

Trường Đại Học Nguyễn Tất Thành
Khoa Công Nghệ Sinh Học



**ỨNG DỤNG CÔNG NGHỆ
GENE TRONG CHĂM SÓC
SỨC KHỎE NGƯỜI**

Giảng viên: TS. TRẦN HOÀNG DŨNG

Tháng 04/2012

Chương 1

CẤU TRÚC GENE, SỰ BIỂU HIỆN GENE VÀ CƠ SỞ PHÂN TỬ CỦA DI TRUYỀN

1. GIỚI THIỆU

Từ khi thuật ngữ “gene” được đặt ra bởi nhà thực vật học Đan Mạch Wilhelm Johannsen vào năm 1909, khái niệm gen được mở ra. Lúc đầu, gen được cho là một thực thể trừu tượng không có một ý nghĩa vật chất – cấu trúc nào. Nó có ý nghĩa đối với những nhà tự nhiên học quan tâm đến sự di truyền của những biến đổi có lợi cung cấp vật liệu cho tiến hóa.

Vào những năm đầu thập niên 50, thí nghiệm của Seymour Benzer trên locus rII của T4 bacteriophage đã giúp định nghĩa gene dưới dạng một đơn vị chức năng, gọi là “cistron”. Khái niệm cistron mô tả cistron là một chuỗi DNA liên tục mã hóa cho một polypeptide thông qua sự phiên mã ra RNA. Nghiên cứu sâu hơn của Charles Yahofsky và Harvey Itano đã cho ra giả thuyết “1 cistron-1 polypeptide”. Khái niệm gene-protein được xác định độc lập bởi Sydney Brenner và Charles Yanofsky và mô hình operon do Francois Jacob và Jacques Monod đưa ra vào những năm đầu thập niên 60 cũng tán thành khái niệm cistron này. Mô hình operon giải thích sự phiên mã một cistron được điều hòa như thế nào, trong khi mô hình gene-protein chứng minh rằng một đột biến trên gene (cistron) gây ra sự biến đổi trình tự amino acid trên protein. Vì vậy, mô hình điều hòa sự biểu hiện cistron (gene) thông qua tương tác promoter-operator đã giúp thống nhất những khía cạnh về cấu trúc và chức năng của gene thành một khái niệm gene duy nhất.

Khái niệm gene này đã được xem xét một lần nữa thông qua những khám phá độc lập được công bố vào năm 1977 bởi Phillip Sharp và Richard Roberts, theo sau đó là một chuỗi những công bố tương tự. Những khám phá này chứng minh rằng gene không nhất thiết tồn tại như là một chuỗi DNA liên tục mà nó còn có thể tồn tại một cách ngắt quãng: vùng mã hóa của một gene (cistron) bị ngắt quãng bởi những trình tự không mã hóa xen kẽ (intron). Gene được phiên mã cho ra một chuỗi dài gọi là “heterogeneous nuclear RNA” (hnRNA) hay “tiền-mRNA” (pre-mRNA). Việc xử lý những “tiền-mRNA” liên quan đến ba sự kiện: gắn mũ chụp (capping), polyadenyl hóa và cắt nối (splicing). Gắn mũ chụp là gắn thêm một cái “mũ” (m^7G) vào base đầu tiên của mRNA ở đầu 5'; polyadenyl hóa là việc gắn thêm một chuỗi dài các nucleotide Adenyl (khoảng 200-250 ở eukaryote) vào đầu 3' của mRNA; và cắt nối (splicing) là việc loại bỏ các đoạn intron tạo thành mRNA trưởng thành. Những vùng gene có hiện diện trên tiền-mRNA mà không hiện diện trên mRNA trưởng thành được gọi là “intron”, những đoạn hiện diện trên mRNA trưởng thành gọi là “exon”. Thuật ngữ exon và intron được đưa ra bởi Walter Gilbert.

Sự phát triển của khái niệm gene từng phần (gồm exon và intron) trên đã không làm mất đi khái niệm cistron, nó vẫn đúng đối với những gene không có

intron như ở những gene prokaryote và một số eukaryote. Thuật ngữ “cistron” hiện nay đã được thay thế bằng thuật ngữ “khung đọc mở” (ORF).

Cùng với sự hiểu biết ngày càng cao về cấu trúc và chức năng của gene, quá trình và sự điều hòa phiên mã, sau phiên mã, dịch mã và sau dịch mã, đã có nhiều khám phá đầy bất ngờ, thách thức khái niệm gene từng phần. Một vài trong số những khám phá này như gene tái cấu trúc, promoter khác thường, những giai đoạn khác nhau của sự cắt nối khác thường bao gồm cả những exon bên cạnh intron, gene chồng lấp (nested genes), trans-splicing mRNA, sự sắp xếp RNA (RNA editing) và cắt nối protein (protein splicing) đã nhấn mạnh hơn nữa tính lưu động của cấu trúc gene eukaryote vượt qua khỏi mô hình exon-intron. Quan điểm truyền thống về sự tương ứng một đối một giữa gene, mRNA và trình tự polypeptide đã không còn là một chủ đề phổ biến nữa; nó có thể thích hợp với nhiều gen eukaryote nhưng không phải tất cả.

Mặc dù có nhiều ngoại lệ đối với quan hệ giữa trình tự gene-mRNA-polypeptide, mô hình exon-intron vẫn là mô hình chủ yếu trong việc tìm hiểu cấu trúc phân tử và chức năng của những gene eukaryote. Bài tiểu luận này sẽ thảo luận về cấu trúc của gene eukaryote điển hình và sự biểu hiện của chúng.

2. CẤU TRÚC GENE

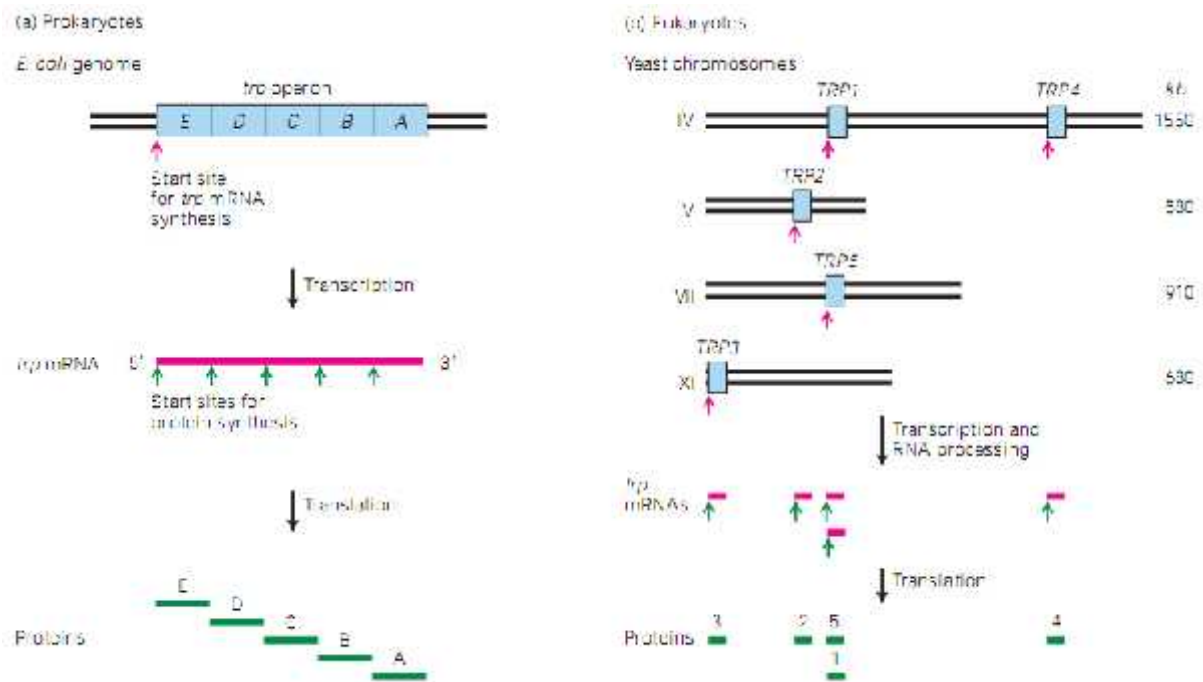
Có thể định nghĩa một gene là toàn bộ trình tự nucleic acid cần cho sự tổng hợp một phân tử sản phẩm có chức năng (polypeptide hay RNA). Dựa vào định nghĩa này, một gen bao gồm nhiều hơn những đoạn nucleotide mã hóa cho trình tự acid amin của một protein. Một gen bao gồm cả những trình tự DNA cần cho việc tổng hợp một phân tử RNA. Ở những gen eukaryote, những vùng điều khiển phiên mã được gọi là enhancer có thể nằm ở vị trí cách vùng mã hóa 50 kb hoặc hơn. Những vùng không mã hóa quan trọng khác ở eukaryote là những trình tự đánh dấu sự cắt ở đầu 3' và sự polyadenyl hóa, gọi là vùng poly (A), và đánh dấu sự cắt nối (splicing) những đoạn RNA phiên mã, được gọi là vùng cắt nối. Đột biến trên những tín hiệu chế biến RNA này làm ngăn chặn sự biểu hiện của một mRNA chức năng và do đó ngăn chặn biểu hiện ra polypeptide. Mặc dù hầu hết các gene đều được phiên mã ra mRNA mã hóa cho protein, tuy nhiên rõ ràng là một số trình tự DNA được phiên mã thành những RNA không mã hóa cho protein (ví dụ, tRNA và rRNA). Tuy nhiên, vì những DNA mã hóa cho tRNA và rRNA có thể gây ra những kiểu hình đặc biệt khi nó bị đột biến, nên những vùng DNA này thường được quy là gene tRNA và rRNA, mặc dù sản phẩm cuối cùng của chúng không phải là protein. Ngoài ra còn có nhiều loại RNA khác cũng được phiên mã từ gene không mã hóa protein.

