

[www.mientayvn.com](http://www.mientayvn.com)

Dịch tiếng anh chuyên ngành khoa học tự nhiên và kỹ thuật.

Dịch các bài giảng trong chương trình học liệu mở của học viện MIT, Yale.

Tìm và dịch tài liệu phục vụ cho sinh viên làm seminar, luận văn.

Tại sao mọi thứ đều miễn phí và chuyên nghiệp ???

**Trao i tr c tuy n t i:**

[www.mientayvn.com/chat\\_box\\_sinh.html](http://www.mientayvn.com/chat_box_sinh.html)

**PGS. TS. Nguyễn Hoàng Lộc (chủ biên)**  
**TS. Lê Việt Dũng - TS. Trần Quốc Dung**

Giáo trình  
**Công nghệ DNA tái tổ hợp**

**NXB Đại học Quốc gia TP Hồ Chí Minh**  
**2007**

# Một số thuật ngữ cơ bản

**Adapter.** Một oligodeoxyribonucleotide tổng hợp tương tự linker, nhưng có một đầu bằng và một đầu lồi 5' tương ứng với một vị trí cắt hạn chế cho phép nối cDNA sợi đôi với các plasmid vector hoặc bacteriophage  $\lambda$  vector có đầu tương đồng (xem thêm linker).

**Adenosine diphosphate (ADP).** Một ribonucleoside 5'-diphosphate được cấu tạo từ adenine, đường ribose (5C) và hai gốc phosphate. ADP có tác dụng nhận phosphate trong chu trình năng lượng của tế bào.

**Adenosine triphosphate (ATP).** Một ribonucleoside 5'-triphosphate được cấu tạo từ adenine, đường ribose (5C) và ba gốc phosphate. ATP là phân tử chứa năng lượng hóa học chính của tế bào, chủ yếu được tập hợp trong ty thể (mitochondria) và lục thể (chloroplast). Các gốc phosphate của ATP có mang các liên kết khi bị thủy phân sẽ phóng thích một năng lượng tự do lớn. Năng lượng của quá trình hô hấp hoặc quang hợp được sử dụng để tạo thành ATP từ ADP. Sau đó, ATP được biến đổi ngược trở lại thành ADP ở nhiều vùng khác nhau của tế bào, năng lượng phóng thích ra được dùng để điều khiển các phản ứng hóa sinh nội bào. Đôi khi cũng xảy ra sự thủy phân tiếp ADP thành những AMP (adenosine monophosphate) để phóng thích năng lượng nhiều hơn.

**Amino acid.** Là một phân tử nhỏ mang một gốc amine ( $-\text{NH}_3$ ) và một gốc carboxyl ( $-\text{COOH}$ ) liên kết với cùng một nguyên tử carbon. Amino acid là đơn vị cấu trúc cơ sở của chuỗi polypeptide. Có 20 amino acid khác nhau trên các chuỗi polypeptide có trong tự nhiên. Trình tự sắp xếp của các amino acid trên chuỗi polypeptide quyết định cấu trúc và chức năng của polypeptide và protein mà nó tạo thành.

**Ampicillin (Amp).** Chất kháng sinh bán tổng hợp được dùng trong môi trường chọn lọc để chọn các tế bào mang đột biến khuyết dưỡng hoặc chọn dòng tế bào (tái tổ hợp) mang đoạn DNA được tạo dòng.

**BAC (bacteria artificial chromosome).** Nhiễm sắc thể nhân tạo của vi khuẩn, dựa trên cơ sở plasmid F-factor, được sử dụng làm vector tạo dòng. BAC có thể tái bản trong *E. coli* với các đoạn chèn DNA có kích thước lên đến 300 kb.

**Bản đồ cắt hạn chế (restriction map).** Trình tự các vị trí nhận biết (recognition sites) của tất cả các enzyme hạn chế (restriction enzyme hay restriction endonuclease, RE) trên một phân tử DNA.

**Bazo đồng đẳng (analog base).** Chất hóa học có cấu trúc phân tử rất giống các base bình thường của DNA. Chúng có thể thay thế các nitrogen base bình thường trong DNA và hoạt động như một tác nhân đột biến. Trong lần sao chép tiếp theo của DNA, base đồng đẳng có thể bắt cặp sai với một base bình thường, tạo nên đột biến điểm. Ví dụ: base đồng đẳng của adenine (A) là 2-aminopurine (AP) có thể gắn vào DNA ở vị trí của adenine; trong lần sao chép tiếp đó có thể bắt cặp với cytosine (C), trong lần sao chép tiếp theo nữa C kết cặp với guanine (G). Như vậy đã diễn ra sự thay thế cặp A-T bằng cặp G-C.

**Bazo nitơ (nitrogen base).** Loại phân tử cấu tạo nên nucleic acid (DNA và RNA). Các nitrogen base có trong nucleic acid là adenine, guanine, cytosine và thymine (DNA) hoặc uracil (RNA). Trình tự sắp xếp của chúng dọc theo phân tử nucleic acid đã tạo nên thông tin di truyền của cơ thể sinh vật.

**Bắt cặp bổ sung (complementary base pairing).** Sự kết hợp thành từng đôi giữa các nitrogen base nằm trên hai mạch đơn của chuỗi xoắn kép DNA-DNA, DNA-RNA hoặc RNA-RNA thông qua các mối liên kết hydrogen. Sự bắt cặp đó mang tính đặc hiệu: guanine bắt cặp với cytosine, còn adenine bắt cặp với thymine trên DNA hoặc uracil trên RNA.

**Biến nạp (transformation).** Là quá trình truyền DNA ngoại lai vào một tế bào nhận, chẳng hạn sphaeroplast hoặc protoplast, và có thể hợp nhất trong nhiễm sắc thể nhờ sự tái tổ hợp tương đồng hoặc được biến đổi trong một đơn vị sao chép tự trị (autonomous replicon). Sự biến nạp có thể xuất hiện trong các điều kiện tự nhiên ở một số vi khuẩn (ví dụ: *Bacillus*, *Haemophilus*, *Neisseria* và *Streptococcus*), nhưng ở nhiều vi khuẩn (ví dụ: *E. coli*) và các cơ thể sinh vật eukaryote sự biến nạp chỉ có thể xuất hiện ở những tế bào “thâm” được DNA bằng các phương pháp nhân tạo như: hóa biến nạp, điện biến nạp...

**Biến nạp bằng điện (electroporation).** Kỹ thuật dùng xung điện tạo ra các lỗ thủng tạm thời trên màng sinh chất để đưa DNA ngoại lai vào bên trong tế bào vật chủ.

**Biến tính (denaturation).** Là hiện tượng chuyển từ dạng mạch kép sang dạng mạch đơn của DNA và RNA thường do nhiệt gây nên. Biến tính của protein là hiện tượng chuyển từ cấu hình hoạt động thành dạng không hoạt động.

**Biểu hiện của gen (gene expression).** Là các quá trình phiên mã (transcription) và dịch mã (translation) của một gen để tạo ra sản phẩm protein của nó.

**Cặp base (base pair, bp).** Là liên kết A-T hoặc C-G trên một phân tử DNA mạch kép, và là đơn vị đo chiều dài của một phân tử DNA.

**Chromosome walking.** Kỹ thuật này dùng để lập bản đồ nhiễm sắc thể từ tập hợp các đoạn DNA cắt hạn chế chồng lên nhau (overlapping). Bắt đầu từ một thư viện trong đó chứa các đoạn DNA nói trên đã được tạo dòng. Một đoạn DNA mang một gen đã biết được lựa chọn và sử dụng như một mẫu dò để nhận dạng (ví dụ: bằng cách lai khuẩn lạc) các đoạn khác, là các đoạn chồng lên nhau chứa cùng một gen. Sau đó, trình tự nucleotide của các đoạn này sẽ được phân tích và nhờ vậy có thể xác định được toàn bộ các đoạn của nhiễm sắc thể. Từ đó, bản đồ của một vùng đặc biệt sẽ được xây dựng dần dần.

**Chu trình sinh tan (lytic cycle).** Một kiểu chu trình sống của thực khuẩn thể (bacteriophage) khi nó xâm nhiễm vi khuẩn, điều khiển các hoạt động sinh sản và sinh trưởng bằng các gen của nó và sinh ra các bacteriophage thế hệ con, chui ra khỏi tế bào vi khuẩn sau khi phá vỡ tế bào đó.

**Chu trình tiềm tan (lysogenic cycle).** Là hiện tượng hệ gen của bacteriophage hiện diện ở trạng thái ổn định và không sinh tan trong tế bào vật chủ sống của nó. Các tế bào vật chủ có thể tiếp tục sinh trưởng và phân chia, và sự sao chép của hệ gen bacteriophage (prophage) được phối hợp với nhiễm sắc thể của vật chủ sao cho khi tế bào phân chia thì prophage cũng được chuyển vào trong cả hai tế bào con. Prophage được duy trì bằng cách hoặc hợp nhất trong nhiễm sắc thể vật chủ (ví dụ: bacteriophage  $\lambda$ , bacteriophage  $\Phi 105$ ) hoặc như là một plasmid bên ngoài nhiễm sắc thể (ví

dụ: bacteriophage P1 và bacteriophage F116). Tế bào vật chủ có thể hoặc không thể biểu hiện ra một kiểu hình biến đổi.

**Chuỗi contig (contiguous sequence).** Một đoạn trình tự dài được hình thành từ một số các đoạn phân tử ngắn chồng lên nhau (overlapping).

**Chuỗi khảm (concatemer).** Phân tử DNA bao gồm nhiều đoạn cá biệt nối với nhau thông qua các đầu dính.

**Chuỗi mã hóa (coding sequence).** Đoạn phân tử DNA mang mã di truyền xác định để phiên mã thành mRNA và sau đó dịch mã thành chuỗi polypeptide.

**Chuyển gen (transgenic).** Quá trình chuyển một đoạn DNA ngoại lai (foreign DNA) bằng các kỹ thuật khác nhau (*Agrobacterium*, vi tiêm, bắn gen, xung điện...) vào một cơ thể vật chủ (vi sinh vật, động vật hoặc thực vật).

**Chuyển nhiễm (transfection).** Kỹ thuật đưa DNA phage hoặc DNA virus vào các tế bào vật chủ.

**Cosmid.** Vector lai (hybrid vector) được cấu thành từ các đoạn trình tự của plasmid và các vị trí *cos* (đầu dính) của bacteriophage  $\lambda$ .

**Công nghệ DNA tái tổ hợp (DNA recombinant technology).** Hệ thống các phương pháp phòng thí nghiệm cho phép cắt đoạn DNA từ một sinh vật để ghép nối vào DNA của một sinh vật khác tạo ra phân tử DNA tái tổ hợp. Phân tử này được đưa vào các sinh vật khác nhau để tạo ra những giống chủng vi sinh vật, thực vật và động vật mới có những phẩm chất đặc biệt, đáp ứng nhu cầu ngày càng cao của sản xuất và đời sống con người. Công nghệ này có ứng dụng rộng rãi trong y học, dược học, nông nghiệp và nhiều ngành công nghiệp khác.

**Công nghệ sinh học (biotechnology).** Theo nghĩa rộng là các quá trình công nghiệp có sử dụng vi sinh vật hoặc các tế bào động vật và thực vật (công nghệ sinh học truyền thống). Theo nghĩa phổ biến hiện nay đó là những quá trình sản xuất sử dụng các giống sinh vật mới, được tạo ra bởi công nghệ DNA tái tổ hợp (công nghệ sinh học hiện đại).

Trong công nghệ sinh học truyền thống (lên men sản xuất rượu, bia, ủ chua thực phẩm, làm phomat, trồng trọt, chăn nuôi...) trước tiên con người chọn lựa các đối tượng sản xuất thích hợp (chủng vi sinh vật, cây trồng và

vật nuôi) thông qua thực tiễn sản xuất và sau này bằng các phương pháp khoa học như gây đột biến, phân lập...

Ngày nay, bằng cách thay đổi gen nhờ công nghệ DNA tái tổ hợp người ta đã tạo ra các đối tượng sản xuất thích hợp hơn, có thể thay đổi cả về lượng và chất nhiều quá trình sản xuất bằng công nghệ sinh học truyền thống trước đây, nâng chúng lên vị trí cao hơn và mở ra những triển vọng lớn cho các lĩnh vực hoạt động của công nghệ sinh học.

**Deoxyribonucleotide triphosphate (dNTP).** Tiền chất đã được triphosphoryl hóa (“năng lượng cao”) cần thiết cho quá trình tổng hợp DNA. N được ký hiệu cho một trong bốn nitrogen base (A, G, T hoặc C).

**Deoxyribonuclease (DNase).** Loại enzyme nuclease thủy phân (phân hủy) DNA sợi đôi hoặc DNA sợi đơn.

**Deoxyribonucleic acid (DNA).** DNA là đại phân tử sinh học có cấu trúc xoắn đôi, tồn tại chủ yếu trong nhân tế bào, trên các nhiễm sắc thể, mang thông tin di truyền của sinh vật. Phân tử DNA gồm hai chuỗi polynucleotide, chuỗi nọ xoắn quanh chuỗi kia tạo nên chuỗi xoắn kép. Trong các nucleotide, theo chiều dọc các gốc phosphate nối xen kẽ với các phân tử đường deoxyribose tạo nên bộ khung bên ngoài của chuỗi xoắn kép, theo chiều ngang mỗi phân tử đường đều kết hợp với một trong bốn nitrogen base: adenine, guanine, cytosine hoặc thymine.

DNA không trực tiếp thể hiện chức năng sinh học mà gián tiếp qua protein do nó mã hóa. DNA tạo RNA, RNA tạo protein. RNA cũng là acid nhân (nucleic acid). Nó có thành phần cấu tạo khá giống DNA, ngoại trừ gốc thymine (T) trong DNA được thay thế bởi gốc uracil (U), và RNA ở dạng sợi đơn chứ không phải ở dạng xoắn kép như DNA.

Quá trình đọc mã di truyền chứa trong DNA để tổng hợp protein gọi là sự phiên mã (transcription) tạo ra RNA mang thông tin di truyền là mRNA (messenger RNA). mRNA kết hợp với một cơ quan tử trong tế bào là ribosome để tạo ra protein trong quá trình dịch mã (translation). Quá trình trên được gọi là quá trình sinh học căn bản.

Năm 1962, Watson (Mỹ) và Crick (Anh) đã chia sẻ Giải Nobel với Wilkins (Anh) về phát minh ra cấu trúc không gian của DNA và ý nghĩa của nó trong việc truyền thông tin di truyền. Điều đáng tiếc là Franklin, người

đã có những đóng góp đáng kể cho phát minh này đã mất trước đó. Theo qui định thì Giải Nobel không được phép tặng cho người đã mất.

**Dịch chuyển điểm đứt (nick translation).** Phương pháp đánh dấu DNA bằng phóng xạ [ $\alpha$ - $^{32}$ P]dCTP nhờ enzyme DNA polymerase I của *E. coli*.

**Dịch mã (translation).** Sự tổng hợp protein trên khuôn mRNA. Quá trình chuyển thông tin di truyền trong trình tự base của mRNA sang trình tự amino acid của chuỗi polypeptide trong tế bào còn gọi là quá trình sinh tổng hợp protein.

**Dịch mã ngược (reverse translation).** Là kỹ thuật phân lập các gen nhờ khả năng của chúng trong việc lai với một đoạn mã oligonucleotide nào đó, đoạn này được chuẩn bị bằng cách dự đoán đoạn mã nucleic acid từ những đoạn mã hóa của protein biết trước.

**Dideoxyribonucleotide triphosphate (ddNTP).** Một đồng phân của dNTP dùng để kết thúc một chuỗi DNA trong các thí nghiệm xác định trình tự gen (sequencing).

**Dimer.** Là một hỗn hợp được tạo ra giữa hai phân tử đồng nhất nhưng khối lượng phân tử thì gấp đôi so với phân tử nguyên thủy.

**DNA bổ sung (complementary DNA, cDNA).** DNA được tổng hợp trên khuôn mẫu mRNA nhờ quá trình phiên mã ngược (reverse transcription). Do vậy, nó có trình tự sắp xếp các nucleotide bổ sung với trình tự các nucleotide trên mRNA. Ví dụ: trên mRNA trình tự đó là UUGAAG thì trên các DNA bổ sung sẽ có trình tự tương ứng là AACTTC. DNA bổ sung được tổng hợp tự nhiên trong chu trình sống của virus mang vật chất truyền là RNA. Ví dụ: HIV, virus cúm và các retrovirus nói chung. DNA bổ sung được tổng hợp nhân tạo trong các phòng thí nghiệm để xây dựng thư viện cDNA (cDNA library).

**DNA khuôn mẫu (template DNA).** Sợi DNA mà trình tự các nucleotide của nó được dùng làm khuôn mẫu để tổng hợp nên sợi DNA mới trong quá trình tái bản (sao chép) hoặc khuếch đại DNA (PCR) hoặc để tổng hợp nên sợi RNA mới trong quá trình phiên mã.

**DNA polymerase.** Enzyme tổng hợp bản sao DNA trên khuôn mẫu DNA bằng cách xúc tác phản ứng gắn từng nucleotide riêng biệt vào đầu chuỗi DNA đang được tổng hợp.



Năm 1959, hai nhà khoa học người Mỹ là Kornberg và Ochoa đã được nhận Giải Nobel về những nghiên cứu đã làm sáng tỏ cơ chế cơ bản của quá trình sao chép DNA liên quan đến DNA polymerase I.

**DNA siêu xoắn (supercoiled DNA).** DNA xoắn lại trên bản thân nó, thường là kết quả của sự gấp khúc, mở xoắn hoặc xoắn lại của chuỗi xoắn kép DNA.

**DNA vệ tinh (satellite DNA).** Là những đoạn DNA mang các trình tự lặp lại nối tiếp có thành phần khác với trị số trung bình của DNA hệ gen. DNA vệ tinh tạo thành băng theo gradient tỷ trọng và dễ dàng phân biệt với băng của phần lớn DNA hệ gen do DNA vệ tinh có tỷ trọng nhỏ hơn. Bản sao của DNA vệ tinh lặp lại hàng triệu lần trong hệ gen, tập trung chủ yếu ở vùng tâm động và hai đầu của nhiễm sắc thể.

**Dòng (clone).** Tập hợp các tế bào hoặc phân tử giống hệt nhau cùng bắt nguồn từ một tế bào hay phân tử ban đầu.

**Dot blot.** Là kỹ thuật trong đó các vết tròn nhỏ (hoặc các điểm) có chứa nucleic acid được đặt lên màng nitrocellulose hoặc nylon để lai với đoạn môi DNA có đánh dấu đồng vị phóng xạ.

**Đa hình độ dài các đoạn cắt hạn chế (restriction fragment length polymorphism, RFLP).** Tính đa hình chiều dài các đoạn cắt hạn chế để chỉ các sai biệt di truyền ở vị trí nhận biết của các enzyme hạn chế (chẳng hạn như do sự thay đổi một nucleotide) dẫn đến sự sai biệt trong chiều dài của các đoạn hình thành từ phản ứng cắt hạn chế DNA với cùng một enzyme. RFLP thường được dùng để thiết lập bản đồ di truyền với một số marker di truyền biết trước.

**Đánh dấu ở đuôi (end labelling).** Bổ sung phân tử phóng xạ vào đuôi của một polynucleotide nhờ enzyme T4 polynucleotide kinase.

**Đầu bằng (blunt end).** Các đầu của DNA sợi đôi không có các đầu 3' hoặc 5' lồi ra (protruding ends).

**Đầu dính (cohesive ends hoặc sticky ends).** Các đầu của phân tử DNA có các trình tự bổ sung ngắn có thể dính kết lại để nối hai phân tử DNA với nhau. Các đầu dính thường do các enzyme hạn chế tạo ra.

**Đầu tận cùng C (C terminus).** Góc carboxyl (COOH) tự do ở vị trí tận cùng của một phân tử protein hoặc chuỗi polypeptide.

**Đầu tận cùng N (N terminus).** Gốc amine ( $\text{NH}_2$ ) ở vị trí tận cùng của một phân tử protein hoặc chuỗi polypeptide. Tất cả các polypeptide đều được tổng hợp từ đầu tận cùng N đến đầu tận cùng C.

**Điểm đứt (nick).** Điểm đứt gãy ở một sợi đơn trên DNA sợi đôi.

**Điện di trên gel (gel electrophoresis).** Kỹ thuật dùng để phân tách các phân tử nucleic acid hoặc protein dựa vào sự dịch chuyển của chúng trên giá thể dạng gel (agarose hoặc polyacrylamide) dưới ảnh hưởng của điện trường. Sự dịch chuyển của các phân tử này phụ thuộc vào điện tích, cấu hình, kích thước và khối lượng phân tử của nucleic acid hoặc protein cũng như dung môi và nồng độ của chất dùng làm giá thể.

**Đoạn cắt hạn chế (restriction fragment).** Các đoạn DNA nhỏ được sinh ra sau khi xử lý đoạn DNA lớn bằng enzyme hạn chế.

**Đoạn kết thúc phiên mã (terminator hay transcription terminator).** Trình tự nucleotide nằm ở cuối gen hoạt động như một tín hiệu kết thúc sự phiên mã. Nó ra hiệu cho RNA polymerase giải phóng phân tử RNA mới được tạo thành ra khỏi gen. Lưu ý không được nhầm với các bộ ba kết thúc (terminator codons hay stop codons: UAG, UAA và UGA), xuất hiện trong mRNA, là tín hiệu dừng của sự dịch mã (xem mã vô nghĩa). Có hai loại terminator phổ biến: Rho-independent terminator (thường là một cấu trúc thân-quai (stem-loop structure) trong RNA được phiên mã) nằm ở đầu của các operons, và Rho-dependent terminator (vùng không có cấu trúc đặc trưng của RNA, khi không được dịch mã, nó được xem như là yếu tố Rho) là nguyên nhân gây ra chiều phân cực của sự dịch mã (translational polarity).

**Đoạn Klenow (Klenow fragment).** Còn gọi là đoạn lớn của DNA polymerase I. Đây là một đoạn của DNA polymerase I (khối lượng phân tử 76.000) của *E. coli* đã bị mất hoạt tính exonuclease  $5' \rightarrow 3'$ .

**Đoạn môi (primer).** Một trình tự DNA hay RNA ngắn, bắt cặp với một mạch của DNA khuôn mẫu và có mang một đầu  $3'$ -OH tự do giúp DNA polymerase bắt đầu tổng hợp một chuỗi DNA mới.

**Đoạn nhồi (stuffer fragment).** Còn gọi là vùng đệm hay vùng trung tâm. Là một phần của phage  $\lambda$  có thể được loại bỏ và thay thế bằng đoạn chèn DNA (insert DNA) mà không ảnh hưởng đến khả năng sinh sản của phage trong tế bào vật chủ.

**Đoạn xuôi ngược như nhau** (palindrome). Đoạn DNA mạch kép có trình tự sắp xếp các base trên hai mạch đơn giống hệt nhau nếu cùng được đọc theo một chiều (chẳng hạn 5'→3'). Ví dụ: các đoạn nhận biết của enzyme hạn chế.

**Đóng dấu (replica plating).** Phương pháp chuyển nguyên mẫu các khuẩn lạc hoặc vết tan từ một đĩa thạch gốc sang đĩa thạch mới bằng cách dùng màng nylon (ví dụ: màng Hybond-N+) vừa khít áp lên mặt thạch của đĩa gốc để dính lấy các tế bào trong các khuẩn lạc (colony) hoặc vết tan (plaque) của đĩa gốc, rồi đưa màng này áp lên mặt thạch mới.

**Đơn vị phiên mã** (transcriptional unit). Đoạn DNA mã hóa cho phân tử RNA, bắt đầu từ điểm khởi đầu phiên mã đến điểm kết thúc phiên mã, nó có thể dài hơn một gen.

**Đơn vị sao chép (replicon).** Đoạn DNA bắt đầu từ điểm khởi đầu sao chép kéo dài về hai phía tới hai điểm kết thúc sao chép.

**Đơn vị tái tổ hợp (recon).** Đoạn DNA của gen có chiều dài đủ ngắn để sự trao đổi chéo không thể diễn ra ở bên trong nó được nữa. Hiện nay, được biết đó là một cặp nucleotide.

**Đuôi polyA (polyA tail).** Đoạn trình tự dài 50-200 nucleotide adenine được bổ sung vào đầu 3' của hầu hết các mRNA eukaryote sau khi phiên mã.

***E. coli* (Escherichia coli).** Vi khuẩn thường có trong ruột non của động vật có xương sống. *E. coli* được coi như sinh vật mẫu cho việc nghiên cứu hoạt động của tế bào. Đây là vi khuẩn Gram âm có kích thước genome khoảng  $4 \times 10^6$  base-pair. Các quá trình biểu hiện gen (phiên mã và dịch mã) đi đôi với nhau, sinh ra sợi mRNA được tổng hợp mới và được sử dụng ngay cho quá trình dịch mã. Không có hiện tượng biến đổi sau dịch mã (post-translation). Vì thế, *E. coli* được xem là một trong những tế bào vật chủ đơn giản nhất. Rất nhiều thí nghiệm tạo dòng gen đang được thực hiện hàng ngày tại các phòng thí nghiệm đều sử dụng *E. coli* làm vật chủ với nhiều chủng khác nhau về mặt di truyền và cho những ứng dụng đặc biệt.

**Endonuclease.** Là enzyme nuclease cắt bên trong phân tử nucleic acid, ngược với exonuclease là enzyme phân giải DNA từ một đầu hoặc cả hai đầu. Nuclease thủy phân những liên kết phosphodiester giữa các

nucleotide của một phân tử nucleic acid. Các nuclease có thể đặc hiệu đối với DNA (deoxyribonuclease) hoặc đặc hiệu đối với RNA (ribonuclease).

**Enzyme.** Chất xúc tác sinh học, là các phân tử sinh học có bản chất protein đóng vai trò chất xúc tác cho các phản ứng biến đổi hóa sinh.

**Enzyme gắn DNA (DNA ligase).** Enzyme dùng để nối các phân tử DNA với nhau bằng cách tạo ra mối liên kết phosphodiester giữa nhóm 5'-phosphate và nhóm 3'-hydroxyl trong quá trình tái bản hoặc sửa chữa DNA.

**Enzyme hạn chế (restriction enzyme, RE).** Loại endonuclease có khả năng cắt DNA tại những đoạn trình tự nhất định mà chúng nhận biết. Enzyme hạn chế được phát hiện vào năm 1970, chúng tồn tại trong tế bào vi khuẩn, có tác dụng cắt DNA ngoại lai (ví dụ: DNA của phage) tại những điểm xác định, để tiêu diệt DNA này. Cho đến nay hơn 900 enzyme hạn chế đã được tìm thấy. Các enzyme hạn chế được sử dụng rộng rãi trong các phòng thí nghiệm thao tác gen như những “chiếc kéo” cắt DNA tại những điểm đặc hiệu. Vị trí cắt phụ thuộc vào loại enzyme hạn chế được lựa chọn.

Năm 1978, Arber (Thụy Sĩ), Nathans (Mỹ) và Smith (Mỹ) đã được nhận Giải Nobel nhờ phát hiện ra enzyme hạn chế và những ứng dụng của chúng để giải quyết nhiều vấn đề quan trọng của sinh học phân tử. Các enzyme này là những “chiếc kéo phân tử” có thể cắt DNA thành những đoạn xác định, đã mở ra một thời kỳ phát triển mới của sinh học hiện đại - Thời kỳ thao tác gen.

**Enzyme phiên mã ngược (reverse transcriptase).** DNA polymerase phụ thuộc RNA (RNA-dependent DNA polymerase) có trong các RNA virus (retrovirus) được dùng để tổng hợp cDNA trong điều kiện *in vitro*.

**Exon.** Các đoạn DNA trong gen có chức năng phiên mã. Exon tồn tại ở cả sinh vật prokaryote và eukaryote. Riêng ở sinh vật eukaryote các exon nằm xen kẽ với các đoạn intron. Các intron chiếm tới 90% tổng số DNA của tế bào eukaryote và không có chức năng phiên mã.

**Eukaryote.** Sinh vật có tế bào mang nhân điển hình (nhân thật) nghĩa là nhân được bao bọc bởi màng nhân và tham gia vào hai cơ chế phân bào quan trọng là nguyên phân và giảm phân.

**Exonuclease.** Loại enzyme nuclease chỉ tác động vào đầu tận cùng của phân tử nucleic acid, cắt ra từng nucleotide một theo thời gian. Chúng có thể chuyển hóa theo đầu 5' hoặc 3' của sợi DNA.

**Ex vivo.** Thuật ngữ dùng để chỉ các thí nghiệm thực hiện trên tế bào nuôi cấy, các tế bào này sau đó sẽ được đưa vào một cơ thể sống.

**$\beta$ -galactosidase.** Enzyme được mã hóa bởi gen *lacZ*. Enzyme này thủy phân lactose thành glucose và galactose.

**Gen (gene).** Là đơn vị di truyền, yếu tố quyết định một tính trạng cơ thể. Thông tin di truyền của các gen được mã hóa trong DNA quyết định tính biến dị của loài và của cá thể. DNA là một chuỗi bao gồm các đơn vị nucleotide, có bốn loại nucleotide mang bốn nitrogen base khác nhau là adenine (A), guanine (G), cytosine (C), và thymine (T). Trình tự các nucleotide của một gen xác định một polypeptide hoặc một RNA. Gen có khả năng bị đột biến. Các gen chủ yếu nằm dọc theo nhiễm sắc thể ở trong nhân tế bào. Mỗi gen chiếm một vị trí xác định trên nhiễm sắc thể gọi là locus. Gen có thể tồn tại ở nhiều dạng gọi là các allele. Các gen biểu hiện thông qua các phân tử do chúng sinh ra là RNA (trong quá trình phiên mã) và protein (trong quá trình dịch mã).

**Gen chỉ thị (reporter gene).** Là một gen mã hóa mà sản phẩm của nó được trắc nghiệm một cách dễ dàng (ví dụ chloramphenicol acetyltransferase). Gen chỉ thị có thể được gắn với bất kỳ một promoter nào sao cho sự biểu hiện của nó có thể được dùng để thử nghiệm chức năng của promoter.

**Gen *lacZ*.** Gen của *E. coli* mã hóa  $\beta$ -galactosidase thích hợp cho chọn lọc thể biến nạp bằng khuẩn lạc xanh ( $\beta$ -galactosidase sẽ kết hợp với IPTG và X-gal được bổ sung trong môi trường nuôi cấy) và khuẩn lạc trắng (đoạn DNA ngoại lai xen vào giữa gen *lacZ* làm cho gen này mất hoạt tính vì thế không sản xuất được  $\beta$ -galactosidase).

**Ghép đôi lệch (mismatch).** Sự ghép đôi không đúng với quy luật bổ sung giữa các nucleotide thuộc hai sợi đơn DNA trong mạch kép.

**Ghép exon hay splicing (RNA).** Quá trình cắt bỏ những intron và nối các exon của sản phẩm phiên mã ban đầu (tiền thân mRNA) để tạo thành mRNA hoàn chỉnh (mature mRNA). Quá trình biến đổi này xảy ra trong nhân tế bào.

**Gốc tái bản (origin, *ori*).** Trình tự nucleotide hoặc vị trí trên DNA mà ở đó bắt đầu sự tái bản (sao chép).

**Gradient.** Biến thiên của một đại lượng theo một hướng nào đó. Một gradient mật độ được xác lập trong một số trường hợp ly tâm. Một gradient proton hoặc ion được tạo ra qua một màng nhờ sự vận chuyển tích cực đòi hỏi năng lượng.

**Hệ gen (genome).** Là tập hợp các gen có trong một tế bào đơn bội eukaryote, trong một tế bào prokaryote hoặc trong một virus. Hệ gen chứa toàn bộ DNA của cơ thể, ví dụ: hệ gen người chứa DNA dài khoảng 1,6 m nhưng chỉ rộng khoảng 5 phần tỉ mm. Ở người, số DNA nói trên được chia làm 46 phần có độ dài ngắn khác nhau gọi là nhiễm sắc thể (chromosome), là tập hợp DNA ở dạng nén chặt đến kích thước đường kính chỉ còn 3-4 phần triệu mét. Tuy nhỏ như vậy, nhưng nhiễm sắc thể người có đến 3 tỷ gốc nucleotide. Sự sắp xếp đặc thù của chúng quyết định bản chất sinh học của cơ thể.

**Hoạt tính phóng xạ đặc hiệu (specific radioactivity).** Là hoạt độ phóng xạ trên một đơn vị nguyên liệu, chẳng hạn: một mẫu dò đánh dấu phóng xạ có thể có hoạt tính đặc hiệu 10<sup>6</sup> lần đếm/phút trên microgram. Hoạt tính đặc hiệu cũng được dùng để xác định hoạt độ của enzyme.

**Huỳnh quang (fluorescence).** Hiện tượng phát một sóng ánh sáng có bước sóng khác với bước sóng đã được hấp thụ trước đó. Một số phân tử được gọi là thể huỳnh quang (ví dụ: enzyme luciferase ở con đom đóm) do có đặc tính này.

**In dấu DNA (DNA fingerprinting) hay in dấu di truyền (genetic fingerprinting).** Là phương pháp dùng các mẫu dò phóng xạ hoặc dùng kỹ thuật PCR để nhận dạng các băng DNA có các đoạn lặp lại với tần số cao. Bản mẫu hình các băng DNA là duy nhất đối với mỗi cá thể, và do vậy có thể dùng để xác định đặc trưng cá thể hoặc quan hệ huyết thống.

**In dấu chân DNA (DNA footprinting).** Phương pháp nhận dạng các vùng DNA mà các protein điều hòa bám vào.

**Intron.** Những đoạn DNA nhỏ ở sinh vật eukaryote không mang thông tin mã hóa amino acid, phân bố rải rác dọc theo phân tử DNA. Sau khi thông tin từ DNA được phiên mã sang mRNA thì các intron trên mRNA bị cắt bỏ, các đoạn mRNA còn lại gồm toàn các exon được nối lại với nhau

và chuyển đến ribosome để dùng làm khuôn mẫu cho quá trình dịch mã. Intron không thấy có ở sinh vật prokaryote.

***In vitro* và *in vivo*.** Là thuật ngữ mô tả thí nghiệm trong ống nghiệm (*in vitro*) và trong cơ thể (*in vivo*). Cùng với sự phát triển ứng dụng của máy tính, hiện nay các nhà khoa học còn tiến hành thí nghiệm mô phỏng bằng computer. Quá trình này gọi là thí nghiệm *in silico*.

**Kéo dài đoạn môi (primer extension).** Sự tổng hợp một bản sao nucleic acid bắt đầu từ đoạn môi. Được sử dụng để đánh dấu phóng xạ đoạn DNA làm mẫu dò hoặc khuếch đại một đoạn DNA bằng kỹ thuật PCR.

**Kháng nguyên (antigen).** Phân tử thường tìm thấy trên bề mặt tế bào, có tác dụng kích thích sự tạo thành kháng thể. Do vậy, nó được dùng để gây nên một phản ứng miễn dịch.

**Kháng thể (antibody).** Một protein (immunoglobulin) do bạch cầu lympho B của hệ thống miễn dịch sản sinh, có tác dụng nhận biết một kháng nguyên ngoại nhập đặc hiệu và gắn với nó, nếu kháng nguyên nằm trên bề mặt tế bào thì việc gắn kết này sẽ dẫn tới sự kết cụm tế bào và làm bất hoạt kháng nguyên.

**Kháng thể đơn dòng (monoclonal antibody).** Kháng thể xuất hiện trong phản ứng miễn dịch có nhiệm vụ gắn và tham gia loại bỏ chất lạ (antigen) lọt vào cơ thể. Thông thường, trong phản ứng miễn dịch có mặt hỗn hợp của nhiều loại kháng thể. Tuy nhiên, nhờ tế bào lai hybridoma người ta có thể tạo ra một loại kháng thể gọi là kháng thể đơn dòng. Kháng thể đơn dòng chủ yếu được sử dụng cho mục đích chẩn đoán bệnh.

**Khuếch đại (amplification.)** Sự sản xuất nhiều bản sao của một trình tự DNA nhờ kỹ thuật PCR.

**Khung đọc mở (open reading frame, ORF).** Là một trình tự mã hóa chuỗi polypeptide được bắt đầu với mã khởi đầu (initiation codon) và kết thúc bằng một mã dừng (stop codon). Một khung đọc mở bị ngăn chặn nếu một stop codon được định vị gần với mã khởi đầu. Mặc dù về lý thuyết mã di truyền được xây dựng dựa trên bộ ba nucleotide, do đó sẽ có ba khung đọc có thể có trên mỗi sợi DNA, tuy nhiên trong thực tế khung đọc chính xác được xác định bởi một điểm bắt đầu cố định.

**Khuyết đoạn (deletion, deficiency).** Đột biến nhiễm sắc thể dẫn đến làm mất một đoạn vật chất di truyền và thông tin di truyền chứa trong nó rời khỏi nhiễm sắc thể.

**Kiểu hoang dại (wild type).** Dạng thường thấy nhất của một gen trong quần thể hoang dại. Allele kiểu hoang dại được ký hiệu bằng một chữ in hoa hoặc thêm dấu cộng sau chữ viết thường, ví dụ: A hay a<sup>+</sup>. Allele kiểu hoang dại thường là trội và cho kiểu hình bình thường.

**Kilobase (kb).** 1000 base (hoặc cặp base), được dùng như đơn vị để đo hoặc xác định chiều dài của các phân tử DNA hoặc RNA.

**Kinase.** Các enzyme xúc tác phản ứng phosphoryl hóa một phân tử nhận nhờ ATP.

**Kỹ thuật di truyền (genetic engineering).** Còn gọi là công nghệ DNA tái tổ hợp. Bao gồm hệ thống các phương pháp di truyền phân tử dùng để thao tác vật chất di truyền, với ba bước chính gồm ba khâu chính. 1) Tách chiết DNA từ những sinh vật khác nhau; 2) Cắt và nối DNA ở những điểm đặc hiệu để tạo ra DNA tái tổ hợp (DNA mang các gen có nguồn gốc khác nhau), ví dụ: DNA plasmid có mang gen của người; 3) Đưa DNA tái tổ hợp vào hoạt động trong các tế bào hoặc cơ thể sống để sinh ra những sản phẩm đặc biệt cần thiết cho con người, ví dụ: DNA plasmid mang gen tạo insulin của người được đưa vào vi khuẩn *E. coli* để sản xuất.

**Lai khuẩn lạc (colony hybridization).** Kỹ thuật lai *in situ* để xác định vi khuẩn mang vector khảm (chimeric vector) mà DNA của vector này tương đồng với một đoạn mã hóa đặc biệt nào đó.

**Lai phân tử (molecular hybridization).** Quá trình trong đó hai mạch nucleic acid bổ sung (A-T, G-C) bắt cặp hình thành nên một mạch kép. Đây là một kỹ thuật hữu ích để phát hiện một trình tự nucleotide chuyên biệt.

**Lai tại chỗ (in situ hybridization).** Quá trình bắt cặp giữa mẫu dò (là một trình tự DNA sợi đơn hay RNA) với DNA của tế bào được cố định trên lam kính.

**Lập bản đồ hạn chế (restriction mapping).** Kỹ thuật dùng để xác định vị trí các điểm cắt hạn chế trên phân tử DNA.

**Linker.** Một oligonucleotide tổng hợp có hai đầu bằng, chứa một hoặc nhiều vị trí cắt hạn chế cho phép nối cDNA sợi đôi với các plasmid vector



hoặc bacteriophage  $\lambda$  vector. cDNA sợi đôi trước đó được xử lý với DNA polymerase của bacteriophage T4 hoặc DNA polymerase I của *E. coli* để tạo đầu bằng. Các linker sau khi gắn với hai đầu bằng của đoạn cDNA nhờ DNA ligase sẽ được cắt hạn chế để tạo ra đầu so le tương đồng với hai đầu của vector. Phản ứng gắn giữa đoạn cDNA có mang linker ở hai đầu với vector cũng được xúc tác nhờ DNA ligase.

**Lysosome.** Một bào quan có màng bao bọc ở trong tế bào chất của những tế bào eukaryote. Lysosome chứa nhiều enzyme thủy phân.

**Ly tâm theo gradient mật độ (density gradient centrifugation).** Kỹ thuật tách các hợp chất dựa vào sự khác nhau về mật độ của chúng được thực hiện bằng phương pháp ly tâm để làm lắng các chất qua một gradient nồng độ của saccharose hoặc CsCl.

**Mã di truyền (codon).** Nhóm ba nucleotide nằm kề nhau (bộ ba) trên phân tử mRNA xác định một amino acid trên chuỗi polypeptide, hoặc là tín hiệu kết thúc việc tổng hợp polypeptide.

**Mã thoái biến (degenerate codon).** Mã di truyền mà ở đó một amino acid được quy định bởi một số bộ ba nitrogen base, chứ không phải chỉ bởi một bộ ba. Thoái biến là đặc điểm vốn có của mã di truyền tồn tại phổ biến ở sinh giới.

**Mã vô nghĩa (nonsense codon) hay mã dừng (stop codon).** Là codon mà ở đó quá trình dịch mã dừng lại (nơi kết thúc của chuỗi polypeptide). Có tất cả ba codon vô nghĩa với các tên gọi là amber (UAG), ochre (UAA) và opal (UGA)

**Maturation.** Quá trình trong đó các mRNA vừa được phiên mã trải qua một số biến đổi hóa học để trở thành mRNA hoàn chỉnh sẵn sàng làm khuôn mẫu cho việc tổng hợp protein.

**Máy đếm nhấp nháy (scintillation counter).** Máy dùng để xác định hoạt tính phóng xạ trong một mẫu thí nghiệm.

**Mẫu dò (probe).** Một đoạn RNA hay DNA chuyên biệt được đánh dấu bằng đồng vị phóng xạ hay bằng hóa chất (chất phát huỳnh quang hoặc enzyme), dùng để định vị một trình tự nucleic acid nhất định thông qua kỹ thuật lai phân tử (xem Northern blot, Southern blot...)

**Mật độ quang (optical density).** Thông số cho phép đo độ hấp thụ ánh sáng ở một bước sóng nào đó của một môi trường hoặc dung dịch.

**Monomer.** Là các phân tử đơn vị nhỏ, có thể liên kết với các phân tử đơn vị giống nó để hình thành những phân tử lớn hơn (polymer). Ví dụ: các nucleotide là các monomer của nucleic acid và các amino acid là monomer của protein.

**Nấm men *Saccharomyces cerevisiae*.** Là một vi sinh vật nhân thật được sử dụng nhiều trong công nghệ DNA tái tổ hợp. Genome của nấm men *S. cerevisiae* khoảng  $1,35 \times 10^7$  base-pair nhiều hơn *E. coli* khoảng 3,5 lần. Nấm men thường được dùng làm tế bào vật chủ để biểu hiện những protein có cấu trúc phức tạp cần quá trình hậu dịch mã mà vi khuẩn *E. coli* không thể đáp ứng.

**Nhân tố kiểm soát phiên mã (transcription control element).** Đoạn nucleotide nằm xung quanh điểm bắt đầu và kết thúc của mỗi gen, tham gia vào sự điều hòa hoạt động biểu hiện của gen.

**Nhiệt độ nóng chảy (melting temperature,  $T_m$ ).** Là nhiệt độ mà ở đó có một nửa số phân tử của một trình tự DNA bị biến tính.

**Northern blot.** Kỹ thuật chuyển và cố định RNA từ formaldehyde agarose gel (sau khi được phân đoạn bằng điện di) lên màng lai bằng nylon hoặc nitrocellulose để lai với mẫu dò được đánh dấu đồng vị phóng xạ [ $\alpha$ - $^{32}$ P]dCTP hoặc digoxigenin-dUTP.

**Nucleic acids.** Những polynucleotide sinh học thiên nhiên, trong đó những đơn vị nucleotide được kết hợp với nhau bởi những liên kết phosphodiester thành trình tự DNA hoặc RNA riêng biệt.

**Nucleoside.** Một hợp chất gồm một base purine hoặc pyrimidine kết hợp đồng hóa trị với một phân tử đường pentose.

**Nucleotide.** Một nucleoside phosphoryl hóa với một trong những hydroxyl của pentose. Phân tử đóng vai trò cấu trúc cơ sở của DNA và RNA, gồm ba phần: đường pentose (ribose trong RNA, deoxyribose trong DNA), nitrogen base và gốc phosphate.

**Nucleolytic.** Phản ứng thủy phân một cầu nối phosphodiester trong một nucleic acid.

**Nuclease Bal 31.** Một loại enzyme exonuclease thủy phân cả hai sợi của phân tử DNA cùng một lúc.

**Oligo.** Tiếp đầu ngữ có nghĩa là “ít”, ví dụ: oligonucleotide (polynucleotide) có ít nucleotide hoặc oligopeptide (polypeptide) có ít peptide.

**Oligo(dT)-cellulose.** Một đoạn ngắn gồm các gốc deoxy-thymidine liên kết với cơ chất cellulose, được sử dụng để tinh sạch mRNA eukaryote bằng phương pháp sắc ký cột.

**Oligomer.** Thuật ngữ chung để chỉ một đoạn ngắn các monomer.

**Oligonucleotide.** Một đoạn ngắn các monomer là nucleotide, thường từ 20-30 nucleotide.

**Ôn hòa (temperate).** Trạng thái tiềm tan của các bacteriophage tế bào vật chủ.

**$\alpha$ -peptide.** Một phần của protein  $\beta$ -galactosidase, được mã hóa bởi đoạn gen *lacZ*.

**Phage.** Viết tắt của bacteriophage (thực khuẩn thể), là loại virus xâm nhiễm và sinh sản bên trong vi khuẩn. Phage thường có vỏ bọc protein, phức hợp bao gồm phần đầu có hình đa diện chứa nucleic acid và đuôi mà qua đó nucleic acid xâm nhập vào vi khuẩn chủ. Sau quá trình nhân lên của nucleic acid của phage, tế bào vi khuẩn chủ thường bị tan biến. Loại phage luôn luôn làm tan tế bào vi khuẩn khi chúng xâm nhiễm vi khuẩn gọi là phage độc. Ví dụ: phage T4. Ngược lại, còn có phage ôn hòa, khi xâm nhiễm vi khuẩn nó gây nên phản ứng tiềm tan, nghĩa là hệ gen của phage gắn vào nhiễm sắc thể vi khuẩn và được sao chép cùng với nhiễm sắc thể đó. Hệ gen của phage ở trạng thái gắn như vậy với nhiễm sắc thể vi khuẩn gọi là prophage.

**Phagemid.** Là một loại plasmid vector có mang các đoạn trình tự của phage.

**Phản ứng chuỗi polymerase (polymerase chain reaction, PCR).** Phương pháp dùng trong phòng thí nghiệm để khuếch đại các đoạn DNA đặc biệt lên hàng triệu lần trong vòng vài giờ thông qua 20-30 chu kỳ nhiệt, mỗi chu kỳ bao gồm ba mức nhiệt độ: biến tính ở 90-95°C, bắt cặp với môi

ở 40-65°C hoặc hơn và tổng hợp mạch mới nhờ DNA polymerase chịu nhiệt (Taq polymerase) ở 70-72°C. PCR có ứng dụng rộng rãi trong chẩn đoán y học, phân tích sự đa dạng sinh học, chọn giống và trong nhiều lĩnh vực khác.

Nhà khoa học Mỹ (Tiến sĩ Mullis) người phát minh ra kỹ thuật PCR đã nhận giải Nobel năm 1993. Cùng chia sẻ Giải Nobel với Mullis là Smith (Canada) do có những đóng góp mang tính nền tảng cho việc gây đột biến điểm định hướng, dựa trên các oligonucleotide và việc phát triển chúng trong các nghiên cứu protein.

**Phân tích trình tự gen (gene sequencing).** Là kỹ thuật xác định trình tự theo cấu trúc bậc một của chuỗi các nucleotide trong một phân tử nucleic acid. Phân tích trình tự của DNA có các phương pháp hóa học của Maxam-Gilbert và phương pháp enzyme của Sanger. Trong những năm gần đây, một số phương pháp xác định trình tự mới nhờ sự hỗ trợ của máy tính đã xuất hiện. Bên cạnh kỹ thuật thông thường sử dụng các polyacrylamide gel để phân ly các phân tử DNA có độ dài khác nhau, các kỹ thuật mới liên quan đến phát hiện huỳnh quang của các nucleotide được đánh dấu, phân tích trình tự DNA bằng khối phổ, điện di mao dẫn hoặc lai với các đoạn oligonucleotide được tổng hợp nhân tạo cũng đã ra đời.

Năm 1980, Sanger (Anh) và Gilbert (Mỹ) đã được trao giải Nobel do đã có những đóng góp quan trọng về phương pháp xác định trình tự các nucleotide trong phân tử DNA. Đóng góp này là mốc lịch sử to lớn trong sinh học phân tử, là nguyên lý của tất cả các máy xác định trình tự DNA tự động đang sử dụng hiện nay trên khắp thế giới.

**Phần cuối (telomere).** Đoạn cuối, phần cuối của một nhiễm sắc thể thẳng của eukaryote, bao gồm những trình tự DNA ngắn được lặp lại nhiều lần.

**Phần tâm (centromere).** Phần co thắt được thấy trên nhiễm sắc thể ở kỳ giữa, đó là nơi hai nhiễm sắc tử dính với nhau.

**Phiên mã (transcription).** Là quá trình được xúc tác bởi enzyme phiên mã RNA polymerase để tổng hợp mRNA từ khuôn mẫu DNA.

**Phiên mã ngược (reverse transcription).** Quá trình tổng hợp DNA từ khuôn mẫu mRNA nhờ enzyme phiên mã ngược (reverse transcriptase).

**Phóng xạ tự ghi (autoradiography).** Kỹ thuật phát hiện các phân tử có đánh dấu phóng xạ thông qua hiệu ứng tạo ảnh của các phân tử này trên phim X-quang.

**Phosphatase kiềm (alkaline phosphatase).** Enzyme loại bỏ các nhóm 5'-phosphate từ đầu của các phân tử DNA và để lại các nhóm 5'-hydroxyl.

**Plasmid.** Là DNA vi khuẩn có cấu trúc mạch vòng kép, nằm trong tế bào chất và ngoài nhân, có khả năng sao chép độc lập đối với nhiễm sắc thể của tế bào. Tồn tại cả ở sinh vật prokaryote và eukaryote. Ngày nay, các plasmid thiết kế nhân tạo được sử dụng rộng rãi như là các vector dùng trong các kỹ thuật tạo dòng và biểu hiện gen.

**Plasmid không tiếp hợp (non-conjugative plasmid).** Ví dụ: plasmid *ColE1*, là loại plasmid không dùng sự tiếp hợp cho quá trình sống, thông thường là các plasmid có kích thước bé, tồn tại với một số lượng nhiều. Cơ chế nhân lên (nhiều bản sao) của chúng cũng khác hoàn toàn với plasmid tiếp hợp.

**Plasmid không tương hợp (incompatible plasmid).** Là những plasmid có thể cùng tồn tại với nhau trong một vài thế hệ ở tế bào vật chủ, sau đó trong quá trình phân chia của tế bào, một số trong chúng sẽ bị thải loại. Muốn tương hợp trong cùng một tế bào vật chủ, các plasmid khác nhau phải có chung một số đặc điểm trong quá trình tồn tại.

**Plasmid tiếp hợp (conjugative plasmid) hay plasmid F (fertility (F) plasmid).** Là các plasmid chuyển giao những bản sao DNA của chúng từ vi khuẩn này sang vi khuẩn khác nhờ phương thức tiếp hợp (do chúng có tổ hợp gen để sản xuất các ống protein hay còn gọi là lông giới tính (sex pile) làm cầu nối giữa hai tế bào vi khuẩn với nhau). DNA của plasmid, thậm chí cả DNA của hệ gen vi khuẩn, thông qua các ống protein này để chuyển từ tế bào vi khuẩn “cho” sang tế bào vi khuẩn “nhận”. Các plasmid, ngoài việc chuyển giao DNA của riêng chúng, còn có khả năng chuyển giao một phần hay nhiều phần hệ gen của tế bào vi khuẩn vật chủ đến tế bào vi khuẩn “nhận” khác. Do vậy, chúng được gọi là các nhân tố giới tính (sex factor) của vi khuẩn vật chủ, có vai trò quan trọng trong bảo tồn di truyền của vi khuẩn theo phương thức này.

**Plasmid tương hợp (compatible plasmid).** Là những loại plasmid có trong cùng một tế bào vật chủ nhưng không ảnh hưởng lẫn nhau. Khi tế bào vật chủ phân chia, chúng cùng đồng thời phân chia và tồn tại vĩnh viễn.

**Polyacrylamide.** Là polymer của acrylamide và bisacrylamide có cấu trúc gồm các liên kết chéo tạo ra một mạng xốp (giống bột biển). Các chất phải chui vào lỗ gel mới ra được, vì vậy những chất nào có khối lượng phân tử nhỏ sẽ ra trước và ngược lại.

**Polynucleotide.** Trình tự những nucleotide nối đồng hóa trị với nhau, trong đó vị trí 3' của pentose của một trong những nucleotide được nối với một liên kết phosphodiester ở vị trí 5' của pentose của nucleotide tiếp theo.

**Polypeptide.** Một chuỗi dài những amino acid nối với nhau bởi những liên kết peptide.

**Prokaryote.** Sinh vật đơn bào không có nhân tế bào điển hình, DNA nằm trong tế bào chất không có màng bao bọc, không có nguyên phân và giảm phân; đại diện điển hình là vi khuẩn.

**Prophage.** Phage ôn hòa đã xen vào nhiễm sắc thể của vi khuẩn tiềm tan. Nó sao chép đồng thời với nhiễm sắc thể của tế bào vi khuẩn chủ.

**Protein.** Một phân tử lớn gồm một hoặc nhiều chuỗi polypeptide, mỗi chuỗi có một trình tự amino acid và một khối lượng phân tử đặc trưng. Protein là hợp chất quan trọng bậc nhất đối với cơ thể sống. Về cấu trúc, protein là phân tử mạch dài gồm các đơn vị cấu trúc nhỏ là các amino acid nối với nhau qua mỗi liên kết peptide. Khối lượng phân tử của protein từ vài nghìn đến vài triệu. Có khoảng 20 loại amino acid. Các loại protein phức tạp hơn có liên kết thêm với các nhóm bổ sung.

**Protein dung hợp (fusion protein).** Là một protein tái tổ hợp lai được mã hóa bởi một gen lai (fusion gene) do sự dung hợp *in vitro* các đoạn gen khác nhau trên plasmid vector và sau đó biến nạp vào vi sinh vật chủ (chẳng hạn *E. coli*). Vì vậy, protein dung hợp sẽ mang trình tự amino acid của hai protein khác biệt được tổng hợp từ đầu N của vector biểu hiện.

**Protein nguyên thể (native protein).** Là một protein tái tổ hợp được mã hóa bởi một gen ngoại lai (foreign gene) trong vi sinh vật chủ. Khác với protein dung hợp, protein nguyên thể được tổng hợp từ đầu N của nó chứ không phải từ đầu N của vector.

**Purine.** Một hợp chất dị vòng, kiềm, có nitrogen, là thành phần của những nucleotide và nucleic acid. Purine chứa một nhân pyrimidine kết hợp với một nhân imidazol.

**Pyrimidine.** Một nitrogen base dị vòng có ở trong các nucleotide và nucleic acid.

**Retrovirus.** Là loại virus RNA chứa enzyme reverse transcriptase và sinh sản dưới dạng DNA mạch kép. Chúng có khả năng xâm nhiễm tế bào vật chủ cao. Khi xâm nhiễm nó có khả năng gắn hệ gen của virus với hệ gen của tế bào vật chủ, là cơ sở để thiết kế các vector liệu pháp gen hiệu quả.

**Ribonuclease.** Enzyme xúc tác đặc hiệu việc phân hủy RNA bằng cách cắt các mối liên kết phosphodiester trên RNA.

**Ribonucleic acid (RNA).** Thường là phân tử đa phân mạch đơn gồm các đơn vị cấu trúc cơ sở là ribonucleotide. Về mặt hóa học RNA rất giống với DNA. RNA là vật chất di truyền của một số virus và là các phân tử trung gian trong quá trình tổng hợp protein mà thông tin về trình tự amino acid của chúng đã được mã hóa trong DNA.

**Ribonucleotide.** Đơn vị cấu trúc cơ sở của RNA, gồm ba thành phần: đường ribose, nitrogen base và nhóm phosphate.

**Ribosome.** Là cơ quan tử khi kết hợp với mRNA tạo ra bộ máy tổng hợp protein. Trong tế bào thường có hàng nghìn ribosome, ribosome của mọi tế bào đều gồm một tiểu đơn vị nhỏ và một tiểu đơn vị lớn. Mỗi tiểu đơn vị có mang nhiều protein và rRNA (trong đó rRNA là thành phần chủ yếu chiếm khoảng 65%) có kích thước khác nhau. Người ta cũng thấy ribosome trong ty thể, ở đó có sự tổng hợp một số protein ty thể.

**RNA bổ sung (complementary RNA).** RNA sinh ra bằng cách phiên mã từ khuôn mẫu sợi đơn DNA tương ứng.

**RNA ligase.** Enzyme nối các đoạn RNA với nhau sau khi các intron được cắt rời khỏi tiền chất của mRNA (pre-mRNA) ở các sinh vật eukaryote, tạo ra mRNA hoàn chỉnh sẵn sàng tham gia vào quá trình dịch mã diễn ra trên ribosome.

**RNA polymerase.** Còn gọi là RNA polymerase phụ thuộc DNA (DNA-dependent RNA polymerase), xúc tác việc tổng hợp RNA trên khuôn mẫu DNA trong quá trình phiên mã.

**RNA ribosome (ribosomal RNA, rRNA).** Là thành phần cơ bản của ribosome, đóng vai trò xúc tác và cấu trúc trong tổng hợp protein. Tùy theo hệ số lãg rRNA được chia thành nhiều loại: ở eukaryote có rRNA 28S; 18S; 5,8S và 5S; còn các rRNA ở *E. coli* có ba loại: 23S, 16S và 5S. rRNA chiếm nhiều nhất trong bốn loại RNA (khoảng 80% tổng số RNA tế bào), tiếp đến là tRNA khoảng 16%, mRNA chỉ khoảng 2%. Ngoài ra, tế bào sinh vật eukaryote còn chứa những phân tử RNA kích thước nhỏ của nhân (small nuclear, snRNA) chiếm khoảng <1% tham gia vào ghép nối các exon.

**RNA thông tin (messenger RNA, mRNA).** Một loại RNA được phiên mã từ một trình tự DNA. mRNA truyền thông tin di truyền từ nhiễm sắc thể tới ribosome để tổng hợp protein. Trong quá trình đó một sợi của chuỗi xoắn kép DNA được dùng làm khuôn mẫu, dọc theo nó các nucleotide của mRNA bổ sung được xếp thành hàng, nối với nhau tạo nên một polynucleotide giống hệt sợi DNA không làm khuôn mẫu ngoại trừ thymine được thay bằng uracil. Quá trình này gọi là phiên mã và phân tử mRNA mang mã di truyền được dùng để điều khiển sự hình thành protein trên ribosome.

**RNA vận chuyển (transfer RNA, tRNA).** Loại RNA mang các amino acid đến ribosome và sắp xếp chúng dọc theo phân tử mRNA đã nằm sẵn ở đó. Tại đây, các amino acid nối với nhau bằng liên kết peptide để tạo thành phân tử protein. Mỗi amino acid có một phân tử RNA vận chuyển riêng với bộ ba đặc trưng và như vậy các amino acid được sắp xếp theo trật tự của các nitrogen base trên mRNA trong quá trình dịch mã.

**RNA kích thước nhỏ của nhân (small nuclear RNA, snRNA).** Ngoài mRNA, tRNA và rRNA, tế bào eukaryote còn chứa những phân tử RNA kích thước nhỏ của nhân (chiếm khoảng <1%) tham gia vào ghép nối các exon.

