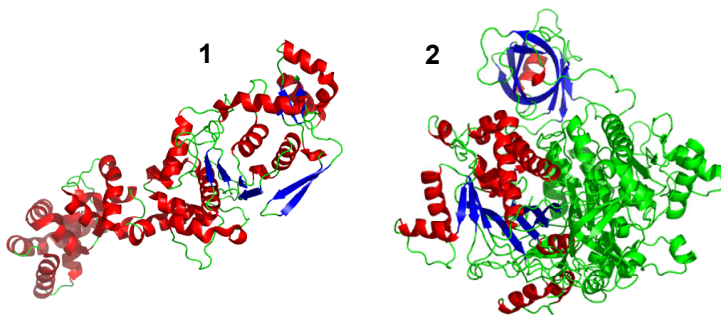
**Рис. 8.4.** Дві хімічні реакції, що каталізуються аміноацил-тРНК-синтетазами:  
1) активування амінокислоти; 2) приєднання амінокислоти до тРНК

Двадцять типів АРСаз, між якими немає майже нічого спільного на рівні первинної структури, можна поділити на два класи, по десять у кожному (табл. 8.2). Це завжди мультидоменні білки зі складною структурою (рис. 8.5), що зумовлено розмаїттям функцій – необхідністю специфічно зв'язати три субстрати й каталізувати дві хімічні реакції. Крім того, еукаріотичні АРСази різних типів взаємодіють між собою і з мембраною ендоплазматичного ретикулуму, утворюючи *кодосому*, що розташована поблизу від рибосом. Мультисубодиничні АРСази (гомодимери, гомо- або гетеротетрамери) містять два ідентичні набори активних центрів і сайтів зв'язування.

**Таблиця 8.2. Порівняльна характеристика двох класів АРСаз**

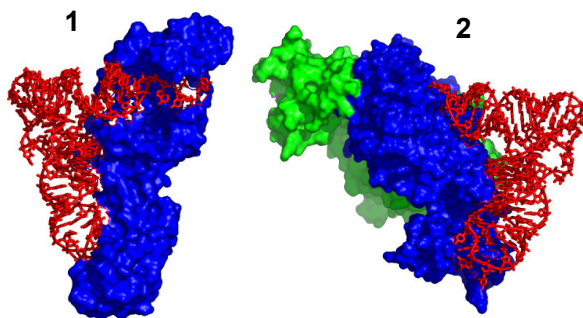
	<b>Клас 1</b>	<b>Клас 2</b>
Кількість субодиниць	1 (іноді 2)	2 або 4
Спільний структурний мотив, що оточує сайт зв'язування амінокислоти	Укладка Россмана	Паралельний $\beta$ -шар із 7 $\beta$ -ділянок
Амінокислоти, що акцептуються	Leu, Ile, Val Cys, Met, Glu, Gln, Arg, Tyr, Trp	His, Pro, Ser, Thr, Phe, Asp, Asn, Lys, Gly, Ala
Група рибози, що аміноацилюється	2'-ОН	3'-ОН

Центральна частина поліпептидного ланцюга АРСази формує сайт зв'язування амінокислоти, в оточенні якого реалізується спільний для ферментів одного класу структурний мотив: для першого класу – укладка Россмана (див. розділ 2), для другого – приблизно плоский паралельний  $\beta$ -шар із семи  $\beta$ -ділянок (рис. 8.5). Як видно з табл. 8.2, амінокислоти з подібними властивостями часто є субстратами АРСаз різних класів залежно від їхнього розміру: більші потрапляють до першого класу, менші – до другого (напр. Tyr і Phe, Glu і Asp, Arg і Lys). Відповідно сайт зв'язування амінокислоти АРСаз першого класу розташований у порівняно неглибокій порожнині на поверхні ферменту, другого класу – у глибшому кармані.



**Рис. 8.5.** Структури АРСаз першого (мономерна Глу-тРНК-синтетаза, 1G59) і другого (Асп-тРНК-синтетаза, гомодимер, 1ASY) класів

Із частини молекули, яка утворює спільний структурний мотив, випетлюються інші структурні домени, які разом із N- або С-кінцевими доменами беруть участь у взаємодії з тРНК. АРСази двох класів упізнають різні елементи структури тРНК: маленький жолобок акцепторного стебла, D-петлю й антикодонову петлю – для першого класу, великий жолобок акцепторного стебла, варіабельну й антикодонову петлі – для другого (рис. 8.6). Специфічність упізнання тРНК залежить передусім від взаємодій з акцепторним стеблом: окрема штучна акцепторна частина здатна специфічно зв'язатися з АРСазою й акцептувати амінокислоту.



**Рис. 8.6.** Комплекси АРСаз першого і другого класів із відповідними тРНК: тРНК<sup>Glu</sup> (перший клас, 1G59) і тРНК<sup>Asp</sup> (другий клас, 1ASY). З димерною АРСазою зв'язані дві молекули тРНК, на представленій проекції видно лише одну з них

Специфічність зв'язування АРСазою такого великого ліганду, як тРНК, не становить значної проблеми: велика кількість контактів дозволяє легко дискримінувати різні типи тРНК за рахунок різниці у вільній енергії зв'язування. Амінокислоти, навпаки, є маленькими лігандами (до того ж іноді з дуже схожою структурою), з якими неможливо реалізувати велику кількість взаємодій. Утім, середня частота помилок при аміноацилюванні тРНК становить приблизно  $10^{-6}$  – такого рівня точності неможливо досягти просто за рахунок різниці у вільній енергії зв'язування, яка не може сильно відрізнятись для різних амінокислот. Відповідно АРСази використовують різноманітні системи корегування (редагування) помилок.

Один із механізмів редагування – *подвійне сито* – схематично зображений на рис. 8.7 на прикладі Пе-тРНК-синтетази. Активний центр, де відбуваються зв'язування й активування амінокислоти, здійснює первинне вибраковування частини амінокислот доволі великого розміру, інші амінокислоти (включаючи Пе) піддаються активуванню. На другому етапі всі аміноацил-аденілати, що містять менші порівняно із Пе амінокислоти, гідролізуються в іншому активному центрі (центрі редагування), а Пе переноситься на тРНК, оскільки розмір ізолейцил-аденілату не дозволяє йому зв'язатися із цим центром. У результаті дуже близький до Пе за структурою Val акцептується тРНК<sup>Le</sup> в одному випадку зі 180 тис.

Подвійне сито належить до категорії *претрансферних* шляхів редагування, коли корегування помилок здійснюється на рівні аміноацил-аденілату, тобто до перенесення амінокислоти на тРНК. Інші механізми претрансферного типу, що реалізуються деякими АРСазами, передбачають також розпізнання помилкових аміноацил-аденілатів і "виштовхування" їх назовні – тоді руйнування цих сполук здійснюється у цитоплазматичному просторі (рис. 8.8). Крім того, часто розпізнаванню помилкового аміноацил-аденілату сприяє споріднена тРНК, зв'язана із АРСазою. Деякі АРСази можуть здійснювати редагування помилок також після приєднання амінокислоти до тРНК – *посттрансферне редагування*, коли помилкова аміноацил-тРНК упізнається як ціле вже після перенесення амінокислоти на тРНК, і неспоріднена амінокислота відщеплюється у спеціальному каталітичному центрі (рис. 8.8).

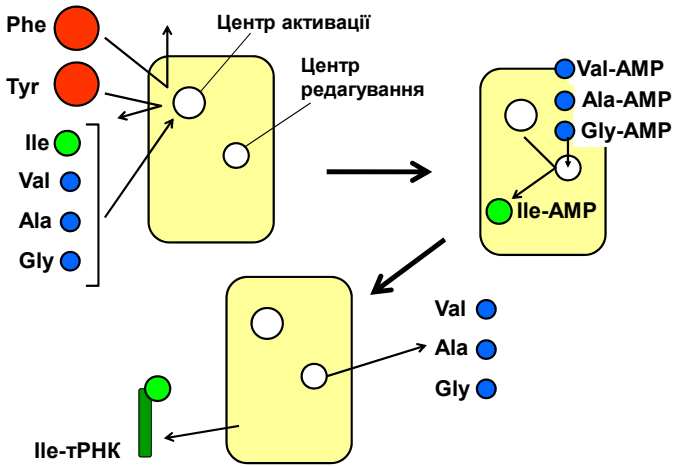


Рис. 8.7. Подвійне сито на прикладі Іле-тРНК-синтетази

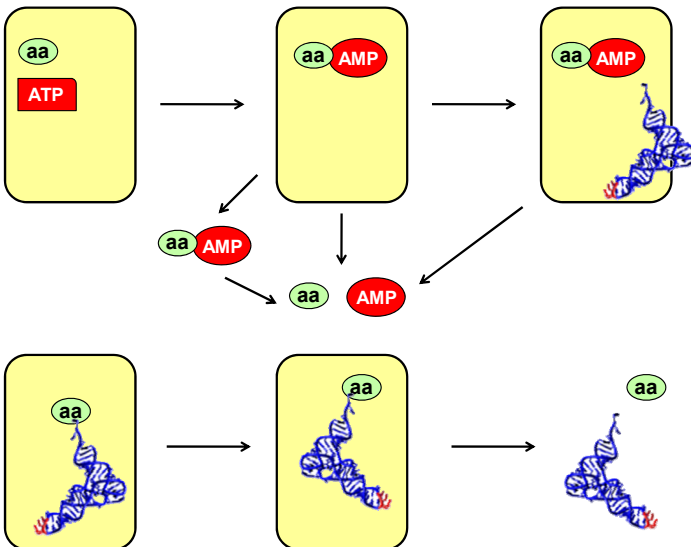
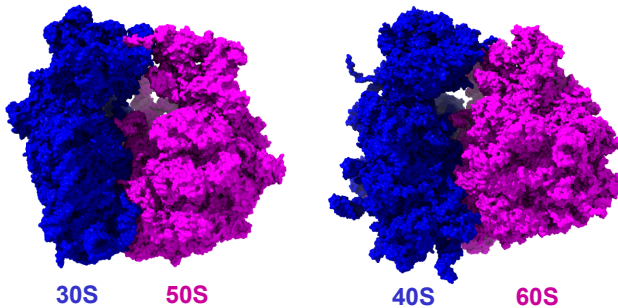


Рис. 8.8. Претрансферні (верхня панель) і посттрансферний (нижня панель) шляхи редагування помилок АРСазами

## Рибосома

Загальний вигляд бактеріальної та еукаріотичної рибосом представлено на рис. 8.9. За деяких відмінностей у деталях, організація рибосом різних доменів життя дуже подібна; так само подібні й принципи їхньої роботи. Загалом білковий синтез є найбільш консервативним біологічним процесом, у головних рисах спільним для усіх живих організмів.



**Рис. 8.9.** Загальний вигляд бактеріальної (ліворуч, 5WIT) і еукаріотичної (праворуч, 6GQB) рибосом. Указано коефіцієнти седиментації субодиниць

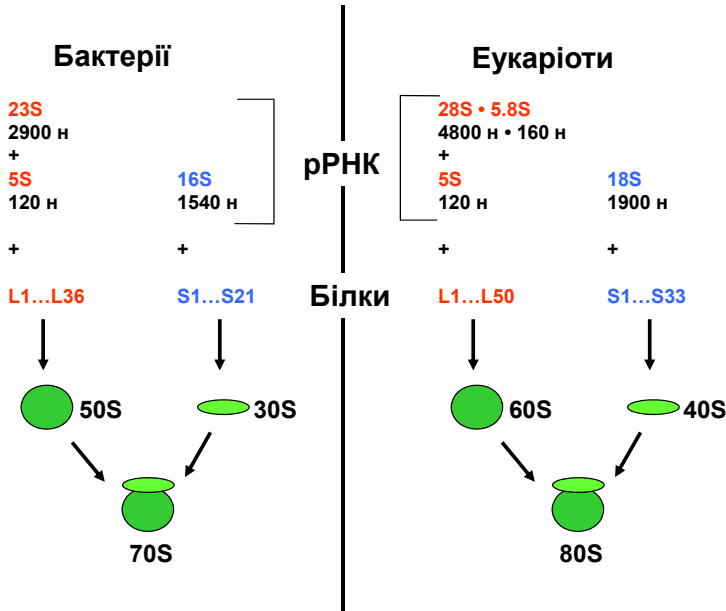
Рибосома завжди складається із двох субодиниць – великої і маленької, які, як і її інші компоненти, позначають *коефіцієнтами седиментації* – коефіцієнтами пропорційності, що показують, наскільки зростає швидкість руху частинки при центрифугуванні зі зростанням відцентрової сили (швидкості обертання ротора центрифуги). Коефіцієнт седиментації використовується як міра рухливості частинки, залежить від її маси, об'єму й форми, вимірюється у сведбергах (S) – позасистемних одиницях, що мають розмірність часу ( $1S = 10^{-13}$  с). Як видно з рис. 8.9, в інтерфейсі взаємодії між двома субодиницями залишається доволі велика порожнина, яка служить місцем зв'язування молекул тРНК. Кожна субодиниця рибосоми, своєю чергою, є рибонуклеопroteїновим комплексом, що складається з певних наборів білків і молекул рРНК.

## Компоненти рибосоми

Маленька субодинаця бактеріальної рибосоми з коефіцієнтом седиментації 30S містить одну молекулу рРНК із коефіцієнтом седиментації 16S і 21 молекулу рибосомних білків, що позначаються як S1–S21 (від *S*mall subunit). Велика субодинаця містить дві молекули рРНК (23S і 5S) і білки L1–L36 (від *L*arge subunit) – цей комплекс седиментує з коефіцієнтом 50S (рис. 8.10). Об'єднання субодинаць дає рибосому з коефіцієнтом седиментації 70S: оскільки коефіцієнт седиментації залежить від форми частинки, то він не є адитивною величиною. Еукаріотична рибосома містить трохи більшу рРНК 18S замість 16S, дві рРНК, що міцно взаємодіють між собою (28S і 5,8S), замість 23S і більшу кількість білків.

Синтез еукаріотичних рРНК 18S, 5,8S і 28S здійснюється у ядерці, яке формується на тандемних повторях кластера відповідних генів рРНК (див. рис. 4.6). Первинний транскрипт, що синтезується РНК-полімеразою I, має константу седиментації 45S і містить три фрагменти майбутніх рРНК, що розділені спейсерами. Процесинг рРНК – деградація спейсерів, а також модифікація певних основ і метилювання 2'-ОН-груп специфічних рибоз – здійснюється за участю *маленьких ядерцевих РНК* (snoRNA – small nucleolar RNA), які визначають специфічні сайти модифікацій (служать "гідом", що спрямовує туди відповідні ферменти). Маленькі ядерцеві РНК беруть участь також у визначенні сайтів хімічних модифікацій у складі тРНК і маленьких ядерних РНК. рРНК 5S синтезується РНК-полімеразою III на окремих кластерах відповідних генів поза ядерцем. У ядерці, практично одночасно із процесингом рРНК, відбувається її поступова взаємодія з рибосомними білками, до ядерця ж дифундує комплекс рРНК 5S із білками – утворюються субодинаці рибосоми, які далі транспортуються до цитоплазми.

Первинний транскрипт, що синтезується на бактеріальному рибосомному опероні, містить ділянки, які відповідають усім трьом бактеріальним рРНК, а також кілька майбутніх тРНК. Часткова деградація транскрипту спричиняє утворення зрілих молекул, які взаємодіють із рибосомними білками, формуючи дві субодинаці рибосоми. *In vivo* остаточне збирання рибосоми із двох субодинаць, як у про-, так і в еукаріотів, відбувається при ініціації трансляції.

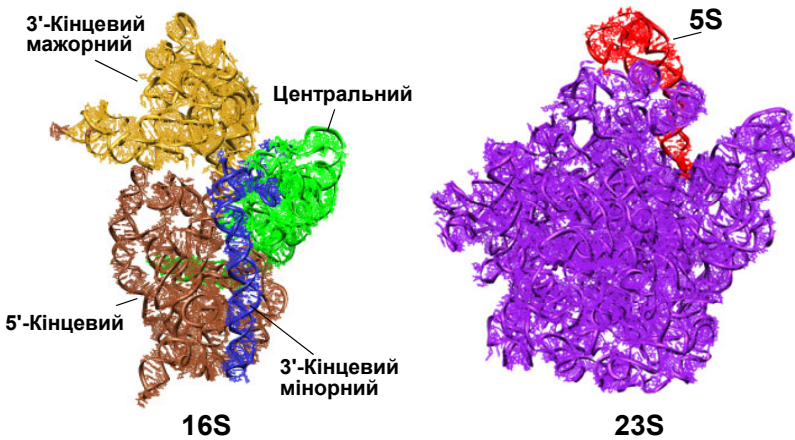


**Рис. 8.10.** Склад бактеріальної та еукаріотичної рибосом.  
rРНК позначено їхніми коефіцієнтами седиментації,  
указано також їхню довжину в нуклеотидах (н)

**Рибосомні РНК** становлять  $\sim 2/3$  маси рибосоми й саме вони визначають її структуру та функції. Полінуклеотидний ланцюг rРНК утворює велику кількість подвійних спіралей, з'єднаних петлями, – переважно це шпильки, проте відбувається і спарювання між віддаленими по ланцюгу ділянками. Загалом ланцюг утворює складну просторову структуру, де подвійні спіралі взаємодіють одна з одною і з одноланцюговими ділянками. Зокрема, аденозини одноланцюгових ділянок доволі часто зв'язуються з маленьким жолобком подвійних спіралей, стабілізуючи їхню структуру.

У складі бактеріальної rРНК 16S близько половини нуклеотидів залучено до  $\sim 60$  коротких дволанцюгових ділянок у середньому завдовжки вісім пар основ. Однак зазвичай у складі 16S РНК виділяють 45 більш-менш суцільних (із короткими неспареними ділянками всередині) подвійних спіралей. Просторова структура (рис. 8.11) чітко розділяється на чотири структурні домени: 5'-кінцевий, центральний, 3'-кінцевий мажорний і

3'-кінцевий мінорний, який складається з довгої спіралі 44 (порядковий номер) і коротшої останньої спіралі 45. Саме ця рРНК визначає структуру маленької субодиниці, показану на рисунках нижче, – на цей РНК-скелет "нарошуються" білки, однак загальна Y-подібна форма зберігається. Еукаріотична рРНК 18S є гомологічною 16S і відрізняється лише кількома вставками.



**Рис. 8.11.** Рибосомні РНК у складі маленької (16S, 2B9O) і великої (23S, 2B9P) субодиниць рибосоми *Thermus thermophilus*. Указано структурні домени рРНК 16S, довга спіраль у складі 3'-кінцевого мінорного домену – спіраль 44. рРНК орієнтовані до глядача інтерфейсами своєї взаємодії у складі рибосоми

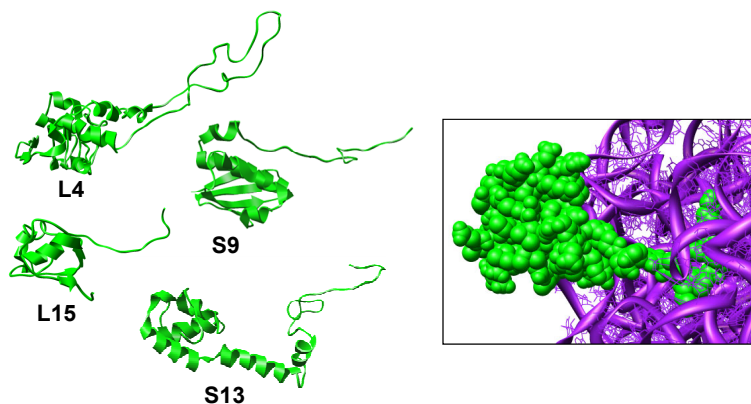
У складі рРНК 23S розрізняють шість структурних доменів, але вони тісно взаємодіють між собою. Порівняно із рРНК 16S, ця структура значно монолітніша (рис. 8.11). Еукаріотична рРНК 5,8S є гомологічною 5'-кінцевій зоні 23S. Узагалі еукаріотичні рРНК 28S і 5,8S, які утворюють міцний комплекс між собою, – це ніби трохи подовжена за рахунок вставок рРНК 23S, розділена на дві нерівні частини.

Рибосомні РНК на рис. 8.11 зображено в тій самій орієнтації, що й субодиниці рибосоми на рис. 8.13 – з боку інтерфейсу взаємодії між ними у складі рибосоми. На цій поверхні майже відсутні рибосомні білки – взаємодія між субодиницями переважно забезпечується контактами РНК-РНК. Усі функціональні сайти

рибосоми – сайти взаємодії із тРНК та інші – також формуються переважно рибосомними РНК. Певною мірою рибосомні РНК і є рибосомою. Однак при цьому стабільна функціонально активна структура рРНК забезпечується рибосомними білками.

**Рибосомні білки** розміщуються на поверхні рРНК (відповідно й на поверхні рибосоми), переважно на зовнішній поверхні обох субодиниць.

У більшості випадків рибосомні білки організовані за спільним планом (рис. 8.12): порівняно невелика глобула (у деяких білків глобула відсутня) і позитивно заряджений відросток (кінцевий хвіст, петля або одна-дві окремі  $\alpha$ -спіралі). Глобула розміщена на поверхні рРНК, а відросток занурюється всередину і стабілізує її структуру (рис. 8.12). Отже, усі рибосомні білки (за винятком деяких, що утворюють білкові відростки рибосомних субодиниць, див. нижче) взаємодіють із рРНК, хоча не всі мають високу спорідненість до вільної рРНК. Збирання субодиниць рибосоми із рРНК і білків (може бути здійснено *in vitro* без участі будь-яких факторів) відбувається поступово, через чотири – п'ять стадій: зв'язування частини білків на кожній стадії індукує конформаційну зміну рРНК у складі комплексу, що викликає спорідненість до нової порції білків.



**Рис. 8.12.** Приклади рибосомних білків у структурі рибосоми (2B9O, 2B9P). На вставці: взаємодія білка L15 із рРНК 23S

Головна роль рибосомних білків – підтримувати функціонально активну структуру рРНК. Проте деякі білки беруть участь у взаємодіях із елементами системи трансляції.

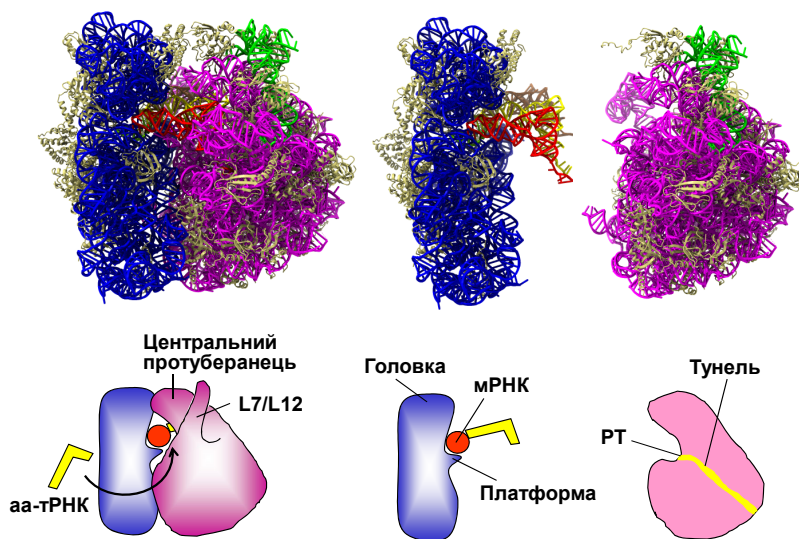
### Структура рибосоми

Структуру бактеріальної рибосоми показано в різних проєкціях на рис. 8.13–8.15, де також схематично зображено зовнішню анатомію субодиниць і їхнє розташування у складі рибосоми відносно одна одної та деяких інших елементів системи трансляції.



**Рис. 8.13.** Структура рибосоми *Thermus thermophilus* та її субодиниць (50S та 30S). Окремі субодиниць орієнтовані інтерфейсами взаємодії між ними до глядача, у складі рибосоми орієнтація великої субодиниць збережена. Білою стрілкою позначено напрям зору, у якому структуру зображено на рис. 8.14, блакитною – на 8.15 (а). Унизу: схематичне зображення субодиниць і їхнього комплексу в тих самих проєкціях, РТ – пептидил-трансферазний центр

Від основного доволі монолітного тіла великої субодиноці відходять три характерні відростки: виріст L1, палець (стебло) L7/12, сформовані відповідними рибосомними білками (ці білки відсутні у структурі на рис. 8.13), і *центральный протуберанець*, утворений комплексом певних білків із рРНК 5S. Два окремі структурні домени – *головка* та *платформа* – відходять від тіла маленької субодиноці. Їхню основу формують відповідно 3'-кінцевий мажорний і центральний домени рРНК 16S (порівн. рис. 8.11).



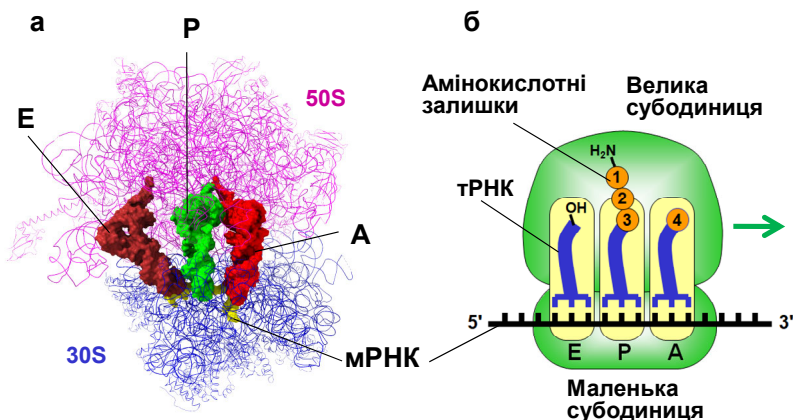
**Рис. 8.14.** Структура рибосоми та її субодиноць у комплексі із трьома молекулами тРНК з рис. 8.13 у іншій проєкції – з боку входу аа-тРНК до А-сайта (по білій стрілці на рис. 8.13). Унизу: схематичне зображення у тій самій проєкції, велика субодиноця праворуч – у розрізі

У складі рибосоми можна виділити три основні зони контактів між субодиноцями: головка – центральний протуберанець; платформа – виріст L1; центральні частини основного тіла обох субодиноць. Усі структурні елементи рухливі: можливі переміщення головки й пальця L7/12, обертання малої субодиноці

навкруг нормалі до поверхні великої субодиноці на  $\sim 10^\circ$  проти годинникової стрілки тощо. Рухи структурних елементів рибосоми в зонах контактів мають важливе значення для її функціонування, оскільки саме на інтерфейсі між субодиноцями містяться всі активні центри й сайти взаємодії з елементами системи трансляції, а саме:

- в основі центрального протуберанця розташований пептидил-трансферазний центр (рис. 8.13, 8.14), який відповідає за каталіз реакції синтезу пептидного зв'язку. Від пептидил-трансферазного центра через тіло великої субодиноці відходить тунель – канал виходу поліпептидного ланцюга, що синтезується (рис. 8.14);
- в основі пальця L7/12 міститься сайт зв'язування факторів трансляції, описаних нижче;
- у щілині між платформою і головою маленької субодиноці розміщується мРНК (рис. 8.13, 8.14);
- сайти зв'язування тРНК – центральний "модуль" системи трансляції – розташовані між двома субодиноцями: антикодонові частини тРНК взаємодіють із мРНК і маленькою субодиноцею, акцепторні частини – з великою субодиноцею (рис. 8.13–8.15). Рибосома містить три такі сайти: **А-сайт**, де відбувається зв'язування aa-тРНК; **Р-сайт**, де з рибосомою взаємодіє пептидил-тРНК (тРНК, до якої приєднаний *пептидил* – ланцюг, що синтезується); **Е-сайт** (від exit), де міститься деаміноацильована тРНК перед її звільненням із рибосоми.

Aa-тРНК потрапляє до рибосоми через щілину між субодиноцями (рис. 8.9, 8.14), розмір якої може змінюватись унаслідок рухливості структурних елементів рибосоми. Акцепторні частини тРНК, розташовані в А- і Р-сайтах, наближені одна до одної (рис. 8.15) і взаємодіють із великою субодиноцею у зоні пептидил-трансферазного центра. Кількість контактів великої субодиноці рибосоми із тРНК у Р-сайті більша, ніж кількість контактів, що утримують молекулу тРНК в А-сайті.



**Рис. 8.15.** Вигляд зверху (по блакитній стрілці на рис. 8.13) на структуру рибосоми в комплексі із трьома молекулами тРНК (а) і функціональна схема рибосоми (б): стрілкою позначено напрямок руху при трансляції, амінокислотні залишки пронумеровано відповідно до порядку їхнього включення до ланцюга

Структура рибосоми чітко вказує на те, що між двома субодинамицями реалізується своєрідне "розділення праці": маленька субодинамиця, яка взаємодіє із мРНК та антикодоновими частинами тРНК, відповідає переважно за декодувальну функцію рибосоми, а велика, яка взаємодіє з акцепторними частинами тРНК, – за каталітичну.

Як було зазначено вище, основну роль у процесі трансляції виконують рРНК, але це не означає, що рибосомні білки залишаються осторонь. Наприклад, білок S1, що міститься поблизу від сайту зв'язування мРНК, сприяє розплітання дволанцюгових шпильок у складі матриці; комплекс S1–S18–S21 взаємодіє із мРНК, а також ініціаторною тРНК при ініціації трансляції; білок L10 сумісно з певною ділянкою рРНК 23S і білком L11 бере участь у формуванні сайту зв'язування факторів трансляції в основі стебла L7/12.

Функціональну схему рибосоми наведено на рис. 8.15, б. Систему зображено в момент перед додаванням четвертої амінокислоти до ланцюга, що синтезується; показано aa-тРНК, пепти-

дил-тРНК і деаміноацильовану тРНК, що зв'язані відповідно з А-, Р- і Е-сайтами. Базуючись на цій схемі, розглянемо, як працює система трансляції під час елонгації, щоб пізніше з'ясувати, як ініціюється процес.

## Елонгаційний цикл

Робота рибосоми під час елонгації трансляції полягає в послідовному (потриpletно) зчитуванні інформації з мРНК і відповідному приєднанні амінокислот до поліпептидного ланцюга. Кожен такий крок складається із трьох операцій, що циклічно повторюються (*елонгаційний цикл*). Цикл розпочинається з такої конфігурації системи, коли в Р-сайті міститься пептидил-тРНК, А-сайт є вільним від тРНК і в його межах на маленькій субодиниці розташований черговий кодон, який має бути впізнаним (рис. 8.16). **Перша операція** циклу – *завантаження aa-тРНК* в А-сайт. Зв'язування aa-тРНК має відбутись із високою специфічністю щодо взаємодій між кодоном і антикодоном – тільки споріднена до кодону тРНК має бути відібрана системою. Процес розміщення aa-тРНК в А-сайті часто супроводжується дисоціацією з Е-сайта деаміноацильованої тРНК, яка може залишитись там із попереднього циклу. Наслідком зв'язування стає належне розташування акцепторних частин aa-тРНК і пептидил-тРНК відносно одна одної та каталітичного активного центра. У результаті рибосома здійснює **другу операцію** – *транспептидацію* – перенесення пептидилу з пептидил-тРНК на амінокислоту у складі aa-тРНК. У результаті система перебудовується: в А-сайті опиняється пептидил-тРНК із подовженим на одну амінокислоту пептидилом, у Р-сайті – деаміноацильована тРНК. **Третя операція** – *транслокація* – полягає у переміщенні рибосоми на один кодон уздовж мРНК (молекули тРНК залишаються зв'язаними зі своїми кодонами, деаміноацильована тРНК може звільнитись з Е-сайту), після чого розпочинається наступний елонгаційний цикл.

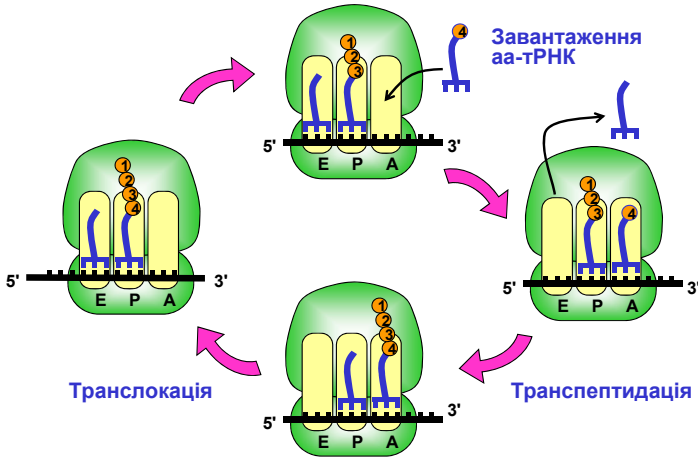


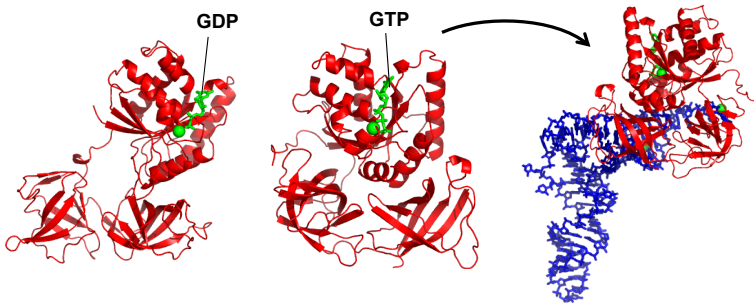
Рис. 8.16. Схема елонгаційного циклу

Відбір серед різних молекул aa-тРНК на першому етапі, коли неспоріднені молекули мають швидко звільнитися, а також рух рибосоми при транслокації, передбачають реалізацію певних "відкритих", не жорстко зафіксованих, рухливих структурних форм рибосоми із частково порушеними взаємодіями між субодиницями. Каталіз транспептидації, навпаки, вимагає жорсткої фіксації субстратів у "закритій", жорсткій і нерухливій формі рибосоми. Ефективне розв'язання цих суперечливих завдань залежить від полегшення реалізації відкритих форм рибосоми на першому й третьому етапах елонгаційного циклу завдяки двом факторам елонгації (EF – Elongation Factors): **EF1** (бактеріальний фактор позначається як **EF-Tu**, еукаріотичний аналог – як eEF1 або eEF1A, eEF – eukariotic Elongation Factor) і **EF2** (або **EF-G** у бактерій і **eEF2** в еукаріотів). Обидва фактори належать до родини GTP-зв'язувальних білків (G-білків). Інші білки цієї родини залучені до багатьох різноманітних процесів, зокрема є елементами асоційованих із мембранами клітинних сигнальних систем.

## Завантаження аміноацил-тРНК

### Елонгаційний фактор EF1

Фактор EF1 – мономерний білок, що має три структурні домени. Він може існувати у двох структурних станах (рис. 8.17), що є типовим прикладом перемикання структурних станів білка лігандом (див. рис. 2.19, 2.23). Співвідношення між вільними енергіями цих станів, а відповідно й імовірність переважної реалізації одного з них, залежать від типу ліганду (GTP чи GDP), що зв'язаний із білком. У комплексі із GDP (гуанозиндифосфат) реалізується відкрита конформація EF1 із порушеними взаємодіями між структурними доменами. Заміна GDP на GTP (гуанозинтрифосфат) спричиняє локальну конформаційну перебудову в межах GTP-зв'язувального домену, унаслідок якої певні амінокислотні групи виводяться до інтерфейсу взаємодії з іншими доменами – структура "закривається". Гідроліз GTP (який здійснюється на рибосомі самим білком, див. нижче) зумовлює заміну ліганду і, відповідно, зворотне перемикання структурного стану.



**Рис. 8.17.** Структура фактора EF-Tu (EF1) у комплексі із GDP (1TUI) і GTP (1TFI) та комплекс EF-Tu·GTP з  $\alpha$ -тРНК (1B23). Зелена кулька – іон  $Mg^{2+}$

Структурна перебудова EF1 є важливою не сама по собі. GDP- і GTP-асоційовані структурні форми мають відповідно низьку й високу спорідненість до  $\alpha$ -тРНК і рибосоми: саме у вигляді потрійного комплексу EF1·GTP– $\alpha$ -тРНК (рис. 8.17) відбувається зв'язування  $\alpha$ -тРНК із рибосомою на першому етапі

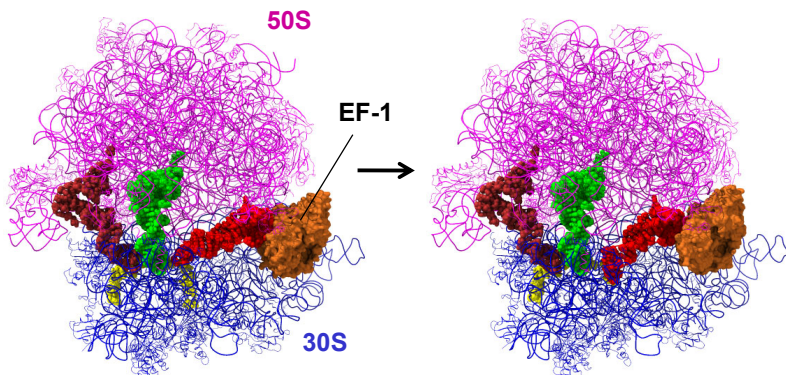
елонгаційного циклу. При цьому вже цей потрійний комплекс відіграє роль ліганду, що перемикає структурні стани рибосоми.

Зв'язування EF1·GTP з акцепторною частиною aa-тРНК (рис. 8.17) відбувається одразу після аміноацилювання тРНК. При завершенні процесу взаємодії з рибосомою (див. нижче) EF1 здійснює гідроліз GTP, що веде до втрати спорідненості й дисоціації EF1·GDP. Після дисоціації молекула GDP замінюється на GTP за допомогою кофакторів (EF-Ts у бактерій, eEF1B в еукаріотів), і знову відбувається зв'язування EF1·GTP з новою молекулою aa-тРНК.

Щоб aa-тРНК могла потрапити до рибосоми, рибосома має опинитися в певному відкритому структурному стані з розширеним каналом між субодинами з боку стебла L7/12 (див. рис. 8.14). Крім того, така відкрита форма рибосоми не допускає дуже міцної взаємодії з aa-тРНК, тобто сприяє її легкій дисоціації за невідповідності між кодоном і антикодоном. Однак зрозуміло, що відкрита форма рибосоми із частково порушеними контактами (енергетично вигідними взаємодіями) між субодинами має характеризуватися підвищеною вільною енергією, тобто малоймовірна. Інший важливий аспект полягає в тому, що на етапі первинного перебирання aa-тРНК із різними антикодонами необхідно повністю виключити для акцепторної частини тРНК можливість випадково потрапити до пептидилтрансферазного центра. Узагалі для такого великого ліганду, як тРНК, існує кінетичний бар'єр асоціації / дисоціації: одночасне утворення / руйнування великої кількості контактів із лігандом є проблематичним – проміжні стани з лімітованим набором контактів мають забезпечити прискорення обох процесів. Роль EF1 саме й полягає у розв'язанні зазначених вище проблем:

- взаємодія EF1 з акцепторною частиною aa-тРНК залишає тільки антикодонову частину вільною для взаємодій із рибосомою (і зв'язаною з нею мРНК);
- при зв'язуванні потрійного комплексу EF1 (який при цьому зв'язаний з акцепторною частиною aa-тРНК) взаємодіє з рибосомою так, що акцепторна частина aa-тРНК розташовується дуже далеко від пептидил-трансферазного центра на великій субодинаці (рис. 8.18, порівн. рис. 8.15, а);

- EF1·GTP має підвищену спорідненість до відкритої форми рибосоми, тобто подібно до того, як GTP перемикає конформацію самого EF1, потрійний комплекс EF1·GTP-aa-тРНК фіксує відкриту форму рибосоми, яка стає енергетично вигідною за рахунок енергії взаємодій з EF1 (рис. 8.19).

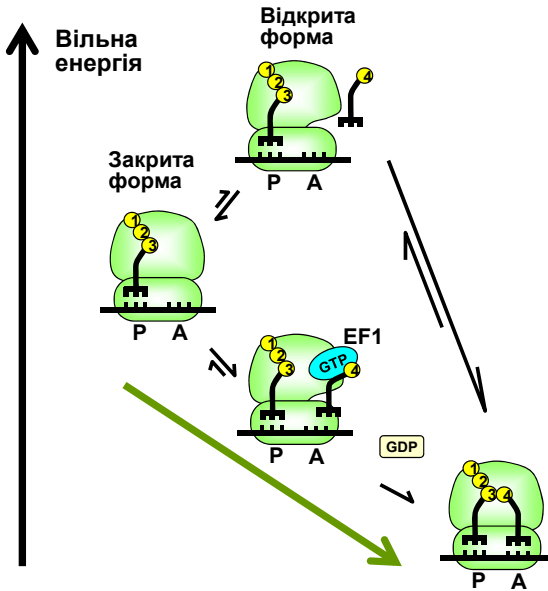


**Рис. 8.18.** Зв'язаний із рибосомою комплекс EF1·GTP-aa-тРНК: первинне зв'язування (ліворуч, 5UUK) і остаточне розташування комплексу (праворуч, 5UYM). У структурі ліворуч ще не відбулось комплементарного спарювання кодону з антикодоном, праворуч – остаточне зв'язування після впізнання кодону антикодоном і структурної перебудови рибосоми (деталі у тексті)

Отже, за наявності EF1·GTP здійснюється не дуже міцне зв'язування aa-тРНК із рибосомою (переважно завдяки взаємодії EF1 із маленькою субодиницею рибосоми, ліва структура на рис. 8.18), яке швидко змінюється дисоціацією; у цей час антикодонова частина тРНК намагається впізнати кодон. Якщо таке впізнання відбувається (див. нижче), то спрацьовує GTP-аза активність EF1 – після гідролізу GTP фактор дисоціює, що дозволяє акцепторній частині aa-тРНК остаточно розміститися в А-сайті.

Зв'язування aa-тРНК є внутрішньою властивістю рибосоми: *in vitro* зв'язування може відбуватися без участі будь-яких факторів, оскільки супроводжується зниженням вільної енергії. Проте безфакторне зв'язування є доволі повільним через наявність на його шляху енергетичного бар'єра – високоенергетичного

відкритого стану рибосоми (рис. 8.19). Отже, EF1 *не зумовлює* зв'язування аа-тРНК, а тільки суттєво *прискорює* його. Можна сказати, що EF1 здійснює каталіз конформаційних перетворень рибосоми за рахунок спорідненості до проміжного інтермедіатного стану. Щодо гідролізу GTP, то він потрібен лише для заміни ліганду з метою позбавитися цієї спорідненості й забезпечити дисоціацію EF1 і повернення рибосоми до закритого стану, необхідного для наступної стадії елонгаційного циклу. Однак це має відбутися лише за умови впізнання кодону антикодоном аа-тРНК.



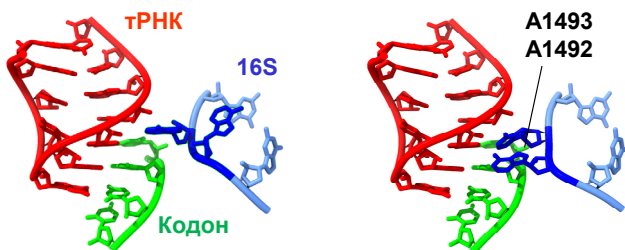
**Рис. 8.19.** Зв'язування з А-сайтом рибосоми аа-тРНК у вільному стані й у комплексі з EF1·GTP. Різні стани системи розміщені на різних рівнях відповідно до їхньої вільної енергії. Зелена стрілка вказує оптимальний шлях процесу. Показано тільки два сайти зв'язування тРНК

### Упізнання кодону

Як зазначено вище, первинне зв'язування комплексу EF1·GTP-аа-тРНК відбувається у такий спосіб, що EF1 взаємодіє з маленькою субодиницею (структура ліворуч на рис. 8.18) –

фактор не має змоги потрапити до свого остаточного сайту в основі стебла L7/12, де тільки й може відбутись гідроліз GTP. Коли під час первинного зв'язування після кількох випробувань у межах А-сайта на маленькій субодиниці опиняється нарешті комплементарний антикодон, між ним і кодоном має утворитись коротка подвійна спіраль. Оскільки вона коротка (три пари основ), то доволі нестабільна: кодон-антикодонова взаємодія за межами рибосоми неефективна. При цьому рибосома (передусім маленька субодиниця) не тільки створює особливе середовище, яке стабілізує кодон-антикодонову спіраль, але й сприяє підвищенню специфічності – ефективно дискримінує споріднені й неспоріднені до кодону антикодони.

При утворенні комплементарної кодон-антикодонової подвійної спіралі з її маленьким жолобком взаємодіють два консервативні аденозини рРНК 16S (з порядковими номерами 1492 і 1493), утворюючи водневі зв'язки із 2'-ОН-групами рибоз нуклеотидів кодону й антикодону (рис. 8.20). Така взаємодія, що реалізується лише за умови комплементарності, стабілізує перші дві (відносно позицій кодону) нуклеотидні пари. Друга нуклеотидна пара додатково стабілізується, також через маленький жолобок, взаємодією з консервативним G530 рРНК 16S. Третя пара кодон-антикодонової спіралі теж взаємодіє з певним цитидином рРНК 16S, але ця взаємодія менш специфічна й допускає неоднозначність спарювання (див. підрозділ, де йдеться про тРНК).



**Рис. 8.20.** Взаємодія в межах А-сайта на маленькій субодиниці між кодоном, антикодоною петлею тРНК і частиною рРНК 16S у випадку неспорідненого (ліворуч, 1N36 – друга кодон-антикодонова пара основ некомплементарна) і спорідненого (праворуч, 1BVM) антикодонів

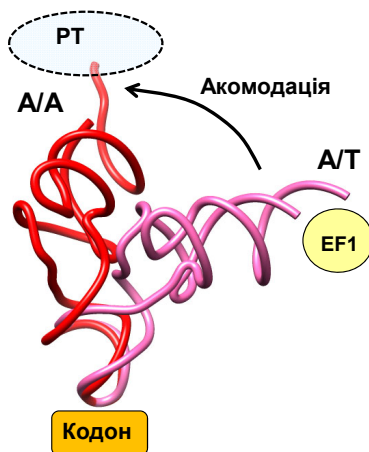
Взаємодії нуклеотидів рРНК 16S із кодон-антикодоною спіраллю не тільки підвищують специфічність упізнання кодо-ну, але й мають інший важливий наслідок. Як показано на рис. 8.20, для реалізації цих взаємодій A1492 і A1493 мають бути переорієнтовані відносно сусідніх нуклеотидів. Така локальна конформаційна перебудова запускає каскад конформаційних змін у інших частинах рРНК 16S за принципом доміно, що спричиняє глобальну конформаційну перебудову всієї маленької субодиниці. Зокрема, головка переміщується у напрямку до великої субодиниці, міцно замикаючи антикодонову частину aa-тРНК на кодоні мРНК. Одночасно маленька субодиниця штовхає зв'язаний з акцепторною частиною aa-тРНК EF1 до його сайту в основі стебла L7/12 (момент, коли EF1 міцно взаємодіє зі своїм сайтом, відображений праворуч на рис. 8.18). У сайті EF1 на великій субодиниці міститься елемент структури рибосоми, що активує GTPазу фактора. Одразу після такої жорсткої фіксації відбувається гідроліз GTP, конформаційна перебудова EF1 із втратою його спорідненості до рибосоми й тРНК і дисоціація фактора.

### **Акомодація aa-тРНК**

Перед гідролізом GTP розміщення aa-тРНК на рибосомі суттєво відрізняється від кінцевої точки процесу зв'язування – локалізації в А-сайті. Оскільки сайти зв'язування тРНК формуються двома субодиницями (мають дві половини), то остаточне положення aa-тРНК в А-сайті, зображене на рис. 8.15, *a*, позначають як А/А. У складі комплексу з EF1 антикодонова частина розміщена в А-сайті на маленькій субодиниці, але акцепторна частина утримується в основі стебла L7/12 на великій – *А/Т-положення* (права структура на рис. 8.18). Процес *акомодації* полягає у русі aa-тРНК з одного положення в інше (рис. 8.21).

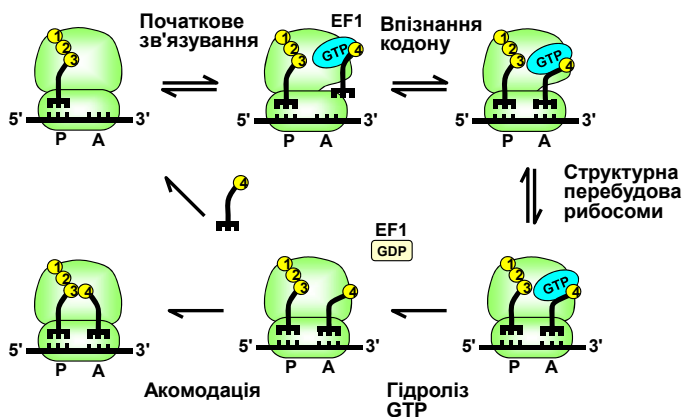
В А/Т-положенні за рахунок фіксації акцепторного стебла зв'язаним із великою субодиницею EF1, а антикодону – кодоном на маленькій субодиниці, антикодонове стебло тРНК суттєво вигинається – реалізується напружена конформація. Після гідролізу GTP, коли EF1 дисоціює, один із фіксаторів зникає – тРНК, ніби пружина, розпрямляється, її акцепторне стебло ав-

томатично опиняється в А-сайті на великій субодиниці – у зоні пептидил-трансферазного центра. Ефективність акомодатії залежить від жорсткості фіксації антикодонової петлі тРНК на маленькій субодиниці. Якщо взаємодія між кодоном і антикодоном не є повністю комплементарною, то aa-тРНК дисоціює від рибосоми – це друга (після первинного зв'язування) можливість виправити помилку впізнання кодону.



**Рис. 8.21.** А/Т-(1QZA) і А/А-(1QZB) положення aa-тРНК на рибосомі до й після акомодатії, відповідно. Кодон і EF1 фіксують напружену конформацію тРНК в А/Т-положенні. РТ – пептидил-трансферазний центр

Отже, загальний сценарій процесу зв'язування aa-тРНК з рибосомою можна описати так (рис. 8.22). На першому етапі початкового відбору aa-тРНК відбувається швидка асоціація / дисоціація потрійних комплексів EF1·GTP-aa-тРНК із відкритою формою рибосоми, яка фіксується за рахунок взаємодій з EF1. Відкрита форма сприяє також дисоціації деаміноацильованої тРНК з Е-сайта рибосоми, де ця тРНК могла залишитись після попереднього елонгаційного циклу (не показано на рис. 8.22).



**Рис. 8.22.** Схема основних подій при зв'язуванні aa-тРНК з А-сайтом рибосоми. Е-сайт тРНК не показано

При утворенні комплементарної кодон-антикодонової подвійної спіралі вона стабілізується за рахунок взаємодій із рРНК 16S. Локальна конформаційна зміна 16S, яка є наслідком цих взаємодій, спрацьовує як тригер структурної перебудови маленької субодиниці, унаслідок якої антикодонова петля тРНК жорстко замикається на маленькій субодиниці, а акцепторне стебло – на великій у основі пальця L7/12; при цьому EF1 набуває GTPазної активності.

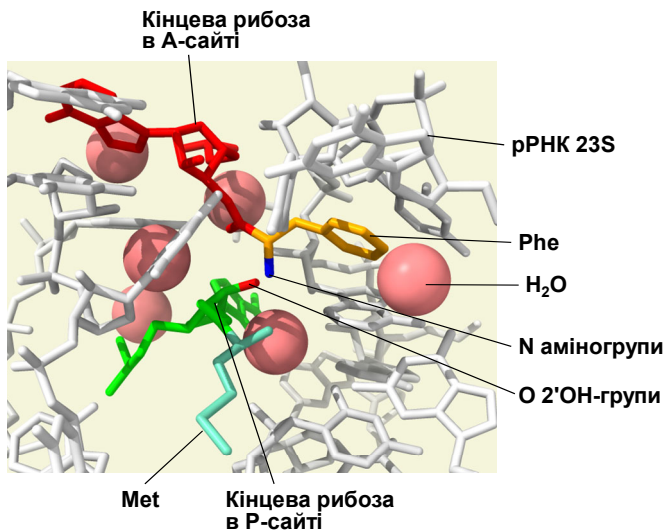
Гідроліз GTP розмикає ланцюг подій на дві частини, роблячи його необоротним. Після гідролізу EF1·GDP дисоціює, залишаючи aa-тРНК у напруженій конформації. Далі процес розгалужується: або, при недостатньо міцних кодон-антикодонових взаємодіях, відбувається дисоціація aa-тРНК, яка повертає систему до вихідного стану, або здійснюється акомодация – релаксація напруженої aa-тРНК, яка виводить її акцепторний кінець до пептидил-трансферазного центра. Таким чином, загальним наслідком процесу зв'язування є повна підготовка системи трансляції до наступного етапу елонгаційного циклу.

## Транспептидація

Реакція перенесення пептидилу – *транспептидація* – полягає у руйнуванні ковалентного зв'язку між карбоксильною групою С-кінцевої амінокислоти пептидилу й 3'-кінцевою ОН-групою пептидил-тРНК і утворенні натомість пептидного зв'язку між звільненою карбоксильною групою й аміногрупою амінокислоти у складі аа-тРНК. Отже, при транспептидації пептидил переноситься з Р- до А-сайта рибосоми: в А-сайті опиняється пептидил-тРНК із подовженим на одну амінокислоту пептидилом, у Р-сайті залишається деаміноацильована тРНК (див. рис. 8.16).

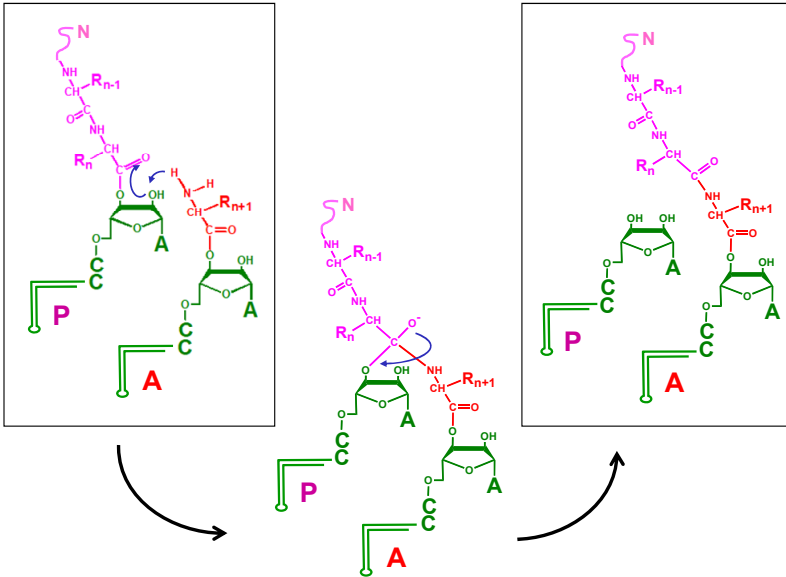
Стандартна вільна енергія гідролізу зв'язку між пептидилом і тРНК оцінюється приблизно в  $-7,5$  ккал/моль, енергія гідролізу пептидного зв'язку  $-0,5$  ккал/моль. Отже, загальний енергетичний баланс транспептидації  $\sim -7,0$  ккал/моль. Це спонтанний процес, який не потребує джерел енергії – енергію було заощаджено при аміноацилюванні тРНК (див. підрозділ, присвячений тРНК).

Процес транспептидації, що відбувається безпосередньо після акомодатії аа-тРНК, є найшвидшим етапом елонгаційного циклу. Каталіз реакції здійснюється пептидил-трансферазним центром великої субодиниці рибосоми, який розташований в основі центрального протуберанця (див. рис. 8.13, 8.14) і формується певною частиною рРНК 23S (28S в еукаріотів). У радіусі  $\sim 18$  Å від субстратів реакції немає жодної білкової хімічної групи (рис. 8.23), тобто *рибосома – це рибозим*. Це, поперше, найбільший із відомих рибозимів, по-друге – єдиний, що здатен каталізувати реакцію полімеризації. Слід, проте, ще раз зауважити, що вільна рРНК 23S (28S) не має каталітичної активності – активна конформація рРНК реалізується лише в комплексі з певним мінімальним набором білків.



**Рис. 8.23.** Субстрати реакції транспептидації у пептидил-трансферазному центрі рибосоми безпосередньо перед початком реакції (1VY4). Показано лише найближче оточення субстратів, що формується нуклеотидами рРНК та іммобілізованими молекулами води. Для двох молекул тРНК показано тільки 3'-кінцеві аденозини з приєднаними до них амінокислотами (пептидил тРНК містить одну амінокислоту)

Каталіз реакції транспептидації на рибосомі є ентропійним (див. обговорення механізмів каталізу в розділі 2): пептидил-трансферазний центр жорстко фіксує субстрати реакції у певній сприятливій конфігурації, що служить головним фактором прискорення реакції. Хімія процесу передбачає перенесення протона з аміногрупи амінокислоти у складі aa-тРНК в А-сайті на наближену до неї 2'-ОН-групу рибози 3'-кінцевого нуклеотиду пептидил-тРНК (рис. 8.23, 8.24). Ця 2'-ОН-група виконує важливу функцію переносника протона, який вона врешті-решт передає через кілька проміжних переносників на 3'-ОН-групу тієї самої рибози при руйнуванні ковалентного зв'язку між нею і пептидилом. Процес передавання протона здійснюється за допомогою системи водневих зв'язків, до якої залучені нуклеотиди рРНК та іммобілізовані в активному центрі молекули води.



**Рис. 8.24.** Спрощена схема процесу перенесення пептидилу з пептидил-тРНК у Р-сайті на аа-тРНК в А-сайті через інтермедіатну сполуку. Амінокислотні бокові залишки R пронумеровані в порядку приєднання їх до поліпептидного ланцюга. Синіми стрілками позначено деякі етапи перенесення протона у процесі реакції

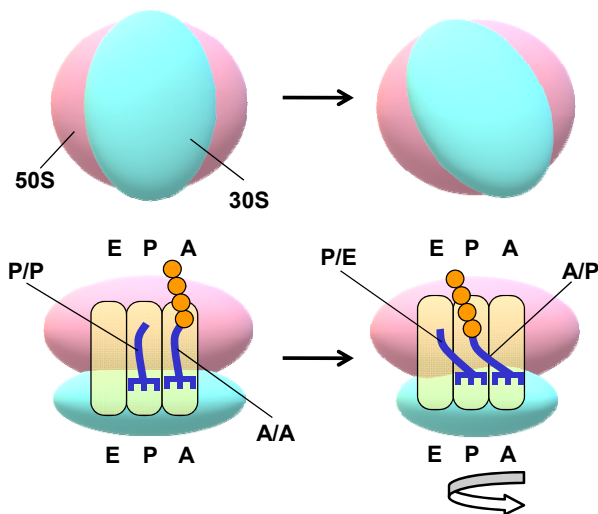
Реакція транспептидації проходить через утворення принаймні двох інтермедіатних сполук, у складі другого інтермедіату формується зв'язок між N аміногрупи аа-тРНК і С, яким пептидил приєднаний до тРНК, атом оксигену набуває при цьому негативного заряду (рис. 8.24). Після перенесення протона на 3'-ОН-групу рибози пептидил-тРНК стає деаміноацильованою – втрачає зв'язок із карбоксильною групою, замість якого остаточно формується пептидний зв'язок.

Отже, механізм каталізу транспептидації цілком аналогічний такому білкових ферментів: рибосома здійснює жорстке взаємне орієнтування субстратів у активному центрі, а також забезпечує стабілізацію проміжного високоенергетичного стану й переміщення протона від одного субстрату до іншого.

