

Tin Sinh học Bioinformatics

Chương 1. Giới thiệu sinh học phân tử

TS. Nguyễn Hồng Quang
Khoa Kỹ thuật máy tính

Leader of Bioinformatics Group, BK.AI center
Trường Công nghệ thông tin và Truyền thông
Trường Đại học Bách Khoa Hà Nội

Nội dung

1.1. Ứng dụng của Bioinformatics

1.2. Hoạt động cơ thể người

1.3. Vai trò và chức năng của protein: enzyme, hormone, kháng thể.

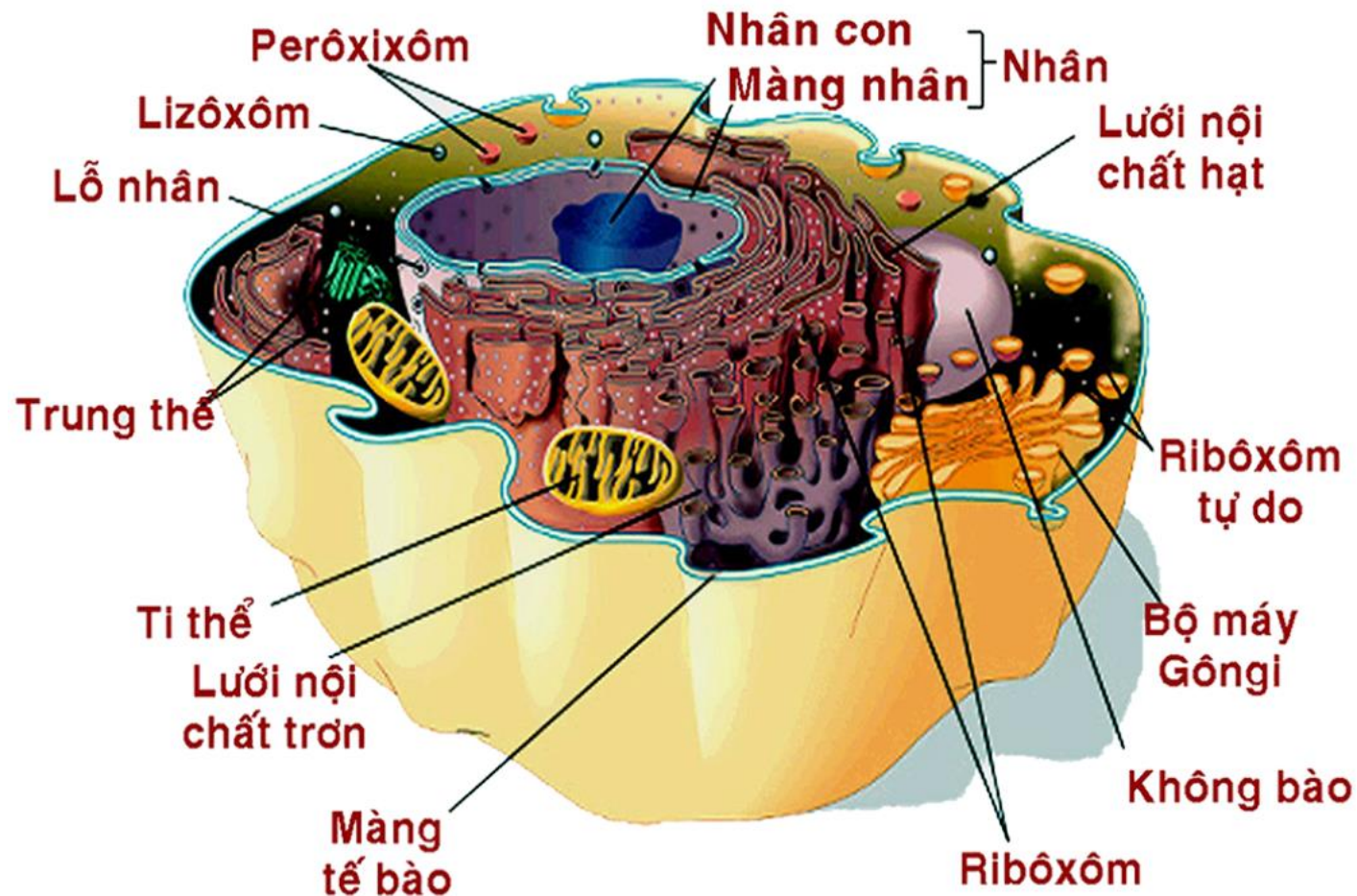
1.4. Cấu tạo của tế bào

1.5. Học thuyết trung tâm (Central Dogma)

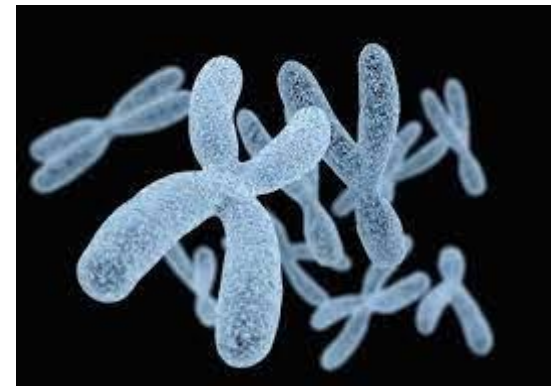
1.6. Các công cụ công nghệ sinh học cơ bản

1.4. Cấu tạo của tế bào

CẤU TRÚC TẾ BÀO ĐỘNG VẬT

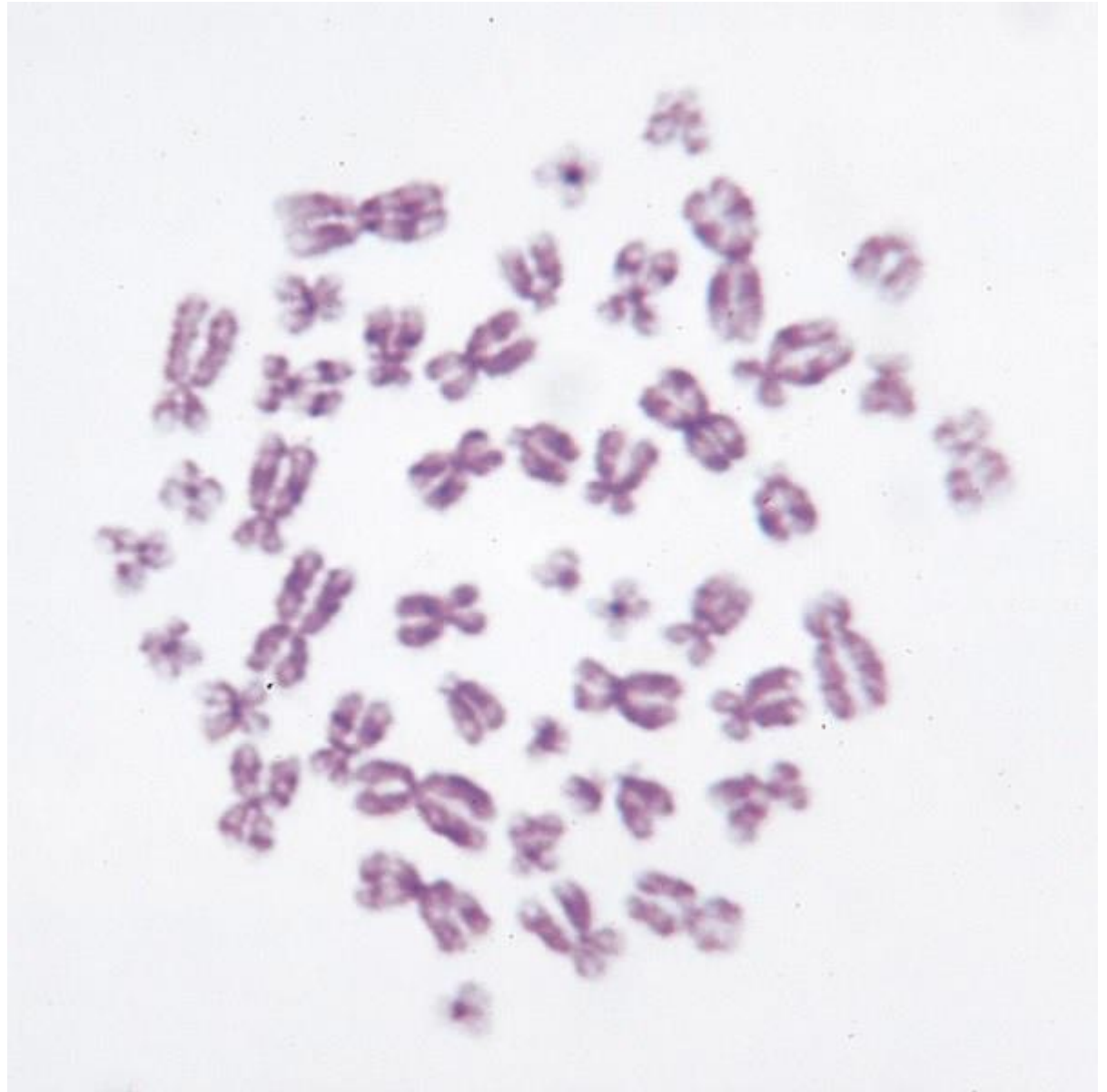


Vai trò nhiễm sắc thể - chromosome



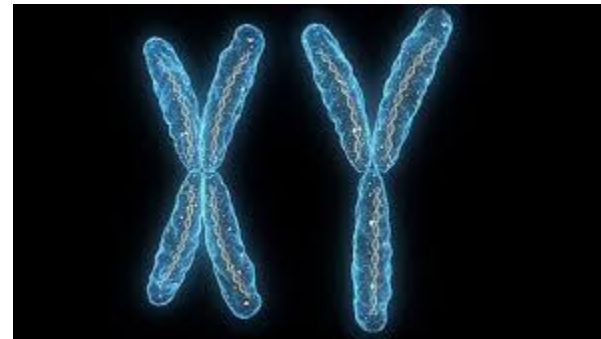
- Nhiễm sắc thể (NST) là vật liệu di truyền tồn tại trong nhân tế bào
- Có khả năng nhuộm màu đặc trưng bằng thuốc nhuộm kiềm tính
- Nhiễm sắc thể là **cấu trúc mang gen** và tự **nhân đôi được**, nhờ đó các tính trạng di truyền được sao chép lại qua các thế hệ cơ thể. gen từ đó sinh ra protein

Quan sát
tiêu bản
nhiễm sắc
thể người
bằng kính
hiển vi



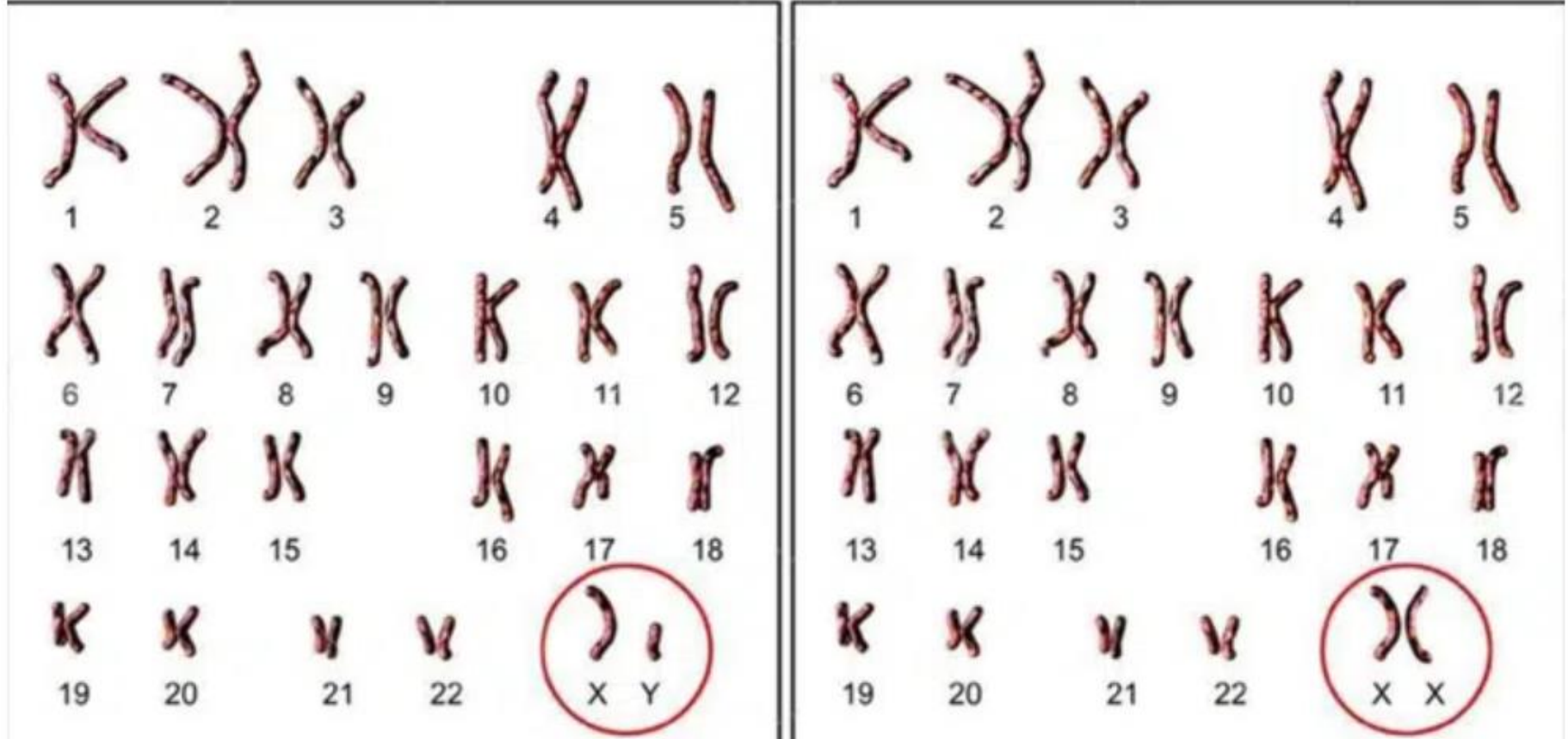
Ví dụ: nhiễm sắc thể giới tính X và Y

- Bình thường trong tế bào mỗi người có một cặp nhiễm sắc thể giới tính.
- Nữ giới có hai nhiễm sắc thể X trong khi nam giới có một nhiễm sắc thể X và một nhiễm sắc thể Y.



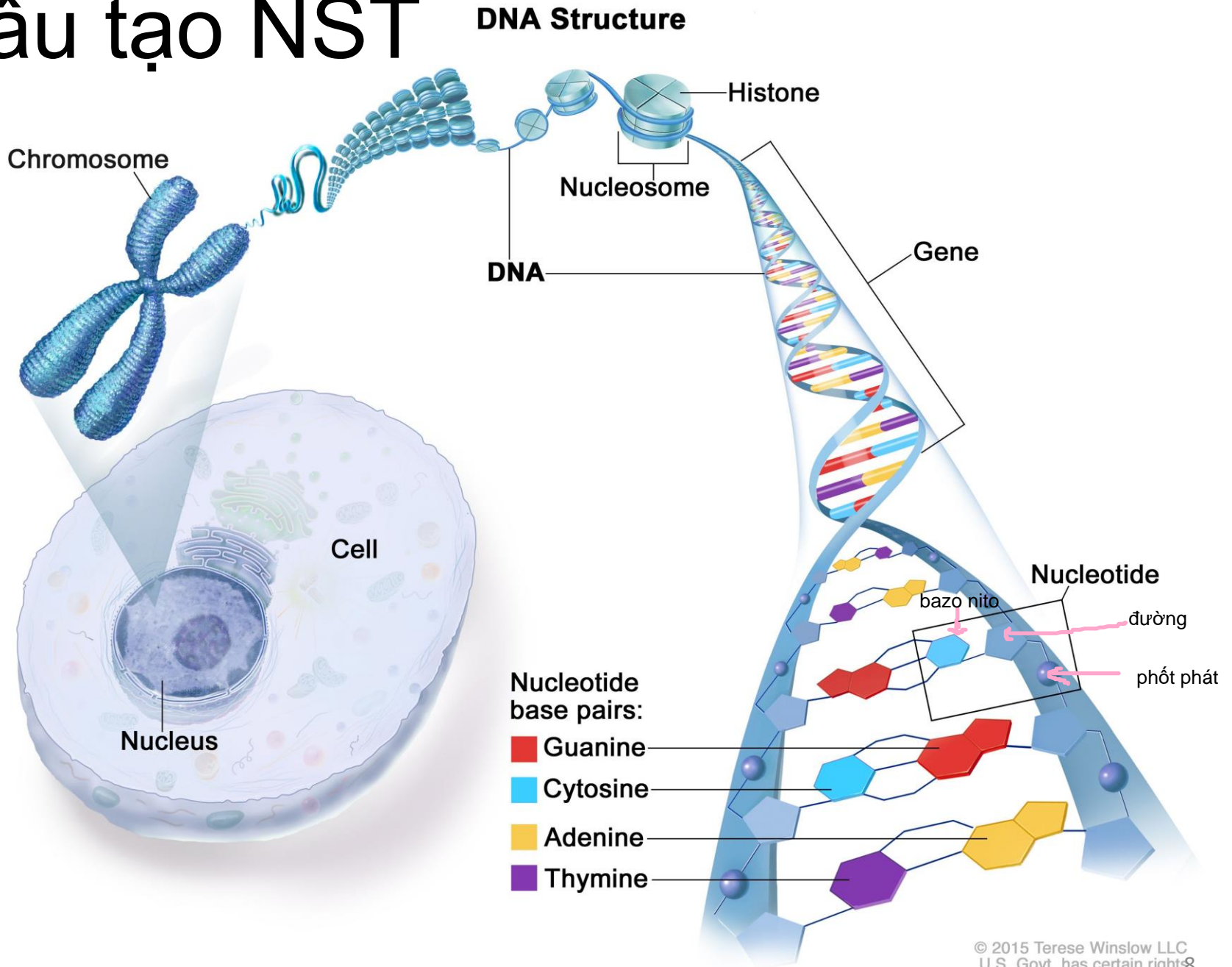
<https://ihope.vn/nhiem-sac-the-x/>

Bảng NST của nam (trái) và nữ (phải)