**Sàng lọc (screening).** Kỹ thuật nhận dạng một dòng DNA trong một thư viện hệ gen (genomic library) hoặc thư viện cDNA (cDNA library) bằng một phương pháp lai mẫu dò có đánh dấu [ $\alpha$ -<sup>32</sup>P]dCTP với các vết tan (trường hợp dùng bacteriophage  $\lambda$  làm vector tạo dòng và cho xâm nhiễm vào vi khuẩn *E. coli*) hoặc khuẩn lạc (dùng plasmid làm vector tạo dòng) của các thư viện đó trên màng nylon hoặc nitrocellulose. Tín hiệu lai được phát hiện bằng phóng xạ tự ghi trên phim X-quang.



**Sinh học phân tử (molecular biology).** Khoa học nghiên cứu các hiện tượng sống ở mức độ phân tử. Lĩnh vực khoa học trẻ tuổi này là điểm gặp nhau của các khoa học kinh điển như di truyền học, hóa sinh học, tế bào học, vật lý học, hóa học hữu cơ và hóa lý. Theo cách hiểu phổ biến hiện nay, sinh học phân tử là khoa học nghiên cứu các gen và hoạt động của chúng ở mức độ phân tử, bao gồm phiên mã, dịch mã, sao chép, điều hòa biểu hiện gen, tái tổ hợp và chuyển gen...

**Sinh tổng hợp protein (protein synthesis).** Phản ứng hóa học diễn ra trên ribosome tạo nên các phân tử protein từ các amino acid trên cơ sở thông tin di truyền nhận được từ trong nhân tế bào thông qua mRNA.

**Southern blot.** Kỹ thuật chuyển và cố định DNA đã biến tính từ agarose gel (sau khi được phân đoạn bằng điện di) lên màng lai bằng nylon hay nitrocellulose để lai với mẫu dò được đánh dấu đồng vị phóng xạ [ $\alpha$ - $^{32}$ P]dCTP hoặc digoxigenin-dUTP.

**Số bản sao (copy number).** (1) Số các phân tử plasmid có trong một tế bào vi khuẩn. (2) Số lượng các bản sao của một gen trong hệ gen của một sinh vật.

**Sơ đồ phóng xạ tự ghi (autoradiogram).** Hình ảnh sinh ra trên phim X-quang do sự phát xạ của các hạt phóng xạ.

**Tái tổ hợp (recombination).** Quá trình mà trong đó nhiễm sắc thể hay phân tử DNA đứt ra rồi các phần đứt được nối lại theo một tổ hợp mới. Quá trình này có thể xảy ra trong tế bào sống (qua sự trao đổi chéo trong phân bào giảm nhiễm) hay trong ống nghiệm nhờ các enzyme cắt và nối DNA.

**Tạo dòng gen (gene cloning).** Còn gọi là nhân dòng, tách dòng hay dòng hóa, là sự sản sinh nhiều bản sao của một phân tử DNA, thường là phân tử DNA tái tổ hợp trong plasmid vector, bằng cách sao chép phân tử đó trong một vật chủ thích hợp chẳng hạn vi khuẩn *E. coli*.

**Terminal transferase.** Enzyme bổ sung các gốc nucleotide vào đầu 3' của oligonucleotide hoặc polynucleotide.

**Tế bào khả biến (competent cell).** Các tế bào vi khuẩn có khả năng tiếp nhận DNA ngoại lai trong quá trình biến nạp.

**Thể biến nạp (transformant).** Tế bào hoặc sinh vật nhận được gen của một sinh vật khác trong quá trình biến nạp và biểu hiện chức năng của gen đó ra kiểu hình.

**Thể đa hình (polymorphism).** Mô tả sự có mặt đồng thời của quần thể trong hệ gen biểu hiện biến dị có tính chất allele, có thể quan sát trên các allele tạo ra các kiểu hình khác nhau, hoặc sự thay đổi của DNA ảnh hưởng đến kiểu cắt hạn chế (restriction patterns).

**Thể tái tổ hợp (recombinant).** Các cá thể hoặc tế bào mang các tổ hợp gen khác với cha mẹ của chúng do các quá trình tái tổ hợp di truyền sinh ra.

**Thông tin di truyền (genetic information).** Thông tin được lưu trữ trong các phân tử DNA của sinh vật ở dạng trình tự sắp xếp của bốn nucleotide ký hiệu là A, T, C và G đóng vai trò như những "chữ cái" của "ngôn ngữ" di truyền. Trong ngôn ngữ này, mỗi từ chỉ có ba chữ cái gọi là một bộ ba. Nghĩa của mỗi từ là một amino acid có mặt trên phân tử protein tương ứng. Mỗi "câu" của ngôn ngữ di truyền là một gen chứa đựng thông tin di truyền để đảm nhiệm một chức năng trọn vẹn. Mỗi chức năng là một đặc tính sinh lý, hình thái hay cấu trúc của sự sống. Do cơ chế sao chép theo kiểu nửa bảo toàn của DNA mà thông tin di truyền được truyền chính xác từ thế hệ nọ sang thế hệ kia hầu như không thay đổi.

**Thư viện cDNA (cDNA library).** Tập hợp các dòng DNA được tạo ra từ mRNA của một tế bào hoặc một mô cụ thể trong bacteriophage vector, đại diện cho thông tin di truyền mà các tế bào đó biểu hiện.

**Thư viện hệ gen (genomic library).** Tập hợp tất cả các đoạn DNA được tạo ra từ phản ứng cắt hạn chế genome trong bacteriophage vector, đại diện được cho toàn bộ cho thông tin di truyền của một hệ gen.

**Trì hoãn gel (gel retardation).** Phương pháp xác định điểm bám của protein trên các đoạn DNA, dựa vào độ di chuyển chậm của chúng so với DNA không bị protein bám trong các thí nghiệm điện di trên gel.

**Trình tự dẫn đầu (leader sequence).** Một trong ba phần chủ yếu của một phân tử mRNA. Trình tự này nằm ở đầu 5' của mRNA và mang thông tin để ribosome và các protein đặc hiệu nhận biết bắt đầu quá trình tổng hợp polypeptide, trình tự dẫn đầu không được dịch mã thành trình tự các amino acid.

**Trình tự điều hòa (regulatory sequence).** Một trình tự của DNA tham gia vào quá trình điều hòa của gen. Ví dụ: trình tự promoter hoặc operator.

**Trình tự khởi động (promoter).** Trình tự nucleotide đặc hiệu nằm trong thành phần operon, có chức năng điều hòa hoạt động của operon, nơi RNA polymerase bám vào để bắt đầu quá trình phiên mã. Trình tự đặc trưng của promoter có khoảng 20-200 nitrogen base.

**Trình tự Shine-Dalgarno (Shine Dalgarno sequence, SD).** Còn gọi là vùng liên kết ribosome (RBS), là một phần của trình tự nucleotide ở đầu 5' của một mRNA prokaryote có thể kết hợp bổ sung cặp base với đầu 3' của 16S rRNA, dùng làm tín hiệu cho sự khởi đầu dịch mã.

**Trình tự tăng cường (enhancer).** Trình tự nucleotide dạng *cis* làm tăng cường độ phiên mã của promoter trong gen eukaryote. Nó có thể nằm cách promoter hàng ngàn cặp base và hoạt động theo cả hai hướng ở bất kỳ vị trí nào so với promoter.

**Ủ để gắn môi (annealing).** Dùng để chỉ sự bắt cặp của khuôn mẫu DNA sợi đơn với primer (môi) để tổng hợp nên một sợi DNA bổ sung bằng cách sử dụng các dNTP có trong môi trường để kéo dài primer nhờ sự xúc tác của enzyme Taq DNA polymerase (trong khuếch đại PCR) hoặc DNA polymerase I (trong tổng hợp cDNA).

**Ức chế amber (amber suppresser).** Là các gen đột biến (mã hóa cho tRNA) mà những anticodons của nó đã được kích hoạt để có thể nhận UAG codon cũng như các codon trước đó.

**Vật chủ (host).** Tế bào dùng để nhân các phân tử DNA lên nhiều lần.

**Vector.** Là các phân tử DNA được sử dụng trong tạo dòng và biểu hiện gen, và nhân bản chúng trong tế bào vật chủ (*E. coli* hoặc nấm men). Có ba nhóm vector chính gồm: (1) Nhóm plasmid, (2) Nhóm phage/phagemid, và (3) Nhóm nhiễm sắc thể nhân tạo (artificial chromosome: BAC, YAC...). Ý tưởng về vector chuyển gen bắt nguồn từ các plasmid của vi khuẩn. Vector chuyển gen là các phân tử DNA có khả năng tự tái sinh, tồn tại độc lập trong tế bào và mang được các gen cần chuyển.

**Vector tạo dòng (cloning vector).** Phân tử DNA mạch kép có khả năng tự sao chép trong tế bào vật chủ. Có thể gắn vào phân tử này một đoạn hoặc một vài đoạn DNA khác nguồn tạo nên phân tử DNA tái tổ hợp dùng để nhân dòng.

**Vector biểu hiện (expression vector).** Phân tử DNA mạch kép có mang các tín hiệu cần thiết (promoter, terminator và vùng liên kết ribosome) cho sự biểu hiện của một khung đọc mở (gen) sẽ được nhân dòng và sản xuất protein tương ứng trong tế bào vật chủ.

**Vết tan (plaque).** Vòng tròn trong suốt xuất hiện trên thảm đục của các vi khuẩn mọc trên môi trường thạch đặc, do sự tan vỡ lặp lại nhiều chu kỳ của các tế bào vi khuẩn bị bacteriophage xâm nhiễm và sinh tan.

**Virus.** Phức hợp chứa nucleic acid (DNA hoặc RNA) nằm trong một vỏ bọc protein, có khả năng gây nhiễm và tái bản bên trong tế bào vật chủ đặc hiệu, tạo ra nhiều virus, lan truyền từ tế bào này sang tế bào khác. Virus là dạng sống không có cấu trúc tế bào, có khả năng xâm nhập vào các tế bào sống xác định và chỉ sinh sản ở bên trong các tế bào đó. Giống như tất cả cá sinh vật khác, virus có bộ máy di truyền của riêng mình, mã hóa việc tổng hợp các hạt virus từ các chất có trong tế bào vật chủ. Như vậy, virus là những vật ký sinh nội bào. Virus phân bố ở khắp nơi trong tự nhiên, xâm nhập vào tất cả các nhóm sinh vật. Người ta đã biết khoảng 500 loại virus xâm nhập động vật máu nóng, 300 loại xâm nhập thực vật bậc cao. Một số khối u ung thư ở động vật và ở người có thể do virus. Virus tồn tại ở hai dạng: dạng nghỉ hay ngoại bào và dạng sinh sản hay nội bào. Kích thước của các hạt virus từ 15-350 nm, chiều dài của một số loại virus có thể đạt tới 2000 nm. Phần lớn virus chỉ nhìn thấy được qua kính hiển vi điện tử. Chất mang thông tin di truyền của virus là nucleic acid: DNA hoặc RNA. Vì vậy, có thể phân virus thành hai loại: loại mang DNA và loại mang RNA.

**Vị trí *cos* (*cos site*) hay đầu *cos* (*cos end*).** Trình tự nucleotide được cắt ra để tạo thành các phần mở rộng sợi đơn, kết dính ở hai đầu cuối của các phân tử DNA mạch thẳng của một phage nào đó (ví dụ: phage  $\lambda$ ).

**Vòng kẹp tóc (hairpin loop).** Vòng chuỗi đơn bổ sung tạo nếp gấp chứa các cặp base tạo thành xoắn kép, được dùng làm mồi (primer) cho quá trình tổng hợp sợi cDNA thứ hai.

**Vùng cùng hướng (downtream).** Đề cập đến vị trí của một đoạn trình tự nào đó nằm ở phía đầu 3' của gen hoặc một đoạn gen quan tâm.

**Vùng liên kết ribosome (ribosome binding sites, RBS).** Là trình tự nucleotide trên phân tử mRNA giúp cho nó bám vào ribosome trong quá trình dịch mã (xem trình tự Shine-Dalgarno).

**Vùng ngược hướng (upstream region).** Vị trí của một trình tự nucleotide nào đó nằm ở phía đầu 5' của phân tử DNA so với gen quan tâm.

**Vùng tạo dòng (multiple cloning sites, MCS) hay vùng đa nối (polylinker hay polycloning sites).** Đoạn DNA ngắn trong một vector thể hệ thứ ba (ví dụ: các nhóm pUC, pGEM, pBluescript...) có chứa các vị trí nhận biết cho một số enzyme cắt hạn chế thông dụng, được thiết kế để chèn đoạn DNA ngoại lai vào đây.

**Western blot.** Kỹ thuật chuyển protein tổng số đã được phân tách bằng điện di SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) lên màng nylon hoặc nitrocellulose để lai với kháng thể một đặc hiệu và sau đó là kháng thể hai có đánh dấu enzyme nhằm phát hiện protein kháng nguyên tương ứng của nó.

**YAC (Yeast artificial chromosome).** Nhiễm sắc thể nhân tạo của nấm men, được dùng làm vector để tạo dòng những đoạn DNA có kích thước rất lớn trong nấm men.

## **Tài liệu tham khảo/đọc thêm**

1. **Ban Từ điển-NXB Khoa học và Kỹ thuật.** 2002. Từ điển Bách khoa Sinh học. *NXB Khoa học và Kỹ thuật*, Hà Nội.
2. **Lawrence E.** 1995. Henderson's Dictionary of Biological Terms. 7<sup>th</sup> ed. *Longman Group Ltd.* Singapore.
3. **Nill K.** 2002. Glossary of Biotechnology Term. 3<sup>rd</sup> ed. *CRC Press LLC*, USA.
4. **Old RW and Primrose SB.** 1994. Principles of Gene Manipulation. *Blackwell Scientific Publications*, Osney Mead, Oxford, UK.
5. **Singleton P and Sainsbury D.** 2001. Dictionary of Microbiology and Molecular Biology. 3<sup>rd</sup> ed. *John Wiley & Sons, Ltd.* UK.

# Lời nói đầu

Công nghệ DNA tái tổ hợp (còn gọi là kỹ thuật di truyền hay kỹ thuật gen) là một bộ phận quan trọng và là công nghệ chìa khóa (key technology) của lĩnh vực công nghệ sinh học. Công nghệ DNA tái tổ hợp được ra đời trên cơ sở các thành tựu của sinh học phân tử và hiện nay đang đóng vai trò cách mạng đối với sự phát triển của sinh học cũng như cải tạo sinh giới.

Các kỹ thuật tái tổ hợp DNA đã cho phép các nhà công nghệ sinh học phân lập và khuếch đại một gen đơn từ genome của một sinh vật để có thể nghiên cứu, biến đổi và chuyển nó vào trong một cơ thể sinh vật khác. Các kỹ thuật này còn được gọi là tạo dòng gen do nó có thể sản xuất ra một số lượng lớn các gen xác định.

Bên cạnh các giáo trình như: sinh học phân tử, nhập môn công nghệ sinh học, công nghệ tế bào, công nghệ chuyển gen... giáo trình công nghệ DNA tái tổ hợp sẽ giúp sinh viên tiếp cận thêm một lĩnh vực khác của công nghệ sinh học thông qua việc cung cấp những kiến thức cơ bản theo hướng tạo dòng và biểu hiện gen như sau:

- Các enzyme dùng trong tạo dòng phân tử.
- Các hệ thống vector.
- Một số kỹ thuật cơ bản trong tạo dòng gen: điện di, PCR...
- Tạo dòng và xây dựng các thư viện genomic DNA và cDNA.
- Biểu hiện các gen được tạo dòng trong *E. coli*.

Do giáo trình này mới được xuất bản lần đầu tiên, hơn nữa lĩnh vực công nghệ DNA tái tổ hợp lại rất phức tạp, nên khó tránh khỏi thiếu sót hoặc chưa đáp ứng được yêu cầu bạn đọc. Vì thế, chúng tôi rất mong nhận được nhiều ý kiến đóng góp để lần xuất bản sau được hoàn thiện hơn.

Chúng tôi chân thành cảm ơn Quỹ Nâng cao chất lượng-Dự án Giáo dục đại học đã hỗ trợ chúng tôi biên soạn giáo trình này, PGS. TS. Lê Trần Bình đã đọc bản thảo và góp nhiều ý kiến quý báu.

**Các tác giả**

# Chương 1

## Các enzyme dùng trong tạo dòng phân tử

### I. Các enzyme hạn chế

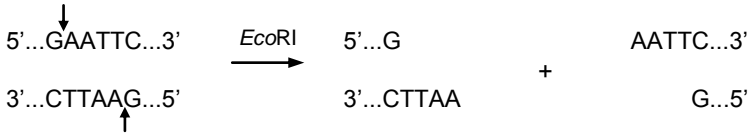
Các enzyme hạn chế được phân lập từ các sinh vật prokaryote, có khả năng phân hủy DNA của bacteriophage để hạn chế khả năng sinh trưởng của chúng ở trong vi khuẩn. Hiện nay, người ta đã tìm thấy hơn 900 enzyme hạn chế khác nhau từ khoảng 250 chủng vi sinh vật.

Các enzyme hạn chế có ba loại (type): I, II và III. Các enzyme được dùng phổ biến hiện nay thuộc type II, có cơ chế tác động đơn giản nhất. Đây là các nuclease cắt ở một vị trí đặc hiệu nằm bên trong sợi DNA (chứ không phân hủy DNA từ hai đầu), nên được gọi là endonuclease. Tên gọi đầy đủ của chúng là các restriction endonuclease type II, hay được gọi đơn giản là enzyme hạn chế (restriction enzyme, RE).

#### 1. Các enzyme hạn chế type II

Cách gọi tên các enzyme hạn chế dựa trên các qui ước quốc tế. Tên chi và tên loài của sinh vật, mà ở đó tìm thấy enzyme, được dùng để đặt cho phần đầu của tên enzyme (viết nghiêng) bao gồm: chữ thứ nhất của tên chi và hai chữ đầu của tên loài. Ví dụ: enzyme được tách chiết từ vi khuẩn *Escherichia coli* thì có tên là *Eco*, còn enzyme được tách chiết từ vi khuẩn *Bacillus globigii* thì viết là *Bgl*... Ngoài ra, tên gọi enzyme hạn chế còn được bổ sung thêm phần sau (viết thẳng), tùy thuộc vào chủng vi khuẩn liên quan và tùy thuộc vào sự có mặt hay không của các yếu tố ngoài nhiễm sắc thể. Ví dụ: RI trong enzyme *EcoRI* có nghĩa như sau: R là viết tắt của chủng RY13, I là bậc xác định đầu tiên trong vi khuẩn (first identified order in bacterium).

Giá trị của enzyme hạn chế là ở tính đặc hiệu của chúng. Các enzyme này cắt DNA ở các vị trí nhận biết riêng biệt bao gồm từ 4-6 cặp nucleotide có trình tự đối xứng đảo ngược nhau, các đoạn ngắn này gọi là palindrome (đoạn đối xứng: là đoạn DNA có hai sợi hoàn toàn đối xứng giống hệt nhau nếu lật ngược đầu đuôi). Ví dụ: enzyme *EcoRI* nhận biết chuỗi hexanucleotide (6 nucleotide):



Giống như *EcoRI*, nhiều enzyme hạn chế đã tạo ra các đoạn DNA với đầu lồi (protruding) 5'. Một số enzyme hạn chế khác (ví dụ: *PstI*) tạo ra các đoạn DNA có đầu lồi 5'. Trong khi đó, một số enzyme hạn chế (ví dụ: *SmaI*) lại cắt ở trục đối xứng để tạo ra các đoạn DNA mang đầu bằng (DNA mạch đôi có hai đầu sợi đơn bằng nhau) (Bảng 1.1).

**Bảng 1.1. Trình tự nhận biết của một số enzyme hạn chế tiêu biểu**

Stt	Tên enzyme	Vị trí nhận biết	Đầu được tạo ra
1	<i>EcoRI</i>	5'...G↓AATTC...3' 3'...CTTAA↑G...5'	5'...G AATTC...3' 3'...CTTAA G...5'
2	<i>HindIII</i>	5'...A↓AGCTT...3' 3'...TTCGA↑A...5'	5'...A AGCTT...3' 3'...TTCGA A...5'
3	<i>PstI</i>	5'...CTGCA↓G...3' 3'...G↑ACGTC...5'	5'...CTGCA G...3' 3'...G ACGTC...5'
4	<i>HpaI</i>	5'...GTT↓AAC...3' 3'...CAA↑TTG...5'	5'...GTT AAC...3' 3'...CAA TTG...5'
5	<i>HpaII</i>	5'...C↓CGG...3' 3'...GGC↑C...5'	5'...C CGG...3' 3'...GGC C...5'
6	<i>HaeIII</i>	5'...GG↓CC...3' 3'...CC↑GG...5'	5'...GG CC...3' 3'...CC GG...5'
7	<i>BamHI</i>	5'...G↓GATCC...3' 3'...CCTAG↑G...5'	5'...G GATCC...3' 3'...CCTAG G...5'
8	<i>BglII</i>	5'...A↓GATCT...3' 3'...TCTAG↑A...5'	5'...A GATCT...3' 3'...TCTAG A...5'
9	<i>SmaI</i>	5'...CCC↓GGG...3' 3'...GGG↑CCC...5'	5'...CCC GGG...3' 3'...GGG CCC...5'
10	<i>XmaI</i>	5'...C↓CCGGG...3' 3'...GGGCC↑C...5'	5'...C CCGGG...3' 3'...GGGCC C...5'

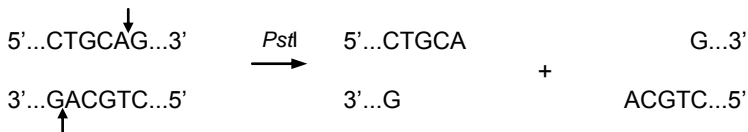


Với bốn loại nitrogen base trong phân tử DNA và giả thiết rằng trình tự sắp xếp của chúng là ngẫu nhiên, thì tần số mong đợi (kỳ vọng) của bất kỳ đoạn trình tự xác định nào, theo tính toán sẽ là  $4^n$ ,  $n$  ở đây là chiều dài của đoạn nhận biết. Từ đó, có thể thấy rằng với các đoạn có bốn nucleotide thì cứ cách 256 cặp base chúng lặp lại một lần, các đoạn sáu nucleotide thì cách 4096 cặp base mới lặp lại. Tất nhiên, các giá trị này có thể dao động rất lớn, nhưng nói chung chiều dài của các đoạn sinh ra đều gần với các giá trị tính toán. Chẳng hạn, một enzyme nhận biết đoạn trình tự bốn nucleotide sẽ sản sinh ra các đoạn ngắn hơn so với đoạn sáu nucleotide.

## 2. Gắn các đầu tận cùng được cắt bởi enzyme hạn chế

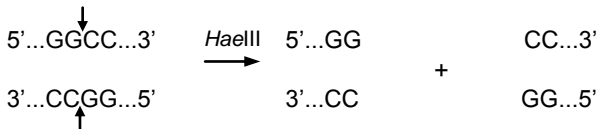
- **Các đầu dính (cohesive ends)**, còn gọi là đầu lồi hay đầu so le.

Ví dụ: đầu dính được tạo ra nhờ *Pst*I



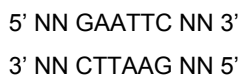
- **Các đầu bằng (blunt ends)**, còn gọi là đầu thô.

Ví dụ: đầu bằng được tạo ra nhờ *Hae*III



- Dùng enzyme DNA ligase của phage T4 có thể gắn lại các đầu dính hoặc đầu bằng. Tuy nhiên, trường hợp đầu bằng thường cho hiệu quả thấp vì thế người ta có thể dùng các biện pháp khác như:

+ *Gắn vào đầu bằng một đoạn bổ sung để tạo ra đầu dính.* Ví dụ: gắn đoạn linker (đoạn nối) bằng cách tổng hợp những trình tự oligonucleotide nhân tạo có từ 10-20 nucleotide để làm linker như sau:



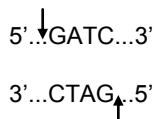
+ *Dùng terminal transferase.* Enzyme này sẽ tạo ở đầu 3' của DNA sợi đôi một homopolymer (xem chương 6).

### 3. Isochizomer

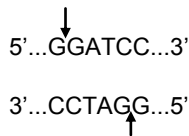
Nói chung, các RE khác nhau nhận biết các trình tự khác nhau (Bảng 1.1), tuy nhiên cũng có một số trường hợp cho thấy có những enzyme được phân lập từ nhiều nguồn khác nhau nhưng lại cắt trong một trình tự, các enzyme đó được gọi là isochizomer. Hơn nữa, một vài enzyme nhận biết các chuỗi tetranucleotide (4 nucleotide) mà trong một số trường hợp các tetranucleotide này lại xuất hiện trong các chuỗi hexanucleotide của các enzyme khác.

Ví dụ:

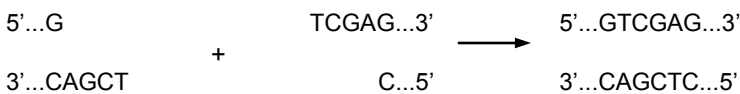
- *MboI* và *Sau3AI* cùng nhận biết trình tự:



- Trong khi đó *BamHI* nhận biết trình tự:

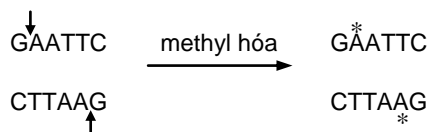


Trong một vài trường hợp khác các đoạn được tạo ra nhờ một enzyme hạn chế này khi gắn với các đoạn được tạo ra nhờ một enzyme hạn chế khác đã hình thành các thể lai (hybrid) mà các enzyme bố mẹ không thể nhận biết được. Ví dụ: *SalI* cắt trình tự (G↓TCGAC) và *XhoI* cắt trình tự (C↓TCGAG), các trình tự được sinh ra này đã gắn với nhau tạo thành thể lai mà các vị trí cắt hạn chế của chúng không thể nhận biết bởi các enzyme *SalI* và *XhoI*:



### 4. Methyl hóa

Sự methyl hóa (methylation) nhằm mục đích bảo vệ DNA của các đoạn palindrome, nghĩa là gắn gốc methyl (CH<sub>3</sub>) ở vị trí cần thiết nên không bị RE cắt. Ví dụ: *EcoRI* nhận biết chuỗi:



Khi có sự methyl hóa thì enzyme không nhận biết được vị trí cắt hạn chế và do đó DNA không bị cắt, những DNA của phage không có gắn gốc methyl sẽ bị cắt bởi RE. Trong kỹ thuật gen, enzyme methyl hóa được gọi là methylase enzyme. Methylase được dùng để bảo vệ đoạn DNA cần gắn vào.

Tất cả các chủng *E. coli* đều chứa hai enzyme methyl hóa DNA là: *dam* và *dcm*.

**- *dam* (DNA adenine methylase)**

Enzyme methylase này gắn các nhóm methyl ở vị trí N<sup>6</sup> của adenine trong chuỗi 5'...G<sup>me</sup>ATC...3'. Nhưng DNA của eukaryote không bị methyl hóa ở vị trí N<sup>6</sup> của adenine.

**- *dcm* (DNA cystosine methylase)**

Enzyme methylase này gắn các nhóm methyl ở vị trí C<sup>5</sup> của cytosine bên trong các chuỗi 5'...C<sup>me</sup>CAGG...3' hoặc 5'...C<sup>me</sup>CTGG...3'. Enzyme chủ yếu ảnh hưởng đến sự methyl hóa *dcm* là *EcoRII*, enzyme *BstNI* có thể nhận biết chính xác cùng một trình tự như *EcoRII* (mặc dù nó cắt DNA ở vị trí xác định khác trong chuỗi này). Nếu không thể thay *BstNI* cho *EcoRII* thì DNA phải được chuẩn bị từ các chủng *E. coli dcm*.

## 5. Cắt DNA bằng enzyme hạn chế

### 5.1. Các loại đệm dùng trong phản ứng cắt DNA

Mỗi enzyme hạn chế có một điều kiện phản ứng tối ưu, nhiệt độ ủ và thành phần của dung dịch đệm là những yếu tố quan trọng. Nhiệt độ yêu cầu khá chính xác còn sự khác nhau giữa các dung dịch đệm thường là không đáng kể. Để thuận tiện trong nghiên cứu người ta chia các enzyme ra làm ba nhóm:

- **Nhóm đệm cao (H)**. Hoạt động tốt nhất ở các dung dịch đệm có hoạt độ ion cao.

- **Nhóm đệm trung bình (M)**. Thích hợp với các dung dịch đệm có hoạt độ ion trung bình.

- **Nhóm đệm thấp (L)**. Thích hợp với các dung dịch đệm có hoạt độ ion thấp.

Như vậy, theo cách phân chia này chỉ có ba loại đệm stock là cần chuẩn bị cho phản ứng cắt DNA bằng RE (Bảng 1.2). Thường các đệm

stock trong bảng 1.2 được pha ở nồng độ ( $\times 10$ ), và có thể giữ ở nhiệt độ  $4^{\circ}\text{C}$ /1-2 tuần hoặc  $-20^{\circ}\text{C}$ /không hạn định.

**Bảng 1.2. Các loại dung dịch đệm cho phản ứng cắt DNA bằng RE**

Đệm	NaCl (mM)	Tris.HCl pH 7,5 (mM)	MgCl <sub>2</sub> (mM)	Dithiothreitol (mM)
Thấp	0	10	10	1
Trung bình	50	10	10	1
Cao	100	50	10	1

*Chú ý*

Riêng enzyme *SmaI* không thể sử dụng các loại đệm ở trên mà cần phải pha dung dịch đệm đặc biệt, bao gồm: 20 mM KCl, 10 mM Tris.Cl (pH 8), 10 mM MgCl<sub>2</sub>, và 1 mM dithiothreitol.

**5.2. Phương pháp cắt DNA bằng enzyme hạn chế**

Các phản ứng đặc trưng dùng khoảng 0,2-1  $\mu\text{g}$  DNA/20  $\mu\text{L}$  dung dịch phản ứng hoặc ít hơn. Ví dụ:

a. Pha dung dịch DNA bằng nước cất hai lần vô trùng trong eppendorf tube vô trùng tới dung tích 17  $\mu\text{L}$ .

b. Bổ sung 2  $\mu\text{L}$  đệm phản ứng ( $\times 10$ ) thích hợp của RE và trộn đều (vortex). Ví dụ:

*BstXI* : Dùng đệm cao  $10\times(\text{H})$

*XbaI* :  $10\times(\text{H})$

*EcoRI* :  $10\times(\text{H})$

c. Bổ sung 1 unit/ $\mu\text{L}$  RE và lắc đều nhẹ. Một unit (đơn vị, ký hiệu là u) RE được xác định như là số lượng yêu cầu đủ để thủy phân hoàn toàn 1 $\mu\text{g}$  DNA plasmid/1 giờ ở dung dịch đệm và nhiệt độ thích hợp trong 20  $\mu\text{L}$  phản ứng.

d. Ủ ở nhiệt độ thích hợp, thời gian tùy thuộc vào yêu cầu. Ví dụ:

*BstXI* : 1-2 giờ/ $50^{\circ}\text{C}$

*Xba*I : 1 giờ/37°C

*Eco*RI : 1 giờ/37°C

e. Dừng phản ứng bằng cách bổ sung 0,5 M EDTA (pH 7,5) để đạt tới nồng độ cuối cùng là 10 mM.

- Nếu DNA sau khi cắt bằng RE, được phân tích trực tiếp trên gel thì bổ sung 1  $\mu$ L ( $\times 10$ ) gel-loading-dye I, trộn đều và chạy điện di.

- Nếu DNA sau khi cắt bằng RE, cần được tinh sạch thì chiết một lần với phenol, một lần với phenol:chloroform (1:1), một lần với chloroform và kết tủa DNA bằng hai thể tích ethanol hoặc một thể tích isopropanol.

#### *Chú ý*

- Kết quả hoạt động của enzyme hạn chế phụ thuộc rất nhiều vào độ sạch của DNA. Các chế phẩm DNA còn lẫn đệm chiết, phenol hoặc EtOH sẽ làm giảm chất lượng hoạt động của enzyme.

- Enzyme hạn chế được giữ ở -20°C (hoặc -80°C nếu bảo quản lâu dài). Nếu lấy enzyme ra khỏi tủ lạnh sâu trong một thời gian ngắn để dùng thì phải đặt trên đá tuyết (ice bath).

## **II. Các enzyme trùng hợp**

### *1. DNA polymerase (DNA-dependent DNA polymerase)*

DNA polymerase được dùng để tổng hợp DNA trong cơ thể hoặc trong điều kiện *in vitro*.

#### 1.1. DNA polymerase I của *E. coli*

Enzyme này dịch chuyển điểm đứt (nick) của DNA, nick là điểm đứt trên một sợi của DNA sợi đôi. Chức năng chính của enzyme là sửa sai dọc theo phân tử DNA. Nếu gặp chỗ sai sót, nó sẽ đi ngược lại để cắt bỏ sau đó mới tổng hợp DNA. DNA polymerase I của *E. coli* mang chuỗi polypeptide có khối lượng phân tử (M) là 109.000 Da với ba hoạt tính riêng biệt:

- **Hoạt tính polymerase 5'→3'**. DNA polymerase I của *E. coli* xúc tác cho sự tổng hợp DNA theo chiều 5'→3' bằng cách bổ sung các gốc nucleotide vào đầu tận cùng 3'-OH được tạo ra khi một sợi của phân tử DNA sợi đôi bị đứt.

- **Hoạt tính exonuclease 3'→5'**. DNA polymerase I của *E. coli* cắt các chuỗi nucleotide ở các đầu tự do của DNA, xúc tác cho sự thoái biến bậc thang từ đầu 3' của cả DNA sợi đôi và sợi đơn khi không có dNTPs. Trong trường hợp có dNTPs, hoạt tính exonuclease trên sợi đôi sẽ bị ức chế bởi hoạt tính polymerase. Trong quá trình tổng hợp DNA, hoạt tính exonuclease thực hiện chức năng đọc sửa (proofreading) bằng cách cắt bỏ những nucleotide lắp ráp sai.

- **Hoạt tính exonuclease 5'→3'**. DNA polymerase I của *E. coli* có thể chuyển chỗ các nucleotide từ phía 5' của điểm đứt và bổ sung tức thời các nucleotide từ phía 3' tạo ra chuyển động của điểm đứt dọc theo DNA.

Hoạt tính	Phản ứng	Cơ chất
5' → 3' DNA polymerase	$\text{DNA}_{\text{OH}} + n\text{dNTP} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} \text{DNA}-(\text{p}d\text{N})_n + n\text{PP}_i$	Khuôn mẫu DNA sợi đơn/đoạn môi DNA mang nhóm 3'-OH
3' → 5' Exonuclease	$\text{dsDNA hoặc ssDNA} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} 5'_p\text{N}_{\text{OH}}$	DNA sợi đơn hoặc sợi đôi có đầu 3'-OH
5' → 3' Exonuclease	$\text{dsDNA} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} 5'_p\text{N}_{\text{OH}} + 5'_p\text{N}(\text{p}N)_{np}\text{N}_{\text{OH}} + \text{ssDNA}$	DNA sợi đôi hoặc thể lai RNA:DNA

### - Ứng dụng chính

- + Đánh dấu đoạn DNA để làm mẫu dò bằng dịch chuyển điểm đứt.
- + Khởi đầu cho sự tổng hợp sợi thứ hai trong tạo vòng cDNA.
- + Xác định trình tự DNA bằng phương pháp dideoxynucleotide.

### 1.2. Đoạn Klenow của DNA polymerase I của *E. coli*

Enzyme này có tác dụng lấp đầy các đầu khuyết 3' của DNA sợi đôi. Do DNA polymerase I có ba hoạt tính, nhưng hoạt tính exonuclease 5'→3' ít sử dụng nên Klenow đã dùng protease để cắt bớt đoạn có hoạt tính đó. Vì

vậy, khối lượng phân tử của enzyme chỉ còn lại 76.000 Da (được gọi là đoạn lớn của DNA polymerase I).

Hoạt tính	Phản ứng	Cơ chất
5' → 3' DNA polymerase	$\text{DNA}_{\text{OH}} + n\text{dNTP} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} \text{DNA}-(\text{p dN})_n + n\text{PP}_i$	Khuôn mẫu DNA sợi đơn/đoạn mồi mang nhóm 3'-OH tự do
3' → 5' Exonuclease	$\text{ds hoặc ss DNA} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} \text{}^5\text{pN}_{\text{OH}}$	DNA sợi đơn hoặc sợi đôi thoái biến từ đầu 3'-OH tự do. Hoạt tính exonuclease trên DNA sợi đôi bị ngăn cản bởi hoạt tính polymerase 5' → 3'

### - Ứng dụng chính

- + Làm đầy đầu khuyết 3' do enzyme hạn chế tạo ra.
- + Đánh dấu đầu tận cùng của các đoạn DNA bằng cách dùng [ $\alpha$ - $^{32}\text{P}$ ]dNTPs để làm đầy đầu khuyết 3'.
- + Đánh dấu đầu tận cùng của các phân tử DNA mang đầu lồi 3'. Đầu tiên, hoạt tính exonuclease 3'→5' loại bỏ đầu lồi 3' để tạo ra đầu khuyết 3'. Sau đó, nhờ sự có mặt ở nồng độ cao của một tiền chất được đánh dấu đồng vị phóng xạ, sự thoái biến bậc thang được cân bằng do sự hợp nhất của các dNTP ở đầu 3'.
- + Tổng hợp sợi thứ hai của cDNA trong tạo dòng cDNA.
- + Tổng hợp DNA sợi đôi từ các khuôn mẫu sợi đơn trong phát sinh đột biến *in vitro*.

### 1.3. DNA polymerase của bacteriophage T4

(T4-infected *E. coli*)

Enzyme này giống đoạn Klenow của DNA polymerase I của *E. coli*. DNA polymerase của phage T4 (M = 114.000 Da) có **hoạt tính polymerase**

**5' → 3'** và **exonuclease 3' → 5'**. Tuy nhiên, hoạt tính exonuclease của DNA polymerase phage T4 lớn hơn của DNA polymerase I đến 200 lần. Đoạn Klenow phải cần có đoạn mồi (primer) mới tổng hợp DNA được, còn DNA polymerase phage T4 thì không cần mồi cũng tổng hợp được DNA.

Hoạt tính	Phản ứng	Cơ chất
5' → 3' DNA polymerase	$\text{DNA}_{\text{OH}} + \text{ndNTP} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} \text{DNA}-(\text{p dN})_n + \text{nPP}_i$	Khuôn mẫu DNA sợi đơn/đoạn mồi mang đầu 3'-OH
3' → 5' Exonuclease	$\text{ssDNA} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} 5'_{\text{p}}\text{NOH}$	Hoạt tính trên DNA sợi đơn hơn trên DNA sợi đôi một cách đáng kể. Hoạt tính exonuclease trên DNA sợi đôi bị ngăn cản bởi hoạt tính polymerase 5' → 3'

### - Ứng dụng chính

- + Làm đầy hoặc đánh dấu đầu khuyết 3' do enzyme hạn chế tạo ra.
- + Đánh dấu đầu tận cùng của các phân tử DNA mang đầu lồi 3', tương tự như ứng dụng của đoạn Klenow.
- + Đánh dấu các đoạn DNA để làm mẫu dò.
- + Biến đổi đầu sole của DNA sợi đôi thành đầu bằng.

#### 1.4. Taq DNA polymerase

(*Thermus aquaticus*)

Taq DNA polymerase (Taq pol) là một loại DNA polymerase chịu nhiệt (M = 65.000 Da) của vi khuẩn *Thermus aquaticus*. Enzyme này được dùng để kiểm tra sự có mặt hoặc không của một gen bằng cách xúc tác cho sự tổng hợp gen đó ở điều kiện *in vitro* nhờ phản ứng chuỗi polymerase (PCR) (xem chương 3). Taq pol thay thế cho DNA polymerase I của *E. coli* do có khả năng chịu nhiệt cao phù hợp với điều kiện phản ứng của PCR, và nó có thể tổng hợp một sợi DNA dài 1 kb trong vòng 30 giây ở 72°C.



- Một trong những hạn chế của Taq pol là sự chính xác không cao trong quá trình sao chép của nó do bị mất cơ chế đọc sửa (hoạt tính exonuclease 3'→5'). Enzyme Taq pol thương mại có tỷ lệ sao chép lỗi 1/10.000 nucleotide, và có thể tạo ra 16% sản phẩm PCR dài 1 kb bị đột biến trong phản ứng khuếch đại. Mặc dù có nhược điểm trên, Taq pol vẫn có thể được dùng trong các thí nghiệm đòi hỏi một trình tự di truyền chính xác (như trong tạo dòng phân tử). Tuy nhiên, kết quả cho ra một vector mang đoạn chèn (sản phẩm PCR) cần phải được kiểm tra bằng sequencing.

- Ưu điểm của Taq pol là sản xuất các đoạn gen mang hai đầu lồi A. Điều này đặc biệt hữu ích trong “TA cloning” là kỹ thuật mà vector (plasmid) được sử dụng đã có sẵn hai đầu lồi T bổ sung với hai đầu lồi A của sản phẩm PCR, nhờ đó đã làm tăng hiệu quả của phản ứng gắn.

Hoạt tính	Phản ứng	Cơ chất
5' → 3' DNA polymerase	$\text{DNA}_{\text{OH}} + \text{ndNTP} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme}} \text{DNA}-(\text{p,dN})_n + \text{nPP}_i$	Khuôn mẫu sợi đơn/ đoạn mồi DNA mang đầu 3'-OH

Ngoài Taq pol, nhiều DNA polymerase chịu nhiệt khác đã được đưa ra thị trường với các chức năng chuyên biệt hay hoàn thiện hơn. Chẳng hạn: Pfu DNA polymerase (Pfu pol) được tách chiết từ *Pyrococcus furiosus*, thường được dùng thay cho, hoặc kết hợp, với Taq pol. Enzyme này ổn nhiệt hơn và có khả năng đọc sửa, vì thế cho tỷ lệ sao chép lỗi thấp. Hoặc Tth DNA polymerase (Tth pol), được tách chiết từ *Thermus thermophilus*, có khả năng hoạt động như một enzyme phiên mã ngược khi có mặt RNA khuôn mẫu và ion  $\text{Mn}^{2+}$ ; nhưng nếu có sự hiện diện của DNA khuôn mẫu và ion  $\text{Mg}^{2+}$ , thì Tth pol lại xúc tác phản ứng khuếch đại DNA. Enzyme này cho phép khuếch đại khuôn mẫu là RNA thông qua sự hình thành cDNA.

## 2. RNA polymerase (DNA-dependent RNA polymerase)

(Bacteriophage SP6-infected *Salmonella typhimurium* LT2, bacteriophage T7 hoặc T3-infected *E. coli*)

Các bacteriophage SP6, T7 và T3 sản xuất enzyme RNA polymerase để khởi đầu tổng hợp RNA trên khuôn mẫu DNA sợi đôi có mang promoter đặc trưng bacteriophage. Quá trình tổng hợp này không cần mồi.

Hoạt tính	Phản ứng	Cơ chất
5' → 3' RNA polymerase	$\text{enzyme}$ $\text{dsDNA} + n\text{rNTP} \xrightarrow{\text{Mg}^{2+}} \text{RNA} + \text{PP}_i + \text{khuôn mẫu}$	DNA sợi đôi mang các promoter của bacteriophage SP6, T3 hoặc T7

### - Ứng dụng chính

- + Sản xuất RNA để làm mẫu dò.
- + Xác định trình tự của một phân tử DNA được tạo dòng trong một vector có mang promoter đặc trưng cho bacteriophage SP6, T7 hoặc T3.
- + Sản xuất một lượng lớn RNA từ một phân tử DNA được tạo dòng để nghiên cứu về cấu trúc, điều hòa và các mối tương tác của phiên bản RNA.

### 3. Enzyme phiên mã ngược (reverse transcriptase, RNA-dependent DNA polymerase)

Đây là enzyme tổng hợp DNA từ khuôn mẫu RNA. Enzyme này có ở trong các virus: AMV (avian myeloblastosis virus), Mo-MLV (moloney murine leukemia virus) thuộc nhóm virus ngược (retrovirus: virus mang RNA) và có các hoạt tính sau:

- **Hoạt tính polymerase 5'→3'**. Tổng hợp DNA theo chiều 5'→3'. Cơ chất của nó là RNA hay DNA (sợi đơn hoặc sợi đôi):

Hoạt tính	Phản ứng	Cơ chất
5' → 3' DNA polymerase	$\text{enzyme}$ $\text{DNA}_{\text{OH}} + n\text{dNTP} \xrightarrow{\text{Mg}^{2+}} \text{DNA}-(\text{p}d\text{N})_n + n\text{PP}_i$ $\text{enzyme}$ $\text{Hoặc: RNA}_{\text{OH}} + n\text{dNTP} \xrightarrow{\text{Mg}^{2+}} \text{RNA}-(\text{p}d\text{N})_n + n\text{PP}_i$	Khuôn mẫu DNA hoặc RNA với đoạn mồi DNA hoặc RNA mang một nhóm 3'-OH

- **Hoạt tính RNase H.** Bao gồm hai hoạt tính exoribonuclease 5'→3' và 3'→5' phân hủy các DNA hoặc RNA trong tổ hợp lai.

Hoạt tính	Phản ứng	Cơ chất
RNase H (5' → 3' và 3' → 5' exoribonuclease)	$  \begin{array}{l}  \text{RNA } 5' \dots \text{pU}_p\text{C}_p\text{C}_p\text{G}_p\text{U}_p\text{A } 3' \\  \text{DNA } 3' \dots \text{A}_p\text{G}_p\text{G}_p\text{C}_p\text{A}_p\text{T } 5'  \end{array}  \xrightarrow{\text{enzyme}}  \begin{array}{l}  5' \dots \text{pU}_p\text{C } 3' \\  3' \dots \text{A}_p\text{G}_p\text{G}_p\text{C}_p\text{A}_p\text{T } 5' \\  + 5' \text{pC}_p\text{G}_p\text{U}_p\text{A } 3'  \end{array}  $	RNA trong thể lai RNA:DNA

Enzyme này được dùng trong kỹ thuật di truyền là nhờ vào khả năng tổng hợp được cDNA của chúng (xem chương 6).

Người ta có thể tách chiết mRNA ở những tế bào tổng hợp protein mạnh, sau đó dùng enzyme reverse transcriptase để tổng hợp cDNA. Có thể tổng hợp nên oligonucleotide rồi dùng nó để xác định trình tự của DNA.

#### 4. Terminal transferase

(Tuyến ức của bê)

Hoạt tính của enzyme này là xúc tác gắn các nucleotide vào đầu 3'-OH tự do của DNA sợi đôi làm lòi ra một đầu ở cuối sợi DNA. Phản ứng gắn này là ngẫu nhiên, do đó thành phần nucleotide của đầu lòi phụ thuộc vào nồng độ của bốn loại nucleotide có trong phản ứng.

Hoạt tính	Phản ứng	Cơ chất
Terminal transferase	$  \text{ssDNA}_{\text{OH}} + \text{ndNTP} \xrightarrow[\text{Mg}^{2+} \text{ hoặc } \text{Mn}^{2+}]{\text{enzyme}} \text{DNA}-(\text{p dN})_n + \text{nPP}_i  $	DNA sợi đơn mang đầu 3'-OH, hoặc: DNA sợi đôi mang đầu lòi 3'-OH

### - Ứng dụng chính

+ Thêm đuôi đồng trùng hợp (homopolymer) để tạo ra đầu so le cho phân tử DNA dùng trong tạo dòng (xem chương 6).

+ Đánh dấu đầu 3'-OH của DNA trong kỹ thuật xác định trình tự gen theo Maxam và Gilbert.

### III. Các enzyme gắn

Enzyme ligase thường dùng là của phage T4 xâm nhiễm trong *E. coli*. Chức năng của enzyme này là nối các đoạn hở bằng liên kết cộng hóa trị, sử dụng năng lượng ATP.

#### 1. Bacteriophage T4 DNA ligase

(Bacteriophage T4-infected *E. coli*)

Enzyme này là một chuỗi polypeptide (M = 68.000 Da) xúc tác cho sự tạo thành liên kết phosphodiester giữa đầu 3'-OH tự do của một phân tử DNA này với đầu cuối 5'-PO<sub>4</sub> của một phân tử DNA khác. DNA ligase có tác dụng cho cả trường hợp đầu so le lẫn đầu bằng, tuy nhiên đối với đầu bằng enzyme đòi hỏi nồng độ cao hơn và điều kiện phản ứng cũng khác.

Hoạt tính	Phản ứng	Cơ chất
Gắn các đầu dính hoặc điểm đứt (nick)	$  \begin{array}{ccc}  5' \dots pA_p C_p G_{OH} & & pA_p A_p T_p T_p C_p G_p T \dots 3' \\  3' \dots T_p G_p C_p T_p T_p A_p A_p & & HO G_p C_p A_p \dots 5' \\  & \downarrow \text{enzyme, ATP, Mg}^{2+} & \\  5' \dots pA_p C_p G_p A_p A_p T_p T_p C_p G_p T \dots 3' & & \\  3' \dots T_p G_p C_p T_p T_p A_p A_p G_p C_p A_p \dots 5' & &   \end{array}  $	(a) Các phân tử DNA sợi đôi có đầu dính bổ sung (b) nicked: DNA
Gắn các đầu bằng	$  \begin{array}{ccc}  5' \dots pC_p G_p A_{OH} & & pC_p G_p T_p A \dots 3' \\  3' \dots G_p C_p T_p & & OH G_p C_p A_p T_p \dots 5' \\  & \downarrow \text{enzyme, ATP, Mg}^{2+} & \\  5' \dots pC_p G_p A_p C_p G_p T_p A \dots 3' & & \\  3' \dots G_p C_p T_p G_p C_p A_p T_p \dots 5' & &   \end{array}  $	Nồng độ cao của các DNA sợi đôi đầu bằng mang nhóm 5'-PO <sub>4</sub> và 3'-OH

## 2. Bacteriophage T4 RNA ligase

(Bacteriophage T4-infected *E. coli*)

Enzyme này xúc tác cho mối liên kết cộng hóa trị giữa nhóm 5'-PO<sub>4</sub> của DNA sợi đơn hoặc RNA với nhóm 3'-OH của DNA sợi đơn hoặc RNA. Do cơ chất thích hợp cho enzyme này là các phân tử nhỏ nên nó thường được dùng để đánh dấu đầu 3' của các phân tử RNA sử dụng làm mẫu dò.

Hoạt tính	Phản ứng	Cơ chất
RNA ligase	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>DNA hoặc RNA</p> <math>5' \dots_p A_p C_p G_{OH} 3'</math> </div> <div style="text-align: center;"> <p>RNA hoặc DNA</p> <math>5' _p A_p A_p T_p T_p C \dots OH 3'</math> </div> </div> <div style="text-align: center; margin: 10px 0;"> <p>enzyme ↓ ATP</p> </div> <div style="text-align: center;"> <math>5' \dots_p A_p C_p G_p A_p A_p T_p T_p C \dots OH 3'</math> </div>	DNA sợi đơn và RNA

## 3. Bacteriophage T4 polynucleotide kinase

(Bacteriophage T4-infected *E. coli*)

Enzyme bacteriophage T4 polynucleotide kinase xúc tác chuyển nhóm  $\gamma$ -phosphate của ATP tới đầu 5' của DNA hoặc RNA. Hai loại phản ứng thường có thể xảy ra là phản ứng thuận và phản ứng trao đổi. Trong phản ứng thuận, nhóm  $\gamma$ -phosphate được chuyển tới đầu 5' của DNA đã được dephosphoryl hóa (mất nhóm phosphate). Trong phản ứng trao đổi, khi có một ADP thừa, enzyme sẽ chuyển nhóm 5' phosphate từ DNA đã được phosphoryl hóa tới ADP, DNA sau đó sẽ được phosphoryl hóa trở lại bằng cách nhận một nhóm  $\gamma$ -phosphate đã được đánh dấu đồng vị phóng xạ từ một phân tử [ $\gamma$ -<sup>32</sup>P]ATP.

### - Ứng dụng chính

+ Đánh dấu phóng xạ đầu 5' của DNA để phân tích trình tự bằng phương pháp Maxam và Gilbert (1977), dùng làm mẫu dò trong các kỹ thuật lai phân tử (xem chương 5).

+ Phosphoryl hóa các đoạn nối (linker) và các đoạn DNA không có nhóm 5' phosphate để chuẩn bị cho phản ứng gắn trong phương pháp tạo dòng.

Hoạt tính	Phản ứng	Cơ chất
Phản ứng thuận	$\text{DNA}_{\text{OH}}^{5'} \text{ hoặc } \text{RNA}_{\text{OH}}^{5'} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme, } [\gamma\text{-}^{32}\text{P}]\text{ATP dithiothreitol}}$ $^{32}\text{P}\text{DNA} \text{ hoặc } ^{32}\text{P}\text{RNA} + \text{ADP}$	DNA sợi đôi hoặc sợi đơn mang đầu 5'-OH, RNA mang đầu 5'-OH
Phản ứng trao đổi	$^5\text{pC}_p\text{G}_p\text{C}\dots^3 + \text{ADP thừa} \xrightarrow[\text{Mg}^{2+}]{\text{enzyme, } [\gamma\text{-}^{32}\text{P}]\text{ATP dithiothreitol}}$ $^5\text{pC}_p\text{G}_p\text{C}\dots^3 + \text{ADP} + \text{ATP}$	DNA sợi đơn mang đầu 5' phosphate

#### 4. Alkaline phosphatase

(*E. coli* và ruột bê)

Cả hai enzyme alkaline của vi khuẩn (bacteria alkaline phosphatase, BAP) và ruột bê (calf intestinal alkaline phosphatase, CIP) đều xúc tác loại bỏ nhóm 5' phosphate khỏi DNA và RNA.

Hoạt tính	Phản ứng	Cơ chất
Phosphatase	$\text{enzyme}$ $^5\text{pDNA} \text{ hoặc } ^5\text{pRNA} \rightarrow ^5\text{OHDNA} \text{ hoặc } ^5\text{OHRNA}$	DNA sợi đôi hoặc sợi đơn và RNA

#### - Ứng dụng chính

+ Loại bỏ nhóm 5' phosphate khỏi DNA hoặc RNA trước khi đánh dấu đầu 5' bằng <sup>32</sup>P.

+ Loại bỏ nhóm 5' phosphate khỏi các đoạn DNA để ngăn cản sự tự gắn. Trong kỹ thuật tạo dòng, khi một vector được mở vòng DNA bằng một RE rồi sau đó được loại bỏ nhóm 5' phosphate thì nó không thể tự gắn lại (tự tái tạo lại vòng). Chỉ khi có đoạn DNA ngoại lai mang các đầu 5' phosphate cần thiết được đưa vào thì phản ứng gắn giữa vector và DNA ngoại lai mới xảy ra.

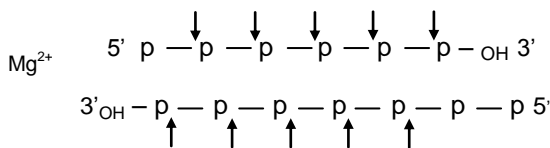
#### IV. Các enzyme phân cắt

Là nhóm các enzyme xúc tác thủy phân liên kết phosphodiester trong phân tử DNA hoặc RNA, bao gồm các endonuclease (cắt bên trong phân tử nucleic acid) và exonuclease (cắt bên ngoài phân tử nucleic acid). Các RE cũng là các endonuclease, nhưng chúng có tính đặc hiệu rất cao cho từng trình tự 4 hoặc 6 nucleotide. Dưới đây là các nuclease chính thường được sử dụng:

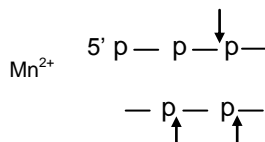
##### 1. Deoxyribonuclease I (DNase I)

(Tự bào)

DNase I xúc tác thủy phân DNA sợi đơn hoặc sợi đôi ở các liên kết phosphodiester nằm bên cạnh các pyrimidine nucleotide, tạo ra các oligonucleotide và các oligonucleotide có đầu tận cùng 5' monophosphate. Khi có mặt  $Mg^{2+}$ , DNase I tác dụng độc lập trên mỗi sợi DNA, và các vị trí cắt được phân bố ngẫu nhiên.



Khi có mặt  $Mn^{2+}$ , DNase I sẽ cắt cả hai sợi DNA gần như ở cùng một vị trí để tạo ra các đoạn DNA đầu bằng hoặc đầu lồi nhưng chỉ nhô ra một hoặc hai nucleotide.



**- Ứng dụng chính**

- + Tạo ra các điểm đứt (nick) trên DNA sợi đôi để đánh dấu mẫu dò đồng vị phóng xạ bằng phương pháp dịch chuyển điểm đứt.
- + Tạo ra các dòng ngẫu nhiên để phân tích trình tự trong bacteriophage M13 vector.
- + Phân tích phức hợp protein:DNA (DNase footprinting).
- + Loại bỏ DNA trong các phân đoạn RNA hay protein.

**2. Nuclease S1**

*(Aspergillus oryzae)*

Enzyme nuclease S1 cắt DNA sợi đơn hoặc RNA để tạo ra đầu 5' monophosphate hoặc các oligonucleotide. DNA sợi đôi và thể lai DNA:RNA ít chịu tác dụng của enzyme này. Tuy nhiên, các nucleic acid sợi đôi sẽ bị nuclease S1 cắt hoàn toàn nếu chúng bị xử lý với một lượng rất lớn enzyme. Một lượng vừa đủ của enzyme sẽ tách các nucleic acid sợi đôi ở các điểm đứt hoặc các lỗ hổng nhỏ thành hai hoặc nhiều đoạn.

Hoạt tính	Phản ứng	Cơ chất
<p>Nuclease đặc hiệu sợi đơn</p>	<p style="text-align: center;">           DNA sợi đơn hoặc RNA <math>\xrightarrow{\text{enzyme}}</math> 5'<sub>p</sub>dN hoặc 5'<sub>p</sub>rN         </p> <p style="text-align: center;">           DNA sợi đôi bị đứt (nick)             </p>	<p>DNA sợi đơn hoặc RNA, hoạt tính trên DNA lớn hơn trên RNA</p>

**- Ứng dụng chính**

- + Phân tích cấu trúc thể lai DNA:RNA.



- + Loại bỏ các phần sợi đơn trên đầu sole ra khỏi đoạn DNA sợi đôi để tạo ra các đầu bằng.
- + Mở vòng cặp tóc xuất hiện trong quá trình tổng hợp cDNA sợi đôi.

Một enzyme có hoạt tính tương tự nuclease S1 là mung-bean nuclease (endonuclease) được tách chiết từ mầm đậu xanh (*Phaseolus aureus*).

### 3. Exonuclease III (Exo III) (*E. coli*)

Exo III xúc tác loại bỏ các 5' mononucleotide từ đầu 3' OH của DNA sợi đôi (DNA mạch thẳng hoặc mạch vòng chứa các điểm đứt hoặc lỗ hổng). Hoạt tính enzyme cho kết quả tạo thành các vùng sợi đơn dài trong DNA sợi đôi. Enzyme này có 3 hoạt tính khác nhau: endonuclease đặc hiệu cho apurinic DNA, RNase H và 3' phosphatase (loại bỏ đầu 3' phosphate nhưng không cắt các liên kết phosphodiester bên trong). Exo III không cắt các DNA sợi đơn hoặc DNA sợi đôi có đầu lồi 3'.

Hoạt tính	Phản ứng	Cơ chất
3' exonuclease	<p>5' P ————— 3' OH 3' OH ————— 5' P</p> <p style="text-align: center;">Mg<sup>2+</sup> ↓ enzyme</p> <p>5' P ————— 3' OH 3' OH ————— 5' P</p> <p>+ 5' pNOH</p>	- Đầu tận cùng 3'-OH của DNA sợi đôi có đầu bằng hoặc đầu 3'-OH ở điểm đứt trong DNA sợi đôi
3' phosphatase	<p>5' ————— 3' P enzyme → 5' ————— 3' OH 3' P ————— 5' → 3' OH ————— 5'</p> <p>+ 2P<sub>i</sub></p>	DNA sợi đôi hoặc sợi đơn có đầu 3' phosphate

### - Ứng dụng chính

+ Tạo ra các cấu trúc sợi đơn ở một số vùng trên phân tử DNA dùng làm cơ chất cho đoạn Klenow của DNA polymerase I (để sản xuất mẫu dò đặc trưng cho từng sợi).

