

HOÀNG TRỌNG PHÁN (chủ biên)
ĐỖ QUÝ HAI

GIÁO TRÌNH

NUCLEIC ACID

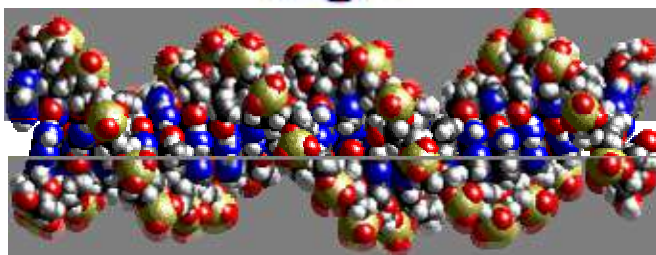


NHÀ XUẤT BẢN ĐẠI HỌC HUẾ

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC HUẾ

HOÀNG TRỌNG PHÁN (Chủ biên) - ĐỖ QUÝ HAI

Giáo trình NUCLEIC ACID



Huế - 2008

Lời nói đầu

Kể từ lúc Oswald T. Avery, MacLeod và McCarty (Đại học Stanford, USA; 1944) chứng minh DNA là vật chất mang thông tin di truyền và đặc biệt là, từ ngày James Watson và Francis Crick khám phá ra cấu trúc phân tử DNA - 25/4/1953 đến nay, Hoá sinh học và Sinh học phân tử đã phát triển với một tốc độ hết sức nhanh chóng. Những thành tựu mới nối tiếp nhau ra đời, đáng kể là sự hoàn thành việc giải mã di truyền bởi hai nhóm nghiên cứu của Marshall Nirenberg và Gobind Khorana vào tháng 6 năm 1966 và sự ra đời của Kỹ thuật Di truyền vào giữa thập niên 1970 là hai sự kiện nổi bật nhất kể từ sau khi sinh học phân tử ra đời. Kể đó, sự thành công của Dự án Bộ gene Người (Human Genome Project = HGP) vào tháng 4 năm 2003 được xem là một trong những kỳ công thám hiểm vĩ đại nhất của loài người. Lần đầu tiên con người có thể đọc được một cách đầy đủ toàn bộ trình tự 3.164.700.000 cặp base trong bộ gene của mình. Tất cả những sự kiện nổi bật này minh chứng một điều rằng: Sự phát triển cùng với những thành tựu đạt được của lĩnh vực nghiên cứu nucleic acid và sinh học phân tử nói chung trong thời gian qua quả là vô cùng to lớn!

Để góp phần đổi mới nội dung giáo trình Nucleic Acid theo hướng cập nhật kiến thức cũng như phương pháp dạy và học bộ môn, chúng tôi đã tham cứu nhiều tài liệu khác nhau và cố gắng biên soạn giáo trình trên tinh thần ấy. Chúng tôi hy vọng rằng giáo trình này sẽ đáp ứng được phần nào nhu cầu học tập của sinh viên trong bối cảnh đổi mới giáo dục hiện nay.

Nội dung giáo trình gồm sáu chương bao quát các kiến thức cơ bản về nucleic acid. Chương 1 đề cập đến Lịch sử và phương pháp nghiên cứu nucleic acid; các chương 2, 3 và 4 tập trung chủ yếu vào các khía cạnh cấu trúc của các nucleotide, polynucleotide, các phân tử DNA và RNA; còn các chương 5 và 6 đi sâu vào cơ chế của các quá trình sinh tổng hợp nucleotide, DNA, RNA và protein. Mặt khác, để đảm bảo tính toàn diện và tính hệ thống của giáo trình (trong khuôn khổ đã định), các kiến thức đại cương về cơ sở phân tử của đột biến và tái tổ hợp DNA cũng được trình bày ở cuối chương 5 như là một biến đổi thiết yếu, mang tính biện chứng của cấu trúc di truyền này. Cuối mỗi chương đều có các phần Câu hỏi và Bài tập và Tài liệu tham khảo để bạn đọc tiện ôn tập và tra cứu.

Giáo trình Nucleic Acid này được ra đời trong khuôn khổ của Dự án Giáo dục Đại học Huế. Vì vậy một số kiến thức nâng cao như phần Công

nghe DNA tái tổ hợp - một lĩnh vực ứng dụng mới mẻ và rộng lớn của sinh học phân tử - theo quy định sẽ được đề cập trong một giáo trình riêng - Công nghệ DNA tái tổ hợp - mà không đi sâu với tư cách là một chủ đề hay một chương riêng. Bên cạnh đó, một số thuật ngữ khoa học được thống nhất sử dụng bằng tiếng Anh để giúp người học dễ dàng hơn trong việc tiếp cận với thông tin qua sách báo nước ngoài hoặc internet.

Giáo trình này do ThS. Hoàng Trọng Phán và PGS.TS. Đỗ Quý Hai - hiện đang công tác tại Khoa Sinh học các trường Đại học Sư phạm và Đại học Khoa học thuộc Đại học Huế - biên soạn, với sự phân công như sau:

ThS. Hoàng Trọng Phán chủ biên và biên soạn các chương 3, 4, 5, 6 và một phần của chương 2; và

PGS.TS. Đỗ Quý Hai biên soạn chương 1 và một phần của chương 2.

Để giáo trình này kịp thời ra mắt bạn đọc, chúng tôi xin trân trọng cảm ơn Dự án Giáo dục Đại học Huế đã tài trợ cho việc biên soạn và xuất bản giáo trình trong khuôn khổ của Dự án Giáo dục Đại học mức B.

Chúng tôi xin bày tỏ lòng cảm ơn đặc biệt đến GS. TSKH. Lê Doãn Diên, Chủ tịch Hội Hoá sinh Việt Nam, Giám đốc Trung tâm INCEDA đã dày công đọc bản thảo và cho nhiều ý kiến quý báu kể từ khi đề cương giáo trình bắt đầu được hình thành.

Do khả năng còn hạn chế, chắc chắn giáo trình còn nhiều thiếu sót. Chúng tôi rất mong nhận được sự phê bình và chỉ bảo của các đồng nghiệp và bạn đọc để giáo trình được hoàn chỉnh hơn trong lần in sau.

Huế, ngày 15 tháng 10 năm 2005

Các tác giả,

HOÀNG TRỌNG PHÁN - ĐỖ QUÝ HAI

Mục lục

Lời nói đầu	3
Chương 1: Lịch sử và Phương pháp Nghiên cứu Nucleic Acid	
	Đỗ Quý Hai 7
I. Lịch sử nghiên cứu	7
II. Các phương pháp nghiên cứu	11
1. Các phương pháp chung	12
2. Các phương pháp tách chiết nucleic acid	14
3. Các phương pháp phân tích định tính và định lượng thô nucleic acid	17
4. Các phương pháp xác định trình tự nucleic acid	19
Chương 2: Cấu trúc của các Nucleotide và Polynucleotide	
	Hoàng Trọng Phán - Đỗ Quý Hai 21
I. Thành phần hoá học của các nucleotide	21
1. Base nitơ	21
2. Đường pentose	24
3. Phosphoric acid	25
II. Cấu trúc của các nucleotide	25
1. Cấu trúc của các nucleoside	25
2. Cấu trúc của các nucleotide	26
3. Cấu trúc của các di- và triphosphate nucleoside	27
III. Cấu trúc của các chuỗi polynucleotide	29
Chương 3: Cấu trúc và Đặc điểm của DNA	
	Hoàng Trọng Phán 33
I. Thành phần hoá học của DNA	33
II. Cấu trúc chuỗi xoắn kép DNA	34
1. Mô hình Watson-Crick	35
2. Các dạng DNA xoắn phải và xoắn trái	38
3. Các DNA mạch vòng sợi kép và sợi đơn	40
III. Định khu, hàm lượng và kích thước của DNA	43
IV. Đặc tính hoá lý của DNA	49
V. Chức năng của DNA	56

Chương 4: Cấu trúc và Chức năng của các RNA

Hoàng Trọng Phán 61

- I. Cấu trúc và chức năng của các mRNA 63
- II. Cấu trúc và chức năng của các tRNA 69
- III. Cấu trúc và chức năng của các rRNA 71

Chương 5: Sinh tổng hợp Nucleotide và DNA

Hoàng Trọng Phán 75

- I. Sinh tổng hợp các nucleotide 75
 - 1. Sinh tổng hợp các nucleotide purine 75
 - 2. Sinh tổng hợp các nucleotide pyrimidine 77
- II. Sinh tổng hợp DNA (Tái bản) 79
 - 1. Những nguyên tắc và đặc điểm chung của tái bản DNA 80
 - 2. Các enzyme tham gia tái bản DNA 82
 - 3. Cơ chế tái bản DNA 84
 - 4. Tái bản của các bộ gene RNA 91
 - 5. Ứng dụng của enzyme tái bản trong kỹ thuật sinh học phân tử 93
- III. Cơ sở phân tử của đột biến và tái tổ hợp DNA 97
 - 1. Cơ sở phân tử của đột biến 97
 - 2. Sửa chữa DNA 105
 - 3. Tái tổ hợp và đại cương về công nghệ DNA tái tổ hợp 107

Chương 6: Sinh tổng hợp RNA và Protein

Hoàng Trọng Phán 117

- I. Sinh tổng hợp RNA (Phiên mã) 117
 - 1. Đặc điểm chung của phiên mã 117
 - 2. Các RNA polymerase ở prokaryote và eukaryote 118
 - 3. Cơ chế phiên mã ở prokaryote và eukaryote 119
- II. Cấu trúc và chức năng của protein 128
 - 1. Cấu trúc của protein 128
 - 2. Chức năng của protein 130
- III. Mã di truyền 132
- IV. Cơ chế của quá trình sinh tổng hợp protein (Dịch mã) 136
 - 1. Hoạt hoá amino acid 136
 - 2. Mở đầu, kéo dài và kết thúc sự tổng hợp chuỗi polypeptide 137
- V. Sự điều hoà sinh tổng hợp protein ở vi khuẩn 141
 - 1. Mô hình operon ở *E. coli* 141
 - 2. Operon lactose và cơ chế điều hoà cảm ứng - âm tính 143
 - 3. Operon tryptophan và cơ chế điều hoà ức chế - âm tính 145

Chương 1

Lịch sử và Phương pháp Nghiên cứu Nucleic Acid

I. Lịch sử nghiên cứu

Nucleic acid là những hợp chất cao phân tử đóng vai trò hết sức quan trọng trong hoạt động sống của mọi cơ thể sinh vật. Chúng tham gia vào các quá trình cơ bản của sự sống như sinh tổng hợp protein, sinh trưởng, sinh sản và di truyền.

Trong một thời gian dài các nhà hóa học và các nhà nghiên cứu về sinh lý dinh dưỡng đã coi protein, lipid và carbohydrate là ba chất quan trọng nhất tạo nên cơ thể sống. Quan điểm cho rằng nucleic acid là những cấu tử trợ của nhân và tế bào chất đã mãi mãi lãng quên từ khi chất thứ tư này – nucleic acid được chứng minh là chất quan trọng hơn so với các chất trước đó.

Năm 1869, lần đầu tiên nhà hóa sinh học trẻ tuổi người Thụy Sĩ Friedrich Miescher (1844 -1895) đã phát hiện nucleic acid trong nhân tế bào. Ông đã đặt tên là nuclein vì nhận thấy nó tồn tại ở trong nhân tế bào (nucleus).

Đối tượng nghiên cứu của Miescher là những bạch cầu (lymphocytes). Khi dùng acid kết tủa dịch chiết xuất từ nhân của tế bào mủ lấy từ bông băng bỏ đi (bằng cách dùng enzyme phân hủy protein của dịch dạ dày là pepsin để tiêu hóa các phần khác giàu protein của tế bào) ông đã vô cùng ngạc nhiên nhận thấy rằng, nhân tế bào chứa một chất không phải mỡ, không phải carbohydrate, cũng không phải protein. Nó cũng không giống một chất sống nào đã biết và chứa phosphor và nitơ, hòa tan thì cho tính acid. Từ “nucleic acid” là do Altman đề nghị năm 1889 (ông đã phát hiện ra rằng đối tượng thuận tiện tốt nhất để chiết rút chất nuclein là những đầu tinh trùng của cá hồi). Và thực chất mỗi đầu tinh trùng của cá hồi hoàn toàn tương ứng với một nhân tế bào. Để điều chế chế phẩm nucleic acid, Miescher đã hòa tan phần đầu của tinh trùng (tế bào sinh dục đực) trong dung dịch muối nồng độ cao, sau đó bằng cách thêm nước vào đã gây ra kết tủa nucleic acid ở dạng sợi. Cần phải giữ chế phẩm lạnh do đó trong các ngày đông tháng giá của mùa đông, ông đã làm việc trong phòng không sưởi.

Thực tế, lịch sử nghiên cứu hóa học của nucleic acid gắn liền

với tên tuổi của Felix Hoppe - Seiler (1825-1895), nhà sinh lý học và hoá học rất nổi tiếng người Đức vì chính ở phòng thí nghiệm của ông ở Tübingen. Miescher đã làm việc với danh nghĩa là người học trò và cộng tác viên khoa học trẻ. Mặc dù Miescher là một cộng tác viên vô cùng cẩn thận và đầy triển vọng song Hoppe - Seiler vẫn thấy cần thiết phải đích thân lặp lại thí nghiệm xem có đúng là một chất mới hay không. Kết quả thực nghiệm không những khẳng định thành tựu của Miescher mà còn thu được thêm nhiều dẫn liệu mới rất đặc trưng. Hoppe - Seiler còn phát hiện trong nhân tế bào nấm men cũng có chất nuclein giống như ở tế bào bạch cầu.

Người kế tục phát triển công trình của Miescher là nhà hóa sinh học (hóa học hữu cơ) người Đức Albrecht Kossel (1853 - 1927). Năm 1882, ông đã phân tích nucleic acid ra những phần nhỏ chứa acid phosphoric, đường và base chứa nitơ (Kossel đã tách được hai pyrimidine và đặt tên là cytosine và thymine, cũng như hai purine với tên adenine và guanine). Do công trình này ông đã đạt giải Nobel về y và sinh lý học vào năm 1910. Một pyrimidine khác là uracil cũng đã được phát hiện sau này.

Trong những năm đầu của thế kỷ XX (1900 - 1932), nhà hóa sinh học người Mỹ gốc Nga Phoebus Aaron Theodore Levene (1869 - 1940) đã xác định được đơn vị cấu tạo của nucleic acid là các nucleotide (hay mononucleotide) và phân biệt được hai loại nucleic acid: deoxyribonucleic acid (DNA) và ribonucleic acid (RNA).

Như vậy, cho đến năm 30 của thế kỷ XX này, người ta đã biết được thành phần của các nucleic acid.

Levene đã đoán trước 4 nucleotide khác nhau liên quan đến RNA là adenylic acid (chứa adenine), guanilic acid (chứa guanine), cytidilic acid (chứa cytosine) và uridic acid (chứa uracil). Ông cũng giả thiết 4 nucleotide tiếp theo liên quan đến DNA là deoxyadenilic acid, deoxyguanilic acid, deoxythymidilic acid (chứa thymidine) và deoxycytidilic acid (chứa cytosine).

Do các giả thiết này của Levene tỏ ra phù hợp với hiểu biết của hóa học cho nên chúng sớm được các nhà hóa học chấp nhận. Nhưng cho đến thời kỳ đầu những năm 1950, chúng vẫn chưa được chứng minh rõ ràng khi nhà hóa sinh học người Anh Alexander Robertus Todd (1907-1997) chưa tổng hợp được các nucleotide phù hợp một cách chính xác với các công thức của Levene. Todd đã xác nhận được rằng các chất này thực tế giống

các hợp chất đã thu nhận được từ nucleic acid. Với công trình này, Todd đã được trao giải thưởng Nobel về hóa học năm 1957.

Trong những năm 1949 - 1953 nhà sinh học người Mỹ gốc Áo Erwin Chargaff (1905 - 2002) khi phân tích DNA đã phát hiện sự cân bằng của các base trong DNA: số nhóm purine bằng số nhóm pyrimidine (purine/pyrimidine = 1), số nhóm adenine bằng số nhóm thymine (adenine/thymine = 1), số nhóm guanine bằng số nhóm cytosine (guanine/cytosine = 1). Về sau này người ta đã đặt quy tắc mang tên ông về tổng lượng các loại nucleotide cấu tạo nên các DNA.

Trong những năm đầu thập niên 1950, hai nhà vật lý học người Anh, Maurice Hugh Frederick Wilkins (1916 - 2004) cùng với Rosalind Franklin (1920 - 1958) bắt đầu tiến hành các nghiên cứu đầu tiên về cấu trúc DNA. Wilkins là con trai của một bác sĩ người New Zealand chuyên học vật lý và vào thời kỳ đầu cuộc Chiến tranh Thế giới thứ hai, ông làm việc cho sự phát triển bom nguyên tử. Nhưng sau cuộc Thế chiến này, ông dùng tất cả sức lực của mình cho công việc nghiên cứu DNA và đã trở thành một trong những nhà bác học tiên phong trong lĩnh vực này nên được coi là nhà lý sinh học. Wilkins và Franklin đã áp dụng *phương pháp nhiễu xạ Ronghen* (X-ray diffraction) vào việc nghiên cứu cấu trúc tinh thể của DNA (lấy từ tuyến ức bê, thymus). Với những bức ảnh chụp cấu trúc tinh thể DNA có dạng chữ thập, hai tác giả này gợi ý rằng DNA có thể gồm hai hoặc ba mạch đơn bện xoắn với nhau.

Năm 1953, từ các nghiên cứu của mình kết hợp với các kết quả nghiên cứu trước đó của Chargaff và Wilkins, hai nhà khoa học ở trường Đại học Tổng hợp Cambridge là Francis Harry Compton Crick (England; 1916-2004) và James Dewey Watson (USA; 1928-2003) đã xây dựng thành công mô hình chuỗi xoắn kép của phân tử DNA. Với công trình này, Watson và Crick đã được trao giải thưởng Nobel về y học và sinh lý học năm 1962. Từ đó Crick đã xây dựng một chuỗi xoắn kép DNA bằng đồng trong vườn nhà ông ở trung tâm Cambridge, sau đó đã viết: "Người ta hỏi tôi là khi nào thì sẽ mạ vàng".

Cùng với việc nghiên cứu cấu trúc DNA, các nhà khoa học cũng đã nỗ lực tìm hiểu vai trò và các chức năng sinh học của nó.

Vào năm 1944, nghĩa là sau 75 năm kể từ khi phát hiện DNA trong nhân tế bào năm 1869, nhà vi khuẩn học người Mỹ Oswald Theodore Avery (1877 - 1955) đã chứng minh DNA là chất liệu di

truyền khi bổ sung DNA chiết xuất từ chủng độc (chủng không độc của phế cầu khuẩn *Diplococcus pneumoniae* biến thành chủng độc gây viêm phổi và làm chết chuột bắt nguồn từ thí nghiệm biến nạp của F. Griffith năm 1928) vào môi trường nuôi cấy và kết hợp với việc xử lý bằng DNase. Điều đó chứng tỏ DNA quyết định tính chất di truyền của phế cầu khuẩn.

Năm 1952, Alfred Day Hershey (1908 - 1997) và Martha Chase (1927 - 2003) cho biết, bằng đồng vị phóng xạ có thể theo dõi sự di chuyển của protein và DNA của virus (cơ sở là DNA chứa phosphor, protein chứa lưu huỳnh mà không chứa phosphor; vì vậy DNA virus có thể được đánh dấu bằng đồng vị phóng xạ phosphor, còn protein virus được đánh dấu bằng đồng vị phóng xạ lưu huỳnh). Khi đánh dấu một số hạt virus ở phần protein và một số hạt virus ở phần DNA rồi đưa vào môi trường nuôi cấy vi khuẩn chủ *E. coli* thì thấy DNA virus nhanh chóng thâm nhập vào tế bào chủ, còn phần protein thì không. Thí nghiệm cũng đã chứng minh tầm quan trọng về mặt di truyền của DNA.

Năm 1955, nhà hóa sinh học Mỹ Fraenkel-Konrad (1910 -) đã chia virus ra làm hai phần và sau đó nối được hai phần đó lại, phần protein không gây nhiễm (không chui vào tế bào vật chủ) còn phần DNA thì gây nhiễm. Thí nghiệm của Krauss trên hồng cầu người chứng minh sự liên quan giữa cấu trúc hóa học của protein với vai trò di truyền của DNA (khi đưa DNA chiết xuất từ hồng cầu non của người lành vào tủy xương bệnh nhân thiếu máu hình lưỡi liềm (có HbS không bình thường) thì thấy hemoglobin bình thường được tổng hợp. Điều đó chứng tỏ DNA quyết định cấu trúc đặc hiệu của protein.

Các thí nghiệm sau đó của các nhà khoa học đã chú trọng vào việc tổng hợp DNA.

Năm 1956 Severo Ochoa (1905 - 1993) và Arthur Kornberg (1918 -) - hai nhà khoa học người Mỹ đã tổng hợp DNA *in vitro* bằng cách trộn enzyme DNA polymerase I được chiết xuất từ *E. coli* với các dNTP, DNA khuôn (DNA tự nhiên) và Mg^{++} , DNA thu được không khác DNA tự nhiên. Hai nhà khoa học này đã được nhận giải thưởng Nobel về y học và sinh lý học vào năm 1959.

Năm năm sau khi công bố công trình cấu trúc xoắn kép của DNA vào năm 1958, Matthew Stanley Meselson (1930 -) và Franklin William Stahl (1928 -) đã chứng minh bằng thực nghiệm dự đoán trên là hoàn toàn có cơ sở. Bằng cách nuôi vi khuẩn *E.*

coli ở môi trường chứa nitơ nặng ^{15}N sau đó bằng ^{14}N bình thường rồi theo dõi qua các thế hệ, xem sự thay đổi tỷ trọng của DNA qua máy ly tâm siêu tốc (siêu ly tâm); hai nhà khoa học đã xác định được cơ chế tái sinh bán bảo tồn của DNA.

Về việc nghiên cứu RNA, cho đến cuối những năm 30 của thế kỷ XX, các nhà khoa học vẫn nghĩ rằng DNA là đặc trưng của các tế bào động vật, còn RNA chỉ có thể gặp được ở trong nấm men và các tế bào thực vật. Nguồn quan trọng nhất của DNA lúc này là tuyến ức (thymus), còn đối với RNA là nấm men.

Các nghiên cứu chính xác hơn sau này đã cho thấy cả hai loại nucleic acid đều có mặt trong tất cả các tế bào động vật và thực vật.

Năm 1939, các nhà nghiên cứu Torbjorn Oskar Caspersson (Thụy điển; 1910 - 1997) và Jean Brachet (Bỉ; 1909 - 1998) đã chứng minh sự tổng hợp mạnh mẽ protein trong bào tương cùng xảy ra với sự tổng hợp mạnh mẽ RNA.

Năm 1956, S. Ochoa (nhà hóa sinh học người Mỹ) đã dùng enzyme RNA polymerase được chiết xuất từ vi khuẩn *Azobacter vineladni* để tổng hợp RNA *in vitro*.

Năm 1956, George Emile Palade (1912 -) đã chứng minh microsome có chứa những hạt có nhiều RNA gọi là thể ribo tức là ribosome (55% RNA, còn lại là protein).

Năm 1961, Francois Jacob (1920 -) và Jacques Lucien Monod (1910 - 1976) nêu vấn đề có một loại RNA khác có đặc điểm chuyển hóa nhanh chóng và có cấu tạo tương tự DNA. Đó là mRNA, là chất trung gian chuyển thông tin từ DNA đến chuỗi polypeptide được tổng hợp.

Trước đó vào năm 1958, M. B. Hoagland (nhà hóa sinh học người Mỹ) đã nghiên cứu vấn đề đưa amino acid vào ribosome. Ông đã xác định là trước khi tham gia tổng hợp protein, amino acid được kết hợp với RNA vận chuyển (tRNA) để vận chuyển amino acid từ bào tương vào ribosome.

Với những phương pháp thí nghiệm khác nhau các phòng thí nghiệm của Marshall Warren Nirenberg (1927 -), Har Gobind Khorana (1922 -) và Ochoa đã xác định được các bộ ba mật mã của toàn bộ 20 amino acid vào những năm 1961 - 1965.

Do vai trò quan trọng của nucleic acid trong hoạt động sống của cơ thể sinh vật (tham gia vào các quá trình cơ bản của sự sống

như sinh tổng hợp protein, sinh trưởng, sinh sản và di truyền), cho nên việc nghiên cứu nucleic acid ngày càng được chú trọng. Từ những năm 70 của thế kỷ XX trở lại đây, nhất là trong 10 năm cuối của thế kỷ và những năm đầu của thế kỷ XXI này, hiểu biết về nucleic acid ngày càng được mở rộng.

Nhờ những thành tựu của sinh học phân tử, một lĩnh vực khoa học mới đã hình thành và phát triển mạnh mẽ, đó là công nghệ gen, một trọng tâm được đầu tư nghiên cứu hàng đầu của công nghệ sinh học.

II. Các phương pháp nghiên cứu

Cũng như sự nghiên cứu các đại phân tử sinh học khác, có nhiều phương pháp và kỹ thuật được sử dụng trong việc nghiên cứu nucleic acid.

Tuy nucleic acid có cấu trúc phức tạp và hoạt động rất chặt chẽ, nhưng nhờ những thành tựu mới về kỹ thuật, đã ra đời nhiều phương tiện và phương pháp nghiên cứu hiện đại, đảm bảo độ chính xác cao, cho phép khám phá thêm nhiều điều quan trọng ở nucleic acid.

1. Các phương pháp chung

1.1. Phương pháp nhiễu xạ Rơnghen

Phương pháp này dựa trên sự tán xạ của tia X (tia Rơnghen) qua cấu trúc tinh thể của chất cần phân tích. Bằng cách đó và pha của tia X bị nhiễu xạ kết hợp với việc xử lý dữ liệu trên máy tính, người ta có thể xác định được chính xác vị trí của một nguyên tử bất kỳ so với các nguyên tử còn lại trong một đại phân tử. Nhờ vậy, người ta có thể thu nhận được một hình ảnh "tĩnh" về cấu trúc phân tử nằm ở trạng thái tinh thể. Tuy nhiên, cần chú ý rằng các chất nằm trong tế bào không bao giờ tồn tại ở trạng thái tinh thể, mà chúng thường nằm ở trạng thái hoà tan hoặc kết hợp với các cấu trúc khác của tế bào. Bởi vậy, người ta thường sử dụng thêm phương pháp phân tích cộng hưởng từ hạt nhân để thu được cấu trúc phân tử của chất cần phân tích ở trạng thái hoạt động sinh học.

Nhờ kết quả nhiễu xạ Rơnghen trên các sợi nucleic acid, người ta đã xác định được cấu trúc không gian của phân tử nucleic acid (chủ yếu là của DNA).

1.2. Phương pháp điện di

Nguyên tắc của phương pháp này là dựa vào tính phân cực của phân tử và khả năng di chuyển về một hướng xác định khi chịu tác động của một điện trường. Trong các dung dịch kiềm và trung tính, các đại phân tử nucleic acid tích điện âm đồng đều trên khắp bề mặt nên trong điện trường, chúng sẽ di chuyển về cực dương của điện trường.

Bằng phương pháp điện di, chúng ta có thể thu được nucleic acid dạng tinh khiết từ dịch chiết nghiên cứu. Phương pháp này cũng được dùng để tách riêng các nucleic acid DNA, RNA, cũng như các loại RNA riêng biệt (mRNA, rRNA, tRNA).

1.3. Phương pháp sắc ký

Phương pháp sắc ký dựa vào khả năng phân bố của các chất chạy sắc ký khi đưa vào dung môi kết hợp với khả năng hấp phụ của vật liệu sắc ký. Các chất khác nhau có khả năng hòa tan vào dung môi khác nhau và được vật liệu sắc ký hấp phụ với lực khác nhau cho nên tốc độ di chuyển của chúng trên sắc ký đồ là khác nhau. Do đó các chất này sẽ được tách ra trong quá trình chạy sắc ký.

Phương pháp này được dùng để tách chiết và tinh chế nucleic acid. Có nhiều phương pháp sắc ký: sắc ký giấy, sắc ký cột, sắc ký trao đổi ion, sắc ký lớp mỏng, sắc ký khí.

1.4. Phương pháp quang phổ

Nguyên tắc của phương pháp này là dựa vào sự hấp thụ mạnh ánh sáng của một chất ở một bước sóng xác định. Nucleic acid hấp thụ mạnh ánh sáng tử ngoại ở bước sóng 260 nm do sự có mặt của base purine và pyrimidine. Giá trị mật độ quang ở bước sóng 260nm (OD_{260nm}) của các mẫu cho phép xác định nồng độ nucleic acid trong mẫu. Người ta cũng xác định được độ tinh khiết của chế phẩm nucleic acid khi xác định độ hấp thụ ánh sáng ở 280 và 260 nm (độ hấp thụ cực đại của protein là ở 280 nm).

1.5. Phương pháp đồng vị phóng xạ

Nguyên tắc của phương pháp này là các chất đồng vị phóng xạ được gắn lên chất nghiên cứu và dùng kỹ thuật phóng xạ tự ghi để theo dõi quá trình biến đổi của các chất khi đưa chúng vào cơ thể hoặc môi trường cần nghiên cứu. Nhờ vậy có thể xác định được cơ chế tham gia quá trình biến đổi của chất cần nghiên cứu. Cũng có thể dùng máy đo phóng xạ để định lượng chất cần nghiên cứu.

Phương pháp đồng vị phóng xạ được dùng để nghiên cứu cơ chế tổng hợp nucleic acid. Ngoài ra việc xác định thành phần cấu trúc cũng như định lượng nucleic acid cũng được tiến hành bằng phương pháp này. Các đồng vị hay dùng cho các nghiên cứu này là: ^3H , ^{14}C , ^{32}P , ^{15}N .

1.6. Phương pháp hóa học

Dùng phương pháp thủy phân thích hợp bằng acid hay kiềm, người ta có thể thu nhận được các phần khác nhau của nucleic acid. Bằng cách kết hợp với các phương pháp khác như sắc ký hay điện di, các thành phần được tách riêng để nghiên cứu cấu trúc, thành phần hóa học và tính chất của chúng. Người ta cũng có thể dùng phương pháp hóa học để định tính nucleic acid trong tế bào nhờ những phản ứng hóa học đặc trưng.

1.7. Phương pháp dùng hệ thống vô bào

Người ta có thể sử dụng đối tượng nghiên cứu là cơ thể toàn vẹn, lát cắt mô, tế bào, nhưng một phương pháp nghiên cứu rất có hiệu quả là phương pháp dùng các hệ thống vô bào (không dùng tế bào nguyên vẹn).

Nghiên vỡ tế bào rồi dùng các phương pháp ly tâm để tách các thành phần khác nhau của tế bào. Dựa trên nhu cầu nghiên cứu, người ta có thể thêm vào hệ thống vô bào các chất khác nhau. Nhờ hệ thống vô bào này chúng ta có thể nghiên cứu quá trình sinh tổng hợp protein, vai trò của các nucleic acid trong quá trình này, cũng như bản thân quá trình tổng hợp các nucleic acid. P. C. Zamecnik và cộng sự lần đầu tiên phát triển các hệ thống vô bào để nghiên cứu sinh tổng hợp protein, đã xác định các hạt ribonucleoprotein (ribosome) là nơi sinh tổng hợp protein và đã phát hiện ra tRNA.

2. Các phương pháp tách chiết nucleic acid

Để đảm bảo cho các nghiên cứu tiếp theo, nucleic acid cần được tách với một lượng đủ lớn và đủ tinh sạch. Chính vì vậy điều cần chú ý trước hết là phải thu nhận các nucleic acid ở trạng thái nguyên vẹn, không bị phân hủy bởi các tác nhân cơ học hay hóa học. Việc nghiền mạnh không đúng quy trình cũng dễ làm tổn thương đến phân tử nucleic acid (dễ bị gãy). Các enzyme nội bào bị phá vỡ cũng có thể thủy phân nucleic acid.

Để ức chế hoạt động của các enzyme nội bào (deoxyribonuclease

= DNase, và ribonuclease = RNase) cần tách chiết nucleic acid ở nhiệt độ thấp, hoặc sử dụng các chất ức chế sự hoạt động của các enzyme này.

2.1. Phương pháp tách chiết DNA

DNA là phân tử có kích thước lớn nên cần tránh các tác nhân có học hoặc hóa học mạnh, tránh làm đứt gãy. Có nhiều phương pháp khác nhau được sử dụng để tách chiết DNA, như: ly tâm gradient tỷ trọng CICs, cột sắc ký trao đổi anion (Kit QIAGEN), v.v. Dưới đây chỉ là một trong các phương pháp tách chiết DNA đó.

Phương pháp tách chiết này gồm 3 bước cơ bản sau:

- Bước 1: Phá vỡ màng tế bào và màng nhân (ở tế bào eukaryote) bằng hỗn hợp chất tẩy (SDS - sodium dodecyl sulfate, sarcosyl) và proteinase. Các DNA sẽ được giải phóng ra môi trường. Các protein liên kết với DNA cũng tự bị phân hủy.

- Bước 2: Loại bỏ thành phần không mong muốn trong mẫu mà chủ yếu là protein bằng dung dịch: phenol chloroform. Dung dịch này làm biến tính protein và không hòa tan nucleic acid. Sau khi ly tâm loại protein (nằm giữa pha nước và pha phenol: chloroform) sẽ thu được nucleic acid (ở pha nước).

- Bước 3: Kết tủa nucleic acid. Mục đích là thu nhận nucleic acid dưới dạng cô đặc và bảo vệ chúng khỏi sự phân hủy của các enzyme cũng như có thể hòa tan trở lại theo nồng độ mong muốn.

- + Dùng ethanol nồng độ cao (2,5 dung tích ethanol/1 dung tích mẫu) trong môi trường có lực ion cao (nồng độ muối cao), ở nhiệt độ thấp. Hầu như toàn bộ nucleic acid đều kết tủa trong điều kiện này.

- + Dùng isopropanol (thể tích dung môi: thể tích mẫu là 1:1), không có sự hiện diện của muối. Các DNA trọng lượng phân tử thấp không bị kết tủa, do đó có thể loại chúng ra khỏi dịch chiết bằng cách này.

Tủa thu được bằng 2 cách trên được thu nhận lại bằng cách ly tâm. Sau đó được rửa lại bằng ethanol 70% để loại bỏ các muối và các dấu vết isopropanol.

Và cuối cùng là, tiến hành xử lý bằng RNase để loại bỏ RNA.

2.2. Phương pháp tách chiết RNA toàn phần và mRNA

Các loại RNA đều là các phân tử không bền, dễ bị phân hủy bởi

các enzyme ribonuclease (RNase). Hoạt tính của các enzyme này rất cao và bền vững với các tác nhân thường dùng để bất hoạt enzyme (như xử lý 90°C trong 1^{h} không làm mất hoạt tính nó). RNase lại có mặt khắp nơi (ví dụ trên đầu ngón tay của người thao tác...). Chính vì vậy, cần phải có nhiều biện pháp thận trọng để tránh các tạp nhiễm bởi các RNase từ môi trường: thao tác trong điều kiện vô trùng, mọi dụng cụ hóa chất đều được khử trùng bằng nhiệt hay được xử lý với DEPC hầu tránh mọi tiếp xúc với dụng cụ bằng tay trần.

- Việc tách chiết RNA toàn phần cũng gồm 3 bước cơ bản như đối với DNA: (Lưu ý rằng, đây cũng chỉ là một trong các phương pháp được sử dụng mà thôi. Hiện nay phương pháp được sử dụng phổ biến nhất để chiết tách RNA là *Tri201*.)

+ Nghiền tế bào mô trong dung dịch có một chất tẩy mạnh (SDS, sarcosyl) ở nồng độ cao, một tác nhân gây biến tính protein mạnh (guanidinium thiocyanate), một chất khử (2-mercaptoethanol). Các chất này có tác dụng ức chế hoạt động của các RNase nội bào và tách các protein liên kết khỏi phân tử RNA.

+ Loại protein bằng xử lý phenol: chloroform và ly tâm.

+ RNA được kết tủa bằng ethanol và thu lại bằng ly tâm. Dưới dạng kết tủa trong ethanol hoặc đông lạnh ở -70°C trong nước có chứa chất ức chế RNase là R-nasine, RNA có thể được bảo quản trên một năm. Và, cuối cùng là bước tách RNA khỏi DNA.

- Tách chiết mRNA (ở các tế bào nhân chuẩn)

Khoảng 90-99% tổng số RNA tế bào là rRNA (80 - 85%), tRNA (15 - 20%), snRNA (<1%). Chúng có kích thước và trình tự xác định và có thể được tách riêng bằng điện di, ly tâm.... mRNA chiếm khoảng 1 - 5% tổng số RNA tế bào. Kích thước và trình tự của loại này vô cùng đa dạng. Tuy vậy, chúng có một đặc điểm chung là có cấu trúc đuôi polyA (có thể lên đến 100A). Người ta có thể tách mRNA ra khỏi mẫu bằng cách dựa vào cấu trúc trên và đặc tính liên kết bổ sung A-T của nucleic acid, sử dụng sắc ký ái lực trên cột oligodT - cellulose. Các bộ mẫu thử (các kit) được xuất hiện trên thương trường hiện nay sử dụng các viên bi từ có mang oligodT trên bề mặt là dựa vào nguyên tắc đã nêu trên. Ràng liên kết bổ sung với oligodT, sau khi các mRNA bám lên bề mặt các viên bi từ, chúng sẽ được thu nhận lại qua ly tâm hoặc sử dụng nam châm và mRNA sẽ được tách khỏi các viên bi và giữ lại. Bằng cách này có thể thu nhận mRNA từ những mẫu có khối lượng rất

nhỏ.

Sau khi tách chiết các nucleic acid có thể được tinh sạch bằng các phương pháp siêu ly tâm, sắc ký hay điện di.

2.3. Phương pháp siêu ly tâm

Siêu ly tâm trên một gradient liên tục cesium chloride (CsCl). Khi ly tâm, dung dịch CsCl đậm đặc trong ống ly tâm sẽ tự động hình thành một gradient tỷ trọng với tỷ trọng tăng dần từ miệng ống xuống đáy ống. Dưới tác dụng của lực ly tâm, các nucleic acid di chuyển trong ống đến vị trí có tỷ trọng bằng với tỷ trọng của chính chúng sẽ đạt thể cân bằng và ngừng lại, hình thành một lớp cố định trong ống. Sau khi ly tâm, lớp này sẽ được thu nhận lại. Phương pháp này thường được dùng để tinh sạch plasmid và phage.

- Siêu ly tâm trên đệm CsCl (gradient không liên tục): Hỗn hợp nhiều nucleic acid có tỷ trọng biết trước khác nhau được đặt trên một lớp dung dịch CsCl (đệm). Trong quá trình ly tâm chỉ có những phân tử có tỷ trọng cao hơn tỷ trọng lớp đệm mới di chuyển qua được lớp đệm. Ở đây, ống ly tâm bao gồm nhiều lớp đệm có tỷ trọng tăng dần từ miệng đến đáy ống. Nucleic acid cần tinh sạch sẽ nằm ở mặt phân cách hai lớp đệm. Phương pháp này thường dùng để tinh sạch một lượng lớn phage.

- Siêu ly tâm trên gradient saccharose: Thường dùng để phân tích thô một hỗn hợp có kích thước chênh lệch nhau nhiều kb. Ứng dụng chúng trong chọn lọc các đoạn DNA có kích thước xác định dùng trong việc thiết lập các ngân hàng gen.

2.4. Phương pháp sắc ký

Có thể dùng các phương pháp sắc ký khác nhau để phục vụ cho các mục đích khác nhau trong tách chiết nucleic acid.

- Sắc ký ái lực trên polyU - sepharose hay trên oligodT - cellulose để tinh sạch mRNA.

- Sắc ký lọc gel trong phân tách các nucleic acid ra khỏi các nucleic tự do sau quá trình đánh dấu DNA hoặc RNA.

- Sắc ký trao đổi ion trên vi cột để thu hồi những lượng rất nhỏ DNA.

- Sắc ký lỏng cao áp (High Performance Liquid Chromatography - HPLC). Phương pháp có độ phân giải cao này được dùng trong tinh sạch các oligonucleotide tổng hợp (độ phân

giải một nucleotide), plasmid, phân tách các đoạn DNA.

3. Các phương pháp phân tích định tính và định lượng thô nucleic acid

Muốn định tính và định lượng các nucleic acid, người ta phải tiến hành thu nhận chúng ở dạng sạch. Các phương pháp thường dùng trong định tính và định lượng chúng là phương pháp đo mật độ quang, điện di, siêu ly tâm, sắc ký.

3.1. Phương pháp định lượng bằng quang phổ kế

Phương pháp này cho phép ước lượng tương đối nồng độ nucleic acid có trong mẫu phục vụ yêu cầu nghiên cứu.

Nguyên tắc của phương pháp đã được trình bày ở phần trước.

Để kiểm tra độ sạch của dung dịch, người ta đo thêm giá trị OD ở 280 nm (OD_{280nm}). Ở bước sóng này, các protein có mức hấp thụ cao nhất. Ngoài ra, các protein cũng hấp thụ ánh sáng ở bước sóng 260 nm như các nucleic acid và do đó làm sai lệch giá trị thật của nồng độ nucleic acid. Một dung dịch nucleic acid được xem là sạch (không tạp nhiễm protein) khi tỷ số OD_{260nm}/OD_{280nm} đối với DNA là $OD_{260nm}/OD_{280nm} \geq 1,8$; còn đối với RNA là $OD_{260nm}/OD_{280nm} \geq 2$.

3.2. Phương pháp điện di gel

Cơ sở của phương pháp này đã được trình bày ở phần trước. Hai loại gel được sử dụng trong nghiên cứu nucleic acid là gel polyacrylamide và gel agarose. Việc chọn loại gel cũng như nồng độ các chất tạo thành gel tùy thuộc kích thước trung bình của các đoạn nucleic acid cần phân tách. Mỗi tương quan giữa nồng độ agarose và acrylamide cần sử dụng để phân tách các trình tự có kích thước xác định được thể hiện qua bảng 1.1.

3.2.1. Gel polyacrylamide

Thường dùng để tách các đoạn kích thước nhỏ dưới 1000 cặp base. So với gel agarose thì thao tác với gel này phức tạp hơn. Vì vậy, gel này chỉ được dùng cho những mục đích đặc hiệu. Ứng dụng chủ yếu của loại gel này là:

- Tinh sạch các oligonucleotide tổng hợp
- Xác định trình tự DNA
- Tách các đoạn DNA có chiều dài dưới 500 cặp base

Gel được đổ giữa hai tấm thủy tinh và điện di thực hiện theo chiều thẳng đứng.

Bảng 1.1 Tương quan giữa nồng độ gel và kích thước các đoạn cần phân tách

% acrylamide	Kích thước các đoạn cần phân tách (cặp base: bp)
4	200 - 800
5	80 - 200
8	40 - 100
11	10 - 50
% agarose	Kích thước các đoạn cần phân tách (kb)
0,6 - 0,8	1 - 20
0,9 - 1,2	0,5 - 7
1,2 - 1,5	0,2 - 5

3.2.2. Gel agarose

Là loại gel thông dụng nhất. Thao tác với loại gel này đơn giản, thường dùng để phân tách những đoạn có kích thước trong khoảng 0,5 - 20kb. Gel được đổ trên một giá thể nằm ngang và điện di được thực hiện theo phương nằm ngang.

Trong gel agarose các nucleic acid sẽ hiện hình dưới tia tử ngoại (UV) nhờ ethidium bromide. Hóa chất này có khả năng gắn xen vào giữa các base của nucleic acid và sẽ phát huỳnh quang dưới tác dụng của tia tử ngoại. Dưới sự chiếu sáng bằng tia tử ngoại ($\lambda = 260 - 360 \text{ nm}$), nucleic acid sẽ hiện hình dưới dạng những vạch màu đỏ cam.

Để ước lượng kích thước các trình tự nucleic acid trong gel agarose người ta sử dụng một yếu tố đánh dấu trọng lượng phân tử (molecular weight marker - MWM). Đó là tập hợp nhiều trình tự DNA có kích thước đã biết (thang DNA - DNA ladder).

Điện di trên gel agarose còn được sử dụng để tinh sạch và thu nhận mẫu. Ở đây, sau khi điện di, các vạch tương ứng với DNA cần tinh sạch được phát hiện và thu nhận lại theo một trong các phương pháp sau:

(i) Cắt phần agarose chứa các vạch DNA, sau đó thu nhận lại DNA bằng cách khuếch tán từ gel agarose vào một dung dịch đệm thích hợp.

(ii) Khoét một “giếng” nhỏ trong agarose ngay dưới vạch DNA. Bơm dung dịch đệm vào đầy “giếng” và điện trường được tái lập. DNA di chuyển vào “giếng” chứa đầy dung dịch đệm và được thu nhận lại.

(iii) Dùng gel agarose đặc biệt như Nusieve hay Seaplaque có điểm nóng chảy rất thấp (65°C) cho điện di. Sau điện di, vạch DNA được cắt ra và ủ trong một dung dịch đệm ở nhiệt độ 65°C . Khi agarose đã hoàn toàn tan chảy DNA được thu nhận lại sau nhiều công đoạn tách chiết và kết tủa.

4. Các phương pháp xác định trình tự của nucleic acid

Các phương pháp phân tích nucleic acid đã nêu đã cung cấp nhiều thông tin về nucleic acid nhưng chưa cho phép kết luận về bản chất của một đoạn nucleic acid cụ thể. Thông tin về sự tương ứng của nucleic acid với gene gì, có chức năng điều hòa hay mã hóa cho protein nào, chỉ có thể rút ra được từ việc xác định trình tự nucleotide của nucleic acid.

Các phương pháp xác định trình tự nucleic acid nói chung dựa vào hai nguyên tắc cơ bản, được tóm tắt dưới đây (về chi tiết, có thể tham khảo trong: Hồ Huỳnh Thuỳ Dương 1997; Đỗ Quý Hai 2001.):

- Nguyên tắc hóa học (Phương pháp Maxam - Gilbert, 1977): dựa vào các phản ứng hóa học thủy phân đặc hiệu phân tử DNA, tạo thành một tập hợp nhiều phân đoạn có kích thước khác nhau.

- Nguyên tắc enzyme học (Phương pháp Sanger (1977) và các phương pháp cải biên): Dựa vào sự tổng hợp mạch bổ sung cho trình tự cần xác định nhờ DNA polymerase. Với việc sử dụng thêm dideoxynucleotide (nhóm 3'-ON được thay bằng H) cùng với các deoxy nucleotide thông thường, kết quả tổng hợp cũng là sự hình thành một tập hợp nhiều đoạn DNA có kích thước khác nhau.

Ở cả hai trường hợp, các phân đoạn DNA sẽ được phân tách qua điện di trên gel polyacrylamide có khả năng phân tách hai trình tự DNA chỉ chênh nhau một nucleotide. Với việc sử dụng một đoạn nucleotide có đánh dấu đồng vị phóng xạ, kết quả trình tự cần xác định được đọc trên bản phóng xạ tự ghi từ bản điện di.

Câu hỏi ôn tập

1. Trình bày tóm tắt lịch sử nghiên cứu các nucleic acid.

2. Mô tả các phương pháp chung thường dùng trong nghiên cứu nucleic acid.
3. Trình bày nguyên tắc và ứng dụng của các phương pháp tách chiết nucleic acid.
4. Mô tả các phương pháp phân tích định tính và định lượng thô nucleic acid.
5. Sơ lược về các phương pháp xác định trình tự của các nucleic acid.

Tài liệu Tham khảo

- Nguyễn Bá Lộc. 2000. *Giáo trình Axit nucleic và Sinh tổng hợp protein*. NXB Giáo dục.
- Đỗ Quý Hai. 2001. *Bài giảng Axit Nucleic*. Trường ĐHKH, Đại Học Huế.
- Phạm Thị Trân Châu và Trần Thị Áng (1992): *Hoá sinh học*. NXB Giáo dục, Hà Nội.
- Hồ Huỳnh Thuỳ Dương. 1997. *Sinh học Phân tử*. NXB Giáo Dục.
- Lehninger L. et al. 1993. *Principles of Biochemistry*. Worth Publishers, New York.
- Stryer L. 1981. *Biochemistry*. W.-H. Freeman and Co., San Francisco.

Chương 2

Cấu trúc của các Nucleotide và Polynucleotide

Việc nghiên cứu cấu trúc của các nucleic acid thực sự diễn ra từ giữa thập niên 1950, sau khi O.T.Avery, MacLeod và McCarty (1944) lần đầu tiên chứng minh: DNA là vật chất mang thông tin di truyền. Cho đến nay, chúng ta biết rằng các *nucleic acid*, bao gồm deoxyribonucleic acid (DNA) và ribonucleic acid (RNA), là những đại phân tử sinh học có trọng lượng phân tử lớn với thành phần gồm các nguyên tố C, H, O, N và P; chúng được cấu thành từ nhiều đơn phân (monomer) là các *nucleotide*. Các đơn phân này nối với nhau bằng các liên kết phosphodiester tạo thành các cấu trúc đa phân (polymer) gọi là các chuỗi, mạch (chain) hay sợi (strand) *polynucleotide* - cấu trúc sơ cấp của các phân tử nucleic acid.

Trong chương này, chúng ta lần lượt tìm hiểu thành phần hoá học và cấu trúc của các nucleotide và polynucleotide của DNA và RNA.

I. Thành phần hóa học của các nucleotide

Vào giữa thập niên 1940, các nhà hoá sinh học đã biết được các cấu trúc hoá học của DNA và RNA. Khi phân cắt DNA thành các tiểu đơn vị, họ phát hiện ra rằng mỗi nucleotide của DNA gồm ba thành phần: một base nitơ (nitrogenous base), một đường deoxyribose, và một phosphoric acid. Tương tự, RNA cho ra các base, phosphoric acid và đường ribose. Các nucleotide cũng có nhiều chức năng khác trong tế bào, ví dụ như các dòng năng lượng, các chất dẫn truyền thần kinh và các thông tin loại hai như tải nạp tín hiệu chẳng hạn.

1. Base nitơ

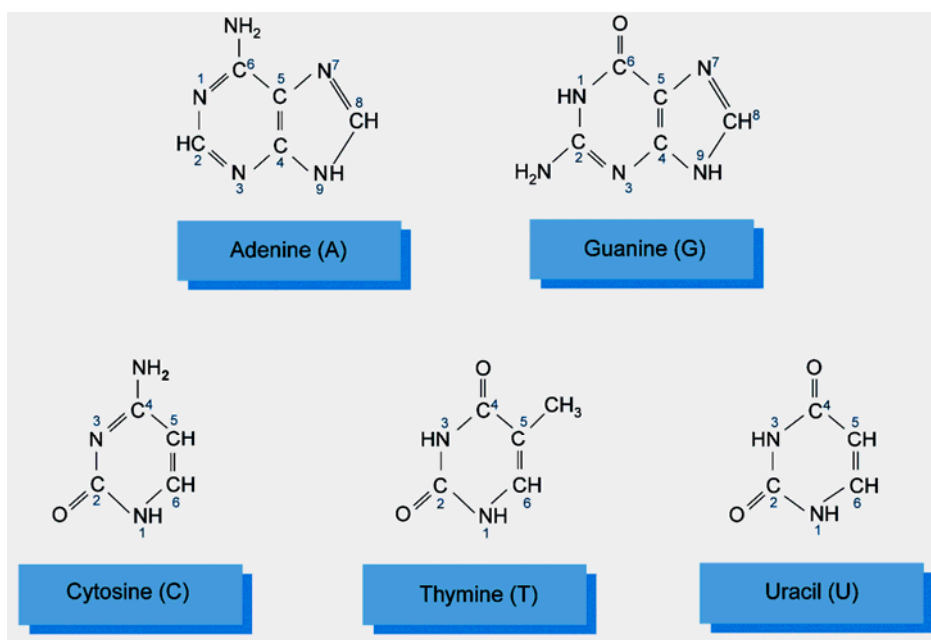
Các base nitơ (gọi tắt là base), thành phần đặc trưng của các nucleotide, là các hợp chất *purine* và *pyrimidine* dị vòng chứa nitơ có tính kiềm. Về cơ bản, các dẫn xuất của purine bao gồm *adenine* (A) và *guanine* (G), còn của pyrimidine gồm có: *thymine* (T), *uracil* (U) và *cytosine* (C).

DNA chứa bốn loại base chính là adenine, guanine, thymine và

cytosine. Trong RNA cũng chứa các base như thế, chỉ khác là *uracil* thay thế *thymine* (Hình 2.1). Cần chú ý rằng purine và pyrimidine là các base dị vòng chứa các nguyên tử nitơ nằm xen với các nguyên tử carbon, nên việc đánh số các vị trí không thêm dấu phẩy trên đầu như trong trường hợp của đường pentose (xem các Hình 2.4 - 2.6).

Bên cạnh các dạng phổ biến nói trên, các purine khác cũng có vai trò quan trọng trong trao đổi chất của tế bào, như: *xanthine*, *hypoxanthine* và *uric acid*; còn đối với pyrimidine đó là các *orotic* và *dihydroorotic acid*.

Ngoài ra còn bắt gặp một số loại base hiếm thuộc cả hai nhóm purine và pyrimidine. Đó là những base biến đổi chủ yếu do hiện tượng *methyl hoá* (methylation) xảy ra ở các vị trí khác nhau, chẳng hạn: 1-methyladenine, 6-methyladenine, 2-methylguanine, 5-methylcytosine v.v.

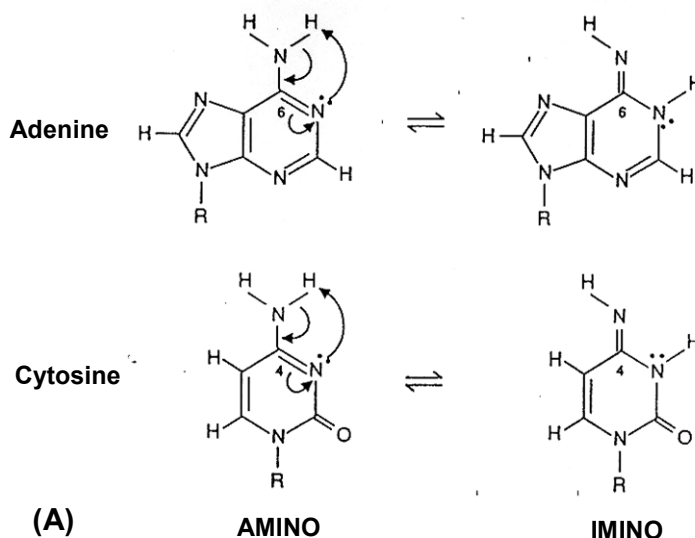


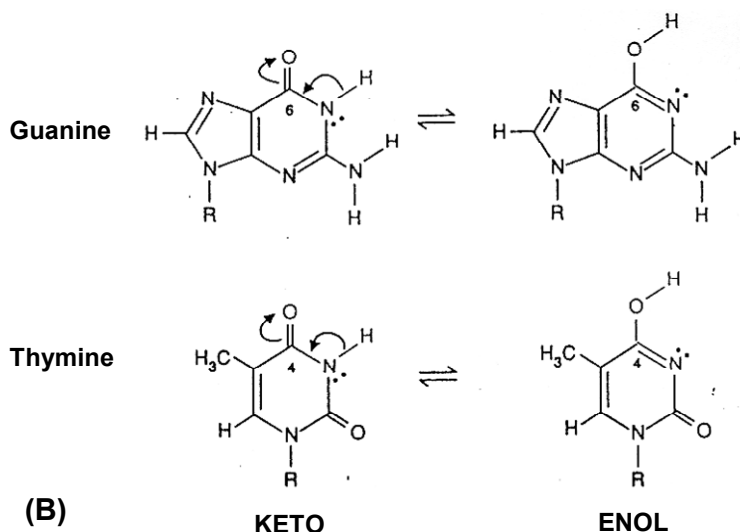
Hình 2.1 Cấu trúc các base của DNA và RNA. Adenine và guanine là các dẫn xuất của purine; còn cytosine, thymine và uracil là các dẫn xuất của pyrimidine; trong đó uracil là đặc thù cho RNA và thymine cho DNA.

Các base purine và pyrimidine có thể tồn tại dưới các dạng *hỗ biến* (tautomeric forms) *amino* và *imino* (đối với adenine và cytosine; Hình 2.2A), hoặc *keto* và *enol* (đối với guanine và thymine; Hình 2.2B). Đó là hai trạng thái tồn tại bền (phổ biến) và

kém bền (ít phổ biến), có thể biến đổi qua lại với nhau do sự dịch chuyển vị trí của các nguyên tử hydro trong các base purine và pyrimidine. Hình 2.2 cho thấy các dạng hồ biến của các base trong DNA. Tương tự, uracil có hai dạng hồ biến: lactam (dạng *keto*) chiếm ưu thế ở pH = 7 và lactim (dạng *enol*) gia tăng khi pH giảm. Chính hiện tượng hồ biến này dẫn tới thay đổi khả năng kết cặp bình thường của các base và làm phát sinh các đột biến gene dạng thay thế một cặp base.

Các base phổ biến trong cả DNA và RNA là tương đối bền vững ở trạng thái hồ biến được gọi là *dạng hồ biến ưu thế* (dominant tautomeric form); có lẽ đó là lý do tại sao chúng được chọn lọc để mang thông tin di truyền. Nói chung, các base này đều ít tan trong nước và có khả năng hấp thu ánh sáng cực đại ở 260-270 nanomet ($1\text{nm} = 10^{-9}\text{m}$). Chúng có thể được tách ra bằng các phương pháp sắc ký và điện di.



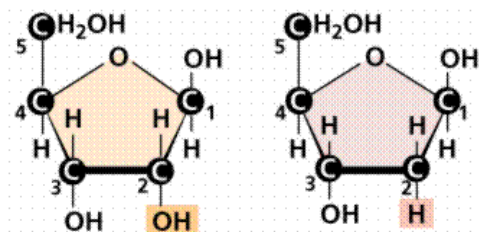


Hình 2.2 Các dạng hỗ biến của các base trong DNA. (A) Các dạng amino (phổ biến) của adenine và cytosine có thể biến đổi thành các dạng imino; và (B) các dạng keto (phổ biến) của guanine và thymine có thể sắp xếp lại thành các dạng enol. Các mũi tên biểu thị sự dịch chuyển vị trí nguyên tử hydro. R là các gốc đường và phosphate.

2. Đường pentose

Các đường chứa năm carbon (pentose) là sản phẩm của quá trình trao đổi chất trong tế bào, với nhiều loại như: arabinose, ribulose, ribose và dẫn xuất của nó là deoxyribose v.v.