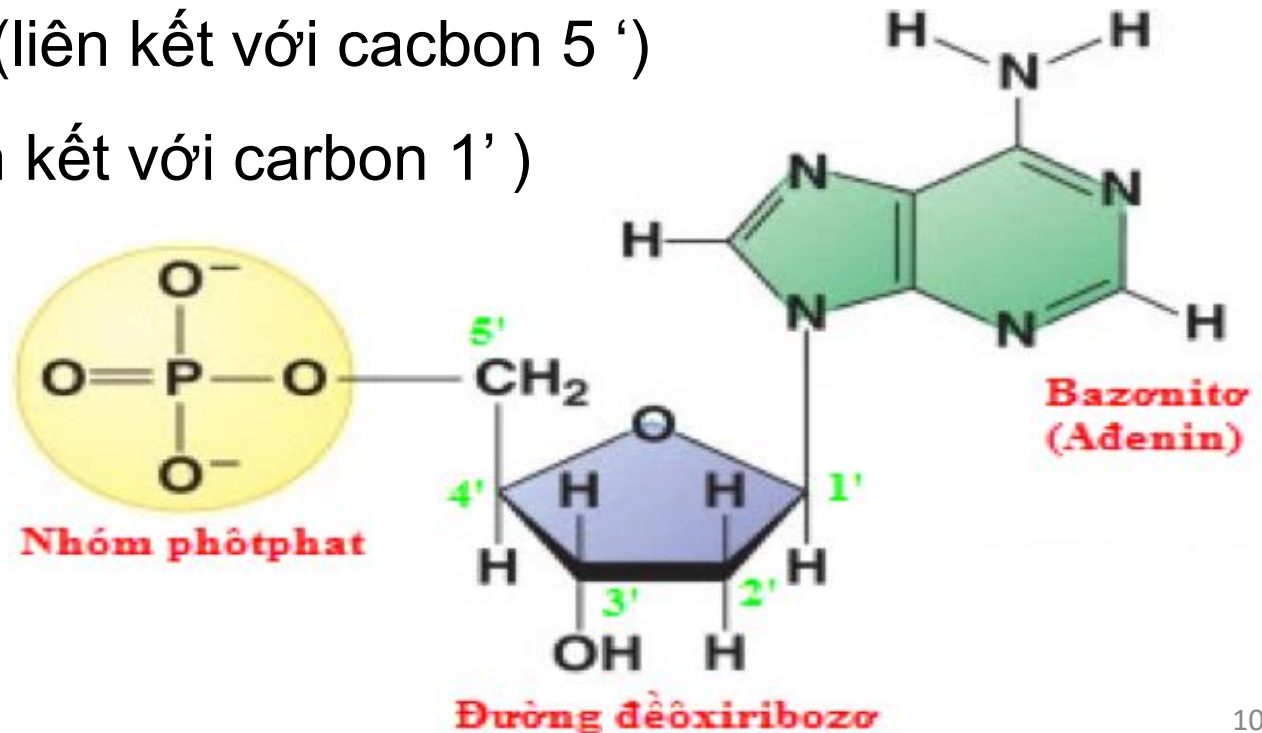
- Một bộ DNA hoàn chỉnh chứa 3 tỷ bazơ, 20.000 gen và 23 cặp nhiễm sắc thể.
- Chúng ta thừa hưởng một nửa DNA từ bố và một nửa từ mẹ.

Cấu tạo NST



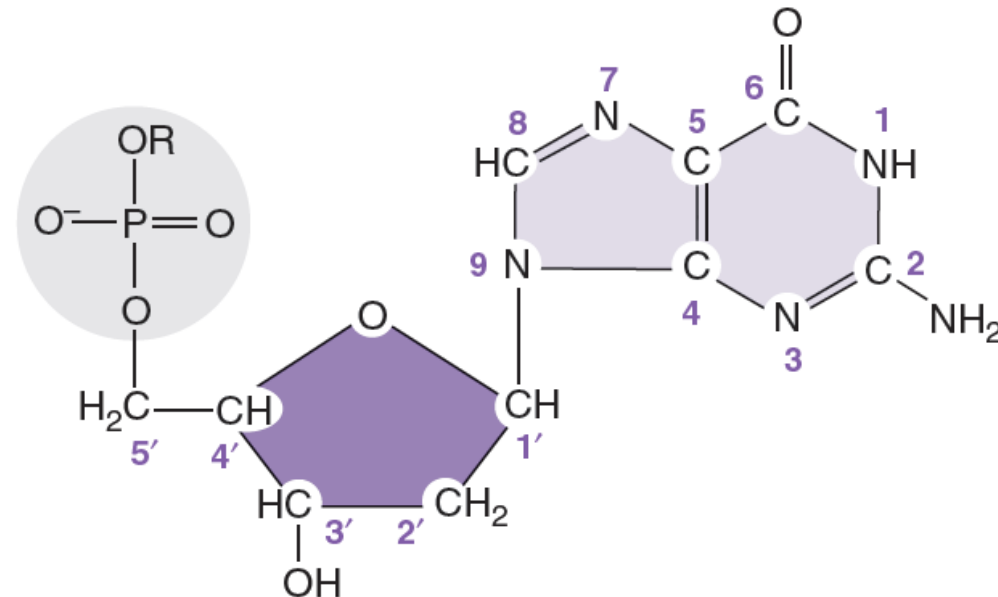
Nucleotide

- DNA được cấu tạo từ 4 loại Nucleotide: Adenine (A), Guanine (G), Thymine (T), Cytosine (C)
- Nucleotide bao gồm ba phần:
 - Deoxyribose (Đường)
 - Phốt phát (liên kết với cacbon 5 ')
 - Bazơ (liên kết với carbon 1')



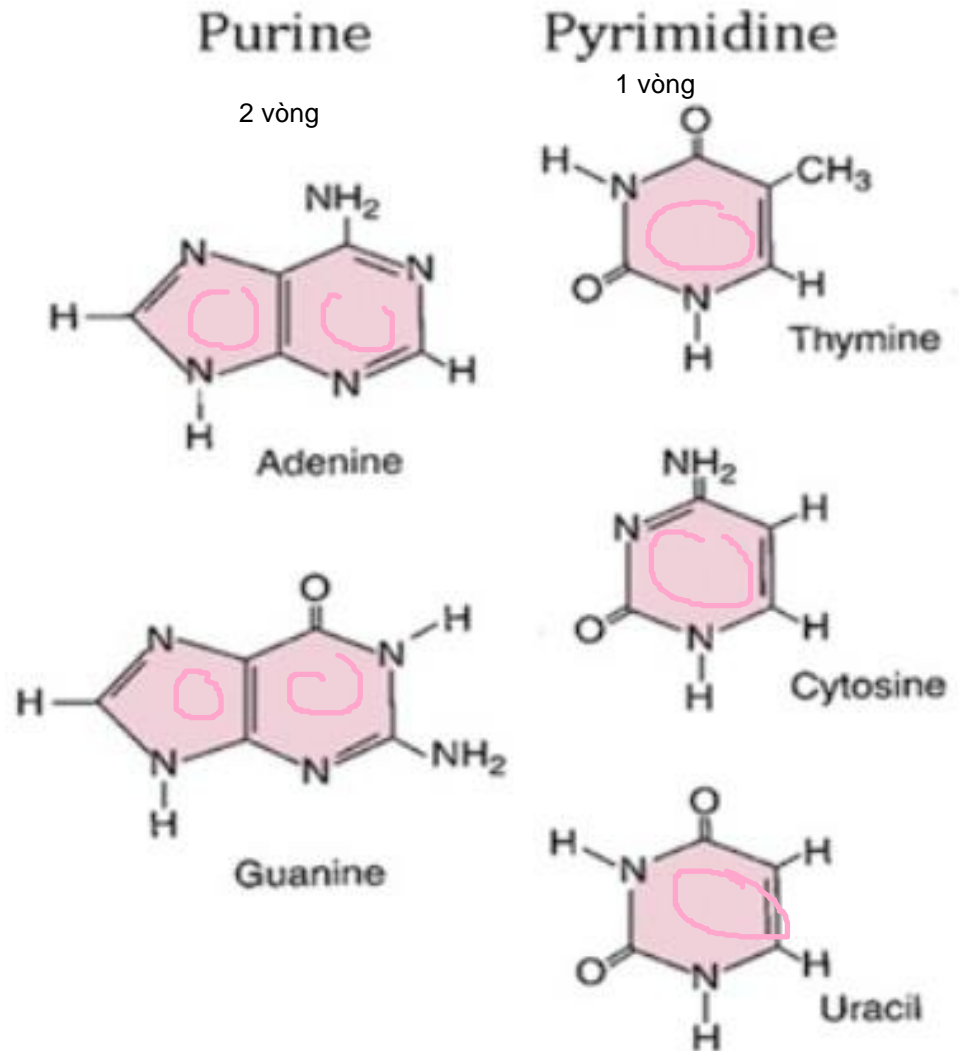
Đánh số vị trí carbon của một nucleotide monophosphate

- Các nguyên tố cơ bản của bazơ nitơ được đánh số từ 1 đến 9.
- Các nguyên tố đường được đánh số từ 1' đến 5'.
- Nhóm photphat liên kết với cacbon 5'
- Carbon 1' giữ bazơ nitơ.

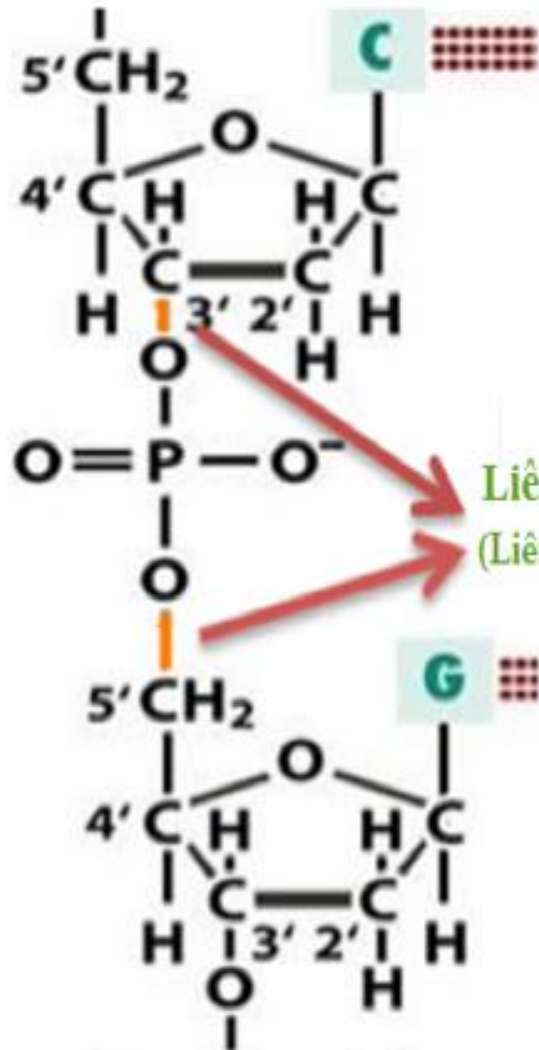


Các loại nucleotide

- A, G được gọi là purin. Chúng có cấu trúc 2 vòng.
- C, T, U được gọi là pyrimidine. Chúng có cấu trúc 1 vòng.

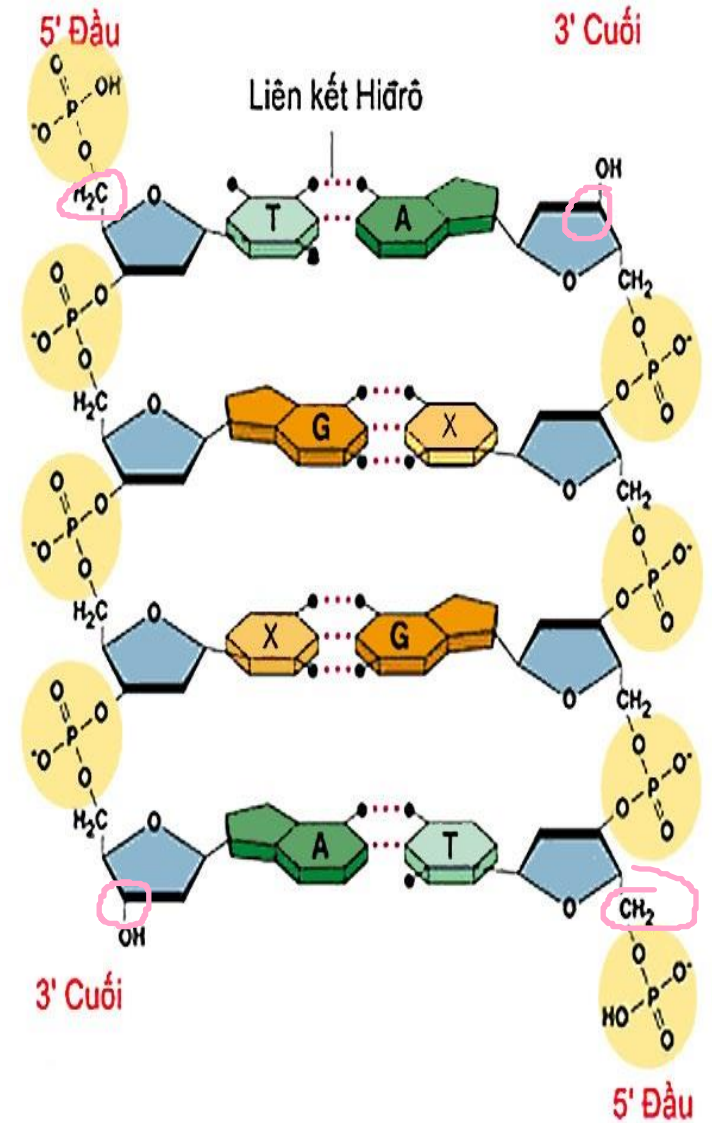


Cấu tạo DNA



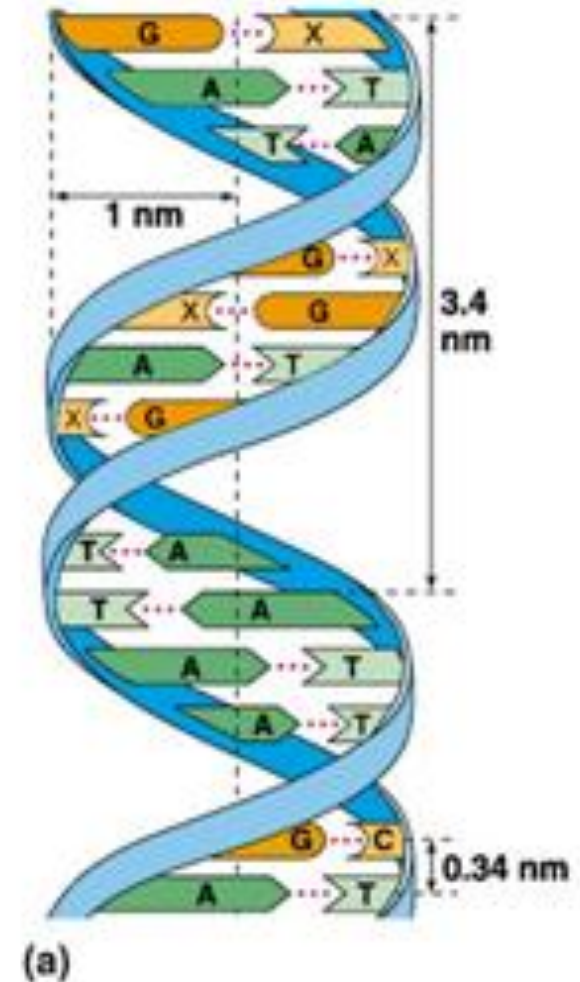
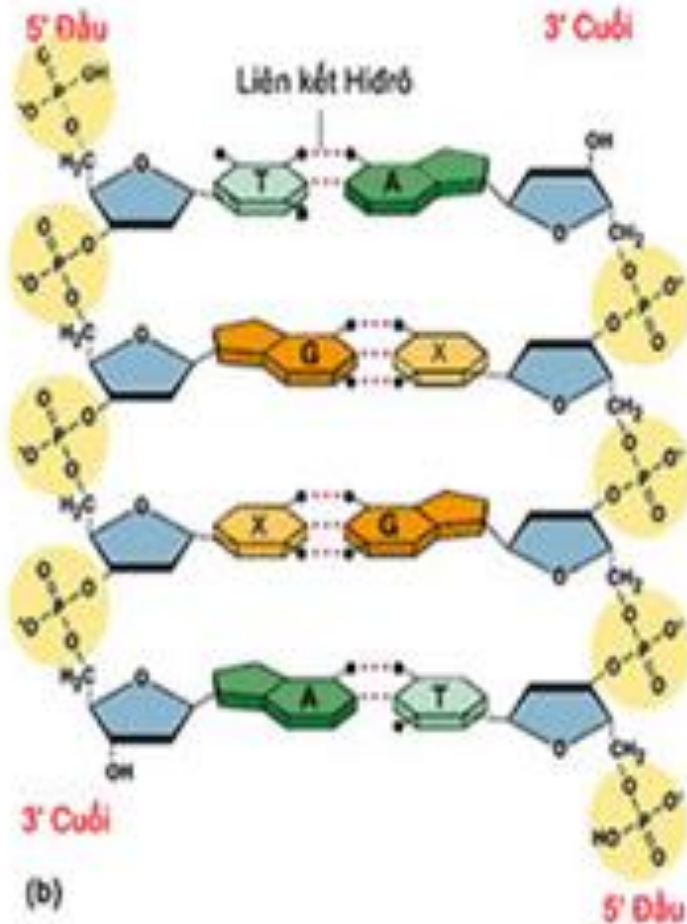
Kết nối giữa 2 nucleotit

