

Một số vấn đề của sinh học phân tử

Võ Thị Hương Lan



NXB Đại học quốc gia Hà Nội 2007. 181 tr.

Từ khoá: ADN, GEN, genome, nhiễm sắc thể, ty thể, lục tạp, geomic, ADN, tái tổ hợp, ngân hàng các ADNc, cDNA, ADN genome, phản ứng PCR, kỹ thuật gen, phương pháp lai, Protein, tổng hợp protein, vận chuyển protein, tín hiệu tế bào, truyền tín hiệu tế bào, Thụ thể tyrosine kinase, Protein G, sinh trưởng, phát triển, hệ gen lưỡng bội, phôi, chu trình tế bào, phân chia tế bào.

Tài liệu trong Thư viện điện tử ĐH Khoa học Tự nhiên có thể được sử dụng cho mục đích học tập và nghiên cứu cá nhân. Nghiêm cấm mọi hình thức sao chép, in ấn phục vụ các mục đích khác nếu không được sự chấp thuận của nhà xuất bản và tác giả.

Mục lục

LỜI NÓI ĐẦU.....	5
Chương 1 ADN VÀ GEN.....	6
1.1 Khái niệm về gen.....	6
1.2 Genome (hệ gen).....	10
1.2.1. Genome của tế bào prokaryot (tế bào nhân sơ).....	11
1.2.2. Genome của tế bào eukaryot (tế bào nhân thực).....	13
1.3 Cấu trúc sợi nhiễm sắc trong tế bào eukaryot.....	14
1.3.1. Histone trong cấu trúc nucleosome.....	15
1.3.2. Methyl hoá ADN.....	17
1.4 Các gen trong genome eukaryot.....	18
1.4.1. Các gen trong cùng một họ gen.....	20
1.4.2. Gen lặp đi lặp lại liên tục.....	21
1.4.3. Pseudogen (gen giả).....	23
1.5 Thành phần ADN lặp lại trong genome eukaryot.....	23
1.5.1. ADN vệ tinh (satellite DNA) và ADN tiểu vệ tinh (minisatellite DNA).....	23
1.5.2. Các đoạn ADN có khả năng di chuyển.....	24
1.6 Tương tác của T-ADN với genome thực vật.....	29
1.7 ADN trong ty thể và lục tạp.....	32

1.7.1.	ADN ty thể	32
1.7.2.	ADN lục lạp.....	33
1.8	Genomics.....	33
1.8.1	So sánh genome	33
1.8.2	Genome người	34
1.8.3	Nghiên cứu Genomics ở thực vật	35
Chương 2	HOẠT ĐỘNG CỦA GEN TRONG TẾ BÀO	38
2.1	Kiểm soát hoạt động của gen khi phiên mã	41
2.1.1	Kiểm soát khởi đầu phiên mã	42
2.1.2	Kiểm soát kết thúc phiên mã	50
2.1.3	Các protein điều khiển (regulatory proteins)	51
2.2	Kiểm soát sau phiên mã	53
2.2.1	Kim hãm dịch mã liên quan đến cấu trúc vùng 5'UTR của phân tử ARNm.....	53
2.2.2	Độ dài của đuôi polyA ảnh hưởng tới độ bền vững của phân tử ARNm.....	54
2.2.3	Độ bền vững của ARNm	54
2.2.4	ARN anti-sense.....	55
2.2.5	Phản ứng đọc sửa ARNm - "RNA editing"	56
2.3	Kiểm soát ở giai đoạn dịch mã và sau dịch mã	57
2.4	Biến đổi phân tử ARNm trong tế bào eukaryot	59
2.4.1	Phản ứng cắt intron và nối exon	60
2.4.2	Các intron có khả năng tự cắt ra khỏi phân tử ARNm-Phản ứng self-splicing	62
2.4.3	Phản ứng trans-splicing nối hai exon của hai phân tử ARNm.....	64
2.4.4	Cấu trúc chung của phân tử ARNm.....	64
Chương 3	KỸ THUẬT ADN TÁI TỔ HỢP.....	66
3.1	Phân cắt, phân ly ADN.....	66
3.2	Đưa các đoạn ADN vào vector	67
3.2.1	Các vector sử dụng trong kỹ thuật tách dòng	68
3.2.2	Đưa ADN vào vector.....	70
3.3	Ngân hàng ADN.....	72
3.3.1	Ngân hàng các ADNc (cDNA library)	72
3.3.2	Ngân hàng ADN genome (genomic DNA library).....	74
3.4	Sàng lọc một dòng từ ngân hàng ADN	76
3.4.1	Phương pháp sàng lọc chung.....	76
3.4.2	Phương pháp sàng lọc phân biệt "differential screening"	77
3.4.3	Phương pháp đi dọc nhiễm sắc thể "chromosome walking"	78
3.4.4	Nhảy bước trên nhiễm sắc thể "jumping on chromosome"	80
3.5	Các phương pháp lai.....	80
3.5.1	Phương pháp Southern blots.....	81
3.5.2	Phương pháp northern blots.....	82
3.5.3	Kỹ thuật lai in-situ	82
3.5.4	Điều kiện phản ứng lai.....	82
3.6	RFLP trong nghiên cứu genome và lập bản đồ gen	83
3.7	Phản ứng PCR (Polymerase Chain Reaction)	86
3.7.1	Các yếu tố ảnh hưởng đến phản ứng PCR.....	87
3.7.2	Một số dạng của phản ứng PCR	88
3.8	Kỹ thuật gen	89
3.8.1	Nghiên cứu vai trò của ADN điều khiển, chức năng của gen hoặc protein.....	89
3.8.2	Thay thế hoặc gây đột biến gen	92
3.8.3	Gây mất hoặc tăng cường chức năng của gen	93
3.8.4	Gen báo cáo "reporter gene"	96
3.8.5	Biến đổi genome thực vật.....	96

Chương 4 TỔNG HỢP VÀ VẬN CHUYỂN PROTEIN.....	98
4.1 Vai trò của ARN vận chuyển (ARNt) trong tổng hợp protein.....	98
4.2 Tổng hợp protein ở bộ máy Ribosome.....	100
4.3 Vận chuyển protein.....	102
4.3.1 Vận chuyển vào mạng lưới nội chất.....	103
4.3.2 Vận chuyển protein cấu trúc màng (membrane proteins).....	105
4.4 Biến đổi sau dịch mã và kiểm tra chất lượng protein trong khoang ER.....	108
4.4.1 Tạo cầu liên kết disulfide (S-S) và cuộn gấp trong khoang ER.....	108
4.4.2 Hình thành cấu trúc multimer từ các chuỗi peptide.....	109
4.4.3 Quá trình đường hoá protein.....	109
4.5 Vận chuyển từ mạng lưới nội chất đến Golgi và Lysosome.....	110
4.6 Vận chuyển từ Golgi đến bề mặt tế bào: Con đường tiết ngoại bào (exocytosis).....	110
Chương 5 TRUYỀN TÍN HIỆU TẾ BÀO.....	112
5.1 Thụ thể trên bề mặt tế bào.....	114
5.2 Thụ thể nối với protein G.....	117
5.2.1 Protein G.....	117
5.2.2 Hoạt hoá hoặc ức chế cAMPase thông qua protein G.....	119
5.3 Protein kinase phụ thuộc cAMP (cAPK hoặc kinase A).....	121
5.4 Thụ thể tyrosine kinase và các protein Ras.....	124
5.4.1 Thụ thể tyrosine kinase (RTKs).....	124
5.4.2 Protein Ras và chuỗi các phản ứng truyền tín hiệu hoạt hoá bởi thụ thể tyrosine kinase.....	127
5.5 Tín hiệu thứ cấp Ca^{+2} trong chuỗi truyền tín hiệu.....	129
5.5.1 Inositol phospholipid.....	130
5.5.2 Inositol triphosphate (IP3) và sự vận chuyển Ca^{+2} ra khỏi ER.....	130
5.5.3 Calmodulin- protein tạo phức với Ca^{+2} ở trong tế bào.....	132
5.6 Khuếch đại các tín hiệu bên ngoài tế bào.....	133
5.7 Truyền tín hiệu qua các thụ thể nối với enzym trên bề mặt tế bào.....	135
5.7.1 Thụ thể guanylyl cyclase.....	135
5.7.2 Các oncogene và tín hiệu dẫn truyền từ thụ thể tyrosine kinase.....	136
5.7.3 Protein MAP kinase.....	136
5.8 Tyrosine kinase phối hợp với thụ thể. Thụ thể Tyrosine phosphatase.....	137
Chương 6 CHU TRÌNH VÀ PHÂN CHIA TẾ BÀO.....	139
6.1 Những đặc tính cơ bản của chu trình tế bào.....	139
6.2 Chu trình tế bào ở giai đoạn phát triển phôi sớm.....	143
6.3 Protein cyclin.....	145
6.4 Nấm men và hệ thống kiểm soát chu trình tế bào.....	147
6.5 Kiểm soát phân bào ở động vật.....	150
6.6 Vai trò của sợi vi ống tubulin trong phân bào.....	152
Chương 7 SINH TRƯỞNG VÀ PHÁT TRIỂN.....	154
7.1 Kiểm soát xác định giới tính.....	155
7.2 Phát triển ở ruồi giấm Drosophila.....	158
7.3 Hoạt động của các gen có nguồn gốc từ mẹ trong quá trình hình thành trục đầu-đuôi và trục lưng-bụng.....	159
7.3.1 Nhóm gen quyết định phát triển của phần đầu và ngực ấu thể (anterior-group genes).....	160
7.3.2 Nhóm gen qui định phát triển phần đuôi (posterior-group genes).....	162
7.3.3 Nhóm gen qui định phát triển trục lưng-bụng (dorsoventral-group genes).....	162
7.3.4 Nhóm gen qui định phát triển các cấu trúc tận cùng của ấu thể (terminal-group genes).....	164

7.4	Hoạt động của các gen trong hệ gen lưỡng bội (phôi)	164
7.3.5.	Các gen tạo đốt "gap"	166
7.3.6.	Các gen cặp đốt "pair-rule"	166
7.3.7.	Các gen phân cực đốt.....	167
7.5	Các gen chọn lọc	167

Lời nói đầu

Với mong muốn chia sẻ cùng bạn đọc mối quan tâm về Sinh học phân tử, một lĩnh vực đang được học tập và nghiên cứu ở Việt Nam, chúng tôi xuất bản cuốn sách "**Một số vấn đề cơ bản của Sinh học phân tử**" nhằm giới thiệu những quá trình quan trọng xảy ra trong tế bào (trình bày trong chương 1, 2, 4, 5, 6 và chương 7) và một số kỹ thuật cơ bản được sử dụng để nghiên cứu những quá trình đó (chương 3). Những quá trình này được nghiên cứu ở mức độ phân tử phần nào làm sáng tỏ sự giống và khác nhau trong cấu trúc của genome, cấu trúc của một gen giữa tế bào prokaryot và eukaryot (chương 1). 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 2). 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. Ngoài ra, chức năng và hoạt tính của những protein tham gia quá trình truyền tín hiệu được trình bày trong chương 5; của protein tham gia chu trình tế bào được trình bày trong chương 6 và những protein tham gia kiểm soát biệt hoá, phát triển, sinh trưởng và hình thành cơ thể được giới thiệu trong chương 7.

Để có thể học được những kiến thức chuyên sâu trong lĩnh vực sinh học phân tử, tôi xin bày tỏ lòng biết ơn sâu sắc tới các thầy, các cô trong Khoa Sinh học Trường Đại học Tổng hợp Hà Nội (nay là Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội). Đồng thời tôi xin chân thành cảm ơn Phó Giáo sư Trương Nam Hải và Giáo sư Nguyễn Mộng Hùng đã có những nhận xét và góp ý quý báu cho cuốn sách.

Lần đầu xuất bản, chắc chắn cuốn sách còn có những thiếu sót, tôi rất mong nhận được sự phê bình, góp ý của bạn đọc và đồng nghiệp.

Với sự cảm ơn chân thành!

Tác giả

Chương 1

ADN VÀ GEN

1.1 Khái niệm về gen

Trải qua một thời gian dài, các khái niệm và định nghĩa về gen dần dần được hình thành dựa vào kết quả thí nghiệm, trước hết là các thí nghiệm di truyền cổ điển. Đầu tiên, 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 alen của một gen. Một gen có thể có nhiều alen. Mức độ biểu hiện của tính trạng phụ thuộc vào sự kết hợp giữa hai alen. Đơn giản nhất là một gen có 2 alen (Aa). Khi đó tính trạng có thể biểu hiện ở 3 mức độ khác nhau: trội (AA), bán trội (Aa), hoặc lặn (aa). Tiếp theo đó, với một loạt các thí nghiệm tiến hành trên ruồi giấm *Drosophila*, Morgan và cộng sự đã nhận thấy một số tính trạng được quyết định không phải do các alen 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, từ kết quả thí nghiệm, các nhà di truyền đã đưa ra 3 đặc tính để xác định một gen:

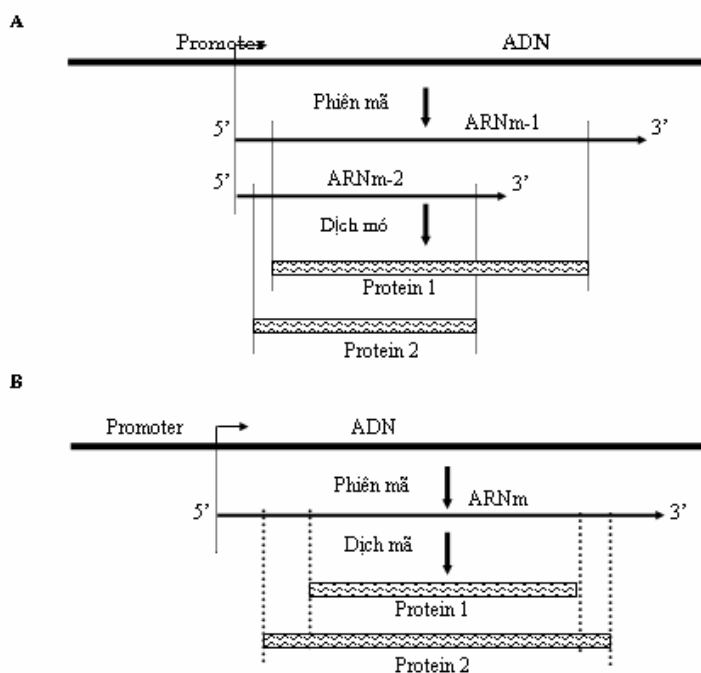
1. Gen qui định một tính trạng có thể quan sát được và chiếm một vị trí trên nhiễm sắc thể.
2. Gen được xem là đơn vị di truyền nhỏ nhất có thể bị đột biến.
3. Gen được xem là đơn vị di truyền nhỏ nhất mà trao đổi chéo không thể xảy ra trong một gen. Trao đổi chéo được thực hiện giữa các gen tương đồng.

Từ những đặc tính này, rõ ràng hai tính trạng không giống nhau có thể phân biệt được thì phải do ít nhất hai gen khác nhau qui 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 vị trí một gen gặp rất nhiều khó khăn do phải sàng lọc các cá thể đột biến từ số lượng cá thể rất lớn ở các thế hệ con cháu qua các phép lai khác nhau. Mặt khác, bằng phân tích trao đổi chéo, vị trí các gen có thể được xác định trên bản đồ di truyền nhưng không phản ánh được chức năng riêng biệt của chúng. 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 qui 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. Nói cách khác, một gen có thể chứa nhiều đột biến khác nhau. Nhà di truyền học Seymour Benzer đã xác định được 199 vị trí đột biến trên gen *rIIA* ở thực khuẩn thể T4. Đặc biệt nhờ vào việc khám phá ra cấu trúc ADN, Charles Yanofsky 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 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.



Hình 1.1:

Hai protein được mã bởi một đoạn ADN duy nhất do điểm bắt đầu (hoặc kết thúc) quá trình phiên mã tổng hợp ARNm xảy ra ở các vị trí khác nhau ngay trên đoạn ADN đó tạo ra các sợi ARNm khác nhau (A) hoặc do điểm khởi đầu dịch mã tổng hợp protein phân bố ở các vị trí khác nhau trên một sợi ARNm (B).

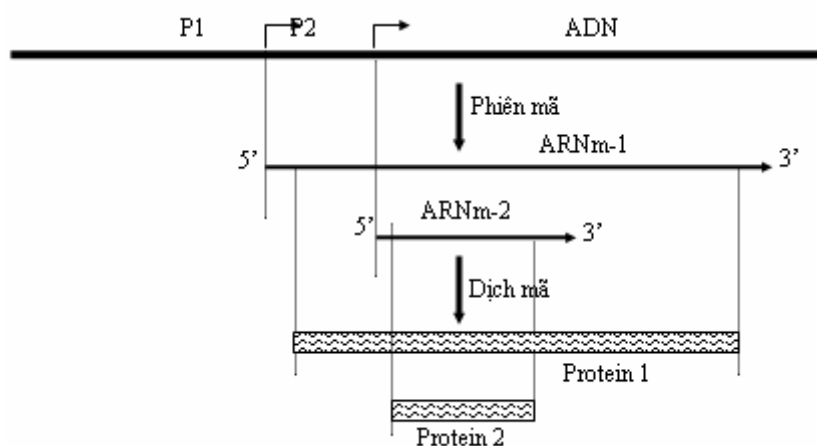
Tuy nhiên, khái niệm gen nêu trên không thể giải thích cho một số hiện tượng như sau:

a/ Hiện tượng các gen gối lên nhau (*overlapping genes*): trên một đoạn ADN hai gen không nằm kế tiếp nhau mà gen này nằm gối đầu lên gen kia. Như thế, phần ADN có 2 gen nằm gối lên nhau chứa mã di truyền cho cả hai gen. Có thể xảy ra các trường hợp sau:

* Hai phân tử ARNm được phiên mã từ các vị trí bắt đầu hoặc kết thúc khác nhau trên một đoạn ADN. Kết quả là hai phân tử protein (được dịch mã từ hai sợi ARNm) có chứa một đoạn acid amin giống hệ nhau mặc dù hai protein đó các chức năng khác nhau trong tế bào (Hình 1.1A).

* Một phân tử ARNm được phiên mã từ một đoạn ADN có thể dùng làm khuôn để tổng hợp hai chuỗi polypeptide khác nhau do điểm bắt đầu dịch mã (*start codon*) phân bố lệch nhau (hiện tượng lệch khung đọc). Hai protein này có thể khác nhau hoàn toàn về trình tự acid amin và chức năng trong tế bào (Hình 1.1B).

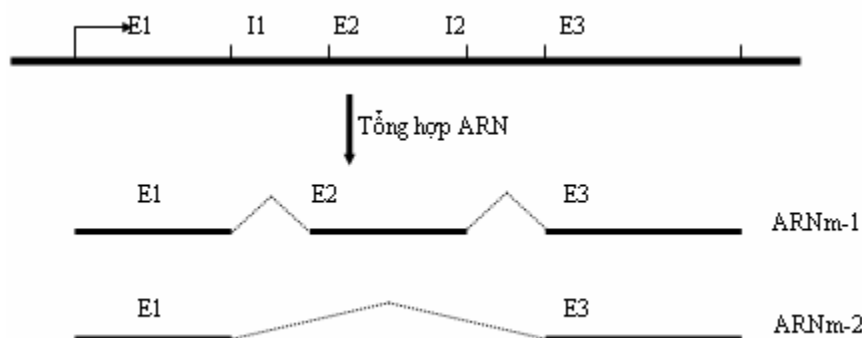
b/ Một đoạn ADN mang mã di truyền của 2 gen nên được phiên mã tổng hợp nên 2 loại ARNm khác nhau. Điều này xảy ra khi mã di truyền phân bố theo các khung đọc khác nhau ngay trên đoạn ADN đó (Hình 1.2). Do đó, hai protein có thành phần acid amin và chức năng khác nhau hoàn toàn được tổng hợp. Một đột biến xảy ra tại một vị trí trên đoạn ADN này có thể gây ảnh hưởng đến một hoặc cả hai gen. Điều đó gây khó khăn cho việc xác lập bản đồ tính trạng.



Hình 1.2:

Hai gen mã cho hai protein cùng nằm trên một đoạn ADN do mã di truyền của hai gen này phân bố theo các khung đọc khác nhau

c/ Đối với sinh vật eukaryot, một gen thường bao gồm các đoạn nucleotide chứa mã di truyền (*exon*) xen kẽ với các đoạn không chứa mã (*intron*). Các exon và intron đều được phiên mã sang phân tử ARN (gọi là phân tử tiền thân ARN thông tin-ARNm). Sau đó, các intron sẽ bị cắt bỏ đi, các exon được nối lại với nhau theo đúng thứ tự để tạo ra phân tử ARNm hoàn chỉnh. Có thể xảy ra trường hợp hoặc là chỉ một số intron hoặc là tất cả các intron đều bị loại đi khỏi phân tử ARNm. Mặt khác có thể xảy ra hoặc tất cả các exon hoặc chỉ một số exon được nối với nhau. Việc lựa chọn intron để cắt sẽ tạo ra các phân tử ARNm khác nhau mặc dù chúng đều xuất phát từ một loại ARNm tiền thân được phiên mã từ một khuôn ADN (Hình 1.3). Đây là hiện tượng cắt nối intron-exon luân phiên (*alternative splicing*).

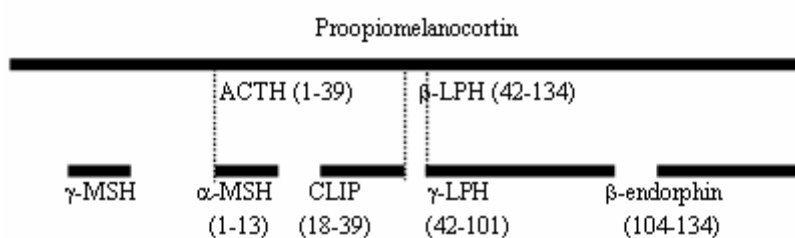


Hình 1.3:

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.

d/ Gen mã cho polyprotein: Polyprotein là sản phẩm đầu tiên của việc dịch mã từ một phân tử ARNm, nhưng sau đó phân tử protein này bị cắt ra thành các đoạn peptide nhỏ hơn. Phân tử polyprotein không có hoạt tính. Chỉ có các đoạn peptide mới có các chức năng khác nhau. Ví dụ, các hormon adrenocorticotropic (ACTH), lipotropic (LPHs), hormon kích hoạt melanocyte (MSHs) và enkephalin được tạo ra từ một phân tử proopiomelanocortin ban đầu (Hình 1.4). Như vậy trên thực tế, một đoạn ADN sao chép ra một loại ARNm nhưng có nhiều loại protein được tạo thành.

e/ Một số gen không mang thông tin di truyền cho protein: Một điều rõ ràng rằng các phân tử ARN ribosome (ARNr), ARN vận chuyển (ARNt) đều được sao chép từ ADN nhưng chúng không được dịch mã. Ngoài ra, trong nhân tế bào eukaryot còn tìm thấy các phân tử ARNsn kích thước nhỏ (*small nuclear RNA*) đảm nhiệm rất nhiều chức năng khác nhau như tham gia vào việc biến đổi phân tử ARNm (cắt intron và nối exon), kiểm tra lại thông tin di truyền trên chúng (cơ chế đọc sửa ARNm), tác động đến độ bền vững của ARNm trong tế bào chất hoặc tham gia vào cơ chế bất hoạt gen (*ARNi-interference RNA* - tạm dịch là ARN nhiễu). Do đó, các đoạn ADN mã cho các loại ARN này phải được xác định như các gen bởi lẽ đột biến trên chúng đều có thể liên quan đến việc xuất hiện các tính trạng lạ.

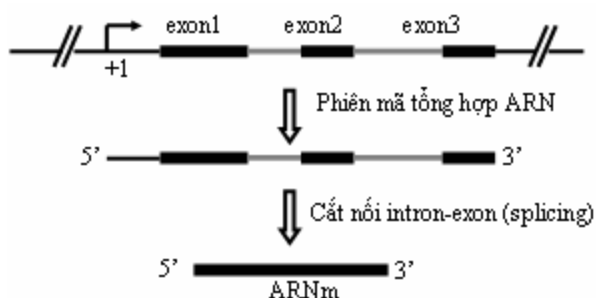


Hình 1.4:

Phân tử proopiomelanocortin được phân cắt để tạo ra các hormon MSH, LPH, CLIP và β -endorphin có hoạt tính.

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 cần thiết cho 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 (Hình 1.5). 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.



Hình 1.5:

Cấu trúc gen mã cho protein ở tế bào nhân thực (eukaryote gene). Vị trí nucleotide đầu tiên được phiên mã sang phân tử ARN được ký hiệu là +1. Nucleotide nằm trước vị trí +1 được ký hiệu -1 (không có vị trí 0). Các nucleotide nằm trước vị trí (+1) thuộc vùng promoter. Các intron nằm xen kẽ các exon. Intron bị loại khỏi phân tử ARNm bởi phản ứng cắt nối intron-exon (splicing). Chiều phiên mã được chỉ bằng mũi tên.

1.2 Genome (hệ gen)

Genome chứa toàn bộ thông tin di truyền lập trình đảm bảo hoạt động sống cho tế bào. Đa số genome vi khuẩn phân bố trên một nhiễm sắc thể có kích thước nhỏ và có dạng vòng khép kín. 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 sinh- *epigenetic information*). Trình tự nucleotide của toàn bộ genome đã được xác định đối với một số sinh vật mô hình (*model organisms*) đại diện cho mỗi giới sinh vật như vi khuẩn *E.coli*, nấm men, ruồi giấm, giun tròn, *Arabidopsis* và người. 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 (*physical 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*) và 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 càng khó xảy ra trao đổi chéo giữa chúng 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 ánh chính xác khoảng cách giữa các vị trí trên bản đồ di truyền. Kết hợp giữa bản đồ vật lý và bản đồ di truyền cho 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, 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 10000 bp thì tổng số chiều dài các gen hoạt động trong một tế bào cũng chỉ chiếm 1-2% genome. Hay nói cách khác chỉ một phần rất nhỏ genome mang thông tin di truyền cần thiết cho hoạt động sống của tế bào. 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 genome của những loài cách xa nhau (tức là có tính phức tạp khác nhau) cho thấy kích thước genome không phải luôn luôn tỷ lệ với tính phức tạp của loài. Ví dụ, genome của người có kích thước khoảng $3,3 \times 10^9$ bp, trong khi đó genome các loài lưỡng cư dài tương tự cỡ $3,1 \times 10^9$ bp hoặc của thực vật có thể đạt đến 10^{11} bp. Có lẽ nào loài lưỡng cư lại có tính phức tạp như cơ thể chúng ta? Mặt khác, ngay trong cùng một loài chúng ta cũng nhận thấy sự mâu thuẫn về kích thước genome. Ví dụ, ruồi sống trong nhà (*Musca domestica*) có genome cỡ $8,6 \times 10^8$ bp, lớn gấp 6 lần kích thước genome ruồi giấm (*D.melanogaster*) với genome cỡ $1,4 \times 10^8$ bp. Ngoài ra, kích thước genome của các quần thể lưỡng cư thay đổi từ 10^9 bp đến 10^{11} bp (khác nhau gấp 100 lần). Vì sao ngay trong cùng một loài kích thước genome lại biến thiên nhiều như vậy ?

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, các gen phân bố không theo qui luật trong genome. Thứ hai, 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 vv... 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.

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

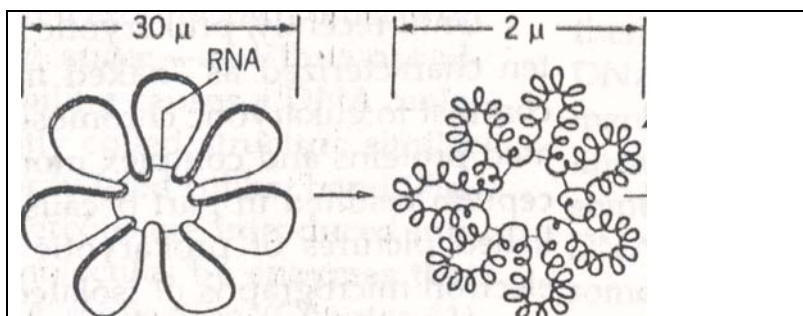
Genome trong tế bào prokaryot không lớn nên số lượng genome của các loài vi khuẩn được xác định trình tự ngày càng nhiều. Nhờ đó các thông tin dữ liệu về cấu trúc hệ gen prokaryot, sự phân bố các gen, cách thức kiểm soát hoạt động cũng như chức năng của chúng ngày càng phong phú và trở nên rõ ràng.

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. Do 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. Hơn nữa, một loại plasmid đôi khi được tìm thấy 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 5 Mb (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 910 kb 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 533 kb 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*.

Những dẫn liệu thực nghiệm nhận được khi phân tích genome và các plasmid ở *Borrelia burgdorferi* đã gây tranh cãi giữa các nhà sinh học khi so sánh genome *Borrelia burgdorferi* với *Treponema pallium*. Theo phân loại, đây là hai loài vi khuẩn có quan hệ gần gũi nhau. Giống như đa số các tế bào prokaryot khác, genome loài thứ hai là một phân tử ADN dạng vòng có kích thước 1138 kb với 1041 gen. Điều thú vị là không một gen nào ở *Treponema pallium* tương đồng với các gen phân bố trên plasmid của loài thứ nhất. Phải chăng các plasmid vừa được tự nhiên biến nạp vào *Borrelia burgdorferi*?

Genome prokaryot không đóng gói trong cấu trúc nucleosome (như genome eukaryot) và không được bao bọc bởi màng nhân. Nhiễm sắc thể dạng vòng có cấu trúc không gian giống như những cánh hoa của bông hoa, mỗi cánh là một đoạn ADN có cấu trúc siêu xoắn (*supercoil*). Các cánh không đều như nhau và được đính vào lõi protein. Genome vi khuẩn có khoảng 40-50 cánh. Cấu trúc kiểu bông hoa này được gọi là *nucleoid* (Hình 1.6). Cấu trúc nucleoid giúp genome chỉ chiếm một thể tích rất nhỏ trong tế bào. Ngoài ra, cấu trúc không gian này của nhiễm sắc thể được duy trì nhờ các phân tử ARN kích thước nhỏ tương tác với protein. Do đó, ngay khi bị đứt gãy, cấu trúc siêu xoắn của nhiễm sắc thể cũng chỉ mở ra một cách cục bộ ở cánh bị tổn thương chứ không xảy ra trên toàn bộ genome. Hai enzym ADN gyrase và ADN topoisomerrase giữ vai trò chính cùng phối hợp với phức protein khác làm nhiệm vụ đóng gói 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 ở dạng tetramer tạo thành lõi được quấn quanh bởi đoạn ADN khoảng 60 bp. Như vậy, protein HU có chức năng tương tự histone trong việc qui 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 1.6:

Mô hình cấu trúc nucleoid ở *E.coli* gồm 40-50 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).

Năm 1995, genome *Haemophilus influenzae* là genome đầu tiên được xác định toàn bộ trình tự. Đến năm 1998 đã có hơn 18 genome vi khuẩn khác được đọc hoàn toàn. Trong số này, *Mycoplasma genitalium* có kích thước nhỏ nhất gồm 580.070 bp và *Mycobacterium tuberculosis* có kích thước lớn tới 4.411.529 bp.

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. Hệ gen *E.coli* gồm 4.639.221 bp với 4.288 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 methionine (*start codon*) và kết thúc bởi một trong số ba mã dừng tổng hợp protein (*stop codon*).

Mặc dù có kích thước nhỏ hơn nhiều so với genome eukaryot, nhưng genome prokaryot có mật độ phân bố các gen cao hơn, số đoạn ADN không chứa mã di truyền ít hơn. Nói cách khác, khoảng cách giữa các gen ngắn hơn. Ví dụ, khoảng cách trung bình giữa hai gen ở *E.coli* là 118 bp. 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 có intron.

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 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 hệ gen 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*. Hơn nữa, việc 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 tăng dần số lượng gen cần bị đột biến và nhận thấy ít nhất cần có 300 gen để đảm bảo sự sống cho vi sinh vật đơn giản nhất. 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*.

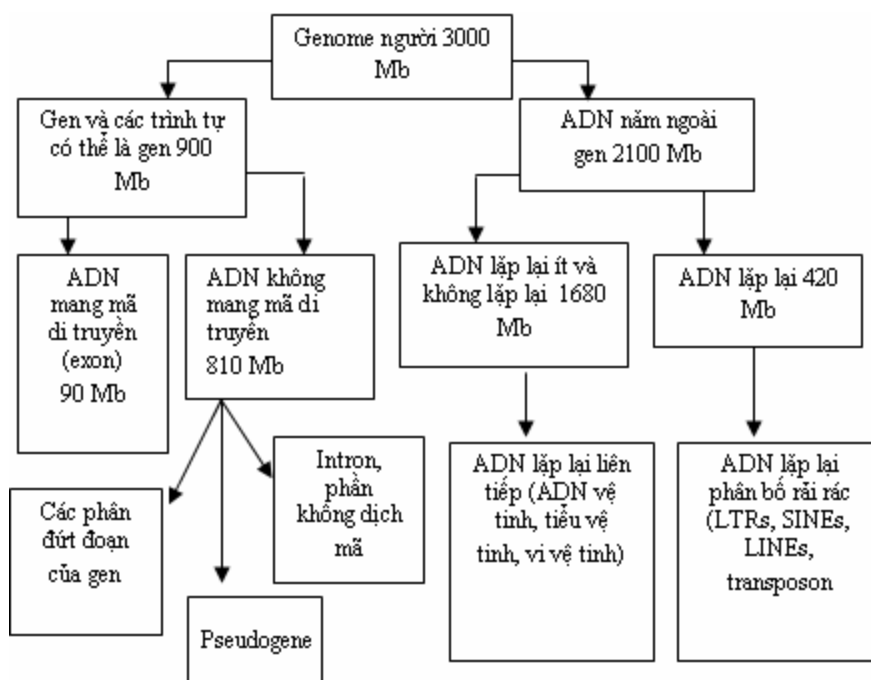
1.2.2. Genome của tế bào eukaryot (tế bào nhân thực)

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, 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 hệ gen 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, hệ gen 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 giải thích được.

Kích thước genome trong nhân eukaryot thay đổi từ 12 Mb (nấm men *S.cerevisiae*) đến 120.000 Mb (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 10^9 bp 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 (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).

1.3 Cấu trúc sợi nhiễm sắc trong tế bào eukaryot

Trong nhân tế bào eukaryot, ADN liên kết với protein tạo ra cấu trúc gọi là **chromatin** (sợi nhiễm sắc). Có thể phân biệt các protein này làm hai nhóm chính: histone và non-histone. Thành phần protein non-histone thay đổi giữa các mô, tổ chức, giữa các loài. Mỗi loại protein non-histone chỉ chiếm một số lượng rất nhỏ so với tổng số protein non-histone hoặc với bất kỳ loại protein histone nào. Tuy nhiên các protein non-histone giữ một vai trò rất quan trọng qui định cấu trúc không gian đặc thù của từng vùng nhiễm sắc thể. Hoạt động của nhiều gen, đặc biệt các gen liên quan đến phát triển phôi, không chỉ phụ thuộc vào trình tự nucleotide mà còn phụ thuộc vào cấu trúc của nhiễm sắc thể. *Thông tin di truyền chứa trong cấu trúc không gian của nhiễm sắc thể được gọi là thông tin di truyền ngoại sinh (epigenetic information).*

Sử dụng dung dịch có liên kết ion yếu, có thể tách ra khỏi nhân tế bào các sợi nhiễm sắc ở dạng sợi đơn, đường kính khoảng 30 nm, gồm các hạt nhỏ giống như chuỗi hạt cườm (đường kính hạt khoảng 10 nm). Các hạt nhỏ này được gọi là **nucleosome**. Chúng không phân bố đồng đều ở mọi vùng trên sợi nhiễm sắc. Khi sợi ADN bị cắt bởi nuclease, các nucleosome tách ra riêng biệt. Mỗi nucleosome gồm một đoạn ADN dài 146 bp quấn 2 vòng quanh lõi protein chứa 8 tiểu phần của 4 loại histone H2A, H2B, H3, H4. Phần đầu NH₂ của histone không nằm trong cấu trúc nucleosome mà tồn tại tự do. Đoạn ADN nằm giữa hai nucleosome được gọi là ADN nối (*linker DNA*). Đoạn này có kích thước khoảng 50-70 bp. Protein histone H1 liên kết với ADN linker nằm giữa 6 nucleosome và đóng gói chúng lại thành một cấu trúc đặc biệt gọi là **solenoid** (giống như một bông hoa 6 cánh). Các solenoid quấn chồng lên nhau thành sợi xoắn. Nhờ cấu trúc đặc biệt đó nên thể tích ADN chiếm trong nhân giảm đi rất nhiều.

Sợi ADN quấn quanh lõi histone trong cấu trúc nucleosome và được đóng gói trong các solenoid có độ trật tự rất cao. Một điều thú vị được đặt ra là các cấu trúc này thay đổi như thế nào khi một đoạn ADN được sử dụng để phiên mã (tạo ARNm) hoặc để sửa chữa khi xảy ra sai hỏng? Liệu khi đó chúng có bị phá vỡ tạm thời như trong quá trình tái bản ADN hay không? Các phân tử histone được giải phóng hay vẫn liên kết với ADN? Giải đáp những câu hỏi này có nhiều kết quả khác nhau cho thấy việc phiên mã không nhất thiết yêu cầu phá vỡ cấu trúc chromatin. Tuy nhiên chắc chắn có xảy ra những thay đổi trong các tương tác protein-ADN, histone-histone, histone-non histone. Điều đặc biệt có ý nghĩa là việc thêm bớt các nhóm chức trên từng phân tử protein tham gia liên kết tạo nucleosome khiến cho cấu trúc nucleosome thay đổi. Lõi histone có thể không bị phân rã thành các tiểu phần nhưng bị dịch chuyển một cách cục bộ trên sợi nhiễm sắc bởi các protein điều biến chromatin (*remodelling chromatin proteins*). Các protein này giúp cho việc tháo gỡ cục bộ sợi ADN khỏi cấu trúc nucleosome mà không ảnh hưởng đến các vùng khác.

ADN được giải phóng ra khỏi nucleosome không có nghĩa chúng ở trạng thái tự do, vì như vậy ADN rất dễ bị phá hủy bởi các tác nhân khác nhau trong tế bào, đặc biệt bởi các nuclease. Lúc đó ADN thường liên kết với các protein đặc hiệu (các factor) cần thiết cho sự phiên mã hoặc các phức cần thiết cho quá trình sửa chữa ADN. Thí nghiệm cho thấy khi các factor phiên mã tương tác với ADN, chúng có khả năng ngăn cản histone liên kết với ADN. Điều này có thể lý giải việc tồn tại những vùng trơ với nuclease nằm xen các vị trí nhạy cảm trong một gen. Khi protein bám vào ADN, ADN được bảo vệ khỏi sự phân cắt của enzym.

1.3.1. Histone trong cấu trúc nucleosome

Mỗi nucleosome có 146 bp ADN quấn hai vòng quanh lõi histone gồm 8 tiểu đơn vị 2x[H3-H4] và 2x[H2A-H2B]. Các histone của lõi có cấu trúc tương tự như nhau, gồm các đoạn peptide (*domain*) tận cùng đầu NH₂ (*N-terminal*), domain chung giúp histone gấp khúc

và phần tận cùng COOH (*C-terminal*). Domain cần cho histone gấp khúc còn liên quan đến tương tác giữa các histone và giữa histone với ADN.

Các nucleosome được gắn với nhau nhờ histone H1. Protein này liên kết lõi histone với ADN nối (*ADN linker*) nằm giữa nucleosome. Độ dài của các ADN linker không cố định như nhau. Histone H1 thiết lập nên cấu trúc có trật tự cao cho sợi nhiễm sắc.

Các histone tham gia cấu trúc lõi đều có phần đầu NH₂ nằm ở ngoài lõi, phân bố tự do theo các hướng khác nhau (Hình 3-GT). Chiều dài của đoạn phân bố tự do thay đổi từ 16 đến 44 acid amin (H3-44; H2B-32; H4-26 và H2A-16 acid amin). Các đoạn này giữ vai trò quan trọng đối với sự co đặc của sợi nhiễm sắc. Nghiên cứu động học quá trình thay đổi cấu hình của sợi nhiễm sắc cho thấy nó có thể tồn tại ở ba dạng: không co đặc (*unfolded*), co đặc ở mức độ trung bình (*moderately folded*) và co đậm đặc (*extensively folded*). Đoạn tự do của H3 và H4 cần thiết để sợi nhiễm sắc có mức độ co đặc vừa phải. Đặc biệt đoạn tự do của H3 không thể thay thế được. Độ dài và vị trí ra khỏi phần lõi của đoạn này cần thiết cho việc hình thành cấu trúc không gian của sợi nhiễm sắc. Như vậy cùng với histone H1, các đoạn tự do của bốn loại histone trong cấu trúc lõi đều cần thiết để duy trì cấu trúc không gian ba chiều cho nucleosome, duy trì trạng thái co đậm đặc của sợi nhiễm sắc cũng như tương tác giữa các sợi nhiễm sắc với nhau. Ngoài ra, chúng còn là các vị trí tương tác với các protein non-histone.

Sau khi được tổng hợp, cả bốn loại histone đều chịu các biến đổi như ubiquitin hoá, phosphoryl hoá, glycosyl hoá và đặc biệt có ý nghĩa là quá trình methyl hoá và acetyl hoá. Hầu hết các biến đổi này xảy ra ở vùng N-terminal. Quá trình phosphoryl hoá và methyl hoá có thể tác động qua lại với nhau, ảnh hưởng đến sự co đặc của nhiễm sắc thể khi bước vào mitose. Riêng trường hợp ubiquitin hoá xảy ra ở phần đuôi C-terminal của histone, giúp cho cấu trúc nucleosome bị phá vỡ tạm thời trong quá trình tái bản hoặc tổng hợp ARN. Như vậy, các biến đổi hoá học của histone tác động đến cấu trúc không gian của nucleosome và hoạt động của gen, trước hết ở quá trình phiên mã.

Các histone trong cấu trúc lõi bị acetyl hoá tại các acid amin lysine đặc hiệu phân bố ở phần N-terminal. Ngoại trừ histone H2A, các histone lõi khác thường có 4 đến 5 vị trí có gắn nhóm acetyl. Một nucleosome có thể có tới 26 vị trí mang nhóm acetyl. Acetyl hoá histone có một vai trò quan trọng, quyết định đến cấu trúc cứng sợi nhiễm sắc. Nhờ đó sợi nhiễm sắc không co đặc, ADN được giải phóng ra khỏi nucleosome, sự tương tác giữa các nucleosome bị phá hủy, gây thay đổi liên kết giữa các domain N-terminal của histone với các protein non-histone, hoặc liên kết giữa các protein với ADN. Những biến đổi này góp phần hoạt hoá phản ứng tổng hợp ARN. Chỉ cần 46% trong tổng số 26 vị trí đặc biệt bị acetyl hoá cũng đủ phá vỡ trật tự cấu trúc của sợi nhiễm sắc và tăng cường quá trình sao chép ARN ở các gen.

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 H2A-H2B 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) cùng với 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 acetyl hoá 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.

Ở động vật có xương sống, bốn loại histone H2A, H2B, H3 và H4 ít thay đổi giữa các loài. Tuy nhiên protein H1 gồm một số dạng được ký hiệu từ H1a đến H1e, H1t và H5. Vị trí phân bố của các loại histone H1 này chưa được xác định rõ ràng. Mặt khác trong các tế bào tinh trùng, histone được thay thế bởi protein protamine. Hơn nữa, histone có tính kiềm do cấu trúc bậc I của chúng có khoảng 20-30% arginine và lysine. Đây là các acid amin tích điện dương (+). Nhờ vậy thay đổi điện tích của histone liên quan chặt chẽ đến khả năng tương tác với ADN và độ bền vững của tương tác đó vì acid nucleic có điện tích âm quyết định bởi nhóm phosphate.

1.3.2. Methyl hoá ADN

Bản thân ADN cũng 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 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 sinh (*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. Sự methyl hoá ADN ở prokaryot được xem như là một cơ chế bảo vệ hệ gen, 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 DNA (*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. Hiện tượng đánh dấu ADN sẽ được xem xét chi tiết ở phần sau.

Methyl hoá ADN có ý nghĩa đặc biệt đối với hoạt động của gen eukaryot, nhất là các gen trong quá trình hình thành phát triển 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 methyl-cytosine 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 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 1-2 kb 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 (*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.

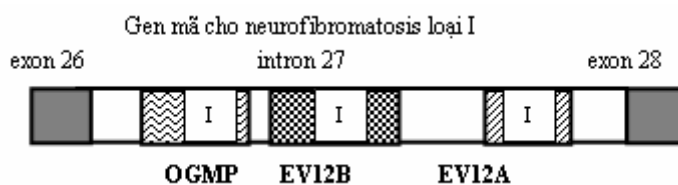
Phản ứng gắn nhóm methyl vào cytosine được xúc tác bởi các enzym methyltransferase. Có thể phân biệt các enzym này thành 2 nhóm. Nhóm thứ nhất làm nhiệm vụ duy trì gốc methyl ở những vị trí cytosine trên sợi ADN vừa được tổng hợp trong quá trình tái bản ADN. Việc gắn gốc methyl mới này dựa vào nhóm ^mCpG 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 có thể đảm nhận phản ứng khử nhóm methyl. 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 ^mC bị loại đi và thay thế bởi cytosine không mang nhóm methyl.

1.4 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 (*interrupted gene*). Hiện tượng này được khám phá lần đầu tiên năm 1977 và được tìm thấy phổ biến ở mọi sinh vật eukaryot. Kỳ lạ là 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 những đoạn mang mã di truyền (gọi là exon), đa số các gen eukaryot còn chứa các intron. Mặc dù không chứa mã di truyền và bị cắt đi khỏi phân tử ARNm, đột biến xảy ra ở intron có thể ngăn cản phản ứng nối các exon với nhau, do đó tạo nên phân tử ARNm sai hỏng không sử dụng được để dịch mã tổng hợp protein.

Khi phân tử ARN được phiên mã từ một gen, nó phải trải qua quá trình loại bỏ các intron, nối các exon với nhau (phản ứng *splicing*). Phản ứng cắt nối này xảy ra với các loại ARNm, ARNr và ARNt. Để tạo ra phân tử ARN hoàn thiện, việc cắt intron, nối các exon tuân theo những qui luật nghiêm ngặt và chính xác để đảm bảo thứ tự của chúng. Điều thú vị là các exon của một phân tử ARNm được nối với nhau. Hiếm trường hợp nối các exon của các phân tử ARNm khác nhau. 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 các đột biến có thể ảnh hưởng đến phản ứng *splicing* khi chúng xảy ra ở các vị trí cần thiết để cắt nối intron-exon. 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 ra ARNm nhưng chúng không được dịch mã. Những phân tử ARNm này vẫn trải qua phản ứng cắt nối intron-exon để tạo ra các đoạn ARN ngắn. Chúng tiếp tục được phân huỷ thành các phân tử ARN kích thước nhỏ 22-25 nucleotides (*miRNAs*: *micro RNAs*). 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 huỷ 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.

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 200 bp trong khi của intron có thể lớn hơn 10 kb hoặc thậm chí đạt tới 50-60 kb. 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. Hơn nữa, một gen này có thể nằm trong một gen khác, tức là gen thứ hai được phân bố trong intron của gen thứ nhất. Ví dụ, điển hình cho trường hợp gen trong gen (*genes-within-genes*) ở genome của người là gen mã cho neurofibromatosis loại I. Intron 27 của gen này có chứa 3 gen khác, mỗi gen đó đều có exon và intron riêng của mình (Hình 1.8).



Hình 1.8:

Cấu trúc gen trong gen ở intron 27 của gen mã cho neurofibromatosis. Intron 27 có chứa 3 gen nhỏ OGMP, EV12B và EV12A. Mỗi gen này đều có intron (I) và exon (phần sẫm màu).

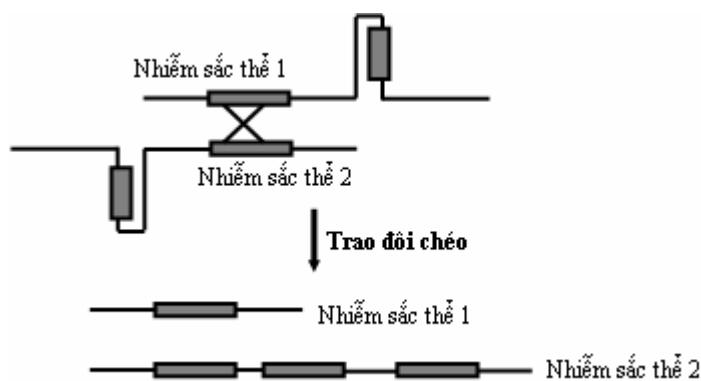
Có thể phân loại các gen tùy theo cấu trúc của gen hoặc theo chức năng của các sản phẩm do chúng mã hoá. Genome đơn bội ở các cơ thể đa bào có khoảng 1/4 đến 1/2 số gen mã cho protein là các gen đơn lẻ (*single copy gene*), không tồn tại bản sao thứ hai. Số gen còn lại thường tồn tại hai hoặc nhiều bản sao trong genome. Các bản sao của một 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 hình thành từ một gen tổ tiên được xếp vào một họ gen (*family genes*). Các gen trong cùng một họ có thể 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 mặc dù các gen này thường hoạt động ở những thời điểm nhất định và trong các loại tế bào biệt hoá khác nhau. Ví dụ, việc tổng hợp các protein globin (được mã bởi các gen trong cùng một họ gen) xảy ra ở những giai đoạn nhất định trong quá trình phát triển phổi

thai và ở cơ thể trưởng thành. 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, gen mã cho ARNr, ARNt, histone vv... Như vậy, các gen eukaryot có thể phân thành các loại chính như sau: gen đơn lẻ, các gen thuộc một họ gen, gen lặp đi lặp lại liên tục và các pseudogen.

1.4.1. Các gen trong cùng một họ gen

Cho đến nay, hầu hết các gen mã cho protein được nghiên cứu ở sinh vật eukaryot đều không phải là những gen đơn lẻ. 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).

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 1.9).

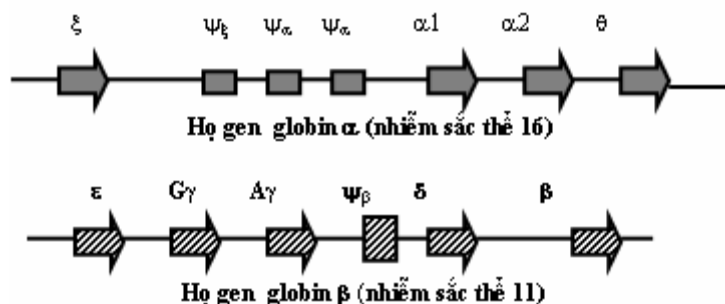


Hình 1.9:

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.

Sản phẩm của các thành viên trong một họ gen có chức năng giống nhau nhưng thường được sử dụng ở những thời điểm phát triển khác nhau hoặc trong các loại tế bào biệt hoá khác nhau. Trình tự acid amin của chúng chỉ tương tự mà không giống nhau hoàn toàn. 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.

Các gen globin là thí dụ điển hình về một họ gen (Hình 1.10). Ở mọi loài động vật, các gen này có cấu trúc tương tự do chúng có cùng nguồn gốc từ một gen tổ tiên. Tế bào trong cơ thể trưởng thành có globin tồn tại ở dạng tetramer gồm hai chuỗi polypeptide α và hai chuỗi β . Các gen mã cho các chuỗi này nằm trên hai nhiễm sắc thể khác nhau. Do đó hoạt động của chúng 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. 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. Ngoài ra, trong mỗi họ còn có các pseudogen (gen giả) và một số thành viên khác mà sản phẩm của chúng đôi khi vẫn được sử dụng.

**Hình 1.10:**

Họ gen globin α và β ở người tập trung thành các nhóm trên hai nhiễm sắc thể. Chúng gồm các gen mã cho globin và các pseudogen (ψ). Các gen hoạt động theo trình tự từ trái sang phải phù hợp với quá trình phát triển từ phôi đến cơ thể trưởng thành.

Họ gen globin α chiếm 28 kb trên nhiễm sắc thể 16, gồm các gen ξ , $\alpha 1$, $\alpha 2$ và θ . Sản phẩm của hai gen $\alpha 1$, $\alpha 2$ giống hệt nhau. Họ gen globin β chiếm 50 kb trên nhiễm sắc thể 11 gồm 5 gen hoạt động (ϵ , $G\gamma$, $A\gamma$, δ , β) và một pseudogen $\psi\beta$. Sản phẩm của hai gen γ chỉ khác nhau duy nhất ở một acid amin tại vị trí 136 (Glicine và Alanine). 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 1.1).

Bảng 1.1:

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

Giai đoạn phát triển	Hemoglobin
Mô phôi (8 tuần)	$\xi 2\epsilon 2$, $\xi 2\gamma 2$, $\alpha 1\epsilon 2$
Thai nhi (3-9 tháng)	$\alpha 2\gamma 2$
Cơ thể trưởng thành (từ khi sinh)	$\alpha 2\delta 2$ (~ 2%), $\alpha 2\beta 2$ (~97%), $\alpha 2\gamma 2$ (~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.

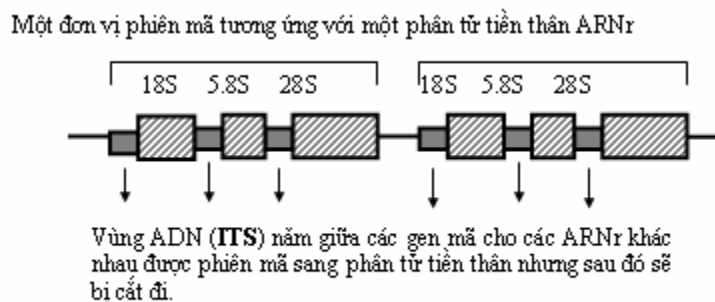
1.4.2. Gen lặp đi lặp lại liên tục

Thông thường các thành viên trong một họ gen không giống nhau hoàn toàn. Sự sai khác giữa chúng đảm bảo tính hoạt động độc lập của từng gen và được duy trì qua chọn lọc. Tuy nhiên cũng có một vài trường hợp cá biệt, số lượng các thành viên trong họ rất lớn và chúng giống hệt nhau, thường tập hợp thành các nhóm phân bố trên các nhiễm sắc thể khác nhau. Mỗi nhóm có thể bao gồm từ hai cho 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 như cần đáp ứng kịp thời các phân tử ARNr cho giai đoạn sinh trưởng nhanh (phôi) hoặc các loại protein histone cho quá trình tái bản ADN.

Gen mã cho ARNr: ARN ribosome chiếm 80-90% tổng số ARN có trong tế bào. Số gen mã cho chúng 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 ADN_r). 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 1.11).

Một 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 ARNr này sẽ bị phân hủy. Mỗi một nhóm gen mã cho ARNr gồm nhiều gen giống hệt nhau, khoảng cách giữa mỗi gen thay đổi tùy 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 30 kb. Khoảng cách này có vai trò trong khởi động quá trình tổng hợp ARNr hoặc giúp cho ARN polymerase dễ dàng bám vào promoter.

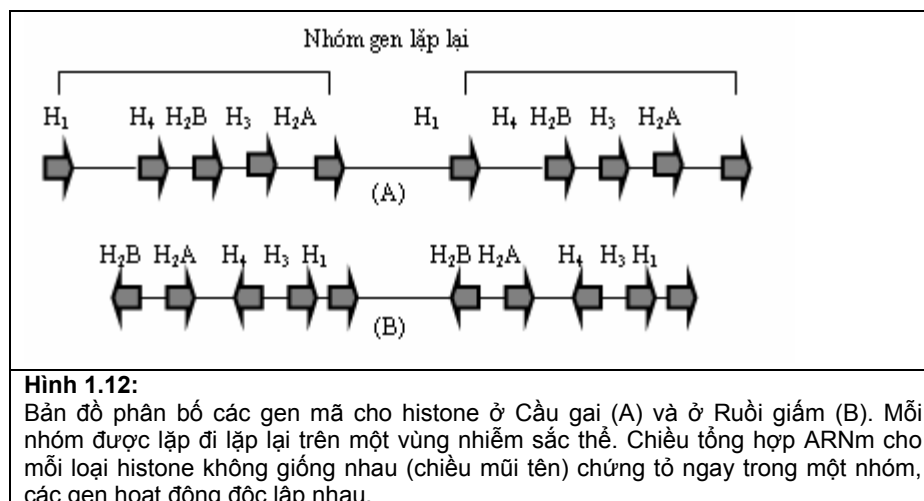


Hình 1.11:

Một đơn vị phiên mã (một gen mã cho ARNr) mang thông tin di truyền cho các phân tử ARNr 18S, 5.8S và 28S. Gen này được lặp đi lặp lại liên tục. Khoảng cách giữa các gen thay đổi tùy theo từng loài sinh vật.

Gen mã cho protein histone: Protein histone tham gia liên kết với ADN để hình thành cấu trúc nucleosome. Có bốn loại histone khác nhau. Histone H2A, H2B, H3 và H4 tương tác với nhau tạo cấu trúc lõi. Lõi này được quấn quanh bởi đoạn ADN 146 bp tạo thành nucleosome. Histone H1 liên kết với ADN linker 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 5-6 kb (ở động vật có xương sống) (Hình 1.12). Cũng giống như nhóm gen mã cho ARNr, khoảng cách giữa các gen trong cùng một nhóm và giữa các nhóm thay đổi giữa các loài, thậm chí ngay trong từng cá thể.

Có thể phân biệt các gen mã cho histone thành hai nhóm. Nhóm 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.



1.4.3. 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 enzyme 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, không có các đoạn nucleotide 5' và 3' nằm trước mã khởi đầu và nằm sau mã kết thúc phản ứng tổng hợp protein. Hai đoạn trước và sau này được gọi là đoạn không dịch mã (*5' and 3' untranslated regions*).

1.5 Thành phần ADN lặp lại trong genome eukaryot

1.5.1. ADN vệ tinh (satellite DNA) và ADN tiểu vệ tinh (minisatellite DNA)

Bên cạnh các họ gen và các gen lặp đi lặp lại liên tiếp, genome trong tế bào eukaryot còn chứa những vùng ADN gồm các oligonucleotide (thường từ 5, 10 đến 150, 300 bp) được lặp đi lặp lại rất nhiều lần. Điều đó tạo ra những đặc tính vật lý riêng biệt của loại ADN này. Dựa vào đó người ta có thể phân đoạn và tách chúng ra khỏi ADN genome. Chúng được gọi là các

ADN vệ tinh (*DNA satellite*). Tỷ lệ ADN vệ tinh thay đổi giữa các loài chiếm từ 10 đến 30% hệ gen.

Trong hầu hết tế bào động vật có vú, ADN vệ tinh thường tập trung xung quanh tâm động (*centromere*) và vùng cuối hai đầu nhiễm sắc thể (*telomere*). Sự phân bố của chúng ở đó có vai trò nhất định trong quá trình phân chia tế bào và đảm bảo độ dài của telomere qua các lần tái bản ADN. Khi các nhiễm sắc thể phân ly về hai cực trong phân bào, các protein đặc hiệu bám dính vào những vị trí đặc biệt ở tâm động để kiểm tra, điều khiển sự di chuyển đó. ADN vệ tinh giữ vai trò của những vị trí đặc biệt này. Nói chung chúng không được phiên mã sang phân tử ARN. Ngoài ra, ADN vệ tinh ở tâm động được nhân bản cuối cùng trong quá trình tái bản nhiễm sắc thể. Rất có thể hiện tượng lặp đi lặp lại của một loại ADN tại tâm động nhằm ngăn cản sự xuất hiện tâm tái bản tại vị trí này.

Ở côn trùng ADN vệ tinh thường bao gồm các đoạn nucleotide rất ngắn (khoảng 5-15 bp), còn ở động vật có vú thành phần này đa dạng hơn và thường phân bố thành từng nhóm trên nhiễm sắc thể. Ở người, có ít nhất hơn 10 loại ADN vệ tinh. Mỗi loại có thể chiếm tới 0,5-1% tổng số genome, tương đương khoảng 10^7 bp. Đối với từng cá thể riêng biệt, trong mỗi loại ADN vệ tinh, các đoạn oligonucleotide có thể lặp lại hoàn toàn chính xác như nhau hoặc có thể xảy ra sự thay thế, loại bỏ hay thêm vào một vài nucleotide. Tuy nhiên những biến đổi này phụ thuộc từng vùng trên nhiễm sắc thể. Chức năng của ADN vệ tinh phân bố rải rác trong genome chưa được sáng tỏ. Những năm cuối của thập kỷ 20, sinh học hiện đại đã chứng minh được các đoạn lặp lại phân bố gần hoặc nằm ngay trong gen có vai trò kiểm soát hoạt động của gen đó. Thông thường các đoạn ADN lặp lại không được phiên mã. Chúng bị bất hoạt do các cytosine và histone H3 bị methyl hoá ở lysine 9 nhưng histone H4 bị khử nhóm acetyl.

Khi các oligonucleotide gồm khoảng 25-50 bp được lặp lại nhiều lần chiếm một đoạn ADN từ 1 đến 5 kb, thậm chí đến 20 kb thì chúng được gọi là ADN tiểu vệ tinh (*minisatellite DNA*) hoặc ADN lặp lại ngẫu nhiên đa hình VNTR (*variable number tandem repeat*). Tương tự như ADN vệ tinh, việc tồn tại của ADN tiểu vệ tinh có liên quan đến cấu trúc nhiễm sắc thể bởi vì loại ADN này thường bắt gặp ở telomere. Tuy nhiên, chức năng của ADN tiểu vệ tinh phân bố rải rác trong genome chưa được làm sáng tỏ.

Ngoài ra khi số nucleotide rất ít (1-4 bp) được lặp lại nhiều lần thành từng đoạn khoảng 200 bp thì chúng được gọi là ADN vi vệ tinh (*microsatellite DNA*). ADN vi vệ tinh thường bao gồm 1 đến 4 nucleotide lặp lại khoảng 10 đến 20 lần. Số lượng loại ADN này rất lớn trong genome, vì vậy chúng được dùng làm chỉ thị phân tử trong việc xác định vị trí của gen trên bản đồ. Ví dụ, trong genome người, ADN vi vệ tinh CA (CACACA...) lặp đi lặp lại chiếm khoảng 0,5% (15Mb), trong khi sự lặp lại của một nucleotide A (AAA...) cũng chiếm đến 0,3%.

Mặc dù chức năng của ADN vi vệ tinh chưa được biết nhưng chúng có một ý nghĩa rất quan trọng trong lập bản đồ toàn bộ genome. Trong mỗi một quần thể, các ADN vi vệ tinh tương tự như nhau, tuy nhiên số lần lặp lại cũng như những biến đổi trong mỗi loại phụ thuộc vào từng cá thể. Nói một cách khác, mỗi loại tiểu vệ tinh tồn tại trong mọi cá thể của quần thể, nhưng số lần lặp lại cũng như các biến đổi trong trình tự nucleotide lại đặc trưng cho từng cá thể. Tính chất này được áp dụng để phân biệt các cá thể khác nhau và phân tích quan hệ huyết thống (kỹ thuật *DNA-fingerprinting*...).

1.5.2. Các đoạn ADN có khả năng di chuyển

Tần số trao đổi chéo giữa các ADN tiểu vệ tinh lớn hơn khoảng 10 lần so với trao đổi chéo xảy ra giữa các đoạn nhiễm sắc thể tương đồng trong phân bào giảm nhiễm. Đó là một trong những nguyên nhân tạo ra sự khác biệt giữa genome của các cá thể trong một loài. Ngoài ra sự đa dạng của genome còn do các đoạn ADN có khả năng di chuyển (thường được gọi là *transposon*).

Khi di chuyển, các transposon gây ra việc sắp xếp, tổ chức lại genome của từng cá thể như tạo các đoạn ADN mới hoặc thay đổi chức năng hoạt động của các đoạn ADN ở vị trí chúng ghép vào và tách ra. Chúng có thể di chuyển tới vị trí bất kỳ và hoàn toàn không yêu cầu mối quan hệ nào giữa hai vị trí mới và cũ. Khi tách ra 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 transposons (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 H3, deacetyl histone H4 vv....

Cách thức di chuyển và ghép vào genome của các đoạn ADN đặc biệt này 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 retroelement 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.

Trong số các loại retroelement, cần lưu ý đến yếu tố ERVs (*endogenous retrovirus*) và các retrotransposons. Chúng đều là những đoạn ADN có khả năng di chuyển trong genome. Tuy nhiên ERVs có chung một đặc điểm là hai đầu được tận cùng bởi hai đoạn nucleotide lặp lại với kích thước lớn (*long terminal repeat-LTRs*). LTRs giữ vai trò quyết định trong quá trình di chuyển. Ngoài ra, retrotransposons bao gồm các yếu tố LINES (*Long Interspersed Nuclear Elements*) hoặc SINES (*Short Interspersed Nuclear Elements*) là những đoạn lặp lại dài hoặc ngắn phân bố rải rác trên các nhiễm sắc thể. Yếu tố LINES không chứa LTRs nhưng có mang gen mã cho reverse transcriptase trong khi SINES không có gen đó nhưng có khả năng "vay mượn" enzym này do các retroelements khác tổng hợp. Trong genome của người, yếu tố LINE-1 có tới 3500 bản sao dài nguyên vẹn 6,1 kb và hàng trăm nghìn bản sao có kích thước ngắn hơn. Bên cạnh đó trình tự *Alu* gồm hàng triệu bản sao là ví dụ điển hình của yếu tố SINES. Mặc dù phân tử ARN được tổng hợp từ *Alu* nhưng sản phẩm protein không được tạo thành. Dù sao sự tồn tại của các ARN này cũng làm tăng cơ hội giúp *Alu* ghép vào genome.

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 retroelement. Ví dụ, ở genome người, chỉ có

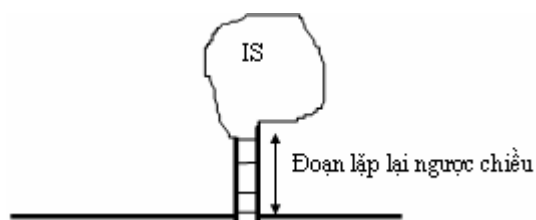
khoảng 100 loại ADN transposon. Tuy nhiên, ADN transposon có một ý nghĩa đặc biệt quan trọng đối với sự đa dạng hoá genome. Một số transposon có mặt trong genome của các loại sinh vật khác nhau. Ví dụ, yếu tố *mariner* có chiều dài 1250 bp được tìm thấy ở ruồi giấm *Drosophila* cũng như rất nhiều động vật khác, kể cả người. Phải chăng các transposon này có thiên chức tự nhiên trong tiến hoá là chuyên chở gen giữa các genome khác nhau?

Các transposon có chung đặc điểm là hai đầu tận cùng của mỗi transposon có chứa hai đoạn oligonucleotide lặp lại ngược chiều (*inverted repeats*). Các transposon có thể chia làm hai loại dựa vào khả năng di chuyển độc lập hay phải phụ thuộc vào sự có mặt của transposon khác.

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

*Loại thứ hai gồm các transposon không có khả năng tự hoạt động, tức là chúng không có khả năng di chuyển do không chứa gen mã cho các enzym cần thiết. Việc di chuyển của transposon ở loại này phụ thuộc vào sự có mặt của transposon có khả năng hoạt động độc lập (*transposon* nhóm 1) cùng nhóm. Hai transposon có thể xếp vào cùng nhóm khi chúng có cấu trúc tương đồng với nhau, đặc biệt là các đoạn oligonucleotide phân bố ở hai đầu transposon. Đây là vị trí để enzym nhận biết và cắt nối transposon ở vị trí cũ và mới. Khi các transposon loại này di chuyển, chúng tạo ra những đột biến bền vững nếu như trong thế hệ nối tiếp chúng đã phân ly độc lập (phân ly theo định luật Mendel) với transposon có khả năng hoạt động độc lập cùng nhóm.

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, $\lambda :: IS1$ 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). Do đó khi sợi đúp IS tách thành hai sợi đơn thì mỗi sợi này có khả năng hình thành liên kết bổ sung tại hai đầu của IS tạo cấu trúc dạng vòng (*stem-loop*) (Hình 1.13).

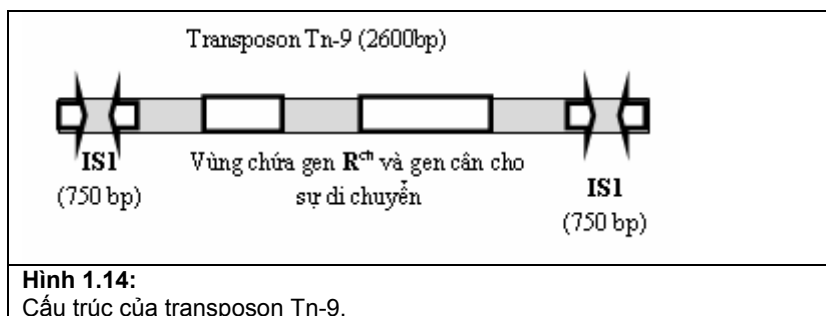


Hình 1.13:

Cấu trúc dạng vòng được tạo ra do liên kết tạo cặp bổ sung giữa hai trình tự lặp lại ngược chiều của một IS trên một sợi đơn ADN.

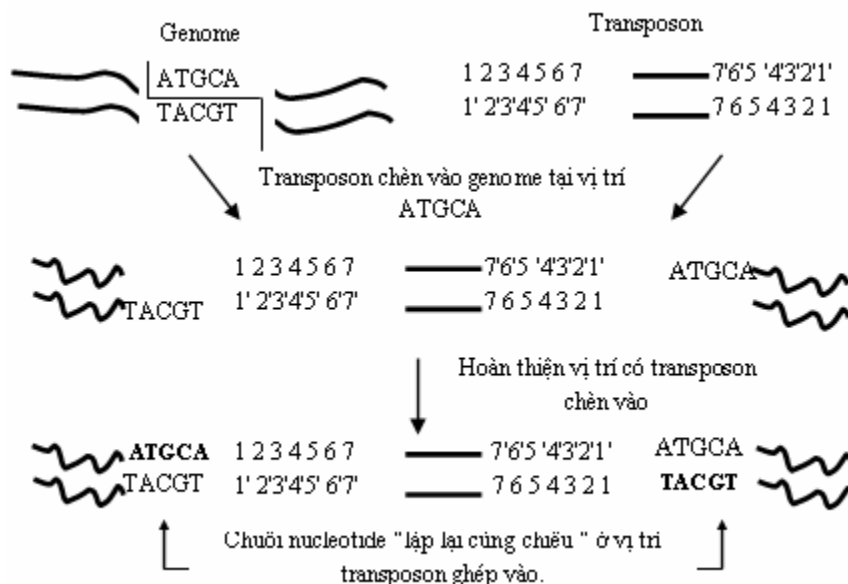
Ngoài các IS, ở vi khuẩn còn có các đoạn ADN có khả năng di chuyển với kích thước dài hơn, gọi là transposon Tn. Các Tn thường phân bố trên plasmid (phân tử ADN dạng vòng, kích thước thường không lớn) và có khả năng ghép xen vào bất kỳ vị trí nào trong genome. Chúng thường mang thông tin di truyền mã 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.



Hình 1.14 mô tả cấu trúc của transposon Tn-9. Transposon này mang hai gen; một mã cho tính chống chịu chloramphenicol (R^{ch}) và gen kia mã cho protein cần thiết cho sự di chuyển. Hai đầu của Tn-9 được giới hạn bởi IS-1 mà trình tự nucleotide của IS này sắp xếp theo cùng một chiều.

Một số transposon chứa gen mã cho các enzym transposase làm nhiệm vụ nhận biết chuỗi nucleotide lặp lại ngược chiều (*inverted repeat*) để cắt transposon. ADN của vị trí mới bị cắt sao cho mỗi sợi đơn lệch nhau vài nucleotide (cắt thành đầu so le). Transposon nối vào các đầu cắt, tạo ra hai khoảng trống (gaps). Khoảng trống được sửa chữa theo nguyên tắc tạo cặp bổ sung. Do đó các nucleotide của đầu so le ở vị trí mới được sao chép thành hai bản, mỗi bản ở một đầu và trình tự sắp xếp các nucleotide giống nhau. Vì vậy chúng được gọi là lặp lại cùng chiều (*direct repeat*) (Hình 1.15). Chiều dài của chúng thường khoảng 7-9 bp. Dựa vào sự có mặt của các đoạn cùng chiều và ngược chiều có thể xác định được vị trí transposon ghép vào hoặc chuyển đi.

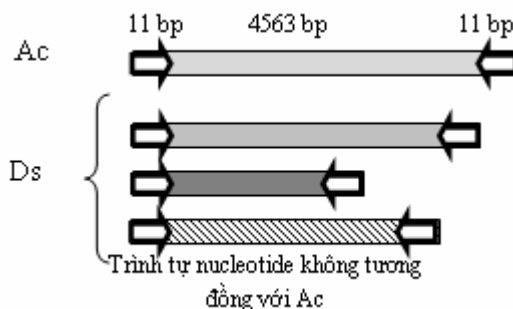


Quá trình di chuyển của một transposon từ vị trí cũ (*donor*) sang vị trí mới (*recipient*) xảy ra theo hai cơ chế khác nhau: Cơ chế sao y bản chính (*transposon* có mặt ở cả hai vị trí) và cơ chế tách ra khỏi vị trí cũ di chuyển đến vị trí mới. Trong cơ chế thứ nhất, trình tự nucleotide

của transposon được sao chép từ vị trí cho và được ghép vào vị trí nhận. Như vậy mỗi lần di chuyển thì số lượng bản sao được tăng lên. Quá trình này liên quan đến hai loại enzym: transposase (tác động vào hai đầu bản gốc *transposon*) và resolvase (tác động lên bản sao). Trong cơ chế thứ hai, một transposon có thể tách ra khỏi vị trí cũ và ghép vào vị trí mới. Như vậy số lượng transposon không thay đổi. Kiểu di chuyển này chỉ đòi hỏi enzym transposase. Khi transposon chuyển đi, vị trí cũ bị gãy. Nó được nối lại nhờ cơ chế sửa chữa ADN trong tế bào.

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

Hai loại transposon Ac và Ds được nghiên cứu khá kỹ ở ngô. 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 4563 bp, được giới hạn hai đầu bởi 11 bp lặp lại ngược chiều, tiếp đến 8 bp lặp lại cùng chiều của genome. Mọi Ds đều có đoạn lặp lại ngược chiều giống nhau mặc dù chiều dài của chúng thay đổi (Hình 1.16).



Hình 1.16:

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

Các transposon ở ngô thường ghép vào gần các gen, làm rối loạn hoạt động của chúng dẫn đến việc xuất hiện tính trạng mới nhưng không gây đột biến chết. Sự di chuyển của transposon ghép vào vị trí allel của một gen bất kỳ trên nhiễm sắc thể xảy ra ở tế bào soma sẽ tác động đến biểu hiện của allel đó trong quá trình phát triển của cây. Trải qua phân bào nguyên nhiễm (*mitose*), con cháu của tế bào chứa allel 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*, yếu tố P có khả năng di chuyển được phát hiện khi tiến hành lai giữa con đực dòng P với con cái dòng M. Hầu hết con lai bị bất dục, nhiễm sắc thể bị đứt gãy, bị 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 yếu tố di chuyển P. Yếu tố dài nhất gồm có 2907 bp có chứa gen mã cho transposase. Điều đáng chú ý là mặc dù có chiều dài khác nhau, các yếu tố P đều có mang các trình tự nhận biết bởi transposase.

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

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 do yếu tố trong tế bào trứng ngăn cản P chuyên chỗ. Tuy nhiên, tế bào trứng dòng M thụ tinh với con đực dòng P cho phép P di chuyển gây ra những rối loạn bất thường trong cấu trúc genome. Điều đó khiến con lai bị bất dục hoặc xuất hiện các tính trạng lạ.

