Chương 1

Các đại phân tử sinh học

I. Nucleic acid

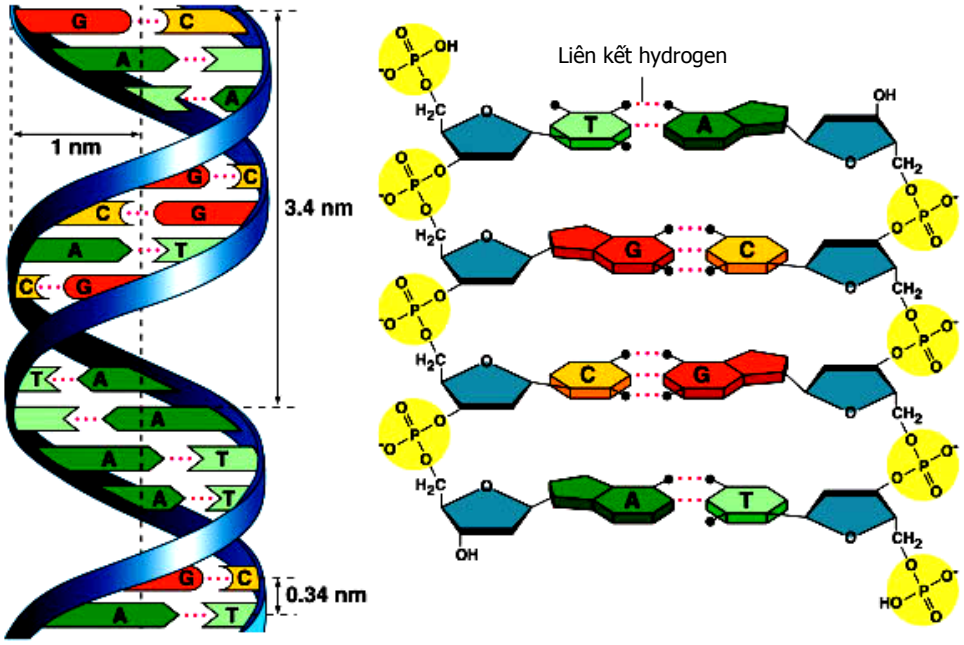
Nucleic acid, vật chất mang thông tin di truyền của các hệ thống sống, là một polymer hình thành từ các monomer là nucleotide. Mỗi nucleotide gồm ba thành phần: nhóm phosphate, đường pentose (đường 5 carbon) và một nitrogen base. Các nitrogen base thuộc hai nhóm: các purine gồm adenine (A) và guanine (G), các pyrimidine gồm thymine (T), cytosine (C) và uracil (U). Các nucleotide được nối với nhau bằng liên kết phosphodiester tạo thành chuỗi dài.

Nucleic acid gồm hai loại phân tử có cấu tạo rất giống nhau là deoxyribonucleic acid (DNA) và ribonucleic acid (RNA).

1. Deoxyribonucleic acid

Phân tử DNA là một chuỗi xoắn kép gồm hai sợi đơn. Mỗi sợi đơn là một chuỗi nucleotide (Hình 1.1). Mỗi nucleotide gồm ba thành phần: nhóm phosphate, đường deoxyribose và một trong bốn base (A, C, G và T) (Hình 1.2). Hai sợi đơn kết hợp với nhau nhờ các liên kết hydrogen hình thành giữa các base bổ sung nằm trên hai sợi: A bổ sung cho T và C bổ sung cho G. Mỗi sợi đơn có một trình tự định hướng với một đầu 5'phosphate tự do, đầu kia là 3' hydroxyl tự do (quy ước là 5'→3'). Hướng của hai sợi đơn trong chuỗi xoắn kép ngược nhau, nên được gọi là hai sợi đối song.

Những phân tích cấu trúc hiện đại đã cho thấy cấu trúc của DNA không phải luôn luôn tương ứng với dạng được gọi là B mà Watson và Crick đã đưa ra. Do sự tác động của các hợp chất có khối lượng nhỏ hoặc protein, dạng B có thể chuyển sang dạng A (nén nhiều hơn) hoặc là dạng Z (xoắn trái). Chúng có thể tự gấp lại hoặc xoắn mạnh, ví dụ một sợi đôi DNA có độ dài là 20 cm được nén trong một chromosome có kích thước là 5 μm.



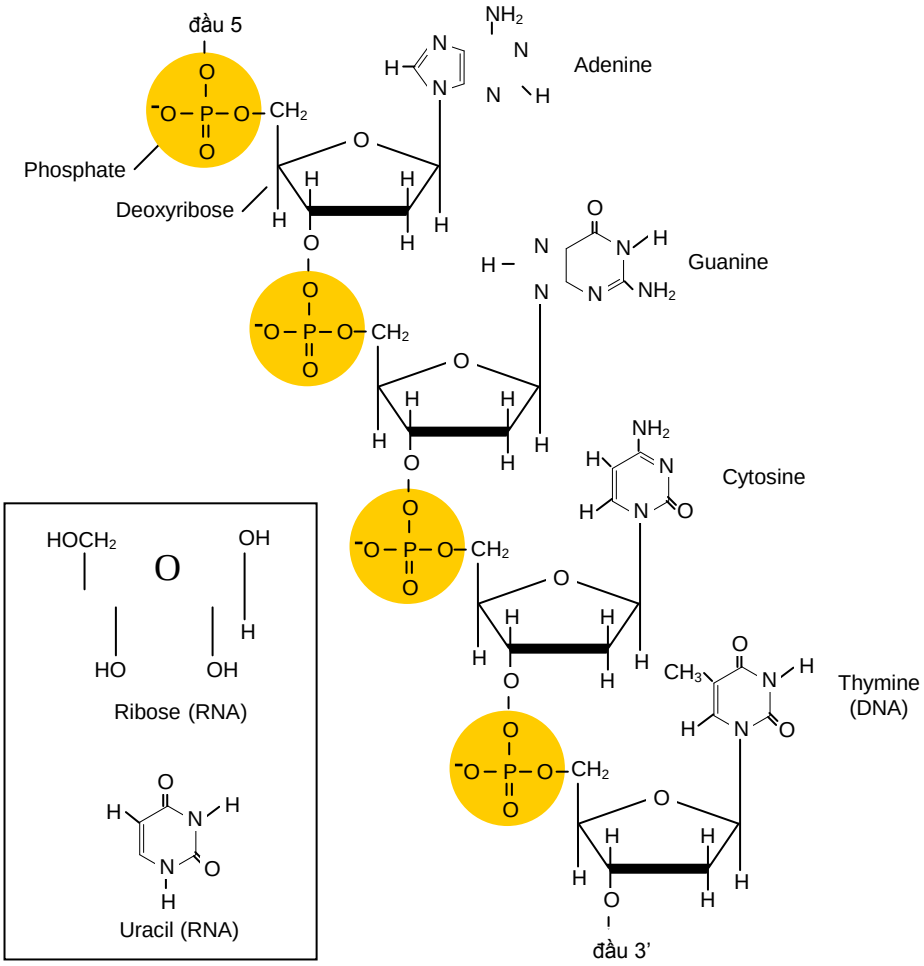
(a) Cấu trúc của DNA

Hình 1.1. Chuỗi xoắn kép của DNA

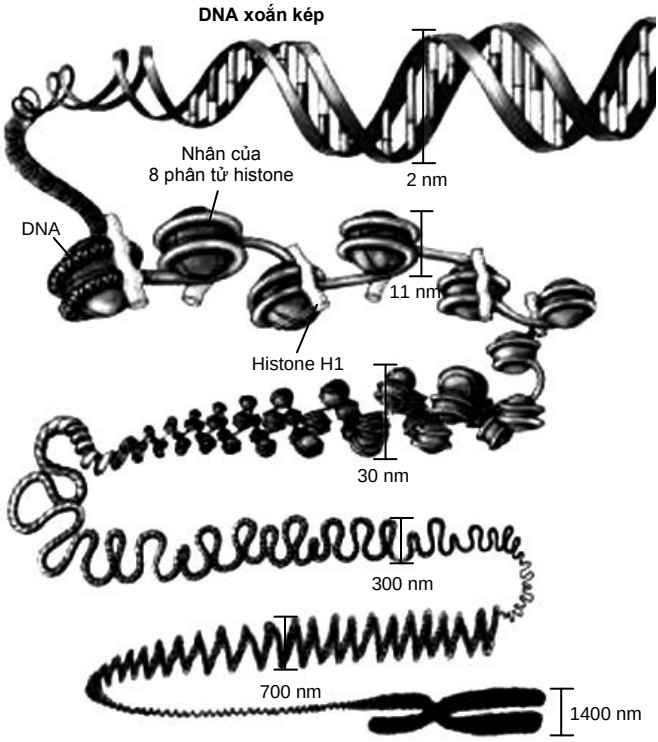
Phân tử DNA trong nhiễm sắc thể của sinh vật eukaryote ở dạng mạch thẳng, còn ở phần lớn tế bào prokaryote (vi khuẩn) phân tử DNA lại có dạng mạch vòng. Tuy nhiên, dù ở dạng nào thì các phân tử DNA này đều tồn tại theo kiểu cuộn chặt. Trong tế bào eukaryote, DNA kết hợp chặt chẽ với các protein là histone.

DNA eukaryote có kích thước rất lớn (Ví dụ: DNA ở người có thể dài đến 1 m) nên vấn đề đặt ra là phân tử này phải được nén như thế nào trong một thể tích rất hạn chế của nhân. Việc nén được thực hiện ở nhiều mức độ, mức độ thấp nhất là nucleosome và mức độ cao nhất là cấu trúc nhiễm sắc chất. Thật vậy, đường kính của chuỗi xoắn DNA chỉ là 20 $\overset{o}{A}$, trong khi sợi nhiễm sắc chất quan sát dưới kính hiển vi điện tử có đường kính 100 $\overset{o}{A}$, đôi khi đạt 300 $\overset{o}{A}$. Điều này chứng tỏ phân tử DNA tham gia hình thành những cấu trúc phức tạp hơn (Hình 1.3).

Sợi nhiễm sắc chất có đường kính 100 $\overset{o}{A}$ là một chuỗi chứa nhiều nucleosome. Đó là những cấu trúc hình thành từ một sợi DNA quấn quanh một lõi gồm 8 phân tử histone (mức độ tổ chức cao nhất của DNA). Sợi có đường kính 100 $\overset{o}{A}$ này có cấu trúc phức tạp hơn sợi có đường kính 300 $\overset{o}{A}$. Trong nhân tế bào, các sợi vừa kể trên kết hợp chặt chẽ với nhiều protein khác nhau và cả với các RNA tạo thành nhiễm sắc chất, mức độ tổ chức cao nhất của DNA.



Hình 1.2. Cấu trúc các nucleotide điển hình



Hình 1.3. Cấu trúc nucleosome và nhiễm sắc thể. Phân tử DNA được cuộn lại trên nhiễm sắc thể làm cho chiều dài ngắn lại hơn 50.000 lần.

Các DNA ở eukaryote có đặc điểm khác với DNA prokaryote. Toàn bộ phân tử DNA prokaryote đều mang thông tin mã hóa cho các protein trong khi đó DNA của eukaryote bao gồm những trình tự mã hóa (các exon) xen kẽ với những trình tự không mã hóa (intron). Các trình tự mã hóa ở eukaryote chìm ngập trong một khối lớn DNA mà cho đến nay vẫn chưa rõ tác dụng được gọi là "DNA rác" (junk DNA). Tùy theo mức độ hiện diện của chúng trong nhân, các trình tự DNA được chia làm ba loại:

- **Các trình tự lặp lại nhiều lần.** Ví dụ: ở động vật có vú các trình tự này chiếm 10-15% genome (hệ gen). Đó là những trình tự DNA ngắn (10-200 kb), không mã hóa, thường tập trung ở những vùng chuyên biệt trên

nhiễm sắc thể như ở vùng tâm động (trình tự CEN) hay ở đầu các nhiễm sắc thể (trình tự TEL). Chức năng của các trình tự này chưa rõ, có thể chúng tham gia vào quá trình di chuyển DNA trên thoi vô sắc (trình tự CEN) hoặc vào quá trình sao chép toàn bộ phần DNA nằm ở đầu mút nhiễm sắc thể (trình tự TEL).

- **Các trình tự có số lần lặp lại trung bình.** Ví dụ: ở genome người các trình tự này chiếm 25-40 %. Chúng đa dạng hơn và có kích thước lớn hơn (100-1.000 kb) các trình tự lặp lại nhiều lần. Các trình tự này phân bố trên toàn bộ genome. Chúng có thể là những trình tự không mã hóa mà cũng có thể là những trình tự mã hóa cho rRNA, tRNA và 5S RNA.

- **Các trình tự duy nhất.** Là các gen mã hóa cho các protein, có trình tự đặc trưng cho từng gen.

Một đặc điểm của phân tử DNA có ý nghĩa rất quan trọng và được sử dụng vào phương pháp lai phân tử, đó là khả năng biến tính và hồi tính. Biến tính là hiện tượng hai sợi đơn của phân tử DNA tách rời nhau khi các liên kết hydrogen giữa các base bổ sung nằm trên hai sợi bị đứt do các tác nhân hóa học (dung dịch kiềm, formamide, urea) hay do tác nhân vật lý (nhiệt). Sau đó, nếu điều chỉnh nhiệt độ và nồng độ muối thích hợp, các sợi đơn có thể bắt cặp trở lại theo nguyên tắc bổ sung, để hình thành phân tử DNA ban đầu, đó là sự hồi tính.

2. Ribonucleic acid

Phân tử RNA có cấu tạo tương tự DNA ngoại trừ ba điểm khác biệt sau:

- Phân tử RNA là chuỗi đơn.

- Đường pentose của phân tử RNA là ribose thay vì deoxyribose.

- Thymine (T), một trong bốn loại base hình thành nên phân tử DNA, được thay thế bằng uracil (U) trong phân tử RNA.

Cấu trúc và chức năng của RNA có sự biến đổi rõ rệt. Về cơ bản RNA chỉ là chất mang thông tin di truyền ở virus, sau đó người ta chứng minh rằng nó không những đóng vai trò cơ bản ở việc chuyển thông tin di truyền mà còn có vai trò cấu trúc khi tạo nên phức hệ RNA-protein.

Theo một lý thuyết tiến hóa mà đại diện là Eigen, RNA là chất mang thông tin di truyền, thành viên trung gian của sự biểu hiện gen, thành phần

cấu tạo và là chất xúc tác. Nhóm OH ở vị trí thứ hai của ribose cần thiết cho đa chức năng làm nhiễu loạn sự tạo thành sợi đôi, qua đó làm tăng độ không bền vững của liên kết phosphodieste.

Trong tế bào có ba loại RNA chính, có các chức năng khác nhau:

2.1. Các RNA thông tin (mRNA)

mRNA là bản sao của những trình tự nhất định trên phân tử DNA, có vai trò trung tâm là chuyển thông tin mã hóa trên phân tử DNA đến bộ máy giải mã thành phân tử protein tương ứng. Các RNA có cấu trúc đa dạng, kích thước nhỏ hơn so với DNA vì chỉ chứa thông tin mã hóa cho một hoặc vài protein và chỉ chiếm khoảng 2-5% tổng số RNA trong tế bào.

Quá trình chuyển thông tin được thể hiện như sau:

Ở *E. coli*, kích thước trung bình của một phân tử mRNA khoảng 1,2 kb.

2.2. RNA vận chuyển (tRNA)

tRNA làm nhiệm vụ vận chuyển các amino acid hoạt hóa đến ribosome để tổng hợp protein từ các mRNA tương ứng. Có ít nhất một loại tRNA cho một loại amino acid. tRNA vận chuyển chứa khoảng 75 nucleotide (có khối lượng khoảng 25 kDa), là phân tử RNA nhỏ nhất. Các tRNA có cấu trúc dạng cỏ ba lá. Cấu trúc này được ổn định nhờ các liên kết bổ sung hiện diện ở nhiều vùng của phân tử tRNA. Hai vị trí không có liên kết bổ sung đóng vai trò đặc biệt quan trọng đối với chức năng của tRNA:

- Trình tự anticodon gồm ba nucleotide.
- Trình tự CCA, có khả năng liên kết cộng hóa trị với một amino acid đặc trưng.

2.3. RNA ribosome (rRNA)

rRNA là thành phần cơ bản của ribosome, đóng vai trò xúc tác và cấu trúc trong tổng hợp protein.

Tùy theo hệ số lắng rRNA được chia thành nhiều loại: ở eukaryote có 28S; 18S; 5,8S và 5S rRNA; còn các rRNA ở *E. coli* có ba loại: 23S, 16S và 5S.

rRNA chiếm nhiều nhất trong ba loại RNA (80% tổng số RNA tế bào), tiếp đến là tRNA khoảng 16% và mRNA chỉ khoảng 2%. Ngoài ra, tế bào sinh vật eukaryote còn chứa những phân tử RNA kích thước nhỏ của nhân (small nuclear, snRNA) chiếm khoảng <1% tham gia vào ghép nối các exon. Ribosome là những phân tử cần thiết cho sự tổng hợp protein, ribosome của mọi tế bào đều gồm một tiểu đơn vị nhỏ và một tiểu đơn vị lớn. Mỗi tiểu đơn vị có mang nhiều protein và rRNA (trong đó rRNA là thành phần chủ yếu chiếm khoảng 65%) có kích thước khác nhau. Người ta cũng thấy ribosome trong ty thể, ở đó có sự tổng hợp một số protein ty thể.

Bảng 1.1. Các phân tử RNA trong *E. coli*

Loại	Tổng số tương đối (%)	Hệ số lắng (S)[1]	Khối lượng phân tử (kDa)	Số lượng nucleotide
rRNA	80	$\begin{cases}23\\16\\5\end{cases}$	$\begin{cases}1{,}2\times10^{3}\\0{,}55\times10^{3}\\3{,}6\times10^{1}\end{cases}$	$\begin{cases}3700\\1700\\120\end{cases}$
tRNA	15	4	$2{,}5\times10^{1}$	75
mRNA	5		Không đồng nhất	

2.3.1. Ribosome của prokaryote

Tế bào được nghiên cứu về ribosome nhiều nhất là *E. coli.* Ribosome (70S) của *E. coli* gồm hai tiểu đơn vị: tiểu đơn vị nhỏ (30S) và tiểu đơn vị

[1] S (Svedberg): đơn vị đo vận tốc lắng. Hệ số lắng của một tiểu đơn vị phụ thuộc không những vào khối lượng của tiểu đơn vị đó mà còn phụ thuộc vào hình dạng và độ rắn của nó, điều này giải thích tại sao sự kết hợp của hai tiểu đơn vị 50S và 30S lại tạo ra một ribosome 70S.

lớn (50S). Căn cứ vào hệ số lắng, người ta phân biệt ba loại rRNA: 23S rRNA, 16S rRNA và 5S rRNA.

- Tiểu đơn vị 30S chứa: 1 phân tử 16S rRNA (có 1540 nu) và 21 ribosomal protein khác nhau.

- Tiểu đơn vị 50S chứa: 1 phân tử 5S rRNA (có 120 nu), 1 phân tử 23S rRNA (có 2900 nu) và 34 ribosomal protein.

Hai tiểu đơn vị nhỏ và lớn khi kết hợp với nhau sẽ tạo ra một rãnh ở chỗ tiếp giáp của chúng để cho mRNA đi qua.

2.3.2. Ribosome của eukaryote

Ribosome của eukaryote (80S) lớn hơn ribosome của prokaryote cũng bao gồm hai tiểu đơn vị: tiểu đơn vị nhỏ (40S) và tiểu đơn vị lớn (60S).

- Tiểu đơn vị 40S chứa: 1 phân tử 18S rRNA (có 1900 nu) và 33 ribosomal protein.

- Tiểu đơn vị 60S chứa: 3 phân tử rRNA (5S; 5,8S và 28S) và 49 ribosomal protein.

Tóm lại, tất cả RNA trong tế bào đều được tổng hợp nhờ enzyme RNA polymerase. Enzyme này đòi hỏi những thành phần sau đây:

- Một khuôn mẫu, thường là DNA sợi đôi.

- Tiền chất hoạt hóa: Bốn loại ribonucleoside triphosphate: ATP, GTP, UTP và CTP.

Sinh tổng hợp RNA giống DNA ở một số điểm, thứ nhất hướng tổng hợp là 5'→3', thứ hai là cơ chế kéo dài giống nhau: nhóm 3'-OH ở đầu cuối của chuỗi tổng hợp là vị trí gắn kết của nucleoside triphosphate tiếp theo. Thứ ba, sự tổng hợp xảy ra do thủy phân pyrophosphate.

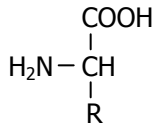
Tuy nhiên, khác với DNA là RNA không đòi hỏi mồi (primer). Ngoài ra, RNA polymerase không có hoạt tính nuclease để sửa chữa khi các nucleotide bị gắn nhầm.

Cả ba loại RNA trong tế bào được tổng hợp trong *E. coli* nhờ một loại RNA polymerase. Ở động vật có vú, các RNA khác nhau được tổng hợp bằng các loại RNA polymerase khác nhau.

II. Protein

1. Cấu trúc của protein

Amino acid là đơn vị cơ sở (monomer) cấu thành protein. Tất cả 20 amino acid có mặt trong protein đều được xây dựng theo một kiểu mẫu chung:



Công thức tổng quát của L-α-amino acid

Trong đó, gốc R (mạch bên) cũng là phần khác duy nhất giữa 20 loại amino acid, quy định tính chất của từng loại.

Nhóm amine (NH_2) đính ở nguyên tử C_2, theo tên cũ là nguyên tử C_α. Vì vậy, người ta gọi là nhóm α-amine. Các amino acid tồn tại chủ yếu trong tự nhiên có nhóm amine đứng ở bên trái trục, được gọi là amino acid dạng L. Dạng D-amino acid chỉ tồn tại riêng biệt, ví dụ trong thành tế bào vi khuẩn.

Các amino acid riêng biệt có những đặc tính khác nhau là do gốc R của chúng. Những amino acid trung tính có một nhóm amine và một nhóm carboxyl. Những protein chứa nhiều amino acid trung tính là những protein trung tính. Khi chiều dài gốc R tăng sẽ hình thành đặc tính kỵ nước. Những protein có chứa nhiều amino acid như valine, leucine, isoleucine có tính chất đặc trưng là kỵ nước. Những amino acid có tính acid trong phần gốc có một nhóm carboxyl. Protein chứa nhiều amino acid có tính acid là những protein acid. Tương tự như vậy đối với protein chủ yếu được hình thành bởi những amino acid có tính kiềm là những protein kiềm. Phần gốc R của amino acid có ý nghĩa quyết định đối với đặc tính của protein mà chúng tạo nên. Điều này không những có ý nghĩa đối với tính chất hóa học mà cả cấu trúc của protein.

Thủy phân hoàn toàn protein, thu được chủ yếu các L-α-amino acid. Mặc dù protein rất đa dạng nhưng hầu hết chúng đều được cấu tạo từ 20 L-α-amino acid. Dựa vào đặc tính của gốc R, amino acid được chia làm bảy nhóm chính sau đây:

- **Amino acid trung tính mạch thẳng.** Bao gồm glycine, alanine, valine, leucine và isoleucine.

- **Các hydroxyl amino acid mạch thẳng.** Bao gồm serine và threonine.

- **Amino acid chứa lưu huỳnh mạch thẳng.** Bao gồm cysteine và methionine. Khi oxy hóa hai nhóm -SH của hai phân tử cysteine tạo thành cystine có chứa cầu (-S-S-).

- **Các amino acid acid và các amide.** Bao gồm aspartic acid và glutamic acid. Trong phân tử của chúng có một nhóm amine và hai nhóm carboxyl. Ở độ pH sinh lý (6-7), các amino acid này tích điện âm. Amine hóa nhóm carboxyl ở mạch bên của aspartate và glutamate tạo thành các amide tương ứng là asparagine và glutamine.

- **Các amino acid kiềm**. Bao gồm lysine và arginine.

- **Iminoacid.** Proline.

- **Các amino acid thơm và dị vòng.** Bao gồm phenylalanine, tyrosine và tryptophan. Do có chứa vòng thơm nên các amino acid này có một số phản ứng đặc trưng.

Các amino acid được nối với nhau bởi các liên kết peptide, liên kết này được hình thành do sự kết hợp nhóm amine của một amino acid với nhóm carboxyl của amino acid kế tiếp. Phản ứng kết hợp giải phóng một phân tử H_2O.

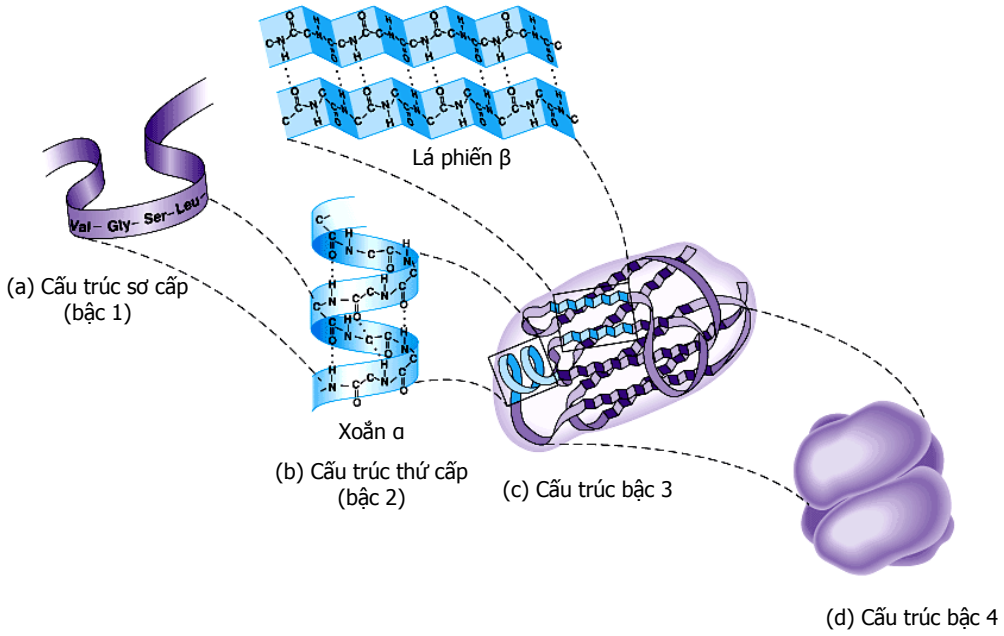
Peptide là một chuỗi nối tiếp nhiều amino acid (số lượng ít hơn 30). Với số lượng amino acid lớn hơn chuỗi được gọi là polypeptide. Mỗi polypeptide có hai đầu tận cùng, một đầu mang nhóm amine tự do, đầu kia mang nhóm carboxyl tự do. Protein được dùng để chỉ đơn vị chức năng, nghĩa là một cấu trúc phức tạp trong không gian chứ không phải đơn thuần là một trình tự amino acid.

Chuỗi polypeptide có thể uốn thành cấu trúc hình gậy như trong các protein hình sợi hay cấu trúc khối cầu như trong các protein dạng cầu hay một cấu trúc gồm cả hai dạng trên. Một protein có thể được hình thành từ nhiều chuỗi polypeptide.

Người ta thường phân biệt cấu trúc của phân tử protein thành bốn bậc như sau (Hình 1.4):

- **Cấu trúc bậc 1.** Là trình tự sắp xếp các gốc amino acid trong chuỗi polypeptide. Cấu trúc này được giữ vững nhờ liên kết peptide (liên kết cộng hóa trị).

Vì mỗi một amino acid có gốc khác nhau, các gốc này có những đặc tính hóa học khác nhau, nên một chuỗi polypeptide ở các thời điểm khác nhau có những đặc tính hóa học rất khác nhau. Tuy nhiên, về tổng quát thì tất cả các chuỗi polypeptide được xây dựng một cách có hệ thống từ các nhóm nguyên tử CO, CH và NH. Việc xây dựng có hệ thống này là cơ sở để tạo nên cấu trúc bậc hai.



Hình 1.4. Các mức độ tổ chức của phân tử protein

- **Cấu trúc bậc 2.** Là tương tác không gian giữa các gốc amino acid ở gần nhau trong chuỗi polypeptide. Cấu trúc được bền vững chủ yếu nhờ liên kết hydrogen hình thành giữa các liên kết peptide ở kề gần nhau, cách nhau những khoảng xác định.

Cấu trúc bậc 2 của phân tử protein: xoắn α (α-helix), lá phiến β và xoắn collagen. Loại α-helix là sợi ở dạng xoắn ốc, cuộn xung quanh một trục, mỗi vòng xoắn có 3,6 gốc amino acid.

Những sợi collagen chạy song song tạo nên những bó sợi dai của gân. Collagen cũng có trong xương và trong các mô nối. Elastin là một protein, gồm những sợi protein tương đối ngắn, gắn kết với nhau nhờ liên kết cộng hóa trị. Những chuỗi polypeptide quay theo dạng xoắn ốc, tự duỗi xoắn khi có áp lực.

- Cấu trúc bậc 3. Là tương tác không gian giữa các gốc amino acid ở xa nhau trong chuỗi polypeptide, là dạng cuộn lại trong không gian của toàn chuỗi polypeptide.

Nhiều chuỗi polypeptide trong cơ thể sống tồn tại không phải ở dạng thẳng mà gấp khúc và qua đó tạo nên cấu trúc không gian ba chiều. Tuy nhiên, cấu trúc này hoàn toàn xác định, chủ yếu là do trình tự các amino acid và môi trường. Khi một chuỗi polypeptide tách ra khỏi ribosome sau khi tổng hợp và được đưa vào trong tế bào chất như là môi trường tạo hình thì nó sẽ hình thành nên cấu trúc tự nhiên rất nhanh, đặc biệt đối với cấu trúc hình cầu, đem lại cho protein những đặc tính sinh lý quan trọng. Có thể do chuyển động nhiệt của các chuỗi polypeptide mà các nhóm của các gốc amino acid tiếp xúc với nhau, dẫn đến có thể kết hợp với nhau. Trong nhiều protein hình cầu có chứa các gốc cysteine, sự tạo thành các liên kết disulfite giữa các gốc cysteine ở xa nhau trong chuỗi polypeptide sẽ làm cho chuỗi bị cuộn lại đáng kể. Các liên kết khác, như liên kết Van der Waals, liên kết tĩnh điện, phân cực, kỵ nước và hydrogen giữa các mạch bên của các gốc amino acid đều tham gia làm bền vững cấu trúc bậc 3. Cấu trúc hình cầu của protein được gọi là cấu trúc bậc ba, đó chính là cấu trúc của enzyme.

- Cấu trúc bậc 4. Là tương tác không gian giữa các chuỗi của các phân tử protein gồm hai hay nhiều chuỗi polypeptide hình cầu. Mỗi chuỗi polypeptide này được gọi là một tiểu đơn vị (subunit). Sự kết hợp giữa các phân tử này lỏng lẻo và chủ yếu là do liên kết hydrogen và kỵ nước. Bằng cách này hai phân tử xác định có thể kết hợp với nhau tạo thành một dimer. Chẳng hạn: hemoglobin được tạo nên từ hai chuỗi α với mỗi chuỗi có 141 gốc amino acid và hai chuỗi β với mỗi chuỗi là 146 gốc amino acid.

Cấu trúc của một hoặc nhiều chuỗi polypeptide có ý nghĩa quan trọng đối với độ hòa tan và chức năng của chúng. Cấu trúc protein được hiểu là sự

sắp xếp của những chuỗi riêng lẻ hoặc nhiều chuỗi. Chúng phụ thuộc nhiều vào độ pH của môi trường. Protein và chuỗi polypeptide hòa tan tốt khi những nhóm ưa nước hướng ra phía ngoài, nhóm kỵ nước hướng vào bên trong. Khi một protein thay đổi cấu trúc thì những nhóm kỵ nước quay ra ngoài, protein mất khả năng hòa tan trong nước, ví dụ trường hợp kết tủa không ở dạng tinh thể của protein sữa trong môi trường chua. Lactic acid được sản sinh do vi khuẩn làm giảm pH sữa, làm thay đổi protein sữa. Nhiều nhóm kỵ nước được hướng ra bên ngoài, protein mất khả năng tan trong nước. Vì vậy, việc thường xuyên duy trì giá trị pH trong tế bào chất rất quan trọng, vì chỉ có như vậy chức năng hoạt động của các enzyme trong tế bào chất mới được đảm bảo.

2. Chức năng của protein

Mỗi một hoạt động trong tế bào phụ thuộc vào một hoặc nhiều phân tử protein đặc hiệu. Một trong các cách phân loại protein là dựa vào chức năng sinh học của chúng. Bảng 1.2 tóm tắt sự phân loại protein theo chức năng và đưa ra một số ví dụ đại diện cho mỗi loại.

Bảng 1.2. C**ác chức năng sinh học của protein và một số ví dụ**

Các nhóm chức năng	Ví dụ
Enzyme	Ribonuclease Trypsin Phosphofructokinase Alcohol dehydrogenase Catalase Malic enzyme
Protein điều khiển	Insulin Somatotropin Thyrotropin *lac* repressor NF1 (nuclear factor 1) Catabolite activator protein (CAP) AP1

Protein vận chuyển	Hemoglobin Serum albumin Glucose transporter
Protein dự trữ	Ovalbumin Casein Zein Phaseolin Ferritin
Protein vận động và co rút	Actin Myosin Tubulin Dynelin Kinesin
Protein cấu trúc	α-Keratin Collagen Elastin Fibroin Proteoglycans
Protein cấu trúc tạm thời (scaffold protein)	Grb 2 crk shc stat IRS-1
Protein bảo vệ	Immunoglobulins Thrombin Fibrinogen Antifreeze proteins Snake and bee venom proteins Diphtheria toxin Ricin
Protein lạ/ngoại lai (exotic protein)	Monellin Resilin Glue proteins

2.1. Chức năng enzyme

Phần lớn protein là enzyme. Hiện nay, có hơn 3.000 loại enzyme đã được biết. Enzyme là chất xúc tác sinh học có vai trò làm tăng tốc độ phản ứng. Mỗi một bước trong trao đổi chất đều được xúc tác bởi enzyme. Enzyme có thể làm tăng tốc độ phản ứng lên 10^{16} lần so với tốc độ phản ứng không xúc tác. Sự kết hợp giữa enzyme và cơ chất xảy ra ở vị trí hoạt động của enzyme.

2.2. Protein điều khiển

Một số protein không thực hiện bất kỳ sự biến đổi hóa học nào, tuy nhiên nó điều khiển các protein khác thực hiện chức năng sinh học, chẳng hạn insulin điều khiển nồng độ đường glucose trong máu. Đó là một protein nhỏ (5,7 kDa), gồm hai chuỗi polypeptide nối với nhau bằng các liên kết disulfite. Khi không đủ insulin thì sự tiếp nhận đường trong tế bào bị hạn chế. Vì vậy, mức đường trong máu tăng và dẫn đến sự thải đường mạnh mẽ qua nước tiểu (bệnh tiểu đường).