Liên kết este thứ 1 và thứ 2
(Liên kết photphodiester giữa 2 đường)



Cấu trúc xoắn kép của DNA

Bộ gen người: có 3G (3 tỷ) cặp bazơ, được tổ chức thành 23 cặp nhiễm sắc thể



James D. Watson and Francis Harry Compton Crick

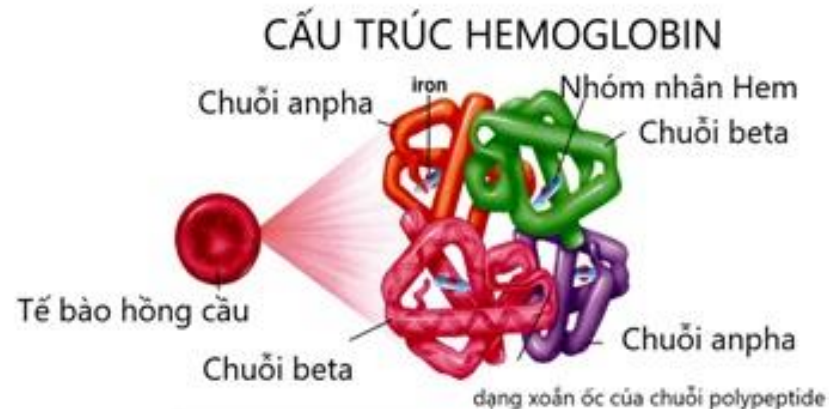


Giải Nobel về sinh lý và y khoa vào năm 1962

Protein

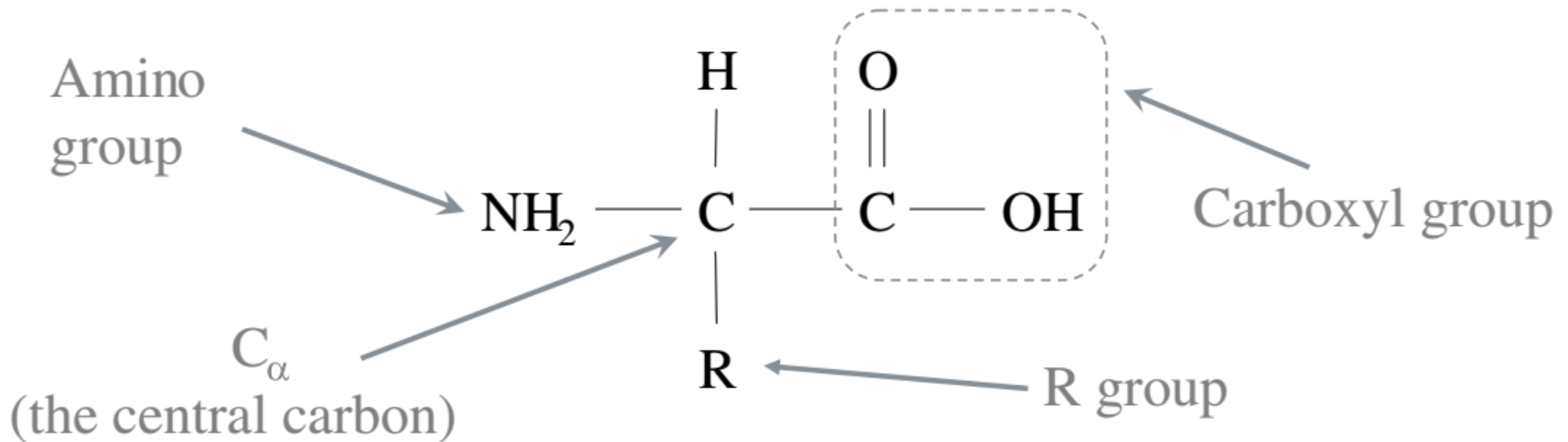
Protein

- Protein là một chuỗi được tạo ra từ 20 axit amin.
- Chiều dài nằm trong khoảng từ 20 đến hơn 5000 axit amin.
- Trung bình, protein chứa khoảng 350 axit amin.
- Protein gấp lại thành hình dạng ba chiều, tạo thành các khối và thực hiện hầu hết các phản ứng hóa học trong tế bào.



Axit amin (Amino acid)

- Mỗi axit amin bao gồm
 - Nhóm amino
 - Nhóm cacboxyl
 - Nhóm R

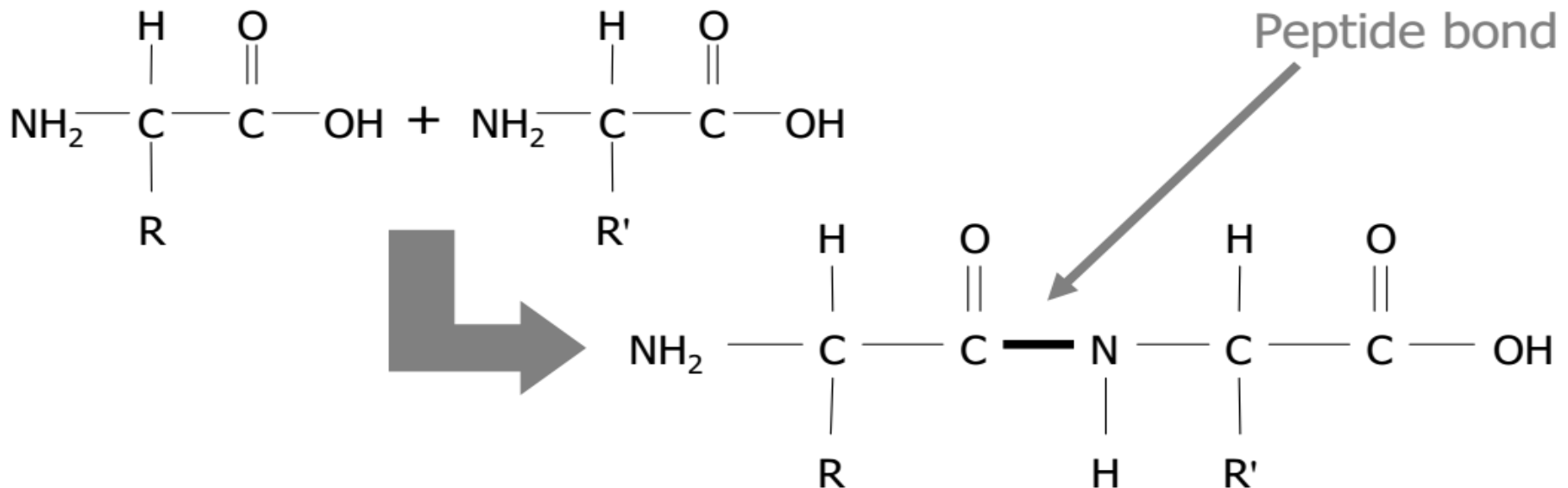


Tóm tắt các tính chất của axit amin

Amino Acid	1-Letter	3-Letter	Avg. Mass (Da)	volume (Å ³)	Side chain polarity	Side chain acidity or basicity	Hydropathy index
Alanine	A	Ala	89.09404	67	non-polar	Neutral	1.8
Cysteine	C	Cys	121.15404	86	polar	basic (strongly)	-4.5
Aspartic acid	D	Asp	133.10384	91	polar	Neutral	-3.5
Glutamic acid	E	Glu	147.13074	109	polar	acidic	-3.5
Phenylalanine	F	Phe	165.19184	135	polar	neutral	2.5
Glycine	G	Gly	75.06714	48	polar	acidic	-3.5
Histidine	H	His	155.15634	118	polar	neutral	-3.5
Isoleucine	I	Ile	131.17464	124	non-polar	neutral	-0.4
Lysine	K	Lys	146.18934	135	polar	basic (weakly)	-3.2
Leucine	L	Leu	131.17464	124	non-polar	neutral	4.5
Methionine	M	Met	149.20784	124	non-polar	neutral	3.8
Asparagine	N	Asn	132.11904	96	polar	basic	-3.9
Proline	P	Pro	115.13194	90	non-polar	neutral	1.9
Glutamine	Q	Gln	146.14594	114	non-polar	neutral	2.8
Arginine	R	Arg	174.20274	148	non-polar	neutral	-1.6
Serine	S	Ser	105.09344	73	polar	neutral	-0.8
Threonine	T	Thr	119.12034	93	polar	neutral	-0.7
Valine	V	Val	117.14784	105	non-polar	neutral	-0.9
Tryptophan	W	Trp	204.22844	163	polar	neutral	-1.3
Tyrosine	Y	Tyr	181.19124	141	non-polar	neutral	4.2

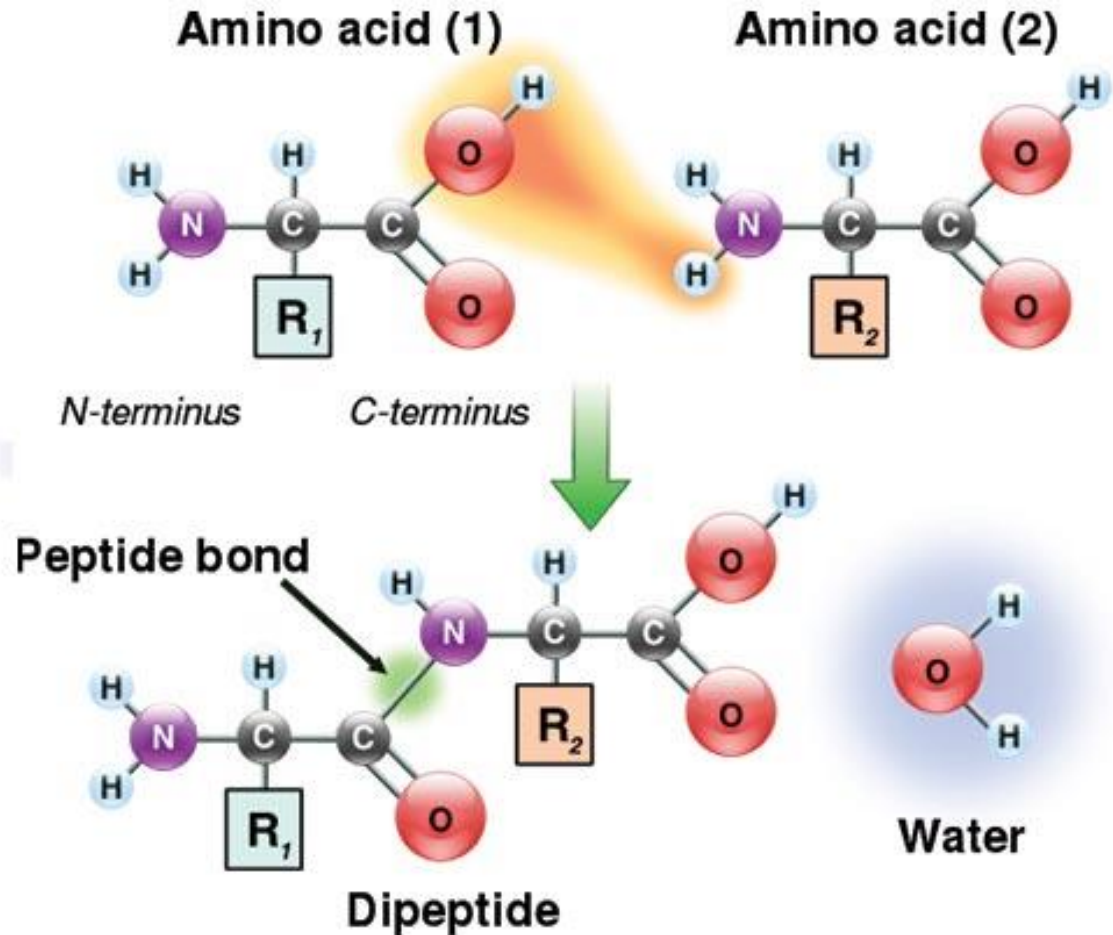
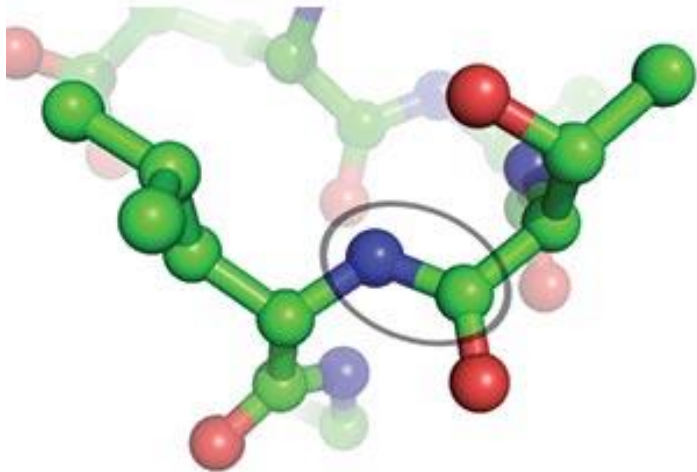
Polypeptide

- Protein hoặc chuỗi polypeptit được hình thành bằng cách nối các axit amin với nhau thông qua liên kết peptit.
- Một đầu của polypeptit là nhóm amin, được gọi là N-endinus.
- Đầu kia của polypeptit là nhóm cacboxyl, được gọi là đầu cuối C.



