VÕ THỊ THƯƠNG LAN

Giáo trình
SINH HỌC PHÂN TỬ TẾ BÀO VÀ ỨNG DỤNG

NHÀ XUẤT BẢN GIÁO DỤC

VÕ THỊ THƯƠNG LAN

Giáo trình

SINH HỌC PHÂN TỬ TẾ BÀO VÀ ỨNG DỤNG

NHÀ XUẤT BẢN GIÁO DỤC

LỜI NÓI ĐẦU

Sinh học phân tử được hiểu là môn học đề cập đến những kiến thức và kỹ thuật cơ bản nhằm nghiên cứu các quá trình sinh học ở mức độ phân tử. Vì vậy, sinh học phân tử được hình thành và phát triển dựa trên kiến thức liên quan đến tất cả các chuyên môn của sinh học, đặc biệt là di truyền, tế bào, vi sinh và sinh hoá. Chính vì vai trò tạo ra mối liên quan khăng khít giữa nhiều chuyên ngành nên sinh học phân tử được ứng dụng rộng rãi trong các lĩnh vực nông lâm nghiệp, thuỷ sản, y tế và xã hội.

Sinh học phân tử là một lĩnh vực mới được học tập và nghiên cứu ở Việt Nam. Giáo trình và tài liệu tham khảo về lĩnh vực này bằng tiếng Việt là yêu cầu cần thiết ở nước ta. Để đáp ứng nhu cầu của độc giả, chúng tôi xuất bản cuốn sách "Sinh học phân tử tế bào và ứng dụng" nhằm cung cấp cho sinh viên và các nhà nghiên cứu sinh học một số kiến thức cơ bản của sinh học hiện đại ở mức độ phân tử. Cuốn sách này giới thiệu những quá trình quan trọng xảy ra trong tế bào, nhờ đó làm sáng tỏ sự giống và khác nhau trong cấu trúc của genome, cấu trúc của một gen ở tế bào prokaryot và eukaryot (chương 1), quá trình tái bản và sửa chữa genome (chương 2). Những cấu trúc đó liên quan đến các cách thức kiểm soát hoạt động của các gen ở giai đoạn phiên mã, sau phiên mã và dịch mã để tổng hợp protein (chương 3). Quá trình tổng hợp protein, những biến đổi cấu trúc protein và những cách thức để nhận biết và vận chuyển protein đặc hiệu đến những vị trí đích khác nhau trong tế bào, hoặc tiết ra bên ngoài được giới thiệu trong chương 4. Một số kỹ thuật cơ bản trong nghiên cứu sinh học phân tử được trình bày trong chương 5.

Lần đầu xuất bản, chắc chắn cuốn sách còn có những thiếu sót, Tác giả mong nhận được ý kiến đóng góp của độc giả để giáo trình ngày càng hoàn thiện. Mọi ý kiến xin gửi về: Công ty Cổ phần Sách Đại học - Dạy nghề, Nhà xuất bản Giáo dục, 25 Hàn Thuyên - Hà Nội.

TÁC GIẢ

MỤC LỤC

Lời nói đầu 3

Chương 1. CẤU TRÚC VÀ CHỨC NĂNG CỦA PHÂN TỬ ACID NUCLEIC

1.1. Cấu trúc nucleotide. Cấu trúc phân tử ADN, ARN 8
1.2. ADN là vật liệu di truyền 10
1.3. Cấu trúc nhiễm sắc thể 12
1.3.1. Nhiễm sắc thể vi khuẩn có cấu trúc hạch nhân 13
1.3.2. Nhiễm sắc thể eukaryot gồm các vùng dị nhiễm sắc và nhiễm sắc thực 14
1.3.3. Tâm động (centromere) của nhiễm sắc thể eukaryot 15
1.3.4. Đầu mút (telomere) của nhiễm sắc thể eukaryot 16
1.4. Genome (hệ gen) 17
1.4.1. Genome của tế bào prokaryot (tế bào nhân sơ) 20
1.4.2. Genome của tế bào eukaryot (tế bào nhân thực) 22
1.4.3. Methyl hoá ADN trong genome eukaryot 23
1.5. ADN trong các bào quan của tế bào eukaryot 25
1.5.1. ADN ty thể 27
1.5.2. ADN lục lạp 27
1.6. Khái niệm gen 28
1.6.1. Các gen trong genome vi khuẩn 31
1.6.2. Các gen trong genome virus 34
1.6.3. Các gen trong genome eukaryot 36
1.7. Phân loại gen 38
1.7.1. Các gen trong cùng một họ gen 38
1.7.2. Gen lặp đi lặp lại liên tục 40
1.7.3. Pseudogen (gen giả) 42
1.8. Thành phần ADN không chứa gen trong genome 43
1.8.1. Transposon trong genome vi khuẩn 45
1.8.2. Transposon trong genome eukaryot 46
1.8.3. T-ADN di chuyển từ genome prokaryot đến genome eukaryot 47
1.9. Sắp xếp lại genome 48
1.9.1. Thay đổi dạng giao phối ở nấm men 49
1.9.2. Thay đổi kháng nguyên bề mặt ở ký sinh trùng đơn bào 50
1.9.3. Nhân bản tạo các bản gen mới từ genome 52
1.10. Chuyển nhiễm ADN lạ (gen) vào tế bào eukaryot 52

Chương 2. TÁI BẢN VÀ SỬA CHỮA ADN

2.1. Replicon 55
2.2. ADN polymerase 57
2.3. Tổng hợp ADN 58
2.4. Đảm bảo tính chính xác trong tái bản ADN 62
2.5. Tâm tái bản (origin) 63
2.6. Kiểm soát khởi động tái bản tại tâm sao chép 64
2.7. Tái bản theo cơ chế vòng tròn lăn 65
2.8. Tái bản ADN ở virus 67
2.8.1. Tái bản ADN của thực khuẩn thể 67
2.8.2. Tái bản từ đầu mút phân tử ADN dạng thẳng 68

2.9. Tái bản telomere của phân tử ADN eukaryot....69
2.10. Tái bản ADN ty thể....70
2.11. Chu trình tế bào....72
2.12. Sửa chữa ADN....74
2.13. Cắt bỏ đoạn ADN bị sai hỏng....76
2.13.1. Sửa chữa ADN theo cơ chế loại bỏ base hỏng....77
2.13.2. Sửa chữa ADN theo cơ chế loại bỏ nucleotide hỏng....78
2.14. Sửa chữa ADN theo cơ chế "Error-Prone"....79
2.15. Sửa chữa ADN ghép đôi lệch....81
2.16. Sửa chữa ADN dựa vào trao đổi chéo....83
2.17. Sửa chữa ADN trong tế bào eukaryot....85

Chương 3. PHIÊN MÃ VÀ KIỂM SOÁT PHIÊN MÃ

3.1. Mã di truyền bộ ba....88
3.2. Promoter....90
3.3. Protein tham gia khởi động phiên mã....92
3.4. Kiểm soát khởi động phiên mã....94
3.4.1. Kiểm soát tiêu cực – vai trò của yếu tố kìm hãm....94
3.4.2. Kiểm soát tích cực – vai trò của yếu tố hoạt hoá....95
3.4.3. Kiểm soát khởi đầu phiên mã theo cơ chế suy giảm – vai trò của cấu trúc cặp tóc....97
3.5. Kiểm soát phiên mã trên gen eukaryot....98
3.5.1. Phản ứng methyl ở promoter....99
3.5.2. Phản ứng acetyl hoá histone....99
3.6. Kiểm soát kết thúc phiên mã ở vi khuẩn....100
3.7. Tín hiệu ngăn cản dừng phiên mã ở prokaryot....101
3.8. Kiểm soát kết thúc phiên mã ở eukaryot....103
3.9. Biến đổi phân tử ARNm trong tế bào eukaryot....104
3.9.1. Polyadenyl hoá đầu 3' của ARNm sơ cấp....104
3.9.2. Phản ứng cắt nối exon-intron....104
3.9.3. Các intron có khả năng tự cắt ra khỏi phân tử ARNm -Phản ứng self-splicing....106
3.9.4. Phản ứng *trans*-splicing nối hai exon của hai phân tử ARNm....108
3.10. Kiểm soát sau phiên mã ở tế bào eukaryot....108
3.10.1. Độ dài của đuôi polyA....108
3.10.2. Độ bền vững của ARNm....109
3.10.3. ARNmi (micro RNA)....111
3.10.4. Phản ứng đọc sửa ARNm - "RNA editing"....111

Chương 4. DỊCH MÃ TỔNG HỢP PROTEIN

4.1. Phản ứng tổng hợp protein....116
4.2. Tính chính xác của phản ứng tổng hợp protein....121
4.3. Protein được vận chuyển về đích khi đang hoặc đã tổng hợp xong....124
4.4. Protein đi vào mạng lưới nội chất....124
4.4.1. Tín hiệu dẫn....125
4.4.2. Translocator – kênh vận chuyển trên màng lưới nội chất....126
4.4.3. Hình thành cấu trúc không gian của các chuỗi peptide trong khoang lưới nội chất....128

4.4.4. Quá trình đường hoá protein 129
4.5. Vận chuyển từ mạng lưới nội chất về đích 130
4.6. Vận chuyển protein màng 131
4.7. Tín hiệu dẫn protein tổng hợp tự do trong tế bào chất về đích 134
4.8. Protein chaperon 136

Chương 5. KỸ THUẬT ADN TÁI TỔ HỢP 139

5.1. Cắt ADN bằng enzym giới hạn 140
5.2. Phân ly các đoạn nucleotide có kích thước khác nhau bằng kỹ thuật điện di 141
5.3. Xây dựng bản đồ vị trí của enzym giới hạn 142
5.4. Các vector trong kỹ thuật tách dòng 144
5.4.1. Plasmid 145
5.4.2. Phage 145
5.5. Đưa ADN vào vector 147
5.6. Ngân hàng ADN tổng số 149
5.7. Ngân hàng các ADNc 151
5.8. Sàng lọc một dòng từ ngân hàng ADN 152
5.8.1. Nguyên tắc của kỹ thuật lai acid nucleic 153
5.8.2. Điều kiện của phản ứng lai 155
5.8.3. Phương pháp đánh dấu đầu dò 156
5.8.4. Phương pháp sàng lọc chung (screening) 158
5.9. Các kỹ thuật lai acid nucleic 158
5.9.1. Phương pháp Southern blots 161
5.9.2. Phương pháp northern blots 162
5.9.3. Kỹ thuật lai tại chỗ (*insitu*) 162
5.10. Xác định trình tự nucleotide 163
5.10.1. Phương pháp hoá học Maxam-Gilbert 164
5.10.2. Phương pháp enzym Sanger (phương pháp tạo đầu tận cùng của chuỗi-chain terminator method) 165
5.10.3. Xác định trình tự nucleotide trên máy tự động 167
5.11. Một số kỹ thuật xác định tương tác giữa ADN và protein 167
5.11.1. Phương pháp "DNA footprint" 168
5.11.2. Phương pháp xác định băng điện di chậm 169
5.12. Phản ứng PCR (Polymerase Chain Reaction) 170
5.12.1. Một số yếu tố ảnh hưởng đến phản ứng PCR 172
5.12.2. Một số ứng dụng của phản ứng PCR 174
5.13. Một số ứng dụng của sinh học phân tử 174
5.13.1. Sản xuất vaccine tái tổ hợp 174
5.13.2. Sản xuất các hợp chất có hoạt tính sinh học phục vụ và nâng cao chất lượng cuộc sống 175
5.13.3. Chẩn đoán mầm bệnh 175
5.13.4. Liệu pháp gen 176
5.13.5. Chuyển gen tạo các động, thực vật có năng suất chất lượng cao 177
5.13.6. Chọn lọc ưu thế lai 178
Giải nghĩa Thuật ngữ chuyên dụng 191
Tài liệu tham khảo 191

Chương 1

CẤU TRÚC VÀ CHỨC NĂNG CỦA PHÂN TỬ ACID NUCLEIC

Sự tồn tại của acid nucleic như một hợp chất giàu nitơ và phospho trong tế bào được phát hiện lần đầu tiên bởi Johann Friedrich Miescher (Thuỵ Sĩ) vào năm 1868. Vào thời gian đó, chỉ có các chất béo được biết là những hợp chất giàu nitơ và phospho. Vì thế khám phá của Friedrich Miescher không được các nhà nghiên cứu quan tâm đến.

Vai trò đặc biệt quan trọng của acid nucleic trong việc lưu trữ và truyền tải thông tin di truyền cho các thế hệ sau bắt đầu được biết đến vào những năm 1940 từ những thí nghiệm về biến nạp, chuyển nhiễm trên vi khuẩn và virus. Đến năm 1953, cấu trúc phân tử acid nucleic đã được công bố bởi hai nhà bác học James Watson và Francis Crick. Trong tế bào, acid nucleic gồm hai loại: ADN và ARN. Chúng có cấu trúc cơ bản gồm các nucleotide liên kết với nhau, mỗi nucleotide gồm một base purin hoặc pirimidin liên kết với phân tử đường pentose.

Trong tế bào, toàn bộ thông tin di truyền được lưu trữ trong một hoặc nhiều phân tử acid nucleic, có thể là ADN hoặc ARN. Các phân tử này có thể tồn tại ở dạng thẳng hoặc vòng tròn. Hầu hết phân tử acid nucleic không tồn tại tự do mà thường liên kết với các loại protein khác nhau, tạo thành các ***nhiễm sắc thể***. Tập hợp tất cả các nhiễm sắc thể cùng với những phân tử ADN tự do hoặc ADN tồn tại trong bào quan (đối với tế bào eukaryot) được gọi là ***genome (hệ gen)***. Điều cần lưu ý là những phân tử ADN tồn tại tự do thường có kích thước rất nhỏ so với nhiễm sắc thể.

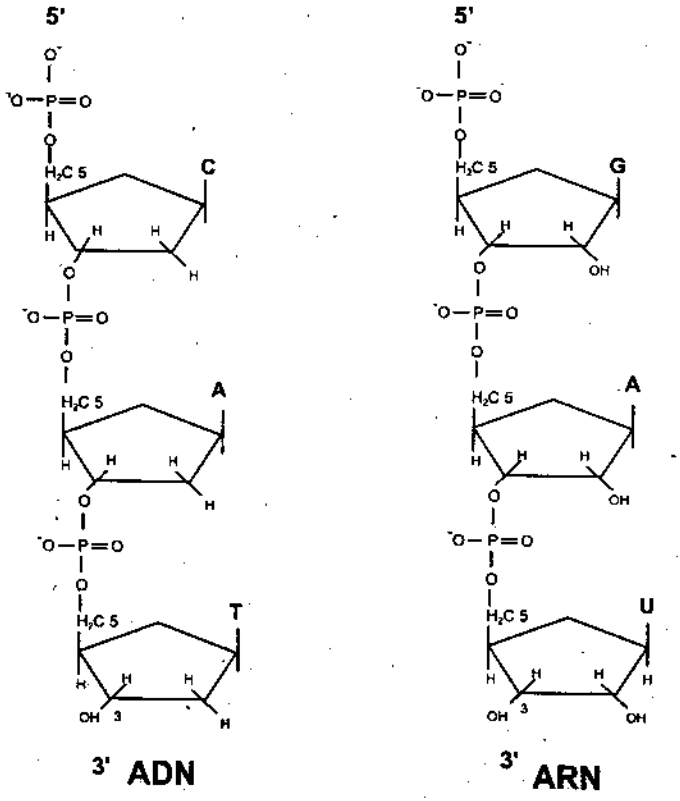
Không phải tất cả trình tự nucleotide trên nhiễm sắc thể đều chứa ***gen***. Gen là đơn vị chứa thông tin di truyền (một trình tự các nucleotide) mã cho một sản phẩm cần thiết đối với hoạt động sống của tế bào. Gen được phiên mã thành các loại ARN (ARN vận chuyển-ARNt, ARN ribosome-ARNr, ARN thông tin-ARNm). Các phân tử ARN thông tin (ARNm-messager RNA) có thể được dịch mã thành

protein. Tuy nhiên một số ARN không được dịch mã mà liên quan đến việc hình thành, duy trì và thậm chí thay đổi trật tự nucleotide cũng như cấu trúc không gian của nhiễm sắc thể. Như vậy, các loại ARN đó có vai trò quan trọng đối với thông tin di truyền ***ngoại gen*** (***epigenetic information***), là dạng thông tin phụ thuộc vào cấu trúc không gian hơn là vào trình tự nucleotide. Sai hỏng ở thông tin ngoại gen có thể gây nên những tính trạng mới mặc dù trình tự nucleotide không bị đột biến.

1.1. CẤU TRÚC NUCLEOTIDE. CẤU TRÚC PHÂN TỬ ADN, ARN

Hai loại phân tử ADN và ARN đều có cấu trúc hoá học tương tự như nhau, gồm các nucleotide (các monomer) liên kết với nhau tạo thành sợi ADN hoặc ARN (sợi polymer). Một phân tử ADN hoặc ARN gồm 4 loại nucleotide, mỗi loại gồm một phân tử đường pentose liên kết với một base nitơ (ở vị trí 1') và với một nhóm phosphate (thường ở vị trí 5'). Trong phân tử ADN, cả 4 loại nucleotide đều chứa đường pentose thuộc loại deoxyribose, còn trong phân tử ARN, bốn loại nucleotide có đường thuộc loại ribose. Như vậy sự khác nhau chính giữa các nucleotide của một phân tử ADN hoặc ARN phụ thuộc vào base nitơ. Trong phân tử ADN, các base nitơ bao gồm adenine (A), guanine (G), cytosine (C) và thymidine (T). Trong phân tử ARN, các base nitơ gồm A, C, G và uraxine (U). Rõ ràng rằng, hai loại phân tử ADN và ARN có cấu trúc hoá học khác nhau ở đường deoxyribose/ribose và ở base nitơ T/U.

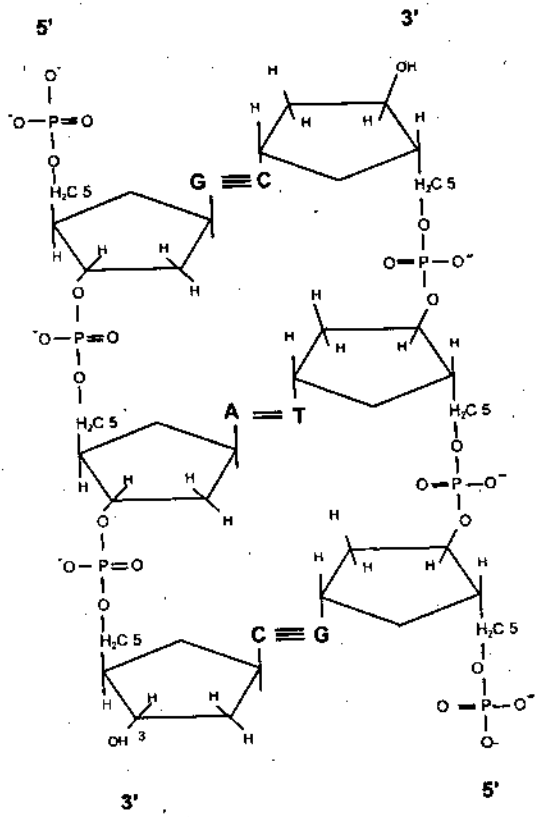
Tính acid của nucleotide được quy định bởi nhóm phosphate trên phân tử đường. Nhóm này bị mất ion hydro ở pH trung tính như pH bên trong tế bào, làm cho nucleotide tích điện âm (-). Khi nucleotide không có nhóm phosphate (chỉ gồm base nitơ và đường pentose) thì chúng được gọi là *nucleoside*. Như vậy, nucleotide là các nucleoside có gắn thêm một, hai hoặc ba gốc phosphate; lúc đó chúng được gọi là *nucleoside monophosphate, di- hoặc triphosphate*. Hai nucleotide liên kết với nhau bởi cầu nối phosphoester giữa nhóm hydroxyl OH ở vị trí 3' với nhóm phosphate ở vị trí 5' của hai phân tử đường. Với cách thức tương tác như vậy, sợi polynucleotide được xem như có hai đầu ký hiệu là 3' và 5'. Đầu 3' có nhóm OH tự do, còn đầu 5' thường có nhóm phosphate tự do. Do đó, sợi polynucleotide chỉ được kéo dài ở một đầu 3' (Hình 1.1). Hướng kéo dài từ 5' sang 3' là tính chất rất quan trọng đối với phân tử acid nucleic.



Hình 1.1: Cấu trúc sợi polynucleotide gồm các base nitơ và đường pentose.

Phân tử ADN gồm 4 loại base nitơ A, C, G, T và đường deoxyribose, còn phân tử ARN gồm A, C, G, U và đường ribose.

Trình tự các nucleotide trên một sợi ADN hoặc ARN được gọi là cấu trúc bậc 1 của phân tử acid nucleic. Mặc dù cấu trúc bậc 1 của ADN và ARN tương tự như nhau nhưng chúng có cấu trúc không gian rất khác nhau do khả năng xoắn, gấp khúc của từng loại. Phân tử ARN chỉ là một sợi đơn polynucleotide nhưng ADN chủ yếu gồm hai sợi đơn có chiều ngược nhau và liên kết hydrogen giữa các base nitơ của hai sợi tạo nên phân tử xoắn kép. Mỗi vòng xoắn gồm 10 cặp base nitơ. Do tính chất đặc biệt, adenine chỉ liên kết với thymine (hai liên kết hydrogen) và cytosine chỉ liên kết với guanine (3 liên kết hydrogen) nên phân tử kép ADN gồm hai sợi đơn bổ sung ngược chiều với nhau. Ví dụ, khi biết trình tự của một sợi đơn theo chiều 5' → 3' là GACTTACG thì dựa vào nguyên tắc tạo cặp bổ sung sẽ suy ra trình tự của sợi đơn kia theo chiều 3'→5' là CTGAATGC (Hình 1.2).



Hình 1.2: Cấu trúc sợi kép ADN gồm hai sợi đơn ngược chiều nhau có base nitơ bổ sung và liên kết với nhau bởi các cầu nối hydrogen.

Base Guanine (G) trên một sợi đơn này sẽ tạo 3 liên kết hydrogen với base Cytosine (C), và base Adenine (A) tạo 2 liên kết với base Thymidine (T) trên sợi đơn kia.

1.2. ADN LÀ VẬT LIỆU DI TRUYỀN

Năm 1928, những nghiên cứu trên vi khuẩn *Pneumococcus* gây bệnh viêm phổi ở chuột đã giúp các nhà sinh học đi đến kết luận vật liệu chứa thông tin di truyền là ADN. Các chủng *Pneumococcus* khác nhau có thể phân biệt nhờ căn cứ vào lớp vỏ polysaccharide bao bọc bên ngoài tế bào. Một số chủng có vỏ polysaccharide khiến bề mặt tế bào nhẵn và chúng có khả năng gây bệnh; do đó chúng được gọi là *Pneumococcus* S (smooth). Ngược lại, những loại không có vỏ polysaccharide thì bề mặt tế bào sần; do đó chúng được gọi là *Pneumococcus* loại R (rough). Chủng

Pneumococcus R không có khả năng gây bệnh; chúng bị phân hủy sau khi xâm nhiễm vào cơ thể vật chủ. Mối liên hệ khăng khít giữa sự xuất hiện của vỏ polysaccharide với tính độc của *Pneumococcus* dễ dàng đưa chúng ta đến kết luận rằng, vỏ polysaccharide là yếu tố duy nhất quyết định tính độc của vi khuẩn.

Tuy nhiên, vấn đề không như chúng ta tưởng. Thí nghiệm trong đó sử dụng *Pneumococcus* loại S và R kết hợp với nhau để gây nhiễm trên chuột đưa đến những kết quả bất ngờ (Bảng 1.1). Khi gây nhiễm cho chuột với chủng *Pneumococcus* có độc tính S nhưng đã bị chết vì nhiệt hoặc gây nhiễm với chủng *Pneumococcus* lành R thì chuột không chết. Tuy nhiên, khi lây nhiễm đồng thời cùng một lúc cả hai cho chuột thì có hiện tượng chuột bị chết, mặc dù loại có độc tính đã bị giết. Các xét nghiệm từ xác chuột chết đã phân lập được chủng *Pneumococcus* có bề mặt sần. Điều này chứng tỏ yếu tố mang thông tin di truyền từ chủng gây chết S đã chuyển sang chủng lành R. Thực nghiệm với dịch chiết tế bào đã chứng minh yếu tố mang thông tin di truyền trong tế bào vi khuẩn *Pneumococcus* là phân tử ADN. Thông tin di truyền được giải mã thành vỏ polysaccharide quy định nên tính độc của *Pneumococcus*. Khám phá này được công bố năm 1944. Quá trình đưa phân tử ADN của chủng R vào tế bào chủng S được gọi là **biến nạp** (***transformation***).

Nghiên cứu trên virus T2 tiếp tục khẳng định phân tử ADN mang thông tin di truyền được tái bản và được chuyển cho các thế hệ tiếp (năm 1952). Virus T2 có khả năng xâm nhiễm tế bào vi khuẩn *E.coli* nên T2 còn được gọi là thực khuẩn thể. Sau khoảng 20 phút nhiễm với T2, tế bào vi khuẩn bị vỡ (trạng thái sinh tan) giải phóng ra các virus T2 mới. Tế bào T2 bao gồm phân tử ADN được bao bọc bởi lớp vỏ protein. Thực nghiệm đã đánh dấu riêng biệt phân tử ADN bằng đồng vị phóng xạ P^{32}, còn vỏ protein bằng S^{35}. Nhờ đó có thể theo dõi được thành phần ADN hay protein có mặt trong tế bào vi khuẩn để hình thành nên các thể T2 mới. Đồng vị phóng xạ P^{32} được phát hiện chủ yếu (70%) trong các tế bào vi khuẩn nhiễm T2. Hơn nữa, khoảng 30% P^{32} chứ không phải S^{35} được phát hiện trong các thể virus T2 giải phóng ra từ tế bào vi khuẩn bị vỡ. Điều này chứng tỏ ADN của virus chứ không phải protein được tái bản trong vi khuẩn và di truyền cho các thế hệ T2 tiếp theo. Như vậy, phân tử ADN là vật liệu di truyền không chỉ đối với vi khuẩn mà với cả virus.

Bảng 1.1: Sự gây nhiễm đồng thời *Pneumococcus* loại có độc tính S (đã chết) cùng với loại lành R là nguyên nhân gây chuột chết

Loại *Pneumococcus* nhiễm cho chuột	Kết quả sau khi nhiễm
Vi khuẩn loại S	Chuột chết
Vi khuẩn loại R	Chuột sống
Vi khuẩn loại S đã chết (do nhiệt)	Chuột sống
Vi khuẩn loại S cùng nhiễm với loại R	Chuột chết
Vi khuẩn S đã chết cùng nhiễm vi khuẩn loại R	Chuột chết

Thí nghiệm chuyển ADN lạ (gen lạ) vào tế bào động vật không đơn giản như đối với vi sinh vật. Phải đến đầu thập kỷ 70 thế kỷ XX thực nghiệm mới thành công trong việc đưa ADN lạ vào tế bào động vật nuôi cấy cũng như vào tế bào trứng làm xuất hiện tính trạng mới. Quá trình này được gọi là **chuyển nhiễm (*transfection*)**. Bản chất của biến nạp ở tế bào vi khuẩn và chuyển nhiễm ở tế bào động vật đều khiến cho ADN lạ tồn tại trong genome của tế bào nhận và làm xuất hiện tính trạng mới. Hơn nữa, ADN lạ và tính trạng mới được di truyền cho thế hệ sau. Điều đó cho thấy ADN là vật liệu di truyền có thể tồn tại trong các loại tế bào của các cơ thể khác nhau mà vẫn giữ được chức năng của ADN.

Ngày nay, chức năng của ADN trong việc lưu trữ thông tin di truyền và chuyển thông tin đó cho các thế hệ sau được biết đến ở rất nhiều cơ thể sinh vật khác nhau từ vi khuẩn đến động, thực vật bậc cao và con người. Tuy nhiên, trong một số virus, phân tử ARN giữ các chức năng đó. Chúng ta sẽ xem xét đến những trường hợp thiểu số này trong chương 2.

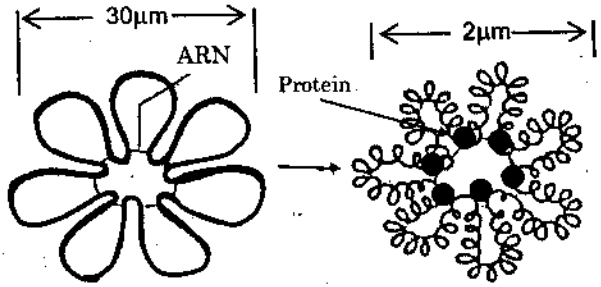
1.3. CẤU TRÚC NHIỄM SẮC THỂ

Tập hợp của tất cả vật liệu di truyền có mặt trong tế bào đảm bảo sự tồn tại và phân chia của tế bào được gọi là ***genome***. Tế bào prokaryot không có màng nhân nên genome tồn tại trong tế bào chất. Ngược lại, hầu hết genome trong tế bào eukaryot phân bố trong nhân cách biệt với tế bào chất bởi màng nhân. Ngoài ra, một phần rất nhỏ nhưng quan trọng của genome eukaryot phân bố trong các bào quan như ty thể, lục lạp. Dù sao, một trong những điểm giống nhau giữa hai loại tế bào là genome (ADN) đều liên kết với protein để tạo nên cấu

trúc không gian đặc thù cho từng loại tế bào. Trong tế bào prokaryot, ADN nằm trong cấu trúc hạch nhân (nucleoid), còn trong tế bào eukaryot ADN nằm trong cấu trúc sợi nhiễm sắc (chromatin). Khi tế bào bước vào phân chia (giảm nhiễm hoặc nguyên nhiễm), các sợi nhiễm sắc co đặc lại và quan sát rõ ràng. Lúc đó, sợi nhiễm sắc được gọi là nhiễm sắc thể. Kích thước của nhiễm sắc thể và sợi nhiễm sắc thay đổi rất lớn. Ví dụ, khi tế bào bước vào phân chia thì nhiễm sắc thể nhỏ nhất của người co đặc lại và có chiều dài khoảng 2μm. Khi tế bào ở giai đoạn gian kỳ (tế bào không phân chia) thì nhiễm sắc thể duỗi ra đạt đến chiều dài khoảng 14.000μm. Ngoài ra, nồng độ ADN trong tế bào rất cao. Ví dụ, nồng độ ADN trong tế bào vi khuẩn vào khoảng 10mg/ml, trong virus lớn hơn (500mg/ml), hoặc trong nhân tế bào eukaryot khoảng 100mg/ml. Trong thực nghiệm, khi ADN tồn tại ở những nồng độ lớn như thế, chúng tạo nên dung dịch có độ nhớt rất cao. Vì vậy, câu hỏi còn bỏ ngỏ chưa được giải đáp là trong tế bào, bằng cách nào mà protein tương tác được với ADN khi ADN có nồng độ đậm đặc như vậy.

1.3.1. Nhiễm sắc thể vi khuẩn có cấu trúc hạch nhân (nucleoid)

Khoảng 80% ADN có mặt trong tế bào vi khuẩn được phân bố trong hạch nhân (nucleoid). Hạch nhân không có màng nhân bao bọc. ADN nằm trong hạch nhân liên kết với protein để hình thành nên cấu trúc không gian đặc thù. Một số protein như HU hoặc H_1 tham gia đóng gói ADN trong cấu trúc hạch nhân. Hạch nhân có thể phân chia thành khoảng 50-100 vùng khác nhau; mỗi vùng được gọi là một "domain". Cấu trúc của một domain gồm khoảng 40kb ở dạng siêu xoắn. Các domain đính kết với nhau ở những đoạn ADN nhất định. Điều đó giúp cho toàn bộ genome (hạch nhân) có cấu trúc như một bông hoa gồm 50-100 cánh đính kết với nhau tại nhị hoa (Hình 1.3). Sự mở xoắn của một cánh (do đứt gãy hoặc do phiên mã tổng hợp ARN) không làm ảnh hưởng đến các cánh khác. Hai enzym ADN gyrase và ADN topoisomerrase giữ vai trò chính cùng phối hợp với phức protein làm nhiệm vụ tạo cánh hoa cho ADN vi khuẩn. Thực nghiệm đã phát hiện được ít nhất 4 protein tham gia phức này, trong đó protein HU có chức năng tương tự như histone ở tế bào eukaryot. Mặc dù có cấu trúc rất khác với histone nhưng HU tạo thành lõi được quấn quanh bởi đoạn ADN khoảng 60bp. Như vậy, protein HU có chức năng tương tự histone trong việc quy định nghiêm ngặt cấu trúc không gian của sợi nhiễm sắc thể. Tuy nhiên, chúng ta chưa xác định được các lõi này có phân bố đều đặn hay chỉ tập trung tại "nhị hoa" nucleoid.



Hình 1.3: Mô hình cấu trúc nucleoid ở *E.coli* gồm 50-100 vòng siêu xoắn kết dính với lõi protein và sợi ARN. Khi có đứt gãy xảy ra ở một vòng siêu xoắn, nhiễm sắc thể chỉ mở xoắn cục bộ ở vùng này (theo Snustad và Simmons, 2000).

1.3.2. Nhiễm sắc thể eukaryot gồm các vùng dị nhiễm sắc (heterochromatin) và nhiễm sắc thực (euchromatin)

Trong nhân tế bào eukaryot, ADN liên kết với protein tạo nên sợi nhiễm sắc. Các sợi nhiễm sắc có cấu trúc xoắn, gấp khúc hình thành cấu trúc sợi trục. Khi tế bào ở giai đoạn gian kỳ (không phân chia), từng sợi nhiễm sắc rất khó phân biệt với nhau. Tuy nhiên, ở giai đoạn tế bào sắp bước vào phân chia thì sợi nhiễm sắc co đặc lại. Lúc đó chúng được gọi là nhiễm sắc thể với đường kính khoảng 30nm.

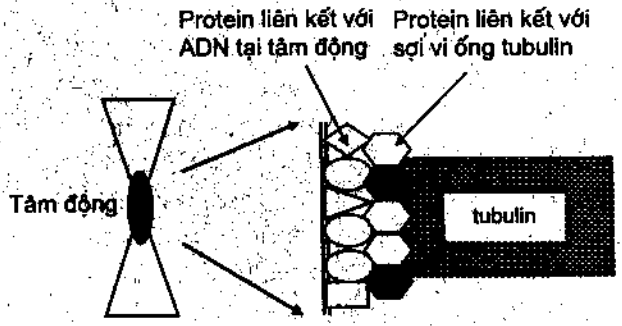
Quan sát dưới kính hiển vi điện tử cho thấy sợi nhiễm sắc gồm các vùng co đậm đặc hoặc không co đặc xen kẽ lẫn nhau. Vùng đậm đặc được gọi là vùng ***dị nhiễm sắc*** (***heterochromatin***). Phần lớn các vùng dị nhiễm sắc hầu như không thay đổi ngay cả khi tế bào bước vào phân chia, chúng được gọi là vùng dị nhiễm sắc bảo thủ (constitutive heterochromatin). Một số vùng dị nhiễm sắc có thể xuất hiện mới hoặc biến mất khi tế bào trải qua các giai đoạn sinh trưởng biệt hoá, chúng được gọi là vùng dị nhiễm sắc bất ổn (facultative heterochromatin). Vùng nhiễm sắc không co đặc được gọi là ***nhiễm sắc thực*** (***euchromatin***). Nhiễm sắc thực chỉ co đặc khi tế bào chuẩn bị phân chia. Trong giai đoạn gian kỳ, vùng nhiễm sắc thực phân bố rải rác khắp trong nhân. Hầu hết các gen phân bố trong vùng nhiễm sắc thực trong khi các đoạn ADN lặp lại và một số rất ít gen (các gen này ở trạng thái không hoạt động) phân bố trong vùng dị nhiễm sắc bảo thủ. Ngoài ra, vùng dị nhiễm sắc sao chép chậm hơn các vùng nhiễm sắc thực trong quá trình tái bản nhiễm sắc thể. Đặc biệt, trao đổi chéo hầu như không xảy ra ở vùng dị nhiễm sắc giữa hai nhiễm sắc thể tương đồng.

Những tính chất đặc biệt trên của vùng dị nhiễm sắc được sinh học hiện đại quan tâm đến vào những năm cuối thập kỷ 20 thế kỷ XX. Sự hình thành và duy trì cấu trúc dị nhiễm sắc phụ thuộc vào mức độ methyl hoá của cytosine và histone, mức độ khử acetyl của histone. Thực nghiệm cho thấy, khi khử bỏ các nhóm methyl trên phân tử ADN và histone thì vùng dị nhiễm sắc giãn ra và một số gen phân bố trong vùng đó có thể được phiên mã.

Bằng kỹ thuật làm tiêu bản tế bào nhuộm với chất màu Giemsa, chúng ta có thể quan sát được nhiễm sắc thể gồm các băng có màu nhuộm ở mức độ khác nhau. Chúng phân bố nằm ngang nhiễm sắc thể và được gọi là các băng G (***G-bands***). Số lượng băng G cũng như độ rộng của từng băng đặc trưng cho từng nhiễm sắc thể. Vì vậy các băng G được sử dụng trong phân biệt kiểu nhân (karyotype). Điều thú vị là các vùng nằm giữa các băng G có thành phần G-C lớn hơn so với chính băng G. Ví dụ như, nhiễm sắc thể động vật có vú có thành phần G-C trung bình khoảng 40%, trong đó 30% phân bố trong các băng G còn khoảng 65% phân bố trong vùng nằm giữa các băng. Tương tự như ở vùng nhiễm sắc thực, hầu hết các gen phân bố ở vùng nằm giữa các băng G (vùng có màu rất nhạt).

1.3.3. Tâm động (centromere) của nhiễm sắc thể eukaryot

Trong quá trình phân bào, các sợi vi ống tubulin kéo nhiễm sắc thể về hai cực của tế bào. Một đầu sợi vi ống được cố định ở trung thể, đầu kia đính vào vị trí đặc biệt trên nhiễm sắc thể thông qua phức protein (Hình1.4).



Hình 1.4: Tâm động gồm trình tự nucleotide cần thiết cho sự phân ly của nhiễm sắc thể trong quá trình phân bào. Phức protein tương tác với tâm động. Phức này liên kết với sợi vi ống tubulin thông qua các protein trung gian khác.

Vị trí đặc biệt này được gọi là ***tâm động (centromere)***. Sự có mặt của tâm động là điều kiện tiên quyết để nhiễm sắc thể phân ly chính xác trong phân bào. Khi nhiễm sắc thể bị đứt gãy thành hai phần, trong đó một phần chứa tâm động, thì chỉ có phần chứa tâm động tồn tại trong các tế bào con. Phần không có tâm động bị mất trong quá trình phân bào. Như vậy, mỗi nhiễm sắc thể chỉ cần một và duy nhất một tâm động mà thôi. Trong một số trường hợp đột biến chuyển đoạn khiến cho một nhiễm sắc thể có hai tâm động. Khi phân ly, các sợi vi ống định vào hai tâm động có thể kéo nhiễm sắc thể theo hai hướng khác nhau gây đứt gãy. Mặt khác khi chuyển đổi vị trí của tâm động từ nhiễm sắc thể này sang nhiễm sắc thể khác thì quá trình phân ly của nhiễm sắc thể với tâm động lạ vẫn không bị ảnh hưởng. Do đó, trình tự nucleotide tại tâm động dường như chỉ liên quan đến nhiệm vụ đảm bảo phân ly chính xác của nhiễm sắc thể.

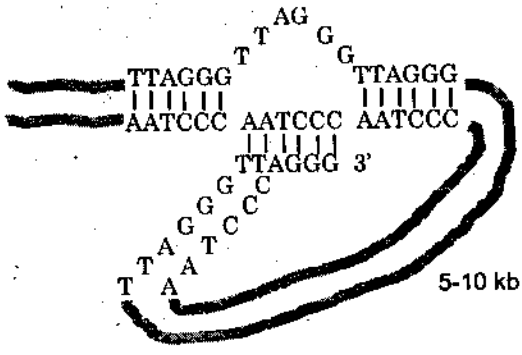
Tâm động eukaryot đơn giản nhất là tâm động nhiễm sắc thể nấm men *S.cerevisiae* chỉ gồm 125bp. Tuy nhiên, hầu hết các tâm động eukaryot đều chứa các đoạn ADN lặp lại liên tiếp. Ví dụ, ở nấm *S.pombe*, ở ruồi giấm *Drosophila*, ở thực vật *Arabidopsis*, các đoạn ADN ngắn lặp lại nhiều lần tạo thành tâm động tương ứng với chiều dài khoảng 40-100kb (*S.pombe*), 200kb (*Dresophila*) và lớn hơn 500kb (*Arabidopsis*).

1.3.4. Đầu mút (telomere) của nhiễm sắc thể eukaryot

Nhiễm sắc thể trong nhân tế bào eukaryot tồn tại ở dạng thẳng. Hai đầu mút nhiễm sắc thể (còn gọi là telomere) bao gồm các trình tự ADN lặp lại và có chứa đoạn nucleotide ở dạng sợi đơn; đoạn này có khả năng hình thành cấu trúc cặp tóc (hair loop). Telomere được xem như là cấu trúc cần thiết để đảm bảo nguyên vẹn chiều dài của nhiễm sắc thể. Trong thực nghiệm, các đoạn nhiễm sắc thể tạo ra do bị đứt gãy vì một tác động bất kỳ có khả năng kết hợp với nhau tạo nên phân tử mới. Tuy nhiên trong tế bào, cấu trúc telomere ngăn cản việc nối ghép những đoạn ADN bất kỳ vào hai đầu nhiễm sắc thể. Thực nghiệm đã thiết kế thành công plasmid dạng thẳng có mang đầu mút telomere của nhiễm sắc thể *Tetrahymena* giúp cho plasmid trở nên bền vững trong các thế hệ tế bào nấm men trải qua nhiều lần phân chia. Khi thiếu các telomere, plasmid dạng thẳng bị đứt gãy nhanh chóng.

Telomere có cấu trúc tương tự như nhau ở sinh vật eukaryot bậc thấp, bậc cao và ở người. Telomere bao gồm 100-1000 đoạn ADN ngắn lặp lại liên tiếp. Đoạn ADN ngắn đó có thể biểu diễn dưới dạng *Cn(A*

hoặc *T)m* cho mọi cơ thể sinh vật eukaryot; trong đó n>1 và m=1-4. Số lần lặp lại của *Cn(A/T)m* có thể tăng hoặc giảm sau mỗi lần phân bào. Nối tiếp ngay sau các đoạn lặp lại là một chuỗi giàu G-T ở dạng sợi đơn dài khoảng 14-16 nucleotide. Chính đoạn giàu G-T dạng sợi đơn có khả năng tạo thành cấu trúc cặp tóc bảo vệ đầu tận cùng của nhiễm sắc thể (Hình 1.5). Chiều dài của đoạn ADN nằm trong cấu trúc cặp tóc khoảng từ 5-10kb. Do đầu mút gồm các đoạn ADN lặp lại nên các đầu mút có khả năng trao đổi chéo khi hai nhiễm sắc thể tương đồng tạo cặp trong phân bào. Để tránh hiện tượng trao đổi chéo không cân khiến một nhiễm sắc thể có telomere dài ra và nhiễm sắc thể kia có telomere ngắn lại; tế bào ngăn cản trao đổi chéo bằng các protein tương tác với telomere hoặc bởi cơ chế sửa chữa ADN.



Hình 1.5: Đoạn nucleotide ở dạng sợi đơn tại đầu 3' của telomere tạo cặp với đoạn bổ sung tách ra từ dạng sợi đúp nằm ngay trước đó. Sự tạo cặp này hình thành nên cấu trúc cặp tóc ở đầu tận cùng của nhiễm sắc thể. Phản ứng tạo cặp được xúc tác bởi protein TRF2 (theo Lewin, 2004).

Trong quá trình tái bản nhiễm sắc thể, đầu telomere được duy trì nhờ enzym telomerase. Hoạt tính của enzym không phát hiện được trong các tế bào đã biệt hoá. Vì một lý do nào đó mà tế bào bước vào phân chia nhưng không có enzym telomerase thì đầu mút nhiễm sắc thể sẽ bị ngắn dần qua các lần phân bào. Do đầu mút ngắn nên nhiễm sắc thể không bền vững, dễ bị đứt đoạn hoặc chuyển đoạn. Vì thế, trong tế bào ung thư (tế bào phân chia liên tục) các nhiễm sắc thể thay đổi hình dạng và xuất hiện nhiều phân tử ADN lạ, có kích thước nhỏ.

1.4. GENOME (HỆ GEN)

Genome chứa toàn bộ thông tin di truyền đảm bảo hoạt động sống cho tế bào. Đa số genome vi khuẩn phân bố trên một nhiễm sắc thể ở

dạng vòng khép kín và có kích thước nhỏ so với genome eukaryot. Ngược lại, phần genome trong nhân tế bào eukaryot thường rất lớn và phân bố trên các nhiễm sắc thể dạng thẳng. 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 gen- epigenetic information).

Những thông tin di truyền liên quan đến tính trạng (gen), đến trình tự nucleotide đặc biệt, đến độ dài của các đoạn ADN, đến vị trí nhận biết của enzym giới hạn cắt đặc hiệu ADN, đến tính đa allen ... được các nhà sinh học đưa vào bản đồ genome. Bản đồ mà khoảng cách giữa các vị trí được tính bằng đơn vị nucleotide được xem là chính xác nhất. Bản đồ này được gọi là bản đồ vật lý (phisical map). Ngoài ra còn có một số loại bản đồ khác. Ví dụ, bản đồ di truyền (genetic map) cho biết mối liên hệ về vị trí của các nhóm gen với nhau hay của các chỉ thị (markers) trên nhiễm sắc thể. Các chỉ thị 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), hoặc đa dạng các đoạn ADN được khuếch đại ngẫu nhiên (randomly amplified polymorphic DNA-RAPD). Khoảng cách giữa các vị trí trên bản đồ di truyền được tính bằng cM (centiMorgan) dựa vào tần số trao đổi chéo. Hai vị trí càng gần nhau thì giữa chúng càng khó xảy ra trao đổi chéo trong phân bào giảm nhiễm. Tuy nhiên, trao đổi chéo không xảy ra như nhau ở mọi vị trí trên nhiễm sắc thể nên đơn vị centiMorgan không phản ảnh chính xác khoảng cách giữa các vị trí trên bản đồ di truyền. Trong thực tế, sinh học phân tử kết hợp giữa bản đồ vật lý và bản đồ di truyền để biết chính xác khoảng cách giữa các gen (tính trạng), giữa các chỉ thị phân tử liên quan đến những tính trạng cần nghiên cứu.

Genome không phải đơn thuần là tập hợp của các gen. Genome của vi khuẩn và sinh vật eukaryot 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, thành phần ADN chứa các gen chỉ chiếm một tỷ lệ rất nhỏ so với toàn bộ genome trong tế bào eukaryot bậc cao. Các gen trong tế bào eukaryot bậc cao thường chứa nhiều intron và phân bố xa nhau. Từ những năm 70 thế kỷ XX, 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, microarray...), cho phép xác định số gen hoạt động trong một tế bào. Ví dụ, 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 10000bp thì tổng số chiều dài các gen hoạt động trong một tế bào eukaryot cũng chỉ chiếm 1-2% genome. Hay nói cách khác chỉ một phần rất nhỏ genome eukaryot mang thông tin di truyền cần thiết mã cho protein tham gia hoạt động sống của tế bào. So sánh 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 kích thước genome của những loài cách xa nhau (tức là có tính phức tạp khác nhau) cho thấy: kích thước genome không phải luôn tỷ lệ với tính phức tạp của loài. Ví dụ, genome của người có kích thước khoảng $3,3x10^9$bp, trong khi đó genome các loài lưỡng cư dài tương tự cỡ $3,1x10^9$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, ruồi sống trong nhà (*Musca domestica*) có genome cỡ $8,6x10^8$bp, lớn gấp 6 lần kích thước genome ruồi giấm (*D.melanogaster*) với genome cỡ $1,4x10^8$bp. Ngoài ra, kích thước genome của các quần thể lưỡng cư thay đổi từ 10^9bp đến 10^{11}bp (khác nhau gấp 100 lần). Vấn đề thu hút sự quan tâm của sinh học hiện đại là vì sao ngay trong cùng một loài kích thước genome lại biến thiên nhiều như vậy ?

Kết quả bước đầu so sánh genome 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 quy luật trong genome. Thứ hai là kích thước genome 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 genome riêng biệt của các bào quan (ty thể, lục lạp) sang genome trong nhân đều chịu ảnh hưởng của thời gian, 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 genome trong nhân.

1.4.1. Genome của tế bào prokaryot (tế bào nhân sơ)

Genome của tế bào prokaryot thường được nghiên cứu trên đối tượng vi khuẩn. Genome vi khuẩn không lớn nên số lượng genome của các loài vi khuẩn được xác định trình tự ngày càng nhiều. Nhờ đó các thông tin dữ liệu về cấu trúc hệ gen prokaryot, sự phân bố các gen, cách thức kiểm soát hoạt động cũng như chức năng của chúng ngày càng phong phú và trở nên rõ ràng.

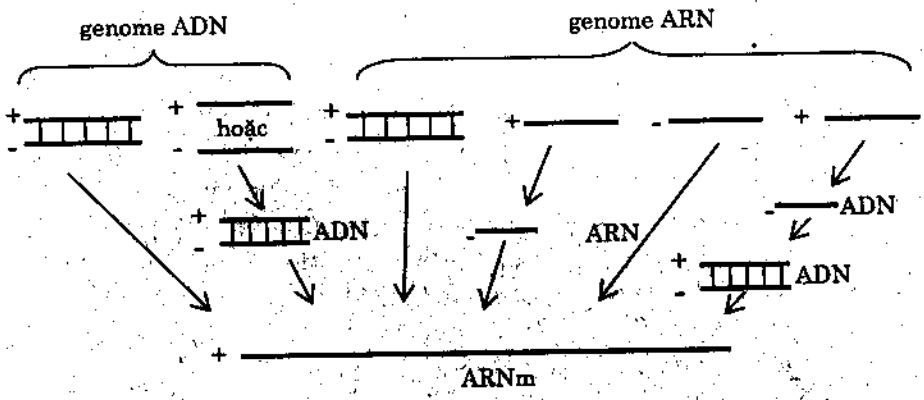
Genome prokaryot có kích thước nhỏ hơn rất nhiều so với genome eukaryot. Bên cạnh nhiễm sắc thể chứa phần lớn thông tin di truyền, tế bào prokaryot còn có nhiều loại plasmid. Trước đây, plasmid được xem là những phân tử ADN dạng vòng chứa các gen không quan trọng. Ví dụ, plasmid thường mang gen liên quan đến tính chống chịu kháng sinh. Lúc đó, tế bào vẫn có thể tồn tại ngay khi thiếu vắng các gen này. Tuy nhiên, khái niệm plasmid được mở rộng ra khi thực nghiệm tìm thấy một số tế bào prokaryot có chứa phân tử ADN kích thước nhỏ, ở dạng thẳng và mang các gen tương tự như plasmid dạng vòng. Vì vậy, plasmid được hiểu là những đoạn ADN kích thước nhỏ mang một số gen không quyết định sự sống còn của tế bào. Trong tự nhiên, một loại plasmid đôi khi có mặt trong các loại tế bào prokaryot khác nhau. Mặt khác, plasmid có khả năng biến nạp từ loại tế bào prokaryot này sang loại khác. Vì vậy, mặc dù có chứa gen nhưng plasmid dường như không được xem là một phần của genome.

Hầu hết genome prokaryot nhỏ hơn 5Mb (5.000.000 nucleotide) và thường được phân bố trên một nhiễm sắc thể dạng vòng. Một số tế bào prokaryot có genome là phân tử ADN dạng thẳng. Đặc biệt, một số genome prokaryot là phân tử ARN, hoặc kết hợp cả hai loại ADN và ARN. Ngoài ra, genome prokaryot có thể bao gồm các gen phân bố trên các đoạn thẳng ADN, hoặc trên cả hai loại phân tử ADN dạng thẳng và dạng vòng. Ví dụ, nhiễm sắc thể dạng thẳng được phát hiện lần đầu tiên ở *Borrelia burgdorferi* vào năm 1989. Nhiễm sắc thể này dài 910kb gồm 853 gen. Bên cạnh đó, tế bào *Borrelia burgdorferi* còn có tới 17 plasmid dạng vòng và dạng thẳng với tổng chiều dài là 533kb liên quan tới 430 gen. Hầu hết các gen phân bố trên plasmid không quan trọng, chỉ có một số ít gen cần thiết cho quá trình tổng hợp purine và protein màng tế bào. Do đó, trong tổng số 17 plasmid, một vài plasmid chứa các gen này được xem là một bộ phận của genome trong tế bào *Borrelia burgdorferi*.

Mặc dù vi khuẩn được xem là đối tượng điển hình cho tế bào prokaryot, virus cũng thu hút sự quan tâm đặc biệt bởi chúng gây ra

những bệnh hiểm nghèo cho động, thực vật và người. Virus gây tan tế bào vi khuẩn (thực khuẩn thể) được phát hiện đầu tiên vào năm 1910. Ngày nay, rất nhiều loại virus khác nhau được phân lập từ các tế bào chủ, hoặc là tệ bào vi khuẩn, hoặc là tế bào eukaryot (động, thực vật và người). Virus được xem là sinh vật nhỏ bé nhất mà con người biết.

Genome virus thường được bao bọc bởi lớp vỏ protein. Đối với một số virus động vật, vỏ protein còn được bao phủ bởi lớp màng lipid có nguồn gốc từ tế bào chủ. Genome virus có thể là phân tử ADN hoặc ARN, hoặc bao gồm cả hai loại ADN/ARN ở dạng sợi đôi hoặc sợi đơn. Do tính chất phức tạp của genome virus nên có thể phân loại virus dựa vào loại phân tử acid nucleic cấu thành nên genome (Hình 1.6). Genome virus không chứa các gen mã cho các protein hoặc ARNr cần thiết cho bộ máy ribosome, cũng như không chứa các gen liên quan đến con đường tổng hợp năng lượng. Virus "chiếm dụng" từ tế bào chủ tất cả những gì cần thiết để tăng sinh tạo ra các thể virus mới.



Hình 1.6: Phân loại virus xâm nhiễm tế bào động vật dựa vào thành phần genome và cách thức hình thành khuôn để phiên mã tạo ARNm virus.

Genome có thể là phân tử ADN sợi kép hoặc sợi đơn. Sau khi virus xâm nhiễm vào tế bào chủ, genome có thể được phiên mã trực tiếp hoặc được làm khuôn để tổng hợp nên khuôn mới cần thiết để phiên mã tạo ARNm. Sợi ký hiệu (-) là sợi có trình tự bổ sung với trình tự trên phân tử ARNm (theo Lodish và cs., 2000).

Quá trình tăng sinh của thực khuẩn thể T4 khi xâm nhiễm tế bào chủ vi khuẩn được nghiên cứu khá chi tiết. Thực khuẩn thể T4 bám trên bề mặt tế bào vi khuẩn. Tiếp đến genome T4 (ADN) được tống vào trong tế bào vi khuẩn còn lớp vỏ protein ở lại bên ngoài. Điều đáng lưu

ý là genome T4 có một số cytosine gắn nhóm hydroxylmethyl CH_2OH ở vị trí 5' (ký hiệu là HMC-hydroxyl-methylcytossine). Hơn nữa, HMC còn bị đường hoá (gắn thêm các dẫn xuất của glucose). Nhờ đó, ADN của T4 tránh được sự phân hủy bởi chính các nuclease do T4 tổng hợp. Những nuclease này tác động lên genome tế bào chủ.

Một số virus có genome là phân tử ARN. Chúng được gọi chung là ***retrovirus***. Ví dụ, virus gây suy giảm miễn dịch (HIV-human immunodeficiency virus) bao gồm hai phân tử ARN giống hệt nhau và enzym phiên mã ngược (enzym tổng hợp ADN trên khuôn ARN) được đóng gói trong lớp vỏ protein.

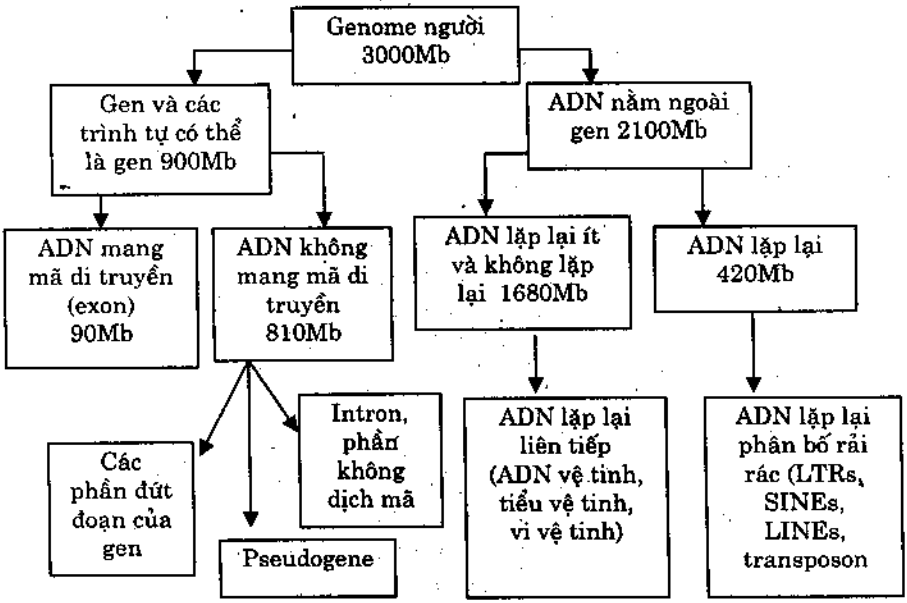
1.4.2. Genome của tế bào eukaryot (tế bào nhân thực)

Genome của tế bào eukaryot bao gồm các nhiễm sắc thể phân bố trong nhân và ADN phân bố trong một số bào quan như lục lạp, ty thể. Tuy nhiên, do hầu hết số lượng ADN cũng như các gen tập trung chủ yếu trong nhân nên ADN (nhiễm sắc thể) phân bố trong nhân được các nhà sinh học quan tâm rất nhiều.

Các nhiễm sắc thể là các phân tử ADN liên kết với protein. Chúng tồn tại ở dạng thẳng. Không có mối liên hệ ràng buộc nào giữa ba thông số sinh học: số lượng nhiễm sắc thể, kích thước genome và tính phức tạp của loài. Ví dụ, nấm men *S. cerevisiae* được xem là sinh vật eukaryot bậc thấp nhưng lại có số lượng nhiễm sắc thể nhiều gấp 4 lần ruồi giấm *D. melanogaster*. Ngoài ra, kích thước genome kỳ nhông lớn gấp 30 lần genome của người nhưng số lượng nhiễm sắc thể chỉ bằng một nửa. Hơn nữa, một số nhiễm sắc thể có kích thước rất nhỏ (các nhiễm sắc thể mini) nhưng có mật độ phân bố các gen rất cao. Ví dụ, genome của gà gồm 39 nhiễm sắc thể, trong đó 6 nhiễm sắc thể bình thường chiếm 66% ADN nhưng chỉ có 25% các gen phân bố trên 6 nhiễm sắc thể đó. Ba mươi ba nhiễm sắc thể còn lại đều là nhiễm sắc thể mini chiếm 1/3 ADN và có tới 75% các gen. Những so sánh lý thú này cho thấy sự bí hiểm giữa tiến hoá và cấu trúc genome trong các sinh vật khác nhau mà hiện tại sinh học chưa tìm ra câu trả lời thích đáng.

Kích thước genome trong nhân eukaryot thay đổi từ 12Mb (nấm men *S. cerevisiae*) đến 120000Mb (thực vật *F. assyriaca*). Genome bao gồm thành phần ADN không lặp lại và ADN lặp lại. Phần lớn các gen phân bố trong thành phần ADN không lặp lại và số lượng của chúng tăng cùng với tính phức tạp của loài. Tuy nhiên, điều đặc biệt lưu ý là tính phức tạp không chỉ phụ thuộc vào số lượng gen mà còn được xác

định bởi thành phần ADN lặp lại. Vì vậy, không phải luôn luôn tồn tại mối tương quan tỷ lệ thuận giữa kích thước genome và tính phức tạp của loài. Ví dụ, kích thước genome của người khoảng 10^9bp trong khi genome một số loài lưỡng cư hoặc thực vật có thể đạt đến 10^{11}bp. Genome của người đã được giải mã hoàn toàn (năm 2001) bao gồm các thành phần ADN được trình bày trên Hình 1.7.



Hình 1.7: Các loại ADN trong genome người (theo Brown, 2001).

1.4.3. Methyl hoá ADN trong genome eukaryot

Bản thân ADN trong genome eukaryot chịu các biến đổi do gắn thêm các nhóm chức khác nhau. Ví dụ, hiện tượng methyl hoá cytosine hoặc adenine. Những thay đổi này có tính đặc thù cho từng vùng nhiễm sắc thể, tác động đến cấu trúc không gian của sợi nhiễm sắc và tham gia kiểm soát hoạt động của các gen. Những đặc thù cấu hình riêng của từng vùng nhiễm sắc thể được di truyền cho thế hệ sau. Sai lệch trong cấu trúc không gian của sợi nhiễm sắc có thể làm xuất hiện tính trạng mới ngay khi trình tự nucleotide không sai hỏng. Do đó, phân tử ADN có chứa hai dạng thông tin: thông tin di truyền (***genetic information***) quyết định bởi trình tự nucleotide và thông tin ngoại gen (***epigenetic information***) quyết định bởi tính phức tạp về cấu

hình không gian của genome. Hiện tượng methyl hoá xảy ra với cả ADN prokaryot và eukaryot. Tuy nhiên methyl hoá ADN ở prokaryot được xem như là một cơ chế bảo vệ genome, trong khi ở eukaryot, methyl hoá đóng vai trò quan trọng trong dạng thông tin thứ hai. Đó chính là một trong các cơ chế kiểm soát hiện tượng đánh dấu ADN (***DNA imprinting***), tức là tính trạng của gen được biểu hiện phụ thuộc vào nguồn gốc di truyền từ bố hay mẹ. Cần lưu ý "DNA imprinting" hoàn toàn khác với di truyền theo giới tính.

Methyl hoá ADN có ý nghĩa đặc biệt đối với hoạt động của gen eukaryot, nhất là các gen trong quá trình phát triển hình thành cá thể. Phản ứng methyl hoá xảy ra ở những vị trí đặc hiệu. Khoảng 2-7% ADN ở tế bào động vật bị methyl hoá. Hầu hết nhóm methyl được tìm thấy ở cytosine (C) phân bố trong cặp nucleotide CpG. Tỷ lệ cytosine bị methyl hoá thay đổi rất khác nhau giữa các loài. Hầu như không phát hiện được cytosine bị methyl hoá ở nấm men *S.cerevisiae*. Khoảng 10% cytosine bị methyl hoá ở động vật có xương sống và 30% ở thực vật. Chỉ đến năm cuối cùng của thập kỷ 20 thế kỷ XX mới khẳng định được có hiện tượng methyl hoá ADN ở *Drosophila*. Tuy nhiên, chỉ có khoảng 0,4% toàn bộ hệ gen ruồi giấm bị methyl hoá. Hơn nữa, cytosine gắn gốc methyl nằm trong cấu trúc CpT và CpA chứ không phải trong trật tự CpG (như đối với động vật bậc cao). Đặc biệt ở thực vật bậc cao, hiện phản ứng methyl hoá có thể xảy ra với cytosine trong mọi cấu trúc CpG, CpNpG và CpNpN, trong đó N = A, T hoặc C.

Khi đưa các gen bị methyl hoá, hoặc bị khử methyl vào genome tế bào nhận (thí nghiệm chuyển gen) thì chỉ những gen không có nhóm methyl mới hoạt động. Mặt khác, vùng ADN không có nhóm methyl thường trùng với vùng có các vị trí nhạy cảm ADNase. Thực nghiệm nhận thấy rất nhiều gen khi đang phiên mã tổng hợp ARN đều không có nhóm methyl ở vùng chứa promoter và exon thứ nhất (đầu 5'), mặc dù các exon tiếp sau và phía đầu 3' có chứa nhóm này. Rõ ràng methyl hoá có tác dụng ngăn cản gen hoạt động. Ngược lại, nếu khử nhóm này thì gen lại được hoạt hoá. Do đó để phân biệt với CpG bị methyl hoá, các cặp CpG không có gốc methyl, lặp đi lặp lại nhiều lần ở phía trước đầu 5' của gen khoảng 1-2kb được gọi là cụm **CpG *(CpG island)***. Khoảng 56% các gen trong genome người được phân bố gần với cụm CpG. Những gen hoạt động trong mọi loại tế bào (gọi là gen quản gia-*housekeeping genes*) đều có cụm CpG không bị methyl hoá. Tuy nhiên, đối với các gen đặc hiệu (chỉ hoạt động trong tổ chức chuyên biệt) thì phản ứng methyl hoá cụm CpG của chúng được kiểm soát chặt chẽ. Cụm này không bị methyl hoá trong tế bào cần đến sản phẩm của gen

nhưng lại bị gắn gốc methyl trong những tế bào mà gen không biểu hiện. Như vậy để một gen hoạt động, ngoài việc xuất hiện các vị trí nhạy cảm với nuclease gần promoter, ADN ở vùng chứa gen cần bị khử nhóm methyl. Khi đưa ADN đã bị methyl hoá vào tế bào, nó tiếp tục bị methyl hóa không ngừng qua mỗi lần nhân đôi ADN. Ngược lại, nếu đưa ADN không có nhóm methyl vào tế bào, chúng không bị methyl hoá sau mỗi lần tái bản. Rõ ràng, methyl hoá cytosine là cơ chế kiểm soát tiêu cực hoạt động của gen. Tuy nhiên, cần lưu ý một số gen đặc biệt vẫn được phiên mã ngay khi vùng điều khiển (ví dụ như promoter) có chứa nhóm methyl. Đó là do vùng điều khiển của gen được nhận biết đặc hiệu bởi protein hoạt hoá mà protein này chỉ tương tác với ADN thông qua nhóm methyl ở cytosine.