Đường pentose của RNA là *D-ribose* và của DNA là *2'-deoxy-D-ribose* (ký hiệu D chỉ dạng đường quay phải trước ánh sáng phân cực để phân biệt với dạng L quay trái không có trong thành phần của các nucleic acid tự nhiên). Các phân tử đường này đều có cấu trúc vòng *furanose* (gọi như thế bởi vì nó giống với hợp chất *furan* dị vòng). Do các nguyên tử carbon ở đây xếp liên tục nên được đánh số thứ tự có dấu phẩy trên đầu, ví dụ C_{1'}, C_{2'} cho đến C_{5'}.



Hình 2.3 Cấu trúc của các phân tử đường ribose (trái) và deoxyribose (phải); chúng khác nhau ở nguyên tử carbon số 2.

Hai phân tử đường này khác nhau ở C_2' ; trong ribose đó là nhóm hydroxyl và trong deoxyribose là một hydro (Hình 2.3). Do các gốc đường khác nhau này đã tạo ra hai loại nucleotide là *ribonucleotide* và *deoxyribonucleotide*, mà từ đó cấu tạo nên hai loại nucleic acid khác nhau tương ứng là RNA và DNA. Và chính sự khác biệt nhỏ nhất về mặt cấu trúc này đã tạo nên các đặc tính hoá lý rất khác nhau giữa DNA và RNA. Dung dịch DNA tỏ ra đặc quánh hơn nhiều do sự trở ngại lập thể (steric hindrance) và miễn cảm hơn với sự thuỷ phân trong các điều kiện kiềm (alkaline), có lẽ điều này giải thích phần nào tại sao DNA xuất hiện như là vật chất di truyền sơ cấp (primary genetic material).

Cần đề ý rằng, trong các phân tử đường này có ba vị trí quan trọng có chứa nhóm hydroxyl (-OH) tự do, đó là: (i) nhóm -OH ở vị trí C_1' có khả năng hình thành liên kết N-glycosid với gốc -NH của các base để tạo thành các nucleoside; (ii) nhóm -OH ở vị trí C_5' có khả năng hình thành liên kết ester với nhóm phosphate để tạo ra các nucleotide; và (iii) nhóm -OH ở vị trí C_3' có khả năng hình thành liên kết phosphodiester với nhóm phosphate của một nucleotide khác để tạo chuỗi polynucleotide. Như vậy, tính phân cực (polarity) trong gốc đường mà từ đó quyết định tính phân cực của các chuỗi polynucleotide được thể hiện ở hai vị trí C_5' và C_3' .

3. Phosphoric acid

Phosphoric acid (H_3PO_4) là acid vô cơ có chứa phosphor (P), một nguyên tố đóng vai trò quan trọng trong trao đổi chất và năng lượng của tế bào. Do có chứa ba nhóm -OH nên acid này có thể hình thành liên kết ester với các gốc đường tại các vị trí C_5' và C_3' để tạo nên các nucleotide và chuỗi polynucleotide.

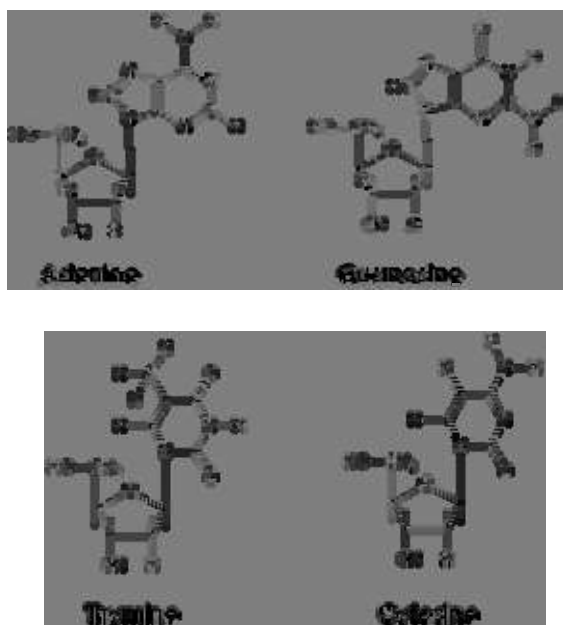
Trong các nucleotide của DNA và RNA, nhóm phosphate liên kết với các nucleoside tại C_5' (xem Hình 2.5). Trong trường hợp phân tử điều hoà *AMP vòng* (cyclic AMP = cAMP), nhóm phosphate tạo liên kết ester với hai nhóm -OH ở C_5' và C_3' trong cùng một nucleotide.

II. Cấu trúc của các nucleotide

1. Cấu trúc của các nucleoside

Các base và đường trong RNA và DNA được nối với nhau thành các đơn vị gọi là *nucleoside*. Mỗi nucleoside được tạo thành do một base nối với một đường pentose tại vị trí C_1' bằng một *liên*

kết β -N-glycosid. Cụ thể là, nguyên tử carbon $C_{1'}$ của đường nối với nguyên tử N_1 của pyrimidine hoặc với nguyên tử N_9 của purine (xem các Hình 2.4 - 2.6).



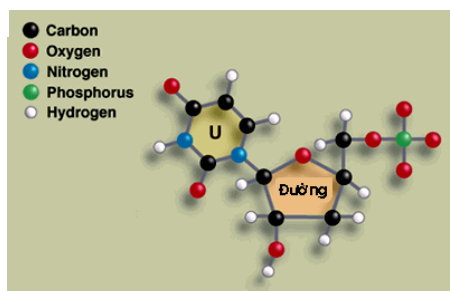
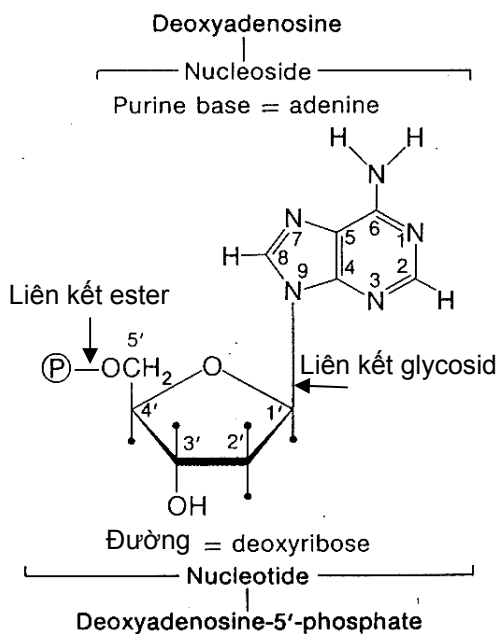
Hình 2.4 Cấu trúc của bốn loại deoxynucleoside trong DNA.

Tên gọi chính thức hay danh pháp của các nucleoside bắt nguồn từ các base tương ứng, trong đó các nucleoside là dẫn xuất của purine có đuôi -osine và các dẫn xuất của pyrimidine có đuôi -idine (Bảng 2.1).

2. Cấu trúc của các nucleotide

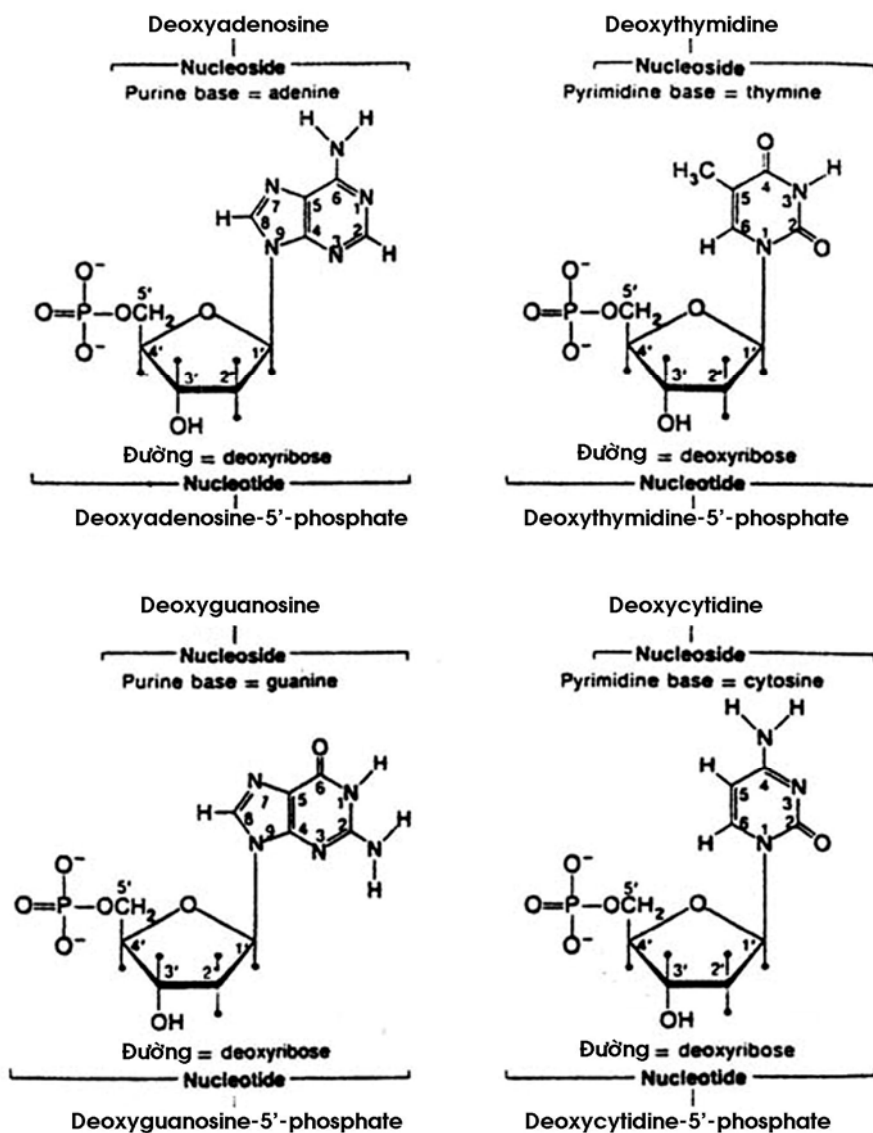
Đơn vị cấu trúc cơ sở của các nucleic acid là các nucleotide. Các nucleotide là những ester phosphate của các nucleoside. Hiện tượng *ester hoá* (esterification) có thể xảy ra ở bất kỳ nhóm hydroxyl tự do nào, nhưng phổ biến nhất là ở các vị trí 5' và 3' trong các nucleic acid.

Về cấu trúc, mỗi nucleotide gồm ba thành phần kết dính với nhau như sau: gốc đường *pentose* nối với một *base* tại $C_{1'}$ bằng một *liên kết β -glycosid* và nối với nhóm *phosphate* tại $C_{5'}$ bằng một *liên kết phosphomonoester* (Hình 2.5 và 2.6).



Hình 2.5 Cấu trúc của một deoxyribonucleotide (dAMP, bên trái) và một ribonucleotide (UMP). Ở đây cho thấy các mối liên kết N-glycosid và ester.

Như thế tính phân cực trong cấu trúc một nucleotide thể hiện ở các nhóm hydroxyl thuộc hai vị trí C_{5'} (tạo liên kết ester với nhóm phosphate trong từng nucleotide) và C_{3'} (tạo liên kết phosphodiester với nucleotide khác trong chuỗi polynucleotide).



Hình 2.6 Cấu trúc chi tiết của bốn loại deoxyribonucleotide trong DNA. Ở đây cũng chỉ ra danh pháp của các nucleoside và nucleotide.

3. Cấu trúc của các nucleoside di- và triphosphate

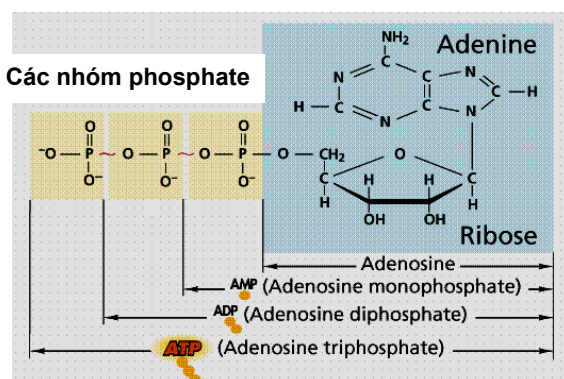
Như đã đề cập, mỗi nhóm phosphate (phosphate group) được nối với vòng của gốc đường bằng một *liên kết phosphomonoester*, và nhiều nhóm phosphate có thể nối nhau thành một dãy bằng các liên kết phosphoanhydride (hình 2.7). Sự ester hoá ở C_{5'} có thể đi

với mono-, di- hoặc triphosphate (nguyên tử phosphor P được đánh dấu tương ứng với các vị trí từ C_{5'} hướng ra ngoài là α , β và γ). Các nucleoside 5'-triphosphate là những hợp chất cho tổng hợp nucleic acid. Hai nhóm hydroxyl cũng có thể được ester hoá bằng cách nối cùng một nhóm phosphate để sinh ra một *nucleotide vòng* (cyclic nucleotide), ví dụ cAMP là 3'-5'-cyclic phosphate (đóng vai trò tải nạp tín hiệu, điều hoà dương tính operon lactose; xem chương 6).

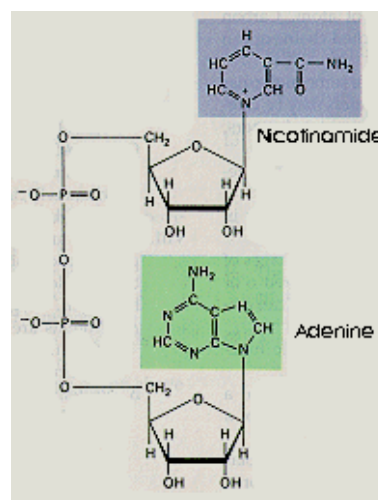
Bảng 2.1 Danh pháp các nucleoside của RNA và DNA

Base	Nucleoside (RNA)	Deoxynucleoside (DNA)
Purine		
Adenine	Adenosine = A	deoxyadenosine = dA
Guanine	Guanosine = G	deoxyguanosine = dG
Hypoxanthine*	Inosine* = I	Không có
Pyrimidine		
Cytosine	Cytidine = C	deoxycytidine = dC
Thymine	Thường không có	(deoxy)thymidine = dT**
Uracil	Uridine = U	Thường không có

Ghi chú: * Đây là dạng hiếm, có mặt trong thành phần của các RNA vận chuyển. ** Bởi vì thymine thường không có trong RNA, nên tiếp đầu ngữ "deoxy" chỉ cho loại deoxynucleoside này thường được lược bỏ và gọi tắt là thymidine. Tuy nhiên, trong các RNA vận chuyển thường có *ribothymidine* chứa đường ribose.



(a)



(b)

Hình 2.7 (a) Cấu trúc chi tiết của các nucleotide adenosine ở ba trạng thái mono-, di- và triphosphate; và (b) *nicotinamide adenosine diphosphate* (NADP).

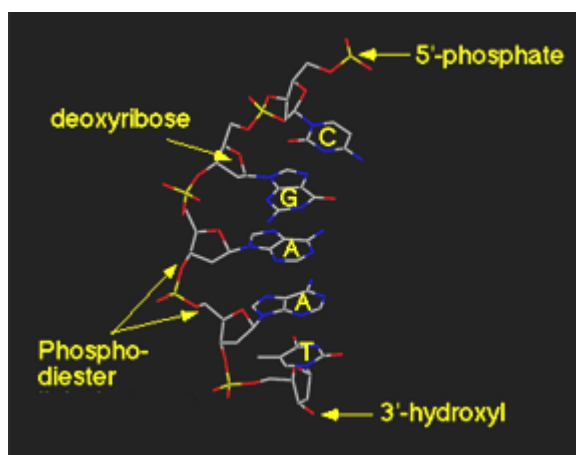
Nói chung, các nucleotide thường có tính acid mạnh và tan trong nước. Các nucleoside monophosphate được xem là các axit đúng như tên gọi phản ánh (ví dụ AMP là adenylic acid hay adenylate); chúng có sự ion hoá sơ cấp với pKa 1-2 và ion hoá thứ cấp với pKa 6,5-7,0, như sau:



Tất cả các phosphate của các nucleoside di- và triphosphate đều ion hoá, nhưng chỉ nhóm tận cùng là có ion hoá thứ cấp. Các nucleotide này đều có ái lực với cation hoá trị hai như Mg^{2++} và Ca^{2++} (chúng tương tác với các nhóm phosphate α và β hoặc β và γ).

III. Cấu trúc của các chuỗi polynucleotide

Các nucleotide trong DNA hoặc RNA nối với nhau bằng các mối liên kết đồng hoá trị (covalent) có tên là *liên kết 3',5'-phosphodiester* giữa gốc đường của nucleotide này với nhóm phosphate của nucleotide kế tiếp, tạo thành chuỗi *polynucleotide*. Vì vậy các chuỗi này bao giờ cũng được kéo dài theo chiều 5'→3' (đầu 5' mang nhóm phosphate tự do và đầu 3' chứa nhóm -OH tự do). Chúng có bộ khung vững chắc gồm các gốc đường và phosphate xếp luân phiên nhau, còn các base nằm về một bên. Trình tự các base vì vậy được đọc theo một chiều xác định 5'→3'. Đây là cấu trúc hoá học sơ cấp của DNA và RNA (Hình 2.8 và 2.9).



Hình 2.8 Mô hình cấu trúc chuỗi polynucleotide DNA. Ở đây cho thấy các vị trí 5'-phosphate và 3'-hydroxyl cũng như đường deoxyribose và liên kết

phosphodiester nối giữa các gốc đường này.

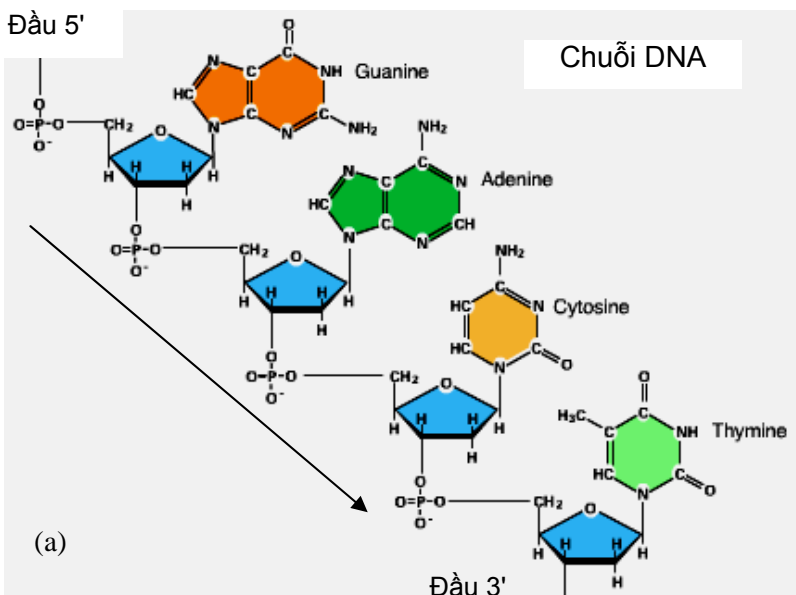
Thông thường người ta biểu diễn trình tự base 5'→3' theo chiều từ trái sang phải. Hình 2.8 cho thấy các chuỗi DNA và RNA chỉ khác nhau bởi base U hoặc T và gốc đường trong các nucleotide của chúng. Nếu bỏ qua sự khác biệt về gốc đường, ta có thể hình dung trình tự các base của hai chuỗi polynucleotide của DNA và RNA đều sinh trưởng theo chiều từ 5' đến 3' (5'→3'), như sau:

Chuỗi DNA: (5') pApApTpTpCpTpTpApApApTpTpC -OH (3')

Chuỗi RNA: (5') pApApUpUpCpUpUpApApApUpUpC -OH

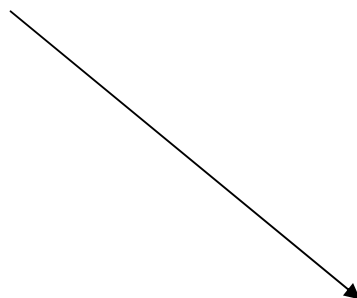
(3')

Đầu 5'



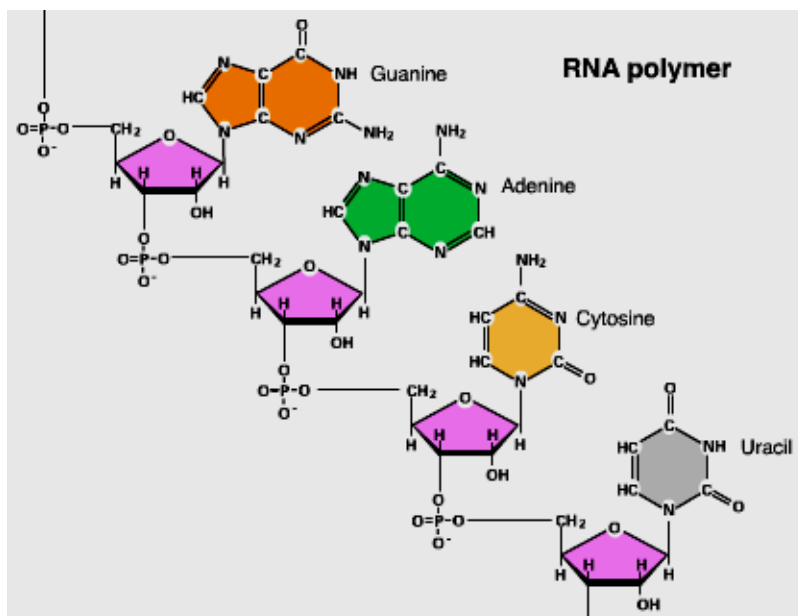
Đầu 5'

Chuỗi RNA



(b)

Đầu 3'



Hình 2.9 Cấu trúc chuỗi polynucleotide của DNA (a) và của RNA (b). Các chuỗi polynucleotide bao giờ cũng được tổng hợp (kéo dài) theo chiều 5'→3'; chúng có bộ khung "đường-phosphate" rất vững chắc và trình tự base được viết theo quy ước từ trái (đầu 5') sang phải (đầu 3') đối với chuỗi DNA ở đây như sau: 5'...dGdAdCdT...3', còn đối với chuỗi RNA là 5'...GACU...3'.

Cần lưu ý rằng, các hợp chất cho polymer hoá là các nucleoside triphosphate, nhưng các monomer của một nucleic acid là monophosphate. Phản ứng trùng hợp này được xúc tác bởi các enzyme tương ứng là *DNA-* hoặc *RNA polymerase* và sinh ra các pyrophosphate (có thể xem thêm trong: *tái bản DNA* và *phiên mã*; chương 5 và 6).

Các *oligonucleotide* là những nucleic acid ngắn (nghĩa là có độ dài dưới 100 nucleotide). Các oligoribonucleotide tồn tại trong tự nhiên và được sử dụng như là những *đoạn mồi* (primer) trong tái bản DNA và cho các mục đích khác nhau trong tế bào (xem chương 5). Các oligonucleotide tổng hợp có thể tạo ra bằng sự tổng hợp hoá học và là nguyên liệu thiết yếu cho các kỹ thuật thí nghiệm [ví dụ như dùng để giải mã di truyền, chương 6; có thể tham khảo thêm các kỹ thuật *xác định trình tự DNA* (DNA sequencing), *phản ứng trùng hợp chuỗi bằng polymerase* (polymerase chain reaction), *lai in situ* (*in situ* hybridization), *mẫu dò nucleic acid* (nucleic acid probe), *lai nucleic acid* (nucleic acid hybridization), *liệu pháp gene* (gene therapy); các chương 3-6].

Câu hỏi và Bài tập

1. Phân tích các thành phần hoá học của các nucleotide và mối liên hệ giữa chúng.
2. Vẽ sơ đồ cấu tạo của: (a) các base purine và pyrimidine có mặt trong thành phần của DNA và RNA; và (b) các đường ribose và deoxyribose. Từ đó so sánh cấu trúc của các nucleotide DNA và RNA.
3. Phân tích cấu trúc một nucleoside và một nucleotide của DNA và RNA. Cho các sơ đồ minh hoạ.
4. Phân tích sự hình thành chuỗi polynucleotide của DNA và RNA và chỉ ra những điểm giống và khác nhau giữa chúng.
5. Tính phân cực (polarity) thể hiện như thế nào trong cấu trúc của các nucleotide và chuỗi polynucleotide của DNA và RNA? Giải thích và cho các sơ đồ minh hoạ.

Tài liệu Tham khảo

Tiếng Việt

- Nguyễn Bá Lộc. 2004. *Giáo trình Axit nucleic và sinh tổng hợp protein* (tái bản). Trung tâm Đào tạo Từ xa - Đại học Huế.
- Hoàng Trọng Phán. 1995. *Một số vấn đề về Di truyền học hiện đại* (Tài liệu BDTX cho giáo viên THPT chu kỳ 1993-1996). Trường ĐHSP Huế.
- Hoàng Trọng Phán. 1997. *Di truyền học Phân tử*. NXB Giáo Dục.
- Hoàng Văn Tiến (chủ biên), Lê Khắc Thận, Lê Doãn Diên. 1997. *Sinh hoá học với cơ sở khoa học của công nghệ gene*. NXB Nông Nghiệp, Hà Nội.
- Nguyễn Tiến Thắng, Nguyễn Đình Huyền. 1998. *Giáo trình Sinh hoá hiện đại*. NXB Giáo Dục.

Tiếng Anh

- Blackburn G.M., Gait M.J. (Eds., 1996): *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford.
- Lehninger, .L. et al. (1993): *Principles of Biochemistry*. Worth Publishers, New York.
- Russell P.J. 2003. *Essential Genetics*. Benjamin/Cummings

Publishing Company, Inc, Menlo Park, CA.

Stryer, L. (1981): *Biochemistry*. W.-H. Freeman and Co., San Francisco.

Tamarin RH. 1999. *Principles of Genetics*. 6th ed, McGraw-Hill, Inc, NY.

Twyman RM. 1998. *Advanced Molecular Biology*. BIOS Scientific Publishers Ltd/ Springer-Verlag Singapore Pte Ltd.

Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. 1987. *Molecular Biology of the Gene*. 4th ed, Benjamin/Cummings Publishing Company, Inc, Menlo Park, CA.

Weaver RF, Hedrick PW. 1997. *Genetics*. 3rd ed, McGraw-Hill Companies, Inc. Wm.C.Brown Publishers, Dubuque, IA.

Chương 3

Cấu trúc và Đặc điểm của DNA

"DNA - phân tử quý giá nhất trong tất cả các phân tử"

(James D. Watson)

Sự khám phá ra cấu trúc phân tử DNA bởi James Watson và Francis Crick năm 1953 với những hệ quả sinh học của nó là một trong những sự kiện khoa học to lớn nhất của thế kỷ XX. Nếu như sự ra đời của tác phẩm *"Nguồn gốc các loài"* (1859) của R.Ch.Darwin đã tạo nên một cuộc cách mạng to lớn trong tư tưởng nhân loại, thì khám phá này thực sự làm biến đổi hiểu biết của chúng ta về sự sống. Toàn bộ câu chuyện về việc phát minh ra phân tử kỳ diệu này đã được thiên tài Watson miêu tả hết sức sinh động trong cuốn hồi ký nhan đề *"Chuỗi xoắn kép"* (1968).

Trong chương này, chúng ta sẽ lần lượt tìm hiểu thành phần hoá học và cấu trúc của DNA cũng như các đặc tính hoá lý của nó. Từ đó bí ẩn của sự sống dần dần hé mở những lời giải đáp thú vị, với biết bao thành tựu to lớn tác động lên mọi mặt của đời sống - xã hội trong suốt hơn 50 năm qua.

I. Thành phần hóa học của DNA

Năm 1944, Oswald T. Avery và các đồng sự của mình chứng minh DNA là vật chất mang thông tin di truyền, chứ không phải protein. Đến năm 1949, Erwin Chargaff áp dụng phương pháp sắc ký giấy vào việc phân tích thành phần hóa học của DNA các loài khác nhau (Bảng 3.1) đã khám phá ra rằng:

Bảng 3.1 Thành phần base của DNA ở một số loài

Sinh vật	A%	T%	G%	C%	$\frac{A+G}{T+C}$	$\frac{A+T}{G+C}$
Phage lambda	21,3	22,9	28,6	27,2	1,00	0,79
Phage T7	26,0	26,0	24,0	24,0	1,00	1,08
<i>Mycobacterium tuberculosis</i>	15,1	14,6	34,9	35,4	1,00	0,42
<i>Escherichia coli</i>	24,7	23,6	26,0	25,7	1,03	0,93
<i>Aspergillus niger</i> (nấm mốc)	25,0	24,9	25,1	25,0	1,00	1,00
<i>Saccharomyces cerevisiae</i>	31,3	32,9	18,7	17,1	1,00	1,79
<i>Triticum</i> (lúa mì)	27,3	27,1	22,7	22,8	1,00	1,19

<i>Zea mays</i> (ngô)	26,8	27,2	22,8	23,2	0,98	1,17
<i>Salmo salar</i> (cá hồi)	29,7	29,1	20,8	20,4	1,02	1,43
<i>Gallus domestica</i> (gà nhà)	29,5	27,7	22,4	20,4	1,08	1,34
<i>Homo sapiens</i> (người)	30,9	29,4	19,9	19,8	1,01	1,52

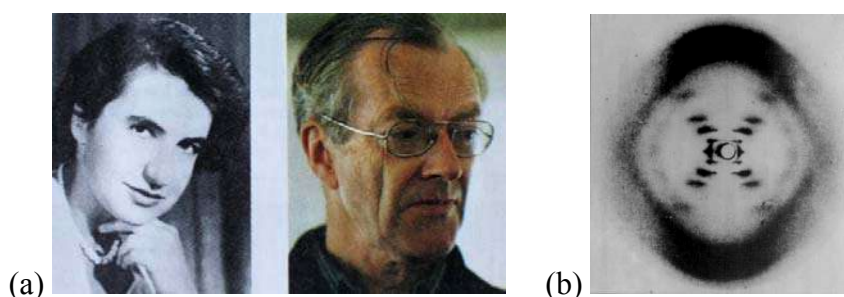
(i) Số lượng bốn loại base trong DNA là không bằng nhau;

(ii) Tỷ lệ tương đối của các base là không ngẫu nhiên; và trong tất cả các mẫu DNA nghiên cứu tồn tại mối tương quan về hàm lượng (%) giữa các base như sau: $A \approx T$ và $G \approx C$, nghĩa là tỷ số $(A+G) / (T+C) \approx 1$; và

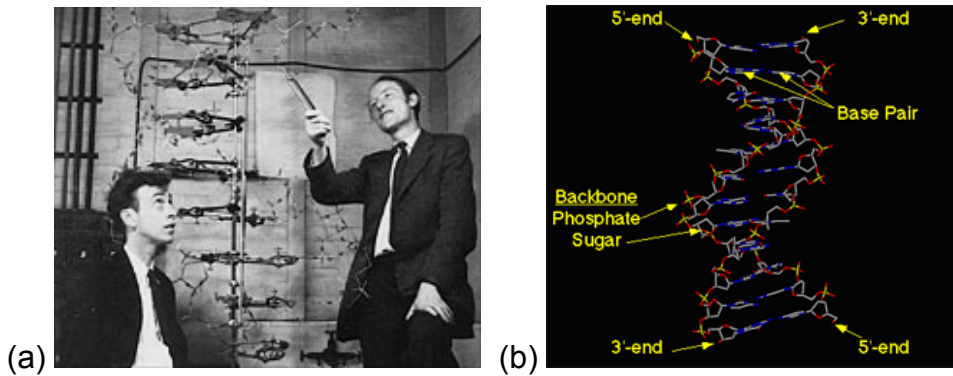
(iii) Mỗi loài có một tỷ lệ $(A+T)/(G+C)$ đặc thù.

II. Cấu trúc chuỗi xoắn kép DNA

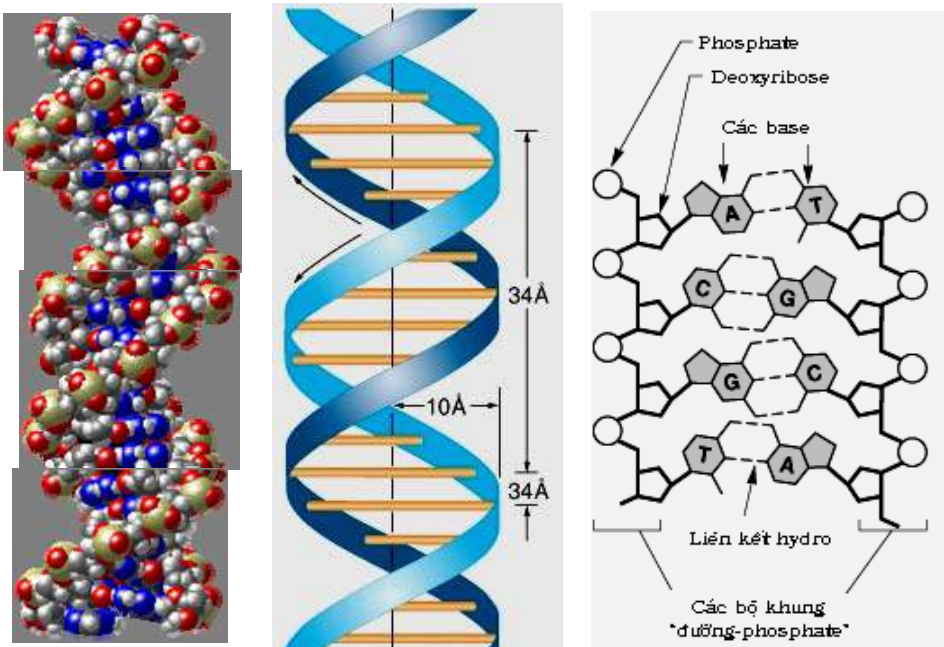
Vào năm 1951-52, việc nghiên cứu cấu trúc ba chiều của DNA bằng phân tích nhiễu xạ tia X được bắt đầu bởi Maurice Wilkins và Rosalind Franklin. Các bức ảnh chụp được 1952 (hình 3.1) gợi ý rằng DNA có cấu trúc xoắn gồm hai hoặc ba chuỗi. Lúc này ở Anh còn có một số nghiên cứu khác nhằm phát triển lý thuyết nhiễu xạ của Linus Pauling để tìm hiểu cấu trúc DNA. Tuy nhiên, giải pháp đúng đắn nhất là *chuỗi xoắn kép* bổ sung do Watson và Crick đưa ra năm 1953 (Hình 3.2 và 3.3). Mô hình này hoàn toàn phù hợp với các số liệu của Wilkins và Franklin cũng như của Chargaff. Sự kiện này mở ra một bước ngoặt mới cho sự ra đời và phát triển với tốc độ nhanh chóng của di truyền học phân tử.



Hình 3.1 (a) R.Franklin (trái) và M.Wilkins; và (b) Ảnh chụp cấu trúc DNA tinh thể bằng tia X của Franklin.



Hình 3.2 (a) J.Watson (trái) và F.Crick; và (b) Mô hình cấu trúc tinh thể DNA.



Hình 3.3 Các mô hình cấu trúc chuỗi xoắn kép DNA.

1. Mô hình Watson-Crick

Mô hình Watson-Crick (DNA dạng B; Hình 3.3) có các đặc điểm sau:

(1) DNA gồm hai chuỗi *đối song song* (antiparallel) cùng uốn quanh một trục trung tâm theo chiều xoắn phải, với đường kính 20\AA ($1\text{\AA} = 10^{-10}\text{m}$), gồm nhiều vòng xoắn lặp lại một cách đều đặn và chiều cao mỗi vòng xoắn là 34\AA , ứng với 10 cặp base (base pair, viết tắt là bp).

(2) Các bộ khung đường-phosphate phân bố ở mặt ngoài chuỗi xoắn và các base nằm ở bên trong; chúng xếp trên những mặt phẳng song song với nhau và thẳng góc với trục phân tử, với khoảng cách trung bình 3,4 Å.

(3) Hai sợi đơn gắn bó với nhau bằng các mối *liên kết hydro* (vốn là lực hóa học yếu) được hình thành giữa các cặp base đối diện theo *nguyên tắc bổ sung* "một purine - một pyrimidine". Cụ thể là, trong DNA chỉ tồn tại hai kiểu kết cặp base đặc thù là A-T (với hai liên kết hydro) và G-C (với ba liên kết hydro) (Hình 3.3 và 3.4).

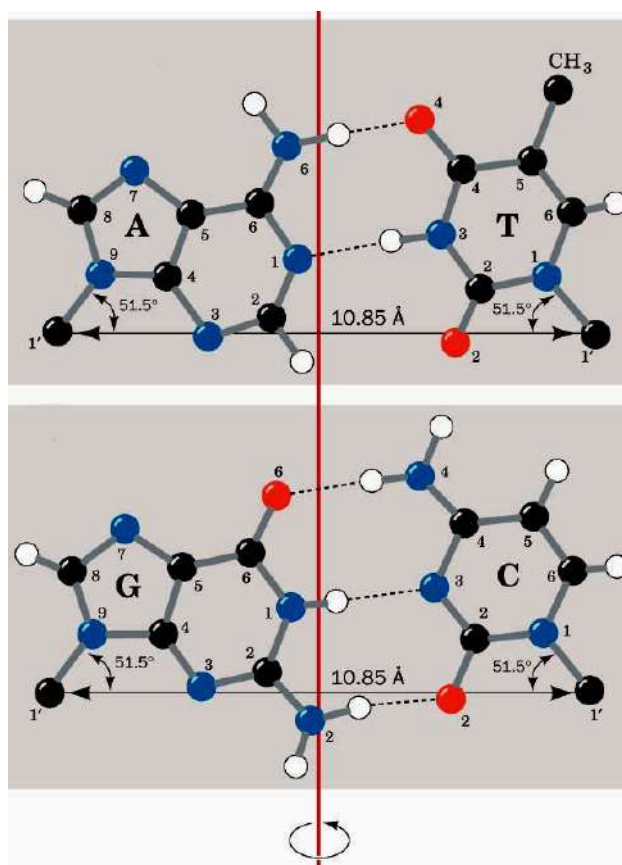
(4) Tính chất bổ sung theo cặp base dẫn đến sự bổ sung về trình tự các base giữa hai sợi đơn của mỗi chuỗi xoắn kép. Vì vậy, trong bất kỳ một phân tử DNA sợi kép nào hoặc một đoạn của nó bao giờ cũng có: $A = T$ và $G = C$; nghĩa là: $[A + G] = [T + C]$ hay $\frac{A+G}{T+C} = 1$ (đây là tỷ số giữa các base purine và các base pyrimidine),

còn tỷ lệ $\frac{A+T}{G+C}$ là đặc thù cho từng loài (thực chất đây là tỷ lệ giữa hai base không bổ sung cho nhau hoặc giữa hai base cùng nhóm, ví dụ A/G hoặc T/C).

Như vậy, mô hình cấu trúc chuỗi xoắn kép của Watson-Crick (1953) hoàn toàn thỏa mãn và cho phép lý giải một cách thỏa đáng các kết quả nghiên cứu của Chargaff (1949). Vì vậy người ta gọi các biểu thức $A = T$ và $G = C$ là các quy luật hay *quy tắc Chargaff* (Chargaff's rules).

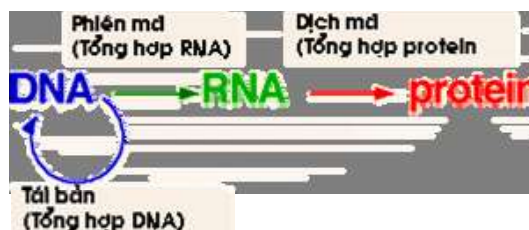
Theo nguyên tắc bổ sung của các cặp base, ta có thể xác định trình tự base ở sợi bổ sung khi biết được trình tự base của một sợi đơn. Ví dụ:

Sợi cho trước: 5'- AATTCTTAAATTC -3'
Sợi bổ sung: 3'- TTAAGAATTTAAG -5'



Hình 3.4 Hai kiểu kết cặp base của DNA. Cặp AT nối với nhau bằng hai liên kết hydro và cặp GC - ba liên kết hydro (biểu thị bằng các đường chấm: ---). Các nguyên tử $C_{1'}$ đại diện cho vị trí của đường và phosphate ở mỗi cặp nucleotide.

Tóm lại, hai đặc điểm quan trọng nhất trong cấu trúc DNA là sự *phân cực ngược chiều* của hai sợi đơn ($5' \rightarrow 3'$ và $3' \rightarrow 5'$) và *nguyên tắc bổ sung* của các cặp base (A-T và G-C). Đây là hai nguyên lý căn bản chi phối các cơ chế di truyền ở cấp độ phân tử (tái bản, phiên mã và dịch mã), mà ta có thể hình dung tổng quát dưới dạng các kênh truyền thông tin di truyền trong tế bào (được gọi là *Giáo lý* hay *Lý thuyết trung tâm*, Central Dogma, của Sinh học phân tử; Hình 3.5) sau đây:

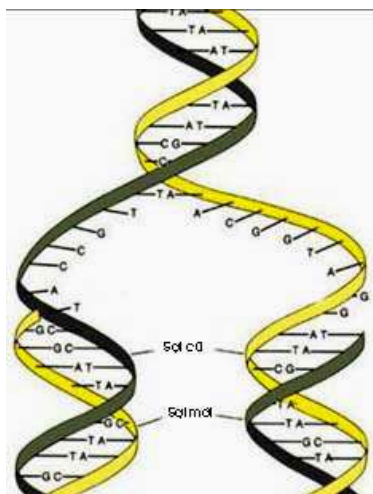


Hình 3.5 Lý thuyết trung tâm của Sinh học phân tử

♦ Về tầm vóc vĩ đại của phát minh cấu trúc phân tử DNA, Lawrence Bragg - Giám đốc Phòng thí nghiệm Cavendish (England) - đánh giá rằng: "Sự phát minh ra cấu trúc DNA với tất cả các hệ quả sinh học của nó là một trong các sự kiện khoa học to lớn nhất của thế kỷ chúng ta..." (Watson 1968, bản Việt dịch của Lê Đình Lương và Thái Doãn Tĩnh, Nxb KH-KT tr.9). Nhờ phát minh vĩ đại đó, Watson và Crick cùng chia sẻ với Wilkins giải thưởng Nobel năm 1962.

Thật vậy, nhìn lại ta thấy rằng Watson và Crick đã công bố phác thảo về mô hình cấu trúc DNA trong bài báo nhan đề "A Structure for Deoxyribose Nucleic Acid" trên tạp chí *Nature* Vol. 171, trang 737 ngày 25-4-1953 (<http://www.nature.com/>). Đây là một bài báo khoa học kinh điển rất ấn tượng và không bình thường tý nào! Một cách chính xác, bài báo này chỉ dài 900 chữ với vốn vẹn 128 dòng, nhưng đằng sau mỗi dòng là cả một lịch sử khoa học kỳ diệu, một câu chuyện thú vị. Bài báo này được công bố rất nhanh, chưa đầy một tháng kể từ sau ngày gửi đăng.

Trên thực tế, Crick muốn làm sáng tỏ các hàm ý sinh học của mô hình này, nhưng Watson thì chẳng hài lòng với cách làm như vậy. Hai ông đã thoả thuận trong một câu mà nó đã trở thành một trong những câu nói giản lược vĩ đại trong tài liệu khoa học: "Chúng ta không thể không nhận thấy rằng nguyên tắc kết cặp base đặc thù mà chúng tôi nêu lên ngay lập tức gợi ra một cơ chế sao chép khả dĩ cho vật chất (di truyền) nói chung". ["It has not escaped our noticed that the specific base pairing we have proposed immediately suggests a possible copying mechanism for the general (genetic) material."].



Hình 3.6 Mô hình tái bản của DNA do Watson gợi ý từ 1953.

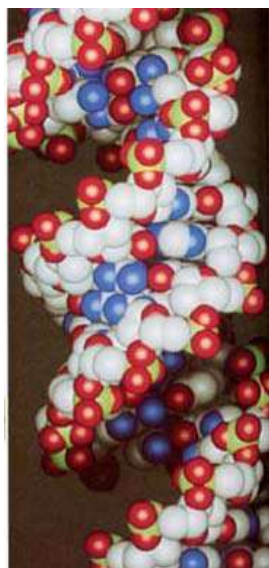
Như câu nói đầy khêu gợi này đã chỉ rõ, mô hình của Watson và Crick quả thực gợi ra một cơ chế sao chép cho DNA. Vì một sợi là *bổ sung* (complement) của sợi kia, nên hai sợi có thể được tách ra và mỗi sợi sau đó có thể dùng làm khuôn (template) cho việc xây dựng nên một sợi mới cặp với nó (Hình 3.6). Bằng cơ chế *tái bản bán bảo toàn* (semiconservative replication) như thế sẽ đảm bảo được hai phân tử DNA con tạo ra có cấu trúc giống hệt DNA cha mẹ, và qua đó các tế bào sinh ra sẽ chứa các gene giống nhau; nghĩa là có thể giải thích được tại sao con cái thường giống cha mẹ. Quả thực đây là sự tiên đoán thiên tài mà sự đúng đắn của nó đã được chứng minh bằng các thực nghiệm khác nhau chỉ sau đó vài năm (xem chương 5).

Đến đây hẳn là chúng ta có thể hiểu tại sao Watson lại đưa ra được những nhận định sắc sảo tuyệt vời và cực kỳ chính xác đến như vậy, chẳng hạn: "*Một cấu trúc tuyệt đẹp như thế, lẽ tự nhiên là phải tồn tại trên thực tế*", hay "*Cách giải quyết đúng không những phải đẹp mà còn phải đơn giản*". Bạn có thể làm sáng tỏ điều này? Và bạn có thể học được gì từ bài báo kinh điển của Watson và Crick với nhan đề "A Structure for Deoxyribose Nucleic Acid" trên tạp chí *Nature* ngày 25-4-1953 (☞ <http://www.nature.com/>) và từ Lời bình độc đáo của Tom Zinnen (2004) về bài báo khoa học này (☞ <http://www.accessexcellence.com/>)?

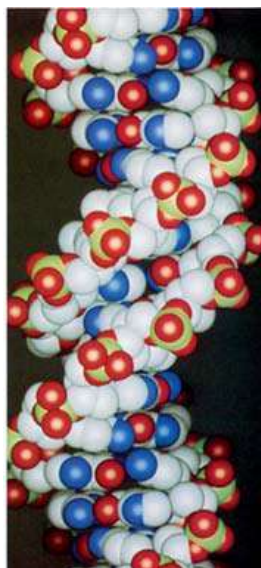
2. Các dạng DNA xoắn phải và xoắn trái

Mô hình Watson-Crick hay *DNA dạng B* là cấu trúc phổ biến.

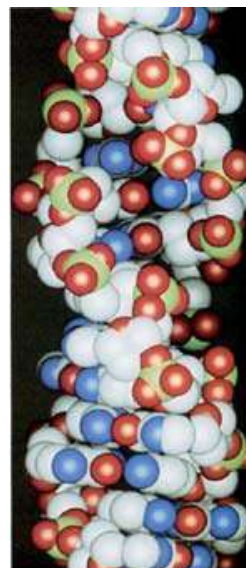
Tuy nhiên, sau này người ta còn phát hiện ra nhiều dạng xoắn phải khác (A, C, D...); chúng có một số biến đổi so với DNA-B (xem Bảng 3.2).



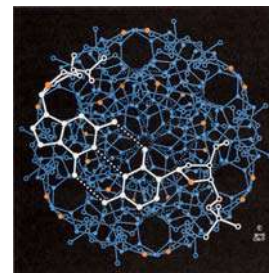
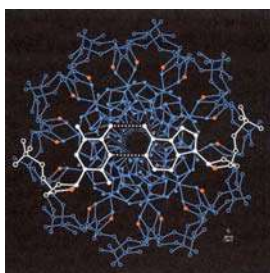
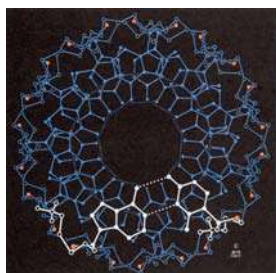
DNA dạng A



DNA dạng B



DNA dạng Z



Hình 3.7 Các mô hình DNA dạng A, B và Z (hình trên) và thiết diện cắt ngang của chúng cho thấy vị trí phân bố của một cặp base.

Bên cạnh các dạng DNA xoắn phải, Alexander Rich và đồng sự (1979) còn phát hiện thêm một dạng DNA xoắn trái duy nhất cho đến nay. Dạng DNA này có bộ khung hình zigzag (nên gọi là DNA-Z, và cũng là chữ cái cuối cùng trong bảng chữ cái Latin) uốn gập khúc theo chiều xoắn trái, mỗi vòng xoắn dài $45,6\text{\AA}$ chứa 12 cặp base. Nhìn chung, so với DNA dạng B, DNA-Z dài và gầy hơn, các rãnh lớn bị dẹt ra phần bề mặt của chuỗi xoắn; còn DNA dạng A ngắn và to mập hơn (Hình 3.7 và Bảng 3.2).

Những vùng nào của DNA có chứa các purine và pyrimidine

sắp xếp xen kẽ nhau trên một sợi thì có thể tiếp nhận cấu hình DNA-Z, ví dụ:

5'--GCGCGCGC--3'

3'--CGCGCGCG--5'

Sự chuyển đổi này cũng được tạo thuận lợi bởi sự có mặt của 5-methylcytosine và bởi trạng thái siêu xoắn nghịch (negative supercoiling). DNA là một phân tử động học và vì vậy nó có thể chuyển từ một cấu hình này sang một cấu hình khác dựa trên các lực bên ngoài trong tế bào. Có thể là sự chuyển đổi từ dạng B sang dạng Z có liên quan đến sự điều hoà biểu hiện gene. Mặc dù Rich khám phá DNA-Z khi nghiên cứu về các hợp chất mô hình, cấu trúc này dường như cũng có mặt trong các tế bào sống ở một tỷ lệ nhỏ song chức năng của nó vẫn còn chưa thật sự hiểu rõ.

Bảng 3.2 Một số đặc điểm chính của các DNA dạng A, B, C và Z

Dạng	Chiều xoắn	Số bp/vòng xoắn	Đường kính chuỗi xoắn
A	Phải	11,0	23A°
B	Phải	10,0	19A°
C	Phải	9,3	19A°
Z	Trái	12,0	18A°

**Nguồn*: J.Kimball (từ internet). Về chi tiết, có thể xem trong Watson *et al* (1987, tr.249) và Twyman (1998; tr.229).

3. Các DNA mạch vòng sợi kép và sợi đơn

Kể từ sau khám phá quan trọng của Watson và Crick, cho đến nay không những đã phát hiện thêm các dạng DNA xoắn phải và xoắn trái, mà trên thực tế còn có các bộ gene được tổ chức theo những thể thức khác, đó là: *DNA sợi kép dạng vòng* có mặt ở hầu hết các bộ gene prokaryote, bộ gene một số virus và bộ gene tế bào chất của các tế bào eukaryote (các phân tử DNA ty thể và lạp thể); *DNA sợi đơn vòng* của một số virus ký sinh ở vi khuẩn; và *bộ gene RNA* của nhiều virus ký sinh ở các thực vật và động vật. Đáng kể là các virus RNA gây ung thư, HIV/AIDS và các virus thuộc họ corona gây viêm phổi cấp (SARS) với nhiều biến thể có khả năng lây lan sang nhiều vật chủ khác nhau và có nguy cơ làm xuất hiện nạn đại dịch trên phạm vi toàn cầu hiện nay. Vấn đề này sẽ được đề cập thêm ở mục III dưới đây và ở Bảng 3.3.

★ *Thảo luận thêm về các bậc cấu trúc của các nucleic acid:*

❶ *Cấu trúc bậc một* của nucleic acid là các chuỗi polynucleotide;

❷ *Cấu trúc bậc hai* của các nucleic acid được sinh ra bởi hai loại tương tác không phải đồng hoá trị: *sự kết cặp base* (base pairing) và *sự co cụm base* (base stacking). Sự kết cặp base liên quan với các liên kết hydro và là lực chiếm ưu thế khiến cho các sợi nucleic acid kết hợp với nhau, nhưng cấu trúc được giữ ổn định bằng các tương tác hydrophobic giữa các base kề sát nhau mang lại bằng các điện tử pi (π) trong các vòng. Các mối *tương tác π - π* này được mô tả như là các lực kéo co cụm base. Cấu trúc bậc hai của DNA được đặc trưng bằng sự kết cặp base giữa các phân tử để sinh ra các phân tử sợi kép hay sợi đôi (double-stranded or duplex; ký hiệu là *dsDNA*). Các cấu trúc bậc hai trong RNA, vốn tồn tại nguyên thủy ở dạng sợi đơn (single-stranded form), nói chung phản ánh các mối tương tác base nội trong phân tử.

- Trong cấu trúc của các chuỗi xoắn kép DNA, quan trọng nhất là sự *kết cặp base bổ sung* (complementary base pairing) A-T và G-C. Các *cặp base Watson-Crick* này (Watson-Crick base pairs) tạo thành cơ sở của hầu hết các tương tác cấu trúc bậc hai trong các nucleic acid, cũng như giải thích cho các quy tắc Chargaff, và chúng đồng thời xác định cách thức DNA có thể hoạt động như là cái khuôn cho tái bản và phiên mã... Trong RNA, uracil thay thế cho thymine, nhưng vì uracil có cấu trúc hoá học tương tự với thymine và hình thành các liên kết hydro với adenine y như thế, cả hai nucleic acid lại theo cùng các quy tắc chung. Tuy nhiên, vì các mối tương tác này có mặt khắp nơi, nên có những sơ đồ kết cặp base biến đổi đôi chút so với các kiểu kết cặp Watson-Crick; chúng đóng các vai trò quan trọng trong việc hình thành các cấu trúc bậc hai và bậc ba.

- Cho đến nay, bên cạnh các cặp base Watson-Crick chiếm ưu thế trong các cấu trúc và chức năng của các nucleic acid, người ta thấy có 28 cách sắp xếp khả dĩ của ít nhất hai liên kết hydro giữa các base; điều này cung cấp cơ sở cho một nhóm đa dạng các tương tác. Có ý nghĩa đáng kể nhất trong số các cấu hình biến đổi này là các *cặp base Hoogsteen* (Hoogsteen base pairs), vốn góp phần vào cấu trúc bậc ba của tRNA và cho phép hình thành các chuỗi xoắn ba (triple helices). Một sự sửa đổi so với các cặp base Watson-Crick là các *cặp linh hoạt* (wobble pairs). Vấn đề này sẽ

được xem xét trở lại ở các chương 4 (cấu trúc và chức năng của *tRNA*) và 6 (mục III: *mã di truyền, giả thuyết linh hoạt*).

- Về các *cấu hình chuỗi xoắn* và *tính mềm dẻo cục bộ trong cấu trúc DNA*: Bên cạnh cấu trúc DNA sợi kép (dsDNA) dạng B phổ biến do Watson và Crick đưa ra năm 1953, còn có các dạng biến đổi khác như đã đề cập ở trên. Một đặc điểm khác nữa đó là *tính mềm dẻo cục bộ* (local flexibility) trong cấu trúc DNA. Nhiều thực nghiệm đã cho thấy rằng DNA dạng B đặc biệt mềm dẻo linh hoạt, nó không tồn tại ở các dạng có cấu hình cứng nhắc mà có thể thay đổi một cách uyển chuyển giữa các cấu hình khác nhau do các hiện tượng đa hình cục bộ gây ra, chẳng hạn như DNA có thể uốn gập và *hoán chuyển chuỗi xoắn* (helical transitions) nội trong một phân tử đơn (ví dụ sự hoán chuyển qua lại giữa các dạng B và Z đã nói ở trên). DNA cuộn gập cũng có thể được cảm ứng bởi các protein và tạo vòng (circularization). Việc cuộn lại do cảm ứng cần thiết cho sự đóng gói DNA trong các nhiễm sắc thể (Hình 3.10) và cho tái bản, tái tổ hợp và phiên mã (chương 5 và 6). Các protein cũng có thể nhận biết DNA được cuộn lại theo thể thức nào đó (ví dụ các *topoisomerase* nhận biết các khởi điểm tái bản).

- *Cấu trúc bậc hai trong RNA và DNA không thuộc dạng sợi đôi*: Trong RNA và các vùng DNA sợi đơn, cấu trúc bậc hai được xác định bằng sự kết cặp base nội phân tử. Cấu trúc bậc hai trong RNA đóng vai trò chính yếu trong biểu hiện gene và điều hoà của nó. Ví dụ: sự kết cặp base giữa rRNA và mRNA kiểm soát việc khởi đầu tổng hợp protein; sự kết cặp base giữa tRNA và mRNA xảy ra trong dịch mã; các cấu trúc *kẹp tóc trong RNA* (RNA hairpin loop) và các *vòng thân* (stem loops) kiểm soát sự kết thúc phiên mã, hiệu quả dịch mã và sự ổn định của mRNA; và sự kết cặp base RNA-RNA cũng đóng vai trò quan trọng trong việc tách bỏ các intron (xem các chương 4 và 6).

③ *Cấu trúc bậc ba* của các nucleic acid phản ánh các mối tương tác góp phần kiến thiết toàn bộ hình dáng ba chiều của DNA và RNA. Điều này bao gồm các tương tác giữa các yếu tố cấu trúc bậc hai khác nhau, các mối tương tác giữa các sợi đơn và các yếu tố cấu trúc bậc hai, và các đặc điểm hình thể của các nucleic acid.

- *Các tương tác sợi bậc ba trong DNA*: Trong DNA, các mối tương tác bậc ba có liên quan tới sự tương tác giữa các sợi đơn với các sợi đôi hoặc tương tác giữa các sợi đôi với các sợi đôi, kết quả là tạo thành các cấu trúc bộ ba và bốn sợi (triple and

quadruple strand structures). Các guanine có thể hình thành các *bộ bốn base* (base tetrads), và các DNA chứa các loạt gốc guanine có thể tạo thành các cấu trúc bộ bốn mà từ đó có thể góp phần vào cấu trúc *telomere* (chương 5). Ví dụ, *DNA dạng H* là một dạng của DNA bộ ba sợi nội phân tử xuất hiện trong các đoạn bất cặp homopurine/homopyrimidine và có liên quan các cặp base Hoogsteen. Bây giờ ta hãy hình dung các cấu trúc bậc ba được gọi là các *vòng R* (R-loop) tạo thành khi RNA được phiên mã từ DNA sợi kép được cố định tại chỗ (*in situ*), xảy ra chẳng hạn trong khi môi hoá cho tái bản ở plasmid ColE1. Các cấu trúc bậc ba bốn sợi, các chỗ nối trong mô hình Holliday, cũng hình thành trong khi tái tổ hợp (có thể xem *tái tổ hợp tương đồng*).