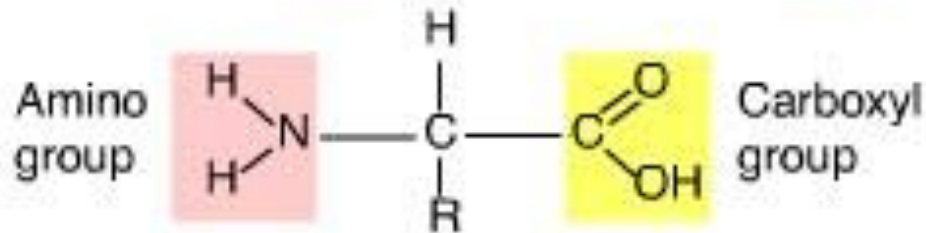
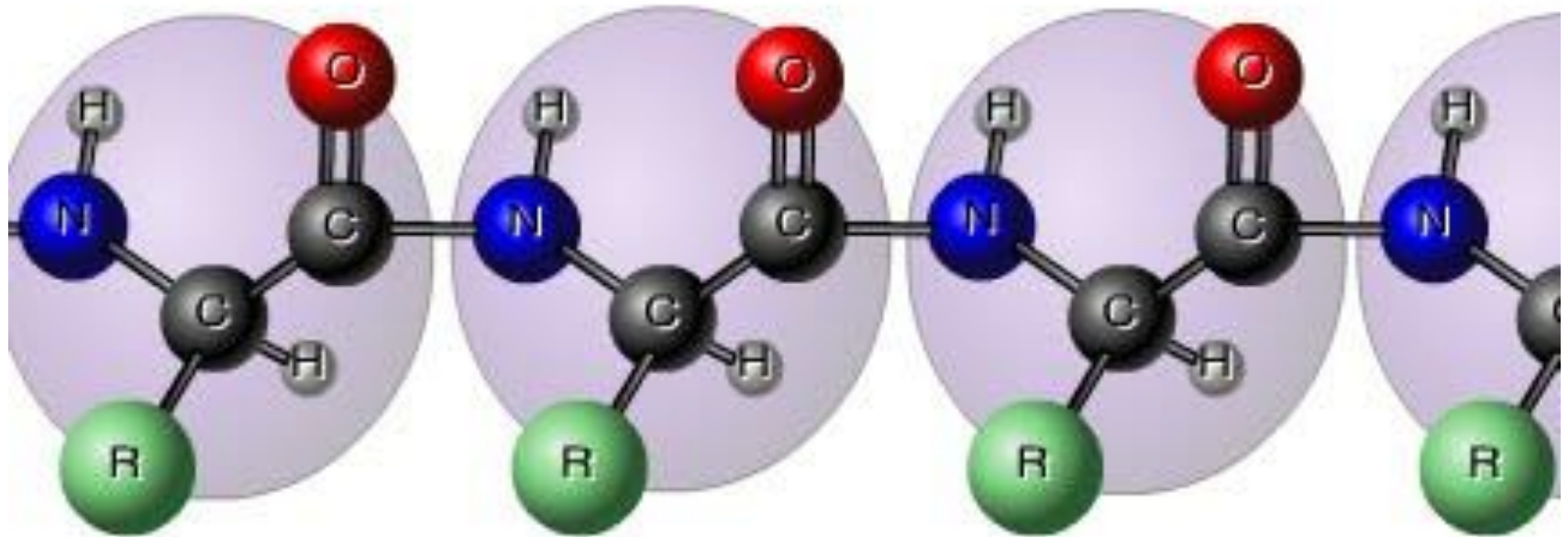
Liên kết peptit

Peptide and Peptide bond



<https://microbenotes.com/peptide-bond/>

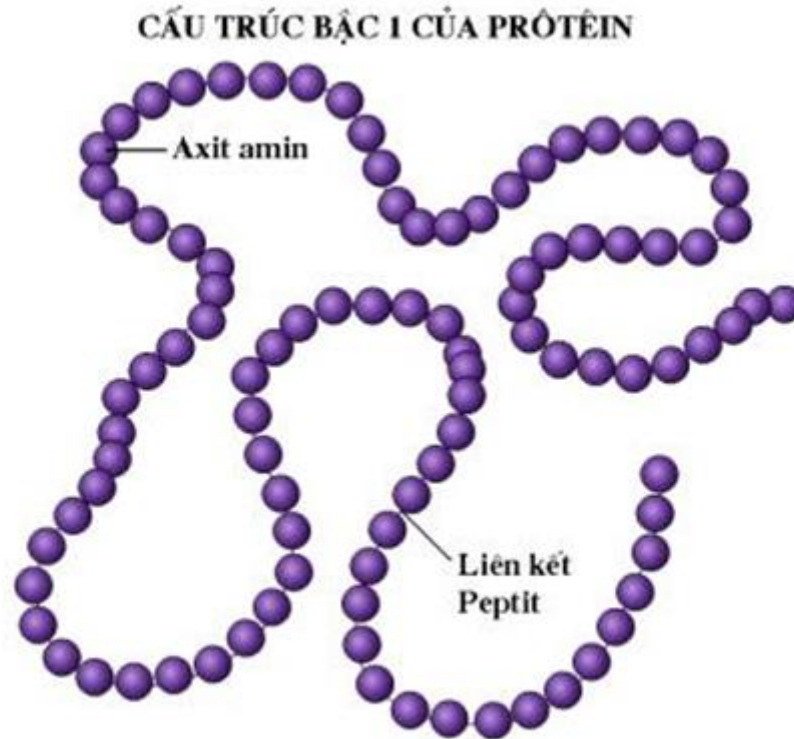
Chuỗi polypeptit



Cấu trúc vật lý của protein

Cấu trúc bậc 1

- Chuỗi polypeptide có dạng mạch thẳng

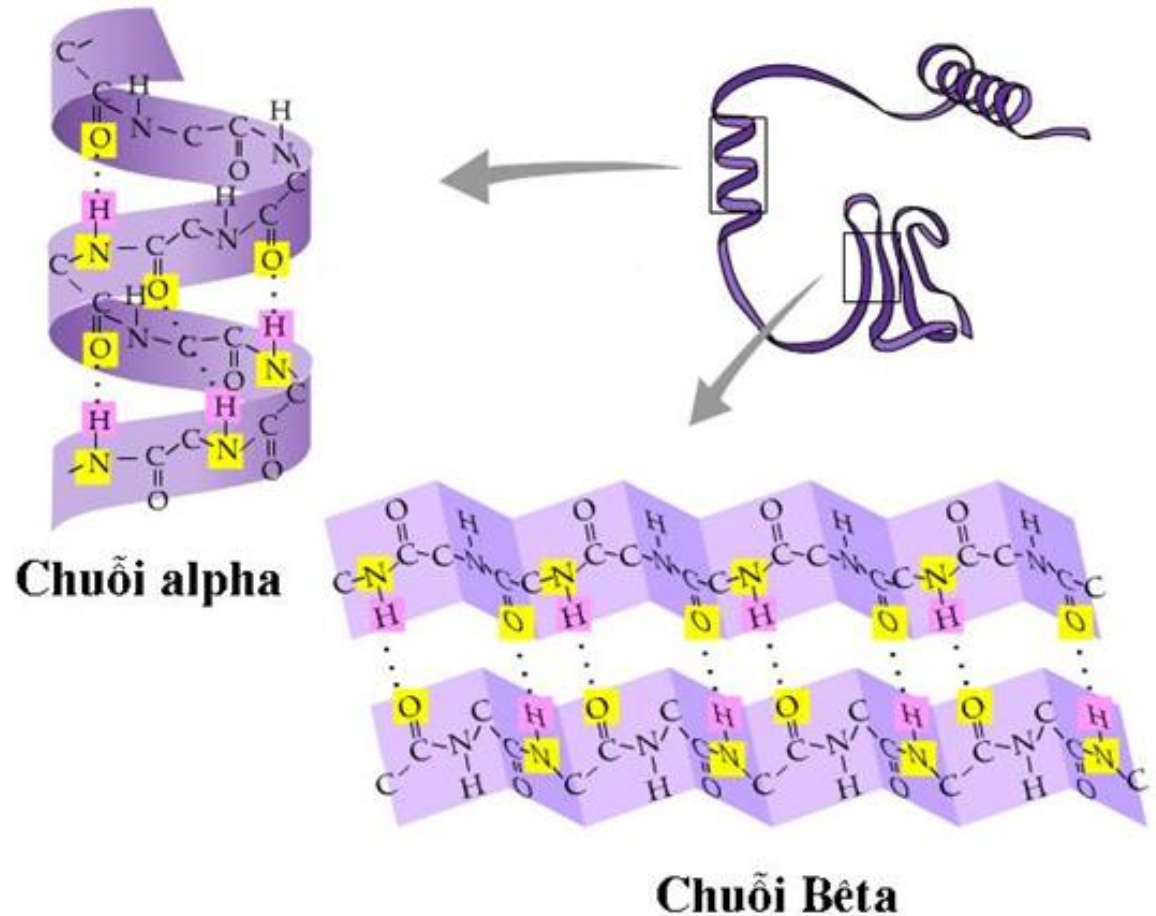


Cấu trúc vật lý của protein

Cấu trúc bậc 2

- Dạng xoắn lò xo (alpha) hay dạng nếp gấp (beta)
- Các nếp gấp và vòng xoắn được cố định bởi các liên kết hidro giữa các amino acids gần nhau

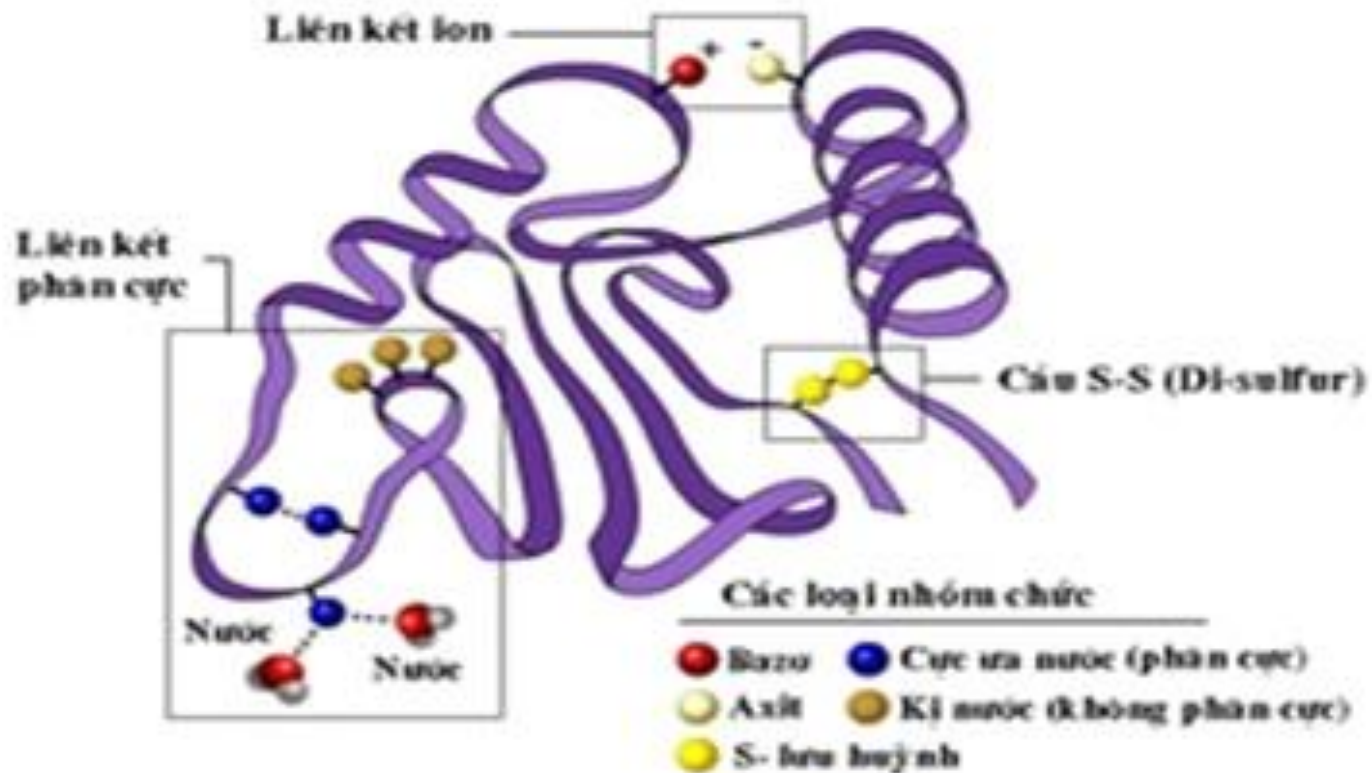
CẤU TRÚC BẬC 2 CỦA PRÔTÊIN



Cấu trúc vật lý của protein

Cấu trúc bậc 3

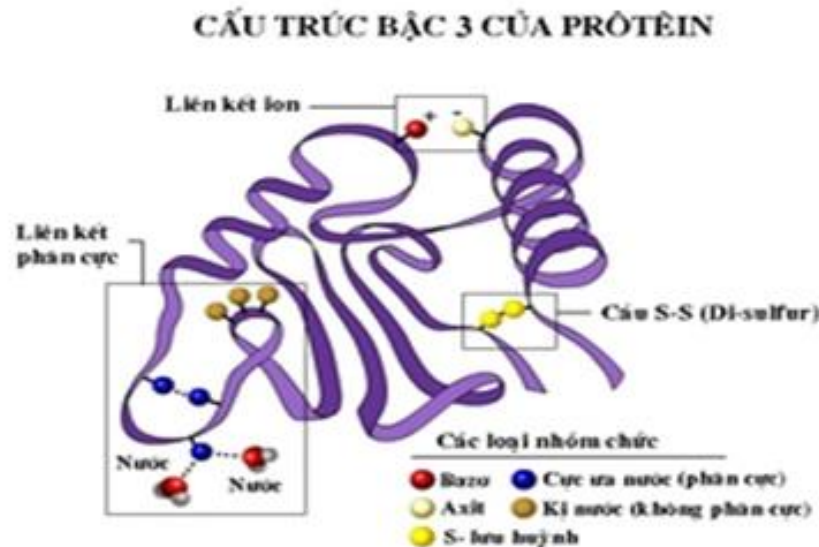
CẤU TRÚC BẬC 3 CỦA PRÔTÊIN



Cấu trúc vật lý của protein

Cấu trúc bậc 3

- Chuỗi xoắn cuộn xếp tạo thành cấu trúc đặc thù trong không gian 3 chiều, tạo nên tính đặc trưng của từng loại protein bằng các liên kết cầu di-sulfur, liên kết ion, ...
- Từ đó tăng tính bền vững của phân tử protein.



LIÊN KẾT TRONG PT PROTEIN

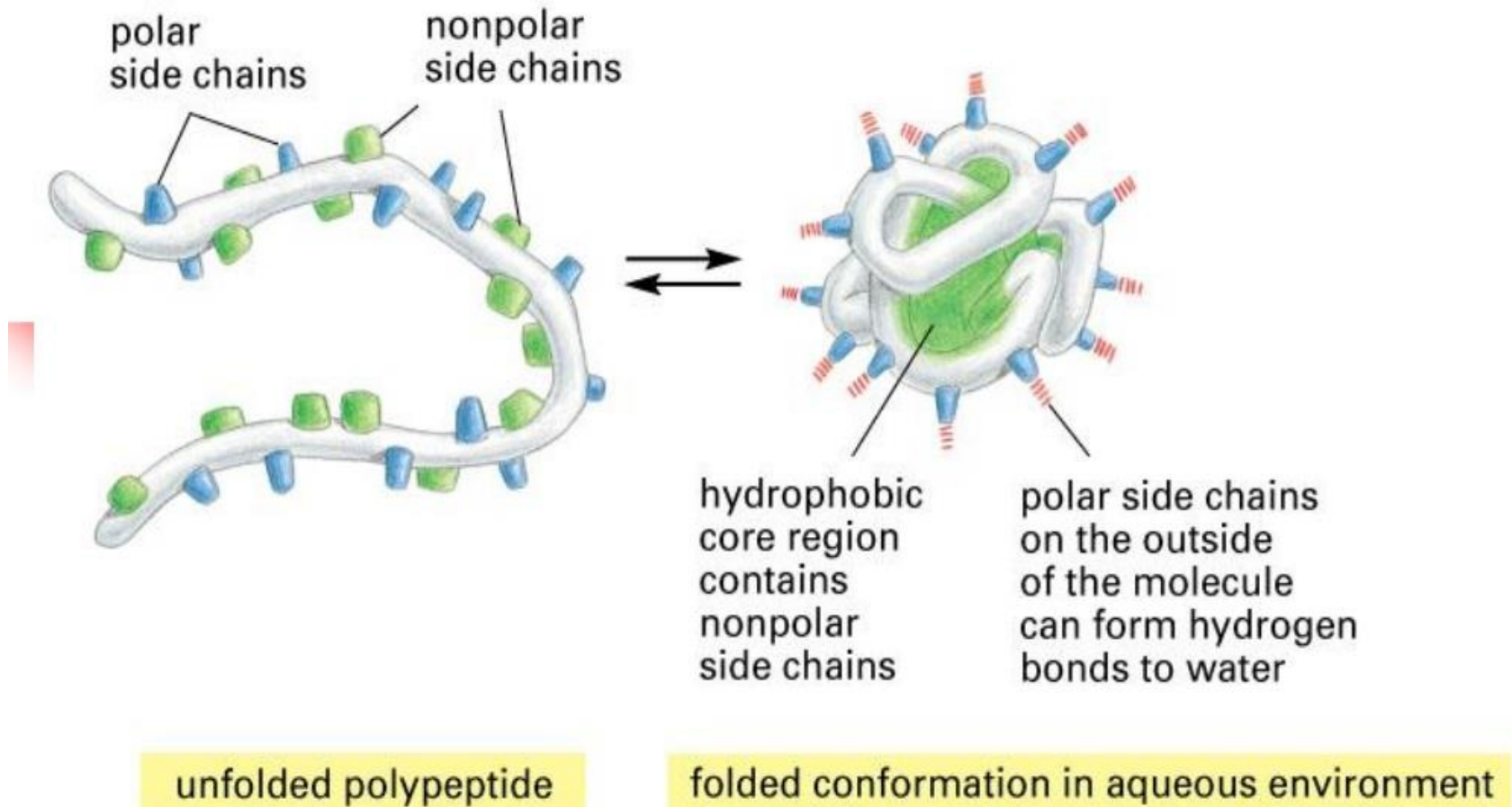
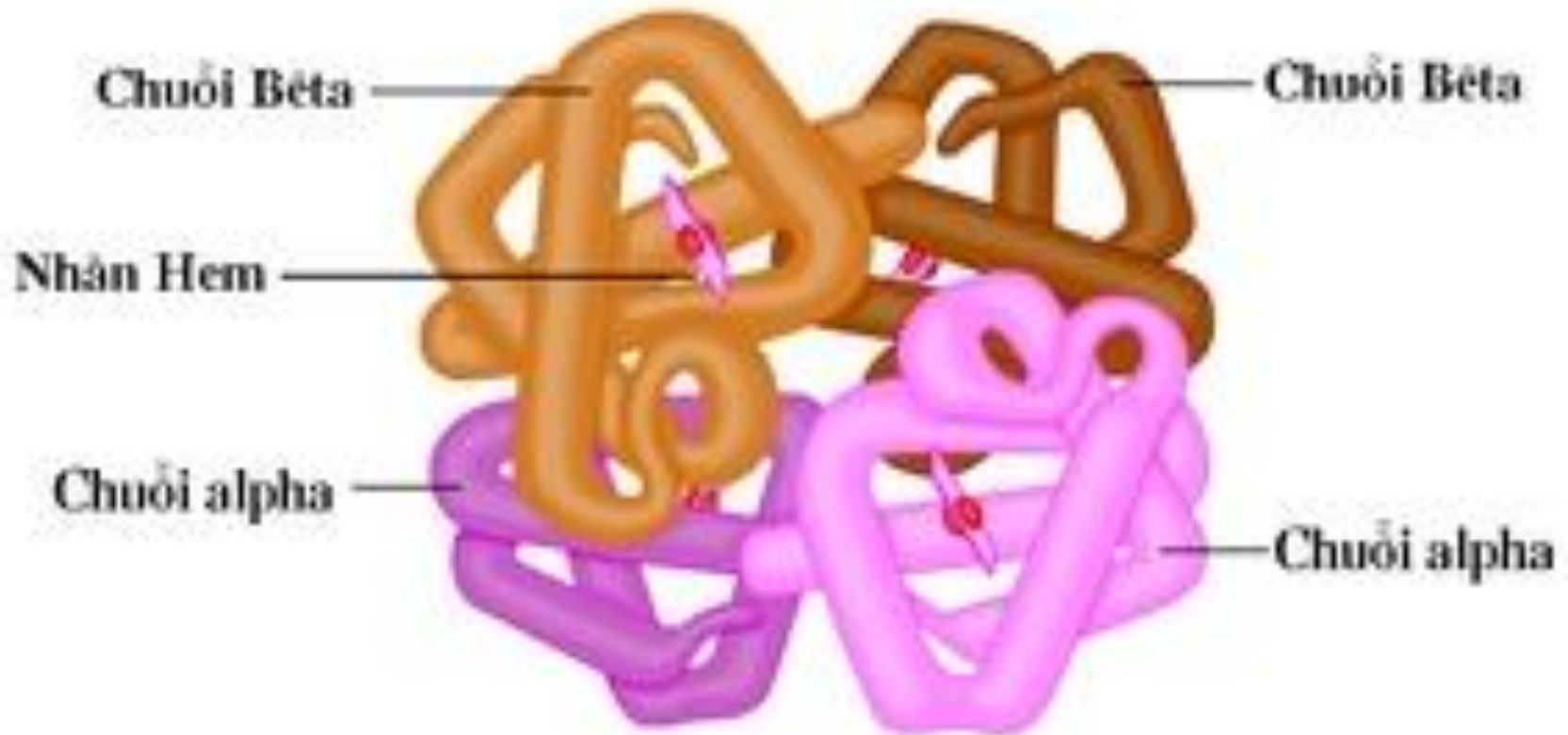


