

ĐẠI HỌC QUỐC GIA HÀ NỘI

VÕ THỊ THƯƠNG LAN

Sinh học phân tử

QGHN

3
L
7

85



NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

VÕ THỊ THƯƠNG LAN

SINH HỌC PHÂN TỬ

(In lần thứ tư)

NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA HÀ NỘI

LỜI NÓI ĐẦU

Sinh học phân tử nghiên cứu ở mức độ phân tử các phản ứng sinh học xảy ra trong tế bào. Hoạt động của từng gen cũng như sự phối hợp giữa các gen; kiểm soát sự phiên mã, dịch mã cũng như sự phân bố của các protein trong tế bào; các phản ứng sinh học đảm bảo hoạt động sống của tế bào cũng như điều hoà hoạt động giữa các tế bào trong một mô, giữa các mô với nhau... đã được khám phá nhờ các kỹ thuật sinh học phân tử. Sinh học phân tử là một môn học mới, đòi hỏi sự hiểu biết và có liên quan chặt chẽ đến các kiến thức cơ bản của di truyền, sinh hoá, vi sinh tế bào...

Giáo trình "Sinh học phân tử" dành cho cử nhân Sinh học gồm sáu chương. Chương thứ nhất đề cập đến một số thông tin về hệ gen (genome); về đặc điểm và tính đa dạng của genome. Chương hai giới thiệu về cấu trúc và hoạt động của gen prokaryot và eukaryot; các cơ chế kiểm soát hoạt động ở mức độ phiên mã và dịch mã. Chương ba đề cập đến một số kiến thức về sửa chữa và tổng hợp ADN. Chương bốn đưa ra các khái niệm chung về tái tổ hợp di truyền. Chương năm nhấn mạnh đến các kỹ thuật cơ bản của sinh học phân tử và chương sáu nêu lên quá trình tổng hợp và vận chuyển, phân bố protein trong các bào quan của tế bào.

Sinh học phân tử là một môn học mới, đòi hỏi kiến thức của nhiều môn học khác. Vì vậy nội dung của giáo trình này không

thể tránh khỏi những thiếu sót. Tác giả đã sửa chữa và bổ sung thêm một số kiến thức mới trong lần tái bản thứ nhất này. Tác giả xin bày tỏ lòng biết ơn sâu sắc đến GS. TS. Nguyễn Như Hiền, PGS. TS. Nguyễn Mộng Hùng và TS. Nông Văn Hải đã có những nhận xét, góp ý quý báu. Ngoài ra, tác giả rất mong nhận được sự phê bình góp ý của bạn đọc và đồng nghiệp.

Với sự cảm ơn chân thành

Tác giả

MỤC LỤC

Mục lục

Chương 1: CẤU TRÚC GENOME	11
1.1. Các thành phần của genome	14
1.2. Tính phức tạp của genome, giá trị C	18
1.3. Thay đổi trật tự các đoạn ADN trong genome.	
Transposons	22
1.4. Tương tác của T-ADN với genome thực vật	36
1.5. Sắp xếp và khuếch đại các gen trong genome	42
1.5.1. Chuyển đổi dạng giao phối của nấm <i>Saccharomyces cerevisia</i>	44
1.5.2. Chuyển đổi gen ở <i>Trypanosome</i>	50
1.5.3. Tăng số lượng các gen mã cho ARNr và một số gen khác ở tế bào eukaryot	54
Chương 2: CẤU TRÚC VÀ HOẠT ĐỘNG CỦA GEN	57
2.1. Cấu trúc gen	57
2.2. Kiểm soát sự bắt đầu tổng hợp phân tử ARNm prokaryot	61
2.2.1. Kiểm soát phiên mã trên lac operon theo cơ chế tiêu cực	61
2.2.2. Kiểm tra phiên mã trên ara operon theo cơ chế tích cực	65

2.2.3. Hoạt động của operon tryptophan	68
2.2.4. Ức chế quá trình dị hoá liên quan đến điều khiển tích cực	70
2.2.5. Điều khiển gen thông qua các promoter khác nhau	73
2.3. Kiểm soát dừng phản ứng tổng hợp ARNm	74
2.4. Xâm nhập của Bacteriophage λ vào tế bào <i>E.coli</i>	79
2.5. Tổng hợp ARNm eukaryot	82
2.6. Kiểm soát sau phiên mã	91
2.6.1. Sinh tổng hợp các protein ribosome được kiểm soát thông qua quá trình dịch mã trên ARNm	92
2.6.2. ARRN anti-sense	93
2.6.3. Phản ứng đọc sửa ARNm - "RNA editing"	95
2.6.4. Protein bám vào đầu 5' của phân tử ARNm kìm hãm quá trình dịch mã Phản ứng "dịch mã tiêu cực "	97
2.6.5. Độ dài của đuôi polyA ảnh hưởng tới độ bền vững của phân tử ARNm	98
2.6.6. Độ bền vững của ARNm	100
Chương 3: CÁC CƠ CHẾ SỬA CHỮA VÀ TỔNG HỢP ADN	102
3.1. Cắt bỏ đoạn ADN bị sai hỏng	106
3.1.1. Sửa chữa ADN loại bỏ base hỏng (BER)	107
3.1.2. Sửa chữa ADN loại bỏ nucleotide hỏng (NER)	108
3.2. Sửa chữa ADN ghép đôi lệch (mismatch repair)	111
3.3. Sửa chữa ADN theo cơ chế "Error-Prone"	113

3.4. Tổng hợp các enzym sửa chữa ADN	114
3.5. Tái bản ADN	115
3.5.1. <i>Đảm bảo tổng hợp ADN chính xác bởi cơ chế "đọc sửa" (Proofreading)</i>	119
3.5.2. <i>Oligo môi trong quá trình tổng hợp ADN</i>	122
3.5.3. <i>Các protein tham gia quá trình tổng hợp ADN</i>	124
3.5.3.1. <i>ADN helicase</i>	124
3.5.3.2. <i>SSB protein (Single Strand Binding Protein)</i>	125
3.5.3.3. <i>Protein có cấu trúc vòng giữ ADN polymerase chuyển động trên sợi ADN</i>	125
3.5.3.4. <i>Vai trò của ADN topoisomerase trong tái bản ADN</i>	126
Chương 4: TÁI TỔ HỢP DI TRUYỀN	127
4.1. Tái tổ hợp chung	128
4.2. Trao đổi gen và tổng hợp ADN có giới hạn	132
4.3. Tái tổ hợp ADN xảy ra ở vị trí đặc hiệu	133
Chương 5: ADN TÁI TỔ HỢP	139
5.1. Phân cắt và xác định trình tự nucleotide	141
5.1.1. <i>Phân cắt ADN. Enzym giới hạn (Restriction enzym)</i>	141
5.1.2. <i>Phân ly các đoạn ADN</i>	144
5.2. Đưa các đoạn ADN vào vector	144
5.3. Xây dựng ngân hàng các cDNA (cDNA library)	146

5.4. Xây dựng ngân hàng ADN genome (genomic DNA library)	148
5.5. Các phương pháp lai	152
5.5.1. <i>Phương pháp Southern blots</i>	152
5.5.2. <i>Phương pháp Northern blots</i>	154
5.5.3. <i>Kỹ thuật lai in-situ</i>	155
5.6. Xác định trình tự nucleotide	156
5.6.1. <i>Phương pháp hoá học Maxam-Gilbert</i>	156
5.6.2. <i>Phương pháp enzym Sanger</i>	157
5.6.3. <i>Xác định trình tự nucleotide trên máy tự động</i>	159
5.6.4. <i>Phương pháp "DNA footprint" cho phép xác định vị trí protein liên kết với phân tử ADN</i>	162
5.7. Hợp nhất tế bào - Kỹ thuật quan trọng trong di truyền tế bào soma	163
5.8. Phản ứng dây chuyền tổng hợp ADN (Polymerase Chain Reaction - PCR)	165
Chương 6: VẬN CHUYỂN PROTEIN TRONG TẾ BÀO	171
6.1. Các thành phần cơ bản tham gia tổng hợp protein	173
6.1.1. <i>Phân tử ARNt (RNA transfers)</i>	175
6.1.2. <i>Ribosome</i>	179
6.1.3. <i>Tính chính xác của quá trình tổng hợp protein</i>	182
6.2. Các con đường vận chuyển	184
6.3. Vận chuyển qua siêu lỗ trên màng nhân	187
6.4. Vận chuyển protein vào ty thể và lục lạp	189

6.5. Mạng lưới nội chất ER (<u>E</u> ndoplasmic <u>R</u> eticulum) và tổng hợp protein	191
6.6. Vận chuyển từ mạng lưới nội chất đến Golgi và Lysosome	196
6.7. Vận chuyển theo con đường thực ẩm bào (endocytosis)	198
6.8. Vận chuyển từ Golgi ra bề mặt tế bào (exocytosis)	200
6.9. Một số phương pháp nghiên cứu protein	201
6.9.1. Điện di trên gel polyacrylamide SDS-PAGE	201
6.9.2. Phân tích protein bằng phản ứng kháng nguyên - kháng thể	202
6.9.3. Nghiên cứu cấu trúc bậc I của protein (trình tự acid amin)	202
Tài liệu tham khảo	204

Chương 1

CẤU TRÚC GENOME

Genome (hệ gen) chứa toàn bộ thông tin di truyền và các chương trình cần thiết cho tế bào hoạt động. Đối với sinh vật nhân thực, 99% hệ gen nằm trong nhân tế bào và phần còn lại nằm trong một số bào quan như lục lạp, ty thể. Đa số genome vi khuẩn và phần chứa trong các bào quan thường có kích thước nhỏ và ở dạng vòng khép kín. Ngược lại, phần genome trong nhân thường rất lớn và phân bố trên các nhiễm sắc thể dạng thẳng. Cấu trúc của hệ gen trong nhân đã và đang thu hút trí tuệ của nhiều nhà sinh học do thông tin di truyền không chỉ nằm trong trình tự nucleotide (genetic information) mà phụ thuộc rất nhiều vào cấu hình không gian của nhiễm sắc thể (di truyền ngoại sinh - epigenetic information).

Genome không phải đơn thuần là tập hợp của các gen. Cấu trúc của hệ gen rất phức tạp và có độ trật tự cao. Ở sinh vật eukariot, thành phần ADN chứa các gen chỉ chiếm một tỷ lệ rất nhỏ so với toàn bộ genome. Cấu trúc của gen cũng như chức năng của chúng cũng khác nhau giữa sinh vật prokariot và eukariot, giữa sinh vật nhân thực bậc thấp và nhân thực bậc cao. Hệ gen của vi khuẩn và sinh vật eukariot bậc thấp thường không lớn và các gen phân bố sát nhau. Hầu hết các gen này chỉ có một bản sao trong genome và rất ít bị gián đoạn bởi các đoạn ADN không chứa mã di truyền (intron). Ngược lại, các gen

trong tế bào eukariot bậc cao thường chứa nhiều intron và phân bố xa nhau.

Genome của một số đối tượng cụ thể (model organisms) đại diện cho mỗi giới sinh vật như vi khuẩn *E.coli*, nấm men, ruồi giấm, giun tròn, *Arabidopsis* và người đã được xác định trình tự nucleotide. Bản đồ hệ gen có thể được xác định theo 3 cách chính:

- Bản đồ tế bào (cytogenetic map) liên quan đến sự xuất hiện của các băng sau khi làm tiêu bản tế bào (nhiễm sắc thể bắt màu). Kỹ thuật kính hiển vi điện tử ngày nay cho phép phát hiện các băng nhuộm màu đặc trưng cho một gen nhất định trên nhiễm sắc thể.

- Bản đồ di truyền (genetic map) đôi khi còn được gọi là bản đồ liên kết (linkage map) cho biết mối liên hệ gần gũi về vị trí của các nhóm gen với nhau hay của các marker với nhóm gen trên nhiễm sắc thể. Các marker này có thể là hình thái (biểu hiện tính trạng), sự đa dạng của protein (protein polymorphisms), đa dạng độ dài của các đoạn giới hạn (restriction fragment length polymorphisms-RFLPs), đa dạng độ dài các trình tự đơn giản (simple sequence length polymorphisms-SSLPs) và đa dạng các đoạn ADN được khuếch đại ngẫu nhiên (randomly amplified polymorphic DNA-RAPD). Hai nhóm gen hay gen với marker càng gần nhau thì càng khó phân ly độc lập trong phân bào giảm phân. Do đó, mối quan hệ này được biểu diễn bởi tần số trao đổi chéo có thể xảy ra giữa chúng trong quá trình phân bào. Tuy nhiên, bản đồ này không phản ánh chính xác khoảng cách thực giữa các gen do trao đổi chéo không xảy ra như nhau ở mọi vị trí trên nhiễm sắc thể.

- Bản đồ vật lý (physical map) được tính theo đơn vị một nucleotide (bp). Vì vậy, việc xây dựng bản đồ vật lý hay xác định trình tự nucleotide của genome hay ít nhất là của toàn bộ các gen có mặt trong hệ gen là mối quan tâm đặc biệt của sinh học hiện đại.

Trình tự genome của những đối tượng sinh vật mô hình có rất nhiều giá trị trong hướng nghiên cứu mới về genome (genomic). Dựa vào đó, các nhà sinh học có thể phân tích cấu trúc, hoạt động và chức năng của các gen, làm sáng tỏ được vai trò của ADN lặp lại, ADN không chứa mã di truyền, ADN nằm giữa các gen v.v... Điều đặc biệt có ý nghĩa là khi so sánh các hệ gen với nhau, có thể hiểu được hoạt động của hệ gen trong các đối tượng sinh vật, mối quan hệ giữa chúng, sự đa dạng sinh học và mức độ tiến hoá.

Mục đích của việc xây dựng bản đồ gen là xác định vị trí của một gen bất kỳ tại một vùng trên nhiễm sắc thể. Tính chính xác của các vị trí cũng như mối liên hệ với các chỉ thị xung quanh là yêu cầu cần đạt được trong việc lập bản đồ gen. Từ bản đồ gen chúng ta có thể biết được trật tự sắp xếp của các gen trên nhiễm sắc thể và thành phần các loại ADN trong toàn bộ genome. Từ đó, có thể xác lập được sự di truyền của các tính trạng và mối liên quan của chúng với cấu trúc nhiễm sắc thể. Sản phẩm của một gen do trình tự nucleotide quyết định. Tuy nhiên sản phẩm này được tạo ra nhiều hay ít còn phụ thuộc vào vị trí của gen trong genome. Sự thay đổi vị trí tác động đến mức độ biểu hiện của tính trạng do gen qui định và đôi khi còn gây ảnh hưởng đến các gen khác. Với những tiến bộ kỹ thuật trong công nghệ gen, số bản đồ chi tiết chính xác của các đối tượng sinh vật khác nhau ngày càng tăng.

Kết quả bước đầu so sánh hệ gen giữa các loài sinh vật với nhau cho phép rút ra ba đặc điểm nổi bật. Thứ nhất là các gen phân bố không theo qui luật trong hệ gen. Thứ hai là kích thước hệ gen thay đổi không tỷ lệ với tính phức tạp của loài và cuối cùng là số lượng nhiễm sắc thể cũng rất khác nhau ngay giữa những loài rất gần nhau. Nếu phân tích chi tiết đối với một gen nhất định thì vị trí các intron, các exon, các đoạn ADN điều khiển hoạt động của gen v.v... đều là những yếu tố quan trọng để so sánh tìm ra mối quan hệ giữa các loài. Ngoài ra, tổng số gen nói chung, số lượng các gen có nhiều bản sao trong genome, tỷ lệ các loại ADN lặp lại và thành phần của chúng cũng như sự di chuyển của các gen từ hệ gen riêng biệt của các bào quan (ty thể, lục lạp) sang genome trong nhân đều chịu ảnh hưởng của thời gian, tức là đều phản ánh quá trình tiến hoá của các loài. Mặt khác, để có được sự so sánh chính xác hơn, toàn diện hơn, cần xét đến cấu trúc sợi chromatin, cấu hình không gian ba chiều của nhiễm sắc thể cũng như của toàn bộ hệ gen trong nhân.

Trong chương này, chúng ta cùng nhau xem xét thành phần và cấu trúc của genome phân bố trong nhân tế bào mà không xét đến phần hệ gen trong các bào quan.

1.1. CÁC THÀNH PHẦN CỦA GENOME

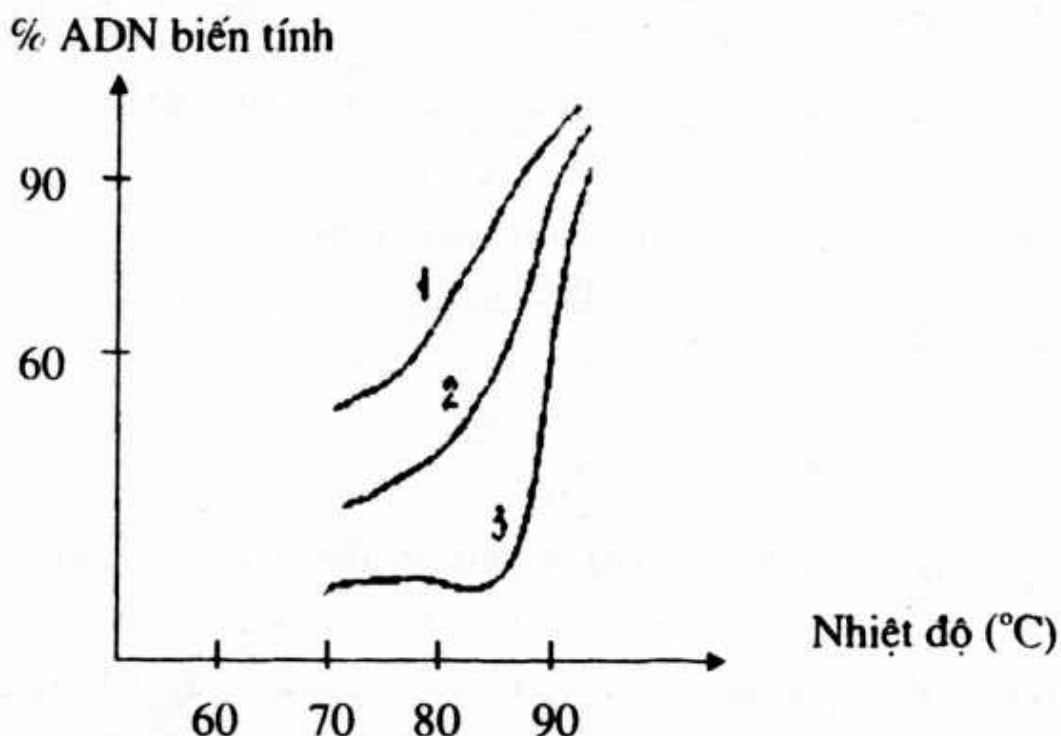
Genome chứa mọi thông tin di truyền đặc trưng cho từng loài, thậm chí cho từng cá thể trong loài. Genome có thể bao gồm các phân tử ADN hoặc ARN. Đối với sinh vật bậc cao, kích thước genome thay đổi từ 10^9 (động vật có vú) đến 10^{11} (thực vật) cặp base (bp). Khác với tế bào prokaryot (tế bào nhân sơ), các gen trong genome eukaryot (tế bào nhân chuẩn) thường tồn tại

nhiều bản sao và thường bị gián đoạn bởi các đoạn không mang thông tin di truyền (các intron). Vì vậy, một trong những vấn đề được các nhà sinh học quan tâm là đánh giá được số lượng các gen khác nhau có mặt trong genome cũng như số lượng gen hoạt động trong từng loại mô, từng giai đoạn phát triển và tỷ lệ các gen so với kích thước genome...

Để giải đáp được những vấn đề nêu trên, một phương pháp hay được sử dụng là nghiên cứu động học phản ứng lai acid nucleic. Chúng ta đều biết rằng phân tử ADN ở dạng sợi kép có thể mở xoắn thành hai chuỗi đơn dưới tác dụng nhiệt. Lúc đó chúng ta nói phân tử ADN bị biến tính (denaturation). Tuy nhiên khi duy trì ở nhiệt độ thích hợp thì hai chuỗi đơn có thể liên kết tạo cặp trở thành chuỗi xoắn kép. Điều kiện để hai sợi đơn liên kết tạo nên sợi kép là số lượng các nucleotide trên hai sợi đó có khả năng tạo cặp bổ sung phải đạt đến một tỷ lệ nhất định. Tỷ lệ này càng lớn thì độ bền vững của sợi kép càng cao. Lúc đó chúng ta có phản ứng phục hồi phân tử ADN (renaturation) hay còn gọi là phản ứng lai ADN. Như vậy, phản ứng lai không phải chỉ xảy ra với hai sợi đơn của một sợi kép mà có thể xảy ra với các phân tử ADN khác nhau. Chúng ta gọi chung các phân tử ADN này là các phân tử tương đồng (về trình tự nucleotide).

Kết quả nghiên cứu động học phản ứng lai ADN cho thấy genome tế bào vi khuẩn chỉ có một thành phần gồm các đoạn ADN không giống nhau (hay còn được gọi là các đoạn ADN đơn bản). Nói cách khác, các gen ở vi khuẩn thường chỉ duy nhất mà không có bản sao. Tuy nhiên genome eukaryot có cấu trúc phức tạp hơn. Chúng gồm ba thành phần ADN khác nhau. Thành phần thứ nhất gồm các đoạn ADN lặp lại nhiều (chiếm khoảng

25%); thành phần thứ hai gồm các đoạn ADN lặp lại ít (chiếm khoảng 30 %) và thành phần thứ ba gồm các đoạn ADN đơn bản (chiếm khoảng 45%). Khi từng thành phần ADN riêng biệt trải qua biến tính và phục hồi lần thứ nhất, sau đó mỗi thành phần tiếp tục trải qua quá trình biến tính nhiệt lần thứ hai, chúng ta thu được kết quả trình bày trên Hình 1.1.



Hình 1.1: Đường cong biến tính
của các thành phần ADN lặp lại và không lặp lại.

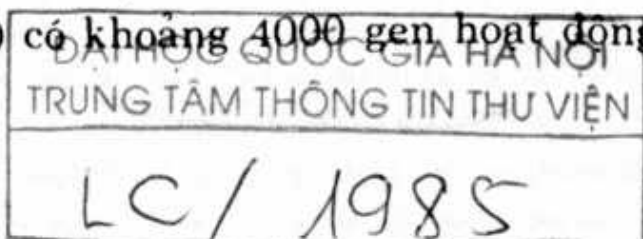
(1): ADN lặp lại. (2): ADN không lặp lại. (3): ADN tổng số ban đầu

So sánh các đường cong biến tính lần thứ hai của thành phần ADN lặp lại (1) và ADN không lặp lại (2) với đường cong biến tính của thành phần ADN tổng số ban đầu (3), chúng ta thấy trong lần biến tính thứ hai thành phần ADN lặp lại biến tính nhanh trong khi thành phần ADN không lặp lại biến tính tương tự như ADN tổng số ban đầu. Điều này chứng tỏ rằng thành phần ADN lặp lại chứa các đoạn có trình tự nucleotide

không hoàn toàn giống nhau mà chỉ tương đồng với nhau. Chúng ta có thể giải thích kết quả đó như sau:

Thành phần ADN lặp lại gồm các đoạn khá tương đồng nhau. Do đó, sau khi biến tính lần thứ nhất, các sợi đơn của những đoạn tương đồng có thể lai với nhau trong phản ứng phục hồi ngay khi không phải mọi nucleotide trên hai sợi đơn đều có khả năng tạo cặp bổ trợ với nhau. Như vậy phân tử ADN lai (sau khi phục hồi) có chứa những khoảng trống (không có liên kết hydrogen giữa hai sợi đơn). Khi gây biến tính lần hai, nhiệt lượng để bẻ gãy các liên kết sẽ ít hơn so với lần đầu do có sẵn những khoảng trống đó. Vì vậy đường cong biến tính lần hai đối với thành phần ADN lặp lại có giá trị T_m nhỏ hơn so với giá trị T_m so với lần biến tính ban đầu hoặc so với khi biến tính ADN tổng số. Ngược lại thành phần ADN không lặp lại chỉ gồm các đoạn ADN không giống nhau. Do đó, sau khi biến tính lần thứ nhất hai sợi đơn của một sợi kép sẽ tạo cặp với nhau và lập lại trạng thái ban đầu. Lúc này không xảy ra hiện tượng lai giữa hai sợi khác nhau vì thành phần này không chứa các sợi tương đồng nhau. Do không có các khoảng trống giữa hai sợi đơn nên nhiệt độ T_m không thay đổi ở lần biến tính thứ hai so với lần đầu.

Không phải mọi đoạn ADN trong genome đều tương ứng với các gen (mã cho protein hoặc một sản phẩm cần thiết cho hoạt động sống của tế bào). Từ những năm 70, bằng các thí nghiệm gây bão hoà đột biến, các nhà di truyền học có thể xác định được số gen nằm trên một đoạn nhiễm sắc thể. Ngày nay các kỹ thuật phân tích ADN hiện đại (các phép lai Southern, Northern, micro array...), cho phép xác định số gen hoạt động trong một tế bào. Ví dụ như ở tế bào nấm men (sinh vật eukaryot bậc thấp) có khoảng 4000 gen hoạt động, còn tế bào



động vật có vú khoảng 10.000 - 15.000 gen. Như vậy, nếu độ dài trung bình của một gen khoảng 10 000 bp thì tổng số chiều dài các gen hoạt động trong một tế bào cũng chỉ chiếm 1-2% genome. Hay nói cách khác chỉ một phần rất nhỏ genome mang thông tin di truyền cần thiết cho hoạt động sống của tế bào. Vậy phần genome còn lại có vai trò gì ? Phải chăng kích thước genome càng lớn thì tính phức tạp của loài (cá thể) càng cao ?

Để giải đáp câu hỏi trên, chúng ta chỉ cần xem xét kích thước genome của một số loài gần nhau trong bậc thang tiến hoá (tức là có độ phức tạp loài tương tự như nhau) cũng như genome của những loài cách xa nhau (tức là có tính phức tạp khác nhau). Ví dụ như genome của người có kích thước khoảng $3,3 \times 10^9$ bp, trong khi đó genome các loài lưỡng cư dài tương tự cỡ $3,1 \times 10^9$ bp hoặc của thực vật có thể đạt đến 10^{11} bp. Có lẽ nào loài lưỡng cư lại có tính phức tạp như cơ thể chúng ta ? Mặt khác, ngay trong cùng một loài chúng ta cũng nhận thấy sự mâu thuẫn về kích thước genome. Ví dụ như ruồi sống trong nhà (*Musca domestica*) có genome cỡ $8,6 \times 10^8$ bp, lớn gấp 6 lần kích thước genome ruồi giấm (*Drosophila.melanogaster*) với genome cỡ $1,4 \times 10^8$ bp. Ngoài ra, kích thước genome của các quần thể lưỡng cư thay đổi từ 10^9 bp đến 10^{11} bp (khác nhau gấp 100 lần). Vì sao ngay trong cùng một loài kích thước genome lại biến thiên nhiều như vậy ? Phải chăng ruồi nhà có cấu tạo phức tạp hơn nhiều so với ruồi giấm?

1.2. TÍNH PHỨC TẠP CỦA GENOME, GIÁ TRỊ C.

Phản ứng lai acid nucleic được tiến hành giữa ADN genome với cADN, ADN với ARNm v.v... Kết quả nghiên cứu động học của các phản ứng này cho thấy hầu hết các gen hoạt động đều

nằm trong thành phần ADN không lặp lại. Như vậy thành phần này có ý nghĩa rất quan trọng trong việc đánh giá tính phức tạp của genome. Hay nói cách khác, dựa vào thành phần ADN không lặp lại có thể biết được kích thước genome cũng như mức độ tiến hoá của loài. Nếu như kích thước genome (ở trạng thái đơn bội) được coi là một tham số động học ký hiệu là C, thì giá trị này đặc trưng cho từng loài và không phải luôn tỷ lệ thuận với tính phức tạp của loài. Ngược lại giá trị C phản ánh các nghịch lý sau:

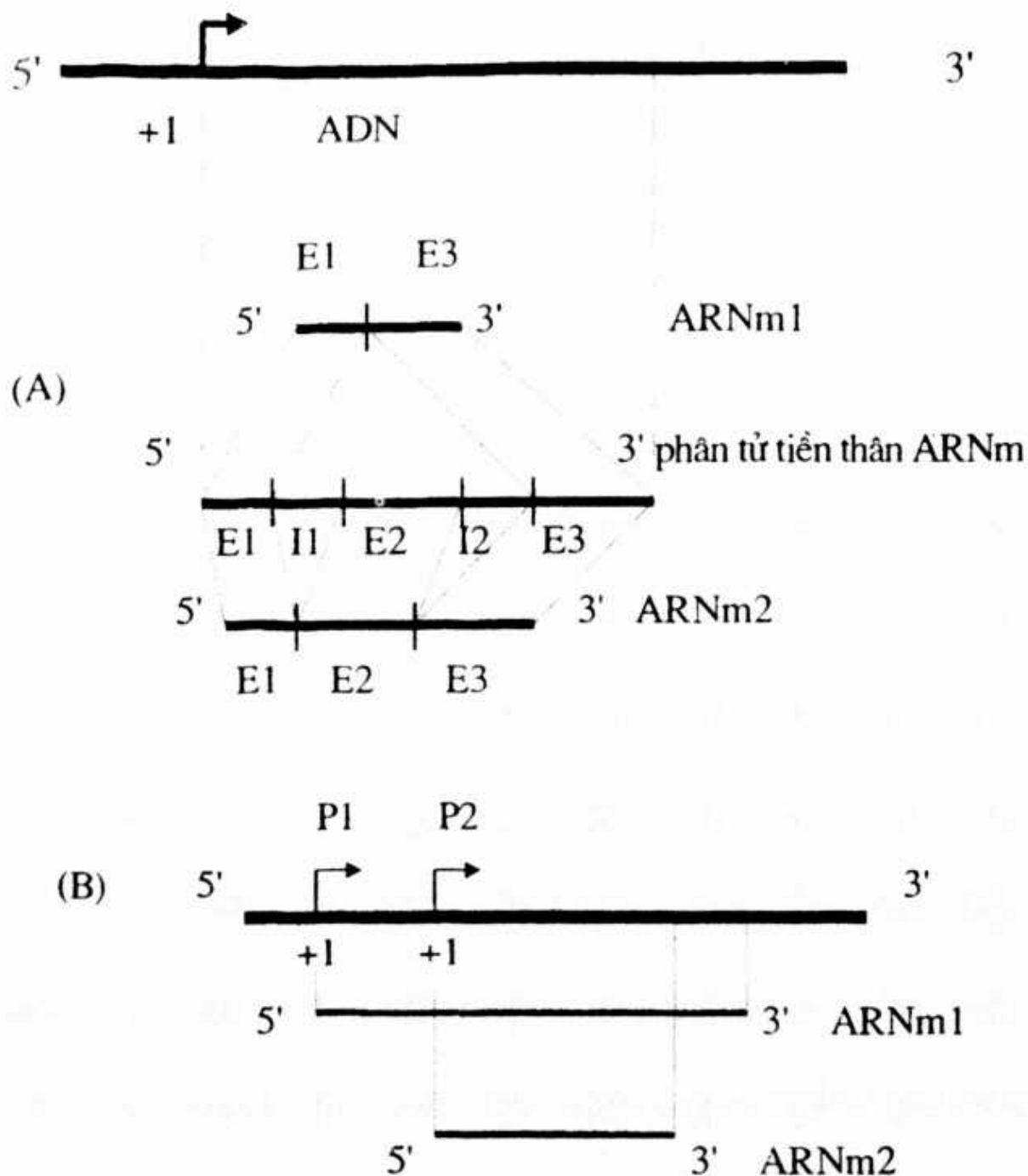
- + Số lượng ADN mã cho các sản phẩm cần thiết đối với hoạt động sống của cơ thể rất nhỏ so với số lượng ADN trong genome.

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

Genome vi khuẩn được xem là chỉ chứa các đoạn ADN không lặp lại và các gen thường tồn tại đơn bản (không có nhiều bản sao của một gen). Ngược lại, genome eukaryot thường chứa các gen có hai hay nhiều bản sao. Hơn nữa trình tự nucleotide của các bản sao này có thể không giống nhau hoàn toàn mặc dù sản phẩm protein mà chúng mã có cùng một chức năng. Các bản sao tương đồng của một gen được xếp chung vào một nhóm gọi là một họ gen. Như vậy ngoài các gen đơn bản giống như ở vi khuẩn, genome eukaryot còn chứa các họ gen. Hầu hết các gen mã cho protein đã được phân lập đều nằm trong các họ gen khác nhau. Các gen trong một họ thường hoạt động theo thời gian và không gian. Điều đó có nghĩa, mỗi thành viên trong họ thường hoạt động ở một thời điểm nhất định trong quá trình hình thành và phát triển cá thể hoặc hoạt động trong các mô chuyên

biệt. Khi một thành viên trong họ bị đột biến (bất hoạt) thì thành viên khác có thể hoạt động thay thế.

Khái niệm về gen được hình thành khi các nhà di truyền học cổ điển nghiên cứu biểu hiện các tính trạng mới do gây đột biến ADN genome vi khuẩn hay thực khuẩn thể *bacteriophage*. Một gen được xem là một đoạn ADN mà bất cứ đột biến nào xảy ra trên đó đều dẫn đến xuất hiện tính trạng mới. Điều này dễ hiểu đối với genome chỉ chứa các gen đơn bản (như genome tế bào prokaryot). Tuy nhiên, ở genome eukaryot, khi có nhiều gen cùng qui định một tính trạng hoặc các gen tương đồng trong một họ có thể hoạt động hỗ trợ thay thế nhau thì đột biến trên một gen không phải lúc nào cũng quan sát được ở mức độ phenotype. Mặt khác, các đoạn ADN tương ứng với mã di truyền (các exon) cho một protein thường bị ngắt quãng bởi các đoạn ADN không chứa thông tin (các intron). Các intron được phiên mã cùng với các exon sang phân tử ARN (gọi là phân tử tiền thân ARNm) nhưng sau đó chúng bị loại bỏ và các exon được nối với nhau tạo thành phân tử ARNm hoàn chỉnh (được dùng vào quá trình tổng hợp protein). Quá trình cắt các intron, nối exon không tuân theo một trật tự bắt buộc mà đa dạng biến hoá tạo ra các phân tử ARNm khác nhau từ một phân tử ARNm tiền thân (Hình 1.2A). Bên cạnh ARNm, các phân tử ARNr, ARNt cũng được hình thành từ các phân tử tiền thân chứa intron. Ngoài ra còn có hiện tượng mã di truyền của gen này nằm xen kẽ với các mã của gen khác (các gen nằm gối lên nhau) hoặc trường hợp dịch chuyển khung đọc ngay trên một đoạn ADN (Hình 1.2B).



Hình 1.2: (A): Cắt intron và nối exon tạo ra hai phân tử ARNm khác nhau nhưng có chung hai exon giống nhau E1 và E3. Chúng mã cho hai chuỗi polypeptit có chức năng khác nhau trong tế bào. (B): Hai gen mã cho hai protein cùng nằm trên một đoạn ADN do mã di truyền phân bố theo các khung đọc khác nhau.

Chúng ta đã biết đến chức năng của các phân tử ARN trong hoạt động sống của tế bào. Bên cạnh 3 loại ARN đã được nghiên cứu khá kỹ (ARN thông tin-ARNm; ARN ribosome - ARNr và,

ARN vận chuyển - ARNt), vai trò của một số loại ARN khác mới được phát hiện vào những năm cuối thế kỷ 20. Chúng kiểm soát hoạt động của gen (hiện tượng bất hoạt gen - gene silencing); tham gia phản ứng đọc sửa thông tin di truyền trên phân tử ARNm (hiện tượng RNA editing) hay quyết định tính bền vững của ARNm (các ribonuclease) v.v... Thậm chí có những phân tử tiền thân ARNm được tổng hợp không phải mã cho protein mà với mục đích phân cắt tạo ra những phân tử ARN có kích thước nhỏ hơn tham gia quá trình kiểm soát hoạt động của các gen khác. Đột biến ở những đoạn ADN mã cho tất cả các loại ARN này thường gắn liền với việc xuất hiện tính trạng mới. Do đó cần xem xét các đoạn nucleotide đó như các gen mặc dù chúng không mã cho protein.

Ngày nay theo quan điểm của sinh học phân tử, ***một gen được xem là một đoạn ADN mã cho một sản phẩm cần thiết đối với hoạt động sống của tế bào.*** Rõ ràng rằng không phải chỉ có ADN mã cho protein mà cả các ADN mã cho ARN ribosome, ARN vận chuyển và các loại ARN khác tham gia vào những con đường kiểm soát hoạt động của genome cũng được xác định là gen.

1.3. THAY ĐỔI TRẬT TỰ CÁC ĐOẠN ADN TRONG GENOME TRANSPOSONS

Chúng ta đều biết rằng kích thước và cấu trúc genome của các loài khác nhau thì khác nhau. Tuy nhiên điều gì qui định tính đa dạng giữa các cá thể ngay trong loài? Sự khác nhau giữa genome của các cá thể trong một loài thường phân bố trong thành phần nào của hệ gen? Những nguyên nhân nào dẫn đến sự khác nhau đó?

Trao đổi chéo giữa các nhiễm sắc thể tương đồng xảy ra trong phân bào meiose là một trong những nguyên nhân dẫn đến sự đa dạng trong loài. Ngoài ra trong mỗi genome, trật tự các đoạn ADN cũng như cấu trúc các gen còn được sắp xếp lại đặc trưng cho từng cá thể. Chính các yếu tố di truyền có khả năng vận động (Mobile Genetic Elements) giữa các vị trí trong một genome hoặc giữa các genome khác nhau góp phần làm đa dạng di truyền giữa các cá thể trong loài.

Các yếu tố di truyền có khả năng vận động được xếp vào 3 nhóm chính tùy thuộc vào tính độc lập của chúng. Nhóm thứ nhất gồm các yếu tố có khả năng di chuyển giữa các vị trí khác nhau trong genome, nhóm thứ hai gồm các yếu tố có khả năng ghép vào và tách ra khỏi genome để tồn tại độc lập trong tế bào (các episome như plasmid F, thực khuẩn thể λ) và nhóm thứ ba chỉ di chuyển dưới sự kiểm soát của tế bào ở những giai đoạn sinh trưởng phát triển nhất định để sắp xếp khởi động một số gen đặc biệt (sự hình thành cassette hoạt động ở nấm men *S. cerevisiae*, ký sinh trùng đơn bào Trypanosome). Di chuyển của các yếu tố ở nhóm thứ hai và thứ ba liên quan đến tái tổ hợp tương đồng hoặc tái tổ hợp ở các vị trí đặc hiệu (homologous or site-specific recombination). Mặc dù không có khả năng tồn tại độc lập bên ngoài genome, nhóm thứ nhất tự kiểm soát sự di chuyển của chúng trong hệ gen và không đòi hỏi sự tương đồng giữa chúng với vị trí ghép vào. Do đó có thể nói chúng di chuyển một cách tự do trong hệ gen. Chúng được gọi chung là các yếu tố di chuyển (transposable elements). Trong khi thực khuẩn thể λ

được xem là episome thì thực khuẩn thể *Mu* và một số retrovirus ở eukaryot lại là transposable element.

Các yếu tố di chuyển có thể phân thành hai loại dựa vào cách thức di chuyển của chúng. Loại thứ nhất được gọi là retroelement phải trải qua hình thức trung gian ARN trong quá trình di chuyển (hệ gen ARN được sao chép nhờ reverse transcriptase tạo ADN bổ sung (ADNc) để ghép vào vị trí mới). Loại thứ hai là các đoạn ADN hoặc bị tách ra hoặc được tái bản tạo thêm bản sao để ghép vào vị trí mới. Thông thường loại thứ hai này được gọi là transposon.

Khi di chuyển, các transposon gây ra việc sắp xếp, tổ chức lại genome của từng cá thể như tạo các đoạn ADN mới hoặc thay đổi chức năng hoạt động của các đoạn ADN ở vị trí chúng ghép vào và tách ra. Chúng có thể di chuyển tới vị trí bất kỳ và hoàn toàn không yêu cầu mối quan hệ nào giữa hai vị trí mới và cũ. Khi tách ra transposon có thể mang theo các đoạn ADN phụ cận, gây sự mất đoạn tại vị trí cũ. Ngược lại, khi ghép vào vị trí mới, chúng gây ra hiện tượng thêm đoạn hoặc chuyển đoạn ở vị trí mới. Do đó, transposon giống như các vector chuyên chở ADN từ nơi này sang nơi khác trong một genome hoặc từ genome này sang genome khác. Ngoài ra, trao đổi chéo giữa các transposon tương đồng ở hai vị trí khác nhau trên một hoặc trên hai nhiễm sắc thể cũng tạo ra những biến đổi tương tự. Những biến đổi đó dẫn đến sắp xếp lại genome, tạo tính đa dạng giữa chúng và tính đặc thù riêng của từng cá thể. Đặc biệt, sự thay đổi vị trí của các transposon còn có thể gây ảnh hưởng đến hoạt động của các gen phân bố xung quanh ngay khi

chúng không làm thay đổi trật tự nucleotide ở những gen này. Tần số ghép của các transposon vào genome khoảng 10^{-5} - 10^{-7} sau mỗi thế hệ. Ngược lại, tần số tách ra khoảng 10^{-6} đến 10^{-10} .

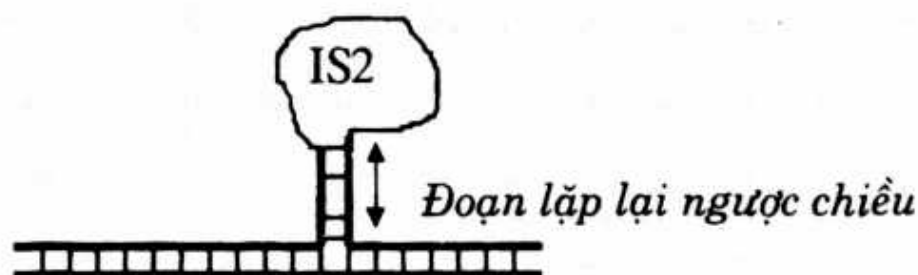
Các transposon có thể chia làm hai nhóm dựa vào khả năng di chuyển độc lập hay phải phụ thuộc vào sự có mặt của transposon khác.

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

* Các transposon không có khả năng tự hoạt động, tức là chúng không có khả năng di chuyển do không mang đủ thông tin di truyền mã cho các enzym cần thiết. Vì vậy các transposon loại này tạo ra những đột biến gen một cách tự phát nhưng là đột biến bền vững. Việc di chuyển của transposon ở nhóm này phụ thuộc vào sự có mặt của transposon có khả năng hoạt động độc lập cùng nhóm. Hai transposon có thể xếp vào cùng nhóm khi chúng có cấu trúc tương đồng với nhau, đặc biệt là các đoạn oligonucleotide phân bố ở hai đầu transposon. Đây là vị trí để enzym nhận biết và cắt nối transposon ở vị trí đi và đến.