Phản ứng gắn nhóm methyl vào cytosine được xúc tác bởi các enzym methyltransferase. Có thể phân biệt các enzym này thành 2 nhóm. *Nhóm thứ nhất* làm nhiệm vụ duy trì gốc methyl ở những vị trí cytosine trên sợi đơn ADN vừa được tổng hợp trong quá trình tái bản ADN. Việc gắn gốc methyl mới này dựa vào nhóm ^{m}CpG trên sợi khuôn. Chúng được gọi chung là các enzym duy trì nhóm methyl (maintenance methyltransferase). *Nhóm thứ hai* gồm các enzym xúc tác phản ứng gắn gốc methyl vào vị trí cytosine trên phân tử ADN mà vị trí này trước đó không có nhóm methyl. Ví dụ, gắn nhóm methyl vào cụm CpG ở promoter khi cần kìm hãm hoạt động của gen. Ngoài ra, quá trình methyl hoá cytosine trong trật tự CpNpG, hoặc CpNpN đòi hỏi protein và các phân tử ARN kích thước ngắn (20-25 nucleotide) để nhận biết những cytosine đó. Nhờ đó, cấu trúc sợi nhiễm sắc cũng như hoạt động của gen bị thay đổi. Enzym demethylase làm nhiệm vụ khử nhóm methyl ở ADN chưa được tìm thấy trong tế bào động vật. Kết quả nghiên cứu gần đây (2002-2005) cho thấy một số enzym tham gia sửa chữa ADN có liên quan đến việc loại bỏ cytosine mang nhóm methyl. Lúc đó, đoạn ADN chứa ^{m}C bị loại đi và thay thế bởi cytosine không mang nhóm methyl.

1.5. ADN TRONG CÁC BÀO QUAN CỦA TẾ BÀO EUKARYOT

Đối với tế bào eukaryot, ADN không chỉ phân bố trong nhân mà còn có mặt ở các bào quan ty thể và lục lạp. ADN phân bố trong bào quan còn được gọi là ADN ngoài nhân (extranuclear genome). Tuy nhiên, ADN ngoài nhân không đồng nghĩa với ADN tồn tại trong tế bào chất (ví dụ, một số plasmid tồn tại trong tế bào chất). ADN của bào quan chứa các gen mã cho các sản phẩm cần thiết cho bào quan.

ADN bào quan không tuân thủ định luật Mendel khi phân ly trong phân bào giảm nhiễm hoặc nguyên nhiễm. Ví dụ, trong quá trình thụ tinh, tế bào trứng chỉ tiếp nhận genome đơn bội mà hầu như không tiếp nhận tế bào chất từ tinh trùng. Do đó, phôi chỉ có ty thể của mẹ mà không tiếp nhận vật liệu di truyền ngoài nhân của bố. Đây là hiện tượng *di truyền theo mẹ (maternal inheritance)*. Ngoài ra, trong phân bào nguyên nhiễm, sự phân chia không đồng đều các ty thể về hai tế bào con có thể xảy ra ngẫu nhiên. Đây là sự phân ly không Mendel hay còn gọi là sự *phân ly sinh dưỡng* (*somatic segregation*). Do điều kiện môi trường ở bào quan và nhân khác nhau nên trong quá trình tiến hoá, ADN ở bào quan có những điểm khác biệt so với ADN trong nhân. Ví dụ, cơ chế tái bản của ADN bào quan khác biệt so với ADN trong nhân, do đó tỷ lệ sai sót trong tái bản cũng khác nhau. Thực nghiệm đã chứng minh trong tế bào động vật có vú, ADN ty thể đột biến nhanh hơn nhiều so với ADN trong nhân. Ngược lại, trong tế bào thực vật, ADN ty thể lại đột biến chậm hơn so với ADN nhân.

Hầu hết phân tử ADN trong các bào quan ở dạng mạch vòng. Tuy nhiên cũng có một số trường hợp, chủ yếu ở sinh vật eukaryot bậc thấp, ADN trong bào quan có thể tồn tại ở cả hai dạng mạch vòng và mạch thẳng. Ví dụ, phân tử ADN trong ty thể của *Paramecium*, *Chlamydomonas* và một số loài nấm men luôn luôn là sợi ADN mạch thẳng.

Mỗi tế bào thường có chứa nhiều ty thể hoặc lục lạp. Hơn nữa, mỗi bào quan có thể có chứa nhiều phân tử ADN. Do đó, số lượng ADN ty thể (ADNmt) hoặc ADN lục lạp (ADNcp) có thể đạt đến hàng nghìn bản sao trong một tế bào. Ví dụ, một tế bào người có tới 800 ty thể, mỗi ty thể có khoảng 10 phân tử ADNmt, như vậy một tế bào có đến 8000 phân tử ADNmt. Tế bào trứng của động vật có vú có chứa tới 10^8 bản sao của ADNmt. Một cách tương tự, số lượng lục lạp trong mỗi tế bào thực vật thay đổi từ 20-40, mỗi lục lạp có chứa số lượng tương tự (20-40) phân tử ADNcp. Ở thực vật bậc thấp, số lượng lục lạp trong một tế bào cũng như số lượng ADNcp trong một lục lạp lớn hơn so với tế bào thực vật bậc cao. Ví dụ, vi tảo *Chlamydomonas* chứa khoảng 1000 phân tử ADN lục lạp trong một tế bào. Ngoài ra, kích thước phân tử ADN ở bào quan không tỷ lệ với tính phức tạp của cá thể. Kích thước phân tử ADNmt có thể thay đổi rất rộng từ 16-17kb ở động vật có xương sống đến 2500kb ở một số thực vật có hoa. Kích thước ADNcp ở thực vật thay đổi từ 200kb ở thực vật bậc thấp đến 140kb ở thực vật bậc cao.

Do kích thước nhỏ hơn nhiều so với genome trong nhân nên ADN ở các bào quan chứa số lượng gen ít hơn và các gen phân bố sát nhau hơn (khoảng cách giữa hai gen rất nhỏ, thậm chí chỉ vài nucleotide). Phân tử ADNmt hay ADNcp chứa những gen mã cho protein thực hiện chức năng chuyên hoá đặc thù của ty thể hay lục lạp. Ngoài ra ADN trong ty thể và lục lạp còn chứa gen mã cho ARNr, ARNt và protein ribosome dùng riêng cho bào quan.

1.5.1. ADN ty thể

Trình tự nucleotide của ADN ty thể (ADNmt) ở một số sinh vật đã được xác định. Kết quả này giúp chúng ta hiểu rõ hơn cấu trúc và sắp xếp các gen trên phân tử ADN của bào quan. Ở động vật có xương sống, ADNmt có kích thước nhỏ gồm các gen không có intron và hầu như không có khoảng trống giữa các gen. Ví dụ, ADNmt của người gồm 16.659bp tương ứng với 37 gen, trong đó 22 gen mã cho các phân tử ARNt, 13 gen mã cho các polypeptide liên quan đến phản ứng oxy hoá khử. Ở nấm men, ADNmt có kích thước lớn hơn so với động vật (78.000bp) do một số gen có intron và khoảng cách giữa các gen khá lớn. ADNmt nấm men có ít nhất 33 gen, trong số này có 2 gen mã cho ARNr, 23 gen mã cho ARNt, 1 gen mã cho protein ribosome và 7 gen mã cho polypeptide tham gia phản ứng oxy hoá khử. Đặc biệt, ADNmt của thực vật có kích thước lớn nhất và cấu trúc phức tạp đa dạng nhất. Trình tự ADNmt của *Marchantia polymorpha*, thực vật nguyên thuỷ không có hệ mao dẫn, đã được xác định hoàn toàn. Đây là phân tử mạch vòng có kích thước 186kb tương ứng với 94 khung đọc mở (ORFs). Trong số 94 ORFs này, thực nghiệm mới xác định được một số gen mà số lượng intron của một gen lên đến 32. Đối với thực vật có hệ mạch, ADNmt còn lớn hơn nhiều. Ví dụ, ở ngô hay dưa hấu, ADNmt tương ứng với 570kb và 300kb. Ở các loài thực vật bậc cao, các gen có thể phân bố ở vị trí khác nhau trên phân tử ADNmt mặc dù sản phẩm của gen có cùng một chức năng trong tế bào.

1.5.2. ADN lục lạp

Thực vật có ba loại lục lạp khác nhau tuỳ thuộc vào hợp chất mà chúng có như tinh bột, các sắc tố hoặc các chất béo. Cả ba loại này đều có chứa phân tử ADN (ADNcp) với kích thước thay đổi từ 85 đến 292kb ở tảo và 120 đến 160kb ở thực vật bậc cao. Đặc biệt ở một số thực vật như tảo xanh *Acetabularia*, ADNcp lớn đến 2000kb. Phân tử ADNcp của một số thực vật đã được xác định trình tự nucleotide. Lục lạp

thuốc lá *Nicotiana tobacum* có ADNcp gồm 155844bp tương ứng với khoảng 150 gen.

Số lượng phân tử ADNcp trong mỗi tế bào phụ thuộc vào số lục lạp trong một tế bào và số ADNcp trong mỗi lục lạp. Ví dụ, tế bào tảo đơn bào *Chlamydomonas reinhardtii* chỉ có một lục lạp chứa khoảng 100 phân tử ADNcp. Số gen phân bố trên ADNcp bao gồm gen mã cho ARNr, ARNt, protein ribosome và một số polypeptide tham gia phản ứng quang hợp, hấp thụ năng lượng ánh sáng mặt trời.

1.6. KHÁI NIỆM GEN

Khái niệm đầu tiên về gen được hình thành dựa vào kết quả thí nghiệm, trước hết là các thí nghiệm di truyền cổ điển của Mendel G.Johann (năm 1865). Từ phép lai giữa các cây đậu có những tính trạng khác nhau và theo dõi sự di truyền của chúng, Mendel đã đưa ra kết luận mỗi tính trạng được quyết định bởi các allen của một gen. Một gen có thể có nhiều allen cùng quy định một tính trạng. Tổ hợp của hai allen quy định mức độ biểu hiện của tính trạng đó. Tiếp theo, với một loạt các thí nghiệm tiến hành trên ruồi giấm *Drosophila,* Morgan T.Hunt và cộng sự (năm 1909) đã nhận thấy một số tính trạng được quyết định không phải do các allen của một gen mà do nhiều gen. Điều quan trọng hơn nữa, dựa vào tần số trao đổi chéo giữa hai nhiễm sắc thể tương đồng trong quá trình phân bào giảm nhiễm (meiosis), Morgan có thể lập được bản đồ di truyền (genetic map) cho phép xác định vị trí của gen trên nhiễm sắc thể. Hai gen càng gần nhau thì tần số trao đổi chéo giữa chúng càng nhỏ. Trên thực tế, bản đồ di truyền cho biết vị trí của những gen liên quan đến các tính trạng, hoặc các đột biến mà khoảng cách giữa chúng được tính bằng tần số trao đổi chéo (cM). Tuy nhiên, trao đổi chéo không xảy ra như nhau ở mọi vị trí trên sợi nhiễm sắc thể khiến cho khoảng cách giữa các vị trí trên bản đồ di truyền không phải lúc nào cũng tỷ lệ với tần số trao đổi chéo.

Trước năm 1940, vị trí các gen trên nhiễm sắc thể được xem như các hạt cườm trong một chuỗi. Trao đổi chéo được xem là chỉ xảy ra giữa các gen mà không thể xảy ra trong một gen. Vì vậy, các nhà di truyền học bước đầu tiên đã kết luận sơ bộ rằng gen là đơn vị di truyền nhỏ nhất có thể bị đột biến và một gen quy định một tính trạng và trao đổi chéo không thể xảy ra trong một gen. Như vậy hai tính trạng không giống nhau có thể phân biệt được thì phải do ít nhất hai gen khác nhau quy định. Rõ ràng, khái niệm về gen ban đầu này chỉ cho

phép xác định mối tương quan theo kiểu một đột biến - một tính trạng - một gen.

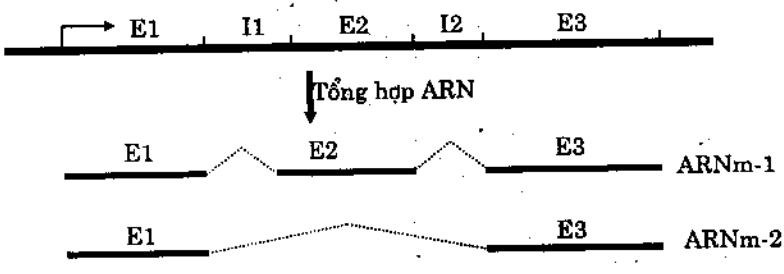
Trong thực tế, việc xác định tần số trao đổi chéo để tìm ra mối liên hệ giữa hai gen gặp rất nhiều khó khăn do phải sàng lọc các cá thể có trao đổi chéo từ số lượng rất lớn các cá thể ở thế hệ con cháu qua các phép lai khác nhau. Mặt khác, một tính trạng có thể được quy định bởi các gen riêng biệt. Nhược điểm này được khắc phục nhờ thí nghiệm bổ trợ chức năng (complementation tests). Ví dụ, khi kết hợp các tế bào nấm men *Neurospora* dạng đơn bội bị đột biến có cùng một biểu hiện là mất khả năng mọc trên môi trường thiếu histidine, các nhà di truyền nhận được một số tế bào lưỡng bội có thể phục hồi khả năng sinh sôi trên môi trường không có histidine. Kết quả phép lai giữa các dòng tế bào đột biến cho phép xác định tính trạng này liên quan đến hai gen khác nhau trong con đường sinh tổng hợp histidine. Dựa vào tần số trao đổi chéo, các gen này có vị trí phân bố ở những điểm khác nhau trên bản đồ di truyền. Như vậy, thí nghiệm bổ trợ chức năng cho phép phân biệt từng gen trong nhóm gen cùng quy định một tính trạng.

Các nghiên cứu tiếp theo do các nhà di truyền Clarence P. Oliver và Melvin M. Green thực hiện trên ruồi giấm đã phát hiện thấy trao đổi chéo có thể xảy ra ngay trong một gen. Đặc biệt nhờ vào việc khám phá ra cấu trúc ADN, Yanofsky C. và cộng sự lần đầu tiên đã đưa ra bằng chứng rõ ràng về trao đổi chéo xảy ra giữa các nucleotide của một gen khi nghiên cứu gen mã cho enzym tổng hợp tryptophan ở *E.coli*. Nhờ các kết quả đặc biệt quan trọng trên mà khái niệm về gen đã được hoàn thiện hơn. Lúc này gen được xem là một đoạn nucleotide mang mã di truyền cho các acid amin của một sợi peptide. Từ khái niệm ban đầu cho rằng, mỗi gen là một hạt cườm của một chuỗi (chuỗi đó chính là một nhiễm sắc thể trong genome) và trao đổi chéo cũng như đột biến chỉ xảy ra giữa các hạt cườm thì các nhà di truyền học đã tìm được mối liên hệ tuyến tính giữa các mã di truyền bộ ba của một gen với trật tự acid amin trên sợi polypeptide.

Khám phá ra mã di truyền bộ ba (*triplet codon*) được công bố vào những năm đầu thập kỷ 60 thế kỷ XX khi chuỗi polyU được sử dụng để tổng hợp *in vitro* chuỗi peptid polyphenylalanine. Ba nucleotide UUU là một mã di truyền của một acid amin phenylalanine. Như vậy, tổ hợp của bốn loại nucleotide (A, C, G, T) sẽ tạo ra 64 mã bộ ba, trong đó 61 mã là của 20 acid amin, 3 mã còn lại (*stop codon*) dùng để kết thúc quá trình tổng hợp protein. Số mã bộ ba nhiều hơn số acid amin nên có

hiện tượng nhiều mã bộ ba mã cho cùng một acid amin. Hiện tượng này được gọi là thoái hoá mã. Ví dụ, alanine được mã bởi 4 mã GCN trong đó N=A, C, G, T. Ngoài ra, trong một số trường hợp hiếm gặp ở vi khuẩn *E.coli*, mã dừng UGA được dịch mã cho acid amin selenocysteine hoặc mã dừng UAG được dịch mã cho pyrrolysine. Đây là các acid amin không nằm trong số 20 acid amin đã biết. Phân tử ARNm mã cho các protein có chứa selenocysteine thường có cấu trúc không gian SECIS (selenocysteine insertion sequence) khiến cho bộ máy tổng hợp protein nhận biết mã UGA không phải là stop codon. Cấu trúc SECIS thường gặp trong vùng chứa mã di truyền đối với ARNm vi khuẩn nhưng lại hay phân bố trong vùng 3' không dịch mã của ARNm eukaryot.

Mã di truyền bộ ba hầu như không thay đổi và thông dụng từ vi khuẩn đến các sinh vật eukaryot bậc cao. Hiện tượng sai lệch của mã xảy ra không nhiều. Ví dụ, ở động vật nguyên bào (mycoplasma), mã dừng UGA được quy định cho tryptophan. Sai lệch mã bộ ba xảy ra có tính hệ thống với các gen của ty thể. Tuy nhiên, do mã di truyền có tính chất thoái hoá nên không phải tất cả các mã của một acid amin có tần suất sử dụng như nhau ngay trong một loài và giữa các loài. Ví dụ, có 6 mã bộ ba cho một acid amin arginine, trong đó mã CGU được vi khuẩn sử dụng nhiều nhất (0,42), được sử dụng rất ít (0,09) ở người.



Hình 1.8: Phiên mã ở gen eukaryot.

Quá trình lựa chọn, cắt các intron (I) và nối các exon (E) theo các thứ tự khác nhau để tạo ra các phân tử ARNm chỉ giống nhau ở một số exon (E1 và E3). Chúng mã cho hai chuỗi polypeptide có chức năng khác nhau trong tế bào.

Những nghiên cứu trên gen eukaryot đã vấp phải những vướng mắc không thể giải thích theo khái niệm về gen đã nêu ở trên. Ví dụ, mã di truyền trên gen eukaryot không nằm kế tiếp nhau liên tục như ở gen vi khuẩn. Những mã đó bị ngắt đoạn bởi những trình tự không

chứa mã. Các đoạn chứa mã được gọi là exon, đoạn không chứa mã được gọi là intron (Hình 1.8). Hơn nữa, một số đột biến nằm trong vùng không chứa mã di truyền nhưng có thể gây ảnh hưởng đến tính trạng. Một số gen không mã cho sản phẩm protein mà chỉ phiên mã thành các sợi ARN. Đột biến ở những gen đó cũng gây tác động đến tính trạng do các gen khác quy định. Vì vậy, từ các khái niệm về gen được hình thành và thay đổi dần để phù hợp với các kết quả thí nghiệm, sinh học phân tử ngày nay định nghĩa một gen như sau: ***Gen là một đoạn ADN cần thiết cho sự tổng hợp một polypeptide có hoạt tính, hoặc một phân tử ARN liên quan đến hoạt động của tế bào.*** Như vậy, một gen không phải chỉ bao gồm vùng chứa mã di truyền (codon region) mà còn gồm các đoạn ADN (các vùng ADN điều khiển (regulatory elements), cần thiết cho việc phiên mã. Mặt khác, có những đoạn ADN có cấu trúc hay trình tự nucleotide rất giống gen nhưng chúng không được phiên mã, hoặc không biểu hiện chức năng gì nên chúng không thể được xem là gen.

Có rất nhiều câu hỏi thú vị được đặt ra như cấu trúc chung của các gen trong genome prokaryot và eukaryot khác nhau như thế nào? Các intron của gen eukaryot có nhiệm vụ gì? Tổng số các gen của cơ thể chiếm bao nhiêu phần trăm kích thước của genome? Có bao nhiêu gen hoạt động chuyên biệt trong một loại tế bào? Có bao nhiêu gen hoạt động trong nhiều loại tế bào biệt hoá khác nhau? Có phải tất cả các gen đều cần thiết cho sự tồn tại của tế bào bởi vì một số gen bị đột biến không gây ra thay đổi trong hoạt động sống của tế bào. Phải chăng những gen đó chỉ liên quan đến phản ứng linh hoạt của cơ thể trong điều kiện sống biến thiên không ngừng? Để có thể phần nào trả lời được những câu hỏi trên, chúng ta có thể sử dụng kỹ thuật xác định trình tự genome để tìm ra các khung đọc mở ORF, tức là các trình tự có khả năng phiên mã. Ngoài ra, chúng ta có thể phân tích tất cả các loại ARN của các gen hoạt động trong từng loại tế bào, tức là hướng nghiên cứu ***transcriptome***, hay phân tích tất cả các protein có mặt trong tế bào, hướng nghiên cứu ***proteome***. Tuy nhiên, điều cần chú ý là: do nhiều loại ARNm eukaryot có thể được hình thành nhờ phiên mã từ một gen và trải qua cắt nối luân phiên (trình bày chi tiết ở chương 2) cho nên số lượng ARN có thể nhiều hơn số lượng gen trong genome. Hơn nữa, một số gen mã cho multiprotein nên số lượng protein nhiều hơn số gen.

1.6.1. Các gen trong genome vi khuẩn

Vi khuẩn sống ký sinh trong tế bào chủ đều có kích thước genome

nhỏ hơn 1,5Mb. Chúng có số lượng gen ít nhất do đã cắt giảm bớt những gen mã cho các sản phẩm chuyển hoá chất dinh dưỡng. Những chất này do tế bào chủ cung cấp. Ví dụ, ký sinh trùng đơn bào *Mycoplasma genitalium* được xem như là sinh vật có genome nhỏ nhất và có số lượng gen tối thiểu cần thiết để duy trì sự sống. So sánh số lượng gen phân bố trong những genome có kích thước nhỏ nhất như *M. genitalium, M. pneumoniae* cho phép đánh giá được số lượng gen tối thiểu cần thiết để duy trì sự sống cho cơ thể đơn giản nhất. *M. genitalium* có 470 gen và *M. pneumoniae* có 679 gen. So sánh các gen và chức năng của chúng ở hai loại vi khuẩn này cho phép ước tính số gen tối thiểu cần có để duy trì sự sống là 256 gen. Tuy nhiên nhờ kỹ thuật di truyền phân tử gây đột biến định hướng chính xác từng gen, thực nghiệm đã xác định rằng ít nhất cần có 300 gen để đảm bảo sự sống cho vi sinh vật đơn giản nhất.

Các vi khuẩn cổ được xem như là dạng tồn tại trung gian giữa prokaryot và eukaryot. Ví dụ, *M. janaschiii* phân bào tương tự như prokaryot nhưng bộ máy kiểm soát hoạt động của các gen "vận hành" giống như trong tế bào eukaryot. Tuy nhiên, kích thước genome cũng như số lượng gen trong tế bào vi khuẩn cổ không nhiều hơn so với vi khuẩn. Kích thước genome vi khuẩn cổ thay đổi từ 1,5-3Mb, tương ứng với số gen từ 1500-2700 gene.

Vi khuẩn chịu nhiệt (thermophile) *Aquyfex aeolicus* được xem là vi sinh vật tồn tại trong tự nhiên có kích thước genome nhỏ nhất 1,5Mb với số lượng gen cần thiết là 1512 gen. Vi khuẩn Gram (-) điển hình *Haemophilus influenzae* có 1743 gen. Do đó, chúng ta có thể xem 1500 gen là số lượng gen tối thiểu cần có đối với vi khuẩn tồn tại trong tự nhiên. Một số vi khuẩn có kích thước genome và số lượng gen lớn tương tự như nấm men (eukaryot bậc thấp). Ví dụ, vi khuẩn *S. meliloti,* hoặc *M.loti* sống cộng sinh cố định nitơ trong rễ cây có genome lớn hơn 7Mb và có khoảng hơn 6000 gen. Đây là những vi khuẩn có genome và số lượng gen lớn nhất.

E.coli được nghiên cứu chi tiết nhất và được xem là đối tượng mô hình của di truyền, hoá sinh và sinh học phân tử. Genome *E.coli* gồm 4.639.221bp với 4288 trình tự có các đặc tính cấu trúc của gen mã cho protein (putative protein coding sequences). Một phần ba số trình tự này đã được xác định là các gen trong khi 38% chưa biết được chức năng. Các trình tự nucleotide được giả định là gen nhưng chưa biết sản phẩm protein mà chúng mã cho thì được gọi chung là khung đọc

mở (*open reading frame-**ORFs***). Một khung đọc mở thường bắt đầu bởi bộ ba mã di truyền cho methynonine (start codon) và kết thúc bởi một trong số ba mã dừng tổng hợp protein (stop codon).

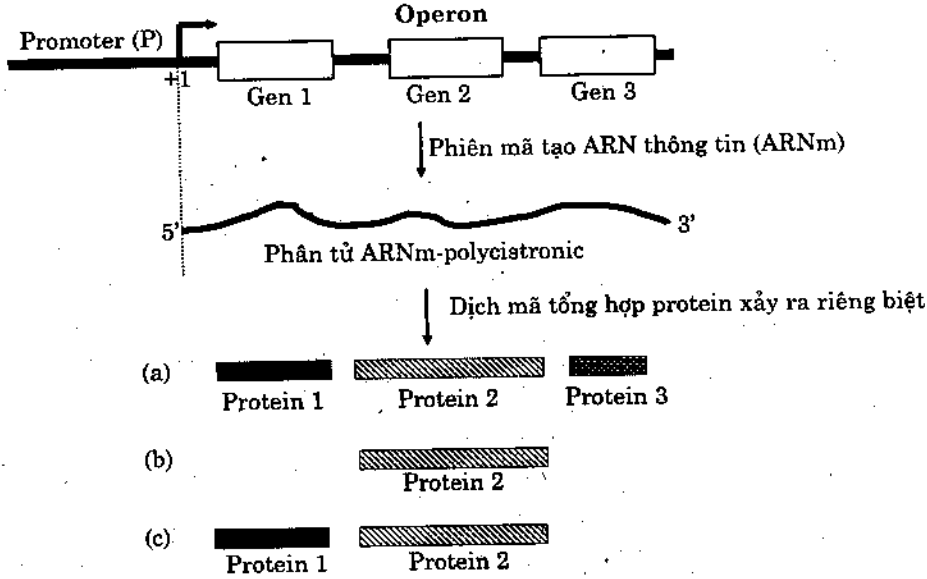
Kết quả so sánh trình tự nucleotide của toàn bộ genome *E.coli* với các trình tự ADN lưu trữ trong ngân hàng dữ liệu cho phép phát hiện 6 gen mới mã cho ARNt, 12 gen liên quan đến sinh tổng hợp và lắp ráp roi (flagella) cũng như 2 gen mã cho các enzym tham gia vào con đường phân hủy các hợp chất hữu cơ vòng. Rõ ràng việc so sánh trình tự genome giữa *E.coli* và các sinh vật prokaryot khác đặc biệt có ý nghĩa trong việc xác định các gen mới cũng như chức năng của chúng. Ngoài ra, khi so sánh số lượng gen tham gia vào một quá trình sinh học ở các vi khuẩn khác nhau cho phép đánh giá số lượng gen tối thiểu cần thiết cho quá trình đó. Ví dụ, quá trình trao đổi chất liên quan đến khoảng 243 gen ở *E.coli*, 112 gen ở *Haemophilus influenzae* nhưng chỉ cần đến 31 gen ở *Mycoplasma genitalium*. Ngoài ra, việc so sánh các gen giống và khác nhau giữa các vi khuẩn có quan hệ gần gũi trong tiến hoá đặc biệt có ý nghĩa để xác định những gen riêng biệt của từng loài, tức là những gen chỉ thị dùng để phân biệt loài này với loài kia. Ví dụ, trong số 470 gen có ở *M. genitalium* thì 350 gen cũng tồn tại ở *Bacillus subtilis*. Như vậy chỉ có 120 gen tạo nên sự khác biệt giữa hai loại vi khuẩn này. Tuy nhiên, những nghiên cứu về cách thức hoạt động của các gen riêng biệt này, chức năng của từng sản phẩm protein mà gen mã cho, cũng như các quá trình hoá sinh mà chúng tham gia chưa đưa đến kết luận rõ ràng về vai trò của 120 gen đặc thù cho *M. genitalium*.

Mặc dù có kích thước nhỏ hơn nhiều so với genome eukaryot, nhưng genome prokaryot có mật độ phân bố các gen cao hơn. Khoảng 85-90% genome chứa các gen. Như vậy số đoạn ADN không chứa mã di truyền ít hơn, tức là khoảng cách giữa các gen ngắn hơn. Ví dụ, khoảng cách trung bình giữa hai gen ở *E.coli* là 118bp. Các gen và ORFs chiếm 87,8%; các gen mã cho ARNs chiếm 0,8%; còn thành phần ADN lặp lại không chứa gen chỉ chiếm có 0,7%. Mặt khác hầu hết các gen ở prokaryot đều tồn tại đơn bản (single-copy gen) và các gen không bị gián đoạn bởi những đoạn không chứa mã di truyền (các intron).

Trong genome prokaryot, một số gen phân bố cạnh 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, 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 1.9).



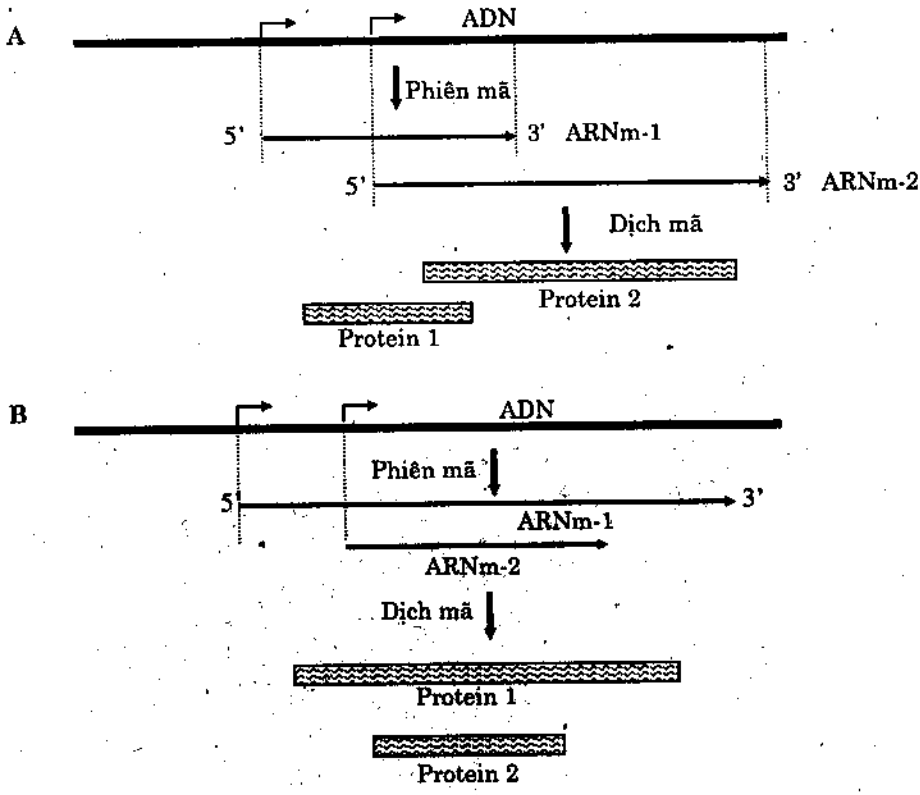
Hình 1.9: Cấu trúc operon trong genome vi khuẩn.

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 phiên mã sang phân tử ARNm; P- promoter).*

1.6.2. Các gen trong genome virus

Khi virus xâm nhiễm tế bào chủ chỉ có genome của virus được đưa vào bên trong tế bào chủ. Tế bào chủ có thể trải qua hai trạng thái,

hoặc bị vỡ, hoặc genome bị chèn thêm genome của virus. Đối với trạng thái đầu tiên (gọi là trạng thái sinh tan), genome virus được phiên mã, tái bản và tạo ra các thể virus mới. Lúc đó tế bào chủ bị tan vì các virus mới giải phóng ra ngoài môi trường. Đối với trạng thái thứ hai (trạng thái tiềm tan), genome virus sẽ chèn vào genome tế bào chủ; do đó có thể di truyền cho các tế bào con. Trạng thái tiềm tan có thể chuyển đổi thành sinh tan trong những điều kiện thích hợp.



Hình 1.10: Hai protein được mã bởi một đoạn ADN duy nhất do hai khung đọc nằm gối lên nhau (A) hoặc do khung đọc của gen thứ hai nằm trong khung đọc của gen thứ nhất (B).

Quá trình sinh tan liên quan đến hai nhóm gen của virus. Nhóm các gen hoạt động ngay sau khi virus vừa xâm nhiễm vào tế bào chủ và nhóm gen hoạt động tiếp theo sau khi genome virus bắt đầu tái bản. Hai nhóm này được gọi là nhóm gen sớm (early genes) và nhóm gen muộn (late genes). Nhóm gen sớm mã cho các nuclease phân huỷ ADN của vi khuẩn và các protein làm nhiệm vụ ức chế hoạt tính của ADN polymerase, ARN polymerase và một số protein của tế bào chủ.

Đặc biệt một số protein của nhóm gen sớm có chức năng biến đổi ARN polymerase của vi khuẩn khiến cho enzym này không nhận biết được promoter của gen vi khuẩn. Vì thế hoạt động của các gen vi khuẩn bị kìm hãm. Hơn thế, ARN polymerase của vi khuẩn tiếp tục bị biến đổi để ngừng phiên mã các gen sớm nhưng nhận biết promoter của các gen muộn của virus.

Genome của thực khuẩn thể ϕX174 là một phân tử ADN sợi đơn dạng vòng gồm 5386 nucleotide. Thực nghiệm cho thấy genome ϕX174 mã cho 11 protein với tổng số 2300 acid amin. Tuy nhiên, toàn bộ genome ϕX174 chỉ có thể tương ứng với 1795 mã bộ ba khi mọi nucleotide đều tham gia các mã (5386 nucleotide/3 nucleotide của một mã =1795 mã bộ ba). Do đó, số mã bộ ba của toàn bộ genome ϕX174 không đủ cho 2300 acid amin. Thực nghiệm đã chứng minh sự thiếu hụt này là do hai gen nằm gối lên nhau (overlapping genes), hoặc do gen này nằm trong gen kia (genes-within-genes). Hiện tượng hai gen nằm gối lên nhau và gen nằm trong gen đều liên quan đến một đoạn ADN có hai khung đọc (ORF) lệch nhau, mỗi khung tương ứng với một protein (Hình 1.10).

1.6.3. Các gen trong genome eukaryot

Một trong những sai khác cơ bản trong cấu trúc gen giữa sinh vật prokaryot và eukaryot là hiện tượng gen bị gián đoạn (interupted gene). Hiện tượng này được khám phá lần đầu tiên năm 1977 và được tìm thấy phổ biến ở mọi sinh vật eukaryot. Sau đó hiện tượng này cũng được phát hiện ở một số thực khuẩn thể (bacteriophage). Khi so sánh trình tự nucleotide trên một gen với phân tử ARNm được phiên mã từ gen đó, các nhà khoa học phát hiện thấy gen có chứa những đoạn không mang mã di truyền. Những đoạn này không tìm thấy trong phân tử ARNm được sử dụng làm khuôn để tổng hợp protein. Chúng được gọi là các intron. Như vậy bên cạnh việc chứa mang mã di truyền (gọi là exon), đa số các gen eukaryot còn chứa các intron. Các intron được phiên mã sang phân tử tiền thân ARN nhưng tiếp đó chúng bị cắt ra khỏi phân tử ARN. Do các intron không mang mã di truyền nên đột biến xảy ra trên chúng thường không được biểu hiện ở cấu trúc của chuỗi polypeptide. Tuy nhiên, một số đột biến xảy ra ở intron có thể ngăn cản phản ứng cắt bỏ intron, do đó tạo nên phân tử ARNm sai hỏng không sử dụng được để dịch mã tổng hợp protein.

Ở sinh vật bậc cao, các gen mã cho protein hay các ARNt, ARNr hầu như đều bị gián đoạn. Độ dài trung bình của exon khoảng 200bp

trong khi của intron có thể lớn hơn 10kb hoặc thậm chí đạt tới 50-60kb. Ngoài ra, hiện tượng các gen nằm gối lên nhau (overlapping genes) rất hiếm xảy ra ở ADN nằm trong nhân tế bào eukaryot. Hiện tượng này hay gặp trong genome prokaryot và với gen phân bố trong các bào quan của tế bào eukaryot.

Kích thước genome prokaryot thường tỷ lệ thuận với số lượng gen phân bố trong genome. Sự tồn tại của các intron kích thước rất lớn so với exon đã phần nào khiến cho sự tỷ lệ thuận đó bị mất đi. Thực tế cho thấy sinh vật eukaryot bậc cao có số lượng gen lớn hơn so với eukaryot bậc thấp nhưng sự tăng số gen không tỷ lệ với sự tăng bất thường kích thước genome. Nấm men *S.pombe* và *S.cerevisiae* được xem là cơ thể đơn bào eukaryot đơn giản nhất có số gen khoảng 5000-6000 gen. Như vậy, số gen của nấm men không nhiều hơn của vi khuẩn cộng sinh nhưng kích thước genome nấm men tăng gần gấp 2 lần (12,5-13Mb). Điều thú vị là chỉ khoảng 5% các gen của *S.cerevisiae* trong khi 43% gen của *S.pombe,* chứa intron. Điều khó lý giải hơn nữa là ruồi giấm *D.melanogaster* được xem là sinh vật phức tạp hơn giun tròn *C.elegans* trong tiến hoá, nhưng số lượng gen của ruồi giấm (13.600 gen) chỉ chiếm khoảng 70% của giun tròn (18.500 gen), mặc dù ruồi giấm có genome lớn hơn. Thực vật bậc cao như lúa có khoảng 8000 gen không xuất hiện ở prokaryot hoặc động vật có vú. Dường như những gen này đảm bảo chức năng đặc hiệu cho thực vật như quá trình quang hợp.

Bên cạnh việc xuất hiện intron làm tăng kích thước gen eukaryot, kích thước chuỗi peptide cũng tăng lên từ prokaryot đến eukaryot. Chuỗi peptide ở vi khuẩn cổ và *E.coli* có kích thước trung bình khoảng 270-300 acid amin, ở nấm men *S. cerevisiae* và giun tròn *C.elegans* khoảng 440-480 acid amin. Số lượng protein có kích thước lớn (500 acid amin) rất hiếm ở prokaryot nhưng chiếm đến 1/3 tổng số protein ở eukaryot. Ngoài ra, khoảng cách giữa các gen khác nhau cũng tạo nên sự tăng kích thước genome trong khi số lượng gen không thay đổi. Các gen prokaryot phân bố rất gần nhau, thậm chí chỉ cách nhau 1 nucleotide. Tuy nhiên, các gen eukaryot phân bố xa nhau từ vài kb đến hàng chục kb. Những trình tự nằm giữa các gen thường liên quan đến các cơ chế kiểm soát hoạt động của gen (trình bày chi tiết trong chương 2).

Phản ứng cắt nối intron-exon xảy ra với các loại ARNm, ARNr và ARNt của tế bào eukaryot. Để tạo ra phân tử ARN hoàn thiện, việc cắt bỏ intron, nối các exon tuân theo những quy luật nghiêm ngặt và

chính xác để đảm bảo thứ tự của chúng. Đối với đa số các gen eukaryot, các exon từ phân tử ARN tiền thân được nối với nhau tạo nên phân tử ARN hoàn chỉnh. Hiếm trường hợp nối các exon của các phân tử ARNm khác nhau. Điều đáng lưu ý với ADN của ty thể và lục lạp, intron của gen này có thể là exon của gen khác và sản phẩm protein của hai gen đó có chức năng hoàn toàn độc lập. Ngoài ra, một số gen được phiên mã tạo phân tử ARNm nhưng chúng không được dịch mã. Những phân tử ARNm này vẫn trải qua phản ứng cắt nối intron-exon tạo nên phân tử ARN hoàn chỉnh. Tuy nhiên, phân tử ARN này tiếp tục được phân huỷ thành các phân tử ARN kích thước nhỏ 21-25 nucleotide (micro RNAs – miRNAs). Các phân tử miRNAs tham gia vào nhiều quá trình kiểm soát hoạt động của một số gen trong genome, chủ yếu ở quá trình sau phiên mã. Trong cơ chế kiểm soát này, miRNAs làm nhiệm vụ nhận biết ARNm của một số gen khác để phân hủy các ARNm này. Đây là một cơ chế kiểm soát hoạt động của gen sau phiên mã được phát hiện vào những năm cuối thập kỷ 20 thế kỷ XX.

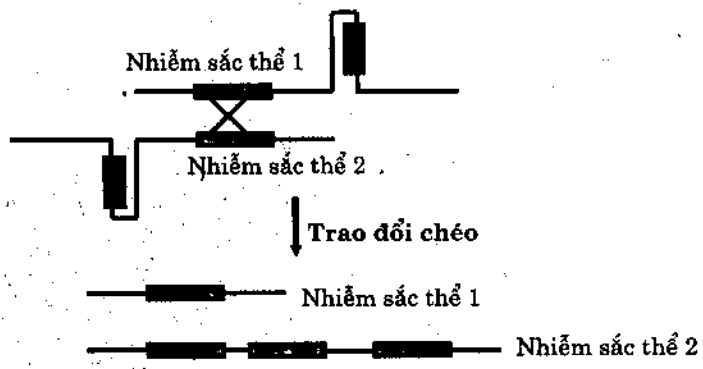
1.7. PHÂN LOẠI GEN

Có thể phân loại gen tùy theo số lượng bản sao của gen có mặt trong genome, hoặc theo chức năng của các sản phẩm do gen mã hoá. Khoảng một nửa các gen đã biết trong genome động vật có xương sống đều có các bản sao giống hệt hoặc tương tự (số bản copy có thể từ 2 đến 20). Khoảng 1/4 số gen mã cho protein là các gen đơn bản (single copy gene), không tồn tại bản sao thứ hai. Các gen có trình tự nucleotide và chức năng tương đồng có thể xếp vào họ gen. Ngoài ra còn có những trình tự nucleotide giống với một gen đã biết nhưng trình tự đó không được phiên mã hoặc không được dịch mã. Chúng được gọi là giả gen (pseudogen). Một số gen gồm nhiều bản sao giống hệt nhau lặp đi lặp lại liên tục trên một vùng nhiễm sắc thể (tandem repeat genes). Ví dụ như, gen mã cho ARNr, ARNt, histone vv... Như vậy, các gen eukaryot có thể phân thành các loại chính như sau: gen đơn bản (single-copy), các gen thuộc một họ gen (family genes), gen lặp đi lặp lại liên tục (tandem repeated genes) và các gen giả (pseudogen).

1.7.1. Các gen trong cùng một họ gen

Các gen hình thành từ một gen tổ tiên ban đầu được xếp vào một họ gen. Các bản sao trong họ gen không bắt buộc phải giống nhau

hoàn toàn do trong quá trình tiến hoá chúng chịu những đột biến như thêm, mất, thay thế hoặc chuyển đoạn các nucleotide. Các gen trong cùng một họ có thể phân bố tập trung thành một nhóm (trên một nhiễm sắc thể) hoặc phân tán trong genome (trên các nhiễm sắc thể khác nhau). Sản phẩm của các thành viên trong một họ có chức năng giống hệt nhau hoặc có liên quan đến nhau. Tuy nhiên các gen này thường hoạt động ở những thời điểm nhất định và trong các loại tế bào biệt hoá khác nhau. Khi một thành viên trong họ gen bị bất hoạt, thành viên khác có thể được hoạt hoá thay thế mặc dù bình thường thành viên thứ hai không hoạt động cùng với gen ban đầu. Hiện tượng tồn tại nhiều bản sao giống hoặc tương tự của một gen có thể gây ra do sai lệch trong trao đổi chéo giữa hai nhiễm sắc thể tương đồng trong phân bào giảm nhiễm (meiosis). Điều đó làm cho một nhiễm sắc thể có số lượng bản copy tăng lên trong khi nhiễm sắc thể kia có số lượng giảm đi (Hình 1.11).



Hình 1.11: Sai lệch trong trao đổi chéo giữa hai nhiễm sắc thể (mỗi nhiễm sắc thể có hai bản sao của một gen) khiến một nhiễm sắc thể chỉ mang một bản sao trong khi nhiễm sắc thể thứ hai mang ba bản sao.

Các gen globin của người là ví dụ điển hình về một họ gen. Tế bào trong cơ thể trưởng thành có globin tồn tại ở dạng tetramer gồm hai chuỗi polypeptide α và chuỗi β. Tế bào máu của phôi cũng chứa globin ở dạng tetramer nhưng gồm hai chuỗi tương tự α và tương tự β. Các gen mã cho chuỗi α và chuỗi tương tự α đều thuộc một họ gen trong khi các gen mã cho chuỗi β và chuỗi tương tự β thuộc họ gen khác. Hai họ gen nằm trên hai nhiễm sắc thể khác nhau. Họ gen globin α chiếm 28kb trên nhiễm sắc thể 16, gồm 4 gen ξ, α1, α2 và θ. Sản phẩm của hai gen α1, α2 giống hệt nhau. Họ gen globin β chiếm 50kb trên

nhiễm sắc thể 11 gồm 5 gen hoạt động (ε, Gγ, Aγ, δ, β) và một pseudogen ψ_β. Sản phẩm của hai gen Aγ và Gγ chỉ khác nhau duy nhất ở một acid amin tại vị trí 136 (Glicine và Alanine). Hoạt động của các gen trong hai họ gen α và β phải được phối hợp đồng thời sao cho số lượng hai loại polypeptide được tạo ra một cách tương đồng với nhau về mặt số lượng. Ngoài ra, trong mỗi họ còn có một số thành viên khác mà sản phẩm của chúng đôi khi vẫn được sử dụng. Các chuỗi polypeptide liên kết với nhau tạo ra các dạng globin không giống nhau và được sử dụng ở những giai đoạn phát triển khác nhau của cơ thể (Bảng 1.2).

Bảng 1.2: Các dạng globin thay đổi trong quá trình phát triển ở người

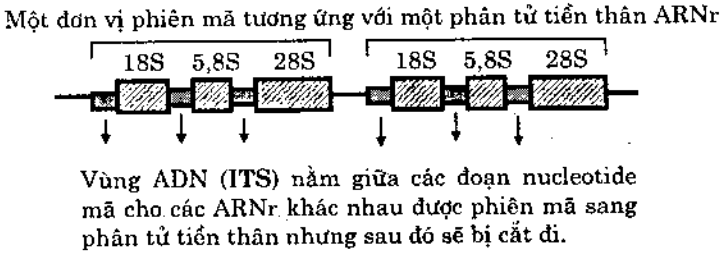
Giai đoạn phát triển	Hemoglobin
Mô phôi (8 tuần)	ξ2ε2, ξ2γ2, α1ε2
Thai nhi (3-9 tháng)	α2γ2
Cơ thể trưởng thành (từ khi sinh)	α2δ2 (~ 2%), α2β 2 (~97%), α2γ2 (~1%)

Bên cạnh họ gen mã cho globin tập trung tại hai vùng trên nhiễm sắc thể 11 và 16, họ gen mã cho aldolase được xem là ví dụ điển hình về sự phân bố rải rác của một họ gen trên các nhiễm sắc thể khác nhau. Họ gen này gồm 5 gen thành viên phân bố trên 5 nhiễm sắc thể 3, 9, 10, 16 và 17. Mặc dù phân tán trong khắp genome, các gen này có độ tương đồng rất cao về trình tự nucleotide cũng như trình tự acid amin tương ứng.

1.7.2. Gen lặp đi lặp lại liên tục

Trong một số trường hợp cá biệt, số lượng các thành viên trong họ gen rất lớn và chúng có trình tự nucleotide giống hệt nhau. Các thành viên của họ gen thường tập hợp thành các nhóm phân bố trên các nhiễm sắc thể khác nhau. Mỗi nhóm có thể bao gồm từ hai đến hàng trăm gen, gen nọ nối tiếp gen kia. Việc lặp đi lặp lại liên tiếp các bản sao của một gen trên một đoạn ADN (trên một vùng nhiễm sắc thể) có thể nhằm mục đích đáp ứng nhanh, đủ số lượng rất lớn sản phẩm của gen khi tế bào yêu cầu. Ví dụ, cần đáp ứng kịp thời các phân tử ARN ribosome cho giai đoạn sinh trưởng nhanh (tổng hợp protein), hoặc các loại protein histone cho quá trình tái bản ADN. Số lượng các gen mã cho ARNr thay đổi từ 7 ở *E.coli*, 100-200 ở eukaryot bậc thấp đến vài

trăm ở động vật bậc cao. Trong nhân tế bào eukaryot, hầu hết các gen mã cho ARNr tập trung thành từng nhóm chiếm một vùng trên nhiễm sắc thể (vùng ADNr). ARNr gồm các loại chính ARNr-5S, ARNr-5,8S, ARNr-18S và ARNr-28S (tương ứng với hai tiểu phần nhỏ và lớn của ribosome). Phân tử ARNr-5S được mã bởi gen riêng biệt và được tổng hợp bởi ARN polymerase III. Genome của người có chứa khoảng 2000 gen mã cho ARNr 5S. Tất cả các gen này đều tập trung trên một vùng của nhiễm sắc thể số 1. Ba loại ARNr 5,8S, 18S và 28S được tổng hợp từ một gen bởi ARN polymerase I (Hình 1.12).

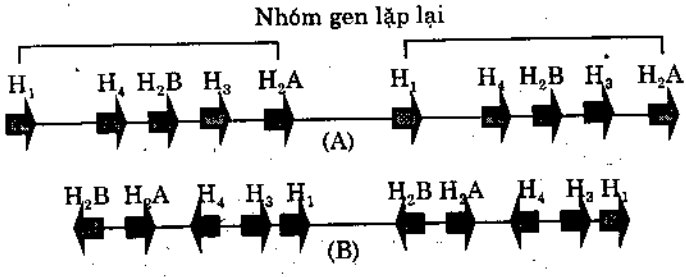


Hình 1.12: Một đơn vị phiên mã (một gen mã cho ARNr) mang thông tin di truyền cho ba loại phân tử ARNr 18S, 5,8S và 28S. Gen này được lặp đi lặp lại liên tục. Khoảng cách giữa các gen thay đổi tuỳ theo từng loài sinh vật.

Hình 1.12 biểu diễn cấu trúc lặp lại liên tiếp của gen mã cho ARN ribosome. Phân tử tiền thân ARNr được phiên mã từ gen, sau đó bị cắt bởi các ribonuclease tạo thành các phân tử ARNr 18S, 5,8S và 28S. Đoạn nucleotide nằm giữa các phân tử ARN này sẽ bị phân hủy. Mỗi một gen mã cho ARNr được gọi là một đơn vị phiên mã. Các đơn vị phiên mã giống hệt nhau và lặp lại liên tiếp, khoảng cách giữa chúng thay đổi tuỳ theo loài, thậm chí ngay trong cùng một loài. Genome ở người có khoảng 280 bản sao của gen mã cho ba loại ARNr, tập trung thành 5 vùng (mỗi vùng có từ 50-70 bản copy), phân bố trên 5 nhiễm sắc thể 13, 14, 15, 21 và 22. Ở động vật có vú, mỗi gen thường chiếm 13kb, nằm cách nhau khoảng 30kb. Khoảng cách này có vai trò trong khởi động quá trình tổng hợp ARNr hoặc giúp cho ARN polymerase dễ dàng bám vào promoter.

Gen mã cho protein histone cũng tồn tại nhiều bản sao lặp lại liên tiếp trong genome. Protein histone tham gia liên kết với ADN để hình thành cấu trúc nucleosome. Có bốn loại histone khác nhau H_2A, H_2B, H_3 và H_4 tương tác với nhau tạo cấu trúc lõi. Lõi này được quấn quanh bởi đoạn ADN 146bp tạo thành nucleosome. Histone H_1 liên kết với

ADN nằm giữa các nucleosome. Histone chiếm khoảng 0,5-1% tổng số protein của tế bào eukaryot. Việc tổng hợp protein này xảy ra trong suốt 1/3 chu kỳ tế bào (ở pha S). Tuy nhiên phân tử ARNm histone có thời gian bán sống ngắn (vài phút). Có lẽ vì lý do đó, có rất nhiều gen mã cho histone (50-500) phân bố thành các nhóm trên nhiễm sắc thể. Chúng nằm nối tiếp nhau, mỗi nhóm chiếm khoảng 5-6kb (ở động vật có xương sống) (Hình 1.13). Cũng giống như nhóm gen mã cho ARNr, khoảng cách giữa các gen trong cùng một nhóm và khoảng cách giữa các nhóm thay đổi từ loài này sang loài khác, thậm chí thay đổi ngay trong từng cá thể.



Hình 1.13: Bản đồ phân bố các gen mã cho histone ở Cầu gai (A) và ở Ruồi giấm (B). Mỗi nhóm được lặp đi lặp lại trên một vùng nhiễm sắc thể. Chiều tổng hợp ARNm cho mỗi loại histone không giống nhau (chiều mũi tên), chứng tỏ ngay trong một nhóm, các gen hoạt động độc lập nhau.

Có thể phân biệt các gen mã cho histone thành hai nhóm. *Nhóm gen thứ nhất* gồm các gen mã cho histone dùng trong quá trình tái bản ADN. Nhóm gen này không có intron và phân tử ARNm phiên mã từ chúng không có đuôi polyA. Đây là điều khác biệt với các ARNm eukaryot. *Nhóm gen thứ hai* gồm những gen mã cho histone tham gia vào quá trình biến đổi cấu trúc không gian của nhiễm sắc thể (liên quan đến thông tin di truyền ngoại sinh). Các gen thuộc nhóm này có chứa intron và phân tử ARNm tương ứng có gắn đuôi polyA.

1.7.3. Pseudogen (gen giả)

Mọi thành viên trong một họ gen đều có thể hoạt động tùy thuộc trạng thái tế bào. Tuy nhiên có những thành viên mà không bao giờ phát hiện được sản phẩm của chúng mặc dù chúng giống hệt hoặc có trình tự nucleotide tương đồng rất cao với các thành viên khác. Những gen đó được gọi là các ***Pseudogen*** (tạm dịch là các gen giả, thường ký hiệu là ψ).

Pseudogen không tạo được sản phẩm cuối cùng là protein, mặc dù chúng có thể được phiên mã tổng hợp ARNm. Cấu trúc pseudogen có thể chỉ gồm toàn exon, hoặc gồm các exon và intron, hoặc có trình tự nucleotide giống hệt hay tương tự các gen hoạt động khác nhưng không có promoter. Thực nghiệm cho thấy đột biến đã xảy ra ở các pseudogen khiến quá trình phiên mã không thể khởi động được, hoặc khiến quá trình tổng hợp ARNm dừng không đúng chỗ, hoặc ngăn cản phản ứng cắt nối intron-exon tạo phân tử ARNm. Thậm chí ngay khi phân tử ARNm được tạo ra, nó đã chứa các tín hiệu làm dừng quá trình tổng hợp protein sớm hơn cần thiết.

Hầu hết các họ gen đều có các pseudogen, mặc dù với số lượng rất nhỏ. Các gen này có thể xuất hiện do sai lệch trong trao đổi chéo giữa các allen của hai nhiễm sắc thể tương đồng. Theo thời gian, các đột biến thêm, bớt, chuyển đoạn hoặc thay thế nucleotide ngày càng tích tụ trên các pseudogen. Ngoài ra không thể loại trừ khả năng enzym reverse transcriptase tổng hợp phân tử ADN trên khuôn mẫu các ARNm và các bản sao ADN này được ghép vào genome. Do đó, pseudogen thường không có promoter, không chứa intron.

1.8. THÀNH PHẦN ADN KHÔNG CHỨA GEN TRONG GENOME

Như đã trình bày trong phần genome và gen, chúng ta biết rằng hầu hết genome prokaryot chứa các gen đơn bản. Điều này có nghĩa, đặc điểm của genome prokaryot là bao gồm chủ yếu các đoạn ADN không lặp lại. Ngược lại, tổng số gen eukaryot kể cả intron chiếm một phần rất nhỏ trong genome. Hơn nữa, không phải các gen phân bố một cách đều đặn trong genome. Có những vùng rộng (~ 500kb) trên nhiễm sắc thể không chứa một gen nào. Hơn 20% genome của người là những vùng ADN không chứa gen. Câu hỏi thú vị đặt ra là ADN không chứa gen có đặc điểm và chức năng như thế nào trong genome, trong tiến hoá?

Đặc điểm của những vùng ADN không có gen là chứa những trình tự nucleotide lặp lại. Chúng có thể lặp lại liên tục trên một vùng nhiễm sắc thể, hoặc phân bố rải rác khắp trong genome. Chiều dài của đoạn lặp lại có thể chỉ từ 1-2 nucleotide đến vài kb. Đặc biệt một số đoạn ADN lặp lại có khả năng di chuyển từ vị trí này sang vị trí khác. Chúng được gọi chung là transposon. Transposon có khả năng di chuyển độc lập và được xem là ở dạng hoạt động (active). Số lượng transposon ở dạng hoạt động rất ít trong genome. Tuy nhiên, số lượng các trình tự tương đồng với dạng hoạt động chiếm tỷ lệ rất lớn; hầu hết

chúng không có khả năng di chuyển độc lập và được xem là ở dạng không hoạt động (inactive). Trong genome của người, transposon hoạt động hoặc không hoạt động chiếm đến 45%. Transposon giữ vai trò quan trọng trong việc hình thành và duy trì cấu trúc không gian cũng như trật tự sắp xếp các nucleotide của nhiễm sắc thể.

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 khỏi vị trí cũ, 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. Do đó hoạt động của các gen liên quan đến sự di chuyển của transposon (thường là các gen nằm trong transposon) được kiểm soát rất chặt chẽ. Cơ chế kiểm soát chủ yếu thông qua biến đổi cấu trúc không gian vùng nhiễm sắc thể chứa transposon như methyl hoá ADN, methyl hoá histone H_3, deacetyl histone H_4 v.v...

Cách thức di chuyển và ghép vào genome của các transposon tuân theo hai cách liên quan đến dạng trung gian ADN hoặc ARN. Những đoạn ADN nào mà sự di chuyển của chúng gắn liền với dạng trung gian ARN được gọi là retroelement, hoặc ADN retrotransposon. Việc di chuyển của retrotransposon xảy ra tương tự với cách thức xâm nhiễm của virus mà genome của chúng là phân tử ARN (những virus này được gọi là retrovirus). Một khi đã xâm nhiễm vào tế bào, ARN của retrovirus được sao chép bởi reverse transcriptase tạo ra ADN. Phân tử ADN này sẽ được ghép vào genome của tế bào chủ. Khi virus sinh sôi, phần ADN đó lại được dùng để phiên mã tạo ra các phân tử ARN mới cần thiết cho việc đóng gói tạo virus mới. Các transposon ADN có khả năng thay đổi vị trí trong genome eukaryot không qua dạng trung gian ARN chiếm tỷ lệ ít hơn so với các retrotransposon. Ví dụ, ở

genome người, chỉ có khoảng 100 loại ADN transposon. Tuy nhiên ADN transposon có một ý nghĩa đặc biệt quan trọng đối với sự đa dạng hoá genome. Đặc điểm của transposon là có hai trình tự nucleotide với độ tương đồng cao nhưng lặp lại ngược chiều nhau ở hai đầu transposon.

1.8.1. Transposon trong genome vi khuẩn

Các transposon đơn giản nhất ở vi khuẩn được gọi là đoạn gắn 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, λ :: IS1 mô tả transposon IS1 gắn vào genome của bacteriophage λ. Transposons vi khuẩn không giữ một chức năng nào trong tế bào. Trình tự nucleotide ở một đầu IS thường lặp lại nhưng ngược chiều so với đầu kia. Hai trình tự ở hai đầu một IS được gọi là trình tự lặp lại ngược chiều (inverted repeat). Ví dụ, cấu trúc của một IS có trình tự như sau: GGTAT-X_n-ATACC (trong đó n là số nucleotide nằm giữa hai đầu lặp lại ngược chiều). Ngoài các IS, ở vi khuẩn còn có các đoạn ADN có khả năng di chuyển với kích thước dài hơn, ký hiệu là transposon Tn. Các Tn thường phân bố trên plasmid (phân tử ADN dạng vòng, kích thước thường không lớn) và có khả năng ghép xen vào bất kỳ vị trí nào trong genome. Chúng thường mang thông tin di truyền mã cho các protein chống chịu kháng sinh. Giữa IS và Tn có mối quan hệ về trình tự các nucleotide. Các Tn thường được giới hạn ở hai đầu bởi một loại IS nào đó. Hai IS ở hai đầu của một transposon có thể cùng chiều nhau (Bảng 1.4).

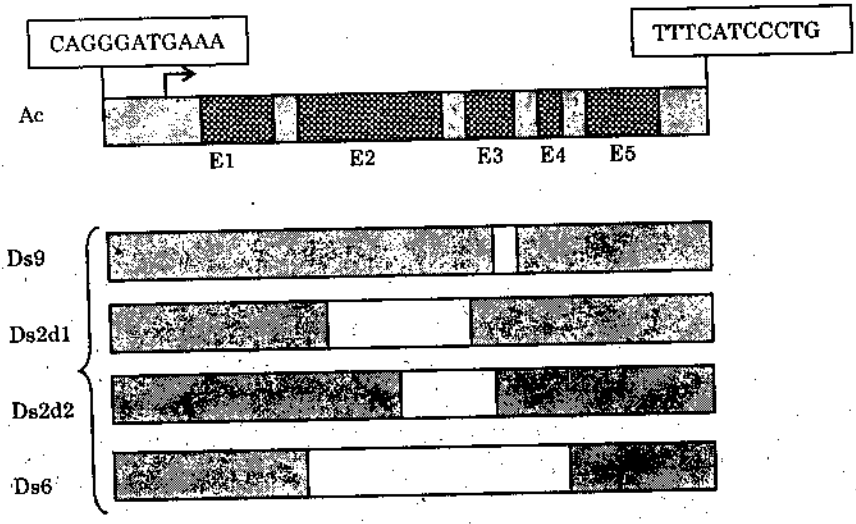
Bảng 1.4: Cấu trúc thường gặp của transposon. IS* không chức năng

Transposon	IS trái (ISL)	Gen marker của Transposon	IS phải (ISR)
Tn903	IS903	*kan*^{R} (kháng kanamicine)	IS903
Tn10	IS10L*	*tet*^{R} (kháng tetraciline)	IS10R
Tn5	IS50L*	*kan*^{R} (kháng kanamicine)	IS50R

Bảng 1.4 mô tả cấu trúc của một số transposon. Các transposon này đều mang gen mã cho tính chống chịu kháng sinh (R) và hai đầu của được giới hạn bởi trình tự nucleotide của hai IS sắp xếp ngược chiều nhau. Điều lưu ý là hai trình tự ở hai đầu có thể có chức năng hoặc không có chức năng IS.

1.8.2. Transposon trong genome eukaryot

Transposon eukaryot được nghiên cứu khá chi tiết đối với ngô và ruồi giấm. Ở ngô, hai transposon Ac và Ds được xem là ví dụ kinh điển về transposons. Chúng cùng thuộc vào một nhóm transposon, đều có hai trình tự lặp lại ngược chiều giống nhau. Di chuyển của các transposon Ds phụ thuộc vào sự có mặt của Ac. Trình tự nucleotide của Ac gồm 4563bp (5 exon), được giới hạn hai đầu bởi 11bp lặp lại ngược chiều. Mọi Ds đều có đoạn lặp lại ngược chiều giống nhau mặc dù chiều dài của chúng thay đổi (Hình 1.15).



Hình 1.15: Cấu trúc của transposon Ac/Ds. Transposon Ac chứa gen mã cho enzym transposase; gen này có 5 exon. Hai đầu Ac giới hạn bởi 11bp lặp lại ngược chiều. Các Ds có chiều dài khác nhau đều do Ac bị đột biến mất đoạn (phần ô trắng). Tuy nhiên tất cả các transposon Ds này đều được giới hạn bởi 11bp 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 ghép vào vị trí allen của một gen bất kỳ trên nhiễm sắc thể xảy ra ở tế bào soma sẽ tác động đến biểu hiện của allen đó trong quá trình phát triển của cây. Trải qua phân bào nguyên nhiễm, 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).

Ở ruồi giấm *Drosophila melanogaster*, transposon 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 F1 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 genome của các cá thể thuộc dòng P có chứa transposon P. Cơ chế kiểm soát sự di chuyển của P phụ thuộc vào yếu tố tồn tại trong tế bào chất của trứng (di truyền theo mẹ). Khi yếu tố 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. 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.

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 genome ruồi trong những năm cuối thế kỷ XX. 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 genome vi khuẩn thông qua quá trình tiếp hợp (transduction).

1.8.3. T-ADN di chuyển từ genome prokaryot đến genome eukaryot

Hiện tượng transposon di chuyển từ genome prokaryot sang genome eukaryot được nghiên cứu chi tiết đối với trường hợp T-ADN di chuyển từ *Agrobacterium* sang thực vật hai lá mầm. 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. Các biểu hiện bệnh ở cây gây ra do hoạt động của một số gen vi khuẩn đã được chuyển vào genome cây chủ. Cũng giống như các khối u động vật, các tế bào thực vật có ADN vi khuẩn ghép vào genome 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.