2.1. Monocistron – Polycistron

Những phân tử mRNA ở vi khuẩn đều là polycistron, nghĩa là một phân tử mRNA (ví dụ mRNA mã hóa từ operon Trp) chứa vùng mã hóa cho một vài protein hoạt động với nhau trong cùng một quá trình sinh học. Ngược lại, hầu hết mRNA ở

eukaryote đều là monocistron, nghĩa là mỗi phân tử mRNA chỉ mã hóa cho một protein duy nhất. Sự khác nhau giữa các mRNA monocistron và polycistron này tương ứng với sự khác nhau cơ bản trong quá trình dịch mã của chúng (sẽ được nói rõ ở các phần sau).

Trong một phân tử mRNA polycistron ở vi khuẩn, vùng gắn ribosome nằm ở gần vị trí bắt đầu vùng mã hóa protein, hay những cistron trong mRNA. Sự khởi sự dịch mã có thể bắt đầu tại bất cứ vị trí nào trong những vị trí này, sản xuất ra nhiều loại protein khác nhau (hình 1a). Tuy nhiên đối với hầu hết những mRNA eukaryote, cấu trúc đầu mũ chụp 5' chỉ thị cho sự gắn vào của ribosome và dịch mã bắt đầu tại vị trí codon AUG gần nhất (hình 1b).



Hình 1: So sánh cấu trúc gene, sự phiên mã và dịch mã ở prokaryote và eukaryote

(a) Operon tryptophan (*trp*) ở *E. coli* gồm 5 gene (màu xanh dương) mã hóa cho những enzyme cần thiết cho sự tổng hợp tryptophan. Toàn bộ operon được phiên mã từ một promoter thành một phân tử mRNA dài và liên tục (màu đỏ). Sự dịch mã mRNA này bắt đầu tại 5 vùng khác nhau, tạo ra 5 protein khác nhau (màu xanh lá). Trật tự của các gene này trong nhiễm sắc thể của vi khuẩn song song tương ứng với thứ tự chức năng liên tiếp của các protein được mã hóa trong con đường tổng hợp tryptophan.

(b) 5 gene mã hóa cho các enzyme cần cho quá trình tổng hợp tryptophan ở nấm men (*Saccharomyces cerevisiae*) nằm trên 4 nhiễm sắc thể khác nhau (IV, V, VII và XI). Mỗi gene được phiên mã từ promoter của chính nó tạo ra đoạn phiên mã sơ cấp được chế biến thành phân tử mRNA chức năng mã hóa cho ra một protein đơn lẻ. Chiều dài của các nhiễm sắc thể trên được thể hiện ở bên phải và tính bằng kilobase (10^3 base).

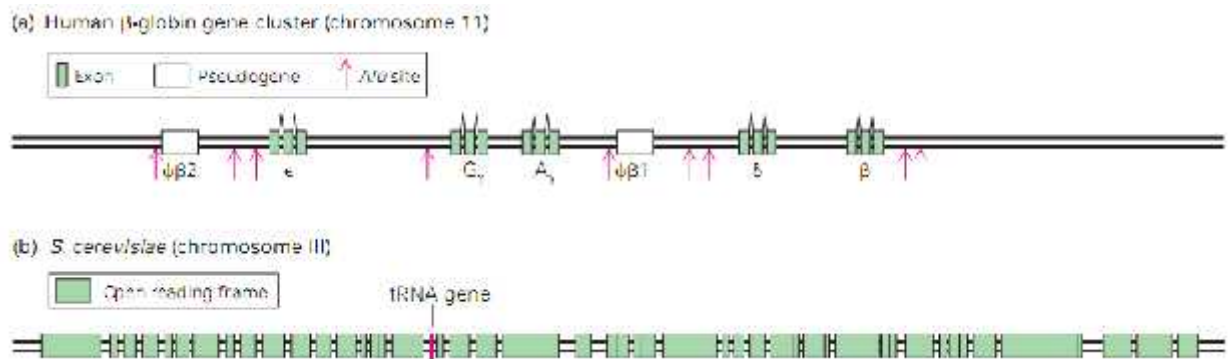
2.2. Intron – Exon

Khác với những gene không chứa intron ở vi khuẩn và nấm men, hầu hết những gene ở động vật và thực vật đa bào đều chứa những đoạn intron được loại bỏ trong suốt quá trình chế biến mRNA. Trong nhiều trường hợp, những đoạn intron trong một gene dài hơn đáng kể so với exon. Ví dụ, trong một gene mã hóa cho một protein có kích thước trung bình chứa khoảng 50.000 cặp base, hơn 95% là intron và những vùng không mã hóa ở đầu 5' và 3'. Nhiều phân tử protein lớn ở những sinh vật bậc cao có những domain lặp lại, được mã hóa bởi những gen bao gồm những đoạn exon tương tự lặp đi lặp lại và xen kẽ bởi những đoạn intron có chiều dài biến thiên.

2.3. Sự tổ chức các gene và DNA không mã hóa trên nhiễm sắc thể

Bộ gene của nhiều sinh vật chứa rất nhiều DNA không chức năng. So sánh ban đầu trên tổng DNA nhiễm sắc thể trong mỗi tế bào ở nhiều loài gợi ý rằng đa số DNA ở các sinh vật không mã hóa cho RNA hay có bất kỳ một chức năng cấu trúc hay điều hòa nào rõ ràng. Ví dụ ở nấm men, ruồi giấm, gà và người được nhận thấy là có lượng DNA trong bộ nhiễm sắc thể tương ứng với cấp độ phức tạp của chúng (lượng DNA lần lượt là 12; 180; 1300; và 3300 Mb). Tuy nhiên, trong số động vật có xương, những loài có lượng DNA trong mỗi tế bào lớn nhất lại là lưỡng cư - những loài chắc chắn là ít phức tạp hơn so với người cả về cấu trúc lẫn hành vi. Rất nhiều loài thực vật cũng có số lượng DNA nhiều hơn đáng kể so với người. Ví dụ như tulip có số lượng DNA gấp 10 lần so với người. Ngoài ra, lượng DNA trong mỗi tế bào cũng biến thiên đáng kể giữa những loài có quan hệ gần gũi với nhau.

Việc giải trình tự và xác định chi tiết các đoạn exon trên DNA nhiễm sắc thể đã chứng minh rằng bộ gen của những sinh vật eukaryote bậc cao chứa lượng lớn DNA không mã hóa. Ví dụ như, chỉ một phần nhỏ (khoảng 80 kb) trên cụm gene β -globin ở người là mã hóa cho protein (hình 2). Hơn nữa, so với những vùng DNA khác ở động vật có xương, cụm gene β -globin thường giàu những trình tự mã hóa cho protein, và những đoạn intron trên gene globin ngắn hơn đáng kể so với những đoạn intron trên nhiều gene khác ở người. Ngược lại, một đoạn DNA 80 kb ở nấm men *S. cerevisiae* chứa nhiều trình tự mã hóa gần nhau, rất ít intron và DNA không mã hóa. Mật độ gene biến thiên rất lớn giữa những vùng DNA nhiễm sắc thể người khác nhau, từ những vùng “giàu gene” như cụm gene β -globin cho đến những vùng “sa mạc” nghèo gene. Trong số 94% DNA bộ gene người được giải trình tự, chỉ khoảng 1,5% là tương ứng với các trình tự mã hóa protein (các đoạn exon). Hầu hết các đoạn exon ở người chứa từ 50-200 cặp base mặc dù đoạn exon ở đầu 3' ở nhiều đơn vị phiên mã dài hơn nhiều. Độ dài những đoạn intron ở người cũng biến thiên đáng kể: nhiều đoạn dài khoảng 90 cặp base, một số khác thường dài hơn rất nhiều, trung bình 3,3 kb. Xấp xỉ 1/3 DNA bộ gene người được cho là được phiên mã thành những đoạn tiền-mRNA, nhưng khoảng 95% những trình tự này đều là intron, được loại bỏ trong quá trình cắt nối RNA.