Các transposon đơn giản nhất ở vi khuẩn được gọi là IS (Insertion Sequences). Chúng có thể nằm trên chromosome hoặc trên các plasmid. Để diễn tả việc ghép của IS vào vị trí nào đó, ký hiệu hai lần dấu hai chấm được sử dụng (::). Ví dụ như $\lambda ::$

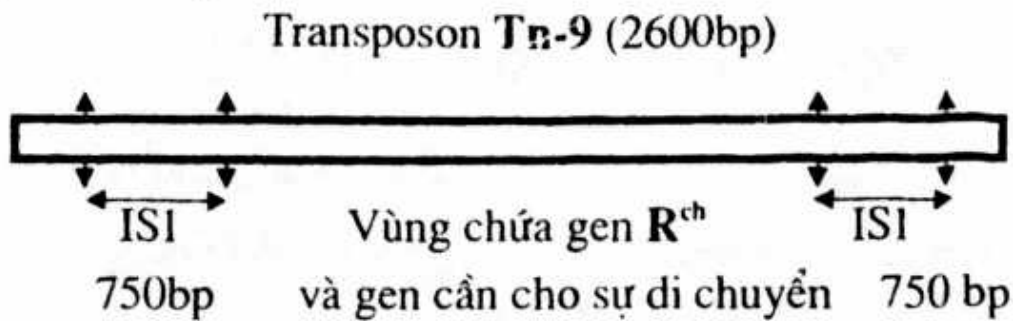
IS1 mô tả transposon IS1 gắn vào genome của bacteriophage λ . Transposons vì khuẩn không giữ một chức năng nào trong tế bào. Trình tự nucleotide ở một đầu IS thường lặp lại nhưng ngược chiều so với đầu kia (gọi là các inverted repeat). Ví dụ như GGTAT- X_n -ATACC. Do đó khi sợi đúp IS tách thành hai sợi đơn thì mỗi sợi này có khả năng hình thành liên kết bổ sung tại hai đầu của IS tạo cấu trúc dạng vòng (stem-loop) (Hình 1.3).



Hình 1.3: Dạng heteroduplex được tạo ra do liên kết tạo cặp bổ sung giữa hai sợi đơn ADN genome của hai phage trong đó một phage có chứa IS2.
 Tại hai đầu của IS2 là các đoạn có trình tự nucleotide ngược nhau

Transposon thường mã cho các enzym transposase làm nhiệm vụ nhận biết chuỗi nucleotide lặp lại ngược chiều (inverted repeat) để cắt transposon và di chuyển. Khi một IS được ghép vào vị trí bất kỳ của genome thì một đoạn ngắn ADN tại vị trí này được nhân đôi, tức là hai đầu của IS được chặn bởi các nucleotide giống nhau, sắp xếp theo cùng một chiều. Vì vậy đoạn ngắn ADN này được gọi là "lặp lại cùng chiều" (direct repeat). Chiều dài của chúng thường khoảng 9 bp (Hình 1.4). Dựa vào sự có mặt của các đoạn cùng chiều và ngược chiều có thể xác định được vị trí transposon ghép vào hoặc chuyển đi.

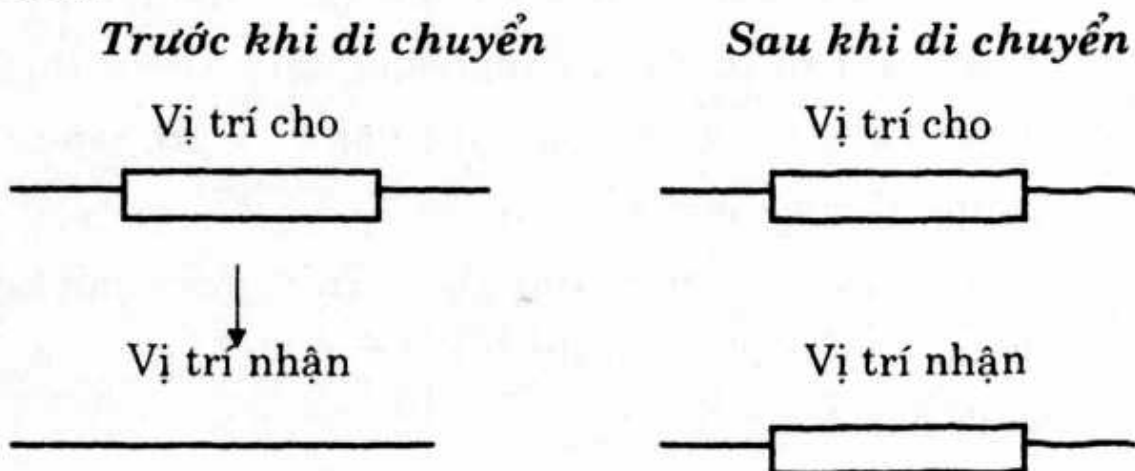
Tn-9 được giới hạn bởi IS-1 mà trình tự nucleotide của IS này sắp xếp theo cùng một chiều.



Hình 1.5: Cấu trúc của transposon Tn-9

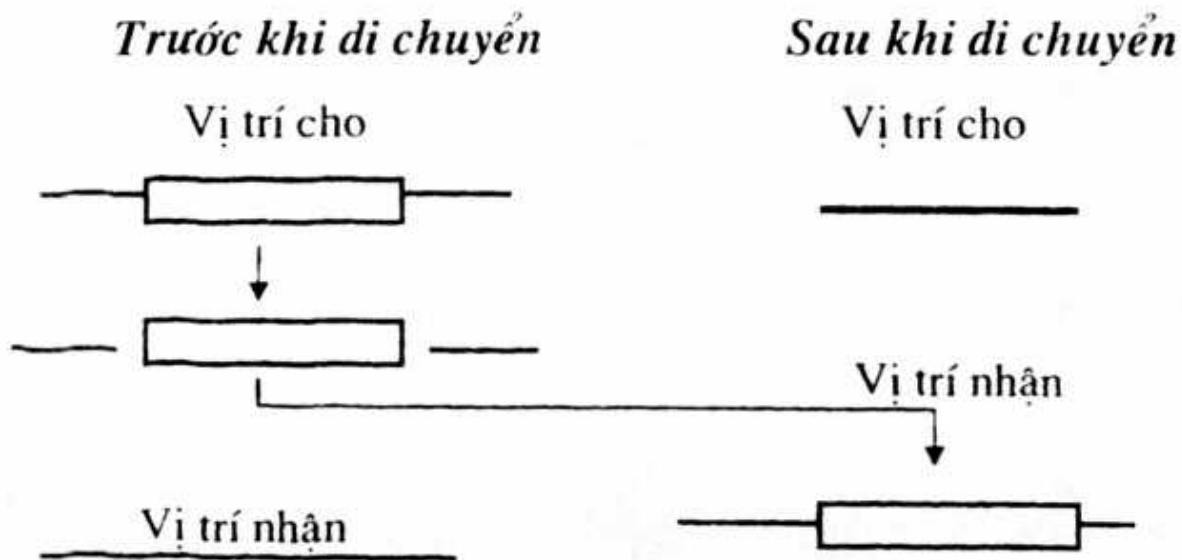
Quá trình di chuyển của một transposon từ vị trí cũ (donor) sang vị trí mới (recipient) xảy ra theo hai cơ chế khác nhau: Cơ chế sao y bản chính (transposon có mặt ở cả hai vị trí) và cơ chế tách ra khỏi vị trí cũ di chuyển đến vị trí mới.

*Cơ chế sao y bản chính được mô tả trên Hình 1.6. Phiên bản được sao chép từ vị trí cho và được ghép vào vị trí nhận. Như vậy mỗi lần di chuyển thì số lượng bản sao được tăng lên. Quá trình này liên quan đến hai loại enzym: transposase (tác động vào hai đầu bản gốc transposon) và resolvase (tác động lên bản sao).



Hình 1.6: Cơ chế sao y bản chính của một transposon từ vị trí cho sang vị trí nhận. Số lượng transposon tăng lên sau mỗi lần sao chép

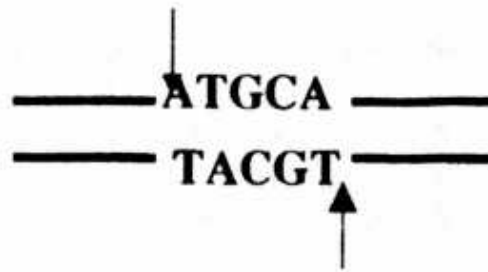
* Ngoài cơ chế trên, một transposon có thể tách ra khỏi vị trí cũ và ghép vào vị trí mới. Như vậy số lượng transposon không thay đổi (Hình 1.7). Kiểu di chuyển này chỉ đòi hỏi enzym transposase. Khi transposon chuyển đi, vị trí cũ bị gãy. Nó được nối lại nhờ cơ chế sửa chữa ADN trong tế bào.



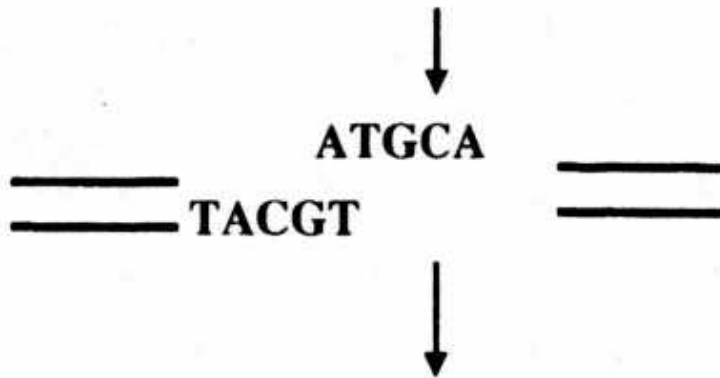
Hình 1.7: Cơ chế chuyển transposon không làm tăng số lượng

Khi transposon ghép vào vị trí mới, một đoạn nucleotide ngắt ở đó được sao thành hai bản, mỗi bản nằm ở một đầu của transposon (direct repeat). Cơ chế của quá trình này được mô tả trên Hình 1.8.

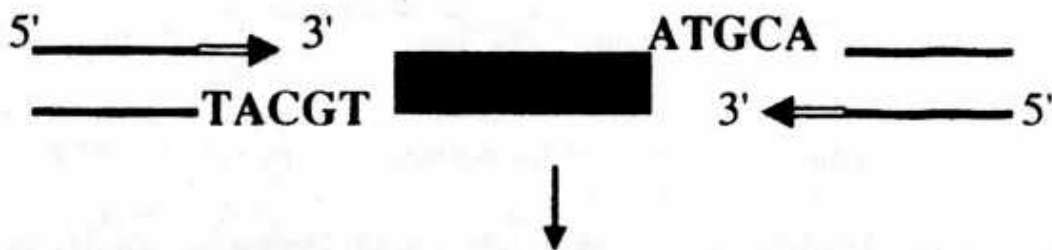
Tại vị trí nhận, mỗi sợi đơn ADN bị cắt lệch nhau vài nucleotide. Transposon nối vào các đầu cắt, tạo ra hai khoảng trống (gaps). Chúng được sửa chữa theo nguyên tắc tạo cặp bổ sung. Do đó đoạn nucleotide nằm giữa hai vết cắt được sao chép thành hai bản, mỗi bản ở một đầu và trình tự sắp xếp các nucleotide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



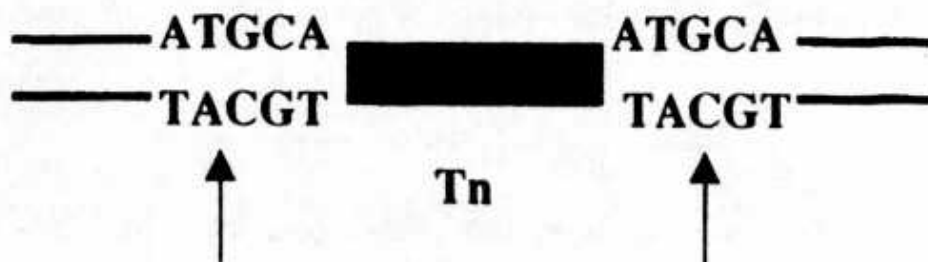
Vị trí transposon ghép vào hệ gen có hai sợi đơn bị cắt cách nhau 5 nucleotide



Transposon ghép vào genome

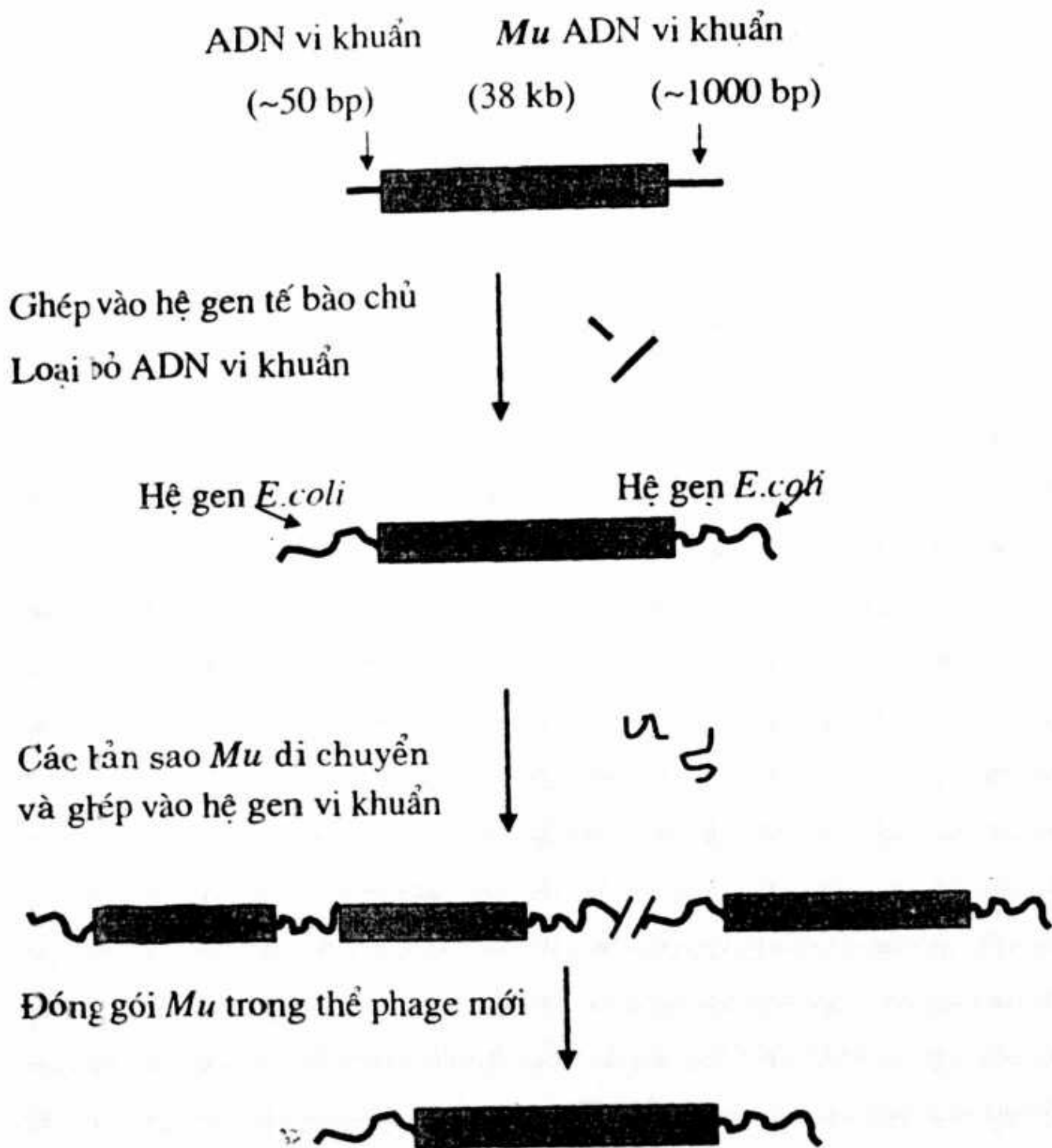


ADN polymerase sửa chữa vết trống



Chuỗi nucleotide "Lặp lại cùng chiều " ở vị trí gắn Tn vào

Hình 1.8: Đoạn ngắn ADN ở vị trí có transposon ghép vào gồm 5 nucleotide (ATGCA) được sao thành hai bản, chặn hai đầu của transposon. Chúng được gọi là các chuỗi nucleotide lặp lại cùng chiều "Direct repeat"



Hình 1.9: Tải bản của thực khuẩn thể *Mu* theo cơ chế sao y bản chính của *transposon*. Thực khuẩn thể *Mu* luôn được đóng gói cùng với hai đoạn ADN vi khuẩn trong thể phage. Khi *Mu* xâm nhiễm và ghép vào hệ gen tế bào chủ, các đoạn ADN này bị loại bỏ. *Mu* tái bản bằng cách sao y bản chính và ghép vào vị trí mới trên hệ gen vật chủ. Hệ gen *E.coli* có thể chứa hàng trăm bản copy *Mu* sau 1 giờ bị nhiễm thực khuẩn thể này

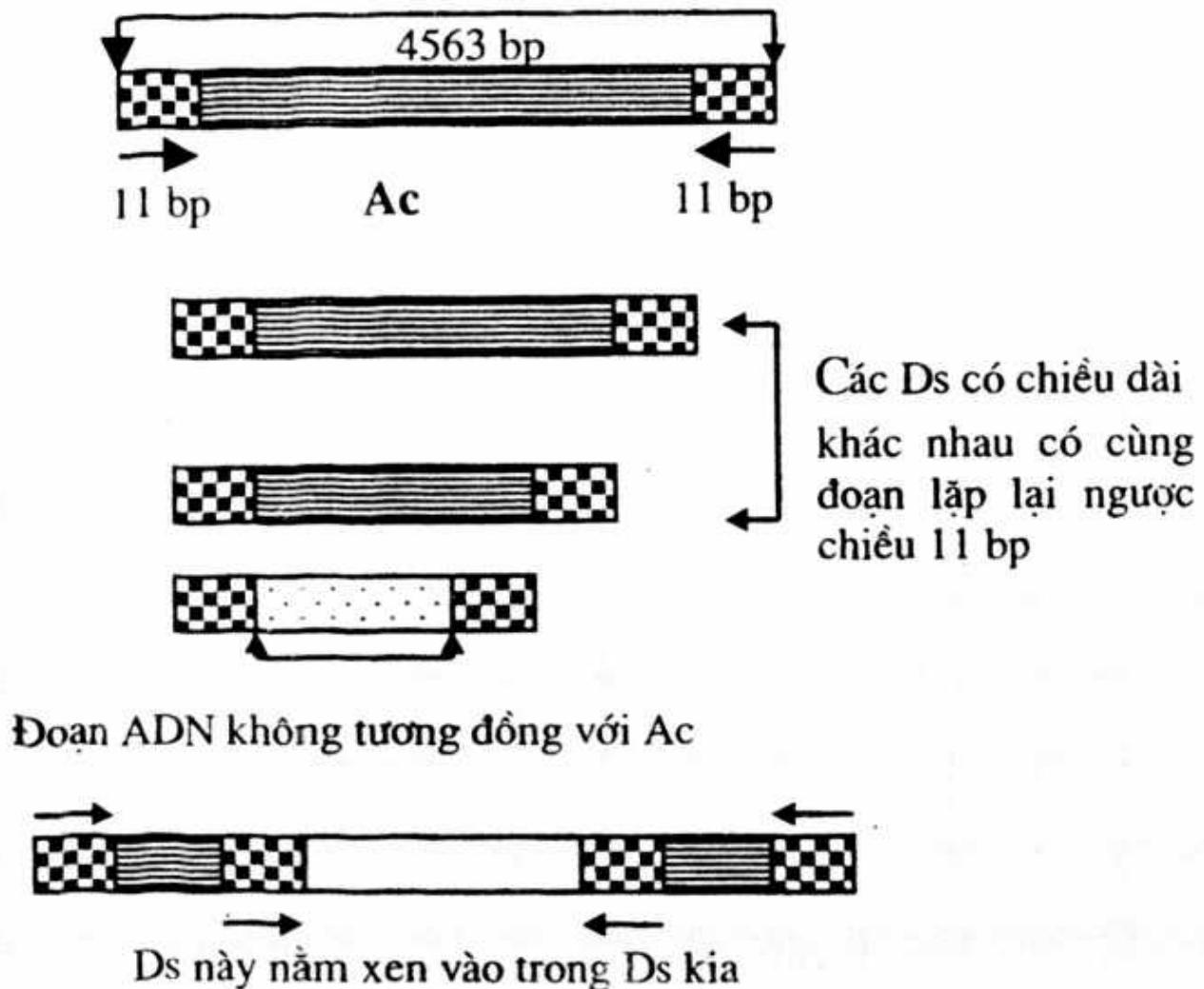
Sự tái bản của thực khuẩn thể *Mu* trong tế bào vi khuẩn được xem là một ví dụ điển hình về transposon. Tương tự như thực khuẩn thể λ , phage *Mu* có khả năng tồn tại ở hai trạng thái tan và tiềm tan. Tuy nhiên, đặc điểm riêng biệt của *Mu* trong trạng thái tan là chúng tái bản ADN theo cách thức sao y bản chính của transposon. Bản thân hệ gen *Mu* (38000 bp) chứa các gen liên quan đến quá trình di chuyển và các gen mã cho protein cấu trúc có chức năng đóng gói ADN tạo thể phage mới.

Giống như mọi transposon khác, *Mu* được giới hạn hai đầu bởi các đoạn nucleotide lặp lại ngược chiều và chúng tạo ra đoạn ngắn lặp lại cùng chiều (5 bp) của hệ gen vật chủ tại vị trí ghép vào. Điều đáng lưu ý là khi đóng gói trong thể phage mới hai đầu hệ gen *Mu* luôn luôn có các đoạn ADN của vi khuẩn tại vị trí *Mu* ghép vào. Các đoạn ADN này sẽ bị phân hủy khi *Mu* tiếp tục sự xâm nhiễm của mình (Hình 1.9).

Ở sinh vật eukaryot, các transposon còn được gọi là yếu tố kiểm soát (controlling elements). Chúng được nghiên cứu từ những năm 1940. Tuy nhiên cơ chế hoạt động của chúng ở mức độ phân tử chỉ mới được sáng tỏ trong những năm gần đây. Các nghiên cứu điển hình được tiến hành với transposon ở ngô và ở ruồi giấm *Drosophila*. Chúng di chuyển, sắp xếp và khởi động các gen ở những thời điểm đặc trưng cho quá trình sinh trưởng phát triển của cá thể.

Hai transposon Ac và Ds được nghiên cứu khá kỹ ở ngô. Chúng cùng thuộc vào một nhóm transposon, trong đó di chuyển của Ds phụ thuộc vào sự có mặt của Ac. Trình tự nucleotide của Ac gồm 4563 bp, được giới hạn hai đầu bởi 11 bp lặp lại ngược chiều, tiếp đến 8 bp lặp lại cùng chiều của genome.

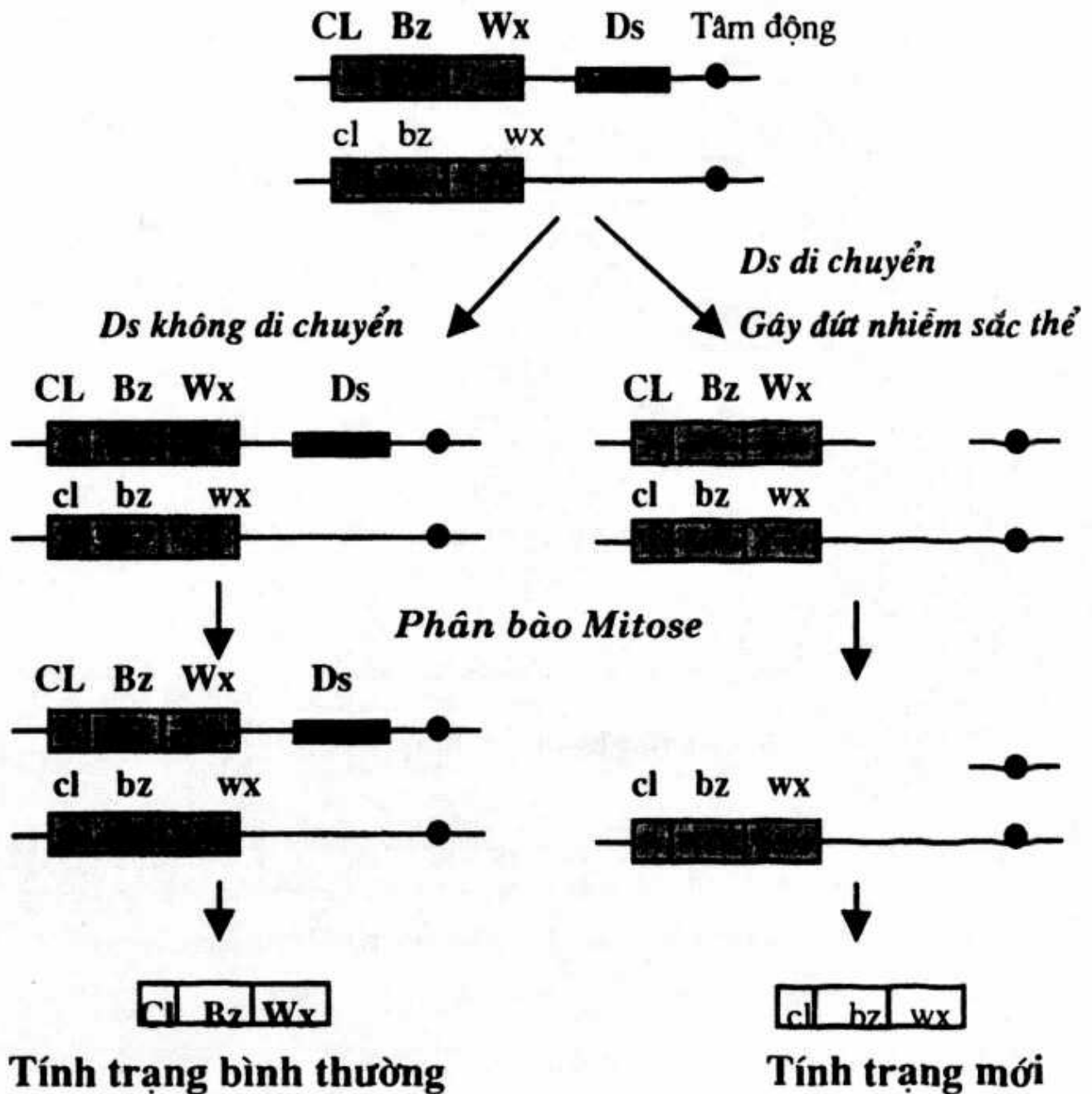
Mọi Ds chịu sự kiểm soát của Ac đều có đoạn lặp lại ngược chiều giống nhau mặc dù chiều dài của chúng thay đổi (Hình 1.10).



Hình 1.10: Cấu trúc của transposon Ac/Ds. Các Ds có chiều dài khác nhau (do Ac bị đột biến mất đoạn) hoặc có thể chứa đoạn ADN hoàn toàn không tương đồng với Ac, hoặc có thể nằm xen vào nhau. Tuy nhiên tất cả các transposon này đều được giới hạn bởi 11 bp lặp lại ngược chiều

Các transposon ở ngô thường ghép vào gần các gen, làm rối loạn hoạt động của chúng dẫn đến việc xuất hiện tính trạng mới nhưng không gây đột biến chết. Sự di chuyển của transposon xảy ra ở tế bào soma gây ảnh hưởng đến biểu hiện của một allele của một gen bất kỳ trong quá trình phát triển của cây. Trải qua

phân bào nguyên phân (mitose), con cháu của tế bào chứa allen đột biến đó sẽ có biểu hiện tính trạng mới (thường quan sát được ở hình dạng, màu sắc của hạt ngô). Thay đổi này xảy ra trong quá trình phát triển soma được gọi là "variegation" hay còn gọi là hiện tượng mosaic (xuất hiện các đốm) (Hình 1.11).



Hình 1.11: Hoạt động của transposon ở ngô. Di chuyển của yếu tố Ds gây đứt đoạn nhiễm sắc thể. Đoạn nhiễm sắc thể không có centromere có thể bị mất đi khi trải qua phân bào mitose. Nếu đoạn mất này mang các allen trội, các tế bào con cháu sẽ biểu hiện tính trạng mới do allen lặn qui định

Ở ruồi giấm *Drosophila melanogaster*, yếu tố P có khả năng di chuyển được phát hiện khi tiến hành lai giữa con đực dòng P với con cái dòng M. Hầu hết con lai bị bất dục, nhiễm sắc thể bị đứt gãy, đột biến. Hiện tượng rối loạn di truyền này chỉ xảy ra theo một chiều, tức là phép lai giữa con cái dòng P với con đực dòng M vẫn tạo ra các con lai bình thường. Hiện tượng này gây ra do hệ gen của các cá thể thuộc dòng P có chứa yếu tố di chuyển P. Yếu tố dài nhất gồm có 2907 bp có chứa gen mã cho transposase. Điều đáng chú ý là mặc dù có chiều dài khác nhau, các yếu tố P đều có mang các trình tự liên quan đến quá trình tương tác với transposase.

Quan sát quần thể ruồi giấm trong thiên nhiên cho thấy số lượng P thay đổi từ vài bản sao đến 50 copy/genome. Hơn nữa, những loài ruồi giấm phát hiện trước năm 1950 đều không có P trong genome. Phải chăng P chỉ mới xuất hiện trong hệ gen ruồi trong những năm cuối thế kỷ 20. Liệu sự có mặt của chúng có phải do virus xâm nhiễm ruồi giấm gây nên? Hiện tượng tương tự cũng được quan sát thấy ở vi khuẩn bị nhiễm thực khuẩn thể mang IS. Yếu tố IS xuất hiện trong hệ gen vi khuẩn thông qua quá trình tiếp hợp (transduction).

Cơ chế kiểm soát sự di chuyển của P phụ thuộc vào factor tồn tại trong tế bào chất của trứng (di truyền theo mẹ). Khi factor này có mặt thì chúng kìm hãm sự di chuyển của P. Vì vậy, tế bào trứng của con cái dòng P thụ tinh với con đực dòng M vẫn cho con lai bình thường do factor trong tế bào trứng ngăn cản P chuyển chỗ. Tuy nhiên, tế bào trứng dòng M thụ tinh với con đực dòng P cho phép P di chuyển gây ra những rối loạn bất thường trong cấu trúc hệ gen. Điều đó khiến con lai bị bất dục hoặc có các tính trạng lạ.

1.4. TƯƠNG TÁC CỦA T-ADN VỚI GENOME THỰC VẬT

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

Bệnh xuất hiện nốt sần hoặc mọc nhiều rễ trên thân chỉ xảy ra khi có mặt *Agrobacteria*. Tuy nhiên sau đó bệnh được duy trì không phụ thuộc sự tồn tại của vi khuẩn. Đó là do một số gen vi khuẩn đã được chuyển vào hệ gen cây chủ và hoạt động gây bệnh. Các gen vi khuẩn có khả năng di chuyển và hoạt động trong tế bào thực vật nằm trên plasmid Ti (Tumor inducing) của *A. tumefaciens* gây bệnh nốt sần hoặc trên plasmid Ri (Root-hairs inducing) của *A. rhizogenes* gây bệnh mọc lông rễ. Cũng giống như các khối u động vật, các tế bào thực vật có ADN vi khuẩn ghép vào hệ gen bị chuyển sang trạng thái mới, ở đó sự phát triển và biệt hoá của chúng hoàn toàn khác với các tế bào bình thường. Đó là do hoạt động của các gen vi khuẩn (prokaryot) trong genome của thực vật (eukaryot). Bình thường những gen này có mặt trong hệ gen vi khuẩn nhưng chúng chỉ được bật mở sau khi ghép vào hệ gen thực vật và chịu sự kiểm soát của tế bào cây chủ. Quá trình này có tính chất đặc hiệu, tức là một loại vi khuẩn chỉ có khả năng gây nốt sần trên một số loại cây chủ này mà không tương tác được với các loại cây khác.

Việc tạo nốt sần hay thực chất quá trình chuyển gen từ vi khuẩn sang hệ gen thực vật dẫn đến biến đổi trạng thái sinh lý của tế bào thực vật đòi hỏi các điều kiện sau:

- Phải có hoạt động của các gen trên 3 vùng *chvA*, *chvB*, *pscA* nằm trên nhiễm sắc thể của vi khuẩn để khởi động việc bám dính vi khuẩn vào thân cây.

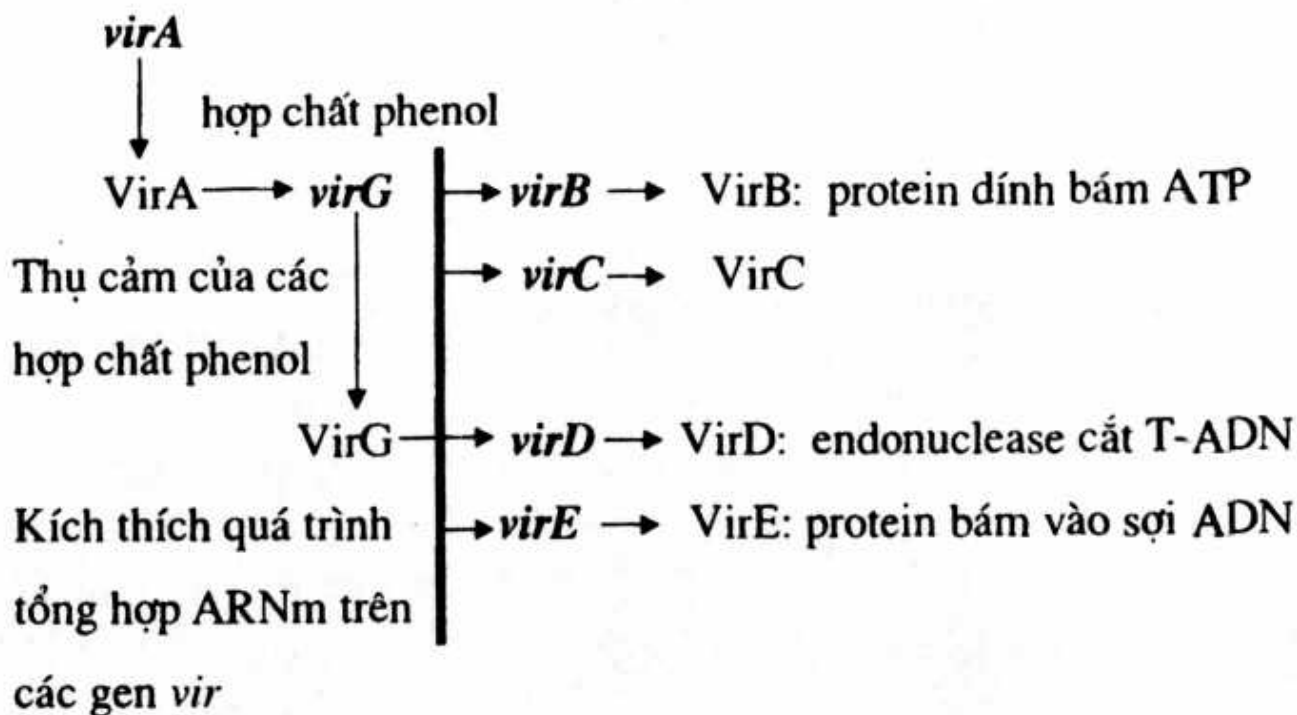
- Plasmid Ti phải mang vùng *vir* - ADN (nằm ngoài đoạn T-ADN). Vùng này mang các gen cần thiết cho việc tách và vận chuyển T-ADN sang tế bào thực vật.

- Các gen trên vùng T-ADN được ghép vào genome thực vật gây biến đổi trạng thái các tế bào này.

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

Protein VirA đóng vai trò quan trọng trong việc quyết định số loại cây chủ bị nhiễm bởi *Agrobacteria*. Trong thực tế, *Agrobacteria* không có khả năng xâm nhập vào cây một lá mầm. Có thể protein VirA không nhận biết được các tín hiệu do cây một lá mầm tiết ra. Tuy nhiên, năm đầu tiên của thế kỷ 21, các nhà khoa học đã thành công trong việc sử dụng T-ADN như vector chuyển gen vào tế bào nuôi cấy Hela có xuất sứ từ khối u tử cung của người.

Sản phẩm của những gen trên vùng *vir* còn liên quan chủ yếu đến việc cắt T-ADN ra khỏi plasmid và vận chuyển nó vào tế bào chủ. Bằng các thí nghiệm bổ sung chức năng (complementation test), thực nghiệm đã phát hiện ít nhất có 21 polypeptide (sản phẩm của các gen *vir*) cũng như xác định được chức năng của hầu hết các protein này trong quá trình vận chuyển T-ADN. Một số protein Vir giữ vai trò quan trọng được minh họa trên Hình 1.12.

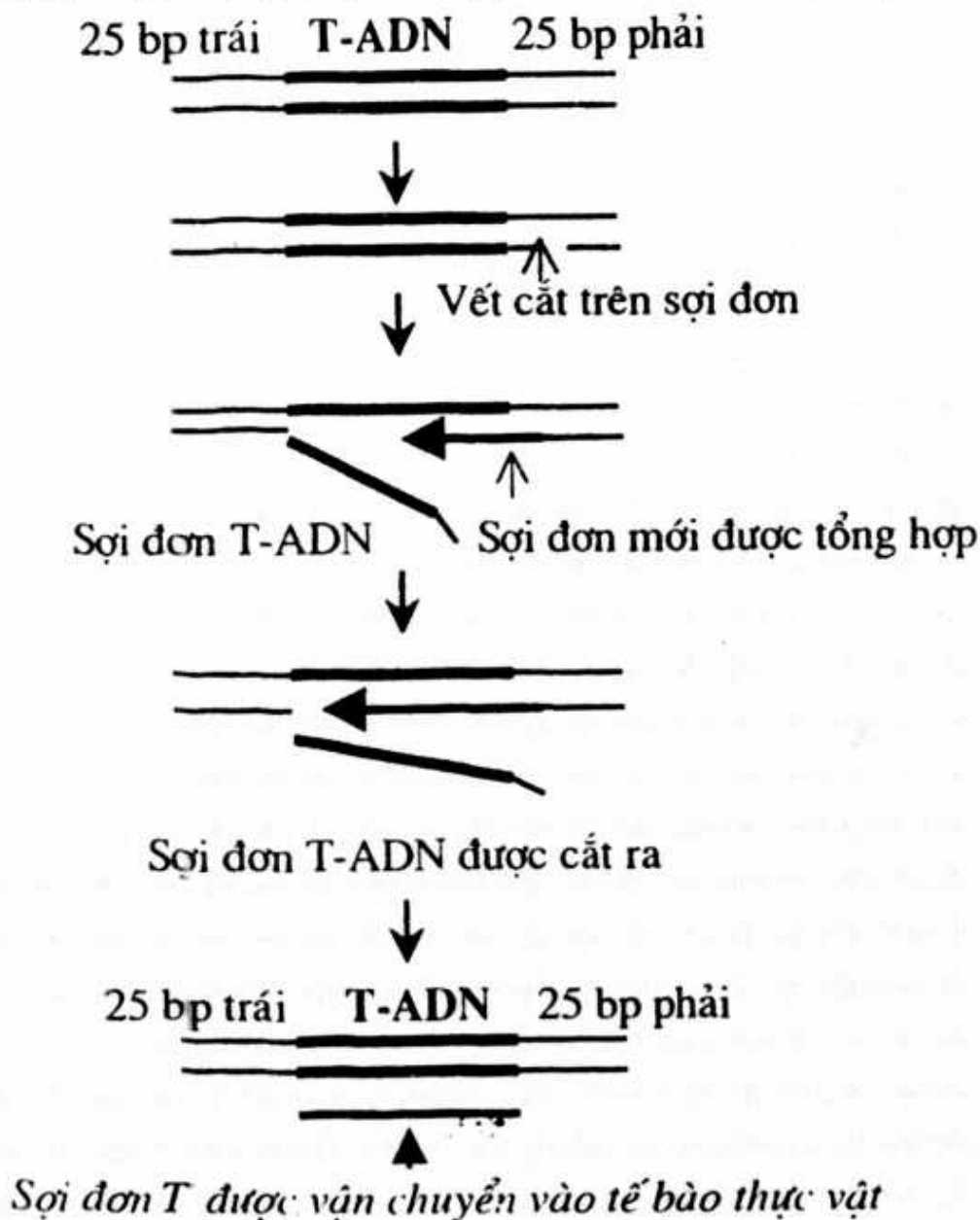


Hình 1.12: Chức năng hoạt động của 6 gen *vir* (trên plasmid *Ti*) trong quá trình vận chuyển T-ADN vào tế bào cây chủ

Protein VirC1 nhận biết và tương tác với các nucleotide nằm ở đầu bên phải của T-ADN. Mặc dù hai đầu T-ADN có trật tự tương đối giống nhau (chỉ sai khác nhau 2 nucleotide trong tổng số 25 nucleotide cần thiết cho sự vận chuyển T-ADN) nhưng các nucleotide đầu bên phải giữ vai trò quyết định cắt T-ADN ra khỏi plasmid. Đột biến ở đầu này khiến T-ADN không được cắt ra khỏi plasmid trong khi đột biến đầu bên trái hoàn toàn không ảnh hưởng đến quá trình vận chuyển T-ADN từ tế bào vi khuẩn vào trong nhân tế bào cây chủ. Điều đó cho thấy việc cắt T-ADN được bắt đầu ở phía bên phải và tiến dần sang bên trái.

Điều đặc biệt lưu ý là chỉ có một sợi đơn T-ADN được cắt ra và vận chuyển sang tế bào thực vật. Sợi đơn đó được gọi là sợi T. Đầu 5' của sợi T tương ứng với đầu bên phải của đoạn T-ADN. Các protein VirD1 và VirD2 liên quan đến phản ứng cắt sợi T ra khỏi plasmid. Tiếp theo đó, protein VirE2 tương tác với sợi T dọc theo chiều dài của sợi. Cơ chế cắt T-ADN được minh họa trên Hình 1.13.

Protein VirD2 giữ vai trò quan trọng trong quá trình vận chuyển. Cấu trúc VirD2 gồm nhiều vùng có hoạt tính khác nhau, liên quan đến các chức năng như cắt, vận chuyển sợi T. Bên cạnh việc tham gia phản ứng cắt sợi T tại đầu bên phải của T-ADN, protein VirD2 còn liên kết với đầu 5' của sợi này tạo thành phức. Nhờ đó T-ADN được vận chuyển dưới dạng phức ra khỏi tế bào vi khuẩn và xâm nhập vào nhân tế bào cây chủ. Bằng các thí nghiệm trên cây chuyển gen, người ta đã phát hiện được protein VirD2 có mặt trong nhân tế bào thực vật.



Hình 1.13: Quá trình cắt và vận chuyển T-ADN ở dạng sợi đơn vào tế bào thực vật

Vận chuyển sợi T ra khỏi tế bào vi khuẩn và ghép vào hệ gen tế bào cây chủ là một quá trình phức tạp, đòi hỏi sự tham gia nhiều protein. Trong số đó, operon *virB* nằm trên vùng *vir* giữ một vai trò đặc biệt. Operon này dài 9,5 kb mã cho 11 proteins, đa số là các protein tiết hoặc phân bố trên màng tế bào. Chúng bao gồm ATPase (VirB11) và các protein kỵ nước tạo nên kênh dẫn trên màng. Các nhà nghiên cứu cho rằng một trong các protein được mã bởi operon này phân bố phía ngoài màng làm nhiệm vụ tương tác với protein của tế bào thực vật, tạo kênh dẫn đưa T-ADN vào tế bào cây chủ.