- *Các tương tác sợi bậc ba trong RNA*: Việc RNA cuộn lại thành các cấu trúc phức tạp có dính dáng tới các tương tác bậc ba giữa các sợi, các vòng (loops) và các sợi đôi. Ví dụ, trong tRNA có các ví dụ về các bộ ba base, các đoạn của chuỗi xoắn ba, các *chỗ nối phần thân* (stem junctions; trong đó hai hoặc nhiều vùng sợi đôi được nối với nhau) và các *mấu giả* (pseudo-knots; tại đó các sợi tương tác với các vòng thân - stem loop).

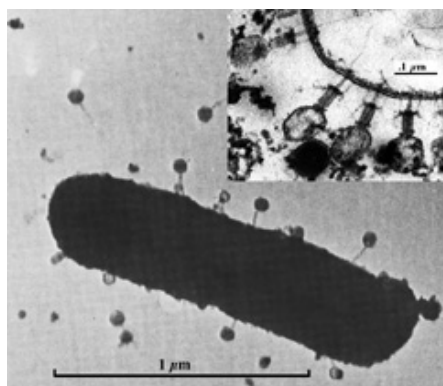
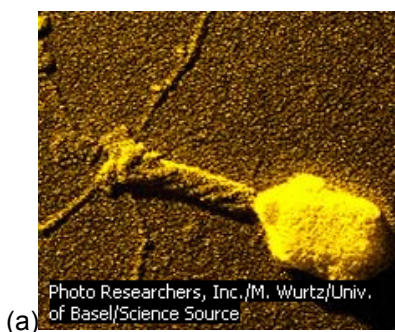
- Đối với các đặc điểm cấu trúc hình học của DNA, nếu như các phân tử DNA có các đầu mút tự do (ví dụ một phân tử mạch thẳng) thì hai sợi mở xoắn quanh nhau theo cách tiện ích nhất về mặt năng lượng và phân tử đó được coi là được *giãn xoắn* (relaxed). Tuy nhiên, trong các DNA mạch vòng, không có các đầu mút tự do và nó chỉ được biến đổi bằng cách cắt mở vòng, chứ không phải bằng cách làm biến dạng nó. Nếu như DNA ở dạng vòng khép kín tiến hành tháo xoắn thì cách duy nhất để làm giãn xoắn kiểu vận xoắn như vậy được tạo ra thông qua sự *siêu xoắn* (supercoiling), tại đó một sự vận xoắn được đưa vào trong chính trục chuỗi xoắn. Trạng thái siêu xoắn là một dạng khác nữa của cấu trúc bậc ba của nucleic acid. Ý nghĩa sinh lý học của sự siêu xoắn là ở chỗ DNA không bị bó chặt thì thường không có hoạt tính sinh học. Trạng thái siêu xoắn nghịch tỏ ra cần thiết cho nhiều quá trình thiết yếu, như: tái bản, phiên mã và kể cả tái tổ hợp, DNA siêu xoắn lưu giữ năng lượng để điều khiển các phản ứng này. Ở các eukaryote vốn chứa các nhiễm sắc thể mạch thẳng, các vùng bó chặt về mặt không gian được bắt đầu bằng cách tổ chức chromatin thành các vòng với các đầu mút được cố định bởi các protein chống đỡ; các nucleosome đưa các cuộn siêu xoắn nghịch vào trong DNA eukaryote (xem Hình 3.10).

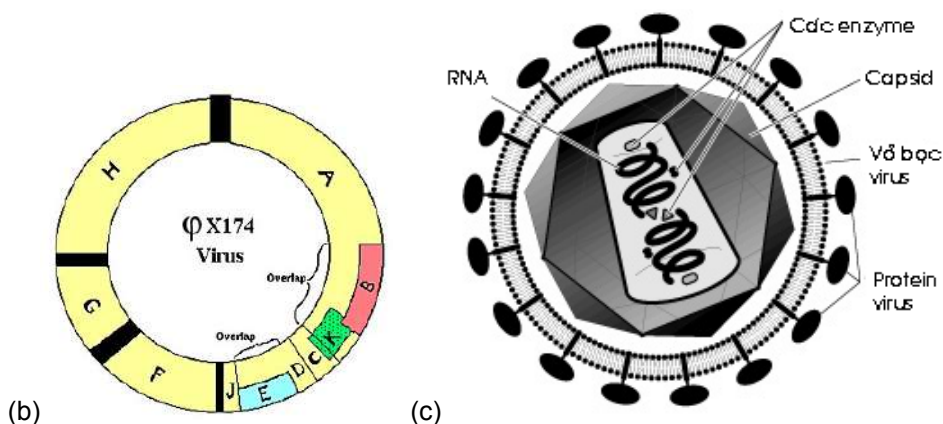
④ **Cấu trúc bậc bốn** của các nucleic acid: Trong nhiều cấu trúc, các nucleic acid tương tác ở cấu hình *trans* (ví dụ, ribosome và spliceosome), và đây có thể xem là bậc bốn của cấu trúc nucleic acid. Các nucleic acid cũng tương tác với một số lượng lớn các protein (ví dụ, các protein cấu trúc bộ gene, các yếu tố phiên mã, các enzyme, các nhân tố splicing). Khá nhiều các protein này gây một tác dụng đáng kể lên cấu hình DNA và RNA. Ví dụ enzyme cắt giới hạn EcoRI có thể bám vào đoạn nhận biết trong DNA và từ đó phát huy hoạt tính cắt bên trong sợi...

III. Định khu, hàm lượng và kích thước của DNA

1. Sơ lược về bộ gene của các virus, prokaryote và eukaryote

Virus là nhóm "sinh vật" bé nhất chưa có cấu tạo tế bào, không tồn tại đơn độc mà ký sinh bắt buộc ở các tế bào *sinh vật tiền nhân* (prokaryote) hoặc *sinh vật nhân chuẩn* (eukaryote); chỉ trong điều kiện đó chúng mới có khả năng tái bản. Các virus ký sinh hoặc gây nhiễm vi khuẩn gọi là *thể thực khuẩn* (bacteriophage) hay *phage*. Cấu trúc của các virus tương đối đơn giản, gồm hai phần chính là lõi acid nucleic và vỏ protein (Hình 3.8). Một số virus ở thực vật hoặc các virus gây ung thư, AIDS, SARS ở người và động vật có bộ gene là RNA. Số còn lại bao gồm nhiều virus ký sinh ở vi khuẩn và động vật có bộ gene là DNA sợi kép hoặc sợi đơn, có mạch thẳng hoặc mạch vòng (Bảng 3.3).





Hình 3.8 (a) Vi ảnh điện tử của phage T4 và sự bám của phage trên màng tế bào *E. coli* (từ trái sang). (b) Bản đồ DNA sợi đơn vòng của phage ΦX174 với một số gene gối nhau. (c) Sơ đồ cấu trúc của HIV - một retrovirus.

Nhóm prokaryote bao gồm các *vi khuẩn* (bacteria) và *vi khuẩn cổ* (archae) là các sinh vật có cấu tạo tế bào đơn giản nhất. Vi khuẩn *Escherichia coli* (hình 3.9) là đối tượng được sử dụng rộng rãi trong các nghiên cứu di truyền phân tử. Bộ gene chính của nó là một phân tử DNA sợi kép vòng có kích thước lớn (4.639.221 cặp base, với 4290 gene mã hóa protein và 53 gene RNA) gọi là nhiễm sắc thể chính. Nó thường tập trung ở một "*vùng nhân*" (nucleoid), không có màng nhân bao bọc, và ở trạng thái *siêu xoắn* (supercoiled DNA) dưới sự kiểm soát của các enzyme *topoisomerase*. Ngoài ra còn có nhiều phân tử DNA sợi kép trần mạch vòng khác có kích thước bé hơn nhiều, gọi là các *plasmid*.

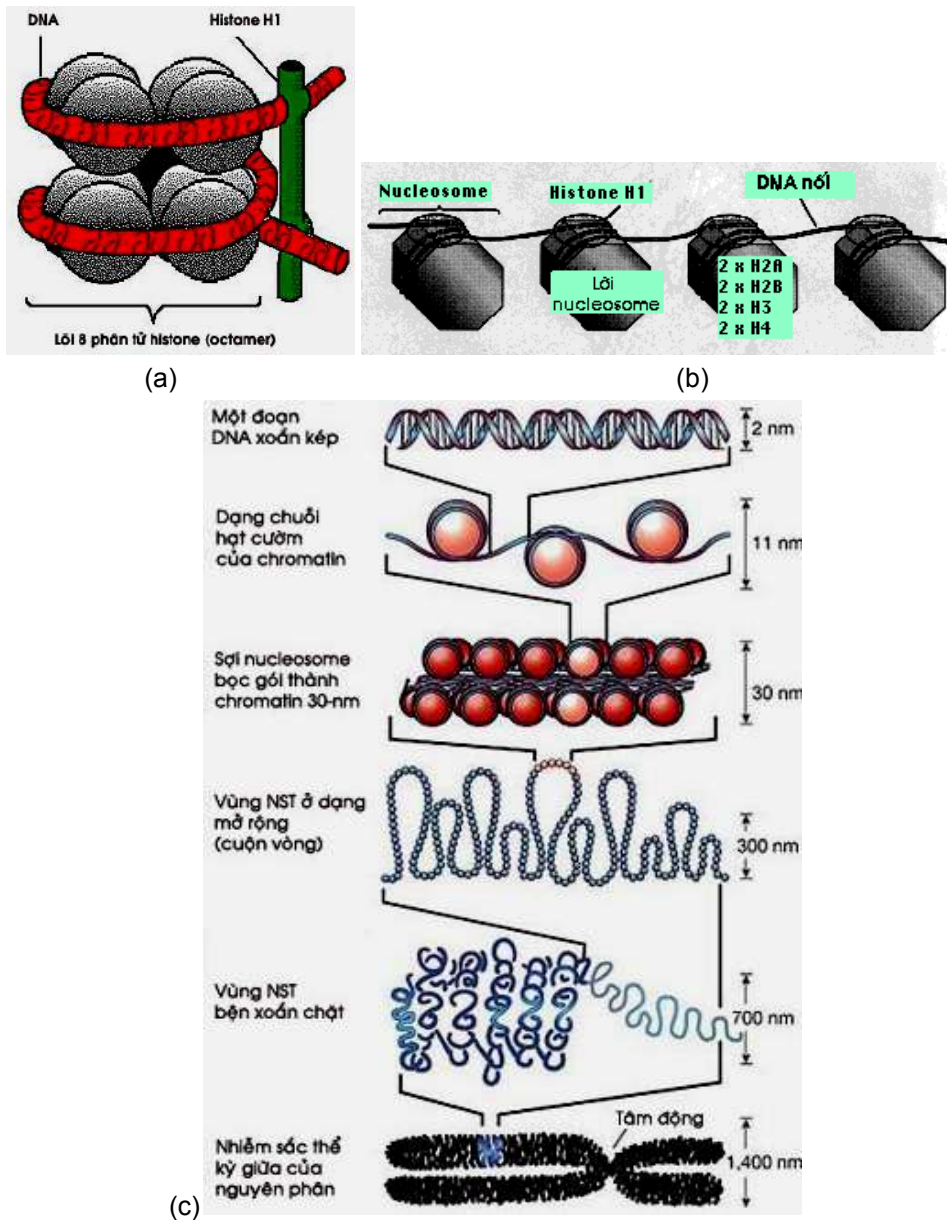
Lưu ý: Từ đầu thập niên 1990 đến nay, người ta phát hiện thấy rằng tổ chức bộ gene DNA sợi kép của nhiều vi khuẩn không chỉ có một phân tử mạch vòng (như ở *Bacillus*, *E. coli*, *Pseudomonas*, v.v.) mà còn có thể có các trường hợp sau: 1 DNA mạch thẳng (*Borella* = 0,91 Mbp); 2 DNA mạch vòng (ví dụ: *V. cholera* = 2,9 + 1,1 Mbp); hoặc 3 DNA vòng (ví dụ: *Paracoccus denitrificans* = 2,0 + 1,1 + 0,64 Mbp); hoặc thậm chí bộ gene của *Agrobacterium tumefaciens* gồm một DNA mạch thẳng (2,1 Mbp) và một DNA mạch vòng (3,0 Mbp). Về phần các plasmid cũng vậy, ví dụ ở chi *Borella* có rất nhiều plasmid vòng và thẳng với kích thước biến thiên từ 5 đến 200 Kbp. (1 Mbp = 10^3 Kbp = 10^6 bp).

[Về chi tiết, có thể xem trong chương 2 của *Giáo trình Di truyền Vi sinh vật* do Hoàng Trọng Phán chủ biên; Nxb Đại Học Huế - 2006.]



Hình 3.9 (a) Ảnh chụp các tế bào *E. coli*. (b) Sơ đồ tổ chức vật chất di truyền của tế bào *E. coli* cùng với vi ảnh điện tử của plasmid pSC101 (hình dưới).

Nhóm eukaryote là nhóm lớn nhất và tiến hóa đa dạng nhất về trình độ tổ chức cơ thể, bao gồm tất cả các sinh vật có cấu tạo tế bào (trừ vi khuẩn và vi khuẩn lam), có thể là đơn bào hoặc đa bào. Trong tế bào chứa hai hệ thống di truyền, bộ gene nhân và bộ gene tế bào chất. Bộ gene tế bào chất bao gồm các phân tử DNA sợi kép vòng, đó là: các DNA ty thể (mitochondrial DNA = mtDNA) có mặt trong tất cả các tế bào eukaryote, còn DNA lục thể (chloroplast DNA = cpDNA hay ctDNA) chỉ có trong các tế bào thực vật (xem mục 4). Trong nhân tế bào eukaryote chứa nhiều nhiễm sắc thể; mỗi nhiễm sắc thể là một phức hợp *nucleoprotein* (còn gọi là *chất nhiễm sắc*, chromatin), gồm một phân tử DNA mạch kép thẳng kết hợp với các phân tử protein cơ sở có tên là các histone (giàu lysine và arginine). Đơn vị tổ chức của cơ sở của nhiễm sắc thể eukaryote là các *nucleosome* (hình 3.10a) có đường kính khoảng 11 nm, gồm một khối cầu tám phân tử histone, $(H2A + H2B + H3 + H4)_2$, gọi là *lõi octamer* và đoạn DNA có kích thước 146 cặp base quấn xung quanh nó $1\frac{3}{4}$ vòng (thường được mô tả là 160-200 bp quấn hai vòng quanh octamer). Một phân tử H1 bám vào các vùng *DNA nối* (linker DNA) bên ngoài nucleosome, giữ vững sự tương tác của DNA với các histone lõi. Các mức độ tổ chức hay sự hoá xoắn của nhiễm sắc thể eukaryote được mô tả ở hình 3.10b.



Hình 3.10 Tổ chức DNA trong nhiễm sắc thể eukaryote. (a-b) Cấu trúc một nucleosome và chuỗi nucleosome; (c) Các mức độ tổ chức của vật chất di truyền ở tế bào eukaryote.

2. Mối quan hệ giữa kích thước bộ gene và tính phức tạp về mặt tiến hóa

Dựa trên các kết quả phân tích bộ gene của các virus và vi

khuẩn cũng như của bộ gene ở eukaryote (bộ nhiễm sắc thể đơn bội), cho phép khái quát như sau: *Kích thước của bộ gene tăng lên tương đối cùng với mức độ phức tạp về tiến hóa.*

Thật vậy, từ bảng 3.3 cho thấy rằng kích thước bộ gene của các virus nói chung là rất nhỏ so với nhóm prokaryote; và kích thước bộ gene của các prokaryote lại tỏ ra quá đơn giản so với ngay cả một eukaryote đơn bào như nấm men.

Trong khi DNA của một vi khuẩn điển hình như *E. coli* hơn 4,6 triệu cặp base và chứa 4.377 gene mã hóa protein (không kể 53 gene RNA), thì phage MS2 là một trong các virus bé nhất, bộ gene RNA của nó cũng chỉ có 3.569 base với tất cả 4 gene; hoặc có kích thước lớn như virus Epstein-Barr (bộ gene DNA sợi kép mạch vòng) cũng chỉ có 172.282 cặp base với tất cả 80 gene.

Nếu xét trên cả hai nhóm prokaryote và eukaryote, ta thấy rằng kích thước các bộ gene biến thiên rất rộng: bộ gene đối với một sinh vật sống tự do (một vi khuẩn) được biết là bé nhất chứa khoảng 600.000 cặp base DNA, trong khi các bộ gen người và chuột là khoảng 3 tỷ. Nếu xét riêng ở nhóm eukaryote, ta đã biết mỗi loài có một số lượng nhiễm sắc thể đặc trưng và nó không phản ánh trình độ tiến hóa của các loài. Tuy nhiên, về mặt nào đó rõ ràng là có sự tương quan thuận giữa hàm lượng DNA của các bộ gene đơn bội và bậc thang tiến hóa của các động-thực vật. Dù vậy vẫn có một số ngoại lệ so với quy tắc này!

3. Hàm lượng DNA và nghịch lý giá trị C

Chẳng hạn, mặc dù *Psilotum nudum*, đôi khi gọi là "dương xỉ lông", là một thực vật đơn giản hơn nhiều so với loài *Arabidopsis thaliana* vốn là một thực vật có hoa thuộc họ cải, nhưng nó lại có kích thước bộ gene lớn hơn tới 3.000 lần. Đó là do trên 80% bộ gene của nó là DNA lặp lại không chứa thông tin di truyền nào cả! Hay một số thực vật (như ngô, loa kèn) hay một số động vật (như lưỡng thê, cá) lại có kích thước bộ gene lớn gấp nhiều lần so với lớp thú (bảng 3.3). Một số lưỡng thê chứa DNA nhiều gấp bộ gene chúng ta đến 30 lần, nhưng chắc chắn không phải là chúng phức tạp gấp chúng ta 30 lần.

Người ta gọi tổng hàm lượng DNA trong bộ gene đơn bội là *giá trị C* (C-value). Với phân tích ở trên cho thấy không hề tồn tại một mối quan hệ kiên định nhất quán giữa giá trị C và tính phức tạp của một sinh vật (như lưỡng thê với thú); cái đó gọi là *nghịch lý giá trị C*

(C-value paradox).

Bảng 3.3 Kích thước bộ gene của một số sinh vật thường gặp

Bộ gene sinh vật	Số bp	Số gene	Ghi chú
Virus			
Phage Ø-X174 (ở <i>E. coli</i>)	5.386	10	DNA sợi đơn vòng
Phage lambda (ở <i>E. coli</i>)	48.502	~61	DNA sợi kép thẳng
Phage T2 hoặc T4 (ở <i>E. coli</i>)	~2x10 ⁵	150-200	DNA sợi kép thẳng
Phage MS2 (ở <i>E. coli</i>)	3.569	4	RNA
SV40 (gây khối u ở khỉ)	5.226		DNA sợi kép vòng
Virus SARS (ví dụ, H5N1)	29.600		RNA; viêm phổi cấp
Epstein-Barr virus (EBV)	172.282	80	DNA sợi kép thẳng
Prokaryote			
<i>Haemophilus influenzae</i>	1.830.138	1.738	Gây nhiễm tai giữa
<i>Chlamydia trachomatis</i>	1.042.519	936	Bệnh lây qua tình dục
<i>Streptococcus pneumoniae</i>	2.160.837	2.236	<i>Pneumococcus</i> ở 2 NST; gây
<i>Vibrio cholerae</i>	4.033.460	3.890	cholera
<i>Mycobacterium tuberculosis</i>	4.411.532	3.959	Gây bệnh lao
<i>Bacillus subtilis</i>	4.214.814	4.779	
<i>Escherichia coli</i> ⁽¹⁾	4.639.221	4.377	Có 4290 cistron
<i>Agrobacterium tumefaciens</i> ⁽²⁾	4.674.062	5.419	Vector hữu ích...
<i>E. coli</i> O157:H7	5.44 x 10 ⁶	5.416	Gây bệnh ở người
Eukaryote			
	12.495.68		
<i>Saccharomyces cerevisiae</i>	2	5.770	Men bia nảy chồi
<i>Schizosaccharomyces pombe</i>	12.462.63		Nấm men phân cắt
	7	4.929	
<i>Plasmodium falciparum</i> ⁽³⁾	22.853.76		Gây sốt rét nguy
	4	5.268	nhất
	38.639.76		
<i>Neurospora crassa</i>	9	10.082	+ 498 gene RNA
	100.258.1	~19.00	
<i>Caenorhabditis elegans</i> ⁽⁴⁾	71	0	Giải tr.tự đầu tiên
<i>Arabidopsis thaliana</i>	115.409.9	25.498	Thực vật có hoa

	49		
	122.653.9		
<i>Drosophila melanogaster</i>	77	13.379	Ruồi giấm
		~25.00	
Người (<i>Homo sapiens</i>) ⁽⁵⁾	~3,2 x 10 ⁹	0	Hoàn tất 4/2003
		~60.00	
Lúa (<i>Oryza sativa</i>)	4,3 x 10 ⁸	0	
Chuột (<i>Mus musculus</i>)	2,2 x 10 ⁹	?	
Lưỡng thê (<i>Xenopus laevis</i>)	10 ⁹ - 10 ¹¹	?	

Chú thích: ⁽¹⁾ Có 4290 gene mã hóa protein, còn lại là RNA; ⁽²⁾ Vector hữu ích để chuyển gen ở thực vật; ⁽³⁾ Cộng với 53 gene RNA; ⁽⁴⁾ Eukaryote đa bào đầu tiên được xác định trình tự; ⁽⁵⁾ Được xác định đầy đủ trình tự vào 4/2003 bởi HGP, với tổng cộng 3.164.700.000 cặp base; và theo ước tính mới nhất công bố ngày 21/10/2004 trên tạp chí *Nature*, bộ gene người chúng ta chứa khoảng 20.000-25.000 gene mã hóa protein chiếm khoảng 2% toàn bộ bộ gene.

4. Kích thước DNA một số bào quan

Kích thước DNA của một số bào quan (bảng 3.4) cho thấy chúng có vẻ đơn giản và không có dấu hiệu tiến hóa rõ rệt. Nói chung, kích thước mỗi phân tử DNA ty thể của người và các động vật có vú thường nằm trong khoảng 15.000-17.000 bp; ví dụ mtDNA người là 16.569 bp. Trong khi đó, kích thước một DNA lặp thể ở phần lớn tế bào các thực vật thường biến thiên trong khoảng 130.000 - 150.000 bp. Chẳng hạn, cpDNA ở lúa tròng (*O. sativa*) thuộc hai nhóm indica và japonica có kích thước tương ứng là 134.494 bp và 135.525 bp, ở lúa mì (*Triticum aestivum*) là 134.545 bp và ở ngô (*Zea mays*) là 140.384 bp, v.v. Còn các plasmid của một số tế bào thực vật thường có kích thước rất bé khoảng 1-2 ngàn cặp base.

Bảng 3.4 Kích thước DNA bào quan ở một số sinh vật nhân chuẩn

DNA ty thể	Số cặp base	Plasmid	Số cặp base
Người (<i>Homo sapiens</i>)	16.569	<i>O. sativa</i> (<i>indica</i>)	1.485
<i>D. melanogaster</i>	19.517	<i>O. sativa</i> (<i>japonica</i>)	2.135
<i>S. cerevisiae</i>	85.779	Ngô (<i>Zea mays</i>)	1.913

DNA lặp thể	Số cặp base		
Lúa <i>O. sativa (indica)</i>	134.494	<i>Brassica</i>	11.640
<i>O. sativa (japonica)</i>	134.525	<i>N. crassa</i> (3 loại)	3.581
Ngô (<i>Zea mays</i>)	140.384		3.675
Mía (S. <i>officinarum</i>)	141.182		7.050

IV. Đặc tính hóa lý của DNA

1. Biến tính và hồi tính của DNA

Một trong những đặc điểm quan trọng nhất của DNA là hai mạch đơn bổ sung của nó gắn với nhau bằng các mối liên kết hydro, vốn là lực liên kết hóa học yếu nên chúng có thể bị phân hủy dưới tác dụng của các enzyme, năng lượng... làm cho hai mạch đơn của chuỗi xoắn kép tách rời nhau, gọi là *biến tính* (denaturation). Nhờ đó DNA mới có thể tái bản, các gene mới có thể biểu hiện ra các sản phẩm của mình. Mặt khác, DNA có thể phục hồi trở lại trạng thái ban đầu theo một quá trình ngược lại, gọi là *hồi tính* (renaturation).

Bằng thực nghiệm, người ta đã chứng minh điều đó bằng cách sử dụng các tác nhân vật lý và hóa học khác nhau. Chẳng hạn, khi đun nóng từ từ các phân tử DNA lên tới nhiệt độ gần 100°C (thường là 90-95°C), thì các liên kết hydro của chúng bị phá hủy hoàn toàn và hai sợi bổ sung tách ra. Ngược lại, khi làm nguội từ từ dung dịch đột nóng chứa DNA bị biến tính hoàn toàn, các sợi đơn thường cặp lại với sợi bổ sung của chúng và làm phục hồi chuỗi xoắn kép như lúc đầu. Rõ ràng đây là các quá trình có tính thuận-ngịch.

1.1. Biến tính hay sự tách hai sợi của chuỗi xoắn kép DNA

Trong khi các tỷ số G với C và A với T trong DNA của một sinh vật là cố định, thì hàm lượng GC (tỷ lệ phần trăm của G + C) có thể sai khác nhau một cách đáng kể giữa các DNA thuộc các loài khác nhau. Ở bảng 3.5 cho thấy hàm lượng GC của DNA nhiều loài sinh vật. Các trị số này biến thiên từ 22% đến 73%, và những sự khác nhau này được phản ánh trong sự sai khác về các đặc tính của DNA.

Ở nhiệt độ vừa phải hoặc khi có mặt các tác nhân gây biến tính

như kiềm hay formamide, thì các phân tử DNA bị biến tính từng phần. Khi đó tại các vùng giàu cặp A-T sẽ tách từng phần trước, trong khi các vùng giàu cặp G-C vẫn giữ nguyên đặc tính xoắn kép (Hình 3.11). Điều này có thể lý giải là do mỗi cặp A-T chỉ có hai liên kết hydro hiển nhiên là kém bền hơn so với mỗi cặp G-C vốn có tới ba liên kết như thế.

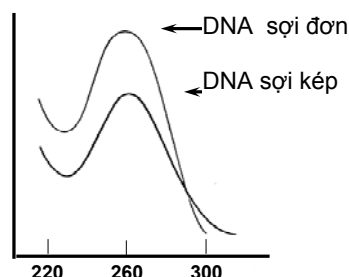
Bảng 3.5 Hàm lượng tương đối (G + C) của các DNA khác nhau

Nguồn DNA	(G+C) %	Nguồn DNA	(G+C) %
<i>Dictyostelium</i> (mốc nhầy)	22	Lách chuột	44
<i>Streptococcus pyogenes</i>	34	Tinh trùng cá hồi	44
<i>Vaccinia virus</i>	36	<i>B. subtilis</i>	44
<i>Bacillus cereus</i>	37	Phage T1	46
<i>B. megaterium</i>	38	<i>Escherichia coli</i>	51
<i>Hemophilus influenzae</i>	39	Phage T7	51
<i>Saccharomyces cerevisiae</i>	39	Phage T3	53
Tuyến ức bê	40	<i>Neurospora crassa</i>	54
Gan chuột (<i>Rattus</i>)	40	<i>Pseudomonas aeruginosa</i>	68
Tinh trùng bò đực	41	<i>Sarcina lutea</i>	72
<i>Streptococcus pneumoniae</i>	42	<i>Micrococcus lysodeikticus</i>	72
Mầm lúa mì	43	Herpes simplex virus	72
Gan gà	43	<i>Mycobacterium phlei</i>	73

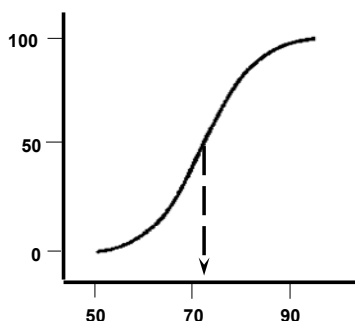
Nhiệt độ mà tại đó các sợi DNA bị biến tính hay tách nhau một nửa được gọi là *nhiệt độ nóng chảy* (melting temperature), hay T_m . T_m là điểm giữa của pha chuyển tiếp (Hình 3.13) và nó tùy thuộc vào hàm lượng GC của DNA, nghĩa là đặc trưng cho DNA mỗi loài (Hình 3.14). Ví dụ, DNA của *E. coli* với 50-51% GC thì có T_m là 69-70°C (Hình 3.14). Tương tự, kết quả xử lý nhiệt đối với DNA phế cầu khuẩn *Streptococcus pneumoniae* và nhiệt độ nóng chảy của nó được đo bằng sự gia tăng độ hấp thụ ở 260 nm cho phép thu được đường cong nóng chảy của vi khuẩn này. T_m cho DNA này dưới những điều kiện như thế là khoảng 85°C



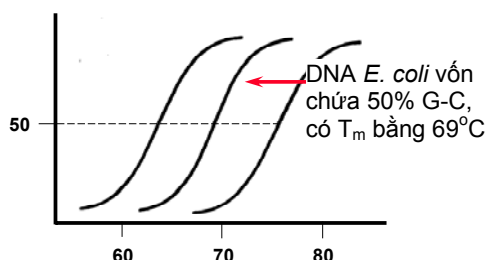
Hình 3.11 Vi ảnh điện tử của DNA bị biến tính từng phần. Các bút sợi đơn (mũi tên dưới) là vùng giàu AT bị biến tính trước, trong khi các vùng sợi kép dày hơn (mũi tên trên) vẫn còn chưa bị biến tính.



Hình 3.12 Hyperchromicity. Sự hấp thụ của một dung dịch DNA (trên trục tung) tăng lên cực đại ở 260 nm (thuộc vùng cực tím của quang phổ) khi chuỗi xoắn kép bị biến tính thành các sợi đơn.



Hình 3.13 Đường cong nóng chảy DNA. Tỷ lệ phần trăm hyperchromicity (ở trục tung) có thể được dùng theo sự biến tính của DNA như là một hàm số của sự gia tăng nhiệt độ (ở trục hoành). Nhiệt độ tại điểm giữa của đường cong nóng chảy được gọi là T_m .



Hình 3.14 Sự phụ thuộc của T_m vào hàm lượng G+C của ba DNA khác nhau. Hàm lượng GC trung bình có thể được xác định từ nhiệt độ nóng chảy của DNA. Ví dụ, đường cong bên trái nói lên một DNA có hàm lượng G+C thấp hơn so với DNA *E. coli* (đường cong ở giữa).

Cần lưu ý rằng hàm lượng tách sợi, hay nhiệt độ nóng chảy, được đo bằng sự hấp thụ của dung môi DNA ở 260 nm (Hình 3.12). Các nucleic acid hấp thụ ánh sáng ở bước sóng này do cấu trúc điện tử trong các base của chúng, nhưng khi hai sợi của DNA lại gần nhau, thì khoảng cách gần gũi của các base trên hai sợi làm giảm bớt phần nào sự hấp thụ này. Khi hai sợi tách ra thì hiện tượng này biến mất và độ hấp thụ tăng lên 30-40%. Hiện tượng này được gọi là *sự dịch chuyển do thừa sắc tố* (hyperchromic shift;

Hình 3.12) và thuật ngữ chính thức chỉ cho hiện tượng này là *hyperchromicity*. Sự tăng tiến độ dốc trên đường cong cho thấy các sợi giữ vững màu cho tới khi nhiệt độ tiệm cận T_m và sau đó nhanh chóng buông ra.

Hàm lượng GC của một DNA có một tác dụng đáng kể lên T_m của nó. Trên thực tế, hàm lượng GC của DNA càng cao thì T_m của nó càng cao (Hình 3.14). Tại sao như vậy? Ta nhớ lại rằng một trong những lực giữ cho hai sợi gắn với nhau là liên kết hydro, trong khi các cặp A-T chỉ có hai liên kết thì các cặp G-C có tới ba liên kết. Vì vậy hai sợi của DNA giàu GC sẽ giữ chặt hơn là hai sợi của DNA giàu AT.

Tóm lại, hàm lượng GC của một DNA có thể biến thiên từ 22% ở nấm mốc nhầy *Dictyostelium* đến 73% ở *Mycobacterium phlei* (Bảng 3.5). Điều này có thể gây một hiệu quả mạnh lên các đặc tính hóa lý của DNA, đặc biệt là lên nhiệt độ nóng chảy tăng tuyến tính với hàm lượng GC. Nhiệt độ nóng chảy (T_m) của một phân tử DNA là nhiệt độ mà tại đó hai sợi bị biến tính hay tách nhau một nửa. Ngoài ra, nồng độ ion thấp và các dung môi hữu cơ cũng thúc đẩy sự biến tính của DNA.

1.2. Sự phục hồi trạng thái nguyên thể của DNA (renaturation)

Một khi hai sợi của DNA tách ra, dưới những điều kiện thích hợp chúng có thể kết hợp trở lại và làm phục hồi trạng thái ban đầu (renaturation, annealing). Góp phần vào hiệu quả "hồi tính" này của DNA có nhiều nhân tố. Dưới đây nêu lên ba nhân tố quan trọng nhất:

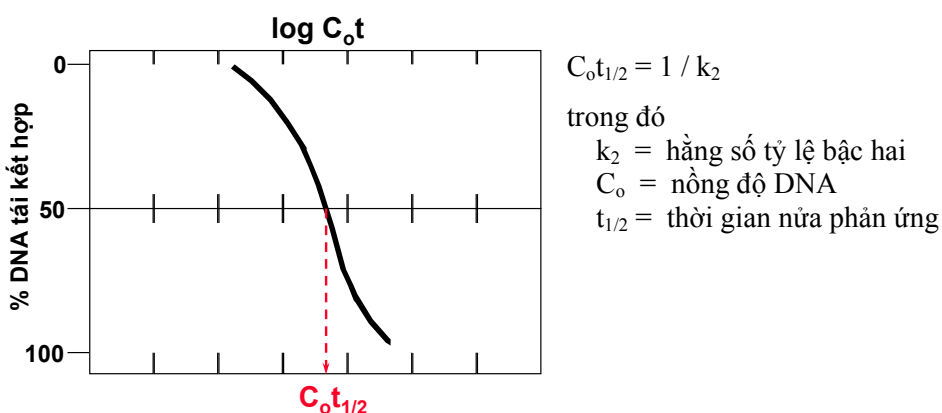
1. **Nhiệt độ.** Nhiệt độ tốt nhất cho sự hồi tính của một DNA là khoảng 25°C dưới nhiệt độ nóng chảy của nó. Nhiệt độ này là đủ thấp để cho sự biến tính không xảy ra nữa, nhưng đủ cao để cho phép khuếch tán nhanh và làm yếu đi liên kết không bền giữa các trình tự kết cặp ngẫu nhiên và các vùng kết cặp base ngắn trong sợi. Điều này gợi ra rằng việc làm nguội nhanh theo sau sự biến tính sẽ cản trở sự hồi tính. Thật vậy, quy trình chung đảm bảo cho DNA bị biến tính dừng biến tính lại là đột ngột chuyển dung dịch DNA vào trạng thái đóng băng. Điều này được gọi là *quenching*.
2. **Nồng độ DNA.** Nồng độ DNA trong dung dịch cũng quan trọng. Trong giới hạn hợp lý, nồng độ DNA càng cao thì hai sợi bổ sung sẽ càng dễ dàng bắt gặp nhau trong một thời

gian nào đó. Nói cách khác, nồng độ càng cao thì sự hàn gắn trở lại càng nhanh.

3. *Thời gian hồi tính.* Rõ ràng là, thời gian cho phép hai sợi hàn gắn trở lại càng dài thì sẽ càng dễ dàng xảy ra.

R.J. Britten và D.E. Kohn phát minh ra thuật ngữ, C_0t , để gộp các nhân tố nồng độ DNA và thời gian. C_0t (đọc là "cot") là sản phẩm của nồng độ DNA ban đầu (C_0) tính theo số mole của các nucleotide trên mỗi lít và thời gian (t) tính bằng giây. Tất cả các nhân tố khác được coi là ngang bằng nhau thì mức độ hồi tính của các sợi bổ sung trong một dung dịch DNA sẽ phụ thuộc vào C_0t ; cái đó gọi là đường cong C_0t . Nếu ta coi tỷ lệ kết hợp lại *tăng lên* theo hướng *đi xuống* trên trục y, thế thì các đường cong đồ dốc biểu thị cho sự hồi tính các DNA.

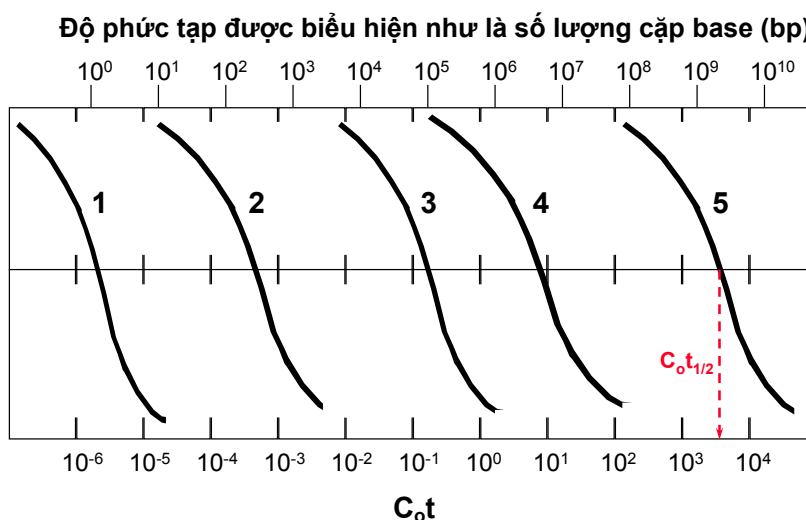
Hình 3.15 và 3.16 (kèm theo các chú thích chi tiết) dưới đây cho thấy phản ứng tái kết hợp DNA (cho một loại DNA sợi đơn) và đường cong C_0t đối với năm mẫu DNA khác nhau có các mức độ phức tạp khác nhau rất là lớn.



Hình 3.15 Phản ứng tái kết hợp DNA hay đường cong C_0t lý tưởng đối với một loại DNA sợi đơn.

Chú thích: Ở đây cho thấy mẫu DNA của *E. coli*. Phản ứng được biểu diễn trên đồ thị như là một sơ đồ bán-log với " $\log C_0t$ " trên trục x và "tỷ lệ % DNA tái kết hợp" trên trục y (0% ở trên và 100% ở dưới). Phản ứng xảy ra theo sự vận động bậc hai lý tưởng. C_0t là sản phẩm của C_0 (nồng độ DNA) và t (thời gian). Như vậy, nếu nồng độ DNA ổn định, trục x đơn thuần là tiến trình thời gian của phản ứng. Tiến trình phản ứng xảy ra như sau: tại thời điểm zero DNA bị biến tính thành các sợi đơn và các điều kiện đó được kết nối để thúc đẩy sự kết cặp base của DNA để cho phép sự hồi tính DNA xảy ra; tại một nửa thời gian của phản ứng ($t_{1/2}$) thì có 50% DNA được tái kết hợp. Điểm này là $C_0t_{1/2}$ của phản ứng mà nó xác định tỷ lệ nghịch của hằng số tỷ lệ tái kết hợp bậc hai, k_2 . Nói

cách khác, hằng số tỷ lệ đối với một mẫu DNA có thể được xác định về mặt thực nghiệm bằng cách đo $C_0t_{1/2}$ của phản ứng. Hằng số tỷ lệ sau đó sẽ cung cấp độ phức tạp của DNA bằng cách so sánh nó với các hằng số tỷ lệ của các mẫu DNA khác có độ phức tạp đã biết.



Hình 3.16 Đường cong C_0t đối với năm mẫu DNA khác nhau có các độ phức tạp rất khác nhau sau đây:

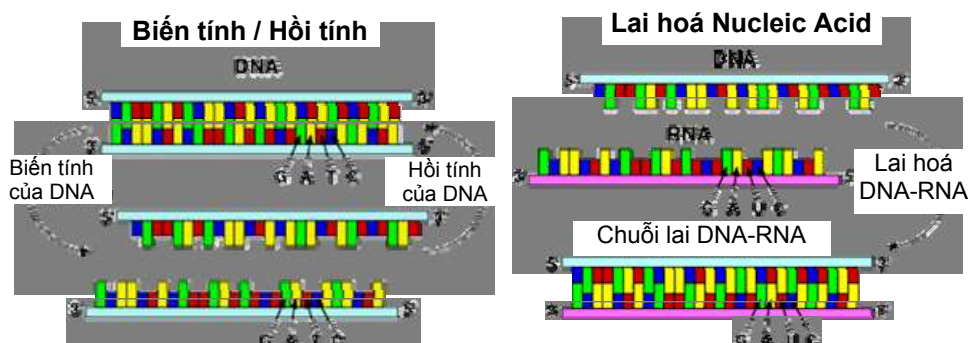
- 1- poly(dA)-poly(dT); 2- DNA vệ tinh của người được tinh chiết;
- 3- DNA phage T4; 4- DNA bộ gene *E. coli*; và
- 5- DNA bản sao đơn của người được tinh chiết.

Chú thích: Mỗi DNA tái kết hợp với một tỷ lệ nhất định, phụ thuộc vào mức độ phức tạp đặc trưng riêng của nó (xem các trị số về độ phức tạp ở hàng ngang phía trên). Mẫu 1 là trình tự đơn giản nhất. Nó là DNA sợi kép gồm các homopolymer poly(dA) và poly(dT) cặp với nhau. Theo định nghĩa có tính chất quy ước, nó có độ phức tạp là "một" bởi vì nó bao gồm các đoạn lặp của chỉ các cặp A-T mà thôi. Các mẫu DNA còn lại (2-5) có độ phức tạp tăng lên cho đến DNA bản sao đơn của người được tinh chiết, vốn cấu thành hầu hết bộ gene người và có độ phức tạp của hơn một tỷ ($>10^9$) cặp base.

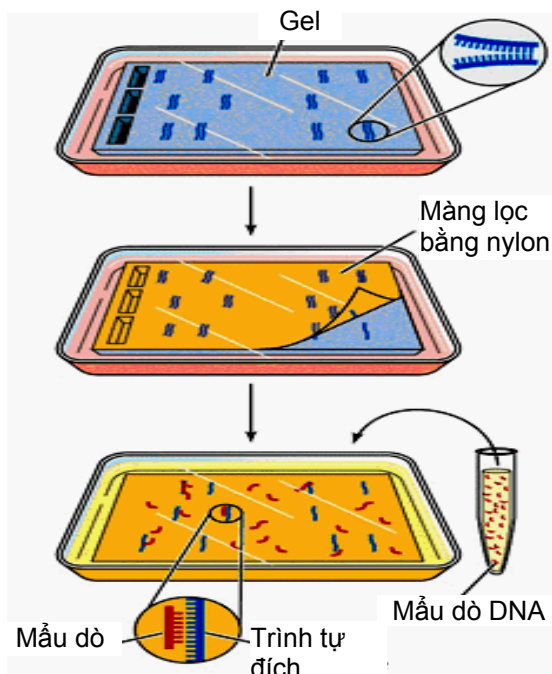
2. Ứng dụng trong lai phân tử

Người ta lợi dụng khả năng nói trên của DNA để tạo ra các phân tử DNA lai nhân tạo bằng cách làm lạnh từ từ hỗn hợp các DNA biến tính từ hai loài khác nhau. *Kỹ thuật lai phân tử* (molecular hybridization; Hình 3.17) này đã được ứng dụng rộng rãi để xác định mức độ tương đồng DNA của các nhóm phân loại khác nhau. Trên nguyên tắc, nếu mức độ tương đồng càng lớn thì số lượng các đoạn lai càng lớn và ngược lại. Ví dụ, các thực nghiệm cho thấy có khoảng 25% tổng số DNA người và chuột có thể lai với

nhau (Watson *et al* 1987). Thông thường mức độ lai được xác định bằng các phương pháp sắc ký hoặc ly tâm, trong đó một loại DNA được đánh dấu bằng đồng vị phóng xạ. Hiện giờ các kỹ thuật này được áp dụng khá rộng rãi trong các nghiên cứu sinh học phân tử cũng như trong các lĩnh vực điều tra hình sự hoặc phát hiện sớm một số bệnh phân tử trước khi các triệu chứng xuất hiện để điều trị kịp thời.



Hình 3.17 Biến tính và hồi tính của DNA (trái) và lai nucleic acid.



Hình 3.18 Sơ đồ minh họa việc sử dụng mẫu dò DNA để tìm đoạn đích.

Tóm lại, theo nghĩa rộng, lai phân tử hay lai nucleic acid được

hiều là sự tương tác giữa các sợi nucleic acid bổ sung. Nó có thể xảy ra giữa hai sợi DNA, hoặc giữa các sợi DNA và RNA, hoặc giữa hai sợi RNA, và đó chính là cơ sở của nhiều kỹ thuật, chẳng hạn như: lai một *mẫu dò có đánh dấu đồng vị phóng xạ* (radioactive probe) để lọc ra DNA hay RNA bám vào màng lai là một trong những thí nghiệm cho nhiều thông tin bổ ích nhất được tiến hành trong di truyền học phân tử. Hai kiểu cơ bản của lai phân tử là *phương pháp lai Southern* (Southern blots) và *phương pháp lai Northern* (Northern blots). Trong đó, kiểu đầu là lai bằng một mẫu dò để phát hiện, định lượng một phân tử DNA xác định; còn kiểu sau dùng để phát hiện, định lượng một phân tử RNA. *Mẫu dò* (probe) là các công cụ sơ cấp được dùng để xác định các trình tự bổ sung cần quan tâm hay *trình tự đích* (target sequence). Đó là một nucleic acid sợi đơn thường được đánh dấu phóng xạ và được dùng để xác định một trình tự nucleic acid bổ sung bám trên màng lọc nitrocellulose hay màng lai bằng nylon (Hình 3.18).

V. Chức năng của DNA

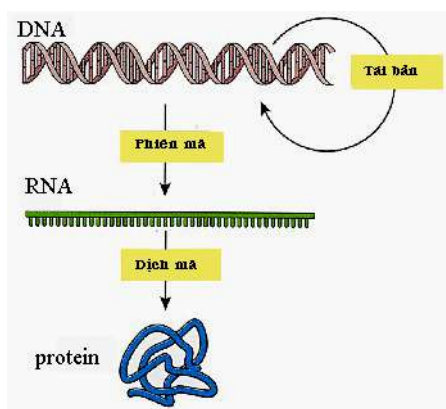
Ngày nay, chúng ta đều biết rõ rằng DNA hay bộ gene của tất cả các sinh vật nói chung có chức năng chính là mang đầy đủ toàn bộ *thông tin di truyền* (genetic information) đặc trưng cho từng loài. Thông tin di truyền này được ghi lại dưới dạng mật mã, gọi là *mã di truyền* (genetic code), và chứa đựng trong các *gene cấu trúc* (structural genes) cũng như các yếu tố kiểm soát di truyền nhằm điều khiển mọi hoạt động sinh trưởng, phân chia và biệt hoá của tế bào (chương 6).

Hơn nữa, DNA hay vật chất di truyền nói chung đều có khả năng tự sao chép một cách chính xác bản thân nó trong một quá trình gọi là *tái bản* (replication) - cơ sở của sự tự nhân đôi nhiễm sắc thể và, do đó, là cơ sở của sự phân chia tế bào (mà thực chất là sự phân chia hay truyền đạt vật chất di truyền; chương 5). Đó còn là các quá trình hoạt động và điều hoà sự biểu hiện của các gene trong bộ gene - *phiên mã* (transcription) và *dịch mã* (translation) - tạo ra các phân tử (RNA và protein) tham gia vào các cấu trúc và hoạt động sống cơ sở của tế bào (chương 6). Nhờ đó mà con cái sinh ra thường giống với cha mẹ, mỗi loài duy trì sự ổn định tương đối bộ gene của mình và, nói rộng ra là, nhờ đó mà sự sống được duy trì một cách liên tục kể từ khi sự sống bắt đầu hình thành trên trái đất cách đây chừng ba tỷ rưỡi năm.

Mặt khác, DNA hay vật chất di truyền nói chung có khả năng

phát sinh các biến đổi trong quá trình phát triển cá thể và sinh sản của sinh vật. Đó là các *đột biến gene* (gene mutations) gây ra bởi tác động của các tác nhân vật lý và hoá học khác nhau, hoặc do chính các sai sót trong quá trình tái bản, hoặc do sự dịch chuyển vị trí của bản thân các gene trong bộ gene - các *yếu tố di truyền vận động* (transposable genetic elements) hay còn gọi là các *gene nhảy* (jumping genes) - gây sự biến động của bộ gene và sự biến đổi ở kiểu hình. Ngoài ra, đó còn là các quá trình *tái tổ hợp di truyền* (genetic recombination) tạo nên các biến dị tổ hợp phong phú và đa dạng trong quá trình sinh sản của sinh vật. Chính các quá trình biến đổi đa dạng này đã không ngừng tạo nên các nguồn biến dị di truyền sơ cấp và thứ cấp cho sự chọn lọc và tiến hoá của sinh giới (chương 5).

Các chức năng và cơ chế truyền đạt thông tin di truyền chính yếu của DNA được mô tả tóm tắt như ở hình 3.19 dưới đây, và sẽ được thảo luận chi tiết trong các chương kế tiếp.



Hình 3.19 Giáo lý trung tâm của sinh học phân tử.

Nói chung, trong một bộ gene sinh vật có chứa các thành phần chức năng khác nhau thuộc hai nhóm chính sau đây:

+ Nhóm thứ nhất bao gồm các gene, tức là các thành phần của bộ gene được biểu hiện thành các sản phẩm cuối cùng là RNA hay protein, đó là: (i) Các gene cấu trúc mã hoá các RNA thông tin, quy định các protein khác nhau, gọi là *các gene mã hoá protein* (protein coding genes); (ii) Các gene mã hoá các RNA vận chuyển; (iii) Các gene mã hoá các RNA ribosome; (iv) Đối với các tế bào eukaryote, còn có các gene mã hoá các RNA chức năng đặc trưng khác như: iRNA, snRNA, snoRNA, scRNA, gRNA, RNA của các enzyme

splicing và telomerase ... (xem chương 4).

+ Nhóm thứ hai bao gồm các yếu tố không phải gene, tức là các thành phần của bộ gene không được biểu hiện thành các sản phẩm mà chỉ đóng vai trò là các tín hiệu điều hoà hoặc kiểm soát hoạt động của bộ gene hoặc các gene. Nhóm này bao gồm nhiều vùng DNA đặc thù khác nhau, chẳng hạn: (i) Các trình tự điều hoà hoạt động tái bản, như: *khởi điểm tái bản* (origin of replication = *ori*) và *kết thúc tái bản* (terminator = *ter*); (ii) Các yếu tố kiểm soát phiên mã, như: *vùng khởi động* (promoter region), *yếu tố chỉ huy* trong các operon ở prokaryote (operator, các *yếu tố tăng cường* (enhancer) hoặc *yếu tố gây bất hoạt gene* (silencer) ở eukaryote, ...; (iii) Các yếu tố kiểm soát dịch mã, ví dụ *trình tự Shine-Dalgarno*, hoặc các vùng không được dịch mã nằm trước và sau mRNA - sản phẩm gene mã hoá protein - gọi là 5'-UTR và 3'-UTR; (iv) Các *đoạn đệm giữa các gene* (intergenic spacer); (v) Các *intron* - các đoạn không mã hoá protein (phân bố xen kẽ với các đoạn mã hoá gọi là các *exon*) của các gene phân đoạn ở eukaryote, ở các vi khuẩn cổ và một số virus lây nhiễm ở sinh vật bậc cao; (vi) Các *gene giả* (pseudogene); (vii) Các trình tự *DNA lặp lại* (repetitive DNA) khác nhau, kể cả các DNA tâm động (centromere) và DNA đầu mút (telomere) của các nhiễm sắc thể eukaryote; v.v.

Như đã đề cập trước đây, trong khi mật độ phân bố các gene trong bộ gene của các prokaryote và các virus là hầu như dày đặc, thì ở các eukaryote mà đặc biệt là ở các sinh vật bậc cao, tỷ lệ này là tương đối nhỏ và thay vào đó, phần lớn bộ gene chứa các vùng DNA thuộc nhóm thứ hai nói trên.

Chẳng hạn, bộ gene người (human genome) có các đặc điểm sau:

(1) *Bộ gene nhân* (nuclear):

- bộ gene đơn bội có khoảng 3,2 tỷ cặp base;
- chứa khoảng 20.000 - 25.000 gene, chiếm khoảng 2% toàn bộ bộ gene (số liệu ước tính mới nhất công bố từ tạp chí *Nature* ra ngày 21/10/2004; chỉ trước đó không lâu con số này là ~30.000 đến 40.000);
- hầu hết các gene trong bộ gene đơn bội đều ở dạng các bản sao đơn;
- các gene đều có chứa từ 1 đến hơn 75 exon;
- các gene sai khác nhau về chiều dài, biến thiên từ dưới 100

đến trên 2.300.000 cặp base;

- các *trình tự Alu* (Alu sequence) có mặt khắp bộ gene.

(2) *Bộ gene ty thể* (mitochondrial):

- bộ gene mạch vòng với 16.569 cặp base (cũng có một số liệu khác cho là 16.571 cặp base);

- chứa dưới 40 gene.

Câu hỏi và Bài tập

1. Liệt kê các đặc điểm của mô hình DNA dạng B và từ mô hình này, hãy cho biết: Dựa vào đâu Watson đã tiên đoán một cách tài tình và chính xác khả năng tất yếu phải xảy ra kiểu tái bản bán bảo toàn của DNA?

2. (a) Về mặt toán học, mô hình Watson-Crick lý giải một cách thoả đáng các kết quả của Chargaff ra sao? (b) Hãy chỉ ra những điểm giống và khác nhau giữa DNA dạng B và DNA dạng Z, và mối quan hệ giữa chúng.

3. Anh (chị) hãy viết một tổng luận về chặng đường ra đời và phát triển của sinh học phân tử trong hơn 50 năm qua kể từ ngày Watson và Crick khám phá ra mô hình cấu trúc chuỗi xoắn kép DNA..

4. Đối với DNA sợi kép, chỉ cần biết trước hàm lượng của một loại nucleotide ta có thể xác định được hàm lượng của các loại còn lại. Tại sao? Hãy vận dụng các quy tắc Chargaff để tính số lượng từng loại nucleotide trong bộ gene người và vi khuẩn *E. coli* (Biết rằng tỷ lệ A/G của DNA người là 1,52; tỷ lệ phần trăm G + C của *E. coli* là 51%; về tổng số nucleotide của mỗi bộ gene cần tra cứu ở Bảng 3.3).

5. Anh (chị) hãy phát biểu những ý tưởng và hiểu biết của mình dựa trên các cặp base của mô hình Watson-Crick.

6. (a) Tại sao trong cấu trúc DNA cũng như các cơ chế di truyền ở cấp độ phân tử vốn quan trọng như vậy nhưng chỉ có hai kiểu kết cặp base căn bản: A-T và G-C? (b) Có thể xảy ra kiểu kết cặp base nào khác nữa hay không? Nếu có, hậu quả là gì? Giải thích và cho ví dụ.

7. Thế nào là biến tính và hồi tính của DNA? Giải thích và cho biết ý nghĩa sinh học cũng như ứng dụng của các hiện tượng này.

8. (a) Phân tích mối quan hệ giữa kích thước bộ gene của các sinh vật và tính phức tạp về mặt tiến hóa của chúng; (b) Nghịch lý giá trị-C là gì? Phân tích khái niệm này và cho ví dụ.

9. Phân tích các mối tương tác trong các cấu trúc bậc hai và bậc ba của các nucleic acid, và nêu các ý nghĩa của chúng.

10. Dựa vào các kiến thức liên quan đến cấu trúc cặp G-C, anh (chị) hãy thể hiện những hiểu biết có thể được của mình về cấu trúc này.

Tài liệu Tham khảo

Tiếng Việt

Hoàng Trọng Phán. 1995. *Một số vấn đề về Di truyền học hiện đại* (Tài liệu BDTX giáo viên THPT chu kỳ 1993-1996). Trường ĐHSP Huế.

Hoàng Trọng Phán. 1997. *Di truyền học Phân tử* (tái bản). Trung tâm ĐTTX Đại học Huế - NXB Giáo Dục.

Hoàng Văn Tiến (chủ biên), Lê Khắc Thận, Lê Doãn Diên. 1997. *Sinh hoá học với cơ sở khoa học của công nghệ gene*. NXB Nông Nghiệp, Hà Nội.

Nguyễn Tiến Thắng, Nguyễn Đình Huyền. 1998. *Giáo trình Sinh hoá hiện đại*. NXB Giáo Dục.

Watson JD. 1968. *Chuỗi xoắn kép* (bản Việt dịch của Lê Đình Lương và Thái Doãn Tĩnh). NXB Khoa học và Kỹ thuật, Hà Nội, 1984.

Tiếng Anh

Benz R. (Ed.). 2004. *Bacterial and Eukaryotic Porins: Structure, Function, Mechanism*. John Wiley & Sons, Inc., UK.

Blackburn G.M., Gait M.J. (Eds., 1996): *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford.

Bolsover SR, Hyams JS, Shephard EA, White HA, Wiedemann CG. 2003. *Cell Biology: A Short Course*, 2nd ed. John Wiley & Sons, Inc., UK.

Campbell PN, Smith AD, Peters TJ. 2005. *Biochemistry Illustrated: Biochemistry and molecular biology in the post-genomic era*. 5th ed., Elsevier Ltd, Edinburgh - London - New York - Oxford - Philadelphia - St Louis - Sydney - Toronto.

Horton, Moran, Ochs, Rawn, Scrimgeour. 2002. *Principles of Biochemistry*. Prentice Hall, Inc.

<http://cwx.prenhall.com/horton/medialib/media_portfolio/index.html>

Lehninger L. et al. (1993): *Principles of Biochemistry*. Worth Publishers, New York.

Mulligan ME. 2002.

<http://www.mun.ca/biochem/cuourses/3017/Topics/bases.html>

Nelson DL and Cox MM. 2000. *Lehninger Principles of Biochemistry*, 3rd ed., Worth Publishers, New York.

O'Brien SJ, Menninger J, Nash WG. 2006. *Atlas of Mammalian Chromosomes*. John Wiley & Sons, Inc., UK.

Stein CA and Krieg AM (Eds), March 1998. *Applied Antisense Oligonucleotide Technology*. John Wiley & Sons, Inc. / BIOS Scientific Publishers Ltd, UK.

Stryer L. (1981): *Biochemistry*. W.-H. Freeman and Co., San Francisco.