Figure 3–6. Molecular Biology of the Cell, 4th Edition.

Cấu trúc vật lý của protein

Cấu trúc bậc 4

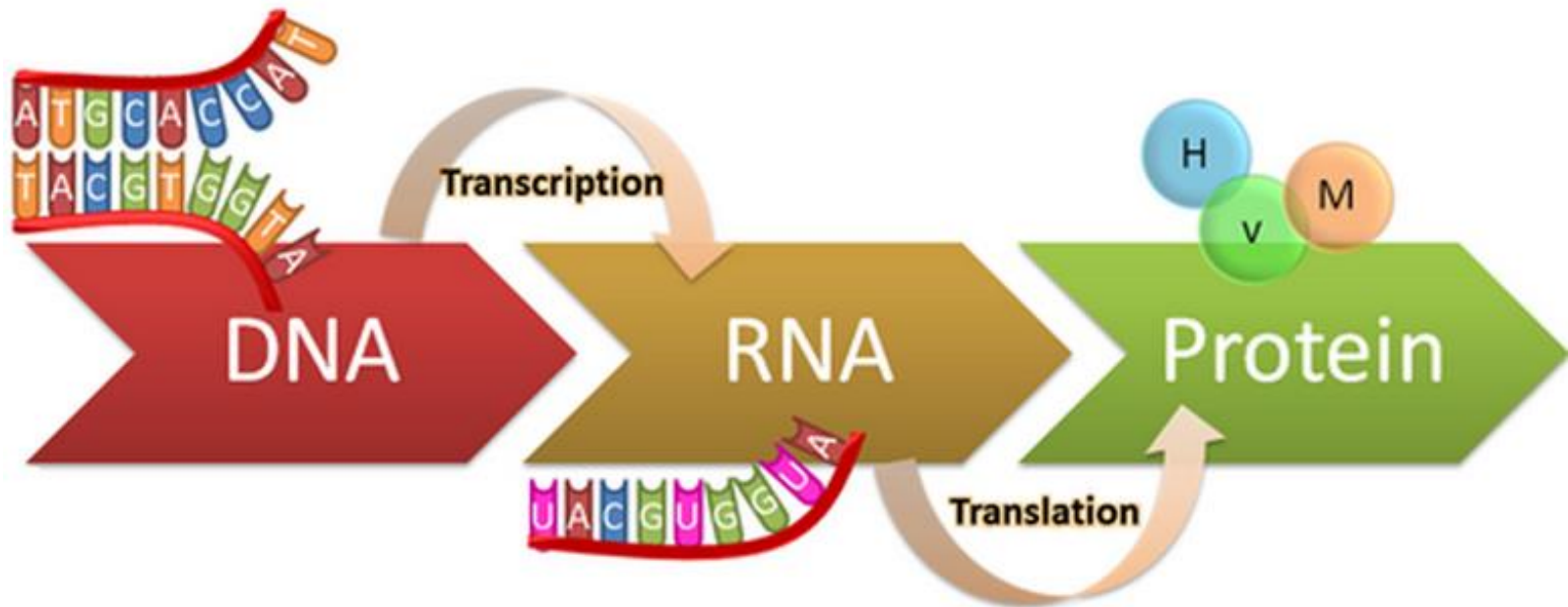
CẤU TẠO HEMÔGLOBIN (BẬC 4)



1.5. Học thuyết trung tâm

Central Dogma

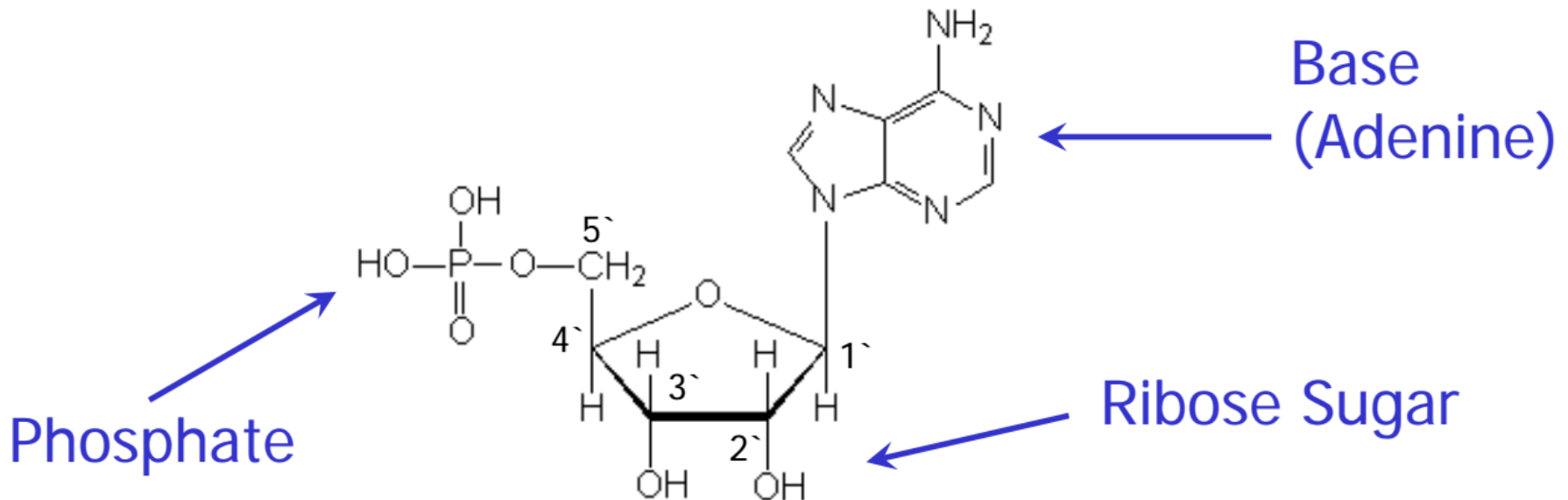
- Protein được tạo ra từ gene
- Gene là một đoạn DNA mã hóa một protein.
- Trong hệ gen của con người, có 30.000 - 35.000 gen.
- Chiều dài trung bình ước tính của một gen: 1000-2000 bp



RNA : Ribonucleic acid



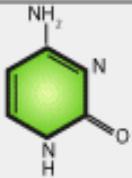
- mRNA: bản sao của gene (trong DNA)
- **Nucleotide** bao gồm ba phần:
 - Đường Ribose (có thêm một nhóm OH ở 2')
 - Phốt phát (liên kết với cacbon 5')
 - Ba zơ (liên kết với carbon 1')



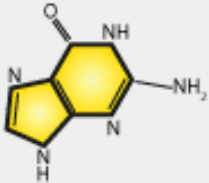
RNA sử dụng bazo U thay vì T

DIFFERENCE BETWEEN DNA AND RNA

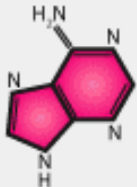
CYTOSINE C



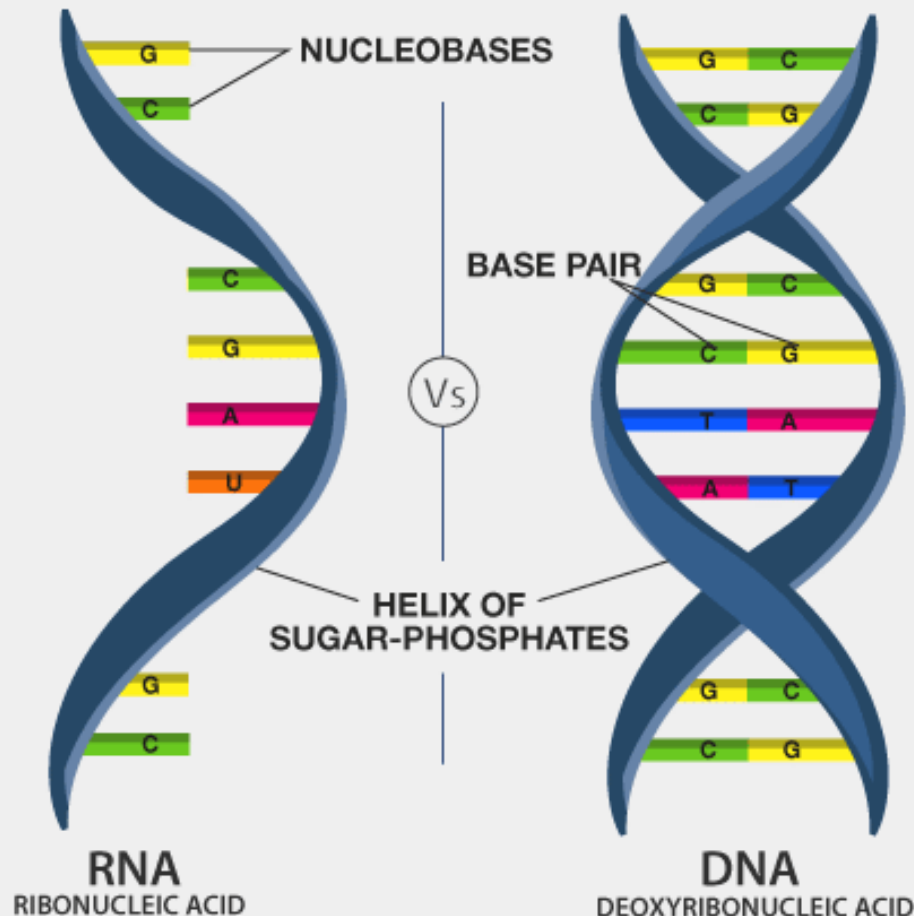
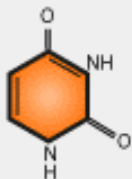
GUANINE G