Một nhóm protein khác tham gia vào sự điều khiển biểu hiện gen. Những protein này có đặc tính là gắn vào những trình tự DNA hoặc để hoạt hóa hoặc ức chế sự phiên mã thông tin di truyền sang mRNA, ví dụ chất ức chế (repressor) đình chỉ sự phiên mã.

2.3. Protein vận chuyển

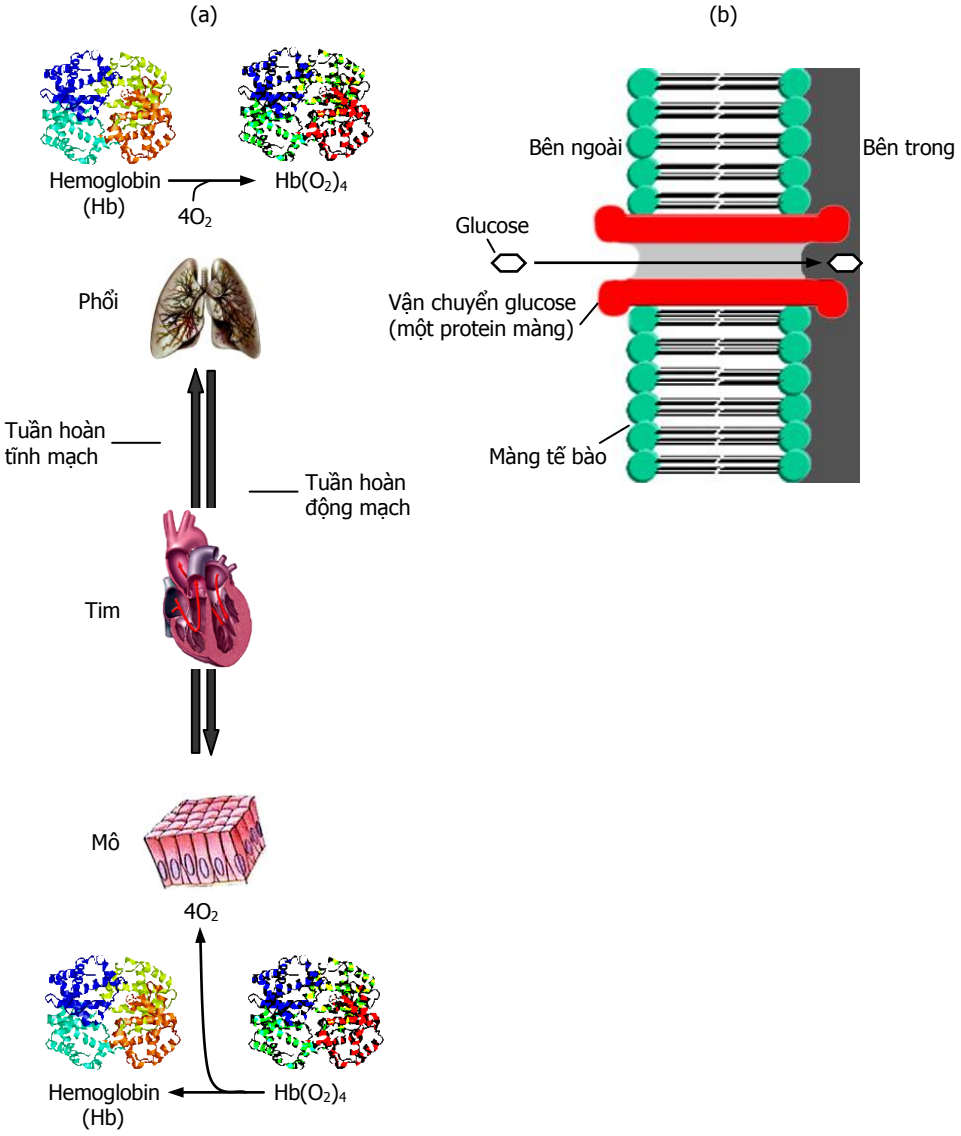
Làm nhiệm vụ vận chuyển chất đặc hiệu từ vị trí này sang vị trí khác, ví dụ vận chuyển O_2 từ phổi đến các mô do hemoglobin hoặc vận chuyển acid béo từ mô dự trữ đến các cơ quan khác nhờ protein trong máu là serum albumin.

Các chất được vận chuyển qua màng được thực hiện bằng các protein đặc hiệu, chẳng hạn vận chuyển glucose hoặc các amino acid qua màng (Hình 1.5).

2.4. Protein dự trữ

Các protein là nguồn cung cấp các chất cần thiết được gọi là protein dự trữ. Protein là polymer của các amino acid và nitrogen thường là yếu tố

hạn chế cho sinh trưởng, nên cơ thể phải có protein dự trữ để cung cấp đầy đủ nitrogen khi cần. Chẳng hạn, ovalbumin là protein dự trữ trong lòng trắng trứng cung cấp đủ nitrogen cho phôi phát triển. Casein là protein sữa cung cấp nitrogen cho động vật có vú còn non. Hạt ở thực vật bậc cao cũng chứa một lượng protein dự trữ lớn (khoảng 60%), cung cấp đủ nitrogen cho quá trình nảy mầm của hạt.



Hình 1.5. Hai kiểu vận chuyển cơ bản. (a): vận chuyển bên trong hoặc giữa các tế bào hoặc mô. (b): vận chuyển vào hoặc ra khỏi tế bào.

Protein cũng có thể dự trữ các chất khác ngoài thành phần amino acid (N, C, H, O và S), ví dụ ferritin là protein tìm thấy trong mô động vật kết hợp với Fe. Một phân tử ferritin (460 kDa) gắn với 4.500 nguyên tử Fe (chiếm 35% khối lượng). Protein có vai trò giữ lại kim loại Fe cần thiết cho sự tổng hợp những protein có chứa Fe quan trọng như hemoglobin.

2.5. Protein vận động và co rút

Một số protein mang lại cho tế bào khả năng vận động, tế bào phân chia và co cơ. Các protein này có đặc điểm: chúng ở dạng sợi hoặc dạng polymer hóa để tạo sợi, chẳng hạn actin và myosin. Tubulin là thành phần cơ bản của thoi vô sắc (sợi xuất hiện khi phân chia các nhiễm sắc thể về các cực).

2.6. Protein cấu trúc

Có chức năng tạo độ chắc và bảo vệ tế bào và mô. Chẳng hạn: α-keratin là protein không tan, cấu tạo nên tóc, sừng và móng. Collagen là protein hình sợi có trong xương. Ở động vật collagen chiếm 1/3 protein tổng số. Fibroin (β-keratin) là thành phần cơ bản của kén tằm.

Một chức năng phổ biến khác của protein là cấu tạo nên màng sinh học.

2.7. Protein bảo vệ

Trong việc giải độc các kim loại nặng, phytochelatin có một ý nghĩa quan trọng, đây là những polypeptide đơn giản có nguồn gốc từ glutation và có công thức chung như sau:

$$(\gamma\text{-glutamyl-cysteinyl})_n\text{-glycine}$$

Do có nhiều nhóm SH nên chúng có khả năng kết hợp chặt với các kim loại nặng, làm cho những kim loại nặng này không thể gây rối loạn trao đổi chất. Sự tổng hợp phytochelatin được kích thích bởi những kim loại nặng như Cd, Cu, Ag, Bi và Au.

Protein bảo vệ có vai trò quan trọng trong các phản ứng miễn dịch. Động vật có xương sống có một cơ chế phức tạp và phát triển cao để ngăn ngừa những tác nhân vi sinh vật gây bệnh. Chức năng này có liên quan đến

đặc tính của chuỗi polypeptide. Khi một protein lạ (có nguồn gốc virus, vi khuẩn hoặc nấm) xâm nhập vào máu hoặc vào mô thì phản ứng tự vệ của cơ thể sẽ xuất hiện rất nhanh. Protein lạ được gọi là kháng nguyên (antigen) chứa một vùng có trật tự xác định các nguyên tử có thể kết hợp với tế bào lympho và kích thích tế bào này sản sinh kháng thể. Những tế bào lympho tồn tại trong hệ thống miễn dịch với số lượng 10^9 và trên bề mặt của nó có những vùng nhận biết nơi mà kháng nguyên sẽ được kết hợp (Hình 1.6). Những vùng nhận biết này rất khác nhau và đặc hiệu cho từng loại kháng nguyên. Trong cơ thể luôn có sẵn một lượng lớn các tế bào lympho khác nhau và chúng có thể tổng hợp rất nhanh các kháng thể đặc hiệu khi kháng nguyên xuất hiện. Mỗi loại kháng thể có một vị trí kết hợp duy nhất đặc trưng với kháng nguyên. Khả năng bảo vệ của hệ miễn dịch đã làm cho protein lạ của tác nhân gây bệnh trở thành vô hại. Những kháng thể này được gọi là globulin miễn dịch. Chúng chiếm khoảng 20% protein tổng số trong máu.

Một nhóm protein bảo vệ khác là protein làm đông máu thrombin và fibrinogen, ngăn cản sự mất máu của cơ thể khi bị thương.

Cá ở các vùng cực của Trái đất có protein chống đông (antifreeze protein) có tác dụng bảo vệ máu khi nhiệt độ xuống dưới 0^oC.

2.8. Protein lạ/ngoại lai

Ví dụ monellin là một loại protein được tìm thấy ở một loại cây ở châu Phi, được coi là chất ngọt nhân tạo cho con người.

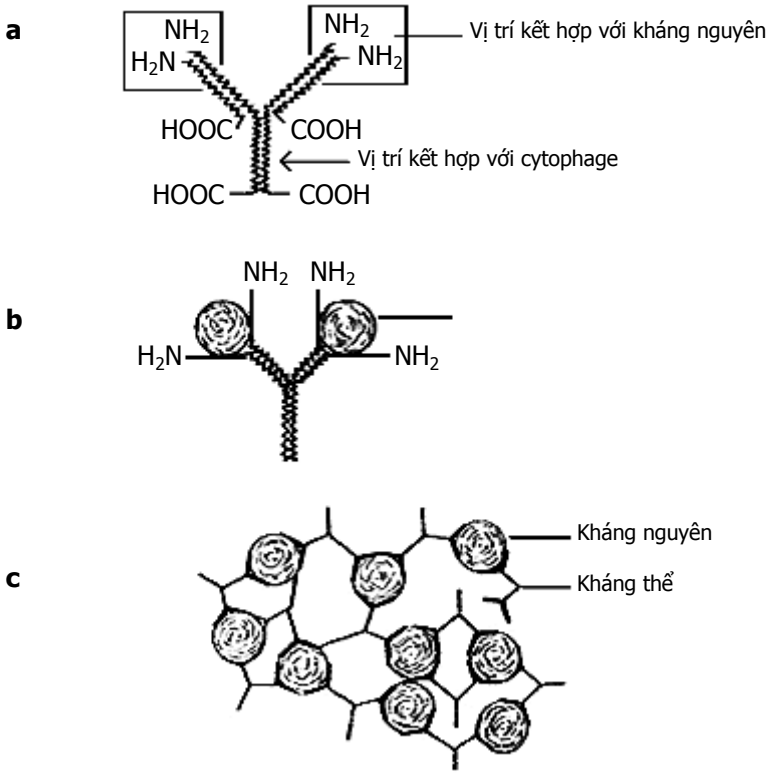
Ở một số sinh vật biển như họ Trai tiết ra loại protein keo (glue protein), cho phép nó gắn chặt lên bề mặt.

III. Lipid

Mặc dù không mang hoạt tính sinh học cao như protein nhưng lipid cũng đóng một vai trò đặc biệt trong hệ thống sống. Chúng là nhân tố chính tạo nên các màng sinh học mà nếu thiếu thì mọi hoạt động của protein sẽ không thể phối hợp nhịp nhàng.

Đơn vị cấu trúc của lipid là các acid béo. Mỗi acid béo được cấu tạo từ một mạch carbohydrate (gồm các nguyên tử C và H) gắn với một nhóm carboxyl có tính acid. Các acid béo khác nhau bởi độ dài của chúng, bởi số

lượng và vị trí các liên kết đôi. Các acid béo không có liên kết đôi được gọi là các acid béo bão hòa, các acid béo không bão hòa có ít nhất một liên kết đôi.



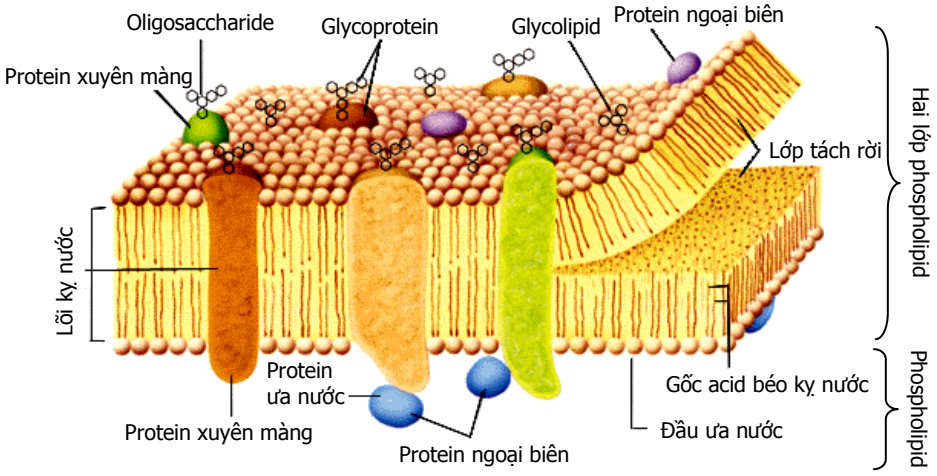
Hình 1.6. Sơ đồ biểu diễn của kháng thể và kháng nguyên. a: kháng thể gồm 4 chuỗi polypeptide. b: kháng thể kết hợp với kháng nguyên. c: kết hợp giữa kháng nguyên và kháng thể.

Màng sinh học có chức năng là giới hạn những vùng trao đổi chất và tham gia vào việc vận chuyển các chất. Màng sinh học cũng có khả năng chuyển đi những tín hiệu. Protein màng cũng có thể là các enzyme. Chức năng này được thể hiện ở màng trong của ty thể và lạp thể. Màng sinh học bao gồm lớp kép lipid với những protein phân bố ở trong đó (Hình 1.7)

Các lipid màng được hình thành từ một chuỗi dài acid béo nối với những nhóm có đặc tính ưa nước cao và được gọi là những phân tử lưỡng cực vì một đầu tương tác với nước, còn đầu kia thì kỵ nước.

Bảng 1.3. Cấu trúc một số acid béo tiêu biểu trong hệ thống sống

Công thức hóa học	Tên thường dùng
Acid béo bão hòa	
$CH_3(CH_2)_{10}$ COOH	Lauric
$CH_3(CH_2)_{14}$ COOH	Palmitic
Acid béo không bão hòa	
$CH_3(CH_2)_7CH=CH\ (CH_2)_7$ COOH	Oleic
$CH_3(CH_2)_4$ CH= CH- CH_2 CH= CH $(CH_2)_7$ COOH	Linoleic



Hình 1.7. Sơ đồ biểu diễn một đoạn cắt của màng sinh học

IV. Polysaccharide

Các polysaccharide có nhiều chức năng quan trọng trong tế bào, chúng tham gia vào cấu tạo tế bào và là nguồn dự trữ năng lượng chủ yếu. Các polysaccharide được hình thành từ nhiều monomer, là các đường đơn giản (monosaccharide) nối với nhau bằng liên kết glycoside. Liên kết này được hình thành từ sự kết hợp giữa C_1 của một phân tử đường với nhóm hydroxyl của phân tử kế tiếp.

Nguồn dự trữ tinh bột ở các tế bào động vật là glycogen, trong khi đó ở thực vật là tinh bột. Một polymer khác của glucose là cellulose thì tạo nên thành tế bào thực vật và là hợp chất hữu cơ hiện dịn nhiều nhất trong sinh quyển.

Chúng ta vừa điểm qua riêng rẽ từng thành phần cấu tạo tế bào chính. Trong thực tế, hoạt động của chúng phối hợp mật thiết với nhau. Các nucleic acid trong tế bào thường kết hợp chặt chẽ với các protein tạo thành nucleoprotein. DNA của tế bào eukaryote thì được bọc bởi những protein đặc hiệu là các histone. Màng tế bào cũng không phải chỉ có phospholipid, chính các protein gắn trong màng đã tạo ra những đặc trưng riêng của màng sinh học. Một điểm cần lưu ý là nếu như cấu trúc và các tính chất hóa lý của nucleic acid, lipid và polysaccharide tương đối đồng nhất thì các protein lại hết sức đa dạng cả về cấu trúc và chức năng. Một phân tử protein thường bao gồm nhiều vùng mang những đặc tính khác nhau: vùng ưa nước hay kỵ nước, vùng gắn một đường, vùng có hoạt tính xúc tác, vùng liên kết với nucleic acid hay với một protein khác. Từ mỗi chức năng của tế bào, từ sự hình thành vật chất mang thông tin di truyền, truyền đạt thông tin di truyền, sự chuyển hóa năng lượng, sự liên lạc giữa các tế bào... đều có sự tham gia của các protein. Điều kỳ diệu của sự sống là toàn bộ các hoạt động vô cùng đa dạng ấy được thực hiện bởi một phân tử duy nhất.

Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. *NXB Giáo dục,* Hà Nội.

2. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

3. Lewin B. 2000. Gene VII. *Oxford University Press,* Oxford, UK.

4. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *WH Freeman and Company*, New York, USA.

5. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Loscik R. 2004. Molecular Biology of the Gene. *The Benjamin Cummings/Cold Spring Habor Laboratory Press*, San Francisco, CA, USA.

6. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company Inc.* New York, USA.

Chương 2

Cấu trúc genome

Genome (hệ gen, bộ gen) là thuật ngữ được dùng với các nghĩa khác nhau như sau:

- Nguyên liệu di truyền của một cơ thể: 1) nhiễm sắc thể trong tế bào vi khuẩn (hoặc một trong mỗi loại nhiễm sắc thể nếu hơn một loại có mặt, ví dụ: các nhiễm sắc thể lớn hoặc bé của *Vibrio cholerae*), 2) DNA hoặc RNA trong một virion, 3) nhiễm sắc thể cùng với mọi plasmid được kết hợp (ví dụ: nhiễm sắc thể và hai plasmid nhỏ trong vi khuẩn *Buchnera*).

- Tất cả các gen (khác nhau) trong tế bào hoặc virion.

- Bộ nhiễm sắc thể đơn bội hoặc genome đơn bội trong tế bào.

Chuỗi genome hoàn chỉnh (nghĩa là trình tự hoàn chỉnh của các nucleotide trong genome) đã được công bố cho một số loài vi khuẩn. Các trình tự khác cũng đã được công bố, ví dụ genome của cây cúc dại (*Arabidopsis thaliana*) và genome người.

Genome chứa toàn bộ thông tin di truyền và các chương trình cần thiết cho cơ thể hoạt động. Ở các sinh vật nhân thật (eukaryote), 99% genome nằm trong nhân tế bào và phần còn lại nằm trong một số cơ quan tử như ty thể và lạp thể. Đa số genome vi khuẩn và phần genome chứa trong các cơ quan tử thường có kích thước nhỏ và ở dạng vòng khép kín. Ngược lại, phần genome trong nhân thường rất lớn và phân bố trên các nhiễm sắc thể dạng thẳng.

Dự án genome là dự án xác định cấu trúc di truyền chính xác của một genome cơ thể sống, nghĩa là trình tự DNA của tất cả các gen của nó. Dự án genome của một số sinh vật mô hình (model organisms) đã được hoàn thành như sau:

- Các genome vi khuẩn. Các trình tự hoàn chỉnh của genome *Escherichia coli* đã được xác định theo phương thức tổ hợp/tập hợp (consortium) của các phòng thí nghiệm. Năm 1995, hai trình tự genome hoàn chỉnh của vi khuẩn *Haemophilus influenzae* và *Mycoplasma genitalium* cũng được hoàn thành. Loài *M. genitalium* có một genome đơn

giản (khoảng 580.067 base), do nó dựa vào vật chủ để vận hành nhiều bộ máy trao đổi chất của mình. Loài *H. influenzae* là một vi khuẩn đặc trưng hơn, và có genome khoảng 1.830.121 base với 1.749 gen.

- Chuỗi genome hoàn chỉnh của nấm men *Saccharomyces cerevisiae* đã được hoàn chỉnh trong năm 1996, nhờ một consortium của các phòng thí nghiệm. Genome của chúng dài 12.146.000 base.

- Các dự án genome ở động vật như: chuột, cừu, lợn, giun tròn (*Caenorhabditis elegans*), ruồi giấm (*Drosophila melanogaster*)…, hoặc ở thực vật như: lúa nước, lúa mì, ngô, táo, cúc dại…, mà nổi bật nhất trong số đó là dự án genome người cũng đã được thực hiện.

Ngày 12. 2. 2001 genome người đã được công bố với khoảng 30.000 gen, ít hơn nhiều so với dự kiến trước đây (hàng trăm ngàn gen), và chỉ gấp hai lần giun tròn hoặc ruồi giấm. Người ta đã xác định hệ gen người giống 98% so với tinh tinh và có đến 99% là giống nhau giữa các dân tộc, các cá thể. Do đó, vấn đề hình thành và phát triển nhân cách, chỉ số thông minh... phải chủ yếu trên cơ sở xã hội và sự rèn luyện của từng người để phát triển tiềm năng sinh học của bản thân.

Trình tự genome của những sinh vật mô hình rất có ý nghĩa trong những nghiên cứu của một chuyên ngành khoa học mới đó là genome học (genomics). Dựa vào đây, các nhà sinh học phân tử có thể phân tích cấu trúc, hoạt động và chức năng của các gen, làm sáng tỏ được vai trò của DNA lặp lại, DNA không chứa mã di truyền, DNA nằm giữa các gen... Điều đặc biệt có ý nghĩa là khi so sánh các genome với nhau, có thể hiểu được hoạt động của genome trong các cơ thể sống, mối quan hệ giữa chúng, sự đa dạng sinh học và mức độ tiến hóa.

Kết quả bước đầu so sánh genome giữa các loài sinh vật với nhau đã cho thấy có ba đặc điểm nổi bật: 1) các gen phân bố trong genome không theo qui luật, 2) kích thước của genome thay đổi không tỷ lệ thuận (tương quan) với tính phức tạp của loài, 3) số lượng nhiễm sắc thể cũng rất khác nhau ngay giữa những loài rất gần nhau.

I. Thành phần và đặc điểm của genome

Genome chứa mọi thông tin di truyền đặc trưng cho từng loài, thậm chí cho từng cá thể trong loài. Genome có thể bao gồm các phân tử DNA

hoặc RNA. Đối với sinh vật bậc cao, kích thước genome thay đổi từ 10^9 bp (động vật có vú) đến 10^{11} bp (thực vật). Khác với tế bào tiền nhân (prokaryote), các gen trong genome của eukaryote thường tồn tại nhiều bản sao và thường bị gián đoạn bởi các đoạn mã mù không mang thông tin di truyền (các intron). Vì vậy, một trong những vấn đề đang được quan tâm là cần phải biết số lượng các gen khác nhau có mặt trong genome cũng như số lượng các gen hoạt động trong từng loại mô, từng giai đoạn phát triển và tỷ lệ các gen so với kích thước genome...

1. Genome của cơ quan tử

Hầu hết genome của cơ quan tử, nhưng không phải luôn luôn, có dạng phân tử DNA mạch vòng đơn của một chuỗi duy nhất.

Genome của cơ quan tử mã hóa cho một số, không phải tất cả, các protein được tìm thấy trong cơ quan tử. Do có nhiều cơ quan tử trong một tế bào, cho nên có nhiều genome của cơ quan tử trên một tế bào. Mặc dù bản thân genome của cơ quan tử là duy nhất. Nhưng nó cấu tạo gồm một chuỗi lặp lại[1] liên quan với mọi chuỗi không lặp lại[2] của nhân. Về nguyên tắc, các gen cơ quan tử được phiên mã và dịch mã bởi các cơ quan tử.

1.1. Genome của ty thể

DNA ty thể (mitochondrial DNA-mtDNA) là một genome độc lập, thường là mạch vòng, được định vị trong ty thể.

- DNA ty thể của tế bào động vật mã hóa đặc trưng cho 13 protein, 2 rRNA và 22 tRNA.

- DNA ty thể của nấm men *S. cerevisiae* dài hơn mtDNA của tế bào động vật năm lần do sự có mặt của các đoạn intron dài.

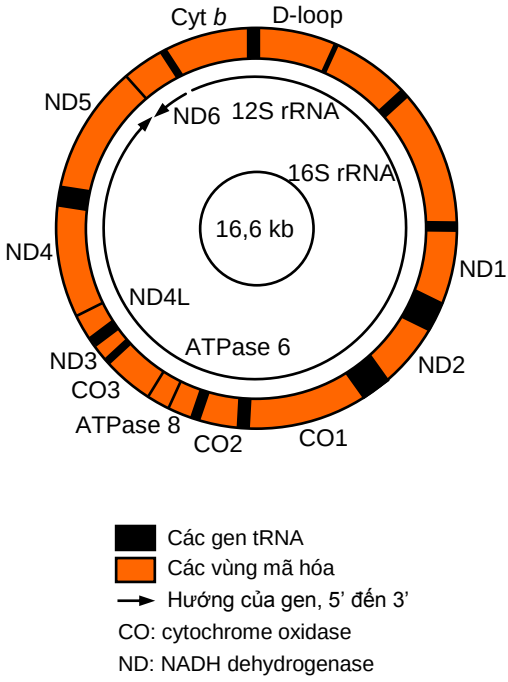
Các genome ty thể có kích thước tổng số rất khác nhau, các tế bào động vật có kích thước genome nhỏ (khoảng 16,5 kb ở động vật có vú) (Hình 2.1). Có khoảng một vài trăm ty thể trên một tế bào. Mỗi ty thể có

[1] DNA lặp lại mô tả các chuỗi hiện diện hơn một bản sao trong mỗi genome.
[2] DNA không lặp lại chứa các chuỗi duy nhất: chỉ có một bản sao trong genome đơn bội.

nhiều bản sao DNA. Số lượng tổng số của DNA ty thể so với DNA nhân là rất nhỏ (<1%).

Trong nấm men *S. cerevisiae*, genome ty thể có kích thước khá lớn (khoảng 80 kb) và khác nhau tùy thuộc vào từng chủng. Có khoảng 22 ty thể trên một tế bào, tương ứng khoảng 4 genome trên một cơ quan tử. Ở những tế bào sinh trưởng, tỷ lệ mtDNA có thể cao hơn (khoảng 18%).

Kích thước của genome ty thể ở các loài thực vật là rất khác nhau, tối thiểu khoảng 100 kb. Kích thước lớn của genome đã gây khó khăn cho việc phân lập nguyên vẹn DNA, nhưng bản đồ cắt hạn chế (restriction map) trong một vài loài thực vật đã cho thấy genome ty thể thường là một chuỗi đơn, được cấu tạo như một mạch vòng. Trong mạch vòng này có những chuỗi tương đồng ngắn và sự tái tổ hợp giữa chúng đã sinh ra các phân tử tiểu genome (subgenome) mạch vòng nhỏ hơn, cùng tồn tại với genome "chủ" (master genome) hoàn chỉnh, đã giải thích cho sự phức tạp của các DNA ty thể ở thực vật.



Hình 2.1. DNA ty thể của người. Bao gồm 22 gen tRNA, 2 gen rRNA, và 13 vùng mã hóa protein.

Bảng 2.1 tóm tắt sự phân công của các gen trong một số genome ty thể. Tổng số gen mã hóa protein là khá ít, và không tương quan với kích thước của genome. Ty thể động vật có vú sử dụng các genome 16 kb của chúng để mã hóa cho 13 protein, trong khi đó ty thể nấm men *S. cerevisiae* dùng các genome từ 60-80 kb mã hóa cho khoảng 8 protein. Thực vật với genome ty thể lớn hơn nhiều mã hóa cho nhiều protein hơn. Các intron được tìm thấy trong hầu hết các genome của ty thể, nhưng lại không có trong các genome rất nhỏ của động vật có vú.

Hai rRNA chính luôn được mã hóa bởi genome ty thể. Số lượng các tRNA được mã hóa bởi genome ty thể dao động từ không cho đến đầy đủ (25-26 trong ty thể). Nhiều protein ribosome được mã hóa trong genome ty thể của thực vật và sinh vật nguyên sinh, nhưng chỉ có một ít hoặc không có trong genome của nấm và động vật.

Bảng 2.1. Các genome ty thể có các gen mã hóa cho các protein, rRNA và tRNA

Ty thể mã hóa cho các RNA và protein			
Loài	**Kích thước (kb)**	**Các gen mã hóa protein**	**Các gen mã hóa RNA**
Nấm	19-100	8-14	10-28
Sinh vật nguyên sinh	6-100	3-62	2-29
Thực vật	186-366	27-34	21-30
Động vật	16-17	13	4-24

1.2. Genome của lạp thể

DNA lạp thể (chloroplast DNA-ctDNA) cũng là một DNA genome độc lập, thường là mạch vòng, được tìm thấy trong lạp thể của thực vật.

- Genome của lạp thể rất khác nhau về kích thước, nhưng đủ lớn để mã hóa cho khoảng 50-100 protein cũng như rRNA và tRNA.

- DNA lạp thể dài từ 120-190 kb. Các genome của lạp thể đã được phân tích trình tự cho thấy có khoảng 87-183 gen. Bảng 2.2 mô tả các chức năng được mã hóa bởi genome lạp thể ở cây trồng.

Bảng 2.2. Genome của lạp thể ở các cây trồng mã hóa cho 4 rRNA, 30 tRNA và khoảng 60 protein

Các lạp thể có hơn 100 gen
Các gen
- Mã hóa RNA
16S rRNA
23S rRNA
4,5S rRNA
5S rRNA
tRNA
- Biểu hiện gen
Các r-protein
RNA polymerase
Khác
- Các chức năng của lạp thể
Rubisco và thylakoids
NADH dehydrogenase

Nói chung, các đặc điểm của genome lạp thể tương tự ở ty thể, ngoại trừ lạp thể mang nhiều gen hơn. Genome lạp thể mã hóa cho tất cả các loại rRNA và tRNA cần thiết trong tổng hợp protein, và cho khoảng 50 protein, bao gồm cả RNA polymerase và các protein ribosome.

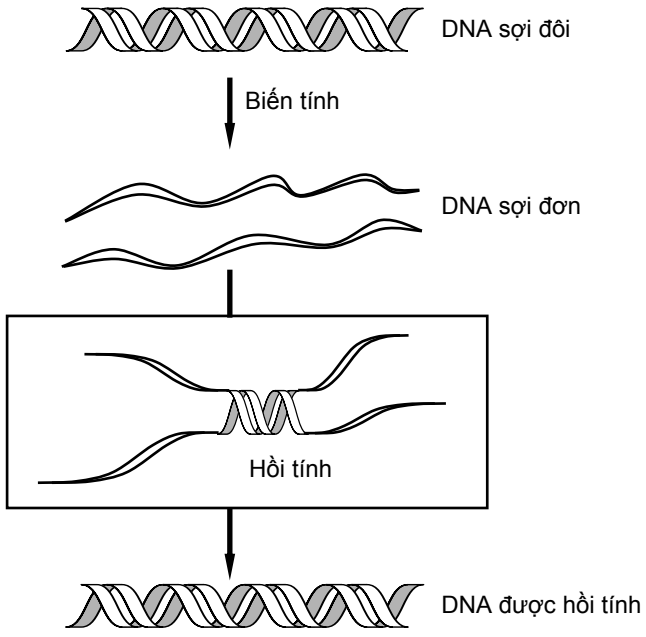
Các intron trong lạp thể được chia thành hai nhóm: 1) những intron ở trên các gen tRNA thường (mặc dù không chắc chắn) được định vị trong vòng anticodon, giống như các intron được tìm thấy trong các gen tRNA

của nhân nấm men *S. cerevisiae*; 2) những intron trong các gen mã hóa protein tương tự với các intron của các gen ty thể.

Vai trò của lạp thể là thực hiện quá trình quang hợp. Do đó, nhiều gen của nó mã hóa cho các protein của các phức hợp định vị trong các màng thylakoid. Một vài phức hợp protein của lạp thể giống các phức hợp protein của ty thể: có một số tiểu đơn vị được mã hóa bởi genome của cơ quan tử và một số khác được mã hóa bởi genome của nhân. Nhưng các phức hợp còn lại được mã hóa hoàn toàn bởi genome lạp thể.

2. Động học của phản ứng lai DNA

Bản chất chung của eukaryotic genome được phản ánh qua động học của sự tái liên kết các DNA (DNA reassociation kinetics) bị biến tính. Sự tái liên kết giữa các chuỗi DNA bổ sung xảy ra nhờ bắt cặp base, ngược lại với quá trình biến tính (denaturation) mà nhờ đó chúng được tách rời (Hình 2.2) để thực hiện sự tái bản hoặc phiên mã. Động học của phản ứng tái liên kết phản ánh sự khác nhau của các chuỗi hiện diện, vì thế phản ứng này có thể được dùng để định lượng các gen và các sản phẩm RNA của chúng.



Hình 2.2. DNA có thể biến tính và hồi tính

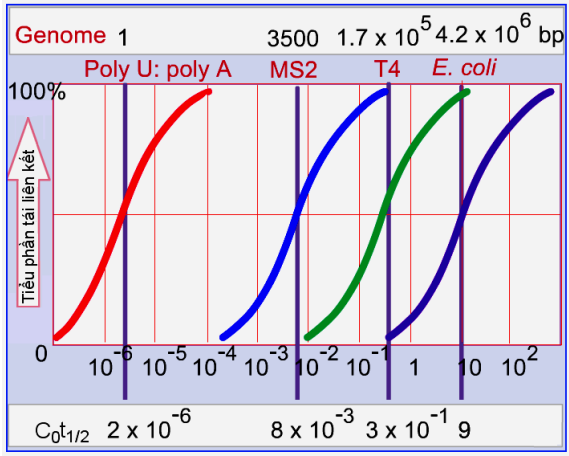
Bảng 2.3 mô tả phản ứng tái liên kết. Sự hồi tính của DNA (renaturation) phụ thuộc vào sự va chạm ngẫu nhiên của các chuỗi bổ sung. Phản ứng của các DNA riêng biệt có thể được mô tả bằng các điều kiện cần thiết cho sự hoàn thành một nửa (half-completion). Đây là tích số của $C_0 \times t_{1/2}$ và được gọi là $C_0t_{1/2}$. Giá trị này tỷ lệ nghịch với hằng số tốc độ. Do $C_0t_{1/2}$ là tích số của nồng độ và thời gian yêu cầu cho một nửa đường, nên một giá trị $C_0t_{1/2}$ lớn hơn dẫn đến một phản ứng chậm hơn.