Twyman RM. 1998. *Advanced Molecular Biology*. BIOS Scientific Publishers Ltd/ Springer-Verlag Singapore Pte Ltd.

Watson JD, Tooze J., Kurtz DT. 1983. *Recombinant DNA (A short course)*. Scientific American Books, W.H. Freeman & Co., New York.

Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. 1987. *Molecular Biology of the Gene*. 4th ed, Benjamin/Cummings Publishing Company, Inc, Menlo Park, CA.

Weaver RF, Hedrick PW. 1997. *Genetics*. 3rd ed, McGraw-Hill Companies, Inc. Wm.C.Brown Publishers, Dubuque, IA.

Chương 4

Cấu trúc và Chức năng của các RNA

Trên nguyên tắc, các RNA được cấu tạo từ các đơn phân là các *ribonucleotide*; các ribonucleotide này nối kết với nhau bằng các liên kết 3',5'-phosphodiester tạo thành các chuỗi *polyribonucleotide* - cấu trúc sơ cấp của các phân tử RNA (như đã đề cập ở chương 2). Đường pentose đặc trưng của RNA là ribose, còn thành phần base, ngoài bốn loại cơ bản adenine (A), uracil (U), guanine (G) và cytosine (C), còn phát hiện khá nhiều dạng base hiếm có mặt chủ yếu trong các tRNA (xem mục III).

Có ba loại phân tử RNA cơ bản tham gia vào quá trình sinh tổng hợp protein của các tế bào, đó là: *RNA thông tin* (messenger RNA, viết tắt: mRNA), *RNA vận chuyển* (transfer RNA, viết tắt: tRNA), và *RNA ribosome* (ribosomal RNA, viết tắt: rRNA).

Nói chung, các phân tử RNA có kích thước bé hơn các phân tử DNA ở bất kỳ sinh vật cụ thể nào. Các phân tử RNA có thể là sợi đơn hoặc sợi kép, mạch thẳng hoặc mạch vòng nhưng phổ biến là dạng sợi đơn, thẳng (nhưng không thấy có các phân tử RNA sợi kép, vòng nào được mô tả). Loại RNA có hàm lượng cao nhất trong các tế bào là rRNA.

Ở Bảng 4.1 cho thấy hàm lượng tương đối (%) và kích thước (trọng lượng phân tử - TLPT và số nucleotide) của các phân tử RNA ở vi khuẩn *Escherichia coli* (*E. coli*).

Bảng 4.1 Các phân tử RNA ở *E. coli*

Loại RNA	Chức năng	(%)	TLPT	Số nucleotide
mRNA	Mã hoá các protein	5	Biến thiên	Biến thiên
tRNA	Mang amino acid	15	$2,5 \cdot 10^1$	~ 75
rRNA	5S Thành phần ribosome	80	$3,6 \cdot 10^1$	120
	16S Thành phần ribosome		$0,55 \cdot 10^3$	1542
	23S Thành phần ribosome		$1,2 \cdot 10^3$	2904

Ngoài ba lớp RNA chính (rRNA, tRNA và mRNA) vốn cũng có mặt trong các prokaryote, các tế bào eukaryote còn có các lớp RNA khác, như: (i) *RNA dị nhân*, hnRNA (heterogenous nuclear RNA), với kích thước sai biệt nhau rất lớn, chúng là tiền thân của các mRNA trưởng thành sau này; (ii) các *RNA nhân kích thước bé*, snRNA (small nuclear RNA), với các loại như U1, U2, U3, U4, U5, U6, U7, U8, U9, U10...trong đó sáu loại đầu

có vai trò quan trọng trong xử lý pre-mRNA của các gene phân đoạn; và (iii) *RNA tế bào chất*, scRNA (small cytoplasmic RNA) 7SL cần cho tổng hợp các protein chế tiết và bám vào màng; và một vài RNA khác nữa được giới thiệu ở Bảng 4.2. Các enzyme chịu trách nhiệm cho tổng hợp các RNA này sẽ được trình bày ở chương 6.

Bảng 4.2 trình bày khái quát về các lớp RNA có chức năng chính yếu và thứ yếu trong các tế bào.

Bảng 4.2 Các lớp RNA chức năng chính và phụ

Lớp RNA	Chức năng
<i>1. Các lớp chính</i>	
mRNA	RNA được phiên mã từ các gene mã hoá protein, mang thông tin cho dịch mã. Một số bản sao tương tự mRNA không được dịch mã, ví dụ XIST, H19 do cơ chế in dấu bộ gene bố mẹ (parental imprinting).
hnRNA (heterogenous nuclear RNA)	mRNA trước khi cắt-nối. Đó là các bản sao chưa được sửa đổi của các gene eukaryote; sở dĩ gọi như vậy bởi vì tính đa dạng lớn về kích thước của nó so với tRNA và rRNA.
tRNA	Phân tử thích ứng (adaptor) thực hiện việc dịch mã. tRNA cũng làm môi cho tái bản DNA trong sự tái bản của các retrovirus.
rRNA	Thành phần cấu trúc chính của các ribosome, cần cho quá trình tổng hợp protein của tế bào.
<i>2. Các lớp phụ</i>	
iRNA (initiator RNA)	Các trình tự RNA ngắn được dùng làm môi cho sự tổng hợp DNA ở sợi ra chậm (xem <i>tái bản</i> , chương 5).
snRNA (small nuclear RNA) hay U-RNA (uridine-rich RNA)	Các phân tử RNA trọng lượng phân tử thấp phát hiện được trong dịch nhân, là thành phần của các enzyme cắt bỏ các intron và các phản ứng xử lý (processing) khác; chúng chứa nhiều gốc uridine được sửa đổi.
snoRNA (small nucleolar RNA)	Các phân tử RNA trọng lượng phân tử thấp phát hiện được trong hạch nhân, có thể tham gia vào quá trình xử lý rRNA (RNA processing).
scRNA (small cytoplasmic RNA)	Các phân tử RNA trọng lượng phân tử thấp phát hiện được trong tế bào chất với các chức năng khác nhau. Ví dụ đó là RNA 7S vốn là thành phần của tiểu phần nhận biết tín hiệu và <i>pRNA</i> (prosomeal RNA), một RNA bé kết hợp với khoảng 20 protein và được bọc gói với mRNA trong <i>mRNP</i> hay <i>thể thông tin</i> .

	(informosome) vốn có tác dụng điều hoà sự biểu hiện của gene.
RNA telomerase	Một RNA nhân có chứa khuôn cho các đoạn lặp <i>telomere</i> và là thành phần của enzyme <i>telomerase</i> (xem chương 5).
gRNA (guide RNA)	Một loại RNA được tổng hợp trong các roi động (kinetoplasts) ở <i>Trypanosoma</i> ; nó cung cấp khuôn cho <i>biên tập RNA</i> (editing RNA).
antisense RNA	<i>RNA ngược nghĩa</i> (antisense RNA) bổ sung với mRNA và có thể tạo thành một sợi đôi với nó để kìm hãm việc tổng hợp protein. Loại RNA này thấy có trong nhiều hệ thống, nhưng rất phổ biến ở vi khuẩn; và cũng được gọi là <i>RNA bổ sung gây nhiễu mRNA</i> .
Các Ribozyme	Các phân tử RNA mà có thể xúc tác cho các phản ứng hoá học, các <i>enzyme chứa RNA</i> (RNA enzymes). Thông thường nó có hoạt tính tự xúc tác (ví dụ các <i>intron tự cắt</i> = self-splicing introns), nhưng một ribonuclease P là một chất xúc tác đích thực (ví dụ <i>xử lý tRNA</i> : tRNA processing). Các RNA khác hoạt động hài hoà với các protein, ví dụ MRP endonuclease trong tái bản DNA ty thể.

I. Cấu trúc và chức năng của các RNA thông tin (mRNA)

Trong nhân các tế bào eukaryote có các RNA nhân kích thước lớn và sai khác nhau rất lớn gọi là hnRNA (*heterogenous nuclear RNA*) vốn là tiền thân của các mRNA, các RNA nhân kích thước bé snRNA (*small nuclear RNA*) có mặt trong thành phần của các enzyme splicing, và các RNA tế bào chất kích thước bé scRNA (*small cytoplasmic RNA*).

Bảng 4.3 Độ phức tạp của các lớp mRNA trong tế bào động vật có vú

Lớp phong phú	Độ phong phú	Số lượng mRNA	
	(số bản sao/ tế bào)	khác nhau	Tổng
cao	12.000	9	108.000
trung gian	300	700	210.000
thấp (hiếm)	15	11.500	172.500
	Cộng	12.209	490.500

Một tế bào eukaryote điển hình chứa chừng ba lớp mRNA phong phú được trình bày ở Bảng 4.3. Thật vậy, trong các tế bào động vật có vú, một vài loại mRNA thì hết sức phong phú, trong khi đó hầu như mức độ phức

tập của mRNA (số lượng loại mRNA khác nhau) được đại diện bởi các mRNA hiếm. Một điểm cần chú ý là sự biểu hiện của một gene là tỷ lệ với độ phong phú của loại RNA tương ứng (một gene biểu hiện càng mạnh khi số bản sao của nó trong tế bào càng lớn).

1. Chức năng của các mRNA

Các mRNA là loại RNA quan trọng nhất được dùng làm khuôn trực tiếp cho quá trình tổng hợp các chuỗi polypeptide trong tế bào chất.

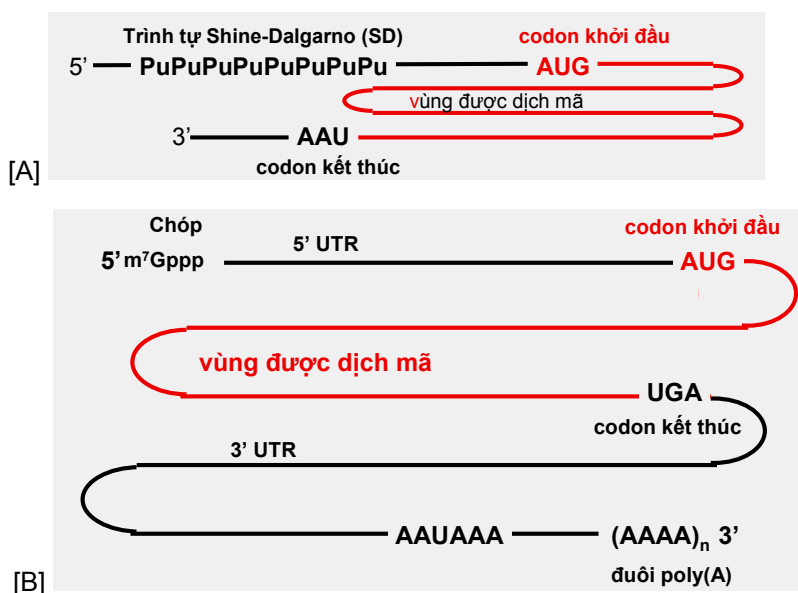
2. Cấu trúc của các mRNA

Nhìn chung, các mRNA có cấu trúc mạch thẳng, với kích thước khác nhau và đều có ba phần chính như sau:



5'-UTR |← vùng mã hóa →| 3'-UTR

(i) *Vùng dẫn đầu* (5'UTR) không được dịch mã nhưng có cấu trúc cần thiết cho sự bám vào của tiểu đơn vị ribosome bé (Hình 4.1).

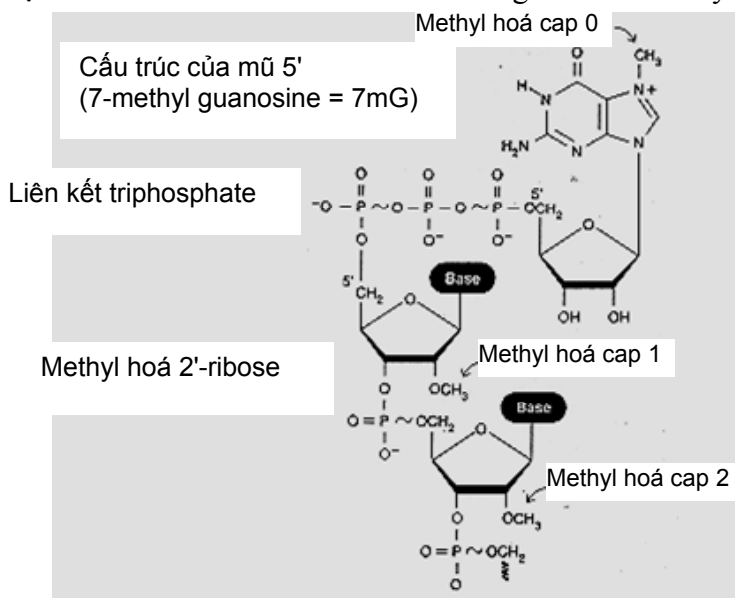


Hình 4.1 Cấu trúc của mRNA prokaryote (A) và mRNA eukaryote (B). Ở cấu trúc mRNA prokaryote cho thấy: (i) vùng 5'-UTR chứa trình tự Shine-Dalgarno (SD, gồm 8 base purine), vị trí tương tác với vùng đặc thù giàu pyrimidine của rRNA 16S trong tiểu đơn vị ribosome bé để khởi đầu tổng hợp protein; (ii) vùng được dịch mã được giới hạn bởi codon khởi đầu và codon kết thúc; và (iii) vùng 3'-UTR nằm sau codon kết thúc. Ở cấu trúc mRNA eukaryote cho thấy rõ "mũ" m⁷Gppp ở đầu mót 5' và đuôi poly(A) ở đầu 3'.

(ii) *Vùng mã hoá* (coding region) nằm kế sau vùng 5'-UTR; nó mang thông tin cấu trúc của một chuỗi polypeptide, nếu là mRNA của eukaryote (monocistronic mRNA) hoặc mang thông tin của nhiều polypeptide khác nhau và cách nhau bởi các đoạn đệm không được dịch mã, nếu là mRNA prokaryote (polycistronic mRNA).

(iii) *Vùng kéo sau* (3'-UTR) nằm ở đuôi mRNA, không được dịch mã.

Ở hình 4.1 cho thấy những điểm khác nhau trong cấu trúc của các mRNA ở prokaryote và eukaryote, và hình 4.2 cho thấy cấu trúc "mũ" đặc trưng có mặt ở đầu 5' của tất cả các mRNA trưởng thành ở eukaryote.



Hình 4.2 Cấu trúc của "mũ" (5' cap) có mặt ở tất cả các mRNA eukaryote.

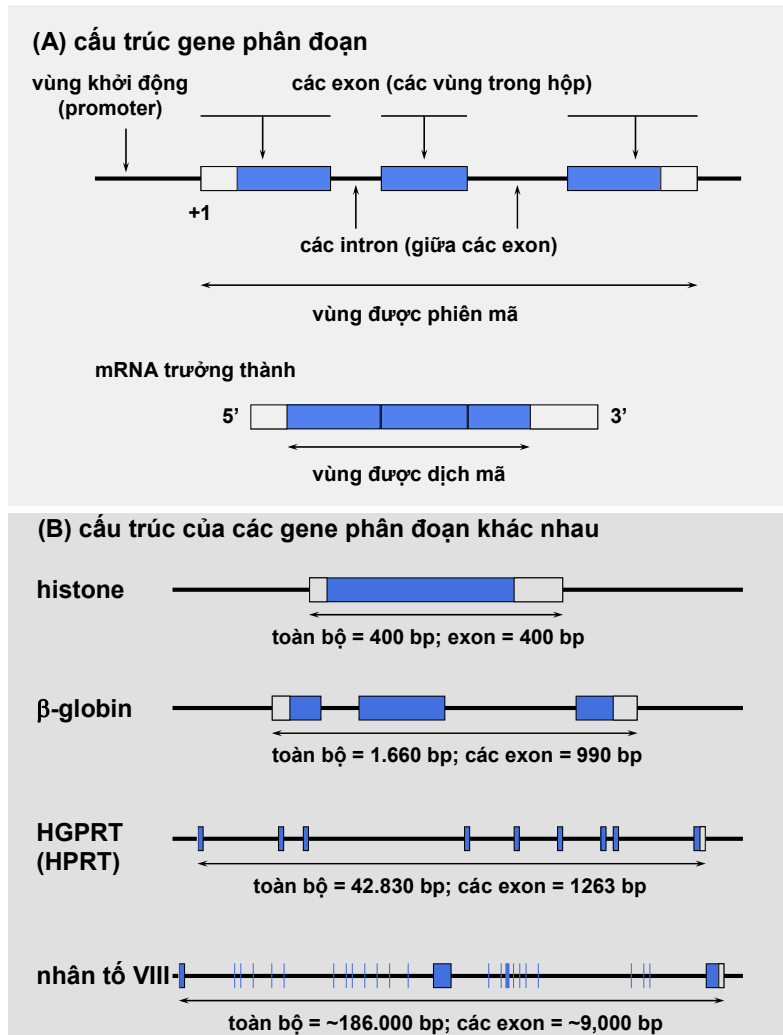
3. Sơ lược cấu trúc gene phân đoạn eukaryote và sự sửa đổi sau phiên mã

Như đã đề cập, trừ mRNA prokaryote ra, tất cả các RNA còn lại dù ở pro- hay eukaryote đều phải trải qua quá trình sửa đổi sau phiên mã với rất nhiều cơ chế tinh vi và phức tạp khác nhau để tạo ra các RNA trưởng thành tham gia vào quá trình dịch mã. Để có cái nhìn hệ thống, ở đây ta hãy tìm hiểu đôi nét về cấu trúc gene quan trọng nhất ở các eukaryote, các gene mã hoá protein và sự xử lý sau phiên mã các bản sao sơ cấp của chúng. Vấn đề này sẽ được đề cập chi tiết hơn ở chương 6.

* Về cấu trúc của các gene mã hoá protein ở eukaryote

Hầu hết các gene mã hoá protein ở eukaryote là các *gene phân đoạn* (split genes), nghĩa là trong vùng mã hoá protein của chúng bao gồm các đoạn mã hoá (gọi là các *exon*) nằm xen kẽ với các đoạn không mã hoá (gọi

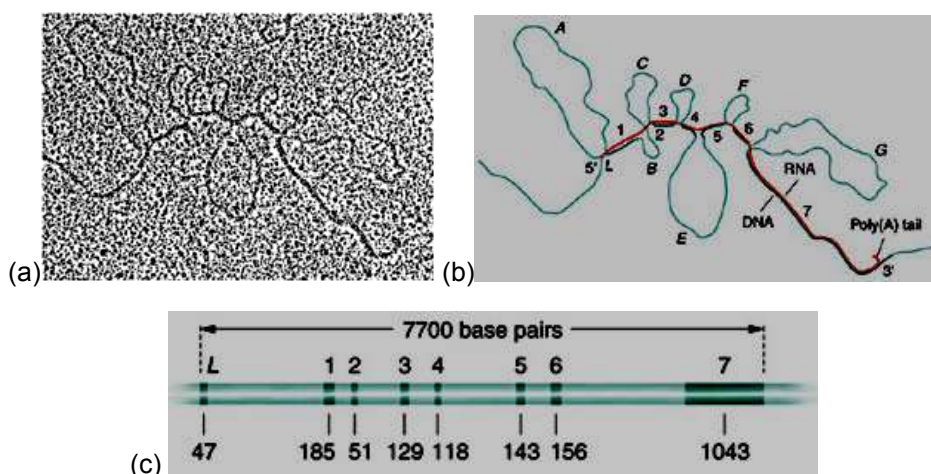
là các *intron*). Sau khi phiên mã, các intron trong bản sao pre-mRNA của các gene này phải được loại bỏ ngay trong nhân cùng với một số sự kiện quan trọng khác. Hình 4.3 cho thấy cấu trúc của một gene điển hình ở eukaryote và một số ví dụ về các gene mã hoá protein trong bộ gene người.



Hình 4.3 (A) Cấu trúc của một gene mã hoá protein điển hình ở eukaryote và mRNA tương ứng của nó. (B) Minh hoạ cấu trúc một số gene mã hoá protein trong bộ gene người. Ở đây cho thấy một vài gene như gene histone chẳng hạn là không có các intron; còn đại bộ phận gene đều có chứa intron, ví dụ: gene β -globin có ba exon và hai intron; gene mã hoá hypoxanthine-guanine phosphoribosyl transferase (HGPRT hoặc HPRT) có chín exon và lớn hơn gene histone trên 100 lần, tuy nhiên mRNA của nó chỉ lớn gấp chừng ba lần mRNA histone (chiều dài toàn bộ các exon là 1.263 bp); và gene của nhân tố VIII gây đông máu có quá nhiều intron (được biểu thị bằng các đường kẻ đứng mảnh).

* Gắn thêm "mũ" m^7Gppp và đuôi $poly(A)$

Để trở thành mRNA trưởng thành trước khi đi ra tế bào chất làm khuôn cho dịch mã, tất cả các pre-mRNA của các gene mã hóa protein của eukaryote đều trải qua hai sự kiện chính yếu trong nhân: (i) Lắp thêm vào đầu 5' một cái "mũ" 7-methylguanosinetriphosphate (m^7Gppp cap; Hình 4.2); và (ii) gắn thêm một cái đuôi $poly(A)$ dài khoảng 150 - 200 base ở đầu 3'; ngoại trừ các mRNA của histone là không có đuôi $poly(A)$. Đuôi $poly(A)$ -3' và cả "mũ"-5' có chức năng bảo vệ mRNA khỏi bị sớm thoái hoá, và trong nhiều trường hợp đuôi $poly(A)$ còn kích thích sự dịch mã. Đối với các gene mã hóa protein không có các intron, ví dụ các gene histone, quá trình hoàn thiện mRNA kết thúc tại đây.

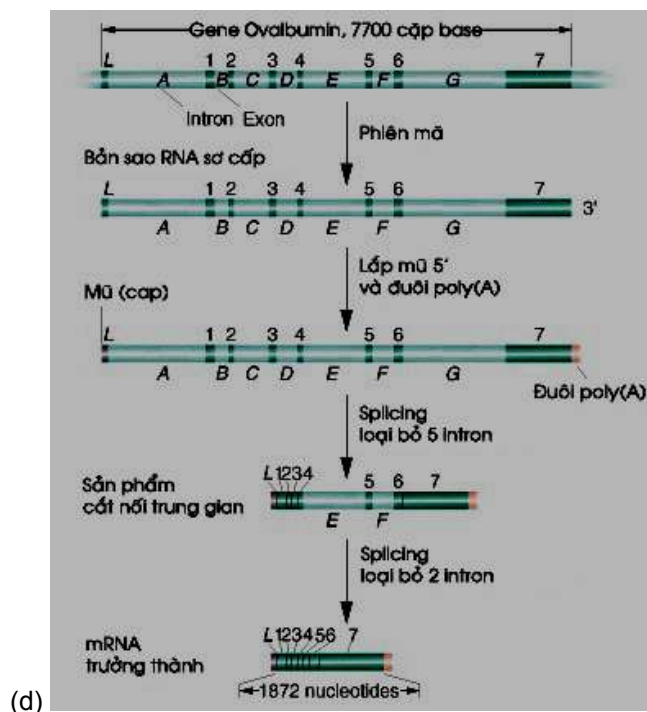


Hình 4.4 (a-b) Vi ảnh điện tử và sơ đồ minh họa sự lai hoá giữa mRNA ovalbumin trưởng thành được đánh dấu với sợi khuôn của gene ovalbumin thuộc DNA bị biến tính. Sự kết cặp bổ sung tạo thành chuỗi xoắn kép lai RNA-DNA được biểu thị bằng các đoạn mã hoá L và 1-7. Các vùng ký hiệu A -G là các intron của sợi khuôn gene, do không có vùng bổ sung tương ứng trên mRNA để kết cặp nên chúng xuất hiện dưới dạng các vòng. (c) Cấu trúc của gene ovalbumin, gồm đoạn mã hoá "leader" (L) với các exon 1-7 (hàng trên) và số lượng cặp base tương ứng (hàng dưới); xen kẽ giữa chúng là các intron.

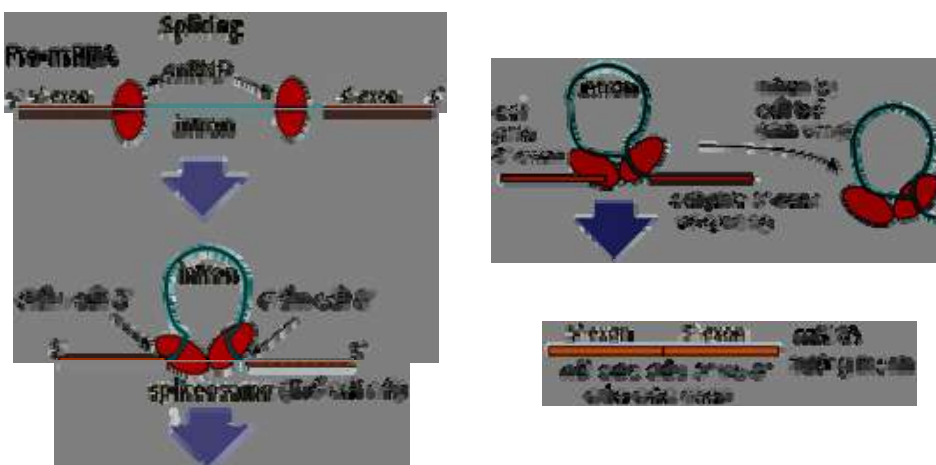
* Loại bỏ các intron và nối các exon

Đối với sản phẩm phiên mã sơ cấp của các gene phân đoạn (pre-mRNA), ngoài hai sự kiện chung nói trên còn có các quá trình loại bỏ các intron và nối các exon với nhau gọi là *splicing* hay *xử lý RNA* (RNA processing). Ví dụ, gene ovalbumin gồm bảy intron xen kẽ giữa tám exon có độ dài 7.700 cặp base đã được E. Chambon phân tích trình tự đầy đủ vào năm 1981 (Hình 4.4). Sau khi enzyme splicing cắt bỏ các intron và nối

tất cả các exon trong một quá trình gọi là *xử lý RNA* (RNA processing) thì mRNA trưởng thành có vùng mã hóa protein dài 1.872 base (hình 4.5).



Hình 4.5 Phiên mã gene ovalbumin và sự tạo thành mRNA trưởng thành.



Hình 4.6 Một mô hình về cơ chế cắt-nối trong quá trình xử lý pre-mRNA.

Có hai sự kiện chính liên quan cơ chế *cắt-nối* (splicing) trong quá trình xử lý pre-mRNA được tóm tắt như sau (về chi tiết, xem chương 6): (1) Ở

hai đầu mút của mỗi intron có hai nucleotide rất ổn định, đó là 5'-GU.....AG-3'; và (2) Ở một số snRNA có mặt trong thành phần của phức hợp enzyme cắt-nối (*spliceosome*) cũng có các trình tự dinucleotide bổ sung với các trình tự chuẩn trong intron. Các trình tự này của snRNA tương tác với các đầu mút intron, kéo chúng xích lại gần nhau tạo ra cấu trúc hình vòng. Nhờ đó enzyme tiến hành loại bỏ intron và nối các exon lại với nhau; và cuối cùng, tạo ra phân tử mRNA trưởng thành (Hình 4.6).

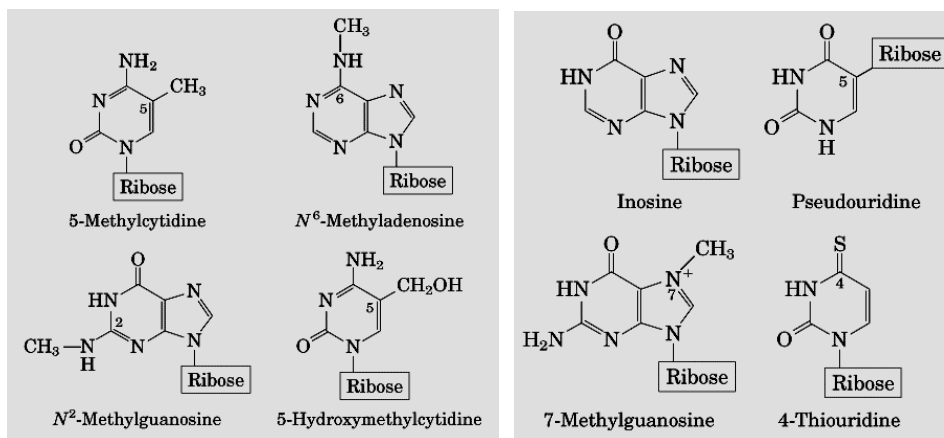
II. Cấu trúc và chức năng của các RNA vận chuyển (tRNA)

1. Chức năng của các tRNA

Mỗi phân tử tRNA có hai chức năng chính là mang amino acid đã được hoạt hoá và đi đến phức hệ "ribosome-mRNA" để tiến hành việc đọc dịch mã cho một codon cụ thể của mRNA.

2. Thành phần hoá học của các tRNA

Trong thành phần nucleotide của các tRNA có khá nhiều base chuẩn bị biến đổi thành các base sửa đổi nhờ hoạt động xúc tác của các enzyme sau phiên mã. Các base này (còn gọi là các base hiếm) tập trung chủ yếu ở các vòng thân (stem loops) như: 5',6'-dihydrouridine (DHU), inosine (I), ribothymidine (T), pseudouridine (Ψ) v.v. (Hình 4.7A).



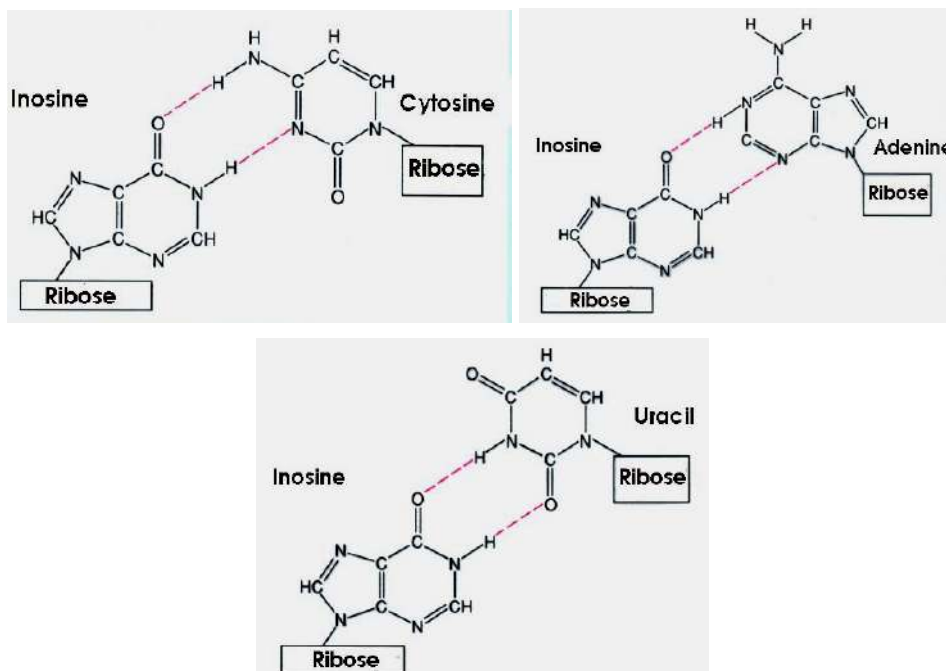
Hình 4.7A Các base hiếm có mặt trong RNA, chủ yếu là các tRNA.

3. Cấu trúc của các tRNA

Có 86 tRNA ở *E. coli*. Hầu hết các tRNA có khoảng 75-80 nucleotide và có cấu trúc bậc hai mở rộng do các tương tác cặp base (A-U và G-C) ở một số đoạn của chúng (Hình 4.8) cũng như cấu trúc bậc ba (không phải dạng siêu xoắn, mà nó có kiểu uốn gập thêm nữa trong không gian ba

chiều). Đây là kiểu cấu trúc "lá ba thuyền" gọn nhẹ và vững chắc phù hợp với các chức năng khác nhau của các tRNA.

Nói chung, các phân tử tRNA thường rất giống nhau ở nhiều đoạn và khác nhau chủ yếu ở *bộ ba đối mã* (anticodon). Cần lưu ý rằng, base hiếm Inosine (I) có mặt ở vị trí 5' của anticodon trong một số phân tử tRNA có thể kết cặp linh hoạt với một trong các base ở vị trí 3' (C, U hoặc A) của các codon đồng nghĩa trong mRNA (Hình 4.7B).

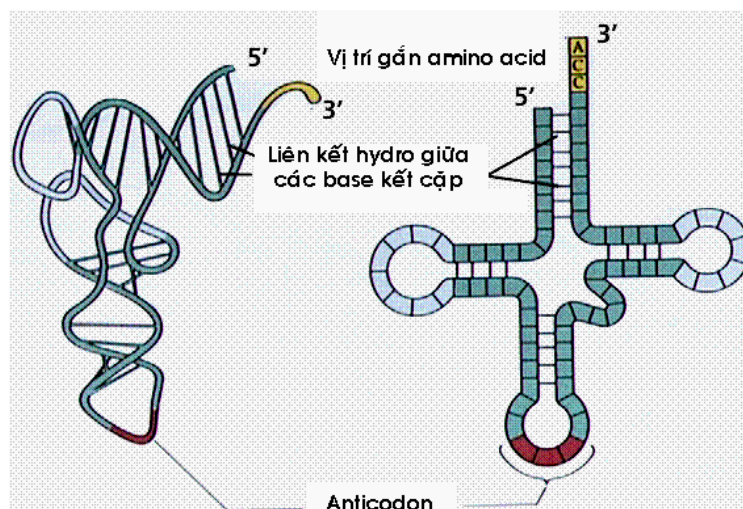


Hình 4.7B Base hiếm Inosine ở vị trí 5' của anticodon trong một số tRNA có thể kết cặp với một trong các base (C, U hoặc A) ở vị trí 3' của các codon đồng nghĩa trong mRNA.

Mỗi tRNA thường có 3-4 vòng trên thân (tính từ đầu 5') với chức năng khác nhau như sau:

- (i) vòng *DHU* nhận biết aminoacyl-tRNA synthetase;
- (ii) vòng *anticodon* đọc mã trên mRNA bằng sự kết cặp anticodon-codon (theo nguyên tắc bổ sung nhưng có sự linh hoạt; xem chương 6);
- (iii) vòng "*phụ*" (extra loop) có thể không có ở một số tRNA; và
- (iv) vòng *T Ψ C* nhận biết ribosome để đi vào đúng vị trí tiếp nhận aminoacyl-tRNA (vị trí A).

Và cuối cùng, đoạn mạch thẳng -CCA ở đầu 3' là vị trí gắn vào của amino acid đã được hoạt hoá để tạo thành các aminoacyl-tRNA.



Hình 4.8 Cấu trúc bậc ba (trái) và bậc hai của một phân tử tRNA.

III. Cấu trúc và chức năng của các RNA ribosome (rRNA)

1. Chức năng của các rRNA

Các rRNA cùng với các protein đặc thù là những thành phần cấu trúc nên các ribosome - "nhà máy" tổng hợp protein của tế bào (Hình 4.9).

2. Cấu trúc của các rRNA và ribosome

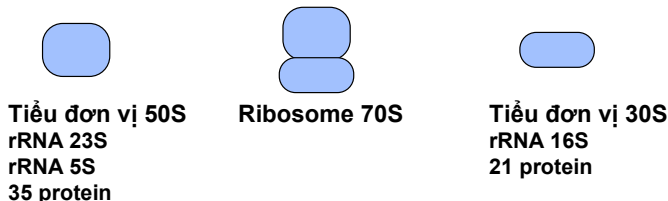
Ở vi khuẩn có 3 loại rRNA có các hệ số lắng là 23S, 16S và 5S, với số lượng nucleotide tương ứng là 2904, 1542 và 120 (xem Bảng 4.1). Ở tế bào eukaryote có 4 loại rRNA với các hệ số lắng là 28S, 18S, 5,8S và 5S. Riêng các tế bào thực vật còn có thêm các rRNA được mã hoá trong các chloroplast DNA (cpDNA).

Các hợp phần cấu tạo nên các ribosome của prokaryote và eukaryote được trình bày ở Bảng 4.4 và Hình 4.9.

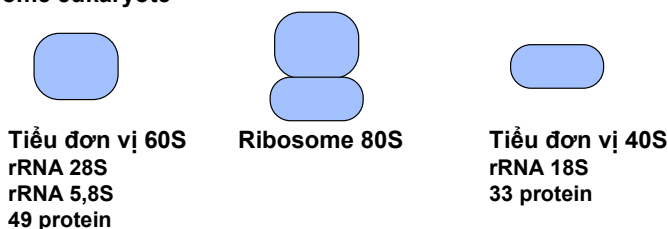
Bảng 4.4 Thành phần cấu tạo của các ribosome (R) ở pro- và eukaryote

	Thành phần	R 70S ở vi khuẩn	R 80S ở eukaryote
Tiểu đơn vị bé	rRNA	16S	18S
	Protein	21 phân tử	33 phân tử
Tiểu đơn vị lớn	rRNA	23S + 5S	28S + 5S + 5,8S
	Protein	35 phân tử	49 phân tử
Đường kính		18-20 nm	20-22 nm

•ribosome prokaryote



•ribosome eukaryote



Hình 4.9 Các hợp phần cấu thành các ribosome của pro- và eukaryote.

Mỗi ribosome hoàn chỉnh có hai *tiểu đơn vị* bé và lớn (small and large subunits). Hai tiểu đơn vị này chỉ kết hợp với nhau tạo ra một ribosome hoạt động khi quá trình dịch mã trên mRNA thực sự bắt đầu. Tiểu đơn vị bé bám vào mRNA trước tiên trong dịch mã. Tiểu đơn vị lớn chứa hai vị trí: vị trí A là nơi bám vào của aminoacyl-tRNA và vị trí P là chỗ dừng tạm của peptidyl-tRNA. Trong tiểu đơn vị lớn có chứa *peptidyl transferase*. Enzyme này có chức năng tách gốc peptidyl ra khỏi tRNA của nó (ở vị trí P) và nối với aminoacyl-tRNA (ở vị trí A) bằng một liên kết peptide làm cho chuỗi polypeptide sinh trưởng dài ra theo chiều N→C (xem chương 6, mục II-1 và IV-2).

Câu hỏi và Bài tập

1. So sánh các lớp RNA trong các tế bào prokaryote và eukaryote.
2. Mức độ phức tạp của các mRNA trong các tế bào động vật có vú được biểu hiện như thế nào?
3. Hãy chỉ ra những đặc điểm giống và khác nhau trong cấu trúc của các mRNA trưởng thành của các tế bào prokaryote và eukaryote.
4. Nêu những điểm chính trong sự sửa đổi sau phiên mã đối với sản phẩm phiên mã sơ cấp của các gene phân đoạn và vai trò của các intron.
5. Phân tích sự phù hợp giữa cấu trúc và chức năng của các mRNA.

6. Phân tích sự phù hợp giữa cấu trúc và chức năng của các tRNA.
7. Có những loại rRNA nào trong các tế bào prokaryote và eukaryote? Chúng đóng vai trò gì trong tế bào?
8. Hãy cho biết sự giống nhau và khác nhau trong thành phần cấu tạo của các ribosome ở các tế bào prokaryote và eukaryote. Cho biết ý nghĩa của sự giống và khác nhau đó.
9. Thế nào là những base sửa đổi dạng hiếm? Chúng có mặt chủ yếu trong loại RNA nào? Vẽ một sơ đồ minh hoạ.
10. Tại sao hàm lượng các rRNA rất phong phú trong các tế bào, trong khi các RNA khác hiếm hơn? Sự ổn định và bảo tồn cao độ của các tRNA và rRNA ở các tế bào prokaryote và eukaryote có ý nghĩa gì trên phương diện tiến hoá?

Tài liệu Tham khảo

Tiếng Việt

Nguyễn Bá Lộc. 2004. *Giáo trình Axit nucleic và Sinh tổng hợp protein* (tái bản). Trung tâm ĐTTX - Đại học Huế.

Hoàng Trọng Phán. 1993. *Giáo trình Di truyền phân tử* (ronéo). Trường ĐHSP Huế.

Hoàng Trọng Phán. 1995. *Một số vấn đề về Di truyền học hiện đại* (Tài liệu BDTX giáo viên THPT chu kỳ 1993-1996). Trường ĐHSP Huế.

Hoàng Trọng Phán. 1997. *Di truyền học phân tử* (tái bản). Trung tâm ĐTTX Đại học Huế - NXB Giáo Dục.

Hoàng Văn Tiến (chủ biên), Lê Khắc Thận, Lê Doãn Diên. 1997. *Sinh hoá học với cơ sở khoa học của công nghệ gene*. NXB Nông Nghiệp, Hà Nội.

Tiếng Anh

Bolsover SR, Hyams JS, Shephard EA, White HA, Wiedemann CG. 2003. *Cell Biology: A Short Course*, 2nd ed. John Wiley & Sons, Inc., UK.

Blackburn GM, Gait MJ (Eds., 1996): *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford.

Campbell PN, Smith AD, Peters TJ. 2005. *Biochemistry illustrated - Biochemistry and molecular biology in the post-genomic era*. 5th ed., Elsevier Limited, Edinburgh - London - New York - Oxford - Philadelphia - St Louis - Sydney - Toronto. (www.elsevierhealth.com)

Horton, Moran, Ochs, Rawn, Scrimgeour. 2002. *Principles of Biochemistry*. Prentice Hall, Inc.

<http://cwx.prenhall.com/horton/medialib/media_portfolio/index.html>

Mulligan ME. 2002.

<http://www.mun.ca/biochem/cuourses/3017/Topics/bases.html>

Lehninger L. *et al.* (1993): *Principles of Biochemistry*. Worth Publishers, New York.

Nelson DL and Cox MM. 2000. *Lehninger Principles of Biochemistry*, 3rd ed., Worth Publishers, New York.

O'Brien SJ, Menninger J, Nash WG. 2006. *Atlas of Mammalian Chromosomes*. John Wiley & Sons, Inc., UK.

Russell PJ. 2003. *Essential Genetics*. Benjamin/Cummings Publishing Company, Inc, Menlo Park, CA.

Stryer, L. (1981): *Biochemistry*. W.-H. Freeman and Co., San Francisco.

Tamarin RH. 1999. *Principles of Genetics*. 6th ed, McGraw-Hill, Inc, NY.

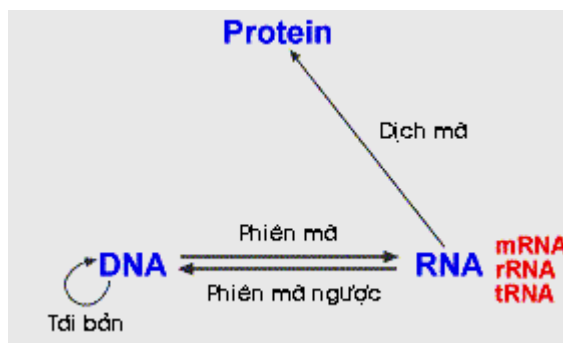
Twyman RM. 1998. *Advanced Molecular Biology*. BIOS Scientific Publishers Ltd/ Springer-Verlag Singapore Pte Ltd.

Weaver RF, Hedrick PW. 1997. *Genetics*. 3rd ed, McGraw-Hill Companies, Inc. Wm.C.Brown Publishers, Dubuque, IA.

Chương 5

Sinh tổng hợp Nucleotide và DNA

Sau khi khám phá ra cấu trúc DNA, Watson và Crick đã đề xuất ba kiểu truyền thông tin di truyền có thể có trong các tế bào: (1) *Tái bản* (replication): DNA→DNA; (2) *Phiên mã* (transcription): DNA→ RNA; và (3) *Dịch mã* (translation): RNA→ Protein. Các kênh truyền thông tin này còn được gọi là *Giáo lý trung tâm* (central dogma) của sinh học phân tử. Sau này đến năm 1970, Baltimore và Temin trên cơ sở nghiên cứu cơ chế hoạt động của bộ gene RNA ở virus Sarcoma đã bổ sung thêm kênh RNA→ DNA, gọi là *phiên mã ngược* (reverse transcription). (Hình 5.1)



Hình 5.1 Giáo lý trung tâm của sinh học phân tử (có sửa đổi).

Trong chương này, chúng ta sẽ tìm hiểu các quá trình sinh tổng hợp các nucleotide và tái bản của DNA, kể cả bộ gene RNA của một số virus. Ngoài ra, chương này cũng đề cập đến cả cơ sở phân tử của các quá trình phát sinh đột biến và tái tổ hợp DNA như là thuộc tính thứ hai của DNA.

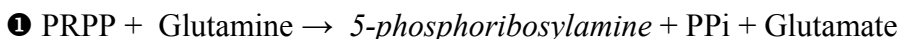
I. Sinh tổng hợp các nucleotide

Các quá trình sinh tổng hợp các ribonucleotide và deoxyribonucleotide có ý nghĩa cực kỳ quan trọng đối với các tế bào, vì chúng là các tiền chất cơ bản cho tổng hợp DNA và RNA. Hơn nữa, các nucleotide đóng vai trò quan trọng trong hầu hết các quá trình trao đổi chất và năng lượng của tế bào. Chẳng hạn, ATP về bản chất là *nucleotide của adenine* - cơ chất toàn năng trong trao đổi năng lượng sinh học, và là thành phần chính của các coenzyme như NAD^+ , NADP^+ , FAD và CoA. Một số nucleotide như cAMP có vai trò nổi bật trong cơ chế điều hoà hoạt động gene và tải nạp tín hiệu (signal transduction) ở các tế bào prokaryote và eukaryote. Các nucleotide khác cũng cần thiết cho các quá trình tổng hợp các hợp chất carbohydrate, lipid, amino acid, nucleic acid và protein.

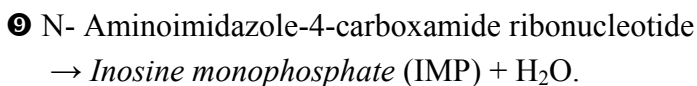
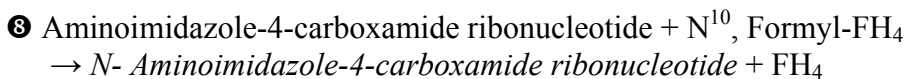
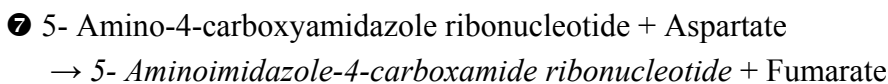
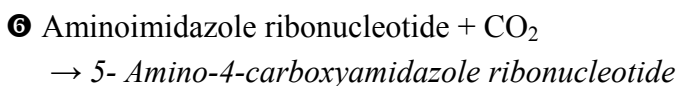
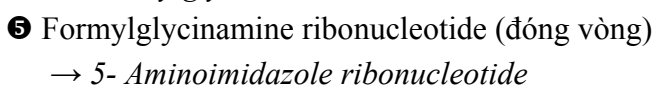
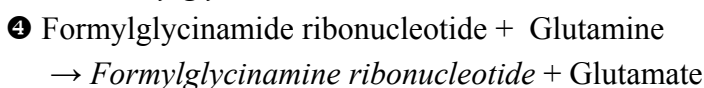
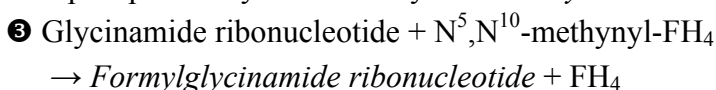
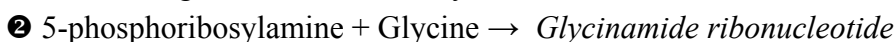
1. Sinh tổng hợp các nucleotide purine

Để tổng hợp mới (*de novo*) purine tạo ra IMP, một nucleotide của hypoxanthine, có tất cả 10 giai đoạn được xúc tác bởi các enzyme mà chất trung gian đều là các ribonucleoside - 5'-monophosphate.

Ở giai đoạn thứ nhất, *5'-phosphoribosine-1-pyrophosphate* (PRPP) chuyển hoá thành trong một phản ứng phụ thuộc vào glutamine và đưa vào N-9 của vòng purine, như sau:



Chất trung gian 5-phosphoribosylamine này lại phản ứng với glycine đưa vào C-4, C-5 và N-7. Các nguyên tử còn lại của vòng purine lần lượt được đưa vào từng cái một, như sau đây:



⑩ IMP có thể chuyển hoá thành *Guanosine monophosphate* (GMP) nhờ sự oxy hoá và amin hoá ở C-2, hoặc thành *Adenosine monophosphate* (AMP) nhờ amine hoá C-6. Trong phản ứng này, aspartate là chất cho nhóm amin và trở thành fumarate. Điều này được minh hoạ như sau:

- * $\text{IMP} + \text{NAD} + \text{H}_2\text{O} \rightarrow \text{Xanthylate (XMP)} + \text{NADH} + \text{H}^+$
- * $\text{XMP} + \text{Glutamine} + \text{ATP} \rightarrow \text{GMP} + \text{Glutamate} + \text{AMP} + \text{PPi}$
- * $\text{IMP} + \text{Aspartate} + \text{GTP} \rightarrow \text{Adenylsuccinate} + \text{GDP} + \text{Pi}$
- * $\text{Adenylsuccinate} \rightarrow \text{AMP} + \text{Fumarate}$

Như vậy việc tổng hợp các purine đòi hỏi phải được cung cấp glutamine như một nguồn nguyên tử nitơ và các chất dẫn xuất của tetrahydrofolate vốn cung cấp các gốc carbon đơn. Hiện tượng amin hoá phụ thuộc glutamine bị ức chế bởi các chất tương tự glutamine và các chất kháng sinh do vi khuẩn *Streptomyces* sinh ra.

2. Sinh tổng hợp các nucleotide pyrimidine

Trong phần này xem xét chủ yếu sáu giai đoạn của quá trình tổng hợp UMP (uridine-5'-monophosphate), được tóm lược như sau:

❶ Tạo thành *carbamyl aspartate* từ glutamine và bicarbonate cùng với 2 phân tử ATP nhờ enzyme carbamyl phosphate synthetase;



❷ Tạo thành *carbamyl aspartate*: Carbamyl phosphate nhường nhóm carbamoyl cho nhóm α -amin của aspartate để tạo thành carbamyl aspartate;

❸ Tạo thành *dihydroorotate* bằng cách loại một phân tử nước;

❹ Tạo thành *orotate* nhờ sự tham gia của coenzyme NAD^+ , giải phóng NADH^+ và H^+ ;

❺ Orotate kết hợp với PRPP tạo ra *Orotidine monophosphate* và giải phóng PPi ;

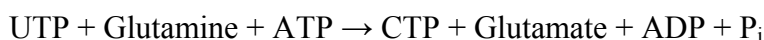
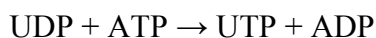
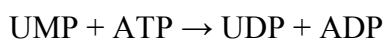
❻ Orotidine monophosphate khử carboxyl tạo thành *Uridine monophosphate* hay uridilic acid (UMP).

* Giống như các nucleotide purine, các nucleotide pyrimidine cũng có thể được tạo thành từ các pyrimidine tự do hoặc từ các nucleoside (con đường trao đổi bổ sung) sau đây:

- $\text{Uracil} + \text{ribose-1-P} \leftrightarrow \text{Uridine} + \text{P}_i$; [nucleotide phosphorylase]
 $\text{Uridine} + \text{ATP} \rightarrow \text{UMP} + \text{ADP}$; [nucleoside kinase]
- $\text{Uracil} + \text{PRPP} \leftrightarrow \text{UMP} + \text{PPi}$; [nucleotide phosphorylase]

Nhờ quá trình phosphoryl hoá, UMP được biến đổi thành UDP và UTP cung cấp cho tổng hợp RNA.

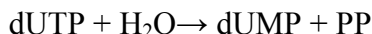
* Các nucleotide cytidine được hình thành trong quá trình amin hoá phụ thuộc vào glutamine và cơ chất UTP, được tạo ra qua hai lần phosphoryl hoá UMP:



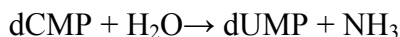
Ở đây UTP nhận nhóm từ NH_3 hoặc glutamine để tạo thành CTP.

* Đối với việc tổng hợp dTMP (deoxythymidine-5'-monophosphate) có thể xảy ra theo một trong hai con đường sau:

(1) dUMP là tiền chất trực tiếp của dTMP, được tạo ra do thủy phân dUTP (phản ứng này ngăn cản việc đưa dUTP vào DNA):

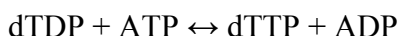
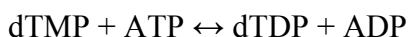
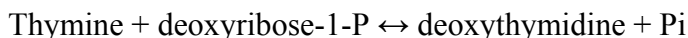


Ngoài ra dUMP cũng có thể được tạo ra bằng cách khử nhóm amin của dCMP theo phản ứng sau:

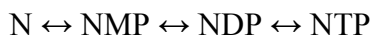


Các dUMP được hình thành theo hai cách nói trên có thể được methyl hoá để trở thành dTMP: $\text{dUMP} \rightarrow \text{dTMP}$.

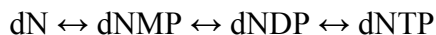
(2) Sự tổng hợp bắt đầu từ thymine và deoxyribose-1-phosphate theo chuỗi phản ứng sau đây:



Trên thực tế có thể xảy ra các phản ứng biến đổi qua lại đối với các ribonucleoside thuộc purine hoặc pyrimidine (ký hiệu: N) như sau:



hoặc các phản ứng của các deoxyribonucleoside thuộc purine hay pyrimidine (dN):



Ngoài ra, các deoxyribonucleotide có thể hình thành trực tiếp từ các ribonucleotide bằng cách khử nguyên tử oxy ở C-2' của đường ribose trong nucleoside diphosphate ($\text{NDP} = \text{ADP}, \text{GDP}, \text{UDP}, \text{CDP}$). Phản ứng này được xúc tác bởi *ribonucleotide reductase*. [Chất khử trung gian là *thioredoxin* có thể cho 2 điện tử cùng với sự oxy hoá 2 phân tử cysteine ($-\text{SH}$)₂ thành một cystine (S-S). Còn thioredoxin oxy hoá thì bị khử bởi NADPH dưới tác dụng của *thioredoxin reductase*.] Từ đây các deoxyribonucleoside diphosphate ($\text{dNDP} = \text{dADP}, \text{dGDP}, \text{dUDP}, \text{dCDP}$) lại được phosphoryl hoá tiếp bởi enzyme *kinase* để tạo thành các dNTP tương ứng dùng làm nguyên liệu cho quá trình sinh tổng hợp DNA (tái bản).

Cần lưu ý rằng, khi các nucleic acid bị phân huỷ sẽ sinh ra các nucleoside monophosphate (NMP) để rồi các NMP này lại bị phân giải tiếp bởi các enzyme nucleotidase, nucleoside phosphorylase và deaminase.

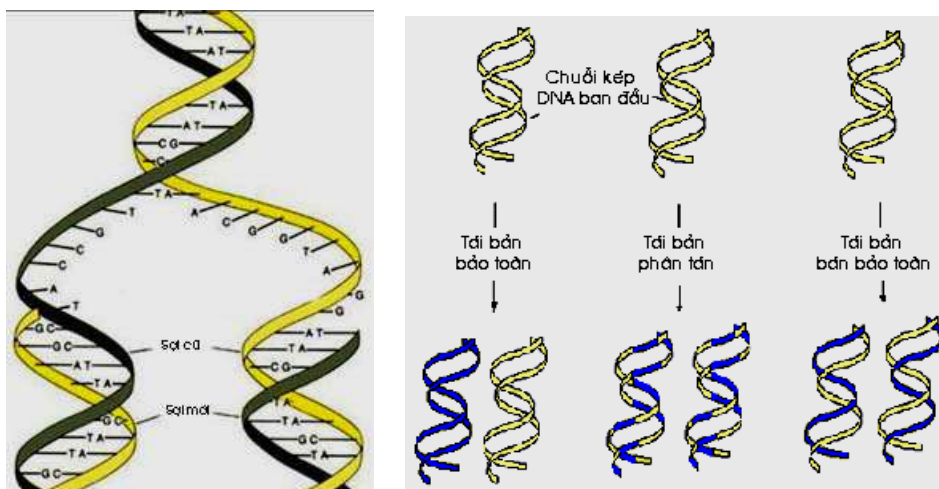
Sau đó các purine và pyrimidine tự do có thể được sử dụng lại để tổng hợp các nucleotide bằng con đường tái sử dụng, trong đó có phản ứng với PRPP do enzyme phosphoribosyl transferase xúc tác.

Một khi các purin bị phân giải đến cùng bởi enzyme xanthine oxydase và tạo thành uric acid quá nhiều sẽ gây ra bệnh Gout. Còn các pyrimidine khi bị phân giải thì được thải ra dưới dạng β -alanine hoặc các dẫn xuất của nó. Chính nhờ các cơ chế kiểm soát kiểu liên hệ ngược (dương tính và âm tính) đối với sự tổng hợp nucleotide mà tế bào có được sự cân đối về hàm lượng bốn loại nucleotide cần cho quá trình sinh tổng hợp DNA.

Những sai sót về mặt di truyền liên quan đến sự trao đổi chất của các purine và pyrimidine là nguyên nhân dẫn tới xuất hiện nhiều căn bệnh như Gout, bệnh Lesh-Nyhan và nhiều bệnh suy giảm miễn dịch khác.

II. Sinh tổng hợp DNA (tái bản)

Tái bản (replication) là một đặc tính quan trọng nhất của vật chất di truyền, nhờ đó sự sống được duy trì liên tục, các loài bảo tồn được tính chất đặc trưng của mình, và con cái thường giống bố mẹ. Vậy DNA và các bộ gene nói chung được tái bản như thế nào?



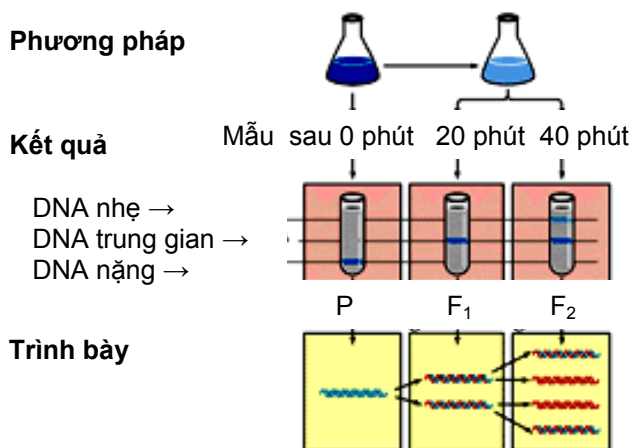
Hình 5.2 Mô hình tái bản bán bảo toàn của DNA do Watson đề xuất (a), và giả định của Luria và Delbruck về ba kiểu tái bản có thể có (b).