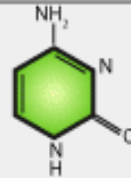
ADENINE A



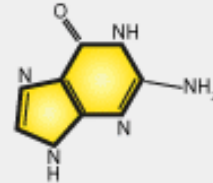
URACIL U



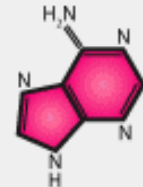
CYTOSINE C



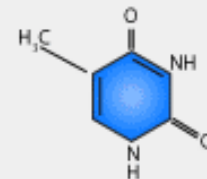
GUANINE G



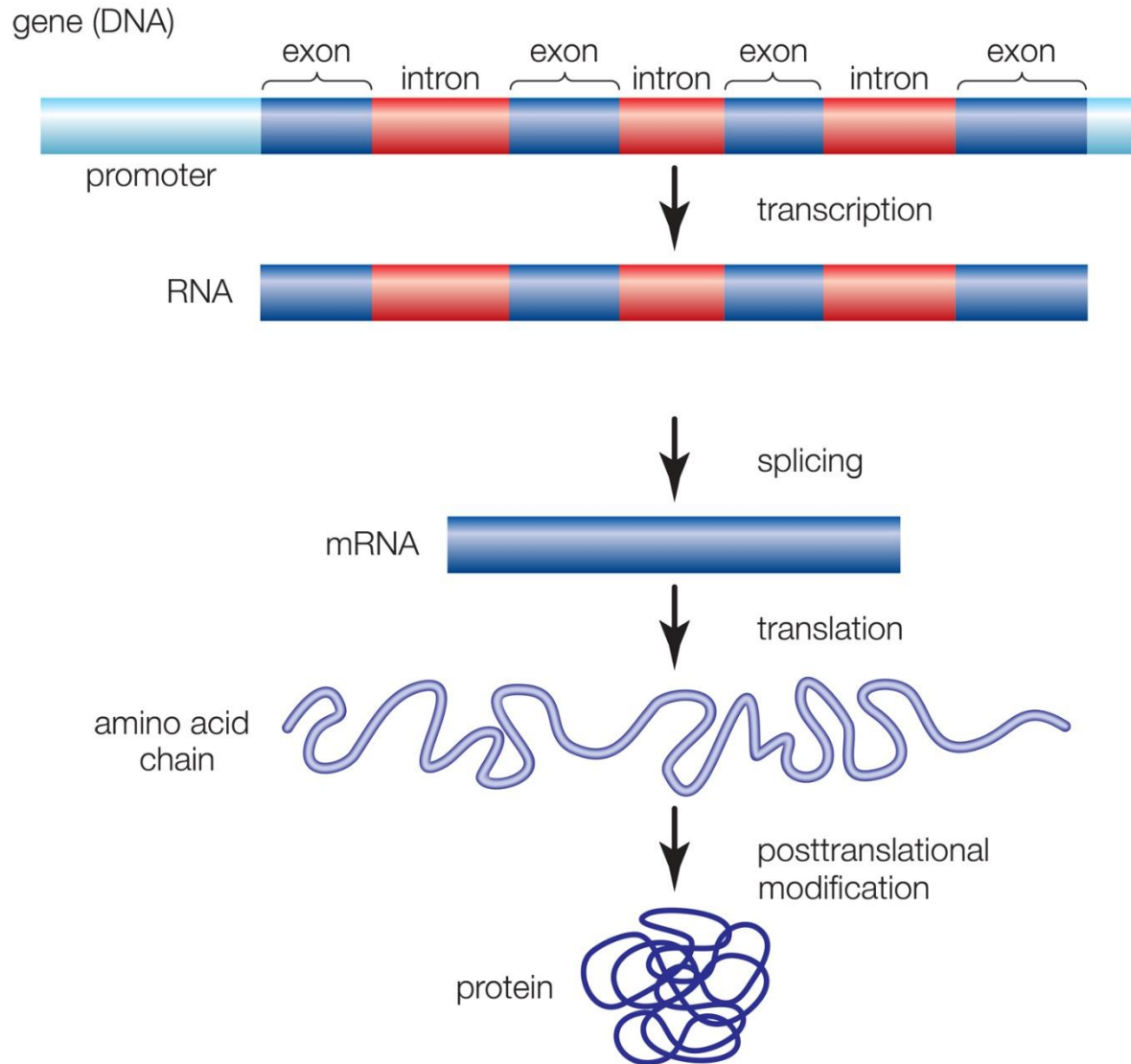
ADENINE A



THYMINE T

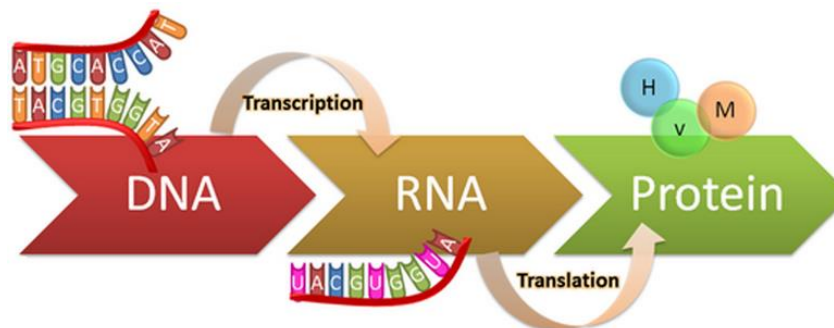


Quá trình hình thành Protein từ gene

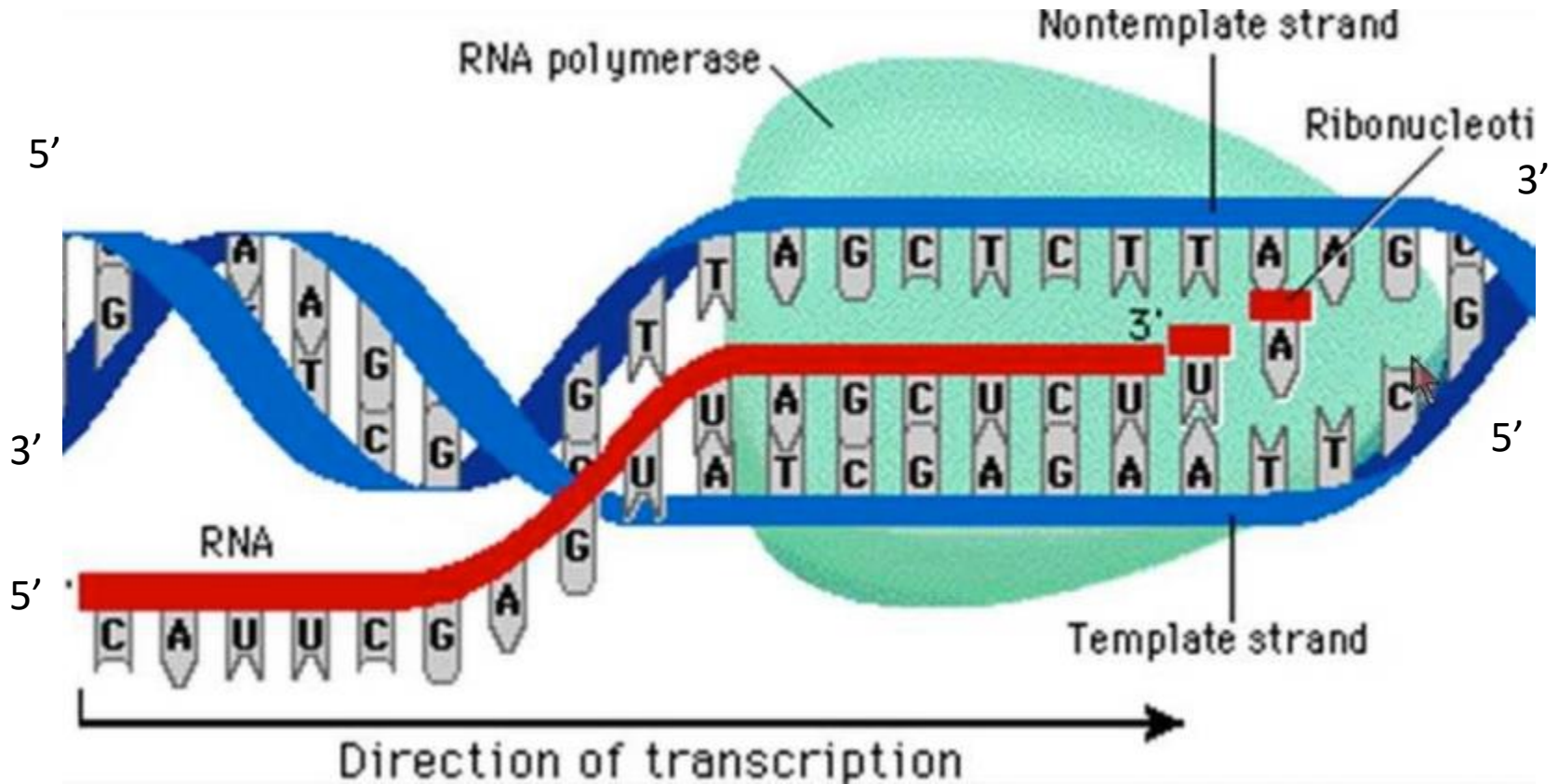


Phiên mã

- Tổng hợp một đoạn pre-mRNA (RNA thông tin, mRNA)
 1. Enzyme RNA polymerase tạm thời phân tách DNA sợi đôi
 2. Bắt đầu phiên mã tại vị trí khởi đầu.
 3. $A \Rightarrow A$, $C \Rightarrow C$, $G \Rightarrow G$ và $T \Rightarrow U$
 4. Khi RNA polymerase đến vị trí kết thúc phiên mã, quá trình phiên mã sẽ dừng lại.

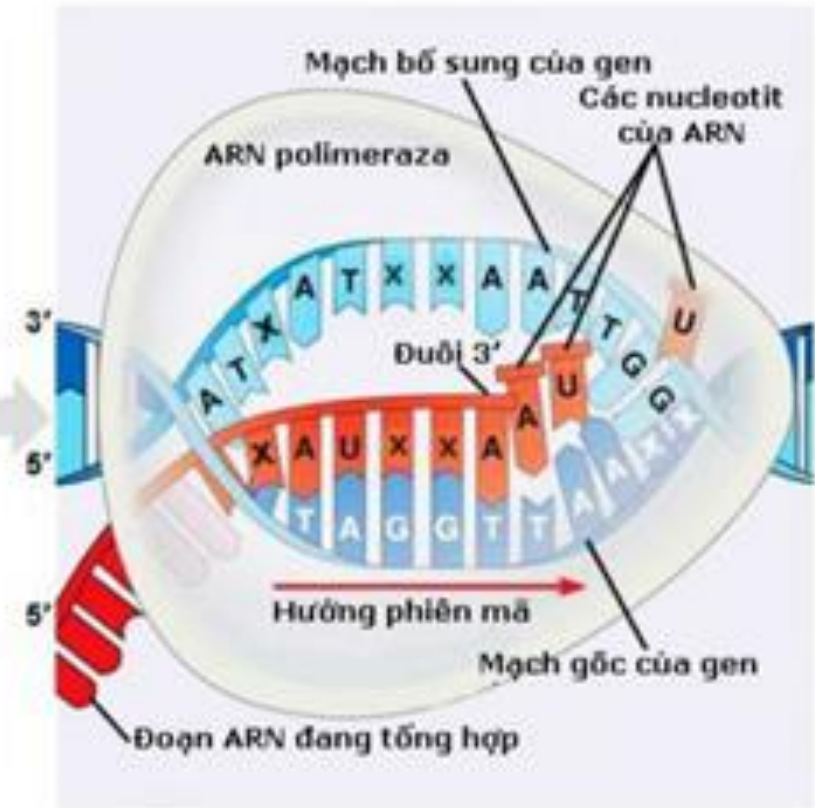
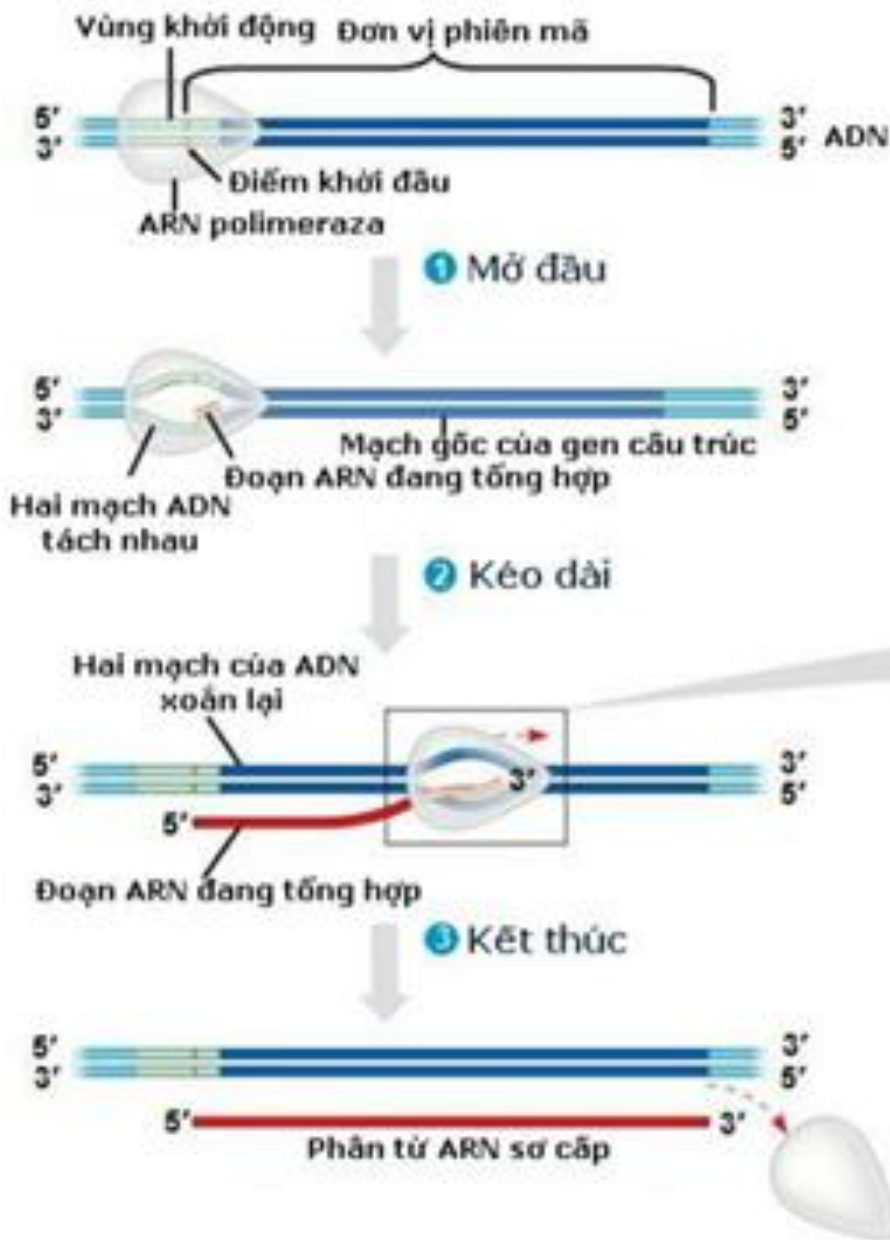


Phiên mã



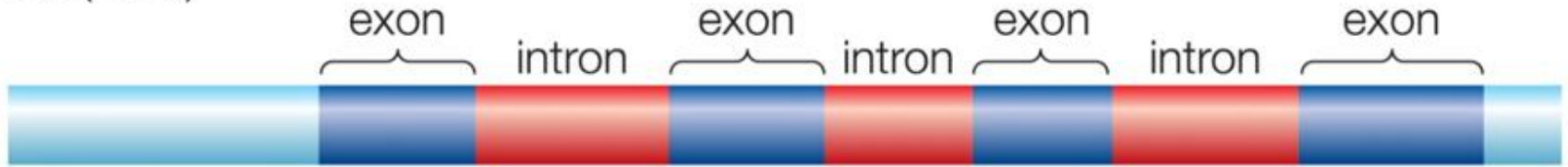
<https://www.sinhhocphantu.org/2018/06/phien-ma.html>

Phiên mã



Phiên mã (Transcription)

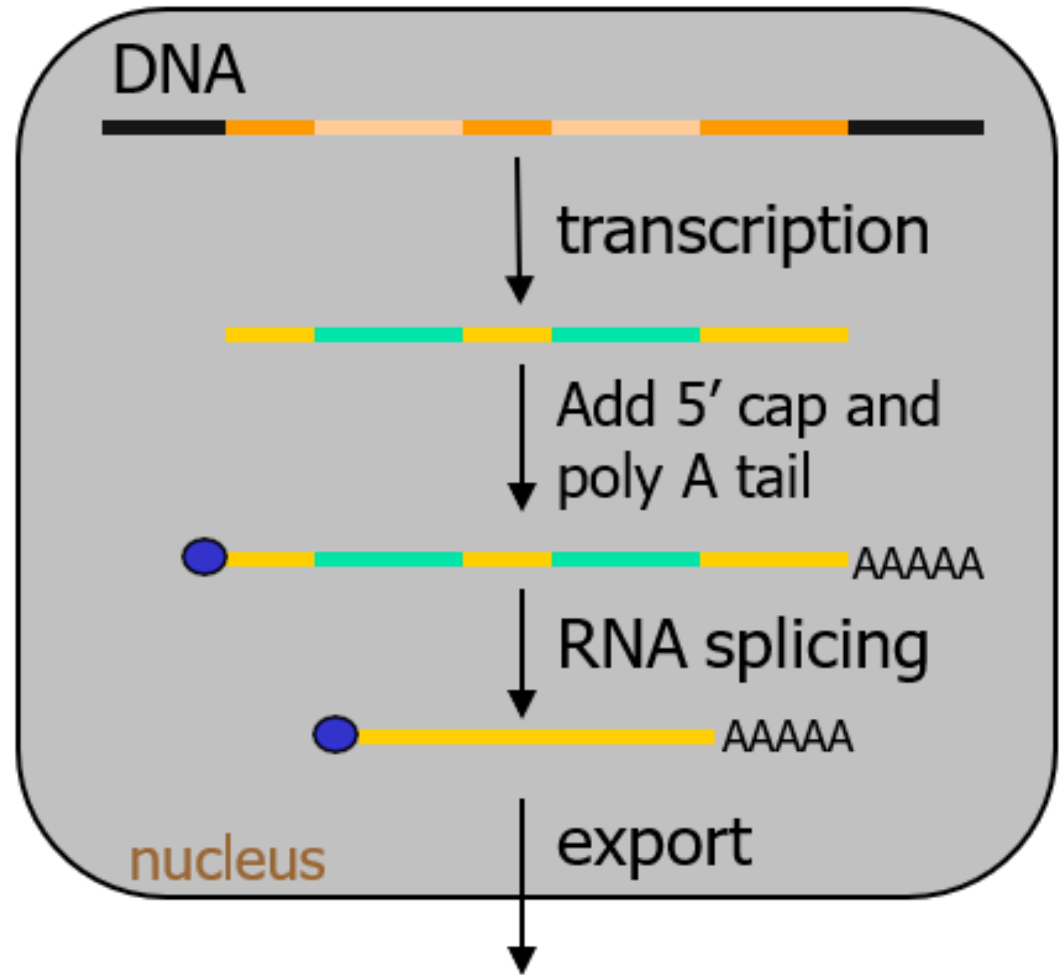
gene (DNA)



- Các gen của sinh vật nhân thực chứa các intron và exon.
- Intron là trình tự sẽ được loại ra khỏi pre-mRNA
- Intron thường thỏa mãn quy tắc GT-AG, tức là intron bắt đầu bằng GT và kết thúc bằng AG.
- Mỗi gen có thể có nhiều intron và mỗi intron có thể có hàng nghìn base.