1.6 Tương tác của T-ADN với genome thực vật

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

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

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

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

b/ Plasmid Ti phải mang vùng *vir* - ADN (nằm ngoài đoạn T-ADN). Vùng này mang các gen cần thiết cho việc tách và vận chuyển T-ADN từ vi khuẩn sang tế bào thực vật. Vi khuẩn xâm nhiễm vào tế bào cây chủ tại vị trí tổn thương trên thân cây. Cây có vết thương do sự hư hỏng ngẫu nhiên của màng tế bào thực vật hoặc do vi khuẩn tiết ra hỗn hợp những chất được mã bởi các gen *vir*. Hoạt động của các gen này được hoạt hoá bởi hợp chất phenolic của cây (ví dụ như acetosyringone, catechol, các dẫn xuất của chalcone...). Ngoài ra, sự có mặt các monosaccharides như glucose, arabinose trong môi trường cũng khiến cho nhóm gen *vir* của vi khuẩn nhạy cảm hơn với các hợp chất phenolic do cây tiết ra. Sản phẩm của những gen trên vùng *vir* còn liên quan chủ yếu đến việc cắt T-ADN ra khỏi plasmid và vận chuyển nó vào tế bào chủ. Bằng các thí nghiệm bổ sung chức năng (*complementation test*), thực nghiệm đã phát hiện ít nhất có 21 polypeptide sản phẩm của các gen *vir* cũng như xác định được chức năng của hầu hết các protein này trong quá trình vận chuyển T-ADN. Protein VirA đóng vai trò quan trọng trong việc qui định tính đặc hiệu giữa các loại cây chủ với *Agrobacteria*. Trong thực tế, *Agrobacteria* không có khả năng xâm nhập vào cây một lá mầm. Có thể protein VirA không nhận biết được các tín hiệu do cây một lá mầm tiết ra. Protein VirC1 nhận biết và tương tác với các nucleotide nằm ở đầu bên phải của T-ADN. Mặc dù hai đầu T-ADN có trật tự tương đối giống nhau (chỉ sai khác nhau 2 nucleotide trong tổng số 25 nucleotide cần thiết cho sự vận chuyển T-ADN) nhưng các nucleotide đầu bên phải giữ vai trò quyết định cắt T-ADN ra khỏi plasmid. Đột biến ở đầu này khiến T-ADN không được cắt ra khỏi plasmid trong khi đột biến đầu bên trái hoàn toàn không ảnh hưởng đến quá trình vận chuyển T-ADN từ tế bào vi khuẩn vào trong nhân tế bào cây chủ. Điều đó cho thấy việc cắt T-ADN được bắt đầu ở phía bên phải và tiến dần sang bên trái. Điều đặc biệt lưu ý là chỉ có một sợi đơn T-ADN được cắt ra và vận chuyển sang tế bào thực vật. Sợi đơn đó được gọi là sợi T. Đầu 5' của sợi T tương ứng với đầu bên phải của đoạn T-ADN. Các protein VirD1 và VirD2 liên quan đến phản ứng cắt sợi T ra khỏi plasmid. Tiếp theo đó, protein VirE2 tương tác với sợi T dọc theo chiều dài của sợi. Protein VirD2 giữ vai trò quan trọng trong quá trình vận chuyển.

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

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

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

T-ADN gồm hai nhóm gen. Nhóm thứ nhất gồm các oncogen mà cơ chế hoạt động của chúng khác biệt giữa *A. tumefaciens* và *A. rhizogenes*. Điều đó dẫn đến sự hình thành các nốt sần hoặc bệnh lông rể. Trong trường hợp xuất hiện nốt sần, T-ADN mang ba oncogen mã cho các enzym tham gia vào phản ứng tổng hợp các hocmon sinh trưởng auxin và cytokinin. Chỉ khi T-ADN được ghép vào genome thực vật, các oncogen nằm trên T-ADN mới hoạt động một cách tự động. Do đó tế bào cây chủ nào có T-ADN ghép vào hệ gen lập tức phát triển không bình thường do rối loạn hocmon sinh trưởng mà T-ADN mã cho. Nốt sần xuất hiện tại vị trí cây bị nhiễm *A. tumefaciens* là tập hợp của các tế bào bình thường và tế bào bị biến đổi hệ gen.

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

Nhóm gen thứ hai có mặt trên đoạn T-ADN gồm các gen mã cho các enzym tham gia tổng hợp những phức chất dinh dưỡng cần thiết cho sinh trưởng và phát triển của vi khuẩn. Các phức chất này 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 genome thực vật.

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

chủ, các gen trên đoạn T-ADN mới được hoạt động. Như vậy điều đáng chú ý là các gen trên T-ADN (*prokaryot*) chỉ hoạt động dưới sự kiểm soát của các yếu tố phiên mã trong genome eukaryot. Nói chung, một gen được điều khiển bởi một promoter.

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

1.7 ADN trong ty thể và lục lạp

Đối với tế bào eukaryot, ADN không chỉ phân bố trong nhân mà còn có mặt ở ty thể và lục lạp. Hầu hết phân tử ADN trong các bào quan này ở dạng mạch vòng. Tuy nhiên cũng có một số trường hợp 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 chứa nhiều ty thể hoặc lục lạp. Hơn nữa, mỗi bào quan có thể có 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 mỗi tế bào người có tới 8000 phân tử ADNmt, trong đó một ty thể có khoảng 10 phân tử. Tế bào trứng của động vật có vú có chứa tới 10^8 bản sao của ADNmt. 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. Phân tử ADNmt có thể thay đổi rất rộng từ 16-17 kb ở động vật có xương sống đến 2500 kb ở một số thực vật có hoa.

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 như các protein tham gia chuỗi hô hấp. Ngoài ra, ADN trong ty thể và lục lạp còn chứa gen mã cho ARNr, ARNt và protein ribosome dùng riêng cho bào quan.

1.7.1. ADN ty thể

Trình tự nucleotide của phân tử ADNmt ở một số sinh vật đã được xác định. Kết quả này giúp chúng ta hiểu rõ hơn cấu trúc và trật tự sắp xếp các gen trên phân tử ADN của bào quan. Ở động vật có xương sống, ADN ty thể 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 16.659 bp 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 (78.000 bp) 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 thủy không có hệ mao dẫn, đã được xác định hoàn toàn. Đây là phân tử mạch vòng có kích thước 186 kb 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 570 kb và 300 kb. 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.

1.7.2. ADN lục lạp

Thực vật có ba loại lục lạp khác nhau tùy 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 292 kb ở tảo và 120 đến 160 kb ở thực vật bậc cao. Đặc biệt ở một số thực vật như tảo xanh *Acetabularia*, ADNcp lớn đến 2000 kb. Phân tử ADNcp của một số thực vật đã được xác định trình tự nucleotide. Lục lạp thuốc lá *Nicotiana tabacum* có ADNcp gồm 155.844 bp 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.

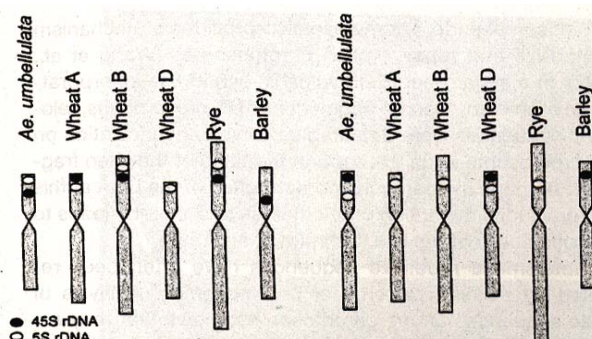
1.8 Genomics

1.8.1 So sánh genome

Dựa vào trình tự nucleotide của một số genome điển hình, các nhà sinh học có thể phân tích cấu trúc, hoạt động và chức năng của các gen, làm sáng tỏ được vai trò của ADN lặp lại, ADN nằm giữa các gen, ADN không chứa mã di truyền (các vùng 5' và 3' không được dịch mã) và các đoạn intron của từng gen vv... Điều đặc biệt có ý nghĩa là khi so sánh các genome với nhau, chúng ta có được những hiểu biết tổng quan về hoạt động của genome ở các sinh vật khác nhau, mối quan hệ giữa chúng, sự đa dạng sinh học và mức độ tiến hoá. Ví dụ, toàn bộ trình tự nucleotide của genome *Arabidopsis* được xác định cuối năm 2000 nhằm mục đích phát hiện, phân lập các gen quan trọng của các cây nông nghiệp dựa vào sự tương đồng của chúng với các gen của *Arabidopsis*. Đây là thực vật đầu tiên có genome được xác định toàn bộ trình tự do kích thước genome tương đối nhỏ (130-140 Mbp, nhỏ hơn khoảng 200 lần so với các thực vật khác). Bộ nhiễm sắc thể đơn bội của *Arabidopsis* gồm 5 nhiễm sắc thể. Ngoài ra, *Arabidopsis* có vòng đời ngắn, dễ trồng và có thể mọc quanh năm. Hình dáng cây nhỏ chiếm rất ít diện tích nên hoàn toàn thích hợp với điều kiện nuôi trồng trong phòng thí nghiệm.

Trình tự nucleotide của genome ở các sinh vật mô hình được đưa vào các loại ngân hàng ADN khác nhau tùy thuộc vào mục đích nghiên cứu. 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 Institute*), 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*). Bên cạnh trình tự toàn bộ hệ gen, các loại ADN khác như cDNA, ADN đích-ESTs (*Expressed Sequence Tags*) vv... cũng được lưu giữ phục vụ cho việc so sánh, phân tích và xác định chức năng của genome, của gen và sản phẩm (protein hoặc ARN) tương ứng.

So sánh genome giữa các loài sinh vật với nhau cho phép rút ra ba đặc điểm nổi bật: Thứ nhất là số lượng nhiễm sắc thể rất khác nhau ngay giữa những loài rất gần nhau. Thứ hai là các gen thường phân bố không theo qui luật. Một gen hoặc một họ gồm nhiều gen mã cho sản phẩm cùng chức năng có thể phân bố trên các nhiễm sắc thể khác nhau, nằm thành nhóm hoặc rải rác trong genome. Ví dụ, sự phân bố của gen mã cho ARNr được trình bày trên hình 1.17. Thứ ba là kích thước genome thay đổi không hoàn toàn tỷ lệ với tính phức tạp của loài. Nhìn chung, kích thước genome thường phản ánh tính phức tạp của loài. Tuy nhiên, điều đó không đồng nghĩa giữa việc tăng số lượng các gen với mức độ tiến hoá. Chỉ khi so sánh trình tự toàn bộ genome của một số sinh vật cũng như hoạt động của một số gen quan trọng trong sinh trưởng phát triển mà các nhà sinh học mới nhận thấy tính phức tạp liên quan chủ yếu đến việc tăng số lượng các đoạn ADN lặp lại. Ví dụ, genome của một số loài lưỡng cư hoặc thực vật có kích thước khoảng 10^{11} bp, trong đó thành phần ADN lặp lại chiếm hơn 60-70%. Genome của người nhỏ hơn, chỉ khoảng 3×10^9 bp. Chắc chắn rằng chỉ riêng kích thước genome không thể quyết định tính phức tạp hay mức độ tiến hoá của các loài.



Hình 1.17:
Phân bố của ADNr tương ứng với ARNr 45S và ARNr 5S trong các loài Triticeae

Bên cạnh so sánh tổng thể toàn bộ genome giữa các loài, việc phân tích chi tiết đối với một gen nhất định còn liên quan đến vị trí các intron, các exon, các đoạn ADN điều khiển hoạt động của gen. Đây 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 ADN riêng biệt trong các bào quan (ty thể, lục lạp) sang genome trong nhân đều chịu ảnh hưởng của thời gian, tức là đều phản ánh quá trình tiến hoá của các loài. Mặt khác, để có được sự so sánh chính xác hơn, toàn diện hơn, cần xét đến cấu trúc sợi nhiễm sắc, cấu hình không gian ba chiều của nhiễm sắc thể cũng như của toàn bộ genome phân bố trong nhân.

1.8.2 Genome người

Dự án xác định trình tự genome người (hệ gen trong nhân) được đề cập đến từ những năm 1984-1988. Dự án được bắt đầu vào đầu thập kỷ 90 với sự tham gia của hơn 20 nhóm nghiên cứu từ các nước Mỹ, Nhật, Đức, Anh, Pháp và Trung quốc do tổ chức quốc tế Genome Người (*Human Genome Organization-HUGO*) và công ty tư nhân *Celera Inc.* cùng tiến hành độc lập với nhau. Dự án được triển khai với ba bước cơ bản: thứ nhất là lập bản đồ của tất cả các gen (khoảng 70.000 đến 100.000 gen), tiếp đến là xác định bản đồ vật lý của 24 nhiễm sắc thể ở mức độ chi tiết nhất (mà các kỹ thuật hiện đại có thể đáp ứng được) và cuối cùng là đọc trình tự nucleotide của toàn bộ genome.

Genome người được xem gồm có hai phần phân bố trong nhân và trong ty thể. Phân tử ADN ty thể có dạng vòng với kích thước 16.569 bp. Kích thước này quá nhỏ, có thể coi là

không đáng kể so với genome trong nhân. Tuy nhiên, do ty thể không có cơ chế sửa chữa ADN nên các đột biến (thêm, mất hoặc đảo đoạn) thường được tích lũy trong phân tử ADN của bào quan này. Mặt khác, mỗi tế bào có khoảng 800 ty thể, mỗi ty thể có hơn 10 phân tử ADN. Các phân tử này không giống nhau do chứa các đột biến tạo nên tính đa dạng rất cao của ADN ty thể giữa các tế bào ngay trong một cơ thể.

Cuối năm 2000, hơn 96% trình tự nucleotide của genome người đã được công bố. Genome người có kích thước khoảng $3,2 \times 10^6$ kb, tức là 3,2 Gb (Gigabase-đơn vị lớn nhất dùng đo chiều dài trên bản đồ vật lý). Trong đó khoảng 2,95 Gb là vùng chất nhiễm sắc (*euchromatin*). Chỉ có 1,1 đến 1,4% chứa gen mã cho khoảng 30.000-40.000 protein, trong đó chỉ mới xác định được 1/3, còn lại là các protein dự đoán (*predicted protein*). Genome người có tới 1,4 triệu chỉ thị SNPs. Thành phần ADN lặp lại (*SINEs*, *LINEs*, *LTRs* và *transposon*) chiếm gần một nửa genome (~43%). Tuy nhiên hầu hết các transposon và LTRs đều ở trạng thái không hoạt động.

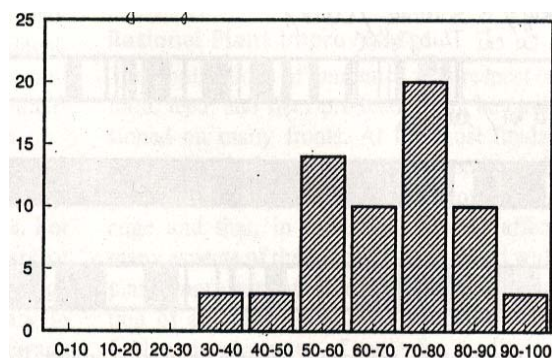
1.8.3 Nghiên cứu Genomics ở thực vật

Số lượng các gen và những thông tin về genome của rất nhiều loài sinh vật nói chung cũng như của thực vật nói riêng lưu trữ trong ngân hàng gen tăng nhanh không ngừng. Những số liệu này rất hữu ích trong định hướng nghiên cứu nhằm lựa chọn được phương pháp thích hợp. Chúng ta cùng nhau tìm hiểu xem các nhà sinh học đã sử lý những thông tin về genome, đặc biệt là của cây mô hình *Arabidopsis* như thế nào trong hướng nghiên cứu genomics đối với thực vật.

Trình tự nucleotide của genome *Arabidopsis* đã được xác định hoàn toàn vào cuối năm 2000. Loài thực vật này rất gần với các cây họ cải (cải bắp, súp lơ, xu hào, củ cải vv...). Đây là những loại rau xanh rất phổ biến trong đời sống con người. Một trong nhiều lý do khiến cho genome *Arabidopsis* được chọn để xác định toàn bộ trình tự là do genome có kích thước tương đối nhỏ (130 Mb) và chứa ít ADN lặp lại. Bên cạnh đó, dự án đọc trình tự genome cây lúa đã được thực thi. Genome của lúa lớn gấp 3,5 lần so với *Arabidopsis* nhưng cũng chỉ bằng 20% genome của ngô và 3% của lúa mì. Điều may mắn là cấu trúc genome của các cây lương thực (lúa, lúa mì, lúa miến, kê, ngô ...) khá giống nhau. Sự khác nhau về kích thước genome chủ yếu do thành phần ADN lặp lại. Không phải genome càng lớn thì số lượng gen khác nhau càng nhiều. Do đó, trật tự các gen, chức năng của chúng cũng như cấu trúc genome của *Arabidopsis* và lúa rất cần thiết để phân lập và điều khiển các gen quan trọng ở các cây có quan hệ gần gũi với chúng. Những kết quả này đặc biệt cần thiết cho chọn lọc, tạo mới nguồn giống cây nông nghiệp có tính trạng mong muốn.

Cùng với genome của *Arabidopsis* và lúa, hơn 127.000 các trình tự biểu hiện hay còn gọi là ADN đích- ESTs có nguồn gốc từ thực vật được lưu trữ trong ngân hàng dữ liệu cho phép so sánh các gen giữa các loài thực vật với nhau. Kết quả cho thấy hầu hết các gen đã biết ở thực vật bậc cao đều có mức độ tương đồng rất lớn. Do đó, từ trình tự nucleotide có thể suy đoán được chức năng của gen và ngược lại. Khi biết một gen bất kỳ ở một loài có thể phân lập được gen có chức năng tương tự ở loài khác. Ví dụ, các nhà sinh học đã tiến hành so sánh mức độ giống nhau của 64 protein được chọn ngẫu nhiên từ lúa và *Arabidopsis*. Chức năng của các protein này cũng như trình tự nucleotide của gen mã cho chúng đã biết. Để tránh hiện tượng so sánh các thành viên trong cùng một họ gen, các gen được chọn đều là những gen chỉ có một bản sao trong genome. Kết quả cho thấy số gen có độ tương đồng 70-80% chiếm tỷ lệ cao nhất. Điều này có nghĩa, mặc dù lúa và *Arabidopsis* không có quan hệ gần gũi với nhau

nhưng không vì thế mà các gen của chúng phải khác nhau. Ngược lại, đa số các gen có cấu trúc và chức năng giống nhau đều tồn tại ở cả hai loài này (Hình 1.18).



Hình 1.18:

So sánh mức độ tương đồng protein giữa *Arabidopsis* và lúa.

Thực vật có hoa được xem là xuất hiện và trải qua tiến hoá từ khoảng 150 triệu năm. Bên cạnh những nét tương đồng về mặt di truyền, các đặc điểm về sinh trưởng phát triển rất phong phú, đa dạng giữa các loài và ngay trong cùng một loài. Tuy nhiên, kết quả so sánh genome của lúa và *Arabidopsis* (không gần nhau) cho thấy sự khác biệt giữa chúng không phải do số gen đặc hiệu cho từng loài mà phụ thuộc rất nhiều đến cách thức kiểm soát hoạt động của một gen hoặc các thành viên trong cùng một họ gen. Các kết quả nghiên cứu khác cũng đưa ra nhận xét sự đa dạng giữa thực vật chủ yếu do thay đổi trình tự nucleotide ở các vùng ADN điều khiển (*cis-regulatory sequence*) và các loại ADN lặp lại.

Đối với các giống trong cùng một loài, thay đổi rất nhỏ trong cấu trúc của một gen hoặc cách thức kiểm soát biểu hiện của nó thường dẫn đến những sự khác biệt rất rõ về tính trạng. Điều này được minh họa bởi ví dụ điển hình với các gen liên quan đến biến đổi các acid béo. Thực vật bậc cao có hơn 200 loại acid béo khác nhau, thường được dự trữ trong hạt. Chỉ xét đến cấu trúc bậc một, các acid béo này khác nhau bởi số liên kết đôi, liên kết ba, số nhóm OH, epoxi vv... Tuy nhiên, những sai khác đó đều được tạo ra bởi một số enzym thuộc họ gen chứ không phải do 200 gen khác nhau. Chúng ta đã biết từ một gen, có thể có nhiều phân tử ARNm khác nhau nhờ cơ chế cắt nối exon-intron luân phiên (*alternative splicing*). Chỉ cần thay đổi 4 acid amin đủ khiến cho dasaterase chuyển thành hydroxylase. Rõ ràng việc hình thành các hợp chất hoá học mới không nhất thiết cần phải xuất hiện gen mới mã cho chúng. Ngoài ra, chúng ta đã biết các thành viên trong họ gen mã cho cytochrome P450s đảm nhận một loạt chức năng khác nhau liên quan đến sinh tổng hợp polysaccharide, kinase, phosphatase, các factor điều khiển gen ... Như vậy, để sử dụng có hiệu quả các dữ liệu trong ngân hàng, trước tiên cần xác định một chức năng cụ thể cần nghiên cứu trong số các chức năng mà một gen đơn bản hoặc các gen trong một họ gen đảm nhận.

Khai thác thông tin trong ngân hàng dữ liệu tạo nên một hướng mới, được gọi là **phylo-genomics** (tạm dịch là phát sinh genome) trong nghiên cứu phát sinh chủng loại. Các nhà sinh học dựa vào các thông tin đã biết để phân tích nguồn gốc của một gen cần quan tâm, phân tích nguyên nhân của sự đa dạng biến đổi về trình tự nucleotide hoặc acid amin (sản phẩm protein) của chính gen đó tồn tại trong các sinh vật khác nhau. Hơn nữa, có thể tiến hành phân tích mức độ tương đồng về ADN hay sản phẩm protein của gen đó với các gen khác. Từ đây có thể xác định chức năng đặc thù của gen này.

Những kết quả đạt được trong nghiên cứu phylo-genomics cho thấy khoảng 54% các gen ở thực vật bậc cao có thể phân thành các nhóm có chức năng khác nhau. Khoảng 13% các gen của *Arabidopsis* liên quan đến các yếu tố điều khiển gen và con đường truyền tín hiệu. Phân

tích 389 gen đã biết hoặc ORFs nằm trên một đoạn 1,9 Mb trong genome *Arabidopsis* cũng có thể xếp chúng vào các nhóm có chức năng khác nhau liên quan đến các con đường chuyển hoá, đến tổng hợp năng lượng, truyền tín hiệu, sinh trưởng và phân chia tế bào. Tuy nhiên, mặc dù có thể biết sản phẩm của gen nhưng vai trò cụ thể của nó trong sinh trưởng phát triển đối với từng cá thể lại rất khó xác định rõ ràng. Để khắc phục được yếu điểm đó, cần sự hỗ trợ mật thiết của các kỹ thuật khác như bẫy gen, đột biến di truyền... Ví dụ khi đưa một đoạn ADN có trình tự nucleotide đã biết vào genome có thể tạo ra các dòng đột biến khác nhau. Căn cứ vào trình tự nucleotide hai bên vị trí ghép (xác định nhờ phản ứng invert-PCR hoặc TAIL-PCR) có thể xác định được gen bị bất hoạt. Dựa vào biểu hiện của tính trạng mới, có thể biết được động học biểu hiện của gen đó ở các thời điểm và giai đoạn khác nhau.

Một khó khăn nữa mà phylo-genomics gặp phải là sự lặp lại của các gen xảy ra khá phổ biến trong genome. Genome *Arabidopsis* có nhiều gen lặp lại nằm sát nhau, do đó khi gây đột biến bởi các tác nhân hoá lý, hầu như không thể tạo được dòng đột biến kép nhờ trao đổi chéo. Điều này rất cần biết khi xác định đột biến liên quan đến một hay nhiều gen cũng như hạn chế hiện tượng các gen cùng chức năng hỗ trợ lẫn nhau khi một gen bị bất hoạt. Chính kỹ thuật sử dụng COs (phức ADN/ARN) đã phần nào khắc phục được khó khăn này. Phức được thiết kế có mang mã dừng tổng hợp protein để gây đột biến điểm ở đoạn gen bảo thủ trong mọi thành viên của họ gen (hoặc các gen lặp lại). Ngoài ra, phương pháp sử dụng sợi đúp ARN gây bất hoạt (*gene silencing*) cũng tỏ ra có nhiều ưu điểm trong việc kìm hãm hoạt động của gen. Hơn nữa, khi bị nhiễm virus, thực vật có khả năng nhận biết và kìm hãm sự nhân bản của ARN virus (*virus-induced gene silencing*). Nếu như bị nhiễm bởi virus tái tổ hợp có mang gen lạ tương đồng với gen của chính thực vật, cây chủ sẽ tăng cường bất hoạt luôn cả gen trong genome của mình.

Chương 2

HOẠT ĐỘNG CỦA GEN TRONG TẾ BÀO

Gen mang thông tin di truyền mã cho sản phẩm cần thiết cho sự sống của tế bào. Một gen được xem là hoạt động khi thông tin đó được sử dụng để tạo ra sản phẩm cuối cùng là protein hay phân tử ARN. Điều đáng ghi nhớ là thông tin chỉ được phiên mã và dịch mã theo một chiều từ ADN sang ARN đến protein. Không xảy ra hiện tượng thông tin truyền từ ADN sang thẳng protein hoặc từ protein ngược trở lại acid nucleic. Việc trao đổi thông tin giữa các dạng acid nucleic (giữa ADN và ARN) có thể xảy ra thuận nghịch. Ví dụ, ADN \rightarrow ARN (phiên mã xuôi từ ADN sang ARN), ARN \rightarrow ADN (phiên mã ngược từ ARN sang ADN - reverse transcription), ADN \rightarrow ADN (tái bản ADN), ARN \rightarrow ARN (nhân bản ARN bởi ARN polymerase dựa trên khuôn ARN).

Hoạt động của gen mã cho protein gồm hai giai đoạn chính là tổng hợp ARNm và tổng hợp protein. Xen giữa chúng gồm nhiều giai đoạn trung gian có nhiệm vụ sàng lọc và điều biến thông tin từ gen đến sản phẩm cuối cùng. Đầu tiên, một gen bắt đầu hoạt động khi cấu trúc của nó trong nhiễm sắc thể được thay đổi (khử methyl ở cytosine, biến đổi một số liên kết ở histone, trở nên nhạy cảm với nuclease, có sự sắp xếp lại các đoạn ADN vv...). Đây là giai đoạn thay đổi ở mức độ ADN. Tiếp đó là giai đoạn tổng hợp ARNm (hình thành các phức hoạt hoá quá trình phiên mã, kiểm tra việc bắt đầu và kết thúc tổng hợp ARNm vv...), giai đoạn biến đổi phân tử ARNm (*RNA processing*-chỉ xảy ra ở tế bào eukaryot), vận chuyển ARNm ra ngoài tế bào chất (*RNA export*), kiểm tra thông tin di truyền trên phân tử ARNm (*editing RNAs*) và quyết định tính bền vững của nó (*RNA turnover*). Cuối cùng là tổng hợp protein (*protein synthesis*), biến đổi chúng thành các sản phẩm có hoạt tính (*protein modification*), vận chuyển và phân bố protein đến các vị trí khác nhau (*protein targeting*) cũng như qui định thời gian tồn tại của chúng trong tế bào (*protein turnover*). Bất kỳ giai đoạn nào cũng được kiểm soát chặt chẽ bởi nhiều cơ chế khác nhau. Tất cả đều nhằm mục đích đảm bảo một cách nghiêm ngặt, chính xác số lượng sản phẩm của từng gen theo đúng yêu cầu của tế bào. Trong mỗi tế bào, các cơ chế điều khiển có thể xảy ra chung với nhiều gen (*global regulation*) hay riêng biệt cho từng gen (*specific regulation*); có thể là tích cực (*positive control*) hay tiêu cực (*negative control*), có thể là tự điều biến (*autoregulation*) hoặc kiểm soát phản hồi (*feedback regulation*). Sự phối hợp điều hoà giữa các cơ chế thường khác nhau ở các tế bào prokaryot và eukaryot (phụ thuộc vào cấu trúc tế bào không nhân hoặc có nhân, cấu trúc không gian của nhiễm sắc thể, cấu trúc các gen thành operon, vị trí của gen trong genome, cấu trúc các intron trong một gen vv...).

Chúng ta đã biết từ chương 1, ADN trong tế bào eukaryot được đóng gói trong cấu trúc nucleosome. Cấu trúc đó sẽ bị thay đổi cục bộ tại vùng ADN chuẩn bị được phiên mã. Ví dụ, tại vùng chuẩn bị phiên mã, một số cytosine ở đầu 5' của gen bị khử nhóm methyl, các acid amin lysine của histone H3 có sự thay đổi mức độ methyl hoá, histone H4 bị acetyl hoá vv... 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 cấu trúc không gian, thông qua cấu trúc nucleosome,

trở thành yếu tố đầu tiên kiểm soát một gen chuyển từ trạng thái bất hoạt (*silent state*) sang trạng thái sẵn sàng cho khởi động phiên mã (*active state*).

Tuy nhiên, khởi động phiên mã có thể xảy ra được hay không còn phụ thuộc vào nhiều yếu tố khác nữa. Ví dụ, khởi động ở các gen chịu sự kiểm soát tích cực (*positive control*) còn phụ thuộc vào sự có mặt của các protein hoạt hoá. Protein hoạt hoá thường liên kết với protein khác, gọi là chất cảm ứng (*inducer*) tạo thành phức. Phức này (*activator+inducer*) có khả năng tương tác với ADN hoặc với ARN polymerase. Nhờ đó, cấu trúc cục bộ của ADN thay đổi, hoặc ái lực liên kết của ARN polymerase với promoter tăng, tạo điều kiện thuận lợi để khởi động phiên mã. Như vậy hoạt động của gen bị kiểm soát theo cơ chế tích cực sẽ tăng hoặc giảm phụ thuộc vào sự có mặt hoặc thiếu vắng các activator và inducer (ví dụ điển hình là hoạt động của operon *ara*). Ngược lại, cơ chế kiểm soát tiêu cực liên quan đến protein ức chế (*repressor*). Repressor cũng thường liên kết với protein khác, gọi là co-repressor để tạo phức. Phức này tương tác với ADN hoặc với ARN polymerase để ngăn cản khởi động phiên mã. Do đó, hoạt động của gen bị kiểm soát theo cơ chế tiêu cực sẽ được bật mở hoặc đóng phụ thuộc vào sự thiếu vắng repressor hoặc có mặt của repressor và co-repressor (ví dụ điển hình là hoạt động của operon *lac*).

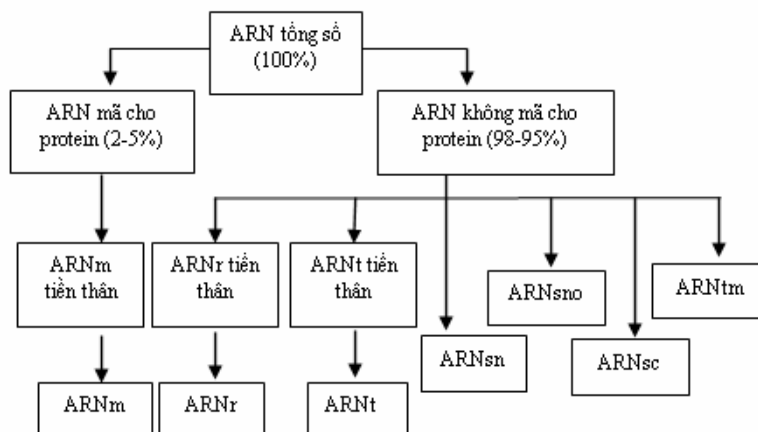
Cả hai cơ chế kiểm soát tiêu cực và tích cực có thể kết hợp cùng nhau để điều biến biểu hiện của gen. Các protein hoạt hoá và ức chế có thể cạnh tranh tương tác để điều biến một cách linh động biểu hiện của gen. Một số gen thường biểu hiện liên tục nhưng chỉ ở mức độ tối thiểu trong tế bào. Các gen này dễ dàng tăng cường hoạt động khi xuất hiện các yếu tố hoạt hoá như activator và inducer. Ví dụ, hoạt động của một số gen mã cho enzym tăng mạnh khi xuất hiện cơ chất. Những gen này được gọi là "*inducible gene*" (gen có thể hoạt hoá). Ngược lại, một số gen thường biểu hiện ở mức độ mạnh. Chúng dễ dàng bị kìm hãm khi xuất hiện protein ức chế-repressor. Ví dụ, hoạt động của những gen tham gia tổng hợp tryptophan trong tế bào *E.coli* bị giảm khi bổ sung tryptophan vào môi trường. Những gen này được gọi là "*repressible gene*" (gen có thể bị kìm hãm). Ở đây, chúng ta hình dung được sự phối hợp, sự cạnh tranh giữa các cơ chế kiểm soát nhằm thay đổi mức độ biểu hiện của gen một cách rất linh động.

Các cơ chế chung điều khiển hoạt động của gen được nghiên cứu khá chi tiết ở giai đoạn bắt đầu tổng hợp ARNm. Tuy nhiên, chúng ta nhận thấy chúng đều có thể thực thi đối với bất kỳ giai đoạn nào trong suốt quá trình chuyển đổi thông tin từ ADN đến protein. Trong tế bào eukaryot, các gen chịu điều khiển theo cơ chế tích cực chiếm ưu thế. Hoạt động của những gen này thường khó xảy ra nếu như thiếu activator. Ngược lại, trong tế bào prokaryot, các gen thường bị kiểm soát bởi các protein ức chế repressor. Đa số các gen prokaryot hoạt động mạnh nên chúng cần các repressor để kìm hãm mức độ biểu hiện của gen.

Trong hầu hết các sinh vật, phân tử ADN có chức năng quan trọng nhất là lưu trữ thông tin di truyền còn các phân tử ARN đảm nhiệm nhiều nhiệm vụ khác nhau trong tế bào. Số lượng ARNm mang mã di truyền để tổng hợp protein chiếm tỷ lệ rất ít so với tất cả các loại ARN khác có mặt trong tế bào (Hình 2.1).

Những phân tử ARN chỉ được phiên mã nhưng không dịch mã (không dùng làm khuôn để tổng hợp protein) được gọi chung là ARN không mang mã (*non coding ARNs*). Thực chất đây là những phân tử ARN không chứa các mã bộ ba tương ứng cho acid amin. Tuy nhiên, thông tin di truyền trên những phân tử này cần thiết cho rất nhiều phản ứng chức năng của tế bào. Chúng tham gia trực tiếp vào quá trình tái bản ADN (ở *telomere*), làm thay đổi trình tự nucleotide, 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, ARN 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 (như bắt đầu sự phiên mã), kiểm soát sau phiên mã như cắt nối intron-exon (*splicing*), đọc sửa ARNm (*RNA editing*), phân cắt ARNm (*posttranscriptional gene silencing*), kìm hãm hoạt động của các gen nhảy (*transposon silencing*) vv...



Hình 2.1:

Các loại ARNs có mặt trong tế bào sinh vật. ARNsn (small nuclear RNA): ARN kích thước nhỏ trong nhân. ARNsno (small nucleolar RNA): ARN kích thước nhỏ trong hạch nhân. ARNsc (small cytoplasmic RNA): ARN kích thước nhỏ trong tế bào chất. ARNtm (transfer-messenger RNA): ARN tương tự ARNt có khả năng tương tác với ARNm (theo Brown, 2001).

Điều đặc biệt cần lưu ý đối với các loại ARN có mặt trong tế bào (được gọi chung là ARN tổng số) là chúng thay đổi về số lượng và chủng loại tùy thuộc vào sự biệt hoá và trạng thái của tế bào. Sự tổng hợp các ARN mới luôn xảy ra cùng với sự phân huỷ các ARN cũ. Tuy nhiên, số lượng của từng loại ARN cũng như số loại phân tử ARN tạo ra thường không giống với số lượng và số loại ARN bị phân huỷ. Điều này thể hiện rất rõ với các loại ARNm. Trong quá trình tế bào phát triển, biệt hoá hoặc trả lời các tín hiệu từ bên ngoài, một số gen có thể được bật mở trong khi một số khác bị bất hoạt; một số gen có thể tăng mức độ phiên mã trong khi các gen khác giảm. Tập hợp tất cả các phân tử ARNm có mặt trong tế bào được gọi chung là **transcriptome**. Transcriptome sẽ qui định các loại protein cần có trong tế bào. Tập hợp tất cả các protein sẽ được gọi chung là **proteome**. Như vậy, transcriptome sẽ thay đổi trong quá trình tế bào sinh trưởng và biệt hoá, trong các phản ứng mà tế bào trả lời tín hiệu khác nhau từ bên ngoài. Sự thay đổi của transcriptome sẽ tương ứng với thay đổi của proteome. Dù biến đổi một cách liên tục, nhưng không bao giờ transcriptome chỉ gồm các phân tử ARNm mới. Trong phân bào, tế bào con đều nhận một phần transcriptome từ tế bào mẹ ngay khi các ARNm đó chưa cần sử dụng để dịch mã. Ví dụ, trong bào tử vi khuẩn hoặc trong tế bào trứng, trong các hạt thực vật đều chứa transcriptome ở dạng dự trữ chưa được dịch mã sang proteome. Khi trứng thụ tinh, khi hạt nảy mầm... transcriptome sẽ được giải mã thành proteome tương ứng.

Phản ứng tổng hợp ARNm được thực hiện nhờ enzyme ARN polymerase dựa trên khuôn mẫu là sợi đơn ADN. Sợi đơn dùng làm khuôn để tổng hợp ARNm được gọi là sợi khuôn (*template strand*). Trình tự nucleotide trên sợi ADN đơn bổ sung với sợi khuôn giống hệt phân tử ARNm. Sợi đó được gọi là sợi mang mã di truyền (*coding strand*). Phản ứng tổng hợp ARNm có thể chia làm 4 giai đoạn như sau:

1. Nhận biết khuôn mẫu: Các phức protein cần thiết khởi động quá trình tổng hợp ARNm của một gen sẽ bám vào những trình tự đặc biệt (có thể phân bố phía trước hoặc phía sau, hoặc ngay trong gen) và enzym ARN polymerase bám vào promoter. Sợi kép ADN mở xoắn tại vị trí promoter, giải phóng sợi khuôn ở dạng tự do.
2. ARN polymerase bắt đầu phản ứng tổng hợp ARN. Đầu tiên, enzym tổng hợp các sợi ARN rất ngắn chỉ khoảng 2-9 ribonucleotide. Sau đó, enzym mới tổng hợp phân tử ARNm. Các ribonucleotide của sợi ARNm tạo liên kết bổ sung với sợi khuôn.
3. ARN polymerase chuyển động dọc theo phân tử ADN và sợi ARNm dài dần do các nucleotide được thêm vào đầu 3' của sợi ARN. Liên kết ADN/ARN xảy ra tại vùng ADN mở xoắn. Khi ARN polymerase chuyển động qua, liên kết đó bị phá vỡ; hai sợi đơn ADN phục hồi thành dạng kép, sợi ARNm tách ra dưới dạng sợi đơn.
4. ARN polymerase cùng với protein phụ trợ nhận biết được vị trí đặc hiệu để kết thúc phản ứng tổng hợp ARN. ARN polymerase tách ra khỏi sợi ADN; sợi ARN tách ra khỏi phức enzym và ADN khuôn.

2.1 Kiểm soát hoạt động của gen khi phiên mã

Trong tế bào prokaryot, chỉ có một loại ARN polymerase xúc tác cho phản ứng tổng hợp các loại ARNr, ARNt và ARNm. Một tế bào *E.coli* có khoảng 7000 phân tử ARN polymerase, trong đó khoảng 2000 đến 5000 phân tử luôn ở trạng thái hoạt động. Phân tử ARN polymerase của tế bào prokaryot gồm hai phần: phần lõi tetramere ($\alpha_2\beta\beta'$) và yếu tố σ . Cấu trúc lõi không thay đổi nhưng có nhiều loại yếu tố σ . Phần lõi có khả năng tổng hợp ARN trên mọi khuôn ADN, còn yếu tố σ cho phép phản ứng xảy ra chính xác tại promoter. Ở vi khuẩn, một số gen như các gen mã cho ARN ribosome (ARNr), ARN vận chuyển (ARNt), protein tạo bộ máy ribosome, các enzym ARN và ADN polymerase thường hoạt động liên tục (*constitutive expression*). Chúng được gọi là các gen “housekeeping” (tạm dịch là gen quản gia). Hầu hết các gen mã cho protein trong tế bào prokaryot có khả năng đáp ứng rất nhanh những tác động thay đổi của môi trường bên ngoài. Hoạt động của một gen thường được kiểm soát ở 5 mức độ kế tiếp nhau: phiên mã tạo ARNm, biến đổi phân tử ARNm (thay đổi cấu trúc, thay đổi nucleotide...), thời gian tồn tại của ARNm, dịch mã và cuối cùng là biến đổi sau dịch mã (biến đổi cấu trúc phân tử protein để chuyển từ dạng không hoạt tính sang có hoạt tính). Tuy nhiên, ở prokaryot, kiểm soát ở giai đoạn đầu tiên (giai đoạn phiên mã) giữ vai trò chủ yếu đối với hoạt động của gen.

Tế bào prokaryot không có nhân, do đó phân tử ARNm vừa được phiên mã trên khuôn ADN vừa được sử dụng để dịch mã tổng hợp protein. Như vậy, cả hai quá trình phiên mã và dịch mã xảy ra đồng thời. Đây là điều khác biệt với tế bào eukaryot, ở đó ARNm được tổng hợp trong nhân, trải qua các biến đổi, được vận chuyển ra ngoài tế bào chất để dùng làm khuôn tổng hợp protein.

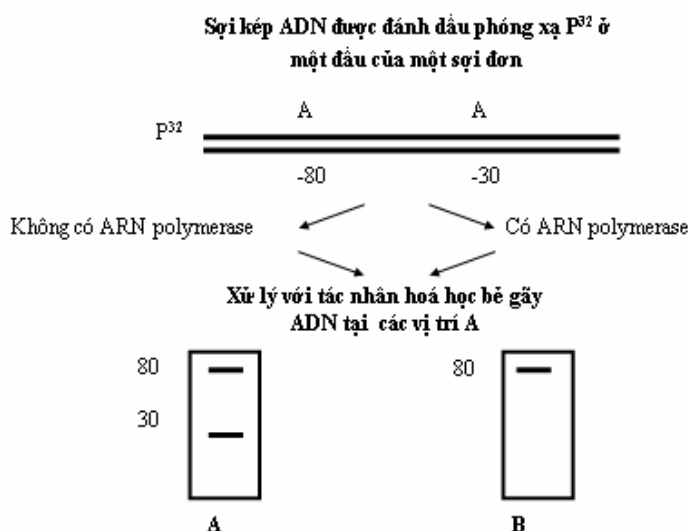
Như đã trình bày trong chương 1, cấu trúc của một gen gồm hai phần chính: phần mang các mã di truyền và phần chứa các nucleotide liên quan đến chức năng điều khiển phiên mã di truyền sang sợi ARN. Vị trí nucleotide đầu tiên phiên mã sang ARN được ký hiệu là +1. Với qui định này, các nucleotide nằm trong phần điều khiển (ở vùng 5' phía trước vị trí +1) sẽ được ký hiệu là vị trí (-).

Promoter của các gen vi khuẩn mã cho protein thường bao gồm các nucleotide ở vị trí -10 và -35. ARN polymerase nhận biết và tương tác với vị trí -35, sau đó vị trí -10 bị mở xoắn để

bắt đầu phiên mã. Kết quả nghiên cứu trình tự nucleotide của hơn 100 promoter ở vi khuẩn cho phép xác định được tính bảo toàn của đoạn nucleotide tại vị trí -10 và -35 như sau: (Các chữ số chỉ % xuất hiện của base trong promoter)

Vị trí -10: T₈₀ A₉₅ T₄₅ A₆₀ A₅₀ T₉₆

Vị trí -35: T₈₂ T₈₄ G₇₈ A₆₅ C₅₄ A₄₅



Hình 2.2:

Thí nghiệm "footprint" cho phép phát hiện vị trí bám của ARN polymerase trên một đoạn ADN. So sánh hai kết quả điện di cho thấy enzym bám vào vị trí -30 nên không xuất hiện vạch -30 ở (B).

Điều nên lưu ý là một số gen không có cấu trúc promoter điển hình hoặc quá trình phiên mã đòi hỏi một số protein khác tham gia cùng ARN polymerase. Ngoài ra, các gen hoạt động ở những điều kiện khác nhau đòi hỏi ARN polymerase liên kết với các yếu tố khác nhau. Ví dụ, tế bào vi khuẩn chỉ có một loại ARN polymerase phiên mã cho các gen khác nhau. Khi nhiệt độ môi trường tăng cao, ARN polymerase của vi khuẩn sẽ chỉ liên kết với yếu tố σ_{54} để nhận biết promoter của các gen hoạt động khi nhiệt độ tăng bất thường (các gen chịu sốc nhiệt). Tế bào eukaryot có nhiều loại ARN polymerase, trong đó ARN polymerase III nhận biết promoter của gen mã cho các phân tử ARN vận chuyển (ARNt) và phân tử 5S ARN ribosome (ARNr). Promoter của những gen này nằm phía sau vị trí +1 (tức là nằm trong phần mang mã di truyền).

Vị trí promoter nằm trên một đoạn ADN có thể xác định bằng thực nghiệm thông qua phương pháp "footprints" (tạm dịch là in dấu vân) dựa vào nguyên tắc: ARN polymerase bám vào promoter và bảo vệ nó khỏi tác dụng phân hủy của DNase hoặc các tác nhân hoá học. Trong phương pháp này, đầu tiên ADN được xử lý với tác nhân hoá học hoặc với DNase, sau đó được điện di trên gen acrylamide. Vùng promoter được bảo vệ sẽ xuất hiện như khoảng trống (*footprints*) (Hình 2.2).

2.1.1 Kiểm soát khởi đầu phiên mã

Các gen vi khuẩn thường tập hợp thành nhóm và được điều khiển bởi một promoter. Cấu trúc này gọi là *operon*. Như vậy, phiên mã trên các gen nằm trong một operon được kiểm tra, điều khiển đồng bộ theo nguyên tắc: hoặc là không có gen nào hoặc là tất cả các gen cùng được phiên mã sang ARNm. Phân tử ARNm phiên mã từ một operon được gọi là ARNm -

polycistronic. Tuy nhiên, thông tin di truyền của mỗi gen nằm trên phân tử ARNm - polycistronic được dịch mã một cách hoàn toàn độc lập. Phân tử ARNm - polycistronic mang các vị trí để ribosome bám vào và tách ra tương ứng với từng gen nằm trong operon. Tốc độ tổng hợp các protein trên một phân tử ARNm polycistronic hoàn toàn khác nhau. Như thế, mặc dù cùng được phiên mã nhưng sản phẩm protein của từng gen nằm trong một operon được tổng hợp một cách độc lập với số lượng khác nhau và tại thời điểm khác nhau.

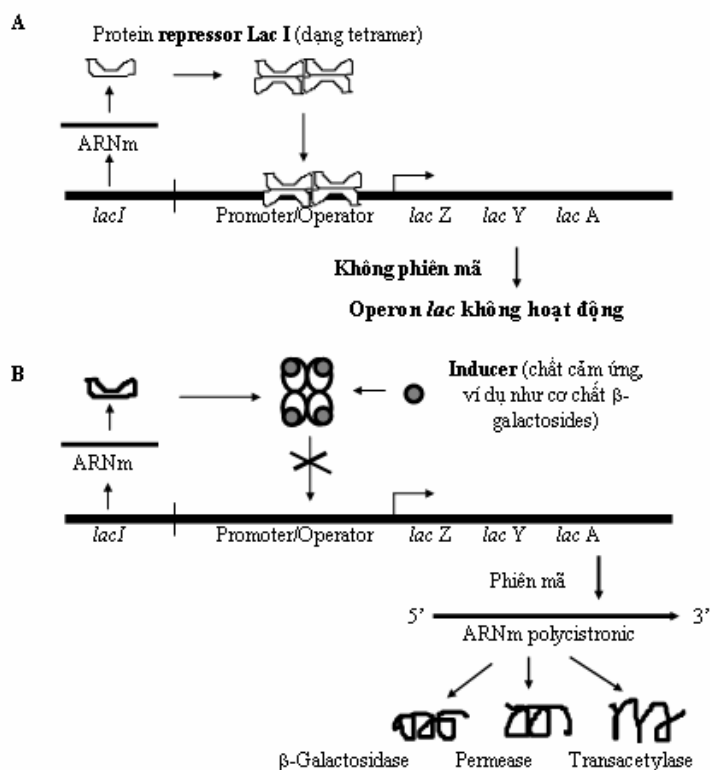
Quá trình khởi đầu phiên mã trong tế bào prokaryot thường được kiểm soát chủ yếu bởi hai cơ chế khác nhau: cơ chế tiêu cực và cơ chế tích cực. Trong cơ chế thứ nhất, hoạt động của gen bị kìm hãm do protein ức chế tương tác với vị trí đặc hiệu nằm gần promoter. Vị trí này được gọi là operator (vị trí ức chế). Khi repressor bám vào operator, ARN polymerase cũng như các protein hoạt hoá khác không thể dịch chuyển hoặc tương tác với sợi ADN khuôn. Do đó quá trình phiên mã tổng hợp ARNm bị kìm hãm, số lượng phân tử ARNm bị giảm. Ngược với cơ chế tiêu cực, hoạt động của gen chịu sự kiểm soát theo cơ chế tích cực chỉ bắt đầu khi có mặt của protein hoạt hoá. Chúng tương tác với các vị trí đặc hiệu ở vùng ADN điều khiển khiến cho cấu trúc ADN tại promoter thay đổi. Nhờ đó ARN polymerase dễ dàng nhận biết được các vị trí cần thiết để khởi động quá trình tổng hợp ARNm.

2.1.1.1 Cơ chế kiểm soát tiêu cực

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. Operon *lac* gồm 3 gen cấu trúc: *lacZ*, *lacY* và *lacA*. Sản phẩm của các gen này giúp tế bào tiếp nhận và chuyển hóa β -galactosides. Gen *lacY* mã cho enzym làm nhiệm vụ vận chuyển β -galactosides vào tế bào; gen *lacA* mã cho enzym chuyển nhóm acetyl từ acetyl-CoA vào β -galactosides và gen *lacZ* mã cho enzym chuyển β -galactosides thành các đường. Cấu trúc và cơ chế kiểm soát tiêu cực đối với hoạt động của operon *lac* được trình bày trên hình 2.3.

Quá trình phiên mã trên *lac* operon được kiểm soát bởi protein ức chế LacI. Protein này là sản phẩm của gen *lacI* nằm cạnh operon *lac*. Gen *lacI* được gọi là gen điều khiển vì sản phẩm của nó làm nhiệm vụ kiểm soát hoạt động của gen (nhóm gen) khác. Protein LacI chỉ có hoạt tính khi tồn tại ở dạng tetramer.

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. 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.

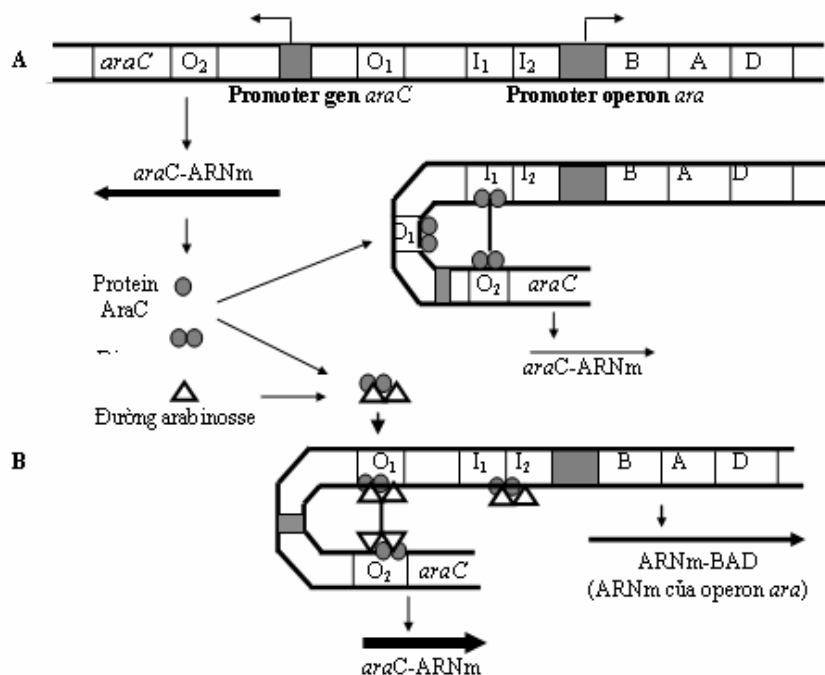


Hình 2.3:

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 khiến protein ức chế không bám được vào operator. ARN polymerase bám được vào vùng promoter để tổng hợp ARNm.

2.1.1.2 Kiểm soát theo cơ chế tích cực

Hoạt động của operon *ara* trong genome vi khuẩn lại được kiểm soát theo cơ chế ngược với cơ chế tiêu cực. Điều này thể hiện ở chỗ operon *ara* chỉ hoạt động khi có mặt của protein hoạt hoá. Một gen hoặc operon hoạt động mạnh hơn khi có mặt của protein hoạt hoá được xem là chịu sự kiểm soát theo cơ chế tích cực. Operon *ara* gồm có 3 gen: gen A mã cho isomerase, gen B mã cho kinase và gen D mã cho epimerase (Hình 2.4). Các enzym này cần thiết cho phản ứng chuyển hoá arabinose thành xylulose-5-phosphate. Khi môi trường có arabinose, hoạt động của operon *ara* tăng mạnh tương tự như lactose kích thích hoạt động của operon *lac*. Tuy nhiên, khi xuất hiện cơ chất arabinose, hoạt động của operon *ara* còn đòi hỏi sự có mặt của protein hoạt hoá AraC. AraC được mã bởi gen *araC* (nằm gần operon *ara*).

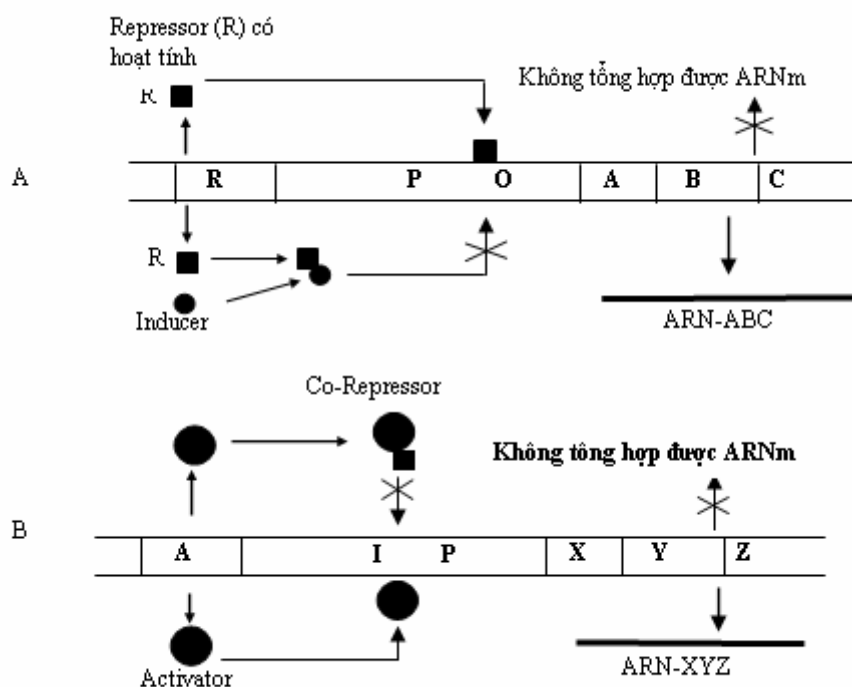


Hình 2.4:

Hoạt động của operon *ara* được điều khiển theo cơ chế tích cực bởi protein hoạt hoá AraC. (A): Khi môi trường không có đường arabinose, AraC bám vào các vị trí O₁, O₂ và I₁ tạo liên kết giữa O₂ và I₁ gây uốn cong sợi ADN. Operon *ara* bị kìm hãm, lượng ARNm tạo ra rất ít. (B) : Khi có mặt arabinose, phân tử đường này sẽ tương tác với protein hoạt hóa AraC. Phức chất bám vào 4 vị trí O₁, O₂, I₁ và I₂. Tương tác của phức chất giữa O₁ và O₂ kích thích hoạt động của gen *araC*. Ngoài ra, tương tác của phức chất giữa I₁ và I₂ khiến hoạt động của operon *ara* tăng lên rất nhiều. (theo Lodish và cs., 2000).

Khi môi trường không có arabinose, AraC bám vào 3 vị trí O₁, O₂ và I₁, nằm phía trước sát với promoter của operon *ara*. Hơn nữa, các phân tử AraC bám vào hai vị trí O₂, I₁ có thể tương tác với nhau (tương tác protein-protein) gây uốn cong sợi ADN. Lúc này operon *ara* không hoạt động. Khi môi trường có arabinose, protein hoạt hoá AraC tương tác với phân tử đường, do đó bị thay đổi cấu hình. Lúc này phức chất AraC và đường bám vào 4 vị trí O₁, O₂, I₁ và I₂. Tiếp đó, AraC ở hai vị trí O₁ và O₂ tương tác với nhau làm cho hoạt động của gen *araC* tăng lên. Mặt khác, phức chất đó khi gắn vào I₁ và I₂ gây hoạt hóa operon *ara*. Quá trình tổng hợp ARNm trên operon này được khởi động.

Điều đáng lưu ý là protein AraC tự điều khiển hoạt động của chính gen tạo ra nó (gen *araC*) bằng việc tương tác với các vị trí O₁ và O₂. Khác với protein ức chế LacI chỉ có tác dụng điều khiển tiêu cực đối với operon *lac*, protein hoạt hoá AraC kiểm soát hoạt động của cả operon *ara* và gen *araC* theo hai cơ chế tùy thuộc vào vị trí liên kết của nó với ADN. Trong khi protein AraC chỉ có khả năng kiểm soát theo cơ chế tích cực với gen *araC* thì vai trò của chúng đối với operon *ara* thay đổi tùy thuộc vào vị trí tương tác với ADN và tương tác giữa chúng với nhau. Hai cơ chế tiêu cực và tích cực xảy ra hoàn toàn đối lập nhau. Sự khác biệt giữa chúng được mô tả trên hình 2.5: các gen chịu sự điều khiển theo cơ chế tiêu cực bị ức chế khi có mặt của repressor. Ngược lại, các gen chịu kiểm soát theo cơ chế tích cực chỉ hoạt động khi có mặt activator.

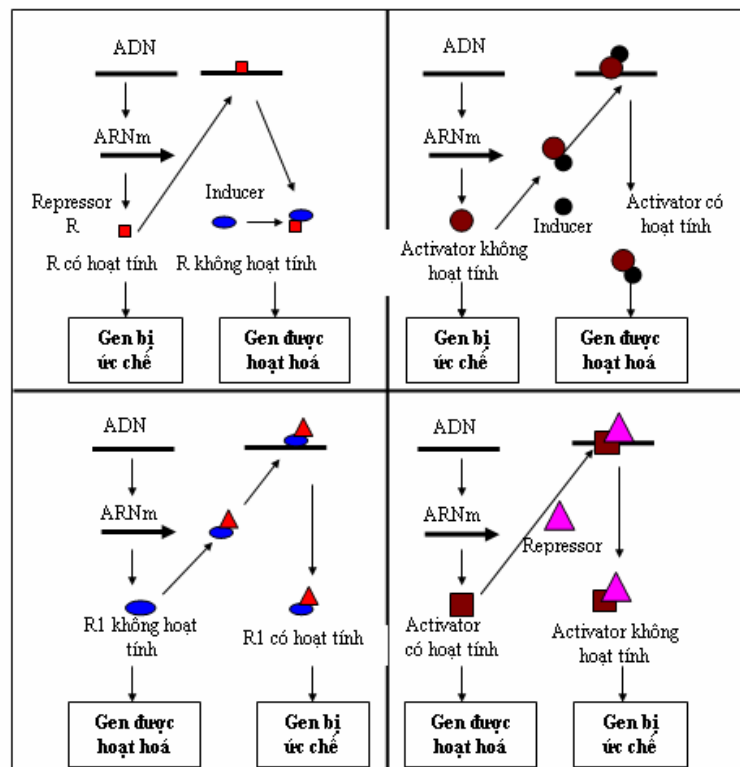


Hình 2.5:

So sánh hai cơ chế điều khiển tiêu cực (A) và điều khiển tích cực (B). Mặc dù tham gia kiểm soát quá trình tổng hợp ARNm theo chiều hướng ngược nhau, hai cơ chế này có thể hoạt động độc lập hoặc phối hợp cùng nhau để điều biến mức độ biểu hiện của gen một cách linh động (theo Lewin, 1993).

Các protein ức chế ngoài việc bám vào ADN để ngăn cản ARN polymerase khởi động việc phiên mã, chúng còn có thể tương tác với nhau để thực thi chức năng kìm hãm ở mức độ tối đa hoặc chúng bám vào ARNm để ngăn cản ribosome thực hiện việc tổng hợp protein. Khi có nhiều repressor cùng phối hợp hoạt động, chúng được gọi chung là các co-repressor.

Căn cứ vào mức độ hoạt động khi có mặt các phân tử protein điều khiển để biết được gen đó được kích thích hoặc bị kìm hãm. Ngoài ra tương tác giữa các protein điều khiển có thể gây đóng hoặc mở gen. Ví dụ, một protein cảm ứng gây mất hoạt tính của protein ức chế hoặc phục hồi hoạt tính của protein hoạt hoá đều dẫn đến khởi động gen. Ngược lại, biểu hiện của gen bị kìm hãm khi các co-repressor kích thích hoạt tính của protein ức chế, hoặc gây mất hoạt tính protein hoạt hoá. Những điều này được trình bày trên hình 2.6.



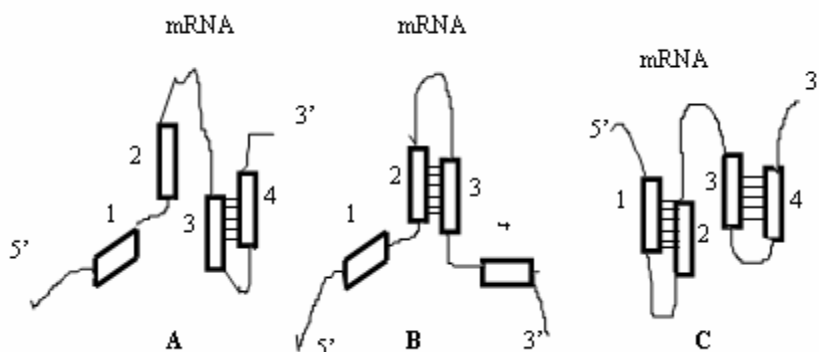
Hình 2.6:

Các cơ chế kiểm tra hoạt động của gen thay đổi một cách linh hoạt nhằm ức chế hoặc hoạt hoá gen. Phối hợp giữa các cơ chế này đảm bảo mức độ biểu hiện của gen rất linh động, đáp ứng yêu cầu của tế bào (theo Lewin, 1993).

2.1.1.3 Kiểm soát theo cơ chế suy giảm (attenuation control)

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 khi tương tác với repressor để ngăn cản hoạt động của operon *tryptophan*. Như vậy, operon này được điều khiển theo cơ chế tiêu cực do chính sản phẩm cuối cùng của operon. Cơ chế kiểm soát đặc biệt này được gọi là **ức chế phản hồi** (*feedback inhibition*). 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 2.7.

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" (gọi là 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. Tùy thuộc vào sự có mặt của tryptophan trong môi trường mà xảy ra những kiểu liên kết tạo cặp giữa hai chuỗi tạo ra các cấu trúc 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.



Hình 2.7:

Cấu trúc bậc hai của phân tử ARNm phiếm mã từ operon *tryptophan*. Cấu trúc này giữ vai trò kiểm soát phiên mã suy giảm khiến 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 Tryptophan trong môi trường. (A)- Khi môi trường thừa Tryptophan. (B)- Môi trường thiếu Tryptophan. (C)- Không có sự tổng hợp enzym.

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 và 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.

2.1.1.4 Quá trình dị hoá liên quan đến điều khiển tích cực

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

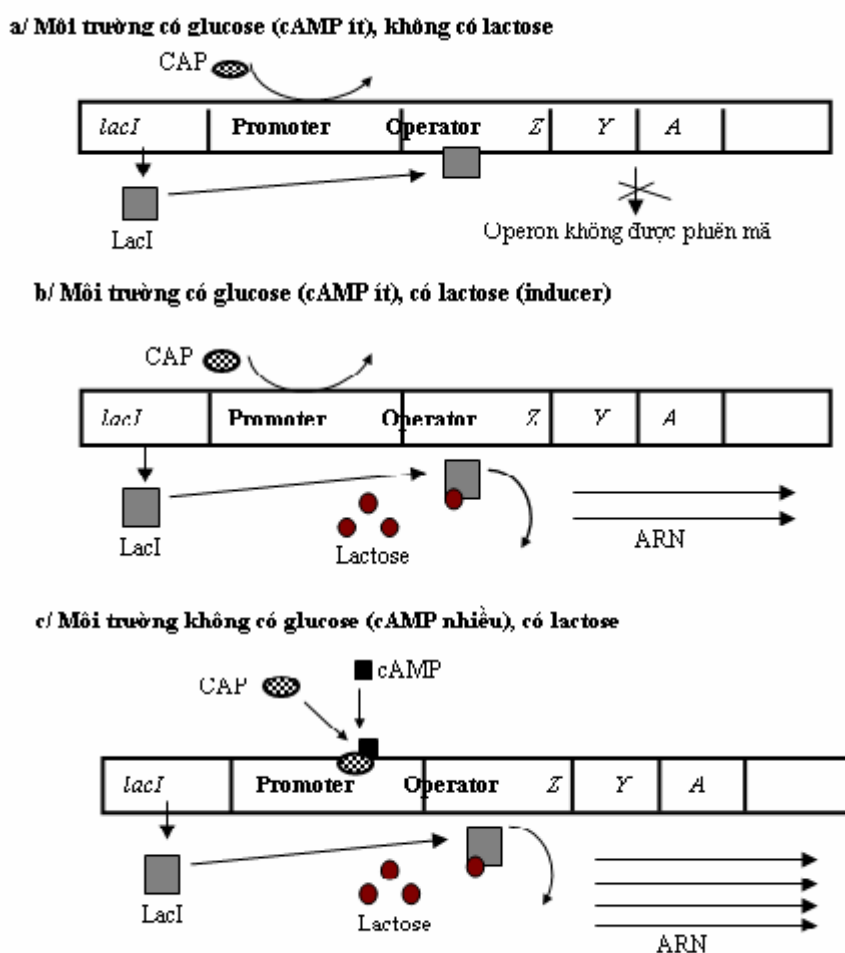
Các nhà hoá học phát hiện rằng khi vi khuẩn thiếu glucose, chúng sẽ tăng cường tổng hợp nucleotide cyclic adenosine-3,5'-monophosphate (cAMP). Số lượng cAMP tăng lên giống như tín hiệu báo động về nguy cơ chuyển hoá. Hơn nữa, khi bổ sung cAMP vào môi trường có chứa glucose và các đường khác thì hoạt 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.

Nghiên cứu các tế bào vi khuẩn đột biến không có khả năng sử dụng bất cứ loại đường nào ngoài glucose, các nhà khoa học phát hiện rằng chúng không tổng hợp được cAMP trong điều kiện môi trường thiếu glucose. Nói cách khác, các tế bào này chịu đột biến ở gen mã cho enzym tổng hợp cAMP (adenylate cyclase). Ngoài ra một số tế bào khác có khả

năng tổng hợp cAMP nhưng vẫn không tổng hợp được enzyme cần thiết để chuyển hoá các loại đường khác glucose. Như vậy phải tồn tại một protein tương tác với cAMP để hoạt hoá các operon.

Thực nghiệm đã phân lập được protein này và gọi là protein hoạt hoá quá trình dị hoá (Catabolite Activator Protein- CAP). Trong thí nghiệm tổng hợp ARNm 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 2.8).

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



Hình 2.8:

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.

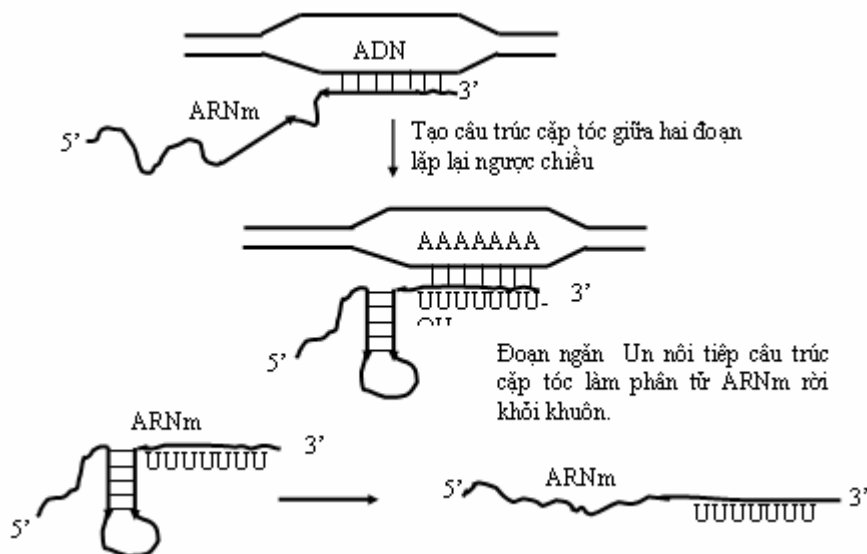
(b) Khi có lactose, LacI không bám được vào operator nhưng do có glucose nên cAMP không được tổng hợp dẫn đến phức chất cAMP-CAP không được tạo thành. Do đó hoạt

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., 2002).

2.1.2 Kiểm soát kết thúc phiên mã

Quá trình tổng hợp ARNm được dừng lại khi ARN polymerase gặp tín hiệu kết thúc (*termination signals*). 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 có thể là cấu trúc đặc biệt của phân tử ARN do trình tự nucleotide nằm phần cuối của gen qui định.

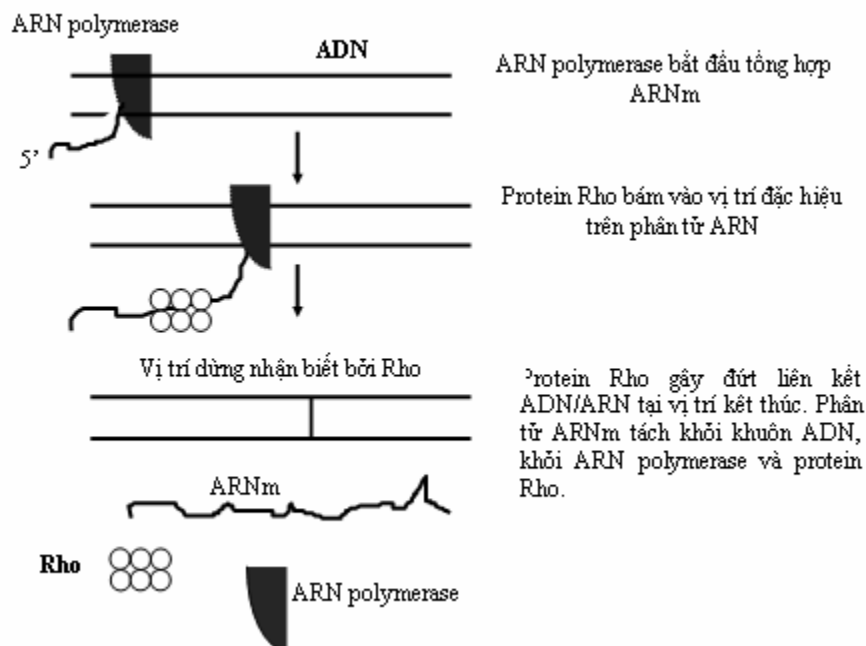


Hình 2.9:

Kết thúc phản ứng tổng hợp ARNm nhờ cấu trúc đặc biệt "hairpin". Cấu trúc này tạo ra do liên kết giữa các chuỗi nucleotide lặp lại ngược chiều "inverted repeat" và được nối tiếp bởi chuỗi 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.

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 2.9).

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 2.10).



Hình 2.10:

Cơ chế hoạt động của protein Rho. Protein Rho nhận biết vị trí đặc hiệu trên ADN và làm ngừng phản ứng tổng hợp ARNm. Nhờ đó, các phân tử ARNm đều có kích thước chính xác như nhau.

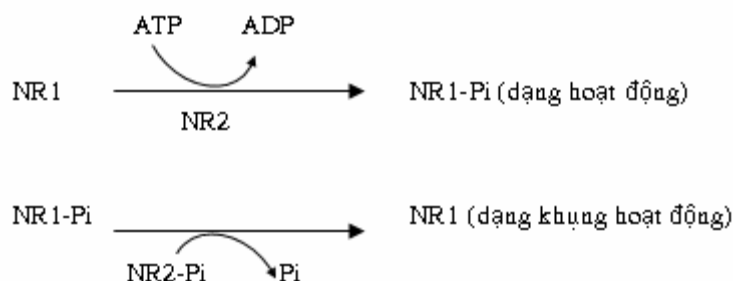
Tín hiệu kết thúc phiên mã có thể bị bỏ qua một khi có các tín hiệu chống kết thúc (*antitermination*). Lúc đó, ARN polymerase tiếp tục tổng hợp kéo dài sợi ARNm cho đến khi gặp tín hiệu kết thúc khác. Đây thực sự là một hình thức kiểm soát để bật mở (có phiên mã) hoặc đóng (không phiên mã nữa) đối với các gen nằm phía cuối operon. Tín hiệu chống phiên mã là protein (ví dụ như protein N ở thực khuẩn thể) có khả năng liên kết với ARN polymerase, giúp enzym vượt qua tín hiệu kết thúc hoặc ngăn cản hoạt tính của protein Rho. Hiện tượng chống kết thúc phiên mã được nghiên cứu khá chi tiết đối với hoạt động của các gen ở thực khuẩn thể (λ). Ngay sau khi vừa xâm nhiễm vào tế bào vi khuẩn *E.coli*, genome thực khuẩn thể tồn tại ở dạng vòng tròn. Phiên mã các gen của phage λ được bắt đầu tại hai promoter P_L và P_R (nằm sát cạnh nhau) bởi chính ARN polymerase vi khuẩn. Sợi ARNm phiên mã từ P_R được dịch mã sang 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 sang protein N. Đây là protein chống kết thúc phiên mã tại hai vị trí kết thúc nằm ngay sau gen mã cho Cro và N. Nhờ đó, ARN polymerase vi khuẩn tiếp tục phiên mã các gen kế tiếp trên genome thực khuẩn thể.

2.1.3 Các protein điều khiển (regulatory proteins)

Các protein điều khiển tham gia kiểm soát hoạt động của gen. Những protein này có thể tương tác với chính ADN hoặc tương tác với các enzym ARN polymerase hoặc tương tác với nhau. Vì vậy, protein điều khiển thường có nhiều domain qui định chức năng khác nhau. Hầu hết các protein điều khiển được nghiên cứu kỹ có chức năng hoạt hoá hoặc ức chế quá trình phiên mã và sau phiên mã. Một domain thường gồm nhiều đoạn peptide nằm không cạnh

nhau trên sợi polypeptide. Nhờ cấu trúc không gian của protein mà những đoạn này được phân bố cạnh nhau tạo nên trung tâm qui định hoạt tính đặc thù. Một protein có thể có nhiều domain. Ví dụ, protein hoạt hoá gen có thể có các domain khác nhau, mỗi domain có vai trò đặc hiệu đảm bảo tương tác với ARN polymerase, với ADN, hoặc với các protein khác.

Ngoài ra, một số protein điều khiển được tổng hợp dưới dạng không hoạt động. Ví dụ, khi tế bào vi khuẩn mọc trên môi trường thiếu nitơ, các operon ntr (nitrogen-regulated operons) được kích thích để tổng hợp khoảng 20-25 proteins khác nhau. Những protein này cho phép tế bào sống được trong điều kiện thiếu nitơ. Promoter của các operon ntr được nhận biết bởi ARN polymerase chứa yếu tố $\sigma 54$. Do đó việc tổng hợp yếu tố này liên quan đến hoạt động của các operon ntr. Ngoài ra chúng còn được điều khiển bởi protein NR1. Protein này bám vào ADN nằm trước vị trí +1. Gen mã cho NR1 hoạt động liên tục giống như gen lacI, do đó NR1 luôn có mặt trong tế bào ở mọi thời điểm. Tuy nhiên khi môi trường đủ nitơ, NR1 tồn tại ở dạng không hoạt động. Khi thiếu nitơ, NR1 được hoạt hoá. Quá trình chuyển đổi được mô tả trên hình 2.11.