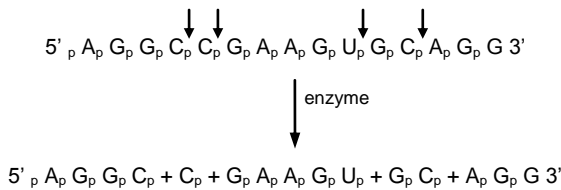
+ Tạo ra các đột biến khuyết đoạn (deletion) tại các trình tự tận cùng của DNA sợi đôi mạch thẳng. Phản ứng này thường phối hợp với mung-bean nuclease hoặc nuclease S1.

Một exonuclease tương tự là nuclease BAL31, có hoạt tính exonuclease 5'→3' và 3'→5', có khả năng tạo các DNA sợi đôi đầu bằng mà không cần sự hiện diện của nuclease S1.

#### 4. Ribonuclease (RNase A)

(Tụy bò)

Enzyme xúc tác thủy phân RNA thành các đoạn nhỏ hơn. RNase A có hoạt tính rất mạnh, hiện diện ở mọi nơi và rất bền vững (không bị mất hoạt tính khi bị xử lý ở 90°C trong 1 giờ). RNase A cắt liên kết phosphodiester nằm ngay sau một pyrimidine của một RNA sợi đơn.



#### - Ứng dụng chính

+ Loại bỏ RNA trong các chế phẩm DNA hay protein.

+ Loại bỏ các vùng không bắt cặp trên RNA trong thể lai RNA:DNA.

#### 5. RNase H

Enzyme RNase H là một loại ribonuclease có khả năng cắt liên kết 3'-O-P của RNA trong sợi đôi của thể lai DNA:RNA để tạo ra các sản phẩm có đầu tận cùng 3'-OH và 5'-PO<sub>4</sub>. RNase H là một endonuclease không đặc

hiệu, xúc tác cắt RNA thông qua cơ chế thủy phân nhờ một ion kim loại hóa trị 2 liên kết với enzyme.

Trong tạo dòng phân tử, RNase H xúc tác cắt đặc hiệu RNA trong thể lai RNA:DNA mà không cắt DNA hoặc RNA không ở trong thể lai, enzyme này thường được dùng để phá hủy khuôn mẫu RNA sau khi tổng hợp sợi cDNA thứ nhất bằng phiên mã ngược, để tiếp tục tổng hợp sợi cDNA thứ hai tạo thành một sợi đôi cDNA.

## V. Các protein liên kết DNA sợi đơn

Các protein liên kết DNA sợi đơn (single stranded DNA-binding proteins, SSB) kết hợp với DNA sợi đơn nhưng không kết hợp với DNA sợi đôi. Chúng được xem như là các chất phản ứng trong tạo dòng phân tử xuất phát từ chỗ chúng có khả năng phá vỡ sự ổn định của các cấu trúc thứ cấp bên trong sợi nucleotide; tăng nhanh sự tái ủ của các polynucleotide bổ sung và tăng hoạt tính của các enzyme DNA polymerase bằng cách loại bỏ các cấu trúc thứ cấp bên trong sợi là những yếu tố ngăn cản sự tiến triển của các enzyme này. Nhờ tính chất này, các protein SSB trở thành các chất phản ứng hữu ích trong xác định trình tự DNA.

### - Ứng dụng chính

- + Phát sinh đột biến điểm trực tiếp bao gồm D-loops.
- + Tăng chiều dài chuỗi của sản phẩm trong mọi phản ứng được xúc tác bởi DNA polymerase trên các khuôn mẫu dài. Các protein SSB đặc biệt có thể kích thích các DNA polymerase đặc hiệu dùng trong các phản ứng phân tích trình tự chuỗi DNA.

## Tài liệu tham khảo/đọc thêm

1. Hồ Huỳnh Thùy Dương. 1998. Sinh học phân tử. NXB Giáo dục, Hà Nội.
2. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K. 2002. Short Protocol in Molecular Biology. Vol 1 and 2. 5<sup>th</sup> ed. John Wiley & Sons, Inc. USA.

3. **Brown TA.** 2001. Gene Cloning-An Introduction. 4<sup>th</sup> ed. *Blackwell Science, Oxford, UK.*
4. **Glick BR and Pasternak JJ.** 2003. Molecular Biotechnology: Principles and Applications of Recombinant DNA. 3<sup>rd</sup> ed. *ASM Press, USA.*
5. **Maniatis T, Fritsch EF and Sambrook J.** 1989. Molecular Cloning-A Laboratory Manual. *Cold Spring Harbor Laboratory Press, USA.*
6. **Ohman DE.** 1989. Experiments in Gene Manipulation. *Prentice Hall, Englewood Cliffs, New Jersey, USA.*
7. **Primrose SB, Twyman R and Old RW.** 2001. Principles of Gene Manipulation. 6<sup>th</sup> ed. *Blackwell Science, Oxford, UK.*

## Chương 3

# Khuếch đại *in vitro* DNA bằng phản ứng chuỗi polymerase (PCR)

PCR (polymerase chain reaction) là một kỹ thuật được sử dụng phổ biến trong công nghệ sinh học hiện đại và đã đóng góp rất lớn cho những tiến bộ về sinh học phân tử, đánh dấu một bước tiến vô cùng quan trọng tương đương với việc khám phá ra các enzyme hạn chế (xem chương 1) và kỹ thuật Southern blot (xem chương 5).

PCR dựa trên cơ sở phản ứng kéo dài primer nhờ enzyme Taq polymerase để khuếch đại *in vitro* các nucleic acid đặc hiệu trong thiết bị điều nhiệt tuần hoàn (thermocycler) còn gọi là máy PCR (Hình 3.1). PCR cho phép khuếch đại theo hàm mũ lên đến hàng triệu lần các đoạn DNA có chiều dài từ 200-3.000 bp. Đoạn DNA được khuếch đại (DNA đích) được nhận dạng nhờ cặp primer đặc hiệu (oligonucleotide) thường có chiều dài khoảng 20 nucleotide.



Hình 3.1. Thiết bị điều nhiệt tuần hoàn

## I. Nguyên tắc của PCR

Taq polymerase (xem chương 1) là một loại enzyme DNA polymerase chịu nhiệt (có ở vi khuẩn chịu nhiệt độ cao *Thermus aquaticus*) được dùng để tổng hợp các đoạn DNA mới trong môi trường có bốn loại

deoxyribonucleotide (dATP, dCTP, dGTP và dTTP) và hai primer, trên cơ sở khuôn mẫu của một đoạn DNA nhất định đã biết hoặc chưa biết trình tự. Các đoạn DNA mới hình thành lại được sử dụng làm khuôn mẫu. Sau nhiều chu kỳ, số lượng đoạn DNA nói trên được nhân lên gấp nhiều lần, nhờ vậy có thể đủ số lượng để tách ra, phân tích trình tự hoặc tạo dòng... Primer ở bên trái tác động trên sợi DNA 3'-5' được gọi là primer thuận (forward primer, ký hiệu là F). Primer ở bên phải tác động trên sợi DNA 5'-3' được gọi là primer ngược (reverse primer, ký hiệu là R).

Nguyên tắc của PCR được trình bày trong hình 3.2 và 3.3. Theo đó, từ chu kỳ thứ hai Taq DNA polymerase (gọi tắt là Taq pol) bắt đầu tạo ra các đoạn DNA có chiều dài xác định. Các primer thường là một oligonucleotide tổng hợp (synthetic oligonucleotide) có khoảng 10-20 nucleotide hoặc hơn. Nếu biết trình tự của đoạn gen cần khuếch đại thì có thể tổng hợp nhân tạo các primer tương ứng để thực hiện PCR và tách chúng ra bằng kỹ thuật điện di. PCR thường tiến hành khoảng 25-35 chu kỳ, qua đó từ  $10^{-6}$   $\mu$ g DNA ban đầu có thể khuếch đại (amplification) lên tới trên 1  $\mu$ g (khoảng 2 kb). Mỗi chu kỳ PCR bao gồm ba giai đoạn có nhiệt độ khác nhau:

**- Gây biến tính (denaturation) ở 90-95°C**

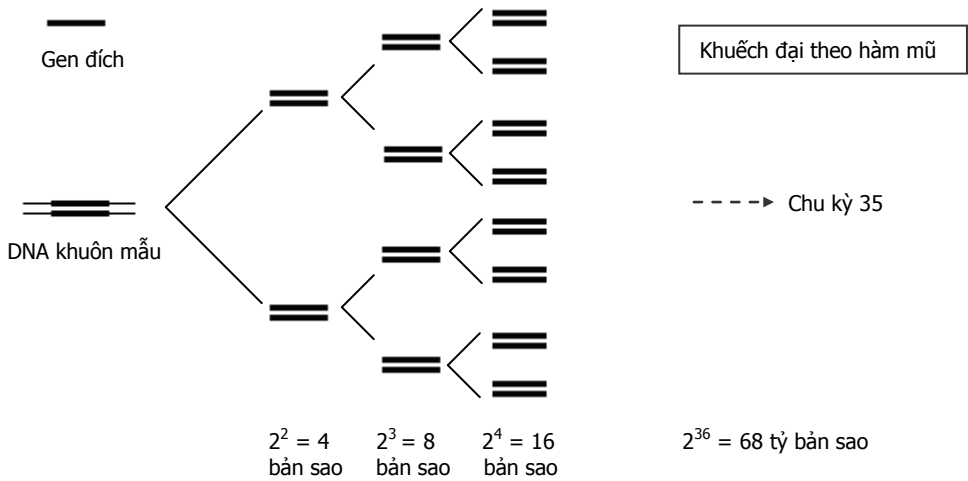
Trong giai đoạn biến tính, phân tử DNA khuôn mẫu ở dạng xoắn kép được tách thành hai sợi đơn (single strands). Tất cả các phản ứng enzyme trong giai đoạn này đều bị dừng lại (ví dụ: phản ứng tổng hợp DNA từ chu kỳ trước đó).

**- Gắn môi (annealing) ở 40-65°C**

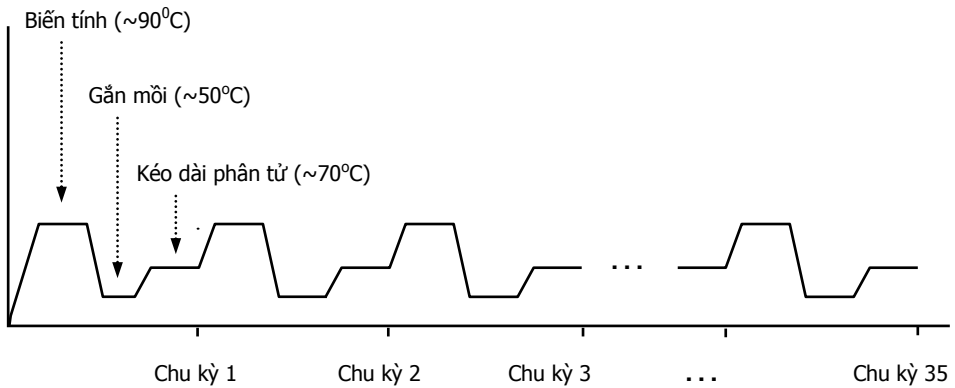
Trong giai đoạn này các primer gắn vào các vị trí có trình tự tương đồng ở DNA khuôn mẫu. Các primer bị lắc nhẹ chung quanh do chuyển động Brown vì thế các liên kết ion được tạo thành và bị đứt gãy liên tục giữa primer sợi đơn và DNA khuôn mẫu sợi đơn. Các liên kết ion ổn định hơn tạo thành một đoạn nhỏ (các primer đã lắp ráp chính xác) và trên các đoạn nhỏ DNA sợi đôi đó (khuôn mẫu và primer) Taq pol có thể bắt đầu quá trình sao chép khuôn mẫu. Ở giai đoạn này phạm vi nhiệt độ được sử dụng có thể rất rộng tùy thuộc vào trình tự nucleotide của primer, thông thường khoảng 55°C, nhưng có khi chỉ 35°C hoặc đôi lúc lên đến 68°C.

**- Kéo dài phân tử (extension) ở nhiệt độ 70-72°C.**

Đây là khoảng nhiệt độ tối thích cho Taq pol tiến hành tổng hợp DNA bằng cách bổ sung các dNTP bắt đầu từ các vị trí có primer theo chiều 5'→3'. Các primer có một vài base gắn vào khuôn mẫu có mối liên kết ion mạnh hơn lực phá vỡ các liên kết này sẽ không bị đứt gãy. Các primer ở các vị trí không bắt cặp chính xác lại bị rời ra (do nhiệt độ cao) khỏi khuôn mẫu đã không tổng hợp được DNA.

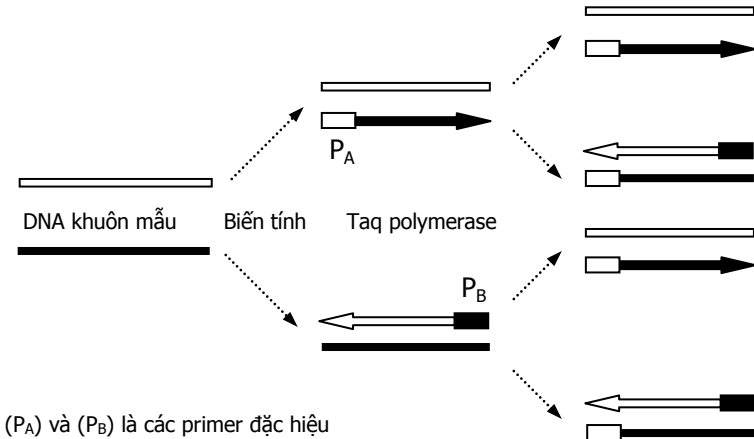


**Hình 3.2. Sơ đồ phản ứng chuỗi polymerase**



**Hình 3.3. Các chu kỳ của kỹ thuật PCR**

Có ba kỹ thuật PCR thông dụng: PCR chuẩn, PCR mở neo và PCR đảo ngược. Trong **PCR chuẩn (standard PCR)**, DNA khuôn mẫu được mở xoắn kép, sau đó hai primer tác động ở hai đầu sợi đơn, rồi DNA mới được tổng hợp nhờ Taq pol với 4 loại deoxyribonucleotide (Hình 3.4).



**Hình 3.4. Sơ đồ kỹ thuật PCR chuẩn**

**PCR mở neo (anchored PCR)**, kỹ thuật này yêu cầu chỉ cần biết trình tự nucleotide ở một đầu của khuôn mẫu và gắn một đuôi đồng trùng hợp nhân tạo (ví dụ: dA) vào đầu kia (chưa biết trình tự). Sau đó đoạn DNA được khuếch đại nhiều lần nhờ một primer có trình tự đã biết trước và một primer thứ hai có oligo(dT) (Hình 3.5).

**PCR đảo ngược (inverse PCR)**, được dùng để khuếch đại các đoạn phân tử kề cận với đoạn có trình tự DNA đã biết trước, phân tử DNA này được cắt hạn chế và nối lại nhờ enzyme DNA ligase để tạo thành một vòng tròn có tính chất monomer, sau đó đoạn DNA được khuếch đại nhờ hai primer tương đồng với đầu của trình tự đã biết trước (Hình 3.6).

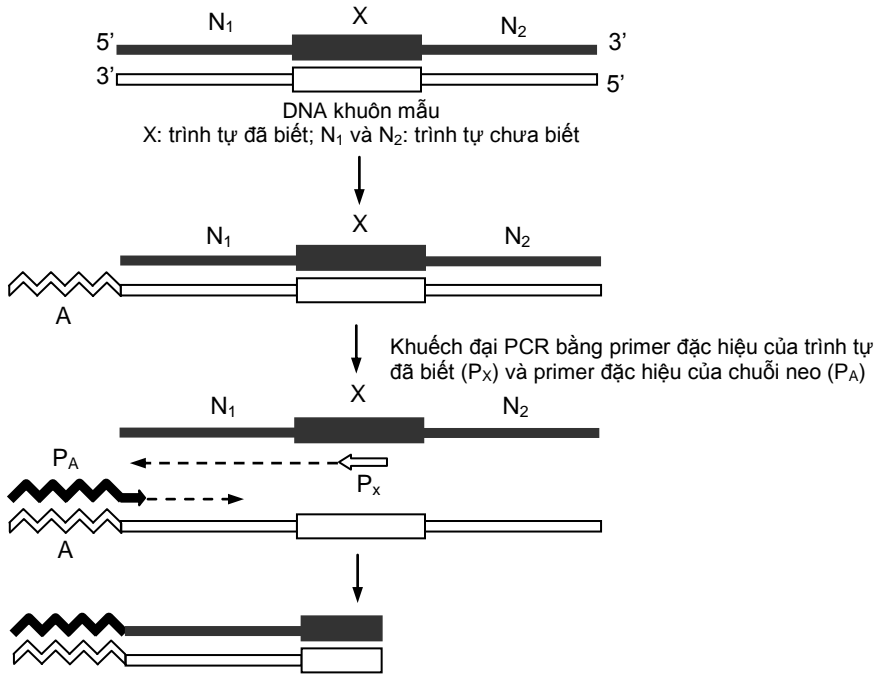
## II. Quy trình PCR

### 1. Thành phần phản ứng

- Phân tử DNA có chứa đoạn DNA cần khuếch đại
- 2 primer (F và R)



- Taq pol
- 4 loại dNTP
- Đệm và các muối khoáng



**Hình 3.5. Sơ đồ kỹ thuật PCR mở neo**

## 2. Thực hiện phản ứng

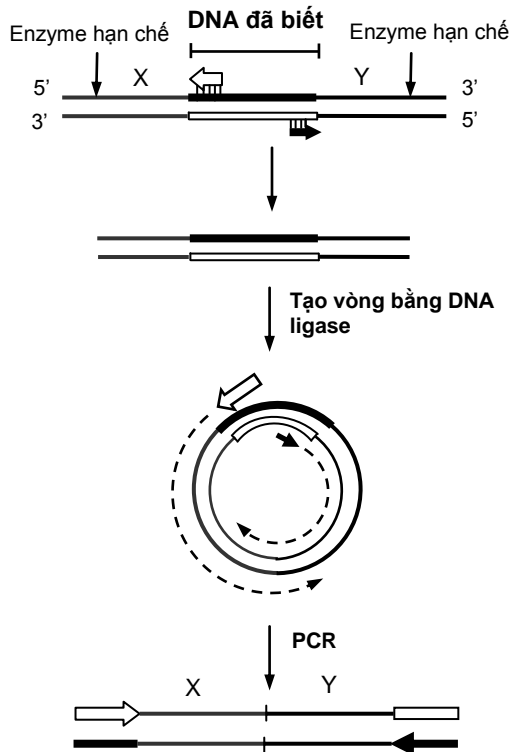
Dưới đây là ví dụ minh họa cho một phản ứng khuếch đại:

2.1. Chuẩn bị dung dịch master mix và cho vào eppendorf tube (E-tube) loại 0,5 mL, trộn đều các thành phần sau:

Đệm 10× PCR	4,5 μL
Hỗn hợp 4 dNTP (2,5 mM mỗi loại)	4 μL
Primer-F (10 pmol/μL)	2,5 μL
Primer-R (10 pmol/μL)	2,5 μL
Thêm H <sub>2</sub> O tới	40 μL

## 2.2. Chuẩn bị dung dịch pha loãng của Taq pol:

Đệm $\times 10$	2 $\mu\text{L}$
Taq pol (5 units/ $\mu\text{L}$ )	1 $\mu\text{L}$
Thêm $\text{H}_2\text{O}$ tới	20 $\mu\text{L}$



**Hình 3.6. Sơ đồ kỹ thuật PCR đảo ngược**

2.3. Bổ sung 5  $\mu\text{L}$  DNA khuôn mẫu (100 ng) vào 40  $\mu\text{L}$  dung dịch của master mix tube.

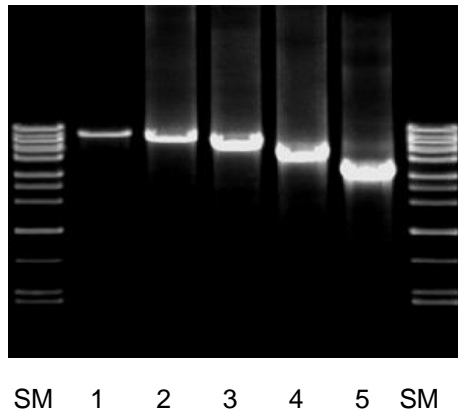
2.4. Bổ sung 2  $\mu\text{L}$ /1 tube dung dịch pha loãng của Taq pol.

2.5. Đặt E-tube đựng dung dịch phản ứng vào heating block của máy PCR để thực hiện chế độ nhiệt theo chu kỳ như sau:

- Start program
- 95°C/5 phút
- Thực hiện 30 chu kỳ: 95°C/30 giây, 50°C/30 giây và 72°C/1 phút
- 72°C/10 phút
- Giữ sản phẩm PCR ở 4°C
- Chạy điện di để kiểm tra kết quả PCR (Hình 3.7)
- Nếu chưa chạy điện di thì bảo quản sản phẩm PCR ở -20°C

*Chú ý*

- Các thành phần và điều kiện của phản ứng như: Độ dài của primer, chế độ nhiệt trong một chu kỳ, số chu kỳ đối với mỗi đối tượng có thể thay đổi ít nhiều bằng thực nghiệm. Các thông số trên chỉ có ý nghĩa tham khảo.
- Nếu làm nhiều mẫu, có thể trộn chung các thành phần khác trừ DNA và Taq pol, sau đó chia ra từng tube rồi cho DNA và Taq pol vào sau cùng.
- Thao tác trong tủ cấy vô trùng (hoặc không).



**Hình 3.7. Điện di các sản phẩm PCR.** SM: Chuẩn kích thước DNA. Các đường số 1, 2, 3, 4 và 5: Các sản phẩm PCR khác nhau.

### III. Tối ưu hóa các điều kiện cho PCR

#### 1. Trình tự của primer

Đối với genome của eukaryote người ta thường dùng các primer dài khoảng 18 nucleotide trở lên. Tuy nhiên, cũng tùy từng trường hợp, có những primer của các chỉ thị phân tử ngẫu nhiên như RAPD<sup>1</sup> rất ngắn (10-16 mer) hoặc STS<sup>2</sup> cần primer dài hơn (20-24 mer). Nói chung, genomic DNA không hoàn toàn là một trình tự ngẫu nhiên mà nó còn mang đặc tính của các họ gen, những nhân tố có tính lặp lại (sự lặp đoạn đơn giản hay phức tạp). Tùy theo loài sinh vật và tùy theo cách thể hiện của trình tự DNA khuôn mẫu, người ta sẽ thiết kế trình tự của primer và đối chiếu cẩn thận nó với số liệu lưu trữ trước đây nhằm loại trừ các sai sót về kỹ thuật. Gần đây, người ta thường sử dụng phần mềm trong computer để thiết kế các primer thích hợp cho từng mục tiêu nghiên cứu và có thể giúp loại bỏ các cặp primer được thiết kế không tối ưu.

Khi thiết kế primer cần chú ý một số điểm như sau: Cố gắng chọn trình tự primer có khoảng 50% GC. Tránh G và C ở đầu 3' của primer vì nó có thể làm tăng cơ hội tạo ra hiện tượng primer-dimers (hai primer bắt cặp với nhau). Tránh chọn các vùng có trình tự tương đồng để hạn chế các primer tự gắn với nhau. Tính toán nhiệt độ nóng chảy ( $T_m$ ) của primer với tổng số 4°C cho GC và 2°C cho AT, sau đó trừ đi 5°C từ giá trị này và đây chính là nhiệt độ ủ ( $T_a$ ) của primer. Nói chung, nhiệt độ ủ có giá trị thấp hơn nhiệt độ nóng chảy của primer. Sự khác nhau từ 4-6°C giữa  $T_m$  primer và  $T_a$  dường như không ảnh hưởng đến hiệu suất của PCR.

## 2. Nhiệt độ của quá trình ủ

Nhiệt độ ủ thích hợp là nhiệt độ sao cho ở đó đó 1/2 số primer sẽ gắn với DNA khuôn mẫu. Công thức dưới đây ứng dụng trong trường hợp primer có khoảng 20 oligonucleotide:

$$T_a = 4(G + C) + 2(A + T) - 5^\circ\text{C} \quad (1)$$

Tuy nhiên, công thức này chỉ có tính tương đối, bằng kinh nghiệm nghiên cứu chúng ta mới có thể có số liệu đúng về nhiệt độ cho quá trình ủ. Số base của primer càng ít thì nhiệt độ này càng thấp và ngược lại.

---

<sup>1</sup> RAPD (random amplified polymorphic DNA): DNA đa hình được khuếch đại ngẫu nhiên.

<sup>2</sup> STS (sequence-tagged site): STS primer được thiết kế trên cơ sở các trình tự của RAPD markers.

### 3. *Magnesium*

Nồng độ của  $Mg^{2+}$  (được cung cấp dưới dạng  $MgCl_2$ ) có thể ảnh hưởng đến nhiệt độ biến tính DNA khuôn mẫu, quá trình ủ của primer, tính đặc hiệu của sản phẩm PCR, hoạt tính của Taq pol và độ chính xác của kết quả. Nồng độ thích hợp của  $Mg^{2+}$  là từ 0,5-2,5 mM ứng với nồng độ dNTP tổng số đã cho. Nồng độ  $Mg^{2+}$  cao sẽ làm cho phân tử DNA sợi đôi ổn định hơn và ngăn ngừa được sự biến chất hoàn toàn (mở xoắn để giải phóng sợi đơn của sản phẩm PCR trong mỗi chu kỳ) làm cho kết quả PCR nghèo đi. Mặt khác, nó còn làm cho hiện tượng bắt cặp giả (tại những vị trí không tương đồng) ổn định hơn dẫn đến xuất hiện những sản phẩm PCR không đặc hiệu với số lượng khá lớn. Ngược lại, nếu nồng độ  $Mg^{2+}$  quá thấp sẽ ảnh hưởng xấu đến quá trình tổng hợp DNA ( $Mg^{2+}$  đóng vai trò là một co-factor của Taq pol). Do đó, cần phải xác định nồng độ tối ưu của  $Mg^{2+}$  nhằm đảm bảo hiệu suất khuếch đại và tính đặc hiệu của sản phẩm PCR.

### 4. *Các deoxyribonucleotide triphosphate (dNTPs)*

Dung dịch stock của các dNTP phải có pH 7, nồng độ thường dùng là 10 mM (2,5 mM mỗi loại) được bảo quản ở  $-20^{\circ}C$ . Hàm lượng của các dNTP trong khoảng 20-200  $\mu M$  cho kết quả ổn định, chính xác và đặc hiệu. Bốn loại dNTP được sử dụng cần có nồng độ tương đương nhau để giảm thiểu tối đa hiện tượng sai biệt do kết hợp sai (misincorporation) mã di truyền nào đó.

### 5. *Enzyme Taq pol*

Nồng độ Taq pol thích hợp là từ 1-2,5 unit cho 100  $\mu L$  dung dịch phản ứng. Thông thường có thể sử dụng 0,5 unit/25  $\mu L$ . Nếu nồng độ Taq pol quá cao, có thể xuất hiện các sản phẩm PCR không đặc hiệu và làm sai lệch kết quả. Nếu nồng độ Taq pol quá thấp, sẽ không đủ lượng enzyme để xúc tác tạo ra sản phẩm PCR mong muốn.

### 6. *Đệm ổn định hoạt động của Taq pol*

Những chất ổn định Taq pol được sử dụng là gelatin và Triston X-100 trong đệm bảo quản enzyme. Nồng độ thường dùng là 0,01% gelatin và 0,1% (v/v) Triston X-100.

## 7. *Nồng độ các primer*

Trong hầu hết các ứng dụng của PCR, hai primer F và R đều phải có nồng độ bằng nhau. Nồng độ cuối cùng của mỗi primer là 0,1  $\mu\text{M}$ , tương đương với 2,5 pmol (16,25 ng của 20-mer) trong 25  $\mu\text{L}$  dung dịch phản ứng. Sử dụng nồng độ primer cao không cần thiết và thường tạo ra bất lợi, vì quá thừa primer sẽ có hiện tượng primer-dimers và hiện tượng gắn primer nhầm vị trí trên khuôn mẫu.

## 8. *Nồng độ DNA khuôn mẫu*

PCR bao gồm hai giai đoạn: giai đoạn sàng lọc và giai đoạn khuếch đại. Nếu các chất trong thành phần phản ứng (bao gồm primer) có nồng độ cao trong khi DNA khuôn mẫu lại có nồng độ thấp thì giai đoạn sàng lọc của PCR sẽ trở nên khó khăn hơn, vì tần suất để primer và DNA khuôn mẫu gặp nhau giảm rõ rệt, trong khi sự tiếp xúc giữa primer và primer lại tăng lên gây ra hiện tượng primer-dimers trong giai đoạn khuếch đại. Thông thường, nồng độ DNA khuôn mẫu được sử dụng trong khoảng từ 10-100 ng/25  $\mu\text{L}$  dung dịch phản ứng.

## 9. *Tỷ lệ giữa primer và DNA khuôn mẫu*

Một trong những yếu tố quan trọng nhất của PCR là tỷ lệ tối ưu giữa primer và DNA khuôn mẫu. Nếu tỷ lệ này quá cao thì hiện tượng primer-dimers sẽ xuất hiện, giống như trường hợp mẫu DNA quá loãng. Nếu tỷ lệ này quá thấp kết quả sản phẩm PCR sẽ không nhiều. Trong hầu hết các trường hợp áp dụng PCR, người ta đều dùng nồng độ primer không quá 0,5  $\mu\text{M}$  (12,5 pmol/25  $\mu\text{L}$ ) và tính toán nồng độ DNA khuôn mẫu để tránh hiện tượng primer-dimers.

## 10. *Khởi động nóng PCR*

Khởi động nóng PCR (“hot-start” PCR) là phương pháp sản xuất các sản phẩm PCR sạch hơn. DNA khuôn mẫu và primer được trộn với nhau và giữ ở nhiệt độ trên ngưỡng của liên kết không đặc hiệu giữa primer và

khuôn mẫu. Tất cả các thành phần của PCR được bổ sung trước ngoại trừ một chất then chốt (thường là Taq pol). Chỉ trước khi đi vào chu kỳ khuếch đại, thành phần chưa có mới được bổ sung để cho phép phản ứng xảy ra ở nhiệt độ cao hơn. Do không có hiện tượng lai không đặc hiệu của primer với khuôn mẫu, nên các băng DNA được khuếch đại trở nên sáng hơn, các đoạn primer không có cơ hội để gắn với nhau. Kỹ thuật này được tiến hành bằng cách làm lạnh các tube trên đá tuyết (ice bath) trong khi bổ sung hỗn hợp thành phần của PCR. Sau đó, đặt tube trong máy PCR đã được làm nóng ngay trước khi bổ sung thành phần cuối cùng. Dưới đây là ví dụ minh họa cho khởi động nóng PCR.

10.1. Chuẩn bị dung dịch master mix và cho vào E-tube loại 0,5 mL và trộn đều các thành phần sau:

Đệm 10× PCR	4,5 μL
Hỗn hợp 4 dNTP (2,5 mM mỗi loại)	4 μL
Primer-F (10 pmol/μL)	2,5 μL
Primer-R (10 pmol/μL)	2,5 μL
Thêm H <sub>2</sub> O tới	40 μL

10.2. Chuẩn bị dung dịch pha loãng của Taq pol:

Đệm ×10	2 μL
Taq pol (5 units/μL)	1 μL
Thêm H <sub>2</sub> O tới	20 μL

10.3. Bổ sung 5 μL DNA khuôn mẫu (100 ng) vào 40 μL dung dịch của master mix tube.

10.4. Đặt E-tube đựng dung dịch phản ứng vào heat block của máy PCR để thực hiện chế độ nhiệt theo chu kỳ như sau:

- Start program
- “Hot start”: Khoảng 80°C/10 giây «pause»
- Bổ sung 2 μL/1 tube dung dịch pha loãng của Taq pol.
- Restart
- 94°C/5 phút

- Thực hiện 30 chu kỳ: 94°C/30 giây, 54°C/30 phút và 72°C/1 phút.
- 72°C/10 phút
- Giữ sản phẩm PCR ở 4°C
- Chạy điện di để kiểm tra kết quả PCR
- Nếu chưa chạy điện di thì bảo quản sản phẩm PCR ở -20°C

#### IV. Thiết kế primer

Việc chọn lựa một primer có một ý nghĩa quan trọng khi muốn ứng dụng PCR có hiệu quả và độ chính xác cao. Vì thế, chúng ta phải có một thiết kế chính xác về oligonucleotide primer. Trình tự của primer được chọn xác định kích thước và vị trí của sản phẩm PCR, cũng như  $T_m$  của vùng được khuếch đại. Primer được thiết kế đúng có thể giúp chúng ta tránh tạo ra những sản phẩm PCR không đặc hiệu (non-specific).

Mục đích của việc thiết kế primer là có được một sự cân bằng giữa hai kết quả: sự đặc hiệu và hiệu suất khuếch đại. Sự đặc hiệu được xem xét trên cơ sở xuất hiện bao nhiêu lần hiện tượng gắn primer nhầm (mis-priming) trên tổng số lần gắn primer (priming) tức là sự lai giữa primer và khuôn mẫu tại vị trí đích của nó. Hiệu suất là sự tiếp cận với lý thuyết về kết quả sản phẩm, trong mỗi chu kỳ PCR, một cặp primer có thể khuếch đại ra một sản phẩm. Với một trình tự DNA đã được cung cấp người ta có thể phân tích primer trên computer để có một kết quả cân bằng giữa hai mục tiêu nói trên.

##### 1. Chiều dài primer

Tính đặc hiệu của sản phẩm PCR thường phụ thuộc vào chiều dài của primer và nhiệt độ ủ. Các oligonucleotide có kích thước khoảng 18-24 base có xu hướng trở thành những trình tự rất đặc hiệu nếu nhiệt độ ủ của PCR được thiết kế sai khác chỉ vài độ so với  $T_m$  của primer (nhiệt độ lý thuyết). Primer có chiều dài càng lớn thì khả năng gắn của primer càng nhỏ. Trong khi khuếch đại, một sự cố nào đó của quá trình ủ cũng sẽ làm giảm kết quả PCR. Để tối ưu hóa PCR, người ta sử dụng primer có chiều dài tối thiểu sao cho đảm bảo  $T_m$  khoảng 54°C hoặc hơn chút ít, điều này sẽ cung cấp những cơ hội tốt nhất để duy trì tính đặc trưng của sản phẩm PCR và hiệu suất phản



ứng cao. Những oligonucleotide ngắn (15 base hoặc ngắn hơn) được sử dụng hạn chế trong nhiều quy trình PCR. Chiều dài primer tối thiểu cho hầu hết các ứng dụng PCR là khoảng 18 nucleotide và người ta thường thiết kế các primer có độ dài 18-24 mer. Primer càng ngắn thì quá trình gắn giữa primer và khuôn mẫu DNA càng nhanh, tạo thành sợi đôi ổn định trong tổng hợp DNA. Thông thường, các primer có 28-35 mer cần cho việc khuếch đại các trình tự có mức độ dị hợp cao.

Đối với những primer ngắn hơn 20 mer, có thể sử dụng công thức tính  $T_m$  liên hệ với các base:  $T_m = 4(G + C) + 2(A + T)$ . Trong khi primer dài hơn công thức này chỉ có giá trị gần đúng.

## 2. Nucleotide cuối cùng của primer

Vị trí đầu 3' của primer nên được xác định cẩn thận, vì nó có tính chất quyết định cho sự thành công của PCR. Khi chúng ta xác định được một amino acid, hai base đầu tiên trong codon của nó hoặc ba base trong trường hợp nó được mã hóa bằng codon đơn (methionin và tryptophan) thì hai hoặc ba base đầu tiên này được dùng như đầu 3'. Những cặp base hoàn chỉnh giữa những đầu 3' của primer và khuôn mẫu cho phép chúng ta có được kết quả mong muốn, và giảm thiểu tối đa hiện tượng bắt cặp sai (mismatch) trong khoảng 5-6 nucleotide tại đầu 3' của primer. Thông thường, việc thêm vào một trình tự không liên quan tại đầu 5' của primer vẫn không làm thay đổi quá trình gắn môi.

Nhiệm vụ của đầu 3' trong primer là kiểm soát hiện tượng gắn primer nhầm và ngăn cản hiện tượng tương đồng trong cùng một cặp primer. Nếu không thận trọng trong việc xác định đầu 3' thì hiện tượng primer-dimers sẽ xảy ra, và với tính chất bổ sung cho nhau sản phẩm PCR lúc bấy giờ sẽ là một sự khuếch đại của chính primer chứ không phải của DNA khuôn mẫu mà ta mong muốn. Trong trường hợp có nhiều cặp primer đưa vào trong cùng một phản ứng (multiplex PCR), chúng ta cần phải kiểm tra gấp đôi hiện tượng bổ sung cho nhau của tất cả primer.

## 3. Hàm lượng GC và nhiệt độ nóng chảy $T_m$

Primer của PCR phải duy trì được hàm lượng GC đến mức có thể được. Những oligonucleotide có 20 base với 50% GC thường có giá trị  $T_m$

trong khoảng 56-62°C sẽ tạo điều kiện để quá trình ủ đạt kết quả tốt. Hàm lượng GC và  $T_m$  phải khớp với một cặp primer được thiết kế. Nếu giá trị  $T_m$  càng lớn thì cơ hội cho hiện tượng gắn primer nhằm càng cao. Do đó, khi thiết kế primer cần phải lưu ý đặc biệt đến hàm lượng GC.

## V. Kỹ thuật RT-PCR

Phản ứng RT-PCR là phản ứng khuếch đại một đoạn khuôn mẫu RNA theo nguyên lý của PCR, bao gồm hai giai đoạn:

**Giai đoạn thứ nhất.** Phiên mã ngược khuôn mẫu RNA thành sợi DNA thứ nhất, sau đó dùng sợi này làm khuôn mẫu để tổng hợp sợi DNA thứ hai (tương tự như giai đoạn thứ nhất và thứ hai của quá trình tổng hợp cDNA-xem chương 6).

**Giai đoạn thứ hai.** Dùng DNA sợi đôi làm khuôn mẫu để thực hiện phản ứng PCR như đã trình bày ở trên.

RNA được cấu tạo nhờ bốn loại nucleotide (adenine, guanine, cytosine và uracil) liên kết với nhau tạo thành một chuỗi đơn. Khi chuỗi nucleotide này được biến đổi thành DNA thì uracil (U) được thay thế bằng thymine (T). Phản ứng biến đổi RNA hệ gen thành sợi DNA thứ nhất phải nhờ đến một enzyme phiên mã ngược (reverse transcriptase) do vậy giai đoạn này được gọi là phiên mã ngược (RT), sau đó quá trình tổng hợp sợi DNA thứ hai lại nhờ một enzyme khác là DNA polymerase I (thường dùng đoạn Klenow). Khi đã có DNA sợi đôi từ khuôn mẫu RNA thì phản ứng tiếp theo sẽ là khuếch đại gen nhờ kỹ thuật PCR được thực hiện với ba mức nhiệt độ phù hợp ở mỗi chu kỳ. Toàn bộ phản ứng khuếch đại một đoạn DNA từ khuôn mẫu RNA trải qua hai giai đoạn nói trên được gọi là RT-PCR.

Do trong tự nhiên một số virus có hệ gen là RNA (retrovirus) nên muốn khuếch đại một đoạn gen của nó thì người ta phải dùng kỹ thuật RT-PCR. Một số nghiên cứu khác về mRNA cũng cần đến RT-PCR để biến đổi sợi RNA thành DNA cho quá trình tạo dòng để phân tích trình tự gen (gene sequencing) hay tái tổ hợp (recombination) trong nghiên cứu biểu hiện gen.

## VI. Real-time PCR

### 1. Giới thiệu chung

Trong PCR truyền thống, sản phẩm khuếch đại được phát hiện nhờ sự phân tích đầu cuối (end-point) bằng cách chạy điện di DNA trên agarose gel sau khi phản ứng kết thúc. Ngược lại, phương pháp real-time PCR cho phép phát hiện và định lượng sản phẩm khuếch đại khi tiến trình phản ứng đang diễn ra dựa trên cơ sở phản ứng huỳnh quang, trong đó sự tăng lên về số lượng DNA tương ứng với sự tăng lên của tín hiệu huỳnh quang.

Trong real-time PCR, người ta thường sử dụng các loại thuốc nhuộm liên kết DNA sợi đôi (SYBR Green) để phát huỳnh quang, và các probe hoặc primer đặc hiệu chuỗi (trình tự) được đánh dấu huỳnh quang (TaqMan PCR). Thiết bị ổn nhiệt chu kỳ (máy PCR) đặc biệt được gắn với một module phát hiện tín hiệu huỳnh quang để theo dõi tiến trình phản ứng khi sự khuếch đại xảy ra. Tín hiệu huỳnh quang đo được phản ánh số lượng sản phẩm được khuếch đại trong mỗi chu kỳ.

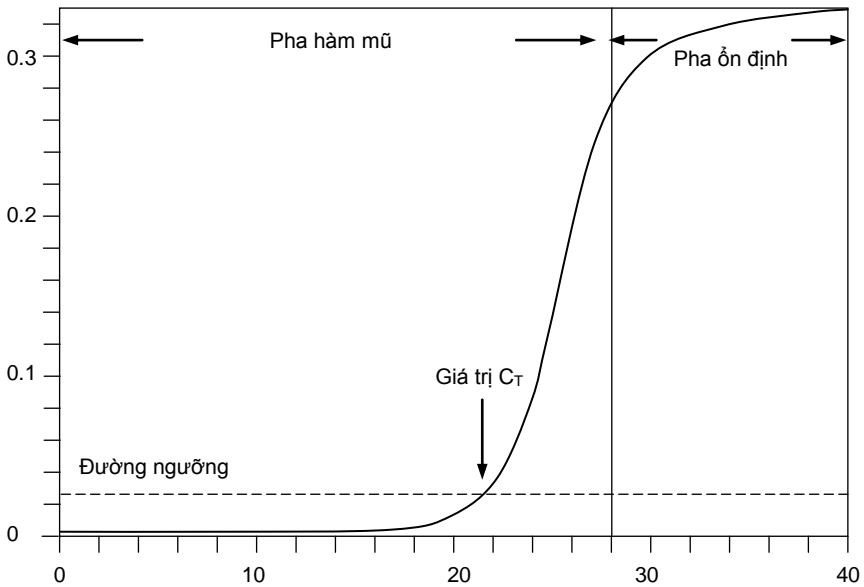
Real-time PCR có ưu điểm chính so với PCR truyền thống là cho phép chúng ta định lượng số copy khởi đầu của khuôn mẫu với độ chính xác và độ nhạy cao trên một phạm vi động học lớn. Các kết quả của real-time PCR có thể là chất lượng (sự có mặt hoặc không của trình tự đích) hoặc định lượng (số bản copy của DNA). Real-time PCR định lượng còn được biết như là qPCR. Số liệu của real-time PCR có thể đánh giá mà không cần điện di gel, thời gian thí nghiệm được rút ngắn và số lượng nguyên liệu đưa vào quá trình được tăng lên. Cuối cùng, do các phản ứng được chạy và số liệu được đánh giá trong một hệ thống tube đóng kín, nên các cơ hội nhiễm bẩn được giảm thiểu và sự cần thiết của các thao tác hậu khuếch đại được loại bỏ.

## *2. Phân tích tổng quát một phản ứng real-time PCR*

Để hiểu được real-time PCR như thế nào, chúng ta bắt đầu bằng một đường cong khuếch đại mẫu (Hình 3.9). Trong hình này, số chu kỳ PCR được trình bày trên trục x, và tín hiệu huỳnh quang từ phản ứng khuếch đại, tỷ lệ với số lượng sản phẩm được khuếch đại trong tube, được trình bày ở trục y.

Đường cong khuếch đại có 2 pha, một pha hàm mũ (exponential phase) được tiếp theo bởi một pha ổn định (plateau phase). Trong suốt pha hàm mũ, số lượng sản phẩm PCR xấp xỉ gấp đôi trong mỗi chu kỳ. Tuy nhiên, khi phản ứng diễn ra thành phần phản ứng sẽ bị tiêu hao, cuối cùng

làm cho một hoặc nhiều thành phần bị hạn chế. Ở điểm này phản ứng chậm lại và đi vào pha ổn định (các chu kỳ 28-40, Hình 3.9).



**Hình 3.9. Đường cong khuếch đại.** Tín hiệu huỳnh quang được trình bày đã trừ đường nền (baseline).

Đầu tiên, tín hiệu huỳnh quang duy trì ở mức độ nền (background), và việc tăng tín hiệu huỳnh quang là không thể phát hiện được (các chu kỳ 1-18, Hình 3.9) cho dù sản phẩm tích lũy theo hàm mũ. Sau đó, sản phẩm khuếch đại tích lũy đủ để đạt tới một hiệu suất cho phép phát hiện được tín hiệu huỳnh quang. Số chu kỳ để đạt được hiệu suất này gọi là chu kỳ ngưỡng (threshold cycle,  $C_T$ ). Do giá trị  $C_T$  được đo trong pha hàm mũ khi các tác nhân phản ứng không bị hạn chế, nên real-time qPCR có thể được sử dụng để tính toán chính xác và tin cậy số lượng khởi đầu của khuôn mẫu hiện diện trong phản ứng.

$C_T$  được xác định chủ yếu bởi số lượng của khuôn mẫu hiện diện ở lúc bắt đầu phản ứng khuếch đại. Khi lượng khuôn mẫu nhiều thì chỉ cần một vài chu kỳ khuếch đại để tích lũy sản phẩm đủ cho một tín hiệu huỳnh quang ở trên background. Vì vậy, phản ứng sẽ có  $C_T$  thấp hơn hoặc sớm. Ngược lại, nếu lượng khuôn mẫu lúc bắt đầu phản ứng quá ít, thì số chu kỳ khuếch đại cần nhiều hơn để tín hiệu huỳnh quang tăng trên background. Vì vậy,

phản ứng sẽ có  $C_T$  cao hoặc muộn. Các dạng quan hệ này là cơ sở cho hướng định lượng của real-time PCR.

### 3. Một số ứng dụng chính của real-time PCR

- **Nghiên cứu biểu hiện gen.** Xác định sự hoạt động của gen và định lượng mức độ biểu hiện của nó thông qua RNA bằng kỹ thuật RT-PCR.

- **Chẩn đoán phân tử.** Nghiên cứu mức độ nhiễm virus và vi khuẩn để chẩn đoán các bệnh viêm nhiễm.

- **Xét nghiệm phân biệt allele.** Là phương pháp xét nghiệm đầu cuối được dùng để xác định kiểu gen của mẫu vật (đồng hợp tử: chỉ có allele 1 hoặc chỉ có allele 2, dị hợp tử: có cả hai allele 1 và 2) và phân biệt sự đa hình nucleotide đơn (single nucleotide polymorphism, SNP).

## VII. Một số ứng dụng của PCR

PCR có phạm vi ứng dụng rất rộng trong công nghệ sinh học hiện đại. Theo nguyên tắc trên, đã có nhiều bổ sung và cải tiến để dùng PCR cho nhiều mục đích khác nhau. Ở đây chỉ giới thiệu một số ứng dụng chính:

### - Trong nghiên cứu genome

- + Nhân bản phân tử bằng PCR
- + PCR tái tổ hợp
- + Kỹ thuật footprinting DNase I
- + Sàng lọc thư viện  $\lambda$ gt11
- + Multiplex PCR...

### - Trong y học

- + Phát hiện tế bào T/virus gây bệnh ung thư máu
- + Phát hiện virus viêm gan B
- + Phát hiện virus Dengue gây bệnh sốt xuất huyết
- + Phát hiện vi khuẩn lao...

## 1. Sử dụng PCR để tạo nhanh các gen tổng hợp

Sử dụng phương pháp PCR hai giai đoạn (two-step PCR) để tạo nhanh các gen tổng hợp. Phương pháp này được xây dựng trên cơ sở những nucleotide chồng lên nhau (overlapping), do đó có thể được sử dụng để tạo ra đoạn DNA tổng hợp thông qua nhiều vòng khuếch đại PCR. Khả năng tạo ra DNA tổng hợp có nhiều ứng dụng tùy thuộc vào thiết kế của chúng ta. Người ta có thể thiết kế một gen tổng hợp có chứa codon được dùng để biểu hiện protein trong cơ thể sinh vật. Phương pháp này được hoàn tất nhờ sự giải mã ngược chuỗi amino acid của protein mong muốn. Để thay đổi cấu trúc của codon, gen tổng hợp có thể được thiết kế có những vị trí cắt hạn chế thuận lợi phục vụ cho thí nghiệm tạo dòng về sau.

Nguyên tắc của PCR hai giai đoạn đó là hai phản ứng PCR xảy ra liên nhau, phản ứng thứ nhất của PCR phát sinh ra sợi khuôn mẫu tương ứng với gen tổng hợp và sau đó nó được khuếch đại trong phản ứng PCR lần thứ hai (Hình 3.10). Trước khi bắt đầu quá trình này, người ta phải thiết kế cấu trúc và xác định trình tự nucleotide của gen tổng hợp mà mình mong muốn. Sau đó, trình tự các oligonucleotide (primers) theo chiều dài của gen phải được thiết kế và tổng hợp. Thông thường, người ta tổng hợp các oligonucleotide theo số chẵn với chiều dài khoảng 16-30 nucleotide.

## 2. Tạo dòng cDNA bằng PCR

Kỹ thuật tạo dòng kinh điển cDNA đòi hỏi nhiều công sức về xây dựng thư viện để bảo quản bacteriophage hoặc plasmid, nó cần một khối lượng công việc chọn lọc một số lượng lớn các phage hoặc plasmid có tính chất tái tổ hợp. Có ba hạn chế trong phương pháp này:

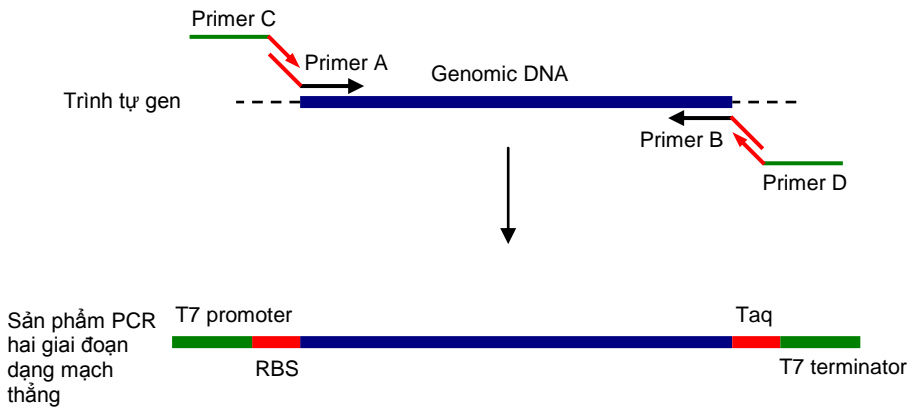
- Cần có một lượng cơ chất (ít nhất 1  $\mu\text{g}$ ) mRNA tinh sạch làm vật liệu khởi đầu để xây dựng thư viện với mức độ đa dạng đầy đủ cho nghiên cứu.

- Khó khăn thuộc về bản chất bên trong của những phản ứng enzyme tiếp nối nhau làm cho việc tạo dòng cDNA có kết quả thấp, và những dòng gen bị đứt đoạn.

- Việc sàng lọc một thư viện với kỹ thuật lai đòi hỏi nhiều thời gian.

Kỹ thuật PCR hiện nay có thể giúp chúng ta khắc phục những nhược điểm nói trên, làm cho việc tạo dòng cDNA trở nên dễ dàng hơn. Sử dụng

hai primer đặc hiệu của gen, một cDNA có trình tự đã được biết rõ, người ta cho khuếch đại trong PCR. Tuy nhiên, cái khó là làm sao phân lập được những bản sao của cDNA hoàn chỉnh (full-length) từ mRNA, trên cơ sở thông tin về trình tự rất hạn chế. Trình tự chưa biết này nằm kề một đoạn nhỏ của trình tự đã biết rõ. Trình tự chưa biết không thể khuếch đại bằng phương pháp PCR thông thường. Do đó, phương pháp PCR mở neo và phương pháp PCR đảo ngược đã được phát triển trong thời gian gần đây để giải quyết vấn đề này.



**Hình 3.10. Sơ đồ minh họa một phản ứng tạo nhanh gen tổng hợp thông qua phương pháp PCR hai giai đoạn.** Gen quan tâm được khuếch đại trong phản ứng PCR thứ nhất với cặp oligonucleotide đặc hiệu gen nhưng có đưa vào vùng chồng lên nhau (primer A và primer B). Trong phản ứng PCR thứ hai, các primer mở rộng C và D gắn với vùng chồng lên nhau của sản phẩm PCR lần thứ nhất và bổ sung tất cả các nhân tố điều hòa cần thiết cho sự phiên mã và dịch mã. Sản phẩm PCR lần thứ nhất được pha loãng 200 lần để làm khuôn mẫu cho phản ứng PCR thứ hai.

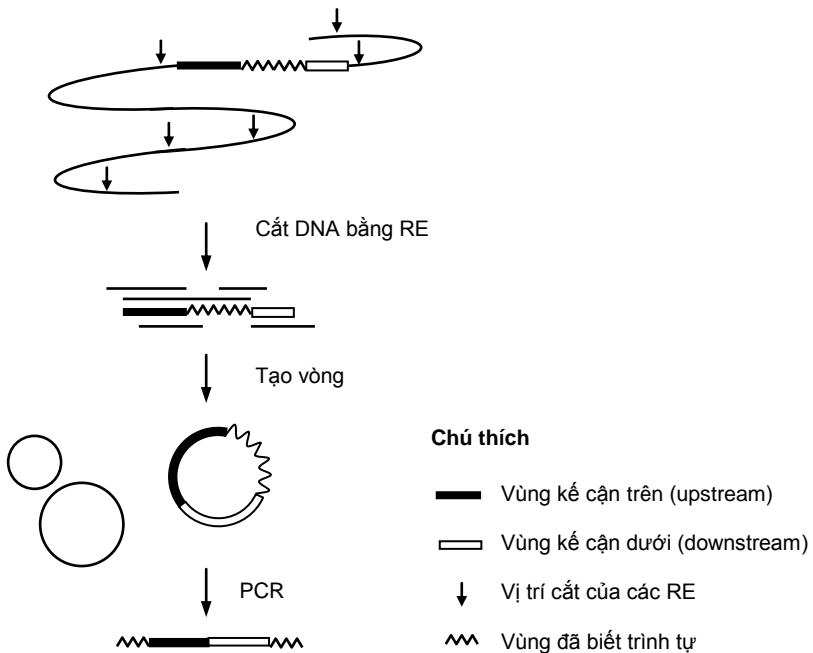
## 2.1. PCR mở neo

Kỹ thuật PCR mở neo có một điểm chung là: Tạo dòng DNA đi từ một đoạn trình tự đã biết rồi tới đoạn trình tự kế cận chưa biết nhờ sự giúp đỡ của một primer đặc hiệu đối với gen tại một đầu sợi đơn và một primer tổng

hợp tại một đầu sợi đơn khác. Do chỉ có primer đặc hiệu đối với gen mới có thể neo trong PCR làm dễ dàng hơn cho việc thu thập một lượng lớn sản phẩm khuếch đại không đặc hiệu so với PCR chuẩn. Kỹ thuật này khác với PCR chuẩn là chỉ cần một primer đặc hiệu cho phản ứng. Trong kỹ thuật PCR mở neo, một đuôi nhân tạo của các gốc dA được thêm vào đoạn cuối của DNA khuôn mẫu. Sự khuếch đại DNA xảy ra khi có primer của một trình tự đã biết trước và một primer thứ hai có các gốc oligo(dT).

## 2.2. PCR đảo ngược

Phương pháp PCR đảo ngược có thuận lợi lớn hơn nhờ khuếch đại trình tự kế cận chưa biết bằng hai primer đặc hiệu của gen. PCR đảo ngược hoạt động theo nguyên tắc tạo dòng sợi đơn một trình tự DNA. Sau đó, DNA này được cắt bằng RE ở vị trí tương ứng để tạo ra DNA đích, tiếp theo nó hoạt động theo chu trình có tính chất monomer. DNA này được khuếch đại lên bằng các primer tương đồng tại hai đầu sợi đơn của DNA có trình tự đã biết.



**Hình 3.11. Ứng dụng PCR đảo ngược.** Hai primer đặc hiệu được gắn vào trình tự đã biết và sau đó DNA được khuếch đại theo chiều mũi tên.



Ban đầu là sợi 5'→3' của mRNA cho vào phản ứng phiên mã ngược thành sợi đơn cDNA 3'→5', tiếp tục tổng hợp sợi đơn của cDNA lần thứ hai, cho ra sợi đôi cDNA (5'→3' và 3'→5'). Sợi này có chứa một trình tự đã biết rõ. Kỹ thuật tạo vòng làm cho cDNA thẳng biến thành cDNA vòng sợi đôi. Tiếp đến, mở vòng nhờ RE cắt trình tự được biết làm đôi, một nửa ở đầu sợi thẳng bên này, một nửa bên kia. Kéo thẳng sợi DNA và khuếch đại trong PCR (Hình 3.11).

### 3. Ứng dụng trong di truyền

PCR giúp đắc lực cho việc lập bản đồ gen (gene mapping) của sinh vật. Phương pháp này có tên là phân tích DNA đa hình được khuếch đại ngẫu nhiên (RAPD). Trong phương pháp RAPD, PCR được thực hiện với các primer chọn ngẫu nhiên. Kết quả là tạo ra một số đoạn DNA có chiều dài khác nhau đặc trưng cho mỗi loài, thậm chí mỗi cá thể. So sánh điện di sản phẩm PCR của nhiều giống có đặc điểm khác nhau trong cùng loài, với nhiều primer khác nhau có thể giúp xác định bản đồ gen của sinh vật.

Hiện nay, ngoài RAPD còn có nhiều loại chỉ thị phân tử (molecular marker) khác đã được phát triển như đa hình chiều dài các đoạn cắt hạn chế (restriction fragment length polymorphism, RFLP), vi vệ tinh (microsatellite) hay còn gọi là sự lặp lại các đoạn nucleotide đơn giản (simple sequence repeats, SSR), và đa hình chiều dài đoạn DNA được khuếch đại chọn lọc (amplified fragment length polymorphism, AFLP). Các kỹ thuật này được xây dựng trên cơ sở PCR (ngoại trừ RFLP) có triển vọng rất lớn trong nghiên cứu đa dạng di truyền (genetic diversity), định vị gen (gene location) và lập bản đồ gen do chúng đơn giản về mặt kỹ thuật, tiết kiệm thời gian và có hiệu quả cao. PCR hiện nay còn được dùng thay thế cho điện di isozyme để phân loại thực vật, xác định mối quan hệ của chúng một cách chính xác.

### 4. Ứng dụng trong chẩn đoán lâm sàng

Trước đây, kỹ thuật phân tích RFLP thường được sử dụng để xác định các đột biến dẫn tới các bệnh di truyền ở người. Tuy nhiên, kỹ thuật này chỉ áp dụng được trong trường hợp đột biến đó dẫn đến sự thay đổi trình tự có thể phát hiện được trên cơ sở độ dài của đoạn DNA. Ngược lại, trường hợp

một số đột biến gây ra bệnh di truyền nhưng không tạo ra các đoạn DNA cắt hạn chế khác nhau khi phân tích RFLP thì chỉ có thể phát hiện nếu xác định được trình tự đoạn DNA chứa điểm đột biến. Như vậy, chúng ta buộc phải xây dựng thư viện hệ gen (xem chương 5) cho mỗi người, sau đó tạo dòng và xác định trình tự nucleotide của dòng gen đột biến, vì thế phương pháp này đòi hỏi tốn nhiều thời gian.