Trong khi khám phá ra mô hình cấu trúc DNA, chính Watson đã đưa ra dự đoán chính xác (dựa trên nguyên tắc bổ sung của các cặp base) rằng sự tái bản DNA phải diễn ra theo kiểu *bán bảo toàn* (*semi-conservative*) như ở hình 5.2a. Đến 1956, Salvador Luria và Max Delbruck đề nghị ba kiểu tái bản có thể có (Hình 5.2b): *bán bảo toàn*, *bảo toàn* (*conservative*)

và *phân tán* (dispersive). Tuy nhiên, nhiều thí nghiệm sau đó đã nhanh chóng khẳng định sự tái bản DNA diễn ra theo kiểu bán bảo toàn.

Điện hình là thí nghiệm của Meselson và Stahl năm 1958 trên đối tượng là *E. coli* bằng phương pháp đánh dấu đồng vị phóng xạ N^{15} (nitơ nặng) kết hợp với *ly tâm siêu tốc* (ultra-centrifugation). Đầu tiên cho vi khuẩn này sinh trưởng trên môi trường N^{15} ; rồi đưa trở lại môi trường N^{14} (nitơ nhẹ) và sau một, hai hoặc ba thế hệ đều có lấy các mẫu DNA. Để tách DNA nặng và nhẹ, người ta cho trộn lẫn các mẫu này với Cesium chloride ($CsCl$) trước khi đem ly tâm. Khi ly tâm, DNA có các tỷ trọng nặng (heavy), nhẹ (light) và trung bình (intermediate) sẽ tách thành các vạch tương ứng khác nhau trong ống nghiệm (Hình 5.3).

Các kết quả cho thấy rằng sau một thế hệ, 100% DNA sợi kép có tỷ trọng trung bình, nghĩa là một sợi nặng (từ dạng cha mẹ) và một sợi nhẹ (được tổng hợp mới). Kết quả này cho phép khẳng định sự tái bản xảy ra theo kiểu bán bảo toàn đúng như Watson dự đoán từ trước.



Hình 5.3 Thí nghiệm Meselson-Stahl về sự tái bản bán bảo toàn của DNA.

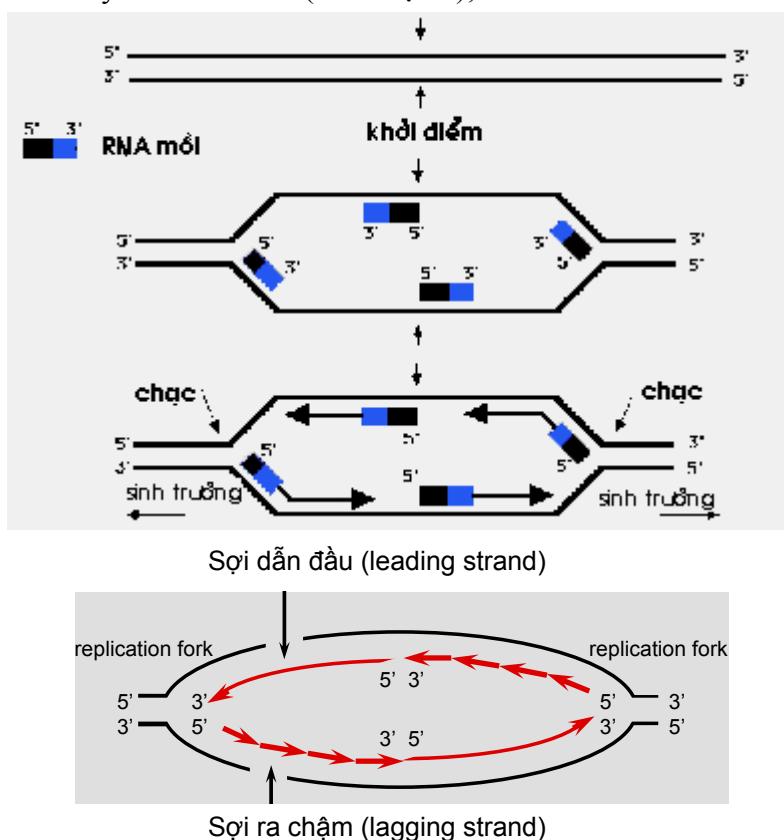
1. Những nguyên tắc và đặc điểm chung của tái bản DNA

(i) Tái bản theo kiểu *bán bảo toàn* và *gián đoạn* (discontinuous);

(ii) Sự tái bản được bắt đầu tại một hoặc nhiều vị trí đặc thù trên phân tử DNA và diễn ra đồng thời theo hai hướng ngược nhau gọi là *khởi điểm tái bản hai hướng* (bidirectional origin of replication). Nghĩa là, từ khởi điểm này DNA sợi kép mở xoắn thành hình vòng sinh trưởng theo hai hướng đối lập nhau tạo ra hai *chạc tái bản* (replication fork). Mỗi khởi điểm tái bản cùng với hai chạc như vậy gọi là một *đơn vị tái bản* (replicating unit, hay *replicon*) được minh họa ở Hình 5.4. Nói chung, đối

với các DNA mạch vòng (bộ gene một số virus, vi khuẩn, các plasmid và các DNA bào quan của eukaryote), mỗi phân tử chỉ có một khởi điểm tái bản; trong khi đó mỗi DNA trong nhiễm sắc thể eukaryote có nhiều khởi điểm tái bản hoạt động theo một trình tự đặc thù (Hình 5.5).

(iii) Quá trình tái bản DNA phụ thuộc vào một hệ thống gồm nhiều protein và enzyme khác nhau (xem mục 2);



Hình 5.4 Một khởi điểm với hai chạc tái bản sinh trưởng theo hai hướng đối lập nhau. Để tổng hợp DNA, các đoạn mồi (RNA primer) phải được tổng hợp trước. Ở hình dưới cùng cho thấy phương thức tái bản nửa gián đoạn xảy ra đồng thời ở cả hai chạc tái bản (replication fork) của mỗi đơn vị tái bản.

(iv) Tại mỗi chạc tái bản, trước tiên xảy ra sự tổng hợp các *đoạn mồi* RNA (primer) bởi vì các DNA polymerase tự nó không thể bắt đầu tổng hợp mới được. Mặt khác, do hai sợi đơn của mỗi chạc phân cực ngược chiều nhau trong khi các enzyme DNA polymerase và RNA polymerase chỉ xúc tác theo chiều $5' \rightarrow 3'$, cho nên phương thức tái bản DNA diễn ra trên hai sợi khuôn là không giống nhau: một sợi liên tục hay còn gọi là *sợi dẫn đầu* (leading strand) và một sợi không liên tục hay còn gọi là *sợi ra*

chậm (lagging strand). Kiểu tái bản như thế gọi là *tái bản nửa gián đoạn* (semi-discontinuous). Sự tổng hợp không liên tục dưới dạng các đoạn có độ dài 1.000-2.000 nucleotide do R. Okazaki phát hiện đầu tiên năm 1969, nên còn gọi là các *đoạn Okazaki* (Okazaki fragments). Sau đó, các đoạn mới sẽ được cắt bỏ và chỗ trống được thay thế bằng DNA, và các đoạn Okazaki được nối lại với nhau bởi enzyme DNA ligase.

2. Các enzyme tham gia tái bản DNA

Mặc dù mỗi nhóm sinh vật có một hệ thống các enzyme và protein tái bản riêng, song chúng có các enzyme với vai trò chung nhất sau đây:

Bảng 5.1 Đặc tính của các DNA polymerase ở *E. coli* và người

DNA polymerase <i>E. coli</i>	Pol I	Pol II	Pol III		
Polymerase 5'→3'	có	có	có		
Exonuclease 3'→5'	có	có	có		
Exonuclease 5'→3'	có	không	không		
DNA polymerase người	α	β	γ	δ	ϵ
Định vị	nhân	nhân	ty thể	nhân	nhân
Tái bản	có	không	có	có	có
Sửa chữa	không	có	không	có	có
Chức năng					
Polymerase 5'→3'	có	có	có	có	có
Exonuclease 3'→5'	không	không	có	có	có
Exonuclease 5'→3'	không	không	không	không	không
Primase	có	không	không	không	không
Kết hợp với PCNA*	không	không	không	có	có
Mẫu trượt (Processivity)	thấp			cao	
Tổng hợp sợi	ra chậm	sửa	cả hai	dẫn đầu	ra chậm

(1) Protein nhận biết và bám khởi điểm để từ đó hình thành nên "phức hợp mở" (open complex). Ở *E. coli*, đó là protein *dnaA*.

(2) *DNA gyrase* mở cuộn DNA siêu xoắn trước mỗi chạc tái bản.

(3) *Helicase* (ở *E. coli*, đó là protein *dnaB*) tháo xoắn DNA sợi kép tại mỗi chạc tạo thành các vùng sợi đơn. Ở *E. coli*, nó cũng gọi là protein *dnaB* hay protein *rep*.

(4) *Protein SSB* (single strand binding protein) bám vào các vùng DNA sợi đơn do helicase tách ra, giữ cho tạm thời không dính trở lại và

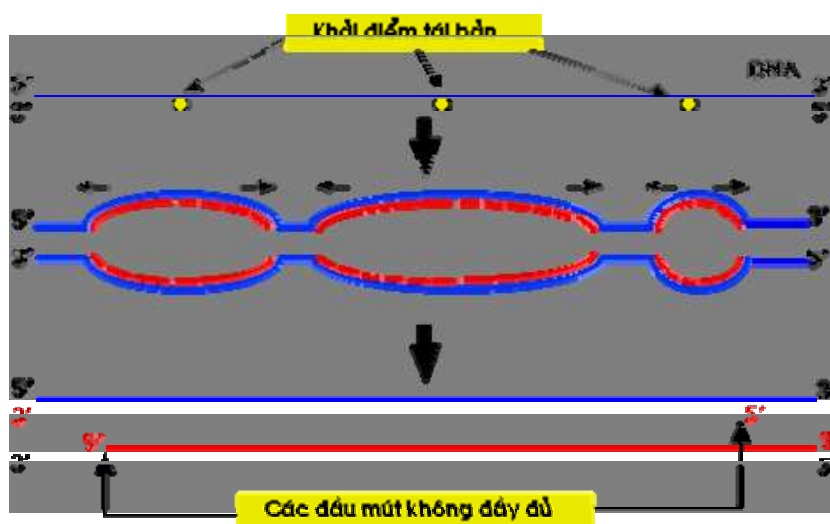
nhờ đó mỗi sợi đơn mới có thể làm khuôn (template) cho tái bản.

(5) *Primase* tổng hợp mồi. Ở *E. coli*, nó còn được gọi là protein *dnaG*.

(6) Các DNA polymerase xúc tác chính cho việc tổng hợp DNA mới nhờ có hoạt tính xúc tác: polymerase $5' \rightarrow 3'$, một số còn có hoạt tính đọc sửa: exonuclease $3' \rightarrow 5'$. Ở *E. coli*, đó là *DNA polymerase III*.

(7) DNA polymerase vừa cắt bỏ dần đoạn mồi đi trước nhờ hoạt tính cắt bỏ exonuclease $5' \rightarrow 3'$, vừa kéo dài đoạn Okazaki theo sau lấp chỗ trống nhờ hoạt tính polymerase $5' \rightarrow 3'$. Ở *E. coli*, đó là *DNA polymerase I*.

(8) *DNA ligase* hàn liền khe hở giữa các đoạn Okazaki (DNA mới tổng hợp) bằng cách hình thành liên kết $3',5'$ -phosphodiester.



Hình 5.5 Nhiều khởi điểm tái bản trên một nhiễm sắc thể eukaryote. Ở đây cũng cho thấy các đầu mút (telomere) 5' không được tổng hợp đầy đủ.

Ở đây chúng ta cần lưu ý thêm một số điểm khác nhau về các enzyme tái bản giữa hai nhóm prokaryote và eukaryote (xem Bảng 5.1):

(i) Ở *E. coli*, cả ba loại protein *dnaB*, *dnaC* và *dnaG* hợp thành một phức hợp có tên là *primasome* (thể mở đầu); trong đó protein *dnaC* bám protein *dnaB*. Bảng 5.1 mô tả các đặc tính của các DNA polymerase ở *E. coli* và người đại diện cho các nhóm tương ứng, prokaryote và eukaryote.

(ii) Tất cả các DNA polymerase đều cần có mồi với một nhóm $3'$ -OH tự do; chúng xúc tác tổng hợp chuỗi theo một hướng $5' \rightarrow 3'$ và chỉ một số enzyme này có *hoạt tính đọc sửa* (proofreading activity) $3' \rightarrow 5'$.

(iii) Khác với *E. coli*, các tế bào người có năm loại DNA polymerase, trong đó ba loại chịu trách nhiệm tái bản DNA nhân là các DNA

polymerase *alpha*, *delta* và *epsilon*. Polymerase δ chịu trách nhiệm chính cho tổng hợp *sợi dẫn đầu* (leading strand) và thậm chí cả *sợi ra chậm* (lagging strand). Polymerase α kết hợp với một RNA primase và được coi là có các hoạt tính cho tổng hợp một đoạn mồi RNA ngắn (ở *E. coli* là 10-12 base). Có điều không chắc chắn lắm khi nói về chức năng của polymerase α trên sợi ra chậm. Vì dường như nó thiếu mất hoạt tính đọc sửa, nên không thể tổng hợp DNA với độ chính xác cao và như vậy có thể nó không phải là polymerase chính trong tổng hợp sợi ra chậm. Polymerase ϵ có thể hoạt động như DNA polymerase I của vi khuẩn. Loại *kháng nguyên nhân làm tăng sinh tế bào* (proliferating cell nuclear antigen = PCNA) có vai trò trong cả tái bản lẫn sửa chữa. Một trong các chức năng của nó là dùng làm *nhân tố trượt* cho DNA polymerase δ và ϵ . PCNA giữ DNA polymerase với sợi khuôn để tổng hợp DNA nhanh.

(iv) Ngoài ra còn có một số enzyme và protein đặc thù tham gia vào khâu kết thúc tái bản, tổng hợp các *đầu mút* (telomere) hoặc tham gia cắt nối trong quá trình tái tổ hợp (xem ở phần sau).

3. Cơ chế tái bản DNA

Trong phần này, chúng ta tìm hiểu chủ yếu cơ chế tái bản ở vi khuẩn *E. coli* và nêu một số vấn đề liên quan eukaryote, đại diện là DNA người.

3.1. Giai đoạn khởi đầu của sự tái bản (initiation of replication)

Đối với nhiễm sắc thể *E. coli*, sự tái bản bắt đầu tại một khởi điểm đặc thù duy nhất gọi là *Ori* (Hình 5.4). Trong khi đó, mỗi nhiễm sắc thể eukaryote có nhiều khởi điểm tái bản hai hướng như thế (Hình 5.5). Quá trình diễn biến tại khởi điểm cho đến lúc hình thành hai chạc có thể tóm tắt (từ Kelman và O'Donnell 1994; xem các Hình 5.6 và 5.7) như sau:

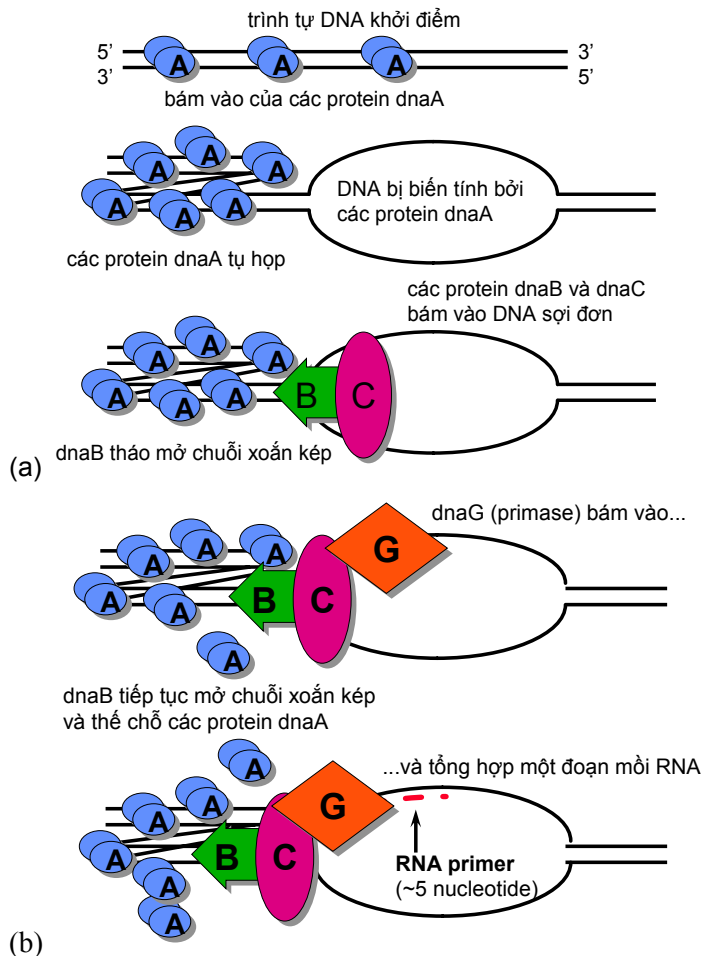
(1) Các protein bám khởi điểm *dnaA* được tổng hợp để bám *Ori* tạo ra *cấu trúc nucleoprotein chuyên hoá của khởi điểm*;

(2) Cấu trúc này mở xoắn vùng DNA giàu AT để hình thành "*phức hợp mở*" (open complex); và

(3) Hai phân tử helicase *dnaB* chui vào phức hợp mở làm mở xoắn khởi điểm theo cả hai hướng, tạo thành hai *chạc tái bản* (replication fork). Khi cả hai sợi đơn của mỗi chạc được tách ra thì các protein SSB bám vào. Nhờ vậy các sợi đơn được dùng làm khuôn hiệu quả cho các enzyme tái bản hoạt động.

Khởi đầu trong tái bản DNA là sự tổng hợp một đoạn mồi ngắn bởi primase (xem Hình 5.4), mà ở *E. coli* là phức hợp *primasome*. Kích thước đoạn mồi ở các sinh vật thường không vượt quá 12 nucleotide. Ở *E. coli*,

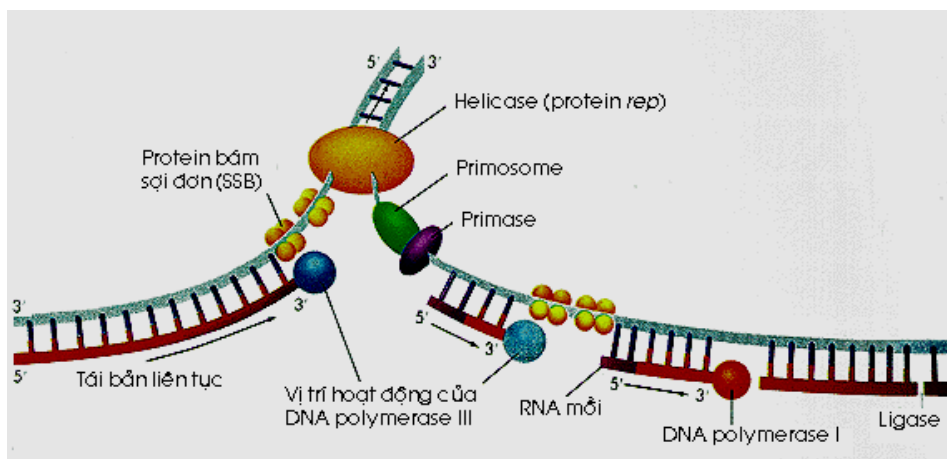
theo kết quả nghiên cứu của bà Okazaki và cs (1985), các đoạn mỗi dài 10-12 nucleotide. [Về số liệu này, một số cho rằng khoảng 5 nucleotide.]



Hình 5.6 Tiến trình khởi đầu tái bản DNA ở *E. coli*. Mở đầu là sự bám vào *ori* của các protein dna A (a) cho đến khi tổng hợp đoạn mồi RNA đầu tiên bởi primasome - phức hợp mở đầu gồm các protein "dnaG + dnaC + dnaB" (b).

3.2. Giai đoạn kéo dài (elongation)

Một khi primasome tổng hợp xong một mồi, đó là lúc sự kéo dài chuỗi DNA mới bắt đầu. Ở *E. coli*, enzyme chủ chốt cho quá trình này là DNA polymerase III hoàn chỉnh (*holoenzyme*), một phức hợp gồm mười polypeptide khác nhau. Mỗi phân tử cho một sợi khuôn. Sự kéo dài trong suốt quá trình tổng hợp DNA ở *E. coli* rõ ràng là cần tới một *replisome* (thể tái bản) do kết hợp giữa primasome và DNA polymerase III hoàn chỉnh. Tốc độ tổng hợp trung bình là 1.000 nucleotide mỗi giây.



Hình 5.7 Cơ chế tái bản nửa gián đoạn ở một chạc của DNA *E. coli*.

Sự tái bản DNA kiểu *nửa gián đoạn* (semi-discontinuous) xảy ra tại mỗi chạc tái bản của DNA *E. coli* như sau (Hình 5.7):

- Trên sợi khuôn dẫn đầu ($3' \rightarrow 5'$): Trước tiên, enzyme primase hay primosome chỉ tổng hợp một đoạn mồi RNA với đầu 3'-OH tự do. Sau đó, enzyme hoàn chỉnh DNA polymerase III (replisome) bắt đầu kéo dài chuỗi DNA mới sinh trưởng theo chiều $5' \rightarrow 3'$ một cách liên tục.

- Trên sợi khuôn ra chậm ($5' \rightarrow 3'$): Sự kéo dài diễn ra không liên tục dưới dạng các đoạn Okazaki. Kích thước trung bình mỗi đoạn Okazaki ở *E. coli* là 1.000 - 2.000 nucleotide (ở các sinh vật nói chung là 100-1.000 nucleotide). Quá trình này có tính chu kỳ, đòi hỏi phải được "mồi hóa" nhiều lần, với sự tham gia lần lượt của bốn enzyme sau đây:

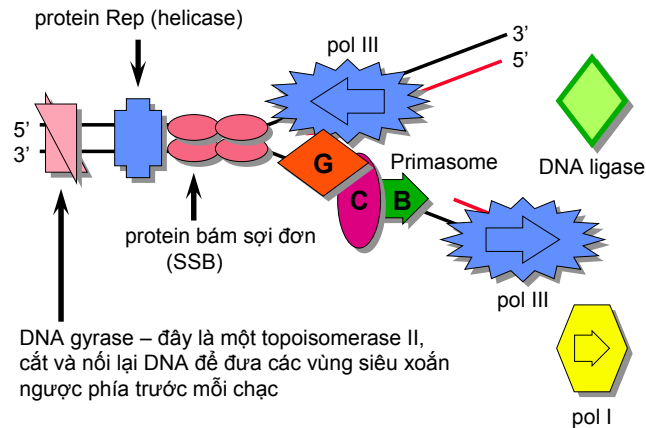
- (i) primase tổng hợp một đoạn mồi RNA bổ sung với DNA;
- (ii) DNA polymerase III hoàn chỉnh kéo dài đoạn Okazaki;
- (iii) DNA polymerase I vừa cắt bỏ dần từng nucleotide của đoạn mồi vừa lấp khoảng trống bằng cách kéo dài dần đoạn Okazaki theo sau nó;
- (iv) DNA ligase hàn liền khe hở còn lại giữa hai đoạn Okazaki kề sát nhau bằng một liên kết 3',5'-phosphodiester.

Quá trình tổng hợp Okazaki trên sợi khuôn ra chậm diễn ra theo chu kỳ hoạt động của bốn enzyme nói trên, và nó diễn ra đồng thời với quá trình tổng hợp liên tục trên sợi khuôn dẫn đầu của mỗi chạc tái bản.

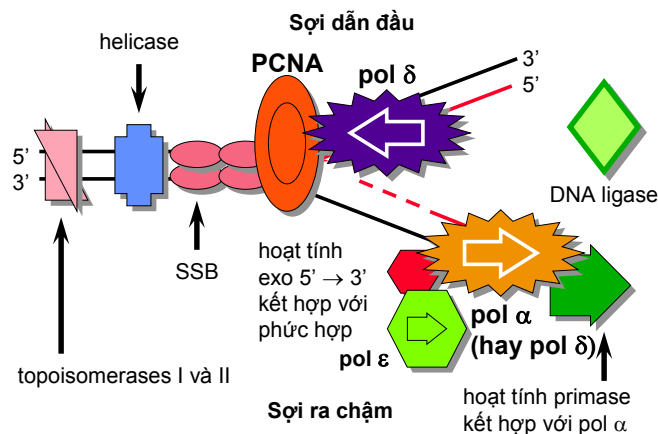
Ở eukaryote, vai trò của các enzyme ở giai đoạn kéo dài như đã được đề cập ở trước: DNA polymerase δ chịu trách nhiệm tổng hợp sợi dẫn đầu, còn DNA polymerase α (và thậm chí cả pol δ) tổng hợp sợi ra chậm.

Hình 5.8 cho thấy sự hoạt động của các protein và enzyme khác nhau

tại mỗi chạc tái bản của *E. coli* và của người.



Hình 5.8A Các enzyme và protein tại mỗi chạc tái bản DNA *E. coli*.



Hình 5.8B Các enzyme và protein tại mỗi chạc tái bản DNA người.

3.3. Giai đoạn kết thúc (termination)

Do đặc điểm cấu trúc nhiễm sắc thể ở hai nhóm prokaryote và eukaryote là hoàn toàn khác nhau, đặc biệt là cấu trúc các đầu mút nhiễm sắc thể eukaryote được phát hiện gần đây (hình 5.9), nên cơ chế kết thúc tái bản của chúng cũng khác nhau.

3.3.1. Vấn đề kết thúc tái bản ở *E. coli*

Để hoàn thành công việc tái bản DNA, tế bào phải lấp đầy các khoảng trống do các RNA mồi bị cắt bỏ để lại. Đối với các DNA mạch vòng như của vi khuẩn chẳng hạn, không có vấn đề lấp tất cả các khoảng trống bởi vì bao giờ cũng có đầu 3' DNA khác nằm phía trước dùng làm mồi. Thật

vậy, cả hai chạc tái bản được bắt đầu từ một khởi điểm duy nhất (*ori*), và di chuyển hầu như cùng tốc độ, theo hai hướng đối lập nhau xung quanh nhiễm sắc thể mạch vòng cho tới khi chúng gặp nhau tại một điểm kết thúc chung đối xứng với *ori*. Theo Bastia và cs (1997) cũng như nhiều tác giả khác, đây là vùng chứa các trình tự đặc thù gọi là các *điểm kết thúc tái bản* (replication termini). Và tại các trình tự đặc thù này có các *protein kết thúc tái bản* (replication terminator protein = RTP) bám vào và các phức hợp protein-DNA này ngăn cản sự di chuyển của các chạc tái bản theo cách phân cực hoặc định hướng đặc thù. Sự ngừng lại của các chạc tái bản tại vùng kết thúc tạo nên bước đầu tiên trong quá trình hoàn thành một vòng tái bản và sau đó, tách rời hai nhiễm sắc thể con một cách có trật tự nhờ xúc tác của *topoisomerase IV*.

3.3.2. Vấn đề kết thúc tái bản ở eukaryote

Như chúng ta đều biết, mỗi nhiễm sắc thể eukaryote chứa một phân tử DNA sợi kép mạch thẳng kết hợp với nhiều loại protein, có các đầu mút đặc trưng gọi là *telomere*. Một khi đoạn mỗi đầu tiên trên mỗi sợi được loại bỏ thì nó không có cách nào để bù đắp lại khoảng trống đó, bởi vì DNA không thể nào nối rộng theo chiều 3'→5', và cũng chẳng có đầu 3' nằm phía trước như trong các DNA mạch vòng. Nếu quả thực như vậy thì các sợi DNA sẽ ngắn bớt đi sau mỗi lần tái bản. Vậy các tế bào eukaryote giải quyết vấn đề này như thế nào?

Các kết quả nghiên cứu đầu tiên của Elizabeth Blackburn và các đồng sự của bà đã giải đáp cho vấn đề này (Hình 5.10). Các telomere không chứa các gene; thay vì thế chúng có cấu trúc đơn giản gồm những trình tự ngắn (6-8 bp) lặp lại nối tiếp cả ngàn lần và đặc thù cho từng loài.

Chẳng hạn, ở *Tetrahymena*, một nhóm động vật nguyên sinh có lông tơ, trình tự này là (TTGGGG)_n; ở *Caenorhabditis*: (CCCTCCC)_n; ở bọ *Oxytrichia* thuộc Euplotes: (CCCCAAA)_n; v.v. Ở người và các động vật có vú là 5'-TTAGGG-3' được lặp lại khoảng 1.000-2.000 lần (theo Yakoob và cs 1999, khoảng biến thiên phát hiện được trong các tế bào ở các giai đoạn khác nhau là 150-2.000 lần).

Các đoạn lặp này được gắn thêm vào đầu 3' của các sợi DNA, không phải bằng tái bản bán bảo toàn, mà bằng một enzyme có tên là *telomerase*. Nếu như telomerase không cho phép một cơ chế tái bản bán bảo toàn, thì nó không thể sử dụng một sợi DNA làm khuôn cho sợi kia. Blackburn cho rằng đặc tính này là do một RNA nhỏ có mặt trong chính telomerase.

Thật vậy, telomerase là một ribonucleoprotein có bản chất của một enzyme *phiên mã ngược* (reverse transcriptase) ở chỗ, tổng hợp DNA từ

một khuôn RNA. Chẳng hạn, telomerase của *Tetrahymena* có một RNA dài 160 nucleotide có chứa trình tự 3'-CAACCCCAA-5' làm khuôn cho tổng hợp các đoạn lặp 5'-TTGGGG-3' trong các telomere của nó (Hình 5.10). Ở người, phân tử RNA trong telomerase dài khoảng 450 nucleotide có chứa trình tự 3'-AAUCCC-5' (nằm trong đoạn ứng với vị trí 46-56) làm khuôn cho sự tổng hợp các đoạn lặp telomere 5'-TTAGGG-3' trên các sợi đơn có đầu mút 3'. Sau đó, ở sợi đối diện, DNA polymerase có thể hoàn thành việc tổng hợp "các đầu mút không đầy đủ" (incomplete ends) trên sợi đối diện sau khi đã được mồi hóa bên trong các telomere đó; và cuối cùng đoạn mồi sẽ bị cắt bỏ.

Các nghiên cứu gần đây cho thấy rằng: enzyme telomerase chỉ có mặt trong các *tế bào mầm* (germ cells), kể cả các *tế bào gốc của phôi* (embryonic stem cells); các *tế bào ung thư* (cancer cells); và các eukaryote đơn bào như *Tetrahymena thermophila*. Trong khi đó, các tế bào soma bình thường của động vật có vú không thấy có telomerase.

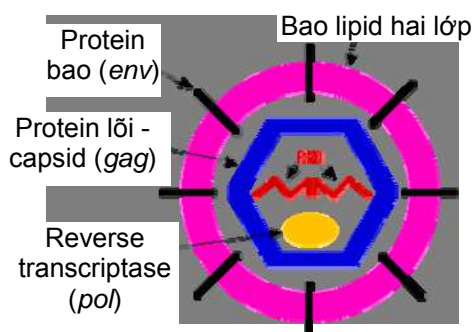
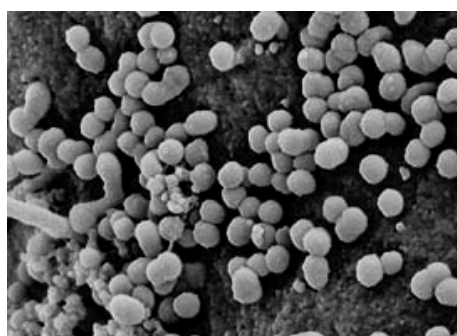
Điều đó cho phép lý giải tại sao các tế bào mầm cũng như các tế bào ung thư có khả năng phân chia gần như là vô hạn; hay nói cách khác, telomerase và sự duy trì chiều dài telomere chính là chìa khóa cho sự bất tử của tế bào. Ngược lại, hậu quả của sự vắng bóng telomerase trong các tế bào soma (do gene mã hóa nó bất hoạt) là ở chỗ: cứ sau mỗi lần *nguyên phân* (mitosis), tất cả 92 telomere của các tế bào soma người đồng loạt mỗi cái bị mất đi chừng 100 cặp base, nghĩa là khoảng 16 đoạn lặp TTAGGG. Theo lý thuyết, với tốc độ này, sau 125 lần nguyên phân các telomere sẽ biến mất hoàn toàn. Trên thực tế, các thí nghiệm khác nhau kể từ khám phá đầu tiên của Leonard Hayflick vào đầu thập niên 1960 cho đến nay (Greider và Blackburn 1996; Hayflick 1997; Shay và cs 2004; ...) đã xác nhận rằng: chỉ sau khoảng 30-50 lần nguyên phân, các tế bào soma mất hẳn khả năng phân chia và bước vào giai đoạn lão hóa (senescence) và chết tự nhiên. Cái đó được gọi là *giới hạn Hayflick* (Hayflick limit).

Chính quá trình rút ngắn telomere theo thời gian (nghĩa là số lần nguyên phân) như thế khiến người ta liên tưởng tới sự tồn tại của cái gọi là "*đồng hồ di truyền cho sự lão hóa*" (genetic clock for aging) và bắt đầu tiến hành các nghiên cứu về "*hiệu ứng vị trí telomere*" (telomere position effect = TPE) (Borek 2002). Đó là lý do tại sao các tế bào soma có số lần phân chia hữu hạn trước khi chúng chết và cũng là lý do tại sao tất cả chúng ta cũng như mọi sinh vật đều phải già lão và chết đi. Như thế, cái chết tất yếu trên phương diện sinh học, cái chết tự nhiên của thân xác đã được lý giải khá rõ ràng và đầy đủ. Từ các nghiên cứu này cũng đã mở ra triển vọng to lớn cho *liệu pháp ung thư* (cancer therapy) cũng như vấn đề

4. Tái bản của các bộ gene RNA

4.1. Đặc điểm tái bản của các bộ gene RNA virus

Ở các virus có bộ gene RNA, phương thức tái bản của chúng rõ ràng là khác với các hệ thống tái bản DNA đã biết. Trước hết, các enzyme tổng hợp RNA không cần có mồi (primer), vì vậy không có cơ hội cho việc *đọc sửa* (proofreading). Ngoài ra, về mặt di truyền RNA là một phân tử kém bền vững so với DNA bởi vì nhóm hydroxyl có mặt ở nguyên tử C_{2'} của gốc đường cho phép cắt đứt (thủy phân) liên kết phosphodiester giữa hai ribonucleotide và tạo thành dạng phosphodiester vòng giữa các nhóm hydroxyl của C_{2'} và C_{3'} trong cùng một gốc đường; sau đó cấu trúc vòng này mở ra và để lại nhóm phosphate ở vị trí C_{3'}. Sự kết hợp của hai đặc điểm trên là nguyên nhân làm hạn chế kích thước các bộ gene RNA. Các bộ gene này không thể sinh trưởng quá lớn, hoặc không thể tránh khỏi tỷ lệ sai sót cao trong quá trình tái bản (10^{-3} - 10^{-4}). Vì vậy, trên thực tế, bộ gene RNA lớn nhất được biết là một RNA sợi đơn với 29.600 base ở các coronavirus (virus gây dịch sốt viêm phổi cấp, SARS, với chủng gây bệnh phổ biến H5N1 thuộc họ virus này được phát hiện hồi tháng 3/2002 tại một số nước châu Á, Canada... hiện có nguy cơ gây nên đại dịch toàn cầu). Đó cũng là lý do tại sao các virus RNA cho nhiều biến thể khác nhau đến như vậy, mặc dù kích thước bộ gene không lớn. Chính điều này mà các virus RNA có thể trở thành mối hiểm họa thực sự bởi vì chúng có thể tiến hóa nhanh để xâm nhập vào các hệ thống miễn dịch của vật chủ. Chẳng hạn, các bộ gene của HIV-1 và HIV-2 hoặc các virus gây bệnh SARS có nhiều biến thể gây khó khăn cho việc bào chế các vaccine và thuốc chống lại tất cả các biến thể của chúng.



Hình 5.11 Vi ảnh điện tử phóng đại phần bề mặt của HIV (trái) và sơ đồ minh họa cấu trúc của một retrovirus điển hình: HIV.

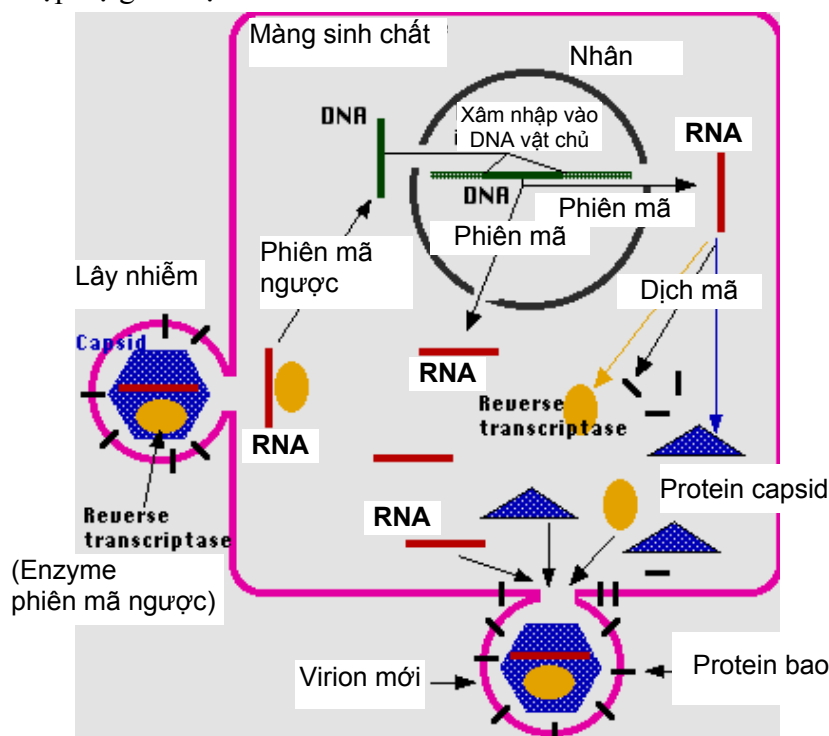
4.2. Tái bản của bộ gene RNA (RNA → RNA)

Kiểu truyền thông tin trực tiếp từ RNA sang RNA xảy ra trong các tế

bào lây nhiễm virus RNA (chẳng hạn virus đốm thuốc lá và nhiều virus thực vật khác, kể cả các phage RNA như MS2...) Bộ gene RNA của các virus này có mang gen mã hoá enzyme tái bản đặc thù. Sau khi được tổng hợp trong tế bào chủ, enzyme này sử dụng bản thân RNA của virus làm khuôn để tổng hợp các phân tử RNA bổ sung. Đến lượt mình các RNA lại làm khuôn để tổng hợp trở lại các phân tử RNA của các virus thế hệ con.

4.3. Phiên mã ngược (RNA → cDNA)

Phiên mã ngược (reverse transcription) là kiểu truyền thông tin từ RNA sang DNA, chỉ xảy ra trong các tế bào động vật và người bị lây nhiễm bởi một số virus mang một sợi RNA có khả năng gây khối u hoặc hai sợi RNA như trường hợp HIV chẳng hạn (Hình 5.11 và 5.12). Các virus này được gọi là các *retrovirus*. Trên mỗi sợi RNA lõi của các virus này có mang một *enzyme phiên mã ngược* (reverse transcriptase, viết tắt là: RTase) và enzyme *integrase* giúp (cDNA sợi kép có bản chất virus) xâm nhập bộ gene vật chủ.



Hình 5.12 Các hoạt động sống của một retrovirus (HIV) trong tế bào người.

Khi xâm nhập vào tế bào chủ, enzyme này sử dụng RNA của virus làm khuôn để tổng hợp sợi *DNA bổ sung* (complementary DNA = cDNA). Sau

đó, sợi cDNA này có thể làm khuôn để tổng hợp trở lại bộ gene của virus (cDNA→RNA), hoặc tổng hợp ra sợi DNA thứ hai bổ sung với nó (cDNA→DNA) như trong trường hợp virus gây khối u mà kết quả là tạo ra một cDNA sợi kép. [Vì vậy, sau này người ta tinh chiết các enzyme phiên mã ngược để phục vụ cho kỹ thuật tạo dòng cDNA tái tổ hợp; xem mục 5.2 bên dưới]. Phân tử DNA sợi kép được tổng hợp trước tiên trong quá trình lây nhiễm có thể xen vào DNA của vật chủ (nhờ enzyme integrase). Trạng thái tồn tại của DNA sợi kép này trong bộ gene vật chủ được gọi là *DNA tiền virus* (DNA provirus). Vì vậy, provirus được truyền lại cho các tế bào con thông qua sự tái bản của DNA vật chủ, nghĩa là các tế bào con cháu của vật chủ cũng bị chuyển sang tình trạng có mầm bệnh ung thư. Các tế bào ung thư này mất khả năng kiểm soát sự sinh trưởng và phân chia điển hình của tế bào bình thường; chúng tăng sinh rất nhanh và tạo ra *khối u* (tumor). Đó chính là cơ chế gây ung thư bởi virus.

5. Ứng dụng của các enzyme tái bản trong kỹ thuật sinh học phân tử

5.1. Khuyếch đại gene - phương pháp PCR

Những tiến bộ của công nghệ sinh học ngày nay có thể cho một DNA của vi sinh vật "sinh trưởng" thậm chí ngay cả khi sinh vật đó khó nuôi cấy. Nhờ đó có đủ các mẫu DNA cho việc xác định hầu như bất kỳ vi sinh vật nào có thể thu được từ một mẫu tiêu bản thậm chí không phải qua nuôi cấy sinh vật đó. Đó chính là nhờ sự phát triển của phương pháp *khuyếch đại gene* hay *PCR* (Gene amplification - Polymerase Chain Reaction).

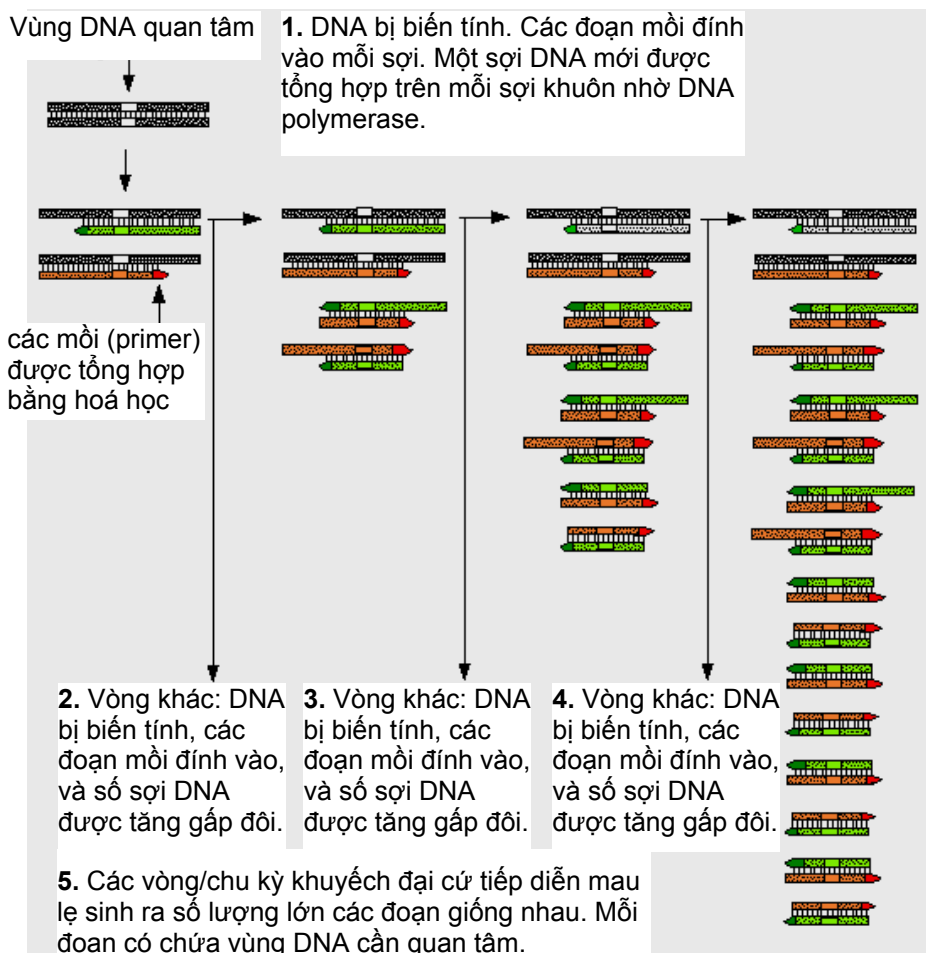
PCR là một kỹ thuật cho phép khuyếch đại nhanh một mẫu DNA cụ thể trong ống nghiệm (hơn là trong các tế bào sống như là *E. coli*). Với quy trình này người ta có thể tạo ra vô số bản sao của một phân tử DNA đơn. Quy trình "tạo dòng *in vitro*" này được tóm tắt như sau (hình 5.13):

- Để thực hiện một PCR, cần phải biết ít nhất một đoạn trình tự của phân tử DNA cần quan tâm (ví dụ một mẫu máu).
- Sau đó phải tổng hợp các *đoạn mồi* (primer), tức các oligonucleotide ngắn (chứa khoảng hai chục nucleotide) mà nó bổ sung chính xác với trình tự ở đầu 3' của mỗi một sợi của DNA cần khuyếch đại.
- Mẫu DNA được đun nóng để tách các sợi đơn (biến tính) và trộn lẫn với các đoạn mồi.
- Nếu như các đoạn mồi tìm thấy các trình tự bổ sung trong DNA, chúng sẽ kết hợp vào các sợi đó.
- Sự tổng hợp bắt đầu (bao giờ cũng theo chiều 5' → 3') bằng cách sử dụng sợi gốc làm khuôn.

- Hỗn hợp phản ứng phải chứa tất cả bốn loại deoxynucleotide triphosphate (dATP, dCTP, dGTP và dTTP) và một DNA polymerase chịu nhiệt (ví dụ, *Taq polymerase* được chiết xuất từ vi khuẩn *Thermus aquaticus* - sống ở suối nước nóng - vẫn giữ được hoạt tính trong khi gây biến tính ở nhiệt độ cao).

- Sự trùng hợp tiếp diễn chừng nào mỗi sợi đơn được tổng hợp mới còn chứa đủ vị trí được nhận biết bởi đoạn mồi khác.

- Lúc này ta có hai phân tử DNA giống hệt phân tử ban đầu.
- Bây giờ ta lấy hai phân tử này cho biến tính và lặp lại quá trình đó.
- Sau mỗi chu kỳ số phân tử DNA lại tăng gấp đôi.



Hình 5.13 Sơ đồ minh họa quy trình kỹ thuật PCR do Kary Mullis đề xuất.

Rõ ràng, phản ứng PCR tái bản DNA theo một cách thức tương tự với

sự tái bản xảy ra trong tế bào, ở chỗ:

(1) DNA sợi kép được tách thành các sợi đơn. Tuy nhiên, trong PCR, thì nhiệt chứ không phải enzyme được dùng để làm biến tính DNA.

(2) Các mẫu nucleotide ngắn được dùng làm mồi để khởi đầu tái bản. Các đoạn mồi của PCR được tổng hợp từ các deoxynucleotide. Chúng bổ sung với các trình tự ở hai đầu của gene quan tâm. Một đoạn mồi kết dính vào đầu 3' của mỗi một sợi DNA bổ sung.

(3) DNA polymerase tiếp tục kéo dài sợi DNA mới từ đầu 3'-OH của các đoạn mồi. Sản phẩm của sự tái bản bán bảo toàn này là hai sợi con.

Ưu thế của PCR là ở chỗ, cứ sau mỗi vòng hay chu kỳ tái bản thì thể hệ DNA sợi kép con lại bị biến tính, nếu tiếp tục bổ sung các đoạn mồi thì quá trình này lại tiếp diễn. Nhờ sử dụng các thiết bị tự động, mỗi chu kỳ tái bản có thể hoàn thành chưa đầy 5 phút. Bằng cách đó, từ một mẫu DNA đơn có thể khuếch đại lên 1.000 lần sau 10 chu kỳ ($2^{10} = 1.024 \approx 10^3$). Và sau 30 chu kỳ, từ một phân tử DNA ban đầu được khuếch đại lên hơn một tỷ bản sao ($2^{30} = 1,02 \times 10^9$).

Ứng dụng của PCR: Trên thực tế, PCR được dùng để tạo các mẫu dò (probe) nucleic acid cho việc xác định các vi sinh vật. Chẳng hạn, nó có thể phát hiện số lượng nhỏ các tác nhân gây bệnh trong các mẫu lâm sàng và qua đó chẩn đoán các bệnh gây ra bởi các tác nhân khó nuôi cấy, như virus HIV/AIDS, *Mycobacterium tuberculosis* sinh trưởng chậm v.v. PCR cũng đã được dùng để khuếch đại các gene người phục vụ cho việc phân tích trình tự trong dự án bộ gene người (HGP), chẩn đoán các bệnh di truyền, và xác định các nghi can thuộc các trường hợp tội phạm (hoặc để chứng minh cho sự vô tội của họ). Thậm chí ngay cả những vết máu nhỏ nhất hoặc một sợi tóc còn sót lại ở hiện trường phạm tội vẫn có thể dùng làm khuôn cho PCR - khuếch đại đủ số DNA cần thiết cho việc kiểm tra DNA. DNA này được phân cắt bằng các *enzyme giới hạn* (restriction enzyme), và các phân đoạn tạo ra được tách riêng theo kích thước và thành phần bằng cách chạy *điện di gel* (gel electrophoresis). [Theo nguyên tắc, sự di chuyển của các phân tử qua bản gel lỗ xốp đáp ứng với dòng điện - các đoạn càng bé di chuyển càng nhanh hơn so với các đoạn lớn hơn]. Sau đó mẫu DNA này được đem so với mẫu của nghi can đang tạm giam. Nếu như nó trùng khớp, có thể dùng làm chứng cứ xác nhận nghi can đã gây án hoặc phạm tội ở hiện trường đó, giống như các dấu vân tay là bằng chứng hợp pháp vậy.

Nói chung, PCR là một kỹ thuật có ưu thế rất lớn và được sử dụng rất phổ biến trong các nghiên cứu liên quan đến công nghệ sinh học. Với

đóng góp to lớn của sự phát minh ra phương pháp PCR này, Kary Mullis đã được trao giải thưởng Nobel năm 1993 (cùng với Michael Smith, người đề xuất phương pháp *gây đột biến định hướng*, site-directed mutagenesis và sự phát triển của nó trong các nghiên cứu protein).

Ở nước ta, trong thời gian gần đây, phương pháp PCR cũng đã được áp dụng rộng rãi trong các nghiên cứu liên quan công nghệ sinh học và giải quyết các vấn đề xã hội khác. Một trong những hướng ứng dụng đó là kết hợp kỹ thuật PCR với các marker phân tử khác nhau để nghiên cứu tính đa dạng di truyền của nguồn tài nguyên động vật và thực vật ở Việt Nam. Chẳng hạn, Nông Văn Hải và cs (1999) đã bước đầu áp dụng kỹ thuật PCR kết hợp với việc phân tích enzyme giới hạn đoạn sản phẩm thu được, D-LOOP (~1,3 Kb), trong vùng điều khiển của DNA ty thể để tiến hành phân loại học phân tử và nghiên cứu vấn đề đa dạng di truyền ở ba loài Gà lôi (*Lophura*) ở Việt Nam. Tương tự, các tác giả cũng đã thông báo các kết quả phân tích các đoạn gene 18S rRNA (~0,25 Kb) ở các chủng Tuyến trùng ký sinh và ở côn trùng ở cây Bình vôi (1,7 Kb). Trần Thị Hoà và Triest (1999) sử dụng kỹ thuật PCR-RAPD trong nghiên cứu đa hình di truyền ở thực vật v.v.

Đối với các nghiên cứu ở cây lúa (*Oryza sativa*), có khá nhiều công trình nghiên cứu ứng dụng kỹ thuật PCR và các marker phân tử khác nhau. Chẳng hạn, trong các nghiên cứu về tính đa dạng phân tử và đánh giá quỹ gene (Nguyễn Đức Thành và cs 1999; Bùi Chí Bửu và Nguyễn Thị Lang 1999). Trong các nghiên cứu chọn tạo giống lúa, có một số công trình như: Tạo dòng phân tử và biểu hiện các gene liên quan *Myb* ở hạt lúa (Nguyễn Hoàng Lộc và cs 1999); đánh giá nguồn gene kháng bệnh đạo ôn phục vụ công tác chọn tạo giống lúa (Lã Tuấn Nghĩa 2001) v.v.

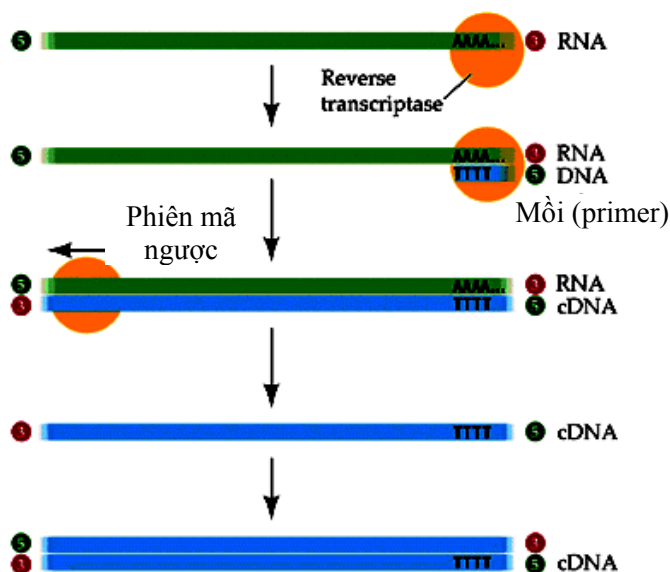
Ngoài ra, gần đây ở nước ta cũng đã thành lập một số Trung tâm phân tích DNA hoạt động trong các lĩnh vực hình sự (thuộc Bộ Công an), tư vấn di truyền (của GS Lê Đình Lương), và trong các xét nghiệm DNA liên quan đến sự chẩn đoán các bệnh ở một số Bệnh viện lớn trong cả nước.

5.2. Tổng hợp cDNA nhờ enzyme phiên mã ngược

Trong trường hợp cần thiết (chẳng hạn, đối với kỹ thuật tạo dòng để biểu hiện sản phẩm của một protein mong muốn), người ta có thể tổng hợp gene của nó dựa trên khuôn mẫu mRNA và *enzyme phiên mã ngược* (reverse transcriptase = RTase) được tinh chiết từ các retrovirus.

Như đã biết, hầu hết các mRNA eukaryote đều có đuôi poly(A) ở đầu 3'. Chính trình tự này tạo điều kiện thuận lợi cho việc tổng hợp sợi DNA bổ sung (cDNA). Khi đem trộn lẫn các đoạn ngắn gồm các nucleotide

thymine (oligo(dT)) với mRNA này sẽ xảy ra sự lai hoá giữa nó với vùng đuôi của mRNA. Đoạn oligo(dT) làm môi cho RTase tổng hợp sợi cDNA, sản phẩm ban đầu là sợi kép lai RNA- cDNA. Ở đầu 3' của sợi cDNA được tổng hợp có cái 'mũ' (tương tự đầu 5' của mRNA). Tiếp theo, bằng cách xử lý với NaOH, sợi mRNA được tách ra; kế đó 'mũ' ở đầu 3' của cDNA lại làm môi cho DNA polymerase I tổng hợp sợi thứ hai bằng cách kéo dài tiếp tục nó dọc theo sợi khuôn vốn có này. Vì vậy sản phẩm cDNA bây giờ có mang một cấu trúc 'vòng' sợi đơn. Sau đó, vòng này được cắt bỏ bằng cách xử lý với nuclease S1 để tạo ra cDNA sợi kép (hình 5.14).



Hình 5.14 Tổng hợp cDNA từ một RNA nhờ sử dụng enzyme phiên mã ngược.

III. Cơ sở phân tử của đột biến và tái tổ hợp DNA

1. Cơ sở phân tử của đột biến

1.1. Đại cương về đột biến gene

- **Định nghĩa:** *Đột biến gene* (gene mutation) là những biến đổi xảy ra bên trong cấu trúc của một gene hoặc một vùng nhỏ của bộ gene, liên quan chủ yếu tới sự thay đổi trình tự nucleotide vốn có của nó (kiểu đại), làm phát sinh các *allele* mới.

- **Nguyên nhân:** Các đột biến gene xảy ra có thể do sai sót trong quá trình tái bản, hoặc do trong bộ gene có các vùng dễ phát sinh đột biến gọi là các "điểm nóng" (hot spots), hoặc các tổn thương tự phát dưới ảnh hưởng của các tác nhân lý-hoá từ môi trường ngoài.

- **Phân loại:** Các đột biến gene có thể được phân loại theo các cách

khác nhau. Chẳng hạn, dựa vào các hiệu quả của chúng lên kiểu hình (các đột biến hình thái, đột biến gây chết, các đột biến soma và dòng mầm) hoặc lên chính bản thân vật chất di truyền DNA (các đột biến điểm). Căn cứ vào nguồn gốc, chúng được chia thành các *đột biến tự phát* (spontaneous) và *đột biến cảm ứng* (induced).

- *Hậu quả*: Các đột biến gene nói chung làm suy yếu hoặc biến đổi chức năng của gene hơn là tăng cường chức năng của nó. Từ đó ảnh hưởng lên các đặc tính sinh lý-hoá sinh của tế bào cũng như sức sống và sinh sản của cơ thể sinh vật nói chung.

- *Vai trò và ý nghĩa*: Đột biến gene vì vậy được xem là cơ sở của hiện tượng đa hình di truyền trong các quần thể và là nguồn biến dị di truyền sơ cấp vô cùng phong phú và đa dạng cho quá trình chọn lọc và tiến hoá. Người ta lợi dụng đặc tính biến đổi này của các sinh vật để xây dựng các phương pháp gây đột biến khác nhau và có thể kết hợp với lai hữu tính hoặc sử dụng kỹ thuật di truyền để cải biến bộ gene của các vật nuôi, cây trồng về các tính trạng cần quan tâm.

- *Tỷ lệ đột biến gene* (ở người): Nếu như *đột biến bộ gene* (genome mutation) và *đột biến nhiễm sắc thể* xảy ra với tần số trung bình tương ứng là 10^{-2} và 6×10^{-4} mỗi lần phân bào, thì với *đột biến gene*: tỷ lệ đột biến này là 10^{-10} cho mỗi cặp base trong mỗi lần phân bào hoặc có thể biến thiên từ 10^{-4} đến 10^{-7} (tính bằng số đột biến / locus / thể hệ) tùy thuộc vào kích thước gene và bản chất của nó. Tỷ lệ này ở hầu hết locus là 10^{-5} - 10^{-6} .