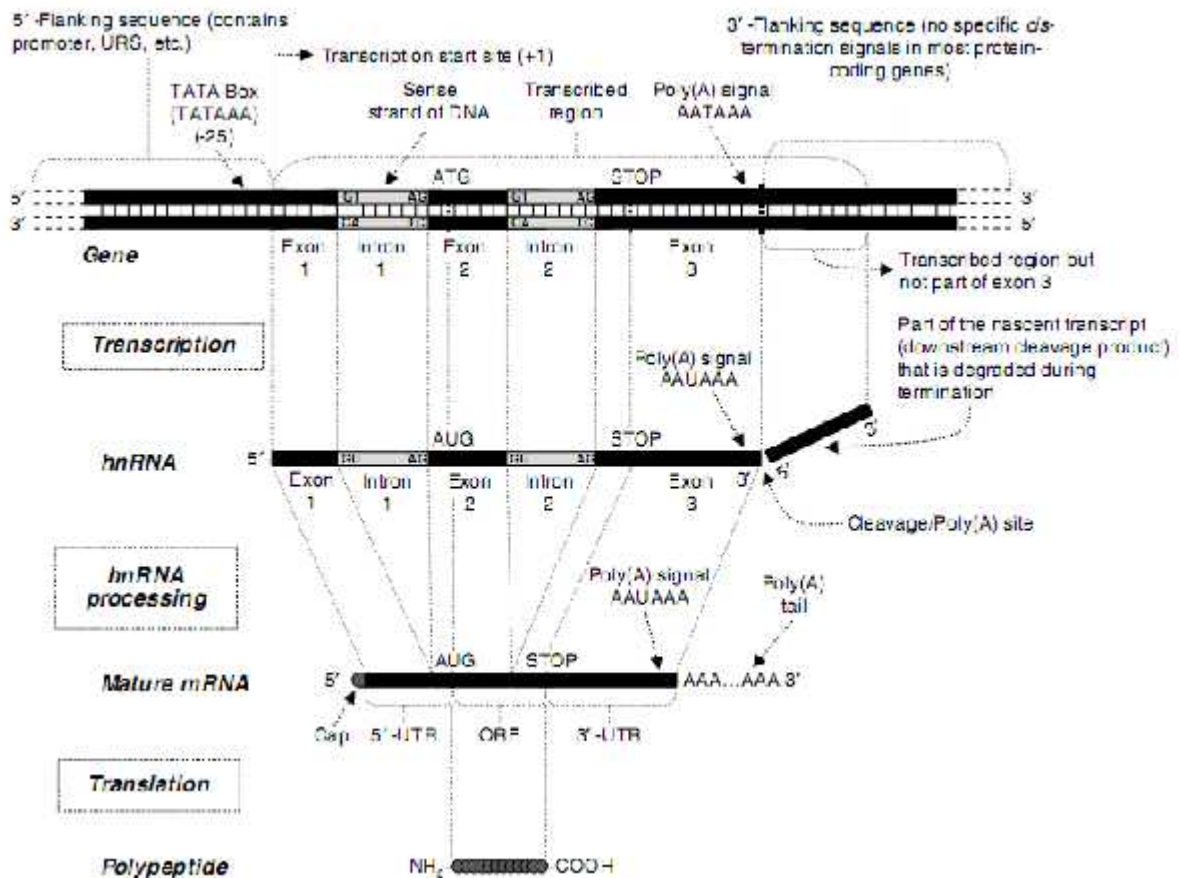
Hình 2: So sánh mật độ gene trên một vùng ≈ 80 kb ở người và nấm men

[Phần (a), xem F. S. Collins và S. M. Weissman, 1984, *Prog. Nucl. Acid Res Mol. Biol.* **31**:315; phần (b), xem S.G. Oliver và cs, 1992, *Nature* **357**:28.]

Sự khác nhau đáng kể về lượng DNA không chức năng ở những sinh vật đơn bào và đa bào có thể là do những áp lực chọn lọc khác nhau trong suốt quá trình tiến hóa. Ví dụ như, những vi sinh vật cần phải cạnh tranh lượng chất dinh dưỡng có giới hạn trong môi trường, do đó việc kiểm soát trao đổi chất là một đặc điểm then chốt. Vì sự tổng hợp DNA không chức năng cần có nhiều thời gian và năng lượng cho nên có lẽ đã có áp lực chọn lọc để loại bỏ DNA không chức năng trong suốt quá trình tiến hóa của vi sinh vật. Mặt khác, chọn lọc tự nhiên ở những loài động vật có xương phần lớn dựa vào hành vi của chúng. Năng lượng đầu tư vào việc tổng hợp DNA là không đáng kể so với năng lượng trao đổi chất cần cho sự vận động của các bắp cơ; do đó có rất ít áp lực chọn lọc để loại bỏ DNA không chức năng này.

2.4. Cấu trúc gene mã hóa protein điển hình ở eukaryote

Cấu trúc của một gene mã hóa protein điển hình ở eukaryote được minh họa ở hình dưới đây. Có 3 phần chính: một vùng cánh 5' (5'-flank) ở đầu 5' của gene; một vùng được phiên mã ở giữa; và một vùng cánh 3' (3' -flank) ở đầu 3' của gene. Promoter nằm ở đầu 5' của gene.



Hình 3: Cấu trúc một gene eukaryote điển hình

Trên hình thể hiện mạch sense, mạch khuôn, TATA box, vị trí mũ chup (vị trí bắt đầu phiên mã), tín hiệu poly(A) (AATAAA ở DNA và AAUAAA ở RNA), vị trí cho và nhận sự ghép nối trên intron (GT...AG ở DNA; GU...AG ở RNA), cũng như quá trình chế biến tiền-mRNA. Trên hình cho thấy đoạn exon 1 và một phần nhỏ ở đầu 5' của đoạn exon 2 là những đoạn không mã hóa, do đó chúng cấu tạo nên 5'-UTR.

2.4.1. Vùng được phiên mã

Vùng phiên mã của một gene gồm 3 phần: một vùng 5' không được dịch mã (5'-UTR – 5' -untranslated region), một vùng mã hóa amino acid (còn được gọi là khung đọc mở hay ORF), và vùng 3' không được dịch mã (3'-UTR).

Đối với bất kỳ một gene nào, chỉ một trong hai mạch của DNA là được phiên mã. Mạch được phiên mã được gọi là mạch khuôn (template) hay mạch antisense. Mạch không được phiên mã kia được gọi là mạch sense vì hai lý do: đầu tiên, trình tự của những base trong mạch không phiên mã tương tự như trình tự base ở mRNA (ngoại trừ T ở DNA thay vì U ở RNA) cho nên trình tự những codon ở mRNA được phản ánh ở trình tự base của mạch không phiên mã; thứ hai, tính phân cực 5' → 3' ở mạch không phiên mã cũng tương tự như mRNA. Tất cả những gene nằm trên cùng một DNA nhiễm sắc thể có thể sẽ không được phiên mã từ cùng một mạch DNA. Đối với một vài gene, một mạch có thể là mạch khuôn, trong khi đối với những gene khác, mạch kia có thể là mạch khuôn. Do sự phiên mã luôn diễn ra theo chiều

5'→3', và do mạch DNA khuôn và RNA được tổng hợp từ nó là đối song song, nên vị trí của promoter tự động xác định mạch nào của DNA có thể được dùng làm mạch khuôn cho sự phiên mã.

Sự biểu hiện đầu 5'- và 3'-UTR có liên quan đến cả mRNA và gene. Vùng 5'-UTR của một gene (và mRNA) là toàn bộ trình tự từ vị trí bắt đầu phiên mã (vùng mũ chụp) cho đến nucleotide trước codon khởi sự dịch mã (ATG ở mạch không phiên mã – sense strand của DNA; AUG ở mRNA). Tương tự, vùng 3'-UTR của một gene (và mRNA) là toàn bộ trình tự bắt đầu sau codon kết thúc dịch mã (TAG/TGA/TAA ở mạch không phiên mã của DNA; UAG/UGA/UAA ở mRNA) cho đến nucleotide trước đuôi poly(A) (hình 3). Do đó, hai vùng 5'- và 3'-UTR của một gene bao gồm tất cả những exon không mã hóa, những phần không mã hóa của exon và thính thoảng gồm cả intron.

2.4.2. Vùng cánh 5' của gene

2.4.2.1. Promoter

Mô hình operon được đề xuất bởi Jacob và Monod đã đưa ra khái niệm promoter như là một phần không thể thiếu của gene hay đơn vị phiên mã, là nơi mà RNA polymerase gắn vào. Với những tiến bộ về kỹ thuật tạo dòng phân tử, người ta đã phân tích và xác định được rất nhiều các trình tự promoter thông qua phân tích xóa bỏ (deletion analysis). Đặc điểm quan trọng nhất của trình tự promoter đó là chúng điều khiển sự khởi sự phiên mã của một gene đặc hiệu, và vị trí của chúng được cố định tương đối so với vị trí bắt đầu phiên mã.

Do sự phiên mã được tiến hành theo chiều 5'→3', và RNA vừa mới tổng hợp có định hướng đối song song với mạch khuôn DNA nên vị trí của promoter sẽ tự động quyết định mạch nào trong hai mạch DNA của gene đó sẽ được phiên mã. Có rất nhiều vùng promoter được gọi là “promoter trung tâm” (core/basal promoter), “promoter lân cận” (proximal promoter) và “promoter ngoại biên” (distal promoter) dựa trên chức năng và khoảng cách của chúng so với điểm khởi đầu phiên mã. Đôi lúc những trình tự điều hòa phiên mã này nằm ở vùng thượng nguồn của promoter trung tâm được gọi chung là những “trình tự promoter thượng nguồn” (upstream promoter elements). Ngoài promoter ra còn có những trình tự DNA khác cũng đóng góp trong việc điều hòa sự biểu hiện gene bao gồm enhancer, silencer, vùng điều khiển locus (LCR) và những trình tự cách ly (insulator elements).