T-ADN là một đoạn ADN có chiều dài khoảng 23kb (tuỳ thuộc vào từng loại *A.tumefaciens*) nằm trên plasmid Ti. Hai đầu của đoạn ADN này có chứa 25bp lặp đi lặp lại giống nhau hoàn toàn chỉ sai khác nhau ở hai nucleotide. 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 mã cho các enzym tham gia vào phản ứng tổng hợp các hoocmon sinh trưởng auxin và cytokinin. Chỉ khi T-ADN được ghép vào genome thực vật, các oncogen nằm trên T-ADN mới hoạt động. Do đó tế bào cây chủ nào có T-ADN ghép vào genome lập tức phát triển không bình thường do rối loạn hoocmon 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 có genome bị biến đổi.

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.

Argobacterium 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, 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.

1.9. SẮP XẾP LẠI GENOME

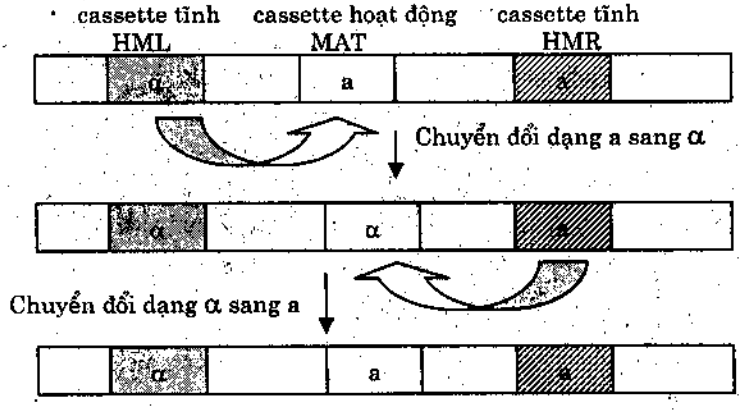
Mặc dù được xem là ổn định và bảo toàn trong các tế bào sinh dưỡng, genome cũng trải qua những biến động thay đổi các đoạn ADN tại những thời điểm nhất định trong phát triển. Những biến động này có thể là sự sắp xếp lại, tăng hoặc thậm chí mất đi một số đoạn ADN của genome, chủ yếu ở genome eukaryot bậc thấp. Sự sắp xếp lại genome hầu như không xảy ra ở tế bào mầm sinh dục mà chỉ

quan sát được ở các tế bào sinh dưỡng. Ở động vật, hiện tượng này rất hiếm gặp ngoại trừ trường hợp các gen liên quan đến phản ứng miễn dịch của cơ thể.

1.9.1. Thay đổi dạng giao phối ở nấm men *S.cerevisiae*

Hiện tượng sắp xếp lại genome đơn giản nhất xảy ra ở nấm men *S.cerevisiae* nhằm mục đích thay đổi dạng giao phối. Tế bào nấm men đơn bội có thể tồn tại ở hai dạng giao phối a và α. Tế bào ở dạng này có thể chuyển đổi thành dạng kia một cách tự phát trong tự nhiên. Dạng giao phối được quyết định bởi trình tự nucleotide ở một vùng nhất định trên nhiễm sắc thể số 3 của nấm men. Vùng này được phiên mã nên gọi là cassette hoạt động (active cassette). Khi tế bào ở dạng giao phối a thì trình tự nucleotide quyết định dạng a sẽ có mặt ở cassette hoạt động. Ngược lại, khi tế bào ở dạng α thì trình tự nucleotide quyết định dạng α sẽ có mặt ở cassette hoạt động. Như vậy, trình tự ở cassette hoạt động bị thay đổi dẫn đến sự thay đổi dạng giao phối.

Cassette hoạt động phân bố trên nhiễm sắc thể số 3. Hai phía của cassette này là hai trình tự quy định dạng giao phối a và α nhưng cả hai đều không được phiên mã. Vì vậy hai trình tự này được gọi là cassette tĩnh (silent cassette). Trình tự của một casette tĩnh quy định dạng giao phối α (hoặc a) có thể được phiên bản và thay thế cho trình tự quy định dạng giao phối a (hoặc α) đang tồn tại ở cassette hoạt động. Nhờ đó, trình tự ở cassette tĩnh vẫn được giữ nguyên nhưng phiên bản của trình tự đó được thay thế cho trình tự ở cassette hoạt động. Không thể xảy ra sự thay thế giữa hai cassette cùng dạng giao phối (Hình 1.16).



Hình 1.16: Chuyển đổi dạng giao phối ở nấm men *S. cerevisiae*.

Phân tích cấu trúc và cơ chế kìm hãm phiên mã ở hai cassette tĩnh cho thấy sự bất hoạt của hai cassette đều gắn liền với hiện tượng methyl hoá acid amin lysine nằm tại vị trí thứ 9 (K9 – tính từ đầu NH_2) của histone H_3, protein tham gia tạo cấu trúc nucleosome. Methyl hoá lysine 9 được xem là chỉ thị phân tử của cấu trúc dị nhiễm sắc vì hiện tượng này luôn quan sát được ở những vùng này. Do đó, vùng nhiễm sắc ở hai cassette này bị co đặc lại khiến cho các gen ở đó không thể phiên mã. Methyl hoá histone tham gia vào cơ chế kiểm soát hoạt động của gen chủ yếu ở mức độ cấu trúc không gian của sợi nhiễm sắc ; cơ chế này được gọi chung ***bất hoạt gen (gene silencing).***

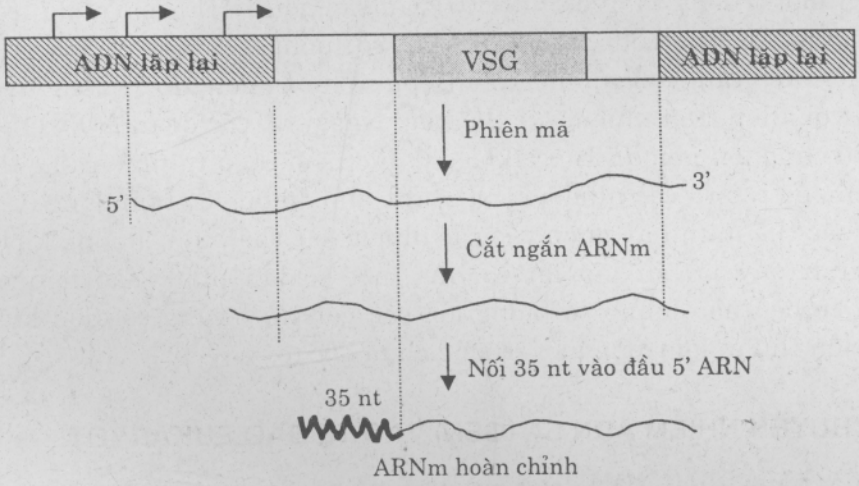
1.9.2. Thay đổi kháng nguyên bề mặt ở ký sinh trùng đơn bào *Trypanosome*

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 $5x10^6$ 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 trong cơ thể vật chủ, *Trypanosome* có khả năng thay đổi kháng nguyên bề mặt, trung bình khoảng 1-2 tuần/lần. Do đó ký sinh trùng tránh được phản ứng miễn dịch của tế bào chủ. Một loại *Trypanosome* có thể tạo ra ít nhất khoảng 100 VGS từ khi nhiễm cho đến khi gây chết vật chủ. Có nhiều loài *Trypanosome* xâm nhập các tế bào chủ khác nhau. *Trypanosoma 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*.

Kháng nguyên bề mặt do gen hoạt động ở một vị trí đặc biệt trong hệ gen nằm gần telomere; vị trí này được gọi là bản sao hoạt động ELC (expression-linked copy). Bản sao hoạt động cũng tương tự như cassette hoạt động của nấm men. Các gen mã cho các dạng VSG khác nhau nằm ở các vị trí khác nhau trong genome. Những vị trí này được gọi là bản gen gốc (basic-copy gene); một bản gen gốc tương tự như một cassette tĩnh. Số lượng bản sao hoạt động khoảng 20/genome và số lượng bản gen gốc có thể nhiều hơn 1000. Tuy nhiên, chỉ có một bản

sao hoạt động tại một thời điểm nhất định, cũng như chỉ một bản gen gốc được lựa chọn để thay thế bản sao khi chuyển đổi kháng nguyên. Quá trình chuyển đổi kháng nguyên bề mặt phụ thuộc vào sự chuyển đổi gen từ một bản gen gốc thay thế cho gen đang tồn tại ở bản sao 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.

Phân tích vị trí của bản sao hoạt động nằm gần telomere cho thấy vị trí này bị giới hạn bởi các đoạn trình tự lặp lại. Đặc biệt, bản sao hoạt động có thể được phiên mã bởi nhiều promoter phân bố từ 4kb đến 60kb phía trước ORF của VSG. Một trình tự có kích thước dao động trong khoảng 2500-3500bp từ bản gen gốc sẽ thay thế trình tự ở bản sao hoạt động bằng cơ chế trao đổi chéo. Kích thước này lớn hơn ORF của VSG (một VSG gồm khoảng 500 acid amin). Hầu hết trình tự không thuộc ORF nằm ở đầu 5' của gen (Hình 1.17). Những trình tự này dường như ảnh hưởng đến promoter nào được khởi động phiên mã cũng như cơ chế cắt nối exon-intron của ARNm.



Hình 1.17: Cấu trúc bản sao hoạt động của VSG.

Bản gen gốc sao chép vào bản sao có kích thước lớn hơn ORF của VSG. Bản sao được phiên mã từ promoter phân bố xa đầu 5' của gen VSG tạo ra phân tử ARN dài chứa cả trình tự lặp lại nằm trước và sau bản sao. Phân tử này sẽ được biến đổi ngắn dần và được thêm 35 nucleotide vào đầu 5' để hình thành ARNm hoàn chỉnh. Phân tử ARNm này được dịch mã tổng hợp VSG.

Phân tích trình tự nucleotide của ARNm của VSG cho thấy 35 nucleotide đầu tiên không phải là các nucleotide đã có mặt ở bản sao hoạt động. Nói cách khác, 35 nucleotide này được nối vào phân tử ARNm của VSG theo cơ chế cắt nối trans-splicing (trình bày chi tiết trong chương 2).

1.9.3. Nhân bản tạo các bản gen mới từ genome

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. Ví 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. Số lượng gen có thể thay đổi từ 40-400 gen.

Quá trình nhân bản số gen *dhfr* gồm nhiều bước, diễn biến của từng bước chưa được xác định rõ ràng. Thực nghiệm cho thấy, khi sàng lọc trên môi trường tăng dần độc tố, bản sao gen DHFR chỉ tăng trên một nhiễm sắc thể tương đồng, tức là số lượng gen *dhfr* nằm cạnh nhau liên tiếp tăng đối với một allen, còn số lượng gen *dhfr* không thay đổi đối với allen trên nhiễm sắc thể kia. Ngoài ra, các đoạn ADN (100-1000kb) chứa 2-4 gen *dhfr* (~31kb/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 nguyên nhiễm, tế bào nào mang số lượng lớn các gen *dhfr* càng có điều kiện phát triển thuận lợi trong môi trường chứa độc tố.

1.10. CHUYỂN NHIỄM ADN LẠ (GEN) VÀO TẾ BÀO EUKARYOT

Quá trình đưa ADN lạ (gen) vào tế bào eukaryot được gọi là chuyển nhiễm (***transfection***). Chuyển nhiễm thường xảy ra với xác suất rất thấp, gồm nhiều bước. Thí nghiệm chuyển nhiễm đầu tiên thực hiện khi bổ sung các nhiễm sắc thể ở metaphase (hoặc các đoạn ADN tinh sạch) vào tế bào nuôi cấy. Một số rất ít tế bào sẽ nhận nhiễm sắc thể; tạo ra những tế bào chuyển nhiễm không bền (***transient transfectants***) do các nhiễm sắc thể lạ bị đào thải nhanh chóng. Một

số đoạn ADN (sinh ra do nhiễm sắc thể lạ bị đứt gãy) được chèn vào genome của tế bào nhận tạo nên dòng tế bào chuyển gen bền.

Trong dòng tế bào chuyển nhiễm không bền, ADN lạ không chèn vào genome nhưng vẫn tồn tại bên trong tế bào. Trong dòng tế bào chuyển nhiễm bền, ADN lạ được duy trì bền sau khi đã chèn vào genome. Trong cả hai trường hợp bền và không bền, ADN lạ nếu là gen thì vẫn hoạt động như nhau. Do đó, cần phải có chỉ thị phân tử để phân biệt chính xác hai dòng tế bào này. Chỉ thị phân tử thường được chuyển nhiễm cùng với gen lạ. Quá trình này được gọi là đồng chuyển nhiễm (***cotransfection***). Một trong các chỉ thị phân tử hay được sử dụng là gen ***tk*** mã cho thymidine kinase, enzym xúc tác phản ứng tổng hợp thymidine triphosphate, tiền chất cần có trong quá trình tổng hợp ADN. Gen ***tk*** sẽ được chuyển nhiễm cùng gen lạ vào tế bào không có gen ***tk***. Gen ***tk*** cùng với gen lạ thường được ghép một cách ngẫu nhiên vào genome tế bào nhận và thường tập trung thành các bản sao lặp lại liên tục. Số lượng bản sao tăng lên khi điều kiện chọn lọc các tế bào bị chuyển nhiễm tăng lên. Điều lưu ý trong mỗi dòng tế bào chuyển nhiễm bền, gen ***tk*** và gen lạ chỉ ghép vào một vị trí, mặc dù vị trí này phân bố ngẫu nhiên trong genome tế bào nhận.

Chương 2

TÁI BẢN VÀ SỬA CHỮA ADN

Phân chia tế bào ở mọi cơ thể từ prokaryot đến eukaryot đều đòi hỏi quá trình nhân đôi (tái bản) 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. Trong tế bào prokaryot, tái bản được bắt đầu từ một vị trí duy nhất trên nhiễm sắc thể và gắn liền với quá trình phân chia tạo ra hai tế bào con. Trong tế bào eukaryot, tái bản xảy ra trong phase S nhưng tế bào còn trải qua phase G2 rồi mới bước vào phân bào.

Tái bản ADN là một quá trình phức tạp, trong đó phân tử ADN được làm khuôn để tổng hợp nên một phân tử mới giống hệt khuôn. Tái bản ADN được nghiên cứu đầu tiên với vi khuẩn *E. coli* bị đột biến, gọi là các tế bào *dna*. Nếu ADN không tái bản thì tế bào *E. coli* không phân chia, tức là không có khả năng duy trì sự sống. Do đó, nghiên cứu tái bản ADN chỉ có thể tiến hành trên các dòng tế bào đột biến nhạy cảm nhiệt độ (temperature-sensitive mutants). Đây là những tế bào có thể tái bản ADN bình thường ở 37°C nhưng tái bản sẽ bị ngừng khi chuyển tế bào sang 42°C. Hầu hết các dòng *dna* thu nhận từ thực nghiệm là những đột biến liên quan đến enzym tham gia trực tiếp vào tổng hợp sợi ADN mới. Một số ít dòng tế bào *dna* có đột biến liên quan đến khuôn và phức protein chuẩn bị cho khởi động tái bản. Tái bản ADN bao gồm các bước cơ bản sau:

1) Phức protein tương tác với trình tự nucleotide đặc biệt để bắt đầu quá trình tái bản. Trình tự này được gọi là ***tâm tái bản*** (***origin***). Tại tâm tái bản, phân tử kép ADN phải được tách thành các sợi khuôn dạng đơn.

2) Phức ***replisome*** gồm nhiều protein tham gia tổng hợp sợi đơn

ADN mới trên khuôn mẫu sợi đơn cũ. Phức replisome không tồn tại ở dạng tự do (tức là không thể tồn tại độc lập như ribosome) mà luôn luôn liên kết với ADN khuôn. Khi phức repliosome di chuyển trên sợi khuôn ADN, hai sợi đơn của khuôn tiếp tục bị mở xoắn, hai sợi đơn mới được tổng hợp dựa vào hai khuôn sợi cũ.

3) Kết thúc tổng hợp ADN. Hai phân tử ADN giống hệt nhau tương tác với phức protein (histone, phi histone...) để có cấu trúc không gian.

2.1. REPLICON

Trình tự nucleotide chỉ tái bản một lần duy nhất trong một chu kỳ tế bào được gọi là một ***replicon*** (tạm dịch là ***đơn vị tái bản***). Replicon thường chứa vị trí bắt đầu tái bản (***origin***) và có thể có cả vị trí kết thúc tái bản. Vị trí origin đặc biệt quan trọng vì nó liên quan đến các yếu tố cần thiết cho quá trình tái bản.

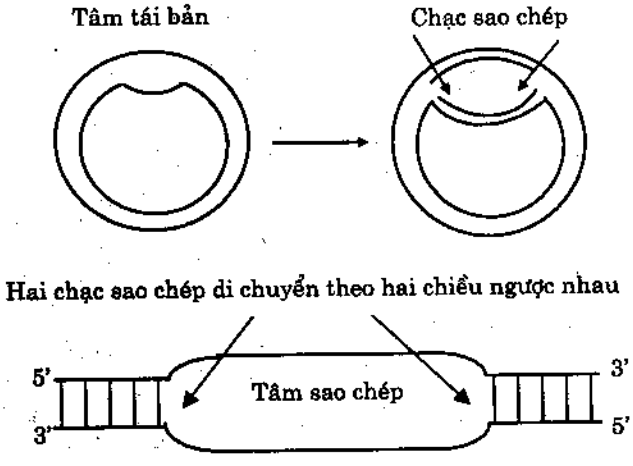
– Mỗi tế bào vi khuẩn chỉ có một nhiễm sắc thể. Nhiễm sắc thể đó được xem là một replicon vì chỉ có một vị trí origin. Do đó khi genome vi khuẩn bắt đầu tái bản thì sẽ kết thúc bằng việc hình thành hai genome (hai nhiễm sắc thể) mới để phân về hai tế bào con. Như vậy, chúng ta có thể xem tái bản nhiễm sắc thể vi khuẩn được kiểm soát bởi cơ chế một bản sao duy nhất (*single-copy control*).

Tuy nhiên, tế bào vi khuẩn còn chứa các loại plasmid là những phân tử ADN dạng vòng có kích thước nhỏ hơn nhiễm sắc thể. Một plasmid có thể chỉ có một bản copy trong một tế bào. Plasmid này chỉ tái bản (nhân đôi) một lần giống như nhiễm sắc thể để phân chia về hai tế bào con. Do đó, tái bản của loại plasmid này cũng được kiểm soát bởi cơ chế một bản sao duy nhất (single-copy control). Ngoài ra, một số plasmid có thể tồn tại nhiều bản copy trong một tế bào; tức là số lượng plasmid nhiều hơn 1 (tức là nhiều hơn số nhiễm sắc thể). Những plasmid này nhân đôi nhiều lần ngay khi tế bào không phân chia. Chúng được kiểm soát bởi cơ chế đa bản sao (*multicopy control*). Genome virus sau khi xâm nhiễm tế bào chủ cũng chịu sự kiểm soát đa bản sao.

– Genome eukaryot gồm nhiều nhiễm sắc thể; mỗi nhiễm sắc thể có nhiều origin. Do đó genome eukaryot gồm nhiều đơn vị tái bản replicon. Các replicon này có thể tái bản tại thời điểm khác nhau. Tuy nhiên tất cả replicon eukaryot đều giống nhau là cũng chỉ tái bản một lần duy nhất trong một lần phân bào. Như vậy, replicon của genome vi

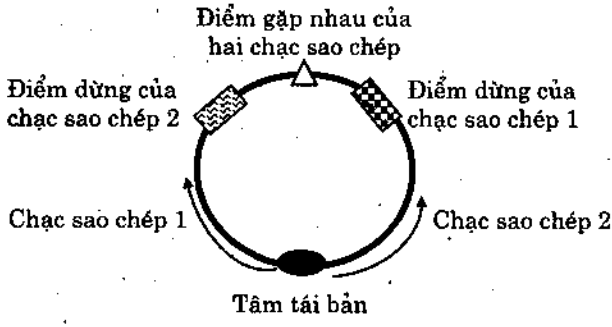
khuẩn và của genome trong nhân tế bào eukaryot đều chịu kiểm soát bởi cơ chế một bản sao.

Replicon có thể tồn tại trên phân tử ADN dạng thẳng hoặc dạng vòng. Vị trí tại đó sợi ADN đang tổng hợp được gọi là ***chạc sao chép*** (***replication fork***). Chạc này bắt đầu xuất hiện tại tâm tái bản và di chuyển dọc theo một, hoặc cả hai chiều trên sợi ADN khuôn. Nếu chỉ có một chạc sao chép xuất phát từ một tâm tái bản; tức là chạc di chuyển theo một chiều trên sợi khuôn thì quá trình tái bản được xem là đơn chiều (*unidirectional replication*). Nếu có hai chạc sao chép xuất phát từ một tâm tái bản và di chuyển theo cả hai phía thì quá trình tái bản được xem là hai chiều (*bidirectional replication*) (Hình 2.1).



Hình 2.1: Chạc sao chép hình thành từ tâm tái bản (origin) và di chuyển về hai phía ngược nhau trên phân tử ADN khuôn. ADN khuôn có thể tồn tại ở dạng vòng hoặc dạng mạch thẳng.

Hai chạc sao chép di chuyển với tốc độ như nhau trên phân tử khuôn dạng vòng (genome vi khuẩn). Mỗi chạc sao chép sẽ có một điểm dừng tái bản. Điều thú vị là chạc sao chép này phải đi qua điểm dừng của chạc sao chép kia trước khi hai chạc gặp nhau và kết thúc quá trình tái bản (Hình 2.2). Vì một lý do nào đó mà một chạc di chuyển nhanh hơn chạc kia thì nó sẽ đi qua điểm dừng chung nhưng bị chặn lại ở điểm kết thúc của riêng mình. Tại đây, chạc di chuyển nhanh sẽ đợi để gặp chạc di chuyển chậm. Cách thức sắp xếp các vị trí dừng tái bản như ở genome vi khuẩn được gọi là "bẫy chạc" (*replication fork trap*).



Hình 2.2: Phân bố các điểm dừng của hai chạc sao chép trên nhiễm sắc thể vi khuẩn.

Chạc sao chép thứ nhất bắt buộc phải đi qua điểm dừng của chạc sao chép thứ hai và ngược lại. Sau đó, hai chạc sẽ gặp nhau tại điểm kết thúc tái bản.

2.2. ADN POLYMERASE

ADN polymerase là enzym xúc tác phản ứng tổng hợp ADN. Tế bào prokaryot và eukaryot đều có nhiều loại ADN polymerase có hoạt tính khác nhau. Chỉ một số ít enzym tham gia tổng hợp mới (*denovo*) sợi ADN còn các ADN polymerase khác giữ chức năng phụ trợ trong tái bản hoặc liên quan đến quá trình sửa chữa các sai hỏng ADN (Bảng 2.1). Hầu hết các ADN polymerase có hoạt tính chỉ khi tạo phức với một số protein khác. Những protein này tham gia mở xoắn sợi kép ADN, duy trì dạng sợi đơn, đảm bảo tương tác bền của enzym với sợi khuôn v.v...

Bảng 2.1 trình bày các loại ADN polymerase và chức năng của chúng ở vi khuẩn *E.coli*. ADN polymerase I có hoạt tính lớn nhất trong dịch chiết từ tế bào *E.coli* cho nên rất khó phát hiện được hoạt tính của các loại ADN polymerase khác. Vì thế, dịch chiết từ tế bào đột biến gen *polA* hay được sử dụng trong thí nghiệm tổng hợp ADN *in vitro*.

Bảng 2.1: Các loại ADN polymerase ở vi khuẩn *E.coli* và tế bào eukaryot

ADN polymerase *E.coli*	Chức năng chính
ADN polymerase I (PolA)	Sửa chữa ADN và chức năng phụ trong tái bản
ADN polymerase II (PolB)	Khởi động lại chạc tái bản khi ADN bị tổn thương trong khi tái bản
ADN polymerase III (PolC - replicase)	Tổng hợp mới sợi ADN trong quá trình tái bản

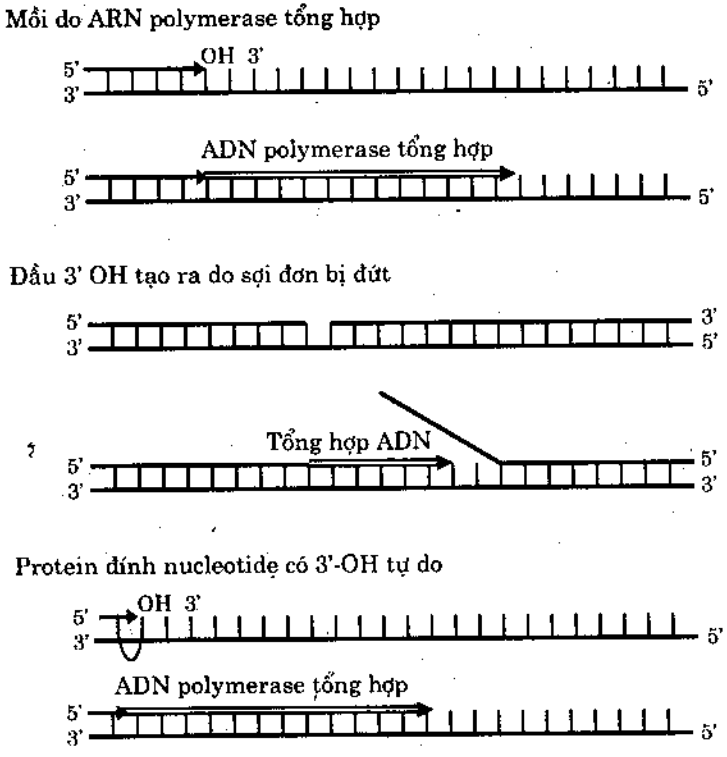
ADN polymerase IV và V	Cho phép tái bản ADN vượt qua những vị trí sai hỏng trên sợi khuôn
ADN polymerase eukaryot	
ADN polymerase α	Tổng hợp ADN trong nhân, có độ chính xác cao
ADN polymerase δ	Tổng hợp ADN trong nhân, có độ chính xác cao
ADN polymerase ε	Tổng hợp ADN trong nhân, có độ chính xác cao
ADN polymerase γ	Tổng hợp ADN trong ty thể, có độ chính xác cao
ADN polymerase β	Sửa chữa ADN (BER) có độ chính xác cao
ADN polymerase ζ	Tổng hợp ADN vượt qua dimer T-T
ADN polymerase η	Sửa chữa ADN
ADN polymerase τ	Cần thiết cho phân bào giảm phân
ADN polymerase κ	Sửa chữa các đột biến mất hoặc thay thế nucleotide

Một số thực khuẩn thể có gen mã cho ADN polymerase, ví dụ, thực khuẩn thể T4, T5, T7 và SPO1. Tất cả các enzym của thực khuẩn thể đều có hoạt tính tổng hợp ADN theo chiều 5'→3' và có hoạt tính đọc sửa (proofreading) exonuclease 3'→5'. Tuy nhiên, ADN polymerase của thực khuẩn thể phải tương tác với protein do thực khuẩn thể mã cho, hoặc protein của tế bào chủ để tạo thành phức có những hoạt tính đó.

Trong nhân tế bào eukaryot, ADN polymerase α cần thiết để bắt đầu tái bản; hai enzym ADN polymerase β (beta) và ε (epsilon) tham gia tổng hợp mới và sửa chữa ADN. Trong ty thể, ADN polymerase γ (gamma) có chức năng tái bản ADN.

2.3. TỔNG HỢP ADN

ADN được tái bản theo cơ chế bán bảo toàn. Phân tử ADN khuôn ban đầu được tách thành hai sợi đơn; mỗi sợi đơn là khuôn để tổng hợp nên một sợi đơn mới. Sợi mới có trình tự nucleotide bổ sung với sợi khuôn. Sau quá trình tái bản sẽ có hai phân tử ADN giống hệt nhau; mỗi phân tử ADN gồm một sợi đơn cũ và một sợi mới.

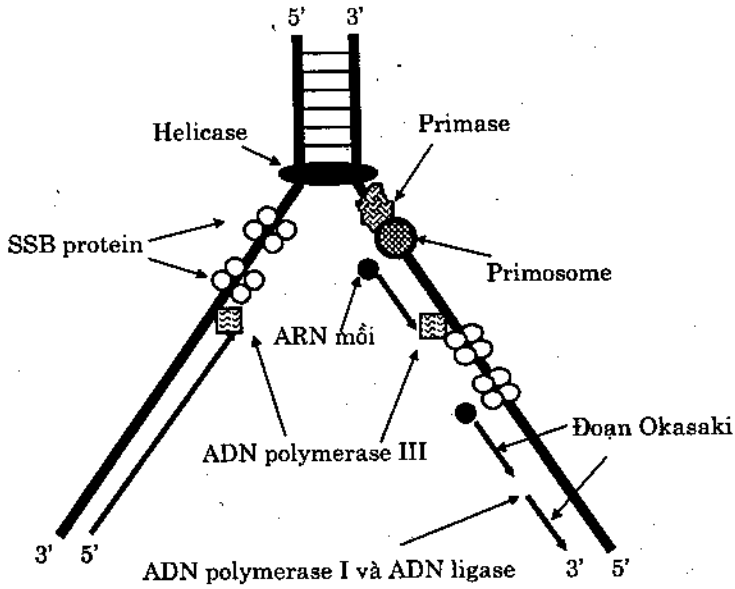


Hình 2.3: Đầu 3' có nhóm OH tự do được tạo ra để cho ADN polymerase thêm nucleotide vào trong phản ứng tổng hợp ADN (theo Lewin; 2004).

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 một đoạn oligonucleotide ngắn có nhóm OH tự do ở đầu 3'. Đoạn ngắn này được gọi là mồi (*primer*). Đầu 3' có nhóm OH tự do có thể tạo ra bằng nhiều cách (Hình 2.3).

Enzym ADN polymerase thêm các nucleotide vào đầu 3' của mồi để tổng hợp nên sợi đơn mới theo một chiều nhất định từ 5'→3'. Do đó trên sợi khuôn thứ nhất (*leading strand*), một sợi đơn được tổng hợp liên tục theo chiều từ 5'→3'. Trên sợi khuôn thứ hai (*lagging strand*) ngược chiều và bổ sung với sợi khuôn thứ nhất, có nhiều sợi ngắn được tổng hợp một cách gián đoạn. Các đoạn ngắn này được gọi là *đoạn Okazaki*. Chúng được nối với nhau bởi ADN ligase để tạo thành một sợi liên tục. 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 2.4). Như vậy, tái bản ADN trên sợi khuôn kép xảy ra một cách bán liên tục (*semidiscontinous replication*).

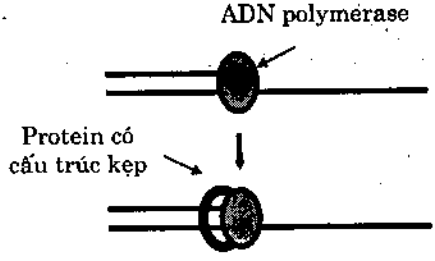


Hình 2.4: Mô hình chạc sao chép ADN. SSB (single strand binding) protein giữ cho sợi đơn duỗi thẳng. Helicase mở xoắn sợi ADN khuôn. Primase tổng hợp mồi ARN. Mồi này sau đó bị thay thế bởi đoạn nucleotide do ADN polymerase I tổng hợp. Các đoạn Okasaki được nối với nhau bởi ADN ligase.

Trong quá trình tái bản, enzym helicase sử dụng ATP mở xoắn phân tử khuôn tạo ra hai sợi đơn. Nhờ đó chạc sao chép di chuyển để tổng hợp sợi ADN mới. Do tổng hợp ADN chỉ thực hiện theo một chiều từ 5'→3' nên một sợi đơn khuôn phải được giữ ở trạng thái tự do đến một độ dài nhất định đủ để tổng hợp sợi lagging. Trạng thái tự do của sợi đơn ADN được duy trì nhờ tương tác với protein. Những protein có khả năng tương tác với sợi đơn ADN được gọi là SSB (*single-strand binding protein*). Vi khuẩn *E.coli* có khoảng 12 loại helicase khác nhau tham gia vào tổng hợp, sửa chữa và tái tổ hợp ADN. Helicase thường tồn tại dưới dạng multilmer tạo ra nhiều vị trí tương tác với 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 enzym khi tổng hợp sợi ADN dài liên tục. Để khắc phục điều này, phân tử enzym được giữ lại trên sợi khuôn nhờ một cấu trúc protein đặc biệt giống như chiếc kẹp. Kẹp này cho phép enzym bám vững chắc khi chuyển động trên sợi khuôn nhưng đồng thời đảm bảo cho enzym rời khỏi khuôn nhanh chóng khi quá trình tổng hợp kết thúc (Hình 2.5).



Hình 2.5: Cấu trúc kẹp protein giữ ADN polymerase bám chắc vào sợi 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 enzym 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 enzym 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.

Tái bản đầu mút các nhiễm sắc thể eukaryot xảy ra hoàn toàn khác biệt so với các vùng khác nhằm đảm bảo tính nguyên vẹn của đầu mút. Ví dụ, thực khuẩn thể T4 hoặc phage λ đã giải quyết vấn đề này bằng cách biến đổi nhiễm sắc thể dạng thẳng sang dạng vòng tròn hoặc tạo ra các phân tử ADN multimer trong quá trình tái bản. Đầu mút nhiễm sắc thể eukaryot thường chứa các trình tự ngắn lặp lại. Số lần lặp lại có thể thay đổi bởi telomerase. Do đó, việc duy trì đầu mút không nhất thiết phải phụ thuộc hoàn toàn vào quá trình tái bản.

2.4. ĐẢM BẢO TÍNH CHÍNH XÁC TRONG TÁI BẢN ADN

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, 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 với 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'. Đây là vị trí nucleotide mới được gắn vào mồi. Chỉ có nucleotide tạo cặp theo nguyên tắc bổ sung với nucleotide trên sợi khuôn mới được gắn vào vị trí 3' nhờ liên kết phosphodiester. 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). Nhờ 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 ADN polymerase tự sửa các sai hỏng. Ngược lại, 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 quy định tính chất đọc sửa exonuclease 3'→5' của ADN polymerase. Vì vậy rất nhiều sai sót xảy ra mà không bị phát hiện để sửa chữa. Tuy nhiên chúng có thể được sửa theo một cơ chế khác gọi là cơ chế đọc 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 kép 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).

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. Để phân biệt giữa sợi mới và sợi khuôn có lẽ phức protein nhận biết vết đứt gãy hay gặp trên sợi ADN 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.

2.5. TÂM TÁI BẢN (ORIGIN)

Hai sợi đơn ADN bắt đầu tách nhau ra tại vị trí đặc biệt gọi là tâm tái bản hay còn gọi là tâm sao chép (***replication origin - ori***). Đối với vi khuẩn, nấm men và một số virus ký sinh trên động vật, genome của chúng thường chỉ có một tâm tái bản ký hiệu là *oriC*. Tâm tái bản của genome vi khuẩn *E.coli* có chứa hai đoạn ADN 9bp và 13bp lặp lại nhiều lần tạo nên tâm tái bản có kích thước 245bp. Tái bản được bắt đầu khi phức gồm 6 protein (DnA, DnB, DnC, HU, Gyrase và SSB) tương tác với đoạn 245bp. Protein DnA nhận biết tâm tái bản; phức protein DnB-DnC có hoạt tính helicase sử dụng ATP để mở xoắn sợi khuôn; protein HU tương tác với ADN làm uốn cong sợi khuôn; protein Gyrase sử dụng ATP để tháo cấu trúc siêu xoắn do chạc sao chép di chuyển trên sợi khuôn.

Tại tâm tái bản, đoạn ngắn ARN được dùng làm mồi để ADN polymerase bắt đầu tổng hợp sợi ADN mới (sợi leading). Đoạn ARN này có thể được phiên mã từ promoter nằm phía trước tâm và kết thúc tại tâm. Nghiên cứu tái bản tại tâm sao chép của thực khuẩn thể cho thấy tái bản chỉ bắt đầu khi ARN polymerase phiên mã tại promoter nằm ngay phía trước tâm tái bản. Như vậy phiên mã xảy ra ở xung quanh tâm có thể có hai chức năng. *Thứ nhất*, phiên mã làm thay đổi cấu trúc của vùng chứa tâm sao chép giúp phức protein tiếp cận với trình tự nucleotide tại đó. *Thứ hai*, phiên mã kết thúc tại tâm sao chép và tạo ra sợi ARN có đầu 3' chứa nhóm OH tự do; nhờ đó ADN polymerase khởi động tổng hợp bằng cách thêm nucleotide vào đầu 3' đó.

Tái bản bắt đầu ở tâm sao chép trên genome thực khuẩn thể λ yêu

cầu protein O và protein P mã bởi hai gen của thực khuẩn thể. Tâm sao chép *oriλ* nằm ngay trong gen mã cho protein O. Phức protein O-P tương tác với trình tự nucleotide ở *oriλ* và tương tác với DnaB (protein của tế bào vi khuẩn). Protein O tương tác với các trình tự lặp lại của *oriλ*, còn protein P tương tác với DnaB để giữ DnaB tại tâm sao chép. Như vậy protein O có chức năng tương tự DnaA còn protein P tương tự DnaC của *E.coli*. Protein P của thực khuẩn thể ức chế hoạt tính helicase của DnaB. Chỉ khi nào protein P rồi khỏi tâm sao chép thì DnaB mới mở xoắn để tổng hợp sợi ADN mới. Tâm sao chép *oriλ* gồm trình tự ngắn (18 nucleotide) lặp lại là vị trí tương tác với protein O và đoạn giàu A-T nằm ngay sau đó. Đoạn giàu A-T rất nhạy cảm với nuclease S1, enzym nhận biết và phân cắt ADN không bắt cặp bổ sung. Như vậy liên kết bổ sung giữa hai sợi khuôn ở đoạn giàu A-T bị đứt hình thành chạc sao chép.

Cấu trúc tâm sao chép của vi khuẩn và thực khuẩn thể cũng như cách thức khởi động tổng hợp ADN tại hai loại tâm đó có những điểm tương đồng nhau. Tâm sao chép của virus động vật SV40 hoạt động tương tự như *oriC* và *oriλ*. Protein antigene T do virus mã tương tác với những trình tự lặp lại ngắn trong genome. Đặc biệt antigene T có hoạt tính helicase phân huỷ ATP nên virus không cần protein có chức năng như DnaB.

Đối với sinh vật eukaryot, trên mỗi nhiễm sắc thể có nhiều tâm tái bản. Các tâm này (20 đến 80) thường tập trung trên một đoạn ADN tạo nên đơn vị sao chép (replication unit). Khoảng cách giữa các tâm tái bản trên một đơn vị thay đổi từ 30kb đến 300kb. 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 dị 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.

2.6. KIỂM SOÁT KHỞI ĐỘNG TÁI BẢN TẠI TÂM SAO CHÉP

Tâm sao chép là vị trí đầu tiên tại đó ADN được tổng hợp. Một khi toàn bộ sợi ADN mới chưa được tổng hợp xong thì tâm sao chép không

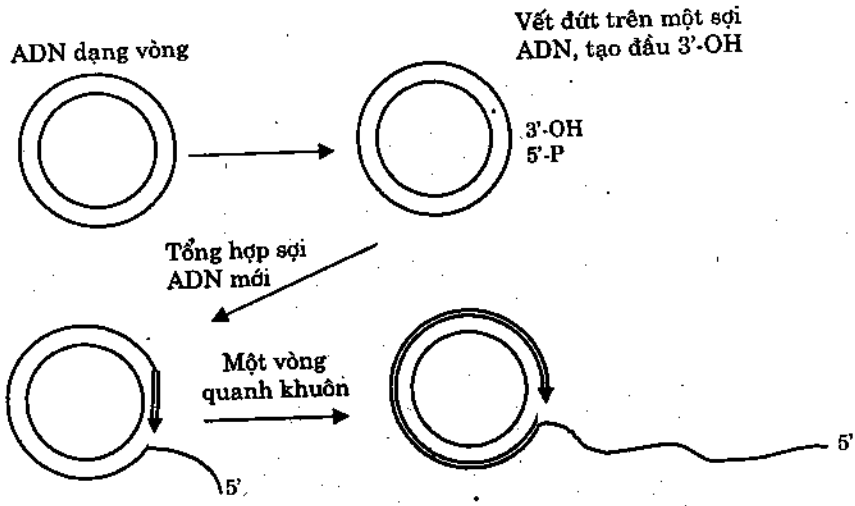
được khởi động cho lần tái bản tiếp theo. Điều này có nghĩa là quá trình tái bản không xảy ra theo kiểu cuốn chiếu; chỉ khi nào phân tử mới ADN được tổng hợp xong thì lần tái bản tiếp sau mới có thể bắt đầu. Cơ chế nào kiểm soát tâm sao chép tuân thủ đúng nguyên tắc đó?

Phân tích trình tự của tâm sao chép *oriC* ở vi khuẩn *E.coli* cho thấy, *oriC* gồm một trình tự ngắn GATC lặp lại 11 lần, trong đó mọi adenine đều bị methyl hoá ở vị trí N_6. Tái bản ADN xảy ra theo cơ chế bán bảo toàn. Vì vậy sau khi tái bản, tâm sao chép chỉ bị methyl hoá bán phần; trình tự GATC trên sợi khuôn có nhóm methyl nhưng trình tự GATC trên sợi mới không bị methyl hoá. Adenine trong trình tự GATC chỉ bị methyl hoá toàn phần sau khoảng 13 phút tính từ khi tổng hợp ADN bắt đầu. Phản ứng methyl hoá đòi hỏi enzym Dam methylase. Phải chăng do chỉ bị methyl hoá bán phần mà tâm sao chép không thể khởi động tổng hợp ADN ? Thực nghiệm cho thấy khi biến nạp plasmid có tâm *oriC* vào tế bào đột biến dam^- (tế bào không có khả năng tổng hợp enzym Dam) thì khả năng nhân đôi của plasmid hoàn toàn phụ thuộc vào *oriC* có bị methyl hoá hay không. Nếu như plasmid chứa *oriC* bị methyl hoá toàn phần thì plasmid sẽ nhân đôi tạo ra các plasmid mới có *oriC* bị methyl hoá bán phần. Sau đó các plasmid này không thể tái bản được nữa. Tuy nhiên, nếu như plasmid mang *oriC* không bị methyl hoá thì plasmid tái bản một cách bình thường giống như trong tế bào có gen dam^+. Như vậy, hiện tượng methyl hoá bán phần tại tâm sao chép ngăn cản tâm này khởi động tái bản. Chỉ khi được methyl hoá toàn phần thì tâm sao chép mới có thể bước vào lần tái bản tiếp theo. Methyl hoá toàn phần *oriC* xảy ra sau khoảng 13 phút tính từ khi tổng hợp ADN bắt đầu tại *oriC*. Trong khoảng thời gian này, phiên mã tại promoter của gen *dam* bị ức chế khiến lượng protein Dam giảm đi.

2.7. TÁI BẢN THEO CƠ CHẾ VÒNG TRÒN LĂN (ROLLING CYCLE REPLICATION)

Trong quá trình tái bản ADN, sợi ADN khuôn có thể tồn tại dưới dạng thẳng hoặc dạng vòng, ở dạng sợi đôi hoặc sợi đơn. Những dạng khuôn này được sử dụng để tổng hợp nên bản sao giống hệt chúng về hình dạng (tròn hoặc thẳng) và cấu trúc (đôi hoặc đơn). Tuy nhiên, phân tử ADN sợi đôi có thể làm khuôn để tổng hợp sợi đơn hoặc ngược lại. Do đó, có nhiều cách thức tái bản ADN khác nhau. Mối liên hệ giữa hình dạng cấu trúc của sợi khuôn và sợi cần nhân đôi là yếu tố quyết định cách thức tái bản ADN.

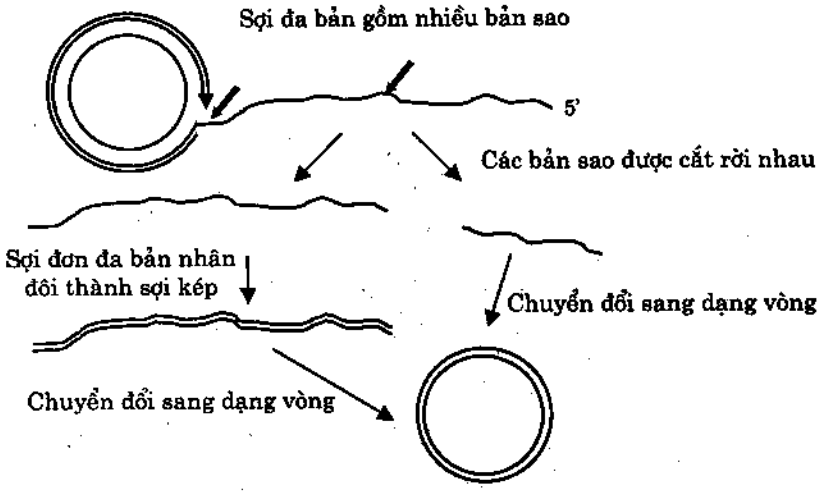
Phân tử ADN dạng vòng trong tế bào vi khuẩn được tái bản theo cơ chế mô tả trên Hình 2.2, trong đó liên kết giữa hai sợi đơn bị phá vỡ tại tâm sao chép. Tuy nhiên, phân tử ADN dạng vòng của thực khuẩn thể được tái bản theo cách thức khác biệt với ADN vi khuẩn. Quá trình tổng hợp sợi ADN mới được bắt đầu tại một vết cắt thuộc tâm sao chép trên một sợi đơn khuôn, giải phóng đầu 3'-OH tự do. Đầu này được sử dụng như mồi để ADN polymerase tổng hợp sợi mới tạo cặp bổ sung với sợi khuôn không bị cắt. Như thế, chạc sao chép sẽ di chuyển liên tục vòng quanh sợi khuôn. Mỗi lần chạc sao chép di chuyển hết một vòng thì một sợi đơn ADN được tổng hợp xong. Sợi mới tách khỏi sợi khuôn để chạc sao chép di chuyển vòng tiếp theo. Cách thức tổng hợp như thế tạo ra sợi đơn ADN dài gồm các bản sao kế tiếp nhau liên tục (Hình 2.6). Cách thức này được gọi là tái bản theo vòng tròn lăn (***rolling cycle replication***).



Hình 2.6: Cơ chế tái bản vòng tròn lăn (rolling cycle replication).
Quá trình tổng hợp sợi ADN mới được bắt đầu tại một vết cắt giải phóng đầu 3'-OH tự do. Đầu này được làm khuôn để tổng hợp sợi đơn mới.

Sợi ADN tổng hợp theo cơ chế vòng tròn lăn gồm nhiều bản sao liên tiếp. Các bản này có thể được cắt rời nhau để tạo thành từng bản riêng biệt; mỗi bản giống hệt phân tử khuôn ban đầu. Tiếp đến, mỗi bản đơn lẻ có thể tạo thành phân tử dạng vòng (Hình 2.7). Tái bản theo cơ chế vòng tròn lăn thường được sử dụng để nhân bản một gen hoặc một genome với số lượng nhiều trong thời gian ngắn. Ví dụ, trứng ếch sau khi thụ tinh phát triển rất nhanh nên yêu cầu rất nhiều ARN ribosome cho bộ máy tổng hợp protein. Số lượng ARNr được đáp ứng

đầy đủ do gen mã cho ARNr được nhân bản theo cơ chế vòng tròn lăn tạo nên số lượng gen rất lớn trong một khoảng thời gian nhất định trong quá trình phát triển phôi. Ngoài ra, genome của thực khuẩn thể cũng được tái bản theo cơ chế này.



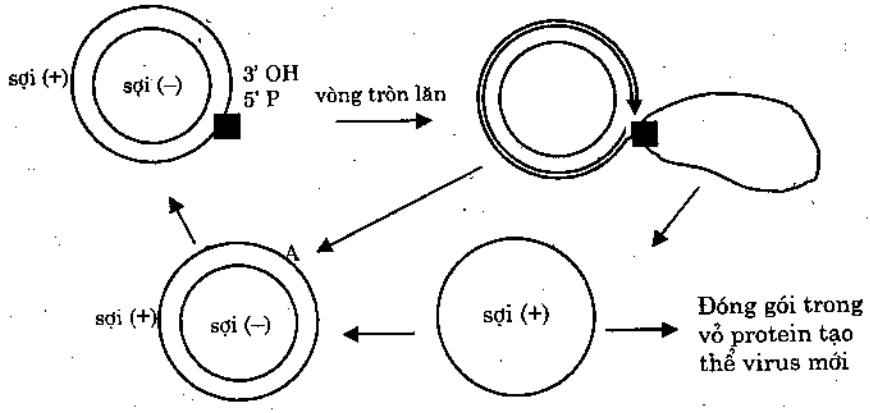
Hình 2.7: Sợi đơn đa bản tổng hợp theo cơ chế vòng tròn lăn được chuyển đổi sang sợi đôi có dạng vòng (theo Lewin; 2004).

2.8. TÁI BẢN ADN Ở VIRUS

2:8.1. Tái bản ADN của thực khuẩn thể

Genome thực khuẩn thể ϕX174 là sợi đơn ADN dạng vòng, ký hiệu là sợi dương (+). Sợi này được dùng làm khuôn để tổng hợp sợi bổ sung, ký hiệu là sợi âm (-). Hai sợi dương và âm tạo thành phân tử sợi đôi dạng vòng. Lúc này, protein A mã bởi gen virus bám vào sợi (+) tại vị trí tâm sao chép và cắt sợi đó để tạo ra đầu 3'-OH tự do. Đầu 5' tự do liên kết với protein A trong suốt quá trình tái bản.

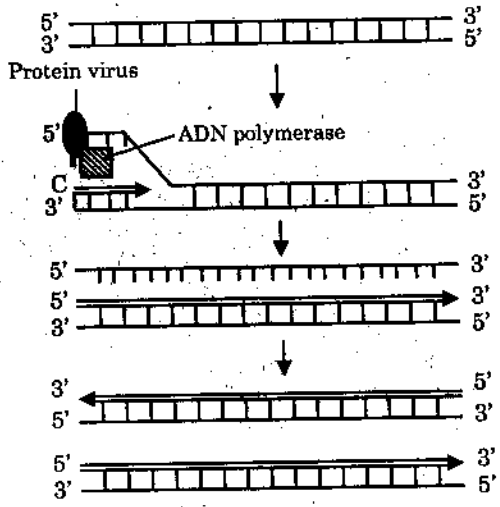
Sử dụng đầu 3' tự do, sợi (+) được tháo dần ra khỏi sợi (-) và sợi (+) mới được tổng hợp thay thế vị trí của sợi (+) cũ theo cơ chế vòng tròn lăn. Kết quả sau một vòng tái bản, một phân tử ADN sợi đôi dạng vòng và một sợi (+) dạng vòng được tổng hợp (Hình 2.8). Sợi (+) tiếp tục làm khuôn để tổng hợp sợi (-) và lặp lại quá trình tổng hợp. Như vậy, số lượng khuôn tăng lên rất nhanh nên số lượng genome virus (số lượng sợi dương) cũng tăng rất nhanh. Sợi (+) được đóng gói trong vỏ protein tạo thể virus mới.



Hình 2.8: Tổng hợp genome ϕX174 sợi đơn dạng vòng theo cơ chế vòng tròn lăn.

2.8.2. TÁI BẢN TỪ ĐẦU MÚT PHÂN TỬ ADN DẠNG THẲNG

Genome adenovirus và ϕ29 là phân tử ADN dạng thẳng. Tái bản được bắt đầu từ một trong hai đầu mút của phân tử ADN thẳng. Đầu tận cùng 5' của sợi khuôn ADN tương tác với protein của virus. Protein này tương tác với nucleotide cytidine (C) và ADN polymerase. Nucleotide cytidine có đầu 3'-OH tự do được dùng để tổng hợp sợi ADN mới. Như vậy, nhờ tương tác với protein, nucleotide C đóng vai trò của mồi để ADN polymerase thực hiện phản ứng. Tổng hợp ADN tạo ra một phân tử mới và một sợi đơn ADN. Sợi này chuyển sang cấu trúc dạng vòng nhờ liên kết bổ sung giữa các trình tự tương đồng ở hai đầu mút (Hình 2.9).



Hình 2.9: Tái bản ADN bắt đầu từ một trong hai đầu mút của phân tử ADN dạng thẳng nhờ protein có khả năng tương tác với ADN ở đầu 5' và tương tác với ADN polymerase.

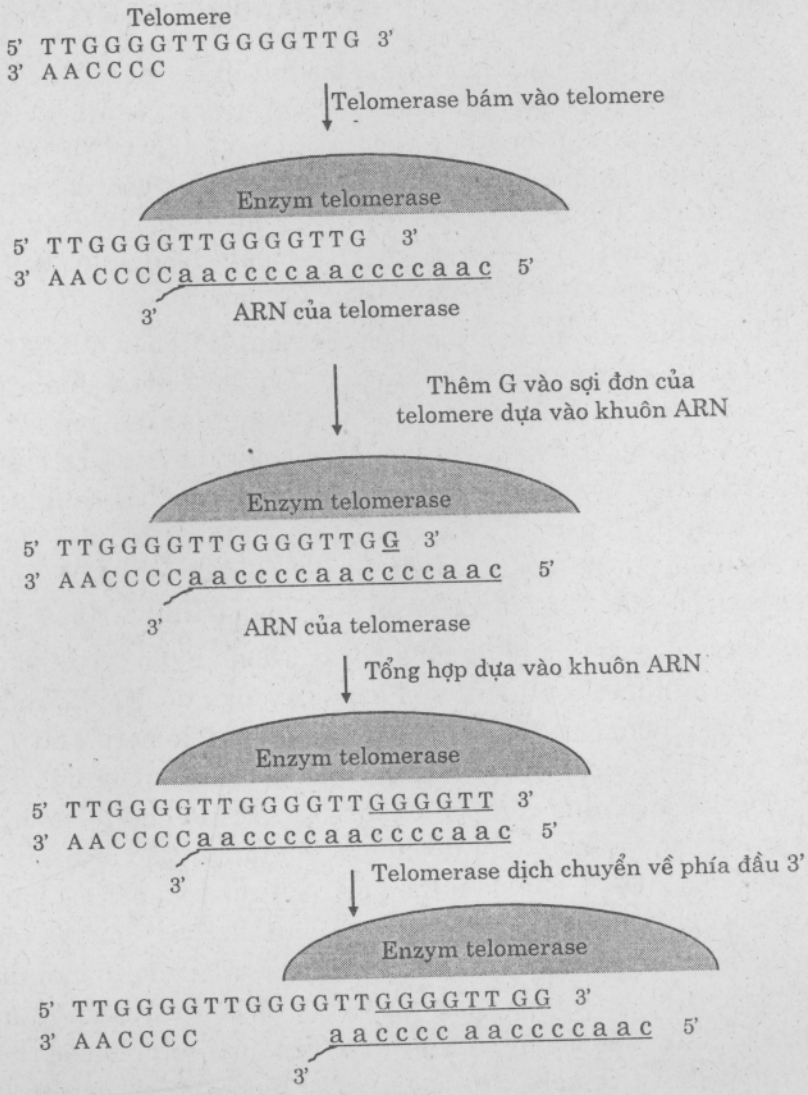
2.9. TÁI BẢN TELOMERE CỦA PHÂN TỬ ADN EUKARYOT

Tất cả các enzym ADN polymerase mà chúng ta biết đều không có khả năng tổng hợp sợi ADN mới từ các nucleotide tự do. Hơn nữa, ADN polymerase chỉ tổng hợp theo một chiều nhất định từ 5'→3'. Do đó, trong quá trình tái bản nhiễm sắc thể eukaryot, các enzym này dường như rất khó đảm bảo nguyên vẹn chiều dài và tính chính xác của đầu tận cùng telomere trên sợi ADN mới (sợi lagging). Trong tế bào, enzym telomerase thực hiện chức năng tái bản các đầu mút telomere.

Telomerase là phức ribonucleoprotein có hoạt tính thêm các nucleotide tự do vào đầu tận cùng của sợi đơn ADN dựa vào khuôn ARN có mặt trong phức. Sợi ARN của telomerase tạo cặp với trình tự giàu C+A lặp lại ở đầu telomere. Một trong những protein tạo nên phức telomerase có hoạt tính phiên mã ngược từ ARN sang ADN. Nhờ đó, các trình tự lặp lại ở đầu mút telomere được kéo dài theo hướng 5'→3'. Sau đấy, đoạn mới này được dùng làm khuôn để ADN polymerase hoàn thiện các đầu tận cùng của sợi ADN mới (Hình 2.10).

Telomerase có khả năng tổng hợp các đoạn ngắn lặp lại ở telomere nhưng không kiểm soát được số lượng các đoạn đó, tức là không kiểm soát được số lượng các đoạn lặp lại. Chiều dài telomere phụ thuộc vào hoạt tính của các protein vào telomere làm ảnh hưởng đến hoạt tính của telomerase. Trong genome eukaryot, gen mã cho telomerase hoạt động trong các tế bào ở giai đoạn sớm của phôi. Các tế bào soma của người tái bản ADN trong điều kiện gen mã cho telomerase không hoạt động. Vì vậy, độ dài telomere bị ngắn dần qua các chu kỳ phân bào. Giữa độ dài của telomere và thời gian sống của tế bào có mối ràng buộc nhất định. Khi telomere càng ngắn thì tế bào càng khó phân chia chính xác. Các hiện tượng đứt gãy, chuyển đoạn nhiễm sắc thể xảy ra càng nhiều khi telomere càng ngắn. Các tế bào nấm men trở nên già cỗi (không phân chia và chết) khi không còn telomere.

Mặc dù telomere gồm các đoạn nucleotide ngắn lặp lại nhưng trao đổi chéo không xảy ra do protein đặc hiệu bám vào telomere. Trong thí nghiệm nuôi cấy tế bào, hầu hết các tế bào đều già hoá và chết sau một số lần phân chia nhất định. Tuy nhiên, một số tế bào bất bình thường vẫn phân bào do nhiễm sắc thể của chúng có telomere dài ra, hậu quả của trao đổi chéo không cân xứng. Những tế bào này thường mang đột biến liên quan đến cơ chế sửa chữa sai hỏng trong tạo cặp bổ sung giữa hai sợi đơn ADN.

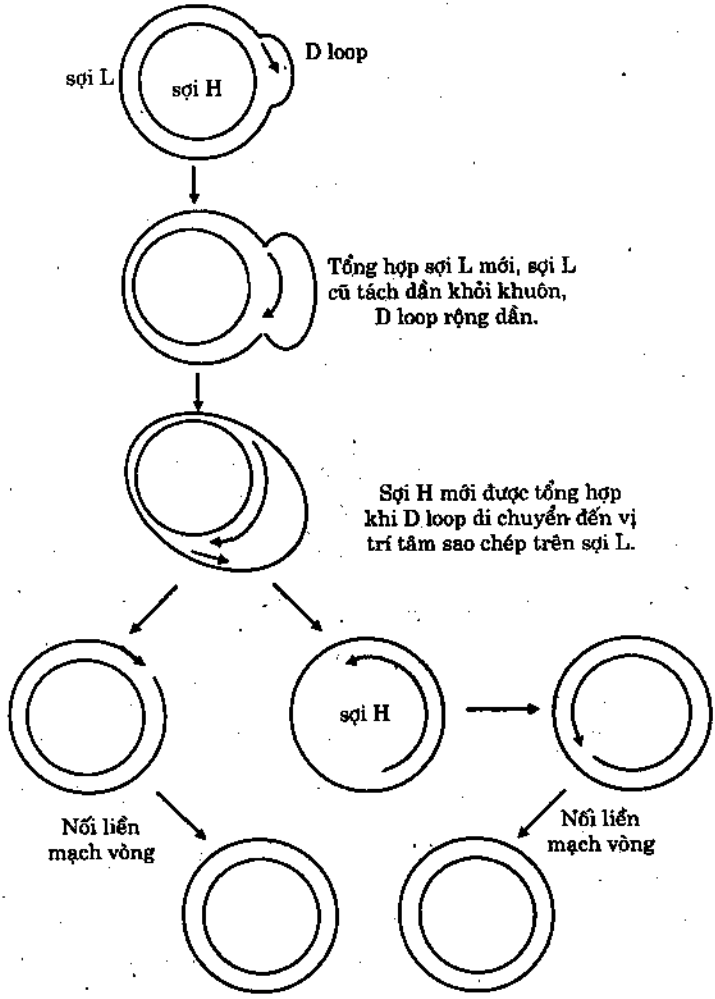


Hình 2.10: Telolerase gồm sợi ARN và phức protein có hoạt tính phiên mã ngược (phiên mã trên khuôn ARN sang ADN). Sợi ARN làm khuôn để tổng hợp trình tự lặp lại ở sợi ADN đơn của telomere (theo Lewin; 2004).

2.10. TÁI BẢN ADN TY THỂ

Tâm sao chép của nhiễm sắc trong vi khuẩn và trong nhân tế bào eukaryot có những tính chất điển hình: đều là trình tự đặc hiệu được nhận biết dưới dạng sợi kép và là vị trí có liên kết bổ sung bị đứt gãy để bắt đầu quá trình tổng hợp. Tuy nhiên, tâm sao chép của phân tử ADN nằm trong ty thể lại có điểm khác biệt với ADN trong nhân.

Trong ty thể, phân tử ADN có cấu trúc sợi đôi dạng tròn. Cả hai sợi đều mang gen mã cho các sản phẩm cần thiết với ty thể. Chiều phiên mã của các gen trên sợi phía ngoài (sợi H) ngược với chiều phiên mã của các gen trên sợi phía trong (sợi L). Đặc biệt, mỗi sợi có một tâm sao chép riêng của mình, trong đó tâm sao chép sợi H có thể tồn tại dưới dạng một đoạn ADN đơn lẻ. Hơn nữa, bắt đầu tái bản không xảy ra đồng thời ở hai tâm riêng biệt (Hình 2.11).



Hình 2.11: Cấu trúc D loop tại tâm sao chép trên sợi đơn H của ADN ty thể động vật.

Mỗi sợi đơn trên phân tử ADN ty thể có một tâm sao chép riêng biệt. Khởi động phiên mã trên hai sợi này không xảy ra tại cùng một thời điểm (theo Lewin, 2004).

Đầu tiên, sợi mồi ARN được tổng hợp tại tâm sao chép trên sợi H khiến sợi L bị tách dần ra khỏi sợi H. Đoạn nucleotide trên sợi L bị tách ra do bắt đầu tổng hợp sợi H, tạo nên cấu trúc gọi là "D loop" (*displacement*). Tái bản được thực hiện với khuôn là sợi H, do đó tiến triển của phản ứng tổng hợp là một sợi L mới được kéo dài và sợi L cũ bị tách dần ra dưới dạng sợi đơn, tức là "D loop" càng rộng dần. Khi D loop giãn rộng đến vị trí tâm sao chép trên sợi L thì tái bản của sợi L mới bắt đầu. Như vậy, hai tâm sao chép khởi động tại thời điểm khác nhau, mỗi tâm tương ứng với một chạc sao chép, chạc này di chuyển ngược chiều chạc kia.

Phân tử ADN ty thể có thể có một hoặc nhiều D loop. ADN ty thể trong tế bào động vật có vú có một D loop với chiều dài khoảng 500-600bp. Sợi ARN mồi tạo cặp không bền với trình tự của D loop. Do thời điểm tái bản của hai sợi đơn H và L không xảy ra đồng thời nên khi sợi H tổng hợp được 1/3 thì sợi L đã tổng hợp xong.

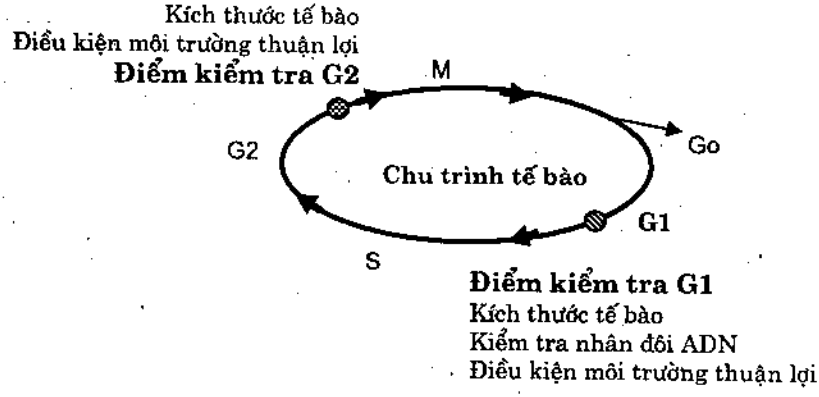
Phân tử ADN trong lục lạp sử dụng tâm sao chép tương tự như D loop để tái bản.

2.11. CHU TRÌNH TẾ BÀO

Trong tế bào eukaryot, tái bản nhiễm sắc thể (xảy ra ở phase S) chỉ là một giai đoạn trong chu trình tế bào. Chu trình tế bào ở sinh vật eukaryot được xem gồm 4 giai đoạn. Ba giai đoạn G1, S và G2 được gọi chung là gian kỳ (interphase). Diễn biến một chu trình tế bào được thực hiện theo đúng trật tự từ G1 sang pha S, đến G2 và kết thúc bởi giai đoạn phân chia tế bào M. Tiến trình này chỉ xảy ra theo một chiều mà không thể đảo ngược lại.