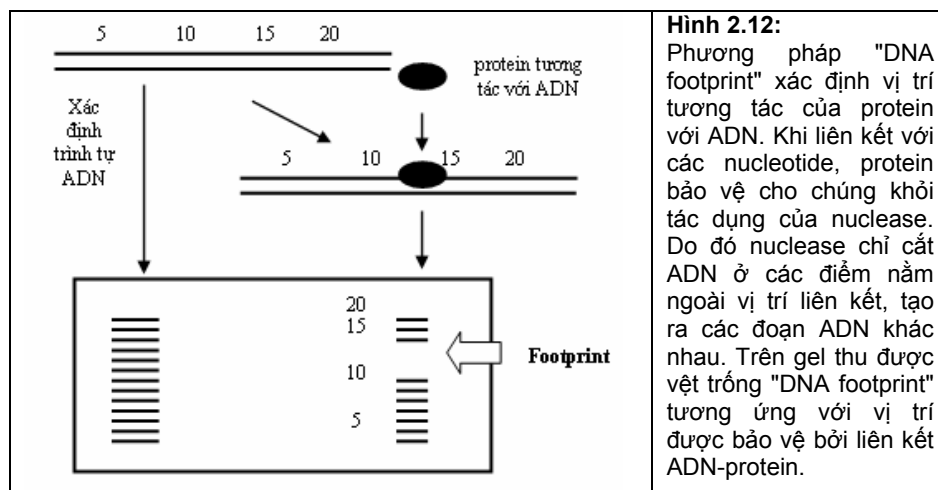


Hình 2.11:

Chu trình chuyển đổi giữa hai dạng hoạt động và dạng không hoạt động của phân tử NR1 nhờ NR2 kinase và P2-phosphatase. Trong tế bào eukaryot, các enzym kinase và phosphatase có vai trò rất quan trọng trong quá trình chuyển trạng thái của các protein điều khiển từ dạng không hoạt động sang dạng hoạt động và ngược lại.

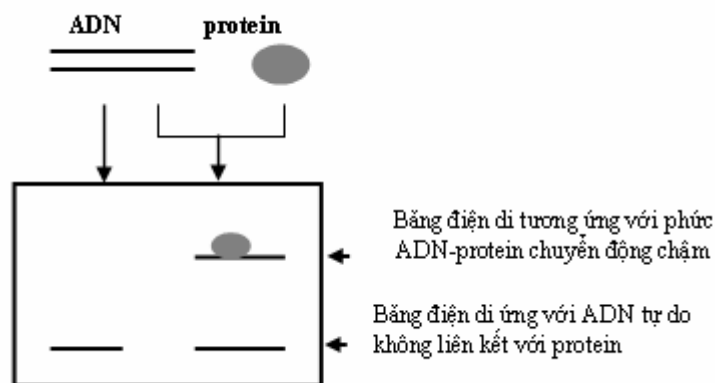
Để nghiên cứu tương tác giữa protein và acid nucleic, ngoài phương pháp dấu vân "DNA footprint" khá phổ biến, một số kỹ thuật khác hay được sử dụng như xác định băng điện di chậm (*gel retardation assay*), test thử khả năng bảo vệ biến đổi (*modification protection assay*) ...

* **Phương pháp "DNA footprint"** cho phép xác định chính xác các nucleotide được protein bảo vệ che chắn khỏi tác dụng phân hủy của ADNase (Hình 2.12). 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.



Hình 2.12: 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 vết trống "DNA footprint" tương ứng với vị trí được bảo vệ bởi liên kết ADN-protein.

* **Phương pháp xác định bằng điện di chậm** (*gel retardation assay*) 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 (*competitor 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 2.13).



Hình 2.13: Phương pháp xác định bằng 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.

2.2 Kiểm soát sau phiên mã

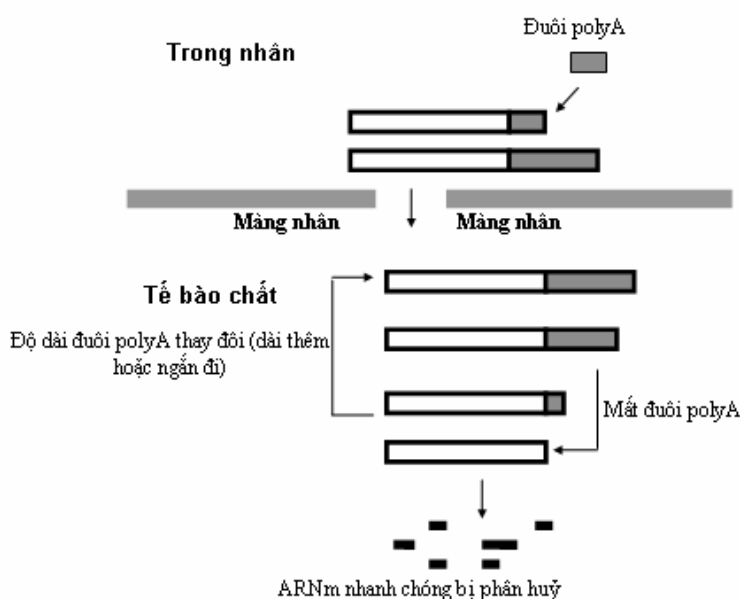
2.2.1 Tìm hãm dịch mã liên quan đến cấu trúc vùng 5'UTR của phân tử ARNm

Phân tử ARNm có các vùng 5' không dịch mã (5' UTR – 5' *Untranslated Region*) nằm phía trước mã khởi động (AUG) và vùng 3' không dịch mã (3' UTR) nằm sau mã dừng. Đầu 5' không dịch mã có thể tạo cấu trúc cặp tóc ngăn cản ribosome khởi động dịch mã. Ngoài ra, phần 5' UTR có thể là vị trí tương tác của một số protein repressor. Khi đó ribosome không

thể dịch chuyển trên sợi ARNm khiến cho quá trình dịch mã bị kìm hãm hoàn toàn. Rõ ràng trong trường hợp này, ARNm vẫn được tổng hợp một cách bình thường nhưng protein không được tạo ra. Cơ chế điều khiển quá trình dịch mã do protein tương tác với đầu 5' của ARNm được gọi là "kiểm soát dịch mã tiêu cực" (*negative translational control*).

Ở tế bào eukaryot, kiểm soát dịch mã tiêu cực 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.

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



Hình 2.14:

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

Phản ứng gắn đuôi polyA (~200 nucleotide 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 nucleotide A. Do đó đuôi polyA ngắn nhất để phân tử ARNm không bị phân hủy có thể là 30 nucleotide A. Đặc biệt, để đảm bảo độ bền vững của đuôi, các nucleotide A còn được thêm vào. Nói cách khác, việc thay đổi độ dài đuôi polyA được kiểm tra rất chặt chẽ (Hình 2.14).

2.2.3 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 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.

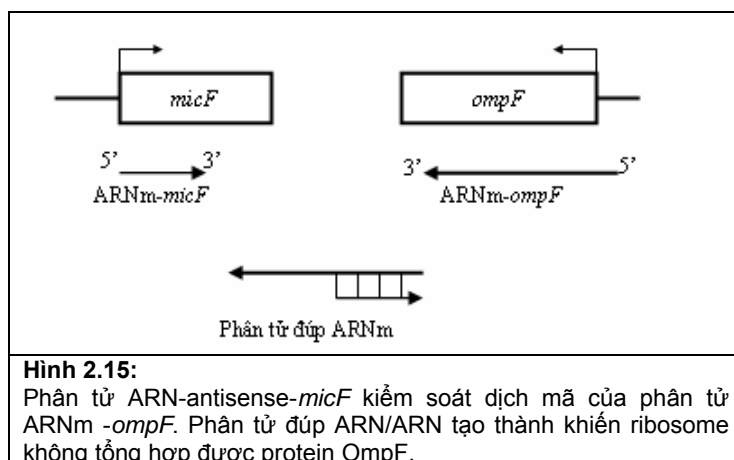
Enzym nuclease chịu trách nhiệm phân hủy ARNm đến nay vẫn chưa được xác định rõ ràng. Đối với một số ARNm polycistronic của các operon như lac hoặc trp, chúng bị phân cắt bởi endonuclease tạo ra các monocistronic tương ứng với từng gen trong operon. Khoảng cách giữa các monocistronic không chứa mã di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

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.

2.2.4 ARN anti-sense

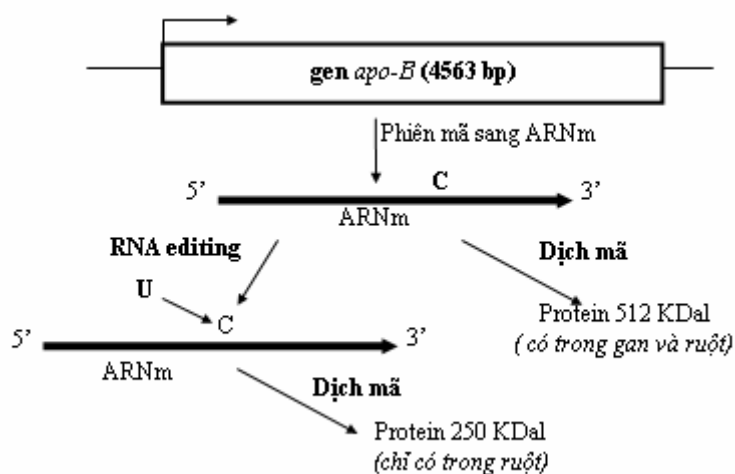
Thông thường, các protein giữ vai trò chủ đạo kiểm soát quá trình phiên mã cũng như dịch mã. Tuy nhiên hoạt động của một số gen lại được điều khiển bởi các ARNs. Cũng giống như các protein điều khiển, các ARNs làm nhiệm vụ kiểm soát hoạt động của gen được tổng hợp một cách độc lập và tương tác với các vị trí đặc hiệu (các đoạn nucleotide). Xét trường hợp cụ thể về vai trò điều khiển sau phiên mã của ARNs đối với việc tổng hợp protein OmpF ở tế bào vi khuẩn E.coli. Protein OmpF nằm phía ngoài màng tế bào làm nhiệm vụ nhận biết những thay đổi về áp suất thẩm thấu trong môi trường. Khi áp suất tăng thì gen điều khiển *micF* hoạt động mạnh hơn. Sản phẩm của gen *micF* là phân tử ARN 174 bases, có thể tạo cặp theo nguyên tắc bổ sung với vị trí nhận biết của ribosome trên phân tử ARNm của gen *ompF*. Như vậy phân tử ARNm của *micF* làm nhiệm vụ ngăn cản việc tổng hợp protein OmpF. Các phân tử ARN có chức năng tương tự *micF* được gọi là ARN-antisense (Hình 2.15).



2.2.5 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.

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.



Hình 2.16:

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.

Ở động vật, các phân tử ARNm ít trải qua phản ứng "RNA editing". Trường hợp đầu tiên được phát hiện đối với gen apolipoprotein B (apo-B) dài 4503 bp, là gen đơn bản trong genome người (Hình 2.16). 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 dừng phản ứng tổng hợp protein. Protein ngắn tương ứng với loại ARNm này có trọng lượng phân tử 250 KDa và chỉ tồn tại ở trong ruột. Phân tử protein 512 KDa tương ứng với ARNm không bị đọc sửa tồn tại ở cả gan và ruột.

Phản ứng "RNA editing" được thực hiện nhờ các phân tử ARN dài khoảng 40 - 80 base. Các ARN này gọi là ARN phụ trợ, 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.

2.3 Kiểm soát ở giai đoạn dịch mã và sau dịch mã

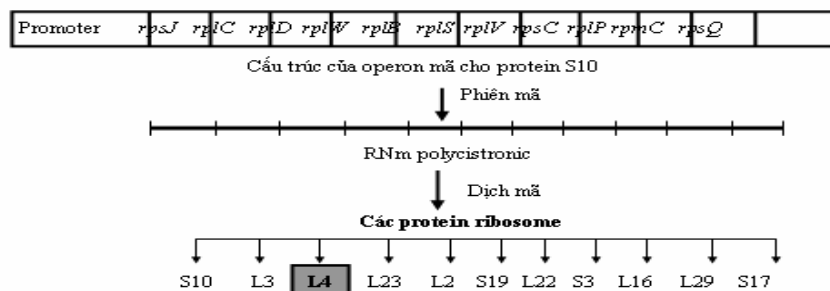
Kết quả nghiên cứu hoạt động của operon lac và operon ara được xem là những thí dụ kinh điển cho cách thức kiểm soát tiêu cực và tích cực ở giai đoạn phiên mã. Ngoài ra, quá trình phiên mã của các gen ở genome thực khuẩn thể Lamda sau khi xâm nhiễm vào tế bào E.coli đã minh họa rõ ràng cho sự phối hợp giữa hai cơ chế kiểm soát tiêu cực và tích cực cùng với một số cơ chế khác như tự điều khiển, kiểm soát phản hồi tiêu cực... Đối với tế bào prokaryot, các phân tử ARNm polycistronic thường mang thông tin di truyền của nhiều gen. Các gen trên cùng một operon được phiên mã đồng thời. Tuy nhiên sản phẩm của các gen này không được tổng hợp với số lượng như nhau. Ví dụ, khi mọc trên môi trường chỉ chứa lactose, một tế bào E.coli có khoảng 3000 phân tử β -galactosidase, 1500 phân tử β -galactoside permease và 600 phân tử β -galactoside transacetylase. Như vậy operon lac chứa 3 gen và sản phẩm của từng gen là khác nhau, mặc dù chỉ có một loại ARNm được tổng hợp. Rõ ràng rằng, quá trình dịch mã trên phân tử ARNm polycistronic được kiểm soát chặt chẽ.

Quá trình tổng hợp các protein ribosome ở E.coli được xem là thí dụ minh họa rõ ràng cho cơ chế kiểm soát sau phiên mã (Hình 2.17). Ribosome E.coli gồm 3 phân tử ARNr và 52 chuỗi polypeptide. Chúng được tạo ra một cách đồng bộ với nhau, không có sự thiếu hay thừa bất cứ thành phần nào nhằm tránh lãng phí năng lượng. Hầu hết các gen mã cho protein ribosome nằm thành nhóm trong cấu trúc operon. Ví dụ, operon chứa gen mã cho protein S10 (protein số 10 trong tiểu đơn vị nhỏ ribosome) gồm 11 gen, tất cả cùng mã cho protein ribosome.

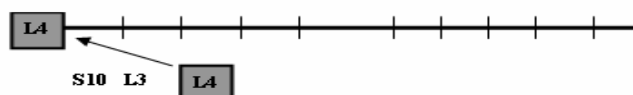
Trong số đó, gen rplD mã cho protein L4 ở tiểu đơn vị lớn ribosome làm nhiệm vụ kiểm soát tiêu cực quá trình dịch mã trên ARNm polycistronic của operon. Bình thường, khi có ARNr tự do trong tế bào, L4 tương tác với ARNr tham gia hình thành ribosome. Tuy nhiên khi vắng mặt ARNr, L4 tương tác với đầu 5' của ARNm polycistronic và ngăn cản phản ứng tổng hợp protein.

Ngăn cản quá trình dịch mã bởi một trong số các protein có mã di truyền trên phân tử ARNm polycistronic xảy ra khá phổ biến ở tế bào prokaryot. Cơ chế này được gọi là điều khiển tiêu cực phản hồi (negative autogenous regulation) hoặc tự điều khiển tiêu cực (negative self-regulation). Khi sản phẩm của gen thuộc các thành phần cấu trúc ở bào quan, cơ chế tự điều khiển thường xảy ra bởi các phân tử dạng đơn phân (monomer) có mặt tự do trong tế bào.

a. Hoạt động của operon mã cho S10 ở *E.coli* khi có ARNr tự do trong tế bào



b. Protein L4 ngăn cản dịch mã khi không có ARNr tự do trong tế bào



Hình 2.17:

Cấu trúc và kiểm soát dịch mã trên operon mã cho protein S10 ở *E.coli*. Operon này gồm có 11 gen mã cho các protein ribosome. Protein L4 của operon làm nhiệm vụ kiểm soát dịch mã trên phân tử ARNm polycistronic. Khi tế bào có ARNr tự do, 11 protein được tổng hợp đồng bộ đảm bảo cho quá trình hình thành ribosome. Tuy nhiên, khi thiếu vắng ARNr, L4 tương tác với đầu 5' của ARNm, ngăn cản không cho quá trình dịch mã xảy ra (theo Lodish và cs., 2000).

Hoạt động của một số gen mã cho các enzym liên quan đến một chu trình chuyển hóa sinh tổng hợp trong tế bào thường được kiểm soát bởi sản phẩm cuối cùng. Một khi sản phẩm cuối cùng được cung cấp đầy đủ thì chúng thường quay lại ức chế hoạt động của các gen mã cho các protein hoạt động trước trong chu trình đó. Đây chính là sự kim hãm phản hồi (feedback inhibition). Nhờ đó, sản phẩm cuối cùng luôn được duy trì ở nồng độ tối ưu mà tế bào yêu cầu. Trong cơ chế kiểm soát này, sản phẩm cuối cùng thường có khả năng tương tác với các sản phẩm hoạt động trước nó. Như vậy, sản phẩm trung gian (thường là các enzym) có chứa đồng thời hai vị trí liên kết với cơ chất và với sản phẩm cuối. Tương tác với sản phẩm cuối khiến cho enzym bị biến đổi cấu hình, do đó ái lực liên kết của enzym với cơ chất bị giảm. Như vậy mặc dù sản phẩm trung gian được tổng hợp, nhưng chúng không thực hiện được chức năng của mình.

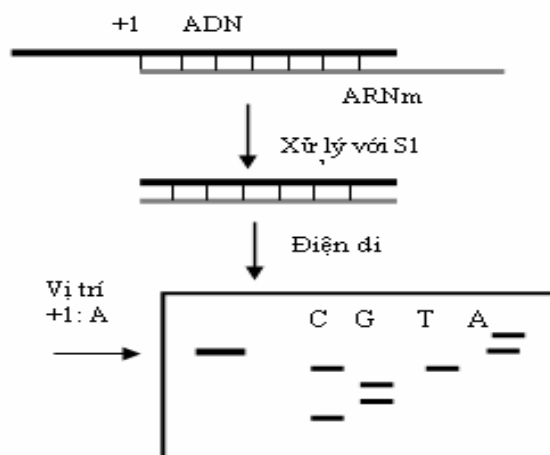
Bên cạnh ức chế phản hồi, hoạt động của một số gen mã cho enzym được kiểm soát theo con đường hoạt hoá phản hồi (feedback activation). Cấu trúc của các enzym thực hiện chức năng này thường có các vị trí tương tác với vài cơ chất hoặc một số phân tử trọng lượng nhỏ. Ngoài ra, một enzym còn có thể tham gia cả hai cơ chế điều khiển ức chế và hoạt hoá phản hồi. Ví dụ, enzym glutamine synthetase xúc tác cho phản ứng cuối cùng trong quá trình sinh tổng hợp acid amin glutamine. Enzym này có cấu trúc đa phần (multimer) liên quan đến cả hai cơ chế kiểm soát sau dịch mã của 16 sản phẩm trung gian trong quá trình tổng hợp acid amin glutamine ở cả tế bào nhân sơ và tế bào nhân chuẩn. Glutamine synthetase thực hiện chức năng điều khiển thông qua việc biến đổi cấu hình không gian của các sản phẩm khiến chúng không có hoạt tính.

Trong tế bào prokaryot, quá trình phiên mã, dịch mã và phân hủy ARNm thường bắt đầu ngay khi ARNm chưa được tổng hợp xong. Như vậy, tại bất kỳ thời điểm nào, mỗi phân tử ARNm có thể chịu sự kiểm soát của cả 3 quá trình trên. Cả 3 quá trình có thể xảy ra một cách đồng thời trên một phân tử ARNm. Đặc biệt tốc độ di chuyển của ribosome cũng như mức độ phân hủy xảy ra ở những vùng đặc biệt của ARNm là khác nhau ngay trên một phân tử ARNm polycistronic. Ngoài ra, phân tử ARNm (transfer-messenger RNA) tương tác với

ARNm có khả năng đánh dấu các protein bị tổng hợp sai bằng việc gắn đoạn peptide ngắn vào các phân tử này để sau đó chúng bị phân hủy.

2.4 Biến đổi phân tử ARNm trong tế bào eukaryot

Trong tế bào eukaryot, quá trình phiên mã phụ thuộc vào từng loại promoter cũng như từng loại enzym ARN polymerase. Các ARNr được tổng hợp nhờ ARN polymerase I, ARNm được tổng hợp bởi ARN polymerase II và các ARNt và 5S ARNr được phiên mã nhờ ARN polymerase III. Promoter cho các ARN polymerase I và II thường nằm trước vùng chứa mã di truyền trong khi một số promoter cho ARN polymerase III nằm phía sau điểm bắt đầu phiên mã (tức là promoter của ARN polymerase III nằm ngay trong vùng chứa mã di truyền). Cùng với các ARN polymerase, phản ứng tổng hợp ARN trong tế bào eukaryot chỉ thực hiện được khi có sự tham gia của nhiều protein khác. Chúng tạo thuận lợi để enzym khởi động phiên mã tại promoter của các gen eukaryot. Đây là điều khác biệt giữa quá trình phiên mã ở tế bào prokaryot và eukaryot. Promoter ở prokaryot thường kim hãm bởi protein ức chế, vì thế chúng được xem là promoter mạnh. Ngược lại, promoter ở eukaryot thường đòi hỏi phải có các protein hoạt hoá; chúng được xem là promoter yếu.



Hình 2.18:

Kỹ thuật xác định vị trí +1 sử dụng S1 nuclease. Đoạn ADN chứa vị trí +1 được lai với phân tử ARNm. Phân tử lai dạng kép này được xử lý với S1 nuclease để phân hủy phần nucleotide không lai. Sau đó, phân tử lai được điện di cùng với mẫu xác định trình tự của chính đoạn ADN ban đầu. Trên Hình 2.12 cho thấy nucleotide A chính là vị trí đầu tiên được phiên mã sang phân tử ARNm.

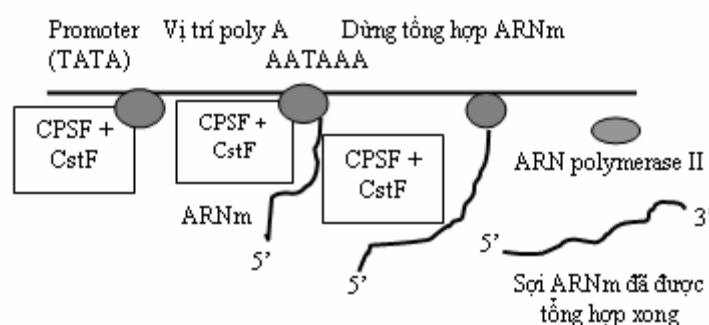
Để xác định vị trí +1, tức là nucleotide đầu tiên được phiên mã sang phân tử ARNm, chúng ta có thể áp dụng kỹ thuật "S1 nuclease mapping" (xác định vị trí +1 bởi S1 nuclease). Đầu tiên, đoạn ADN (sàng lọc từ genome) có chứa phần 5' của gen được lai với ARNm của chính gen đó. Tiếp theo, phần nucleotide ADN hoặc ARNm không lai với nhau mà ở dạng sợi đơn sẽ bị phân hủy bởi S1 nuclease. Chỉ có phân tử lai dạng kép được điện di trên gel acrylamide cùng với phản ứng xác định trình tự của chính đoạn ADN ban đầu. Nhờ đó, vị trí nucleotide đầu tiên +1 được xác định chính xác (Hình 2.18).

Quá trình tổng hợp ARNm eukaryot xảy ra nhờ ARN polymerase II. Tuy nhiên enzym này không tự bắt đầu phản ứng mà đòi hỏi sự phối hợp của nhiều yếu tố phiên mã. Nhờ tương tác với ARN polymerase II hoặc với ADN ở promoter hay ở các vị trí khác ngoài promoter,

các yếu tố này giúp enzym nhận biết dễ dàng các vị trí bảo toàn trên promoter để bắt đầu quá trình phiên mã.

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) khoảng 0,5 đến 2 kb. 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 2.19).

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" Gm7 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.



Hình 2.19:

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.

Sau khi có đuôi polyA và mũ Gm7, 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.

2.4.1 Phản ứng cắt intron và nối exon

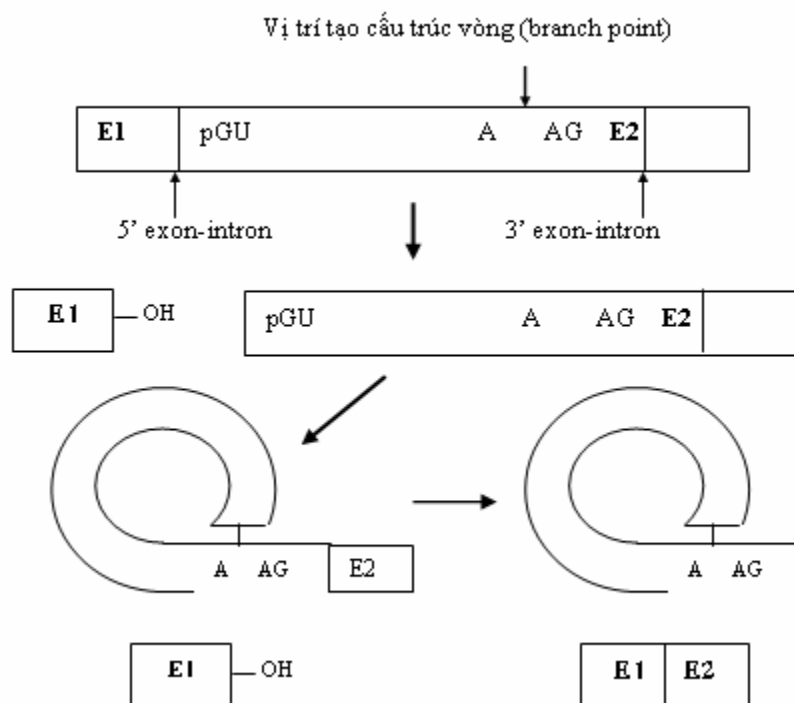
Phản ứng cắt intron và nối hai exon với nhau được thực hiện theo các bước cơ bản như sau (Hình 2.20):

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

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

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

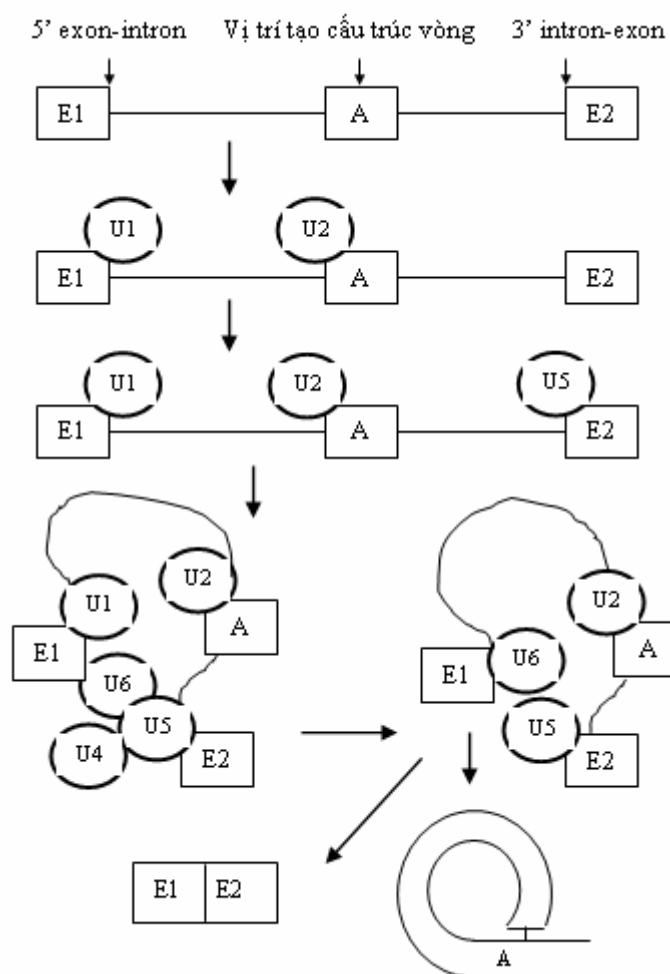
Phản ứng cắt bỏ intron đòi hỏi sự tham gia của một số phân tử snARNs kích thước nhỏ (small nuclear RNA). Chúng liên kết với protein tạo thành phức ribonucleoprotein (snRNPs). Một snRNP thường chứa một phân tử ARN và khoảng 10 protein khác nhau. Một số protein đặc hiệu cho từng loại snRNP, một số khác chung cho các snRNPs. Phân tử snARN của snRNP có khả năng tạo liên kết theo nguyên tắc bổ sung với các đoạn nucleotide ở vùng biên giới 5' hoặc 3' giữa exon và intron. Một số snARNs 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.



Hình 2.20:

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.

Thực nghiệm đã tìm ra ít nhất có 6 phân tử snARN kí hiệu từ U1 đến U6 tương ứng với 6 snRNPs. Phân tử snARN U1 liên kết với 15-20 nucleotide tại nơi tiếp giáp 5' exon-intron, snARN U2 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 (Hình 2.21).



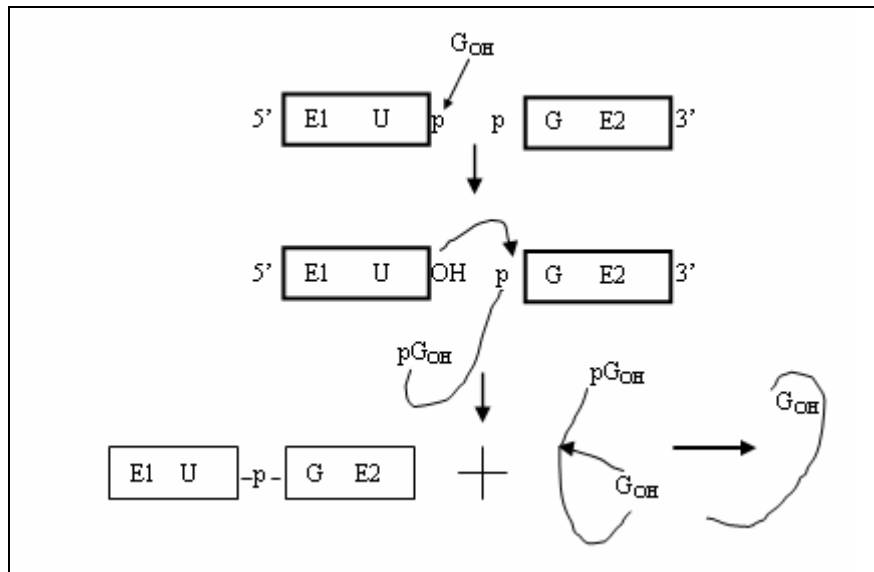
Hình 2.21:

Tương tác giữa các snRNP tạo spliceosome và hoạt động của chúng trong phản ứng loại intron, nối các 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 điện tử gọi là các spliceosome. Chúng có kích thước tương tự như ribosome, bao gồm các liên kết ARN-ARN, ARN-protein và protein-protein. Một số protein của spliceosome là các enzym helicase phân giải ATP, giải phóng năng lượng làm thay đổi cấu trúc không gian cần thiết của phân tử ARNm.

2.4.2 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 snRNPs 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, tùy thuộc vào cách thức xảy ra phản ứng.

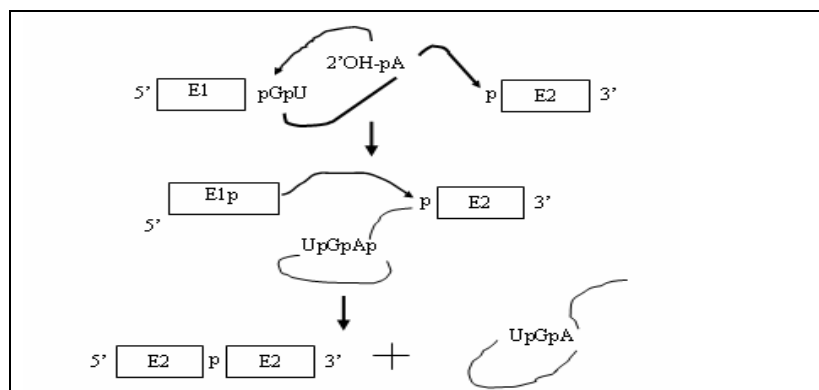


Hình 2.22:

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 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 2.22).

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 snRNPs (Hình 2.23)



Hình 2.23:

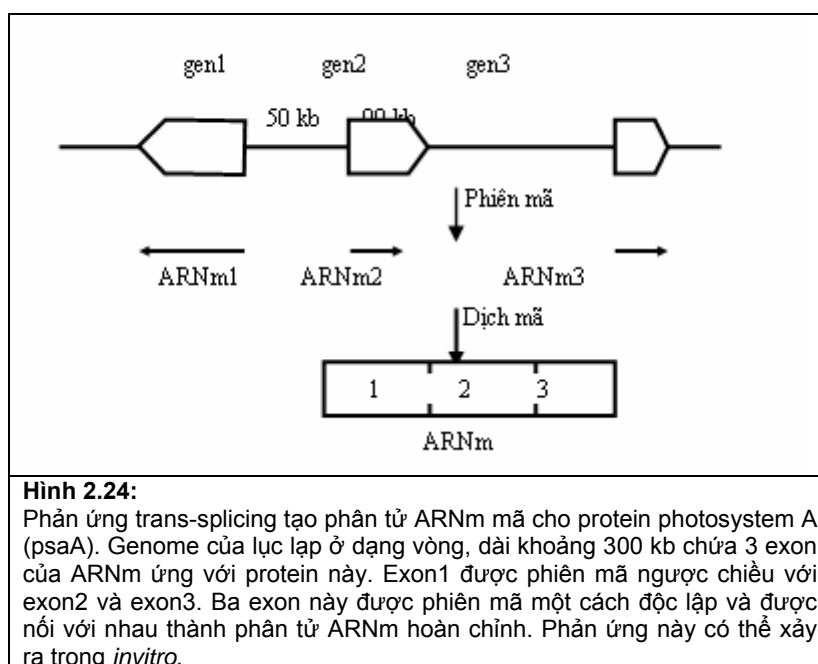
Phản ứng self-splicing của các intron nhóm II. Nhóm 2'OH của A nằm gần 3' intron-exon tương tác với đầu 3' của exon 1. Nơi tiếp giáp 5' exon-intron bị bẻ gãy tạo cấu trúc vòng cho intron. Tiếp theo phản ứng nối hai exon với nhau, intron được giải phóng ra dưới dạng vòng.

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. 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 snRNPs trong spliceosome.

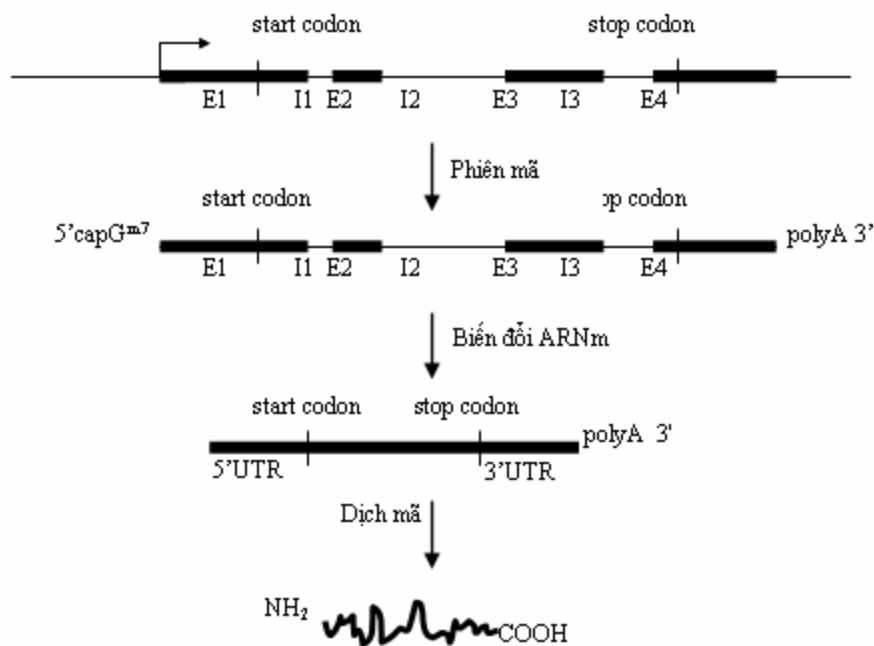
2.4.3 Phản ứng trans-splicing nối hai exon của hai phân tử ARNm

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 (Hình 2.24).



2.4.4 Cấu trúc chung của phân tử ARNm

Ngoài phản ứng cắt intron nối exon, phân tử ARNm eukaryot còn chịu các biến đổi ở đầu 5' và 3'. Đầu 5' của ARNm có 7-methylguanylate (m7G) liên kết với nucleotide đầu tiên của ARNm. Mũ m7G này đóng vai trò giúp ribosome nhận biết ARNm trong quá trình tổng hợp polypeptide. Đầu 3' của ARNm có đuôi polyA được xem là có vai trò trong việc vận chuyển ARNm từ trong nhân ra ngoài tế bào chất. Một phân tử ARNm eukaryot hoàn chỉnh chỉ chứa exon, mang mũ Gm7 ở đầu 5', đuôi polyA ở đầu 3'. Phân tử ARNm có chứa vùng 5' không dịch mã (5' UTR) và vùng 3' không dịch mã (3'UTR). Phần 5'UTR nằm trước mã bắt đầu tổng hợp peptide "start codon" còn phần 3'UTR nằm sau mã dừng tổng hợp protein "stop codon" (Hình 2.25). Các vùng này liên quan đến kiểm soát tính bền vững của ARNm cũng như kiểm soát quá trình dịch mã trên sợi ARNm. Vùng 5' không dịch mã thường chứa các đoạn oligonucleotide có thể tương tác với các protein đặc biệt hoặc có thể tạo cấu trúc bậc hai (tạo các liên kết bổ sung giữa chúng). Những tương tác đó ngăn cản ribosome di chuyển trên ARNm. Mặt khác, vùng 3' UTR của một số ARNm có các vị trí nhận biết bởi endonuclease hoặc chứa các đoạn nucleotide đặc biệt (giàu AU) kích thích sự phân hủy ARNm.



Hình 2.25:

Quá trình phiên mã và dịch mã. Gen chứa các exon (E) và các intron (I) được phiên mã sang phân tử ARNm tiền thân. Quá trình biến đổi ARNm thành phân tử hoàn chỉnh để sử dụng làm khuôn dịch mã trong quá trình tổng hợp protein.

Phân tử ARNm eukaryot khá bền vững. Thời gian bán sống cũng như khả năng dùng làm khuôn dịch mã phụ thuộc phần nào vào độ dài đuôi polyA. Trong thực tế, không phát hiện được phân tử ARNm có đuôi polyA ngắn hơn 30A. Độ dài của đuôi liên quan đến tính bền vững của ARNm nên việc thêm hoặc bớt các nucleotide A vào đuôi được kiểm soát rất chặt chẽ (cần lưu ý đuôi polyA được gắn vào ARNm ngay sau khi phiên mã ở trong nhân. Tuy nhiên thay đổi độ dài của đuôi có thể xảy ra trong tế bào chất). Đặc biệt trong quá trình phát triển mô phôi, rất nhiều loại ARNm được phiên mã từ các gen của mẹ và được dự trữ trong tế bào trứng. Chúng thường có đuôi polyA rất ngắn (10-30 A) nên không được dịch mã. Chỉ sau khi trứng thụ tinh, các ARNm này được thêm đuôi polyA và trở thành khuôn để tổng hợp protein.

Chương 3

KỸ THUẬT ADN TÁI TỔ HỢP

Bắt đầu từ những năm 1970 trở đi, các phương pháp nghiên cứu, phân tích ADN và gen được phát triển mạnh mẽ. Nhờ những kỹ thuật hiện đại, gọi chung là kỹ thuật tách dòng và tái tổ hợp, một gen bất kỳ có thể được phân lập và được tách ra khỏi genome. Gen này được biến đổi theo ý muốn và được đưa trở lại tế bào nuôi cấy. Tuy khó khăn phức tạp hơn, gen mang đột biến được đưa vào tế bào sinh dục. Chúng hoạt động trong genome của cá thể và được di truyền cho các thế hệ sau.

Kỹ thuật tách dòng liên quan đến việc tạo ra phân tử ADN tái tổ hợp từ các đoạn ADN có nguồn gốc khác nhau. Để cắt và nối các đoạn ADN với nhau, hay nói cách khác để có thể thao tác với các đoạn ADN, nhất thiết phải sử dụng các enzym giới hạn (*restriction endonucleases*) và ADN ligase. Ví dụ, một đoạn ADN có chứa gen được cắt ra từ genome người và được ghép vào plasmid 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*).

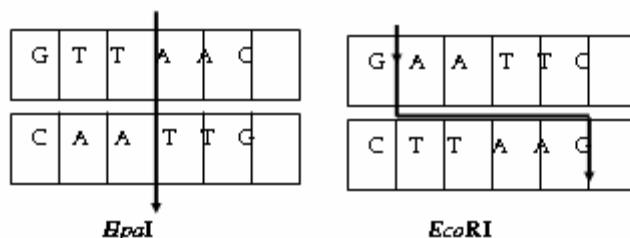
Các đoạn ADN đưa vào vector có thể xuất phát từ các nguồn khác nhau. Nguồn sơ cấp (*primary source*) có thể là ADN tách chiết trực tiếp từ genome hoặc lấy từ ngân hàng ADN. Nguồn thứ cấp (*secondary source*) thường là đoạn ADN phân lập từ nguồn sơ cấp được cắt thành các đoạn nhỏ hơn tiếp tục đưa vào các loại vector khác (*subcloning*).

Thông thường, để tách được một gen bất kỳ từ genome cần phải biết trước một đoạn ADN liên quan đến gen đó (hoặc nằm gần nó). Đoạn này được sử dụng như mồi hay còn gọi là đầu dò (*probe*) để tìm ra gen cần nghiên cứu. Khi sản phẩm của gen đã biết, có thể xuất phát từ protein để suy diễn trình tự nucleotide mã cho nó. Đó là những thông tin đầu tiên cần thiết giúp cho nghiên cứu ARNm và ADN (*gen*) tương ứng với protein.

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 *in vitro* và *in vivo* 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? vv... 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.

3.1 Phân cắt, phân ly ADN

Ngày nay, chúng ta có thể dễ dàng cắt genome của tế bào eukaryot gồm các phân tử ADN có kích thước rất lớn (khoảng 10^6 đến 10^{11} nucleotide) thành các đoạn ADN có kích thước theo ý muốn nhờ các enzym giới hạn (*restriction enzymes* hoặc *restriction endonucleases*). Vị trí cắt có thể nằm cách xa hoặc ngay tại vị trí nhận biết. Các enzym nhận biết và cắt ngay tại một vị trí được sử dụng nhiều trong kỹ thuật ADN tái tổ hợp. Chúng nhận biết các vị trí đặc hiệu gồm 4 đến 8 nucleotide trên phân tử ADN và cắt các phân tử này thành các đoạn khác nhau. Hiện nay các hãng thương mại đã giới thiệu gần 600 enzym giới hạn được sản xuất theo công nghệ ADN tái tổ hợp. Chúng được tìm ra và tách chiết chủ yếu từ vi khuẩn.



Hình 3.1:

Phân tử ADN bị cắt bởi enzym giới hạn. Các enzym này nhận biết trình tự sắp xếp của 4 nucleotide và cắt ADN ở các vị trí này. Một số enzym cắt ADN thành các đoạn có đầu bằng (các đầu "Blunt ends") như *HpaI*. Ngoài ra còn có các enzym cắt ADN tạo ra các đầu dính hay còn gọi là đầu so le (các đầu "Cohensive ends") như *EcoRI*.

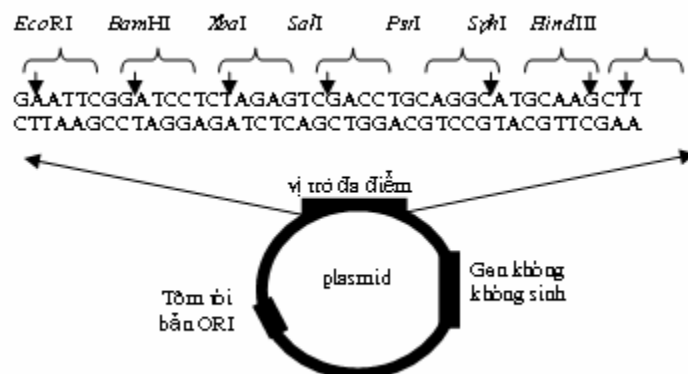
Các đoạn ADN có kích thước khác nhau được phân ly bằng phương pháp điện di trên gel agarose (điện di ngang) hoặc gel polyacrylamide (điện di đứng). Phân tử ADN tích điện âm nên chúng chuyển động từ cực âm sang cực dương trong điện trường một chiều. Các phân tử ADN có kích thước lớn ($\sim 10^5$ - 10^7 bp) được phân ly bằng điện di trên trường xung đẩy "pulsed-field". Những đoạn dài từ 200 bp đến 10.000 bp 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 200 bp sai khác nhau không quá 10 bp 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). Sau khi điện di, các đoạn ADN 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 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 ADN rất nhỏ (cỡ ng).

3.2 Đưa các đoạn ADN vào vector

Vector được sử dụng để lưu giữ và chuyên chở ADN vào các tế bào nhận. Chúng có thể tồn tại trong tế bào vi sinh vật (ví dụ 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). Vì vậy, việc lựa chọn và thiết kế vector phụ thuộc vào mục đích nghiên cứu.

3.2.1 Các vector sử dụng trong kỹ thuật tách dòng

Tùy thuộc vào kích thước của các đoạn ADN cần lưu giữ mà chọn các vector khác nhau. Phổ biến nhất là các plasmid, genome của bacteriophage (gọi tắt là phage hay λ), cosmid, các vector BAC, YAC. Các plasmid dùng làm vector thường có kích thước nhỏ, do đó chúng không mang được các đoạn ADN có kích thước lớn (<7 kb). Tuy nhiên vector nhân tạo YAC có thể mang được đoạn ADN cỡ 1 Mb.



Hình 3.2:

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. Các enzyme này có nguồn gốc từ vi sinh vật. *EcoRI*-*Escherichia coli* RY13, *BamHI*-*Bacillus amyloliquefaciens* H, *XbaI*-*Xanthomonas badrii*, *SmaI*-*Streptomyces albus*, *PstI*-*Providencia stuartii*, *SphI*-*Streptomyces phacochromogenes* và *HindIII*-*Haemophilus influenzae* Rd.

Cấu tạo của vector sử dụng trong kỹ thuật tách dòng thường có chứa 3 yếu tố cơ bản: thứ nhất, tâm tái bản ADN (*origin of replication*) cho phép vector tái bản trong tế bào nhận. Thứ hai, chỉ thị chọn lọc (*selectable marker gene*) hay còn gọi là gen chỉ thị, thường là gene qui định khả năng chống chịu kháng sinh. Nhờ có gen này mà chúng ta có thể sàng lọc được những tế bào mang phân tử ADN tái tổ hợp. Cuối cùng là vị trí gắn đoạn ADN lạ vào (vị trí đa điểm tách dòng-*polycloning site*) (Hình 3.2).

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 1 kb 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 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 qui định tính chống chịu tetracycline (*tet^r*) và một vị trí *EcoRI* duy nhất để ghép ADN lạ vào. Ngày nay, có rất nhiều loại plasmid khác nhau 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. 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. Ví 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 có số bản copy nhiều hay được sử dụng trong các kỹ thuật ADN thông thường. Chúng thích hợp với các đoạn ADN không dài quá 7 kb. Khi đoạn ADN có kích thước lớn hơn, trải qua một số lần nhân bản, đoạn này bị ngắt dần hoặc bị loại khỏi plasmid do tế bào có xu hướng đào thải ADN không cần thiết.

Bacteriophage (vector thực khuẩn thể): Hầu hết các vector thực khuẩn thể được thiết kế dựa vào nhiễm sắc thể của thực khuẩn thể λ . Nhiễm sắc thể thực khuẩn thể λ gồm 48.502 bp mã cho 46 gen. Khi xâm nhiễm vào tế bào vi khuẩn, nhiễm sắc thể λ tồn tại ở dạng vòng do liên kết của các vị trí *cos* ở hai đầu của phân tử ADN thẳng. 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 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 15 kb). 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 15 kb. 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.

Cosmid: Nhiễm sắc thể λ chuyển từ dạng thẳng sang dạng vòng nhờ các vị trí *cos* ở hai đầu của sợi ADN thẳng. Do đó trong thí nghiệm *in-vitro*, một phân tử ADN có kích thước thích hợp mang vị trí *cos* này có thể được "đóng gói" trong vỏ protein. Cosmid là các plasmid mang các vị trí *cos* và được "đóng gói" *in-vitro* giống như thực khuẩn thể λ . Khi các phage mang cosmid này xâm nhập vào tế bào vi khuẩn, cosmid sẽ chuyển thành dạng vòng và tồn tại như plasmid trong tế bào nhận. Tuy nhiên phage chứa vector cosmid không còn khả năng xâm nhiễm tế bào vi khuẩn vì không mang những gen cần thiết cho quá trình này. Vì vậy, cosmid được đưa vào tế bào chủ nhờ các phage phụ trợ. Cosmid có thể mang các ADN lạ dài khoảng 35 đến 45 kb. Ngoài ra, còn có loại vector **Fosmid** mang tâm tái bản của plasmid F (*E.coli*) và các trình tự *cos* của cosmid. Tuy nhiên, do tính chất của tâm tái bản nên fosmid có số lượng bản sao ít trong tế bào *E.coli*. Việc sử dụng vector Fosmid giúp hạn chế phần nào những sai sót xảy ra với ADN lạ xảy ra trong quá trình vector tái bản.

Vector bacteriophage P1: tương tự như vector thiết kế từ nhiễm sắc thể lambda λ , vector P1 được thiết kế dựa vào việc loại bỏ những đoạn ADN không cần thiết trong nhiễm sắc thể thực khuẩn thể P1. Do genome P1 lớn hơn genome của lambda λ , vector P1 có khả năng nhận đoạn ADN lạ dài tới 125 kb. Ngoài ra, khi kết hợp các đặc tính của vector P1 và vector BAC, loại vector PAC (P1-derived Artificial Chromosome) được thiết kế phù hợp với các đoạn ADN khoảng 300 kb.

Vector BAC (Bacterial Artificial Chromosomes): được thiết kế nhân tạo có những ưu điểm của vector YAC, tức là có khả năng chứa các đoạn ADN kích thước lớn cỡ 10^5 bp. Tuy nhiên, do không phức tạp như YAC nên thao tác với BAC có phần đơn giản hơn. BAC có thể

tái bản dễ dàng trong *E.coli* giống như plasmid. Những năm cuối thập kỷ 20, vector BAC đang dần dần thay thế YAC trong các nghiên cứu với đoạn ADN kích thước lớn như việc xây dựng ngân hàng, xác định bản đồ của toàn bộ nhiễm sắc thể...

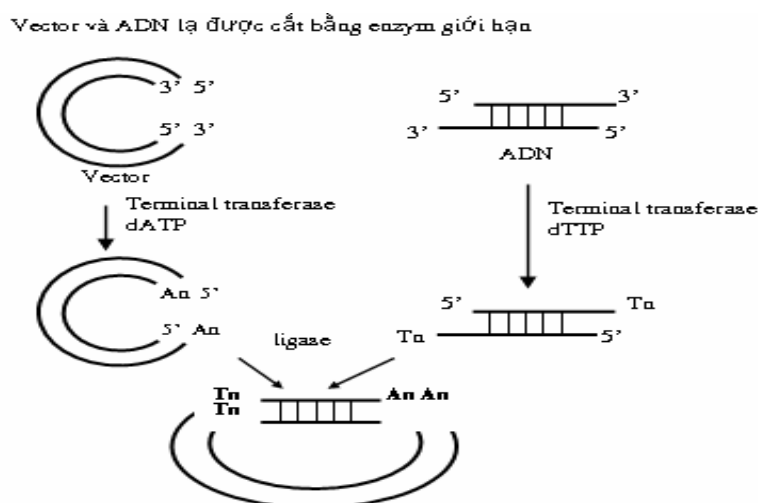
Vector YAC (Yeast Artificial Chromosome): Đây là vector được thiết kế nhân tạo dựa vào cấu trúc của nhiễm sắc thể nấm men. Vector này có chứa gen chỉ thị, vị trí đa điểm tách dòng, trình tự nucleotide của tâm tái bản ARS (Autonomously Replicating Sequence), tâm động và telomere (trình tự đầu mút nhiễm sắc thể). Như vậy, thực chất YAC là nhiễm sắc thể mini của tế bào nấm men. Thông thường, sản phẩm của gen chỉ thị trên vector YAC giúp tế bào nhận có khả năng bổ khuyết sự thiếu hụt một chất nào đó trong môi trường. Ví dụ, vector YAC mang gen chỉ thị URA^+ giúp tế bào khuyết dưỡng URA^- mọc được trên môi trường thiếu uracil. Vector YAC có thể mang ADN lạ dài tới 1 Mb (10^6 bp).

3.2.2 Đưa ADN vào vector

Để đưa một đoạn ADN vào vector, phản ứng nối giữa vector với đoạn ADN đó giữ vai trò quyết định. Hai đầu nối của vector và ADN thường liên kết với nhau theo nguyên tắc tạo cặp bổ sung. Một số phương pháp hay sử dụng như sau:

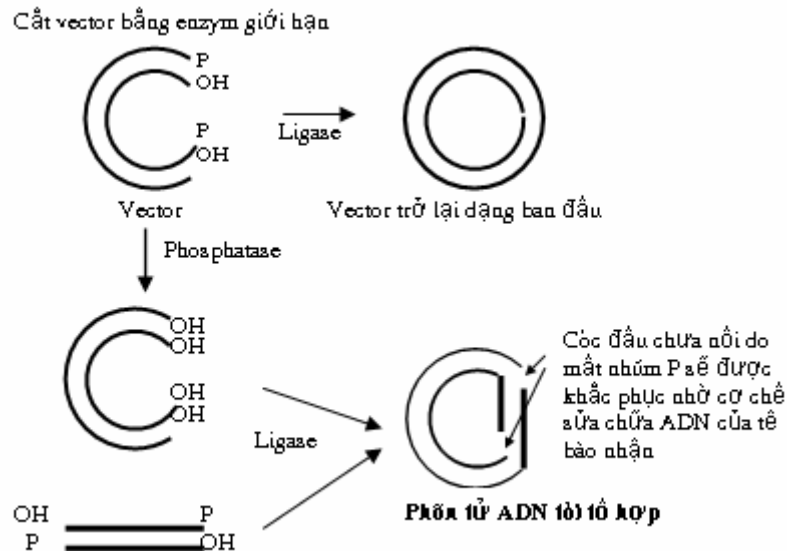
* **Sử dụng các homopolymer:** Các homopolymer có thể liên kết với nhau theo nguyên tắc tạo cặp bổ sung. Chúng được nối vào các đầu của vector và đoạn ADN. Ví dụ, oligo (polyT) được nối vào đầu 3' của ADN lạ trong khi oligo (polyA) được nối vào đầu 3' của vector. Độ dài các oligo thường khoảng từ 10 đến 40 nucleotide. Enzym terminal transferase có khả năng thêm một loại nucleotide vào các đầu cắt của vector hoặc ADN. Sau đó được gắn thêm các polyA, poly T, vector và đoạn ADN lạ được nối với nhau nhờ liên kết bổ sung giữa A/T và nhờ ADN ligase (Hình 3.3).

* Cắt ADN lạ và vector bằng enzym giới hạn tạo đầu bằng (*blunt ends*). Chúng được nối với nhau bởi T_4 ADN ligase. Enzym này có khả năng nối hai đoạn ADN có đầu bằng với nhau. Phương pháp này không đòi hỏi các nucleotide tạo cặp bổ sung ở các đầu cắt. Tuy nhiên, ADN lạ đã đưa vào vector có thể không lấy ra nguyên vẹn do vị trí nhận biết bởi enzym giới hạn ban đầu bị thay đổi.



Hình 3.3: Enzym terminal transferase thêm các đuôi homopolymer vào các đầu của vector và ADN lạ. Chúng được nối với nhau tạo phân tử tái tổ hợp ADN nhờ ligase.

* Cắt ADN và vector bởi một loại enzym giới hạn tạo các đầu so le (*cohesive ends*). Sau đó chúng được nối với nhau bởi enzym ligase. 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, vector (ở dạng thẳng) hoặc đoạn ADN sau khi bị cắt bởi enzym giới hạn, được xử lý với phosphatase để loại nhóm P ra khỏi đầu 5' (Hình 3.4).

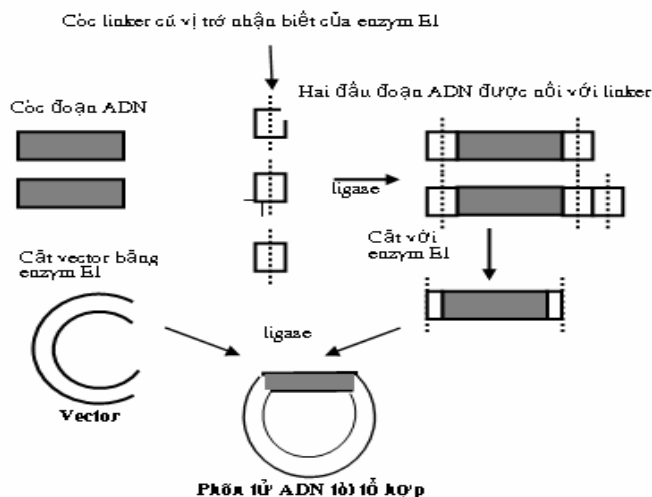


Hình 3.4:

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.

* **Phương pháp sử dụng linker:** Một số oligonucleotide được tổng hợp nhân tạo có chứa vị trí nhận biết của các enzym giới hạn được gọi là các *linker*. Đoạn ADN lạ có đầu bằng có thể được nối với linker trước khi đưa vào vector. Thông thường, enzym có vị trí nhận biết trong linker sẽ được sử dụng để cắt vector (Hình 3.5).

* **Phương pháp sử dụng linker đôi:** Phương pháp này thường được sử dụng trong quá trình xây dựng ngân hàng ADNc. Các bước chi tiết được mô tả trên hình 3.7. Phân tử ADNc được tổng hợp từ khuôn ARNm và được nối với linker S (có vị trí nhận biết của enzym S). Đầu cong "loop" được tạo ra do tính chất của enzym phiên mã ngược reverse transcriptase sẽ bị cắt bởi S1 nuclease và được sửa lại thành đầu bằng nhờ ADN polymerase. Sau đó đầu bằng này được nối với linker R (chứa vị trí nhận biết của enzym giới hạn R). Đoạn ADN tiếp tục được xử lý với hai enzym S và R tạo ra các đầu so le. Vector cũng bị cắt bởi hai enzym S và R. Bằng kỹ thuật này, có thể điều khiển được chiều ghép ADN vào vector.



Hình 3.5:

Sử dụng linker để đưa ADN vào vector. Enzym T4 ADN ligase nối các linker có chứa vị trí nhận biết của enzym E1 vào hai đầu đoạn ADN lạ. Sau đó dùng chính enzym này để cắt ADN và cắt vector. Chỉ có các đoạn ADN mang vị trí nhận biết của E1 được gắn vào vector nhờ ADN ligase.

3.3 Ngân hàng ADN

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 thường 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 và lưu giữ tất cả các đoạn trong 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 ngân hàng chỉ cung cấp thông tin có mặt trong các sợi ARNm của một tổ chức tế bào biệt hoá. 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, tùy thuộc vào mục đích sử dụng.

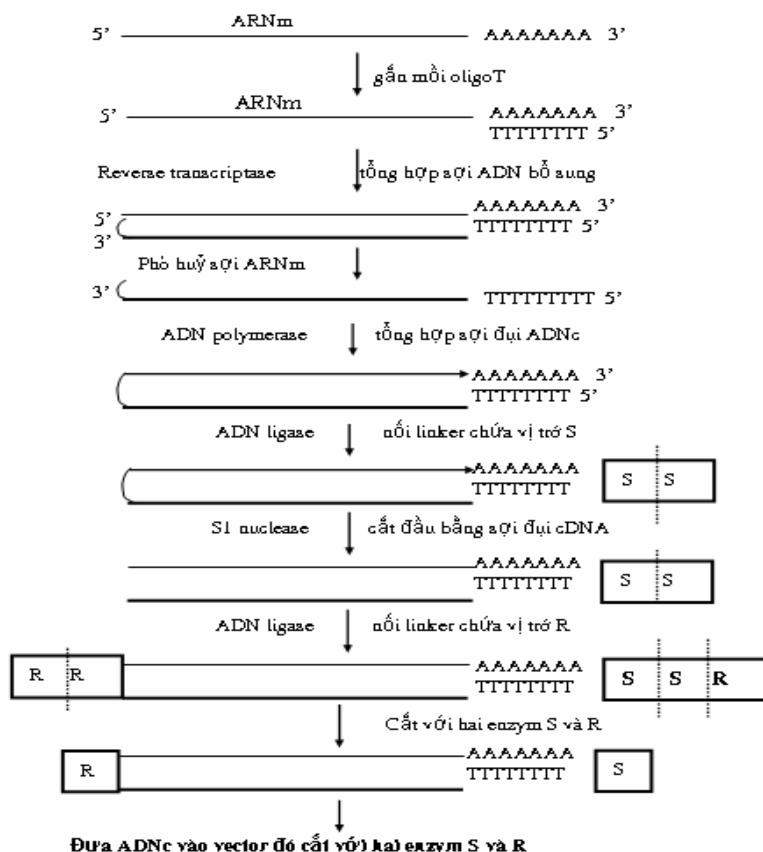
3.3.1 Ngân hàng các ADNc (cDNA library)

Tập hợp tất cả các phân tử ARNm tách chiết từ một tổ chức có thể đại diện cho tất cả các gen hoạt động trong tổ chức đó. Các ARNm chỉ chứa exon nên ngắn hơn nhiều so với các gen tương ứng. 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. Tuy nhiên, 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. Các bước cụ thể của quá trình tạo ADNc được mô tả trên hình 3.6.

Tập hợp tất cả các ADNc tương ứng với mọi ARNm của một loại mô được đưa vào vector. Mỗi vector nhận một phân tử ADNc. Tập hợp tất cả các vector chứa ADNc được chuyển vào tế bào nhận. Tập hợp tất cả các tế bào nhận chứa đầy đủ ADNc của một loại mô

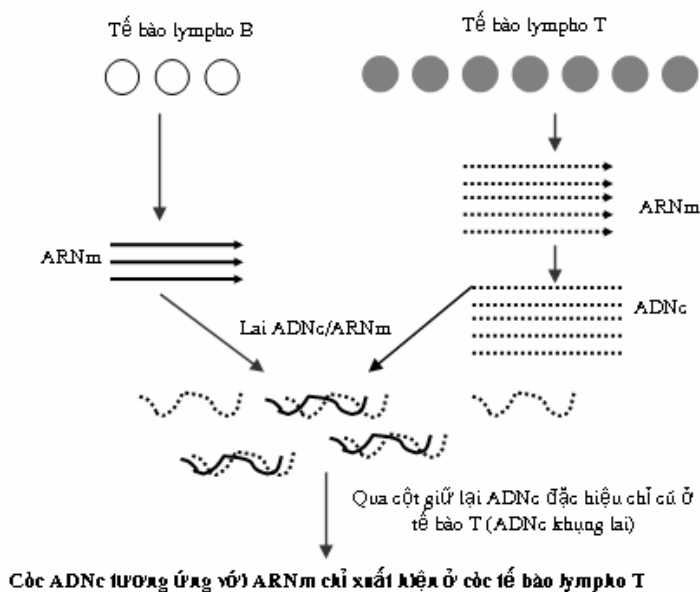
tạo thành ngân hàng ADNc của loại mô đó. Lúc này mỗi ADNc trong ngân hàng thường được gọi là một dòng (một clone). Để sàng lọc bất kỳ một clone ADNc từ ngân hàng, cần phải có môi (đầu dò) đặc hiệu đối với ADNc này. Ví dụ, chúng ta có thể chọn được ADNc ứng với một đoạn polypeptide nếu chúng ta biết trình tự acid amin. Dựa vào trình tự acid amin chúng ta suy ra trình tự các nucleotide tương ứng dựa vào mã di truyền bộ ba. Đoạn nucleotide này được tổng hợp nhân tạo và sử dụng như môi để sàng lọc ADNc ứng với polypeptide đó. Kỹ thuật sàng lọc sẽ được mô tả chi tiết ở phần sau.



Hình 3.6:

Tổng hợp ADNc từ ARNm. Enzym reverse transcriptase tổng hợp sợi đơn ADNc trên khuôn ARNm (sử dụng môi 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. Do đó khi đưa vào vector, ADNc được giữ đúng chiều.

Các vector nhận ADNc được đưa vào tế bào nhận. Quá trình này được gọi là biến nạp (*transformation*). Tế bào có khả năng nhận vector được gọi là tế bào khả biến (*competent cells*). Tế bào đã nhận vector được gọi là thể biến nạp (*transformants*). Tập hợp tất cả các tế bào chứa vector tạo thành một ngân hàng ADNc đại diện cho tổ chức mô dung tách chiết ARNm. Đối với một số phân tử ARNm chỉ xuất hiện với số lượng không lớn, chúng thường được làm giàu trước khi tổng hợp ADNc. Phương pháp lai loại trừ (*subtractive hybridization*) hay được sử dụng để phát hiện các loại ARNm đặc hiệu chỉ được tổng hợp trong một loại mô nhất định. Phương pháp này lần đầu tiên được ứng dụng để xác định ADNc tương ứng với các protein thụ thể (*receptor*) chỉ xuất hiện trên bề mặt tế bào lympho T mà không có ở lympho B (Hình 3.7).

**Hình 3.7:**

Phương pháp lai loại trừ được sử dụng để làm sạch những phân tử ADNc tương ứng với ARNm chỉ xuất hiện ở tế bào lympho T (không có ở lympho B). Phương pháp này cho phép làm giàu những phân tử đặc hiệu khác biệt giữa hai loại tế bào.

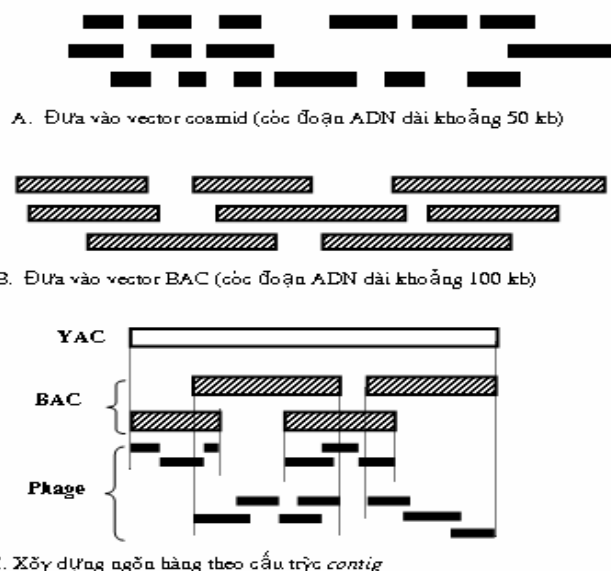
Đầu tiên, ARNm tách từ tế bào lympho T được dùng để tổng hợp các ADNc. Sau đó các ADNc (với số lượng lớn) được lai với ARNm tách từ tế bào lympho B sao cho mọi ADNc tương ứng với ARNm xuất hiện ở cả hai loại tế bào đều bị lai với nhau; chỉ còn lại ADNc tương ứng với ARNm riêng biệt của lympho T. Sau phản ứng lai, sản phẩm được đi qua cột lọc để tách riêng những phân tử ADNc ở dạng sợi đơn. Tiếp theo, chúng được dùng để tổng hợp ADNc dạng sợi kép và xây dựng ngân hàng ADNc. Ngân hàng này đặc hiệu cho riêng loại tế bào lympho T vì chỉ chứa thông tin dành riêng cho loại tế bào này. Những ngân hàng ADNc chuyên biệt như thế rất thuận lợi khi cần xác định hoạt động khác nhau của gen giữa hai loại tế bào.

Các ADNc có thể được đưa vào vector có mang sẵn promoter, thông thường là những promoter mạnh để tạo ra số lượng ARNm lớn. Những vector này được gọi là vector biểu hiện gen (*expression vector*). Dưới sự điều khiển của promoter, ADNc được phiên mã tạo ra ARNm. Đặc điểm của những sợi ARNm này là không chứa intron. Do đó, khi đưa vector biểu hiện chứa ADNc vào tế bào prokaryot, protein ứng với ADNc có thể được tổng hợp trong các thể biến nạp. Như vậy, khi đưa vector biểu hiện gen mang một ADNc hoàn chỉnh vào tế bào khả biến thì sản phẩm tương ứng của ADNc có thể được tổng hợp với số lượng lớn trong một thời gian ngắn. Ngoài ra, vector biểu hiện có thể mang đoạn nucleotide mã cho một protein đánh dấu, ví dụ như protein phát huỳnh quang (Green Fluorescence Protein- GFP). Đoạn nucleotide này nằm sau promoter và sẽ nối với ADNc. Như vậy, promoter sẽ kiểm soát biểu hiện của protein dung hợp (fusion protein) gồm đoạn peptide đánh dấu nối với protein mã bởi ADNc. Nhờ sự phát huỳnh quang của GFP mà có thể biết được hàm lượng protein sự phân bố và chức năng của nó trong tế bào sống.

3.3.2 Ngân hàng ADN genome (genomic DNA library)

Một phân tử ADNc chỉ chứa các exon của một gen, trong khi gen của tế bào eukaryot thường gồm các intron. Vì vậy, ngoài việc xây dựng ngân hàng ADNc (tương ứng với vùng

mang mã di truyền), các nhà sinh học còn xây dựng ngân hàng chứa các gen hoàn chỉnh. Điều đó giúp họ nghiên cứu sâu hơn hoạt động của các gen, ví dụ như nghiên cứu vùng ADN điều khiển, các ADN tăng cường (*enhancer*), nghiên cứu vai trò của các intron đối với điều hoà hoạt động của gen vv...



Hình 3.8:

Xây dựng các loại ngân hàng ADN genome của cùng một đối tượng. Các 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.

Để đưa toàn bộ các gen của genome vào một ngân hàng, trước hết genome phải bị cắt bởi enzym giới hạn thành các đoạn ADN có kích thước thích hợp với loại vector sẽ nhận chúng. 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 (4-6 nucleotide). Số đoạn ADN trong trường hợp thứ nhất sẽ ít hơn so với trường hợp thứ hai. Ngoài ra, cần lưu ý không phải dễ dàng nhận được một gen nguyên vẹn nằm gọn trên một đoạn ADN. Trong thực tế, một gen 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 chỉ xảy ra từng phần 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.

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 3.8). 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. Ngược lại, mỗi cơ thể có nhiều loại ngân hàng ADN đặc trưng cho loại mô tổ chức dùng để tách chiết

ARNm. Tất cả các ngân hàng ADNc chỉ chứa vùng exon của các gen. Như vậy, 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.

3.4 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 10^5 kb được cắt ra thành các đoạn trung bình 40 kb. 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 qui 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.

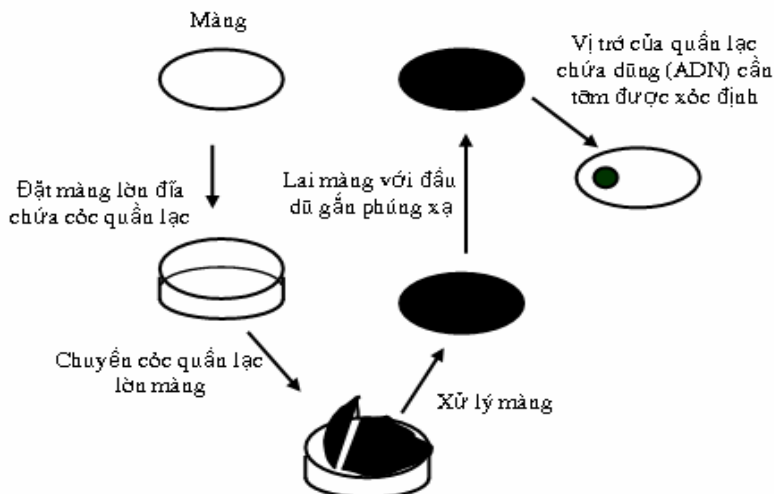
Ngoài ra, các gen tham gia một số con đường chuyển hoá trong tế bào cũng có thể sàng lọc dễ dàng theo cách thức chọn lọc di truyền. Ví dụ, muốn chọn được các gen mã cho enzyme tham gia sinh tổng hợp histidine ở *S.cerevisiae* thì có thể xây dựng ngân hàng ADN genome nấm men gồm các vector tái tổ hợp 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.

3.4.1 Phương pháp sàng lọc chung

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 (*probe*) 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.



Hình 3.9:

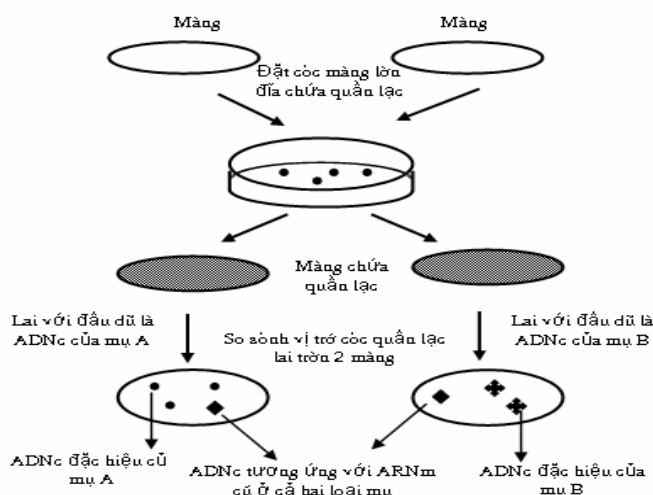
Sàng lọc quần lạc mang dòng cần tìm từ ngân hàng. Màng chứa các quần lạc được gọi là các replica. Chúng được xử lý với kiềm và nhiệt độ trước khi lai với đầu dò gắn phóng xạ hoặc phát huỳnh quang. Sau khi lai, vị trí quần lạc lai với đầu dò được phát hiện trên phim tự ghi phóng xạ hoặc dựa vào tín hiệu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

Phương pháp phổ biến được áp dụng trước đây để sàng lọc (*screening*) một dòng từ ngân hàng được mô tả trên hình 3.9. 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 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. Sử dụng đầu dò đánh dấu phóng xạ cho phép xác định nhanh, nhạy, chính xác dòng mong muốn trong số hàng triệu dòng. Quá trình này thường được lặp lại 3 lần với các điều kiện lai thay đổi để tăng khả năng sàng lọc, loại trừ các dòng lai nhiễu do đầu dò có khả năng lai với cả các đoạn ADN có trình tự tương đồng với đầu dò.

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.

3.4.2 Phương pháp sàng lọc phân biệt "differential screening"

Phương pháp sàng lọc phân biệt phù hợp với mục đích phân lập các ADNc xuất phát từ ARNm đặc hiệu cho từng loại mô hoặc khi các ARNm chỉ xuất hiện trong từng giai đoạn phát triển cơ thể. Phương pháp này cho phép xác định được các dòng giống nhau trong hai hay nhiều loại ngân hàng ADNc. Đồng thời, các dòng đặc hiệu cho từng loại mô cũng được phân biệt rõ ràng. Các bước cụ thể của phương pháp sàng lọc phân biệt được mô tả trên hình 3.10. Các dòng tế bào của một ngân hàng được trải trên đĩa petri. Sau đó các quần lạc được chuyển sang hai màng. Mỗi màng được lai với đầu dò phóng xạ khác nhau. Đầu dò thứ nhất bao gồm các ADNc tương ứng với ARNm của loại mô A. Đầu dò thứ hai gồm các ADNc ứng với ARNm của loại mô B. So sánh sự có mặt hoặc độ đậm nhạt của các vết lai trên phim tự ghi phóng xạ có thể xác định được các ARNm chỉ xuất hiện trong một hoặc trong cả hai loại mô.



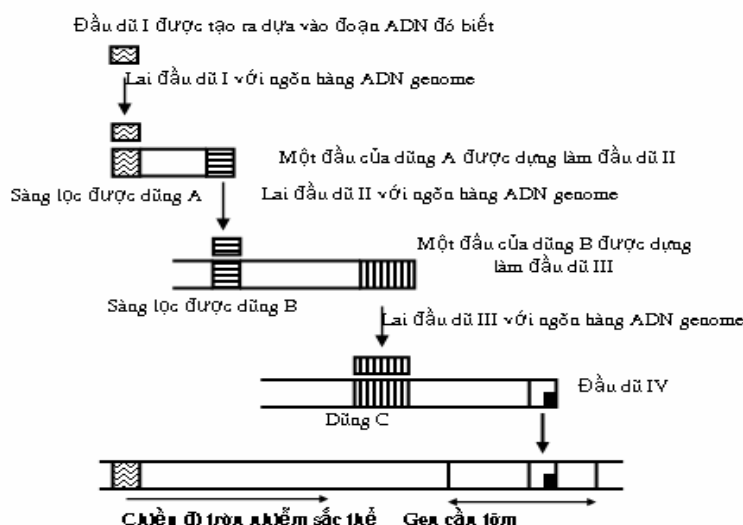
Hình 3.10:

Phương pháp sàng lọc phân biệt. Lần lượt đặt hai màng lên trên mặt mỗi đĩa petri mang các quần lạc của ngân hàng ADNc sao cho vị trí của các quần lạc trên hai màng giống hệt nhau. Sau đó mỗi màng được lai với đầu dò phóng xạ khác nhau. Đầu dò thứ nhất bao gồm các ADNc tương ứng với ARNm của loại mô A. Đầu dò thứ hai gồm các ADNc ứng với ARNm của loại mô B. So sánh sự có mặt hoặc độ đậm nhạt của các vết lai (trên phim tự ghi phóng xạ hoặc màu phát ra) có thể xác định được các ARNm chỉ xuất hiện trong một hoặc trong cả hai loại mô. Nhờ đó, phân lập được các dòng ADNc đặc hiệu cần nghiên cứu.