Thông thường, vị trí bắt đầu phiên mã được quyết định bởi hộp TATA và trình tự khởi đầu, hoặc đối với những promoter không có hộp TATA, vị trí này được quyết định bởi trình tự khởi đầu và trình tự promoter hạ nguồn, tất cả đều nằm bên trong promoter trung tâm. Vùng promoter trung tâm giúp hình thành phức hợp tiền khởi sự phiên mã (PIC) ở gần vị trí bắt đầu phiên mã. Phức hợp PIC bao gồm RNA polymerase và những nhân tố phiên mã chung (GTFs) – những nhân tố cần cho sự khởi sự phiên mã bởi RNA pol II. Tuy nhiên, hiệu quả và tính đặc hiệu trong việc nhận biết promoter phụ thuộc vào một vài trình tự khác (và những protein tương tác

với chúng) nằm xa hơn về phía thượng nguồn trong promoter lân cận. Vùng promoter lân cận là nơi gắn của một nhóm các nhân tố phiên mã khác – các nhân tố hoạt hóa phiên mã. Những protein hoạt hóa này tương tác với những cấu trúc cơ bản. Một vài các nhân tố hoạt hóa có thể có tính đặc hiệu với mô và được gọi là “những nhân tố phiên mã đặc hiệu” hoặc “nhân tố hoạt hóa đặc hiệu mô”.

a) Promoter trung tâm

Promoter trung tâm là trình tự tiếp giáp, dẫn dắt sự khởi đầu phiên mã chính xác bởi RNA poly II. Nó là vị trí gắn của RNA poly II và các GTF. Thông thường nó dài khoảng 35 cặp base và kéo dài cả về phía thượng nguồn lẫn hạ nguồn của vị trí bắt đầu phiên mã (-35 đến +35). Promoter trung tâm có thể chứa hai hay nhiều hơn trong số các motif sau: hộp TATA, trình tự khởi đầu (Inr) và trình tự promoter hạ nguồn (DPE).

b) Promoter lân cận

Promoter lân cận dài khoảng 250 cặp base và có thể kéo dài cả hai phía của vị trí bắt đầu phiên mã (-250 đến +250). Tuy nhiên, theo tài liệu, những trình tự xa hơn -250 về phía thượng nguồn cũng được gọi là “trình tự promoter lân cận”. Thông thường nó là nơi gắn các nhân tố phiên mã đặc hiệu hoặc nhân tố hoạt hóa. Hai trình tự hoạt hóa phiên mã nằm trong promoter này gồm hộp CAAT và hộp GC. Hộp CAAT gắn nhân tố phiên mã NF-I (nuclear factor I, còn gọi là NF-Y, CTF và CBF), nằm ở vị trí khoảng nucleotide 75 phía thượng nguồn tính từ vị trí bắt đầu phiên mã và có trình tự thỏa hiệp là GG(T/C)CAATCT. Hộp GC có trình tự thỏa hiệp là GGGCGG, nằm ở vị trí nucleotide 90 phía thượng nguồn tính từ vị trí bắt đầu phiên mã, là nơi gắn nhân tố phiên mã Sp1 (specificity protein 1). Hộp CAAT và GC hoạt động giống như các trình tự enhancer vì chúng có thể hoạt hóa sự phiên mã khi được đặt ở một trong hai hướng bên trong promoter lân cận.

c) Promoter ngoại biên

Thuật ngữ “promoter ngoại biên” được dùng để chỉ các trình tự nằm xa hơn về phía thượng nguồn của promoter trung tâm. Có rất nhiều ví dụ về sự kết hợp giữa promoter lân cận và promoter ngoại biên trong việc điều hòa phiên mã. Cả hai promoter của gene đều thể hiện sự đặc hiệu đối với mô và chứa nhiều các trình tự điều hòa phiên mã đặc biệt, được nhận biết bởi các nhân tố phiên mã đặc hiệu mô.

Ngoài ra, thỉnh thoảng những trình tự bên trong intron cũng đóng vai trò không thể thiếu trong việc điều hòa tăng cường phiên mã được gọi là những trình tự giống promoter bên trong intron (Salguero S. và cs, 2000). Nhóm tác giả cho rằng một vài intron cũng có đặc điểm tương tự promoter, như trình tự giống TATA, hộp CAAT.

2.4.2.2. Các trình tự điều hòa khác (Enhancer, Silencer, LCR, Insulator)

Rất nhiều các trình tự điều hòa phiên mã có thể nằm ở vị trí cách xa gene khoảng một vài kb cả về hai phía tính từ vị trí bắt đầu phiên mã. Một vài trong số những trình tự này kích hoạt sự phiên mã như enhancer và LCR, trong khi một số khác hoạt động như những tác nhân ức chế phiên mã như silencer. Những vùng này

chứa những trình tự nhận biết nhiều loại protein gắn DNA đặc hiệu với trình tự có liên quan đến sự điều hòa phiên mã.

a) Enhancer và Silencer

Enhancer có thể giúp tăng cường tốc độ phiên mã bằng cách tăng cường vận dụng promoter. Chúng là vị trí gắn của các nhân tố hoạt hóa phiên mã đặc hiệu. Một đoạn enhancer có thể điều khiển sự phiên mã của nhiều hơn một gene theo kiểu không phụ thuộc vào vị trí và chiều hướng. Những trình tự enhancer có thể nằm ở vị trí gần với vị trí bắt đầu phiên mã, phía thượng nguồn lẫn hạ nguồn và thậm chí còn có thể nằm bên trong intron. Hộp CAAT và GC đề cập ở trên là những trình tự enhancer đóng vai trò không thể thiếu trong promoter lân cận ở hầu hết các gene. Người ta cho rằng enhancer đem những nhân tố hoạt hóa gắn vào nó tương tác với những nhân tố phiên mã gắn với promoter bằng cách uốn cong DNA. Từ đó, chúng làm tăng nồng độ các nhân tố hoạt hóa gần promoter và những nhân tố này tương tác trực tiếp hoặc gián tiếp với promoter để khởi sự quá trình phiên mã.

Ngược lại với enhancer là silencer, chúng ức chế sự phiên mã bằng cách gắn với những nhân tố ức chế phiên mã, từ đó hoạt động như là yếu tố điều hòa âm. Silencer có thể hoạt động không phụ thuộc vào chiều, vị trí và khoảng cách, và chúng cũng có thể nằm bên trong intron.

b) Vùng điều khiển locus (LCR)

LCR (locus control region) là một loại trình tự tăng cường phiên mã khác. LCR tăng cường sự phiên mã của một cụm các gene liên kết với nhau bằng cách tạo ra một hình dáng nhiễm sắc thể mở bên cạnh locus. Từ đó, LCR có thể tác động mạnh đến mức độ hoạt động của một vùng nhiễm sắc chất điển hình (euchromatic) của một nhiễm sắc thể. LCR được xác định đầu tiên ở locus β -globin ở người.

c) Insulator

Insulator (trình tự cách ly) là một trình tự ranh giới gene quan trọng. Khi gắn vào những protein gắn insulator, chúng sẽ bảo vệ promoter khỏi những tác động của các yếu tố điều hòa gần đó. Insulator có hai dạng chức năng: chức năng ngăn chặn enhancer và chức năng hàng rào dị nhiễm sắc chất (heterochromatin).

Chức năng ngăn chặn enhancer liên quan đến việc ngăn chặn sự tương tác giữa một enhancer và promoter khi insulator nằm ở vị trí giữa hai trình tự này, và từ đó ngăn chặn sự hoạt hóa của enhancer đối với promoter. Vì một enhancer có thể tác động đến nhiều hơn một promoter nên chức năng ức chế có thể ngăn chặn tác động bừa bãi của một enhancer lên nhiều promoter và bắt buộc chúng chỉ tác động lên một promoter đặc hiệu.

Chức năng hàng rào dị nhiễm sắc chất (heterochromatin barrier) liên quan đến việc bảo vệ một vùng nhiễm sắc chất điển hình (bao gồm những gene với những promoter hoạt động) khỏi việc trở thành dị nhiễm sắc chất bằng cách bất hoạt tác động xâm lấn của dị nhiễm sắc chất gần kề. Một số insulator sở hữu cả hai chức năng ngăn chặn và hàng rào, trong khi một số khác chỉ có một chức năng.

2.4.3. Vùng cánh 3' của gene

Vùng cánh 3' của gene kéo dài xa hơn vùng 3'-UTR và chứa những tín hiệu kết thúc phiên mã.

Ở vi khuẩn (prokaryote) có hai kiểu kết thúc phiên mã: phụ thuộc rho và không phụ thuộc vào rho. Kết thúc phiên mã không phụ thuộc rho cần có hai cấu trúc đặc trưng cần cho cả hai kiểu kết thúc: (i) một trình tự có thể tạo ra cấu trúc thông lọng (stem-loop) và (ii) một vùng giàu U ở phía cuối của cấu trúc thông lọng. Cấu trúc thông lọng thường chứa một vùng giàu GC nằm cách vùng giàu U khoảng 10 base. Sự hình thành cấu trúc thông lọng này làm cho RNA poly ngừng lại, từ đó kết thúc sự phiên mã.

Kết thúc phiên mã phụ thuộc vào rho được phát hiện ở khoảng phân nửa các gene của E. coli. Rho là một nhân tố kết thúc phiên mã gắn vào RNA, là một protein có hoạt tính helicase và ATPase phụ thuộc RNA. Vùng kết thúc phiên mã phụ thuộc vào rho thường là giàu C và nghèo G. Trong suốt quá trình phiên mã, nhân tố rho gắn lên trên RNA đang được kéo dài ở vị trí dài khoảng 75 nucleotide và phía thượng nguồn của vùng kết thúc phiên mã. Sử dụng hoạt tính ATPase của nó, rho di chuyển dọc theo RNA và có lẽ với vận tốc nhanh hơn RNA pol di chuyển trên mạch khuôn. Khi RNA pol ngừng tại vị trí kết thúc, rho bắt kịp nó và sử dụng hoạt tính helicase để tách đoạn mạch lai giữa DNA-RNA, từ đó kết thúc sự phiên mã và giải phóng RNA và RNA pol.