Trong giai đoạn G1, ngay sau giai đoạn phân bào nguyên nhiễm (mitose), tế bào hoàn thiện kích thước riêng để đảm bảo có đủ mọi điều kiện cần thiết cho giai đoạn nhân đôi ADN. Ở giai đoạn G2 sau S, tế bào kiểm tra quá trình nhân đôi ADN, đảm bảo quá trình đó xảy ra không có bất kỳ một sai sót nào. Sau đó tế bào mới chuyển sang phân chia. Như vậy, tế bào thực thi hệ thống kiểm soát chặt chẽ chu trình tế bào (*cell-cycle control system*), đặc biệt tại các điểm kiểm tra nằm giữa hai giai đoạn (*checkpoints*) để quyết định sự chuyển tiếp. Nhiệm vụ của hệ thống kiểm soát là đảm bảo tính chính xác của quá trình nhân đôi ADN, phân ly nhiễm sắc thể về hai cực và hình thành hoàn hảo hai tế bào con. Khi tiếp nhận các tín hiệu không bình thường (từ các diễn biến trong quá trình phân bào hoặc từ môi trường bên ngoài), hệ thống

kiểm soát có thể điều khiển chu kỳ phân bào chậm hoặc dừng lại ở một điểm kiểm tra. Điểm dừng trong pha G1 được gọi là điểm bắt đầu "Start" (ở nấm men), hoặc đơn giản gọi là điểm kiểm tra G1. Khi hệ thống kiểm soát vượt qua điểm G1, chu trình phân bào chuyển sang pha S. Tương tự như ở pha G1, khi hệ thống vượt qua điểm kiểm tra G2, chu trình phân bào được phép bước vào pha M (Hình 2.12).



Hình 2.12: Các tín hiệu môi trường và tín hiệu phản hồi từ các diễn biến xảy ra trong quá trình phân chia cho phép tế bào vượt qua các điểm kiểm tra G1 và G2. Tại các điểm G1 và G2, hệ thống kiểm soát phân tích các điều kiện cần và đủ trước khi cho phép tế bào chuyển sang pha tiếp theo.

Hệ thống kiểm soát hoạt động dựa vào phức gồm hai họ protein. Đó là các protein kinase phụ thuộc cyclin (***cyclin-dependent protein kinase***), viết tắt là **Cdk** và nhóm protein cyclin. Cdk có hoạt tính phosphoryl hoá protein đặc hiệu tại các acid amin serine/threonine. Tuy nhiên khả năng này của Cdk được kiểm soát bởi cyclin. Mỗi Cdk có thể liên kết với nhiều cyclin khác nhau để tạo thành phức Cdk-cyclin có hoạt tính đặc hiệu, tức là quyết định protein bị phosphoryl hoá. Rõ ràng ba sự kiện: tạo phức Cdk-cyclin; hoạt hoá phức và phân rã phức là những bước chuyển rất quan trọng điều khiển toàn bộ chu trình phân bào. Các protein cyclin có thể phân làm hai nhóm chính: các cyclin liên kết với Cdk trong pha G2 để tế bào bước vào mitose (nên gọi là *mitose cyclin*) và các cyclin liên kết với Cdk trong pha G1 để chu trình chuyển sang pha S (nên còn gọi là *G1 cyclin*). Những nghiên cứu đầu tiên trên nấm men cho thấy một nhóm protein Cdk tham gia hệ thống kiểm soát chu trình tế bào ở cả hai điểm kiểm tra G1 và G2. Tuy nhiên ở động vật có vú, mỗi điểm kiểm tra được kiểm soát bởi ít nhất một nhóm Cdk riêng biệt. Hoạt tính của Cdk góp phần quyết định tế

bào tồn tại ở một pha nhất định hoặc chuyển pha. Vì một lý do nào đó mà nhóm Cdk phụ trách ở một điểm kiểm tra được hoạt hoá thì tế bào bước sang pha tiếp theo bất chấp các nguyên liệu cần thiết (protein, ADN...) để hoàn thiện pha đó có đầy đủ hay không. Rõ ràng sai sót ở hệ thống kiểm soát có thể khiến tế bào phân chia liên tục.

2.12. SỬA CHỮA ADN

Như chúng ta đã biết, hầu hết sai sót trong quá trình tổng hợp ADN được nhận biết và sửa chữa nhờ hoạt tính đọc sửa của ADN polymerase. Tuy nhiên, các tác nhân ở bên trong tế bào (các chất oxy hoá, các gốc tự do... không liên quan đến tái bản) cũng như các tác nhân môi trường (chiếu xạ, sốc nhiệt, các chất độc...) đều có khả năng gây tổn thương sai hỏng ADN. Xuất phát từ quan điểm của các nhà vật lý và hoá học, 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 cytosinè thành uraxine 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.

Tuy nhiên, trong thực tế, phân tử ADN có tỷ lệ đột biến tự phát rất thấp, tức là thông tin di truyền trên phân tử ADN được bảo toàn rất cao. Chúng ta có thể đánh giá tỷ lệ đột biến xảy ra tự phát căn cứ vào trình tự acid amin của một loại protein có mặt 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 phả hệ). Sự sai khác đó do các đột biến ở mức độ gen, tức là ở các nucleotide. Từ đây có thể đánh giá được thời gian cần thiết để xảy ra một đột biến ở mức độ ADN. Gần đây, kỹ thuật xác định trình tự cho phép so sánh trình tự nucleotide ở một số vùng không chứa gen trong genome của các loài có quan hệ gần gũi trong tiến hoá. 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. Một

gen có khoảng 10^3bp mang mã di truyền (tương ứng với các protein có kích thước trung bình trong tế bào) chỉ 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. 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. Rõ ràng, tỷ lệ đột biến cao sẽ khiến cho quá trình tiến hoá phải dừng lại ở những loài sinh vật không thể đa dạng và phức tạp hơn ruồi giấm.

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. Tính bền vững của tế bào sinh dục và tế bào soma đều phụ thuộc vào tính bền vững của ADN. 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 đồng thời tiến hành sửa chữa sai hỏng bằng nhiều cách thức khác nhau.

Hầu hết các đột biến tự phát trên phân tử ADN xảy ra hàng ngày hoặc do các tác nhân bên ngoài gây ra (UV, tác động vật lý, hoá học...) được khắc phục bởi 3 cách thức sửa chữa chính. Thứ nhất là sửa chữa phục hồi trực tiếp (direct repair); thứ hai là cắt bỏ sai hỏng và sửa chữa lại dựa vào trình tự bổ sung (excision repair, recombination repair) và thứ ba là chống chịu các tổn thương trong các trường hợp cần thiết như đang tái bản hoặc đang phiên mã (inducible damage tolerance). Cả ba cách thức sửa chữa đều liên quan đến sản phẩm của nhiều gen. Khoảng 130 gen tham gia sửa chữa các loại tổn thương trong genome người (Bảng 2.2).

Sửa chữa trực tiếp thường liên quan đến enzym photolyase và hai loại sai hỏng trên phân tử ADN 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. Photolyase có khả năng 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 thực vật.

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.

Bảng 2.2: Một số cơ chế và các gen liên quan đến sửa chữa tổn thương ADN trong genome người (theo Lewin, 2004).

Cơ chế sửa chữa ADN tổn thương	Số lượng gen liên quan
Sửa chữa trực tiếp	1
Cắt bỏ các base sai hỏng	15
Cắt bỏ các nucleotide sai hỏng	28
Sửa chữa tạo cặp bổ sung sai lệch	11
Sửa chữa dựa vào trình tự tương đồng	14
Sửa chữa không dựa vào các đoạn ADN tương đồng	5
Sử dụng enzym có hoạt tính khác nhau	16

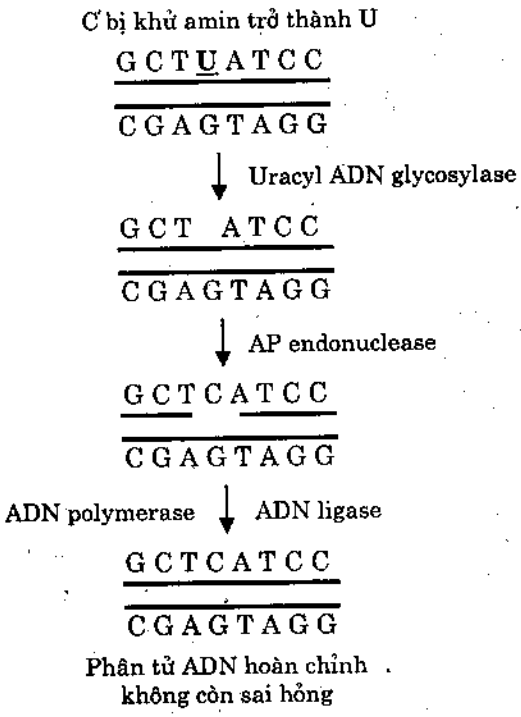
2.13. CẮT BỎ ĐOẠN ADN BỊ SAI HỎNG

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 ADN glycosylase. 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, 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 liện 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.

2.13.1. SỬA CHỮA ADN THEO CƠ CHẾ LOẠI BỎ BASE HỎNG (BASE EXCISION REPAIR - BER)

Các base sai hỏng được nhận biết và cắt ra khỏi phân tử đường bởi ADN glycosydase. 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 uracin khỏi liên kết sai lệch U-G được mô tả trên Hình 2.13. Tuy nhiên enzym này không nhận biết được các cytosine bị methyl hoá ở vị trí C_5. 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

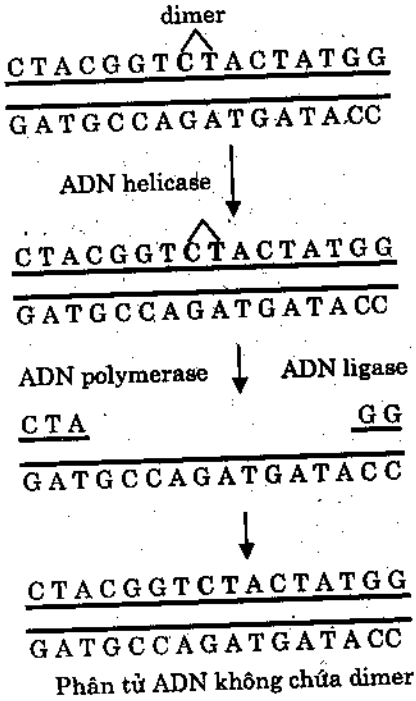


Hình 2.13: Cơ chế loại bỏ base hỏng bởi ADN glycosylase. Enzym uracil ADN-glycosylase nhận biết và cắt U khỏi liên kết U-G (lúc này ta có hiện tượng cytosine bị khử amin). Đường deoxyribose thiếu liên kết với base được nhận biết bởi AP endonuclease và được sửa chữa.

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.

2.13.2. Sửa chữa ADN theo cơ chế loại bỏ nucleotide hỏng (nucleotide excision repair - NER)

Loại bỏ nucleotide sai hỏng là 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. Cơ chế này có sự tham gia của nhiều protein để nhận biết tìm ra sai hỏng trên phân tử ADN. 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 2.12.1. chỉ cắt từng base riêng biệt). Đoạn này sau đó được tổng hợp mới nhờ ADN polymerase và ADN ligase (Hình 2.14).



Hình 2.14: Sửa chữa ADN theo cơ chế loại bỏ ADN hỏng bởi ADN helicase. Sau khi phức các enzym nhận biết sai hỏng dimer (pyrimidine dimer), ADN helicase cắt hai phía của sợi đơn chứa dimer, loại bỏ một đoạn ADN. ADN polymerase và ADN ligase có nhiệm vụ thay thế đoạn bị cắt. Phân tử ADN trở lại nguyên vẹn không có dimer.

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. Phức này chỉ 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. Phức UvrBC sẽ cắt sợi ADN ở hai phía của vị trí có dimer. Vết cắt ở phía đầu 3' sẽ cách vị trí sai hỏng khoảng 3-4 nucleotide. Vết cắt ở phía đầu 5' sẽ cách vị trí sai hỏng khoảng 7 nucleotide. Như vậy đoạn sợi đơn chứa dimer dài khoảng 12 nucleotide bị loại ra khỏi phân tử ADN nhờ protein UvrD. Protein UvrD có hoạt tính helicase khiến phân tử ADN mở xoắn cục bộ, loại bỏ đoạn bị cắt ra. 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 để nhận biết và thực hiện cắt bỏ đoạn sợi đơn theo cơ chế NER. Hầu hết (99%) các trường hợp sửa chữa theo cơ chế NER ở vi khuẩn *E.coli* đều cắt một đoạn sợi đơn dài 12 nucleotide gồm cả nucleotide hỏng. 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 gồm 3 thành phần ARN polymerase + ADN khuôn (bị hỏng) +ARN. Nhờ đó phức này có thể vượt qua vị trí hỏng trong quá trình phiên mã. 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 nucleotide.

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

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 tái bản AND, hoặc phiên mã 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. Trong khi tỷ lệ sai sót của cơ chế đọc sửa trong tái bản hoặc sửa chữa BER và NER rất thấp thì sửa chữa Error-Prone làm tăng đột biến; tuy nhiên những đột biến này làm tăng khả năng chống chịu của cơ thể đối với những tác nhân gây tổn thương genome.

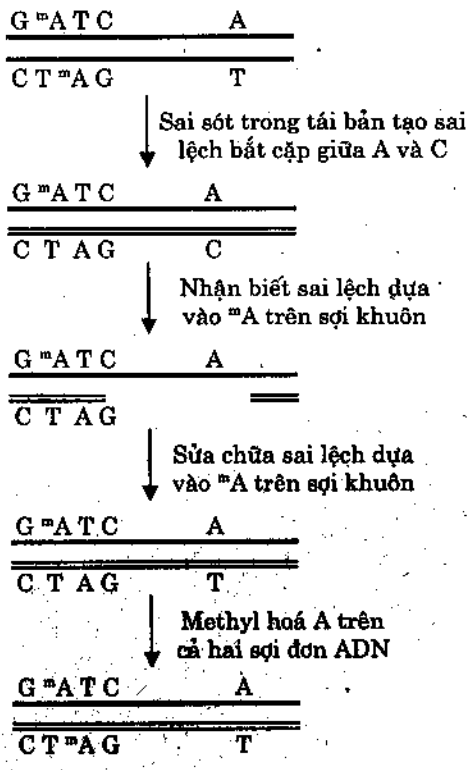
Sửa chữa Error-Prone được phát hiện lần đầu tiên khi đưa ADN thực khuẩn thể bị tổn thương vào tế bào *E.coli* đã bị chiếu xạ tia tử ngoại UV. Phân tử ADN thực khuẩn thể được sửa chữa nhưng đồng thời xuất hiện các đột biến mới. Kết quả này cho thấy chiếu xạ UV đã hoạt hoá các gen liên quan đến việc gây đột biến. Sản phẩm của những gen này đã tác động lên ADN của *E.coli* cũng như của thực khuẩn thể. Bằng các thí nghiệm di truyền, vai trò quan trọng của hai protein UmuC và UmuD ở *E.coli* tham gia cơ chế sửa chữa "Error-Prone" được phát hiện. Đột biến hai gen này khiến cho tế bào *E.coli* bị chết sau khi chiếu xạ. Như vậy, mặc dù protein UmuC và UmuD làm tăng tỷ lệ đột biến nhưng chúng tăng khả năng sống sót cho tế bào khi genome bị tổn thương.

Protein UmuC và UmuD 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. Các enzym ADN polymerase IV và ADN polymerase V có khả năng vượt qua sai hỏng trên sợi khuôn (như vượt qua vị trí có dimer) để tiếp tục tổng hợp ADN.

Ngoài ra khi cả hai sợi đơn ADN đều chịu đột biến nhưng nếu trong genome còn có ít nhất một bản sao 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.

2.15. SỬA CHỮA ADN GHÉP ĐÔI LỆCH (MISMATCH REPAIR)

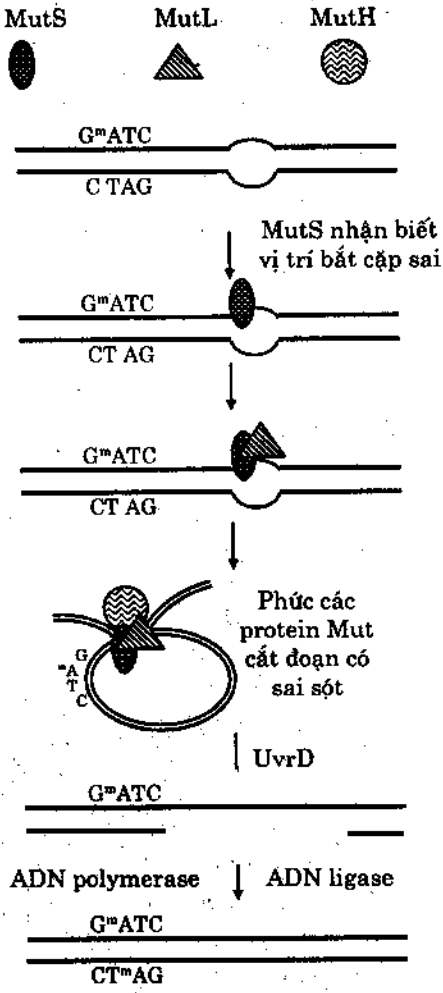
Tế bào luôn kiểm tra tính chính xác về cấu trúc và trình tự nucleotide của phân tử ADN trước khi bước vào phase S, tức là trước khi tái bản ADN. Một số sai hỏng có thể xảy ra từ lần tái bản trước (sai sót không bị phát hiện bởi cơ chế đọc sửa) hoặc xuất hiện sau khi tái bản (không bị phát hiện bởi cơ chế BER, NER). Những sai hỏng này khiến cho hai nucleotide trên hai sợi đơn không bổ sung với nhau hoặc gây nên sự thay đổi cấu trúc cục bộ do một vài nucleotide bớt đi hay chèn thêm vào một sợi đơn. Sự thêm, bớt một vài nucleotide có thể xuất hiện do sự trượt của sợi khuôn cũng như sợi đang tổng hợp ở những đoạn ADN lặp lại trong phản ứng tổng hợp ADN. Những sai hỏng này dẫn đến sự sai lệch trong bắt cặp giữa hai sợi ADN bổ sung. Do đó, tế bào có cơ chế nhận biết và sửa chữa sai lệch bắt cặp. Tuy nhiên ở đây cần phải phân biệt được sợi đơn đúng với sợi đơn có nucleotide sai, hoặc có vài nucleotide bị thay thế, hoặc thêm bớt.



Hình 2.15: Nhận biết và sửa chữa sai lệch bắt cặp ADN sinh ra trong quá trình tái bản dựa vào hiện tượng methyl hoá adenine trong trình tự GATC trên sợi khuôn.

Nucleotide sai hỏng có thể không bị phát hiện bởi cơ chế đọc sửa của ADN polymerase trong quá trình tái bản. Điều đó dẫn đến sai lệch bắt cặp giữa sợi ADN khuôn và sợi ADN đang được tổng hợp. Sai lệch bắt cặp này được phát hiện do sợi đang được tổng hợp chưa bị methyl hoá trong khi adenine A ở trình tự GATC trên sợi khuôn bị methyl hoá. Do đó có thể phân biệt sợi mới với sợi khuôn. Ở vi khuẩn *E.coli*, adenine A trong chuỗi GATC bị methyl hoá tại vị trí N^6 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 (Hình 2.15). Vi khuẩn *E.coli* mang đột biến ở gen *dam* mã cho ADN methylase có tỷ lệ đột biến tự phát tăng cao. Enzym này cùng phối hợp với các protein MUT trong cơ chế sửa chữa bắt cặp sai lệch. Bên cạnh tái bản, trao đổi chéo có thể làm xuất hiện nucleotide sai hỏng gây sai lệch bắt cặp (các dị ghép đôi-heteroduplex). Trong trường hợp này cả hai sợi đơn đều chứa nhóm methyl nên sửa chữa xảy ra với hiệu suất thấp và tính chính xác không cao.

Trong tế bào *E.coli*, cơ chế sửa chữa ghép đôi lệch đòi hỏi sản phẩm của bốn gen ký hiệu là *mutH, mutL, mutS* và *uvrD* (*mutator*). Chúng được gọi là "*mutator*" bởi 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. Protein MutS có hai vị trí tương tác với ADN. Vị trí thứ nhất nhận biết bắt cặp sai lệch. Vị trí thứ hai giúp MutS di chuyển dọc sợi ADN đến nơi có trình tự GATC. Do MutS đồng thời tương tác với ADN tại hai vị trí nên đoạn ADN giữa hai vị trí đó bị uốn cong có dạng vòng. MutS liên kết với MutL tạo phức MutSL. Khi gặp trình tự GATC, phức MutSL tương tác với MutH. Protein MutH có hoạt tính endonuclease phân cắt đầu 5' ngay tại vị trí G của chuỗi GATC không chứa nhóm methyl. Enzym này đòi hỏi phức 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 2.16.



Phân tử ADN không có các nucleotide bắt cặp sai

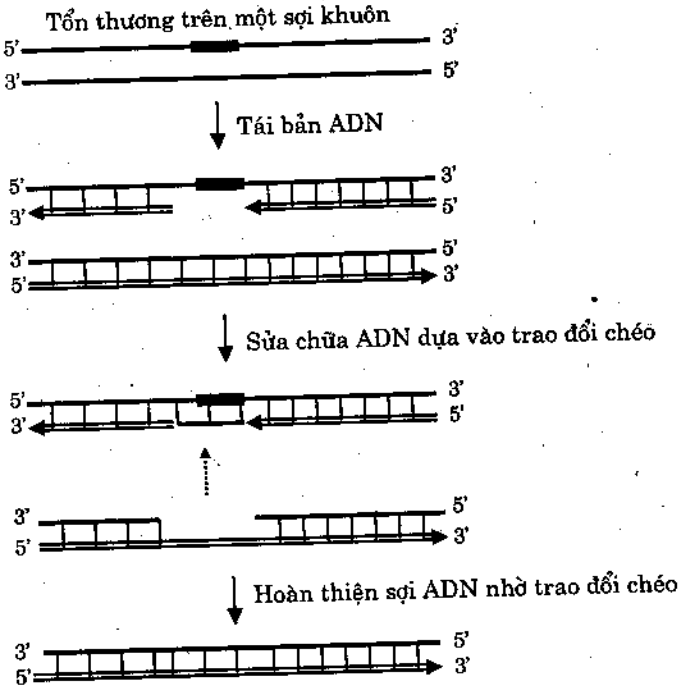
Hình 2.16: Cơ chế sửa chữa bắt cặp sai lệch ở *E.coli*.

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. Protein MutH cắt sợi đơn tại vị trí GATC và protein UvrD mở xoắn đưa đoạn chứa lỗi ra. Đoạn này sau đó bị phân hủy bởi endonuclease. Khoảng trống được sửa chữa phục hồi nhờ ADN polymerase và ADN ligase.

2.16. SỬA CHỮA ADN DỰA VÀO TRAO ĐỔI CHÉO (RECOMBINATION REPAIR SYSTEM)

Tất cả các loại tế bào đều có nhiều cách thức sửa chữa phân tử ADN bị tổn thương. Việc lựa chọn con đường nào phụ thuộc vào kiểu tổn thương và trạng thái sinh lý của tế bào. Cơ chế cắt bỏ đoạn chứa

sai hỏng (NER và BER) được sử dụng chủ yếu trong nhiều trường hợp tổn thương khác nhau. Tuy nhiên, trong quá trình tái bản ADN, nếu sợi khuôn bị tổn thương thì sợi tổng hợp trên sợi khuôn hỏng được sửa chữa theo cơ chế dựa vào trao đổi chéo giữa hai trình tự tương đồng.



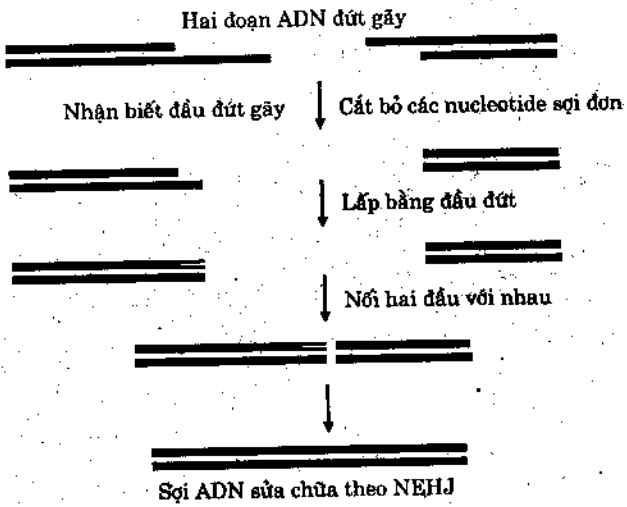
Hình 2.17: Sửa chữa sai hỏng (thiếu đoạn) trên sợi được tổng hợp dựa vào trình tự không sai trên sợi tương đồng. Đoạn ADN không tổn thương của sợi khuôn thay thế cho đoạn bị thiếu.

Trong quá trình tái bản ADN, hai phân tử giống hệt nhau được tổng hợp dựa vào hai sợi khuôn. Khi một sợi khuôn bị tổn thương thì các loại ADN polymerase đặc biệt có thể tiếp tục tổng hợp vượt qua vị trí sai hỏng trên khuôn. Tuy nhiên, sợi được tổng hợp có thể chứa các sai hỏng khác nhau (thiếu hụt hoặc thay thế nucleotide). Do đó, sau tái bản, phân tử ADN dạng sợi đôi sẽ gồm 2 sợi đơn dị hợp tử (heteroduplex sequences). Những sai hỏng đó được khắc phục nhờ trao đổi chéo giữa hai phân tử ADN sau tái bản (Hình 2.17). Trong tế bào *E.coli*, quá trình tổng hợp ADN bị dừng do sai hỏng trên phân tử khuôn đều tạo ra các tín hiệu cấp cứu (***SOS response***). Những tín hiệu này 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 hủy protein ức chế các gen liên quan đến cơ chế SOS, giúp cho quá trình tổng hợp ARNm và tổng hợp

protein tương ứng 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.

2.17. SỬA CHỮA ADN TRONG TẾ BÀO EUKARYOT

Các cơ chế sửa chữa ADN mà chúng ta vừa xem xét được nghiên cứu chi tiết ở vi khuẩn *E.coli*. Những cơ chế này cũng được biểu hiện ở sinh vật eukaryot *S.cerevisiae*. Khi chiếu xạ gây tổn thương ADN và phân tích các dòng nấm men đột biến cho thấy các gen liên quan đến sửa chữa ADN ở nấm men có thể xếp vào 3 nhóm chính. Nhóm thứ nhất gồm các gen RAD3 tham gia con đường cắt bỏ ADN sai hỏng (excision repair). Nhóm thứ hai gồm các gen RAD6 tham gia sửa chữa ADN sau tái bản (post-replication repair). Nhóm gen thứ ba gồm các gen RAD52 tham gia sửa chữa ADN theo cơ chế tương tự trao đổi chéo tương đồng (recombination-like repair). Đặc biệt, ADN đứt gãy cả hai sợi còn được sửa chữa theo cơ chế không phụ thuộc vào trình tự tương đồng (***non-homologous end-joining - NHEJ***). Theo cơ này, hai đoạn ADN đứt gãy được nối với nhau (Hình 2.18).



Hình 2.18: Nối hai đoạn ADN đứt gãy cả hai sợi đơn theo cơ chế không phụ thuộc vào trình tự tương đồng (NHEJ).

Tế bào *E.coli* có 3 loại ADN polymerase (I, II, III) tham gia tái bản ADN trên sợi khuôn bình thường. Tuy nhiên, vi khuẩn còn có ADN polymerase IV (mã bởi gen *dinB*) và ADN polymerase V (mã bởi *umuCD*); những enzym này có thể thay thế ADN polymerase bình thường tại vị trí sai hỏng trên sợi khuôn. Nhờ đó, tái bản ADN vẫn tiếp tục ngay khi sợi khuôn bị tổn thương. Các enzym ADN polymerase đặc biệt cũng được tìm thấy trong tế bào eukaryot. Ví dụ, gen *rad30* ở nấm men hoặc gen *XPV* ở người đều mã cho ADN polymerase η có khả năng vượt qua nucleotide sai hỏng. Ví dụ, hai thymidine (T) trên một sợi khuôn có thể liên kết với nhau tạo dimer T-T. Dimer T-T được nhận biết và sửa chữa bởi cơ chế cắt bỏ đoạn sai hỏng. Tuy nhiên khi dimer T-T chưa được sửa thì ADN polymerase η sẽ bất chấp tiếp tục tổng hợp. Sự khác nhau giữa các loại ADN polymerase đặc biệt của vi khuẩn và eukaryot ở chỗ enzym vi khuẩn thêm nucleotide bất kỳ (theo kiểu "error-prone"), còn enzym eukaryot thêm những nucleotide nhất định vào vị trí cần sửa chữa. Ví dụ, khi gặp liên kết giữa hai thymidine (T) trên một sợi khuôn (dimer T-T), enzym vi khuẩn sẽ tự động thêm nucleotide bất kỳ vào sợi đang tổng hợp để vượt qua vị trí hỏng. Trong trường hợp này, enzym ở eukaryot sẽ thêm chính xác cặp A-A vào sợi đang tổng hợp, tương ứng với T-T trên sợi khuôn.

Điều thú vị là sửa chữa ADN ở nấm men có liên quan đến phiên mã. Sửa chữa các gen đang phiên mã (gen hoạt động) được ưu tiên trước so với những gen không hoạt động. Thậm chí ngay tại một gen, sợi đơn chứa mã di truyền (được phiên mã) được ưu tiên trước so với sợi bổ sung. Thực nghiệm chứng minh rằng protein RAD3 có hoạt tính helicase trong sửa chữa cắt bỏ đoạn sai hỏng còn là yếu tố liên kết với ARN polymerase II trong quá trình phiên mã.

Chương 3

PHIÊN MÃ VÀ KIỂM SOÁT PHIÊN MÃ

Phiên mã là quá trình tổng hợp nên các loại ARN dựa trên khuôn là một trong hai sợi đơn của phân tử kép ADN. Quá trình phiên mã tạo ra các loại ARN chức năng như ARN ribosome (ARNr), ARN vận chuyển (ARNt)... Những loại ARN này không chứa mã di truyền cho protein nên chúng không được dịch mã mặc dù sự tồn tại của chúng gắn liền với quá trình dịch mã. Chỉ có các ARNm được dịch mã thành protein. Ngoài ra, trong tế bào còn nhiều loại ARN có kích thước nhỏ và chức năng khác nhau. Chúng được gọi chung là ARN không mang mã (*non coding ARNs*). Mặc dù không chứa các mã bộ ba cho acid amin nhưng thông tin di truyền trên những phân tử này cần thiết cho hoạt động khác nhau của tế bào. Chúng tham gia trực tiếp vào quá trình tái bản ADN (ở telomere), làm thay đổi trình tự nucleotide, kiểm soát biểu hiện của các gen, thay đổi vị trí các đoạn ADN trên nhiễm sắc thể cũng như thiết lập cấu trúc không gian của ADN ở các vùng dị nhiễm sắc (heterochromatin formation). Do đó, các ARN không chứa mã di truyền có tác động trực tiếp đến phân tử ADN mang mã di truyền. Ngoài ra, ARN tham gia các cơ chế kiểm soát quá trình phiên mã, kiểm soát sau phiên mã như cắt nối intron-exon (*RNA splicing*), đọc sửa ARN (*RNA editing*), phân cắt ARNm (*post-transcriptional gene silencing*), kìm hãm hoạt động của các gen nhảy (*transposon silencing*) v.v...

Quá trình phiên mã gồm nhiều bước nối tiếp nhau, mỗi bước đều chịu sự kiểm tra chặt chẽ bởi các cơ chế đan xen lẫn nhau. Ví dụ, trong nhân tế bào eukaryot, một gen bắt đầu hoạt động khi cấu trúc của nó trong nhiễm sắc thể phải được thay đổi (cơ chế khử methyl ở cytosine, biến đổi một số liên kết ở histone, có sự sắp xếp lại các đoạn ADN v.v...). Những thay đổi đó khiến một promoter trở nên nhạy cảm với DNase. Vì vậy, promoter được xem là đang tồn tại ở trạng thái sẵn sàng hoạt động khi nó không bị methyl hoá và dễ dàng bị phân cắt bởi DNase. Chúng ta nhận thấy kiểm soát cấu trúc không gian, thông qua cấu trúc nucleosome, trở thành yếu tố đầu tiên cho phép một gen

chuyển từ trạng thái bất hoạt (silent state) sang trạng thái sẵn sàng cho khởi động phiên mã (active state).

Trong chương này, chúng ta đặc biệt quan tâm đến phiên mã và kiểm soát phiên mã tạo ra các sợi ARN mang mã di truyền của protein: các phân tử ARN thông tin (ARNm). Kiểm soát phiên mã liên quan đến tương tác giữa các yếu tố cần thiết cho khởi động phiên mã, các phức kiểm tra việc bắt đầu và kết thúc tổng hợp ARNm. Điểm khác nhau cơ bản giữa phiên mã ở tế bào prokaryot và eukaryot là khởi động phiên mã trên gen eukaryot đòi hỏi phức gồm nhiều protein. Do cấu trúc tế bào prokaryot và eukaryot khác nhau nên ARNm eukaryot được phiên mã trong nhân nhưng phải vận chuyển ra tế bào chất để làm khuôn cho quá trình dịch mã. Từ nhân ra đến ngoài tế bào chất, phân tử ARNm được biến đổi và chịu sự kiểm tra về số lượng cũng như thông tin di truyền có trên ARNm.

Tế bào *E.coli* chỉ có một loại ARN polymerase xúc tác cho phản ứng tổng hợp tất cả các loại ARN, kể cả ARNm. Trong tế bào eukaryot, phản ứng tổng hợp ARNm được xúc tác bởi ARN polymerase II, trong khi ARNr (18S, 5,8S và 28S) được tổng hợp bởi ARN polymerase I và ARNt được tổng hợp bởi ARN polymerase III. Phản ứng tổng hợp ARNm gồm các giai đoạn khởi đầu phiên mã tại promoter, tổng hợp phân tử ARNm và kết thúc tổng hợp tại vị trí dừng thích hợp. Chúng ta cần phân biệt một số khái niệm sau:

1) Yếu tố phiên mã (*transcription factor*) là protein cần thiết để khởi động phiên mã. Yếu tố phiên mã có thể tương tác với trình tự nucleotide đặc biệt, tương tác với yếu tố phiên mã khác hoặc tương tác với ARN polymerase.

2) Bộ máy phiên mã cơ bản (*basal transcription apparatus*) : phức gồm các yếu tố phiên mã đã định vị tại promoter trước khi ARN polymerase bám vào đó để khởi động phiên mã.

3.1. MÃ DI TRUYỀN BỘ BA

Một mã di truyền gồm 3 nucleotide tương ứng với một acid amin; vì vậy mã di truyền còn được gọi là mã bộ ba. Tổ hợp của bốn loại nucleotide khác nhau tạo thành 64 mã bộ ba, trong đó 61 mã ứng với các acid amin, 3 mã dùng để kết thúc phản ứng tổng hợp protein. 61 mã này được nhận biết bởi các đối mã trên phân tử ARNt đặc hiệu cho acid amin, còn 3 mã dừng (stop codon) được nhận biết bởi protein có

vai trò của yếu tố kết thúc dịch mã. 61 mã di truyền chỉ tương ứng với 20 acid amin.

Như vậy, nhiều mã bộ ba mã cho một acid amin (Bảng 3.1). Các mã này được gọi là mã thoái hoá. Rất có thể số mã nhiều hơn số acid amin giúp tế bào có khả năng dự trữ thông tin di truyền, tăng tính thích nghi với môi trường. Khi một mã bị đột biến thì có thể duy trì sự sống nhờ các mã di truyền thoái hoá khác.

Bảng 3.1: Mã di truyền bộ ba và hiện tượng thoái hoá mã

Thay đổi nucleotide thứ ba của mã			
UUU, UUC — Phe	UCU, UCC, UCA, UCG — Ser	UAU, UAC — Tyr	UGU, UGC — Cys
UUA, UUG, CUU, CUC, CUA, CUG — Le		UAA, UAG — mã dừng	UGA — mã dừng
			UGG — Trp
	CCU, CCC, CCA, CCG — Pro	CAU, CAC — Gln	CGU, CGC, CGA, CGG — Ar
		CAA, CAG — His	
AUU, AUC, AUA — Ile	ACU, ACC, ACA, ACG — Thr	AAU, AAC — As	AGU, AGC — Ser
		AAA, AAG — Lys	AGA, AGG — Ar
AUG — Met			
GUU, GUC, GUA, GUG — Val	GUU, GUC, GUA, GUG — Ala	GAC, GAC — As	GGU, GGC, GGA, GGG — Gly
		GAA, GAG — Glu	

Tương tác giữa mã bộ ba với đối mã trên ARNt không đòi hỏi sự tạo cặp chính xác giữa 3 nucleotide của một mã di truyền (trên phân tử ARNm) với 3 nucleotide của đối mã (trên phân tử ARNt). Ví dụ, nucleotide U đầu tiên của đối mã có thể nhận biết nucleotide thứ ba A hoặc G của mã di truyền (Bảng 3.2).

Bảng 3.2: Tương tác giữa mã bộ ba và đối mã không dựa vào sự tạo cặp bổ sung

Vị trí thứ nhất của đối mã	Vị trí thứ ba của mã bộ ba
U	A hoặc G
C	G
A	U
G	C hoặc U

Mã di truyền trình bày trong Bảng 3.1 là những mã phổ biến với mọi đối tượng từ prokaryot đến eukaryot. Tuy nhiên, trong một số trường hợp, mã di truyền thay đổi. Hầu hết những thay đổi xảy ra với mã dừng (stop codon). Ví dụ, ở động vật đơn bào *Mycoplasma capricolum*, mã UGA không được dịch là mã dừng (stop codon) mà lại trở thành mã cho tryptophan. Điều này có thể liên quan đến đột biến ở gen mã cho yếu tố nhận biết mã dừng. Ngoài ra, mã di truyền có thể thay đổi như trường hợp CUG mã cho leusine bị đọc thành mã của serine (ở nấm *Candida*). Hiện tượng này liên quan đến đột biến ở gen mã cho ARNt. Trong thực tế, đột biến xảy ra ở đối mã của ARNt có thể nhận biết đột biến trong mã bộ ba (gen) như một mã bình thường. Lúc đó, trình tự acid amin trên chuỗi peptide hoàn toàn bình thường mặc dù đột biến vẫn tồn tại trên gen mã cho peptide đó. Trong trường hợp này, đột biến trên gen ARNt ức chế biểu hiện của đột biến trên gen mã cho protein. Hiện tượng đột biến thứ hai tác động đến biểu hiện của đột biến thứ nhất được gọi là đột biến ức chế (***suppressor***).

3.2. PROMOTER

Như đã trình bày trong Chương 1, cấu trúc của một gen gồm 2 phần chính: phần mang các mã di truyền và phần chứa các nucleotide liên quan đến chức năng điều khiển phiên mã từ ADN (gen) sang sợi ARN. Vị trí nucleotide đầu tiên trên ADN được phiên mã sang ARN được ký hiệu là +1. Với quy định này, các nucleotide nằm trong phần điều khiển (ở vùng 5' phía trước vị trí +1) sẽ được ký hiệu là vị trí (-). Vùng điều khiển của gen bao gồm promoter và các trình tự đặc biệt (*response element*) được nhận biết bởi các yếu tố phiên mã (*transcription factor*).

Promoter là trình tự nucleotide cần thiết để ARN polymerase bám vào bắt đầu phiên mã. Trình tự này nằm ngay trước vị trí (+1) ở các gen prokaryot. Tuy nhiên, promoter của gen eukaryot thường được

xem xét gồm vị trí nhận biết bởi ARN polymerase và các trình tự đặc biệt nhận biết bởi yếu tố phiên mã. Đó là do phiên mã trên gen prokaryot có thể xảy ra ngay khi một mình ARN polymerase bám vào promoter trong khi điều này hầu như không xảy ra với gen eukaryot.

Tiến hành thí nghiệm gây đột biến thay thế hoặc mất đoạn các nucleotide của promoter cho phép xác định chính xác các nucleotide không thể thiếu đối với khởi đầu phiên mã. Kết quả cho thấy promoter của gen mã cho protein thường gồm 2 trình tự riêng biệt phân bố ngay trước điểm khởi đầu phiên mã. Trình tự cơ bản của promoter (*core promoter*) thường gồm các nucleotide ở vị trí -10 đến vài nucleotide sau +1 đối với gen vi khuẩn hoặc gồm các nucleotide ở vị trí -40 đến +20 đối với gen eukaryot. Hầu hết trình tự cơ bản ở gen prokaryot và eukaryot đều có chứa một đoạn ngắn giàu A-T gọi là hộp TATA. Hộp TATA ở vị trí -10 ở promoter vi khuẩn nhưng phân bố ở vị trí -25 ở promoter eukaryot. Khi promoter chỉ có trình tự cơ bản thì mức độ phiên mã xảy ra rất thấp. Mức độ phiên mã tăng rất nhiều khi promoter có thêm trình tự nằm phía trước trình tự cơ bản. Trình tự thứ hai này thường nằm ở vị trí -35 đối với gen vi khuẩn, hoặc vị trí -70 đến -90 ở gen eukaryot. ARN polymerase vi khuẩn có thể tự nó nhận biết promoter và khởi đầu phiên mã. Tuy nhiên ARN polymerase II không thể tự mình khởi động phiên mã nếu như không có sự trợ giúp của các yếu tố phiên mã.

Điều cần lưu ý là một số gen không có cấu trúc promoter điển hình, hoặc các gen hoạt động ở những điều kiện khác nhau đòi hỏi ARN polymerase liên kết với các yếu tố khác nhau. Đối với vi khuẩn chỉ có một loại ARN polymerase, nhưng lõi enzym có thể liên kết với các yếu tố sigma (σ) khác nhau để phiên mã. Ví dụ, ở điều kiện bình thường, ARN polymerase vi khuẩn có chứa yếu tố σ_{70} phiên mã. Khi nhiệt độ môi trường tăng cao, ARN polymerase của vi khuẩn sẽ liên kết với yếu tố σ_{32} để nhận biết promoter của các gen hoạt động khi nhiệt độ tăng bất thường (các gen chịu shock nhiệt). Khi môi trường thiếu nguồn dinh dưỡng nitơ thì yếu tố σ_{54} được sử dụng để phiên mã. Đặc biệt ARN polymerase chứa σ_{54} nhận biết promoter từ vị trí -24 đến -12 ; trong khi với các tố sigma khác, enzym nhận biết promoter từ vị trí -35 và -10. Dường như yếu tố σ_{54} có vai trò như yếu tố phiên mã đối với gen eukaryot. Các yếu tố sigma tham gia cấu trúc của ARN polymerase để nhận biết các promoter khác nhau được trình bày trên Bảng 3.3.

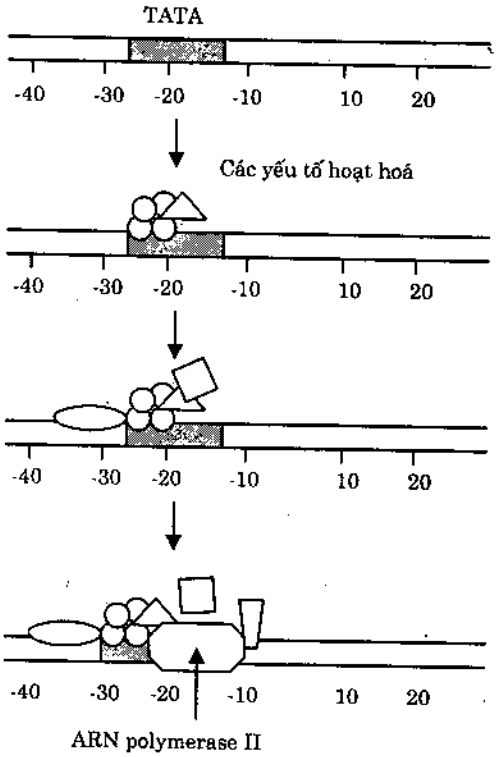
Bảng 3.3: Các yếu tố sigma của *E.coli* tham gia ARN polymerase khởi động phiên mã trong các điều kiện môi trường khác nhau

Gen	Yếu tố sigma	Điều kiện sử dụng
rpoD	σ_{70}	Bình thường
rpoE	σ_E	Sốc nhiệt
rpoH	σ_{32}	Sốc nhiệt
rpoN	σ_{54}	Thiếu nitơ
rpoS	σ_S	Stress
fliA	σ_{28} (σ_F)	Tạo roi

Promoter của gen eukaryot mã cho protein được nhận biết bởi ARN polymerase II. Tuy nhiên enzym chỉ tương tác với promoter sau khi hầu hết các yếu tố phiên mã đã có mặt ở promoter (Hình 3.1). Ngoài ra, phiên mã ở các gen eukaryot còn đòi hỏi các trình tự tăng cường (***enhancer***). Trình tự tăng cường thường đặc thù cho từng gen hoặc nhóm gen, trình tự này có thể phân bố phía trước, sau hoặc ngay trong gen. Trình tự tăng cường là vị trí tương tác với yếu tố hoạt hoá. Điều này có thể dẫn đến thay đổi cục bộ cấu trúc đoạn ADN chứa promoter, hoặc làm tăng nồng độ các yếu tố hoạt hoá tại promoter. Như vậy, vai trò của trình tự tăng cường là tăng mức độ phiên mã trên promoter.

3.3. PROTEIN THAM GIA KHỞI ĐỘNG PHIÊN MÃ

Quá trình phiên mã trên các gen eukaryot được kiểm soát chủ yếu ở giai đoạn khởi đầu. Giai đoạn này phụ thuộc vào tương tác giữa các yếu tố phiên mã với nhau, hoặc với các trình tự nucleotide đặc biệt. Kết quả của những tương tác đan chéo dẫn đến tăng cường hoặc ức chế ARN polymerase tổng hợp ARN. Các yếu tố có chức năng tăng cường hoạt động phiên mã được gọi là yếu tố hoạt hoá (*activator*). Các yếu tố có chức năng ức chế hoạt động phiên mã được gọi là yếu tố kìm hãm (*repressor*). Các yếu tố hoạt hoá và kìm hãm đều có thể liên kết với các protein khác để điều chỉnh hoạt tính của mình. Protein liên kết với yếu tố hoạt hoá giúp yếu tố hoạt hoá thực thi nhiệm vụ được gọi là yếu tố đồng hoạt hoá (*co-activator*). Tương tự, protein liên kết với yếu tố kìm hãm để ức chế phiên mã thì được gọi là yếu tố đồng kìm hãm (*co-repressor*). Hầu hết các yếu tố phiên mã có nhiều domain. Mỗi domain có vai trò đặc hiệu đảm bảo tương tác với ARN polymerase, với ADN, hoặc với các protein khác.



Hình 3.1: Các yếu tố hoạt hoá tạo phức tại promoter để ARN polymerase II bắt đầu phiê mã trên gen mã cho protein trong tế bào eukaryot.

Yếu tố hoạt hoá (activator) có khả năng nhận biết và tương tá với trình tự nucleotide đặc biệt như promoter (nằm trong vùng điều khiển phía trước vùng mang mã di truyền của gen), hoặc tương tá với trình tự tăng cường (enhancer) nằm cách xa gen (phía trước, phía sau hoặc ngay trong gen). Sự có mặt của yếu tố hoạt hoá giúp ARN polymerase II dễ dàng tương tác với promoter và bắt đầu phản ứng tổng hợp ARNm. Như vậy, chức năng của activator là tăng cường tốc độ khởi động phiên mã tại promoter. Một số activator hoạt động liên tục. Tuy nhiên rất nhiều activator chỉ hoạt động tại những thời điểm và chỉ có mặt trong một số mô tổ chức nhất định. Đặc biệt những activator kích hoạt các gen mà sản phẩm của chúng cần thiết trong giai đoạn sinh trưởng phát triển của cơ thể. Hầu hết các activator hoạt động theo thời gian và không gian sẽ nhận biết và tương tác vớ những trình tự đặc hiệu của riêng từng nhóm gen hoặc từng gen riêng biệt.

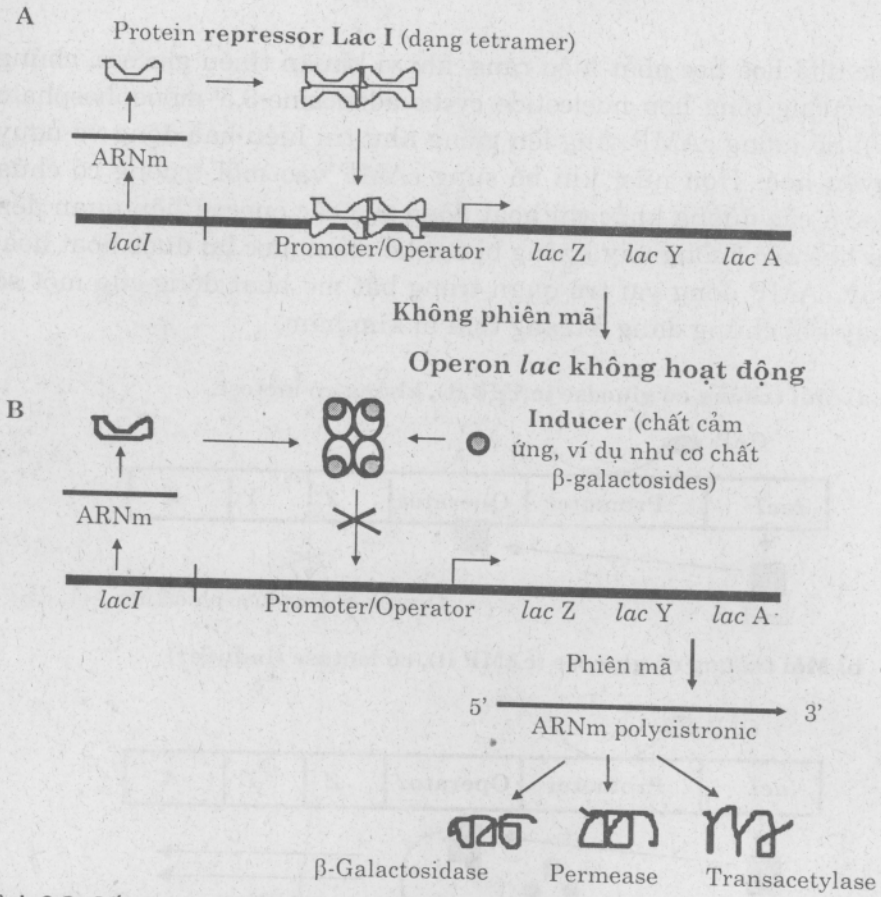
Khác với yếu tố hoạt hoá, yếu tố kìm hãm (repressor) luôn luôn tương tác với trình tự đặc hiệu trên phân tử ADN. Tại vị trí đó, repressor có thể tương tác với ARN polymerase để ngăn cản enzym khởi động phiên mã. Ví dụ, yếu tố kìm hãm tương tác với vị trí ***operator*** nằm gối lên promoter của operon *lac* ở *E.coli*. Yếu tố kìm hãm tách khỏi ADN khi có mặt của yếu tố cảm ứng hay còn gọi là yếu tố kích thích (*inducer*). Một promoter có thể có nhiều vị trí cho yếu tố hoạt hoá cũng như nhiều vị trí cho yếu tố kìm hãm. Tuỳ thuộc vào vị trí trên ADN, hoặc sự có mặt của các inducer mà yếu tố kìm hãm ức chế phiên mã ở các mức độ khác nhau.

3.4. KIỂM SOÁT KHỞI ĐỘNG PHIÊN MÃ

3.4.1. Kiểm soát tiêu cực – vai trò của yếu tố kìm hãm

Operon *lac* ở *E.coli* là operon đầu tiên được phát hiện và nghiên cứu khá chi tiết từ năm 1961 bởi Jacob và Monod. Hoạt động của operon này được kiểm soát theo ***cơ chế tiêu cực***, tức là phiên mã trên operon *lac* bị ức chế bởi yếu tố kìm hãm: protein LacI. Protein này là sản phẩm của gen *lac*I nằm ngay phía trước operon *lac*. Protein LacI chỉ có hoạt tính khi tồn tại ở dạng tetramer. Cấu trúc và cơ chế kiểm soát operon *lac* được trình bày trên Hình 3.2.

Khi tồn tại ở dạng tetramere, protein LacI có 2 vị trí tương tác. Vị trí thứ nhất tương tác với đoạn nucleotide thuộc vùng promoter của operon *lac*. Đoạn này chính là ***operator*** (O). Vị trí thứ hai có thể tương tác với cơ chất β-galactosides, hoặc các chất có cấu trúc hoá học tương tự cơ chất, hoặc một số phân tử đặc biệt như isopropyl-thiogalactoside (IPTG). Thực nghiệm nhận thấy hoạt động của operon *lac* tăng rất mạnh khi có mặt của những chất có khả năng tương tác với vị trí thứ hai trên protein LacI. Các chất có đặc tính kích thích hoạt động của các gen được gọi chung là chất cảm ứng (***inducer***). Khi không có chất cảm ứng trong môi trường, quá trình phiên mã trên operon *lac* bị kìm hãm do protein LacI bám vào operator. Tuy nhiên khi có mặt chất cảm ứng, protein LacI tương tác với chất cảm ứng nên bị thay đổi cấu trúc không gian. Do đó, một khi đã tương tác với chất cảm ứng thì protein LacI không còn khả năng bám vào vị trí ức chế nữa. Lúc này khởi động phiên mã trên operon *lac* không còn bị kìm hãm nên số lượng phân tử ARNm tăng nhanh.

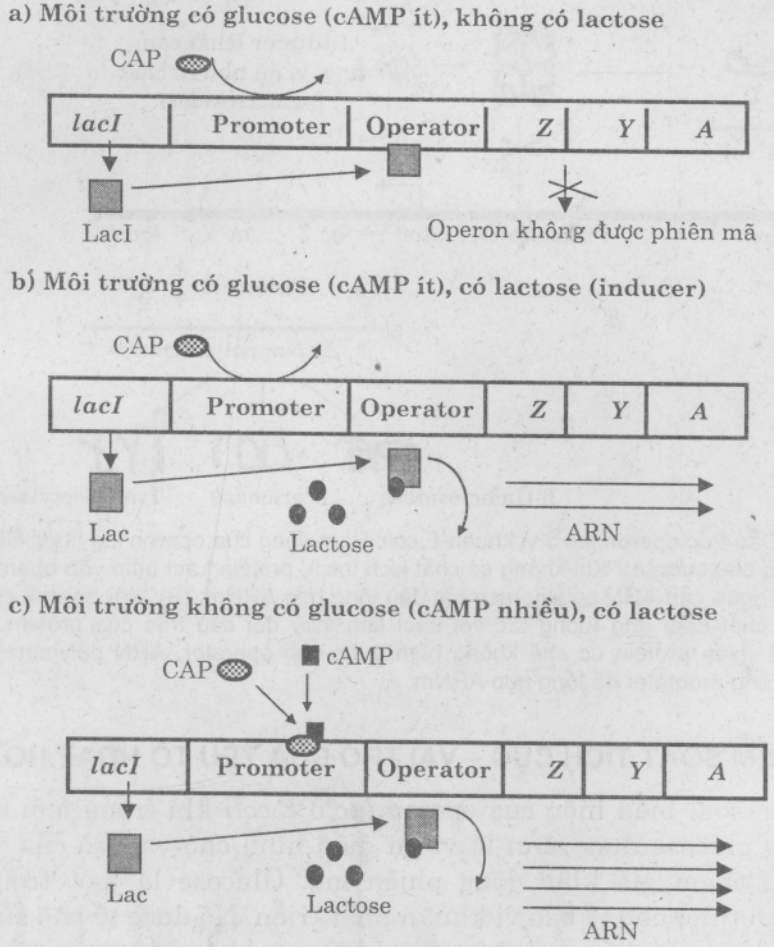


Hình 3.2: Cấu trúc operon lac ở vi khuẩn *E.coli*. Hoạt động của operon lac được kiểm soát theo cơ chế tiêu cực. (A): Khi không có chất kích thích, protein LacI bám vào operator của operon lac ngăn cản ARN polymerase bắt đầu tổng hợp ARNm. (B): Khi có chất cảm ứng xuất hiện, chất cảm ứng tương tác với LacI làm thay đổi cấu trúc của protein ức chế (repressor) khiến protein ức chế không bám được vào operator. ARN polymerase bám được vào vùng promoter để tổng hợp ARNm.

3.4.2. KIỂM SOÁT TÍCH CỰC – VAI TRÒ CỦA YẾU TỐ HOẠT HOÁ

Kiểm soát biểu hiện của operon *lac* ở *E.coli* khi trong môi trường không có glucose được xem là ví dụ điển hình cho vai trò của yếu tố hoạt hoá tham gia khởi động phiên mã. 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 mà 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 (*catabolite reppression*).

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.

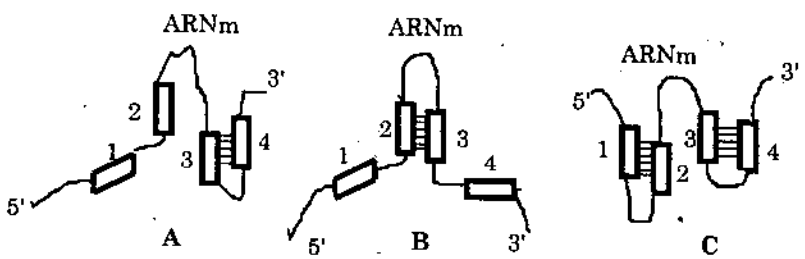


Hình 3.3: Kiểm tra hoạt động của operon lac bởi hai protein LacI và CAP.
(a) Nếu môi trường không có lactose, ARNm-lac không được tổng hợp ngay khi có hoặc không có glucose vì LacI bám vào operator. (b) Khi có lactose, LacI không bám được vào operator nhưng do có glucose nên cAMP không được tổng hợp 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 (theo Alberts và cs., 1994).

Phân tích các dòng tế bào đột biến chỉ có khả năng sử dụng glucose cho thấy cAMP cần tương tác với một protein đồng hoạt hoá khác. 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 *invitro* 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 3.3). Như vậy operon *lac* chịu sự kiểm soát của hai cơ chế. Quá trình chuyển đổi giữa hai cơ chế này phụ thuộc vào điều kiện trong tế bào và môi trường.

3.4.3. KIỂM SOÁT KHỞI ĐẦU PHIÊN MÃ THEO CƠ CHẾ SUY GIẢM – VAI TRÒ CỦA CẤU TRÚC CẶP TÓC

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** (fèedback inhibition). Ngoài ra, operon *tryptophan* còn chịu sự kiểm soát ngay khi phiên mã đã bắt đầu theo cơ chế suy giảm trình bày trên Hình 3.4.



Hình 3.4: Cấu trúc bậc hai của phân tử ARNm phiên mã từ operon tryptophan. Cấu trúc này giữ vai trò kiểm soát phiên mã suy giảm tạo ra phân tử ARNm không hoàn chỉnh (do phiên mã bị kết thúc sớm). Tuy nhiên kiểm soát phiên mã suy giảm phụ thuộc vào sự có mặt của acid amin tryptophan trong môi trường. (A)- Khi môi trường thừa tryptophan. (B)- Môi trường thiếu tryptophan. (C)- Không có sự tổng hợp enzym.

Phân tích trình tự nucleotide trên phân tử ARNm polycistronic của operon *tryptophan* 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 đoạn gây suy giảm "attenuator" có chứa 4 chuỗi nucleotide ngắn có trình tự tương đồng nhưng ngược chiều. Do đó 4 chuỗi này có thể liên kết tạo cặp với nhau, tạo cấu trúc "***hairpin***" (tạm dịch là ***cấu trúc cặp tóc***). Tuy nhiên, do mức độ tương đồng khác nhau nên khi xảy ra liên kết tạo cặp giữa chúng thì tính bền của từng cặp khác nhau. Tuỳ thuộc vào sự có mặt của tryptophan trong môi trường mà xảy ra các kiểu liên kết tạo cặp giữa hai chuỗi tạo ra các cấu trúc cặp tóc 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.

Chuỗi nucleotide **1** trên đoạn suy giảm "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 cặp tóc 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 cặp tóc 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 nên tạo liên kết với chuỗi **3**. Tuy nhiên cấu trúc cặp tóc này còn cách xa vị trí bám 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.

3.5. KIỂM SOÁT PHIÊN MÃ TRÊN GEN EUKARYOT

Kiểm soát khởi đầu phiên mã trên gen eukaryot không đơn thuần phụ thuộc vào tương tác giữa các yếu tố phiên mã với promoter hay với enhancer. Phiên mã còn liên quan chặt chẽ đến sự thay đổi cục bộ và tạm thời cấu trúc không gian của sợi nhiễm sắc. Nhờ đó các yếu tố phiên mã có thể tiếp cận với promoter hoặc enhancer. Chúng ta hãy xem xét cơ chế thay đổi ở mức độ phân tử đối với ADN và lõi histone, hai thành phần chính của sợi nhiễm sắc.

3.5.1. Phản ứng methyl ở promoter

Kiểm soát phiên mã thông qua methyl hoá ADN có thể quan sát bằng nhiều cách. Sử dụng các dạng đồng phân của enzym giới hạn (các endonuclease cắt ADN tại những trình tự nucleotide nhất định – xem chương 5) để cắt promoter của những gen đang phiên mã (hoạt động) hoặc gen không phiên mã (bất hoạt). Kết quả cho thấy ADN bị phân cắt phụ thuộc vào nhóm methyl ở cytosine có mặt trong vị trí cắt. Promoter của gen hoạt động không có nhóm methyl nên bị cắt dễ dàng trong khi promoter của gen không phiên mã bị methyl hoá nên không bị cắt.

Bên cạnh việc sử dụng enzym giới hạn, có thể dùng hoá chất 5-azacytidine. Tế bào có thể dùng chất này thay thế cytidine trong quá trình tái bản ADN. Tuy nhiên, sợi ADN mới không thể bị methyl hoá được nữa. Bằng cách này có thể làm cho các gen không bị methyl hoá và trở nên hoạt động. Thực nghiệm cho thấy khi bổ sung 5-azacytidine trong môi trường, các gen trên nhiễm sắc thể X bất hoạt chuyển sang trạng thái hoạt động. Khi loại bỏ tất cả các gốc methyl ở cytosine trong vùng promoter của gen γ-globin thì gen này mới hoạt động. Mặt khác, promoter của những gen hoạt động liên tục (gen quản gia – housekeeping gene) đều không chứa nhóm methyl. Như vậy, methyl hoá ADN là một cách thức kiểm soát tiêu cực phiên mã của các gen eukaryot. Để chuyển từ trạng thái bất hoạt sang hoạt động, nhóm methyl bị khử cytosine. Tuy nhiên, enzym này chưa được tìm thấy trong tế bào động vật. Kết quả nghiên cứu gần đây (2002-2005) cho thấy một số enzym tham gia sửa chữa ADN có liên quan đến việc loại bỏ cytosine mang nhóm methyl. Lúc đó, đoạn ADN chứa ^{m}C bị loại đi và thay thế bởi cytosine không mang nhóm methyl.

3.5.2. Phản ứng acetyl hoá histone

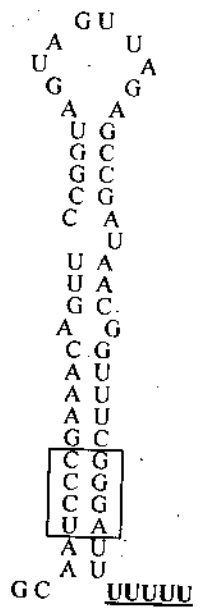
Thông thường acetyl hoá và khử acetyl ở histone liên quan đến hoạt hóa hay kìm hãm hoạt động của gen. Mỗi loại histone có thể được gắn nhóm acetyl ở những vị trí đặc hiệu bởi các enzym riêng biệt. Điều đó gây ra những tác động khác nhau đến biểu hiện của gen. Ngoài ra, quá trình acetyl hoá còn làm thay đổi cấu trúc của phức điều biến chromatin (remodeling chromatin complexes). Phức này có chức năng phá vỡ tạm thời cấu trúc lõi histone hay dịch chuyển nucleosome trên sợi nhiễm sắc. Chúng thường tương tác với vùng N-terminal của histone. Khi vùng này có mang nhóm acetyl, phức điều biến chromatin có thể làm cho các histone H_2A-H_2B bị di chuyển ra khỏi cấu trúc lõi nucleosome. Nhờ đó, các promoter được bộc lộ, cho phép quá trình tổng hợp ARNm được bắt đầu.

Động học của phản ứng acetyl hoá và khử acetyl rất linh động, phức tạp, phụ thuộc vào hoạt tính của các enzym liên quan. Biến đổi thuận nghịch giữa hai dạng acetyl hoá và khử acetyl của histone phụ thuộc vào hai loại enzym histone acetyl transferase (HAT) và histone deacetylase (HDAC). Hai enzym này đòi hỏi các protein đồng hoạt hóa (coactivator) với HAT hoặc đồng ức chế (corepressor) với HDAC. Rõ ràng, hai quá trình acetyl hoá và khử acetyl có tác dụng ngược nhau trong việc làm thay đổi cấu trúc sợi nhiễm sắc và hoạt động của các gen. Các enzym deacetylase HDAC làm giảm mức độ acetyl hoá histone, dẫn đến kìm hãm quá trình phiên mã. Ngược lại, enzym acetyl transferase HAT tăng cường mức độ acetyl, kích thích quá trình phiên mã. Mặt khác, cạnh tranh giữa hai phản ứng acetyl hoá và khử acetyl giúp sợi nhiễm sắc thay đổi cấu trúc linh hoạt, đáp ứng kịp thời với tăng cường hoặc kìm hãm hoạt động của gen.