Bảng 2.3. Một phản ứng tái liên kết của DNA được mô tả bởi $C_0t_{1/2}$

Phản ứng lai phụ thuộc vào C_0t
Tốc độ phản ứng
Phản ứng theo phương trình bậc hai $\frac{dC}{dt} = -kC^2$ *C* là nồng độ của DNA sợi đơn ở thời điểm t. *K* là hằng số tốc độ tái liên kết.
Tiến độ phản ứng
Lấy tích phân phương trình tốc độ giữa các giới hạn: nồng độ ban đầu của DNA = C_0 ở thời điểm t = 0; nồng độ duy trì sợi đơn = C sau thời gian t $\frac{C}{C_0} = \frac{1}{1 + k.C_0t}$
Thông số tới hạn là $C_0t_{1/2}$
Khi phản ứng hoàn thành một nửa ở thời điểm t = ½ $\frac{C}{C_0} = \frac{1}{2} = \frac{1}{1 + k.C_0t_{1/2}}$ Vì thế $C_0t_{1/2} = C_0t_{1/2} = \frac{1}{k}$

Sự hồi tính của DNA thường có dạng đường cong C_0t, đường cong biểu diễn đồ thị phân số của DNA được tái liên kết ($1-C/C_0$) theo log của C_0t. Hình 2.3 trình bày đường cong C_0t của một số genome đơn giản. Các đường cong có dạng tương tự nhau, nhưng giá trị $C_0t_{1/2}$ của mỗi đường là khác nhau.

Các genome trong hình 2.3 đại diện cho các nguồn DNA khác nhau (PolyU:PolyA, thực khuẩn thể MS2, thực khuẩn thể T4 và vi khuẩn *E. coli*). $C_0t_{1/2}$ liên quan trực tiếp với lượng DNA trong genome. Điều này phản ánh tình trạng khi genome trở nên phức tạp hơn, thì sẽ có thêm một số bản sao của một vài chuỗi đặc biệt trong một lượng DNA có trước. Ví dụ: nếu C_0 của DNA là 12 pg, thì nó sẽ chứa khoảng 3.000 bản sao của mỗi trình tự trong genome vi khuẩn.



Hình 2.3. $C_0t_{1/2}$ phụ thuộc vào độ phức tạp của genome. PolyU:PolyA, thực khuẩn thể MS2, thực khuẩn thể T4 và vi khuẩn *E. coli*.

3. Kích thước của genome

Không phải tất cả các đoạn DNA trong genome đều tương ứng với các gen (mã hóa cho protein hoặc một sản phẩm cần thiết cho hoạt động sống của tế bào). Từ những năm 1970, bằng các thí nghiệm gây bão hòa đột biến người ta đã có thể xác định được số gen nằm trên một đoạn nhiễm sắc thể. Ngày nay, nhờ các kỹ thuật phân tích DNA và RNA hiện đại (Southern blot, Northern blot, microarray...) các nhà khoa học có thể xác định số gen hoạt

động trong một tế bào. Ví dụ: ở tế bào nấm men *S. cerevisiae* (sinh vật eukaryote bậc thấp) có khoảng 4.000 gen hoạt động, còn tế bào động vật có vú khoảng 10.000-15.000 gen. Như vậy, nếu độ dài trung bình của một gen khoảng 10 kb thì tổng số chiều dài các gen hoạt động trong một tế bào cũng chỉ chiếm 1-2% genome. Hay nói cách khác, chỉ một phần rất nhỏ genome mang thông tin di truyền cần thiết cho hoạt động sống của tế bào. Vậy phần genome còn lại có vai trò gì, và tính phức tạp của loài có liên quan gì với kích thước genome hay không?

Để làm sáng tỏ vấn đề trên, chúng ta cần xem xét kích thước genome của một số loài gần nhau trong bậc thang tiến hóa (có độ phức tạp loài tương tự nhau) cũng như genome của những loài xa nhau (có tính phức tạp khác nhau). Chẳng hạn:

- Genome của người có kích thước khoảng $3{,}3\times10^9$ bp, trong khi đó genome của những loài lưỡng cư dài khoảng $3{,}1\times10^9$ bp hoặc thực vật có thể lên đến 10^{11} bp. Như vậy, có phải là các loài lưỡng cư có tính phức tạp tương tự cơ thể chúng ta?

- Hay là ngay trong cùng một loại, chúng ta cũng nhận thấy có sự mâu thuẫn về kích thước genome? Ví dụ: ruồi nhà (*Musca domestica*) có genome khoảng $8{,}6\times10^8$ bp, lớn gấp sáu lần kích thước genome của ruồi giấm khoảng $1{,}4\times10^8$bp. Ngoài ra, trong các loài lưỡng cư kích thước genome của chúng cũng thay đổi khá lớn từ 10^9-10^{11} bp. Vì sao ngay trong cùng một loại mà kích thước genome lại biến thiên nhiều như vậy, có phải ruồi nhà có cấu tạo phức tạp hơn nhiều so với ruồi giấm?

Từ những dữ liệu trên, chúng ta có thể nhận định rằng tính phức tạp của loài không liên quan đến kích thước của genome. Tuy nhiên, vai trò của phần genome còn lại (phần không mã hóa) đến nay vẫn chưa được biết nhiều.

4. Tổng số gen được biết ở một số loài eukaryote

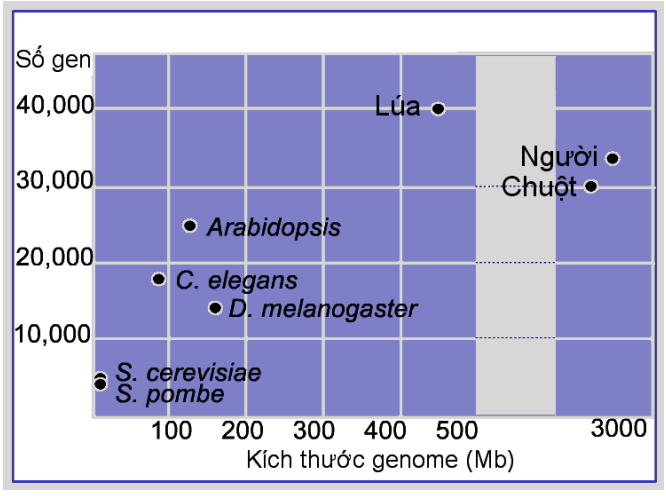
Có 6.000 gen ở nấm men *S. cerevisiae*, 18.500 gen ở giun tròn, 13.600 gen ở ruồi giấm, 25.000 gen ở *Arabidopsis*, và có khả năng 30.000 gen ở chuột và < 30.000 gen ở người.

Như chúng ta đã biết, mối quan hệ giữa kích thước genome và số lượng gen đã không còn nữa. Genome của các sinh vật eukaryote đơn bào

có cùng phạm vi kích thước với genome của vi khuẩn lớn nhất. Các eukaryote bậc cao có nhiều gen hơn, nhưng số lượng không tương quan với kích thước genome.

Hình 2.4 cho thấy genome của loài nấm men *S. cerevisiae* dài 13.500 kb và loài nấm men *S. pombe* là 12.500 kb, có khoảng 6.000 gen và 5.000 gen tương ứng. Khung đọc mở trung bình khoảng 1,4 kb, vì thế khoảng 70% genome được chiếm giữ bởi các vùng mã hóa. Sự khác nhau chủ yếu giữa chúng là chỉ 5% gen của *S. cerevisiae* có intron, so với 43% của *S. pombe*.

Genome của giun tròn có khoảng 18.500 gen. Mặc dù genome của ruồi giấm lớn hơn genome của giun tròn, nhưng chúng lại có số gen ít hơn. Đến nay, chúng ta chưa hiểu tại sao ruồi giấm-một cơ thể phức tạp hơn nhiều-chỉ có 70% số gen so với giun tròn. Điều này đã cho thấy không có một mối quan hệ chính xác giữa số gen và tính phức tạp của cơ quan.



Hình 2.4. Số lượng gen của sinh vật eukaryote rất khác nhau. Thay đổi từ 6.000-40.000 nhưng không tương quan với kích thước genome hoặc độ phức tạp của cơ thể.

Cây *Arabidopsis* có kích thước genome trung gian giữa giun tròn và ruồi giấm, nhưng lại có số gen lớn hơn cả hai (25.000). Điều này một lần nữa cho thấy không có một quan hệ rõ ràng, và cũng nhấn mạnh nét đặc biệt

của thực vật, là có thể có nhiều gen hơn (do sự nhân đôi của ông bà tổ tiên truyền lại) các tế bào động vật. Đa số genome *Arabidopsis* được tìm thấy trong các đoạn được nhân đôi, gợi ý rằng có một sự nhân đôi xa xưa trong genome (tạo ra một dạng tứ bội). Chỉ 35% các gen của *Arabidopsis* hiện diện như các bản sao đơn.

Genome của lúa lớn hơn *Arabidopsis* khoảng 4 lần, nhưng số gen chỉ lớn hơn khoảng 50%, có khả năng khoảng 40.000 gen. DNA lặp lại chiếm khoảng 42-45% genome. Hơn 80% gen tìm thấy trong *Arabidopsis* được hiện diện trong lúa. Trong số những gen chung này, khoảng 8.000 có trong *Arabidopsis* và lúa nhưng không thấy ở bất kỳ genome động vật hoặc vi khuẩn nào (những gen đã được phân tích trình tự). Có khả năng đây là tập hợp các gen mã hóa cho các chức năng đặc trưng của thực vật, chẳng hạn như quang hợp.

II. Tính phức tạp của genome

Kết quả nghiên cứu động học của các phản ứng lai được tiến hành giữa genomic DNA với cDNA (complementary DNA-DNA bổ sung), giữa DNA với mRNA… cho thấy hầu hết các gen hoạt động đều nằm trong thành phần DNA không lặp lại. Như vậy, thành phần này có ý nghĩa rất quan trọng trong việc đánh giá tính phức tạp của genome. Hay nói cách khác, dựa vào thành phần DNA không lặp lại có thể biết được kích thước genome cũng như mức độ tiến hóa của loài. Nếu như kích thước genome (ở trạng thái đơn bội) được coi là một thông số động học (ký hiệu C), thì giá trị này đặc trưng cho từng loài và không phải luôn luôn tỷ lệ thuận với tính phức tạp của loài. Ngược lại, giá trị C phản ánh các đặc điểm sau:

- Số lượng DNA mã hóa cho các sản phẩm cần thiết đối với hoạt động sống của cơ thể rất nhỏ so với số lượng DNA có trong genome.

- Có sự biến đổi rất lớn của giá trị C giữa một số loài mà tính phức tạp của chúng không khác nhau nhiều.

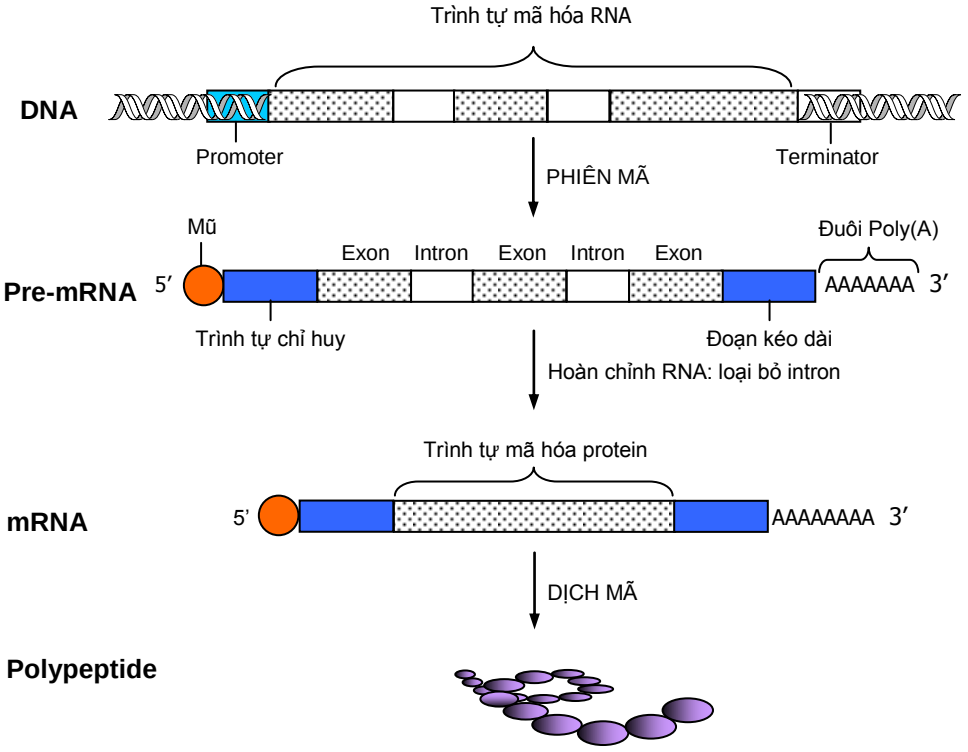
Genome vi khuẩn được xem là chỉ chứa các đoạn DNA không lặp lại và các gen thường tồn tại bản sao đơn. Ngược lại, genome của eukaryote thường chứa các gen có hai hoặc nhiều bản sao. Hơn nữa, trình tự nucleotide của các bản sao này có thể không giống nhau hoàn toàn mặc dù sản phẩm protein mà chúng mã hóa có cùng một chức năng. Các bản sao tương đồng của một gen được xếp chung vào một nhóm gọi là một họ gen

(gene family). Như vậy, ngoài các gen có một bản sao giống như ở vi khuẩn, genome của eukaryote còn chứa các họ gen. Hầu hết các gen mã hóa cho protein đã được phân lập đều nằm trong các họ gen khác nhau. Các gen trong một họ thường hoạt động theo thời gian và không gian. Điều đó có nghĩa, mỗi thành viên trong họ thường hoạt động ở một thời điểm nhất định trong quá trình hình thành và phát triển cá thể hoặc hoạt động trong các mô chuyên biệt. Khi một thành viên trong họ bị đột biến (bất hoạt) thì thành viên khác có thể hoạt động thay thế.

Khái niệm về gen được hình thành khi các nhà di truyền học cổ điển nghiên cứu biểu hiện các tính trạng mới do gây đột biến DNA của genome vi khuẩn hay thực khuẩn thể (bacteriophage). Một gen được xem là một đoạn DNA mà bất cứ đột biến nào xảy ra trên đó đều dẫn đến xuất hiện tính trạng mới. Điều này dễ hiểu đối với những genome chỉ chứa các gen có một bản sao (như genome của prokaryote). Tuy nhiên, ở genome của eukaryote, khi có nhiều gen cùng qui định một tính trạng hoặc các gen tương đồng trong một họ có thể hoạt động hỗ trợ thay thế nhau thì đột biến trên một gen không phải lúc nào cũng quan sát được ở mức độ phenotype. Mặt khác, các đoạn DNA tương ứng với trình tự mã hóa (coding sequence, exon) cho một protein thường bị ngắt quãng bởi các đoạn DNA không chứa thông tin di truyền (non-coding sequence, intervening sequence, intron). Các intron được phiên mã cùng với các exon sang phân tử RNA gọi là phân tử tiền thân mRNA (pre-mature mRNA hay pre-mRNA) nhưng sau đó chúng bị loại bỏ và các exon được nối lại với nhau tạo thành phân tử mRNA hoàn chỉnh (mature mRNA hay mRNA) được dùng cho quá trình sinh tổng hợp protein. Quá trình cắt các intron, nối exon không tuân theo một trật tự bắt buộc mà biến hóa đa dạng tạo ra các phân tử mRNA khác nhau từ một phân tử pre-mRNA (Hình 2.5). Bên cạnh mRNA, các phân tử rRNA và tRNA cũng được hình thành từ các phân tử tiền thân chứa intron. Ngoài ra, còn có hiện tượng mã di truyền của gen này nằm xen kẽ với các mã di truyền của gen khác (các gen nằm gối lên nhau-overlapping) hoặc trường hợp dịch chuyển khung đọc ngay trên một đoạn DNA.

Chúng ta đã biết đến chức năng của các phân tử RNA trong hoạt động sống của tế bào. Bên cạnh ba loại RNA đã được nghiên cứu khá kỹ (mRNA, rRNA và tRNA), vai trò của một số loại RNA khác mới được phát hiện vào những năm cuối thế kỷ 20. Chúng kiểm soát sự hoạt động của gen (hiện

tượng bất hoạt gen-gene silencing), tham gia phản ứng đọc sửa thông tin di truyền trên phân tử mRNA (hiện tượng RNA editing) hay quyết định tính bền vững của mRNA (các ribonuclease)… Thậm chí có những phân tử pre-mRNA được tổng hợp không phải mã hóa cho protein mà với mục đích phân cắt tạo ra những phân tử RNA có kích thước nhỏ hơn tham gia quá trình kiểm soát hoạt động của các gen khác. Đột biến ở những đoạn DNA mã hóa cho tất cả các loại RNA này thường gắn liền với việc xuất hiện tính trạng mới. Do đó, cần xem xét các đoạn nucleotide đó như các gen mặc dù chúng không mã hóa cho protein.



Hình 2.5. Các gen bị gián đoạn được biểu hiện thông qua RNA tiền thân. Các intron được loại bỏ, trong khi đó các exon được nối lại với nhau. mRNA chỉ có các trình tự của exon được dịch mã thành chuỗi polypeptide.

Ngày nay, theo quan điểm của sinh học phân tử, một gen được xem là một đoạn DNA mã hóa cho một sản phẩm cần thiết đối với hoạt động sống

của tế bào. Rõ ràng rằng không phải chỉ có DNA mã hóa cho protein mà cả các DNA mã hóa cho rRNA, tRNA và các loại RNA khác tham gia vào những phương thức kiểm soát hoạt động của genome cũng được xác định là gen.

III. Thay đổi trật tự của các đoạn DNA trong genome-Transposon

1. Sự thay đổi trật tự của các đoạn DNA trong genome

Như đã biết kích thước và cấu trúc genome của các loài rất khác nhau. Nguyên nhân của sự đa dạng này là do: 1) trao đổi chéo giữa các cặp nhiễm sắc thể tương đồng xảy ra trong phân bào giảm nhiễm đã dẫn đến sự đa dạng trong loài; 2) trong genome trật tự các đoạn DNA cũng như cấu trúc các gen được sắp xếp lại, đặc trưng cho từng cá thể. Chính các yếu tố di truyền có khả năng di chuyển giữa các vị trí trong một genome hoặc giữa các genome khác nhau đã góp phần làm đa dạng di truyền giữa các cá thể trong loài.

Các yếu tố di truyền có khả năng di chuyển được xếp vào ba nhóm chính tùy thuộc vào tính độc lập của chúng:

- Nhóm thứ nhất gồm các yếu tố có khả năng di chuyển giữa các vị trí khác nhau trong genome.

- Nhóm thứ hai gồm các yếu tố có khả năng ghép vào và tách ra khỏi genome để tồn tại độc lập trong tế bào (các episome như plasmid F, bacteriophage).

- Nhóm thứ ba chỉ di chuyển dưới sự kiểm soát của tế bào ở những giai đoạn sinh trưởng phát triển nhất định để sắp xếp khởi động một số gen đặc biệt (hình thành cassette hoạt động ở nấm men *S. cerevisiae*, ký sinh trùng đơn bào *Trypanosome*).

Sự di chuyển của các yếu tố ở nhóm thứ hai và thứ ba liên quan tới tái tổ hợp tương đồng hoặc tái tổ hợp ở các vị trí đặc hiệu. Mặc dù không có khả năng tồn tại độc lập bên ngoài genome, nhóm thứ nhất tự kiểm soát sự di chuyển của chúng trong genome và không đòi hỏi sự tương đồng giữa chúng với vị trí ghép vào. Do đó, có thể nói chúng di chuyển một cách tự do trong genome và được gọi tên chung là các yếu tố di chuyển (transposable elements, transposons). Trong khi thực khuẩn thể λ được xem là episome,

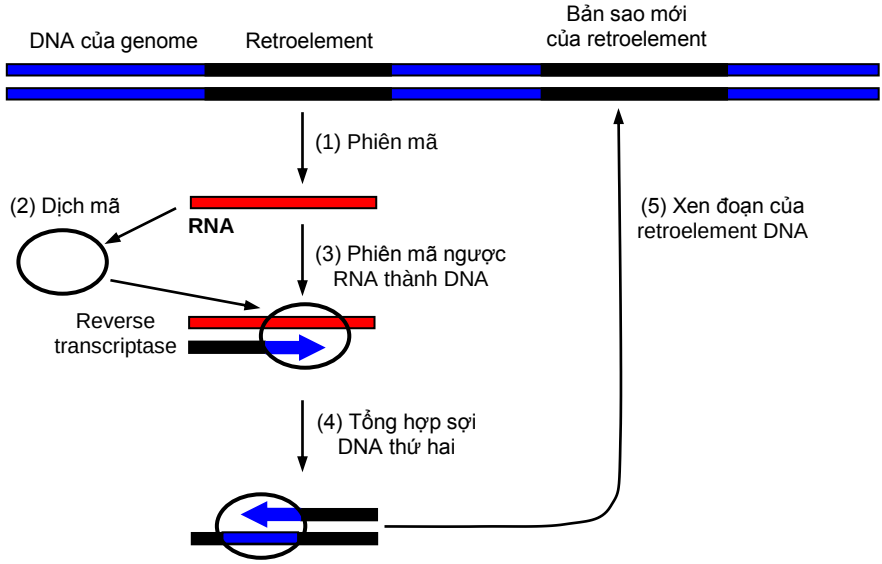
thì thực khuẩn thể *Mu* và một số retrovirus ở eukaryote được xem là transposable elements.

2. Các transposon

Các yếu tố di chuyển có thể được chia làm hai loại dựa vào cách thức di chuyển của chúng:

- Loại thứ nhất được gọi là retroelement, chúng phải trải qua hình thức trung gian RNA trong quá trình di chuyển (RNA genome được sao chép nhờ reverse transcriptase tạo ra cDNA để ghép vào vị trí mới) (Hình 2.6).

- Loại thứ hai là các đoạn DNA hoặc bị tách ra hoặc được tái bản tạo thêm bản sao để ghép vào vị trí mới. Thông thường, loại thứ hai này được gọi là transposon.



Hình 2.6. Retroelement

Sự tồn tại của các transposon (còn gọi là gen nhảy) là nét đặc trưng của tế bào thực vật. Ở sinh vật eukaryote, các transposon còn được gọi là yếu tố kiểm soát (controlling elements). Transposon có thể di chuyển từ nhiễm sắc thể này sang nhiễm sắc thể khác hoặc ở các vị trí khác nhau trên cùng một nhiễm sắc thể. Chúng có thể "nhảy" vào giữa dãy mã của một gen đang hoạt động làm cho gen này bất hoạt hoặc ngược lại. Ở cây ngô, người

ta đã phát hiện 10 nhóm transposon và một số trong chúng đã được giải mã. Tất cả chúng đều có kích thước khoảng 4.200 nucleotide và khác nhau chủ yếu ở đoạn mã tận cùng của gen.

Khi di chuyển, các transposon gây ra việc sắp xếp và tổ chức lại genome của từng cá thể như tạo các đoạn DNA mới ở vị trí chúng ghép vào và tách ra. Chúng có thể di chuyển tới vị trí bất kỳ và hoàn toàn không cần một mối quan hệ nào giữa hai vị trí mới và cũ. Khi tách ra transposon có thể mang theo các đoạn DNA bên cạnh, gây ra sự mất đoạn tại vị trí cũ. Ngược lại, khi ghép vào vị trí mới, chúng gây ra hiện tượng thêm đoạn hoặc chuyển đoạn ở vị trí mới. Do đó, transposon giống như các vector chuyên chở DNA từ nơi này sang nơi khác trong một genome hoặc từ genome này sang genome khác.

Ngoài ra, trao đổi chéo giữa các transposon tương đồng ở hai vị trí khác nhau trên một hoặc trên hai nhiễm sắc thể cũng tạo ra những biến đổi tương tự. Những biến đổi đó dẫn đến việc sắp xếp lại genome, tạo ra tính đa dạng giữa chúng và tính đặc thù riêng của từng cá thể. Đặc biệt, sự thay đổi vị trí của các transposon còn có thể ảnh hưởng đến hoạt động của các gen phân bố xung quanh ngay khi chúng không làm thay đổi trật tự nucleotide ở những gen này. Tần số ghép của các transposon vào genome khoảng 10^{-5}-10^{-7} sau mỗi thế hệ. Ngược lại, tần số tách ra khoảng 10^{-6} đến 10^{-10}.

Các transposon có thể chia làm hai nhóm dựa vào khả năng di chuyển độc lập hay phải phụ thuộc vào sự có mặt của transposon khác:

- Nhóm thứ nhất gồm các đoạn DNA có khả năng di chuyển độc lập. Chúng chứa gen mã hóa cho các protein điều khiển quá trình đó, ví dụ như enzyme nhận biết hai đầu transposon để tách chúng ra khỏi vị trí cũ và ghép vào vị trí mới. Do đó, quá trình này được thực hiện hoàn toàn độc lập. Nhờ khả năng này, chúng tạo ra các đột biến không bền vững.

- Nhóm thứ hai gồm các transposon không có khả năng tự hoạt động, tức là chúng không có khả năng di chuyển do không mang đủ thông tin di truyền mã hóa cho các enzyme cần thiết. Vì vậy, các transposon loại này tạo ra những đột biến gen một cách tự phát nhưng là đột biến bền vững. Việc di chuyển của transposon ở nhóm này phụ thuộc vào sự có mặt của transposon có khả năng hoạt động độc lập cùng nhóm. Hai transposon có thể xếp vào cùng nhóm khi chúng có cấu trúc tương đồng nhau, đặc biệt là các đoạn

oligonucleotide phân bố ở hai đầu transposon. Đây là vị trí để enzyme nhận biết và cắt nối transposon ở vị trí đi và đến.

Các transposon đơn giản nhất ở vi khuẩn được gọi là IS (insertion sequences). Chúng có thể nằm trên chromosome hoặc trên các plasmid. Transposon vi khuẩn không giữ một chức năng nào trong tế bào. Trình tự nucleotide ở một đầu IS thường lặp lại nhưng ngược chiều so với đầu kia (inverted repeat). Ví dụ như GGTAT-X_n-ATACC. Do đó, khi sợi đôi IS tách thành hai sợi đơn thì mỗi sợi này có khả năng hình thành liên kết bổ sung tại hai đầu của IS tạo cấu trúc thân-quai/cán-thòng lọng (stem-loop).

Transposon thường mã hóa cho các enzyme transposase làm nhiệm vụ nhận biết chuỗi nucleotide lặp lại ngược chiều (inverted repeat) để cắt transposon và di chuyển. Khi một IS được ghép vào vị trí bất kỳ của genome thì một đoạn ngắn DNA tại vị trí này được nhân đôi, tức là hai đầu của IS được chặn bởi các nucleotide giống nhau, sắp xếp theo cùng một chiều. Vì vậy, đoạn ngắn DNA này được gọi là "lặp lại xuôi chiều" (direct repeat). Chiều dài của chúng thường khoảng 9 bp (Hình 2.7). Dựa vào sự có mặt của các đoạn cùng chiều và ngược chiều có thể xác định được vị trí transposon ghép vào hoặc chuyển đi.

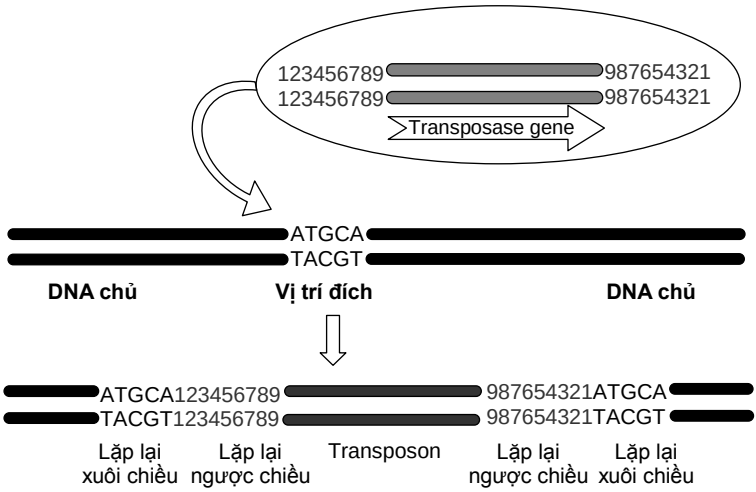
Ngoài các IS, ở vi khuẩn còn có các đoạn DNA có khả năng di chuyển với kích thước dài hơn, gọi là T_n. Các T_n thường phân bố trên plasmid (phân tử DNA dạng vòng, kích thước thường không lớn) và có khả năng chèn vào bất kỳ vị trí nào trong genome. Chúng thường mang thông tin di truyền mã hóa cho các protein kháng kháng sinh. Giữa IS và T_n có mối quan hệ về trình tự các nucleotide. Các T_n thường được giới hạn ở hai đầu bởi một loại IS nào đó.

Quá trình di chuyển của một transposon từ vị trí cũ (vị trí cho, donor) sang vị trí mới (vị trí nhận, recipient) xảy ra theo hai cơ chế khác nhau:

- **Cơ chế sao y bản chính (transposon có mặt ở cả hai vị trí).** Theo cơ chế này, phiên bản sau khi được sao chép từ vị trí cho sẽ được ghép vào vị trí nhận. Như vậy, mỗi lần di chuyển thì số lượng bản sao được tăng lên. Quá trình này liên quan đến hai loại enzyme: transposase (tác động vào hai đầu bản gốc transposon) và resolvase (tác động lên bản sao).

- **Cơ chế tách ra khỏi vị trí cũ di chuyển đến vị trí mới.** Theo cơ chế này, một transposon có thể tách ra khỏi vị trí cũ và ghép vào vị trí mới. Như vậy, số lượng transposon không thay đổi. Kiểu di chuyển này chỉ cần

enzyme transposase. Khi transposon chuyển đi, vị trí cũ bị gãy và nó được nối lại nhờ cơ chế sửa chữa DNA trong tế bào.



			Chiều dài tổng số	Chọn lọc đích
IS1	9 bp	23 bp	768 bp	ngẫu nhiên
IS2	5 bp	41 bp	1327 bp	các điểm nóng
IS4	11-13 bp	18 bp	1428 bp	$AAAN_{20}TTT$
IS5	4 bp	16 bp	1195 bp	các điểm nóng
IS10R	9 bp	22 bp	1329 bp	NGCTNAGCN
IS50R	9 bp	9 bp	1531 bp	các điểm nóng
IS903	9 bp	18 bp	1057 bp	ngẫu nhiên

Hình 2.7. Một transposon có đoạn lặp lại ngược chiều gồm 9 nucleotide (123456789) gắn vào vị trí có 5 nucleotide (ATGCA). Đoạn ngắn ATGCA có mặt ở cả hai đầu của IS và có trình tự nucleotide sắp xếp theo cùng một chiều.

Khi transposon ghép vào vị trí mới, một đoạn nucleotide ngắn ở đó được sao thành hai bản, mỗi bản nằm ở một đầu của transposon (direct repeat). Tại vị trí nhận, mỗi sợi đơn DNA bị cắt lệch nhau vài nucleotide. Transposon nối vào các đầu cắt, tạo ra hai khoảng trống (gaps). Chúng được sửa chữa theo nguyên tắc tạo cặp bổ sung. Do đó, đoạn nucleotide nằm giữa hai vết cắt được sao chép thành hai bản, mỗi bản ở một đầu và trình tự sắp xếp các nucleotide giống nhau. Vì vậy, chúng được gọi là lặp lại xuôi chiều.

IV. Tương tác của T-DNA với genome thực vật

Sự di chuyển DNA từ genome vi khuẩn sang genome thực vật được nghiên cứu khá kỹ đối với tương tác giữa *Argobacterium tumefaciens* hoặc *A. rhizogenes* với hầu hết các cây hai lá mầm. Hiện tượng di chuyển DNA này gây những biến đổi về mặt di truyền, biểu hiện ở việc xuất hiện các khối u trên thân cây (*A. tumefaciens*) hoặc mọc rất nhiều rễ tơ (*A. rhizogenes*) tại nơi bị nhiễm vi khuẩn.

A. tumefaciens và *A. rhizogenes* là hai loài vi khuẩn gây bệnh ở thực vật. Tuy nhiên, sau đó bệnh được duy trì lại không phụ thuộc sự tồn tại của vi khuẩn. Đó là do một số gen của vi khuẩn đã được chuyển vào genome cây chủ và hoạt động gây bệnh.

Các gen vi khuẩn có khả năng di chuyển và hoạt động trong tế bào thực vật nằm trên Ti-plasmid (tumor inducing plasmid) của *A. tumefaciens* hoặc trên Ri-plasmid (hairy-root inducing plasmid) của *A. rhizogenes*. Cũng như các khối u động vật, các tế bào thực vật có DNA vi khuẩn ghép vào genome bị chuyển sang trạng thái mới, ở đó sự phát triển và biệt hóa của chúng hoàn toàn khác với các tế bào bình thường. Đó là do hoạt động của các gen vi khuẩn trong genome của thực vật. Bình thường, những gen này có mặt trong genome vi khuẩn nhưng chúng chỉ được hoạt động (mở) sau khi hợp nhất vào genome thực vật và chịu sự kiểm soát của tế bào cây chủ. Quá trình này có tính chất đặc hiệu vật chủ, tức là một loại vi khuẩn chỉ có khả năng gây khối u trên một số loại cây chủ này mà không tương tác được với các loại cây khác.

Việc tạo khối u hay thực hiện quá trình chuyển gen từ vi khuẩn sang genome thực vật dẫn đến biến đổi trạng thái sinh lý của tế bào thực vật đòi hỏi các điều kiện sau:

- Phải có hoạt động của các gen trên ba vùng *chvA, chvB, pscA* nằm trên nhiễm sắc thể của vi khuẩn để khởi động việc bám dính vi khuẩn vào thân cây.

- Ti-plasmid hoặc Ri-plasmid phải mang vùng *vir* (nằm ngoài vùng T-DNA). Vùng này mang các gen cần thiết cho việc tách và vận chuyển T-DNA sang tế bào thực vật.

- Các gen trên vùng T-DNA của Ti-plasmid hoặc Ri-plasmid được ghép vào genome thực vật gây biến đổi trạng thái các tế bào này.

1. Ti-plasmid và Ri-plasmid

Plasmid là các vòng DNA tự sinh sản độc lập ở bên ngoài nhân. Ở vi khuẩn và động-thực vật, plasmid liên quan tới yếu tố giới tính của tế bào, đến khả năng chống chịu các loại kháng sinh... Đặc điểm quan trọng của plasmid là chúng có thể liên kết vào nhiễm sắc thể nhưng cũng có thể tồn tại bên ngoài nhiễm sắc thể một cách độc lập.