Ở Eukaryote người ta vẫn chưa biết nhiều về sự kết thúc phiên mã. Tuy nhiên người ta có thể khái quát nó dựa trên những hiểu biết hiện tại. Mỗi loại RNA pol sử dụng một cơ chế kết thúc phiên mã khác nhau. Đối với pol I và III, việc dừng lại ở trình tự kết thúc ở vùng cánh 3' có vẻ đóng vai trò quan trọng trong việc kết thúc phiên mã và giải phóng RNA và pol. Những nhân tố giúp dừng phiên mã có thể không nhất thiết là những nhân tố giải phóng. Những tín hiệu kết thúc cũng như những nhân tố kết thúc và giải phóng có thể rất khác nhau ở những loài eukaryote khác nhau.

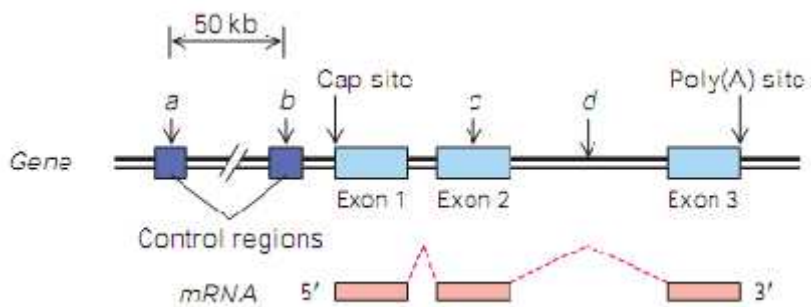
3. SỰ BIỂU HIỆN GENE

3.1. Đơn vị phiên mã

Ở vi khuẩn, một đơn vị phiên mã bao gồm một cụm các gen tạo nên một operon, được phiên mã từ một promoter xác định thành một đoạn phiên mã duy nhất. Nói cách khác, ta có thể phân biệt được gene và đơn vị phiên mã ở prokaryote. Ngược lại, ở eukaryote hầu hết các gene và đơn vị phiên mã là như nhau, và hai thuật ngữ này thường có thể được sử dụng thay thế cho nhau. Tùy thuộc vào số phận của đoạn phiên mã sơ cấp, đơn vị phiên mã ở eukaryote được phân thành 2 dạng: đơn vị phiên mã đơn giản và đơn vị phiên mã phức tạp.

Một đơn vị phiên mã đơn giản (simple transcription unit) tạo ra một đoạn phiên mã sơ cấp, đoạn phiên mã này sẽ được chế biến để tạo ra một dạng mRNA duy nhất, mã hóa cho một protein đơn lẻ. Tất cả đột biến trên các đoạn exon, intron

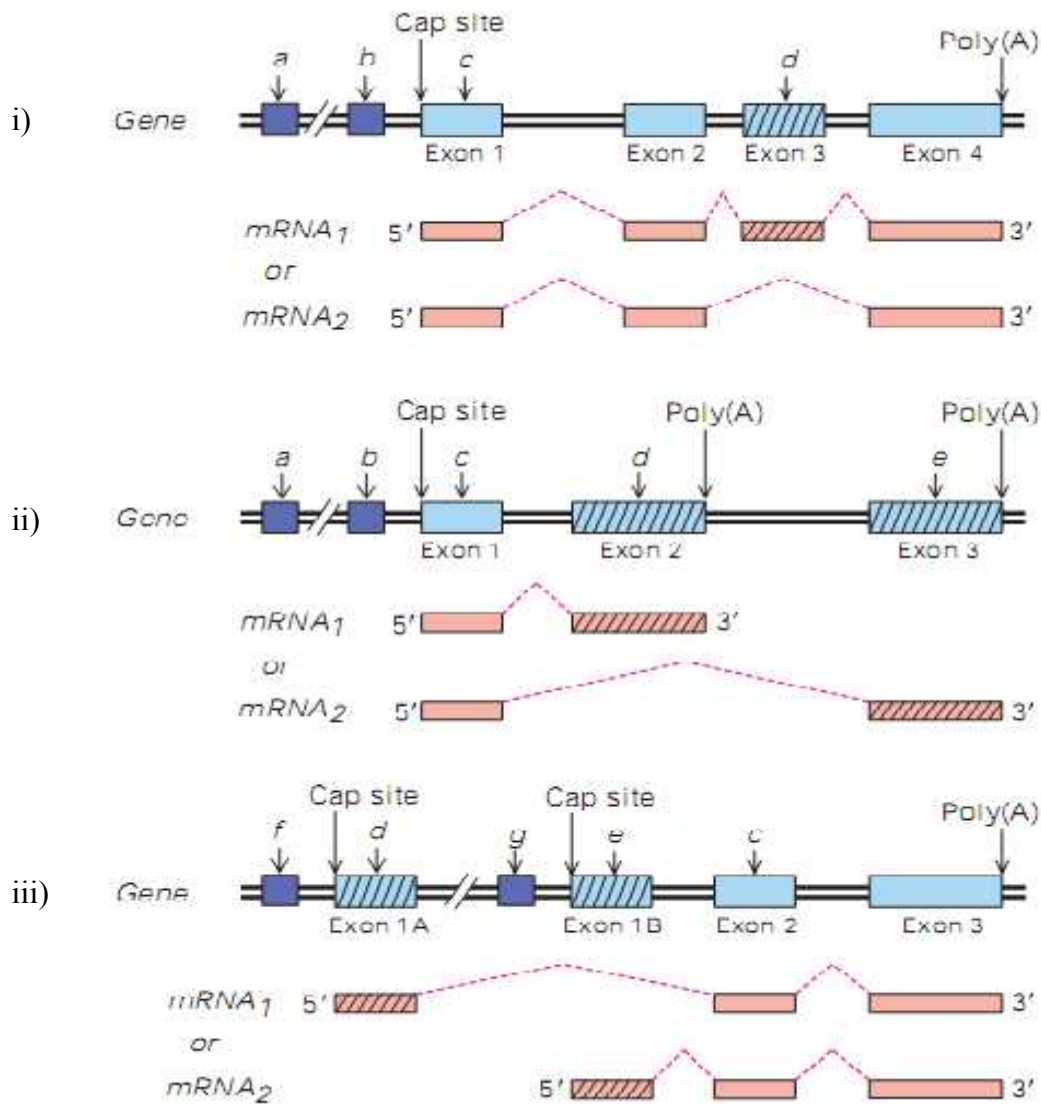
và những vùng điều hòa phiên mã đều có thể ảnh hưởng đến sự biểu hiện của protein được mã hóa bởi một đơn vị phiên mã đơn giản (hình 4).



Hình 4: Đơn vị phiên mã đơn (simple transcription unit)

Một đơn vị phiên mã đơn giản gồm một vùng mã hóa cho một protein, kéo dài từ mũ chụp đầu 5' cho đến vùng poly(A) ở đầu 3' và các vùng điều hòa có liên quan. Các đoạn intron nằm xen kẽ giữa các đoạn exon (hình chữ nhật màu xanh) và được loại bỏ trong suốt quá trình chế biến các tiền-mRNA và do đó không hiện diện trên phân tử mRNA chức năng đơn cistron (monocistron). Những đột biến trên vùng điều hòa phiên mã (đột biến a, b) có thể làm giảm hay ngăn chặn sự phiên mã, từ đó làm giảm hay xóa bỏ sự tổng hợp protein được mã hóa. Một đột biến bên trong đoạn exon (đột biến c) có thể tạo ra một protein bất thường. Một đột biến bên trong đoạn intron (đột biến d) cho ra vị trí cắt nối mới, kết quả tạo ra một mRNA bị cắt nối không bình thường, mã hóa cho ra một protein mất chức năng.

Trong khi đó, đơn vị phiên mã phức (complex transcription unit) khá phổ biến ở sinh vật đa bào và đoạn phiên mã RNA sơ cấp có thể được chế biến theo nhiều cách, dẫn đến hình thành các mRNA chứa những đoạn exon khác nhau. Mỗi mRNA đều là monocistron và được dịch mã thành một polypeptide duy nhất, với sự dịch mã khởi sự ở AUG đầu tiên trên mRNA. Trong hình 6, có nhiều loại mRNA khác nhau có thể được tạo thành theo ba cách: (i) Sử dụng những vị trí nối khác nhau tạo ra những mRNA có cùng những exon ở đầu 5' và 3' nhưng khác nhau ở những exon bên trong; (ii) Sử dụng những vị trí poly(A) khác nhau tạo ra những mRNA có cùng những exon đầu 5' nhưng khác ở những exon đầu 3'; (iii) Sử dụng những promoter khác nhau tạo ra những mRNA khác nhau ở những exon đầu 5' nhưng giống nhau ở những exon đầu 3'. Một gene được biểu hiện một cách có chọn lọc ở hai hay nhiều loại tế bào khác nhau thì thường được phiên mã từ những promoter khác nhau đặc hiệu với loại tế bào.