3.6. KIỂM SOÁT KẾT THÚC PHIÊN MÃ Ở VI KHUẨN

Quá trình tổng hợp ARNm được dừng lại khi ARN polymerase gặp tín hiệu kết thúc (***terminator***). Tại vị trí có tín hiệu kết thúc, liên kết giữa 3 yếu tố (sợi ADN khuôn, sợi ARNm và ARN polymerase) kém bền vững. Ở *E.coli*, tín hiệu này liên quan đến cấu trúc đặc biệt của phân tử ARN do trình tự nucleotide nằm phần cuối của gen quy định.

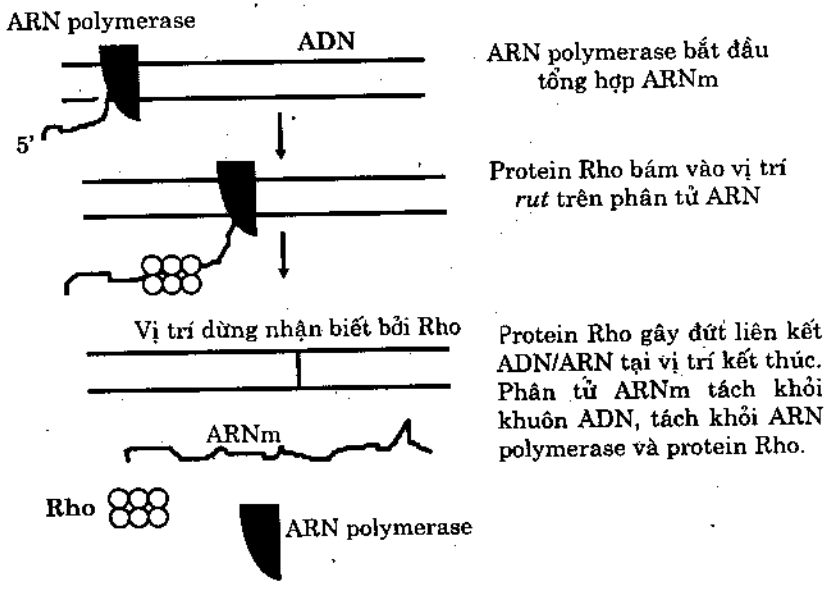
Phiên mã ở khoảng 50% các gen vi khuẩn thường kết thúc ở vị trí sợi khuôn ADN có đoạn nucleotide lặp lại ngược chiều giàu G và C. Các nucleotide lặp lại ngược chiều có thể liên kết với nhau theo nguyên tắc bổ sung tạo cấu trúc cặp tóc. Cấu trúc này làm ARN polymerase dừng hoặc chuyển động chậm (thời gian dừng khoảng 60 giây). Tuy nhiên để làm enzym tách ra khỏi khuôn ADN, tiếp theo cấu trúc cặp tóc là các nucleotide A 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 3.5).



Hình 3.5: Cấu trúc cặp tóc "hairpin" và chuỗi polyU tại vị trí dừng (terminator) đối với phiên mã ở vi khuẩn.

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 (theo Lewin, 2004).

Bên cạnh cấu trúc cặp tóc, phiên mã của một số phân tử ARNm được kết thúc với sự tham gia của protein Rho có hoạt tính helicase. Protein này giúp ARN polymerase kết thúc tổng hợp ARN tại vị trí chính xác. Do đó các phân tử ARNm có kích thước như nhau. Trong trường hợp này, vị trí kết thúc cũng thường chứa cấu trúc cặp tóc nhưng sau đó không có chuỗi nucleotide A. Cơ chế kết thúc phiên mã bằng protein Rho đến nay còn chưa được xác định một cách rõ ràng. Tuy nhiên, rất có thể hoạt tính helicase của protein Rho cắt đứt liên kết giữa ADN/ARN khiến ARN polymerase dừng tổng hợp ARN (Hình 3.6).



Hình 3.6: Cơ chế hoạt động của protein Rho.

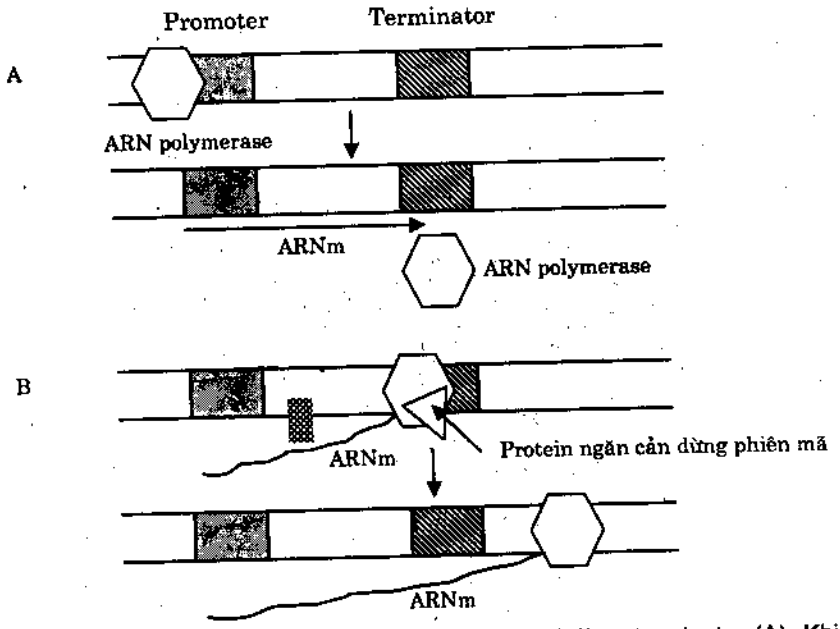
Protein Rho nhận biết vị trí đặc hiệu (rut site) trên ADN và làm ngừng phản ứng tổng hợp ARNm. Nhờ đó, các phân tử ARNm đều có kích thước chính xác như nhau.

3.7. TÍN HIỆU NGĂN CẢN DỪNG PHIÊN MÃ Ở PROKARYOT

Ngăn cản dừng phiên mã ngay khi ARN polymerase đã gặp tín hiệu kết thúc là một cơ chế kiểm soát phiên mã các gen ở thực khuẩn thể phage và operon của vi khuẩn. Tín hiệu kết thúc phiên mã có thể bị bỏ qua một khi có các tín hiệu ngăn cản dừng phiên mã (anti-termination). Lúc đó, ARN polymerase tiếp tục tổng hợp kéo dài sợi ARNm cho đến khi gặp tín hiệu kết thúc khác. Đây thực sự là một hình thức kiểm soát để bật mở (có phiên mã) hoặc đóng (không phiên

mã nữa) đối với các gen nằm nối tiếp nhau như ở genome thực khuẩn thể hoặc các gen nằm phía cuối operon ở vi khuẩn.

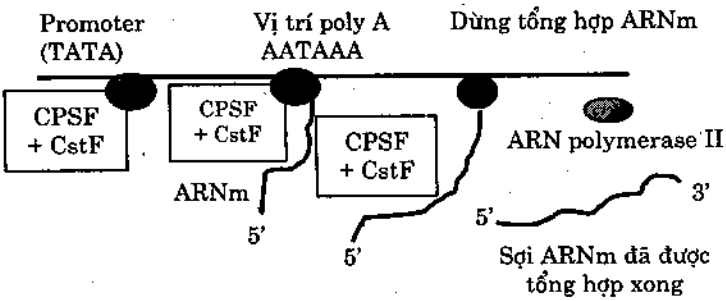
Cơ chế ngăn cản dừng phiên mã được phát hiện lần đầu ở thực khuẩn thể. Ngay sau khi nhiễm vào tế bào *E.coli*, hai gen sớm nhất (*immediate early genes*) của phage được phiên mã từ hai promoter P_L và P_R (nằm sát cạnh nhau) bởi ARN polymerase của *E.coli* và được dịch mã ngay lập tức. Sợi ARNm phiên mã từ P_R được dịch mã thành protein Cro cần thiết để thực khuẩn thể xác định hình thức xâm nhiễm sinh tan hay tiềm tan. Sợi ARNm phiên mã từ P_L được dịch mã thành protein N. Đây là protein chống kết thúc phiên mã. Do đó ARN polymerase không dừng ở vị trí kết thúc mặc dù hai gen *N* và *cro* có vị trí kết thúc được nhận biết bởi protein Rho (của tế bào chủ). Nhờ đó, ARN polymerase vi khuẩn tiếp tục phiên mã các gen kế tiếp trên genome thực khuẩn thể. Vị trí nhận biết bởi protein chống dừng phiên mã thường nằm trước vị trí kết thúc. Khi tương tác với vị trí của mình (vị trí *nut*), protein N sẽ tương tác với ARN polymerase, giúp enzym vượt qua tín hiệu kết thúc hoặc ngăn cản hoạt tính của prótein Rho (Hình 3.7).



Hình 3.7: ARN polymerase kết thúc phiên mã tại vị trí dừng terminator (A). Khi có mặt protein ngăn cản dừng phiên mã, ARN polymerase vượt qua vị trí kết thúc tiếp tục phiên mã tổng hợp phân tử ARNm dài hơn (B). Trên genome thực khuẩn thể, protein chống dừng phiên mã nhận biết vị trí *nut* nằm phía trước terminator.

3.8. KIỂM SOÁT KẾT THÚC PHIÊN MÃ Ở EUKARYOT

Tế bào eukaryot sử dụng ít nhất 3 loại ARN polymerase để phiên mã. Mặc dù vị trí dừng chính xác cho mỗi loại ARN polymerase chưa được xác định rõ ràng như đối với *E.coli* nhưng cơ chế dừng có nhiều điểm tương đồng giữa vi khuẩn và tế bào eukaryot. Quá trình tổng hợp ARNr bởi ARN polymerase I được kết thúc bởi protein có chức năng tương tự Rho. Protein Rho của vi khuẩn tương tác với ARN polymerase nhưng yếu tố dừng tổng hợp ARNr ở eukaryot tương tác với trình tự nucleotide nằm ngay trong gen. Phản ứng tổng hợp ARNt bởi ARN polymerase III sẽ kết thúc khi enzym gặp chuỗi poly U nhưng không đòi hỏi cấu trúc cặp tóc.



Hình 3.8: Phối hợp giữa vị trí polyA (AATAAA) và phức protein trong quá trình tổng hợp ARNm. Phức CPSF và CstF liên kết với ARN polymerase II. Đến vị trí polyA, factor bị phân ly khỏi enzym khiến enzym không thể tiếp tục tổng hợp ARNm. Đồng thời Poly(A) transferase thêm đuôi polyA vào đầu 3' của ARNm.

Cho đến nay vị trí dừng chính xác quá trình tổng hợp ARNm trong 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). Trong thí nghiệm *invitro*, việc loại bỏ vị trí polyA hoặc gây các đột biến tại đó đều khiến cho phân tử ARNm được tổng hợp rất dài do enzym không xác định được vị trí dừng. Cơ chế kiểm soát dừng tổng hợp ARNm khi ARN polymerase II đi qua vị trí polyA vẫn chưa sáng tỏ. Những kết quả thực nghiệm bước đầu cho thấy đây là vị trí liên kết của phức gồm protein CPSF (Cleavage and Polyadenylation Specificity Factor) và protein CstF (Cleavage Stimulation Factor). Phần COOH của tiểu đơn vị lớn nhất của ARN polymerase II cần thiết cho việc thêm đuôi polyA. Do đó, việc kết thúc phiên mã và thêm đuôi có thể xảy ra đồng thời trong nhân tế bào eukaryot (Hình 3.8).

3.9. BIẾN ĐỔI PHÂN TỬ ARNm TRONG TẾ BÀO EUKARYOT

Sau khi được phiên mã, phân tử tiền thân ARNm eukaryot phải trải qua ba thay đổi cơ bản trước khi được dùng làm khuôn để tổng hợp protein. Trước hết ARNm được gắn đuôi polyA vào đầu 3', gắn "mũ" ^{m7}G vào đầu 5' (ở vị trí 2' của đường ribose) và cắt nối intron-exon thành phân tử ARNm hoàn chỉnh. Phản ứng methyl hoá tạo cấu trúc mũ giúp cho ribosome nhận biết được phân tử ARNm và bám vào nó. Cấu trúc mũ còn giúp bảo vệ đầu 5' của ARNm không bị phân cắt bởi exonuclease.

3.9.1. Polyadenyl hoá đầu 3' của ARNm sơ cấp (Pre-mRNA)

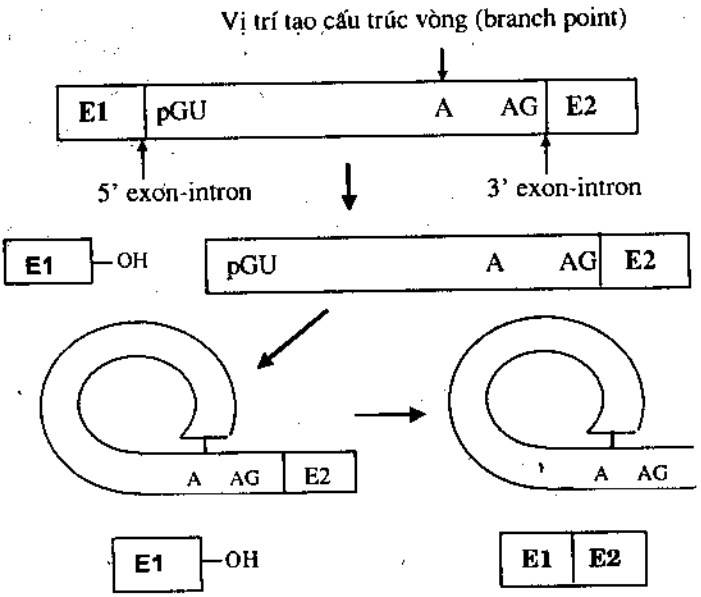
Trong tế bào động vật, tất cả phân tử ARNm, trừ ARNm của histone, đều bị polyadenyl hoá ở đầu 3', tức là mọi ARNm đều được gắn thêm đuôi polyA sau khi đầu 3' bị cắt bởi endonuclease. Phản ứng gắn polyA đòi hỏi trình tự AAUAAA nằm phía trước vị trí gắn đuôi từ 10-35 nucleotide. Các gen mang đột biến ở trình tự này phiên mã tạo ARNm không đuôi và bị phân huỷ rất nhanh. Trình tự AAUAAA được gọi là vị trí polyA (polyA site). Đầu tiên, khoảng 12 nucleotide A được gắn vào từ từ. Tiếp đến, khoảng 200-250 nucleotide A được thêm vào rất nhanh tạo thành đuôi polyA.

Sau khi có đuôi polyA và mũ ^{m7}G, ARNm phải trải qua quá trình cắt bỏ các intron trước khi được vận chuyển ra ngoài tế bào chất. Khoảng 30-40 nucleotide nằm ở biên giới giữa intron và exon rất quan trọng để phản ứng cắt intron xảy ra chính xác. Đột biến các nucleotide nằm ở trung tâm intron không ảnh hưởng đến phản ứng này. Đặc biệt trong cấu trúc của mỗi intron, hai nucleotide đầu tiên GU ở đầu 5' và hai nucleotide AG tận cùng ở đầu 3' giữ vai trò quyết định. Chúng được bảo toàn ở mọi intron và là vị trí nhận biết để cắt intron ra khỏi phân tử tiền thân ARNm.

3.9.2. Phản ứng cắt nối exon-intron

Phân tử ARNm sơ cấp có chứa các trình tự không mang mã di truyền (các intron). Các intron được cắt đi và các exon được nối với nhau theo đúng trật tự từ đầu 5'→ 3' tạo thành phân tử ARNm hoàn chỉnh. Đầu tiên biên giới exon-intron ở đầu 5' của phân tử tiền thân ARNm bị cắt. Tiếp đến, 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. Sau cùng, biên giới intron-exon ở đầu 3' của phân tử ARNm bị cắt, intron bị loại ra và hai exon được nối với nhau (Hình 3.9).



Hình 3.9: Phản ứng loại bỏ intron ra khỏi phân tử 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' 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.

Phản ứng cắt bỏ intron đòi hỏi sự tham gia của một số phân tử snARN kích thước nhỏ (small nuclear RNA). Chúng liên kết với protein tạo thành phức ribonucleoprotein (snRNP). Mỗi phức 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 snRNP. Phân tử snARN của 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 snARN có thể liên kết với intron hoặc với các snARN khác. Những liên kết này giữ vai trò quan trọng trong phản ứng loại bỏ intron.

Thực nghiệm đã tìm ra ít nhất có 6 phân tử snARN ký hiệu từ U_1

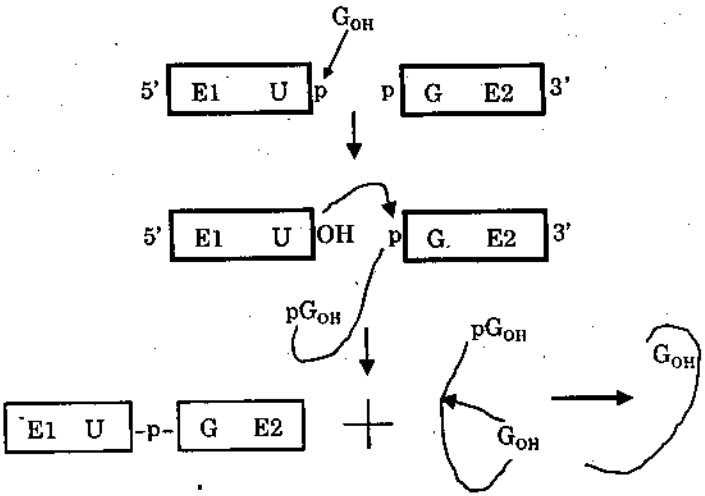
đến U_6 tương ứng với 6 snRNP. Phân tử snARN U_1 liên kết với 15-20 nucleotide tại nơi tiếp giáp 5' exon-intron, snARN U_2 liên kết với khoảng 20 nucleotide tại đoạn chứa **A** (vị trí tạo cấu trúc vòng của intron), còn snARN U5 liên kết với các nucleotide tại vùng biên 3' intron-exon. Ngoài ra các snARN U4, U6 có khả năng tương tác với nhau và tương tác với U2, U5. Mọi tương tác giữa các snRNP đều làm thay đổi cấu trúc không gian của phân tử ARNm, mang các đầu nối của exon lại gần nhau, tạo cấu trúc dạng vòng của intron, xúc tác cho phản ứng cắt bỏ intron và nối exon với nhau.

Trong thực tế, mọi snRNP và các protein tham gia phản ứng cắt bỏ intron tạo nên một cấu trúc có thể quan sát dưới kính hiển vi điển 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 của phân tử ARNm.

3.9.3. Các intron có khả năng tự cắt ra khỏi phân tử ARNm - Phản ứng self-splicing

Trên đây chúng ta đã xét phản ứng cắt bỏ intron, nối hai exon với nhau tạo phân tử ARNm hoàn chỉnh nhờ xúc tác của các snRNP và protein (trong cấu trúc spliceosome). Ngoài ra, một số phân tử tiền thân ARNm ở ty thể hoặc lục lạp có khả năng loại bỏ các intron mà không cần sự tham gia của các snRNP hoặc protein. Nói cách khác, các intron này có thể tự cắt ra khỏi các phân tử tiền thân ARNm và các exon được nối với nhau. Chúng được chia làm hai nhóm khác nhau, tuỳ thuộc vào cách thức xảy ra phản ứng.

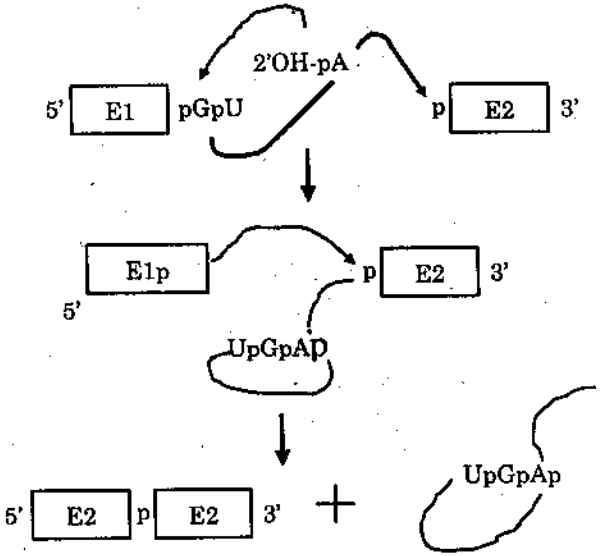
Intron nhóm I: Phản ứng cắt intron nhóm I chỉ yêu cầu phân tử tiền thân ARNm (có chứa intron) và nucleotide G có nhóm OH ở vị trí 3'. Sự có mặt của guanosine này dẫn đến việc cắt tại biên giới 5' exon-intron, đồng thời G được gắn vào đầu intron vừa được giải phóng (bằng việc chuyển liên kết phosphate từ đường này sang đường khác mà không ảnh hưởng đến liên kết phosphodiester). Đầu tự do 3'OH của exon sẽ tương tác với nơi tiếp giáp 3' intron-exon. Tiếp theo hai exon được nối với nhau và intron tồn tại ở dạng cấu trúc vòng. Cấu trúc này sau đó sẽ chuyển sang dạng thẳng (Hình 3.10).



Hình 3.10: Phản ứng self-splicing của các intron nhóm I.

Phản ứng này chỉ đòi hỏi sự tham gia của G_{OH}. Trong quá trình tự cắt intron, phosphodiester được chuyển từ phân tử đường này sang phân tử đường khác (phản ứng transesterification).

Intron nhóm II: Các intron thuộc nhóm này cũng có khả năng tự cắt ra khỏi phân tử tiền thân ARNm. Chúng có chứa GU ở đầu 5' và cơ chế tự cắt xảy ra tương tự phản ứng đòi hỏi xúc tác của snRNP (Hình 3.11). Tuy nhiên, trong phản ứng tự cắt này không hề có sự tham gia của bất kỳ factor nào ngoài bản thân phân tử tiền thân ARNm.



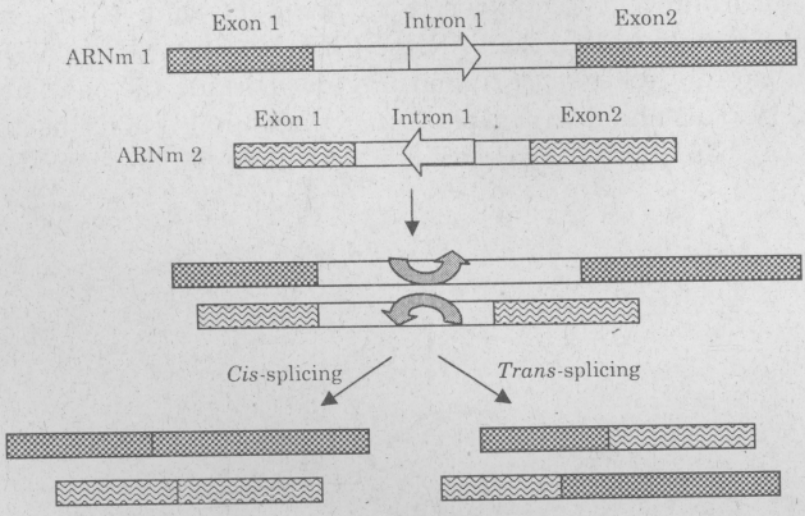
Hình 3.11: Phản ứng self-splicing của các intron nhóm II.

Nhóm 2'OH của A nằm gần 3' intron-exon tương tác với đầu 3' của exon 1. Nơi tiếp giáp 5' exon-intron bị bẻ gãy tạo cấu trúc vòng cho intron. Tiếp theo phản ứng nối hai exon với nhau, intron được giải phóng ra dưới dạng vòng.

Điều đáng lưu ý, các intron nhóm II có cấu trúc không gian rất phức tạp. Các đoạn nucleotide trong intron có thể tương tác tạo cặp với nhau. Sự thay đổi cấu trúc này làm cho phân tử ARNm có khả năng tự xúc tác cho phản ứng cắt bỏ một đoạn nucleotide của chính nó. Như vậy chính cấu trúc không gian phức tạp của intron nhóm II đóng vai trò của các snRNP trong spliceosome.

3.9.4. Phản ứng *trans*-splicing nối hai exon của hai phân tử ARNm

Phản ứng này không đòi hỏi các snARN hoặc các protein. Chính phân tử ARN cho exon thứ nhất đóng vai trò của các snRNP. Mặt khác phân tử ARN này có khả năng nhận biết vị trí 5' exon-intron ở phân tử ARN thứ hai. Như vậy các ARN tham gia phản ứng trans-splicing tự xúc tác cho phản ứng. Phản ứng này xảy ra với hầu hết ARNm của *trypanosome*, giun tròn hoặc lục lạp trong tế bào thực vật. Phản ứng trans-splicing thường xảy ra với những phân tử tiền thân ARNm có intron chứa trình tự nucleotide lặp lại ngược chiều (Hình 3.12).



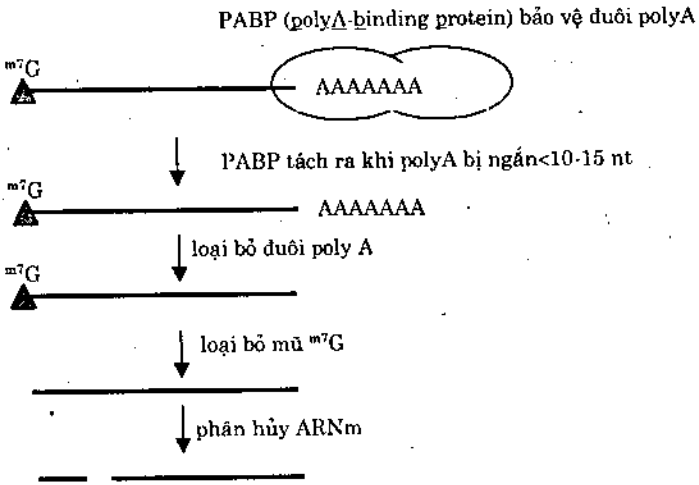
Hình 3.12: *Cis*-splicing và *Trans*-splicing có thể xảy ra với những phân tử tiền thân ARNm có trình tự nucleotide lặp lại ngược chiều phân bố trong intron.

3.10. KIỂM SOÁT SAU PHIÊN MÃ Ở TẾ BÀO EUKARYOT

3.10.1. Độ dài của đuôi polyA

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. Trong tế bào nấm men, đuôi polyA được bảo vệ bởi protein bám vào đuôi (PABP). Khi đuôi bị ngắn đi, PABP tách ra khỏi đuôi khiến ARNm bị khử polyA. Tiếp đến mũ ^{m7}G cũng bị khử bỏ và phân tử ARNm nhanh chóng bị phân huỷ từ đầu 5' (Hình 3.13).



Hình 3.13: Độ dài đuôi polyA ảnh hưởng đến độ bền vững của phân tử ARNm. Trong tế bào nấm men, đuôi polyA bị cắt ngắn đi sẽ kéo theo việc khử hoàn toàn đuôi này cùng với khử mũ ^{m7}G. Phân tử ARNm mất đầu và đuôi nhanh chóng bị phân hủy từ đầu 5'.

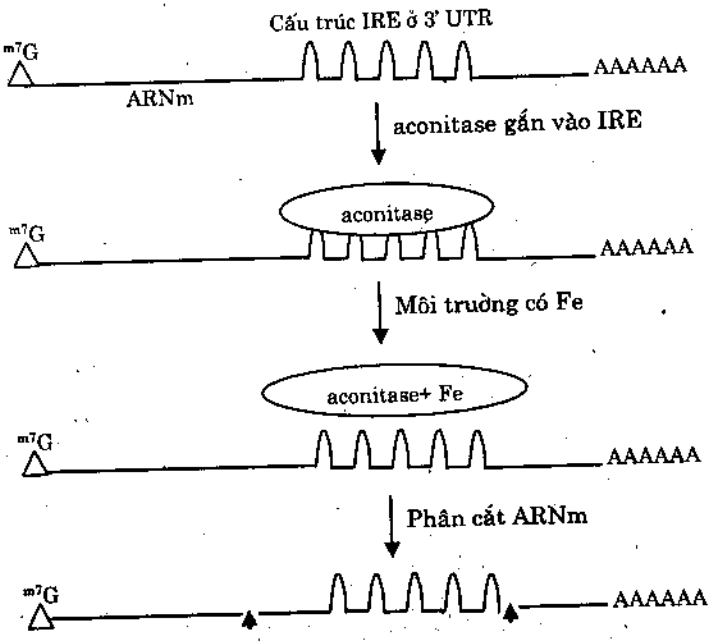
3.10.2. Độ bền vững của ARNm

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 yêu cầu một số enzym, ARNm cần thiết để 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, 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.



Hình 3.14: Cấu trúc IRE ở vùng 3'UTR trên phân tử ARNm. Protein gắn vào đó bảo vệ ARNm. Khi protein tách khỏi IRE, ARNm bị phân cắt bởi nuclease.

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.

Chúng ta biết rằng hai đầu của ARNm có các vùng 5' và 3' không dịch mã (UTR-untranslated region). Đầu 5' không dịch mã có thể là vị trí tương tác của một số protein có chức năng kìm hãm quá trình dịch mã. 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. Một cách tương tự, đầu 3' UTR của một số ARNm có cấu trúc kiểu cặp tóc. Đây cũng là vị trí tương tác với protein nhằm bảo vệ ARNm khỏi nuclease. Ở tế bào eukaryot, kiểm soát sau phiên mã theo cách thức trên được phát hiện đối 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 3.14).

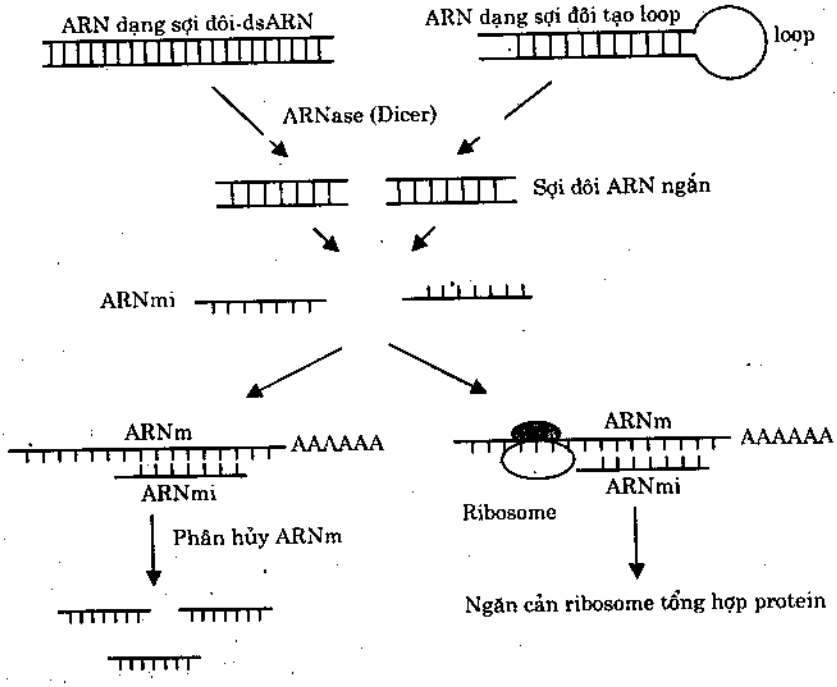
3.10.3. ARNmi (micro RNA)

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 ARN có kích thước nhỏ 21-27 nucleotide. Chúng được hình thành khi số lượng ARNm của gen đích (gen chịu sự điều khiển) tăng quá mức trong tế bào. Cơ chế hình thành các ARNmi được mô tả trên Hình 3.15. ARNmi bắt nguồn từ các sợi đôi ARN, hoặc sợi ARN chứa các trình tự lặp lại nên có thể tạo thành cấu trúc sợi đôi tương.tự cặp tóc. Những phân tử ARN dạng sợi đôi này này bị cắt ngắn bởi ARNase III (Dicer). Các sợi ngắn phân ly thành sợi đơn, được gọi là ARNmi, để tạo phức. Phức này sẽ phân cắt ARNm thông tin hoặc ức chế ribosome tổng hợp protein. Hoạt tính của phức phụ thuộc vào tương tác bổ sung giữa ARNmi và ARNm (phân tử ARN đích).

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

Phản ứng đọc sửa ARNm 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.

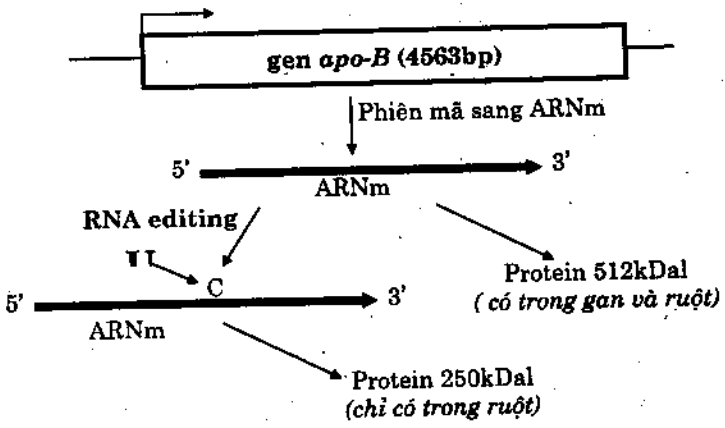


Hình 3.15: Phân tử ARNmi được hình thành từ sợi đôi ARN.

ARNmi ở dạng sợi đơn, khoảng 21-27 nucleotide, có trình tự tương đồng với ARNm. ARNmi tham gia phức protein để phân hủy ARNm hoặc ngăn cản ribasome tổng hợp prôtêin.

Các nghiên cứu tiếp theo cho thấy 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. 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, ký hiệu là ARNg (guide RNA). 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 ARNg tham gia sửa đổi một phân tử ARNm.



Hình 3.16: 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).

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 4563bp, là gen đơn bản trong genome người (Hình 3.16). 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 (stop codon) dừng phản ứng tổng hợp protein. Protein ngắn tương ứng với loại ARNm này có khối lượng phân tử 250kDal và chỉ tồn tại ở trong ruột. Phân tử protein 512kDal tương ứng với ARNm không bị đọc sửa tồn tại ở cả gan và ruột.

Chương 4

DỊCH MÃ TỔNG HỢP PROTEIN

Protein có thể là một hoặc nhiều chuỗi peptide tương tác với nhau; mỗi chuỗi gồm các acid amin liên kết với nhau (Bảng 4.1). Phản ứng tổng hợp chuỗi peptide xảy ra ở ribosome định vị trong tế bào chất. Từng mã di truyền bộ ba trên phân tử ARNm được nhận biết bởi đối mã trên phân tử ARNt nằm trong phức aminoacyl-ARNt. Tương tác giữa mã bộ ba và đối mã chỉ được thực hiện trong bộ máy ribosome với sự trợ giúp của các yếu tố cần thiết. Khi hai mã này khớp nhau chính xác, ribosome chuyển một acid amin từ ARNt sang sợi peptide. Việc gắn chính xác acid amin vào ARNt được thực hiện nhờ các enzym aminoacyl-tRNA synthetase. Trong quá trình tổng hợp protein, cả 3 loại ARN chính trong tế bào (ARNr, ARNt, ARNm) tập trung và tương tác với nhau trong bộ máy ribosome.

Bảng 4.1: Các acid amin và cách viết tắt

Acid amin	Viết tắt		Acid amin	Viết tắt	
Serine	Ser	S	Isoleucine	Ile	I
Threonine	Thr	T	Methyonine	Met	M
Cysteine	Cys	C	Phenylalanine	Phe	F
Tyrosine	Tyr	Y	Tryptophan	Trp	W
Asparagine	Asn	N	Proline	Pro	P
Glutamine	Gln	Q	Aspartic acid	Asp	D
Glycine	Gly	G	Glutamic acid	Glu	E
Alanine	Ala	A	Lysine	Lys	K
Valine	Val	V	Arginine	Arg	R
Leucine	Leu	L	Histidine	His	H

Phân tử protein có các cấu trúc khác nhau. Cấu trúc bậc một chính là trình tự acid amin tạo nên phân tử đó. Cấu trúc bậc hai liên quan đến cấu hình không gian cục bộ từng phần của protein. Ví dụ, cấu trúc xoắn α-helices hay cấu trúc lá β-sheet... Các liên kết hydro

giữ vai trò quan trọng trong việc hình thành và duy trì cấu trúc bậc 2 của protein. Các vùng khác nhau trên phân tử protein có thể sắp xếp trong cấu trúc xoắn α và cấu trúc lá β. Cấu trúc bậc ba phản ánh cấu hình không gian ba chiều của toàn bộ phân tử protein. Cấu trúc này thể hiện ở sự gấp khúc cuộn lại của phân tử protein phụ thuộc chủ yếu vào các loại liên kết yếu như liên kết cầu disulfide (giữa hai acid amin có chứa gốc sulfua). Cấu trúc bậc bốn của protein đặc trưng cho cấu hình không gian khi các chuỗi peptide tương tác với nhau tạo ra protein multimer. Tuỳ thuộc vào cấu trúc bậc 3 và bậc 4, có thể phân biệt protein thành các dạng protein hình cầu, protein hình sợi... Một phân tử protein có thể chuyển từ cấu hình không gian này sang cấu hình không gian kia tương ứng với các hoạt tính khác nhau của protein.

Một số protein có trình tự acid amin tương đồng cao tại một hay vài đoạn oligopeptide. Đoạn đó thường liên quan đến một chức năng chuyên biệt của riêng những protein chứa đoạn đó. Đoạn tương đồng cao này được gọi là vùng bảo thủ (***conserved region***). Chúng ta cần phân biệt các khái niệm về ***motif, modul và domain*** của một phân tử protein. Bất cứ vùng bảo thủ nào (ở các gen hoặc protein) đều có thể gọi là box hay motif. Khi motif gắn với một chức năng nhất định thì được gọi là modul. Một modul có thể gồm một vài motif nằm cạnh nhau trong cấu trúc bậc một. Một domain được xem là một phần của phân tử protein có xét đến cấu trúc bậc bốn nên không nhất thiết phải gồm các đoạn peptide nằm cạnh nhau (tức là có thể không phụ thuộc vào cấu trúc bậc một). Sự phân biệt rõ ràng cho phép thiết kế; điều khiển cũng như hiểu được chức năng của các protein chimeric. Một modul có thể được thêm vào hoặc lấy đi khỏi một protein bằng cách gây đột biến thêm đoạn hoặc mất đoạn ADN mang mã di truyền ứng với modul đó. Tuy nhiên, không thể áp dụng tương tự để thay đổi các domain của protein.

Tính bền vững của protein phụ thuộc vào chức năng khác nhau của chúng. Các protein điều khiển hoạt động của gen có thể chỉ tồn tại vài phút, trong khi các protein cấu trúc hoặc protein dự trữ (collagen, hemoglobin) có thể bền vững trong vài tuần. Quá trình phân hủy đổi mới protein phụ thuộc vào phản ứng của tế bào với các tín hiệu từ môi trường hoặc từ các tế bào khác trong cơ thể. Một số protein không bền thường có chứa đoạn peptide giàu proline, glutamine, serine và threonine, gọi là "PEST site". Đoạn này được nhận biết để phân cắt protein. Ngoài ra, protein còn bị phân hủy khi tương tác với ubiquytin

(protein có khối lượng phân tử nhỏ), hoặc bằng con đường tự thực bào (microautophagy).

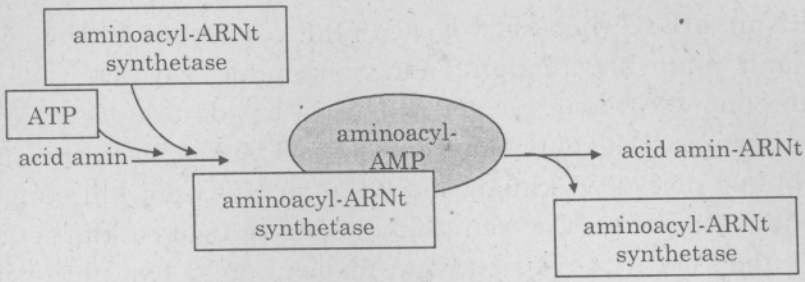
4.1. PHẢN ỨNG TỔNG HỢP PROTEIN

Protein tổng hợp trong bộ máy ribosome. Ribosome là phức của các loại ARNr với protein (Bảng 4.2). Protein tham gia cấu thành ribosome được gọi là *r-protein*. Mặc dù có sự sai khác về kích thước phân tử ARNr và số lượng protein nhưng ribosome có tính bảo thủ cao về cấu trúc trong tiến hoá từ vi khuẩn, đến động, thực vật bậc cao (bao gồm cả ribosome trong tế bào chất và trong các bào quan). Trong mọi đối tượng, ribosome gồm 2 tiểu đơn vị lớn và nhỏ. Trong tế bào *E.coli*, tiểu đơn vị nhỏ (30S) gồm phân tử ARNr 16S, và 21 r-protein; tiểu đơn vị lớn gồm ARNr-23S, ARNr 5S và 31 r-protein. Hầu hết (51) r-protein được mã bởi một gen trong genome *E.coli*. Trong tế bào eukaryot, tiểu đơn vị nhỏ gồm phân tử ARNr 18S và 33 r-protein, trong khi tiểu đơn vị lớn gồm các loại ARNr 28S, 5,8S, 5S và 49 r-protein. Tiểu đơn vị nhỏ đảm nhận các chức năng quan trọng của ribosome vì tại đây có các vị trí liên kết với ARNt (vị trí A và P), với ARNm và tiểu đơn vị lớn. Tiểu đơn vị lớn có hoạt tính enzym xúc tác phản ứng tạo liên kết giữa các acid amin (petidyl transferase). Các ribosome trong bào quan có thể thay đổi số lượng phân tử ARNr từ 30% đến 70%. Cấu trúc phức tạp giúp các vùng cục bộ của ribosome thực hiện chức năng đặc thù như trung tâm xúc tác phản ứng tổng hợp, vị trí tương tác với aminoacyl-ARNt, ARNm và chuỗi peptide đang tổng hợp.

Bảng 4.2: Các phân tử cấu thành nên ribosome ở prokaryot và eukaryot

Ribosome	Prokaryot		Eukarryot	
	ARNr	Số lượng protein	ARNr	Số lượng protein
Tiểu đơn vị nhỏ	16S	21	18S	33
Tiểu đơn vị lớn	5S, 23S	31	5S; 5,8S; 28S	49

Mỗi ARNt có hai vị trí đặc hiệu: phía đầu 5' có trình tự "anticodon" là vị trí tương tác với mã bộ ba trên phân tử ARNm, còn đầu 3' là vị trí liên kết với nhóm carboxyl (COOH) của acid amin tương ứng với mã bộ ba đó. Liên kết này được thực hiện qua 2 bước nhờ xúc tác của enzym aminoacyl-ARNt synthetase (Hình 4.1). Bước thứ nhất là acid amin được hoạt hoá, bước tiếp theo là gắn acid amin vào ARNt đặc hiệu cho acid amin đó.



Hình 4.1: Quá trình gắn acid amin vào ARNt thông qua 2 bước.

Thứ nhất: acid amin được hoạt hoá tạo phức trung gian aminoacyl~AMP. Phức này giữ tương tác với enzym để chuyển sang bước gắn acid amin vào ARNt. Cả hai bước này được xúc tác bởi enzym aminoacyl-ARNt synthetase.

Phản ứng tổng hợp protein có thể chia làm 3 giai đoạn chính: bắt đầu tổng hợp chuỗi peptide, kéo dài chuỗi và kết thúc phản ứng. Giai đoạn đầu tiên yêu cầu ribosome tương tác với ARNm và phân tử ARNt đã gắn acid amin đầu tiên. Giai đoạn tiếp theo lần lượt tạo liên kết peptide giữa các acid amin theo đúng mã di truyền trên ARNm. Giai đoạn kết thúc liên quan đến việc tách ribosome ra khỏi chuỗi peptide và ARNm.

Để bắt đầu phản ứng tổng hợp, tiểu phần ribosome nhỏ bám vào ARNm với sự hỗ trợ của các yếu tố khởi động (*initiation factors*-IFs). Ở vi khuẩn, tương tác giữa tiểu phần nhỏ 30S ribosome/ARNm phụ thuộc vào sự tạo cặp bổ sung giữa ARNr và ARNm. Trình tự nucleotide gần đầu 3' của 16S ARNr và trình tự phân bố phía trước mã bộ ba của acid amin đầu tiên (trình tự Shine-Dalgarno) trên ARNm. Đầu 5' của phân tử ARNm có thể mang nhiều mã codon của methyonine nằm cạnh nhau. Tuy nhiên chỉ có một trong số đó sẽ là mã bắt đầu chuỗi peptide. Chính một vài nucleotide nằm ngay sau mã này sẽ quyết định sự lựa chọn chính xác. Đoạn Shine-Dalgarno (UAAGGAGG) thường nằm trước mã đầu tiên ứng với Methyonine khoảng 7 đến 13 nucleotide.

Trong tế bào eukaryot, các phân tử ARNm đều có cấu trúc mũ ^{7m}G ở đầu 5'. Hầu hết ARNm virus cũng có mũ ở đầu 5' để tận dụng bộ máy tổng hợp protein của tế bào chủ. Mũ ^{7m}G được nhận biết bởi phức các yếu tố cần thiết để bắt đầu dịch mã. Phức này gồm protein eIE4E (nhận biết mũ), eIE4A (mở xoắn cấu trúc không gian ở đầu 5'

của ARNm), eIE4G (liên kết với các thành phần khác để thúc đẩy sự hợp thành phức khởi động). Thông qua phức này mà tiểu phần ribosome nhỏ 40S tương tác với ARNm và bắt đầu dò tìm mã bộ ba của methyonine. Đối với tế bào eukaryot, mã bộ ba cho methyonine là AUG, nhưng đối với vi khuẩn, mã GUG và UUG đôi khi cũng được dùng cho methyonine. Các yếu tố khởi động sẽ tách ra khi tiểu phần nhỏ tìm thấy mã AUG và tiểu phần nhỏ sau đó sẽ hợp nhất với tiểu phần lớn ribosome tạo ra bộ máy tổng hợp hoàn chỉnh. Phân tử ARNt đặc biệt (*initiator* tRNA – tạm dịch ARNt khởi động) mang methyonine (acid amin đầu tiên) là phân tử duy nhất có khả năng gắn vào vị trí P mà không đi qua vị trí A. Như vậy mọi chuỗi polypeptide đều được bắt đầu bằng methyonine. Lúc này, mã bộ ba cho acid amin thứ hai trên phân tử ARNm nằm vào vị trí A sẵn sàng liên kết với đối mã trên ARNt mang acid amin thứ hai. Chuỗi peptide chuẩn bị được kéo dài.

Mọi chuỗi peptide đều bắt đầu bằng acid amin methyonine. Để phân biệt methyonine đầu tiên với các methyonine nằm trong chuỗi peptide, tế bào *E.coli* cũng như ty thể gắn ARNt khởi động với methyonine bị formyl hoá. Phân tử này được ký hiệu fARNt-fMet (hoặc fMet-ARNt^{Met}). Phản ứng formyl hoá xảy ra khi methyonine gắn vào ARNt. Nhờ đó phức fARNt-fMet không thể tham gia vào giai đoạn kéo dài chuỗi do vị trí liên kết với acid amin tiếp theo đã gắn nhóm formyl. Tuy nhiên, việc gắn thêm nhóm formyl có ý nghĩa làm tăng hiệu quả của tổng hợp protein vì methyonine không có nhóm formyl vẫn có thể bắt đầu chuỗi peptide. Tính đặc hiệu duy nhất của fARNt-fMet gắn vào vị trí P do phân tử fARNt quyết định. Tế bào *E.coli* có hai loại ARNt nhận biết fMet và Met. Cấu trúc cục bộ của hai loại ARNt này khác nhau ở phần liên quan đến liên kết với acid amin. Khi gây đột biến các nucleotide trong vùng này, phân tử fARNt có khả năng vận chuyển methyonine trong giai đoạn kéo dài chuỗi peptide. Tính đặc hiệu của ARNt thể hiện rõ trong tế bào eukaryot, ở đó methyonine đầu tiên của chuỗi peptide không bị formyl hoá. Cấu trúc đặc biệt của ARNt khởi động được nhận biết bởi yếu tố bắt đầu dịch mã IF2 (*initiation factor 2*). Yếu tố này giúp phức fARNt-fMet vào thẳng vị trí P (bỏ qua vị trí A) ở ribosome. Nhóm formyl ở methyonine sẽ bị khử và sau đó chính methyonine bị cắt khỏi chuỗi peptide. Phản ứng cắt bỏ methyonine xảy ra khi chuỗi peptide có

khoảng 15 acid amin hoặc khi chuỗi đã đạt quá nửa chiều dài cần tổng hợp.

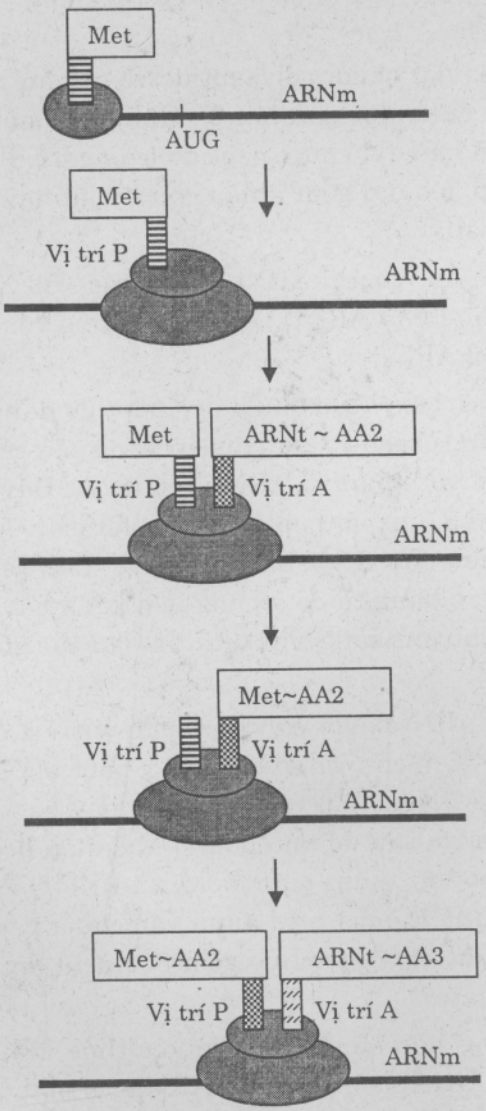
Giai đoạn kéo dài chuỗi polypeptide về cơ bản là giống nhau ở tế bào prokaryot và eukaryot (Hình 4.2). Giai đoạn này đòi hỏi sự tham gia của các yếu tố kéo dài chuỗi peptide (*elongation factors* - EFs). Có thể xem giai đoạn kéo dài gồm nhiều chu kỳ lặp lại, mỗi chu kỳ có ba bước cơ bản như sau:

Bước 1: Phức aminoacyl-ARNt tương tác với vị trí A. Ba nucleotide của anticodon (trên ARNt) tạo cặp bổ sung với ba nucleotide của codon trên phân tử ARNm.

Bước 2: Đầu carboxyl của chuỗi polypeptide đang gắn với phân tử ARNt ở vị trí P được tách ra và chuyển sang tạo cầu nối peptide với acid amin liên kết với phân tử ARNt ở vị trí A. Đây là phản ứng chủ chốt được xúc tác bởi enzym peptidyl transferase với sự tham gia của các ARNr thuộc tiểu đơn vị lớn của ribosome. Lúc này ARNt ở vị trí P được giải phóng ra ở dạng tự do, không liên kết với acid amin. Phân tử ARNt này được chuyển sang vị trí E để sau đó giải phóng ra khỏi ribosome.

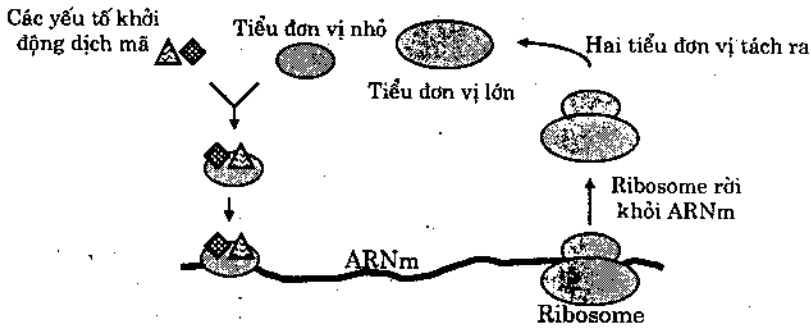
Bước 3: Phức ARNt-chuỗi polypeptide ở vị trí A được chuyển sang vị trí P. Ribosome dịch chuyển đi chính xác một mã bộ ba (một codon) dọc theo phân tử ARNm. Bước dịch chuyển này bao gồm dịch chuyển của tiểu đơn vị lớn rồi sau đó tiểu đơn vị nhỏ mới dịch theo. Ribosome dịch chuyển đòi hỏi sự tham gia của yếu tố EF-G có hoạt tính thủy phân GTP. Phản ứng thêm 1 acid amin vào chuỗi peptide chỉ hết 0,05 giây ở *E. coli*. Do đó, chuỗi peptide gồm 300 acid amin được tổng hợp trong khoảng 15 giây.

Quá trình tổng hợp chuỗi peptide kết thúc khi một trong 3 mã dừng (UAA, UAG, UGA) trên phân tử ARNm có mặt ở vị trí A. Các mã này được nhận biết bởi các yếu tố dừng tổng hợp (RFs-release factor). Khi nhận biết mã dừng, các RFs có mặt ở vị trí A sẽ thay đổi hoạt tính của peptydyl transferase khiến enzym thêm phân tử H_2O vào đầu COOH của chuỗi peptide. Chuỗi peptide tách khỏi phân tử ARNt ở vị trí P; sợi ARNm rời khỏi ribosome, hai tiểu phần ribosome tách rời nhau. Tiểu đơn vị nhỏ tương tác với ARNm nhờ sự trợ giúp của các yếu tố khởi động dịch mã. Sau khi tìm được mã đầu tiên, tiểu đơn vị nhỏ hợp nhất với tiểu đơn vị lớn hình thành bộ máy ribosome hoàn chỉnh, tiếp tục quá trình tổng hợp chuỗi peptide mới (Hình 4.3).



Hình 4.2: Quá trình tổng hợp protein.

Đầu tiên ARNt mang methyonine vào vị trí P trên tiểu đơn vị nhỏ ribosome. Sau đó, tiểu đơn vị nhỏ tương tác với và dịch chuyển trên sợi ARNm đến vị trí mã AUG đầu tiên. Tiểu đơn vị lớn ribosome sẽ tạo phức với tiểu đơn vị nhỏ tạo thành bộ máy ribosome hoàn chỉnh để tổng hợp chuỗi peptide.



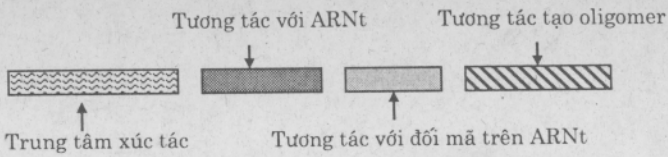
Hình 4.3: Sự quay vòng của ribosome trong quá trình tổng hợp protein.

Phức hai tiểu đơn vị lớn và nhỏ tách nhau sau khi ribosome tổng hợp xong chuỗi peptide. Tiểu đơn vị nhỏ tương tác với yếu tố khởi động dịch mã để sau đó tương tác với ARNm. Khi đã tìm thấy mã khởi đầu AUG, yếu tố khởi động dịch mã tách khỏi tiểu đơn vị nhỏ và hai tiểu đơn vị hợp nhất tiếp tục một chu trình mới.

Một sợi ARNm có thể được dịch mã đồng thời bởi nhiều ribosome. Ví dụ, trên sợi ARNm mã cho peptide fibroin (protein ở trong sữa có khối lượng phân tử 200kDa) có tới 50-80 ribosome đồng thời dịch mã. Quá trình tổng hợp protein tiêu tốn năng lượng hơn bất cứ phản ứng sinh hoá nào xảy ra trong tế bào. Ít nhất có bốn phân tử cao năng bị phân hủy để tạo ra một cầu nối peptide: hai phân tử dùng để gắn acid amin vào ARNt, phân tử thứ ba dùng trong bước thứ nhất và phân tử thứ tư dùng trong bước thứ ba. Mặt khác một phân tử ARNm có thể mang thông tin cho ba khung đọc khác nhau (tương ứng với ba chuỗi polypeptide khác nhau), cho nên việc chọn chính xác khung đọc khi bắt đầu phản ứng tổng hợp protein rất quan trọng.

4.2. TÍNH CHÍNH XÁC CỦA PHẢN ỨNG TỔNG HỢP PROTEIN

Tần số sai sót xảy ra trong phản ứng tổng hợp protein thấp, chỉ khoảng 10^{-5} –10^{-4}. Điều này cho thấy phản ứng tổng hợp protein được kiểm soát chặt chẽ từ phản ứng gắn acid amin vào ARNt đến 3 giai đoạn trong quá trình tổng hợp. Phản ứng gắn acid amin vào ARNt được xúc tác bởi enzym aminoacyl-ARNt synthetase. Cấu trúc phân tử của enzym gồm nhiều domain với chức năng khác nhau (Hình 4.4). Một enzym sẽ nhận biết đúng acid amin đặc hiệu mà nó cần vận chuyển trong số các acid amin khác cùng tồn tại trong tế bào chất. Nói cách khác, một enzym riêng biệt sẽ tương ứng với acid amin.

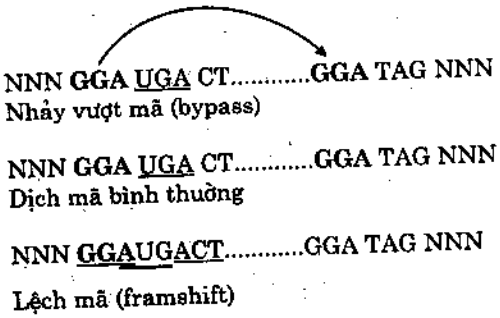


Hình 4.4: Enzym aminoacyl-ARNt synthetase có thể bao gồm 3 hoặc 4 domain có chức năng khác nhau. Riêng các enzym dạng multimer mới có domain tạo oligomer.

Sai sót có thể xảy ra khi enzym gắn nhầm acid amin vào ARNt hoặc ngược lại. Ví dụ như, cysteine có thể bị gắn nhầm vào ARNt của alanine (cysteyl-ARNt^{Ala}). Tuy nhiên những sai sót này được phát hiện và loại bỏ nhờ sự không tương thích về cấu trúc khi phức acid amin-ARNt đi vào ribosome. Đây là hoạt tính đọc sửa (***proofreading***) của enzym synthase. Một khi đã đọc sửa xong thì ARNt có nhiệm vụ mang đúng acid amin đó vào ribosome; đối mã trên ARNt tương tác với mã trên ARNm và acid amin được chuyển sang chuỗi peptide. Bằng con đường hoá học, có thể thay thế acid amin đúng bằng acid amin sai trong phức acid amin-ARNt. Ví dụ, cysteyl-ARNt^{Cys} (ARNt mang cysteine ở trạng thái hoạt hoá) có thể thay thế thành alanyl-ARNt^{Cys}. Trong phản ứng tổng hợp protein *in vitro*, alanine (acid amin sai) được chuyển từ phức alanyl-ARNt^{Cys} vào chuỗi peptide, mặc dù đối mã trên ARNt^{Cys} tạo cặp với mã bộ ba cho cysteine trên ARNm. Điều này cho thấy tính chính xác của các acid amin trong chuỗi peptide phụ thuộc vào phản ứng gắn acid amin với ARNt. Nếu sai sót không bị phát hiện thì acid amin sai vẫn nối vào chuỗi peptide ngay khi đối mã của ARNt và mã bộ ba của ARNm hoàn toàn tương xứng với nhau.

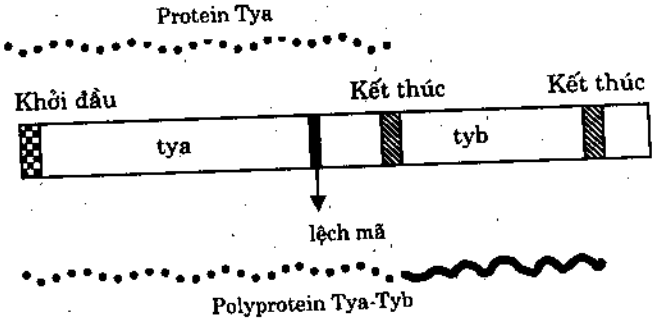
Sai sót xảy ra trong quá trình tổng hợp ở ribosome chủ yếu do ribosome trượt lệch không đúng 3 nucleotide (một mã bộ ba) hoặc bỏ qua mã bộ ba trên phân tử ARNm. Điều này gây sai sót tương tự như đột biến lệch khung đọc hoặc đột biến làm mất vài acid amin (Hình 4.5). Hiện tượng trượt phụ thuộc chủ yếu vào cấu trúc phân tử ARNm. Ngoài ra, ribosome có thể chấp nhận acid amin không đúng với mã bộ ba trên ARNm. Điều này xảy ra khi ARNt bị đột biến hoặc có thay đổi trong cấu trúc không gian. Ví dụ, một vài ARNt có khả năng nhận biết mã dừng tổng hợp như một mã bộ ba của acid amin nào đó. Thực nghiệm cho thấy, tần suất tương tác khi chỉ có 2 nucleotide của mã bộ ba và đối mã bổ sung với nhau chỉ sai khác khoảng 10^{-1} đến 10^{-2} lần so

với khi cả 3 nucleotide của mã và đối mã đều bổ sung với nhau. Như vậy, tương tác giữa mã bộ ba trên ARNm và đối mã trên ARNt không phải là điều kiện duy nhất quyết định tính chính xác. Thực tế, tương tác tạo cặp bổ sung tạo điều kiện để ARNr aminoacyl-ARNt tương tác với ARNr. Nếu không xảy ra tương tác giữa ARNr 16S và aminoacyl-ARNt, phức aminoacyl-ARNt sẽ bị đào thải ra khỏi vị trí A trên ribosome.



Hình 4.5: Một số sai sót trong quá trình dịch mã tổng hợp protein gây ra do hiện tượng ribosome di chuyển lệch mã trên sợi ARNm.

Hiện tượng ribosome trượt mã di truyền trên sợi ARNm được virus, nấm men sử dụng để kiểm soát dịch mã tạo ra các polyprotein. Khi đó, cơ chế kiểm soát được gọi là dịch mã luân phiên (*alternative transcription*). Trong tế bào nấm men, hai gen *tya* và *tyb* nằm cạnh nhau. Hiện tượng dịch mã luân phiên cho phép từ một sợi ARNm tổng hợp được các protein Tya và polyprotein Tya-Tyb (Hình 4.6).



Hình 4.6: Cơ chế dịch mã luân phiên làm dịch khung đọc để tổng hợp hai protein Tya và polyprotein Tya-Tyb từ một phân tử ARNm phiên mã từ *Ty* trong tế bào nấm men.

4.3. PROTEIN ĐƯỢC VẬN CHUYỂN VỀ ĐÍCH KHI ĐANG HOẶC ĐÃ TỔNG HỢP XONG

Trong tế bào, protein được tổng hợp để phân phối về các đích khác nhau: hoặc tiết ra ngoài tế bào, hoặc phân bố trong bào quan, trong nhân, trên màng, hoặc tồn tại tự do trong tế bào chất. Hầu hết các protein tiết ra ngoài hoặc phân bố trên màng được tổng hợp bởi ribosome bám trên mạng lưới nội chất. Con đường vận chuyển những protein này về đích được bắt đầu bằng việc đưa chúng vào trong khoang mạng lưới nội chất ngay khi đang được tổng hợp (***co-translation translocation***). Từ mạng lưới nội chất, protein chuyển sang Golgi rồi được phân phối về các đích. Các protein có đích phân bố ở trong nhân, ty thể, lục lạp được tổng hợp bởi các ribosome tự do trong tế bào chất. Con đường vận chuyển những protein này về đích chỉ bắt đầu sau khi protein đã hoàn toàn được tổng hợp xong (***post-translation transslocation***). Dù bất cứ kiểu vận chuyển nào, protein đều cần có một tín hiệu dẫn về đích của riêng mình. Đặc biệt với các protein đi qua mạng lưới nội chất, chúng đều có tín hiệu dẫn đi vào mạng lưới nội chất. Sau đó những protein này có thể cần tín hiệu dẫn đặc hiệu cho từng đích riêng, hoặc được vận chuyển thụ động đến màng tế bào.

4.4. PROTEIN ĐI VÀO MẠNG LƯỚI NỘI CHẤT

Tín hiệu dẫn (***signal sequence***) cần thiết để protein đi vào mạng lưới nội chất ngay khi ribosome đang tổng hợp chúng. Vì thế các ribosome "bám sát" trên bề mặt của lưới nội chất khiến cho bề mặt lưới nội chất trở nên "sần sùi" không trơn. Do đó những vùng lưới nội chất có ribosome bám vào được gọi là lưới nội chất hạt, còn vùng không có ribosome gọi là lưới nội chất trơn. Vùng lưới nội chất trơn còn được gọi là vùng lưới nội chất chuyển tiếp (transitional ER), vì đây là nơi hình thành các túi vận chuyển protein và lipid đi từ lưới nội chất đến Golgi. Trong một số loại tế bào như tế bào gan, vùng lưới nội chất trơn chiếm phần lớn mạng lưới nội chất và là nơi tập trung các enzym tham gia chuyển hoá, tổng hợp lipoprotein và các enzym xúc tác phản ứng khử độc các chất hoà tan trong lipid. Không có sự khác nhau giữa các ribosome tự do và ribosome bám trên mạng. Mọi ribosome đều có thể bắt đầu tổng hợp một chuỗi peptide bất kỳ ở trong tế bào chất. Chính tín hiệu vận chuyển đặc hiệu nằm ở đầu NH_2 của chuỗi peptide sẽ đưa ribosome đến bề mặt với lưới nội chất. Do nhiều ribosome cùng tham

gia tổng hợp peptide trên một sợi ARNm nên sợi đó luôn gắn với bề mặt lưới nội chất hạt.