Các gen trên T-ADN: T-ADN là một đoạn ADN có chiều dài khoảng 23 kb (tùy thuộc vào từng loại *A.tumefaciens*) nằm trên plasmid Ti. Hai đầu của đoạn ADN này có chứa 25 bp lặp đi lặp lại giống nhau hoàn toàn chỉ sai khác nhau ở hai nucleotide (imperfect repeat sequence). Các nucleotide đầu bên phải giữ vai trò quan trọng trong việc cắt T-ADN. Các nucleotide đầu bên trái đóng vai trò trong việc ghép T-ADN vào genome cây chủ.

T-ADN gồm hai nhóm gen. Nhóm thứ nhất gồm các oncogen mà cơ chế hoạt động của chúng khác biệt giữa *A. tumefaciens* và *A. rhizogenes*. Điều đó dẫn đến sự hình thành các nốt sần hoặc bệnh lông rề. Trong trường hợp xuất hiện nốt sần, T-ADN mang ba oncogen mã cho các enzym tham gia vào phản ứng tổng hợp các hormone sinh trưởng auxin và cytokinin. Chỉ khi T-ADN được ghép vào genome thực vật, các oncogen nằm trên T-ADN mới hoạt động một cách tự động, độc lập với quá trình điều hoà, kiểm soát chung diễn ra trong các tế bào thực vật. Do đó tế bào cây chủ nào có T-ADN ghép vào hệ gen lập tức phát triển không bình thường do rối loạn hormone sinh

trưởng mà T-ADN mã cho. Tại vị trí cây bị nhiễm *A. tumerfacien* xuất hiện các nốt sần, là tập hợp của các tế bào bình thường và tế bào bị biến đổi hệ gen.

Trong trường hợp với bệnh mốc nhiều lông rễ, R-ADN của *A. rhizogenes* có chứa các oncogen mà sản phẩm của chúng làm thay đổi ngưỡng nhạy cảm của tế bào thực vật đối với nồng độ hormone có mặt trong môi trường. Từ đó gây rối loạn sự phát triển của các tế bào có R-ADN ghép vào khiến cho rất nhiều rễ xuất hiện tại vị trí nhiễm.

Nhóm gen thứ hai có mặt trên đoạn T-ADN gồm các gen mã cho các enzym tham gia tổng hợp những phức chất dinh dưỡng cần thiết cho sinh trưởng và phát triển của vi khuẩn. Chúng được gọi chung là opines. Vi khuẩn sử dụng opines như nguồn cacbon và nitơ. Đặc biệt, khi T-ADN mang gen mã cho một loại opine nào đó thì ngay trên plasmid Ti, nằm ở ngoài đoạn T-ADN, có các gen tham gia quá trình chuyển hoá loại opine này, giúp cho vi khuẩn sinh trưởng và phát triển. Điều đáng lưu ý là opine được tổng hợp lại trở thành tín hiệu kích thích hoạt động của operon chứa các gen đồng hoá opine đó nằm trên plasmid Ti. Vì opine là protein được mã bởi các gen vi khuẩn, sự có mặt của chúng trong tế bào thực vật được xem là chỉ thị để phát hiện sự chuyển ghép thành công của T-ADN vào hệ gen thực vật.

T-ADN được vận chuyển vào trong nhân tế bào chủ và được ghép vào hệ gen. Có thể xuất hiện nhiều bản sao của T-ADN trong một genome. Dạng vòng của T-ADN đôi khi được tìm thấy trong tế bào cây chủ. Đây là dạng trung gian hay chỉ là sự liên kết ngẫu nhiên giữa hai đầu trái và phải của T-ADN đang là vấn đề cần làm sáng tỏ. Khi đã ghép vào genome vật chủ, các

gen trên đoạn T-ADN mới được hoạt động. Như vậy điều đáng chú ý là các gen trên T-ADN (prokaryot) chỉ hoạt động dưới sự kiểm soát của các factor sao chép mã trong hệ gen eukaryot. Nói chung, một gen được điều khiển bởi một promoter.

Agrobacterium có khả năng đưa các gen lạ vào genome thực vật. Vì vậy, chúng được sử dụng như các vector chuyên chở gen một khi các gen gây nốt sần trên T-ADN bị thay thế bởi gen nghiên cứu. Ngoài ra, promoter của các gen trên T-ADN đều là những promoter hoạt động mạnh trong tế bào nhận. Vì vậy, chúng được sử dụng làm promoter báo cáo hoặc promoter điều khiển gen lạ trong kỹ thuật chuyển gen. Kỹ thuật này được ứng dụng rất rộng rãi trong nông nghiệp. Ví dụ như đưa các gen chống chịu sâu bệnh, gen chịu được môi trường trồng trọt khắc nghiệt... vào các cây trồng quý hiếm hoặc cho năng suất cao.

1.5. SẮP XẾP VÀ KHUYẾT ĐẠI CÁC GEN TRONG GENOME

Chúng ta biết rằng genome có cấu trúc, tổ chức bền vững. Trao đổi chéo trong phân bào meiose là một trong những nguyên nhân gây ra biến đổi của hệ gen. Tuy nhiên, điều này xảy ra chủ yếu trong tế bào sinh dục mà không có trong các tế bào soma. Do đó việc phát hiện ra các cơ chế tự nhiên khác nhằm điều chỉnh lại genome là một điều hết sức lý thú, mặc dù những thay đổi này không xảy ra thường xuyên.

Việc sắp xếp lại genome dẫn đến những thay đổi sau:

- * Tạo ra những gen mới cần thiết cho từng hoàn cảnh mới.
- * Bật mở hoặc đóng gen. Đây chính là một trong những cơ chế điều hoà hoạt động của gen.

Trường hợp đơn giản nhất là hiện tượng sắp xếp lại các gen trong genome của nấm *Saccharomyces cerevisiae*. Ở trạng thái đơn bội, nấm này có thể tồn tại hai dạng giao phối (dạng α và α) và có khả năng chuyển từ dạng giao phối này sang dạng giao phối kia. Mỗi dạng được xác định bởi sự có mặt của các gen qui định dạng giao phối tại một vùng đặc biệt trên nhiễm sắc thể. Vị trí này được gọi là vị trí hoạt động hay cassette hoạt động MAT. Ngoài ra, các gen đó còn tồn tại ở hai vị trí khác trong genome. Tuy nhiên ở đó chúng đều bị bất hoạt. Hai vị trí này được gọi là vị trí tĩnh (cassette tĩnh HML và HMR). Mỗi vị trí tĩnh chỉ mang các gen qui định cho một dạng giao phối. Khi các gen được sao chép từ một vị trí tĩnh vào vị trí hoạt động thì ARNm mới được tổng hợp từ các gen đó. Như vậy, quá trình sao chép quyết định dạng giao phối của nấm. Bản gốc luôn được bảo tồn ở vị trí tĩnh, nó được sao chép và bản thứ hai xuất hiện ở vị trí hoạt động. Rõ ràng nếu vị trí tĩnh có mang đột biến thì chúng sẽ được sao chép vào vị trí hoạt động. Tuy nhiên nếu xảy ra đột biến ở cassette MAT thì tính trạng mới xuất hiện không bền. Một khi dạng giao phối chuyển đổi thì các gen cũ ở vị trí MAT bị thay thế bởi bản sao của vị trí tĩnh khác. Lúc đó đột biến bị loại đi, tính trạng mới biến mất.

Bên cạnh nấm *S. cerevisiae*, loài ký sinh trùng đơn bào *Trypanosome* cũng là một ví dụ điển hình về khả năng sắp xếp lại hệ gen trong quá trình gây bệnh ngủ ở vật chủ. *Trypanosome* có khả năng tránh được hệ thống miễn dịch của cơ thể nhờ thay đổi liên tục các kháng nguyên trên bề mặt tế bào của mình. Mỗi loại kháng nguyên được tổng hợp nhờ hoạt động của một gen tương ứng tại vị trí hoạt động. Gen này có thể bị thay thế bởi một gen mã cho loại kháng nguyên khác nằm ở một vị trí tĩnh

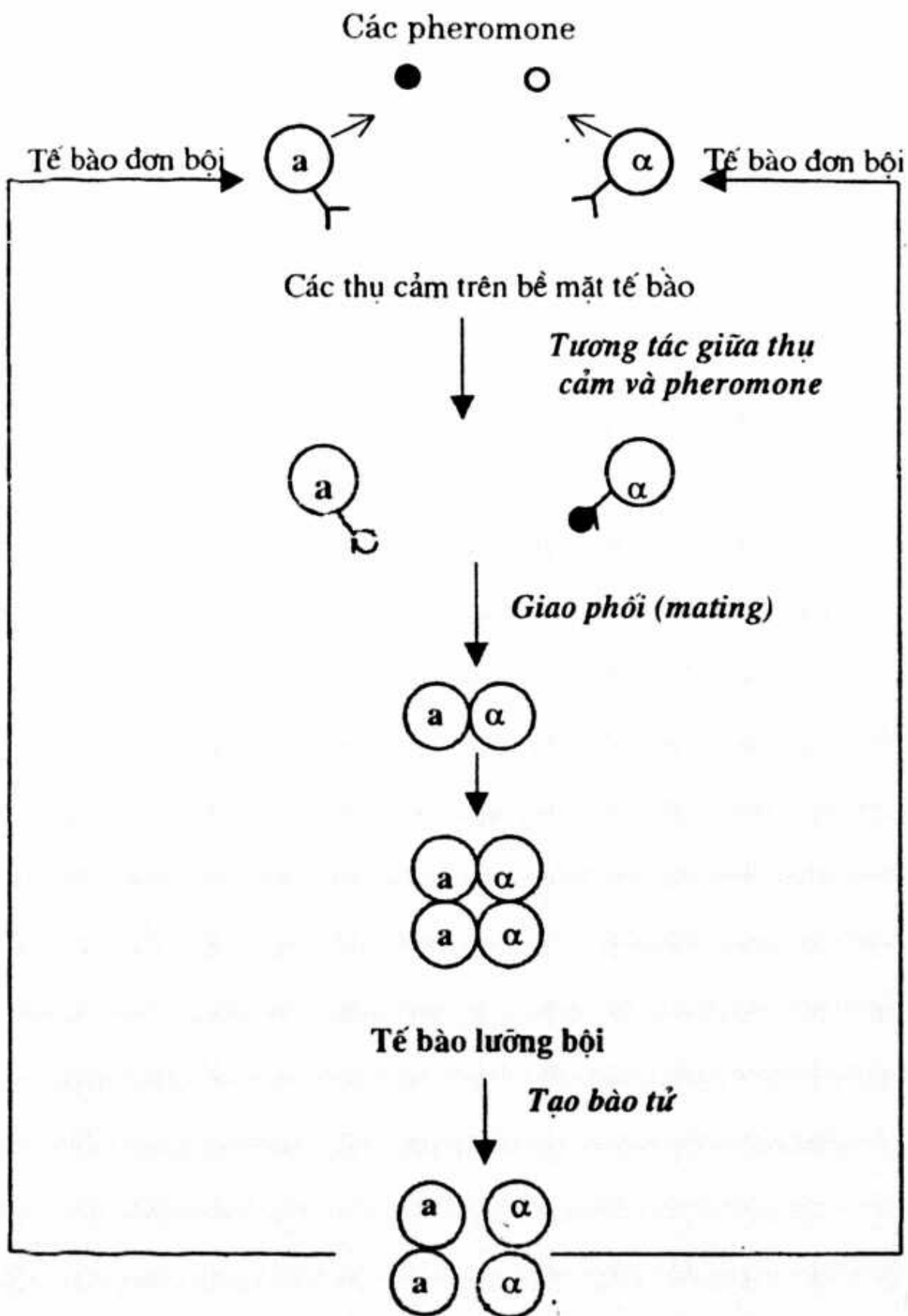
nào đó trong genome. Mỗi vị trí tĩnh chứa một gen ở trạng thái không hoạt động. Gen này chỉ được bật mở khi chuyển đến vị trí hoạt động. Trong genome *Trypanosome*, có rất nhiều vị trí tĩnh nhưng chỉ có một vị trí hoạt động.

Ngoài ra, trong quá trình sinh trưởng phát triển, số lượng bản sao của một số gen được tăng lên tạm thời trong quá trình phát triển các tế bào soma. Ví dụ như các gen mã cho ARNr có số lượng tăng khoảng 200 lần ở trứng ếch đang phát triển. Ngoài ra ở ruồi giấm *D. melanogaster*, các gen mã cho protein bề mặt chorion (~100 kb) cũng được nhân lên khoảng 50 lần ở các tế bào nang trứng (ovarian follicle) trong quá trình phát triển nhộng.

Khi ADN lạ được đưa vào tế bào eukaryot, nó có thể tồn tại dưới dạng nhiễm sắc thể tự do hoặc được ghép vào genome. Nếu xảy ra theo trường hợp thứ hai, genome sẽ mang những đột biến bền vững (có thể di truyền) và nhiều khi ADN lạ vẫn tiếp tục hoạt động làm xuất hiện các tính trạng mới. Việc đưa các gen lạ vào tế bào soma hoặc tế bào sinh dục mà vẫn duy trì hoạt động của những gen đó được gọi là chuyển nhiễm (transfection). Cá thể biểu hiện tính trạng mới nhờ hoạt động của gen lạ đưa vào tế bào sinh dục được gọi là cá thể chuyển gen.

1.5.1. Chuyển đổi dạng giao phối của nấm Saccharomyces cerevisia.

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

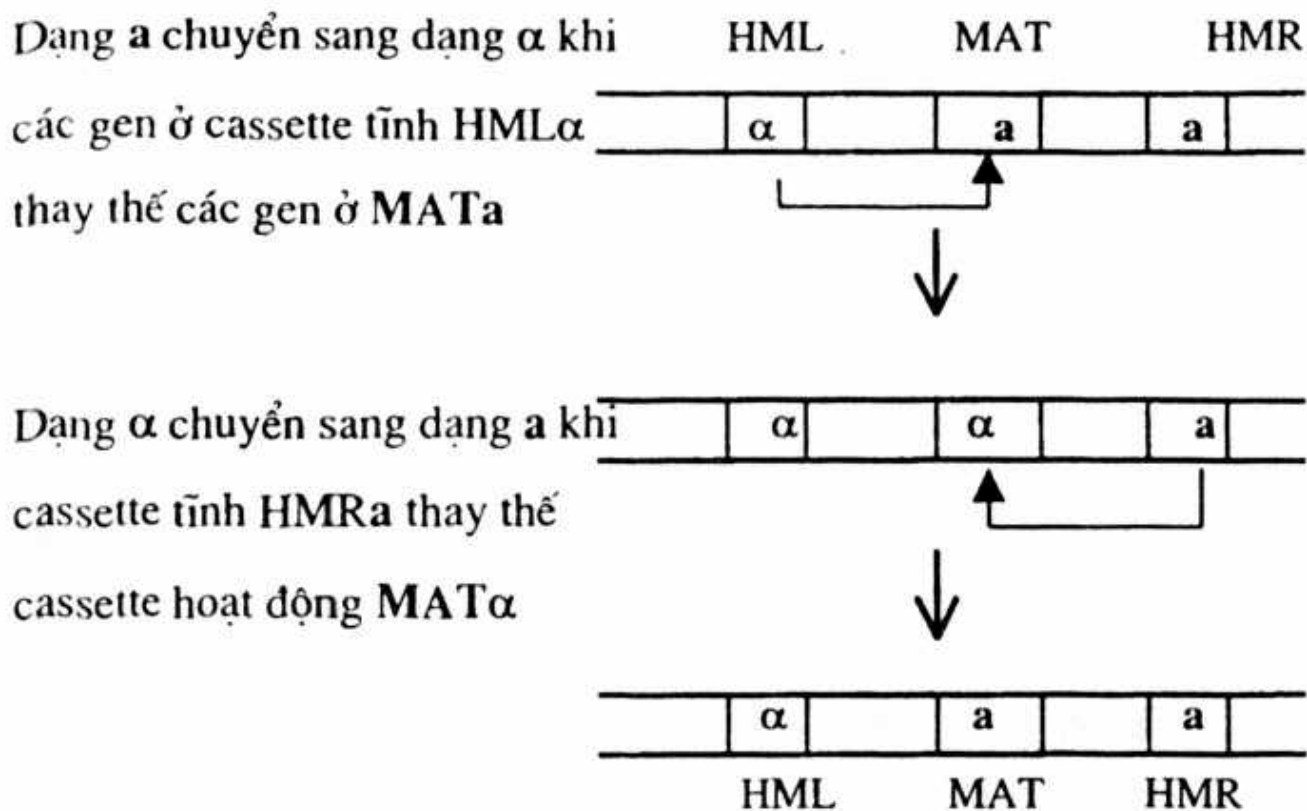


Hình 1.14: Chu kỳ sống của nấm *S. cerevisiae*. Các tế bào đơn bội loại a và α giao phối với nhau tạo lưỡng bội dị hợp tử

Dạng giao phối của một tế bào đơn bội được xác định nhờ thông tin di truyền có mặt trên một đoạn nhiễm sắc thể số 3. Vùng này được gọi là vùng hoạt động MAT. Khi allele MAT_a tồn tại ở vùng này thì tế bào thuộc loại a; còn khi allele MAT_α có mặt ở đó sẽ tương ứng với tế bào loại α. Khả năng giao phối chỉ xảy ra giữa các tế bào của hai loại. Chúng nhận biết nhau nhờ các pheromone do tế bào tiết ra. Tế bào loại a tiết ra a-factor, còn tế bào loại α tiết α-factor. Trên bề mặt tế bào loại a có thụ cảm nhận biết α-factor và ngược lại. Khi hai tế bào thuộc hai loại nhận biết nhau, chúng ngừng ở pha G1 của chu trình phát triển tế bào, sau đó tế bào và nhân của chúng kết hợp với nhau tạo tế bào lưỡng bội a/α (Hình 1.12).

Đối với một số loài nấm mang gen trội HO, dạng giao phối của chúng được chuyển đổi tự động sau mỗi chu kỳ tế bào. Do đó không cần quan tâm đến dạng giao phối ban đầu, trải qua một số nhân đôi, xuất hiện cả hai loại giao phối trong quần thể nấm, dẫn đến việc xuất hiện dạng lưỡng bội. Khả năng đó chứng tỏ thông tin di truyền cần thiết cho cả hai dạng giao phối đều tồn tại trong một genome, nhưng chỉ có một dạng được biểu hiện khi thông tin di truyền quyết định dạng đó có mặt tại cassette hoạt động MAT.

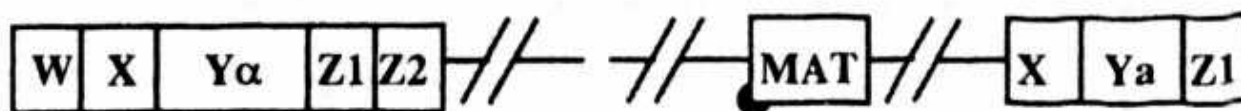
Ngoài ra còn có hai vùng khác (nằm trên cùng nhiễm sắc thể với vùng MAT) cần thiết cho quá trình chuyển đổi: vùng HML_α cần để chuyển sang dạng α còn vùng HMR_a cần để chuyển sang dạng a. Hai vùng này được gọi là các cassette tĩnh. Chúng chứa các gen qui định dạng a hoặc α nhưng các gen này không hoạt động. Quá trình chuyển đổi diễn ra theo cơ chế "cassette" như trên Hình 1.15.



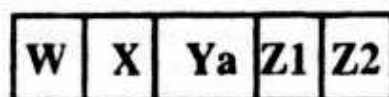
Hình 1.15: Chuyển đổi dạng giao phối khi cassette tĩnh thay thế cassette hoạt động. Các gen ở hai cassette chuyển đổi phải khác nhau

Trình tự các nucleotide ở vùng tĩnh và vùng hoạt động chỉ khác nhau một đoạn ngắn ký hiệu là Y_a và Y_α (Hình 1.14). Enzym HO-endonuclease nhận biết vị trí đặc hiệu tại ranh giới phân cách giữa Z và Y của cassette hoạt động MAT và cắt cả hai sợi ADN tại đó. Điều đặc biệt là enzym này không cắt ADN khi chúng có mặt ở cassette tĩnh mặc dù vị trí đặc hiệu tồn tại ở các vùng này. Vấn đề thu hút sự quan tâm đặc biệt và đã được sáng tỏ phần nào ở mức độ phân tử là sự gây bất hoạt các gen ở vị trí tĩnh. Chúng liên quan đến cấu hình không gian của nhiễm sắc thể, các yếu tố điều biến cấu trúc sợi nhiễm sắc. Nói cách khác, các gen ở vị trí tĩnh chịu kiểm soát của di truyền ngoại sinh (epigenetic).

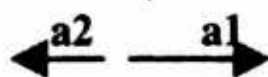
Nhiễm sắc thể số 3



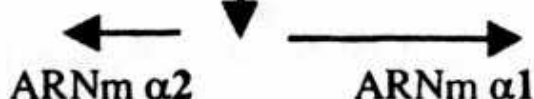
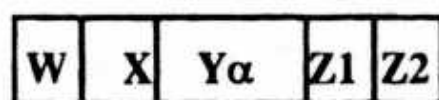
MAT α (dạng tế bào a)



Tổng hợp ARNm

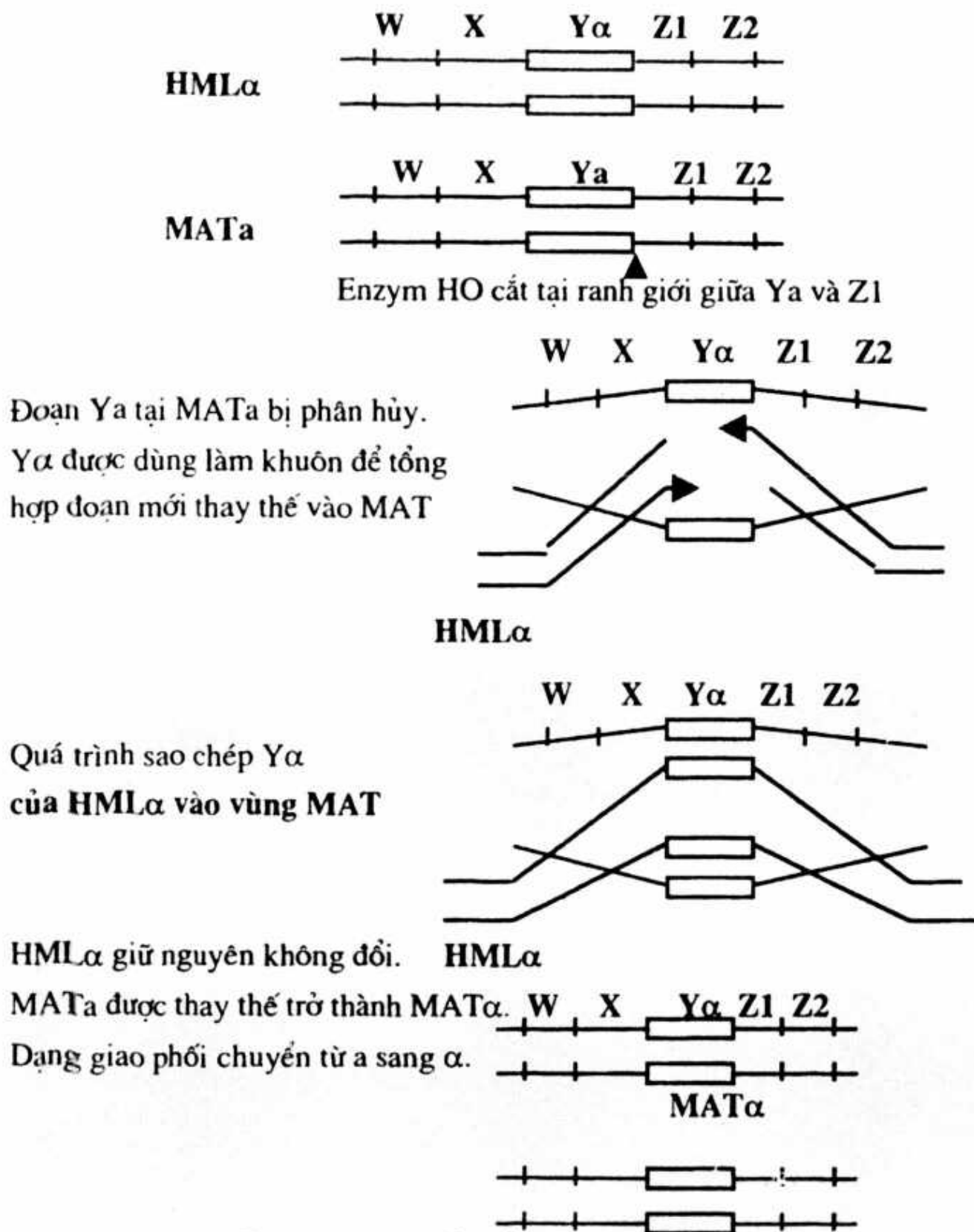


MAT α (dạng tế bào α)



Hình 1.16: Cấu trúc ba vùng HML α , MAT (a hoặc α) và HMRA trên nhiễm sắc thể số 3, xác định dạng giao phối a hoặc α ở nấm *S. cerevisiae*. Vùng MAT có thể tồn tại dưới hai dạng khác nhau quyết định bởi đoạn Y (Ya hoặc Y α), do đó sẽ tương ứng với các ARNm khác nhau

Sau khi đoạn Y của vùng MAT bị phân hủy hết, đoạn Y của một trong hai cassette tĩnh được dùng làm khuôn mẫu để sao chép vào vị trí bị phân hủy (Hình 1.15). Nếu đột biến xuất hiện ở vùng MAT (đoạn Y), tính trạng mới chỉ biểu hiện tạm thời. Một khi dạng giao phối chuyển đổi, các gen bình thường được sao chép vào vùng MAT và đột biến bị loại đi.



Hình 1.15: Quá trình chuyển đổi dạng giao phối từ a sang α

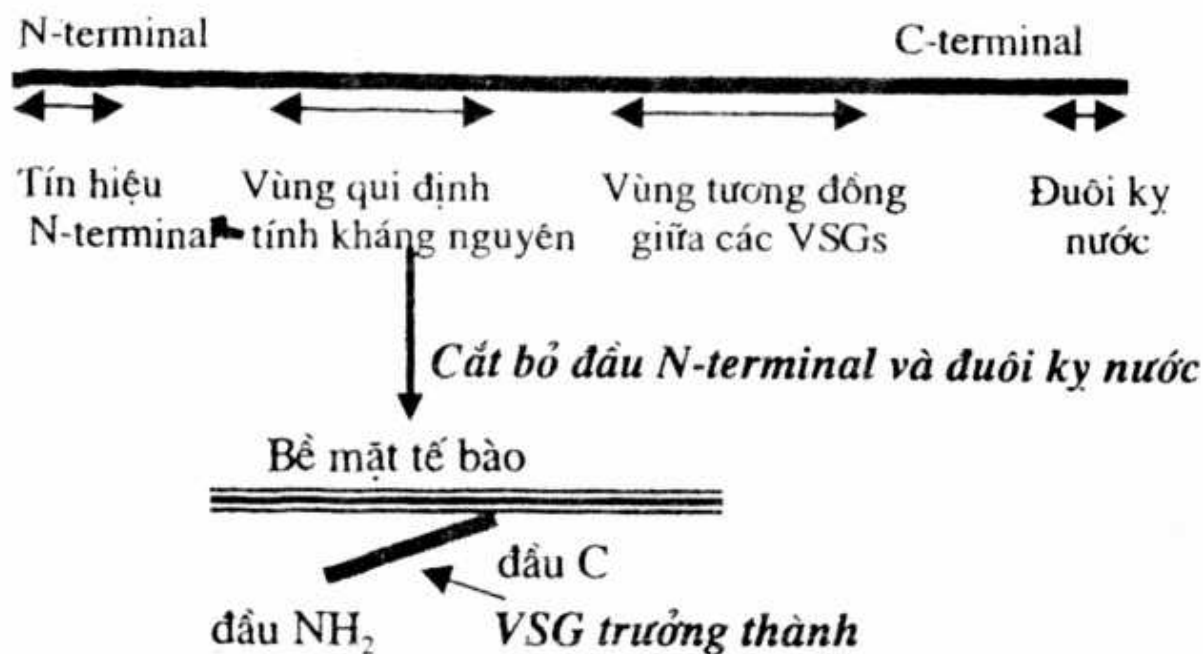
nhờ trao đổi gen giữa vùng MATa và HMLα trên nhiễm sắc thể số 3.

Vùng Yα của MAT bị cắt và phân hủy. Đoạn Yα của HMLα được sao chép vào chỗ trống nhờ ADN polymerase. HMLα giữ nguyên không đổi và được tiếp tục sử dụng cho các lần chuyển đổi sau

1.5.2. Chuyển đổi gen ở *Trypanosome*

Trypanosome là loài ký sinh trùng đơn bào gây bệnh ngủ ở người và một số động vật. Chúng trải qua một số lần biến đổi hình thái khi được truyền từ ruồi châu Phi sang vật chủ. Bề mặt tế bào *Trypanosome* được bao bọc một lớp đơn gồm 5×10^6 phân tử của một loại glycoprotein (ký hiệu là VSG-variable surface glycoprotein). Đây chính là kháng nguyên bề mặt của *Trypanosome* khi chúng xâm nhập vào vật chủ. Điều đáng chú ý là chúng có khả năng thay đổi kháng nguyên bề mặt, do đó tránh được phản ứng miễn dịch của tế bào chủ. Quá trình thay thế kháng nguyên bề mặt phụ thuộc vào sự chuyển đổi các gen mã cho chúng xảy ra ở một vị trí đặc biệt trong hệ gen (vị trí hoạt động). Chuyển đổi gen mã cho kháng nguyên bề mặt nhằm mục đích hoạt hoá gen mã cho kháng nguyên bề mặt mới thay thế cho kháng nguyên tồn tại trước đó. Khi một gen đang hoạt động bị thay thế bởi một gen khác thì tương ứng với việc xuất hiện kháng nguyên mới và loại bỏ kháng nguyên cũ.

Có nhiều loài *Trypanosome* xâm nhập các tế bào chủ khác nhau. *Trypanosome brucei*, nhiễm trên động vật nhưng không gây bệnh cho người được nghiên cứu khá kỹ. Chúng có thể thay đổi VSG một cách ngẫu nhiên với tần số 10^{-4} đến 10^{-6} qua một lần phân bào và khả năng này không phụ thuộc hệ thống miễn dịch của vật chủ. Cơ thể không sản xuất kịp thời các kháng thể để phản ứng lại các kháng nguyên khi chúng liên tục được đổi mới trên bề mặt *Trypanosome*.



Hình 1.16: Cấu trúc chung của một VSG.

Sau khi được tổng hợp, phân tử này trải qua biến đổi ở cả hai đầu NH_2 và COOH trước khi đính vào màng tế bào

Cấu trúc chung của một VSG được mô tả trên Hình 1.16. Một VSG vừa được tổng hợp dài khoảng 500 acid amin gồm tín hiệu N-terminal, tiếp theo là đoạn peptide quyết định tính kháng nguyên; đoạn peptide bảo thủ giữa các VGS và đuôi kỵ nước. Phân tử này được tổng hợp dưới dạng pre-protein. Do đó chúng phải trải qua biến đổi ở cả hai đầu NH_2 và COOH để trở nên dạng trưởng thành (mature form). Dạng này được đính vào màng tế bào ở đầu COOH .

Một loại *Trypanosome* có thể tạo ra ít nhất khoảng 100 VGS từ khi nhiễm cho đến khi gây chết vật chủ. Số gen mã cho VSG có thể nhiều hơn 1000. Tất cả các gen này đều nằm trong genome. Tuy nhiên, tại một thời điểm bất kỳ chỉ có một gen hoạt động tổng hợp nên một loại VSG. Do đó sự thay đổi kháng nguyên tương ứng với sự thay đổi hoạt động của gen. Khi một gen mới được bật mở, gen hoạt động trước nó phải bị ức chế

hoàn toàn. Lúc đó một kháng nguyên mới sẽ thay thế kháng nguyên tồn tại trước nó.

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

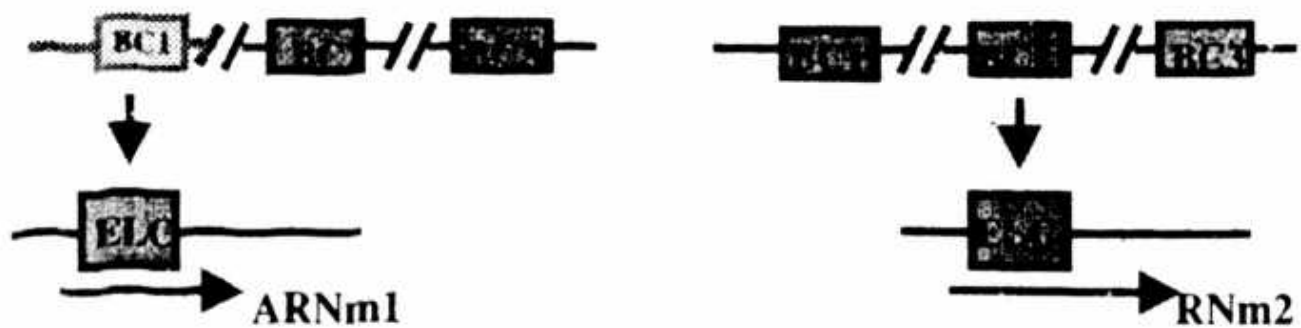
- * Các gen nằm ở telomere (khoảng 5-15 kb) > 200 gen

- * Các gen nằm cách telomere hơn 50 kb.

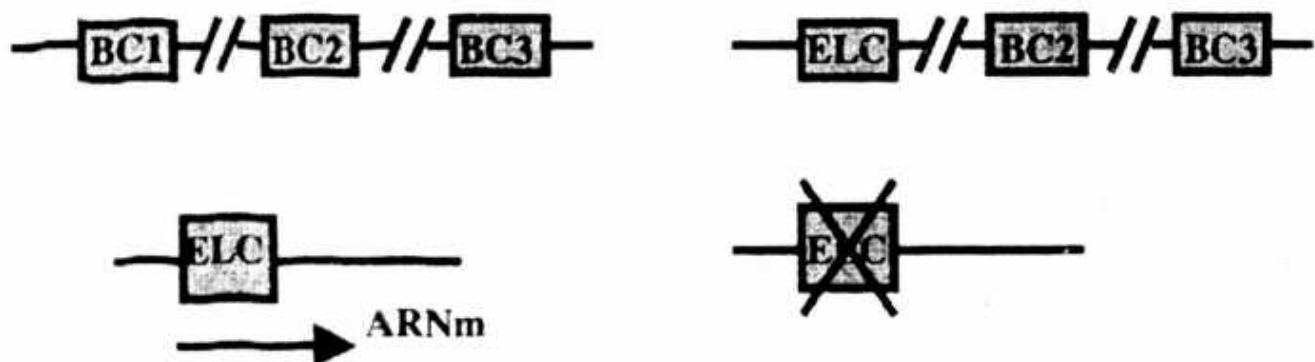
Tương tự như ở nấm *S. cerevisiae*, các gen mã cho VSG nằm rải rác trong genome và ở trạng thái không hoạt động. Một gen bất kỳ trở nên hoạt hoá chỉ khi nó được sao chép vào vị trí hoạt động (expression site) trong khi nguyên bản của nó vẫn được bảo tồn ở vị trí tĩnh. Bản sao của bản gen gốc vào vị trí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

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

- * Vị trí hoạt động thay đổi: Khi cần tổng hợp một VSG mới, gen ở vị trí hoạt động cũ bị ngừng và gen ở một vị trí khác (gần telomere) được khởi động (Hình 1.17).



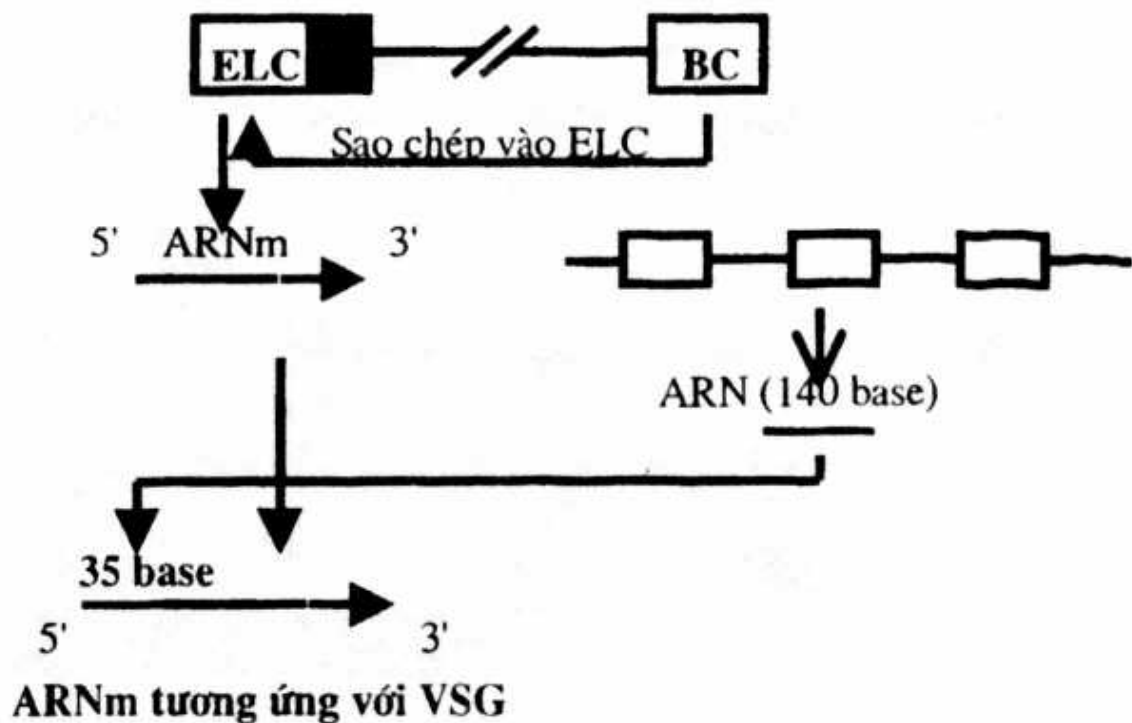
a/ Các gen ở vị trí tĩnh (còn gọi là bản gen gốc) nằm gần telomere hoặc phía trong telomere được sao chép vào ELC (vị trí hoạt động) và được hoạt hoá.



b/ Vị trí hoạt động thay đổi: vị trí cũ ngừng hoạt động, vị trí mới gần telomere được khởi động. Vị trí này có thể là một bản gen gốc được hoạt hoá hoặc là một gen gốc nằm phía trong telomere được sao chép vào vị trí mới.

Hình 1.17: Gen VSG được tạo ra nhờ sao chép từ nguyên bản gốc vào vị trí hoạt động hoặc nhờ hoạt hoá gen nằm ở telomere

Một số phân tử ARNm tương ứng với các VSGs khác nhau được phân lập và được xác định trình tự nucleotide (thông qua cADN). Điều đáng ngạc nhiên là phần oligonucleotide ở đầu 3' của mọi gen VSG đều khác với phần 3' của các ARNm được phiên mã từ các gen đó. Mặt khác các gen này không có phần 5' giống như các ARNm. Như vậy các ARNm không được tổng hợp hoàn toàn trên khuôn mẫu các gen này. Phần 3' của các ARNm tương ứng với phần 3' của vị trí hoạt động (ELC), trong khi phần 5' (gồm 35 nucleotides) được tổng hợp từ những đoạn ADN khác và được gắn vào ARNm (hiện tượng trans-splicing) (Hình 1.18).



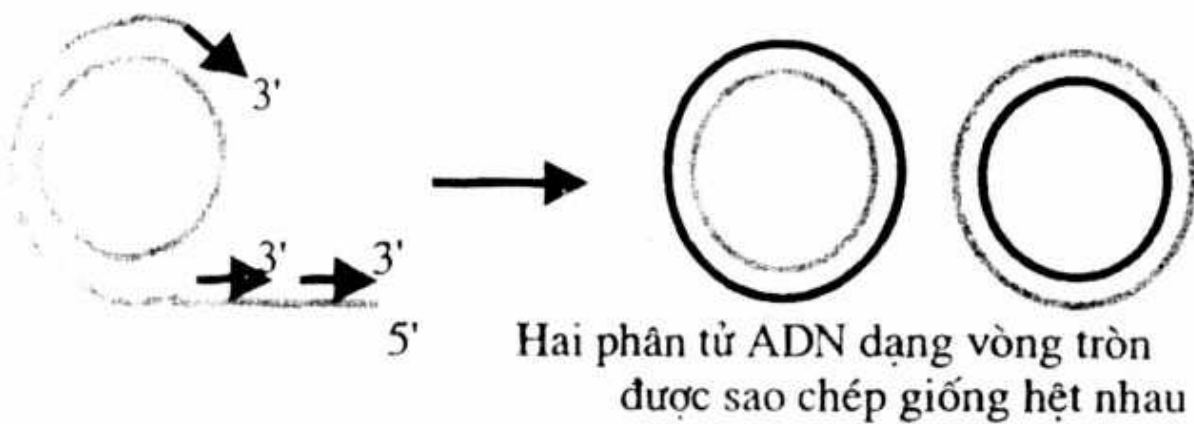
Hình 1.18: Phân tử ARNm tương ứng với một VSG.

Đoạn 35 nucleotide (phiên mã từ ADN nằm trong genome) được gắn vào phân tử ARNm phiên mã từ gen ở vị trí hoạt động

Thay đổi kháng nguyên không phải là hiện tượng duy nhất ở *Trypanosome*. Vi khuẩn *Borrelia hermsii* (gây bệnh ở người và động vật) cũng có khả năng này. Như vậy vi sinh vật gây bệnh prokaryot và eukaryot đều tránh được hệ thống miễn dịch nhờ thay đổi các kháng nguyên bề mặt

1.5.3. Tăng số lượng các gen mã cho ARNr và một số gen khác ở tế bào eukaryot

Việc tăng số lượng của một gen đặc biệt nào đó phụ thuộc từng điều kiện cụ thể của tế bào và xảy ra không phổ biến. Các bản sao có thể nằm tập trung thành một nhóm gồm bản sao này nối tiếp bản sao khác hoặc có thể tồn tại như những đoạn ADN có khả năng tái bản độc lập. Ví dụ điển hình là sự nhân bản của gen mã cho ARNr ở trứng ếch. Trứng ếch có đường kính khoảng 2-3 mm, dự trữ rất nhiều ARNr. Chúng được phiên mã từ rất nhiều gen ADNr. Các gen này được nhân lên (khoảng 2000 lần) theo cơ chế "vòng tròn lăn" trong quá trình phát triển và tồn tại dưới dạng các vòng tròn khép kín (Hình 1.19).