Процес транспептидації загалом не потребує допоміжних факторів, але за одним винятком. При синтезі олігопролінових треків виникають стеричні проблеми для їхнього розташування на вході в тунель усередині великої субодиниці (схематично зображений на рис. 8.14). Ця обставина суттєво дестабілізує пептидил-тРНК у Р-сайті. Додатково стабілізувати її та подолати стеричні перешкоди допомагає особливий фактор елонгації, що зв'язується з Е-сайтом рибосоми – EF-P (у бактерій) або eIF-5A (в еукаріотів).

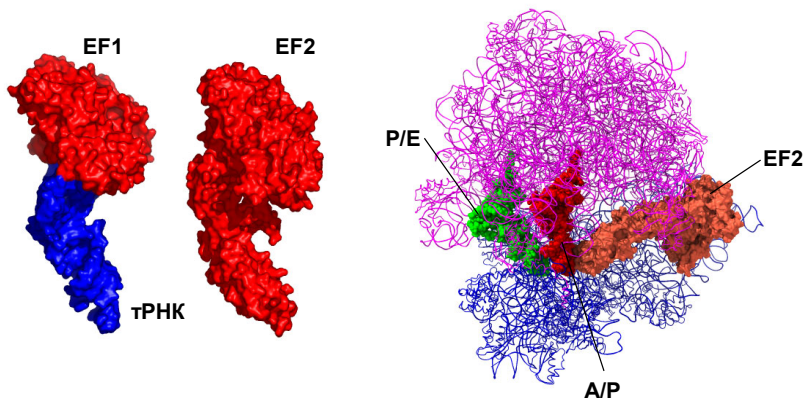
## Транслокація

Результатом транспептидації є значне перегрупування лігандів, зв'язаних із рибосомою: пептидил-тРНК опиняється в А-сайті, деаміноацильована тРНК – у Р-сайті. Унаслідок такого перегрупування, а також підвищеної спорідненості акцепторної частини пептидил-тРНК до Р-сайта, структура стає менш жорсткою: реалізується обертальний рух двох субодиниць одна відносно одної навкруг нормалі до інтерфейсу між ними. Обертання малої субодиниці відносно великої на  $\sim 10^\circ$  проти годинникової стрілки спричиняє виникнення *гібридних сайтів* тРНК А/Р і Р/Е – положень тРНК, коли їхні акцепторні частини переміщуються в Р- (або Е-)сайт на великій субодиниці, але антикодонові частини залишаються в А- (або Р-)сайті на маленькій (рис. 8.25). Оскільки такий структурний стан рибосоми асоційований з енергетично не вигідним послабленням взаємодій між субодиницями, то він є менш імовірним. Функцію прискорення процесу транслокації, першим етапом якої є стан із гібридними сайтами, виконує елонгаційний фактор EF2 (EF-G у бактерій).

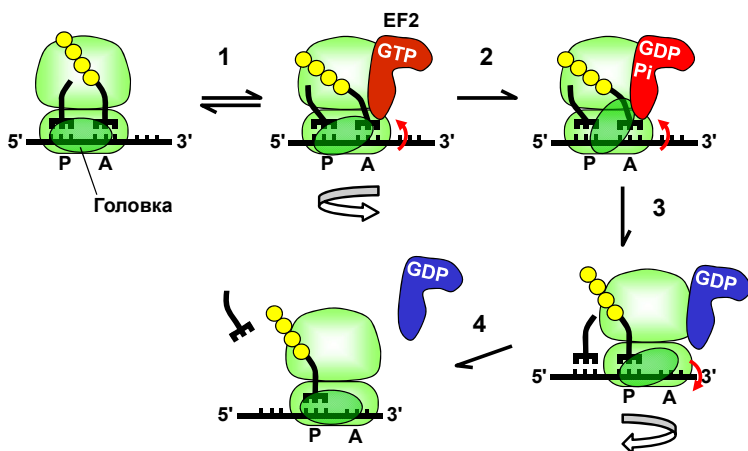


**Рис. 8.25.** Обертальний рух субодиниць рибосоми перед транслокацією і виникнення гібридних сайтів тРНК. Верхня панель: напрямок зору вздовж нормалі до інтерфейсу між субодиницями (як на рис. 8.13). Нижня панель: вид зверху (як на рис. 8.15)

Частина молекули EF2 гомологічна білку EF1: також має сайт зв'язування GTP/GDP і взаємодіє з великою субодиницею рибосоми в основі пальця L7/12. Однак у складі білка є ще один додатковий домен, який мімікрує під структуру антикодонової частини тРНК, – структура білка загалом нагадує комплекс EF1-тРНК (рис. 8.26). На першому етапі транслокації EF2 стабілізує стан рибосоми з оберненими одна відносно одної субодиницями (а отже, і гібридними сайтами тРНК). При цьому відбувається також обертальний рух головки маленької субодиниці в напрямку, позначеному на рис. 8.27, що спричиняє послаблення взаємодій головки з антикодоновими частинами тРНК.



**Рис. 8.26.** Поверхня комплексу EF-Tu (EF1) із тРНК (структура з рис. 8.18) і білка EF-G (EF2, 2EFG) та комплекс EF-G із рибосомою (7K54)



**Рис. 8.27.** Спрощена схема послідовності основних подій при транслокації рибосоми. 1 – зв'язування EF2-GTP із рибосомою, фіксація стану, у складі якого маленька субодиниця обертається відносно великої (біла стрілка), а головка маленької субодиниці також здійснює обертальний рух (червона стрілка); 2 – гідроліз GTP, розблокування рибосоми; 3 – дисоціація неорганічного фосфату  $P_i$ , зворотні обертальні рухи й маленької субодиниці, переміщення рибосоми на один кодон відносно мРНК і обох молекул тРНК; 4 – дисоціація EF2-GDP, можлива дисоціація деаміноацильованої тРНК з E-сайта, остаточне розміщення пептидил тРНК у P-сайті

Доволі швидко після зв'язування EF2 здійснює гідроліз GTP, при цьому неорганічний фосфат залишається асоційованим із білком – ця структурна форма EF2 найефективніше розблоковує структуру рибосоми в ділянці, де мРНК і антикодонові частини тРНК взаємодіють із маленькою субодиноцею. Головка маленької субодиноці продовжує обертальний рух, що сприяє переміщенню рибосоми відносно тРНК і мРНК. Домен EF2, схожий на антикодонову частину тРНК, при цьому витісняє антикодонову частину пептидил-тРНК із ділянки А-сайта на маленькій субодиноці, що сприяє руху рибосоми у напрямку до 3'-кінця мРНК. При дисоціації неорганічного фосфату рух рибосоми завершується, а головка маленької субодиноці й сама маленька субодиноця здійснюють зворотне обертання до свого "стандартного" положення. Унаслідок цього пептидил тРНК остаточно розміщується в Р-сайті (Р/Р-положення), деаміноацильована тРНК – в Е-сайті. Після дисоціації фосфату EF2 втрачає спорідненість до рибосоми й також дисоціює (рис. 8.27). Отже, роль фактора елонгації EF2 полягає в тому, що він стабілізує стан рибосоми з оберненою відносно великої маленькою субодиноцею, розблоковує (після гідролізу GTP) кодон-антикодонові комплекси мРНК із тРНК і забезпечує рух рибосоми в потрібному напрямку – до 3'-кінця мРНК.

У результаті транслокації система стає готовою до наступного елонгаційного циклу: новий кодон опинився в А-сайті, сам А-сайт вільний від тРНК. Деаміноацильована тРНК, яка в результаті транслокації опинилася в Е-сайті рибосоми, може дисоціювати звідти одразу або залишитися до етапу зв'язування наступної aa-тРНК.

Слід ще раз зауважити, що під елонгаційними факторами EF1/2 скрізь розумілись бактеріальні фактори EF-Tu/G, відповідно, – молекулярні механізми елонгації трансляції найкраще вивчені для бактерій. В еукаріотів ці механізми, напевно, суттєво не відрізняються, хоча детальної інформації щодо еукаріотичної системи трансляції до сих пір не вистачає.

Описані механізми функціонування рибосоми під час елонгаційного циклу демонструють, що рибосома працює за принципами молекулярної машини (див. розділ 2), яка з високою

швидкістю (у середньому 10–15 елонгаційних циклів за секунду) і точністю (частота помилок  $\sim 10^{-4}$ ) здійснює білковий синтез. Під час елонгаційного циклу рибосома осцилює між закритим структурним станом, який забезпечує каталіз реакції нарощування амінокислотного ланцюга, і відкритими станами, у яких здійснюється специфічний відбір амінокислот згідно з генетичним кодом і переміщення рибосоми вздовж матриці. Перемикавання конформацій залежить від взаємодій із факторами елонгації, які фіксують проміжні відкриті стани рибосоми, забезпечуючи прискорення перебудов. Гідроліз двох молекул GTP під час елонгаційного циклу використовується для звільнення цих факторів і повернення рибосоми до закритої жорсткої конформації. Рушійною силою для переміщення рибосоми та її частин є хаотичний тепловий рух, але молекулярна конструкція рибосоми, її взаємодії із тРНК, матрицею і факторами елонгації, зміни цих взаємодій унаслідок хімічних реакцій (транспептидація, гідроліз GTP) задають певну траєкторію рухів і каналізують їх у певних напрямках. Аналогічні механізми використовуються при ініціації і термінації трансляції.

## Ініціація трансляції

Розпочинатися процес елонгації трансляції має зі стартового кодону мРНК, і в першому елонгаційному циклі, коли немає пептидил-тРНК у Р-сайті, якась стартова – ініціаторна – aa-тРНК має виконувати цю роль. Отже, сутність ініціації трансляції полягає в:

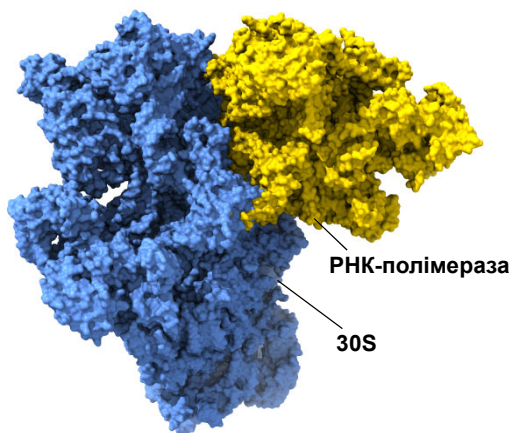
- упізнанні стартового кодону, який задає початок і рамку зчитування інформації;
- завантаженні ініціаторної aa-тРНК на стартовий кодон і водночас у Р-сайт рибосоми;
- збиранні рибосоми із двох субодиниць на мРНК у зоні стартового кодону.

Стартовим кодоном виступає здебільшого метіоніновий кодон AUG. Відповідно ініціаторною є завжди Met-тРНК<sup>Met</sup>. Індекс "і" вказує на те, що це – саме *ініціаторна* метіонінова

tРНК, тобто вона відрізняється за структурою від tРНК<sup>Met</sup>, яка використовується при елонгації. У бактерій, на відміну від еукаріотів, аміногрупа метіоніну ініціаторної tРНК формільована (до ініціаторної tРНК приєднаний N-форміл-метіонін, f-Met). Проте це далеко не єдина відмінність – механізми ініціації трансляції суттєво відрізняються у бактерій і еукаріотів.

### **Ініціація трансляції у бактерій**

Прокаріотична трансляція тісно узгоджена із транскрипцією і відбувається під час синтезу мРНК (див. розділи 5, 6). Більше того, рибосома здатна безпосередньо взаємодіяти із РНК-полімеразою, "підхоплюючи" мРНК одразу після її синтезу (рис. 8.28). При цьому мРНК, що синтезується на оперонах, часто містить кілька послідовних рамок зчитування і кілька стартових кодонів. Ініціація трансляції відбувається окремо на кожному з них усередині мРНК, і впізнання цих стартових кодонів не залежить від їхнього розташування відносно 5'-кінця матриці – у бактерій відбувається *внутрішня ініціація*.



**Рис. 8.28.** Комплекс бактеріальної РНК-полімерази з маленькою субодиноцею рибосоми (5M<sub>Y</sub>1)

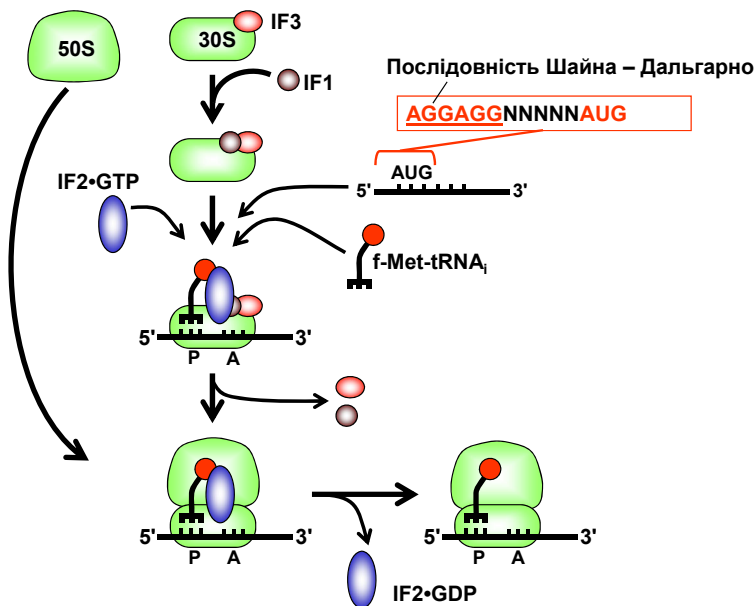
Послідовність подій при ініціації трансляції у бактерій показано на рис. 8.29. Рибосома вступає у процес у вигляді двох окремих субодиноць, дисоційований стан яких підтримується за-

вдяки білковому фактору IF3 (IF – Initiation Factor), зв'язаному з маленькою субодиноцею. У присутності IF3 з маленькою субодиноцею зв'язується також білок IF1, який блокує зону майбутнього А-сайта тРНК.

Потрійний комплекс маленької субодиноці з IF3 та IF1 взаємодіє із трьома іншими елементами:

- фактором ініціації IF2 – G-білком, який зв'язується з маленькою субодиноцею рибосоми в комплексі із GTP і нагадує за структурою елонгаційний фактор EF-G;
- ініціаторною молекулою тРНК – f-Met-тРНК<sub>i</sub><sup>Met</sup>, антикодонова частина якої взаємодіє з Р-сайтом маленької субодиноці, оскільки А-сайт заблокований фактором IF1;
- мРНК, яка взаємодіє з маленькою субодиноцею у такий спосіб, що стартовий кодон розміщується в зоні Р-сайта. Загалом із маленькою субодиноцею взаємодіє ділянка мРНК завдовжки ~30 нуклеотидів. Специфічне розміщення стартового кодону в Р-сайті забезпечується комплементарним упізнанням між ділянкою рРНК 16S і консервативною послідовністю Шайна – Дальгарно (John Shine, Lynn Dalgarno), яка розміщена в мРНК за 5–9 нуклеотидів від стартового кодону в напрямку до 5'-кінця. Саме наявність послідовності Шайна – Дальгарно робить зазначений кодон AUG стартовим, відрізняючи його від звичайного метіонінового.

Зв'язування трьох зазначених елементів, яке спричиняє утворення преініціаторного комплексу, відбувається незалежно один від одного й у будь-якому порядку, зокрема ініціаторна тРНК взаємодіє із преініціаторним комплексом незалежно від стартового кодону. Наступним кроком є впізнання кодону антикодоном у Р-сайті, якому сприяють фактори IF2 і IF3. Утворення кодон-антикодонової подвійної спіралі індукує конформаційну зміну маленької субодиноці, унаслідок чого відбувається звільнення факторів IF1, IF3 і зв'язування великої рибосомної субодиноці. Взаємодія між великою субодиноцею та IF2 індукує GTPазну активність останнього: відбувається гідроліз GTP і дисоціація IF2.



**Рис. 8.29.** Послідовність основних подій при ініціації трансляції у бактерій

У результаті система трансляції готова для першого елонгаційного циклу: А-сайт наготові прийняти aa-tРНК із другою амінокислотою поліпептидного ланцюга, на яку при транспептидації буде перенесено f-Met з ініціаторної tРНК.

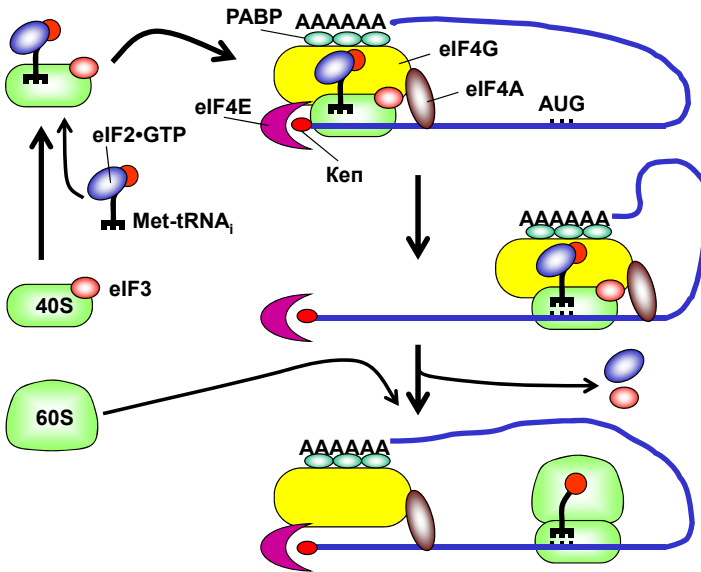
Слід зауважити, що частина бактеріальних мРНК не містить послідовності Шайна – Дальгарно. Механізми ініціації трансляції на таких мРНК ще недостатньо з'ясовані.

### **Ініціація трансляції в еукаріотів**

На відміну від прокаріотів, еукаріотична мРНК після синтезу транспортується з ядра до цитоплазми, де й відбувається ініціація. Крім того, кожна молекула мРНК містить тільки одну рамку зчитування (один стартовий кодон). Основний шлях ініціації трансляції в еукаріотів передбачає взаємодію елементів системи ініціації спочатку з кепом на 5'-кінці молекули, після чого відбувається сканування матриці з метою пошуку стартового коду – ініціація здійснюється за *сканувальним механізмом*.

Загалом у процесі беруть участь 12 або більше факторів ініціації (не всі з них остаточно ідентифіковані). Еукаріотична рибосома поза процесом трансляції також дисоційована на дві субодиниці завдяки взаємодії маленької із фактором eIF3 (складається із 9–11 субодиниць). Крім того, дисоційований стан субодиниць підтримується кількома іншими факторами, що взаємодіють із маленькою і великою субодиницями. Із комплексом маленька субодиниця-eIF3 зв'язується потрійний комплекс ініціаторної метіонінової тРНК з eIF2·GTP (eIF2 є аналогом фактора елонгації eEF1).

Паралельно (рис. 8.30) відбувається впізнання кепа на 5'-кінці мРНК факторами ініціації групи eIF4: субодиниця eIF4E зв'язується з кепом, eIF4G є структурним модулем, що забезпечує цілісність комплексу та його взаємодію з іншими елементами ініціації, eIF4A – АТФ-залежна РНК-геліказ, фактор eIF4B (не показаний на рис. 8.30) взаємодіє із мРНК, стимулює АТФазну активність гелікази та при скануванні мРНК бере участь у впізнанні стартового кодону.



**Рис. 8.30.** Послідовність основних подій при ініціації трансляції в еукаріотів

Наступним кроком є об'єднання кеп-асоційованого комплексу з таким, що сформований за участю маленької субодиноці (ініціаторна тРНК має бути наявною на маленькій субодиноці, щоб остання могла приєднатися до кеп). Ефективність утворення преініціаторного комплексу на кепі підсилюється за рахунок взаємодії polyA-зв'язаних білків (РАВР) з eIF4G: polyA-послідовність на 3'-кінці мРНК також приєднується до преініціаторного комплексу, замикаючи мРНК у кільце.

Далі за рахунок геліказної активності eIF4A преініціаторний комплекс (залишаючи на кепі лише eIF4E) починає транслокацію: eIF4A здійснює гідроліз АТФ і пересувається уздовж мРНК у напрямку до 3'-кінця, одночасно руйнуючи дволанцюгові шпильки, якщо вони зустрічаються. Під час транслокації преініціаторний комплекс сканує послідовність мРНК, перевіряючи її на наявність стартового кодону. Зовсім не обов'язково при цьому перший кодон AUG, що зустрічається, сприймається як стартовий. Упізнання стартового кодону залежить від контексту послідовності, у якій він розташований. Найкращим контекстом, який максимально сприяє ініціації трансляції, є *послідовність Козак* (Marilyn Kozak): GCC(A/G)CCAAUGG.

Упізнання стартового кодону викликає відповідні конформаційні зміни комплексу, звільнення факторів eIF3 і eIF4 (останній повертається на кеп разом із polyA-ділянкою мРНК), приєднання великої рибосомної субодиноці й дисоціацію eIF2·GDP після гідролізу GTP. Дисоціація eIF3 і eIF2 полегшується фактором eIF5. Рибосома після цього вступає в етап елонгації трансляції, а на кепі починається збирання наступної рибосоми.

Отже, на мРНК одночасно може працювати кілька рибосом, які утворюють *полісому*. Зациклення полісоми: 1) сприяє залученню зруйнованої при термінації трансляції у 3'-кінцевій зоні мРНК рибосоми до нового раунду ініціації; 2) забезпечує додатковий захист 3'-кінця мРНК (унаслідок його взаємодії з кеп-асоційованим комплексом) від нуклеазної деградації; 3) підвищує надійність системи трансляції: оскільки деградація мРНК може розпочинатися тільки із 3'-кінця (див. нижче), то лише повноцінні

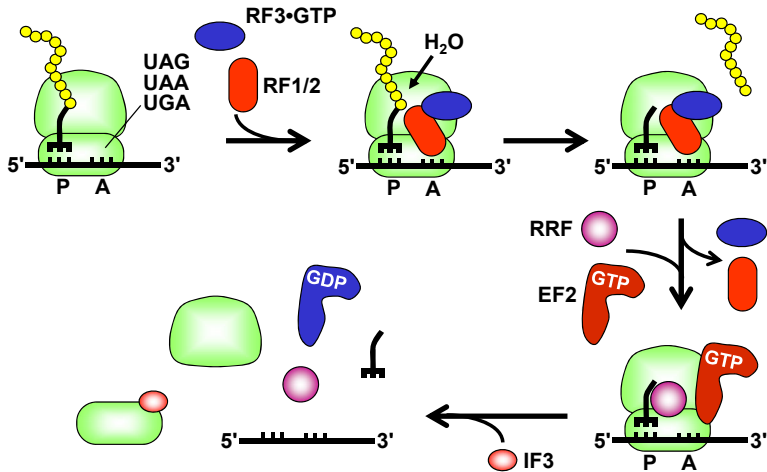
матриці, що містять polyA-послідовність (і автоматично – кодуювальну частину), залучаються до білкового синтезу.

Крім описаного вище сканувального механізму, в еукаріотів реалізується внутрішня ініціація трансляції – на 5–10 % мРНК здійснюється незалежно від кепа завантаження рибосоми. Найперед це властиво вірусним мРНК, але й доволі значній частині власних мРНК клітини. Внутрішня ініціація в еукаріотів залежить від упізнання маленькою субодиницею у складі преініціаторного комплексу особливої просторової структури у 5'-кінцевій нетрансльованій зоні мРНК – сайту внутрішнього завантаження рибосоми (IRES, Internal Ribosome Entry Site).

## Термінація трансляції

Термінація білкового синтезу здійснюється за цілком подібною для бактерій і еукаріотів схемою (рис. 8.31). Коли у бактерій після чергового елонгаційного циклу (який стане останнім) в А-сайті опиняється один із трьох стоп-кодонів, він упізнається одним із двох факторів термінації трансляції – RF1 або RF2 (RF – Release Factor). Жодна тРНК не містить відповідних антикодонів. Фактор RF1 упізнає стоп-кодони UAG/UAA, RF2 – UGA/UAA. Фактор RF1/2 також рекрутує до рибосоми комплекс RF3·GTP (аналог EF1), який потім допомагає дисоціації RF1/2.

Фактори RF1/2 не тільки впізнають стоп-кодон, а й індукують активність пептидил-трансферазного центра рибосоми. Оскільки в А-сайті при цьому немає звичайного субстрату транспептидації, то відбувається перенесення пептидилу на молекулу води – гідроліз зв'язку С-кінцевої амінокислоти з рибозою у складі пептидил-тРНК. На цьому закінчується власне термінація трансляції – з рибосоми звільняється синтезований поліпептид. Наступні події стосуються вже підготовки рибосоми до нового раунду трансляції.



**Рис. 8.31.** Послідовність основних подій при термінації трансляції у бактерій

Після визволення поліпептиду відбувається дисоціація RF1/2 і, після гідролізу GTP, – RF3. Натомість із рибосомою зв'язуються фактор відновлення рибосоми RRF (Ribosome Recycling Factor) і фактор елонгації EF-G у комплексі із GTP. Сумісна дія цих факторів у координації з гідролізом GTP спричиняє дисоціацію рибосомних субодиниць. Із маленькою субодиницею замість дисоційованих RRF і EF2 зв'язується фактор ініціації трансляції IF3 (який, таким чином, можна вважати й фактором термінації). IF3 сприяє визволенню тРНК і мРНК із маленької субодиниці й запобігає взаємодії між субодиницями. У результаті субодиниці рибосоми знов можуть залучитися до ініціації трансляції.

В еукаріотів усі три стоп-кодони упізнаються фактором eRF1, який зв'язується з рибосомою у складі комплексу з фактором eRF3 (останній у комплексі із GTP). Після впізнання кодону й гідролізу GTP eRF1 індукує активність пептидилтрансферазного центра, що спричиняє звільнення поліпептиду. Далі особливі еукаріотичні фактори беруть участь у відновленні рибосоми до нового раунду трансляції.

Із загального правила щодо відсутності тРНК, які могли б упізнавати стоп-кодони, є один виняток. Таким винятком є залучення до синтезу кількох важливих білків (це наявні в усіх організмів ферменти, що належать до класу оксидоредуктаз) модифікованої амінокислоти селеноцистеїну, яка є елементом активного центра зазначених ферментів. Селеноцистеїнова тРНК (продукт особливого тРНК-гена) здатна впізнавати внутрішній стоп-кодон UGA, фланкований (у напрямку до 3'-кінця) специфічною шпилькою. До цієї тРНК спочатку приєднується Ser, який далі перетворюється на селеноцистеїн відповідними ферментами (відбувається заміна ОН-групи на групу SeH, див. рис. 2.2). Селеноцистеїн-тРНК зв'язується зі специфічним аналогом елонгаційного фактора EF1 (позначається SelB у бактерій і SECIS – в еукаріотів). Шпилька в мРНК упізнається або самим специфічним фактором елонгації (бактерії), або ще одним білком, з яким цей фактор елонгації взаємодіє (еукаріоти), що й забезпечує зв'язування селеноцистеїн-тРНК з А-сайтом рибосоми і, відповідно, включення селеноцистеїну до ланцюга, що синтезується.

## Регуляція трансляції

Прокаріотична мРНК має доволі невеликий час життя – практично вона існує під час транскрипції й одразу після неї. Регуляція експресії гена у прокаріотів здійснюється переважно на рівні транскрипції, хоча білковий синтез може бути залученим до такої регуляції (див. розділ 6). Для еукаріотів регуляція трансляції є, навпаки, важливим окремим елементом загальної регуляції експресії генів.

**Регуляція на рівні ініціації трансляції** – один із найважливіших механізмів. Ефективність ініціації трансляції залежить насамперед від:

- контексту послідовності, у якій міститься стартовий кодон – відхилення цього контексту від послідовності Козак (див. вище) утруднює впізнання стартової точки і, відповідно, зумовлює необхідність позитивної регуляції;

- відстані стартового кодону від 5'-кінця мРНК – тривале сканування матриці під час ініціації підвищує імовірність руйнування преініціаторного комплексу;
- наявності / відсутності дволанцюгових шпильок у 5'-кінцевій зоні мРНК, що не транслюється, які гальмують процес сканування;
- присутності / відсутності регуляторних білків, які зв'язуються у 5'-кінцевій зоні.

Один із прикладів участі білків у негативній регуляції ініціації трансляції наведено на рис. 8.32. Білок, який утримує залізо у клітині, – феритин – має з'являтися негайно після проникнення в цитоплазму цього необхідного, але отруйного у вільному стані мікроелемента. Тому мРНК феритину завжди міститься в цитоплазмі, за відсутності заліза – у неактивному стані. Інактивація забезпечується за рахунок зв'язування репресора (який блокує процес сканування) зі шпилькою у 5'-кінцевій зоні.

Репресором виступає аконітаза – один із ферментів циклу Кребса (Hans Adolf Krebs). Залізо зв'язується аконітазою, що викликає її звільнення від мРНК і, як наслідок, активацію трансляції феритину. Синтезований феритин зв'язує залізо, забираючи його в аконітази, яка повертається на мРНК і зупиняє трансляцію.

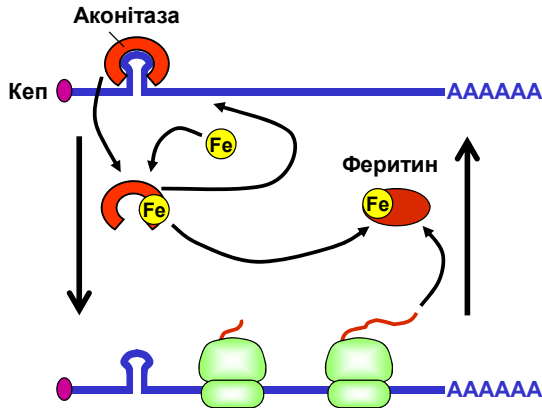


Рис. 8.32. Регуляція синтезу феритину

Мішенню глобальної регуляції на рівні ініціації є також фактори ініціації. Наприклад, фосфорилування / дефосфорилування eIF-2 спричиняє, відповідно, його деактивацію / активацію.

Крім регуляторних білків, до регуляції трансляції залучаються регуляторні РНК. Серед них є такі, що мають стимулюючий вплив на трансляцію: РНК деяких вірусів, РНК-продукти транскрипції Alu-повторів гальмують фосфорилування eIF-2, сприяючи активації трансляції. Інші регуляторні РНК, навпаки, гальмують трансляцію (зокрема на етапі сканування матриці) за рахунок комплементарної взаємодії із мРНК або забезпечують внутрішнє, незалежне від сканування, завантаження рибосоми.

Особливо важливими для регуляції трансляції є мікроРНК (див. розділ 6). Основною мішенню мікроРНК у цитоплазмі є 3'-кінцева зона мРНК, що не транслюється. Зв'язування мікроРНК у цій зоні може індукувати розривання мРНК або стимулювати її деградацію (див. нижче). Однак, оскільки мРНК замкнена в кільце, то зв'язування мікроРНК у 3'-кінцевій зоні може впливати й на ефективність ініціації трансляції.

**Елонгація трансляції** також є мішенню регуляторних впливів. Швидкість елонгації варіює в доволі широких межах відносно середнього значення, іноді рибосома може зупинитися на певний час. Паузи відбуваються на кодонах, які відповідають мінорним тРНК, зміна концентрації яких є одним зі шляхів регуляції швидкості трансляції. Підвищення загальної швидкості трансляції пов'язане з адаптацією наборів aa-тРНК – узгодженням між частотою найуживаніших кодонів і відповідних ізоакцепторних тРНК.

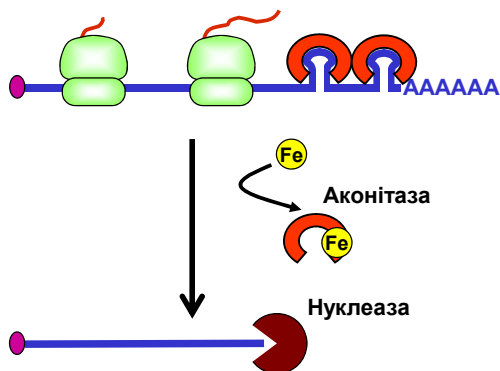
Наявність шпильок усередині кодувальної частини мРНК також гальмує елонгацію, як і репресори, які впізнають елементи послідовності мРНК.

Однією з мішеней регуляції є фактори елонгації, які, як і фактори ініціації, піддаються певним модифікаціям, що впливає на рівень їхньої активності. Екстремальний приклад – дифтерійний токсин, який інактивує фактор eEF-2 і повністю викликає білковий синтез.

**Зміна часу життя мРНК** – ще один важливий шлях регуляції трансляції. Час життя мРНК у цитоплазмі залежить пе-

реважно від ступеня захищеності 3'-кінця від екзонуклеаз, оскільки кеп до них нечутливий. При скороченні поліА-хвоста 3'-екзонуклеазами запускається процес деградації мРНК: основний шлях знищення мРНК здійснюється 3'-екзонуклеазами у складі спеціальних комплексів – *екзосом*. Скорочення поліА-хвоста може стимулювати відщеплення кепа спеціальним набором ферментів, після якого деградація може відбуватись і з 5'-кінця.

Отже, деградація мРНК у цитоплазмі починається зі скорочення її 3'-кінця, але ступінь його захищеності визначається наявністю / відсутністю тих чи інших факторів. Доволі типову ситуацію ілюструє приклад регуляції трансляції іншого білка, який має відношення до обміну заліза, – рецептора трансферину (рис. 8.33).



**Рис. 8.33.** Регуляція синтезу рецептора трансферину

Трансферин є позаклітинним переносником заліза, транспорт заліза в цитоплазму залежить від кількості його рецепторів у мембрані. За відсутності заліза всередині клітини на мРНК рецептора здійснюється синтез білка, мРНК при цьому стабілізована: аконітаза зв'язана зі шпильками у 3'-кінцевій зоні, що блокує нуклеазну деградацію. Зростання концентрації заліза викликає дисоціацію аконітази, і починається швидка деградація мРНК.

## Формування просторової структури білка

Термінація білкового синтезу на рибосомі ще не означає утворення функціонально активної молекули білка. Серед важливих операцій, яким піддається синтезований поліпептидний ланцюг, слід згадати такі, які можна об'єднати під назвою *білкового процесингу* – дозрівання поліпептидного ланцюга, що може включати такі процеси:

- часткову протеолітичну деградацію – відщеплення кінцевих ділянок ланцюга, частіше на N-кінці, або (іноді) розрізання ланцюга на окремі фрагменти. Часткова деградація особливо характерна для секреторних білків і ферментів, які мають гідролітичні активності, – така деградація переводить білок у активну форму;
- утворення дисульфідних містків – ковалентних S-S-зв'язків між двома наближеними у просторі залишками Cys. Формування містків прискорюється спеціальним ферментом – дисульфід-ізомеразою;
- ковалентні посттрансляційні модифікації амінокислотних залишків (фосфорилування, ацетилювання, глікозилювання тощо), а також приєднання полісахаридів, ліпідів, простетичних груп і кофакторів небілкової природи;
- білковий сплайсинг – доволі екзотичну операцію, описану для кількох десятків білків різних таксономічних груп, від бактерій до хребетних. У процесі такого сплайсингу центральна частина ланцюга (*інтеїн*) вирізається, а два кінцеві фрагменти (*екстеїни*) зшиваються між собою в автокаталітичній реакції. Реакція не потребує не тільки зовнішнього ферменту, а й джерела енергії: як і сплайсинг мРНК (розділ 7), білковий сплайсинг здійснюється шляхом заміни одних зв'язків іншими.

Кожна з перелічених операцій не є універсальною. Проте будь-який поліпептидний ланцюг, незалежно від того, потребує він процесингу чи ні, має сформувати просторову структуру для набуття певної активності (зрозуміло, що не йдеться про неструктуровані білки, згадані в розділі 2).

## Закономірності укладання білкової глобули

Перше й головне твердження щодо процесу укладання поліпептидного ланцюга в нативну глобулу – *білкового фолдингу* (folding) – полягає в тому, що *просторова структура нативного білка і шлях її формування повністю визначаються амінокислотною послідовністю, і тільки амінокислотною послідовністю!*

Процес укладання білка в нативну структуру – це пошук конформації, яка відповідає мінімуму вільної енергії для цієї послідовності амінокислот і, відповідно, є найстабільнішою. Певне уявлення про своєрідний "енергетичний ландшафт", що відповідає загальному набору різноманітних конформацій амінокислотного ланцюга, дає рис. 8.34 (зрозуміло, що насправді кількість конформаційних координат значно більше – конформаційний простір є багатовимірним). Нативна структура глобулярного білка завжди відокремлена від інших конформацій особливо глибоким мінімумом (перебуває в енергетичній ямі, рис. 8.34). Власне, наявність такої ями й відрізняє природні амінокислотні послідовності від великої кількості інших випадкових послідовностей (див. розділ 2). Глибока енергетична яма, яка відповідає стабільному щільно упакованому гідрофобному ядру, забезпечує надійність функціонування білка: теплові флуктуації не здатні вивести структуру із глибокого енергетичного мінімуму.



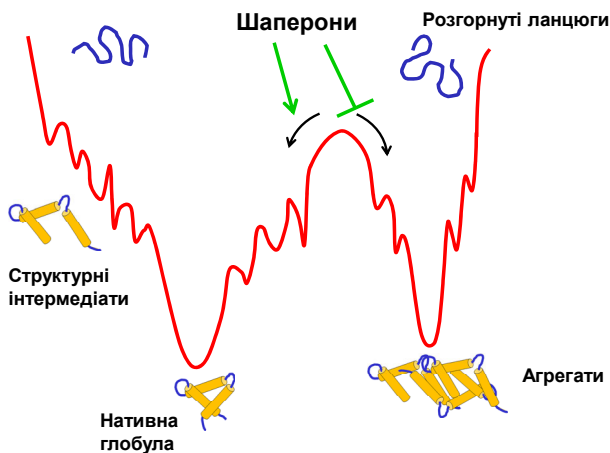
**Рис. 8.34.** Схематичний одинимірний "зріз" через конформаційний простір поліпептидного ланцюга – залежність вільної енергії від параметра (по осі абсцис), що характеризує конформацію ланцюга

Величезна кількість конформацій, теоретично доступних для поліпептидного ланцюга, зумовила проблему, відому як *парадокс Левінтала* (Cyrus Levinthal). Якщо кожен амінокислотний залишок може набувати  $\sim 10$  конформацій, то ланцюг зі 100 залишків –  $10^{100}$  конформацій. За найнижчою оцінкою, перемикання однієї конформації відбувається за  $\sim 10^{-13}$  с, тобто щоб перебрати всі конформації, потрібно щонайменше  $10^{80}$  років, що значно перевищує вік нашого Всесвіту. Однак аналогія конформаційного простору з ландшафтом одразу розв'язує цей парадокс: річка не перебирає весь ландшафт, вона просто тече одним або кількома альтернативними низькими руслами, впадаючи врешті-решт у найнижчу яму.

Так само й для поліпептидного ланцюга існують певні "виділені" шляхи укладання, які відповідають ієрархії структури білка (див. розділ 2). Спочатку за рахунок локальних взаємодій утворюються елементи вторинної структури з гідрофобними поверхнями. Завдяки гідрофобним взаємодіям вони "злипаються" між собою, утворюючи розплавлену глобулу. Цей процес відбувається доволі швидко – за кілька мікросекунд. На другому етапі, який потребує значно більшого часу (кілька секунд чи навіть хвилин), реалізується щільне укладання глобули за рахунок вандерваальсових взаємодій з утворенням твердої нативної молекули, що відповідає глобальному мінімуму вільної енергії. Конформація ланцюга змінюється, прямуючи через локальні мінімуми енергії, які відокремлені один від одного невисокими бар'єрами (рис. 8.34). Бар'єри легко долаються за рахунок теплових флуктуацій, і ланцюг нарешті опиняється у глибокому глобальному мінімумі.

Однак на цьому шляху можливі енергетичні пастки (рис. 8.35) – доволі глибокі локальні мінімуми, з яких неможливо вибратися за прийнятний час, оскільки глибина такої локальної ями перевищує енергію теплових флуктуацій. Головною причиною виникнення енергетичних пасток є агрегація поліпептидних ланцюгів. У стані розплавленої глобули, яка відрізняється від нативної великомасштабними рухами своїх частин, час від часу відбувається вихід на поверхню значної частини гідрофобних залишків. Отже, недоструктуровані поліпептидні ланцюги будуть прагнути утворити великі неспецифічні агрегати з метою закра-

нувати неполярні групи від молекул води. Агрегат і є пасткою – опиняючись у мінімумі вільної енергії, він практично не допускає дисоціації своїх елементів і не дозволяє їм продовжити пошук нативної структури. Зрозуміло, що ефективність агрегації зростає при підвищенні концентрації поліпептидів, а також при зростанні температури, яке підсилює і рухливість глобули, і гідрофобні взаємодії (див. розділ 1).



**Рис. 8.35.** Енергетичні пастки на шляху укладання поліпептидного ланцюга: шаперони допомагають уникнути таких пасток, чим сприяють пошуку нативної структури

Повертаючись до сформульованої вище догми про визначення шляху укладання білка його первинною структурою, слід зауважити, що процес укладання може просто не добігти кінця. Численні експерименти *in vitro* свідчать про те, що догма повністю підтверджується: практично будь-який білок у денатурованому стані можна без жодних специфічних факторів ренатурувати – відновити його нативну структуру й функцію. Однак успішна ренатурація можлива, якщо штучно створити умови й режим її виконання. Передусім бажано проводити ренатурацію поступово, повільно повертаючи зовнішні умови до фізіологічних. У результаті на проміжних етапах ренатурації будь-який локальний мінімум енергії не буде надто глибоким і не зможе

спрацювати як енергетична пастка: ланцюг швидко перебере різні не дуже стабільні конформації, реалізуючи конформаційну рівновагу, зсунуто до певного найнижчого мінімуму енергії. Другою важливою умовою є здійснення ренатурації за низької концентрації білка – з метою запобігти неспецифічній агрегації. Крім цих загальних правил, у кожному випадку часто доводиться підбирати умови ренатурації окремо: визначати оптимальний режим зміни складу розчинника, температуру, концентрації субодиниць білка із четвертинною структурою тощо. Особливо проблематичним процес ренатурації є для білків великого розміру: легко і швидко ренатурують білки з ланцюгом завдовжки до ~100 амінокислотних залишків.

Зрозуміло, що вказані вище умови ренатурації важко забезпечити *in vivo* – достатньо сказати, що загальна середня концентрація білків у клітині становить ~300 мг/мл або більше, а середня довжина ланцюга для білків людини – близько 550 залишків. Отже, клітинні умови дуже агресивні для поліпептидного ланцюга, що перебуває у процесі пошуку своєї нативної конформації, і без спеціальних механізмів більшість білків просто не встигли б після свого синтезу таку конформацію знайти, швидко опинившись у складі агрегату. Такі механізми – створення рівноважних умов поступового пошуку нативної структури й захист від агрегації – забезпечуються спеціальними білками – *шаперонами* (chaperon (фр.) – "капюшон", або "особа, що супроводжує незайману дівчину").

*Шаперони не визначають ані нативну структуру білка, ані шлях її укладання – те та інше детермінується амінокислотною послідовністю. Головна функція шаперонів – забезпечити умови для швидкого пошуку нативної конформації, створюючи своєрідний інкубатор для неструктурованого поліпептидного ланцюга, що запобігає включенню такого ланцюга до складу агрегатів.*

Серед доволі великої кількості шаперонів є специфічні, зокрема такі, що обслуговують процеси збирання мультибілкових комплексів і субодиничних білків. Більш універсальні шаперони відносять до родини **Hsp** – *білків теплового шоку* (heat shock proteins). Білки Hsp є загальним компонентом клітин для всіх

таксономічних груп. У пойкилотермних організмів транскрипція на генах *hsp* активується у відповідь на зростання температури до 30–40 °С при гальмуванні транскрипції багатьох інших генів (тепловий шок). Крім того, шаперони ще одного універсального класу зв'язані з рибосомою, зустрічаючи поліпептидний ланцюг на виході з неї.

### **Котрансляційний білковий фолдинг**

Тунель у тілі великої субодиниці, через який синтезований ланцюг виходить за межі рибосоми, має довжину  $\sim 100$  Å й середній діаметр  $\sim 15$  Å – у тунелі можуть розміститися тільки витягнутий ланцюг завдовжки  $\sim 30$  амінокислотних залишків або сформуватись елементи  $\alpha$ -спіралі. Отже, повноцінний фолдинг не може починатися в тунелі, а тільки на виході з нього. Однак є суттєва різниця між таким фолдингом і ренатурацією *in vitro*: у другому випадку весь ланцюг одночасно здійснює пошук своєї структури, у першому – фолдинг починається в N-кінцевій частині ланцюга, яка вже вийшла з тунелю, тоді як C-кінцева частина ще не може бути залученою до цього процесу. Таким чином, поліпептидний ланцюг, який ще синтезується, вже починає структуруватись зі свого N-кінця. При цьому рибосома (просто тому, що ланцюг залишається приєднаним до неї) сприяє ізоляції ланцюга від його оточення, створюючи умови для успішного пошуку просторової структури.

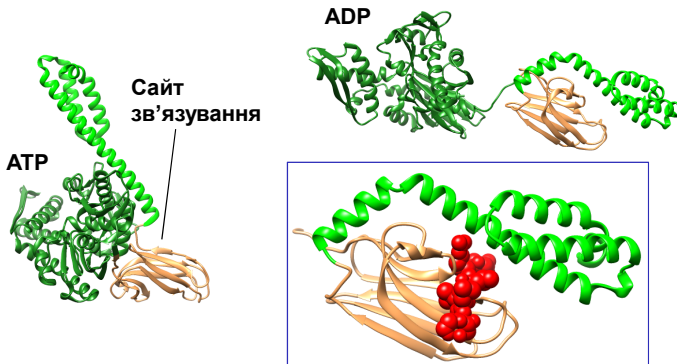
Більше того, з рибосомою на виході з тунелю взаємодіють допоміжні фактори, які служать шаперонами: тригер-фактор (TF, Trigger Factor) у бактерій і білкові комплекси RAC (Ribosome-Associated Complex) та NAC (Nascent-Chain-Associated Complex) в еукаріотів. Ці білки взаємодіють із гідрофобними ділянками поліпептиду, що виходить із рибосоми, додатково перешкоджаючи агрегації. Для багатьох білків маленького розміру з однодоменною структурою такого простого механізму виявляється достатньо для ефективного утворення нативної просторової структури у процесі білкового синтезу. Однак білки більшого розміру потребують "допомоги" щодо утворення своєї просторової структури вже після звільнення з рибосоми.

## Шаперони Hsp70

Білки Hsp70 (тут і далі цифрами позначено молекулярну вагу в кілодальтонах) зв'язують недоструктурований поліпептидний ланцюг, тобто такий, де залишаються не занурені всередину гідрофобні ділянки, одразу після його звільнення з рибосоми.

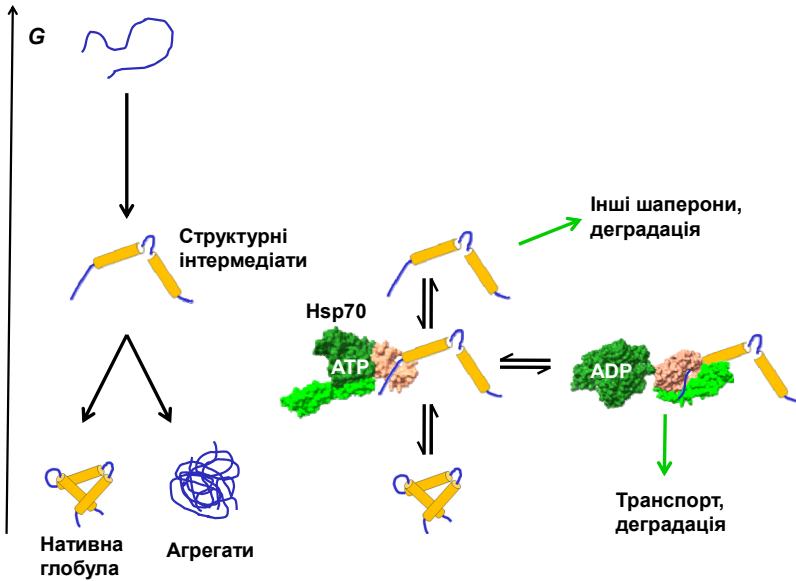
Молекула Hsp70 має два структурні домени, один з яких АТРазний, а інший, що складається із двох субдоменив, – пептид-зв'язувальний, який взаємодіє з гідрофобною ділянкою розгорнутого поліпептидного ланцюга (рис. 8.36). Зв'язування з гідрофобними ділянками підтримує розгорнутий стан поліпептиду – ланцюг вкривається "шубою" молекул шаперону. Власне, *підтримання розгорнутого стану ланцюга при запобіганні агрегації* – одна з основних функцій Hsp70.

Молекула шаперону може існувати у двох структурних станах залежно від типу зв'язаного ліганду – АТР чи АDР (рис. 8.36). У комплексі з АТР реалізується відкрита структурна форма, яка допускає швидку рівновагу між зв'язаним / дисоційованим поліпептидом. Гідроліз АТР (здійснюється АТРазним доменом Hsp70 за сприяння кошаперону Hsp40) викликає структурну зміну з міцною фіксацією поліпептиду пептид-зв'язувальним доменом. За допомогою інших кофакторів АDР знову замінюється на АТР.



**Рис. 8.36.** Бактеріальний аналог Hsp70 (білок DnaK) у відкритій (у комплексі з АТР, 4В9Q) і закритій (у комплексі з АDР, 2KHOK) структурних формах. Субдомени білка забарвлені у різні кольори. На вставці: комплекс пептид-зв'язувального домену із фрагментом поліпептиду (червоний) завдовжки 7 амінокислотних залишків (1DKZ)

У процесі такої циклічної АТР-залежної зміни спорідненості шаперону до поліпептиду створюються умови для рівноважного пошуку нативної конформації під час короткострокової дисоціації (рис. 8.37). Цей механізм цілком аналогічний такому, що розглядався в розділі 4 у зв'язку зі збиранням хроматину за допомогою проміжних переносників гістонів, які також називають *хроматиновими шаперонами* (див. рис. 4.10). Власне, відкрита форма Hsp70 і виконує основну роботу зі створення рівноважних умов: звільнений поліпептид із нестабільною структурою (з великою кількістю гідрофобних залишків на поверхні) знов зв'язується із Hsp70 і здійснює нову спробу при наступній дисоціації – у такій схемі для агрегатів не залишається місця (порівн. ліву і праву частини рис. 8.37). Для багатьох білків опосередкований Hsp70 пошук просторової структури – основний шлях фолдингу в клітині.



**Рис. 8.37.** Формування просторової структури глобулярного білка за відсутності (ліворуч) і за наявності Hsp70. G – вільна енергія

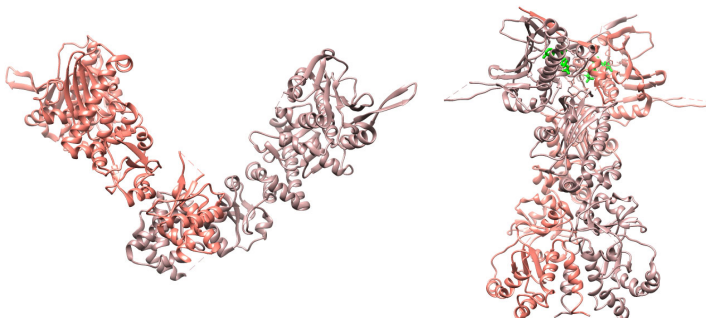
У складі міцного комплексу із Hsp70·ADP виключаються і фолдинг, і агрегація. При тривалому існуванні такі комплекси стають мішенями для протеолітичної деградації. Крім того, у

складі комплексу із Hsp70·ADP поліпептид транспортується крізь мембрану до інших компартментів клітини, усередину мітохондрій, до місця збирання мультибілкових комплексів тощо.

Крім зворотного зв'язування з відкритою формою Hsp70, для дисоційованого поліпептиду є ще дві можливості: частина таких недоструктурованих поліпептидів піддаються протеолітичній деградації, частина – переносяться на інші шаперони. Під іншими тут маємо на увазі специфічні шаперони – фактори збирання мультибілкових комплексів, а також дві системи універсальних шаперонів, які розглянемо нижче.

### Система Hsp90

Білок Hsp90 (гомодимер) може існувати в кількох структурних формах, залежних від ATP/ADP (рис. 8.38). Подібно до Hsp70, він може зв'язувати неструктуровані ділянки поліпептидів і звільняти їх. Hsp90, який діє за принципом шаперонів, виконує важливу функцію у сигнальних системах клітини: зв'язує дещо недоструктуровані (і, як наслідок, неактивні) білки й утримує їх у такому стані, а коли надходить певний сигнал – звільняє активний структурований білок. Приклад було наведено в розділі 6 у зв'язку з обговоренням транскрипційних факторів, які відповідають на дію стероїдних гормонів (гормонових рецепторів, див. рис. 6.14). Сигналом, що індукує визволення рецептора від Hsp90, у цьому випадку стає стероїдний гормон.



**Рис. 8.38.** Відкрита форма бактеріального аналога Hsp90 за відсутності ATP/ADP (2IQQ) і закрита форма еукаріотичного Hsp90 в комплексі із забарвленим зеленим аналогом ATP (2CG9)

Інший приклад стосується регуляції транскрипції генів білків теплового шоку пойкилотермних організмів. За низьких температур РНК-полімераза заарештована в паузі під час утворення кепа (див. розділ 7), а фактор транскрипції HSF (Heat Shock Factor), потрібний для її розблокування, перебуває в неактивній формі – у недоструктурованому стані в комплексі з білками Hsp40, Hsp70, Hsp90, які синтезуються за рахунок фонової активності своїх генів. Підвищення температури, яке зумовлює дестабілізацію багатьох білків і експонування гідрофобних груп у їхньому складі, викликає зв'язування Hsp із цими білками – HSF звільняється, набуває остаточної структури і зв'язується із промоторами генів *hsp*, активуючи їхню транскрипцію. Підвищення концентрації білків Hsp знов інактивує HSF.

### **Шапероніни**

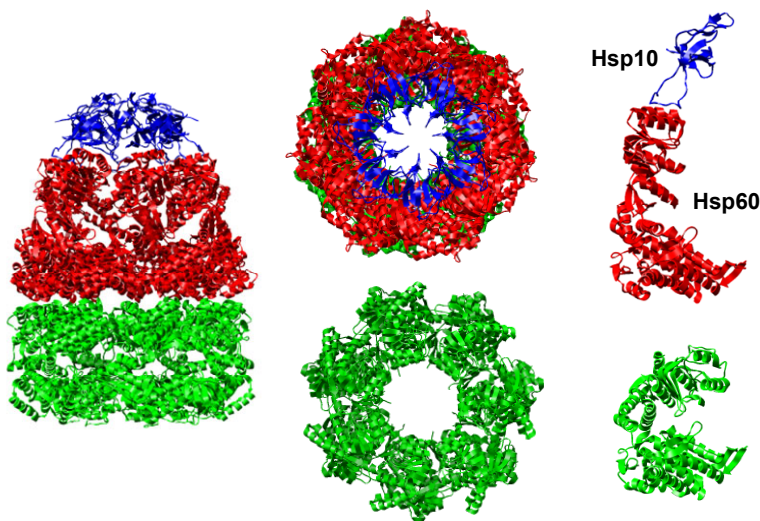
Особливий клас універсальних шаперонів, побудованих із білків Hsp60, об'єднують у групу шаперонінів. Шапероніни відповідають за фолдинг ~10 % білків у бактерій і еукаріотів. Бактеріальний шаперонін (шаперонін I) містить два комплекси, кожен з яких сформований сімома молекулами Hsp60 (інша назва GroEL). Комплекс має форму кільця з каналом ~45 Å у діаметрі, утворюючи своєрідну нанопробірку (рис. 8.39), яка може закриватися кришечкою із 7 молекул Hsp10 (інша назва GroES). До класу I належать також шапероніни мітохондрій і хлоропластів.

Еукаріотичні й архейні аналоги (шапероніни II, зокрема еукаріотичний TRiC) мають подібну будову, але кожне кільце формується із восьми або дев'яти субодиниць, додаткові структурні домени яких утворюють кришечку, що закривається / відкривається в результаті структурних перебудов.

Кожна субодиниця кільця, і на цьому базується принцип роботи шапероніну, існує принаймні у двох структурних формах (рис. 8.39). Перша (у двох варіантах, що трохи розрізняються) реалізується в комплексі з АТФ або АДФ і характеризується: 1) високою спорідненістю до Hsp10 (що сприяє утворенню кришечки); 2) низькою гідрофобністю своєї внутрішньої (всередині нанопробірки) поверхні. Дисоціація ліганду (АДФ після гідролізу АТФ) індукує структурну перебудову зі взаємним пе-

реміщенням трьох структурних доменів. При цьому втрачається спорідненість до Hsp10 (відкривання кришечки) і на внутрішню поверхню білка (усередині каналу нанопробірки) експонуються гідрофобні групи.

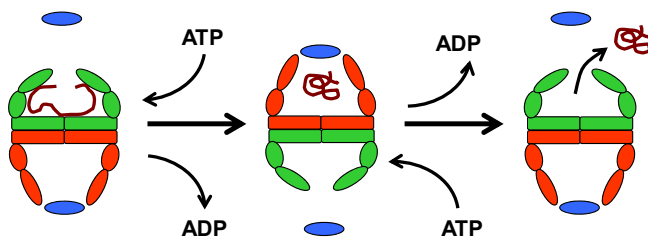
Зв'язування і гідроліз ATP, дисоціація ADP і відповідні перебудови відбуваються синхронно для семи субодиниць одного кільця (*позитивна* кооперативність у межах кільця), але два кільця працюють за принципом *негативної* кооперативності: перший стан одного кільця сприяє реалізації другого стану в іншому кільці (як на рис. 8.39).



**Рис. 8.39.** Структура шапероніну GroEL (1WE3) у різних проєкціях і його окремих субодиниць у двох структурних станах: зв'язаний з ADP Hsp60 (червоний) у комплексі з Hsp10 (синій); вільний від ліганду Hsp60 (зелений)

Робочий цикл шапероніну показано на рис. 8.40. Через відкритий отвір нанопробірки всередину потрапляє поліпептидний ланцюг. Кільце Hsp60 (верхнє на рис. 8.40, розглянемо саме його) перебуває при цьому у стані, що не зв'язує жодного ліганду: кришечка відкрита, поліпептид фіксується всередині каналу в розгорнутому вигляді за рахунок гідрофобних взаємодій із внут-

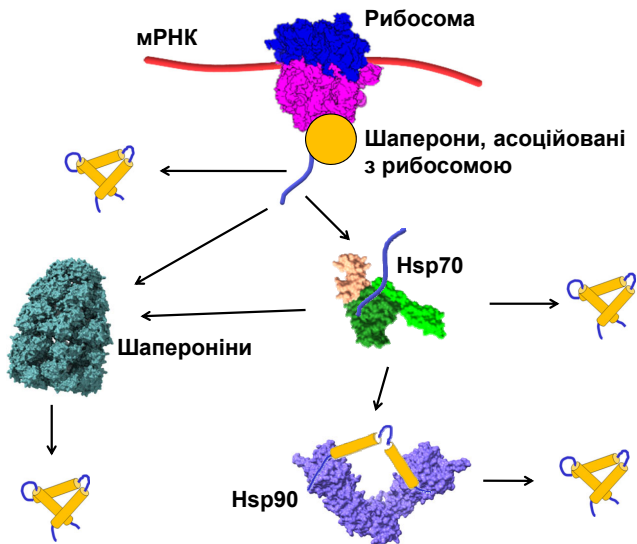
рішньою поверхнею пробірки. Зв'язування семи молекул АТР викликає структурну перебудову: кришечка закривається, гідрофобність внутрішньої поверхні зникає – поліпептид (повністю захищений від зовнішнього середовища) може здійснювати укладання. На це він має ~20 с – час, поки існує зв'язаний АТР, після чого відбувається його гідроліз (за рахунок активності Hsp60) і відтак залишається зв'язаний ADP. Зв'язування АТР із нижнім кільцем індукуює дисоціацію ADP від субодиниць верхнього: кришечка відкривається, і білок зі сформованою нативною структурою – гідрофобних груп практично немає на його поверхні й він не прилипає до стінок – виходить через отвір нанопробірки. Для більшості білків 20 с цілком вистачає, щоб знайти нативну конформацію. Якщо цього не відбулося, то білок залишається у нанопробірці й процес може повторитися.