Kỹ thuật PCR cho phép đưa ra một chọn lựa khác đơn giản hơn cho phép thu nhận thông tin về trình tự gen rất nhanh bằng cách khuếch đại đoạn gen chứa điểm đột biến và sau đó phân tích trực tiếp các sản phẩm PCR thu được. Khả năng nhận dạng nhanh các đột biến không chỉ quan trọng trong chẩn đoán lâm sàng, mà còn đẩy nhanh việc nghiên cứu các bệnh di truyền.

Độ nhạy cao của kỹ thuật PCR cho phép ứng dụng dễ dàng trong các chẩn đoán bệnh nhiễm trùng. Ví dụ: việc khuếch đại một số lượng lớn DNA virus trong các mẫu bệnh cho khả năng chẩn đoán bệnh sớm hơn trước khi triệu chứng bệnh bắt đầu xuất hiện. Điều này giúp cho thầy thuốc có phác đồ điều trị bệnh thích hợp, đặc biệt các bệnh ung thư do virus gây ra (ví dụ: ung thư vòm họng do papillomavirus người gây ra) và thông thường có kết quả hơn nếu như bắt đầu điều trị ở giai đoạn sớm.

## **Tài liệu tham khảo/đọc thêm**

**1. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K.** 2002. Short Protocol in Molecular Biology. Vol 1 and 2. 5<sup>th</sup> ed. *John Wiley & Sons, Inc.* USA.

**2. Chen BY and Janes HW.** 2002. PCR Cloning Protocols. 2<sup>nd</sup> ed. *Humana Press Inc.* New Jersey, USA.

**3. Dieffenbach CW and Dveksler GS.** 2003. PCR Primer: A Laboratory Manual. 2<sup>nd</sup> Edition. *Cold Spring Harbor Laboratory Press*, New York, USA.

**4. Innis MA, Gelfand DH, Sninsky JJ and White TJ.** 1990. PCR Protocols: A Guide to Methods and Applications. *Academic Press*, New York, USA.

**5. Maniatis T, Fritsch EF and Sambrook J.** 1989. Molecular Cloning-A Laboratory Manual. *Cold Spring Harbor Laboratory Press*, USA.

**6. Rapley R and Walker JM.** 1998. Molecular Biomethods Handbook. *Humana Press Inc.* New Jersey, USA.

**7. Surzycki S.** 2000. Basic Techniques in Molecular Biology. *Springer-Verlag*, Berlin, Heidelberg, Germany.

## Chương 4

# Các hệ thống vector

Có ba nhóm vector chính được sử dụng trong tạo dòng DNA và nhân bản chúng trong tế bào vật chủ (*E. coli* hoặc nấm men), bao gồm: 1) Nhóm plasmid, 2) Nhóm phage/phagemid, và 3) Nhóm nhiễm sắc thể nhân tạo (artificial chromosome: BAC và YAC). Ý tưởng về vector chuyển gen bắt nguồn từ các plasmid của vi khuẩn. Vector chuyển gen là phân tử DNA có khả năng tự tái sinh, tồn tại độc lập trong tế bào và mang được các gen cần chuyển. Các vector chuyển gen phải thỏa mãn các yêu cầu tối thiểu sau:

- Có trình tự khởi đầu sao chép (*ori*) để có thể tự sao chép mà tồn tại độc lập.

- Có các vị trí nhận biết đơn (single recognition sites-SRS), nơi mà các enzyme hạn chế nhận biết để cắt rời làm chỗ lắp ráp đoạn gen ngoại lai vào. Các trình tự này thường nằm xa điểm khởi đầu sao chép để tránh bị cắt nhầm.

- Có trình tự khởi động (promoter) để tiến hành phiên mã gen ngoại lai.

- Đảm bảo sự di truyền bền vững của DNA tái tổ hợp ở dạng độc lập hay gắn vào nhiễm sắc thể của tế bào vật chủ.

- Có các gen chỉ thị (marker genes) để dễ dàng phát hiện ra thể tái tổ hợp.

Ngoài ra, chúng còn phải có những đặc tính bổ sung khác để cho việc tạo dòng dễ thực hiện như là:

- Chứa các gen làm vô hiệu hóa đoạn DNA không mong muốn được gắn vào.

- Có nhiều bản sao để được tách ra khỏi tế bào với số lượng lớn và đảm bảo sự khuếch đại của gen ngoại lai.

- Ngoài promoter cần có thêm các trình tự nucleotide khác cần thiết cho sự biểu hiện của gen, chẳng hạn như: RBS (ribosome binding sites-vùng liên kết với ribosome để dịch mã).

Giá trị của vector ở chỗ nó được thiết kế sao cho thật thuận tiện với các mục đích sử dụng khác nhau. Không có vector toàn năng cho chuyển gen, mà cần có sự chọn lựa tùy mục đích và kích thước của đoạn gen được tạo dòng. Chúng được thiết kế với nhiều chức năng chuyên biệt để mang được các trình tự nucleotide như mong muốn. Các vector chuyển gen có năm ứng dụng quan trọng chủ yếu:

- Tạo dòng (cloning) và khuếch đại (amplification) một đoạn trình tự của DNA (nhiều bản sao giống nhau).

- Nghiên cứu sự biểu hiện của một đoạn trình tự DNA.

- Đưa gen vào các tế bào vi sinh vật hay các động-thực vật.

- Sản xuất RNA.

- Sản xuất protein từ gen được tạo dòng.

Chương này giới thiệu sáu loại vector phổ biến (thuộc ba nhóm vector: plasmid, phage/phagemid và nhiễm sắc thể nhân tạo) dùng trong tạo dòng gen (gene cloning): plasmid, bacteriophage  $\lambda$ , cosmid, bacteriophage M13, BAC và YAC.

## **I. Plasmid vector**

### *1. Đặc điểm của plasmid*

Plasmid là các thông tin di truyền ngoài nhân, được tìm thấy trong nhiều loài vi khuẩn khác nhau. Chúng là những phân tử DNA mạch vòng, sợi đôi, kích thước từ 1-  $\geq$  200 kb, có khả năng tự sao chép để tồn tại độc lập trong tế bào. Các plasmid có thể mang một số gen của vi khuẩn và các gen này có thể biểu hiện ra protein. Một số kiểu hình khác nhau của plasmid như:

- Kháng các chất kháng sinh

- Kháng kim loại nặng

- Sản xuất các chất kháng sinh

- Thủy phân các hợp chất hữu cơ phức tạp

- Sản xuất độc tố đường ruột enterotoxin

- Sản xuất chất diệt khuẩn bacteriocin

- Sản xuất các colicin<sup>1</sup>
- Sản xuất haemolysin<sup>2</sup>
- Sản xuất các enzyme sửa đổi<sup>3</sup> và enzyme hạn chế<sup>4</sup>.

Một loại vi khuẩn vật chủ thường chứa hai hoặc nhiều loại plasmid cùng tồn tại. Chúng có thể là các plasmid tương hợp (compatible plasmid) hoặc không tương hợp (incompatible plasmid). Người ta có thể chia plasmid làm hai loại dựa trên đặc điểm sinh sản và phương thức sống của chúng, đó là: plasmid tiếp hợp (conjugative plasmid) và plasmid không tiếp hợp (non-conjugative plasmid).

Trong tự nhiên, plasmid xâm nhập vào tế bào vật chủ nhờ sự tiếp hợp (conjugation) của vi khuẩn, nhưng trong nghiên cứu thực nghiệm các plasmid được chuyển nạp vào vi khuẩn bằng một quy trình nhân tạo gọi là biến nạp gen (gene transformation). Kiểu hình mới của thể nhận (recipient) do plasmid đem đến (Ví dụ: plasmid kháng kháng sinh) cho phép chọn lọc xác định các nhân tố biến nạp và duy trì plasmid trong quần thể vi khuẩn. Ở các vị trí cắt hạn chế của plasmid, DNA ngoại lai (foreign DNA) có thể chèn vào và được xác định với gen mã hóa cho các marker chọn lọc. Một trong các vector hay được sử dụng trước đây là pBR 322 mang hai gen kháng ampicillin (*Amp<sup>r</sup>*) và tetracycline (*Tet<sup>r</sup>*) cùng một số vị trí cắt hạn chế thuận lợi. Có hai dạng phân chia từ pBR 322 thích hợp cho sự tái bản là các vector pAT 153 và pXf 3.

Các plasmid kích thước nhỏ có số lượng bản sao lớn hơn và miễn cảm hơn với vi khuẩn. Tuy nhiên, việc giảm kích thước của plasmid có thể làm mất các vị trí tạo dòng (multiple cloning sites, MCS) hữu ích. Ví dụ: pXf 3 thiếu các vị trí *Bam*I và *Ava*I (các vị trí này có trong pBR 322 và pAT 153).

Một số plasmid không được sử dụng rộng rãi như pBR 322, mà chỉ được sử dụng trong các mục đích đặc biệt, ví dụ:

---

<sup>1</sup> Colicin: Một nhóm bất kỳ của các protein được sản xuất bởi các vi khuẩn khác nhau ức chế sinh trưởng hoặc giết chết các vi khuẩn khác.

<sup>2</sup> Haemolysin: dung huyết tố, chất gây tan máu, chất tiêu hồng cầu.

<sup>3</sup> Enzyme sửa đổi (modification enzymes): là các enzyme tác động qua lại với DNA để sửa đổi cấu trúc hoặc sửa chữa các tổn thương đã xảy ra với phân tử DNA.

<sup>4</sup> Enzyme hạn chế (restriction enzymes hay restriction endonuclease, RE): xem chương 1.

- **pMK 16**: chứa các vị trí nhận biết đơn *Sma*I và *Xho*I trong gen mã hóa tính kháng kanamycin.

- **pKC 7** và **pACYC 184**: mang các vị trí tạo dòng hữu ích và gen mã hóa tính kháng chloramphenicol.

- Các plasmid có kích thước lớn **pCR1** (11,4 kb) và **pSC 101** (9,9 kb): không được dùng thường xuyên trong các thí nghiệm tạo dòng.

Thông thường, sự sao chép DNA plasmid được tiến hành nhờ các enzyme tái bản nhiễm sắc thể vi khuẩn. Ở một số plasmid dạng stringent control<sup>5</sup> sự sao chép của chúng khá ít hoặc bị giới hạn, chỉ gấp đôi sự sao chép của tế bào vật chủ và chỉ một hoặc một số bản sao của chúng có mặt trong mỗi tế bào vi khuẩn. Các plasmid dạng relaxed control<sup>6</sup> có số lượng bản sao trong mỗi tế bào nhiều hơn stringent plasmid, khoảng 10-20 và có khi lên đến hàng ngàn bản sao nếu như sự tổng hợp protein bị dừng lại (Ví dụ: bằng cách xử lý chloramphenicol). Nếu thiếu sự tổng hợp protein thì sự sao chép của các relaxed plasmid được tiếp tục, trong khi sự sao chép của DNA nhiễm sắc thể và của các stringent plasmid bị dừng lại. Các plasmid dùng làm vector tạo dòng phải có một số tính chất sau:

- Kích thước tương đối nhỏ và phải tái bản trong dạng relaxed plasmid.

- Mang một hoặc nhiều marker chọn lọc (selectable marker) cho phép xác định các nhân tố biến nạp và duy trì plasmid trong quần thể vi khuẩn.

- Chứa các vị trí nhận biết đơn cho một hoặc nhiều enzyme hạn chế.

Đến nay, các plasmid đã được cải tiến để ngày càng thuận tiện hơn cho kỹ thuật tái tổ hợp DNA, trải qua ba thế hệ:

- **Thế hệ đầu tiên**. Là các plasmid tự nhiên đến nay hầu như không còn sử dụng nữa.

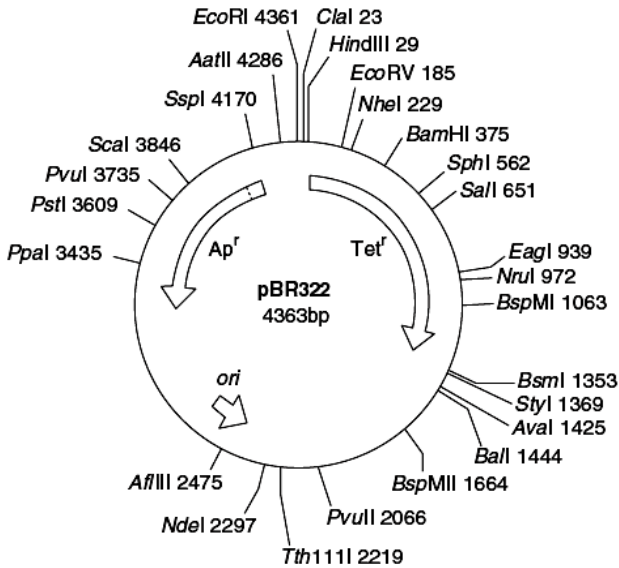
- **Thế hệ thứ hai**. Là các plasmid được cấu tạo phức tạp hơn. Một trong những plasmid thường được sử dụng là pBR 322 (Hình 4.1) có nguồn

---

<sup>5</sup> Stringent control: Còn gọi là kiểm soát sao chép chặt chẽ (stringent replication control), mô tả sự giới hạn nhân bản của các plasmid với sự sao chép của chromosome vi khuẩn.

<sup>6</sup> Relaxed control: Còn gọi là kiểm soát sao chép lỏng lẻo (relaxed replication control): liên quan đến khả năng của một số plasmid tiếp tục sao chép sau khi vi khuẩn ngừng phân chia.

gốc từ một plasmid nhỏ *ColE1*. Nó được thiết kế từ nhiều đoạn của các plasmid khác nhau để vừa có các gen kháng thuốc ( $Amp^r$  và  $Tet^r$ ), trình tự khởi đầu sao chép (*ori*), vừa có các vị trí nhận biết cho các enzyme hạn chế như *EcoRI*, *HindIII*, *BamHI*, *Sall*,... Ở đoạn gen  $Amp^r$  có bốn điểm nhận biết, ở gen  $Tet^r$  có tám điểm nhận biết, những điểm này giúp dễ phát hiện các plasmid có gen ngoại lai gắn vào. Ví dụ: nếu cắt plasmid bằng enzyme *BamHI* (375) rồi gắn DNA ngoại lai vào chỗ cắt thì gen  $Tet^r$  bị phân đôi do chèn đoạn DNA lạ, vì thế tế bào mất khả năng kháng tetracycline nhưng vẫn kháng ampicillin. Plasmid này có khả năng sao chép độc lập với tế bào *E. coli* và tồn tại với số lượng trung bình 20-30 bản sao cho mỗi tế bào. Trong những điều kiện nuôi cấy nhất định có thể khuếch đại có chọn lọc làm tăng số plasmid đến hơn 1.000 bản sao cho một tế bào.



**Hình 4.1. Plasmid vector pBR 322.**  $Ap^r$  (hay  $Amp^r$ ) và  $Tet^r$ : gen kháng ampicillin và tetracycline, *ori*: trình tự khởi đầu sao chép, và một số vị trí nhận biết cho các RE.

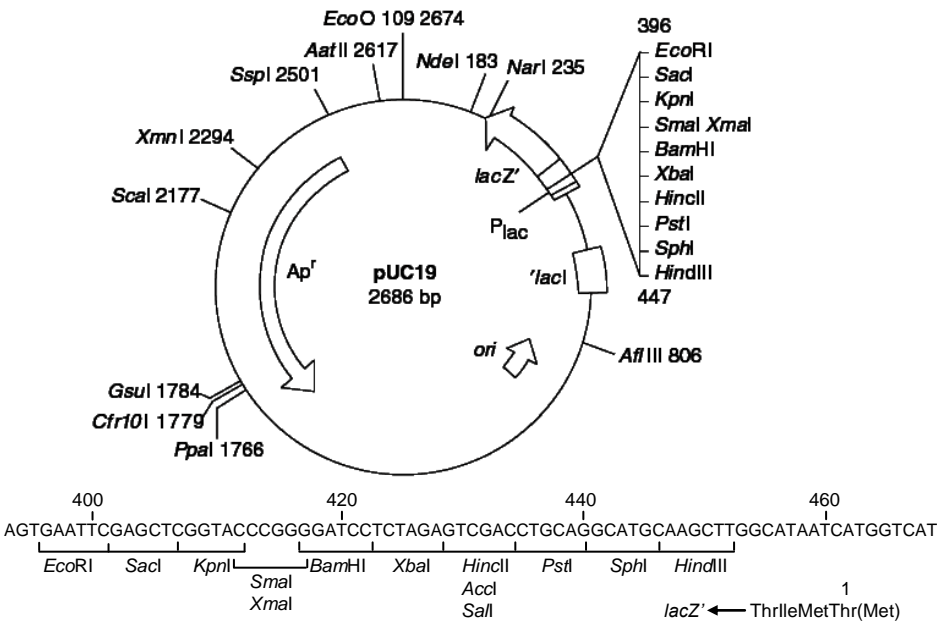
- **Thế hệ thứ ba.** Là các plasmid đa năng (polycloning plasmid) và chuyên dụng. Để tiện cho việc sử dụng nhiều loại RE khác nhau, nhiều trình tự nhận biết của chúng được xếp nối tiếp nhau thành một đoạn dài gọi là



polylinkers (vùng đa nối) hay multiple cloning sites (các vị trí tạo dòng). Các plasmid vi khuẩn có thể chứa đoạn DNA ngoại lai khoảng 3-10 kb. Có 3 nhóm plasmid thông dụng hiện đang được bán rộng rãi trên thị trường như:

+ **Nhóm các plasmid pUC.** Kích thước khoảng 2.600 bp, mang gen  $Ap^r$  và một phần gen  $lacZ'$ , xen vào giữa gen  $lacZ'$  là polylinker (Hình 4.2). Các thành viên của nhóm này chỉ khác nhau do độ dài và chiều của polylinker. Các ưu điểm của nhóm này:

- ◆ Kích thước nhỏ.
- ◆ Sự có mặt của gen  $lacZ'$ <sup>7</sup> rất thuận tiện cho việc phát hiện vector tái tổ hợp.

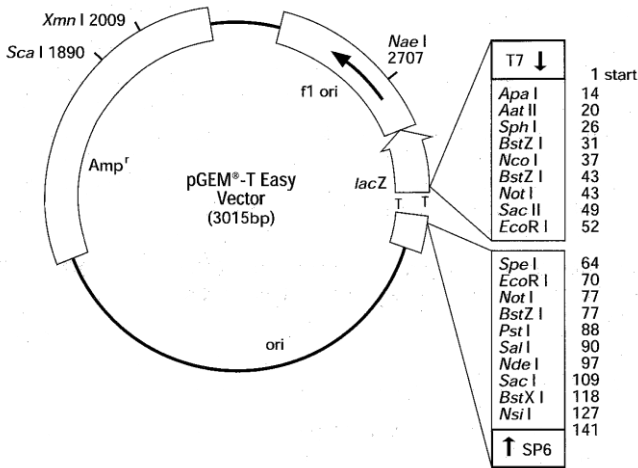


**Hình 4.2. Plasmid vector pUC19.** Vị trí tạo dòng (từ 396-447) được gắn vào gen  $lacZ'$ , nhưng không can thiệp vào chức năng của gen.

<sup>7</sup> Gen  $lacZ'$ : gen của *E. coli* mã hóa  $\beta$ -galactosidase thích hợp cho chọn lọc thể biến nạp bằng khuẩn lạc xanh ( $\beta$ -galactosidase sẽ kết hợp với IPTG và X-gal được bổ sung trong môi trường nuôi cấy) và khuẩn lạc trắng (đoạn DNA ngoại lai xen vào giữa gen  $lacZ'$  làm cho gen này mất hoạt tính vì thế không sản xuất được  $\beta$ -galactosidase).

♦ Vùng polylinker cho phép chèn vào gần như bất kỳ một trình tự DNA lạ nào.

+ **Nhóm các plasmid pGEM:** Kích thước khoảng 3.000 bp, mang gen *Amp<sup>r</sup>*, gen *lacZ*, polylinker và promoter đặc trưng cho các RNA polymerase (SP6, T7) ở hai bên vùng polylinker cho phép phiên mã đoạn DNA gắn trong vector thành nhiều RNA (Hình 4.3). Các RNA này thường dùng làm probe (mẫu dò) hay dùng trong nghiên cứu cấu trúc và chức năng của RNA.



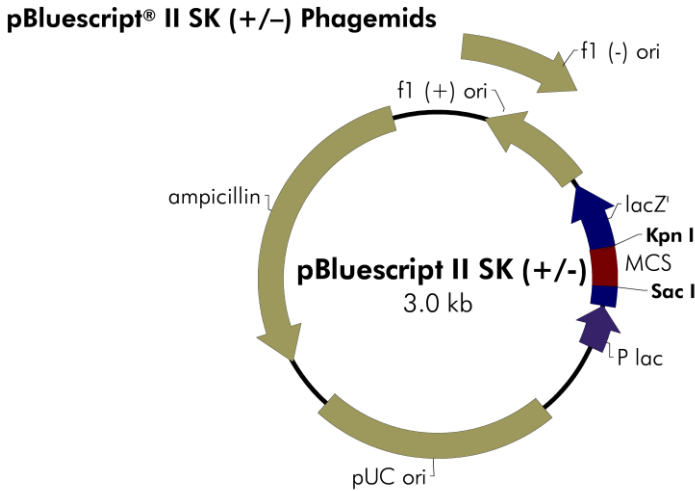
**Hình 4.3. Plasmid vector pGEM.** Nhóm vector này được mở vòng sẵn mang 2 đầu T ở vùng MCS, đặc trưng cho việc gắn các sản phẩm PCR mang 2 đầu A.

+ **Nhóm các plasmid pBluescript:** Kích thước khoảng 3.000 bp, các plasmid này kết hợp được tất cả những ưu điểm của mấy nhóm vừa kể và được xem là nhóm có tiềm năng ứng dụng lớn nhất hiện nay (Hình 4.4).

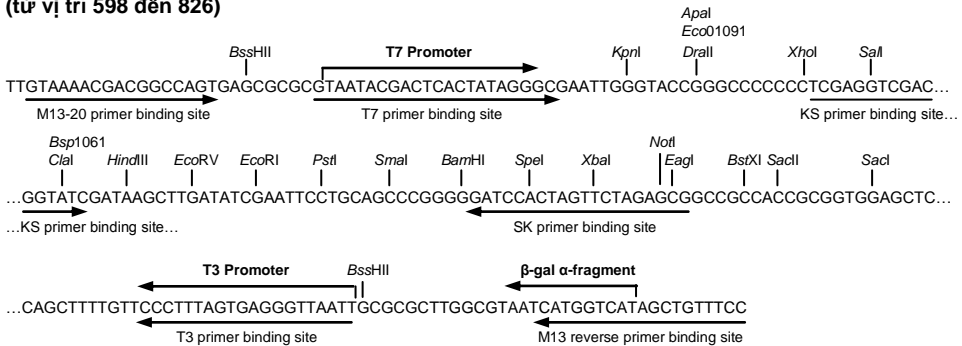
## 2. Tạo dòng trong plasmid

Về nguyên tắc, DNA plasmid bị cắt bởi enzyme hạn chế và gắn *in vitro* với đoạn DNA ngoại lai tạo ra các plasmid tái tổ hợp, sau đó chúng được dùng để biến nạp vào vi khuẩn. Khó khăn lớn nhất là phân biệt giữa các plasmid chứa các đoạn DNA ngoại lai-thể tái tổ hợp (recombinant) và các phân tử DNA vector đã tái tạo lại vòng (recircularization). Sự tái tạo lại

vòng của plasmid có thể được hạn chế bằng cách điều chỉnh nồng độ DNA ngoại lai và DNA vector trong phản ứng gắn (ligation). Một số phương thức được dùng để phân biệt giữa các thể tái tổ hợp và tái tạo lại vòng như sau:



**Vùng tạo dòng MCS của pBluescript II SK (+/-)**  
(từ vị trí 598 đến 826)

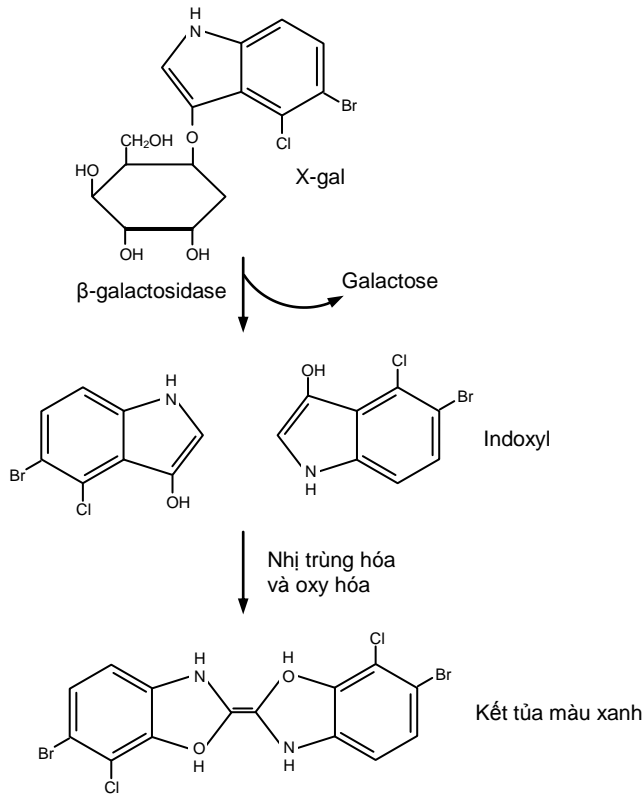


**Hình 4.4. Plasmid vector pBluescript II SK (+/-).** Vector này mang vùng tạo dòng ở vị trí từ 598-826 trên gen *lacZ'*, gen kháng ampicillin, các promoter T3 và T7, và các vị trí gắn cho các cặp primer khác nhau dùng trong phân tích trình tự đoạn DNA ngoại lai.

### 2.1. Khử hoạt tính bằng chèn đoạn (insertional inactivation)

Phương pháp này dùng cho các plasmid mang hai hoặc nhiều marker (ví dụ: *Amp<sup>r</sup>*, *lacZ'*). DNA ngoại lai và DNA plasmid được thủy phân cùng

một loại enzyme hạn chế và tinh sạch, sau đó gắn hai DNA với nhau, hỗn hợp gắn được dùng để biến nạp vào *E. coli* miễn cảm Amp để chọn lọc thể biến nạp nhờ tính kháng Amp của plasmid và hoạt tính  $\beta$ -galactosidase của gen *lacZ*. Các khuẩn lạc chứa các plasmid tái tổ hợp sinh trưởng trên môi trường có mặt Amp và isopropyl-thiogalactoside (IPTG) cùng với cơ chất nhiễm sắc thể X-gal sẽ có màu trắng do đoạn DNA ngoại lai xen vào giữa gen *lacZ* làm gen này mất hoạt tính. Trong khi đó các khuẩn lạc chứa DNA plasmid tái tạo lại vòng sẽ có màu xanh do gen *lacZ* không bị mất hoạt tính (Hình 4.5 và 4.6).

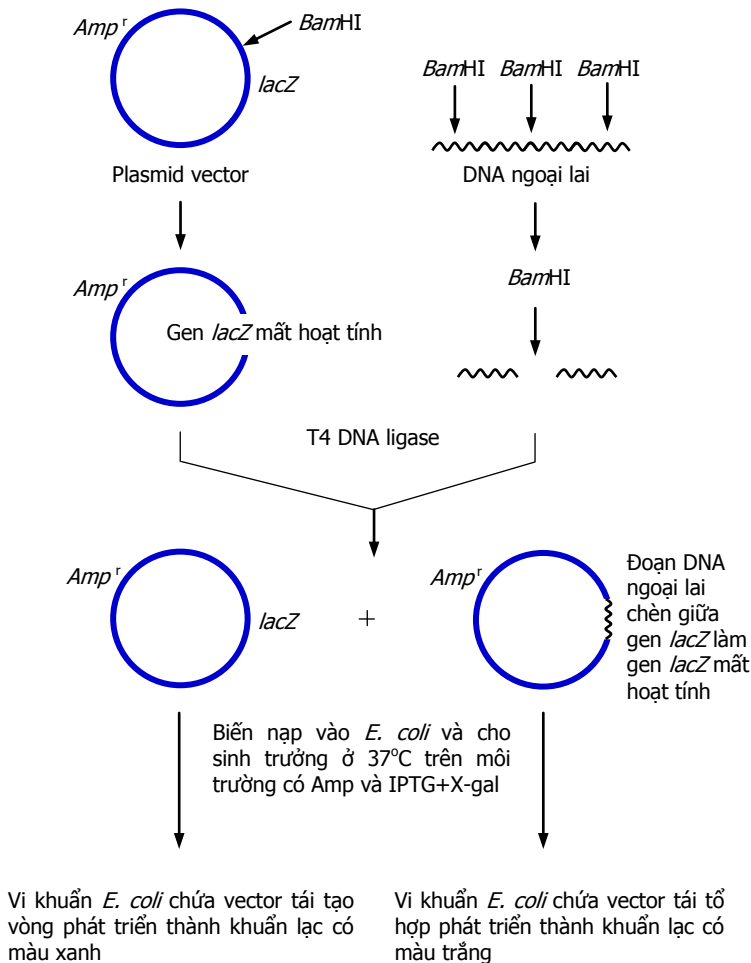


**Hình 4.5. Cơ chế tác dụng của  $\beta$ -galactosidase**

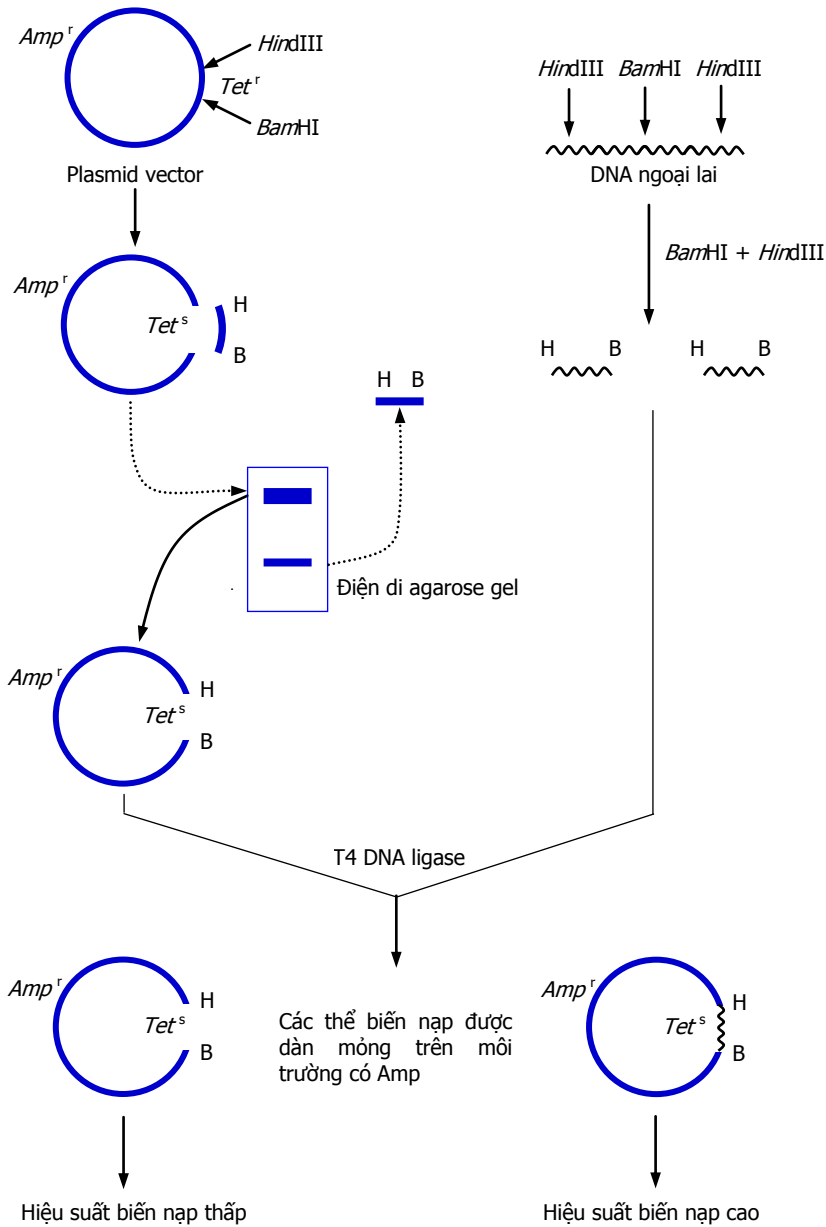
## 2.2. Tạo dòng định hướng (directional cloning)

Hầu hết các plasmid vector mang hai hoặc nhiều vị trí nhận biết enzyme hạn chế, ví dụ: vector pBR 322 chứa các vị trí nhận biết đơn *Hind*III và *Bam*HI sau khi cắt bằng hai enzyme hạn chế tương ứng, đoạn

DNA plasmid lớn hơn có thể được tinh sạch bằng điện di agarose gel và gắn với một đoạn DNA ngoại lai mang các đầu kết dính tương đồng với nó cũng được cắt bởi *Bam*HI và *Hind*III. Kết quả, thể tái tổ hợp dạng vòng mang tính kháng Amp sau đó được dùng để biến nạp vào *E. coli*. Do thiếu sự hỗ trợ giữa các đầu lồi *Hind*III và *Bam*HI nên đoạn vector lớn hơn không thể tái tạo lại vòng một cách hiệu quả được vì thế nó biến nạp vào *E. coli* rất kém. Dĩ nhiên, các tổ hợp khác của enzyme cũng có thể được sử dụng tùy thuộc vào các vị trí nhận biết (RS-recognition sites) trong vector và của đoạn DNA ngoại lai (Hình 4.7).



**Hình 4.6. Tạo dòng theo phương thức khử hoạt tính bằng chèn đoạn.** *Amp<sup>r</sup>*: gen kháng ampicillin, *lacZ*: gen mã hóa enzyme  $\beta$ -galactosidase.



**Hình 4.7. Tạo dòng định hướng.**  $Tet^r$ : gen kháng tetracycline,  $Tet^s$ : gen kháng tetracycline bị khuyết đoạn và mất hoạt tính.

### 2.3. Biến nạp vector tái tổ hợp vào vi khuẩn/tế bào vật chủ

Sau khi tạo được vector tái tổ hợp mang gen ngoại lai, việc tiếp theo là biến nạp nó vào tế bào vật chủ. Trong trường hợp này tế bào vật chủ thường

được sử dụng là vi khuẩn *E. coli* để khuếch đại một lượng lớn DNA plasmid tái tổ hợp dùng cho các phân tích về sau.

Hai phương pháp được dùng để biến nạp vector tái tổ hợp vào *E. coli* là điện biến nạp (electroporation transformation) và hóa biến nạp (chemical transformation).

### 2.3.1. Điện biến nạp

Đây là kỹ thuật hiệu quả nhất để biến nạp vi khuẩn. Hai thông số quan trọng của phương thức này là loại tế bào vi khuẩn và tần số xung điện cần thiết. Thể tích của dung dịch tế bào thường được dùng là 30  $\mu\text{L}$  (trùng ứng với nồng độ  $10^{10}$  tế bào *E. coli*/mL) có bổ sung 5 ng plasmid trong một cuvette có khoảng trống điện cực (electrode gap) 0,1 cm. Hiệu suất biến nạp của phương thức này lớn hơn  $1 \times 10^9$  thể biến nạp/ $\mu\text{g}$  plasmid siêu xoắn và khoảng  $1 \times 10^8$  thể biến nạp/ $\mu\text{g}$  plasmid được dùng trong phản ứng gắn. Tần số biến nạp khoảng 0,02 cho cả hai loại plasmid. Tần số biến nạp thấp đã ngăn cản được sự đồng biến nạp (co-transformation) vào vi khuẩn của hai hoặc nhiều phân tử plasmid.

Phương pháp điện biến nạp có một số ưu điểm sau:

- Hiệu suất biến nạp cao.
- Có thể dùng một thể tích dịch tế bào nhỏ. Thể tích dịch tế bào khoảng 20  $\mu\text{L}$  có thể cho hiệu suất khoảng  $10^9$  thể biến nạp.
- Phương pháp chuẩn bị tế bào biến nạp rất đơn giản, không sử dụng các kỹ thuật phức tạp và tốn thời gian. Hơn nữa, các tế bào dùng để biến nạp có thể được chuẩn bị trước và bảo quản vô hạn định mà không mất tính khả biến.
- Tần số điện biến nạp với DNA siêu xoắn và DNA mạch vòng là giống nhau. Do đó, không cần thiết phải dùng vector được tinh sạch cao trong các phản ứng gắn.
- Hiệu suất biến nạp phân tử cho DNA mạch vòng rất cao đối với các plasmid có kích thước lên đến 50 kb.

Tuy nhiên, nhược điểm của phương pháp này là đòi hỏi thiết bị biến nạp đắt tiền.

### 2.3.2. Hóa biến nạp

Đây là phương thức kinh điển để biến nạp DNA plasmid vào tế bào *E. coli*. Các tế bào được ủ trong dung dịch  $\text{CaCl}_2$  để trở thành tế bào khả biến giúp cho chúng dễ tiếp nhận DNA plasmid. DNA plasmid đưa vào bằng cách shock nhiệt nhanh (40-50 giây), các tế bào biến nạp sau đó được chọn lọc bằng phương pháp chọn lọc dương tính trên đĩa agar chứa môi trường LB với kháng sinh thích hợp. Mỗi khuẩn lạc trên đĩa kháng sinh đại diện cho một thể biến nạp đơn. Các tế bào chứa plasmid mang DNA ngoại lai có thể xác định bằng mắt trên đĩa môi trường có bổ sung thêm cơ chất nhiễm sắc thể cho  $\beta$ -galactosidase (X-gal) vì chúng là các khuẩn lạc không màu do sự khử hoạt tính của enzyme bằng cách chèn đoạn DNA ngoại lai.

Phương pháp chuẩn bị và bảo quản tế bào khả biến trong hóa biến nạp cũng rất đơn giản. Hiệu suất biến nạp của phương pháp này trong khoảng  $10^4$ - $10^6$  thể biến nạp/ $\mu\text{g}$  plasmid, tùy thuộc vào kích thước của đoạn DNA chèn (insert DNA) và chủng vi khuẩn được sử dụng. Hiệu suất này thích hợp cho các phương thức tạo dòng truyền thống. Đối với các phương thức cần hiệu suất biến nạp cao hơn (ví dụ: xây dựng thư viện cDNA, xây dựng thư viện phân tích trình tự DNA...) thì tốt hơn hết là dùng phương pháp điện biến nạp. Tuy nhiên, nếu không có sẵn thiết bị biến nạp bằng điện thì vẫn có thể thu được hiệu suất biến nạp cao bằng cách dùng các chủng vi khuẩn thích hợp hơn cho mục đích này và có thể thu được hiệu suất  $10^9$  thể biến nạp/ $\mu\text{g}$  plasmid.

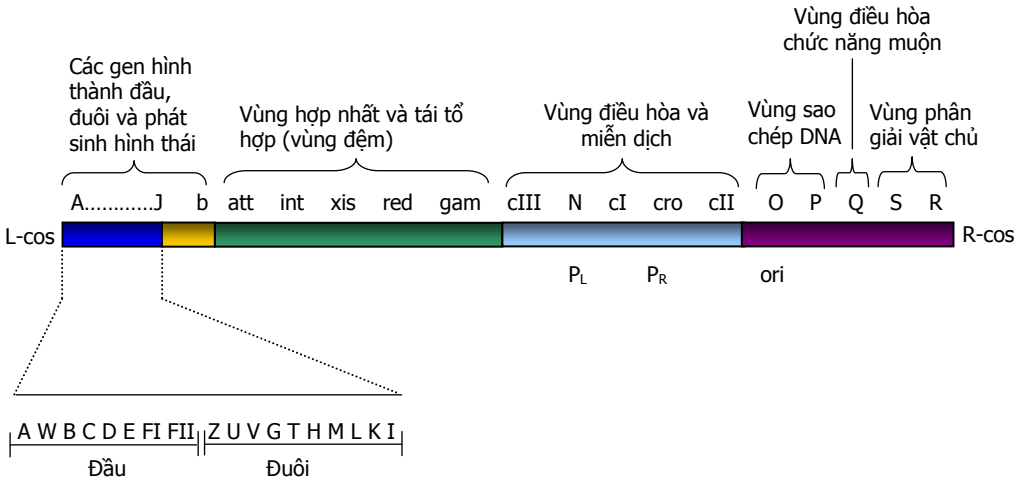
## II. Bacteriophage $\lambda$ vector

Phage  $\lambda$  là một virus mang DNA sợi đôi có kích thước khoảng 50 kb, DNA của nó ở dạng phân tử mạch thẳng có hai đầu hỗ trợ sợi đơn dài 12 nucleotide (được dùng làm đầu *cos*). Sau khi xâm nhiễm vào vi khuẩn vật chủ, DNA tạo vòng ngay lập tức bằng cách ghép hai đầu kết dính *cos* nhờ enzyme DNA ligase của vật chủ (Hình 4.8). Các giai đoạn tiếp đó có thể dẫn đến việc hoặc làm tan (lysis) tế bào vật chủ (bao gồm việc sản xuất và giải phóng các phage thế hệ con) hoặc là gây ra hiện tượng tiềm tan (lysogeny) (trong giai đoạn này  $\lambda$  DNA hợp nhất vào trong chromosome của vật chủ).

Các phage có khả năng thực hiện tải nạp mang gen từ tế bào vi khuẩn cho (donor) sang tế bào nhận (recipient). Vector phage  $\lambda$  được sử dụng rộng rãi để xây dựng thư viện gen (gene library), vì nó mang được đoạn DNA



ngoại lai lớn hơn (15-23 kb) so với plasmid, dễ bảo quản, dễ tách ra để phân tích. Ưu điểm nổi bật của các phage  $\lambda$  là chúng có hệ thống tự động xâm nhập và sinh sản trong tế bào vi khuẩn với một hiệu quả cao hơn nhiều so với việc đưa plasmid vào tế bào vi khuẩn bằng biến nạp. Tuy nhiên các thao tác ban đầu có phức tạp hơn.



**Hình 4.8. Bản đồ đơn giản của bacteriophage  $\lambda$  genome.** Các gen liên quan đến các chức năng khác nhau đã được trình bày trên sơ đồ. *cIII*, *N*, *cI*, *cro* và *cII*: các gen liên quan đến hoạt động điều hòa, miễn dịch tiền phage, siêu nhiễm. *O* và *P*: các gen tổng hợp DNA. *Q*: gen điều hòa chức năng muôn. *S* và *R*: các gen phân giải tế bào vật chủ.  $P_L$ : promoter bên trái,  $P_R$ : promoter bên phải, L-cos: đầu kết dính bên trái, R-cos: đầu kết dính bên phải.

### 1. Đặc điểm của bacteriophage $\lambda$

Trước đây, người ta nghĩ rằng việc sử dụng phage  $\lambda$  làm vector là rất bất tiện do DNA của nó quá dài (50 kb), muốn cắt ra nhiều đoạn thì phải có nhiều trình tự cắt. Tuy nhiên, về sau người ta đã phát hiện ở giữa DNA của phage  $\lambda$  có vùng đệm (stuffer) hay còn gọi là vùng trung tâm có chiều dài bằng khoảng 1/3 chiều dài của phage  $\lambda$ , vùng này không quan trọng và nếu cắt nó đi thì vẫn không ảnh hưởng đến khả năng sinh sản của phage  $\lambda$  trong tế bào vi khuẩn chủ và khả năng làm tan tế bào chủ. Vì thế, người ta đã cắt bỏ đoạn stuffer để lắp các đoạn gen ngoại lai vào. Do DNA phage  $\lambda$  có dạng mạch thẳng nên khi bị cắt ở giữa sẽ tách rời làm thành hai nhánh (arms):

phải và trái. Đoạn gen ngoại lai được gắn vào giữa sao cho DNA tái tổ hợp không lớn hơn 105% hay nhỏ hơn 78% của bộ gen bình thường của phage  $\lambda$ . Khả năng có thể mang một DNA ngoại lai dài hơn của plasmid là một ưu thế của phage  $\lambda$ . Điều đó có nghĩa rằng khi lấy đoạn stuffer ra (chỉ còn 60% chiều dài bình thường của phage  $\lambda$ ) thì nó quá ngắn để lắp vào đầu, nó chỉ lắp được khi DNA ngoại lai gắn thêm vào chỗ trống. Nhờ vậy, để xác định DNA ngoại lai đã gắn vào phage  $\lambda$  hay chưa và đồng thời tính chất này cũng giúp loại bỏ các phage  $\lambda$  khi đoạn DNA ngoại lai bị cắt rời ra.

Trên vector phage  $\lambda$  người ta cũng đã chọn được các vị trí nhận biết cho các enzyme cắt hạn chế, và xây dựng phương pháp đóng gói *in vitro* (*in vitro* packing) để đưa đoạn DNA đã được gắn vào đầu của phage.

## 2. Tạo dòng trong bacteriophage $\lambda$

Bacteriophage  $\lambda$  với genome phức tạp và lớn được dùng như một vector, DNA của nó chứa một số vị trí thích hợp cho các RE cần thiết để tạo dòng. Các bước tạo dòng trong phage  $\lambda$  được tóm tắt như sau (Hình 4.9 và 4.10):

- Tinh sạch DNA vector bằng ly tâm theo gradient tốc độ hoặc điện di gel sau khi đã được cắt bằng RE.

- Các nhánh phải và trái sau đó được liên kết với DNA ngoại lai kích thước khoảng 20 kb có đầu kết thúc tương đồng với đầu của các nhánh.

- Kết quả các DNA tái tổ hợp được đóng gói *in vitro* trong các phage  $\lambda$ .

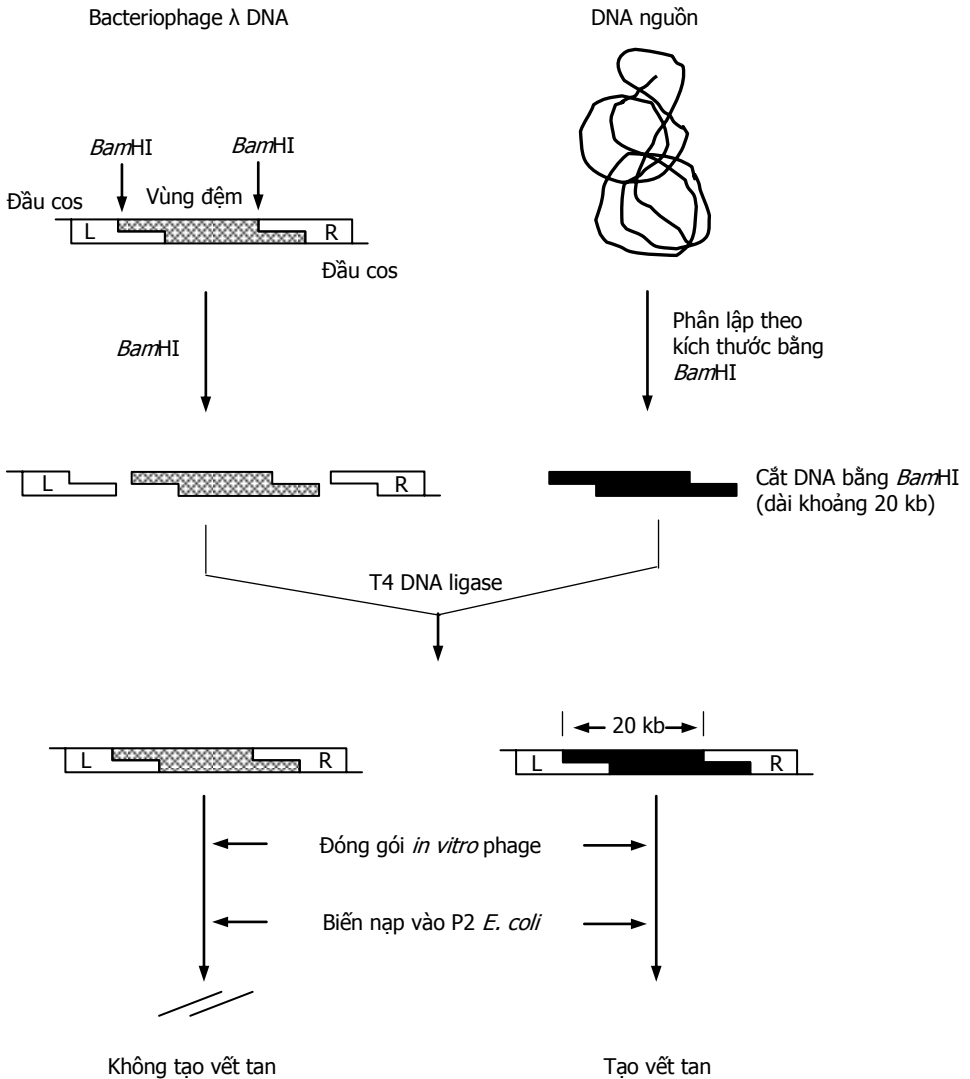
- Các phage tái tổ hợp mang các đoạn DNA ngoại lai mong muốn được xác định bằng phương pháp lai nucleic acid.

- Tinh sạch và chọn tế bào vật chủ (*E. coli*) cho nó sinh sản để duy trì và phân tích các đoạn DNA ngoại lai.

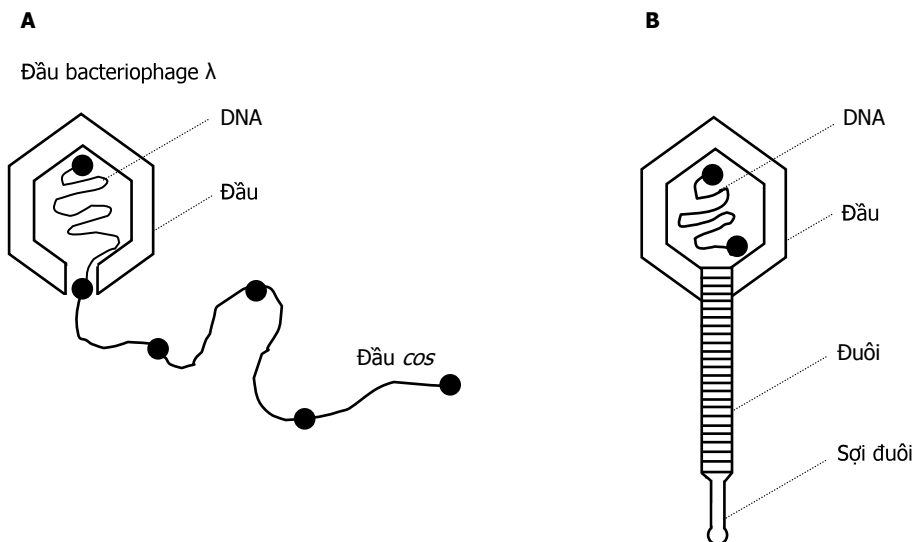
### 2.1. Chọn vector $\lambda$ thích hợp

Có hai yếu tố ảnh hưởng đến sự chọn lọc, đó là enzyme cắt hạn chế được sử dụng và kích thước DNA ngoại lai. Chỉ khoảng 60% genome (nhánh trái và nhánh phải) là cần thiết cho sự sinh tan của phage. Khả năng sống sót của phage giảm khi DNA dài hơn 105% hoặc ngắn hơn 78% so với genome tự nhiên đã được đóng gói. Điều quan trọng là chọn vector và DNA

ngoại lai như thế nào để kích thước của phage tái tổ hợp nằm trong giới hạn cho phép.



**Hình 4.9. Các bước tạo dòng trong bacteriophage λ.** DNA nguồn được cắt bằng *Bam*HI và được phân đoạn theo kích thước (khoảng 20 kb). Bacteriophage λ cũng được phân cắt bằng *Bam*HI. Hai mẫu này được trộn vào nhau và xử lý bằng T4 DNA ligase. Phản ứng gắn xảy ra tạo một bacteriophage λ mới (tái tổ hợp): đoạn stuffer được thay bằng đoạn DNA ngoại lai. Những phân tử này đóng gói *in vitro* trong đầu của phage λ, và chúng có thể gây nhiễm sau khi được thêm phần đuôi.



**Hình 4.10. Phân tử DNA của phage  $\lambda$ .** A: Tự tái bản DNA từ dạng vòng của  $\lambda$  làm cho nó trở thành dạng dây thẳng, hình thành các đoạn chồng lấp lên nhau, với độ dài khoảng 50 kb. B: Mỗi đầu đều chứa đầy 50 kb đơn vị của  $\lambda$ DNA, trước khi đuôi được tổng hợp gắn vào sau.

Một số vector  $\lambda$  đã được thiết kế thuận lợi cho sinh trưởng của chúng trong *E. coli* mang prophage P2, phenotype này được gọi là  $Spi^+$ . Các chủng  $\lambda$  thiếu hai gen cần thiết trong tái tổ hợp (*red* và *gam*) biểu hiện phenotype  $Spi^-$  sẽ sinh trưởng tốt trong các thể tiềm tan P2 chủng nào mà chúng còn mang chuỗi *chi* (*chi* là cơ chất cho sự tái tổ hợp được xúc tác bởi hệ thống *recA*).

Tạo dòng ở các vector  $\lambda$  L47.1,  $\lambda$  1059 hoặc  $\lambda$  BF101 đã được tiến hành bằng cách loại bỏ đoạn stuffer chứa *red* và *gam*. Kết quả thu được phage tái tổ hợp là  $Spi^-$  và chúng có thể được nhận biết hoặc chọn lọc bằng khả năng sinh trưởng của chúng trên các thể tiềm tan P2 *recA*<sup>+</sup> của *E. coli*.

Hầu hết các vector thay thế mất gen *gam* khi đoạn stuffer bị loại bỏ. Các thể tái tổ hợp được tạo thành với các vector như thế phải được sinh sản trong vi khuẩn *recA*<sup>+</sup>. Phage  $\lambda$ sep6-lac5 mang các gen *red* và *gam* trong nhánh phải của nó và các thể tái tổ hợp được tạo thành với vector này có phenotype  $Spi^+$  có thể được nhân lên trong vi khuẩn mang đột biến *recA*<sup>-</sup>.

## 2.2. Những vector $\lambda$ được cải tiến

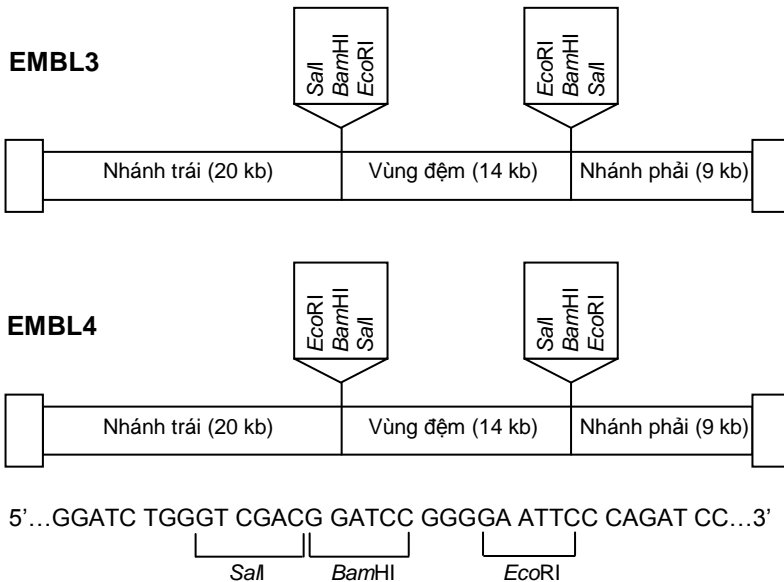
Những vector  $\lambda$  thường được cải tiến nhằm mục đích:

- Tăng khả năng thích ứng của các đoạn DNA ngoại lai khác nhau với các RE.

- Cho phép chọn lựa phương thức tái tổ hợp mong muốn.

- Cho phép có được các RNA probe sẵn sàng cho việc chuyển mã của DNA ngoại lai gắn vào vector. Điều này giúp chúng ta dễ dàng sàng lọc các thư viện trong quá trình chromosome walking, ví dụ như  $\lambda$  ZAP.

- Phát triển các vector trong quá trình gắn vào cDNA của sinh vật bậc cao, dưới hình thức một polypeptide dung hợp với  $\beta$ -galactosidase. Hình thức này rất hữu ích trong việc chọn lọc kháng thể, ví dụ như vector  $\lambda$ gt11.



**Hình 4.11. Vector EMBL3 và EMBL4 được thiết kế từ phage  $\lambda$  mang các vị trí nhận biết cho các enzyme *SalI*, *BamHI* và *EcoRI***

Thế hệ gần đây nhất của vector  $\lambda$  là EMBL3 và EMBL4 (Hình 4.11), chúng mang các trình tự có tính chất polylinker nằm gần đoạn có thể thay thế được. Các phage với những thể insert đính bên trong nó có thể được chọn lọc bằng kiểu hình *Spi*<sup>-</sup>, chính là *chi*<sup>+</sup>. Thể cải tiến từ EMBL3 có

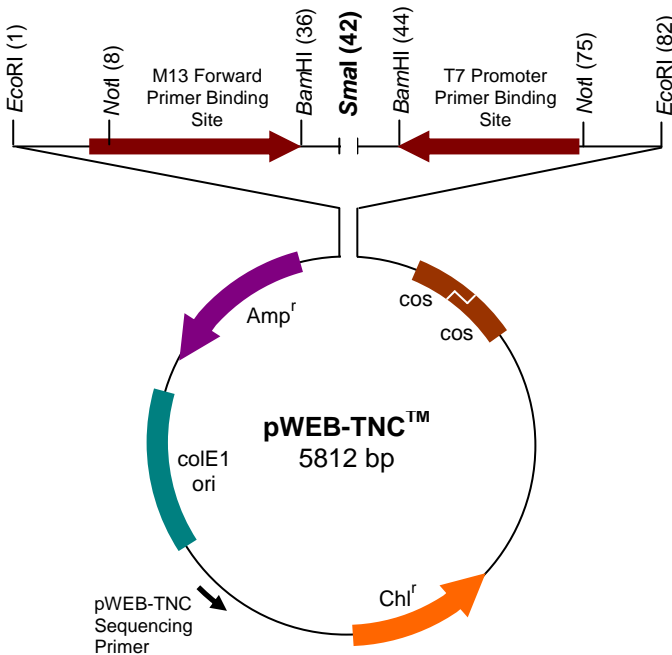
những đột biến EMBL3 *Sam*, EMBL3 *Aam Sam*. Những đột biến trong vector không chỉ làm gia tăng khả năng chứa của nó, mà còn được sử dụng trong hệ thống chọn lọc những trình tự DNA được phân lập, liên kết với các gen ức chế (suppressor).

### III. Cosmid vector

#### 1. Đặc điểm của cosmid

Cosmid là các vector đặc biệt được xây dựng bởi plasmid + đầu *cos* dùng để tạo dòng các đoạn DNA kích thước lớn của eukaryote (Hình 4.12). Các thành phần không thể thay thế của các cosmid vector bao gồm:

- Các marker kháng kháng sinh và một trình tự *ori* của plasmid.
- Đoạn DNA mang đầu dính *cos* của phage  $\lambda$ .
- Kích thước nhỏ sao cho các đoạn DNA eukaryote dài khoảng 45 kb có thể thích ứng.



**Hình 4.12.** Bản đồ và vùng tạo dòng của vector cosmid pWEB-TNC. Vị trí tạo dòng *SmaI* nằm bên cạnh các cặp *BamHI*, *NotI* và *EcoRI*. *Chl<sup>r</sup>*: gen kháng chloramphenicol.