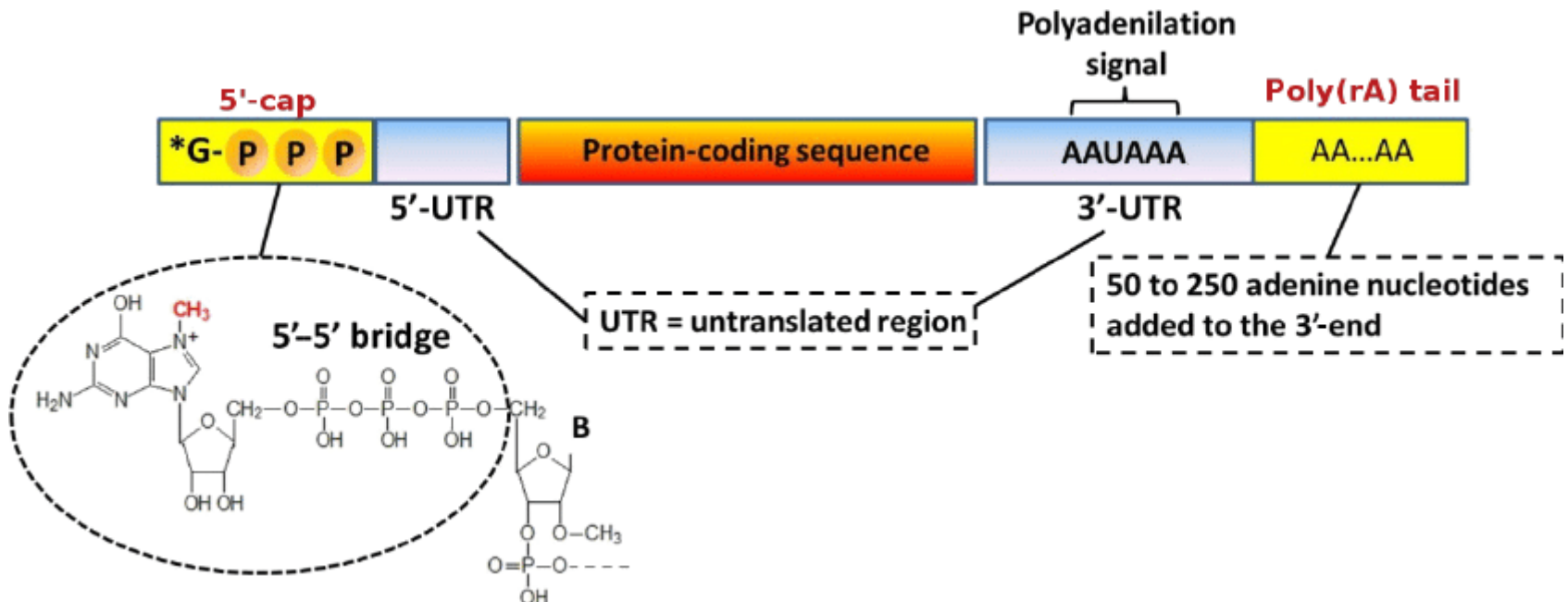
Phiên mã (Transcription)

- Phiên mã tạo ra pre-mRNA chứa cả intron và exon
- Nắp 5' và đuôi poly-A được thêm vào pre-mRNA
- RNA splicing loại bỏ các intron và mRNA được tạo ra.
- mRNA được vận chuyển ra khỏi nhân



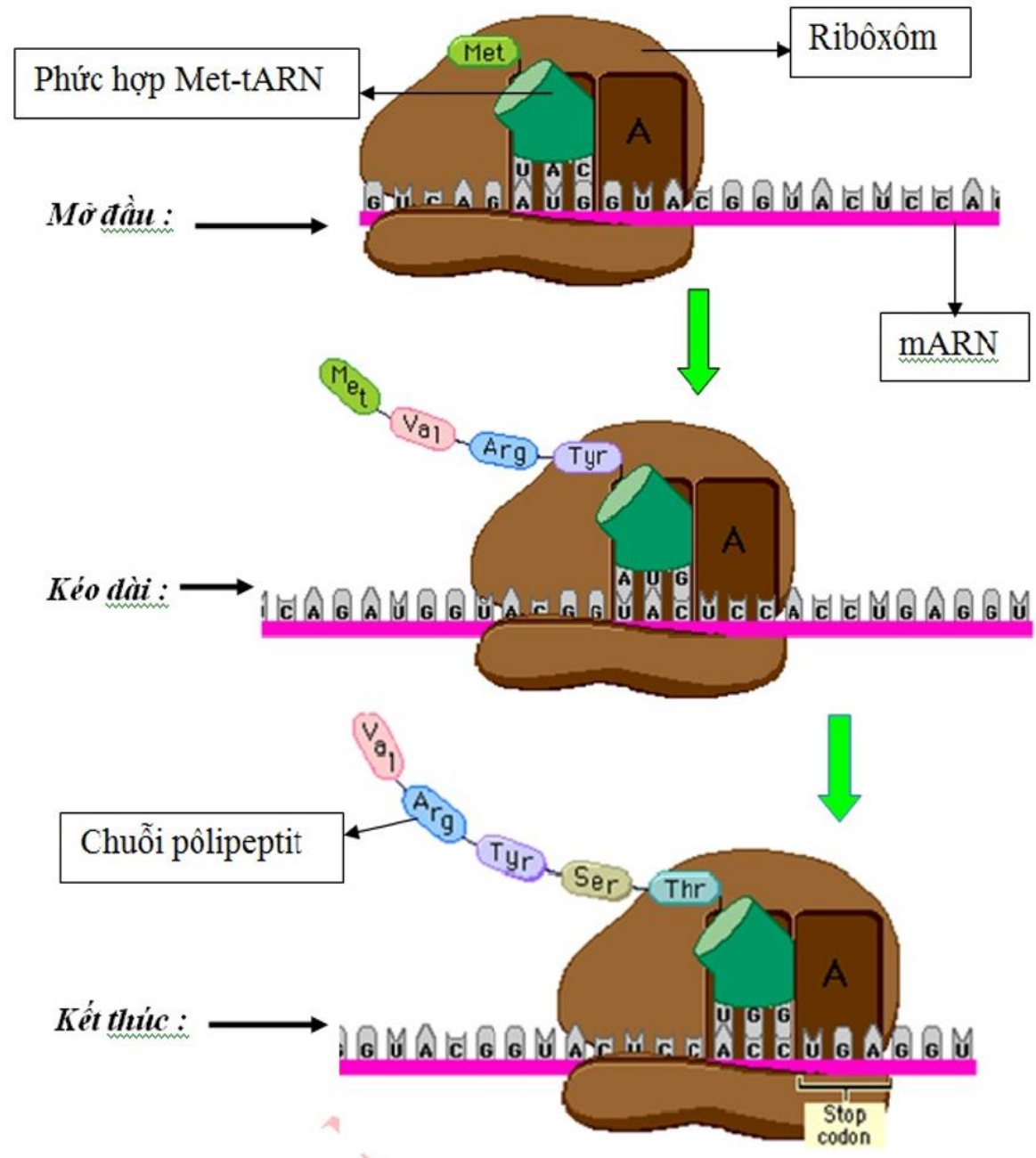
Nắp 5' và đuôi poly-A được thêm vào tiền mRNA

- Hỗ trợ chuyển mRNA từ trong nhân ra tế bào chất.
- Bảo vệ mRNA khỏi tác dụng của các enzyme thủy phân.
- Giúp ribosome gắn được vào đầu 5' của mRNA khi dịch mã.



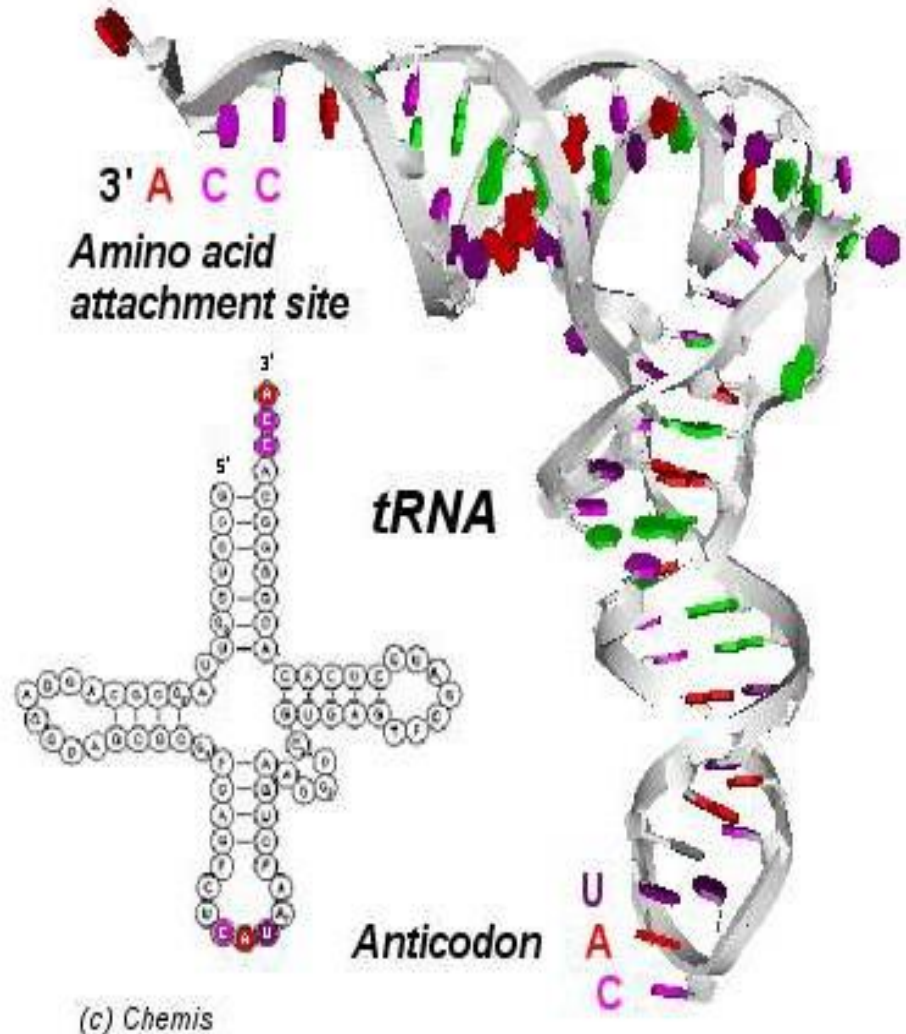
Dịch mã

- Dịch tổng hợp một protein từ mRNA.
- Trên thực tế, mỗi axit amin được mã hóa bởi **trình tự liên tiếp của 3 nucleotide**, được gọi là **codon**.
- Bảng giải mã từ codon thành axit amin được gọi là mã di truyền.



tRNA: RNA vận chuyển

- Có 61 tRNA khác nhau, mỗi tRNA tương ứng với một codon
- Mỗi tRNA gấp lại để tạo thành cấu trúc hình lá cỏ ba lá
- Một bên giữ một phản codon (anticodon)
- Bên còn lại chứa axit amin thích hợp



Dịch mã

- Có $4^3 = 64$ codon khác nhau. Do đó, các codon không tương ứng 1-1 với 20 axit amin.
- Tất cả các sinh vật sử dụng cùng một bảng giải mã!
- Các codon mã hóa cùng một axit amin có xu hướng có cùng nucleotit thứ nhất và thứ hai.
- Nhắc lại rằng các axit amin có thể được phân thành 4 nhóm. Một sự thay đổi bazơ đơn lẻ trong codon thường không đủ để tạo ra một codon mã hóa cho một axit amin trong nhóm khác.

Genetic code


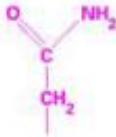
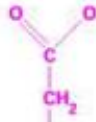
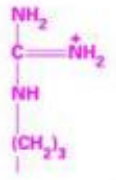

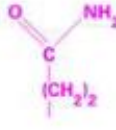
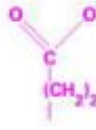

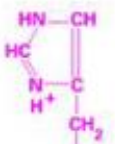
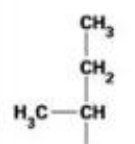
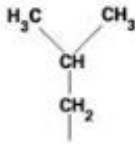

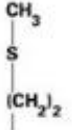

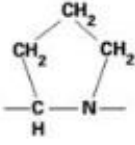

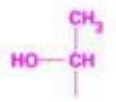
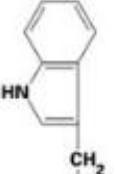

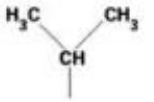
- Start codon: ATG (also code for M)
- Stop codon: TAA, TAG, TGA

Bảng giải mã

	T	C	A	G	
T	TTT Phe [F]	TCT Ser [S]	TAT Tyr [Y]	TGT Cys [C]	T
	TTC Phe [F]	TCC Ser [S]	TAC Tyr [Y]	TGC Cys [C]	C
	TTA Leu [L]	TCA Ser [S]	TAA Ter [end]	TGA Ter [end]	A
	TTG Leu [L]	TCG Ser [S]	TAG Ter [end]	TGG Trp [W]	G
C	CTT Leu [L]	CCT Pro [P]	CAT His [H]	CGT Arg [R]	T
	CTC Leu [L]	CCC Pro [P]	CAC His [H]	CGC Arg [R]	C
	CTA Leu [L]	CCA Pro [P]	CAA Gln [Q]	CGA Arg [R]	A
	CTG Leu [L]	CCG Pro [P]	CAG Gln [Q]	CGG Arg [R]	G
A	ATT Ile [I]	ACT Thr [T]	AAT Asn [N]	AGT Ser [S]	T
	ATC Ile [I]	ACC Thr [T]	AAC Asn [N]	AGC Ser [S]	C
	ATA Ile [I]	ACA Thr [T]	AAA Lys [K]	AGA Arg [R]	A
	ATG Met [M]	ACG Thr [T]	AAG Lys [K]	AGG Arg [R]	G
G	GTT Val [V]	GCT Ala [A]	GAT Asp [D]	GGT Gly [G]	T
	GTC Val [V]	GCC Ala [A]	GAC Asp [D]	GGC Gly [G]	C
	GTA Val [V]	GCA Ala [A]	GAA Glu [E]	GGA Gly [G]	A
	GTG Val [V]	GCG Ala [A]	GAG Glu [E]	GGG Gly [G]	G

Codon usage

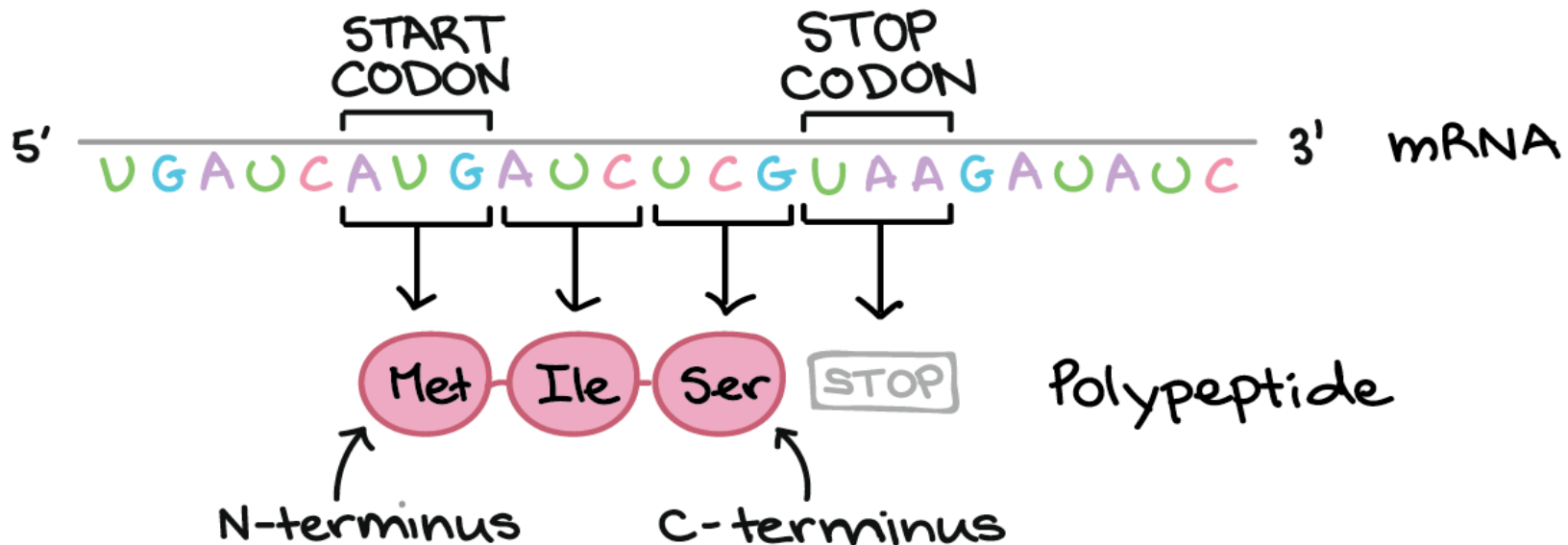
- Tất cả trừ 2 axit amin (W và M) được mã hóa bởi nhiều hơn một codon.
- S được mã hóa bởi 6 codon khác nhau.
- Các sinh vật khác nhau thường thích một codon cụ thể để mã hóa một axit amin cụ thể.