3.4.3 Phương pháp đi dọc nhiễm sắc thể “chromosome walking”

Dựa vào kết quả phân tích các đột biến di truyền hoặc các phép lai nucleotide, vị trí một gen có thể được xác định trên nhiễm sắc thể (bản đồ nhiễm sắc thể), mặc dù sản phẩm của gen này chưa biết. Vấn đề đặt ra là cần phân lập và xác định được chức năng của gen mã đó.

Một khi đã biết bản đồ nhiễm sắc thể và một đoạn ADN nằm cạnh gen cần nghiên cứu, gen đó có thể phân lập được từ ngân hàng ADN genome nhờ phương pháp "đi dọc nhiễm sắc thể" (Hình 3.11). Phương pháp này dựa vào quá trình chọn lọc từ ngân hàng ADN genome các dòng chứa các đoạn ADN nằm gối lên nhau; phân tích dọc theo chúng để đến được gen cần nghiên cứu. Mặc dù trình tự cũng như sản phẩm của gen chưa biết nhưng dựa vào các đoạn ADN đã biết phân bố trên cùng nhiễm sắc thể và không xa với gen cần nghiên cứu, chúng ta có thể tách được gen này từ ngân hàng ADN genome.



Hình 3.11:

Phương pháp đi dọc nhiễm sắc thể được sử dụng để chọn lọc các dòng ADN nằm gộp lên nhau theo một trật tự nhất định. Các dòng này bao phủ một vùng nhiễm sắc thể chứa gen cần nghiên cứu. Để bước nhanh, ngân hàng genome cần chứa các dòng có kích thước lớn. Chỉ một đoạn ngắn ADN ở đầu các dòng được dùng làm đầu dò. Các đầu dò này không giống nhau và không chứa ADN lặp lại.

Đầu tiên, đoạn ADN đã biết được dùng làm đầu dò để sàng lọc ngân hàng ADN genome (Hình 3.11). Một đầu của đoạn ADN lai với đầu dò được cắt ra và sử dụng làm đầu dò thứ hai để tiếp tục sàng lọc đoạn ADN khác lai với nó. Kết quả thu được đoạn ADN thứ hai nằm gộp lên đoạn thứ nhất. Đoạn thứ hai ADN được tinh sạch và một đầu của nó (nằm về phía có gen cần tìm) lại được cắt ra và dùng làm đầu dò tiếp theo để sàng lọc ngân hàng chọn tiếp đoạn ADN thứ ba nằm gộp lên đoạn thứ hai. Quá trình được lặp đi lặp lại nhiều lần cho đến khi thu được đoạn ADN chứa gen cần tìm. Với phương pháp này, chúng ta có thể đi dọc về hai phía trên nhiễm sắc thể xuất phát từ đoạn ADN đã biết trước từ ban đầu (đoạn ADN dùng làm đầu dò thứ nhất).

Tuy nhiên khi áp dụng phương pháp này, cần lưu ý đến sự tồn tại của các trình tự nucleotide lặp đi lặp lại trong genome. Nếu các đầu dò có chứa đoạn lặp lại thì không thể tiến được trên nhiễm sắc thể vì đầu dò sẽ lai với nhiều đoạn ADN nằm rải rác trong genome. Các đoạn này đều có chứa trình tự lặp lại tương đồng với đầu dò. Do đó kết quả sàng lọc thu được các đoạn ADN không nằm gộp lên nhau. Vì vậy, đầu dò sử dụng để đi từ đoạn này sang đoạn khác nhất thiết phải là duy nhất. Ngày nay kỹ thuật này được thay thế bằng các phương pháp khác nhanh và ít công sức hơn như PCR hoặc “positional cloning” - tạm dịch là tách dòng theo vị trí (dựa vào vị trí trên bản đồ).

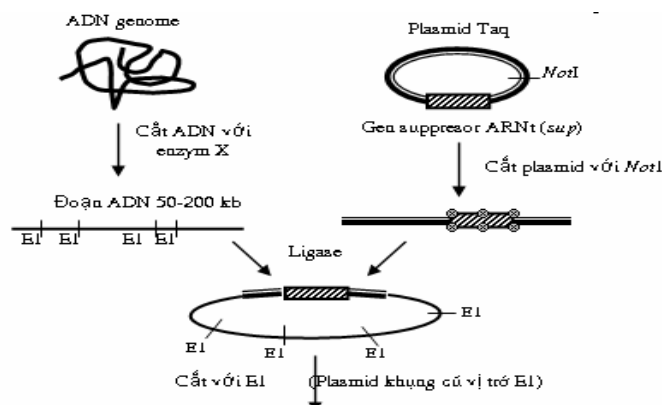
Khi sử dụng kỹ thuật PCR để tìm ra các đoạn ADN nằm gộp lên nhau, các cặp mồi được thiết kế dựa vào trình tự nucleotide của đoạn ADN đã biết ban đầu. Phản ứng PCR được thực hiện độc lập với các vector của ngân hàng. Vector nào cho sản phẩm PCR sẽ chứa đoạn ADN nằm gộp lên đoạn ADN đã biết. Đoạn ADN mới được phân lập và đầu cuối của nó lại được xác định trình tự để dựa vào đó thiết kế các cặp mồi tiếp theo.

Mặc dù kết hợp với phản ứng PCR, trong thực tế tiến độ đi dọc trên nhiễm sắc thể thực hiện rất chậm và mất nhiều công sức để sắp xếp 15-20 đoạn ADN vào một *contig* (gồm các đoạn ADN nằm gộp lên nhau - Hình 3.8). Phương pháp tách dòng theo vị trí đã khắc phục phần nào những yếu điểm trên. Nguyên tắc của phương pháp này là xuất phát từ một vị trí đã được xác định trên bản đồ nhiễm sắc thể. Vị trí đó chỉ cách xa gen cần phân lập một vài Mb. Như vậy

tách dòng theo vị trí cần phải kết hợp các loại bản đồ với nhau như bản đồ vị trí enzym giới hạn, bản đồ chỉ thị ADN microsatellite vv... để tìm ra vị trí làm mốc cũng như xác định chính xác khoảng cách đến gen cần tìm.

3.4.4 Nhảy bước trên nhiễm sắc thể “jumping on chromosome”

Phương pháp đi dọc nhiễm sắc thể không thích hợp với những đoạn nhiễm sắc thể dài hơn 1000 kb. Để khắc phục khó khăn đó, hai đầu của một đoạn ADN genome có kích thước rất lớn được nối với một đoạn ADN đã biết tạo thành phân tử dạng vòng khép kín (Hình 3.12).



Hình 3.12:

Phương pháp nhảy bước trên nhiễm sắc thể. ADN genome được cắt bằng enzym có ít vị trí nhận biết trong genome. Do đó thu được những đoạn ADN rất lớn. Phản ứng ligase xảy ra trong điều kiện có lợi cho việc tạo phân tử dạng vòng (nồng độ vector >> nồng độ ADN). Phân tử ADN tái tổ hợp mang hai đoạn ADN nằm ở hai đầu của đoạn ADN lớn. Hai đoạn này rất xa nhau trong genome.

Tiếp theo phân tử này bị cắt bởi enzym giới hạn mà đoạn ADN đã biết không chứa vị trí nhận biết của enzym này. Như vậy, hai đầu của đoạn ADN genome và đoạn ADN đã biết cùng nằm trên một phân tử ADN mới. Phân tử này được đưa vào vector và hai đầu của đoạn ADN genome lúc này nằm gần nhau trong cùng một vector (với kích thước thích hợp do việc trộn enzym giới hạn) mặc dù trong thực tế chúng rất xa nhau.

Khi hai đầu của một đoạn ADN rất lớn nằm trong cùng một vector, nếu đã biết một đầu này thì dễ dàng tìm được đầu kia mà không cần đi dọc nhiễm sắc thể (không cần tìm các đoạn nằm gối lên nhau, bao phủ lên cả vùng ADN lớn). Ngân hàng chứa các đầu của những đoạn ADN lớn cắt ra từ genome cho phép áp dụng kỹ thuật nhảy bước, tiến xa và nhanh trên nhiễm sắc thể.

3.5 Các phương pháp lai

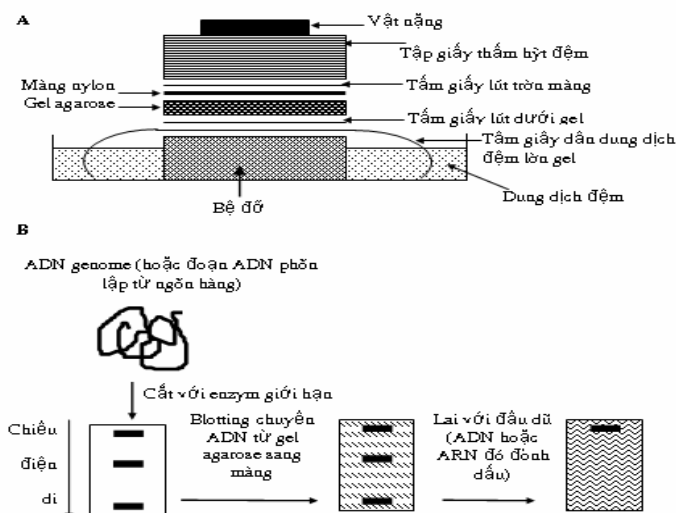
Nguyên tắc của phương pháp lai acid nucleic là dựa vào phản ứng tạo ra phân tử lai nhờ liên kết bổ sung giữa hai sợi đơn ADN hoặc hai sợi đơn ARN hay ADN/ARN. Điều kiện của phản ứng lai phụ thuộc chủ yếu vào nhiệt độ và nồng độ muối trong đệm. Thông thường, sợi đơn làm đầu dò được đánh dấu bằng phóng xạ (P^{32}) hoặc bằng các chất phát huỳnh quang. Do đó, sợi đích lai với đầu dò được phát hiện nhờ phim có độ nhạy với phóng xạ hoặc các kỹ thuật nhuộm màu. 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.

3.5.1 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 nitrocellulose). Bước chuyển này được gọi là kỹ thuật blotting. Blotting có thể thực hiện bằng cách thấm thấu mao dẫn (đòi hỏi nhiều thời gian-thường tiến hành qua đêm) (Hình 3.13A), 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 lai với đầu dò đặc hiệu đã đánh dấu, thường bằng phóng xạ P^{32} (hoặc các chất phát màu). Vị trí của đoạn ADN lai với đầu dò được xác định nhờ phim hoặc hoá chất đặc biệt (Hình 3.13B). Phương pháp này rất nhạy cho phép phát hiện được đoạn ADN xuất hiện một lần (đơn bản) trong genome ($\sim 1/10^6$) khi chỉ dùng 5 μg ADN genome.



Hình 3.13:

(A)-Blotting chuyển ADN từ gel agarose sang màng bằng phương pháp thấm 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 blot sử dụng để phát hiện đoạn ADN mong muốn.

Lai Southern được áp dụng để phát hiện số lượng bản sao của một gen trong genome; phát hiện thay đổi trật tự sắp xếp của các đoạn ADN, sự có mặt của các đoạn ADN lạ, sự thay

đổi vị trí của các transposon; phân biệt một gen trong họ gen, xác định các intron, exon của một gen vv.... 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.

3.5.2 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.

Trong những năm đầu thế kỷ 21, kỹ thuật lai northern được sử dụng để phát hiện các phân tử ARNs kích thước nhỏ (vài chục nucleotide). Những phân tử này đóng vai trò quan trọng trong quá trình hình thành cấu trúc dị nhiễm sắc (*heterochromatin formation*), trong kiểm soát bất hoạt gen (*gene silencing*). Tuy nhiên, do kích thước rất ngắn, nên các ARNs phải được tách khỏi các loại ARN kích thước lớn. Sau đó, ARNs được phân ly trên gel polyacrylamide và được chuyển sang màng nhờ máy. Đặc biệt đầu dò sau khi đánh dấu cũng phải được phân cắt thành các đoạn ngắn. Mọi bước tiếp theo tương tự như với lai Southern.

3.5.3 Kỹ thuật lai in-situ

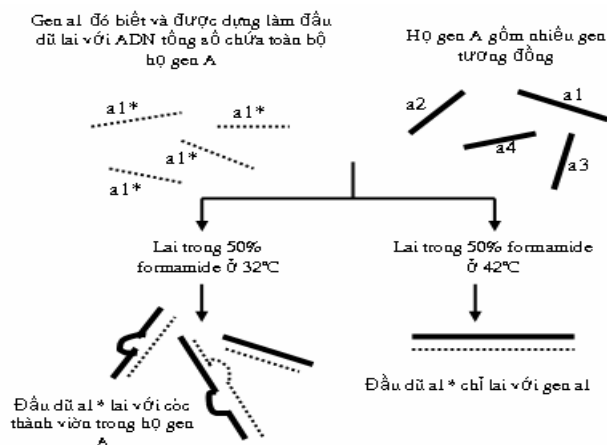
Phát triển từ phép lai Southern blots, kỹ thuật lai *in situ* sử dụng các đầu dò đánh dấu (đoạn nucleotide hoặc kháng thể) để lai với ADN, ARN hoặc với protein ở ngay trong tế bào mà không cần tách chiết chúng. Các đầu dò oligonucleotide thường được đánh dấu bằng phương pháp hoá học (chất phát huỳnh quang) thay cho việc đánh dấu bằng phóng xạ. 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 môi. 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.

3.5.4 Điều kiện 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 lai giữa các sợi đơn ADN/ADN, ADN/ARN và ARN/ARN tạo thành sợi kép có thể thay đổi từ mức độ trình tự nucleotide khá tương đồng với nhau đến yêu cầu chúng phải hoàn toàn tuyệt đối tạo cặp bổ

sung. Do đó ở điều kiện nghiêm ngặt của phản ứng, có thể xác định được một gen đột biến trong số các thành viên của một họ gen. Khi điều kiện phản ứng lại cho phép lai giữa các sợi tương đồng thì có thể phát hiện các gen trong một họ gen nếu một thành viên trong một họ gen đã biết và được dùng làm đầu dò (Hình 3.14).



Hình 3.15:

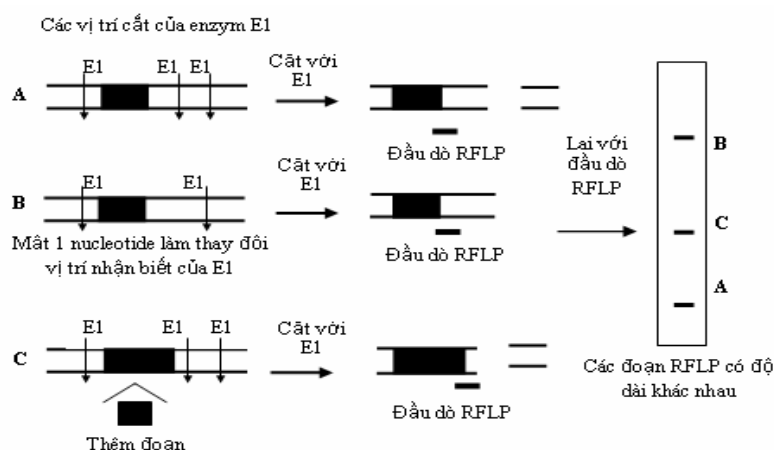
Thay đổi điều kiện lai trong phản ứng Southern sẽ thu được các kết quả khác nhau. Trong điều kiện khắc nghiệt, đầu dò chỉ lai với đúng đoạn tạo cặp bổ sung chính xác. Trong điều kiện dễ dàng hơn, đầu dò có thể lai với các đoạn tương đồng.

Hơn 3000 bệnh di truyền ở người do đột biến xảy ra ở những gen chỉ tồn tại một bản sao duy nhất trong genome (đơn bội). Những đột biến này thường ở trạng thái lặn và chỉ biểu hiện khi cả hai allele của gen đều bị đột biến. 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 allele 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 khi ADN tách ra từ tế bào của bào thai (*amniocentesis*) lai với đầu dò là hai oligo trong phép lai Southern, dạng bình thường và dạng đột biến của gen mã cho β globin có thể phân biệt được. Thai nhi mang hai allele 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 Southern thực hiện trong điều kiện kém chặt chẽ (đầu dò chính là thành viên đã biết). Lúc đó các sợi đơn ADN không hoàn toàn bổ sung với đầu dò vẫn lai được với nó (Hình 3.14). 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.

3.6 RFLP trong nghiên cứu genome và lập bản đồ gen

Bản đồ genome có thể được xác định theo hai cách khác nhau: bản đồ vật lý và bản đồ liên kết di truyền (*physical map* và *genetic linkage map*). Đơn vị đo khoảng cách (giữa các gen) trên bản đồ vật lý được tính bằng số lượng các nucleotide. Ngược lại, bản đồ liên kết di truyền được xây dựng dựa vào cách thức di truyền các tính trạng. Bản đồ liên kết di truyền dựa vào tần số cùng di truyền của hai hay nhiều tính trạng trong các phép lai. Các tính trạng này được sử dụng như các chỉ thị (*marker*) di truyền. Mỗi một marker là một vị trí trên bản đồ. Đơn vị đo khoảng cách giữa 2 vị trí trên bản đồ di truyền là centiMorgan (tính theo tần số trao đổi chéo giữa 2 vị trí đó). Khi có sự thay đổi trên bản đồ genome, nếu sự thay đổi xuất hiện rất ít và đột ngột thì xem như xảy ra đột biến. Khi thay đổi xảy ra phổ biến theo tiến hoá thì được gọi là đa dạng (*polymorphism*).



Hình 3.16:

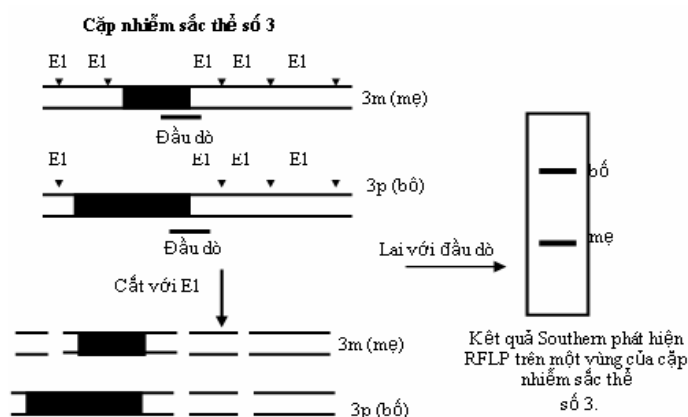
Tính đa dạng các đoạn ADN bị cắt bởi enzym giới hạn

(A) Hầu hết các cá thể trong quần thể có 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 sử dụng đầu dò RFLP

(B) Khi có sự thay đổi của nucleotide tại một vị trí cắt, enzym không nhận biết được vị trí đó nữa, nên chỉ cắt ở hai vị trí kia tạo một đoạn ADN dài (lúc này được xem như một RFLP)

(C) Khi xảy ra đột biến thêm đoạn, đoạn ADN bị cắt dài hơn. Loại RFLP này thường phổ biến ở những vùng nhiễm sắc thể chứa các đoạn ngắn nucleotide lặp đi lặp lại như GTGTGT... Số lần lặp lại thay đổi rất nhiều trong quần thể. Tại một vùng nhất định, số lần lặp lại đặc trưng cho từng cá thể. Đó là các minisatellite, microsatellite hoặc còn gọi là các VNTR (*variable number of tandem repeat*) (theo Alberts & cs., 2002).

Thông thường, bản đồ liên kết di truyền được xác định nhờ biểu hiện của các đột biến như màu mắt, nhóm máu. Gần đây, các phương pháp tái tổ hợp ADN sử dụng các marker di truyền mới - các đoạn ADN có độ dài khác nhau giữa các cá thể bình thường, để lập bản đồ di truyền. Đặc điểm của các đoạn ADN này là sự sai khác giữa chúng không hề có biểu hiện tính trạng. Tuy nhiên, thay đổi nhỏ giữa chúng 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 (Hình 3.15). Tương tự, sự thêm bớt các nucleotide vào một vùng ADN bất kỳ có thể không gây ra thay đổi tính trạng nhưng làm thay đổi kích thước của các đoạn ADN bị cắt ra từ vùng đó. Sự sai khác này được gọi là tính đa dạng của các đoạn ADN bị cắt bởi enzym giới hạn, viết tắt là RFLP (Restriction Fragment Length Polymorphism).



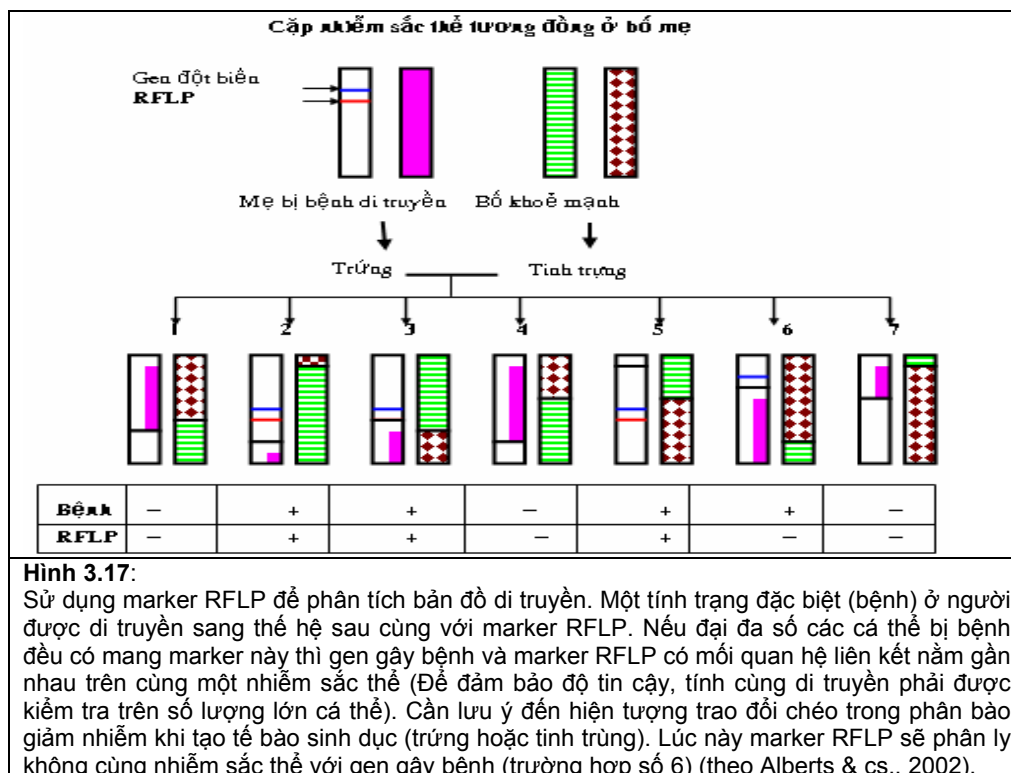
Hình 3.16:

Kỹ thuật Southern blots được sử dụng để phát hiện một RFLP. Sự khác nhau của một vùng trên nhiễm sắc thể số 3 di truyền từ cha và mẹ có thể phân biệt dựa vào kết quả của phép lai. RFLP được sử dụng rất hữu hiệu trong phân tích quan hệ huyết thống.

Khi một RFLP rất phổ biến trong quần thể, sự khác nhau giữa hai cá thể có thể xác định được nhờ RFLP đó. Các RFLP thường chứa các đoạn ADN ngắn, lặp đi lặp lại, có chiều dài khác nhau do số lần lặp lại thay đổi (từ 4 đến 40 lần). Sự khác nhau về kích thước của các đoạn ADN mà một cá thể được thừa hưởng từ cha và mẹ có thể phát hiện dễ dàng nhờ phép lai Southern. Đầu dò lai phải tạo liên kết bổ xung với một chuỗi nucleotide nằm trong các đoạn ADN đó (Hình 3.16).

Khi hai marker di truyền nằm trên hai nhiễm sắc thể khác nhau, xác suất để một cá thể nhận được hai marker này chỉ chiếm 50% (khi hai nhiễm sắc thể cùng đi về một cực trong phân bào giảm nhiễm). Ngay với hai marker nằm ở hai phía của một nhiễm sắc thể (đối diện nhau qua tâm động), xác suất phân ly của chúng khá cao do xảy ra trao đổi chéo trong phân bào giảm nhiễm. Các marker càng gần nhau trên một nhiễm sắc thể, xác suất để chúng cùng được truyền cho một cá thể con cháu càng lớn.

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, nhờ đó các protein tương ứng được phân tích rõ ràng (Hình 3.17).



Trên hình 3.17 mô tả một marker RFLP nằm gần một gen gây bệnh. Khi marker nằm càng gần gen thì càng dễ phát hiện được gen đó. Để phát hiện một gen bất kỳ dựa vào RFLP thì phải biết được vị trí của các marker RFLP trên nhiễm sắc thể chứa gen đó. Vị trí của tất cả các marker RFLP (số lượng đến hàng nghìn) tạo thành bản đồ RFLP của genome. Trên bản đồ này, khoảng cách giữa hai marker khoảng 10^6 bp (tương đương với $\sim 0,1$ cM). Ở khoảng cách này chúng cùng di truyền cho con cháu với xác suất 99%.

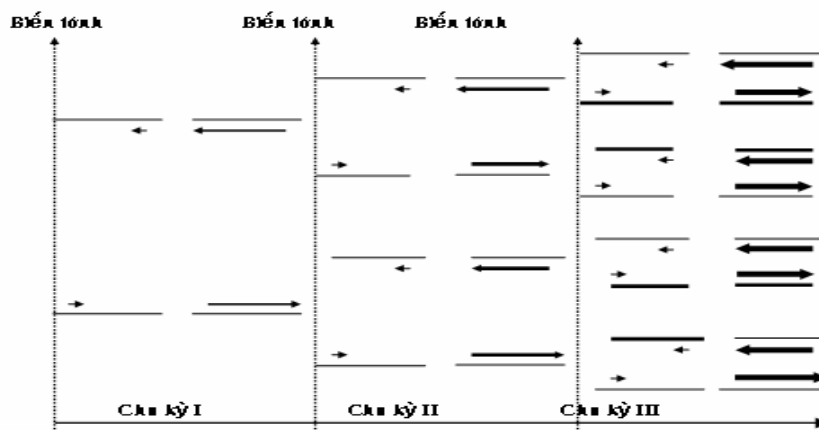
Dựa vào bản đồ RFLP, khi một tính trạng cùng được di truyền với một marker RFLP thì vị trí của gen qui định tính trạng đó có thể xác định được dễ dàng. Nhờ đó, việc sàng lọc các clone ADN tương ứng từ ngân hàng genome và phân lập được gen liên quan đến tính trạng trở nên thực thi.

3.7 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 môi-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 môi 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} μ g).

Một phản ứng PCR gồm nhiều chu kỳ (Hình 3.18). Mỗi chu kỳ phản ứng đòi hỏi ba bước. Đầu tiên, ADN khuôn được xử lý nhiệt độ để tách 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 phức 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 60 và mỗi chu kỳ thường kéo dài khoảng 5 phút.



Hình 3.18:

Phản ứng PCR. Ba chu kỳ đầu tiên của phản ứng được mô tả trên hình vẽ. Từ hai sợi đơn ADN biến tính, sau 3 chu kỳ thu được 16 sợi đơn, trong đó 8 sợi có kích thước đúng bằng khoảng cách giữa hai mồi oligo. 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 sợi có chiều dài không xác định

x: Số sợi khuôn ban đầu ; **N:** Số sợi tổng hợp được trong phản ứng.

3.7.1 Các yếu tố ảnh hưởng đến phản ứng PCR

* Các mồi oligonucleotide: Tỷ lệ GC ở các mồi thường chiếm khoảng 40-75%. Khoảng cách giữa hai mồi thường không dài quá 3 kb. Các mồi không được tạo cấu trúc bậc hai do liên kết giữa các nucleotide ngay trong một mồi. Độ dài của mồi (số nucleotide) từ 15 đến không quá 30. Những mồi dài hoặc ngắn hơn ảnh hưởng đến khả năng bắt cặp chính xác với ADN khuôn hoặc nhân bản không đặc hiệu. 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 0,1 μ M đến 1 μ M. Nồng độ thường được xác định dựa vào giá trị OD đo ở 260 nm.

* 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 ủ mồi với sợi khuôn không cần thiết quá lâu. Nhiệt độ ở giai đoạn này phụ thuộc chiều dài của mồi và hàm lượng GC. Khi oligo mồi gồm 12 đến 15 base, nhiệt độ ủ khoảng 40-45°C.

Nhiệt độ 72°C thích hợp cho phản ứng tổng hợp ADN bằng Taq polymerase (là ADN polymerase có hoạt tính ở nhiệt độ cao). Lúc này cần lưu ý các mồi ngắn thường bị biến tính

tách ra khỏi sợi khuôn. Ngoài ra, trong trường hợp mỗi được tổng hợp xuất phát từ trình tự acid amin thì liên kết giữa mỗi với ADN khuôn không hoàn toàn bổ sung do tính chất thoái hoá của mã di truyền. Do đó, mỗi 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 mỗi có thể kéo dài hơn so với mỗi có trình tự chính lấy từ đoạn cần nhân bản vì Taq polymerase có thể tổng hợp ADN ở nhiệt độ gắn mỗi. Sau đó mới tăng nhiệt độ lên 72°C. Rõ ràng nhiệt độ và thời gian ở giai đoạn gắn mỗi đặc biệt quan trọng; nhiệt độ quá cao gây biến tính, nhiệt độ quá thấp tạo ra các liên kết không đặc hiệu (mỗi 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 mỗi 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 mỗi đã liên kết với khuôn thì Taq polymerase sẽ kéo dài sợi mỗi thành sợi ADN mới bắt chấp mỗi chưa liên kết hoàn toàn với khuôn ở đầu 5'.

* Nồng độ một số ion cũng như bốn loại nucleotide dNTPs trong môi trường đều ảnh hưởng đến phản ứng PCR. Ví dụ, ion Mg^{+2} kích thích phản ứng tổng hợp ADN trong khi muốn nhân được đoạn ADN dài cần loại bỏ hoàn toàn KCl khỏi môi trường. ADN khuôn 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.

3.7.2 Một số dạng của phản ứng PCR

* **PCR lồng (Nested-PCR)**: Sản phẩm ADN của phản ứng PCR thứ nhất được sử dụng làm khuôn cho phản ứng PCR thứ hai với khoảng cách (tính bằng bp) giữa hai mỗi lần sau ngắn hơn lần trước. Cách tiến hành như vậy nhằm làm tăng tính đặc hiệu, chính xác cho sản phẩm cần nhân bản. Các đoạn ADN không đặc hiệu có thể được tạo ra trong lần PCR thứ nhất nhưng rất khó trở thành khuôn trong lần PCR thứ hai do các sản phẩm đặc hiệu chiếm ưu thế tuyệt đối.

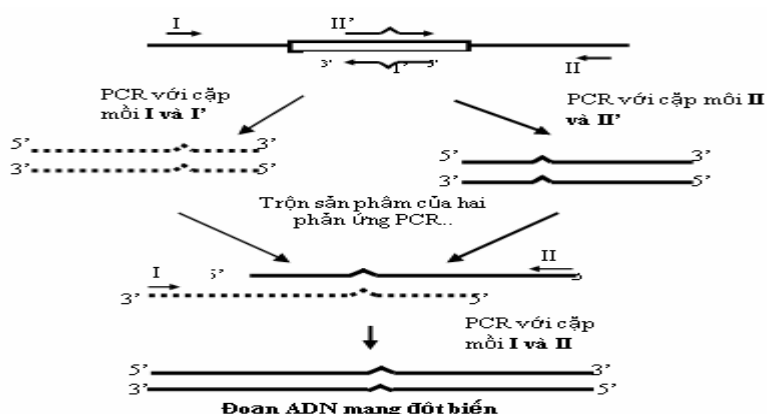
* **PCR phức (Multiplex PCR)**: Phương pháp này sử dụng nhiều cặp mỗi trong một phản ứng PCR để thu được cùng một lúc các băng đặc hiệu cho từng cặp. Nhờ đó giảm được thời gian thí nghiệm cũng như hạn chế việc nhiễm mẫu do thao tác. Ví dụ, bệnh teo cơ Duchenne (DMD) gây ra do thiếu hụt các đoạn ADN ở gen tương ứng. Gen này dài khoảng 2000 kb gồm 70 exon (kích thước trung bình của intron khoảng 35 kb). Sự mất đoạn chủ yếu xảy ra ở hai vị trí trên gen (các vị trí nhạy cảm đột biến). Sử dụng các cặp mỗi đặc hiệu của từng exon trong cùng một phản ứng PCR phức, có thể phát hiện được 80-90% bệnh nhân teo cơ do mất đoạn.

* **PCR đảo (Inverse-PCR)**: Dùng phương pháp này để phân lập những phần nằm trước hoặc sau một đoạn ADN đã biết. Từ đó phân tích được chuỗi nucleotide nằm trước hoặc sau của một gen. Đầu tiên ADN genome được cắt bởi enzym giới hạn tạo ra các đoạn ADN khác nhau. Sau đó, các đoạn ADN được nối lại thành dạng vòng và sử dụng như ADN khuôn cho phản ứng PCR với hai oligo xuất phát từ phần ADN đã biết. Phương pháp này tương tự như phương pháp đi trên nhiễm sắc thể. Nó được ứng dụng để xác định promoter hoặc các đoạn nucleotide làm nhiệm vụ điều khiển hoạt động của gen.

* **PCR ngược (Reversed-PCR)**: Phản ứng này nhằm nghiên cứu số lượng ARNm, làm giàu các ARNm hiếm trong tế bào. Đầu tiên ARN tổng số hoặc ARNm được dùng làm khuôn để tổng hợp ADNc nhờ enzym reverse transcriptase. Sau đó các sợi đơn ADNc được tổng hợp thành sợi kép nhờ Taq polymerase. ADNc dạng sợi kép được sử dụng làm khuôn trong phản

ứng PCR. Các oligo trong phản ứng này thường bắt nguồn từ các exon để dễ dàng phân biệt các đoạn ADN chứa intron tạo ra do nhiễm ADN genome với các đoạn nhân lên từ ARNm (ADNc). Chỉ cần 1 μg ARN tổng số ($\sim 10^6$ tế bào) đủ để nhân bản phân tử ARNm hiếm (1-10 copies/tế bào).

* **PCR tái tổ hợp**: 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. Kỹ thuật này được mô tả cụ thể trong hình 3.19.



Hình 3.19:

Kết hợp các phản ứng PCR để thay thế các nucleotide trong một đoạn ADN đã biết. Mỗi thiết kế mang các nucleotide thay thế. Kỹ thuật này cho phép gây đột biến tại vị trí mong muốn trên một đoạn ADN (gen).

* **Nhân ngẫu nhiên các đoạn ADN bằng PCR (RAPD-PCR) (Random Amplified Polymorphic DNA-PCR)**: Dựa vào các minisatellite hoặc các microsatellite (VNTRs) đã biết để tìm được các đoạn nucleotide nằm trước và sau các VNTR này. Chúng được dùng làm môi trong phản ứng PCR. Sản phẩm thu được là những băng ADN có kích thước khác nhau. Phản ứng có thể dùng một môi riêng biệt. Phương pháp này hay được sử dụng để xây dựng cây phân loài.

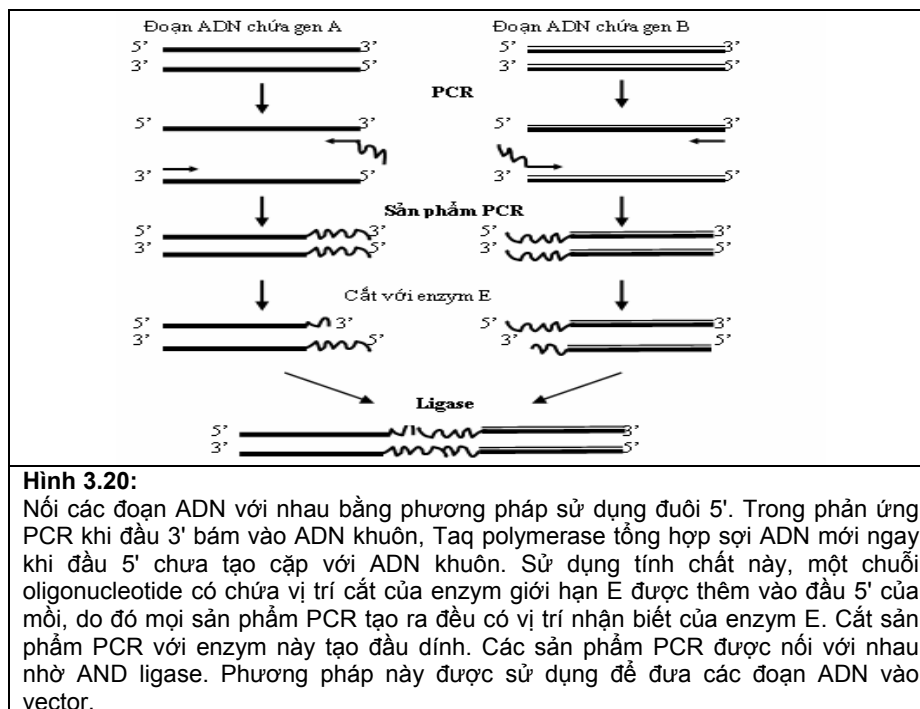
3.8 Kỹ thuật gen

Chúng ta đã cùng nhau thiết kế một phân tử ADN tái tổ hợp. Đơn giản nhất khi phân tử đó bao gồm vector và một đoạn ADN lạ. Trên thực tế, chúng ta có thể ghép hai hoặc nhiều đoạn ADN có nguồn gốc khác nhau (từ các loài khác nhau) hoặc xuất phát từ các gen khác nhau của cùng một cơ thể rồi đưa chúng vào cùng một vector. Bằng cách đó chúng ta có thể thu được phân tử ADN có chuỗi nucleotide theo ý muốn mà không nhất thiết phải tổng hợp chúng bằng con đường nhân tạo (Hình 3. 20). Phân tử ADN tái tổ hợp giữ vai trò đặc biệt quan trọng trong việc sản xuất protein dung hợp (*fusion protein*), biến đổi cấu trúc protein thông qua việc thay thế trình tự gen, trong nghiên cứu các domain, các vùng có chức năng khác nhau trên phân tử protein vv...

3.8.1 Nghiên cứu vai trò của ADN điều khiển, chức năng của gen hoặc protein

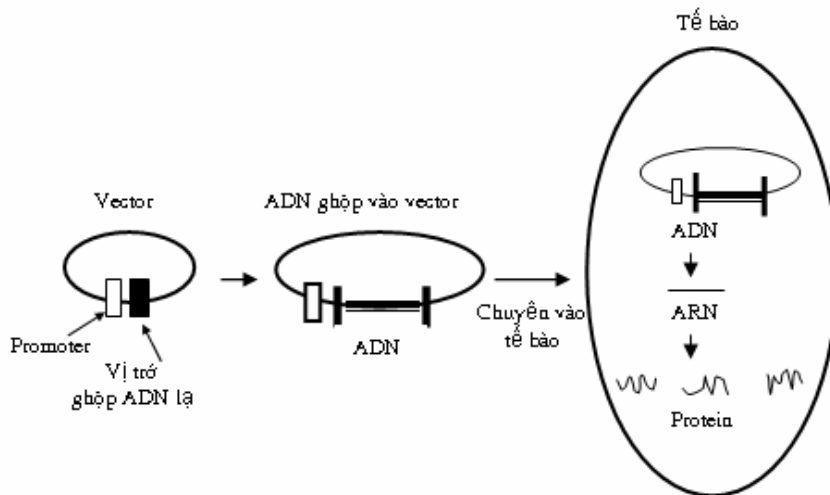
Phiên mã từ một gen được kiểm soát bởi vùng ADN điều khiển (*promoter, enhancer*) và phụ thuộc vào các tín hiệu trong từng loại tế bào ở các điều kiện khác nhau. Các đoạn ADN tăng cường (*enhancer*) có thể phân bố ngay trong gen (vùng intron) hoặc ở phía trước hoặc sau gen hàng nghìn nucleotide và cùng nhau phối hợp hoạt động. Với một gen bất kỳ,

các đoạn ADN điều khiển có thể xác định được khi ghép đoạn ADN nằm trước hoặc sau gen đó với một gen đã biết mã cho một protein đặc hiệu. Gen đã biết được gọi là gen báo cáo (*reporter gene*). Sản phẩm của gen báo cáo được gọi là protein báo cáo (*reporter protein*). Gen báo cáo hay được dùng mã cho sản phẩm là enzym hay chất phát màu. Do đó protein báo cáo dễ dàng phát hiện nhờ phản ứng enzym hoặc nhuộm tế bào. Khi đưa phân tử ADN tái tổ hợp xây dựng theo cách thức này vào tế bào, gen báo cáo có thể ghép vào vị trí bất kỳ. Nếu vị trí đó nằm sau một promoter thì chúng ta có thể xác định được thời gian, mức độ hoạt động cũng như các cơ chế kiểm soát của promoter thông qua hoạt động của gen báo cáo, tức là thông qua thời gian và mức độ biểu hiện của protein báo cáo.



Khi các phân tử ARNm của một gen đã biết được tổng hợp rất ít trong tế bào, việc tách và tinh sạch chúng rất khó khăn. Để khắc phục nhược điểm đó, phân tử ADN tái tổ hợp được xây dựng gồm gen đã biết nối với promoter virus. Các promoter này thường hoạt động mạnh. Promoter này được nhận biết bởi ARN polymerase của virus. Phân tử ADN tái tổ hợp được ủ với enzym và 4 loại ribonucleoside triphosphate để tổng hợp *in vitro* các phân tử ARNm với số lượng mong muốn. Những promoter đã biết được dùng để nghiên cứu chức năng của gen ghép với nó được gọi chung là promoter báo cáo (*reporter promoter*).

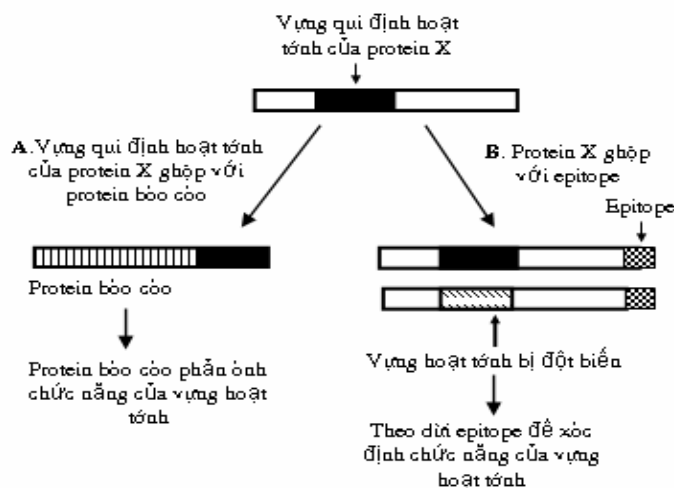
Để nghiên cứu chức năng, cấu trúc protein tương ứng với một gen đã biết, phương pháp dùng vector biểu hiện (*expression vector*) hay còn gọi là phương pháp biểu hiện gen được áp dụng để thu được số lượng lớn ARNm và protein trong tế bào sống (Hình 3.21). Vector này mang promoter hoạt động mạnh nằm gần vị trí ghép gen. Khi đưa vector vào tế bào sống (tế bào vi khuẩn, virus, nấm men, tế bào động, thực vật), ARNm tương ứng với gen trong vector cũng như protein được tổng hợp với số lượng lớn. Để tránh hiện tượng protein lạ gây ảnh hưởng tới sự phát triển của tế bào, trong thực tế thường sử dụng các tế bào nhạy cảm nhiệt độ. Phương pháp này rất có hiệu quả trong việc sản xuất vaccine hoặc bất kỳ protein nào với số lượng lớn.



Hình 3.21:

Số lượng lớn protein được tạo ra nhờ tách dòng (*cloning*) đoạn ADN chứa mã di truyền vào vector hoạt động. Vector này mang promoter hoạt động mạnh, do đó số lượng ARNm và protein tương ứng được tổng hợp nhiều.

Protein thường có một đoạn peptide phân bố ở phần đầu (NH_2 hoặc COOH) làm nhiệm vụ cố định protein ở vị trí thích hợp trong tế bào hoặc đảm bảo độ bền vững của nó sau khi tổng hợp. Để nghiên cứu vai trò của đoạn này, nó thường được nối với protein báo cáo đã bị cắt bỏ phần tương ứng. Đây là phân tử protein hợp nhất hay còn gọi là protein dung hợp. Kỹ thuật tái tổ hợp ADN cho phép xây dựng phân tử ADN gồm những đoạn nucleotide khác nhau của một gen ghép với gen báo cáo. Từ đó sản phẩm protein dung hợp cho phép xác định chức năng của các đoạn quan trọng trong chuỗi polypeptide có liên quan đến vận chuyển, cố định hoặc liên quan đến hoạt tính của protein.



Hình 3.22:

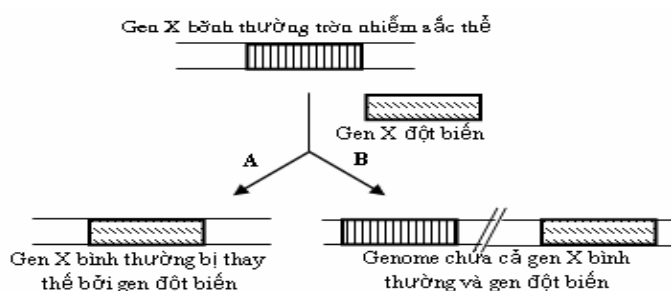
Hai kỹ thuật khác nhau cho phép phân tích cấu trúc và chức năng của protein X
(A) Theo dõi chức năng mới xuất hiện do ghép protein báo cáo với một đoạn của protein X cần nghiên cứu
(B) Những thay đổi do đột biến có thể so sánh được giữa protein bình thường và protein đột biến nhờ quan sát "epitope" gắn vào chúng.

Không phải mọi phần quan trọng của một protein đều có thể ghép được với protein báo cáo. Ví dụ, để vận chuyển protein từ Golgi đến lisosome, các đoạn peptide liên quan đến tín hiệu vận chuyển của protein này nằm xa nhau trên phân tử protein và phải được gấp lại theo

một cấu trúc đặc biệt khiến chúng trở nên gần nhau. Để nghiên cứu vai trò của các đoạn này, kỹ thuật đích epitope "epitope tagging" được sử dụng. Toàn bộ protein cần nghiên cứu được ghép với một đoạn peptide ngắn 8-12 acid amin. Đoạn này phát hiện dễ dàng nhờ kháng thể. Vị trí, chức năng cũng như sự di chuyển trong tế bào của protein lại được theo dõi nhờ kháng thể, do đó mọi ảnh hưởng do thay thế bất kỳ acid amin nào trong protein đều có thể xác định được (Hình 3.22).

3.8.2 Thay thế hoặc gây đột biến gen

Genome trong tế bào vi khuẩn và một số eukaryot bậc thấp thường tồn tại ở dạng đơn bội. Thực nghiệm cho thấy đối với những cơ thể này, có thể thay thế một gen bình thường (đang hoạt động) bằng gen đột biến nhờ hiện tượng trao đổi chéo tương đồng. Hiện tượng này xảy ra với tần số không quá nhỏ (Hình 3.23A).



Hình 3.23:

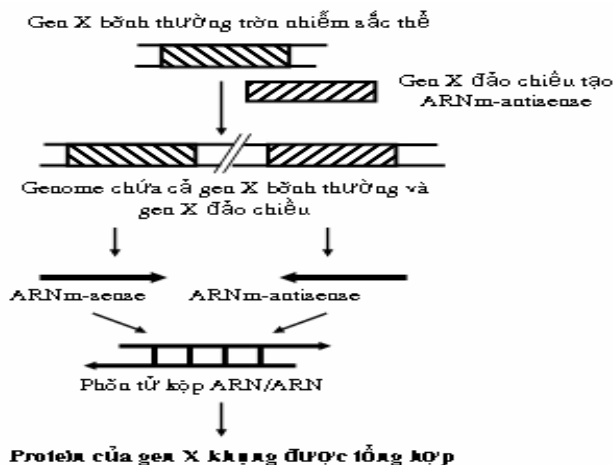
Thay thế hoặc bổ sung gen vào genome

- (A) Đối với vi khuẩn hoặc các cơ thể đơn bào eukaryot bậc thấp, gen đột biến thay thế gen bình thường nhờ trao đổi chéo giữa hai chuỗi nucleotide tương đồng
(B) ở cơ thể eukaryot bậc cao, gen đột biến thường được ghép vào nhiễm sắc thể, do đó genome mang cả gen đột biến và gen bình thường.

Thông thường, các gen đột biến được chọn sao cho sản phẩm của chúng nhạy cảm với nhiệt độ, tức là gen chỉ hoạt động ở một nhiệt độ thích hợp. Khi chuyển sang nhiệt độ cao hơn hoặc thấp hơn, chúng sẽ bị ức chế hoàn toàn. Loại tế bào nhạy cảm với nhiệt độ được sử dụng rất rộng rãi trong nghiên cứu chức năng, vai trò của protein.

Đối với sinh vật eukaryot bậc cao như động vật có vú, genome thường tồn tại ở dạng lưỡng bội (các cặp nhiễm sắc thể tương đồng). Quá trình thay thế gen xảy ra rất hiếm với những lý do chưa được làm sáng tỏ. Tuy nhiên khi đưa một gen đột biến vào tế bào, gen này thường được ghép vào một vị trí bất kỳ trong genome. Do đó bên cạnh gen bình thường, một bản sao của gen cùng tồn tại trong genome nhưng ở dạng đột biến (Hình 3.23B).

Thực nghiệm hay sử dụng phương pháp tạo sợi ARNm ngược nghĩa (*ARNm-antisense*) gây kim hãm hoàn toàn hoạt động của gen bình thường. Chúng ta biết rằng khi phiên mã, sợi kép ADN mở xoắn, một trong hai sợi đơn được dùng làm khuôn để tổng hợp ARNm. Nếu gen đưa vào tế bào được chọn sao cho nó có trình tự nucleotide chính xác như sợi đơn bổ sung với sợi khuôn thì việc sao chép thực hiện trên gen đó sẽ tạo ra những phân tử ARN có khả năng lai với phân tử ARNm của gen bình thường. Do đó quá trình tổng hợp protein bị ức chế hoàn toàn (Hình 3.24).



Hình 3.24:

Bất hoạt gen bằng phân tử ARNm-antisense

Phân tử ARNm-antisense có trình tự nucleotide bổ sung với phân tử ARNm bình thường (ARNm sense)

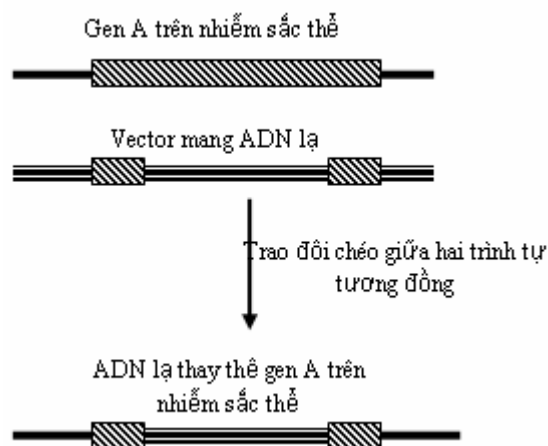
Hai phân tử này liên kết tạo phân tử dạng kép, ngăn cản quá trình tổng hợp protein trên phân tử ARNm sense.

Ngoài ra có thể đưa các phân tử ADN hoặc ARN ngăn được tổng hợp nhân tạo (bằng phương pháp hoá học hoặc enzym) vào tế bào sống để tạo phân tử lai với gen hoặc với ARNm của gen cần nghiên cứu. Do đó, quá trình phiên mã hoặc tổng hợp protein trên những ARNm này bị kìm hãm hoàn toàn. Tuy nhiên trong thực tế, việc lai giữa ARNm-antisense và ARNm-sense không phải luôn luôn thực hiện được dễ dàng. Để khắc phục điều đó, các nghiên cứu thường được tiến hành trên protein.

Trong tế bào, protein thường có hoạt tính khi ở dạng phức chất do liên kết giữa hai hoặc nhiều chuỗi polypeptide. Nếu như protein giữ vai trò quan trọng trong hoạt động sống của tế bào, việc gây mất hoạt tính của nó sẽ dẫn đến tế bào bị chết. Do đó để có thể quan sát được tính trạng thì các gen mang đột biến chỉ hoạt động trong những điều kiện nhất định. Ví dụ, sử dụng các tế bào nhạy cảm nhiệt độ, khi chuyển chúng sang nhiệt độ không thích hợp các gen đột biến mới hoạt động (cùng với gen bình thường) tạo phân tử protein không có hoạt tính. Nhờ đó vai trò, chức năng của protein được xác định.

3.8.3 Gây mất hoặc tăng cường chức năng của gen

Một trong những phương pháp điều biến hoạt động của gen thông thường nhất là áp dụng hiện tượng trao đổi chéo giữa hai trình tự tương đồng để gây bất hoạt. Đây là nguyên tắc của kỹ thuật “loss of function” (tạm dịch là gây mất chức năng). Chúng ta đã biết đầy đủ và chính xác trình tự nucleotide của gen. Do đó, chỉ cần thiết kế đoạn ADN lạ có hai đầu tận cùng là hai chuỗi nucleotide của gen. Đưa đoạn ADN đó vào tế bào. Trao đổi chéo xảy ra giữa hai đầu đoạn ADN lạ và gen có mặt trong genome khiến cho trình tự nguyên vẹn của gen bị gián đoạn bởi ADN lạ. Lúc đó gen bị đột biến nên không thể xảy ra phản ứng tổng hợp protein (Hình 3.25). Thiếu hụt sản phẩm của gen thường dẫn đến sự xuất hiện tính trạng mới.



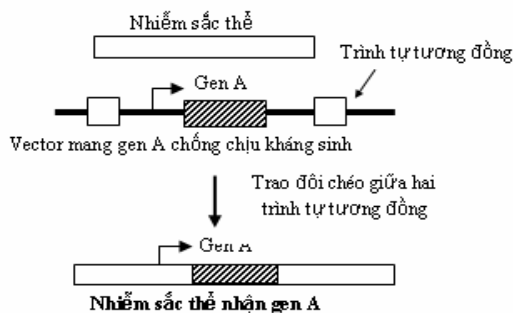
Hình 3.25:

Gây đột biến bất hoạt gen bằng trao đổi chéo tương đồng. Vector chuyên chở đoạn ADN lạ có hai trình tự nucleotide của gen đã biết. Trao đổi chéo xảy ra giữa gen nguyên vẹn trong genome với các trình tự tương đồng trên đoạn ADN lạ. Lúc đó gen bị gián đoạn bởi ADN lạ và bị bất hoạt.

Kỹ thuật trên thực sự đơn giản đối với cơ thể đơn bào (như vi sinh vật hoặc nấm men). Tuy nhiên với cơ thể đa bào, chúng ta không mong muốn chỉ một tế bào đơn lẻ mà toàn bộ các tế bào trong cơ thể đều mang gen đột biến. Để thỏa mãn yêu cầu đó, vector chuyên chở gen đột biến có thể được vi tiêm vào nhân của trứng vừa thụ tinh trước khi hai nhân đơn bội (của tinh trùng và trứng) chưa kết hợp thành nhân lưỡng bội. Sau đó, trứng thụ tinh được đưa trở lại dạ con. Ngay trong giai đoạn tế bào trứng bắt đầu phân cắt, ADN lạ được ghép vào nhiễm sắc thể của genome lưỡng bội theo cơ chế trao đổi chéo tương đồng. Một tỷ lệ rất nhỏ các tế bào đầu tiên của phôi có chứa gen đột biến. Các tế bào đó có thể biệt hoá thành dòng tế bào sinh dục. Ngoài ra có thể tiến hành kỹ thuật gây đột biến gen trực tiếp trên tế bào mầm phôi (Embryonic Stem cell - ES cell). Tế bào ES mang gen đột biến được cấy trở lại phôi ở giai đoạn sớm. Các tế bào mầm chưa biệt hoá nên chúng có tiềm năng phát triển thành bất cứ loại tế bào nào trong quá trình sinh trưởng phát triển, tức là chúng có khả năng biệt hoá thành dòng tế bào sinh dục. Như vậy, động vật phát triển từ phôi bị cấy ES chứa ẩn tiềm năng mang dòng tế bào sinh dục có gen đột biến.

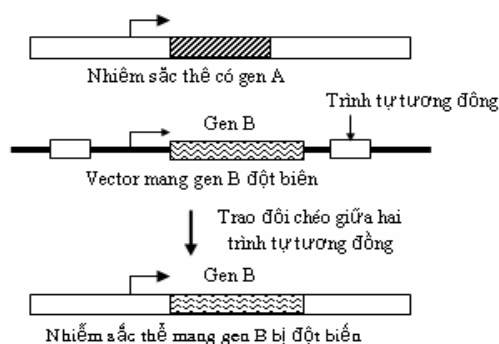
Những thí nghiệm tạo động vật chuyển gen thường được tiến hành trên chuột nhắt. Chuột sinh ra từ phôi cấy ES hoặc từ trứng bị vi tiêm được gọi là thế hệ Go. Chúng có sự pha trộn giữa các tế bào bình thường và các tế bào chuyển gen. Thế hệ Go giao phối với nhau sẽ sinh ra chuột thuần chủng (thế hệ G1) đối với gen đột biến. Điều này có nghĩa tất cả tế bào trong cá thể đó đều mang gen bất hoạt. Chúng được gọi chung là chuột “knock-out” (tạm dịch là chuột bị đánh bật gen). Tuy nhiên, chuột knock-out chỉ có ý nghĩa đặc biệt quan trọng khi nghiên cứu các gen nếu bị bất hoạt cũng không gây chết.

1. Đưa gen chống chịu kháng sinh vào genome



Gen chống chịu kháng sinh ghép vào genome. Chọn lọc tế bào chuyên gen có khả năng sống trong môi trường bổ sung kháng sinh để dùng trong bước thứ hai.

2. Thay thế gen đột biến cho gen chống chịu kháng sinh



Gen đột biến B được ghép vào genome. Chọn lọc các tế bào chuyên gen nhạy cảm với kháng sinh.

Hình 3.26:

Kỹ thuật thay thế gen theo hai bước

Đầu tiên gen chỉ thị (chống chịu kháng sinh) được ghép vào nhiễm sắc thể nhờ trao đổi chéo tương đồng. Chọn lọc các tế bào sống trong môi trường có kháng sinh.

Gen đột biến hoặc gen có promoter mạnh thay thế cho gen chỉ thị nhờ trao đổi chéo giữa trình tự tương đồng

Sau cùng là chọn các tế bào nhạy cảm với kháng sinh để cấy vào mô phôi (theo Brown, 2000)

Đột biến làm mất chức năng của gen dựa vào trao đổi chéo giữa các trình tự tương đồng, thường là trình tự ở vùng chứa mã di truyền. Tuy nhiên, có thể tác động đến hoạt động của gen theo chiều hướng ngược lại, tức là tăng mức độ biểu hiện của gen. Kỹ thuật này được gọi là “gain of function” (tạm dịch là tăng cường chức năng) liên quan chủ yếu đến vùng ADN điều khiển (*promoter*). Lúc này, vùng ADN chứa mã di truyền (cần bảo toàn nguyên vẹn) được lắp ghép với promoter mạnh và sau đó được đưa vào vector dưới dạng nhiều bản sao. Vector được đưa vào tế bào ES và chọn lọc sao cho mỗi tế bào ES có thể có tới 40-200 copy của gen cần nghiên cứu. Hoạt động quá mức bình thường của những bản sao này thường gây rối loạn các quá trình sinh học trong tế bào nên dẫn tới sự biểu hiện tính trạng mới.

Xác suất để vector (ADN lạ) ghép vào genome nhờ trao đổi chéo giữa hai trình tự nucleotide tương đồng thường rất nhỏ. Vì vậy để chọn lọc được những tế bào bị mất hoặc được tăng cường chức năng, gen chỉ thị (thường là những gen qui định tính chống chịu kháng sinh) được đưa vào genome cùng với gen nghiên cứu. Có thể tiến hành kỹ thuật chuyển gen theo hai bước (Hình 3.26). Bước đầu tiên là chuyển gen chỉ thị và chọn lọc các tế bào có khả năng sống trong môi trường có kháng sinh. Tiếp đến là chuyển gen đột biến hoặc ghép gen

với promoter mạnh. Lúc này cần chọn lọc các tế bào nhạy cảm với kháng sinh do gen nghiên cứu đã thay thế cho gen chỉ thị.

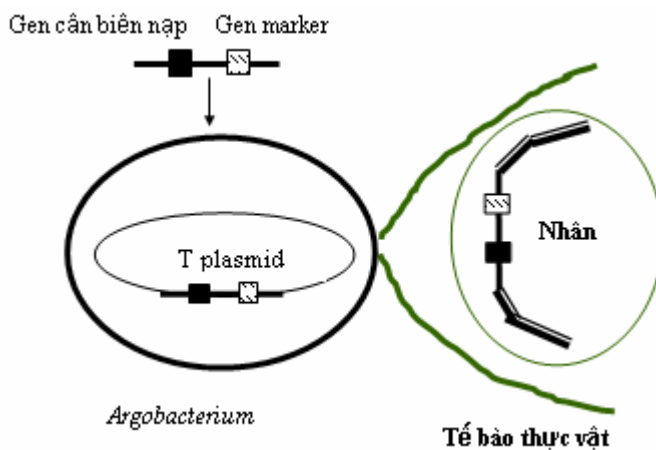
3.8.4 Gen báo cáo “reporter gene”

Kỹ thuật làm mất hoặc tăng cường chức năng cho phép nghiên cứu vai trò của một gen khi chỉ biết trình tự nucleotide. Ngoài ra, có thể căn cứ vào thời điểm promoter được hoạt hoá để xác định sự hoạt động theo thời gian và không gian của gen. Từ đó biết được tính đặc hiệu của gen trong từng giai đoạn sinh trưởng và biệt hoá tế bào. Một promoter được xem là hoạt động khi xuất hiện các phân tử ARNm hoặc sản phẩm protein tương ứng với vùng mã di truyền nằm sau promoter. Do đó, promoter cần nghiên cứu thường được ghép với vùng chứa mã di truyền cho protein có khả năng phát huỳnh quang hoặc tham gia phản ứng tạo màu. Nhờ vậy thời điểm cũng như vị trí xuất hiện protein trong các tổ chức mô có thể phát hiện được một cách dễ dàng. Vùng ADN tương ứng với mã di truyền của protein đã biết được gọi là gen báo cáo. Ở đây, reporter gene được dịch là gen báo cáo để tránh sự nhầm lẫn với khái niệm gen chọn lọc hoặc gen chỉ thị (*marker gene*). Một số gen báo cáo thường gặp như các gen mã cho β -galactosidase (*lacZ*), β -glucuronidase (*uidA*), luciferase (*lux*), Green Fluorescent Protein (GFP). Ngoài ra, có thể ghép một phần của gen báo cáo với vùng chứa mã di truyền cho một protein nào đó (tạo nên protein tái tổ hợp) để nghiên cứu sự phân bố cũng như vai trò của protein này đến hoạt động sống trong tế bào. Sản phẩm của các gen báo cáo như *lacZ* hoặc *uidA* chỉ quan sát thấy khi sử lý tế bào hoặc mô chuyển gen với một số hoá chất. Do đó, các tế bào đều bị chết. Tuy nhiên, protein GFP sẽ phát màu khi chiếu ánh sáng xanh hoặc tia tử ngoại yếu mà không cần bất cứ sự hỗ trợ nào khác. Do đó, có thể phát hiện tế bào mang gen mã cho GFP một cách dễ dàng mà không gây tổn thương chúng. Căn cứ vào thời điểm, vị trí xuất hiện GFP cũng như cường độ phát sáng có thể theo dõi sự tổng hợp, phân bố và di chuyển của protein tái tổ hợp trong tế bào hoặc mô sống. Bằng cách gây đột biến từng nucleotide của gen mã cho GFP, kỹ thuật ADN tái tổ hợp đã tạo ra các biến thể khác nhau của protein GFP phát ra các màu khác nhau, với cường độ mạnh gấp hàng chục lần so với các dạng GFP tồn tại trong tự nhiên.

3.8.5 Biến đổi genome thực vật

Khi thực vật bị tổn thương, các tế bào đã biệt hoá bước vào phân chia làm tăng số lượng tế bào. Điều đặc biệt là dù đã biệt hoá, những tế bào thực vật vẫn có khả năng phân chia tạo ra các loại tế bào khác nhau. Thậm chí ở một số loài cây, tế bào biệt hoá vẫn có khả năng phát triển thành cây hoàn chỉnh cho các giao tử. Tính chất đặc biệt này của tế bào thực vật được ứng dụng để tái tạo cây hoàn chỉnh từ các tế bào nuôi cấy.

Khi nuôi cấy những phần nhỏ của mô thực vật trong môi trường đầy đủ dinh dưỡng và các chất điều hoà sinh trưởng, rất nhiều tế bào bị kích thích phân chia liên tục một cách vô tổ chức tạo ra số lượng lớn các tế bào không biệt hoá gọi là mô sẹo (*callus*). Khi điều chỉnh nồng độ các chất trong môi trường thích hợp, ngọn và rễ cây xuất hiện từ *callus*. Đối với nhiều loài thực vật, khi tách các tế bào *callus* rời nhau và nuôi cấy trong môi trường vô trùng, từ một tế bào riêng lẻ có thể phát triển thành cây con hoàn chỉnh. Bằng các kỹ thuật hiện đại, ADN lạ được đưa vào genome tế bào thực vật. Nuôi cấy các tế bào đó sẽ nhận được cây chuyển gen hoàn chỉnh. Thực tế thường sử dụng *Agrobacterium* như là vector chuyên chở gen vào genome thực vật (Hình 3.27).



Hình 3.27:

Sử dụng plasmid tái tổ hợp vận chuyển gen vào genome tế bào thực vật. Đoạn T-ADN của plasmid Ti được biến đổi, bỏ bớt những gen nằm giữa hai đầu trái và phải, đồng thời nhận thêm gen cần chuyển. Khi *Argobacterium* xâm nhiễm tế bào thực vật, đoạn T-ADN mang gen lạ được chuyển vào genome thực vật.

Ưu việt của thực vật chuyển gen tạo ra những bước tiến nhanh trong chọn giống, kết hợp các tính trạng di truyền có lợi của các loài khác nhau trong một cá thể. Thực vật chuyển gen đáp ứng nhu cầu xã hội về năng suất và hiệu quả. Tuy nhiên, tác động của thực vật chuyển gen đối với cân bằng sinh thái, môi trường đang còn là vấn đề tranh cãi.

Chương 4

TỔNG HỢP VÀ VẬN CHUYỂN PROTEIN

Genome chứa mọi thông tin di truyền và lưu trữ các chương trình cần thiết để tế bào sinh trưởng, duy trì và phát triển. Protein chiếm hơn một nửa trọng lượng khô của tế bào. Sự tồn tại, đổi mới của chúng giữ vai trò quyết định nhất để duy trì sự sống cũng như đảm bảo sinh trưởng và phát triển của cơ thể. Toàn bộ protein trong tế bào được mã bởi các gen thuộc genome được gọi chung là **proteome**. Trong một tế bào động vật, số protein khác nhau dao động trong khoảng 10.000-20.000 loại, tương ứng với chừng 10 tỷ (10^{10}) phân tử. Số gen mã cho proteome chỉ tương ứng với phần ADN mang mã di truyền, do đó chỉ đại diện cho một phần mà không phải toàn bộ genome. Hơn nữa, từ một gen có thể có nhiều sản phẩm protein.

Hoạt động của một gen được nhìn nhận như một quá trình gồm hai giai đoạn: phiên mã và dịch mã. Trong các chương trước, chúng ta đã xét đến các bước chuẩn bị được hoàn thiện ở trong nhân tế bào và các cơ chế kiểm soát ở giai đoạn thứ nhất để tạo ra sản phẩm ARNm. Các phân tử ARNm được vận chuyển ra ngoài tế bào chất để dịch mã tổng hợp proteome. Quá trình này được thực hiện bởi bộ máy Ribosome với sự tham gia của ARNt. Sau khi được tổng hợp, protein bị biến đổi trong quá trình vận chuyển đến những vị trí khác nhau tùy thuộc vào chức năng của chúng. Ví dụ, các enzym được vận chuyển về lysosome, protein điều hoà hoạt động gen được đưa vào nhân vv... Có thể xem hoàn thiện cấu trúc bậc I (trật tự acid amin) là bước cuối cùng liên quan đến hoạt động của một gen. Tuy nhiên, hoạt tính của protein hoàn toàn phụ thuộc vào cấu trúc không gian bậc hai, bậc ba và bậc bốn. Cấu trúc bậc hai liên quan đến cấu hình không gian cục bộ từng phần của chuỗi peptide. Cấu trúc bậc ba phản ánh cấu hình không gian ba chiều của toàn bộ chuỗi và cấu trúc bậc bốn đặc trưng cho cấu hình không gian khi các chuỗi polypeptide tương tác với nhau tạo ra protein có hoạt tính. Phân tử protein có thể chuyển từ cấu hình không gian này sang cấu hình không gian kia gắn liền với các hoạt tính khác nhau của chúng.

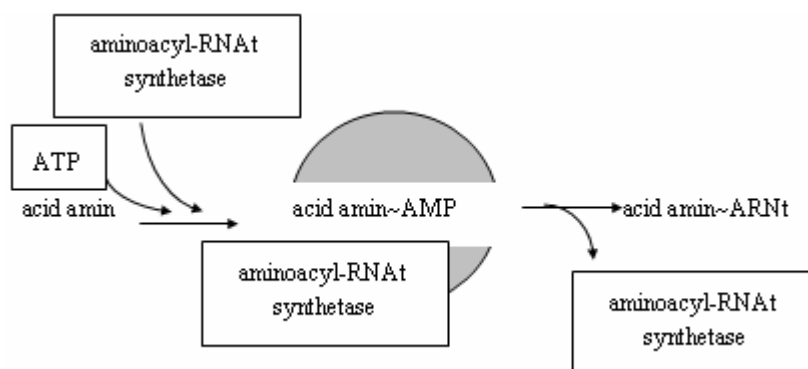
Bảng 4.1: Các phân tử cấu thành nên Ribosome ở prokaryot và eukaryot				
Ribosome	Prokaryot		Eukaryot	
	ARNr	Số lượng protein	ARNr	Số lượng protein
Tiểu đơn vị nhỏ	16S	21	18S	33
Tiểu đơn vị lớn	5S, 23S	31	5S; 5.8S; 28S	49

Quá trình tổng hợp protein đòi hỏi sự có mặt của 3 loại ARN. Ngoài sự đa dạng của ARNm, 3 đến 5 loại phân tử ARNr (*ARN ribosome*) khác nhau kết hợp với protein để tạo thành cấu trúc Ribosome (Bảng 4.1). Tế bào có khoảng 35-60 phân tử ARNt (ARN vận chuyển) giữ chức năng “adaptor” chuyển đổi mã di truyền trên ARNm sang acid amin. Mỗi acid amin được mã bởi ba nucleotide (mã bộ ba). Mỗi mã bộ ba trên ARNm được đọc bởi đối mã (*anticodon*) nằm trên phân tử ARNt. Khi hai mã này khớp nhau, 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. Có thể có nhiều ribosome đồng thời dịch mã trên một sợi ARNm tạo nên cấu trúc polyribosome.

4.1 Vai trò của ARN vận chuyển (ARNt) trong tổng hợp protein

Phân tử ARNt có kích thước nhỏ (70 đến 90 nucleotide), làm nhiệm vụ vận chuyển acid amin tương ứng với mã di truyền trên sợi ARNm. Một số phân tử ARNt có cùng khả năng vận chuyển một acid amin. Vì vậy, để chuyên chở 21 acid amin, trong tế bào prokaryot và eukaryot có 35 và 60 phân tử ARNt khác nhau. Acid amin thứ 21 là selenocysteine. Đây là acid amin hiếm gặp, có cấu trúc tương tự như cysteine nhưng selenium thay thế cho sulfur. Hầu hết các acid amin của protein có sự biến đổi (thêm, bớt, thay thế các nhóm chức...) sau khi chuỗi polypeptide đã tổng hợp. Tuy nhiên, selenocysteine được tạo ra nhờ biến đổi xảy ra ở serine, sau đó được nhận biết bởi ARNt đặc hiệu để ghép vào chuỗi peptide. Mã di truyền của selenocysteine là UGA, mặc dù đây cũng là mã dừng tổng hợp đối với hầu hết các protein. Các phân tử ARNm mã cho các protein có chứa selenocysteine thường có cấu trúc không gian đặc biệt khiến cho bộ máy tổng hợp protein nhận biết mã UGA không phải là mã dừng.

Phân tử ARNt được phiên mã bởi ARN polymerase I và có chứa intron. Điều đáng lưu ý, intron của ARNt có khả năng tự cắt (intron thuộc nhóm I). Chức năng của một ARNt phụ thuộc rất nhiều vào cấu trúc không gian ba chiều - tức là phụ thuộc vào tương tác tạo cặp giữa các nucleotide bổ sung, hay thậm chí không bổ sung, để tạo nên sự gấp khúc chính xác của phân tử. Đa số các phân tử ARNt thường có cấu trúc gồm các thùy hình "lá nhép" và các nucleotide thường bị biến đổi (gắn thêm các nhóm chức) sau khi được phiên mã từ gen. Chính những biến đổi này giúp cho sự hình thành liên kết giữa các nucleotide không bổ sung đặc trưng cho ARNt.



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 acid amin~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-tRNA synthetase.

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 hoạt hoá acid amin, bước tiếp theo là gắn nó vào ARNt đặc hiệu cho acid amin đó. Phân tử ARNt có thể chọn đúng acid amin đặc hiệu mà nó cần vận chuyển trong số các acid amin khác cùng tồn tại trong tế bào chất? Đó là nhờ hoạt tính của aminoacyl-ARNt synthetase. Một acid amin sẽ tương ứng với một enzym riêng biệt. Enzym xúc tác cho phản ứng hoạt hoá acid amin bằng việc sử dụng năng lượng của ATP để tạo ra phức acid amin~AMP. Phức này gắn với enzym cho đến khi tìm được ARNt đặc hiệu. Enzym sẽ chuyển acid amin từ phức sang đầu 3' của ARNt.

Phân tử ARNt mang acid amin đi 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

thể 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} tương ứng với mã bộ ba cho cysteine trên ARNm. Điều đó cho thấy tính chính xác của quá trình tổng hợp protein phụ thuộc vào các phân tử ARNt.

Mặc dù không tồn tại mã thoái hoá cho methyonine nhưng có hai loại ARNt^{Met}. Trong tế bào vi khuẩn hoặc trong các bào quan như ty thể, lục lạp, acid amin đầu tiên của chuỗi polypeptide luôn luôn là methyonine bị formyl hoá tại nhóm amin. Như vậy, *initiator* RNAt^{Met} nhận biết methyonine bị formyl hoá và đưa nó vào vị trí đầu tiên của chuỗi peptide. Phân tử RNAt^{Met} bình thường nhận biết các mã AUG cho methyonine nằm trong chuỗi. Tuy nhiên, trong tế bào eukaryot, methionine đầu tiên không bị formyl hoá. Do đó, hai loại ARNt^{Met} này chỉ khác nhau ở cấu trúc không gian.