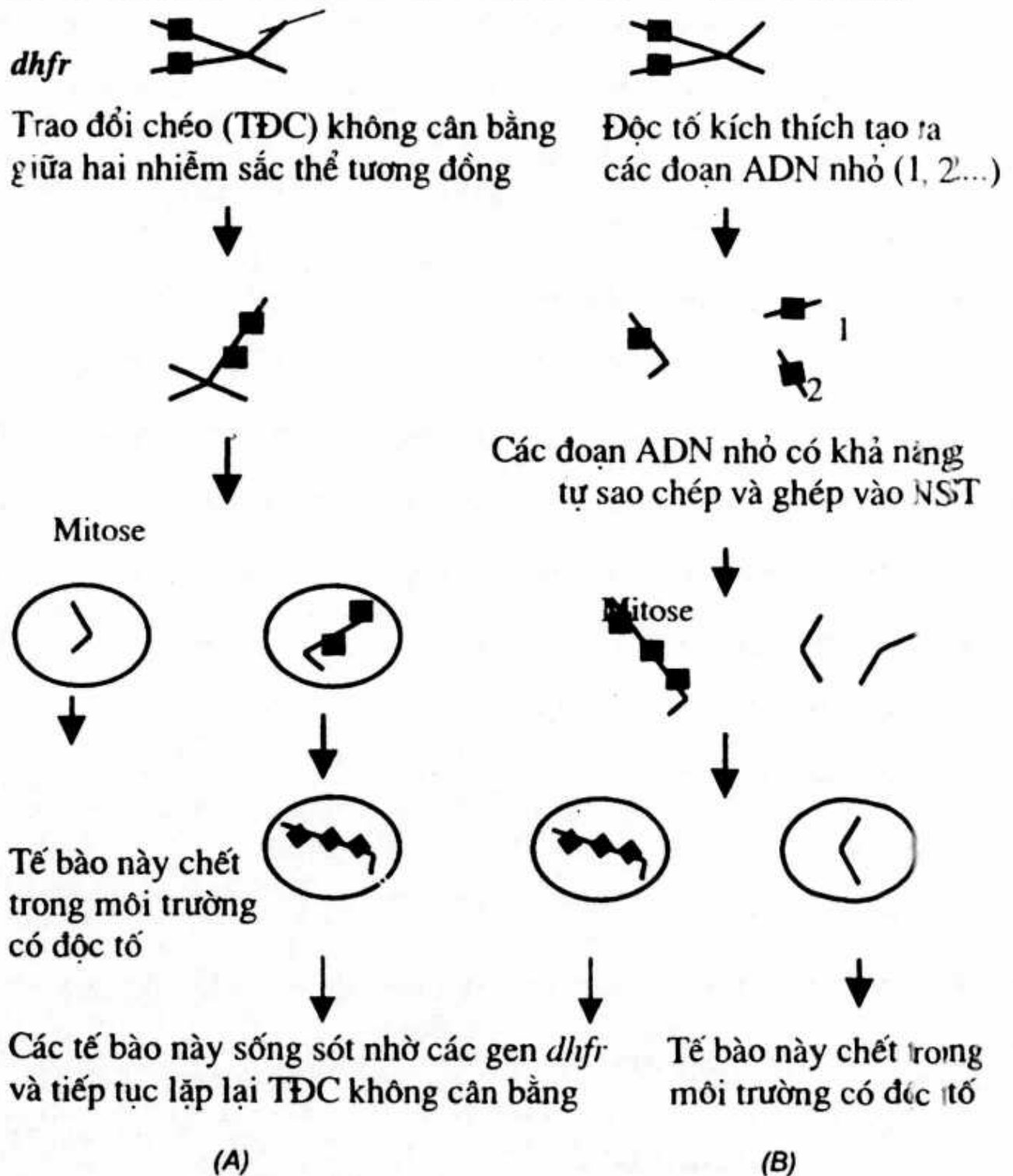
Hình 1.19: Tăng số lượng gen mã cho ARNr theo cơ chế "vòng tròn lăn" (rolling circle)

Khi nuôi cấy các tế bào động vật có vú trong môi trường đặc biệt, ADN tại một số vị trí trong genome được nhân lên. Thí dụ điển hình nhất là nuôi cấy các tế bào ung thư trong môi trường chứa độc tố methotrexate. Chất này ức chế hoạt tính của enzym dihydrofolate reductase (DHFR) giữ vai trò trong tổng hợp các nucleotide của ADN. Các tế bào ung thư nuôi cấy trong môi trường có chất độc này phát triển thành các quần lạc kháng lại độc tố. Khi nồng độ chất độc tăng dần, nồng độ DHFR cũng tăng theo, có thể đạt tới 1000 lần lớn hơn mức bình thường. Nồng độ enzym tăng do số lượng các gen mã cho chúng tăng. Cơ chế chính xác của hiện tượng này chưa rõ ràng, nhưng có thể xảy ra theo hai cách:

- * Trao đổi chéo không cân bằng giữa hai sợi nhiễm sắc tử (chromatide) của nhiễm sắc thể dẫn đến một số tế bào không có gen *dhfr* và một số khác có 2 copy của gen này. Trong môi trường có độc tố, trao đổi chéo không cân bằng được lặp đi lặp lại và các tế bào chứa nhiều gen *dhfr* vẫn phát triển tốt trong môi trường này (Hình 1.20A).

- * Các đoạn ADN (100-1000 kb) chứa 2-4 gen *dhfr* (~31 kb/gen) được sao chép từ nhiễm sắc thể bình thường tạo ra các nhiễm sắc thể rất nhỏ, không có tâm động. Các nhiễm sắc thể mini này ghép vào các nhiễm sắc thể bình thường khác. Quá

trình này lặp đi lặp lại và qua một số lần phân bào mitose, tế bào nào mang số lượng lớn các gen *dhfr* càng có điều kiện phát triển thuận lợi trong môi trường chứa độc tố (Hình 1.20B).



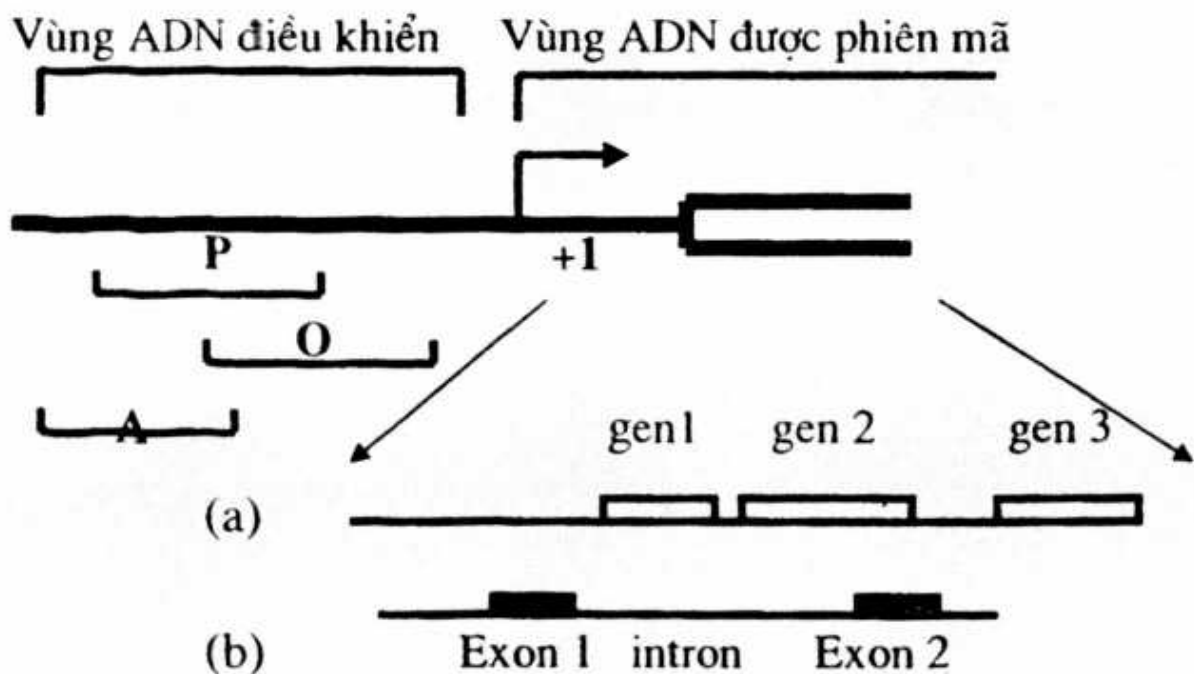
Hình 1.20: Hai cơ chế tăng số lượng gen *dhfr* trong môi trường nuôi cấy tế bào ung thư chứa độc tố ức chế hoạt tính enzym DHFR. Độc tố tạo điều kiện thuận lợi để chọn lọc các tế bào chứa số lượng lớn gen này

Chương 2

CẤU TRÚC VÀ HOẠT ĐỘNG CỦA GEN

2.1. CẤU TRÚC GEN

Một gen thường được xem gồm có hai phần, phần mang mã di truyền được phiên mã sang phân tử ARN và phần ADN làm nhiệm vụ điều khiển hoạt động của gen (Hình 2.1). Cả hai phần đều có cấu trúc cơ bản khá giống nhau ở mọi sinh vật prokaryot và eukaryot. Tuy nhiên gen eukaryot còn có cấu trúc đặc thù liên quan đến các cơ chế kiểm soát khác nhau đối với hoạt động của gen.



Hình 2.1: Cấu trúc chung của gen prokaryot (a) và eukaryot (b)

Vùng ADN điều khiển thường chứa các vị trí nhận biết cho enzym ARN polymerase (vị trí promoter-P), cho các protein hoạt hoá (vị trí kim hãm-O). Đặc biệt các gen eukaryot còn có các vùng ADN tăng cường (enhancers) nằm trước, nằm sau hoặc ngay trong gen tham gia điều khiển hoạt động của gen. Phần mang mã di truyền ở gen eukaryot thường bị ngắt quãng bởi các intron

Đoạn ADN làm nhiệm vụ điều khiển hoạt động của gen thường chứa các vị trí nhận biết cho enzym ARN polymerase (vị trí promoter-P), cho các protein khác nhau làm nhiệm vụ hoạt hoá (vị trí hoạt hoá-A), hoặc ức chế gen (vị trí kìm hãm-O). Nếu gọi vị trí của nucleotide đầu tiên được phiên mã sang phân tử ARN là +1 thì các nucleotide thuộc đoạn ADN điều khiển nằm phí trước vùng mang mã được đánh dấu (-). Ví dụ như vùng promoter bao gồm đoạn ADN nằm ở vị trí -35 đến -10 ở gen prokaryot và ở vị trí -25-30 đến -75 -90 ở gen eukaryot. Ngoài ra các vị trí A hoặc O đều được nhận biết bởi các protein điều khiển. Những protein này có thể liên kết với ADN, với ARN polymerase hoặc với các protein khác làm tăng cường hoặc kìm hãm hoạt động của gen ở mức độ phiên mã ARN.

Vùng ADN mang mã di truyền được phiên mã sang phân tử ARNm. Quá trình này được thực hiện theo chiều 5' --->3' trên sợi ARNm đang được tổng hợp. Không phải mọi mã di truyền trên phân tử ARNm đều được dịch mã sang phân tử protein. Hai đầu 5' và 3' của phân tử ARNm gồm một số nucleotide không được dịch mã mà liên quan đến tính bền vững của ARNm hoặc tham gia kiểm soát quá trình dịch mã. Hai đoạn này được gọi là đoạn không dịch mã 5' và 3' (untranslated region). Đoạn không dịch mã 5' nằm trước điểm khởi đầu dịch mã (ba nucleotide mã cho Methyonine đầu tiên của chuỗi peptide) và đoạn không dịch mã 3' nằm ở sau điểm kết thúc dịch mã (stop codon). Do tế bào prokaryot không có cấu trúc nhân nên quá trình phiên mã (tổng hợp ARNm) và dịch mã (tổng hợp protein) xảy ra đồng thời. Còn các phân tử ARNm eukaryot được phiên mã trong nhân, sau đó phải cắt bỏ các intron, chịu biến đổi tại các đầu 5' và 3' trước khi được vận

chuyển ra ngoài tế bào chất để dùng làm khuôn tổng hợp protein.

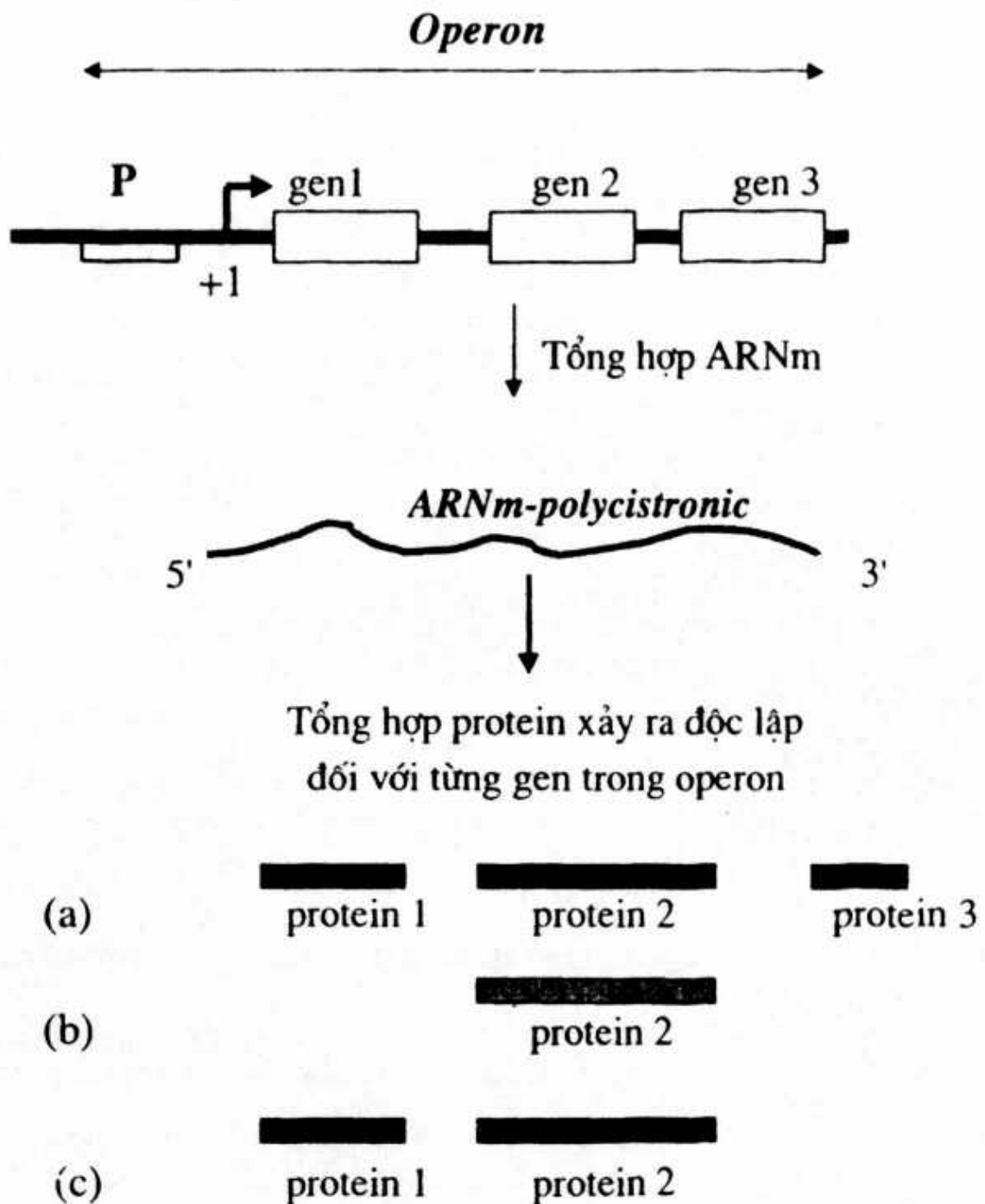
Hoạt động của một gen được đánh giá thông qua quá trình phiên mã sang phân tử ARNm và quá trình dịch mã tổng hợp protein. Hoạt động này được kiểm soát rất chặt chẽ bằng các cơ chế khác nhau ở mọi giai đoạn như bắt đầu và kết thúc phiên mã, quá trình biến đổi ARNm, quyết định tính bền vững và kiểm tra lại thông tin di truyền trên các phân tử này v.v... Do cấu trúc sắp xếp các gen prokaryot khác với các gen eukaryot nên sự phối hợp giữa các cơ chế điều khiển mang tính chất riêng biệt cho từng hệ gen. Chúng ta hãy xét cấu trúc hay gặp của trong hệ gen prokaryot và eukaryot và một số cơ chế điều khiển điển hình.

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

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

Tuy nhiên điều cần ghi nhớ là quá trình dịch mã trên phân tử ARNm-polycistronic xảy ra hoàn toàn độc lập với nhau. Mỗi đoạn tương ứng với một gen trên phân tử này đều có vị trí bám của ribosome, có mã bắt đầu và kết thúc tổng hợp chuỗi

polypeptide riêng biệt. Do đó tốc độ tổng hợp các protein trên một phân tử ARNm polycistronic hoàn toàn khác nhau (Hình 2.2).



Hình 2.2: Cấu trúc operon trong hệ gen vi khuẩn. Quá trình phiên mã xảy ra đồng thời, tạo phân tử ARNm-polycistronic nhưng quá trình dịch mã xảy ra hoàn toàn độc lập. Tất cả các gen

(a) hoặc từng gen (b) hoặc một số gen (c) trên phân tử ARNm-polycistronic được dịch mã theo yêu cầu của tế bào. (Vị trí +1 chỉ nucleotide đầu tiên được sao chép sang phân tử ARNm; P- promoter)

2.2. KIỂM SOÁT BẮT ĐẦU TỔNG HỢP PHÂN TỬ ARNm PROKARYOT

Quá trình tổng hợp ARNm trong tế bào prokaryot thường được kiểm soát theo hai cơ chế khác nhau: cơ chế tiêu cực và cơ chế tích cực. Ở cơ chế thứ nhất, hoạt động của gen bị kìm hãm khi có mặt của protein ức chế - **repressor**. Chúng tương tác với vị trí đặc hiệu nằm gần promoter gọi là **operator**. Khi protein ức chế bám vào operator, số phân tử ARNm được tổng hợp bị giảm, hay nói cách khác quá trình phiên mã bị kìm hãm. Ở cơ chế thứ hai, hoạt động của gen chỉ bắt đầu khi có mặt của loại protein hoạt hoá. Các protein này tương tác với vài vị trí đặc biệt trên vùng ADN điều khiển- các vị trí **activator**. Lúc đó ARN polymerase mới có khả năng khởi động quá trình phiên mã. Chúng ta hãy xét thí dụ điển hình cho từng cơ chế này thông qua hoạt động của operon *lac* (xảy ra theo cơ chế tiêu cực-negative control) và operon *ara* (xảy ra theo cơ chế tích cực-positive control).

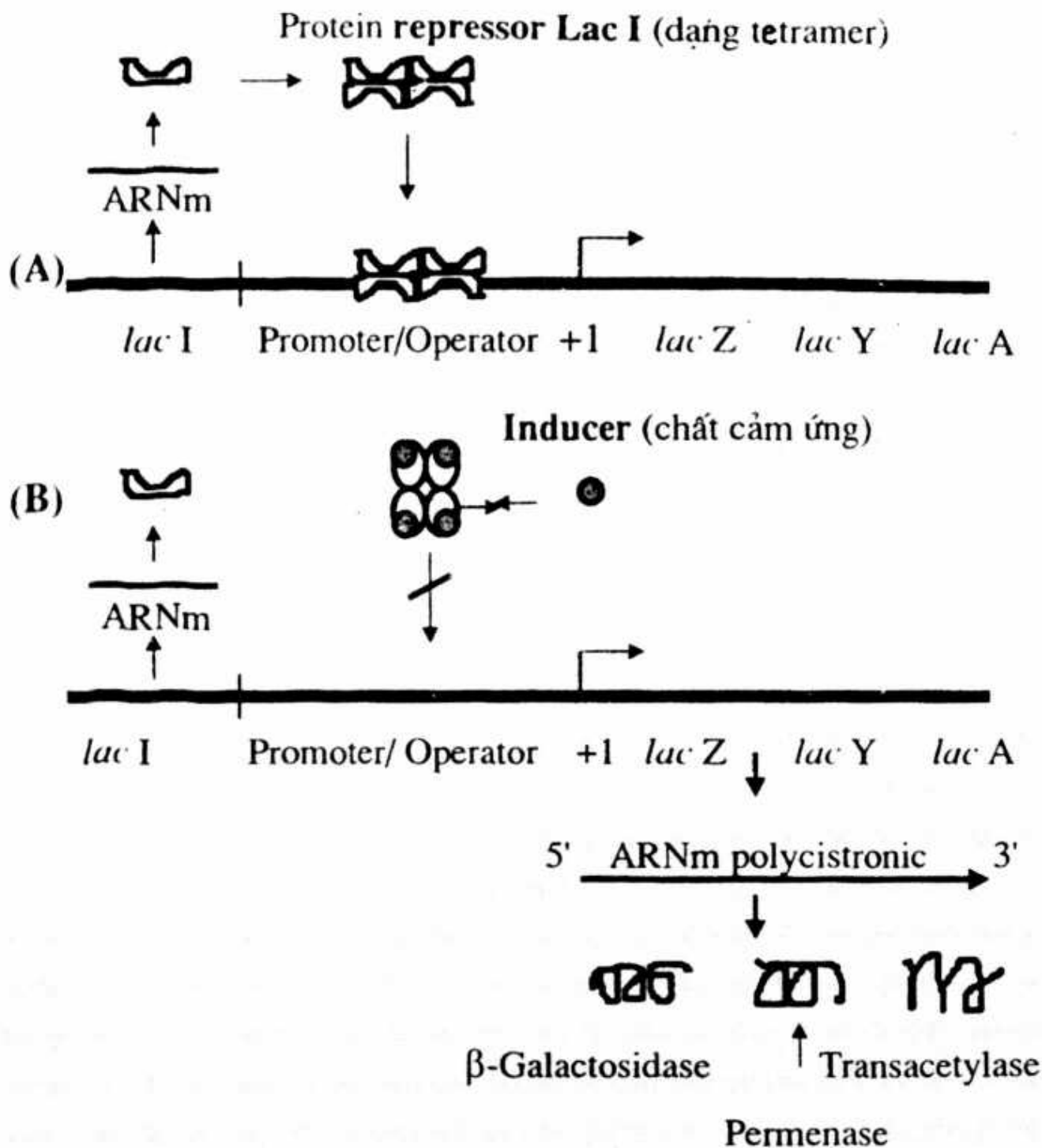
2.2.1. Kiểm soát phiên mã trên operon *lac* theo cơ chế tiêu cực

Operon *lac* gồm 3 gen cấu trúc: *lacZ*, *lacY*, *lacA*. Sản phẩm của các gen này giúp tế bào tiếp nhận và chuyển hóa β -galactosides (ví dụ như lactose). Gen *lacZ* mã cho enzym β -galactosidase có tác dụng chuyển hoá β -galactosides thành các đường (ví dụ chuyển hoá lactose thành glucose và galactose). Gen *lacY* mã cho β -galactoside permease làm nhiệm vụ vận chuyển β -galactosides vào tế bào. Gen *lacA* mã cho β -galactoside transacetylase chuyển nhóm acetyl từ acetyl-CoA vào β -galactosides.

Khi nuôi cấy vi khuẩn trên môi trường không có β -galactosides, số lượng enzym β -galactosidase nhỏ hơn 5 phân tử trong một tế bào, tức là operon *lac* hoạt động rất yếu. Tuy nhiên khi bổ sung cơ chất vào môi trường, số lượng enzym tăng rất nhanh: sau 2-3 phút có thể đạt đến 5000 phân tử/ tế bào. Quá trình tăng phản ứng tổng hợp enzym nhờ sự có mặt của cơ chất hoặc các chất tương tự được gọi là quá trình kích thích (induction).

Phân tử *lac*- ARNm rất không bền, thời gian bán hủy (half-life) chỉ khoảng 3 phút. Do đó, khi cơ chất không còn trong môi trường, phản ứng tổng hợp ARNm bị ngừng và các phân tử ARNm bị phân hủy rất nhanh. Số lượng enzym giảm về mức tối thiểu.

Quá trình phiên mã trên *lac* operon được kiểm soát bởi protein ức chế LacI - sản phẩm của gen *lacI* nằm cạnh operon *lac*. Như vậy gen *lacI* là gen điều khiển. Khi gen này bị đột biến, protein LacI không được tạo ra khiến cho operon *lac* hoạt động liên tục. Thực nghiệm nhận thấy hoạt động của operon *lac* cũng tăng rất mạnh khi trong môi trường xuất hiện cơ chất β -galactosides hoặc các chất có cấu trúc hoá học tương tự. Đặc biệt một số phân tử kích thước nhỏ có khả năng kích thích operon. Các chất có đặc tính này cũng như cơ chất của enzym được gọi là chất cảm ứng (inducer). Ví dụ như isopropylthiogalactoside (IPTG) không phải là cơ chất của β -galactosidase nhưng nó kích thích rất có hiệu quả hoạt động của operon *lac*.



Hình 2.3: Hoạt động của operon lac theo cơ chế tiêu cực

(A): Khi không có chất kích thích, protein LacI bám vào operator của operon lac ngăn cản ARN polymerase bắt đầu tổng hợp ARNm. (B): Khi chất cảm ứng xuất hiện, chúng tương tác với LacI làm repressor thay đổi cấu trúc không bám được vào operator. Phản ứng tổng hợp ARNm được khởi động

Bằng các thí nghiệm khác nhau, đặc biệt là, các thí nghiệm bổ sung (complementation test) cơ chế hoạt động của operon lac

đã được phát hiện như mô tả trên Hình 2.3. Khi không có chất cảm ứng trong môi trường, quá trình phiên mã trên operon *lac* bị kìm hãm do protein repressor LacI (ở dạng hoạt động) bám vào operator của operon. Tuy nhiên khi có mặt cơ chất, chất ức chế LacI tương tác với cơ chất nên bị thay đổi cấu trúc không gian, chuyển từ dạng hoạt động sang dạng không hoạt động và không còn khả năng tương tác với operator. Lúc này phản ứng tổng hợp phân tử ARNm polycistronic trên operon *lac* được bắt đầu.

Điều đáng lưu ý ngay khi không có chất cảm ứng, operon *lac* vẫn hoạt động ở mức độ rất yếu (0,1% so với khi có chất cảm ứng). Điều đó đảm bảo sự vận chuyển cơ chất vào tế bào xảy ra rất nhanh ngay khi cơ chất vừa có mặt ở môi trường bên ngoài tế bào. Tuy nhiên, chất cảm ứng có thể xâm nhập vào tế bào bằng nhiều con đường khác nhau. Hơn nữa, operon *lac* không chỉ chịu sự kiểm soát của cơ chế tiêu cực mà còn một số cơ chế khác nữa.

Protein repressor LacI chỉ có hoạt tính khi tồn tại dưới dạng dimer. Vì một lý do nào đó mà hai tiểu phần monomer không liên kết tạo dimer thì hoạt động của operon *lac* nằm ngoài sự kiểm soát của cơ chế tiêu cực. Khi ở dạng dimer protein LacI có hai vị trí hoạt động khác nhau, một tương tác với ADN, vị trí kia tương tác với protein cảm ứng (inducer). Khi gen mã cho LacI bị đột biến, sản phẩm của nó có cấu trúc thay đổi sao cho không thể tương tác với ADN (nhưng có thể vẫn tương tác được với inducer) thì operon *lac* hoạt động một cách liên tục. Đó là do LacI không thể liên kết với ADN ngay khi cơ chất hay các chất cảm ứng vắng mặt trong môi trường. Ngược lại, nếu repressor chỉ tương tác với ADN mà không còn khả

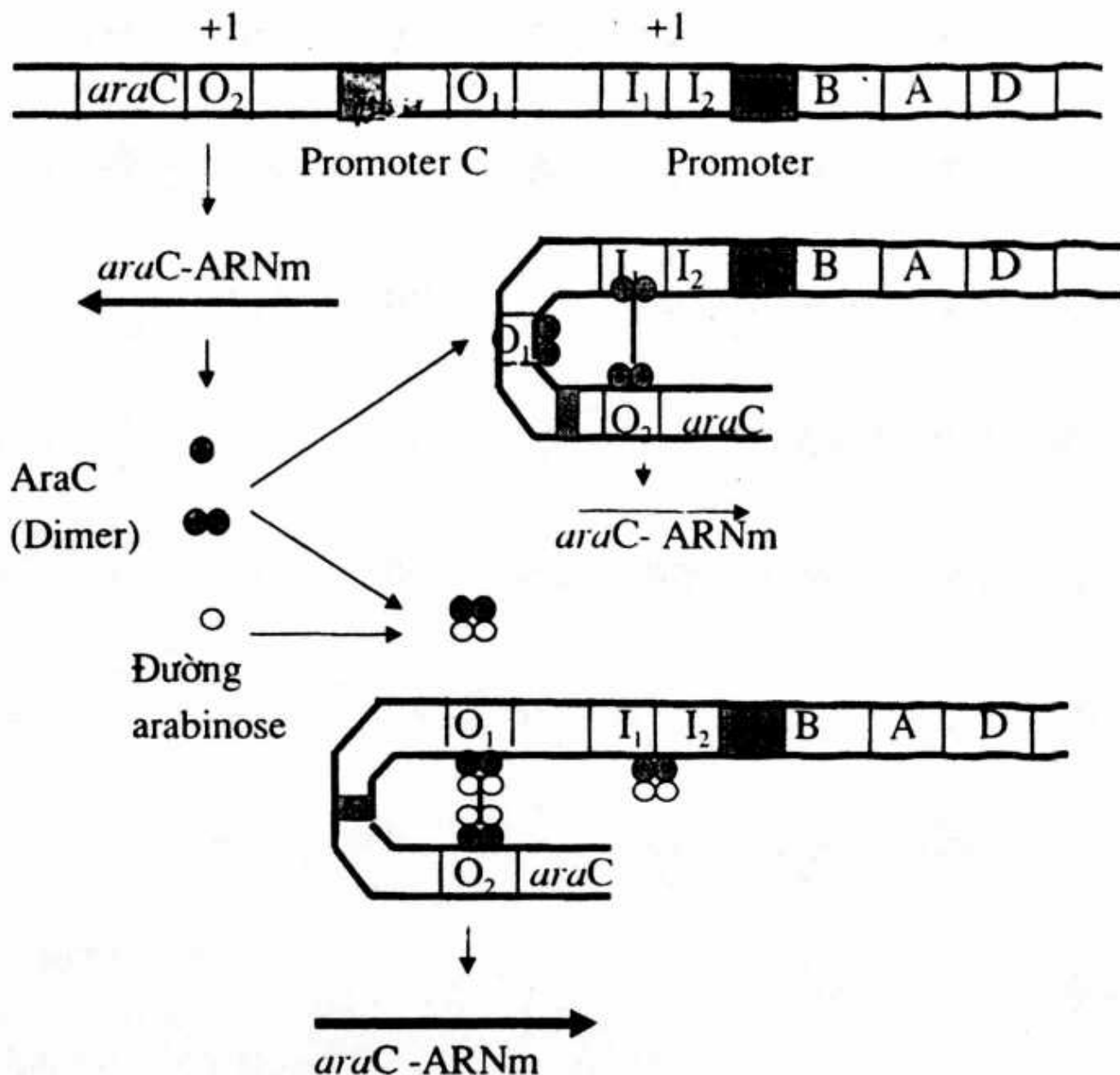
năng tương tác với inducer thì operon *lac* không hoạt động ngay khi môi trường có cơ chất.

2.2.2. Kiểm tra phiên mã trên operon *ara* theo cơ chế tích cực

Khi tế bào *E.coli* mọc trên môi trường chứa arabinose, chúng tổng hợp ba enzym cần thiết để biến đổi đường này thành xylulose-5-phosphate. Chất này được sử dụng trong con đường chuyển hoá glucose để bổ sung năng lượng cho tế bào. Cả ba enzym được mã bởi các gen nằm trong operon *ara*: gen A mã cho isomerase, gen B mã cho kinase và gen D mã cho epimerase. Khi môi trường có arabinose, hoạt động của operon *ara* tăng mạnh tương tự như lactose kích thích hoạt động của operon *lac*. Tuy nhiên cơ chế hoạt động của hai operon này hoàn toàn khác nhau.

Hoạt động của operon *ara* được điều khiển bởi protein hoạt hóa AraC do- sản phẩm của gen *araC* (nằm gần *ara* operon). Khi gen *araC* bị đột biến, operon *ara* không hoạt động ngay khi môi trường có mặt arabinose, mặc dù cả ba gen ABD đều nguyên vẹn. Điều này trái ngược với gen *lacI* khi bị đột biến thì operon *lac* hoạt động liên tục. Như vậy cơ chế kiểm soát hoạt động của operon *ara* đòi hỏi sự có mặt đồng thời của cả cơ chất arabinose và protein AraC. Protein AraC có vai trò như yếu tố kích hoạt sự phiên mã trên operon *araC*. Một gen (hoặc operon) được xem là hoạt động theo cơ chế tích cực khi chúng được khởi động bởi các protein hoạt hóa như AraC.

Các thí nghiệm đột biến di truyền đã xác định vị trí bám của AraC vào ADN trên operon *ara*. Protein AraC có thể bám đồng thời vào nhiều các vị trí khác nhau gây tác dụng điều khiển khác nhau (Hình 2.4).



Hình 2.4: Hoạt động của operon ara được điều khiển theo cơ chế tích cực bởi protein hoạt hoá AraC.

(a): Khi môi trường không có đường arabinose, AraC bám vào các vị trí O_1 , O_2 và I_1 , tạo liên kết giữa O_2 và I_1 , gây uốn cong sợi ADN. Operon ara bị kìm hãm, lượng ARNm tạo ra rất ít. (b) : Khi có mặt arabinose, phân tử đường này sẽ tương tác với protein hoạt hóa AraC. Phức chất bám vào 4 vị trí O_1 , O_2 , I_1 và I_2 . Tương tác của phức chất giữa O_1 và O_2 kích thích hoạt động của gen araC. Ngoài ra, tương tác của phức chất giữa I_1 và I_2 khiến hoạt động của operon ara tăng lên rất nhiều.

Khi môi trường không có arabinose, AraC bám vào 3 vị trí O_1 , O_2 và I_1 , nằm phía trước sát với promoter của operon ara.

Hơn nữa, các phân tử AraC bám vào hai vị trí O_2 , O_1 có thể tương tác với nhau (tương tác protein-protein) gây uốn cong sợi ADN. Lúc này operon *ara* không hoạt động.

Khi môi trường có arabinose, protein hoạt hoá AraC tương tác với phân tử đường, do đó bị thay đổi cấu hình. Lúc này phức chất AraC và đường bám vào 4 vị trí O_1 , O_2 , I_1 và I_2 . Tiếp đó, AraC ở hai vị trí O_1 và O_2 tương tác với nhau làm cho hoạt động của gen *araC* tăng lên. Mặt khác, phức chất đó khi gắn vào I_1 và I_2 gây hoạt hóa operon *ara*. Quá trình tổng hợp ARNm trên operon này được khởi động.

Điều đáng lưu ý là protein AraC tự điều khiển hoạt động của chính gen tạo ra nó (gen *araC*) bằng việc tương tác với các vị trí O_1 và O_2 . Khác với protein ức chế LacI chỉ có tác dụng điều khiển tiêu cực đối với operon *lac*, protein hoạt hoá AraC kiểm soát hoạt động của cả operon *ara* và gen *araC* theo hai cơ chế tùy thuộc vào vị trí liên kết của nó với ADN. Trong khi protein AraC chỉ có khả năng kiểm soát theo cơ chế tích cực với gen *araC* thì vai trò của chúng đối với operon *ara* thay đổi tùy thuộc vào vị trí tương tác với ADN và tương tác giữa chúng với nhau.

Hai cơ chế tiêu cực và tích cực xảy ra hoàn toàn đối lập nhau. Các gen chịu sự điều khiển theo cơ chế tiêu cực bị kìm hãm không hoạt động khi có mặt protein ức chế. Ngược lại, các gen chịu kiểm soát theo cơ chế tích cực chỉ hoạt động khi có mặt protein hoạt hoá. Các protein ức chế ngoài việc bám vào ADN để ngăn cản ARN polymerase khởi động việc phiên mã, chúng còn có thể bám vào ARNm để ngăn cản ribosome thực hiện việc tổng hợp protein. Có thể có nhiều repressor cùng phối hợp hoạt động, chúng được gọi chung là các co-repressor. Hơn nữa, nhiều cơ chế có thể phối hợp cùng nhau để kiểm soát hoạt động của gen hay

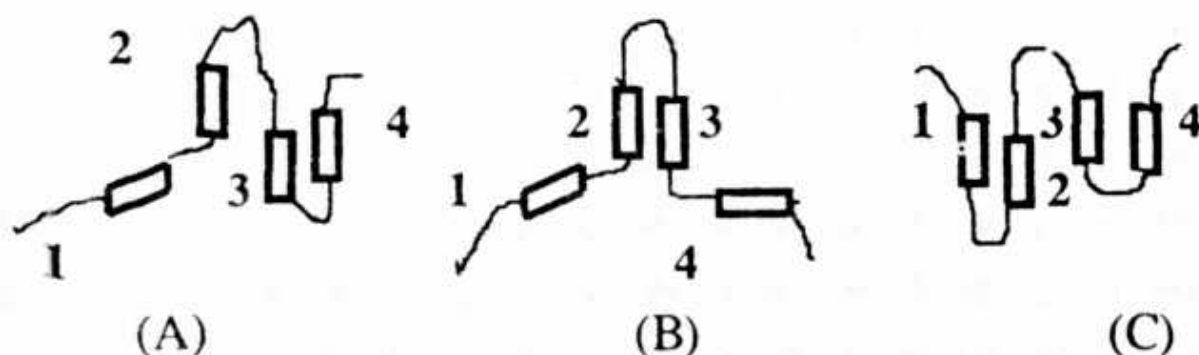
operon. Do đó, sự cạnh tranh tương tác giữa các protein ức chế và hoạt hoá với trình tự đích trên vùng ADN điều khiển sẽ điều biến một cách linh hoạt biểu hiện của gen. Nhờ vậy tế bào và cơ thể trả lời nhanh nhạy và chính xác đối với những tín hiệu trong môi trường.

Căn cứ vào phản ứng của operon với các phân tử protein điều khiển để biết được hoạt động của operon đó được kích thích hoặc bị ức chế. Ngoài ra tương tác giữa các loại protein với nhau có thể gây đóng hoặc mở gen. Ví dụ như một protein cảm ứng (inducer) gây mất hoạt tính của protein ức chế (repressor) hoặc phục hồi hoạt tính của protein hoạt hoá (activator) đều dẫn đến khởi động operon. Ngược lại, operon bị ức chế khi các co-repressor hoạt hóa protein ức chế (repressor) hoặc gây mất hoạt tính protein hoạt hoá (activator)....

2.2.3. Hoạt động của operon tryptophan

Operon *tryptophan* gồm 5 gen mã cho các enzym liên quan đến việc tổng hợp tryptophan. Khi môi trường thiếu acid amin này, hoạt động của operon tăng gấp 10 lần. Như vậy, sự có mặt của tryptophan trong môi trường kìm hãm operon *tryptophan*. Tuy nhiên, tryptophan không phải là tác nhân gây ức chế biểu hiện của operon. Chúng giữ vai trò của chất đồng ức chế (co-repressor) khi tương tác với repressor để ngăn cản hoạt động của operon *tryptophan*. Như vậy, operon này được điều khiển theo cơ chế tiêu cực do chính sản phẩm cuối cùng của operon. Cơ chế kiểm soát đặc biệt này được gọi là **ức chế phản hồi (feedback inhibition)**. Trong một số trường hợp khác, gen hay operon được hoạt hoá bằng chính sản phẩm do chúng mã cho. Lúc đó chúng được điều khiển bởi cơ chế hoạt hoá phản hồi.

Phân tích trình tự nucleotide trên phân tử ARNm polycistronic của operon *tryptophan* làm sáng tỏ cơ chế ức chế phản hồi ở mức độ phân tử. Kết quả cho thấy đoạn nucleotide gồm 160 base nằm trước điểm dịch mã đầu tiên (mã cho Methyonine) đóng vai trò điều khiển quá trình phiên mã tạo ARNm trên operon *tryptophan*. Đoạn này được gọi là "attenuator" có chứa 4 chuỗi nucleotide ngắn có trình tự tương đồng ngược chiều. Hai chuỗi ngược chiều có thể liên kết với nhau tạo cấu trúc dạng vòng. Tùy thuộc vào sự có mặt của tryptophan trong môi trường mà xảy ra những kiểu liên kết tạo cặp giữa hai chuỗi tạo ra các cấu trúc dạng vòng khác nhau. Nhờ đó, sự tổng hợp các enzym cần thiết cho phản ứng tạo ra tryptophan được kiểm soát phù hợp với sự đòi hỏi của tế bào đối với loại acid amin này (Hình 2.5).



Hình 2.5: Cấu trúc bậc hai của phân tử tryptophan- ARNm đóng vai trò điều khiển operon tryptophan trong phản ứng tạo ra các enzym tham gia tổng hợp tryptophan. Hoạt động của operon này bị kiểm tra bởi cơ chế tiêu cực và bởi đoạn "attenuator".

(A)- Khi môi trường thừa Tryptophan. (B)- Môi trường thiếu Tryptophan.

(C)- Không có sự tổng hợp enzym.