## 2. Tạo dòng trong cosmid

Các nguyên tắc cơ bản của phương pháp tạo dòng trong cosmid được trình bày ở hình 4.13. Trước hết cosmid được cắt với RE thứ nhất (*ScaI*) để tạo thành mạch thẳng, sau đó phân tử mạch thẳng được cắt tiếp với RE thứ hai (*BamHI*) để làm thành hai nhánh (nhánh lớn và nhánh nhỏ). Một nhánh chứa đầu *cos*, gen kháng kháng sinh và vùng *ori*; nhánh còn lại chỉ chứa đầu *cos* thứ hai. Hai nhánh này kết hợp với genomic DNA đã được cắt với RE (cùng loại với RE thứ hai cắt cosmid thành hai nhánh) nhờ DNA ligase, cuối cùng tất cả chúng được đóng gói *in vitro* trong đầu của các tiểu thể phage  $\lambda$  trưởng thành. Sự đóng gói các phân tử tái tổ hợp trong các tiểu thể phage chỉ thực hiện cho các thể tái tổ hợp có kích thước khoảng 40-50 kb. Các cosmid vector được thiết kế tối ưu nhất có chiều dài từ 4-6 kb và vì thế chúng có thể thích hợp với DNA ngoại lai có kích thước khoảng 45 kb.

Mặc dù có những thuận lợi về kích thước, các cosmid dùng làm vector cũng có một số khó khăn sau:

- Gắn vector với vector: sự tái tổ hợp trong phân tử (intramolecular) giữa hai hoặc nhiều vector trong một cosmid đơn có thể dẫn đến cắt bỏ DNA của cosmid vector và cuối cùng mất cosmid do phân ly.

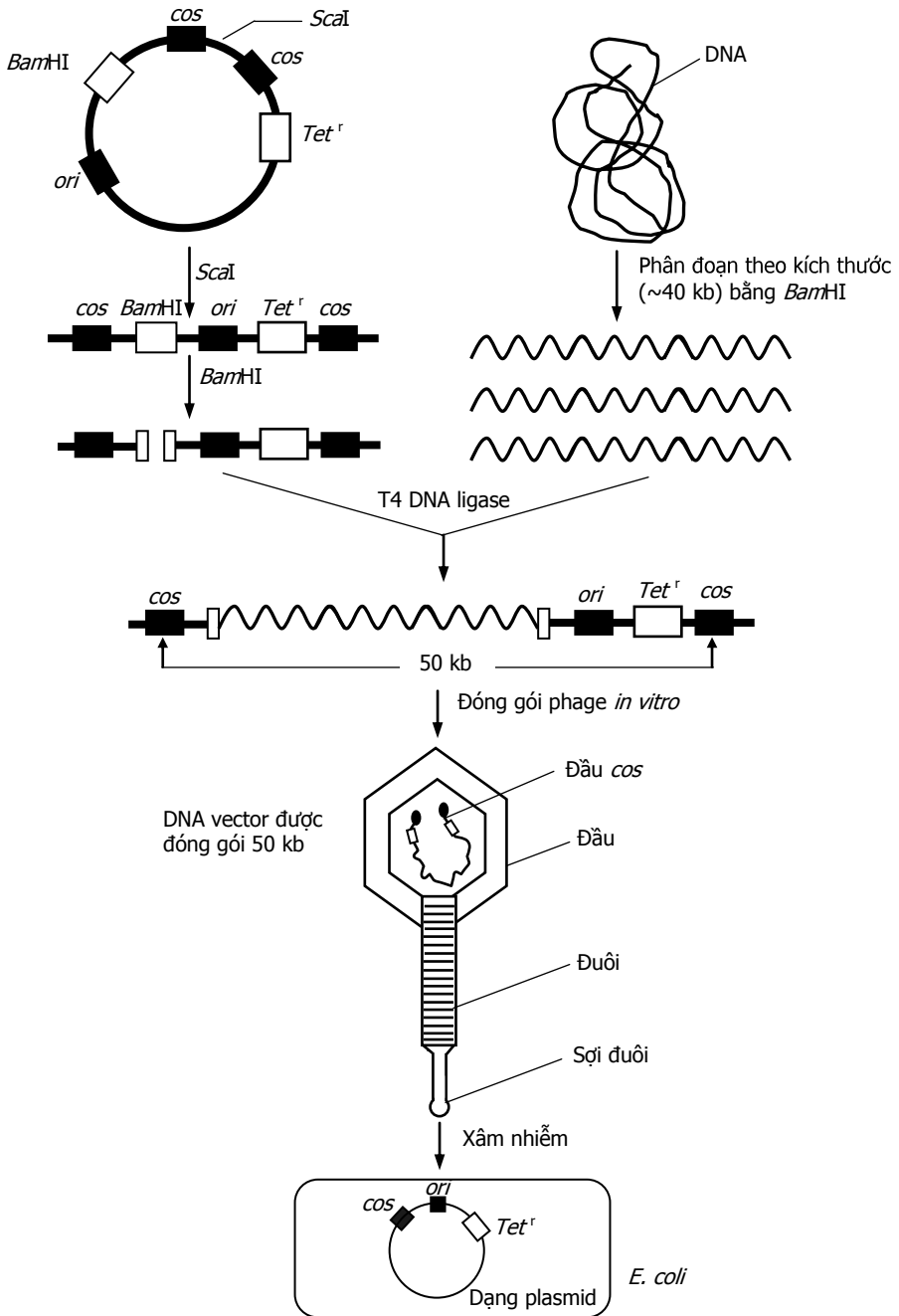
- Sự chát đồng lộn xộn (scrambling): xuất hiện do sự xâm nhiễm vào trong cùng một vector của hai hoặc nhiều đoạn không ở cạnh với đoạn DNA mong muốn trong bộ gen gốc (original genome).

- Khó khăn trong việc sàng lọc một số lượng lớn khuẩn lạc vi khuẩn.

Sự tiện lợi của phương pháp tạo dòng các đoạn DNA lớn trong cosmid đã được Royal và cs. (1979) công bố đầu tiên. Ông có thể phân lập từ thư viện cosmid của DNA gà hai đoạn chồng lên nhau (overlapping) mang gen ovalbumin và hai gen giống ovalbumin. Gần đây hơn, các thư viện gen dùng cosmid vector mang DNA ruồi giấm, DNA chuột, DNA người cũng đã được xây dựng.

Hình 4.13 cho thấy cosmid có chứa vị trí *ori* của *E. coli* (cho phép cosmid được duy trì như một plasmid trong *E. coli*). Hai đầu *cos* gần vị trí cắt hạn chế *ScaI* và *BamHI*. DNA nguồn được cắt bởi *BamHI* để phân đoạn có kích thước khoảng 40 kb. Gen kháng Tet (*Tet<sup>r</sup>*) định vị gần *cos*. Phân tử dạng plasmid được cắt bởi *BamHI* và *ScaI*. Ba mẫu DNA này được trộn vào nhau và được gắn bằng T4 DNA ligase. Sau khi gắn xong, những phân tử

này được đóng gói trong phần đầu của phage  $\lambda$ , và những phần tử có thể lây nhiễm sẽ được hình thành sau khi tạo đuôi.



**Hình 4.13. Tạo dòng trong cosmid**



## IV. Bacteriophage M13 vector

### 1. Đặc điểm của bacteriophage M13

M13 là phage dạng sợi (filamentous bacteriophage) với DNA vòng đóng, dài khoảng 6.500 nucleotide (Hình 4.14). M13 được xem như là một vật truyền tạo dòng (cloning vehicle), các vector M13 và các DNA primer đặc biệt đã được xây dựng cho phép tạo dòng đoạn DNA lớn hơn 350 bp từ một dòng đơn. Tạo dòng các đoạn DNA dài được tiến hành bằng cách tạo dòng từng đoạn ngắn chồng lên nhau.

### 2. Tạo dòng trong bacteriophage M13

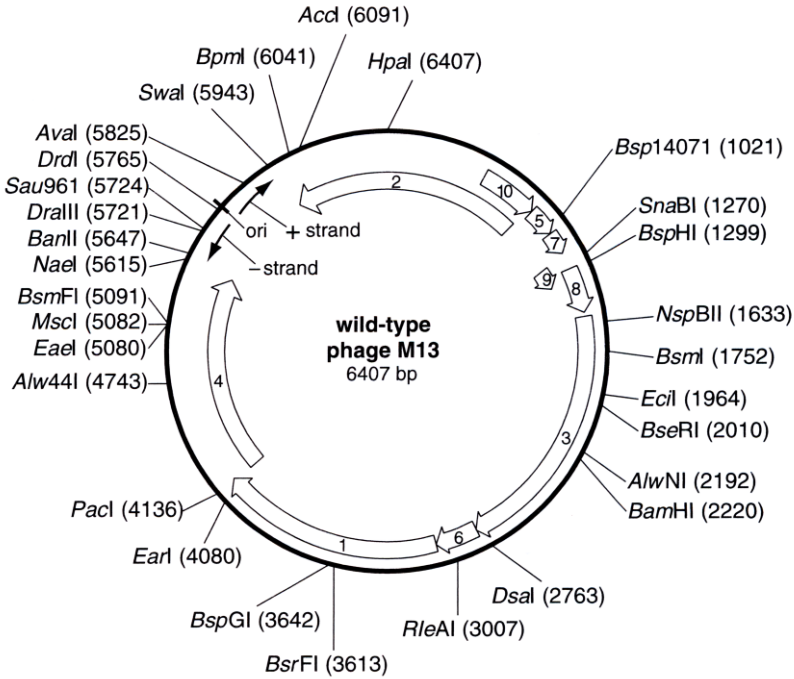
Vấn đề chính trong tạo dòng M13 là sự không ổn định của các đoạn DNA ngoại lai có kích thước lớn, các trình tự lớn hơn 1.000 nucleotide thường không ổn định và làm tăng các đoạn khuyết trong quá trình sinh sản của bacteriophage. Tuy nhiên, các đoạn DNA ngoại lai gồm 300-400 nucleotide thường là ổn định để tiến hành các ứng dụng như đã nói ở trên.

Ngoại trừ vùng 507 nucleotide được biết như là chuỗi liên gen (intergenic sequence-IS) còn tất cả hệ gen của M13 đều chứa thông tin di truyền không thay thế cho sự sao chép của virus. Vùng IS nhận các đoạn DNA ngoại lai nhưng không ảnh hưởng đến khả năng sống sót của tế bào. Trong các dạng phân chia của M13 có hai chuỗi (trình tự) xâm nhiễm:

- Đầu tiên là chuỗi operon lac của *E. coli* chứa gen điều hòa (regulator) và mã hóa thông tin cho 146 amino acid đầu tiên của gen  $\beta$ -galactosidase (Z). Phần amino ở đầu tận cùng của protein  $\beta$ -galactosidase có thể hỗ trợ (hỗ trợ  $\alpha$ ) gen  $\beta$ -galactosidase sai hỏng (defective) có mặt trong episome F trong tế bào vật chủ. Sự hỗ trợ này sản xuất  $\beta$ -galactosidase hoạt động làm tăng màu xanh khi phage và các tế bào sinh trưởng trong môi trường có nhân tố cảm ứng isopropyl-thiogalactoside (IPTG) và cơ chất nhiễm sắc thể X-gal.

- Loại chuỗi thứ hai là đoạn DNA nhỏ, polylinker có chứa một số vị trí cắt hạn chế đã được gắn trong đầu tận cùng amino của gen  $\beta$ -galactosidase. Sự xâm nhiễm này không ảnh hưởng đến khả năng hỗ trợ của chuỗi peptide  $\beta$ -galactosidase cho đột biến  $\beta$ -galactosidase. Các phage mang

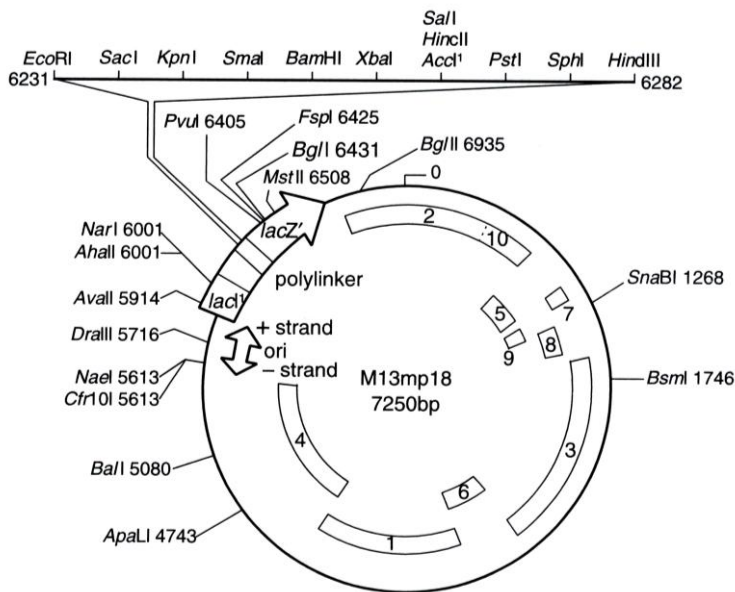
các đoạn chèn đã làm tăng các vết không màu khi sinh trưởng trên môi trường có mặt IPTG và X-gal.



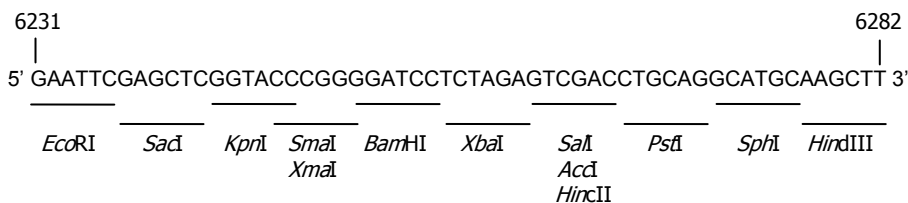
**Hình 4.14. Dạng tự nhiên của bacteriophage M13.** Các gen từ 1-10 có các chức năng khác nhau liên quan đến cấu trúc vỏ protein và phát sinh hình thái của phage.

Vector tạo dòng M13 đúng tiêu chuẩn (phagemid) được trình bày ở hình 4.15, các vector tạo dòng M13 khác loại vector kể trên là ở các vị trí cắt hạn chế trong polylinker.

Đặc biệt, các đoạn DNA nhỏ có vai trò như một primer cũng được tạo dòng trên plasmid hoặc được tổng hợp hóa học. Các primer như vậy tương đồng với vector M13 trong đoạn gen  $\beta$ -galactosidase ở cạnh đoạn polylinker.



**Vùng tạo dòng (MCS) của M13mp18**



**Hình 4.15. Vector M13mp18 (phagemid).** Đây là một trong các vector được sử dụng phổ biến của nhóm vector M13mp.

**V. BAC vector**

*1. Đặc điểm của BAC*

Hệ thống nhiễm sắc thể nhân tạo của vi khuẩn (bacterial artificial chromosome-BAC) dựa trên cơ sở plasmid F-factor, nó có thể tái bản trong *E. coli* với các đoạn chèn có kích thước lên đến 300 kb. Thể tái tổ hợp BAC được biến nạp vào trong tế bào vi khuẩn bằng xung điện (electroporation)<sup>8</sup>. Các BAC thuận lợi cho việc xây dựng thư viện, lập bản đồ và phân tích

<sup>8</sup> Ở điện áp cao tính thấm của màng tế bào được tăng lên sẽ giúp cho DNA ngoại lai dễ dàng đi vào bên trong.

genome. Chúng có hiệu suất tạo dòng cao, các đoạn chèn DNA được thao tác dễ dàng và duy trì ổn định.

Một plasmid F cơ bản bao gồm bốn vùng cần thiết có chức năng ổn định plasmid và quyết định số lượng bản sao. Đó là *parA*, *parB*, *oriS* và *repF*. Cả hai *parA* và *parB* đều cần cho việc phân đoạn và ổn định plasmid. Ngoài ra, *parB* còn cần cho việc kết hợp với các F-factor khác, *oriS* là gốc tái bản DNA, *repF* mang tín hiệu của protein E cần cho quá trình tự tái bản từ *oriS* và quá trình kiểm soát số lượng bản sao. Trên plasmid F người ta còn bổ sung thêm gen kháng chloramphenicol ( $CM^r$  hoặc  $Chl^r$ ) và gen *lacZ* để làm marker chọn lọc các thể biến nạp. Chủng *E. coli* được sử dụng phổ biến cho kỹ thuật tạo dòng BAC là DH10B, đặc điểm chính của nó là mang những đột biến ngăn cản:

- Sự phân cắt DNA lạ bởi hệ enzyme nội sinh (*hsd/RMS*).
- Sự phân cắt DNA có gốc methyl (5'-methyl cytosine, 5'-hydromethylcytosine, gốc methyladenine) (*mcrA*, *mcrB*, *mcrC* và *mrr*).
- Sự tái tổ hợp (*recAI*).

Nhìn chung, các BAC vector có các đặc điểm tương tự với các plasmid vector tiêu chuẩn của *E. coli*, như :

- Chứa vùng *ori* cần thiết cho tái bản của plasmid.
- Có nguồn gốc từ một plasmid lớn trong tự nhiên: F-factor.
- Có số lượng bản sao thấp (1-2 bản sao/tế bào).
- Hiệu suất biến nạp thấp đã được khắc phục bằng phương pháp xung điện để đưa DNA tái tổ hợp vào tế bào *E. coli*.

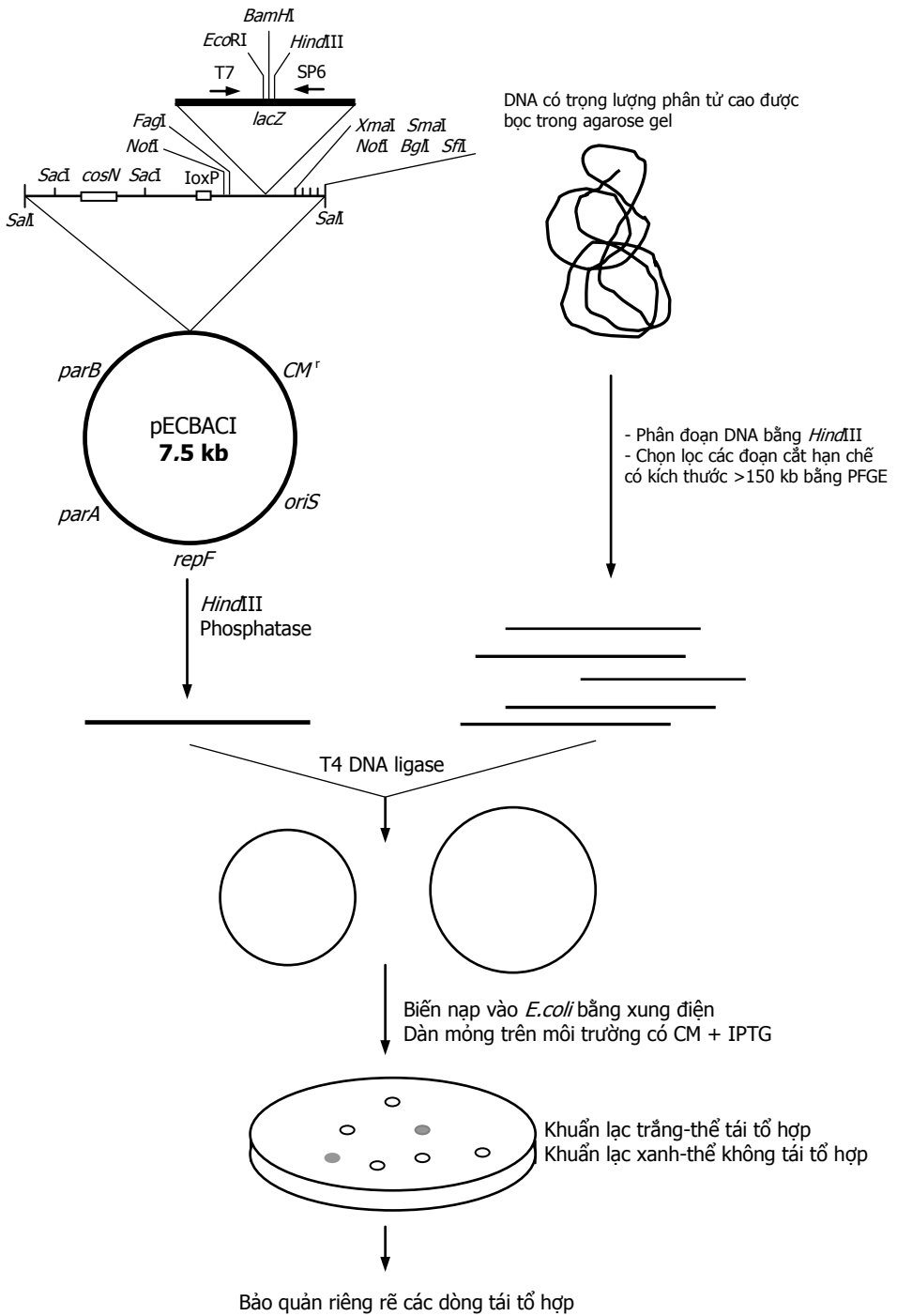
Một số ưu điểm của hệ thống BAC so với YAC:

- Ổn định.
- Dễ biến nạp.
- Tốc độ sinh trưởng của vật chủ *E. coli* nhanh.
- Dễ tinh sạch.

Vì thế, hầu hết genome người đều được phân tích trình tự bằng cách dùng các dòng BAC hơn là các dòng YAC.

## 2. Tạo dòng trong BAC

Quá trình tạo dòng trong BAC vector được trình bày trong hình 4.16, bao gồm các bước sau:



**Hình 4.16. Tạo dòng trong BAC vector**

- BAC vector được cắt bằng enzyme hạn chế *HindIII* (có vị trí nhận biết nằm trên gen *lacZ*), sau đó hai đầu *HindIII* của vector được dephosphoryl hóa bằng phosphatase để ngăn cản sự tái tạo lại vòng.

- DNA có khối lượng phân tử cao được bọc trong agarose và sau đó phân đoạn cũng bằng *HindIII* như trường hợp vector.

- Gắn vector và các đoạn DNA có kích thước >150 kb bằng enzyme T4 DNA ligase, sau đó biến nạp vào *E. coli* bằng xung điện.

- Chọn lọc thể tái tổ hợp trên môi trường nuôi cấy có bổ sung CM và IPTG. Các khuẩn lạc có màu trắng là khuẩn lạc mang thể tái tổ hợp, còn các khuẩn lạc có màu xanh là khuẩn lạc không mang thể tái tổ hợp.

## VI. YAC vector

### 1. Đặc điểm của YAC

Nhiễm sắc thể nhân tạo của nấm men (yeast artificial chromosome-YAC) là các vector tạo dòng mạch thẳng dựa trên cấu trúc nhiễm sắc thể tự nhiên của nấm men. Chúng có thể được cắt thành hai đoạn hoặc hai nhánh nhiễm sắc thể để nhận các đoạn rất lớn có kích thước lên đến 2.000 kb. Cấu trúc của YAC sau đó có thể đưa vào các tế bào nấm men có thành tế bào đã bị loại bỏ (spheroplast), ở đó chúng được duy trì ổn định như các nhiễm sắc thể tự trị (autonomous). Thư viện YAC được lắp ráp và sử dụng cho việc lập bản đồ các vùng có kích thước lớn của DNA người. Do kích thước lớn của chúng cho nên kỹ thuật phân tích thích hợp hơn cả là điện di gel trường gián đoạn (pulsed field gel electrophoresis-PFGE). Nhược điểm của YAC là hiệu suất tạo dòng của một vài hệ thống tương đối thấp, sự xuất hiện các dòng khảm (các tế bào biến nạp chứa hai mẫu DNA không liền kề nhau), sự không ổn định của đoạn chèn và thao tác chúng tương đối khó khăn so với các hệ thống vi khuẩn.

Trước năm 1987, các hệ thống tạo dòng thích hợp chủ yếu dựa trên *E. coli* và chỉ có thể nhận các đoạn DNA tương đối nhỏ. Các vector vi khuẩn có khả năng lớn nhất cũng chỉ có thể nhận các đoạn DNA trong phạm vi kích thước từ 35-45 kb. Tuy nhiên, các YAC vector có thể tạo dòng các đoạn DNA có kích thước từ 100-2.000 kb. Sau này, người ta cũng đã phát triển các vector vi khuẩn mới có khả năng tạo dòng các đoạn DNA lớn hơn, nhưng cho đến nay vẫn chưa thể đạt được khả năng như các YAC. Khả năng

rất lớn này của YAC đã được khai thác để xây dựng các bản đồ vật lý thể hệ thứ nhất và thứ hai của genome người.

**Bảng 4.1. So sánh một số đặc điểm của các vector tạo dòng**

Vector	Năm đưa vào sử dụng	Vật chủ	Cấu trúc	Kích thước đoạn chèn (kb)
Plasmid	1974	<i>E. coli</i>	Mạch vòng	<10
Phage $\lambda$	1977	<i>E. coli</i>	Mạch thẳng	15-23
Cosmid	1978	<i>E. coli</i>	Mạch vòng	35-45
P1 <sup>9</sup>	1990	<i>E. coli</i>	Mạch vòng	70-100
BAC	1992	<i>E. coli</i>	Mạch vòng	<300
PAC <sup>10</sup>	1994	<i>E. coli</i>	Mạch vòng	100-300
YAC	1987	<i>S. cerevisiae</i>	Nhiễm sắc thể mạch thẳng	100-2000

## 2. Tạo dòng trong YAC

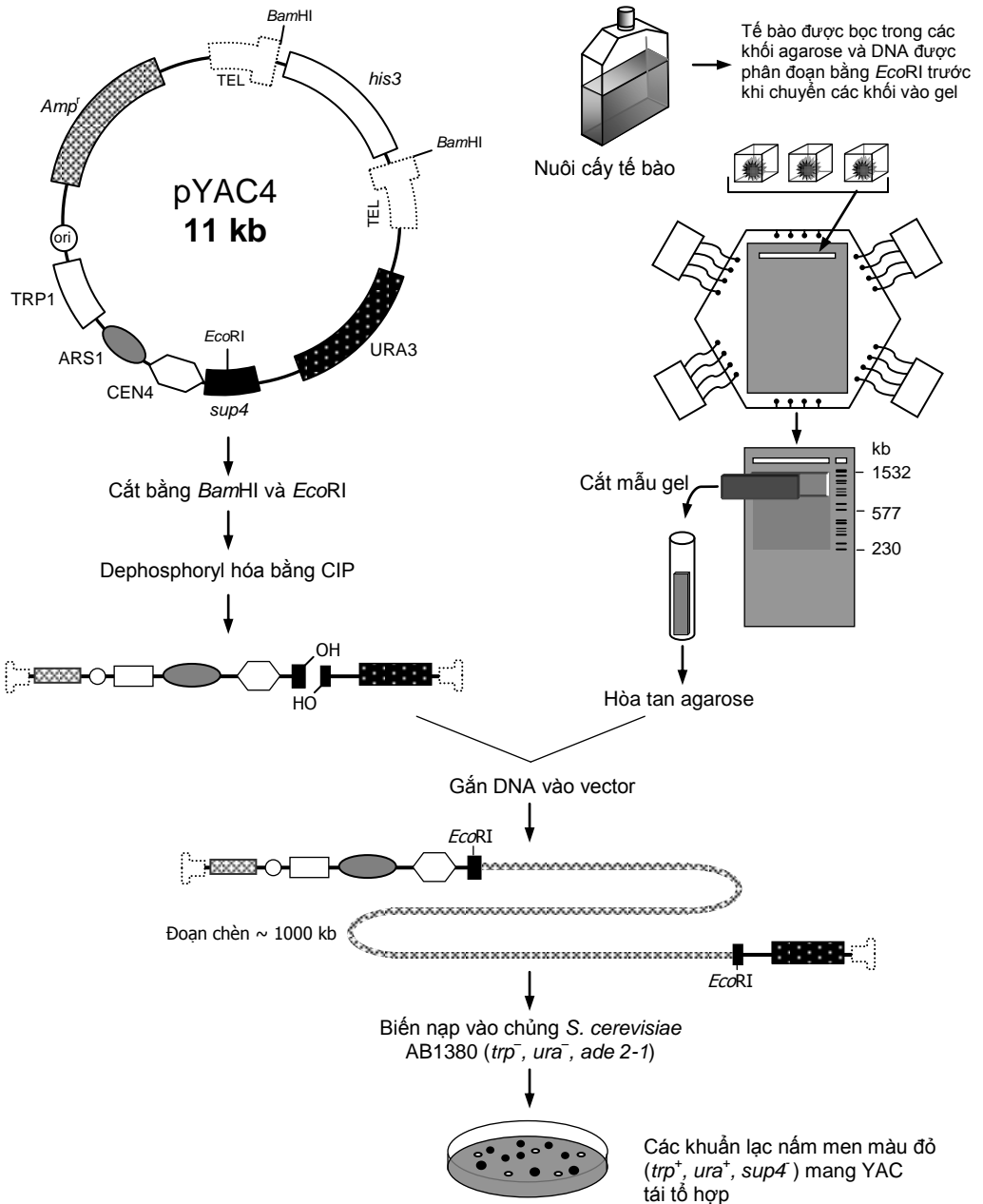
Vector YAC được sử dụng rộng rãi nhất để xây dựng các thư viện YAC là pYAC4 (Hình 4.17). Cắt plasmid mạch vòng dài 11 kb bằng enzyme hạn chế *Bam*HI tạo thành một phân tử mạch thẳng có các trình tự telomere (TEL) ở cả hai đầu. Đoạn nằm giữa hai vị trí chứa gen *his3* là đoạn “stuffer” dài 1,8 kb đã được loại bỏ.

Vị trí tạo dòng để chèn đoạn DNA ngoại lai là *Eco*RI. Đây là vị trí cắt phân tử mạch thẳng thành hai nhánh của YAC, nhánh phải là một đoạn dài 3,4 kb và nhánh trái dài 6 kb chứa các nhân tố CEN4 và ARS1. Vị trí *Eco*RI nằm trong gen *sup4*. Sản phẩm của gen này là một tRNA đột biến ức chế sự đột biến ochre (nonsense) trong gen *ade2*. Khi một đoạn DNA được tạo dòng trong gen *sup4* thì sự ức chế bị đảo thái và quá trình chuyển hóa adenine bị gián đoạn, dẫn đến tích lũy tiền chất trao đổi có màu đỏ

<sup>9</sup> P1 bacteriophage vector.

<sup>10</sup> Nhiễm sắc thể nhân tạo có nguồn gốc từ P1 (P1-derived artificial chromosome).

(phosphoribosyl-aminoimidazole) cho ra các dòng tái tổ hợp màu đỏ để phân biệt.



**Hình 4.17. Vector pYAC4.** Vị trí tạo dòng *Eco*RI ở trong gen *sup4*. *trp1* và *ura3* là các marker chọn lọc ở nhánh trái và phải của YAC tương ứng. *TEL* là hai trình tự telomere và *CEN4* cung cấp chức năng phân tâm.



Hai marker chọn lọc của nấm men, *trp1* và *ura3* ở các nhánh trái và phải tương ứng, mã hóa các enzyme cho phép các tế bào nấm men mang YAC để sản xuất tryptophan và uracil ở điều kiện *de novo*. Dòng nấm men vật chủ *S. cerevisiae* AB 1380 (MAT $\alpha$ ψ+, *ura3*, *trp1*, *ade2-1*, *can1-100*, *lys2-1*, *his5*) không có khả năng này, và sinh trưởng của nấm men này trong môi trường thiếu các chất chuyển hóa như thể cho phép chọn lọc dương tính đối với các phân tử tái tổ hợp chứa cả hai nhánh của YAC.

## Tài liệu tham khảo/đọc thêm

1. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K. 2002. Short Protocol in Molecular Biology. Vol 1 and 2. 5<sup>th</sup> ed. John Wiley & Sons, Inc. USA.
2. Glick BR and Pasternak JJ. 2003. Molecular Biotechnology: Principles and Applications of Recombinant DNA. 3<sup>rd</sup> ed. ASM Press. USA.
3. Maniatis T, Fritsch EF and Sambrook J. 1989. Molecular Cloning-A Laboratory Manual. Cold Spring Harbor Laboratory Press, USA.
4. Primrose SB, Twyman R and Old RW. 2001. Principles of Gene Manipulation. 6<sup>th</sup> ed. Blackwell Science, Oxford, UK.
5. Rapley R and Walker JM. 1998. Molecular Biomethods Handbook. Humana Press Inc. New Jersey, USA.
6. Singleton P and Sainsbury D. 2001. Dictionary of Microbiology and Molecular Biology. 3<sup>rd</sup> ed. John Wiley & Sons, Ltd. UK.

## Chương 5

# Tạo dòng genomic DNA của eukaryote và xây dựng thư viện genomic DNA

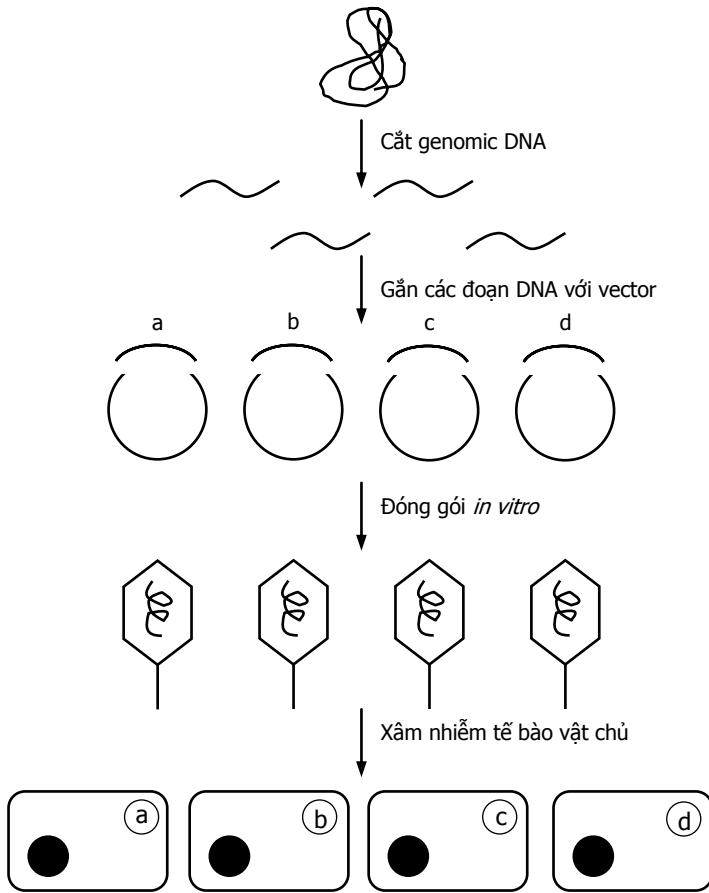
Thư viện genomic DNA là một tập hợp các đoạn DNA cùng đại diện cho một genome nguyên vẹn (hoặc gần nguyên vẹn) của cá thể mà DNA được bắt nguồn từ đó. Các đoạn này nằm trong các vector tự sao chép cho phép chúng được duy trì và sinh sản cùng với tế bào của cơ thể vi sinh vật, như vi khuẩn *Escherichia coli* hoặc nấm men *Saccharomyces cerevisiae*. Những thư viện như thế có thể bảo quản như là một nguồn cố định của các đoạn DNA đại diện cho một cơ thể đặc biệt. Một phòng thí nghiệm này có thể thu thập thư viện genome từ một phòng thí nghiệm khác hoặc từ một nguồn thương mại như là một cách lựa chọn để xây dựng thư viện riêng của mình.

Một khi thu được, các thư viện được dùng chủ yếu hoặc để sàng lọc theo các phương pháp khác nhau nhằm phân lập một (các) trình tự nucleotide quan tâm đặc biệt, hoặc để xác định vị trí và thứ tự của các trình tự trong genome mà từ đó thư viện được xây dựng (physical mapping).

Xây dựng một thư viện genomic DNA bao gồm bốn giai đoạn chính (Hình 5.1): Cắt DNA, gắn DNA vào vector, đóng gói các dòng tái tổ hợp trong vỏ protein của virus để làm bước trung gian trước khi xâm nhiễm vào tế bào vật chủ, và chuyển cấu trúc đó vào tế bào vật chủ.

### I. Cắt genomic DNA bằng các enzyme hạn chế

Để đưa toàn bộ các gen của genome vào một thư viện, trước hết genome phải được cắt bởi enzyme hạn chế thành các đoạn DNA có kích thước thích hợp tùy thuộc vào loại vector nhận chúng. Genomic DNA thường được cắt bằng enzyme hạn chế có trình tự nhận biết 4 bp, ví dụ như *MboI* nhận biết trình tự 5'-GATC-3'. Tuy nhiên, không phải dễ dàng thu được một gen nguyên vẹn nằm trên một đoạn DNA. Trong thực tế, một gen thường bị cắt thành nhiều đoạn DNA do vị trí nhận biết của enzyme nằm ngay trong gen.



**Hình 5.1. Xây dựng thư viện genomic DNA.** Các giai đoạn trong xây dựng thư viện DNA trong vector bacteriophage  $\lambda$ . a, b, c và d trình bày sự khác nhau (nhưng chồng lên nhau) của các đoạn DNA trong genome, được hợp nhất trong các vector riêng rẽ và đại diện như là các dòng riêng biệt trong một thư viện.

Thông thường, một phản ứng cắt từng phần genomic DNA được tiến hành bằng cách dùng một lượng tối thiểu enzyme và ủ trong một thời gian ngắn sao cho mọi vị trí nhận biết không bị cắt đồng thời. Do các vị trí cắt được phân bố ngẫu nhiên nên một gen vẫn có thể được bảo toàn nguyên vẹn ngay khi nó chứa vị trí nhận biết trong gen. DNA sau đó được phân đoạn và chỉ có các đoạn có kích thước nằm trong phạm vi mong muốn được chọn lọc. Các đoạn DNA được đưa vào vector thích hợp. Tập hợp các vector chứa các đoạn genomic DNA tạo thành thư viện genomic DNA. Mỗi đoạn DNA

chứa trong vector của thư viện được gọi là một dòng (clone). Tùy thuộc vào kích thước của các đoạn DNA mà chọn vector là phage (có thể chứa đoạn DNA dài từ 15-23 kb), cosmid (45 kb) hay BAC (>100 kb).

Có thể sử dụng DNA tách từ các mô khác nhau để xây dựng thư viện genomic DNA. Các thư viện genomic DNA của một cơ thể chỉ chứa các đoạn DNA dài ngắn khác nhau nhưng thông tin di truyền không thay đổi. Ngược lại, thư viện cDNA được xây dựng từ các mRNA. Trong cơ thể có những gen chỉ hoạt động trong những loại tế bào nhất định, do đó giữa các mô khác nhau có thể thu được các phân tử mRNA khác nhau. Vì vậy, một cơ thể có nhiều thư viện cDNA khác nhau đặc trưng cho từng loại tổ chức chuyên hóa (xem chương 6).

## II. Tạo dòng các đoạn DNA để xây dựng thư viện genome

Các đoạn DNA thu được từ phản ứng cắt hạn chế được gắn đồng hóa trị *in vitro* vào vector tạo dòng bằng cách dùng enzyme DNA ligase. Trước khi gắn, vector được cắt ra ở một vị trí đơn, thông thường là cùng với enzyme được dùng để cắt DNA nguồn, nhằm tạo các đầu kết dính bổ trợ cho các đầu của các đoạn chèn DNA. Một đôi khi, vector và DNA nguồn được cắt với tổ hợp hai enzyme khác nhau để ngăn cản mỗi loại phân tử tự gắn lại. Sự tự tái tạo lại vòng của vector cũng có thể giảm thiểu bằng cách xử lý với enzyme alkaline phosphatase để loại nhóm 5' phosphate khỏi các đầu của vector.

Một số lượng lớn các vector tạo dòng được phát triển cho việc xây dựng các thư viện DNA, chọn lựa vector nào phụ thuộc vào phạm vi kích thước của các đoạn DNA được tạo dòng, và các ứng dụng tiếp theo.

### 1. Các plasmid vector

Các plasmid là các phân tử DNA mạch vòng đóng cộng hóa trị, siêu xoắn và có kích thước vài kilobase (chương 4). Mặc dù các plasmid xuất hiện tự nhiên, chủ yếu ở vi khuẩn, nhưng những plasmid dùng trong xây dựng các thư viện DNA thường được thiết kế có chủ định với các cấu trúc nhân tạo. Các vector tạo dòng này chứa các vị trí nhận biết đơn cho một enzyme hạn chế mà tại đó các DNA ngoại lai có thể được chèn vào (insertion). Các vị trí cho nhiều enzyme khác nhau có thể được tập hợp lại trong vùng tạo dòng (multiple cloning sites) hoặc vùng đa nối (polylinker).

Các plasmid mang bổ sung một hoặc nhiều gen chỉ thị (marker gene), như các gen kháng kháng sinh, cho phép chỉ có các tế bào chứa plasmid sinh trưởng trên môi trường chọn lọc, và các gen cho enzyme như  $\beta$ -galactosidase cho phép các phân tử tái tổ hợp chứa đoạn chèn DNA được phân biệt với các thể không tái tổ hợp.

Nhược điểm của các plasmid được dùng làm các vector tạo dòng là khả năng chứa của chúng bị giới hạn chỉ thích hợp với những đoạn DNA ngoại lai có kích thước nhỏ vài kilobase (kb) và hiệu suất biến nạp của chúng vào tế bào vật chủ tương đối thấp.

## 2. Các bacteriophage $\lambda$ vector

Các bacteriophage là các virus xâm nhiễm vào vi khuẩn. Genome của chúng có thể là DNA sợi đôi mạch vòng hoặc mạch thẳng, và được bọc bằng vỏ protein làm trung gian cho sự xâm nhập của chúng vào tế bào vật chủ.

Bacteriophage  $\lambda$  là một trong các vector tạo dòng có nguồn gốc virus thông dụng nhất được dùng trong xây dựng thư viện. Genome của nó bao gồm một phân tử DNA sợi đôi mạch thẳng dài khoảng 50 kb, phân tử này xâm nhiễm vào *E. coli* và tạo vòng bằng cách ghép các đầu dính (xem chương 4). Trong chu kỳ sinh tan (lytic cycle), các phân tử dạng vòng của  $\lambda$  genome sẽ được sao chép và sau đó được cắt ở vị trí đặc biệt (vị trí *cos*) để tạo ra các  $\lambda$  monomer sẽ đóng gói trong các đầu protein hoàn chỉnh. Cũng có khi phage  $\lambda$  không đi vào chu kỳ tan, mà chuyển sang giai đoạn tiềm tan (lysogenic), trong suốt giai đoạn này nó hợp nhất ổn định trong nhiễm sắc thể của tế bào vật chủ *E. coli*. Các gen đòi hỏi cho chức năng sinh tan được tập hợp lại trong đoạn stuffer, và được loại bỏ khi xây dựng các vector tạo dòng có nguồn gốc  $\lambda$  để có thể nhận các đoạn DNA kích thước lên đến 20 kb. Các vector  $\lambda$  tái tổ hợp bao gồm nhánh trái, nhánh phải và đoạn DNA ngoại lai được gắn vào giữa hai nhánh. Kết quả các thể tái tổ hợp sau đó được đóng gói trong vỏ protein bằng cách dùng hệ thống đóng gói *in vitro* cho phép xâm nhiễm vào tế bào vật chủ với hiệu suất cao.

Đối với genome đặc trưng của động vật có vú dài  $3 \times 10^9$  bp, xấp xỉ khoảng  $7 \times 10^5$  thể tái tổ hợp  $\lambda$  mang các đoạn chèn có kích thước trung bình khoảng 20 kb sẽ đảm bảo xác suất 99% mọi đoạn DNA cần thiết hiện diện trong thư viện.

### 3. Các cosmid vector

Cosmid là plasmid vector mang đầu *cos* của  $\lambda$  (chương 4). Vì thế, các cosmid có thể được đóng gói trong các tiểu thể  $\lambda$  với hiệu suất cao. Sau khi xâm nhiễm vào tế bào vật chủ, DNA tái tổ hợp mạch thẳng tạo vòng ở đầu *cos* và tiếp đó được nhân lên như là một plasmid lớn. Giống như plasmid, cosmid tương đối dễ thao tác, nhưng có thể nhận các đoạn chèn genomic DNA có kích thước khoảng 40-45 kb, lớn hơn khả năng của phage  $\lambda$ .

### 4. Các BAC vector

Các BAC vector hiện nay được xem như công cụ hữu hiệu nhất trong việc xây dựng các thư viện genome, do chúng có các ưu điểm sau:

- Có khả năng gắn các đại phân tử DNA 100-300 kb.
- Ít hình thành thể khảm (chimera).
- Có hiệu quả cao trong tạo dòng và thu hồi DNA.
- Duy trì sự ổn định của các DNA được gắn vào vector.

Nhờ những ưu điểm trên nên hệ thống BAC thường được dùng để xây dựng bản đồ vật lý. DNA có thể dễ dàng được phân lập trong các dòng BAC, các dòng phân lập được từ lai khuẩn lạc sẽ được kiểm tra bằng kỹ thuật DNA fingerprinting thông qua phản ứng cắt của enzyme hạn chế *HindIII*. Kỹ thuật fingerprinting cung cấp cho chúng ta thông tin về những phân trùng lặp (overlapping) và không trùng lặp của BAC. Nhờ vậy, có thể hoàn thiện bản đồ vật lý có chất lượng cao, phục vụ cho việc phân tích genome và chuyển nạp gen sau này.

### 5. Các YAC vector

Các YAC là các nhiễm sắc thể nhân tạo của nấm men và cũng là các vector tạo dòng được sử dụng trong phân tích genome. Chúng mang các đoạn telomere (phần cuối) và centromere (phần tâm) của một nhiễm sắc thể trong nấm men. Cả hai nhân tố này cần thiết cho sự tái bản của nhiễm sắc thể trong các tế bào nấm men: nếu không có telomere, thì sau đó các đầu của nhiễm sắc thể có khả năng bị rời ra hoặc gắn vào các nhiễm sắc thể khác. Nếu không có centromere, thì sau đó các nhiễm sắc thể mới được tạo

thành sẽ không được đẩy vào trong các tế bào mới của quá trình phân chia tế bào. Ngoài ra, còn phải có một gốc tái bản để sao chép DNA.

Các nhân tố này được đặt trong một đoạn DNA đơn, được dùng như một vector để tạo dòng DNA ngoại lai ở trong nấm men. Ưu điểm của YAC là có thể nhận các đoạn DNA có kích thước lớn. Như vậy, trong khi tạo dòng vi khuẩn theo phương pháp truyền thống bằng cách dùng bacteriophage hoặc các plasmid thường bị giới hạn bởi kích thước của các đoạn DNA được tạo dòng (chỉ vài chục kilobase), thì các YAC có thể tạo dòng các đoạn DNA dài hàng ngàn kilobase. Điều này giúp cho việc lập bản đồ genome hoàn chỉnh dễ dàng hơn nhiều.

Nhược điểm của các YAC đó là kỹ thuật thao tác còn nhiều khó khăn, và một vấn đề chung là DNA từ hai vùng khác nhau của genome gốc có thể kết thúc trong một YAC, dẫn đến xuất hiện sự liên kết sai mà kết quả là tạo ra một dòng khảm.

### **III. Đóng gói *in vitro* DNA tái tổ hợp và xâm nhiễm vào tế bào vật chủ**

Trường hợp sử dụng vector mang gen ngoại lai là bacteriophage  $\lambda$  hoặc cosmid để xây dựng thư viện genome, thì thể tái tổ hợp trần (naked recombinant) của các loại vector này không thể chuyển vào trong vi khuẩn được. Do đó, người ta cần phải đóng gói *in vitro* DNA của chúng trong một đầu rỗng của phage  $\lambda$ , rồi sau đó mới cho chúng xâm nhiễm vào các tế bào vi khuẩn.

Nguyên lý của sự đóng gói *in vitro* (*in vitro* packaging) là nhờ vào khả năng mang các đầu *cos* của phage  $\lambda$  ở hai đầu 5' và 3' của các vector tái tổ hợp để đóng gói chúng khi có mặt các đầu rỗng của phage và các protein đóng gói. Phương thức này đòi hỏi các phân tử vector tái tổ hợp phải có chiều dài ít nhất là 38 kb và không được lớn hơn 52 kb. Các đầu phage được đóng gói sẽ hoàn thiện trong các tiểu thể phage xâm nhiễm *in vitro* khi có mặt các sản phẩm của các gen W và FII, và các đuôi của phage. Tất cả protein cần thiết cho đóng gói có thể thu được từ các chủng *E. coli* BHB2688 và BHB2690. Các hỗn hợp đóng gói chỉ cần chuẩn bị một lần và có thể bảo quản ở  $-80^{\circ}\text{C}$ .

Tùy thuộc vào loại vector mà có thể dùng một lượng thích hợp chủng *E. coli* cho việc xâm nhiễm. Các chủng này được cho “đói” (starve) trước khi sử dụng, hoặc bằng cách sinh trưởng trên môi trường tối thiểu và sau đó xử lý với dung dịch  $\text{CaCl}_2$  0,1 M, hoặc bằng cách rửa trong dung dịch  $\text{CaCl}_2$  0,1 M sau khi nhân lên trong môi trường LB<sup>1</sup>. Phương thức xử lý này cảm ứng sự tổng hợp các  $\lambda$  receptor (protein *lamB*) trên bề mặt tế bào và tăng đáng kể sự hấp thụ phage vào các tế bào *E. coli*. Các tiểu thể phage xâm nhiễm, thu được khi có mặt các protein phage và các hợp phần của đuôi phage, sẽ được dùng để xâm nhiễm các vi khuẩn mẫn cảm (susceptible bacteria). Các dòng tái tổ hợp được chọn lọc trên cơ sở tính kháng kháng sinh của chúng và sau đó có thể được khuếch đại tiếp.

Sự xâm nhiễm được thực hiện bằng cách ủ hỗn hợp các tiểu thể phage xâm nhiễm thu được sau khi đóng gói *in vitro* các phân tử DNA tái tổ hợp với các tế bào mẫn cảm (sensitive cell) của vi khuẩn được tiền xử lý. Tiếp theo bước này là khuếch đại thư viện sau khi đã chuẩn độ (titration) các thể tái tổ hợp.

Thư viện gen sau khi được xây dựng có thể bảo quản trong một vài năm dưới dạng dịch huyền phù ở 4°C, hoặc lâu hơn trong dung dịch stock có bổ sung glycerol (khoảng 30%) ở -80°C.

#### **IV. Phân tích genomic DNA bằng lai Southern**

Việc phát hiện và xác định các chuỗi nucleic acid đặc trưng là công việc thường xuyên trong nghiên cứu sinh học phân tử. Nguyên tắc của kỹ thuật này là dựa vào sự lai phân tử (molecular hybridization), dưới các điều kiện thích hợp hai chuỗi nucleic acid đơn tạo thành một phân tử lai. Phản ứng lai phụ thuộc rất nhiều vào mức độ tương đồng của hai trình tự nucleotide. Việc tạo thành phân tử sợi đôi như thế xảy ra chủ yếu thông qua liên kết hydrogen giữa các base G với C, và A với T. Thành phần và sự phân bố của các base không giống rõ rệt với các phân tử nucleic acid cho

---

<sup>1</sup> LB: lysogeny broth, một môi trường giàu dinh dưỡng được dùng chủ yếu cho sự sinh trưởng của vi khuẩn. Môi trường LB còn được gọi là Luria broth hoặc Luria-Bertani broth.



kết quả các tính chất lai khác nhau, vì thế liên kết hydrogen (hoặc lai phân tử giữa các base tương đồng) khi được ghép cặp thích hợp đã cung cấp một công cụ có giá trị để xác định các chuỗi liên quan hoặc đồng nhất với chuỗi nucleic acid quan tâm (mẫu dò).

Trong số các kỹ thuật khác nhau sử dụng lai phân tử để phân tích nucleic acid, thì kỹ thuật được sử dụng phổ biến nhất là lai mẫu dò nucleic acid được đánh dấu với nucleic acid đích được cố định trên một vật đỡ rắn, thường là màng nitrocellulose hoặc nylon.

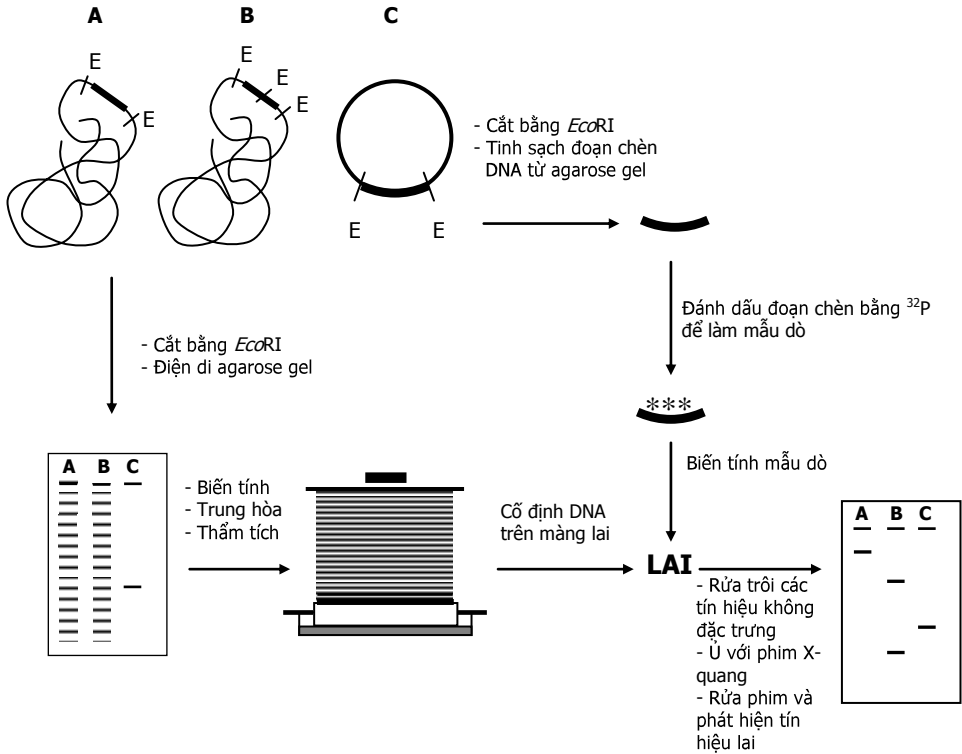
Trình tự của kỹ thuật lai Southern blot bao gồm các bước sau: (1) phân tách các đoạn cắt hạn chế của genomic DNA bằng điện di agarose gel, (2) chuyển DNA từ agarose gel lên màng lai, (3) lai các mẫu dò được đánh dấu đồng vị phóng xạ  $^{32}\text{P}$  với các nucleic acid được cố định trên màng lai (Hình 5.2).

### *1. Phân tách các đoạn cắt hạn chế của genomic DNA bằng điện di agarose gel*

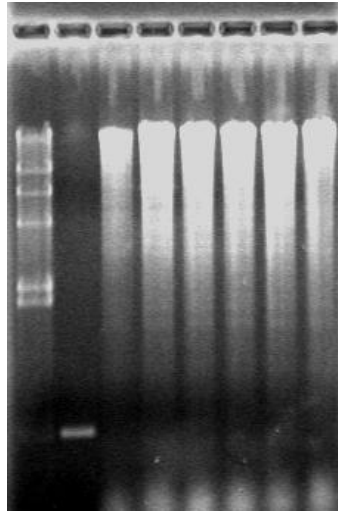
DNA mang chuỗi đích (chẳng hạn genomic DNA) được cắt bằng một hoặc một vài enzyme hạn chế sẽ cho một tập hợp các đoạn có các chiều dài khác nhau được phân đoạn theo kích thước bằng điện di trên agarose gel. Ở trường hợp genomic DNA hình ảnh điện di sẽ cho một vệt dài (smear) trên gel (Hình 5.3).

Lượng DNA dùng cho phản ứng cắt hạn chế tùy thuộc vào mức độ phong phú của chuỗi đích (target) có trong mẫu. Trường hợp genomic DNA, ta cần thủy phân một lượng tương đối lớn DNA (2-10  $\mu\text{g}$ ). Nhưng nếu mẫu DNA được tạo dòng trong các plasmid hoặc phage thì chỉ cần một lượng DNA ít hơn nhiều (một vài picogram) là đủ.

Ảnh gel nhuộm EtBr được xếp thẳng hàng với thước huỳnh quang trước khi thẩm tích là bước quan trọng để đánh giá các kích thước của các băng DNA được lai với chỉ thị kích thước chuẩn của DNA (DNA size-marker). DNA của bacteriophage  $\lambda$  được phân cắt bằng *Hind*III hoặc 1-kb ladder DNA là các chỉ thị kích thước chuẩn của DNA thông dụng nhất cho Southern blot.



**Hình 5.2. Sơ đồ của kỹ thuật lai Southern blot.** Các mẫu DNA đích và genomic DNA trong ví dụ này được cắt bằng *EcoRI* và được phân đoạn bằng điện di trên agarose gel. Các vị trí cắt hạn chế liên quan và trình tự bổ sung với mẫu dò (đoạn dày) cũng đã được trình bày (A và B). DNA của plasmid mang đoạn chèn DNA dùng làm mẫu dò (C) được cắt và điện di để làm đối chứng dương tính. Sau khi biến tính và trung hòa, DNA sợi đơn được chuyển lên màng lai, bằng phương pháp thẩm tích mao dẫn (capillary) hoặc bằng phương pháp thẩm tích nửa khô (semi-dry) nhờ dòng điện, và được cố định. Để chuẩn bị mẫu dò, đoạn chèn DNA (đoạn dày đậm trong C) được tinh sạch từ vector bằng cách cắt và thu hồi sau khi điện di trên agarose gel, đánh dấu bằng  $^{32}P$  và gây biến tính. Tiếp theo, màng lai đã liên kết DNA được lai với mẫu dò, tín hiệu không đặc trưng bị loại bỏ, ủ màng lai với phim X-quang. Sau khi rửa phim, các đoạn DNA bổ sung với mẫu dò được phát hiện như là các băng phóng xạ tự ghi. Trên hình này, nguyên lý cơ bản của đa hình chiều dài các đoạn cắt hạn chế đã được mô tả. Các DNA được tách chiết từ hai mẫu riêng biệt (A và B), trong đó mẫu B có thêm một vị trí *EcoRI* ở chuỗi đích. Sự khác nhau như thế trong genome sẽ được phát hiện bởi các kiểu lai khác nhau. Nguyên tắc này được khai thác trong nhiều ứng dụng của sinh học phân tử.



SM PC 1 2 3 4 5 6

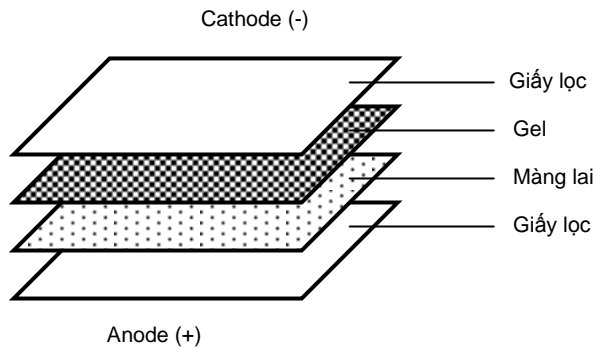
**Hình 5.3. Hình ảnh điện di của genomic DNA sau khi được phân cắt bằng enzyme hạn chế.** SM: Chỉ thị kích thước chuẩn của DNA. PC: đối chứng dương tính (đoạn DNA được dùng để đánh dấu  $^{32}\text{P}$  làm mẫu dò). Các đường 1, 2, 3, 4, 5 và 6: các mẫu genomic DNA được cắt bằng enzyme hạn chế.