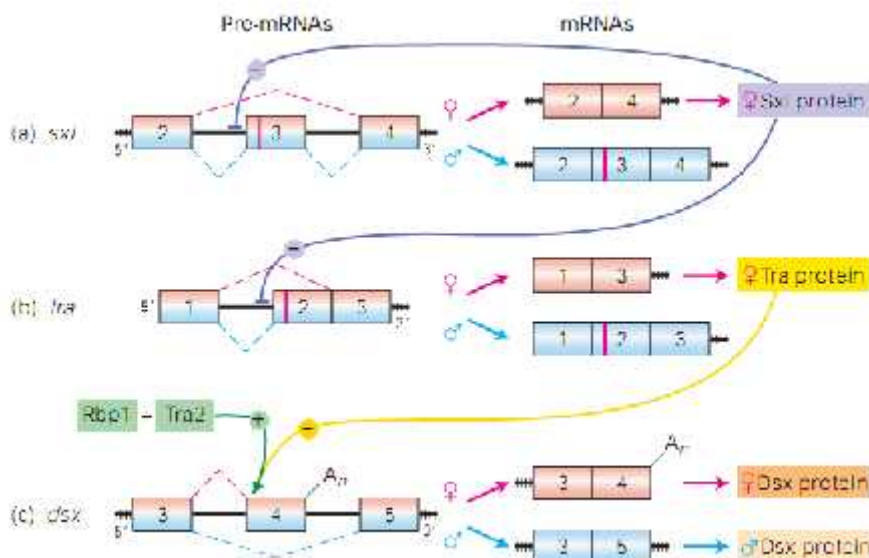


Hình 5: Đơn vị phiên mã phức (complex transcription unit)

Đơn vị phiên mã phức tạo ra những đoạn phiên mã sơ cấp có thể được chế biến theo nhiều cách khác nhau: (i) Nếu một đoạn phiên mã sơ cấp có chứa những vị trí cắt nối khác nhau, nó có thể được chế biến thành những mRNA với cùng những exon ở đầu 5' và 3' nhưng khác nhau ở những exon bên trong; (ii) Nếu một đoạn phiên mã sơ cấp có hai vị trí poly(A), nó có thể được chế biến thành các mRNA với những exon đầu 3' khác nhau; (iii) Nếu các promoter khác nhau (f hoặc g) được kích hoạt ở những dạng tế bào khác nhau, thì mRNA₁ - được sản xuất ở một dạng tế bào mà trong đó promoter f được kích hoạt - sẽ có một exon (1A) khác với mRNA₂ - được sản xuất ở một dạng tế bào khác mà trong đó promoter g được kích hoạt và exon 1B được sử dụng thay cho 1A. Những đột biến ở những vùng điều hòa (a và b) và những đột biến bên trong exon (c) giống nhau ở những mRNA khác nhau có ảnh hưởng đến những protein mã hóa bởi cả hai mRNA được chế biến khác nhau đó. Ngược lại, những đột biến bên trong các đoạn exon (d và e) khác nhau ở những mRNA khác nhau chỉ có tác động ảnh hưởng đến protein được mã hóa từ mRNA đó. Đối với những gene được

phiên mã từ những promoter khác nhau ở các dạng tế bào khác nhau, những đột biến ở các vùng điều hòa khác nhau (f và g) chỉ có tác động ảnh hưởng ở dạng tế bào mà trong đó đoạn promoter đó được hoạt hóa.

Hình dưới đây mô tả ba cách chế biến RNA khác nhau xảy ra trong quá trình biệt hóa giới tính ở *Drosophila*. Thông thường một mRNA được tạo ra từ một đơn vị phiên mã phức ở một vài dạng tế bào, và một mRNA khác được tạo ra ở những dạng tế bào khác. Ví dụ như sự khác nhau ở cách ghép nối ở những đoạn phiên mã fibronectin sơ cấp ở tế bào fibroblast và tế bào gan sẽ quyết định liệu protein tiết ra có bao gồm những domain đóng vai trò bám vào bề mặt tế bào hay không.



Hình 6: Chuỗi điều hòa cắt nối điều khiển sự định đoạt giới tính ở phôi *Drosophila*

Chuỗi điều hòa cắt nối điều khiển sự định đoạt giới tính thông qua sự biểu hiện của các gene *sxl* (*sex-lethal*), *tra* (*transformer*) và *dsx* (*double-sex*) ở phôi *Drosophila*. Hình chỉ thể hiện các đoạn exon (hình chữ nhật) và intron (đường gạch đen) - nơi xảy ra sự điều hòa cắt nối. Đường đứt khúc màu đỏ thể hiện sự cắt nối (*splicing*) tiền-mRNA ở con cái và tương tự đường màu xanh thể hiện sự cắt nối ở con đực. Đường dọc màu đỏ bên trong các đoạn exon biểu thị các codon stop bên trong khung đọc - cái ngăn chặn sự tổng hợp protein chức năng. Chỉ những phôi cái mới sản xuất protein Sxl chức năng - protein ức chế việc cắt nối giữa exon 2 và 3 trong tiền-mRNA *sxl* (a) và giữa exon 1 và 2 trong tiền-mRNA *tra* (b). (c) Ngược lại, việc gắn mang tính điều phối của protein Tra và hai protein SR, Rbp1 và Tra2, hoạt hóa việc cắt nối giữa exon 3 và 4 và sự cắt/polyadenyl hóa A_n ở đầu 3' của exon 4 ở tiền-mRNA *dsx* ở phôi cái. Ở phôi đực - phôi thiếu Tra chức năng - protein SR không gắn vào exon 4, và do đó exon 3 được nối vào exon 5. Các protein Dsx khác nhau sản sinh ở phôi đực và cái - kết quả của chuỗi điều hòa cắt nối - có tác dụng ức chế sự phiên mã các gene cần thiết cho sự biệt hóa giới tính của giới tính kia. [Sửa lại từ M. J. Moore và cs, 1993, in R. Gesteland và J. Atkins, eds., *The RNA World*, Cold Spring Harbor Press, trang 303-357.]

Đối với trường hợp đơn vị phiên mã phức, quan hệ giữa một đột biến và một gene không phải lúc nào cũng trung thực hay trực tiếp. Một đột biến ở cùng một vùng điều hòa hay một đoạn exon ở những mRNA khác nhau sẽ ảnh hưởng đến tất cả những protein khác nhau được mã hóa bởi cùng một đơn vị phiên mã phức. Mặt khác, những đột biến trên một exon chỉ hiện diện ở một trong số những mRNA khác nhau thì chỉ ảnh hưởng đến protein được mã hóa bởi mRNA đó.

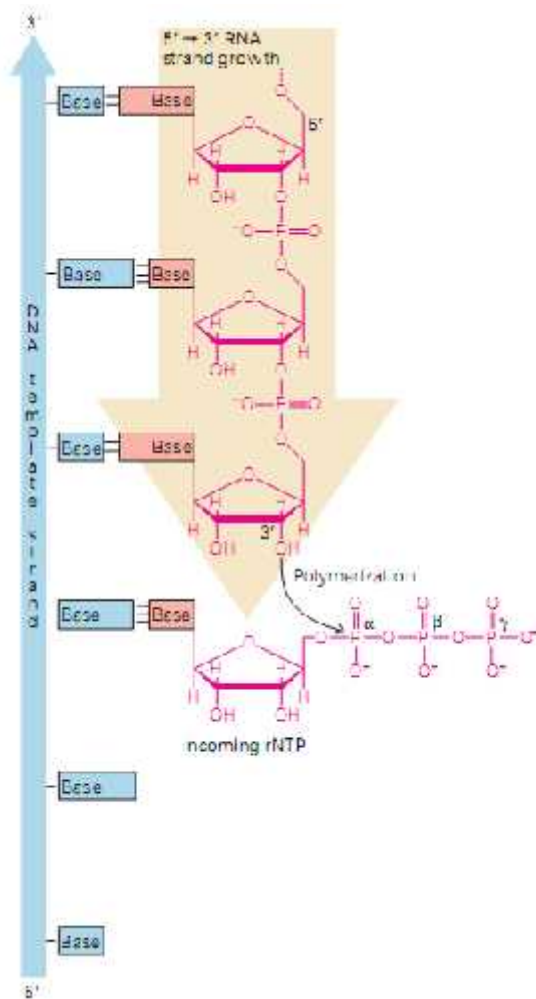
3.2. Sự phiên mã ở gene mã hóa protein và sự hình thành mRNA chức năng

Khái niệm đơn giản nhất của gene đó là một “đơn vị DNA bao gồm thông tin chỉ định sự tổng hợp của một chuỗi polypeptide hay một phân tử RNA chức năng (như tRNA).” Và một phần rất lớn trong gene mang thông tin để tạo nên phân tử protein, và chính những bản sao RNA của những gene mã hóa protein này cấu thành nên những phân tử mRNA của tế bào. Ở virus một phân tử DNA chỉ chứa một vài gene, trong khi ở những động thực vật bậc cao một phân tử DNA ở mỗi nhiễm sắc thể có thể chứa đến vài ngàn gene.

Quá trình tổng hợp RNA (phiên mã) chỉ đơn giản là sự sao chép lại ngôn ngữ gồm bốn base của DNA bao gồm A, G, C và T thành ngôn ngữ gồm bốn base tương tự của RNA, chỉ trừ U thay thế cho T. Mặt khác, quá trình tổng hợp protein là sự phiên dịch lại ngôn ngữ trên thành ngôn ngữ 20 amino acid của protein. Dưới đây là quá trình hình thành mRNA chức năng từ các gene mã hóa protein điển hình.