Chuỗi nucleotide 1 trên đoạn "attenuator" có chứa các mã di truyền của tryptophan (Trp). Khi môi trường có nhiều Trp, các ARNt vận chuyển Trp cho ribosome dịch mã trên chuỗi 1 này dễ dàng tìm thấy Trp trong tế bào. Do đó ribosome vượt qua chuỗi 1 nhanh và bao trùm lên chuỗi 2. Lúc này chỉ có chuỗi 3 và 4 liên kết với nhau tạo cấu trúc vòng gây tín hiệu ngừng tổng

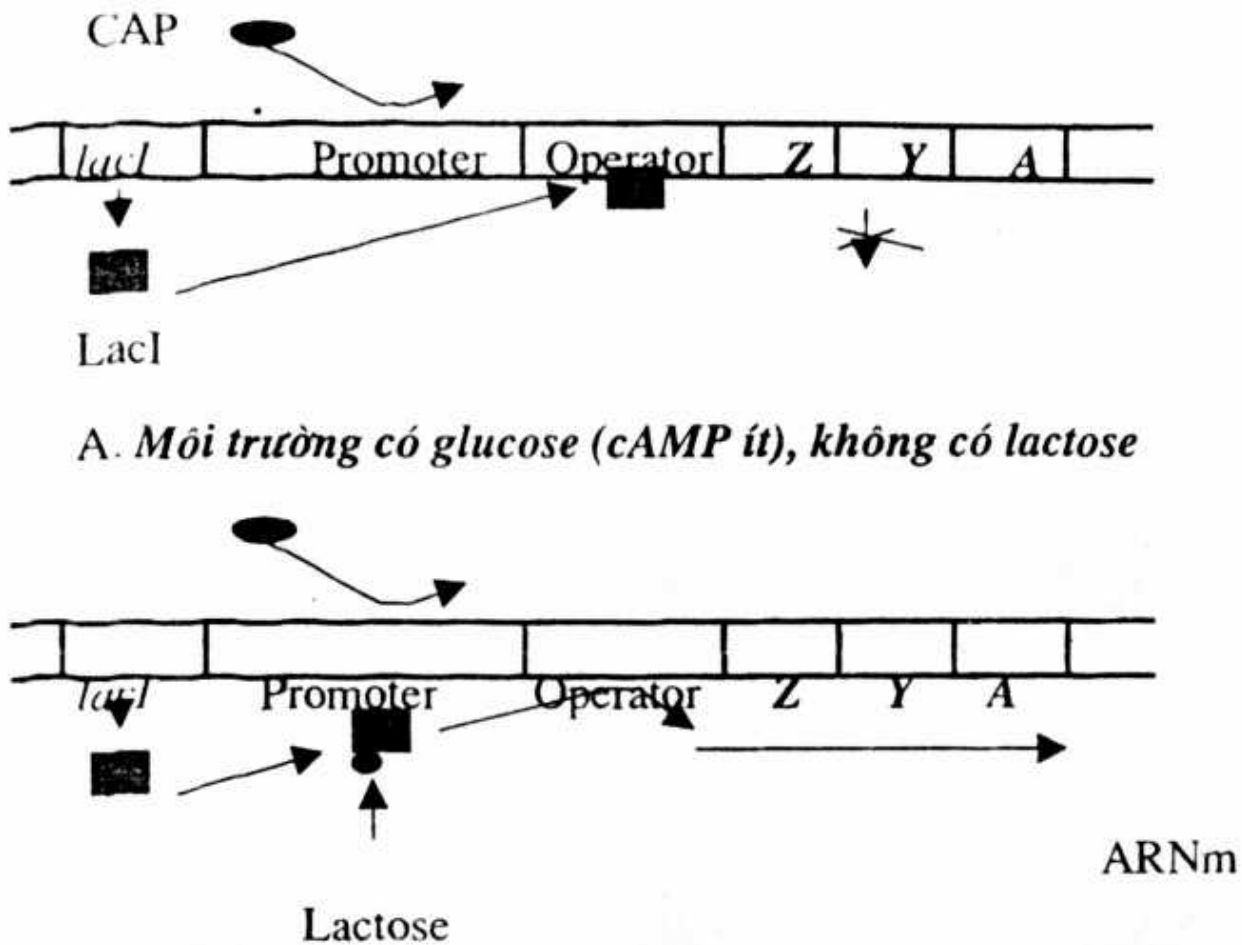
hợp phân tử ARNm (Lưu ý rằng ở vi khuẩn quá trình phiên mã bởi ARN polymerase xảy ra đồng thời với quá trình dịch mã bởi ribosome. Cấu trúc dạng vòng trên phân tử ARNm là một trong những yêu cầu cần có để ARN polymerase nhận biết vị trí dừng quá trình phiên mã). Ngược lại, khi môi trường thiếu Trp, các ARNt khó tìm thấy Trp, do đó ribosome phải chờ đợi lâu ở chuỗi 1. Lúc này chuỗi 2 tự do và liên kết với chuỗi 3. Tuy nhiên cấu trúc dạng vòng này nằm cách xa vị trí của ARN polymerase nên không ảnh hưởng đến việc tổng hợp ARNm. Nhờ đó phân tử ARNm của operon *tryptophan* được tạo ra một cách trọn vẹn.

Một số operon mã cho các enzym cần thiết cho phản ứng sinh tổng hợp một số chất trong tế bào (operon *phenylalanine* hoặc operon *histidine*) cũng được kiểm tra theo cơ chế giống như đối với operon *tryptophan*.

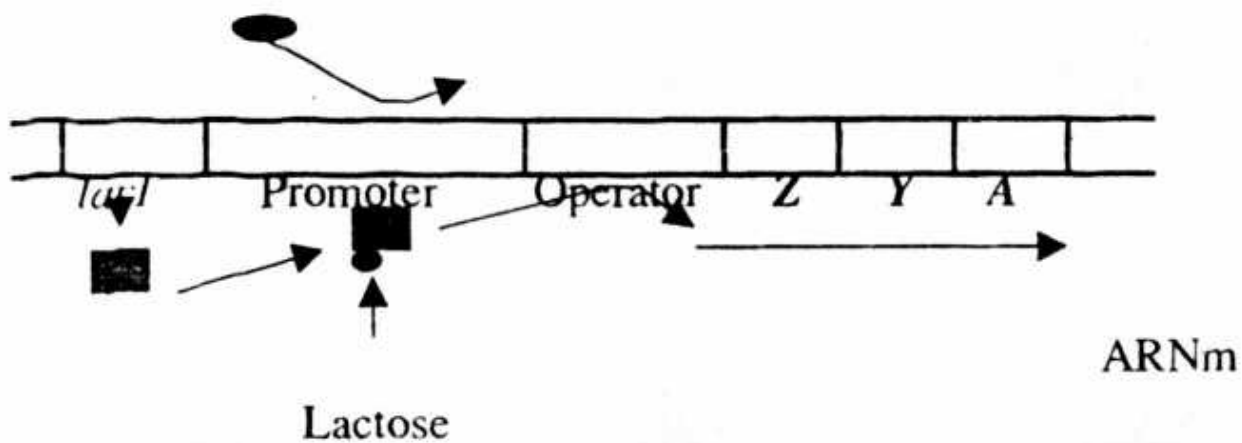
2.2.4. Ức chế quá trình dị hoá liên quan đến điều khiển tích cực

Glucose là một trong các đường cần thiết cho tế bào vi khuẩn phát triển. Nó được tế bào sử dụng trực tiếp, nhanh chóng và đơn giản nhất không yêu cầu tổng hợp mới bất kỳ enzym nào. Khi glucose cùng với một số đường khác như lactose, arabinose cùng có mặt trong môi trường, tế bào chỉ sử dụng glucose và kìm hãm hoạt động của các operon liên quan đến chuyển hóa các loại đường khác. Hiện tượng này được gọi là ức chế quá trình dị hoá.

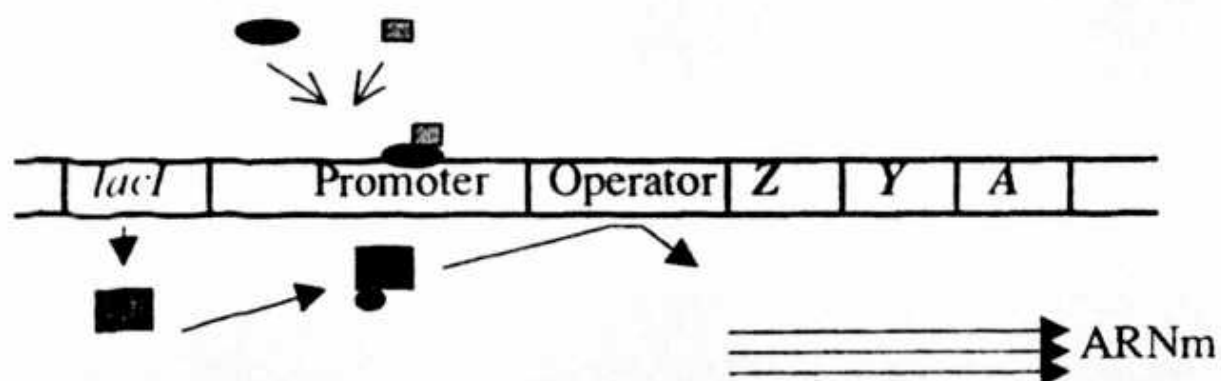
Các nhà hoá học phát hiện rằng khi vi khuẩn thiếu glucose, chúng sẽ tăng cường tổng hợp nucleotide cyclic adenosine-3,5'-monophosphate (cAMP). Số lượng cAMP tăng lên giống như tín hiệu báo động về nguy cơ chuyển hoá. Hơn nữa, khi bổ sung cAMP vào môi trường có chứa glucose và các đường khác thì hoạt động của các operon liên quan đến chuyển hoá các đường này không bị ức chế mà ngược lại được hoạt hoá. Như vậy cAMP đóng vai trò quan trọng bật mở hoạt động của một số gen ngay khi chúng đang ở trạng thái bị kìm hãm.



A. Môi trường có glucose (cAMP ít), không có lactose



B. Môi trường có glucose (cAMP ít), có lactose (inducer)



C. Môi trường không có glucose (cAMP nhiều), có lactose

Hình 2.6: Kiểm tra hoạt động của operon lac bởi hai protein LacI và CAP.

- (A) Nếu môi trường không có lactose, ARNm-lac không được tổng hợp ngay khi có hoặc không có glucose vì LacI bám vào operator. (B) Khi có lactose, LacI không bám được vào operator nhưng do có glucose nên cAMP không được tổng hợp dẫn đến phức chất cAMP-CAP không được tạo thành. Do đó hoạt động của operon yếu. (C) Khi môi trường không có glucose nhưng có lactose, operon được hoạt hoá theo cả hai cơ chế, do đó lượng ARNm-lac được tổng hợp đạt cực đại.

Nghiên cứu các tế bào vi khuẩn đột biến không có khả năng sử dụng bất cứ loại đường nào ngoài glucose, các nhà khoa học phát hiện rằng chúng không tổng hợp được cAMP trong điều kiện môi trường thiếu glucose. Nói cách khác, các tế bào này chịu đột biến ở gen mã cho enzym tổng hợp cAMP (adenylate cyclase). Ngoài ra một số tế bào khác có khả năng tổng hợp cAMP nhưng vẫn không tổng hợp được enzym cần thiết để chuyển hoá các loại đường khác glucose. Như vậy phải tồn tại một protein tương tác với cAMP để hoạt hoá các operon.

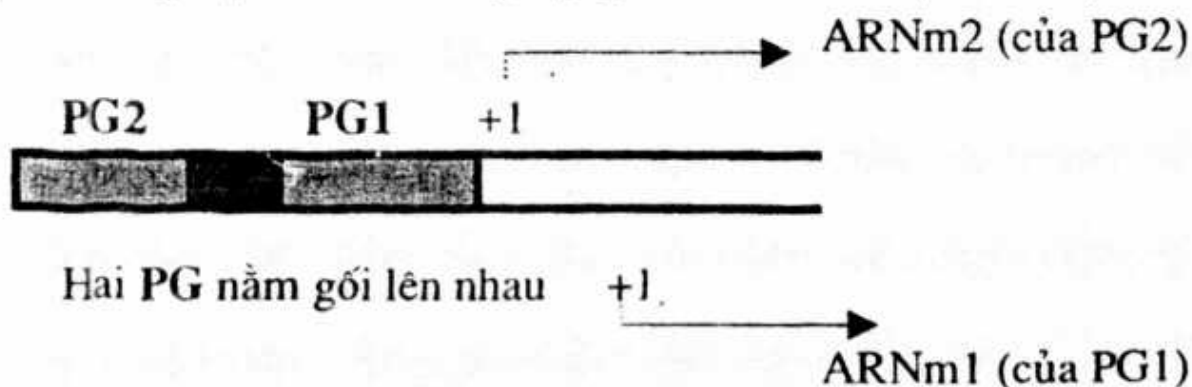
Thực nghiệm đã phân lập được protein này và gọi là protein hoạt hoá quá trình dị hoá (catabolite activator protein-CAP). Trong thí nghiệm tổng hợp ARNm *in vitro* đối với operon *lac*, số lượng ARNm tăng rất nhiều khi thêm cAMP và CAP vào phản ứng. Nếu chỉ thêm một trong hai chất đó, số lượng ARNm chỉ chiếm 5% so với khi có mặt cả hai. Như vậy ngoài việc bị điều khiển theo cơ chế tiêu cực bởi protein LacI, operon *lac* còn chịu sự kiểm soát theo cơ chế tích cực nhờ phức chất cAMP-CAP. Phức này tương tác với promoter gây uốn cong sợi ADN, bộc lộ vị trí bám cho ARN polymerase (Hình 2.6).

Không phải chỉ tham gia kiểm soát hoạt động của operon *lac*, protein CAP còn đóng vai trò protein hoạt hoá trong quá trình tổng hợp ARNm của nhiều operon khác, ví dụ như với operon *ara*, *galactose*... Như vậy operon *ara* cần hai protein hoạt hoá khác nhau (AraC và CAP) để đạt được mức độ hoạt động cực đại. Với thí dụ về operon *ara*, chúng ta đã thấy tính phức tạp đan chéo nhau của các cơ chế kiểm soát hoạt động của gen (operon).

2.2.5. Điều khiển gen thông qua các promoter khác nhau

Operon *galactose* (*gal*) gồm các gen mã cho ba enzym cần thiết để chuyển hoá galactose khi tế bào phát triển trên môi trường có chứa loại đường này. Khi operon *gal* hoạt động ở mức độ tối đa, số lượng ARNm tạo ra chỉ cao hơn 10-15 lần so với khi không có galactose. Điều đó khác hẳn operon *lac* khi được kích thích, lượng enzym β -galactosidase có thể đạt gấp 1000 lần so với khi operon không hoạt động. Mặt khác, khi môi trường có glucose, biểu hiện của operon *lac* giảm khoảng 100 lần bởi cơ chế ức chế quá trình dị hoá trong khi operon *gal* chỉ giảm hoạt động khoảng 10-15 lần. Điều này chứng tỏ các gen trên operon *gal* hoạt động độc lập và không ngừng ở một mức độ nhất định bất chấp môi trường chứa loại đường nào. Một trong những lý do buộc operon *gal* phải hoạt động liên tục do nó mang gen mã cho enzym *epimerase* cần thiết cho quá trình tổng hợp màng tế bào. Vì vậy nồng độ enzym này phải được đảm bảo theo yêu cầu của tế bào mà không phụ thuộc vào sự có mặt của glucose trong môi trường.

Một loạt các thí nghiệm được tiến hành nhằm phân tích trình tự nucleotide ở vùng ADN điều khiển của operon *gal*, gây đột biến điểm và đột biến mất đoạn, thực hiện các thử nghiệm bổ sung v.v... cho thấy operon *gal* có hai promoter nằm gối lên nhau và có hai vị trí riêng biệt bắt đầu tổng hợp phân tử ARNm (Hình 2.7).



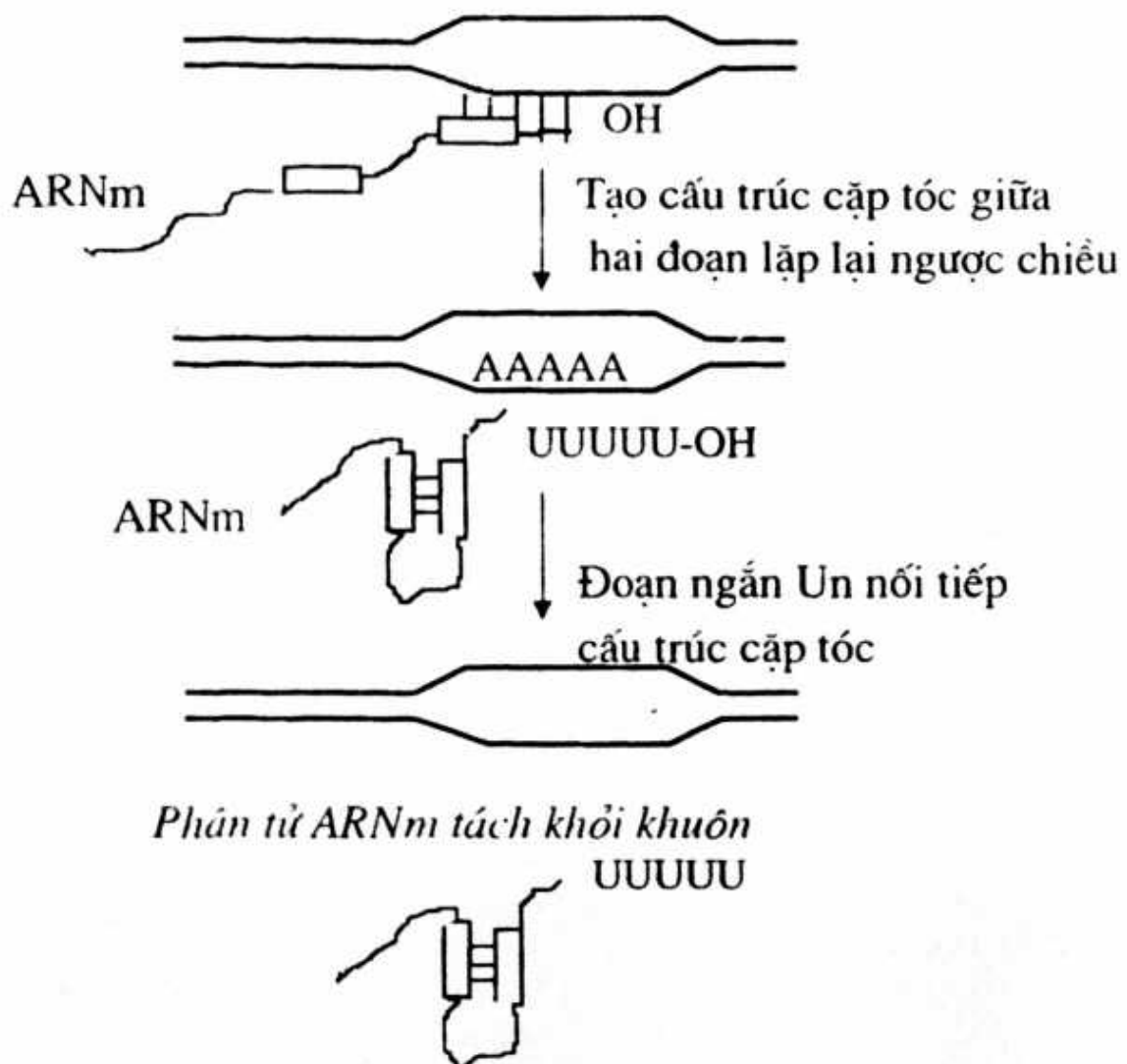
Hình 2.7: Hai promoter PG1 và PG2 hoạt động riêng biệt trong quá trình điều khiển operon *gal*. PG₂ hoạt động liên tục dù có hay không có glucose. PG₁ chỉ hoạt động khi có galactose, cAMP và CAP ở nồng độ cao.

Promoter PG_2 cho phép tổng hợp ARNm ngay trong điều kiện môi trường chứa glucose nhưng không chứa galactose. Như vậy operon *gal* hoạt động liên tục tạo ARNm từ promoter PG_2 . Tuy nhiên khi nồng độ cAMP và CAP cao, ARN polymerase bám vào PG_2 khó khăn hơn. Ngược lại, trong điều kiện đó enzym này tương tác dễ dàng với promoter PG_1 . Vì vậy, PG_1 chỉ tổng hợp ARNm khi môi trường không có glucose. Tính nhạy cảm của ARN polymerase đối với PG_1 cao hơn với PG_2 nên lượng ARNm được tổng hợp từ promoter PG_1 (khi không có glucose) lớn hơn nhiều so với khi chỉ có PG_2 hoạt động.

2.3. KIỂM SOÁT DỪNG PHẢN ỨNG TỔNG HỢP ARNm

Các cơ chế điều khiển tiêu cực, tích cực hoặc điều khiển bằng nhiều promoter khác nhau đối với một gen (hay một operon) đều nhằm kiểm tra quá trình bắt đầu tổng hợp ARNm. Tuy nhiên kết thúc phản ứng này cũng được kiểm tra chặt chẽ nhằm đảm bảo dừng quá trình tổng hợp tại vị trí chính xác. Một số cơ chế điều khiển việc dừng tổng hợp phân tử ARNm như sau:

1) Phản ứng tổng hợp ARNm được kết thúc do hình thành cấu trúc đặc biệt "hairpin". Trong thí nghiệm tổng hợp ARNm *invitro*, enzym ARN polymerase luôn dừng lại tại những đoạn ARN có chuỗi nucleotide lặp lại ngược chiều (inverted repeat) giàu G và C (thời gian dừng khoảng 60 giây). Các chuỗi lặp lại ngược chiều liên kết với nhau theo nguyên tắc bổ sung tạo cấu trúc "hairpin" gọi là cặp tóc (Hình 2.8). Cấu trúc này làm ARN polymerase dừng hoặc chuyển động chậm. Tuy nhiên để làm enzym tách ra khỏi khuôn ADN, tiếp theo cấu trúc cặp tóc cần có các nucleotide U nối nhau liên tiếp. Liên kết A-U giữa ARN và ADN là các liên kết yếu nên cần ít năng lượng để bẻ gãy chúng.

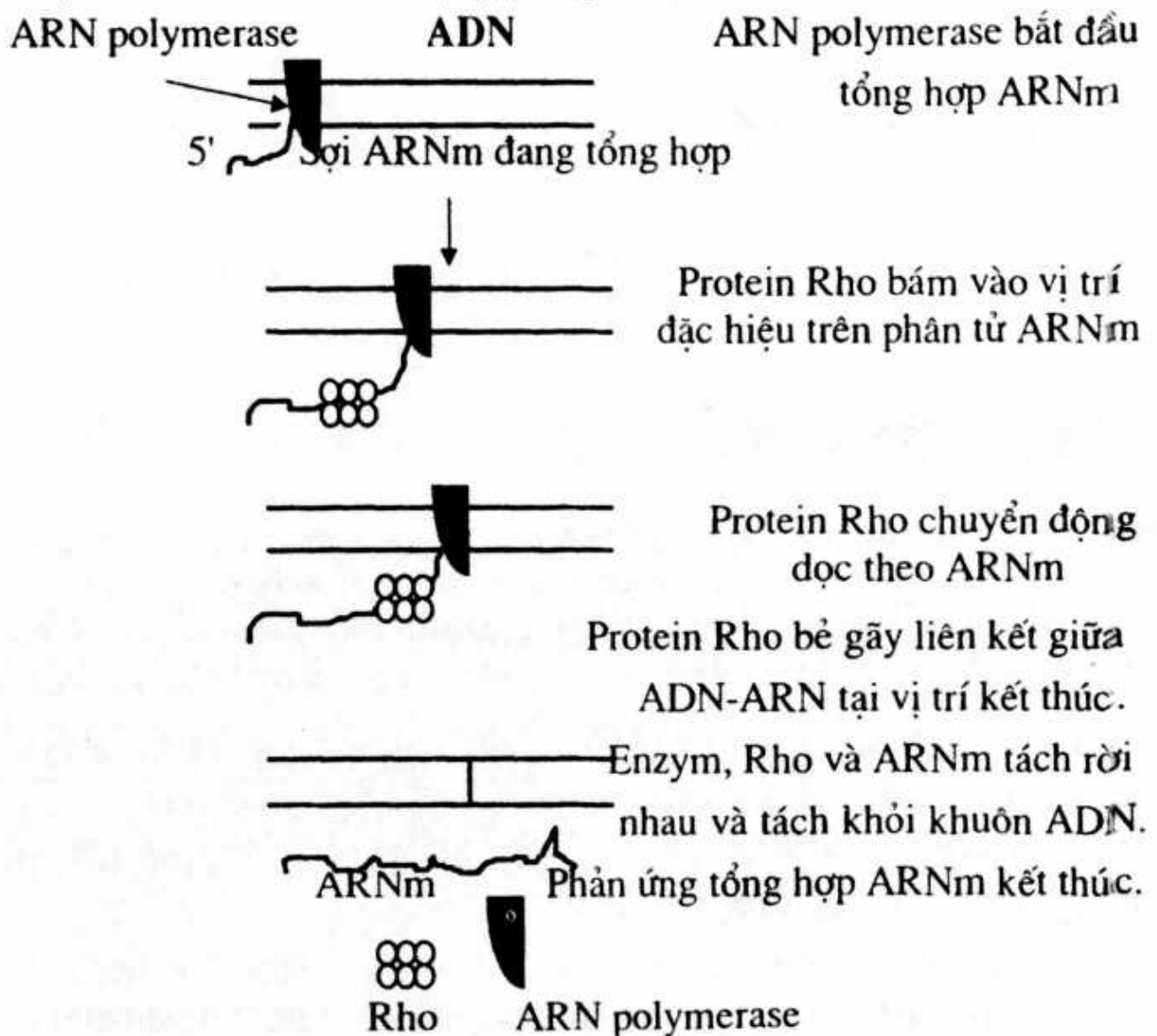


Hình 2.8: Kết thúc phản ứng tổng hợp ARNm nhờ cấu trúc đặc biệt "hairpin". Cấu trúc này tạo ra do liên kết giữa các chuỗi nucleotide lặp lại ngược chiều "inverted repeat" và được nối tiếp bởi chuỗi base U. Enzym ARN polymerase bị dừng lại và rời khỏi khuôn tại vị trí đặc hiệu này. Phân tử ARNm được tổng hợp hoàn chỉnh.

Thực nghiệm cho thấy khi đột biến xảy ra ở đoạn nucleotide U tiếp sau cấu trúc cặp tóc thì ARN polymerase vẫn tiếp tục tổng hợp ARNm. Do đó vị trí kết thúc quá trình phiên mã phụ thuộc hai yếu tố:

- Cấu trúc cặp tóc của các đoạn "inverted repeat" giàu G, C trên phân tử ARNm làm enzym ARN polymerase dừng chuyển động.
- Đoạn ngắn nucleotide gồm các base U tiếp ngay sau cấu trúc cặp tóc khiến enzym tách rời khuôn và phân tử ARNm được tổng hợp hoàn chỉnh.

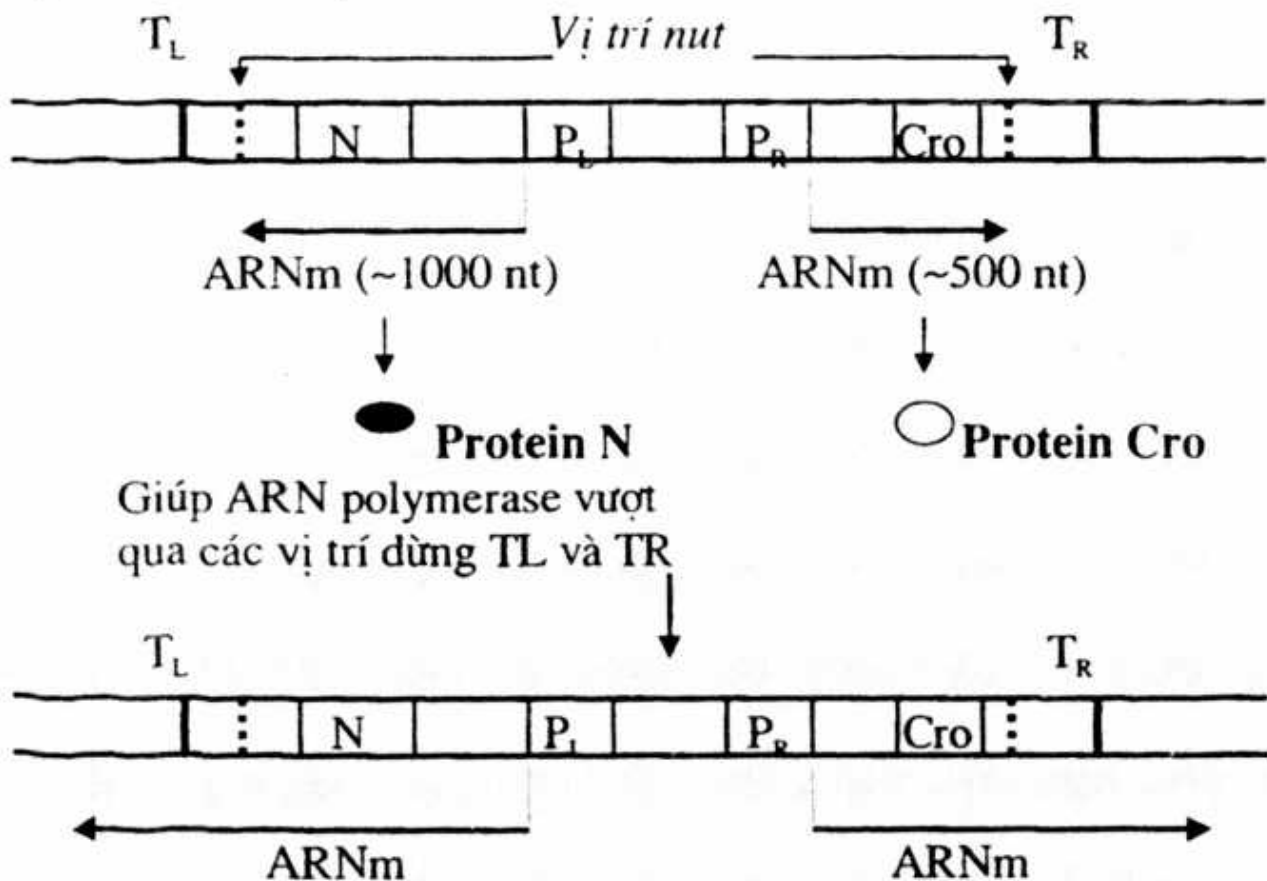
2) Phản ứng tổng hợp ARNm được kết thúc bởi protein đặc hiệu **Rho** (Hình 2.9). Trong phản ứng tổng hợp ARNm *in vitro* nếu không có dịch tách chiết từ tế bào thì thường nhận được các phân tử ARNm có kích thước khác nhau do enzym ARN polymerase không dừng đúng chỗ. Khi dịch tách chiết được thêm vào môi trường thí nghiệm, enzym kết thúc chính xác phản ứng, cho các phân tử ARNm có chiều dài duy nhất. Protein chịu trách nhiệm dừng phản ứng tại vị trí đặc hiệu có hoạt tính ATPase và được gọi là protein **Rho**.



Hình 2.9: Cơ chế hoạt động của protein Rho làm ngừng phản ứng tổng hợp ARNm tại vị trí chính xác được nhận biết bởi protein này

Protein ngăn cản dừng tổng hợp ARNm

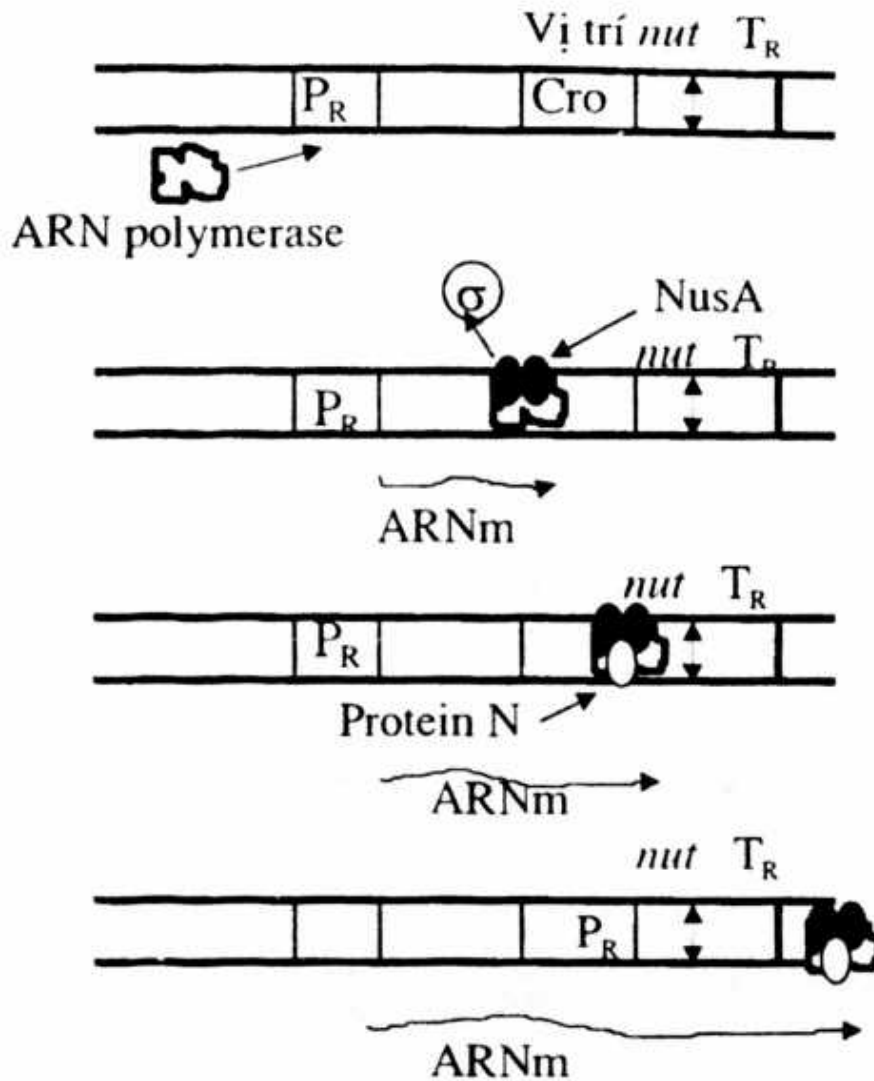
Protein **Rho** bị mất hoạt tính bởi sự có mặt của protein **N**. Đây là protein đặc hiệu ngăn cản việc dừng phản ứng tổng hợp ARNm, giúp cho ARN polymerase vượt qua các vị trí dừng. Tác dụng của protein **N** được phát hiện và nghiên cứu kỹ đối với phản ứng tổng hợp các ARNm của bacteriophage khi chúng xâm nhập vào tế bào vật chủ *E.coli*.



Hình 2.10: Tác dụng của protein **N** giúp ARN polymerase vượt qua các vị trí dừng trong phản ứng tổng hợp ARNm khi thực khuẩn thể phage xâm nhiễm tế bào chủ. Hai vị trí dừng T_L và T_R được nhận biết bởi protein **Rho**. Do tác dụng của **N**, protein **Rho** bị mất hoạt tính, nhờ đó enzym vượt qua hai vị trí này để tổng hợp các phân tử ARNm dài hơn tương ứng với các gen nằm xa hai promoter P_L và P_R .

Khi xâm nhập vào tế bào vi khuẩn *E.coli*, genome của bacteriophage ở dạng vòng. Quá trình tổng hợp ARNm được bắt đầu ở hai promoter trái và phải P_L và P_R và tiếp tục xa dần sang hai phía. Tuy nhiên trong giai đoạn đầu tiên bắt đầu xâm nhập, các phân tử ARNm được tổng hợp có kích thước rất ngắn (khoảng 1000 base tính từ P_L , khoảng 500 base tính từ P_R) do ARN polymerase dừng lại ở hai vị trí T_L và T_R . Phân tử ARNm sao chép đầu tiên từ P_L mã cho protein N có tác dụng ngăn cản việc dừng tổng hợp ở hai vị trí này. Nhờ đó các phân tử ARNm được tổng hợp dài hơn tương ứng với các gen khác của genome λ (Hình 2.10).

Cơ chế hoạt động của protein N được mô tả trên Hình 2.11. Protein N tương tác với protein NusA. Khi enzym ARN polymerase đã bắt đầu tổng hợp ARNm, phân tử σ tách ra khỏi enzym, phần còn lại sẽ liên kết với protein NusA. Khi enzym chuyển động đến vị trí đặc hiệu *nut* trên khuôn ADN, protein NusA tương tác với protein N. Vị trí này thường nằm gần vị trí nhận biết bởi protein **Rho**. Ngoài ra còn có các protein khác như NusB, NusE, NusG tương tác với protein NusA và protein N tạo phức hợp giúp enzym vượt qua các vị trí dừng trong phản ứng tổng hợp ARNm.



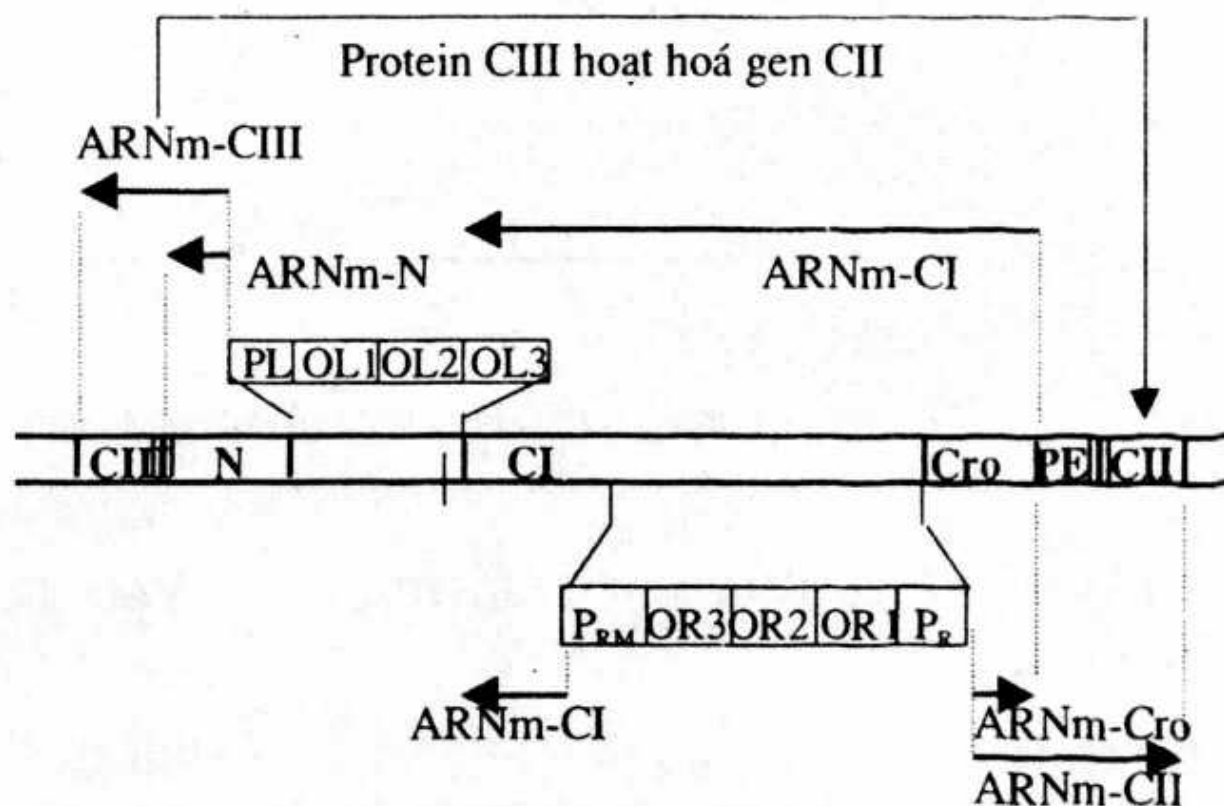
Hình 2.11: Cơ chế hoạt động của protein N tại vị trí T_R ngăn cản việc ngừng tổng hợp ARN_m trên genome của bacteriophage λ

2.4. XÂM NHẬP CỦA BACTERIOPHAGE λ VÀO TẾ BÀO *E. COLI*

Chúng ta đã biết các cơ chế kiểm soát khác nhau trong phản ứng tổng hợp phân tử ARN_m (tiêu cực, tích cực, điều khiển phản hồi "feedback", điều khiển tự động...). Bây giờ chúng ta xét đến sự phối hợp của các cơ chế này trong quá trình điều khiển hoạt động của toàn bộ hệ gen thực khuẩn thể λ khi chúng xâm nhập vào tế bào *E. coli*.

Khi bacteriophage λ nhiễm vào tế bào *E. coli*, một trong hai sự kiện khác nhau sẽ xảy ra: Thứ nhất là ADN của λ được nhân

lên tự do trong tế bào chủ, các thể phage mới được tạo thành và tế bào vi khuẩn nhanh chóng bị phá vỡ (trạng thái litic hay còn gọi là trạng thái sinh tan). Thứ hai là ADN của λ được ghép vào nhiễm sắc thể của tế bào chủ và được nhân lên cùng với ADN của vi khuẩn (trạng thái lisogenic hay là trạng thái tiềm tan). Việc quyết định tồn tại ở trạng thái nào phụ thuộc nhiều yếu tố khác nhau ở cả tế bào chủ và thể phage λ . Tuy nhiên những điều kiện thuận lợi cho trạng thái tiềm tan được quyết định bởi chính protein ức chế CI do phage λ tổng hợp. Protein này có hoạt tính ở dạng dimer. Protein CI làm nhiệm vụ điều khiển hoạt động của hai promoter chính P_L , P_R của toàn bộ hệ gen thực khuẩn thể và promoter P_{RM} của riêng gen mã cho repressor CI.



Hình 2.12: Cấu trúc các gen của thực khuẩn thể λ được sao chép ngay khi xâm nhập vào tế bào E.coli.

Các protein CI, CII, CIII, Cro và N giữ vai trò quan trọng điều khiển quá trình tổng hợp ARNm trên hệ gen thực khuẩn thể. Hai promoter P_L và P_R hoạt động liên tục ngay khi phage λ nhiễm tế bào chủ, tạo ARNm của protein Cro và N. Protein N tác dụng lên các vị trí dừng T_L và T_R cho phép tổng hợp ARNm của protein CII, CIII. Protein CII làm nhiệm vụ hoạt hoá P_E , quyết định trạng thái tiềm tan xảy ra. Promoter P_{RM} được tự động điều khiển bằng chính nồng độ CI. Protein này bám vào các operator O_{R1} , O_{R2} , O_{R3} cũng như O_{L1} , O_{L2} , O_{L3} .

Khi xâm nhiễm vào tế bào vi khuẩn, hệ gen của phage tồn tại ở dạng vòng. Các phân tử ARNm bắt đầu được tổng hợp từ hai promoter P_L và P_R . Lúc đó trong tế bào *E.coli* không có protein CI. Sản phẩm đầu tiên tạo ra từ P_L là ARNm mã cho protein N ngăn cản việc dừng tổng hợp ARNm tại vị trí kết thúc T_L . Nhờ đó, ARNm mã cho protein CIII được tổng hợp. Đối với promoter P_R , các ARNm ứng với protein Cro, CII lần lượt được tổng hợp. Ngoài ra hai promoter P_E và P_{RM} của gen mã cho protein CI cũng hoạt động. Tuy nhiên P_{RM} bị ức chế bởi protein Cro nên ARNm tạo ra từ promoter này rất ít. ARNm của protein CI được tổng hợp chủ yếu từ promoter P_E (Hình 2.12).

Các protein cần thiết để ADN phage ghép vào nhiễm sắc thể *E.coli* (trạng thái tiềm tan) được mã bởi các gen nằm bên trái, sau gen CIII. Các gen liên quan đến trạng thái sinh tan nằm bên phải sau gen CII. Việc một trong hai trạng thái xảy ra phụ thuộc vào tốc độ tổng hợp ARNm từ hai promoter P_L , P_R cũng như hoạt tính các protein tạo ra. Khi nồng độ protein CI cao, trạng thái tiềm tan được xác lập. Ở giai đoạn đầu tiên, P_E được hoạt hoá bởi CII, do đó lượng CI tăng dần. Protein này lần lượt tương tác với các operator O_{R1} , O_{R2} , O_{R3} với độ nhạy cảm $O_{R1} \gg O_{R2} \gg O_{R3}$.

Khi protein CI bám vào O_R , hoạt động của promoter P_R bị kìm hãm. Do đó lượng protein Cro cũng như sản phẩm của các gen bên phải P_R bị giảm. Mặt khác, protein Cro là repressor nên nó có khả năng tương tác với các operator O_R . Tuy nhiên, độ nhạy cảm giữa Cro với ADN giảm theo trình tự $O_{R3} \gg O_{R2} \gg O_{R1}$. Khả năng tương tác với ADN của CI lớn gấp 10 lần so với Cro. Khi lượng Cro giảm, hoạt động của promoter P_{RM} mạnh lên và lượng CI cũng tăng theo.

Protein CI bám vào O_R còn gây ra các ảnh hưởng khác nhau đối với hoạt động của P_{RM} . Hoạt động của promoter này tăng mạnh khi CI bám vào O_{R1} và O_{R2} làm cho nồng độ CI trong tế bào càng tăng. Khi lượng CI đạt đến một giá trị nhất định, CI bám cả vào O_{R3} gây kìm hãm hoạt động của promoter P_{RM} . Như vậy tùy thuộc vào nồng độ CI, tùy thuộc vào vị trí tương tác của nó với các operator O_R mà protein CI có vai trò tự điều khiển hoạt động của chính gen mã cho nó thông qua cả hai cơ chế tích cực và tiêu cực. Ngoài ra, protein này còn là repressor đối với hai promoter P_R và P_L , đặc biệt đối với P_R . Vì vậy các gen liên quan đến trạng thái sinh tan bị ức chế hoàn toàn, tạo điều kiện thuận lợi cho trạng thái tiềm tan xảy ra.