## 2. Chuyển DNA từ agarose gel lên màng lai

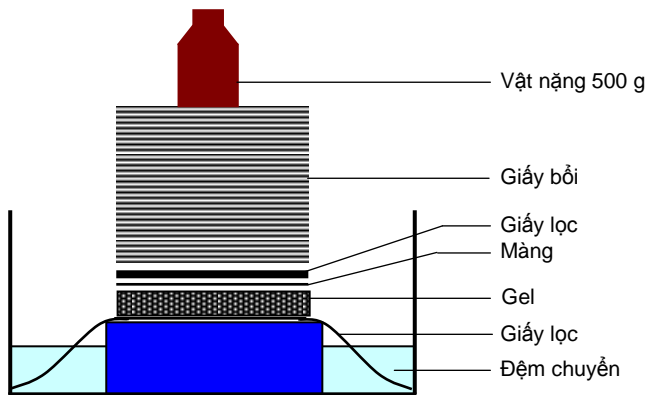
Sau khi điện di gel và trước khi thẩm tích, DNA sẽ được khử purin (depurination) bằng dung dịch HCl loãng để cắt nhỏ DNA tạo điều kiện dễ dàng để chuyển các đoạn DNA có kích thước lớn lên màng. Nhìn chung, bước này khó kiểm soát và có thể làm mất các đoạn DNA nhỏ hơn. Vì thế, nếu có cách khắc phục việc đánh giá hình ảnh phóng xạ tự ghi thì có thể bỏ qua bước này. Tuy nhiên, khử purin vẫn cần thiết khi phân tích Southern các đoạn DNA có kích thước rất lớn.

Các loại màng lai được sử dụng trong Southern là màng nitrocellulose hoặc màng nylon. Tuy nhiên màng nylon có sức chịu đựng cao hơn màng nitrocellulose vì thế có thể tái sử dụng một vài lần nên chúng được sử dụng phổ biến hơn. Hơn nữa, màng nylon có thể gắn các đoạn DNA ngắn (<500 bp) hiệu quả hơn màng nitrocellulose, và DNA sau khi được thẩm tích lên màng có thể được liên kết đồng hóa trị bằng chiếu xạ UV trong thời

gian ngắn (khoảng 1-2 phút) ở 2000 J trong khi đó màng nitrocellulose cần đun nóng ở 80°C trong điều kiện chân không khoảng 2 giờ. Màng nylon tích điện cũng thuận lợi và thích hợp cho thấm tích kiềm (alkali-blotting), ví dụ: NaOH, do DNA liên kết cộng hóa trị với màng dưới các điều kiện này mà không cần bất kỳ một xử lý nào nữa. Trong quá trình thấm tích, thông thường đệm chuyển có nồng độ muối cao được dùng cho genomic DNA, trong khi đó DNA của plasmid hoặc các đoạn PCR được dùng với các đệm có nồng độ muối thấp.



**Hình 5.4. Sơ đồ thấm tích nửa khô bằng điện chuyển DNA từ agarose gel lên màng lai**



**Hình 5.5. Sơ đồ thấm tích mao dẫn chuyển DNA từ agarose gel lên màng lai**

### 3. Lai các mẫu dò được đánh dấu đồng vị phóng xạ với các nucleic acid được cố định trên màng lai

Các điều kiện tiền lai và lai được thiết kế để tăng tối đa phản ứng lai đặc hiệu của mẫu dò với các chuỗi đích và giảm thiểu các liên kết không đặc hiệu với màng và DNA không phải chuỗi đích. Nói chung, các đệm lai cần có cường lực ion cao (nhân tố quan trọng đối với khả năng ổn định lai). Một vài loại tác nhân ngăn chặn có thể được dùng để ức chế các liên kết không đặc hiệu, bằng cách đó đã hạn chế tín hiệu nền. Dung dịch Denhardt và sữa khô không có chất béo (nonfat dried milk) được sử dụng phổ biến để ngăn cản (blocking) liên kết giữa mẫu dò với màng lai, chất tẩy SDS cũng được dùng trong trường hợp này. Salmon sperm DNA (DNA tinh trùng cá hồi) và calf thymus DNA (DNA tuyến ức của bê) được dùng để ngăn cản các tương tác không đặc hiệu giữa các DNA không phải chuỗi đích gắn trên màng lai với mẫu dò.

Trong giới hạn thực hành chỉ có những nhân tố ảnh hưởng đến sự ổn định nhiệt của các phân tử nucleic acid lai mới thể hiện vai trò quyết định trong hầu hết thí nghiệm lai trên màng. Nhiệt độ nóng chảy ( $T_m$ ) của nucleic acid sợi đôi cho biết nhiệt độ mà ở đó DNA sợi đôi bị biến tính 50% dưới các điều kiện đã cho, và nó là kết quả của việc đo trực tiếp khả năng ổn định của việc lai nucleic acid.

Nhiệt độ nóng chảy của DNA sợi đôi có thể được đánh giá thông qua biểu thức 1:

$$T_m(^{\circ}\text{C}) = 81,5^{\circ}\text{C} + 16,6 (\log_{10}M) + 0,41 (\%G + C) - 0,72 (\%f) - 500/n \quad (1)$$

Trong khi biểu thức 2 có thể được dùng cho thể lai RNA-DNA:

$$T_m(^{\circ}\text{C}) = 79,8^{\circ}\text{C} + 18,5 (\log_{10}M) + 0,58 (\%G + C) - 11,8 (\%G + C)^2 - 0,56 (\%f) - 820/n \quad (2)$$

*Trong đó*

$M$  là nồng độ phân tử của các cation hóa trị 1 (đặc trưng là  $\text{Na}^+$ )

$(\%G + C)$  là nồng độ phần trăm của guanine và cytosine

$(\%f)$  là phần trăm của formamide

$n$  là chiều dài của thể lai (theo base).

Các biểu thức này chỉ ra khả năng ổn định của các sợi đôi nucleic acid liên quan trực tiếp với nồng độ muối, tỷ lệ của các nucleotide tạo ra liên kết ba với hydrogen, và chiều dài thể lai.  $T_m$  cũng tương quan nghịch với nồng độ của tác nhân làm mất ổn định của xoắn đôi, như formamide. Mặc dù việc sử dụng formamide cho phép phản ứng lai được tiến hành ở các nhiệt độ thấp truyền thống, thậm chí khi mẫu dò rất dài và chứa một tỷ lệ cao các base G và C, thì việc sử dụng formamide vẫn không được khuyến khích lắm vì nó rất độc. Biểu thức 1 và 2 thích hợp cho các chuỗi bổ sung 100%. Tuy nhiên, sự xuất hiện hiện tượng ghép đôi không tương xứng (mismatch) đã ảnh hưởng lớn đến khả năng ổn định của chuỗi xoắn.

Cường lực dùng trong các thí nghiệm lai là số nghịch đảo của giá trị chuẩn  $C$  (criterion value) được đưa ra bởi biểu thức 3:

$$C = T_m - T_i \quad (3)$$

*Trong đó*

$T_i$  là nhiệt độ thí nghiệm, khi  $T_i$  thấp thì  $C$  cao và lúc đó cường lực sẽ thấp, và ngược lại. Các chuỗi DNA đích có độ tương đồng giống y hệt hoặc gần giống với mẫu dò có thể được xác định dưới các điều kiện cường lực cao (ví dụ:  $5^\circ\text{C} < C < 10^\circ\text{C}$ ), các điều kiện cường lực thấp thích hợp cho các chuỗi ít tương đồng.

Thông thường, phản ứng lai và các bước rửa đầu tiên được tiến hành ở điều kiện cường lực thấp (ví dụ:  $T_m = 30^\circ\text{C}$ ) và cường lực được tăng lên dần trong các bước rửa tiếp theo.

Biểu thức 1 và 2 chỉ chính xác trong trường hợp các DNA sợi đôi dài hơn 100-200 nucleotide. Các mẫu dò oligonucleotide được dùng thường xuyên để sàng lọc các thư viện cDNA hoặc genomic DNA và phân tích Northern, và đôi khi chúng cũng được dùng trong các thí nghiệm lai Southern. Các phân tử lai ngắn có độ ổn định thấp (ngay cả khi độ tương đồng 100%), đặc biệt trong suốt các bước rửa, bởi vì nồng độ mẫu dò trong đệm rửa hầu như là bằng không. Như vậy, các bước rửa cần tiến hành trong thời gian ngắn (2-5 phút).

Hơn nữa, độ ổn định của các sợi đôi oligonucleotide bị ảnh hưởng lớn bởi thành phần nucleotide và bởi hiện tượng ghép đôi không tương xứng.  $T_m$

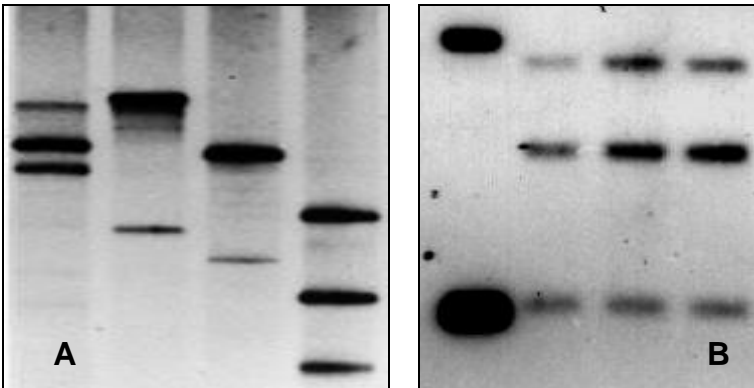
của các mẫu dò oligonucleotide ngắn hơn 50 nucleotide thường được tính bằng biểu thức 4:

$$T_m(^{\circ}\text{C}) = 4(G + C) + 2(A + T) \quad (4)$$

*Trong đó*

A, C, G và T là số các nucleotide tương ứng. Tuy nhiên, thông thường các điều kiện thí nghiệm của phản ứng lai đích thực với các mẫu dò oligonucleotide (đặc biệt khi phân tích các genome lớn) phải được xác định trên cơ sở kinh nghiệm.

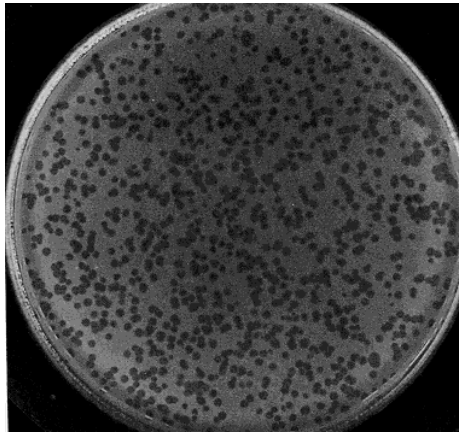
Các thí nghiệm lai trên màng thường được tiến hành bằng cách đánh dấu mẫu dò bằng  $^{32}\text{P}$  (mặc dù các đồng vị phóng xạ khác như  $^{35}\text{S}$  cũng có thể được sử dụng). Phản ứng lai của Southern đòi hỏi hoạt tính đặc hiệu của mẫu dò tối thiểu phải là  $10^9$  dpm/ $\mu\text{g}$ , mặc dù hoạt tính  $10^8$  dpm/ $\mu\text{g}$  có thể được xem như là chấp nhận được trong các ứng dụng cần cường lực thấp. Các phương pháp không dùng đồng vị phóng xạ (ví dụ: digoxigenin-dUTP) ngày càng được sử dụng phổ biến hơn mặc dù độ nhạy của chúng là không thể so với các phương thức dùng đồng vị phóng xạ (Hình 5.6). Nhưng các kỹ thuật không dùng đồng vị phóng xạ an toàn hơn cho nghiên cứu viên, tạo ra các mẫu dò có thể được bảo quản trong thời gian dài trước khi dùng và kết quả phát hiện các tín hiệu lai nhanh hơn.



**Hình 5.6. Hình ảnh phân tích Southern blot của hai thí nghiệm khác nhau.** (A) Hình ảnh phóng xạ tự ghi trên phim X-quang với mẫu dò được đánh dấu  $^{32}\text{P}$ . (B) Hình ảnh bắt màu trên màng lai Hybond-N với mẫu dò được đánh dấu digoxigenin-dUTP. Các băng màu đen là tín hiệu lai giữa DNA đích và mẫu dò.

## V. Sàng lọc thư viện genomic DNA

DNA được tạo dòng trong các vector có nguồn gốc plasmid hoặc các vector nhiễm sắc thể nhân tạo (BAC, YAC) sản xuất ra các khuẩn lạc (vi khuẩn hoặc nấm men) khi các nuôi cấy biến nạp được dàn mỏng trên đĩa agar chứa môi trường sinh trưởng và nuôi dưới những điều kiện thích hợp. Trong khi đó, các vector có nguồn gốc virus (bacteriophage) sẽ sinh tan tế bào bị chúng xâm nhiễm và sản xuất ra các plaque (vết tan) có dạng hình tròn chu vi khoảng 2-3 mm có màu sáng trên thảm vi khuẩn (bacterial lawn) của đĩa agar (Hình 5.7).



**Hình 5.7.** Các vết tan của phage  $\lambda$  trên thảm vi khuẩn *E. coli*. Các phage  $\lambda$  được trộn với các tế bào *E. coli* trong môi trường nuôi rồi dàn mỏng (plating) lên đĩa petri chứa bottom agar (nuôi cấy top agar). Sau khi ủ qua đêm ở 37°C, các tế bào vi khuẩn mọc tạo nên thảm vi khuẩn, ở trên đó các vùng bị nhiễm phage  $\lambda$  hình thành nên các vết trong, do hiện tượng sinh tan vi khuẩn của phage, gọi là vết tan.

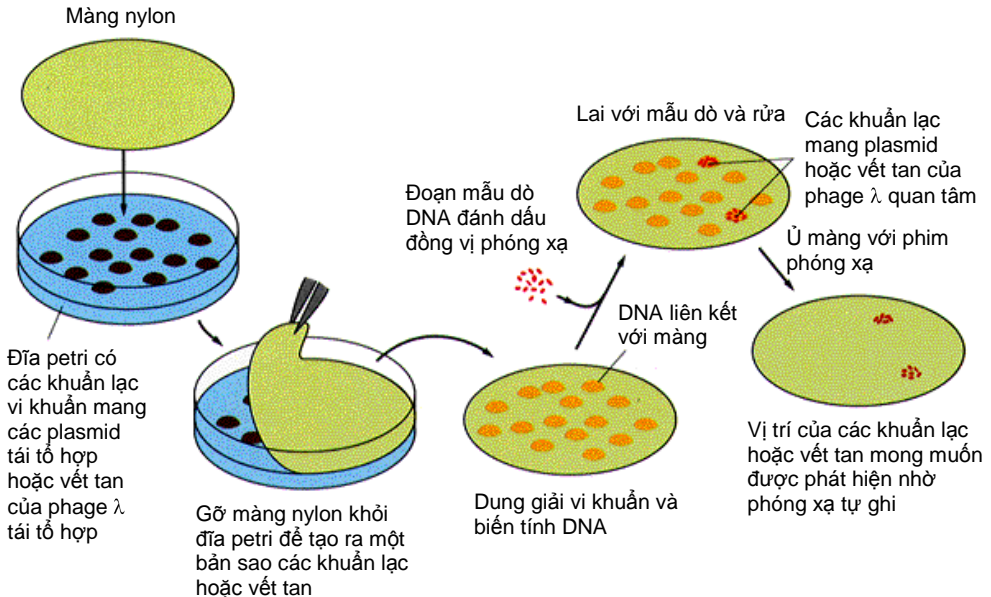
Các vector, như mô tả ở trên, thường chứa các gen chỉ thị cho phép chọn lọc những tế bào vật chủ mang vector (thể biến nạp). Thông thường các chỉ thị này là gen kháng kháng sinh và các tế bào biến nạp sinh trưởng trên môi trường chứa kháng sinh tương ứng.

Ngoài ra, các vector còn chứa các gen bổ sung để phân biệt các tế bào biến nạp chứa đoạn chèn của DNA ngoại lai với các tế bào chứa các vector tự tái tạo lại vòng. Ví dụ: vector mã hóa gen  $\beta$ -galactosidase xúc



tác sản sinh ra các sản phẩm màu xanh từ cơ chất không màu X-gal cho kết quả sinh trưởng của các khuẩn lạc màu xanh. Đoạn chèn của DNA ngoại lai đã làm mất hoạt tính của gen cho kết quả sản xuất ra các khuẩn lạc màu trắng.

Một dòng chứa các chuỗi DNA quan tâm đặc biệt có thể được xác định bởi lai khuẩn lạc. Một lượng nhỏ khuẩn lạc biến nạp hoặc vết tan được chuyển lên màng nitrocellulose hoặc nylon đã được phủ lên (overlay) trên đĩa agar trước đó. DNA được biến tính và cố định trên màng bằng cách đun nóng (baking) hoặc chiếu tia cực tím (UV-crosslinking), và sau đó được lai trong đệm chứa mẫu dò được đánh dấu đồng vị phóng xạ có trình tự bổ sung một phần của chuỗi được xác định. Ví dụ: Có thể một oligonucleotide tổng hợp nhân tạo, có nguồn gốc từ genomic DNA từng phần, cDNA hoặc chuỗi protein hoặc sản phẩm PCR. Một vài trường hợp mẫu dò có thể được thiết kế dựa trên trình tự bắt nguồn từ gen tương đồng của các loài khác và thường được lai với cường lực thấp. Các mẫu dò thừa được rửa khỏi màng để ủ với phim X-quang. Theo hướng của phim (sau khi rửa) với sự đối chiếu đĩa agar gốc, sau đó sẽ có khả năng gắn kết các khuẩn lạc thực tế với các khuẩn lạc lai dương tính tương ứng trên phim X-quang (Hình 5.8).



**Hình 5.8. Sàng lọc (screening) thư viện gen trong khuẩn lạc vi khuẩn hoặc vết tan của bacteriophage λ**

Gần đây, một phương pháp sàng lọc khác đã được thiết kế. Theo hướng này các khuẩn lạc vi khuẩn hoặc nấm men được nhặt riêng rẽ và chắm ngay ngắn thành hàng lên màng hoặc trong giếng của đĩa microtiter. Mỗi dòng có thể được xác định bằng tọa độ đường kẻ riêng rẽ của nó. So với các phương pháp truyền thống, phương pháp mới này dễ dàng hơn trong việc tự động hóa, sắp xếp các thư viện và đối chiếu số liệu giữa các phòng thí nghiệm khác nhau.

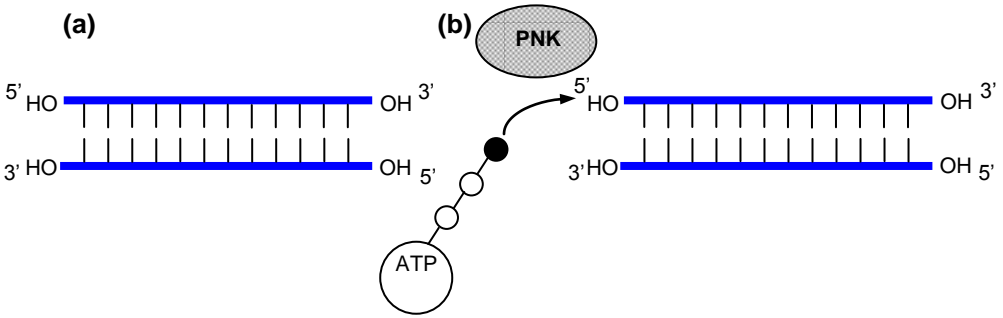
Nếu dòng mong muốn biểu hiện một protein đặc biệt hoặc một đoạn protein, thì sau đó bằng cách xây dựng thư viện với một vector biểu hiện chứa các tín hiệu phiên mã và dịch mã người ta có thể sàng lọc trực tiếp đối với protein được biểu hiện. Ví dụ: nếu protein tương ứng đầy đủ là thích hợp, thì kháng thể đặc hiệu có thể được tạo ra, được đánh dấu và sử dụng để phát hiện yếu tố quyết định kháng nguyên trên protein được biểu hiện từ các khuẩn lạc được dung ly. Nếu đoạn chèn được yêu cầu mã hóa một hoạt tính sinh học, thì sau đó việc sàng lọc có thể được thực hiện bằng các thử nghiệm cho hoạt tính đó, hoặc bằng cách chọn lọc trực tiếp, ví dụ: chọn lọc trên môi trường sinh trưởng không hoàn chỉnh đối với enzyme sinh tổng hợp cần thiết cho sự sinh trưởng. Tuy nhiên, các hướng dựa trên sự biểu hiện nói chung dễ ứng dụng cho thư viện cDNA hơn là thư viện genomic DNA.

## **VI. Các phương pháp đánh dấu đoạn DNA bằng phóng xạ**

Mục đích của bước này là sản xuất các đoạn DNA có độ phóng xạ và hoạt tính đặc hiệu cao để làm mẫu dò dùng trong các thí nghiệm lai phân tử. Trong trường hợp này dấu phóng xạ thường được sử dụng là  $^{32}\text{P}$  phát xạ  $\beta$  năng lượng cao. Dưới đây là một số phương pháp đánh dấu phổ biến được trong kỹ thuật gen.

### *1. Đánh dấu ở đuôi*

Enzyme polynucleotide kinase xúc tác để chuyển nhóm phosphate nằm cuối ATP sang gốc 5'-OH của các phân tử nucleic acid đã được dephosphoryl hóa. Nếu như ATP được đánh dấu phóng xạ, thì nó sẽ tạo ra nucleic acid được đánh dấu phóng xạ. Tuy nhiên, hoạt tính đặc hiệu của chúng tương đối thấp, do chỉ có phần đuôi của mỗi phân tử DNA được đánh dấu (Hình 5.9).



**Hình 5.9. Đánh dấu ở đuôi của đoạn DNA nhờ enzyme polynucleotide kinase (PNK).** (a) DNA được dephosphoryl hóa bằng enzyme phosphatase để tạo ra các nhóm 5'-OH. (b) Tiếp đó, đuôi phosphate của  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  (vòng tròn đậm trên hình) được chuyển sang gốc 5'-OH nhờ PNK. Đây là phản ứng trao đổi các nhóm 5'-PO<sub>4</sub>.

## 2. Đánh dấu bằng dịch chuyển điểm đứt

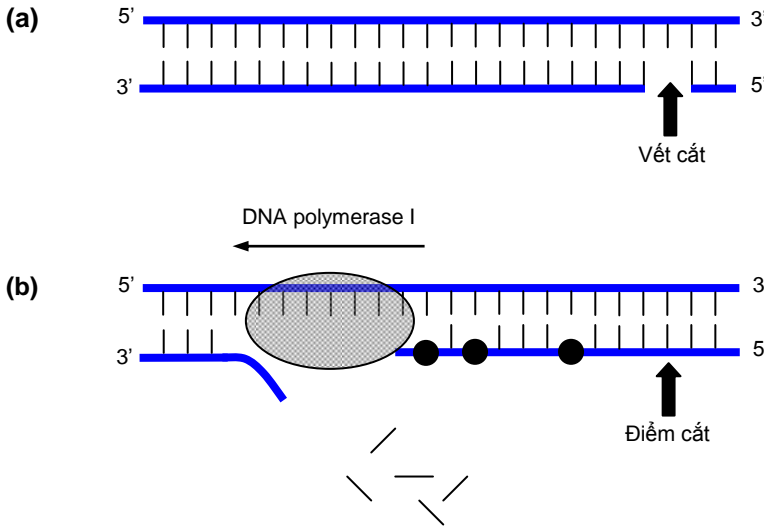
Phương pháp này dựa vào enzyme DNA polymerase I của *E. coli* có khả năng dịch chuyển điểm đứt của DNA (xem chương 1). Các điểm đứt có thể xuất hiện tự nhiên và cũng có thể do tác dụng của enzyme DNase I<sup>2</sup> ở nồng độ thấp trong hỗn hợp phản ứng. Enzyme DNA polymerase I xúc tác phản ứng thay thế sợi bằng cách xen các dNTP mới vào chuỗi DNA. Nếu một trong các dNTP đã đánh dấu phóng xạ (ví dụ:  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ ), thì kết quả là phân tử DNA sẽ được đánh dấu với hoạt tính đặc hiệu cao (Hình 5.10).

## 3. Đánh dấu bằng kéo dài đoạn môi

Đoạn DNA cần đánh dấu được biến tính bằng nhiệt, sau đó các đoạn môi oligonucleotide (thường là các phân tử hexadeoxyribonucleotide) được gắn vào các DNA sợi đơn. Sử dụng enzyme DNA polymerase I có thể tổng hợp nên bản sao mới của sợi khuôn mẫu. Nếu như một dNTP đã đánh dấu phóng xạ (ví dụ:  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ ) được gắn vào, thì bản sao DNA mang hoạt tính đặc hiệu rất cao sẽ được tạo ra (Hình 5.11). Cũng có thể dùng kỹ

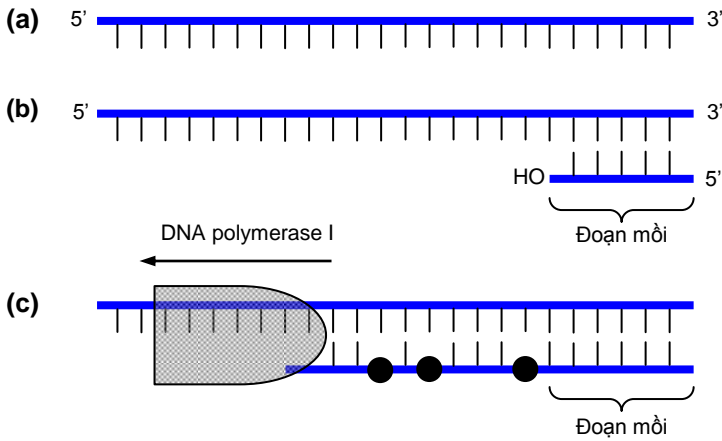
<sup>2</sup> DNase I (tụy của bò): enzyme xúc tác phản ứng thủy phân liên kết nằm ngay sau một pyrimidine trên chuỗi DNA. Trong trường hợp DNA sợi đôi, chỗ cắt có thể xảy ra trên một hay trên cả hai sợi (xem chương 1).

thuật PCR để đánh dấu phóng xạ đoạn DNA trong trường hợp muốn tạo ra một số lượng lớn mẫu dò. Lúc này, enzyme DNA polymerase I sẽ được thay bằng DNA Taq polymerase.



**Hình 5.10. Đánh dấu DNA bằng dịch chuyển điểm đứt.** (a) Dùng DNase I tạo ra một điểm đứt trên sợi đơn của chuỗi DNA sợi đôi. (b) Tiếp đó enzyme DNA polymerase I tổng hợp một bản sao mới của sợi khuôn mẫu khi phân hủy sợi có điểm đứt nhờ hoạt tính exonuclease 5' → 3' của nó. Nếu [ $\alpha$ - $^{32}$ P]dCTP được cung cấp thì nó sẽ gắn vào bản sao mới (các hình tròn bôi đen).

Trong phản ứng đánh dấu phóng xạ, thông thường cần tách DNA đã được đánh dấu khỏi các nucleotide không được đánh dấu còn thừa trong hỗn hợp phản ứng. Hoạt tính phóng xạ của mẫu dò cần được đánh giá bằng phương pháp đếm nhấp nháy lỏng (liquid scintillation). Các số liệu sau đó được dùng để tính toán lượng mẫu dò cần thiết cho các phản ứng lai phân tử (ví dụ: Southern blot, Northern blot, screening...).



**Hình 5.11. Đánh dấu DNA bằng cách kéo dài đoạn mồi.** (a) DNA được biến tính để tạo ra các phân tử sợi đơn. (b) Một đoạn mồi oligonucleotide được bổ sung để gắn với sợi DNA khuôn mẫu. (c) Enzyme DNA polymerase I xúc tác tổng hợp một bản sao mới của sợi khuôn mẫu bằng cách kéo dài đoạn mồi, trong quá trình đó các  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$  (các vòng bôi đen) sẽ được đính vào bản sao ở các vị trí có G.

## VII. Ứng dụng của thư viện genomic DNA

Thư viện genomic DNA có các ứng dụng trong phạm vi rộng để lập bản đồ vật lý của DNA và xác định các gen gây bệnh hoặc các chuỗi DNA quan tâm cho những phân tích xa hơn nữa.

Sản xuất ra các dòng mang các đoạn chèn DNA khác nhau nhưng gói chồng lên nhau có nhiều thuận lợi và thư viện có thể được sử dụng trong quá trình chromosome walking. Phương thức này cho phép xúc tiến từ điểm khởi đầu trên nhiễm sắc thể (ví dụ: gen chỉ thị liên kết với bệnh) tới locus không xác định ở gần đó (ví dụ: gen gây ra chính bệnh đó) bằng cách tiến hành các chu kỳ lặp lại của tạo dòng, mô tả đặc điểm và lai phân tử bằng cách dùng các dòng như là mẫu dò để xác định các dòng xa hơn với thư viện mang chuỗi gần kề từ DNA gốc. Chromosome walking thường được thực hiện với thư viện của cosmid,  $\lambda$  hoặc YAC.

Chromosome walking cũng cho phép xây dựng một cấu trúc “clone contig” (một chuỗi tuần tự các dòng DNA gói chồng lên nhau cùng với sự hiện diện của vùng liền kề của genomic DNA). Cấu trúc clone contig

thường tạo thành một phần cần thiết của việc phân tích các chuỗi DNA được tạo ra trong suốt quá trình xây dựng bản đồ vật lý của DNA và xác định các gen gây bệnh bằng tạo dòng vị trí (positional cloning). Cấu trúc clone contig được xây dựng bằng các vector khác nhau, bao gồm  $\lambda$ , BAC, YAC...

Xác định gen gây xơ nang đã minh họa cho việc sử dụng thư viện genomic DNA để xác định bằng cách tạo dòng vị trí. Các thư viện genomic DNA cũng hữu ích trong việc xác định các gen gây bệnh bằng cách tạo dòng chức năng (functional cloning). Theo hướng này, thông tin về chức năng của gen được khai thác để phân lập gen mong muốn từ thư viện. Một oligonucleotide có trình tự dựa trên chuỗi amino acid từng phần được dùng như là một mẫu dò để phân lập dòng cDNA bằng cách sàng lọc thư viện cDNA. Dòng cDNA này sau đó có thể được dùng trong sàng lọc thư viện genome để phân lập các dòng genomic DNA và cho phép quan sát đặc điểm của chuỗi genome hoàn chỉnh. Hướng này được dùng để xác định gen hemophilia A (nhân tố VIII). Các mẫu dò oligonucleotide dựa trên chuỗi amino acid của protein nhân tố VIII của lợn được sử dụng trước để phân lập dòng từ genome (nhân tố VIII của lợn), sau đó nó được dùng như là một mẫu dò cho thư viện DNA người để xác định gen ở người.

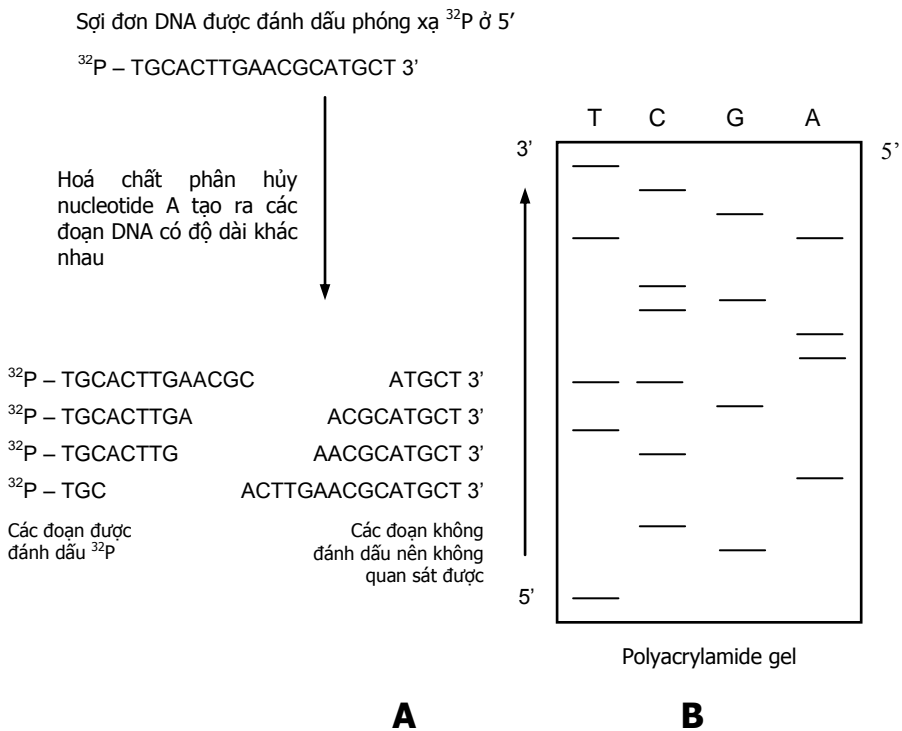
## **VII. Phân tích trình tự của đoạn DNA được tạo dòng**

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

### *1. Phương pháp hóa học Maxam-Gilbert*

Đầu tiên phân tử DNA sợi đôi được đánh dấu  $^{32}\text{P}$  tại đầu 5'. Sau đó, hai sợi đơn tách rời nhau do biến tính nhiệt. Chúng được xử lý với các chất hóa học sao cho chỉ một loại nucleotide bị phá hủy. Trên hình 5.12 đó là nucleotide A. Nồng độ các chất hóa học được kiểm tra chặt chẽ sao cho chỉ một nucleotide bị phá hủy trên mỗi sợi đơn DNA. Kết quả hàng loạt các đoạn DNA có kích thước khác nhau được tạo thành.

Các đoạn DNA tạo ra do bị bẻ gãy được phân ly trên polyacrylamide gel và chỉ có đoạn nào bị gắn  $^{32}\text{P}$  ở đầu 5' mới quan sát được. Kích thước của chúng tương ứng với khoảng cách từ đầu 5' có đánh dấu đến vị trí của các nucleotide bị bẻ gãy trên phân tử DNA. Sử dụng các chất hóa học khác nhau phá hủy bốn loại nucleotide bằng bốn phản ứng riêng biệt. Sản phẩm thu được của bốn phản ứng này được phân ly trên cùng một gel. Đoạn DNA ngắn nhất có đánh dấu phóng xạ chạy nhanh nhất dưới tác dụng của điện trường. Căn cứ vào vị trí các vạch trên gel mà xác định trình tự các nucleotide (Hình 5.12).



**Hình 5.12. Xác định trình tự nucleotide bằng phương pháp hóa học.** A: tiến hành bốn phản ứng độc lập, sử dụng bốn loại hóa chất khác nhau, mỗi hóa chất chỉ phá hủy một loại nucleotide nhất định (trong hình vẽ đó là A). B: Sản phẩm của bốn phản ứng được chạy điện di đồng thời trên polyacrylamide gel. Trình tự nucleotide đọc từ dưới lên theo chiều 5'→3'.

## 2. Phương pháp enzyme Sanger

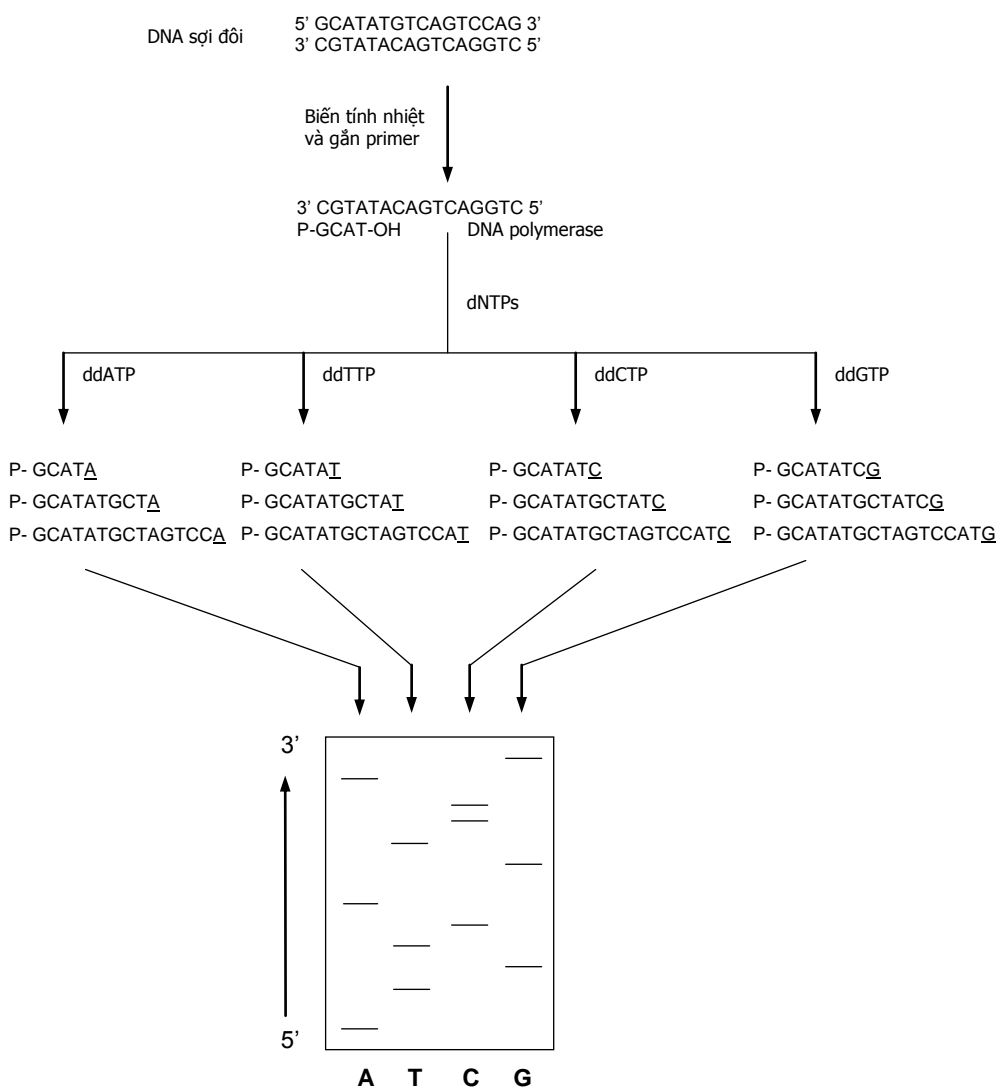
Đây là phương pháp tạo đầu tận cùng của chuỗi (chain terminator method). Nguyên lý của phản ứng này là sử dụng dideoxynucleotide không có nhóm OH ở vị trí 3' trong phản ứng tổng hợp DNA. Do đó, khi DNA polymerase gắn chúng vào sợi DNA thì quá trình tổng hợp bị ngừng lại. Vì vậy, phương pháp này còn gọi là phương pháp dideoxy (dideoxy method).

Đầu tiên sợi đôi DNA (sợi khuôn cần đọc trình tự) bị biến tính tách thành hai sợi đơn và được lai với primer. Primer là một sợi đơn gồm vài chục nucleotide có thể liên kết với một trong hai sợi đơn DNA theo nguyên tắc tạo cặp bổ sung. Sau đó, bốn phản ứng riêng rẽ được tiến hành đồng thời. Mỗi phản ứng là hỗn hợp của DNA khuôn mẫu, primer, DNA polymerase, một loại dideoxynucleotide và bốn loại deoxynucleotide theo tỷ lệ 1:100. Các sợi đơn DNA được tổng hợp có độ dài khác nhau do dideoxynucleotide gắn ngẫu nhiên làm dừng phản ứng. Sản phẩm của bốn phản ứng cùng được phân ly trên polyacrylamide gel cho phép phân biệt hai sợi đơn DNA hơn kém nhau một nucleotide. Các vạch DNA quan sát được nhờ có gắn phóng xạ  $^{32}\text{P}$  vào mỗi hoặc vào một trong bốn loại nucleotide bình thường (Hình 5.13).

## 3. Xác định trình tự nucleotide trên máy tự động

Trong những năm gần đây, một số phương pháp xác định trình tự mới đã xuất hiện. Bên cạnh kỹ thuật thông thường sử dụng các gel để phân ly các phân tử DNA có độ dài khác nhau, các kỹ thuật mới liên quan đến phát hiện huỳnh quang của các nucleotide được đánh dấu, phân tích trình tự DNA bằng khối phổ, điện di mao quản hoặc lai với các đoạn oligonucleotide được tổng hợp nhân tạo. Những tiến bộ nhanh chóng trong lĩnh vực kính hiển vi điện tử cho phép quan sát chi tiết cấu trúc bề mặt của phân tử nucleic acid và phân biệt từng nucleotide. Ngoài ra, kỹ thuật lai sử dụng tấm chip gắn cố định các oligonucleotide cho phép phân biệt được hơn 50.000 phân tử DNA khác nhau. Trong tương lai gần, kỹ thuật lai này có thể gắn hàng nghìn oligo trên một tấm chip nhỏ và trình tự nucleotide được phân tích tự động bằng các chương trình của computer. Điều đó cho phép xác định trình tự một cách nhanh chóng.





**Hình 5.13. Xác định trình tự nucleotide bằng phương pháp enzyme (Sanger).**

Dideoxynucleotide triphosphate (ddNTP) gắn vào sợi DNA đang tổng hợp sẽ làm ngừng phản ứng. Nhờ đó thu được các đoạn DNA hơn kém nhau chỉ một nucleotide. Các đoạn này được quan sát trên polyacrylamide gel khi primer được đánh dấu phóng xạ <sup>32</sup>P hoặc đánh dấu huỳnh quang.

Phương pháp tạo đầu tận cùng của chuỗi được cải tiến cho phép xác định trình tự trên máy đọc tự động (automatic sequencer), sử dụng các bộ

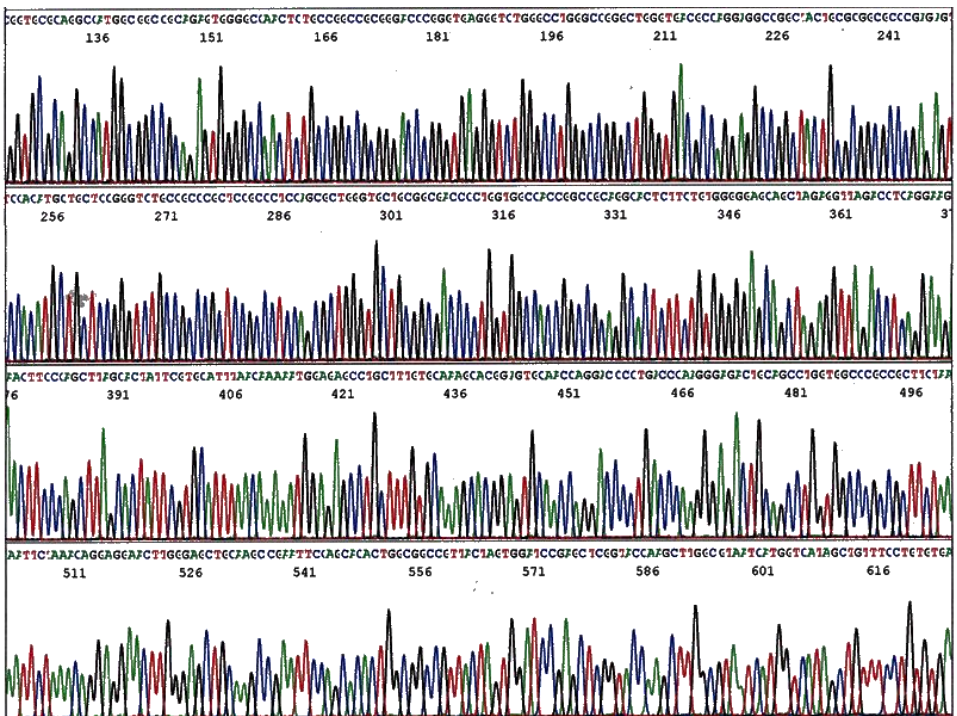
Kit khác nhau, trong đó đầu tận cùng được đánh dấu bằng các chất phát huỳnh quang màu (dye terminator). Phản ứng gắn các nucleotide đánh dấu vào sợi DNA tổng hợp trên khuôn mẫu được tiến hành trong máy PCR. Vì vậy, kỹ thuật này còn được gọi là phản ứng chuỗi đọc trình tự. Sau đó, sản phẩm PCR được điện di trên polyacrylamide gel có độ phân giải cao, cho phép phân biệt được các sợi đơn DNA hơn kém nhau một nucleotide. Quá trình chạy điện di được thực hiện trên máy tự động và kết quả được phân tích bằng các chương trình computer chuyên dụng (Hình 5.14).

Trên các máy đọc tự động hiện đại, thông thường bốn loại nucleotide được đánh dấu bằng các chất phát huỳnh quang khác nhau. Đầu đọc laser trong máy sẽ kích hoạt các chất này khiến chúng hấp thụ và phát ra các màu, mỗi màu ứng với một loại nucleotide và được hiển thị bằng một đỉnh. Thông thường, phản ứng gắn nucleotide vào sợi DNA tổng hợp trên sợi khuôn mẫu được thực hiện bằng PCR. Hai loại nucleotide dGTP và dTTP được thay thế bằng dITP và dUTP nhằm hạn chế sự trùng lặp các đỉnh. Sản phẩm PCR thường được tinh sạch, loại bỏ các nucleotide và primer thừa trước khi đưa vào máy đọc tự động. Trình tự đọc được trên máy tự động thường dài khoảng 600-1.000 bp. Ưu điểm của phương pháp này là kết quả đọc được đưa trực tiếp vào chương trình computer, giảm bớt sai sót do đọc bằng mắt gây ra. Hơn nữa, phản ứng PCR không đòi hỏi lượng DNA khuôn mẫu (dạng sợi đôi) nhiều nhưng các sợi đơn DNA cần đọc lại được khuếch đại lên rất nhiều lần.



**Hình 5.14. Máy phân tích trình tự nucleotide tự động và hình ảnh điện di các băng DNA phát huỳnh quang quan sát được trên computer**

Trình tự hệ gen cũng như các loại DNA được xác định ngày càng nhiều ở các sinh vật khác nhau. Do đó, việc lưu trữ số liệu, phân tích, so sánh và sắp xếp chúng đã thúc đẩy hình thành một hướng nghiên cứu mới đó là tin sinh học (bioinformatics). Mỗi liên quan mật thiết giữa trình tự nucleotide và protein đã khiến lĩnh vực mới này phát triển rất nhanh chóng. Ba ngân hàng dữ liệu chính hiện nay lưu trữ hầu hết các thông tin về DNA là EMBL (thuộc European Informatics Institute), GenBank (thuộc US National Centre for Biotechnology Information) và DDBJ (thuộc DNA Database Bank of Japan). Một ngân hàng dữ liệu khác lưu trữ thông tin về protein là Swiss Protein Database (Thụy Sĩ).



**Hình 5.15.** Minh họa trình tự nucleotide một đoạn DNA đã được phân tích trình tự

## Tài liệu tham khảo/đọc thêm

1. **Võ Thị Thương Lan.** 2002. Sinh học phân tử. *NXB Đại học Quốc gia Hà Nội*, Hà Nội.
2. **Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K.** 2002. Short Protocol in Molecular Biology. Vol 1 and 2. 5<sup>th</sup> ed. *John Wiley & Sons, Inc.* USA.
3. **Brown TA.** 2001. Gene Cloning-An Introduction. 4<sup>th</sup> ed. *Blackwell Science*, Oxford, UK.
4. **Glick BR and Pasternak JJ.** 2003. Molecular Biotechnology: Principles and Applications of Recombinant DNA. 3<sup>rd</sup> ed. *ASM Press*, USA.
5. **Maniatis T, Fritsch EF and Sambrook J.** 1989. Molecular Cloning-A Laboratory Manual. *Cold Spring Harbor Laboratory Press*, USA.
6. **Ohman DE.** 1989. Experiments in Gene Manipulation. *Prentice Hall*, Englewood Cliffs, New Jersey, USA.
7. **Primrose SB, Twyman R and Old RW.** 2001. Principles of Gene Manipulation. 6<sup>th</sup> ed. *Blackwell Science*, Oxford, UK.
8. **Rapley R and Walker JM.** 1998. Molecular Biomethods Handbook. *Humana Press Inc.* New Jersey, USA.
9. **Surzycki S.** 2000. Basic Techniques in Molecular Biology. *Springer-Verlag*, Berlin, Heidelberg, Germany.

## Chương 6

# Tạo dòng và xây dựng thư viện cDNA

## I. Tách chiết, tinh sạch và phân tích RNA

### 1. Tách chiết và tinh sạch RNA

#### 1.1. Tách chiết RNA tổng số

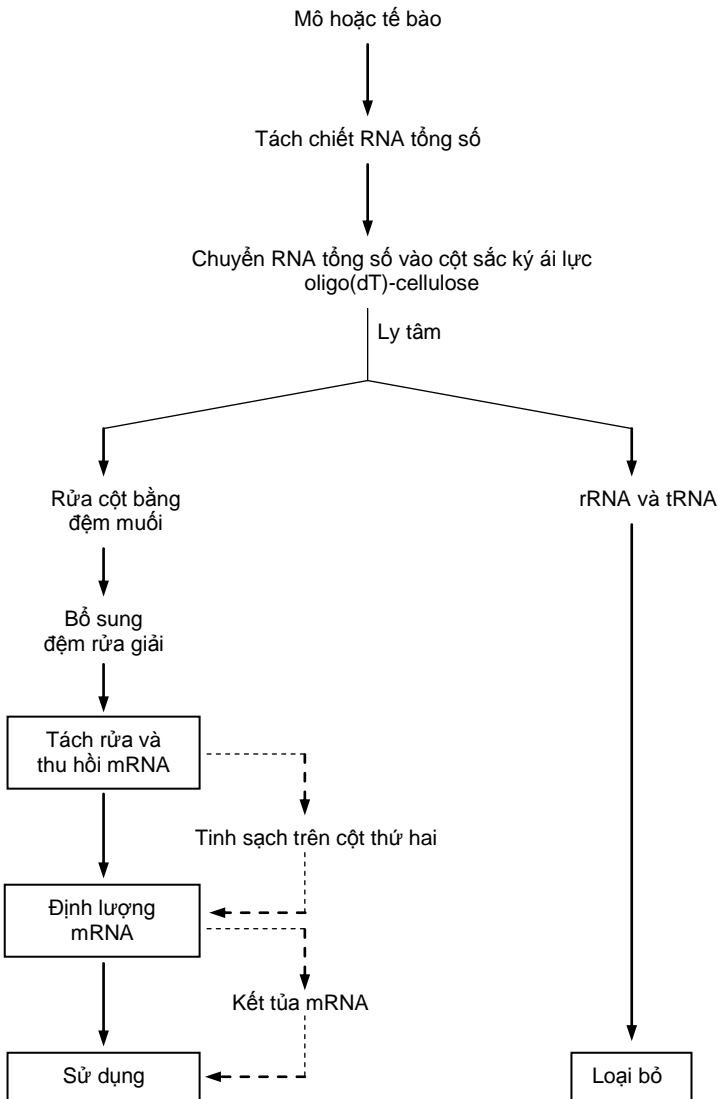
- **Kiểm soát hoạt tính ribonuclease.** Các phân tử RNA không bền, dễ bị phân hủy bởi các ribonuclease (RNase). Để thu được dịch chiết RNA có chất lượng tốt, cần phải giảm thiểu hoạt tính của RNase được giải phóng trong suốt quá trình sinh tan tế bào hoặc từ các nguồn tiềm tàng khác trong phòng thí nghiệm (dụng cụ, bàn làm việc, bàn tay của kỹ thuật viên...) bằng các nhân tố ức chế RNase, bao gồm các nhân tố ức chế protein của RNase, các phức hợp vanadyl-ribonucleoside hoặc macaloid.

- **Nguyên lý.** Phương pháp tách chiết RNA tổng số bao gồm các bước cơ bản giống như tách chiết DNA. Tế bào hoặc mô được nghiền trong dung dịch có chứa chất tẩy mạnh là SDS (sodium dodecyl sulphate, sarcocyl) ở nồng độ cao, các dung dịch muối mạnh (guanidinium thiocyanate-GTC) để hòa tan RNA và kết tủa các protein bị biến tính ở cùng một thời gian, và một chất khử ( $\beta$ -mercaptoethanol). Hai loại chất sau có tác dụng ức chế các RNase nội bào và tách các protein liên kết khỏi phân tử RNA. Bên cạnh đó, có thể bổ sung các protease (chẳng hạn protease K) để gây biến tính các thành phần protein của phức chất RNA-protein. Các protein được loại bỏ khỏi mẫu bằng cách chiết (ly tâm) với phenol, phenol:chloroform và chloroform. RNA hòa tan trong pha nước được kết tủa bằng ethanol (hoặc isopropanol) và được thu nhận lại qua ly tâm. RNA được bảo quản trên một năm ở  $-70^{\circ}\text{C}$  trong nước có chứa RNasin (nhân tố ức chế RNase).

#### 1.2. Phân lập RNA poly(A)<sup>+</sup>

RNA tổng số của tế bào chứa khoảng 90-99% rRNA và tRNA, trong đó rRNA chiếm khoảng 80-85% và tRNA chiếm khoảng 15-20%. Các RNA này có kích thước và trình tự xác định và có thể được tách riêng dễ dàng bằng điện di hoặc ly tâm. Riêng mRNA chỉ chiếm khoảng 1-5% RNA tổng

số của tế bào, lại có kích thước và trình tự rất đa dạng. Tuy nhiên, chúng có một điểm chung là cấu trúc đuôi polyA (có thể lên tới 100A). Dựa vào cấu trúc này và đặc tính liên kết bổ sung A-T của nucleic acid, có thể tách mRNA ra khỏi mẫu bằng sắc ký ái lực trên cột oligo(dT)-cellulose. Các mRNA sẽ bám lên cột thông qua liên kết bổ sung với oligo(dT) và được thu nhận lại qua ly tâm. Kỹ thuật này cho phép thu nhận mRNA từ những mẫu có khối lượng rất nhỏ.

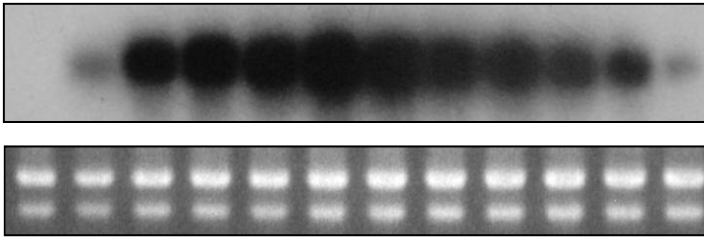


**Hình 6.1. Sơ đồ quy trình tinh sạch mRNA từ RNA tổng số**

## 2. Phân tích RNA

### 2.1. Lai Northern (Northern hybridization)

Phân tích RNA là công việc quan trọng của nhiều nghiên cứu về sinh học phân tử, vì nó có thể cung cấp thông tin về sự biểu hiện của gen trong cơ thể sống. Lai bằng màng lọc (nylon hoặc nitrocellulose) các RNA được phân đoạn với đoạn mồi là nucleic acid đánh dấu đồng vị phóng xạ  $^{32}\text{P}$  (hoặc digoxigenin-dUTP) được gọi là phương pháp thâm tích Northern (Northern blot). Phương thức này bắt nguồn từ phân tích Southern blot (xem chương 5), dựa trên cơ sở khả năng bổ sung của các nucleic acid sợi đơn để tạo thành các phân tử lai. Một cách tóm tắt, RNA được phân chia theo kích thước trên agarose gel dưới các điều kiện biến tính, thâm tích và liên kết không thuận nghịch với màng nylon hoặc nitrocellulose, lai với đoạn mồi nucleic acid được đánh dấu  $^{32}\text{P}$ . Sau khi rửa trôi đoạn mồi không liên kết hoặc liên kết không đặc hiệu, các phân tử RNA lai được phát hiện như là các băng phóng xạ tự ghi trên phim X-quang (Hình 6.2).



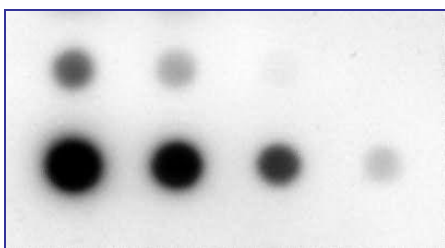
**Hình 6.2. Phân tích sự biểu hiện của gen bằng lai Northern.** Hình phía dưới là điện di RNA tổng số trên agarose gel được biến tính bằng formamide. Hình phía trên là tín hiệu phóng xạ thu được từ phản ứng lai giữa RNA tổng số với mẫu dò được đánh dấu bằng  $^{32}\text{P}$ .

Thâm tích Northern là phương pháp quan trọng để nghiên cứu biểu hiện của gen, do các RNA hiện diện trong tế bào hoặc loại mô quy định mô tả khái quát một phần của genome được biểu hiện trong tế bào hoặc mô đó. Phân tích Northern cho thấy mức độ ổn định của sự tích lũy chuỗi RNA qui định (thường là mRNA) trong mẫu nghiên cứu. Vì thế, phân tích Northern cho phép người ta liên kết các mức độ biểu hiện mRNA với các tính chất hình thái và sinh lý của cơ thể sống. Kỹ thuật lai Northern là một phương

pháp được ứng dụng rộng rãi cho việc phân tích biểu hiện gen. Ví dụ: kỹ thuật này đã được sử dụng cho việc mô tả đặc điểm của các cDNA và gen được tạo dòng, nghiên cứu đặc trưng cơ quan/mô của sự biểu hiện gen, phân tích hoạt tính của các gen nội và ngoại sinh trong cơ thể chuyển gen, và nghiên cứu sự biểu hiện của gen trong mối liên quan với sự phát triển và các yếu tố vô sinh và hữu sinh.

## 2.2. Lai Dot (Dot hybridization)

Lai dot được Kafatos và cs. (1979) thực hiện đầu tiên bằng cách chấm các mẫu nhỏ của dung dịch RNA lên màng nitrocellulose, sau đó màng được làm khô, lai với mẫu dò đặc trưng RNA hoặc DNA có đánh dấu  $^{32}\text{P}$ , và ủ với phim X-quang. Mặc dù khó định lượng chính xác do kích thước của các chấm lớn và khác nhau, nhưng trong nhiều trường hợp có thể thu được một ý tưởng tốt về cường độ biểu hiện của một gen đặc biệt nào đó trong các mô đặc trưng hoặc các tế bào nuôi cấy (Hình 6.3).



**Hình 6.3. Phân tích dot để định lượng RNA.** Hàng phía dưới là các nồng độ RNA chuẩn đã biết. Hàng trên là các nồng độ RNA khác nhau được tách chiết từ mô/cơ quan.

## II. Tổng hợp cDNA (complementary DNA)

Phương pháp tổng hợp cDNA có nhiều thuận lợi do: (1) Lấy mRNA đúng của gen. (2) Không tốn công phân tích những đoạn gen trùng lặp. (3) Chọn lọc dễ dàng trong một số trường hợp.

Ở đây chỉ trình bày phương pháp tổng hợp cDNA. Quá trình tổng hợp cDNA qua ba giai đoạn như sau: (1) Tổng hợp sợi thứ nhất bằng enzyme reverse transcriptase. (2) Biến tính và cắt RNA trong thể lai mRNA-cDNA



tuần tự bằng nhiệt và enzyme RNase H của *E. coli*. Thay thế khuôn mẫu là chuỗi RNA bằng chuỗi cDNA thứ nhất và tổng hợp sợi cDNA thứ hai nhờ DNA polymerase I của *E. coli*. (3) Cắt vòng cặp tóc của cDNA sợi đôi nhờ nuclease S1 và sửa chữa hai đầu bằng đoạn Klenow (Hình 6.4).

## 1. Tổng hợp sợi cDNA thứ nhất

### 1.1. Enzyme phiên mã ngược (reverse transcriptase)

Nhân tố quan trọng nhất trong tổng hợp cDNA là chất lượng của enzyme phiên mã ngược được sử dụng trong phản ứng. Mặc dù chất lượng của enzyme được sử dụng nói chung là tốt, song vẫn khó loại trừ được RNase bản. Trở ngại này có thể khắc phục bằng cách dùng các nhân tố ức chế mạnh RNase (RNase inhibitor), ví dụ như: vanadyl-ribonucleoside hoặc RNasin, trong phản ứng phiên mã ngược.