- *Hiện tượng đa hình (polymorphism) trong bộ gene*: Ở người, con số các đa hình tồn tại là ~1 trên mỗi 500 cặp base. Như vậy có khoảng 5,8 triệu sai khác trong mỗi bộ gene đơn bội là do các đột biến gây ra.

1.2. Các đột biến điểm (point mutations)

1.2.1. Các đột biến thay thế base (base substitution)

- Đồng hoán và dị hoán

Đồng hoán (transition) là dạng đột biến trong đó một purine này được thay bởi một purine khác ($A \rightarrow G$ hoặc $G \rightarrow A$) hoặc một pyrimidine này được thay bởi một pyrimidine khác ($T \rightarrow C$ hoặc $C \rightarrow T$).

Dị hoán (transversion) là dạng đột biến trong đó một purine được thay thế bằng một pyrimidine, hoặc ngược lại: $A \rightleftharpoons T$; $A \rightleftharpoons C$; $G \rightleftharpoons C$; $G \rightleftharpoons T$.

- Các đột biến nhầm nghĩa và vô nghĩa

+ *Đột biến nhầm nghĩa* (missense mutation) - biến đổi nghĩa của một codon kéo theo sự thay thế một amino acid. Hiệu quả của nó phụ thuộc vào vị trí và tính chất của amino acid bị biến đổi (bảng 5.2).

Bảng 5.2 Các đột biến nhằm nghĩa điển hình ở chuỗi β -globin người

Chuỗi β	Vị trí amino acid	Thay thế amino acid	Thay đổi ở β -mRNA	Kiểu thay thế ở gene β -globin
HbS	6	Glu \rightarrow Val	GAA \rightarrow GUA	Dị hoán A \rightarrow T
HbC	6	Glu \rightarrow Lys	GAA \rightarrow AAA	Đồng hoán G \rightarrow A
HbE	26	Glu \rightarrow Lys	GAA \rightarrow AAA	Đồng hoán G \rightarrow A

Một ví dụ tuyệt vời cho trường hợp này là *bệnh hồng cầu hình liềm* (sickle-cell disease; Hình 5.15a). Bằng cách phân tích và so sánh trình tự amino acid các chuỗi β -globin (146 amino acid) của hai protein này, Vernon Ingram phát hiện rằng chúng chỉ khác nhau một amino acid ở vị trí thứ sáu, trong HbA là glutamic acid và trong HbS là valine. Rõ ràng đây là đột biến dị hoán: A \rightarrow T (Hình 5.15b).

+ *Đột biến vô nghĩa* (nonsense mutation) - biến đổi một base ở *codon* có nghĩa (sense codon) bên trong trình tự mã hoá của gene tạo thành *codon kết thúc* chuỗi. Đột biến này gây ra chuỗi polypeptide ngắn bất thường và không có hoạt tính. Ví dụ, các dạng *thiếu máu vùng biển β -thalassemie* mà không tổng hợp được β -globin là do đột biến điểm ở các codon 17 hoặc 39 của gene β -globin.

Ngược lại, nếu đột biến làm biến đổi codon kết thúc thành codon có nghĩa sẽ dẫn tới hậu quả là chuỗi polypeptide được tổng hợp dài hơn bình thường. Đây là trường hợp đặc biệt của đột biến nhằm nghĩa. Điển hình là dạng *hemoglobin Constant Spring* (HbCS). Phân tử mRNA của chuỗi alpha bình thường có codon kết thúc là UAA với polypeptide dài 141 amino acid, còn ở mRNA dạng bệnh này nó là CAA mã hoá cho glutamine, với chuỗi alpha có tới 172 amino acid.



(a)

HbA	Thr	Pro	Glu	Glu
...	A C T	C C T	G A G	G A G...
Codon #	4	5	6	7
HbS	Thr	Pro	Val	Glu
...	A C T	C C T	G T G	G A G...

(b)

Hình 5.15 (a) Các tế bào hồng cầu bình thường (hình trên) và các tế bào dạng lưỡi liềm với kích thước phóng đại gần gấp đôi. (b) Một đoạn trình tự của các chuỗi β -globin và sợi đối khuôn của gene ở người bình thường và người bệnh.

- Các đột biến im lặng và trung tính

Các đột biến làm biến đổi thành phần base của một codon nhưng vẫn mã hoá cùng một amino acid, gọi là các *đột biến im lặng* (silent mutation) hay *đột biến đồng nghĩa* (synonymous). Phần lớn các đột biến này xảy ra ở vị trí thứ ba của các codon, và chúng xảy ra với tốc độ cao hơn rất nhiều so với các vị trí khác. Đây được coi là một trong những bằng chứng quan trọng của *Thuyết trung tính về sự tiến hoá phân tử* (xem Kimura 1983).

1.2.2. Đột biến dịch khung (frameshift mutation)

Dạng đột biến mất hoặc xen thêm một base trong gene dẫn đến khung đọc mã bị dịch chuyển một base, kể từ vị trí bị biến đổi cho đến cuối gene, được gọi là *đột biến dịch khung*. Hậu quả là vùng protein được tổng hợp tương ứng có trình tự amino acid bị biến đổi đáng kể. Điển hình là dạng *hemoglobin Wayne* (HbWI); do mất một base thứ ba của codon ở vị trí 138 trên chuỗi α (141 amino acid) gây hậu quả là ba amino acid ở các vị trí 139 - 141 sát đầu C của polypeptide bị biến đổi, đồng thời do codon kết thúc ở vị trí 142 cũng bị biến đổi đã kéo dài thêm 5 amino acid.

1.3. Các đột biến tự phát (spontaneous mutations)

Về nguyên tắc, tất cả các đột biến đều phải có một nguyên nhân, nhưng đôi khi các đột biến xảy ra mà không có các *tác nhân gây đột biến* (mutagens). Các đột biến tự phát như vậy có nhiều nguyên nhân khác nhau gây ra mà không phải từ bên ngoài.

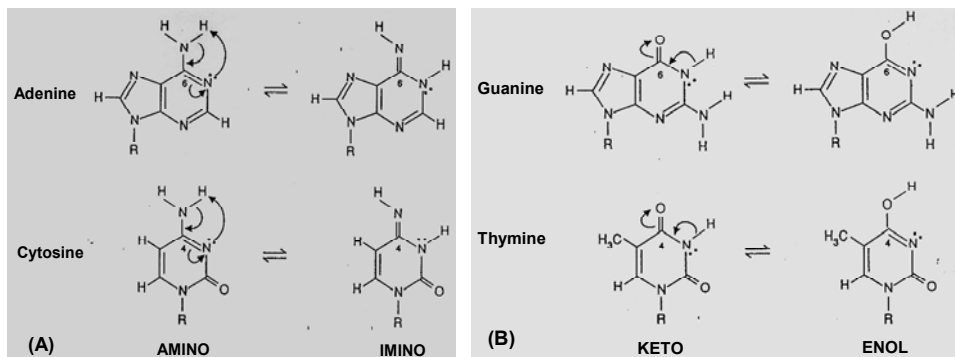
1.3.1. Các đột biến gây ra bởi bộ máy tái bản DNA

Dù cho sự tái bản được coi là chính xác, nhưng không phải là hoàn hảo. Như thế một số đột biến nào đó xảy ra đơn thuần là do tổng hợp sai. Tần số kết cặp base sai có thể là 10^{-5} và bản thân các DNA polymerase có khả năng *đọc sửa* trong khi tổng hợp DNA, nhờ vậy hạn chế đáng kể tần số phát sinh đột biến do tái bản ($10^{-5} \times 10^{-5} = 10^{-10}$).

Hiệu quả của tái bản không chính xác đặc biệt rõ ở các nòi đột biến của *E. coli* gọi là các *gene gây đột biến* (mutator). Các vi khuẩn này thường có nhiều sai sót trong khi tái bản, tỷ lệ đột biến của chúng cao hơn bình thường. Hiện giờ người ta đã lập được bản đồ gene cho một số đột biến gene mutator. Chẳng hạn, đó là các gene *mutD* mã hoá cho tiểu đơn vị epsilon của DNA polymerase III hoàn chỉnh. Đây là polypeptide cho enzyme hoạt tính đọc sửa exonuclease 3'→5'. Nếu không có hoạt tính này thì sự đọc sửa không thể xảy ra và do đó sẽ để lại nhiều đột biến. Các đột biến trong các gene *muH*, *mutL* và *mutS* cũng là các đột biến mutator. Chúng không sửa được các base kết cặp sai.

1.3.2. Sai sót trong khi tái bản do các base DNA

Các base trong DNA thường tồn tại ở một trong hai trạng thái biến đổi qua lại giữa chúng gọi là các *dạng hỗ biến* (tautomer). *Hiện tượng hỗ biến* (tautomerism) xảy ra có thể đơn thuần là do sự sắp xếp lại của các liên kết và do sự thay đổi vị trí của các nguyên tử hydro trong các base.



Hình 5.16 Các dạng hỗ biến của các base trong DNA. (A) Các dạng amino và imino của adenine và cytosine; và (B) Các dạng keto và enol của guanine và thymine. Chú ý các dạng amino và keto (bên trái mỗi hình) là phổ biến, còn các dạng imino và enol (bên phải mỗi hình) là các dạng hiếm.

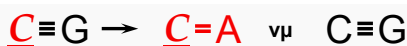
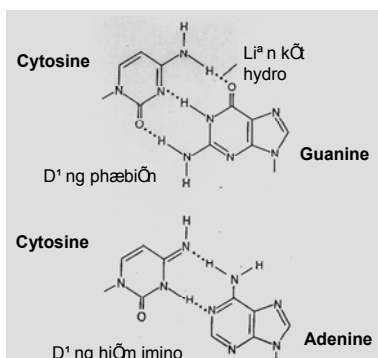
Các dạng phổ biến (bền vững) của adenine và cytosine là các dạng **amino**; chúng có thể sắp xếp lại thành các dạng hiếm **imino** kém bền hơn (hình 5.16A). Chính sự thay đổi vị trí của các hydro làm thay đổi đặc tính kết cặp của các base. Hậu quả của nó là dẫn tới sự kết cặp nhầm giữa các base trong tái bản, do đó trong những lần tái bản tiếp theo sẽ tạo ra các thể đột biến đồng hoán tương ứng. Ví dụ, DNA cha mẹ chứa cặp CG, trong lúc tái bản C chuyển sang dạng imino (hiếm) và cặp nhầm với A (thay vì với G) tạo thành cặp C-A trong một DNA con. Trong lần tái bản tiếp theo, A sẽ cặp bình thường với T tạo ra cặp T-A thay chỗ cặp CG trước đây. Đây là cơ chế của đột biến đồng hoán CG \rightarrow TA (hình 5.17).

Hình 5.16B cũng cho thấy các dạng **keto** (phổ biến) và dạng **enol** (hiếm) của guanine và thymine. Các dạng hỗ biến hiếm như enol của guanine chẳng hạn, nếu xảy ra nó sẽ kết cặp với thymine (thay vì với cytosine) trong khi tái bản. Vì vậy trong lần tái bản tiếp theo thymine ở vị trí vốn bị kết cặp sai trong lần tái bản đầu sẽ kết cặp với adenine tạo ra một thể đột biến đồng hoán GC \rightarrow AT (hình 5.17). Với phân tích tương tự, ta thấy các dạng đột biến tự phát này chủ yếu là các đồng hoán.

1.3.3. Các đột biến dịch khung tự phát trong khi tái bản

Các tác nhân xen vào giữa (intercalating agents) là nhóm tác nhân

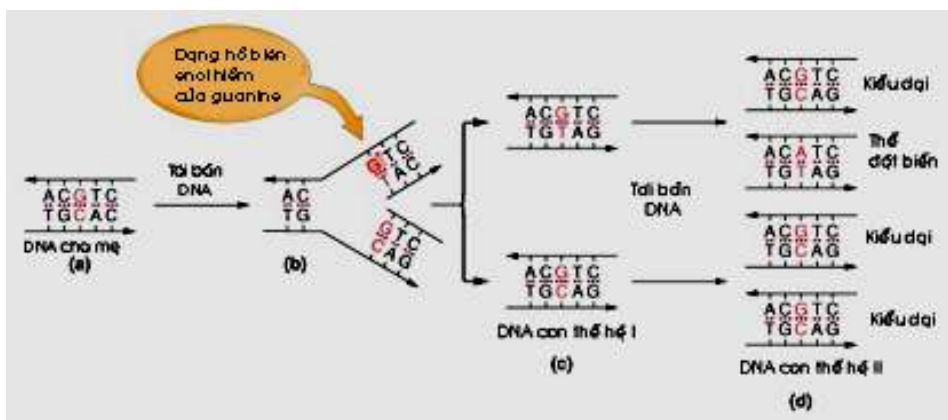
quan trọng khác gây biến đổi DNA, bao gồm các thuốc nhuộm acridine. Chúng là các phân tử mô phỏng các base và có thể xen vào giữa các base nitơ của chuỗi xoắn kép DNA. Bằng cách đó chúng gây ra sự xen thêm hoặc mất một cặp nucleotide trên gene (vì acridine có kích thước gấp đôi một nucleotide), dẫn đến khung đọc mã thay đổi và các protein được tổng hợp thường không có hoạt tính.



Sử hình thành hệ biến C trong t, i b' n sĩ g' y ra k' c' p sai
vũ sử xen vào kh' ng' óng của A trong mét sĩ i của DNA con



Tổ @y cả th' g' y ra @ét bi' n' @ng ho, n C-G th' n' T-A trong
l' n t, i b' n k' o' t' i' p, ho' @' @' i c k' c' p i' i kh' ng' óng c, ch



Hình 5.17 Đột biến đồng hoán CG → TA gây ra bởi sự h' b' i' n' của cytosine (hình trên) và đồng hoán GC → AT gây ra bởi h' b' i' n' của guanine (hình dưới).

1.3.4. Các đột biến tự phát gây ra bởi deamine hoá

Các đột biến tự phát có thể xảy ra bằng các cơ chế kết cặp sai khác trong tái bản DNA. Chẳng hạn, các base đặc biệt là cytosine, có xu hướng mất đi nhóm amin của chúng trong một quá trình gọi là *deamine hoá* (deamination). Kiểu deamin hoá phổ biến nhất là đảo cytosine thành uracil, thường thì không dẫn tới đột biến, bởi vì các tế bào có một cơ chế tách bỏ uracil (nhờ enzyme uracil-DNA glycosylase). Enzyme này cắt liên kết giữa uracil và deoxyribose của nó; để rồi một enzyme khác tái bổ sung một cytosine vào gốc đường này để nó cặp với guanine ở sợi đối diện.

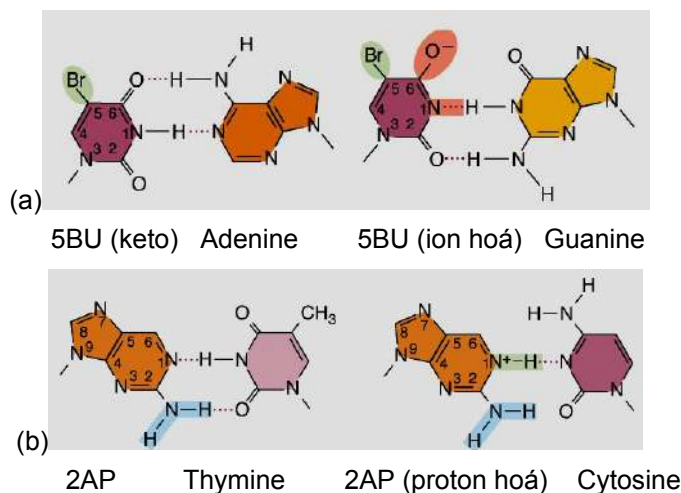
DNA của một số sinh vật có chứa một số lượng nhỏ các base sửa đổi thêm vào 4 loại thông thường. Phổ biến nhất là *5-methylcytosine* (5mC); nó có thể cặp chính xác với guanine như các cytosine chuẩn. Tuy nhiên, các vị trí chứa 5mC trong DNA có thể trở thành các "điểm nóng" cho các đột biến tự phát thông qua deamine hoá. Điều này đã được chứng minh bằng cách sử dụng HNO_2 để deamine hoá cytosine thành uracil; sau đó uracil kết cặp với adenine trong tái bản và kết quả là tạo ra đột biến đồng hoán GC \rightarrow AT. Mặt khác, nó deamin hoá adenine thành hypoxanthine (có thể cặp với C giống như G) và sẽ gây đồng hoán ngược lại, AT \rightarrow GC.

1.4. Các tác nhân gây đột biến

1.4.1. Các tác nhân hoá học

- Đột biến gây ra bởi các chất tương tự nucleoside

Một số hợp chất hoá học có thể làm tăng tần số hồ biến và qua đó gây ra các đột biến. Ví dụ kinh điển là *5-bromodeoxyuridine* (*BrdU* hay *5-BU*), một chất giống như thymidine ngoại trừ thay một nguyên tử bromine (Br) cho một nhóm methyl ở vị trí 5. Tuy nhiên, một khi 5-BU được kết hợp vào chỗ của thymidine, nó có thể gây rắc rối. Rắc rối bắt nguồn từ chỗ 5-BU có xu hướng tăng cường bật sang dạng hồ biến enol để có thể kết cặp base giống như C thay vì T, tạo ra cặp 5-BU-G (hình 5.18). Và ở lần tái bản sau tạo thành cặp GC thay chỗ AT. Kết quả là tạo ra đồng hoán AT \rightarrow GC. Do hồ biến mà 5-BU có thể gây đồng hoán hai chiều, AT \rightleftharpoons GC.



Hình 5.18 Các chất tương tự nucleoside. (a) Dạng keto phổ biến của 5-BU là chất tương tự thymidine, có thể cặp với A; còn ở dạng ion hoá - enol, nó có thể cặp với G. (b) 2-aminopurine là chất tương tự adenosine, có thể cặp với T; khi ở dạng proton hoá, nó có thể cặp với C.

Một hóa chất gây đột biến khác là *2-aminopurine* (2-AP), chất tương tự của A, có thể cặp với T. Khi bị proton hóa, 2-AP có thể kết cặp nhầm với C để gây đột biến đồng hoán AT → GC ở lần tái bản sau (hình 5.18).

- Đột biến gây ra bởi sự alkyl hoá của các base

Một số chất trong môi trường là các *chất có ái lực với điện tử* (electrophyle); chúng tìm các trung tâm điện tích âm trong các phân tử khác và bám vào. Nhiều chất khác thuộc môi trường được chuyển hoá trong cơ thể thành ra các hợp chất có ái lực điện tử. Một trong những trung tâm điện tích âm rõ nhất trong sinh học là phân tử DNA. Mỗi nucleotide chứa một điện tích âm toàn phần trên phosphate và các điện tích âm từng phần trên các base. Khi các chất ái lực điện tử bắt gặp các trung tâm điện tích âm này, chúng sẽ tấn công, thông thường là gắn thêm các nhóm chứa carbon gọi là các *nhóm alkyl*. Quá trình đó gọi là *alkyl hoá* (alkylation).

Nhiều *tác nhân gây ung thư* (carcinogen) là các chất ái lực điện tử đường như hoạt động bằng cách tấn công DNA và alkyl hoá nó. Nhiều tác nhân đột biến được sử dụng trong phòng thí nghiệm để gây tạo các đột biến cũng là các tác nhân alkyl hoá, ví dụ ethylmethane sulfonate (EMS). Khi xử lý bằng EMS, hoá chất này nhường nhóm ethyl ($\text{CH}_3\text{-CH}_2$) cho DNA mà cụ thể là cho oxy O^6 của guanine tạo ra O^6 -alkylguanine. Base được alkyl hoá này cặp với thymine thay vì cytosine. Kết quả là sinh ra đồng hoán GC → AT ở lần tái bản sau.

1.4.2. Các tác nhân vật lý (đột biến cảm ứng-phóng xạ)

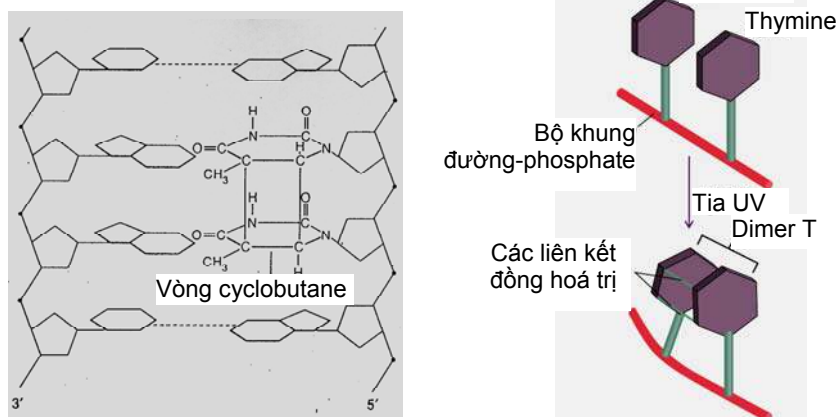
Trong số tác nhân phóng xạ gây đột biến, các tia tử ngoại, tia gamma và tia X là phổ biến trong tự nhiên và trong các thí nghiệm gây đột biến. Chúng khác nhau chủ yếu về năng lượng, do đó các kiểu tổn thương DNA do chúng gây ra cũng khác nhau.

+ *Bức xạ tử ngoại* (ultraviolet radiation = UV) là tương đối yếu, vì vậy tổn thương do nó gây ra là tương đối nhỏ; ví dụ hai pyrimidine nằm kề nhau trên cùng một sợi liên kết ngang với nhau tạo thành các dimer, thông thường là *dimer thymine* (hình 5.19). Tổn thương này biểu hiện ở sự biến dạng của chuỗi xoắn kép và cản trở sự kết cặp của các base purine trong tái bản. Hậu quả là có thể làm chết tế bào hoặc tạo ra đột biến do việc lắp sai nucleotide vào vị trí đối diện với dimer trong khi tái bản.

Loại bức xạ cực tím làm tổn thương DNA nặng nhất có bước sóng khoảng 260 nm, tức là vùng sóng mà DNA hấp thụ mạnh nhất. Bức xạ này có rất nhiều trong ánh sáng mặt trời và có nguy cơ gây ung thư da.

+ *Các tia gamma và tia X* chứa nhiều năng lượng hơn các tia UV, có thể tương tác trực tiếp với phân tử DNA. Chúng ion hoá các phân tử xung

quanh DNA và tạo thành các *gốc tự do* có phản ứng cao có thể tấn công DNA, làm biến đổi các base hoặc làm đứt các sợi.



Hình 5.19 Sự hình thành dimer thymine bởi tia cực tím (UV).

◆ *Hồi biến* (reversion) hay *đột biến ngược* (back-mutation) là đột biến quay lại kiểu hình ban đầu. Các đột biến gây ra bởi các base tương đồng hoặc HNO_2 có thể hồi biến bằng cách tiếp tục xử lý chính tác nhân đó. Tuy nhiên một số chất như hydroxylamine vì chỉ phản ứng đặc thù với cytosine và làm phát sinh đồng hoán một chiều $\text{GC} \rightarrow \text{AT}$, nên không thể hồi biến bằng chính nó.

◆ Áp dụng phương pháp đột biến trong chọn giống thực vật ở nước ta, có thể nói phát triển mạnh trong các thập niên 1970 - 1990, đã góp phần tạo ra nhiều giống mới ở lúa, ngô, đậu tương, v.v. Phần lớn các công trình nghiên cứu này được tiến hành ở các Viện Khoa học Kỹ thuật Nông nghiệp Việt Nam, Viện Di truyền Nông nghiệp, Viện Cây lương thực và thực phẩm, Viện Bảo vệ thực vật, các trường ĐHTH Hà Nội và ĐHSP Hà Nội. Chẳng hạn, gần đây, bằng phương pháp chiếu xạ tia gamma (^{60}Co) lên hạt khô và kết hợp lai hữu tính, các nhà khoa học Viện DTNN đã chọn tạo được nhiều giống lúa cho năng suất cao và có một số đặc tính quý như chín sớm, chống chịu sâu bệnh... như các giống DT17, DT21, v.v. Bằng phương pháp chiếu xạ tia gamma (^{60}Co), với liều lượng 10 - 20 krad lên hạt nảy mầm ở các thời điểm thích hợp 66 - 75 giờ, nhóm nghiên cứu của PGS.TS. Nguyễn Minh Công (ĐHSP Hà Nội), TS. Đỗ Hữu Ất (Viện DTNN), và TS. Phạm Văn Ro (Viện lúa Đồng bằng sông Cửu long) đã chọn tạo thành công các giống lúa Tám thơm đột biến, Tép hành đột biến (THDB) và Tài nguyên đột biến TNDB-100 (Do Huu At et al 2000) v.v.

2. Sửa chữa DNA (DNA repair)

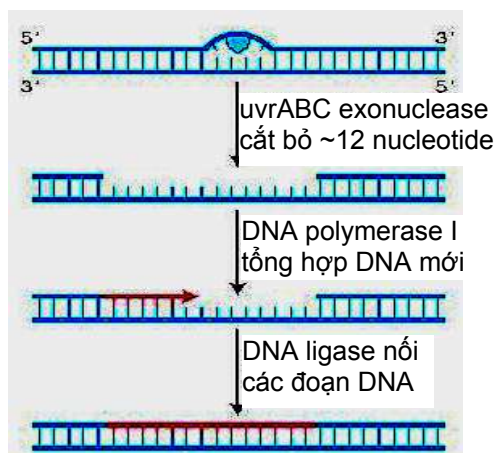
Sửa chữa DNA là đặc tính thiết yếu của tế bào sống, nhằm phục hồi cấu trúc DNA bị tổn thương do các tác nhân lý-hoá hoặc tự phát trong quá trình tái bản DNA. Có nhiều hệ thống sửa sai trong các tế bào. Tựu trung có các cơ chế cơ bản sau: *đọc sửa* trong lúc tái bản nhờ một số DNA polymerase (như đã đề cập), *quang phục hoạt*, *cắt bỏ* base hoặc nucleotide (và *sửa chữa sau tái bản* - không xét ở đây). Tỷ lệ đột biến tự nhiên nói chung thấp là nhờ tính hiệu quả của các hệ thống sửa sai này.

2.1. Sửa chữa bằng quang phục hoạt

Quang phục hoạt (photoreactivation) là kiểu sửa chữa xảy ra dưới tác dụng của ánh sáng, nên còn gọi là *sửa chữa ngoài sáng* (light repair). Nó được xúc tác bởi enzyme *DNA photolyase*. Cơ chế sửa chữa dimer thymine bởi enzyme này được hình dung như sau: (i) phát hiện và bám vào vị trí DNA bị tổn thương; (ii) enzyme hấp thụ ánh sáng có bước sóng 320-370 nm, và được kích hoạt để có thể cắt đứt các liên kết vốn tạo ra dimer, làm phục hồi trạng thái bổ sung giữa các base của hai sợi; và (iii) enzyme rời khỏi DNA, và hư hại được chữa xong.

2.2. Sửa chữa bằng cắt bỏ

Sự *sửa chữa bằng cắt bỏ* (excision repair) được thực hiện bởi các enzyme đặc thù và không cần ánh sáng, nên gọi là *sửa chữa trong tối* (dark repair). Sự tách bỏ vùng DNA hư hại và thay bằng DNA mới xảy ra bằng cách hoặc *cắt bỏ base* hoặc *cắt bỏ nucleotide*.



Hình 5.20 Các bước sửa sai bằng cách cắt bỏ nucleotide.

Đối với sự sửa chữa bằng cách cắt bỏ nucleotide (nucleotide excision repair), chẳng hạn, tổn thương base hàng loạt, kể cả dimer thymine, cũng có thể tách bỏ trực tiếp mà không cần sự giúp đỡ của DNA glycosylase. Ở

E. coli, enzyme chủ chốt của quá trình này là *urvABC endonuclease* (vì nó chứa ba polypeptide của các gene *urvA*, *urvB* và *urvC*). Enzyme này cắt một đoạn dài 12-13 nucleotide trên một sợi, tùy thuộc vào sai hỏng ảnh hưởng lên một nucleotide (alkyl hoá) hoặc hai nucleotide (dimer). Khoảng trống này sẽ được lấp đầy nhờ *DNA polymerase I*. Và cuối cùng, *DNA ligase* sẽ hàn gắn các khe hở (hình 5.20).

3. Tái tổ hợp và đại cương về công nghệ DNA tái tổ hợp

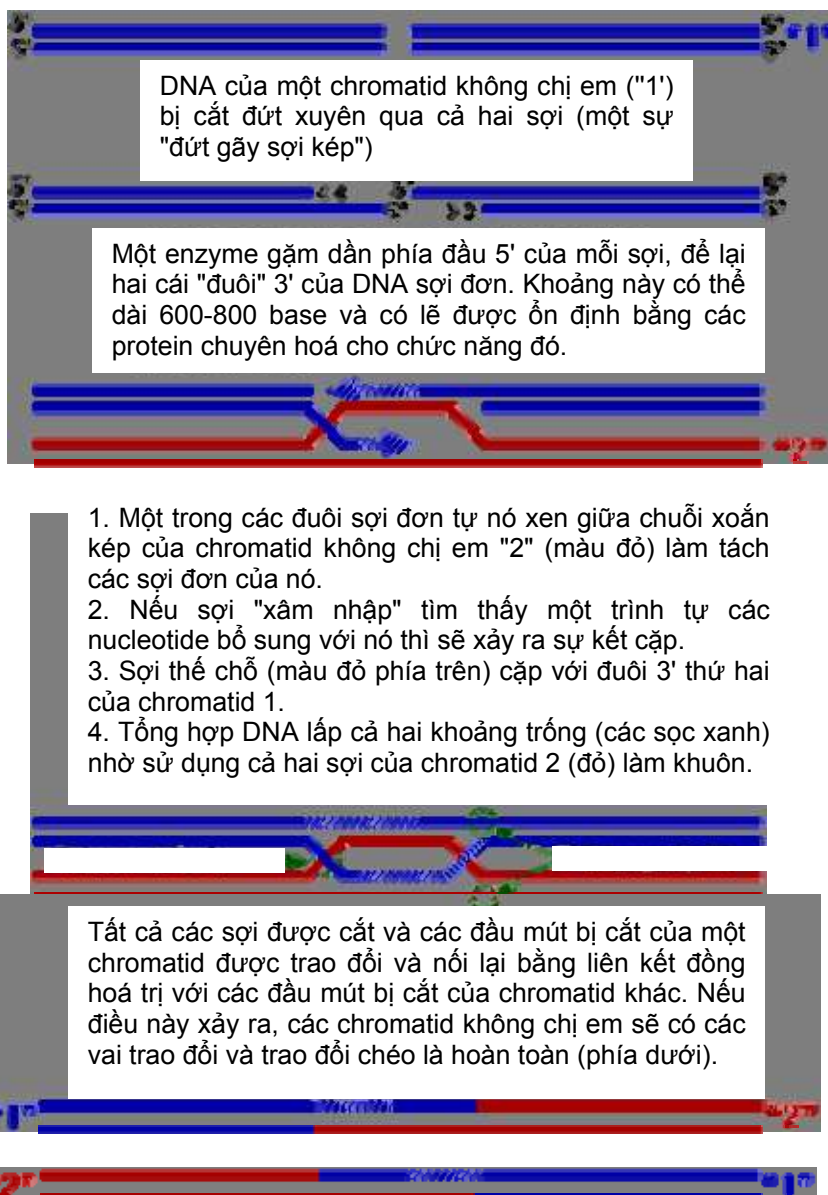
3.1. Cơ chế tái tổ hợp (recombination)

Tái tổ hợp là sự trao đổi vật liệu di truyền tạo ra các tổ hợp gene mới trong các phân tử DNA thể hệ con, là nhân tố quan trọng trong tiến hoá. Nói đến tái tổ hợp là nói đến sự *trao đổi sợi* hay trao đổi chéo có thể xảy ra nội trong phân tử (do sự vặn xoắn và đứt-nối của một sợi) hoặc có liên quan giữa hai phân tử với cơ chế thuận nghịch hoặc không thuận nghịch.

Ngoài ra, có ba kiểu tái tổ hợp khác nhau về căn bản: (1) *Tái tổ hợp vị trí đặc hiệu* (site-specific recombination) phụ thuộc vào trình tự tương đồng được giới hạn giữa các DNA tái tổ hợp. Điển hình là sự xen của DNA phage λ vào DNA vật chủ. (2) *Tái tổ hợp không chính thức* (illegitimate recombination) cũng cần một sự tương đồng nhỏ nào đó giữa các DNA tái tổ hợp, nhưng không đặc thù. Các yếu tố di truyền vận động thường sử dụng con đường tái tổ hợp ngoại lệ này. (3) *Tái tổ hợp tương đồng* (homologous recombination) hay đôi khi được gọi là *tái tổ hợp chung* (generalized recombination) là loại tái tổ hợp xảy ra trong giảm phân. Đây là dạng phổ biến nhất của tái tổ hợp, phụ thuộc vào độ tương đồng về trình tự rộng rãi giữa các phân tử DNA tham gia. Dưới đây chúng ta chỉ tìm hiểu sơ lược mô hình Holliday về tái tổ hợp tương đồng.

Các mô hình hiện nay về cơ chế tái tổ hợp tương đồng thống nhất gồm các bước thiết yếu nhưng không nhất thiết theo thứ tự, như sau: (1) kết cặp của hai DNA tương đồng; (2) đứt gãy của hai trong số các sợi DNA tương đồng; (3) tái tạo các liên kết phosphodiester để nối liền hai sợi tương đồng (hoặc các vùng của cùng một sợi trong tái tổ hợp nội phân tử); (4) đứt gãy hai sợi khác và nối chúng lại. Cơ chế cơ sở do Robin Holliday đưa ra vào năm 1964 (gọi là mô hình Holliday).

Tổng quát, mô hình Holliday dựa trên việc tạo ra một cầu chéo có thể di chuyển dọc theo hai chuỗi xoắn kép khác nhau và sau đó cắt đứt cấu trúc trung gian theo một trong hai cách để tạo ra các kiểu phân tử DNA tái tổ hợp khác nhau (Hình 5.21).



Hình 5.21 Con đường tái tổ hợp tương đồng với sự tham gia của các protein tái tổ hợp RecBCD. Đầu tiên, RecBCD tạo các vết đứt; sau đó các protein RecA và SSB xúc tiến sự xâm nhập sợi, tạo ra bắt chéo hay là chỗ nối Holliday; RuvA và B thúc đẩy sự di chuyển theo nhánh bên; RuvC tạo các vết nứt làm dung giải cấu trúc này thành ra các DNA tái tổ hợp.

3.2. Đại cương về công nghệ DNA tái tổ hợp

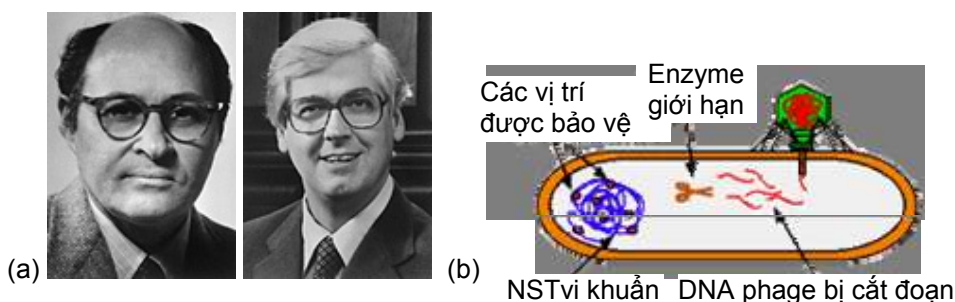
Các nghiên cứu của di truyền vi sinh vật về các enzyme giới hạn, các

plasmid và việc sử dụng chúng để tạo ra các phân tử DNA tái tổ hợp *in vitro* trong những năm đầu của thập niên 1970 là cơ sở cho sự ra đời của *công nghệ DNA tái tổ hợp* (recombinant DNA technology).

Dưới đây chúng ta chỉ tìm hiểu sơ lược về các enzyme giới hạn và vector - những công cụ thiết yếu của kỹ thuật di truyền, phương pháp cơ bản của việc xây dựng DNA tái tổ hợp *in vitro* và một quy trình đơn giản về sự tạo dòng gene ở vi khuẩn và một số ứng dụng của nó.

3.2.1. Các enzyme giới hạn (restriction endonuclease)

Công trình nghiên cứu đầu tiên của Daniel Nathan và Hamilton Smith năm 1969 ở *Haemophilus influenzae* xác nhận các tế bào vi khuẩn là những hệ thống chứa các enzyme sửa đổi và enzyme giới hạn (Hình 5.2). Các enzyme đều có đối tượng nhận biết là các đoạn trình tự DNA vật chủ và DNA ngoại lai, nhưng có vai trò khác nhau. Các *enzyme sửa đổi* (methylase) đóng vai trò bảo vệ DNA vật chủ bằng cách gắn thêm nhóm methyl ở một số base nhất định trong *đoạn nhận biết* hay *đoạn đích* (recognition/target sequence), trong khi các *enzyme giới hạn* (restriction endonuclease) đóng vai trò vô hiệu hoá hoạt tính của các DNA lạ bằng cách phân cắt ở các vị trí đặc thù (Hình 5.22).



Hình 5.22 (a) D.Nathan (trái) và H.Smith. (b) Enzyme giới hạn cắt vụn DNA của phage lạ khi nó xâm nhập vào tế bào vi khuẩn.

* Danh pháp và tính chất của các enzyme giới hạn

- Các enzyme giới hạn chỉ phát hiện được ở các vi khuẩn vì vậy, tên gọi của các enzyme giới hạn được biểu thị bằng ba hoặc bốn chữ cái viết tắt của vi khuẩn mà từ đó enzyme được chiết xuất. Chữ cái đầu tiên được viết hoa để chỉ *chi* (genus) và hai chữ cái tiếp theo viết thường để chỉ *loài* (species), và nếu cần thêm chữ cái thứ tư để chỉ *nòi* hoặc *chủng* (strain, type). Ngoài ra, để phân biệt các enzyme của cùng một nòi dùng thêm số La Mã kèm theo sau (xem Bảng 5.3).

- Tính chất quan trọng nhất của các enzyme giới hạn là *tính đặc hiệu vị*

trí, nghĩa là chúng có thể nhận biết và cắt tại trình tự DNA đặc thù.

- Đoạn đích là có kích thước ngắn 4-8 cặp base và có tính *đối xứng xuôi ngược* (palindrome).

- Các enzyme giới hạn khác nhau có hai kiểu cắt: cắt lệch và cắt thẳng. Với kiểu cắt lệch, tạo ra các đoạn DNA có các đầu sợi đơn bổ sung gọi là các *đầu dính* (cohesive/sticky ends); ví dụ EcoRI. Với kiểu cắt thẳng, tức cắt cùng vị trí trên cả hai sợi của DNA sợi kép, tạo ra các đoạn DNA có các *đầu bằng* (blunt ends); ví dụ AluI (Bảng 5.3).

Bảng 5.3 Các trình tự nhận biết và vị trí cắt của các enzyme giới hạn được chọn lọc (mũi tên chỉ vị trí cắt; các trình tự ở đây được chỉ ra trên sợi 5'→3')

Nguồn vi sinh vật	Tên enzyme	Trình tự nhận biết
<i>Arthrobacter luteus</i>	AluI	AG↓CT
<i>Bacillus amyloliquefaciens</i> H	BamHI	G↓GATCC
<i>Escherichia coli</i> RY13	EcoRI	G↓AATTC
<i>Haemophilus influenzae</i> Rd	HindII	GTPy↓PuAC
<i>Haemophilus influenzae</i> Rd	HindIII	A↓AGCTT
<i>Nocardia otitidis-caviarum</i>	NotI	GC↓GGCCGC
<i>Providencia stuartii</i>	PstI	CTGCA↓G
<i>Serratia marcescens</i>	SmaI	CCC↓GGG
<i>Xanthomonas malvaccarum</i>	XmaI	C↓CCGGG

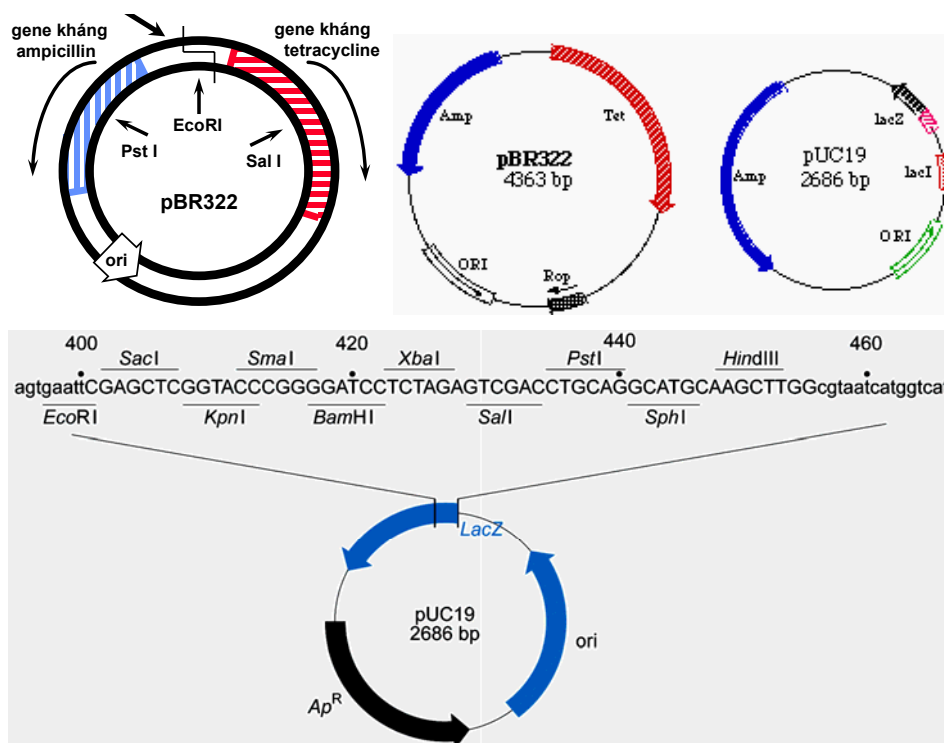
Bên cạnh các enzyme giới hạn, còn có các enzyme khác thuộc ba nhóm sau đây được sử dụng trong lĩnh vực công nghệ này: (i) Các *polymerase*: DNA polymerase và RNA polymerase, kể cả reverse transcriptase; (ii) Các *ligase*: DNA ligase và RNA ligase; và (iii) Các *nuclease*: DNase và RNase.

3.2.2. Các vector

DNA tái tổ hợp (recombinant DNA) là phân tử DNA được tạo ra trong ống nghiệm bằng cách kết hợp các DNA từ các nguồn khác nhau, theo một quy trình kỹ thuật nhất định (gọi là kỹ thuật tái tổ hợp DNA).

Thông thường một phân tử DNA tái tổ hợp bao gồm một phân tử DNA có bản chất là *plasmid* hoặc *phage* nguyên vẹn gọi là *vector* (thể tải) và một đoạn DNA từ nguồn khác mang một gene hoặc yếu tố điều hòa mong muốn được cho xen vào gọi là *DNA ngoại lai* (foreign DNA).

Trong số các vector, các plasmid của vi khuẩn được sử dụng rộng rãi hơn cả, nhờ có các đặc điểm sau: (i) Mỗi plasmid là một phân tử DNA sợi kép thường ở dạng vòng và chỉ có một khởi điểm tái bản riêng; (ii) Có khả năng xâm nhập vào tế bào vật chủ và hoạt động bình thường; (iii) Có kích thước bé, thường chỉ vài ngàn cặp base nên dễ dàng tinh chiết; (iv) Số bản sao trong mỗi tế bào vi khuẩn thường khá cao; (v) Một số plasmid có chứa các gene kháng thuốc tiện lợi cho việc theo dõi và phát hiện sự có mặt của plasmid tái tổ hợp trong vi khuẩn chủ (Hình 5.23).



Hình 5.23 Các plasmid pBR322 và pUC19. Ở đây cho thấy kích thước, khởi điểm (Ori) và các vùng chứa gene kháng ampicillin (Amp^R) và tetracyclin (Tet^R)... cũng như vị trí cắt của một số enzyme giới hạn trong các vùng này.

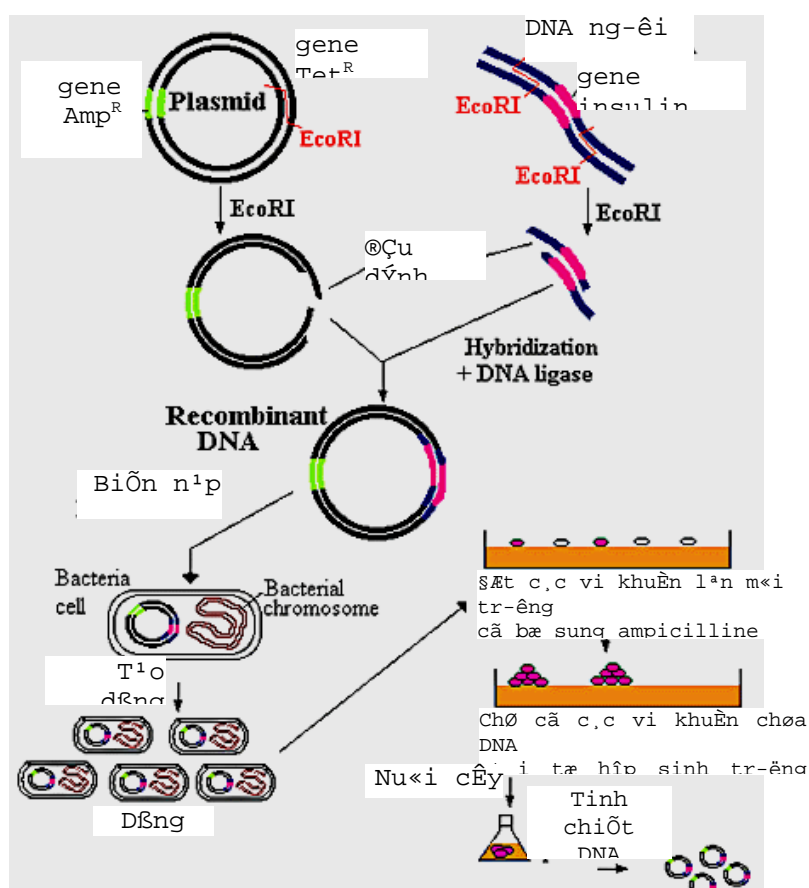
3.2.3. Xây dựng phân tử DNA tái tổ hợp *in vitro*

Bất kỳ đoạn DNA nào nếu được cắt bởi cùng một loại enzyme giới hạn (ví dụ, *EcoRI*) cho các đầu dính thì có thể bổ sung với nhau và được nối bởi DNA ligase (hình 5.24). Phương pháp thành lập phân tử DNA tái tổ hợp kiểu này lần đầu tiên được đưa ra bởi J.Mert và R.Davis năm 1972 bằng thực nghiệm trên các virus. Và sau đó, lần đầu tiên năm 1973, H.Boyer và nhóm nghiên cứu của S.Cohen đã tạo ra được phân tử DNA tái tổ hợp gồm vector là plasmid nhỏ pSC101 của *E. coli* và DNA "ngoại

lai" là một plasmid khác. Chính sự kiện này đã đặt nền móng và mở ra triển vọng to lớn cho kỹ thuật DNA tái tổ hợp sau này.

3.2.4. Tạo dòng gene ở vi khuẩn

Về nguyên tắc, kỹ thuật DNA tái tổ hợp hay *tạo dòng* (cloning) gồm các bước chung nhất như sau: (1) Tách chiết và tinh sạch DNA thuộc các nguồn khác nhau (gồm vector và DNA mang gene mong muốn); (2) Kiến tạo phân tử DNA tái tổ hợp *in vitro*; (3) Đưa phân tử DNA tái tổ hợp vào trong tế bào nhận, thường là *E. coli* hoặc nấm men; và (4) Phát hiện và phân lập các dòng DNA tái tổ hợp đặc hiệu.



Hình 5.24 Tạo dòng vi khuẩn mang DNA tái tổ hợp có chứa gene insulin người.

Trong tế bào chủ, phân tử DNA tái tổ hợp có thể biểu hiện gene mong muốn hoặc tái bản độc lập nhiều lần để tạo ra hàng loạt bản sao của nó, và khi tế bào chủ phân chia sẽ kéo theo hiện tượng *tạo dòng* (cloning). Mặt khác, do tốc độ phân chia rất nhanh của các vi khuẩn nên có thể tạo

hàng triệu bản sao mong muốn. Vì thế nhà khoa học có thể tách dòng bất kỳ một gene nào dùng để nghiên cứu hoặc sản xuất trên quy mô công nghiệp một số lượng lớn các protein vốn là những chế phẩm y-sinh học nào đó.

Hình 5.24 giới thiệu một quy trình kỹ thuật tạo dòng ở vi khuẩn mang gene insulin người. Việc phát hiện DNA tái tổ hợp dựa trên khả năng kháng thuốc do vector plasmid mang lại (chứa hai gene kháng Amp^R và Tet^R; và giả thiết gene Tet^R có chứa điểm cắt của EcoRI).

Một số sản phẩm được tạo dòng ở vi khuẩn liên quan đến các ứng dụng khác nhau trong điều trị bệnh ở người được giới thiệu ở bảng 5.4.

[Về chi tiết, có thể xem trong các giáo trình *Công nghệ DNA tái tổ hợp* do Nguyễn Hoàng Lộc chủ biên; *Di truyền vi sinh vật và ứng dụng* do Hoàng Trọng Phán chủ biên, NXB Đại học Huế - 2006.]

Bảng 5.4 Các sản phẩm sinh ra thông qua các vi khuẩn chứa các gene người được tạo dòng

Sản phẩm	Áp dụng
Các interferon	Dùng để điều trị các bệnh lây nhiễm virus và ung thư
Interleukin 2	Kích thích hệ thống miễn dịch và có thể dùng trong điều trị ung thư và các rối loạn hệ thống miễn dịch
Insulin	Được dùng để điều trị bệnh tiểu đường
Somatotropin	Được dùng để điều trị dị tật lùn thuộc về tuyến yên
Chất kích hoạt plasminogen	Làm tan các cục đông máu cho điều trị và ngăn chặn các tình trạng tắc nghẽn tim mạch
Nhân tố hoại tử khối u	Tấn công và giết các khối u ung thư
Các nhân tố XI và VIII	Được dùng để điều trị bệnh máu khó đông
Erythropoietin	Kích thích tạo các tế bào hồng cầu cho điều trị bệnh thiếu máu (anemia)
Beta endorphin	Các chất giảm đau tự nhiên do cơ thể tạo ra; có thể được dùng làm thuốc giảm đau
Các enzyme	Được dùng rộng rãi từ việc điều khiển các phản ứng hoá học trong các quá trình kỹ nghệ cho đến việc bổ sung các enzyme trong khẩu phần ăn của người
Các vaccine tiểu đơn vị	Kích thích khả năng miễn dịch của cơ thể đối với một hoặc hai kháng nguyên then chốt của một tác nhân gây bệnh; giảm nguy cơ rủi ro của các vaccine thông thường

Câu hỏi và Bài tập

1. Trình bày các giai đoạn của quá trình sinh tổng hợp nucleotide purine IMP và các dẫn xuất AMP và GMP.
2. Nêu các giai đoạn của quá trình sinh tổng hợp nucleotide pyrimidine UMP và các dẫn xuất của nó.
3. Tại sao nói tái bản là một trong những đặc tính quan trọng nhất của DNA và vật chất di truyền nói chung? Nguyên tắc chung của tái bản DNA ở các prokaryote và eukaryote là gì?
4. Phân tích vai trò của các enzyme và protein tham gia tái bản ở *E. coli*. Giải thích cơ chế tái bản nửa gián đoạn và vẽ một sơ đồ minh họa.
5. Hãy chỉ ra những điểm giống nhau và khác nhau trong: (a) khởi điểm tái bản; (b) enzyme tham gia tái bản; và (c) các cơ chế mở đầu, kéo dài, và kết thúc tái bản giữa prokaryote và eukaryote.
6. Thế nào là "giới hạn Hayflick"? Hãy cho biết mối liên quan giữa telomere, telomerase với hiện tượng ung thư và lão hoá của tế bào.
7. Phân tích cấu trúc của telomere và cơ chế duy trì sự ổn định của đầu mút nhiễm sắc thể eukaryote.
8. Sự tái bản của các bộ gene RNA ở một số virus khác với các hệ thống tái bản DNA ở những điểm nào? Tại sao kích thước các bộ gene RNA thường nhỏ, nhưng chúng lại có khả năng gây nguy hiểm thực sự đối với vật chủ? Hãy giải thích một cơ chế phát triển ung thư do virus.
9. Phương pháp PCR được phát triển dựa trên hiểu biết về sự tái bản của DNA như thế nào? Trình bày các nguyên lý của quy trình khuếch đại gene và các ứng dụng của nó.
10. Thế nào là phiên mã ngược? Trình bày cơ chế tái bản của các retrovirus và ứng dụng của hiểu biết này trong kỹ thuật di truyền.
11. Phân biệt các cặp khái niệm đột biến sau: đột biến cảm ứng và đột biến tự phát; đột biến nhằm nghĩa và vô nghĩa; đột biến thay thế và đột biến dịch khung; đột biến trung tính và đột biến có hại. Cho ví dụ.
12. Tại sao trong số các đột biến thay thế base bất gặp chủ yếu các dạng đồng hoán hơn là dị hoán? Giải thích cơ sở của một số bệnh di truyền liên quan các đột biến hemoglobin.
13. Hãy cho biết các hệ thống sửa chữa DNA của tế bào. Phân biệt các cơ chế sửa chữa ngoài sáng và trong tối đối với dimer thymin.
14. Các enzyme giới hạn và các vector plasmid có những tính chất và đặc điểm nào mà chúng được xem là các công cụ đắc lực của công nghệ

DNA tái tổ hợp? Trình bày một quy trình tạo dòng gene ở vi khuẩn, và cho biết các ứng dụng của nó.

Tài liệu Tham khảo

Tiếng Việt

Bùi Chí Bửu, Nguyễn Thị Lang. 1999. Ứng dụng DNA marker trong đánh giá quỹ gene cây lúa. Trong: *Báo cáo khoa học - Proceedings*: Hội nghị CNSH toàn quốc, Hà Nội 1999. NXB KH & KT, tr.1216-1225.

Phạm Thị Trân Châu (Chủ biên), Trần Thị Áng. 1992. *Hoá sinh học*. NXB Giáo Dục.

Nông Văn Hải và cs. 1999. Ứng dụng công nghệ DNA trong nghiên cứu tài nguyên động vật và thực vật Việt Nam. Trong: *Báo cáo khoa học (tlđd)*, tr.1197-1204.

Trần Thị Hoà, Triest L. 1999. Sử dụng kỹ thuật PCR-RAPD trong nghiên cứu đa hình di truyền ở thực vật. Trong: *Báo cáo khoa học (tlđd)*; tr.1305-1310.

Kimura. 1983. *Thuyết tiến hoá phân tử trung tính* (Bản dịch của Hoàng Trọng Phán). NXB Thuận Hoá, Huế. (Phần IV, tr.34-40)

Nguyễn Bá Lộc. 2004. *Giáo trình Axit nucleic và Sinh tổng hợp protein*. Trung tâm Đào tạo Từ xa, Đại học Huế.

Nguyễn Hoàng Lộc, Suzuki A, Takaiwa F. 1999. Tạo dòng phân tử và biểu hiện các gen liên quan *Myb* ở hạt lúa. Trong: *Báo cáo khoa học (tlđd)*; tr.1266-1273.

Lã Tuấn Nghĩa. 2001. *Nghiên cứu sự đa dạng di truyền quần thể nấm đạo ôn (Pyricularia oryzae cavara) và nguồn gen kháng bệnh ở một số giống lúa phục vụ cho công tác chọn tạo giống*. Luận án Tiến sĩ Nông nghiệp. Viện KH-KT Nông nghiệp Việt Nam.

Hoàng Trọng Phán. 1997. *Di truyền học Phân tử*. NXB Giáo Dục.

Nguyễn Đức Thành, Phan Thị Bầy, Lê Hồng Điệp. 1999. Phát triển và ứng dụng các chỉ thị phân tử trong nghiên cứu đa dạng phân tử ở lúa. Trong: *Báo cáo khoa học (tlđd)*; tr.1205-1215.

Nguyễn Tiến Thắng (Chủ biên), Nguyễn Đình Huyền. 1998. *Giáo trình Sinh hoá hiện đại*. NXB Giáo Dục.

Hoàng Văn Tiến (chủ biên), Lê Khắc Thận, Lê Doãn Diên. 1997. *Sinh hoá học với cơ sở khoa học của công nghệ gene*. NXB Nông Nghiệp, Hà Nội.

Tiếng Anh

Campbell PN, Smith AD, Peters TJ. 2005. *Biochemistry illustrated - Biochemistry and molecular biology in the post-genomic era*. 5th ed., Elsevier Limited, Edinburgh - London - New York - Oxford - Philadelphia - St Louis - Sydney - Toronto. (www.elsevierhealth.com)

Do Huu At, BH Thuy, NV Bich, TD Quy, NM Cong. 2000. The use of induced mutation combined with crossing in high quality rice breeding. In: *Seminar on Methodology for plant mutation breeding for quality effective use of physical/chemical mutagens* (for regional nuclear cooperation in Asia). Oct 9-13, 2000, Hanoi. pp76-81.

Bastia D, Manna AC, Sahoo T. 1997. Termination of DNA replication in prokaryotic chromosomes. In: *Genetic Engineering, Vol. 19*. (Setlow JK ed.) pp 101-119. Plenum Press, New York, USA.

Blackburn G.M., Gait M.J. (Eds., 1996): *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford.

Borek C. 2002. *Telomere Control and Cellular Aging*:

[http://www.lef.org/magazine/mag2002/oct2002 report telo 01.html](http://www.lef.org/magazine/mag2002/oct2002%20report%20telo%2001.html).

Greider CW and Blackburn EH. 1996. Telomere, Telomerase and Cancer. *Scientific American*, 2/96, p.92: <http://www.genethik.de/telomerase.htm>.

Hayflick L. 1997. *Mortality and Immortality at the Cellular Level. A Review*. Copyright 2005BioProtNetwork:webmaster@protein.bio.msu.su.

Kelman Z, O'Donell M. 1994. DNA replication: enzymology and mechanisms. In: *Current Opinion in Genetics & Development* (Stillman B and Green M, eds.), Vol.4(2): 185-195. Current Biology Ltd, UK.