Với bất cứ một lý do nào đó làm cho lượng protein CI giảm (ví dụ như lượng protease tăng hoặc tác dụng của tia tử ngoại sẽ phân hủy các dimer CI làm chúng mất hoạt tính), operator O_{R3} được giải phóng khỏi CI nhưng bị tương tác bởi Cro. Do đó promoter P_{RM} bị ức chế làm cho nồng độ CI càng giảm mạnh. Một khi các operator O_{R1} , O_{R2} , O_{L1} , O_{L2} không liên kết với Cro, promoter P_L được kích thích dẫn đến lượng protein N tăng và nồng độ các ARNm tổng hợp từ P_R cũng tăng theo. Vì vậy sản phẩm của các gen nằm bên phải P_R tăng, khởi động cho việc chuyển sang trạng thái sinh tan.

2.5. TỔNG HỢP ARNm EUKARYOT

Các phân tử ARN prokaryot (ARNm, ARNr, ARNt) được tổng hợp nhờ một loại enzym ARN polymerase. Trong tế bào eukaryot, quá trình phiên mã phụ thuộc vào từng loại promoter cũng như từng loại ARN polymerase. Các ARNr được tổng hợp nhờ ARN polymerase I, ARNm được tổng hợp bởi ARN polymerase II và các ARNt được phiên mã nhờ ARN polymerase

III. Promoter cho các ARN polymerase I và II thường nằm trước vùng chứa mã di truyền trong khi một số promoter cho ARN polymerase III nằm ngay trong vùng này (phía sau điểm bắt đầu phiên mã +1).

Cùng hoạt động với các enzym còn có các factor cần thiết tham gia vào phản ứng tổng hợp ARN. Có thể chia các factor phối hợp cùng ARN polymerase II để tổng hợp các ARNm làm 3 nhóm khác nhau:

- Các factor kết hợp cùng ARN polymerase tạo phức nhận biết vị trí đầu tiên (+1) để phiên mã.

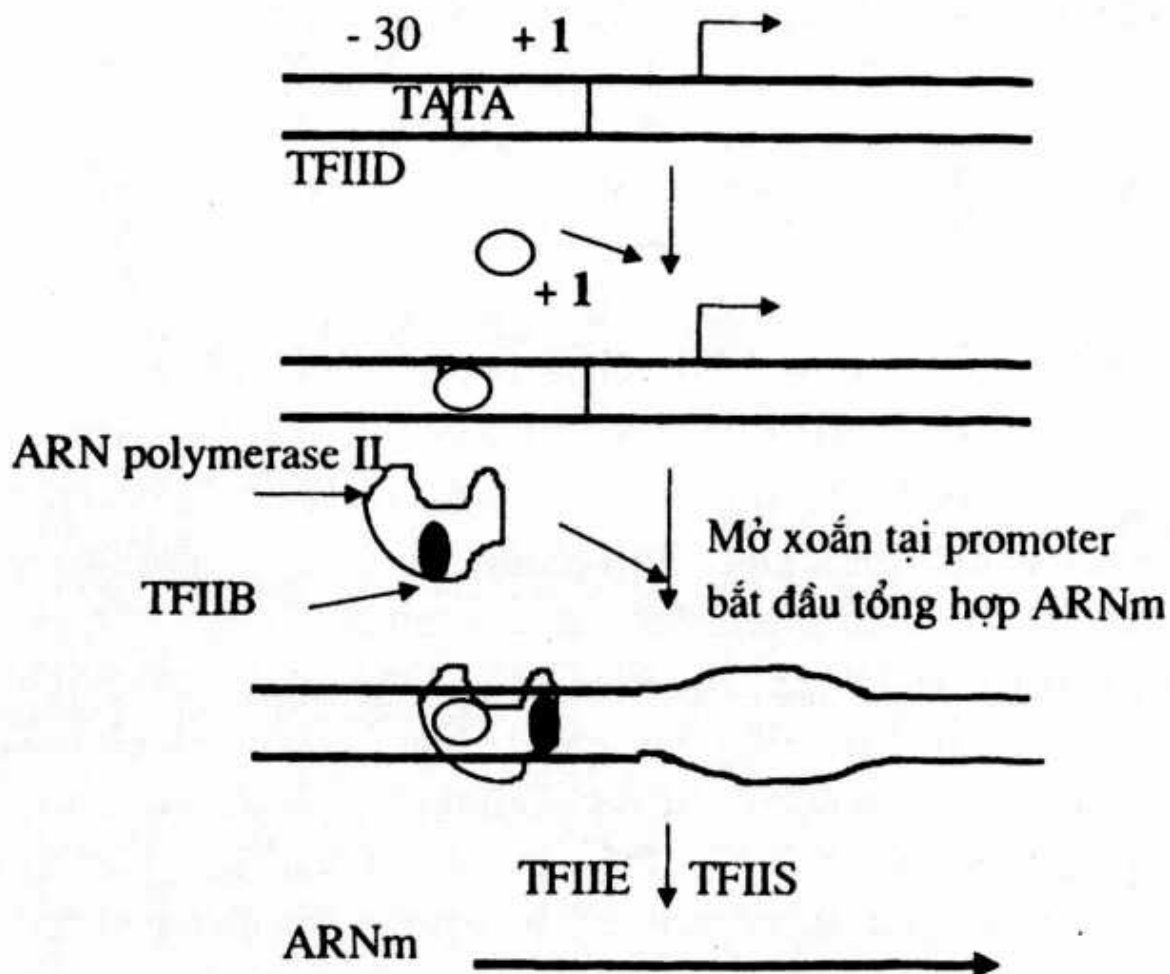
- Các factor (thường là các protein) bám vào vùng ADN điều khiển nằm trước vị trí đầu tiên được phiên mã.

- Các factor làm nhiệm vụ điều khiển quá trình phiên mã theo thời gian và không gian. Chúng được sinh tổng hợp và hoạt hóa tại những thời điểm nhất định và trong các mô, các tế bào đặc hiệu. Các factor này được gọi là factor kích thích. Chúng bám vào những vị trí ADN đặc biệt được gọi là các yếu tố trả lời (response elements).

Điều khác biệt giữa quá trình phiên mã ở tế bào prokaryot và eukaryot được biểu hiện rõ ràng là promoter eukaryot yêu cầu nhiều factor tham gia vào việc khởi động quá trình phiên mã. Các factor cần thiết để tổng hợp ARNm (mà không phải là các thành phần của enzym) được gọi chung là các factor phiên mã (transcription factor). Có thể xảy ra tương tác giữa các factor, tương tác giữa factor với các hợp phần khác của enzym (tương tác protein -protein) và tương tác giữa factor với ADN. Ngoài ra các cơ chế kết thúc phản ứng tổng hợp ARNm ở tế bào eukaryot chưa được xác định rõ ràng như ở prokaryot. Phân tử ARNm eukaryot được tổng hợp trong nhân, phải trải qua nhiều biến đổi trước khi được sử dụng làm khuôn tổng hợp protein trong tế bào chất. Do đó quá trình phiên mã (xảy ra trong nhân

tế bào) và quá trình dịch mã (xảy ra trong tế bào chất) hoàn toàn cách biệt và không xảy ra đồng thời như đối với tế bào prokaryot.

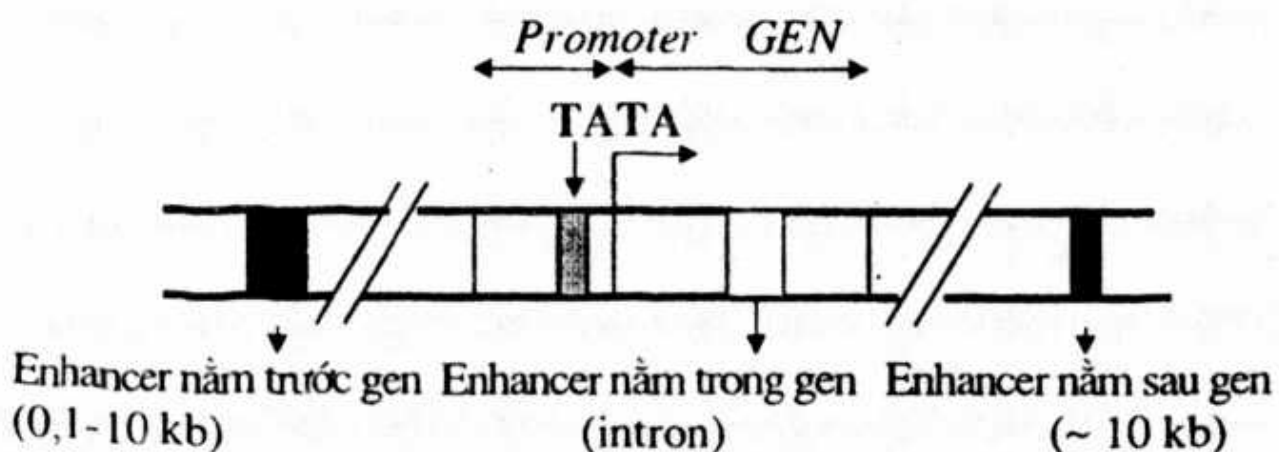
Trong tế bào eukaryot, quá trình tổng hợp phân tử ARNm được xúc tác nhờ enzym ARN polymerase II. Tuy nhiên enzym này không tự bắt đầu phản ứng mà đòi hỏi sự phối hợp của nhiều factor phiên mã. Các factor này giúp enzym nhận biết các vị trí bảo toàn trên promoter hoặc làm tăng khả năng bắt đầu phản ứng do chúng liên kết với enzym hoặc với ADN ở các vị trí khác ngoài promoter (Hình 2.13).



Hình 2.13: Các factor phiên mã tham gia phản ứng tổng hợp ARNm. Chúng có thể liên kết với ADN (như TFIID) hoặc chỉ liên kết với enzym ARN polymerase II (như TFIIB) hoặc liên kết với các protein (như TFIIE...).

Đoạn nucleotide ngắn được bảo toàn nằm ở vị trí -10 (với hầu hết promoter prokaryot) và ở vị trí -30 đến -25 (với promoter eukaryot) phía trước nucleotide đầu tiên (+1) được phiên mã. Đoạn này được gọi là **hộp T₈₂A₉₇T₉₃A₈₇** (các chữ số chỉ % xuất hiện của nucleotide trong các promoter đã được nghiên cứu). Ngoài ra, khi phân tích hơn 150 promoter của các gen động thực vật, các đoạn nucleotide ở vị trí -75: **CCAAT** và ở vị trí -90: **GGGCGG** giữ vai trò quan trọng trong quá trình nhận biết vị trí đầu tiên +1. Chúng thường tương tác với các factor phiên mã.

Đặc biệt, một số đoạn nucleotide (được gọi là các đoạn tăng cường-enhancer) ở gần (khoảng 100 base) hoặc ở xa gen (5kb-40 kb); nằm trước hoặc nằm sau, thậm chí nằm ngay trong gen (ở các intron) có tác dụng tăng hiệu quả phản ứng tổng hợp ARNm. Các đoạn tăng cường thường liên kết với một số protein làm thay đổi cấu trúc của sợi ADN, nhờ đó các vị trí cần thiết như promoter được bộc lộ ra, do đó enzyme ARN polymerase dễ dàng bắt đầu phản ứng (Hình 2.14).



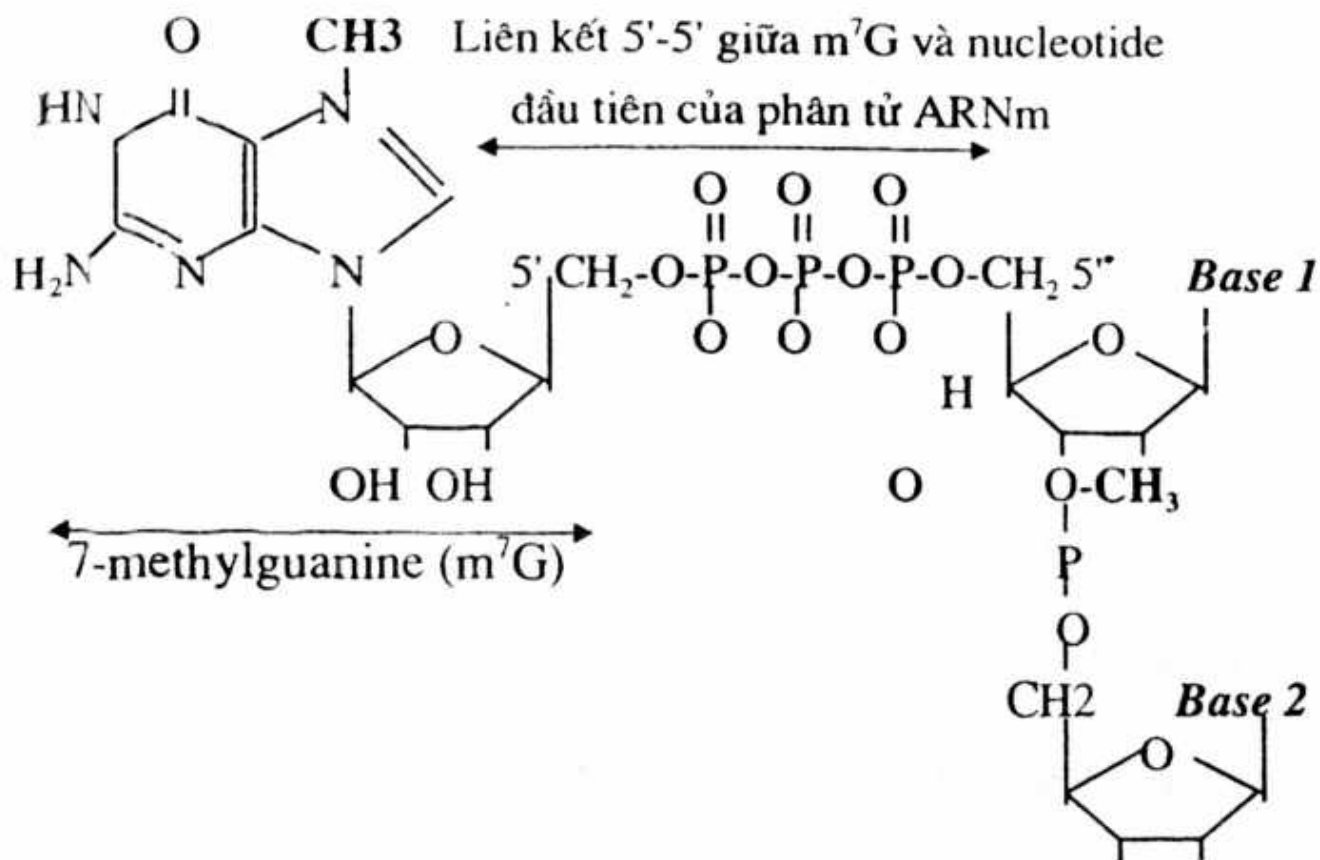
Hình 2.14: Các đoạn nucleotide tăng cường "enhancer" hoạt hoá phản ứng tổng hợp ARNm. Chúng có thể nằm gần hoặc xa, phía trước hay phía sau hoặc nằm ngay trong gen. Chúng có thể hoạt động theo cả hai chiều trên phân tử ADN.

Cho đến nay vị trí dừng chính xác quá trình tổng hợp ARNm ở tế bào eukaryot vẫn chưa được xác định. Tuy nhiên với hầu hết các gen mã cho protein, phản ứng được dừng lại sau khi ARN polymerase II vượt qua vị trí đặc biệt AATAAA (gọi là vị trí polyA) khoảng 0,5 đến 2 kb. Trong thí nghiệm *invitro*, việc loại bỏ vị trí polyA hoặc gây các đột biến trong đoạn đó đều dẫn tới tổng hợp các phân tử ARNm rất dài do enzym không xác định được vị trí dừng. Cơ chế khiến enzym dừng tổng hợp ARNm khi đi qua vị trí này vẫn chưa sáng tỏ. Rất có thể tồn tại factor AT làm nhiệm vụ chống dừng phản ứng. Factor này liên kết với ARN polymerase II trong quá trình phiên mã. Khi đến vị trí polyA, nó tách ra khỏi enzym khiến cho enzym không thể tiếp tục phản ứng.

Phân tử ARNm eukaryot hầu hết là các monocistronic, tức là một phân tử ARNm mang mã di truyền cho một chuỗi polypeptide. Chỉ có rất ít trường hợp (phát hiện được ở virus động vật), một phân tử ARNm eukaryot có hai hay nhiều vị trí bắt đầu tổng hợp protein (điều này xảy ra phổ biến với các ARNm polycistronic ở prokaryot).

Điều khác biệt quan trọng giữa các phân tử ARNm eukaryot và prokaryot được thể hiện ở các điểm sau:

a) Phân tử ARNm eukaryot mang mũ m⁷Gppp ở đầu 5'. Các phân tử ARNm tách chiết được đều có 7-methylguanylate (m⁷G) liên kết 5'-5' với nucleotide đầu tiên của ARNm. Mũ m⁷G này đóng vai trò giúp ribosome nhận biết ARNm trong quá trình tổng hợp polypeptide (Hình 2.15).



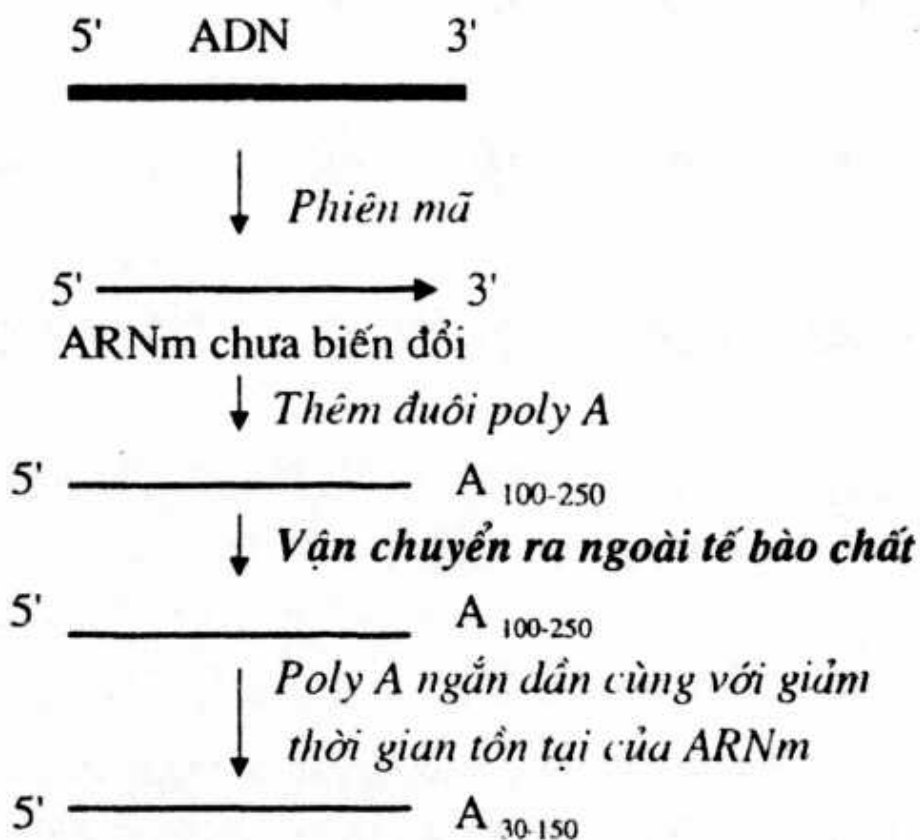
Hình 2.15: Cấu trúc mũ m^7Gppp của phân tử ARNm eukaryot. Các phân tử ARNm không có đầu 5' tự do. Chúng còn bị methyl hoá ở vị trí 2' của đường ribose. Cấu trúc mũ giúp cho ribosome nhận biết được phân tử ARNm.

b) Đuôi polyA được thêm vào đầu 3' của các phân tử ARNm eukaryot. Chiều dài của đuôi polyA thay đổi tùy thuộc loài, cá thể riêng biệt, thậm chí giữa các gen khác nhau trong một cơ thể. Đuôi polyA không tồn tại trong phân tử ADN, nó cũng không được thêm vào các ARNt, ARNr. Các ARNm sau khi gắn thêm đuôi được chuyển ra ngoài tế bào chất. Ở đó đuôi này bị ngắn dần đi và thường còn lại khoảng 30 đến 150 base (Hình 2.16).

Đuôi polyA tham gia quá trình vận chuyển ARNm từ nhân ra ngoài tế bào chất. Mặt khác độ dài của đuôi này gắn liền với tính bền vững của phân tử ARNm. Các ARNm không có hoặc có đuôi polyA ngắn thường bị phân hủy nhanh chóng. Tuy nhiên

đuôi này không phải là yêu cầu tuyệt đối cho mọi ARNm eukaryot. Trên thực tế các ARNm mã cho protein histone không có đuôi polyA (giống như ARNm prokaryot).

Nhờ có đuôi polyA, các phân tử ARNm có thể tách được ra khỏi các ARNr, ARNt khi cho ARN tổng số đi qua cột sắc ký chứa polyU (hoặc polyT). Các polyT sẽ tạo cặp với polyA và giữ các ARNm lại, cho các loại ARN khác đi qua cột. Sau đó ARNm được tách rửa khỏi cột trong điều kiện nồng độ các ion thấp (không cho phép tồn tại các liên kết giữa A và T).



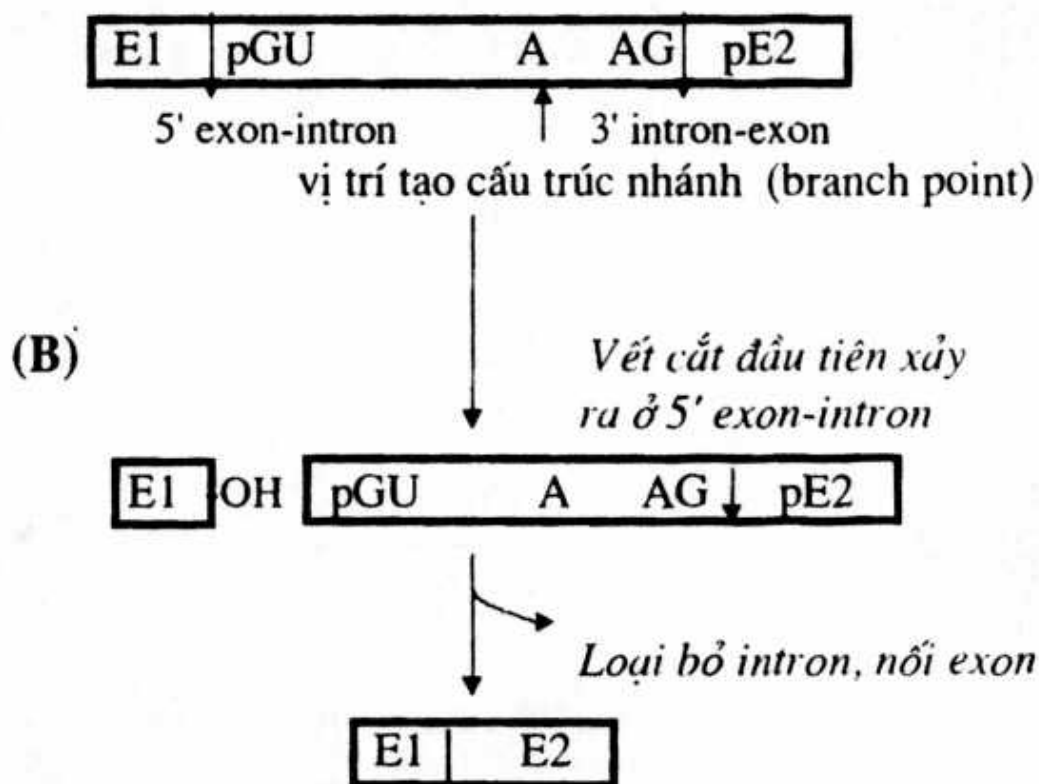
Hình 2.16: Thêm đuôi polyA vào phân tử ARNm trong tế bào eukaryot. Đuôi polyA không được mã trong ADN mà được thêm vào đầu 3' của ARNm nhờ enzym polyA polymerase. Ở ngoài tế bào chất đuôi này ngắn dần cùng với việc giảm dần thời gian sống của ARNm.

Phân tử ARNm sau khi được thêm mũ G' và thêm đuôi polyA phải trải qua quá trình cắt bỏ các intron (các đoạn nucleotide không chứa mã di truyền). Khoảng 30-40 nucleotide nằm ở biên giới giữa các intron và exon cần thiết cho phản ứng cắt intron xảy ra chính xác. Việc loại bỏ các nucleotide nằm ở trung tâm intron không ảnh hưởng đến phản ứng này. Đặc biệt hai nucleotide đầu tiên **GU** ở đầu 5' và **AG** tận cùng ở đầu 3' của mỗi intron đóng vai trò quyết định. Chúng được bảo toàn trong mọi intron và là vị trí nhận biết để cắt intron ra khỏi phân tử tiền thân ARNm (Hình 2.17A).

Phản ứng cắt bỏ intron ra khỏi các phân tử ARNm đòi hỏi sự tham gia của một số phân tử snARNs kích thước nhỏ (small nuclear ARN) liên kết với protein tạo thành các ribonucleoprotein (snRNPs). Một snRNP thường chứa một phân tử ARN và khoảng 10 protein khác nhau. Một số protein đặc hiệu cho từng loại snRNP, một số khác chung cho các snRNPs. Phân tử snARN của một số snRNP có khả năng tạo liên kết theo nguyên tắc bổ sung với các đoạn nucleotide ở vùng biên giới 5' hoặc 3' giữa exon và intron. Một số snARNs khác có thể liên kết liên kết với nhau hoặc với intron. Những liên kết này giữ vai trò quan trọng trong phản ứng loại bỏ intron. Phản ứng cắt intron và nối hai exon với nhau được thực hiện theo các bước cơ bản như sau (Hình 2.17B):

	exon 5'		intron								3' exon		
(A)	A	G	G	U	A	A	G	U	Giàu Pyr	C	A	G	G
	62	77	100	100	60	70	84	50	~ 15 base	78	100	100	55

Phân tử tiền thân ARNm



Hình 2.17: (A) Xác suất các nucleotide xuất hiện tại biên giới giữa exon-intron ở phân tử tiền thân ARNm eukaryot. Các nucleotide **GU** và **AG** (in đậm) được bảo toàn trong mọi intron. Vùng giàu Pyrimidine (Pyr) nằm gần đầu 3' chiếm khoảng 63-91% trong các intron. (B) Phản ứng loại bỏ intron ra khỏi phân tử tiền thân ARNm. Sau khi biên giới giữa exon-intron ở đầu 5' bị cắt, cấu trúc vòng của intron được tạo ra do liên kết 5'-2' giữa G của đầu 5' intron với A nằm gần đầu 3' của intron. Việc cắt tại nơi tiếp giáp giữa intron-exon ở đầu 3' cũng như việc nối hai exon với nhau xảy ra đồng thời.

+ Cắt tại biên giới exon-intron ở đầu 5' của phân tử tiền thân ARNm, đầu 5' của exon được giải phóng.

+ Nucleotide 5'pG ở đầu intron vừa bị cắt liên kết với 2'OH của adenine nằm cách đầu kia (phía 3') khoảng 25 base. Liên kết này khiến cho intron có cấu trúc vòng.

+ Cắt tại biên giới intron-exon ở đầu 3' của phân tử ARNm. Intron được loại ra và hai exon được nối với nhau.

Trong thực tế, mọi snRNP và các protein liên quan đến phản ứng cắt bỏ intron tạo nên một cấu trúc có thể quan sát dưới kính hiển vi điện tử gọi là các spliceosome. Chúng có kích thước tương tự như ribosome, bao gồm các liên kết ARN-ARN, ARN-protein và protein-protein. Một số protein của spliceosome là các enzym helicase phân giải ATP, giải phóng năng lượng làm thay đổi cấu trúc không gian cần thiết của phân tử ARNm.

Phản ứng cắt nối exon-intron có thể xảy ra một cách tự động không đòi hỏi bất cứ protein hoặc snARNs nào (phản ứng ***self-splicing***). Chính cấu trúc không gian phức tạp của các intron, đặc biệt tương tác giữa các chuỗi nucleotide ngược chiều trong intron giữ vai trò của các snARN trong quá trình tự cắt nối. Ngoài ra, exon của phân tử ARNm này có thể được nối với exon của phân tử khác (hiện tượng ***trans-splicing***) mà chúng ta đã được biết đến trong quá trình tổng hợp protein bề mặt của *Trypanosome*.

2.6. KIỂM SOÁT SAU PHIÊN MÃ

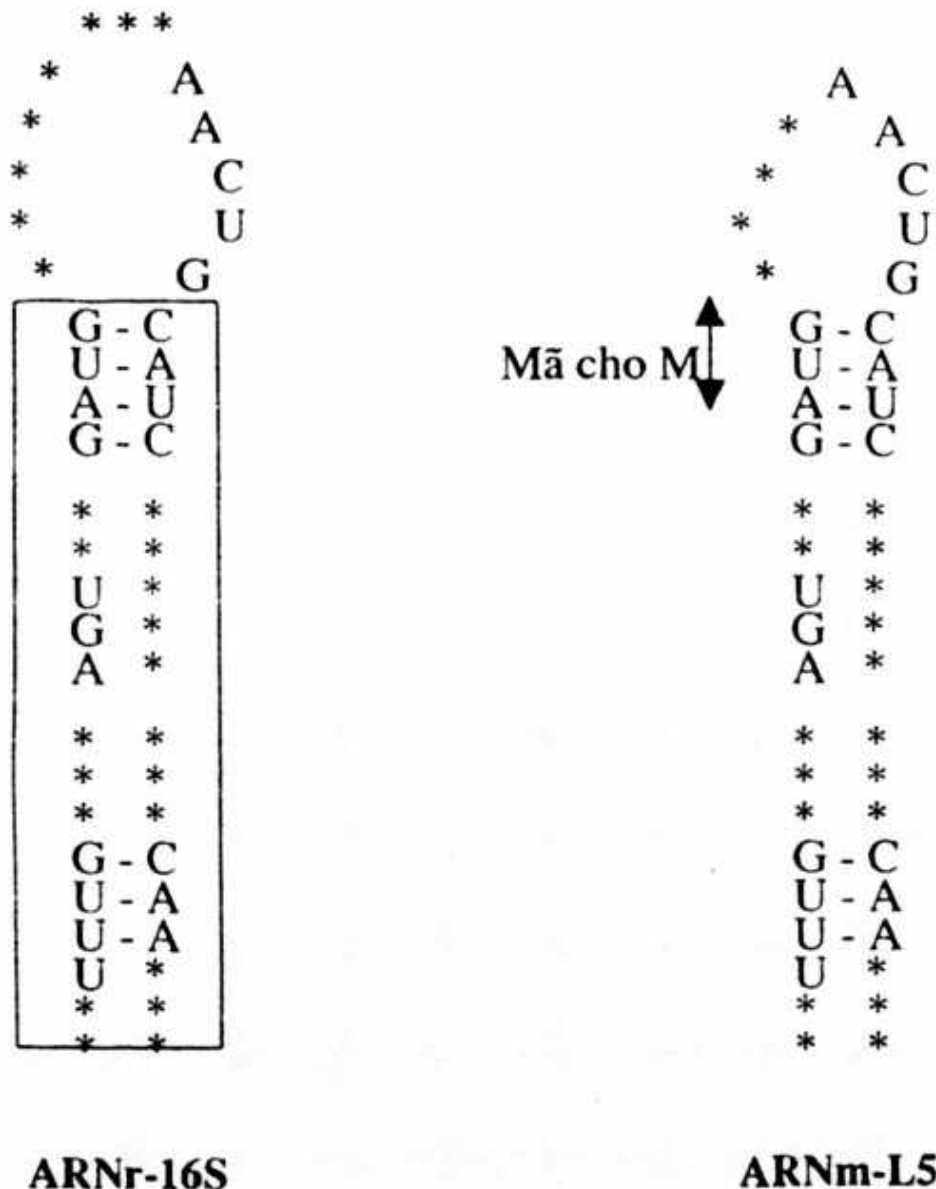
Sản phẩm cuối cùng của gen thường là các protein. Để đáp ứng nhanh nhạy, chính xác mọi yêu cầu của tế bào về số lượng protein, hoạt động của một gen được điều khiển xuyên suốt từ

tổng hợp ARNm đến tổng hợp protein. Trước khi được dịch mã, phân tử ARNm trải qua các cơ chế kiểm soát tính bền vững, kiểm soát số lượng cần thiết sử dụng làm khuôn v.v...

2.6.1. Sinh tổng hợp các protein ribosome được kiểm soát thông qua quá trình dịch mã trên ARNm

Tỷ lệ giữa ARNr và các protein ribosome trong tế bào luôn luôn tương đồng với nhau và phụ thuộc vào sự phát triển của tế bào. Trong thí nghiệm tổng hợp protein ribosome *invitro*, khi nồng độ các protein này quá cao, một số gen mã cho chúng vẫn tiếp tục phiên mã tạo phân tử ARNm. Tuy nhiên các phân tử này không được sử dụng để tổng hợp protein. Điều này chứng tỏ quá trình dịch mã từ ARNm sang các protein ribosome được quyết định bởi sự dư thừa của chính các protein này.

Thực nghiệm đã chứng tỏ rằng protein S8 (protein số 8 liên kết với ARNr 16s nằm trong tiểu phần nhỏ của ribosome) ức chế quá trình tổng hợp protein L5 (protein số 5 thuộc tiểu đơn vị lớn của ribosome). Khi so sánh trình tự ribonucleotide giữa hai phân tử ARNr 16s và ARNm của L5 thì thấy chúng có chứa các đoạn khá tương đồng (Hình 2.18). Do đó, nếu nồng độ S8 quá lớn thì nó sẽ liên kết với cả ARNm của protein L5, vì vậy ngăn cản việc dịch mã trên phân tử ARNm này. Do các phân tử ARNm trong tế bào vi khuẩn bị phân hủy nhanh, liên kết của S8 với ARNm của L5 khiến cho nồng độ phân tử này giảm mạnh trong tế bào.



Hình 2.18: Cấu trúc bậc một và cấu trúc bậc hai của một đoạn ARNr 16s và phân tử ARNm của protein L5. Vị trí bám của protein S8 vào phân tử ARNr 16s được đóng khung. Sự giống nhau về cấu trúc cũng như sự tương đồng nucleotide khiến cho ARNr 16s và ARNm cạnh tranh với nhau trong liên kết với protein S8.

2.6.2. ARRN anti-sense

Thông thường, các protein giữ vai trò chủ đạo kiểm soát quá trình phiên mã cũng như dịch mã. Tuy nhiên hoạt động của một số gen lại được điều khiển bởi các ARNs. Cũng giống như

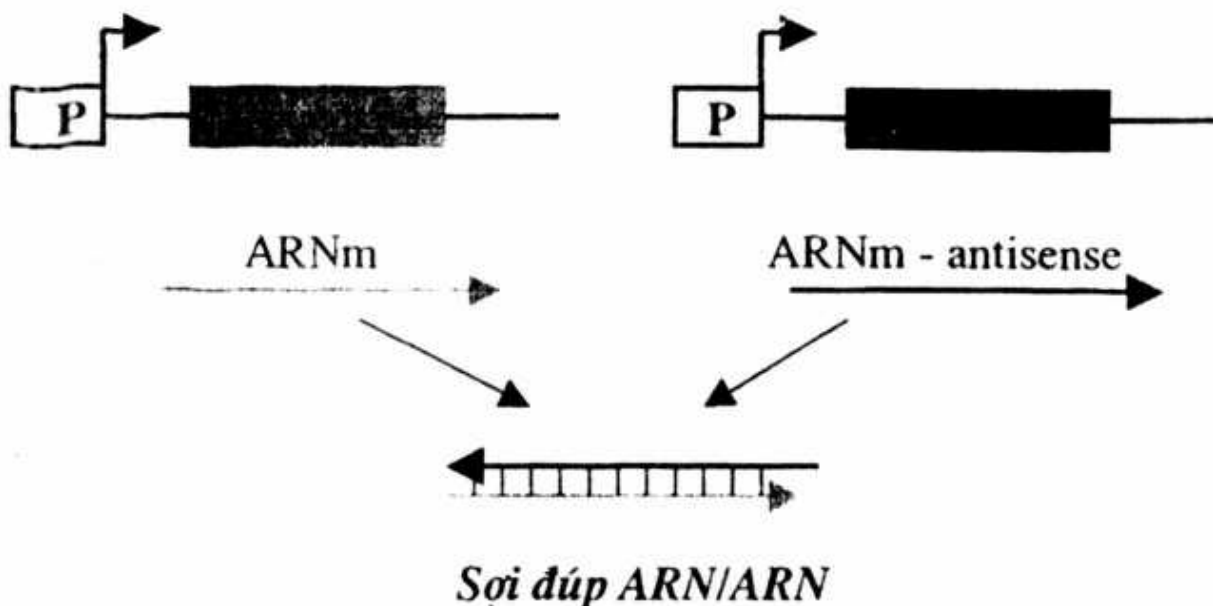
các protein điều khiển, các ARNs làm nhiệm vụ kiểm soát hoạt động của gen được tổng hợp một cách độc lập và tương tác với các vị trí đặc hiệu (các đoạn nucleotide). Chúng tạo cặp theo nguyên tắc bổ sung với các ARNm. Hoạt động của các ARNs này có thể xảy ra theo các cơ chế như:

- Tạo ra các cấu trúc sợi kép để che dấu các đoạn nucleotide cần thiết cho dịch mã. Ví dụ như để che vị trí bắt đầu dịch mã trên sợi ARNm.

- Tạo ra các vùng có cấu trúc sợi kép gây thay đổi cấu trúc không gian của những vùng khác trên phân tử ARNm, làm ảnh hưởng đến chức năng của những vùng này.

- Tạo ra các sợi đúp ARN/ARNm được nhận biết và phân cắt bởi các ribosome.

Xét trường hợp cụ thể về vai trò điều khiển sau phiên mã của ARNs đối với việc tổng hợp protein OmpF ở tế bào vi khuẩn *E.coli*. Protein này nằm phía ngoài màng tế bào làm nhiệm vụ nhận biết những thay đổi về áp suất thẩm thấu trong môi trường. Khi áp suất tăng thì gen điều khiển *micF* hoạt động mạnh hơn. Sản phẩm của gen *micF* là phân tử ARN 174 bases, có thể tạo cặp theo nguyên tắc bổ sung với vị trí nhận biết của ribosome trên phân tử ARNm của gen *ompF*. Như vậy phân tử ARNm của *micF* làm nhiệm vụ ngăn cản việc tổng hợp protein OmpF. Các phân tử ARN có chức năng tương tự *micF* được gọi là ARN-antisense. Kỹ thuật sinh học hiện đại cho phép tạo ra các gen tổng hợp ARN-antisense bằng việc đổi chiều của một gen nào đó (tính từ promoter) (Hình 2.19). Sử dụng các ARN-antisense là một trong những phương pháp điều khiển hoạt động của gen.



Hình 2.19: Phân tử ARN-antisense có thể được tạo ra bằng việc đối chiều của gen tính từ promoter (P)

Biến đổi phân tử tiền thân ARNm xảy ra ở trong nhân tế bào. Sau đó các phân tử ARNm được vận chuyển ra ngoài tế bào chất. Ngay ở ngoài tế bào chất, ARNm tiếp tục trải qua các biến đổi trước khi được sử dụng vào quá trình tổng hợp protein. Nói một cách khác, quá trình dịch mã (tổng hợp protein) được kiểm tra một cách rất chặt chẽ thông qua kiểm tra lại thông tin di truyền mang trên phân tử ARNm, độ bền vững của phân tử này v.v...

2.6.3. Phản ứng đọc sửa ARNm - "RNA editing"

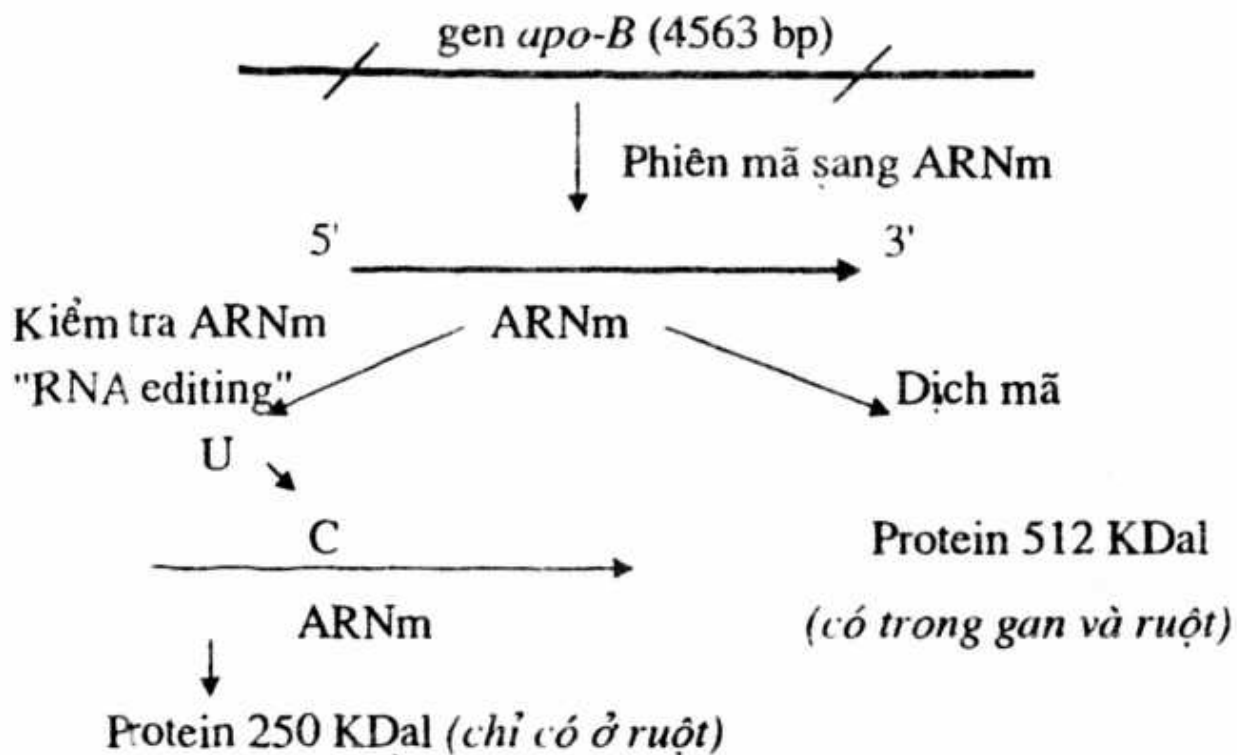
Đây là phản ứng làm thay đổi trình tự nucleotide của phân tử ARNm. Phản ứng này xảy ra trong tế bào chất. Các nucleotide của một số phân tử ARNm bị loại bỏ hoặc bị thay thế, một số nucleotide khác được thêm vào. Như vậy trình tự nucleotide trên phân tử ARNm khác với gen (ADN) mà từ đó phân tử ARNm được phiên mã. Phản ứng này được phát hiện đầu tiên đối với ARNm mã cho protein ở ty thể của

Trypanosome. Các phân tử ARNm này mang thêm một vài nucleotide U khác với trình tự trên gen ADN.