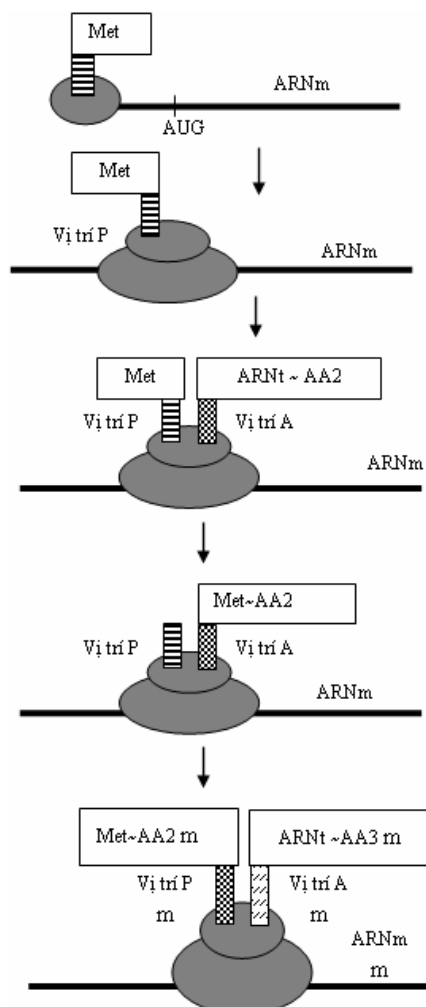
4.2 Tổng hợp protein ở bộ máy Ribosome

Ribosome bao gồm ARNr tương tác với các protein (Bảng 4.1). Một số protein ribosome giữ vai trò tạo cấu trúc khung cho bộ máy tổng hợp protein. Chính các ARNr quyết định hoạt tính của các phản ứng xảy ra trong ribosome, còn protein trong cấu trúc ribosome có tính chất phụ trợ, thúc đẩy các phản ứng đó.

Ribosome mang các vị trí tương tác đặc biệt: ba vị trí cho ARNt và một vị trí cho ARNm. Vị trí **P** (**p**eptidyl-RNAt-binding site) liên kết với một phân tử ARNt mà phân tử này được gắn với chuỗi polypeptide đang được tổng hợp. Vị trí **A** (**a**minoacyl-RNAt-binding site) là nơi tương tác của phân tử ARNt thứ hai vừa mang một acid amin vào. Vị trí **E** (*e*xit site) là nơi ARNt đã làm xong nhiệm vụ, chuẩn bị được giải phóng khỏi ribosome. Các ARNt này chỉ được giữ chặt ở các vị trí P hoặc A một khi anticodon của chúng tạo cặp bổ sung với các codon trên phân tử ARNm.

Quá trình tổng hợp protein ở ribosome được chia làm 3 giai đoạn: bắt đầu tổng hợp chuỗi peptide, kéo dài và kết thúc chuỗi peptide. Đầu tiên, 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 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, nhờ vào cấu trúc mũ ^{7m}G ở đầu 5' của phân tử ARNm mà tiểu đơn vị ribosome nhỏ tìm được phân tử ARNm. Một số protein liên kết với mũ này, tiếp đến các yếu tố khởi động liên kết với những protein đó tạo thành phức với ARNm. Thông qua phức này mà tiểu phần ribosome nhỏ 40S tương tác với ARNm và bắt đầu dò tìm mã AUG của 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) mang acid amin đầu tiên (*methyonine*) 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 được kéo dài (Hình 4.2).

Giai đoạn kéo dài chuỗi polypeptide về cơ bản là giống nhau ở tế bào prokaryot và eukaryot. 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:



Hình 4.2:

Quá trình tổng hợp protein. Đầu tiên ARNt mang Methionine 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.

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 enzyme 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 codon dọc theo phân tử ARNm. Bước này 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 peptidyl transferase khiến enzym thêm phân tử H₂O 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.

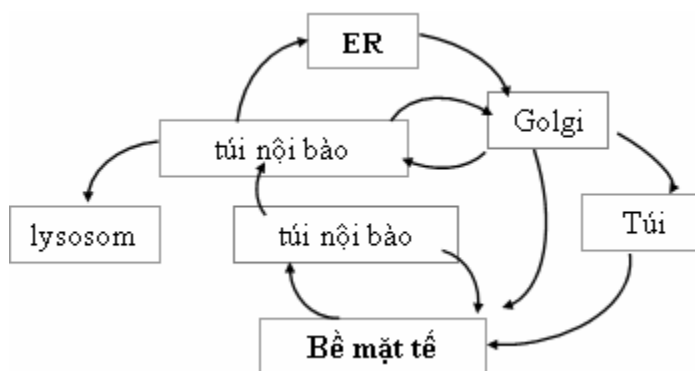
Một sợi ARNm có thể được dịch mã đồng thời bởi nhiều ribosome. Ví dụ, trên sợi ARNm mã cho peptide fibroin (protein ở trong sữa có trọng lượng phân tử 200 KDa) có tới 50-80 ribosome đồng thời dịch mã. Quá trình tổng hợp protein tiêu tốn năng lượng hơn bất cứ phản ứng sinh hoá nào xảy ra trong tế bào. 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.

4.3 Vận chuyển protein

Phần lớn protein ở tế bào vi khuẩn không bị biến đổi sau khi được tổng hợp. Chúng thực hiện ngay chức năng của mình ở bên trong hoặc được tiết ra bên ngoài tế bào. Tế bào tiết protein ra ngoài nhằm đáp ứng yêu cầu về chất dinh dưỡng, truyền tin bảo vệ hoặc duy trì vật liệu tạo cấu trúc bên ngoài màng tế bào. Những protein có chức năng, hoạt tính ở bên ngoài tế bào được gọi chung là protein tiết. Cách thức đưa protein tiết ra ngoài màng tế bào được gọi là các con đường tiết. Tùy thuộc cấu trúc màng mà protein được vận chuyển qua màng theo những cách thức khác nhau. Ví dụ, ở tế bào vi khuẩn Gram dương, protein tiết được vận chuyển vượt qua màng tế bào chất để ra ngoài. Tuy nhiên, ở tế bào vi khuẩn gram âm, protein tiết phải đi qua màng trong, lớp phân cách giữa hai màng và cuối cùng đi qua màng ngoài.

Không giống như protein ở vi khuẩn, hầu hết protein của tế bào eukaryot sau khi được tổng hợp còn trải qua những biến đổi như glycosyl hoá, acetyl hoá, gắn thêm các nhóm phospho, gốc sulfat... trên con đường vận chuyển đến đích. Những biến đổi này được thực hiện chủ yếu ở mạng lưới nội chất và bộ máy Golgi.

Hầu hết các protein đi vào màng lưới nội chất ER (Endoplasmic Reticulum) ngay khi chúng đang được tổng hợp. Vì thế trên bề mặt của ER luôn có nhiều ribosome bám vào khiến cho bề mặt ER trở nên "sần sùi" không trơn. Do đó những vùng ER có ribosome bám vào được gọi là ER hạt, còn vùng không có ribosome là ER trơn. Vùng ER trơn còn được gọi là vùng ER chuyển tiếp (*transitional ER*) vì đây là nơi hình thành các túi vận chuyển protein và lipid đi từ ER đến Golgi. Trong một số loại tế bào như tế bào gan, vùng ER trơn chiếm phần lớn mạng lưới nội chất 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. Chúng ta cần lưu ý rằng ribosome tồn tại tự do trong tế bào chất vẫn tham gia tổng hợp protein. Không có sự khác nhau giữa các ribosome tự do và ribosome bám trên ER. Mọi ribosome đều bắt đầu tổng hợp một chuỗi peptide bất kỳ ở trong tế bào chất. Chính tín hiệu vận chuyển đến ER nằm ở đầu NH₂ chuỗi peptide sẽ đưa ribosome đến bề mặt ER. Do nhiều ribosome cùng tham gia tổng hợp peptide trên một sợi ARNm nên sợi đó luôn gắn với bề mặt ER. Từ màng lưới ER, các protein đi đến Golgi và tiếp tục được đưa đến các vị trí khác nhau như các bào quan hoặc tiết ra khỏi tế bào (Hình 4.3).



Hình 4.3:

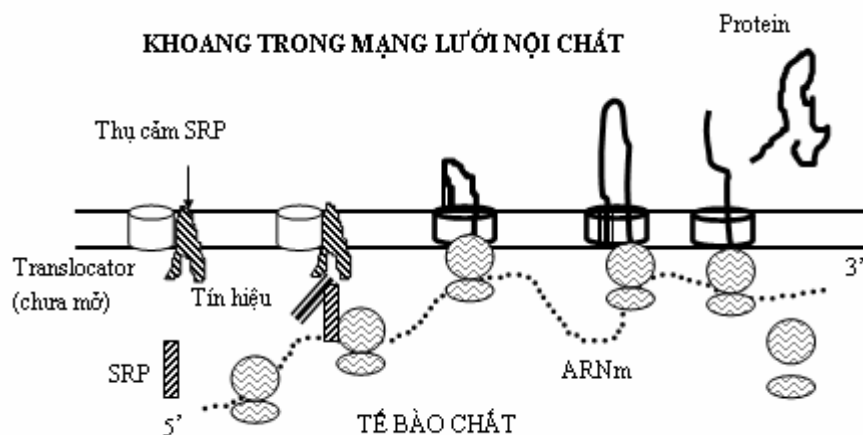
Các con đường vận chuyển protein từ ER đến các vị trí đích khác nhau và từ bề mặt tế bào đến mạng lưới nội chất ER đi qua các bước trung gian ở túi nội bào sớm và túi nội bào muộn.

Trên đường vận chuyển, protein chịu những biến đổi về cấu trúc không gian và thành phần hoá học tùy thuộc vào đích cần đến. Cần chú ý rằng các protein có đích phân bố trong nhân hoặc ty thể, lập thể sẽ được tổng hợp bởi ribosome tự do trong tế bào chất. Những protein đó được tổng hợp trọn vẹn và có cấu trúc không gian hoàn chỉnh ở ngoài tế bào chất, sau đó mới được vận chuyển đến đích.

4.3.1 Vận chuyển vào mạng lưới nội chất

Con đường vận chuyển protein từ mạng lưới nội chất ER sang Golgi và đi về các đích khác nhau (các bào quan ở trong hoặc bên ngoài tế bào) được gọi là con đường tiết (*secretory pathway*). Bất chấp vị trí phân bố cuối cùng khác nhau, các protein đi vào ER nhờ tín hiệu dẫn nằm ở đầu NH_2 . Tín hiệu này được gọi là tín hiệu bắt đầu vận chuyển (*start-transfer signal*). Phản ứng tổng hợp chuỗi peptide được bắt đầu nhờ các ribosome tự do trong tế bào chất. Tuy nhiên, khi đầu NH_2 của sợi peptide có chứa tín hiệu vận chuyển ra khỏi ribosome, tín hiệu được nhận biết bởi phức SRP (*Signal Recognition Particle*). Liên kết giữa tín hiệu và SRP làm dừng tạm thời phản ứng tổng hợp protein, giúp ribosome có thời gian đến bề mặt ER. Chuỗi peptide nằm tiếp sau tín hiệu vận chuyển được đưa vào khoang trong của ER.

Tín hiệu vận chuyển là đoạn peptide khoảng 30 acid amin trong đó có 6 đến 12 acid amin kỵ nước thường nằm cạnh nhau. Đoạn kỵ nước giữ vai trò quan trọng đối với tương tác giữa tín hiệu bắt đầu vận chuyển và thụ cảm tương ứng nằm trên màng ER. 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 ER. Vai trò của tín hiệu vận chuyển được minh chứng bằng thí nghiệm tổng hợp protein *in vitro* với sự có mặt của microsomes. Microsomes là những phần ER tạo ra do ER bị đứt gãy khi tế bào bị nghiền đồng thể. Các phần ER nhỏ có khả năng tạo thành các tiểu phần dạng tròn (đường kính ~100-200 nm). Khi không có microsomes, phân tử protein tổng hợp *in vitro* có đoạn peptide mang tín hiệu vận chuyển. Nếu cho thêm protease vào môi trường thì protein bị phân hủy. Tuy nhiên khi bổ sung microsomes vào môi trường trước khi protein được bắt đầu tổng hợp thì ở bên trong microsomes xuất hiện phân tử protein không còn mang tín hiệu vận chuyển. Do đó, protein không bị phân hủy nếu bổ sung thêm protease. Như vậy, tín hiệu vận chuyển giúp protein đi qua màng ER vào khoang trong. Tín hiệu này bị cắt khỏi protein và bị phân hủy thành các acid amin bởi peptidase và protease có trong khoang ER (Hình 4.4).



Hình 4.4:

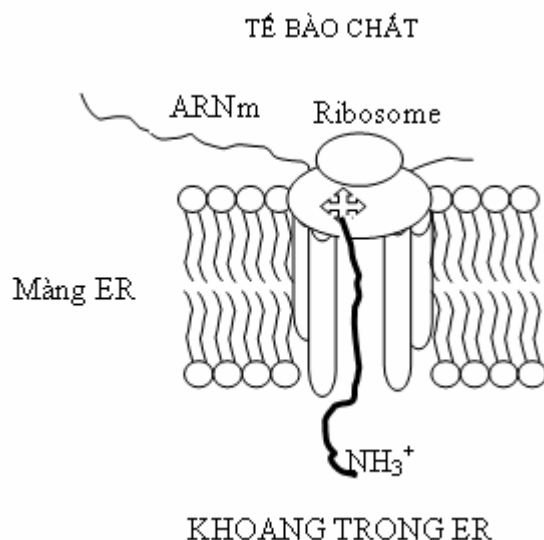
Ribosome bám vào màng lưới nội chất (ER) nhờ thụ cảm SRP nằm ở trên màng. Thụ cảm tương tác với phức SRP. Phức SRP gồm 6 chuỗi peptide và một phân tử ARN (có khả năng tạo cặp với ARNr). SRP có hai vị trí liên kết với tín hiệu dẫn và Ribosome. SRP tồn tại trong tế bào chất liên kết với đoạn peptide (đầu NH₂) do Ribosome tổng hợp. Sau khi SRP đã tạo phức với Ribosome và protein, SRP tương tác với thụ cảm của nó trên màng ER (theo Alberts & cs., 2002).

Kỹ thuật ADN tái tổ hợp đã thiết kế các phân tử protein mang đầu NH₂ bất kỳ nhưng đảm bảo độ dài và tính kỵ nước thì chúng đều có thể đi qua màng ER để vào khoang bên trong. Thực nghiệm xác định được quá trình vận chuyển vào khoang ER được bắt đầu khi đoạn peptide ra khỏi ribosome có chiều dài khoảng 70 acid amin. 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. Đầu NH₂ được đưa vào trong màng lưới nội chất ngay khi sợi peptide chưa được tổng hợp xong.

Phức SPR nhận biết một cách đặc hiệu tín hiệu vận chuyển. Phức này gồm có 6 protein khác nhau và một phân tử ARN (~300 nucleotide) có trình tự bổ sung với ARNr. Khi tín hiệu vận chuyển xuất hiện và chuỗi peptide đang tổng hợp có đủ độ dài nhất định thì SPR tương tác với tín hiệu, với ribosome và với thụ thể nằm trên màng ER. Tương tác đó giúp ribosome dường như bị dính sát trên bề mặt ER. Thụ thể của SRP được gọi là "**docking protein**" gồm hai chuỗi polypeptide α (nằm trên màng ER) và β (nằm xuyên qua màng ER). Thí nghiệm tổng hợp protein *in vitro* cho thấy khi không có microsome thì phức SRP sẽ gây ngừng quá trình tổng hợp protein khi đầu NH₂ ra khỏi ribosome. Như vậy, SRP không chỉ làm nhiệm vụ mang các chuỗi peptide đang được ribosome tổng hợp đến bề mặt ER mà còn có chức năng ngăn cản quá trình tổng hợp protein hoàn chỉnh nếu như thiếu vắng ER. Như vậy, phức SRP có ba domain qui 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.

Ribosome tiếp tục tổng hợp protein sau khi SRP đã liên kết với thụ thể của riêng nó trên màng ER. Sau đó, SRP tách khỏi ribosome và chuỗi peptide được chuyển sang phức protein nằm trong cấu trúc màng ER. Phức này được gọi là **translocator** gồm 3-4 tiểu phần khác nhau; mỗi tiểu phần gồm 3 protein nằm xuyên qua màng ER. 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 ER 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 ER. Trong suốt quá trình đưa chuỗi peptide vào ER, ribosome nằm sát translocator bị kín sao cho không có bất cứ thành phần nào ở bên trong khoang ER, đặc biệt là các cation Ca⁺², lọt ra ngoài. Protein tham gia phức translocator phân bố trên màng ER được xác định

bằng thí nghiệm tổng hợp protein *in vitro* (Hình 4.5). Thí nghiệm được tiến hành với phân tử ARNm có mang mã di truyền cho lysin phân bố ở đoạn giữa của ARNm. Mã này sẽ được đọc bởi phân tử ARNt đặc biệt làm nhiệm vụ vận chuyển lysin có gắn thêm chất đánh dấu. Do đó, chuỗi peptide được tổng hợp có mang lysin đánh dấu. Khi chuỗi được vận chuyển đi qua translocator, chất đánh dấu có khả năng tương tác với màng ER nên chuỗi peptide không vào được trong khoang ER mà bị giữ lại ở translocator (do chất đánh dấu tương tác với màng tại vị trí chuỗi peptide đi qua). Nhờ đó xác định được các protein tham gia cấu trúc phức translocator. Một trong các protein của translocator là Sec61p. Protein này đóng kênh khi không có protein đi qua. Khi có mặt ribosome, Sec61p tương tác với tiểu phần lớn của ribosome, do đó sợi peptide được đi qua trong khi ribosome bịt chặt kênh không cho các phân tử khác tự do đi qua.



Hình 4.5:
Cấu trúc của translocon tạo kênh dẫn cho phép chuỗi polypeptide được vận chuyển qua màng ER vào trong khoang khi đang được ribosome tổng hợp (theo Lodish và cs., 2000).

Không phải mọi protein vận chuyển vào ER 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 ER thông qua translocator. Ở vi khuẩn, protein *SecA* ATPase tham gia cấu trúc của translocator. Protein này phân bố ở mặt ngoài ER, 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 lỗ. 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 ER. Như vậy, các protein có cấu trúc không gian không thể đi qua translocator để vào khoang trong ER.

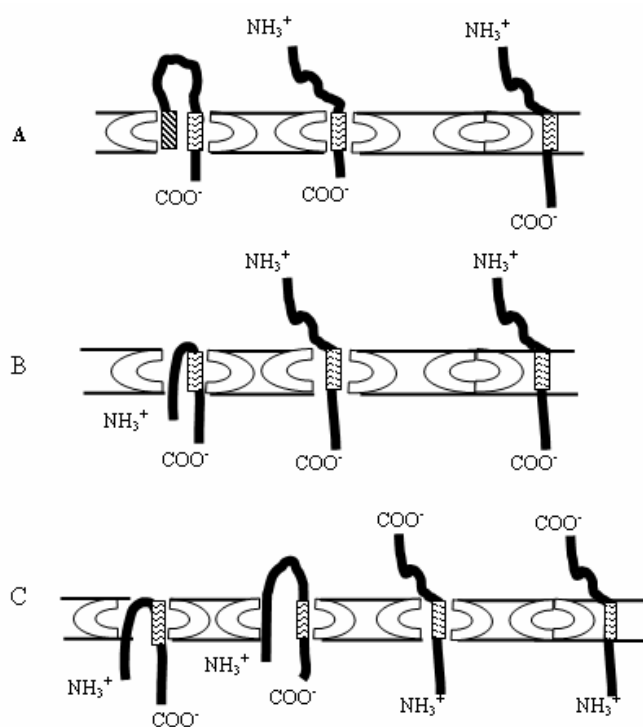
Vận chuyển phân tử protein đang tổng hợp vào khoang ER giúp protein không ở trạng thái tự do trong tế bào chất. Phải chăng đây là một cơ chế bảo đảm an toàn cho tế bào chất, ngăn cản không cho các protein tiết hoặc các enzym tồn tại ở dạng tự do trong tế bào chất, tránh phá hủy các thành phần nội bào. Với một số tế bào đặc biệt làm nhiệm vụ tiết hydrolase, tế bào chất chứa các chất ức chế hydrolase để bảo vệ tế bào không bị phân hủy bởi các enzym này.

4.3.2 Vận chuyển protein cấu trúc 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 ER. Tuy nhiên, chúng không được giải phóng vào khoang trong ER mà vẫn giữ lại trên màng ER. Chiều phân cực của những protein này ở trên màng ER, tức là vị trí các đầu NH_3^+ và COO^- ở trong khoang ER hay ngoài tế bào chất, 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 ER 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 tổng hợp trên ER.

Chúng ta hãy xét trường hợp đơn giản nhất là sự vận chuyển của protein màng mà chúng chỉ nằm xuyên qua màng một lần. Trong trường hợp này, một số protein, ví dụ như glycoporphin trên màng tế bào hồng cầu hay thụ thể của insulin, có phần đầu COO^- nằm trong tế bào chất còn phần đầu NH_3^+ nằm phía ngoài màng. Tuy nhiên, một số protein khác, ví dụ như thụ thể của transferrin, phân cực ngược lại; chúng có đầu COO^- nằm phía ngoài màng còn phần đầu NH_3^+ nằm phía trong tế bào chất. Đối với trường hợp thứ nhất, chuỗi peptide có hai tín hiệu đặc biệt: tín hiệu vận chuyển qua ER như đã mô tả phần trên và tín hiệu thứ hai gồm khoảng 20-25 acid amin kỵ nước tạo nên cấu trúc xoắn α . Trong quá trình chuỗi peptide vừa đang tổng hợp vừa đi vào khoang ER, tín hiệu ER bị cắt khỏi chuỗi peptide, giải phóng đầu NH_3^+ vào trong khoang ER. Chuỗi peptide tiếp tục được đưa vào khoang cho đến khi tín hiệu thứ hai đi qua translocator thì bị giữ lại.

Do đó đoạn peptide nằm 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 (đây là phức protein) và nằm xen vào lớp phospholipid của màng ER. 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 4.6).



Hình 4.6:**Vận chuyển protein màng**

Các protein thuộc cấu trúc màng được giữ lại ở màng ER 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 ER (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.

Đối với protein màng có đầu COO^- nằm bên ngoài màng và đầu NH_3^+ nằm phía trong tế bào chất, ví dụ như thụ thể của transferrin, 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 vận chuyển ER và chức năng cố định tại ER. 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 ER như đã mô tả ở mục 4.3.1. Tín hiệu đi vào translocator và đoạn peptide nằm phía sau tín hiệu đi vào trong khoang ER. 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 của translocator và xuyên vào 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.

Khi chuỗi peptide là protein cấu trúc màng của bào quan khác và chỉ xuyên qua màng một lần thì chính đoạn acid amin phân bố qua màng có thể đảm nhận đồng thời các chức năng vận chuyển, dừng và giữ chuỗi đó trên màng ER. Giả sử chuỗi này có đầu NH_3^+ phân bố bên ngoài bào quan (tức là đầu NH_3^+ nằm trong tế bào chất) thì đầu COO^- sẽ bị đưa vào trong khoang ER và túi chứa peptide này hình thành từ màng ER sao cho đầu NH_3^+ sẽ nằm phía ngoài túi tiết. Sau đó, túi tiết sẽ đi đến bộ máy Golgi và cuối cùng protein màng được đưa đến vị trí đích. Túi tiết chở protein tiết sẽ tiếp hợp với màng bào quan và đảm bảo giữ đúng chiều phân cực bên trong hay bên ngoài của hai đầu NH_3^+ và COO^- . 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 khoang trong ER. 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 ER đã được qui định phù hợp với sự phân bố ở vị trí đích cuối cùng (Hình 4.6).

Một số protein thuộc cấu trúc màng, ví dụ như protein tạo cấu trúc bơm ion, kênh vận chuyển ion, có thể nằm xuyên qua màng nhiều lần. Phần đầu NH_3^+ của những protein này thường chứa hai tín hiệu vận chuyển (*start transfer signal*) và tín hiệu dừng vận chuyển (*stop transfer signal*) nối tiếp nhau. Tiếp theo tín hiệu dừng đầu tiên là các tín hiệu giữ (*membrane-anchor signal*) và tín hiệu dừng khác nằm kế tiếp nhau. Số tín hiệu tương ứng với số lần xuyên qua màng của protein. Những tín hiệu này đều là những đoạn peptide có cấu trúc xoắn α và kỵ nước.

Đối với những protein có vị trí phân bố ở bên trong khoang ER, chúng được giữ lại sau khi được đưa vào ER nhờ tín hiệu đặc biệt gồm 4 acid amin KDEL (Lys-Asp-Glu-Leu). Sau khi được tổng hợp và vận chuyển vào khoang, chúng được giữ lại nhờ tín hiệu này. Tín hiệu đó được gọi là tín hiệu sở hữu "*retention signal*"; tín hiệu này thường nằm ở đầu carboxyl của phân tử protein. Tín hiệu KDEL được nhận biết bởi thụ cảm nằm trên màng ER. Đôi khi, chúng được tìm thấy trên nang vận chuyển làm nhiệm vụ giữ lại các protein mang tín hiệu KDEL. Có thể xảy ra trường hợp vì một lý do nào đó mà các protein này bị đóng gói nhầm

vào nang thay cho việc ở lại trong khoang ER. Lúc đó tế bào sửa chữa sai sót bằng cách bắt giữ các protein có tín hiệu KDEL và đưa chúng quay trở lại ER.

Một trong những protein quan trọng được giữ lại trong khoang ER là disulfide isomerase (PDI) có chức năng xúc tác phản ứng oxi hoá nhóm SH của acid amin cystein để tạo cầu disulfide S-S. Hầu hết các cystein ở đoạn protein phân bố phía ngoài màng tế bào đều có liên kết cầu disulfide. Tuy nhiên, phần protein phân bố trong tế bào chất không có cầu disulfide. Bên cạnh PDI, protein chaperon BiP cũng được giữ lại trong khoang ER. Giống như các chaperon khác, BiP nhận biết những protein gấp khúc sai hoặc những tiểu phần tự do của một phức protein. Những tiểu phần này vì một lý do nào đó chưa được tạo thành phức multimer. BiP sẽ liên kết với những protein sai hỏng về mặt cấu trúc này để giữ chúng lại trong ER. Ngoài ra, BiP còn có hoạt tính phân hủy ATP giúp cho các protein có thể gấp khúc khi chúng được vận chuyển vào ER.

4.4 Biến đổi sau dịch mã và kiểm tra chất lượng protein trong khoang ER

Vào trong khoang ER, các protein được gấp khúc tạo cấu trúc không gian riêng biệt và chính xác của từng protein. Quá trình này liên quan đến việc tạo các cầu nối disulfide, gấp khúc chuỗi polypeptide và lắp ráp các monomer với nhau để tạo phân tử protein multimer. Ngoài ra, một số proprotein có thể còn được phân cắt thành các chuỗi nhỏ ngay trong khoang ER. Tuy nhiên quá trình này không phổ biến ở ER mà xảy ra chủ yếu trong Golgi. Mọi phân tử protein không gấp khúc hoặc có sai sót trong việc gấp khúc hay bị sai hỏng khi lắp ráp tạo multimer đều bị giữ lại trong ER. Thậm chí vì lý do nào đó mà chúng vẫn được chuyển từ ER đến Golgi thì chúng lại bị phát hiện tại Golgi và đưa trở về ER. Các protein sai hỏng sẽ bị đẩy ra khỏi ER thông qua kênh translocon và bị phân hủy trong tế bào chất. Những protein bị đẩy ra ngoài tế bào chất thường được gắn với ubiquitin và sau đó bị phân hủy bởi protease. Các protease thường phân bố ngay phía ngoài màng ER để sẵn sàng thực hiện nhiệm vụ của mình.

4.4.1 Tạo cầu liên kết disulfide (S-S) và cuộn gấp trong khoang ER

Các cầu nối disulfide có vai trò quyết định trong việc thiết lập cấu trúc bậc ba và bậc bốn của phân tử protein. Đặc biệt liên kết S-S giữa hai cysteine giữ vai trò quan trọng đảm bảo cấu trúc thích hợp cũng như hoạt tính của các enzym và hormone. Do đó, bên trong khoang ER, phản ứng tạo cầu disulfide được thực hiện trước khi hoàn thiện cấu trúc không gian cho protein. Cầu disulfide có thể được hình thành trên chuỗi peptide đang vận chuyển vào khoang ER. Ngược lại, các protein được tổng hợp tự do trong tế bào chất hoặc trên vùng ER trơn thường không có cầu disulfide. Chúng phải phụ thuộc vào nhiều mối tương tác với các protein khác để thiết lập cấu trúc không gian của mình.

Enzym disulfide isomerase (PDI) phân bố trong khoang ER và xúc tác cho phản ứng oxy hoá nhóm SH tạo cầu disulfide giữa hai cystein của cùng một chuỗi peptide hoặc giữa các chuỗi khác nhau. Một protein có thể có nhiều cầu disulfide. Ví dụ, chuỗi nhẹ immunoglobulin Ig có hai cầu S-S. Khi chuỗi này đang được tổng hợp và vận chuyển vào khoang ER, cầu S-S được thiết lập cho hai cystein nằm gần đầu NH₂ ngay khi cystein thứ ba chưa được gắn vào chuỗi Ig. Điều đó đảm bảo cho sự tạo cầu S-S chính xác. Một số cầu disulfide được hình thành một cách ngẫu nhiên khiến cho cấu trúc không gian của protein bị sai lệch. Trong trường hợp này, PDI sẽ chỉnh sửa lại những cầu S-S không đúng. Bên cạnh phản ứng tạo cầu nối S-S, chuỗi peptide được gấp cuộn theo một trật tự đặc thù riêng của peptide đó. Chính cầu S-S giúp cho việc gấp khúc được chính xác. Một số protein khác của ER như các enzym

peptidyl-prolyl isomerase thúc đẩy sự quay quanh cầu liên kết peptidyl-prolyl của đoạn peptide chưa gấp khúc, giúp chúng gấp khúc dễ dàng hơn.

4.4.2 Hình thành cấu trúc multimer từ các chuỗi peptide

Các protein có cấu trúc multimer được hình thành ngay trong khoang ER. Nghiên cứu sự biến đổi cấu hình của protein Hemagglutinin HA (protein cấu trúc vỏ virus gây bệnh cúm) từ dạng monomer sang trimer xảy ra bên trong ER của tế bào chủ nhiễm virus là ví dụ điển hình cho quá trình hình thành cấu trúc không gian thích hợp đối với từng protein. Bằng việc sử dụng các kháng thể khác nhau, mỗi loại chỉ có khả năng tương tác đặc hiệu với HA ở dạng monomer hoặc dạng trimer, người ta có thể nhận biết được cấu trúc của từng phân tử HA ở trong khoang ER. Kết quả thí nghiệm cho thấy phân tử HA chuyển từ dạng monomer sang trimer trong khoảng thời gian 7 phút. Ngay sau đó, các cầu disulfide trên mỗi sợi được tạo thành và ba sợi tương tác với nhau tạo phân tử HA dạng trimer hoàn chỉnh. Dạng này được vận chuyển tiếp đến Golgi. Khi chuỗi peptide 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 ER.

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 ER. Tương tự như trường hợp HA, kết quả nghiên cứu protein α 1-antiprotease do tế bào gan tiết vào máu nhằm ức chế hoạt tính enzym elastase cho thấy, khi α 1-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 ER mà không được tiếp tục vận chuyển để tiết ra ngoài. Lúc đó tế bào gan bị trương phồng lên bởi mạng lưới ER của nó chứa đầy protein α 1-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 disulfide isomerase... Một khi tế bào gan không cung cấp đủ α 1-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. Bệnh di truyền này do đột biến làm cho α 1-antiprotease không có cấu trúc thích hợp để được tiết ra ngoài.

Sau khi đã có cấu trúc không gian hoàn thiện, protein được đóng gói trong các túi để vận chuyển đến Golgi. Một số protein có thể được gắn thêm các nhóm đường trước khi bị đưa tiếp đến Golgi. Tuy nhiên phản ứng đường hoá tiếp tục xảy ra ở Golgi.

4.4.3 Quá trình đường hoá protein

Sau khi đã vào trong ER, 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 ER nhờ liên kết với dolichol. Khi phân tử protein đi vào khoang ER, cả chuỗi đường được chuyển sang protein. Phản ứng được thực hiện nhờ enzym oligosaccharyl transferase gắn ở phía trong màng ER. Enzym chuyển oligosaccharide $(\text{Glc})_3(\text{Man})_9(\text{GlcNAc})_2$ 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 $(\text{Glc})_3$ bởi 3 enzym khác nhau. Lúc đó oligosaccharide trên chuỗi polypeptide chỉ còn lại $(\text{Man})_8(\text{GlcNAc})_2$. Đường 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 đến Golgi và Lysosome

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 ER sẽ được vận chuyển đến Golgi nhờ các nang tải có nguồn gốc từ một số vùng đặc biệt trên màng ER.

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 ER 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. Nhờ đó, chúng được "đánh dấu" ở mannose 6 phosphate (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 màng phía *trans*-Golgi nơi có pH=7 đảm bảo duy trì được tương tác giữa thụ thể với M6P. Tiếp đến nang chứa enzym hình thành từ *trans*-Golgi được dung hợp với nang muộn (*late endosome*) có pH khoảng 5,5. Ở trong nang muộn, các enzym tách khỏi thụ thể, nhóm phosphate bị khử khỏi đường mannose nhằm ngăn cản thụ thể tương tác trở lại với enzym. Các enzym được tự do trong nang muộn 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.

4.6 Vận chuyển từ Golgi đến bề mặt tế bào: Con đường tiết ngoại bào (exocytosis)

Trong tế bào vi khuẩn, các protein phân bố trên màng hoặc cần phải tiết ra ngoài thường được tổng hợp dưới dạng preprotein với tín hiệu vận chuyển nằm ở đầu NH₂. Tín hiệu này được gọi là "*leader peptide*", thường không quá 25 acid amin. Có rất nhiều phân tử protein chaperone bám trên chuỗi polypeptide đang được tổng hợp. Chaperon làm nhiệm vụ giữ cho protein không bị gấp khúc đồng thời tham gia vào vận chuyển, tiết protein ra ngoài tế bào. Leader peptide bị cắt khỏi protein khi nó được vận chuyển đến đích.

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*). Có hai con đường vận chuyển hay còn gọi là hai con đường tiết protein khác nhau:

- Vận chuyển liên tục (tiết liên tục): chủ yếu xảy ra với các protein cấu trúc.
- Vận chuyển chọn lọc: thường xảy ra với các protein hoà tan, một số hormon, các chất dẫn truyền thần kinh hoặc các enzym. Thông thường, các protein này chỉ được tổng hợp hoặc giải phóng ra khỏi các nang tiết khi tế bào có yêu cầu.

Đối với các tế bào biệt hoá làm nhiệm vụ tiết, các chất 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.

Bên cạnh nang phủ clathrin làm nhiệm vụ vận chuyển protein từ *trans*-Golgi hoặc từ màng tế bào đến nang muộn, thực nghiệm còn phát hiện được ít nhất hai loại nang phủ khác làm nhiệm vụ chuyên chở protein giữa các bào quan. Nang phủ COPI vận chuyển protein giữa các nếp gấp của thể Golgi theo hướng ngược chiều từ *trans* về *cis* và từ *cis*-Golgi quay trở lại mạng lưới nội chất ER. Nang phủ COPII vận chuyển protein từ ER đi *cis*-Golgi.

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 NH₂ (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.

Chương 5

TRUYỀN TÍN HIỆU TẾ BÀO

Ở cơ thể đa bào, các tín hiệu được truyền đi giữa các tế bào cũng như ngay trong một tế bào. Tín hiệu truyền giữa các tế bào cho phép chúng cùng phối hợp hoạt động, giữ cân bằng cho cơ thể thành một khối thống nhất. Các tín hiệu giúp từng tế bào xác định được vị trí cũng như chức năng chuyên biệt của chúng. Điều đó đảm bảo tế bào chỉ bước vào phân chia khi các tế bào bên cạnh cần đến. Vì bất cứ lý do gì khiến cho sự phối hợp hoạt động giữa chúng bị sai lệch, tế bào sẽ phân chia không tuân theo sự kiểm soát chung và điều đó có thể dẫn đến xuất hiện ung thư giết chết cơ thể.

Các kỹ thuật hiện đại cho phép nghiên cứu quá trình truyền tín hiệu trong một tế bào hoặc giữa chúng với nhau. Mỗi tế bào eukaryot đều có một hệ thống truyền tín hiệu phức tạp nhưng cực kỳ hoàn hảo, bao gồm các protein làm nhiệm vụ truyền và trả lời tín hiệu. Đặc biệt quan trọng là các thụ thể phân bố trên bề mặt màng tế bào và bên trong tế bào (trong tế bào chất và nhân), các protein kinase, các phosphatase, protein Gs vv... Ở tế bào động vật, tín hiệu gửi đi dưới nhiều dạng khác nhau như các phân tử protein, đoạn nhỏ peptide, các acid amin, các nucleotide, các dẫn xuất của axit béo, các steroid, thậm chí các chất khí như oxit nitơ, các monoxit cacbon... Tín hiệu ban đầu kích hoạt một loạt các phản ứng truyền tin nối tiếp hoặc xảy ra đồng thời với nhau. Mọi tín hiệu đầu tiên được gọi chung là **ligand**. Các thụ thể nhận biết ligand rất nhạy cảm, chúng tương tác với ligand ở những nồng độ rất nhỏ (thường $< 10^{-8}$ M). Ngay sau khi tiếp nhận tín hiệu, một loạt các phản ứng truyền tín hiệu xảy ra dẫn đến thay đổi hoạt động của tế bào nhằm trả lời tín hiệu. Trong một số trường hợp khi thụ thể nằm bên trong tế bào, các ligand phải đủ nhỏ và có tính kỵ nước mới có khả năng khuếch tán vượt qua màng vào tế bào chất và tương tác với thụ thể. Trong mối tương tác thụ thể-ligand, vai trò chủ yếu của ligand là làm thay đổi cấu hình của thụ thể thông qua liên kết. Sự thay đổi cấu hình này giống như tín hiệu thông báo cho tế bào cần có phản ứng đối với ligand.

Trong cơ thể đa bào, một tế bào bất kỳ luôn tiếp xúc với hàng trăm tín hiệu khác nhau. Các tín hiệu này có thể là chất hoà tan, hoặc bám vào chất nền hoặc bám trên bề mặt các tế bào lân cận. Chúng có thể kết hợp với nhau cùng hoạt động điều biến phản ứng trả lời tín hiệu. Do đó mỗi tế bào đều có một chương trình riêng biệt để chọn lọc các tín hiệu và phản ứng đáp lại chúng. Các tế bào khác nhau sẽ trả lời khác nhau đối với cùng một tín hiệu.

Tế bào thường phải trả lời cho tập hợp của nhiều chứ không phải từng tín hiệu riêng lẻ. Câu trả lời là kết quả của một chuỗi các phản ứng làm thay đổi hoạt động của tế bào; quyết định tế bào sống hay chết, bước vào phân chia hay tồn tại ở trạng thái tĩnh. Đặc biệt có những protein trong hệ thống truyền tín hiệu có khả năng phân tích và xử lý tín hiệu dẫn đến những hiệu ứng sinh học thích ứng giúp tế bào giữ được trạng thái cân bằng.

Điều đáng lưu ý là các thụ thể nhận biết hormon rất khó xác định và tinh sạch vì nồng độ của chúng rất nhỏ trong tế bào. Số lượng thụ thể đặc hiệu cho một hormon thường dao động từ 10.000 đến 20.000 phân tử trên bề mặt tế bào, tức là chỉ chiếm khoảng 10^{-6} tổng số protein có trong tế bào hoặc cỡ 10^{-4} tổng số protein trên bề mặt màng. Trong thực tế, sự xuất hiện của thụ thể và hàm lượng của chúng được phát hiện và xác định nhờ thí nghiệm chức năng dựa vào tính chất liên kết đặc hiệu giữa chúng với các hormon gắn phóng xạ hoặc chất phát màu. Các hormon này vẫn giữ được hoạt tính tuy được đánh dấu bằng các cách khác nhau.

Các thụ thể hormon nằm trên bề mặt thường được xác định bằng phương pháp đánh dấu ái lực "affinity labeling". Hormon gắn phóng xạ bám vào thụ thể được xử lý bằng các chất hoá học để tạo liên kết cộng hoá trị giữa hormon và thụ thể của nó. Nhờ đó, liên kết này bền vững ngay khi có mặt các chất khử hoặc các chất gây biến tính được sử dụng để tách thụ thể ra khỏi bề mặt tế bào.

Rất nhiều thụ thể phân bố trên bề mặt được tách ra khỏi màng tế bào mà vẫn giữ được hoạt tính nhờ sắc ký đánh dấu ái lực. Tuy nhiên, có những trường hợp khi số lượng thụ thể quá nhỏ thì không thể tinh sạch bằng phương pháp này. Kỹ thuật tách dòng và tái tổ hợp ADN cho phép khắc phục nhược điểm trên. Hình 5.1 minh họa phương pháp tách dòng biểu hiện (*expression cloning*), cho phép thu được số lượng lớn thụ thể của hormon mà không cần phải tinh sạch chúng từ protein tổng số.

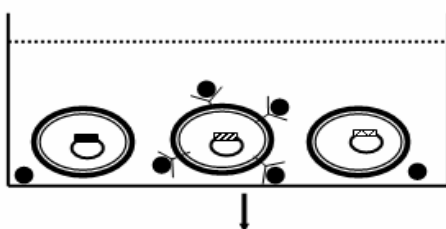
A. Đưa cDNAs vào vector biểu hiện gen



B. Biến nạp vector tái tổ hợp vào tế bào nhận



C. Thụ thể xuất hiện trên bề mặt tế bào nhận vector mang ADN mã cho thụ thể



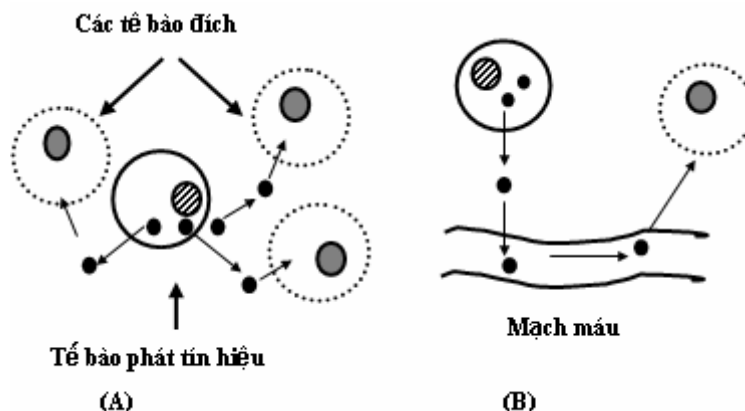
Tách dòng được tế bào mang ADN mã cho thụ thể

Hình 5.1:

Phương pháp tách dòng biểu hiện xác định các ADNc hiếm, mã cho thụ thể bề mặt tế bào. ADNc đưa vào vector biểu hiện mang promoter hoạt động mạnh. Các vector tái tổ hợp được đưa vào dòng tế bào đột biến không tổng hợp được thụ thể. Thụ thể chỉ xuất hiện trên bề mặt tế bào nào nhận vector chứa ADNc mã cho thụ thể và có khả năng liên kết với ligand.

Trong phương pháp này, hormon (*ligand*) chỉ bám vào tế bào mang ADNc mã cho thụ thể. Hormon thường được đánh dấu phát huỳnh quang, nhờ đó phát hiện dễ dàng. Trong cơ thể, hầu hết hormon là những phân tử nhỏ, có thể hoà tan trong lipid hoặc trong nước. Một số ít hormon hoà tan trong lipid có thụ thể nằm ngoài màng tế bào. Rất nhiều hormon hoà tan trong lipid có thụ thể nằm trong tế bào chất. Chúng phải khuếch tán qua màng nguyên sinh chất để tương tác với thụ thể. Ngược lại, hầu hết hormon hoà tan trong nước có thụ thể phân bố trên mặt ngoài của tế bào do hormon không có thể khuếch tán qua màng tế bào.

Các ligand do tế bào tiết ra được dẫn truyền đi theo nhiều cách khác nhau. Tín hiệu có thể chỉ truyền đến các tế bào nằm ngay trong một khu vực cùng với tế bào gửi tín hiệu. Quá trình này được gọi là *paracrine* (Hình 5.2A). Đặc điểm của cách truyền này là tín hiệu không được khuếch tán đi xa. Chúng bị phân hủy bởi các enzym nằm ngoài tế bào hoặc bị giữ lại trong chất nền giữa các tế bào.



Hình 5.2:

Hai con đường truyền tín hiệu thông qua các phân tử tiết

(A)- Truyền tín hiệu paracrine

(B)- Truyền tín hiệu thông qua hệ mạch dẫn. Sự khác nhau giữa chúng thể hiện ở tốc độ truyền cũng như tính đặc hiệu của tế bào nhận.

Trong cơ thể đa bào, để phối hợp nhịp nhàng hoạt động giữa các tế bào nằm cách xa nhau, có những tế bào chuyên hoá giữ vai trò đặc biệt trong việc truyền tín hiệu đi toàn cơ thể. Đó là các tế bào thần kinh (các neuron) với những axon. Khi tiếp nhận tín hiệu từ môi trường hoặc từ các tế bào thần kinh khác, neuron gửi các xung điện dọc theo sợi axon đến xi náp nằm tận cùng sợi axon đó. Xi náp nhận xung điện và tiết ra các tín hiệu hoá học đặc biệt gọi là các chất dẫn truyền thần kinh (*neurotransmitter*). Chất này được gửi đến tế bào nhận rất nhanh và đặc hiệu. Cách thức truyền tín hiệu qua các tế bào thần kinh được đề cập kỹ trong giáo trình thần kinh.

Bên cạnh tế bào thần kinh, còn có các tế bào đặc biệt làm nhiệm vụ gửi các tín hiệu đi kiểm soát hoạt động của cơ thể, đảm bảo sự phối hợp cân bằng nhịp nhàng cho một thực thể thống nhất. Đó là các tế bào nội tiết. Chúng tiết ra hormon gửi vào mạch máu (ở động vật) hoặc vào nhựa cây (ở thực vật), nhờ đó tín hiệu được truyền đi khắp cơ thể (Hình 5.2B). Tuy nhiên tín hiệu do tế bào tiết gửi vào hệ mạch được truyền đi chậm hơn so với tế bào thần kinh. Ngược lại, các hormon tiết lại có tác dụng ngay khi nồng độ của chúng rất thấp ($<10^{-8}$ M) trong khi các chất dẫn truyền thần kinh có hoạt tính ở nồng độ khoảng 10^{-4} M.

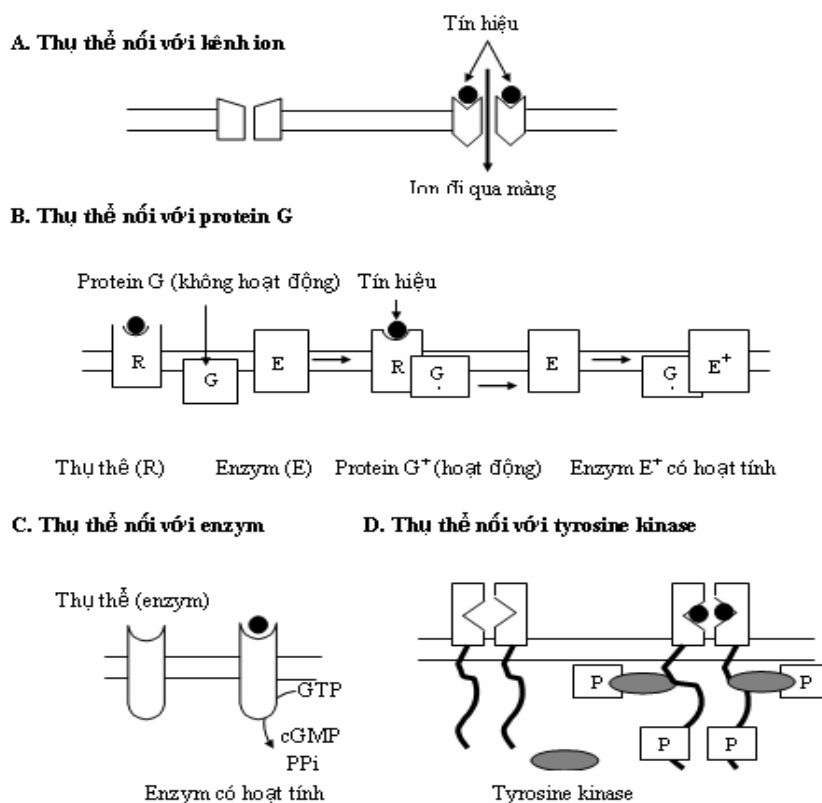
5.1 Thụ thể trên bề mặt tế bào

Có nhiều loại thụ thể khác nhau. Trong chương này chúng ta chú ý đến các thụ thể protein nằm trên bề mặt tế bào và cách thức chúng tham gia chuỗi các phản ứng truyền tín hiệu. Hầu hết ligand tương tác với thụ thể trên bề mặt tế bào đều là protein hoà tan trong nước.

Các protein thụ thể trên bề mặt được xếp vào 4 loại dựa vào cơ chế truyền tín hiệu: thụ thể nối với kênh truyền ion, thụ thể nối với protein G, thụ thể nối với tyrosine kinase và thụ thể nối với enzym (Hình 5.3). Tuy nhiên, hoạt động của chúng xảy ra theo cùng một phương thức: Sau khi tiếp nhận tín hiệu (bên ngoài màng tế bào), thụ thể truyền tín hiệu cho các chất trung gian. Điều đó gây nên sự thay đổi nồng độ của một số chất truyền tín hiệu trung gian (nếu chúng là enzym), hoặc làm thay đổi cấu trúc của chất truyền tín hiệu (nếu như chúng là tyrosine kinase), hoặc làm thay đổi tính thấm ion của màng tế bào (nếu chúng thuộc kênh dẫn ion)... Đến lượt mình, các chất truyền trung gian sẽ làm thay đổi hoạt động của các protein đặc hiệu khác. Cuối cùng tín hiệu được chuyển vào nhân, thường dưới dạng các yếu tố phiên

mã, kiểm soát mức độ biểu hiện của gen. Như vậy, từ một tín hiệu ban đầu, một loạt thay đổi xảy ra giúp tế bào trả lời tín hiệu nhanh, nhạy và chính xác.

a) Thụ thể nối với kênh ion: Các tín hiệu được nhận biết bởi thụ thể nối với kênh ion thường là các chất dẫn truyền thần kinh. Do đó các thụ thể loại này thường nằm trên tế bào thần kinh. Chất dẫn truyền thần kinh mở hoặc đóng tạm thời các kênh ion thông qua việc tương tác với các protein thụ thể nằm ngay trên kênh này.



Hình 5.3:

Bốn loại thụ thể nằm trên bề mặt tế bào nhận các tín hiệu từ bên ngoài

A- Thụ thể nối với kênh ion thường phân bố ngay trên kênh

B-Thụ thể nối với protein G hoạt hoá protein này (G⁺). G⁺ hoạt hoá enzym xúc tác phản ứng tạo chất truyền trung gian

C- Thụ thể nối với tyrosine kinase hoạt hoá kinase. Enzym kinase phosphoryl hoá tyrosine của thụ thể. Chất trung gian tương tác với tyrosine bị phosphoryl hoá và tiếp tục truyền tín hiệu

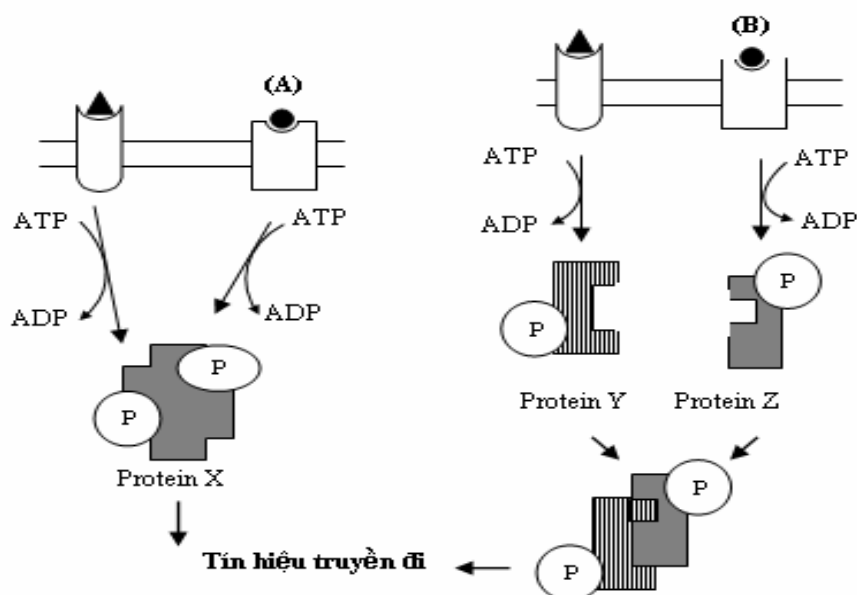
D-Thụ thể nối với enzym xúc tác cho phản ứng truyền tín hiệu trung gian (theo Alberts & cs., 2002).

b) Thụ thể nối với protein G: Sau khi nhận được tín hiệu, các thụ thể loại này sẽ truyền tiếp tín hiệu cho các chất trung gian khác thông qua protein G (*trimeric GTP binding regulatory protein*). Cấu trúc và hoạt tính của protein G được nêu trong phần 4.2. Mọi thụ thể nối với protein G có cấu trúc khá giống nhau và chúng đều nằm vắt qua màng 7 lần.

c) Thụ thể nối với enzym: Khi nhận tín hiệu, các thụ thể loại này hoặc hoạt động giống enzym hoặc kết hợp cùng enzym. Hầu hết chúng chỉ vắt qua màng tế bào một lần và có vị trí liên kết với tín hiệu ở bên ngoài màng, còn vị trí có hoạt tính enzym nằm bên trong màng. Các thụ thể loại này không giống nhau và chủ yếu là các kinase hoặc protein kết hợp với kinase.

d) Thụ thể nối với tyrosine kinase: Tương tác giữa thụ thể với ligand dẫn đến việc tạo dimer của thụ thể. Sự thay đổi cấu trúc của thụ thể sẽ hoạt hoá tyrosine kinase nằm trong tế

bào chất. Enzym này gắn gốc phosphat vào tyrosine của thụ thể. Chất trung gian liên kết với tyrosine có gốc phosphat và tiếp tục bị phosphoryl hoá bởi các kinase khác hoặc bởi chính mình. Nhờ đó tín hiệu tiếp tục truyền đi.



Hình 5.4:

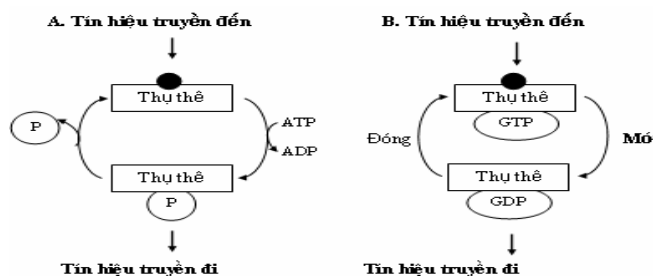
Phối hợp các tín hiệu trong con đường truyền tín hiệu

A - Hai tín hiệu dẫn đến phản ứng phosphoryl hóa protein X ở các vị trí khác nhau. Chỉ khi cả hai vị trí trên X được gắn gốc phosphate thì X mới có hoạt tính. Như vậy bắt buộc phải có đồng thời hai tín hiệu thì protein X mới chuyển trạng thái hoạt động

B - Hai tín hiệu hoạt hoá đồng thời hai protein Y và Z để tạo ra phức có hoạt tính.

Chuỗi các phản ứng phosphoryl hoá bắt đầu bằng thụ thể nối tyrosine kinase được thực hiện chủ yếu bởi các kinase serine/threonine hoặc tyrosine (Hình 5.4). Đây là các enzym đặc biệt, xúc tác cho phản ứng gắn nhóm phosphate vào các acid amin serine, threonine (ít hơn) hoặc tyrosine phân bố trên phân tử protein. Các nhà nghiên cứu đã ước tính có khoảng 1% gen của cơ thể người mã cho kinase. Một tế bào động vật có vú chứa khoảng 100 loại enzym kinase khác nhau chủ yếu là serine/threonine kinase. Mặc dù số protein bị phosphoryl hoá bởi tyrosine kinase chỉ chiếm 0,1% tổng số protein trong tế bào, tyrosine kinase đóng vai trò then chốt trong hệ thống truyền tín hiệu.

Tín hiệu thu nhận bởi hai loại thụ thể nối với protein G và nối tyrosine kinase thường được truyền qua một số chất chuyển tải trung gian và cuối cùng là đến các yếu tố phiên mã. Các yếu tố đó sẽ điều biến mức độ biểu hiện của gen sao cho tế bào có phản ứng tương thích với tín hiệu. Một số protein được biệt hoá chỉ hoạt động trong các hệ thống truyền tín hiệu. Chúng thường tồn tại ở hai trạng thái: không hoạt tính (trước khi nhận tín hiệu) và có hoạt tính (sau khi nhận tín hiệu). Khi nhận tín hiệu chúng chuyển sang trạng thái hoạt động. Ví dụ, phản ứng phosphoryl hoá và mức độ phosphoryl hoá (nhận một hoặc nhiều nhóm phosphate) xảy ra với tyrosine kinase liên quan mật thiết đến biểu hiện hoạt tính của protein này. Ở trạng thái hoạt động chúng truyền tín hiệu đi bằng cách gây phosphoryl hoá các protein tiếp theo, tạo nên chuỗi các phản ứng phosphoryl hoá. Sau đó chúng trở về trạng thái không hoạt động nhờ phản ứng khử nhóm phosphate. Đối với, thụ thể liên kết với GTP, chúng truyền tín hiệu cho các phần tử trung gian và trở về dạng không có hoạt tính khi liên kết với GDP. Như vậy, chức năng nhận biết và truyền tín hiệu của cả hai loại thụ thể đều phụ thuộc vào sự thay đổi cấu hình của chúng thông qua phản ứng phosphoryl hoá hoặc nhờ tương tác với GTP/GDP (Hình 5.5).



Hình 5.5:

Hai cơ chế chủ đạo truyền tín hiệu trong tế bào. Đối với cả hai trường hợp, protein truyền tín hiệu được hoạt hoá bởi phản ứng phosphoryl hoá (hoặc khử gốc phosphate). (A): Truyền tín hiệu thông qua các protein kinase. (B): Truyền tín hiệu thông qua protein G.

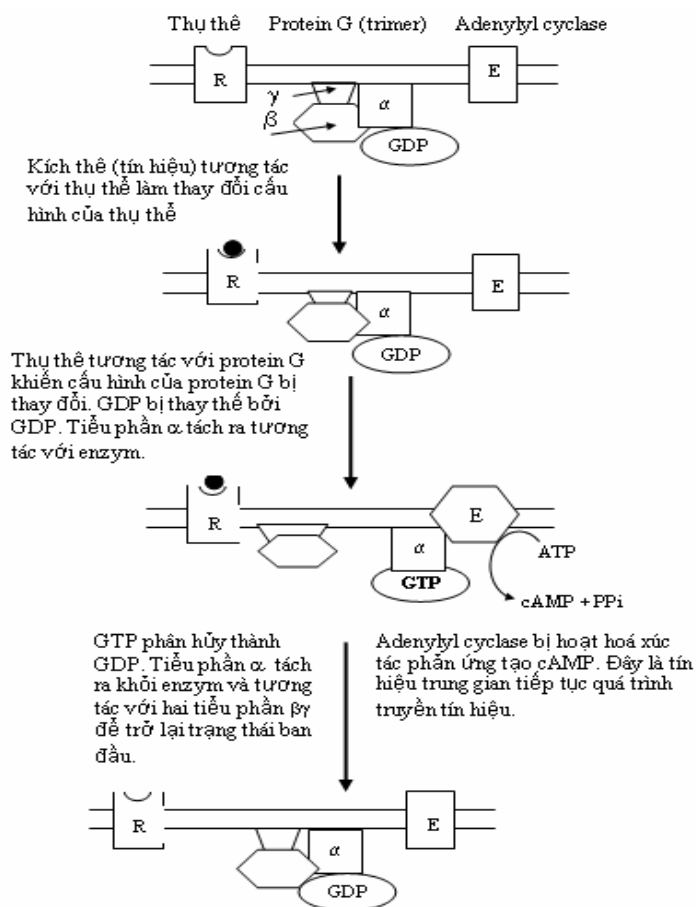
5.2 Thụ thể nối với protein G

Hơn 100 thụ thể nối với protein G được phát hiện ở động vật có vú, phần lớn nhờ kỹ thuật tách dòng tương đồng (*homology cloning*) dựa vào phản ứng lai ở điều kiện không quá nghiêm ngặt giữa môi ADNc đã biết và ADN của đối tượng cần nghiên cứu. Ngoài ra các thụ thể còn được tìm thấy nhờ kỹ thuật tách dòng biểu hiện dựa vào các tín hiệu xuất hiện khi tế bào bị kích thích. Trong phương pháp này, mô có chứa thụ thể cần nghiên cứu được sử dụng để tách ARNm và tổng hợp ADNc, xây dựng ngân hàng ADNc trong vector biểu hiện gen (vector có chứa promoter). Sau đó, các dòng của ngân hàng được đưa vào trứng ếch *Xenopus* để tổng hợp nên các protein tương ứng với ADNc. Những protein lạ được gắn vào màng tế bào và được phát hiện nhờ các protein được đánh dấu có khả năng liên kết với chúng.

5.2.1 Protein G

Protein G (*trimeric GTP binding protein*) có hoạt tính phân hủy GTP (Hình 5.6). Chúng luôn phối hợp hoạt động cùng với các thụ thể nằm trên bề mặt tế bào tạo thành một cặp truyền tín hiệu. Tín hiệu được truyền qua protein G vào trong tế bào thông qua các enzym hoặc các kênh ion trên màng nguyên sinh chất. Protein G khác với monomeric GTPase ở trong tế bào eukaryot. Monomeric GTPase tham gia kiểm soát các phản ứng sinh học và vận chuyển túi tiết. Tuy nhiên, protein G cũng có hoạt tính GTPase. Nó ở trạng thái hoạt động khi liên kết với GTP nhưng bị mất hoạt tính khi liên kết với GDP. Khi chưa có tín hiệu, các thụ thể không tương tác với protein G.

Đồng thời lúc đó bản thân protein G ở trạng thái liên kết với GDP (tức là ở dạng không hoạt tính). Khi thụ thể nhận tín hiệu, cấu trúc không gian của thụ thể bị thay đổi. Phức thụ thể - ligand sẽ tương tác với protein G khiến GDP bị tách ra và GTP thay vào đó. Lập tức protein G tương tác với chất trung gian tiếp theo (enzym hoặc protein). Liên kết giữa protein G và chất nhận tín hiệu tiếp theo khiến cho GTP bị phân hủy thành GDP bởi chính hoạt tính GTPase của protein G. Phản ứng phân hủy GTP khiến protein G tách rời khỏi chất trung gian. Protein G trở về trạng thái ban đầu (mang GDP). Chất trung gian truyền tín hiệu đi.



Hình 5.6:
Mô hình hoạt động của protein G trong quá trình truyền tín hiệu
Protein G duy trì ở trạng thái hoạt hoá một khi ligand vẫn bám vào thụ thể. Vì vậy

Protein G gồm ba chuỗi polypeptide khác nhau: α , β và γ . Chuỗi α (α s) có hoạt tính GTPase. Chuỗi này sẽ liên kết với GTP và thủy phân GTP gây hoạt hoá adenylyl cyclase. Chuỗi β và γ tạo phức $\beta\gamma$ giữ cho protein G bám trên bề mặt tế bào bên phía tế bào chất. Khi GDP liên kết với chuỗi α , protein G không có hoạt tính. Khi GDP bị thay thế bởi GTP, chuỗi α tách ra khỏi phức và liên kết với adenylyl cyclase. Enzym này được hoạt hoá và đến lượt nó xúc tác cho phản ứng tạo cAMP. Như vậy, protein G phân rã thành các tiểu đơn vị α và $\beta\gamma$ khi được hoạt hoá.

Có thể có nhiều loại thụ thể nối protein G nhưng tất cả chỉ nhận biết một loại tín hiệu. Ngược lại, có thể có nhiều thụ thể nối protein G, chúng nhận biết các tín hiệu khác nhau. Hơn nữa, từ protein G, mỗi tín hiệu có thể được truyền theo con đường riêng biệt khiến tế bào có câu trả lời đặc hiệu với từng loại tín hiệu. Hầu hết các thụ thể nối protein G hoạt hoá cho một chuỗi các phản ứng làm thay đổi nồng độ của một hoặc vài phân tử nhỏ trong tế bào. Những phân tử này làm nhiệm vụ trung chuyên và được gọi chung là các chất truyền tín hiệu trung gian (hoặc còn gọi là các tín hiệu thứ cấp). Hai trong số các chất trung gian hay gặp nhất trong tế bào là cAMP và ion Ca^{+2} . Nồng độ của chúng biến đổi do nhiều tín hiệu khác nhau, trong đó hầu hết là tín hiệu nhận biết bởi các thụ thể nối protein G.

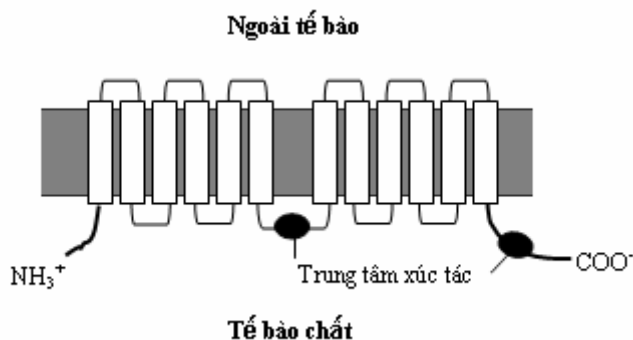
Sở dĩ tế bào trả lời rất nhanh, nhạy đối với sự biến động về liều lượng của các tín hiệu bên ngoài là do hoạt tính thuận nghịch của adenylyl cyclase. Điều này xảy ra nhờ chuỗi α s chỉ tồn tại ở trạng thái hoạt động trong thời gian rất ngắn. Khi α s liên kết với adenylyl cyclase,

hoạt tính GTPase của chuỗi α s được kích thích làm cho GTP bị phân hủy thành GDP rất nhanh. Chuỗi α s tách khỏi adenylyl cyclase. Lúc này cả α s và adenylyl cyclase bị mất hoạt tính. Chuỗi α s liên kết trở lại với phức $\beta\gamma$ tạo thành phân tử protein G ở dạng không hoạt động. Trong tế bào động vật, có ít nhất 6 loại đồng phân adenylyl cyclase đều phân bố trên màng tế bào phối hợp với các thụ thể nhận những tín hiệu khác nhau. Việc chuyển đổi trạng thái từ hoạt hoá sang không có hoạt tính của những enzym này đều phụ thuộc vào protein G. Đặc biệt trên màng tế bào thần kinh, enzym này được hoạt hoá bởi ion Ca^{+2} nằm trong phức với calmodulin.

5.2.2 Hoạt hoá hoặc ức chế cAMPase thông qua protein G

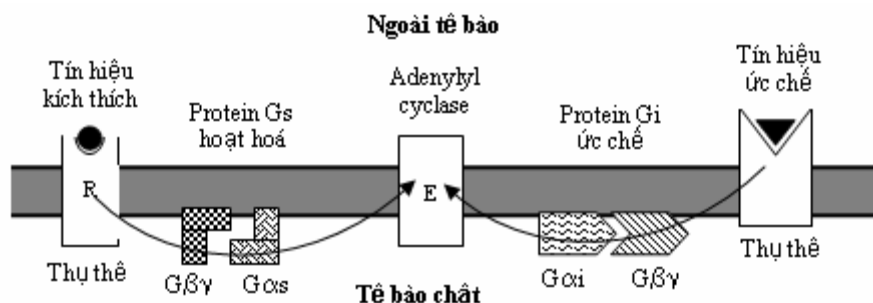
Từ năm 1959, phân tử AMP dạng vòng ký hiệu là cAMP (*cyclic adenosine-3',5'-monophosphate cAMP*) được xác định như là chất truyền tín hiệu trung gian trong tế bào. Từ đó đến nay, thực nghiệm nhận thấy cAMP tham gia vào nhiều quá trình sinh hoá khác nhau trong tế bào prokaryot cũng như tế bào động vật. Nồng độ của cAMP trong tế bào eukaryot nhỏ hơn 10^{-7} M. Khi bị kích thích, nồng độ này tăng gấp 5 lần trong một giây. Điều đó đòi hỏi phản ứng tổng hợp cAMP phải xảy ra rất nhanh, đồng thời phải được dừng tức thời khi không còn tín hiệu kích thích. Phân tử cAMP được tổng hợp từ ATP nhờ enzym adenylyl cyclase (Hình 5.7). Mặt khác, cAMP bị phân hủy liên tục nhờ cAMP-phosphodiesterase. Nhiều tín hiệu bên ngoài tham gia kiểm soát nồng độ cAMP thông qua việc điều biến hoạt tính của adenylyl cyclase hơn là thay đổi hoạt tính của cAMP-phosphodiesterase. Trong một loại tế bào nhất định, mọi tín hiệu gây hoạt hoá adenylyl cyclase thường có chung hiệu ứng. Ví dụ, trong các tế bào mỡ ít nhất có 4 loại hormon hoạt hoá adenylyl cyclase cùng dẫn đến việc tăng cường chuyển hoá triglyceride (dạng dự trữ của chất béo) thành acid béo. Các thụ thể cho những hormon này hoạt động khi liên kết với protein G. Cá thể mắc bệnh suy giảm di truyền trong việc tổng hợp protein G thường rối loạn chuyển hoá, phát triển xương và phát triển trí não bất bình thường.

Thí nghiệm minh họa rõ nhất về sự tạo cặp cùng hoạt động giữa thụ thể và adenylyl cyclase là các thụ thể β -adrenergic (liên quan đến hoạt động của *adrenaline* và *noradrenaline*). Khi cơ thể hoạt động liên tục hoặc chịu sự căng thẳng, các mô đòi hỏi bổ sung glucose và acid béo. Các nguyên liệu này được tế bào gan và các tế bào dự trữ chất béo (*lipolysis*) cung cấp vào máu rất nhanh khi các tín hiệu *adrenaline* và *noradrenaline* tương tác với thụ thể β -adrenergic nằm trên bề mặt các tế bào này. Ngoài ra, liên kết của *adrenaline* với thụ thể β -adrenergic ở các tế bào cơ tim làm tăng tần số co bóp tim, dẫn đến tăng lượng máu đi tới các mô. Liên kết giữa *adrenaline* với thụ thể ở tế bào cơ trơn của ruột gây co giãn cơ. *Adrenaline* còn có khả năng liên kết với thụ thể α -adrenergic khiến các động mạch co thắt lại, ngăn cản máu lưu thông đến các vùng lân cận (nơi không có tương tác giữa *adrenaline* và thụ thể). Những tác dụng khác nhau của hormon *adrenaline* đều nhằm mục đích cuối cùng là bổ sung năng lượng cho các cử động nhanh của cơ trong những hoạt động căng thẳng mà cơ thể đang chịu đựng.

**Hình 5.7:**

Cấu trúc của adenylyl cyclase gồm 2 tiểu phần phân bố trong màng tế bào, mỗi tiểu phần gồm 6 đoạn α helix. Hai trung tâm hoạt động của enzym nằm phía trong tế bào chất sẽ phân hủy cơ chất khi enzym nhận tín hiệu truyền từ protein G.

Khi adrenaline liên kết với thụ thể β -adrenergic, nó hoạt hoá adenylyl cyclase. Tuy nhiên, khi adrenaline liên kết với α_2 -adrenergic, nó ức chế enzym. Điều đó cho thấy các protein G khác nhau có tác dụng khác nhau (Hình 5.8).

**Hình 5.8:**

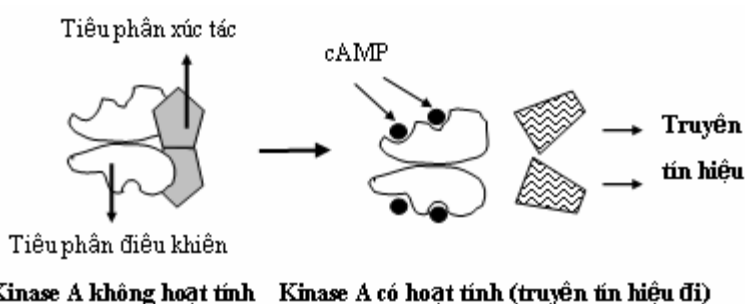
Các loại protein G tham gia hoạt hoá (*protein Gs*) hoặc ức chế (*protein Gi*) enzym adenylyl cyclase. Các tiểu phần $G\beta\gamma$ là như nhau trong các loại protein này. Chúng chỉ khác nhau ở tiểu phần $G\alpha$ và thụ thể mà chúng tương tác (theo Alberts & cs., 2002).

Adenylyl cyclase có thể được hoạt hoá hoặc bị ức chế thông qua các protein G khác nhau. Trong phản ứng hoạt hoá, thụ thể β -adrenergic cùng tạo cặp hoạt động với G_s . Trong phản ứng ức chế, thụ thể tạo cặp với protein G_i . Protein G_i có cấu trúc tương tự G_s chỉ khác ở tiểu phần α . Ví dụ, tương tác giữa adenosine và prostaglandin với thụ thể của chúng sẽ kích hoạt protein G_i . Lúc đó, tiểu phần $G\alpha_i$ tương tác với GTP và tách khỏi protein G_i . Phức $G\alpha_i$ -GTP ức chế (chứ không phải là hoạt hoá) adenylyl cyclase khiến nồng độ cAMP giảm xuống. Ngoài ra, không phải chỉ riêng $G\alpha_i$ mà cả $G\beta\gamma$ cũng có tác dụng ức chế adenylyl cyclase. Tiểu phần $G\alpha_i$ ức chế theo con đường gián tiếp trong khi phức $G\beta\gamma$ ức chế theo hai cách: hoặc là phức $G\beta\gamma$ liên kết với enzym hoặc liên kết với bất kỳ tiểu phần α s tự do có trong tế bào, do đó hạn chế chúng hoạt hoá enzym. Ngoài ra protein G_i còn kích thích việc mở kênh K^+ trên màng tế bào. Như vậy các protein G rất linh hoạt trong việc truyền tín hiệu, thực hiện nhiệm vụ trung gian giữa thụ thể và các tín hiệu thứ cấp trong tế bào. Ở các thí nghiệm đã tiến hành, hoặc tiểu phần $G\alpha$ hoặc phức $G\beta\gamma$ hoặc cả hai đều tham gia phản ứng. Tuy nhiên, phức $G\beta\gamma$ làm tăng hoạt tính của một số dạng adenylyl cyclase chỉ sau khi enzym đã được hoạt hoá bởi tiểu phần $G\alpha_s$.

5.3 Protein kinase phụ thuộc cAMP (cAPK hoặc kinase A)

Trong các tế bào động vật, chức năng chủ yếu của cAMP là hoạt hoá enzym kinase. Những enzym này được gọi chung là cAPK (*cAMP-dependent protein kinase*), hoặc protein kinase A. Các cAPK xúc tác cho phản ứng chuyển nhóm phosphate tận cùng từ ATP sang acid amin serine hoặc threonine của protein đặc hiệu. Khi bị phosphoryl hoá, hoạt tính của protein bị thay đổi (từ trạng thái hoạt động sang trạng thái bất hoạt hoặc ngược lại). Protein kinase A được tìm thấy trong mọi tế bào động vật và được xem như liên quan đến hầu hết các chức năng của cAMP. Cơ chất của kinase A đa dạng, phụ thuộc vào từng loại tế bào. Điều đó giải thích vì sao tác dụng của cAMP cũng thay đổi theo từng loại tế bào.

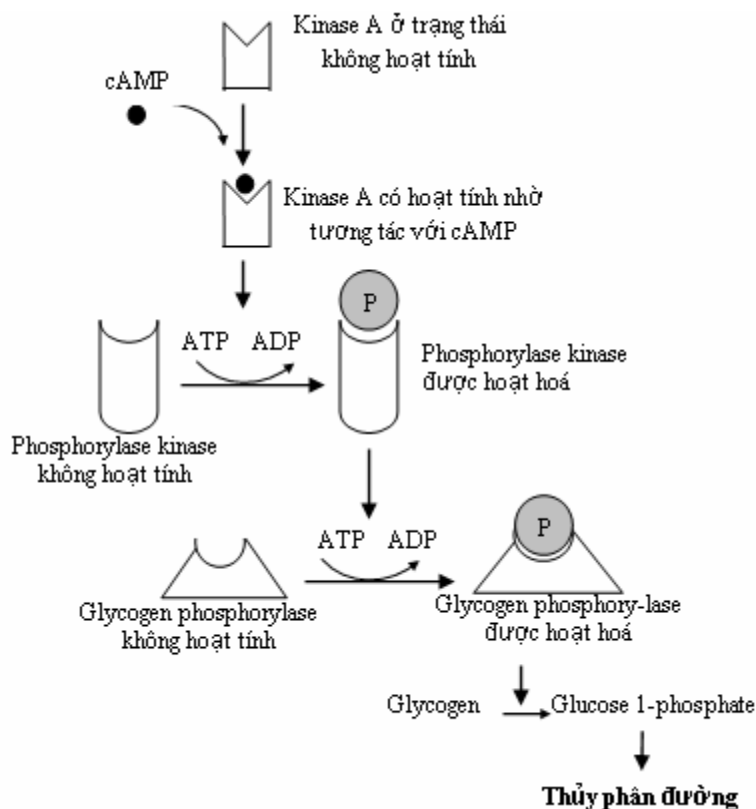
Ở trạng thái không hoạt động, kinase A là phức tetramer gồm hai tiểu phần có hoạt tính enzym và hai tiểu phần kiểm soát việc liên kết với cAMP. Liên kết giữa hai tiểu phần này với cAMP làm thay đổi cấu trúc không gian của hai tiểu phần kia, khiến chúng tách khỏi phức hệ (4 tiểu phần). Lúc đó hai tiểu phần có hoạt tính enzym được hoạt hoá tham gia xúc tác cho phản ứng phosphoryl hoá protein đặc hiệu (Hình 5.9).