Lingner J and Cesh TR. 1998. Telomerase and chromosome end maintenance. In: *Current Opinion in Genetics & Development* (Allis CD and Gasser SM, eds.), Vol.8(2): 226-232. Current Biology Ltd, UK.

Mullis KB. 1990. The unusual origin of the Polymerase chain reaction. *Scientific American* July, 56-65.

McKane L., Kandel J. 1996. *Microbiology: Essentials and Applications*. International edition, 2nd edn., McGraw-Hill, Inc., New York.

Shay J, Schneider E, Masoro EJ. 2004. *Is there a genetic clock for aging?:* www.infoaging.org/b-tel-home.html.

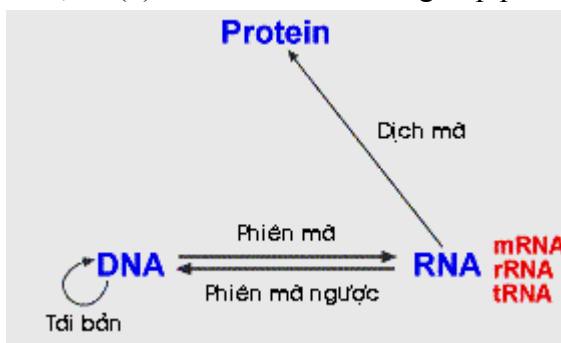
Twyman RM. 1998. *Advanced Molecular Biology*. BIOS Scientific Publishers Ltd/ Springer-Verlag Singapore Pte Ltd.

Weaver RF, Hedrick PW. 1997. *Genetics*. 3rd ed, McGraw-Hill Companies, Inc. Wm.C.Browm Publishers, Dubuque, IA.

Chương 6

Sinh tổng hợp RNA và Protein

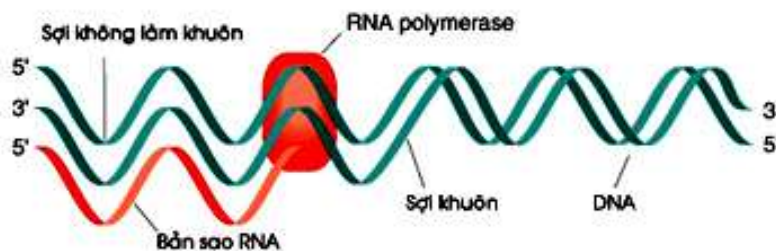
Các quá trình tái bản DNA và phiên mã ngược ở bộ gene RNA của một số virus đã được xem xét ở chương trước. Trong chương này, chúng ta sẽ lần lượt tìm hiểu về protein và các quá trình biểu hiện gene nhằm hoàn tất những hiểu biết cơ bản về các nguyên lý di truyền ở cấp độ phân tử được nêu ra ở sơ đồ bên dưới, đó là: (i) sinh tổng hợp RNA hay phiên mã (transcription) và sơ lược về sửa đổi sau phiên mã; (ii) cấu trúc và chức năng của protein; (iii) bản chất của mã di truyền; (iv) sinh tổng hợp protein hay dịch mã; và (v) sự điều hoà sinh tổng hợp protein ở vi khuẩn.



I. Sinh tổng hợp RNA (Phiên mã)

1. Đặc điểm chung của phiên mã

Phiên mã (transcription) là quá trình tổng hợp các RNA khác nhau từ thông tin di truyền chứa đựng trong DNA. Trừ các gene mã hóa protein trong các operon ở vi khuẩn, nói chung, các RNA mới được tổng hợp chỉ là các *bản sao sơ cấp* (primary transcript) gọi là các pre-RNA. Các pre-RNA này phải trải qua một quá trình sửa đổi để trở thành các RNA trưởng thành (mature) trước khi tham gia vào quá trình sinh tổng hợp protein của tế bào.



Hình 6.1 Sự tổng hợp RNA trên một sợi khuôn của gene (DNA) dưới tác dụng của RNA polymerase.

Quá trình phiên mã DNA các đặc điểm chung sau đây (hình 6.1).

(i) Diễn ra dưới tác dụng của các enzyme *RNA polymerase*.

(ii) Vùng DNA chứa gene được mở xoắn cục bộ, và chỉ một sợi đơn gọi là *sợi có nghĩa* (sense strand) được dùng làm *khuôn* (template) cho tổng hợp RNA.

(iii) Phản ứng tổng hợp RNA diễn ra theo *nguyên tắc bổ sung* và được kéo dài theo chiều 5'→3', ngược với chiều của sợi khuôn như sau:

Sợi khuôn: 3'-A-T-G-C-5'

Sợi RNA: 5'-U-A-C-G-3'

(iv) Nguyên liệu cho tổng hợp là bốn loại *ribonucleoside triphosphate*: ATP, UTP, GTP và CTP.

(v) Sản phẩm của phiên mã là các *RNA sợi đơn* (single RNAs).

(vi) Sự khởi đầu và kết thúc phiên mã phụ thuộc vào các tín hiệu điều hoà là các trình tự DNA đặc thù nằm trước và sau gene được phiên mã.

(vii) Quá trình phiên mã có thể chia làm ba bước: *Mở đầu* (initiation) là sự tương tác giữa RNA polymerase với vùng promoter nhằm xác định sợi khuôn của gene và tổng hợp vài nucleotide; *Kéo dài* (elongation) là giai đoạn sinh trưởng tiếp tục của chuỗi RNA dọc theo sợi khuôn cho đến cuối gene; và *Kết thúc* phiên mã (termination) đặc trưng bằng sự giải phóng sợi RNA và RNA polymerase ra khỏi khuôn DNA.

2. Các RNA polymerase ở prokaryote và eukaryote

Bảng 6.1 Các thành phần cấu trúc của RNA polymerase prokaryote

Tiểu đơn vị	Số lượng	Vai trò
α	2	chưa rõ
β	1	hình thành liên kết phosphodiester
β'	1	bám khuôn DNA
σ	1	nhận biết promoter và xúc tiến khởi đầu phiên mã
$\alpha_2\beta\beta'\sigma$	\leftrightarrow	$\alpha_2\beta\beta'$ + σ
Holoenzyme		Lõi enzyme Yếu tố sigma

Ở các prokaryote, đại diện là *E. coli*, RNA polymerase hoàn chỉnh (holoenzyme) là một protein nhiều tiểu đơn vị, gồm hai phần chính là yếu tố sigma (σ) và lõi enzyme. *Yếu tố sigma* (sigma factor) giúp cho RNA polymerase nhận biết và bám chặt vào promoter để có thể bắt đầu phiên mã tại vị trí chính xác, và *lõi enzyme* (polymerase core) đóng vai trò chính

trong tổng hợp sợi RNA. Tất cả các lớp RNA ở *E. coli* đều được phiên mã bởi chỉ một RNA polymerase (Bảng 6.1).

Ở các eukaryote, có ba loại RNA polymerase I, II và III với sự phân bố và chức năng chuyên hóa khác nhau đối với bộ gene trong nhân, như sau: *RNA polymerase I* ở trong hạch nhân (nucleolus) phiên mã phức hợp gene rRNA cho sản phẩm gồm các rRNA 18S, 28S và 5,8S; *RNA polymerase II* có trong dịch nhân (nucleoplasm) phiên mã các gene mã hóa protein cho sản phẩm là các hnRNA - tiền chất của các mRNA và cả gene cho các kiểu snRNA; *RNA polymerase III* có trong dịch nhân phiên mã các gene tRNA, rRNA 5S, và cả snRNA U6. Ngoài ra, RNA polymerase ty thể ở trong ty thể và chịu trách nhiệm tổng hợp tất cả các RNA của ty thể (Bảng 6.2).

Bảng 6.2 Các RNA polymerase của người

Polymerase	Định khu	Sản phẩm
RNA polymerase I	hạch nhân	các rRNA 18S, 28S và 5,8S
RNA polymerase II	dịch nhân	các hnRNA/mRNA, các snRNA U1, U2, U3, U4, U5
RNA polymerase III	dịch nhân	các tRNA, RNA 5S, snRNA U6, RNA 7S (hay 7SL)
RNA polymerase ty thể	ty thể	tất cả các RNA của ty thể
Độ miễn cảm của các RNA polymerase nhân với α-amanitin* (* Đây là một octapeptide vòng từ nấm độc <i>Amanita phalloides</i>)		
RNA polymerase I	kháng	
RNA polymerase II	độ miễn cảm cao (bám với $K = 10^{-8}M$)	
RNA polymerase III	độ miễn cảm thấp (bám với $K = 10^{-6}M$)	

3. Cơ chế phiên mã ở prokaryote và eukaryote

3.1. Các promoter ở các prokaryote và eukaryote

Các vùng khởi động (promoter) nói chung nằm kề trước gene và có chứa các đoạn trình tự đặc thù cho phép RNA polymerase nhận biết và bám chặt vào để khởi đầu phiên mã tại vị trí chính xác trên sợi khuôn của gene. Vấn đề này tương đối phức tạp ở các eukaryote, vì vậy ở đây ta chỉ xét các promoter của các gene mã hóa protein mà không đề cập các loại promoter của các gene mã hóa các RNA khác. Các promoter của các gene mã hóa protein eukaryote và của operon vi khuẩn nói chung có cấu trúc khá tương đồng nhau. Đoạn trình tự quan trọng nhất của promoter được gọi là *hộp*

RNA polymerase nhất định, được gọi là các *trình tự điều hòa* (consensus sequences). Các đột biến thay thế base tại các hộp TATA, nghĩa là làm cho nó bớt giống với trình tự được bảo tồn (ví dụ, TATAAAT → TGTAAAT) do đó sẽ làm yếu khả năng phiên mã của promoter (*down mutation*). Ngược lại, các đột biến làm cho các trình tự promoter trở nên giống với các trình tự điều hòa (ví dụ, TATCTT → TATAAAT), sẽ làm mạnh khả năng phiên mã của promoter (*up mutation*).

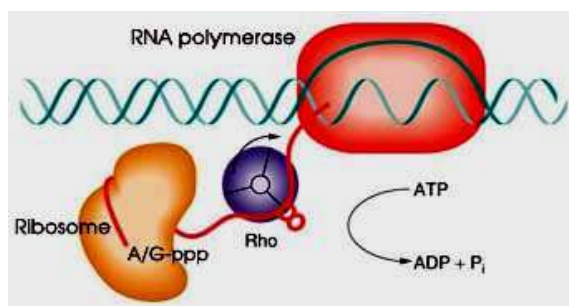
3.2. Các giai đoạn của quá trình phiên mã

Ở đây chỉ đề cập một mô hình đơn giản về quá trình phiên mã ở *E. coli*.

- *Giai đoạn bám và khởi đầu*: Trước tiên, enzyme RNA polymerase hoàn chỉnh (holoenzyme) nhận biết và bám chặt vào promoter sẽ tháo xoắn một vùng có kích thước khoảng 12 cặp base. Sau khi RNA polymerase hoàn chỉnh tổng hợp được một vài nucleotide, nhân tố sigma tách ra để đi vào một chu kỳ phiên mã khác, gọi là *chu kỳ sigma* (sigma cycle).

- *Giai đoạn kéo dài*: Enzyme lõi tiến hành kéo dài sợi RNA dọc theo sợi khuôn. RNA polymerase lõi tiến đến đâu thì DNA được mở xoắn và phiên mã đến đấy; và vùng DNA đã được phiên mã đóng xoắn trở lại.

- *Giai đoạn kết thúc*: Khi quá trình phiên mã tổng hợp xong hai đoạn kết thúc giàu GC và AT nằm đằng sau gene thì tại vùng đuôi sợi RNA hình thành cấu trúc "kẹp tóc" làm dừng sự phiên mã của lõi RNA polymerase. Sau đó, dưới tác dụng của nhân tố *rho* (ρ) có bản chất protein, sợi RNA vừa được tổng hợp và enzyme lõi được giải phóng ra khỏi DNA khuôn.



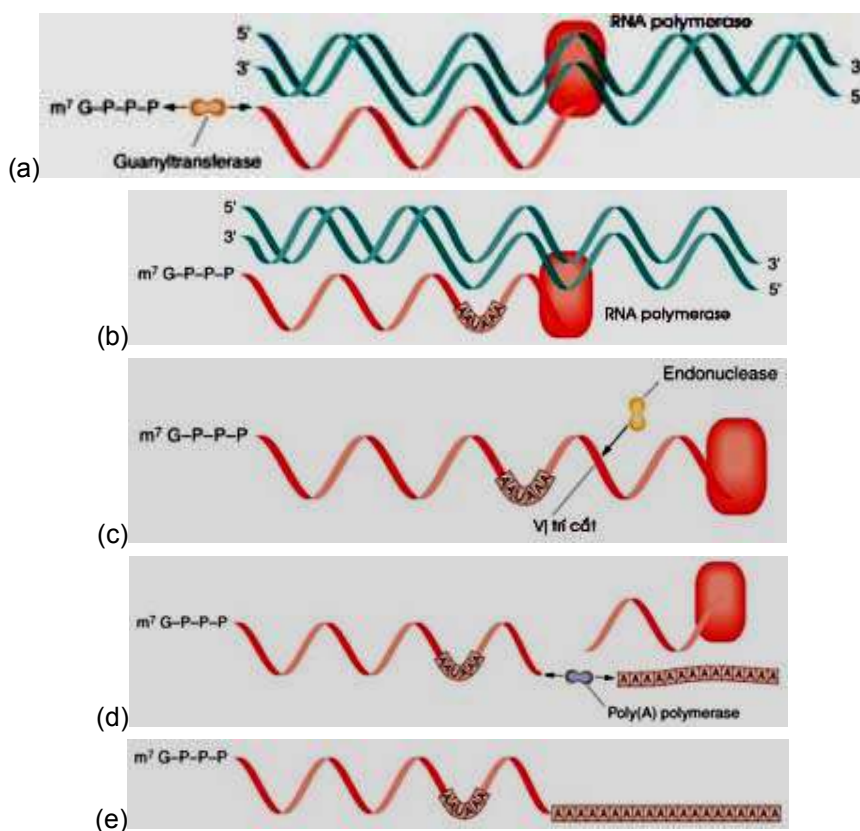
Hình 6.3 Phiên mã ở *E. coli*, với sự hình thành cấu trúc "kẹp tóc" và tham gia của yếu tố rho ở giai đoạn kết thúc phiên mã. Lưu ý, ở vi khuẩn sự dịch mã mRNA được bắt đầu trong khi phiên mã đang còn tiếp diễn.

* Về cơ chế tổng hợp RNA, cần lưu ý một số điểm sau: (i) tổng hợp RNA thường được khởi đầu bằng nucleotide đầu tiên là ATP hoặc GTP; (ii) chuỗi RNA được tổng hợp theo hướng 5' đến 3'; (3) Việc kết thúc một số bản sao cần tới *nhân tố Rho*; và (4) nói chung là rất giống với cơ chế tổng hợp DNA bởi DNA polymerase (Hình 6.3).

3.3. Sự sửa đổi sau phiên mã đối với các pre-mRNA của eukaryote

Như đã đề cập, trừ mRNA prokaryote ra, tất cả các RNA còn lại dù ở pro- hay eukaryote đều phải trải qua quá trình sửa đổi sau phiên mã với rất nhiều cơ chế tinh vi và phức tạp khác nhau để tạo ra các RNA trưởng thành tham gia vào quá trình dịch mã; ở eukaryote, các quá trình này xảy ra trong nhân. Để có cái nhìn hệ thống, ở đây ta hãy tìm hiểu một ít về các cơ chế hoàn thiện các bản sao sơ cấp mRNA (pre-mRNA) ở các tế bào eukaryote.

3.3.1. Gắn thêm "mũ" m⁷Gppp và "đuôi" poly(A)

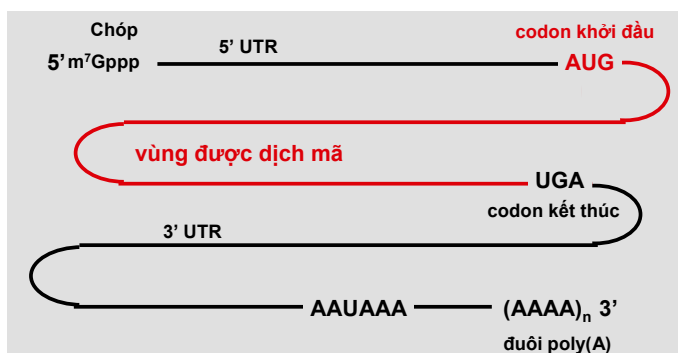


Hình 6.4 Quá trình tổng hợp pre-mRNA ở eukaryote và lắp thêm "mũ" m⁷Gppp và đuôi poly(A) vào các đầu 5' và 3' của nó (xem giải thích trong bài).

Để trở thành phân tử mRNA trưởng thành trước khi đi ra tế bào chất làm khuôn cho dịch mã, tất cả các pre-mRNA của các gene mã hóa protein khác nhau ở tế bào eukaryote, đều được lắp thêm "mũ" (cap) là 7-methylguanosinetriphosphate (m⁷Gppp) vào đầu 5' và "đuôi" poly(A) vào đầu 3' (Hình 6.4 và 6.5). Đối với các gene mã hóa protein có vùng mã hóa là liên tục (không bị gián đoạn bởi các exon) như các gene histone chẳng hạn,

quá trình hoàn thiện mRNA dừng lại tại đây. Loại này chiếm khoảng 10%. Nói chung, tất cả các mRNA trưởng thành của eukaryote đều có mũ 5' và hầu hết có đuôi poly(A), ngoại trừ các mRNA của các histone.

- **Sự hình thành mũ 5':** Quá trình này đòi hỏi nhiều bước, bước thứ nhất xảy ra ngay lập tức sau khởi đầu phiên mã. Nucleotide khởi đầu có một đầu 5'-ribose được thủy phân thành một đầu monophosphate và pyrophosphate vô cơ (PPi). Sau đó enzyme *guanylyltransferase* chuyển một gốc G cho đầu monophosphate bằng cách sử dụng GTP như là một cơ chất, làm giải phóng phosphate vô cơ (Pi). Liên kết được hình thành là một liên kết 5'-5' triphosphate. Tiếp đó enzyme *methyltransferase* gắn nhóm methyl (-CH₃) vào vị trí 7 của guanosine ở đầu mũ, tạo thành 7-methyl guanosine (7mG) hay nói đầy đủ là methyl-7-guanosine triphosphate (m⁷Gppp). Kế đó, xảy ra sự *methyl hoá* (methylation) ở hai nucleotide tại các vị trí 2' ribose. Mũ 5' có chức năng bảo vệ đầu 5' của mRNA (gia tăng tính ổn định của mRNA) và cần thiết cho sự khởi đầu tổng hợp protein.



Hình 6.5 Cấu trúc điển hình của một mRNA trưởng thành ở eukaryote. Ở đây cho thấy đầy đủ các vị trí khác nhau, tính từ đầu 5' như sau: (1) mũ m⁷Gppp + vùng 5'-UTR; (2) Vùng được dịch mã giới hạn bởi codon khởi đầu AUG và codon kết thúc (UGA, UAG hoặc UAA); và (3) Vùng 3'-UTR mà sau nó là vị trí polyadenyl hoá và đuôi poly(A) dài 150-200 base.

- **Sự hình thành đuôi poly(A):** Cần lưu ý rằng, ở đầu 3' của hầu hết các gene mã hóa protein eukaryote có chứa trình tự AATAAA đóng vai trò là tín hiệu cho việc gắn "đuôi" poly(A) vào đầu 3' của mRNA. Sự phiên mã thông thường vẫn còn tiếp diễn sau khi đi qua vị trí *polyadenyl hoá* này (polyadenylation site). Trình tự tương ứng ở vùng cuối 3' mRNA được phiên mã là AAUAAA báo hiệu cho *endonuclease* nhận biết và cắt chuỗi RNA tại một điểm xác định nằm sau nó khoảng 10-30 base. Sau đó, một enzyme khác là *poly(A)-polymerase* sẽ lắp thêm vào đầu 3' của mRNA một dãy adenine dài khoảng 150-200 base gọi là *đuôi poly -A* (Hình 6.5). Đuôi poly(A) có chức năng bảo vệ mRNA khỏi bị suy thoái; và trong nhiều

trường hợp nó còn kích thích cả sự dịch mã (Weaver và Hedrick 1997).

Nói chung, tất cả các mRNA trưởng thành của eukaryote đều có mũ 5' (5'cap) và tất cả các mRNA (ngoại trừ các mRNA histone) đều chứa đuôi poly(A). Mũ 5' và đuôi poly(A) ngăn cản sự suy thoái của mRNA.

3.3.2. Sự cắt-nối đối với pre-mRNA của các gene phân đoạn

Gene phân đoạn (split genes) hay *gene khảm* (mosaic genes) là những gene mã hoá protein mà bên trong trình tự mã hoá của chúng gồm các đoạn không mã hoá (gọi là các *intron*) nằm xen kẽ với các đoạn mã hoá (gọi là các *exon*). Kiểu tổ chức đặc trưng này của các gene mã hoá protein được khám phá đầu tiên vào năm 1977 bởi Richard J. Roberts và Phillip A. Sharp ở adenovirus - một loại virus lây nhiễm các tế bào sinh vật bậc cao. Tuy nhiên, sau này người ta thấy các gene phân đoạn có mặt phổ biến trong các bộ gene của các eukaryote, và gần đây thấy chúng có mặt cả ở các vi khuẩn cổ (archaea hay archaeobacteria).

Việc lý giải cơ chế *cắt-nối* (splicing) trong quá trình xử lý bản sao pre-mRNA dựa chủ yếu trên hai sự kiện sau (Hình 6.6 và 6.7):

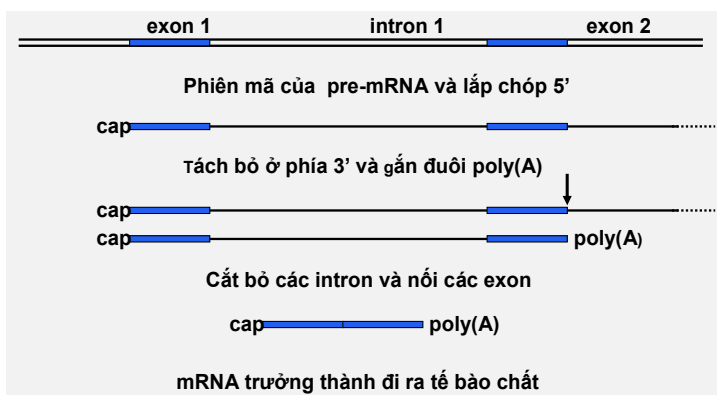
(i) Kết quả phân tích trình tự base tại các chỗ tiếp giáp exon/intron (vị trí cho) và intron/exon (vị trí nhận) cho thấy: ở hai đầu mút của mỗi intron có hai nucleotide rất ổn định, gọi là các "trình tự chuẩn", đó là 5'-GU.....AG-3' (Bảng 6.3). Mỗi intron nói chung có độ dài khoảng 10^2 - 10^4 nucleotide. Từ đây nảy sinh câu hỏi: Bằng cách nào bộ máy cắt-nối có thể xác định một cách chính xác và hiệu quả các vị trí cắt nối nếu như xảy ra sự biến đổi trong các vị trí này. Điều đó cũng đặt ra khả năng là sự biến đổi hay đột biến của các vị trí cắt nối này tại các gốc then chốt có thể làm rối loạn chức năng của chúng.

Bảng 6.3 Tần số các base ở mỗi vị trí của các điểm cắt-nối

Các trình tự cho (donor)																
	exon				intron											
%A	30	40	64	9	0	0	62	68	9	17	39	24				
%U	20	7	13	12	0	100	6	12	5	63	22	26				
%C	30	43	12	6	0	0	2	9	2	12	21	29				
%G	19	9	12	73	100	0	29	12	84	9	18	20				
			A	G	<u>G</u>	<u>U</u>	A	A	G	U						
Các trình tự nhận (acceptor)																
	intron												exon			
%A	15	10	10	15	6	15	11	19	12	3	10	25	4	100	0	22 17
%U	51	44	50	53	60	49	49	45	45	57	58	29	31	0	0	8 37
%C	19	25	31	21	24	30	33	28	36	36	28	22	65	0	0	18 22
%G	15	21	10	10	10	6	7	9	7	7	5	24	1	0	100	52 25
	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	A	G	G

Dãy polypyrimidine (Y = U hoặc C; N = nucleotide bất kỳ)

(ii) Ở một số snRNA chứa trong thành phần của enzyme splicing cũng có các trình tự dinucleotide bổ sung với các trình tự chuẩn trong intron. Các snRNA (U1-U6 như đã đề cập ở Bảng 6.2) trong phức hợp snRNP (*small nuclear ribonucleoprotein*) tương tác với các đầu mút của mỗi intron, kéo hai đầu mút xích lại gần nhau tạo ra cấu trúc hình vòng, nhờ đó enzyme tiến hành cắt bỏ các intron và nối các exon lại với nhau.



Hình 6.6 Các bước tổng quát trong quá trình xử lý mRNA (mRNA processing): (1) lắp mũ 5' (xảy ra cùng lúc phiên mã); (2) tách bỏ và gắn đuôi poly(A); và (3) splicing (xảy ra trong nhân trước khi mRNA đi ra tế bào chất)

Hình 6.6 giới thiệu tổng quát ba bước của cơ chế cắt-nối (splicing) pre-mRNA eukaryote và Hình 6.7 mô tả chi tiết của cơ chế này.

Vấn đề đặt ra là sự có mặt của các intron trong các gene phân đoạn như thế có ý nghĩa gì? Bởi vì với lối tổ chức như vậy làm cho trình tự mã hoá của gene bị gián đoạn, và trong quá trình xử lý pre-mRNA của nó có thể xảy ra dù là một sai sót nhỏ cũng đủ để tạo ra mRNA có khung đọc mã bị thay đổi. Người ta cho rằng có khoảng 90% bản sao mRNA bị suy thoái và chỉ để lại khoảng 10% mRNA trưởng thành đi ra tế bào chất.

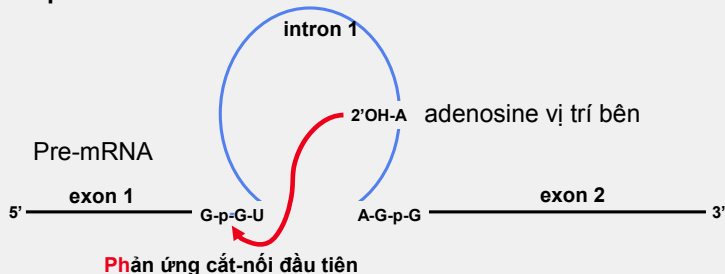
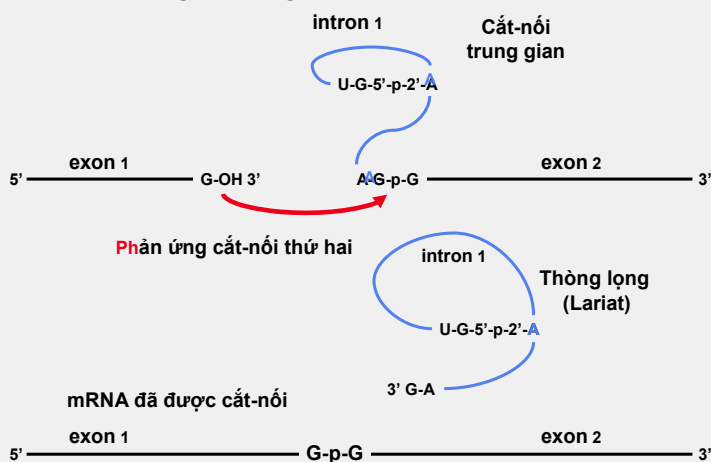
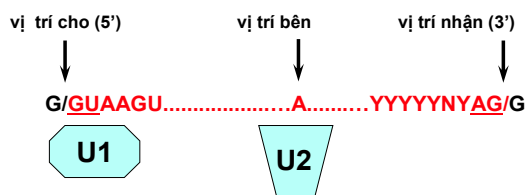
Theo quan niệm hiện nay, các intron có thể có các vai trò sau:

(i) các intron như là các *đoạn đệm* (spacer) tạo thuận lợi cho sự tái tổ hợp trong gene (giữa các exon);

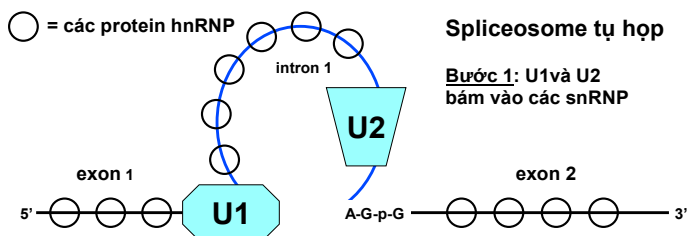
(ii) các intron như là các vùng đệm tách biệt các vùng chức năng của một số protein; và

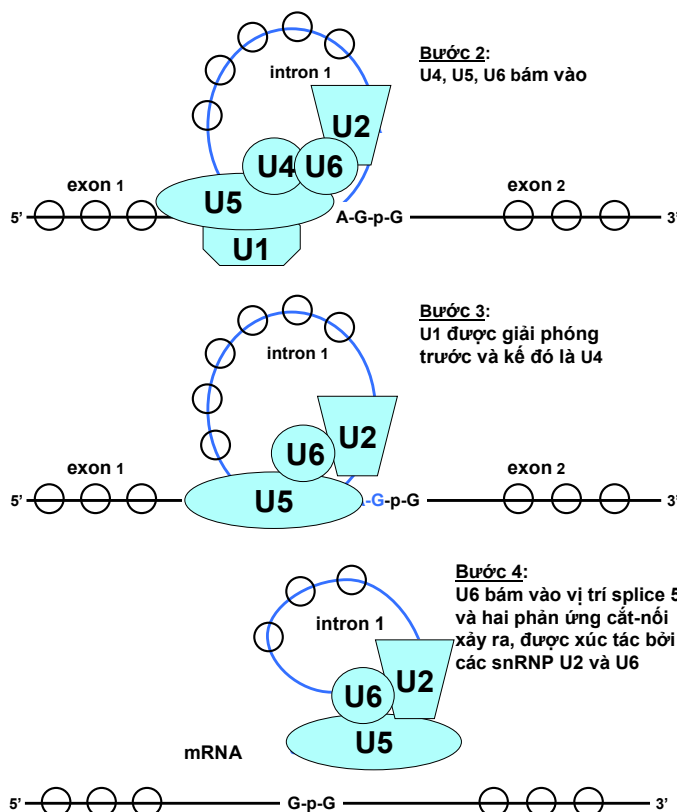
(iii) các intron như là các vùng phân cách cho phép các exon có thể được *cắt-nối có chọn lọc* (alternative splicing) để tạo ra các mRNA trưởng thành khác nhau và dịch mã thành các polypeptide khác nhau từ một gene. Con đường sử dụng exon có chọn lọc này mang tính đặc thù cho từng mô ở các eukaryote bậc cao.

(A)

Hoá học của cắt-nối mRNA(B) **Nối các exon giải phóng intron**(C) **Nhận biết các vị trí cắt-nối**(D) **Spliceosome - sự tụ hợp của bộ máy splicing**

- các snRNA được kết hợp với các protein (snRNP)
- các snRNA splicing - U1, U2, U4, U5, U6





Hình 6.7 Cơ chế chi tiết của quá trình splicing pre-mRNA eukaryote.

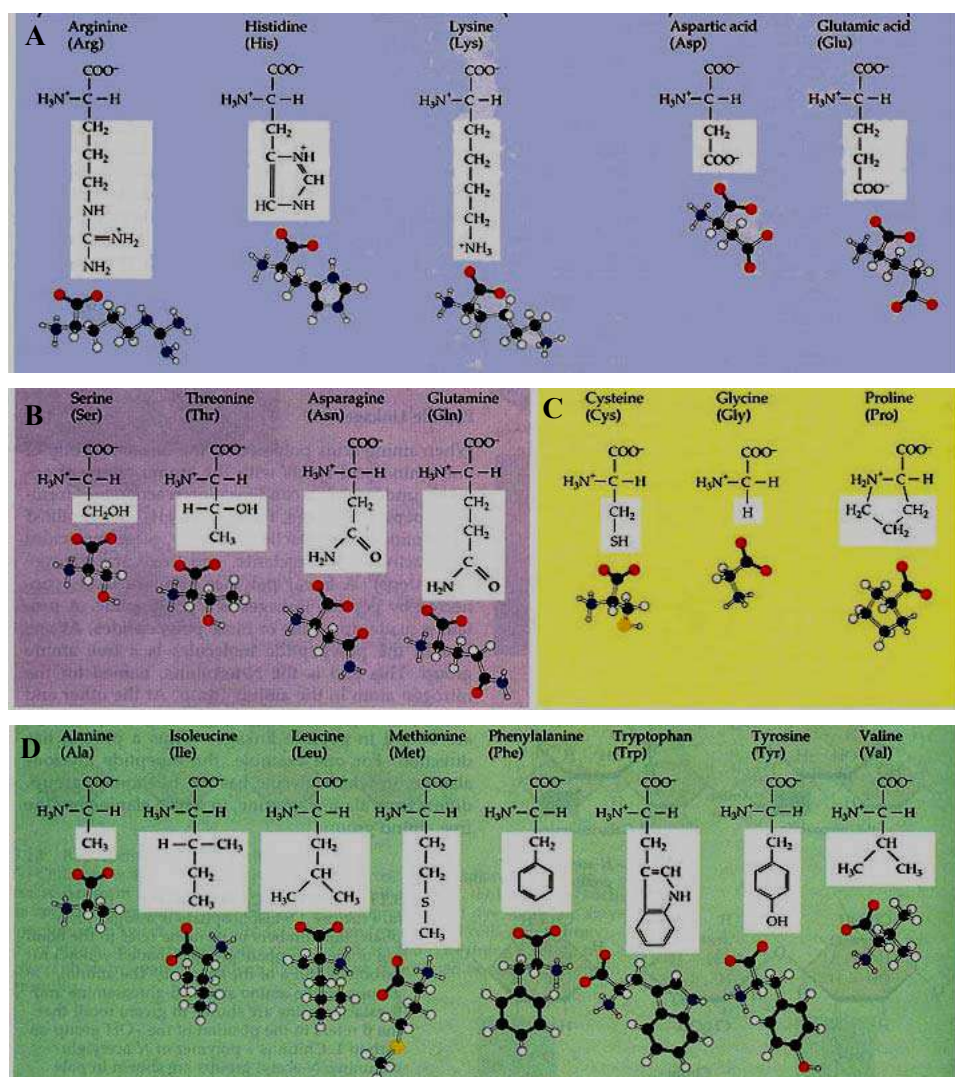
- A.** Bản chất hoá học của sự cắt-nối mRNA gồm hai phản ứng ester hoá chéo - trao đổi một liên kết phosphodiester cho một cái khác – không được xúc tác bởi các enzyme thông thường, chú ý adenosine ở vị trí bên tạo thành liên kết 2', 5' *phosphodiester* với guanosine ở đầu 5' của intron.
- B.** Nối các exon và giải phóng intron dưới dạng RNA vòng (thông lọng: Lariat).
- C.** Dinucleotide GU và AG “chuẩn” ở hai đầu mút mỗi intron cho thấy các vị trí *cho* và *nhận* nằm bên trong các trình tự liên ứng được bảo tồn; snRNA U1 nhận biết vị trí *cho* và snRNA U2 bám vào vị trí bên (theo kiểu tương tác cặp base). Y = U hoặc C chỉ cho pyrimidine; N = nucleotide bất kỳ.
- D.** Sự tụ hợp của bộ máy splicing (spliceosome), với 4 bước: (1) U1 và U2 bám vào các snRNP; (2) Sự bám vào của U4, U5, U6; (3) U1 được giải phóng trước và kế đó là U4; (4) U6 bám vào vị trí splice 5' và hai phản ứng cắt-nối xảy ra, được xúc tác bởi các snRNP U2 và U6.

Chính điều này gây khó khăn thêm cho việc định nghĩa gene cũng như nỗ lực phân loại một cách rành mạch các đơn vị di truyền học, đặc biệt là ở các sinh vật bậc cao. Theo hiểu biết hiện nay, ở các eukaryote, một gene phân đoạn không chỉ xác định một polypeptide mà còn có thể sinh ra nhiều polypeptide khác nhau nhưng có quan hệ với nhau.

II. Cấu trúc và chức năng của protein

1. Cấu trúc của protein

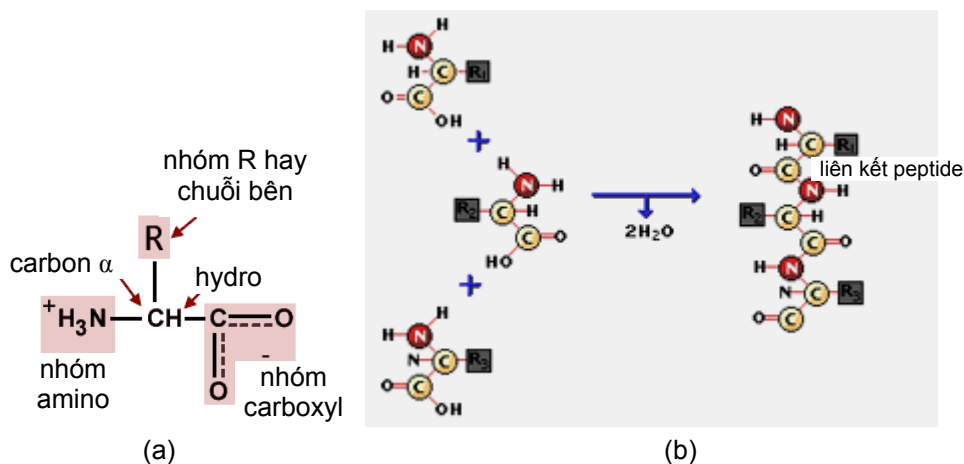
Các protein là những polymer sinh học được tạo ra bởi sự kết nối của các *amino acid* với nhau bằng các *liên kết peptide*. Có 20 loại L- α -amino acid được phát hiện trong các protein của các tế bào (Hình 6.8).



Hình 6.8 Hai mươi loại amino acid phát hiện được trong các protein, với bốn nhóm: A. Các amino acid có chuỗi bên tích điện dương (3 bên trái) và âm (2 bên phải); B. Các amino acid có chuỗi bên không tích điện; C. Các trường hợp đặc biệt; và D. Các amino acid có chuỗi bên kỵ nước.

Về cấu trúc, nói chung, mỗi amino acid gồm có một nguyên tử carbon alpha ($C\alpha$) ở vị trí trung tâm, đính xung quanh nó là một nhóm amin ($-NH_2$), một nhóm carboxyl ($-COOH$), một nguyên tử hydro ($-H$) và một gốc R hay chuỗi bên đặc trưng cho từng loại amino acid (Hình 6.9a).

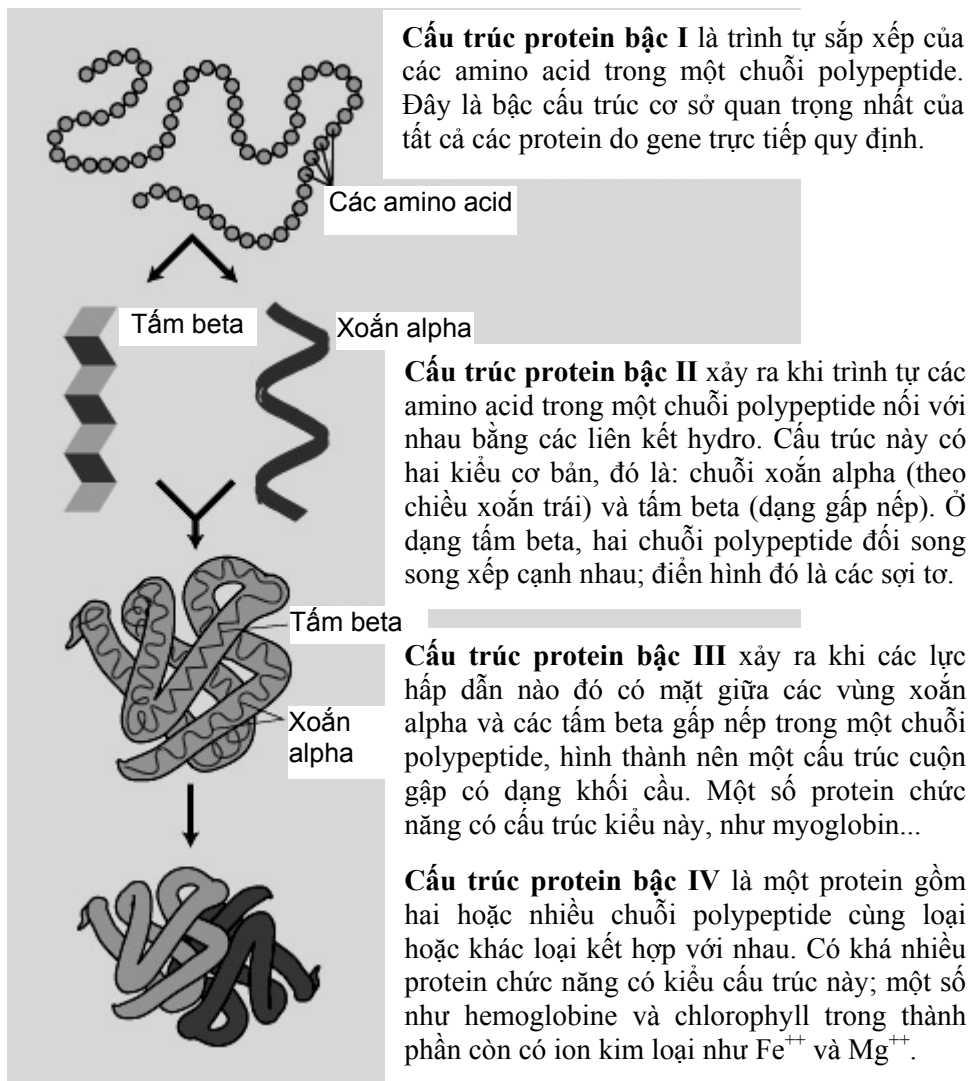
Khi ở trạng thái dung dịch, các nhóm amin và carboxyl thường phân ly thành trạng thái ion, tương ứng là $^+H_3N-$ và $-COO^-$. Hai amino acid nối với nhau bằng một liên kết peptide ($-CO-NH-$) giữa nhóm carboxyl của amino acid này với nhóm amin của amino acid kế tiếp và loại trừ một phân tử nước; cứ như thế các amino acid kết nối với nhau tạo thành một chuỗi gồm nhiều amino acid, thường được gọi là *polypeptide* (Hình 6.9b). Mỗi chuỗi polypeptide luôn luôn có chiều xác định $^+H_3N \rightarrow COO^-$ (do tác dụng của *peptidyl-transferase*) và được đặc trưng về số lượng, thành phần và chủ yếu là trình tự sắp xếp của các amino acid (hay còn gọi là cấu trúc sơ cấp, cấu trúc quan trọng nhất của tất cả các protein do gene quy định).



Hình 6.9 Cấu trúc tổng quát của một amino acid (a) và sự hình thành chuỗi polypeptide mà ở đây là ba amino acid (b).

Có bốn mức độ cấu trúc của các protein được trình bày ở Hình 6.10. Trật tự sắp xếp thẳng hàng của các amino acid tạo thành *cấu trúc bậc I* (primary structure) của protein. Cách thức các amino acid này tương tác với các amino acid lân cận bằng các mối liên kết hydro hình thành nên *cấu trúc bậc II* (secondary structure) của protein; hai dạng phổ biến của cấu trúc bậc II là: *chuỗi xoắn alpha* (α -helix) và *tấm beta* (β -pleated sheet). Còn hình dáng không gian ba chiều của một chuỗi polypeptide chính là *cấu trúc bậc III* (tertiary structure) của nó; hầu hết các protein đều lấy dạng này mà ta gọi là hình cầu (globular). Và nhiều protein có cấu trúc gồm hai hoặc nhiều polypeptide cùng hợp nhất trong một protein phức tạp,

gọi là *cấu trúc bậc IV* (quaternary structure). Đây là mức cấu trúc cao nhất của protein; chúng thường chứa nhiều vùng cấu trúc cuộn chặt gọi là các *domain*, như trong hemoglobin hoặc các kháng thể (Hình 6.11).



Hình 6.10 Bốn bậc cấu trúc của protein.

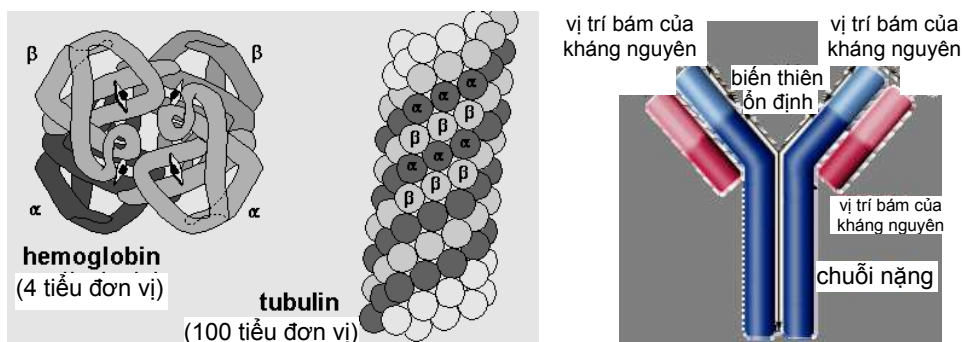
2. Chức năng của protein

Nói chung, protein là các hợp chất hữu cơ vốn là cơ sở của sự sống, với các chức năng thiết yếu sau đây:

(i) Các protein là *thành phần cấu tạo* cơ sở của các tế bào, bao gồm các màng tế bào, các bào quan, bộ máy di truyền của chúng. Đó cũng là

các protein dạng sợi làm thành các cơ quan, bộ phận trên cơ thể các động vật, như: collagen làm nên xương, sụn, gân và da; keratin cấu tạo nên các lớp ngoài cùng của da và tóc, móng, sừng và lông;

(ii) Các *enzyme* đóng vai trò xúc tác cho tất cả các phản ứng hóa học trong tế bào và cơ thể đều là những protein hình cầu. Quan trọng nhất là các enzyme tham gia vào các con đường chuyển hóa và các enzyme tham gia vào các quá trình truyền thông tin di truyền trong tế bào.



Hình 6.11 Cấu trúc bậc IV điển hình của hemoglobin, tubulin và immunoglobulin. Ở đây cho thấy số chuỗi polypeptide và các vùng chức năng đặc trưng của chúng, cụ thể là ở kháng thể immunoglobulin kiểu IgG.

(iii) Các *kháng thể* (antibodies) trong hệ thống miễn dịch, còn gọi là các immunoglobulin, làm ra hàng ngàn protein khác nhau vốn được sinh ra trong huyết thanh máu phản ứng lại với các *kháng nguyên* (antigens). Chúng đóng vai trò bảo vệ cơ thể chống lại sự xâm nhập của các vật lạ.

(iv) Các *hormone* protein bắt nguồn từ các tuyến nội tiết thì không hoạt động như các enzyme. Thay vì kích thích các cơ quan đích, chúng kiểm soát các hoạt động quan trọng, như tốc độ chuyển hóa và sản xuất các enzyme tiêu hóa và sữa chẳng hạn. Insulin (từ tuyến tụy) điều hòa sự chuyển hóa carbohydrate bằng cách kiểm soát các mức glucose trong máu. Thyroglobulin (từ tuyến giáp) điều hòa các quá trình chuyển hóa nói chung; calcitonin cũng từ tuyến giáp làm hạ thấp mức calci máu.

(v) Ngoài ra, các protein còn là *nguồn dinh dưỡng* chính cung cấp năng lượng cho tế bào và cơ thể duy trì các hoạt động trao đổi chất và lớn lên; các protein như *hemoglobin* mang các sinh chất theo máu đi khắp cơ thể; các *fibrinogen* và fibrin được biến đổi từ nó vốn có trong máu cần thiết cho quá trình đông máu. Bên cạnh đó, các *protein cơ* mà chủ yếu là myosin phối hợp với actin tạo thành actomyosin, chịu trách nhiệm cho hoạt động cơ cơ, v.v.

III. Mã di truyền

Gene (DNA) được cấu tạo từ bốn loại nucleotide, trong khi đó protein được cấu tạo bởi 20 loại amino acid. Vấn đề đặt ra là, các gene mã hóa cho các sản phẩm protein của chúng bằng cách nào?

Bằng suy luận, ta có thể suy đoán rằng mỗi amino acid không thể được xác định bởi đơn vị mã gồm một, hai hoặc bốn nucleotide bởi một đẳng còn chưa đủ và một đẳng khác lại quá dư thừa. Có lẽ nó phải là một nhóm gồm ba nucleotide ($4^3 = 64$). Với 64 kiểu bộ ba hoá ra là đủ thừa để mã hoá cho 20 loại amino acid. Như thế, một amino acid được xác định bởi trung bình ba bộ ba khác nhau. Vậy phải chăng mã di truyền là mã bộ ba?

Năm 1961, S.Brenner, F.Crick và L.Barnett đã phân tích chi tiết nhiều thể đột biến của phage T4 nhận được bằng cách xử lý acridin, tác nhân gây các đột biến mất hoặc thêm một cặp base, đã khẳng định mã di truyền là *mã bộ ba* (triplet code) đúng như dự đoán. Như vậy, *đơn vị mã* (coding unit) gồm ba nucleotide xác định một amino acid gọi là *codon*.

1. Giải mã di truyền

Việc tiếp theo là xác định xem mỗi amino acid cụ thể được mã hoá bởi một hoặc một số bộ ba nào. Cũng trong năm 1961, M.Nirenberg và H. Matthaei lần đầu tiên sử dụng mRNA nhân tạo có thành phần base biết trước được tổng hợp bằng enzyme *polynucleotide phosphorylase* (do Ochoa tìm ra năm 1959) và hệ thống tổng hợp là dịch chiết tế bào *E. coli* bao gồm đầy đủ các yếu tố (ribosome, tRNA, amino acid, enzyme, ATP...) cần thiết cho tiến hành giải mã di truyền *in vitro*. Với mRNA chỉ chứa toàn U, poly(U), chuỗi polypeptide sinh ra chỉ chứa toàn phenylalanine (Phe). Điều đó chứng tỏ UUU là bộ ba mã hoá của Phe.

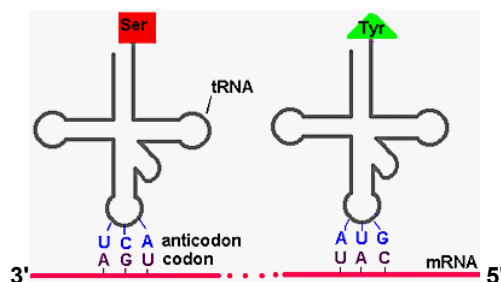
Sau đó, Har Gobind Khorana đã tiến hành các thí nghiệm sử dụng các mRNA tổng hợp có chứa hai, ba hoặc bốn nucleotide được kết nối theo kiểu lặp lại để tiến hành giải mã. Ví dụ: (i) Với mRNA nhân tạo chứa hai base là poly(UC) hay UCUCUC..., sẽ chứa hai codon xen kẽ UCU và CUC (chú ý rằng sự dịch mã *in vitro* khởi đầu tại vị trí ngẫu nhiên). Kết quả là thu được một polypeptide gồm hai amino acid xen kẽ nhau là serin và leucin, poly(Ser-Leu); (ii) Với mRNA tổng hợp gồm các bộ ba lặp lại sẽ được dịch thành các homopolypeptide. Ví dụ, poly(UUC) có thể được đọc là (UUC-UUC), hoặc (UCU-UCU), hoặc (CUU-CUU) tùy thuộc vào vị trí bắt đầu dịch mã. Và kết quả là có ba loại polypeptide được tổng hợp, poly(Phe) hoặc poly(Ser) hoặc poly(Leu) v.v.

Từ các kết quả thu được bằng cách đó Khorana đã xác định được 'nghĩa' của phần lớn các codon có thành phần base không đồng nhất, và

việc giải toàn bộ hệ thống *mã di truyền* (genetic code) được hoàn tất vào tháng 6 năm 1966. Với công lao to lớn đó Khorana và Nirenberg được trao giải thưởng Nobel năm 1968. Từ đây cho phép xây dựng nên bảng mã di truyền (Bảng 6.4) với các đặc tính được trình bày ở dưới đây.

Bảng 6.4 Mã di truyền (cho các codon trên mRNA theo chiều 5'→3')

	U	C	A	G
U	UUU Phenylalanine UUC Phenylalanine UUA Leucine UUG Leucine	UCU Serine UCC Serine UCA Serine UCG Serine	UAU Tyrosine UAC Tyrosine UAA Stop codon UAG Stop codon	UGU Cysteine UGC Cysteine UGA Stop codon UGG Tryptophan
C	CUU Leucine CUC Leucine CUA Leucine CUG Leucine	CCU Proline CCC Proline CCA Proline CCG Proline	CAU Histidine CAC Histidine CAA Glutamine CAG Glutamine	CGU Arginine CGC Arginine CGA Arginine CGG Arginine
A	AUU Isoleucine AUC Isoleucine AUA Isoleucine AUG Methionine; initiation codon	ACU Threonine ACC Threonine ACA Threonine ACG Threonine	AAU Asparagine AAC Asparagine AAA Lysine AAG Lysine	AGU Serine AGC Serine AGA Arginine AGG Arginine
G	GUU Valine GUC Valine GUA Valine GUG Valine	GCU Alanine GCC Alanine GCA Alanine GCG Alanine	GAU Aspartic acid GAC Aspartic acid GAA Glutamic acid GAG Glutamic acid	GGU Glycine GGC Glycine GGA Glycine GGG Glycine



Hình 6.12 Các phân tử tRNA mang amino acid Ser (trái) và Tyr (phải) đọc mã trên mRNA bằng cách khớp anticodon của chúng với codon của mRNA.

2. Các đặc tính của mã di truyền

- Mã di truyền là *mã bộ ba* (triplet code). Các bộ ba của mRNA gọi là *codon* (mã) và bộ ba đặc trưng của tRNA có thể khớp với codon của mRNA theo nguyên tắc bổ sung gọi là *anticodon* (đối mã) (Hình 6.12).

- Mã di truyền *không gối lên nhau* (non-overlapping). Mỗi codon là một đơn vị độc lập, và thông tin của mRNA được đọc lần lượt qua các codon theo chiều 5'→3' bắt đầu từ codon khởi đầu.

- Mã di truyền có tính liên tục, *không bị ngắt quãng* (unpunctuated).

- Mã di truyền có các codon *khởi đầu* (initiation) và *kết thúc* (termination) nằm ở hai đầu 5' và 3' của mRNA đóng vai trò là tín hiệu khởi đầu và kết thúc tổng hợp chuỗi polypeptide.

- Mã di truyền có tính *đơn trị*, rõ ràng (unambiguous). Mỗi codon xác định một amino acid duy nhất, hoặc là tín hiệu xác định sự kết thúc dịch mã.

- Mã di truyền có tính *thoái hóa* (degenerate). Với tất cả 61 *codon có nghĩa* (sense codon) trong khi chỉ có 20 loại amino acid, vì vậy mỗi amino acid (hay tín hiệu kiểm soát dịch mã) có thể được xác định bởi nhiều hơn một codon. Các codon cùng xác định một amino acid như thế gọi là các codon đồng nghĩa; chúng thường khác nhau ở base cuối và base 3' đó được gọi là base thoái hóa. Ví dụ, các amino acid Arg, Ser và Leu mỗi cái có tới sáu codon đồng nghĩa (xem Bảng 6.4).

- Mã di truyền có tính *phổ biến* (universal), chung cho toàn bộ sinh giới.

3. Những ngoại lệ so với mã di truyền "phổ biến"

Bên cạnh tính *phổ biến* (universal) của hệ thống mã di truyền nói trên, các nghiên cứu gần đây cho thấy một vài chệch hướng mà hầu hết là xảy ra trong các bộ gene ty thể (Bảng 6.5).

Bảng 6.5 Các ngoại lệ so với mã "phổ biến"

Nguồn	Codon	Nghĩa phổ biến	Nghĩa mới
Ty thể ruồi giấm	UGA	Kết thúc	Tryptophan
	AGA & AGG	Arginine	Serine
	AUA	Isoleucine	Methionine
Ty thể động vật có vú	AGA & AGG	Arginine	Kết thúc
	AUA	Isoleucine	Methionine
	UGA	Kết thúc	Tryptophan
Ty thể nấm men	CUN ⁽¹⁾	Leucine	Threonine
	AUA	Isoleucine	Methionine
	UGA	Kết thúc	Tryptophan
Ty thể thực vật bậc cao	UGA	Kết thúc	Tryptophan
	CGG	Arginine	Tryptophan
Các nhân của Protozoa ⁽²⁾	UAA & UAG	Kết thúc	Glutamine
<i>Mycoplasma</i>	UGA	Kết thúc	Tryptophan

⁽¹⁾ N = U, C, A hoặc G; ⁽²⁾ Protozoa: Động vật nguyên sinh.

Tuy nhiên, ở các bào quan thực vật, sự *biên tập RNA* (RNA editing) là phổ biến và không rõ ràng, ở chỗ: Liệu phải chăng tất cả các trường hợp sai lệch so với mã phổ biến ở thực vật là các biến đổi thật hay là hậu quả của sự biên tập RNA trước khi dịch mã. Một vài thay đổi thỉnh thoảng cũng xảy ra ở các bộ gene vi khuẩn và bộ gene nhân của các eukaryote, nhưng thường thì liên quan với các codon kết thúc. Sự phân bố phát sinh chủng loại của các thay đổi này chỉ ra rằng mã di truyền vẫn còn tiến hóa.

4. Sự linh hoạt trong việc kết cặp anticodon-codon

Mặc dù có 61 codon có nghĩa nhưng trên thực tế trong mỗi tế bào prokaryote và eukaryote chỉ có khoảng 40 phân tử tRNA khác nhau. Tuy nhiên, trong tế bào chỉ có 20 loại amino acid được xác định bởi mã di truyền, nên một số loại tRNA phải mang cùng một loại amino acid. Các bản sao tRNA như thế có thể có trình tự anticodon giống nhau, trong trường hợp đó chúng có thể thay thế chức năng cho nhau. Các tRNA khác thì mang các trình tự anticodon khác nhau và do vậy nhận biết các codon khác nhau; chúng được gọi là *isoaccepting tRNA*, và độ giàu tương đối của chúng có thể ảnh hưởng tới cách thức sử dụng codon.