alanine (ala) A	asparagine (asn) N	aspartate (asp) D	arginine (arg) R
 GCU GCC GCA GCG	 AAU AAG	 GAU GAC	 CGU CGC CGA CGG AGA AGG
cysteine (cys) C	glutamine (gln) Q	glutamate (glu) E	glycine (gly) G
 UGU UGC	 CAA CAG	 GAA GAG	 GGU GGC GGA GGG
histidine (his) H	isoleucine (ile) I	leucine (leu) L	lysine (lys) K
 CAU CAC	 AUU AUC AUA	 UUA UUG CUU CUC CUA CUG	 AAA AAG
methionine (met) M	phenylalanine (phe) F	proline (pro) P	serine (ser) S
 AUG	 UUU UUC	 CCU CCC CCA CCG	 AGU AGC UCU UCC UCA UCG
threonine (thr) T	tryptophan (trp) W	tyrosine (tyr) Y	valine (val) V
 ACU ACC ACG ACA	 UGG	 UAU UAC	 GUU GUC GUG GUA
	STOP UGA	STOP UAA UAG	

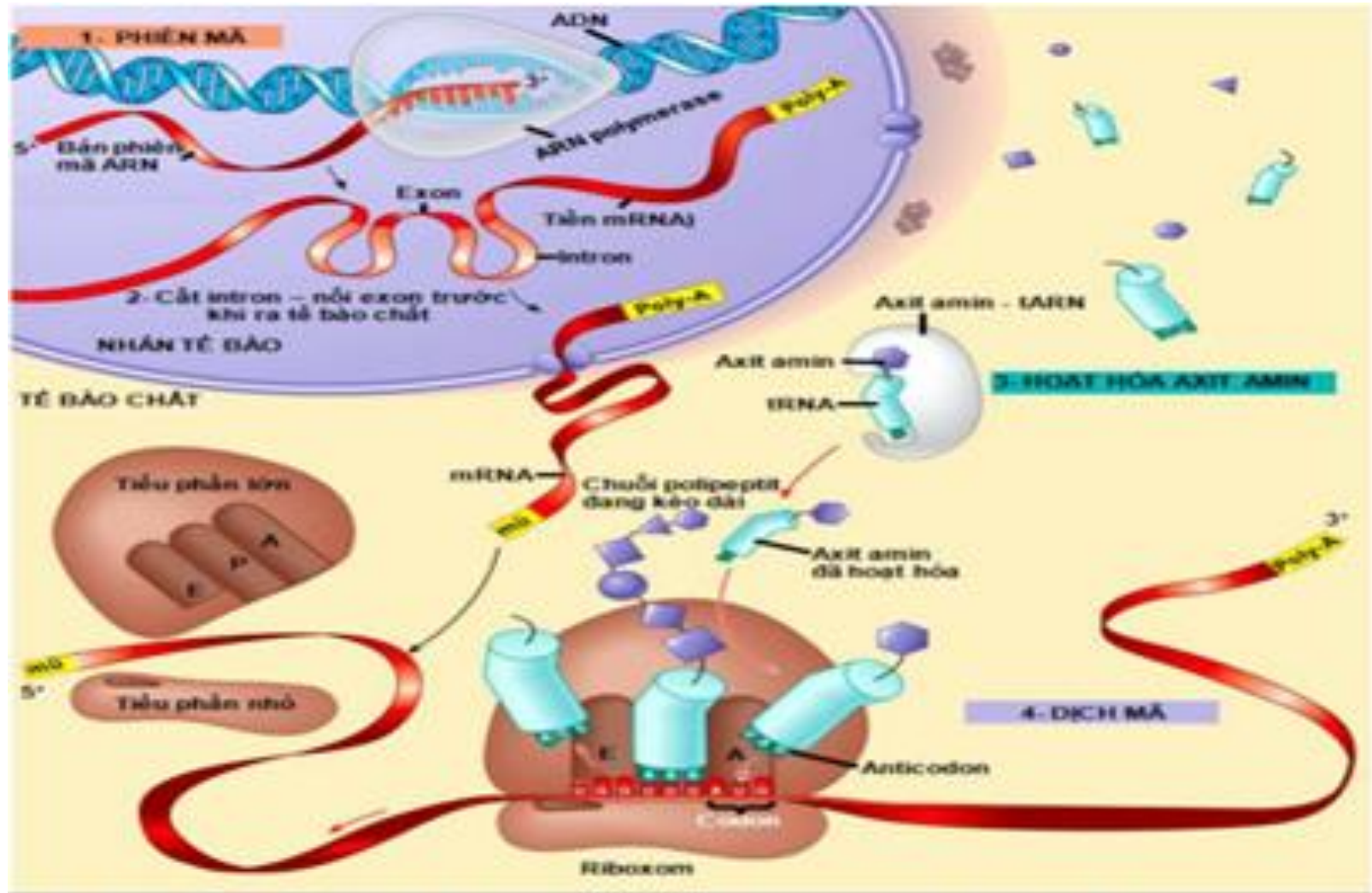
Dịch mã

Vùng mã hóa của gene

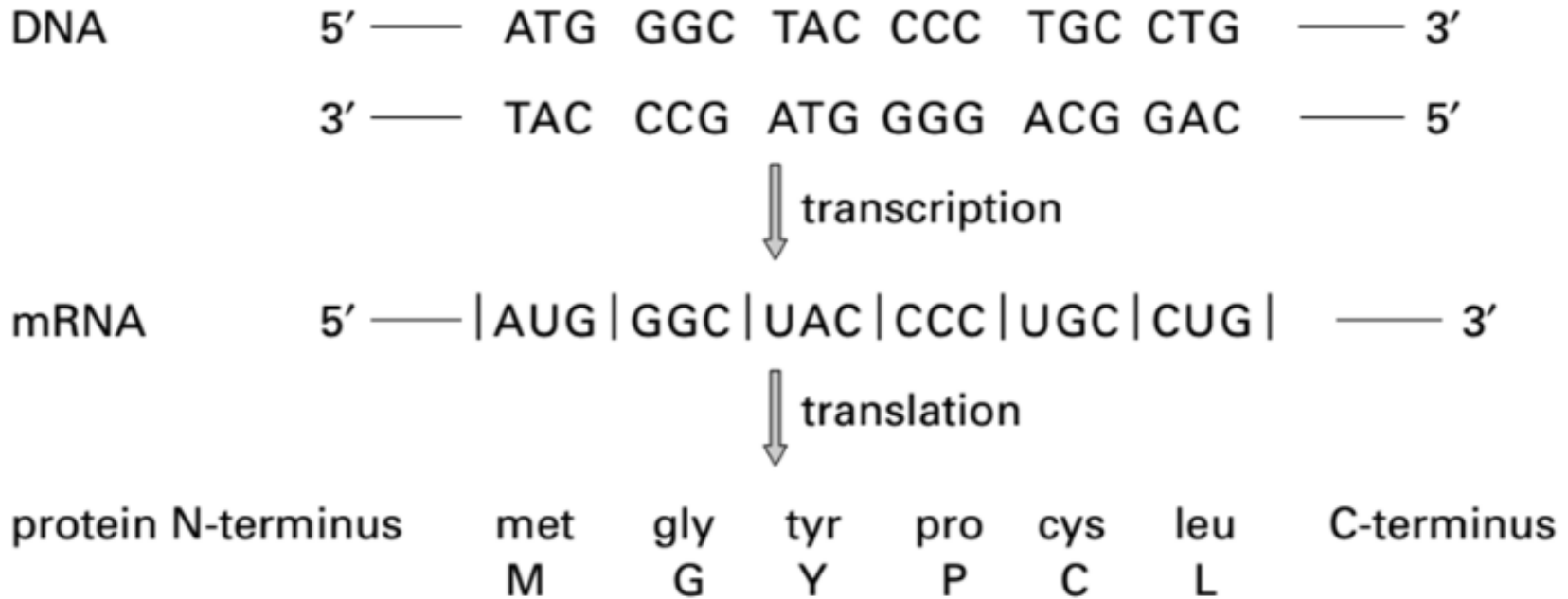
- Vùng mã hóa của gene chứa các codon cho protein, được gọi là khung đọc mở (ORF: open reading frame), với chiều dài là bội số của 3.
- Vùng này phải bắt đầu bằng codon bắt đầu, kết thúc bằng codon kết thúc và phần còn lại của các codon không phải là codon kết thúc.



Tổng hợp protein từ gene trong DNA

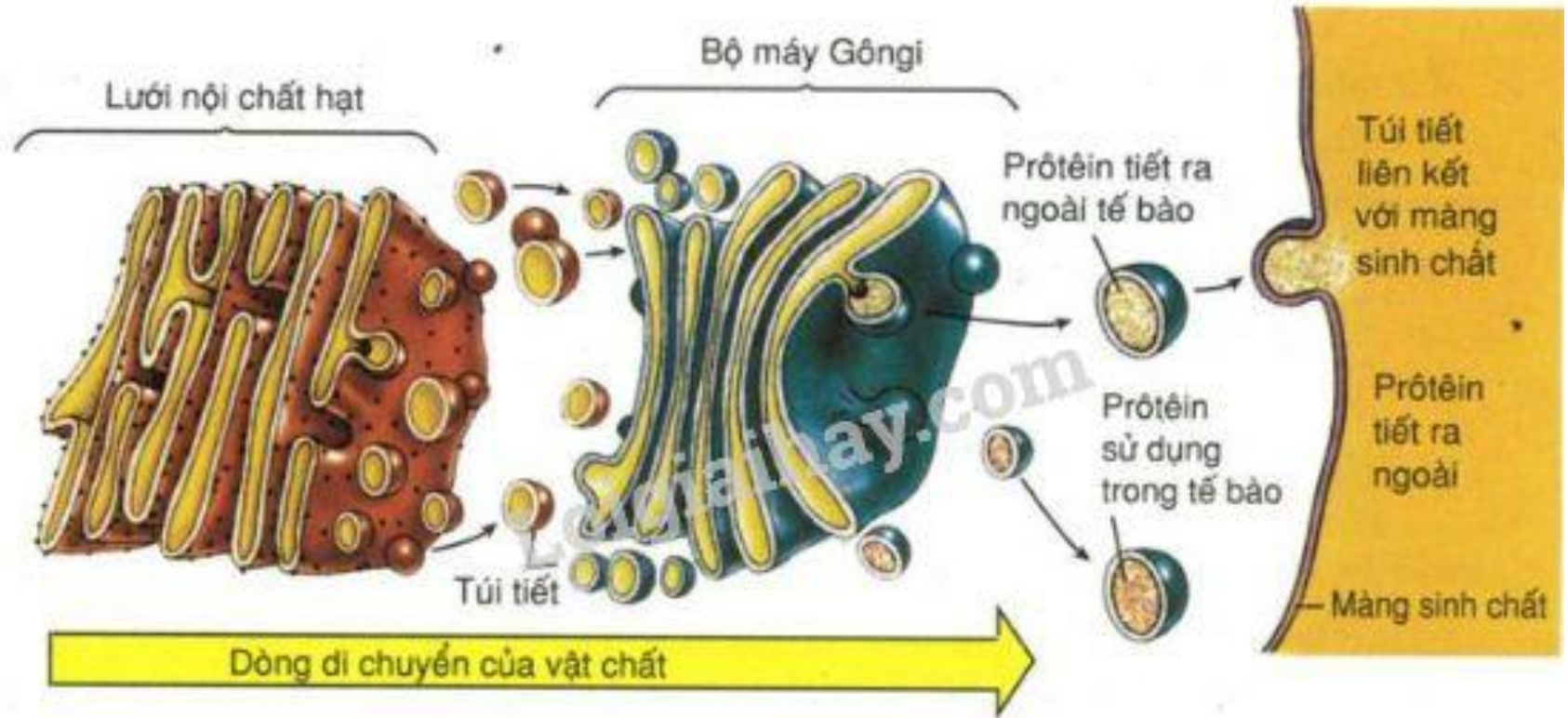


Tổng hợp protein



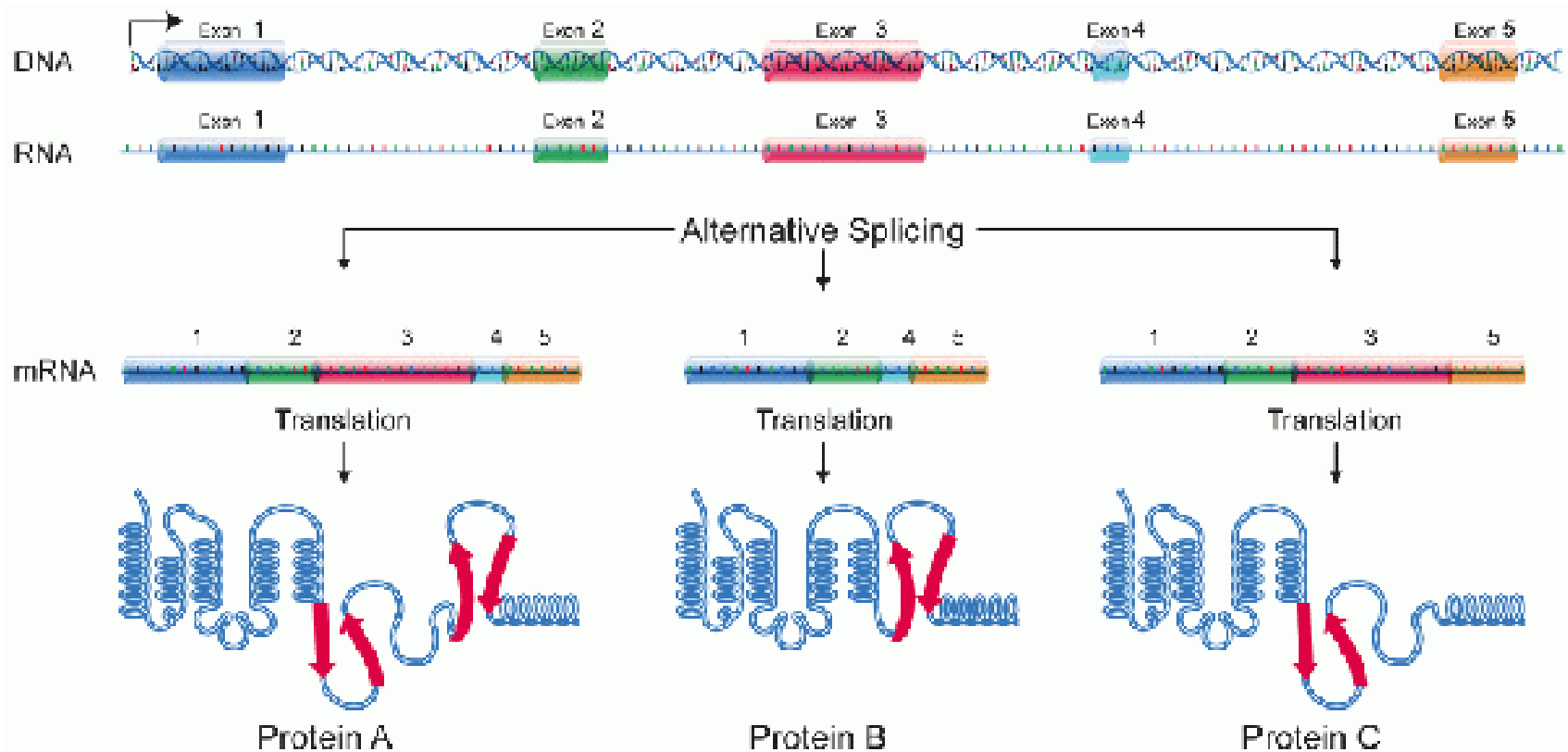
Chiều 3' - 5' trên mạch DNA được phiên mã thành phân tử mRNA và được dịch mã thành protein. Chú ý, mRNA được tổng hợp theo chiều 5' - 3' và protein được tổng hợp theo chiều từ đầu N.

Tổng hợp protein trong tế bào



Hình 8.2. Cấu trúc và chức năng của bộ máy Gôngi

Một gene có thể tạo thành nhiều protein



Có nhiều trường hợp trong tế bào của sinh vật nhân thực, cùng 1 gen được phiên mã tạo thành ARN nhưng lại tổng hợp ra nhiều loại protein khác nhau vì do trong quá trình cắt intron, có sự sắp xếp lại của các exon theo các cách khác nhau

Bài tập: Tìm kiếm và phân tích trình tự nucleotide và protein trên NCBI

- a. Hãy truy cập trang NCBI, tìm kiếm các trình tự DNA, RNA liên quan đến một loại gene như insulin, hemoglobin, collagen, các gen mã hóa enzym trong quá trình tiêu hóa, các gen liên quan đến bệnh:
 - Đột biến gen KRAS liên quan đến 95% ung thư tụy và 45% ung thư đại trực tràng
 - Đột biến gen EGFR trong bệnh ung thư phổi.
 - Đột biến gen BRAF trong bệnh ung thư đại trực tràng hoặc ung thư tuyến giáp
- b. Download dữ liệu tìm được ở các định dạng cơ bản như FASTA, GENBANK
- c. Viết báo cáo mô tả các thông tin liên quan đến trình tự tìm được (tối thiểu 1 gene).
- Tài liệu tham khảo: <https://www.youtube.com/watch?v=7lOvjC6eO4>