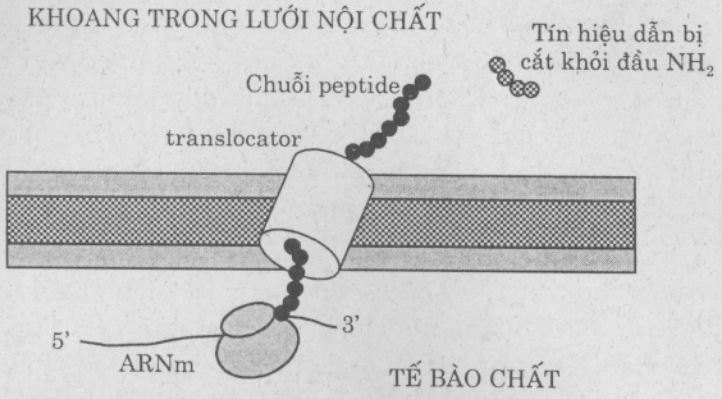
4.4.1. Tín hiệu dẫn

Tín hiệu vận chuyển gồm khoảng 12-30 acid amin trong đó có khoảng 6-12 acid amin kỵ nước. Đoạn kỵ nước giữ vai trò quan trọng đối với tương tác giữa tín hiệu bắt đầu vận chuyển và thụ cảm tượng ứng nằm trên màng lưới nội chất. Nếu gây đột biến loại bỏ vài acid amin trong đoạn kỵ nước, hoặc thay thế chúng bằng các acid amin khác nhằm gây biến đổi điện tích thì protein không được vận chuyển vào trong khoang nội chất. Không có sự tương đồng giống nhau giữa các tín hiệu vận chuyển của các protein mặc dù chức năng chính của tín hiệu là đưa protein đang tổng hợp vào trong khoang lưới nội chất. Tín hiệu sẽ bị cắt khỏi protein và bị phân hủy bởi protease có mặt ở trong khoang.

Thực nghiệm cho thấy, bất cứ protein nào được ghép với tín hiệu thì protein đó sẽ đi vào khoang nội chất. Ví dụ, globin là protein phân bố trong tế bào chất. Khi đầu NH_2 của globin được ghép với tín hiệu thì globin bị tiết ra khỏi tế bào. Thực nghiệm xác định được quá trình vận chuyển vào khoang nội chất được bắt đầu khi đoạn peptide ra khỏi ribosome có chiều dài khoảng 70 acid amin. Đầu NH_2 được đưa vào trong mạng lưới nội chất ngay khi sợi peptide chưa được tổng hợp xong. Như vậy, quá trình tổng hợp và vận chuyển với hầu hết các protein có chiều dài lớn hơn 100 acid amin sẽ xảy ra một cách đồng thời.

Liên kết đầu tiên giữa ribosome và màng nội chất đòi hỏi phức SRP (*signal recognition particle*). Phức này nhận biết một cách đặc hiệu tín hiệu vận chuyển của protein. Phức SRP gồm có 6 protein khác nhau và một phân tử 7S ARN (~300 nucleotide) có trình tự bổ sung với ARNr. Sáu protein thành phần không thể tạo phức khi thiếu phân tử 7S ARN. Khi tín hiệu vận chuyển xuất hiện và chuỗi peptide đang tổng hợp có đủ độ dài nhất định thì SRP tương tác với tín hiệu, với ribosome và với thụ thể nằm trên lưới nội chất. Tương tác đó giúp ribosome dường như bị dính sát trên bề mặt lưới nội chất. Thụ thể của SRP được gọi là "***docking protein***" gồm hai chuỗi polypeptide α (nằm trên lưới nội chất) và β (nằm xuyên qua màng lưới nội chất). Thí nghiệm tổng hợp protein *in vitro* cho thấy, khi không có microsome (cấu trúc hình thành từ lưới nội chất trong quá trình tách chiết) thì phức SRP sẽ gây ngừng quá trình tổng hợp protein khi đầu NH_2 ra

khỏi ribosome. Như vậy, SRP không chỉ làm nhiệm vụ mang vác chuỗi peptide đang được ribosome tổng hợp đến bề mặt lưới nội chất mà còn có chức năng ngăn cản quá trình tổng hợp protein hoàn chỉnh nếu như thiếu vắng lưới nội chất. Phức SRP có ba domain quy định ba chức năng khác nhau: tương tác với tín hiệu trên protein, tương tác với docking protein và tương tác với ribosome làm dừng tạm thời quá trình tổng hợp protein (Hình 4.7).



Hình 4.7: Ribosome bám vào mạng lưới nội chất nhờ thụ cảm SRP nằm ở trên màng. Phức SRP tồn tại trong tế bào chất liên kết với tín hiệu dẫn (đầu NH_2) của chuỗi peptide và với ribosome. Sau khi SRP đã tạo phức với ribosome và peptide, SRP tương tác với thụ cảm của nó trên màng lưới nội chất (theo Lewin., 2004).

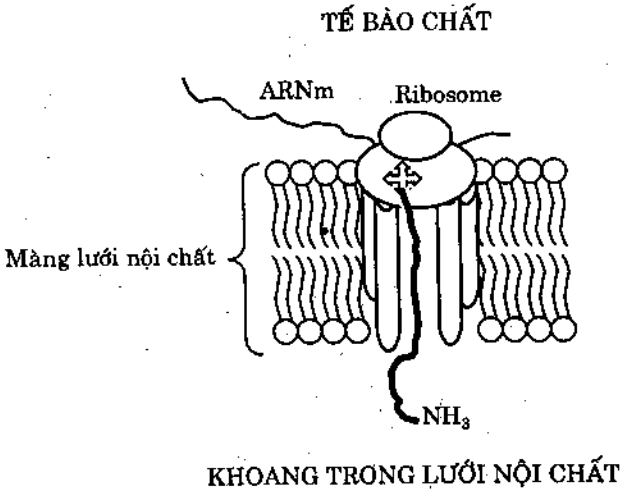
4.4.2. Translocator – kênh vận chuyển trên màng lưới nội chất

Sau khi SRP đã liên kết với thụ thể của riêng nó trên màng nội chất, ribosome tiếp tục tổng hợp protein. Phức SRP tách khỏi ribosome và chuỗi peptide được chuyển sang phức protein nằm trong cấu trúc màng nội chất. Phức này được gọi là **translocator**, có chức năng như kênh dẫn cho phép protein đi vào trong khoang nội chất.

Translocator gồm 3-4 tiểu phần khác nhau; mỗi tiểu phần gồm 3 protein nằm xuyên qua màng nội chất. Thực chất translocator là phức tạo cấu trúc kênh cho phép polypeptide đi xuyên qua màng. Bình thường khi không có chuỗi polypeptide cần đưa vào trong khoang thì translocator sẽ đóng. Translocator nhận biết tín hiệu vận chuyển trên chuỗi peptide và sẽ mở ra cho phép chuỗi polypeptide được đưa dần vào trong khoang. Trong suốt quá trình đưa chuỗi peptide vào khoang nội chất, ribosome nằm sát translocator bịt kín sao cho không có bất

cứ thành phần nào ở bên trong khoang, đặc biệt là các cation Ca^{+2}, lọt ra ngoài (Hình 4.8).

Quá trình đưa protein vào trong khoang nội chất đòi hỏi GTP. Cả SRP và thụ thể của nó đều có hoạt tính phân hủy GTPase. Khi protein chuyển từ SRP sang translocator trên màng thì SRP tách khỏi thụ thể của nó. Cả hai bước chuyển này đều cần năng lượng do SRP và thụ thể của nó phân huỷ GTP tạo ra.



Hình 4.8: Cấu trúc của translocator tạo kênh dẫn cho phép chuỗi polypeptide được vận chuyển qua màng lưới nội chất vào trong khoang khi đang được ribosome tổng hợp (theo Lodish và cs., 2000).

Không phải mọi protein vận chuyển vào lưới nội chất trong khi đang được tổng hợp. Một số protein, ví dụ như các protein có khối lượng phân tử nhỏ, được tổng hợp trong tế bào chất. Sau đó, chúng được đưa vào lưới nội chất thông qua translocator. Ở vi khuẩn, protein SeA (sec-sec<u>r</u>etation) có hoạt tính ATPase tham gia cấu trúc của translocator. Protein này phân bố ở mặt ngoài lưới nội chất, tiếp xúc với tế bào chất. Mỗi khi một phân tử ATP bị phân hủy, một phần của protein SecA sẽ đẩy sợi peptide vừa được tổng hợp vào kênh. Đối với một số protein kích thước nhỏ, chúng thường được tổng hợp trọn vẹn bởi các ribosome phân bố tự do trong tế bào chất. Khi đó, các protein chaperon tồn tại sẵn trong tế bào chất sẽ liên kết với đoạn peptide ra khỏi ribosome. Nhờ đó chuỗi peptide không có cấu trúc gấp khúc và được vận chuyển vào trong khoang lưới nội chất. Như vậy, các protein

có cấu trúc không gian không thể đi qua kênh translocator để vào bên trong màng lưới nội chất.

4.4.3. Hình thành cấu trúc không gian của các chuỗi peptide trong khoang lưới nội chất

Cấu trúc không gian được thiết lập nhờ các liên kết giữa hai acid amin trên chuỗi. Đặc biệt liên kết disunfide (S-S) giữa hai cysteine giữ vai trò quan trọng đảm bảo cấu trúc thích hợp cũng như hoạt tính của các enzym và hoocmon. Liên kết S-S được hình thành nhờ enzym disulfide isomerase (PDI) phân bố bên trong khoang mạng lưới nội chất. Enzym này xúc tác cho phản ứng oxy hoá nhóm SH tạo cầu disunfide giữa hai cystein của cùng một chuỗi peptide hoặc giữa các chuỗi khác nhau. Chuỗi nhẹ immunoglobin Ig có hai cầu S-S. Khi chuỗi này đang được tổng hợp và vận chuyển vào khoang nội chất, cầu S-S được thiết lập cho hai cystein nằm gần đầu NH_2 ngay khi cystein thứ ba chưa được gắn vào chuỗi Ig. Điều đó đảm bảo cho sự tạo cầu S-S chính xác.

Bên cạnh phản ứng tạo cầu nối S-S, chuỗi peptide được gấp cuộn theo một trật tự đặc thù riêng của peptide đó. Ví dụ, cấu trúc multimer của protein được hình thành ngay trong khoang lưới nội chất. Chính cầu S-S giúp cho việc gấp khúc được chính xác. Kết quả thí nghiệm cho thấy 3 chuỗi monomer Hemagglutinin HA (protein cấu trúc vỏ virus gây bệnh cúm) liên kết tạo thành phân tử HA dạng trimer hoàn chỉnh ở trong khoang lưới nội chất sau khi các cầu disulfide trên mỗi sợi được tạo thành. Khi một chuỗi monomer HA mang các đột biến mất, thêm hoặc thay thế acid amin, mọi dạng monomer cũng như trimer có cấu trúc không thích hợp đều bị giữ lại trong khoang nội chất chứ không được vận chuyển về đích.

Các protein vào trong khoang còn được sửa chữa cấu trúc không gian sai lệch, hoặc cấu trúc polymer không thích hợp nhờ các chaperone (BiP) và một số enzym phân bố trong khoang nội chất. Tương tự như trường hợp HA, kết quả nghiên cứu protein α1-antiprotease do tế bào gan tiết vào máu nhằm ức chế hoạt tính enzym elastase cho thấy, khi α1-antiprotease có glutamate ở vị trí 342 bị thay thế bằng lysine thì chuỗi peptide này không có cấu hình thích hợp. Do đó chúng bị giữ lại trong khoang mạng lưới nội chất mà không được tiếp tục vận chuyển để tiết ra ngoài. Lúc đó tế bào gan bị trương phồng lên bởi mạng lưới nội chất của nó chứa đầy protein α1-antiprotease ở dạng cấu trúc sai lệch. Sự tích lũy các protein có cấu trúc không đúng

sẽ hoạt hoá gen mã cho những protein làm nhiệm vụ sửa chữa sai hỏng như chaperon, peptidyl-prolyl isomerase và disunfide isomerase... Một khi tế bào gan không cung cấp đủ α1-antiprotease, elastase vẫn có hoạt tính. Enzym này sẽ phân hủy mạng lưới hấp thụ oxy ở phổi. Vì vậy bệnh nhân mắc bệnh di truyền Caucasian với các triệu chứng khó thở, tổ chức phổi bị thoái hóa bởi elastase.

4.4.4. Quá trình đường hoá protein

Sau khi đã vào trong khoang lưới nội chất, các phân tử protein bị đường hoá để dễ dàng phân loại trong quá trình vận chuyển đến từng đích. Chuỗi oligosaccharide liên kết với nhóm amin của acid amin asparagine thông qua N-acetylglucosamine (GlcNAc). Do liên kết với nhóm amin, chuỗi oligo được gọi là N-oligosaccharide. Chuỗi N-oligosaccharide gồm 14 gốc đường $(Glc)_3(Man)_9(GlcNAc)_2$, trong đó Glc-Glucose; Man-Mannose; GlcNAc-Gluco-N acetyl glucosamine. Acid amin asparagine bị đường hoá thường có vị trí asparagine-X-serine/threonine, trong đó X là acid amin bất kỳ. Ba acid amin Asn-X-Ser/Thr ở vị trí đường hoá được gọi là ***sequon***. Không phải mọi sequon đều được gắn oligosaccharide. Do đó, các acid amin bên cạnh sequon hoặc bản thân cấu trúc không gian của protein có liên quan đến quá trình đường hoá này. Ngoài ra, chuỗi oligosaccharide có thể liên kết với nhóm hydroxyl của acid amin serine hoặc threonine thông qua N-acetylgalactosamine (GalNac). Chuỗi này cũng có thể liên kết với nhóm hydroxyl của acid amin hydroxylysine thông qua galactose. Do liên kết với nhóm hydroxyl nên chuỗi này được gọi là O-oligosaccharide. Tuy nhiên, rất ít protein bị đường hoá ở nhóm hydroxyl và phản ứng này, nếu có, thường xảy ra ở Golgi.

Chuỗi N-oligosaccharide gồm 14 nhóm đường được giữ ở bên trong khoang nội chất nhờ liên kết với dolichol. Khi phân tử protein đi vào trong khoang nội chất, cả chuỗi saccharide được chuyển sang protein. Phản ứng được thực hiện nhờ enzym oligosaccharyl-transferase gắn ở phía trong màng ER. Enzym chuyển oligo-saccharide $(Glc)_3(Man)_9(GlcNAc)_2$ sang chuỗi peptide ngay khi đang được vận chuyển vào khoang. Tuy nhiên ngay sau đó, glycoprotein lại chịu biến đổi tiếp. Chúng bị cắt bớt đi ba gốc đường $(Glc)_3$ bởi 3 enzym khác nhau. Lúc đó oligosaccharide trên chuỗi polypeptide chỉ còn lại $(Man)_8(GlcNAc)_2$. Dường như ba gốc đường này có vai trò đánh dấu những chuỗi N-oligo đã được hoàn chỉnh. Sau khi loại bỏ 3 gốc đường, phân tử protein được chuyển sang *cis*-Golgi.

4.5. VẬN CHUYỂN TỪ MẠNG LƯỚI NỘI CHẤT VỀ ĐÍCH

Sau khi đã hoàn thiện cấu trúc không gian thích hợp và trở thành glycoprotein nhờ gắn oligosaccharide vào asparagine ở sequon, các protein trong khoang mạng lưới nội chất sẽ được vận chuyển đến Golgi nhờ các nang tải có nguồn gốc từ một số vùng đặc biệt trên màng lưới nội chất.

Ở trong Golgi, chuỗi oligosaccharide tiếp tục chịu những biến đổi tùy thuộc vào từng loại protein. Ngoài ra, phản ứng đường hoá có thể xảy ra với các acid amin serine, threonine. Cả hai quá trình đường hoá và khử đường xảy ra ở trong khoang lưới nội chất và Golgi của tế bào eukaryot mà không có trong tế bào prokaryot. Rất có thể sự có mặt của oligosaccharide giúp cho các protein tránh được tác dụng của các protease trong khi sự khử các đường khác nhau chuẩn bị cho protein được vận chuyển có chọn lọc về các vị trí của riêng mình.

Từ Golgi, protein được vận chuyển chọn lọc đến các đích khác nhau trong tế bào như một số bào quan, màng tế bào hoặc được tiết ra ngoài. Ví dụ, để vận chuyển các enzym về lysosome, các enzym này phải được phosphoryl hoá tại phân tử đường mannose trong chuỗi oligosaccharide (M6P). Vì vậy, các enzym được phân biệt với các protein khác và được nhận biết bởi thụ thể đặc hiệu nằm trên màng *trans* của Golgi. Thụ thể tương tác với nhóm M6P và "đóng gói" các enzym vào trong nang tải. Nang này tách ra khỏi mạng phía *trans*-Golgi. Trên đường vận chuyển các enzym về lysosome, nang có thể dung hợp với một số loại nang khác. pH trong các nang thay đổi từ trung tính (pH=7) đến khoảng 5,5. Sự thay đổi pH này khiến liên kết giữa thụ thể và enzym trở nên không bền vững. Các enzym được tự do trong nang và dung hợp vào lysosome, giải phóng các enzym vào bên trong (nơi có pH=5).

Sự thiếu vắng một vài enzym trong lysosome là nguyên nhân gây nên một số bệnh di truyền. Gen mã cho các enzym này hoàn toàn không bị đột biến. Tuy nhiên, enzym không có mặt ở lysosome là do thiếu sót trong quá trình sàng lọc nhận diện chúng để vận chuyển về lysosome. Ví dụ, nếu không được phosphoryl hoá ở M6P thì các enzym sẽ không được đưa về lysosome mặc dù chúng có mặt trong Golgi.

Đối với tế bào eukaryot, các thụ thể, các protein cấu trúc màng tế bào, các proteoglycan, glycoprotein cũng như một số protein khác làm nhiệm vụ truyền tín hiệu giữa các tế bào được vận chuyển theo con

đường tiết ngoại bào (***exocytosis***). Hầu hết các protein cấu trúc được vận chuyển liên tục (tiết liên tục), trong khi các protein hoà tan, một số hoocmon, các chất dẫn truyền thần kinh hoặc các enzym chủ yếu vận chuyển một cách chọn lọc. Các protein vận chuyển chọn lọc thường được tổng hợp và tích trữ trong các nang tiết (*secretory vesicle*). Đó là những nang đặc biệt được bao phủ lớp clathrin ở phía ngoài. Các protein chứa trong nang tiết được gọi là protein tiết. Nồng độ protein tiết trong nang rất đậm đặc (có thể tăng gấp 200 lần so với nồng độ trong Golgi). Nhờ đó tế bào có thể tiết ra một số lượng lớn ngay khi cơ thể yêu cầu. Một nang tiết chọn lọc có thể chứa nhiều loại protein tiết với nồng độ khác nhau. Các nang tiết sẽ dung hợp với màng tế bào và nhả các chất tiết ra ngoài.

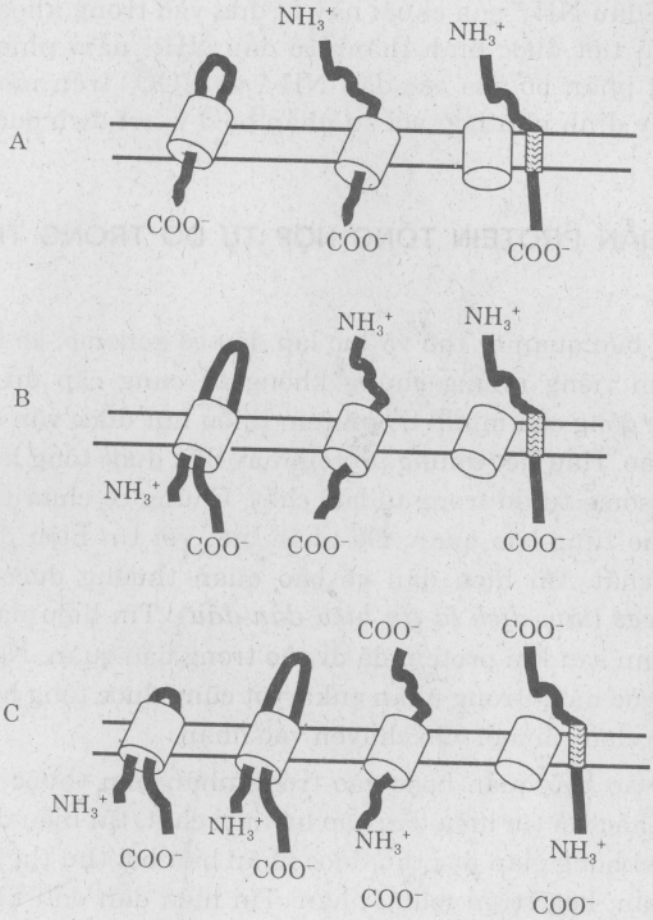
Khi vừa tổng hợp, đa số protein tiết đều ở dạng không có hoạt tính do chúng mang các đoạn peptide không cần thiết ở các đầu COOH và NH_2 (nên gọi là các ***pre-protein***), hoặc do chúng được tổng hợp dưới dạng polyprotein (gọi chung là các ***pro-protein***), sau đó được phân cắt tạo ra các phân tử có hoạt tính. Chúng chỉ có hoạt tính khi có mặt trong nang tiết hay thậm chí chỉ khi đã tiết ra khỏi tế bào. Điều này tạo điều kiện thuận lợi cho vận chuyển các protein có kích thước rất ngắn. Ở dạng polyprotein (các đoạn peptide giống hệt nhau được lặp đi lặp lại liên tục) chúng được nhận biết và đóng gói dễ dàng hơn vào các nang tiết. Mặt khác đây có lẽ cũng là biện pháp bảo đảm an toàn cho tế bào tránh tác dụng của các enzym. Chúng chỉ có hoạt tính khi được cất giữ trong nang hoặc được tiết ra khỏi tế bào.

4.6. VẬN CHUYỂN PROTEIN MÀNG (MEMBRANE PROTEINS)

Các protein tham gia hình thành cấu trúc màng chất nguyên sinh hay màng các bào quan được tổng hợp và vận chuyển đồng thời vào màng lưới nội chất. Tuy nhiên, chúng không được giải phóng vào trong khoang nội chất mà vẫn giữ lại trên màng. Chiều phân cực của những protein này ở trên màng nội chất, tức là chiều phân bố đầu NH_3^+ (hoặc COO^-) ở bên trong khoang nội chất hay bên ngoài tế bào chất, sẽ giống hệt như sự phân cực tại vị trí trên màng đích (màng đích có thể là màng nguyên sinh chất hay màng của các bào quan). Trong quá trình vận chuyển từ lưới nội chất đến vị trí đích, tính phân cực của protein màng không thay đổi. Như vậy, sự phân cực của protein thuộc cấu trúc màng được hình thành ngay khi chúng đang được đưa vào lưới nội chất.

Trường hợp đơn giản nhất là vận chuyển protein tới màng mà protein này chỉ nằm xuyên qua màng một lần. Trong trường hợp này, một số protein có phần đầu COO^- nằm trong tế bào chất còn phần đầu NH_3^+ nằm phía ngoài màng. Tuy nhiên, một số protein khác phân cực ngược lại; chúng có đầu COO^- nằm phía ngoài màng còn phần đầu NH_3^+ nằm phía trong tế bào chất. Chúng ta xem xét trường hợp vận chuyển chuỗi peptide có đầu COO^- nằm trong tế bào chất còn phần đầu NH_3^+ nằm phía ngoài màng. Chuỗi sẽ có hai tín hiệu đặc biệt. Tín hiệu thứ nhất là tín hiệu dẫn như đã mô tả phần trên. Khi chuỗi peptide đi vào khoang, tín hiệu dẫn bị cắt khỏi chuỗi peptide, giải phóng đầu NH_3^+ vào trong khoang. Chuỗi pepitde tiếp tục được đưa vào khoang cho đến khi tín hiệu thứ hai đi qua translocator thì bị giữ lại. Do đó đoạn pepitde nằm kế tiếp sau tín hiệu thứ hai (chứa đầu COO^-) không được đưa vào khoang nữa. Tín hiệu thứ hai bị giữ ở translocator cho đến khi tổng hợp chuỗi peptide kết thúc. Đầu COO^- của chuỗi nằm bên ngoài tế bào chất. Sau đó, tín hiệu thứ hai sẽ di chuyển ra khỏi translocator và nằm xen vào lớp phospholipid của màng nội chất. Lúc này chuỗi peptide được cố định tại màng nhờ tín hiệu thứ hai xuyên qua lớp phospholipid. Tín hiệu thứ hai này được gọi là tín hiệu dừng vận chuyển (***stop-transfer signal***). Tuy nhiên nó còn có chức năng gắn giữ chuỗi peptide xuyên qua màng (membrane-anchor signal) (Hình 4.9).

Đối với protein màng có đầu COO^- nằm bên ngoài màng và đầu NH_3^+ nằm phía trong tế bào chất, chúng được vận chuyển và cố định tại màng khi đang được tổng hợp. Tuy nhiên, những protein này chỉ có một tín hiệu, tín hiệu này đồng thời đảm nhiệm cả hai chức năng: chức năng dẫn và chức năng cố định trên màng nội chất. Điều cần lưu ý lúc này tín hiệu không phải là đoạn peptide nằm ngay phía đầu NH_3^+. Do đó, chuỗi peptide nằm phía trước tín hiệu được tổng hợp tự do trong tế bào chất. Khi đoạn tín hiệu được tổng hợp đi ra khỏi ribosome, nó được nhận biết bởi phức SPR đặc hiệu và chuỗi peptide được đưa đến sát màng nội chất. Tín hiệu đi vào translocator và đoạn peptide nằm phía sau tín hiệu đi vào trong khoang nội chất. Tín hiệu được giữ lại tại translocator cho đến khi tổng hợp chuỗi peptide kết thúc. Phần đầu COO^- được giải phóng vào trong khoang. Sau đó, tín hiệu sẽ di chuyển ra khỏi phức protein ở translocator và định vị trong lớp phospholipid của màng. Lúc này, chuỗi peptide nằm xuyên qua màng có đầu NH_3^+ nằm ngoài tế bào chất và đầu COO^- phân bố phía trong màng.



Hình 4.9: Vận chuyển protein màng.

Các protein thuộc cấu trúc màng được giữ lại ở màng nội chất nhờ tín hiệu dừng nằm sau tín hiệu vận chuyển (A), hoặc nhờ một tín hiệu đảm nhận đồng thời hai chức năng vận chuyển và dừng vận chuyển (B và C). Đầu NH_3^+ bị đưa vào trong khoang nội chất (B) hay phân bố phía ngoài tế bào chất (C) phụ thuộc vào sự phân cực của các acid amin nằm phía trước và sau tín hiệu.

Protein phân cực trên màng nội chất sẽ được chuyển đến màng tế bào nhờ các túi tiết. Túi tiết hình thành từ vùng màng nội chất chứa protein cố định nằm xuyên qua. Như vậy, nếu protein có đầu NH_3^+ nằm trong tế bào chất và đầu COO^- ở trong khoang thì túi tiết được hình thành sao cho đầu NH_3^+ sẽ nằm phía ngoài túi tiết. Túi tiết chở protein tiết sẽ tiếp hợp với màng và đảm bảo protein giữ đúng chiều phân cực giống như ở trên màng nội chất. Nếu chuỗi peptide là protein cấu trúc màng nguyên sinh chất và có đầu NH_3^+ phân bố bên

ngoài màng, thì đầu NH_3^+ của chuỗi này bị đưa vào trong khoang nội chất. Sau đó, túi tiết được hình thành có đầu NH_3^+ nằm phía trong túi. Rõ ràng, sự phân bố của các đầu NH_3^+ và COO^- trên màng nội chất đã được quy định phù hợp với sự phân bố ở vị trí đích cuối cùng (Hình 4.9).

4.7. TÍN HIỆU DẪN PROTEIN TỔNG HỢP TỰ DO TRONG TẾ BÀO CHẤT VỀ ĐÍCH

Mặc dù hai bào quan ty thể và lục lạp đều có genome, có bộ máy tổng hợp protein riêng nhưng chúng không tự cung cấp đủ tất cả protein cho hoạt động của mình. Số protein thiếu hụt được vận chuyển từ tế bào chất vào. Hầu hết những protein này đều được tổng hợp trọn vẹn bởi các ribosome tự do trong tế bào chất. Chúng có chứa tín hiệu dẫn đặc hiệu cho từng bào quan. Để phân biệt với tín hiệu dẫn vào mạng lưới nội chất, tín hiệu dẫn về bào quan thường được gọi là ***leader sequences*** (*tạm dịch là tín hiệu dẫn đầu*). Tín hiệu này cũng bị cắt khỏi protein sau khi protein đã đi vào trong bào quan. Ngoài ra, các protein có chức năng trong nhân eukaryot cũng được tổng hợp trọn vẹn trong tế bào chất rồi mới vận chuyển vào nhân.

Protein đi vào bào quan hoặc vào trong nhân phụ thuộc vào tín hiệu dẫn đầu. Khác với tín hiệu dẫn vào lưới nội chất, tín hiệu dẫn đầu phải có cấu trúc không gian đặc thù được nhận biết bởi thụ thể trên bề mặt của bào quan, hoặc trên màng nhân. Tín hiệu dẫn đầu không có tính bảo thủ đối với các protein đi vào ty thể, hoặc đi vào lục lạp. Tuy nhiên, tín hiệu này đủ để đảm bảo dẫn protein về đúng đích. Thực nghiệm cho thấy, khi ghép tín hiệu này với protein phân bố trong tế bào chất thì protein này sẽ đi về bào quan. Sau khi xảy ra tương tác giữa thụ thể và tín hiệu, protein được đẩy từ bên ngoài và kéo từ bên trong để đi qua màng. Năng lượng phân huỷ từ GTP được dùng để "đẩy" và để "kéo" ở hai phía của màng.

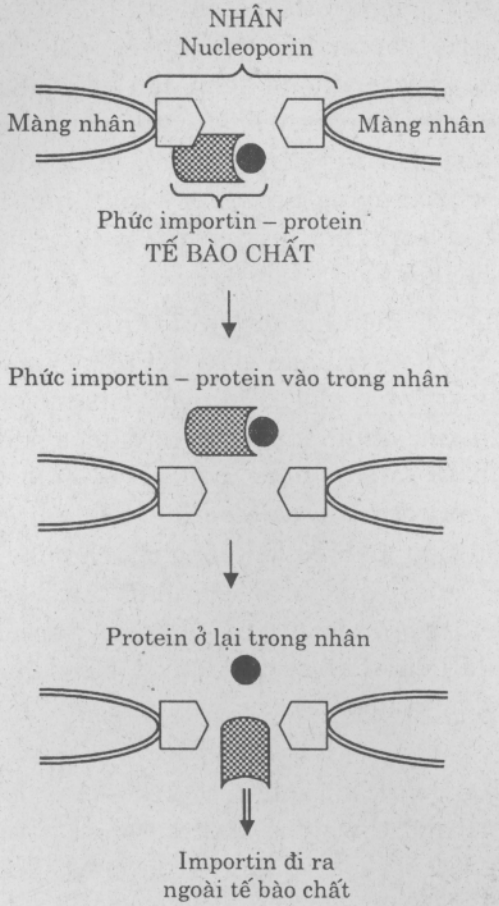
Cấu trúc đặc biệt của ty thể và lục lạp gồm nhiều khoang. Protein của hai bào quan này có thể phân bố ở khoang trong, khoang ngoài hoặc nằm giữa các khoang. Do đó protein cần có thêm các tín hiệu khác để về đúng vị trí. Các tín hiệu này thường nằm kế tiếp tín hiệu dẫn đầu. Tuy nhiên, chỉ sau khi tín hiệu dẫn đầu bị cắt thì tín hiệu tiếp theo mới được bộc lộ. Bản thân tín hiệu thứ hai cũng bị cắt khỏi protein sau khi protein đi đến đích. Như vậy, chúng ta hình dung

protein đi vào ty thể hay lục lạp có thể có nhiều tín hiệu, mỗi tín hiệu là một đích trung gian trên con đường vận chuyển. Các tín hiệu được nhận biết bởi thụ thể đặc hiệu và đi qua kênh dẫn khác nhau. Ví dụ, có sự khác biệt rõ ràng về cấu trúc của translocator nằm trên các màng bên trong ty thể và màng ngoài giữa ty thể với tế bào chất. Protein phải đi qua translocator trên màng ngoài rồi mới đi qua translocator trên màng trong để về các vị trí khác nhau trong ty thể. Điều này cũng xảy ra tương tự đối với lục lạp.

Chúng ta xét trường hợp điển hình đối với vận chuyển protein vào nhân tế bào eukaryot. Nhân được phân cách với tế bào chất bởi màng nhân. Màng nhân có cấu trúc hai lớp. Trên màng nhân có những vị trí tại đó hai lớp màng tiếp xúc với nhau. Vị trí này được gọi là cấu trúc siêu lỗ của màng nhân (***nuclear pore complex***). Các protein cấu thành nên siêu lỗ được gọi chung là ***nucleoporin***. Màng nhân tế bào động vật có khoảng 3000 siêu lỗ. Siêu lỗ không chỉ cho phép protein đi vào nhân mà còn để các loại ARN hoặc các phức ARN-protein đi ra ngoài tế bào chất. Ví dụ, r-protein đi vào nhân và tạo phức với các ARNr tạo thành các tiểu đơn vị ribosome lớn và ribosome nhỏ riêng biệt. Sau đó, từng tiểu đơn vị được vận chuyển qua siêu lỗ ra ngoài tế bào chất.

Cấu trúc siêu lỗ cho phép các protein có khối lượng phân tử nhỏ hơn 50kDa qua lại một cách thụ động theo sự chênh lệch nồng độ. Đối với những protein có khối lượng phân tử lớn hơn 50kDa bắt buộc chịu kiểm soát của cơ chế vận chuyển tích cực (có tiêu hao năng lượng). Điều cần chú ý là, dù vận chuyển tích cực hay thụ động, protein đi qua siêu lỗ đều phải có tín hiệu dẫn cho phép đi vào (gọi là tín hiệu định vị trong nhân – ***nuclear localization signal - NLS***), hoặc cho phép đi ra (tín hiệu đi ra khỏi nhân – ***nuclear export signal - NES***). Thụ thể tương ứng của tín hiệu NLS là ***importin***, còn thụ thể đặc hiệu cho tín hiệu NES là ***exportin***. Trình tự acid amin của các tín hiệu này không có tính chất bảo toàn giữa các loại protein đi vào nhân hoặc giữa các protein đi ra. Do đó, cấu trúc không gian cục bộ xung quanh tín hiệu NLS và NES ảnh hưởng đến tính đặc hiệu của protein đi qua siêu lỗ.

Thụ thể importin và exportin không định vị trên màng nhân. Chúng sàng lọc nhận biết protein đi vào hoặc đi ra qua siêu lỗ để rồi đưa protein đến gần siêu lỗ. Đến lượt mình, các thụ thể tương tác với nucleoporin. Tiếp đến, protein được đẩy qua siêu lỗ (Hình 4.10).



Hình 4.10: Vận chuyển protein vào nhân đi qua siêu lỗ.

Thụ thể importin tương tác với NLS, sau đó tương tác với nucleoporin ở siêu lỗ. Protein ở lại trong nhân còn importin đi ra ngoài tế bào chất (theo Lewin, 2004).

4.8. PROTEIN CHAPERON

Hoạt tính của protein không chỉ phụ thuộc vào trình tự acid amin hay cấu trúc bậc 1 mà được quyết định bởi cấu trúc không gian (bậc 2, 3 và 4). Để đạt được một cấu hình chính xác, quá trình gấp khúc và sắp xếp các chuỗi peptide với nhau đòi hỏi sự trợ giúp của các protein có tên gọi là ***chaperon***. Ngay trong khi đang được tổng hợp, sợi peptide được chaperon bám vào để ngăn cản liên kết giữa các acid amin của các sợi gây sai lệch hoặc ảnh hưởng đến cấu trúc không gian. Hơn nữa, chaperon có thể tương tác với protein để duy trì cấu hình không gian mà protein đã có ; nhờ đó kiểm soát hoạt tính của protein

hoặc tạo thuận lợi để vận chuyển protein về đích. Ngoài ra, trong một số trường hợp, chaperon ngăn cản nguy cơ lắng đọng protein, hoặc phức các protein khi nồng độ của chúng tăng cao trong tế bào chất.

Hsp70 (heat shock protein) là họ các protein chaperon phổ biến trong tế bào vi khuẩn cũng như tế bào eukaryot (phân bố ở tế bào chất, trong khoang mạng lưới nội chất, ty thể và lục lạp). Protein Hsp70 có hai vùng với chức năng khác nhau : vùng phía đầu NH_2 có hoạt tính ATPase, còn phía đầu COOH có khả năng tương tác với cơ chất (protein). Khi liên kết với ATP, Hsp70 bám vào cơ chất và tách ra rất nhanh. Tuy nhiên, khi ATP phân huỷ thành ADP, quá trình này xảy ra chậm. Thay đổi liên kết của Hsp70 giữa ATP và ADP phụ thuộc vào hai protein Hsp40 và GrpE. Hsp40 giúp phức Hsp70-ATP tương tác với cơ chất (protein). ATP phân huỷ thành ADP để thay đổi sự gấp khúc của protein. ADP được thay thế bởi ATP nhờ GrpE. Chu trình sẽ lặp lại trên dọc chiều dài sợi peptide nhằm đảm bảo tính chính xác cho cấu trúc không gian của protein.

Chương 5

KỸ THUẬT ADN TÁI TỔ HỢP

Kỹ thuật ADN tái tổ hợp liên quan đến việc cắt, nối các đoạn ADN có nguồn gốc khác nhau tạo thành phân tử ADN có chức năng; tức là phân tử này có thể được phiên mã, dịch mã, tái bản tạo số lượng lớn trong tế bào hoặc trong *invitro*. Phân tử ADN bao gồm các đoạn ADN mà trật tự sắp xếp của chúng không tồn tại trong tự nhiên được gọi là ADN tái tổ hợp.

Thiết kế phân tử ADN tái tổ hợp được bắt đầu với các enzym. Ví dụ, enzym giới hạn endonuclease (*restriction endo-nucleases*) có khả năng cắt các đoạn ADN lớn thành các đoạn nhỏ hơn. Tiếp đến, các đoạn ADN được nối với nhau nhờ ADN ligase. Hơn nữa, nếu như phân tử ADN có mang tâm tái bản thì số lượng của chúng tăng cao sau khi được đưa vào tế bào vi khuẩn. Ví dụ, một đoạn ADN có chứa gen được cắt ra từ genome người và được ghép vào plasmid (có tâm sao chép) tạo thành phân tử ADN tái tổ hợp. Phân tử này được đưa trở lại tế bào vi khuẩn. Sau 24 giờ nuôi cấy, khoảng 10^{12} phân tử ADN giống hệt phân tử ban đầu được tạo ra. Do đó, số lượng đoạn ADN mang gen của người cũng được nhân bản lên theo. Lúc này đoạn ADN đó xem như đã được tách dòng và các plasmid mang ADN lạ được gọi là vector tách dòng (cloning vector). Nếu như vector có mang promoter và các tín hiệu kết thúc phiên mã phù hợp thì đoạn ADN được phiên mã và dịch mã tổng hợp protein theo các mã di truyền trên đoạn ADN. Nếu như vector tách dòng được đưa vào tế bào động vật, ADN lạ có thể ghép vào genome và tế bào biểu hiện tính trạng mới.

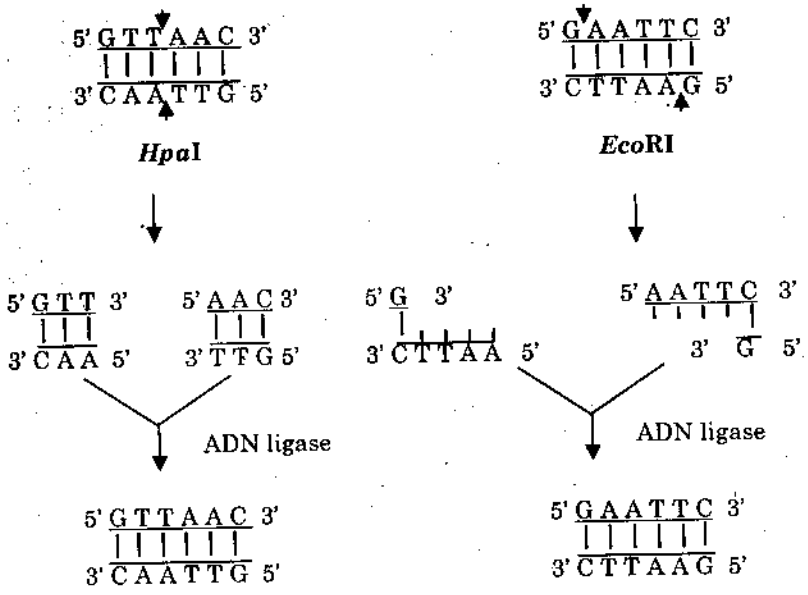
Vấn đề đặt ra làm thế nào tách được một gen riêng biệt ra khỏi genome? Làm thế nào xác định được phân tử ARNm tương ứng với một gen? Khi đã có gen, chúng ta có thể tổng hợp được protein *invitro* và *invivo* hay không? Mặt khác chúng ta có thể điều khiển hoạt động của gen sau khi đã thay đổi (gây đột biến) trình tự nucleotide hay không? Trong cơ thể, đa số các gen hoạt động một cách chuyên biệt, tức là

chúng thường biểu hiện ở một số, hoặc thậm chí trong một loại tế bào biệt hoá ở những thời điểm nhất định, với mức độ biểu hiện khác nhau. Làm thế nào để tìm ra được các gen đó và biết được chức năng của chúng? Ngoài ra có những căn bệnh do đột biến gen gây ra (bệnh di truyền) nhưng sản phẩm của gen chưa xác định được. Làm thế nào để phân lập được gen và tìm ra những biến đổi ở mức độ ADN liên quan đến bệnh? v.v... Tất cả những vấn đề nêu trên đã và đang được sinh học phân tử giải quyết, đặc biệt nhờ áp dụng các kỹ thuật tách dòng và tái tổ hợp ADN.

5.1. CẮT ADN BẰNG ENZYM GIỚI HẠN (RESTRICTION ENDONUCLEASE)

Enzym giới hạn (*restriction endonuclease* – RE) có khả năng nhận biết và phân cắt ADN tại vị trí có các nucleotide sắp xếp theo một trật tự. Tuỳ thuộc vào khoảng cách giữa vị trí nhận biết và vị trí cắt để phân biệt enzym giới hạn thành các loại khác nhau. Các enzym nhận biết và cắt ngay tại một vị trí được xếp vào loại II. Đây là những enzym được sử dụng nhiều trong kỹ thuật ADN tái tổ hợp. Chúng chủ yếu được tách chiết từ các vi sinh vật. Việc gọi tên của các enzym tuân theo quy tắc như sau: Chữ đầu tên enzym là chữ đầu tên của giống vi sinh vật từ đó enzym được tách chiết. Chữ này phải được viết hoa. Hai chữ tiếp theo không viết hoa bắt nguồn từ tên loài vi sinh vật. Cả 3 chữ đầu tiên đều in nghiêng. Tiếp theo 2 chữ này có thể là một chữ viết hoa nữa để chỉ chủng vi sinh vật đã sử dụng. Cuối cùng thường là chữ số La Mã chỉ thứ tự enzym được phát hiện từ loài vi sinh vật đó (do một loài vi sinh vật có nhiều loại enzym giới hạn). Ví dụ, *Eco*RI và *Eco*RII là các enzym giới hạn tách chiết đầu tiên (I) và thứ hai (II) từ vi khuẩn *Escherichia coli* Ry**13**.

Enzym giới hạn nhận biết và cắt trình tự đặc hiệu gồm 4 đến 8 nucleotide. Đặc điểm của các trình tự này là đối xứng (palindrome); tức là trình tự nucleotide tại vị trí cắt trên hai sợi đơn đều giống nhau theo cùng một chiều (5'→ 3' hoặc 3'→ 5'). Enzym có thể cắt hai sợi đơn ADN tại cùng một điểm tạo ra hai đầu cắt bằng phẳng. Những enzym đó được gọi là enzym cắt đầu bằng. Bên cạnh đó, tại vị trí nhận biết, enzym giới hạn có thể cắt hai sợi đơn ADN lệch nhau vài nucleotide tạo ra hai đầu có các sợi đơn so le nhau. Những enzym này được gọi là enzym cắt đầu so le (Hình 5.1). Enzym tạo đầu so le rất thông dụng trong ADN tái tổ hợp do ADN ligase nối các đoạn ADN đầu so le với hiệu suất cao hơn rất nhiều so với các đoạn đầu bằng.



Hình 5.1: Enzym giới hạn nhận biết và cắt ADN ở trình tự nucleotide sắp xếp đặc hiệu tạo ra hai đoạn có đầu bằng (*Hpa*I) hoặc các đầu đầu so le (*Eco*RI). Các đoạn ADN được nối với nhau nhờ ADN ligase. Đặc biệt hai đoạn ADN có đầu so le tạo ra bởi một enzym sẽ nối với nhau dễ dàng hơn nhờ liên kết bổ sung tại hai đầu so le.

Xác suất phân bố các vị trí nhận biết của enzym giới hạn sẽ giảm tỷ lệ nghịch với số nucleotide ở vị trí nhận biết. ADN chứa bốn loại nucleotide A, T, C, G khác nhau. Xác suất để 4 nucleotide sắp xếp theo một trật tự sẽ là $(1/4)^4 = 1/256$. Như vậy, trung bình một đoạn ADN dài 256 nucleotide sẽ có một vị trí của enzym giới hạn nhận biết 4 nucleotide sắp xếp theo một trật tự xác định. Xác suất này giảm xuống 1/4000 khi vị trí gồm 6 nucleotide. Như vậy, khi phân cắt một genome có kích thước xác định, số lượng và chiều dài của các đoạn tạo thành phụ thuộc vào việc lựa chọn số nucleotide ở vị trí nhận biết của enzym giới hạn. Ngoài ra, cần chú ý đến hiện tượng methyl hoá ADN ức chế một số enzym nhận biết vị trí có cytosine hoặc adenine bị methyl hoá. Ví dụ, *Hpa*II và *Msp*I đều nhận biết CCGG nhưng hoạt tính của chúng thay đổi tuỳ thuộc vào nhóm methyl gắn vào C đầu tiên hay C tiếp theo.

5.2. PHÂN LY CÁC ĐOẠN NUCLEOTIDE CÓ KÍCH THƯỚC KHÁC NHAU BẰNG KỸ THUẬT ĐIỆN DI

Phân tử acid nucleic tích điện âm nên chúng chuyển động từ cực âm sang cực dương trong điện trường một chiều. Các đoạn ADN (hoặc

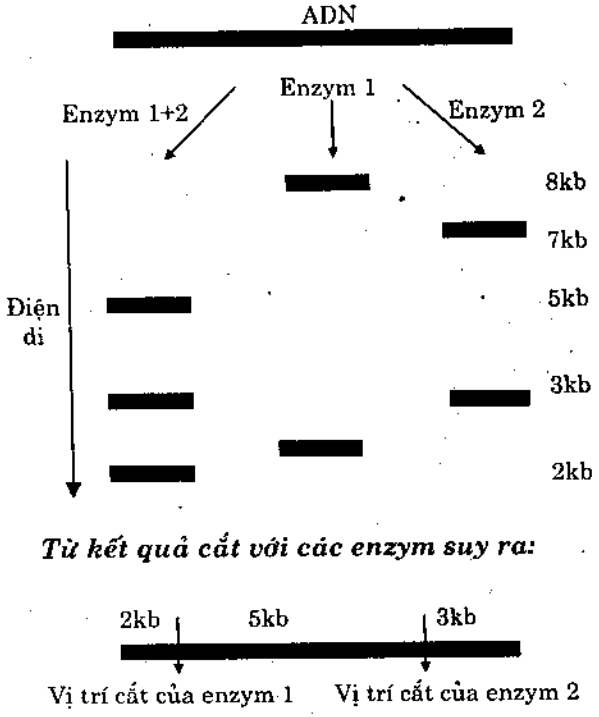
ARN) có kích thước khác nhau được phân ly bằng phương pháp điện di trên gel agarose (điện di ngang), hoặc gel polyacrylamide (điện di đứng). Các phân tử ADN có kích thước lớn (~ 10^5-10^{7bp}) được phân ly bằng điện di trên trường xung đẩy "pulsed-field". Những đoạn dài từ vài chục nucleotide đến 10.000bp được phân ly trên gel agarose hoặc acrylamide với nồng độ khác nhau. Nồng độ gel lớn phù hợp với các đoạn ADN ngắn. Ví dụ, những đoạn ADN ngắn 200bp sai khác nhau không quá 10bp có thể phân biệt trên gel agarose 4%. Tuy nhiên, nồng độ agarose không thể tăng cao do gel giòn, dễ gãy và có điện trở lớn. Vì vậy, gel polyacrylamide có ưu thế hơn so với gel agarose trong việc phân ly các đoạn ADN ngắn, có số lượng nucleotide chênh lệch nhau ít, mặc dù khâu chuẩn bị và thời gian điện di với gel polyacrylamide lâu hơn. Gel polyacrylamide có khả năng phân giải những đoạn ADN hơn kém nhau một nucleotide (phương pháp xác định trình tự nucleotide). Ngoài ra, ARN thường điện di trên gel có thêm formadehyde ngăn cản cấu trúc ghép đôi ở sợi đơn ARN và ức chế ARNase. Sau khi điện di, các đoạn ADN, ARN trên gel có thể quan sát được dưới tia tử ngoại khi chúng được nhuộm với ethidium bromide (EtBr), hoặc có thể quan sát bằng mắt thường khi nhuộm bạc hay được đánh dấu bằng các chất phát màu khác nhau. Ngoài ra, ADN, ARN còn được phát hiện bằng việc gắn phóng xạ (P^{32}, H^3 hoặc S^{35}). Phương pháp đánh dấu phóng xạ rất nhạy, cho phép phát hiện lượng mẫu rất nhỏ (cỡ ng).

5.3. XÂY DỰNG BẢN ĐỒ VỊ TRÍ CỦA ENZYM GIỚI HẠN

Các đoạn ADN bị cắt bởi enzym giới hạn được gọi là các đoạn giới hạn (restriction fragment). Xác định vị trí của các enzym giới hạn trên phân tử ADN cho phép xây dựng bản đồ vị trí giới hạn. Bản đồ này rất có ý nghĩ trong việc xác định thứ tự sắp xếp của các phân đoạn ADN thu nhận được từ một đoạn lớn hơn (Hình 5.2).

Giả sử chúng ta có một đoạn ADN phân lập được từ ngân hàng (xem chi tiết phần 5.7) nhưng chưa biết vị trí của các enzym giới hạn trên đoạn này. Sử dụng riêng biệt hoặc cùng phối hợp các enzym khác nhau để cắt đoạn ADN đó thành các phân đoạn nhỏ hơn. Điện di các phân đoạn để xác định kích thước của chúng. Căn cứ vào sự phối hợp giữa các kích thước đó, chúng ta biết được trật tự phân bố của các enzym trên đoạn ADN, tức là chúng ta xác định được bản đồ vị trí giới hạn của đoạn đó (Hình 5.2). Việc lập bản đồ vị trí giới hạn cho phép so sánh cùng một đoạn ADN trong genome của các cá thể khác nhau. Ví dụ, so sánh các vị trí giới hạn trên đoạn ADN mã cho ARNr có ý nghĩa

trong nghiên cứu tính đa dạng di truyền và quá trình tiến hoá. Ngoài ra, bản đồ vị trí giới hạn rất hữu ích cho kỹ thuật tách dòng và công nghệ gen (DNA clonning and engineering).



Hình 5.2: Kích thước các đoạn ADN cùng bị cắt bởi hai hoặc từng enzym riêng biệt giúp cho việc xác định trật tự các vị trí cắt của các enzym đó và xác định bản đồ giới hạn của phân tử ADN.

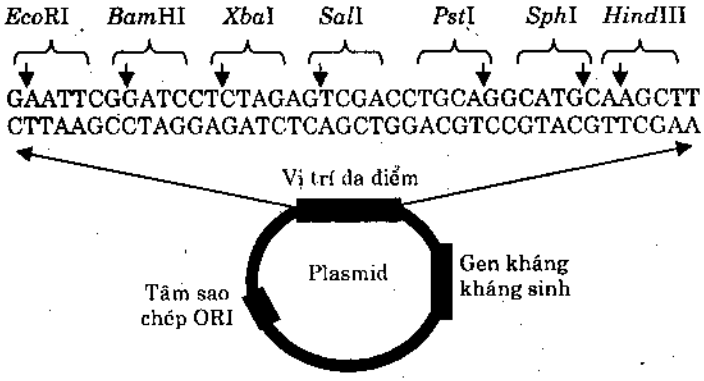
5.4. CÁC VECTOR TRONG KỸ THUẬT TÁCH DÒNG

Để có thể lưu trữ và nhân bản một đoạn ADN (hoặc 1 gen) đến một số lượng mong muốn, một trong các phương pháp thông dụng là gắn đoạn ADN đó vào vector. Sau đó vector được đưa vào tế bào nhận. Sử dụng bộ máy tái bản ADN của tế bào nhận, vector được tái bản tạo ra số lượng lớn các bản sao giống hệt vector ban đầu. Với những nhiệm vụ như vậy, vector cần phải có các đặc tính cơ bản sau:

– Có vị trí để ghép nối đoạn ADN lạ vào (vị trí đa điểm tách dòng-*polycloning site*). Đây là những vị trí duy nhất trên vector (Hình 5.3).

– Có tâm sao chép (*origin of replication*) được nhận biết bởi bộ máy tái bản ADN của tế bào nhận, cho phép vector tái bản một cách tích cực trong tế bào nhận.

- Có gen chỉ thị (*selactable marker gene*) giúp cho việc phân biệt tế bào nhận vector mang ADN lạ với các tế bào khác. Gen chỉ thị thường là gen quy định khả năng chống chịu kháng sinh.



Hình 5.3: Các yếu tố cơ bản và cấu trúc vị trí đa điểm tách dòng của vector.

Vị trí cắt của các enzyme giới hạn được ký hiệu bởi các mũi tên. *EcoRI-Escherichia coli* RY13, *BamHI-Bacillus amyloliquefaciens* H, *XbaI-Xanthomonas badrii*, *SalI-Streptomyces albus*, *PstI-Providencia stuartii*, *SphI-Streptomyces phacochromogenes* và *HindIII-Haemophilus influenzae* R_d.

Vector có thể tồn tại trong tế bào vi sinh vật (ví dụ như plasmid trong tế bào vi khuẩn), hoặc trong tế bào eukaryot (vector YAC trong tế bào nấm men). Ngoài ra, vector được thiết kế có khả năng tồn tại ở trong các loại tế bào prokaryot và eukaryot (vector con thoi). Tuỳ thuộc vào kích thước của các đoạn ADN cần lưu giữ mà chọn các vector khác nhau. Vì vậy, việc lựa chọn và thiết kế vector phụ thuộc vào mục đích nghiên cứu. Hầu hết các vector sử dụng hiện nay là các phân tử ADN tái tổ hợp. Chúng được thiết kế nhằm đạt được hiệu suất sử dụng tối ưu nhất cho từng mục đích nghiên cứu. Phổ biến nhất là các plasmid, genome của bacteriophage (gọi tắt là phage hay λ), cosmid, các vector BAC, YAC (Bảng 5.1).

Bảng 5.1: Một số loại vector thông dụng trong kỹ thuật tách dòng

Vector	Nhận ADN lạ tối đa
Plasmid	7-9kb
Phage	10-15kb
Cosmid	45-50kb
BAC	100-300kb
YAC	1000kb (1Mb)

5.4.1. PLASMID

Plasmid là những phân tử ADN dạng vòng, tồn tại trong tế bào vi khuẩn và nằm ngoài nhiễm sắc thể. Plasmid có kích thước thay đổi trong khoảng 1kb đến vài trăm kb. Trong tự nhiên, plasmid có thể tái bản đồng thời với nhiễm sắc thể, hoặc có khả năng tái bản độc lập. Thông thường, các plasmid tái bản độc lập thường có kích thước nhỏ (2,5-4kb) và có số lượng bản sao lớn trong một tế bào vi khuẩn. Plasmid thường mang những gen chống chịu kháng sinh, vì vậy những gen đó được sử dụng như các chỉ thị (marker) để chọn lọc tế bào chứa plasmid.

Plasmid đầu tiên được sử dụng trong kỹ thuật tách dòng là pSC101 chỉ có một gen chỉ thị quy định tính chống chịu tetracyline (*tet*^{r}) và một vị trí *Eco*RI duy nhất để ghép ADN lạ vào. Đây là plasmid tồn tại sẵn có trong tự nhiên. Ngày nay có rất nhiều loại plasmid khác nhau được thiết kế có mang hai hoặc nhiều gen chỉ thị và vị trí đa điểm tách dòng gồm nhiều vị trí nhận biết của các enzyme giới hạn. Kích thước của các plasmid được thiết kế sao cho không lớn giúp plasmid nhân bản rất nhanh trong tế bào nhận. Đặc biệt, vị trí đa điểm của plasmid thường được thiết kế nằm trong một gen chỉ thị. Khi đoạn ADN xen vào tại một trong các điểm này, chúng gây bất hoạt gen chỉ thị. Do đó, tế bào nhận plasmid gắn ADN lạ sẽ không tổng hợp được sản phẩm của gen chỉ thị. Sự thiếu vắng sản phẩm của gen chỉ thị có thể liên quan đến sự chuyển màu của môi trường nuôi cấy. Vì thế, dựa vào màu sắc, chúng ta chọn lọc được tế bào nhận plasmid gắn ADN lạ. Ví dụ, các plasmid pUC có vị trí đa điểm nằm trong trình tự của gen chỉ thị mã cho β-galactosidase. Khi pUC có gắn ADN lạ, enzym không được tổng hợp. Nếu chúng ta cho vào môi trường nuôi cấy cơ chất của enzym, cơ chất không bị chuyển hoá nên tế bào nhận pUC gắn ADN lạ sẽ có màu trắng. Ngược lại, những tế bào nhận pUC có gen chỉ thị nguyên vẹn, chúng sẽ có màu xanh tạo ra do cơ chất bị chuyển hoá bởi β-galactosidase.

Các plasmid kích thước nhỏ thường thích hợp với các đoạn ADN không dài quá 7-9kb. Khi đoạn ADN có kích thước lớn hơn, trải qua một số lần nhân bản, đoạn này bị ngắn dần hoặc bị loại khỏi plasmid do tế bào có xu hướng đào thải ADN không cần thiết. Quá trình đưa plasmid vào vector được gọi là biến nạp. Tế bào có khả năng tiếp nhận plasmid được gọi là tế bào khả biến.

Khi vector có chứa promoter nhằm mục đích phiên mã đoạn ADN lạ trong tế bào nhận thì sẽ được gọi là vector biểu hiện (***expression***

vectors). Thông thường các vector được thiết kế có chứa promoter mạnh không đòi hỏi các protein hoạt hoá, do đó lượng ARN phiên mã từ ADN lạ được tổng hợp rất nhiều.

5.4.2. PHAGE

Bacteriophage (thực khuẩn thể) có ưu điểm đặc biệt là khả năng xâm nhiễm tế bào chủ rất cao và số lượng thể phage mới tăng rất nhanh. Do đó, hầu hết các vector thực khuẩn thể được thiết kế dựa vào nhiễm sắc thể của thực khuẩn thể λ (lamda). Nhiễm sắc thể thực khuẩn thể λ gồm 48.502bp mã cho 46 gen. Khi xâm nhiễm vào tế bào vi khuẩn, ADN của phage được "đóng gói " trong vỏ protein. Khi "đóng gói", kích thước nhiễm sắc thể λ không cần phải chính xác mà có thể thay đổi nhưng không nhiều quá 5% hoặc ít hơn 25%. Sử dụng tính chất tự nhiên này, đoạn ADN lạ được thêm vào nhiễm sắc thể của phage. Lúc này nhiễm sắc thể λ trở thành vector. Các vector phage mang ADN lạ được cho xâm nhiễm vào tế bào vi khuẩn. Chúng tái bản và gây trạng thái sinh tan (phá hủy các tế bào chủ) và giải phóng ra các phage mới. Bằng hình thức đó, ADN lạ được nhân lên cùng với nhiễm sắc thể phage.

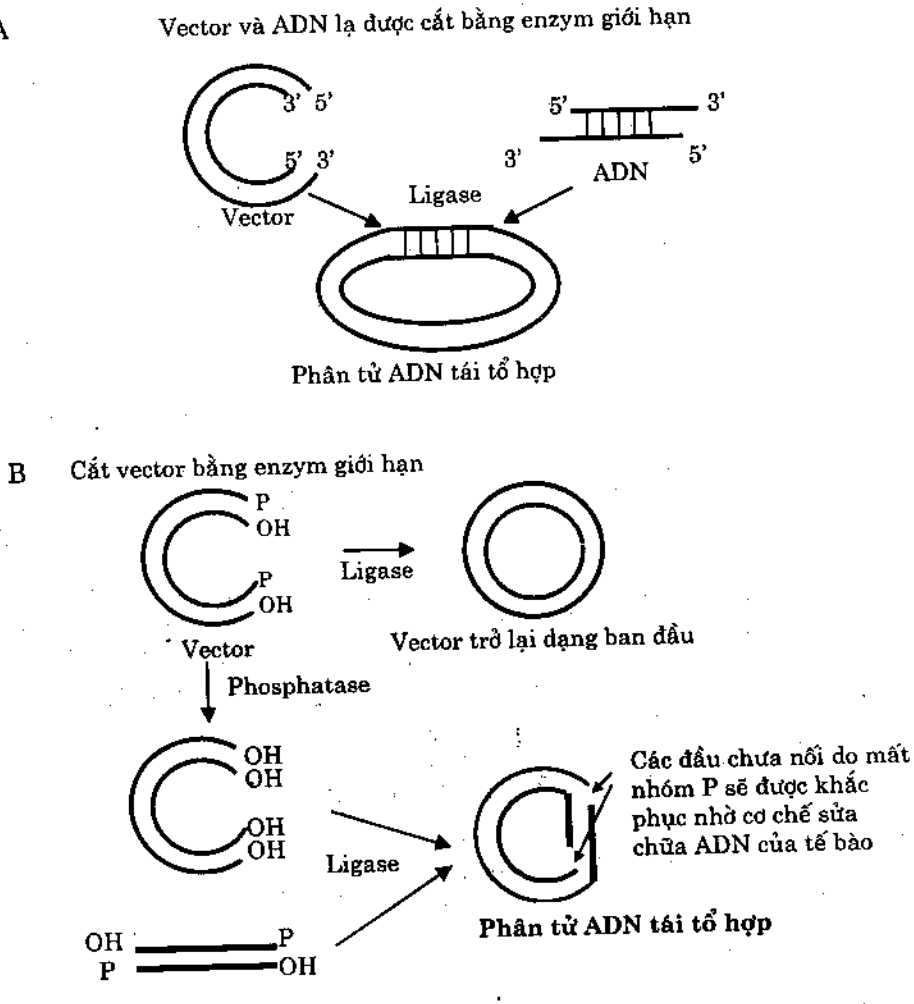
Cấu trúc và chức năng của tất cả các gen phân bố trên nhiễm sắc thể λ đã được xác định hoàn toàn. Các gen liên quan đến trạng thái tiềm tan chiếm khoảng 1/3 nhiễm sắc thể λ (khoảng 15kb). Chúng có thể bị loại đi mà không làm ảnh hưởng đến hoạt động của những gen gây trạng thái sinh tan. Vì thế chúng được thay thế bởi ADN lạ. Do đó độ dài của ADN đưa vào nhiễm sắc thể λ có thể đạt đến 15kb. Khi nhiễm vào tế bào chủ (tế bào vi khuẩn), các vector phage mang ADN mất khả năng tồn tại ở trạng thái tiềm tan. Chúng tái bản và gây trạng thái sinh tan, giải phóng ra nhiều thể phage mới.

Vector phage có khả năng tiếp nhận đoạn ADN có kích thước lớn đến 15kb. Đặc biệt, phage M13 là phân tử ADN sợi đơn có kích thước 6,4kb. Do tính chất đặc biệt này, vector phage M13 thường được sử dụng khi xác định trình tự đoạn ADN lạ chèn vào vector.