**Рис. 8.40.** Робочий цикл шапероніну.

Структуру схематично зображено в розрізі: для кожного кільця показано по дві субодиниці, кожна складається із трьох доменів. Червоний і зелений кольори відповідають різним структурним станам кільця (як на рис. 8.39), синій – Hsp10

Основні шляхи, якими шаперони допомагають поліпептидним ланцюгам знайти нативну просторову структуру, підсумовує рис. 8.41. Для значної частини білків (до 70 % у бактерій) достатньо асоційованих із рибосомою шаперонів – практично з рибосом звільняється повністю готовий білок. Поза рибосомою центральну функцію виконує шаперон Hsp70: він і допомагає багатьом білкам сформувати нативну структуру, і передає поліпептидні ланцюги на сигнальні шаперони Hsp90, шапероніни та специфічні шаперони, і бере участь у транспортуванні поліпептидів в інші компартменти клітини.



**Рис. 8.41.** Основні шляхи білкового фолдингу за допомогою шаперонів

### Контрольні запитання

1. Назвіть основні структурні елементи тРНК. Як вони розташовані у просторі? Які взаємодії стабілізують просторову структуру тРНК?

2. Якою хімічною групою тРНК акцептується амінокислота? Назвіть групу амінокислоти, що задіяна в цьому зв'язку.

3. Які хімічні реакції каталізує АРСаза? Що таке активування амінокислоти?

4. Як реалізується специфічність АРСаз щодо тРНК і амінокислот?

5. Що таке рибосома? Із чого вона складається? Опишіть основні риси просторової будови рибосоми.

6. Які функціональні сайти має рибосома? Опишіть їхнє розташування відносно один одного та щодо субодиниць рибосоми.

7. Як розподілені функції рибосоми між двома субодиницями?

8. Порівняйте структурні й функціональні особливості рибосомних РНК і білків.
9. З яких етапів складається елонгаційний цикл рибосоми?
10. Опишіть структуру й функціональне значення факторів елонгації EF1 і EF2.
11. У чому полягає роль гідролізу GTP при елонгації трансляції?
12. Як рибосома здійснює декодувальну функцію?
13. Що таке акомодация aa-тРНК на рибосомі та як саме вона відбувається?
14. У чому полягає реакція подовження поліпептидного ланцюга на одну амінокислоту? Як відбувається каталіз цієї реакції?
15. Як відбувається транслокація рибосоми вздовж мРНК?
16. Яка стадія елонгаційного циклу найшвидша?
17. Порівняйте системи ініціації трансляції у бактерій і еукаріотів.
18. Опишіть основні стадії термінації трансляції.
19. За якими механізмами здійснюється регуляція білкового синтезу?
20. Назвіть основні стадії процесу укладання білкової глобули. Якою вільною енергією має характеризуватися нативна конформація білка порівняно з іншими конформаціями? Що таке енергетична пастка, як вона може виникати?
21. У чому полягає функціональна роль шаперонів?
22. Опишіть робочий цикл шаперонів Hsp70.
23. Охарактеризуйте функціональну роль шаперону Hsp90.
24. Опишіть структурну організацію і робочий цикл шаперонів.

## Рекомендована література

Andersen, G.R., Nissen, P. and Nyborg, J. (2003) Elongation factors in protein biosynthesis. *Trends Biochem. Sci.* 28, 434–441.

Balchin, D., Hayer-Hartl, M. and Hartl, F.U. (2016) In vivo aspects of protein folding and quality control. *Science* 353, aac4354.

Bukau, B. and Horwich, A.L. (1998) The hsp70 and hsp60 chaperone machines. *Cell* 92, 351–366.

Burbaum, J.J. and Schimmel, P. (1991) Structural relationships and the classification of aminoacyl-tRNA-synthetases. *J. Biol. Chem.* 266, 16965–16968.

Deuring, E. and Bukau, B. (2004) Chaperone-assisted folding of newly synthesized proteins in the cytosol. *Crit. Rev. Biochem. Mol. Biol.* 39, 261–277.

Dever, T.E., Dinman, J.D. and Green, R. (2018) Translation elongation and recoding in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 10, a032649.

Frank, J., Sengupta, J., Gao, H., et al. (2005) The role of tRNA as a molecular spring in decoding, accommodation, and peptidyl transfer. *FEBS Letters* 579, 959–962.

Genest, O., Wickner, S. and Doyle, S.M. (2019) Hsp90 and Hsp70 chaperones: collaborators in protein remodeling. *J. Biol. Chem.* 294, 2109–2120.

Hellen, C.U.T. (2018) Translation termination and ribosome recycling in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 10, a032656.

Laursen, B.S., Sørensen, H.P., Mortensen, K.K. and Sperling-Petersen, H.U. (2005) Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* 69, 101–123.

Merrick, W.C. and Pavitt, G.D. (2018) Protein synthesis initiation in eukaryotic cells. *Cold Spring Harb. Perspect. Biol.* 10, a033092.

Mirande, M. (2017) The aminoacyl-tRNA synthetase complex. *Subcell. Biochem.* 83, 505–522.

Mitra, K. and Frank, J. (2006) Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biochem. Biomol. Struct.* 35, 299–317.

Moore, P.B. and Steitz, T.A. (2003) The structural basis of large ribosomal subunit function. *Annu. Rev. Biochem.* 72, 813–850.

Ogle, J.M. and Ramakrishnan, V. (2005) Structural insights into translational fidelity. *Annu. Rev. Biochem.* 74, 129–177.

O'Reilly, F.J., Xue, L., Graziadei, A., et al. (2020) In-cell architecture of an actively transcribing-translating expressome. *Science* 369, 554–557.

Pearl, L.H. and Prodromou, C. (2006) Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.* 75, 271–294.

Phizicky, E.M. and Hopper, A.K. (2010) tRNA biology charges to the front. *Genes Dev.* 24, 1832–1860.

Prabhakar, A., Choi, J., Wang, J., et al. (2017) Dynamic basis of fidelity and speed in translation: coordinated multistep mechanisms of elongation and termination. *Protein Sci.* 26, 1352–1362.

Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell* 108, 557–572.

Rodnina, M.V. (2016) The ribosome in action: tuning of translational efficiency and protein folding. *Protein Sci.* 25, 1390–1406.

Rodnina, M.V. (2018) Translation in prokaryotes. *Cold Spring Harb. Perspect. Biol.* 10, a032664.

Rodnina, M.V., Korniy, N., Klimova, M., et al. (2020) Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res.* 48, 1056–1067.

Rubio Gomez, M.A. and Ibba, M. (2020) Aminoacyl-tRNA synthetases. *RNA* 26, 910–936.

Spirin, A.A. (2002) Ribosome as a molecular machine. *FEBS Letters* 514, 2–10.

Steitz, T.A. (2008) A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* 9, 242–253.

Wilden, B., Savelsbergh, A., Rodnina, M.V. and Wintermeyer, W. (2006) Role and timing of GTP binding and hydrolysis during EF2-dependent tRNA translocation on the ribosome. *Proc. Natl. Acad. Sci. USA* 103, 13670–13675.

Wintermeyer, W., Peske, F., Beringer, M., et al. (2004) Mechanisms of elongation on the ribosome: dynamics of a macromolecular machine. *Biochem. Soc. Trans.* 32, 733–737.

Young, J.C., Agashe, V.R., Siegers, K. and Hartl, F.U. (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nat. Rev. Mol. Cell Biol.* 5, 781–791.

## Розділ 9

# РЕПЛІКАЦІЯ ДНК

It has not escaped our notice that the specific base pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.  
*J. Watson, F. Crick*

Процес подвоєння ДНК – *реплікація* (replication) – забезпечує відтворення спадкової інформації і передавання її до дочірніх клітин при клітинному поділі. Синтез ДНК відбувається при реплікації з використанням обох полінуклеотидних ланцюгів як матриць за *напівконсервативним механізмом*: дві дочірні молекули-копії містять один материнський ланцюг (що служив матрицею) і один ланцюг, синтезований *de novo*. Включення нуклеотидів до ланцюга, що синтезується, детермінується матрицею за принципом комплементарності. Такий механізм реплікації став зрозумілим одразу після того, як Уотсоном і Кріком була запропонована модель подвійної спіралі ДНК.

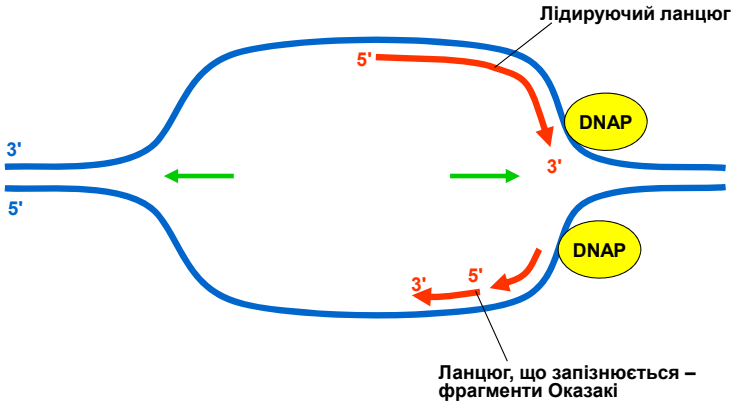
Як і при синтезі РНК (розділ 5), зростання ланцюга ДНК відбувається в напрямку від 5' до 3'-кінця. Субстратами реакції є 3'-кінцева ОН-група дезоксирибози зростаючого ланцюга й дезоксирибонуклеозидтрифосфати (dNTP). Фермент, що каталізує цю реакцію, – *ДНК-залежна ДНК-полімераза* (DNA dependent DNA Polymerase, DNAP).

Молекулярні принципи реплікації є спільними для всіх живих організмів. На початку розділу буде розглянуто переважно бактеріальну систему реплікації, далі – особливості реплікації в еукаріотів, а також репаративні процеси, що в більшості випадків пов'язані із синтезом ДНК.

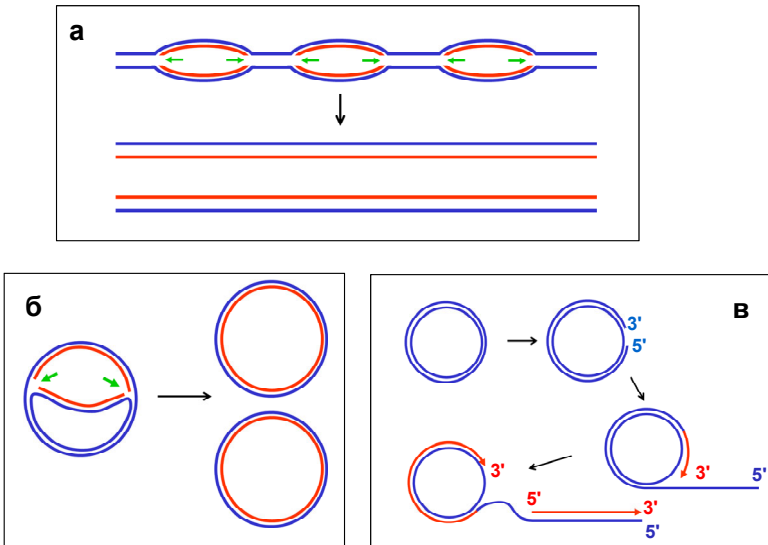
## Реплікони

Реплікація ДНК починається з невеликої ділянки – *ориджину* (origin), де здійснюється ініціація процесу, головним моментом якої є розходження ланцюгів ДНК. Далі в ході реплікації такий *реплікативний міхур* (рис. 9.1) розростається у двох протилежних напрямках. На кожному боці міхура існує *реплікативна вилка*, в основі якої відбувається синтез ДНК. Ділянку ДНК, де здійснюється реплікація, що розпочинається з одного ориджину, називають *репліконом*. Бактеріальна хромосома часто містить тільки один ориджин, тобто є єдиним репліконом (зокрема, монорепліконною є хромосома *E. coli*). У деяких бактерій може бути два реплікони на хромосому. Еукаріотична хромосома є полірепліконом – вона містить велику кількість точок ініціації (рис. 9.2).

Оскільки синтезуються два ланцюги ДНК, то в кожній реплікативній вилці мають працювати принаймні дві ДНК-полімерази. Оскільки два ланцюги антипаралельні (див. розділ 3) і синтез відбувається тільки в напрямку від 5'- до 3'-кінця, то синтез тільки одного з ланцюгів може відбуватися (і відбувається) безперервно, починаючись від ориджину (рис. 9.1). Цей ланцюг називають *лідуючим* (leading strand), його 3'-кінець розташований поблизу від основи реплікативної вилки. Синтез іншого ланцюга розпочинається від реплікативної вилки: синтезуються окремі фрагменти – *фрагменти Оказакі* (Reiji Okazaki), які згодом з'єднуються між собою. Для синтезу кожного із фрагментів треба спочатку звільнити відповідний простір на матричному ланцюзі – пересунути реплікативну вилку вперед (рис. 9.1), тому фрагментарний ланцюг називають *ланцюгом, що запізнюється* (lagging strand). Середня швидкість синтезу ДНК при реплікації становить ~1000 нуклеотидів за секунду в бактерій, ~300 нуклеотидів за секунду в археїв (загалом архейна реплікативна молекулярна система дуже подібна до еукаріотичної) і ~30 нуклеотидів за секунду в еукаріотів. Синтез бактеріальної хромосоми відбувається за ~50 хв, повна реплікація ДНК еукаріотичної клітини потребує близько 8 год.



**Рис. 9.1.** Реплікативний міхур – дві реплікативні вилки, що переміщуються у протилежних напрямках (ланцюги, що синтезуються, і ДНК-полімерази (DNAP) показано тільки для однієї з них)



**Рис. 9.2.** Реплікація еукаріотичної хромосоми (а), циркулярної бактеріальної хромосоми (б) і циркулярної ДНК деяких бактеріофагів (в)

У більшості репліконів реплікація здійснюється так, як зображено на рис. 9.1, тобто в обох напрямках. Сусідні реплікони еукаріотичної хромосоми врешті-решт зустрічаються, у результаті чого утворюються дві копії ДНК хромосоми (рис. 9.2, *а*). Циркулярна бактеріальна хромосома також реплікується у двох напрямках з утворенням ідентичних циркулярних молекул ДНК (рис. 9.2, *б*). Аналогічно реплікуються окремі автономні реплікони бактерій – циркулярні плазміди. Для деяких із них реплікація відбувається тільки в одному напрямку: в ориджині формується лише одна реплікативна вилка, яка рухається навколо кільця до вихідної точки.

Особливий випадок – реплікація циркулярних ДНК за механізмом *кільця, що котиться* (rolling circle), який реалізується при реплікації ДНК бактеріофагів і деяких плазмід. Один із варіантів такого механізму (що здійснюється, зокрема, для бактеріофага  $\lambda$ ) показано на рис. 9.2, *в*. В одному з ланцюгів циркулярної ДНК індукується одноланцюговий розріз, 3'-кінець, що утворився, добудовується з використанням інтактного циркулярного ланцюга як матриці. Продовження цього процесу створює довгий одноланцюговий 5'-кінцевий хвіст, що складається з тандемних повторів – копій кільцевої молекули. Кожен із цих повторів використовується як матриця для синтезу іншого ланцюга. Після цього лінійні дволанцюгові фрагменти вирізаються й упаковуються до частинок бактеріофага.

## Бактеріальні ДНК-полімерази

У клітині *E. coli* працюють ДНК-полімерази трьох основних типів. Усі вони мають дві ферментативні активності: власне *полімеразну*, за рахунок якої до 3'-кінця ланцюга, що синтезується, приєднуються нуклеотиди, і *3'-екзонуклеазну*, яка використовується для редагування помилок – відщеплення помилкових нуклеотидів, щойно приєднаних до 3'-кінця.

- **ДНК-полімераза I**, або полімераза Корнберга (Arthur Kornberg), – мономерний білок із мультидоменною структурою. На відміну від інших ДНК-полімераз, вона має до-

даткову 5'-екзонуклеазну активність. ДНК-полімераза I використовується як допоміжна полімераза при реплікації (див. нижче) та інших процесах синтезу ДНК.

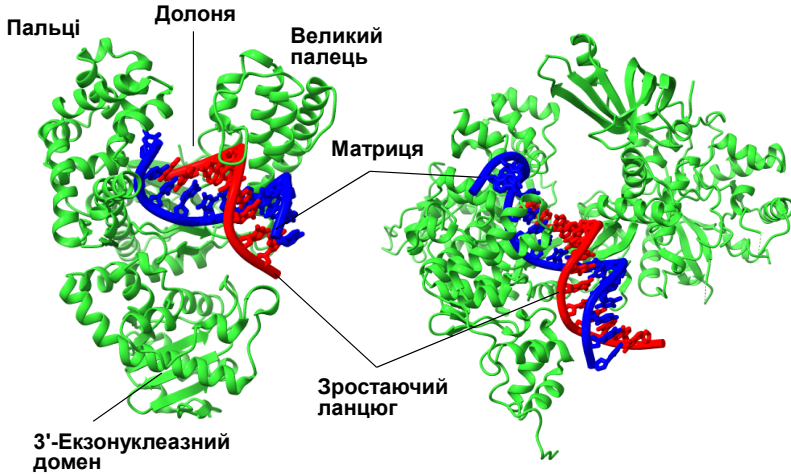
- **ДНК-полімераза II** також залучена до синтезу ДНК при реплікації, хоча й не є основним реплікативним ферментом. Від основної реплікативної ДНК-полімерази III ця полімераза відрізняється нижчою швидкістю синтезу, але підвищеною точністю. Вважають, що вона може замінювати основну реплікативну полімеразу у випадку, коли та зупиняється за різних причин або продукує велику кількість помилок.
- **ДНК-полімераза III** – основна реплікативна полімераза. Складається із трьох субодиниць: субодиниця  $\alpha$  відповідає за полімеразну активність,  $\epsilon$  – за 3'-екзонуклеазну,  $\theta$  – виконує структурну функцію.

Крім перелічених полімераз, у бактеріальних клітинах наявні також ДНК-полімерази IV та V – полімерази низької точності синтезу. Їхня основна функція полягає в тому, щоб забезпечити синтез ДНК у разі пошкодження матриці: зустрічаючи пошкодження, реплікативна машинерія високої точності (яка включає полімеразу III) зупиняється. Полімераза низької точності дозволяє долати пошкодження, які згодом можуть бути виправлені шляхом репарації, і знов замінюється на полімеразу III, що продовжує високоточний синтез ДНК.

### **Структура і полімеразна активність ДНК-полімераз**

Структуру основної частини ДНК-полімерази I – *фрагмента Кленова* (Hans Klenow) – показано на рис. 9.3 (видалений із цієї полімерази домен відповідає за 5'-екзонуклеазну активність). Полімеразний активний центр розташований у межах *долони* (palm) – щілини, яка оточена двома характерними структурними доменами: рухливими *пальцями* (fingers) і *великим пальцем* (thumb). Великий палець взаємодіє з маленьким жолобком подвійної спіралі, утвореної матричним ланцюгом і таким, що синтезується. Пальці взаємодіють із матричним ланцюгом, який при цьому трохи вигинається навкруг них, експонуючи чергову

азотисту основу для взаємодії з NTP. Подібну архітектуру, що нагадує долоню правої руки, мають також інші бактеріальні й еукаріотичні ДНК-полімерази. Приклад однієї з них представлено на рис. 9.3, див. також рис. 9.5.

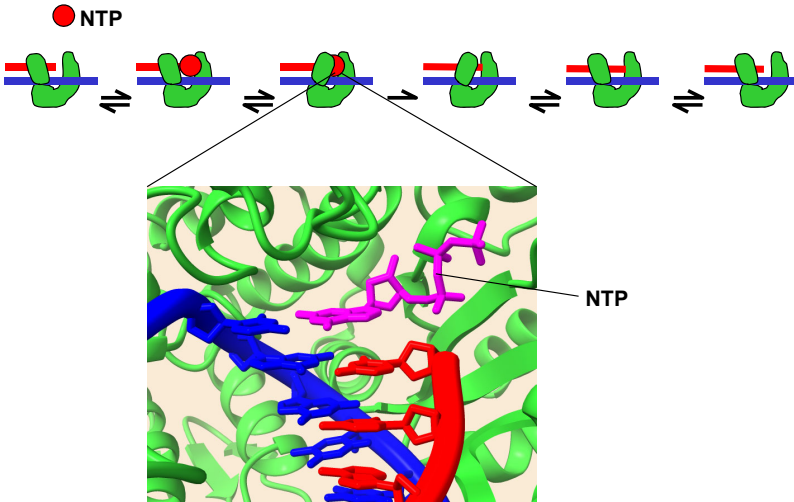


**Рис. 9.3.** Комплекси з ДНК бактеріальних ДНК-полімераз: ліворуч – фрагмент Кленова ДНК-полімерази I *E. coli* (1L3T); праворуч – полімеразна каталітична субодиноця  $\alpha$  ДНК-полімерази III *Geobacillus kaustophilus* (3F2B)

Конформаційна рухливість пальців передбачає два основні структурні стани полімерази: відкритий, коли фермент здатен здійснювати швидке перебирання нуклеозидтрифосфатів і рухатись уздовж матриці, і закритий, коли нуклеозидтрифосфат жорстко фіксується в активному центрі, що забезпечує каталіз реакції приєднання чергового нуклеотиду. Хімія цього процесу цілком подібна до реакції приєднання нуклеотидів при синтезі РНК. Механізм каталізу ДНК-полімеразної реакції також цілком аналогічний: ключова функція у каталізі належить двом іонам  $Mg^{2+}$ , що утримуються в активному центрі трьома залишками Asp (див. рис. 5.8).

Принципова схема елонгаційного циклу ДНК-полімерази також практично не відрізняється від такої РНК-полімерази (рис. 9.4). На першій стадії відбувається зв'язування NTP із відкритою формою полімерази. Якщо нуклеотид виявляється

комплементарним матриці, то пара основ, що виникла, індукує конформаційну зміну ферменту (найповільніша лімітувальна стадія циклу) із переміщенням пальців, наслідком чого стає фіксація NTP у активному центрі й нерухомість полімерази відносно ДНК. На третій стадії здійснюється каталіз реакції. Після приєднання нуклеотиду відбувається дисоціація пірофосфату й розмикання полімерази. Оскільки утворився новий подовжений 3'-кінець, який має спорідненість до активного центра, то на останній стадії здійснюється транслокація – переміщення полімерази в її відкритій формі на один нуклеотид уздовж матриці.

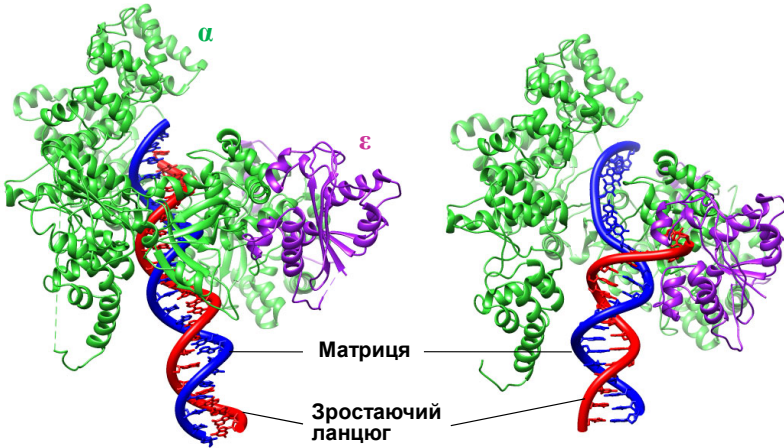


**Рис. 9.4.** Елонгаційний цикл ДНК-полімерази.  
На вставці: нуклеозидтрифосфат у активному центрі  
ДНК-полімерази (із правої структури на рис. 9.3)

### Редагування помилок

Упізнання комплементарної пари основ активним центром ДНК-полімерази перед каталізом полімеразної реакції забезпечує частоту помилкового включення нуклеотидів на рівні  $\sim 10^{-5}$ – $10^{-6}$  (основне джерело помилок – таутомерія азотистих основ, див. рис. 5.10). Як і РНК-полімерази, ДНК-полімерази здійснюють редагування помилок – у цьому випадку ефективний механізм редагування пов'язаний із 3'-екзонуклеазною активністю.

Розташований в окремому структурному домені або окремій субодиноці 3'-екзонуклеазний активний центр знаходиться на відстані до ~30–60 Å від полімеразного (рис. 9.5). Два каталітичні центри конкурують за 3'-кінець ланцюга, що синтезується.



**Рис. 9.5.** Дві субодиноці ДНК-полімерази III *E. coli*:  
 ліворуч – 3'-кінець ланцюга, що синтезується, міститься  
 в полімеразному активному центрі на субодиноці  $\alpha$  (5FKV);  
 праворуч – 3'-кінець у 3'-екзонуклеазному центрі  
 на субодиноці  $\epsilon$  (5M1S)

Якщо внаслідок приєднання помилкового нуклеотиду утворилася некомплементарна (нестабільна) пара основ, то її спорідненість до полімеразного центра знижується, а 3'-кінець переміщується до екзонуклеазного центра. Подвійна спіраль ДНК при цьому прокручується навколо своєї осі й дещо розплітається. Помилковий нуклеотид відщеплюється, унаслідок чого відновлюється стабільність кінцевої пари основ. Більш висока спорідненість стабільної пари до полімеразного центра викликає повернення туди 3'-кінця, і ДНК-полімераза здійснює нову спробу його подовження. У результаті такої осциляції полімерази з перемиканням активності між двома центрами рівень помилок знижується до  $\sim 10^{-8}$ . Оскільки ДНК синтезується "раз і назавжди" перед її передаванням нащадкам, то такий рівень не може вважатися задовільним. Остаточний рівень помилок знижується до  $\sim 10^{-10}$  за рахунок активності систем репарації.

## Особливості ДНК-полімерази порівняно з РНК-полімеразою

Незважаючи на високу подібність базових механізмів роботи двох типів полімераз, що здійснюють синтез нуклеїнових кислот, існують принципові відмінності між ними. Головна полягає в тому, що для ДНК-полімерази **ДНК є одночасно і матрицею, і продуктом реакції**, що створює суттєві проблеми.

Оскільки при синтезі РНК в активному центрі РНК-полімерази наявна гібридна подвійна спіраль ДНК-РНК (див. розділ 5), то РНК-полімераза може легко дискримінувати гібрид від звичайної подвійної спіралі ДНК. Висока спорідненість оточення активного центра РНК-полімерази до гібрида забезпечує високу процесивність ферменту – здатність працювати без дисоціації після однократного акта ініціації транскрипції. ДНК-полімераза має подвійну спіраль ДНК як в оточенні свого активного центра, так і скрізь поза полімеразним комплексом. Відповідно існує висока ймовірність її дисоціації: процесивність ДНК-полімерази дуже низька – вона може синтезувати до дисоціації лише ділянку завдовжки 10–20 нуклеотидів. Отже, потрібен певний додатковий механізм підвищення процесивності.

Висока спорідненість РНК-полімерази до гібрида ДНК-РНК має ще один наслідок: вона дозволяє легко руйнувати подвійну спіраль ДНК по ходу руху полімерази при елонгації транскрипції – транскрипт просто витісняє нематричний ланцюг ДНК із дуплекса. Для ДНК-полімерази такий механізм неможливий: дуплекси ДНК у комплексі з полімеразою і попереду від неї нічим не відрізняються один від одного, тобто ДНК-полімераза потребує наявності одноланцюгової матричної ДНК, яка має бути вилучена з подвійної спіралі.

Третя проблема полягає в тому, що ДНК-полімераза здатна виконувати тільки одну операцію – подовжувати (редагуючи) 3'-кінець ланцюга ДНК. Вона не може *ініціювати* синтез, створити перший фосфодієфірний зв'язок. Це означає, що певна коротка ділянка має бути створена якось інакше, щоб далі ДНК-полімераза могла продовжувати її синтез. Таку ділянку, без якої неможлива робота ДНК-полімерази, називають *праймером* (primer). Насправді кор-фермент РНК-полімерази також не здатен ініцію-

вати синтез РНК: первинний короткий транскрипт синтезується лише за участю факторів ініціації, які допомагають додатково зафіксувати перші нуклеотиди в активному центрі (див. розділ 5). При синтезі ДНК було знайдено інше еволюційне рішення: роль праймера виконує коротка ділянка РНК, що синтезується специфічною ДНК-залежною РНК-полімеразою – *праймазою*. Використання саме РНК у ролі праймера на початку синтезу ДНК має важливе значення для підвищення точності синтезу. Короткі первинні ділянки неможливо синтезувати з дуже високою точністю. На подальших етапах реплікації РНК-праймери можна легко відрізнити від ДНК, вилучити їх і заповнити прогалини між сусідніми ділянками ДНК за допомогою ДНК-полімерази.

Отже, ДНК-полімераза є хоч і головним, але не достатнім елементом системи реплікації. У реплікативній вилці працює складний мультибілковий комплекс – *реплісома*, до якої, крім ДНК-полімераз, входять компоненти, що забезпечують розплітання ДНК, підвищення процесивності й виконують інші важливі допоміжні операції.

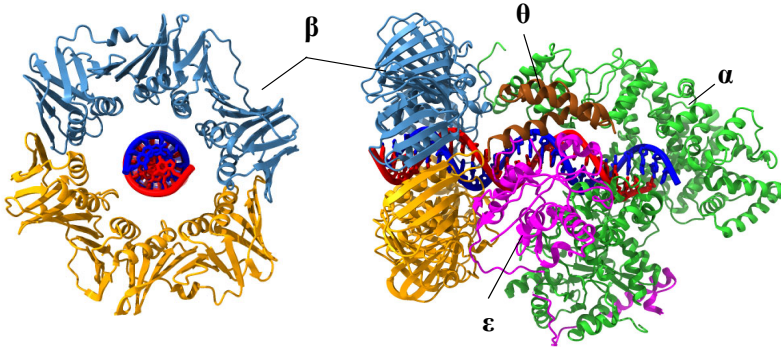
## **Бактеріальна реплісома та інші елементи системи реплікації**

### **Голофермент ДНК-полімерази**

Голофермент ДНК-полімерази, який є основою реплісоми, крім ДНК-полімерази III (три субодиниці – кор-фермент), містить  $\tau$ -білок,  $\gamma$ -комплекс (п'ять субодиниць) і  $\beta$ -білок. Білок  $\tau$  виконує основну структурну роль – взаємодіє із ДНК-полімеразою і  $\gamma$ -комплексом, кілька його молекул взаємодіють між собою, утримуючи разом і забезпечуючи координацію роботи всіх компонентів реплісоми (див. нижче).

Підвищення процесивності ДНК-полімерази здійснюється за допомогою  $\beta$ -білка, який утворює *ковзний обруч* (*хомут, затискач, sliding clamp*). Дві однакові субодиниці обруча, щільно взаємодіючи за принципом "голова до хвоста", утворюють циркулярну тороподібну структуру з отвором приблизно 35 Å в діаметрі (рис. 9.6). Усередині такого кільця легко розміщується по-

двійна спіраль ДНК, яка практично не взаємодіє із  $\beta$ -білком. Кільце може вільно рухатись уздовж ДНК, але не може дисоціювати. При цьому ДНК-полімераза III взаємодіє з обручем. У результаті обруч тримає полімеразу "на повідку" – допускає певну свободу її рухів, але не дозволяє дисоціювати від ДНК.

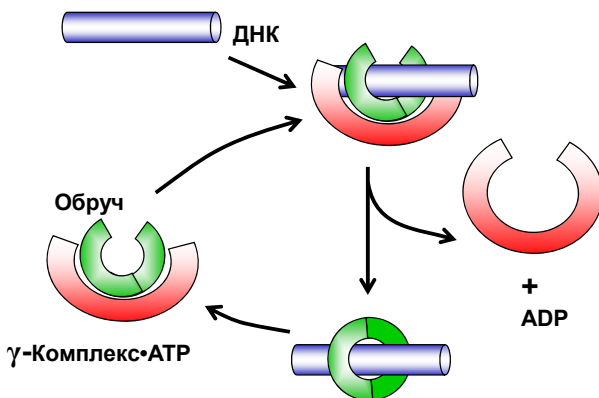


**Рис. 9.6.** Ковзний обруч (завантажений на ДНК бактеріальний  $\beta$ -білок) і комплекс обруча із ДНК-полімеразою III, яка складається із трьох субодиниць. Зображення є комбінацією структур 5FKV і 5M1S

Оскільки обруч є білковим кільцем, що оточує ДНК, то його завантаження на подвійну спіраль виконується АТР-залежним шляхом. Ця функція належить  $\gamma$ -комплексу, що складається із п'яти субодиниць і має U-подібну структуру (рис. 9.7).  $\gamma$ -Комплекс зв'язує АТР і змінює свою структуру у такий спосіб, що взаємодіє з обручем, порушуючи його цілісність (розмикає кільце). Далі  $\gamma$ -комплекс упізнає дуплекс, забезпечуючи посадку на нього відкритого кільця. Взаємодія з дуплексом індукує АТРазну активність  $\gamma$ -комплексу. У результаті гідролізу АТР  $\gamma$ -комплекс втрачає спорідненість до  $\beta$ -білка, що спричиняє замикання обруча. Далі  $\gamma$ -комплекс знову зв'язує АТР, проте обруч має вищу спорідненість до ДНК-полімерази, ніж до  $\gamma$ -комплексу, тому він залишається на ДНК і взаємодіє з полімеразою та забезпечує її процесивність.

Завантаження обруча відбувається на гібридну подвійну спіраль ДНК і РНК-праймера: один раз на лідируючий ланцюг при ініціації реплікації і кожного разу, коли на ланцюзі, що запізню-

ється, з'являється праймер, з якого починається черговий фрагмент Оказакі. Синтез праймера виконує спеціалізована ДНК-залежна РНК-полімераза (праймаза), зв'язана з реплікативною ДНК-геліказою, яка руйнує подвійну спіраль ДНК в основі реплікативної вилки. Комплекс гелікази та праймази – *праймосома* – є ще одним невід'ємним компонентом реплісоми. Через інші структурні модулі праймосома зв'язана із ДНК-полімеразами та працює у тісній координації з ними.



**Рис. 9.7.** Схема АТР-залежного завантаження ковзного обруча  $\gamma$ -комплексом

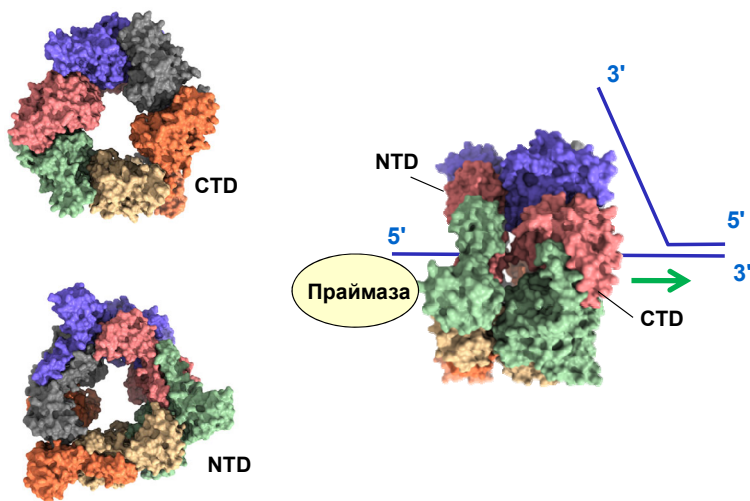
### **Праймосома: геліказа та праймаза**

Реплікативною геліказою бактерій є білок DnaB – кільцевий комплекс шести однакових субодиниць (рис. 9.8). У каналі всередині кільця розміщується один полінуклеотидний ланцюг, уздовж якого відбувається транслокація гексамеру в напрямку 5'-3' (тобто по ланцюгу, який запізнюється), що спричиняє руйнування подвійної спіралі й рух реплікативної вилки.

Кожна субодиниця гелікази складається із двох структурних доменів, з'єднаних доволі гнучким лінкером, – N- (NTD) і С-кінцевого (СТD). N- і С-кінцеві домени утворюють два кільця, накладені одне на одне (рис. 9.8). С-кінцеві домени мають сайти зв'язування і гідролізу АТР. Із шести таких сайтів два, розташовані у кільці навпроти один одного, у кожен момент перебувають в одному із трьох можливих станів: два порожні сайти;

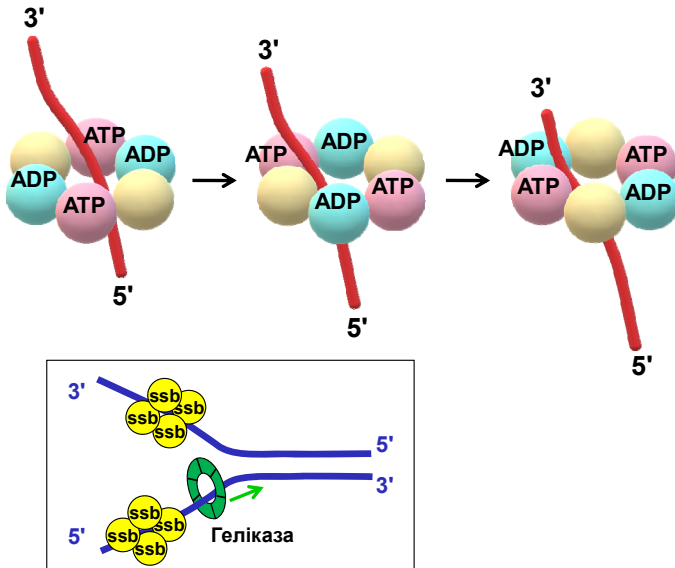
два сайти, де відбулось зв'язування АТФ; два сайти, де відбувся гідроліз АТФ (рис. 9.9). Зв'язування АТФ із порожнім сайтом викликає гідроліз АТФ у сусідньому сайті, а цей гідроліз – дисоціацію АДФ із наступного по кільцю сайта. Структурні перебудови субодиниць у відповідь на зв'язування та гідроліз АТФ і дисоціацію АДФ змінюють інтерфейс взаємодії з одноланцюговою ДНК, що зумовлює транслокацію гексамерного комплексу (рис. 9.9).

Енергія гідролізу АТФ використовується геліказою для здійснення енергетично не вигідного процесу розплітання дуплекса, тобто одноланцюгова ДНК, що виникає у реплікативній вилиці внаслідок активності гелікази, має бути тимчасово стабілізована в одноланцюговому стані. Цю функцію виконують білки SSB (Single Strand Binding), які мають високу спорідненість до одноланцюгової ДНК. Білки SSB зв'язуються з полінуклеотидним ланцюгом у вигляді гомотетрамеру, ефективно підтримуючи одноланцюговий стан ДНК позаду реплікативної вилки (рис. 9.9). SSB-білки взаємодіють також з іншими компонентами реплісоми, підсилюючи їхню активність, зокрема й активність ДНК-полімераза.



**Рис. 9.8.** Бактеріальна реплікативна геліказа DnaB у трьох проєкціях і схема її розташування відносно ланцюгів ДНК і праймази (2R6D).

Зелена стрілка – напрямок руху гелікази, NTD – N-кінцеві, CTD – C-кінцеві домени кожної із шести субодиниць



**Рис. 9.9.** Схема механізму транслокації гексамерної гелікази вздовж одноланцюгової ДНК.

На вставці: реплікативна вилка, що виникає як результат дії гелікази та підтримується білками SSB

Геліказне кільце, що сформовано N-кінцевими доменами субодиниць, має спорідненість до праймази – білка DnaG (рис. 9.8). Робота праймази стимулюється геліказою і SSB-білками. У деяких бактерій праймаза взаємодіє з геліказою достатньо міцно, у *E. coli* – тимчасово. Розпочинаючи від основи реплікативної вилки, праймаза, яка має низьку процесивність, синтезує *праймер* – фрагмент РНК завдовжки 9–14 нуклеотидів. Після цього ДНК-полімераза III, зв'язана з обручем, завантаженим на праймер, подовжуючи 3'-кінець праймера, починає синтезувати решту фрагмента Оказакі, а праймаза знову зв'язується із геліказою і згодом розпочинає синтез нового праймера.

На відміну від ДНК-полімерази, праймаза може з однаковим успіхом включати до праймера як дезоксирибо-, так і рибонуклеотиди. Праймер є фрагментом саме РНК просто тому, що внутрішньоклітинна концентрація рибонуклеозидтрифосфатів вища.

## **Загальна схема організації бактеріальної реплісоми**

Хоча мінімально в реплісомі мають бути дві ДНК-полімерази для синтезу двох ланцюгів ДНК, бактеріальна реплісома містить три молекули ферменту. Головну структурну функцію в організації реплісоми виконує  $\tau$ -білок, який взаємодіє з ДНК-полімеразою, геліказою, праймазою й утримує на собі  $\gamma$ -комплекс. Численні взаємодії  $\tau$ -білка забезпечують функціональну координацію між елементами реплісоми.

Три молекули  $\tau$ -білка взаємодіють між собою і кожна утримує кор-фермент ДНК-полімерази III, дві з яких, своєю чергою, зв'язані з обручем (рис. 9.10). Одна з полімераз здійснює синтез лідируючого ланцюга, інша – фрагмента Оказакі. Обидві полімерази орієнтовані в реплісомі однаково. Оскільки вони при цьому виконують синтез на антипаралельних ланцюгах, то той, що запізнюється, утворює реверсну петлю – конфігурація реплікативної вилки нагадує тромбон. Третя полімераза або є резервною (замінює ту, яка працює на ланцюзі, що запізнюється, після втрати її взаємодії з обручем), або завантажується на наступний праймер ще до закінчення синтезу фрагмента Оказакі – певний час на ланцюзі, що запізнюється, можуть працювати дві полімерази.

На рис. 9.10 зображено момент, коли праймаза почала синтез чергового праймера, а  $\gamma$ -комплекс готовий завантажити черговий обруч. Коли полімераза, яка працює на ланцюзі, що запізнюється, завершить синтез фрагмента Оказакі (зустріне попередній фрагмент), вона втратить зв'язок з обручем, але збереже взаємодію з  $\tau$ -білком. Втрата спорідненості полімерази до обруча визначається одноланцюговим розривом між фрагментами Оказакі. Новий обруч буде при цьому вже завантажений на черговий праймер, полімераза (та сама або резервна) зв'яжеться з ним і розпочне синтез нового фрагмента Оказакі. Обруч, який щойно втратив зв'язок із полімеразою, взаємодітиме з  $\gamma$ -комплексом (останній працює також як розвантажувач обруча, див. рис. 9.7), знімаючи його із ДНК і забезпечуючи згодом його посадку на праймер.

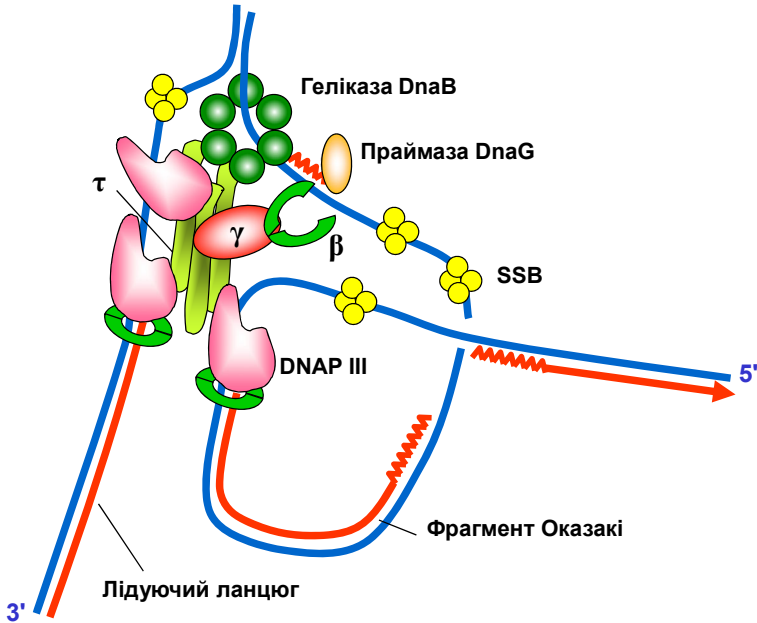
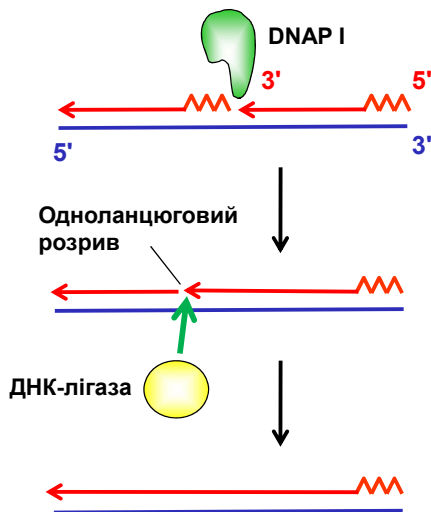


Рис. 9.10. Схема організації бактеріальної реплісоми

### **Завершення синтезу ланцюга, що запізнюється**

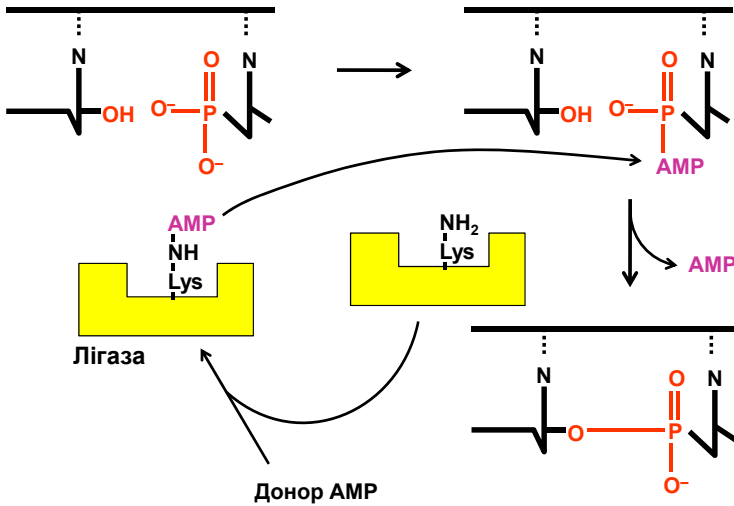
Ланцюг, що запізнюється, у процесі синтезу має вигляд фрагментів Оказакі завдовжки по 1–2 тис. нуклеотидів, кожен з яких розпочинається РНК-праймером. Зрозуміло, що праймер на початку кожного фрагмента Оказакі має бути замінений на відповідну послідовність ДНК. Цю роботу в бактеріальній системі реплікації виконує ДНК-полімераза I (рис. 9.11). За рахунок своєї 5'-екзонуклеазної активності полімераза видаляє праймер і одночасно, використовуючи 3'-кінець попереднього фрагмента Оказакі, подовжує його, заповнюючи прогалину.



**Рис. 9.11.** Робота ДНК-полімерази I і ДНК-лігази на ланцюзі, що запізнюється

Після виконання останньої операції між двома фрагментами Оказакі залишається одноланцюговий розрив (рис. 9.11). ДНК-полімераза не може його зашити, оскільки здатна тільки переносити нуклеотиди з NTP на 3'-кінцеву ОН-групу. Утворення фосфодієфірного зв'язку між сусідніми фрагментами Оказакі, яке потребує джерела енергії, здійснюється в реакції, яка каталізується ДНК-лігазою (DNA ligase).

Схему роботи лігази показано на рис. 9.12. До активного центра ферменту, ковалентно приєднуючись до аміногрупи Lys, переноситься аденозинмонофосфат АМР. Донором АМР стає макроергічна сполука: у бактерій цю функцію виконує нікотинамідаденіндинуклеотид  $NAD^+$ , еукаріотична лігаза використовує АТР. Далі лігаза переносить АМР до 5'-кінцевого фосфату в місці розриву (своєрідна активація, що нагадує активацію амінокислоти АРСазою, див. розділ 8). Видалення АМР супроводжується утворенням фосфодієфірного зв'язку.



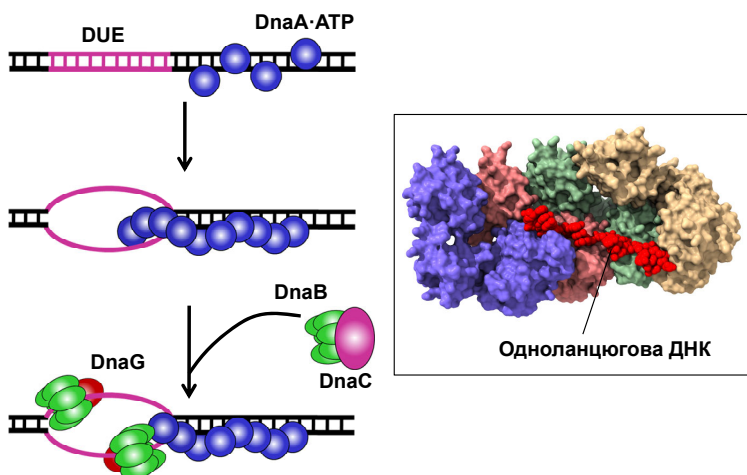
**Рис. 9.12.** Схема роботи ДНК-лігази.  
Донором AMP для бактеріальної лігази служить NAD<sup>+</sup>

## Ініціація і термінація реплікації у бактерій

Сайт ініціації реплікації *E. coli* – ориджин *oriC* – є ділянкою завдовжки 245 пар основ, яка містить чотири сайти по дев'ять пар основ кожний, що мають спорідненість до білка DnaA, а також кілька АТ-збагачених ділянок по 13 пар основ, які формують ділянку зі зниженою стабільністю подвійної спіралі DUE (DNA-unwinding element). Передумовою ініціації реплікації є створення за рахунок активності гірази певного рівня негативної надспіралізації в циркулярній хромосомі. По-перше, негативна надспіралізація сприяє локальному руйнуванню подвійної спіралі в межах ориджину. По-друге, оскільки введення надспіралізації можливе тільки за умови цілісності полінуклеотидних ланцюгів (хоча б один одноланцюговий розрив знімає будь-яку еластичну напругу), це є ефективним засобом перевірки готовності матриці (по всій довжині) до реплікації.

Ключову роль в ініціації реплікації відіграє білок DnaA, який містить кілька структурних доменів: один із них є АТФ-зв'язу-

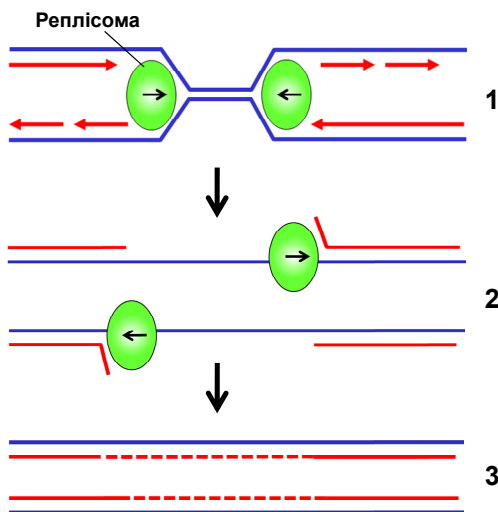
вальним (належить до родини AAA+ – ATPases associated with various cellular activities), інший – упізнає нонамерні сайти в ділянці ориднину. За відповідних сигналів регуляції клітинного циклу клітинна концентрація білка зростає і він зв'язує АТР – тільки у такій формі білок є активним. Зв'язування DnaA з ориджином спричиняє кооперативне заповнення доволі довгої ділянки білковим філаментом із 10–20 субодиниць білка (рис. 9.13). Утворення філаменту можливе тільки за наявності АТР у відповідному сайті білка. У складі філаменту DnaA стабілізує напружену конформацію подвійної спіралі, а експансія філаменту на ділянку DUE спричиняє плавлення подвійної спіралі. Плавлення є АТР-залежним тільки в тому сенсі, що саме АТР-зв'язана форма білка DnaA може його індукувати – від гідролізу АТР плавлення не залежить. Таким чином, руйнування подвійної спіралі ДНК при ініціації реплікації забезпечується певним рівнем негативної надспіралізації і взаємодіями з білком DnaA. Після закінчення процесу ініціації гідроліз АТР, який активується певними кофакторами, спричиняє переведення DnaA у неактивну форму і його дисоціацію.



**Рис. 9.13.** Ініціація реплікації у *E. coli* за участю білка DnaA. DUE – ділянка зі зниженою стабільністю (DNA-unwinding element), DnaB – геліказа, DnaC – завантажувач гелікази, DnaG – праймаза. На вставці: структура комплексу чотирьох субодиниць DnaA з одноланцюговою ДНК (3R8F)

Білок DnaA рекрутує в ділянку ориджину білок DnaC, який у АТР-залежний спосіб забезпечує завантаження у розплавлену зону гелікази DnaB – розкриває гексамерне кільце, порушуючи взаємодії між субодинамиціями. Геліказа, своєю чергою, зв'язує праймазу DnaG (рис. 9.13). Утворення двох праймосом ініціює збирання двох реплісом в основі двох реплікативних вилок, які починають рухатись у протилежних напрямках.

На діаметрально протилежному щодо *oriC* боці циркулярної бактеріальної хромосоми розташовано 6 ділянок по 22 пари основ – сайтів термінації реплікації (Ter A, B, C, D, E, F), які утворюють комплекс із білком Tus (Terminus utilization substance). На сайтах термінації реплікативні вилки зупиняються на певний час. Деталі складного процесу завершення реплікації ще недостатньо з'ясовані, але зрозуміло, що дві реплісоми (або просто дві гелікази) продовжують рух назустріч одна одній і заходять трохи далі у "не свою" зону (рис. 9.14), що викликає руйнування подвійної спіралі в ділянці термінації. Після цього гелікази дисоціюють, а прогалини, що залишилися на двох дочірніх молекулах, заповнюються ДНК-полімеразами.

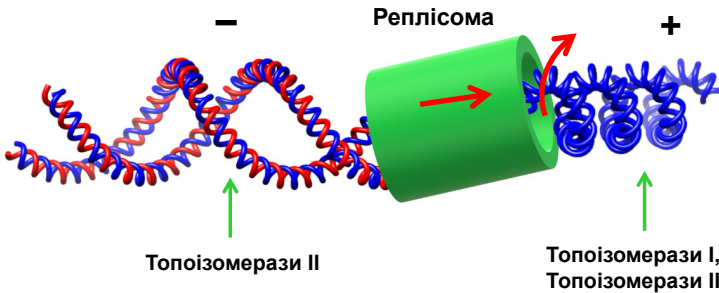


**Рис. 9.14.** Приблизна схема основних подій при термінації реплікації: зустріч двох реплікативних вилок (1); продовження руху геліказ, руйнування подвійної спіралі (2); заповнення прогалин на двох дочірніх молекулах ДНК (3)

Руйнування подвійної спіралі в зоні термінації не означає, що дві дочірні молекули при цьому розділюються – дві циркулярні дочірні хромосоми залишаються об'єднаними у катенан (див. рис. 3.31). Необхідність декатенації на завершальному етапі реплікації становить одну з топологічних проблем, що виникають під час цього процесу.

## Топологічні проблеми, пов'язані з реплікацією

Топологічні проблеми під час реплікації ДНК спільні для бактерій і еукаріотів. Рух реплікативної вилки з одночасним руйнуванням подвійної спіралі ДНК супроводжується прокручуванням дуплекса через реплісому, що створює позитивні надспіральні витки – еластичні напруження – попереду від вилки й негативні – позаду (див. рис. 3.25). У випадку реплікації позаду знаходяться дочірні ланцюги ДНК, що закручуються один навкруг одного (рис. 9.15).



**Рис. 9.15.** Рух реплісоми створює позитивну надспіралізацію попереду й негативну, яка спричиняє взаємне закручування дочірніх ланцюгів, – позаду

Позитивні напруження попереду двох вилок, що рухаються назустріч одна одній, накладаються. Це стосується як еукаріотичних систем у зоні між двома сусідніми репліконами (рис. 9.2, *a*), так і прокаріотичних циркулярних хромосом: зустрічний рух двох реплікативних вилок (рис. 9.2, *б*) створює позитивні на-

пруження на ділянці, яка ще не реплікована. Еластична напруга має час від часу зніматися, інакше згодом вона повністю зупинить рух реплісом. Так само перекручені ланцюги позаду від реплісоми треба час від часу розділяти. Відповідно процес реплікації обов'язково потребує допоміжної дії ДНК-топоізомераз (див. розділ 3), основна функція яких саме й полягає у релаксації надспіралізованої ДНК: штучне вимикання активності топоізомераз повністю блокує реплікацію.

В еукаріотичній клітині завдання релаксації позитивної надспіралізації попереду реплікативної вилки виконують як топоізомерази I, так і топоізомерази II. Прокаріотична топоізомераза I (топоізомераза Ia) не підходить для виконання цієї ролі, оскільки здатна знімати тільки негативну надспіралізацію. Релаксація позитивної надспіралізації при реплікації у бактерій забезпечується ДНК-гіразою – ферментом, що належить до класу топоізомераз II. Гіраза вносить негативну надспіралізацію у циркулярну ДНК, що й компенсує позитивні напруження.

Розділення двох ланцюгів позаду реплікативної вилки потребує створення тимчасового дволанцюгового розрізу в одному з них. Таку операцію здійснюють топоізомерази класу II – топоізомераза II еукаріотів і її аналог – топоізомераза IV – у бактерій.

Ще одна топологічна проблема, згадана вище, виникає при закінченні реплікації – як у циркулярному бактеріальному репліконі, так і при зустрічі реплікативних вилок двох сусідніх репліконів у еукаріотів. Рис. 9.16 ілюструє останні стадії термінації реплікації у бактерій. Як описано вище, у зоні термінації відбувається руйнування подвійної спіралі, яка ще не піддалась реплікації. Однак це неможливо зробити без відповідного перекручування двох дочірніх кільцевих молекул ДНК десь у іншому місці. Одноланцюгові прогалини заповнюються ДНК-полімеразами, після чого дві остаточно синтезовані циркулярні молекули залишаються топологічно зчепленими одна з одною – утворюється катенан (рис. 9.16). Таке саме перекручування двох дочірніх молекул ДНК реалізується при закінченні реплікації на двох сусідніх еукаріотичних репліконах.

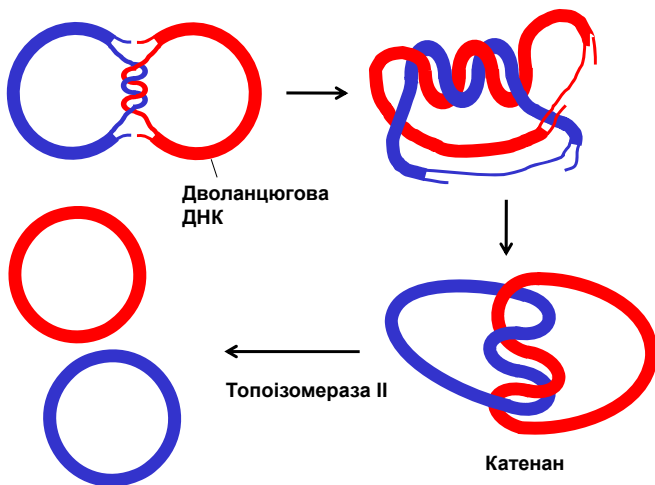


Рис. 9.16. Термінація реплікації у циркулярному репліконі

Зрозуміло, що декатенація – розділення двох молекул – потребує дволанцюгового розриву в одній із них. За цю операцію також відповідають топоізомерази класу II, без активності яких повне завершення реплікації неможливе.