Hình 5.9:

Hoạt hoá kinase A bởi cAMP. Enzym kinase A có cấu trúc tetramer. Liên kết giữa cAMP với hai tiểu phần điều khiển làm thay đổi cấu trúc không gian của kinase A, khiến hai tiểu phần có hoạt tính enzym tách ra và được hoạt hoá. Mỗi tiểu phần điều khiển có hai vị trí liên kết với cAMP. Khi cả hai vị trí đều liên kết với cAMP thì hai tiểu phần kia mới tách ra và xúc tác cho phản ứng phosphoryl hoá. Có ít nhất hai loại kinase A trong tế bào động vật có vú. Loại thứ nhất tồn tại chủ yếu trong tế bào chất. Loại thứ hai, dính vào màng sinh chất, màng nhân và các sợi vi ống (microtubulin) nhờ hai tiểu phần điều khiển. Đối với cả hai loại kinase A, khi các tiểu phần có hoạt tính được hoạt hoá, chúng có thể di chuyển vào nhân gây phosphoryl hoá các protein điều khiển hoạt động của gen trong khi hai tiểu phần kia vẫn tồn tại trong tế bào chất.

Phản ứng phosphoryl hoá xảy ra nhờ vai trò trung gian của cAMP được phát hiện lần đầu tiên khi nghiên cứu quá trình trao đổi glycogen trong tế bào cơ xương. Glycogen là dạng dự trữ của glucose. Cả hai quá trình tổng hợp và phân hủy glycogen được kiểm soát bởi adrenaline. Khi động vật bị đánh hoặc gặp bất kỳ tổn thương nào, tuyến adrenaline tiết hormon vào máu gây tín hiệu báo động đến các tổ chức khác trong cơ thể. Adrenaline tuần hoàn trong máu, kích thích các tế bào cơ phân giải glycogen thành glucose, đồng thời gây ngừng quá trình tổng hợp glycogen. Glucose bị oxy hóa tạo ATP cung cấp cho cơ. Bằng cách này, adrenaline chuẩn bị cho tế bào cơ tham gia hoạt động căng thẳng. Hoạt tính của adrenaline được thể hiện khi nó liên kết với thụ thể β -adrenergic ở trên bề mặt tế bào cơ, gây ra sự tăng nồng độ cAMP trong tế bào chất. Lúc đó cAMP sẽ hoạt hoá kinase A và enzym này phosphoryl hoá hai enzym khác (Hình 5.10).



Hình 5.10:

Phân tử cAMP kích thích chuyển hoá glycogen thành glucose trong tế bào cơ. Nhờ liên kết với cAMP, kinase A được hoạt hoá, xúc tác cho phản ứng phosphoryl hoá và dẫn đến hoạt hoá hai enzym phosphorylase kinase và glycogen phosphorylase. Nhờ đó, glycogen phân giải thành glucose. Kinase A trực tiếp và gián tiếp làm tăng quá trình phosphoryl hoá enzym glycogen synthase. Do bị phosphoryl hoá, glycogen synthase bị mất hoạt tính, dẫn đến ức chế tổng hợp glycogen (theo Alberts & cs., 2002).

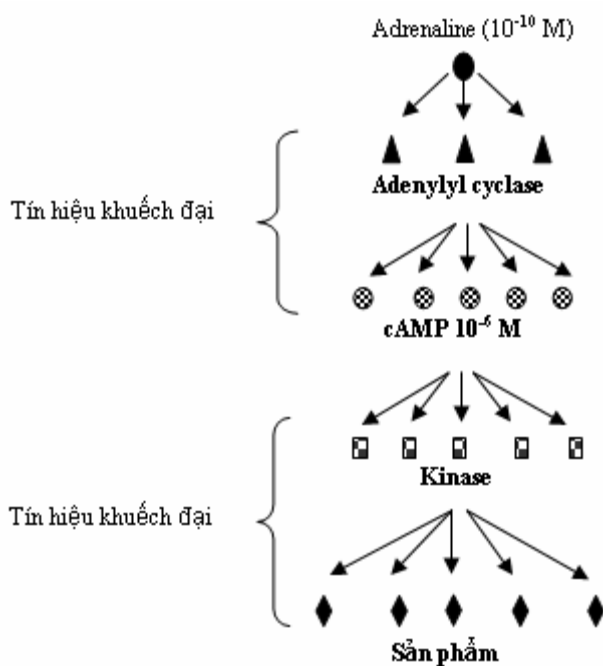
Như vậy, sự tăng nồng độ cAMP dẫn đến một loạt phản ứng vừa kích thích chuyển hoá glycogen thành glucose vừa ức chế quá trình chuyển glucose thành glycogen đảm bảo cho nồng độ glucose đạt cực đại trong tế bào. Tổng hợp và phân hủy glycogen phụ thuộc nồng độ cAMP xảy ra chủ yếu ở tế bào gan và cơ, nơi dự trữ glycogen. Tuy nhiên, cAMP còn tham gia truyền tin khi có mặt của một số hormon. Ví dụ, trong tế bào tạo mỡ, nồng độ cAMP tăng cao sẽ kích hoạt lipase thông qua cAPK, do đó tổng hợp acid béo được tăng cường. Một cách tương tự, trong tế bào buồng trứng, cAMP tăng khiến cho cAPK kích thích phản ứng tổng hợp 2 hormon estradiol và progesterone. Ngoài ra, các cAPK trong tế bào thần kinh tham gia điều biến hoạt động của các kênh ion liên quan đến việc hình thành trí nhớ trong quá trình phát triển và biệt hoá. Các enzym cAPK thường được cố định tại một vùng trong tế bào nhờ protein liên kết với cAPK. Những protein này thường có hai phần, một phần liên kết với cAPK, phần kia cố định lại tại một vị trí tương tự như cái “mỏ neo” để giữ cAPK không khuếch tán trong tế bào chất.

Trong một số tế bào động vật, nồng độ cAMP tăng sẽ kích thích hoạt động của một số gen đặc biệt. Ví dụ, trong tế bào tiết hormon somatostatin, nồng độ cAMP đạt đến mức nhất định sẽ bật mở gen mã cho hormon này. Quá trình hoạt hoá gen mã cho somatostatin được thực hiện như sau: Gen có chứa một đoạn ADN ngắn, gọi là vùng CRE (cAMP response element) là vị trí tương tác của protein điều khiển, gọi là protein CREB (CRE-binding protein). Tuy nhiên protein CREB ở dạng bị phosphoryl hoá mới liên kết được với CRE. Khi bị phosphoryl hoá bởi kinase A, protein CREB liên kết với vùng CRE, hoạt hoá gen mã cho hormon. Từ ví dụ cụ thể của kinase A, chúng ta thấy một chuỗi các phản ứng phosphoryl hoá

cũng như khử phosphoryl xảy ra kế tiếp nhau. Trong chuỗi phản ứng này, kết quả của mỗi một phản ứng là sự hoạt hoá hoặc ức chế protein đặc hiệu nhờ sản phẩm của phản ứng xảy ra trước đó.

Chuỗi các phản ứng dây chuyền thường được điều khiển bởi nồng độ của một loại phân tử ban đầu, ví dụ bắt đầu bởi cAMP. Mặt khác, các phản ứng có thể xảy ra thuận nghịch khi nồng độ cAMP thay đổi. Hơn nữa, chuỗi các phản ứng còn cho phép khuếch đại tín hiệu nhỏ bé ban đầu (ở nồng độ rất nhỏ). Ví dụ, nồng độ của adrenaline cần thiết trong máu để kích thích phản ứng phân giải glycogen khoảng 10^{-10} M. Tương ứng với lượng adrenaline rất ít này, nồng độ cAMP khoảng 10^{-6} M. Sau ba phản ứng liên tiếp xảy ra, nồng độ glucose tăng 50% (Hình 5.11). Tỷ lệ nồng độ giữa ba loại enzym kinase, glycogen phosphorylase kinase và glycogen phosphorylase đạt được sau chuỗi phản ứng là 1: 10 : 240. Như vậy tác dụng hoạt hoá của adrenaline và cAMP đã được khuếch đại lên rất nhiều lần.

Hoạt tính của cAMP chỉ có tác dụng tạm thời. Tế bào nhất định phải khử gốc phosphate của những protein đã bị phosphoryl hoá bởi kinase A. Nhìn chung, phản ứng khử phosphate cho những protein bị phosphoryl hoá ở các acid amin serine hoặc threonine được xúc tác nhờ các enzym serine/threonine phosphatase. Các enzym này được xếp vào 4 nhóm: I; IIA; IIB và IIC. Enzym nhóm I giữ vai trò quan trọng trong phản ứng trả lời đối với thay đổi nồng độ cAMP. Enzym nhóm IIA xúc tác cho rất nhiều phản ứng khử nhóm phosphate do các enzym kinase serine/threonine gây ra. Nó giữ vai trò quan trọng trong kiểm soát chu trình tế bào. Enzym nhóm IIB được hoạt hoá nhờ Ca^{+2} và xuất hiện nhiều trong não.

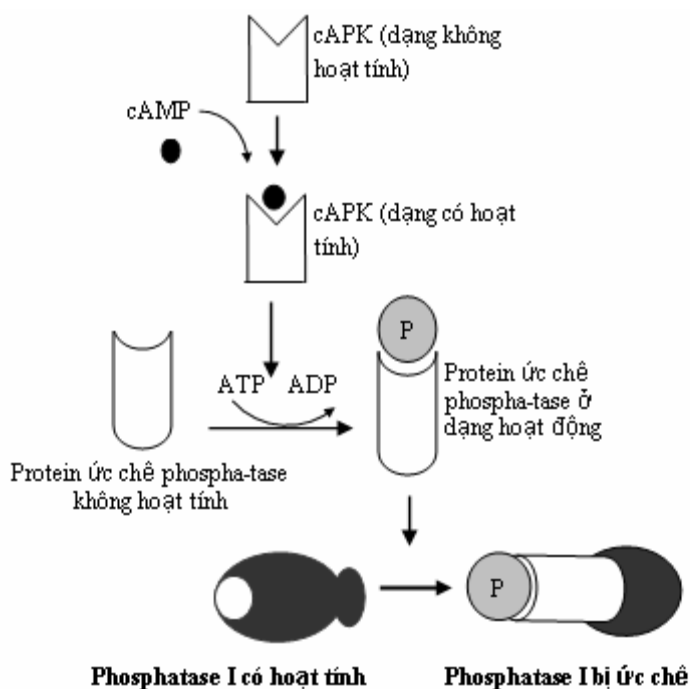


Hình 5.11:

Truyền và khuếch đại tín hiệu từ bên ngoài tế bào qua thụ thể nối protein G
Phân tử adrenaline tương tác với thụ thể làm nồng độ cAMP tăng và hoạt hoá các enzym trong chuỗi các phản ứng dây chuyền. Nhờ vậy tín hiệu được khuếch đại.

Hoạt tính của bất kỳ protein nào liên quan đến phản ứng phosphoryl hoá đều phụ thuộc vào sự cân bằng giữa hoạt tính của kinase và phosphatase. Phosphatase I khử phosphate ở những protein bị phosphoryl hoá bởi kinase A. Ví dụ, hoạt tính của CREB bị ức chế do enzym nhóm I khử nhóm phosphate, vì vậy hoạt động của gen điều khiển bởi CREB bị ngừng do CERB mất khả năng liên kết với CRE. Trong các tế bào cơ bị kích thích bởi adrenaline, hoạt

tính của phosphatase I bị giảm do protein ức chế. Khi kinase A xúc tác phản ứng phosphoryl hoá đối với protein ức chế, protein này liên kết với phosphatase I và làm enzym mất hoạt tính (Hình 5.12).



Hình 5.12:

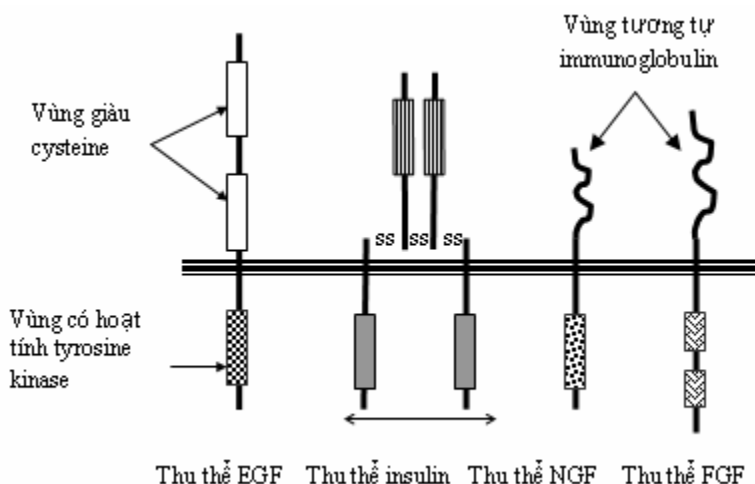
Vai trò điều khiển của phosphatase I trong phản ứng tạo glycogen thông qua cAMP. Phosphatase I bị ức chế bởi cAMP (trái ngược với phản ứng phosphoryl hoá được hoạt hoá bởi cAMP). cAMP hoạt hoá kinase A (cAPK) gây phosphoryl hoá protein ức chế phosphatase. Nhờ đó, protein ức chế có hoạt tính và đến lượt nó kìm hãm hoạt động của phosphatase I.

5.4 Thụ thể tyrosine kinase và các protein Ras

Hầu hết thụ thể tyrosine kinase nhận biết các factor tăng trưởng và biệt hoá tế bào. Ví dụ, thụ thể của factor tăng trưởng xuất phát từ tiểu cầu (*PDGF-platelet derived growth factor*), từ sợi nguyên bào fibroblast (*FGFs-fibroblast growth factor*), từ tế bào biểu bì (*EGF-epidermal growth factor*) vv... Tương tác giữa tín hiệu ban đầu (*ligand*) với thụ thể tyrosine kinase (RTKs) dẫn đến phản ứng phosphoryl hoá acid amin tyrosine và tiếp sau đó là một loạt các phản ứng phosphoryl hoá khác trong chuỗi truyền tín hiệu. Chúng ta cùng xem xét cách thức hoạt động của những thụ thể này và những protein nhận tín hiệu từ chúng để tiếp tục truyền tin.

5.4.1 Thụ thể tyrosine kinase (RTKs)

Tất cả các thụ thể RTKs đều gồm hai phần. Phần thứ nhất phân bố trên bề mặt tế bào làm nhiệm vụ tương tác đặc hiệu với tín hiệu (*ligand*). Phần thứ hai nằm trong tế bào chất có hoạt tính tyrosine kinase. Phần thứ hai này xúc tác phản ứng phân hủy ATP thành ADP và gắn gốc phosphate vào tyrosine. Hai phần của RTKs được nối với nhau bởi đoạn peptide helix α kỵ nước. Cấu trúc của một số loại thụ thể tyrosine kinase được trình bày trên hình 5.13.

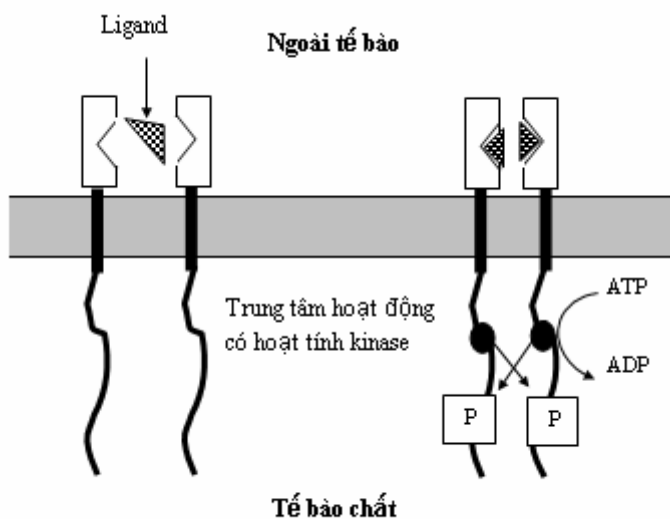


Hình 5.13:

Các thụ thể tyrosine kinase

Chức năng của vùng giàu cysteine và vùng tương tự immunoglobulin chưa được xác định. EGF – yếu tố tăng trưởng biểu bì (*epidermal growth factor*); NGF – yếu tố tăng trưởng thần kinh (*nerve growth factor*); FGF – yếu tố tăng trưởng nguyên bào (*fibroblast growth factor*).

Hầu hết thụ thể RTKs tồn tại ở dạng monomer khi chưa có tín hiệu. Tương tác với ligand làm thay đổi cấu trúc của thụ thể khiến chuỗi đơn monomer RTKs chuyển sang cấu trúc dimer. Tiếp đó, cấu trúc dimer khiến trung tâm xúc tác của phần nằm trong tế bào chất chuyển sang trạng thái có hoạt tính kinase. Hoạt tính kinase của chuỗi monomer này sẽ gắn gốc phosphate vào tyrosine của chuỗi monomer kia. Như thế, mỗi chuỗi đơn RTK sẽ chỉ có hoạt tính kinase đối với tyrosine của sợi đơn kia sau khi hai sợi đó tương tác với nhau tạo cấu trúc dimer. Đây là phản ứng tự phosphoryl hoá.



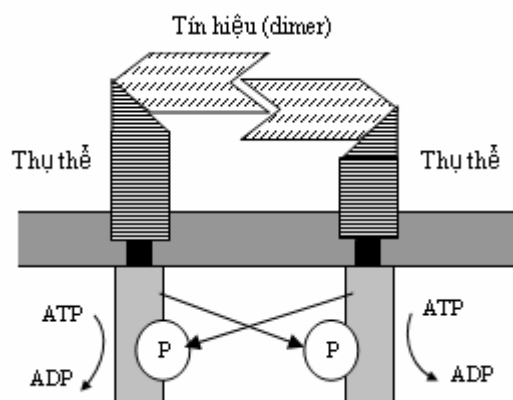
Hình 5.14:

hay đổi cấu trúc của thụ thể tyrosine kinase trước và sau khi tương tác với tín hiệu đầu tiên (*ligand*). Thụ thể (dạng *monomer*) tương tác với ligand sẽ tạo dimer. Hoạt tính kinase của mỗi sợi monomer sẽ phân hủy ATP và gắn gốc phosphate vào tyrosine trên sợi monomer kia. Đây là phản ứng tự phosphoryl hoá.

Yếu tố tăng trưởng EGF (yếu tố kích thích các tế bào biểu bì cũng như một số tế bào khác bước vào phân chia) được nhận biết bởi thụ thể tyrosine kinase. Thụ thể của EGF nằm xuyên

qua màng. Phần thụ thể nằm ngoài tế bào liên kết với EGF nên cấu trúc không gian của phần nằm ngoài tế bào bị thay đổi cho phép chúng tạo liên kết dimer. Chuyển đổi từ monomer sang dimer đã hoạt hoá phần có hoạt tính tyrosine kinase nằm phía trong tế bào chất. Phản ứng dimer hoá là cơ chế chung để hoạt hoá thụ thể tyrosine kinase (Hình 5.14). Một khi đã hoạt hoá, thụ thể xúc tác phản ứng tự phosphoryl hoá chéo nhau. Góc phosphate được chuyển từ ATP sang acid amin tyrosine đặc biệt của chính thụ thể hoặc của một số protein khác thuộc con đường dẫn truyền tín hiệu.

Ngoài ra, tín hiệu có thể ở dạng dimer như tín hiệu nhận biết bởi thụ thể tăng trưởng có nguồn gốc từ tiểu cầu PDGF. Khi tín hiệu liên kết với thụ thể, hai phân tử thụ thể tương tác với nhau và gây phosphoryl hoá lẫn nhau (Hình 5.15).



Hình 5.15:

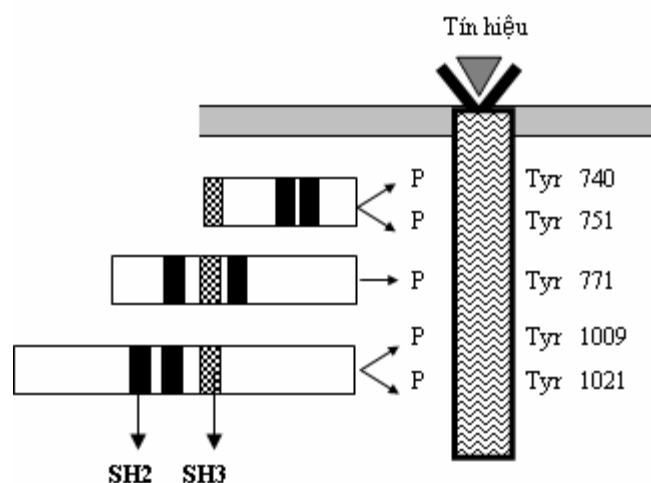
Yếu tố tăng trưởng có nguồn gốc tiểu cầu (*PDGF – platelet driven growth factor*) tồn tại ở dạng dimer gồm hai vị trí liên kết với thụ thể. Do đó tương tác của yếu tố PDGF với thụ thể dẫn đến việc tạo dimer giữa hai thụ thể với nhau và khởi động quá trình truyền tín hiệu vào trong tế bào. Một số yếu tố tăng trưởng khác cũng hoạt động ở dạng dimer thúc đẩy dimer hoá của hai phân tử thụ thể tyrosine với nhau.

Cơ chế truyền tín hiệu bắt đầu từ các thụ thể của EGF, PDGF, FGF đã được nghiên cứu khá kỹ. Trong từng trường hợp, tyrosine của thụ thể bị phosphoryl hoá trở thành vị trí có ái lực liên kết cao đối với các protein làm nhiệm vụ truyền tín hiệu. Liên kết giữa mỗi protein với các vị trí tyrosine (bị phosphoryl hoá) khác nhau trên thụ thể phụ thuộc vào các đoạn polypeptide nằm xung quanh vị trí đó. Nhờ tương tác với thụ thể tại vị trí đặc hiệu, nhiều protein có khả năng tự phosphoryl hoá các tyrosine của riêng mình. Như vậy, những protein này sẽ có hoạt tính kinase chỉ khi tương tác với kinase đã bị phosphoryl hoá trước đó. Rõ ràng phản ứng tự phosphoryl hoá tyrosine giữ vai trò chính trong con đường truyền tín hiệu, chuyển các thành viên sang trạng thái hoạt động, đảm bảo tín hiệu được truyền vào sâu trong tế bào. Các thụ thể tyrosine kinase khác nhau thường tương tác với tổ hợp các protein khác nhau, do đó hoạt hoá các phản ứng trả lời khác nhau.

Xét trường hợp thụ thể cho insulin và yếu tố tăng trưởng tương tự insulin IGF1 (Insulin like Growth Factor-1). Các thụ thể này ở dạng tetramer, do đó khi liên kết với tín hiệu, cấu trúc của thụ thể bị thay đổi nhưng không xảy ra hiện tượng dimer hoá. Mặt khác khi có insulin bám vào, thụ thể tự phosphoryl hoá tại vùng có hoạt tính kinase (phần nằm phía trong màng tế bào). Vùng này tiếp tục gây phosphoryl hoá các tyrosine của protein khác. Các tyrosine này là vị trí liên kết và hoạt hoá các protein trong con đường dẫn truyền tín hiệu.

Protein có tyrosine bị phosphoryl hoá được nhận biết bởi các protein khác có chứa đoạn peptide đặc hiệu gọi là vùng SH2. Rất nhiều protein liên kết với tyrosine có gắn nhóm

phosphat đều có cấu trúc khá bảo toàn tại hai vùng gọi là SH2 và SH3 (Src homology regions 2 & 3-tên gọi bắt nguồn từ các vùng 2 và 3 của protein Src). Những protein có hai vùng SH2 và SH3 còn được gọi là protein tương hợp (*adaptor protein*). Vùng SH2 nhận biết các tyrosine đã bị phosphoryl hoá, do đó những protein có chứa vùng này đều có khả năng liên kết với thụ thể tyrosine kinase hoặc các protein bị phosphoryl hoá ở tyrosine (Hình 5.16). Vùng SH3 của protein tương hợp tương tác với các đoạn peptide giàu proline. Như vậy, tương tác ở vùng SH3 gây nên những thay đổi cấu trúc cần thiết đảm bảo cho tương tác đặc hiệu giữa vùng SH2 và tyrosine có gắn gốc phosphat.



Hình 5.16:

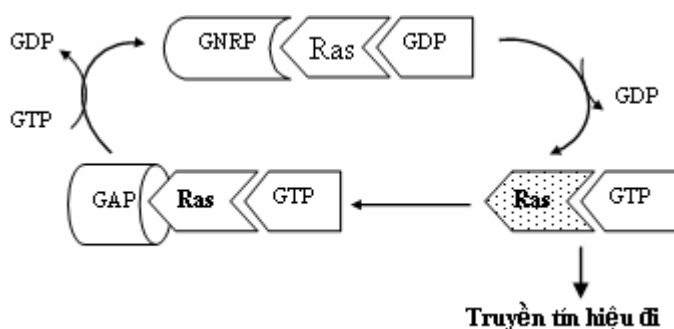
Tương tác giữa protein truyền tín hiệu mang vùng SH2, SH3 với thụ thể của PDGF

Thụ thể có chứa 5 acid amin tyrosine có khả năng tự phosphoryl hoá. Đây là vị trí liên kết với 3 protein đặc hiệu (phía bên trái hình vẽ) có mang các vùng SH2 và SH3. Ngoài ra thụ thể của PDGF còn có hai tyrosine khác nằm sát phía trong màng nguyên sinh chất cũng có khả năng tự phosphoryl hoá

Chúng trở thành vị trí tương tác với tyrosine kinase không phải là thụ thể. Các chữ số chỉ vị trí của tyrosine trong chuỗi peptide. Các vị trí liên kết được xác định nhờ kỹ thuật tái tổ hợp ADN khi gây đột biến từng tyrosine.

5.4.2 Protein Ras và chuỗi các phản ứng truyền tín hiệu hoạt hoá bởi thụ thể tyrosine kinase

Các protein Ras nằm trong nhóm monomer GTPase; chúng có hoạt tính tương tự tiểu phần $G_{\alpha s}$ của protein G. Protein Ras tồn tại ở hai trạng thái khác nhau phụ thuộc vào cấu trúc không gian: trạng thái hoạt động khi liên kết với GTP và trạng thái bất hoạt khi liên kết với GDP. Protein Ras thủy phân GTP chậm hơn ít nhất 100 lần so với tiểu phần $G_{\alpha s}$ của protein G. Ở phía trong màng tế bào, nồng độ GTP luôn cao hơn nồng độ GDP gấp 10 lần. Vì vậy, khi Ras liên kết với GTP, nó sẽ duy trì trạng thái hoạt động nếu như không có các protein khác kiểm soát. Trong tế bào có hai nhóm protein truyền tín hiệu kiểm soát hoạt động của Ras thông qua điều khiển quá trình chuyển trạng thái của protein này giữa dạng hoạt động và bất hoạt. Để ức chế hoạt tính của Ras, các protein kiểm soát sẽ hoạt hoá GTPase. Những protein này được gọi là các protein GAPs (*GTPase-activating protein*). Hoạt động của GAPs dẫn đến tỷ lệ GTP bám vào Ras bị phân hủy tăng, do đó Ras nhanh chóng chuyển sang trạng thái không hoạt động. Tuy nhiên hoạt động của GAPs lại bị ức chế bởi các protein giải phóng guanine (*guanine nucleotide releasing proteins GNRPs*). GNRPs thúc đẩy phản ứng thay thế GDP bằng GTP, do đó chúng lại hoạt hoá Ras (Hình 5.17).



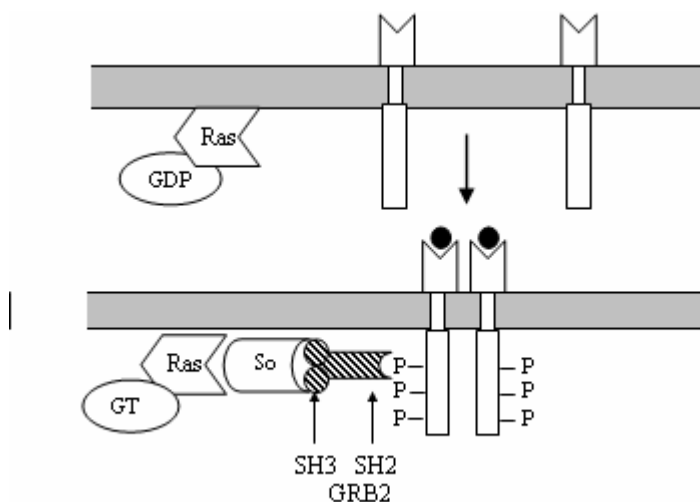
Hình 5.17:

Kiểm soát hoạt động của protein Ras

Các protein hoạt hoá (GAPs) làm mất hoạt tính của Ras bằng việc kích thích phản ứng thủy phân GTP liên kết với Ras. Các protein GNRPs hoạt hoá Ras thông qua phản ứng thay thế GDP bằng GTP. Do nồng độ GTP trong tế bào cao hơn 10 lần nồng độ GDP, Ras luôn luôn có xu thế duy trì liên kết với GTP. Ở động vật có vú, hai loại protein GAPs được phát hiện là p120^{GAP} và neurofibromin

Hoạt động của chúng phụ thuộc vào từng loại tế bào nhằm duy trì Ras ở trạng thái không hoạt động.

Ras tham gia vào con đường truyền tín hiệu từ thụ thể tyrosine kinase. Tuy nhiên, Ras không liên kết trực tiếp với thụ thể; Ras nhận tín hiệu thông qua protein tương hợp. Những thí nghiệm nuôi cấy tế bào có bổ sung yếu tố EGF đã làm sáng tỏ hoạt động của Ras trong chuỗi truyền tín từ thụ thể tyrosine kinase. Đầu tiên, yếu tố EGF được thụ thể tyrosin kinase tiếp nhận. Tiếp đến, các tyrosine của thụ thể tự phosphoryl hoá. Những tyrosine này sẽ liên kết với vùng SH2 của protein GRB2. Đến lượt mình, GRB2 liên kết với protein Sos và hoạt hoá protein này. Chức năng của Sos là thay thế GDP của Ras bằng GTP. Cuối cùng Ras phân hủy GTP để truyền tín hiệu đi (Hình 5.18).



Hình 5.18 :

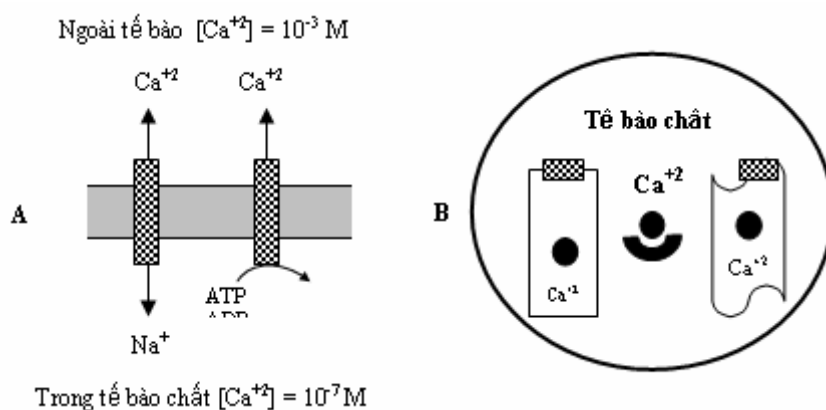
Protein Ras tiếp nhận tín hiệu từ thụ thể tyrosine kinase thông qua hai protein tương hợp GRB2 và Sos. Vùng SH2 của GRB2 tương tác với tyrosine bị phosphoryl hoá của thụ thể. Vùng SH3 của GRB2 tương tác với Sos. Sos hoạt hoá Ras khiến GDP bị thay thế bởi GTP; GTP bị phân hủy để tiếp tục truyền tín hiệu.

Cùng với các protein Rho và Rac, Ras tham gia truyền tín hiệu từ thụ thể trên mặt tế bào đến hệ thống mạng lưới sợi actin. Chúng phối hợp hoạt động cùng với các protein Rab liên quan đến điều khiển vận chuyển các nang trong tế bào. Giống như hầu hết các monomer GTPase, protein Ras được giữ chặt một đầu vào phía trong màng sinh chất. Chức năng của

chúng là truyền tín hiệu từ thụ thể tyrosine kinase trên bề mặt tế bào vào đến nhân để hoạt hoá các gen liên quan đến tăng trưởng hoặc biệt hoá tế bào. Khi hoạt động của Ras bị ức chế, tế bào không trả lời với các kích thích của bên ngoài. Tuy nhiên khi protein Ras bị đột biến khiến Ras liên tục tồn tại ở trạng thái có hoạt tính hoặc được tổng hợp quá nhiều sẽ khiến tế bào tăng trưởng và phát triển không ngừng, ngay khi không có tín hiệu bên ngoài. Trong thực tế, các protein Ras được phát hiện khi chúng hoạt động quá mạnh do gen mã cho Ras bị đột biến. Lúc đó tế bào không chịu sự kiểm soát của quá trình phân bào và biệt hoá, chúng sinh sôi không ngừng dẫn đến phát triển ung thư. Khoảng 30% ung thư ở người là do đột biến xảy ra ở gen *ras*.

5.5 Tín hiệu thứ cấp Ca^{+2} trong chuỗi truyền tín hiệu

Chúng ta đã biết phân tử protein G trung chuyển tín hiệu từ thụ thể đến adenylyl cyclase dẫn đến thay đổi nồng độ cAMP trong tế bào. Chúng ta sẽ xét đến vai trò khác của protein G trong việc phối hợp hoạt động giữa thụ thể với enzyme phospholipase C. Enzyme này làm thay đổi nồng độ ion Ca^{+2} trong tế bào chất, khởi động chuỗi các phản ứng truyền tín hiệu khác nhau. Ion Ca^{+2} là tín hiệu thứ cấp được tế bào sử dụng rộng rãi hơn cAMP trong quá trình truyền tín hiệu.



Hình 5.19:

Kiểm soát nồng độ Ca^{+2} trong tế bào chất. Sơ đồ biểu diễn một số phương thức cơ bản đảm bảo duy trì nồng độ Ca^{+2} trong tế bào chất ở nồng độ thấp. (A): Ca^{+2} được bơm ra ngoài tế bào bằng bơm trao đổi Na và Ca hoặc bằng bơm Ca^{+2} -ATPase. (B): Ca^{+2} được bơm từ tế bào chất vào mạng lưới nội chất ER hoặc vào ty thể. Ngoài ra trong tế bào chất còn tồn tại các phân tử khác nhau liên kết với Ca^{+2} làm giảm nồng độ Ca^{+2} ở trạng thái tự do. Ion Ca^{+2} được bơm vào ty thể chỉ khi nồng độ Ca^{+2} ở tế bào chất rất cao gây nguy hiểm cho tế bào.

Nồng độ Ca^{+2} tự do trong tế bào chất của bất kỳ tế bào nào cũng luôn được kiểm soát chặt chẽ không quá $<10^{-7}$ M, trong khi ở ngoài tế bào và ở màng lưới nội chất ER nồng độ Ca^{+2} cao hơn rất nhiều (10^{-3} M). Vì vậy, gradient nồng độ ion Ca^{+2} luôn tồn tại theo chiều từ ngoài vào trong tế bào hoặc từ khoang trong ER ra ngoài tế bào chất. Khi chưa có tín hiệu, tế bào thực thi các cơ chế kiểm soát để đảm bảo sự chênh lệch nồng độ này. Khi có tín hiệu, kênh truyền ion Ca^{+2} mở ra, các ion này tràn vào trong tế bào chất làm nồng độ Ca^{+2} ở đó tăng vọt lên gây hoạt hoá các protein liên quan đến Ca^{+2} .

Có nhiều cơ chế khác nhau để duy trì nồng độ ion Ca^{+2} ở mức rất thấp trong tế bào chất. Mọi tế bào eukaryot đều có phức Ca^{+2} - ATPase trên màng nội chất, sử dụng năng lượng do thủy phân ATP để bơm Ca^{+2} ra khỏi tế bào. Một số tế bào như tế bào cơ, tế bào thần kinh còn có bơm trao đổi Ca^{+2} và Na^{+} . Bơm này chỉ bơm Ca^{+2} ra khỏi tế bào khi nồng độ Ca^{+2} ở trong tế bào vượt quá 10 lần nồng độ cho phép. Ngoài ra trên màng ER cũng có bơm Ca^{+2} . Khi tế bào ở trạng thái tĩnh, nồng độ Ca^{+2} khoảng 10^{-7} M, còn khi tế bào hoạt động nồng độ này có

thể đạt đến 5×10^{-6} M. Khi tế bào tổn thương không có khả năng bơm Ca^{+2} , nồng độ Ca^{+2} có thể tăng đến mức gây nguy hiểm cho tế bào (10^{-5} M). Lúc này ngay cả bơm Ca^{+2} trên màng ty thể cũng hoạt động, sử dụng năng lượng của các phản ứng truyền điện tử để bơm Ca^{+2} ra khỏi tế bào chất. Các con đường thải Ca^{+2} được mô tả trên hình 5.19.

5.5.1 Inositol phospholipid

Rất nhiều hormon khi tương tác với thụ thể trên bề mặt tế bào gây nên sự tăng nồng độ Ca^{+2} trong tế bào chất, thậm chí ngay cả khi ion Ca^{+2} không có mặt trong môi trường xung quanh. Ở các tế bào gan, tế bào mỡ và một số tế bào khác, thực nghiệm đã phát hiện rằng ion Ca^{+2} xuất hiện trong tế bào chất do được giải phóng ra từ mạng lưới nội chất ER hoặc từ một số nang trong tế bào chứ không phải từ môi trường bên ngoài vào. Câu hỏi đặt ra là làm thế nào để khi có tương tác giữa hormon và thụ thể trên bề mặt tế bào, tín hiệu được truyền đến ER và từ đó ion Ca^{+2} được bơm ra tế bào chất?

Cho đến đầu năm 1980, thí nghiệm đã chỉ ra rằng nồng độ Ca^{+2} trong tế bào chất tăng lên liên quan đến phản ứng phosphoryl hoá phân tử phosphatidylinositol (PI) tạo nên inositol phospholipid. Phospholipid PI nằm ở màng sinh chất về phía tế bào chất. Hai dạng inositol phospholipid hay gặp nhất trong hệ thống dẫn truyền tín hiệu là PI 4-phosphate (PIP) và PI 4,5-bisphosphate (PIP₂). Khi có tín hiệu bên ngoài gửi đến, thụ thể được hoạt hoá. Thông qua protein G (gọi là G_q), tín hiệu gửi đến enzym phospholipase C- β . Enzym này được hoạt hoá, xúc tác phản ứng phân cắt PIP₂ thành hai sản phẩm: 1,2-diacylglycerol (được giữ lại trên màng) và inositol 1,4,5-triphosphate (IP₃) tan trong nước. Cả hai sản phẩm trên đều tham gia vào hai nhánh truyền tín hiệu khác nhau.

Mặc dù con đường truyền tín hiệu thông qua inositol phospholipid chưa được nghiên cứu chi tiết như với cAMP, ít nhất có khoảng hơn 25 thụ thể phân bố trên bề mặt tế bào nối với protein G_q tham gia vào con đường truyền tín hiệu thông qua inositol phospholipid.

5.5.2 Inositol triphosphate (IP₃) và sự vận chuyển Ca^{+2} ra khỏi ER

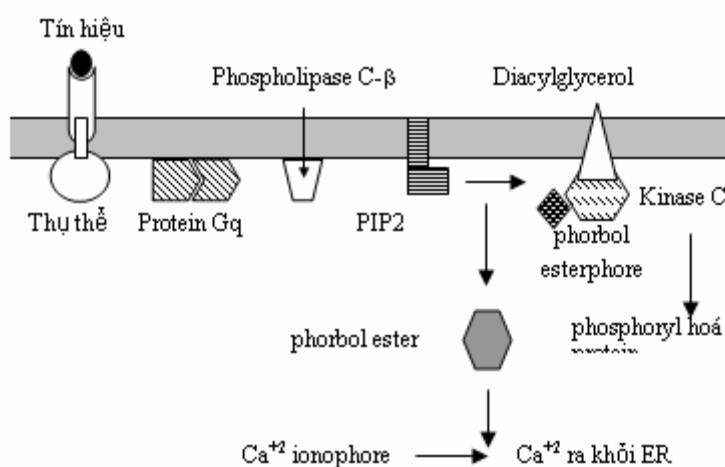
Inositol triphosphate IP₃ là sản phẩm của phản ứng phân hủy PIP₂. Đây là những phân tử kích thước nhỏ tan trong nước và khuếch tán rất nhanh vào tế bào chất. Ở đó, IP₃ liên kết với các kênh đặc biệt ở trên màng ER (IP₃-gated Ca^{+2}). Kênh này được điều khiển bởi cơ chế phản hồi tích cực. Khi Ca^{+2} ra khỏi kênh, nó tương tác trở lại với kênh khiến kênh được mở rộng hơn. Do đó số ion Ca^{+2} được giải phóng ra khỏi ER tăng lên rất nhanh. Ion Ca^{+2} được giải phóng ra rất nhanh và đột ngột theo kiểu "tất cả hoặc không có gì" (*all-or-none*).

Khi lượng Ca^{+2} tăng lên trong tế bào chất, enzym đặc hiệu gây phosphoryl hoá IP₃ cũng được hoạt hoá. Do đó một số IP₃ bị phosphoryl hoá tạo IP₄. Chất này giúp cho tế bào duy trì phản ứng trả lời tín hiệu được lâu và chậm hơn. Đồng thời IP₄ kích thích việc tích lũy Ca^{+2} trong tế bào chất bằng việc lấy Ca^{+2} từ ngoài vào. Như vậy, nồng độ Ca^{+2} trong tế bào chất có thể tăng bằng hai con đường: chuyển Ca^{+2} từ ER ra và lấy Ca^{+2} từ ngoài tế bào vào.

Sau khi trả lời tín hiệu bên ngoài, tế bào phản ứng lại sự tăng nồng độ Ca^{+2} theo hai cơ chế. Cả hai đều nhằm mục đích giảm số lượng ion Ca^{+2} trong tế bào chất trở lại nồng độ ban đầu. Trong cơ chế thứ nhất, IP₃ nhanh chóng bị khử nhóm phosphate (chuyển sang dạng không hoạt động) bởi enzym phosphatase đặc hiệu. Do đó, quá trình giải phóng Ca^{+2} ra khỏi ER bị ngừng. Với cơ chế thứ hai, ion Ca^{+2} nhanh chóng được bơm ra khỏi tế bào bằng bơm Ca^{+2} -ATPase.

Cùng với IP3 làm thay đổi nồng độ Ca^{+2} trong tế bào còn có sản phẩm khác của PIP2 là diacylglycerol. Một trong những vai trò quan trọng mà diacylglycerol đảm nhiệm trong quá trình truyền tín hiệu là hoạt hoá protein kinase serine/threonine. Enzym được hoạt hoá bởi diacylglycerol được gọi chung là kinase C (hoặc PKC) do hoạt tính của nó phụ thuộc vào Ca^{+2} . Khi nồng độ Ca^{+2} trong tế bào chất tăng lên do tác dụng của IP3, protein kinase C biến đổi cấu trúc giúp nó di chuyển trong tế bào chất đến sát màng sinh chất. Ở đó nó được hoạt hoá bởi ion Ca^{+2} và diacylglycerol.

Ở động vật có vú, có ít nhất 8 dạng kinase C khác nhau. Chúng gây phosphoryl hoá protein tại các acid amin serine hoặc threonine đặc hiệu. Cơ chất của kinase C rất khác nhau, phụ thuộc vào từng loại tế bào. Nồng độ kinase C cao nhất được tìm thấy trong não, ở đó chúng phosphoryl hoá các kênh ion trong tế bào thần kinh dẫn đến thay đổi tính chất của màng sinh chất và kích thích màng này.



Hình 5.20:

Hai nhánh truyền tín hiệu của inositol phospholipid

Thụ thể liên kết với protein G (Gq) gây hoạt hoá phospholipase C-β. Enzym này cắt PIP2 tạo IP3 và diacylglycerol. Sản phẩm thứ hai lại hoạt hoá kinase C. Hai enzym phospholipase C-β và kinase C tan trong nước. Khi di chuyển đến sát mặt trong của màng nguyên sinh chất, hai enzym này mới bộc lộ hoạt tính của mình.

Trong rất nhiều tế bào, hoạt động của kinase C làm tăng quá trình phiên mã của một số gen đặc biệt. Các gen có thể được hoạt hoá theo hai con đường khác nhau. Trong con đường thứ nhất, kinase C hoạt hoá các protein kinase trong chuỗi các phản ứng phosphoryl hoá, dẫn đến hoạt hoá các protein điều khiển gen. Trong con đường thứ hai, hoạt động của kinase C gây phosphoryl hoá protein ức chế, nhờ đó các protein điều khiển được giải phóng ra ở dạng tự do trong tế bào chất, di chuyển vào nhân và bật mở các gen đặc hiệu.

Các con đường truyền tín hiệu của inositol phospholipid được mô tả trên hình 5.20. Một số chất khi được bổ sung vào môi trường nuôi cấy tế bào có thể khởi động các con đường đó. Ví dụ, IP3 có hoạt tính khi có mặt của phức Ca^{+2} -ionophore, chất này cho phép nồng độ Ca^{+2} tăng lên trong tế bào chất. Ngoài ra phorbol ester cũng gây ra những tác dụng tương tự diacylglycerol. Chất này liên kết với kinase C và hoạt hoá enzym.

Trong các tế bào, các cơ chế truyền tín hiệu được kết hợp cùng nhau để tạo ra phản ứng tối ưu nhất cho tế bào. Ví dụ, để kích thích các tế bào nuôi cấy bước vào phân chia, cả hai chất Ca^{+2} -ionophore và chất hoạt hoá kinase C đều được thêm vào môi trường nuôi cấy. Nếu chỉ thêm một trong hai chất này thì tế bào không phân chia.

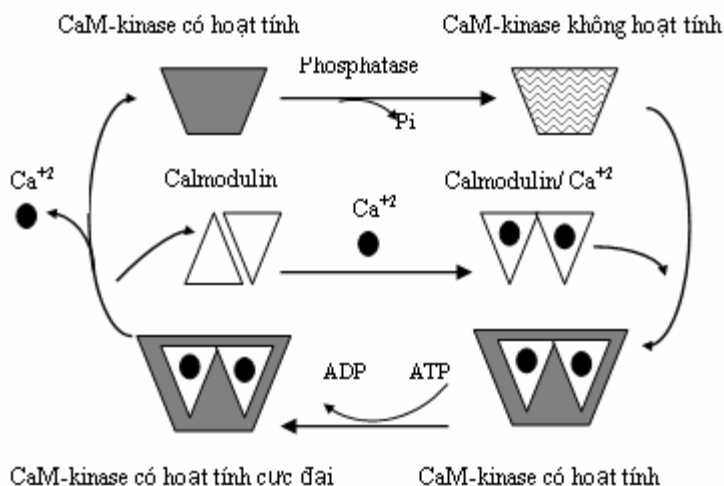
5.5.3 Calmodulin- protein tạo phức với Ca^{+2} ở trong tế bào

Nồng độ Ca^{+2} trong tế bào chất luôn phải nhỏ hơn 10^{-7} M và không bao giờ vượt quá 5×10^{-6} M ngay khi các kênh dẫn truyền Ca^{+2} mở. Chính vì vậy để tăng cường khả năng kiểm soát nồng độ Ca^{+2} , trong tế bào chất còn tồn tại các protein liên kết đặc hiệu và có ái lực cao với Ca^{+2} , đảm bảo cho sự tăng nồng độ Ca^{+2} chỉ xảy ra trong khoảng thời gian ngắn, chính xác. Trong các tế bào eukaryot, protein tương tác đặc hiệu với Ca^{+2} hay gặp nhất là calmodulin. Một tế bào động vật có khoảng 10^7 phân tử calmodulin, tương đương với 1% tổng số protein có trong tế bào. Calmodulin có nhiều chức năng khác nhau, có thể liên kết với Ca^{+2} hoặc làm trung gian cho một số phản ứng sinh học. Phân tử calmodulin là một chuỗi polypeptid gồm 150 acid amin có 4 vị trí đặc hiệu tương tác với Ca^{+2} . Cấu trúc không gian của Calmodulin thay đổi khi nó liên kết với Ca^{+2} .

Phức chất Ca^{+2} /calmodulin không có hoạt tính enzym. Phức này làm nhiệm vụ kiểm soát hoạt tính của protein liên kết với chúng. Thông thường phức Ca^{+2} /calmodulin tương tác với những protein giữ vai trò vận chuyển các chất qua màng. Khi nồng độ Ca^{+2} trong tế bào tăng cao vì bất kỳ một lý do gì thì bơm Ca^{+2} được kích thích bởi phức Ca^{+2} /calmodulin để bơm Ca^{+2} ra khỏi tế bào. Bên cạnh đó, hoạt tính của một số kinase hoàn toàn phụ thuộc vào phức Ca^{+2} /calmodulin. Vì thế hoạt động chính của phức là hoạt hoá những kinase đặc biệt này. Chúng được gọi chung là CaM-kinase (*Ca⁺²/calmodulin-dependent protein kinase*).

Protein CaM-kinase xúc tác cho phản ứng phosphoryl hoá acid amin serine/threonine của một số protein đặc hiệu. Cũng giống như trường hợp của cAMP, sự tăng nồng độ Ca^{+2} tự do trong tế bào chất phụ thuộc vào protein đặc hiệu mà những protein này bị kiểm soát bởi CaM-kinase. Đặc biệt, CaM-kinase II có mặt trong mọi tế bào động vật và nhiều nhất ở các tế bào thần kinh. Tại đó CaM-kinase II chiếm 2% tổng số protein tập trung ở các synape. Khi synape hoạt động, dòng ion Ca^{+2} vượt qua kênh điện thế Ca^{+2} (*voltage-gated Ca^{+2} channels*) vào bên trong tế bào, kích thích tế bào tiết ra các chất dẫn truyền thần kinh. Dòng ion Ca^{+2} này sẽ hoạt hoá CaM-kinase II. Đến lượt mình, CaM-kinase II phosphoryl hoá tyrosine hydroxylase khiến enzym này có hoạt tính. Đây là enzym kiểm soát tổng hợp các chất dẫn truyền thần kinh. Nhờ đó, quá trình tiết và tổng hợp các chất dẫn truyền được kích thích khi tế bào thần kinh ở trạng thái hoạt hoá do nồng độ Ca^{+2} tăng.

Đặc điểm nổi bật của các CaM-kinase là chúng có thể hoạt động như các phần tử nhớ: chúng chuyển sang trạng thái hoạt động khi có mặt của phức Ca^{+2} /calmodulin và giữ được hoạt tính ngay khi nồng độ Ca^{+2} giảm. Đó là do sau khi bị hoạt hoá bởi phức Ca^{+2} /calmodulin, CaM-kinase phosphoryl hoá các protein khác và có khả năng tự phosphoryl hoá chính nó. Như vậy CaM-kinase tự hoạt hoá mình. Ở trạng thái tự hoạt hoá, enzym CaM-kinase giữ được hoạt tính ngay khi không có Ca^{+2} . Nhờ đó thời gian hoạt động của enzym được kéo dài cho đến khi phosphatase áp đảo khiến enzym mất tác dụng (Hình 5.21).



Hình 5.21:

Hoạt động của CaM-kinase. Phân tử enzym gồm 12 tiểu đơn vị. Khi không có phức Ca^{+2} /calmodulin, enzym không có hoạt tính do vùng qui định hoạt tính liên kết với vùng ức chế. Khi phức Ca^{+2} /calmodulin tương tác với enzym, cấu trúc không gian của enzym thay đổi, enzym tự phosphoryl hoá vùng ức chế, giải phóng vùng có hoạt tính. Tự phosphoryl hoá cho phép kéo dài hoạt tính enzym theo hai cách: Thứ nhất, giữ phức Ca^{+2} /calmodulin không tách ra khỏi enzym cho đến khi nồng độ Ca^{+2} trong tế bào giảm về nồng độ cho phép (ít nhất trong 10 giây). Thứ hai, enzym được duy trì ở trạng thái hoạt hoá không phụ thuộc nồng độ Ca^{+2} ngay cả khi phức Ca^{+2} /calmodulin tách ra khỏi enzym. Hoạt tính duy trì cho đến khi phản ứng phosphatase áp đảo được phản ứng phosphoryl hoá. Kinase CaMIII đóng vai trò quan trọng trong hoạt động thần kinh. Chuột bị đột biến một trong các protein của khâu dẫn truyền tín hiệu mô tả trên hình đều mất trí nhớ về vị trí không gian (theo Alberts & cs, 2002).

5.6 Khuếch đại các tín hiệu bên ngoài tế bào

Thông thường khi nồng độ tín hiệu thay đổi, trả lời của tế bào cũng thay đổi theo một cách gần như tuyến tính. Ví dụ, phản ứng của tế bào với sự thay đổi nồng độ hormon xảy ra có tính chất đồng biến. Đó là do mỗi thụ thể chỉ tương tác với một phân tử hormon và hoạt động của các gen liên quan là độc lập với nhau. Khi nồng độ hormon tăng, nồng độ phức hormon-thụ thể cũng tăng theo một cách tuyến tính. Do đó, số phức chất bám vào ADN để hoạt hoá gen cũng tăng theo tương tự. Kết quả là trả lời của tế bào với hormon tăng tỷ lệ thuận hoặc không đổi.

Tuy nhiên, trong một số trường hợp, trả lời của tế bào thường xảy ra đột ngột khi nồng độ tín hiệu tăng dần. Ở nồng độ nhỏ hơn ngưỡng nhất định, tế bào không phát hiện được. Nhưng khi nồng độ vượt qua ngưỡng, phản ứng có thể đạt ngay giá trị cực đại. Vì sao nồng độ tín hiệu tăng dần dần nhưng trả lời của tế bào xảy ra rất đột ngột, gần giống như cách thức "tất cả hay không có gì"?

Một trong những yêu cầu để phản ứng tuân theo cách thức trên là cần có hai hay nhiều phân tử (hoặc một phức chất) tương tác với các đại phân tử để kích thích phản ứng trả lời. Ví dụ, trong một số phản ứng trả lời kích thích của hormon steroid, không phải chỉ một phức thụ thể-hormon mà phải có hai phức trở lên cùng tương tác với ADN để hoạt hoá một gen. Do đó khi nồng độ hormon vừa tăng, gen bị hoạt hoá rất đột ngột và đạt ngưỡng hoạt động cực đại. Cơ chế hoạt hoá kinase A và calmodulin xảy ra một cách tương tự. Nhiều ion Ca^{+2} cùng tương tác với calmodulin để gây thay đổi cấu trúc không gian của protein này khiến nó bị hoạt hoá rất nhanh. Như vậy, khi nồng độ Ca^{+2} tự do tăng 10 lần thì hoạt tính của calmodulin tăng 50 lần.

Phản ứng trả lời của tế bào cũng xảy ra theo cách thức đột ngột khi tín hiệu hoạt hoá enzym thứ nhất, đồng thời ức chế enzym thứ hai, mà hai enzym này xúc tác cho hai phản ứng

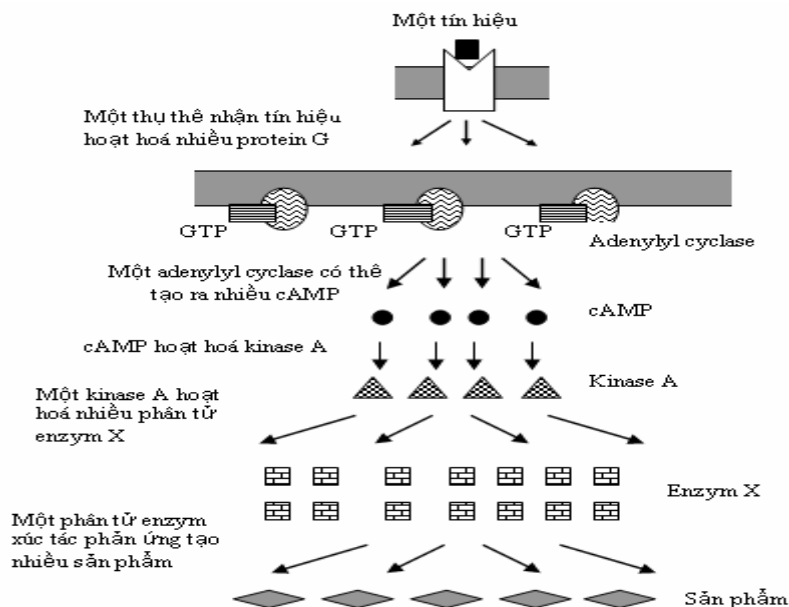
có chức năng ngược nhau của cùng một quá trình truyền tín hiệu. Ví dụ, một tín hiệu đồng thời liên quan đến hai phản ứng: phản ứng thứ nhất gây phosphoryl hoá nhưng phản ứng thứ hai lại khử gốc phosphate. Chẳng hạn như trong tế bào cơ xương, việc tăng nồng độ cAMP vừa hoạt hoá phosphorylase kinase vừa ức chế phosphatase. Do đó phản ứng phân giải glycogen xảy ra nhanh, đột ngột so với việc tăng nồng độ chất kích thích ban đầu.

Với một số ví dụ điển hình về cách thức truyền tín hiệu, chúng ta đã hình dung được phần nào hệ thống các kênh dẫn truyền đan xen vào nhau, kiểm soát qua lại đối với nhau. Các thụ thể nối với protein G có thể gây hoạt hoá hoặc gây mất hoạt tính một cách gián tiếp các enzym bám trên màng sinh chất hoặc trên các kênh dẫn truyền ion. Một số thụ thể khác đảm nhận chức năng hoạt hoá hoặc ức chế adenylyl cyclase làm thay đổi nồng độ của chất dẫn truyền tín hiệu trung gian cAMP trong tế bào. Một số thụ thể liên quan đến hoạt hoá phospholipase C- β . Enzym này thủy phân PIP₂ (*phosphatidylinositol bisphosphate*) tạo ra hai chất truyền tín hiệu trung gian là IP₃ (*inositol trisphosphate*) và diacylglycerol. Vai trò của IP₃ làm tăng nồng độ Ca⁺² trong tế bào chất bằng việc kích thích giải phóng Ca⁺² từ ER. Còn diacylglycerol được giữ lại trên màng sinh chất làm nhiệm vụ hoạt hoá kinase C. Các enzym kinase A, kinase C và kinase CaM gây phosphoryl hoá các protein đặc hiệu ở các acid amin serine hoặc threonine, làm thay đổi hoạt tính của các protein này. Mỗi loại tế bào chứa các protein đặc hiệu đặc trưng cho loại tế bào đó. Chúng đảm bảo cho tế bào trả lời chính xác với từng kích thích.

Tín hiệu thường được khuếch đại lên rất nhiều lần, do đó phản ứng trả lời của tế bào có thể xảy ra rất nhanh và đạt ngưỡng cực đại. Tuy nhiên, phản ứng đó kết thúc rất nhanh ngay khi ngừng kích thích bên ngoài. Đó là do bản thân phân tử protein G có khả năng tự thủy phân GTP cũng như phân tử IP bị khử gốc phosphate bởi phosphatase rất nhanh vv... Việc thay đổi trạng thái liên tục của các tín hiệu thứ cấp trong tế bào đảm bảo nồng độ của chúng đáp ứng với kích thích bên ngoài mà không cần phải tổng hợp mới.

Khi một tín hiệu bên ngoài tế bào tương tác với thụ thể, thụ thể có thể hoạt hoá đồng thời nhiều phân tử protein Gs. Một phân tử Gs lại có thể hoạt hoá nhiều phân tử adenylyl cyclase. Mỗi phân tử Gs có thể tồn tại vài giây ở trạng thái hoạt hoá trước khi nó thủy phân GTP trở về trạng thái bất hoạt. Thời gian này cũng đủ cho phân tử adenylyl cyclase tương tác với Gs đang ở trạng thái hoạt hóa và xúc tác tạo ra số lượng lớn cAMP từ ATP (Hình 5.22).

Một cơ chế khuếch đại tín hiệu tương tự xảy ra trong con đường truyền tín hiệu inositol-phospholipid. Nồng độ tín hiệu bên ngoài tế bào cỡ 10^{-10} M đủ để kích thích tạo ra nồng độ 10^{-6} M của các tín hiệu thứ cấp như cAMP hoặc Ca⁺². Các tín hiệu thứ cấp này có chức năng hoạt hóa các enzym đặc hiệu hoặc các kênh dẫn truyền ion. Khi tế bào chỉ tiếp nhận một phân tử tín hiệu bên ngoài, hàng trăm phân tử khác trong tế bào có thể bị thay đổi cấu trúc, hoạt tính hoặc nồng độ...



Hình 5.22:

Chuỗi phản ứng khuếch đại tín hiệu được kích thích bởi một tín hiệu ban đầu. Bước khuếch đại đầu tiên đòi hỏi phân tử tín hiệu phải tương tác với thụ thể trong thời gian đủ để phức này hoạt hoá các phân tử Gs. Trong một số trường hợp khác, ligand phải tách ra khỏi phức rất nhanh để quá trình khuếch đại bắt đầu xảy ra (theo Alberts & cs., 2002).

Với bất kỳ chuỗi truyền và khuếch đại tín hiệu xảy ra nhanh và nhạy thì đều phải đảm bảo sự quay về trạng thái cân bằng của từng phản ứng trong chuỗi khi ngừng tín hiệu kích thích. Vì vậy trong tế bào luôn tồn tại cơ chế rất hiệu lực đảm bảo việc phân huỷ cAMP, giảm nồng độ Ca^{+2} trong tế bào chất cũng như gây mất hoạt tính của những enzym hoặc protein truyền tín hiệu trung gian đã bị hoạt hóa trước đó. Cơ chế này không chỉ ngừng phản ứng trả lời tín hiệu mà còn phải lập lại rất nhanh trạng thái tĩnh giống như trước khi bị kích thích. Nói chung, khi phản ứng trả lời với kích thích xảy ra nhanh thì cơ chế giúp tế bào quay lại trạng thái tĩnh cũng được thực hiện một cách mau chóng tương tự.

5.7 Truyền tín hiệu qua các thụ thể nối với enzym trên bề mặt tế bào

Cũng giống như các thụ thể nối với protein G, thụ thể nối với enzym thường là các protein xuyên màng (*trans-membrane protein*). Những thụ thể này có phần nhận biết tín hiệu nằm ở phía ngoài màng. Tuy nhiên, phần nằm phía trong màng không liên kết với protein G mà nó có hoạt tính enzym hoặc tương tác trực tiếp với enzym.

Có 5 loại thụ thể nối với enzym: 1/ Thụ thể guanylyl cyclase xúc tác cho phản ứng tạo cGMP trong tế bào chất. 2/ Thụ thể kinase tyrosine gây phosphoryl hoá các acid amin tyrosine đặc hiệu của các protein dẫn truyền tín hiệu trung gian. 3/ Thụ thể phối hợp cùng kinase tyrosine. 4/ Thụ thể phosphatase tyrosine chuyển nhóm phosphate khỏi acid amin tyrosine của protein đặc hiệu. 5/ Thụ thể serine/threonine kinase gây phosphoryl hoá các acid amin serine hoặc threonine của một số protein. Chúng ta lần lượt xét từng loại thụ thể trên.

5.7.1 Thụ thể guanylyl cyclase

Các thụ thể guanylyl cyclase sử dụng cGMP như tín hiệu trung gian tương tự như các thụ thể nối với protein G sử dụng cAMP. Tuy nhiên, thụ thể vừa nhận biết tín hiệu đồng thời đóng

vai trò của enzym xúc tác phản ứng tổng hợp cGMP. Ví dụ điển hình cho hoạt động của thụ thể loại này là tương tác giữa thụ thể với tín hiệu ANPs (*atrial natriuretic peptides*) được tiết ra từ các tế bào cơ trong tâm nhĩ của tim khi huyết áp tăng. Kết quả cuối cùng của tương tác này là ion Na^+ và H_2O được thải ra khỏi tế bào cơ trơn. Thành mạch máu giãn ra, nhờ đó huyết áp giảm. Phần thụ thể nằm ngoài màng tế bào có vị trí liên kết với ANPs còn phần nằm trong có hoạt tính guanylyl cyclase. Khi ANPs liên kết với thụ thể, tín hiệu được truyền vào bên trong tế bào và phần thụ thể có hoạt tính enzym xúc tác phản ứng tổng hợp cGMP. Các phân tử cGMP lập tức hoạt hoá các protein kinase phụ thuộc cGMP (gọi chung là các G-kinase). Bằng cách này tín hiệu được truyền vào trong tế bào. Trong thực tế, các tín hiệu guanylyl cyclase ít gặp trong các nghiên cứu hoạt động và chức năng của tế bào.

5.7.2 Các oncogene và tín hiệu dẫn truyền từ thụ thể tyrosine kinase

Ở động vật bậc cao, tế bào chỉ bước vào phân chia khi chúng được kích thích bởi các factor tăng trưởng do các tế bào khác tiết ra. Tuy nhiên đối với các tế bào ung thư, chúng phân chia liên tục ngay khi không có các tín hiệu bên ngoài. Như vậy, chúng không chịu sự kiểm soát chung của quá trình sinh trưởng. Một trong những nguyên nhân dẫn đến sự mất khả năng kiểm soát là do đột biến xảy ra ở những gen mã cho protein tham gia con đường dẫn truyền tín hiệu (ví dụ các protein Ras, Src...). Những gen khi hoạt động bất thường dẫn đến rối loạn quá trình phát triển, gây ung thư được gọi là các oncogen.

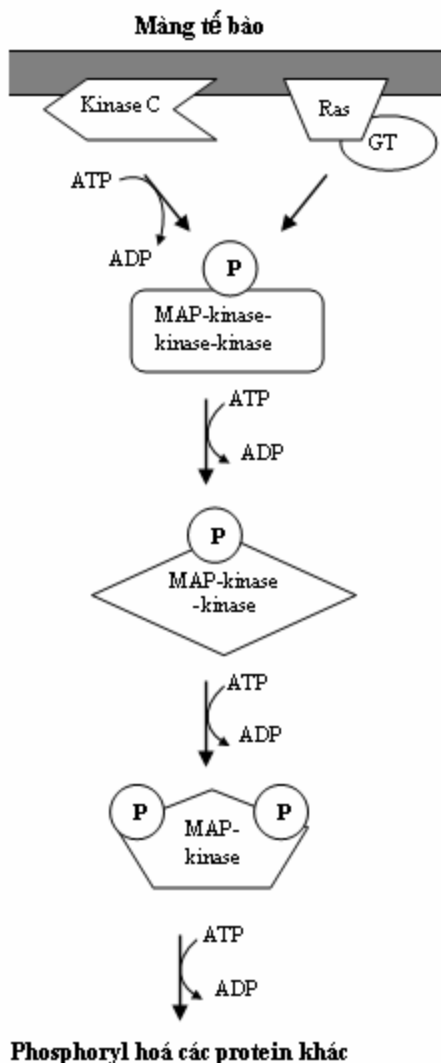
Bất kỳ một đột biến nào liên quan đến việc tạo ra sản phẩm protein có hoạt tính không bình thường trong con đường dẫn truyền tín hiệu đều có thể khởi động quá trình hình thành khối u. Tế bào bình thường trở thành tế bào ung thư sẽ phân chia liên tục mà không yêu cầu các tín hiệu bên ngoài gửi đến. Ví dụ, oncogen *erbB* mã cho thụ thể của factor EGF có vùng qui định hoạt tính tyrosine kinase luôn luôn ở trạng thái hoạt hoá. Tế bào mang oncogen này sẽ phân chia liên tục. Nói chung, sự phân chia vô tổ chức thường do tế bào mang oncogen mã cho protein điều khiển mà protein này luôn luôn có hoạt tính. Ngoài ra, một số hormon kích thích tế bào bước vào phân chia khi liên kết với thụ thể nối với protein G hơn là với thụ thể tyrosine kinase.

5.7.3 Protein MAP kinase

Quá trình phosphoryl hoá tyrosine ở Ras khiến protein này được hoạt hoá rất nhanh. Tyrosine bị khử gốc phosphate nhờ enzym đặc hiệu tyrosine phosphatase và Ras bị mất hoạt tính khi GTP (liên kết với Ras) bị phân huỷ thành GDP. Để kích thích tế bào sinh sôi hoặc biệt hoá, nhất thiết các tín hiệu ban đầu phải tạo nên chuỗi các phản ứng xảy ra liên tiếp để truyền tín hiệu vào trong nhân. Hệ thống truyền tín hiệu liên quan đến một loạt phản ứng phosphoryl hoá serine/threonine. Phản ứng xảy ra trước sẽ hoạt hoá phản ứng sau và chúng đều xảy ra lâu hơn phản ứng phosphoryl hoá tyrosine. Một nhóm protein serine/threonine kinase gồm ít nhất 5 thành viên đóng vai trò đặc biệt quan trọng trong chuỗi phản ứng truyền tín hiệu. Đó là các protein kinase kích thích quá trình phân bào mitose, được gọi là các protein MAP (*Mitose Activator Proteins*) hay là MAP kinase.

Hoạt tính kinase của MAP chỉ đạt cực đại khi cả hai acid amin threonine và tyrosine của nó cùng bị phosphoryl hoá. Hai acid amin này chỉ cách nhau bởi 1 acid amin. Kinase xúc tác cho cả hai phản ứng phosphoryl hoá trên MAP kinase được gọi là MAP kinase-kinase. Việc yêu cầu cả hai acid amin cùng được phosphoryl hoá đảm bảo cho MAP kinase luôn được giữ ở trạng thái không hoạt động trừ khi bị hoạt hoá bởi chính MAP kinase-kinase (lưu ý rằng MAP kinase không có khả năng tự phosphoryl hoá mà nó có hoạt tính kinase chỉ khi cả hai

acid amin bị phosphoryl hoá bởi MAP kinase-kinase). Bản thân MAP kinase-kinase có hoạt tính khi các serine/threonine của nó được gắn gốc phosphate. Phản ứng đó được xúc tác bởi MAP kinase-kinase-kinase. Đến lượt mình, MAP kinase-kinase-kinase được hoạt hoá có lẽ nhờ tương tác với protein Ras (Hình 5.23).



Hình 5.23:

Chuỗi các phản ứng dây chuyền phosphoryl hoá serine/threonine được hoạt hoá bởi protein Ras hoặc kinase C. Tín hiệu được nhận biết đầu tiên nhờ thụ thể tyrosine kinase và được truyền thông qua Ras. (MAP kinase-kinase-kinase còn được gọi là Raf khi tín hiệu truyền từ thụ thể nối protein G qua kinase C) (theo Alberts & cs., 2002).

Một khi MAP kinase hoạt động, nó truyền tín hiệu đi nhờ phosphoryl hóa các protein khác nhau gồm các kinase hoặc các protein điều khiển. Thông thường, quá trình phiên mã của một số gen xảy ra sau vài phút khi có mặt các factor tăng trưởng. Phức hợp các protein giữ vai trò quan trọng trong hoạt hoá các gen khi chúng liên kết với ADN đặc hiệu. Khi được hoạt hoá, MAP kinase di chuyển từ tế bào chất vào nhân và phosphoryl hoá các protein của phức hợp đó. Lúc đó phức hợp mới có khả năng tương tác với ADN và bật mở các gen.

5.8 Tyrosine kinase phối hợp với thụ thể. Thụ thể Tyrosine phosphatase

Có nhiều thụ thể trên bề mặt tế bào, ví dụ như các thụ thể đặc hiệu cho kháng nguyên trên tế bào lympho B và T, không phải là protein của kênh ion cũng không phải là thụ thể nối với protein G. Hơn nữa chúng không có vùng mang hoạt tính kinase. Chúng hoạt động nhờ phối hợp với các tyrosine kinase. Khi thụ thể liên kết với ligand, lập tức tín hiệu được truyền đến tyrosine kinase và các enzym này sẽ phosphoryl hoá các protein khác để truyền tiếp tín hiệu

đi. Những tyrosine kinase kết hợp cùng thụ thể được nghiên cứu nhiều nhất là nhóm Src hoặc Janus.

Nhóm tyrosine kinase Src gồm ít nhất 8 protein khác nhau nhưng chúng đều chứa vùng SH2, SH3 và chúng đều nằm ở phía trong màng tế bào chất. Chúng được cố định vào màng nhờ tương tác với chuỗi lipid và tương tác với thụ thể. Mỗi protein trong nhóm có thể liên kết với nhiều loại thụ thể khác nhau. Có thể là thụ thể tyrosine kinase hoặc các thụ thể không phải là tyrosine kinase. Do đó tín hiệu bắt đầu từ các thụ thể này có thể thông qua Src được truyền đi theo nhiều con đường khác nhau.

Như trên đã nêu, tyrosine bị phosphoryl hoá lập tức lại bị khử gốc phosphate rất nhanh nhờ tyrosine phosphatase. Các enzym này chỉ khử gốc phosphate ở tyrosine của một số protein. Tính đặc hiệu đó đảm bảo nồng độ tyrosine mang gốc phosphate trong tế bào rất nhỏ. Như vậy, các enzym này có vai trò rất quan trọng trong các con đường truyền tín hiệu. Bên cạnh đó, tyrosine phosphatase còn đóng vai trò thụ thể với một số ligand đặc biệt.

Một ví dụ điển hình cho thụ thể tyrosine phosphatase là protein CD45 được tìm thấy trên bề mặt bạch cầu giữ vai trò chủ yếu trong quá trình hoạt hoá các lympho B và T bởi các kháng nguyên. CD45 là một glucoprotein nằm xuyên qua màng, vùng có hoạt tính tyrosine phosphatase nằm phía trong màng. Khi CD45 liên kết với các kháng thể (bên ngoài màng tế bào), vùng có hoạt tính được hoạt hoá, xúc tác cho phản ứng chuyển nhóm phosphate khỏi tyrosine của một số protein đặc hiệu, thường là các kinase. Nhờ đó, những protein này có hoạt tính tiếp tục gây phosphoryl hoá, truyền tín hiệu vào trong tế bào.

Chương 6

CHU TRÌNH VÀ PHÂN CHIA TẾ BÀO

Hầu hết các tế bào eukaryot đều trải qua quá trình phân chia mà kết quả là các nhiễm sắc thể được nhân đôi và phân ly về hai tế bào con giống hệt nhau. Đây là một quá trình gồm nhiều giai đoạn, tất cả đều được kiểm soát rất chặt chẽ, đảm bảo sự phát triển chính xác và đồng bộ trong một cơ thể đa bào hoàn chỉnh. Hầu hết các sai lệch không tuân thủ sự kiểm soát đều dẫn đến hiện tượng tế bào chết hoặc phát triển thành ung thư.

Sự phối hợp các kỹ thuật sinh hoá, di truyền và đặc biệt là ADN tái tổ hợp đã cho phép nghiên cứu chu trình tế bào ở các đối tượng sinh vật khác nhau và rút ra được những nét đặc trưng chung cho mọi tế bào eukaryot. Đặc biệt kết quả nghiên cứu đã xác định được các phức protein làm nhiệm vụ kiểm soát hai giai đoạn chính của chu trình phân bào là tái bản ADN (pha S) và phân chia tế bào (pha M). Trong chương này, chúng ta quan tâm đến các thí nghiệm, các kỹ thuật được áp dụng để nghiên cứu tìm hiểu bản chất và hoạt tính của các phức protein. Từ đó chúng ta hiểu được cơ chế kiểm soát chung chu trình phân bào ở tế bào eukaryot.

6.1 Những đặc tính cơ bản của chu trình tế bào

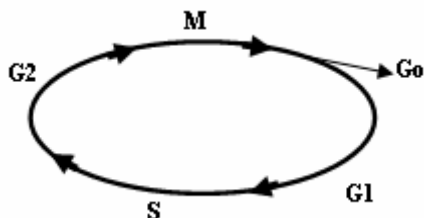
Chu trình tế bào là quá trình tế bào sinh sản bằng việc nhân đôi mọi vật chất chứa trong tế bào và phân chia thành hai tế bào mới. Đối với sinh vật đơn bào như vi khuẩn; nấm men, mỗi chu trình tế bào đều tạo thêm cơ thể mới. Sinh vật đa bào đòi hỏi rất nhiều lần phân chia tế bào để tạo nên cơ thể hoàn chỉnh. Ngoài ra, sự phân bào vẫn luôn xảy ra trong cơ thể trưởng thành để đổi mới hoặc thay thế các tế bào bị tổn thương hoặc bị chết (do tác động bên ngoài hay do chính chương trình gây chết tế bào được lập trình sẵn trong bộ máy di truyền). Mỗi một giây, cơ thể chúng ta tạo ra hàng triệu tế bào mới để duy trì trạng thái cân bằng. Vì một lý do nào đó, ví dụ như cơ thể bị chiếu xạ mạnh, phân chia tế bào bị ngừng sẽ dẫn đến tử vong trong vài ngày.