Tỷ lệ enzyme phiên mã ngược dùng cho khuôn mẫu mRNA cũng có ảnh hưởng khác nhau đến hiệu suất tổng hợp cDNA hoàn chỉnh. Nếu số lượng khuôn mẫu nhiều thì hiệu suất của sản phẩm phiên mã cDNA sẽ tăng lên cùng với việc tăng số lượng enzyme reverse transcriptase. Trong nghiên cứu, hiệu suất tối đa của các sản phẩm phiên mã cDNA hoàn chỉnh đạt đến 80 tiểu phần enzyme/ $\mu\text{g}$  của khuôn mẫu, tỷ lệ enzyme cao như thế đòi hỏi phải sử dụng enzyme tinh sạch cao và bao gồm cả các nhân tố ức chế RNase trong phản ứng.

### 1.2. Oligo dT primer

Các oligo dT primer được sử dụng để làm mồi cho phản ứng tổng hợp sợi cDNA thứ nhất dựa trên khuôn mẫu của mRNA. Các oligo dT primer sẽ liên kết với đuôi polyA của mRNA.

### 1.3. Deoxynucleoside triphosphate (dNTP)

Nồng độ của bốn loại dNTP là yếu tố đặc biệt quan trọng ảnh hưởng đến hiệu suất tổng hợp cDNA. Nếu nồng độ của một trong bốn dNTP giảm xuống dưới 10-50  $\mu\text{M}$  thì hiệu suất của sản phẩm phiên mã giảm rõ rệt. Khi sử dụng RNA của avian myeloblastosis virus làm khuôn mẫu thì việc sản xuất các cDNA hoàn chỉnh tăng tối đa ở nồng độ 75  $\mu\text{M}$

của cả bốn loại dNTP. Nồng độ của dNTP từ 200-250  $\mu\text{M}$  đang được dùng phổ biến.

#### 1.4. Các cation hóa trị 1 và hóa trị 2

Các điều kiện ion ảnh hưởng thật sự đến hiệu quả phiên mã của các khuôn mẫu khác nhau. Đối với các sản phẩm phiên mã dài thì ion  $\text{K}^+$  có hiệu quả cao hơn ion  $\text{Na}^+$ . Nồng độ  $\text{K}^+$  tối ưu cho quá trình tổng hợp và kéo dài kích thước cDNA là khoảng từ 140-150 mM. Các cation hóa trị 2 là yêu cầu tuyệt đối cho hoạt tính của enzyme phiên mã ngược. Không có hoạt tính nào được tìm thấy ở nồng độ  $<4$  mM của ion  $\text{Mg}^{2+}$ , nồng độ ion  $\text{Mg}^{2+}$  tối ưu cho quá trình sản xuất các sản phẩm phiên mã hoàn chỉnh là khoảng 6-10 mM.

#### 1.5. pH của dung dịch phản ứng

Giá trị pH 8,3 là tối ưu để sản xuất hiệu quả các sản phẩm phiên mã cDNA hoàn chỉnh. Trong quá trình sản xuất các sản phẩm phiên mã, một số đệm đã được thử nghiệm nhưng không tốt hơn đệm Tris.HCl.

### 2. Tổng hợp sợi cDNA thứ hai

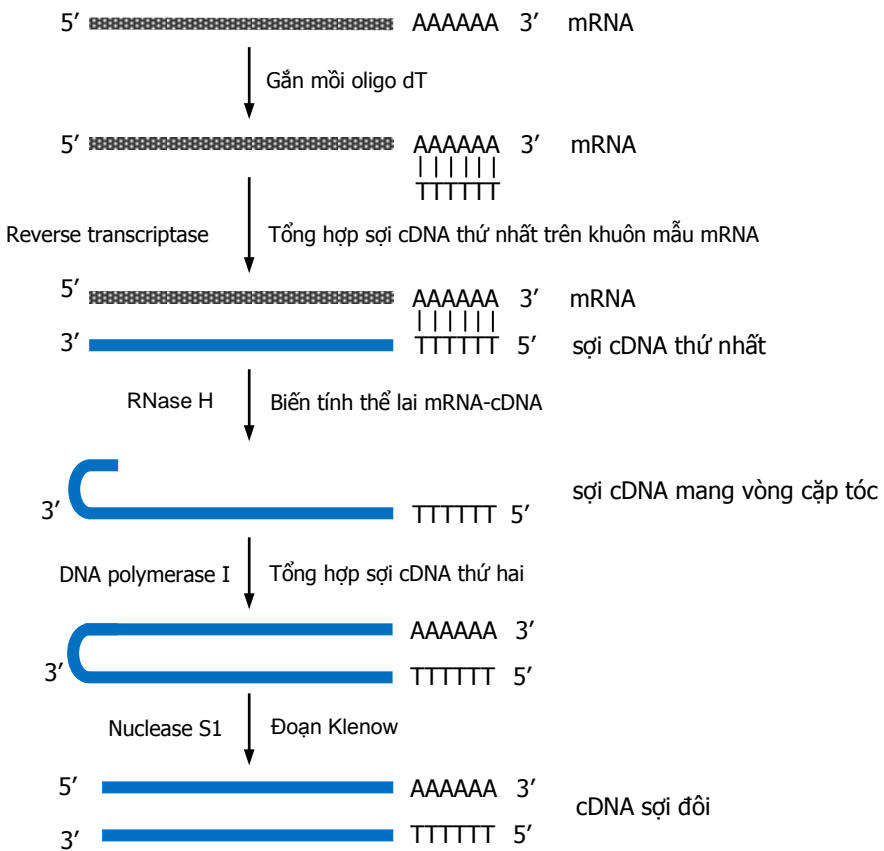
Gây biến tính (đun sôi) thể lai mRNA-cDNA để cắt RNA bằng RNase H của *E. coli*. Thông thường, đầu tận cùng 3' của các cDNA sợi đơn có khả năng tạo thành các cấu trúc vòng cặp tóc (hairpin loop)<sup>1</sup> và vì thế có thể được sử dụng để làm mồi (primer) cho quá trình tổng hợp sợi cDNA thứ hai bằng DNA polymerase I của *E. coli* hoặc reverse transcriptase (Hình 6.4). Mặc dù người ta đã dự đoán cấu trúc vòng đôi ở đầu tận cùng của các cDNA và cơ chế phát sinh của chúng, nhưng hiện tượng tổng hợp sợi cDNA thứ hai vẫn chưa được nghiên cứu một cách hệ thống.

Tổng hợp sợi cDNA thứ hai bằng DNA polymerase I đã được sử dụng rộng rãi. Đoạn Klenow của DNA polymerase I thiếu hoạt tính exonuclease 5 $\rightarrow$ 3' cũng được sử dụng thành công để tổng hợp sợi cDNA thứ hai.

---

<sup>1</sup> Vòng cặp tóc: hình dạng của phân tử DNA hoặc RNA được dùng làm mồi cho quá trình tổng hợp sợi thứ hai.

Nhiều tác giả đã sử dụng reverse transcriptase để tổng hợp sợi cDNA thứ hai. Mặc dù có tác giả cho rằng AMV reverse transcriptase không thể dùng để tổng hợp sợi thứ hai của cDNA immunoglobulin, nhưng thành công của nhiều thí nghiệm dùng reverse transcriptase để tổng hợp sợi cDNA thứ hai đã cho thấy việc sử dụng cả hai enzyme đều có thể cho kết quả tốt. DNA polymerase I và reverse transcriptase có thể tạm ngừng hoặc ngừng lại ở các chuỗi khác nhau. Vì vậy, các sợi thứ hai được tổng hợp một cách đặc biệt, chúng được sản xuất nhờ một enzyme và được mở rộng hoàn toàn nhờ một enzyme khác.



**Hình 6.4. Sơ đồ tổng hợp đoạn cDNA từ khuôn mẫu mRNA**

### 3. Cắt vòng cặp tóc nhờ nuclease S1

Sau khi tổng hợp cDNA hoàn toàn, sợi thứ nhất và thứ hai được liên kết cộng hóa trị bởi vòng cặp tóc và vòng cặp tóc dễ bị cắt bởi nuclease S1. Sau đó, đoạn cDNA được sửa chữa bằng enzyme Klenow, kết quả là hai đầu tận cùng là đầu bằng.

Sợi đôi cDNA sau đó được tách thành các tiểu phần theo kích thước và các phân tử lớn nhất được gắn vào các plasmid của vi khuẩn. Hoặc là một tập hợp đầy đủ các kích thước của cDNA sợi đôi được tạo dòng trong bacteriophage  $\lambda$  để xây dựng thư viện cDNA (cDNA library). Tuy nhiên, việc đưa các đầu bằng vào vector sẽ gây khó khăn trong việc lấy chúng ra khỏi vector một cách nguyên vẹn sau này. Do đó các linker thường được nối vào hai đầu của các cDNA nhờ DNA ligase. Linker là những đoạn nucleotide ngắn có chứa vị trí nhận biết của một loại RE (ví dụ: *EcoRI*) được tổng hợp nhân tạo tương ứng với vị trí nhận biết RE (ví dụ: *EcoRI*) của vector. Sau đó, các cDNA mang linker và vector sẽ được cắt bởi cùng một enzyme (ví dụ: *EcoRI*). Nhờ đó các cDNA và vector đều có đầu sole tương đồng (đầu dính) và cDNA sẽ dễ dàng gắn cũng như lấy ra khỏi vector một cách nguyên vẹn.

## III. Tạo dòng phân tử của cDNA sợi đôi

Có nhiều phương pháp khác nhau được dùng để liên kết cDNA sợi đôi với các plasmid vector. Đa số được dùng theo các phương pháp sau:

- **Bổ sung các đuôi đồng trùng hợp (homopolymetric tailing)** cho cDNA sợi đôi và cho DNA của plasmid vector. DNA vector và cDNA sau đó được nối bằng liên kết hydrogen giữa các đuôi đồng trùng hợp bổ sung. Sự tạo thành vòng DNA đóng bằng enzyme gắn *in vitro* (DNA ligase) là cần thiết để hình thành nên các plasmid tái tổ hợp trong *E. coli*.

- **Bổ sung các đoạn nối nhân tạo (synthetic linkers)** cho đầu tận cùng của DNA sợi đôi. Sau khi phân cắt bằng RE thích hợp, các phân tử DNA được chuyển vào trong plasmid DNA cũng đã được cắt với cùng một enzyme.

### 1. Đuôi đồng trùng hợp

#### 1.1. Đuôi dA:dT

Enzyme terminal transferase xúc tác cho sự bổ sung của các deoxyrinucleotide triphosphate (dNTP) vào đầu 3'-OH của DNA sợi đôi

hoặc sợi đơn, được dùng để đưa DNA tái tổ hợp vào trong *E. coli* bằng cách nối dA:dT. Thông thường từ 50-150 gốc dA được bổ sung vào vector DNA và một số tương ứng của các gốc dT vào cDNA sợi đôi, cDNA sợi đôi được đưa vào trong các plasmid vector qua phương thức nối dA:dT. Tuy nhiên, đuôi đồng trùng hợp dA:dT ít khi được dùng để tạo dòng cDNA, lý do chính là không có phương thức thích hợp để cắt cDNA gắn trong plasmid nhờ đuôi dA:dT.

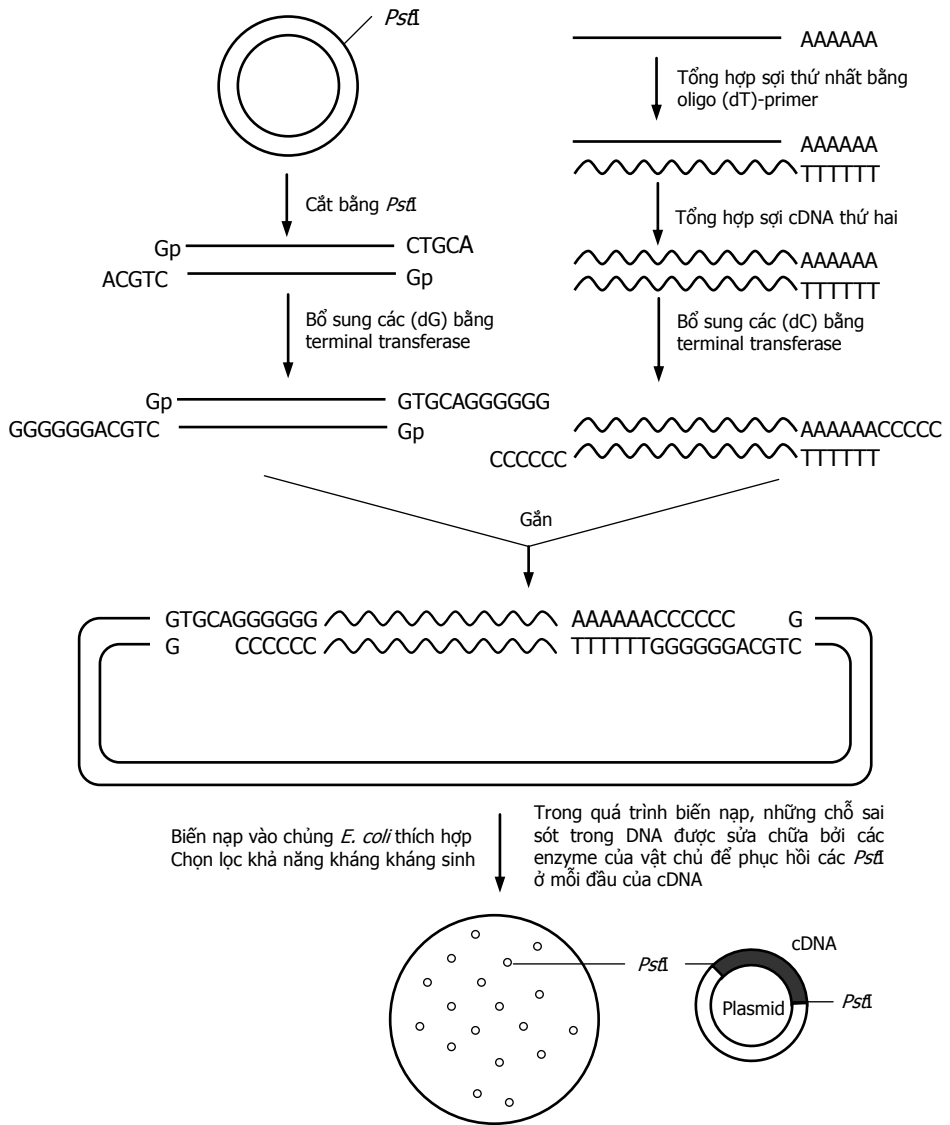
## 1.2. Đuôi dC:dG

Phương pháp được sử dụng rộng rãi hơn để tạo dòng các cDNA bằng đuôi đồng trùng hợp đòi hỏi bổ sung các đuôi dC cho cDNA sợi đôi và các đuôi dG được bổ sung vào plasmid vector đã được cắt hạn chế bằng *Pst*I. Enzyme *Pst*I cắt chuỗi 5'...CTGCAG...3' tạo ra đầu 3' tận cùng là cơ chất lý tưởng cho việc bổ sung các đuôi đồng trùng hợp. Các dòng cDNA mang các đuôi dC:dG có thể dễ dàng tách ra khỏi plasmid bằng cách thủy phân nhờ *Pst*I (Hình 6.5).

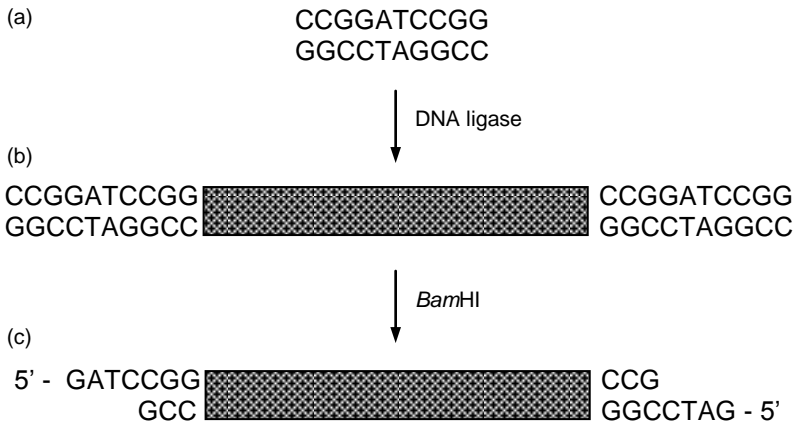
Số lượng các gốc dA:dT và dC:dG cần thiết để cho hiệu suất gắn tối thích cDNA vào plasmid vector đã được xác định, số lượng các gốc trên plasmid và cDNA phải xấp xỉ nhau, với khoảng 100 gốc được bổ sung tới mỗi DNA để nối dA:dT và khoảng 20 gốc để nối dC:dG.

## 2. Các linker và adapter nhân tạo

Các linker chứa một hoặc nhiều vị trí cắt hạn chế cho phép nối cDNA sợi đôi với các plasmid vector hoặc bacteriophage  $\lambda$  vector. cDNA sợi đôi được xử lý với DNA polymerase của bacteriophage T4 hoặc DNA polymerase I của *E. coli*, các enzyme này loại bỏ đầu tận cùng 3' sợi đơn so le bằng hoạt tính exonuclease 3'  $\rightarrow$  5' và lấp đầy các đầu tận cùng 3'-OH bị khuyết bằng hoạt tính trùng hợp (polymerization). Sự phối hợp của các hoạt tính này đã tạo ra các phân tử DNA đầu bằng, sau đó các cDNA này được ủ với một số lượng lớn các phân tử linker với sự có mặt của bacteriophage T4 DNA ligase (enzyme xúc tác cho quá trình gắn của các phân tử DNA đầu lồi với linker) (Hình 6.6).



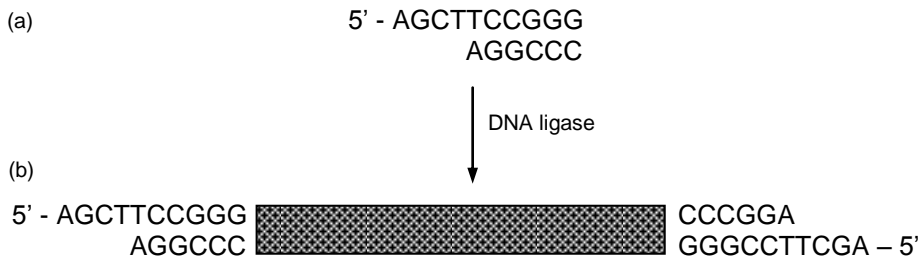
**Hình 6.5. Tạo dòng cDNA sợi đôi bằng đuôi đồng trùng hợp dG:dC.** Enzyme terminal transferase có hoạt tính tạo nhóm homopolymer ở đầu 3'-OH của DNA sợi đôi làm lòi ra một đầu ở cuối cơ chất của nó là DNA sợi đơn.



**Hình 6.6. Minh họa một linker.** (a) Đoạn 5'-CCGGATCCGG-3' mang vị trí nhận biết cho *Bam*HI. (b) Linker này được gắn vào cDNA đầu bằng nhờ DNA ligase. (c) Cấu trúc này sau đó được cắt bằng *Bam*HI để tạo ra đầu 5' lồi.

Các phân tử DNA sợi đôi mang các đầu kết dính nhân tạo được cắt ở vị trí cắt hạn chế trong linker, tinh sạch, và sau đó gắn với vector cũng được cắt bằng RE tương ứng tạo ra các đầu dính tương đồng với các đầu của linker. Có thể hạn chế sự tái tạo lại vòng của plasmid vector (không tái tổ hợp) bằng cách xử lý các vector được cắt bằng enzyme phosphatase (calf intestinal alkaline phosphatase-CIAP) trước khi thực hiện phản ứng gắn với cDNA.

Việc sử dụng adapter có hiệu quả hơn linker. Các adapter có kích thước ngắn, là các oligonucleotide sợi đôi mang một đầu bằng (để gắn với cDNA sợi đôi) và một đầu tận cùng kết dính (để gắn với đầu tận cùng tương ứng trong vector). Không giống như linker, các adapter không đòi hỏi phải cắt bằng các RE sau khi chúng được gắn với cDNA sợi đôi. Tuy nhiên, các phân tử cDNA mang các adapter được phosphoryl hóa (phosphorylation) có thể tạo thành các phân tử mạch vòng đóng bằng liên kết cộng hóa trị (dạng không thể tạo dòng) hoặc các phân tử mạch thẳng dạng khảm (dạng hoàn toàn không mong muốn) trong suốt phản ứng gắn tuần tự với sự có mặt của vector DNA đã được dephosphoryl hóa (Hình 6.7).



**Hình 6.7. Minh họa một adapter.** (a) Adapter, có đầu tận cùng 5' tương ứng với enzyme *Hind*III, được gắn vào cDNA đầu bằng nhờ DNA ligase. (b) Đầu 5' của adapter có thể được dephosphoryl hóa để ngăn cản sự tự kết nối lại.

Khi nối các linker hoặc adapter với các phân tử cDNA sợi đôi đầu bằng, phản ứng gắn cần được tiến hành trong một dung tích tối thiểu (để duy trì một nồng độ cao của linker/adapter). Nồng độ phân tử của linker/adapter phải lớn hơn nồng độ của đầu tận cùng của cDNA ít nhất là 100 lần. Cuối cùng, trước khi đoạn cDNA được gắn vào vector, các adapter không có phản ứng và các sản phẩm có trọng lượng phân tử thấp (do phản ứng cắt hạn chế tạo ra) cần được loại bỏ bằng sắc ký cột, điện di agarose hoặc polyacrylamide gel.

#### IV. Các phương pháp tạo dòng cDNA khác

##### 1. Tạo dòng mRNA-cDNA

Một phương pháp khác để tạo dòng cDNA là biến nạp vào *E. coli* các thể lai mRNA-cDNA đã được gắn với các plasmid vector. Các vi khuẩn vật chủ loại bỏ mRNA và thay thế nó bằng DNA. Sau khi sợi cDNA đầu tiên được tổng hợp theo cách thông thường, các gốc dA được bổ sung cho thể lai mRNA-cDNA sau đó được ủ với plasmid mang các đuôi dT. Do đầu tận cùng 3'-OH của RNA kém ít nhất 10 lần trong phản ứng tương đồng với DNA nên hầu hết các gốc dA được bổ sung cho thể lai đều phối hợp ở đầu 3' của DNA. Việc nối các đầu khác của thể lai với vector có khả năng thành công do môi liên kết hydrogen giữa poly(A) tự nhiên ở đầu 3' của mRNA và plasmid có đuôi dT. Phương pháp này có một số thuận lợi sau:



- Không cần tổng hợp sợi cDNA thứ hai.
- Cắt vòng cặp tóc DNA bằng nuclease S1 là không cần thiết.

Tuy nhiên, hạn chế của phương pháp này là có hiệu quả kém ít nhất 10 lần so với phương pháp tạo dòng cDNA sợi đôi và vì thế không thích hợp cho việc xây dựng một số lượng lớn các dòng cDNA.

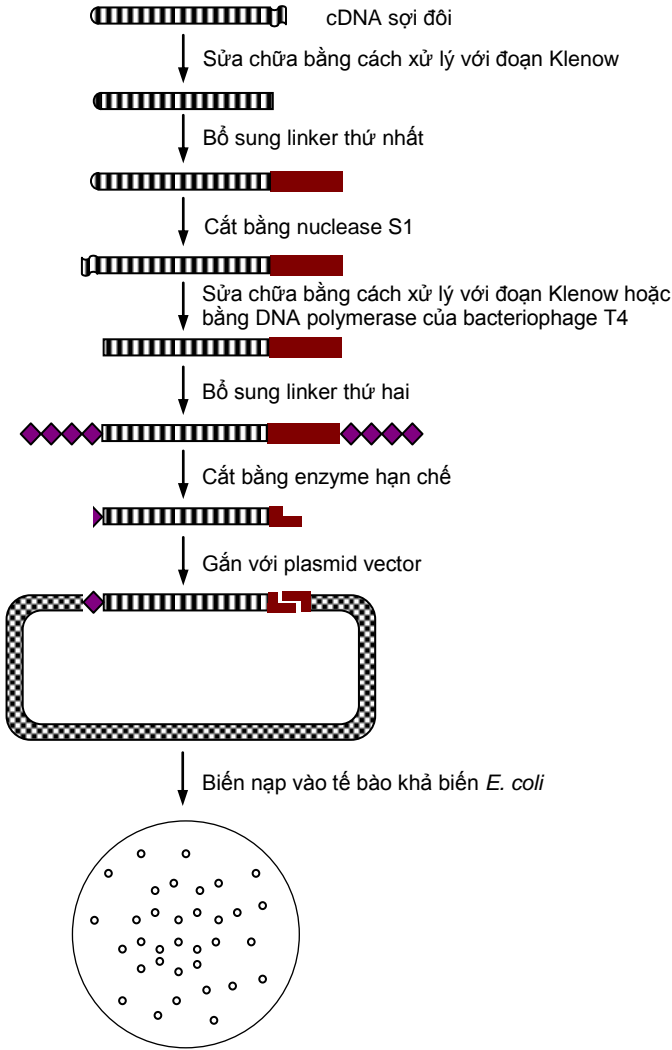
## 2. Bổ sung tuần tự các linker khác nhau

cDNA sợi đôi, được tổng hợp bằng cơ chế tự mồi (self-priming), được gắn vào một loại linker nhân tạo trước khi vòng cặp tóc trong cDNA bị cắt. Vì thế, các linker chỉ được bổ sung ở một đầu của cDNA sợi đôi (đầu tương ứng với đầu tận cùng 3' của mRNA). Sau đó cDNA được tinh sạch khỏi các linker thừa, vòng cặp tóc được cắt bằng nuclease S1 và sửa chữa bằng đoạn Klenow của DNA polymerase I của *E. coli* trước khi loại linker thứ hai được gắn vào cDNA. Các linker này sẽ được gắn vào cả hai đầu của cDNA. cDNA được gắn linker kép sau đó được xử lý với các RE thích hợp và gắn vào trong vector bằng phương pháp tạo dòng định hướng (Hình 6.8).

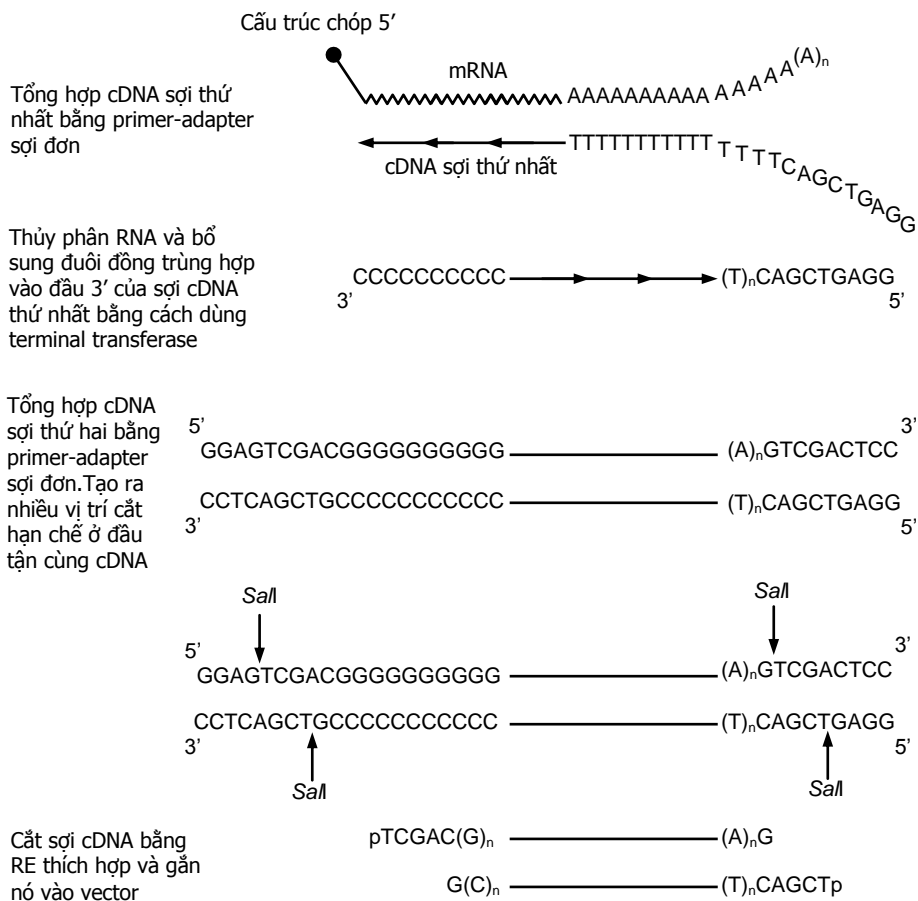
Phương pháp này được dùng để gắn cDNA theo hướng chính xác của các promoter cho phép biểu hiện các đoạn chèn (inserted sequences) trong vi khuẩn. Các bacteriophage  $\lambda$  vector cho phép tạo dòng định hướng cũng đã được phát triển trong thời gian gần đây.

## 3. Tạo dòng cDNA bằng các primer-adapter

Ngày nay, việc sản xuất dễ dàng các oligonucleotide primer của các trình tự xác định đã cho phép phát triển các phương pháp tạo dòng sử dụng primer-adapter chứa một vùng của DNA đồng trùng hợp ở đầu tận cùng 3' và một vị trí cắt hạn chế ở đầu tận cùng 5'. Các chuỗi đồng trùng hợp được dùng làm mồi để tổng hợp sợi thứ nhất và sợi thứ hai của cDNA, và các vị trí cắt hạn chế bên cạnh được sử dụng để đưa các phân tử cDNA sợi đôi cuối cùng vào trong một vector thích hợp. Trong mô hình đơn giản nhất, phương pháp này cho phép các cDNA được tạo dòng với hiệu suất cao (Hình 6.9).



**Hình 6.8. Tạo dòng cDNA bằng cách bổ sung tuần tự các linker nhân tạo.** Đoạn Klenow tạo đầu bằng cho cDNA sợi đôi và enzyme nuclease S1 cắt vòng cặp tóc. Các linker thứ nhất và thứ hai được gắn tuần tự vào hai đầu của cDNA, sau đó đoạn cDNA này sẽ được cắt cùng enzyme hạn chế với vector tạo dòng có vị trí nhận biết trên hai linker nhân tạo. Cuối cùng, đoạn cDNA có hai đầu tương đồng được gắn với vector và biến nạp vào *E. coli*.



**Hình 6.9. Tổng hợp cDNA sợi đôi bằng phương pháp primer-adaptor**

Đoạn cDNA được tổng hợp theo phương pháp trên khi gắn vào trong các vector biểu hiện như  $\lambda$ gt20 và  $\lambda$ gt22 chỉ có 1/6 cơ hội biểu hiện trong vi khuẩn, lý do: chỉ một nửa các phân tử cDNA được đưa vào trong vector theo hướng chính xác đối với promoter *lacZ*, và chỉ một trong ba phân tử được gắn ở hướng phải sẽ ở trong khung đọc chính xác để sản xuất protein hợp nhất.

## V. Các bacteriophage $\lambda$ vector dùng trong tạo dòng cDNA

### 1. Các bacteriophage $\lambda$ vector được dùng phổ biến

Hai loại vector biểu hiện của bacteriophage  $\lambda$  thường dùng để xây dựng thư viện cDNA là  $\lambda$ gt10 và  $\lambda$ gt11. Vector  $\lambda$ gt10 dùng để xây dựng các

thư viện chỉ được sàng lọc bằng các mẫu dò nucleic acid, trong khi vector  $\lambda$ gt11 dùng để xây dựng các thư viện được sàng lọc bằng các mẫu dò miễn dịch để phân lập các chuỗi DNA mã hóa các kháng nguyên đặc hiệu.

### 1.1. Vector $\lambda$ gt10

$\lambda$ gt10 là vector được thiết kế để nhận các đoạn DNA ngoại lai trong vùng mã hóa của gen ức chế  $cI$ <sup>2</sup>. Các vector này mang gen  $cI$ , giống như bacteriophage  $\lambda$ , nó tạo thành các vết tan mờ đục trên hầu hết các chủng *E. coli*. DNA của  $\lambda$ gt10 mang vị trí nhận biết đơn *EcoRI* nằm trong vùng mã hóa của gen  $cI$  là nơi mà các đoạn DNA có thể gắn vào. Kết quả chèn đoạn DNA sẽ làm bất hoạt gen  $cI$ , sản sinh ra các bacteriophage  $cI$  tái tổ hợp và tạo thành các plaque màu sáng, dễ dàng phân biệt với các plaque mờ đục của  $\lambda$ gt10 bố mẹ.

DNA của  $\lambda$ gt10 bố mẹ xấp xỉ 43 kb và như vậy có thể nhận các đoạn DNA ngoại lai dài tới 7,6 kb. Các thư viện cDNA được xây dựng trong  $\lambda$ gt10 luôn chứa hỗn hợp các bacteriophage tái tổ hợp và không tái tổ hợp. Hai loại bacteriophage này có thể phân biệt kiểu hình nhờ vào khả năng tạo thành các plaque của chúng.

### 1.2. Vector $\lambda$ gt11

$\lambda$ gt11 là vector biểu hiện mang bản sao của gen *lacZ* của *E. coli*, có vị trí nhận biết đơn *EcoRI* nằm ở vùng ngược hướng 53 bp của codon kết thúc dịch mã của gen *lacZ*. DNA ngoại lai có kích thước xấp xỉ 7,2 kb có thể được gắn ở vị trí này. Các chuỗi mã hóa gắn trong khung đọc theo hướng chính xác sẽ được biểu hiện để sản xuất các protein dung hợp (fusion protein) có đầu tận cùng amino chứa các chuỗi  $\beta$ -galactosidase và đầu tận cùng carboxyl chứa mạch polypeptide ngoại lai. Một số protein dung hợp này sẽ thể hiện các yếu tố quyết định kháng nguyên (epitopes) được phát hiện bởi khả năng phản ứng với các kháng thể của chúng.

Các protein dung hợp mang các yếu tố quyết định kháng nguyên ngoại lai có thể được phát hiện bằng cách sàng lọc các plaque với mẫu dò miễn dịch. Bacteriophage được dàn trải ở 42°C trên *E. coli*. Sau khoảng 4 giờ, các

---

<sup>2</sup>  $cI$  repressor: gen ức chế có sản phẩm protein liên kết với vùng operator hoặc promoter của gen, và kìm hãm phiên mã bằng cách ngăn chặn sự liên kết của RNA polymerase.

đĩa petri được chuyển đến nhiệt độ 37°C và được phủ màng lọc vô trùng nitrocellulose có thấm isopropylthio- $\beta$ -D-galactoside (IPTG), chất làm bất hoạt gen ức chế *lac* và cảm ứng biểu hiện protein ngoại lai. Sau một vài giờ nuôi ở 37°C, các màng lọc được ủ với kháng thể để liên kết với kháng nguyên quan tâm. Các kháng thể này sau đó được phát hiện bằng các xét nghiệm hóa học phóng xạ (radiochemistry) hoặc hóa học mô (histochemistry).

## 2. Một số bacteriophage $\lambda$ vector khác

### 2.1. Vector $\lambda$ gt18 và $\lambda$ gt19

Hai loại vector  $\lambda$ gt18 và  $\lambda$ gt19 có nguồn gốc từ vector  $\lambda$ gt11 được cải tiến mang vùng tạo dòng (polycloning sites) ở vị trí *EcoRI* đơn của  $\lambda$ gt11. Có thể sử dụng hai loại vector này để tạo dòng định hướng và không định hướng cDNA. Hơn nữa, ít nhất một trong các vị trí của vùng tạo dòng (*SalI*) ít khi xuất hiện trong DNA động vật có vú (trung bình 1 vị trí/đoạn 100 kb), và vì thế ít có khả năng các dòng cDNA có chứa vị trí *SalI*. Như vậy, các cDNA sợi đôi có đầu bằng mang các linker *SalI* không cần phải được bảo vệ bằng cách methyl hóa (methylation) trước khi được thủy phân bằng RE. Khi các đoạn linker được loại bỏ, sẽ dẫn đến kết quả quần thể các phân tử cDNA có thể được gắn trực tiếp với các nhánh của  $\lambda$ gt18 hoặc  $\lambda$ gt19.

### 2.2. Vector $\lambda$ gt20 và $\lambda$ gt21

Các vector  $\lambda$ gt20 và  $\lambda$ gt21 có nguồn gốc tương ứng từ  $\lambda$ gt18 và  $\lambda$ gt19, và như vậy có nhiều tính chất giống như đã mô tả ở  $\lambda$ gt18 và  $\lambda$ gt19. Những thay đổi đặc biệt so với  $\lambda$ gt18 và  $\lambda$ gt19 như sau:

- Vị trí tổng hợp *chi* được đưa vào vector cho phép loại bỏ sự chọn lọc đoạn chèn cDNA chứa các vị trí *chi*.

- Các vị trí *SacI* và *XbaI* được loại bỏ khỏi các nhánh của vector sao cho vị trí *XbaI* ở trong vùng tạo dòng là duy nhất, các thao tác này làm khuyết một đoạn khoảng 500 bp trong DNA của bacteriophage  $\lambda$  ở phía bên phải gen *lac5*. Đặc điểm này cho phép nhận các đoạn chèn DNA lớn hơn (tới 8,2 kb) so với các vector  $\lambda$ gt18 và  $\lambda$ gt19.

### 2.3. Vector $\lambda$ gt22 và $\lambda$ gt23

Các vector này cũng bắt nguồn từ  $\lambda$ gt18 và  $\lambda$ gt19, mang trình tự tổng hợp *chi* và các vị trí tạo dòng trong một khung khác gần đầu 3' của vùng mã

hóa gen *lacZ*. Các vector  $\lambda$ gt22 và  $\lambda$ gt23 giống hệt nhau ngoại trừ hướng của các vị trí tạo dòng. Các vị trí *Xba*I và *Sac*I ở các nhánh của  $\lambda$ gt18 và  $\lambda$ gt19 bị loại bỏ, như vậy các vector này chỉ mang năm vị trí cắt hạn chế duy nhất để tạo dòng là: *Not*I, *Xba*I, *Sac*I, *Sal*I và *Eco*RI. Các đoạn cDNA gắn vào một trong năm vị trí có thể được biểu hiện như là một protein dung hợp LacZ. Các vector  $\lambda$ gt22 và  $\lambda$ gt23 được thiết kế để cho phép tạo dòng định hướng cDNA vào trong các vị trí *Sal*I và *Not*I. Tổng hợp sợi thứ nhất và thứ hai của cDNA bằng phương pháp primer-linker cho phép cDNA sợi đôi được cắt bằng *Sal*I và *Not*I và gắn trực tiếp trong các nhánh của vector theo hướng liên quan với promoter *lacZ*. Một trong ba thể tái tổ hợp sẽ mang phân tử cDNA trong khung đọc chính xác để sản xuất protein dung hợp.

#### 2.4. Vector $\lambda$ ZAP

Vector  $\lambda$ ZAP mang vùng tạo dòng cùng hướng với promoter *lacZ* của *E. coli*. Đoạn cDNA dài 10 kb có thể được gắn vào trong các vị trí này và biểu hiện hoặc trong vi khuẩn được xâm nhiễm hoặc trong các thể tiềm tan được cảm ứng, như mô tả ở  $\lambda$ gt11. Hơn nữa, các protein dung hợp  $\lambda$ ZAP có thể được biểu hiện từ các plasmid có số lượng bản sao lớn. Vector  $\lambda$ ZAP có một số đặc điểm như sau:

- Có sáu vị trí RE trong vùng tạo dòng được sử dụng để tạo dòng định hướng các phân tử cDNA, trong đó có một vị trí cắt *Eco*RI tương tự như  $\lambda$ gt11.

- Có các promoter của bacteriophage T3 và T7 nằm bên cạnh vùng tạo dòng.

- Có các trình tự có nguồn gốc từ bacteriophage f1 để chuyển đổi *in vivo* đoạn chèn DNA từ bacteriophage  $\lambda$  vector tới Bluescript plasmid, các trình tự của nó được chứa trong  $\lambda$ ZAP. Quá trình cắt bỏ được thực hiện dễ dàng nhờ sự bố trí các trình tự của bacteriophage f1 (khởi đầu tổng hợp DNA sợi đơn) về một phía của Bluescript plasmid DNA mang vùng tạo dòng và sự bố trí các trình tự bacteriophage f1 (kết thúc tổng hợp DNA sợi đơn) về một phía khác của plasmid DNA.

Vector  $\lambda$ ZAP là dạng đầu tiên của vector thể hệ mới được thiết kế nhằm cung cấp một trình tự các vị trí tiềm năng để tạo dòng và biểu hiện cDNA, một phương pháp tiện lợi và đơn giản để thu hồi và thực hiện các thao tác tiếp theo của cDNA. Do tính linh hoạt của chúng, các vector này

thay thế cho  $\lambda$ gt10 và  $\lambda$ gt11 và được xem là vector thích hợp để xây dựng các thư viện cDNA. Vector  $\lambda$ ZAPII cũng tương đồng với  $\lambda$ ZAP.

## VI. Nhận dạng các dòng cDNA quan tâm

### 1. Các phương pháp sàng lọc

Có ba phương pháp sàng lọc thư viện cDNA cho các dòng quan tâm:

- Lai nucleic acid.
- Phát hiện các kháng nguyên miễn dịch đặc hiệu.
- Chọn lọc sib các dòng cDNA.

Dưới đây chỉ trình bày hai phương pháp phổ biến đó là lai nucleic acid và phát hiện các kháng nguyên miễn dịch đặc hiệu.

#### 1.1. Lai nucleic acid

Đây là phương pháp được sử dụng phổ biến và đáng tin cậy nhất khi sàng lọc các thư viện cDNA để tìm kiếm các dòng quan tâm. Sàng lọc bằng phương pháp lai nucleic acid cho phép phân tích đồng thời nhiều dòng và nhanh, không đòi hỏi các dòng cDNA phải hoàn chỉnh và sản phẩm được tổng hợp trong tế bào vật chủ phải có hoạt tính sinh học hoặc kháng nguyên. Hơn nữa, cơ sở lý thuyết của kỹ thuật lai nucleic acid đã được hiểu biết đầy đủ. Điều này cho phép phát triển một số lượng lớn các kỹ thuật khác nhau có thể điều chỉnh các mẫu dò của nucleic acid có các chiều dài và đặc điểm khác nhau.

- **Các mẫu dò tương đồng.** Các mẫu dò tương đồng chứa ít nhất một phần của chuỗi nucleic acid chính xác của dòng cDNA mong muốn. Chúng được dùng trong nhiều trường hợp khác nhau, ví dụ: khi một dòng bộ phận của cDNA hiện có được sử dụng để phân lập một dòng hoàn chỉnh từ thư viện cDNA. Thông thường, đoạn bắt nguồn từ một đầu hoặc đầu khác của dòng hiện có được phân lập, đánh dấu phóng xạ *in vitro* và dùng để thăm dò thư viện. Lai với các mẫu dò tương đồng thường được tiến hành dưới các điều kiện nghiêm ngặt (stringency).

- **Các mẫu dò tương đồng từng phần.** Các mẫu dò tương đồng từng phần được dùng để phát hiện các dòng cDNA có quan hệ họ hàng, nhưng không giống hệt nhau hoàn toàn, với các trình tự của mẫu dò. Nếu các mẫu

dò kháng thể lẫn nucleic acid không có sẵn, thì có thể thay đổi bằng một vài phương pháp. Ví dụ: Nếu dùng gen đã được tạo dòng từ các loài khác hoặc nếu gen liên quan đã được tạo dòng từ cùng loài, thì cần tiến hành một loạt các thí nghiệm kiểm tra xem có sự bảo toàn đầy đủ của trình tự nucleic acid hay không để cho phép sàng lọc thư viện cDNA bằng cách lai. Điều này được thực hiện dễ dàng nhất bằng cách tổ chức một loạt các phản ứng lai Southern và Northern ở các mức độ nghiêm ngặt khác nhau. Dùng một lượng tối thiểu (5-10  $\mu\text{g}$ ) của genomic DNA đã được cắt bằng RE để chạy điện di trên agarose gel 0,8%, các phân đoạn sau đó được chuyển vào màng lai nitrocellulose. Màng được cắt thành từng mảnh nhỏ, mỗi mảnh được lai dưới các điều kiện khác nhau với các lượng mẫu dò có hoạt tính phóng xạ giống nhau. Một loạt các phản ứng lai tương tự có thể tiến hành với các mRNA được tiêu phân hóa bằng điện di và chuyển lên giá thể rắn. Mục đích của cả hai trường hợp là thiết lập các điều kiện cho phép các gen được tạo dòng trước đó được sử dụng như là mẫu dò cho cDNA quan tâm.

**- Các oligonucleotide mẫu dò nhân tạo.** Các oligonucleotide mẫu dò nhân tạo là các đoạn nucleotide của trình tự xác định được tổng hợp *in vitro*. Trình tự của các mẫu dò này được suy luận, bằng cách dùng mã di truyền, từ các vùng ngắn của chuỗi amino acid đã biết của protein quan tâm. Do sự thoái biến (degeneracy) của mã di truyền, chuỗi amino acid đề xuất có thể không được đặc trưng chính xác bởi các oligonucleotide đơn được dự đoán cho trình tự. Mặc dù, trong rất nhiều trường hợp, trình tự các amino acid giống nhau có thể được đặc trưng bởi nhiều oligonucleotide khác nhau. Không có cách nào để biết chắc chắn các oligonucleotide này trên thực tế có được dùng trong gen quan tâm hay không. Ba giải pháp được đưa ra cho vấn đề này là như sau:

+ Một họ oligonucleotide có thể được tổng hợp chứa tất cả các trình tự có khả năng mã hóa cho một trình tự đã cho của chuỗi amino acid. Số lượng các thành viên trong họ này tùy thuộc vào mức độ thoái biến của mã bộ ba cho các amino acid đặc biệt. Tuy nhiên, do tất cả các trình tự oligonucleotide có khả năng được hiện diện, nên ít nhất một trong số các thành viên sẽ tương hợp với dòng cDNA quan tâm. Duy trì kích thước của mỗi họ trong tỷ lệ dễ sử dụng, các oligonucleotide ngắn (14-17 base) thường được sử dụng, một kích thước tối thiểu được tiến hành cho phản ứng lai.



+ Oligonucleotide dài hơn (40-60 base) của trình tự duy nhất có thể được tổng hợp bằng cách dùng các mã bộ ba được sử dụng phổ biến nhất cho mỗi amino acid (tránh dùng dinucleotide CpG, vì nó được hiện diện lại trong hầu hết các cDNA của eukaryote). Tất nhiên, oligonucleotide này sẽ không tương hợp một cách chính xác với trình tự trong cDNA, nhưng nó sẽ cố định đủ tốt để phát hiện bằng cách lai dưới các điều kiện không nghiêm ngặt (non-stringency).

+ Một oligonucleotide có thể được tổng hợp chứa một base như là inosine ở các vị trí thoái biến tiềm tàng. Inosine có thể bắt cặp với tất cả bốn loại base truyền thống mà không làm tổn thương nghiêm trọng khả năng ổn định của các thể lai có kết quả. Vì thế, nó có khả năng sản sinh ra các họ oligonucleotide dài hơn được làm giảm số lượng và còn lai với gần như tất cả các dòng cDNA có thể mã hóa cho protein quan tâm.

Cuối cùng, nếu trình tự protein có sẵn từ đầu tận cùng amino của protein, thì thư viện cDNA được sàng lọc phải có chất lượng cao để chắc chắn rằng hầu hết đầu tận cùng 5' của mRNA được đại diện.

## 1.2. Phát hiện miễn dịch các kháng nguyên đặc hiệu

Các thư viện cDNA được xây dựng trong các vector biểu hiện như  $\lambda$ gt11,  $\lambda$ gt18-23,  $\lambda$ ZAP... có thể được sàng lọc bằng kháng thể theo hướng chống lại protein quan tâm. Các màng lọc nitrocellulose in các vết tan của vi khuẩn được ngâm trong dung dịch chứa kháng thể. Sau khi rửa, màng lọc được ủ với protein A của *Staphylococcus aureus* hoặc bằng kháng thể thứ hai theo hướng chống lại các yếu tố quyết định kháng nguyên đặc hiệu loài của kháng thể thứ nhất. Ở các mô tả đầu tiên của phương pháp này, phối tử (ligand) thứ hai được đánh dấu đồng vị phóng xạ với  $^{125}\text{I}$ . Ngày nay, phối tử thứ hai được liên kết hóa trị với enzyme mà hoạt tính của enzyme có thể được phát hiện bằng hóa tổ chức học (ví dụ: alkaline phosphatase).

Chìa khóa thành công của phương pháp này nằm trong chất lượng của kháng thể. Điều cần thiết là kháng thể nhận diện được protein bị biến tính (Ví dụ: nó phải cho các tín hiệu mạnh trong phân tích Western blot-xem chương 7). Hơn nữa, quá trình sàng lọc được tiến hành dễ dàng hơn và miễn cảm hơn nếu kháng thể được bắt nguồn từ huyết thanh đa dòng (polyclonal antiserum) độ chuẩn cao. Bởi vì các huyết thanh như thế phản ứng bình thường với nhiều yếu tố quyết định kháng nguyên khác nhau, cơ hội để phát

hiện dòng cDNA biểu hiện đoạn protein quan tâm được tăng lên. Các huyết thanh đa dòng thường chứa các kháng thể phản ứng chéo nhận diện các thành phần không tái tổ hợp của vi khuẩn tiềm tan và chúng phải được loại bỏ trước khi quá trình sàng lọc được thực hiện.

Phản ứng liên kết không đặc hiệu sẽ thấp hơn nhiều khi sử dụng kháng thể đơn dòng làm mẫu dò. Tuy nhiên, số lượng các thể tái tổ hợp có thể được phát hiện cũng bị giảm bởi vì mỗi kháng thể đơn dòng riêng biệt có thể phản ứng với chỉ một yếu tố quyết định kháng nguyên đơn. Vì thế, mẫu dò miễn dịch lý tưởng có thể chứa một hỗn hợp của nhiều kháng thể đơn dòng khác nhau, mà mỗi kháng thể phản ứng mạnh với protein bị biến tính.

## 2. Các phương pháp xác nhận các dòng cDNA

Các thư viện cDNA thường được dàn trải ở mật độ cao để sàng lọc với kháng thể hoặc các mẫu dò của nucleic acid, và mỗi dòng phản ứng dương tính trong vòng đầu tiên đòi hỏi một số chu kỳ dàn trải và sàng lọc bổ sung trước khi chúng được xem như thuần khiết. Tuy nhiên, khả năng phản ứng thích hợp với mẫu dò đặc biệt là không đầy đủ để chứng minh dòng cDNA thu được bắt nguồn từ mRNA quan tâm. Sự chứng minh chỉ tuyệt đối khi cho thấy dòng cDNA chứa một khung đọc mở mã hóa cho trình tự amino acid hoàn toàn của protein. Một số vấn đề sau cần được lưu ý để đảm bảo mức độ chính xác của các dòng cDNA thu được:

- Biểu hiện của protein từ cDNA hoàn chỉnh trong các tế bào prokaryote hoặc eukaryote thể hiện các hoạt tính sinh học hoặc enzyme chính xác.

- Sự tương ứng giữa các phần của chuỗi nucleotide của cDNA và các chuỗi amino acid của các peptide có nguồn gốc từ protein được tinh sạch.

- Sự tương ứng giữa các bản đồ peptide của chuỗi polypeptide được tổng hợp *in vitro* bằng sự phiên mã của dòng cDNA và các bản đồ peptide của protein đích thực.

- Sự kết tủa miễn dịch của polypeptide được tổng hợp *in vitro* hoặc *in vivo* từ sự phiên mã dòng cDNA bằng các kháng thể được tăng khả năng chống lại protein quan tâm. Sự chặt chẽ của thử nghiệm này tăng lên khi nó được tiến hành bởi một chuỗi các kháng thể đơn dòng xác nhận các yếu tố quyết định kháng nguyên khác nhau trên phân tử protein.

- Kết tủa miễn dịch protein đích thực với các kháng thể tăng khả năng chống lại các peptide tổng hợp mà các trình tự của chúng được xác định bởi trình tự nucleic acid của cDNA được tạo dòng.

## VII. Xây dựng thư viện cDNA trong bacteriophage $\lambda$ vector

Sử dụng thư viện cDNA có hai ưu điểm sau:

- Các dòng cDNA chứa trình tự mã hóa liên tục của một gen. Nhiều gen ở eukaryote là gián đoạn, có chứa nhiều đoạn intron. Sau khi cắt tiền thân mRNA (pre-mRNA) và nối lại, các đoạn intron đã bị loại và mRNA hoàn thiện (mature mRNA) có trình tự mã hóa liên tục được tạo thành. Do cDNA được phiên mã ngược từ khuôn mẫu mRNA hoàn thiện nên các dòng cDNA có thể tổng hợp protein cần thiết với số lượng lớn như mong muốn.

- Nhiều protein được tổng hợp với số lượng lớn bởi những tế bào chuyên hóa và như vậy trong các tế bào này mRNA của protein đó sẽ có tỷ lệ cao và thư viện cDNA được tạo ra từ các tế bào này sẽ có nhiều cDNA mã hóa cho các protein tương ứng. Sự dồi dào cDNA của một vài loại mRNA nào đó làm giảm nhẹ đáng kể việc xác định đúng dòng mong muốn từ thư viện gen.

Phương pháp tổng hợp gen từ mRNA ngày càng được phát triển, nó kết hợp với các phương pháp khác của sinh học phân tử cho phép ứng dụng trong nhiều lĩnh vực.

Các bước chính của quá trình xây dựng thư viện cDNA như sau:

### 1. Chọn lọc kích thước của cDNA

cDNA sau khi cắt hạn chế được phân đoạn bằng sắc ký cột Sepharose để loại bỏ những linker thừa và các phân tử cDNA có kích thước nhỏ hơn 500 bp. Cột sắc ký thường có kích thước 27×0,3 cm thích hợp cho việc phân đoạn các cDNA.

### 2. Gắn các cDNA với các nhánh của bacteriophage $\lambda$

Thông thường cần phải tiến hành thử một loạt các phản ứng gắn và phản ứng đóng gói. Trong các phản ứng này, một lượng không đổi của các nhánh

bacteriophage  $\lambda$  được gắn với các lượng khác nhau của cDNA, mục đích là để xác định lượng cDNA tạo ra ít nhất là  $5 \times 10^6$  bacteriophage tái tổ hợp.

Phản ứng gắn cần được thiết kế sao cho tối thiểu một bacteriophage tái tổ hợp đơn sẽ chứa nhiều hơn một phân tử cDNA bằng cách dùng tỷ lệ của các nhánh được phosphoryl hóa và cDNA để chỉ 5% bacteriophage được tái tổ hợp. Nếu sử dụng các nhánh dephosphoryl hóa, thì các bacteriophage không tái tổ hợp bị ức chế hiệu quả, và vì thế không thể xác định lượng cDNA cần thiết để tạo ra một quần thể bacteriophage chứa 5% thể tái tổ hợp.

Việc sử dụng các nhánh dephosphoryl hóa hoặc các nhánh có đầu tận cùng không tương thích, tạo ra bằng cách lấp đầy từng phần, được khuyến cáo sử dụng khi xây dựng thư viện cDNA trong các bacteriophage như  $\lambda$ gt11,  $\lambda$ gt18-23 và  $\lambda$ ZAP, là những vector không có hệ thống cho phép chọn lọc dựa theo các bacteriophage bố mẹ. Số lượng các thể không tái tổ hợp có thể được giảm xuống khoảng 100 lần bằng cách dùng các nhánh dephosphoryl hóa. Trong hệ thống này, các bacteriophage không tái tổ hợp tạo thành các plaque màu xanh trên các chủng *E. coli* thích hợp khi có mặt của IPTG và X-gal, trong khi đó các plaque tái tổ hợp là không màu.

### 3. Phân tích các đoạn chèn cDNA

Thu thập khoảng mười hai plaque của bacteriophage tái tổ hợp và chuẩn bị DNA để cắt bằng RE thích hợp. Phân tích kích thước của các đoạn chèn cDNA bằng điện di trên agarose gel 1%, dùng DNA marker có các đoạn từ 500 bp tới 5 kb. Nếu các bacteriophage tái tổ hợp chứa các đoạn chèn có kích thước khác nhau và nếu kích thước trung bình của chúng xấp xỉ 1 kb hoặc lớn hơn, thì có thể tiến hành các bước tiếp theo (ví dụ: tạo ra thư viện cDNA hoàn chỉnh). Nếu kích thước trung bình của các đoạn chèn nhỏ hơn 1 kb một cách rõ rệt, thì chất lượng của thư viện là không cao. Trong trường hợp này cần quay lại phân tích chất lượng của mRNA khởi đầu và các phản ứng tổng hợp cDNA.

### 4. Tạo thư viện cDNA hoàn chỉnh

Nếu sử dụng các nhánh dephosphoryl hóa để xây dựng thư viện cDNA, thì tính toán lượng cDNA cho kết quả tăng gấp 10 lần số lượng của

các plaque không tái tổ hợp. Nếu sử dụng các nhánh phosphoryl hóa, thì tính toán lượng cDNA cho kết quả quần thể bacteriophage chứa xấp xỉ 5% các thể tái tổ hợp. Thiết kế một phản ứng gắn quy mô lớn, tăng số lượng của tất cả các thành phần ở một tỷ lệ thích hợp. Chắc chắn rằng thiết kế phản ứng gắn đối chứng chỉ chứa các nhánh của bacteriophage  $\lambda$  mà không chứa cDNA. Đóng gói tất cả các hỗn hợp gắn bằng cách thiết kế một chuỗi các phản ứng gắn, mỗi phản ứng chứa 0,5  $\mu\text{g}$  các nhánh của bacteriophage  $\lambda$ . Gộp các tiêu thể đóng gói và xác định độ chuẩn của bacteriophage gốc trên các chủng vi khuẩn thích hợp.

### 5. Khuếch đại thư viện cDNA

- **Vector  $\lambda\text{gt}10$ .** Để thiết lập một sự cung cấp lâu dài của thư viện, thì nó cần phải được khuếch đại bằng sự sinh trưởng trên chủng *E. coli* thích hợp (chẳng hạn chủng BNN102) trên đĩa agar. Nếu thư viện chỉ được sàng lọc một lần, bước khuếch đại này có thể bỏ qua và các bacteriophage có thể được dàn mỏng trực tiếp trên chủng BNN102 ở một mật độ tối ưu để tiến hành sàng lọc.

- **Vector  $\lambda\text{gt}11$  và các dẫn xuất của nó và  $\lambda\text{ZAP}$ ,  $\lambda\text{ZAPII}$ .** Các thư viện cDNA được xây dựng trong các vector biểu hiện  $\lambda\text{gt}11$ ,  $\lambda\text{gt}18$ ,  $\lambda\text{gt}20$  và  $\lambda\text{gt}22$  phải được khuếch đại trên chủng *E. coli* Y1090*hdsR* (hoặc nếu là vector  $\lambda\text{ZAP}$  thì trên chủng BB4, vector  $\lambda\text{ZAPII}$  thì trên chủng XL1-Blue). Các chủng này là không hoàn hảo cho sự cắt hạn chế kiểm soát vật chủ nhưng lại mang một hệ thống methyl hóa hoạt động. Các vị trí tiềm tàng cho phản ứng cắt hạn chế bằng hệ thống *EcoK* sẽ được methyl hóa trong suốt quá trình khuếch đại sao cho, nếu cần thiết, các thể tái tổ hợp có thể được dùng để gây nhiễm chủng *E. coli* khả biến-cắt hạn chế. Trong suốt quá trình khuếch đại, điều quan trọng là ức chế sản xuất các protein độc tố tiềm tàng bởi các bacteriophage tái tổ hợp.

### Tài liệu tham khảo/đọc thêm

1. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K. 2002. Short Protocol in Molecular Biology. Vol 1 and 2. 5<sup>th</sup> ed. John Wiley & Sons, Inc. USA.