Các plasmid của *Agrobacterium* mang các vùng T-DNA, *vir* (virulence region), gen chuyển hóa opine, và vùng *ori* khởi đầu sao chép trong tế bào vật chủ.

Sự khác nhau giữa Ti-plasmid và Ri-plasmid ở chỗ, vùng T-DNA của Ti-plasmid chứa các gen tổng hợp auxin, cytokinin và opine. Trong khi đó, vùng T-DNA của Ri-plasmid chỉ mang gen tổng hợp auxin và opine. Đây là điểm khác nhau gây ra hiệu quả tác động khác nhau lên thực vật. Cơ chế gây bệnh của các *Agrobacterium* là sau khi xâm nhiễm vào tế bào, chúng gắn đoạn T-DNA vào bộ máy di truyền của tế bào thực vật, làm rối loạn các chất sinh trưởng nội sinh, tạo ra khối u (trường hợp *A. tumefaciens*) hoặc rễ tơ (trường hợp *A. rhizogenes*). Khả năng chuyển các gen này đã được khai thác để chuyển gen ngoại lai vào bộ máy di truyền của tế bào thực vật theo ý muốn.

2. T-DNA

Vùng T-DNA được nghiên cứu rất kỹ. Đó là một đoạn DNA có kích thước 25 kb trong đó chứa gen mã hóa cho sinh tổng hợp auxin, cytokinin và opine (trường hợp Ti-plasmid) hoặc auxin và opine (trường hợp Ri-plasmid). Trong plasmid, vị trí của T-DNA được giới hạn bằng biên phải (right border-RB) và biên trái (left border-LB), mỗi biên có chiều dài 25 bp.

3. Vùng vir

Trong các vùng DNA của Ti-plasmid và Ri-plasmid, ngoài T-DNA, được nghiên cứu nhiều hơn cả là vùng DNA phụ trách khả năng lây nhiễm còn gọi là vùng *vir* (virulence). Sản phẩm hoạt động của các gen nằm trong vùng *vir* dưới tác động kích thích của các hợp chất phenol tiết ra từ vết thương là một loạt các protein đặc hiệu như *vir*A, *vir*E2, *vir*B, *vir*D, *vir*D2, *vir*C1... Các protein này nhận biết các vết thương ở các cây chủ thích hợp (hầu hết là cây hai lá mầm), kích thích sản sinh ra các đoạn T-DNA, bao bọc

che chở các đoạn DNA này và giúp chúng tiếp cận với genome của cây chủ một cách an toàn.

4. Quá trình chuyển T-DNA vào tế bào thực vật

Quá trình cắt đoạn T-DNA ra khỏi plasmid và vận chuyển nó vào tế bào thực vật trước hết phụ thuộc vào sản phẩm của các gen *vir*. Vi khuẩn xâm nhiễm tại bất kỳ vị trí tổn thương nào trên thân cây. Cây có vết thương do sự hỏng ngẫu nhiên của màng tế bào thực vật hoặc do vi khuẩn tiết ra hỗn hợp những chất được mã hóa bởi các gen *vir*. Hoạt động của các gen này được hoạt hóa bởi hợp chất phenolic của cây (ví dụ như acetosyringone, catechol, các dẫn xuất của chalcone...). Ngoài ra, các monosaccharide như glucose, arabinose có mặt trong môi trường cũng làm tăng khả năng gây cảm ứng nhóm gen *vir*.

Protein VirA đóng vai trò quan trọng trong việc quyết định loại cây chủ bị nhiễm bởi *Agrobacterium*. Trong thực tế, *Agrobacterium* không có khả năng xâm nhập vào cây một lá mầm, có thể protein VirA không nhận biết các tín hiệu do cây một lá mầm tiết ra. Nói chung, *Agrobacterium* chỉ gây hại ở cây hai lá mầm, vì vậy người ta cho rằng chúng chỉ có thể đưa T-DNA vào hệ gen các cây hai lá mầm. Tuy nhiên, gần đây nhiều tác giả đã chứng minh khi nhiễm vi khuẩn, các cây một lá mầm cũng có thể sản xuất opine và vì thế có thể khai thác khả năng biến nạp gen của *Agrobacterium* vào cây một lá mầm. Hiện nay, người ta cũng đã thành công trong việc chuyển gen vào một số cây một lá mầm như lúa, mía... qua trung gian *A. tumefaciens.*

Khi cây nhiễm *A. tumefaciens*, do T-DNA hợp nhất vào trong genome của cây chủ bắt đầu hoạt động và sản xuất ra auxin, cytokinin và opine, toàn bộ sinh trưởng của cây bị rối loạn, các tế bào phân chia vô tổ chức và tạo ra các khối u. Opine được vi khuẩn sử dụng như một loại "thức ăn" nhờ gen chuyển hóa opine trên Ti-plasmid. Cơ chế lây nhiễm của *A. rhizogenes* đối với cây hai lá mầm cũng tương tự, nhưng trong vùng T-DNA của *A. rhizogenes* chỉ có gen sản sinh ra auxin, vì thế sự thay đổi hình thái chính của thực vật là chúng tạo ra rất nhiều rễ tơ khi bị nhiễm bệnh.

Để gắn T-DNA vào tế bào thực vật, đầu tiên vi khuẩn *A. tumefaciens* phải tiếp xúc với thành tế bào thực vật bị tổn thương. Quá trình này được thực hiện nhờ các gen *chvA* và *chvB*. Gen *chvB* mã hóa một protein liên

quan đến hình thành β-1,2 glucan mạch vòng, trong khi đó gen *chvA* xác định một protein vận chuyển, định vị ở màng trong của tế bào vi khuẩn. Protein vận chuyển giúp vận chuyển β-1,2 glucan vào khoảng giữa thành tế bào và màng sinh chất. β-1,2 glucan giữ vai trò quan trọng để vi khuẩn *Agrobacterium* tiếp xúc với thành tế bào thực vật. Nếu không có sự tiếp xúc này, sẽ không có sự dẫn truyền T-DNA.

Các sản phẩm protein của vùng *vir* có tác dụng cho việc dẫn truyền T-DNA từ vi khuẩn vào tế bào thực vật. Các loại protein đó rất cần thiết cho quá trình cắt T-DNA khỏi plasmid, cảm ứng thay đổi màng tế bào thực vật mà chúng tiếp xúc, tham gia di chuyển phần T-DNA qua màng vi khuẩn tới tế bào chất của tế bào thực vật, vận chuyển tới nhân rồi cuối cùng xâm nhập vào genome của cây chủ.

Thực chất, chỉ riêng T-DNA của plasmid được chuyển vào genome tế bào thực vật, mà không còn phần nào khác. Quá trình dẫn truyền chỉ do sản phẩm của các gen *vir* và gen *chv* quyết định mà không liên quan đến các gen khác trên T-DNA. Tuy nhiên, chuỗi DNA 25 bp (RB và LB của T-DNA) có vai trò là vị trí cảm ứng cho các sản phẩm của tổ hợp các gen vùng *vir*, đặc biệt là protein từ gen *vir*E mang chúng dẫn truyền vào tế bào thực vật. Chúng hoạt động như các tín hiệu nhận biết và khởi động quá trình dẫn truyền. Trước hết gen *vir*A trong tổ hợp gen vùng *vir* được phosphoryl hóa nhờ tác động của các hợp chất phenol như acetosyringone giải phóng ra từ các tế bào thực vật tổn thương. Sản phẩm của quá trình này lại tiếp tục phosphoryl hóa gen *vir*G. Sản phẩm của gen *vir*G liên tiếp làm hoạt hóa toàn bộ các gen *vir* còn lại, mà hai gen cuối cùng được hoạt hóa là gen *vir*B và *vir*E. Trước đó, khi gen *vir*D được hoạt hóa, sản phẩm của nó cảm ứng nhận biết RB và LB của T-DNA và làm đứt phần T-DNA ra khỏi DNA của plasmid thành các sợi đơn. Đồng thời, quá trình phosphoryl hóa này cũng làm thay đổi thẩm suất màng tế bào thực vật, màng tế bào bị mềm ra và bị thủng. Các sợi đơn T-DNA được gắn vào protein do gen *vir*E tổng hợp và dịch chuyển về phía màng tế bào vi khuẩn. Ngay sau đó, sợi T-DNA được trượt từ vi khuẩn vào tế bào thực vật. Cầu nối chính là sự tiếp hợp (conjugation) giữa hai tế bào do cảm ứng sản phẩm gen *vir*B mà thành. Khi T-DNA đã được chuyển giao vào tế bào thực vật, chúng nhanh chóng hợp nhất vào genome tế bào thực vật được ổn định và di truyền như các gen bình thường khác.

V. Sắp xếp và khuếch đại các gen trong genome

1. Sắp xếp lại các gen

Nói chung, genome có cấu trúc và tổ chức bền vững. DNA của genome thường không bị biến đổi bởi sự phát triển vô tính, nhưng thỉnh thoảng các trình tự của chúng cũng có thể bị chuyển chỗ trong gen, bị cải biến, khuếch đại hoặc thậm chí biến mất, như là một trường hợp tự nhiên.

Trao đổi chéo trong phân bào giảm nhiễm là một trong những nguyên nhân gây ra biến đổi của genome. Tuy nhiên, điều này xảy ra chủ yếu trong tế bào sinh dục mà không có trong các tế bào soma. Việc sắp xếp lại genome sẽ dẫn đến những thay đổi sau:

- Tạo ra các gen mới cần thiết cho sự biểu hiện trong các trường hợp đặc biệt.

- Sự tái sắp xếp có thể đáp ứng cho việc mở hoặc đóng gen. Đây cũng chính là cơ chế của sự điều hòa biểu hiện gen.

Một ví dụ điển hình là hiện tượng sắp xếp lại các gen trong genome của nấm men *S. cerevisiae* và của ký sinh trùng *Trypanosome* châu Phi khi gây chứng ngủ li bì ở vật chủ.

1.1. Chuyển đổi dạng giao phối của nấm men

Nấm men có thể tồn tại ở cả hai dạng đơn bội hoặc lưỡng bội. Các dạng lưỡng bội là dị hợp tử ở trường hợp locus kiểu giao phối, và các tế bào đơn bội có thể hoặc là MATa hoặc MATα. Việc chuyển đổi trạng thái được thông qua giao phối, kết hợp các bào tử đơn bội thành lưỡng bội và qua việc tái tạo giao tử mới. Tuy nhiên, giao phối chỉ xảy ra giữa hai loại tế bào đơn bội a và α. Các tế bào đơn bội cùng loại không thể kết hợp với nhau tạo thành tế bào lưỡng bội.

MAT (mating type locus) là vị trí hoạt động hay cassette hoạt động: một vùng đặc biệt của nhiễm sắc thể số 3 chứa các gen qui định dạng giao phối (a và α). Các gen này còn tồn tại ở hai vị trí khác trong genome. Tuy nhiên, ở đó chúng đều bị bất hoạt. Hai vị trí này được gọi là hai vị trí tĩnh (hay cassette tĩnh HML và HMR). Mỗi vị trí tĩnh chỉ mang các gen qui định cho một dạng giao phối. Khi các gen được sao chép từ một vị trí tĩnh vào vị trí hoạt động thì mRNA mới được tổng hợp từ các gen đó. Như vậy, quá

trình sao chép quyết định dạng giao phối của nấm. Bản gốc luôn được bảo tồn ở vị trí tĩnh và bản thứ hai xuất hiện ở vị trí hoạt động.

Nếu vị trí tĩnh có mang đột biến thì chúng sẽ được sao chép vào vị trí hoạt động. Tuy nhiên, nếu xảy ra đột biến ở cassette MAT thì tính trạng mới xuất hiện không bền. Một khi dạng giao phối chuyển đổi thì các gen cũ ở vị trí MAT bị thay thế bởi bản sao của vị trí tĩnh khác. Lúc đó, đột biến bị loại đi và tính trạng mới sẽ biến mất.

So sánh giữa các dạng cùng sợi nấm (homothallic) và khác sợi nấm (heterothallic) người ta nhận thấy dạng heterothallic có gen *HO* hoạt động và có thể chuyển đổi tự động giữa các kiểu giao phối, và vì thế một bào tử đơn có thể làm tăng quần thể tự phối, trong khi dạng homothallic không có gen *HO* và duy trì cùng một kiểu giao phối trong suốt chu trình sinh trưởng đơn bội.

Phân tích di truyền cho thấy các gen sau đây cần cho việc chuyển đổi kiểu giao phối:

- *MAT*
- *HO,* mã hóa cho enzyme endonuclease
- *HML*α (cho MATa chuyển thành MATα)
- *HMR*a (cho MATα chuyển thành MATa)

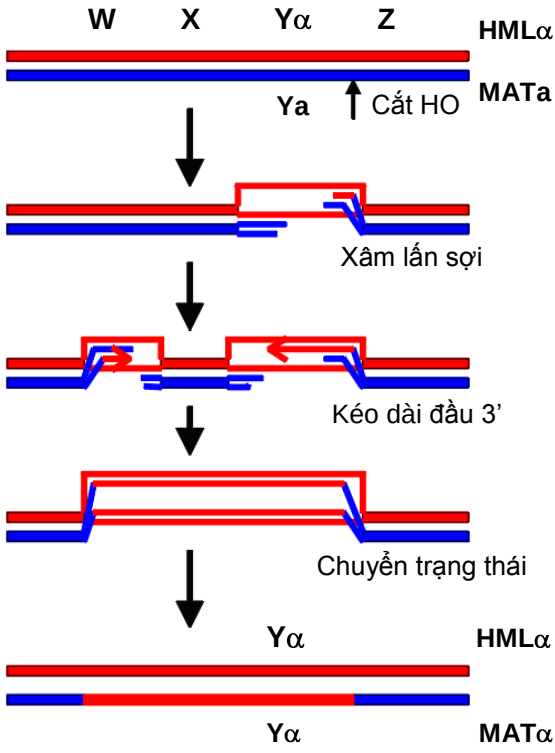
Trình tự các nucleotide ở vị trí tĩnh và vị trí hoạt động chỉ khác nhau một đoạn ngắn ký hiệu là Ya và Yα (Hình 2.8). Enzyme HO-endonuclease nhận biết vị trí đặc hiệu tại ranh giới phân cách giữa Z và Y và của cassette hoạt động MAT và cắt cả hai sợi DNA tại đó. Điều đặc biệt là enzyme này không cắt DNA khi chúng hiện diện ở cassette tĩnh.

Sau khi đoạn Y của vùng MAT bị phân hủy hết, đoạn Y của một trong hai cassette tĩnh được dùng làm khuôn mẫu để sao chép vào vị trí bị phân hủy (Hình 2.8). Nếu đột biến xuất hiện ở vùng MAT (đoạn Y), tính trạng mới chỉ biểu hiện tạm thời. Một khi dạng giao phối chuyển đổi, các gen bình thường được sao chép vào vùng MAT và đột biến bị loại đi.

1.2. Chuyển đổi gen ở *Trypanosome*

Trypanosome có khả năng lẩn tránh được hệ thống miễn dịch của vật chủ thường bằng cách thay đổi kháng nguyên bề mặt (surface antigen) của chúng. Mỗi loại kháng nguyên được tổng hợp nhờ hoạt động của một gen

tương ứng tại vị trí hoạt động. Gen này có thể bị thay thế bởi một gen mã hóa cho loại kháng nguyên khác nằm ở một vị trí tĩnh nào đó trong genome. Mỗi vị trí tĩnh chứa một gen ở trạng thái không hoạt động. Gen này chỉ được mở khi chuyển đến vị trí hoạt động. Trong genome của *Trypanosome*, có rất nhiều vị trí tĩnh nhưng chỉ có một vị trí hoạt động.



Hình 2.8. Quá trình chuyển đổi dạng giao phối từ a sang α nhờ trao đổi gen giữa vùng MATa và HMLα trên nhiễm sắc thể số 3

Bình thường, *Trypanosome* sẽ trải qua một số lần biến đổi hình thái khi được truyền từ ruồi châu Phi sang vật chủ. Bề mặt của tế bào *Trypanosome* được bao bọc một lớp đơn gồm 5×10^6 phân tử của một loại glycoprotein (VSG-variable surface glycoprotein). Đây chính là kháng nguyên bề mặt của *Trypanosome* khi chúng xâm nhập vào vật chủ. Điều đáng chú ý là chúng có khả năng thay đổi kháng nguyên bề mặt, do đó tránh được phản ứng miễn dịch của tế bào vật chủ. Quá trình thay thế kháng nguyên bề mặt phụ thuộc vào sự chuyển đổi các gen mã hóa cho chúng xảy

ra ở một vị trí đặc biệt trong genome (vị trí hoạt động). Chuyển đổi gen mã hóa kháng nguyên bề mặt nhằm mục đích hoạt hóa gen mã hóa kháng nguyên bề mặt mới thay thế cho kháng nguyên tồn tại trước đó. Khi một gen đang hoạt động bị thay thế bởi một gen khác sẽ tương ứng với việc xuất hiện kháng nguyên mới và loại bỏ kháng nguyên cũ.

- Cấu trúc của một VSG ở *Trypanosome*

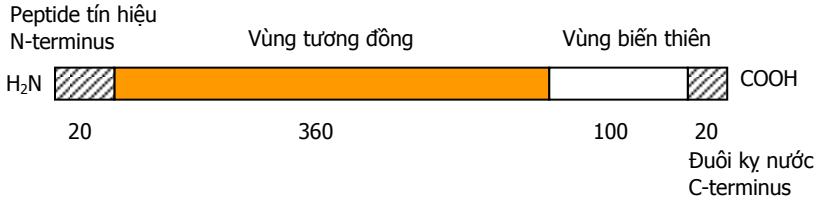
Cấu trúc chung của một VSG được mô tả trên hình 2.9 và 2.10. Một VSG vừa được tổng hợp dài khoảng 500 amino acid gồm tín hiệu N-terminus, tiếp theo là đoạn peptide quyết định tính kháng nguyên; đoạn peptide bảo thủ giữa các VSG và đuôi kỵ nước. Phân tử này được tổng hợp dưới dạng protein tiền thân (pre-protein). Do đó, chúng phải trải qua biến đổi ở hai đầu NH_2 và COOH để trở thành dạng protein hoàn chỉnh (mature form). Dạng này được đính vào màng tế bào ở đầu COOH.

Một loại *Trypanosome* có thể tạo ra ít nhất khoảng 100 VSG từ khi nhiễm cho đến khi gây chết vật chủ. Số gen mã hóa cho VSG có thể nhiều hơn 1.000 gen, tất cả các gen này đều nằm trong genome. Tuy nhiên, tại một thời điểm bất kỳ chỉ có một gen hoạt động tổng hợp nên một loại VSG. Do đó, sự thay đổi kháng nguyên tương ứng với sự thay đổi hoạt động của gen. Khi một gen mới được mở, gen hoạt động trước nó phải bị ức chế hoàn toàn. Lúc đó, một kháng nguyên mới sẽ thay thế kháng nguyên tồn tại trước nó.

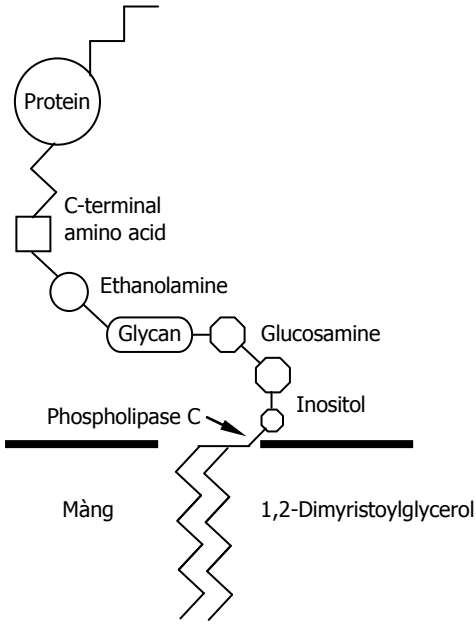
Hình 2.9 cho thấy chuỗi polypeptide chứa khoảng 500 amino acid. N-terminus chứa một peptide tín hiệu cho sự vận chuyển qua ER (lưới nội sinh chất, endoplasmic reticulum) và tới màng plasma, được tách ra khỏi protein hoàn chỉnh. Vùng biến thiên là khác nhau ở mỗi VSG, điều đó cho thấy các VSG có ít hoặc không có tính đồng nhất. Hướng tới phần C-terminus, chuỗi polypeptide được bảo toàn tốt hơn và phần này được gọi là vùng tương đồng. Đuôi kỵ nước chứa một tín hiệu nhận biết để gắn với mỏ neo glycolipid (glycolipid anchor) (Hình 2.10). Khi mỏ neo được gắn thì 20 amino acid cuối cùng sẽ được tách ra.

Glycoprotein biến đổi bề mặt (VSG) được gắn với màng thông qua một mỏ neo glycolipid chứa ethanolamine, một cấu trúc glycan mang một số gốc mannose (mannose moiety), một glucosamine và một phosphoinositol liên kết với 1,2-dimyristoyglycerol có gai trong màng plasma. Gốc glycolipid là yếu tố quyết định phản ứng lai (cross-reacting)

(CRD) được nhận biết bằng các kháng thể phản ứng với tất cả dạng biến đổi của VSG, nhưng chỉ khi VSG được phóng thích khỏi màng. Màng liên kết với VSG không được nhận biết bởi các kháng thể anti-CRD. Sự phóng thích VSG được xúc tác bởi hoạt tính của trypanosome-specific phospholipase C được giả định là hiện diện ở mặt bên trong (inner face) của màng plasma. Người ta không biết rằng enzyme có thể cắt liên kết phosphoester trên mặt khác của màng như thế nào.



Hình 2.9. Sơ đồ minh họa chuỗi protein của một VSG đặc trưng



Hình 2.10. Cấu trúc của mỏ neo glycolipid của VSG

- Hoạt động của gen VSG

Gen mã hóa cho một VSG được gọi là bản gen gốc (basic copy gene). Các gen này được phân thành hai nhóm tùy thuộc vào vị trí của chúng trên nhiễm sắc thể.

+ Các gen nằm ở telomere (khoảng 5-15 kb)>200 gen.

+ Các gen nằm cách telomere hơn 50 kb.

Tương tự như ở nấm men, các gen mã hóa cho VSG nằm rải rác trong genome và ở trạng thái không hoạt động. Một gen bất kỳ được hoạt hóa chỉ khi nó được sao chép vào vị trí hoạt động (expression site) trong khi nguyên bản của nó vẫn được bảo tồn ở vị trí tĩnh. Bản sao của bản gen gốc vào vị trí hoạt động được gọi là ELC (bản sao hoạt động-expression linked copy). Vị trí hoạt động nằm ở gần telomere. Như vậy, việc chọn một bản gen gốc để sao chép tạo bản sao hoạt động sẽ phụ thuộc vào vị trí của bản gen gốc ở telomere hoặc nằm phía trong telomere. Có hai giả thiết để một gen trở nên hoạt hóa như sau:

- Vị trí hoạt động không thay đổi mà chỉ có các ELC thay thế cho nhau. Bản sao của một bản gen gốc sẽ thay thế cho bản sao của một gen khác ở tại vị trí đó, điều này xảy ra tương tự như các gen ở cassette tĩnh được sao chép vào cassette hoạt động trong trường hợp với nấm men.

- Vị trí hoạt động thay đổi, khi cần tổng hợp một VSG mới, gen ở vị trí hoạt động cũ bị ngừng và gen ở một vị trí khác (gần telomere) được khởi động.

Một số phân tử mRNA tương ứng với các VSG khác nhau được phân lập và được xác định trình tự nucleotide (thông qua cDNA). Điều ngạc nhiên là phần oligonucleotide ở đầu 3’ của mọi gen VSG đều khác với phần 3’ của các mRNA được phiên mã từ các gen đó. Mặt khác, các gen này không có phần 5’ giống như các mRNA. Như vậy, các mRNA không được tổng hợp hoàn toàn trên khuôn mẫu các gen này. Phần 3’ của các mRNA tương ứng với phần 3’ của vị trí hoạt động ELC, trong khi phần 5’ (gồm 35 nucleotides) được tổng hợp từ những đoạn DNA khác và được gắn vào mRNA (hiện tượng trans-splicing).

2. Khuếch đại các gen

Số lượng bản sao của một số gen cũng có thể được tăng lên tạm thời trong quá trình phát triển các tế bào soma. Việc tăng số lượng của một gen đặc biệt nào đó phụ thuộc vào từng điều kiện cụ thể của tế bào và xảy ra không phổ biến. Các bản sao có thể nằm tập trung thành một nhóm gồm bản sao này nối tiếp bản sao khác hoặc có thể tồn tại như những đoạn DNA có khả năng tái bản độc lập. Chẳng hạn:

- Sự nhân bản của gen mã hóa cho rRNA ở trứng ếch. Trứng ếch có đường kính khoảng 2-3 mm, dự trữ rất nhiều rRNA. Chúng được phiên mã từ rất nhiều gen rDNA. Các gen này được nhân lên (khoảng 2.000 lần) theo cơ chế "vòng tròn quay" (xem chương 4) trong quá trình phát triển và tồn tại dưới dạng các vòng tròn khép kín.

- Khi nuôi cấy các tế bào động vật có vú trong môi trường đặc biệt, DNA tại một số vị trí trong genome được nhân lên. Ví dụ: nuôi cấy các tế bào ung thư trong môi trường chứa độc tố methotrexate. Chất này ức chế hoạt tính của enzyme dihydrofolate reductase (DHFR) giữ vai trò trong tổng hợp các nucleotide của DNA. Các tế bào ung thư nuôi cấy trong môi trường có chất độc này phát triển thành các quần lạc tế bào kháng lại độc tố. Khi nồng độ chất độc tăng dần, nồng độ DHFR cũng tăng theo, có thể đạt tới 1.000 lần lớn hơn mức bình thường. Nồng độ enzyme tăng do số lượng các gen mã hóa cho chúng tăng. Cơ chế chính xác của hiện tượng này chưa rõ ràng, nhưng có thể xảy ra theo hai cách:

- Trao đổi chéo không cân bằng giữa hai nhiễm sắc tử (chromatid) của nhiễm sắc thể dẫn đến một số tế bào không có gen *dhfr* và một số khác có hai bản sao của gen này. Trong môi trường có độc tố, trao đổi chéo không cân bằng được lặp đi lặp lại và các tế bào chứa nhiều gen *dhfr* vẫn phát triển tốt trong môi trường này.

- Các đoạn DNA (100-1.000 kb) chứa 2-4 gen *dhfr* (~31 kb/gen) được sao chép từ nhiễm sắc thể bình thường tạo ra các nhiễm sắc thể rất nhỏ, không có tâm động. Các nhiễm sắc thể nhỏ này ghép vào các nhiễm sắc thể bình thường khác. Quá trình này lặp đi lặp lại và qua một số lần phân bào nguyên nhiễm, tế bào nào mang số lượng lớn các gen *dhfr* càng có điều kiện phát triển thuận lợi trong môi trường có chứa độc tố.

3. Biến nạp gen

Một phương thức tăng khả năng di truyền là ứng dụng các tương tác vật chủ cộng sinh-ký sinh, trong đó DNA lạ được chuyển vào tế bào vật chủ từ một vi khuẩn. Cơ chế này tương tự với sự tiếp hợp của vi khuẩn. Sự biểu hiện của DNA vi khuẩn trong vật chủ mới của nó sẽ làm thay đổi kiểu hình

của tế bào. Ví dụ điển hình là vi khuẩn *A. tumefaciens* cảm ứng tạo khối u ở tế bào thực vật bị chúng xâm nhiễm.

Khi DNA lạ được đưa vào tế bào eukaryote, nó có thể tồn tại ngoài nhiễm sắc thể hoặc được hợp nhất trong genome. Nếu xảy ra theo trường hợp thứ hai, genome sẽ mang những đột biến di truyền và nhiều khi DNA lạ vẫn tiếp tục hoạt động làm xuất hiện các tính trạng mới. Việc đưa các gen lạ vào tế bào soma hoặc tế bào sinh dục mà vẫn duy trì hoạt động của những gen đó được gọi là chuyển nhiễm (transfection). Cá thể biểu hiện tính trạng mới nhờ hoạt động của gen lạ đưa vào tế bào sinh dục được gọi là cá thể chuyển gen. Quá trình này có thể làm thay đổi sự ổn định của genome. DNA sau khi được tiêm vào trong các tế bào trứng động vật có thể được hợp nhất trong genome và truyền cho thế hệ sau như một thành phần di truyền bình thường. Khả năng đưa một gen chức năng đặc trưng có thể trở thành một kỹ thuật y học sử dụng cho việc chữa bệnh di truyền.

Tài liệu tham khảo/đọc thêm

1. Võ Thị Thương Lan. 2002. Sinh học phân tử. *NXB Đại học Quốc gia Hà Nội*, Hà Nội.

2. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc*. New York, USA.

3. Cantor CR and Smith CL. 1999. Genomics. *John Wiley & Sons, Inc.* New York, USA.

4. Dale JW and Von Schanzt M. 2002. From Gene to Genome. *John Wiley & Sons, Ltd.* West Sussex, UK.

5. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

6. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *WH Freeman and Company*, New York, USA.

7. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Loscik R. 2004. Molecular Biology of the Gene. *The Benjamin Cummings/Cold Spring Habor Laboratory Press*, San Francisco, CA, USA.

8. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company Inc.* New York, USA.

Chương 3

Cấu trúc và chức năng của gen

I. Định nghĩa gen

Chúng ta có thể điểm qua những mốc chính trong lịch sử nghiên cứu về gen như sau:

Mendel (1865) là người đầu tiên đưa ra khái niệm **nhân tố di truyền**. Johansen (1909) đã đề xuất thuật ngữ **gen** (từ **genos**, nghĩa là sản sinh, nguồn gốc) để chỉ nhân tố di truyền xác định một tính trạng nào đó. Sau đó, Morgan trong những năm 1920 đã cụ thể hóa khái niệm về gen, khẳng định nó nằm trên nhiễm sắc thể và chiếm một locus nhất định, gen là đơn vị chức năng xác định một tính trạng.

Vào những năm 1940, Beadle và Tatum đã chứng minh gen kiểm tra các phản ứng hóa sinh và nêu giả thuyết **một gen-một enzyme**. Tuy nhiên, trường hợp hemoglobin là một protein nhưng lại gồm hai chuỗi polypeptide do hai gen xác định, do đó giả thuyết trên buộc phải điều chỉnh lại là **một gen-một polypeptide**.

Vào những năm 1950, DNA (deoxyribonucleic acid) được chứng minh là vật chất di truyền. Mô hình **cấu trúc DNA** của Watson và Crick được đưa ra và **lý thuyết trung tâm** (central dogma) ra đời. Gen được xem là một đoạn DNA trên nhiễm sắc thể mã hóa cho một polypeptide hay RNA.

Cuối những năm 1970, việc phát hiện ra gen gián đoạn ở sinh vật eukaryote cho thấy có những đoạn DNA không mã hóa cho các amino acid trên phân tử protein. Vì thế, khái niệm về gen lại được chỉnh lý một lần nữa: Gen là một đoạn DNA đảm bảo cho việc tạo ra một polypeptide, nó bao gồm cả phần phía trước là vùng 5’ không dịch mã (5’ untranslation) hay còn gọi là vùng ngược hướng (upstream) và phía sau là vùng 3’ không dịch mã (3’ untranslation) hay còn gọi là vùng cùng hướng (downstream) của vùng mã hóa cho protein, và bao gồm cả những đoạn không mã hóa (intron) xen giữa các đoạn mã hóa (exon).

Hiện nay, có thể định nghĩa gen một cách tổng quát như sau: Gen là đơn vị chức năng cơ sở của bộ máy di truyền chiếm một locus nhất định trên

nhiễm sắc thể và xác định một tính trạng nhất định. Các gen là những đoạn vật chất di truyền mã hóa cho những sản phẩm riêng lẻ như các mRNA được sử dụng trực tiếp cho tổng hợp các enzyme, các protein cấu trúc hay các chuỗi polypeptide để gắn lại tạo ra protein có hoạt tính sinh học. Ngoài ra, gen còn mã hóa cho các tRNA, rRNA và snRNA...

Bảng 3.1. Tóm tắt lịch sử nghiên cứu về di truyền học

Mốc thời gian	Năm	Các sự kiện chính
1850	1865	Gen là các nhân tố hạt
	1871	Khám phá ra nucleic acid
1900	1903	Nhiễm sắc thể là các đơn vị di truyền
	1910	Gen nằm trên nhiễm sắc thể
	1913	Nhiễm sắc thể là các dãy sắp xếp mạch thẳng của gen
	1927	Đột biến là những thay đổi vật lý của gen
	1931	Sự tái tổ hợp xuất hiện bởi hiện tượng vắt chéo
	1944	DNA là vật liệu di truyền
	1945	Gen mã hóa cho protein
1950	1951	Trình tự protein đầu tiên
	1953	DNA có dạng xoắn kép
	1958	DNA tái bản theo phương thức bán bảo thủ
	1961	Mã di truyền là bộ ba
	1977	Các gen của sinh vật eukaryote bị gián đoạn
	1977	DNA có thể được phân tích trình tự
	1995	Genome của vi khuẩn được phân tích trình tự
2000	2001	Genome người được phân tích trình tự

II. Lý thuyết trung tâm

1. Sự xác định di truyền cấu trúc bậc một của protein

Cấu trúc không gian của chuỗi polypeptide được xác định bởi trình tự sắp xếp của các amino acid tức cấu trúc bậc một. Như vậy, mặc dù có nhiều mức độ cấu trúc không gian khác nhau, nhưng cấu trúc bậc một tức trình tự sắp xếp các amino acid chi phối toàn bộ các mức độ cấu trúc khác. Việc xác định di truyền phân tử protein ở trạng thái tự nhiên có đầy đủ hoạt tính sinh học chỉ quy tụ lại chủ yếu ở xác định cấu trúc bậc một là đủ.

2. Các enzyme mất hoạt tính do đột biến

Nhiều nghiên cứu cho thấy, việc mất hoạt tính enzyme nhiều khi không phải do vắng mặt của enzyme, mà chỉ do các biến đổi trên phân tử (modification). Có trường hợp đột biến dẫn đến những thay đổi tinh vi, enzyme vẫn có hoạt tính nhưng sẽ biểu hiện khác nếu thay đổi điều kiện. Chẳng hạn:

Ở nấm mốc *Neurospora crassa,* enzyme tyrosinase do gen *T* xác định, xúc tác cho phản ứng chuyển hóa tyrosine thành dihydroxyphenylalanine. Alelle T^+ của dòng hoang dại sản xuất tyrosinase có hoạt tính ở nhiệt độ bình thường và cả ở 60^oC. Một đột biến T^S sản xuất tyrosine có hoạt tính ở nhiệt độ bình thường, nhưng lại mất hoạt tính ở 60^oC.