Chu trình tế bào xảy ra ở bất cứ sinh vật nào cũng đòi hỏi sự nhân đôi nhiễm sắc thể và các bào quan. Các quá trình xảy ra trong nhân (nhân đôi ADN) và trong tế bào chất (tạo các bào quan mới) phối hợp chặt chẽ nhịp nhàng trong suốt một chu trình. Cơ chế nào và yếu tố gì quyết định điều đó?

Trước đây các hiện tượng liên quan đến chu trình phân chia tế bào như nhân đôi ADN, di chuyển các nhiễm sắc thể về hai cực phân bào đã được nghiên cứu khá kỹ về mặt tế bào học. Trong những năm gần đây, các nhà khoa học quan tâm đến hệ thống kiểm soát chu trình tế bào (*cell-cycle control system*). Kết quả phân tích các protein tham gia hệ thống này cho thấy chúng được bảo toàn trong quá trình tiến hoá. Một thành phần của hệ thống kiểm soát ở nấm men có thể thực hiện chức năng của mình một cách hoàn hảo trong các cơ thể khác như ở người và ngược lại. Chúng ta xem xét đến hệ thống kiểm soát chu trình tế bào và sự phối hợp giữa các phản ứng sinh học để thực hiện sự kiểm soát đó.

Chu trình tế bào ở sinh vật eukaryot được xem gồm 4 giai đoạn hay còn gọi 4 pha (Hình 6.1). Ba giai đoạn G1, S và G2 được gọi chung là gian kỳ (*interphase*). Trong pha G1 ngay

sau pha M (*mitose*), tế bào hoàn thiện kích thước riêng để đảm bảo mọi điều kiện cần thiết tiếp tục pha S (nhân đôi ADN). Ở pha G2 sau pha 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 pha M.



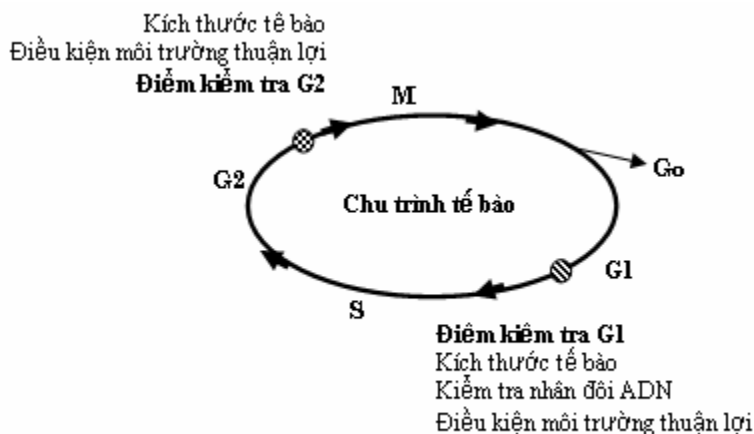
Hình 6.1:

Chu trình tế bào eukaryot gồm 4 pha M, G1, S và G2. Tế bào phát triển trong giai đoạn gian kỳ (gồm các pha G1, S, G2) và phân chia trong pha M.

Chu trình tế bào xảy ra nhanh hay chậm phụ thuộc vào từng loại tế bào ngay trong một cơ thể. Ở động vật có vú, với những mô sinh trưởng nhanh, toàn bộ chu trình phân bào chỉ xảy ra trong khoảng 12 đến 24 giờ. Đặc biệt chu trình phân bào xảy ra rất nhanh ở giai đoạn phát triển phôi sớm (khoảng 8 đến 60 phút). Lúc đó hai pha G1 và G2 hầu như không xảy ra và toàn bộ chu trình tế bào chỉ có hai pha S và M. Khi không có nhu cầu phân chia liên tục, tế bào ngừng lại ở pha G1 và sau đó bước sang một trạng thái đặc biệt gọi là G0. Ở pha G0, tế bào giữ nguyên trạng thái trong vài ngày, vài tuần, vài năm thậm chí trong suốt cả thời gian sống.

Chu trình tế bào có những điểm then chốt - gọi là các điểm kiểm tra (*checkpoints*) quyết định sự chuyển tiếp giữa các pha. Toàn bộ chu trình được điều khiển bởi trung tâm kiểm soát. Trung tâm có nhiệm vụ xử lý tín hiệu diễn biến trong từng pha gửi về. Tại các điểm kiểm tra, tín hiệu phản hồi từ pha trước đó sẽ được trung tâm phân tích để quyết định tế bào có chuyển sang pha tiếp theo hay không. Chúng ta cần phân biệt sự khác nhau giữa các thành phần chủ chốt tham gia trung tâm kiểm soát với toàn bộ nguyên liệu cần thiết cấu thành nên bộ máy để thực thi các giai đoạn cơ bản của chu trình. Ví dụ, các phức của trung tâm kiểm soát có chức năng điều hành diễn biến của pha S sẽ khác với các enzym, các protein tham gia quá trình tái bản ADN ở giai đoạn này. Điều này tương tự như trong một doanh nghiệp cổ phần: hội đồng quản trị kiểm soát chung hoạt động của ban giám đốc; ban giám đốc điều hành cụ thể các mục tiêu do hội đồng quản trị đề ra.

Hệ thống kiểm soát bao gồm một số protein giữ vai trò trọng yếu trong từng pha và phối hợp giữa các pha, đặc biệt ở các điểm kiểm tra. 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 6.2).



Hình 6.2:

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.

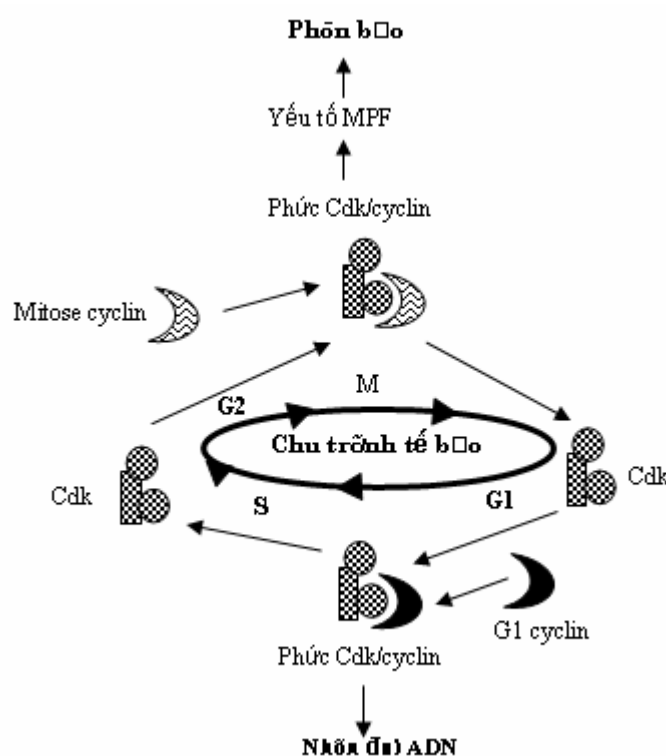
Các thí nghiệm đầu tiên dựa vào hiện tượng dung hợp các tế bào động vật nuôi cấy đã cho phép xác định các yếu tố quan trọng của hệ thống kiểm soát. Tế bào nuôi cấy ở các pha khác nhau có thể dung hợp với nhau tạo tế bào mới. Ví dụ, tế bào ở gian kỳ có thể dung hợp với tế bào ở pha M. Lúc đó trong tế bào dung hợp, nhiễm sắc thể từ tế bào ở gian kỳ trở nên co đặc lại, màng nhân bao bọc xung quanh bị vỡ. Hiện tượng này xảy ra tương tự như nhân tế bào chuẩn bị bước vào phân chia. Như vậy trong tế bào chất có tồn tại yếu tố gây ra những biến động trên. Một loạt các thí nghiệm dung hợp tế bào ở các pha khác nhau, kết hợp với nghiên cứu chu trình phân bào ở nấm men và kỹ thuật phân tích noãn bào trứng ếch đã phát hiện ra các phức protein của hệ thống kiểm soát phân bào.

Hệ thống kiểm soát phân bào hoạt động dựa vào phức dị dimer (*heterodimer*) gồm hai họ protein giữ vai trò chủ đạo. Đó là các protein kinase phụ thuộc cyclin (*cyclin-dependent protein kinase*), viết tắt là **Cdk** và nhóm protein cyclin. Cdk hoạt hoá chuỗi các phản ứng nhờ khả năng 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 kinase đặc hiệu, tức là quyết định protein riêng biệt sẽ bị phosphoryl hoá bởi phức. 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 kết quả nghiên cứu cho thấy hệ thống kiểm soát gắn liền với sự hình thành và hoạt hoá phức Cdk-cyclin ở ba giai đoạn: pha G₁ (phức G₁Cdk), pha S (phức SCdk) và pha M (phức MCdk). Khi tế bào được kích thích chuẩn bị tái bản ADN thì phức G₁Cdk hoạt động nhằm hoạt hoá các yếu tố phiên mã (*transcription factor*) đối với những gen mã cho các enzym cần cho tái bản ADN và các protein cần cho tạo phức SCdk. Khi mới được hình thành, các phức SCdk được giữ ở dạng không có hoạt tính bởi các protein ức chế. Chỉ đến cuối pha G₁, protein ức chế bị phân huỷ, phức SCdk được giải phóng để hoạt hoá tế bào bước sang pha S. Một cách tương tự, phức MCdk được tổng hợp trong pha S và trong pha G₂, nhưng chúng chỉ có hoạt tính sau khi toàn bộ nhiễm sắc thể đã tái bản trọn vẹn. Các phức MCdk thúc đẩy quá trình co đặc của nhiễm sắc thể, phá vỡ màng nhân, xuất hiện tâm động, thoi vô sắc và sắp xếp các nhiễm sắc thể ở mặt phẳng xích đạo. Chỉ khi hoạt tính của phức MCdk bị giảm thì nhiễm sắc thể mới giãn ra, màng nhân xuất hiện xung quanh hai nhân con và tế bào phân chia thành hai tế bào mới.

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 bào Mitosis. Tiến trình này chỉ xảy ra theo một chiều mà không thể đảo ngược lại. Ở cơ thể bậc cao, yếu tố ngoại bào (*mitogen*) kích thích tế bào bắt đầu chu trình phân chia bằng việc tổng hợp và hoạt hoá các phức G1Cdk.

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*) (Hình 6.3). 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.



Hình 6.3:

Các thành phần cơ bản (Cdk và các loại cyclin) của hệ thống kiểm soát chu trình phân bào. Hoạt tính của Cdk phụ thuộc vào sự có mặt của cyclin (theo Alberts & cs., 2002).

Ở tế bào động vật có vú, hoạt động của hệ thống kiểm soát ở điểm G2 được nghiên cứu khá chi tiết. Thực nghiệm nhận thấy nồng độ mitose cyclin tăng dần trong pha G2. Cyclin này liên kết với Cdk tạo phức được gọi là yếu tố khởi động phân bào MPF (*Mitosis Promoting Factor*). Tuy nhiên, phức MPF ở dạng không hoạt tính. Phức chuyển sang dạng hoạt động khi bị phosphoryl hoá tại một số vị trí đặc hiệu. Một khi xuất hiện phức MPF ở dạng hoạt động, nồng độ MPF có hoạt tính tăng nhảy vọt. Sự thay đổi đột ngột hàm lượng MPF liên quan đến cơ chế điều khiển phản hồi tích cực (*positive feedback mechanism*). Điều này có nghĩa, sự xuất hiện các enzyme kinase hoặc phosphatase (được hoạt hoá bởi dạng hoạt động của MPF) sẽ thúc đẩy sự chuyển trạng thái của MPF. Do đó, khi xuất hiện dạng hoạt động MPF thì lượng enzyme được hoạt hoá tăng khiến cho lượng MPF chuyển từ dạng không hoạt tính sang

dạng hoạt động cũng tăng theo. Nói cách khác, càng có nhiều MPF ở trạng thái hoạt động thì lượng enzym có hoạt tính gây hoạt hoá MPF càng tăng. Khi nồng độ MPF dạng hoạt động đạt đến một giá trị ngưỡng, một loạt các phản ứng sinh học xảy ra khiến tế bào bước sang pha M. Khi chu trình tế bào ở biên giới giữa metaphase và anaphase, MPF bị mất hoạt tính đột ngột do mitose cyclin bị khử gốc phosphate. Lập tức tế bào ra khỏi pha M. Như vậy, phản ứng chuyển giữa hai trạng thái hoạt động và không hoạt động của phức cycline-Cdk trong hệ thống kiểm soát tương ứng với chuyển pha trong chu trình tế bào. Cơ chế hoạt động của hệ thống kiểm soát tại điểm G1 chưa được nghiên cứu kỹ như ở điểm G2. Tuy nhiên, những nguyên tắc chính đều được tuân thủ ở cả hai điểm này. Điều đáng lưu ý là các phản ứng xảy ra sau các điểm kiểm tra G1 (trong pha S) và G2 (trong pha M) hoàn toàn khác nhau.

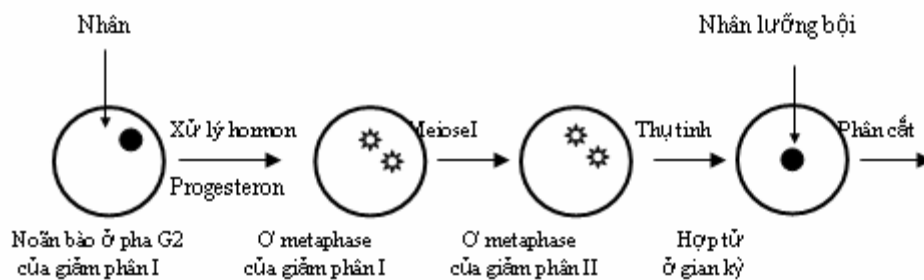
6.2 Chu trình tế bào ở giai đoạn phát triển phôi sớm

Để bước vào phân chia, tế bào phải phát triển đạt kích thước nhất định, số lượng ADN cũng như mọi thành phần khác trong tế bào chất phải được nhân đôi. Thời gian hoàn thiện mọi công tác chuẩn bị này lâu hay nhanh phụ thuộc vào môi trường xung quanh. Do đó các cơ chế kiểm soát chu trình tế bào phải đảm bảo sao cho mọi chỉ tiêu cần thiết được thực hiện trọn vẹn, chính xác theo đúng trật tự trước khi tế bào bước vào phân chia tạo thành hai tế bào con.

Trường hợp đặc biệt tế bào phân chia không trải qua đủ bốn pha G1, S, G2 và M xảy ra ở giai đoạn phát triển phôi sớm. Lúc đó tế bào phân chia liên tục mà không tuân thủ qui định chặt chẽ của các cơ chế kiểm soát. Trong giai đoạn này phân bào xảy ra nhanh và hầu như không dừng lại ở các điểm kiểm tra G1 và G2. Thực chất chu trình tế bào lúc này chỉ thực hiện nhiệm vụ cơ bản nhất là nhân đôi genome và phân ly chúng về hai tế bào con. Chúng ta hãy xét xem yếu tố nào giữ vai trò quan trọng trong hệ thống kiểm soát chu trình tế bào đối với phát triển phôi cũng như hoạt động của chúng liên quan như thế nào đến các giai đoạn đặc thù của phân bào.

Các kết quả đạt được khá chi tiết khi nghiên cứu sự phát triển của trứng ếch sau thụ tinh. Trứng ếch là một tế bào khổng lồ hình cầu, đường kính đạt tới vài mm và chứa lượng tế bào chất nhiều gấp 100.000 lần so với các tế bào bình thường trong cơ thể. Noãn bào phải trải qua phân bào giảm nhiễm để phát triển thành trứng chín (trứng có khả năng thụ tinh). Như chúng ta đã biết, noãn bào ếch dừng lại ở pha G2 hay còn gọi là prophase của giảm nhiễm (*meiosis I*). Điểm dừng này nằm trước pha M và tương ứng với điểm kiểm tra G2 trong chu trình phân bào thông thường. Để phát triển thành trứng chín, hormon sẽ tác động vào noãn bào khiến nó vượt qua điểm dừng G2 để bước vào phân chia (*meiosis II*). Lúc đó trứng chín (Hình 6.4). Sau khi thụ tinh, tế bào trứng phân cắt liên tục mà không phát triển về kích thước tạo hàng nghìn tế bào mới ngày càng nhỏ. Lần phân cắt đầu tiên kéo dài khoảng 90 phút, tiếp theo là 11 lần phân cắt (30 phút /lần) tạo 4096 tế bào mới (2^{12}) trong vòng khoảng 7 giờ. Trong mỗi lần phân cắt, pha M chiếm 15 phút và pha S chiếm 15 phút, còn hai pha G1 và G2 không xảy ra.

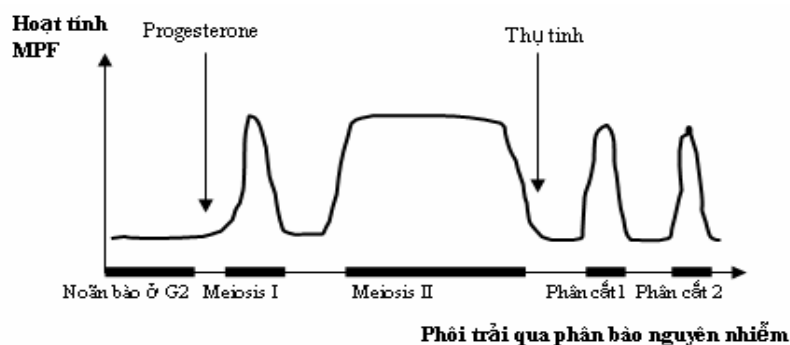
Kỹ thuật phân tích noãn bào ếch (*xenopus oocyte assay technique*) cho phép lấy tế bào chất của trứng chín chưa thụ tinh tiêm vào noãn bào. Lập tức noãn bào không dừng ở pha G2 mà chuyển sang pha M và phát triển thành trứng chín. Rõ ràng không cần kích thích bởi hormon progesteron mà vẫn tạo được trứng chín. Như vậy, phân chia tế bào trứng ếch trong giai đoạn đầu tiên được kiểm soát bởi cơ chế phụ thuộc vào tế bào chất. Phương pháp này cho phép xác định được tác nhân tồn tại trong tế bào chất của trứng chín đóng vai trò quyết định tế bào chuyển sang pha M. Tác nhân này được gọi là tác nhân kích chín do nó quyết định sự thành thực của noãn bào thành trứng chín (*maturation promoting factor*).



Hình 6.4:

Sự hoạt hoá trứng ếch *Xenopus*. Xử lý với hormon progesteron khiến noãn bào thực hiện giảm nhiễm thứ nhất và phát triển thành trứng chín. Thụ tinh giúp trứng này hoàn thành lần giảm nhiễm thứ hai và bước vào chu trình phân chia đầu tiên của phôi (theo Alberts & cs., 2002).

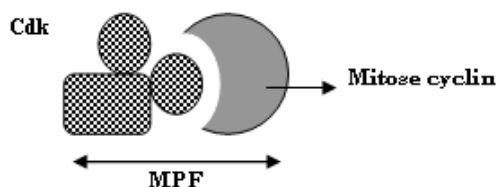
Nghiên cứu diễn biến xảy ra từ khi trứng chín đến thụ tinh và bước vào phân cắt cho thấy tác nhân kích chín có hoạt tính mạnh khi noãn bào chuyển từ G2 sang metaphase (meiosis I) và từ meiosis I sang meiosis II. Hoạt tính này giảm ngay khi trứng vừa thụ tinh nhưng sau đó lại tăng mạnh khi trứng thụ tinh bước vào lần phân cắt đầu tiên, bắt đầu quá trình phát triển phôi. Trải qua 12 lần phân cắt liên tiếp, hoạt tính của tác nhân kích chín biến đổi theo chu kỳ mà cực đại ứng với thời điểm tế bào bước vào phân bào (Hình 6.5).



Hình 6.5:

Hoạt tính MPF dao động theo giảm nhiễm (meiosis) và nguyên nhiễm (mitosis) trong noãn bào và hợp tử bắt đầu bước vào phân cắt.

Vai trò của tác nhân kích chín được phát hiện đầu tiên khi nghiên cứu quá trình phát triển trứng ếch sau thụ tinh. Những thí nghiệm tiếp sau đó tiến hành với các tế bào phân chia nguyên nhiễm ở các loài sinh vật khác nhau đã khẳng định tầm quan trọng của yếu tố tồn tại trong tế bào chất tham gia kiểm soát chu trình tế bào. Ví dụ, thí nghiệm với các tế bào động vật nuôi cấy bị xử lý với colchicine (chất ức chế phân bào, ngăn cản sự hình thành các sợi thoi vô sắc) cho thấy tế bào chất chiết ra từ những tế bào bị xử lý có tác dụng làm cho noãn bào trở thành trứng chín. Như vậy, yếu tố thúc đẩy phân chia nguyên nhiễm (*mitosis promoting factor-MPF*) ở tế bào soma nuôi cấy cũng có hoạt tính tương tự yếu tố khởi động chín (*maturation promoting factor*) ở tế bào noãn. Ngoài ra, cùng với phân tích noãn bào, kỹ thuật dung hợp tế bào cho phép một tế bào ở pha M được dung hợp với tế bào ở giai đoạn gian kỳ. Nhờ đó nhân tế bào ở giai đoạn gian kỳ được đưa vào tế bào chất của tế bào khác ở pha M. Kết quả cho thấy nhân bước vào giai đoạn tái bản và tế bào phân chia. Như vậy tác nhân tồn tại trong tế bào chất đã kích thích quá trình phân bào. Nói cách khác thì *mitosis promoting factor* có mặt trong tế bào chất có khả năng thúc đẩy tế bào soma chuyển từ giai đoạn gian kỳ sang pha M, tức là kích thích tế bào bước vào phân chia nguyên nhiễm. Tập hợp các kết quả nghiên cứu cho phép đưa đến kết luận *maturation promoting factor* cũng chính là *mitosis promoting factor*. Điều thú vị là cả hai khái niệm này có cùng cách viết tắt là MPF.



Hình 6.6:
Hai tiểu phần hình thành nên yếu tố MPF: Cdk và cyclin.

Kết quả tinh sạch và phân tích hoạt tính cho thấy MPF là phức có hoạt tính kinase gây phosphoryl hoá ở acid amin serine/threonine của một số protein đặc biệt. Phức MPF gồm hai tiểu phần: tiểu phần có hoạt tính enzym chính là protein kinase phụ thuộc cyclin và tiểu phần mitose-cyclin kiểm soát hoạt tính kinase (Hình 6.6). Phức MPF có hoạt tính cực đại trong pha M và cực tiểu ở gian kỳ đối với mọi tế bào eukaryot từ nấm men đến động vật có vú. Vì vậy, điều đáng quan tâm trước hết là biết được thời điểm xuất hiện và vai trò cụ thể của mitose cyclin trong việc kiểm soát hoạt tính của phức MPF.

6.3 Protein cyclin

Protein cyclin liên quan đến pha M trong quá trình phân bào được tinh sạch lần đầu tiên từ phôi cầu gai (*sea urchin embryo*). Trứng cầu gai vừa thụ tinh được ủ với acid amin đánh dấu đồng vị phóng xạ. Protein tổng số được tách chiết ở các thời điểm cách đều nhau (10 phút) tính từ khi bắt đầu ủ. Điện di đồng thời các mẫu đó có thể phát hiện được những protein mới và diễn biến nồng độ của chúng trong giai đoạn phát triển phôi sớm. Trong số những protein này chỉ có một protein có nồng độ biến đổi theo chu trình phân chia tế bào: nồng độ cực đại khi bắt đầu pha M và đột ngột giảm xuống cực tiểu trong giai đoạn anaphase. Protein đặc biệt này được gọi là **cyclin B**.

Kỹ thuật tách dòng và tái tổ hợp ADN đã phân lập được ADNc tương ứng với gen mã cho cyclin B ở cầu gai. Đoạn ADNc này được dùng làm đầu dò để lai Southern với ngân hàng ADNc của trứng ếch. Nhờ đó đã phân lập được dòng ADNc mã cho cyclin B của *Xenopus*. Khi đưa ADNc này vào vector biểu hiện thì thu được protein. Đưa protein vào thử để tạo kháng thể. Điều thú vị là kháng thể này tương tác với MPF tách chiết từ trứng ếch trong phép lai Western. Như vậy, cyclin B chính là tiểu phần của phức MPF. Vậy thì cyclin B kiểm soát hoạt tính của MPF như thế nào?

Thực nghiệm tiến hành thất trứng ếch làm hai phần ngay sau khi trứng vừa thụ tinh chưa bước vào lần phân cắt thứ nhất: một phần có nhân và phần kia chứa tế bào chất. Phần có nhân sẽ tiếp tục phân cắt liên tục tạo ra các tế bào nhỏ dần trong khi phần thứ hai cũng trải qua các biến động có tính chu kỳ. Điều này cũng được quan sát thấy rất rõ ở trứng cầu gai sau khi thụ tinh bị lấy nhân đi nhưng vẫn tiếp tục phân cắt liên tiếp tạo các tế bào nhỏ hơn và tất cả chúng đều không có nhân. Như vậy, yếu tố thúc đẩy phân bào tồn tại trong tế bào chất và hoạt tính của nó không phụ thuộc vào nhân (ADN). Vậy thì yếu tố này được dự trữ trong trứng trước khi thụ tinh hay là được tổng hợp mới trong giai đoạn phát triển phôi sớm?

Phân tích dịch chiết từ trứng ếch chưa thụ tinh, thực nghiệm phát hiện được protein non-histone, histone, các enzym liên quan đến tái bản ADN và mọi thành phần cần thiết để tổng hợp protein như ribosome, ARNm. Vì tế bào có thể phân chia khi không có nhân, tức là tế bào có thể không sử dụng đến các thành phần cần thiết cho tái bản ADN. Vậy thì tế bào có sử dụng các thành phần liên quan đến quá trình tổng hợp protein khi phân chia hay không?

Thí nghiệm ủ nhiễm sắc thể của tinh trùng (n) ở pha gian kỳ với dịch chiết từ trứng ếch chưa thụ tinh cho phép quan sát một loạt diễn biến xảy ra với bộ nhiễm sắc thể đơn bội. Đầu tiên, màng nhân hình thành xung quanh nhiễm sắc thể tạo thành nhân đơn bội. Tiếp đến, nhiễm sắc thể co đặc lại, màng nhân bị vỡ ra và ADN tinh trùng tái bản thành lưỡng bội (2n). Sau đó, tất cả cyclin B trong dịch chiết đột ngột bị phân hủy tương tự như tế bào ở giai đoạn anaphase (*mitose*). Lúc này nhiễm sắc thể giãn ra và màng nhân lại xuất hiện. Cả quá trình này được lặp lại đều đặn với chu kỳ khoảng 20 phút. Cuối mỗi chu kỳ, cyclin B bị phân hủy. Vào đầu chu kỳ tiếp theo, lượng cyclin B tăng dần do được tổng hợp mới từ ARNm tương ứng đã dự trữ sẵn trong trứng. Khi hàm lượng cyclin B đạt đến giá trị cực đại thì nhiễm sắc thể bắt đầu co đặc, màng nhân đứt gãy vv... Rõ ràng tế bào đã thực hiện phản ứng tổng hợp mới protein từ ARNm để cung cấp cyclin B cần thiết cho phân chia. Sự tăng hàm lượng cyclin B tỷ lệ với sự tăng hoạt tính của MPF. Khi sử dụng cycloheximide, chất ức chế quá trình tổng hợp protein thì cyclin B không được tạo thành. Đồng thời cũng không phát hiện được hoạt tính của MPF và không quan sát được hiện tượng nhiễm sắc thể co đặc cũng như màng nhân bị đứt gãy. Ngoài ra, khi xử lý dịch chiết từ trứng chưa thụ tinh với RNase ở nồng độ thấp nhằm phân hủy toàn bộ ARNm mà không làm ảnh hưởng đến ARNr và ARNt (hai loại ARN này chỉ bị phân hủy bởi RNase nồng độ cao), thì màng nhân vẫn xuất hiện bao quanh nhiễm sắc thể tinh trùng và ADN vẫn được tái bản. Tuy nhiên, sau đó các diễn biến tương tự như phân chia tế bào không lặp lại nữa. Nếu bổ sung thêm ARNm mã cho cyclin B vào thì các chu kỳ biến động tuần hoàn với thời gian 20 phút lại diễn ra. Kết quả này chứng tỏ chỉ có một loại ARNm duy nhất có mặt trong thí nghiệm (*ARNm cyclin B*), được dùng làm khuôn để tổng hợp cyclin B. Ở cuối phân bào mitose, cyclin B bị biến tính thông qua ubiquitin, tức là cyclin B có mang đoạn peptide tương tác với ubiquitin tạo tín hiệu nhận biết cho protease. Nếu trong thí nghiệm trên, bổ sung ARNm cyclin B có mang đột biến ở phần mã cho tín hiệu nhận biết sao cho cyclin B không bị gắn ubiquitin, thì cyclin B không bị phân hủy bởi protease và quan sát thấy nhiễm sắc thể không giãn ra và màng nhân không xuất hiện. Rõ ràng, cyclin B được tổng hợp mới (*de novo*) có vai trò quyết định hoạt tính của MPF. Hàm lượng cyclin B nhiều hay ít kéo theo sự tăng hay giảm hoạt tính của MPF. Như vậy cyclin B quyết định hoạt tính của MPF và kiểm soát mọi diễn biến xảy ra ngay sau khi trứng thụ tinh.

Những kết quả thu được với các tế bào động vật khác đã tìm được ít nhất ba loại cyclin có chức năng tương tự cyclin B ở trứng ếch. Chúng được gọi chung là mitose cyclin. Sự tổng hợp các cyclin này cũng quan trọng như sự phân hủy chúng để tế bào bước vào hoặc ra khỏi pha M. Thông thường, cyclin bị phân hủy đột ngột ở bước chuyển tiếp giữa metaphase và anaphase. Khi cyclin bị đột biến sao cho nó vẫn còn khả năng kích thích hoạt động của MPF nhưng không bị phân hủy do thiếu đoạn peptide đặc hiệu để tương tác với ubiquitin thì tế bào bước vào pha M nhưng không thể chuyển từ pha M sang G1. Rõ ràng, hoạt hoá phức MPF để bước vào pha M đòi hỏi tổng hợp cyclin nhưng bất hoạt phức MPF để bước sang pha G1 lại yêu cầu phân hủy cyclin. Cyclin là thành phần quan trọng của MPF có mặt trong mọi tế bào eukaryot. Có nhiều loại protein cyclin được mã bởi các gen trong cùng một họ gen.

Để điều khiển tế bào bước vào pha M, phức MPF gây ra nhiều thay đổi như nhiễm sắc thể co đặc lại, màng nhân bị phá vỡ và kết cấu khung tế bào phải được tổ chức lại để tạo thoi phân bào. Tất cả những thay đổi này phụ thuộc vào hoạt tính kinase của MPF gây phosphoryl hoá một số thành phần cơ bản tạo nên cấu trúc tế bào. Ví dụ, cấu trúc nhân bị phá vỡ khi protein lamin của nhân bị phân rã. Phản ứng này được xúc tác bởi MPF gây phosphoryl hoá serine của lamina. Khi lamina không có các serine đặc biệt này (do gen mã cho lamin bị đột biến), chúng sẽ không bị phân hủy khi tế bào ở mitose, mặc dù màng nhân đứt gãy tạo ra các nang và các diễn biến trong pha M vẫn xảy ra bình thường.

Protein histone H1 có chức năng "đóng gói" ADN thành cấu trúc nucleosome cũng bị phosphoryl hoá bởi MPF. Có lẽ phản ứng này dẫn đến việc nhiễm sắc thể co đậm lại. Ngoài ra MPF còn gây phosphoryl hoá một số protein tương tác với sợi vi ống microtubule (dùng để tạo thoi phân bào). Ở nấm men, Cdk còn được gọi là Cdc (*cell-division-cycle*). Cả hai thành phần mitose cyclin và Cdc2 kinase của MPF đều có thể liên kết với tâm động. Có thể mitose cyclin nhận biết được các thành phần của tâm động và giúp cho Cdc2 liên kết đúng vị trí.

Cyclin B và các cyclin chức năng tương tự như cyclin B bị phân hủy ở phase muộn anaphase khi các nhiễm sắc tử đã tách nhau ra và được kéo về hai cực. Phía đầu NH₂ của hầu hết cyclin này đều có đoạn peptide nhận biết bởi ubiquitin. Đoạn này được gọi là hộp phân hủy (*destruction box*). Nồng độ cyclin B giảm khiến tế bào nhanh chóng ra khỏi pha M.

6.4 Nấm men và hệ thống kiểm soát chu trình tế bào

Những thí nghiệm tiến hành trên trùn ếch làm sáng tỏ vai trò của cyclin B trong phức MPF. Tuy nhiên cyclin chỉ là tiểu phần làm nhiệm vụ điều khiển còn hoạt tính của phức do tiểu phần protein kinase đảm nhiệm. Những thí nghiệm nghiên cứu bản chất các kinase và sự thay đổi hoạt tính của chúng trong chu trình tế bào, đặc biệt ở pha S và M được tiến hành trên nấm men nảy chồi *S.cerevisiae* và nấm phân chia *S.pombe*. Tế bào *S.cerevisiae* dạng hình cầu, sinh sản bằng phương thức nảy chồi. Tế bào con hình thành nhỏ hơn tế bào mẹ, do đó chúng phải tăng trưởng để đạt đến kích thước nhất định trước khi tiếp tục phân chia. Tế bào *S.pombe* có dạng hình que với chiều dài tăng dần trong quá trình phát triển. Khi bước vào phân chia, tế bào sẽ phân chia làm hai phần có kích thước như nhau. Dung hợp tế bào và phân tích noãn bào là các kỹ thuật chính được áp dụng để nghiên cứu đối với trùn ếch, trong khi nghiên cứu ở nấm men liên quan đến những kỹ thuật di truyền nhằm sàng lọc và phân tích các dạng đột biến không có khả năng vượt qua các điểm kiểm tra trong chu trình tế bào.

Khác với các tế bào ở giai đoạn phát triển phôi sớm, tế bào nấm men cũng giống như đa số các tế bào trong cơ thể eukaryot đều phải phát triển đạt kích thước nhất định mới bước vào phân bào mitose. Thời gian tế bào tồn tại ở giai đoạn gian kỳ chiếm phần lớn chu trình tế bào. Đặc biệt ở hai pha G1 và G2, kiểm soát chu trình tế bào phải đảm bảo đồng bộ giữa tốc độ phát triển và thời gian thực hiện một chu trình phân bào. Hệ thống kiểm soát sẽ kiểm tra kích thước tế bào tại một số vị trí đặc biệt (*size check point*) thường phân bố ở hai pha G1 và G2. Chu trình tế bào sẽ bị giữ lại tại những điểm này để tế bào đạt kích thước tới hạn hoặc genome được tái bản hoàn chỉnh. Kiểm tra ở điểm G1 quyết định ADN có được tái bản hay không. Tương tự như thế, kiểm tra ở điểm G2 cho phép tế bào bước vào phân chia. Tùy thuộc vào từng loại tế bào và điều kiện môi trường mà xu hướng giữ tế bào ở điểm G1 chiếm ưu thế hơn ở điểm G2 hoặc ngược lại. Cơ chế kiểm tra tại điểm G1 và điểm G2 được sáng tỏ nhờ áp dụng các kỹ thuật di truyền phân tử ở nấm nảy chồi *S.cerevisiae* và nấm phân chia *S.pombe*.

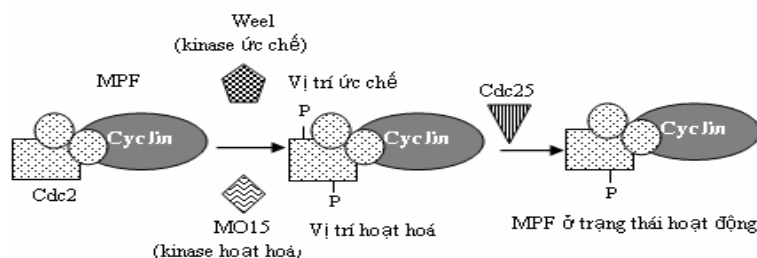
Để nghiên cứu chu trình tế bào và các cơ chế kiểm soát phân bào, nấm men được chọn làm đối tượng điển hình để tìm ra các đột biến gen mã cho các sản phẩm tham gia hệ thống kiểm soát cũng như bộ máy phân bào. Khi một gen thuộc hệ thống kiểm soát phân bào bị đột biến, tế bào không thể thực hiện trọn vẹn một chu trình phân bào nên số lượng tế bào không tăng. Do đó, để có thể chọn và duy trì được các tế bào này, các đột biến chỉ được biểu hiện trong những điều kiện đặc biệt. Hầu hết đó là những đột biến nhạy cảm nhiệt độ (*temperature-sensitive mutant*). Chúng chỉ biểu hiện tính trạng đột biến khi nhiệt độ cao (*restrictive condition*), còn ở nhiệt độ thấp (*permissive condition*) thì tế bào vẫn phân chia và phát triển bình thường.

Chủ yếu các nghiên cứu ở nấm phân chia *S.pombe* tập trung vào hai dạng đột biến nhạy cảm nhiệt độ. Dạng thứ nhất gồm những gen mà khi bị đột biến sẽ làm cho tế bào dừng lại ở tại một điểm trong chu trình tế bào. Chúng được gọi chung là các gen *cdc* (*cell-division-cycle*). Dạng thứ hai gồm những gen mà khi bị đột biến sẽ cho phép tế bào vượt qua các điểm kiểm tra kích thước để bước vào phân bào. Như vậy tế bào sẽ phân chia ngay khi kích thước còn chưa đạt yêu cầu bình thường. Các gen này được gọi chung là *wee* (theo tiếng Scotlen có nghĩa là nhỏ). Hai nhóm gen *cdc* và *wee* có chức năng trái ngược nhau: đột biến gen *cdc* khiến tế bào không qua được điểm kiểm tra trong khi đột biến gen *wee* giúp tế bào vượt qua một cách dễ dàng. Vì vậy, các gen *cdc* thúc đẩy tế bào vượt qua các điểm kiểm tra còn các gen *wee* kìm hãm quá trình này.

Điểm kiểm tra kích thước tế bào nấm men nằm trong pha G2 (điểm bước vào mitose) tương tự như điểm kiểm tra G2 tại đó MPF được hoạt hoá trong chu trình phân bào ở giai đoạn phát triển sớm của phôi *Xenopus*. Ở nấm men *S. pombe*, tiểu phần có hoạt tính kinase tương tự như Cdk của MPF. Tiểu phần này được mã bởi gen *cdc2*. Protein Cdc2 giữ vai trò quyết định nhất, thúc đẩy tế bào nấm men bước vào mitose. Khi Cdc2 bị mất hoạt tính thì không xảy ra quá trình phân bào. Tuy nhiên, khi Cdc2 không bị kiểm soát bởi cyclin khiến nó luôn ở trạng thái hoạt động thì tế bào phân chia liên tục. Gen *cdc2* được tách dòng từ *S. pombe* có khả năng phục hồi kiểu hình hoang dại (*wild type*) cho dạng đột biến *cdc2⁻* trong thí nghiệm bổ sung chức năng. Hơn nữa, các gen tương đồng với *cdc2* phân lập được từ một số sinh vật eukaryot khác cũng có khả năng phục hồi tế bào nấm men bị đột biến *cdc2⁻*. Khi chuyển gen tương đồng với *cdc2* phân lập từ người vào tế bào nấm men bị đột biến gen *cdc2⁻* thì các tế bào này lại sống sót và phân chia như tế bào bình thường. Như vậy hoạt động của các yếu tố kiểm soát phân bào rất giống nhau giữa nấm men và động vật có xương sống. Rõ ràng ở cả nấm men và động vật có xương sống, việc tế bào bước vào phân chia hoàn toàn phụ thuộc vào các kinase và cyclin. Điều này chứng tỏ cơ chế kiểm soát tế bào bước vào pha M được bảo toàn trong tiến hoá đối với mọi tế bào eukaryot.

Protein Cdc2 có hoạt tính kinase. Hoạt động của protein Cdc2 được kiểm soát bởi ba protein khác (mã bởi các gen *wee1*, *cdc13* và *cdc25*). Trong số đó, protein Cdc13 có độ tương đồng cao với mitose cyclin, ví dụ như cyclin B của *Xenopus*. Thực nghiệm đã chứng tỏ rằng Cdc2 và Cdc13 là hai tiểu phần của MPF ở nấm men *S.pombe*.

Phân tích các đột biến *cdc⁻* và *wee⁻* khác cho thấy hoạt tính của phức MPF (Cdc2-Cdc13) ở *S.pombe* còn phụ thuộc vào một số protein kinase khác. Ví dụ, khi Cdc2 tương tác với Cdc13, Cdc2 trở thành cơ chất cho hai protein kinase Wee1 và MO15. Hai protein này điều khiển hoạt tính kinase của chính Cdc2. Kinase MO15 gắn nhóm phosphate vào threonine (acid amin 161) nằm trên phân tử Cdc2. Chỉ khi bị phosphoryl hoá ở vị trí này thì phức MPF mới có hoạt tính cực đại. Ngược lại, kinase Wee1 gắn nhóm phosphate vào tyrosine (acid amin 15) trên Cdc2 gây ức chế hoạt tính kinase của Cdc2 và ngăn cản phức Cdc2-Cdc13 chuyển sang trạng thái hoạt động, ngay khi vị trí 161 đã bị phosphoryl hoá. Điều này có nghĩa khi bị phosphoryl hoá ở hai vị trí 15 và 161 thì phức MPF sẽ luôn trạng thái không có hoạt tính. Chính protein Cdc25 (ở nấm) làm nhiệm vụ khử gốc phosphate ở tyrosine, giúp cho phức có hoạt tính thúc đẩy tế bào bước vào phân chia. Những thay đổi cấu hình không gian của Cdc2 được mô tả trong Hình 6.7.



Hình 6.7:

Hoạt tính của MPF. Tương tác với cyclin cho phép Cdc2 bị phosphoryl hoá bởi Wee 1 và MO15. MPF ở dạng không hoạt tính khi bị phosphoryl hoá bởi Wee1. MPF chuyển sang dạng hoạt động nhờ kinase Cdc25. Đến lượt mình, các phân tử MPF có hoạt tính sẽ gây hoạt hoá Cdc25 và ức chế Wee1, nhờ đó chúng tăng cường hoạt động của chính mình (theo Alberts & cs., 2002).

Đối với nấm men *S.cerevisiae*, tế bào sẽ vượt qua điểm kiểm tra G1 để chuyển sang pha S một khi thoả mãn ba điều kiện: đảm bảo kích thước tế bào, môi trường có đủ các chất dinh dưỡng và tế bào có khả năng bước vào sinh sản hữu tính. Các gen khi đột biến làm tế bào không thể bước sang pha S ngay khi có đủ ba điều kiện cũng được gọi là gen *cdc*. Điều đáng lưu ý là các gen không đột biến ở *S. pombe* được ký hiệu bằng chữ thường in nghiêng (*cdc2*) nhưng ở *S.cerevisiae* thì được viết bằng chữ hoa in nghiêng (*CDC28*). Gen *CDC28* có độ tương đồng rất cao với *cdc2*. Điều lý thú là *cdc2* quyết định tế bào *S.pombe* vượt qua điểm kiểm tra G2 chuyển sang pha M, thì *CDC28* quyết định tế bào *S.cerevisiae* vượt qua điểm kiểm tra G1 (start) để chuyển sang pha S. Đột biến *cdc2* khiến tế bào *S. pombe* dừng lại ở G2 trong khi đột biến *CDC28* khiến tế bào *S.cerevisiae* dừng lại ở G1. Tuy nhiên thí nghiệm bổ sung chức năng (*complementary test*) cho thấy sản phẩm của *cdc2* và *CDC28* có thể thay thế cho nhau, tức là *CDC28* giúp tế bào *S.pombe* mang gen đột biến *cdc2* phân chia bình thường và ngược lại. Hệ thống kiểm soát trong hai loại tế bào này đều chỉ có một protein kinase phụ thuộc cyclin (*cdc2* hoặc *CDC28*). Mặc dù có thể thay thế cho nhau nhưng mỗi kinase có tác động khác nhau trong chu trình phân bào đối với từng loại tế bào nấm men. Đối với *S.pombe*, hệ thống kiểm soát phân bào bắt đầu sử dụng kinase Cdc2 tại điểm G2 còn ở *S.cerevisiae* trung tâm kiểm soát điều hành Cdc28 tại điểm G1. Do đó khi chuyển các tế bào đột biến nhạy cảm nhiệt độ từ điều kiện nhiệt độ cho phép sang nhiệt độ cao, trung tâm kiểm soát ở hai loại tế bào nấm men chịu tác động ở các giai đoạn khác nhau. Vì vậy, sự thiếu hụt kinase Cdc2 và Cdc28 được biểu hiện ở các pha khác nhau trong chu trình phân bào. Như vậy hai loại kinase này đều cần thiết để tế bào bước vào pha S và pha M ở *S.pombe* hoặc *S.cerevisiae*.

Để có thể điều khiển chu trình tế bào ở cả hai điểm kiểm tra G1 và G2, protein Cdc28 tương tác với các cyclin khác nhau. Ở pha G1, Cdc28 tương tác với ba loại cyclin (G1 cyclin), được gọi là Cln1, Cln2, Cln3, tạo nên phức có chức năng kích thích nhiễm sắc thể nhân đôi (tế bào bước vào pha S). Thí nghiệm đánh bật gen (*gene knockout*) cho thấy *S.cerevisiae* vẫn có thể mọc trên môi trường giàu dinh dưỡng khi chỉ mang một trong ba gen *CLN*. Đột biến *CLN3* khiến tế bào tồn tại ở pha G1 lâu hơn vài giờ, tức là bước vào pha S chậm hơn so với bình thường. Nếu cả ba gen bị đột biến thì tế bào chết. Ở giai đoạn G2, Cdc28 liên kết với các mitose cyclin (Clb1, Clb2 ...) tạo phức MPF thúc đẩy tế bào vào mitose. Các phức của Cdc28 với mỗi loại cyclin đều có khả năng phosphoryl hoá các protein khác nhau. Như vậy, hoạt tính riêng biệt của Cdc2 ở hai điểm kiểm tra phụ thuộc vào loại cyclin mà nó liên kết.

Một chu trình phân bào gồm bốn giai đoạn phải được hoàn thành trước khi tế bào bước sang chu trình mới. Các cơ chế kiểm soát phản hồi (*feedback control*) có nhiệm vụ dừng chu kỳ phân bào ở một số vị trí để tế bào hoàn thành giai đoạn trước rồi mới chuyển sang giai đoạn sau. Khi sử dụng các chất hoá học ức chế nhân đôi ADN hoặc khi gây tổn thương ADN,

cơ chế phản hồi sẽ ngăn cản tế bào phân chia. Điều đó đảm bảo an toàn cho các thể hệ tế bào tiếp theo nhận được thông tin di truyền chính xác, nguyên vẹn. Thực nghiệm cho thấy các cơ chế phản hồi tiêu cực giữ vai trò chính so với phản hồi tích cực. Điều này có nghĩa tín hiệu phản hồi gây ngừng hệ thống kiểm soát phân bào hơn là chuyển tế bào sang giai đoạn mới.

6.5 Kiểm soát phân bào ở động vật

Hầu hết những nghiên cứu hệ thống kiểm soát phân bào ở động vật được thực hiện đối với tế bào nuôi cấy. Khi thiếu các factor kích thích sinh trưởng (*mitogen*), tế bào dừng lại ở pha G₀. Nếu bổ sung các factor này vào môi trường, tế bào sẽ chuyển sang pha S sau khoảng 14-16h và hoàn thành tái bản ADN sau 6-8h. Sau đó tế bào sẽ phân chia. Tương tự như ở nấm men, chu trình phân bào ở động vật cũng tồn tại các điểm kiểm soát G₁, G₂. Nếu chuyển từ môi trường đầy đủ sang môi trường không có chất kích thích khi tế bào ở trước điểm kiểm tra G₁ thì chúng không phân chia. Tuy nhiên nếu đã vượt qua điểm G₁ thì mặc dù tồn tại trong môi trường thiếu mitogen, tế bào vẫn thực hiện trọn vẹn một chu trình phân bào.

Giống như nấm men, chu trình phân bào của tế bào động vật bậc cao cũng được điều khiển bởi các kinase tạo phức với cyclin. Nồng độ của các kinase được giữ ổn định nhưng nồng độ các cyclin thay đổi rất phức tạp ở các điểm khác nhau trong chu trình. Ngoài ra, khác với nấm men chỉ có một loại Cdc, tế bào động vật có nhiều loại kinase phụ thuộc cyclin (*cyclin dependent kinase Cdk*), ví dụ như protein tương đồng Cdk2 tham gia hệ thống kiểm soát phân bào. Trong khi phức giữa Cdc2 và cyclin B thúc đẩy tế bào động vật có xương sống bước vào mitose thì phức giữa Cdk2 và cyclin E thúc đẩy tế bào vượt qua điểm kiểm tra G₁ còn phức giữa Cdk2 và cyclin A cần thiết để hoạt hoá bộ máy tái bản ADN (bước vào pha S).

Tế bào nấm men (cơ thể đơn bào) chỉ cần một loại kinase Cdk để kiểm soát phân bào. Ngược lại trong cơ thể đa bào, chu trình phân bào đòi hỏi sự kiểm soát rất chặt chẽ nghiêm ngặt với sự tham gia của nhiều loại Cdk (Cdk1, Cdk2, Cdk3...) và cyclin khác nhau (A, B, D, E...). Protein Cdk1 được phân lập đầu tiên ở người, có độ tương đồng cao với Cdc2 ở *S. pombe*. Kinase Cdk1 có thể bổ sung chức năng cho tế bào đột biến *cdc2*. Tiếp đến, protein Cdk2 của người được tìm thấy nhờ khả năng phục hồi chức năng cho Cdc28 của *S. cerevisiae*. Ở cơ thể trưởng thành, đa số các tế bào không phân chia và giữ nguyên ở một trạng thái nhất định để thực hiện chức năng biệt hoá của tổ chức mà chúng cấu tạo nên. Nếu như điều kiện các chất dinh dưỡng đầy đủ có thể thúc đẩy các tế bào vi khuẩn, nấm phát triển sinh sản, thì trong cơ thể đa bào, không phải nguồn dinh dưỡng mà chính các tín hiệu gửi đi giữa các tế bào đóng vai trò quan trọng. Phần lớn các tín hiệu này là các yếu tố kích thích tăng trưởng (*growth factor*) được nhận biết bởi các thụ thể nằm trên bề mặt tế bào. Nồng độ của chúng rất nhỏ chỉ khoảng 10^{-9} đến 10^{-11} M. Tương tác đặc hiệu giữa tín hiệu và thụ thể có tác dụng thúc đẩy tế bào bước vào phân chia. Như vậy khác với nấm men có chu trình phân bào được kiểm soát chủ yếu bởi các cơ chế phản hồi tiêu cực, phân chia tế bào trong cơ thể đa bào được điều khiển chủ yếu bởi các cơ chế kiểm soát tích cực. Vậy thì bản chất của các tín hiệu điều khiển tích cực là gì? Cơ chế nào để chúng tác động đến phát triển và phân chia tế bào?

Yếu tố tăng trưởng đầu tiên được phân lập có nguồn gốc từ tiểu cầu (các mảnh tế bào tuần hoàn trong máu giúp máu đông cục tại vết thương). Chúng được gọi là PDGF (*platelet-derived growth factor*). Khi máu đông, các tiểu cầu tương tác với nhau tiết ra các chất dự trữ trong các nang tiết của chúng, trong đó có PDGF. Factor này đóng vai trò quan trọng kích thích tế bào phân chia trong suốt thời gian phục hồi vết thương.

PDGF chỉ là một trong số khoảng 50 protein đã biết có hoạt tính của factor kích thích tăng trưởng. Mỗi loại tế bào biệt hoá thực hiện các chức năng chuyên biệt sẽ có các thụ thể

đặc hiệu với các factor khác nhau. Tuy nhiên, PDGF có khả năng kích thích đối với nhiều loại tế bào như tế bào thần kinh, tế bào cơ trơn, các nguyên bào sợi (*fibroblast*) ... trong khi factor *erythropoietin* chỉ kích thích duy nhất các tế bào tiền hồng cầu bước vào phân chia.

Khi factor kích thích tăng trưởng tương tác với thụ thể nằm trên bề mặt tế bào thì phần thụ thể nằm phía trong màng tế bào sẽ xúc tác cho chuỗi phản ứng truyền tín hiệu. Tính phức tạp của các phản ứng dẫn truyền tín hiệu trong tế bào được thể hiện ở chỗ một tín hiệu trung gian có thể khởi động cho nhiều phản ứng xảy ra đồng thời. Mỗi phản ứng này lại gây ra các phản ứng tiếp theo khác nhau. Như vậy, hệ thống dẫn truyền một tín hiệu ban đầu tạo ra nhiều tín hiệu trung gian đan xen nhau và có tác động qua lại với nhau.

Khi có mặt factor kích thích tăng trưởng, tế bào chuyển từ Go sang phân chia. Có thể sắp xếp những gen bị hoạt hoá bởi mitogen thành hai nhóm. Nhóm thứ nhất gồm các gen trả lời tín hiệu sớm (*early-response genes*). Chúng được hoạt hoá sau khi có tín hiệu khoảng 15 phút và không đòi hỏi sự tổng hợp protein. Đó là do các factor phiên mã của những gen này đã tồn tại sẵn trong tế bào ở trạng thái Go. Tuy nhiên chúng phải được biến đổi, ví dụ như bị phosphoryl hoá, để chuyển sang dạng có hoạt tính. Một số gen thuộc nhóm thứ nhất mã cho factor phiên mã đối với gen thuộc nhóm thứ hai. Vì vậy, hoạt động của các gen trả lời chậm (*delayed-response genes*) thuộc nhóm hai phụ thuộc vào quá trình tổng hợp protein. Nhóm thứ hai thường hoạt hoá sau khi có tín hiệu khoảng 1h. Như vậy, sản phẩm của gen trả lời sớm hoạt hoá gen trả lời chậm, trong đó có các gen mã cho protein tham gia hệ thống kiểm soát phân bào như Cdk2, Cdk4, Cdk6 và cyclin D, cyclin E. Một trong những factor phiên mã liên quan chặt chẽ đến giai đoạn tế bào chuyển từ G1 sang pha S là E2Fs.

E2F là factor phiên mã của những gen mã cho một số Cdk, cyclin; các protein tham gia tổng hợp deoxyribonucleotide và tái bản ADN. Hoạt tính của factor này được kiểm soát nhờ tương tác với các protein Rb, p107 và p130. Hơn nữa, khi tương tác với Rb, E2F chuyển từ dạng có hoạt tính kích thích sang dạng có chức năng kìm hãm phiên mã do (Rb - E2F) liên kết với phức khử acetyl cho histone (khử nhóm acetyl có liên quan đến hiện tượng co đặc nhiễm sắc thể). Về phần mình, khả năng tương tác của Rb phụ thuộc vào kinase. Một khi Rb bị phosphoryl hoá, E2F có thể tồn tại ở dạng tự do để thực hiện chức năng khởi động phiên mã. Nhờ đó mà tế bào bước vào tái bản ADN.

Trong một số trường hợp ADN bị tổn thương, tế bào dừng lại ở các pha G1 hoặc G2. Dừng lại ở G1 nhằm ngăn cản việc tái bản ADN chứa các sai hỏng. Dừng lại ở G2 giúp tế bào sửa chữa các đứt gãy ADN trước khi phân ly về các tế bào con trong pha M. Cơ chế dừng tế bào ở hai pha này liên quan đến protein p53 và một số protein ức chế khối u khác (*tumor-suppressor protein*).

Trong tế bào bình thường, protein p53 rất không bền. Vì thế lượng p53 rất khó đạt đến ngưỡng để khởi động phiên mã đối với một số gen, ví dụ như gen mã cho p21^{CIP}. Đây là protein có khả năng tương tác và kìm hãm hoạt tính của hầu hết các phức Cdk-cyclin. Do đó, khi hàm lượng protein p53 ít, p21^{CIP} không được tổng hợp. Tuy nhiên, khi ADN bị tổn thương, p53 trở nên bền vững (do tác động của cơ chế chưa rõ). Lúc đó gen mã cho p21^{CIP} được hoạt hoá và sản phẩm của gen này kìm hãm hoạt tính của các phức Cdk-cyclin khiến tế bào bị dừng lại ở các điểm G1 hoặc G2 cho đến khi ADN được sửa chữa phục hồi nguyên vẹn. Nếu như ADN bị tổn thương quá nặng hoặc các cơ chế sửa chữa ADN không khắc phục được sai hỏng thì p53 tham gia thúc đẩy tế bào chết theo chương trình đã được lập sẵn trong genome (chết theo chương trình-apoptosis). Chương trình tự chết được bật mở có tác dụng loại trừ tế bào mang ADN tổn thương bước vào phân chia, ngăn cản việc khuếch đại các sai

hông. Như vậy, apoptosis làm giảm thiểu khả năng tế bào mang ADN tổn thương phát triển thành ung thư.

6.6 Vai trò của sợi vi ống tubulin trong phân bào

Phân bào nguyên nhiễm được bắt đầu bằng việc nhiễm sắc thể co đậm đặc lại trong nhân và thoi phân bào xuất hiện ở ngoài nhân (giai đoạn prophase). Thoi phân bào bao gồm chủ yếu các sợi vi ống microtubule làm nhiệm vụ đẩy trung thể (*centrosome*) ra xa nhau và kéo các nhiễm sắc thể về hai cực. Một loạt diễn biến liên quan đến việc hình thành và liên kết sợi thoi với tâm động, kéo nhiễm sắc thể về mặt phẳng xích đạo và tiếp đó đẩy chúng về các cực vv... phụ thuộc rất nhiều vào động học các phản ứng xảy ra ở hai đầu sợi vi ống.

Sợi vi ống microtubules có cấu trúc dạng ống gồm 13 sợi nhỏ kết hợp lại, mỗi sợi được cấu tạo từ các phân tử tubulin. Mỗi phân tử tubulin gồm hai tiểu phần (*heterodimer*) α -tubulin và β -tubulin. Do tính đa dạng của các tiểu đơn vị α và β -tubulin nên có nhiều loại tubulin. Các sợi vi ống rất mềm dẻo nhưng bền, phân bố thành mạng lưới trong tế bào. Chúng đảm nhận nhiều chức năng khác nhau như góp phần tạo nên khung tế bào; tham gia vận chuyển các nang, các phức protein và đặc biệt trong quá trình phân bào. Sợi vi ống có tính phân cực phụ thuộc vào động học quá trình polymer và khử polymer ở hai đầu của sợi. Một đầu của sợi [gọi là đầu (+)] có thể dài ra rất nhanh nhờ phản ứng polymer hoá với tốc độ gấp ba lần đầu kia [gọi là đầu (-)]. Tuy nhiên sợi vi ống có khả năng ngăn lại đột ngột do mất đi các phân tử tubulin. Chính hoạt tính phân hủy GTP của tiểu phần β -tubulin quyết định tính không bền của sợi. Tương tác giữa β -tubulin với GTP cần thiết để phân tử tubulin ở dạng dị hợp tử gắn được vào hai đầu của sợi microtubules (quá trình polymer hoá), trong khi phân hủy GTP cần thiết để phá vỡ liên kết giữa các tubulin và tách chúng ra khỏi sợi (khử polymer). Hơn nữa, phản ứng polymer gắn phức tubulin-GTP vào các đầu xảy ra nhanh hơn phản ứng phân hủy GTP \rightarrow GDP khi khử polymer. Ngoài ra ái lực tương tác giữa các tubulin mang GTP lớn hơn giữa tubulin mang GDP. Như vậy, chênh lệch nồng độ GTP và GDP cũng như tốc độ polymer và khử polymer ảnh hưởng rất lớn đến động học của sợi vi ống microtubules.

Sau khi đã gắn vào sợi microtubules, tiểu đơn vị α -tubulin chịu biến đổi acetyl hoá ở các acid amin lysine đồng thời loại bỏ tyrosine ở đầu carboxyl (*post-translational modification*). Các biến đổi này không xảy ra với các phân tử tubulin tự do chưa polymer hoá. Phản ứng acetyl hoá và khử tyrosine liên quan đến khả năng liên kết của microtubules với một số protein để tăng độ bền vững của sợi cũng như tạo thuận lợi cho tương tác với các thành phần khác trong tế bào. Những protein này được gọi chung là MAPs (Microtubule-Associated Proteins).

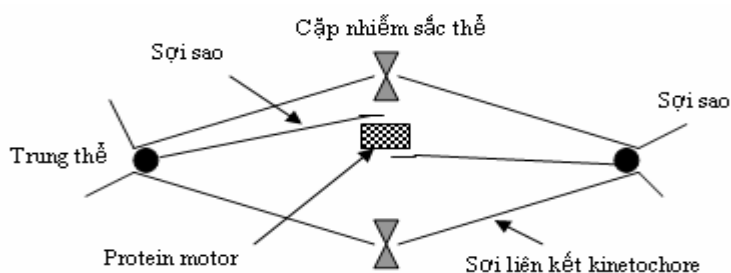
Có thể phân loại các MAPs thành hai nhóm tùy thuộc vào trọng lượng phân tử. Nhóm HMW (*High Molecular Weight*) và Tau proteins. Cấu trúc của các protein trong cả hai nhóm đều gồm hai domain; một domain liên kết với microtubules và domain kia tương tác với các thành phần khác. Khi các MAPs liên kết với đầu sợi microtubules, chúng sẽ ngăn cản quá trình khử polymer, do đó giữ cho sợi được bền vững. Ngoài ra cần phải lưu ý đến các MAPs có khả năng sử dụng năng lượng phân hủy ATP để di chuyển dọc theo sợi microtubules. Những protein MAPs đặc biệt đó được gọi chung là *microtubule motor*.

Phân bào nguyên nhiễm phải đảm bảo tuyệt đối mọi thành phần có trong tế bào được phân chia như nhau về hai tế bào con. Do các bào quan như ty thể, lục lạp và ngay cả thể Golgi, mạng lưới ER không có khả năng tự hình thành từ các nguyên liệu riêng biệt trong tế bào chất, cho nên các bào quan này không thể xuất hiện trong các tế bào con nếu như chúng

không được di truyền từ tế bào mẹ dưới dạng đặc biệt. Trong một tế bào, đa số bào quan thường tồn tại hai hay nhiều copy. Do đó trong tế bào con, các bào quan đó được duy trì nhờ nhân đôi từ một bào quan ban đầu. Đặc biệt đối với Golgi và ER, các tổ chức này bị đứt gãy thành các phần nhỏ trong phân bào và chúng được kéo về hai tế bào con nhờ các sợi vi ống của thoi phân bào.

Sợi vi ống tham gia quá trình phân bào được chia làm ba loại. Sợi cực (*polar microtubules*) có các đầu tận cùng gặp nhau ở mặt phẳng xích đạo làm nhiệm vụ đẩy các cực xa nhau. Sợi liên kết với kinetochore (*kinetochore microtubules*)- phức protein đặc biệt tương tác với tâm động, làm nhiệm vụ điều khiển sự chuyển động của nhiễm sắc thể trong phân bào. Loại thứ ba là sợi sao hay còn gọi là sợi trung thể *astral microtubules* do chúng xuất phát từ trung thể, hướng ra mọi phía và thường ngắn hơn hai loại sợi kia, giữ vai trò tăng cường lực đẩy phân chia hai cực (Hình 6.8).

Ngay khi màng nhân còn nguyên vẹn trong giai đoạn prophase, một số sợi vi ống đã xuất hiện ở trung thể và gặp nhau tại mặt phẳng xích đạo. Một số protein *microtubule motor* làm nhiệm vụ liên kết các sợi microtubules với nhau tại đầu (+) của sợi, tạo lực đẩy và kéo các cực xa nhau. Các sợi microtubules liên tục polymer hoá dài ra và đồng thời bị khử polymer ngắn lại. Tỷ lệ giữa tốc độ polymer và khử polymer không có một giá trị nhất định mà thay đổi phụ thuộc vào từng giai đoạn của phân bào. Cơ chế phân tử của động học biến đổi sợi microtubules chưa rõ ràng mặc dù có liên quan đến MPF trong phản ứng phosphoryl hoá và khử phosphoryl một số protein liên kết với sợi microtubules.



Hình 6.8:
Ba loại sợi vi ống cấu tạo nên thoi phân cực trong phân bào mitosis.

Bước vào phân bào nguyên nhiễm, mỗi sợi nhiễm sắc thể được nhân đôi tạo hai nhiễm sắc tử (*sister chromatid*) dính nhau ở tâm động. Mỗi bên tâm động đều có phức protein tương tác với ADN ở vùng đó tạo nên cấu trúc kinetochore. Chính đầu (+) của sợi vi ống sẽ liên kết với kinetochore thông qua các protein "motor" để điều khiển sự chuyển động của nhiễm sắc thể, đẩy nhiễm sắc tử xa nhau và đưa chúng về các cực đối diện. Ngay khi đã tương tác với kinetochore, đầu (+) của microtubules vẫn tiếp tục polymer hoá và depolymer (tức là vẫn dài thêm và ngắn bớt do thêm vào và đồng thời mất đi các tiểu phần tubulin).

Chương 7

SINH TRƯỞNG VÀ PHÁT TRIỂN

Quá trình sinh trưởng phát triển ở động vật có thể được xem như bắt đầu từ sự thụ tinh của trứng mặc dù trước đó mọi bước chuẩn bị đã phải hoàn tất để quá trình đó được thực hiện một cách hoàn hảo. Ví dụ, một số gen của mẹ được phiên mã và dịch mã. Tuy nhiên, những sản phẩm này hoàn toàn được dự trữ trong trứng như nguồn dinh dưỡng và tập hợp các yếu tố kiểm soát phát triển của hợp tử ngay khi trứng vừa thụ tinh. Sau khi thụ tinh, hợp tử trải qua các giai đoạn phát triển phôi nang, phôi dạ và hình thành các bộ phận của một cơ thể hoàn chỉnh. Ở giai đoạn phôi nang, hầu hết các tế bào chưa biệt hoá. Tuy nhiên, bước sang giai đoạn phôi dạ, phôi thai bắt đầu hình thành với 3 lớp tế bào: nội bì, trung bì và ngoại bì. Từ 3 lớp tế bào này, các tế bào đã có số phận khác nhau, chúng phát triển biệt hoá theo những chương trình khác nhau tạo thành các bộ phận riêng biệt của một cơ thể hoàn chỉnh.

Thật thú vị khi từ một tế bào trứng ban đầu, một cơ thể đa bào với nhiều loại tế bào biệt hoá khác nhau được hình thành. Ở mức độ phân tử, sự khác biệt giữa các loại tế bào được xem như phụ thuộc vào các cách thức hoạt động của gen và tương tác giữa các sản phẩm của chúng. Hoạt động của gen liên quan chủ yếu đến các protein. Chúng có thể làm thay đổi cấu trúc của đoạn ADN chứa promoter, khởi động quá trình sao chép thông tin, điều khiển hoạt động của các enhancer hoặc ức chế các factor phiên mã.

Quá trình điều khiển bật mở hoặc đóng các gen do protein đảm nhiệm dẫn đến việc hình thành các bộ phận của một cơ thể hoàn chỉnh được nghiên cứu khá kỹ trên đối tượng ruồi giấm *D.melanogaster* và giun tròn *C.elegans*. Sau khi thụ tinh, một loạt diễn biến xảy ra trong trứng ruồi giấm, bắt đầu từ sự phân bố không đồng nhất của các chất trong tế bào chất dẫn đến mức độ biểu hiện khác nhau của các gen thay đổi theo thời gian và không gian. Điều đó khiến cho các phần cục bộ phân bố bên trong trứng có những đặc tính riêng biệt. Ở bất kỳ giai đoạn nào trong quá trình phát triển, việc đóng hoặc mở một gen đều liên quan đến hoạt động của các gen khác trong giai đoạn trước đó. Nói một cách khác, một protein điều khiển này sẽ kiểm tra hoạt động của gen mã cho protein điều khiển khác. Như vậy, trong giai đoạn mô phôi, hoạt động của gen xảy ra ở một thời điểm nhất định giữ vai trò chủ đạo trong biệt hoá phát triển.

Nghiên cứu sinh trưởng phát triển ở động vật thường dựa vào phân tích di truyền, đòi hỏi các đối tượng thỏa mãn những yêu cầu như số lượng cá thể nhiều, vòng đời ngắn, dễ dàng thu nhận được các thể đột biến trong tự nhiên cũng như bằng kỹ thuật chuyển gen vv... Ruồi giấm *Drosophila melanogaster* và giun tròn *Caenorhabditis elegans* đáp ứng được những yêu cầu trên nên được sử dụng như những sinh vật mô hình để nghiên cứu về hoạt động của gen trong quá trình phát triển, đặc biệt ở giai đoạn phôi sớm. Hệ gen của hai đối tượng mô hình này không lớn (170.000 kb ở ruồi giấm, 80.000 kb ở giun tròn), tạo thuận lợi cho việc phân lập các gen liên quan. Một khi đã biết trình tự nucleotide của một gen đặc hiệu, hoạt động của gen này theo thời gian và phụ thuộc vào tổ chức mô chuyên hoá được xác định nhờ các kỹ thuật lai acid nucleic hoặc tạo kháng thể. Tất cả những phương pháp sinh học hiện đại đã giúp chúng ta hiểu được thời điểm và vị trí mà các gen bắt đầu hoạt động trong quá trình phát triển phôi. Ngoài ra, các gen phân lập được từ hai đối tượng mô hình này có thể được dùng để tạo ra các cá thể chuyển gen. Từ đó, chức năng đặc thù của gen cũng như các quá trình sinh học

liên quan đến sản phẩm của gen được hiểu một cách đầy đủ hơn. Có thể nói những nghiên cứu trên giun tròn và ruồi giấm giúp các nhà sinh học có những hiểu biết chi tiết và tổng thể về các cách thức kiểm soát sinh trưởng phát triển. Trong khi những gen có vai trò quyết định quá trình phát triển ở ruồi giấm là những gen mã cho protein tham gia vào biến đổi ARN thì ở giun tròn lại là những gen mã cho các yếu tố điều khiển phiên mã. Tuy nhiên, cách thức kiểm soát xác định giới tính ở giun tròn và ruồi giấm giống nhau: giới tính được xác định phụ thuộc vào tỷ lệ giữa số lượng nhiễm sắc thể giới tính X và số lượng nhiễm sắc thể thường. Khi tỷ lệ này bằng hoặc lớn hơn 1 (≥ 1), phôi phát triển thành con cái. Khi tỷ lệ này bằng hoặc nhỏ hơn 0,5 ($\leq 0,5$) thì phôi phát triển thành con đực.

Giun tròn *C.elegans* được các nhà di truyền học quan tâm rất muộn (1960) so với ruồi giấm *Drosophila* (1909). Giun tròn có thể nuôi trong đĩa petri với nguồn thức ăn chính là vi khuẩn *E.coli*. Trong điều kiện môi trường dinh dưỡng đầy đủ, vòng đời của giun tròn chỉ trong khoảng 3 ngày. Đặc biệt, *C.elegans* là động vật lưỡng tính; trứng và tinh trùng sinh ra từ một cá thể có khả năng tự thụ tinh phát triển thành một cơ thể mới. Vì thế dễ dàng chọn lọc được các đột biến lặn dạng đồng hợp tử. Ngoài ra, cơ thể giun tròn trong suốt nên có thể nhìn thấy các tế bào và theo dõi được số phận của chúng trong quá trình phát triển từ trứng thụ tinh. Phả hệ của tất cả các tế bào tạo ra trong suốt quá trình từ trứng thụ tinh đến cơ thể trưởng thành đã được xác định với giun tròn khiến cho *C.elegans* trở nên mô hình lý tưởng trong nghiên cứu sinh trưởng phát triển.

Ruồi giấm *Drosophila* có toàn bộ chu trình sống chỉ kéo dài 9 đến 10 ngày. Trứng ruồi giấm sau thụ tinh phân chia rất nhanh tạo thành các tế bào không có màng. Vì vậy, giai đoạn đầu ngay sau thụ tinh có thể xem trứng là một tế bào có rất nhiều nhân giống nhau. Trạng thái này của trứng được gọi là hỗn bào. Sau 9 lần phân chia nguyên nhiễm, tổng số 512 nhân di chuyển trong tế bào chất về phía màng tế bào trứng và tiếp tục phân chia thêm 4 lần nữa. Tiếp đến tất cả các nhân này đều được bao bọc bởi màng tạo thành các tế bào riêng biệt. Những tế bào này nằm sát màng tế bào trứng (phía tế bào chất) tạo thành lớp đơn bào (gọi là lớp bì phôi) trên bề mặt của phôi. Phôi ruồi giấm còn phải trải qua một loạt biến đổi hình thái từ ấu trùng, nhộng đến cơ thể trưởng thành.

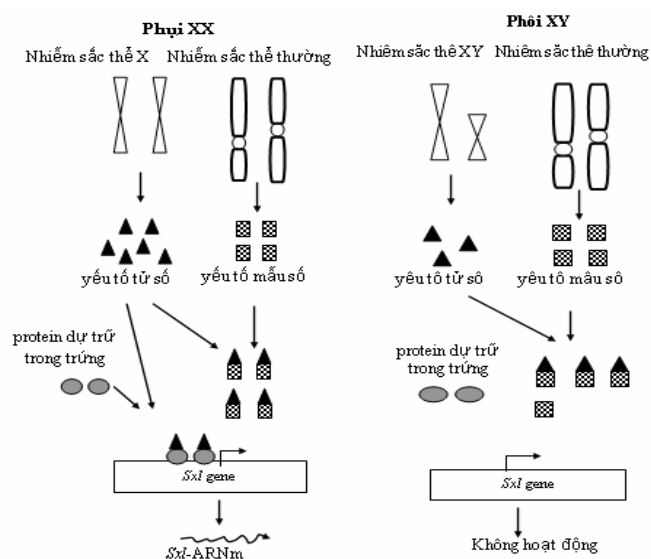
7.1 Kiểm soát xác định giới tính

Ở giun tròn *C.elegans*, giun đực chỉ có một nhiễm sắc thể X, giun cái có 2 nhiễm sắc thể XX. Cách thức xác định giới tính ở *C.elegans* liên quan ít nhất đến sản phẩm của 10 gen khác nhau. Đột biến ở các gen này gây nên những lệch lạc trong phát triển giới tính. Ví dụ, đột biến gây mất chức năng (*loss-of-function*) ở gen *her-1* (*hemaphrodite-1*) khiến phôi XO (chỉ có một nhiễm sắc thể X) phát triển thành cá thể lưỡng tính. Trong phôi XX, protein này không được tổng hợp. Tuy nhiên, trong phôi XO, protein HER-1 có chức năng của protein điều khiển (*regulatory protein*) gây kìm hãm hoạt động của một loạt gen liên quan đến phát triển giới tính cái. Như vậy gen *her-1* mã cho protein cần thiết để phôi XO phát triển thành giun đực. Số lượng nhiễm sắc thể X và số lượng nhiễm sắc thể thường (ký hiệu là A) có mặt trong hợp tử phải được xác định chính xác vì tỷ số X/A giữa hai số lượng đó quyết định biệt hoá giới tính của phôi. Việc xác định chính xác số lượng nhiễm sắc thể giới tính phụ thuộc vào các yếu tố tử số (*numerator elements*) và việc xác định chính xác số lượng nhiễm sắc thể thường phụ thuộc các yếu tố mẫu số (*denominator elements*). Một số yếu tố tử số đã được phát hiện nhưng sự tồn tại của yếu tố mẫu số chưa được chứng minh bằng thực nghiệm.