## Особливості еукаріотичної системи реплікації

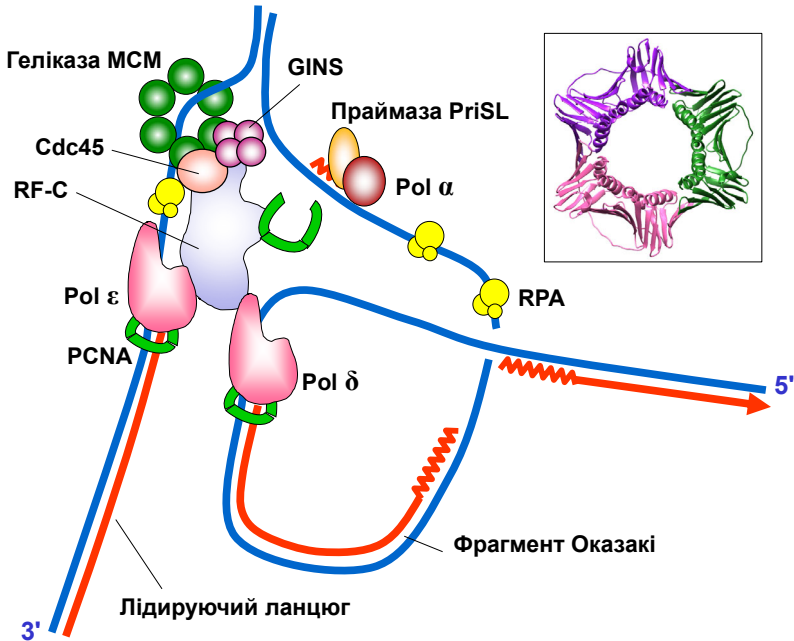
Загальні принципи й механізми реплікації є спільними для про- та еукаріотів, але розрізняються за деякими важливими аспектами.

Так само, як і у бактерій, у реплікативній вилці, одноланцюговий стан ДНК у якій підтримується білками **RPA** (Replication Protein A) – аналог бактеріальних білків SSB), працює мультибілковий комплекс – реплісома. Еукаріотична реплісома (рис. 9.17), на відміну від бактеріальної (порівн. рис. 9.10), містить не три, а дві трохи різні ДНК-полімерази. Як і у бактерій, процесивність цих полімераз забезпечується ковзним циркулярним обручем. Різниця лише в тому, що обруч (**PCNA** – Proliferating Cell Nuclear Antigene) складається із трьох субоди-

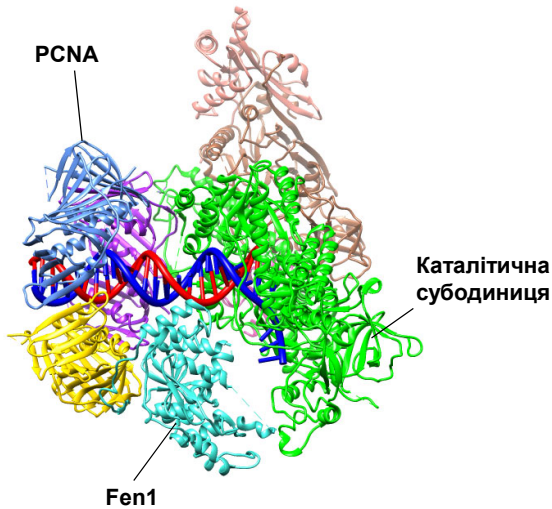
ниць (рис. 9.17). Крім забезпечення процесивності ДНК-полімерази, PCNA взаємодіє ще з багатьма білками, залученими до виконання різноманітних операцій під час реплікації.

Основним структурним модулем еукаріотичної реплісоми є мультибілковий комплекс **RF-C** (**R**eplіcаtіo**n F**аcтo**r C**), що виконує функції, аналогічні таким бактеріальному  $\tau$ -білку і  $\gamma$ -комплексу (рис. 9.17), тобто забезпечує цілісність реплісоми та координацію роботи її компонентів і завантажує обруч PCNA на праймер.

Як і у бактерій, реплікативні ДНК-полімерази еукаріотів мають 3'-екзонуклеазну активність, що використовується для редагування помилок. Дві реплікативні ДНК-полімерази еукаріотів спеціалізовані щодо двох ланцюгів: **ДНК-полімераза  $\epsilon$**  (Pol  $\epsilon$ , чотири субодиниці) веде синтез лідируючого ланцюга, **ДНК-полімераза  $\delta$**  (Pol  $\delta$ , 3–5 субодиниць у різних видів) – фрагментів Оказакі в ланцюзі, що запізнюється (рис. 9.18, порівн. з рис. 9.6).



**Рис. 9.17.** Схема організації еукаріотичної реплісоми. На вставці: структура PCNA – ковзного обруча (5МММ)



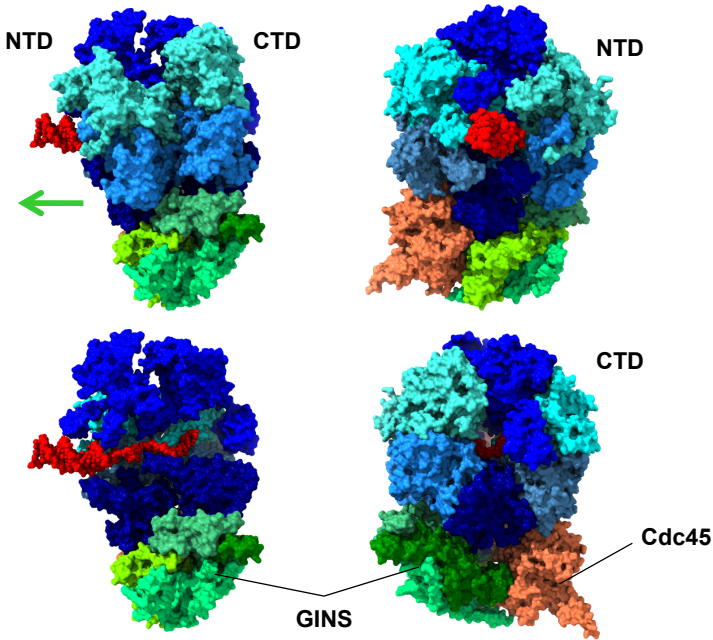
**Рис. 9.18.** ДНК-полімераза  $\delta$  людини (4 субодинаці) у комплексі з ДНК, PCNA-обручем і ендонуклеазою Fen1, що залучена до видалення праймерів ( $\delta$ TNZ). 3'-Кінець зростаючого ланцюга (червоний) знаходиться у полімеразному активному центрі, інший домен каталітичної субодинаці відповідає за 3'-екзонуклеазну активність

Довжина еукаріотичних фрагментів Оказаки становить 150–200 нуклеотидів. Розпочинаються вони, як і у бактерій, із РНК-праймера завдовжки  $\sim 12$  нуклеотидів, який створює еукаріотична *праймаза PriSL* (гомологічна праймаза наявна також у археїв). В еукаріотів ця праймаза зв'язана в єдиному комплексі з низькопроцесивною *ДНК-полімеразою  $\alpha$*  – остання підхоплює 3'-кінець праймера і подовжує його у вигляді ДНК на  $\sim 25$  нуклеотидів, після чого замінюється на полімеразу  $\delta$ . На відміну від бактерій, видалення РНК-праймера при завершенні синтезу ланцюга, що запізнюється, здійснюється завдяки активності спеціалізованої *ендонуклеази Fen1*, зв'язаної із ДНК-полімеразою  $\delta$  та її PCNA-обручем (рис. 9.18): ДНК-полімераза, синтезуючи фрагмент Оказаки, витісняє 5'-кінцеву ділянку із праймером попереднього фрагмента Оказаки, цей од-

ноланцюговий виріст відщеплюється нуклеазою, а одноланцюговий розрив, що залишився, зашивається АТР-залежною ДНК-лігазою.

Крім згаданих ДНК-полімераз  $\alpha$ ,  $\delta$ , і  $\epsilon$ , залучених до синтезу ДНК при реплікації у клітинному ядрі, в еукаріотичних клітинах наявні інші ДНК-полімерази: полімераза  $\beta$  використовується при екзизійній репарації основ (див. нижче); полімераза  $\gamma$  – реплікативна ДНК-полімераза мітохондрій (дві субодиниці, крім ДНК-полімеразної, має обидві, 3'- та 5'-екзонуклеазні, активності); полімерази  $\lambda$  і  $\mu$  залучені до репарації дволанцюгових розривів. В еукаріотичних клітинах наявні також ДНК-полімерази низької точності синтезу ( $\zeta$ ,  $\eta$ ,  $\iota$ ,  $\kappa$ ). Їхня основна функція полягає в тому, щоб забезпечити синтез ДНК у разі пошкодження матриці: зустрічаючи пошкодження, реплікативна машинерія високої точності (яка включає полімерази  $\delta$  чи  $\epsilon$ ) зупиняється. Полімерази низької точності дозволяють долати пошкодження, які пізніше можуть бути виправлені шляхом репарації, і знов замінюються на полімерази  $\delta/\epsilon$ , що продовжують високоточний синтез ДНК.

Повертаючись до організації еукаріотичної реплісоми, слід згадати ще один її невід'ємний модуль – реплікативну геліказу, що забезпечує руйнування подвійної спіралі й рух реплікативної вилки. Еукаріотична реплікативна **геліказа МСМ** (MiniChromosome Maintenance) є гетерогексамером і складається із шести різних субодиниць – білків МСМ2-7 (рис. 9.19). Ці субодиниці не мають нічого спільного з бактеріальним білком DnaB, але, подібно до нього, містять N- (NTD) і C-кінцевий (CTD) структурні домени (останній має АТРазну активність). Подібність загальної організації бактеріальної та еукаріотичної геліказ передбачає також деяку подібність механізмів їхньої дії.



**Рис. 9.19.** Активна еукаріотична реплікативна геліказа – комплекс CMG шести субодиниць MCM із білком Cdc45, тетрамером GINS і ДНК – у трьох проєкціях (5U8S). NTD – N-кінцеві, CTD – С-кінцеві домени білків MCM. Стрілка позначає напрямок руху гелікази. На одній із проєкцій видалено дві субодиниці MCM для візуалізації одноланцюгової ДНК всередині гелікази

Дійсно, руйнування подвійної спіралі відбувається завдяки транслокації гелікази вздовж одноланцюгової ДНК. Проте, на відміну від бактеріальної гелікази, по-перше, еукаріотична геліказа рухається по лідируючому ланцюгу, відповідно в напрямку 3'–5' (рис. 9.17). По-друге, еукаріотична геліказа рухається в бік N-кінцевих доменів своїх субодиниць (рис. 9.19, порівн. з рис. 9.8). По-третє, гексамер MCM2-7 сам по собі не здатен руйнувати подвійну спіраль – геліказа активується завдяки взаємодії з білком Cdc45 (*c*dc – *c*ell *d*ivision *c*ontrol) і тетрамерним комплексом GINS (від японських назв цифр 5-1-2-3 – *g*o-*i*chi-*n*i-*s*an,

якими позначаються субодиниці комплексу). Активна форма гелікази, яка міститься у складі реплісоми (рис. 9.17, 9.19), позначається як **комплекс CMG** (Cdc45–MCM2–7–GINS). Завантаження цього комплексу на еукаріотичний ориджин є ключовою подією ініціації реплікації.

## Ініціація реплікації в еукаріотів

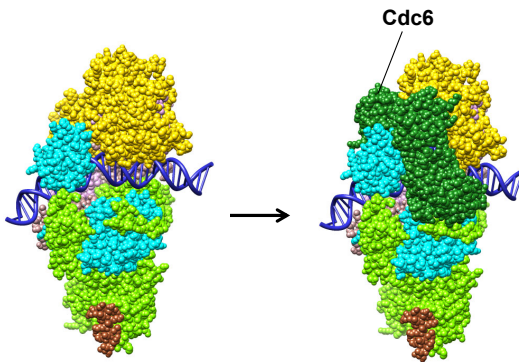
Обставиною, яка радикально відрізняє еукаріотичну систему реплікації від бактеріальної, є те, що кожна хромосома є полірепліконом: загалом геном, наприклад ссавців, містить близько 40 тис. точок ініціації – ориджинів (розмір еукаріотичного реплікона варіює від 50 до 200 тис. пар основ). У дріжджів ориджини є АТ-збагаченими консенсусними елементами нуклеотидної послідовності, що упізнаються факторами ініціації реплікації. У *S. cerevisiae* із біля 12 тис. таких елементів використовуються як ориджини ~400. Для ефективного впізнання ориджинів мають значення також особливості структури хроматину. У багатоклітинних будь-яка подібність нуклеотидних послідовностей ориджинів відсутня. Спільною рисою активних ориджинів є те, що вони зазвичай розташовані в зонах, вільних від нуклеосом, подібних до таких саме зон у стартових точках транскрипції (розділ 6); іноді ці ділянки можуть збігатися. Детальні механізми впізнання ориджинів залишаються недостатньо з'ясованими.

Еукаріотичний ориджин маркується ключовим фактором ініціації реплікації – комплексом із шести білків, який позначається як **ORC** (OriRigin Recognition Complex). Структуру зв'язаного із ДНК ORC зображено на рис. 9.20. У який момент клітинного циклу і як цей комплекс завантажується на ориджин, залишається невідомим. Взаємодія ORC із ДНК не дуже міцна, але наприкінці фази G1 клітинного циклу до ORC приєднується білок Cdc6, який жорстко фіксує ДНК у сайті зв'язування (рис. 9.20). Cdc6, внутрішньоклітинна концентрація якого відносно невисока, зв'язується тільки із частиною ORC. Ориджини, на яких це відбулось, потенційно можуть бути залучені до наступних етапів ініціації. Білок Cdc6 і 5 із 6 субодиниць ORC містять

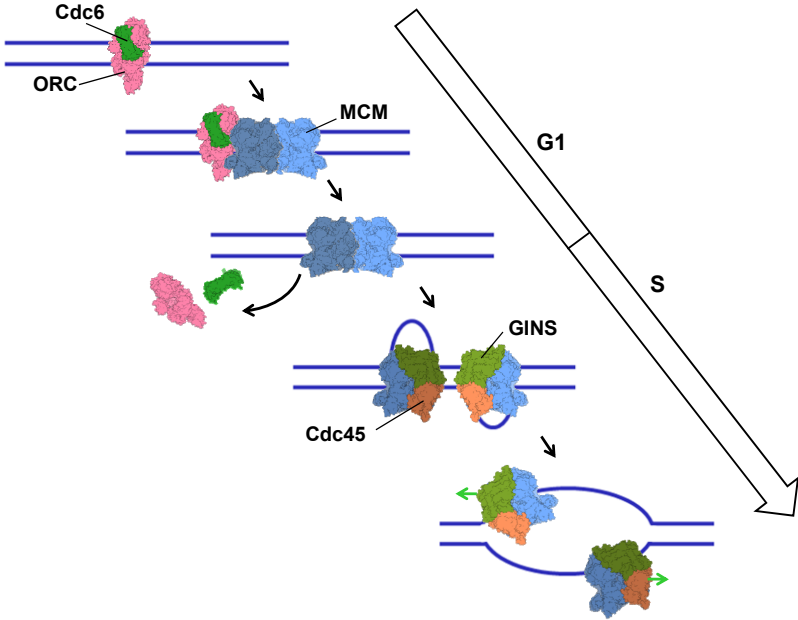
АТР-зв'язувальні структурні домени, що належать до родини AAA+ – тієї самої, якій належить бактеріальний фактор ініціації DnaA. Однак ORC, на відміну від DnaA, не здатен індукувати плавлення подвійної спіралі.

На наступному етапі (рис. 9.21), після відповідних сигналів системи регуляції клітинного циклу у вигляді фосфорилювання певних білків циклінзалежними кінзазами, комплекс ORC-Cdc6 рекрутує в ориджин два гексамери MCM разом із певними кофакторами, не показаними на рис. 9.21. Ті ориджини, де відбулось зв'язування Cdc6 і MCM, стають *лицьзованими* – готовими розпочати процес ініціації. Значна частина таких ориджинів залишається неактивною, на інших відбувається активація геліказ, і тільки такі ориджини стають активними.

На відміну від бактеріальної системи, де завантаження геліказ на одноланцюгову ДНК здійснюється завдяки АТР-залежному відкриванню гексамерних кілець фактором DnaC, завантаження гексамерів MCM не потребує особливих зусиль: у вільному вигляді гексамерне кільце має щілину між двома субодинами, через яку дволанцюгова ДНК потрапляє у внутрішній канал кільця. Два гексамери MCM зв'язуються послідовно один за одним і взаємодіють між собою N-кінцевими доменами своїх субодинаць. У процесі зв'язування після гідролізу АТР відбувається дисоціація Cdc6 і компонентів ORC, а також замикання кілець MCM.



**Рис. 9.20.** Шість субодинаць ORC у комплексі з ориджином і білком Cdc6 (7JGR)



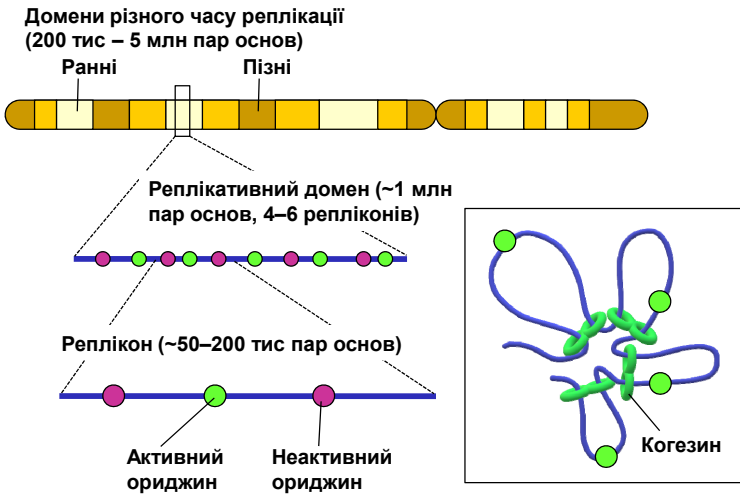
**Рис. 9.21.** Основні події на початковому етапі ініціації реплікації.  
G1, S – фази клітинного циклу

На початку S-фази клітинного циклу відбувається активація геліказ (рис. 9.21): фосфорилювання MCM, руйнування зв'язку між двома гексамерами за допомогою певних кофакторів, розходження гексамерів на деяку відстань у протилежні боки та зв'язування з обома гексамерними кільцями білків Cdc45 і тетрамерів GINS (ці фактори ініціації реплікації залишаються потім у реплісомі, рис. 9.17) – утворюються активні гексамерні комплекси CMG. У складі CMG відбувається виштовхування одного з ланцюгів ДНК назовні, після чого дві гелікази рухаються назустріч одна одній (у бік своїх N-кінцевих доменів) по різних ланцюгах, що зумовлює утворення розплавленої ділянки (рис. 9.21). Цей напрямок руху потім зберігається для двох реплікативних вилок.

На завершальному етапі здійснюється збирання двох реплісом: до основ двох щойно утворених вилок, одноланцюговий стан яких підтримується білками RPA, рекрутуються комплекси праймаза-ДНК-полімераза  $\alpha$ , що починають синтез праймерів лі-

дируючих ланцюгів; з основою вилки зв'язується RF-C, який завантажує обруч PCNA, полімераза  $\alpha$  замінюється на полімеразу  $\epsilon$  (перемикання полімераз). Надалі MCM продовжує розділення ланцюгів ДНК, а полімераза  $\alpha$  здійснює синтез праймерів на ланцюзі, що запізнюється, і кожного разу замінюється на полімеразу  $\delta$ .

Загальна система регуляції реплікації в еукаріотів залишається далекою від остаточного розуміння – тривалий (близько 8 год) процес подвоєння величезної кількості генетичного матеріалу надзвичайно складний. Увесь геном можна розділити на зони завдовжки до 5 млн пар основ, які реплікуються на початку, у середині й наприкінці S-фази клітинного циклу – на *ранніх*, *середніх* і *пізніх* стадіях (рис. 9.22). Ці зони, своєю чергою, складаються із реплікативних доменів, у межах яких об'єднуються кілька сусідніх репліконів. Реплікація запускається синхронно на репліконах одного домену. Імовірно, основою формування реплікативного домену є об'єднання кількох хроматинових петель (див. розділ 4) у єдину хроматинову структуру високого порядку (рис. 9.22).



**Рис. 9.22.** Схема розподілу ділянок, що реплікуються, в еукаріотичному геномі: кілька репліконів об'єднуються у реплікативний домен, сусідні домени формують зону, що реплікується на ранніх, середніх або пізніх стадіях S-фази. На вставці: гіпотетична схема організації реплікативного домену – кожна петля є репліконом

Визначальним фактором розподілу різних активних ориджинів по часових проміжках S-фази є їхня доступність до машинерії ініціації реплікації. Ця доступність залежить від структури хроматину в тій чи іншій ділянці, а структура тісно пов'язана з функціональними характеристиками ділянки: на ранніх стадіях здійснюється реплікація еухроматинових зон, які активно транскрибуються, на пізніх – транскрипційно мовчазного гетерохроматину. При цьому ориджини однієї категорії (ранні, середні й пізні) вмикаються випадково – так, що часовий розподіл ініціації реплікації на різних ориджинах одного типу варіює від одного клітинного циклу до іншого.

Отже, ориджини ранньої S-фази розташовані в зонах, що активно транскрибуються. З іншого боку, ранні ориджини характеризуються порівняно низькою швидкістю ініціації реплікації. Справа в тім, що деякі фактори ініціації (Cdc45 і деякі інші) наявні на початку S-фази в доволі низьких концентраціях, тобто лімітують процес. Ключовим фактором визначення ранніх ориджинів є відкрита структура хроматину у транскрипційно активних ділянках: це сприяє ефективному зв'язуванню в таких ділянках і факторів транскрипції (розділ 6), і факторів ініціації реплікації. Після закінчення реплікації на ранніх репліконах лімітувальні фактори звільняються і, разом із новими лімітувальними факторами, концентрація яких зростає протягом S-фази, мають змогу шукати ориджини наступної часової категорії (рис. 9.23). При цьому ймовірність ініціації підвищена на ориджинах, сусідніх із репліконом, де вже відбувається синтез ДНК.

Отже, реплікація відбувається за принципом доміно. Реплікація, що здійснюється в кількох ранніх репліконах, індукує ініціацію реплікації на сусідніх ориджинах. Принаймні один із факторів такого впливу може полягати в підвищенні структурної динаміки хроматину на ділянках, що фланкують активний реплікон. У результаті ориджини стають більш доступними, там також розпочинається реплікація, що, своєю чергою, індукує ініціацію на сусідніх ділянках. Процес продовжується лавиноподібно (із підвищенням ефективності ініціації) і виключає наявність прогалін, які б залишились нереплікованими.

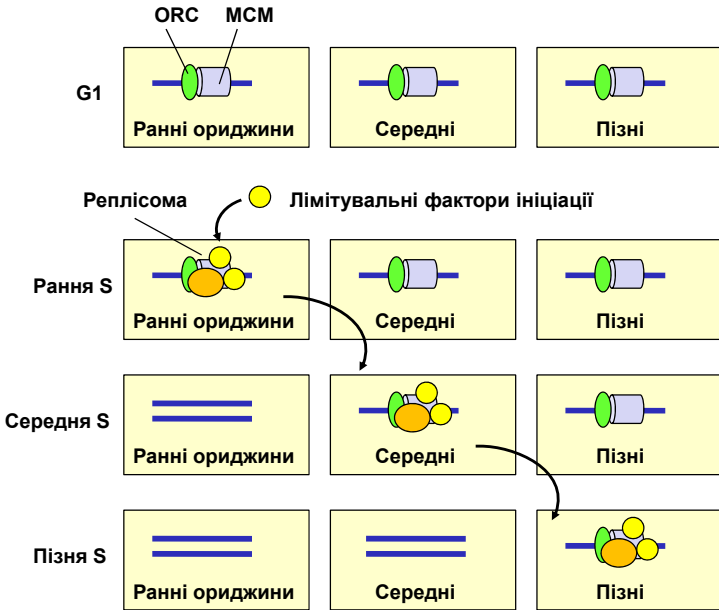


Рис. 9.23. Запуск реплікації ориджинів різних часових категорій

## Подовження кінців еукаріотичної хромосоми

Ще одна принципова відмінність еукаріотичної системи реплікації від бактеріальної полягає в тому, що еукаріотична хромосома, на відміну від циркулярної бактеріальної, є лінійною, тобто має два кінці. Унаслідок цієї простої обставини на 3'-кінцях матричних ланцюгів ДНК залишаються одноланцюгові хвости (рис. 9.24): два РНК-праймери на 5'-кінцях синтезованих ланцюгів видаляються, а прогалина не може бути заповненою, оскільки немає 3'-кінця, який міг би бути використаний як праймер. Одноланцюгові хвости піддаються швидкій нуклеазній деградації, і після кожної реплікації хромосома вкорочується. Це вкорочення не є фатальним, оскільки кінцеві ділянки ДНК еукаріотичної хромосоми – теломери – складаються з невеликих беззмистовних елементів послідовності, що тандемно повторю-

ються, – теломерних повторів. Однак певне критичне скорочення теломер, яке відбувається в соматичних диференційованих клітинах після кількох десятків клітинних поділів, є одним із механізмів активації програми їхньої загибелі.

Проте у стовбурових клітинах (а також зляккісно трансформованих) є активний спеціальний фермент – *теломераза*, – за допомогою якого здійснюється подовження теломер. Теломераза є РНК-залежною ДНК-полімеразою – РНК-матриця входить до складу самого ферменту й містить ділянку, комплементарну теломерному повтору (рис. 9.25). Використовуючи цю ділянку як матрицю і 3'-кінець як праймер, теломераза покроково добуває до 3'-кінця кілька копій теломеразного повтору. Далі подовжений одноланцюговий хвіст використовується як матриця для синтезу іншого ланцюга за звичайним реплікативним механізмом. Видалення РНК-праймера після цього не є проблемою, оскільки хромосома вже подовжена. 3'-Кінець, що виступає, при цьому залишається – він "запаковується" у подвійну спіраль у ділянці теломери за допомогою низки спеціальних білків. Така упаковка необхідна для того, щоб цей кінець не сприймався системами репарації ДНК, інакше можливим стало б зшивання кінців різних хромосом (див. далі обговорення систем репарації).

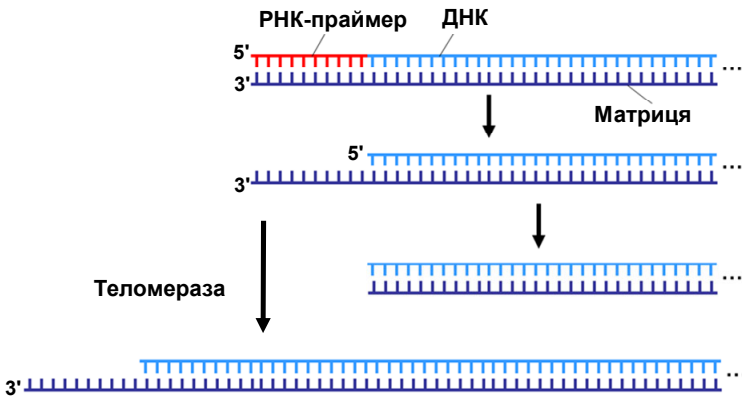
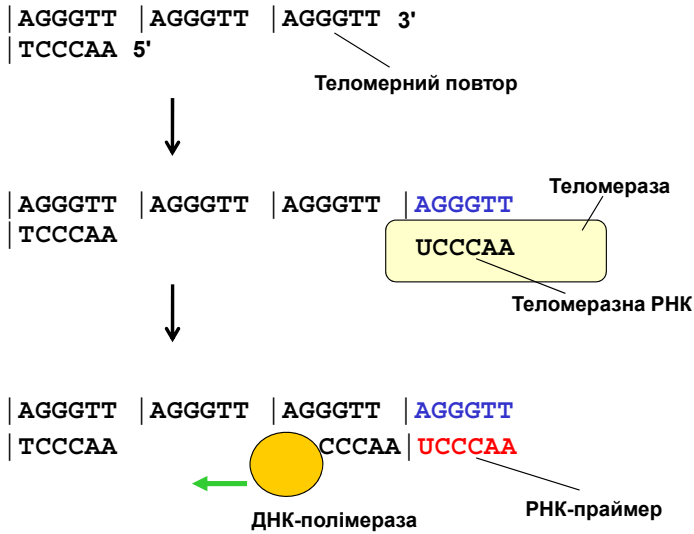


Рис. 9.24. Кінці дочірніх хромосом після реплікації



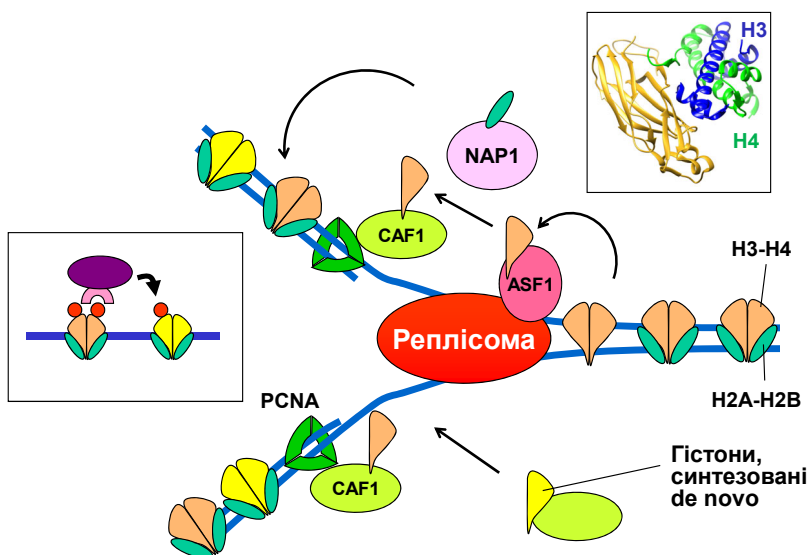
**Рис. 9.25.** Подовження кінців хромосоми за допомогою теломерази

## Проходження реплісоми через хроматин

Ще однією принциповою особливістю еукаріотичної системи реплікації є те, що процес відбувається у нуклеосомному середовищі. Зрозуміло, що нуклеосоми є суттєвим бар'єром на шляху реплісоми (так само як і для РНК-полімерази, див. розділ 4).

Попереду реплікативної вилки відбувається тимчасове руйнування нуклеосом у два етапи (рис. 9.26): спочатку видаляються димери гістонів H2A-H2B, потім найміцніше зв'язані із ДНК димери H3-H4 (ця пара гістонів знімається із ДНК саме у вигляді димерів). Процес видалення гістонових комплексів забезпечується активністю комплексів ремоделювання хроматину (розділ 6) і наявністю проміжних акцепторів гістонів – гістонових шаперонів (розділ 4). Такі шаперони зв'язують гістонові комплекси, що витісняються реплісомою, а потім виконують роль факторів збирання нуклеосом позаду реплікативної вилки. Необхідність таких факторів визначається надзвичайно високою спорідненіс-

тю гістонів до ДНК (див. рис. 4.10). Основну роль під час реплікації виконують проміжні переносники NAP1 (Nucleosome Assembly Protein), що мають підвищену спорідненість до димерів H2A-H2B, і ASF1 (Anti-Silencing Function), споріднений до гістонів H3-H4 та інкорпорований у реплісому. З ASF1 димери H3-H4 передаються на CAF1 (Chromatin Assembly Factor) – модуль, який має спорідненість до PCNA і забезпечує завантаження двох димерів у вигляді тетрамеру (H3-H4)<sub>2</sub> на дочірні ланцюги ДНК позаду реплікативної вилки. Отже, відновлення нуклеосом також здійснюється у дві стадії: після завантаження тетрамеру (H3-H4)<sub>2</sub> до нього приєднуються два димери H2A-H2B, що переносяться шапероном NAP1.



**Рис. 9.26.** Руйнування та відновлення нуклеосом у реплікативній вилці. ASF1, NAP1 – гістонові шаперони, CAF1 – фактор збирання тетрамерів (H3-H4)<sub>2</sub>. На вставках: структура комплексу димеру H3-H4 з ASF1 (2HUE) і механізм відновлення на нових гістонах хімічних модифікацій, які несуть батьківські гістони

По двох дочірніх ланцюгах ДНК гістонові комплекси розподіляються випадково, але при появі двох ланцюгів ДНК замість одного потрібно вдвічі більше гістонів: під час реплікації у цитоплазмі відбувається інтенсивний синтез гістонів *de novo*, і ці нові гістони, що переносяться у ядро гістоновими шаперонами у вигляді димерів, додаються до "старих" на дочірніх ланцюгах.

Отже, "старі" гістони, що містять певні характерні модифікації (див. розділи 4 і 6), повертаються на ту саму ділянку ДНК обох ланцюгів, на якій вони були на материнській молекулі. До ділянок хроматину позаду від реплікативної вилки рекрутуються відповідні ферменти, які здійснюють аналогічні модифікації щойно синтезованих гістонів, використовуючи "зчитувальні головки" (бромодомени, хромодомени тощо) – патерн модифікацій відновлюється, що сприяє збереженню певного функціонального стану ділянки хроматину в дочірніх клітинах. Принцип відновлення модифікацій на нових гістонах із використанням батьківських як матриць зображено на вставці рис. 9.26. Відновлення патерну модифікацій гістонів стосується передусім гістонів H3 та H4 (доволі інтенсивний обмін димерами H2A-H2B між різними ділянками хроматину робить ці гістони менш важливими для епігенетичної спадковості).

Важливу роль в інтеграції процесів відновлення нуклеосом та епігенетичних маркерів під час реплікації відіграє PCNA: крім гістонових шаперонів, він має спорідненість і до багатьох ферментів, що здійснюють модифікації гістонів. Крім того, саме PCNA рекрутує до реплікативної вилки підтримувальну ДНК-метилтрансферазу, яка відновлює патерн метилювання динуклеотидів CpG у ДНК (див. рис. 6.32).

## Репарація ДНК

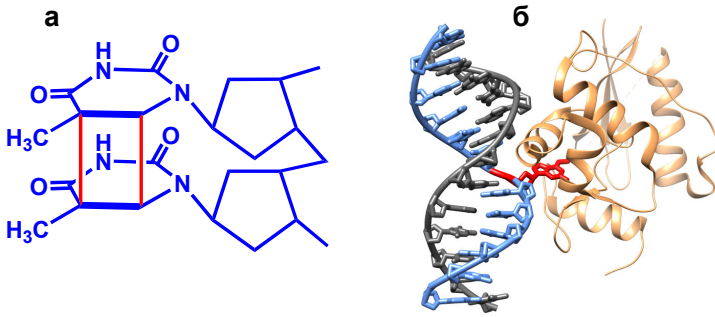
Репарація (repair) ДНК – один із загальних біологічних процесів, спрямований на виправлення помилок синтезу ДНК при реплікації, а також численних пошкоджень, які виникають у ДНК унаслідок дії хімічних агентів і випромінювання різних типів. До таких пошкоджень відносять різноманітні хімічні мо-

дифікації азотистих основ, ковалентні зшивки сусідніх піримідинів (утворення піримідинових, найчастіше тимінових, димерів) під дією ультрафіолетового випромінювання, одно- і дволанцюгові розриви під дією іонізуючої радіації та вільних радикалів тощо. Часто системи репарації працюють під час або одразу після реплікації. Більшість репараційних процесів передбачають видалення пошкодженої одноланцюгової ділянки з подальшим синтезом ДНК за допомогою ДНК-полімераз. Однак існують процеси, пов'язані з безпосереднім виправленням пошкодженого елемента за рахунок прямої дії певних ферментів.

### **Пряма репарація**

Найочевиднішим випадком прямої репарації є зшивання одноланцюгового розриву ДНК лігазою (див. рис. 9.12). Іншим спільним для всіх живих організмів (за винятком ссавців) шляхом прямої репарації є *фотореактивація* – руйнування піримідинових димерів (рис. 9.27, *а*), які були індуковані ультрафіолетовим світлом, ферментом *фотоліазою*. Фотоліаза (або її власні амінокислотні залишки, або зв'язані з білком простетичні групи) здатна поглинати світло, що викликає активацію ферменту, тобто світло, яке викликає утворення піримідинових димерів, одночасно активує фотоліазу, що каталізує розрив ковалентних зв'язків між сусідніми піримідинами, а отже, відновлення структури ДНК.

Одним із пошкоджувальних впливів на ДНК є *алкілювання* азотистих основ – ковалентне приєднання метильних чи етильних груп до атомів O або N. Пряма репарація таких пошкоджень можлива за рахунок активності специфічних метилтрансфераз, що відщеплюють метильні групи (рис. 9.27, *б*). Таким шляхом репаруються O<sup>6</sup>-метилгуанін та O<sup>4</sup>-метилтимін. Ці метилтрансферази не є справжніми ферментами: вони відщеплюють метильну групу й необернено ковалентно приєднують її до залишку Cys у своєму активному центрі – для нового акту деметилування необхідна нова молекула білка.

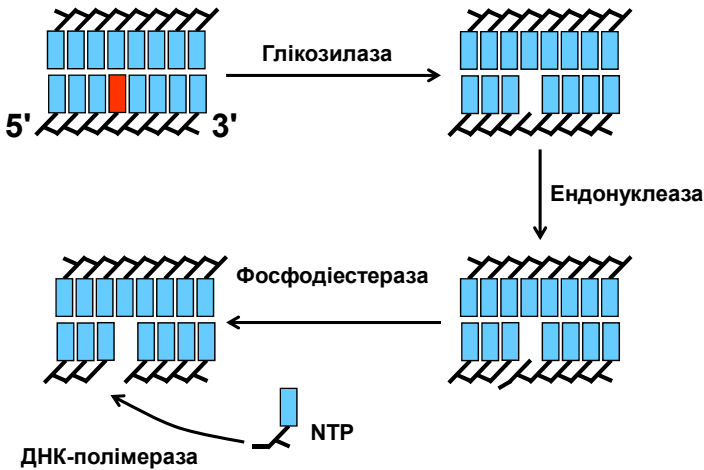


**Рис. 9.27.** Тиміновий димер (а) і структура комплексу ДНК із ферментом репарації *O6*-метилгуанін-ДНК-метилтрансферазою (б, 1Т38) – *O6*-метилгуанін (червоний) міститься в активному центрі ферменту

### Ексцизійна репарація

Більш радикальним і ефективним шляхом виправлення пошкоджень нуклеотидів є *ексцизійна репарація* (excision repair), коли пошкоджена одноланцюгова ділянка вирізається із ДНК, а інший ланцюг використовується потім як матриця для нового синтезу. Існують два варіанти такої репарації.

При *ексцизійній репарації азотистих основ* (Base Excision Repair – BER), що відбувається в усіх організмів, модифікована азотиста основа видаляється ферментом *глікозилазою*. Існує певна кількість специфічних глікозилаз, що розпізнають різноманітні модифіковані основи. Глікозилаза руйнує глікозидний зв'язок між основою і С1'-атомом дезоксирибози (рис. 9.28). У ДНК залишається АП-сайт (апуриновий / апіримідиновий), який упізнається ендонуклеазою, що гідролізує фосфодієфірний зв'язок між 5'-фосфатом звільненої від основи дезоксирибози й попереднім нуклеотидом. Нарешті фосфодіестераза відщеплює цю фосфодезоксирибозу, і в ДНК залишається прогалина завдовжки в один нуклеотид.



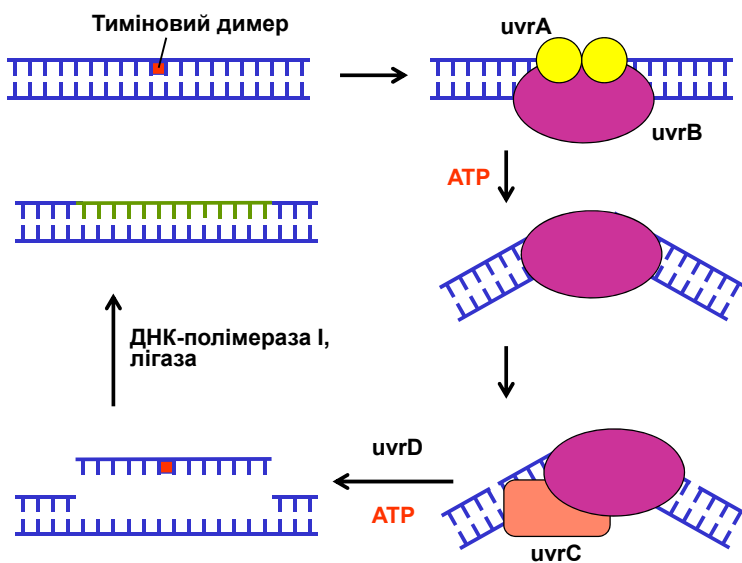
**Рис. 9.28.** Екцизійна репарація азотистих основ.  
Пошкоджена основа забарвлена червоним

Ця прогалина заповнюється ДНК-полімеразою  $\beta$  (в еукаріотів), яка приєднує нуклеотид до 3'-ОН-групи попереднього нуклеотиду ланцюга. Фосфодієфірний зв'язок приєданого нуклеотиду з наступним нуклеотидом ланцюга відновлюється лігазою. У бактерій заповнення прогалини здійснюється ДНК-полімеразою I. При цьому, за рахунок своєї 5'-екзонуклеазної активності, полімераза може руйнувати певну ділянку з 5'-кінця прогалини, одночасно продовжуючи 3'-кінець.

**Екцизійна репарація нуклеотидів** (Nucleotide Excision Repair – NER) – процес, пов'язаний із вирізанням ділянки ДНК, яка містить пошкодження (модифіковану основу, тиміновий димер тощо).

У клітинах *E. coli* за цей шлях відповідає система *uvrABC* (*uvr* – *ultra violet repair*). Комплекс білка *uvrB* і двох білків *uvrA* впізнає пошкодження і зв'язується із ДНК у цьому місці (рис. 9.29). На наступному кроці відбувається АТФ-залежна зміна конформації *uvrB*, вигинання ДНК і дисоціація *uvrA*. До комплексу рекрутується білок *uvrC*. Обидва білки у складі комплексу набувають ендонуклеазної активності: *uvrC* робить однолан-

цюговий розріз у пошкодженому ланцюзі за кілька нуклеотидів у напрямку до 5'-кінця від пошкодження (ліворуч на рис. 9.29), *uvrV* – розріз з іншого боку від пошкодження. Довжина ділянки між розрізами дорівнює 12 (або 13 у випадку для тимінового димеру) нуклеотидам. Далі геліказа *uvrD* руйнує подвійну спіраль між двома розрізами, тобто видаляє пошкоджену ділянку. Прогалина, що залишилася, заповнюється ДНК-полімеразою I, лігаза остаточно відновлює цілісність ланцюга.



**Рис. 9.29.** Система ексцизійної репарації нуклеотидів *uvrABC* у *E. coli*

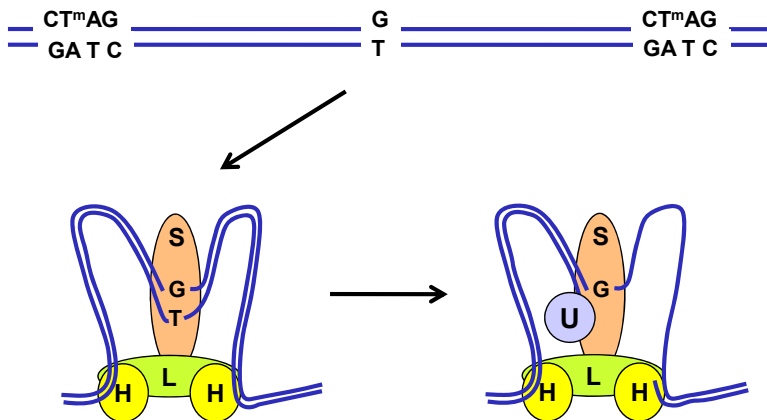
Аналогічна система ексцизійної репарації працює в еукаріотичних клітинах. До неї залучено близько 17 білків, причому за руйнування подвійної спіралі відповідає геліказна частина загального фактора транскрипції TFIIH (див. розділ 5). Це забезпечує тісну координацію репарації із транскрипцією: і підвищена ймовірність пошкоджень, і першочергова необхідність виправляти їх виникає насамперед у транскрипційно активних ділян-

ках. Пошкодження розпізнаються або особливими білковими факторами, або РНК-полімеразою, яка робить зупинку на пошкодженому нуклеотиді. Після цього геліказа розплітає ділянку подвійної спіралі завдовжки 24–32 пари основ, пошкоджена одноланцюгова ділянка вирізається ендонуклеазами, і прогалина заповнюється реплікативною ДНК-полімеразою  $\delta$ .

### **Репарація некомплементарних пар основ – місметчів**

Незважаючи на редагування помилок під час реплікації, певна кількість неправильно спарених основ залишається в синтезованих ланцюгах ДНК. Зрозуміло, що при репарації таких *місметчів* (mismatch) із двох некомплементарних нуклеотидів замінити слід саме той, що входить до *синтезованого*, а не *матричного* ланцюга.

У бактеріальній клітині за репарацію місметчів відповідає система mutHLSU. По бактеріальному геному розподілені (на середній відстані 256 пар основ) короткі паліндромні послідовності STAG, у складі яких аденін піддається постреплікативному метилюванню. Однак певний час після синтезу метильованим є лише матричний (материнський) ланцюг. Саме за цей час спрацьовує система репарації (рис. 9.30): білок, що позначається як mutS, упізнає місметч і рекрутує білок mutL; останній взаємодіє із двома білками mutH, які зв'язуються із тетра-нуклеотидними паліндромними сайтами по обидва боки від місметча. У складі утвореного комплексу mutH набуває ендонуклеазної активності й робить одноланцюговий розріз у *неметильованому* ланцюзі в межах одного із сайтів (один із двох сайтів обирається випадково). Далі геліказа mutU (той самий білок, що й *uvrD*) розплітає подвійну спіраль, а екзонуклеаза руйнує ланцюг від розрізу до місметча й трохи далі. Нарешті прогалина заповнюється ДНК-полімеразою III, і одноланцюговий розрив зашивається лігазою.



**Рис. 9.30.** Бактеріальна система репарації місметчів mutHLSU

Система mutHLSU консервативна, гомологічні білки наявні також і в еукаріотів. Метилювання аденіну не використовується для дискримінації ланцюгів: білки еукаріотичної системи репарації місметчів пов'язані з реплісомою і ланцюгами ДНК, що синтезуються, тобто спрацьовують безпосередньо під час реплікації (рис. 9.31). Один із цих білків упізнає місметч і починає транслокацію до реплікативної вилки. Там розпізнається кінець ланцюга. Наявність кінця означає, що ланцюг синтезується, тобто саме в ньому слід замінювати нуклеотид. Екзонуклеаза руйнує ланцюг до місметча й далі, потім туди рекрутується реплікативна машинерія (полімерази  $\alpha$  і  $\delta$ , PCNA), яка відновлює ділянку ланцюга, що містила місметч.

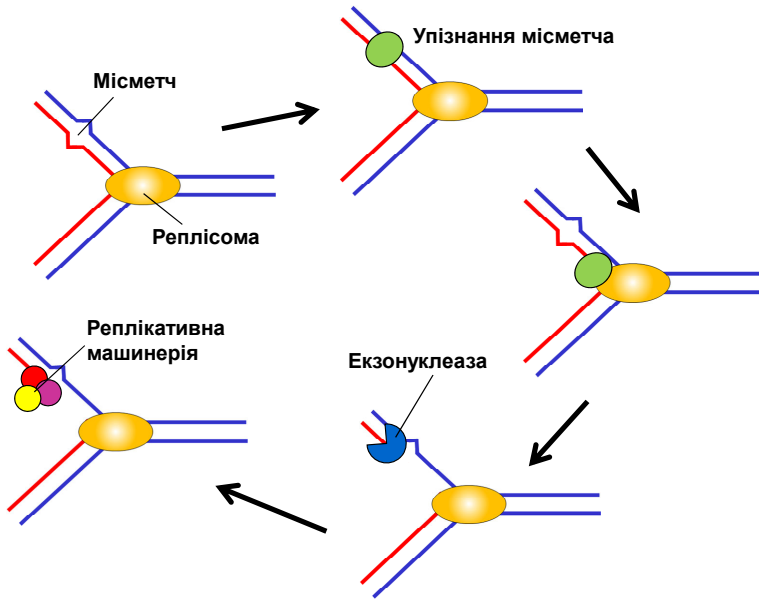


Рис. 9.31. Репарація місметчів у еукаріотів

### Репарація без репарації

Іноді у клітині активуються процеси, які прийнято також називати репарацією, хоча насправді вони є засобом здійснити реплікативний синтез ДНК, незважаючи на пошкодження її структури. Реплікативна машинерія зазвичай зупиняється, зустрічаючи пошкодження у складі матриці. Якщо таких пошкоджень надто багато і справжні репараційні системи не встигають їх виправити, то перемикання на неточний синтез ДНК дає клітині шанс на виживання. Пошкодження при цьому залишаються і, як наслідок, дають велику кількість мутацій. Усі процеси такого типу зазвичай об'єднують під назвою *SOS-репарації* або, що точніше, механізмів синтезу ДНК, толерантних до пошкоджень (damage tolerance mechanisms). Перемикання на неаккуратний синтез, який дозволяє долати перешкоди, відбувається

завдяки заміні реплікативних ДНК-полімераз на полімерази низької точності синтезу, які згадувались вище.

Інший шлях здійснити реплікацію через пошкодження, наприклад тиміновий димер у складі матриці, отримав назву **пост-реплікативної (рекомбінаційної) репарації**. За наявності тимінового димеру реплісома може його обійти, залишивши прогалину в ланцюзі, що синтезується (рис. 9.32). У цьому випадку на місце прогалини шляхом гомологічної рекомбінації (див. наступний розділ) вставляється ділянка сестринської молекули ДНК. Прогалина, що залишається при цьому в сестринській молекулі, легко заповнюється ДНК-полімеразою. Отже, як і при SOS-репарації, пошкодження залишається і може бути виправлене пізніше завдяки ексцизійній репарації.

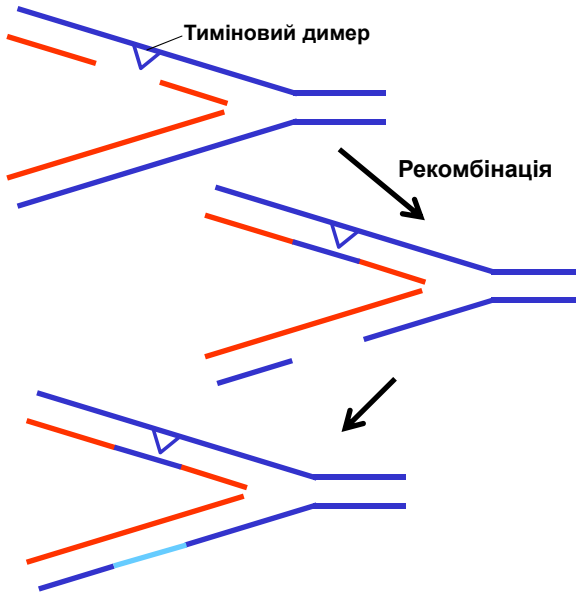


Рис. 9.32. Постреплікативна (рекомбінаційна) репарація

## Репарація дволанцюгових розривів

Точна репарація дволанцюгового розриву ДНК можлива лише під час або одразу після реплікації (у S- або G2-фазах клітинного циклу), коли розірвана ділянка може бути відновлена завдяки використанню сестринської молекули як матриці за механізмом гомологічної рекомбінації, яка розглядатиметься в наступному розділі.

В інших випадках цілісність розірваних молекул ДНК відновлюється в консервативному для всіх організмів процесі *негомологічного з'єднання кінців* (NHEJ – Non-Homologous End Joining), загальну схему якого представлено на рис. 9.33.

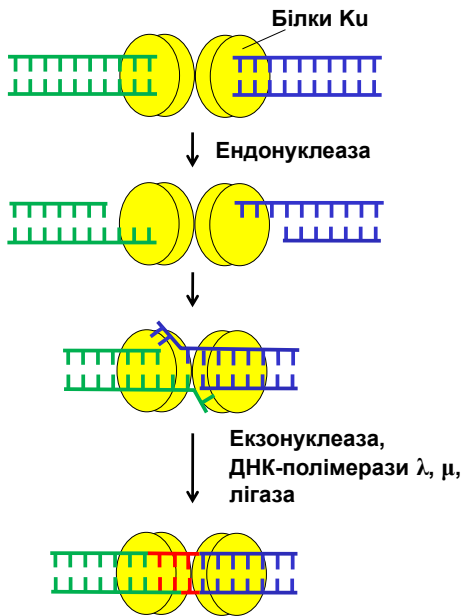


Рис. 9.33. Негомологічне з'єднання кінців ДНК

Ключову роль у процесі NHEJ відіграють білки Ku, які (у гетеродимерній формі) упізнають кінці ДНК, об'єднують їх у нековалентний комплекс і рекрутують до цього комплексу низку інших білків, зокрема специфічну ендонуклеазу, лігазу,

ДНК-полімерази  $\lambda$  та  $\mu$ . У складі комплексу відбувається відщеплення невеликих кінцевих ділянок ДНК і комплементарне спарювання одноланцюгових виростів різних молекул на невеликих ділянках (1–3 пари основ) (рис. 9.33). У результаті утворюється короткий дуплекс, зайві одноланцюгові ділянки відщеплюються екзонуклеазами, що спричиняє втрату кількох пар основ, а прогалини, що залишаються, заповнюються ДНК-полімеразами  $\lambda$  та  $\mu$ . Ці полімерази здатні приєднувати до 3'-кінця 1–2 випадкові нуклеотиди в безматричний спосіб, що може призводити до вставок кількох пар основ у місці з'єднання двох кінців. Нарешті, лігаза зшиває одноланцюгові розриви.

Отже, об'єднання двох кінців ДНК зумовлює вставки і/або втрати кількох пар основ – індукуються точкові мутації типу інсерцій / делецій (indels – insertions / deletions). Ця особливість NHEJ-системи використовується при рекомбінації імуноглобулінових генів (див. наступний розділ) з метою підвищення розмаїття їхніх послідовностей, а також у сучасних підходах до редагування геномів, які будуть розглянуті у розділі 11.

Оскільки при реалізації процесу NHEJ з'єднуються будь-які кінці будь-яких молекул ДНК, то за наявності великої кількості розривів це спричиняє також хромосомні аберації – об'єднання ділянок різних хромосом, транслокацій тощо.

## Контрольні запитання

1. Що таке реплікон? Охарактеризуйте основні типи репліконів.
2. Дайте визначення реплікативної вилки. Яка різниця між двома ланцюгами ДНК, що синтезуються під час реплікації? Що таке фрагменти Оказакі?
3. Які ферментативні активності мають бактеріальні ДНК-полімерази? Порівняйте особливості й функціональне значення ДНК-полімераз I і III.
4. Опишіть основні риси просторової структури ДНК-полімераз.
5. Які спільні риси мають РНК- і ДНК-полімерази? У чому полягають принципові відмінності між ними?
6. Як здійснюється редагування помилок при синтезі ДНК?

7. Що таке реплісома? Назвіть її основні компоненти та їхнє функціональне значення. Порівняйте організацію реплісом у бактерій і еукаріотів.

8. Дайте визначення гелікази. Яку роль вона відіграє в реплікації? Що таке білки SSB?

9. Порівняйте реплікативні гелікази бактерій і еукаріотів.

10. Що таке праймер? Навіщо він потрібен, яка його хімічна природа, який елемент реплісоми його синтезує?

11. Як видаляється праймер на початку фрагментів Оказакі у бактерій і еукаріотів?

12. Яку роль виконує лігаза при реплікації ДНК?

13. Як забезпечується висока процесивність ДНК-полімерази?

14. Які топологічні проблеми виникають під час реплікації і за допомогою чого вони розв'язуються?

15. Що таке ориджин? Як здійснюється ініціація реплікації у бактерій?

16. Укажіть особливості еукаріотичної системи реплікації порівняно із прокаріотичною.

17. Які ДНК-полімерази працюють в еукаріотичній клітині? Яка полімераза здійснює ініціацію синтезу ДНК? Порівняйте процеси ініціації реплікації у бактерій і еукаріотів.

18. Яку функціональну роль виконує теломераза? Опишіть механізм роботи цього ферменту.

19. Завдяки яким механізмам забезпечується спадкування епігенетичних маркерів при клітинному поділі?

20. Порівняйте ексцизійну репарацію основ і нуклеотидів.

21. Як здійснюється репарація місметчів у ДНК? Порівняйте цей процес у бактерій і еукаріотів.

20. Як відбувається репарація дволанцюгових розривів ДНК?

## Рекомендована література

Agarwal, S., Tafel, A.A. and Kanaar, R. (2006) DNA double-strand break repair and chromosome translocations. *DNA Repair* 5, 1075–1081.

Bailey, S., Eliason, W.K. and Steitz, T.A. (2007) Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* 318, 459–463.

Bell, S.D. (2019) Initiating DNA replication: a matter of prime importance. *Biochem. Soc. Trans.* 47, 351–356.

Bleichert, F., Botchan, M.R. and Berger, J.M. (2017) Mechanisms for initiating cellular DNA replication. *Science* 355, eaah6317.

Boehm, E.M., Goldenberg, M.S. and Washington, M.T. (2016) The many roles of PCNA in eukaryotic DNA replication. *Enzymes* 39, 231–254.

Burnham, D.R., Kose, H.B., Hoyle, R.B. and Yardimci, H. (2019) The mechanism of DNA unwinding by the eukaryotic replicative helicase. *Nat. Commun.* 10, 2159.

Chagin, V.O., Stear, J.H. and Cardoso, M.C. (2010) Organization of DNA replication. *Cold Spring Harb. Perspect. Biol.* 2, a000737.

Chang, H.H.Y., Pannunzio, N.R., Adachi, N. and Lieber, M.R. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell. Biol.* 18, 495–506.

Chatterjee, N. and Walker, G.C. (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* 58, 235–263.

Cortez, D. (2019) Replication-coupled DNA repair. *Mol. Cell* 74, 866–876.

Costa, A., Hood, I.V. and Berger, J.M. (2013) Mechanisms for initiating cellular DNA replication. *Annu. Rev. Biochem.* 82, 25–54.

Dewar, J.M. and Walter, J.C. (2017) Mechanisms of DNA replication termination. *Nat. Rev. Mol. Cell. Biol.* 18, 507–516.

*DNA replication: from old principles to new discoveries.* (2017) H.Masai and M.Foiani, eds. Singapore : Springer.

Donmez, I. and Patel, S.S. (2006) Mechanisms of a ring shaped helicase. *Nucleic Acids Res.* 34, 4216–4224.

Ekundayo, B. and Bleichert, F. (2019) Origins of DNA replication. *PLoS Genet.* 15, e1008320.

Evans, R.J., Davies, D.R., Bullard, J.M., et al. (2008) Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc. Natl. Acad. Sci. USA* 105, 20695–20700.

Fernandez-Leiro, R., Conrad, J., Scheres, S.H. and Lamers, M.H. (2015) Cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and  $\tau$ . *Elife* 4, e11134.

Fernandez-Leiro, R., Conrad, J., Yang, J.C., et al. (2017) Self-correcting mismatches during high-fidelity DNA replication. *Nat. Struct. Mol. Biol.* 24, 140–143.

Fragkos, M., Ganier, O., Coulombe, P. and Méchali, M. (2015) DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell. Biol.* 16, 360–374.

Georgescu, R., Yuan, Z., Bai, L., et al. (2017) Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc. Natl. Acad. Sci. USA* 114, E697–E706.

Hamdan, S.M. and Richardson, C.C. (2009) Motors, switches, and contacts in the replisome. *Annu. Rev. Biochem.* 78, 205–243.