Như vậy, trong đa số trường hợp, đột biến của một gen không làm biến mất enzyme mà chỉ biến đổi cấu trúc dẫn đến thay đổi hoạt tính. Các đột biến của cùng một gen có thể gây ra những biến đổi khác nhau trên enzyme. Các hiện tượng đó chứng tỏ rằng cấu trúc của enzyme chịu sự kiểm soát trực tiếp của gen.

3. Bản chất các biến đổi di truyền của protein

Bản chất đó chính là quan hệ một gen-một polypeptide.

Như đã nêu trên, người ta khám phá ở người có những gen tạo ra hemoglobin (Hb) khi biến dị sẽ tạo ra những hemoglobin bất thường do sai hỏng ở các chuỗi polypeptide α hoặc β (Bảng 3.2 và 3.3) và gây ra các bệnh di truyền.

Bảng 3.2. Các loại hemoglobin ở chuỗi polypeptide α

Thứ tự amino acid	1	2	16	57	58	68	116	141
Tên amino acid	Val	Leu	Lys	Glu	His	Asp	Glu	Arg
Loại hemoglobin *Hb I*			Asp					
Hb Nocfolk				Asp				
Hb M Boston					Tyr			
Hb B Philadelphia						Lys		
Hb O Indonesia							Lys	

Ở trường hợp này, thứ tự các amino acid có thể bị sai lệch. Từ những dữ liệu trên ta thấy các dạng đột biến tạo một Hb bất thường đó là thay một amino acid này bằng một amino acid khác.

Bảng 3.3. Các loại hemoglobin ở chuỗi polypeptide β

Thứ tự amino acid	1	2	3	6	26	67	121	146
Tên amino acid	Val	His	Leu	Glu	Glu	Val	Glu	His
Loại hemoglobin *Hb S*				Val				
Hb C				Lys				
Hb E					Lys			
Hb M Minvauki						Glu		
Hb O Ả Rập							Lys	

Qua hai chuỗi polypeptide α và β chúng ta thấy có một số dạng hemoglobin bất thường ở người. Trên mỗi chuỗi, chỉ trình bày những amino acid đã bị thay đổi ở dạng đột biến. Số thứ tự chỉ vị trí của amino acid trong chuỗi polypeptide. Mỗi hemoglobin bất thường có thể được đặt cho một chữ cùng tên (nếu có) của địa phương được tìm thấy.

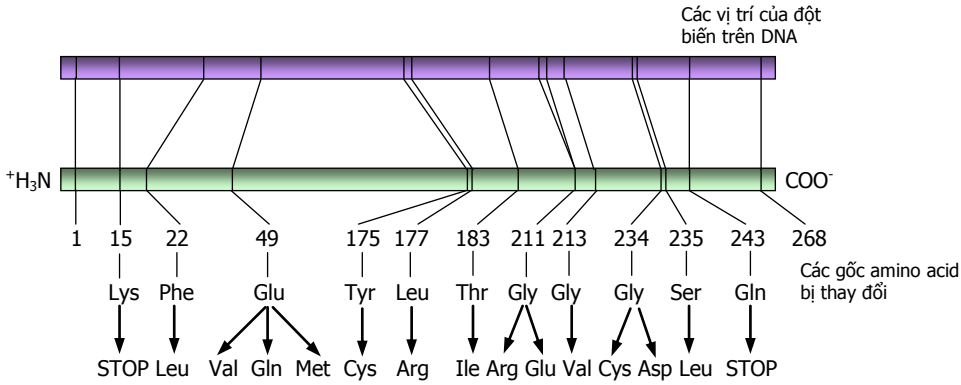
Đột biến được biểu hiện bởi sự thay thế vị trí của một amino acid này bằng một amino acid khác.

4. Sự tương quan đồng tuyến tính gen-polypeptide

4.1. Đột biến tryptophan synthetase-sự đồng tuyến tính giữa gen và chuỗi polypeptide

Nghiên cứu trên enzyme tryptophan synthetase xúc tác cho phản ứng tổng hợp tryptophan của *E. coli* người ta nhận thấy có nhiều đột biến xảy ra trên cùng một gen mã hóa cho tryptophan synthetase.

Thực hiện tái tổ hợp trong gen (nguyên tắc là gen ở các vị trí càng xa nhau trên nhiễm sắc thể càng dễ tái tổ hợp), người ta đã nhận được các dạng biến dị có tính chất khác nhau, và tính được khoảng cách tương đối giữa những điểm khác nhau của đột biến đã được xác định. Vị trí biến dị trên thể nhiễm sắc tương ứng với vị trí của amino acid trên chuỗi polypeptide. Như vậy, có thể cho rằng có sự đồng tuyến tính giữa gen và chuỗi polypeptide (Hình 3.1).



Hình 3.1. Tương quan đồng tuyến tính giữa gen và enzyme tryptophan synthetase của *E. coli* thông qua các vị trí đột biến và các gốc amino acid bị thay đổi

Nhiều dạng đột biến của tryptophan synthethase đã được tạo ra. Bằng cơ chế tái tổ hợp, những khoảng cách tương đối giữa những điểm khác nhau

của đột biến đã được xác định. Sản phẩm protein của mỗi dạng đột biến đã được phân tích, và những thay đổi các amino acid khác cũng được xác định. Người ta đã tìm thấy mối tương quan hoàn toàn giữa những khoảng cách của các đột biến được tìm thấy trên gen với khoảng cách của amino acid bị thay đổi trong phân tử protein.

4.2. Đột biến

4.2.1. Khái niệm

Một gen (DNA) có 4 loại base và một phân tử protein có 20 loại amino acid[1], nhưng giữa chúng có mối tương quan như thế nào. Đầu tiên, người ta cho rằng một base qui định một amino acid, nhưng những tính toán cho thấy không hợp lý. Vì chỉ có 4 base trong DNA và 20 amino acid trong protein, cho nên mỗi codon phải chứa ít nhất 3 base. Hai base cũng không thể làm thành một codon bởi vì chỉ có $4^2 = 16$ cặp hợp lý của 4 base. Nhưng 3 base thì có thể bởi vì sẽ có $4^3 = 64$ bộ ba hợp lý. Vì số lượng bộ ba hợp lý lớn hơn 20, cho nên sẽ có trường hợp một vài codon chỉ định cùng một amino acid. Ví dụ: UCU, UCC, UCA, UCG, AGU và AGC đều cùng mã hóa cho serine.

Từ đó, người ta đưa ra khái niệm mã di truyền (tín hiệu di truyền). Mã di truyền cho phép đọc thứ tự trên DNA để biết thứ tự trên chuỗi polypeptide. Mã di truyền không mơ hồ, có nghĩa với một trình tự chẳng hạn ATA ta biết nó ghi mã cho một amino acid gì, và cũng thấy rằng có nhiều mã di truyền xác định cho một amino acid (Bảng 3.4).

4.2.2. Đột biến điểm

Là đột biến chỉ tác động một vị trí, nói rõ hơn đó là một base. Khi thay đổi một base trên DNA sẽ tạo ra sự thay đổi một amino acid (Hình 3.2).

---

[1] Hai mươi amino acid được tìm thấy trong các phân tử protein là: Alanine (Ala), Arginine (Arg), Asparagine (Asn), Aspartic acid (Asp), Cysteine (Cys), Glutamic acid (Glu), Glutamine (Gln), Glycine (Gly), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Phenylalanine (Phe), Proline (Pro), Serine (Ser), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr) và Valine (Val).

Đột biến dĩ nhiên xảy ra trên DNA và được sao lại trên mRNA trong phiên mã, rồi trên protein trong dịch mã.

Bảng 3.4. Mã di truyền chung

Vị trí thứ nhất	Vị trí thứ hai				Vị trí thứ ba
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
U	Leu (L)	Ser (S)	STOP	STOP	A
U	Leu (L)	Ser (S)	STOP	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
C	Leu (L)	Pro (P)	His (H)	Arg (R)	C
C	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
C	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
A	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
A	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
G	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
G	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

Chú thích

Những đơn vị mã (codon) được đọc theo chiều 5'→3'.

STOP: codon kết thúc (còn gọi là vô nghĩa).

Đột biến điểm có các dạng sau:

- Đột biến sai nghĩa. Thay đổi một amino acid trong protein, có thể dẫn đến một trong ba kết quả sau:

+ Không hậu quả nào cả, vì amino acid không nằm trong vị trí hoạt động hoặc không có vai trò trong cấu trúc enzyme.

+ Có biến đổi nhẹ ở chuỗi polypeptide sẽ tạo ra tính mẫn cảm yếu với nhiệt, làm giảm sự ổn định chuỗi polypeptide.

+ Mất hẳn hoạt tính enzyme nếu đúng ngay vị trí hoạt động của enzyme đó.

- Đột biến vô nghĩa. Thay đổi một base. Nếu đó là một codon vô nghĩa sẽ làm ngừng kéo dài (tổng hợp) chuỗi polypeptide ở vị trí amino acid này. Tức là nếu codon này nằm ở đầu sẽ không có chuỗi polypeptide hoạt động.

- Đột biến acridine hoặc đột biến dịch khung. Đột biến này do chất acridine màu da cam tạo ra (hoặc còn gọi là đột biến dịch khung, frameshift, do thêm vào hoặc bớt đi một base) (Hình 3.2 E và D). Như vậy, một đột biến trên khung đọc khi thêm vào (C) hoặc mất đi (A) thường sẽ dẫn đến xuất hiện một **codon stop** làm ngừng chuỗi polypeptide và enzyme sẽ không có hoạt tính.

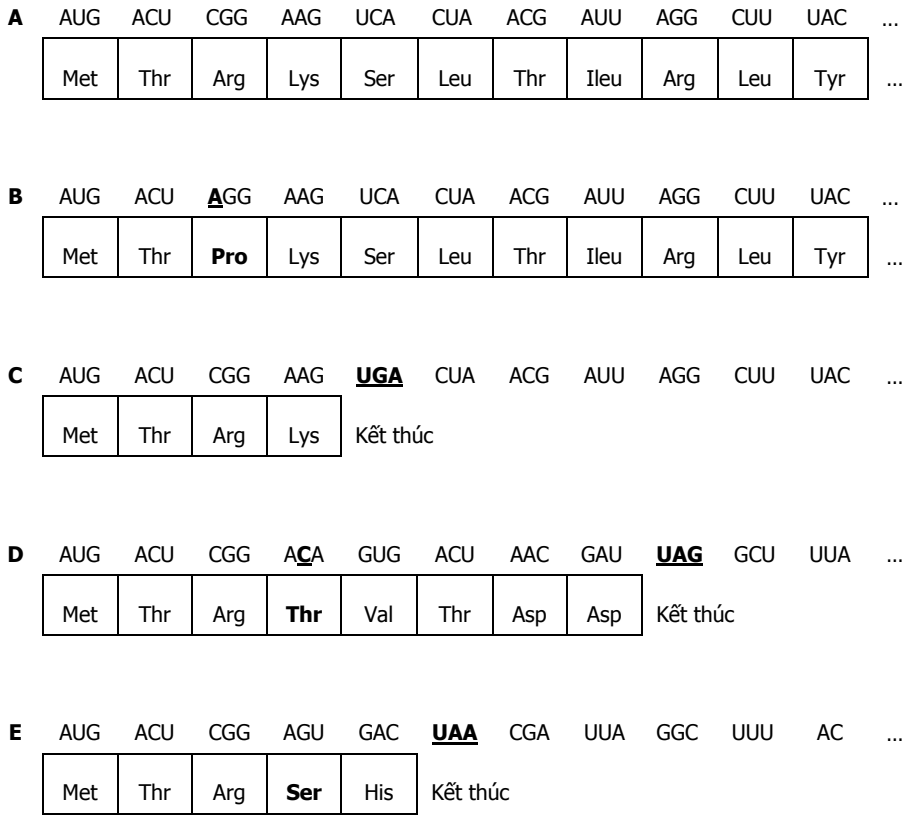
4.2.3. Đột biến kìm hãm

Đến nay, người ta nhận thấy mọi sai lệch trong việc tổng hợp protein nếu có đều xảy ra từ DNA, còn quá trình diễn ra từ RNA đến polypeptide luôn luôn đúng. Nghiên cứu một vài kiểu protein đột biến ta thấy:

- Đột biến sai nghĩa. Làm xuất hiện một bất thường trong trình tự amino acid. Kết quả protein mất hoạt tính. Hoạt tính này có thể được phục hồi, hoặc do một đột biến ngược để cho lại protein cấu trúc ban đầu.

- Đột biến vô nghĩa. Làm mất đi một phần chuỗi polypeptide, phần còn lại không có hoạt tính, và hoạt tính này có thể có lại được nhờ đột biến trong một codon đã bị đột biến.

Thông thường, những gen kìm hãm đột biến vô nghĩa không nằm ở gần vị trí của đột biến ấy. Đó là những gen làm biến đổi hệ thống dịch mã khi tổng hợp protein.



Hình 3.2. Các dạng đột biến điểm

A: trình tự các codon và các amino acid tương ứng ở dạng tự nhiên.

B: thay đổi một base (C thành A) làm thay đổi một amino acid (Arg thành Pro) gây nên đột biến sai nghĩa.

C: thay đổi một base (C thành G) sinh ra một codon vô nghĩa (UGA).

D: thêm một base (C) gây nên đột biến acridine hoặc đột biến dịch khung, có sự dời khung đọc.

E: mất một base (A), gây nên đột biến acridine, chấm dứt đọc.

5. Lý thuyết trung tâm của sinh học phân tử

Tổng hợp protein trong tế bào có các đặc điểm sau:

- Các phân tử thông tin như nucleic acid và protein được tổng hợp theo khuôn. Tổng hợp theo khuôn vừa chính xác vừa ít tốn enzyme. Tuy

nhiên, căn cứ vào hàng loạt tính chất hóa học các protein không thể làm khuôn mẫu cho sự tổng hợp của chính chúng. Vì vậy, khuôn mẫu để tổng hợp nên protein không phải là protein.

- Sinh tổng hợp protein tách rời về không gian với chỗ chứa DNA. Nhiều nghiên cứu cho thấy tổng hợp protein có thể xảy ra khi không có mặt DNA. Sự kiện này thể hiện rõ ràng nhất ở những tế bào eukaryote. Trong những tế bào này, hầu như toàn bộ DNA tập trung ở nhiễm sắc thể trong nhân, còn tổng hợp protein chủ yếu diễn ra ở tế bào chất. Tảo xanh đơn bào *Acetabularia* khi bị cắt mất phần chứa nhân vẫn tổng hợp được protein và sống vài tháng nhưng mất khả năng sinh sản. Rõ ràng, nơi chứa DNA mang thông tin di truyền và chỗ sinh tổng hợp protein tách rời nhau về không gian.

- DNA không phải là khuôn mẫu trực tiếp để tổng hợp protein, do đó phải có chất trung gian chuyển thông tin từ DNA ra tế bào chất và làm khuôn để tổng hợp protein. Chất đó phải có cả trong nhân và tế bào chất với số lượng phụ thuộc vào mức độ tổng hợp protein.

- Chất trung gian đó được xem chính là RNA nhờ các đặc điểm sau:

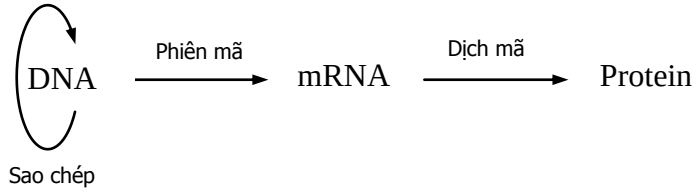
+ RNA được tổng hợp ngay ở trong nhân có chứa DNA, sau đó nó đi vào tế bào chất cho tổng hợp protein.

+ Những tế bào giàu RNA tổng hợp protein nhiều hơn.

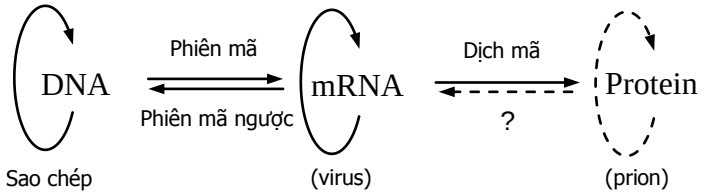
+ Về phương diện hóa học RNA gần giống DNA: chuỗi polyribonucleotide thẳng cũng chứa 4 loại ribonucleotide A, G, C và uracil (U). Nó có thể nhận được thông tin từ DNA qua bắt cặp bổ sung.

Nói chung, trong tế bào không thể tìm thấy chất nào khác ngoài RNA có thể đóng vai trò trung gian cho tổng hợp protein. Mối quan hệ này chính là thông tin di truyền đi từ DNA qua RNA rồi đến protein và được biểu diễn ở hình 3.3. Mối quan hệ này còn được gọi là **lý thuyết trung tâm (central dogma)**, được Crick đưa ra từ 1956 đến nay về căn bản vẫn đúng.

Vào những năm 1970, người ta đã phát hiện quá trình phiên mã ngược từ RNA tổng hợp nên DNA nhờ enzyme reverse transcriptase. Đến nay, việc sao chép (tổng hợp) RNA trên khuôn mẫu RNA cũng đã được chứng minh ở nhiều loại virus. Ngoài ra, thông tin từ protein cũng có thể được truyền sang protein (prion của bệnh bò điên). Riêng dòng thông tin từ protein ngược về mRNA/DNA thì chưa được tìm thấy (Hình 3.4).



Hình 3.3. Lý thuyết trung tâm của Crick



Hình 3.4. Những bổ sung mới vào lý thuyết trung tâm của Crick

6. DNA và mã di truyền

Chúng ta đã biết có sự liên quan đồng tuyến tính giữa DNA và phân tử protein, từ đó dễ dàng dự đoán rằng trình tự đặc hiệu của các amino acid trên phân tử protein sẽ được mã hóa bằng nhóm các nucleotide trên phân tử DNA. Có tất cả 4 loại base, nếu các base có nhóm đôi tức 2 nucleotide mã hóa cho một loại amino acid thì tất cả chỉ có 16 tổ hợp, không đủ cho 20 loại amino acid. Như vậy, đơn vị mã hóa (codon) phải gồm 3 nucleotide (xem phần 4.2).

Vấn đề tiếp theo là xác định chính xác các codon nào mã hóa cho từng amino acid. Nirenberg và Matthaei đã sử dụng enzyme để tổng hợp nhân tạo RNA. Khi dùng chỉ một loại nucleotide là U sẽ nhận được RNA là poly(U), nếu chỉ dùng A sẽ nhận được poly(A).

Năm 1961, Nirenberg và Matthaei đã dùng poly(U) thay cho khuôn mẫu mRNA để tổng hợp protein trong hệ thống vô bào (có amino acid, enzyme tổng hợp protein, nhưng không có DNA...), sản phẩm thu được là chuỗi polypeptide polyphenylalanine chỉ chứa một loại amino acid là phenylalanine. Điều đó chứng tỏ codon UUU mã hóa cho phenylalanine. Đây là codon đầu tiên được xác định. Sau đó, họ cũng chứng minh được rằng AAA mã hóa cho lysine, GGG cho glycine và CCC cho proline.

Năm 1964, Khorana tìm ra phương pháp tổng hợp mRNA nhân tạo với trình tự lặp lại (như AAG AAG AAG...) và nhờ nó giải quyết xong các vấn đề còn chưa rõ ràng.

Bảng mã di truyền (Bảng 3.4) cho thấy trong 64 codon, có 3 codon UAA, UAG, UGA không mã hóa cho amino acid được gọi là vô nghĩa (non-sense), đồng thời là codon kết thúc (termination) tức dấu chấm câu, chấm dứt chuỗi polypeptide.

Mã di truyền có tính suy biến (degeneration) tức một amino acid có nhiều codon mã hóa, chỉ trừ methionine và tryptophane chỉ có một codon (tương ứng là ATG và TGG). Các codon đồng nghĩa tức mã hóa cho cùng một amino acid thường có hai base đầu tiên giống nhau, nhưng khác nhau ở cái thứ ba. Ví dụ: CCU, CCC, CCA và CCG tất cả đều mã hóa cho proline. Trên thực tế, U và C luôn luôn tương đương nhau ở vị trí thứ ba, còn A và G tương đương nhau trong 14 trên 16 trường hợp.

Trừ một số ngoại lệ, mã di truyền có tính phổ biến (universal) tức toàn bộ thế giới sinh vật có chung bộ mã di truyền.

III. Cấu trúc và chức năng của gen

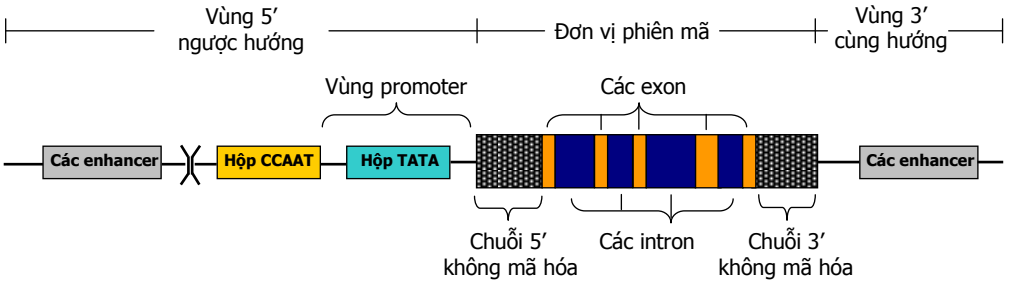
Khi nghiên cứu các quy luật di truyền Mendel và học thuyết di truyền nhiễm sắc thể, gen được quan niệm như một điểm trên nhiễm sắc thể, vừa là đơn vị chức năng xác định một tính trạng, vừa là đơn vị đột biến, vừa là đơn vị tái tổ hợp. Cùng với sự phát triển của di truyền học, khái niệm về gen được cụ thể hóa thêm, cấu trúc và chức năng của gen được hiểu chi tiết hơn.

1. Cấu trúc gen

Một gen thường được xem gồm có hai phần, phần mang mã di truyền được phiên mã sang phân tử mRNA và phần DNA làm nhiệm vụ điều khiển hoạt động của gen (Hình 3.5). Cả hai phần đều có cấu trúc cơ bản khá giống nhau ở mọi sinh vật prokaryote và eukaryote. Tuy nhiên, gen eukaryote còn có cấu trúc đặc thù liên quan đến các cơ chế kiểm soát khác nhau đối với hoạt động của gen.

Cấu trúc một gen điển hình nói chung đều có promoter, vị trí để RNA polymerase hoạt động khởi đầu phiên mã. Đôi khi nằm rất xa gen, có thể thấy các enhancer (vùng tăng cường) hoặc silencer (vùng ức chế) có vai trò

liên quan đến quá trình phiên mã. Độ dài của một gen thay đổi tùy theo số lượng và độ dài của các intron chứa trong nó.



Hình 3.5. Cấu trúc chung của một gen

Vùng DNA mang mã di truyền sẽ được phiên mã sang phân tử mRNA. Quá trình này thực hiện theo chiều 5'→3' trên sợi mRNA đang được tổng hợp. Không phải mọi phiên mã di truyền trên phân tử mRNA đều được dịch mã sang phân tử protein. Hai đầu 5' và 3' của phân tử mRNA gồm một số nucleotide không được dịch mã mà lại liên quan đến tính bền vững của phân tử mRNA hoặc tham gia kiểm soát quá trình dịch mã. Hai đoạn này gọi là vùng không dịch mã 5' và 3' (untranslated region). Vùng không dịch mã 5' nằm trước điểm khởi đầu dịch mã (3 nucleotide AUG mã hóa cho methionine đầu tiên của chuỗi polypeptide) và vùng không dịch mã 3' nằm sau điểm kết thúc dịch mã (stop codon có thể là UAA, UGA và UAG). Do tế bào prokaryote không có cấu trúc nhân nên quá trình phiên mã (tổng hợp mRNA) và dịch mã (tổng hợp protein) xảy ra đồng thời. Còn phân tử mRNA của eukaryote được phiên mã trong nhân, sau đó phải cắt bỏ intron và gắn các exon lại, chịu biến đổi tại các đầu 5' và 3' trước khi vận chuyển ra ngoài tế bào chất để dùng làm khuôn mẫu tổng hợp protein.

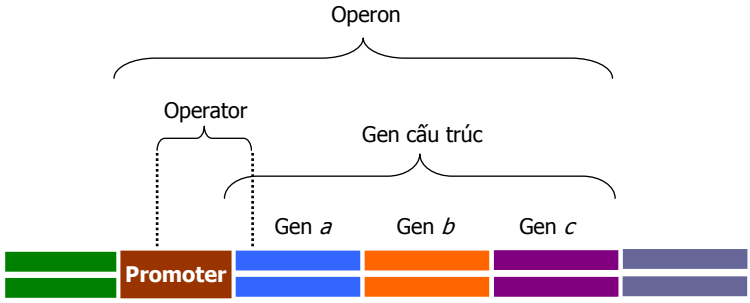
Hoạt động của một gen được đánh giá thông qua quá trình phiên mã (tổng hợp mRNA) và quá trình dịch mã (tổng hợp protein). Hoạt động này được kiểm soát rất chặt chẽ bằng các cơ chế khác nhau ở mọi giai đoạn, như bắt đầu và kết thúc phiên mã, quá trình biến đổi mRNA, quyết định tính bền vững và kiểm tra lại thông tin di truyền trên các phân tử này... Do cấu trúc

sắp xếp các gen prokaryote khác với gen eukaryote nên sự phối hợp giữa các cơ chế điều khiển mang tính chất riêng biệt cho từng loại genome.

Các gen prokaryote thường sắp xếp nằm gần nhau và chịu sự điều khiển chung của một promoter, tức là chúng được phiên mã sang cùng một phân tử mRNA. Cấu trúc này được gọi là operon. Như vậy, một operon gồm hai hay nhiều gen nằm cạnh nhau trên một nhiễm sắc thể. Thông thường, đó là các gen cùng tham gia vào một con đường chuyển hóa, ví dụ như các gen mã hóa cho các enzyme cần thiết cho quá trình chuyển hóa glucose.

Do có chung promoter điều khiển cho mọi gen nằm trong một operon cho nên chỉ có một loại phân tử mRNA được tổng hợp từ một operon (mang thông tin di truyền của tất cả các gen nằm trong đó). Nói cách khác, quá trình phiên mã của các gen trong một operon xảy ra đồng thời và phân tử mRNA đặc trưng cho operon được gọi là mRNA-polycistron.

Tuy nhiên, điều cần ghi nhớ là quá trình dịch mã trên các phân tử mRNA-polycistron xảy ra hoàn toàn độc lập với nhau. Mỗi đoạn tương ứng với một gen trên phân tử này đều có vị trí bám của ribosome, có mã bắt đầu và kết thúc tổng hợp chuỗi polypeptide riêng biệt. Do đó, tốc độ tổng hợp các protein trên các phân tử mRNA-polycistron hoàn toàn khác nhau (Hình 3.6).



Hình 3.6. Cấu trúc operon trong genome vi khuẩn. Một operon là một đơn vị phiên mã đơn bao gồm một chuỗi các gen cấu trúc (structural genes), một promoter và một operator.

2. Sự phân chia nhỏ của gen

Khái niệm locus được đưa ra để chỉ vị trí của gen trên nhiễm sắc thể, là vị trí của tất cả các allele của dãy đa allele. Bản thân hiện tượng đa allele

cho thấy gen có cấu tạo phức tạp, sự biến đổi của gen có thể dẫn đến nhiều trạng thái allele khác nhau.

2.1. Hiện tượng allele giả

Theo quan niệm cổ điển gen là đơn vị tái tổ hợp. Nếu cá thể mang hai allele lặn a_1/a_2 của một dãy đa allele sẽ tạo thành hai loại giao tử là a_1 và a_2, lai phân tích với bố mẹ đồng hợp tử lặn sẽ chỉ cho kiểu hình đột biến a_1 và a_2 mà không có dạng tái tổ hợp hoang dại. Ví dụ:

Bố mẹ	a_1/a_2	×	a_1/a_1
Giao tử	a_1 và a_2		a_1
Các con lai	a_1/a_1 và a_2/a_1 - cả hai đều là kiểu hình đột biến		

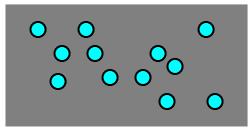
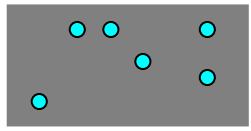
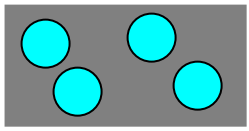
Tuy nhiên, nhiều thí nghiệm cho thấy nếu tăng số cá thể thí nghiệm lên 10.000 hoặc 100.000, thì có thể phát hiện có dạng kiểu hình hoang dại do tái tổ hợp.

Ví dụ: Trường hợp locus mắt quả trám ở ruồi giấm, có 18 allele. Khi tăng số cá thể nghiên cứu lên nhiều lần, người ta phát hiện các allele xếp thành 3 nhóm A, B và C. Các allele của cùng một nhóm, khi lai lẫn nhau, không cho kiểu hình tái tổ hợp hoang dại mắt bình thường, mà chỉ có kiểu hình mắt quả trám. Nhưng lai allele của nhóm này với allele của nhóm khác sẽ có xuất hiện kiểu hình hoang dại do tái tổ hợp. Hiện tượng này được gọi là allele giả.

Hiện tượng allele giả cho thấy gen phân chia nhỏ về mặt tái tổ hợp, có thể xảy ra tái tổ hợp giữa các phần trong gen. Lúc đầu, hiện tượng allele giả được coi là trường hợp ngoại lệ, nhưng khi tăng số cá thể nghiên cứu lên nhiều lần thì rõ ràng đó là hiện tượng phổ biến. Nó được tìm thấy ở nhiều đối tượng khác nhau như nấm men *S. cerevisiae*, ngô, bồ câu, chuột, bacteriophage...

2.2. Locus *rII* của bacteriophage T_4

Nghiên cứu chi tiết về các đột biến *rII* của bacteriophage T_4 đã làm sáng tỏ hơn về cấu trúc gen. Bacteriophage T_4 ở dạng hoang dại r^+ có khả năng xâm nhiễm đồng thời hai chủng *E. coli* B và K, trong khi các đột biến *rII* chỉ xâm nhiễm chủng B mà không xâm nhiễm chủng K (Hình 3.7).

	Chủng vi khuẩn *E. coli*	
	B	K
Dạng hoang dại	 Vết tan (plaque) nhỏ	 Vết tan nhỏ
Đột biến *rII*	 Vết tan lớn	Không có vết tan

Hình 3.7. Kiểu hình của các đột biến *rII* của phage T_4

Benzer (1955) đã thu được vài nghìn đột biến *rII* có nguồn gốc độc lập với nhau. Ông cho lai các đột biến này với nhau và căn cứ vào sự xuất hiện các dạng tái tổ hợp hoang dại r^+ mà lập bản đồ các điểm đột biến (mutation sites).

Trước thí nghiệm của ông, *rII* được coi là một locus. Tuy nhiên, thí nghiệm của ông đã cho thấy các đột biến xếp thành hai nhóm *rII*A và *rII*B. Lai các đột biến *rII*A × *rII*B sẽ có r^+, nhưng lai *rII*A × *rII*A và *rII*B × *rII*B thì kiểu hình đột biến là *r*.

Cho đến nay, chúng ta định nghĩa một gen là nhờ dựa trên kiểu hình đột biến và vị trí trên bản đồ của nó. Bacteriophage là một mô hình di truyền đơn giản (genome của *E. coli* dài khoảng 4.600.000 bp, trong khi bacteriophage T_4 là 165.000 bp và bacteriophage λ khoảng 46.500 bp), chúng có thể sinh sản một số lượng lớn rất nhanh (10^{10} hoặc hơn thế) và dễ dàng phân tích. Các thí nghiệm thực hiện với đột biến *rII* của T_4 được thiết lập dựa trên cơ sở sau:

- Các gen có một phạm vi và ranh giới hạn chế.

- Các gen có thể chia được, có thể có sự tái tổ hợp giữa hai allele trong một gen đơn.

- Hoạt động của gen có thể được phân tích bởi sự phân tích bổ sung.

Kết quả thí nghiệm cho thấy, gen có thể phân chia nhỏ về mặt đột biến. Các đoạn *rII*A và *rII*B được gọi là cistron, đơn vị chức năng nhỏ nhất đảm bảo khả năng xâm nhiễm chủng K. Thuật ngữ cistron thực chất là gen, ngày nay nó chỉ có tính chất lịch sử, ít được sử dụng. Theo quan niệm hiện nay, *rII*A và *rII*B là hai locus. Hai khái niệm mới được đưa ra là muton-đơn vị đột biến và recon-đơn vị tái tổ hợp.

Benzer đã tìm thấy 2.000 điểm đột biến trên đoạn gen được nghiên cứu, chúng phân bố không đều nhau, có những điểm tập trung nhiều đột biến hơn. Chiều dài gen khoảng 900 nucleotide. Đơn vị đột biến muton ở đây tương ứng với 900/2.000. Số đột biến ghi nhận có thể thấp hơn so với thực tế nên muton có thể tương ứng với một cặp nucleotide. Giống như vậy recon có thể tương ứng với một cặp nucleotide.

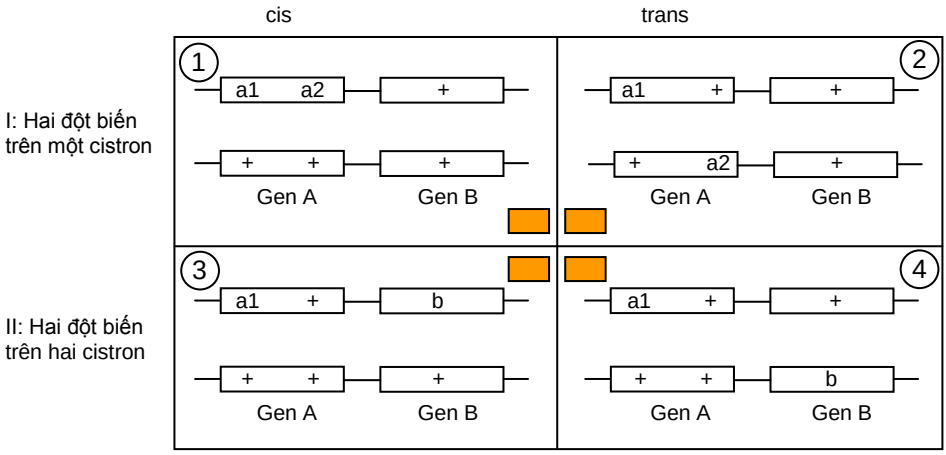
Tóm lại, gen là đơn vị chức năng, có thể chia nhỏ bởi các đơn vị đột biến tái tổ hợp.

3. Thử nghiệm chức năng allele

Muốn nghiên cứu cấu trúc bên trong một gen, phải tìm hiểu nhiều allele của gen đó. Nhiều đột biến có kiểu hình giống nhau nhưng không allele với nhau. Thử nghiệm chức năng allele được sử dụng để xác định xem hai đột biến có allele với nhau không. Đây chính là thử nghiệm mà Benzer dùng để lập bản đồ locus *rII*.