Các phân tử ARNm trong ty thể của tế bào thực vật hầu như đều trải qua phản ứng "RNA editing". Tuy nhiên đối với các ARNm này, chỉ có sự thay thế nucleotide C bằng U mà không có hiện tượng thêm vào hoặc loại bỏ các nucleotide khác. Trình tự các acid amin ở protein có thể thay đổi đến 10% hoặc hơn nữa so với protein được mã bởi phân tử ARNm chưa trải qua phản ứng này.

Ở động vật, các phân tử ARNm ít trải qua phản ứng "RNA editing". Trường hợp đầu tiên được phát hiện đối với gen *apolipoprotein B (apo-B)* dài 4503 bp, là gen đơn bản trong genome (Hình 2.20). Từ gen này hai loại phân tử ARNm được tổng hợp. Chúng chỉ khác nhau bởi một nucleotide C bị thay thế bởi U. Do đó làm xuất hiện một mã dừng (stop codon) phản ứng tổng hợp protein. Protein ngắn tương ứng với loại ARNm này có trọng lượng phân tử 250 KDal và chỉ tồn tại ở trong ruột. Phân tử protein 512 KDal tương ứng với ARNm không bị đọc sửa tồn tại ở cả gan và ruột.

Phản ứng "RNA editing" được thực hiện nhờ các phân tử ARN dài khoảng 40 - 80 base. Các ARN này gọi là ARN phụ trợ. Chúng được tổng hợp độc lập với nhau. Đầu 5' của chúng có thể tạo liên kết với ARNm cần sửa đổi, còn đầu 3' mang đuôi polyU. Các nucleotide U được chuyển trực tiếp từ đuôi này sang ARNm. Quá trình đọc sửa diễn ra từ đầu 3' và tiếp tục về đầu 5' của ARNm. Có thể có nhiều ARN phụ trợ tham gia sửa đổi một phân tử ARNm.



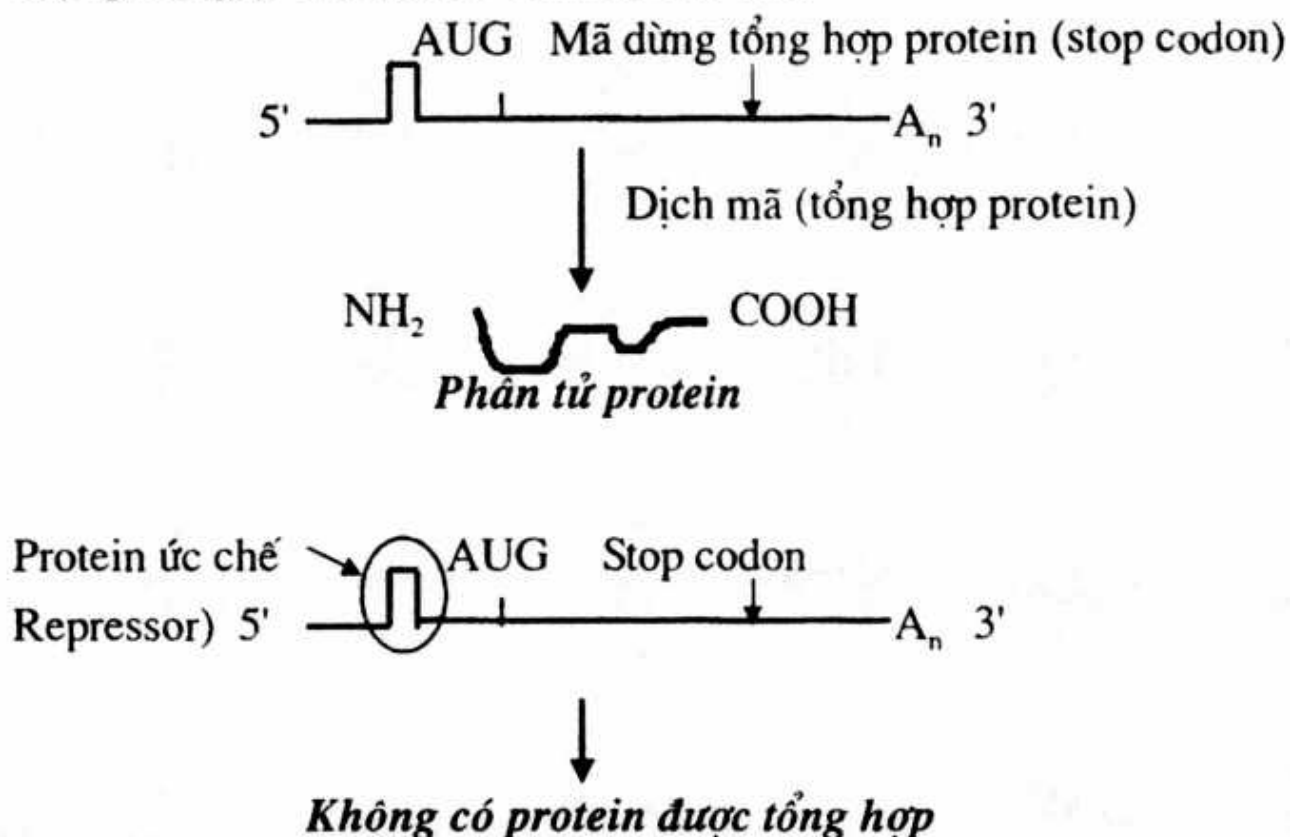
Hình 2.20: Thay đổi thông tin di truyền trên phân tử ARNm của gen apo-B bởi cơ chế "ARN editing". Chiều dài phân tử ARNm không đổi nhưng nucleotide U thay thế C tạo ra một mã dừng tổng hợp protein (stop codon)

2.6.4. Protein bám vào đầu 5' của phân tử ARNm kìm hãm quá trình dịch mã - Phản ứng "dịch mã tiêu cực"

Chúng ta biết rằng hai đầu của ARNm có các vùng 5' và 3' không dịch mã. Đầu 5' không dịch mã có thể là vị trí tương tác của một số protein repressor. Khi đó ribosome không thể dịch chuyển trên sợi ARNm khiến cho quá trình dịch mã bị kìm hãm hoàn toàn. Rõ ràng trong trường hợp này, ARNm vẫn được tổng hợp một cách bình thường nhưng protein không được tạo ra. Cơ chế điều khiển quá trình dịch mã do protein repressor tương tác với đầu 5' của ARNm được gọi là "kiểm soát dịch mã tiêu cực" (negative translational control) (Hình 2.21).

Ở tế bào eukaryot, kiểm soát dịch mã tiêu cực được phát hiện đầu tiên với phản ứng tổng hợp ferritin - protein làm nhiệm vụ

tạo phức với các nguyên tử sắt. Khi môi trường không có sắt, các phân tử ARNm ferritin bị aconitase (protein repressor) bám vào đầu 5', do đó không có protein ferritin được tổng hợp trong tế bào. Khi có sắt trong môi trường, aconitase liên kết với sắt và rời khỏi ARNm ferritin. Lúc này ferritin được tổng hợp với số lượng lớn gấp 100 lần so với lúc ban đầu.



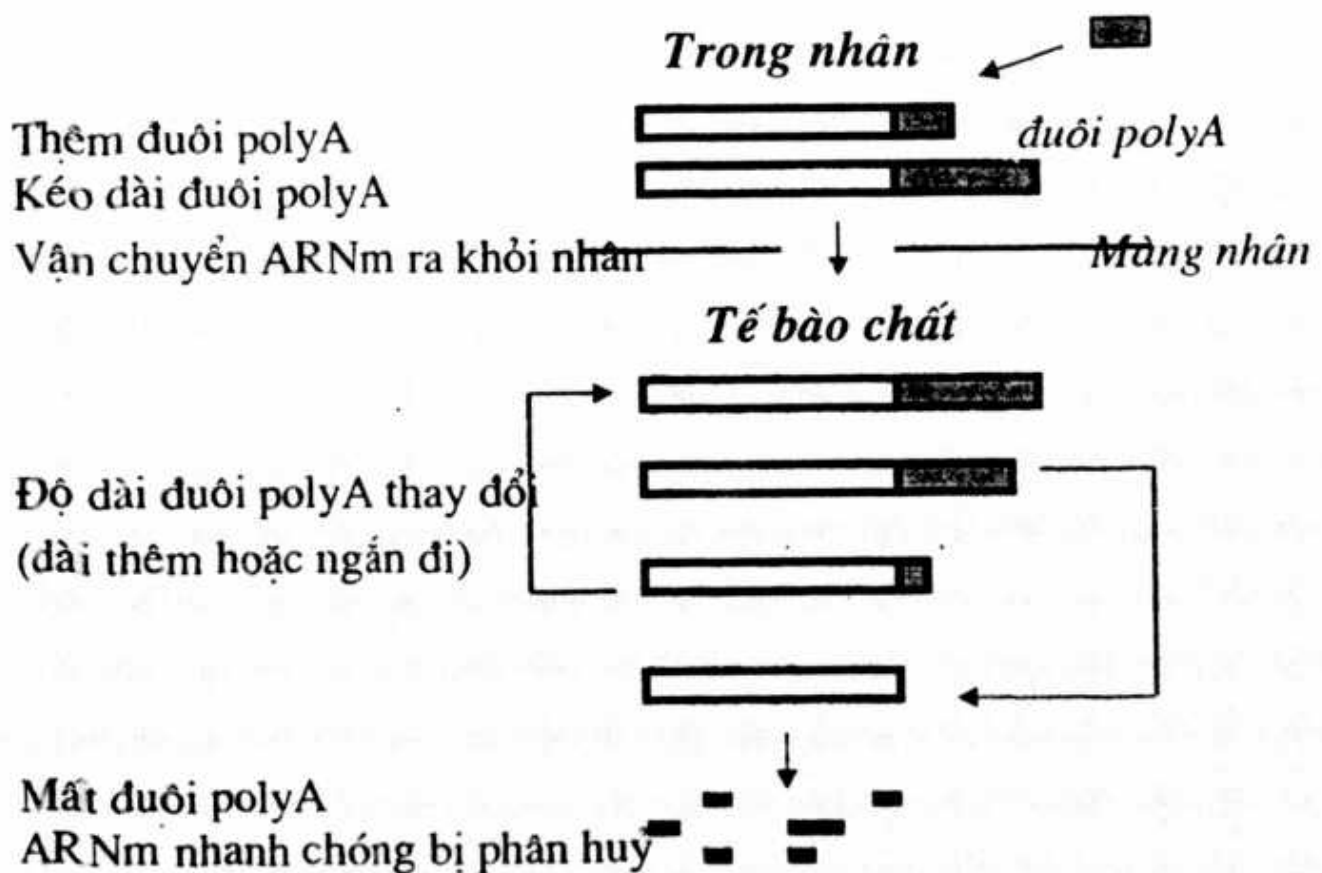
Hình 2.21: Cơ chế "kiểm soát dịch mã tiêu cực". Protein repressor đặc hiệu bám vào đầu 5' của ARNm gây ngừng quá trình tổng hợp protein do ribosome không thể bám vào ARNm

2.6.5. Độ dài của đuôi polyA ảnh hưởng tới độ bền vững của phân tử ARNm

Phản ứng gắn đuôi polyA (~200 A) vào các phân tử ARNm eukaryot xảy ra trong nhân tế bào. Một khi đã chuyển ra ngoài tế bào chất, đuôi này bị cắt ngắn dần. Tuy nhiên trong thực tế, không phát hiện được đuôi polyA nào ngắn hơn 30 A. Do đó đuôi

polyA ngắn nhất để phân tử ARNm không bị phân hủy có thể là 30 A. Đặc biệt, để đảm bảo độ bền vững của đuôi, các nucleotide A còn được thêm vào. Nói cách khác, việc thay đổi độ dài đuôi polyA được kiểm tra rất chặt chẽ (Hình 2.22).

Ví dụ điển hình đối với các tế bào trứng chuẩn bị thụ tinh có chứa rất nhiều loại ARNm. Các phân tử này chỉ mang đuôi polyA gồm 10-30 A, do đó chúng không được dịch mã. Chỉ sau khi thụ tinh, trứng phát triển đòi hỏi các protein với số lượng lớn. Lúc này đuôi polyA được dài thêm và các ARNm được dùng làm khuôn để tổng hợp protein.



Hình 2.22: Độ dài đuôi polyA ảnh hưởng đến độ bền vững và quá trình dịch mã trên phân tử ARNm. Đuôi polyA có thể được thêm vào hoặc cắt ngắn đi, nhưng độ dài của nó không được ít hơn 30 A

2.6.6. Độ bền vững của ARNm

Trong tế bào vi khuẩn, các phân tử ARNm thường bị phân hủy rất nhanh. Thời gian bán hủy của chúng thay đổi từ 3-5 phút ngay khi có hoặc không có các chất cảm ứng (inducer) trong môi trường. Do đó khi tế bào vi khuẩn mọc và phân chia trong khoảng thời gian 20-30 phút, các ARNm cần được tái tạo mới 5-10 lần trong mỗi lần phân chia. Khi tế bào cần thiết một số enzym, ARNm liên quan đến việc tổng hợp các enzym đó được phiên mã nhanh chóng. Khi tế bào không còn nhu cầu, quá trình phiên mã dừng rất nhanh và các phân tử ARNm lập tức bị phân hủy.

Enzym nuclease chịu trách nhiệm phân hủy ARNm đến nay vẫn chưa được xác định rõ ràng. Đối với một số ARNm polycistronic của các operon như *lac* hoặc *trp*, chúng bị phân cắt bởi endonuclease tạo ra các monocistronic tương ứng với từng gen trong operon. Khoảng cách giữa các monocistronic không chứa mã di truyền, do đó không được che chắn bởi ribosome nên chúng dễ bị cắt bởi endonuclease.

Các ARNm prokaryot được tổng hợp cũng như bị phân hủy rất nhanh, do đó vi khuẩn có thể thích ứng nhanh nhạy với những thay đổi của môi trường. Ngược lại, các ARNm eukaryot khá bền vững. Ví dụ như ARNm mã cho β -globin có thời gian bán hủy khoảng hơn 10h. Tuy nhiên cũng có những loại ARNm có thời gian bán hủy chỉ 30 phút hoặc ít hơn. Thường đó là những ARNm mã cho các protein làm nhiệm vụ điều khiển hoạt động của gen.

Các phân tử ARNm eukaryot kém bền do chúng mang các đoạn nucleotide đặc biệt tại đầu 3' kích thích sự phân hủy ARNm. Thực nghiệm cho thấy khi đoạn nucleotide giàu A và U ở vùng 3' không dịch mã của phân tử ARNm kém bền được ghép vào một số ARNm bền vững khác, thì những phân tử này trở nên kém bền do đuôi polyA bị cắt ngắn nhanh. Như vậy đoạn nucleotide giàu AU kích thích sự phân hủy ARNm thông qua việc giảm độ dài đuôi polyA. Ngoài ra, một số ARNm có chứa vị trí đặc hiệu ở đầu 3' được nhận biết bởi endonuclease.

Chương 3

CÁC CƠ CHẾ SỬA CHỮA VÀ TỔNG HỢP ADN

Chỉ có thể đánh giá được một cách gián tiếp tỷ lệ đột biến xảy ra tự phát. Một trong những cách đánh giá đó là xem xét trình tự acid amin của một loại protein ở các loài sinh vật khác nhau. Tỷ lệ số acid amin khác nhau được so sánh với số năm mà hai loài cùng xuất phát từ một nguồn gốc ban đầu (dựa theo cây phân loài). Sự sai khác đó do các đột biến ở mức độ gen, tức là ở các nucleotide. Từ đây có thể đánh giá được số thời gian cần thiết để xảy ra một đột biến ở mức độ ADN.

Gần đây, các kỹ thuật sinh học phân tử hiện đại cho phép so sánh trình tự nucleotide ở một số vùng genome không chứa mã di truyền, hầu hết ở những vùng không mã hóa protein. Kết quả thu được phù hợp với phương pháp dựa vào sự sai khác các acid amin của một protein. Cả hai phương pháp đều cho thấy để xảy ra sự thay đổi tự phát đối với một acid amin của một protein có 400 acid amin thì cần một thời gian khoảng 200.000 năm. Tỷ lệ đột biến rất thấp như vậy chứng tỏ rằng trình tự nucleotide của phân tử ADN rất bền vững.

Nghiên cứu một quần thể có số lượng rất lớn (ruồi giấm, các tế bào vi khuẩn trong nuôi cấy...) cho thấy số cá thể đột biến xuất hiện trong quần thể đó, cũng như những thay đổi xảy ra ở

một protein của tế bào nuôi cấy có thể đặc trưng cho tỷ lệ đột biến. Tỷ lệ đó chỉ khoảng 10^{-9} đối với mỗi thể hệ. Như vậy một gen có khoảng 10^3 bp mang mã di truyền (tương ứng với các protein có kích thước trung bình trong tế bào) sẽ có một đột biến sau 10^6 thế hệ. Kết quả này phù hợp với tính toán theo quá trình tiến hoá, tức là phù hợp với tỷ lệ đột biến xảy ra ở một gen là 200.000 năm/một đột biến.

Nghiên cứu đột biến xảy ra tự phát trong genome cho thấy hầu hết đột biến là do mất các nucleotide. Như vậy chọn lọc tự nhiên và qui luật tiến hoá không cho phép đột biến kiểu này xảy ra với tần số cao. Chúng ta sẽ thấy vì sao. Ngay với tỷ lệ đột biến rất thấp thì số protein quan trọng trong tế bào sẽ có khoảng 60.000 loại khác nhau. Chỉ cần tỷ lệ này tăng gấp 10 lần thì số protein khác nhau chỉ còn khoảng 6000 loại. Như vậy quá trình tiến hoá đã phải dừng lại ở những loài sinh vật không đa dạng phức tạp hơn ruồi giấm.

Như vậy để bảo toàn các đặc tính của loài, thông tin di truyền chứa trong các tế bào sinh dục nhất định phải được bảo vệ tránh không xảy ra các đột biến tự phát với tần số cao. Mặt khác để đảm bảo an toàn cho từng cá thể, thông tin di truyền trong các tế bào soma cũng được bảo vệ tránh những đột biến gây rối loạn phát triển hoặc thậm chí gây chết. Chúng ta sẽ xét đến những đột biến xảy ra với tần số cao ở tế bào soma dẫn đến xuất hiện ung thư. Tính bền vững của tế bào sinh dục và tế bào soma đều phụ thuộc vào tính bền vững của ADN.

Tuy nhiên, xuất phát từ quan điểm của các nhà vật lý và hoá học thì phân tử ADN chịu rất nhiều đột biến tự phát do biến động nhiệt hoặc tác động của các phân tử dung môi. Thực vậy, mỗi ngày ADN của một tế bào trong cơ thể chúng ta mất đi

khoảng 5000 base purine (adenine và guanine) do xảy ra phản ứng khử amin (liên kết N-glycosyl với deoxyribose bị bẻ gãy do dao động nhiệt). Tương tự như vậy sự biến đổi tự phát từ cytosine thành uracine xảy ra với tỷ lệ 1000 base /genome trong một ngày. Ngoài ra, tia tử ngoại của mặt trời có thể thúc đẩy phản ứng tạo cặp bền vững giữa hai pyrimidine nằm cạnh nhau trong chuỗi ADN (ví dụ như tạo dimer giữa hai thymine). Rõ ràng cấu trúc ADN xét theo liên kết hoá học phải chịu các đột biến tự phát với tần số khá cao. Điều này trái ngược với kết quả tính toán nêu ở trên. Vậy cơ chế nào đảm bảo tính bền vững của ADN trước các tác động nhiệt, tác động hoá học v.v...? Phải chăng cơ thể duy trì sự sống của mình cũng như đảm bảo sự trường tồn của thế hệ sau bằng các cơ chế tái bản ADN ngay khi tổn thương đang tồn tại đồng thời tiến hành sửa chữa sai hỏng bằng nhiều cách thức khác nhau.

Hầu hết các đột biến tự phát xảy ra hàng ngày hoặc do các tác nhân bên ngoài gây ra (UV, tác động vật lý, hoá học...) đối với phân tử ADN được khắc phục bởi 3 cách thức sửa chữa chính. Thứ nhất là sửa chữa phục hồi trực tiếp (direct reversal repair); thứ hai là cắt bỏ sai hỏng và sửa chữa lại dựa vào trình tự bổ sung (damage excision and repair using complementary sequence) và thứ ba là chống chịu các tổn thương có thể có (inducible damage tolerance).

Sửa chữa trực tiếp thường liên quan đến hai loại sai hỏng trên phân tử ADN do tia tử ngoại gây ra là: cyclobutane-pyrimidine dimer (CPDs) và pyrimidine (6-4) pyrimidone (6-4PPs). Cả hai loại sai hỏng này đều gây biến dạng cấu trúc xoắn của ADN. CPDs và 6-4 PPs được nhận biết và sửa chữa nhờ photolyase. Các enzym này sử dụng năng lượng ánh sáng để

thực hiện phản ứng làm thay đổi các liên kết hoá học để nucleotide trở lại dạng bình thường. Phản ứng sửa chữa ADN bằng photolyase xảy ra trong rất nhiều sinh vật prokaryot và eukaryot. Tuy nhiên, quá trình này chưa được phát hiện ở động vật có vú. Ở người chưa phát hiện được bất kỳ loại photolyase nào.

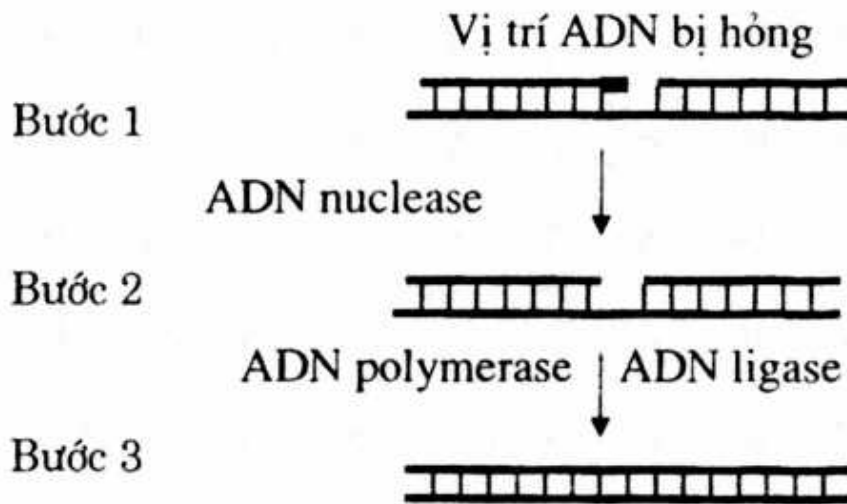
Đa số sai hỏng ADN được sửa chữa bằng con đường thứ hai. Cơ chế này phải sử dụng thông tin di truyền chứa ở một trong hai sợi đơn ADN. Khi trình tự nucleotide trên một sợi bị thay đổi thì sợi thứ hai (liên kết bổ sung với sợi thứ nhất) được dùng làm khuôn để sửa chữa sai sót. Trong trường hợp cả hai sợi đơn đều bị tổn thương, sai hỏng có thể được sửa chữa nhờ đoạn nucleotide tương đồng tồn tại đâu đó trong hệ gen. Cơ chế chống chịu tổn thương sử dụng một số loại ADN polymerase đặc biệt và các protein liên quan (thuộc họ protein Y) có khả năng vượt qua những sai hỏng trên sợi khuôn trong quá trình tái bản ADN. Như vậy, các sai hỏng vẫn tồn tại nhưng quá trình tái bản ADN vẫn xảy ra. Tuy nhiên tính chính xác của những enzym này trong phản ứng tổng hợp ADN có nhiều hạn chế.

Các bước cơ bản để sửa chữa ADN theo cơ chế cắt bỏ các sai hỏng được mô tả trên Hình 3.1 gồm:

Bước 1: Phần ADN mang đột biến được nhận biết và cắt bỏ bởi enzym ADN nuclease làm nhiệm vụ sửa chữa (DNA repair nuclease). Liên kết phosphodiester giữa nucleotide đột biến và nucleotide bình thường bị cắt đứt tạo ra một khoảng trống trên sợi ADN.

Bước 2: Enzym ADN polymerase nhận biết gốc 3'-OH ở khoảng trống và tổng hợp một đoạn nucleotide thay thế tại vị trí đó theo nguyên tắc tạo cặp bổ sung.

Bước 3: Enzym ADN ligase nối sợi ADN với đoạn vừa tổng hợp tạo sợi ADN hoàn chỉnh giống hệt sợi trước khi bị đột biến.



Hình 3.1: Sửa chữa ADN theo cơ chế cắt bỏ sai hỏng gồm 3 bước cơ bản. Bước 1: cắt bỏ đoạn ADN hỏng. Bước 2: tổng hợp đoạn mới.

Bước 3: nối đoạn mới vào sợi ADN. Các enzym tham gia vào quá trình này gồm ADN nuclease, ADN polymerase và ADN ligase.

3.1. CẮT BỎ ĐOẠN ADN BỊ SAI HỎNG

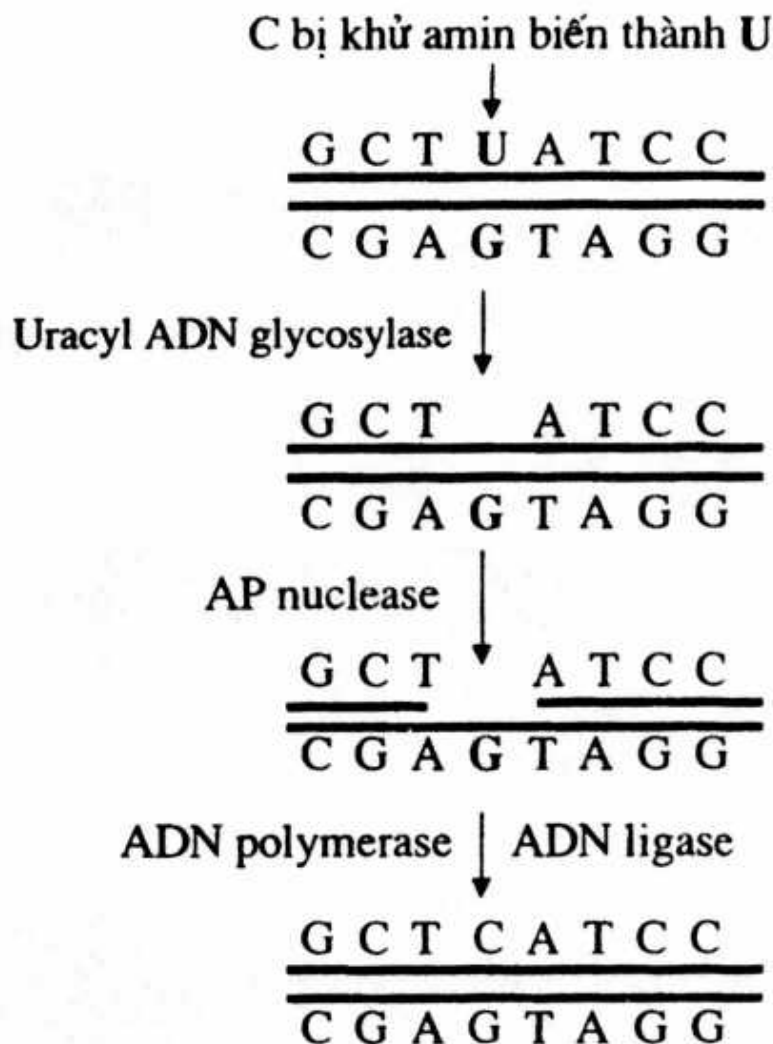
Tuỳ thuộc vào loại sai hỏng mà đoạn ADN được sửa chữa bằng cách loại bỏ các base (base excision repair-BER) hoặc loại bỏ nucleotide (nucleotide excision repair-NER). Trong cơ chế BER, một số dạng sai hỏng ADN hoặc các base không thích hợp bị loại bỏ bởi các enzym DNA glycosylases. Enzym làm nhiệm vụ phân hủy liên kết N-glycosidic giữa base và phân tử đường. Khi đường deoxyribose của ADN bị mất base purine (APurinic) hoặc bị mất base pyrimidin (APyrimidic) thì chúng được nhận biết bởi enzym đặc hiệu có tên gọi AP endonuclease. Ví dụ như đột biến khử purine xảy ra với tần số cao nhất làm phân tử đường bị mất liên kết với base purine. AP endonuclease cắt liên kết phosphodiester ở vị trí 5' của đường để loại bỏ đường ra khỏi

ADN. Sau đó ADN polymerase và ADN ligase tham gia sửa chữa vị trí bị cắt bỏ đó.

3.1.1. Sửa chữa ADN bằng việc loại bỏ base hỏng (BER).

Các base sai hỏng được nhận biết và cắt ra khỏi phân tử đường bởi ADN glycosylase. Mỗi loại enzym ADN glycosylase nhận biết một loại base đột biến và cắt base đó ra khỏi ADN bởi phản ứng phân hủy cầu nối *N-glycosidic* giữa đường và base. Hiện nay đã phát hiện được 6 loại ADN glycosylase gồm các enzym nhận biết và cắt bỏ các base bị khử amin, các base có vòng liên kết bị hở hay có liên kết đôi C=C bị đột biến thành liên kết đơn C-C v.v... Khi ADN glycosylase cắt bỏ base hỏng, phân tử đường được bộc lộ tạo thành vị trí có tên gọi là AP. Vị trí này được nhận biết bởi AP endonuclease. Sau đó ADN được sửa chữa nhờ ADN polymerase và ADN ligase.

Ví dụ về hoạt động của Uracil ADN glycosylase loại bỏ uracil khỏi liên kết sai lệch U-G được mô tả trên Hình 3.2. Tuy nhiên enzym này không nhận biết được các cytosine bị methyl hoá ở vị trí C₅. Các cytosine này khi bị khử amin thì trở thành Thymine (chứ không phải Uracil). Do đó xảy ra liên kết sai T-G. Các T đột biến này không được nhận biết bởi Uracil ADN glycosylase hay bất kỳ enzym nào. Do đó trong quá trình nhân đôi ADN, sợi khuôn có T đột biến sẽ tạo nên liên kết bổ sung với A trên sợi mới. Vì vậy đột biến được truyền cho các thế hệ sau. Cytosine có mang nhóm 5'-methyl chính là một trong các điểm nóng (hot spots) dễ tạo ra các đột biến tự phát.



Phân tử ADN hoàn chỉnh không sai hỏng

Hình 3.2: Cơ chế loại bỏ base hỏng bởi ADN glycosylase.

Enzym uracil ADN glycosylase nhận biết và cắt U khỏi liên kết U-G (lúc này ta có hiện tượng cytosine bị khử amin). Đường deoxyribose thiếu liên kết với base được nhận biết bởi AP endonuclease và được sửa chữa.

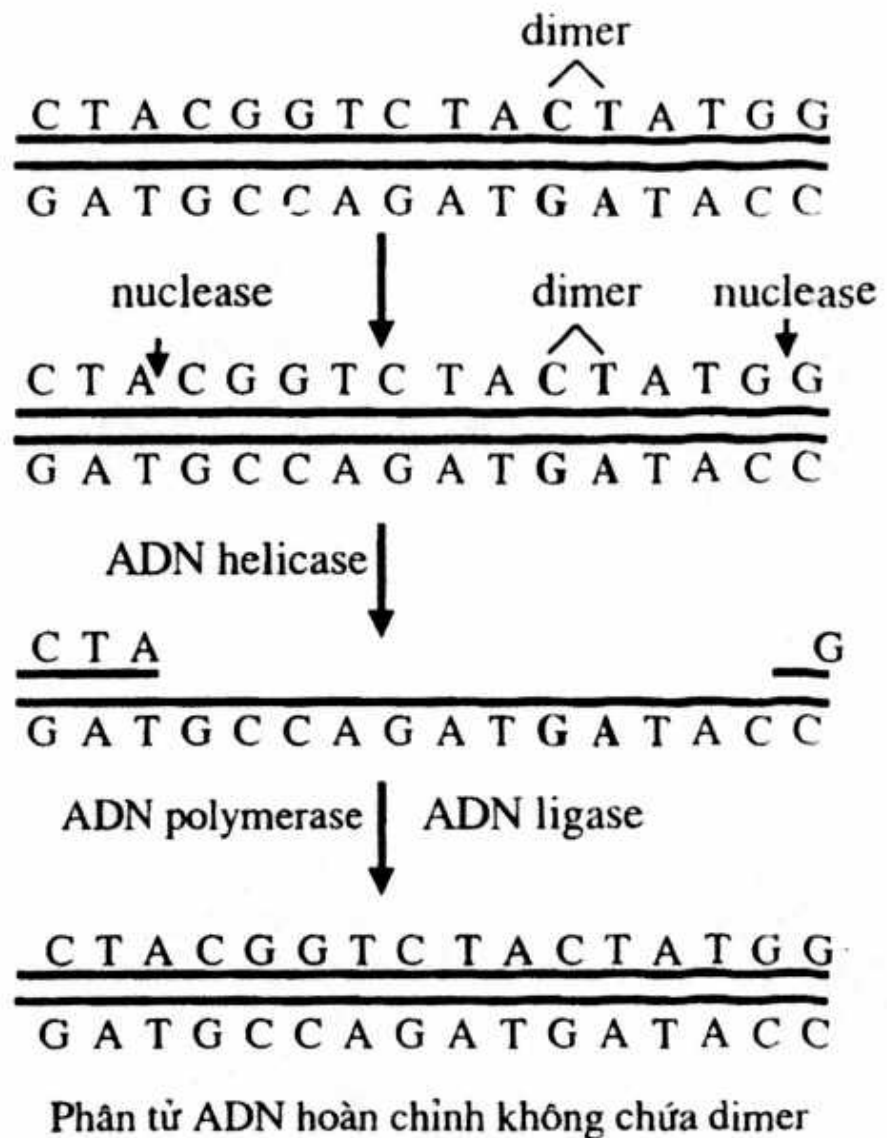
3.1.2. Sửa chữa ADN loại bỏ nucleotide hỏng (NER)

Cơ chế này cho phép loại bỏ và sửa chữa bất kỳ sai hỏng nào của ADN bao gồm các liên kết dimer T-T, T-C hoặc C-C, thậm chí cả đoạn gồm nhiều nucleotide không đúng. Ở đây có sự tham gia của nhiều enzym để kiểm soát ADN tìm ra sai hỏng. Một khi các lỗi được phát hiện, enzym ADN helicase sẽ cắt hai

phía của sợi đơn mang lỗi (khác với cơ chế nêu trong 3.1.1. chỉ cắt từng base riêng biệt). Đoạn này sau đó được tổng hợp mới nhờ ADN polymerase và ADN ligase (Hình 3.3).

Ở vi khuẩn, có 3 enzym chính liên quan đến cơ chế sửa chữa NER là UvrA, UvrB và UvrC. Chúng có thể kết hợp với nhau tạo phức có hoạt tính endonuclease phụ thuộc vào ATP (ATP-dependent endonuclease). Đầu tiên, hai phân tử UvrA liên kết với một UvrB tạo phức tương tác với ADN sai hỏng. Tiếp đến năng lượng do ATP bị phân hủy làm cho UvrA tách ra khỏi ADN. Như vậy sự có mặt của UvrA nhằm giúp cho UvrB có khả năng bám dính vào ADN. Sau khi UvrA tách ra, UvrC tương tác với UvrB và với ADN làm cho cấu trúc không gian của ADN bị thay đổi. Nhờ đó UvrB cắt sợi ADN ở phía đầu 3' của vị trí sai hỏng. Tiếp đến, UvrC cắt ở phía đầu 5', thường cách chỗ hỏng khoảng 8 nucleotide. Như vậy phức UvrABC cắt sợi đơn ở hai phía, tạo ra một đoạn gồm 12 nucleotide. Đoạn này bị loại ra khỏi cấu trúc ADN bởi UvrD có hoạt tính helicase. Vết hở thiếu 12 base trên sợi đơn của phân tử ADN được sửa chữa bởi ADN polymerase và được nối liền bởi ADN ligase.

Phức endonuclease UvrABC cùng phối hợp với một số protein mã bởi các gen thuộc cơ chế SOS có khả năng cắt và thay thế đoạn ADN chứa lỗi dài 12 nt đến 2000 nt. Khi hệ gen bị tổn thương, các gen đang phiên mã được ưu tiên sửa chữa trước. Phức protein mã bởi các gen *mfd* có khả năng liên kết với phức ARN polymerase/ADN khuôn (bị hỏng)/ARN giúp cho phức này vượt qua vị trí hỏng. Ngoài ra protein Mfd có thể tương tác với UvrA, thúc đẩy phức UvrA-UvrB liên kết với vị trí ADN hỏng.



Cơ chế sửa chữa NER ở sinh vật eukaryot xảy ra tương tự như ở vi khuẩn. Đoạn ADN bị cắt thường dài 30 nucleotide và rất hiếm trường hợp đạt đến 1500. Trong khi đó ở *E.coli*, đoạn bị cắt thay đổi từ 12 đến 2000 nt.

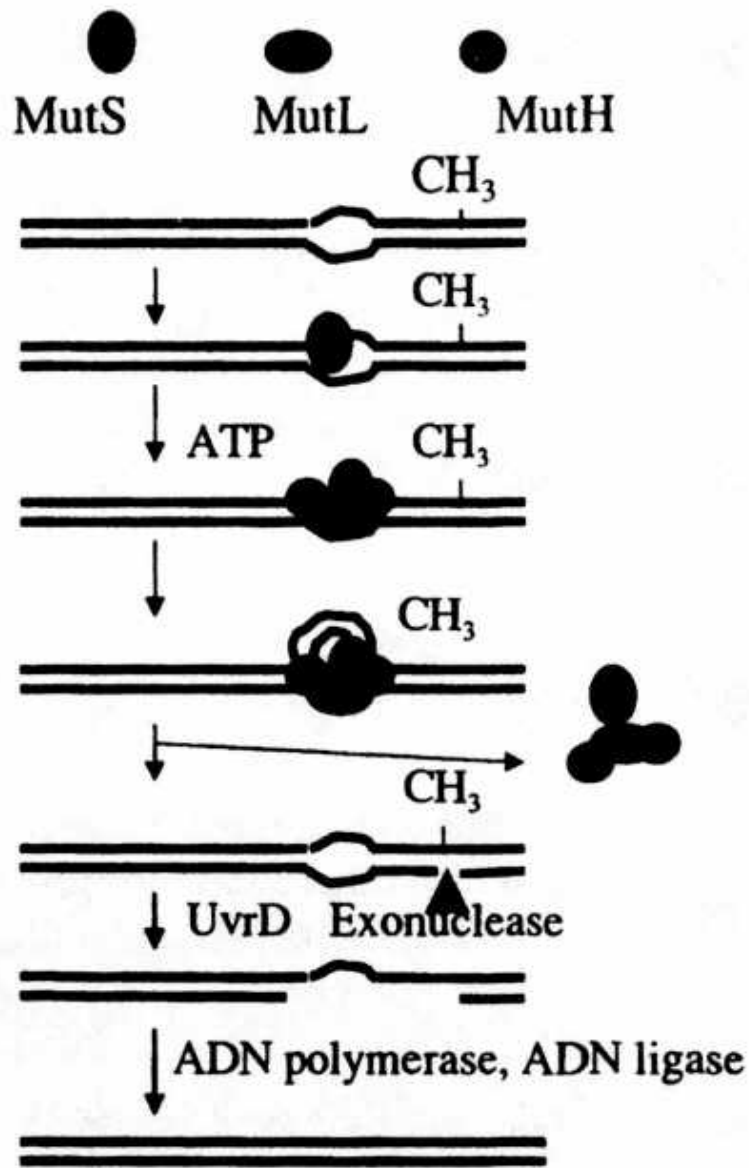
3.2. SỬA CHỮA ADN GHÉP ĐÔI LỆCH (*MISMATCH REPAIR*)

Sự ghép đôi lệch giữa các base trên hai sợi đơn hoặc sự chèn vào bớt đi của một vài nucleotide vào một sợi đơn làm thay đổi cấu trúc không gian của phân tử ADN. Những sai lệch này có thể do nucleotide không thích hợp được ghép vào sợi ADN trong quá trình tái bản ADN hay trao đổi chéo. Ngoài ra, sự thêm bớt một vài nucleotide có thể do sự trượt của sợi khuôn cũng như sợi đang tổng hợp ở những đoạn ADN lắp lại trong phản ứng tổng hợp ADN. Tuy nhiên ở đây không có các base hoặc các nucleotide sai hỏng như trong hai cơ chế BER và NER. Do đó, cơ chế sửa chữa ghép đôi lệch khác với hai cơ chế trên và phải phân biệt được sợi đơn đúng với sợi đơn bị thay thế hoặc bị thêm bớt nucleotide.

Trong quá trình tái bản ADN, nucleotide trên sợi mới bị ghép cặp sai lệch với sợi khuôn được phát hiện do sợi đang được tổng hợp chưa bị methyl hoá. Do đó có thể nhận biết và phân biệt chúng với sợi khuôn. Ở vi khuẩn *E.coli*, adenine A trong chuỗi GATC bị methyl hoá tại vị trí N⁶ bởi enzym Dam methylase (DNA adenine methylase). Khoảng thời gian sợi ADN mới vừa được tổng hợp nhưng chưa bị methyl hoá đủ để cơ chế sửa chữa ghép đôi lệch nhận biết sai hỏng. Trong trường hợp lỗi tạo cặp do trao đổi chéo gây ra (các dị ghép đôi-heteroduplex) khi cả hai sợi đơn đều chứa nhóm methyl thì sửa chữa xảy ra với hiệu suất thấp và tính chính xác không cao.

Sản phẩm của bốn gen *mutH*, *mutL*, *mutS* và *uvrD* không thể thiếu được trong cơ chế sửa chữa ghép đôi lệch. Chúng được gọi là các gen *mutator* do đột biến ở những gen này tác động đến tính bền vững chính xác của quá trình tái bản ADN. Protein MutS nhận biết vị trí ghép đôi lệch. Khả năng này phụ thuộc vào loại sai lệch, ví dụ G:T và A:C dễ nhận biết hơn G:G. Ngoài ra, protein MutH có hoạt tính endonuclease phân cắt đầu 5' ngay tại vị trí G của chuỗi GATC không chứa nhóm methyl. Enzym này yêu cầu MutS, MutL, ATP và một số cation hoá trị 2

để đảm bảo hoạt tính phân cắt trong khi UvrD (helicase) cần thiết để loại bỏ đoạn chứa nucleotide không tạo cặp. UvrD là loại 3' → 5' helicase nên exonuclease loại bỏ đoạn bị lỗi nhất định phải có hoạt tính cắt bỏ các nucleotide theo chiều 3' → 5'. Cơ chế sửa chữa ghép đôi lệch được minh họa trên Hình 3.4.



Sợi đúp ADN không còn ghép đôi lệch

Hình 3.4: Cơ chế sửa chữa ghép lệch đôi trong *E.coli*. Protein MutS nhận biết vị trí lỗi và phức protein MutS/MutH/MutL tương tác với sợi đúp ADN tạo cấu trúc vòng tại đoạn có lỗi. MutH cắt sợi đơn tại vị trí GATC và UvrD mở xoắn. Đoạn mở xoắn chứa lỗi bị phân hủy bởi endonuclease. Khoảng trống được sửa chữa phục hồi nhờ ADN polymerase và ADN ligase.