Hanawalt, P.C. and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell. Biol.* 9, 958–970.

Hübscher, U., Maga, G. and Spadari, S. (2002) Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 71, 133–163.

Kuriyan, J. and O'Donnell, M. (1993) Sliding clamps of DNA polymerases. *J. Mol. Biol.* 234, 915–925.

Lai, W.K.M. and Pugh, B.F. (2017) Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell. Biol.* 18, 548–562.

Marians, K.J. (2000) Crawling and wiggling on DNA: structural insights to the mechanism of DNA unwinding by helicases. *Structure* 8, 227–235.

McInerney, P., Johnson, A., Katz, F. and O'Donnell, M. (2007) Characterization of a triple DNA polymerase replisome. *Mol. Cell* 27, 527–538.

Oakley, A.J. (2019) A structural view of bacterial DNA replication. *Protein Sci.* 28, 990–1004.

O'Donnell, M. (2006) Replisome architecture and dynamics in *Escherichia coli*. *J. Biol. Chem.* 281, 10653–10656.

O'Donnell, M., Langston, L. and Stillman, B. (2013) Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* 5, a010108.

Parker, M.W., Botchan, M.R. and Berger, J.M. (2017) Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* 52, 107–144.

Reardon, J.T. and Sancar, A. (2005) Nucleotide excision repair. *Prog. Nucleic Acid Res. Mol. Biol.* 79, 183–235.

Sauer, P.V., Gu, Y., Liu, W.H., et al. (2018) Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Res.* 46, 9907–9917.

Smith, E.M., Pendlebury, D.F. and Nandakumar, J. (2020) Structural biology of telomeres and telomerase. *Cell Mol. Life Sci.* 77, 61–79.

Stewart-Morgan, K.R., Petryk, N. and Groth, A. (2020) Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* 22, 361–371.

Xu, Z.Q. and Dixon, N.E. (2018) Bacterial replisomes. *Curr. Opin. Struct. Biol.* 53, 159–168.

Yuan, Z., Schneider, S., Dodd, T., Riera A, et al. (2020) Structural mechanism of helicase loading onto replication origin DNA by ORC-Cdc6. *Proc. Natl. Acad. Sci. USA* 117, 17747–17756.

## Розділ 10

# РЕКОМБІНАЦІЯ ДНК

Bond didn't defend the practice.  
He simply maintained that the more effort  
and ingenuity you put into gambling, the more you took out.  
*I. Fleming. Casino Royale*

Суттєвим моментом існування ДНК у живих системах є не тільки розглянуті в попередньому розділі процеси відновлення і збереження інформації, що міститься в послідовності нуклеотидів, а й різноманітні операції, спрямовані на перетасування цієї інформації, наслідком чого стає створення нових комбінацій генів. Нові комбінації зумовлюють розмаїття особин у межах одного виду (достатньо подивитись на людей, що вас оточують, щоб у такому розмаїтті переконатись).

Серед процесів рекомбінації (recombination), які спричиняють генетичне розмаїття і будуть розглянуті у цьому розділі, можна назвати такі:

- гомологічна рекомбінація – обмін ділянками між доволі довгими молекулами ДНК із подібними (гомологічними) послідовностями пар основ; відбувається в усіх організмах, які розмножуються статевим шляхом, між гомологічними хромосомами при мейозі (наслідком є кросинговер), а у прокаріотів – наприклад, після кон'югації двох бактеріальних клітин і проникнення ДНК з однієї в іншу;
- сайт-специфічна рекомбінація – вирізання / вбудовування однієї молекули ДНК із/в іншу або зміна орієнтації (інверсія) фрагмента ДНК у межах однієї молекули. Рекомбінація цього типу здійснюється за рахунок упізнання специфічними білками коротких елементів послідовності ДНК;

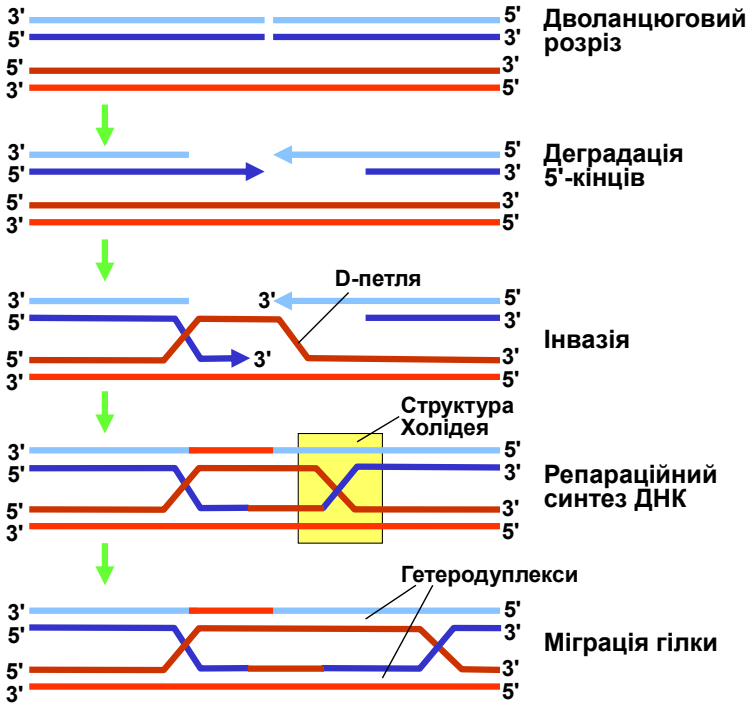
- незаконна рекомбінація – об'єднання двох молекул ДНК, які не мають ані гомології, ані специфічних елементів послідовності;
- особливі типи рекомбінації при дозріванні генів імуноглобулінів у імунокomпетентних клітинах і переміщення в межах геному мобільних елементів можна розглядати як варіанти сайт-специфічної рекомбінації із внеском додаткових механізмів.

## Гомологічна рекомбінація

Необхідною умовою для здійснення гомологічної рекомбінації є загальна гомологія між двома молекулами ДНК по всій довжині, тобто висока подібність (зовсім не обов'язково повна ідентичність) нуклеотидних послідовностей.

Загальну схему початкового етапу гомологічної рекомбінації зображено на рис. 10.1. Ініціувальною подією є *дволанцюговий розріз* в одній з гомологічних молекул. Цей розріз розширюється шляхом 5'-екзонуклеазної деградації ДНК, у результаті чого в місці розрізу залишаються два одноланцюгові 3'-хвости. Один із них здійснює *інвазію* – утворює подвійну спіраль з антипаралельним ланцюгом інтактної гомологічної молекули ДНК. Інший ланцюг останньої виштовхується із дуплекса у вигляді одноланцюгової D-петлі. D-петля (від *displacement*) разом із 3'-хвостом здатна переміщуватись у пошуку гомології – максимальної комплементарності в межах подвійної спіралі, що утворилася між ланцюгами двох гомологічних молекул ДНК. На наступному кроці відбувається репараційний синтез ДНК: два 3'-кінці розірваної молекули ДНК подовжуються ДНК-полімеразами з використанням як матриць двох ланцюгів інтактної молекули. До цього моменту схема на рис. 10.1 є одночасно *схемою точної репарації дволанцюгового розриву* в одній із двох сестринських молекул ДНК під час реплікації. У такому випадку дві молекули, що здійснюють рекомбінацію, є не просто гомологічними, а повністю ідентичними.

Під час рекомбінації відновлення цілісності ДНК є тільки завершенням початкового етапу. У результаті інвазії і репараційного синтезу дві молекули ДНК об'єднуються в чотириланцюгову структуру із двома перехрестями ланцюгів – *структури Холідея* (Robin Holliday). Кожна структура Холідея може переміщуватись (*міграція гілки*), результатом чого є подовження *гетеродуплекса* – подвійної спіралі між двома майже комплементарними ланцюгами двох гомологічних молекул ДНК.



**Рис. 10.1.** Початкові стадії процесу гомологічної рекомбінації

Вважається, що за схемою на рис. 10.1 відбуваються початкові стадії гомологічної рекомбінації в більшості про- та еукаріотів. Молекулярні механізми цих процесів найкраще вивчені для бактерій. У *E. coli* два кінці розірваної молекули ДНК упізнаються комплексом трьох білків – **recBCD** (рис. 10.2), який має геліказну й екзонуклеазні активності. Комплекс починає рухати-

ся, руйнуючи подвійну спіраль і одночасно деградує ДНК за рахунок своїх 5'- і 3'-екзонуклеазних активностей. У  $\chi$ -сайті (сайт підвищеної частоти рекомбінації в *E. coli* – послідовність завдовжки 8 пар основ, що зустрічається із середньою періодичністю у 5 тис. пар основ) 3'-екзонуклеазна активність гальмується, результатом чого стає утворення одноланцюгового 3'-кінцевого хвоста. Цей хвіст одразу вкривається білком **recA**. Останній має два сайти зв'язування із ДНК. За рахунок першого з них відбувається (за присутності АТР як кофактора) кооперативна взаємодія з одноланцюговою ДНК: молекули білка оточують полінуклеотидний ланцюг, формуючи на його поверхні праву спіраль. Саме цей комплекс здійснює інвазію в подвійну спіраль інтактної гомологічної молекули ДНК. Другий сайт взаємодії **recA** з ДНК використовується на цій стадії для утворення комплексу із другим ланцюгом; зв'язаний із **recA** гетеродуплекс перебуває у стані суттєво деформованої подвійної спіралі (рис. 10.3). За умови комплементарності двох ланцюгів здійснюється гідроліз АТР, після чого **recA** втрачає спорідненість до ДНК, залишаючи сформований гетеродуплекс.

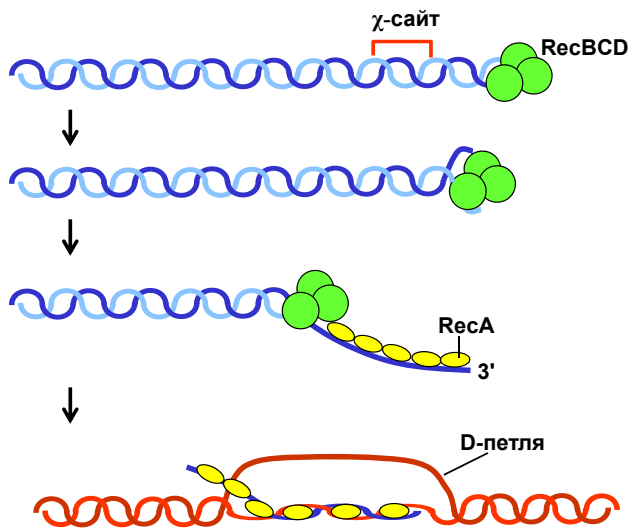
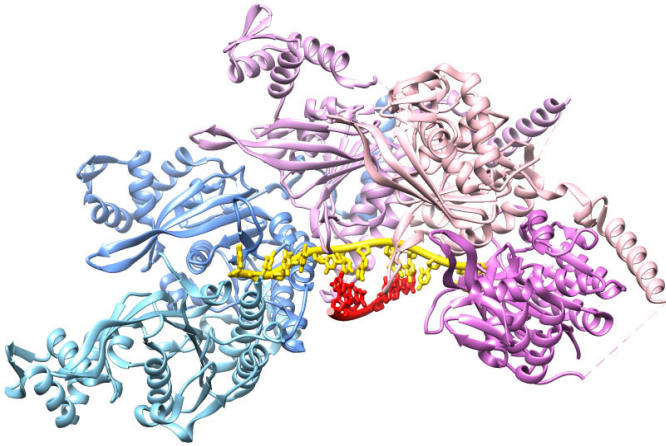


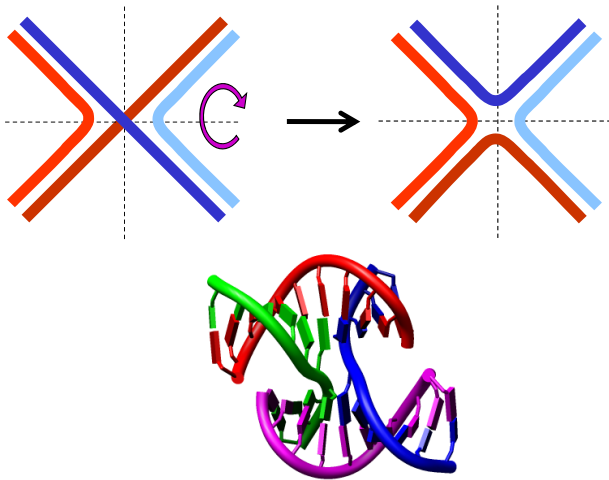
Рис. 10.2. Механізм початкових стадій гомологічної рекомбінації в *E. coli*



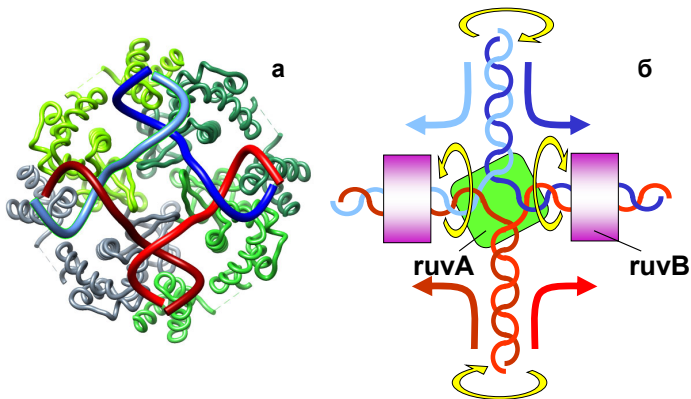
**Рис. 10.3.** Структура комплексу п'яти субодиниць recA з гетеродуплексом (3CMT)

Повертаючись до останньої панелі на рис. 10.1, розглянемо окрему структуру Холідея, яку часто зображують саме в такому вигляді – зі схрещеними ланцюгами. Як показано на рис. 10.4, її можна піддати формальній ізомеризації, повернувши два дволанцюгові кінці на  $180^\circ$ . Насправді структура зі схрещеними ланцюгами не може існувати – штучні *чотириланцюгові з'єднання* (four-way junctions), які служать моделями структур Холідея, набувають форми, що зображена на рис. 10.4. У формі без перехрестя ланцюгів (як праворуч на рис. 10.4) структура Холідея фіксується завдяки її взаємодії з гомотетрамерним білком **ruvA**: білок зв'язується із центром хреста, утримуючи чотири одноланцюгові ділянки у приблизно планарній квадратній конфігурації (рис. 10.5, *a*).

З **ruvA** та двома дуплексами, що виходять із хреста у протилежних напрямках, взаємодіє білок **ruvB** (шість субодиниць, які оточують подвійну спіраль кільцем). RuvB працює як АТР-залежна геліказа: два гексамери **ruvB** генерують обертальні рухи подвійної спіралі у протилежних напрямках, що спричиняє протягування ланцюгів через комплекс **ruvA/ruvB** (рис. 10.5, *b*). Саме активність **ruvB** забезпечує переміщення структури Холідея – міграцію гілки з одночасним подовженням гетеродуплекса.



**Рис. 10.4.** Ізомеризація структури Холідея і структура чотириланцюгового з'єднання ДНК за відсутності білків (467D)

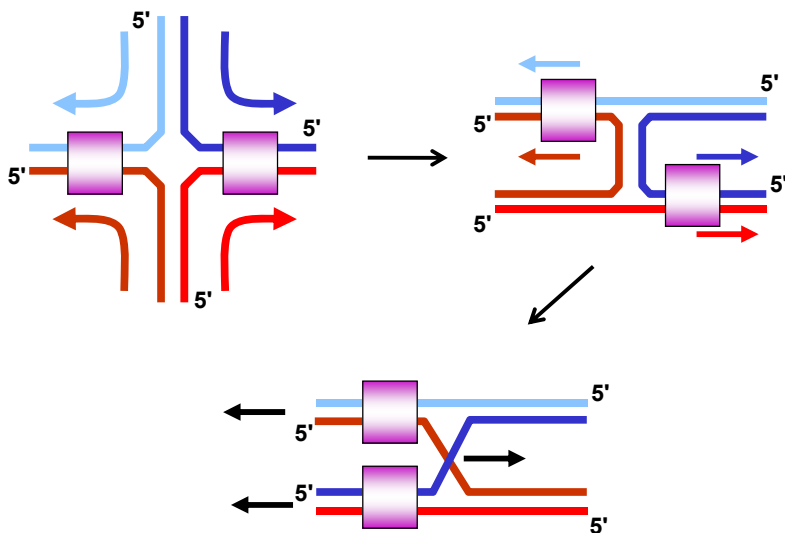


**Рис. 10.5.** а – комплекс тетрамеру *ruvA* зі структурою Холідея (1BDX);  
б – схема переміщення структури Холідея за рахунок активності *ruvB*: стрілки вказують напрямки обертання дуплексів і руху ланцюгів через комплекс *ruvA/ruvB* (колір стрілок збігається із таким ланцюгів)

Зі схеми на рис. 10.6 видно, як можна ототожнити протягування чотирьох полінуклеотидних ланцюгів із міграцією гілки. На схемі у два етапи проведено формальну ізомеризацію планарної конфігурації структури Холідея у конфігурацію з перехрестям: на першому етапі два ланцюги у складі хреста випрямляються, на другому – нижня частина отриманої конфігурації обертається на  $180^\circ$ . Очевидно, що зображене на рис. 10.5, 10.6 протягування ланцюгів через комплекс  $\text{tuvA/tuvB}$  є еквівалентним зсуву перехрестя у правий бік.

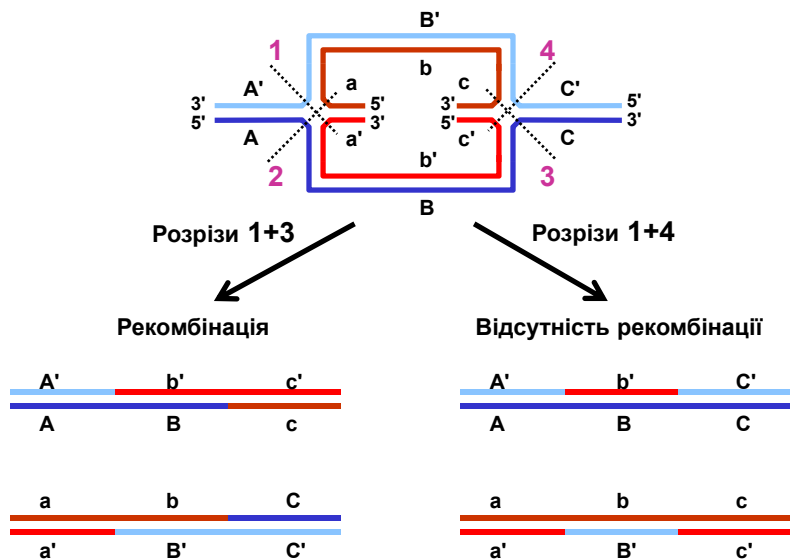
Останньою подією рекомбінації є розділення (resolution) структури Холідея *резольвазою* – білком **ruvC**. Резольваза є ендонуклеазою, дві молекули якої взаємодіють із комплексом  $\text{tuvA/tuvB}$  і двома ланцюгами хреста, розташованими один навпроти одного: є два рівноймовірні варіанти такого зв'язування. Отже, резольваза робить дволанцюговий розріз через хрест двома можливими шляхами (рис. 10.7). Після наступного зшивання розривів лігазою залишаються дві дволанцюгові молекули ДНК.

На рис. 10.7 показано ізомеризовану чотириланцюгову структуру з рис. 10.1 – дві структури Холідея з перехрестями перетворені на планарні (ділянки, синтезовані шляхом репарації, не позначено) – і результат її розділення. Із чотирьох можливих комбінацій розділення двох структур Холідея показано дві. Одна із цих комбінацій викликає рекомбінацію (аналогічно рекомбінантною є також пара розрізів  $2 + 4$ ): дві гомологічні молекули ДНК обмінялися ділянками, і умовна "абетка", якою позначено ділянки ДНК, змінює регістр – великі літери замінюються на маленькі, і навпаки. Пари розрізів  $1 + 4$  і  $2 + 3$  не спричиняють рекомбінації. Отже, рекомбінація при розділенні структур Холідея відбувається з імовірністю 50 %.



**Рис. 10.6.** Спрощена схема комплексу *guvB* зі структурою Холідея з рис. 10.5, б та її формальні перетворення на дві топологічно еквівалентні конфігурації. Остання конфігурація збігається із правою структурою Холідея з рис. 10.1. Кольорові стрілки вказують напрямок руху ланцюгів через *guvB*, чорні – напрямком руху дуплексів і перехрестя

Незалежно від того, чи відбулася рекомбінація, усі продукти містять гетеродуплекси (середня частина всіх остаточно отриманих молекул на рис. 10.7). Оскільки гетеродуплекси складаються з *майже* комплементарних ланцюгів, то вони містять місметчі. Відповідно кінцевою операцією, яка завершує процес гомологічної рекомбінації, стає репарація місметчів описаною в попередньому розділі системою. На відміну від того, як ця система спрацьовує після реплікації, після рекомбінації ланцюг, де відбувається заміна нуклеотидів, обирається випадково. Наприклад, центральний фрагмент першої молекули на рис. 10.7 перетворюється або на *V/V'*, або на *v/b'*. Якщо на цій ділянці розташований ген, то шляхом репарації обирається один із його алелів – *V* або *v*. Отже, побічним ефектом рекомбінації є відоме в генетиці явище *конверсії гена*.



**Рис. 10.7.** Схема розділення двох структур Холідея.

Угорі: конфігурація чотирьох ланцюгів, еквівалентна конфігурації на нижній панелі з рис. 10.1. Літерами позначено ділянки ланцюгів (великі й маленькі літери відповідають гомологічним ділянкам двох молекул, літери зі штрихом і без – комплементарним ділянкам вихідних дуплексів). Цифрами 1–4 позначено можливі розрізи резольвазою. Унизу: дві пари дволанцюгових молекул після розділення структур Холідея, отримані в результаті відповідних розрізів

Еукаріотичні клітини містять білки, гомологічні бактеріальним білкам, залученим до гомологічної рекомбінації. Наприклад, білок Rad51 дріжджів і людини є гомологічним білку RecA й виконує подібні функції. Цей білок використовується еукаріотичною клітиною у процесі мітотичної рекомбінації, тобто при репарації дволанцюгових розривів за рекомбінаційним механізмом; при рекомбінації у процесі мейозу він діє сумісно зі своїм гомологом Dmc1. Кілька еукаріотичних білків виконують важливі допоміжні функції. Зокрема, білок Rad54, член родини АТР-залежних геліказ SWI/SNF (розділ 6), взаємодіє з Rad51, сприяє пошуку гомології на ранніх стадіях рекомбінаційного

процесу, утворенню D-петлі, міграції гілки й ремоделюванню хроматину в процесі рекомбінації. Ініціюючий рекомбінацію дволанцюговий розріз індукується у дріжджів білком Spo11, що належить до родини ДНК-топоізомераз II. Гомологічні білки знайдено в інших еукаріотів і бактерій. Отже, принципові механізми гомологічної рекомбінації спільні для всіх біологічних систем, хоча й можуть розрізнятися за деякими деталями.

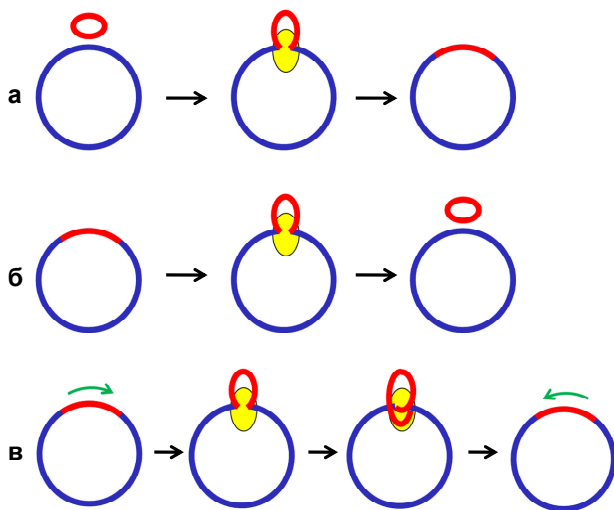
## Сайт-специфічна рекомбінація

Сайт-специфічна рекомбінація, на відміну від гомологічної, базується на специфічному впізнанні певними білками відносно невеликих елементів нуклеотидної послідовності (сайтів). Процеси сайт-специфічної рекомбінації поширені насамперед у бактеріофагів. Ключову роль у цих процесах відіграють різноманітні *рекомбінази*, які забезпечують вбудовування (інсерцію, інтеграцію) фагових ДНК у бактеріальний геном (відносно невеликої циркулярної ДНК у циркулярну бактеріальну хромосому), вирізання (ексцизію) фрагмента ДНК з його одночасною циркуляризацією, зміну напрямку фрагмента ДНК (рис. 10.8).

Прикладом процесу першого типу є інтеграція ДНК бактеріофага  $\lambda$  у бактеріальну хромосому за допомогою *інтегрази* – рекомбінази, яка є продуктом одного з генів рекомбінації бактеріофага (див. також розділ 6). У складі фагової ДНК є сайт attP – ділянка завдовжки 270 пар основ, у складі бактеріальної ДНК – сайт attB (23 пари основ). Обидва сайти мають невеличкі ділянки однакової послідовності. Сайти упізнаються інтегразою, допоміжну роль виконують ще кілька білків, зокрема бактеріальний білок IHF (Integration Host Factor), який індукує різке вигинання у ДНК і стабілізує структуру проміжних комплексів, що виникають у процесі рекомбінації (див. рис. 10.10).

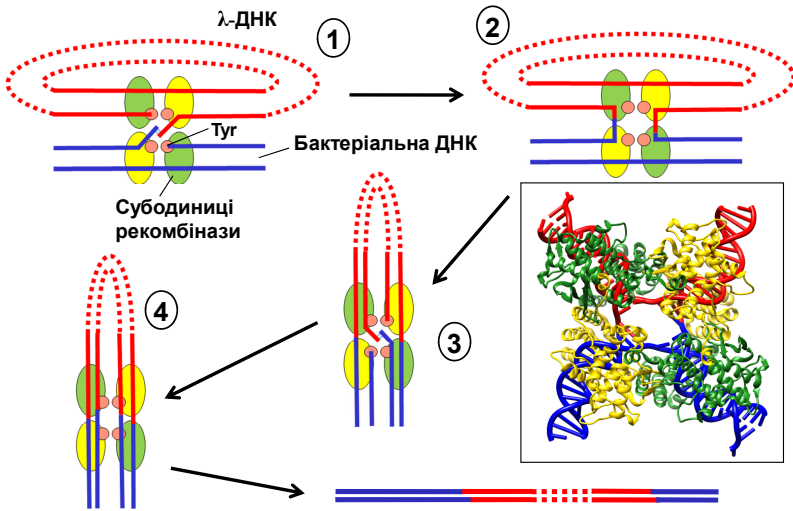
За механізмом дії інтеграза, як і інші фагові рекомбінази, є сайт-специфічною ДНК-топоізомеразою I (див. розділ 3). Чотири субодиниці білка містять активні центри, у кожному з яких є залишок Tyr. На першому етапі інтеграції два активні центри роблять два одностанцюгові розрізи в бактеріальній і фаговій ДНК,

одночасно ковалентно приєднуючи кінцеві фосфати до залишків Туг (рис. 10.9). Далі відновлюється фосфодієфірний зв'язок, але з відповідним кінцем іншого дуплекса – відбувається з'єднання ланцюгів, які належать різним молекулам. На цьому етапі утворюється конфігурація чотирьох ланцюгів, еквівалентна структурі Холідея. Потім відбувається ізомеризація структури Холідея, друга пара субодиниць робить другу пару розривів (на відстані семи пар основ від першого розриву в кожному ланцюзі) і здійснюється друга пара обмінів ланцюгами, що й спричиняє інтеграцію (рис. 10.9).

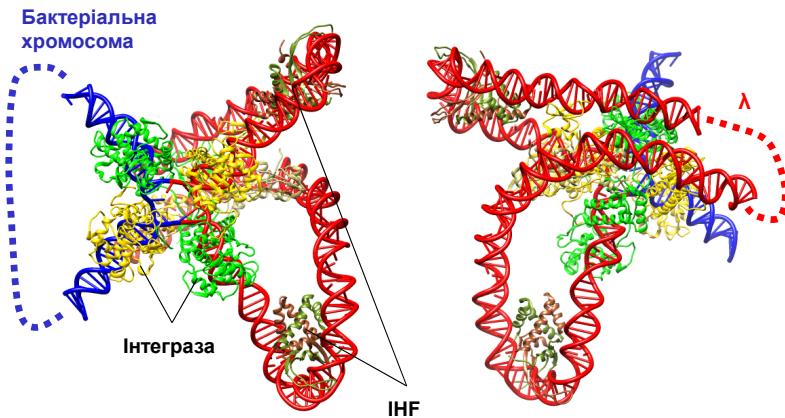


**Рис. 10.8.** Типи активності рекомбіназ бактеріофагів: інсерція (а), вирізання (б) і зміна напрямку (інверсія) фрагмента ДНК (в)

Перебіг процесу, зображеного на рис. 10.9, у зворотному напрямку спричиняє вирізання фрагмента ДНК із його одночасною циркуляризациєю. При цьому також виникає структура Холідея як інтермедіат рекомбінації. Саме так – за допомогою тієї самої інтегрази – здійснюється вирізання  $\lambda$ -ДНК із бактеріального геному (як на рис. 10.8, б). Інтермедіат рекомбінації у цьому процесі зображено на рис. 10.10.



**Рис. 10.9.** Інтеграція ДНК бактеріофага  $\lambda$  у бактеріальний геном.  
 1 – утворення першої пари тимчасових одноланцюгових розрізів;  
 2 – зшивання ланцюгів різних молекул (утворення структури Холідея);  
 3 – ізомеризація структури Холідея і здійснення другої пари розрізів;  
 4 – зшивання ланцюгів.  
 На вставці: інтеграза у комплексі зі структурою Холідея (5J0N)



**Рис. 10.10.** Структура інтермедиату у процесі вирізання  $\lambda$ -ДНК із бактеріальної хромосоми у двох проєкціях (5J0N)

Інший приклад такого самого типу – вирізання і циркуляризація мономерної ДНК бактеріофага P1 із великого попередника рекомбіназою Cre, структура якої дуже схожа на таку інтегрази бактеріофага  $\lambda$ . Попередник, який з'являється внаслідок реплікації фагової ДНК за механізмом кільця, що котиться (рис. 9.2, в), містить велику кількість фрагментів ДНК бактеріофага, які тандемно повторюються і на границях між якими містяться стандартні елементи послідовності – loxP-сайти. Ці сайти упізнаються рекомбіназою Cre. Система рекомбінації Cre-loxP широко використовується як дослідницький засіб – вставлення в геном еукаріотичної клітини loxP-сайтів і гена рекомбінази дозволяє видаляти відповідний геномний фрагмент після активації експресії гена.

Подібно до інтегрази й рекомбінази Cre працює *інвертаза* – сайт-специфічна топоізомераза I, яка здійснює інверсію ділянки ДНК бактеріофага  $\mu$ . Ділянка завдовжки 3 тис. пар основ, що піддається інверсії, містить на кінцях короткі однакові елементи послідовності. Чотири субодиниці інвертази роблять чотири тимчасові розрізи – відбувається обмін кінців і відновлення зв'язків, результатом чого стає зміна напрямку (рис. 10.8, в). Інвертована ділянка містить два гени, що перекриваються, на різних ланцюгах ДНК (тобто такі, що піддаються транскрипції в різних напрямках). У результаті інверсії той чи інший ген (залежно від виду бактерії, у яку потрапив фаг) підставляється під промотор і транскрибується.

В еукаріотів два типи процесів – переміщення мобільних елементів і дозрівання імуноглобулінових генів – здійснюються за допомогою механізмів, що нагадують сайт-специфічну рекомбінацію (без участі спеціалізованих топоізомераз).

Розглянуті вище два типи рекомбінації залежать від подібності нуклеотидних послідовностей двох молекул ДНК або за великою довжиною (гомологічна рекомбінація), або в коротких сайтах (сайт-специфічна рекомбінація). Третім типом є *незаконна рекомбінація*, коли з'єднання молекул ДНК без гомології відбувається в будь-якому місці послідовності. Загальним механізмом незаконної рекомбінації є репарація дволанцюгових розривів системою негомологічного з'єднання кінців (див. рис. 9.33),

яка може спричинити з'єднання чужорідних молекул ДНК. Такий процес, на додаток до сайт-специфічної рекомбінації, також здійснюється при дозріванні генів імуноглобулінів.

## V(D)J-рекомбінація імуноглобулінових генів

Основою імунної відповіді при появі чужорідної молекули або частинки – антигена – є синтез імуноглобуліну певного типу, який має до цього антигена високу специфічну спорідненість. Кількість типів імуноглобулінів, як і антигенів, практично необмежена. Зрозуміло, що закодувати таке розмаїття у вигляді окремих імуноглобулінових генів неможливо. Еволюційним розв'язанням цієї проблеми став особливий рекомбінаційний процес дозрівання імуноглобулінових генів.

Молекула імуноглобуліну складається із двох важких і двох легких поліпептидних ланцюгів (рис. 10.11). Кожен із ланцюгів має константну С-кінцеву (спільну для всіх імуноглобулінів) і варіабельну N-кінцеву частини. Варіабельні частини кожної пари важкого й легкого ланцюгів формують два антиген-зв'язувальні сайти на поверхні молекули.

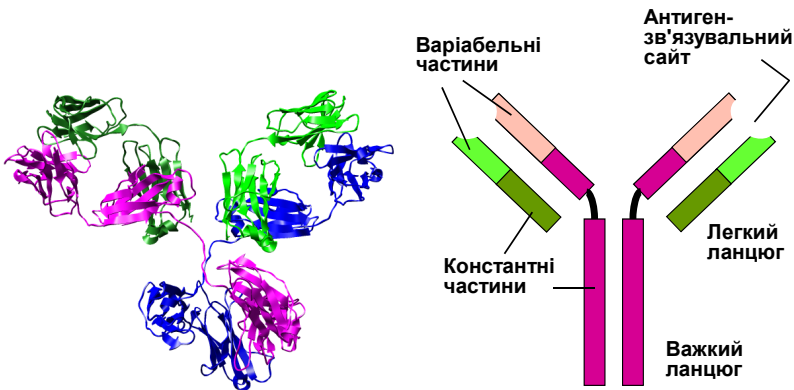
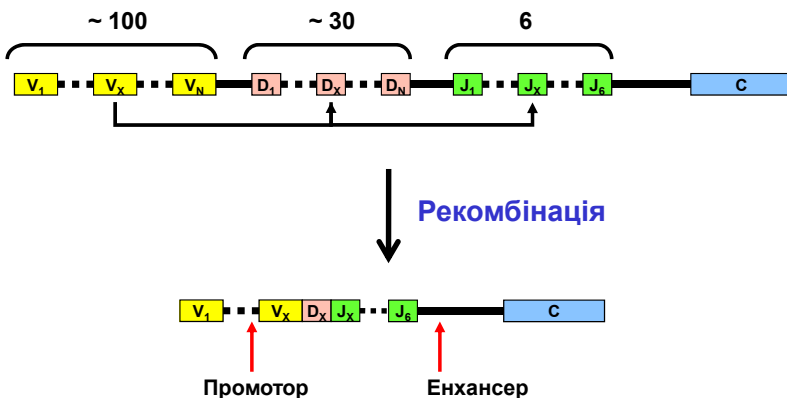


Рис. 10.11. Структура імуноглобуліну IgG1 (1IGY)  
і схема будови молекули

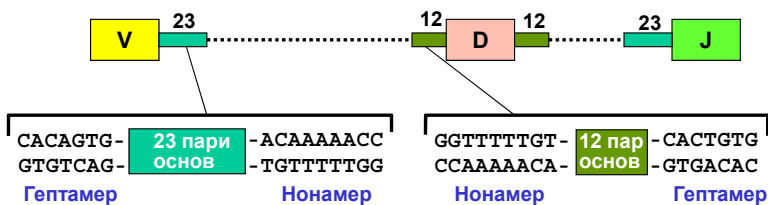
Незрілі імуноглобулінові локуси мають вигляд кластерів окремих елементів нуклеотидної послідовності – блоків, з яких шляхом рекомбінації в імунокомпетентній клітині збирається активний імуноглобуліновий ген. Кластер важкого ланцюга містить ~100 V-сегментів (від *variable*), що тандемно повторюються (послідовності всіх сегментів розрізняються між собою), ~30 D-сегментів (від *diversity*), 6 J-сегментів (від *joining*) і C-ділянку, яка кодує константну частину ланцюга (рис. 10.12). Аналогічно побудований кластер легкого ланцюга, який містить тільки два типи варіабельних сегментів (~100 V- і 4 J-сегменти). Активний ген збирається із сегментів трьох (або двох) типів, як із кубиків, шляхом V(D)J-рекомбінації: один випадковий D-сегмент з'єднується з випадковим J-сегментом (ділянка між ними вирізається), до них приєднується один із V-сегментів (рис. 10.12). Аналогічно для легкого ланцюга об'єднуються один V- з одним із J-сегментів. Перед кожним V-сегментом розташований промотор, у спейсері перед C-ділянкою – енансер. Їхнє зближення після рекомбінації активує транскрипцію, зайві спейсери та інтрони константної частини видаляються із мРНК шляхом сплайсингу.

Для важкого ланцюга можуть реалізуватися ~18 тис. комбінацій між сегментами трьох типів, для легкого – ~400 комбінацій, а загалом за рахунок рекомбінації для обох ланцюгів утворюються ~7 млн варіантів послідовності. Додаткова варіабельність забезпечується за рахунок індукції мутацій у межах V-сегментів, а також завдяки застосуванню механізмів незаконної рекомбінації при з'єднанні сегментів (див. нижче).



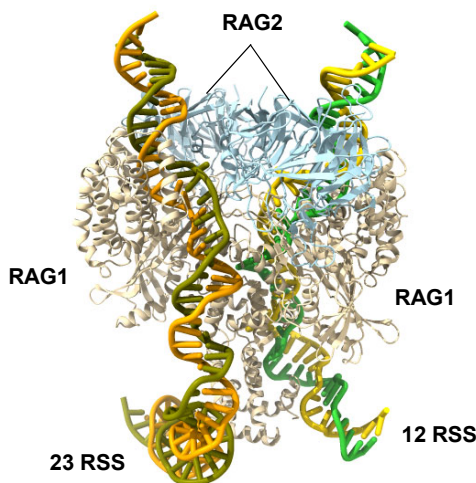
**Рис. 10.12.** Будова кластера генів важкого ланцюга імуноглобуліну та схема збирання активного гена. Константна частина (C) містить інтрони, які не показано

Рекомбінація сегментів залежить від сигнальних послідовностей RSS (Recombination Signal Sequence), що фланкують із двох боків кожен із сегментів (рис. 10.13) і є сайтами рекомбінації. Сигнальні послідовності містять дві консервативні ділянки по сім і дев'ять пар основ, розділені неконсервативним спейсером. Послідовність спейсера не має значення – важливий лише його розмір: два типи спейсерів завдовжки 23 і 12 пар основ визначають два типи RSS. Об'єднання сегментів при рекомбінації можливе лише за умови виконання правила "12/23": об'єднуються тільки сегменти, фланковані RSS різних типів. Завдяки цьому виключається об'єднання однотипних сегментів (напр. V-V) чи пропущення D-сегмента.



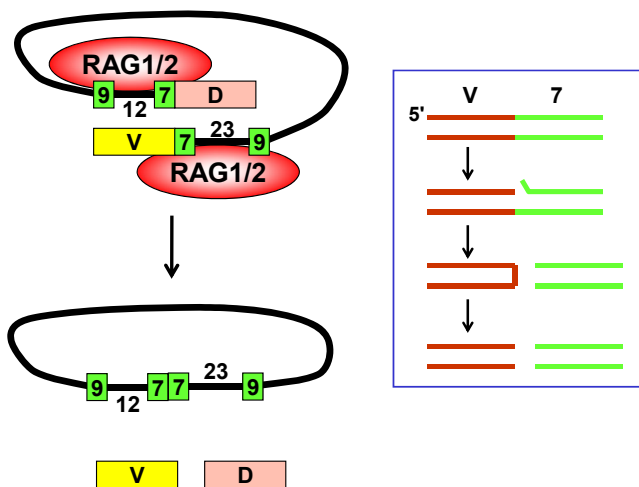
**Рис. 10.13.** Організація сигнальних послідовностей (RSS), які контролюють рекомбінацію в імуноглобулінових генах

Консервативні елементи сигнальних послідовностей упізнаються білками RAG1 і RAG2 (продукти генів *RAG* – Recombination Activating Gene, активних лише у клітинах-попередниках В- і Т-лімфоцитів). Білки об'єднують у єдиний комплекс дві сигнальні послідовності за виконання умови "12/23" (рис. 10.14, 10.15) і каталізують низку реакцій, які схематично показано на рис. 10.15. Перша реакція – одноланцюгове розрізання між сегментом і сигнальним гептамером. Далі відбувається трансестерифікація – розривання зв'язку в іншому ланцюзі й ковалентне замикання кінців сегмента у шпильку. Унаслідок таких самих реакцій у другій RSS вирізається, замикається в кільце й видаляється ділянка між двома сегментами.



**Рис. 10.14.** Комплекс білків RAG1/2 із двома RSS (6OEO)

Ковалентні шпильки на кінцях сегментів нестабільні – вони швидко розмикаються (рис. 10.15). При цьому необов'язково утворюється тупий (без одноланцюгових виростів) кінець, оскільки шпилька може розімкнутися асиметрично.



**Рис. 10.15.** Залежне від білків RAG вирізання ділянки між двома сегментами імуноглобулінового гена. На вставці: послідовність подій на границі між V-елементом і гептамером сигнальної послідовності

На останніх стадіях 3'-кінці сегментів добудовуються дезоксирибонуклеотидил-трансферазою (рис. 10.16) – ферментом, який без участі матриці приєднує випадкові нуклеотиди. Зрозуміло, що цей процес значно підвищує розмаїття послідовностей зрілих генів імуноглобулінів. Частина таких генів буде неактивною завдяки стоп-кодонам, що з певною імовірністю виникають при такому випадковому синтезі, проте це є прийнятною платою за загальне збільшення варіантів.

Нарешті кінці сегментів з'єднуються за допомогою системи репарації дволанцюгових розривів NHEJ (див. рис. 9.33): відбувається пошук мікрогомології між двома 3'-кінцевими виростами, внесення інсерцій / делецій і репаративний синтез ДНК, що остаточно заповнює прогалини, тобто на останніх стадіях рекомбінації процес з'єднання сегментів імуноглобулінових генів навмисно загрублюється з метою підвищення варіабельності.

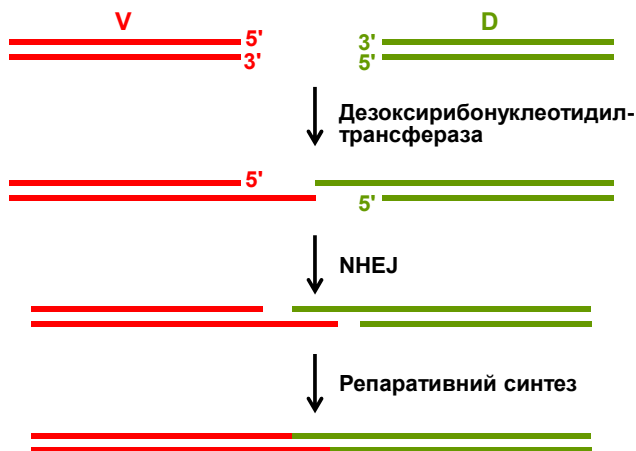


Рис. 10.16. Завершальні стадії процесу з'єднання сегментів імуноглобулінового гена

Отже, за своїми механізмами залежна від білків RAG сайт-специфічна рекомбінація імуноглобулінових генів не має нічого спільного з такою рекомбінацією у бактеріофагів. Відповідно й самі білки RAG1/2 не схожі на фагові рекомбінази, вони є гомологами *транспозаз* – ферментів, які забезпечують переміщення (транспозиції) мобільних елементів.

## Транспозиції мобільних елементів

Переміщення мобільних (інтерсперсних) елементів (див. також розділ 4) є ще одним особливим типом рекомбінації ДНК, яка відбувається в усіх організмів. Мобільні елементи можна поділити на два основні класи (рис. 10.17): *ДНК-транспозони*, що переміщуються за принципом "cut and paste", і ретротранспозони – елементи, переміщення яких відбувається шляхом синтезу РНК із подальшим синтезом ДНК на РНК-матриці. Основними типами мобільних елементів, які належать до другого класу, є *LTR-ретротранспозони* з довгими кінцевими повторами (Long Terminal Repeats), елементи *LINE* (Long Interspersed Elements) і елементи *SINE* (Short Interspersed Elements).

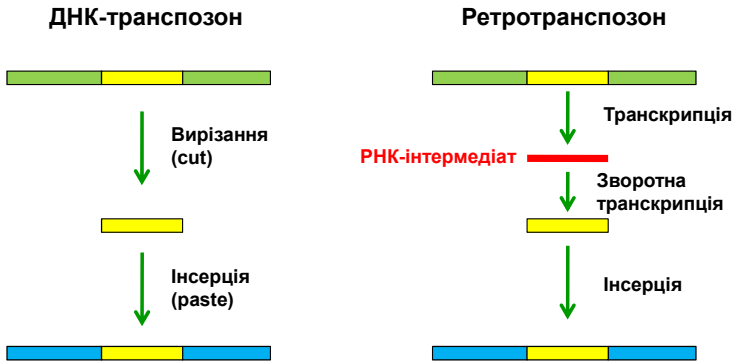


Рис. 10.17. Два класи мобільних елементів

Переміщення мобільних елементів є доволі рідкісною подією – у бактерій відбувається одна транспозиція на  $10^5$ – $10^6$  клітин. Проте активність мобільних елементів має важливе біологічне значення. Вбудовування мобільних елементів у кодувальні частини генів спричиняє мутації, у регуляторні – зміну транскрипційної активності. У процесі переміщення мобільний елемент може захопити той чи інший регуляторний елемент послідовності та вбудувати його в іншу регуляторну зону. Переміщення генетичного матеріалу разом із мобільними елементами розглядається як один із найважливіших еволюційних факторів.

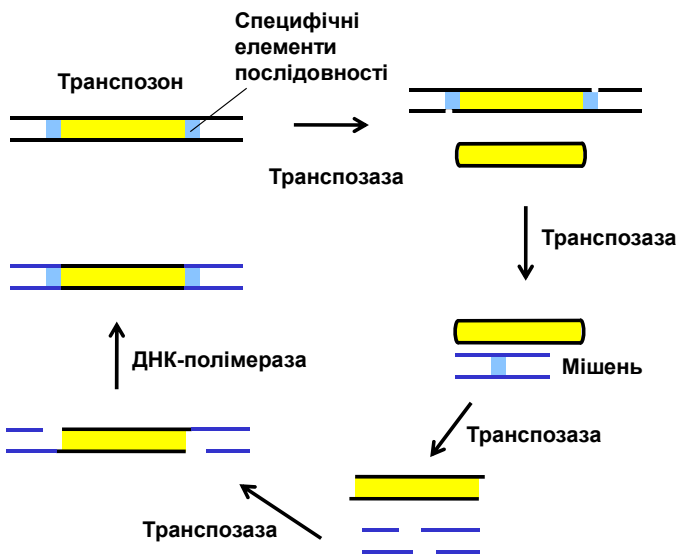
### ДНК-транспозони

Мобільні елементи цього типу часто містять один або два гени, що кодують один або два білки під спільною назвою *транспозаза*. Транспозази здійснюють каталіз хімічних реакцій, що забезпечують транспозицію елемента – його вирізання з донорного сайту та вбудовування у сайт-мішень. Крім генів транспозази, мобільний елемент може містити інші гени, або навпаки – гени транспозази можуть бути пошкодженими. В останньому випадку транспозиція елемента відбувається з використанням транспозази, закодованої іншим ДНК-транспозоном.

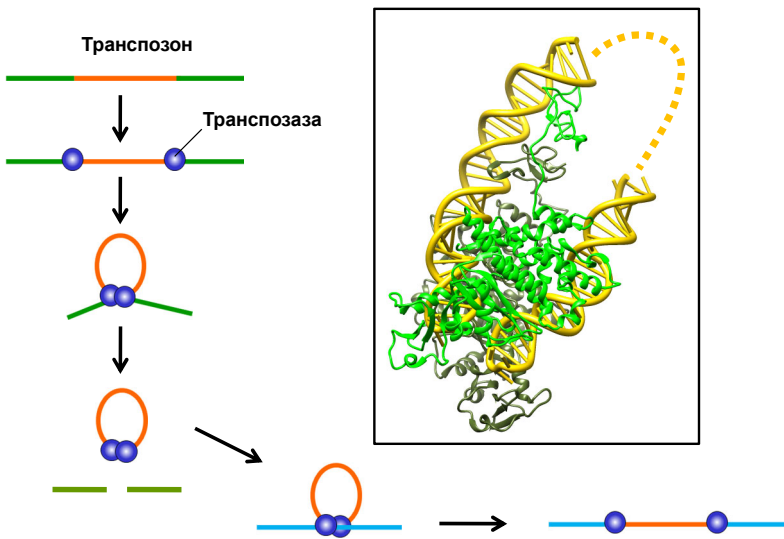
Кодувальна частина транспозона фланкована невеликими елементами послідовності, які розпізнаються двома молекулами транспозази, що індують у цих сайтах одноланцюгові розрізи

в різних ланцюгах (рис. 10.18, 10.19). Потім відбувається реакція трансестерифікації із розрізанням у іншому ланцюзі й ковалентним замиканням двох кінців у шпильку – так само, як при рекомбінації імунoglobulinових генів (див. рис. 10.15) – і вилученням транспозона. Власне, білки RAG є результатом своєрідної "доместикації" транспозази з метою пристосувати її до виконання специфічних корисних для клітини функцій.

Вирізаний транспозон зі шпильками на кінцях у складі комплексу з димером транспозази (рис. 10.19) переноситься до мішені. Сайт-мішень – невелика специфічна послідовність ДНК, яка теж упізнається транспозазою, і в її межах теж індукуються одноланцюгові розрізи (рис. 10.18). Далі транспозаза каталізує розмикання шпильок і ковалентне приєднання 3'-кінців транспозона до цих виростів, прогалини заповнюються шляхом репаративного синтезу.



**Рис. 10.18.** Типова схема подій при переміщенні ДНК-транспозона: створення одноланцюгових розривів і замикання кінців у шпильки; розкривання шпильок і здійснення розривів у мішені при зустрічі з нею; вбудовування транспозона в мішень; репаративний синтез ДНК із заповненням прогалин



**Рис. 10.19.** Переміщення ДНК-транспозона за допомогою транспозази.

На вставці: структура комплексу димеру транспозази зі шпильками на кінцях вирізаного транспозона (6X68)

Отже, транспозиція є, по суті, одним із варіантів сайт-специфічної рекомбінації. У результаті цього процесу залишається дволанцюговий розріз у місці, де був транспозон. Іноді транспозаза залишає тупі кінці (розрізи у двох полінуклеотидних ланцюгах розташовані точно один проти одного), іноді – невеличкі одностанцюгові вирости. У розрізі, що залишився у донорному сайті, може бути різна "доля". Якщо залишились взаємно комплементарні вирости, то може відбутись їхнє спарювання і зашивання розрізів лігазою. Проте зазвичай після того, як транспозон "стрибає" в інше місце, розріз піддається репарації за механізмом негомологічного з'єднання кінців. Нарешті, є третя можливість: доволі часто транспозиція відбувається під час реплікації – тоді розріз репарується за рекомбінаційним механізмом (див рис. 10.1): сестринська молекула ДНК використовується як матриця, і ділянка, що містила транспозон, відновлюється – транспозон "розмножується".

### **LTR-ретротранспозони**

Кодувальна ділянка LTR-ретротранспозона (5–8 тис. пар основ) містить кілька генів: *зворотної транскриптази* (РНК-залежна ДНК-полімераза), *інтегрази* (гомолог транспозази, не має нічого спільного з інтегразою бактеріофага  $\lambda$ ), РНКазы Н. На кінцях розташовані довгі кінцеві повтори (200–600 пар основ), від яких і пішла назва цих мобільних елементів.

Першим етапом переміщення мобільного елемента є транскрипція ретротранспозона – синтез мРНК РНК-полімеразою II (рис. 10.20). Ця мРНК транспортується до цитоплазми, де піддається трансляції. Зворотна транскриптаза, яка є продуктом трансляції, у цитоплазмі здійснює синтез ДНК: мРНК використовується як матриця, 3'-кінець молекули тРНК – як праймер. Комплекс синтезованої ДНК-копії ретротранспозона з інтегразою повертається до ядра, де відбувається інсерція цієї ДНК у геном.

Процес переміщення LTR-ретротранспозона нагадує життєвий цикл ретровірусів: різниця полягає в тому, що після проникнення ретровірусу у клітину одразу відбувається зворотна транскрипція і вбудовування ДНК-копії у геном клітини-хазяїна за допомогою ретровірусної інтегрази, а у ретровірусній РНК містяться також гени білків вірусної оболонки – після транскрипції у цитоплазмі відбуваються їхня трансляція і збирання вірусних частинок. Отже, походження LTR-ретротранспозонів від ретровірусів є очевидним. На рис. 10.21 зображено структуру комплексу ретровірусної інтегрази з ретровірусною ДНК і ДНК-мішенню – аналогічний вигляд повинна мати інтеграза LTR-ретротранспозонів.

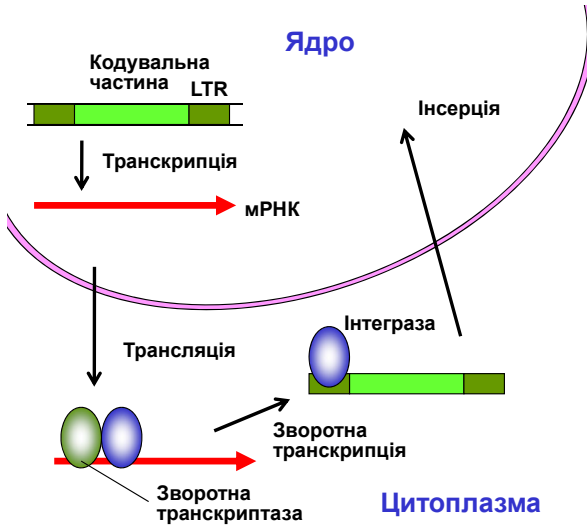


Рис. 10.20. Механізм переміщення LTR-ретротранспозона

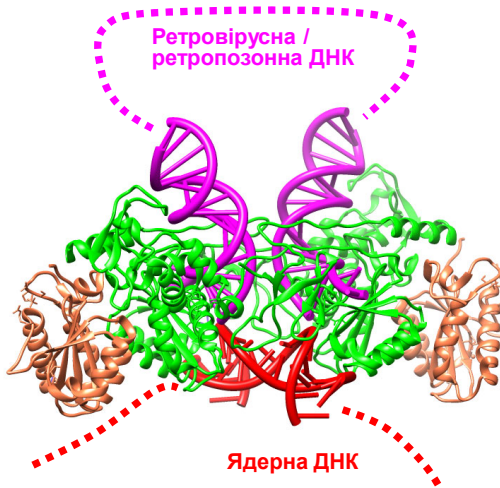


Рис. 10.21. Гомолог ретротранспозонної інтегрази – інтеграза ретровірусу імунодефіциту людини (ВІЛ) у комплексі з ретровірусною ДНК і ДНК-мішенню (3OS1)

## Мобільні елементи LINE та SINE

Мобільні елементи LINE завдовжки приблизно 6–7 тис. пар основ містять принаймні два гени – зворотної транскриптази й ендонуклеази – і невеликі специфічні послідовності на кінцях. Процес переміщення копії елемента починається так само, як і у випадку LTR-ретротранспозонів – із транскрипції РНК-полімеразою II (рис. 10.22).

Після трансляції в цитоплазмі синтезовані білки зв'язуються із мРНК, цей комплекс повертається до ядра, де й відбувається зворотна транскрипція: ендонуклеаза робить одноланцюговий розріз у геномній ДНК, 3'-кінець у місці розрізу використовується як праймер. Після синтезу другого ланцюга ДНК нова копія елемента остаточно вбудовується у геном.

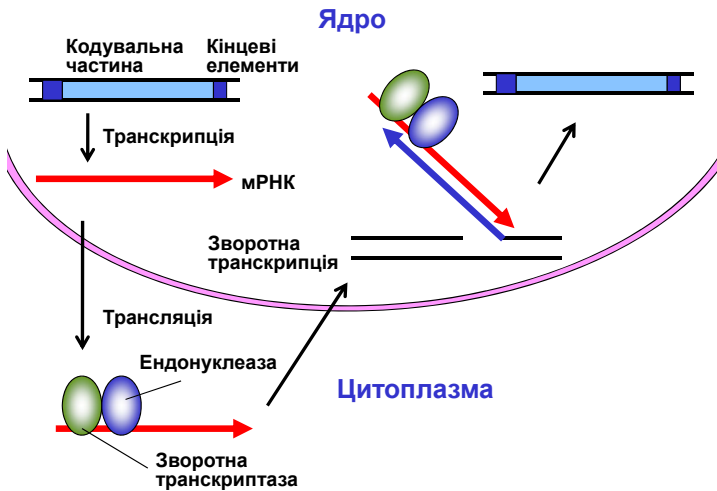


Рис. 10.22. Механізм переміщення мобільного елемента LINE

Синтез мРНК при транскрипції елемента LINE, як і для більшості інших еукаріотичних мРНК, закінчується на polyA-сигналі (див. розділ 7). Цей polyA-сигнал є слабким, що дозволяє елементу вбудовуватись у інтрони звичайних генів без особливих перешкод для їхньої експресії: система процесингу часто не помічає слабкий polyA-сигнал. Елементи LINE є найпоширенішими

мобільними елементами в геномі вищих еукаріотів (~20 % геному людини представлено послідовностями LINE) і основним джерелом зворотної транскриптази у клітинах. Із часом накопичуються мутації, окремі елементи LINE стають неактивними (тобто вже не кодують зворотну транскриптазу), але переміщуються, використовуючи білки, що кодуються активними елементами. Найчастіше в геномі людини зустрічається елемент типу LINE, який позначають L1 (кілька сотень тисяч копій).

Ноді мобільні елементи є не просто ділянками "егоїстичних ДНК", що автономно розмножуються в геномі, а виконують певні конкретні біологічні функції. Наприклад, у дрозофіли відсутня теломеразна система й елемент LINE певного типу виконує функцію подовження кінців хромосом після реплікації: зворотна транскриптаза діє як теломераза, мРНК мобільного елемента – як теломеразна матрична РНК (див. рис. 9.25). Власне, теломераза ссавців є гомологом зворотної транскриптази елементів LINE: саме від них походить теломеразна система.

Мобільні елементи SINE – короткі (100–300 пар основ) елементи послідовності, які не кодують білків і містять лише промотори для РНК-полімерази III. Процес їхнього переміщення є таким самим, як зображено на рис. 10.22: елементи SINE "паразитує" на системі LINE, використовуючи її ферменти. На відміну від усіх інших мобільних елементів, елементи SINE сконцентровані в ділянках геному, збагачених на гени.

## Контрольні запитання

1. Як починається гомологічна рекомбінація? Як здійснюється точна репарація дволанцюгових розривів за рекомбінаційним механізмом?

2. Що таке гетеродуплекс? Яку функцію виконує білок гесА при гомологічній рекомбінації?

3. Що таке структура Холідея? У якій формі вона реалізується під час гомологічної рекомбінації? За рахунок чого стабілізується така форма?

4. За яким механізмом відбувається міграція гілки під час гомологічної рекомбінації?