Thử nghiệm này còn được gọi là thử nghiệm bổ sung (complementary test) vì nó cho biết sai hỏng chức năng ở hai đột biến có bổ sung tức bù trừ cho nhau được không.

Phương pháp thử này cũng được gọi là thử nghiệm đều-lệch (*cis-trans* test). Sở dĩ như vậy là vì phép thử nghiệm này so sánh hiệu quả kiểu hình của các gen đột biến ở hai vị trí khác nhau trên nhiễm sắc thể tương đồng. Ở vị trí lệch (*trans*) các đột biến nằm trên hai nhiễm sắc thể, còn ở vị trí đều (*cis*) các đột biến nằm trên cùng một nhiễm sắc thể. Trường hợp sai hỏng ở hai gen khác nhau nên có thể bổ sung được, còn trường hợp sai hỏng ở cùng một gen không bù đắp được sẽ dẫn đến kiểu hình đột biến.



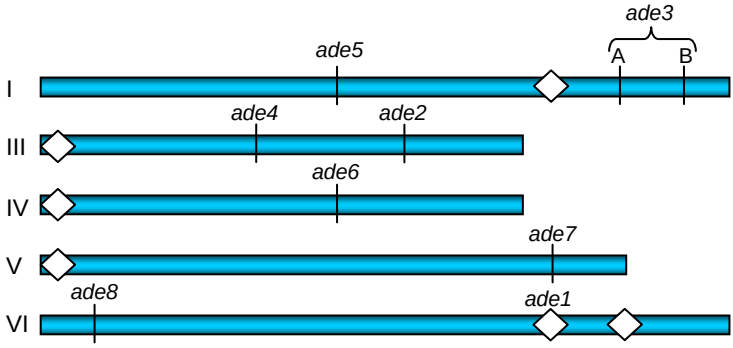
Hình 3.8. Thử nghiệm chức năng allele. I: có kiểu hình đột biến do sai hỏng cùng một gen nên không bù đắp được. II: có kiểu hình hoang dại do sai hỏng khác gen nên bù trừ được cho nhau.

4. Gen là đơn vị chức năng nhỏ nhất

Thử nghiệm chức năng allele có thể được thực hiện dễ dàng trên các đối tượng vi sinh vật với các đột biến hóa sinh, thường là các đột biến khuyết dưỡng (auxotroph mutant: mất khả năng tổng hợp một chất nào đó). Ví dụ: Ở nấm mốc *Neurospora crassa* có nhiều đột biến mất khả năng tổng hợp adenine (*Ade*$^-$). Các đột biến này dễ phát hiện vì có khuẩn lạc màu đỏ. Có hai dạng đột biến *Ade*$_x$ và *Ade*$_y$, nếu dị hợp tử *Ade*$_x$/*Ade*$_y$ có kiểu hình đột biến tức là cho khuẩn lạc màu đỏ, thì *Ade*$_x$ và *Ade*$_y$ là hai allele của một gen. Thử nghiệm chức năng allele cho thấy các đột biến *Ade* ở *N. crassa* tạo thành 9 nhóm. Điều đó chứng tỏ có 9 gen tổng hợp adenine ở loài nấm này: *ade*$_1$, *ade*$_2$, *ade*$_3$... trong đó *ade*$_3$ có hai locus nằm kề sát nhau là *ade*$_3$A và *ade*$_3$B (Hình 3.9).

Qua nghiên cứu các gen sản xuất adenine người ta nhận thấy quá trình tổng hợp adenine có liên quan đến 9 nhóm đột biến của gen này hoặc gen kia, và đều cùng có một kết quả là mất khả năng tổng hợp adenine. Như vậy, khái niệm gen không chỉ kiểm tra di truyền cả chu trình tổng hợp adenine, mà gen là đơn vị chức năng nhỏ nhất, kiểm tra một giai đoạn cơ bản nào đó

của chu trình. Nếu biến đổi di truyền làm sai hỏng chức năng đó sẽ dẫn đến xuất hiện đột biến mất khả năng tổng hợp adenine.



Hình 3.9. Vị trí các gen *ade* trên các nhiễm sắc thể của nấm mốc *Neurospora crassa*

Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. *NXB Giáo dục,* Hà Nội.

2. Phạm Thành Hổ. 2003. Di truyền học. *NXB Giáo dục,* Hà Nội.

3. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

4. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

5. Pierce BA. 2003. Genetics: A Conceptual Approach. *W.H. Freeman & Co*. New York, USA.

6. Watson JD, Hopkins NH, Roberts JW and Weiner AM. 2004. Molecular Biology of the Gene. *The Benjamin/Cummings Publishing Company, Inc*. California, USA.

Chương 4

Tái bản DNA

I. Chứng minh tái bản DNA theo cơ chế bán bảo thủ

1. Cơ chế tái bản bán bảo thủ

1.1. Cơ chế tái bản ở prokaryote

Đặc điểm cơ bản của sự tái bản đó là tái bản theo phương thức bán bảo thủ (semiconservative replication). Tái bản bán bảo thủ nghĩa là trong hai chuỗi của tất cả các phân tử DNA bao giờ cũng có:

- Một chuỗi của DNA cũ (từ một trong hai chuỗi của DNA mẹ).

- Một chuỗi của DNA mới (mới được tổng hợp).

Mỗi một lần tái bản đều có sự tách rời của hai chuỗi của DNA mẹ, đồng thời mỗi chuỗi mẹ tiến hành sao chép để cho một chuỗi con, chuỗi này sau đó lại kết hợp với chuỗi mẹ.

Vị trí mở xoắn kép và tổng hợp DNA mới cùng một lúc trên DNA gọi là chạc ba tái bản (replication fork) do cấu trúc của vùng tái bản có hình chữ Y. Sự tổng hợp DNA mới gắn liền với việc mở xoắn DNA cũ.

1.2. Cơ chế tái bản ở eukaryote

Sự tái bản ở tế bào eukaryote phức tạp hơn so với ở tế bào prokaryote nhưng cơ chế của sự tái bản ở eukaryote cũng tương tự như ở prokaryote và tiến hành theo các nguyên tắc:

- Hai hướng.

- Bổ sung, đối song song, theo chiều 5’→ 3’.

- Không liên tục ở một trong hai chuỗi.

- Cần những RNA primer.

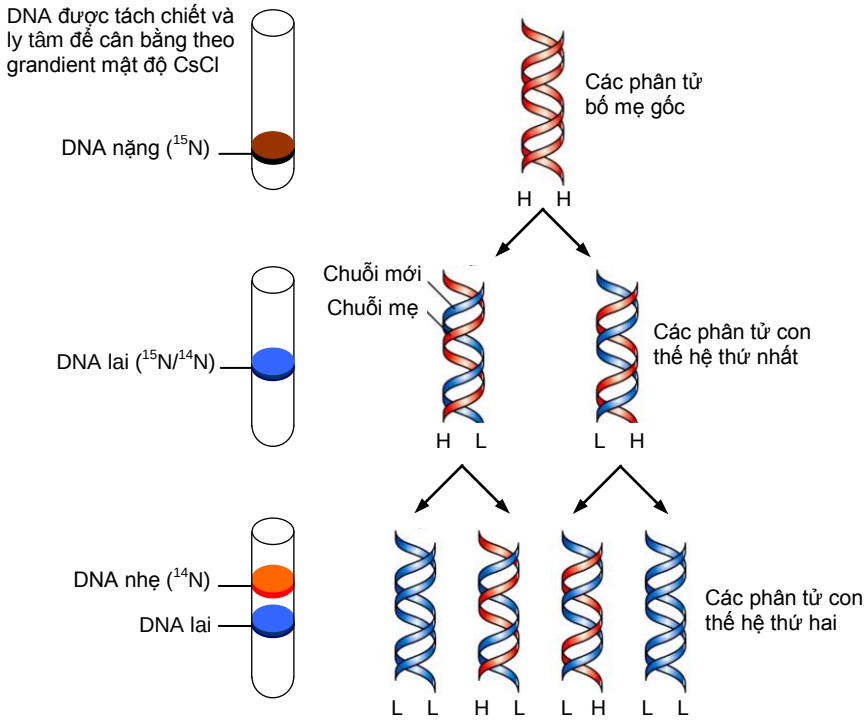
Tuy nhiên, có một số điểm khác như sau:

- Trong khi ở prokaryote chỉ có một điểm khởi đầu, thì sự tái bản ở eukaryote bắt đầu cùng một lúc ở nhiều điểm khởi đầu. Điều này là cần thiết do DNA của eukaryote có chiều dài rất lớn.

- Vận tốc phát triển của chạc ba tái bản ở eukaryote (khoảng 50 nucleotide/s) chỉ bằng 1/10 so với ở *E. coli*.

2. Thí nghiệm của Meselson và Stahl

Những thí nghiệm của Meselson và Stahl (1957) đã chứng minh lý thuyết tái bản DNA theo kiểu bán báo thủ (Hình 4.1). Các tác giả trên đã nuôi cấy *E. coli* trong nhiều thế hệ trong một môi trường chứa $^{15}NH_4Cl$ làm nguồn cung cấp nitrogen duy nhất. Bằng cách này, DNA được tổng hợp có ^{15}N (^{15}N là một chất phóng xạ nặng hơn chất phóng xạ thông thường ^{14}N). Ở một thời điểm nhất định (thời điểm 0), các tác giả này đã chuyển nuôi cấy vào một môi trường chứa $^{14}NH_4Cl$. Tiếp đến, sau từng thời gian đều đặn, họ phân tích DNA chiết xuất từ vi khuẩn bằng phương pháp ly tâm theo gradient CsCl.



Hình 4.1. Minh họa sự tái bản bán bảo thủ. Sơ đồ trình bày sự hợp thành các sợi đôi DNA sau 0, 1, và 2 vòng sao chép. H: chuỗi nặng (^{15}N), L: chuỗi nhẹ (^{14}N).

Kết quả thực nghiệm cho thấy:

- Ở thời điểm 0: chỉ có một phân tử tương ứng với DNA nặng ^{15}N.

- Sau một thế hệ trong môi trường chứa ^{14}N: những phân tử DNA gồm một chuỗi nặng ^{15}N (chuỗi mẹ) và một chuỗi nhẹ ^{14}N (mới được tổng hợp).

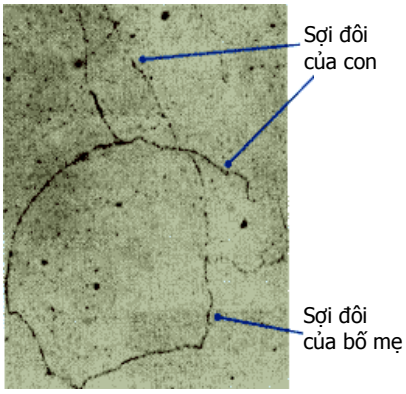
- Sau hai thế hệ trong môi trường chứa ^{14}N: có hai phân tử lai (gồm một chuỗi nặng và một chuỗi nhẹ) và hai phân tử đều gồm những chuỗi nhẹ không có chuỗi nặng.

II. Mô hình tái bản DNA-chạc ba tái bản

1. Mô hình tái bản

Mô hình tái bản được nghiên cứu trên thể nhiễm sắc của vi khuẩn *E. coli*, DNA có dạng mạch vòng sợi đôi (Hình 4.2). Để tự tái bản DNA phải tháo ra đơn giản ở một vị trí nhất định và nơi đó xuất hiện chạc ba tái bản (Hình 4.3).

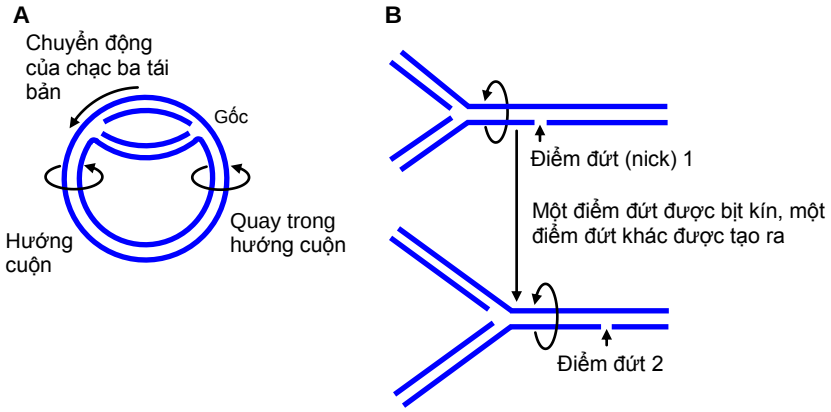
Thí nghiệm của Cairns. Sử dụng nucleotide được đánh dấu bằng đồng vị phóng xạ trong môi trường đang phân chia, ta sẽ biết được tiến trình tái bản do hạt bạc xuất hiện dưới kính hiển vi điện tử.



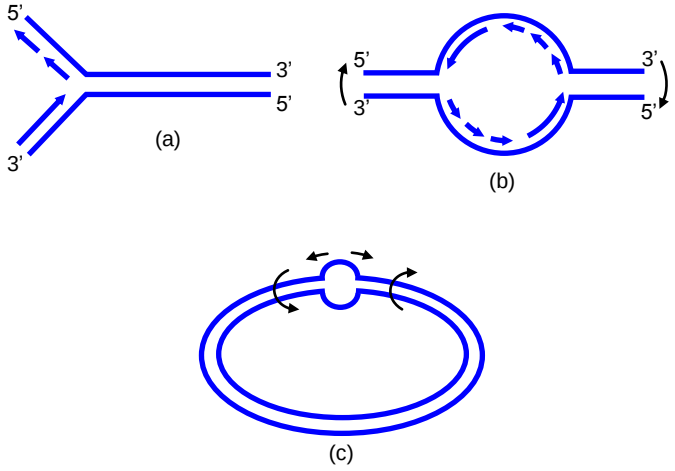
Hình 4.2. DNA dạng mạch vòng sợi đôi. Chiều dài thực tế 1,6 mm ($4,7\times10^6$ bp).

2. Chạc ba tái bản

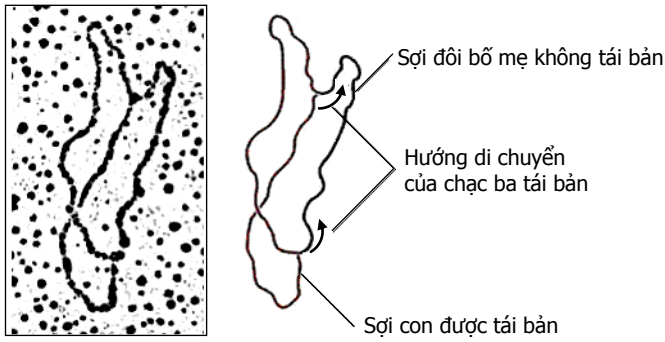
Hình 4.4 mô tả cấu trúc của chạc ba tái bản. Nhờ vào phương pháp phóng xạ ảnh tự ghi, người ta nhận thấy sự tái bản thực hiện theo hai hướng (bidirectional synthesis). Đồng thời cũng chứng minh được vi khuẩn *E. coli* (prokaryote) có duy nhất một điểm gốc tái bản, đó là điểm mà hai chuỗi xoắn kép của DNA mẹ được tách ra, tương ứng với hai chạc ba tái bản phát triển ngược chiều nhau. Chuỗi DNA sau khi tách ra được dùng làm khuôn mẫu cho sự tổng hợp DNA mới (Hình 4.5).



Hình 4.3. Sự tái bản của phân tử DNA sợi đôi mạch vòng của *E. coli*. A: sự chuyển động không cuộn lại của các nhánh trong quỹ đạo tái bản, không có các vị trí quay tự do, gây ra sự cuộn lại quá chặt của phần không được tái bản. B: cơ chế sợi đơn bị đứt (nick) phía trước của chạc ba tái bản cho phép sự quay xảy ra.

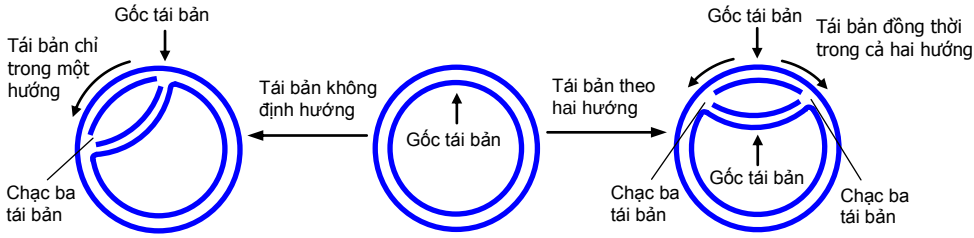


Hình 4.4. Sơ đồ cấu trúc của các chạc ba tái bản. (a): Chạc ba đơn trình bày sợi chủ (leading strand) được tổng hợp liên tục và sợi thứ (lagging strand) được tổng hợp gián đoạn. (b): Chạc ba đôi, phổ biến trong hầu hết mọi sự tái bản DNA của genome. (c): Các hướng hình học của sự tái bản DNA, mũi tên ngắn chỉ sự chuyển động dịch mã của chạc ba, mũi tên dài và cong chỉ sự quay vòng DNA cần thiết quanh các chạc ba.



Hình 4.5. Hình ảnh dưới kính hiển vi điện tử. Phân tử DNA mạch vòng nhỏ của *E. coli* có chiều dài thực tế 0,01 mm (3.000 bp) tái bản bằng kiểu θ. Các đoạn DNA bố mẹ và con được trình bày trong hình vẽ.

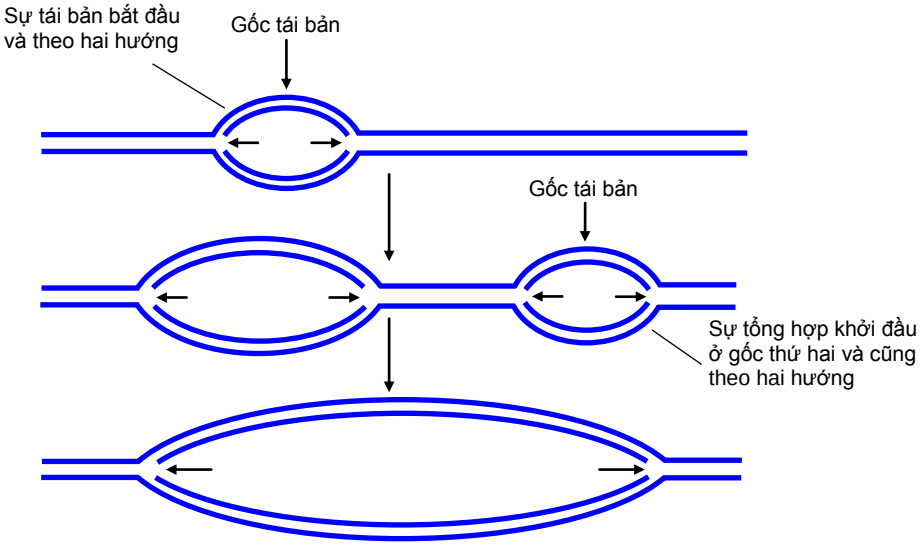
Hình 4.6 so sánh sự khác nhau giữa tái bản DNA không định hướng và tái bản theo hai hướng. Trong tái bản không định hướng, chỉ có một chạc ba tái bản. Trong khi tái bản theo hai hướng yêu cầu hai chạc ba tái bản. Mũi tên cong chỉ hướng chuyển động của các chạc ba. Hầu hết DNA tái bản theo hai hướng.



Hình 4.6. Tái bản DNA không định hướng và theo hai hướng

Hình 4.7 và 4.8 mô tả phương thức hợp nhất các vòng tái bản DNA ở ruồi giấm (*D. melanogaster*). Quá trình tái bản diễn ra đồng thời trên hàng chục ngàn vị trí khác nhau của phân tử DNA và tạo thành các vòng tái bản, các vòng tái bản sau đó sẽ mở rộng theo hai hướng để cuối cùng hợp nhất với nhau tạo thành hai phân tử DNA.

Hình 4.7. Hình ảnh dưới kính hiển vi điện tử của một đoạn nucleotide ở ruồi giấm. Phân tử DNA sợi đôi dài 30 kb cho thấy có 7 vòng tái bản.

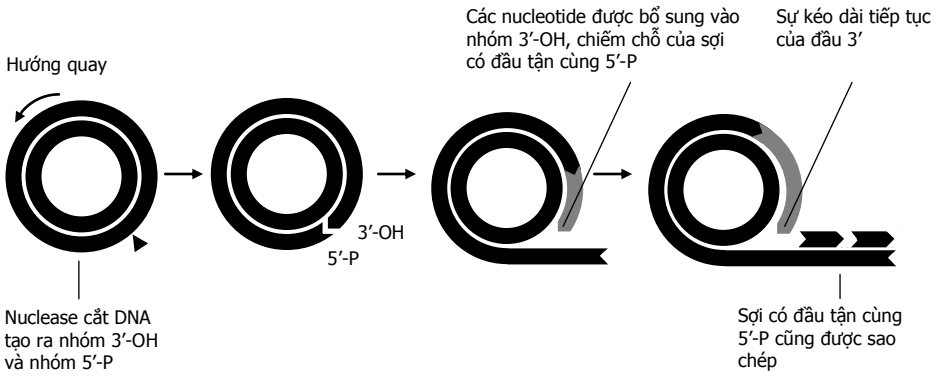


Hình 4.8. Phương thức hợp nhất các vòng tái bản DNA của ruồi giấm. Hai gốc tái bản được trình bày trên hình vẽ, các mũi tên nhỏ chỉ hướng chuyển động của các chạc ba tái bản.

3. Tái bản DNA theo vòng tròn quay

Trường hợp bacteriophage λ (thực khuẩn thể λ) có vật chất di truyền là một phân tử DNA mạch thẳng sợi đôi, khi ta chuyển chúng vào vi khuẩn thì các đầu dính kết DNA (*cos*) của nó gắn lại theo dạng vòng tròn. Sự dính

kết lại này là do hoạt động của enzyme DNA ligase giúp tạo lại dạng xoắn. Khi đó, sự tái bản DNA tiến hành theo cơ chế vòng tròn quay (Hình 4.9).



Hình 4.9. Tái bản vòng tròn quay ở bacteriophae λ. DNA được tổng hợp mới có màu nhạt. Sợi thay thế được tái bản trong các đoạn ngắn.

III. Bản chất xoắn của DNA-Các giai đoạn của sự tái bản

Hình 4.10 mô tả toàn bộ quá trình tái bản DNA. Quá trình này trải qua ba giai đoạn chính sau:

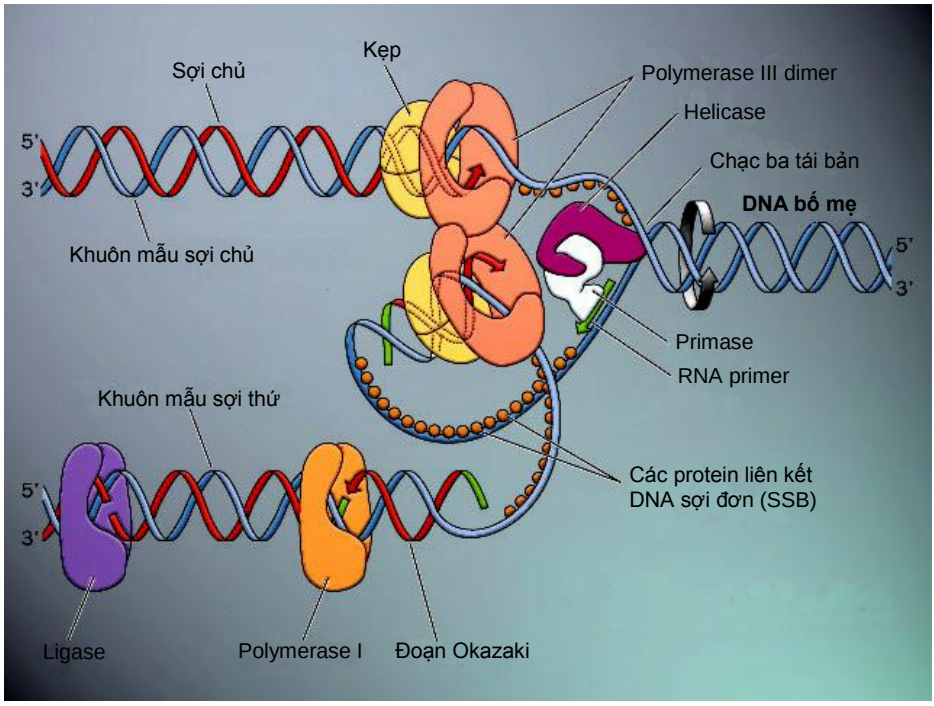
1. Mở xoắn

Trước tiên ta thấy quá trình mở xoắn của hai sợi DNA cần thiết phải có một enzyme rất quan trọng đó là helicase (còn gọi là enzyme mở xoắn).

Chẳng hạn, trên phân tử *E. coli* chỉ có một gốc tái bản (ký hiệu là *ori* C) dài 245 bp. Có ít nhất 8 enzyme hoặc protein tham gia vào giai đoạn khởi đầu của sự tái bản. Những enzyme-protein này mở xoắn DNA ở gốc *ori* C và thiết lập một phức hợp tiền mồi (prepriming complex) để chuẩn bị cho những phản ứng của giai đoạn sau.

Một phức hợp khoảng 20 protein Dna A được kết hợp với một vùng của *ori* C bắt đầu cho quá trình mở xoắn và khởi đầu sự tái bản. Phản ứng này cần ATP và một protein giống như histone của vi khuẩn (HU). Tiếp đó, protein Dna C giúp protein Dna B gắn vào vùng *ori* C, Dna B chính là helicase sẽ mở xoắn DNA theo hai hướng để tạo ra hai chạc ba tái bản. Các phân tử protein liên kết sợi đơn (single stranded binding protein, protein SSB) gắn vào chuỗi đơn DNA để ổn định chuỗi này. Enzyme gyrase (DNA topoisomerase II) mở xoắn để tạo siêu xoắn trái (–). Enzyme primase

(protein Dna G) sẽ xúc tác tổng hợp RNA primer để gắn vào khuôn mẫu DNA.



Hình 4.10. Quá trình tái bản DNA

2. Kéo dài-Tổng hợp chuỗi Okazaki

Giai đoạn kéo dài bao gồm sự tổng hợp cùng một lúc hai chuỗi DNA. Một chuỗi được tái bản cùng hướng phát triển của chạc ba sẽ liên tục (sợi chủ), còn chuỗi kia sẽ không liên tục bao gồm các đoạn Okazaki (sợi thứ). Các nucleotide gắn vào đầu 3' tự do, và kéo dài chuỗi ra nhờ enzyme DNA polymerase III theo chiều 5'→3'. Trong giai đoạn này, nhiều enzyme đã tham gia vào sự tổng hợp hai chuỗi DNA tại chạc tái bản. DNA helicase tiếp tục tách hai chuỗi DNA mẹ, DNA gyrase mở xoắn, protein SSB ổn định những chuỗi DNA đơn đã được tách ra, DNA ligase gắn các đoạn Okazaki trên sợi thứ.

Sinh tổng hợp DNA trong tế bào cùng một lúc diễn ra trên rất nhiều vị trí (20.000-60.000) suốt chiều dài khổng lồ của DNA khuôn mẫu. Ở các vị trí đó, enzyme helicase giúp DNA chuyển từ dạng siêu xoắn sang dạng dãn

(hai sợi DNA tách ra) với tốc độ 10 μm/giờ. Với tốc độ này, để helicase có thể chạy suốt chiều dài chuỗi DNA trong một tế bào cây đậu ngựa (DNA dài tổng cộng khoảng 9 m) phải mất hơn một thế kỷ. Trong thực tế, công việc này chỉ mất khoảng 15-40 giờ.

Chuỗi DNA xoắn kép như vậy được tách đôi ở các vị trí sẽ xảy ra sinh tổng hợp, hình thành các chạc ba. Ở chạc ba, cả hai chuỗi DNA mạch đơn đều được sử dụng làm khuôn mẫu cùng một lúc. Trên một chuỗi, sinh tổng hợp sẽ diễn ra theo chiều từ 5'→3' cùng chiều với chiều phát triển của chạc tái bản. Trên sợi còn lại, sinh tổng hợp vẫn diễn ra theo chiều 5'→3' nhưng ngược chiều với chiều phát triển của chạc tái bản, và chỉ thực hiện được từng đoạn ngắn độ 1.000 nucleotide (đoạn Okazaki). Trong khi di chuyển từng quãng, enzyme primase sẽ tổng hợp những đoạn RNA primer ngắn (11±1 nucleotide) để từ đó DNA được tiếp nối nhờ DNA polymerase III. Khi đoạn Okazaki mới được hoàn chỉnh, RNA primer được tách ra nhờ DNA polymerase I (hoạt tính exonulease 5'→3') và được thay thế bởi DNA cũng nhờ tác dụng của cùng enzyme. Các khe hở còn lại giữa các đoạn Okazaki được DNA ligase gắn và DNA trở lại dạng chuỗi xoắn kép: một phân tử DNA mới được hình thành (Hình 4.11).

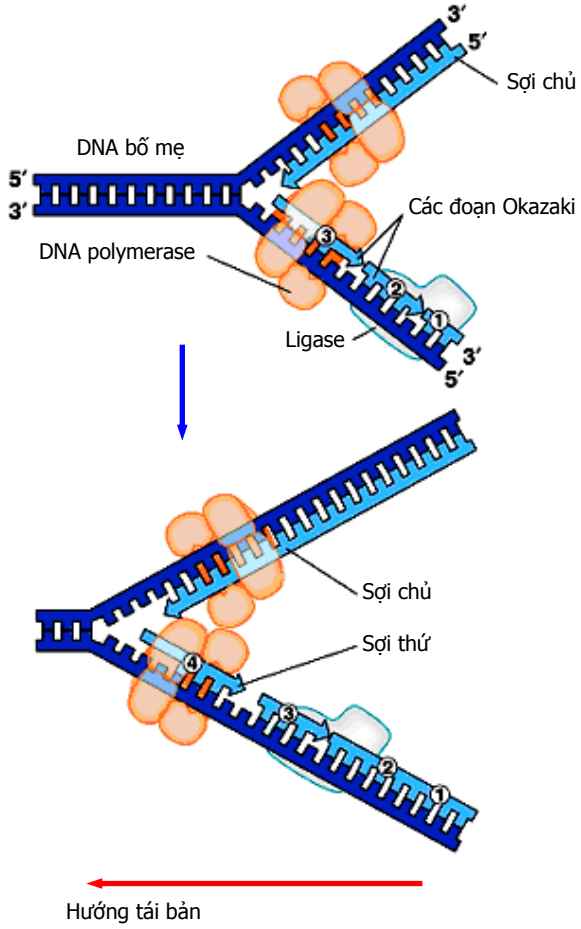
3. Kết thúc

Cuối cùng, hai chạc ba tái bản gặp nhau ở phía đối diện của nhiễm sắc thể vòng của *E. coli*. Người ta biết rất ít về những phản ứng của giai đoạn này. Có thể hoạt động của một loại DNA topoisomerase là cần thiết để tách hai phân tử DNA vòng đã được tổng hợp. Quá trình phân đôi hai phân tử DNA trong tế bào con khi phân chia tế bào cũng chưa được biết rõ.

IV. Khái niệm mồi

Tái bản DNA chỉ xảy ra khi có sự khởi động mồi. Thường mồi (primer) là đoạn RNA ngắn gắn trước đầu DNA.

Quan sát chạc ba tái bản ta thấy DNA polymerase gắn nhóm phosphate vào đầu 3'-OH tự do. Tuy nhiên, lúc đầu chưa có đầu 3'-OH tự do, do đó ở ngay điểm gốc tái bản enzyme RNA polymerase (enzyme primase) sẽ hoạt động tạo nên một đoạn mồi ngắn để có đầu 3'-OH tự do, nhờ đó DNA polymerase mới bắt đầu hoạt động tái bản.



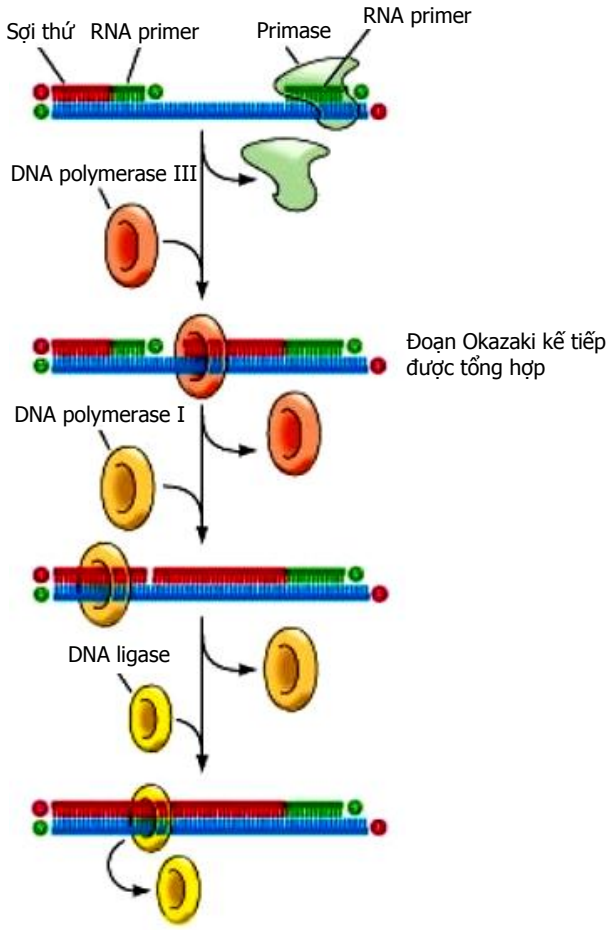
Hình 4.11. Hai sợi DNA mới được tổng hợp theo hai kiểu khác nhau

Nghiên cứu tái bản ở bacteriophage M13 (DNA một sợi vòng). Gọi DNA + (M13) là sợi (+). Từ sợi (+) tạo ra dạng tái bản (–), sợi (–) này sẽ dùng làm khuôn tạo ra sợi (+) khác.

DNA polymerase sẽ không hoạt động, nếu không có mồi ghép đôi vào sợi (+) đầu tiên. Enzyme RNA polymerase sẽ tổng hợp đoạn mồi bằng cách bổ sung một số nucleotide vào vị trí khởi đầu tái bản, tạo nên đầu 3'-OH tự do. RNA polymerase bao gồm các đơn vị α_2, β, β' và σ. Chất kháng sinh rifampicin có tác dụng ức chế đơn vị β, khi đó RNA polymerase sẽ bị ức chế và M13 không có hiện tượng tái bản. Như vậy, chính RNA polymerase giữ vai trò khởi động để sau đó DNA polymerase mới bắt đầu hoạt động tái bản

được. Vì vậy, đầu tiên phải có một đoạn mồi gắn với DNA bằng liên kết cộng hóa trị, và đoạn mồi này do RNA polymerase tổng hợp nên.