3.2.1. Sự polymer hóa ribonucleotide

Như đã đề cập ở trên, trong suốt quá trình phiên mã, một mạch của DNA đóng vai trò là mạch khuôn, quyết định thứ tự của các monomer ribonucleoside triphosphate (rNTP) để hình thành một chuỗi RNA bổ sung. Các base trên mạch DNA khuôn bắt cặp với các rNTP bổ sung, cái mà sau đó được gắn lại với nhau trong phản ứng polymer hóa được xúc tác bởi RNA polymerase. Sự polymer hóa liên quan đến sự ăn mòn bởi 3' oxygen trên chuỗi RNA đang dài ra lên α phosphate của nucleotide tiếp theo, hình thành nên liên kết phosphodiester và giải phóng pyrophosphate (Ppi). Theo cơ chế này thì phân tử RNA luôn luôn được tổng hợp từ đầu 5' \rightarrow 3' (hình 7).



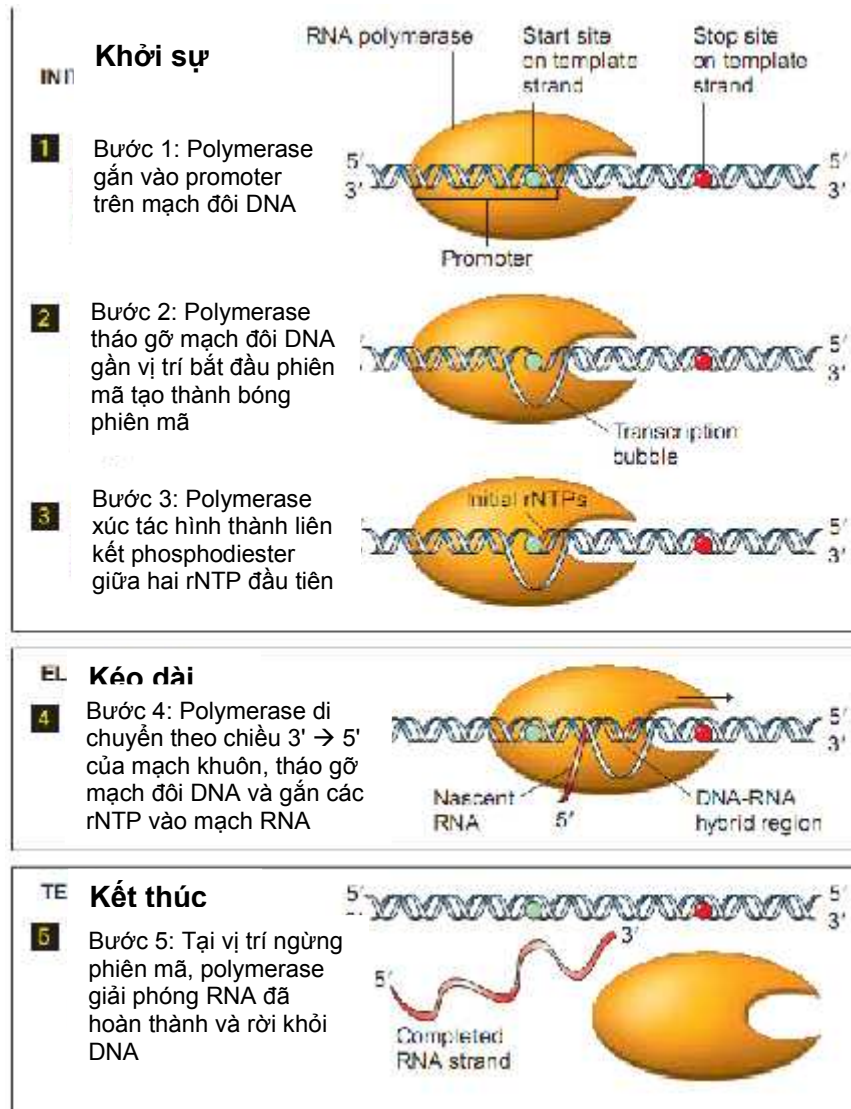
Hình 7: Sự polymer hóa ribonucleotide bởi RNA polymerase trong quá trình phiên mã

Các ribonucleotide sắp được gắn vào đầu 3' của mạch RNA được chỉ định bởi sự bắt cặp giữa base tiếp theo trên mạch DNA khuôn và ribonucleoside triphosphate (rNTP) bổ sung. Một liên kết phosphodiester được hình thành khi RNA polymerase xúc tác một phản ứng giữa 3' O của mạch đang dài ra với α phosphate của một rNTP bắt cặp chính xác. Mạch RNA luôn được tổng hợp theo hướng 5' \rightarrow 3' và ngược lại với chiều phân cực của mạch khuôn DNA.

3.2.1. Các giai đoạn phiên mã ^[1]

Trong suốt quá trình khởi sự phiên mã, RNA polymerase nhận biết và gắn vào một vị trí đặc hiệu – promoter trên DNA mạch đôi (Hình 9-bước 1). Những RNA polymerase nhân cần rất nhiều nhân tố protein khác nhau – các nhân tố phiên mã chung – để giúp chúng định vị promoter và khởi đầu sự phiên mã. Sau khi gắn vào promoter RNA polymerase tháo gỡ hai mạch DNA để các ribonucleotide triphosphate có thể tiếp cận được với các base trên mạch khuôn. Các RNA polymerase trong tế bào tháo gỡ khoảng 14 cặp base xung quanh vị trí bắt đầu phiên mã trên DNA (bước 2). Sự khởi sự phiên mã được cho là hoàn tất khi hai

ribonucleotide của một chuỗi RNA được nối với nhau bằng liên kết phosphodiester (bước 3).



Hình 8: Ba bước của quá trình phiên mã

Trong suốt quá trình khởi sự phiên mã, RNA polymerase tạo nên một bóng phiên mã và bắt đầu sự polymer hóa các ribonucleotide (rNTP) tại vị trí bắt đầu nằm bên trong vùng promoter. Khi một vùng của DNA được phiên mã, mạch còn lại đã tách ra được phục hồi trở lại thành cấu trúc xoắn kép. Đầu 5' của mạch RNA rời khỏi RNA polymerase thông qua một rãnh bên trong enzyme. Sự kết thúc dịch mã xảy ra khi polymerase gặp một trình tự kết thúc đặc hiệu.

Sau khi một vài ribonucleotide đã được gắn vào, RNA pol tách khỏi promoter và các nhân tố phiên mã chung. Trong suốt giai đoạn kéo dài mạch, RNA pol chạy dọc theo mạch DNA khuôn từng base một, đồng thời tháo gỡ mạch đôi DNA ở phía trước nó và phục hồi mạch đôi ở phía nó vừa đi qua (bước 4). Lần lượt từng ribonucleotide một được thêm vào đầu 3' của mạch RNA mới sinh trong giai đoạn kéo dài phiên mã bởi polymerase. Enzyme này vẫn duy trì một vùng tháo gỡ khoảng 14 cặp base, được gọi là bóng phiên mã. Khoảng chừng 8 nucleotide ở đầu

3' của mạch RNA đang tổng hợp vẫn còn bắt cặp base với mạch DNA khuôn trong bóng phiên mã. Phức hợp kéo dài phiên mã bao gồm RNA polymerase, DNA khuôn và mạch RNA đang dài ra. Phức hợp này cực kỳ ổn định.

Trong quá trình kết thúc phiên mã, phân tử RNA đã hoàn thành, hay đoạn phiên mã sơ cấp, được giải phóng khỏi RNA polymerase và polymerase tách khỏi mạch khuôn DNA (bước 5). Những trình tự đặc biệt trên DNA khuôn ra tín hiệu cho RNA polymerase kết thúc phiên mã. Sau khi được giải phóng, RNA pol tự do có thể phiên mã cho cùng một gene một lần nữa hoặc một gene khác.

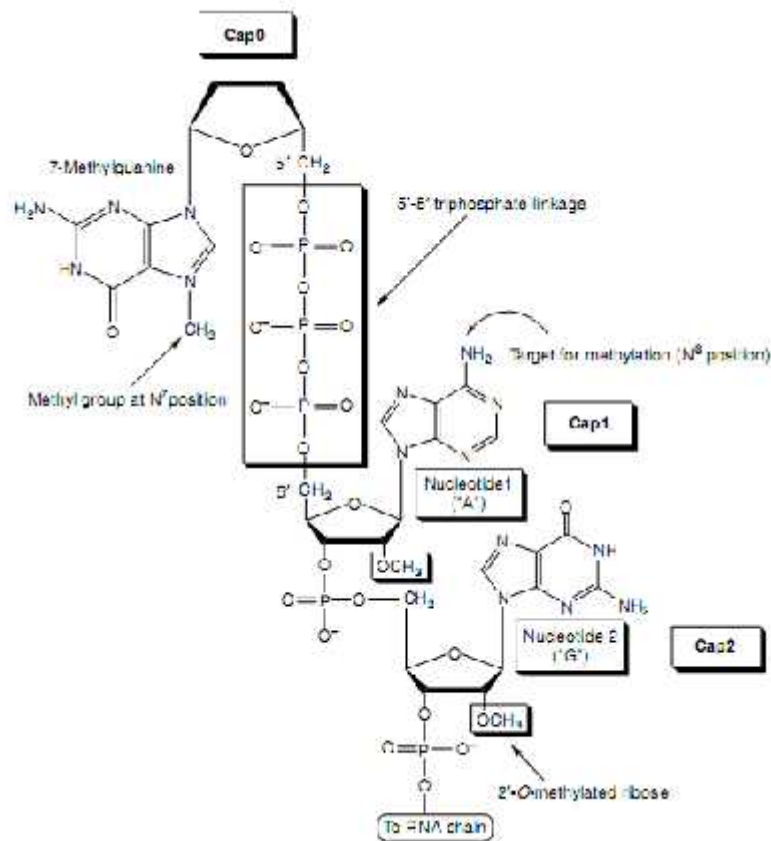
3.3. Biến đổi tiền-mRNA thành mRNA chức năng ở Eukaryote ^{[1] [3]}