5. Як здійснюється розділення структури Холідея під час гомологічної рекомбінації? З якою імовірністю відбувається обмін ділянками між двома гомологічними молекулами ДНК і від чого залежить реалізація такого обміну?

6. Що відбувається з гетеродуплексом після завершення гомологічної рекомбінації?

7. Чим відрізняється сайт-специфічна рекомбінація від гомологічної?

8. Назвіть варіанти сайт-специфічної рекомбінації у бактеріофагів і наведіть відповідні приклади.

9. Опишіть молекулярний механізм сайт-специфічної рекомбінації у бактеріофагів. До якої родини білків належать фагові рекомбінази?

10. Що таке незаконна рекомбінація?

11. Як побудовані кластери імуноглобулінових генів? Які події при дозріванні імуноглобулінового гена можна віднести до сайт-специфічної рекомбінації?

12. Якими механізмами забезпечується необмежене зростання розмаїття послідовностей імуноглобулінових генів при їхньому дозріванні?

13. Порівняйте процеси сайт-специфічної рекомбінації у бактеріофагів, при дозріванні імуноглобулінових генів і переміщенні ДНК-транспозонів.

14. Опишіть молекулярні механізми переміщення ДНК-транспозонів.

15. Яка різниця між механізмами переміщення LTR-ретро-транспозонів і мобільних елементів типу LINE?

16. Чим візняються мобільні елементи типів LINE і SINE?

## **Рекомендована література**

Amunugama, R. and Fishel, R. (2012) Homologous recombination in eukaryotes. *Prog. Mol. Biol. Transl. Sci.* 110, 155–206.

Boeke, J.D. and Chapman, K.B. (1991) Retrotransposition mechanisms. *Curr. Opin. Cell. Biol.* 3, 502–507.

Chen, Q., Luo, W., Veach, R.A., et al. (2020) Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nat. Commun.* 11, 3446.

Chen, X., Cui, Y., Best, R.B., et al. (2020) Cutting antiparallel DNA strands in a single active site. *Nat. Struct. Mol. Biol.* 27, 119–126.

Drost, H.G. and Sanchez, D.H. (2019) Becoming a selfish clan: recombination associated to reverse-transcription in LTR retrotransposons. *Genome Biol. Evol.* 11, 3382–3392.

Hickman, A.B. and Dyda, F. (2016) DNA transposition at work. *Chem. Rev.* 116, 12758–12784.

Kaniecki, K., De Tullio, L. and Greene, E.C. (2018) A change of view: homologous recombination at single-molecule resolution. *Nat. Rev. Genet.* 19, 191–207.

Kazazian, H.H. and Moran, J.V. (1998) The impact of L1 retrotransposons on the human genome. *Nature Gen.* 19, 19–24.

Kuzimov, A. (1996) Unraveling the late stages of recombinational repair: metabolism of DNA junctions in *E. coli*. *BioEssays* 18, 757–765.

Lanciano, S. and Mirouze, M. (2018) Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Curr. Opin. Genet. Dev.* 49, 106–114.

Landy A. (2015) The  $\lambda$  integrase site-specific recombination pathway. *Microbiol. Spectr.* 3, MDNA3-0051-2014.

Laxmikanthan, G., Xu, C., Brilot, A.F., et al. (2016) Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *Elife* 5, e14313.

Platt, R.N., Vandewege, M.W. and Ray, D.A. (2018) Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* 26, 25–43.

Roth, DB. (2014) V(D)J Recombination: Mechanism, errors, and fidelity. *Microbiol. Spectr.* 2, MDNA3-0041-2014.

Ru, H., Zhang, P. and Wu, H. (2018) Structural gymnastics of RAG-mediated DNA cleavage in V(D)J recombination. *Curr. Opin. Struct. Biol.* 53, 178–186.

San Filippo, J., Sung, P. and Klein, H. (2008) Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* 77, 229–257.

San-Segundo, P.A. and Clemente-Blanco, A. (2020) Resolvases, dissolvases, and helicases in homologous recombination: clearing the road for chromosome segregation. *Genes* 11, 71.

Sun, Y., McCorvie, T.J., Yates, L.A. and Zhang, X. (2020) Structural basis of homologous recombination. *Cell. Mol. Life Sci.* 77, 3–18.

Van Duyne, G.D. (2015) Cre recombinase. *Microbiol. Spectr.* 3, MDNA3-0014-2014.

West, S.F. (1992) Enzymes and molecular mechanisms of genetic recombinations. *Annu. Rev. Biochem.* 61, 603–640.

West S. (1997) Processing of recombination intermediates by the Ruv ABC proteins. *Annu. Rev. Genet.* 31, 213–244.

Wright, W.D., Shah, S.S. and Heyer, W.D. (2018) Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10524–10535.

## Розділ 11

# МЕТОДИ МОЛЕКУЛЯРНОЇ БІОЛОГІЇ

There is no method but to be very intelligent.

*T. S. Eliot*

Розвиток молекулярної біології – це насамперед розвиток різноманітних методичних підходів, що дозволяють з'ясувати молекулярні механізми біологічних процесів. Цей розділ присвячено огляду основних сучасних методів молекулярної біології – лише огляду і тільки основних, оскільки детальний опис усіх методик потребує багатьох томів.

Методичні підходи, що використовуються у молекулярній біології, дозволяють аналізувати суміші макромолекул, отримувати біологічні макромолекули – природні або змінені за бажанням дослідника – у достатніх кількостях, визначати їхні структури (від нуклеотидних і амінокислотних послідовностей до просторової структури), досліджувати механізми взаємодій між ними, аналізувати як властивості окремих макромолекул, так і проводити системний аналіз їхньої великої сукупності (геномів, транскриптомів, протеомів тощо), конструювати нові макромолекули й геномні ділянки як *in vitro*, так і в живих клітинах. Сучасні методи молекулярної біології є не тільки дослідницькими інструментами, а й основою для розвитку молекулярних біотехнологій і молекулярної медицини.

## Розділення і аналіз сумішей нуклеїнових кислот

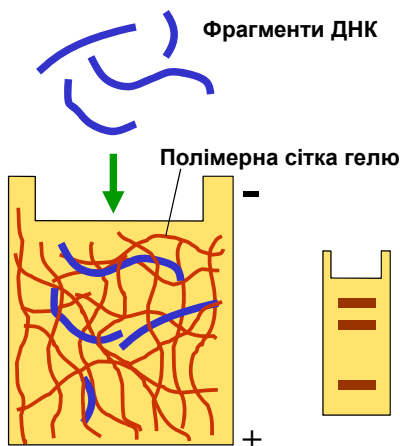
### Гель-електрофорез

Для розділення складних сумішей різноманітних фрагментів ДНК різної довжини найчастіше застосовують гель-електрофорез – надзвичайно важливий метод у молекулярній біології.

Принцип електрофорезу дуже простий (рис. 11.1). Суміш фрагментів ДНК (або РНК) наносять у лунку пластинки гелю – середовища, сформованого полімерною сіткою. Зазвичай гель виготовляють на основі агарози або поліакриламідну. Через нього пропускають електричний струм, завдяки чому негативно заряджені молекули нуклеїнової кислоти рухаються в електричному полі. Полімерна сітка гелю створює для цього руху суттєвий опір, якій долається тим легше, чим меншим є розмір фрагмента. Через певний час фрагменти розділяються на окремі зони – смуги гелю, які містять гомогенний набір фрагментів одного розміру. Для візуалізації смуг гель забарвлюють флуоресцентним барвником (найчастіше використовують бромистий етидій), який зв'язується із ДНК. Після цього смугу можна вирізати з гелю та екстрагувати ДНК.

Гель-електрофорез широко використовують також як аналітичний засіб із метою з'ясування складу суміші. Для візуалізації смуг (особливо якщо розділюють невеликі кількості ДНК) часто застосовують радіоактивне (шляхом уведення радіоактивного ізотопу  $^{32}\text{P}$  у склад 5'-кінцевих фосфатних залишків) або флуоресцентне мічення ДНК. У випадку радіоактивного мічення після електрофорезу здійснюється *авторадіографія*: на гель накладають фотопластинку, яка засвічується напроти смуг унаслідок їхньої радіоактивності. Альтернативою авторадіографії є експонування гелю на спеціальний фосфорний екран, який потім сканують сполученим із комп'ютером приладом – *фосфоріміджером*. Крім вищої, якщо порівнювати з авторадіографією, чутливості, використання фосфоріміджера має ту перевагу, що дозволяє зі значно вищою точністю оцінювати кількість матеріалу в смугі.

У найпростішому варіанті для розділення фрагментів дволанцюгової ДНК за розміром найчастіше використовують агарозні гелі. Проте поліакриламідний гель (ПААГ) має підвищену роздільну здатність і значно різноманітніше застосування. Зокрема, саме ПААГ застосовують для розділення сумішей білків (див. відповідний підрозділ нижче).



**Рис. 11.1.** Розділення фрагментів ДНК за розміром шляхом гелі-електрофорезу.

Праворуч – окремі зони гелю, що містять розділені фрагменти

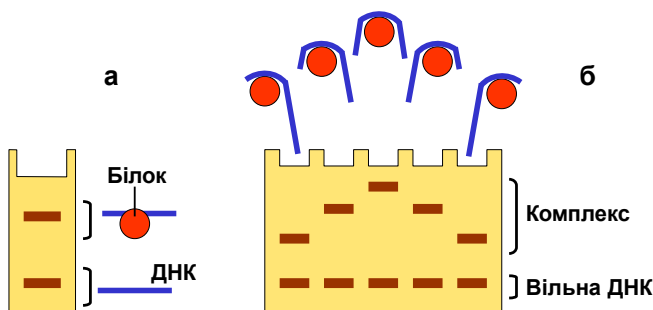
На рухливість макромолекул під час електрофорезу впливає також їхня форма. За допомогою електрофорезу можна, зокрема, розділити:

- молекули ДНК, що містять перманентні вигини, зумовлені особливою послідовністю пар основ – вигнута ДНК більшою мірою гальмується полімерною сіткою гелю;
- циркулярні й лінійні молекули ДНК, а також окремі циркулярні топоізмери (див. розділ 3), оскільки вони розрізняються за ступенем компактності;
- молекулу вільної ДНК і комплекс цієї ДНК із білком – зрозуміло, що комплекс більше гальмується сіткою гелю.

Останній підхід (*шифт-тест*, shift assay) ілюструє рис. 11.2: ДНК-білковий комплекс (якщо він утворюється) суттєво гальмується в гелі порівняно з вільною ДНК, унаслідок чого після електрофорезу зазвичай виникають дві смуги (рис. 11.2, *a*). При цьому якщо білок, як це часто буває, вигинає ДНК у сайті взаємодії, то рухливість комплексу буде залежати від позиції білка на фрагменті ДНК: максимальну рухливість мають комплекси, у складі яких білок розташований на одному з кінців фрагмента,

мінімальну – ті, де білок займає центральну позицію (рис. 11.2, б). Причиною такої закономірності є те, що довші (і тому рухливіші) кінцеві ділянки вільної ДНК легше проходять крізь пори гелю. Це дозволяє оцінити локалізацію білка відносно кінців фрагмента ДНК або розділити комплекси, що розрізняються за такою локалізацією.

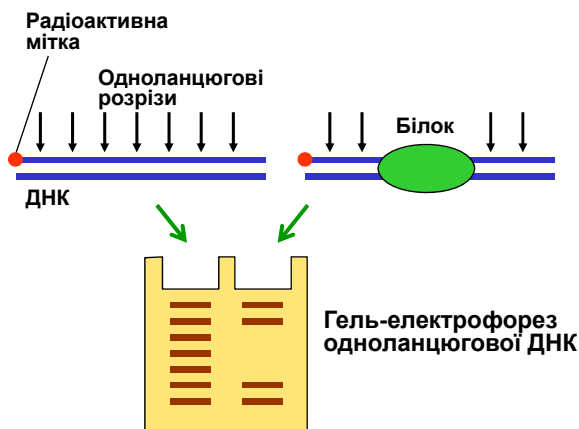
Особливим (і дуже важливим) варіантом електрофорезу нуклеїнових кислот є електрофорез одноланцюгових фрагментів ДНК в умовах, що забезпечують денатурацію (підвищена температура, висока концентрація сечовини). Якщо гель (поліакриламідний) доволі тонкий (~0,4 мм), то він дозволяє розділити фрагменти, довжина яких різниться на один нуклеотид. Такий гель називають *секвенувальним*, оскільки саме він застосовувався у перших методах секвенування – установлення нуклеотидних послідовностей (див. нижче). Широковживаним альтернативним підходом є електрофорез не у пластинках гелю, а в тонких силіконових капілярах, зазвичай заповнених поліакриламідом. *Капілярний електрофорез* відрізняється, за високої роздільної здатності, значно підвищеною швидкістю.



**Рис. 11.2.** Гальмування руху ДНК-білкового комплексу при гель-електрофорезі (а) і залежність ступеня гальмування від позиції білка на ДНК у випадку вигинання ДНК на поверхні білка (б)

Іншим прикладом використання електрофорезу в умовах денатурації є метод точної локалізації білка на фрагменті ДНК – *футпринтинг* ДНК (DNA footprinting). Принцип методу схематично зображено на рис. 11.3. Кінець одного з ланцюгів мітять

радіоактивною міткою. Якщо внести в цю ДНК обмежену кількість одноланцюгових розрізів (напр. один розріз на фрагмент у середньому), то отримаємо випадковий набір мічених фрагментів різної довжини, які можна розділити за допомогою гелелектрофорезу. Як агенти, що вносять розриви, найчастіше використовують ДНКазу I або гідроксил-радикали: обидва агенти розрізають один із ланцюгів ДНК через маленький жолобок. У випадку, коли із фрагментом ДНК взаємодіє білок, ДНК виявляється захищеною в сайті взаємодії: на гелі після електрофорезу можна побачити зону, де відсутні смуги ДНК або їхня інтенсивність нижча порівняно з контрольним результатом деградації вільної ДНК. Довжини верхньої і нижньої смуг, які оточують цю зону, є границями сайту взаємодії.



**Рис. 11.3.** Футпринтинг ДНК у складі білково-нуклеїнового комплексу

### Блот-гібридизація

Потужним засобом аналізу складних сумішей ДНК щодо наявності там специфічних елементів послідовності є *саузерн-блотинг* – блот-гібридизація на нітроцелюлозних фільтрах за Саузерном (Edward Southern). Назва процедури, яку схематично зображено на рис. 11.4, походить від слова blotting (промаку-

вання): фрагменти ДНК розділюються за допомогою гелелектрофорезу (залишаючись невидимими в гелі), після чого на гель накладають нітроцелюлозний фільтр, а під і над цим "сэндвічем" розміщують фільтрувальний папір і занурюють нижній шар паперу в лужний розчин. Під дією капілярних сил розчин піднімається до верхнього шару паперу, захоплюючи при цьому ДНК і переносячи її з гелю на нітроцелюлозу, де вона необоротно фіксується. Одночасно ДНК денатурується лугом. У результаті одноланцюгова ДНК опиняється на фільтрі, який є точною реплікою вихідного гелю. Так само можна перенести ДНК на нітроцелюлозу, застосувавши електрофорез у поперечному напрямку – із гелю на фільтр. Далі проводять оброблення фільтра зондом – одноланцюговим фрагментом ДНК певної послідовності, який містить радіоактивну мітку. Зонд гібридується з комплементарною ДНК у певних досі невидимих смугах, що можна зафіксувати за допомогою авторадіографії.

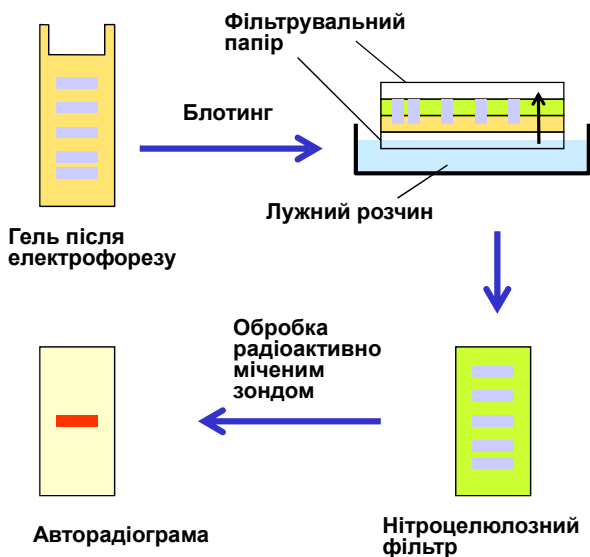


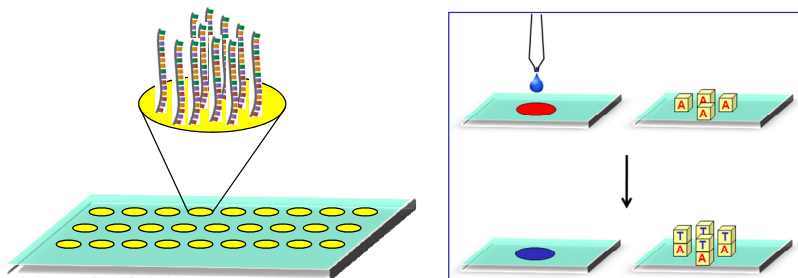
Рис. 11.4. Блот-гібридизація

У такий спосіб можна встановити, наприклад, наявність у складних сумішах нуклеїнових кислот специфічних послідовностей ДНК. Одне із застосувань блотингу за Саузерном полягає у дослідженні ступеня захищеності ДНК білками в конкретних геномних зонах живої клітини. Першим етапом процедури, що називається *непрямим кінцевим міченням*, є обмежена деградація ДНК нуклеазою у клітинних ядрах – чутливість до нуклеази залежить від ступеня доступності ДНК. Після цього виділяють сумарний препарат фрагментів ДНК, розділюють шляхом електрофорезу, здійснюють блотинг і обробляють нітроцелюлозний фільтр радіоактивно міченим зондом до певної ділянки геному. На авторадіограмі візуалізуються (або ні, якщо ДНК була доступною і піддалась деградації) фрагменти ДНК із цієї зони.

Варіацією саузерн-блотингу є *нозерн-блотинг*, який відрізняється лише тим, що на гель для електрофорезу наносять препарат виділеної РНК. Гібридизація з міченим зондом на нітроцелюлозному фільтрі дозволяє ідентифікувати певну РНК у препараті. Назва методу (nothern) є жартівливою аналогією з буквальним значенням прізвища Саузерна.

## **ДНК-чіпи**

Проаналізувати одночасно велику кількість фрагментів ДНК або РНК (у т. ч. тотальний препарат РНК, що продукується клітинами певного типу) дозволяють методи, що базуються на використанні *ДНК-чіпів* (DNA-microchips), відомих також як *ДНК-мікроареї* (DNA-microarrays). Чіп є силіконовою поверхнею, до якої на мікроділянці пришиті однакові одноланцюгові фрагменти ДНК завдовжки до 1 тис. пар основ. Сітка чіпа містить кілька тисяч таких мікроплям, кожна з яких відповідає певній відомій геномній ділянці. Більш популярними є чіпи, на поверхні яких синтезуються короткі (до 20 нуклеотидів) олігонуклеотиди: велика кількість однакових олігонуклеотидів, що походять з однієї ділянки геному, синтезуються поряд, отже, невелика зона містить ДНК, комплементарну індивідуальній геномній ділянці (рис. 11.5).



**Рис. 11.5.** ДНК-чіп і схема його виготовлення під дією мікроструменя розчину нуклеотидів певного типу на поверхню

*Черепичні арей (tiling arrays)* – це ДНК-чіпи, на поверхні яких розташовано кілька мільйонів проб, кожна з яких складається з олігонуклеотидів одного типу. Послідовності олігонуклеотидів є невеличкими фрагментами геному, які перекриваються (усі проби покривають повну послідовність геному або його зони із кроком, наприклад, у чотири нуклеотиди).

Використання ДНК-чіпів базується на гібридизації на них сумішей одноланцюгових флуоресцентно мічених нуклеїнових кислот. Факт гібридизації легко встановити за флуоресценцією мікроплями, причому інтенсивність флуоресценції свідчить про рівень наявності нуклеїнової кислоти певної нуклеотидної послідовності у вихідній суміші. Одну з типових схем зображено на рис. 11.6. Сумарна мРНК, отримана із клітин певного типу, використовується як матриця в реакції зворотної транскрипції. Поряд зі звичайними, до реакційної суміші додається флуоресцентний аналог одного з нуклеозидтрифосфатів. У результаті отримують препарат флуоресцентно мічених кДНК (комплементарних ДНК-копій молекул мРНК). Після гібридизації із кДНК чіп аналізують за допомогою флуоресцентного мікроскопа: наявність флуоресцентної плями свідчить про активність певного гена, інтенсивність флуоресценції – про рівень цієї активності.

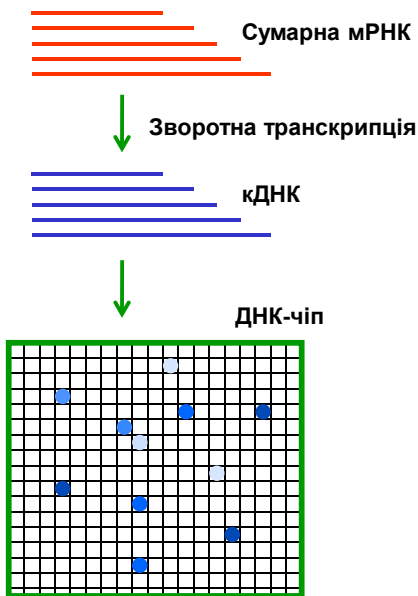


Рис. 11.6. Аналіз сумарної мРНК за допомогою ДНК-чипа

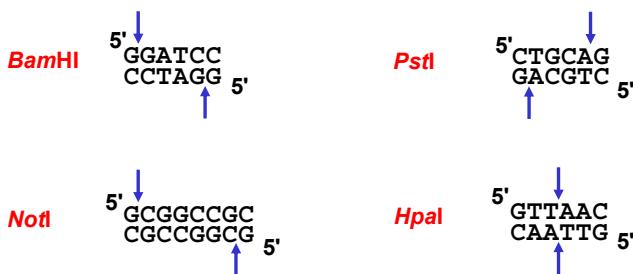
## Клонування ДНК

Перша проблема, що зазвичай постає перед дослідником, – отримати достатню кількість об'єкта дослідження. Ефективним підходом, що дозволяє "розмножити" будь-який конкретний фрагмент ДНК, є техніка клонування ДНК у бактеріальних клітинах: потрібний фрагмент вбудовують у ДНК-вектор, отриману в результаті *рекомбінантну молекулу* ДНК вводять у клітину (*трансформація*). Далі залишається зачекати зростання бактеріальної культури – бактерію використовують як біореактор.

Основними інструментами для створення рекомбінантних ДНК (*рекомбінантних технологій, генетичної інженерії*) служать природні ферменти, що каталізують реакції деградації або синтезу нуклеїнових кислот. Зокрема, важлива роль належить ендонуклеазам рестрикції – *рестриктазам* – ферментам, що здійснюють специфічне впізнання невеликих елементів нуклео-

тидної послідовності й розрізання ДНК. Назви рестриктаз (їх існує кілька тисяч, кілька сотень доступні як комерційні препарати) утворюють за таким принципом: перша велика літера позначає рід мікроорганізму, дві маленькі – вид, римські цифри, а іноді великі літери, – порядковий номер рестриктази серед інших рестриктаз досліджуваної бактерії. Наприклад, *EcoRI* – рестриктаза RI із *Escherichia coli*. Рестриктази розділяють на три класи. До першого належать ферменти, що впізнають певну специфічну нуклеотидну послідовність і роблять розріз молекули ДНК на невизначеній відстані від неї. Рестриктази другого типу (саме вони використовуються найширше) здійснюють розріз усередині сайту, який специфічно впізнають. До третього класу належать рестриктази, що розрізають ДНК поза межами сайту впізнання, але на чітко визначеній відстані від нього.

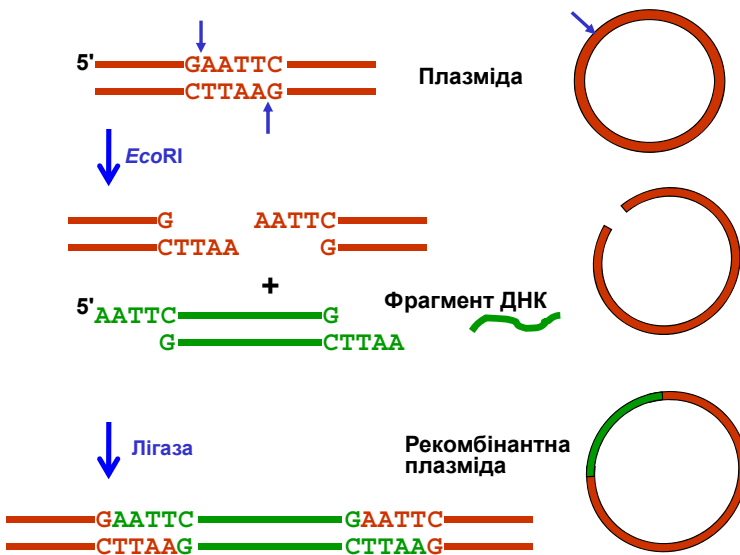
Сайтами рестрикції, що упізнаються рестриктазами другого класу, є невеликі (чотири, шість, іноді трохи більше пар основ) паліндромні послідовності. Два розрізи, що здійснюються певною рестриктазою, можуть бути розташовані точно один напроти одного у двох ланцюгах, що спричиняє утворення тупих (blunt) кінців. Частіше рестриктази залишають взаємно комплементарні 5'-кінцеві (іноді 3'-кінцеві) одностанцюгові вирости – липкі (sticky) взаємно комплементарні кінці (рис. 11.7).



**Рис. 11.7.** Приклади рестриктних сайтів (стрілками позначено місця розрізів).

Рестриктази *BamHI* та *NotI* залишають липкі кінці із 5'-кінцевими, *PstI* – із 3'-кінцевими виступами; *HpaI* – тупі кінці

Інший важливий ензиматичний інструмент – *ДНК-лігаза* (див. розділ 9), що каталізує утворення фосфодієфірного зв'язку. На рис. 11.8 наведено найпростіший приклад використання рестриктази й лігази для створення рекомбінантної молекули ДНК. У цьому прикладі деякий фрагмент ДНК отримано за допомогою рестриктази *EcoRI*, унаслідок чого він має липкі кінці. Деяка бактеріальна плазміда (циркулярна молекула ДНК, здатна до автономної реплікації у клітині) містить тільки один сайт рестрикції *EcoRI* – оброблення плазміди цим ферментом розрізає її з утворенням таких самих липких кінців. При об'єднанні двох препаратів ДНК відбувається гібридизація комплементарних липких кінців, а зашивання одноланцюгових розривів лігазою створює рекомбінантну молекулу ДНК. Власне, плазміда й використовується як вектор-носії вбудованого в неї фрагмента.



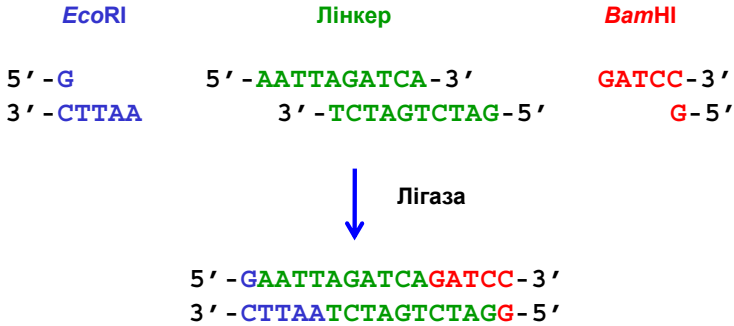
**Рис. 11.8.** Виготовлення рекомбінантної ДНК: розрізання (рестрикція) циркулярної плазміди рестриктазою *EcoRI*, додавання фрагмента ДНК, отриманого за допомогою тієї самої рестриктази, вбудовування фрагмента у плазміду за допомогою лігази

Фрагменти ДНК із тупими кінцями також можна вбудувати у вектор за допомогою лігази, хоча така реакція на порядок менш ефективна. За допомогою *термінальної нуклеотидилтрансферази* (фермент, який без участі матриці приєднує нуклеотиди до 3'-кінця) до 3'-кінців фрагмента ДНК можна приєднати, наприклад, одностанцюговий poly(dA)-хвіст, а до 3'-кінців лінійного вектора – poly(dT). При змішуванні між комплементарними одностанцюговими кінцями відбувається спарювання, остаточне зшивання завершує ДНК-лігаза.

Можна також утворити тупі кінці на фрагментах із липкими кінцями. Для цього або видаляють одностанцюгові вирости за допомогою нуклеази S1 (нуклеаза, що гідролізує тільки одностанцюгову ДНК), або липкі кінці забудовують за допомогою фрагмента Кленова ДНК-полімерази I *E. coli* (див. розділ 9). Далі до утвореного фрагмента з тупими кінцями додають вирости poly(dA) і вбудовують його у вектор.

У випадку, коли дві молекули ДНК, які потрібно об'єднати, не мають взаємно комплементарних липких кінців (напр., були отримані за допомогою різних рестриктаз), може бути застосований короткий двостанцюговий лінкер, що має відповідні вирости на кінцях, комплементарні виростам на кінцях фрагментів (рис. 11.9). Такий лінкер створюється шляхом гібридизації двох олігонуклеотидів, які, своєю чергою, синтезуються хімічним шляхом. *Твердофазний синтез* олігонуклеотидів відбувається за допомогою автоматичних синтезаторів – можна запрограмувати будь-яку довжину й нуклеотидну послідовність. Синтетичні олігонуклеотиди широко застосовують і при реалізації багатьох інших методик, що описані нижче.

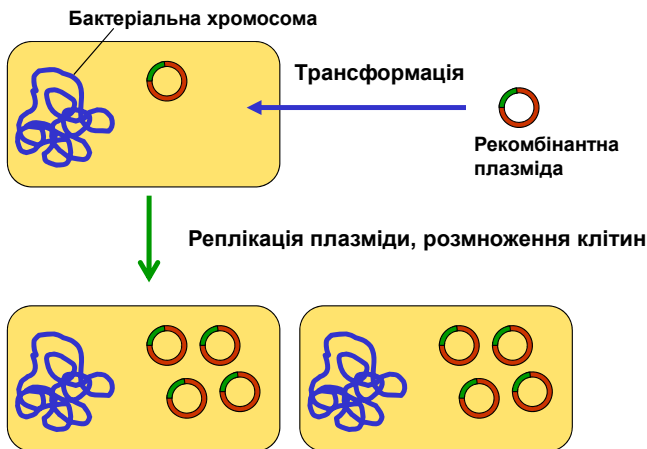
Основними вимогами до плазміди як вектора є наявність у її складі ориджину реплікації, унікального (одного на плазміді) сайту, що упізнається певною рестриктазою, і гена стійкості до одного з антибіотиків. Найзручнішими є плазмідні вектори, які містять *полілінкер* – ділянку з певним набором унікальних рестриктних сайтів, що дозволяє підібрати одну з рестриктаз, найпридатнішу для конкретного випадку.



**Рис. 11.9.** Зшивання двох фрагментів ДНК лігазою за допомогою лінкера

Після створення рекомбінантної плазміди її використовують для *трансформації* бактеріальних клітин. Трансформацію здійснюють зазвичай у розчині  $\text{CaCl}_2$  або шляхом електропорації (через суспензію клітин проводять короткий імпульс електричного струму). В обох випадках підвищується проникність клітинної стінки – і рекомбінантна плазміда потрапляє усередину. Зрозуміло, що далеко не всі бактерії отримують плазміду при трансформації. І тут стає у нагоді ген стійкості до антибіотика: достатньо обробити ним бактеріальну культуру, щоб залишити тільки трансформовані клітини.

Після трансформації клітини висівають у чашки Петрі, розтираючи маленьку порцію суспензії по поверхні твердого поживного середовища, що містить потрібний антибіотик. Далі відбуваються автономна реплікація плазміди й розмноження самих клітин, що зумовлює значне зростання загальної кількості плазмід (рис. 11.10). Розмноження клітин породжує появу колоній, кожна з яких веде походження від однієї клітини, тобто є її клоном.



**Рис. 11.10.** Розмноження рекомбінантної плазміди в бактеріальних клітинах

Отримані рекомбінантні клони можна зберігати й підтримувати нескінченно довго. За потреби клоновані плазміди виділяють із бактеріальної культури. Метод виділення плазмідної ДНК (лужна екстракція) базується на тому, що плазміда є порівняно невеликим замкненим кільцем: при денатурації у лузі, хоча подвійна спіраль руйнується, два ланцюги ДНК не здатні розійтися. При швидкій ренатурації довгі одноланцюгові фрагменти бактеріальної хромосоми безладно гібридизуються й утворюють нерозчинний клубок (осаджуються при центрифугуванні), тоді як плазміди легко відновлюють свій нативний стан (залишаються в надосадовій рідині).

Оброблення виділених плазмід тією самою рестриктазою, що була використана при виготовленні рекомбінантної молекули, дозволяє вирізати з вектора клонований фрагмент ДНК, який після цього можна відокремити за допомогою гелелектрофорезу.

Трансформація бактерій плазмідами тим ефективніша, чим меншою за розміром є плазміда. Відповідно існує обмеження розмірів фрагментів, які можна клонувати описаним шляхом, – до 10 тис. пар основ. Альтернативною, але цілком аналогічною

технікою, що дозволяє працювати із фрагментами завдовжки ~20 тис. пар основ, є клонування ДНК із використанням векторів на основі бактеріофага  $\lambda$  (див. розділ 6). За допомогою рестриктази й лігази фрагмент ДНК вбудовується у фагову ДНК замість ділянки, що містить гени, які відповідають за інтеграцію фагової ДНК у бактеріальний геном (гени лізису й білків оболонки при цьому працюють нормально). До рекомбінантної ДНК (важливо тільки, щоб її розмір – ~50 тис. пар основ – приблизно збігався із розміром ДНК бактеріофага) додаються порожні фагові оболонки, і відбувається збирання фагових частинок *in vitro*. Рекомбінантними бактеріофагами (які не здатні реалізувати шлях лізогенії) заражають бактеріальну культуру, де вони розмножуються.

Використовуючи вектори на основі *космід*, можна клонувати фрагменти ДНК завдовжки ~40 тис. пар основ. Косміда є плазмідною, яка, крім ориджину, сайтів рестрикції і генів стійкості до антибіотиків містить два *cos*-сайти: липкі кінці лінійної молекули ДНК бактеріофага  $\lambda$ . Саме за рахунок *cos*-сайтів лінійна молекула фагової ДНК циркуляризується після її проникнення в бактеріальну клітину. У лінійну космідку з такими липкими кінцями вбудовується фрагмент ДНК, який бажано клонувати. Рекомбінантна косміда (розміром близько 50 тис. пар основ) упаковується *in vitro* у фагові частинки, якими обробляють бактеріальну культуру. У цьому випадку бактеріофаг використовується як ефективний засіб трансформації: лінійна косміда проникає у клітину, де циркуляризується за рахунок *cos*-сайтів і бактеріальної лігази. Далі циркулярна косміда розмножується як звичайна плазмідка (див. схему на рис. 11.10).

Для клонування фрагментів ДНК від 100 тис. пар основ застосовують спеціально сконструйовані вектори ВАС (Bacterial Artificial Chromosome) і YAC (Yeast Artificial Chromosome). ВАС-вектори створено на основі F-плазмід бактерій, YAC-вектор є штучною дріжджовою мініхромосоною, яка містить центромеру, теломери й точку початку реплікації. У такий вектор можна ввести чужорідний фрагмент ДНК розміром понад 100 тис. пар основ. Така мініхромосома, уведена в дріжджову клітину, буде реплікуватися й поводитися аналогічно іншим дріжджовим хромосомам при мітотичному поділі.

## Бібліотеки клонів

Вище було розглянуто принципи клонування деякого фрагмента ДНК. Однак звідки береться цей фрагмент? Для того, щоб працювати з певною ділянкою ДНК (напр. певним геном), спочатку виконують клонування великої кількості різноманітних фрагментів геному – створюють бібліотеку клонів (*геномну бібліотеку*), серед яких шукають потрібний.

Стандартна процедура створення бібліотеки – *метод дробовика* (*shotgun*) – розпочинається із часткового (протягом обмеженого часу) оброблення всієї геномної ДНК певною рестриктазою у такий спосіб, щоб отримати набір різноманітних фрагментів певної середньої довжини, які перекриваються (рис. 11.11). Усі такі фрагменти вбудовують у вектори, здійснюють трансформацію і висівають трансформовані бактерії на тверде середовище. Набір колоній – це набір різноманітних клонів, кожен з яких містить один із фрагментів геному.

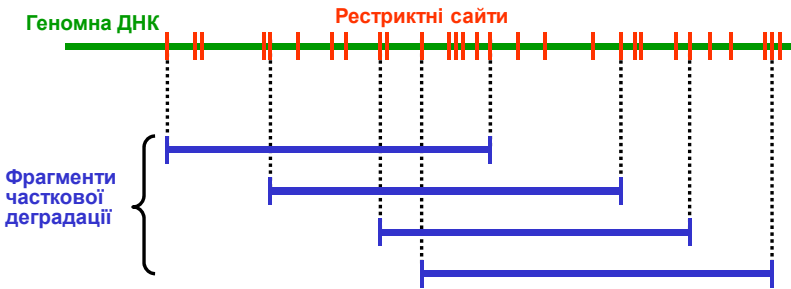
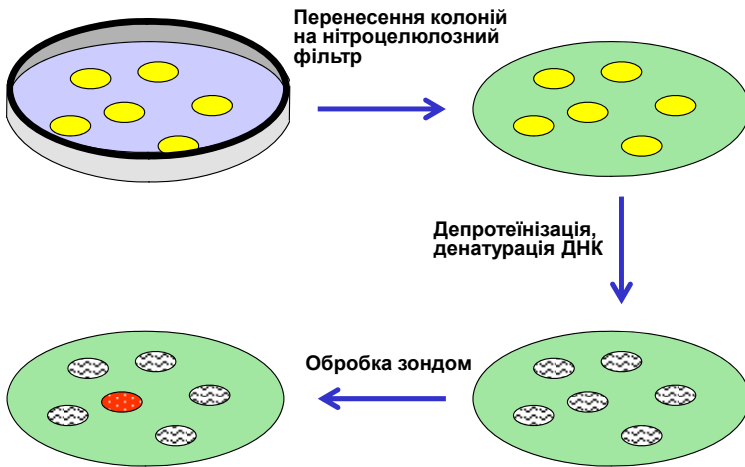


Рис. 11.11. Метод дробовика

Аналогічно створюють бібліотеки клонів кДНК. Для цього спочатку виділяють сумарну мРНК із клітин певного типу. Використовуючи зворотну транскриптазу й оліго(dT), що комплементарні 3'-кінцевим polyA-хвостам мРНК, як праймери, на цих мРНК синтезують комплементарні ланцюги ДНК. Видаляють ланцюг РНК РНКазою і добудовують другий ланцюг ДНК за допомогою фрагмента Кленова ДНК-полімерази I. Далі отримані молекули кДНК піддають клонуванню. На відміну від геном-

ної бібліотеки, бібліотека клонів кДНК містить тільки кодувальні послідовності (екзони) генів і тільки таких, які активні у клітинах даного типу.

Один зі способів знайти потрібну нуклеотидну послідовність серед бібліотеки клонів – гібридизація клонованої ДНК із радіоактивно міченим ДНК-зондом (рис. 11.12). Колонії бактерій у чашці Петрі частково переносять на нітроцелюлозний фільтр – роблять своєрідну репліку. Здійснюється лізис клітин, депротейнізація і денатурація ДНК у лужному розчині. Після висушування фільтра одноланцюгова ДНК необоротно фіксується на ньому. Фільтр обробляють радіоактивно міченою ДНК певної послідовності – зондом, який гібридується із тим клоном, до якого він комплементарний. Детектування гібридизації за допомогою авторадіографії дозволяє ідентифікувати розшукуваний клон.



**Рис. 11.12.** Гібридизація клонованої ДНК із радіоактивним зондом на нітроцелюлозному фільтрі

Зрозуміло, що описаний скринінг бібліотеки клонів залежить від наявності зонда. Одна з можливостей приготувати зонд базується на відомостях про амінокислотну послідовність білка

того гена, клон якого потрібно відшукати (або білка-гомologa іншого організму). Хімічний синтез короткого олігонуклеотиду (~20 нуклеотидів) із послідовністю, що відповідає фрагменту амінокислотної послідовності білка й тому має бути комплементарною ділянці ДНК клону, дозволяє отримати зонд. Після внесення радіоактивної мітки такий олігонуклеотид можна використати для гібридизації. Проте, зважаючи на виродженість генетичного коду, бажано синтезувати набір послідовностей, які відповідають усім можливим комбінаціям кодонів для даної ділянки послідовності амінокислот. Суміш таких альтернативних олігонуклеотидів, один із яких є очевидно комплементарним, і використовують як зонд.

Замість синтезу набору олігонуклеотидів можна обійтися синтезом лише одного зонда, використовуючи базу даних EST (Expressed Sequence Tag). База є загальнодоступною колекцією великої кількості коротких (200–400 пар основ) фрагментів послідовностей кДНК багатьох організмів. Щоб приготувати зонд, достатньо знайти в EST-базі коротку ділянку, яка відповідає фрагменту послідовності білка, і синтезувати її.

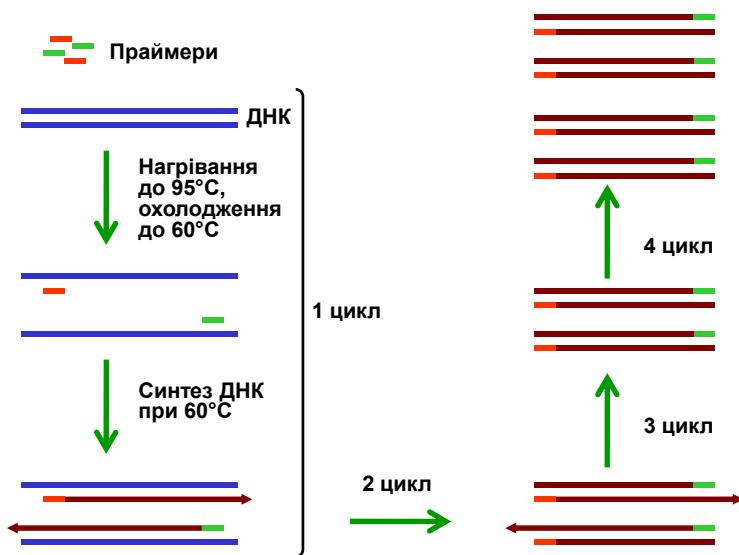
## Полімеразна ланцюгова реакція

Метод полімеразної ланцюгової реакції (ПЛР, або PCR – Polymerase Chain Reaction), що був розроблений як альтернатива для клонування, є сьогодні основним засобом багатократного збільшення кількості бажаного фрагмента ДНК – ампліфікації. ПЛР – це реакція синтезу ДНК *in vitro*, яка повторюється багато разів: синтезовані ланцюги стають матрицями для синтезу в наступному циклі реакції. Для здійснення ПЛР треба знати принаймні короткі послідовності ДНК на кінцях того фрагмента, що має бути ампліфікований.

Реакційна суміш містить слідову кількість вихідної ДНК, надлишкові кількості двох синтетичних праймерів до кінців фрагмента, нуклеозидтрифосфати чотирьох типів і ДНК-полімеразу *Taq* термофільної бактерії *Thermus aquaticus* (або її аналог іншого термофільного організму). Використання саме полімерази *Taq*

зумовлено тим, що вона здатна працювати за підвищеної температури й не піддається денатурації за дуже високих температур.

У першому циклі ПЛР (рис. 11.13) відбувається нагрівання суміші до  $\sim 95^\circ\text{C}$  (денатурація ДНК) і охолодження до  $\sim 60^\circ\text{C}$  (точні значення обох температур підбираються у кожному конкретному випадку, виходячи з нуклеотидної послідовності праймерів). Після охолодження до такої усе ще доволі високої температури ланцюги ДНК не ренатурують, але праймери (оскільки вони наявні у високих концентраціях) знаходять свої комплементарні ділянки та зв'язуються з ними. Полімераза *Taq* починає працювати, подовжуючи ці праймери.



**Рис. 11.13.** Полімеразна ланцюгова реакція.

Наведено приблизні значення температур.

Деякі матриці, довжина яких виходить за межі ділянки, обмеженої двома праймерами, не показано (кількість таких матриць після багатьох циклів залишається дуже малою)

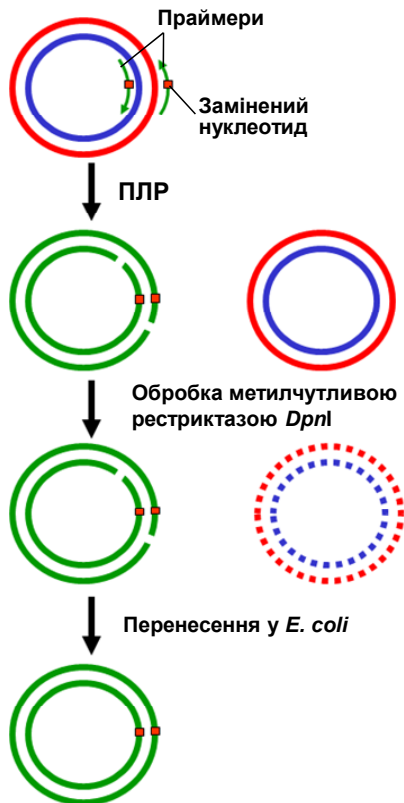
Усі наступні цикли ПЛР (реакція здійснюється в автоматичному режимі на приладі, що називається ампліфікатором) полягають у нагріванні суміші до  $95^\circ\text{C}$ , яке змінюється охолоджен-

ням до 60 °С. Після другого циклу будуть знов синтезовані дві такі самі молекули, як після першого, а також дві молекули, які зображено на рис. 11.13; після третього циклу – дві такі самі молекули, що й після першого, дві такі самі, як після другого, а також дві молекули, довжина яких чітко обмежена двома праймерами (рис. 11.13). Далі кількість таких коротких матриць буде зростати, і після 30-го циклу їх буде в мільйон разів більше, ніж вихідних довгих матриць.

Метод ПЛР має надзвичайно широке застосування. Він використовується кожного разу, коли є необхідність детектувати й дослідити невелику кількість ДНК, зокрема й у складі неочищеного біологічного матеріалу. ПЛР застосовують також у комбінації із клонуванням: ампліфікований ДНК-продукт можна клонувати й використати, наприклад, для синтезу білка (див. нижче); ампліфікація за допомогою ПЛР стає у нагоді для збільшення невеликої кількості клонованої ДНК.

Прикладом одного із численних застосувань ПЛР для аналізу ДНК є сучасний варіант *фінгерпринтингу ДНК* (DNA fingerprinting). Метод базується на наявності в еукаріотичних геномах мінісателітних повторів (див. розділ 4) – невеликих елементів послідовності, що тандемно повторюються в різних місцях геному кілька разів. Розподіл локусів за кількістю повторів індивідуальний – так само, як відбитки пальців. З метою ідентифікації особини (чи особи – у криміналістиці тощо) невелику кількість геномної ДНК використовують для ампліфікації таких локусів. Продукти ампліфікації розділяють за допомогою електрофорезу: на електрофореграмі виникає специфічний для особини набір фрагментів різної довжини, тобто з різною кількістю повторів мінісателіта – своєрідний молекулярний відбиток (DNA fingerprint).

Серед різноманітних застосувань ПЛР слід окремо виділити *спрямований мутагенез* – редагування послідовності ампліфікованого фрагмента ДНК шляхом внесення в неї нуклеотидних замін. Одну з популярних процедур внесення таких замін – QuikChange® – схематично зображено на рис. 11.14.



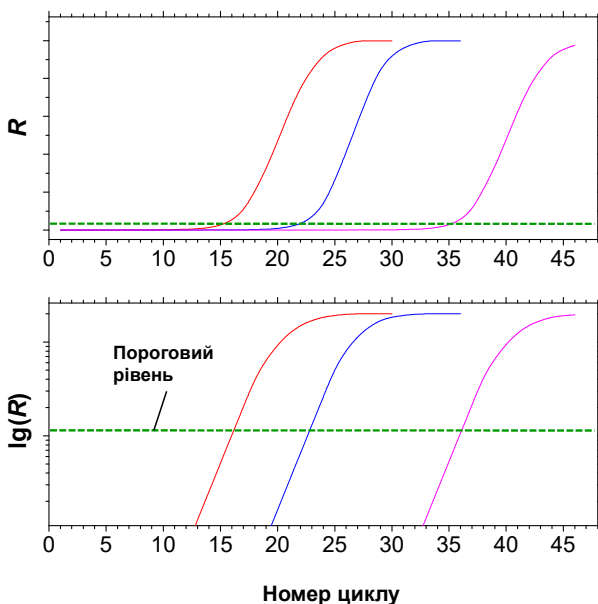
**Рис. 11.14.** Спрямований мутагенез за допомогою ПЛР: процедура QuikChange® для внесення нуклеотидних змін

Як матрицю для ампліфікації використовують циркулярну ДНК (рекомбінантну плазмиду), яка містить послідовність-мішень. Обидва праймери, комплементарні різним ланцюгам на одній ділянці, містять замінені нуклеотиди, що формують замінену нуклеотидну пару. Вихідна циркулярна ДНК (виділяється із відповідного штаму *E. coli*) при цьому метильована по сайтах рестрикції *DpnI* (тільки метильовані сайти є субстратом для зазначеної рестриктази). Після ПЛР утворюються циркулярні молекули з одноланцюговими розрізами, що містять замінену нуклеотидну пару (рис. 11.14); зрозуміло, що продукти ампліфікації

неметильовані. Оброблення рестриктазою дозволяє позбутись вихідних молекул без нуклеотидної заміни. Після перенесення у бактеріальну клітину одноланцюгові розриви в циркулярних молекулах репаруються (або застосовується лігаза перед трансформацією) і відбувається клонування послідовності-мішені, яка містить замінену нуклеотидну пару.

ПЛР застосовують також для ампліфікації РНК у вигляді її кДНК-копії після синтезу кДНК на РНК за допомогою зворотної транскриптази – варіант ПЛР, що позначається як RT-PCR (Reverse Transcriptase Polymerase Chain Reaction). Абревіатуру RT-PCR використовують також у іншому значенні – Real Time Polymerase Chain Reaction (інше позначення QRT-PCR – Quantitative Real Time Polymerase Chain Reaction): процедура, у якій ампліфікація ДНК супроводжується визначенням її кількості. Такий кількісний варіант ПЛР застосовують із метою оцінити вміст (щодо контрольного зразка) певного фрагмента ДНК у вихідному матеріалі.

Типовий вигляд кривих ампліфікації, отриманих у процесі кількісної ПЛР, зображено на рис. 11.15. Оцінювання вмісту ДНК у вихідному зразку базується на такому принципі: чим більше ДНК у вихідному матеріалі, тим раніше (за меншу кількість циклів ампліфікації) можна отримати її високу концентрацію за рахунок ПЛР. Для такого оцінювання криві ампліфікації будують у логарифмічному масштабі – лінійна ділянка відповідає експоненціальному зростанню кількості ДНК (у два рази після кожного циклу); згодом настає виснаження реакційної суміші й темпи зростання знижуються. Приблизно через центр такої ділянки проводять горизонтальну *порогову лінію*. На осі ординат вона відсікає певну порогову кількість ДНК  $R_T$ , перетин лінії із кривою ампліфікації дає порогове значення номеру циклу  $C_T$ , на якому ця кількість досягається. Якщо кількість ДНК у вихідному зразку становить  $R_0$ , то, оскільки у кожному циклі кількість ДНК зростає удвічі,  $R_T = R_0 \times 2^{C_T}$ .



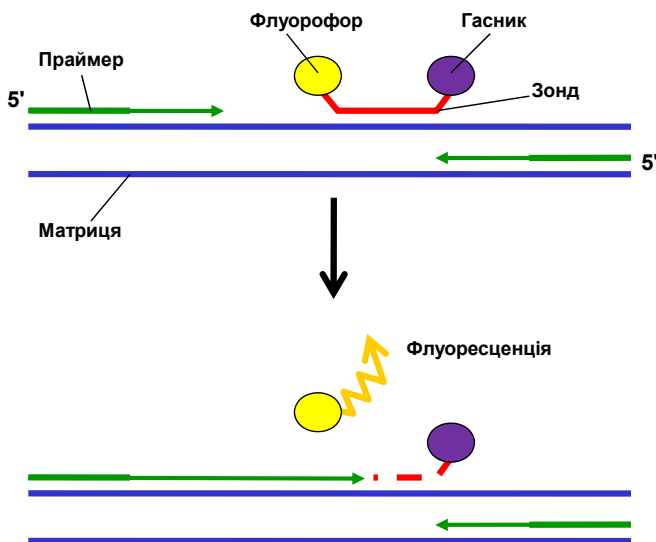
**Рис. 11.15.** Ідеалізовані криві ампліфікації (у двох варіантах) трьох препаратів ДНК (вихідна кількість знижується зліва направо) під час кількісної ПЛР.  $R$  – концентрація ДНК у реакційній суміші в умовних одиницях. Пороговий рівень – горизонтальна лінія, що проводиться приблизно через центр лінійної ділянки логарифмічних кривих; точка перетину кривої і горизонтальної лінії дає пороговий номер циклу  $C_T$

Таке саме рівняння можна записати для деякого стандартного зразка (одна із кривих на рис. 11.15 може бути таким стандартом), де замість  $R_0$  буде вихідна кількість у стандартному зразку  $R_S$ , а замість  $C_T$  – номер порогового циклу для стандартної кривої  $C_S$ . Тоді відносна кількість ДНК у досліджуваному зразку  $R_0 / R_S = 2^{\Delta C_T}$ , де  $\Delta C_T = C_S - C_T$ . Насправді замість двійки у цих рівняннях має фігурувати ефективність ампліфікації  $E$  – величина, зазвичай дещо менша двох, яка може бути визначена шляхом ампліфікації низки розведень стандартного зразка.

Оцінювання концентрації ДНК у реакційній суміші під час кількісної ПЛР здійснюється завдяки використанню флуоресцентних барвників. Найбільш популярним є застосування барвни-

ка типу SYBR<sup>®</sup>Green. Барвник зв'язується із дволанцюговою ДНК за рахунок інтеркаляції, багатократно підвищуючи при цьому інтенсивність своєї флуоресценції. Отже, крива ампліфікації відображає інтенсивність флуоресценції, яка є пропорційною до концентрації дволанцюгової ДНК.

Детекція за допомогою SYBR<sup>®</sup>Green не є високоспецифічною – певна частина дволанцюгової ДНК у суміші може бути продуктом неспецифічної ампліфікації. Інші методи детекції базуються на використанні специфічних зондів. Одним із популярних підходів такого типу є застосування зондів TaqMan<sup>®</sup>. Зонд є олігонуклеотидом, комплементарним одному з ланцюгів усередині фрагмента, що піддається ампліфікації (рис. 11.16). До кінців зонда приєднані флуоресцентний барвник і гасник, який пригнічує флуоресценцію, – вільний чи зв'язаний із матрицею зонд практично не випромінює. Під час ПЛР зонд піддається деградації за рахунок 5'-екзонуклеазної активності *Taq*-полімерази (яка є гомологом ДНК-полімерази I *E. coli*, див. розділ 9) – барвник звільнюється, що супроводжується його флуоресценцією.



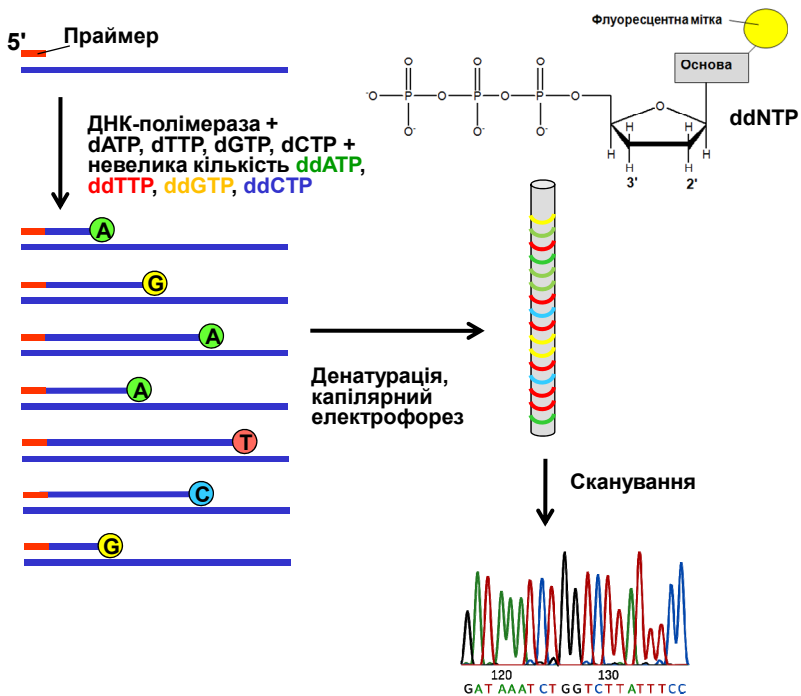
**Рис. 11.16.** Оцінювання концентрації ДНК під час кількісної ПЛР за допомогою зонда TaqMan<sup>®</sup>

## Секвенування ДНК

Клонований або ампліфікований фрагмент ДНК можна дослідити різними способами, однак найвичерпнішу інформацію дає встановлення нуклеотидної послідовності (sequence) фрагмента – *секвенування*. Методи секвенування ДНК лежать в основі багатьох дослідницьких підходів у молекулярній біології, а також широко використовуються для молекулярної діагностики.

### Метод Сангера

Розроблений у 1977 році, метод Сангера (Frederick Sanger), що належить до категорії методів *секвенування шляхом синтезу*, залишається доволі популярним і сьогодні. На рис. 11.17 показано схему цього класичного методу у його сучасному варіанті, що реалізується автоматизованими секвенаторами. Ампліфікований за допомогою ПЛР фрагмент денатурують: власне секвенуванню піддається одностанцюгова ДНК. До реакційної суміші додають праймер, повний набір дезоксинуклеозидтрифосфатів (dNTP), ДНК-полімерази й невелику кількість *дидезоксинуклеозидтрифосфатів* (ddNTP) усіх чотирьох типів. Дидезоксинуклеотид вирізняється тим, що містить атом Н замість ОН-групи не тільки при 2'-, а й при 3'-атомі пентози (рис. 1.17). Включення такого нуклеотиду в ланцюг, що синтезується, спричиняє зупинку синтезу внаслідок відсутності 3'-ОН-групи на його кінці. Оскільки ddNTP наявні у невеликій концентрації, то така подія буде відбуватися на різних матрицях у різних точках ланцюга – в усіх, де, наприклад, аденін розташований напроти тиміну у складі матриці.



**Рис. 11.17.** Секвенування ДНК за Сангером: схема синтезу ДНК у присутності ddNTP – дидезоксинуклеозидтрифосфатів. Одноланцюгові продукти синтезу аналізуються за допомогою капілярного електрофорезу – розподіл смуг дозволяє прочитати послідовність

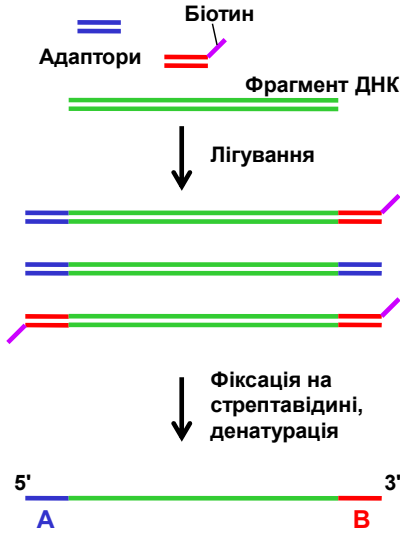
Після закінчення реакції синтезу всі продукти піддають денатурації й отримують набір одноланцюгових фрагментів від праймера до 3'-кінцевого нуклеотиду. При цьому кінцеві нуклеотиди є флуоресцентно міченими: ddNTP кожного типу модифікований специфічною для нього флуоресцентною міткою, яка випромінює у певному спектральному діапазоні (випромінювання чотирьох міток має різний колір). Колір випромінювання визначає тип кінцевого нуклеотиду, а довжина фрагмента – порядковий номер цього нуклеотиду у складі ланцюга. На завершальному етапі проводять капілярний електрофорез і сканують

гель лазерним променем, що збуджує флуоресценцію міток у смугах. Як показано на рис. 11.17, таке сканування дозволяє прочитати нуклеотидну послідовність.