Tiếp tục, đoạn mồi RNA sẽ bị loại ra bởi DNA polymerase (khác với DNA polymerase tái bản ở trên). DNA polymerase này bắt đầu gắn các nucleotide vào đầu 3'-OH tự do, và sau cùng enzyme DNA ligase nối các đoạn tái bản này lại (Hình 4.12).



Hình 4.12. Tổng hợp các đoạn Okazaki. Quá trình này đòi hỏi sự gắn mồi, kéo dài, loại bỏ RNA, làm đầy các khoảng trống và hàn các điểm đứt. Primase tổng hợp đoạn mồi RNA, DNA polymerase III kéo dài đoạn mồi RNA thành đoạn Okazaki ở sợi thứ, DNA polymerase I sử dụng dịch mã điểm đứt để thay thế đoạn mồi RNA bằng DNA, và cuối cùng DNA ligase hàn điểm đứt.

Như vậy, việc tái bản DNA cần các enzyme sau:

- Helicase: mở xoắn để tách sợi DNA đang xoắn.
- SSB: gắn trên DNA một sợi để sợi luôn luôn ở tình trạng mở.
- RNA polymerase (primase): tác động hình thành đoạn mồi RNA.
- DNA polymerase I: loại bỏ đoạn mồi RNA.
- DNA polymerase III (khác với loại trên) tác động tổng hợp DNA bằng cách kéo dài đoạn mồi RNA.
- DNA ligase: gắn các đoạn Okazaki đã tổng hợp với nhau.

Từ các nghiên cứu về tái bản DNA của bacteriophage M13, người ta đã đưa ra sơ đồ giả thuyết về sự biến đổi DNA một sợi của M13 thành dạng sao chép tái bản (RF).

Và quan sát trên chạc ba tái bản đang phát triển, ta có thể theo dõi được sự tái bản liên tục và không liên tục của các đoạn DNA mới bổ sung cho sợi DNA khuôn mẫu.

V. Enzyme tái bản

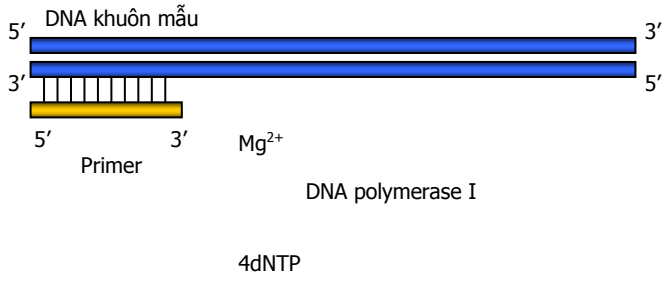
1. DNA polymerase

Đây là enzyme chủ yếu của sự tái bản, chịu trách nhiệm tổng hợp hai chuỗi DNA ở chạc tái bản. Có ba loại DNA polymerase khác nhau, được ký hiệu là I, II và II. Cơ chế và vai trò của các enzyme DNA polymerase cũng rất khác nhau.

$$(DNA)_n + dNTP \xrightarrow[\text{DNA polymerase}]{Mg^{2+}} (DNA)_{n+1} + PP_i$$

1.1. DNA polymerase I

Do Kornberg phát hiện năm 1955. DNA polymerase I của *E. coli* mang chuỗi polypeptide có khối lượng phân tử là 109.000 Da với ba hoạt tính riêng biệt. Nhiệm vụ chính của DNA polymerase I là xúc tác cho quá trình lắp ráp các dNTP từng cái một vào đầu 3’ tự do của đoạn mồi đang gắn trên DNA khuôn mẫu. Kết quả là sợi DNA mới sẽ dài dần về phía đầu 3’ (Hình 4.13).



Hình 4.13. Các thành phần cần thiết cho sinh tổng hợp DNA

DNA polymerase I của *E. coli* còn có hoạt tính exonuclease (hoạt tính cắt các chuỗi nucleotide ở các đầu tự do của DNA) 3'→5' xúc tác cho sự thoái hóa bậc thang từ đầu 3' của cả DNA sợi đôi và sợi đơn khi không có dNTPs. Trong trường hợp có dNTPs, hoạt tính exonuclease trên sợi đôi sẽ bị ức chế bởi hoạt tính polymerase. Trong quá trình tổng hợp DNA, hoạt tính exonuclease thực hiện chức năng đọc sửa bằng cách cắt bỏ những nucleotide lắp ráp sai.

Ngoài ra, enzyme này còn có hoạt tính exonulease 5'→3' xúc tác cho sự chuyển chỗ của các nucleotide từ đầu 5' của điểm đứt và bổ sung tức thời các nucleotide từ đầu 3' tạo ra chuyển động của điểm đứt dọc theo DNA.

1.2. DNA polymerase II và DNA polymerase III

Khởi đầu, enzyme DNA polymerase I được xem như giữ vai trò chủ yếu trực tiếp trong tái bản DNA. Nhưng sau đó, người ta quan sát thấy ở *E. coli* đột biến không có DNA polymerase I mà vẫn có sự tái bản, và phát hiện rằng chính DNA polymerase III giữ vai trò tái bản DNA. Nghiên cứu tính chất và so sánh ba loại enzyme trên người ta nhận thấy DNA polymerase III mới là enzyme thực sự điều khiển tổng hợp DNA. Trong khi đó DNA polymerase I chỉ có vai trò loại bỏ RNA mồi nhờ cắt đầu 5'→3' và thay vào chỗ RNA mồi bằng DNA khác, còn DNA polymerase II chưa rõ tác dụng, có lẽ nó tham gia vào quá trình sửa chữa (thay một đoạn DNA hỏng bằng một đoạn DNA bình thường) hoặc thay thế DNA polymerase I khi có khó khăn trong tổng hợp DNA.

Về tốc độ làm việc, các DNA polymerase rất khác nhau. Trong một giây, DNA polymerase I gắn được 10 dNTP trong khi DNA polymerase II chỉ gắn được 0,5 dNTP và DNA polymerase III gắn tới 150 dNTP.

2. Các topoisomer và DNA topoisomerase

2.1. Topoisomer

Giả sử có hai phân tử DNA mạch vòng có cùng trình tự nucleotide. Nhưng hai phân tử này có thể có số vòng (linking number-Lk) khác nhau trong phân tử. Số vòng ở đây được định nghĩa là số lần của một chuỗi DNA quấn xung quanh một chuỗi khác. Những trường hợp này được gọi là topoisomer. Topoisomer có các dạng sau:

2.1.1. Dạng lỏng lẻo (relaxed DNA)

Ở dạng này sức căng của xoắn kép là tối thiểu. Đó là dạng cấu trúc ổn định nhất của phân tử.

2.1.2. Dạng siêu xoắn (supercoiled DNA)

Trục của xoắn kép có thể cuộn xung quanh mình tạo thành một siêu xoắn. Có hai dạng siêu xoắn sau:

- Siêu xoắn dương (+). Số vòng tăng, xoắn kép xoắn cùng chiều (xoắn phải) tạo thành dạng siêu xoắn dương (positive supercoil).

- Siêu xoắn âm (–). Số vòng giảm, xoắn kép xoắn theo chiều ngược lại (chiều trái) tạo thành dạng siêu xoắn âm (negative supercoil),

Phần lớn những phân tử DNA trong tự nhiên ở dạng siêu xoắn âm.

2.2. DNA topoisomerase

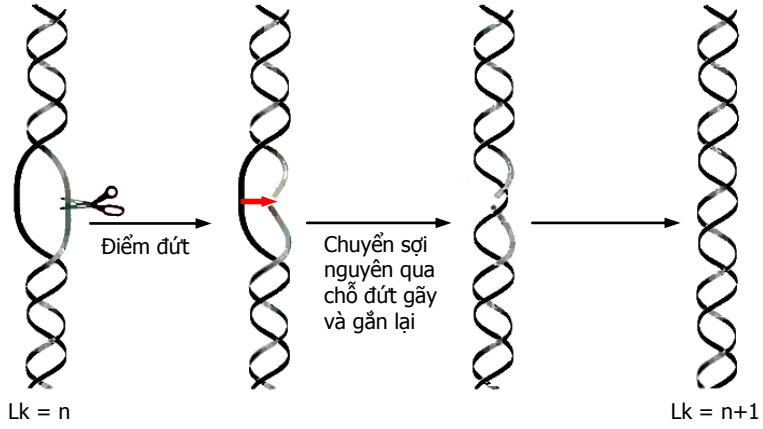
Enzyme DNA topoisomerase (là một enzyme nuclease thuận nghịch) có tác dụng thay đổi số Lk của DNA. Các DNA topoisomerase có tác dụng thêm vào hoặc loại bớt những siêu xoắn trong phân tử DNA xoắn kép. Có hai loại DNA topoisomerase sau:

2.2.1. DNA topoisomerase I

DNA topoisomerase I tồn tại ở cả prokaryote lẫn eukaryote và có tác dụng sau:

- Cắt một chuỗi DNA và tháo một vòng xoắn sau vị trí cắt.

- Nối lại chuỗi DNA đã bị cắt bởi liên kết phosphodiester mới. DNA sau khi được tái tạo lại có cấu trúc lỏng lẻo hơn (Hình 4.14).



Hình 4.14. Cơ chế hoạt động của topoisomerase I. Enzyme cắt một sợi đơn của DNA sợi đôi, chuyển sợi nguyên vẹn qua chỗ đứt gãy, sau đó hàn chỗ đứt gãy lại. Quá trình này tăng số Lk lên 1.

2.2.2. DNA topoisomerase II

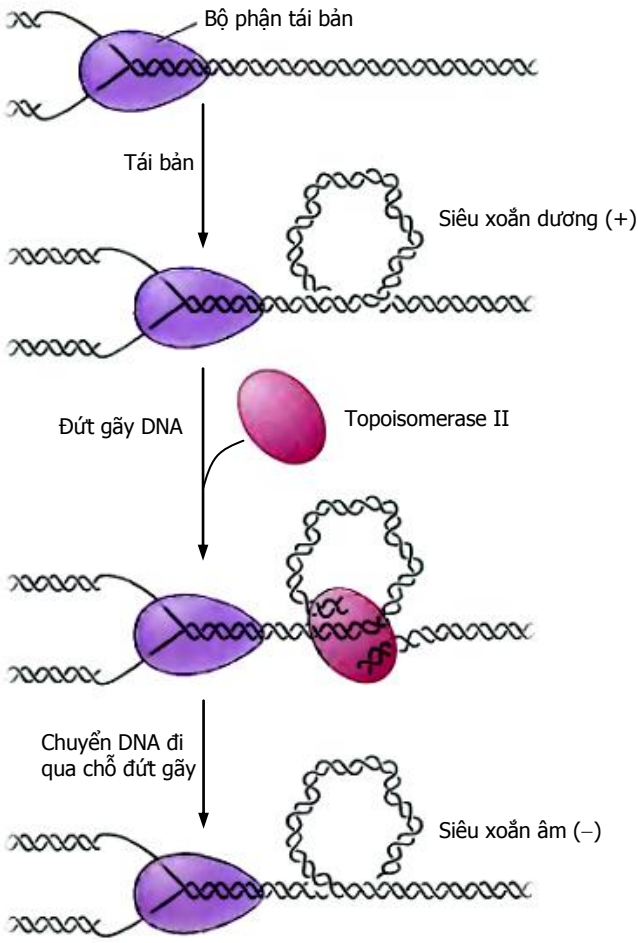
Ở prokaryote, DNA topoisomerase II có tên là DNA gyrase. DNA topoisomerase II có tác dụng cắt tạm thời hai chuỗi của DNA, có tác dụng sắp xếp lại siêu xoắn, tạo ra siêu xoắn trái (–) của chuỗi DNA xoắn kép (Hình 4.15). Phản ứng này cần 1 ATP.

Ở eukaryote, DNA topoisomerase II cũng được tìm thấy nhưng ít được nghiên cứu.

3. Helicase và protein SSB

3.1. Helicase

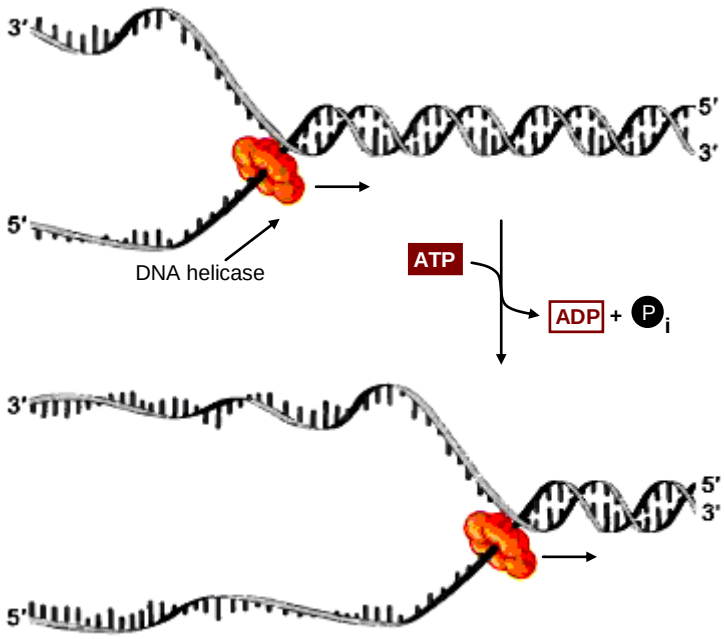
Enzyme helicase (còn có tên là enzyme deroulase) có nhiệm vụ giúp chuỗi DNA từ dạng siêu xoắn sang dạng dãn thành hai sợi đơn bằng cách cắt các liên kết hydrogen giữa những base bổ sung. Dạng cấu trúc này cần thiết cho các đoạn trên chuỗi DNA mà ở đó có nhu cầu sinh tổng hợp. Phản ứng này cần sự có mặt của ATP (Hình 4.16).



Hình 4.15. Hoạt động của topoisomerase II ở chạc ba sao chép

3.2. Protein SSB

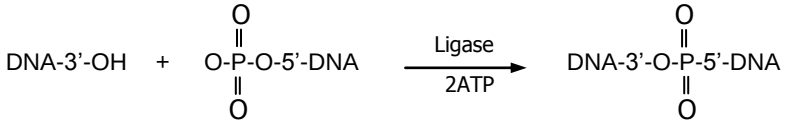
Các protein SSB liên kết phối hợp với DNA sợi đơn nhưng không liên kết với DNA sợi đôi. Chúng được xem như là các chất phản ứng trong tạo dòng phân tử xuất phát từ chỗ chúng có khả năng phá vỡ sự ổn định của các cấu trúc thứ cấp bên trong sợi nucleotide; tăng nhanh sự tái ủ của các polynucleotide bổ trợ và tăng hoạt tính của các DNA polymerase bằng cách loại bỏ các cấu trúc thứ cấp bên trong sợi là những yếu tố ngăn cản sự tiến triển của các enzyme này (Hình 4.17). Nhờ tính chất này, các protein SSB trở thành các chất phản ứng hữu ích trong xác định trình tự DNA.

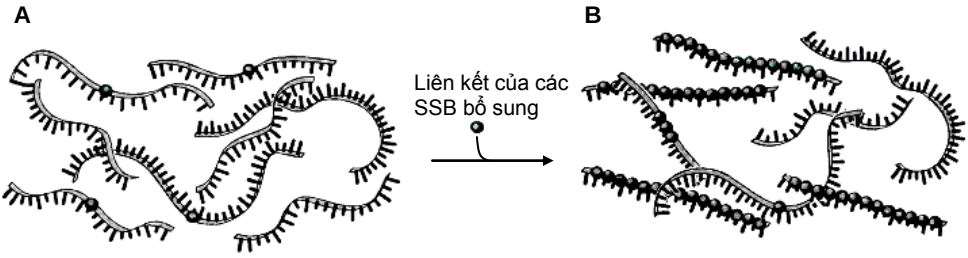


Hình 4.16. Hoạt động của enzyme helicase. DNA helicase tách rời hai sợi của chuỗi xoắn kép. Khi ATP được bổ sung vào DNA helicase được liên kết với sợi đơn, thì helicase di chuyển với một độ phân cực xác định trên DNA sợi đơn. Độ phân cực nghĩa là DNA helicase được liên kết với khuôn mẫu sợi thứ trên chạc ba tái bản.

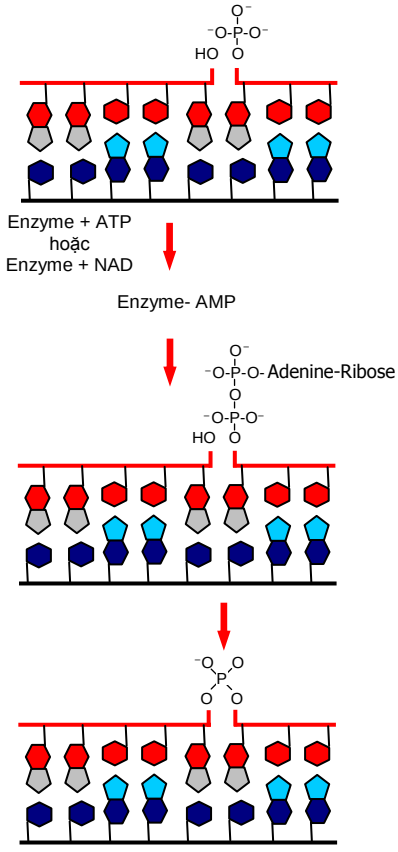
4. DNA ligase

Có nhiệm vụ nối hai đầu 3'-OH và 5'-PO_4 của hai đoạn DNA rời thành một đoạn liên tục. Một mối hàn như vậy cần một năng lượng là 2ATP (đối với eukaryote) hoặc NAD^+ (đối với vi khuẩn).





Hình 4.17. Liên kết của protein SSB lên DNA sợi đơn. Một lượng giới hạn của SSB liên kết với 4 trong 9 phân tử DNA sợi đơn. Khi bổ sung thêm protein SSB, nó sẽ liên kết với các protein SSB đã liên kết từ trước. Chỉ sau khi protein SSB bao bọc hoàn toàn các phân tử DNA sợi đơn ban đầu, thì nó mới tiếp tục liên kết với các phân tử DNA sợi đơn khác.



Hình 4.18. DNA ligase hàn các điểm đứt giữa các nucleotide gần nhau

Đặc điểm của DNA ligase là không làm việc với các sợi DNA đơn (single strand) mà chỉ hàn nối các đoạn DNA ở dạng chuỗi xoắn kép. Khi nối DNA, có thể gặp 2 trường hợp: đầu so le-đầu dính (protruding ends-cohesive ends) hoặc đầu bằng (blunt ends). Trường hợp đầu so le, các nucleotide nằm trên phần so le của hai đoạn DNA phải tương đồng thì DNA polymerase mới hoạt động được.

Tài liệu tham khảo/đọc thêm

1. Phạm Thành Hổ. 2003. Di truyền học. *NXB Giáo dục,* Hà Nội.

2. Lê Đức Trình. 2001. Sinh học phân tử của tế bào. *NXB Khoa học và Kỹ thuật*, Hà Nội.

3. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD. 2002. Molecular Biology of the Cell. 3rd ed. *Garland Publishing, Inc.* New York, USA.

4. Karp G. 2002. Cell and Molecular Biology: Concepts and Experiments. 3rd ed. *John Wiley & Sons, Inc*. New York, USA.

5. Lewin B. 2000. Gene VII. *Oxford University Press*, Oxford, UK.

6. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL and Darnell J. 2004. Molecular Cell Biology. 5th ed. *Freeman and Company*, New York, USA.

7. Watson JD, Hopkins NH, Roberts JW and Weiner AM. 2004. Molecular Biology of the Gene. *The Benjamin/Cummings Publishing Company, Inc*. California, USA.

8. Weaver RF. 2003. Molecular Biology. 2nd ed. *McGraw-Hill Company*, New York, USA.

Chương 5

Phiên mã

Phiên mã là quá trình tổng hợp RNA từ khuôn mẫu DNA. Quá trình này về phương diện hóa học và enzyme rất giống với quá trình tái bản DNA. Cả hai đều liên quan đến các enzyme tổng hợp một chuỗi nucleic acid mới bổ sung với khuôn mẫu DNA. Tất nhiên, hai quá trình này có những khác biệt quan trọng, mà đáng chú ý nhất là chuỗi mới trong quá trình phiên mã được tạo thành từ các ribonucleotide thay vì các deoxyribonucleotide. Các nguyên tắc cơ bản của quá trình phiên mã được thiết lập dựa vào nghiên cứu trên prokaryote (*E. coli*) nhưng dường như các nguyên tắc này cũng có tính phổ biến cho cả eukaryote. Tuy nhiên, do sự khác biệt về cấu trúc genome và hệ thống enzyme nên sự phiên mã ở prokaryote và eukaryote cũng có những đặc trưng nhất định.

I. Các đặc điểm cơ bản của quá trình phiên mã

1. Sự phiên mã tạo ra RNA bổ sung với một sợi DNA

Các ribonucleotide nối tiếp của RNA được xác định dựa trên nguyên tắc bổ sung với các nucleotide trên DNA khuôn mẫu. Khi sự bắt cặp chính xác xảy ra, ribonucleotide tiếp theo được liên kết đồng hóa trị với chuỗi RNA đang hình thành nhờ phản ứng có enzyme xúc tác. Như vậy, sản phẩm phiên mã được kéo dài từng ribonucleotide một và bổ sung chính xác với sợi DNA được dùng làm khuôn mẫu.

Quá trình phiên mã dù có tính chính xác cao nhưng vẫn kém hơn nhiều so với quá trình tái bản DNA (tỷ lệ mắc lỗi là 1/10.000 nucleotide so với 1/10.000.000 trong tái bản). Đó là do sự thiếu một cơ chế sửa sai hữu hiệu, mặc dù trong quá trình tổng hợp RNA cũng có hai dạng sửa sai tồn tại. Tuy nhiên, vì các RNA được phiên mã không bao giờ được sao chép lại nên các sai sót xảy ra không ảnh hưởng đến việc truyền đạt thông tin cho thế hệ sau.

2. Sự phiên mã là một phản ứng enzyme

Những enzyme chịu trách nhiệm cho quá trình phiên mã ở cả tế bào prokaryote và eukaryote đều được gọi là RNA polymerase phụ thuộc DNA (DNA-dependent RNA polymerase), gọi tắt là RNA polymerase.

RNA polymerase xúc tác hình thành các cầu nối phospho-diester để nối các ribonucleotide thành một chuỗi thẳng. Enzyme dịch chuyển từng bước dọc theo DNA khuôn mẫu và kéo dài chuỗi RNA theo hướng từ 5'→3', nghĩa là các ribonucleotide được thêm vào đầu 3' của chuỗi RNA đang hình thành. Các cơ chất được sử dụng để tổng hợp RNA là ATP, GTP, CTP và UTP. Cũng giống như trong sự tái bản DNA, năng lượng cho phản ứng được cung cấp từ sự thủy phân các cầu nối giàu năng lượng của các cơ chất nói trên.

3. Sự phiên mã chỉ sao chép chọn lọc một số phần của genome và tạo ra nhiều bản sao

Sự lựa chọn vùng nào để phiên mã không phải xảy ra ngẫu nhiên. Mỗi vùng phiên mã điển hình bao gồm một hoặc nhiều gen, có những trình tự DNA đặc hiệu hướng dẫn khởi đầu và kết thúc phiên mã.

Đối với một vùng được chọn phiên mã, có thể có một đến hàng trăm thậm chí cả nghìn bản sao RNA được tạo ra. Sự tổng hợp phân tử RNA sau được bắt đầu trước khi phân tử RNA trước hoàn thành. Từ một gen đơn độc, trong vòng một giờ có thể tổng hợp được hơn một nghìn phân tử RNA (đối với eukaryote).

Sự lựa chọn vùng nào để phiên mã và mức độ phiên mã đều được điều hòa. Vì vậy, trong những tế bào khác nhau hoặc trong cùng một tế bào nhưng ở những thời điểm khác nhau sẽ có những nhóm gen khác nhau được phiên mã.

4. Chỉ một trong hai sợi đơn của phân tử DNA được dùng làm khuôn mẫu

Việc gắn của RNA polymerase vào promoter của gen sẽ quyết định việc lựa chọn sợi nào trong hai sợi đơn của DNA làm khuôn mẫu. Promoter, ngoài việc mang vị trí gắn RNA polymerase còn chứa đựng thông tin xác định sợi nào trong hai sợi đơn của DNA được phiên mã và xác định vị trí bắt đầu phiên mã.

5. Sự phiên mã được khởi phát không cần mồi

RNA polymerase có thể khởi đầu sự tổng hợp RNA trên khuôn mẫu DNA mà không cần mồi như DNA polymerase. Điều này đòi hỏi ribonucleotide đầu tiên được mang đến vị trí bắt đầu phiên mã phải được giữ ổn định trên DNA khuôn mẫu trong khi ribonucleotide thứ hai đang được đưa đến để xảy ra phản ứng trùng hợp.

II. Các giai đoạn của quá trình phiên mã

RNA polymerase tiến hành quá trình phiên mã một gen thông qua nhiều bước và được chia thành ba giai đoạn: khởi đầu, kéo dài, và kết thúc.

1. Giai đoạn khởi đầu

Đầu tiên, RNA polymerase gắn với promoter của gen (cùng với các yếu tố khởi đầu) để tạo thành phức hợp promoter-polymerase. Một khi được hình thành, phức hợp này sẽ có sự thay đổi cấu trúc cần thiết cho việc tiến hành giai đoạn khởi đầu.

Giai đoạn khởi đầu này bao gồm ba bước như sau:

1.1. Phức hợp đóng (closed complex)

Khi RNA polymerase vừa mới gắn với promoter thì phức hợp tạo thành ở trạng thái đóng. Ở trạng thái này, DNA vẫn là chuỗi kép và enzyme gắn vào một phía của vòng xoắn.

1.2. Phức hợp mở (open complex)

Lúc này DNA xung quanh điểm bắt đầu phiên mã được mở xoắn và liên kết giữa các cặp base bổ sung bị phá vỡ. Hai sợi DNA với độ dài khoảng 14 bp xung quanh vị trí bắt đầu phiên mã tách rời nhau ra và tạo nên vòng phiên mã. Việc mở DNA đã giải phóng sợi khuôn mẫu. Hai nucleotide đầu tiên được mang đến vị trí bắt đầu đã được hoạt hóa, chúng xếp hàng trên khuôn mẫu và được nối với nhau. RNA polymerase bắt đầu chuyển dịch dọc theo sợi khuôn mẫu, mở xoắn phía trước vị trí trùng hợp và tái gắn hai sợi DNA phía sau. Bằng cách này, các ribonucleotide tiếp theo được gắn

vào chuỗi RNA đang phát triển. Đoạn RNA ban đầu với số lượng ribonucleotide ít hơn 10 sẽ bị loại bỏ và enzyme bắt đầu tổng hợp lại.

1.3. Phức hợp tam nguyên bền (stable ternary complex)

Một khi enzyme đi được khoảng hơn 10 bp, nó được xem là "thoát" khỏi promoter. Lúc này một phức hợp gồm ba thành phần là enzyme, DNA và RNA được hình thành và quá trình phiên mã chuyển sang giai đoạn kéo dài.

2. Giai đoạn kéo dài

Một khi RNA polymerase tổng hợp được một đoạn RNA khoảng 10 base, nó chuyển sang giai đoạn kéo dài. Sự chuyển giai đoạn này đòi hỏi những biến đổi hơn nữa về cấu hình của RNA polymerase làm cho nó càng giữ chặt khuôn mẫu hơn nữa. Trong quá trình kéo dài, enzyme này thực hiện một loạt nhiều nhiệm vụ rất quan trọng ngoài việc tổng hợp RNA. Nó mở xoắn DNA trước mặt và tái gắn DNA phía sau, nó tách dần chuỗi RNA đang hình thành ra khỏi khuôn mẫu khi đang di chuyển dọc theo gen, và nó còn thực hiện chức năng đọc sửa sai (ở quá trình tái bản DNA, các chức năng này do nhiều enzyme khác nhau thực hiện).

3. Giai đoạn kết thúc

Khi RNA polymerase phiên mã hết chiều dài của gen (hoặc các gen), nó dừng lại và giải phóng sản phẩm RNA. Trong một số tế bào, có những trình tự đặc hiệu thúc đẩy giai đoạn kết thúc. Trong một số tế bào khác, người ta vẫn chưa rõ yếu tố nào điều khiển enzyme ngừng phiên mã và giải phóng RNA khỏi khuôn mẫu.

III. Phiên mã ở prokaryote

1. Enzyme RNA polymerase ở prokaryote

Tế bào prokaryote chỉ chứa một loại RNA polymerase. Enzyme này được cấu tạo bởi năm tiểu đơn vị, gồm hai tiểu đơn vị α và các tiểu đơn vị β, β', ω. Các tiểu đơn vị này liên kết với nhau để tạo thành enzyme lõi.

Người ta làm thí nghiệm tinh sạch enzyme lõi từ tế bào vi khuẩn rồi cho vào dung dịch chứa DNA vi khuẩn và ribonucleotide thì enzyme sẽ gắn với DNA và tổng hợp RNA. Tuy nhiên, RNA thu được trong thí nghiệm này không giống với RNA trong tế bào vì enzyme đã gắn vào các vị trí không thích hợp trên DNA. Nếu cho thêm một loại polypeptide tinh sạch được gọi là yếu tố σ trước khi enzyme gắn với DNA thì sự phiên mã sẽ bắt đầu tại những vị trí chính xác như trong tế bào.

Việc gắn yếu tố σ vào enzyme lõi làm tăng ái lực của enzyme với vị trí khởi động trên DNA, đồng thời làm giảm ái lực đối với các vùng khác. Khi enzyme lõi được gắn thêm yếu tố σ sẽ trở thành dạng holoenzyme.

2. Promoter của gen ở prokaryote

Promoter của gen vi khuẩn nằm ngay trước vị trí khởi đầu phiên mã. Bằng cách phân tích các trình tự DNA này của một số lượng lớn các gen vi khuẩn người ta nhận thấy có hai đoạn ngắn tương đối giống nhau giữa gen này và gen khác, mỗi đoạn gồm sáu nucleotide. Đoạn thứ nhất nằm cách vị trí khởi đầu khoảng 35 bp, là sự biến đổi của trình tự TTGACA. Đoạn thứ hai nằm cách vị trí khởi đầu khoảng 10 bp, là sự biến đổi của trình tự TATAAT. Chúng lần lượt được gọi là vùng −35 và vùng −10 (hộp Pribnow).

Đối với các promoter mạnh (ví dụ ở gen mã hóa rRNA), người ta còn tìm thấy yếu tố UP làm tăng sự gắn của RNA polymerase vào DNA. Có một số promoter thiếu vùng −35 và được thay bằng yếu tố −10 mở rộng, bao gồm vùng −10 chuẩn và thêm một đoạn ngắn ở đầu 5’ của nó (ví dụ ở gen *gal* của *E. coli*).

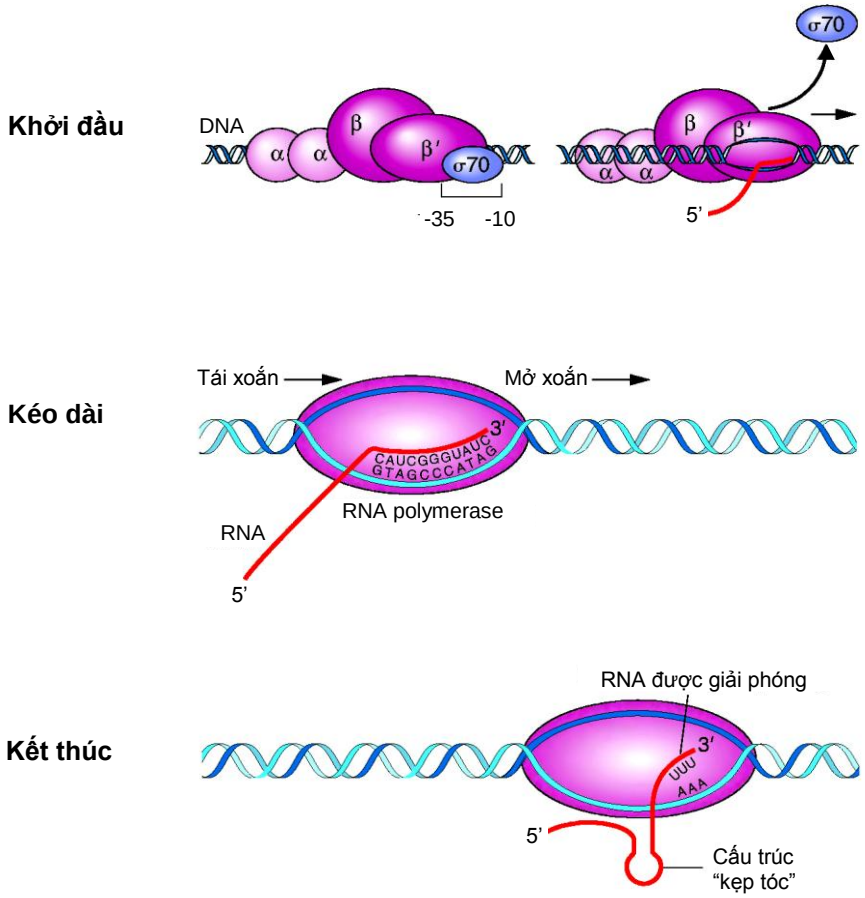
Yếu tố σ nhận dạng các vùng −35 và −10 hoặc yếu tố −10 mở rộng nhờ các cấu trúc đặc biệt của chúng. Riêng yếu tố UP không được nhận dạng bởi σ mà được nhận dạng bởi một vùng ở đầu C tận cùng của tiểu đơn vị α, gọi là α CTD (carboxyl terminal domain).

3. Vai trò của enzyme RNA polymerase và promoter trong quá trình phiên mã

Khi khởi đầu sự phiên mã, RNA polymerase gắn với DNA ở vùng −35 một cách lỏng lẻo tạo thành phức hợp đóng. Sau đó, enzyme gắn vào

promoter chặt hơn và phức hợp chuyển sang trạng thái mở. Một vùng DNA được mở xoắn nằm giữa hai vị trí –11 và +3, nghĩa là bao gồm cả vị trí bắt đầu phiên mã +1. Sợi đơn DNA làm khuôn mẫu được bộc lộ để sinh tổng hợp RNA.

Một khi RNA polymerase tổng hợp được 10 nucleotide (sau lần khởi đầu) thì yếu tố σ rời khỏi enzyme và có thể đến gắn vào promoter khác để khởi đầu một quá trình phiên mã mới. Sự tách rời yếu tố σ cho phép hình thành một kênh thoát mà thông qua đó phần RNA đã được tổng hợp chui ra khỏi enzyme và được kéo dài (Hình 5.1).