3.3. SỬA CHỮA ADN THEO CƠ CHẾ "ERROR-PRONE"

Khi phân tử ADN chịu các đột biến, một số protein đặc biệt làm nhiệm vụ dừng quá trình tổng hợp ADN để sửa chữa theo các cơ chế khác nhau. Một điều dễ hiểu là nếu như tế bào tiếp tục phân chia (tức là ADN tiếp tục được nhân lên) thì tỷ lệ đột biến cao khiến số tế bào sống sót giảm.

Phân tử ADN có thể phải chịu những thiệt hại sai hỏng nặng nề như khi bị chiếu tia tử ngoại (UV) hoặc khi chịu tác dụng của các chất gây ung thư (carcinogens). Lúc đó không phải chỉ một sợi đơn ADN bị hỏng mà cả hai sợi bị đứt gãy mất đi một số nucleotide. Thông tin di truyền của đoạn hỏng bị mất hoàn toàn nên không có sợi khuôn để sửa chữa theo các cơ chế thông thường. Trong trường hợp này, tế bào còn một giải pháp duy nhất để tăng khả năng sống sót: đó là sửa chữa ADN một cách ngẫu nhiên với tỷ lệ sai sót, đột biến cao nhưng dù sao vẫn duy trì được sự sống. Đó chính là cơ chế sửa chữa ADN "Error-Prone". Rõ ràng rằng chấp nhận các nucleotide sai đưa vào phân tử ADN để quá trình nhân đôi ADN vẫn được duy trì còn hơn là bị ngừng hoàn toàn. Như vậy sửa chữa ADN theo cơ chế này cũng chính là một trong những nguyên nhân gây nên tính đa dạng của ADN.

Bằng các thí nghiệm di truyền, vai trò quan trọng của hai protein UmuC và UmuD trong cơ chế sửa chữa "Error-Prone" được phát hiện. Chúng có khả năng gắn các nucleotide vào sợi ADN bất chấp không tồn tại sợi khuôn. Như vậy hai protein này có thể ức chế hoạt tính tự sửa exonuclease 3'-->5' của ADN polymerase hoặc chính chúng là một loại ADN polymerase đặc biệt. Khi đột biến xảy ra trên gen mã cho hai protein đó, các tế

bào vi khuẩn sống sót ít hơn các tế bào bình thường dưới tác dụng của tia tử ngoại .

Ngoài ra khi cả hai sợi đơn ADN đều chịu đột biến nhưng nếu trong genome còn có ít nhất một copy giống hệt đoạn bị hỏng, lúc đó đoạn hỏng có thể được phục chế nhờ cơ chế trao đổi chéo dùng bản copy làm khuôn.

3.4. TỔNG HỢP CÁC ENZYM SỬA CHỮA ADN

Các enzym tham gia sửa chữa ADN được tổng hợp ngay sau khi đột biến xảy ra cho thấy tính chính xác của thông tin di truyền luôn được bảo toàn để duy trì hoạt động sống của tế bào. Cơ chế trả lời các tín hiệu cấp cứu (**SOS response**) ở *E.coli* được nghiên cứu khá chi tiết. Ở loài vi khuẩn này quá trình tổng hợp ADN bị dừng do sai hỏng trên phân tử khuôn đều tạo ra các tín hiệu làm tăng cường hoạt động của hơn 15 gen bao gồm các gen tham gia sửa chữa ADN. Đầu tiên các tín hiệu hoạt hoá protein RecA. Ở dạng hoạt động, RecA phân huỷ protein ức chế các gen thuộc cơ chế SOS, giúp cho quá trình tổng hợp ARNm trên các gen này được tăng cường. Các protein được tổng hợp có hai tác dụng. Thứ nhất, tăng số lượng các enzym cần cho quá trình sửa chữa ADN nên số tế bào sống sót tăng. Thứ hai, khi nồng độ một số protein thuộc cơ chế SOS tăng tạm thời sẽ kéo theo tỷ lệ đột biến tăng (do bản thân các gen này cũng chịu các đột biến). Điều này không có lợi cho trước mắt. Tuy nhiên trong lâu dài nó tạo ra những thuận lợi làm tăng tính đa dạng di truyền trong quần thể các tế bào vi khuẩn. Nhờ đó cơ hội tạo ra các tế bào mang đột biến nhưng khoẻ mạnh và thích ứng với môi trường cũng tăng theo.

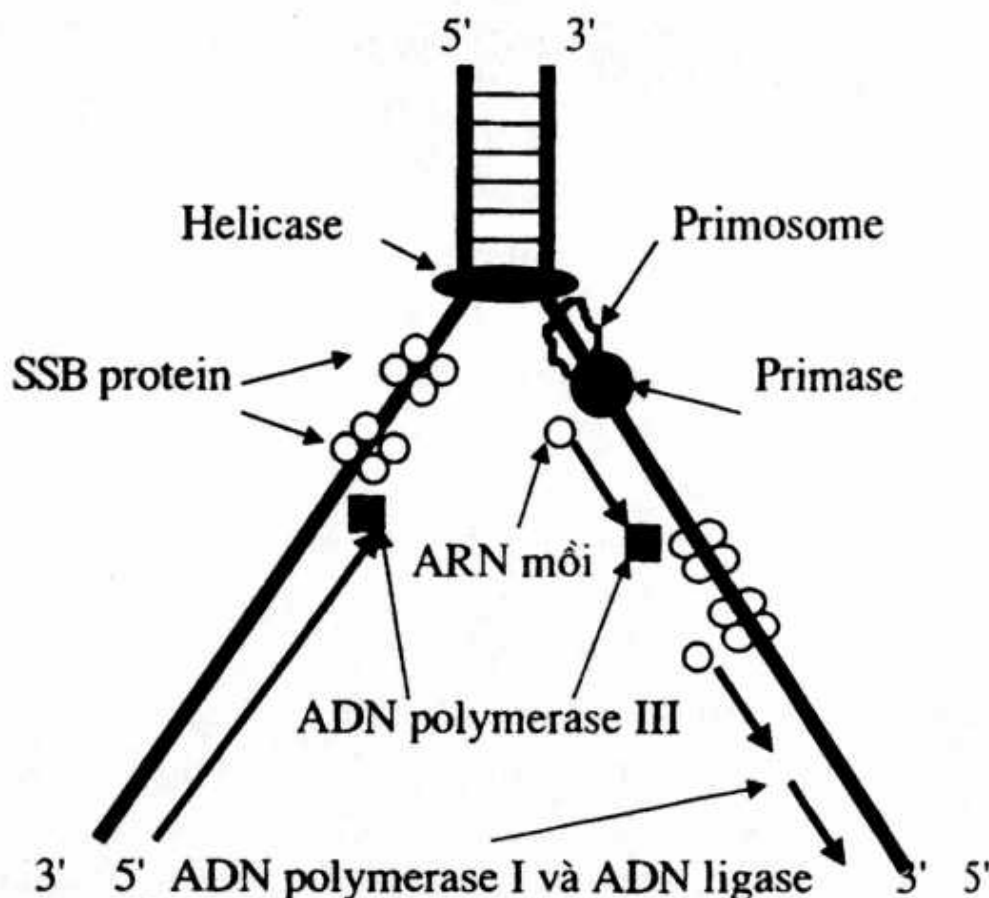
3.5. TÁI BẢN ADN

Phân chia tế bào ở mọi cơ thể từ prokaryot đến eukaryot đều đòi hỏi quá trình nhân đôi ADN xảy ra chính xác. Tốc độ tái bản đạt khoảng 500 nucleotide/giây ở vi khuẩn và 50 nucleotide/giây ở động vật. Có nhiều enzym và protein tham gia quá trình này tạo thành bộ máy tổng hợp ADN, trong đó enzym ADN polymerase đóng vai trò quan trọng nhất.

ADN được tái bản theo cơ chế bán bảo toàn (sau quá trình tái bản, mỗi phân tử ADN gồm một sợi đơn cũ và một sợi mới). Đầu tiên liên kết giữa hai sợi đơn của ADN khuôn bị bẻ gãy, sợi kép mở xoắn tạo chạc sao chép có cấu trúc hình chữ Y (**replication fork**). ADN polymerase không có khả năng bắt đầu phản ứng từ các nucleotide riêng lẻ mà đòi hỏi phải có mỗi oligonucleotide. Enzym làm nhiệm vụ xúc tác phản ứng gắn nucleotide vào đầu 3' của mỗi để tổng hợp sợi mới theo nguyên tắc tạo cặp bổ sung với sợi khuôn. Do đó quá trình tái bản chỉ thực hiện theo một chiều nhất định từ 5'→3' trên sợi mới. Vì thế một sợi đơn được tổng hợp liên tục còn sợi kia tổng hợp một cách gián đoạn, gồm nhiều đoạn ngắn (đoạn Okazaki) nối với nhau bởi ADN ligase. Ở prokaryot, đoạn Okazaki dài khoảng 1000 đến 2000 nucleotide, còn ở sinh vật eukaryot đoạn này chỉ dài khoảng 100 đến 200 nucleotide (Hình 3.5).

Enzym ADN polymerase không có khả năng nối hai nucleotide riêng lẻ với nhau để bắt đầu tổng hợp một sợi ADN mới. Quá trình tổng hợp ADN đòi hỏi phải có mỗi. Đó là một đoạn oligonucleotide ngắn do ADN primase tổng hợp. Bản chất mỗi là một đoạn ARN. Enzym ADN polymerase thêm các nucleotide vào sợi mỗi để tổng hợp nên sợi đơn mới. Sau đó, mỗi ARN bị phân huỷ và đoạn nucleotide tại vị trí mỗi được hoàn thiện nhờ cơ chế sửa chữa ADN. Tuy nhiên trong trường hợp đối với các phân tử ADN dạng thẳng ở một số virus, quá trình tái bản ADN không cần mỗi ARN mà đòi hỏi phức protein-deoxycytidine triphosphate. Quá trình tổng hợp bắt đầu tại một

đầu của sợi đôi theo chiều $5' \rightarrow 3'$ và chỉ xảy ra trên một sợi đơn khuôn để tạo ra một phân tử ADN. Sau đó sợi khuôn thứ hai thay đổi cấu trúc và được dùng để tổng hợp phân tử ADN thứ hai. Cơ chế tổng hợp này vẫn tuân theo nguyên tắc bán bảo toàn.



Hình 3.5 : Mô hình chạc sao chép ADN

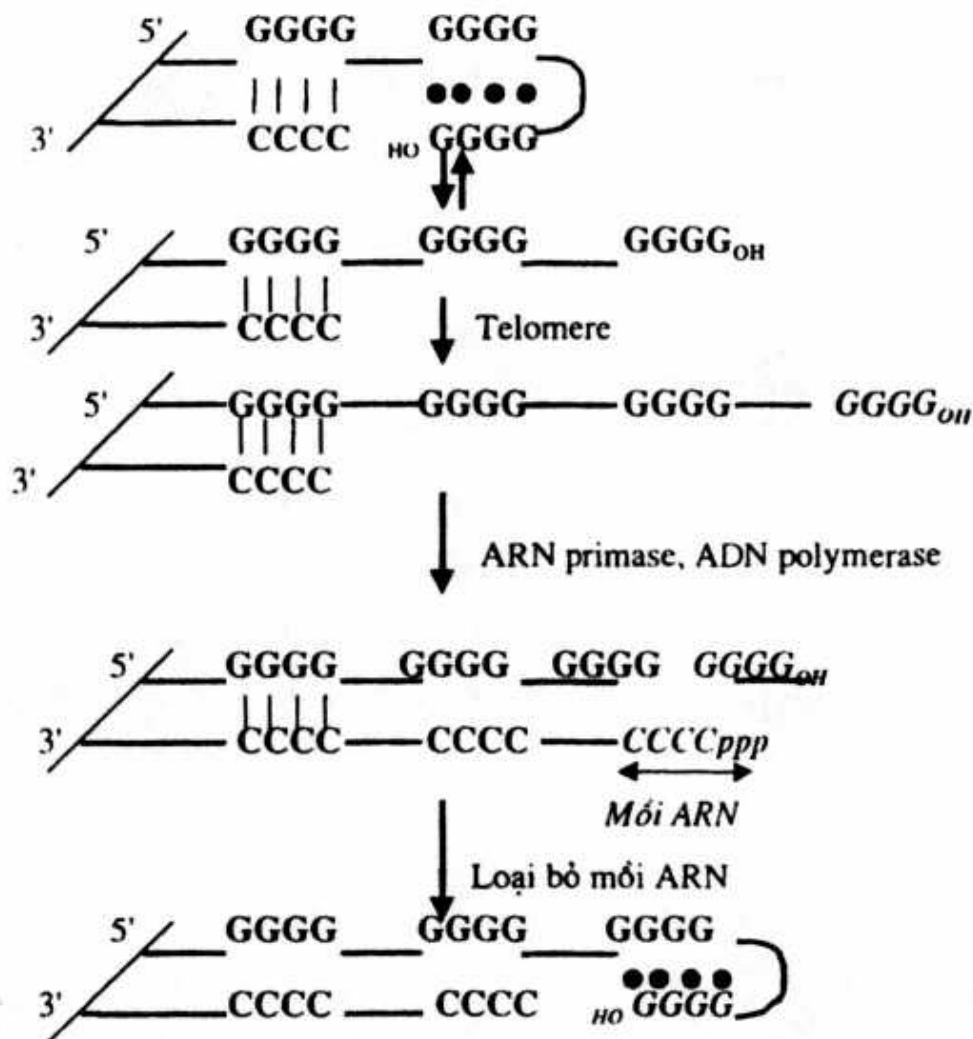
Hai sợi đơn ADN bắt đầu tách nhau ra tại vị trí đặc biệt gọi là tâm sao chép (replication origins). Đối với vi khuẩn, nấm men và một số virus ký sinh trên động vật, hệ gen của chúng thường chỉ có một tâm sao chép và dài khoảng 300 bp. Khi quá trình nhân đôi ADN bắt đầu, hai chạc sao chép hình thành tại tâm và tiến về hai hướng ngược nhau.

Đối với sinh vật eukaryot, trên mỗi nhiễm sắc thể có nhiều tâm sao chép. Các tâm này (20 đến 80) thường tập trung trên một đoạn ADN tạo nên đơn vị sao chép (replication units). Khoảng cách giữa các tâm sao chép trên một đơn vị thay đổi từ 30 kb đến 300 kb. Ngoài ra, không phải mọi đơn vị sao chép đều hoạt động cùng một lúc. Ngay trên cùng một nhiễm sắc thể, các

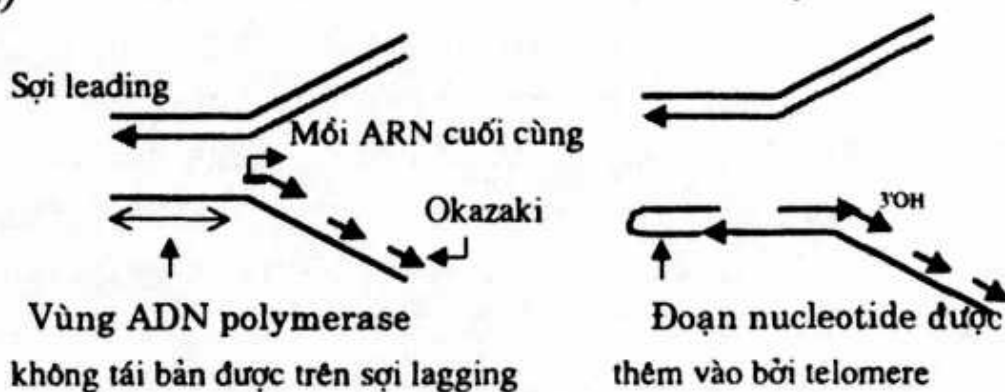
vùng khác nhau được tái bản ở các thời điểm khác nhau. Một điều lý thú là hoạt động của chúng không xảy ra ngẫu nhiên mà theo một trật tự nhất định. Trong phase S, các vùng ADN di nhiễm sắc được nhân đôi muộn trong khi vùng chứa gen được nhân đôi sớm. Hơn nữa, các gen hoạt động trong mọi loại mô tế bào (được gọi là các gen quản gia-"housekeeping genes") được nhân đôi rất sớm trong khi các gen chỉ hoạt động trong một số mô tế bào nhất định sẽ nhân đôi sớm trong tế bào đó nhưng nhân đôi rất muộn trong các tế bào mà gen không hoạt động.

Để tổng hợp được trọn vẹn các đầu telomere tận cùng của phân tử ADN dạng thẳng, enzym telomerase nhận biết được đoạn ngắn ADN giàu nucleotide G tại đầu telomere và kéo dài đầu đó theo hướng 5'→3' (như vậy đầu telomere dài thêm ra). Bản thân telomerase có chứa phân tử ARN được dùng làm khuôn để tổng hợp nên đoạn kéo dài đó. Sau đấy, đoạn mới này được dùng làm khuôn để ADN polymerase hoàn thiện các đầu tận cùng của sợi ADN mới.

Telomere của *Tetrahymena* đã được sử dụng thành công lần đầu tiên trong việc thiết kế plasmid ở dạng thẳng chứ không phải ở dạng khép kín, có khả năng tái bản trong tế bào nấm men. Trình tự nucleotide của telomere ở *Tetrahymena* có chứa chuỗi lặp lại $(T_2G_4)_n$ theo chiều 5'→3' từ trong ra ngoài đầu cùng của telomere. Đầu tận cùng của telomere không tồn tại ở dạng sợi đúp mà có một phần ở dạng sợi đơn. Phần sợi đơn này giàu nucleotide G có khả năng tạo liên kết G-G (không theo nguyên tắc tạo cặp bổ sung của Watson-Crick). Cấu trúc đặc biệt này được nhận biết bởi telomerase. Enzym này phân lập từ *Tetrahymena* có thể gắn thêm $(T_2G_4)_n$ vào telomere của *Tetrahymena*, *Oxytricha*, nấm men. Như vậy enzym giữ vai trò quyết định việc thêm các đoạn lặp lại vào đầu tận cùng của nhiễm sắc thể. Đoạn giàu nucleotide G ở telomere và hoạt động của telomerase trong việc kéo dài đầu này nhằm đảm bảo tính vẹn toàn đầu tận cùng của nhiễm sắc thể được minh họa trên Hình 3.6.



(B)



Hình 3.6: (A). Cấu trúc của telomere. Các nucleotide G ở đầu tận cùng có khả năng tạo cặp với nhau. Enzym telomerase kéo dài thêm đầu tận cùng (các nucleotide in nghiêng) của sợi giàu G. ARN primase và ADN polymerase tổng hợp đoạn mồi trên khuôn sợi giàu C. Cuối cùng, mồi ARN bị loại bỏ, các nucleotide trên sợi giàu G tạo cặp với nhau. Telomere được kéo dài ra. (B). Telomerase đảm bảo sự vẹn toàn của nhiễm sắc thể ở các đầu telomere. Enzym gắn thêm chuỗi lặp lại G tạo đầu 3'OH cho phép ADN polymerase tổng hợp đoạn cuối cùng trên khuôn mẫu sợi lagging trong quá trình tái bản ADN.

Đối với rất nhiều virus cũng như ở một số trường hợp khác, số lượng của gen trong genome được tăng lên rất nhiều tại các thời điểm đặc biệt. Phân tử ADN thường được nhân đôi theo cơ chế vòng tròn lăn (rolling circle mechanism). Quá trình tổng hợp sợi ADN mới được bắt đầu tại một vết cắt thuộc tâm sao chép trên một sợi đơn khuôn - sợi này được gọi là sợi (+). Tại vết cắt, đầu 5' được tháo dần và trở thành khuôn để tổng hợp các sợi Okazaki. Đầu 3' tại vết cắt được thêm các nucleotide mới tạo cặp bổ sung với các nucleotide trên sợi đơn kia. Nhờ đó hai phân tử ADN được tổng hợp và luôn tuân theo cơ chế bán bảo toàn.

3.5.1. Đảm bảo tổng hợp ADN chính xác bởi cơ chế "đọc sửa" (Proofreading)

Sai sót xảy ra trong quá trình tổng hợp ADN chỉ khoảng 10^{-9} (cứ 10^9 nucleotide được tái bản thì có một sai sót). Tần số này nhỏ hơn rất nhiều tần số tạo liên kết sai lệch giữa các base. Ví dụ như liên kết giữa T-G hay C-A có tần số khoảng 10^{-5} . Đó là do các nucleotide khi gắn vào sợi ADN đang tổng hợp phải chịu sự kiểm tra theo cơ chế "đọc sửa". Nhờ đó bất kỳ một nucleotide nào tạo liên kết không đúng nguyên tắc bổ sung đều bị phát hiện và loại bỏ.

Cơ chế kiểm tra "đọc sửa" phụ thuộc vào tính chất đặc biệt của enzym ADN polymerase. Khác với ARN polymerase, ADN polymerase không có khả năng nối hai nucleotide với nhau để bắt đầu tổng hợp sợi mới. Enzym này đòi hỏi phải có một đoạn nucleotide ngắn có đầu tận cùng mang nhóm OH ở vị trí 3'.

Đoạn này được gọi là sợi mồi (primer strand). Các nucleotide tạo cặp theo nguyên tắc bổ sung với các base trên sợi khuôn được gắn vào vị trí 3' nhờ liên kết phosphodiester giữa nhóm OH ở vị trí 3' và gốc phosphate ở vị trí 5' của hai nucleotide liên tiếp nhau. Phản ứng này do ADN polymerase đảm nhiệm. Mọi sai sót không tuân theo nguyên tắc bổ sung đều bị enzym phát hiện và loại ra. Tính chất exonuclease này của enzym chỉ xảy ra theo hướng 3'--->5' trên sợi đang tổng hợp giúp cho enzym có khả năng tự sửa sai (self correcting), do đó tần số đột biến rất nhỏ.

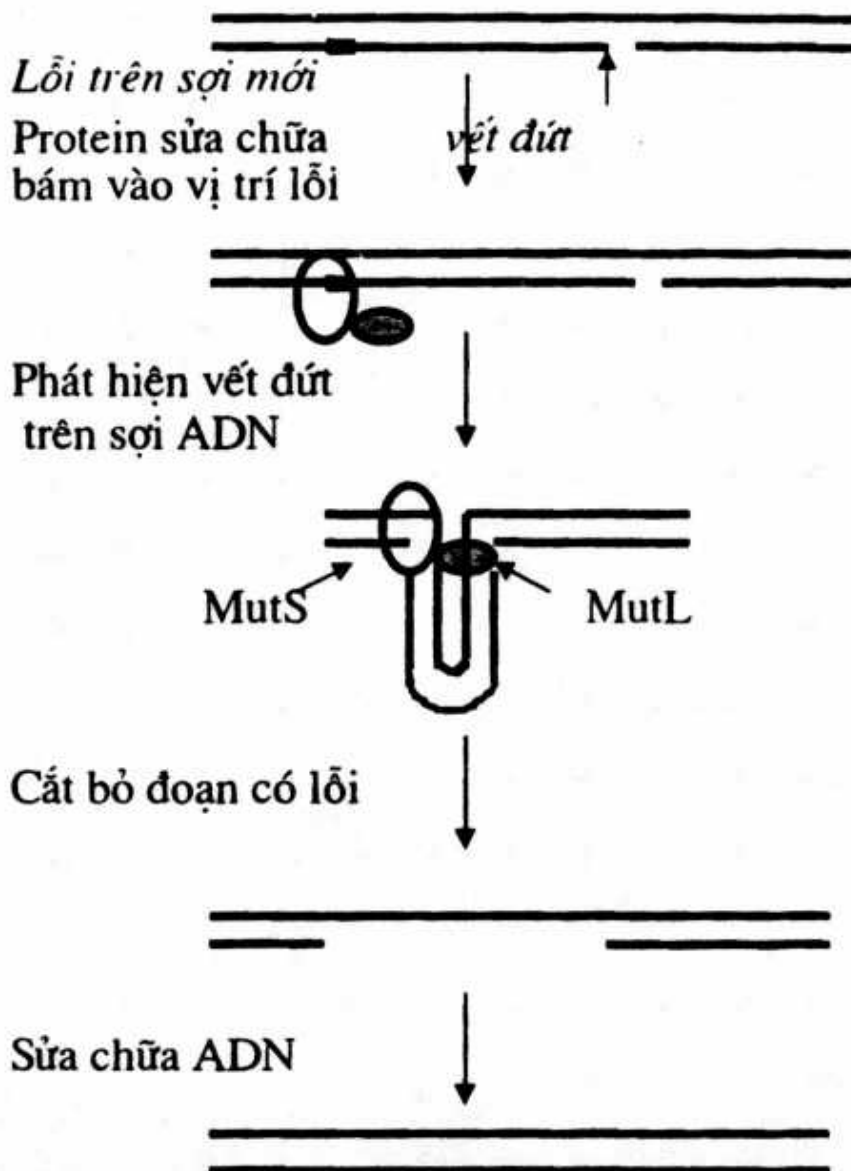
Liên kết theo nguyên tắc bổ sung của nucleotide cuối cùng tại đầu 3' của mồi với khuôn đóng vai trò quyết định giúp enzym tự sửa các sai hỏng. Ngược lại, enzym ARN polymerase xúc tác quá trình tổng hợp ARNm không có tính chất tự "đọc sửa". Điều may mắn là các sai sót trong phân tử ARNm không thể chuyển sang các thế hệ sau. Hơn nữa sai sót ngẫu nhiên trong phân tử ARN không có vai trò quyết định các hậu quả di truyền. ARN polymerase có thể tổng hợp nhiều phân tử ARNm từ một gen mà không cần đoạn oligo mồi ban đầu. Tần số đột biến tự phát xảy ra trong cả hai quá trình tổng hợp ARNm và tổng hợp protein khoảng 10^{-4} .

Ở vi khuẩn *E.coli*, một số đột biến làm tăng mạnh tần số biến đổi tự phát các nucleotide trên sợi ADN mới. Đó là các đột biến xảy ra ở các gen qui định tính chất đọc sửa exonuclease 3'--->5' của DNA polymerase. Vì vậy rất nhiều sai sót xảy ra mà không bị phát hiện để sửa chữa theo cơ chế đọc sửa của enzym. Tuy nhiên chúng có thể được sửa theo một cơ chế khác gọi là cơ

chế độ sửa các liên kết tạo cặp sai (mismatch proofreading hoặc mismatch repair). Cơ chế này khác với cơ chế sửa chữa ADN (DNA repair) vì nó không phụ thuộc vào sự nhận biết nucleotide lạ mà dựa vào sự méo mó cấu trúc sợi đúp do các liên kết tạo cặp sai giữa các base không bổ sung với nhau. Điều đặc biệt lý thú là tính chất đọc sửa theo cơ chế thứ hai chỉ cắt bỏ nucleotide không bổ sung ở sợi mới mà không động chạm đến nucleotide trên sợi khuôn (tức là nhận biết được nucleotide ghép đôi sai lệch trên sợi vừa tổng hợp).

Nghiên cứu cho thấy mọi base A trong dãy nucleotide GATC của ADN *E.coli* đều được gắn nhóm methyl. Chỉ có các base A trên đoạn ADN vừa tổng hợp chưa kịp gắn nhóm này. Nhờ đó cơ chế đọc sửa tạo cặp sai phân biệt được sợi mới và sợi khuôn. Nói cách khác cơ chế này hoạt động ngay khi quá trình tái bản đang xảy ra.

Gần đây một số protein có vai trò tương tự các protein tham gia cơ chế đọc sửa tạo cặp sai ở vi khuẩn được phát hiện ở sinh vật eukaryot. Ở nấm men, khi các gen mã cho những protein này bị đột biến, tỷ lệ sai sót trong tái bản ADN có thể tăng hơn 100 lần. Tuy nhiên các base A của ADN eukaryot không chứa gốc methyl, do đó phân biệt giữa sợi mới và sợi khuôn có lẽ dựa vào các vết đứt gãy hay có ở sợi vừa được tổng hợp. Vì sao xuất hiện các vết này là một vấn đề cần làm sáng tỏ. Tuy nhiên thực nghiệm cho thấy chúng giống như các tín hiệu giúp cho cơ chế đọc sửa nhận ra sợi ADN mới (Hình 3.7).

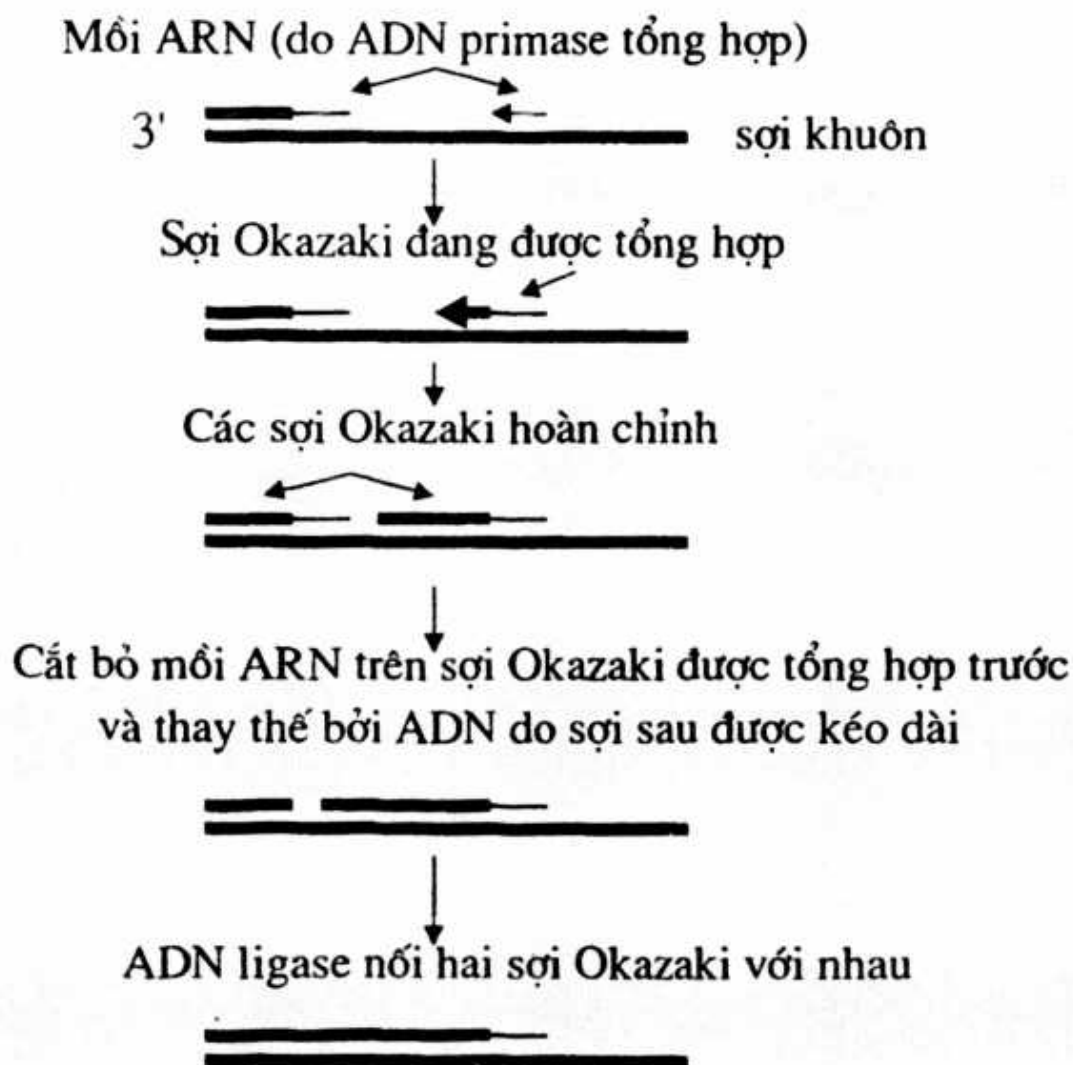


Hình 3.7: Cơ chế đọc sửa sự tạo cặp sai ở ADN trong tế bào eukaryot. Các phản ứng này được phát hiện trong quá trình tái bản invitro (cell-free system). Protein MutS liên kết đặc hiệu vào base tạo cặp sai, protein MutL tìm dò vết đứt gãy nằm gần base đó và gây đứt gãy đoạn nucleotide đơn nằm giữa base tạo cặp sai và vết đứt. Đoạn nucleotide mới được tổng hợp thay thế theo cơ chế sửa chữa ADN (DNA repair).

3.5.2. Oligo mới trong quá trình tổng hợp ADN

Do ADN polymerase chỉ tổng hợp sợi ADN theo một chiều duy nhất từ 5' sang 3' nên một sợi ADN được tổng hợp liên tục còn sợi kia do nhiều đoạn Okazaki được tổng hợp rồi nối với nhau. Khoảng cách giữa hai sợi Okazaki khoảng 2000 base và thời gian trung bình để tổng hợp một sợi chiếm khoảng 4 giây.

Tổng hợp bất cứ đoạn Okazaki nào cũng đòi hỏi oligo mỗi. Bản chất của mỗi này là các ribonucleoside trisphosphate được nối với nhau nhờ enzym ADN primase. Do đó các oligo này còn được gọi là các mồi ARN (RNA primer). Chúng dài khoảng 10 nucleotide. Chúng bị phân hủy sau khi đoạn Okazaki được tổng hợp xong. Cơ chế sửa chữa ADN sẽ thay thế chúng bởi các deoxyribonucleotides theo nguyên tắc tạo cặp bổ sung. Cuối cùng ADN ligase sẽ nối các đoạn Okazaki với nhau tạo sợi ADN mới (Hình 3.8).



Hình 3.3: Tổng hợp sợi Okazaki trong quá trình tái bản ADN. Ở sinh vật eukaryot, các mồi ARN dài khoảng 10 base được tổng hợp bởi enzym ADN primase. Khoảng cách giữa các mồi chừng 200 base. ADN polymerase tổng hợp đoạn Okazaki bằng cách gắn các nucleotide vào mồi ARN. Các mồi này sau đó sẽ bị phân hủy và cơ chế sửa chữa ADN sẽ tổng hợp đoạn nucleotide thay thế mồi. Enzym ADN ligase nối các đoạn Okazaki với nhau tạo sợi ADN mới.

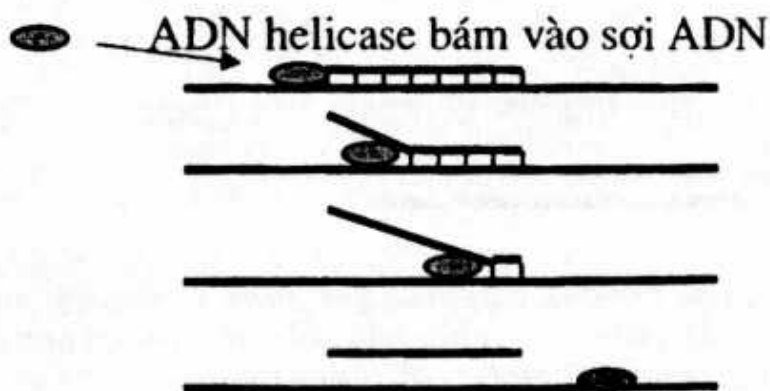
Vì sao mỗi ADN không được tổng hợp để giảm bớt các bước tổng hợp mỗi ARN, phân hủy và sửa chữa thay thế mỗi này trong quá trình tái bản ADN ? Một điều xuất phát từ thực tế cho thấy ADN polymerase có đặc tính tự sửa chữa thì không thể bắt đầu tổng hợp sợi mới từ các nucleotide riêng biệt. Trong khi đó ARN polymerase không cần mẫu thì cũng không có khả năng tự sửa chữa. Do đó những mẫu ARN được xem là những đoạn hỏng cần loại bỏ.

3.5.3. Các protein tham gia quá trình tổng hợp ADN

Để trở thành khuôn cho phản ứng nhân đôi ADN, phân tử ADN dạng đúp nhất thiết phải mở xoắn, tức là liên kết bổ sung giữa các base bị đứt gãy và hai sợi đơn được tách nhau. Mở xoắn và giữ hai sợi đơn ở dạng tự do không có cấu trúc bậc hai được thực hiện nhờ một số enzym, protein đặc hiệu.

3.5.3.1. ADN helicase

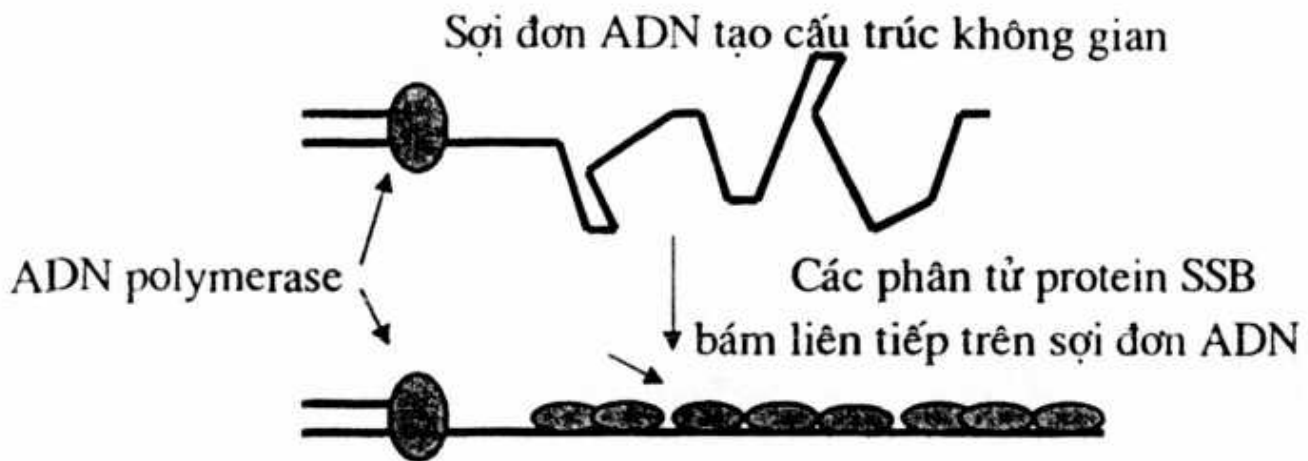
ADN helicase là protein có khả năng phân hủy ATP một khi đã bám vào sợi đơn ADN. Nhờ phản ứng này ADN helicase có thể di chuyển dọc theo chiều dài sợi đơn làm sợi đôi phải tiếp tục mở xoắn tách thành hai sợi đơn (Hình 3.9). Thông thường, tại chạc sao chép chữ Y có hai phân tử ADN helicase di chuyển trên hai sợi đơn theo hai hướng ngược nhau. Chúng có thể thuộc hai loại helicase khác nhau do tính chất di chuyển ngược chiều này.



Hình 3.9: Thử nghiệm hoạt tính mở xoắn tạo hai sợi đơn ADN của enzym ADN helicase

3.5.3.2. SSB protein (Single Strand Binding Protein)

Chức năng của các protein này là giữ cho sợi đơn ADN ở dạng thẳng không có khả năng tạo cặp bổ sung với sợi đơn khác hoặc không tạo liên kết ngay giữa các base của một sợi.



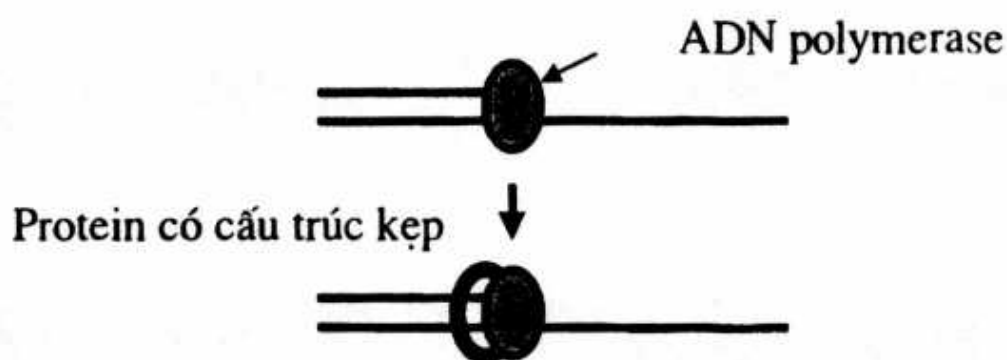
Hình 3.10: Vai trò của SSB protein trong việc giữ sợi đơn ADN không có cấu trúc không gian, tạo thuận lợi cho ADN polymerase di chuyển trên khuôn trong quá trình tổng hợp ADN.

Nhờ đó ADN polymerase di chuyển trên sợi đơn làm khuôn không gặp trở ngại gì trong quá trình tổng hợp sợi mới. Mặc dù các SSB protein không có khả năng mở xoắn nhưng chúng giúp cho helicase giữ được trạng thái mở xoắn và đặc biệt tránh được liên kết tạo cấu trúc vòng của sợi đơn khuôn khi tổng hợp các đoạn Okazaki (Hình 3.10).

3.5.3.3. Protein có cấu trúc vòng giữ ADN polymerase chuyển động trên sợi ADN

Hầu hết các phân tử ADN polymerase có xu hướng tổng hợp một đoạn ngắn các nucleotide rồi tách ra khỏi khuôn. Điều này có giúp cho enzym nhanh chóng rời khỏi đoạn Okazaki đã tổng hợp xong và quay vòng rất nhanh để tổng hợp đoạn tiếp theo ngay trên một sợi khuôn. Tuy nhiên, xu thế tách khỏi sợi

khuôn gây bất lợi cho enzyme khi tổng hợp sợi ADN dài liên tục. Để khắc phục điều này, phân tử enzyme được giữ lại trên sợi khuôn nhờ một cấu trúc protein đặc biệt giống như chiếc kẹp. Kẹp này cho phép enzyme bám vững chắc khi chuyển động trên sợi khuôn nhưng đồng thời đảm bảo cho enzyme rời khỏi khuôn nhanh chóng khi quá trình tổng hợp kết thúc (Hình 3.11).



Hình 3.11: Cấu trúc kẹp protein giữ ADN polymerase bám chắc vào sợi ADN.

3.5.3.4. Vai trò của ADN topoisomerase trong tái bản ADN

Do phân tử ADN có cấu trúc dạng xoắn nên trong quá trình nhân đôi cứ 10 base mở xoắn để tổng hợp tại chạc sao chép thì sợi đôi ở phía trước chạc lại bị xoắn 1 vòng. Như vậy khi chạc chuyển động dọc theo sợi khuôn sẽ khiến sợi này xoắn lại rất nhanh. Hiện tượng đó cũng xảy ra tương tự khi mở xoắn để tổng hợp ARN. Chính enzyme ADN topoisomerase đóng vai trò mở xoắn để khắc phục sự xoắn tít lại của phân tử ADN khuôn. ADN topoisomerase là một loại nuclease đặc biệt bẻ gãy liên kết phosphodiester giữa các base trên một sợi đơn tạo ra một vết đứt, giải phóng hai đầu sợi ở dạng tự do. Các đầu đó sẽ quay để mở xoắn. Sau đó liên kết phosphodiester lại được phục hồi và enzyme được tách khỏi phân tử ADN. Có hai loại ADN topoisomerase. Loại I chỉ tạo vết đứt trên một sợi đơn còn loại II tạo vết đứt trên cả hai sợi của phân tử ADN kép.