Ở ruồi giấm *Drosophila*, tương tự như ở giun tròn *C.elegans*, quá trình xác định giới tính cũng đòi hỏi phải xác định tỷ số giữa số lượng nhiễm sắc thể X và số lượng nhiễm sắc thể

thường (A). Thông tin tỷ số X/A sẽ chuyển đổi thành các tín hiệu để hoạt hoá các nhóm gen khác nhau liên quan đến hình thành các biểu hiện đặc thù của từng giới. Số lượng nhiễm sắc thể X và số lượng nhiễm sắc thể thường được xác định ở giai đoạn rất sớm của phôi. Trong giai đoạn phôi sớm, một số protein được mã bởi các gen phân bố trên nhiễm sắc thể X và một số khác được mã bởi các gen nằm trên nhiễm sắc thể thường. Các protein của phôi tương tác với nhau và tương tác với protein được dự trữ trong tế bào chất của trứng. Số lượng protein do gen nằm trên nhiễm sắc thể X trong phôi XX sẽ nhiều gấp 2 lần so với phôi XY. Do đó, sự chênh lệch nồng độ sản phẩm của các gen liên kết nhiễm sắc thể X so với các gen trên nhiễm sắc thể thường cho phép “đếm” chính xác số lượng nhiễm sắc thể XX. Do có số lượng nhiều hơn nên protein của gen trên nhiễm sắc thể X trong phôi XX dễ dàng tương tác với protein dự trữ sẵn trong tế bào chất của trứng. Tương tác này sẽ dẫn đến khởi động những gen quyết định phát triển giới tính (Hình 7.1). Như vậy, những protein mã bởi gen liên kết giới tính đảm nhận chức năng của yếu tố tử số (*numerator elements*). Ngược lại, những protein mã bởi gen trên nhiễm sắc thể thường đảm nhận chức năng của yếu tố mẫu số. Tỷ số giữa nồng độ yếu tố tử số và nồng độ yếu tố mẫu số tỷ lệ với tỷ số X/A. Do đó, giới tính được quyết định phụ thuộc vào chênh lệch nồng độ sản phẩm của các gen nằm trên nhiễm sắc thể X với sản phẩm các gen trên nhiễm sắc thể thường. (Cần nhắc lại là các gen này hoạt động rất sớm trong giai đoạn phát triển của phôi).

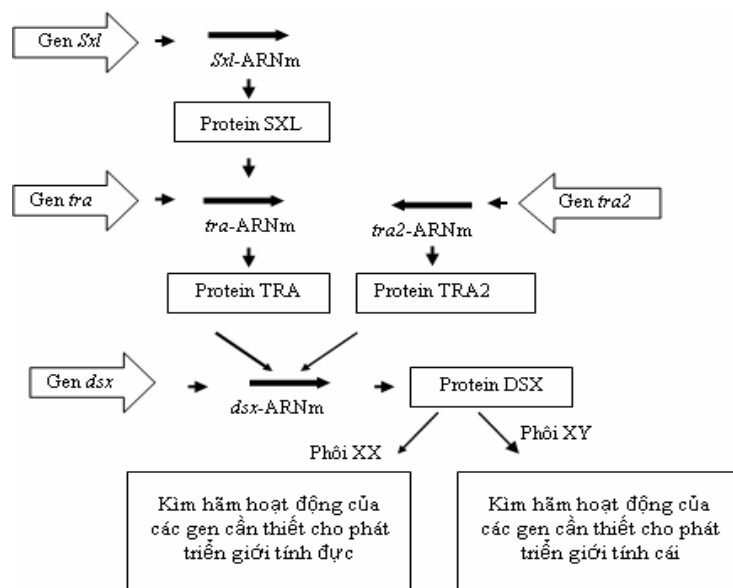
Một khi tỷ số X/A được xác định, giá trị của tỷ số sẽ chuyển đổi thành tín hiệu điều khiển hoạt động của gen *Sxl* (*Sex-lethal gene*) tại promoter P_E . Gen *Sxl* nằm trên nhiễm sắc thể X. Protein SXL của gen *Sxl* chỉ được phát hiện trong giai đoạn sớm ở phôi XX. Sau đó, phiên mã từ P_E được thay thế bởi promoter P_M . Điều đặc biệt thú vị là promoter P_E cũng hoạt động ở phôi XY. Tuy nhiên, quá trình biến đổi (cắt nối *exon-intron*) của phân tử ARNm phiên mã từ P_M chỉ được thực hiện hoàn hảo khi có mặt của protein SXL (nhắc lại là protein này được dịch mã từ ARN của promoter P_E trong phôi XX). Trong phôi XY, cơ chế cắt nối luân phiên khi không có mặt protein SXL bị lệch lạc làm xuất hiện mã dừng (*stop codon*) không đúng chỗ trên phân tử ARNm. Như vậy, sự vắng mặt của protein SXL tương ứng với cả 2 promoter khiến cho phôi phát triển thành ruồi đực.



Hình 7.1:

Tỷ số giữa số lượng nhiễm sắc thể giới tính X và số lượng nhiễm sắc thể thường quyết định hoạt động của gen *Sxl*; protein SXL kiểm soát các gen cần thiết cho xác định giới tính của phôi *Drosophila* (theo Snustad, 2000).

Trong phôi XX, protein SXL đảm nhận chức năng của protein kiểm soát theo cơ chế tích cực quá trình biến đổi phân tử tiền thân ARNm. Điều này có nghĩa, sự có mặt của SXL cho phép số lượng protein SXL tăng lên nhờ cắt nối luân phiên chính xác. Như vậy, protein SXL kiểm soát tích cực hoạt động của chính gen *Sxl*. Ngoài ra, SXL còn kiểm soát sau phiên mã đối với ARNm của gen *tra* (*transformer gene*). Các phân tử ARNm tiền thân phiên mã từ gen *tra* được cắt nối luân phiên. Trong phôi XY, cắt nối luân phiên của *tra*-ARN tạo ra mã dừng ngay trong exon thứ hai. Trong phôi XX, sự có mặt của SXL giúp *tra*-ARNm cắt nối đúng để tiếp tục dịch mã tạo protein TRA. Như vậy, sự có mặt của protein SXL cho phép xuất hiện protein TRA trong phôi XX.



Hình 7.2:

Kiểm soát xác định giới tính ở phôi ruồi giấm *Drosophila*. Protein SXL của gen *Sxl* bật mở chuỗi các bước kiểm soát hoạt động của các gen xác định giới tính thông qua kiểm soát cắt nối luân phiên các phân tử tiền thân ARNm (theo Snustad, 2000).

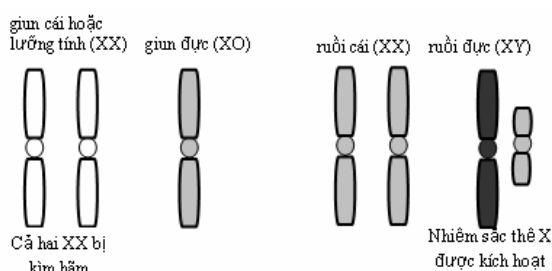
Protein TRA, đến lượt mình, phối hợp với protein TRA2 (mã bởi gen nằm trên nhiễm sắc thể thường) để kiểm soát hoạt động của gen *dsx* (*doublesex gene*). Gen *dsx* mã cho 2 loại protein DSX; chúng được dịch mã từ 2 loại phân tử ARNm khác nhau tạo ra do cắt nối luân phiên từ một *dsx*-ARNm tiền thân ban đầu. Trong phôi XX có mặt protein TRA, TRA tác động đến cắt nối luân phiên của *dsx*-ARNm. Do đó, loại protein DSX của phôi XX có chức năng kìm hãm hoạt động của những gen cần thiết cho phôi phát triển thành con đực. Trong phôi XY không có protein TRA, *dsx*-ARNm khác với *dsx*-ARNm trong phôi XX. Loại protein DSX có trong phôi XY có chức năng kìm hãm hoạt động của những gen cần thiết cho phôi phát triển thành con cái. Như vậy, 2 protein DSX xuất phát từ một gen *dsx*. Tuy nhiên, mỗi loại DSX quyết định con đường phát triển giới tính duy nhất cho phôi. Từ sự lựa chọn này, hàng loạt các gen sẽ biểu hiện khác nhau giữa phôi XX và phôi XY trong quá trình biệt hoá phát triển giới tính (Hình 7.2).

Trong giai đoạn sớm của phát triển phôi, hoạt động của các gen trên nhiễm sắc thể X sẽ được kiểm soát theo cách thức điều tiết xuất liều (*dosage compensation*). Ví dụ, ở giun tròn *C.elegans*, phức protein chỉ xuất hiện trong tế bào phôi có hai nhiễm sắc thể XX và tương tác với chúng để kìm hãm hoạt động của các gen trên nhiễm sắc thể X. Điều đó đảm bảo điều tiết được liều lượng protein của những gen này trong tế bào có 2 XX (giun cái) giống như trong tế bào chỉ có một X (giun đực). Ngược lại giun tròn, trong tế bào ruồi đực (XO) *Drosophila* có phức protein tương tác với một nhiễm sắc thể X để tăng liều xuất hoạt động của những gen

nằm trên nhiễm sắc thể X. Như vậy trong tế bào phôi *Drosophila* có hai nhiễm sắc thể XX, không xảy ra hiện tượng bất hoạt một nhiễm sắc thể X như ở động vật bậc cao. Các gen trên 2 nhiễm sắc thể XX đều hoạt động nhưng mức độ biểu hiện được kiểm soát theo cơ chế kim hãm. Trong các tế bào phôi *Drosophila* chỉ có một nhiễm sắc thể X thì hoạt động của các gen trên nhiễm sắc thể này được kích hoạt tăng liều xuất. Điều đó đảm bảo số lượng protein trong phôi XX tương tự như trong phôi XO (Hình 7.3).

Điều tiết liều xuất ở *Drosophila* đòi hỏi ít nhất sản phẩm của 4 gen khác nhau, gọi chung là những gen *msl* (male-specific lethal genes). Trong phôi đực, phức protein MSL bám trên nhiễm sắc thể X được xác định nhờ kỹ thuật lai sử dụng kháng thể đối với MSL. Đột biến ở những gen này khiến phôi đực bị chết do nhiễm sắc thể X duy nhất của phôi không được kích hoạt. Trong số 4 protein MSL, protein MLE (sản phẩm của gen *mle-maleless*) có hoạt tính helicase gây mở xoắn chuỗi kép ADN. Điều này gợi ý nhiễm sắc thể X duy nhất trong phôi đực mở xoắn để các gen tăng cường hoạt động. Điều đặc biệt là phức MSL có liên kết với ARN để tương tác với nhiễm sắc thể X. Hai gen *roX1* và *roX2* nằm trên nhiễm sắc thể X (*RNA on the chromosome X*) phiên mã tạo *roX1*-ARN và *roX2*-ARN. Hai loại ARN này không được dịch mã mặc dù chúng đều trải qua biến đổi thêm đuôi poly A và cắt nối exon-intron. Chúng phân bố trên nhiễm sắc thể X cùng với phức MSL. Hoạt động của 2 gen mã cho *roX*-ARNs chịu kiểm soát tiêu cực của protein SXL. Hai gen này chỉ phiên mã khi gen *sxl* không hoạt động. Như vậy, các phân tử *roX*-ARNs chỉ xuất hiện trong phôi XY.

Khi phôi XX có đột biến trên cả 2 allele của gen *sxl*, phôi không phát triển thành ruồi đực như dự đoán mà phôi bị chết. Tuy nhiên, phôi chết không phải do rối loạn trong quá trình chuyển đổi giới tính (do gen *sxl* đột biến) mà do rối loạn gây ra trong điều tiết liều xuất của những gen nằm trên nhiễm sắc thể X trong phôi XX. Trong phôi đồng hợp tử đột biến *Sxl*, các gen trên hai nhiễm sắc thể XX không được kim hãm nên dẫn đến sự quá liều các sản phẩm của những gen đó. Điều này khiến cho phôi chết. Như vậy, bên cạnh nhiệm vụ kiểm soát biểu hiện của các gen trong quá trình xác định giới tính, protein SXL còn điều tiết liều xuất của chính các gen nằm trên nhiễm sắc thể X.

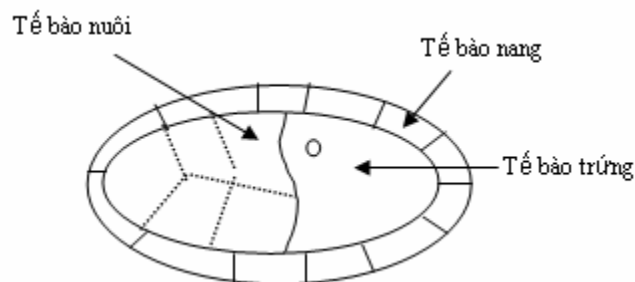


Hình 7.3:

Cơ chế điều tiết liều xuất đối với hoạt động của các gen nằm trên nhiễm sắc thể X ở giun tròn và ruồi giấm. Hai cách thức điều tiết xảy ra ngược nhau đối với hai loài động vật này và khác với cách thức bất hoạt một nhiễm sắc thể X ở động vật có vú.

7.2 Phát triển ở ruồi giấm *Drosophila*

Ở ruồi giấm *Drosophila*, trứng được hình thành từ noãn bào. Một noãn bào có mối liên hệ với 15 tế bào nuôi (*nurse cell*) thông qua các cầu nối tế bào chất (Hình 7.4). Nhờ đó mà các chất dinh dưỡng và ARNm tổng hợp trong tế bào nuôi được vận chuyển sang noãn bào. Những nguyên liệu này cần thiết cho noãn bào phát triển thành trứng cũng như đối với sự phát triển của mô phôi ở giai đoạn đầu.



Hình 7.4:

Các tế bào nuôi nối với nhau và nối với một cực của tế bào trứng thông qua cầu nối tế bào chất. Các tế bào nang là tế bào soma trong khi tế bào trứng và các tế bào nuôi xuất phát từ tế bào sinh dục.

Hầu hết các gen liên quan đến quá trình phát triển của *Drosophila* được phát hiện nhờ các đột biến gây chết ở giai đoạn phát triển phôi sớm hoặc khiến cho một bộ phận cơ thể phát triển không bình thường. Từ các kết quả nghiên cứu di truyền phân tử, thực nghiệm đã phân loại các đột biến thành ba nhóm chính tùy theo hoạt động của chúng cũng như chức năng sản phẩm tương ứng. Một số đặc điểm chung của từng nhóm đột biến như sau:

1. Đột biến xảy ra với các gen có nguồn gốc từ mẹ (*maternal genes*). Các gen này hoạt động trong giai đoạn phát triển trứng ở trong cơ thể mẹ. Đột biến trên những gen này không gây ảnh hưởng đến đời sống bình thường của ruồi mẹ, nhưng lại có biểu hiện tính trạng ở thế hệ con. Chúng có thể hoạt động trước hoặc sau khi trứng chín.
2. Đột biến ở các gen tạo đốt (*gap genes*). Các gen này hoạt động sau khi trứng đã thụ tinh. Đột biến trên chúng làm thay đổi số lượng hoặc tính phân cực của các đốt.
3. Đột biến xảy ra ở các gen chọn lọc (*selector genes*) hay còn gọi là gen homeotic (*homeotic genes*). Các đột biến này không làm thay đổi số lượng, kích thước hoặc tính phân cực của các đốt. Các gen homeotic qui định những đặc tính riêng biệt của từng đốt. Do đó, những đột biến xảy ra làm cho một phần của cơ thể mang những tính trạng của phần khác. Ví dụ đốt ngực biểu hiện các đặc tính của đốt bụng; chân hoặc cánh có thể xuất hiện ở phần đầu v.v...

Các gen của mỗi nhóm hoạt động nối tiếp nhau để hoàn thiện các tính chất riêng có tính đặc thù ngày càng cao của các phần trong phôi. Các gen có nguồn gốc từ mẹ phân định các vùng trong trứng. Trong mỗi vùng, sản phẩm của các gen có nguồn gốc từ mẹ phân bố khác nhau. Điều đó dẫn đến sự hoạt hoá khác nhau của các gen phân đốt trong vùng. Nhờ đó mà các đốt được hình thành. Trong mỗi đốt, các gen homeotic sẽ hoạt động để xác định các tính chất đặc thù của đốt, tức là hình thành các cơ quan có vị trí phân bố nhất định tại đốt đó.

7.3 Hoạt động của các gen có nguồn gốc từ mẹ trong quá trình hình thành trục đầu-đuôi và trục lưng-bụng

Đối với động vật, cơ thể luôn có hai trục nhằm phân biệt phần đầu với phần đuôi và phần lưng với phần bụng. Cả hai trục này được hình thành rất sớm, thậm chí ở một số loài có thể được định hình ngay trước khi trứng thụ tinh. Trên đối tượng ruồi giấm *Drosophila*, quá trình hình thành hai trục đầu-đuôi và lưng-bụng được nghiên cứu khá chi tiết dựa vào các loại đột biến ảnh hưởng đến giai đoạn phát triển rất sớm của phôi.

Phân tích tổng thể các đột biến liên quan đến giai đoạn phát triển đầu tiên của phôi ruồi giấm cho phép xác định bốn nhóm gen có nguồn gốc từ mẹ tham gia kiểm soát quá trình hình

thành các trục đầu-đuôi, lưng-bụng của phôi. Để hiểu rõ ở mức độ phân tử những cơ chế liên quan đến sản phẩm của gen có nguồn gốc từ mẹ, sinh học hiện đại đặc biệt là di truyền phân tử đã tiến hành phân lập gen, xác định hoạt động của gen ở mức độ tổng hợp ARNm, xác định sự phân bố của sản phẩm tương ứng tại các phần khác nhau ở phôi cũng như tác động của sản phẩm đến chức năng của các protein khác. Thực nghiệm nhận thấy các gen có nguồn gốc từ mẹ giữ vai trò rất quan trọng trong giai đoạn phát triển phôi sớm. Đột biến xảy ra ở những gen này không làm xuất hiện tính trạng mới đối với mẹ mà chỉ biểu hiện ở thế hệ sau. Trứng của những con mẹ mang đột biến ở các gen đó thường bị hỏng, phôi bị chết trong quá trình phát triển.

Các gen có nguồn gốc từ mẹ hoạt động trước khi trứng được thụ tinh. Sản phẩm của chúng có thể được sử dụng ngay hoặc được dự trữ trong trứng dưới dạng ARNm hoặc protein. Ngay trong trứng đã xuất hiện gradient của các protein hoặc ARNm phân bố theo hai trục đầu-đuôi và lưng-bụng. Các gradient chịu trách nhiệm hình thành trục đầu-đuôi xuất hiện trước, còn gradient chịu trách nhiệm trục lưng-bụng xuất hiện chậm hơn.

Các gen có nguồn gốc từ mẹ được chia làm bốn nhóm khác nhau: nhóm gen lưng-bụng (*dorsoventral-group genes*), nhóm gen hai vùng tận cùng đầu và đuôi (*terminal-group genes*), nhóm gen phần đầu bao gồm đầu và ngực ruồi giấm (*anterior-group genes*) và nhóm gen phần đuôi (*posterior-group genes*). Mỗi nhóm gồm những gen cùng quyết định chung cho một cách thức phát triển của từng vùng. Trong mỗi vùng, các diễn biến cục bộ xảy ra đầu tiên, có thể ở ngoài hoặc trong trứng. Điều đó dẫn đến phân vùng của các tín hiệu ở trong trứng. Tiếp đến, sự phân bố cục bộ của các protein có thể dẫn đến hình thành một cấu trúc riêng biệt tại vùng đó. Nói một cách khác thì tại mỗi vùng, nồng độ protein đặc hiệu đạt ngưỡng nhất định sẽ quyết định số phận phát triển của vùng đó. Những protein này được gọi là **morphogen** (tạm dịch là protein *mã liều*). Ở giai đoạn đầu tiên phát triển phôi, sự hình thành vùng đầu-đuôi và vùng lưng-bụng được quyết định bởi sự phân bố cục bộ của gần 30 morphogen. Thực nghiệm đã tìm được hơn 30 gen liên quan đến sinh tổng hợp, vận chuyển và phân bố các morphogen. Sự phân vùng xảy ra hoàn toàn độc lập với nhau.

7.3.1. Nhóm gen quyết định phát triển của phần đầu và ngực ấu thể (*anterior-group genes*)

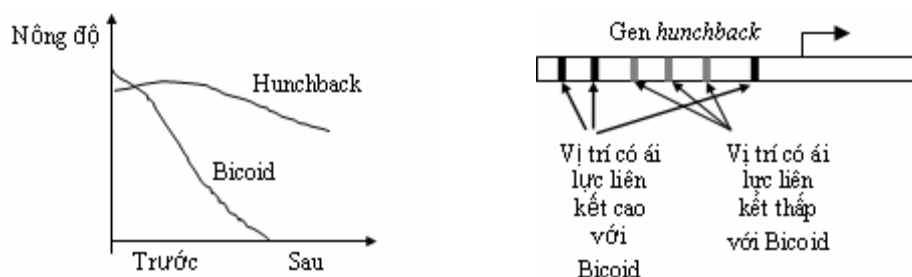
Khi chọc thủng một lỗ nhỏ tại đầu trước của trứng (nơi tiếp giáp với các tế bào nuôi) để tế bào chất của phần đó thoát ra ngoài, phôi sẽ phát triển không có đầu. Nếu lấy tế bào chất ở phần sau của trứng bổ sung vào phần đầu thì phôi phát triển sẽ có hai đuôi (mà không có đầu). Nếu lấy tế bào chất của phần đầu tiêm vào vị trí khác trên trứng, tại đó cấu trúc đầu sẽ hình thành. Các kết quả này chứng tỏ rằng phần đầu ấu thể được quyết định bởi các chất nằm tại một cực của trứng. Di truyền phân tử đã tìm được những con mẹ bị đột biến ở gen *bicoid* sẽ đẻ trứng phát triển thành phôi không có đầu. Nếu lấy tế bào chất ở cực trước của trứng bình thường tiêm vào trứng của con mẹ đột biến thì phôi của trứng đó lại phát triển có đầu. Như vậy chính sản phẩm của gen *bicoid* dự trữ ở cực trước của trứng giữ vai trò kiểm tra quá trình phát triển phần đầu ấu thể.

Kết quả phép lai *insitu* cho thấy ARNm của gen *bicoid* được tổng hợp trong các tế bào nuôi và đưa vào tế bào trứng theo cầu nối tế bào chất. Chúng được giữ lại ở cực trước do phần không dịch mã ở đầu 3' tương tác với các thành phần của khung tế bào tại cực đó. Các phân tử ARNm này chỉ được sử dụng để tổng hợp protein Bicoid sau khi trứng thụ tinh. Nồng độ protein Bicoid lớn nhất tại cực trước và giảm dần về phía đuôi. Gradient nồng độ thay đổi khi genome của con mẹ được ghép thêm các bản sao của gen *bicoid*. Do số bản sao của gen

bicoid trong cơ thể mẹ tăng, nồng độ ARNm cũng như protein Bicoid trong trứng cũng tăng. Điều này dẫn đến phôi phát triển có phần đầu rộng lớn ở phía sau. Như vậy protein Bicoid là một morphogen. Phân đầu của phôi được qui định do sự chênh lệch nồng độ của morphogen này. Protein Bicoid bám vào promoter của một số gen và hoạt hoá chúng. Một trong những gen này là gen *hunchback*.

Hoạt động của *hunchback* phụ thuộc nồng độ protein Bicoid. Gen *hunchback* chỉ bật mở khi nồng độ Bicoid đạt một giá trị nhất định và hoạt động mạnh hay yếu phụ thuộc vào số lượng gen *bicoid*, tức số lượng protein Bicoid có nhiều hay ít. Promoter của gen *hunchback* chứa 3 vị trí có ái lực liên kết cao và 3 vị trí có ái lực yếu đối với protein Bicoid. Ruồi giấm chuyển gen mang gen báo cáo (*reporter gene*) gắn với promoter *hunchback* bị loại bỏ các vị trí có ái lực liên kết khác nhau đã chứng tỏ mức độ tương tác của Bicoid với những vị trí đó sẽ quyết định mức độ hoạt động của gen *hunchback*. Do đó, gradient nồng độ Bicoid sẽ ứng với mức độ biểu hiện khác nhau của gen *hunchback* theo không gian. Mặt khác ở ngay cùng một nồng độ Bicoid, do các vị trí ở vùng ADN điều khiển của gen *hunchback* có ái lực tương tác khác nhau với Bicoid nên sự cạnh tranh liên kết cũng ảnh hưởng đến hoạt động của gen này. Nói một cách khác, sự khác nhau về số lượng morphogen tham gia khởi động các gen được chuyển thành sự khác nhau về bản chất, cấu trúc tế bào trong giai đoạn phát triển phôi (Hình 7.6).

Điều đáng lưu ý là các phân tử ARNm-*hunchback* được tạo ra từ hai nguồn. Nguồn thứ nhất từ quá trình phiên mã trên gen *hunchback* được kiểm soát bởi nồng độ Bicoid. Các phân tử ARNm này phân bố cục bộ ở vùng đầu của phôi. Nguồn thứ hai có sẵn trong tế bào chất của trứng, chúng được vận chuyển từ các tế bào nuôi sang tế bào trứng và phân bố đồng đều trong tế bào chất. Mặc dù các ARNm-*hunchback* từ nguồn thứ hai có mặt ở mọi vùng trong trứng nhưng protein tương ứng Hunchback không xuất hiện ở phần đuôi của phôi. Đó là do sự mặt của protein Nanos. Gen mã cho protein này có nguồn gốc từ mẹ, tức là ARNm-*nanos* được dự trữ sẵn trong trứng. Cũng giống như hầu hết ARNm của các gen có nguồn gốc từ mẹ khác, ARNm-*nanos* phân bố cục bộ ở phía cuối của phôi. Do đó, xuất hiện gradient protein Nanos theo chiều ngược lại với gradient protein Bicoid. Hơn nữa, chức năng của hai protein này cũng ngược nhau. Trong khi Bicoid hoạt hoá thì Nanos kìm hãm hoạt động của gen *hunchback*. Protein Nanos ngăn cản quá trình dịch mã từ các phân tử ARNm-*hunchback* có mặt ở vùng đuôi. Hai protein Bicoid và Nanos đều tham gia kiểm soát hoạt động của *hunchback*. Tuy nhiên cần lưu ý, Bicoid là factor phiên mã trong khi Nanos tác động đến quá trình dịch mã.



Hình 7.6:

Gradient nồng độ của hai protein Bicoid và Hunchback theo trục trước-sau của phôi *Drosophila*. Protein Bicoid hoạt hoá gen *hunchback* thông qua tương tác với các vị trí đặc hiệu nằm trong vùng điều khiển của gen *hunchback*. Các tương tác này có ái lực khác nhau tương ứng với biểu hiện ở mức độ khác nhau của *hunchback*.

Một loạt các protein khác được tổng hợp từ các gen có nguồn gốc từ mẹ tham gia vào giai đoạn phát triển đầu tiên của phôi ruồi giấm. Ví dụ, gen *pumilio* mã cho protein có chức năng

trung tự Nanos, tức là kim hãm quá trình tổng hợp Hunchback. Khả năng kim hãm của Nanos và Pumilio phụ thuộc vào vùng 3' không dịch mã trên phân tử ARNm- *hunchback* bao gồm cả đuôi polyA. Thực nghiệm quan sát thấy ở phôi ruồi giấm bình thường, đuôi polyA được dài thêm trước khi phân tử ARNm- *hunchback* được dịch mã. Protein Nanos kích thích phản ứng cắt ngắn đuôi polyA khiến việc dịch mã trên ARNm- *hunchback* bị ức chế.

7.3.2. Nhóm gen qui định phát triển phần đuôi (posterior-group genes)

Kết quả các thí nghiệm di truyền cho thấy đột biến ở một số gen có nguồn gốc từ mẹ dẫn đến việc phôi phát triển không có phần đuôi và không có các tế bào sinh dục. Tiến hành các thí nghiệm tương tự như đã làm với cực trước của trứng (thay tế bào chất ở cực sau của trứng bình thường vào trứng bị đột biến hoặc tiêm tế bào chất này vào một vị trí bất kỳ...) đều khôi phục được cấu trúc phần đuôi của ấu thể. Thực nghiệm đã tìm ra protein Nanos đóng vai trò morphogen trong việc hình thành cấu trúc đuôi. Gen *nanos* hoạt động trong tế bào nuôi và ARNm của nó được vận chuyển qua cầu nối tế bào chất đến cực sau của trứng. Chúng được giữ ở cực sau nhờ phần không dịch mã ở đầu 3'. Có lẽ phần này được nhận biết và tương tác với sản phẩm của một gen hoạt động trước *nanos* và cũng phân bố cực bộ ở cực sau.

Quá trình hình thành các tế bào sinh dục, cùng với các gen cần thiết cho việc thiết lập cấu trúc phần đuôi còn đòi hỏi một số gen có nguồn gốc từ mẹ mà sản phẩm của chúng được tích lũy tại cực sau của trứng. Gen *oskar* đóng vai trò then chốt trong việc hình thành tế bào sinh dục. Khi gen này bị đột biến, phôi phát triển không có tế bào sinh dục. ARNm và protein Oskar tập trung ở cực sau. Chúng xâm nhập vào một số tế bào ở cực này và quyết định số phận của chúng phát triển thành tế bào sinh dục. Đưa ARNm của gen *oskar* vào cực đầu, các tế bào sinh dục xuất hiện ở cực này.

7.3.3. Nhóm gen qui định phát triển trục lưng-bụng (dorsoventral-group genes)

Sản phẩm của nhóm gen này điều khiển sự hình thành trục lưng-bụng từ khi trứng thụ tinh đến giai đoạn xuất hiện các tế bào phôi bì (*blastoderm*). Đột biến ở bất kỳ gen nào trong nhóm đều dẫn đến hiện tượng phôi phát triển không có cấu trúc bụng. Tại vị trí của phần bụng xuất hiện cấu trúc lưng. Các đột biến kiểu này đều được khắc phục khi đưa tế bào chất của phôi phát triển bình thường vào phôi phát triển từ trứng của con mẹ chứa các gen đột biến thuộc nhóm trục lưng bụng.

Đầu tiên các nhà di truyền đã dùng tác nhân hoá học gây đột biến từng nhiễm sắc thể của ruồi giấm. Rất nhiều gen mang đột biến và tính trạng liên quan đã được phát hiện. Trong số đó, gen *dorsal* được xem đóng vai trò quyết định đến việc hình thành trục lưng bụng. Protein Dorsal được tổng hợp từ mẹ và dự trữ trong tế bào chất của trứng. Đến giai đoạn phôi bì, Dorsal di chuyển vào trong nhân của những tế bào phân bố ở vùng bụng. Lúc đó Dorsal bật mở hoạt động của hai gen *twist* và *snail*, đồng thời kim hãm gen *zerknüllt* và *decapentaplegic*. Nhờ đó, các tế bào vùng bụng sẽ biệt hoá thành trung bì. Diễn biến hoạt động của bốn gen trên trong các tế bào phân bố ở vùng lưng xảy ra ngược lại. Do protein Dorsal không di chuyển vào nhân các tế bào phân bố vùng lưng nên trong các tế bào này hai gen *zerknüllt* và *decapentaplegic* hoạt động còn hai gen *twist* và *snail* bị kim hãm hoàn toàn. Các tế bào này sẽ biệt hoá thành biểu bì.

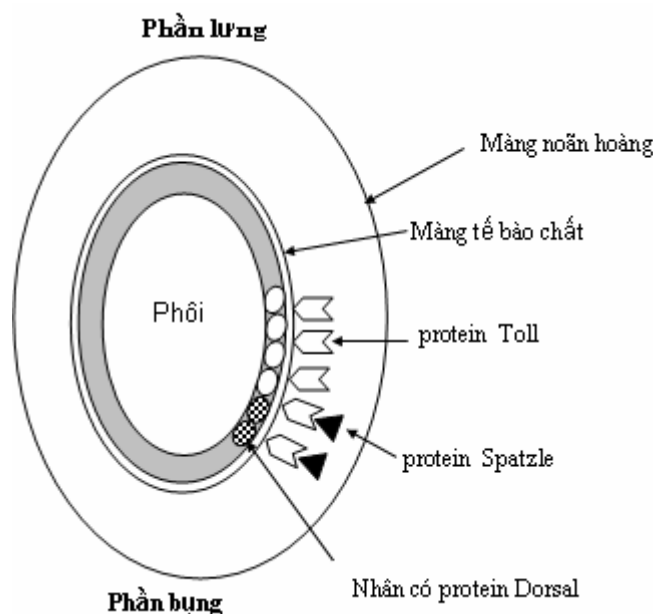
Điều đáng quan tâm là vì sao chỉ xảy ra sự di chuyển vị trí của Dorsal trong tế bào vùng bụng? Tín hiệu quyết định sự chuyển chỗ phát đi từ bên trong trứng hay các tế bào nuôi? Thực nghiệm đã cho thấy tương tác giữa hai protein Toll và Spatzle nằm trên bề mặt trứng,

nơi tiếp giáp với vùng bụng của phôi là dấu hiệu khởi động sự di chuyển của Dorsal. Protein Toll phân bố đồng đều trên bề mặt phía trong màng nguyên sinh bao phủ phôi. Protein Spatzle phân bố trong các khoang giữa màng nguyên sinh và màng noãn hoàng (Hình 7.7).

Protease mã bởi gen *easter* chỉ phân cắt Spatzle phân bố ở vùng bụng phôi thành polypeptide có khả năng tương tác với Toll. Phức Toll- Spatzle được nhận biết bởi Dorsal khiến cho protein này di chuyển vào nhân. Vì sao protease Easter chỉ có hoạt tính với Spatzle ở vùng bụng? Phải chăng gen *easter* chỉ được hoạt hoá ở vùng này hay bản thân sản phẩm của gen đã được dự trữ ở đó và được hoạt hoá một cách cục bộ?

Hoạt động của các gen thuộc nhóm trục lưng-bụng liên quan đến các tín hiệu được trao đổi giữa tế bào trứng và các tế bào nuôi. Tương tác ở mức độ phân tử giữa các tín hiệu này chưa được hoàn toàn sáng tỏ. Tuy nhiên sau khi thụ tinh, kết quả tương tác gây nên sự di chuyển của nhân tế bào trứng từ cực đuôi về phía vùng lưng của phôi. Điều này dẫn đến sự biệt hoá của các tế bào nuôi vùng lưng và vùng bụng. Nhờ đó, protein Spatzle phân bố ở vùng bụng được nhận biết và bị phân cắt bởi protease Easter.

Protein Toll được mã bởi gen *toll* đóng vai trò quyết định trong việc phân trục lưng-bụng. Mọi đột biến ở gen *toll* đều gây rối loạn trong quá trình hình thành trục này. Các gen khác liên quan đến quá trình phân trục đều chỉ tham gia điều hoà hoạt động của protein Toll mà không một gen nào thay thế được nó trong việc khởi động quá trình. Điều đáng lưu ý là protein Toll phân bố đồng đều trong tế bào chất của trứng. Do protein Spatzle chỉ bị phân cắt ở phần dưới (phía bụng) của tế bào trứng nên các tín hiệu (các peptide tạo ra từ Spatzle) hoặc bị receptor Toll tương tác rất nhanh hoặc không có khả năng khuếch tán nên chúng chỉ hoạt hoá được các phân tử Toll phân bố ở phần dưới. Sau khi Toll tương tác với Spatzle, một loạt các phản ứng truyền tín hiệu xảy ra trong trứng: Toll → Tube → Pelle → Cartus → Dorsal. Chức năng của protein Tube chưa được xác định trong khi Toll đóng vai trò thụ cảm còn Pelle có hoạt tính kinase. Protein Toll hoạt hoá gen *tube*; sản phẩm gen *tube* hoạt hoá gen *pelle*. Protein Pelle có chức năng kinase phân huỷ phức chất Cartus-Dorsal, nhờ đó Dorsal được giải phóng ở dạng tự do. Từ tế bào chất, Dorsal di chuyển vào nhân, hoạt hoá các gen có liên quan đến phát triển cấu trúc bụng.



Hình 7.7:

Hình ảnh cắt ngang phôi ruồi giấm *Drosophila* cho biết vị trí phân bố và tương tác giữa protein Toll và

Spatzle khởi đầu quá trình hình thành trục lưng-bụng ở phôi *Drosophila*. Tương tác này xảy ra cục bộ ở vùng bụng của phôi trong khoang giữa màng nguyên sinh và màng noãn hoàng bao quanh phôi. Kết thúc của quá trình này là protein Dorsal được di chuyển vào trong nhân gây hoạt hoá hoặc ức chế các gen *twist*, *snail*, *zerknüllt* và *decapentaplegic*. Hoạt động của các gen này liên quan đến sự hình thành trục lưng-bụng (theo Snustad, 2000).

Protein Dorsal phân bố đồng đều trong tế bào chất của trứng. Sau khi các nhân di chuyển đến bề mặt phôi tạo phôi bì thì các tế bào nằm phía lưng có Dorsal phân bố ở tế bào chất còn các tế bào nằm phía bụng có Dorsal phân bố trong nhân. Tuy nhiên, đầu tiên Dorsal không di chuyển được vào nhân do tương tác với Cartus. Chỉ khi nhận tín hiệu từ Toll, protein Pelle mới có hoạt tính kinase gây phosphoryl hóa Cartus khiến Dorsal được giải phóng ra ở dạng tự do. Lúc đó Dorsal mới di chuyển vào nhân. Chịu sự kiểm soát của Dorsal, các gen *twist* và *snail*, *zerknüllt* và *decapentaplegic* được hoạt hoá hoặc kìm hãm tùy thuộc vào nồng độ của protein này. Các tế bào phía bụng có nồng độ Dorsal trong nhân cao nhất gây hoạt hoá các gen phát triển cấu trúc bụng. Các tế bào phía lưng có nồng độ Dorsal trong nhân thấp nhất cho phép các gen phát triển cấu trúc lưng hoạt động (tức là các gen này bị ức chế ở nồng độ cao của Dorsal). Ở vùng giữa, nồng độ Dorsal đủ cao để gây ức chế một số gen nhưng lại quá thấp để bật mở các gen khác. Điều đó dẫn đến sự phát triển đặc hiệu của ngoại bì.

7.3.4. Nhóm gen qui định phát triển các cấu trúc tận cùng của ấu thể (terminal-group genes)

Phần cấu trúc đặc biệt nằm tận cùng ở các đầu không phân đốt của phôi (cấu trúc *acron* ở phần đầu, cấu trúc *telson* ở phần đuôi) được qui định bởi hoạt động của một số gen. Hoạt động của nhóm gen này cũng tương tự như nhóm gen phân trục lưng-bụng. Sau khi trứng thụ tinh, protein Torso được tổng hợp từ ARNm dự trữ trong tế bào trứng. Protein này nhận các tín hiệu ở các cực của phôi và trở nên hoạt hoá, mở đầu cho các phản ứng truyền tín hiệu khác. Các protein Tarless, Huckebein được tổng hợp từ các ARNm dự trữ ở các cực, chúng tham gia điều khiển bật mở các gen có liên quan đến cấu trúc của các cực tận cùng.

Như vậy các cơ chế hoạt động khác nhau của các gen có nguồn gốc từ mẹ dẫn đến hình thành các trục lưng-bụng và đầu -đuôi của phôi. Đây là những bước đầu tiên trong quá trình phát triển mô phôi để xác định hướng và cấu trúc không gian của phôi. Đối với trục đầu-đuôi, phân bố cục bộ của ARNm, còn đối với trục lưng-bụng phân bố cục bộ của protein, đóng vai trò quyết định. Các ARNm và protein này đều là sản phẩm của các gen có nguồn gốc từ mẹ. Sau khi đã phân trục, hoạt động tiếp theo của các gen thuộc hệ gen lưỡng bội (*zygote*) sẽ quyết định sự hình thành cấu trúc đốt cũng như các đặc tính riêng biệt của từng đốt.

Các gen có nguồn gốc từ mẹ được phiên mã từ cơ thể mẹ và ARNm tương ứng được dự trữ trong trứng. Ngoài ra, chúng còn được hoạt hoá từ hệ gen lưỡng bội của phôi (cá thể con). Ví dụ, gen *hunchback* có ARNm xuất phát từ hai nguồn. Sản phẩm của gen có nguồn gốc từ mẹ cùng với một số gen trong hệ gen lưỡng bội tham gia kiểm soát các gen phân đốt (*gap genes*).

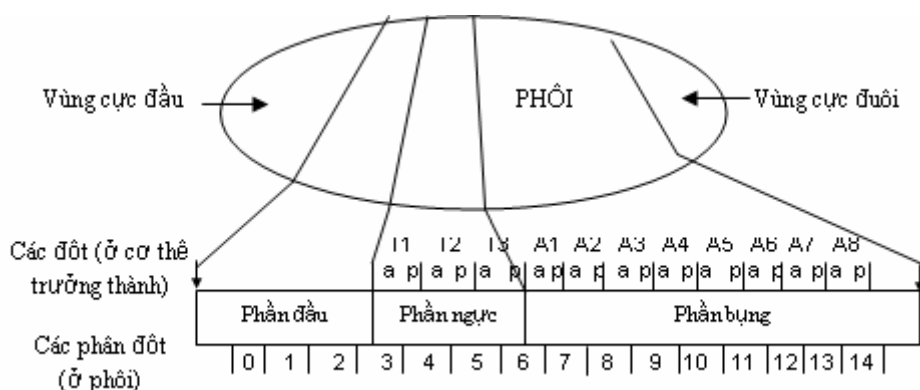
7.4 Hoạt động của các gen trong hệ gen lưỡng bội (phôi)

Ở ruồi giấm trưởng thành, ngoài phần đầu, cơ thể cấu tạo gồm 3 đốt ngực và 8 đốt bụng. Mỗi đốt có thể chia làm 2 tiểu phần, gọi là tiểu phần trước và sau (a và p). Ở giai đoạn phát triển phôi và ấu trùng cũng xảy ra sự tạo đốt. Thực nghiệm quan sát được 14 đốt và nhận thấy mỗi đốt ở phôi sẽ ứng với hai tiểu phần của 2 đốt ở cơ thể trưởng thành. Mỗi đốt ở phôi gồm tiểu phần sau (p) của đốt này và tiểu phần trước (a) của đốt liền kề ở cơ thể trưởng thành (Hình 7.8). Vì vậy các đốt của phôi được gọi là phân đốt. Có thể quan sát được các phân đốt

này sau khi phôi phát triển được 6h. Tuy nhiên, sau 9h thì phân đốt biến mất. Biên giới giữa hai phân đốt trở thành phần giữa của đốt mới. Phân đốt thứ tư sẽ ứng với tiểu phần P của đốt ngực T1 và tiểu phần A của đốt ngực T2 (Hình 7.8).

Khi phôi phân chia đến lần thứ 11, các gen của hệ gen hợp tử (phôi) bắt đầu hoạt động. Hoạt động khác nhau của các gen này giữa các vùng hoặc ngay trong một vùng khiến cho các sản phẩm của chúng cũng phân bố khác nhau. Các gen phân đốt "gap genes" hoạt động chia phôi thành những vùng lớn, trong đó xảy ra sự phối hợp đa dạng của các protein nhằm điều khiển sự biệt hoá của các vùng. Điều đó quyết định số phận của mỗi vùng sẽ phát triển thành một số đốt riêng biệt. Tuy nhiên định mệnh này chỉ được hoàn thiện nhờ hoạt động của các gen tạo cặp đốt, các gen phân cực đốt.

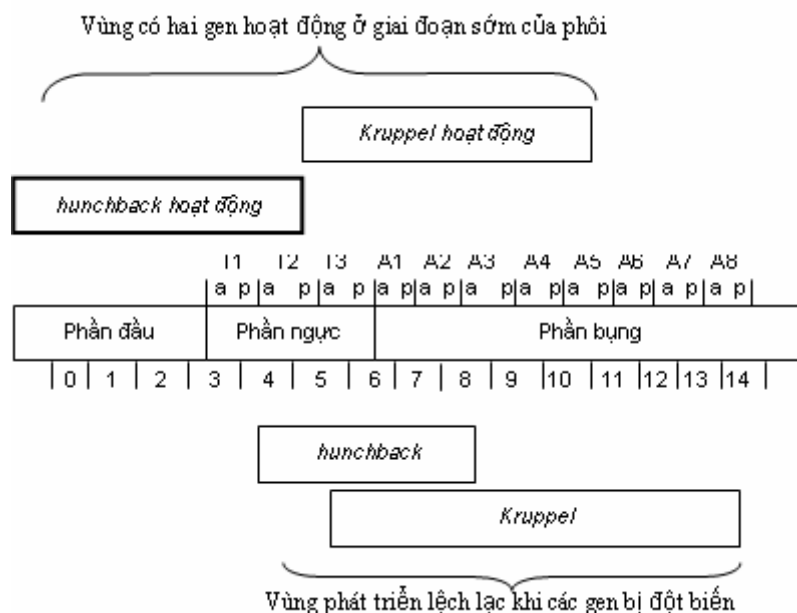
Ấu thể ruồi giấm cũng như cơ thể trưởng thành cấu tạo gồm các đốt. Quá trình hình thành các đốt liên quan đến hoạt động của khoảng 30 gen. Đột biến ở một gen bất kỳ đều ảnh hưởng đến số lượng các đốt hoặc cấu tạo của từng đốt. Các gen phân đốt hoạt động sau các gen có nguồn gốc từ mẹ. Các gen này đều thuộc hệ gen lưỡng bội của phôi. Chúng được chia làm 4 nhóm: **nhóm gen tạo đốt, nhóm gen cặp đốt, nhóm gen phân cực đốt và nhóm gen chọn lọc**. Việc phân nhóm phụ thuộc vào biểu hiện tính trạng khi gen bị đột biến và thời điểm cũng như vị trí mà gen hoạt động (Hình 7.9). Chúng thường bật mở sau 2 giờ trứng được thụ tinh, trước khi hình thành các tế bào trong phôi.



Hình 7.8:

Vị trí tương ứng của các phân đốt xuất hiện trên phôi sau 6h phát triển và các đốt xuất hiện sau 9h. Ngoài phần đầu, các đốt ngực gồm T1-T3 và các đốt bụng gồm A1-A8 (theo Lodish & cs., 2000).

Phần lớn sản phẩm của các gen có nguồn gốc từ mẹ và của một số gen phân đốt đóng vai trò yếu tố phiên mã đối với nhóm gen phân đốt. Những yếu tố này cùng lúc đảm nhận hai chức năng vừa hoạt hoá một số gen phân đốt này đồng thời vừa kìm hãm một số gen khác. Do đó nồng độ ngưỡng của một factor cũng như tỷ lệ nhất định giữa factor hoạt hoá và factor ức chế sẽ quyết định mức độ biểu hiện của từng gen phân đốt. Ví dụ nồng độ protein Hunchback, sản phẩm của gen phân đốt đầu tiên mà cao thì quá trình phiên mã của gen *kruppel* bị kìm hãm nhưng nếu nhỏ dưới một ngưỡng nhất định thì hoạt động của gen này lại được hoạt hoá. Như vậy nồng độ ngưỡng của Hunchback quyết định ranh giới xuất hiện của protein Kruppel. Gradient nồng độ Hunchback giảm dần từ đầu đến đuôi của phôi. Do đó, nồng độ Kruppel sẽ tăng dần theo chiều đó. Một cách tương tự, Kruppel khi đạt đến ngưỡng nào đó sẽ quyết định hoạt động của những gen phân đốt khác tiếp theo.



Hình 7.9:

Vùng hoạt động của các gen *hunchback* và *kruppel*. Hai gen này mã cho protein điều khiển. Chúng hoạt động ở các phân đốt phần đầu và ngực (dựa vào sự phân bố của ARNm tại những phân đốt này trong phôi). Tuy nhiên, các protein Hunchback và Kruppel lại ảnh hưởng đến biệt hoá phát triển của những đốt nằm sau các đốt có hai gen này phiên mã. Sự phối hợp của các protein điều khiển kiểm soát sự biệt hoá khác nhau của các vùng trong phôi (theo Lodish & cs., 2000).

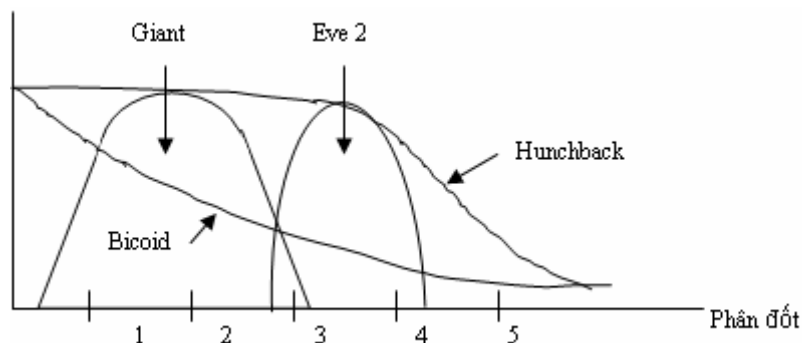
7.3.5. Các gen tạo đốt "gap"

Đột biến ở những gen này gây thiếu hụt một hoặc một số đốt nằm cạnh nhau. Các gen *hunchback*, *kruppel* được xem là những gen tạo đốt do đột biến ở những gen này làm cho các đốt phần đuôi không được tạo thành. Đột biến đơn lẻ thường không gây tác động nghiêm trọng đến việc xuất hiện các đốt phần đuôi nhưng đột biến ở cả hai gen *bicoid* và *kruppel* khiến cho phôi không có toàn bộ các đốt vùng đuôi.

7.3.6. Các gen cặp đốt "pair-rule"

Đột biến ở chúng gây thiếu hụt các đốt chẵn hoặc lẻ làm cho số đốt giảm đi một nửa. Sản phẩm của mỗi gen tạo cặp đốt sẽ xuất hiện trong 7 đốt chẵn hoặc 7 đốt lẻ. Trong mỗi đốt, gen được kiểm soát bởi các yếu tố hoạt hóa và kim hãm một cách độc lập với nhau. Hầu hết các factor phiên mã này là sản phẩm của gen có nguồn gốc từ mẹ hoặc gen tạo đốt hoạt động trước đó.

Hoạt động của gen *eve* tạo đốt chẵn số 2 được xem là ví dụ về sự phối hợp kiểm soát của các factor mã bởi gen *bicoid* và các gen tạo đốt *hunchback*, *kruppel* và *giant*. Các protein này tương tác với những vị trí đặc hiệu nằm trước promoter gen *eve*. Bicoid và Hunchback kích thích hoạt động của *eve* trong phạm vi rộng nhưng Kruppel và Giant lại kim hãm *eve* trong khoảng hẹp xung quanh ranh giới giữa hai tiểu phần trước và sau. Gradient nồng độ mỗi factor điều khiển này biến đổi theo qui luật đặc thù riêng cho từng protein (Hình 7.10). Vì vậy, chúng kiểm soát chặt chẽ ranh giới giữa các tiểu phần của từng đốt.



Hình 7.10:

Biến đổi nồng độ protein Eve cũng như các factor hoạt hoá Bicoid, Hunchback và yếu tố kim hãm Giant trong quá trình kiểm soát hoạt động của gen eve. Phối hợp giữa các protein xác định chính xác ranh giới giữa các tiểu phần trước và sau của đốt thứ hai. Một cách tương tự, sự xuất hiện của các đốt khác được quyết định do tổ hợp các factor được mã bởi gen có nguồn gốc từ mẹ và gen phân đốt.

7.3.7. Các gen phân cực đốt

Gen phân cực đốt qui định trật tự các vùng ngay trong một đốt. Đây là những gen mà khi bị đột biến thường không làm thay đổi số lượng đốt nhưng một phần nào đó của đốt sẽ bị loại bỏ và thay thế bởi phần còn lại của đốt ở vị trí đối xứng (hình ảnh gương). Ví dụ, đột biến trên gen *gooseberry* khiến cho tiểu phần sau của mỗi đốt không được hình thành. Thay vào đó là cấu trúc của tiểu phần trước nhưng sắp xếp theo kiểu hình ảnh gương.

Hoạt động của các gen phân đốt xảy ra tương tự như các gen có nguồn gốc từ mẹ, tức là sản phẩm của gen này sẽ bật mở các gen tiếp theo hoặc ức chế các gen hoạt động trước đó. Như vậy, các gen phân đốt hoạt động theo không gian và thời gian. Hoạt động của các gen phân đốt sẽ phân chia phôi thành những phần nhỏ dần dọc theo trục đầu-đuôi. Trong mỗi phần nhỏ, hoạt động của các gen chọn lọc sẽ qui định các đặc tính riêng biệt của từng phần.

7.5 Các gen chọn lọc

Các gen liên quan đến sự hình thành các đốt ở phôi ruồi giấm được điều khiển theo cách thức sao cho sản phẩm của một số gen hoạt động trước sẽ quyết định mức độ biểu hiện của các gen tiếp theo. Nhờ đó, từng nhóm tế bào sẽ có chung số phận quyết định sự phân vùng dọc theo trục đầu-đuôi. Tuy nhiên, trong mỗi vùng, điều gì khiến các tế bào biệt hoá khác nhau để tạo nên cấu trúc đặc thù riêng của vùng đó? Nói cách khác vì sao mỗi cơ quan chỉ phân bố tại một vùng nhất định trên cơ thể? Liệu chúng ta có thể gây ảnh hưởng đến hoạt động của các gen để thay đổi vị trí của từng bộ phận được hay không? Ví dụ, ở ruồi giấm, có thể gây đột biến để mắt xuất hiện ở vùng bụng hoặc chân xuất hiện ở vùng đầu hay không?

Bằng kỹ thuật gây đột biến và sàng lọc các cá thể có sự sai lệch về vị trí của một bộ phận trên cơ thể, di truyền học phân tử đã xác định được nhóm gen đặc biệt gọi chung là gen chọn lọc (*selector genes*). Đây là những gen hoạt động liên tục từ giai đoạn phát triển phôi đến cơ thể trưởng thành với chức năng đảm bảo và duy trì tính đặc thù của từng vùng dọc theo các trục đầu-đuôi, lưng-bụng trong suốt quá trình sinh trưởng và phát triển. Đầu tiên những gen này được gọi là gen *homeotic* do đột biến ở những gen này gây nên sự sai lệch vị trí của một số bộ phận trên cơ thể. Tuy nhiên, tổng hợp kết quả nghiên cứu trên những gen liên quan đến sự duy trì tính biệt hoá đặc thù của tế bào tại mỗi vùng mà chúng được gọi là gen chọn lọc. Khác với các nhóm gen có nguồn gốc từ mẹ hay gen phân đốt chỉ hoạt động nhất thời, các gen

chọn lọc một khi đã bật mở thì phải hoạt động liên tục. Như vậy hoạt động của các gen chọn lọc nhằm xác định cấu trúc của các phần cơ thể khác nhau dọc theo các trục trong suốt thời gian cá thể tồn tại.

Các gen chọn lọc cùng hoạt động hoặc hoạt động sau các gen phân cực đốt. Đột biến xảy ra ở những gen này làm cho một phần của cơ thể biểu hiện các tính trạng của phần khác hoặc các bộ phận của đốt này lại xuất hiện ở đốt khác. Ví dụ, chân mọc ở đầu hoặc xuất hiện hai cặp cánh giống hệt nhau. *Cần lưu ý rằng khi các gen chọn lọc bị đột biến hoặc hoạt động của chúng xảy ra không đúng thời điểm hoặc ở vị trí không thích hợp, chỉ có vị trí các bộ phận bị sai lệch chứ không hề tạo ra bộ phận mới.* Định mệnh phát triển của các tế bào được qui định bởi các gen hoạt động trước các gen chọn lọc, tuy nhiên định mệnh này sẽ không được thực hiện trọn vẹn khi hoạt động của các gen chọn lọc xảy ra không đúng vị trí. Phần lớn các gen này mã cho các protein điều khiển (liên kết với ADN tại các enhancer hoặc promoter). Chúng làm nhiệm vụ bật mở các gen ở những mức độ khác nhau. Do đó, một đột biến xảy ra đối với một gen chọn lọc không chỉ gây bất hoạt gen đó mà còn thay đổi hoạt động của các gen chọn lọc khác.

Cho đến nay, 8 gen chọn lọc được phát hiện ở ruồi giấm và chúng được xếp vào 2 nhóm Bithorax (BX-C) và Antennapedia (ANT-C) cùng phân bố trên nhiễm sắc thể số 3, cách nhau khoảng 10 cM. Nhóm Antennapedia có 5 gen (*lab, pb, Dfd, Scr, Antp*) kiểm tra sự sai khác giữa các đốt đầu và đốt ngực (từ phân đốt 1-5) còn nhóm Bithorax có 3 gen (*Ubx, abdA, AbdB*) làm nhiệm vụ kiểm tra sự khác nhau giữa các đốt bụng và đốt ngực (phân đốt 5-14). Đột biến ở những gen này thường biểu hiện tính trạng ở cấu trúc biểu bì, hệ cơ, thần kinh dọc theo trục đầu-đuôi. Tất cả 8 gen đều có tính chất chung là vùng không chứa mã di truyền (*intron*) rất lớn. Những vùng này có vai trò rất quan trọng trong kiểm soát mức độ biểu hiện của gen theo thời gian và không gian dọc trục đầu-đuôi.

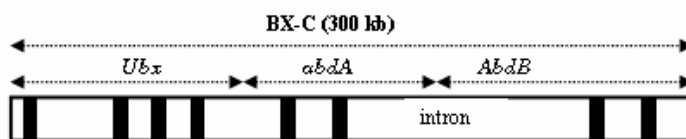
Những gen chọn lọc có vai trò tương tự BX-C và ANT-C đã được phát hiện ở động vật có vú. Tuy nhiên, chúng phân bố rải rác thành bốn nhóm trên các nhiễm sắc thể khác nhau. Tất cả những gen này ở động vật đều được gọi chung là nhóm gen *Hox*. Riêng đối với ruồi giấm thì các gen *Hox* (thuộc hai nhóm BX-C và ANT-X) còn được gọi là gen *Hom*. Nhìn chung các gen *Hox* đều có cùng những đặc điểm sau:

+ Vùng ADN điều khiển của gen *Hox* chứa một đoạn ADN dài khoảng 180 bp có tính chất bảo toàn. Đoạn này được gọi là homeo box. Homeo box mã cho đoạn amino axit (khoảng 60 a.a) có khả năng tương tác với ADN. Đoạn amino axit này được gọi là homeo domain.

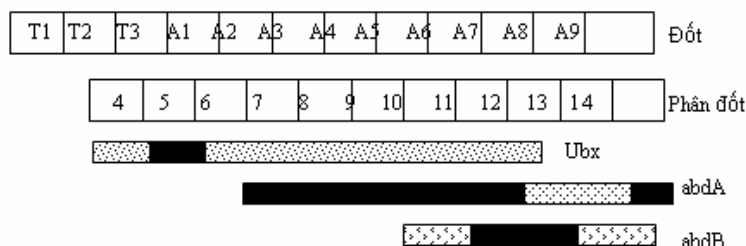
+ Các exon của một gen thường nằm xa nhau, phân cách bởi các đoạn ADN dài không chứa mã di truyền nhưng đóng vai trò điều khiển hoạt động của gen.

+ Các gen trong mỗi nhóm nằm dọc nhiễm sắc thể theo một thứ tự nhất định, chính xác như trình tự hoạt động của chúng dọc theo chiều dài cơ thể từ đầu đến đuôi. Các gen này được hoạt hoá nối tiếp nhau nhờ các tín hiệu truyền dọc nhiễm sắc thể, phù hợp với chỉ dẫn về khoảng cách giữa các phần của cơ thể (tính theo trục đầu-đuôi).

A. Cấu trúc của nhóm BX-C trên nhiễm sắc thể số 3



B. Hoạt động theo không gian của các gen chọn lọc thuộc nhóm BX-C



Hình 7.11:

Cấu trúc của các gen chọn lọc thuộc nhóm BX-C. (A). Các exon được ký hiệu bởi các đen đậm. Các intron kích thước lớn đóng vai trò rất quan trọng trong kiểm soát hoạt động của gen theo không gian và thời gian. (B). Mức độ biểu hiện (ARNm/protein) của các gen thuộc nhóm BX-C trong các phân đốt 4 đến 14. Các vùng đậm chỉ mức độ biểu hiện cao của gen trong từng vùng dọc theo trục đầu đuôi (theo Alberts & cs., 2002).

Nghiên cứu chi tiết vai trò của các gen *Hox* thuộc nhóm BX-C cho thấy thứ tự hoạt động của 3 gen tuân theo trật tự sắp xếp trên nhiễm sắc thể. Sản phẩm của chúng phân bố dọc theo trục đầu đuôi cũng tuân theo qui luật nghiêm ngặt. Hoàn toàn không có sự phân bố đồng đều mà tồn tại sự chênh lệch nồng độ protein của 3 gen ở giữa các phân đốt (Hình 7.11).

Khi cả 3 gen của nhóm BX-C bị đột biến thì phôi bị chết. Phân tích lớp biểu bì của phôi chết phát hiện thấy toàn bộ các phân đốt 5 đến 13 đều giống hệt phân đốt 4. Nói cách khác, đột biến toàn bộ nhóm BX-C khiến cho các phân đốt 5-13 không được tạo thành. Như vậy, nhóm Bithorax chỉ có 3 gen nhưng chịu trách nhiệm về sự sai khác giữa 10 phân đốt. Đó có thể là kết quả của phản ứng biến đổi phân tử ARNm (cắt intron theo các trật tự khác nhau) dẫn đến nhiều loại phân tử ARNm được tổng hợp từ một gen. Mặt khác rất nhiều đột biến được tìm thấy phân bố trên những đoạn ADN không chứa mã di truyền. Trật tự các đột biến này trên nhiễm sắc thể cũng tương ứng với trình tự sắp xếp của các vùng cơ thể mà chúng có ảnh hưởng. Điều này cho thấy sự khác biệt giữa các vùng của cơ thể không chỉ do sản phẩm của các gen *Hox* quyết định mà còn phụ thuộc các đoạn ADN không chứa mã di truyền (các intron). Nói cách khác, vị trí của tế bào chỉ có ý nghĩa trong phát triển phôi nhờ hoạt động của các gen *Hox* cũng như cách thức điều khiển hoạt động của chúng.

Hiện nay vấn đề đang được các nhà nghiên cứu sinh học quan tâm đến chính là xác định cơ chế nào giúp các gen *Hox* nhớ được vị trí hoạt động của chúng. Cơ chế điều khiển phản hồi có thể có liên quan. Thực nghiệm đã chứng minh được sản phẩm của một số gen *Hox* hoạt hoá phản ứng tổng hợp ARNm trên gen của chính nó. Ví dụ, gen *dfd* trong nhóm *Antennapedia* có nhiều vị trí liên kết với chính protein Dfd. Sự liên kết này đủ để duy trì hoạt động của gen. Ngoài ra, nhóm gen *polycom* gồm các gen cần thiết để kìm hãm hoạt động của các gen *Hom*. Bất kỳ một gen nào đó trong nhóm này bị đột biến, các gen *Hom* sẽ hoạt động, mới đầu ở mức độ bình thường và đúng vị trí, nhưng sau đó chúng hoạt động không chịu điều khiển và hoạt động trên mọi vùng của phôi. Protein Polycom bám vào chromatin của gen *Hom* làm thay đổi cấu trúc chromatin. Do đó sự biến dạng chromatin có thể giúp cho gen *Hom* nhớ được vị trí hoạt động của mình.

Dùng đoạn homeo box của các gen *Hom* ở ruồi giấm làm mồi cho các phép lai Southern với ADN genome tách ra từ các loài động vật khác nhau, kết quả cho thấy các gen *Hox* đều tồn tại trong genome ếch, chuột và người. Đặc biệt các gen này đều hoạt động trong quá trình phát triển phôi. Khi đưa các gen *hox* bị đột biến của chuột vào phôi ruồi giấm, chúng gây ra sự phát triển bừa bãi giống như khi các gen *Hom* của ruồi bị đột biến. Các gen *Hox* ở động vật cũng đều hoạt động theo thời gian và không gian tương ứng với trình tự sắp xếp của chúng trên nhiễm sắc thể và theo trật tự nhất định từ đầu đến chân. Tuy nhiên giữa các gen *Hox* và gen *Hom* có sự khác nhau: Các gen *Hox* thường nhỏ không quá dài và số lượng gen trong mỗi nhóm nhiều hơn. Có thể tồn tại nhiều bản sao của một gen *Hox* hoặc nhiều bản sao của các nhóm gen trong genome, do đó khó phát hiện được những đột biến biểu hiện ra tính trạng.

Cùng với các gen BX-C và ANT-C, di truyền phân tử đã phát hiện hàng loạt các gen homeotic khác có liên quan đến sự hình thành của từng cơ quan riêng biệt phân bố ở từng vùng nhất định trên cơ thể. Ví dụ điển hình là gen *eyeless* qui định vị trí của mắt ở ruồi giấm. Sản phẩm của gen này có chứa homeo domain. Protein mã bởi *eyeless* đóng vai trò yếu tố phiên mã liên quan đến biểu hiện của rất nhiều gen khác qui định sự hình thành mắt ở đúng vị trí. Thực nghiệm đã tạo ra ruồi giấm chuyển gen mang gen *eyeless* chịu sự điều khiển của promoter lạ mà promoter này chỉ hoạt động trong mô không liên quan đến việc hình thành mắt. Sản phẩm của gen *eyeless* xuất hiện không đúng vị trí khiến cho mắt có thể xuất hiện ở chân, cánh hoặc vùng ngực. Những mắt đặc biệt này có thể nhận biết ánh sáng như mắt bình thường khác. Động vật có vú có chứa gen tương đồng với gen *eyeless* của ruồi giấm. Ví dụ gen tương đồng ở chuột được gọi là gen *Small eye* do đột biến ở gen này khiến mắt chuột bị nhỏ lại. Lặp lại thí nghiệm chuyển gen *Small eye* của chuột vào ruồi giấm (tương tự thí nghiệm nêu trên) thì ruồi giấm chuyển gen cũng có mắt xuất hiện sai lệch vị trí. Rõ ràng gen *Small eye* ở chuột là gen homeotic và mã cho protein chứa homeo domain. Gen *Small eye* cũng có chức năng qui định việc hình thành mắt đúng vị trí. Tuy nhiên, ruồi giấm mang gen *Small eye* lại có thêm mắt ruồi chứ không phải mắt chuột. Như vậy sản phẩm của hai gen *Small eye* và gen *eyeless* có chức năng tương tự như nhau, tức là có khả năng điều khiển một loạt gen khác của ruồi giấm để dẫn đến kết quả cuối cùng là mắt xuất hiện ở vị trí mới. Vì vậy nhất định mắt ruồi chứ không thể là mắt chuột được hình thành. Gen tương đồng với hai gen *Small eye* và gen *eyeless* đã được phát hiện ở người. Đột biến ở gen này khiến tròng đen của mắt bị nhỏ hoặc mất hẳn.

Kết quả nghiên cứu của di truyền phân tử trong những năm cuối thế kỷ 20 trên đối tượng mô hình *Drosophila* đã làm sáng tỏ phần nào quá trình phát triển của trứng sau thụ tinh. Hoạt động kế tiếp nhau theo một trật tự nghiêm ngặt của các gen có nguồn gốc từ mẹ, gen tạo đốt, gen chọn lọc đã quyết định việc hình thành các trục của cơ thể, sự phân vùng phân đốt, sự biệt hoá của các tế bào, sự xuất hiện các cơ quan trong từng vùng... Đặc biệt sự tương đồng của các gen homeotic được tìm thấy ở ruồi giấm cũng như động vật có vú và con người. Phải chăng những gen này đều có chung nguồn gốc từ một gen ban đầu. Trải qua thời gian tiến hoá, chúng vẫn phần nào hỗ trợ cho nhau để thực hiện một chức năng cụ thể trong sinh trưởng và phát triển?

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.
- allele: Một trong nhiều dạng của gen phân bố trên nhiễm sắc thể tương đồng. Một gen có 2 hoặc nhiều allel (đa allel).
- allele-specific oligonucleotide (ASO) hybridization: Lai xác định một trong hai dạng sợi đơn của một phân tử ADN kép bằng oligonucleotide.
- allosteric transition: Sự thay đổi cấu trúc bậc 3 hoặc bậc 4 của protein do tương tác với phân tử trọng 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.
- amphipathic: Phân tử hoặc cấu trúc bao gồm hai phần kỵ nước và ưa nước.
- annealing: gắn môi. 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 phân 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.
- apoptosis: chết theo chương trình. Quá trình gây chết tế bào theo chương trình được mã trong hệ gen với những biểu hiện hình thái đặc thù riêng.
- 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 excision 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.

- blunt end: đầu bằng. Đầu tận cùng của sợi kép ADN kết thúc bằng hai nucleotide liên kết bổ sung với nhau.
- ADNc (*complementary DNA*): phân tử ADN bổ sung được phiên mã từ ARNm bằng enzym phiên mã ngược. ADNc 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.
- chimera: cơ thể có chứa hai hay nhiều loại tế bào có nguồn gốc di truyền khác nhau.
- chromatin: nhiễm sắc chất, bao gồm ADN, histone và protein nonhistone phân bố trong nhân tế bào eukaryot. Trong giai đoạn phân bào ở metaphase, chất nhiễm sắc co đặc 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.
- chromosome walking: đi trên nhiễm sắc thể. Kỹ thuật cho phép xác định thứ tự sắp xếp của các đoạn ADN trên nhiễm sắc thể với điều kiện các đoạn này phải nằm gồi lên nhau.
- 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 có 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 qui 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 gồi 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.
- degradosome: Phức enzym có chức năng phân hủy ARNm ở vi khuẩn.
- 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 như 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.
- dosage compensation: hiệu quả của một gen tăng hoặc giảm phụ thuộc vào số lượng bản sao của gen đó trong genome.
- dot-blot hybridization: kỹ thuật lai giữa đầu dò ADN hoặc ARN với ADN hoặc ARN với những nồng độ đã biết cố định trên màng.
- embryonic stem (ES) cell: tế bào toàn năng có nguồn gốc từ phôi thai.
- 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 phân tử kháng nguyên tương tác với kháng thể.
- euchromatin: chất nhiễm sắc thực. Là những đoạn ít co đặc của chất nhiễm sắc trong giai đoạn gian kỳ. Chất nhiễm sắc thực 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 ADNc 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 ADNc hoặc đoạn ADN của hệ gen (genome) dựa vào các đặc tính chức năng của protein được mã bởi

ADNc 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 ADNc hay gen tương ứng với protein đó.

- expression vector: Những plasmid hoặc genome virus chuyên chở gen hoặc ADNc vào tế bào nhận thích hợp. Protein do ADNc 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.

- F plasmid: yếu tố giới tính F kiểm soát chuyển gen bằng con đường tiếp hợp giữa vi khuẩn.

- 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 polypeptide được mã bởi các gen khác nhau.

- G protein: Các protein có cấu trúc trimer (gồm 3 tiểu đơn vị) có khả năng liên kết và phân hủy GTP. G protein tham gia con đường truyền tín hiệu trong tế bào.

- gain-of-function mutation: đột biến tạo tính trạng mới.

- 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 conversion: Chuyển hoán gen. Hiện tượng các allele 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 allen của một gen hoặc rộng hơn là toàn bộ các kết cấu di truyền của một tế bào hoặc một cơ thể.
- germ cell: Tế bào mầm (thường chỉ dòng tế bào sinh dục). Đây là loại tế bào phát triển thành giao tử.
- germ line: Dòng tế bào mầm. Khái niệm này còn được dùng chỉ vật chất di truyền từ thế hệ trước sang thế hệ sau thông qua giao tử.
- 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.
- heterochromatin: 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.
- homeobox: Trình tự ADN bảo thủ mã cho một đoạn oligopeptide gọi là homeodomain, của protein kiểm soát phiên mã. Đoạn này tương tác với ADN giúp protein thực hiện chức năng của chúng. Gen mã cho những protein có vùng homeodomain được gọi là “homeotic gene”.
- homeosis: Sự biến đổi một bộ phận cơ thể thành bộ phận khác gây ra do đột biến hoặc do rối loạn hoạt động của một số gen quan trọng cho biệt hoá phát triển phôi thai.
- 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.
- Hox complex: Nhóm các gen chọn lọc tương đồng với nhau. Các gen này quyết định sự hình thành các bộ phận khác nhau của cơ thể.
- hybridization: Sự kết hợp của 2 sợi đơn nucleic bổ xung 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.
- hybridoma: Dòng tế bào lai tồn tại vĩnh cửu và tạo ra kháng thể đơn dòng. Dòng tế bào này được tạo ra nhờ sự dung hợp giữa tế bào lympho B bình thường và tế bào u tủy.
- intron: Những phần của sợi ARN đầu tiên vừa được phiên mã từ ADN sẽ bị loại bỏ đi 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.
- kinase: Enzym chuyển nhóm phosphate cuối cùng của ATP sang cơ chất. Kinase đóng vai trò đặc biệt quan trọng trong kiểm soát hoạt tính của các protein trong tế bào.
- kilobase pair: 1000 cặp base (1000 bp).
- kinetochore: Cấu trúc protein ba lớp 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.
- knockin gene: Kỹ thuật trong đó trình tự mã hoá của gen này bị thay thế bởi trình tự của gen khác.
- knockout gene: Kỹ thuật gây bất hoạt gen một cách chọn lọc bằng cách thay thế gen bình thường của cơ thể bởi gen đột biến.
- 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.
- leaky mutation: đột biến tác động một phần đến tính trạng.
- Ligand: kích thể. Phân tử thường có trọng lượng nhỏ, tương tác đặc hiệu với thụ thể hoặc đại phân tử tạo nên phức. Phức này khởi động chuỗi các phản ứng khiến tế bào trả lời lại sự có mặt của kích thể.
- 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 H1.
- locus: Vị trí của gen trên nhiễm sắc thể. Mọi allele của một gen đều phân bố tại một locus.
- LTR element: đoạn ADN lặp lại tận cùng bởi 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.
- maternal-effect gene: gen có nguồn gốc từ mẹ. Những gen này được phát hiện ở ruồi giấm *Drosophila*. ARNm của những gen này được tích lũy trong trứng và được dịch mã trong giai đoạn phát triển sớm của phôi.
- metastasis: Di căn. Sự di chuyển của các tế bào ác tính từ vị trí phát sinh ban đầu đến vị trí mới và phát triển vô tổ chức tại đó.
- 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 lượng 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).
 - mitogen: tác nhân kích thích tế bào bước vào chu kỳ phân chia.
 - 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.
 - morphogen: Mã lượng. Phân tử can thiệp đến đặc tính của tế bào trong phát triển biệt hoá (giai đoạn phôi thai) tùy thuộc vào nồng độ của phân tử đó.
 - motif: Thường dùng để chỉ cấu trúc không gian cục bộ của protein mà cấu trúc đó liên quan đến một chức năng.
 - motor protein: Nhóm enzym phân hủy ATP để trượt trên vi sợi myosin hoặc trên vi ống dynein và kinesin.
 - 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 146 bp 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.
 - oncogene: Gen gây ung thư. Sản phẩm của gen gây ung thư có tác dụng biến đổi tế bào lành thành ung thư. Hầu hết các gen gây ung thư xuất phát từ các gen tiền ung thư (*proto-oncogene*). Các gen tiền ung thư mã cho các yếu tố kiểm soát tăng sinh hoặc phân chia tế bào. Khi bị đột biến, phần lớn gen tiền ung thư trở thành gen gây ung thư.
 - 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.
- PEST sequences: trình tự acid amin gây ảnh hưởng đến độ phân hủy của protein chứa trình tự PEST đó.
- polymerase chain reaction (PCR): phản ứng trùng hợp nhân bản ADN. Kỹ thuật cho phép nhân bản 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 phân tử ADN.
- 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.

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

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

- RACE (*Rapid Amplification of ADNc ends*): Kỹ thuật PCR nhân bản phân đầu của phân tử ADNc.

- Ras protein: Phân tử monomer tương tác với GTP, phân bố trong tế bào chất, tham gia vào con đường truyền tín hiệu sau khi được hoạt hoá bởi liên kết giữa kích thể và thụ thể trên bề mặt tế bào.

- 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. 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 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).

- replication origin: tâm sao chép (tâm tái bản): Trình tự nucleotide duy nhất tại đó sao chép ADN được bắt đầu. Trên một nhiễm sắc thể eukaryot có nhiều tâm sao chép nhưng một nhiễm sắc thể prokaryot chỉ có một.

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

- reporter gene: gen báo cáo. Dựa vào tính trạng đã biết của gen báo cáo để phân tích các đoạn nucleotide kiểm soát hoạt động của gen này.

- 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 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.

- ribozym: Phân tử hoặc một đoạn ARN có hoạt tính xúc tác phản ứng.
- 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.
- selfish DNA: DNA không có bất cứ chức năng gì trong tế bào.
- sequence tagged site (STS): Trình tự nucleotide duy nhất trong genome.
- shotgun approach: cách thức xác định trình tự của một phân tử ADN bằng việc xác định trình tự các phân đoạn khi phân tử đó bị cắt đứt một cách ngẫu nhiên.
- silencer sequence: trình tự nucleotide khởi động sự co đặc cục bộ ở 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.
- stem cell: tế bào gốc. Tế bào có tiềm năng phân chia tạo thành những dòng tế bào biệt hoá khác nhau trong sinh trưởng phát triển.
- 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 ARN 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.
- transgenic: động thực vật chuyển gen.
- 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 đánh 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 đại.