Метод Сангера дає змогу з високою точністю отримати послідовність фрагмента завдовжки до 1000 нуклеотидів. Він є зручним тоді, коли треба виконати секвенування одного чи невеликої кількості фрагментів ДНК. Однак у випадку секвенування цілих геномів метод виявляється занадто тривалим і витратним. Для одночасного секвенування мільйонів фрагментів ДНК нині використовують методи другого покоління, які є значно швидшими й дешевшими.

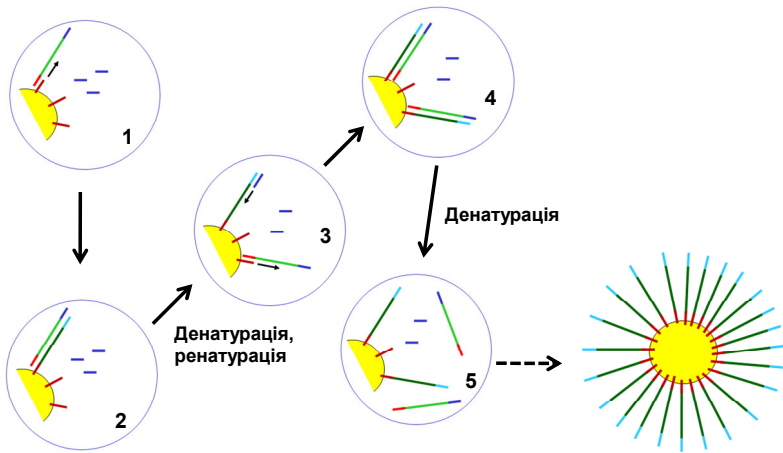
### **Методи секвенування другого покоління**

Усі методи секвенування другого покоління передбачають подібні підготовчі процедури з метою одночасної ампліфікації великої кількості різноманітних, отриманих у той чи інший спосіб, одоланцюгових фрагментів ДНК. Для встановлення нуклеотидної послідовності геному на підготовчому етапі геномна ДНК подрібнюється на невеликі фрагменти завдовжки 300–500 пар основ. Потім до кінців усіх фрагментів пришивають лігазою невеликі стандартні ділянки ДНК двох типів – *адаптори*. Далі необхідно відібрати тільки ті фрагменти, до двох кінців яких пришито *два різні* адаптори (зрозуміло, що після лігування буде отримана суміш трьох типів продуктів, рис. 11.18). Це завдання розв'язується завдяки модифікації кінцевої ділянки одного з ланцюгів одного з адапторів біотином (рис. 11.18). Біотин (вітамін Н) і бактеріальний білок стрептавідин вступають у надзвичайно міцну взаємодію, що часто використовується для міцного пришивання макромолекул до різноманітних поверхонь. У даному випадку суміш отриманих продуктів лігування пропускають через колонку, заповнену кульками, укритими стрептавідином. Як зрозуміло зі схеми на рис. 11.18, тільки два типи продуктів затримуються на колонці, а після денатурації дволанцюгової ДНК тільки один тип ланцюгів з неї змивається. У такий спосіб отримують набір (бібліотеку) різноманітних одноланцюгових фрагментів ДНК, до кінців яких пришиті два різні стандартні адаптори А та В.



**Рис. 11.18.** Приготування бібліотеки фрагментів ДНК для подальшого секвенування

У методі *піросеквенування* на наступному етапі всі ці фрагменти змішуються з компонентами, необхідними для здійснення ПЛР (у т. ч. праймерами, комплементарними адапторам А). Крім того, до суміші додають великий надлишок мікрокульок, до поверхні яких пришиті праймери, комплементарні адапторам В, і емульгують розчин в олії. Кожна кулька служить зародком, навкруг якого утворюється мікрокрапля водного розчину – своєрідний мікрореактор, що містить усі компоненти для ПЛР і тільки один певний фрагмент ДНК (не більше одного завдяки великому надлишку кульок). Далі в такій емульсії, нагріваючи й охолоджуючи її, проводять ПЛР. Перші етапи реакції показано на рис. 11.19. У результаті відбувається ампліфікація фрагментів (до кількох мільйонів копій), комплементарних вихідному, пришитих до поверхні мікрокульки.



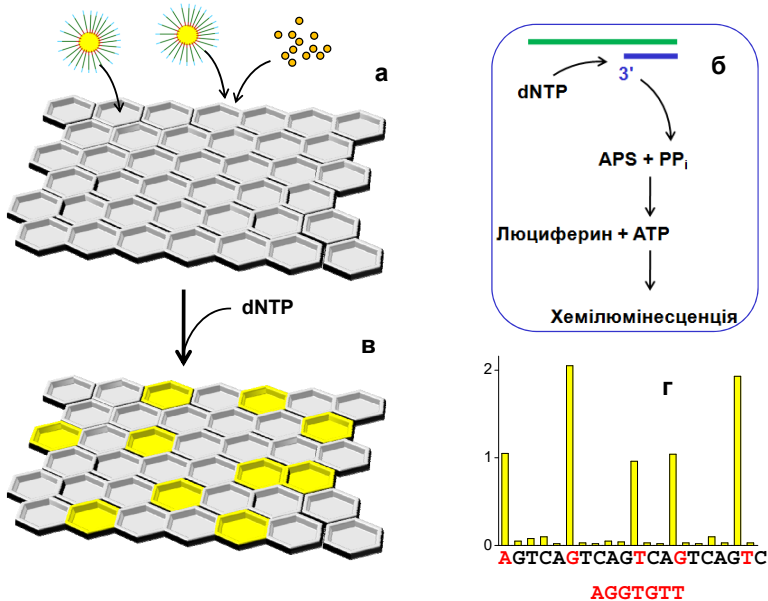
**Рис. 11.19.** Перші стадії ампліфікації ДНК на поверхні мікрокульки: 1 – адаптор В фрагмента ДНК гібридується з комплементарним праймером на поверхні кульки; 2 – синтезується комплементарна копія фрагмента, приєднана до поверхні кульки; 3 – після нагрівання й охолодження вихідний фрагмент гібридується з іншим праймером на поверхні кульки, праймер у реакційній суміші – з кінцевою ділянкою щойно синтезованого фрагмента; 4 – відбувається синтез відповідних ланцюгів; 5 – після денатурації маємо два ідентичні фрагменти, приєднані до кульки, і дві комплементарні ним матриці для наступних раундів ампліфікації

Далі всі кульки з ампліфікованими фрагментами ДНК завантажують у мікрокомірки реактора (по одній кульці на комірку) разом із кульками меншого розміру, до поверхні яких пришиті два ферменти: сульфурілаза й люцифераза (рис. 11.20, а). Реактор омивається розчином, який містить ДНК-полімеразу, праймер (що гібридується з адаптором А) і аденозинфосфосульфат (АПС). Нуклеозидтрифосфати подаються в реакційну суміш імпульсно один за одним. Коли нуклеотид певного типу виявляється комплементарним матриці та включається у зростаючий ланцюг у даній комірці, у розчин вивільняється пірофосфат  $PP_i$  –

саме ця обставина використовується для детекції приєднання нуклеотиду, звідки й походить назва методу. За рахунок активності сульфурілази пірофосфат замінює сульфат у складі APS, перетворюючи цю сполуку на АТР. У присутності АТР люцифераза каталізує реакцію за участю люциферину, яка супроводжується випромінюванням хемілюмінесценції (рис. 11.20, б). Світловий сигнал фіксується оптичною системою – наявність такого сигналу від комірки свідчить про приєднання відповідного нуклеотиду (рис. 11.20, в). При цьому оптична система вимірює інтенсивність випромінювання: удвічі більша інтенсивність сигналу свідчить про приєднання двох однакових нуклеотидів поспіль, утричі – трьох тощо. Після детекції сигналу реакційну суміш відмивають від залишкових реагентів і до неї подають наступний нуклеотид. Послідовність люмінесцентних сигналів від кожної комірки (*флауграма*, рис. 11.20, г) читається як нуклеотидна послідовність.

Після встановлення послідовностей усіх фрагментів (які отримані випадково й тому перекриваються) пошук однакових нуклеотидних ділянок на кінцях дозволяє визначити порядок розташування фрагментів. Описана процедура дає змогу встановити послідовність приблизно 200 млн пар основ за 4,5 год.

Дуже схожа на піросеквенування, але значно дешевша й швидша технологія – *іонне напівпровідникове секвенування* (ion torrent (semiconductor) sequencing). Перші стадії збігаються зі щойно описаними, але ця методика відрізняється способом детекції: кожна комірка реактора з'єднана з напівпровідниковим мікро-рН-метром. Крім пірофосфату, при приєднанні нуклеотиду до зростаючого ланцюга в розчин вивільняється протон  $H^+$  – саме ця подія, яка детектується рН-метром, служить сигналом про таке приєднання у комірку.



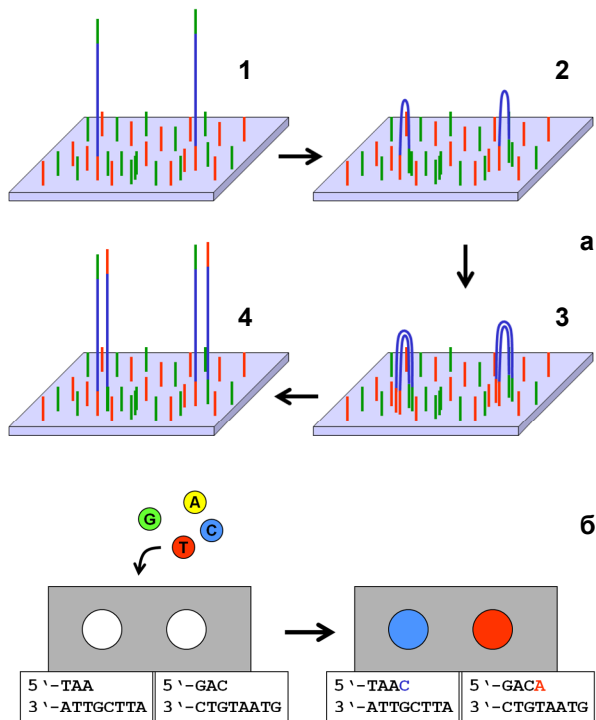
**Рис. 11.20.** Завершальна стадія піросеквенування.

У комірки реактора завантажуються кульки з ампліфікованими фрагментами ДНК і кульки меншого розміру із пришитими до них сульфурілазою та люциферазою (а); у відповідь на приєднання нуклеотиду випромінюється світло (б, в); цикл повторюється багато разів з усіма типами нуклеотидів для отримання флауграми, яка дозволяє прочитати послідовність (г)

**Метод Illumina** – інша популярна процедура швидкого секвенування другого покоління, розроблена компанією "Solexa" (сьогодні реалізується компанією "Illumina", США). У цьому випадку все починається також із отримання одноланцюгових фрагментів із двома типами адапторів на кінцях. Потім ці фрагменти пришивають до поверхні у проточній камері реактора, до якої щільно пришиті два типи праймерів, комплементарні адапторам.

Ампліфікація фрагментів відбувається шляхом твердофазної ПЛР. Кожен фрагмент утворює місток, комплементарно взаємодіючи своїм вільним кінцем із відповідним праймером (рис. 11.21, *a*). ДНК-полімераза подовжує цей праймер і, у такий спосіб, після першого раунду ПЛР синтезується комплементарний ланцюг, пришитий до поверхні. Після багатьох раундів фрагмент ампліфікується, утворюючи на поверхні пляму, що складається із пришитих одноланцюгових комплементарних фрагментів – генерується кілька мільйонів таких плям.

Процедура секвенування полягає в тому, що до поверхні, яка міститься у проточній камері, додають модифіковані нуклеозидтрифосфати всіх чотирьох типів та інші необхідні компоненти ДНК-полімеразної реакції, зокрема й однаковий для всіх праймер до одного з адапторів (у межах плями як матриця використовується тільки половина ланцюгів). Модифікація dNTP полягає в тому, що всі вони містять заблоковану 3'-ОН-групу дезоксирибози (синтез не може бути продовжений після приєднання такого нуклеотиду), а до основи кожного типу приєднана специфічна флуоресцентна мітка, яка випромінює у певному спектральному діапазоні. Після проходження реакції всі dNTP, що не прореагували, відмивають: якщо приєднання нуклеотиду певного типу відбулося в певній плямі (на певному фрагменті ДНК), то цей факт фіксують за флуоресцентним сигналом, причому за довжиною хвилі випромінювання (кольором) установлюють тип приєданого нуклеотиду (рис. 11.21, *б*). Далі відбувається відщеплення усіх флуоресцентних міток, хімічним шляхом розблоковується 3'-ОН-група кінцевого нуклеотиду й повторюється наступний раунд полімеразної реакції.



**Рис. 11.21.** Схема секвенування ДНК методом Illumina.  
 а – ампліфікація фрагментів ДНК – одноланцюгові фрагменти пришиваються до поверхні із праймерами (1), формують місток (2), після першого раунду синтезу ДНК (3) і денатурації (4) утворюються два комплементарні фрагменти, розташовані поряд;  
 б – дві плями ампліфікованої ДНК на поверхні; додаються флуоресцентно мічені нуклеозидтрифосфати, приєднання яких детектується за флуоресцентним сигналом

### Секвенування індивідуальних молекул

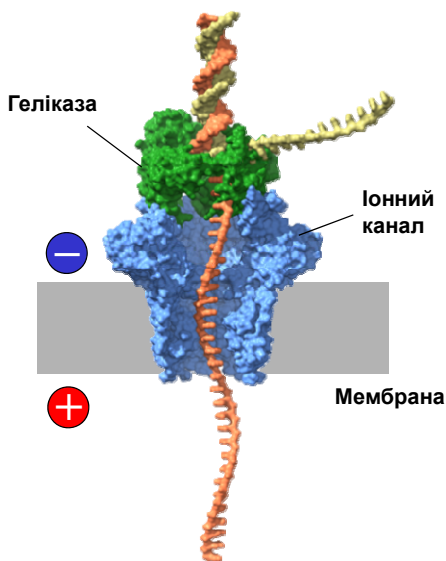
Методи секвенування індивідуальних молекул (SMRT-sequencing – Single Molecule Real Time) спрямовані на визначення послідовностей *індивідуальних* фрагментів ДНК у реальному часі й не потребують стадії ампліфікації.

Одним із таких підходів є *секвенування за допомогою індивідуальних молекул ДНК-полімерази*. Метод реалізується на

спеціальній наноструктурній матриці: на поверхню скла накладається алюмінієва сітка, комірки якої мають розмір у кілька десятків нанометрів. До поверхні скла в кожній комірці пришивається одна молекула ДНК-полімерази, зв'язаної з індивідуальним одноланцюговим фрагментом ДНК (матриця, що секвенується) і приєднаним до неї праймером. До реакційної суміші подаються нуклеозидтрифосфати, що несуть приєднану до кінцевого фосфату флуоресцентну мітку (як і в попередніх методиках, своя мітка на кожен із чотирьох типів нуклеотидів). Скло опромінюється світлом, що може збуджувати флуоресценцію міток, але, оскільки розмір комірок менший за довжину хвилі цього світла, то воно здатне проникати не більше як на ~30 нм углиб комірки (подібний ефект використовують, ставлячи металеву сітку на дверцятах мікрохвильової пічки для захисту від високочастотного випромінювання). Більшість міток у реакційній суміші залишаються невидимими. Коли нуклеозидтрифосфат виявляється комплементарним матриці, він залучається до реакції і затримується в активному центрі полімерази на кілька мілісекунд. Саме тоді тільки його мітка, яка знаходиться у безпосередній близькості до поверхні, сприймає збуджувальний світловий потік і починає випромінювати. Реєстрація цього випромінювання дає інформацію про тип нуклеотиду, що приєднується до зростаючого ланцюга. Щойно приєднання відбулося, мітка разом із пірофосфатом відщеплюється, дифундує у розчин і стає невидимою. Отже, нуклеотидна послідовність зчитується у реальному часі впродовж синтезу як послідовність імпульсів флуоресценції різного кольору.

Ще один популярний підхід – *секвенування за допомогою нанопор*. Нанопора має розмір, який дозволяє проходити крізь неї одноланцюговій нуклеїновій кислоті (ДНК або РНК) або окремим нуклеотидам. Зазвичай вона є іонним каналом, вбудованим у синтетичну силіконову мембрану, по обидва боки якої розміщені електроди. З одного боку мембрани з іонним каналом зв'язана або екзонуклеаза, що відщеплює нуклеотиди один за одним, або ДНК-геліказа, яка забезпечує розплітання дволанцюгового фрагмента ДНК. Один із ланцюгів після цього протягується крізь пору під дією різниці електричних потенціалів (рис. 11.22). Кож-

на пора оздоблена датчиком електропровідності: електропровідність залежить від типу нуклеотиду, що в певний момент міститься в порі (послідовність електричних сигналів прочитується як нуклеотидна послідовність). У варіанті цього підходу, реалізованого компанією "Oxford Nanopore Technologies", компактний прилад, де здійснюється описана процедура, приєднується до портативного комп'ютера – секвенування може бути виконано навіть у польових умовах.



**Рис. 11.22.** Нанопора на основі іонного каналу (у розрізі), що використовується для секвенування ДНК (на основі структур 6Sl7 і 1XHZ)

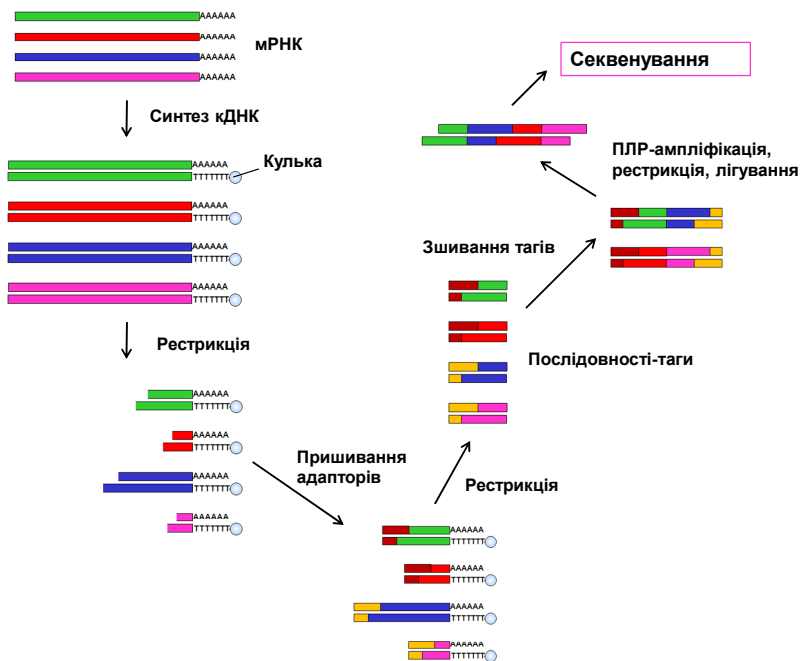
Загалом прогрес у методах секвенування ДНК вражає: якщо на початку XXI століття встановлення повної послідовності геному коштувало біля 100 млн дол., то сьогодні вартість такої процедури наближується до 1 тис. дол. Це відкриває нові можливості у молекулярній діагностиці захворювань, а також дає потужний дослідницький інструмент.

## Аналіз транскриптомів

Одну з можливостей дослідити повну програму експресії геному – установити *транскриптом* (повний набір транскриптів) – дають згадані вище ДНК-чіпи: загальну схему, яка починається з виділення тотального препарату мРНК, представлено на рис. 11.6.

Інший ефективний метод оцінювання рівня експресії багатьох генів одночасно – *серійний аналіз експресії генів* – SAGE (Serial Analysis of Gene Expression). Тут також усе починається з виділення тотального препарату мРНК, яка піддається зворотній транскрипції (рис. 11.23). Модифіковані біотином праймери для синтезу кДНК (polydT) імобілізуються на вкритих стрептавідинном мікрокульках. Після оброблення рестриктазою на кульках залишаються фрагменти різної довжини й весь матеріал розділюється на дві частини: у кожній до кінців фрагментів пришиваються стандартні адаптори одного із двох типів. Далі використовують рестриктазу третього класу, що має сайт упізнання у межах адаптора, а розріз робить на відстані 15 пар основ від нього (друга процедура рестрикції на рис. 11.23). У результаті залишаються короткі ділянки – *таги* (tags), пришиті до адапторів. Увесь підхід базується на тому, що для ідентифікації мРНК конкретного типу достатньо встановити нуклеотидну послідовність такої короткої ділянки, що служить специфічним маркером (у цьому випадку ділянка знаходиться ближче до 3'-кінця мРНК).

На останніх етапах процедури таги попарно зшивають між собою і піддають ПЛР-ампліфікації, використовуючи праймери до стандартних адапторів. Нарешті, рестриктазою відрізають пари тагів від адапторів і зшивають отримані фрагменти між собою, отримуючи доволі довгі молекули ДНК, що складаються з коротких ділянок, які походять від певних мРНК. Усі ці молекули на останньому етапі піддають секвенуванню одним із методів другого покоління: чим частіше в результатах секвенування зустрічається певний таг, тим вище рівень відповідної мРНК у вихідному препараті, тобто рівень експресії відповідного гена (зрозуміло, що повна нуклеотидна послідовність геному при цьому відома).



**Рис. 11.23.** Принцип серійного аналізу експресії генів (SAGE), пояснення в тексті

Схожа методика – *кеп-аналіз експресії генів* (CAGE, Cap Analysis of Gene Expression) – базується на аналізі коротких послідовностей-тагів на 5'-кінцях мРНК.

Крім того, установлення повної картини експресії генів можливе шляхом безпосереднього секвенування кДНК, отриманої з тотального препарату мРНК. Оскільки геномна послідовність відома, то немає необхідності секвенувати молекули кДНК повністю – достатньо встановити послідовності невеликих ділянок на кінцях, щоб знайти джерело відповідного транскрипту в геномі.

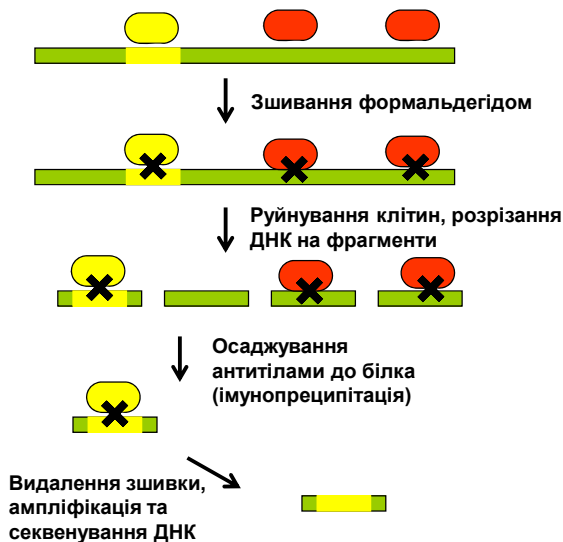
## Повногеномний аналіз ДНК-білкових взаємодій

Вивчення тонких механізмів білково-нуклеїнових взаємодій здебільшого реалізується різноманітними фізичними методами (див. нижче). У цьому підрозділі описано підходи, що дозволяють одночасно аналізувати велику кількість таких взаємодій у живій клітині.

Потужним засобом установлення наявності білка на певній ділянці ДНК *in vivo* є **метод імунопреципітації хроматину** (ChIP – Chromatin Immuno Precipitation) (рис. 11.24). Перший крок – оброблення клітин (або клітинних ядер) формальдегідом із метою ковалентного пришивання усіх білків до ДНК. Далі здійснюють лізис клітин, виділяють ДНК і частково її фрагментують. Ковалентні ДНК-білкові комплекси осаджують антитілами, специфічними до конкретного білка (або пропускають через колонку, на якій іммобілізовані антитіла). Отримані комплекси руйнують, видаляючи зшивку хімічним шляхом, і аналізують ДНК, яка була зв'язана у клітині з білком, що досліджується. Найефективнішим варіантом аналізу є секвенування отриманих фрагментів ДНК одним із методів другого покоління з метою знайти в геномі ділянки (ділянку), зв'язані із цим білком. Якщо зробити зшивання формальдегідом у різні моменти після певного регуляторного сигналу, то можна встановити послідовність зв'язування білків (транс-факторів і кофакторів) з регуляторними ділянками (*цис*-елементами) промотору певного гена.

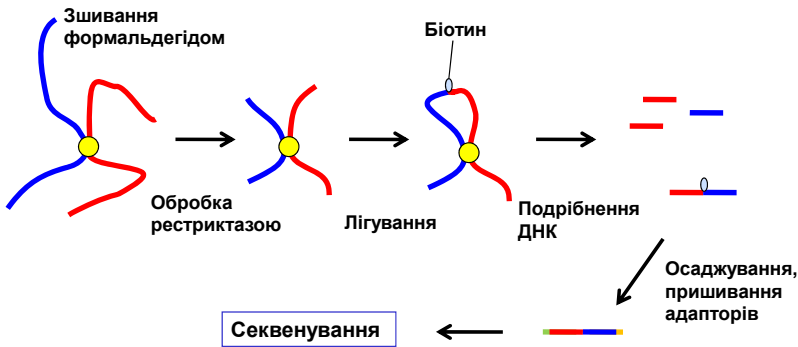
Цілком аналогічний підхід використовують для повногеномного **картування нуклеосом** – установлення їхніх позицій у геномі. Після оброблення формальдегідом, завдяки чому всі нуклеосоми фіксуються у своїх позиціях, виділяють із клітин зшитий матеріал і обробляють мікрококовою нуклеазою. Ця ендонуклеаза робить дволанцюгові розрізи, відповідно доступна для неї тільки лінкерна ДНК, оскільки у складі нуклеосоми один із ланцюгів завжди зв'язаний із гістонами (див. рис. 4.9). Отже, мікрококова нуклеаза вирізає нуклеосомні частинки. Після руйнування зшивок отримують великий набір нуклеосомних ДНК, які секвенують методом другого покоління. Чим частіше в результатах секвенування зустрічається певний фрагмент, тим вище частота наявнос-

ті нуклеосоми у цій ділянці. Саме у такий спосіб отримано, зокрема, дані щодо розташування нуклеосом відносно стартових точок транскрипції, представлені на рис. 6.17.



**Рис. 11.24.** Імунопреципітація хроматину

Картування нуклеосом – це з'ясування "первинної" структури хроматину. На використанні зшивання формальдегідом і секвенування великої кількості фрагментів ДНК базуються також методи дослідження просторової структури хроматину на вищих рівнях його організації. Один із найпотужніших сучасних методів такого типу позначається *Hi-C* і належить до родини підходів під загальною назвою *фіксація (захоплення) конформації хроматину* (chromatin conformation capture). Схему методу представлено на рис. 11.25: тут усе також починається з оброблення клітин формальдегідом, завдяки чому ділянки деякої хромосоми (або різних хромосом), що контактують одна з одною, фіксуються ("захоплюються") у такому контакті.



**Рис. 11.25.** Схема визначення частоти контактів між двома ділянками хроматину методом Hi-C

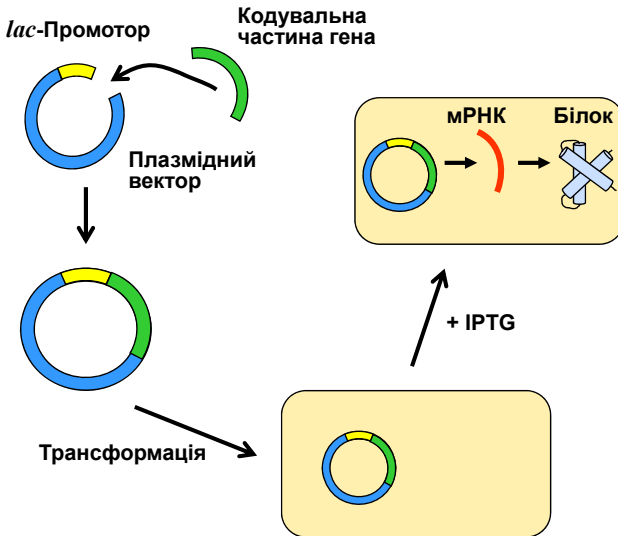
Далі зшитий матеріал обробляють деякою рестриктазою, яка залишає тільки зшиті фрагменти відносно невеликого розміру, а потім – лігазою. Перед з'єднанням лігазою кінці фрагментів модифікують біотином. Після подрібнення ДНК зшиті біотинильовані фрагменти осаджують стрептавідином, до їхніх кінців пришиваються адаптори і здійснюється їхнє секвенування: чим частіше серед секвенованих зшитих фрагментів зустрічається той, який виник унаслідок контакту між двома ділянками, тим частіше ці дві ділянки контактують між собою у клітині. Сучасні уявлення щодо вищих рівнів організації хроматину (петельні й топологічно асоційовані домени, компартменти хроматину), розглянуті в розділі 4, значною мірою базуються на даних, отриманих методом Hi-C і деякими його варіаціями.

## Синтез рекомбінантних білків

Оскільки переважна більшість білків синтезується клітиною в невеликих кількостях, то будь-які дослідження таких білків після їхнього простого виділення практично неможливі. Описані вище технології рекомбінантних ДНК розв'язують цю проблему. Ген, що кодує той чи інший білок, можна клонувати, вбудувати (напр. у бактеріальну плазмиду) і змусити бактеріальну культуру

продукувати цей білок. Виділення білка з культури й очищення його біохімічними методами вже не є принциповою проблемою завдяки його великій кількості.

Одну з найпростіших схем синтезу білка в *E. coli* зображено на рис. 11.26. Зрозуміло, що еукаріотичний ген не має сенсу вводити у прокаріотичну клітину – прокаріоти не мають системи сплайсингу. Тому беруть лише кодувальну частину гена, яку можна отримати, наприклад, з бібліотеки клонів кДНК. Використовуючи придатну рестриктазу й лігазу, потрібну кДНК вбудовують у плазмідний вектор для експресії поряд із промотором, наприклад лактозного оперона (див. розділ 6). Здійснюють трансформацію бактеріальних клітин, до бактеріальної культури додають синтетичний індуктор *lac*-оперона IPTG (ізопропіл- $\beta$ -D-1-тіогалактопиранозид – хімічний аналог алолактози). Промотор активується, і відбуваються транскрипція гена і трансляція білка.



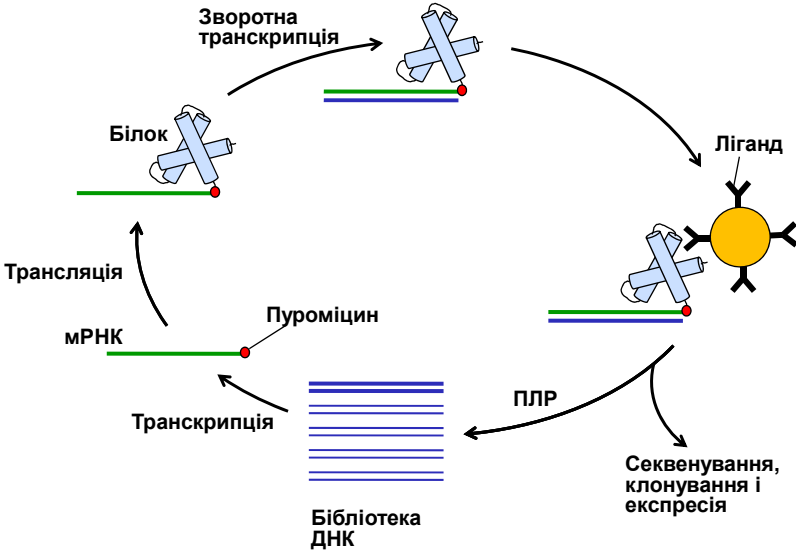
**Рис. 11.26.** Експресія чужорідного гена в бактеріальній клітині.  
*lac*-Промотор – промотор лактозного оперона,  
IPTG – синтетичний індуктор оперона

Більш ефективна двостадійна система експресії використовує промотор РНК-полімерази бактеріофага T7. У клітину вводять два плазмідні вектори: один містить *lac*-промотор і ген РНК-полімерази T7, інший – сильний промотор полімерази T7 разом із геном, що має бути експресований. IPTG індукуює експресію гена полімерази, яка зв'язується тільки із промотором у складі другої плазміди (інших таких промоторів у бактеріальному геномі немає) і забезпечує синтез великої кількості білка.

Вивчення структури й особливостей функціонування багатьох білків було б неможливим без такої штучної експресії. Крім того, використання рекомбінантних ДНК для експресії білків відкриває принципово нові можливості: послідовність ДНК можна практично як завгодно змінити й отримати мутантну форму білка. До білкового гена можна пришити ген іншого білка, який буде виконувати роль мітки. Наприклад, широко використовують із такою метою флуоресцентні білки (див. нижче підрозділ щодо візуалізації макромолекул). Частина гена, що відповідає частині білка (напр. структурному домену), можна вилучити й отримати редукований білок чи окремий домен для структурних досліджень. Будь-який нуклеотид можна замінити на інший методами спрямованого мутагенезу (див. рис. 11.14), завдяки чому отримати білок з амінокислотною заміною. Такі підходи мають велике значення для вивчення механізмів активності ферментів, специфічного впізнання білками інших молекул тощо. Практично в усіх описаних операціях ідеться про **білкову інженерію** – створення білків із новими властивостями.

Методи сучасної білкової інженерії дозволяють отримати ферменти для каталізу реакцій, що взагалі не відбуваються у біологічних системах. Одним із таких підходів є своєрідна молекулярна селекція: створення великої бібліотеки випадкових мутантних нуклеотидних послідовностей (стартуючи зазвичай від якогось природного аналога), їхня експресія і відбір продуктів, що мають бажані властивості. Існують кілька схем такого відбору: на рис. 11.27 зображено як приклад одну з них – схему, спрямовану на пошук білків із високою спорідненістю до якогось ліганду, що базується на утворенні ковалентних кон'югатів білків і відповідних молекул мРНК. У подібний спосіб створено,

наприклад, протеази з новими субстратними специфічностями, полімерази, здатні синтезувати різноманітні модифіковані нуклеїнові кислоти (такі, що складаються із неприродних синтетичних аналогів нуклеотидів) тощо.



**Рис. 11.27.** Молекулярна селекція білків із високою спорідненістю до певного ліганду: створюється велика бібліотека фрагментів ДНК, що несуть випадкові мутації; шляхом транскрипції *in vitro* синтезуються відповідні мРНК, до 3'-кінців яких ковалентно приєднується пуроміцин; при трансляції *in vitro* пуроміцин ковалентно пришивається до С-кінця синтезованого поліпептидного ланцюга; шляхом зворотної транскрипції на кон'югатах мРНК-білок синтезуються комплементарні ланцюги ДНК; кон'югати ДНК/мРНК-білок пропускаються через колонку з імобілізованими лігандами певного типу; із затриманого на колонці матеріалу вилучаються ланцюги ДНК, які піддаються ампліфікації шляхом ПЛР, і формуються бібліотека ДНК для наступного раунду. Після кількох раундів отримують білок, що найміцніше взаємодіє з лігандом, – відповідну ДНК ампліфікують, секвенують, вбудовують у вектор для експресії й отримують бажаний білок у великій кількості

Велику кількість білків (у т. ч. тих, що застосовують у медицині та біотехнологіях) виробляють сьогодні шляхом синтезу в бактеріальних клітинах. Поряд із прокаріотичними використовують також системи синтезу рекомбінантних білків в еукаріотичних клітинах. Особливо важливі еукаріотичні системи експресії для еукаріотичних білків, які не можуть бути синтезовані бактеріальною клітиною в активній формі. Зокрема, це стосується білків, які піддаються суттєвим посттрансляційним модифікаціям, наприклад глікопротеїдів.

Підходи для штучної експресії чужорідних генів в еукаріотичних клітинах розділяють на дві категорії. Перша – це *транзйентна експресія* (transient gene expression). Для транзйентної експресії виконують *трансфекцію*<sup>1</sup> культури клітин плазмідним вектором, у який вбудований деякий ген (*трансген*), оздоблений еукаріотичним промотором. Оскільки плазміда не вбудовується у геном клітини, то вона працює тимчасово, протягом кількох діб, а потім зникає після клітинних поділів і/або внаслідок нуклеазної деградації. Транзйентну експресію частіше використовують не для виділення білка, що синтезується, а як потужний засіб дослідження ефектів появи такого білка у живих клітинах. Наприклад, синтезований у такий спосіб білок може бути мічений флуоресцентним білком, що дозволить спостерігати його розподіл у клітині, швидкість дифузії, взаємодію з іншими білками тощо.

З метою *стабільної експресії* введений у клітину трансген може бути інтегрований у ядерний геном – у цьому випадку йдеться про генетичну модифікацію або редагування геному клітини.

---

<sup>1</sup> Термін *трансформація*, що зустрічався вище, використовують для позначення перенесення ДНК у бактеріальні й рослинні клітини; у випадку тваринних клітин його замінює термін *трансфекція*, оскільки під трансформацією тваринних клітин розуміють їхнє перетворення на злоякісні.

## Редагування геномів

Здійснити інтеграцію трансгена у геном клітини можна за рахунок гомологічної рекомбінації (рис. 11.28): ген у складі плазміди, якою трансфікують клітину, фланкують ділянками ДНК, що збігаються із певними ділянками у складі деякої хромосоми. Після введення такої плазміди всередину клітини там, з певною імовірністю, спрацьовує система гомологічної рекомбінації (див. розділ 10), що забезпечує інсерцію гена в бажану ділянку.

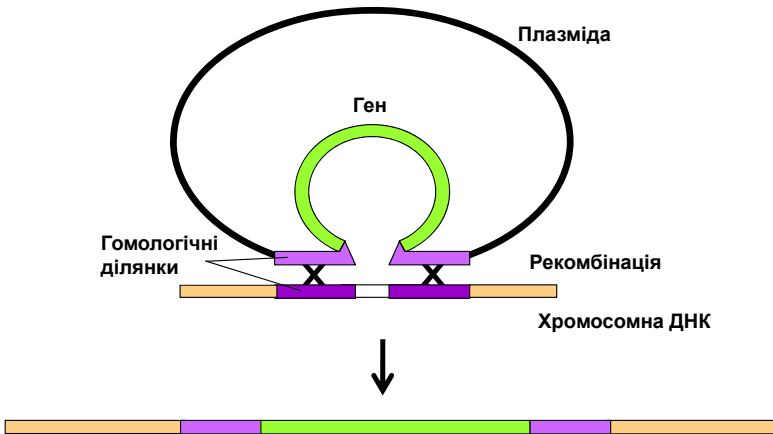


Рис. 11.28. Інтеграція трансгена у хромосому

За схемою на рис. 11.28 можна, навпаки, видалити певний ген із геному клітини-мішені, тобто здійснити *генний нокаут*. У такому випадку замість трансгена у клітину вводять безмістовний фрагмент ДНК, фланкований ділянками, що оточують цей ген. Із яйцеклітини, з якою виконано таку операцію, можна отримати модельну (*нокаутну*) тварину (зазвичай ідеться про мишей). Генний нокаут – надзвичайно важливий метод дослідження функціональної ролі конкретного гена. Більш ефективні підходи до генного нокауту та інших операцій редагування конкретних ділянок геному пов'язані із застосуванням методики *CRISPR-Cas*, яка революціонізувала генетичну інженерію.

Методика базується на використанні компонентів своєї імунної системи бактерій проти бактеріофагів. У бактеріальних геномах містяться особливі ділянки – кластери регулярно розміщених коротких паліндромних повторів (CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats), з якими асоційовано кілька білкових генів *cas* (CRISPR-associated genes). CRISPR є набором стандартних паліндромних повторів, розділених унікальними спейсерами, кожен із яких за походженням є ділянкою ДНК якогось бактеріофага. Продуктом транскрипції CRISPR і подальшого процесингу є набір молекул РНК, кожна з яких містить стандартний елемент (продукт паліндромного повтору) і ділянку завдовжки 20 нуклеотидів, комплементарну ДНК бактеріофага. Частина білків *cas* забезпечують вбудовування фагової ДНК у CRISPR, інша частина є ефекторами, що містять геліказні й ендонуклеазні активності. Ефекторні білки утворюють комплекс з одним із РНК-продуктів CRISPR. РНК у цьому комплексі виконує роль "гіда", спрямовуючи комплекс на ДНК бактеріофага у випадку інфекції, а ця ДНК розрізається ендонуклеазами у складі комплексу. Існують кілька різновидів системи CRISPR-Cas, один із яких має за ефекторний білковий компонент тільки один мультидоменний білок – Cas9. За ендонуклеазні активності у складі білка відповідають два різні структурні домени, що вносять розрізи у два різні ланцюги ДНК. Сама система CRISPR-Cas9 найширше використовується для редагування геномів.

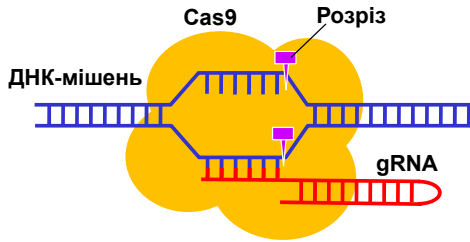
З метою такого редагування у клітину того чи іншого типу вводять плазмідний вектор для транзиторної експресії (тимчасової експресії цілком достатньо для досягнення бажаного ефекту), який несе ген білка Cas9 і сконструйовану ділянку, продуктом транскрипції якої є РНК-гід (gRNA). У такий спосіб комплекс Cas9-gRNA можна спрямувати практично будь-куди – у цьому місці відбувається локальне плавлення ДНК за рахунок геліказної активності Cas9, здійснюється комплементарна взаємодія одного з ланцюгів із gRNA і продукується дволанцюговий розріз ДНК за рахунок ендонуклеазних активностей білка (рис. 11.29). Єдина вимога для зв'язування комплексу з мішенню полягає в тому, що сайт впізнання у ДНК має бути фланкованим

з одного боку невеличким елементом послідовності (зазвичай завдовжки 3 пари основ) – так званим PAM (Protospacer-Adjacent Motif). Наприклад, для Cas9 із *Streptococcus pyogenes* послідовність PAM має вигляд NGG, що дозволяє спрямувати комплекс у середньому на кожні ~12 пар основ у геномі ссавців.

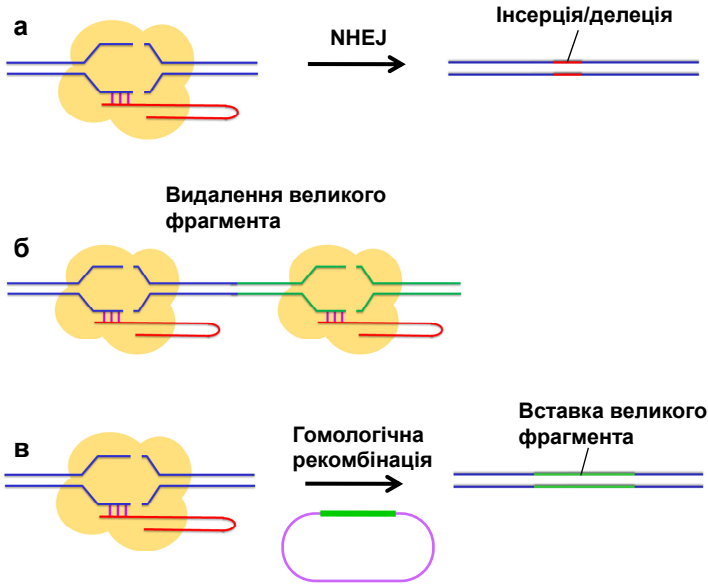
Отже, у найпростішому випадку за допомогою системи CRISPR-Cas можна ввести дволанцюговий розріз у визначене місце геному живої клітини. Унаслідок цього можуть бути реалізовані дві можливості. Перша базується на тому, що дволанцюговий розріз є мішенню для системи репарації за механізмом негомологічного з'єднання кінців NHEJ (див. рис. 9.33). У процесі NHEJ у місці залікованого розрізу завжди виникають мутації типу невеликих вставок і/або делецій, які призводять до пошкодження гена – практично йдеться про зручну методику генного нокауту (рис. 11.30, а). Застосування двох комплексів CRISPR-Cas із різними РНК-гідами дозволяє здійснити два розрізи на певній відстані, тобто вирізати велику ділянку з геному (рис. 11.30, б).

Крім того, дволанцюгове розрізання – це подія, яка ініціює процес гомологічної рекомбінації (за наявності гомологічної ДНК-матриці, що використовується для точної репарації розрізу, див. розділ 10). Якщо ввести у клітину відповідну матрицю у складі плазміди, то система гомологічної рекомбінації забезпечує інсерцію нуклеотидної послідовності, яка відповідає матриці, у місце розрізу (рис. 11.30, в).

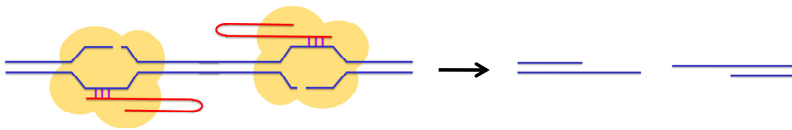
Ділянка РНК-гіда завдовжки 20 нуклеотидів не завжди може забезпечити високу специфічність – у геномі, крім бажаної мішені, зазвичай є такі самі або подібні випадкові послідовності в різних локусах. Тому в підходах на основі CRISPR-Cas використовують також мутантні варіанти Cas9, позбавлені (за рахунок амінокислотних замін) однієї з ендонуклеазних активностей, що перетворює Cas9 на *ніказу*, яка вносить одноланцюговий розріз (*нік*) у той чи інший ланцюг у визначеному місці. При використанні двох таких мутантних комплексів із двома сконструйованими відповідним чином РНК водять два ніки в різні ланцюги ДНК на невеликій відстані – це продукує дволанцюговий розріз із виступаючими одноланцюговими виростами (рис. 11.31). Сенс такої процедури полягає у суттєвому підвищенні специфічності щодо нуклеотидної послідовності.



**Рис. 11.29.** Дволанцюговий розріз, що є наслідком взаємодії комплексу білка Cas9 і РНК-гіда (gRNA) з послідовністю-мішенню у складі геномної ДНК



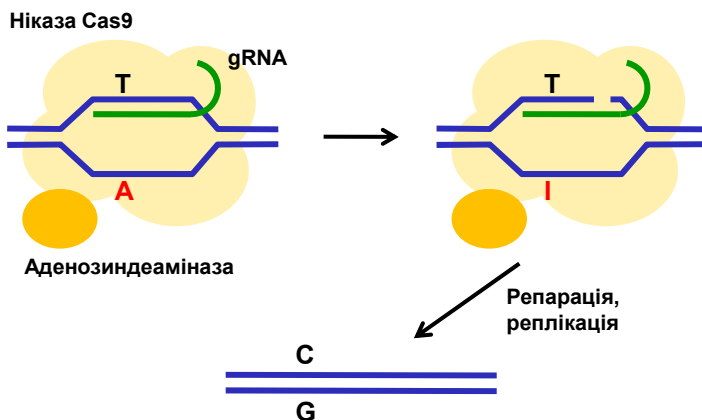
**Рис. 11.30.** Варіанти використання технології CRISPR-Cas:  
 а – пошкодження гена внаслідок інсерцій / делецій після зашивання дволанцюгового розрізу репараційної системою NHEJ;  
 б – вирізання великого фрагмента двома комплексами;  
 в – вставлення фрагмента ДНК у місце розрізу за рахунок гомологічної рекомбінації (за наявності відповідної матриці)



**Рис. 11.31.** Використання мутантних форм (ніказ) Cas9 із відсутніми ендонуклеазними активностями: два комплекси з відповідними РНК спрямовуються на дві близько розташовані ділянки, результатом чого стає дволанцюговий розріз із одноланцюговими виростами

Ще одним варіантом є використання мутантів Cas9, позбавлених обох ендонуклеазних активностей. У цьому випадку комплекс Cas9-gRNA просто зв'язується в певному визначеному місці геному. Якщо це місце наближене до якоїсь функціональної ділянки (напр. промотору), а до Cas9 приєднано регуляторний білок (транскрипційний фактор, кофактор, що запускає процес хімічних модифікацій у хроматині, тощо), то у такий спосіб можна впливати на рівень експресії певного гена.

Нікази Cas9 із приєднаними до них іншими білками використовують також для адресного внесення нуклеотидних змін та інших операцій редагування нуклеотидних послідовностей у живій клітині. На рис. 11.32 представлено схему заміни пари АТ на ГС. У цьому випадку до нікази Cas9 приєднано аденозиндеаміназу – фермент, що за участю молекули води перетворює аденін на інозин (І), неканонічний пурін, часто наявний у складі тРНК (див. розділ 8). Природна деаміназа здатна працювати тільки з РНК (субстратом служить рибоаденозин). Фермент, субстратом якого є дезоксирибоаденозин, було створено шляхом молекулярного добору у спосіб, подібний до зображеного на рис. 11.27. Отже, аденозиндеаміназа перетворює А на І (який потім сприймається системами репарації і реплікації як гуанін), а ніказа робить одноланцюговий розріз в іншому ланцюзі. Далі у клітині спрацьовує репарація місметчів, яка могла б замінити або Т на С, або І на А. Проте наявність розрізу в ланцюзі, де залишився Т, робить саме цей ланцюг переважною мішенню репарації (див. розділ 9): після репарації і реплікації у цьому місці залишається ГС-пара.

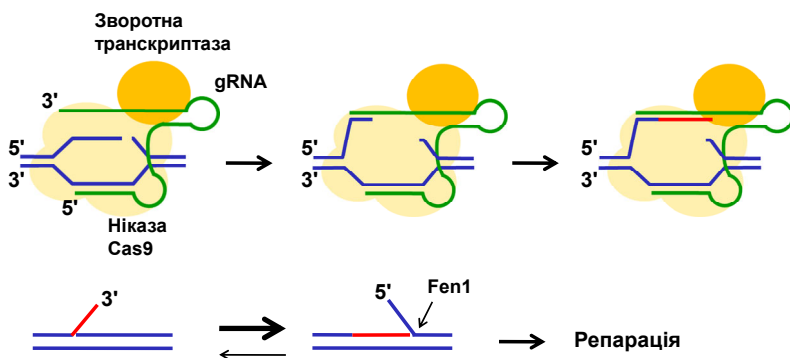


**Рис. 11.32.** Використання CRISPR-Cas для внесення нуклеотидних змін: ніказа Cas9 продукує односторонній розріз, а приєднаний до неї фермент перетворює аденін на інозин в іншому ланцюзі. Після репарації і реплікації відбувається стабільна заміна на пару GC

Цілком аналогічною є процедура заміни пари GC на AT: до нікази Cas9 приєднують цитидиндеаміназу (перетворює цитозин на урацил, який потім сприймається ДНК-полімеразами як тимін) та інгібітор урацилглікозилази (ферменту, що вирізає урацили з ДНК), а ніказа робить розріз у ланцюзі, що містить гуанін.

Нуклеотидні заміни за описаними процедурами не завжди можливі: потрібно, щоб на певній відстані від бажаного місця заміни був тринуклеотид PAM і поряд не було такого самого нуклеотиду, який може стати побічним субстратом для ферментів, що здійснюють заміну. Більш ефективна процедура (Prime Editing) одночасного внесення у визначену ділянку кількох нуклеотидних змін, а також інсерцій (до 44 пар основ) і делецій (до 80 пар основ), передбачає використання приєднаної до нікази Cas9 зворотної транскриптази (рис. 11.33). Матрицею (у яку і вносять бажані зміни нуклеотидної послідовності) для зворотної транскрипції служить 3'-кінцева ділянка гідової РНК. Після розрізання одного ланцюга ДНК ніказою його 3'-кінцева ділянка гібридується з РНК і служить праймером для зворотної транс-

крипції. Після синтезу ДНК у місці розрізу залишаються дві одноланцюгові ділянки: 3'-кінцева (відредагована) і 5'-кінцева, що містить вихідну нуклеотидну послідовність. Одна із двох ділянок є зайвою і може бути відрізана ендонуклеазою. Однак наявність у клітині реплікативної ендонуклеази Fen1 (яка в еукаріотів залучена до видалення праймерів на ланцюзі, що запізнюється, див. розділ 9) робить переважним відщеплення невідредагованої 5'-кінцевої ділянки (рис. 11.33). Зрозуміло, що після закінчення такої процедури в ДНК залишаються два ланцюги, серед яких тільки один відредагований. Система репарації далі використовує цей ланцюг для внесення змін в інший ланцюг: щоб саме інший ланцюг піддавався змінам, у нього вноситься одноланцюговий розріз, що забезпечується додатковим комплексом нікази Cas9, спрямованим на ділянку неподалік від місця редагування.



**Рис. 11.33.** Одночасне внесення нуклеотидних замін / інсерцій / делецій за допомогою нікази Cas9 і приєднаної до неї зворотної транскриптази. Верхня панель: після внесення одноланцюгового розрізу 3'-кінець гібридується зі сконструйованим виштовом гідової РНК; зворотна транскриптаза синтезує ДНК на РНК-матриці. Нижня панель: після дисоціації комплексу можуть бути реалізовані показані на рисунку дві можливості, але за рахунок ендонуклеази Fen1 рівновага зсунута в бажаний бік – залишається синтезована (відредагована) послідовність одного з ланцюгів. За рахунок внесення додатковим комплексом ще одного розрізу в невідредагований ланцюг саме він піддається репарації

## Аналіз протеомів

Кінцевою точкою експресії спадкової інформації є набір білків. Аналіз цього набору (*протеому*), відносної кількості окремих білків і складної картини взаємодій між ними є одним із найактуальніших завдань молекулярної біології.

Найбільш стандартним засобом розділення білкових сумішей є електрофорез, який найчастіше відбувається в поліакриламідному гелі. Найпопулярнішим є варіант електрофорезу білків у присутності додецилсульфату натрію (**S**odium **D**odecyl **S**ulfate – SDS) – негативно зарядженого детергенту, який розгортає поліпептидний ланцюг і вкриває його своєрідною "шубою". У результаті відбувається розділення ланцюгів різної довжини за принципом, що ілюструє рис. 11.1. Для специфічного визначення конкретного білка у складній суміші, як і для нуклеїнових кислот, використовують процедуру блотингу, яку, за аналогією зі згаданими вище назвами саузерн- і нозерн-блотингів, називають *вестерн-блотингом*. Вестерн-блотинг передбачає перенесення білків із гелю після електрофорезу на фільтр і подальше оброблення фільтра антитілами, специфічними до певного білка, а далі вторинними антитілами до цих антитіл. Вторинні антитіла мічені (флуоресцентною міткою, ферментом, що каталізує деяку кольорову реакцію тощо), завдяки чому можна детектувати наявність конкретного білка в тотальному білковому препараті й оцінити його кількість.

Ефективним підходом для аналізу суміші, що містить велику кількість різноманітних білків, 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 ізоелектричне фокусування – електрофорез у градієнті рН, коли здійснюється розділення білків за їхнім сумарним зарядом (ізоелектричною точкою), у другому – електрофорез у присутності додецилсульфату натрію, тобто розділення за молекулярною масою. Двовимірний електрофорез дозволяє ефективно розділити до 1 тис. білків одночасно.

Зручним і потужним методом розділення білкових сумішей є **високоєфективна рідинна хроматографія** (High Performance Liquid Chromatography – HPLC). Найчастіше розчин денатурованих білків наносять на колонку, заповнену твердими пористими частинками, які адсорбують білки за рахунок гідрофобних взаємодій. Змивання (елюція) розчинником, який пропускають через колонку під високим тиском і полярність якого варіюється (напр. лінійно зростає), дозволяє розділити окремі білки, що незначною мірою розрізняються за ступенем гідрофобності. Метод використовують як для аналізу, так і з метою виділення й очищення білків.

Потужним засобом аналізу складних білкових сумішей є **мас-спектрометрія**, принцип якої полягає в розділенні пучка заряджених частинок у 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, зокрема й олігопептиди та білки. Білок з молекулярною вагою до 200 тис. у концентрації  $10^{-15}$  моль/л може бути детектований за допомогою сучасних мас-спектрометрів.

Проте визначення маси білка ще не дозволяє точно ідентифікувати його. Із цією метою білок піддають обмеженій протеолітичній деградації. Отримана комбінація олігопептидів, маса яких визначається за допомогою мас-спектрометрії, є своєрідним "відбитком" білка. Наступний аналіз геномних послідовностей (які транслюються в амінокислотні) дозволяє знайти відповідну амінокислотну послідовність білка, що практично позбавляє необхідності її хімічного визначення.

**Тандемна мас-спектрометрія** дозволяє отримати пряму інформацію щодо послідовності олігопептидів завдяки їхній вторинній фрагментації в камері спектрометра й подальшому визначенню розподілу отриманих фрагментів за масою.

Наступним кроком у вивченні протеому є з'ясування складної картини білок-білкових взаємодій. Одним із популярних методів дослідження таких взаємодій є **pull-down**: молекула білка, що використовується як пастка (bait), адсорбується на колонці завдяки пришитому до нього специфічному ярлику (tag). Через колон-

ку пропускають білкову суміш, утворені білкові комплекси аналізують за допомогою гель-електрофорезу.

За допомогою мас-спектрометрії можна також здійснити ідентифікацію білків, вилучених із комплексів і розділених електрофорезом. Аналогічно до ДНК-мікроареїв у дослідженнях білок-білкових взаємодій застосовують білкові мікроареї – скельця із пришитими до них білками (понад 5 тис. білків), які дозволяють виконувати скринінг протеому.

Ще один ефективний підхід – *визначення профілю рибосом* (ribosome profiling) – дозволяє отримувати інформацію про протеом під час його створення рибосомами. Метод полягає у виділенні полісом, обробленні їх РНКазою й отриманні захищених рибосомами фрагментів мРНК завдовжки ~30 нуклеотидів, які конвертуються у кДНК і секвенуються (рис. 11.34). Накладання результатів секвенування на відомий транскриптом дозволяє оцінити частоту наявності рибосом на конкретних ділянках конкретних транскриптів. Аналіз таких рибосомних профілів дозволяє характеризувати протеом клітин певного типу, вивчати швидкість трансляції і механізми її регуляції.

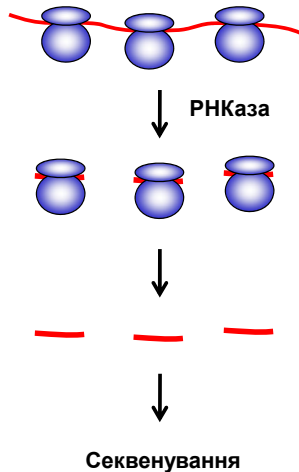


Рис. 11.34. Схема отримання профілю рибосом

## Фізичні методи дослідження біологічних макромолекул

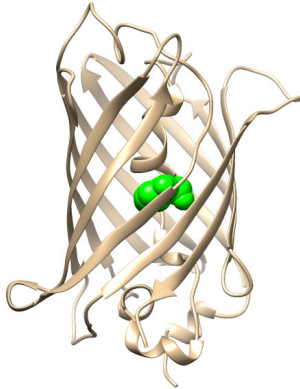
Широкого застосування в молекулярній біології набули різноманітні фізичні методи, важливість яких важко переоцінити. Детальний опис принципів, іноді доволі складних, фізичних методів дослідження структури й функціонування біополімерів та їхніх комплексів виходить за межі цієї книги. Нижче наведено огляд основних таких методів із зазначенням деяких особливостей і характеру інформації, яку можна отримати за їхньою допомогою.