2. **Brown TA.** 2001. Gene Cloning-An Introduction. 4<sup>th</sup> ed. *Blackwell Science*, Oxford, UK.
3. **Calladine CR and Drew HR.** 1997. Understanding DNA: The Molecule and How It Works. 2<sup>nd</sup> ed. *Academic Press*, London, UK.
4. **Glick BR and Pasternak JJ.** 2003. Molecular Biotechnology: Principles and Applications of Recombinant DNA. 3<sup>rd</sup> ed. *ASM Press*, USA.
5. **Maniatis T, Fritsch EF and Sambrook J.** 1989. Molecular Cloning-A Laboratory Manual. *Cold Spring Harbor Laboratory Press*, USA.
6. **Ohman DE.** 1989. Experiments in Gene Manipulation. *Prentice Hall*, Englewood Cliffs, New Jersey, USA.
7. **Primrose SB, Twyman R and Old RW.** 2001. Principles of Gene Manipulation. 6<sup>th</sup> ed. *Blackwell Science*, Oxford, UK.
8. **Rapley R and Walker JM.** 1998. Molecular Biomethods Handbook. *Humana Press Inc.* New Jersey, USA.
9. **Surzycki S.** 2000. Basic Techniques in Molecular Biology. *Springer-Verlag*, Berlin, Heidelberg, Germany.

## Chương 7

# Biểu hiện gen tái tổ hợp trong *Escherichia coli*

Vector biểu hiện là vector có thể mang các gen ngoại lai mong muốn cho phép thực hiện sự phiên mã của các bản sao được tạo dòng và sự dịch mã các mRNA của chúng trong *Escherichia coli*. Những vector như thế được dùng để biểu hiện các gen của eukaryote trong *E. coli* hoặc tăng hiệu suất các sản phẩm của gen ở prokaryote. Mặc dù *E. coli* không phải là một vật chủ lý tưởng để biểu hiện các gen của eukaryote do chúng không có khả năng sửa đổi hậu dịch mã (post-translation modification) cho các chuỗi polypeptide của các protein có cấu trúc phức tạp. Tuy nhiên, trong một số trường hợp người ta vẫn có thể sử dụng vi khuẩn *E. coli* cho những protein có cấu trúc đơn giản hơn. Đối với những protein có cấu trúc phức tạp, vật chủ thường được sử dụng là các cơ thể eukaryote (ví dụ: nấm men *Saccharomyces cerevisiae*, nấm mốc *Aspergillus niger*...) do chúng có thể hậu dịch mã được.

Để biểu hiện tất cả các gen ngoại lai trong *E. coli* phải bắt đầu bằng việc gắn đoạn gen ngoại lai vào trong vector biểu hiện (thường là plasmid). Vector này phải có đủ các cấu trúc cần thiết sau:

- Trình tự khởi đầu sao chép (*ori*) để có thể tạo ra nhiều bản sao trong tế bào vật chủ.

- Các trình tự mã hóa gen chỉ thị chọn lọc (selectable marker) để đảm bảo duy trì vector trong tế bào.

- Một promoter kiểm soát phiên mã (ví dụ: *lac*, *trp* hoặc *tac*) cho phép sản xuất một lượng lớn mRNA từ các gen được tạo dòng.

- Các trình tự kiểm soát dịch mã như trình tự liên kết ribosome được bố trí thích hợp và codon khởi đầu AUG.

- Một polylinker để đưa gen ngoại lai vào trong một hướng chính xác với promoter.

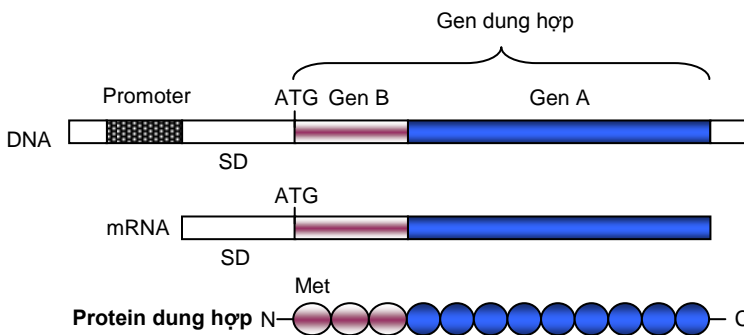
Chỉ khi được cấu trúc đầy đủ như thế, các vector biểu hiện mang gen ngoại lai mới được biến nạp vào chủng *E. coli* thích hợp.

Chương này mô tả các phương pháp và các plasmid vector được sử dụng để sản xuất các protein dung hợp (fusion protein) và các protein nguyên thể (native protein) trong vi khuẩn. Các protein dung hợp có thể được sản xuất với một lượng lớn, dễ dàng tinh sạch và có thể được dùng để gây ra một đáp ứng miễn dịch. Các protein nguyên thể có thể được sản xuất trong vi khuẩn bằng cách gắn một promoter mạnh và một trình tự liên kết ribosome hiệu quả ngược hướng với gen được tạo dòng (vùng 5' không dịch mã). Thông thường, các protein được biểu hiện ở mức độ cao trong các thể vùi không hòa tan.

## I. Sản xuất các protein dung hợp

### 1. Protein dung hợp

Protein dung hợp (protein lai) được mã hóa bởi một gen lai do sự dung hợp *in vitro* hai đoạn gen khác nhau. Vì vậy, protein dung hợp sẽ mang trình tự amino acid của hai protein khác biệt (Hình 7.1).



**Hình 7.1. Protein dung hợp.** Bao gồm gen được tạo dòng (gen A) và một gen khác của vi khuẩn (gen B), do đó protein dung hợp có chứa một phần protein của vi khuẩn. SD: đoạn Shine-Dalgarno.

Sự dung hợp gen được thực hiện bằng cách gắn phần mã hóa của gen được tạo dòng ở gần đầu cuối 3' của gen vi khuẩn (ví dụ: gen *lacZ*). Protein dung hợp có những ưu điểm sau:



- Protein dung hợp thường được sản xuất với hàm lượng lớn do sự khởi đầu phiên mã và dịch mã được điều khiển bởi các trình tự tiêu chuẩn của *E. coli*.

- Dung hợp các trình tự ngoại lai với các gen của *E. coli* thường cho kết quả các sản phẩm ổn định hơn các protein ngoại lai nguyên thể.

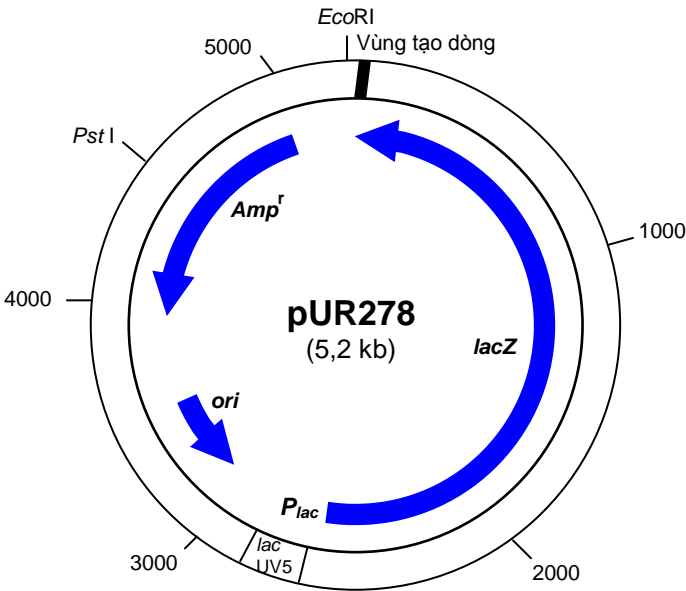
- Protein dung hợp có khối lượng phân tử lớn hơn so với hầu hết các protein của *E. coli* và vì thế dễ dàng nhận biết trong điện di polyacrylamide gel của protein. Bằng protein dung hợp có thể được cắt ra khỏi gel, đông khô (lyophilize) sau đó nghiền thành bột và dùng như một kháng nguyên.

- Protein dung hợp có thể được tiết ra môi trường nhờ biểu hiện của vi khuẩn nếu gen eukaryote được gắn với trình tự mã hóa cho peptide tín hiệu (signal peptide), khi đó peptide sẽ này tạo ra sự chuyển dịch protein qua màng. Tuy nhiên, hiện tượng này chỉ xảy ra ở một vài trường hợp.

## 2. Các hệ thống vector biểu hiện các gen dung hợp với gen *lacZ*

Một số hệ thống vector được phát triển để biểu hiện các gen dung hợp với gen *lacZ*. Chẳng hạn, các vector họ pUR có các vị trí tạo dòng *Bam*HI, *Sal*I, *Pst*I, *Xba*I, *Hind*III và *Cla*I ở đầu 3' của gen *lacZ* (Hình 7.2). Bằng cách chọn lựa các vector và vị trí cắt hạn chế thích hợp người ta có thể tiến hành dung hợp cho hầu hết gen được tạo dòng. Trong một số trường hợp, sự dung hợp có thể thực hiện bằng cách loại bỏ đầu tận cùng lõi hoặc làm đầy đầu tận cùng bị khuyết trước khi gắn.

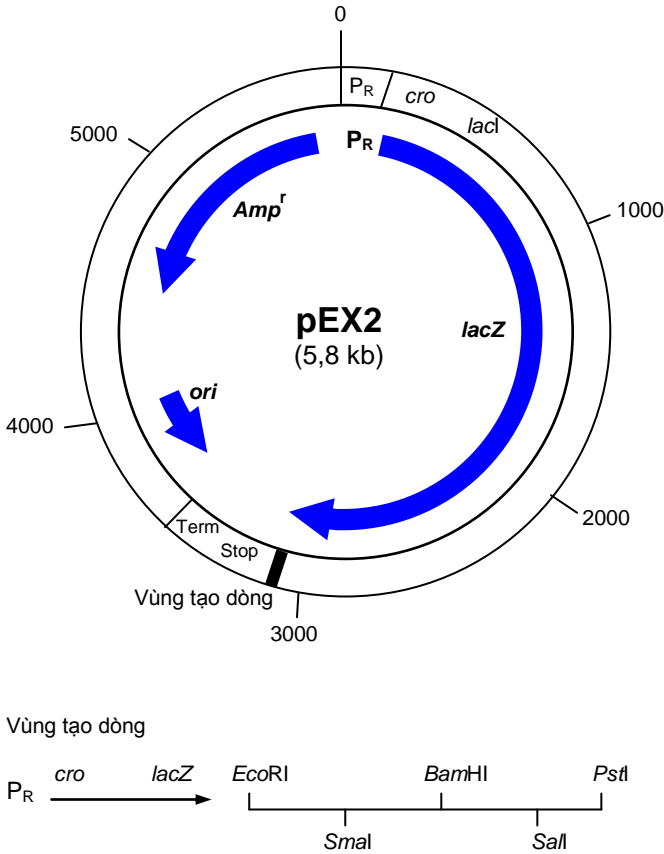
Một hệ thống vector tương tự khác cũng được phát triển để sản xuất các protein dung hợp, đó là các vector biểu hiện pEX1-3. Promoter  $P_R$  được điều hòa và cảm ứng trong cùng một kiểu như promoter  $P_L$  của bacteriophage  $\lambda$ . Các vector pEX1-3 có các vị trí tạo dòng mang các điểm nhận biết *Eco*RI, *Sma*I, *Bam*HI, *Sal*I và *Pst*I nằm ở đầu 3' của gen *lacZ*. Các codon kết thúc dịch mã và các tín hiệu kết thúc phiên mã được đặt cùng hướng (downstream) với vị trí tạo dòng (Hình 7.3).



Vùng tạo dòng

	→	<i>lacZ</i>								
pUR278	TGT CAA AAA GGG GAT CCG TCG ACT CTA GAA AGC TTA TCG ATG									
		<i>Bam</i> HI	<i>Sal</i> I	<i>Xba</i> I	<i>Hind</i> III	<i>Clal</i>				
pUR288	TGT CGG GGA TCC GTC GAC TCT AGA AAG CTT ATC GAT GAT									
		<i>Bam</i> HI	<i>Sal</i> I	<i>Xba</i> I	<i>Hind</i> III	<i>Clal</i>				
pUR289	TGT CAG GGG ATC CGT CGA CTC TAG AAA GCT TAT CGA TGA									
		<i>Bam</i> HI	<i>Sal</i> I	<i>Xba</i> I	<i>Hind</i> III	<i>Clal</i>				
pUR290	TGT CAA AAA GGG GAT CCG TCG ACC TGC AGC CAA GCT TAT CGA TGA									
		<i>Bam</i> HI	<i>Sal</i> I	<i>Pst</i> I	<i>Hind</i> III	<i>Clal</i>				
pUR291	TGT CGG GGA TCC GTC GAC CTG CAG CCA AGC TTA TCG ATG									
		<i>Bam</i> HI	<i>Sal</i> I	<i>Pst</i> I	<i>Hind</i> III	<i>Clal</i>				
pUR292	TGT CAG GGG ATC CGT CGA CCT GCA GCC AAG CTT ATC GAT									
		<i>Bam</i> HI	<i>Sal</i> I	<i>Pst</i> I	<i>Hind</i> III	<i>Clal</i>				

**Hình 7.2. Các vector họ pUR.** Đây là các vector dung hợp với gen *lacZ*, có các vị trí tạo dòng *Bam*HI, *Sal*I, *Pst*I, *Xba*I, *Hind*III và *Clal* ở đầu 3' của gen *lacZ*. Chèn đoạn DNA ngoại lai (cDNA) vào trong các vị trí tạo dòng thích hợp cho phép sản xuất protein dung hợp của  $\beta$ -galactosidase hoạt động với chuỗi peptide được mã hóa bởi DNA ngoại lai. Các vector pUR278, pUR288 và pUR289 chỉ chứa một vị trí *Pst*I trong gen *Amp*<sup>r</sup>. Các vector pUR290, pUR291 và pUR292 chứa một vị trí *Pst*I trong vùng tạo dòng do vị trí *Pst*I trong gen *Amp*<sup>r</sup> đã bị loại bỏ. Sử dụng các plasmid vector này rất tiện lợi khi đoạn DNA ngoại lai quan tâm được tạo dòng trong vị trí *Pst*I bằng phương pháp đuôi GC.



**Hình 7.3. Vector pEX2.** Plasmid vector này dài khoảng 5,8 kb được thiết kế để biểu hiện đoạn DNA ngoại lai (cDNA) được dung hợp ở đầu 3' của gen *lacZ*. Phần tận cùng amino của gen *lacZ* được thay thế bằng một số trình tự của gen *cro* của bacteriophage  $\lambda$  và gen *lacI* của *E. coli*. Promoter *P<sub>R</sub>* của bacteriophage  $\lambda$  (dùng để biểu hiện protein dung hợp) được điều hòa bởi gen ức chế *cIts857* của bacteriophage  $\lambda$ . Vùng tạo dòng hiện diện ở đầu 3' của gen *lacZ*, tiếp theo là các codon kết thúc dịch mã (Stop) và tín hiệu kết thúc phiên mã (Term) của bacteriophage fd.

### 3. Xây dựng plasmid biểu hiện và phát hiện các protein dung hợp

Gắn plasmid vector (pUR hoặc pEX) thích hợp với đoạn DNA ngoại lai để tạo ra một sự dung hợp trong khung đọc. Biến nạp các vector này vào *E. coli* (chủng K12 71/18 hoặc JM 103 cho các vector pUR, chủng M5219 cho các vector pEX). Kiểm tra các khuẩn lạc đơn mang đoạn DNA ngoại lai mong muốn bằng cách tách chiết DNA (DNA miniprep) của plasmid vector,

sau đó cắt bằng enzyme hạn chế thích hợp và điện di kiểm tra trên agarose gel. Sàng lọc các khuẩn lạc sản xuất protein dung hợp.

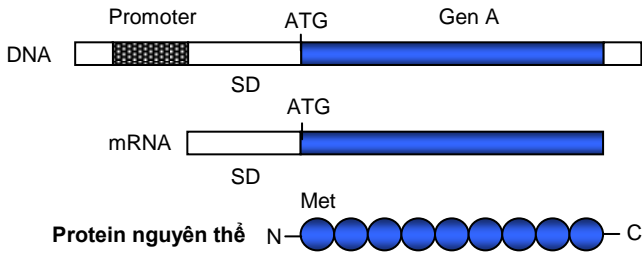
Phương pháp sàng lọc khuẩn lạc để phát hiện protein dung hợp được tiến hành như sau: Nuôi cấy qua đêm ở 37°C (một lượng nhỏ) từ 5-10 khuẩn lạc trong môi trường LB có chứa ampicillin. Sau đó, cảm ứng nuôi cấy bằng cách bổ sung thêm isopropylthio- $\beta$ -galactoside (IPTG) cho vector pUR và tiếp tục nuôi ở 37°C, đối với vector pEX thì chuyển nuôi cấy 37°C lên nhiệt độ 40°C và tiếp tục nuôi cấy (có sục khí). Sau các khoảng thời gian nuôi cấy khác nhau (1, 2, 3 và 4 giờ...), lấy 1 mL dịch nuôi cấy ly tâm nhanh để thu tiểu thể, tái huyền phù chúng trong đệm 1 $\times$  SDS, biến tính ở 100°C trong 3 phút, ly tâm 12.000 g trong 1 phút ở nhiệt độ phòng. Tiến hành điện di SDS trên polyacrylamide gel 6%. Dùng dịch huyền phù tế bào chỉ mang một mình vector làm đối chứng. Protein dung hợp sẽ xuất hiện như là một băng mới dịch chuyển chậm hơn băng  $\beta$ -galactosidase trong đối chứng.

#### 4. Tách chiết các protein dung hợp để sản xuất kháng thể

Protein dung hợp có thể được tách chiết theo một số cách: dùng dịch chiết urea, sắc ký ái lực aminophenylthiogalactoside, điện di SDS-polyacrylamide gel hoặc phối hợp giữa các cách này. Phương pháp đơn giản nhất là điện di SDS-PAGE (sodium dodecylsulfate-polyacrylamide gel electrophoresis), như trình bày ở mục 3 (nuôi cấy một lượng lớn trong 200 mL). Nhuộm gel và phát hiện băng protein mới, sau đó cắt băng này ra khỏi gel, đông khô khoảng 2 ngày rồi nghiền thành bột mịn. Bột này sẽ được dùng tiêm vào thỏ để sản xuất kháng thể.

## II. Sản xuất các protein nguyên thể

Các protein nguyên thể có thể được sản xuất trong *E. coli* bằng cách sử dụng promoter điều hòa mạnh và một trình tự liên kết ribosome hiệu quả. Để biểu hiện gen prokaryote có trình tự liên kết ribosome mạnh chỉ cần cung cấp một promoter là đủ. Trong khi đó, để biểu hiện một gen eukaryote (hoặc một gen prokaryote với một trình tự liên kết ribosome yếu) cần phải cung cấp cả promoter lẫn trình tự liên kết ribosome (Hình 7.4). Các mức độ biểu hiện có thể khác nhau từ dưới 1% cho tới hơn 30% protein hòa tan tổng số (total soluble protein) của tế bào.



**Hình 7.4. Protein nguyên thể.** Gen tạo dòng (gen A) được đặt sau promoter và trình tự SD của vi khuẩn. mRNA chỉ mã hóa các amino acid đặc hiệu của đoạn chèn để tạo ra loại protein nguyên thể.

### ▪ Biểu hiện của các gen ở prokaryote: Promoter

Bước đầu tiên khi biểu hiện các protein của eukaryote trong vi khuẩn là chọn một vector biểu hiện mang promoter mạnh (strong promoter) của prokaryote. Promoter là trình tự DNA hướng dẫn RNA polymerase liên kết với DNA và khởi đầu sự tổng hợp RNA. Các promoter khác nhau có hiệu suất khác nhau, nói chung, promoter mạnh giúp các mRNA được khởi đầu ở tần số cao. Các promoter khác nhau đều mang hai đoạn bảo toàn cao, một được xác định khoảng 10 bp và một khoảng 35 bp nằm ở vùng ngược hướng (upstream) với điểm khởi đầu của sự phiên mã (còn gọi là vùng 5' không dịch mã-5' untranslated region). Hai đoạn này được đánh giá rất quan trọng trong việc xác định cường độ của promoter.

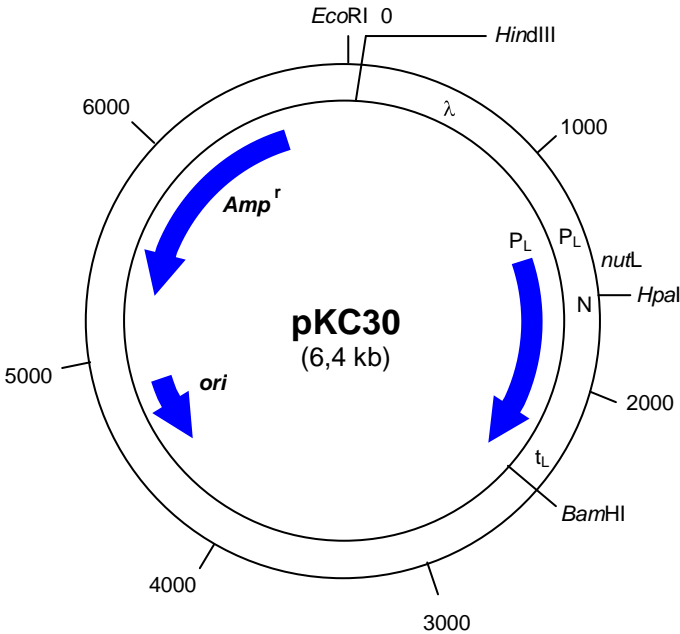
Các promoter thích hợp nhất cho biểu hiện của gen ngoại lai ở *E. coli* là những promoter mà cả hai cường độ và khả năng điều hòa thể hiện mạnh. Nếu sản phẩm của gen được tạo dòng là độc (toxin) đối với các tế bào *E. coli* thì sau đó sự nhân đôi gen sẽ làm cho cường độ promoter yếu và promoter không điều hòa được. Sự phiên mã ở mức độ cao có thể gây trở ngại cho việc sao chép DNA của plasmid và dẫn đến tính không ổn định của plasmid.

Phần này giới thiệu các vector biểu hiện mang promoter  $P_L$  của bacteriophage  $\lambda$ , promoter (lai) *trp-lac* và promoter bacteriophage T7.

#### 1. Promoter $P_L$ của bacteriophage $\lambda$

Promoter  $P_L$  của phage  $\lambda$  (Hình 7.5) là một promoter mạnh, có khả năng điều hòa tốt và được dùng trong một số vector biểu hiện. Ở nhiệt độ thấp ( $31^\circ\text{C}$ ), promoter  $P_L$  được duy trì ở trạng thái bị ức chế bởi sản phẩm

của gen *cI*. Sau khi hoạt tính của gen ức chế bị phá hủy bằng cách tăng nhiệt độ nuôi cấy thì promoter  $P_L$  sẽ xúc tiến phiên mã phần lớn mRNA. Một vài vector sử dụng promoter  $P_L$  của  $\lambda$  như: pPLa 2311, pPLa 8 và pKC30.



**Hình 7.5. Vector pKC30.** pKC30 là plasmid có kích thước xấp xỉ 6,4 kb, mang promoter  $P_L$  của bacteriophage  $\lambda$  và vị trí nhận biết *HpaI* nằm ở vùng cùng hướng (có 321 nucleotides) với vị trí khởi đầu phiên mã của  $P_L$ . Plasmid này là dạng phân chia của vector pBR322 và chứa đoạn *HindIII-BamHI* có nguồn gốc từ bacteriophage  $\lambda$  được chèn vào giữa các vị trí *HindIII* và *BamHI* của vector. Đoạn chèn cDNA mang tín hiệu promoter ( $P_L$ ), một vị trí được nhận biết bởi sản phẩm của gen *N* (*nutL*), bản thân gen *N*, và tín hiệu kết thúc phiên mã phụ thuộc *rho* ( $t_L$ ). Vị trí nhận biết *HpaI* nằm với vùng mã hóa của gen *N*. Các trình tự DNA được chèn vào trong vị trí *HpaI* có thể được điều chỉnh bằng cách chuyển plasmid tái tổ hợp vào thể tiềm tan của bacteriophage miễn cảm với nhiệt độ (*cIts857*). Các tế bào được sinh trưởng đến giữa pha log ở 30°C và sau đó thay đổi tới 40°C để bất hoạt sản phẩm của gen *cI* và mở promoter  $P_L$ . Vector này được dùng để biểu hiện protein *cII* của bacteriophage  $\lambda$  với một mức độ khoảng 4% protein hòa tan tổng số của tế bào.

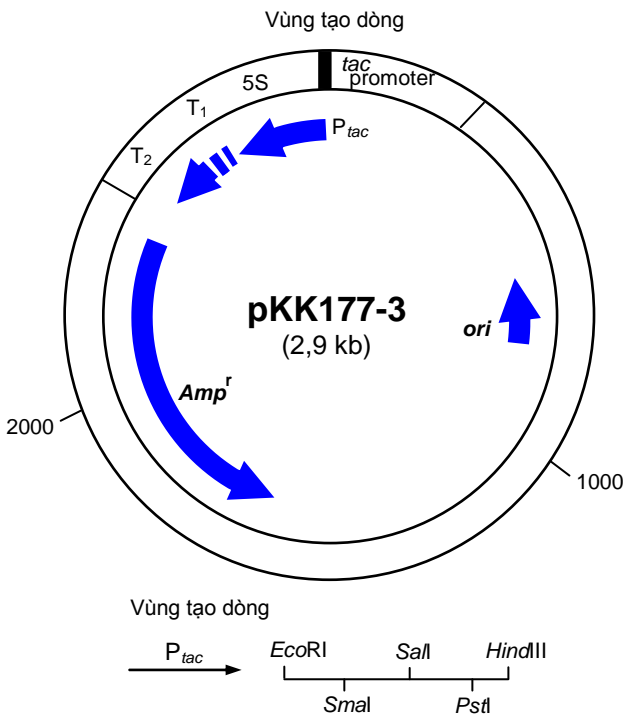
## 2. Promoter *trp-lac*

Sự biểu hiện của các gen được tạo dòng trong vi khuẩn *E. coli* còn được điều hòa bằng các promoter khác. Chẳng hạn promoter *tac* (một dạng promoter lai giữa promoter *trp* và promoter *lac*) đã được sử dụng thành công để sản xuất một lượng lớn protein trong *E. coli* (Hình 7.6).

- Promoter *trp* được điều hòa bởi gen ức chế *trp* và có thể được cảm ứng bởi sự bổ sung 3-IAA (3-indolylacetic acid) vào môi trường, hoặc bằng cách thiếu tryptophan.

- Promoter *lac* được điều hòa bởi gen ức chế *lac* và vì thế có thể được cảm ứng bởi sự bổ sung nhân tố cảm ứng IPTG cho nuôi cấy vi khuẩn.

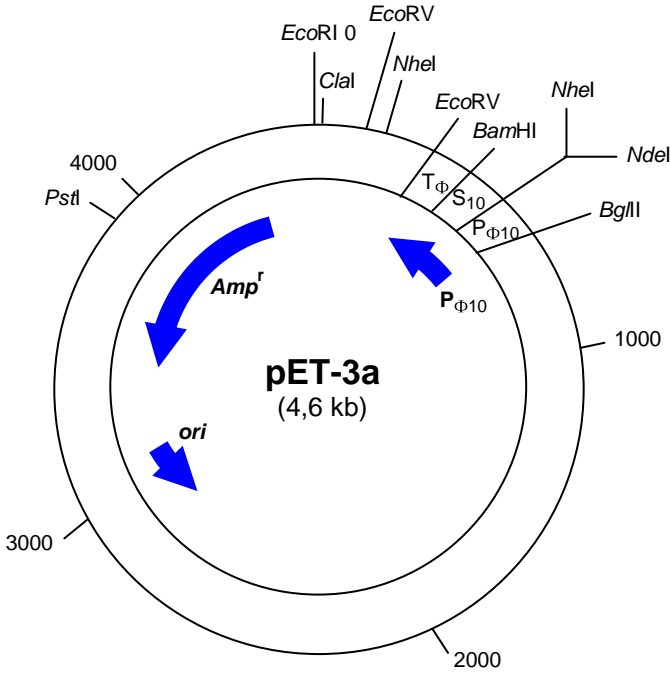
- Cuối cùng promoter lai *trp-lac* chứa đoạn *trp-35* gắn với đoạn *lac-10* và operator *lac* được Boer và cs thiết kế năm 1982, promoter lai *trp-lac* được điều hòa bởi gen ức chế *lac* (*lac* repressor).



**Hình 7.6. Vector pKK177-3.** pKK177-3 là một *tac* vector chứa các vị trí tạo dòng gen ngoại lai cùng hướng với promoter *tac*. Cùng hướng với các vị trí tạo dòng là *rrnB* mang gen 5S của *E. coli* và hai nhân tố kết thúc phiên mã T<sub>1</sub> và T<sub>2</sub>.

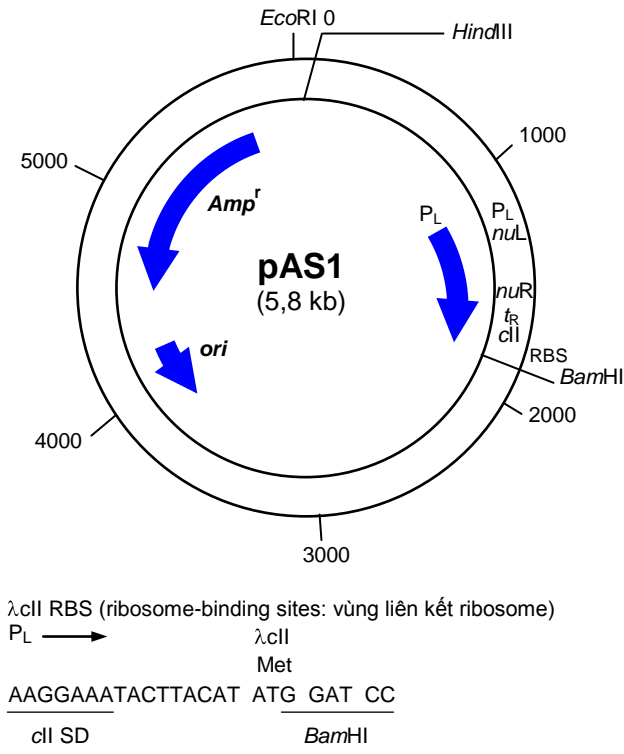
### 3. Promoter bacteriophage T7

Một hệ thống biểu hiện khác đã được phát triển bởi Tabor và Richardson (1985), Studier và Moffatt (1986) đó là hệ thống bacteriophage T7 RNA polymerase/promoter (Hình 7.7). Hệ thống này được thiết kế cho các biểu hiện có chọn lọc của gen được tạo dòng. Chúng cho phép mức độ biểu hiện cao của một vài gen không được biểu hiện hiệu quả trong các hệ thống khác.



**Hình 7.7. Vector pET-3 mang promoter (P<sub>Φ10</sub>) của bacteriophage T7 Φ10 và yếu tố kết thúc phiên mã Φ (T<sub>Φ</sub>).** Yếu tố kết thúc phiên mã có thể tạo ra các thể phiên mã có sức đề kháng mạnh hơn đối với hoạt tính exonuclease. Vector pET-3a là dạng phân chia của vector pET-3 trong đó vùng khởi đầu dịch mã (S<sub>10</sub>) của bacteriophage T7 Φ10 (protein chính của vỏ bacteriophage T7) có mang vị trí BamHI ở codon 11 được đã chèn vào. Vị trí NdeI (CATATG) được đặt ở vùng khởi đầu dịch mã và có thể được sử dụng để xây dựng plasmid biểu hiện các protein nguyên thể.





**Hình 7.8. Vector pAS1.** Vector pAS1 là một plasmid dài khoảng 5,8 kb mang promoter P<sub>L</sub> của bacteriophage λ và vị trí cắt hạn chế duy nhất BamHI định vị ở codon khởi đầu ATG của gen cII của bacteriophage λ. Plasmid này có nguồn gốc từ pKC30, trong đó gen cII của bacteriophage λ được chèn vào ở vị trí HpaI. Gen cII sau đó được cắt bỏ bởi enzyme exonuclease cho tới khi chỉ còn lại codon khởi đầu ATG (G của ATG là nucleotide đầu tiên của vị trí BamHI). Để biểu hiện gen bị mất codon khởi đầu, pAS1 được cắt bằng BamHI và sau đó xử lý với enzyme mung-bean nuclease hoặc nuclease S1 để loại bỏ đầu lồi (đầu tận cùng sợi đơn). Gắn DNA đầu bằng này với đoạn DNA đầu bằng bắt đầu với codon thứ hai của gen được biểu hiện đặt gen đó trong khung với ATG. Các gen được chèn vào theo kiểu này được điều hòa bằng cách chuyển plasmid tái tổ hợp vào trong thể tiềm tan miễn cảm nhiệt độ của bacteriophage λ (cIts857). Các tế bào sinh trưởng tới pha log muộn ở 30°C và sau đó nâng lên 40°C để bất hoạt gen ức chế (repressor) và mở promoter P<sub>L</sub>. Gen được chèn vào cũng có thể được điều hòa bằng hoạt động của protein N ở nutL và nutR để ngăn cản kết thúc ở t<sub>R</sub>.

Hiệu suất dịch mã của mRNA chịu sự chi phối của một vài yếu tố:

- Mức độ bổ sung giữa chuỗi SD và đầu 3' của 16S rRNA.
- Khoảng cách giữa chuỗi SD và codon AUG để chuỗi DNA của eukaryote định vị.
- Nucleotide theo sau AUG ảnh hưởng sự liên kết ribosome.

Đưa codon ATG vào trong gen được tạo dòng và duy trì vị trí liên kết ribosome (RBS) của vi khuẩn bằng cách dùng vector pAS1 (Hình 7.8). Vector pAS1 mang promoter  $P_L$  và RBS của gen *cII* của bacteriophage  $\lambda$ , codon khởi đầu ATG được dung hợp trực tiếp với phần mã hóa đầu tận cùng amino của gen eukaryote muốn biểu hiện. Cách làm này có thể hạn chế việc tối ưu khoảng cách giữa chuỗi SD của vi khuẩn và codon khởi đầu ATG của gen eukaryote để có được sự biểu hiện hiệu quả của gen. Đoạn DNA có gắn promoter và chuỗi SD sau đó được biến nạp vào các chủng *E. coli* thích hợp. Sàng lọc thể biến nạp để thu các khuẩn lạc có mức độ phiên mã cao.

### III. Xác định mức độ biểu hiện của gen được tạo dòng

Nói chung có ba cách thường được dùng để đánh giá mức độ biểu hiện protein ngoại lai của gen được tạo dòng:

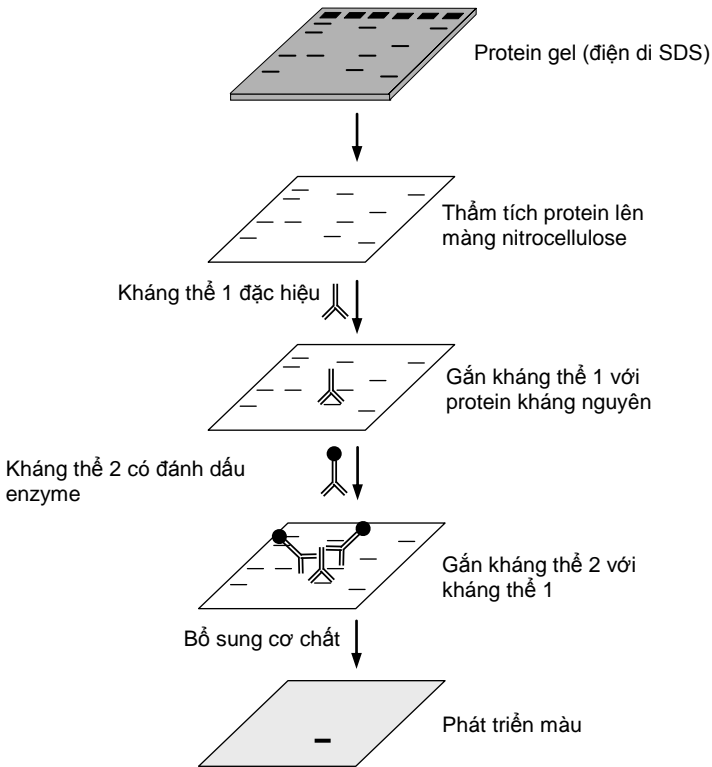
- Điện di polyacrylamide gel để xác định protein có kích thước thích hợp được sản xuất ở mức độ cao trong các tế bào mang vector biểu hiện. Thông thường, protein quan tâm có thể quan sát bằng cách nhuộm gel với Coomassie Brilliant Blue hoặc bằng thuốc nhuộm bạc. Nếu không thấy có băng protein mới khi dùng các thuốc nhuộm này, thì đánh dấu sự trao đổi chất với 100  $\mu\text{Ci}$  của  $[^{35}\text{S}]\text{Met}$  hoặc  $[^{35}\text{S}]\text{Cys}$  trên 1 mL dịch nuôi cấy trong 5 phút. Điện di SDS-polyacrylamide gel và thực hiện phóng xạ tự ghi có thể cho phép phát hiện protein quan tâm.

- Phân tích Western blot bằng cách dùng các kháng thể liên kết đặc hiệu với protein quan tâm đã được thấm tích lên màng nitrocellulose sau khi thực hiện điện di SDS-PAGE (xem chương 2).

- Nếu mức độ biểu hiện thấp thì nên đặt gen *lacZ* cùng hướng với gen được biểu hiện. Như vậy, nếu sự phiên mã hoặc dịch mã hạn chế biểu hiện của gen thì những thay đổi trong hệ thống biểu hiện có thể được kiểm soát bằng những thay đổi trong hoạt tính của  $\beta$ -galactosidase.

## 1. Western blot

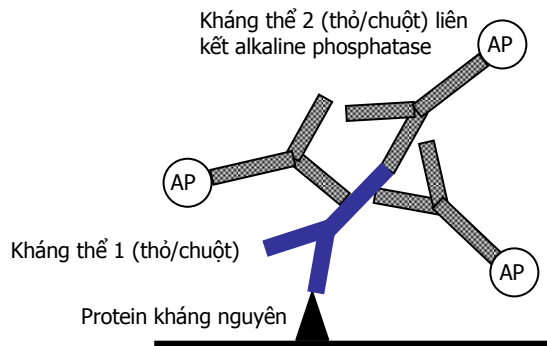
Phản ứng liên kết kháng nguyên-kháng thể có tính đặc hiệu rất cao. Vì vậy, có thể áp dụng phản ứng này để phát hiện sự có mặt và tinh sạch protein. Kháng thể (antibody) được sản xuất khi đưa kháng nguyên vào tổ và được tinh sạch từ máu tổ sau khi gây nhiễm. Những kháng thể tạo ra bằng cách này là những kháng thể đa dòng (polyclonal antibodies-do các tế bào lympho khác nhau tiết ra), do đó chúng có khả năng nhận biết một số kháng nguyên. Ngược lại, kháng thể đơn dòng (monoclonal antibodies) chỉ tương tác với một kháng nguyên nhất định.



**Hình 7.9. Sơ đồ kỹ thuật Western blot**

Kháng thể được đánh dấu bằng enzyme (hoặc bằng chất phát huỳnh quang) để phát hiện protein đặc hiệu (thường được thẩm tích lên màng nitrocellulose sau khi chạy điện di SDS-PAGE, và cố định ở đó) thông

qua kỹ thuật Western blot (hoặc immunoblot) (Hình 7.9). Cơ chế của phản ứng liên kết kháng nguyên-kháng thể được trình bày ở hình 7.10. Sau khi protein trên màng lai gắn với kháng thể thứ nhất đặc hiệu và tiếp đến là kháng thể thứ hai có đánh dấu enzyme (alkaline phosphatase, horse-radish peroxidase...) thì phức hợp này sẽ được liên kết với cơ chất để tạo màu. Sự hiện diện của protein ngoại lai (sản phẩm dịch mã của gen ngoại lai được chuyển vào tế bào vật chủ) sẽ được phát hiện nhờ sự xuất hiện màu của phản ứng lai.



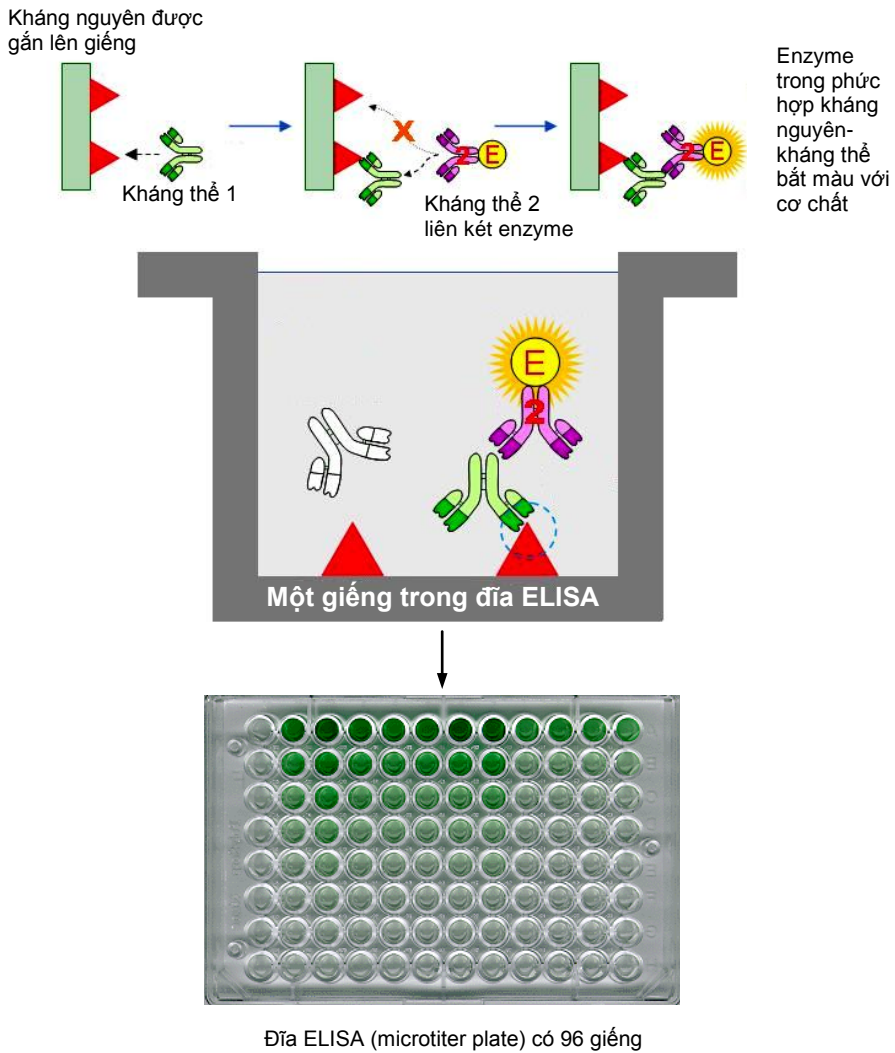
**Hình 7.10. Sơ đồ mô tả liên kết protein (kháng nguyên) với kháng thể thứ nhất đặc hiệu và kháng thể thứ hai có đánh dấu enzyme trong Western blot**

Sự phân bố của protein đặc hiệu trong tế bào và tổ chức mô cũng có thể phát hiện bằng kỹ thuật lai *in situ* (*in situ* hybridization) với nguyên tắc tương tự Western blot. Ngoài ra, kháng thể được sử dụng để tinh sạch protein đặc hiệu bằng kết tủa miễn dịch hoặc sắc ký ái lực (affinity chromatography). Kháng thể đánh dấu được dùng để định lượng kháng nguyên trong kỹ thuật xét nghiệm hấp thụ miễn dịch liên kết enzyme (enzyme-linked immunosorbent assay) gọi tắt là ELISA.

## 2. ELISA

ELISA là một kỹ thuật hóa sinh, cũng dựa trên phản ứng liên kết kháng nguyên-kháng thể, được dùng chủ yếu trong miễn dịch học để phát hiện sự có mặt của kháng thể hoặc kháng nguyên trong mẫu. Tương tự kỹ thuật Western blot, trong ELISA người thường dùng hai loại kháng thể. Một

kháng thể đặc hiệu với kháng nguyên protein (kháng thể thứ nhất). Một kháng thể khác, phản ứng với các phức hợp kháng thể-kháng nguyên, được kết hợp với enzyme (kháng thể thứ hai). Kháng thể thứ hai nhờ có liên kết với enzyme (như tên của kỹ thuật xét nghiệm) nên có thể bắt màu với cơ chất để tạo ra tín hiệu (Hình 7.11).



**Hình 7.11. Sơ đồ kỹ thuật ELISA**

Do ELISA có thể thực hiện để đánh giá sự có mặt của kháng nguyên hoặc kháng thể trong mẫu bằng phương pháp quang phổ (đo độ hấp thụ quang), cho nên nó là công cụ hữu ích để xác định nồng độ (định lượng) kháng nguyên và kháng thể.

#### **IV. Tăng mức độ biểu hiện của gen được tạo dòng**

Các mức độ biểu hiện của gen ngoại lai trong *E. coli* thường được kiểm tra bằng thử nghiệm chức năng hoặc bằng các xét nghiệm miễn dịch khác nhau. Tuy nhiên, nếu cố đạt đến mức tối đa sự biểu hiện của sản phẩm gen thì các phân tích như thế là phức tạp. Để khắc phục khó khăn này khi sản phẩm gen eukaryote không dung hợp được biểu hiện người ta đã sử dụng phương pháp cho DNA ngoại lai được dung hợp với đoạn tương ứng của gen *lacZ* cho phép gen ngoại lai sao chép lại và dịch mã một cách hiệu quả. Plasmid được dùng trong phương pháp này là pLG mang gen *lacZ* mã hóa đầu tận cùng carboxyl có hoạt tính  $\beta$ -galactosidase. Tuy nhiên,  $\beta$ -galactosidase vẫn chưa được tổng hợp do còn thiếu promoter và trình tự SD ở plasmid. Do đó, một promoter di động (portable promoter) sẽ được gắn phía trước gen dung hợp và điều chỉnh sự biểu hiện ở mức độ cao protein dung hợp có hoạt tính  $\beta$ -galactosidase. Các plasmid có các đoạn promoter được đặt tối ưu có thể nhận biết bằng cách biến nạp vi khuẩn  $Lac^-$  và thu được hoạt tính  $\beta$ -galactosidase trên các đĩa chỉ thị lactose (lactose indicator). Các plasmid này sau đó có thể được dùng để tái cấu trúc gen eukaryote không hợp nhất mà nó được biểu hiện ở mức độ cao.

Mức độ biểu hiện thấp của gen được tạo dòng có thể là do RNA không ổn định, sự kết thúc chưa hoàn chỉnh, sự dịch mã không hiệu quả, hoặc chính protein không ổn định.

#### **V. Tinh sạch protein**

##### *1. Thể vùi và phương pháp hòa tan thể vùi*

###### **1.1. Thể vùi**

Protein có mức độ biểu hiện cao trong *E. coli* thường tạo ra các hạt nhỏ không hòa tan ở trong tế bào chết (thể vùi), có thể quan sát bằng kính hiển ngược pha và tách khỏi dịch tan của tế bào sống bằng phương pháp ly

tâm. Các tế bào biểu hiện mức độ cao các protein ngoại lai được cô lại bằng ly tâm và phá vỡ bằng các kỹ thuật cơ học, siêu âm, hoặc dùng lysozyme kết hợp với chất tẩy. Các thể vùi (inclusion) được tạo tiểu thể bằng ly tâm và rửa với Triston X-100 và EDTA hoặc urea.

Để thu được chất hòa tan, protein hoạt động, các thể vùi được rửa phải hòa tan và sau đó phải được tái cuộn xoắn. Mỗi protein có thể cần một phương pháp hòa tan khác nhau và thường được xác định theo kinh nghiệm. Các điều kiện khác nhau (ví dụ: guanidine HCl [5-8 M], urea [6-8 M], SDS, alkaline pH, hoặc acetonitrile/propanol) có thể được sử dụng để hòa tan các thể vùi. Trong nhiều trường hợp, chỉ cần pha loãng đơn giản dịch chiết hòa tan trong một đệm thích hợp là đủ. Nếu protein chứa các cầu nối disulfide, người ta cần phải tiến hành oxy hóa và khử glutathione. Một đôi khi cần bổ sung đồng dung môi (co-solvent) như PEG, hoặc các chất tẩy rửa như Triton X-100, Tween 20 hoặc Zwittergent 3-16.

Sau khi hòa tan thành công, có thể thử nghiệm các phương pháp tái cuộn xoắn khác nhau bao gồm sự pha loãng hoặc thẩm tách. Hiệu suất tạo ra protein hoạt động hoặc protein có các liên kết disulfite giống như protein gốc phụ thuộc vào nồng độ, độ tinh sạch và kích thước của chuỗi polypeptide; độ pH và cường lực ion của dung môi; và tỷ lệ tái cuộn xoắn của chúng. Các nhân tố khác bao gồm số lượng của các liên kết disulfite và bản chất của chính protein. Trong một số trường hợp, các protein hòa tan có thể rất khó tái cuộn xoắn và người ta thấy rằng việc bổ sung chaperonins có thể giúp ích cho các quá trình này. Chaperonins là các protein cần để đảm bảo cuộn xoắn chính xác các *in vivo* protein, với khả năng thuận lợi của chaperonin tái tổ hợp chúng được dùng để xúc tác cho quá trình tái cuộn xoắn chính xác của các *in vitro* protein.

Nguyên nhân tạo thành các thể vùi chưa được biết đầy đủ, vì không phải tất cả protein biểu hiện ở mức độ cao đều tạo thành thể vùi. Tế bào vật chủ cũng thể hiện một vai trò quan trọng. Cấu trúc chính xác của protein tái tổ hợp cũng có thể ảnh hưởng sự tạo thành các thể vùi. Một nghiên cứu được thực hiện với  $\gamma$ -interferon người tái tổ hợp đã cho thấy chỉ một vài thay đổi amino acid cũng có thể ảnh hưởng đến sự chuyển hóa giữa các biểu hiện của protein hòa tan và không hòa tan trong *E. coli*.

## 1.2. Phương pháp hòa tan thể vùi

### 1.2.1. Làm tan tế bào

Ly tâm thu tiểu thể tế bào *E. coli*, tái huyền phù tiểu thể bằng phenylmethylsulfonylfluoride (PMSF) và lysozyme, đặt ở nhiệt độ phòng (RT) khoảng 20 phút (thỉnh thoảng khuấy). Sau đó, bổ sung deoxycholic acid, đặt ở 37°C và khuấy cho tới khi dịch tan trở nên sền sệt, bổ sung DNase I và để 30 phút ở RT.

### 1.2.2. Tinh sạch và rửa thể vùi

- **Phương pháp 1.** Ly tâm dịch tan thu được ở bước trên, tái huyền phù tiểu thể bằng Tris-ton X-100 và EDTA, giữ ở RT 5 phút. Ly tâm 15 phút lấy tiểu thể và tái huyền phù trong nước. Trộn dung dịch với đệm 2× SDS gel-loading dye và phân tích bằng điện di polyacrylamide gel để xác định protein quan tâm có trong tiểu thể không.

- **Phương pháp 2.** Ly tâm dịch tan, tái huyền phù tiểu thể bằng nước. Sau đó, ly tâm lại và tái huyền phù tiểu thể trong Tris.HCl có bổ sung urea. Ly tâm và tái huyền phù tiểu thể bằng nước. Trộn dung dịch với đệm 2× SDS gel-loading dye và phân tích điện di polyacrylamide gel để xác định điều kiện rửa thích hợp cho protein.

### 1.2.3. Hòa tan các thể vùi

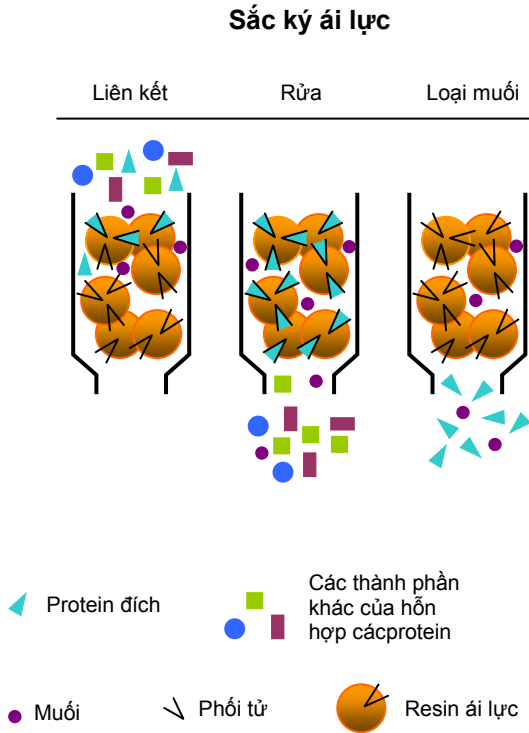
Treo các tiểu thể được rửa trong đệm dung ly chứa PMSF và urea, để 1 giờ ở RT. Bổ sung dung dịch có  $\text{KH}_2\text{PO}_4$ , EDTA và NaCl để 30 phút ở RT, duy trì pH 10,7 bằng KOH. Điều chỉnh pH tới 8,0 bằng HCl để 30 min ở RT. Ly tâm thu tiểu thể và tái huyền phù bằng đệm 1× SDS gel-loading dye. Phân tích điện di để xác định mức độ hòa tan.

## 2. Các đuôi ái lực

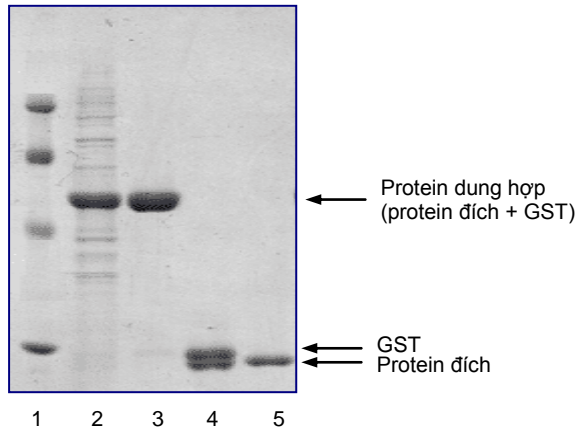
Khái niệm đuôi ái lực xuất hiện khi thiết kế di truyền với mục đích giúp cho sự tinh sạch protein hoặc enzyme hiệu quả hơn. Gen của protein quan tâm được dung hợp với chuỗi DNA mã hóa cho một số trình tự amino acid sẽ đơn giản hóa quá trình tinh sạch protein, bằng cách biến đổi các tính chất của nó trong một kiểu có thể dự đoán. Một trong các trường hợp đầu tiên là dung hợp di truyền của một số gốc arginine với C-terminus của



urogastrone. Gen này sẽ sản xuất một protein liên kết mạnh với khuôn trao đổi ion cho cation. Đuôi polyarginine sau đó được loại bỏ bằng cách dùng enzyme bất động carboxypeptidase A.



**Hình 7.12. Tinh sạch protein đích bằng sắc ký ái lực.** Cho hỗn hợp protein tái tổ hợp đi qua cột sắc ký có chứa một phôi tử liên kết đặc hiệu với các protein mang một đuôi ái lực (ví dụ: His, Histidine hoặc GST, glutathione-S-transferase). Các chất bẩn sẽ được rửa trôi khỏi cột, và các protein được liên kết trên cột sau đó sẽ được rửa giải ở dạng tinh sạch. Đuôi ái lực có nhiều ưu điểm trong tinh sạch protein. Trong IMAC (immobilized metal ion affinity chromatography), His sẽ liên kết chọn lọc cao với  $\text{Ni}^{2+}$  hoặc các kim loại chuyển tiếp khác được bất động trên phôi tử, protein có gắn đuôi ái lực có thể được rửa giải chọn lọc bằng imidazole. Protein được gắn đuôi GST liên kết với glutathione như một phôi tử, và được rửa giải với các dung dịch glutathione. Các protein có đuôi dung hợp GST mang hoạt tính enzyme chỉ có thể được tinh sạch dưới các điều kiện tự nhiên. Ngược lại, các protein gắn đuôi His có thể được tinh sạch dưới các điều kiện tự nhiên hoặc biến tính.



**Hình 7.13. Phân tích SDS-PAGE và nhuộm bằng Coomassie Blue các sản phẩm protein sau khi tinh sạch bằng sắc ký ái lực.** 1: Chuẩn khối lượng phân tử của protein. 2: Protein hòa tan tổng số. 3: Protein dung hợp (protein đích và GST) được rửa giải khỏi cột glutathione sepharose. 4: Protein dung hợp được cắt bằng PresCission Protease cho ra 2 băng là GST và protein đích. 5: Protein đích đã được tinh sạch sau khi cho đi qua IMAC.

Khó khăn chủ yếu ở các dung hợp ái lực để tinh sạch protein là phải loại bỏ thành công đuôi ái lực và các tác nhân được sử dụng. Vấn đề này có thể được giải quyết từng phần bằng sự biến nạp di truyền các điểm đặc hiệu cao cho protease hoặc cho sự cắt bỏ các liên kết acid không bền ở chỗ nối của protein tái tổ hợp và đuôi ái lực. Nhiều phương pháp phân cắt đã được gợi ý. Sử dụng các protease đặc hiệu là thích hợp hơn cả, bởi vì có thể ứng dụng trong các điều kiện “nhẹ nhàng”, chủ yếu là thrombin, enterokinase và Factor X<sub>a</sub>. Tất cả những protein này hoạt động ở 37°C và có thể phân cắt ở các vị trí bên trong protein quan tâm, hoặc do mất tính đặc hiệu hoặc từ sự nhiễm bản các protease. Cuối cùng, các bước sắc ký tiếp theo sẽ được yêu cầu để loại bỏ đuôi ái lực được phân cắt và protease.

Trong sự phát triển gần đây của lĩnh vực này, người ta đã sử dụng dạng tái tổ hợp của protease 3C từ rhinovirus của người. Protease này có kích thước nhỏ (20 kDa) và có một tính đặc hiệu rất hạn chế. Nó được biểu hiện như là một protein tái tổ hợp được dung hợp với glutathione-S-transferase. Điều này có một vài ưu điểm, bản thân protease có thể được tinh

sạch bằng sắc ký ái lực trên glutathione-Sepharose. Nếu protein đích được biểu hiện như là một sự dung hợp với glutathione-S-transferase thì nó có thể được tinh sạch trên cột glutathione-Sepharose. Sau khi xử lý với protease, protein đích có thể được phân tách khỏi đuôi glutathione-S-transferase và protease bằng cách chuyển qua cột glutathione-Sepharose thứ hai.

Một cách khác, phản ứng phân cắt có thể được đặt trên cột glutathione-Sepharose thứ nhất bằng cách bổ sung protease vào đệm của cột, vì thế tránh được sự cần thiết cho một cột thứ hai. Protease này có sẵn ở dạng thương mại dưới tên PreCission Protease do Amersham Pharmacia Biotech sản xuất (Hình 7.12 và 7.13).

### **Tài liệu tham khảo/đọc thêm**

**1. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K.** 2002. Short Protocol in Molecular Biology. Vol 1 and 2. 5<sup>th</sup> ed. *John Wiley & Sons, Inc.* USA.

**2. Maniatis T, Fritsch EF and Sambrook J.** 1989. Molecular Cloning-A Laboratory Manual. *Cold Spring Harbor Laboratory Press*, USA.

**3. Ohman DE.** 1989. Experiments in Gene Manipulation. *Prentice Hall*, Englewood Cliffs, New Jersey, USA.

**4. Primrose SB, Twyman R and Old RW.** 2001. Principles of Gene Manipulation. 6<sup>th</sup> ed. *Blackwell Science*, Oxford, UK.

**5. Rapley R and Walker JM.** 1998. Molecular Biomethods Handbook. *Humana Press Inc.* New Jersey, USA.

**6. Rosenberg IM.** 1996. Protein Analysis and Purification-Benchtop Techniques. *Birkhäuser*, Boston, USA.

**7. Surzycki S.** 2000. Basic Techniques in Molecular Biology. *Springer-Verlag*, Berlin, Heidelberg, Germany.

**8. Walker JM.** 2002. The Protein Protocols Handbook. 2<sup>nd</sup> ed. *Humana Press Inc.* Totowa, New Jersey, USA.