Ở các tế bào prokaryote không có nhân, sự dịch mã một phân tử mRNA thành protein có thể bắt đầu từ đầu 5' của mRNA ngay trong lúc đầu 3' vẫn đang được tổng hợp bởi RNA polymerase. Nói cách khác, sự phiên mã và dịch mã có thể xảy ra đồng thời ở prokaryote. Tuy nhiên, ở tế bào eukaryote nhân được tách khỏi tế bào chất – nơi dịch mã xảy ra nên không gian của hai quá trình phiên mã và dịch mã là khác nhau. Hơn nữa các đoạn phiên mã sơ cấp mới chỉ là những tiền-mRNA, cần phải trải qua một vài biến đổi – gọi chung là quá trình chế biến RNA để hình thành nên một phân tử mRNA chức năng (RNA trưởng thành). Phân tử mRNA này sau đó phải được vận chuyển ra tế bào chất trước khi nó có thể được dịch mã thành protein. Vì vậy, sự phiên mã và dịch mã không thể nào xảy ra đồng thời ở những tế bào eukaryote.

3.3.1. Gắn mũ chụp 5' ^{[1] [3]}

Đầu tiên, tất cả các tiền-mRNA ở eukaryote đều được biến đổi ở hai đầu của nó, và những biến đổi này được giữ lại trên mRNA trưởng thành. Tại đầu 5' của chuỗi RNA vừa mới được tách khỏi RNA pol II, ngay lập tức nó được tác động bởi một vài các enzyme kết hợp tổng hợp nên mũ chụp 5', một 7-methylguanylate liên kết với nucleotide cuối cùng của RNA bằng một liên kết 5', 5' triphosphate.

Mũ chụp được viết là m^7Gppp , viết tắt là m^7G hay m^7G . Cấu trúc mũ chụp m^7G kinh điển được gọi là “cap0”, hiện diện ở tất cả các mRNA eukaryote. Cap0 được thêm vào theo một phản ứng gồm ba bước: đầu tiên, đầu phosphate tận cùng (γ -phosphate) được thủy phân bởi RNA 5' triphosphatase từ triphosphate của nucleotide đầu tiên của tiền-mRNA thành diphosphate; tiếp theo, RNA guanylyltransferase xúc tác sự kết hợp giữa một GMP và diphosphate của nucleotide đầu tiên này thông qua liên kết 5'-5', nhờ đó khôi phục lại triphosphate ở tận cùng; cuối cùng, RNA (guanine-7)-methyltransferase xúc tác sự methyl hóa vị trí N⁷ của guanine. Ở động vật có vú, hoạt động của triphosphatase nằm ở đầu N và hoạt động của guanylyltransferase nằm ở đầu C của cùng một polypeptide, nhưng ở nấm men, chúng được xúc tác bởi những enzyme riêng biệt.



Hình 9: Những cấu trúc mũ chụp trên mRNA ở eukaryote

Ở eukaryote bậc thấp, quá trình gắn mũ chụp chỉ tạo ra cap0. Ở eukaryote bậc cao, quá trình gắn mũ chụp có thể tạo ra cap1 và cap2 bằng những biến đổi như methyl hóa 2'-O trên ribose của nucleotide 1 và 2 của tiền-mRNA, sự methyl hóa N⁶ của base đầu tiên trên tiền-mRNA nếu như base đầu tiên là adenine.

Những enzyme gắn mũ chụp được đưa đến tiền-mRNA bằng cách gắn vào domain đầu C (CTD) được phosphoryl hóa của RNA polymerase II (RNA pol II). Đuôi CTD chứa nhiều đoạn gồm 7 peptide lặp đi lặp lại có tính bảo tồn cao với trình tự thỏa hiệp (consensus sequence) là Tyr-Ser-Pro-Thr-Ser-Pro-Ser (YSPTSPS). Số lượng lặp lại biến thiên từ 26 ở nấm men đến 52 ở động vật có vú. Năm trong số bảy amino acid trong motif thỏa hiệp này là những điểm tiếp nhận phosphate, và sự phosphoryl hóa là một sự biến đổi sau dịch mã chủ yếu của đuôi CTD *in vivo*. Sự mất khả năng đi vào quá trình chế biến của những đoạn phiên mã tạo ra bởi RNA pol I và III được cho là do sự mất đuôi CTD ở những enzyme này.

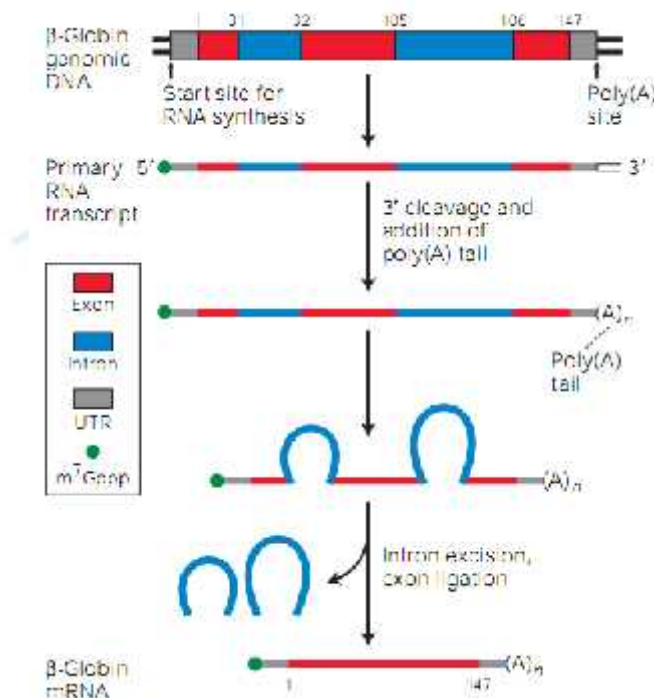
Việc gắn mũ chụp có hai chức năng chủ yếu: (1) ngăn chặn sự phân hủy mRNA trước khi trưởng thành, do đó làm tăng tính ổn định của mRNA; (2) đóng góp vào quá trình khởi sự dịch mã – một quá trình phụ thuộc vào mũ chụp. Ngoài ra, những phát hiện gần đây chỉ ra rằng mũ chụp cũng đóng một vai trò chủ yếu trong việc ức chế dịch mã điều hòa bởi miRNA (microRNA-mediated translational silencing).

3.3.2. Gắn đuôi poly(A) ^[1]

Sự biến đổi ở đầu 3' của tiền-mRNA liên quan đến sự phân cắt bởi một endonuclease để có được một nhóm 3'-hydroxyl tự do – vị trí gắn một chuỗi các adenylic acid bởi enzyme poly(A) polymerase. Kết quả tạo ra một đuôi poly(A) khoảng 100-250 base đối với động vật có xương. Poly(A) polymerase là một phần của phức hợp các protein có thể định vị và phân cắt một đoạn phiên mã tại vị trí đặc hiệu và sau đó thêm vào các adenyl theo một quá trình mà không cần có mạch khuôn.

3.3.3. Cắt nối (splicing) ^[1]

Bước cuối cùng trong quá trình biến đổi của nhiều loại mRNA ở eukaryote đó là cắt nối RNA (RNA splicing): quá trình cắt bỏ những đoạn intron bên trong đoạn phiên mã và tiếp sau đó là quá trình nối lại các đoạn exon mã hóa.



Hình 10: Tổng quan quá trình chế biến RNA tạo RNA trưởng thành ở eukaryote

Gene β -globin chứa ba exon mã hóa protein (vùng mã hóa, màu đỏ) và hai intron không mã hóa xen kẽ (màu xanh). Những đoạn intron làm đứt quãng trình tự mã hóa protein giữa các codon cho amino acid 31 và 32, 105 và 106. Sự phiên mã các gene mã hóa protein ở eukaryote bắt đầu trước trình tự mã hóa cho amino acid đầu tiên và kéo dài qua khỏi trình tự mã hóa cho amino acid cuối cùng, kết quả tạo ra những vùng không mã hóa (màu xám) ở cuối đoạn phiên mã sơ cấp (tiền-mRNA). Những vùng không được dịch mã (UTR) này được giữ lại trong suốt quá trình trên. Đầu 5' được gắn thêm mũ chụp (m^7Gppp) trong suốt quá trình hình thành đoạn phiên mã RNA. Đoạn phiên mã này kéo dài qua khỏi vùng poly(A) (poly(A) site). Sau quá trình cắt tại vùng poly(A) và quá trình thêm vào một chuỗi A tại đầu 3', quá trình cắt nối (splicing) tháo bỏ các đoạn intron và nối các đoạn exon lại với nhau tạo thành phân tử mRNA β -globin trưởng thành gồm 147 codon mã hóa cho amino acid.