5.5. ĐƯA ADN VÀO VECTOR

Để đưa một đoạn ADN vào vector, vector phải được cắt bằng enzym giới hạn tại một vị trí duy nhất trên vector. Hai đầu vết cắt sẽ được nối với đoạn ADN bằng enzym ADN ligase để tạo thành phân tử tái tổ hợp. Khi hai đầu đoạn ADN và hai đầu cắt của vector có các nucleotide tương đồng thì hiệu quả nối ghép được tăng lên nhiều. Vì vậy, đoạn ADN và

vector thường được cắt với cùng một loại enzym giới hạn tạo đầu so le. Các đầu này có trình tự bắt cặp dính với nhau tạo điều kiện thuận lợi cho ADN ligase nối chúng với nhau. Ưu điểm của phương pháp này là dùng chính enzym giới hạn đó lấy lại được đoạn ADN từ vector chứa chúng sau khi vector này được nhân lên trong tế bào. Tuy nhiên trong phản ứng nối các đầu cắt với nhau, có thể chính hai đầu của vector được nối lại, hoặc hai hay nhiều đoạn ADN nối với nhau trước khi ghép vào vector. Để tránh những nhược điểm đó, hoặc là vector (ở dạng thẳng), hoặc là đoạn ADN sau khi bị cắt bởi enzym giới hạn, được xử lý với phosphatase để loại nhóm P ra khỏi đầu 5' (Hình 5.4).



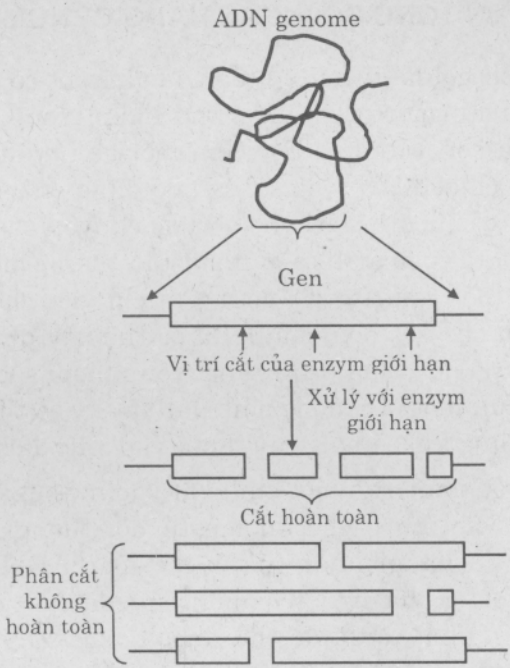
Hình 5.4: (A) Vector và ADN được cắt bởi một enzym giới hạn tạo các đầu so le thuận lợi cho việc đưa đoạn ADN vào vector. (B) Sử dụng phosphatase khử nhóm phosphat (P) ở đầu 5' của vector để tránh sự tạo lại dạng vòng của vector không chứa ADN lạ.

5.6. NGÂN HÀNG ADN TỔNG SỐ (NGÂN HÀNG GENOME)

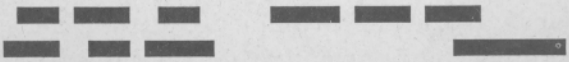
Genome (hay gòn gọi là hệ gen) ở tế bào eukaryot có kích thước rất lớn. Điều đó khiến chúng ta gặp nhiều khó khăn trong thao tác thực nghiệm khi chúng ta chủ yếu làm việc trên một gen nhất định. Vì vậy, chúng ta có thể tách chiết ADN tổng số, sau đó dùng enzym giới hạn cắt thành các đoạn; đưa các đoạn ADN vào vector, chuyển các vector vào tế bào nhận và tập hợp tất cả các tế bào nhận tạo thành một ngân hàng. Khi cần nghiên cứu một gen nào đó, chúng ta chỉ việc chọn gen này từ ngân hàng. Một cách chi tiết hơn, chúng ta có thể xây dựng ngân hàng của từng nhiễm sắc thể hoặc của từng vùng trên nhiễm sắc thể. Như vậy thông tin di truyền của một đối tượng sinh vật bất kỳ có thể được cất giữ trong các loại ngân hàng khác nhau, tuỳ thuộc vào mục đích sử dụng.

Để đưa toàn bộ genome của một sinh vật vào một ngân hàng, trước hết phải tách chiết và tinh sạch genome từ đối tượng đó. Tiếp đến, genome được cắt với enzym giới hạn thành các đoạn ADN có kích thước thích hợp với loại vector mà người nghiên cứu đã lựa chọn. Với một genome có kích thước nhất định, để thu được đoạn ADN dài cần chọn enzym nhận biết số nucleotide lớn (8 nucleotide). Ngược lại, để thu được các đoạn ADN ngắn thì nên chọn enzym nhận biết số nucleotide ít (4 nucleotide). Số đoạn ADN trong trường hợp thứ nhất sẽ ít hơn so với trường hợp thứ hai. Hơn nữa, một gen có thể bị cắt thành các đoạn ADN do vị trí nhận biết của enzym nằm ngay trong gen. Để hạn chế khả năng đó, thủ thuật của phản ứng cắt ADN là cần phải khống chế thời gian cũng như nồng độ của enzym sao cho phản ứng cắt xảy ra không hoàn toàn để mọi vị trí nhận biết không bị cắt đồng thời. Do đó, phản ứng cắt xảy ra có tính ngẫu nhiên với tất cả các vị trí nhận biết của enzym. Nhờ đó một gen có thể được bảo toàn ngay khi nó chứa vị trí cắt trong gen. Sau phản ứng, các đoạn ADN được đưa vào vector thích hợp. Vector cũng được cắt bằng chính enzym sử dụng cắt ADN genome. Tập hợp của mọi vector chứa ADN genome được đưa vào tế bào nhận và tạo thành ngân hàng ADN genome (Hình 5.5).

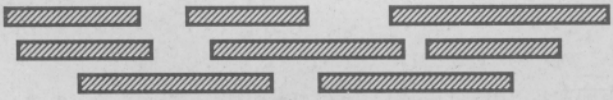
Mỗi vector của ngân hàng được gọi là một dòng (clone). Đối với một cá thể, sử dụng các loại vector khác nhau, có thể xây dựng được các ngân hàng ADN genome chứa các đoạn ADN dài, ngắn khác nhau (Hình 5.6). Tuy nhiên các ngân hàng này không hề có sự khác biệt về thông tin di truyền. Nhờ sử dụng đồng thời nhiều loại ngân hàng, việc lập bản đồ di truyền, bản đồ nhiễm sắc thể, nghiên cứu cấu trúc và chức năng một gen trở nên dễ dàng. Ngân hàng ADN tổng số được sử dụng rất phổ biến để phân lập một gen nguyên vẹn gồm exon và intron cũng như các vùng ADN điều khiển nằm phía đầu 5' của gen.



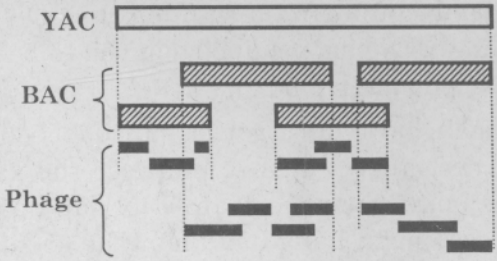
Hình 5.5: Enzym giới hạn phân cắt ADN tổng số không hoàn toàn nên các vị trí nhận biết không bị cắt đồng thời. Do đó, phản ứng cắt xảy ra có tính ngẫu nhiên với tất cả các vị trí nhận biết của enzym. Điều đó tránh một gen có thể bị cắt thành các đoạn nhỏ do vị trí nhận biết của enzym nằm ngay trong gen.



A. Đưa vào vector cosmid (các đoạn ADN dài khoảng 50kb)



B. Đưa vào vector BAC (các đoạn ADN dài khoảng 100kb)



C. Xây dựng ngân hàng theo cấu trúc *contig*

Hình 5.6: Xây dựng các loại ngân hàng ADN genome của cùng một đối tượng. Các dòng (clone) nằm gối lên nhau tạo điều kiện thuận lợi cho việc phân lập gen, xác định bản đồ genome. Cấu trúc một *contig* gồm các dòng tương ứng với những phân đoạn nhỏ cắt ra từ một đoạn ADN lớn hơn. Các ngân hàng dưới dạng contig đặc biệt cần thiết cho việc xác định giải mã toàn bộ genome.

5.7. NGÂN HÀNG CÁC ADNc (cDNA LIBRARY)

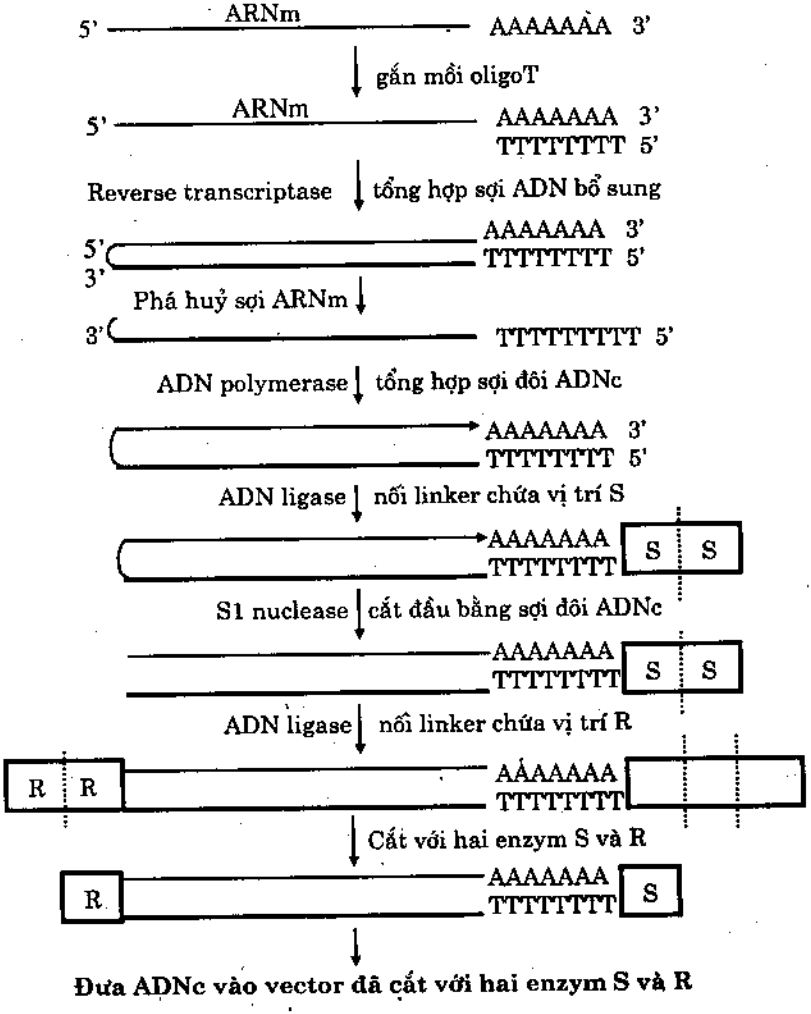
Ngân hàng ADN tổng số chứa tất cả các loại ADN có mặt trong genome. Đối với genome eukaryot, thành phần ADN chứa gen chỉ chiếm 1-2%. Vì vậy, số lượng các dòng chứa gen chiếm tỷ lệ tương tự trong ngân hàng ADN tổng số. Hơn nữa, các gen eukaryot thường có intron với kích thước lớn hơn rất nhiều so với exon. Điều đó góp phần. phức tạp hoá quy trình phân lập một gen (một dòng) từ ngân hàng tổng số. Xuất phát từ thực tế, các phân tử ARNm chỉ chứa các exon nên kích thước ARNm không lớn (dao động xung quanh 3000 nucleotide). Tập hợp tất cả các phân tử ARNm tách chiết từ một loại tế bào xác định đại diện cho tất cả các gen hoạt động trong loại tế bào đó. Vì vậy, thay cho việc thao tác trên gen (ADN), chúng ta khai thác từ ARNm một số thông tin về cấu trúc, chức năng của gen, của sản phẩm protein mà gen mã cho. Vấn đề là cần sao chép thông tin từ ARNm sang ADN do số lượng mỗi loại ARNm không nhiều và ARNm rất dễ biến tính nên thao tác với chúng không thuận lợi.

Để khắc phục nhược điểm này, chúng ta sử dụng enzym phiên mã ngược Reverse transcriptase để sao chép chính xác thông tin từ sợi ARNm sang ADN. Enzym này tổng hợp sợi đơn ADN bổ sung dựa vào sợi khuôn ARNm. Sợi kép ARN/ADN được xử lý với kiềm hoặc ribonuclease H (phân hủy sợi khuôn ARN) chuyển thành sợi đơn ADNc (complementary DNA). Tiếp sau đó, ADN polymerase I tổng hợp sợi đơn ADNc thứ hai để có được sợi đôi ADNc hoàn chỉnh. Phân tử ADN bổ sung dạng sợi đôi được gọi là ADNc. Cuối cùng, các phân tử ADNc được đưa vào vector.

Phân tử ADNc có hai đầu bằng. Do hiệu suất quá trình đưa các đoạn ADN đầu bằng vào vector thấp nên hai đầu của ADNc được xử lý trước khi đưa vào vector. Phân tử ADNc đầu bằng được nối với các oligonucleotide ngắn nhờ enzym ADN ligase. Những oligo này được tổng hợp nhân tạo có mang trình tự nhận biết của một hoặc vài enzym giới hạn tạo đầu so le. Sau đó, ADNc được phân cắt với enzym để tạo đầu so le. Bước cuối cùng là đưa ADNc vào vector đã cắt bằng chính enzym đã xử lý với ADNc. Do kích thước ADNc không lớn nên plasmid được chọn để nhận ADNc. Các cụ thể của quá trình tạo ADNc được mô tả trên Hình 5.7.

Tập hợp tất cả các ADNc tương ứng với mọi ARNm của một loại tế bào (một loại mô) được đưa vào plasmid. Mỗi vector nhận một phân tử ADNc. Tập hợp tất cả các plasmid chứa ADNc được chuyển vào tế bào nhận. Tập hợp tất cả các tế bào nhận chứa đầy đủ ADNc của một loại tế bào (loại mô biệt hoá) tạo thành ngân hàng ADNc của loại tế bào (loại mô) đó. Lúc này mỗi ADNc trong ngân hàng thường được gọi là

một dòng (một clone). Điểm khác biệt giữa ngân hàng ADN genome và ADNc là mỗi cơ thể có nhiều loại ngân hàng ADNc khác nhau, mỗi ngân hàng đặc trưng cho một loại mô tổ chức dùng để tách chiết ARNm. Sự khác nhau chủ yếu giữa hai loại ngân hàng ADN genome và ADNc là ở thành phần ADN. Bên cạnh đó số lượng dòng (clone) cũng như kích thước các đoạn ADN trong ngân hàng ADNc nhỏ hơn nhiều so với ngân hàng ADN genome.



Hình 5.7: Tổng hợp ADNc từ ARNm.

Enzym reverse transcriptase tổng hợp sợi đơn ADNc trên khuôn ARNm (sử dụng mồi oligoT tạo cặp bổ sung với đuôi polyA). Sợi đơn ARNm bị phân hủy trong điều kiện kiềm hoặc bởi ribonuclease H. Sợi ADNc thứ hai được tổng hợp trên khuôn ADNc thứ nhất nhờ ADN polymerase I. Phân tử ADNc (dạng sợi kép) có các đầu bằng được nối với hai loại linker S và R. Linker là đoạn oligonucleotide tổng hợp nhân tạo có mang vị trí nhận biết của enzym giới hạn. Do đó khi đưa vào vector, ADNc được giữ đúng chiều.

5.8. SÀNG LỌC MỘT DÒNG TỪ NGÂN HÀNG ADN

Để đảm bảo chọn được một dòng cần nghiên cứu từ ngân hàng, chúng ta cần tính toán số dòng cần thiết lấy ra từ ngân hàng dùng vào việc sàng lọc. Ví dụ, genome ruồi giấm có kích thước khoảng 10^5kb được cắt ra thành các đoạn trung bình 40kb. Ngân hàng chứa các đoạn ADN này cần có ít nhất N =11.513 dòng để đảm bảo một đoạn ADN bất kỳ sẽ có mặt trong ngân hàng với xác suất P = 99%. Số N được tính theo công thức sau:

$$N = [\ln (1-P)]/[\ln (1-1/n)] = [\ln (1-0,99)]/ [\ln [1-(40/100\ 000)]]$$

Trong đó: n là số đoạn ADN cắt ra từ genome. N là số dòng cần lấy từ ngân hàng để đảm bảo sàng lọc được một dòng với xác suất P. Như vậy, để sàng lọc một dòng từ ngân hàng genome ruồi giấm, việc đầu tiên là chúng ta phải lấy từ ngân hàng ít nhất là 11.513 dòng (Đây là 11.513 khuẩn lạc hoặc là 11.513 transformants nên dễ dàng đếm được khi pha loãng nuôi cấy trên môi trường).

Sàng lọc một dòng trong số hàng triệu dòng từ ngân hàng ADN của sinh vật bậc cao đòi hỏi nhiều thao tác và thời gian. Sàng lọc những gen mã cho enzym (amylase, protease...), hay gen quy định tính chống chịu (kháng lại kháng sinh, độc dược...), hoặc gen khuyết dưỡng có thể thực hiện dễ dàng hơn dựa vào hoạt tính enzym, khả năng mọc trên môi trường có độc dược, hoặc bổ khuyết chức năng. Quá trình sàng lọc những gen này được gọi chung là chọn lọc di truyền (*genetic selection*).

Sàng lọc gen chống chịu kháng sinh được xem là trường hợp điển hình của chọn lọc di truyền. Ví dụ, gen kháng lại peniciline được sàng lọc dễ dàng từ ngân hàng ADN genome của *Salmonella typhimurium* khi các vector của ngân hàng được biến nạp vào tế bào nhận *E.coli*. Những tế bào này được trải trên môi trường có chứa peniciline. Duy nhất dòng tế bào nhận vector mang gen pen^r có khả năng tạo khuẩn lạc trên môi trường này.

Ngoài ra, các gen tham gia một số con đường chuyển hoá trong tế bào cũng có thể sàng lọc dễ dàng theo cách thức chọn lọc di truyền nhờ thí nghiệm bổ khuyết chức năng. Ví dụ, muốn chọn được các gen mã cho enzyme tham gia sinh tổng hợp histidine ở *S. cerevisiae* thì có thể xây dựng ngân hàng ADN genome nấm men gồm các vector tái tổ hợp được biến nạp vào tế bào *E.coli* khuyết dưỡng không có khả năng tổng hợp được acid amin này. Quá trình sàng lọc chỉ tuyển chọn dòng tế bào tạo khuẩn lạc trên môi trường không có histidine.

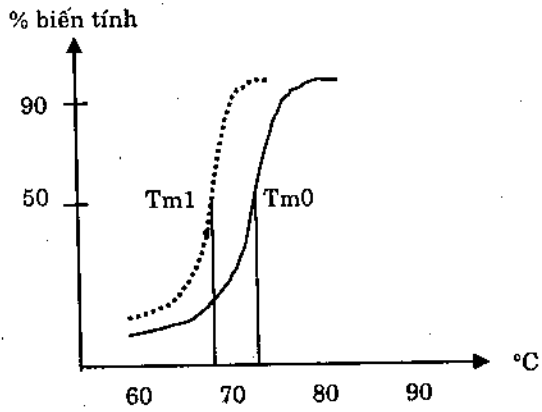
Khi phương pháp chọn lọc di truyền không thích hợp trong tuyển chọn, kỹ thuật lai acid nucleic dựa vào sự tạo cặp giữa các trình tự tương đồng và kỹ thuật miễn dịch dựa vào tương tác kháng nguyên-kháng thể có một vai trò quan trọng để sàng lọc được gen hoặc ADNc của gen cần nghiên cứu. Chúng ta cùng nhau xem xét nguyên tắc của kỹ thuật lai acid nucleic và ứng dụng để sàng lọc một dòng từ ngân hàng ADN.

5.8.1. Nguyên tắc của kỹ thuật lai acid nucleic

Chúng ta đều biết rằng, các 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 điều kiện thích hợp thì hai chuỗi đơn có thể liên kết tạo cặp trở thành chuỗi xoắn kép. Lúc đó chúng ta có phản ứng phục hồi phân tử ADN (renaturation). Điều kiện để tạo nên sợi kép là số lượng các nucleotide trên hai sợi đơn có khả năng tạo cặp bổ sung phải đạt đến một tỷ lệ nhất định. Chúng ta gọi tỷ lệ này là ***mức độ tương đồng*** (về trình tự nucleotide). Mức độ tương đồng càng lớn thì độ bền vững của sợi đôi càng cao. Mức độ tương đồng này thể hiện ở nhiệt độ cần thiết để gây biến tính 50% (nhiệt độ Tm) phân tử dạng sợi đôi (Hình 5.8). Chúng ta gọi chung các phân tử dạng sợi đôi này là các phân tử lai. 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, ARN khác nhau, hoặc giữa một sợi ADN và một sợi ARN. Ví dụ, phân tử lai ADN/ADN tạo ra giữa hai sợi đơn ADN có nguồn gốc từ các gen trong một họ gen. Phân tử lai ADN/ARN có thể tạo ra giữa trình tự của một gen (ADN) với phân tử ARNm phiên mã từ gen đó hoặc từ gen có độ tương đồng cao với gen của ARNm.

Nghiên cứu động học phản ứng lai (dựa vào sự biến đổi Tm) cho phép phân tích thành phần các loại ADN tham gia phản ứng. Ví dụ, gây biến tính rồi lại phục hồi ADN tổng số (ADN genome) tách chiết từ tế bào eukaryot, chúng ta sẽ thu được các phân đoạn khác nhau. Hai phân đoạn chính gồm thành phần ADN không lặp lại và thành phần ADN lặp lại. Khi chúng ta tiếp tục biến tính từng phân đoạn riêng biệt, Tm của phân đoạn thứ nhất hầu như không thay đổi qua các lần biến tính phục hồi lặp lại. Tuy nhiên, Tm của phân đoạn thứ hai luôn nhỏ hơn giá trị Tm đầu tiên. 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 Tm nhỏ hơn so với giá trị Tm 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 Tm không thay đổi ở lần biến tính thứ hai so với lần đầu.

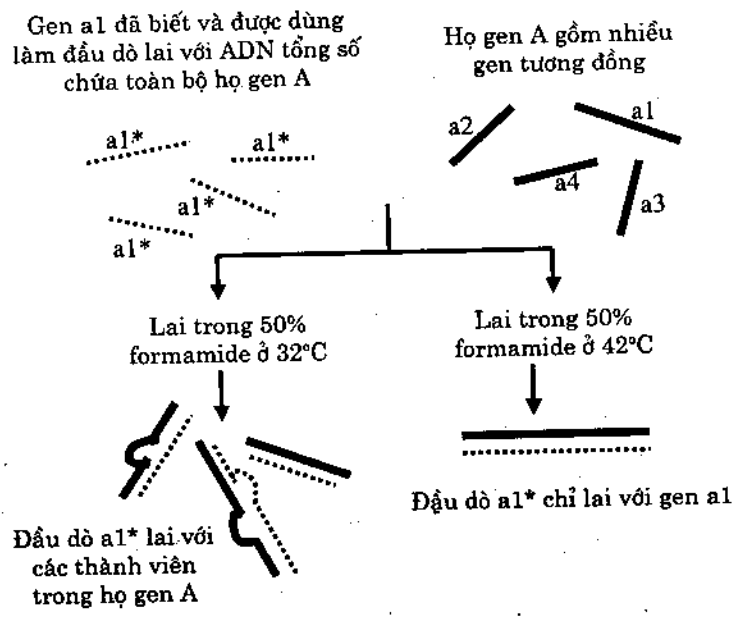


Hình 5.8: Động học phản ứng lai của thành phần ADN lặp lại (ADN gồm các trình tự nucleotide tương đồng).
Giá trị Tm0 của lần biến tính thứ nhất lớn hơn Tm1 của lần biến tính các phân tử lai. Phân tử lai hình thành trong phản ứng lai giữa các sợi đơn của lần biến tính thứ nhất.

5.8.2. Điều kiện của phản ứng lai

Khi thay đổi nhiệt độ hoặc thành phần, nồng độ các chất trong dung dịch của phản ứng lai, tức là thay đổi tính chặt chẽ nghiêm ngặt của điều kiện lai, khả năng tạo ra các phân tử lai bền giữa các sợi đơn ADN/ADN, ADN/ARN và ARN/ARN phụ thuộc rất nhiều vào

mức độ tương đồng. Trong điều kiện lai ngặt nghèo, phân tử lai chỉ tồn tại khi hai sợi đơn có độ tương đồng rất cao. Ví dụ, ARNm chỉ lai với một sợi đơn của gen phiên mã tạo ra ARNm đó. Trong điều kiện phản ứng lai bình thường, phân tử lai tạo ra từ các sợi đơn có độ tương đồng nhất định. Ví dụ, các gen trong một họ gen có thể lai với nhau. Khi một thành viên trong một họ gen đã biết, chúng ta có thể phát hiện được cả họ gen căn cứ vào các phân tử lai với thành viên này. Như vậy, khi biết một đoạn ADN có trình tự xác định, đoạn đó được dùng làm ***đầu dò*** (*probe*) để phát hiện các trình tự tương đồng lai với đầu dò (Hình 5.9).



Hình 5.9: Thay đổi điều kiện lai sẽ thu được các kết quả khác nhau.
Trong điều kiện khắc nghiệt, đầu dò a1* chỉ lai với đúng đoạn tạo cặp bổ sung chính xác. Trong điều kiện dễ dàng hơn, đầu dò có thể lai với các đoạn tương đồng.

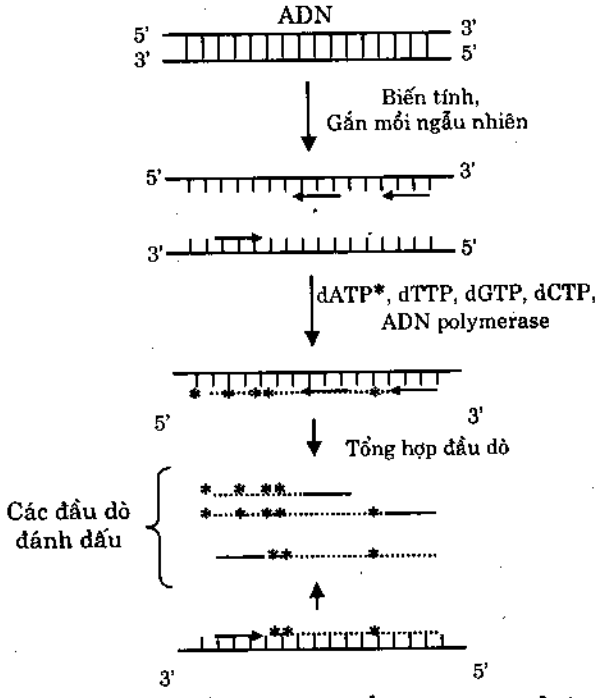
Thay đổi điều kiện lai được áp dụng cho phép phát hiện đột biến chỉ là một nucleotide. Ví dụ, bệnh thiếu máu hồng cầu hình lưỡi liềm xảy ra bởi thay đổi của một nucleotide duy nhất (GAG → GTG) ở gen mã cho chuỗi β globin. Để chẩn đoán thai nhi mang hai allen đột biến của gen này (một nhận từ bố, một nhận từ mẹ), hai chuỗi oligonucleotide (~20 base) được tổng hợp nhân tạo. Oligo thứ nhất ứng với gen bình thường, oligo thứ hai ứng với đoạn gen mang đột biến.

Như vậy hai đầu dò này chỉ sai khác nhau một nucleotide. Trong điều kiện lai rất nghiêm ngặt, các oligo chỉ lai với đoạn nucleotide tương ứng khi các liên kết bổ sung chính xác tuyệt đối. Do đó, tách ADN từ tế bào của bào thai và cho lai với đầu dò là hai oligo cho phép phân biệt dạng bình thường và dạng đột biến của gen mã cho β globin. Thai nhi mang hai allen khiếm khuyết thì ADN tách ra chỉ lai với một loại oligo tương ứng với đoạn đột biến.

Trong quá trình tiến hoá, một gen mới có thể được tạo ra do tổ hợp nhiều phần khác nhau của các gen, hoặc do thay đổi dần dần từ một gen ban đầu. Hầu hết các gen trong một họ gen có chức năng liên quan đến nhau và thường phân bố rải rác trong genome. Khi một thành viên trong họ đã được phân lập, các gen khác trong họ được xác định nhờ phép lai thực hiện trong điều kiện kém chặt chẽ (đầu dò chính là thành viên đã biết). Phương pháp lai này được sử dụng rất hữu ích trong nghiên cứu sinh học phân tử. Ví dụ, toàn bộ các gen mã cho protein tương tác với ADN làm nhiệm vụ điều khiển hoạt động của gen trong phát triển mô phôi ở ruồi giấm *Drosophila* được phát hiện nhờ phép lai ở điều kiện không ngặt nghèo. Sử dụng các đoạn nucleotide của các gen có nguồn gốc từ ruồi giấm làm đầu dò, các họ gen tương tự ở genome người và chuột được phát hiện nhờ chính kỹ thuật lai.

5.8.3. PHƯƠNG PHÁP ĐÁNH DẤU ĐẦU DÒ

Để phát hiện phân tử lai, đầu dò thường được đánh dấu bằng cách gắn với chất phát màu (huỳnh quang), hoặc với đồng vị phóng xạ. Mỗi kiểu đánh dấu có những ưu nhược điểm bổ trợ cho nhau. Ví dụ, đánh dấu bằng đồng vị phóng xạ nhạy rẻ tiền hơn so với đánh dấu huỳnh quang. Tuy nhiên, thao tác và xử lý chất thải phóng xạ đòi hỏi thiết bị bảo vệ và chi phí tốn kém. Các đồng vị phóng xạ thông dụng trong kỹ thuật đánh dấu gồm P^{32}, S^{35} và H^{3}. Mỗi loại đồng vị có khả năng xuyên sâu, năng lượng chùm tia phát ra và thời gian bán rã khác nhau. Do đó, tuỳ mục đích sàng lọc (mục đích phép lai) mà chọn đồng vị thích hợp. Ví dụ, lai đầu dò trực tiếp với lát cắt tế bào (lai tại chỗ-insitu) thì S^{35} được lựa chọn do chùm tia phát ra có năng lượng thấp, không có khả năng phát tán rộng nên tín hiệu rõ nét. Phương pháp đánh dấu đầu dò bằng chất phát huỳnh quang đòi hỏi chi phí cao hơn nhưng đảm bảo an toàn.



Hình 5.10: Tạo đầu dò bằng phương pháp tổng hợp ADN sử dụng mồi ngẫu nhiên và nucleotide đánh dấu (*). Sau phản ứng, số lượng đầu dò đánh dấu tăng nhiều so với lượng khuôn ban đầu.

Phương pháp thông dụng để đánh dấu đầu dò là tổng hợp ADN với một trong bốn loại nucleotide đã gắn với đồng vị phóng xạ hoặc với chất phát huỳnh quang. Đoạn ADN chọn làm đầu dò được dùng làm khuôn trong phản ứng tổng hợp ADN sử dụng mồi ngẫu nhiên chỉ gồm 6 nucleotide. Mồi ngẫu nhiên có khả năng tạo cặp ở những vị trí bất kỳ trên sợi khuôn. ADN polymerase sẽ thêm nucleotide đánh dấu vào đầu 3' của mồi ngẫu nhiên để tổng hợp nên sợi ADN mới có chứa nucleotide đánh dấu. Phương pháp này cho số lượng lớn các đầu dò đánh dấu nhưng chỉ thích hợp khi đoạn ADN khuôn có kích thước vài trăm nucleotide trở lên (Hình 5.10).

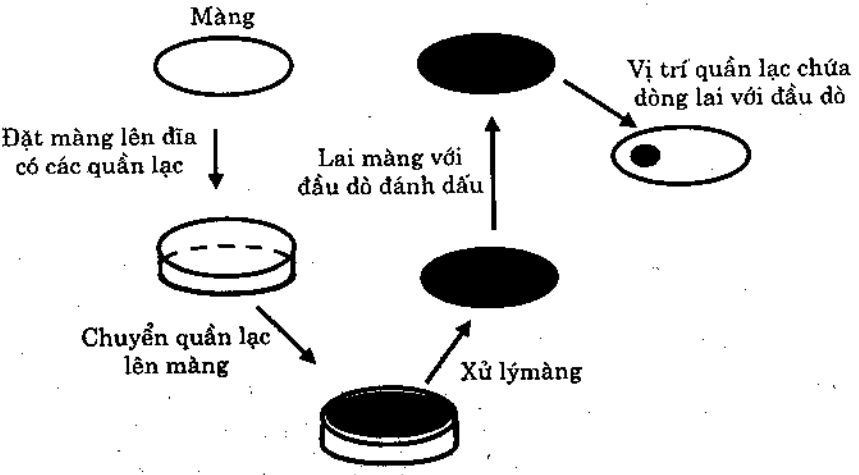
5.8.4. PHƯƠNG PHÁP SÀNG LỌC CHUNG (SCREENING)

Khi áp dụng kỹ thuật lai acid nucleic để lấy được một dòng (một đoạn ADN) từ ngân hàng (ADN genome hoặc ADNc), nhất thiết phải có đầu dò đặc hiệu. Đó là một đoạn ADN có trình tự nucleotide giống hệt hoặc có độ tương đồng cao với một phần bất kỳ của đoạn ADN cần tìm. Nếu cần sàng lọc một gen mà sản phẩm protein của gen đã biết, đầu dò có thể được tổng hợp nhân tạo dựa vào trình tự acid amin của

protein. Do tính chất thoái hoá của mã di truyền nên chúng ta không nhận được trình tự nucleotide chính xác. Điều này không ảnh hưởng nhiều tới kết quả sàng lọc do phép lai chỉ đòi hỏi độ tương đồng nhất định giữa đầu dò và ADN cần tìm.

Phương pháp phổ biến được áp dụng trước đây để sàng lọc (screening) một dòng từ ngân hàng được mô tả trên Hình 5.11. Số lượng N tế bào vi khuẩn (mang plasmid), hoặc các phage (mang vector λ hay cosmid) của ngân hàng được nuôi cấy trên các đĩa petri, mọc thành các quần lạc. Đặt lên trên bề mặt mỗi đĩa một màng sao cho các quần lạc được chuyển từ đĩa sang màng (thường là màng nylon có ái lực liên kết cao với ADN). Xử lý màng với kiềm để phá vỡ tế bào, gây biến tính ADN và cố định chúng trên màng. Sau đó các màng được ủ với dung dịch chứa đầu dò đặc hiệu. Đầu dò này được đánh dấu phóng xạ hoặc hoá học. Đầu dò sẽ lai với ADN có trình tự bổ sung với chính nó. Vị trí của quần lạc mang dòng lai sẽ được phát hiện và được tách ra khỏi các quần lạc khác. Các quần lạc này tiếp tục được nuôi cấy và quá trình sàng lọc như trên được lặp lại trong điều kiện lai chặt chẽ hơn nhằm giảm thiểu các dòng lai không đặc hiệu. Dung dịch lai chứa đầu dò dùng trong lần sàng lọc đầu tiên được sử dụng lại trong các lần sàng lọc lặp lại. Như vậy, sàng lọc cho phép xác định chính xác dòng mong muốn trong số hàng triệu dòng có mặt trong ngân hàng.

Điều cần lưu ý khi sàng lọc ngân hàng ADNc, đầu dò phải chứa trình tự của exon. Điều kiện sàng lọc một dòng từ ngân hàng ADN genome thường không đòi hỏi khắt khe như sàng lọc từ ngân hàng ADNc.



Hình 5.11: Sàng lọc quần lạc mang dòng cần tìm từ ngân hàng

Màng chứa các quần lạc được gọi là các replica. Chúng được xử lý với kiềm và nhiệt độ trước khi lai với đầu dò gắn phóng xạ hoặc phát huỳnh quang. Sau khi lai, vị trí quần lạc lai với đầu dò được phát hiện trên phim tự ghi phóng xạ hoặc dựa vào tín hiệu phát màu.

Ngoài ra nếu kháng thể đã biết, có thể sử dụng kỹ thuật miễn dịch để sàng lọc gen (ADNc) mã cho kháng nguyên. Lúc đó cần sử dụng ngân hàng ADNc chứa các vector biểu hiện gen (expression vector). Kháng nguyên được tổng hợp từ dòng mang ADNc mã cho kháng nguyên đó. Kháng thể được đánh dấu (bằng phóng xạ hoặc hóa học) làm đầu dò. Đầu dò lai với kháng nguyên và vị trí dòng tế bào được phát hiện. Một khi dòng ADNc được phân lập, có thể sử dụng nó làm đầu dò để sàng lọc tiếp gen tương ứng trong ngân hàng ADN genome. Hiện nay, để sàng lọc một hay nhiều dòng trong cùng một lần, kỹ thuật PCR hoặc microarray có rất nhiều ưu điểm, cho phép chọn được các dòng mong muốn một cách nhanh chóng và chính xác.

5.9. CÁC KỸ THUẬT LAI ACID NUCLEIC

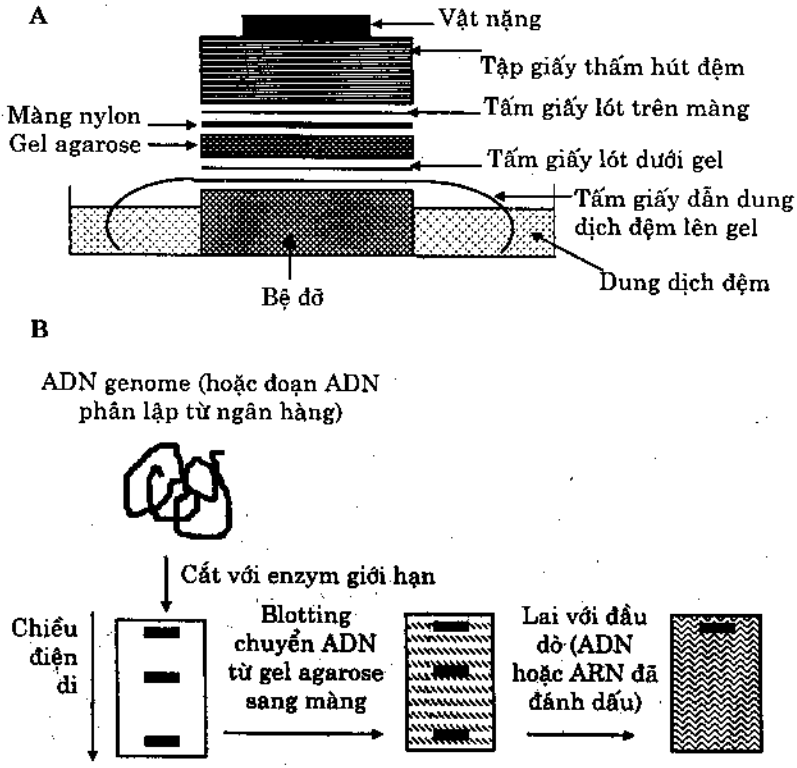
Nguyên tắc của phương pháp lai acid nucleic là dựa vào phản ứng tạo ra phân tử lai nhờ liên kết bổ sung giữa hai sợi đơn ADN, hoặc hai sợi đơn ARN hay ADN/ARN. Điều kiện của phản ứng lai phụ thuộc chủ yếu vào nhiệt độ và nồng độ muối trong đệm. Sợi đích lai với đầu dò đánh dấu bằng phóng xạ (P^{32}) được phát hiện nhờ phim có độ nhạy với phóng xạ. Mặc dù giá thành cao nhưng đảm bảo an toàn cho người sử dụng nên các chất đánh dấu không phóng xạ ngày càng chiếm ưu thế trong các phòng thí nghiệm hiện đại.

Phương pháp lai ADN với đầu dò được sử dụng chủ yếu để phân biệt một đoạn ADN trong số các đoạn khác mà kích thước của chúng không lớn. Ví dụ, sau khi phân lập được một đoạn ADN (một dòng) từ ngân hàng, chúng ta dùng kỹ thuật lai để định vị chính xác gen phân bố trên đoạn ADN đó, xác định cấu trúc của gen (phân bố của exon, intron). Ngoài ra, lai ADN genome (cắt bằng các enzym giới hạn) với đầu dò đặc hiệu của một gen cho phép đánh giá số bản copy của gen đó trong genome. Phép lai giữa ARNm với đầu dò cho phép xác định kích thước phân tử ARNm, định lượng số bản phiên mã (transcripts) và xác định hoạt động của gen ở các tổ chức biệt hoá.

5.9.1. Phương pháp Southern blots

Phương pháp lai Southern blots do Edward Southern phát minh nhằm phân biệt một đoạn ADN nhất định trong số các đoạn khác bị cắt bởi enzym giới hạn. Có thể gọi đoạn ADN cần tìm là đoạn ADN đích. Phương pháp này đòi hỏi đầu dò đặc hiệu lai với ADN đích. Đầu dò có thể là ADN hoặc ARN và có trình tự nucleotide tương đồng cao với ADN đích.

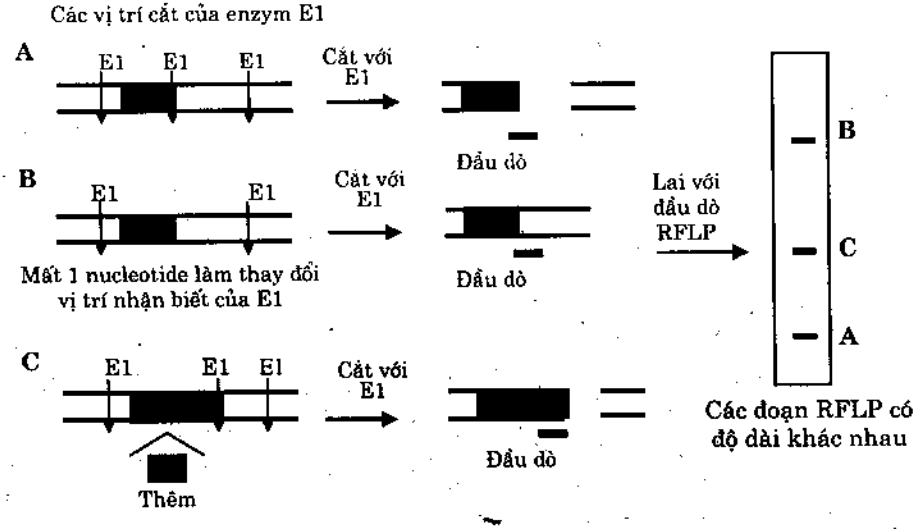
Đầu tiên ADN bị cắt bởi enzym giới hạn thành các đoạn có thể phân ly trên gel agarose. Sau đó các sợi kép ADN (trên gel) bị biến tính (do xử lý với kiềm) và được chuyển sang một màng (thường là màng nylon hoặc nitrocellulosse). Bước chuyển này được gọi là kỹ thuật blotting. Blotting có thể thực hiện bằng cách thẩm thấu mao dẫn (đòi hỏi nhiều thời gian-thường tiến hành qua đêm) (Hình 5.12A), hoặc sử dụng máy hút chân không (chỉ hết khoảng vài giờ). Sau đó màng chứa ADN được xử lý nhiệt (thường 2h ở 80°C), hoặc được chiếu tia tử ngoại (vài phút) để cố định ADN trên màng. Màng này được ủ trong dung dịch chứa đầu dò đặc hiệu đã đánh dấu. Vị trí của đoạn ADN lai với đầu dò được xác định nhờ phim hoặc hoá chất đặc biệt (Hình 5.12B). Phương pháp này rất nhạy cho phép phát hiện được đoạn ADN xuất hiện một lần (đơn bản) trong genome (~$1/10^6$) khi chỉ dùng 5μg ADN genome.



Hình 5.12: (A)-Blotting chuyển ADN từ gel agarose sang màng bằng phương pháp thẩm thấu mao dẫn. Sau khi chuyển, màng chứa ADN giống hệt như trên gel. Màng được xử lý nhiệt hoặc chiếu tia UV để cố định ADN. (B)- Kỹ thuật Southern blots sử dụng để phát hiện đoạn ADN mong muốn.

Lai Southern được ứng dụng vào việc đánh giá các bản copy của gen trong genome, xác định các exon, intron của một gen, phát hiện đột biến thêm, mất đoạn, xác định tính đa hình độ dài các đoạn cắt bởi enzym giới hạn.... Khi sử dụng các chỉ thị phân tử (marker ADN) khác nhau làm đầu dò, chúng ta có thể phân biệt được các loài, các loài đồng hình, các cá thể dưới loài.

Hình 5.13 mô tả ứng dụng của kỹ thuật lai Southern trong việc phát hiện tính đa dạng các đoạn ADN cắt bằng enzym giới hạn (Restriction Fragment Length Polymorphism - RFLP). Trong thực tế, khi so sánh genome của các thể trong một loài, chúng ta nhận thấy thay đổi trình tự nucleotide ở một số đoạn ADN không hề có biểu hiện tính trạng. Tuy nhiên, thay đổi đó có thể làm thay đổi vị trí nhận biết của enzym giới hạn. Ví dụ, thay đổi một nucleotide có thể làm mất vị trí nhận biết, do đó khi cắt bởi enzym thì thu được các đoạn ADN dài ngắn khác nhau. Sự tồn tại của những đoạn này trong tự nhiên tạo nên tính đa dạng các đoạn có chiều dài khác nhau, được gọi là RFLP.

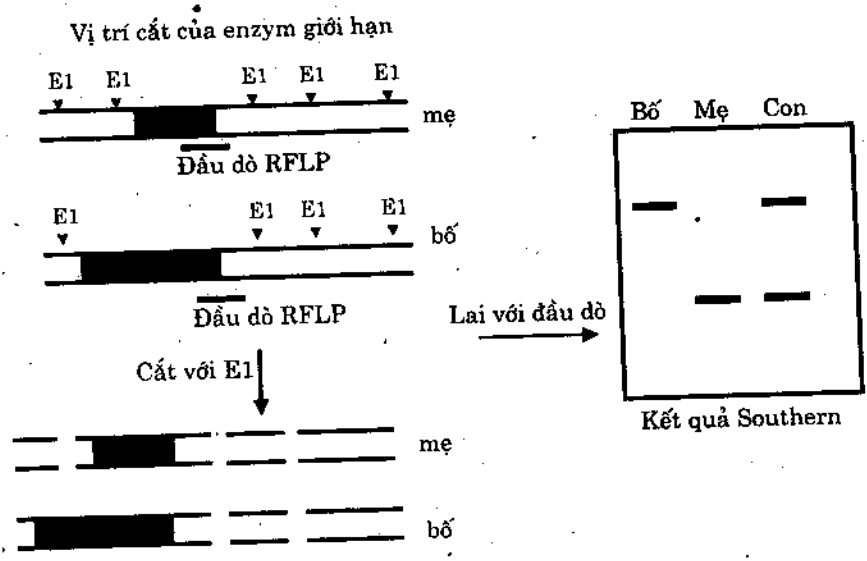


Hình 5.13: Tính đa dạng các đoạn ADN bị cắt bởi enzym giới hạn.

(A) Đoạn nhiễm sắc thể bị enzym giới hạn E1 cắt tại 3 vị trí tạo hai đoạn ADN. Một trong hai đoạn này được phát hiện nhờ phương pháp Southern blots. (B) Sự thay đổi của nucleotide tại một vị trí cắt khiến enzym không nhận biết được vị trí đó nữa. Chỉ có một đoạn ADN (lúc này được xem như một RFLP) được phát hiện qua phép lai. (C) Khi xảy ra đột biến thêm đoạn, đoạn ADN bị cắt dài hơn.

Khi một RFLP rất phổ biến trong quần thể, sự khác nhau giữa hai cá thể có thể xác định được nhờ RFLP đó. Vì vậy, các RFLP được xem là chỉ thị phân tử ADN để phân biệt sự đa dạng của genome. Chúng ta

có thể sử dụng RFLP để tạo đầu dò cho phép lai Southern. nhằm mục đích phân tích mối quan hệ huyết thống giữa các cá thể (Hình 5.14).



Hình 5.14: Kỹ thuật Southern blots sử dụng đầu dò RFLP trong phân tích quan hệ huyết thống.

Khi một RFLP phân bố cạnh một gen thì xác suất để RFLP phân ly cùng với gen trong phân bào giảm nhiễm phụ thuộc vào khoảng cách giữa RFLP với gen đó. Nếu RFLP và gen càng gần nhau, xác suất để chúng cùng được truyền cho một cá thể con cháu càng lớn. Do đó, RFLP được sử dụng làm đầu dò để phân lập gen đột biến. Thông thường để phân lập được gen liên quan đến bệnh di truyền, có thể dùng các marker RFLP để xác định vị trí của gen đó dựa vào phả hệ lớn gồm nhiều người. Sàng lọc các cá thể mà RFLP và tính trạng được thể hiện đồng thời. Rất nhiều gen gây bệnh ở người đã được tìm ra bằng phương pháp sử dụng RFLP.

5.9.2. Phương pháp northern blots

Các bước của phương pháp này tương tự như Southern blots nhưng ở đây ARNs được lai với đầu dò ADN hoặc ARN. Phương pháp này được áp dụng để xác định số lượng cũng như kích thước các phân tử ARNm có trong tế bào. Đầu tiên ARN tổng số được tách chiết khỏi tế bào. Sau đó, ARN được chuyển sang màng và được lai với đầu dò đặc hiệu. Đầu dò có thể là ARN hoặc ADNc. Kích thước cũng như số lượng ARNm lai với đầu dò được phát hiện dựa vào vị trí của băng lai và độ đậm nhạt của nó. Nhờ phương pháp này, có thể so sánh số lượng ARNm của một gen hoạt động bình thường và khi bị đột biến cũng như

nghiên cứu mức độ hoạt động của một gen ở các tổ chức biệt hoá khác nhau hay chịu các tác động bên ngoài.

Đối với những gen mà số lượng ARNm được tổng hợp ít, trước khi tiến hành lai Northern, có thể làm giàu ARNm bằng việc tách riêng chúng ra khỏi các loại ARN khác. Khi cho ARN tổng số qua cột gắn polyT, các ARNm mang đuôi polyA sẽ được giữ lại trên cột trong khi các ARN khác bị loại. Rửa cột bằng dung dịch thích hợp để thu ARNm.

5.9.3. Kỹ thuật lai tại chỗ (lai *insitu*)

Phát triển từ phép lai Southern blots, kỹ thuật lai *insitu* sử dụng các đầu dò đánh dấu (đoạn nucleotide hoặc kháng thể) để lai với ADN, ARN hoặc với protein ở ngay trong tế bào mà không cần tách chiết chúng. Các đầu dò oligonucleotide thường được đánh dấu bằng phương pháp hoá học (chất phát huỳnh quang). Chúng được tổng hợp từ các nucleotide mang những liên kết hoá học đặc biệt. Các đầu dò sau khi lai được phát hiện nhờ các protein khác nhận biết các liên kết hoá học này.

Sự phân bố các phân tử ARNm trong tế bào được phát hiện nhờ phản ứng lai *insitu*. Khi nghiên cứu với ARN, nhiễm sắc thể không ủ ở pH cao, do đó chúng giữ nguyên ở trạng thái sợi đôi và không lai được với mồi. Nhờ phương pháp này mà các biểu hiện của gen ở mức độ phiên mã có thể quan sát được trong tế bào. Ví dụ, với mô phôi ruồi giấm *Drosophila*, các tế bào ở những vị trí khác nhau trong quá trình phát triển cá thể được phát hiện nhờ dựa vào kiểu cách hoạt động của các gen liên quan.

5.10. XÁC ĐỊNH TRÌNH TỰ NUCLEOTIDE

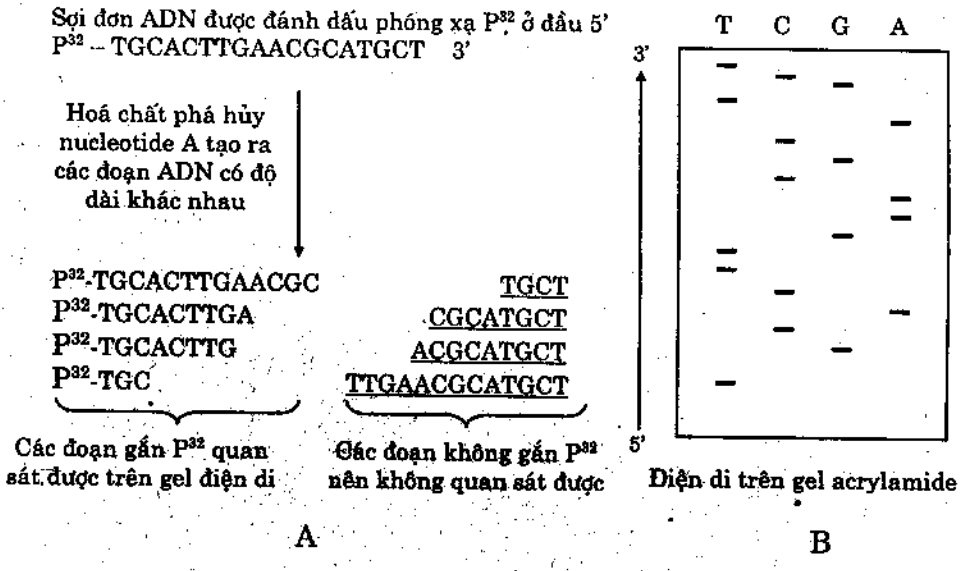
Từ cuối những năm 1970, trình tự các nucleotide trên phân tử ADN được xác định một cách đơn giản và nhanh chóng nhờ sự ra đời của hai phương pháp khác nhau: phương pháp hoá học và phương pháp enzym. Tuy nhiên, từ những năm cuối của thập kỷ 20 thế kỷ XX, trình tự nucleotide được xác định trên máy đọc tự động ngày càng phổ biến. Ưu điểm của phương pháp này là giảm các thao tác, tiết kiệm hoá chất và trình tự đọc được dài hơn hẳn so với các phương pháp trước đó.

Các phương pháp xác định trình tự đều liên quan đến kỹ thuật điện di các đoạn ADN dạng sợi đơn trên gel polyacrylamide. Đặc điểm của những đoạn ADN này là có kích thước sai khác nhau 1 nucleotide. Trên các máy xác định trình tự hiện đại, gel acrylamide được nhồi trong các mao quản và các thông số của quá trình điện di như điện thế, nhiệt độ, thời gian cũng như lưu cất dữ liệu... được kiểm soát tự động.

5.10.1. Phương pháp hoá học Maxam-Gilbert

Phương pháp này sử dụng các chất hoá học cắt đứt sợi ADN tại một loại nucleotide đặc hiệu (Hình 5.15). Đầu tiên phân tử ADN dạng kép được đánh dấu P^{32} tại đầu 5'. Sau đó hai sợi đơn tách rời nhau do biến tính nhiệt. Chúng được xử lý với các chất hoá học sao cho chỉ một loại nucleotide bị phá hủy. Trên hình vẽ đó là nucleotide A. Nồng độ các chất hoá học được kiểm tra chặt chẽ sao cho chỉ một nucleotide bị phá hủy trên mỗi sợi đơn ADN. Kết quả hàng loạt các đoạn ADN có kích thước khác nhau được tạo thành.

Các đoạn ADN tạo ra do bị bẻ gãy được phân ly trên gel polyacrylamide và chỉ có đoạn nào bị gắn P^{32} ở đầu 5' mới quan sát được. Kích thước của chúng tương ứng với khoảng cách từ đầu 5' có đánh dấu đến vị trí của các nucleotide bị bẻ gãy trên phân tử ADN. Sử dụng các chất hoá học khác nhau phá hủy 4 loại nucleotide bằng 4 phản ứng riêng biệt. Sản phẩm thu được của 4 phản ứng này được phân ly cùng nhau trên gel. Đoạn ADN ngắn nhất có đánh dấu phóng xạ chạy nhanh nhất dưới tác dụng của điện trường. Căn cứ vào vị trí các vạch quan sát được trên gel mà xác định trình tự các nucleotide (Hình 5.15).

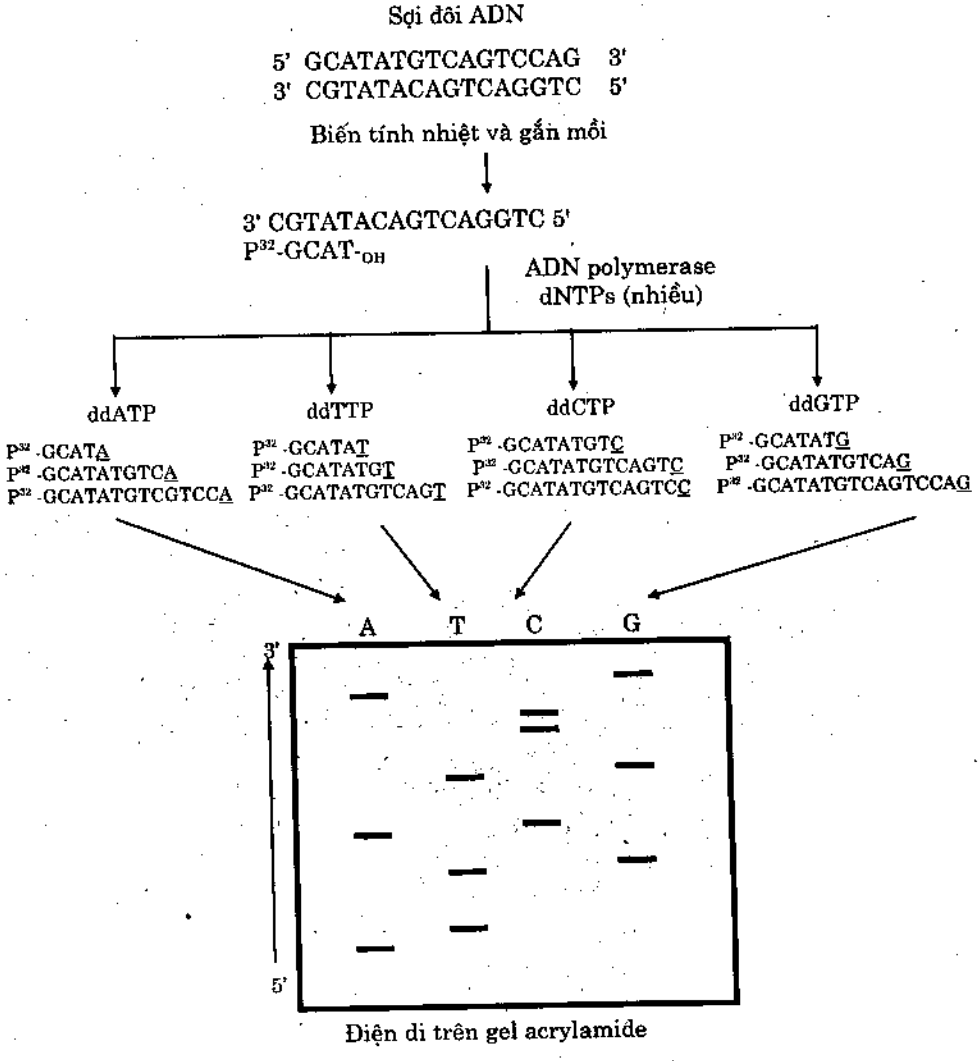


Hình 5.15: Xác định trình tự nucleotide bằng phương pháp hoá học.

(A) Tiến hành 4 phản ứng độc lập, sử dụng 4 hoá chất khác nhau, mỗi hoá chất chỉ phá hủy một loại nucleotide nhất định (trong hình vẽ đó là A (adenine)). (B) Sản phẩm của 4 phản ứng được chạy điện di đồng thời trên gel polyacrylamide. Trình tự nucleotide đọc từ dưới lên theo chiều 5'→3'.

5.10.2. Phương pháp enzym Sanger (phương pháp tạo đầu tận cùng của chuỗi- chain terminator method)

Chìa khoá của phản ứng này là sử dụng dideoxynucleotide không có nhóm OH ở vị trí 3' trong phản ứng tổng hợp ADN. Do đó khi enzym ADN polymerase gắn chúng vào sợi ADN thì quá trình tổng hợp bị ngừng lại. Vì vậy phương pháp này còn được gọi là phương pháp dideoxy (dideoxy method).



Hình 5.16: Xác định trình tự nucleotide bằng phương pháp enzym (phương pháp Sanger). Dideoxyribonucleoside triphosphate (ddNTP) gắn vào sợi ADN đang tổng hợp sẽ làm ngừng phản ứng. Nhờ đó thu được các đoạn ADN hơn kém nhau chỉ một nucleotide. Các đoạn này được quan sát trên gel acrylamide khi mồi được đánh dấu phóng xạ P^{32} hoặc đánh dấu huỳnh quang.

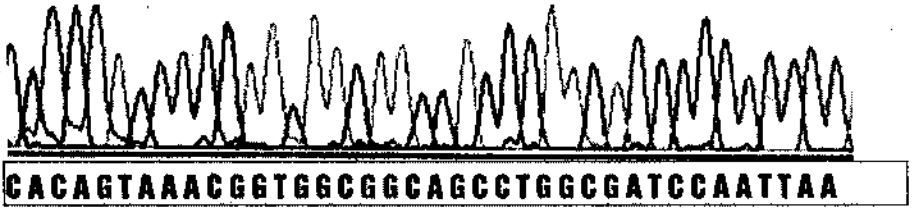
Đầu tiên, sợi đôi ADN (sợi khuôn cần đọc trình tự) bị biến tính tách thành hai sợi đơn và được lai với mồi. Mồi là một sợi đơn gồm vài chục nucleotide có thể liên kết với một trong hai sợi đơn ADN theo nguyên tắc tạo cặp bổ sung. Sau đó 4 phản ứng riêng rẽ được tiến hành đồng thời. Mỗi phản ứng là hỗn hợp của ADN khuôn, mồi, enzym ADN polymerase, một loại dideoxynucleotide và 4 loại deoxynucleotide theo tỷ lệ 1:100. Các sợi đơn ADN được tổng hợp có độ dài khác nhau do dideoxynucleotide gắn ngẫu nhiên làm dừng phản ứng. Sản phẩm của 4 phản ứng cùng được phân ly trên gel polyacrylamide cho phép phân biệt hai sợi đơn ADN hơn kém nhau một nucleotide. Các vạch ADN quan sát được nhờ có gắn phóng xạ P^{32} vào mồi hoặc vào một trong bốn loại nucleotide bình thường (Hình 5.16).

5.10.3. XÁC ĐỊNH TRÌNH TỰ NUCLEOTIDE TRÊN MÁY TỰ ĐỘNG

Trong những năm gần đây, một số phương pháp xác định trình tự mới đã xuất hiện. Bên cạnh kỹ thuật thông thường sử dụng các gel để phân ly các phân tử ADN có độ dài khác nhau, các kỹ thuật mới liên quan đến phát hiện huỳnh quang của các nucleotide bị đánh dấu, phân tích trình tự ADN bằng khối phổ hoặc lai với các đoạn oligonucleotide được tổng hợp nhân tạo. Những tiến bộ nhanh chóng trong lĩnh vực kính hiển vi điện tử cho phép quan sát chi tiết cấu trúc bề mặt của phân tử acid nucleic và phân biệt từng nucleotide. Ngoài ra, kỹ thuật lai sử dụng tấm chip gắn cố định các oligonucleotide cho phép phân biệt được hơn 50000 phân tử ADN khác nhau. Trong tương lai gần, kỹ thuật lai này có thể gắn hàng nghìn oligo trên một tấm chip nhỏ và trình tự nucleotide được phân tích tự động bằng các chương trình phần mềm máy tính. Điều đó cho phép xác định trình tự một cách nhanh chóng.

Phương pháp tạo đầu tận cùng của chuỗi được cải tiến cho phép xác định trình tự trên máy đọc tự động (automatic sequencing), sử dụng các KIT khác nhau, trong đó đầu tận cùng được đánh dấu bằng các chất phát huỳnh quang màu (dye terminator sequencing). Phản ứng gắn các nucleotide đánh dấu vào sợi ADN tổng hợp trên sợi khuôn được tiến hành trên máy PCR. Vì vậy, kỹ thuật này còn được gọi là chuỗi đọc trình tự (cycle sequencing). Sau đó, sản phẩm PCR được điện di trên gel acrylamide có độ phân giải cao, cho phép phân biệt được các

sợi đơn ADN hơn kém nhau 1 nucleotide. Quá trình chạy điện di được thực hiện trên máy tự động và kết quả được phân tích bằng các chương trình máy tính chuyên dụng.



Hình 5.17: Trình tự nucleotide xác định trên máy tự động (Perkin-Elmer-377).

Mỗi nucleotide được biểu hiện bởi một đỉnh, tương ứng với một màu. Một phản ứng đọc tự động có khả năng xác định được khoảng 1000 nucleotide. Trình tự này được phân tích bằng các chương trình phần mềm trên máy tính nối trực tiếp với máy đọc tự động.

Trên các máy đọc tự động hiện đại, thông thường bốn loại nucleotide được đánh dấu bằng các chất phát huỳnh quang khác nhau. Đầu đọc laser trong máy sẽ kích hoạt các chất này khiến chúng hấp thụ và phát ra các màu, mỗi màu ứng với một loại nucleotide và được hiển thị bằng một đỉnh (Hình 5.17). Thông thường, phản ứng gắn nucleotide vào sợi ADN tổng hợp trên sợi khuôn được thực hiện trên máy PCR. Hai loại nucleotide dGTP và dTTP được thay thế bởi dITP và dUTP nhằm hạn chế sự trùng lặp các đỉnh. Sản phẩm PCR thường được tinh sạch, loại bỏ các nucleotide và primer thừa trước khi đưa vào máy đọc tự động. Trình tự đọc được trên máy tự động thường dài khoảng 600 đến 1000bp. Ưu điểm của phương pháp này là kết quả đọc được đưa trực tiếp vào chương trình máy tính, giảm bớt các sai sót do đọc bằng mắt gây ra. Hơn nữa, phản ứng PCR không đòi hỏi lượng ADN khuôn nhiều nhưng các sợi đơn ADN cần đọc được khuếch đại lên rất nhiều lần.