Năm 1966, F.Crick đưa ra một cách để giải thích về sự kết cặp "lỏng lẻo" có thể xảy ra ở vị trí thứ ba của các codon đồng nghĩa. Theo Crick, hai base đầu tiên của một codon phải có sự kết cặp chính xác với anticodon (theo nguyên tắc bổ sung), còn base cuối của codon thì có thể "*linh hoạt*" (wobble), ít đặc thù hơn so với vị trí bình thường của nó để hình thành nên sự kết cặp base bất thường với anticodon. Đề nghị này được gọi là *giả thuyết linh hoạt* (wobble hypothesis).

Bảng 6.6 Các nguyên tắc kết cặp linh hoạt anticodon-codon

Base 5' của anticodon	Base 3' của codon
G	C hoặc U
C	G
A	U
U	A hoặc G
I	U, C hoặc A

Crick gợi ý rằng một base G trong anticodon có thể cặp không chỉ với C ở vị trí thứ ba của một codon (vị trí linh hoạt), mà còn với U. Hơn nữa, ông còn lưu ý một trong số các nucleoside bất thường có mặt trong tRNA là *inosine* (I), có cấu trúc tương tự với guanosine. Nucleoside này thông

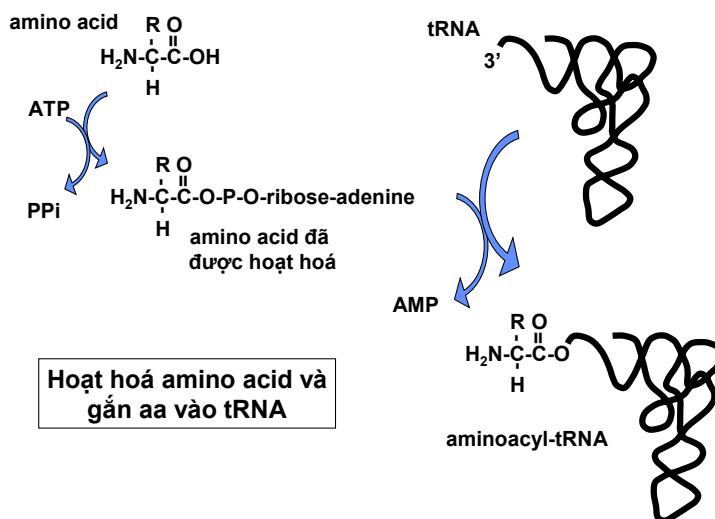
thường có thể kết cặp như G, vì vậy kỳ vọng nó sẽ cặp với C (kết cặp bổ sung) hoặc U (kết cặp linh hoạt) ở vị trí thứ ba của codon. Nhưng ông còn lưu ý rằng I vẫn còn có thể có kiểu kết cặp linh hoạt khác, bây giờ ta biết đó là cặp với A ở vị trí thứ ba của codon. Điều đó có nghĩa là, một anticodon có I ở vị trí thứ nhất về tiềm năng có thể cặp với ba codon khác nhau có base cuối là C, U hoặc A (về chi tiết, xem ở chương 4). Các thí nghiệm sau này khẳng định điều dự đoán của Crick là hoàn toàn đúng và cho thấy các khả năng kết cặp base linh hoạt ở Bảng 6.6.

Rõ ràng là, hiện tượng kết cặp linh hoạt này làm giảm đáng kể số lượng các tRNA cần thiết để dịch mã di truyền. Ví dụ, để dịch mã các codon (5'→3') UUU và UUC mã hoá cho phenylalanine chỉ cần một tRNA^{Phe} mang anticodon (3'→5') AAG.

IV. Cơ chế của quá trình sinh tổng hợp protein (Dịch mã)

Dịch mã (translation) hay tổng hợp protein là một quá trình sinh học quan trọng diễn ra trong tế bào chất, và phụ thuộc vào nhiều yếu tố; quan trọng nhất là các mRNA, tRNA và ribosome. mRNA mang thông tin quy định trình tự kết hợp các amino acid vào chuỗi polypeptide, mà việc dịch mã mRNA được thực hiện bởi các aminoacyl-tRNA, còn ribosome đóng vai trò ổn định việc kết hợp giữa mRNA với các tRNA (Hình 6.14). Quá trình này được chia làm hai giai đoạn dưới đây.

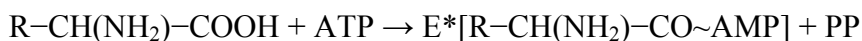
1. Hoạt hoá amino acid



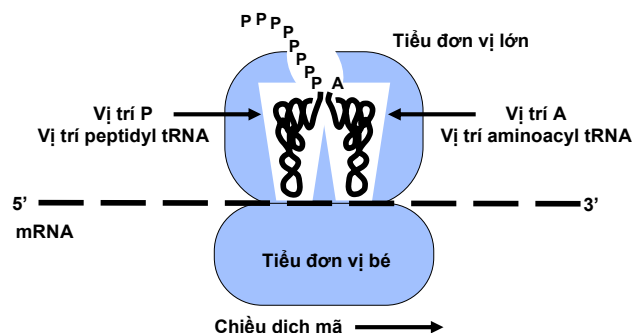
Hình 6.13 Sự hoạt hoá amino acid và gắn amino acid đã hoạt hoá vào tRNA.

Quá trình này diễn ra trong bào tương và tạo nguồn các tRNA mang

các amino acid sẵn sàng tham gia dịch mã. Mỗi amino acid được đính vào tRNA thích hợp nhờ một enzyme *aminoacyl-tRNA synthetase* đặc thù. Trước tiên, enzyme này (E) xúc tác cho phản ứng ATP hoạt hoá amino acid, với sự có mặt của Mg^{2+} , tạo ra phức hợp [aminoacyl-AMP + E].



Tiếp theo, cũng dưới tác dụng của enzyme đó, phức hợp này kết hợp với tRNA thích hợp bằng liên kết đồng hoá trị để tạo ra *aminoacyl-tRNA*.

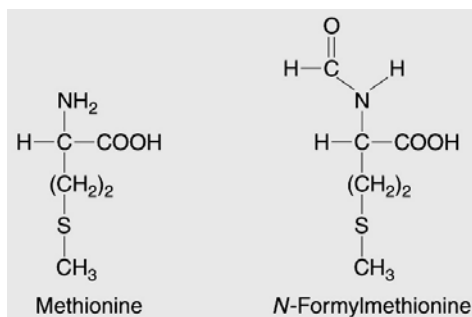


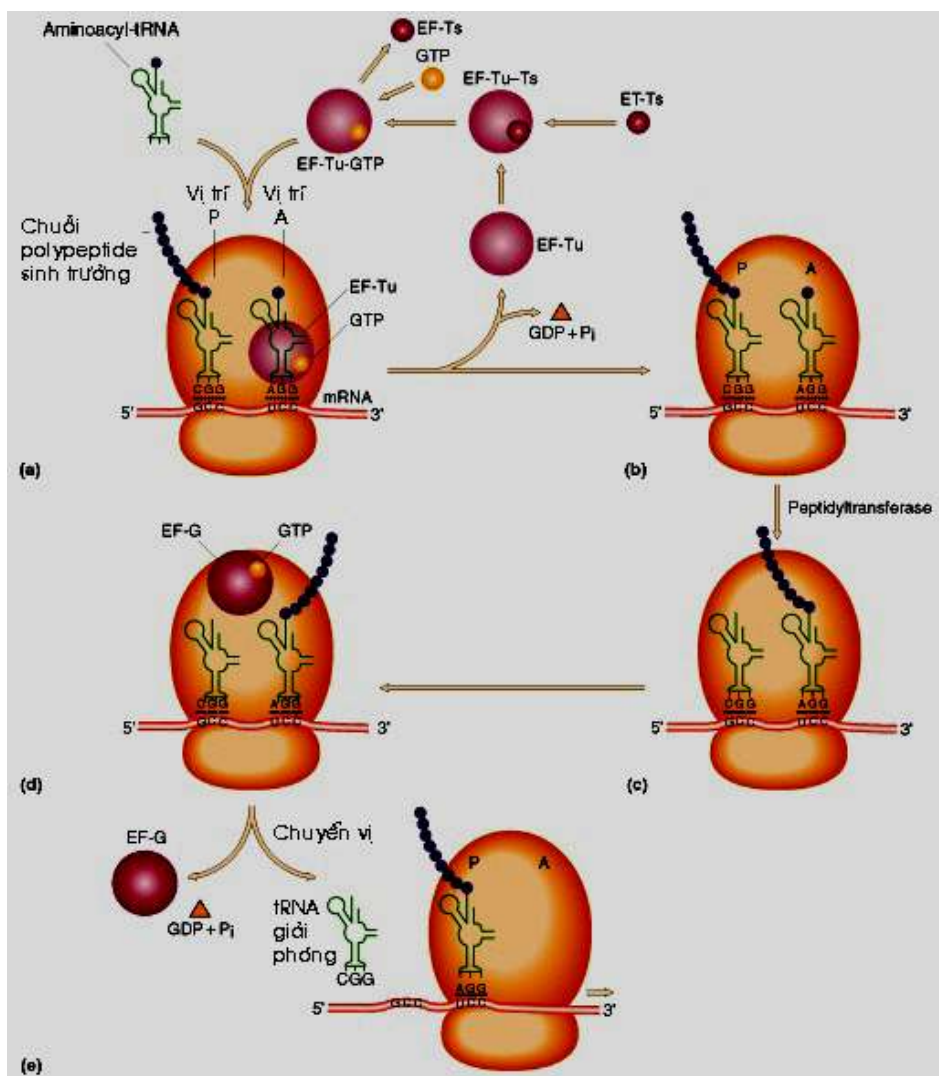
Hình 6.14 Cấu trúc ribosome với sự bám vào của các tRNA và mRNA.

2. Mở đầu, Kéo dài và Kết thúc sự tổng hợp chuỗi polypeptide

Bước 1: Mở đầu (initiation)

Quá trình dịch mã bắt đầu khi một tiểu đơn vị ribosome bé bám vào mRNA tại vị trí của codon khởi đầu AUG. Lúc này một phân tử tRNA khởi đầu đặc thù mang methionine (ở vi khuẩn là formyl-Met; cấu trúc của Met và f-Met được giới thiệu dưới đây) đi vào và khớp anticodon của nó với codon mở đầu của mRNA. Kế đó, tiểu đơn vị ribosome lớn bám vào tiểu đơn vị bé tạo ra một ribosome hoạt động hoàn chỉnh. Lúc này Met-tRNA ở vị trí P và vị trí A để trống; một tRNA thứ hai (aa_2 -tRNA) đi vào vị trí A và khớp với codon thứ hai.



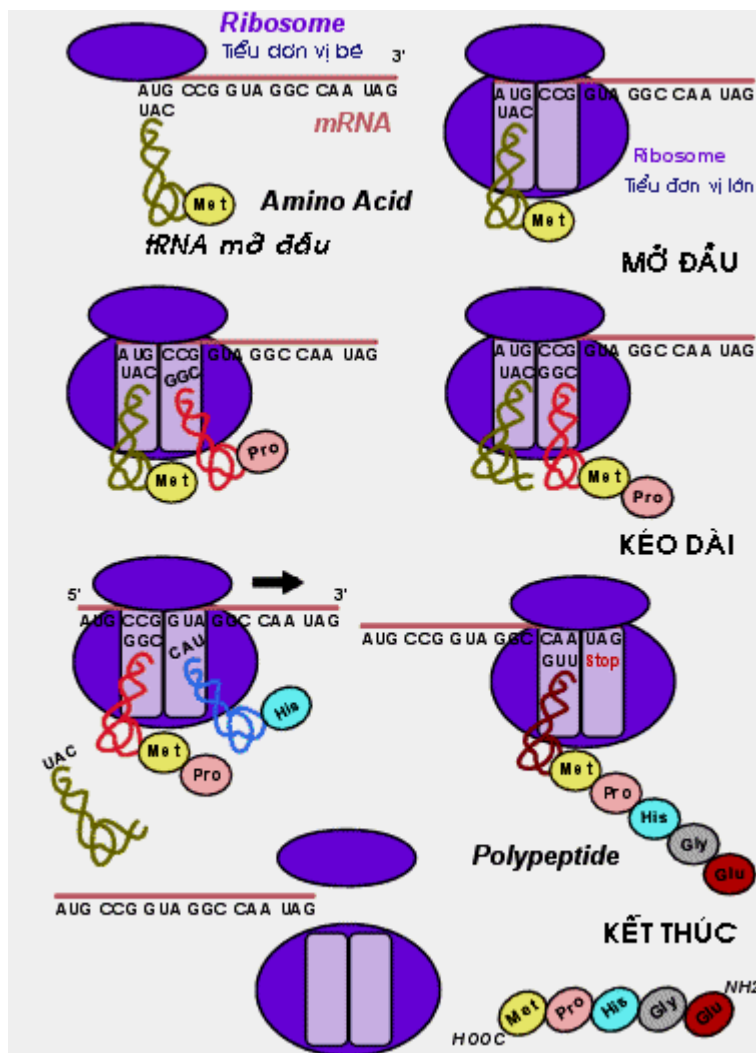


Hình 6.15 Quá trình tổng hợp kéo dài chuỗi polypeptide.

Bước 2: Kéo dài (elongation)

Quá trình kéo dài bắt đầu sau khi liên kết peptide đầu tiên được hình thành. Phản ứng này được xúc tác bởi enzyme *peptidyl transferase*, và kết quả là tạo ra một peptidyl-tRNA ở vị trí A (fMet-aa₂-tRNA). Sau đó, ribosome lập tức chuyển dịch sang một codon mới dọc theo mRNA theo chiều 5'→3'. Phản ứng này đẩy phân tử tRNA tự do vốn ở vị trí P ra ngoài; lúc này peptidyl-tRNA (fMet-aa₂-tRNA) nằm ở vị trí P và vị trí A lại để trống. Một chu kỳ dịch mã mới lại bắt đầu, một aminoacyl-tRNA thứ ba (aa₃-tRNA) đi vào và khớp anticodon của nó với codon đang để

trống ở vị trí A, một liên kết peptid thứ hai được hình thành (fMet-aa₂-aa₃-tRNA), và ribosome lại dịch chuyển sang codon kế tiếp. Quá trình nói trên diễn ra một cách tuần tự dọc theo mRNA theo chu kỳ ba bước nói trên (được minh họa Hình 6.15), và làm cho chuỗi polypeptide dài dần ra cho đến dịch mã xong codon có nghĩa cuối cùng.



Hình 6.16 Các sự kiện cơ bản của quá trình tổng hợp chuỗi polypeptide.

Bước 3: Kết thúc (termination)

Quá trình tổng hợp chuỗi polypeptide sẽ dừng lại khi một codon kết thúc được đưa đối diện với vị trí A để trống, vốn được nhận biết bởi một protein kết thúc gọi là *nhân tố giải phóng* RF (release factor). Sự có mặt

của nó cùng với transferase cắt rời chuỗi polypeptide ra khỏi tRNA cuối cùng và phóng thích hai tiểu đơn vị ribosome cũng như chuỗi polypeptide và tRNA ra khỏi mRNA.

Hình 6.16 tóm tắt các sự kiện cơ bản của ba bước mở đầu, kéo dài và kết thúc sự tổng hợp chuỗi polypeptide.

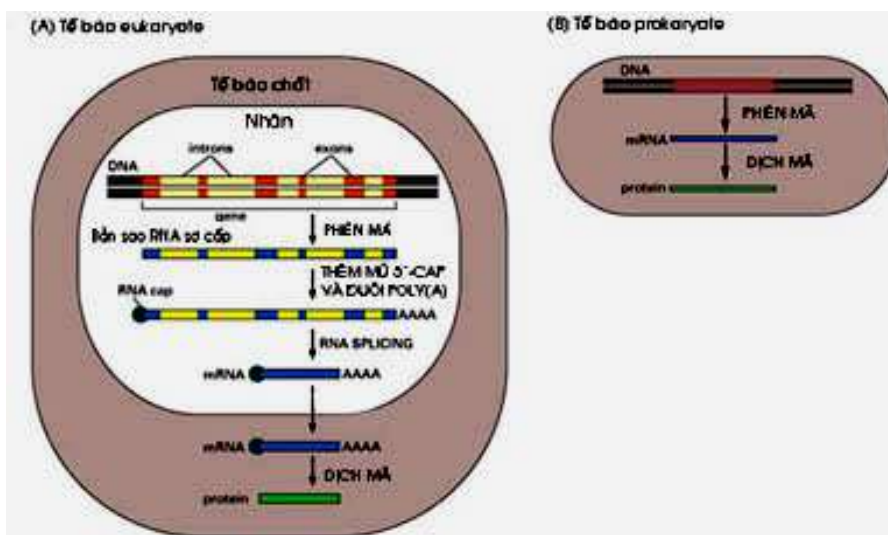
* Một số điểm cần lưu ý thêm:

(1) Trên đây mới chỉ phân tích hoạt động của một ribosome trên mRNA. Thực ra, trên một mRNA có rất nhiều ribosome cùng hoạt động, gọi là polyribosome hay *polysome*, tạo ra nhiều polypeptide giống nhau.

(2) Trên nguyên tắc, amino acid mở đầu sẽ được cắt bỏ khỏi chuỗi polypeptide (sau khi tổng hợp được vài amino acid như trong trường hợp các vi khuẩn) hoặc trước khi chuỗi được tổng hợp đầy đủ (như ở trường hợp eukaryote). Tuy nhiên, ở các eukaryote không phải lúc nào amino acid mở đầu này cũng bị tách bỏ, mà trong một số protein nó vẫn được giữ lại.

(3) Sau khi được tổng hợp, các chuỗi polypeptide sơ cấp này sẽ được sửa đổi và chuyển sang các bậc cấu trúc cao hơn theo cách đặc thù để trở thành các protein hoạt động chức năng.

(4) Tham gia vào các bước mở đầu, kéo dài và kết thúc còn có các yếu tố protein, với tên gọi tương ứng là các *nhân tố mở đầu* (IF: initiation factor), *nhân tố kéo dài* (EF: elongation factor; gồm EF-Tu và EF-G), và *nhân tố giải phóng* (release factor) với GTP và các ion Mg^{2+} , K^+ và NH_4^+ .



Hình 6.17 Các cơ chế tổng hợp và dịch mã mRNA ở các tế bào eukaryote (A) prokaryote (B).

(5) Trong các tế bào prokaryote, do không có màng nhân và các mRNA đa cistron vốn dĩ không phải qua sửa đổi sau phiên mã, cho nên các ribosome và các aminoacyl-tRNA sẽ bám vào đầu 5' của mRNA để bắt đầu quá trình dịch mã ngay trong khi ở đầu 3' của nó quá trình phiên mã đang còn tiếp diễn. Ngược lại, ở các tế bào eukaryote vì có màng nhân phân cách và các pre-mRNA còn phải trải qua các công đoạn sửa đổi phức tạp sau phiên mã (tất cả đều diễn ra trong nhân), còn dịch mã diễn ra sau đó ở trong tế bào chất (hình 6.17). Vì thế cho nên phiên mã và dịch mã rõ ràng là hai quá trình tách biệt nhau cả về cả không gian lẫn thời gian.

V. Sự điều hoà sinh tổng hợp protein ở vi khuẩn

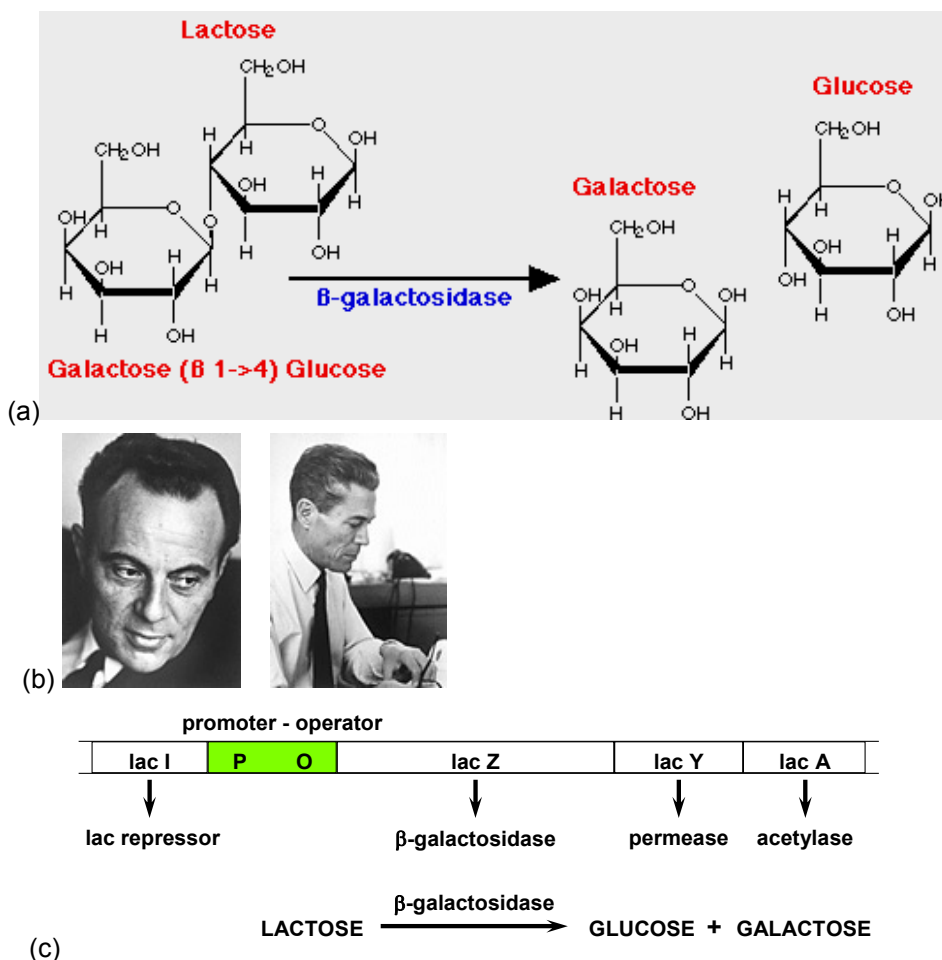
Trên thực tế, các gene không tồn tại riêng rẽ và hoạt động như những thực thể biệt lập với một cường độ ổn định. Trái lại, giữa các gene trong bộ gen có sự kiểm soát lẫn nhau và phụ thuộc vào các điều kiện môi trường. Như vậy mỗi bộ gene của tế bào là một hệ thống mở có khả năng tự điều chỉnh, đảm bảo sự hoạt động của các gene trong từng giai đoạn phát triển của cơ thể diễn ra hợp lý trước điều kiện cụ thể của môi trường. Sự điều hoà hoạt động của các gene biểu hiện ở nhiều mức độ (phiên mã, sau phiên mã, dịch mã và sau dịch mã) theo những cơ chế đặc thù của từng nhóm loài. Trong phần này, chúng ta chỉ tìm hiểu một số cơ chế điều hoà biểu hiện gene vi khuẩn ở mức phiên mã.

1. Mô hình Operon ở E. coli

Phần lớn các gene trong bộ gen vi khuẩn được tổ chức thành các đơn vị hoạt động chức năng đặc trưng, gọi là các *operon*. Các gene cấu trúc trong một operon được điều hoà chung trong quá trình chuyển hoá một hợp chất nhất định của tế bào. Mô hình operon được đưa ra lần đầu tiên là operon lactose (*lac* operon) bởi Francois Jacob và Jacques Monod năm 1961, vốn được nghiên cứu kỹ nhất cho đến nay (Hình 6.18).

Tham gia vào điều hoà hoạt động của một operon gồm có bốn yếu tố thuộc hai thành phần chính: (i) các *locus cấu trúc* (structural loci) và (ii) các *locus điều hoà* (regulatory loci); trong đó nhóm sau bao gồm yếu tố chỉ huy, vùng khởi động và gene điều hoà.

- Một nhóm các *gene cấu trúc* (structural genes) liên quan về mặt chức năng, xếp cạnh nhau, khi phiên mã sẽ tạo ra một phân tử mRNA chung gọi là *mRNA đa cistron* (polycistronic mRNA). Đối với operon-*lac*, đó là ba gene: *lacZ*, *lacY* và *lacA*; trong đó *lacZ* mã hoá β galactosidase (thủy phân lactose thành galactose và glucose), *lacY* xác định permease (vận chuyển lactose qua màng) và *lacA* mã hoá transacetylase.



Hình 6.18 (a) Phân tử đường lactose và sự phân giải thành hai phân tử đường đơn bởi enzyme β -galactosidase; (b) Francois Jacob (trái) và Jacques Monod; và (c) Mô hình operon lactose ở *E. coli*.

- Một *yếu tố chỉ huy* (operator = O): trình tự DNA nằm kế trước nhóm gene cấu trúc, là vị trí tương tác với chất ức chế. Đối với operon-lac, đó là đoạn trình tự DNA dài 34 cặp base cách gene Z chừng 10 cặp base về phía trước. Nó chứa trình tự 24 cặp base đối xứng xuôi ngược, giúp chất ức chế có thể nhận biết và bám vào bằng cách khuếch tán dọc theo DNA từ cả hai phía (Hình 6.21c).

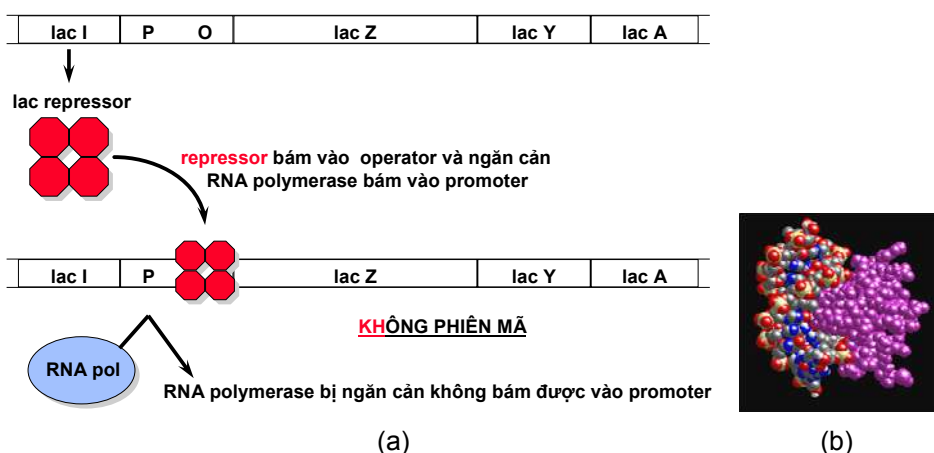
- Một *vùng khởi động* (promotor region = P): trình tự DNA nằm trước yếu tố chỉ huy và có thể trù lên một phần hoặc toàn bộ vùng này, là vị trí bám vào của RNA polymerase để có thể khởi đầu phiên mã tại vị trí chính xác của sợi khuôn. Đối với operon-lac, đó là đoạn DNA dài chừng 90 cặp base nằm trước và trù lên yếu tố chỉ huy 7 cặp base. Nó chứa hai vị trí

tương tác với RNA polymerase và với protein hoạt hoá dị hoá (catabolite activator protein = CAP, hoặc CRP). Điểm khởi đầu phiên mã là vị trí gần cuối của vùng khởi động P nằm trong đoạn chỉ huy O.

- Một *gene điều hoà* hay còn gọi là *gene ức chế* (regulatory/ inhibitory gene = R/I): Gene này sinh ra loại protein điều hoà gọi là *chất ức chế* (repressor) điều hoà hoạt động của nhóm gene cấu trúc thông qua sự tương tác với yếu tố chỉ huy. Đối với operon-lac, gene I nằm trước vùng khởi động mã hoá một protein ức chế gồm bốn polypeptide giống nhau (tứ phân) đều chứa 360 amino acid. Mặc dù mỗi gene điều hoà có một vùng khởi động và không có yếu tố chỉ huy riêng, đôi khi người ta vẫn coi chúng là một operon điều hoà (có tính chất cơ định). Ví dụ, để điều hoà operon-lac nói trên còn có operon-*lacI*.

Tóm lại, một operon là đơn vị điều hoà hoạt động gene của prokaryote, phức hợp liên kết giữa vùng khởi động cùng với yếu tố chỉ huy và nhóm gene cấu trúc do nó kiểm soát.

2. Operon lactose (*lac operon*) và cơ chế điều hoà cảm ứng - âm tính

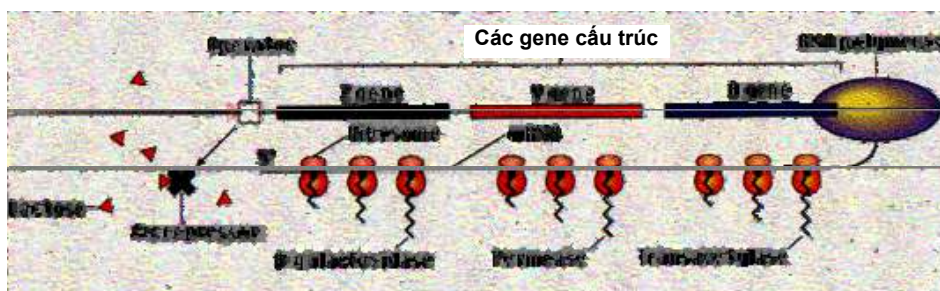


Hình 6.19 (a) Chất ức chế bám chặt *lac operator* gây ức chế phiên mã; và (b) Mô hình cấu trúc *lac operator* (rìa trái) bị bám chặt bởi protein ức chế (rìa phải).

Khi trong môi trường nuôi cấy *E. coli* không có lactose (chất cảm ứng) thì operon không hoạt động, nghĩa là các enzyme tham gia phân giải lactose không được sinh ra. Nguyên nhân là do chất ức chế bám chặt vào yếu tố chỉ huy gây ức chế sự phiên mã của các gene cấu trúc (Hình 6.19).

Ngược lại, nếu bổ sung lactose vào môi trường thì một thời gian sau vi khuẩn sẽ bắt đầu hấp thụ và phân giải nó, nghĩa là các enzyme liên quan đã được sinh ra. Sự kiện này được lý giải như sau: *Chất cảm ứng*

(inducer), ở đây là allolactose - dạng biến đổi của lactose - tương tác với *chất ức chế* (repressor) làm biến đổi hình dáng của chất này. Vì vậy chất ức chế mất ái lực và không thể bám vào yếu tố chỉ huy. Lúc này các gene cấu trúc được phiên mã và tổng hợp các enzyme tương ứng giúp vi khuẩn hấp thụ và phân giải đường lactose như một nguồn năng lượng và carbon (Hình 6.20). Lactose vì vậy là tác nhân gây cảm ứng (hoạt hoá) operon.



Hình 6.20 Chất cảm ứng bám vào chất ức chế làm biến đổi hình dáng của nó, do đó nó không bám vào được *lac* operator. Kết quả là các gene cấu trúc của operon lactose được phiên mã và các enzyme được tạo ra.

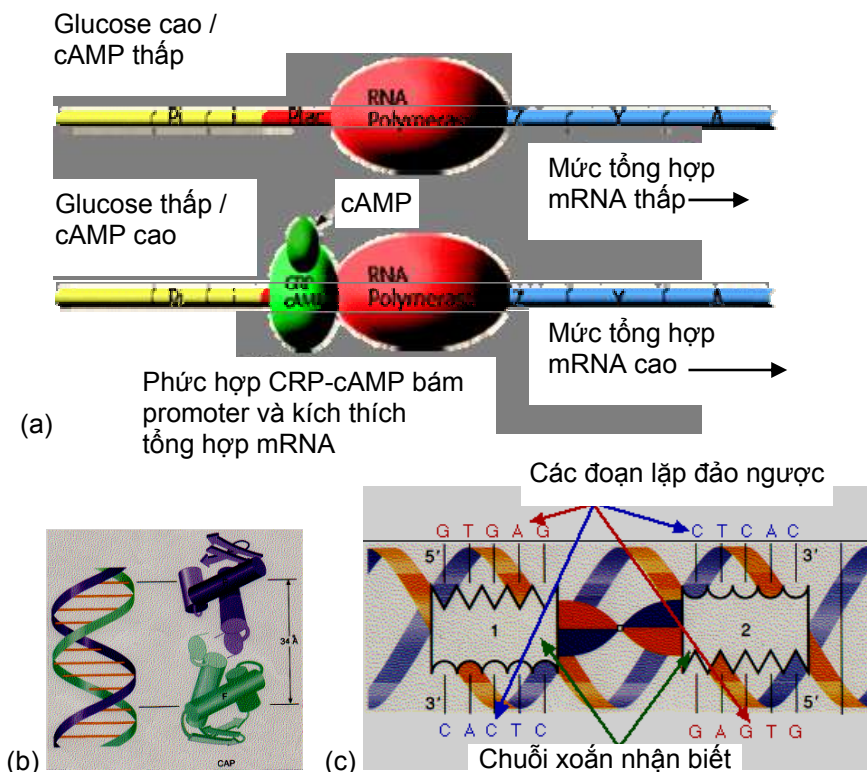
Phương thức điều hoà như thế được gọi là điều hoà cảm ứng - âm tính, bởi vì chất ức chế operon-lac một khi bám vào yếu tố chỉ huy thì gây ngừng phiên mã (gây hiệu quả âm tính lên sự biểu hiện của các gene), và hoạt động chức năng của protein này lại phụ thuộc vào phân tử hiệu ứng (ở đây là chất cảm ứng) có khả năng hoạt hoá operon.

** Điều hoà dương tính (positive control)*

Hoạt động của operon-lac còn chịu sự kiểm soát của một protein điều hoà dương tính liên quan với sự có mặt của glucose. Cụ thể, khi trong môi trường có mặt đồng thời cả lactose và glucose thì operon lac tạm thời ngưng hoạt động (ức chế dị hoá).

Người ta nhận thấy rằng, khi glucose có mặt ở nồng độ cao thì hàm lượng *AMP vòng* (cyclic AMP = cAMP) trong tế bào rất thấp; và ngược lại khi không có glucose hoặc có không đáng kể thì hàm lượng cAMP tăng cao. Vì vậy, cAMP được xem là chất chỉ thị của sự vắng mặt glucose. Ngoài ra còn phát hiện một loại protein điều hoà dương tính có tên là *protein hoạt hoá dị hoá* (catabolite activator protein = CAP, hoặc CRP). Protein CAP gồm hai tiểu đơn vị giống nhau gọi là *homodimer*; nó chỉ hoạt động khi môi trường nội bào có hàm lượng cAMP cao. Lúc này cAMP kết hợp với CAP tạo ra phức hợp CAP-cAMP hoạt động; phức hợp này có khả năng nhận biết và bám vào một đoạn 16 cặp base về phía trước của vùng khởi động, với các đoạn lặp đảo ngược, gọi là vị trí CAP (Hình

6.21b-c). Qua đó RNA polymerase được kích thích bám vào vị trí P và bắt đầu phiên mã ở mức cao (Hình 6.21).



Hình 6.21 (a) Phức hợp CRP- hoặc CAP-cAMP bám vào promoter và kích thích tổng hợp mRNA ở mức cao; (b) CAP gồm hai monomer giống nhau, mỗi monomer nhận biết một trình tự DNA nhờ vùng xoắn alpha được đánh dấu F; và (c) Trình tự đối xứng của vị trí CAP được xem là các đoạn lặp đảo ngược.

Như vậy, khác với kiểu điều hoà âm tính do sự tương tác giữa "chất ức chế và yếu tố chỉ huy", ở đây sự tương tác xảy ra giữa protein điều hoà thuộc phức hợp CAP-cAMP với vùng khởi động dẫn đến sự tăng cường hoạt động phiên mã (điều hoà dương tính) mà chủ yếu là điều chỉnh tốc độ khởi đầu phiên mã.

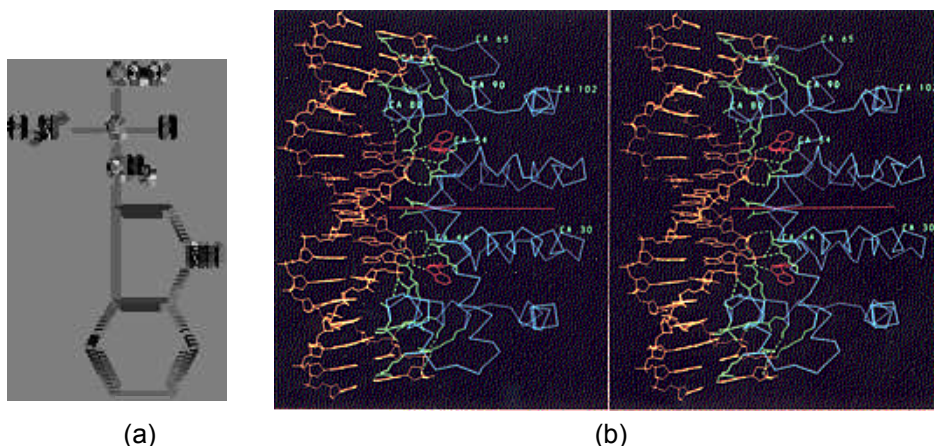
3. Operon tryptophan (*trp* operon) và cơ chế điều hoà ức chế - âm tính

Đại diện cho tất cả các operon của các loại amino acid và các vitamine là operon tryptophan (*trp* operon; *trp* đọc là "trip") ở *E. coli*.

3.1. Cấu trúc của *trp* operon

Operon tryptophan của *E. coli* có chứa năm gene cấu trúc (*trpE*, *trpD*, *trpC*, *trpB* và *trpA*) mã hoá cho các enzyme đồng hoá (anabolic) tham gia

vào quá trình sinh tổng hợp amino acid tryptophan (Hình 6.22a). Ngoài ra, nó có một số đặc điểm khác như: trình tự operator nằm lọt trong promoter, còn gene điều hoà nằm cách xa operon về phía trước. *Trp* operon cũng chịu sự điều hoà âm tính như *lac* operon; nó chỉ hoạt động khi môi trường nội bào thiếu hụt amino acid này và không hoạt động khi dư thừa tryptophan, sản phẩm cuối của con đường sinh tổng hợp. Vì vậy *trp* operon được gọi là *operon ức chế* (inducible) hay *operon đồng hoá*.



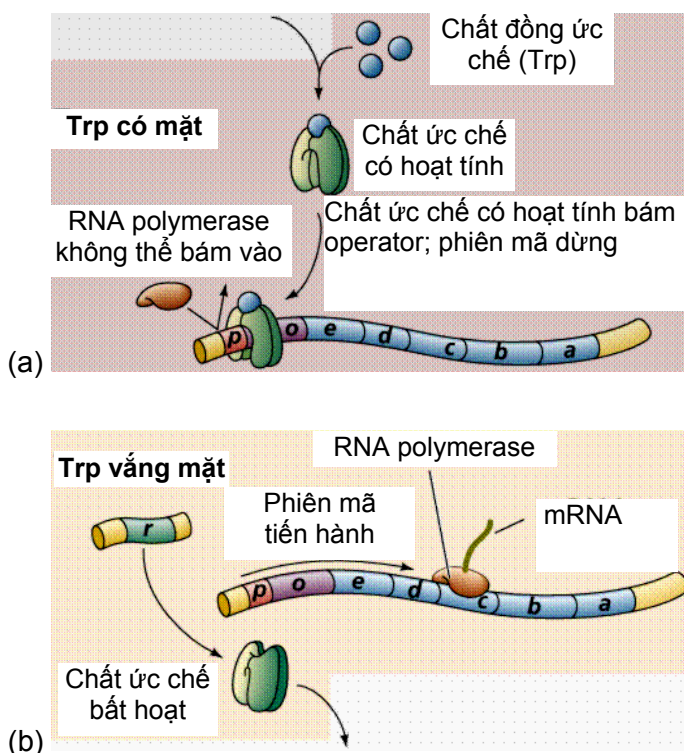
Hình 6.22 (a) Cấu trúc amino acid tryptophan. (b) Cấu trúc lập thể của chất ức chế *trp* operon (bên phải mỗi hình) bám vào DNA operator (bên trái mỗi hình). Chất ức chế này là một dimer gồm 2 polypeptide giống nhau (tương trưng bằng vạch ngang màu đỏ). Sự bám vào DNA chỉ xảy ra khi một phân tử tryptophan (các vòng đỏ) được bám dính vào mỗi monomer của chất ức chế này.

3.2. Cơ chế điều hoà âm tính của *trp* operon

Khi trong tế bào *E. coli* dư thừa amino acid tryptophan (sản phẩm cuối cùng của con đường chuyển hoá) thì *trp* operon ngừng hoạt động và do đó các enzyme tương ứng không được sinh ra. Sự kiện này được giải thích như sau: Chất ức chế bình thường của operon này tồn tại ở dạng bất hoạt (*trp* repressor hay *aporepressor*; Hình 6.22b), không có ái lực đối với yếu tố chỉ huy (*trp* operator). Nhưng khi các amino acid này dư thừa kết hợp vào chất ức chế sẽ tạo ra phức hợp có hoạt tính, nghĩa là có ái lực với yếu tố chỉ huy (tryptophan vì vậy được gọi là *chất đồng ức chế*, corepressor). Phức hợp này bám vào yếu tố chỉ huy làm kìm hãm phiên mã của *trp* operon (Hình 6.23a).

Ngược lại, khi trong tế bào vắng mặt hay thiếu hụt amino acid này, chất ức chế vốn ở trạng thái bất hoạt nên không bám vào yếu tố chỉ huy được. Vì vậy các gene cấu trúc xảy ra sự phiên mã và kết quả là các enzyme tham gia tổng hợp tryptophan được sinh ra (Hình 6.23b). Một khi

hàm lượng amino acid này được tổng hợp ở mức dư thừa sẽ tác động ngược trở lại, kìm hãm hoạt động của *trp* operon.



Hình 6.23 Operon tryptophan ở trạng thái bị kìm hãm (a) và hoạt động (b).

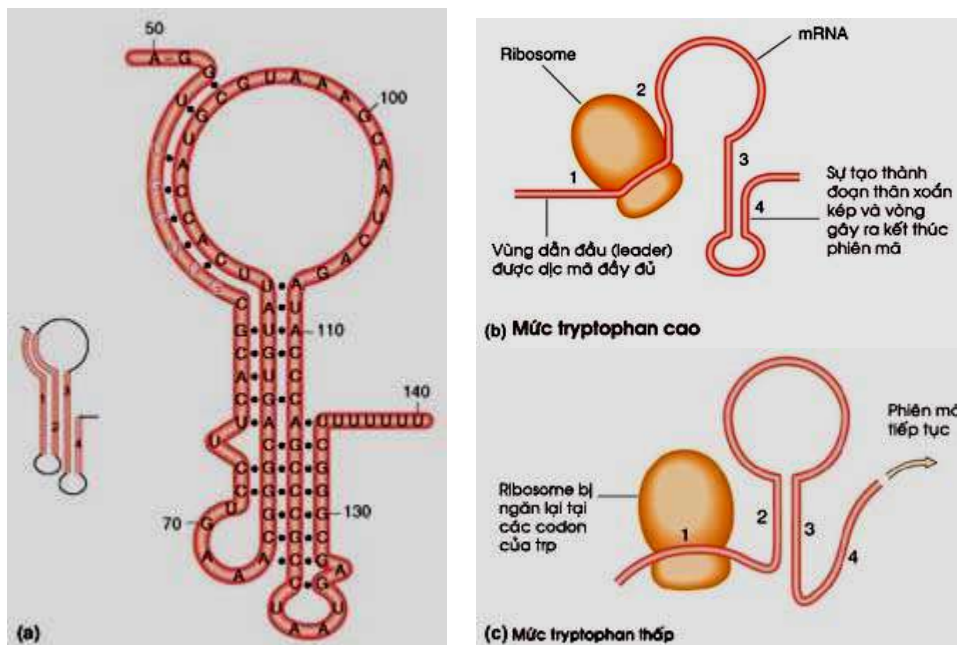
Tóm lại, phương thức điều hòa hoạt động gene theo các cơ chế liên hệ ngược hay phản hồi như thế (feed-back mechanisms) đảm bảo cho bộ gene các vi khuẩn hoạt động một cách hợp lý và nhờ đó các vi khuẩn thích ứng và phát triển trước các điều kiện môi trường luôn thay đổi.

3.3. Sự kết thúc phiên mã sớm ở *trp* operon

Phiên mã dờ (*attenuation*) là một cơ chế điều hoà gây ra sự *kết thúc phiên mã sớm* dưới những điều kiện nhất định, bằng cách đó ngăn cản sự biểu hiện của mRNA cần cho sự biểu hiện của các sản phẩm gene tương ứng. Phiên mã dờ tạo thành mRNA uốn gấp một cách điển hình thành các cấu trúc bậc hai xen kẽ (alternative secondary structures), mà một trong số đó là yếu tố kết thúc độc lập ρ (Rho-independent terminator).

Đối với operon tryptophan, đó là việc sử dụng dịch mã để điều khiển sự phiên mã. Khi có mặt tryptophan trong môi trường nội bào, thậm chí ở nồng độ thấp, sẽ xảy ra sự dịch mã một phần ở vùng dẫn đầu (leader) của

mRNA đang được tổng hợp. Kết quả là làm dừng sự phiên mã trước khi gene cấu trúc đầu tiên (*trpE*) của operon được phiên mã.



Hình 6.24 (a) Cấu trúc đoạn dẫn đầu - *TrpL* của *trp* operon. (b) Khi mức tryptophan cao, xảy ra sự kết thúc phiên mã sớm tại *trp attenuator* với một cái đuôi 3' gồm 8 uridine. (c) Khi mức tryptophan thấp, sự phiên mã tiếp diễn.

Sự kết thúc phiên mã sớm ở operon tryptophan là kết quả của sự tương tác bổ sung nội phân tử giữa các trình tự DNA bên trong vùng leader của bản sao RNA. Kết quả của sự kết thúc phiên mã sớm này tạo ra một mRNA chứa 140 base (hình 6.24a). Tại vùng đầu mút 3' của nó xảy ra sự tự bổ sung ở đoạn giàu GC tạo thành một cấu trúc hình vòng trên thân RNA và gây ra sự kết thúc phiên mã sớm (hình 6.24b). Vùng này được gọi là *đoạn phiên mã dở trp attenuator* và ở phần đuôi của mRNA này cũng có 8 base uridine. Kiểu cấu trúc "nút cài tóc" này là tín hiệu kiểm soát kết thúc phiên mã ở prokaryote nói chung.

Với kiểu cấu trúc đặc thù ở đoạn dẫn đầu của *trp* operon như vậy làm cho nó có ý nghĩa quan trọng trong điều hoà phiên mã dở, ở chỗ: (i) tổng hợp một peptide dẫn đầu chứa 14 amino acid; (ii) trên mRNA của đoạn peptide này chứa hai codon của Trp ở các vị trí 10 và 11; (iii) ở bốn vùng được đánh số 1-4 xảy ra sự tự bổ sung giữa các vùng 1 và 2 và giữa 3 và 4; và ở một số trường hợp có thể xảy ra sự kết cặp giữa các vùng 2 và 3.

Do trong trình tự mã hóa của trình tự dẫn đầu *trpL* có hai codon Trp,

nên sự dịch mã đoạn này tỏ ra nhạy cảm với số lượng tRNA^{trp} đưa vào. Nếu môi trường cung cấp đầy đủ Trp, ribosome trượt qua các codon Trp để đi vào vùng 2. Và sự có mặt của ribosome ở vùng 2 ngăn cản vùng này kết cặp với vùng 3. Khi đó vùng 3 sẽ cặp với vùng 4 và tạo ra điểm kết thúc phiên mã sớm (xảy ra sau khi tổng hợp xong 8 uridine ở ngay sau vùng 4). Khi số lượng tRNA^{trp} đưa vào không đầy đủ, sự dịch mã đoạn dẫn đầu dừng lại đột ngột ở các codon Trp của nó (hình 6.24c). Điều này ngăn cản ribosome tiến vào vùng 2, do đó vùng này sẽ cặp với vùng 3 gây cản trở việc tạo thành cấu trúc phiên mã dở (*trp* attenuator). Kết quả là phân tử mRNA đa cistron của operon tryptophan được tạo thành một cách đầy đủ.

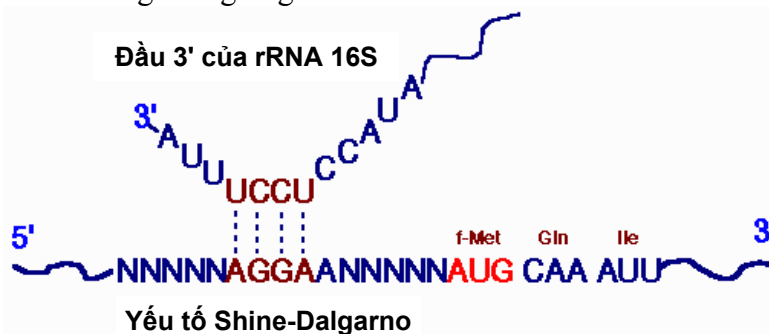
★ Operon ở eukaryote - một ngoại lệ thú vị!

Khác với tất cả các eukaryote, *Caenorhabditis elegans* và có lẽ cả một số giun tròn khác cũng có một tỷ lệ lớn các gene được tổ chức theo kiểu operon. Ở *C. elegans*, ít nhất 2.300 gene của nó (chiếm khoảng 15% bộ gene) có mặt trong các operon, mỗi operon chứa từ 2 đến 8 gene. Giống như các prokaryote, tất cả các gene trong một operon được phiên mã từ một promoter đơn sinh ra một bản sao sơ cấp đơn. Một số gene trong các operon này dường như có liên quan đến cùng chức năng sinh hoá như ở các prokaryote, nhưng không phải là trường hợp cho tất cả. Các operon của *C. elegans* cũng khác với các operon ở prokaryote ở chỗ, mỗi pre-mRNA được xử lý thành một mRNA riêng cho mỗi gene hơn là được dịch mã như một đơn vị (Kimball 2004).

★ Sơ lược về sự điều hoà ở mức dịch mã

Hiệu quả của sự khởi đầu dịch mã thường phụ thuộc vào trình tự giàu purine ở vùng 5'-UTR. Đó là 6-8 base (thường gặp là AGGAGGU) nằm ngay trước codon khởi đầu AUG của mRNA. Đoạn này bám vào tiểu đơn vị ribosome bé và được J.Shine và L.Dalgarno (Austria) xác định lần đầu tiên năm 1974. Vì vậy nó được gọi là *trình tự Shine-Dalgarno* (Hình 6.25). Các tác giả này cho rằng hầu như có sự bổ sung chính xác giữa đoạn trình tự này (ở đầu 5' của mRNA) và vùng tương ứng ở đầu 3' của rRNA 16S. Điều đó phù hợp với hiện tượng cố định bước đầu phân tử mRNA trên tiểu đơn vị 30S. Thông thường, ở các mRNA được dịch mã có hiệu quả nhất thì vùng bám vào ribosome thường nằm cách codon khởi đầu khoảng 8 nucleotide về phía trước. Nếu như đột biến xảy ra ở vùng này thì có thể làm giảm đột ngột hiệu quả dịch mã trên mRNA. Tuy nhiên, chỉ riêng sự có mặt của trình tự Shine-Dalgarno phân bố chuẩn vẫn chưa đủ đảm bảo cho sự khởi đầu dịch mã. Trên thực tế, có nhiều trình tự như

thể bị che khuất dưới dạng "nút cài tóc", vì vậy nó không thể tham gia tương tác với vùng tương ứng của rRNA 16S.



Hình 6.25 Sự tương tác giữa yếu tố Shine-Dalgarno của một mRNA và trình tự tương ứng ở đầu 3' của rRNA 16S có mặt trong tiểu đơn vị ribosome bé (theo M.W.King 1996).

Các số liệu thu được cho thấy hiệu quả của việc sử dụng trình tự Shine-Dalgarno nhất định có thể "chế tác" các protein mà đến lượt chúng lại bám vào trình tự đó và ngăn cản nó. Được nghiên cứu chi tiết nhất là các protein của ribosome ở *E. coli*. Khi tốc độ tổng hợp các protein này vượt quá mức sản sinh các rRNA thì sẽ xảy ra sự tích lũy các protein ribosome tự do. Số protein dư thừa này được gọi là các protein "khóa"; chúng bám vào trình tự Shine-Dalgarno trong các mRNA tương ứng. Nhờ vậy, tốc độ tổng hợp các protein ribosome được duy trì ở mức không vượt quá khả năng sử dụng chúng để cấu thành các ribosome. Có thể nói, sự điều hoà ở mức dịch mã là sự "cạnh tranh" giữa rRNA và mRNA của các protein ribosome, gây ra sự kết hợp với các protein "khóa" này. Khi sự dịch mã mRNA của các protein ribosome không thể tiếp diễn được nữa (do sự bám dính bởi các protein "khóa") thì các mRNA này sẽ bị thoái hoá nhanh hơn bình thường.

Câu hỏi và Bài tập

1. Phân tích các nguyên tắc và đặc điểm chung của quá trình sinh tổng hợp RNA. Từ đó trình bày tóm tắt cơ chế của quá trình này.
2. Trình bày vai trò của các yếu tố tham gia vào quá trình sinh tổng hợp protein của tế bào và phân tích cơ chế sinh tổng hợp protein dựa trên sự tương tác giữa mRNA, các aminoacyl-tRNA và ribosome.
3. So sánh tổ chức của các đơn vị phiên mã ở prokaryote và eukaryote.
4. Giả sử tổng hợp được một mRNA có thành phần 75%U và 25%G.

Khi sử dụng mRNA này để tổng hợp protein *in vitro* đã thu được các amino acid trong các protein với các tần số như sau :

Phe : Val : Leu : Cys : Gly : Trp = 1,00 : 0,44 : 0,33 : 0,33 : 0,15 : 0,11.

Hãy trình bày phương pháp và chỉ ra kết quả của việc giải đoán các codon cho mỗi amino acid nói trên (*không sử dụng bảng mã di truyền*). Biết rằng các codon cùng xác định một amino acid thường có hai nucleotide đầu giống nhau, và Cys được xác định bởi UGU.

5. Thế nào là các gene phân đoạn? Các intron có vai trò gì? Hãy phân tích cơ chế sửa đổi sau phiên mã đối với pre-mRNA của các gene mã hoá protein ở eukaryote.

6. Nếu sử dụng các phân tử mRNA nhân tạo có thành phần gồm các cụm gồm ba hoặc bốn nucleotide lặp lại dưới đây để tiến hành tổng hợp protein trong ống nghiệm, thì thành phần amino acid thu được từ các polypeptide sẽ như thế nào? Có trường hợp nào không tổng hợp được protein hay không? tại sao?

(a) (UUC)_n ; (b) (UAC)_n ; (c) (GAUA)_n ; (d) (GUAA)_n.

7. Phân biệt các cặp khái niệm sau: (a) chất cảm ứng với chất ức chế; (b) điều hoà âm tính với điều hoà dương tính; (c) điều hoà theo kiểu cảm ứng - âm tính với ức chế - âm tính. Hãy vẽ sơ đồ một operon và giải thích mối quan hệ giữa các yếu tố điều hoà hoạt động của một operon.

8. Operon là gì? Hãy nêu các đặc điểm giống nhau và khác nhau trong các cơ chế điều hoà âm tính đối với operon-lac và operon-trp. Từ đó rút ra ý nghĩa của các cơ chế điều hoà này.

9. Hãy giải thích các tình trạng đóng-mở của *lac* operon dưới các điều kiện sau đây và cho các hình vẽ minh hoạ: (a) chỉ có glucose; (b) chỉ có lactose; (c) không có chất đường nào cả; và (d) có cả glucose và lactose.

10. So sánh các quá trình sinh tổng hợp protein (biểu hiện của gene) ở các tế bào prokaryote và eukaryote và hãy cho biết các ý nghĩa sinh học từ sự giống nhau và khác nhau đó.

Tài liệu Tham khảo

Tiếng Việt

Hoàng Trọng Phán. 1995. *Một số vấn đề về di truyền học hiện đại* (Tài liệu BDTX giáo viên THPT chu kỳ 1993-1996). ĐHSPT Huế. Tr.28-72.

Hoàng Trọng Phán. 1997. *Di truyền học phân tử* (tái bản). Trung tâm Đào tạo Từ xa - Đại học Huế / NXB Giáo Dục. Tr.23-60.

Hoàng Trọng Phán. 2005. *Di truyền học*. Trung tâm Đào tạo Từ xa - Đại

học Huế / NXB Đà Nẵng.

Hoàng Văn Tiến (chủ biên), Lê Khắc Thận, Lê Doãn Diên. 1997. *Sinh hoá học với cơ sở khoa học của công nghệ gene*. NXB Nông Nghiệp, Hà Nội.

Nguyễn Tiến Thắng, Nguyễn Đình Huyền. 1998. *Giáo trình Sinh hoá hiện đại*. NXB Giáo Dục.

Tiếng Anh

Blackburn G.M., Gait M.J. (Eds., 1996): *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford.

Campbell PN, Smith AD, Peters TJ. 2005. *Biochemistry illustrated - Biochemistry and molecular biology in the post-genomic era*. 5th ed., Elsevier Limited, Edinburgh - London - New York - Oxford - Philadelphia - St Louis - Sydney - Toronto. (www.elsevierhealth.com)

Cooper DN. 2004. *Gene structure, function and expression*. http://www.cardiff.ac.uk/medicine/medical_genetics/study/medical_teaching/

Doolittle R.F. 1989. Redundancies in protein sequences. In *Predictions of Protein Structure and the Principles of Protein Conformation* (Fasman GD, Ed.) Plenum Press, New York, pp. 599-623.

Horton, Moran, Ochs, Rawn, Scrimgeour. 2002. *Principles of Biochemistry*. Prentice Hall, Inc.

<http://cwx.prenhall.com/horton/medialib/media_portfolio/index.html>

Kimball J. 2004. <http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/>
Lehninger, .L. et al. (1993): *Principles of Biochemistry*. Worth Publishers, New York.

Lewin B. 1999. *Genes VI*. Oxford University Press, Oxford.

Lewis R. 2003. *Human Genetics: Concepts and Applications*. 5th ed, McGraw-Hill, Inc, NY.

Mulligan ME. 2002.

<http://www.mun.ca/biochem/courses/3017/Topics/bases.html>

Nelson DL and Cox MM. 2000. *Lehninger Principles of Biochemistry*, 3rd ed., Worth Publishers, New York.

Russell PJ. 2003. *Essential Genetics*. Benjamin/Cummings Publishing Company, Inc, Menlo Park, CA.

Stryer, L. (1981): *Biochemistry*. W.-H. Freeman and Co., San Francisco.

Twyman RM. 1998. *Advanced Molecular Biology*. BIOS Scientific Publishers Ltd/ Springer-Verlag Singapore Pte Ltd.

Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM. 1987. *Molecular Biology of the Gene*. 4th ed, Benjamin/Cummings Publishing Company, Inc, Menlo Park, CA.

Weaver RF, Hedrick PW. 1997. *Genetics*. 3rd ed, McGraw-Hill Companies, Inc. Wm.C.Brown Publishers, Dubuque, IA.