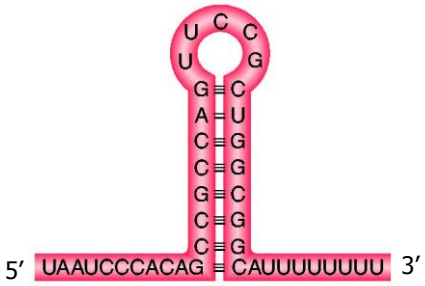
Hình 5.1. Các giai đoạn phiên mã ở prokaryote

4. Tín hiệu kết thúc

Tín hiệu kết thúc là những trình tự trên DNA có vai trò thúc đẩy sự tách RNA polymerase ra khỏi DNA và giải phóng chuỗi RNA mới được tổng hợp. Ở vi khuẩn, có hai loại tín hiệu kết thúc là tín hiệu kết thúc không phụ thuộc Rho và tín hiệu kết thúc phụ thuộc Rho.

4.1. Tín hiệu kết thúc không phụ thuộc Rho

Đó là một cấu trúc đặc biệt bao gồm hai trình tự đối xứng bổ sung nhau giàu GC, tiếp theo sau là một loạt 8 adenine (chúng sẽ được phiên mã thành 8 uracil). Những trình tự này không ảnh hưởng đến RNA polymerase cho đến sau khi chúng được phiên mã. Khi đó các trình tự đối xứng bổ sung trên RNA sẽ bắt cặp với nhau và tạo nên cấu trúc hình chiếc kẹp tóc (Hình 5.2). Cấu trúc kẹp tóc này đã phá vỡ phức hợp kéo dài hoặc bằng cách thúc đẩy mở kênh thoát cho RNA trên RNA polymerase hoặc bằng cách phá vỡ mối tương tác RNA-khuôn mẫu DNA



Hình 5.2. Cấu trúc kẹp tóc

Cấu trúc kẹp tóc này chỉ hoạt động hiệu quả khi nó được theo sau bởi một loạt U. Tại thời điểm cấu trúc kẹp tóc được hình thành, chuỗi RNA đang phát triển vẫn được giữ trên khuôn mẫu chỉ bằng các liên kết giữa A-U. Vì liên kết A-U yếu hơn G-C nên chúng dễ dàng bị phá vỡ bởi hiệu quả của cấu trúc kẹp tóc, và vì thế RNA dễ dàng được giải phóng.

4.2. Tín hiệu kết thúc phụ thuộc Rho

Tín hiệu này là những thành phần RNA mà muốn hoạt động, đòi hỏi phải có yếu tố Rho. Yếu tố Rho là một protein hình nhẫn (ở *E. coli* có khối

lượng phân tử khoảng 50 kDa) gồm sáu tiểu đơn vị giống hệt nhau. Yếu tố này có hoạt tính ATPase: khi gắn vào sản phẩm phiên mã, nó sẽ sử dụng năng lượng từ sự thủy phân ATP để kéo RNA ra khỏi khuôn mẫu và polymerase.

Yếu tố Rho điều khiển phân tử RNA đặc biệt nhờ tính đặc hiệu. Thứ nhất là tính đặc hiệu tại những vị trí nó gắn, đó là những đoạn khoảng 40 nucleotide mà không gấp lại thành cấu trúc bậc hai, chúng thường giàu C. Tính đặc hiệu thứ hai là yếu tố Rho giảm gắn với bất kỳ sản phẩm phiên mã nào đang được dịch mã (nghĩa là những sản phẩm đang gắn ribosome). Ở vi khuẩn, phiên mã và dịch mã được song hành chặt chẽ: dịch mã khởi phát ngay trên RNA đang được tổng hợp kể từ khi RNA bắt đầu chui ra khỏi RNA polymerase. Vì vậy Rho chỉ kết thúc những sản phẩm vẫn đang được phiên mã ở quá đầu tận cùng của gen hoặc operon.

IV. Quá trình phiên mã ở eukaryote

1. Cấu trúc gen ở eukaryote

Cấu trúc một gen mã hóa cho protein của eukaryote bao gồm các vùng sau:

1.1. Vùng 5’ kiểm soát biểu hiện gen

Vùng này bao gồm các trình tự nucleotide điều hòa biểu hiện gen và hoạt hóa sự phiên mã, bao gồm:

- Promoter. Là những trình tự định vị ở đầu 5’ không dịch mã của gen có chức năng xác định vị trí bắt đầu phiên mã, kiểm soát số lượng mRNA và đôi khi cả tính đặc hiệu mô (tissue-specific). Promoter có thể dài đến vài nghìn bp. Vùng này thường chứa một trình tự bảo thủ gọi là hộp TATA nằm cách vị trí phiên mã khoảng 25-30 bp. Hộp này giúp xác định chính xác vị trí bắt đầu phiên mã. Ngoài ra, còn có hộp CCAAT nằm cách vị trí bắt đầu phiên mã khoảng 75-80 bp. Hộp này ít phổ biến hơn hộp TATA, nó có chức năng tăng hiệu quả phiên mã. Một số gen “quản gia” mã hóa cho các enzyme hiện diện ở tất cả các tế bào thường thiếu cả hai hộp này và promoter rất giàu GC. Ngoài ra, còn có các thành phần đặc hiệu khác.

- Vị trí gắn vùng đặc hiệu mô. Là trình tự DNA tương tác với protein đặc hiệu mô để chỉ huy gen cấu trúc sản xuất ra từng protein đặc hiệu của từng mô.

- Vị trí gắn với vùng tăng cường phiên mã (enhancer). Còn gọi là gen tăng cường, là những trình tự DNA được gắn với các tác nhân hoạt hóa để kích thích phiên mã của các gen kề bên, chúng có thể tác động qua một khoảng cách xa và có thể tác động theo hai phía (từ 5' hoặc 3' tới).

1.2. Vùng được phiên mã

Bao gồm các exon và intron nằm xen kẽ. Đây là một đặc điểm để phân biệt với gen của prokaryote. Các exon và intron đều được phiên mã nhưng chỉ có các exon là được dịch mã. Các intron được bắt đầu bằng GT và kết thúc bằng AG. Các intron chiếm phần lớn trong mỗi gen và chúng sẽ được loại bỏ khỏi RNA mới được tổng hợp, còn các exon được nối với nhau để tạo nên mRNA hoàn chỉnh (mature mRNA).

Ở hai đầu của vùng được phiên mã (coding region) còn có vùng 5' không dịch mã (5' untranslation region) và vùng 3' không dịch mã (3' untranslation region). Vùng 5' không dịch mã được tính từ vị trí bắt đầu phiên mã cho đến codon khởi đầu ATG. Vùng 3' không dịch mã bắt đầu từ codon kết thúc đến vị trí gắn đuôi poly(A).

1.3. Vùng 3' không dịch mã

Chức năng chưa rõ, ở một số gen vùng này mang các trình tự điều hòa chuyên biệt.

2. Enzyme RNA polymerase của eukaryote

Tế bào eukaryote có đến ba loại RNA polymerase là RNA polymerase I (pol I), RNA polymerase II (pol II), và RNA polymerase III (pol III). Trong đó, RNA polymerase II đảm nhận việc phiên mã cho hầu hết các gen, chủ yếu là gen mã hóa cho protein. RNA polymerase I phiên mã các gen tiền thân của rRNA lớn và RNA polymerase III phiên mã các gen tRNA, một số gen RNA kích thước nhỏ của nhân (small nuclear, snRNA) và RNA 5S. Các RNA polymerase của eukaryote cũng bao gồm nhiều tiểu đơn vị,

trong đó có những tiểu đơn vị tương ứng với các tiểu đơn vị trong enzyme của vi khuẩn.

Bảng 5.1. Các tiểu đơn vị của RNA polymerase

Prokaryote		Eukaryote		
Vi khuẩn	Archaea	RNA pol I	RNA pol II	RNA pol III
β’	A’/A”	RPA1	RPB1	RPC1
β	B	RPA2	RPB2	RPC2
α’	D	RPC5	RPB3	RPC5
α”	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	(+6)	(+9)	(+7)	(+11)

3. Các yếu tố giúp RNA polymerase khởi đầu phiên mã

Trong khi RNA polymerase của prokaryote chỉ cần thêm một yếu tố khởi đầu là σ để khởi động phiên mã thì RNA polymerase của eukaryote phải cần nhiều yếu tố khởi đầu, các yếu tố này được gọi là các yếu tố phiên mã tổng quát. Mặt khác, sự khởi đầu phiên mã ở eukaryote còn phải đối mặt với hiện tượng các DNA được đóng gói trong nucleosome và các dạng cao hơn của cấu trúc chromatin. Điều này đòi hỏi phải có nhiều yếu tố khác thêm vào để giúp khởi động quá trình phiên mã.

3.1. Vai trò của các yếu tố phiên mã tổng quát

Các yếu tố phiên mã tổng quát giúp RNA polymerase vào đúng vị trí trên promoter, giúp tách hai sợi đơn của DNA ra để phiên mã được bắt đầu, và giải phóng RNA polymerase khỏi promoter một khi phiên mã đã được khởi động xong để đi vào giai đoạn kéo dài.

Các yếu tố này được gọi là “tổng quát” vì chúng gắn trên tất cả các promoter được sử dụng bởi RNA polymerase II. Do đó, chúng được viết tắt là TFII (transcription factor for polymerase II), bao gồm TFIIA, TFIIB...

Như đã mô tả ở trên, nhiều promoter của eukaryote chứa hộp TATA. Hộp này được nhận dạng bởi một tiểu đơn vị của TFIID là TBP (TATA binding protein: protein gắn TATA). Phức hợp TBP-DNA tạo nên một cái nền để thu hút các TFII khác và RNA polymerase đến promoter. *In vitro*, các yếu tố phiên mã tổng quát khác đến gắn vào promoter theo thứ tự sau: TFIIA, TFIIB, TFIIF cùng RNA polymerase II, TFIIE và TFIIH. Sau đó, vùng promoter được mở xoắn. Khác với vi khuẩn, sự mở xoắn ở đây cần có năng lượng cung cấp từ sự thủy phân ATP nhờ TFIIH (yếu tố này có hoạt tính giống helicase).

Sau đó cũng xảy ra hiện tượng khởi đầu sẩy giống như ở prokaryote cho đến khi có sự biến đổi về cấu hình nhằm giải phóng RNA polymerase II khỏi promoter và đi vào giai đoạn kéo dài. Tuy nhiên, ở đây có một bước mà không tìm thấy ở prokaryote đó là sự gắn thêm các gốc phosphate vào đuôi của RNA polymerase (đuôi này còn được gọi là CTD: carboxyl terminal domain). Sự phosphoryl hóa này cũng được xúc tác bởi TFIIH nhờ hoạt tính protein kinase của nó (Hình 5.3).

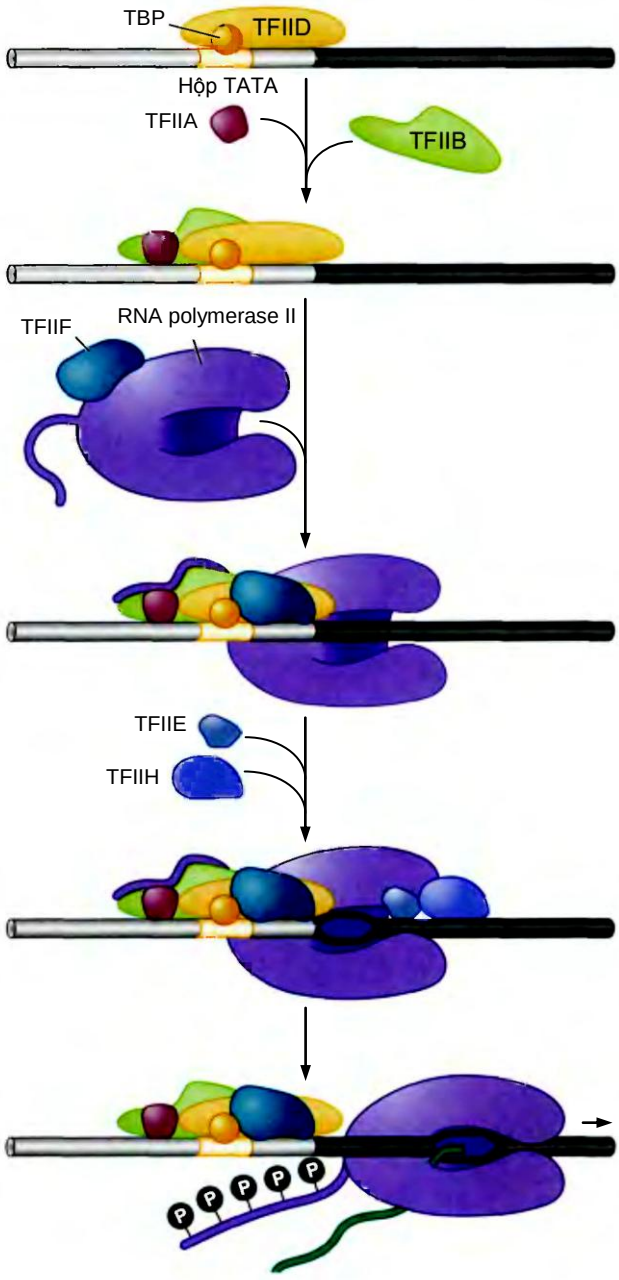
3.2. Vai trò của các tác nhân hoạt hóa, phức hợp trung gian và enzyme biến đổi chromatin.

Ngoài các yếu tố phiên mã tổng quát, RNA polymerase II còn cần sự hỗ trợ của các yếu tố khác (Hình 5.4).

Đầu tiên, cần có các tác nhân hoạt hóa phiên mã (transcriptional activator) đến gắn vào các trình tự đặc hiệu trên DNA (các enhancer) để hấp dẫn RNA polymerase II đến vị trí bắt đầu phiên mã. Sự hấp dẫn này cần thiết để RNA polymerase II và các yếu tố phiên mã tổng quát vượt qua trở ngại khi gắn với DNA được đóng gói trong chromatin.

Tiếp đến, sự khởi đầu phiên mã *in vivo* còn cần sự hiện diện của các protein tạo thành phức hợp trung gian (mediator complex). Phức hợp này cho phép tác nhân hoạt hóa tác động tốt lên RNA polymerase II và các yếu tố phiên mã tổng quát.

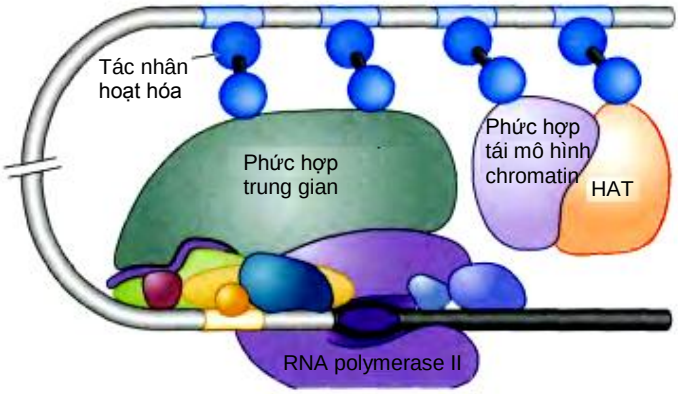
Cuối cùng, sự phiên mã còn cần các enzyme biến đổi chromatin, bao gồm phức hợp tái tạo mô hình chromatin (chromatin remodeling complex) và enzyme histone acetylase. Cả hai có tác dụng giúp cho bộ máy khởi đầu phiên mã có thể gắn vào DNA trong chromatin một cách dễ dàng.



Hình 5.3. Các yếu tố phiên mã tổng quát giúp khởi đầu phiên mã

Như vậy, có rất nhiều protein gắn vào promoter để khởi động sự phiên mã ở eukaryote. Thứ tự gắn của các protein này thay đổi đối với các gen khác nhau. Thực tế, một số có thể gắn với nhau ở xa DNA rồi được mang

đến DNA dưới dạng phức hợp. Ví dụ: phức hợp trung gian, RNA polymerase II, và một số yếu tố phiên mã tổng quát có thể gắn với nhau trong nhân tương rồi được mang đến DNA.



Hình 5.4. Các tác nhân hoạt hóa, phức hợp trung gian và các thành phần biến đổi nucleosome trong phiên mã

4. Các yếu tố kích thích RNA polymerase II hoạt động trong giai đoạn kéo dài

Một khi RNA polymerase II bắt đầu chuyển sang giai đoạn kéo dài thì các yếu tố khởi đầu được loại bỏ như yếu tố phiên mã tổng quát và phức hợp trung gian. Thay vào đó, các yếu tố kích thích giai đoạn kéo dài được thu hút đến, bao gồm TFIIS và hSPT5. Ngoài ra, còn có các yếu tố khác được huy động để phục vụ cho quá trình biến đổi RNA mới được tổng hợp.

Giai đoạn kéo dài trong phiên mã được song hành chặt chẽ với các bước biến đổi RNA mới được tổng hợp. Như đã phân tích ở trên, có một bước quan trọng diễn ra khi chuyển từ giai đoạn khởi đầu sang kéo dài là sự phosphoryl hóa đuôi CTD của RNA polymerase II. Sự phosphoryl hóa này không những giúp RNA polymerase II thoát khỏi các protein ở vị trí bắt đầu phiên mã mà còn cho phép các protein khác đến kết hợp với đuôi để giúp quá trình kéo dài phân tử RNA và biến đổi RNA tiền thân.

5. Quá trình biến đổi các RNA mới được tổng hợp

Ở eukaryote, sự phiên mã chỉ là bước đầu tiên trong một loạt các phản ứng bao gồm cả sự biến đổi hai đầu của mRNA tiền thân và loại bỏ các intron để tạo thành mRNA hoàn chỉnh.

5.1. Sự gắn mũ vào đầu 5'

Ngay khi RNA polymerase II vừa mới tạo ra khoảng 25 ribonucleotide của RNA, đầu 5' của phân tử RNA này được biến đổi bằng cách gắn thêm một cái mũ là guanine có biến đổi hóa học. Phản ứng gắn mũ được thực hiện bởi ba loại enzyme là:

- Phosphatase có tác dụng loại một gốc phosphate khỏi đầu 5' của RNA mới sinh.

- Guanylyl transferase gắn GMP bằng liên kết đảo ngược (5' với 5' thay vì 5' với 3') vào đầu 5' của RNA đang được tổng hợp.

- Methyl transferase gắn nhóm methyl vào guanosine (Hình 5.5).

Trong nhân, mũ gắn với một phức hợp protein được gọi là CBC (CAP-binding complex: phức hợp gắn mũ) giúp RNA được xử lý tốt và được chuyển ra ngoài. Mũ 5' còn có vai trò quan trọng trong quá trình dịch mã mRNA trong bào tương.

5.2. Sự gắn đuôi poly(A) vào đầu 3' và kết thúc phiên mã

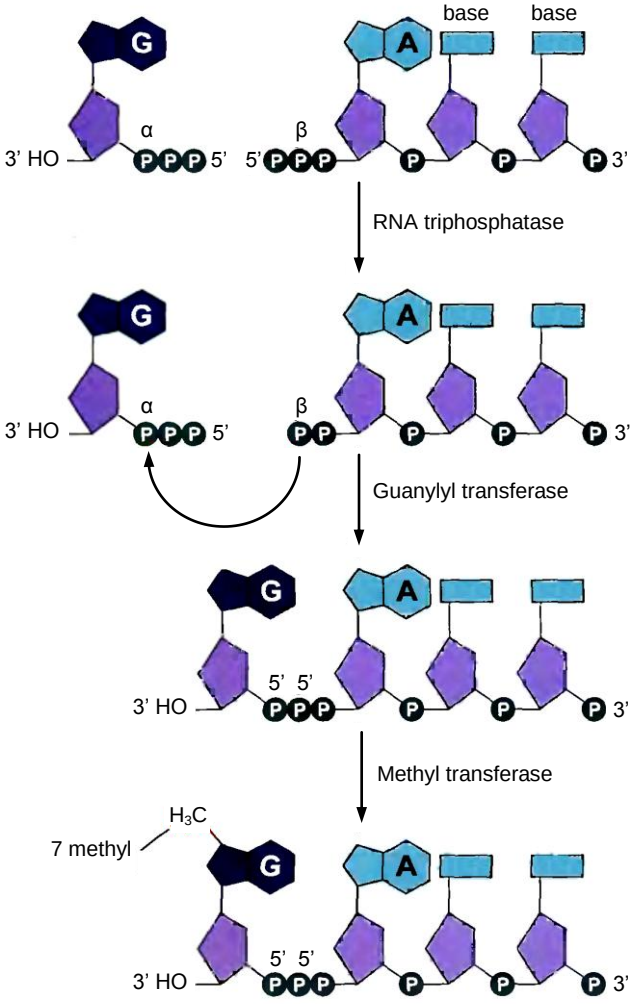
Sự gắn đuôi poly(A) vào đầu 3' được liên kết với sự kết thúc phiên mã. Các enzyme cần thiết cho các quá trình này được tập trung trên đuôi CTD của RNA polymerase II, bao gồm CPSF (cleavage and polyadenylation specificity factor: yếu tố đặc hiệu tách RNA và gắn đuôi polyA) và CstF (cleavage stimulation factor: yếu tố kích thích tách RNA).

Khi RNA polymerase di chuyển đến cuối gen, nó gặp những trình tự đặc hiệu (được gọi là tín hiệu polyA). Trình tự này được phiên mã thành RNA, thúc đẩy chuyển CPSF và CstF đến gắn vào đoạn RNA này. Sau đó các protein khác sẽ được tập trung đến để khởi đầu sự tách rời RNA và gắn đuôi poly(A).

Sự gắn đuôi được thực hiện bởi enzyme poly(A) polymerase (PAP). Enzyme này thêm khoảng từ vài chục đến 200-250 adenine vào đầu 3' của

RNA đã được tách ra. Enzyme PAP sử dụng ATP làm cơ chất và hoạt động giống RNA polymerase, tuy nhiên không có khuôn mẫu. Người ta vẫn chưa rõ yếu tố nào quyết định chiều dài của đuôi nhưng quá trình này có liên quan đến các protein gắn đặc hiệu với đuôi poly(A).

Sau khi RNA được tách ra và gắn đuôi poly(A), RNA polymerase vẫn chưa kết thúc phiên mã ngay. Nó còn tiếp tục di chuyển dọc theo khuôn mẫu và tạo ra một phân tử RNA thứ hai có thể dài hàng trăm nucleotide trước khi kết thúc. Sau đó, RNA polymerase được tách ra khỏi khuôn mẫu, giải phóng RNA mới và RNA này sẽ bị giáng hóa.



Hình 5.5. Phản ứng gắn mũ vào đầu 5' của RNA

5.3. Quá trình cắt nối gen (splicing)

Đây là quá trình loại bỏ các intron và nối các exon lại với nhau. Quá trình này được thực hiện phần lớn bởi các RNA thay vì protein.

Các phân tử RNA này tương đối ngắn (dưới 200 nucleotide) được gọi là snRNA. Có năm loại liên quan đến dạng cắt nối chính là U1, U2, U4, U5 và U6. Mỗi snRNA được kết hợp với nhiều protein để hình thành snRNP.

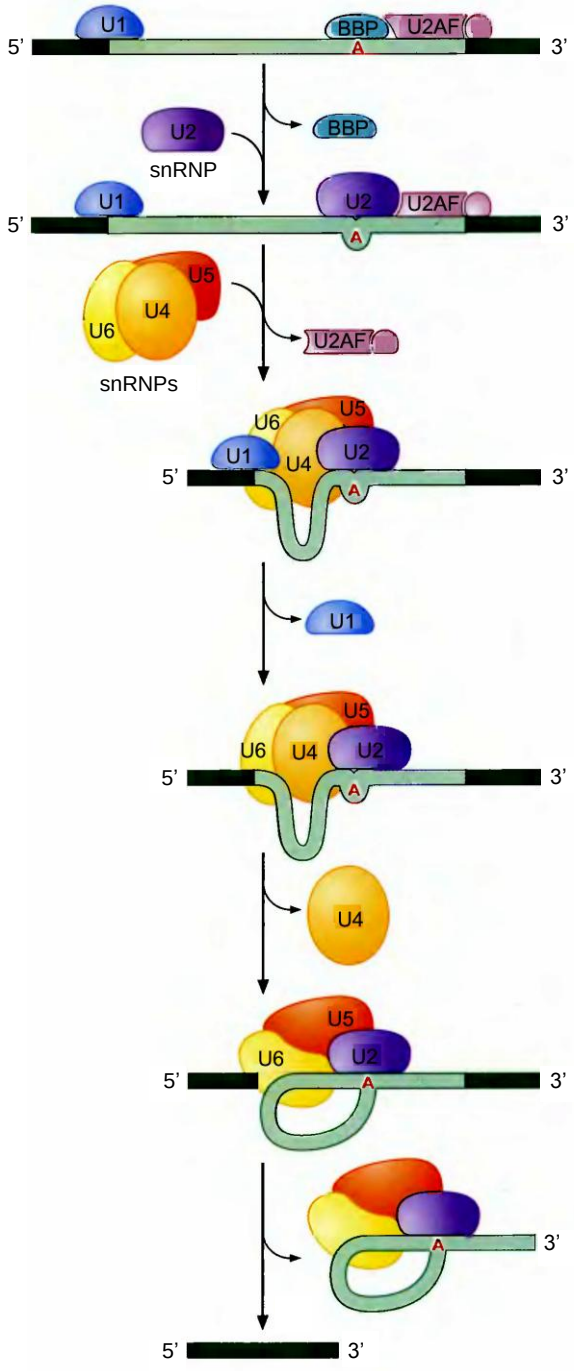
Có ba trình tự nằm trên intron đóng vai trò quan trọng trong quá trình cắt nối là: vị trí cắt nối đầu 5', vị trí phân nhánh là một trình tự giàu các pyrimidine bao quanh một nucleotide adenine ở gần đầu 3', và vị trí cắt nối đầu 3'.

Đầu tiên, vị trí cắt nối đầu 5' được nhận dạng bởi U1 snRNP bằng sự bắt cặp các base bổ sung. Vị trí phân nhánh cũng được nhận dạng bởi BBP (branch-point binding protein: protein gắn vị trí phân nhánh) và U2AF. U2 snRNP đến thay thế BBP (nhờ U2AF giúp đỡ). Sự bắt cặp base giữa U2 snRNA và vị trí phân nhánh làm thừa ra một gốc A không bắt cặp và sẵn sàng phản ứng với vị trí 5'. Sau đó U4 và U6 snRNP cùng U5 snRNP đến gắn với phức hợp trên.

Dạng U1 rời khỏi phức hợp và U6 thế chỗ U1 tại vị trí cắt nối đầu 5'. U4 được giải phóng để U6 tương tác với U2. Điều này làm cho vị trí cắt nối đầu 5' và vị trí phân nhánh được đặt kề nhau và tạo điều kiện cho phản ứng cắt nối xảy ra. Nucleotide adenine đặc hiệu ở vị trí phân nhánh tấn công và cắt intron ở vị trí đầu 5'. Đồng thời có một liên kết đồng hóa trị xảy ra giữa A ở vị trí phân nhánh và đầu 5' của intron tạo nên một cấu trúc hình thòng lọng. Đầu 3' của exon trước nối với đầu 5' của exon kế tiếp và thòng lọng intron được giải phóng (Hình 5.6).

V. Phiên mã ngược

Phiên mã ngược là quá trình tổng hợp DNA dựa trên khuôn mẫu RNA. Quá trình này được xúc tác bởi một loại enzyme đặc biệt gọi là enzyme phiên mã ngược (RT: reverse transcriptase). Enzyme này có cả hoạt tính DNA polymerase và hoạt tính RNase H. Cũng giống như DNA polymerase trong quá trình tái bản DNA, enzyme phiên mã ngược đòi hỏi phải có primer (mồi) đặc biệt để tổng hợp nên DNA mới.



Hình 5.6. Quá trình cắt nối gen

Enzyme phiên mã ngược được phát hiện lần đầu tiên ở retrovirus. Sau khi các retrovirus đi vào tế bào vật chủ, genome của RNA của nó sẽ được phiên mã ngược thành DNA sợi đôi, rồi tích hợp vào DNA của vật chủ. Quá trình phiên mã ngược này khá phức tạp, người ta có thể tóm tắt thành 10 bước như sau:

- Bước 1: Một tRNA của tế bào có tính đặc hiệu với retrovirus đóng vai trò primer để khởi phát quá trình phiên mã ngược. Primer lai với vùng bổ sung trên genome của RNA của retrovirus gọi là vị trí gắn primer (PBS: primer-binding site).

- Bước 2: Một đoạn DNA được kéo dài từ tRNA có trình tự bổ sung với trình tự của genome RNA của retrovirus.

- Bước 3: Trình tự R đầu 5’ và U5 của virus bị loại bỏ bởi RNase H.

- Bước 4: Đổi khuôn mẫu lần thứ nhất (first template exchange) còn được gọi là “nhảy” lần một: DNA đến lai với trình tự R còn lại ở đầu 3'.

- Bước 5: Một sợi DNA được kéo dài từ đầu 3’.

- Bước 6: Hầu hết RNA virus bị loại bỏ bởi RNase H.

- Bước 7: Sợi DNA thứ hai được kéo dài từ RNA còn lại của virus.

- Bước 8: Cả tRNA và phần RNA còn lại của virus bị loại bỏ bởi RNase H.

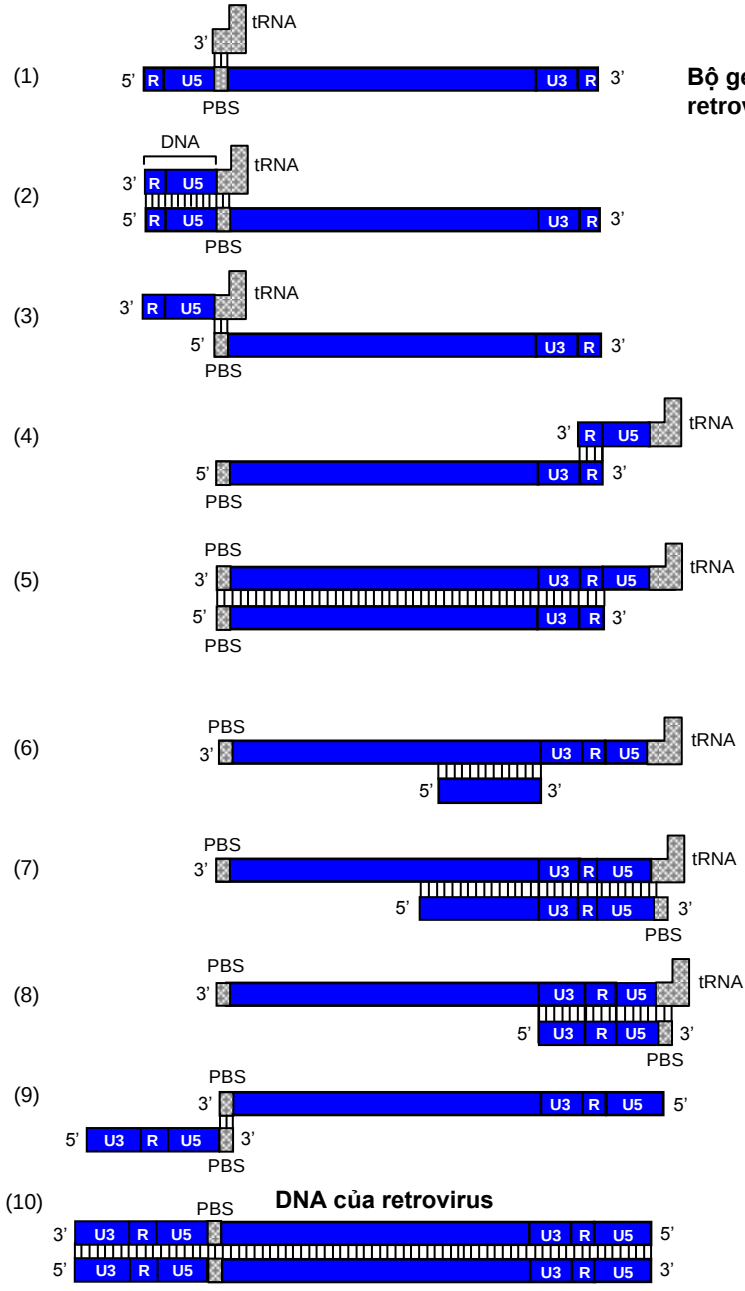
- Bước 9: Đổi khuôn mẫu lần thứ hai (“nhảy” lần hai): vùng PBS của sợi DNA thứ hai đến lai với vùng PBS của sợi thứ nhất.

- Bước 10: Kéo dài cả hai sợi DNA.

R là những trình tự lặp lại trên genome của retrovirus (ở đầu 5’ và 3’), U5 và U3 là những vùng mã hóa cho những tín hiệu tích hợp ở đầu 5’ và 3’ (Hình 5.7).

Ngày nay, người ta cũng phát hiện enzyme phiên mã ngược ở các virus động vật khác như hepadnavirus, và ở các virus thực vật như caulimovirus. Tất cả chúng được gọi là retrovirus. Ngoài ra, người ta còn phát hiện hoạt tính enzyme phiên mã ngược ở một số dòng của myxobacteria và *E. coli*. Enzyme phiên mã ngược đã trở thành một công cụ không thể thiếu trong sinh học phân tử. Nó giúp các nghiên cứu viên phiên mã ngược mRNA của tế bào thành cDNA (complementary DNA), rồi sau đó có thể khuếch đại, tạo dòng và biểu hiện bằng các phương pháp đặc biệt. Sự phát hiện enzyme phiên mã ngược ở nhiều loại virus và đặc biệt là ở một

số vi khuẩn cũng có ý nghĩa quan trọng trong việc nghiên cứu sự tiến hóa của hệ thống sinh giới.



Hình 5.7. Quá trình phiên mã ngược của retrovirus

Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. *NXB Giáo dục*, Hà Nội.

2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. 2002. Molecular Biology of the Cell. 4th ed. *Garland Science Publishing, Inc.* New York, USA.

3. Flint SJ. 2000. Principles of Virology: Molecular Biology, Pathogenesis, and Control. *ASM Press*, Washington, USA.

4. Gelehrter TD, Collins FS and Ginsburg D. 1998. Principles of Medical Genetics. 2nd ed. *Williams & Wilkins Company*, Baltimore, USA.

5. Griffiths AJF, Wessler SR, Lewontin RC, Gelbart WM, Suzuki DT and Miller JH. 2005. An Introduction to Genetic Analysis. *WH Freeman & Co.* New York, USA.

6. Karp G. 2002. Cell and Molecular Biology. *John Wiley & Sons, Inc.* New York, USA.

7. Watson JD, Baker TA, Bell SP, Gann A, Levine M and Losick R. 2004. Molecular Biology of the Gene. *Pearson Education, Inc./Benjamin Cummings Publishing*, San Francisco, USA.