Trong các chương trình nghiên cứu xác định trình tự hệ gen của một số đối tượng cụ thể như dự án gen người, arabidopsis v.v..., genome được cắt từng phần bởi enzym giới hạn thành các đoạn có kích thước lớn có các đầu nằm gối lên nhau (xem phần ngân hàng genomic DNA). Chúng được đưa vào vector (thường là phagemid hoặc BAC) và được lấy ra một cách ngẫu nhiên để xác định trình tự. Số liệu được phân tích trên máy tính để tìm ra các đoạn trùng với nhau. Bất kỳ

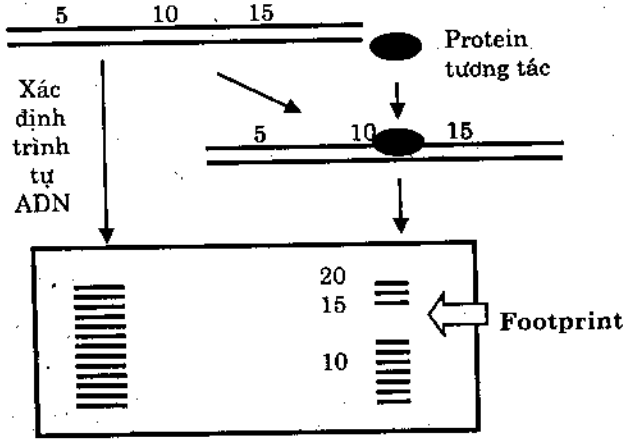
khoảng trống (gap) nào giữa các đoạn đã biết sẽ được xác định khi thiết kế primer dựa vào trình tự ở hai đầu tận cùng của khoảng trống.

Trình tự hệ gen cũng như các loại ADN (các đoạn ADN genome, ADNc, các đoạn đích ADNc-expressed sequence tags-EST ...) được xác định ngày càng nhiều ở các sinh vật khác nhau. Do đó việc lưu trữ số liệu, phân tích, so sánh và sắp xếp chúng đã thúc đẩy sự hình thành một hướng nghiên cứu mới, tạm dịch là tin sinh học (bioinformatics). Mối liên quan mật thiết giữa trình tự nucleotide và protein đã khiến lĩnh vực mới này phát triển rất nhanh chóng. Ba ngân hàng dữ liệu chính hiện nay lưu trữ hầu hết các thông tin về ADN là EMBL (thuộc Viện Tin học châu Âu-European Informatics Institude), GenBank (thuộc Trung tâm Công nghệ Sinh học của Mỹ-US National Centre for Biotechnology) và DDBS (thuộc Ngân hàng dữ liệu ADN của Nhật-DNA Database of Japan).

5.11. MỘT SỐ KỸ THUẬT XÁC ĐỊNH TƯƠNG TÁC GIỮA ADN VÀ PROTEIN

5.11.1. Phương pháp "DNA footprint"

Phương pháp này cho phép xác định chính xác các nucleotide được protein bảo vệ che chắn khỏi tác dụng phân hủy của ADNase (Hình 5.18).



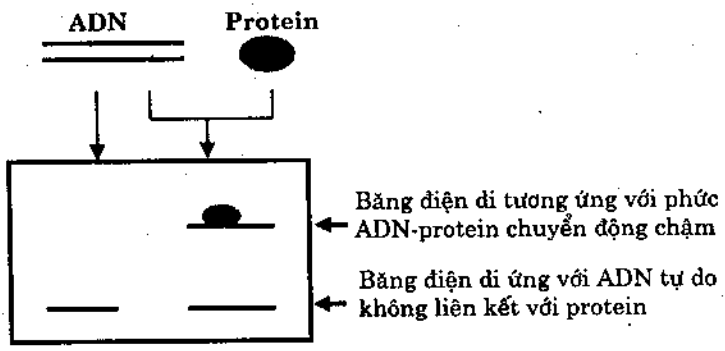
Hình 5.18: Phương pháp "DNA footprint" xác định vị trí tương tác của protein với ADN

Khi liên kết với các nucleotide, protein bảo vệ cho chúng khỏi tác dụng của nuclease. Do đó nuclease chỉ cắt ADN ở các điểm nằm ngoài vị trí liên kết, tạo ra các đoạn ADN khác nhau. Trên gel thu được vệt trống "DNA footprint" tương ứng với vị trí được bảo vệ bởi liên kết ADN-protein.

Nhờ đó khi điện di trên gel acrylamide, các đoạn ADN tương ứng với vị trí tương tác không xuất hiện trên gel, tạo ra những vết trống "footprint". Phương pháp này tương tự như kỹ thuật xác định vị trí tương tác của ARN polymerase trên sợi ADN khuôn trong quá trình tổng hợp ARN.

5.11.2. Phương pháp xác định băng điện di chậm (gel retardation assay)

Phương pháp này cho phép xác định phức ADN hoặc ARN liên kết với protein khi so sánh tốc độ di động của ADN, hoặc ARN tự do với phức ADN-protein hoặc ARN-protein dưới tác dụng của điện trường không đổi. Do tương tác với protein nên phức chuyển động chậm (vì vậy gọi là gel retardation). Tiếp sau đó, protein liên kết với ADN, hoặc ARN có thể được xác định bằng các kháng thể. Khi có mặt kháng thể, tốc độ di động của phức kháng thể-protein-ADN càng bị chậm hơn. Đoạn nucleotide tại vị trí tương tác cũng được xác định chính xác khi sử dụng các oligonucleotide cạnh tranh (competor oligonucleotide). Nếu chuỗi oligonucleotide cạnh tranh có trình tự bổ sung với các base trên sợi ADN hoặc ARN, chuỗi sẽ cạnh tranh với protein trong phản ứng liên kết với ADN hay ARN. Do đó, protein không thể liên kết với ADN hoặc ARN nữa. Trên gel không quan sát được những băng di động chậm (Hình 5.19).



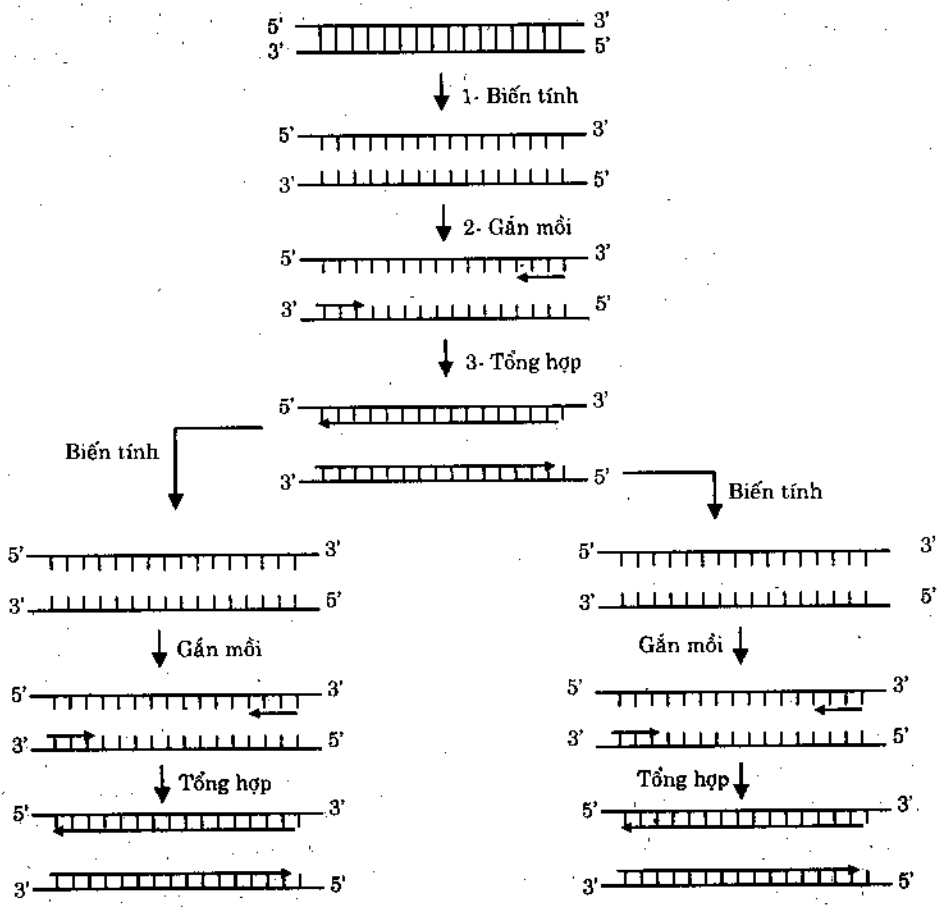
Hình 5.19: Phương pháp xác định băng điện di chậm cho phép phân biệt đoạn ADN tự do và phức ADN-protein.

Khi có liên kết với protein, băng ADN chuyển động trên gel chậm hơn so với băng ADN tự do không liên kết.

5.12. PHẢN ỨNG PCR (POLYMERASE CHAIN REACTION)

Một đoạn ADN bất kỳ có thể được nhân lên nhanh chóng hàng tỷ lần mà không cần đưa vào tế bào vi khuẩn. Điều đó xảy ra trong phản ứng tổng hợp *in-vitro* sử dụng ADN polymerase và các oligonucleotide (các mồi-primers). Đó chính là kết quả tuyệt vời của phản ứng PCR, tạm dịch là chuỗi phản ứng tổng hợp. Kỹ thuật PCR được Kary Mullis tìm ra đã giúp các nghiên cứu về ADN nói riêng, về sinh học phân tử nói chung đạt được những thành tựu đáng kinh ngạc. Nhờ phản ứng này, một đoạn ADN ở một vùng bất kỳ trong genome được khuếch đại lên rất nhiều lần khi trình tự nucleotide ở hai đầu đoạn ADN đó đã biết. Dựa vào trình tự đó, các cặp oligonucleotide được tổng hợp nhân tạo, mỗi oligo tạo liên kết bổ sung với từng sợi đơn. Chúng được sử dụng làm mồi để tổng hợp ADN *invitro* nhờ enzym ADN polymerase. Đặc biệt phản ứng PCR chỉ đòi hỏi một lượng ADN làm khuôn ban đầu rất nhỏ (cỡ 10^{-3}μg).

Một phản ứng PCR gồm nhiều chu kỳ. Mỗi chu kỳ phản ứng đòi hỏi ba bước (Hình 5.20). Đầu tiên, ADN khuôn được xử lý nhiệt độ gây biến tính thành hai sợi đơn. Bước tiếp theo, nhiệt độ giảm xuống cho phép các mồi (số lượng rất lớn) liên kết tạo cặp bổ sung với hai sợi đơn ADN khuôn. (Do số lượng mồi lớn nên liên kết giữa mồi và ADN khuôn xảy ra chiếm ưu thế so với liên kết phục hồi giữa hai sợi đơn). Trong bước thứ ba, hỗn hợp ADN và mồi được ủ với enzym Taq polymerase và bốn loại deoxyribonucleotide triphosphate. Đoạn ADN nằm giữa hai mồi được tổng hợp. Khi các chu kỳ được lặp lại, các đoạn ADN vừa được tổng hợp ở các chu kỳ trước trở thành khuôn mẫu cho chu kỳ sau. Số lượng phân tử ADN tạo ra tăng gấp đôi sau mỗi chu kỳ. Chỉ sau vài chu kỳ đầu, sản phẩm tạo ra chủ yếu là các đoạn ADN có kích thước bằng đúng khoảng cách giữa hai mồi. Số chu kỳ lặp lại trong mỗi phản ứng PCR thường là không quá 50 và mỗi chu kỳ thường kéo dài khoảng 5 phút. Khi số chu kỳ lớn hơn 50, nồng độ dNTP, mồi và hoạt tính của Taq bị suy giảm nghiêm trọng trong khi nồng độ sợi khuôn tăng cao. Điều đó ức chế phản ứng PCR.



Hình 5.20: Phản ứng PCR.

Hai chu kỳ đầu tiên của phản ứng được mô tả trên hình vẽ. Từ một phân tử ADN làm khuôn, sau mỗi chu kỳ gồm 3 bước: biến tính (1), gắn mồi (2) và tổng hợp (3) thu được 4 phân tử. Sau một số chu kỳ, số lượng sợi ADN có kích thước duy nhất chiếm ưu thế tuyệt đối trong phản ứng.

$$N = (2^n - 2n).x$$

n: Số chu kỳ; 2n: Số phân tử có chiều dài không xác định; x: Số phân tử khuôn ban đầu; N: Số sợi tổng hợp được trong phản ứng; 2^n: số phân tử có chiều dài bằng khoảng cách giữa hai mồi.

5.12.1. Một số yếu tố ảnh hưởng đến phản ứng PCR

a) Các mồi (primer)

Mồi dùng trong phản ứng PCR thường có chiều dài từ 10 đến 40 nucleotide, tuỳ thuộc vào mức độ khuếch đại đặc hiệu đối với trình tự

đích (trình tự cần khuếch đại). Việc lựa chọn trình tự các mồi trên một đoạn ADN (một gen) có tính chất quyết định đối với phản ứng PCR. Trong đa số trường hợp, mồi dùng cho PCR nên thoả mãn một số yêu cầu như sau:

– Tỷ lệ GC ở các mồi thường chiếm khoảng 40-75% để đảm bảo liên kết gắn mồi với khuôn bền vững. Tuy nhiên, trình tự dinucleotie GC không lặp lại quá nhiều trong một mồi.

– Khoảng cách giữa hai mồi thường không dài quá 3kb.

Các mồi không có các nucleotide bổ sung để tránh hiện tượng hai mồi bắt cặp với nhau. Hơn nữa, trình tự của từng mồi không được tạo cấu trúc cặp tóc do liên kết giữa các nucleotide ngay trong một mồi.

– Nhiệt độ Tm của hai mồi không quá chênh lệch nhau nhằm đạt được sự bắt cặp tối ưu giữa các mồi với khuôn ở cùng một nhiệt độ gắn mồi.

– Sợi ADN được tổng hợp khi đầu 3' của mồi đã liên kết với khuôn mặc dù đầu 5' có thể tự do. Do đó đầu 5' của mồi có thể được gắn thêm linker, hoặc đoạn nucleotide điều khiển (promoter). Nồng độ mồi khoảng 0,1μM đến 0,5μM. Nồng độ thường được xác định dựa vào giá trị OD đo ở 260nm.

b) Liên kết giữa mồi và ADN khuôn

Mồi gắn vào sợi khuôn ADN nhờ liên kết tạo cặp bổ sung giữa chúng. Liên kết này phụ thuộc thời gian, nhiệt độ, nồng độ mồi cũng như nồng độ sợi khuôn. Khi nồng độ mồi rất lớn so với nồng độ khuôn, thời gian gắn mồi với sợi khuôn không cần thiết quá lâu. Nhiệt độ ở giai đoạn này phụ thuộc chủ yếu vào chiều dài của mồi và hàm lượng GC. Khi oligo mồi gồm 12 đến 15 nucleotide, nhiệt độ gắn mồi khoảng 40-45^{0}C.

Nhiệt độ 72°C là nhiệt độ thích hợp đối với hoạt tính của Taq polymerase. Đây là enzym ADN polymerase có hoạt tính tổng hợp ADN và bền ở nhiệt độ cao. Khi nhiệt độ thay đổi từ giai đoạn gắn mồi lên 72°C ở giai đoạn tổng hợp; các mồi ngắn thường bị biến tính tách ra khỏi sợi khuôn. Ngoài ra, trong trường hợp mồi được tổng hợp xuất phát từ trình tự acid amin thì liên kết giữa mồi với ADN khuôn không hoàn toàn bổ sung do tính chất thoái hoá của mã di truyền. Do đó, mồi loại này dễ dàng tách ra khỏi sợi khuôn ở nhiệt độ tổng hợp ADN 72°C.

Vì vậy trong trường hợp này, thời gian gắn mồi có thể kéo dài hơn so với mồi có trình tự xuất phát trực tiếp từ đoạn cần nhân bản. Rõ ràng nhiệt độ và thời gian ở giai đoạn gắn mồi đặc biệt quan trọng; nhiệt độ quá cao gây biến tính, nhiệt độ quá thấp tạo ra các liên kết không đặc hiệu (mồi có thể bám vào vị trí bất kỳ có độ tương đồng chứ không đảm bảo chính xác). Ngoài ra khi ADN khuôn chiếm tỷ lệ rất nhỏ (ví dụ một đoạn ADN trong genome), thời gian cần thiết để mồi tìm đúng vị trí liên kết đặc hiệu và tạo cặp với khuôn thường khoảng 2 phút trong một số chu kỳ đầu tiên. Điều đáng lưu ý, khi chỉ vài nucleotide ở đầu 3' của mồi đã liên kết với khuôn thì Taq polymerase sẽ kéo dài sợi mồi thành sợi ADN mới bất chấp mồi chưa liên kết hoàn toàn với khuôn ở đầu 5'.

c) ADN khuôn

Phản ứng PCR có ưu điểm là chấp nhận ADN khuôn không tinh sạch (các mẫu hình sự), không đòi hỏi điều kiện bảo quản lạnh (các mẫu sinh phẩm khô), lượng ADN cần rất ít. ADN rất đa dạng, có thể là ADN genome tách ra từ tế bào, ADN từ các ngân hàng genome hoặc ADNc, các ADN bất kỳ; ARN tổng số hoặc ARNm. Nồng độ ADN khuôn thường < ng với ADNc, hoặc <μg với ADN genome.

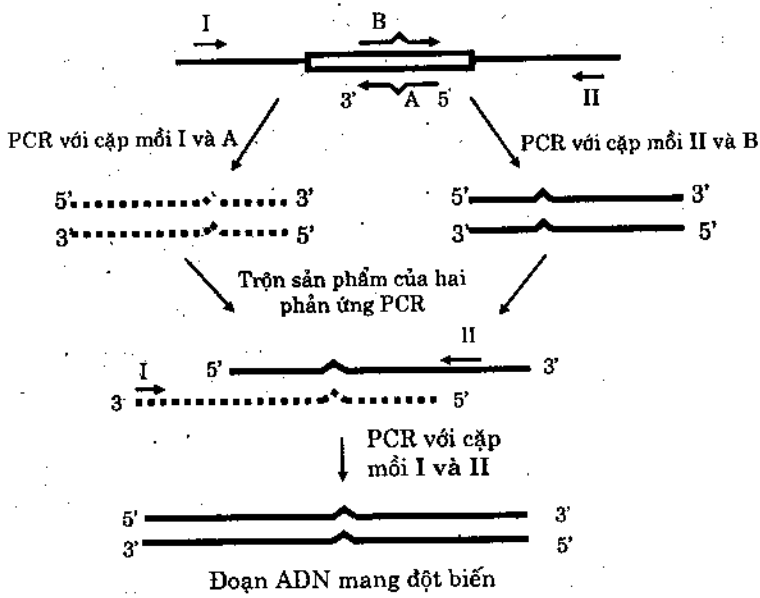
d) Nồng độ một số ion

Trong môi trường, một số ion có ảnh hưởng đến phản ứng PCR. Ví dụ, ion Mg^{+2} kích thích phản ứng tổng hợp ADN, trong khi KCl ức chế quá trình khuếch đại đoạn ADN dài.

5.12.2. Một số ứng dụng của phản ứng PCR

a) RT-PCR (Reverse transcriptase-PCR)

Phản ứng này được tiến hành nhằm nhân bản các ADNc từ khuôn ban đầu là ARNm. Đầu tiên, ARN tổng số hoặc ARNm được làm khuôn để tổng hợp các ADNc sợi đơn nhờ enzym phiên mã ngược Reverse transcriptase. Tiếp đến các ADNc này được nhân bản trong phản ứng PCR với cặp mồi đặc hiệu. Như vậy, chúng ta không nhất thiết phải sàng lọc ADNc từ ngân hàng cũng thu được một phân tử ADNc đặc hiệu cho một gen. Đối với một cặp mồi đặc hiệu cho một gen, số lượng sản phẩm RT-PCR ứng với các mẫu ARN từ các loại tế bào khác nhau cho phép so sánh mức độ phiên mã của gen đó. RT-PCR được xem là phản ứng bán định lượng đối với ARNm.



Hình 5.21: Kết hợp các phản ứng PCR để thay thế các nucleotide trong một đoạn ADN đã biết.

Mồi thiết kế (mồi A và B) mang các nucleotide thay thế. Kỹ thuật này cho phép gây đột biến tại vị trí mong muốn trên một đoạn ADN (gen).

b) RAPD-PCR (Random Amplified Polymorphics DNA -PCR)

Kỹ thuật này cho phép nhân bản ngẫu nhiên các đoạn ADN có khả năng bắt cặp với mồi mà không đòi hỏi sự chính xác đặc hiệu. Do đó, nếu hai mẫu ADN khuôn có nhiều trình tự tương đồng thì sau phản ứng RAPD-PCR, chúng ta sẽ thu được các đoạn ADN có kích thước giống nhau. Do không yêu cầu tính đặc hiệu nên mồi dùng trong phản ứng này thường là mồi chỉ có 10 nucleotide và nhiệt độ gắn mồi không cao. Kỹ thuật RAPD-PCR thường được ứng dụng trong phân tích quan hệ họ hàng, xây dựng cây phân loài.

c) PCR tạo đột biến định hướng

Được áp dụng để xây dựng phân tử ADN tái tổ hợp, đưa vào, hoặc bỏ đi, hoặc thay thế các nucleotide trong một đoạn ADN. Các nucleotide cần gây đột biến được thiết kế nằm trong trình tự của mồi. Kỹ thuật này phối hợp sản phẩm của các phản ứng PCR với nhau như mô tả cụ thể trong Hình 5.21.

5.13. MỘT SỐ ỨNG DỤNG CỦA SINH HỌC PHÂN TỬ

5.13.1. Sản xuất vaccine tái tổ hợp

Vaccine được đưa vào cơ thể nhằm mục đích làm tăng khả năng đáp ứng miễn dịch, đặc biệt với các tác nhân gây bệnh vi sinh vật. Những vaccine kinh điển được sản xuất dựa trên vi khuẩn, virus đã bị chết hoặc độc lực của chúng bị suy giảm để không gây bệnh. Tuy nhiên, những vaccine kinh điển này vẫn có một tỷ lệ nhỏ tác hại không mong muốn cho người sử dụng. Vì vậy, vaccine thế hệ mới sử dụng công nghệ ADN tái tổ hợp để thu nhận vaccine có hoạt tính đặc hiệu cao và đảm bảo an toàn. Công nghệ ADN tái tổ hợp sử dụng vào sản xuất vaccine gồm những bước cơ bản như sau:

– Đưa gen mã cho kháng nguyên của vi khuẩn hoặc virus gây bệnh vào vector biểu hiện có promoter mạnh; rồi đưa vector vào tế bào nhận. Tế bào nhận có thể là prokaryot hoặc tế bào động vật nuôi cấy. Thu nhận và tinh sạch kháng nguyên từ những tế bào này để sản xuất vaccine.

– Gây đột biến, hoặc loại bỏ các gen quy định độc tính của chủng vi khuẩn hoặc virus gây bệnh. Nhờ đó những chủng đột biến không còn nguy hiểm nhưng vẫn có khả năng tạo kháng nguyên đặc hiệu cao. Chúng được sử dụng để sản xuất vaccine.

5.13.2. Sản xuất các hợp chất có hoạt tính sinh học phục vụ và nâng cao chất lượng cuộc sống

– Sản xuất kháng sinh: Nguồn kháng sinh trong tự nhiên chủ yếu được tìm thấy ở vi sinh vật. Quá trình chọn lọc môi trường nuôi cấy cũng như sàng lọc các chủng vi sinh vật cho kháng sinh đặc hiệu và có năng suất cao rất tốn công sức và chi phí. Kỹ thuật tách dòng ADN cho phép tạo ra toàn bộ các dòng gồm những gen tham gia vào quá trình tổng hợp kháng sinh. Nhờ vậy, chúng ta kiểm soát và điều chỉnh một cách chủ động toàn bộ quy trình sản xuất kháng sinh ở quy mô công nghiệp.

– Sản xuất hoocmon: Một số bệnh di truyền gây ra do đột biến ở gen mã cho hoocmon gây thiếu hụt hoocmon. Kỹ thuật gen đã phân lập ADNc của những gen mã cho các hoocmon đó và đưa chúng vào vector có promoter mạnh. Vector tái tổ hợp được biến nạp vào tế bào nhận; từ đó các protein tương ứng được tách chiết và tinh sạch.

– Sản xuất kháng thể đơn dòng: Kháng thể đơn dòng có rất nhiều ứng dụng trong chẩn đoán và điều trị bệnh, đặc biệt là điều trị ung thư. Chúng ta có thể phân biệt tế bào ung thư với tế bào lành nhờ kháng nguyên trên bề mặt tế bào ung thư. Kháng thể đơn dòng đặc hiệu với kháng nguyên được sản xuất nhờ kỹ nghệ ADN tái tổ hợp. Kháng thể này được ghép nối với độc tố, hoặc với đồng vị phóng xạ rồi được đưa vào cơ thể. Nhờ kháng thể chỉ liên kết đặc hiệu với kháng nguyên trên bề mặt tế bào ung thư nên chỉ những tế bào này bị tiêu diệt.

5.13.3. Chẩn đoán mầm bệnh

– Chẩn đoán bệnh truyền nhiễm: Các kỹ thuật kinh điển trong chẩn đoán bệnh do vi sinh vật gây ra chủ yếu dựa vào nuôi cấy tế bào hay xét nghiệm sinh hoá. Điều này đòi hỏi thời gian thao tác và mật độ vi sinh vật trong mẫu bệnh phẩm phải đạt đến một số lượng nhất định. Hiện nay, kỹ thuật PCR, kỹ thuật lai ADN được áp dụng để xét nghiệm nhanh, chính xác sự có mặt của vi sinh vật gây bệnh. Có thể nói, kỹ thuật PCR là xét nghiệm không thể thiếu khi cần chẩn đoán các bệnh truyền nhiễm ở các phòng xét nghiệm hiện đại. Điều cần lưu ý, do độ nhạy của PCR rất cao nên các thao tác thực hiện phải đảm bảo không bị nhiễm nguồn gen lạ với mẫu bệnh phẩm.

– Chẩn đoán các đột biến gen và tư vấn trước sinh: Các bệnh di truyền gây ra do đột biến gen hầu hết là đột biến lặn và thường để lại hậu quả rất nặng nề cho gia đình và xã hội. Một số bệnh di truyền do thiếu hụt hoocmon hay enzym chuyển hoá có thể khắc phục được nếu như được chẩn đoán sớm. Ngoài ra, một số bệnh di truyền chưa có biện pháp khắc phục giảm thiểu hậu quả. Vì vậy, chẩn đoán đột biến gen ở giai đoạn thai nhi sớm rất có ý nghĩa trong tư vấn di truyền trước sinh đối với các gia đình có tiền sử bệnh.

5.13.4. LIỆU PHÁP GEN

Liệu pháp gen được hiểu là thay thế gen hỏng (gây bệnh) bằng gen lành. Liệu pháp gen có thể thực hiện ở hai mức độ: thực hiện với tế bào soma (tế bào sinh dưỡng) hoặc trên tế bào sinh dục.

Liệu pháp gen thực hiện với tế bào soma người được thực hiện lần đầu tiên vào năm 1990. Một bé gái 4 tuổi mắc bệnh không có khả năng đáp ứng miễn dịch do đột biến gen mã cho enzym adenosine deaminase

(ADA). Khi không có enzym này, tế bào lympho T bị chết do cơ chất tích tụ. Tế bào bạch cầu của bé gái được nuôi cấy và nhận gen lành mã cho enzym. Sau đó những tế bào chuyển gen được truyền cho bé gái. Em bé gái này đã nhận dòng tế bào của chính mình, những tế bào có khả năng tổng hợp enzym ADA. Nhờ đó khả năng đáp ứng miễn dịch của bé gái trở nên bình thường.

Tiếp sau đó, nhiều bệnh nhân của căn bệnh này cũng được điều trị bằng liệu pháp gen. Do thời gian sống của tế bào bạch cầu ngắn nên bệnh nhân phải được truyền tế bào liên tục. Để khắc phục nhược điểm này, tế bào mầm tuỷ xương được lấy ra từ bệnh nhân và được chuyển gen ADA. Sau đó những tế bào mầm đã nhận gen được đưa trở lại bệnh nhân. Đây là những tế bào có khả năng sản sinh ra các tế bào bạch cầu mới. Liệu pháp gen với tế bào mầm đã được áp dụng thành công trong điều trị cho 2 bệnh nhi vào năm 1993.

5.13.5. Chuyển gen tạo các động, thực vật có năng suất chất lượng cao

– Động vật chuyển gen: Chuột chuyển gen là động vật mô hình sử dụng trong nghiên cứu phục vụ y học như tác dụng của thuốc, miễn dịch, cấy ghép tạng, tìm hiểu cơ chế gây bệnh, các liệu pháp gen... Ngoài ra, một số động vật như bò, cừu, cá... chuyển gen thể hiện những tính trạng mong muốn như chất lượng thịt và sữa tốt, sản lượng lông tăng cao, tăng trưởng nhanh.

– Thực vật chuyển gen: Thực vật chuyển gen thể hiện những tính trạng mong muốn như có năng suất cao, có tính chống chịu sâu bệnh hay điều kiện môi trường (hạn, mặn...) khắc nghiệt. Những gen quy định các tính trạng này được phân lập từ vi sinh vật hay thực vật khác loài và được đưa vào tế bào thực vật chủ yếu thông qua *Agrobacterium tumerfaciens*. Ví dụ, các gen *cry* quy định tính kháng côn trùng sâu bệnh phân lập từ vi khuẩn *Bacillus thuringiensis* được chuyển vào nhiều loại thực vật như bông, thuốc lá, khoai tây.... tạo nên các giống có tính chống chịu sâu bệnh.

Hoạt động không mong muốn của một số gen ở một số thời điểm sinh trưởng có thể bị ức chế nhờ kỹ thuật chuyển gen để phiên mã tạo các sợi antisense-RNA. Áp dụng chuyển gen bằng kỹ thuật này lần đầu tiên vào năm 1994 đã tạo ra những giống cà chua quả cứng, chín chậm rất thuận tiện cho thu hoạch và vận chuyển. Enzym polygalacturonase có hoạt tính làm đứt gãy màng tế bào làm cà chua

mềm và chín. Gen mã cho enzym này được chuyển vào tế bào cà chua theo chiều ngược lại. Do đó, các phân tử ARN đối nghĩa (antisense RNA) được phiên mã từ gen chuyển và sẽ tạo phân tử kép với ARNm ngăn cản ARNm được dịch mã.

5.13.6. Chọn lọc ưu thế lai

Các giống cây trồng, vật nuôi có tính trạng quý (nhưng phải quan sát được) thường do các phép lai di truyền kinh điển đòi hỏi rất nhiều thời gian và công sức sàng lọc. Ngày nay, các chỉ thị phân tử, ví dụ, RFLP, ADN vi vệ tinh (microsatellite DNA), được sử dụng để chọn lọc những tính trạng mong muốn (ưu thế lai). Sự không phân ly độc lập của một tính trạng với chỉ thị phân tử di truyền cho phép sử dụng chỉ thị đó để sàng lọc các cá thể có ưu thế lai. Ưu điểm của cách chọn này là tính trạng không nhất thiết phải thể hiện ra bên ngoài.

GIẢI NGHĨA THUẬT NGỮ CHUYÊN DỤNG

• activator: protein hoạt hoá. Protein tương tác với ADN làm tăng tốc độ khởi động phiên mã ở promoter.

• allen: một trong nhiều dạng của gen phân bố trên nhiễm sắc thể tương đồng. Một gen có 2 hoặc nhiều allen (đa allen).

• allosteric transition: sự thay đổi cấu trúc bậc 3 hoặc bậc 4 của protein do tương tác với phân tử khối lượng nhỏ. Tương tác này làm thay đổi hoạt tính của protein.

• alpha (α) helix: cấu trúc bậc 2 của phân tử protein tạo nên do một đoạn peptide cuộn xoắn nhờ liên kết giữa nhóm carboxyl và amin.

• annealing: gắn mồi. Tương tác của đoạn oligonucleotide với khuôn ADN hoặc ARN.

• antibody: kháng thể. Là phân tử protein tương tác với kháng nguyên tại một vị trí đặc hiệu và kích thích cơ thể loại bỏ kháng nguyên đó bằng các cơ chế khác nhau.

• anticodon: đối mã. Là trình tự 3 acid nucleic trên phân tử ARNt tạo cặp bổ sung với một mã di truyền bộ ba trên phân tử ARNm. Liên kết bổ sung cho phép acid amin do ARNt vận chuyển được gắn vào chuỗi peptide đang được tổng hợp.

• antigen: kháng nguyên. Bất cứ phần tử nào khi xâm nhập vào cơ thể có khả năng kích thích sự tạo ra kháng thể và tương tác đặc hiệu với kháng thể được gọi là kháng nguyên.

• antisense ARN: phân tử ARN đối nghĩa. Phân tử này có trình tự bổ sung với trình tự ARN khác, thường là ARNm để hạn chế sự tạo ra ARN đó hoặc ngăn cản quá trình dịch mã trên ARNm.

• autonomously replicating sequence (ARS): tâm tái bản (ở nấm men). Trình tự ADN trên nhiễm sắc thể đảm bảo khả năng tái bản cho nhiễm sắc thể đó.

• bacterial artificial chromosome (BAC): vector tách dòng có khả năng nhận đoạn ADN kích thước lớn, được thiết kế dựa vào plasmid F của vi khuẩn *E.coli*.

• bacteriophage (phage): virus có khả năng xâm nhập vào tế bào vi khuẩn.

• basal promoter: tâm promoter. Vị trí nằm trong promoter eukaryot, tại đó phức khởi động phiên mã bắt đầu hình thành.

• base excition repair: sửa chữa loại base. Cơ chế sửa chữa ADN trong đó base sai hỏng loại bỏ và thay thế mới.

• base pair: cặp base. Hai nucleotide bổ sung với nhau bởi liên kết cầu hydrogen. Khi được viết tắt "bp" có nghĩa là đơn vị chiều dài ngắn nhất của phân tử ADN kép.

• blund end: đầu bằng. Đầu tận cùng của sợi kép ADN kết thúc bằng hai nucleotide liên kết bổ sung với nhau.

• cDNA (complementary DNA): phân tử ADN bổ sung được phiên mã từ ARNm bằng enzym phiên mã ngược. cDNA không chứa intron, chỉ gồm exon so với ADN (gen) ở trong hệ gen (genome).

• centromere: tâm động. Là đoạn nhiễm sắc thể tại đó hai nhiễm sắc tử dính vào nhau trong phân bào nguyên nhiễm. Tâm động là vị trí để phức protein kinetochore bám vào và tương tác với các sợi thoi vô sắc kéo nhiễm sắc thể về các cực.

• chaperone: tên gọi protein có khả năng hoặc ngăn cản sự cuộn gấp sai của protein đích hoặc tạo thuận lợi cho protein đích gấp cuộn chính xác.

• chaperonin: phức protein trợ giúp việc gấp khúc tạo cấu trúc không gian của những protein khác.

• checkpoint: vị trí (điểm) trong chu trình tế bào tại đó sự chuyển giữa 2 phase được quyết định phụ thuộc vào sự đáp ứng đầy đủ các điều kiện cần có cho phase tiếp theo.

• chromatin: nhiễm sắc chất, bao gồm ADN, histone và protein nohistone phân bố trong nhân tế bào eukaryot. Trong giai đoạn phân bào ở metaphase, chất nhiễm sắc co dặc thành nhiễm sắc thể.

• chromatin-remodeling: điều biến chất nhiễm sắc. Quá trình làm thay đổi cấu trúc và vị trí của nucleosome trên sợi nhiễm sắc.

• *cis*-acting: tác động nội sinh. Thường dùng để chỉ đoạn ADN kiểm soát hoạt động của gen nằm trên cùng một nhiễm sắc thể. Ở

tế bào prokaryot, yếu tố tác động nội sinh thường nằm gần gen bị kiểm soát. Tuy nhiên, ở tế bào eukaryot, *cis*-acting có thể phân bố rất xa gen.

• clone: thường để chỉ một nhóm tế bào chứa cùng một loại phân tử ADN tái tổ hợp.

• cloning vector: phân tử ADN có khả năng mang đoạn ADN lạ và khả năng tái bản trong tế bào nhận phân tử đó.

• competent: thường dùng cho các tế bào vi khuẩn có khả năng nhận phân tử ADN lạ.

• consensus sequence: trình tự bảo thủ. Đoạn ADN có trình tự tương đồng cao nhưng không giống nhau hoàn toàn và quy định một chức năng đặc biệt.

• constitutive heterochromatin: dị nhiễm sắc thể bền. Vùng nhiễm sắc thể có cấu trúc co đặc bền vững.

• constitutive gene: gen luôn luôn ở trạng thái hoạt động trong mọi loại tế bào của cơ thể.

• contig: tập hợp của những đoạn ADN nằm gối lên nhau.

• cotransformation: đồng biến nạp. Hai hay nhiều gen cùng nằm trên một phân tử ADN được biến nạp vào tế bào vi khuẩn.

• crossing over: trao đổi chéo. Quá trình trao đổi vật chất di truyền giữa hai nhiễm sắc thể xảy ra do đứt gãy, trao đổi và nối lại với nhau.

• CpG island: đảo CpG. Đoạn ADN giàu CG nằm trước gen eukaryot. Hầu hết các gen hoạt động đều có đảo CpG không bị methyl hoá ở cytosine.

• degenerate: thoái hoá. Thường dùng để chỉ trình tự nucleotide mã cho các acid amin khi những acid amin này được mã bởi hai (hoặc nhiều hơn) mã bộ ba.

• delayed-onset mutation: đột biến chậm. Những đột biến này chỉ có ảnh hưởng ở giai đoạn muộn trong thời gian sống của cơ thể.

• *de novo* methylation: thêm các gốc methyl vào cytosine mà trước đó không bị methyl hoá.

• DNA fingerprint: các băng lai nhận được từ phép lai ADN southern blotting giữa hệ gen của đối tượng sinh vật với đầu dò đặc hiệu.

• DNA photolyase: enzym phát hiện được ở vi khuẩn và thực vật

có khả năng sử dụng năng lượng ánh sáng trong phản ứng sửa chữa ADN.

• DNA shuffling: kỹ thuật phát triển dựa trên phản ứng PCR để tạo nên những phân tử ADN tái tổ hợp có chức năng sinh học hoàn thiện hơn. Ví dụ, tạo nên gen mã cho protein có phổ hoạt động rộng hơn và có hoạt tính cao hơn enzym tồn tại trong tự nhiên.

• domain shuffling: kỹ thuật tạo gen mới. Kỹ thuật này sắp xếp nối các đoạn ADN của một hoặc nhiều gen với nhau; mỗi đoạn mang mã di truyền của một vùng chức năng của một hoặc các protein khác nhau.

• dot-blot hybridization: kỹ thuật lai giữa đầu dò ADN hoặc ARN với ADN hoặc ARN ở những nồng độ đã biết cố định trên màng.

• endocytosis: nhập nội bào. Quá trình thu nhận vật chất từ bên ngoài tế bào nhờ những túi có nguồn gốc từ màng tế bào.

• endogenous retrovirus (ERV): hệ gen của retrovirus ở dạng hoạt động hoặc không hoạt động ghép xen trong nhiễm sắc thể của tế bào chủ.

• enhancer: đoạn tăng cường, làm nhiệm vụ kiểm soát hoạt động của gen và có thể phân bố cách xa gen. Đoạn này rất hiếm ở hệ gen prokaryot.

• episome: plasmid có khả năng chèn ghép vào nhiễm sắc thể của tế bào nhận.

• epitope: quyết định kháng nguyên. Là phần đặc biệt thuộc phần tử kháng nguyên tương tác với kháng thể.

• euchromatin: chất nhiễm sắc. Là những đoạn ít co đặc của chất nhiễm sắc trong giai đoạn gian kỳ. Chất nhiễm sắc chứa hầu hết các gen ở trạng thái phiên mã.

• exocytosis: xuất ngoại bào. Quá trình tiết xuất các chất ra bên ngoài tế bào nhờ các túi tiết dung hợp với màng tế bào.

• exon: là những phần của một gen có mặt trong phân tử ARN hoàn chỉnh. Những phân tử ARN đó tồn tại trong tế bào chất, bao gồm ARNm, ARNr và ARNt.

• expressed sequence tag (EST): trình tự xuất phát từ cADN được dùng để xác định nhanh các gen trong hệ gen.

• expression cloning: kỹ thuật ADN tái tổ hợp được sử dụng để phân lập cADN hoặc đoạn ADN của hệ gen (genome) dựa vào các đặc tính chức năng của protein được mã bởi cADN, hoặc đoạn ADN đó mà

không cần phải tinh sạch protein này. Kỹ thuật ADN tái tổ hợp tạo ra số lượng lớn protein nhờ tách dòng cADN hay gen tương ứng với protein đó.

• expression vector: những plasmid hoặc genome virus chuyên chở gen hoặc cADN vào tế bào nhận thích hợp. Protein do cADN hoặc gen mã cho được tổng hợp trong tế bào nhận.

• facultative heterochromatin: vùng dị nhiễm sắc không bền. Cấu trúc dị nhiễm sắc thường xuất hiện ở một số gen không hoạt động trong một số tế bào hoặc ở một số thời điểm của chu trình tế bào.

• fluorescent *in situ* hybridization (FISH): kỹ thuật lai huỳnh quang giữa các dấu (marker) gắn huỳnh quang và nhiễm sắc thể để xác định vị trí của chỉ thị trên nhiễm sắc thể.

• footprinting: dấu vân tay. Kỹ thuật xác định trình tự ADN hoặc ARN tương tác với protein. Phân tử acid nucleic đã đánh dấu (bằng phóng xạ hoặc bằng chất phát huỳnh quang) bị cắt với nuclease khi thiếu hoặc khi có mặt của protein. Khi có mặt protein, đoạn nucleotide tương tác với protein được bảo vệ nên không bị cắt bởi nuclease. So sánh sự xuất hiện các băng cắt trong hai trường hợp, cho phép xác định được trình tự nucleotide tương tác với protein.

• frameshifting: chuyển khung đọc. Sự di chuyển có kiểm soát của ribosome từ một khung đọc này sang khung đọc khác ở vị trí phân bố trong gen.

• fusion protein: protein dung hợp. Protein tạo nên do sự dung hợp của hai polypeptide, hoặc hai hay nhiều phần của các chuỗi polyppetide được mã bởi các gen khác nhau.

• gel retardation analysis: kỹ thuật xác định vị trí tương tác của protein với ADN dựa vào độ di động của ADN và phức ADN-protein trên gel điện di.

• gene: đơn vị mang thông tin di truyền. Ở mức độ phân tử, gen là đoạn ADN bao gồm cả exon, intron và đoạn nucleotide làm nhiệm vụ kiểm soát phiên mã, cần thiết để tạo ra sản phẩm chức năng cho hoạt động sống của tế bào. Vì vậy, gen mã cho protein và các phân tử ARN.

• gene control: các cơ chế kiểm soát hoạt động của gen, bao gồm giai đoạn phiên mã, sau phiên mã, dịch mã và sau dịch mã.

• gene convesion: chuyển hoán gen. Hiện tượng các allen của một

gen chuyển hoán cho nhau trong quá trình trao đổi chéo ở phân bào giảm nhiễm.

• gene expression: biểu hiện của gen. Quá trình chuyển đổi thông tin di truyền từ gen thành tính trạng.

• genes-within-genes: gen trong gen. Khái niệm dùng để chỉ hiện tượng gen thứ hai nằm trong intron của gen thứ nhất.

• genetic code: mã bộ ba nucleotide trên ADN, hoặc ARN mã cho một acid amin của phân tử protein.

• genome: hệ gen. Tổng số thông tin di truyền (ADN hoặc ARN) có mặt trong tế bào hoặc cơ thể.

• genomic DNA: hệ gen ở dưới dạng ADN.

• genotype: kiểu gen. Khái niệm dùng cho alen của một gen hoặc rộng hơn là toàn bộ các kết cấu di truyền của một tế bào hoặc một cơ thể.

• growth factor: phân tử polypeptide (tồn tại ở bên ngoài tế bào) tương tác với thụ thể trên bề mặt tế bào gây ra chuỗi phản ứng truyền tín hiệu vào bên trong tế bào, dẫn đến sự tăng sinh, biệt hoá tế bào.

• guide RNA: đoạn ARN ngắn tham gia đọc sửa ARNm bằng việc xác định vị trí trên ARNm, tại đó một, hoặc vài nucleotide được thêm vào làm thay đổi trật tự mã di truyền trên ARNm so với mã trên ADN.

• hetechromatin: vùng nhiễm sắc thể, tại đó nhiễm sắc thể co đặc và các gen thường ở trạng thái không hoạt động trong giai đoạn gian kỳ.

• heteroduplex: sợi kép ADN có chứa các nucleotide không tạo cặp bổ sung với nhau.

• heterokaryon: tế bào có từ hai nhân trở lên, tạo ra do dung hợp của hai hay nhiều tế bào khác nhau.

• heterozygous: dị hợp tử. Tế bào lưỡng bội hoặc cơ thể có hai allen khác nhau của một gen.

• homology: sự tương đồng về trình tự nucleotide, hoặc acid amin, hoặc cấu trúc của một bộ phận biểu hiện cùng chung nguồn gốc ban đầu. Cần lưu ý, analogy cũng phản ánh sự tương đồng về cấu trúc hoặc chức năng nhưng không có sự liên quan trong tiến hoá.

• homozygous: đồng hợp tử. Tế bào lưỡng bội hoặc cơ thể có hai allen của một gen giống hệt nhau.

• hybridization: sự kết hợp của 2 sợi đơn nucleic bổ sung với nhau

để tạo ra phân tử kép. Sự kết hợp đó có thể xảy ra giữa 2 sợi đơn ADN, giữa sợi đơn ADN và ARN hoặc giữa 2 sợi ARN.

• intron: những phần của sợi ARN đầu tiên vừa được phiên mã từ ADN sẽ bị loại bỏ trong quá trình biến đổi phân tử ARN đó thành phân tử có chức năng sinh học.

• intein: đoạn ngắn trên sợi polypeptide bị cắt đi sau khi dịch mã kết thúc. Đây là một cách thức biến đổi sau dịch mã để có được phân tử protein có hoạt tính.

• isoform: những protein có hoạt tính giống nhau mặc dù chúng có sự khác nhau về trật tự acid amin.

• karyotype: số lượng, kích thước, hình dáng của toàn bộ các nhiễm sắc thể trong tế bào nhân thực ở giai đoạn metaphase.

• kilobase pair: 1000 cặp base (1000bp).

• kinetochore: cấu trúc protein phân bố tại tâm động, tại đó sợi thoi vô sắc bám vào để kéo nhiễm sắc tử về các cực. Kinetochore đóng vai trò rất quan trọng trong việc di chuyển nhiễm sắc thể về hai cực trong anaphase.

• knockout gene: kỹ thuật gây bất hoạt gen một cách chọn lọc bằng cách thay thế gen đột biến cho gen bình thường của cơ thể.

• label: kỹ thuật đánh dấu một phân tử bằng chất phát quang hoặc bằng đồng vị phóng xạ. Nhờ đó có thể xác định vị trí, hoặc để tinh sạch phân tử, hoặc để theo dõi quá trình di chuyển của nó trong chuỗi các phản ứng hoá học.

• LINE: trình tự nucleotide lặp lại phân bố rải rác trong hệ gen, thường có hoạt tính transposase.

• linker DNA: ADN nằm giữa hai nucleosome, tương tác với histone H_1.

• locus: Vị trí của gen trên nhiễm sắc thể. Mọi allen của một gen đều phân bố tại một locus.

• LTR element: đoạn ADN lặp lại tận cùng bởi hai trình tự lặp lại dài (long terminal repeats).

• maintenance methylation: methyl hoá bảo toàn. Duy trì gốc methyl trên sợi ADN vừa được tái bản dựa vào gốc methyl sẵn có trên sợi khuôn.

• MAP kinase: enzym có hoạt tính chuyển gốc phosphat từ phân

tử ATP vào cơ chất (hoạt tính kinase) trong chuỗi phản ứng tế bào trả lời kích thích đối với các yếu tố kích thích sinh trưởng.

• mapping: xác định vị trí nhận biết của các enzym giới hạn (restriction map), hoặc xác định trật tự sắp xếp của các gen (genetic mapping), hoặc vị trí chính xác của các gen (physical mapping) trên nhiễm sắc thể.

• microsatellite DNA: ADN vi vệ tinh. Trình tự gồm 1-4 nucleotide lặp lại liên tục trên nhiễm sắc thể. ADN vi vệ tinh còn được gọi là trình tự đơn giản lặp lại liên tục (STR-simple tandem repeats).

• minisatellite DNA: ADN tiểu vệ tinh. Trình tự gồm vài chục nucleotide lặp lại liên tục trên nhiễm sắc thể. Loại ADN này còn được gọi là trình tự thay đổi và lặp lại liên tục (VNTR-variable number of tandem repeats).

• mitosis-promoting factor (MPF): yếu tố khởi động phân bào. MPF là protein dị lưỡng cực gồm cyclin và kinase phụ thuộc cyclin (Cdk), thúc đẩy sự co đặc nhiễm sắc thể, phá vỡ màng nhân, bắt tế bào bước vào phân chia. Mitosis-promoting factor còn được gọi là maturation-promoting factor (yếu tố khởi động chín).

• mobile DNA element: đoạn ADN không có một vị trí cố định trên nhiễm sắc thể trong tế bào của mọi cá thể trong một loài.

• monoclonal antibody: kháng thể đơn dòng được tạo ra từ các tế bào xuất phát từ một tế bào B ban đầu. Kháng thể đơn dòng có tính đặc hiệu với chỉ một kháng nguyên.

• motif: thường dùng để chỉ cấu trúc không gian cục bộ của protein mà cấu trúc đó liên quan đến một chức năng.

• multigene family: nhóm các gen, nằm cạnh nhau hoặc phân bố rải rác trên nhiễm sắc thể, có trình tự nucleotide tương tự với nhau.

• mutagen: tác nhân (hoá hoặc lý) gây đột biến.

• nonhomologous end-joining (NHEJ): cơ chế sửa chữa phân tử ADN khi cả hai sợi đơn đều bị đứt gãy tổn thương.

• northern blotting: kỹ thuật xác định phân tử ARN đặc hiệu từ các ARN khác bằng việc lai đầu dò đánh dấu (labeled probe) với các ARN.

• nucleolus: hạch nhân. Cấu trúc nằm trong nhân tế bào eukaryot tại đó ARNr được phiên mã và biến đổi để kết hợp với protein tạo ra tiểu phần ribosome.

• nucleosome: cấu trúc gồm 146bp ADN quấn quanh lõi protein histone, tạo nên cấu hình không gian của sợi nhiễm sắc.

• oligonucleotide-directed mutagenesis: đột biến định vị. Kỹ thuật gây đột biến gen bằng việc thay thế một (một số) nucleotide của gen bằng những nucleotide theo ý muốn, thông qua sử dụng oligo tổng hợp nhân tạo có chứa sẵn các nucleotide mong muốn này.

• operator: điểm kiểm soát tiêu cực của gen vi khuẩn hoặc virus. Điểm này thường là trình tự ngắn các nucleotide phân bố ngay trước vùng mã di truyền của một gen. Đây là vị trí tương tác với protein ức chế ngăn cản khởi đầu phiên mã của gen.

• operon: tập hợp các gen nằm cạnh nhau được kiểm soát bởi một promoter, do đó operon được phiên mã tạo nên một phân tử polycistronic ARNm. Cấu trúc operon chỉ gặp ở vi khuẩn.

• origin of replication: vị trí trên trình tự ADN tại đó quá trình tái bản ADN được bắt đầu khởi động.

• polymerase chain reaction (PCR): phản ứng trùng hợp nhân bản ADN. Kỹ thuật cho phép nhân bản đặc hiệu một đoạn ADN trong số tập hợp rất nhiều đoạn khác nhau.

• phage (thực khuẩn thể phage): virus xâm nhiễm vi khuẩn và có thể gây chết tế bào vi khuẩn.

• phage display: kỹ thuật cho phép xác định các protein có khả năng tương tác với nhau.

• phenotype: kiểu hình (tính trạng). Các đặc tính có thể quan sát được của tế bào hoặc cơ thể tương ứng với một kiểu gen.

• pheromone: tín hiệu truyền tin. Là những phân tử kích thước nhỏ phát đi từ một cá thể làm thay đổi tập tính, hoặc hoạt động của gen ở cá thể khác cùng loài.

• pinocytosis: nhập túi bào. Hiện tượng các túi trong tế bào chất thu nhận không đặc hiệu các giọt lỏng từ bên ngoài tế bào vào.

• plaque assay: kỹ thuật cho phép xác định số lượng các thể virus xâm nhiễm trong mẫu vật bằng việc nuôi cấy các nồng độ pha loãng của mẫu vật trên bề mặt tế bào chủ của virus và tính đếm các vết tan do virus gây ra.

• plasmid: phân tử ADN kích thước nhỏ, dạng vòng nằm ngoài nhiễm sắc thể, có khả năng tự tái bản trong tế bào. Plasmid thường dùng làm vector tách dòng trong kỹ thuật ADN tái tổ hợp.

• point mutation: đột biến điểm. Thay thế hoặc thêm, bớt một nucleotide làm thay đổi hoạt động của gen ở mức độ phiên mã (nếu đột biến ở vùng điều khiển của gen), hoặc làm xuất hiện mã dừng tổng hợp (stop codon), hoặc làm thay đổi khung đọc mở (ORF) nếu đột biến nằm ở vùng mang mã di truyền của gen.

• polyadenylation: thêm đuôi polyA vào đầu 3' của phân tử ARNm eukaryot.

• polyadenylation editing: thêm đuôi polyA tạo nên mã dừng tổng hợp protein trên phân tử ARNm trong ty thể động vật. Đây được xem là cơ chế đọc sửa ARNm bằng việc thêm đuôi polyA vào đầu 3' của phân tử ARNm kết thúc là U hoặc UA.

• positional cloning: xác định vị trí và phân lập gen đột biến dựa vào liên kết di truyền và các chỉ thị ADN phân tử.

• pre-ARNm: bản phiên mã ARNm sơ cấp. Phiên bản được biến đổi thành ARNm dùng làm khuôn để dịch mã.

• primary transcript: bản phiên mã sơ cấp. Bản phiên mã này còn trải qua quá trình biến đổi để có chức năng hoạt tính sinh học đặc trưng cho từng loại phân tử ARN.

• primer: mồi. Đoạn nucleotide ngắn chứa nhóm OH tự do ở đầu tận cùng 3', có trình tự bổ sung với sợi khuôn đơn và là nơi bắt đầu để tổng hợp sợi mới.

• primosome: phức protein tham gia tái bản ADN.

• probe: đầu dò. Đoạn ADN hoặc ARN gắn đồng vị phóng xạ hoặc chất phát quang, dùng để phát hiện trình tự nucleotide tương đồng với đoạn đó. Đầu dò thường dùng trong các phép lai acid nucleic.

• promoter: điểm khởi đầu phiên mã.

• proofreading: chỉnh sửa. Khái niệm này hay dùng trong tái bản ADN để chỉ hoạt tính exonuclease 3'→ 5' của ADN polymerase loại bỏ các nucleotide không tạo cặp với khuôn nhưng bị gắn nhầm vào sợi đang tổng hợp.

• prophage: chỉ trạng thái tiềm tan của genome thực khuẩn thể khi genome này ghép vào nhiễm sắc thể tế bào chủ ở trạng thái tiềm tan.

• proteome: tập hợp tất cả các protein có mặt trong tế bào.

• pseudogene: gen giả. Một gen không có chức năng sinh học.

• reading frame: khung đọc. Trình tự nucleotide gồm bội số của 3 nucleotide (mã bộ ba) trên phân tử ARNm, tính từ vị trí bắt đầu dịch mã (start codon) đến vị trí dừng tổng hợp protein (start codon). Một số phân tử ARNm có thể ứng với 2 khung đọc khác nhau.

• receptor: thụ thể. Bất cứ phân tử protein nào tương tác đặc hiệu với kích thể (ligand) ở bên ngoài tế bào và khởi động chuỗi phản ứng bên trong tế bào trả lời kích thể.

• recombinant DNA: ADN tái tổ hợp. Phân tử ADN được tạo ra do sự ghép nối của các đoạn ADN có nguồn gốc khác nhau. Kỹ thuật tạo phân tử ADN tái tổ hợp thường dựa vào việc cắt ADN có nguồn gốc khác nhau bằng enzym giới hạn và nối chúng bằng enzym ligase.

• recombination: tái tổ hợp. Quá trình trao đổi các đoạn nucleotide giữa hai phân tử ADN, hoặc nhiễm sắc thể tạo ra một tổ hợp mới.

• repetitive DNA: trình tự ADN lặp lại nhiều hơn hai lần trong một gen hoặc trong hệ gen (genome).

• replicon: vùng ADN được sao chép từ một tâm sao chép.

• restriction enzyme: enzym giới hạn. Chúng nhận biết trình tự nucleotide đặc hiệu và cắt phân tử ADN sợi kép thành các đoạn.

• restriction fragment length polymorphism (RFLP): đa dạng các đoạn ADN được cắt bởi enzym giới hạn do sự đa dạng về vị trí nhận biết của enzym.

• retrotransposon: đoạn ADN có khả năng di chuyển trong genome thông qua dạng trung gian là ARN nhờ quá trình phiên mã ngược.

• retrovirus: virus xâm nhiễm tế bào eukaryot có genome là ARN. Tái bản genome của retrovirus trải qua bước phiên mã ngược tạo ADN. ADN này được ghép vào nhiễm sắc thể tế bào chủ. Tại đó, ADN được phiên mã tạo ARN genome và ARNm để dịch mã tổng hợp protein của virus.

• reverse transcriptase: enzym phiên mã ngược có hoạt tính tổng hợp sợi kép ADN từ sợi đơn ARN.

• Rho-dependent terminator: vị trí kết thúc phiên mã trên ARNm ở prokaryot, tại đó sự kết thúc phụ thuộc vào protein Rho.

• RNA-depedent RNA polymerase: enzym phiên mã tổng hợp sợi ARN trên khuôn ARN.

• RNA editing: đọc sửa ARN. Trình tự nucleotide trên phân tử ARNm bị thay đổi.

• RNA splicing: cắt intron, nối exon.

• satellite DNA: ADN vệ tinh. Những trình tự nucleotide lặp lại tạo thành các băng vệ tinh trong ly tâm theo gradient.

• secretory vesicle: túi tiết. Những bào quan nhỏ chứa các chất được tiết ra ngoài tế bào.

• sequence tagged site (STS): trình tự nucleotide duy nhất trong genome.

• silencer sequence: trình tự nucleotide khởi động sự co đặc cục bộ trên sợi nhiễm sắc thể, ngăn cản protein kiểm soát phiên mã tương tác với vùng điều khiển của gen nằm cách đó vài trăm bp.

• simple-sequence DNA: đoạn nucleotide ngắn lặp lại liên tiếp, thường phân bố ở tâm động (centromere) và đầu mút nhiễm sắc thể (telomere).

• southern blotting: kỹ thuật xác định phân tử ADN từ những đoạn ADN khác bằng việc lai đầu dò đánh dấu (labeled probe) với những ADN đó.

• spliceosome: phức ribonucleoprotein bám vào phân tử pre-ARNm để thực hiện phản ứng cắt nối intron-exon.

• TATA box: hộp TATA. Đoạn nucleotide bảo thủ ở promoter eukaryot, tại đó phức khởi động phiên mã tương tác với promoter hoặc với nhau.

• *trans*-acting: tác động ngoại sinh. Thường dùng để chỉ sản phẩm của một gen khi sản phẩm đó có khả năng khuếch tán trong tế bào gây tác động đến hoạt động của một số gen khác trên cùng một nhiễm sắc thể, hoặc trên các nhiễm sắc thể.

• transcription unit: đơn vị phiên mã. Đoạn ADN được giới hạn bởi điểm khởi đầu và điểm kết thúc phiên mã cho một sợi ARN. Sợi ADN này có thể còn trải qua quá trình biến đổi thành các phân tử ARN chức năng.

• transfection: chuyển nhiễm. Quá trình chuyển gen hoạt động vào tế bào nuôi cấy để nghiên cứu tác động của gen đó.

• transformation: biến nạp. Quá trình đưa ADN lạ vào tế bào nhằm gây biến đổi hệ gen của tế bào một cách bền vững và được truyền cho thế hệ sau.

• transgene: Gen được tách dòng và được chuyển vào hệ gen của tế bào động, thực vật, tại đó gen lạ tồn tại bền vững và được di truyền cho thế hệ tiếp sau.

• transposon: ADN di chuyển. Đoạn ADN có khả năng di chuyển trong hệ gen.

• vector (dùng trong giới hạn sinh học phân tử): tác nhân mang ADN để chuyển vào tế bào hoặc cơ thể.

• Western blotting: kỹ thuật cho phép xác định protein đặc hiệu bằng kháng thể dánh dấu.

• wild type: dạng bình thường, tồn tại trong tự nhiên không mang bất kỳ đột biến. Khái niệm này được dùng cho một phân tử, tế bào hoặc một cơ thể. Đối với cơ thể, có thể gọi là loài dại.

TÀI LIỆU THAM KHẢO

1. Hồ Huỳnh Thùy Dương (2003). *Sinh học phân tử*. Nhà xuất bản Giáo dục, Hà Nội
2. Võ Thị Thương Lan (2006). *Một số vấn đề cơ bản của Sinh học phân tử*. Nhà xuất bản Đại học Quốc gia, Hà Nội.
3. Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P. (2002). *Molecular Biology of the Cell (Fourth edition)*. Garland Publishing, New York.
4. Brown T.A. (2002). *Genomes (Second edition)*. BIOS Scientific Publishers, Ltd.
5. Glick B.R., Pasternal J.J. (2003). *Molecular Biotechnology. Principles and applications of recombinant DNA (Third edition)*. ASM Press. Washington, DC.
6. Lewin B., (2004). *Genes VIII*. Oxford University Press.
7. Lodish H., Berk A., Zipursky S.L., Matsudaira P., Baltimore D., Darnell J. (2000). *Molecular Cell Biology (Fourth edition)*. Freeman and Company, NewYork.
8. Snustad D.P., Simmons M.J. (2000). *Principles of Genetics (Second edition)*. John Wiley & Sons, Inc, New York.

Chịu trách nhiệm xuất bản:
Chủ tịch HĐQT kiêm Tổng Giám đốc NGÔ TRẦN ÁI
Phó Tổng Giám đốc kiêm Tổng biên tập NGUYỄN QUÝ THAO
Tổ chức bản thảo và chịu trách nhiệm nội dung:
Chủ tịch HĐQT kiêm Giám đốc Công ty cổ phần Sách ĐH-DN
TRẦN NHẬT TÂN

Biên tập và sửa bản in:
NGUYỄN HỒNG ÁNH
Trình bày bìa:
MẠNH DỨA
Chế bản:
ĐINH XUÂN DŨNG

GIÁO TRÌNH SINH HỌC PHÂN TỬ TẾ BÀO VÀ ỨNG DỤNG
Mã số : 7K687 M6–DAI

In 1000 bản, khổ 16 x 24 cm tại Công ty cổ phần In Phúc Yên
Giấy phép xuất bản số: 251 - 2006/CXB/25 - 510/GD
In xong và nộp lưu chiểu quí II năm 2006.

TÌM ĐỌC SÁCH THAM KHẢO
CỦA NHÀ XUẤT BẢN GIÁO DỤC

1. Đất ngập nước	GS. Lê Văn Khoa
2. Động vật học có xương sống	GS. TS. Lê Vũ Khôi
3. Sinh thái học côn trùng	PGS. TS. Phạm Bình Quyền
4. Cơ sở hoá sinh	PGS. TS. Trịnh Lê Hùng
5. Tài nguyên nước Việt Nam	Nguyễn Thanh Sơn
6. Virut học	PGS. TS. Phạm Văn Ty
7. Bộ sách về công nghệ sinh học	
Tập một : Sinh học phân tử và tế bào Cơ sở khoa học của công nghệ sinh học	PGS. TS. Nguyễn Như Hiền
Tập hai : Công nghệ tế bào động vật và thực vật	GS. TS. Vũ Văn Vụ PGS. TS. Nguyễn Mộng Hùng
Tập ba : Enzym và ứng dụng	GS. TS. Phạm Thị Trân Châu TS. Phan Tuấn Nghĩa
Tập bốn : Công nghệ di truyền	TS. Trịnh Đình Đạt
Tập năm : Công nghệ sinh học vi sinh và công nghệ môi trường	GS. TS. Phạm Văn Ty
8. Cơ sở lý thuyết và kỹ thuật sản xuất thực phẩm	TS. Nguyễn Xuân Phương TSKH. Nguyễn Văn Thoa
9. Di truyền học	Đỗ Lê Thăng
10. Sinh học tế bào	GS. TS. Nguyễn Như Hiền
11. Sinh học phân tử tế bào và ứng dụng	TS. Võ Thị Thương Lan

Bạn đọc có thể tìm mua tại các Công ty Sách và Thiết bị trường học ở địa phương hoặc các cửa hàng sách của Nhà xuất bản Giáo dục :
25 Hàn Thuyên, 187 Giảng Võ, 23 Tràng Tiền - Hà Nội
15 Nguyễn Chí Thanh - TP. Đà Nẵng
240 Trần Bình Trọng, Quận 5 - TP. Hồ Chí Minh

8 934980 641019

Giá: 20.500đ