### Візуалізація біологічних макромолекул

Дослідження біологічних макромолекул у клітині або *in vitro* часто потребують їхньої візуалізації з різною роздільною здатністю, не кажучи вже про просте бажання побачити предмет дослідження. Зрозуміло, що йдеться про різноманітні типи мікроскопії. При цьому мікроскопія найнижчої роздільної здатності – оптична – не тільки не втрачає свого значення, а й переживає своєрідний ренесанс. Зокрема, важливу роль відіграє флуоресцентна мікроскопія. Передусім це пов'язано із двома обставинами: створенням *конфокальних мікроскопів* – оптико-електронних пристроїв, що дозволяють робити тонкий віртуальний "зріз" досліджуваного зразка, і відкриттям *флуоресцентних білків*. Першим таким білком став зелений флуоресцентний білок (GFP, Green Fluorescent Protein), виділений із тихоокеанської медузи *Aequorea victoria* (рис. 11.35). Біологічна роль цього білка у життєдіяльності медузи залишається не зовсім зрозумілою, але він став важливим інструментом досліджень.

Загалом у білках є два амінокислотні залишки, Трп і Тур, спроможні до слабкої флуоресценції в ультрафіолетовому діапазоні. Однак у складі GFP реалізується особливий випадок: у процесі укладання просторової структури за рахунок автокаталітичної хімічної реакції між трьома амінокислотними залишками (Ser, Тур, Gly) формується особливий флуорофор – сполука, здатна до інтенсивної флуоресценції у видимій області (зеленого кольору). Після відкриття GFP подібні білки були знайдені та-

кож в інших організмів: зокрема, вони визначають різноманітні кольори коралових рифів. На основі таких білків шляхом амінокислотних замін було створено цілу палітру рекомбінантних флуоресцентних білків, що випромінюють по всьому видимому спектру – від блакитного до темно-червоного.

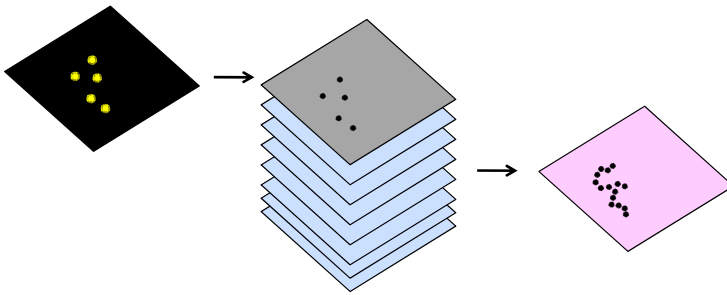


**Рис. 11.35.** Зелений флуоресцентний білок (GFP)  
*Aequorea victoria* (1GFL).  
Зеленим кольором виділено флуорофор

Широке використання флуоресцентних білків обумовлено двома обставинами. По-перше, на відміну від численних синтетичних флуоресцентних міток, флуоресцентні білки (оскільки це білки) є природними біологічними макромолекулами, які не чинять жодного шкідливого впливу на клітини. По-друге, методами генетичної інженерії ген флуоресцентного білка можна приєднати до іншого білкового гена – у результаті у клітині буде синтезуватись (або стабільно, якщо таку модифікацію реалізовано в ядерному геномі, або тимчасово, унаслідок транзиторної експресії плазмідного вектора) той чи інший флуоресцентно мічений білок. Застосування конфокального мікроскопа дозволяє спостерігати за такими білками зокрема й у живих клітинах. У такий спосіб досліджують внутрішньоклітинні розподіли різноманітних білків і динаміку цих розподілів.

Особливим варіантом використання флуоресцентних білків є метод **FRAP – відновлення флуоресценції після фотознебарвлення** (Fluorescence Recovery After Photobleaching), що дозволяє визначати кількісні параметри внутрішньоклітинної дифузії флуоресцентно мічених білків. Для цього невеличку зону в межах клітини опромінують інтенсивним лазерним променем, який руйнує флуорофори, – у цій зоні залишається чорна пляма. Далі за допомогою конфокального мікроскопа, який дозволяє кількісно оцінювати інтенсивність флуоресценції від певної області, вимірюють швидкість заповнення плями за рахунок тих флуоресцентно мічених білків, що були в інших місцях і флуорофори яких не піддавались такому руйнівному впливу. Одним із численних прикладів застосування FRAP є дослідження кінетики процесу ініціації транскрипції флуоресцентно міченими РНК-полімеразами, що дозволило отримати оцінки проміжків часу, які полімераза проводить у тому чи іншому стані, і ймовірностей переходів від одного стану до іншого (див. розділ 5).

Добре відомо, що оптична мікроскопія має принципове обмеження щодо роздільної здатності, яка не може бути більшою ніж кілька сотень нанометрів (довжина хвилі видимого світла). Проте нині існують методи оптичної (флуоресцентної) мікроскопії високої роздільної здатності (іноді до 5 нм), що можуть обходити це обмеження. Один із таких підходів – **STORM** (Stochastic Optical Reconstruction Microscopy) – базується на збудженні флуоресценції зразка імпульсом світла низької інтенсивності. Після імпульсу спостерігається і запам'ятовується зображення, де випромінують кілька випадкових флуоресцентно мічених молекул, що перебувають на великій відстані (більше 250–300 нм) одна від одної. Власне, така відстань і дозволяє подолати обмеження щодо роздільної здатності. Комп'ютерна програма визначає позиції центрів випромінювання і запам'ятовує їх. Процедура повторюється багато разів (зрозуміло, що використовують зафіксований зразок), а накладання зображень дозволяє отримати фінальну картину з високим розділенням (рис. 11.36).



**Рис. 11.36.** Принцип мікроскопії STORM: накладання багатьох зображень, отриманих після випромінювання обмеженої кількості випадкових флуоресцентно мічених макромолекул

Зрозуміло, що методи мікроскопії не обмежуються оптичним діапазоном. Історично першим методом спостереження макромолекулярних комплексів із роздільною здатністю у нанометровому діапазоні стала *електронна мікроскопія*. При застосуванні цього методу зразок розміщують у вакуумній камері мікроскопа, опромінюють променем електронів, який далі проєктується на флуоресцентний екран. Зразок при цьому має бути зафарбований певною речовиною, непрозорою для електронів.

Одна з поширених технік полягає в тому, що макромолекули наносять на карбонову поверхню (підкладку), практично прозору (як і макромолекули) для електронного променя, і фіксують на ній. Потім підкладку промивають розчином солі важкого металу, наприклад ураніацетатом. Після висушування 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, мультибілкові комплекси тощо. Розділення при негативному контрасті лімітоване розміром частинок важкого металу.

Висушені зразки у вакуумній камері електронного мікроскопа можуть бути пошкоджені інтенсивним електронним променем. Цю проблему можна обійти за допомогою особливого варіанта електронної мікроскопії – *кріоелектронної мікроскопії*,

яка передбачає швидке заморожування тонкого шару (~100 нм) розчину макромолекул у присутності кріопротекторів до температури рідкого азоту. Опромінення склоподібного замороженого зразка не надто інтенсивним променем електронів і подальше оброблення отриманих зображень дозволяють візуалізувати макромолекулярні об'єкти у більш властивому ним середовищі. Більше того, кріоелектронна мікроскопія стала потужним методом установлення просторових структур біополімерів (розглянемо це питання у наступному підрозділі).

Важливою альтернативою електронній мікроскопії є *атомно-силова мікроскопія* (або *мікроскопія атомних сил* AFM – *Atomic Force Microscopy*). Макромолекули адсорбуються на деякій поверхні (підкладці), після чого зразок може бути досліджений як після висушування, так і тоді, коли залишається шар рідини. Поверхня сканується *кантилівером* – голкою з надзвичайно тонким вістрям, яке за рахунок міжатомних (вандерваальсових) сил взаємодіє з поверхнею, відстежуючи її "ландшафт" (рис. 11.37). Рух голки реєструється оптичною системою (лазерний промінь, що віддзеркалюється від голки на фотодіод) і перетворюється комп'ютером на зображення поверхні. Роздільна здатність порівнянна з такою електронного мікроскопа, однак, на відміну від електронної мікроскопії, AFM не потребує фіксації і контрастування зразка, дозволяє працювати з макромолекулами, що перебувають в умовах, близьких до фізіологічних. Крім того, оскільки прилад реєструє висоту голки відносно поверхні, то отримане зображення є тривимірним.

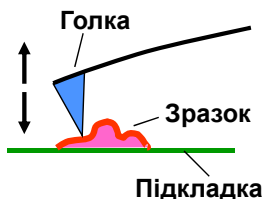


Рис. 11.37. Принцип атомно-силової мікроскопії (AFM)

## Визначення просторової структури макромолекул

Історично першою і до цього часу 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 упорядковану тривимірну решітку. Кристали, придатні для рентгеноструктурного аналізу (розміром щонайменше 0,3 мм), готують із перенасичених розчинів макромолекул, контролюючи швидкість зростання кристала за рахунок варіювання концентрації солі, температури, додаючи органічні розчинники тощо. У кристалах білків і нуклеїнових кислот зазвичай залишається значна кількість упорядкованих молекул води.

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

Картина рентгеновського розсіювання містить інформацію щодо просторових позицій атомів у кристалі. Вилучення цієї інформації є складним математичним завданням, хоча сьогодні комп'ютерна техніка дозволяє аналізувати рентгенограми доволі швидко: практично обмежувальним кроком рентгеноструктурного аналізу є вирощування кристалів. Безпосереднім результатом аналізу рентгенограми є складна тривимірна карта електронної щільності. Інтерпретація цієї карти можлива за умови, що послідовність біополімеру у складі кристала відома. Шляхом численних підгонкових ітерацій на комп'ютері ця послідовність (чи послідовності кількох елементів досліджуваного комплексу) узгоджується з електронною картою, результатом чого стає встановлення просторової координати кожного атома в макромолекулі.

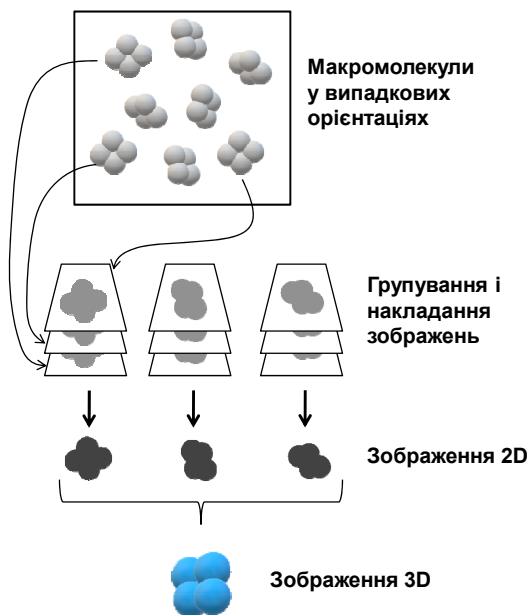
Зазвичай великі макромолекули виявляються надто складними для того, щоб рентгеновське розсіювання на їхніх кристалах дозволило отримати тривимірну структуру. Для визначення структури більшості білків, нуклеїнових кислот і комплексів застосовують *метод ізоморфних похідних*, який потребує набору крис-

талів макромолекули: вихідний кристал і принаймні два – три кристали, які були б ідентичними, але містили 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. В останньому випадку методами спрямованого мутагенезу здійснюють заміну однієї амінокислоти на Cys, до SH-групи якого можна легко приєднати атом важкого металу. Отже, розвиток методів генетичної інженерії і синтезу рекомбінантних білків став важливою передумовою структурних досліджень: рекомбінантні технології дали змогу не тільки отримувати достатню кількість високоочищених білків, придатних для подальшої кристалізації, а й готувати їхні ізоморфні похідні. Порівняння рентгенограм, отриманих від немодифікованого кристала і його похідних, дозволяє врешті-решт розрахувати структуру макромолекули.

Рентгеноструктурний аналіз залишається основним джерелом структурної інформації – більше 87 % із більш ніж 160 тис. структур, представлених сьогодні у базі даних PDB (її назва Protein Data Bank не означає, що там відображені тільки білки), установлені саме завдяки йому (як і більшість структур, наведених у цій книзі). Однак метод має суттєві обмеження, пов'язані передусім із тим, що далеко не всі біологічні макромолекули можна отримати у кристалічній формі. Зокрема, погано кристалізуються макромолекули, що містять невпорядковані ділянки, мембранні білки, мультимолекулярні комплекси великого розміру. За деякими винятками (нуклеосоми, рибосоми) рентгеноструктурний аналіз найбільше придатний для установлення структур із молекулярною масою до ~150 кДа. Інший недолік методу полягає в тому, що за рахунок сил кристалічної упаковки іноді можливі деякі деформації просторової структури у кристалах. Крім того, кристал дає статичну картину молекулярної структури, не дозволяючи досліджувати конформаційну динаміку макромолекул за умов, наближених до фізіологічних.

Подолати вказані обмеження допомагають два важливі альтернативні підходи, що сьогодні швидко розвиваються, – криоелектронна мікроскопія (~5 % структур у PDB) і спектроскопія двовимірного ядерного магнітного резонансу (більше 7 %).

**Кріоелектронна мікроскопія** має ту перевагу, що вона не потребує кристалізації макромолекул і придатна для структурних досліджень комплексів великого розміру (зазвичай не менше  $\sim 150$  кДа, що є обмеженням для цього підходу). Типова роздільна здатність кріоелектронної мікроскопії наближається до  $3 \text{ \AA}$  (іноді менше), що гірше роздільної здатності рентгеноструктурного аналізу ( $1\text{--}2 \text{ \AA}$ ). Схему застосування кріоелектронної мікроскопії для структурних досліджень представлено на рис. 11.38. Зрозуміло, що в тонкому шарі замороженого зразка макромолекули, які дають "відбиток" на екрані мікроскопа, орієнтовані випадково. На першому етапі всі зображення, отримані від подібних орієнтацій, об'єднуються у групи й накладаються одне на одне. Після такого накладання комп'ютерна програма розраховує спочатку двовимірні зображення високої роздільної здатності для різних орієнтацій, а наступне об'єднання цих зображень дає тривимірну картину.



**Рис. 11.38.** Установлення просторової структури за допомогою кріоелектронної мікроскопії

Другий альтернативний метод структурних досліджень – **ядерний магнітний резонанс (ЯМР)** – придатний для відносно невеликих макромолекул (зазвичай до  $\sim 80$  кДа), але це єдиний на сьогодні метод, за допомогою якого можна встановити структуру макромолекули в розчині. Спектроскопія ЯМР, яка широко використовується в органічній і біоорганічній хімії, базується на тому, 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 спричиняє зміну магнітних характеристик зразка, яку можна зафіксувати. Значення резонансної частоти залежить від того, до складу якої хімічної групи належить гідроген і якого типу хімічні групи містяться поруч, тобто спектр ЯМР, який показує залежність ефективності поглинання від частоти, містить інформацію про типи хімічних груп, їхнє взаємне розташування і ступінь рухливості у складі молекули.

Особливо цінною для молекулярної біології є техніка **двовимірного ЯМР**, коли сигнали резонансу детектують після складної комбінації кількох радіочастотних імпульсів і відображають як функцію двох частотних координат. На спектрі такого типу можна розрізнити сигнали гідрогенів різних залишків і виміряти зсуви цих сигналів, зумовлені іншими наближеними залишками. Величина зсуву є мірою відстані між залишками: порівняння цієї інформації з амінокислотою або нуклеотидною послідовністю дозволяє з'ясувати просторову структуру білка або нуклеїнової кислоти. Широко використовують також техніки, що передбачають одночасну реєстрацію спектрів ЯМР від ядер двох типів (напр.  $^1\text{H}$  і  $^{13}\text{C}$  або  $^{15}\text{N}$ ). На такому спектрі окремі піки відображають пари зв'язаних атомів, що належать різним залишкам, що також дозволяє інтерпретувати спектр у структурних термінах. Оскільки вимірювання спектра виконують у розчині, то результатом двовимірного ЯМР є набір дуже подібних одна до одної структур, що відображає конформаційну рухливість досліджуваного об'єкта (рис. 11.39).



**Рис. 11.39.** Ансамбль структур комплексу деякого білка із дволанцюговою РНК, отриманий за допомогою двовимірного ЯМР (6WLH)

### Оптична спектроскопія

Спектральні методи не дозволяють (за винятком двовимірного ЯМР) установити просторову структуру макромолекул, але дають змогу вивчити особливості їхньої конформаційної рухливості та взаємодій у розчині. Загальний принцип будь-якої спектроскопії полягає в тому, що розчин опромінюють тим чи іншим типом електромагнітного випромінювання, яке *мінімально* взаємодіє зі зразком, тобто жорстке іонізуюче випромінювання, яке здатне руйнувати ковалентні зв'язки, непридатне для спектральних досліджень. Різні типи молекулярної спектроскопії використовують діапазони випромінювання, які характеризуються не дуже високою енергією – від ближнього ультрафіолету і видимого світла (оптична спектроскопія) до інфрачервоного випромінювання (коливальна спектроскопія) і високих радіочастот (магнітний резонанс).

Зазвичай лише деякі хімічні групи зразка, які можна розглядати як реперні, здатні до специфічної взаємодії з випромінюванням певного типу. Ключовою обставиною при цьому є те, що характер взаємодії залежить від мікрооточення таких груп. Певні параметри випромінювання, які змінюються унаслідок його взаємодії з реперними групами, реєструються. Їхня інтерпретація дозволяє отримати інформацію про мікрооточення груп, тобто

структуру макромолекули або, частіше, про її перебудову внаслідок зміни умов, взаємодій з іншими молекулами тощо.

Найчастіше для досліджень макромолекул застосовують різні типи оптичної спектроскопії – *спектрофотометрію*, *круговий дихроїзм* і *флуоресцентну спектроскопію*.

Спектрофотометр – прилад для вимірювання ефективності поглинання світла в ультрафіолетовій і видимій областях – є неодмінним атрибутом будь-якої молекулярно-біологічної лабораторії. Передусім *спектрофотометрію* використовують із метою визначення концентрацій молекул: ефективність поглинання пропорційна концентрації. Однак ефективність поглинання може залежати також від мікрооточення хімічних груп, які відповідають за поглинання, – *хромофорів*. Найяскравіше така залежність виявляється для нуклеїнових кислот: поглинання азотистих основ в ультрафіолеті значно знижується (*гіпохромний ефект*) за умови реалізації між ними стекінг-взаємодій (див. розділ 3). Дослідження денатурації нуклеїнових кислот (плавлення подвійної спіралі) за допомогою протилежного, *гіперхромного*, ефекту (зростання поглинання при руйнуванні стекінг-взаємодій) дає інформацію про стабільність подвійних спіралей залежно від зовнішніх умов, нуклеотидного складу, іноді – особливостей послідовності тощо.

*Спектроскопія кругового дихроїзму* (КД) базується на різниці в поглинанні оптично активними молекулами (якими є, зокрема, амінокислоти й нуклеотиди) світла, що поляризоване по колу ліворуч і праворуч. Для амінокислот і нуклеотидів така різниця залежно від довжини хвилі (спектр КД) виявляється у ближній ультрафіолетовій зоні (220–300 нм). Форма спектра КД залежить від взаємної орієнтації однотипних оптично активних молекул (амінокислот чи нуклеотидів) у складі макромолекули. Відповідно аналіз спектра дозволяє оцінити відносний вміст різних типів вторинної структури в молекулі білка, а також зміни цього вмісту при структурних перебудовах, і зареєструвати переходи між структурними формами подвійних спіралей нуклеїнових кислот.

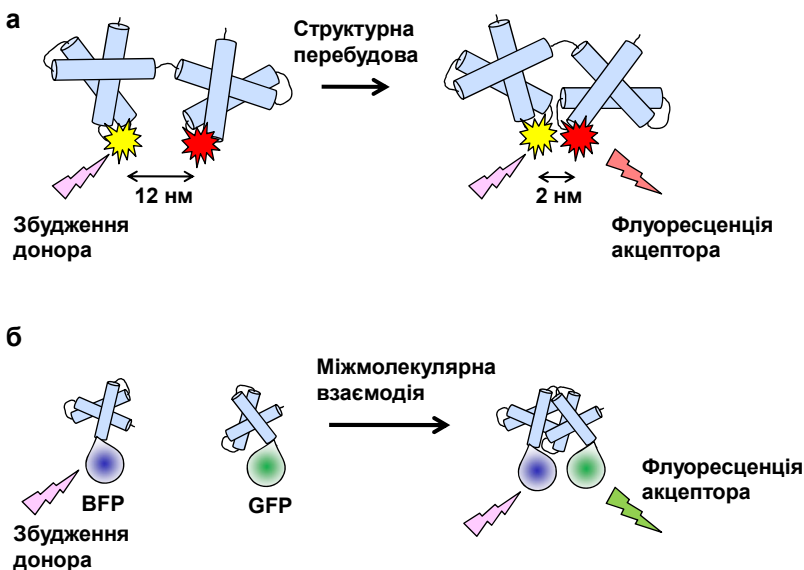
Певні хромофори після поглинання світла (збудження хромофора з переходом електрона на вищій енергетичний рівень) здатні втрачати енергію збудження шляхом випромінювання. Випромінювання такого типу, яке називається *флуоресценцією* (цей

термін доволі часто зустрічався вище), є основою **флуоресцентної спектроскопії**. У складі біополімерів до флуоресценції в ультрафіолеті спроможні дві амінокислоти – Трп і Туг. Дослідження спектрів власної білкової флуоресценції є важливим підходом, однак широке використання флуоресцентної спектроскопії зумовлено переважно великою кількістю синтетичних флуоресцентних зондів, які нековалентно взаємодіють із макромолекулами, і міток, які ковалентно пришиваються до макромолекул, а також використанням згаданих вище флуоресцентних білків. Висока інтенсивність флуоресценції зондів і міток дозволяє, по-перше (про що вже йшлося вище), детектувати молекули певного типу в невеликих кількостях: у процесі кількісного ПЛР, у гелях після електрофорезу, при гібридизації з ДНК-мікроареями, у живих клітинах за допомогою флуоресцентного мікроскопа. По-друге, різноманітні параметри флуоресценції залежать від мікрооточення зондів, міток і природних білкових хромофорів, що дає багату інформацію про характер конформаційної рухливості макромолекул, структурні перебудови, взаємодію між макромолекулами тощо.

Вимірювання кінетики флуоресценції під час певного процесу (коли процес запускається шляхом швидкого змішування реагентів) дозволяє вивчати тонкі механізми каталітичної активності ферментів, білково-нуклеїнового впізнання, роботи молекулярних моторів. Наприклад, використання флуоресцентних похідних aa-tРНК дозволило детально з'ясувати послідовність стадій елонгаційного циклу рибосоми (див. рис. 8.22, 8.27).

Серед різних флуоресцентних підходів особливе місце займає метод **резонансного перенесення енергії** (FRET – Fluorescence Resonance Energy Transfer). Перенесення енергії можливе на відстані для певних пар здатних до флуоресценції хромофорів: зазвичай використовують флуоресцентні мітки, які пришивають до конкретних хімічних груп макромолекули. Одна з міток є донором, інша – акцептором енергії. За відсутності перенесення енергії збудження донора викликає його флуоресценцію у певному спектральному діапазоні. Перенесення полягає в тому, що енергія збудження передається акцептору: він починає випромінювати при збудженні донора. Отже, перенесення енергії можна легко зафіксувати за зменшенням інтенсивності флуоресценції донора та/або зростанням флуоресценції акцептора. Оскільки ефектив-

ність перенесення залежить від відстані (для різних донорно-акцепторних пар відстань, на якій ефективність перенесення становить 50 %, варіює в межах  $\sim 2\text{--}5$  нм), то метод є ефективною молекулярною "лінійкою", що дозволяє вимірювати відстані в нанометровому діапазоні та зміни в цих відстанях при структурних перебудовах (рис. 11.40, *a*). Одним із численних прикладів застосування такого підходу є дослідження взаємних рухів субодиниць рибосоми при її транслокації (див. рис. 8.25).



**Рис. 11.40.** Резонансне перенесення енергії (FRET) у дослідженнях внутрішньомолекулярних структурних перебудов і міжмолекулярних взаємодій: *a* – дві мітки (донор і акцептор) приєднані до білка; за великої відстані між ними збудження донора приводить до його флуоресценції, акцептор не випромінює; скорочення відстані приводить до зниження інтенсивності флуоресценції донора й підвищення інтенсивності флуоресценції акцептора; *б* – до двох білків приєднані синій (BFP) і зелений (GFP) флуоресцентні білки, перший служить донором, другий акцептором; за відсутності взаємодії між білками збудження донора приводить до його випромінювання у синій частині спектра, утворення комплексу детектується за появою зеленої флуоресценції унаслідок перенесення енергії на акцептор

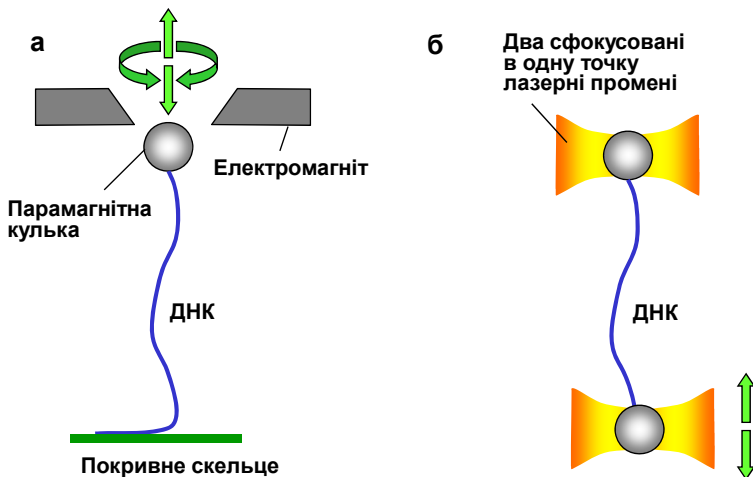
Цілком аналогічно завдяки перенесенню енергії можна зареєструвати утворення комплексу (зокрема й у живій клітині) між двома макромолекулами, до яких приєднані донор і акцептор (рис. 11.40, б). Донорно-акцепторною парою у цьому випадку зазвичай служать флуоресцентні білки: наприклад, синій флуоресцентний білок (BFP, Blue Fluorescent Protein) стає донором (збуджується світлом із довжиною хвилі 380 нм, максимум випромінювання 445 нм), а зелений флуоресцентний білок (GFP) – акцептором, флуоресценція якого (з максимумом 510 нм) з'являється при збудженні донора за умови утворення комплексу.

### **Методи дослідження індивідуальних макромолекул**

Більшість методів дослідження передбачають роботу з доволі великою кількістю (ансамблем) однакових молекул. Однак існують підходи, що дозволяють спостерігати за поведінкою одичних (зазвичай доволі великих) макромолекул і макромолекулярних комплексів: ДНК та її комплексів із білками, хроматиновою фібрили, актоміозиновою фібрили тощо. Одним із таких підходів є описаний вище FRET (за схемою на рис. 11.39, а), застосований до макромолекули, імобілізованої на деякій поверхні: сучасні флуоресцентні мітки й конфокальні мікроскопи дозволяють вимірювати флуоресценцію від однієї молекули.

Широке застосування мають також методи механічних маніпуляцій з окремими макромолекулами. Є два основні підходи для таких маніпуляцій: *магнітний пінцет* (magnetic tweezers) і *оптична пастка* (optical trap), яку називають також *оптичним пінцетом* (рис. 11.41).

У найпростішому випадку в експерименті з магнітним пінцетом молекулу ДНК завдовжки кілька тисяч пар основ пришивають за один кінець до покривного скельця мікроскопа, за інший – до парамагнітної кульки ~2 мкм у діаметрі. Кулька перебуває в полі електромагніта, який можна переміщувати або обертати, як показано на рис. 11.41, а, прикладаючи тим самим зовнішню силу до кінця молекули. Розмір кульки дозволяє спостерігати за нею (тобто за індивідуальною молекулою ДНК) за допомогою звичайного оптичного мікроскопа. При цьому реєструються два параметри – величина розтягувальної сили й відстань між двома кінцями молекули ДНК.

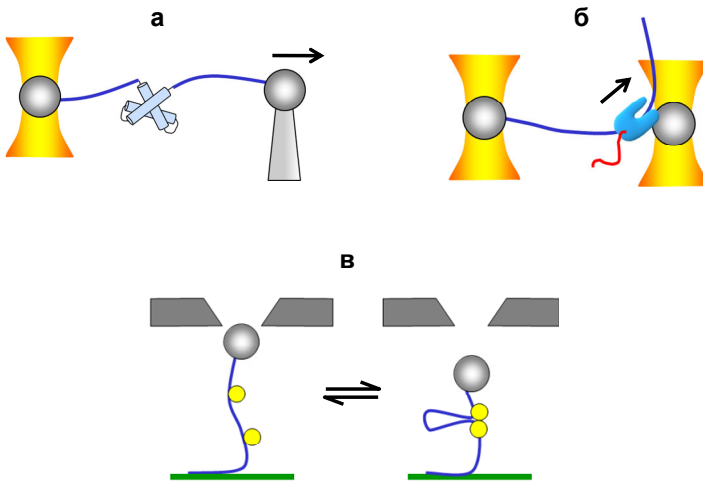


**Рис. 11.41.** Принципова схема експерименту з ДНК за допомогою магнітного пінцета (а) й оптичних пасток (б)

Аналогічно пришивання пластикової кульки до кінця молекули ДНК дозволяє зафіксувати цей кінець за допомогою оптичної пастки і спостерігати за ним за допомогою мікроскопа. Оптична пастка є точкою, куди сфокусовані два лазерні промені: завдяки градієнту електричного поля, який виникає у цій зоні, об'єкт (пластикова кулька), коефіцієнт заломлення якого відрізняється від коефіцієнта заломлення оточуючого середовища, утримується в зоні фокуса з певною силою. Зазвичай один кінець молекули міститься в оптичній пастці, інший – або у ще одній пастці (рис. 11.41, б), або фіксується на поверхні мікропіпетки, яку можна переміщувати, прикладаючи тим самим розтягувальну силу.

Зрозуміло, що молекула ДНК у таких експериментах може бути зв'язана з різноманітними білками, тоді оптична пастка й магнітний пінцет стають ефективними інструментами для дослідження білково-нуклеїнових взаємодій. На молекулі ДНК можуть працювати РНК-полімераза або інші ферменти, за роботою яких можна спостерігати в реальному часі. Дві молекули ДНК можуть бути використані як лінкери, між якими вшитий деякий білок. Можна реалізувати денатурацію такого білка розтягувальною си-

лою і, скинувши цю силу, досліджувати процес білкового фолдингу. Деякі приклади застосування методів маніпуляцій з індивідуальними молекулами представлено на рис. 11.42: за допомогою підходів такого типу отримано цінну інформацію про жорсткість ДНК щодо розтягування і кручення, фізичні механізми денатурації / ренатурації білків, структурні особливості деяких білково-нуклеїнових комплексів, конформаційні перебудови нуклеосом, механізми ініціації, елонгації і термінації транскрипції тощо.



**Рис. 11.42.** Приклади використання оптичної пастки і магнітного пінцета:

- а – молекула білка вбудована між двома ДНК-лінкерами, розтягування за допомогою мікропіпетки (по стрілці) дозволяє здійснити механічну денатурацію білка, яка буде супроводжуватись зростанням відстані між двома кульками;
- б – дві кульки в оптичних пастках: одна на кінці молекули ДНК, на іншій іміобілізована РНК-полімераза, рух якої (по стрілці) супроводжується зростанням відстані між двома кульками;
- в – якщо два зв'язані з ДНК білка взаємодіють між собою, то формується петля, що супроводжується вкороченням відстані між кінцями молекули; за процесом утворення / руйнування петлі можна спостерігати за допомогою магнітного пінцета

## Біоінформатика і молекулярне моделювання

Молекулярна біологія – експериментальна наука, але важливу допоміжну роль у ній виконують різноманітні розрахункові методи – і як засоби аналізу експериментальних даних, і як самостійні дослідницькі підходи. Сьогодні існує можливість на доповнення до результатів, що отримують *in vivo* та *in vitro*, розв'язувати різноманітні проблеми за допомогою комп'ютера – *in silico*.

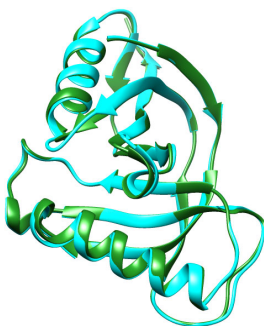
Передусім тут слід згадати **біоінформатику**, що спрямована насамперед на збереження, пошук і аналіз нуклеотидних і амінокислотних послідовностей за допомогою комп'ютерів. Біоінформатика нині є невід'ємною складовою частиною молекулярної біології. Без її методів неможливо, наприклад, установити повну нуклеотидну послідовність геному, оскільки саме за допомогою комп'ютерних програм знаходять порядок розташування у геномі окремих відсеквенованих фрагментів ДНК. На основі біоінформатичного аналізу встановлюють функціональні ролі та структуру генів, аналізують організацію геномів, повногеномні профілі епігенетичних маркерів і профілі транскрипційної активності, мережі регуляції транскрипції, філогенетичні відносини, структурно-функціональні особливості білків, білок-білкові взаємодії тощо.

Найбільші загальнодоступні в інтернеті бази даних нуклеотидних і амінокислотних послідовностей розташовано на порталах NCBI (National Center of Biotechnology Information, США, <https://www.ncbi.nlm.nih.gov/>) і Європейського інституту біоінформатики при Європейській лабораторії молекулярної біології (EMBL, <https://www.ebi.ac.uk/>). Крім власне баз даних, оздоблених інструментами пошуку, на цих та інших порталах можна знайти програмні засоби порівняння послідовностей, пошуку змістовних частин генів, промоторів, стартових точок транскрипції, інтронів та екзонів, структурних мотивів, сайтів зв'язування з лігандами й сайтів посттрансляційних модифікацій у білках тощо. Серед інших біоінформатичних ресурсів слід згадати портал швейцарського Інституту біоінформатики (<https://www.expasy.org/>), базу даних UniProt (<https://www.uniprot.org/>), що інтегрує різнома-

нітну інформацію про білки, бази даних просторових структур білків (Protein Data Bank, <https://www.rcsb.org/>) і нуклеїнових кислот (Nucleic Acid Database, <http://ndbserver.rutgers.edu/>).

Важлива роль належить також **молекулярному моделюванню *in silico***. Таке моделювання допомагає, зокрема, зрозуміти можливі механізми взаємодії як між макромолекулами, так і макромолекул із різноманітними низькомолекулярними лігандами. Пошук найкращої (з мінімальною енергією) структури комплексів – **молекулярний докінг (docking)** – є не тільки дослідницьким інструментом, а й необхідним кроком у сучасному процесі створення нових лікарських препаратів.

Найбільш амбітним завданням у сфері молекулярного моделювання тривалий час було розроблення методів **передбачення просторової структури білків** виходячи з їхньої амінокислотної послідовності. Останнім часом у цій галузі було досягнуто суттєвого успіху – створено алгоритм AlphaFold на основі штучного інтелекту, який дозволяє здійснювати таке передбачення з високою точністю (рис. 11.43). Командою AlphaFold (яка працює в рамках усім відомої компанії "Google") уже здійснено передбачення структур великої кількості білків протеому людини і ще 20 організмів – ці структури розміщено у спеціальній базі даних (AlphaFold Protein Structure Database, <https://alphafold.ebi.ac.uk/>).



**Рис. 11.43.** Приклад передбачення структури білка за допомогою AlphaFold: накладання експериментально встановленої структури інгібітору ДНК-гірази *E. coli* (1JYN, зелений колір) і результату передбачення (блакитний колір, структура з AlphaFold Protein Structure Database, <https://alphafold.ebi.ac.uk/>)

Ще одним важливим доповненням до експериментальних методів вивчення структури макромолекул є комп'ютерні розрахунки в межах 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 або експериментальними методами, або шляхом передбачення. Знаючи силу, яка діє на кожен атом (на основі міжатомних відстаней та енергії нековалентних взаємодій різних типів), можна визначити прискорення кожного атома в системі – отримати систему диференціальних рівнянь руху. Наступне інтегрування цих рівнянь дозволяє обчислити залежності траєкторій і швидкостей руху атомів від часу. Отже, розрахунки молекулярної динаміки дають детальну інформацію про флуктуації і конформаційні перетворення білків і нуклеїнових кислот. Застосування методу лімітується потужністю комп'ютерів: оскільки їхня швидкість постійно зростає, то метод молекулярної динаміки набуває все більшого поширення.

### **Контрольні запитання**

1. Поясніть принцип і охарактеризуйте основні типи гелелектрофорезу.
2. Що таке блот-гібридація? Яка різниця між саузерн-, нозерн- і вестерн-блотингом?
3. Як аналізують активність геному за допомогою ДНК-мікроареїв?
4. Що таке футпринтинг ДНК?
5. У чому полягає процес клонування ДНК? Яким вимогам має відповідати плазмідний вектор для клонування і чому?
6. Що таке рестриктази і як їх використовують у рекомбінантних технологіях?
7. Охарактеризуйте особливості клонування за допомогою бактеріофага  $\lambda$  і космід.
8. Що таке геномна бібліотека? Як створюються такі бібліотеки?

9. Дайте визначення кДНК. Як створюють бібліотеку клонів кДНК?
10. Опишіть основні принципи полімеразної ланцюгової реакції.
11. Як здійснюють фінгерпринтинг ДНК?
12. Як оцінюють відносні кількості ДНК за допомогою ПЛР у реальному часі? Як детектується ДНК у реакційній суміші?
13. Як здійснюють спрямований мутагенез за допомогою ПЛР?
14. Опишіть принципи секвенування ДНК за Сангером.
15. Опишіть схему ампліфікації і секвенування ДНК методами піросеквенування та іонного потоку.
16. Поясніть принцип ампліфікації і секвенування ДНК методом Illumina.
17. У чому полягає принцип секвенування ДНК за допомогою нанопор?
18. Як і з якою метою здійснюють імунопреципітацію хроматину?
19. Назвіть приклади методів, для яких одночасне секвенування великої кількості фрагментів ДНК є останнім кроком.
20. Опишіть принципову схему синтезу рекомбінантних білків у бактеріальних клітинах.
21. Охарактеризуйте систему CRISPR-Cas і принципи її використання для редагування геномів.
22. Опишіть особливості різних варіантів застосування CRISPR-Cas.
23. Охарактеризуйте методи встановлення просторової структури макромолекул.
24. Опишіть принципи атомно-силової мікроскопії й оптичної мікроскопії високої роздільної здатності.
25. Що таке флуоресцентні білки? Як їх використовують?
26. Охарактеризуйте методи оптичної спектроскопії.
27. У чому полягає роль біоінформатики й молекулярного моделювання як дослідницьких підходів молекулярної біології?

## Рекомендована література

Algar, W.R., Hildebrandt, N., Vogel, S.S. and Medintz, I.L. (2019) FRET as a biomolecular research tool – understanding its potential while avoiding pitfalls. *Nat. Methods* 16, 815–829.

Anzalone, A.V., Randolph, P.B., Davis, J.R., et al. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157.

Bai, L., Santangelo, T.J. and Wang, M.D. (2006) Single-molecule analysis of RNA polymerase transcription. *Annu. Rev. Biophys. Biomol. Struct.* 35, 343–360.

Bayguinov, P.O., Oakley, D.M., Shih, C.C., et al. (2018) Modern laser scanning confocal microscopy. *Curr. Protoc. Cytom.* 85, e39.

Bier, F.F., von Nickisch-Roseneck, M., Ehrentreich-Förster, E., et al. (2008) DNA microarrays. *Adv. Biochem. Eng. Biotechnol.* 109, 433–453.

Bustamante, C., Alexander, L., Maciuba, K. and Kaiser, C.M. (2020) Single-molecule studies of protein folding with optical tweezers. *Annu. Rev. Biochem.* 89, 443–470.

Bustamante, C.J., Chemla, Y.R., Liu, S. and Wang, M.D. (2021) Optical tweezers in single-molecule biophysics. *Nat. Rev. Methods Primers* 1, 25.

Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical chemistry*. New York : W.H. Freeman and Company.

Carson, S., Miller H., Srougi, M. and Witherow, D.S. (2019) *Molecular biology techniques*. 4<sup>th</sup> ed. Cambridge : Academic Press.

Cheng, Y. (2018) Single-particle cryo-EM – how did it get here and where will it go. *Science* 361, 876–880.

Chudakov, D.M., Matz, M.V., Lukyanov, S. and Lukyanov, K.A. (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* 90, 1103–1163.

Cooke, R.M. and Cambell, I.D. (1988) Protein structure determination by nuclear magnetic resonance. *Bioessays* 8, 52–56.

Darzacq, X., Yao, J., Larson, D.R., et al. (2009) Imaging transcription in living cells. *Annu. Rev. Biophys.* 38, 173–196.

Das, P.M., Ramachandran, K., van Wert, J. and Singal, R. (2004) Chromatin immunoprecipitation assay. *Biotechniques* 37, 961–969.

Deamer, D., Akeson, M. and Branton, D. (2016) Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524.

De Vlaminck, I. and Dekker, C. (2012) Recent advances in magnetic tweezers. *Annu. Rev. Biophys.* 41, 453–472.

Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096.

Eid, J., Fehr, A., Gray, J., et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138.

Feng, H., Wang, X., Xu, Z., et al. (2018) Super-resolution fluorescence microscopy for single cell imaging. *Adv. Exp. Med. Biol.* 1068, 59–71.

Gerstein, M., Lan, N. and Jansen, R. (2002) Proteomics: integrating interactomes. *Science* 295, 284–287.

Herbert, K.M., Greenleaf, W.J. and Block, S.M. (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* 77, 149–176.

Hsieh, T.S., Fudenberg, G., Goloborodko, A. and Rando, O.J. (2016) Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* 13, 1009–1011.

Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1263–1278.

Ingolia, N.T., Hussmann, J.A. and Weissman, J.S. (2019) Ribosome profiling: global views of translation. *Cold Spring Harb. Perspect. Biol.* 11, a032698.

Ishikawa-Ankerhold, H.C., Ankerhold R. and Drummen, G.P. (2012) Advanced fluorescence microscopy techniques – FRAP, FLIP, FLAP, FRET and FLIM. *Molecules* 17, 4047–4132.

Joo, C., Balci, H., Ishitsuka, Y., et al. (2008) Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* 77, 51–76.

Koch, M.D. and Shaevitz, J.W. (2017) Introduction to optical tweezers. *Methods Mol. Biol.* 1486, 3–24.

Kodzius, R., Kojima, M., Nishiyori, H., et al. (2006) CAGE: cap analysis of gene expression. *Nat. Methods* 3, 211–222.

Koonin, E.V., Makarova, K.S. and Zhang, F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* 37, 67–78.

Kuhlman, B. and Bradley, P. (2019) Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell. Biol.* 20, 681–697.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.

Liu, T.L., Upadhyayula, S., Milkie, D.E., et al. (2018) Observing the cell in its native state: imaging subcellular dynamics in multicellular organisms. *Science* 360, eaq1392.

Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402.

Metzker, M.L. (2009) Sequencing technologies – the next generation. *Nat. Rev. Genet.* 11, 31–46.

Moffitt, J.R., Chelma, Y.R., Smith, S.B. and Bustamante, C. (2008) Recent advances in optical tweezers. *Annu. Rev. Biochem.* 77, 205–228.

Nakane, T., Kotecha, A., Sente, A., et al. (2020) Single-particle cryo-EM at atomic resolution. *Nature* 587, 152–156.

*PCR technology: principle and applications for DNA amplification* (1992) H. Erlich, ed. New York : W.H. Freeman and Company.

Powell, J. (2000) SAGE. The serial analysis of gene expression. *Methods Mol. Biol.* 99, 297–319.

Rao, S.S., Huntley, M.H., Durand, N.C., et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.

Rees, H.A. and Liu, D.R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (2012) *Molecular cloning: a laboratory manual*. 4<sup>th</sup> ed. Cold Spring Harbor : Cold Spring Harbor Laboratory Press.

Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355.

Senior, A.W., Evans, R., Jumper, J., et al. (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710.

Shi, Y. (2014) A glimpse of structural biology through X-ray crystallography. *Cell* 159, 995–1014.

Tam, J. and Merino, D. (2015) Stochastic optical reconstruction microscopy (STORM) in comparison with stimulated emission depletion (STED) and other imaging methods. *J. Neurochem.* 135, 643–658.

Taylor, S.C., Nadeau, K., Abbasi, M., et al. (2019) The ultimate qPCR experiment: producing publication quality, reproducible data the first time. *Trends Biotechnol.* 37, 761–774.

Torrison, M., Pollastri, G. and Le, Q. (2020) Deep learning methods in protein structure prediction. *Comput. Struct. Biotechnol. J.* 18, 1301–1310.

Tugarinov, V., Hwang, P.M. and Kay, L.E. (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu. Rev. Biochem.* 73, 107–146.

Uchihashi, T. and Ganser, C. (2020) Recent advances in bioimaging with high-speed atomic force microscopy. *Biophys. Rev.* 12, 363–369.

Vlachakis, D., Bencurova, E., Papangelopoulos, N. and Kossida, S. (2014) Current state-of-the-art molecular dynamics methods and applications. *Adv. Protein Chem. Struct. Biol.* 94, 269–313.

Yip, K.M., Fischer, N., Paknia, E., et al. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature* 587, 157–161.

Zeymer, C. and Hilvert, D. (2018) Directed evolution of protein catalysts. *Annu. Rev. Biochem.* 87, 131–157.

Zhang, Z., Wu, S., Stenoi, D.L. and Paša-Tolić, L. (2014) High-throughput proteomics. *Annu. Rev. Anal. Chem.* 7, 427–754.

## ПІСЛЯМОВА

Єдиним ефективним шляхом розвитку науки є поступовий рух від вивчення відносно простих об'єктів до більш складних, і молекулярна біологія не є винятком із цього загального правила. Умовно часом її народження як науки можна вважати 1953 рік, коли Уотсон і Крік побудували модель структури подвійної спіралі ДНК. З'ясування принципів структурної організації і функціонування біологічних макромолекул було й залишається основною метою молекулярної біології, однак поступове ускладнення об'єктів дослідження перетворило її на системну дисципліну, у центрі уваги якої перебувають не тільки окремі біополімери, а й загальні сукупності біологічних макромолекул у клітині й цілісному організмі.

Сьогодні системна молекулярна біологія спрямована на вивчення геномів, транскриптомів, протеомів і загальної картини взаємодій між біологічними макромолекулами у масштабі цілої клітини. При цьому досліджуються і тонкі механізми функціонування окремих ланок таких складних систем – молекулярна біологія стала основним підходом, що дозволяє шукати відповіді на фундаментальні питання щодо сутності феномену життя.

Молекулярна біологія є також чудовим прикладом того, як фундаментальні дисципліни врешті набувають прикладного значення. По-перше, її методи стали підґрунтям для розвитку різноманітних біотехнологій – промислового виробництва рекомбінантних білків, генетичної модифікації сільськогосподарських культур, отримання нових матеріалів, виготовлення біосенсорів тощо. По-друге, молекулярна біологія стала основою для розвитку нових підходів у медицині. Зокрема, саме розвиток структурної молекулярної біології зумовив появу нових можливостей розроблення специфічних інгібіторів активних центрів білків із метою створення нових ліків (drug design).

Найважливіше досягнення останнього часу – розроблення методу точного передбачення білкових структур AlphaFold – може розглядатись як ключова умова ефективного розвитку на цьому шляху. Дизайн ліків, конструювання вакцин, молекулярна діагностика захворювань – усе це можна об'єднати під назвою *молекулярна медицина*.

Перспективним напрямом молекулярної медицини є генна терапія спадкових захворювань. Основна її ідея полягає у лікуванні патологій за допомогою уведення у клітини відповідних генетичних конструкцій. Одним із найважливіших досягнень у цій сфері стало розроблення підходів до редагування геномів на основі технології CRISPR-Cas. Зрозуміло, що тут виникає велика кількість проблем, однак сьогодні на стадії розроблення перебувають кілька сотень проєктів генної терапії.

Важливим напрямом розвитку молекулярної медицини є також розроблення методів доставки препаратів спрямованої дії до цільових клітин. Для цього використовують певні вірусні системи й різноманітні наночастинки. Прогрес у розумінні принципів функціонування біологічних молекулярних машин зумовив народження *біонанотехнології*, яка зараз робить лише перші кроки. Створення нанороботів, які б могли лікувати хвороби безпосередньо в місці, де розвиваються патологічні процеси, і репарувати пошкодження, що виникли в організмі, ще не є реальністю, хоча вже й не здається абсолютно фантастичним. Живі клітини мають тисячі працюючих наномашин, здатних перебудовувати матерію на атомному рівні: мотори молекулярного розміру, сенсори, транспортні й інформаційні пристрої та інші корисні механізми. Завданням біонанотехнології є розвиток у бажаних напрямках властивостей таких машин шляхом їхньої модифікації і конструювання нових аналогів. Це дозволить створювати нові матеріали з точно контрольованою структурою на атомарному рівні, наночіпи й комп'ютерні пристрої мінімальних розмірів, прецизійні сенсорні системи. Біонанотехнології будуть насамперед застосовуватися в медицині, адже біонаномашини найкраще працюють у живих клітинах. Молекулярні комплекси, що здатні знаходити злякисні клітини або здійснювати локаль-

ну лікувальну дію, нині вже існують. Діагностичні біонаносенсиори, молекулярні машини для виконання терапевтичних завдань, інші медичні застосування біонанотехнології – справа недалекого майбутнього.

За трохи більше ніж півстоліття молекулярна біологія стала однією із ключових галузей фундаментального природознавства й основою майже всіх біологічних наук. Саме на молекулярно-біологічних підходах базуються сучасна біологія клітини, фізіологія, таксономія, антропологія тощо. Швидкий розвиток молекулярних біотехнологій, народження на наших очах біонанотехнологій і молекулярної медицини роблять молекулярну біологію однією з найважливіших прикладних дисциплін.

## ЗМІСТ

<b>Від автора</b> .....	3
<b>Вступ</b> .....	4
<b>Розділ 1. Фізико-хімічні основи молекулярної біології</b> .....	9
Енергетика взаємодії між атомами й молекулами.....	9
Вільна енергія.....	9
Рівновага утворення комплексів молекул у розчині.....	13
Ковалентний зв'язок між атомами в молекулах.....	16
Властивості ковалентного зв'язку.....	16
Механізм гнучкості полімерного ланцюга .....	18
Перебудови системи ковалентних зв'язків у хімічних реакціях .....	19
Нековалентні взаємодії між атомами й молекулами.....	21
Вандерваальсові взаємодії.....	21
Електростатичні (іонні) взаємодії .....	23
Водневий зв'язок .....	25
Гідрофобні взаємодії.....	28
<b>Розділ 2. Структура і принципи функціонування білків</b> .....	33
Хімічна будова білків .....	34
Амінокислоти .....	34
Пептидний зв'язок і поліпептидний ланцюг.....	37
Вторинна структура.....	39
$\alpha$ -Спіраль.....	39
$\beta$ -Структура .....	41
Спіраль $3_{10}$ .....	42
$\beta$ -Поворот.....	42

Глобулярна структура .....	42
Роль вторинної структури в утворенні глобули .....	44
Стабілізація глобули .....	44
Структурна класифікація глобулярних білків .....	46
Типи неглобулярних білків .....	53
Фібрилярні білки .....	53
Мембранні білки .....	54
Неструктуровані білки .....	55
Конформаційна рухливість білків .....	56
Принципи функціонування білків .....	58
Механізм ферментативного каталізу .....	59
Механізми використання енергії гідролізу АТР .....	62
<b>Розділ 3. ДНК .....</b>	<b>70</b>
Хімічна будова нуклеїнових кислот .....	70
Нуклеотиди .....	70
Полінуклеотидний ланцюг .....	74
Нуклеази .....	75
Подвійна спіраль .....	76
Стабілізація подвійної спіралі .....	76
Конформаційні параметри подвійної спіралі .....	80
Структурні форми ДНК .....	82
Конформація В-ДНК .....	85
Білково-нуклеїнові взаємодії .....	87
Структурна класифікація білків, які взаємодіють із ДНК .....	87
Принципи білково-нуклеїнового впізнання .....	95
Циркулярна ДНК .....	100
<b>Розділ 4. Організація ДНК у клітинах:</b>	
геноми та структура хроматину .....	113
Організація геномів .....	113
Генетичний код .....	113
Гени .....	116
Геноми .....	117

Молекулярна організація хроматину .....	127
Нуклеосома .....	128
Посттрансляційні модифікації гістонових хвостів.....	133
Гістон H1 і компактизація хроматинової фібрили .....	137
Петельні й топологічно асоційовані домени хроматину.....	141
Компартменти хроматину .....	143
Організація хроматину в масштабі клітинного ядра .....	144
Мітотична хромосома.....	145
<b>Розділ 5. Транскрипція .....</b>	<b>151</b>
Загальний сценарій процесу транскрипції .....	151
Бактеріальна РНК-полімераза .....	154
Бактеріальний промотор та ініціація транскрипції.....	158
Елонгація транскрипції .....	161
Термінація транскрипції у бактерій .....	167
Еукаріотичні РНК-полімерази .....	170
РНК-полімераза II .....	171
Ініціація транскрипції РНК-полімеразою II .....	173
Промотор РНК-полімерази II.....	173
Загальні фактори транскрипції РНК-полімерази II і преініціаторний комплекс .....	175
Процес ініціації транскрипції .....	181
Елонгація РНК-полімерази через хроматин.....	183
РНК-полімерази I і III .....	187
Ініціація транскрипції генів рибосомної РНК .....	187
Ініціація транскрипції РНК-полімеразою III .....	188
<b>Розділ 6. Регуляція транскрипції .....</b>	<b>195</b>
Регуляція транскрипції у бактерій .....	196
Лактозний оперон .....	196
Триптофановий оперон: атенюація .....	198
РНК-перемикачі .....	200
Регуляція транскрипції бактеріофага $\lambda$ .....	202

Еукаріотичний транскриптом .....	206
Транскрипційні фактори й регуляторні РНК в еукаріотів .....	210
Транскрипційні фактори.....	210
Регуляторні РНК .....	214
Зовнішня регуляція активності транскрипційних факторів .....	217
Структура хроматину й регуляція транскрипції.....	219
Хімічні модифікації хроматину .....	220
Розташування нуклеосом у промоторах .....	222
Ремоделювання промоторів .....	225
Активація транскрипції на відстані: енхансери.....	231
Конститутивна репресія транскрипції: гетерохроматин .....	235
Субтеломерні зони у <i>S. cerevisiae</i> .....	236
НР1-залежна система репресії .....	238
Метильовання ДНК .....	240
Білки групи полікомб і некодувальні РНК .....	243
Загальна схема регуляції транскрипції в еукаріотів .....	245
<b>Розділ 7. Процесинг еукаріотичних мРНК .....</b>	<b>252</b>
Кепування .....	253
Сплайсинг .....	257
Сплайсосома: механізм сплайсингу .....	258
Поліаденілювання мРНК і термінація транскрипції .....	265
Редагування мРНК.....	269
Альтернативний процесинг .....	270
Транс-сплайсинг .....	275
<b>Розділ 8. Синтез білків.....</b>	<b>279</b>
Транспортні РНК .....	280
Структура тРНК .....	280
Аміноацилювання тРНК.....	284
Рибосома.....	290
Компоненти рибосоми.....	291

Структура рибосоми .....	295
Елонгаційний цикл .....	299
Завантаження аміноацил-тРНК .....	301
Елонгаційний фактор EF1 .....	301
Упізнання кодону .....	304
Акомодація aa-тРНК .....	306
Транспептидація .....	309
Транслокація .....	312
Ініціація трансляції .....	316
Ініціація трансляції у бактерій .....	317
Ініціація трансляції в еукаріотів .....	319
Термінація трансляції .....	322
Регуляція трансляції .....	324
Формування просторової структури білка .....	328
Закономірності укладання білкової глобули .....	329
Котрансляційний білковий фолдинг .....	333
Шаперони Hsp70 .....	334
Система Hsp90 .....	336
Шапероніни .....	337
<b>Розділ 9. Реплікація ДНК .....</b>	<b>344</b>
Реплікони .....	345
Бактеріальні ДНК-полімерази .....	347
Структура і полімеразна активність	
ДНК-полімераз .....	348
Редагування помилок .....	350
Особливості ДНК-полімерази	
порівняно з РНК-полімеразою .....	352
Бактеріальна реплісома та інші елементи	
системи реплікації .....	353
Голофермент ДНК-полімерази .....	353
Праймосома: геліказа та праймаза .....	355
Загальна схема організації	
бактеріальної реплісоми .....	358
Завершення синтезу ланцюга, що запізнюється .....	359

Ініціація і термінація реплікації у бактерій.....	361
Топологічні проблеми, пов'язані з реплікацією.....	364
Особливості еукаріотичної системи реплікації.....	366
Ініціація реплікації в еукаріотів.....	371
Подовження кінців еукаріотичної хромосоми.....	376
Пройходження реплісоми через хроматин.....	378
Репарація ДНК.....	380
Пряма репарація.....	381
Ексцизійна репарація.....	382
Репарація некомплементарних пар основ – місметчів.....	385
Репарація без репарації.....	387
Репарація дволанцюгових розривів.....	389
<b>Розділ 10. Рекомбінація ДНК.....</b>	<b>395</b>
Гомологічна рекомбінація.....	396
Сайт-специфічна рекомбінація.....	404
V(D)J-рекомбінація імуноглобулінових генів.....	408
Транспозиції мобільних елементів.....	413
ДНК-транспозони.....	414
LTR-ретротранспозони.....	417
Мобільні елементи LINE та SINE.....	419
<b>Розділ 11. Методи молекулярної біології.....</b>	<b>424</b>
Розділення і аналіз сумішей нуклеїнових кислот.....	424
Гель-електрофорез.....	424
Блот-гібридизація.....	428
ДНК-чіпи.....	430
Клонування ДНК.....	432
Бібліотеки клонів.....	439
Полімеразна ланцюгова реакція.....	441
Секвенування ДНК.....	448
Метод Сангера.....	448
Методи секвенування другого покоління.....	450
Секвенування індивідуальних молекул.....	456

Аналіз транскриптомів .....	459
Повногеномний аналіз ДНК-білкових взаємодій .....	461
Синтез рекомбінантних білків .....	463
Редагування геномів .....	468
Аналіз протеомів.....	475
Фізичні методи дослідження біологічних макромолекул.....	478
Візуалізація біологічних макромолекул .....	478
Визначення просторової структури макромолекул .....	483
Оптична спектроскопія.....	487
Методи дослідження індивідуальних макромолекул .....	491
Біоінформатика і молекулярне моделювання .....	494
<b>Післямова.....</b>	<b>502</b>

Навчальне видання

СИВОЛОБ Андрій

# МОЛЕКУЛЯРНА БІОЛОГІЯ

Підручник

Редактор *Н. Земляна*

Оригінал-макет виготовлено ВПЦ "Київський університет"



Формат 60x84<sup>1/16</sup>. Ум. друк. арк. 29,76. Наклад 100. Зам. № 223-10735.  
Гарнітура Times New Roman. Папір офсетний. Друк офсетний. Вид. № Б5.  
Підписано до друку 12.10.23

Видавець і виготовлювач  
ВПЦ "Київський університет"  
Б-р Тараса Шевченка, 14, м. Київ, 01601, Україна  
☎ (044) 239 32 22; (044) 239 31 72; тел./факс (044) 239 31 28  
e-mail: vpc@knu.ua; vpc\_div.chief@univ.net.ua; redaktor@univ.net.ua  
http: vpc.knu.ua  
Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02