Lecture 3
Protein and Nucleic Acids

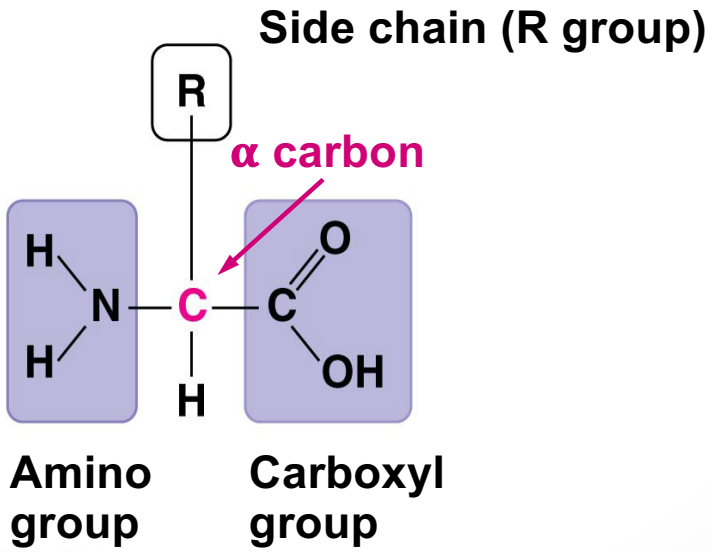
Yang Yang, PhD

Protein

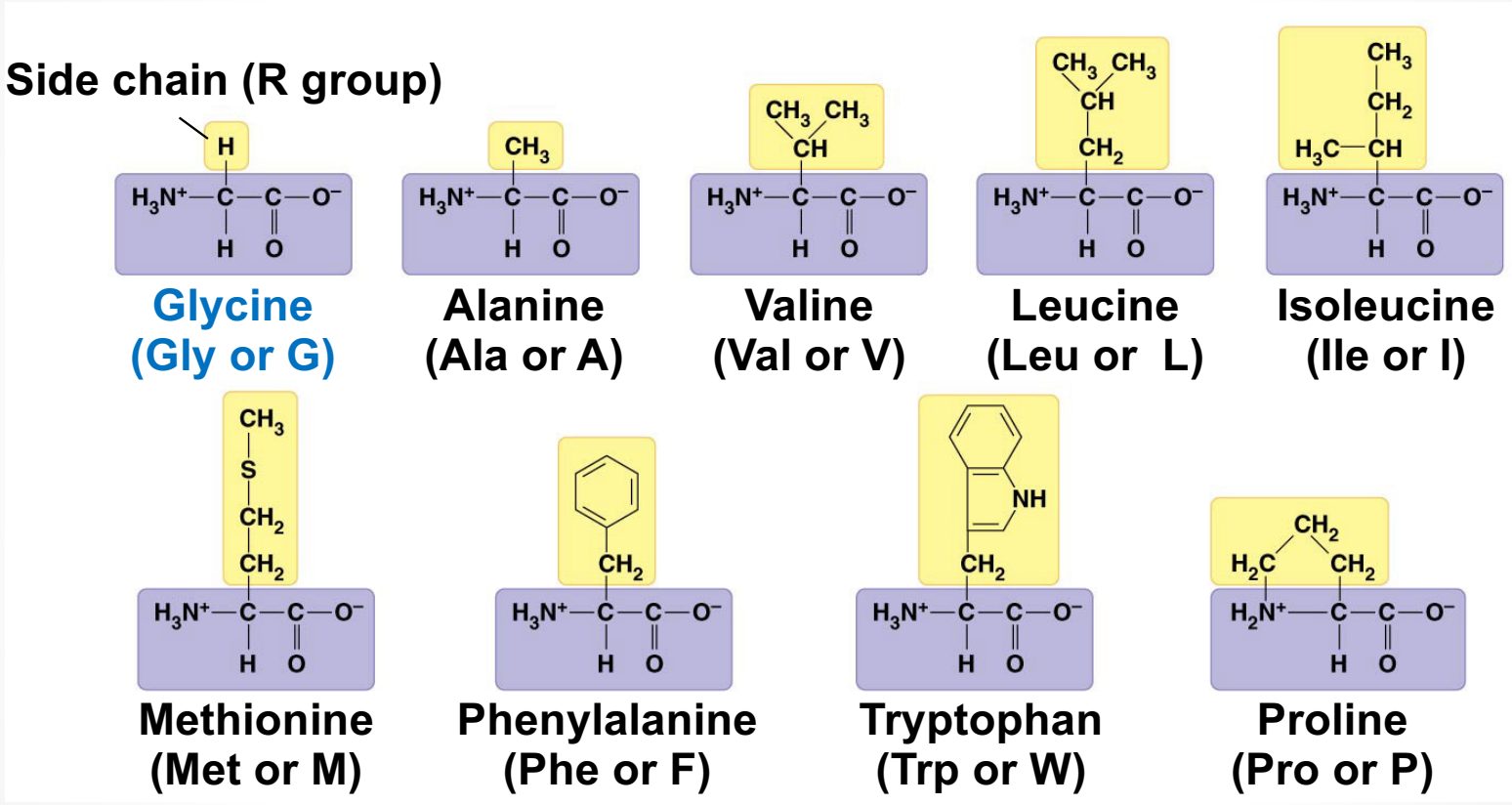
- **Amino acids** 氨基酸
- Reversible: **Dehydration reaction**脱水合成 **<-> hydrolysis**水解
  - To form or break down **peptide bonds** 肽键
- **Protein structure**:
  - Primary, Secondary, Tertiary, Quaternary 一~四级结构
- Side chains and protein structure: determine **protein function**

Amino Acids

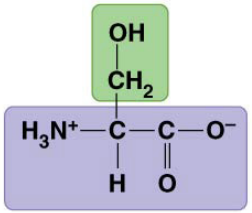
- Amino acids are organic molecules with **amino 氨基 and carboxyl groups羧基**
- Amino acids differ in their properties due to differing side chains (R) 侧链R基团
- All known living organisms use **L-form amino acids 左旋氨基酸**for protein synthesis



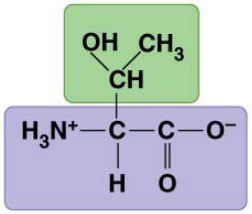
Nonpolar side chains: hydrophobic 疏水



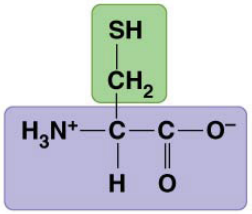
Polar side chains: hydrophilic 亲水



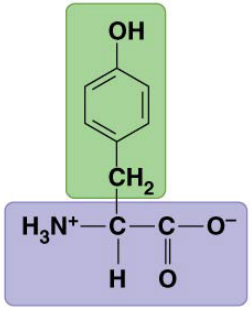
Serine
(Ser or S)



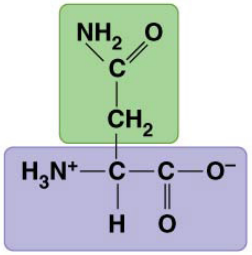
Threonine
(Thr or T)



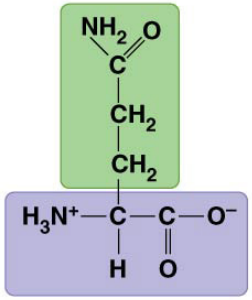
Cysteine
(Cys or C)



Tyrosine
(Tyr or Y)

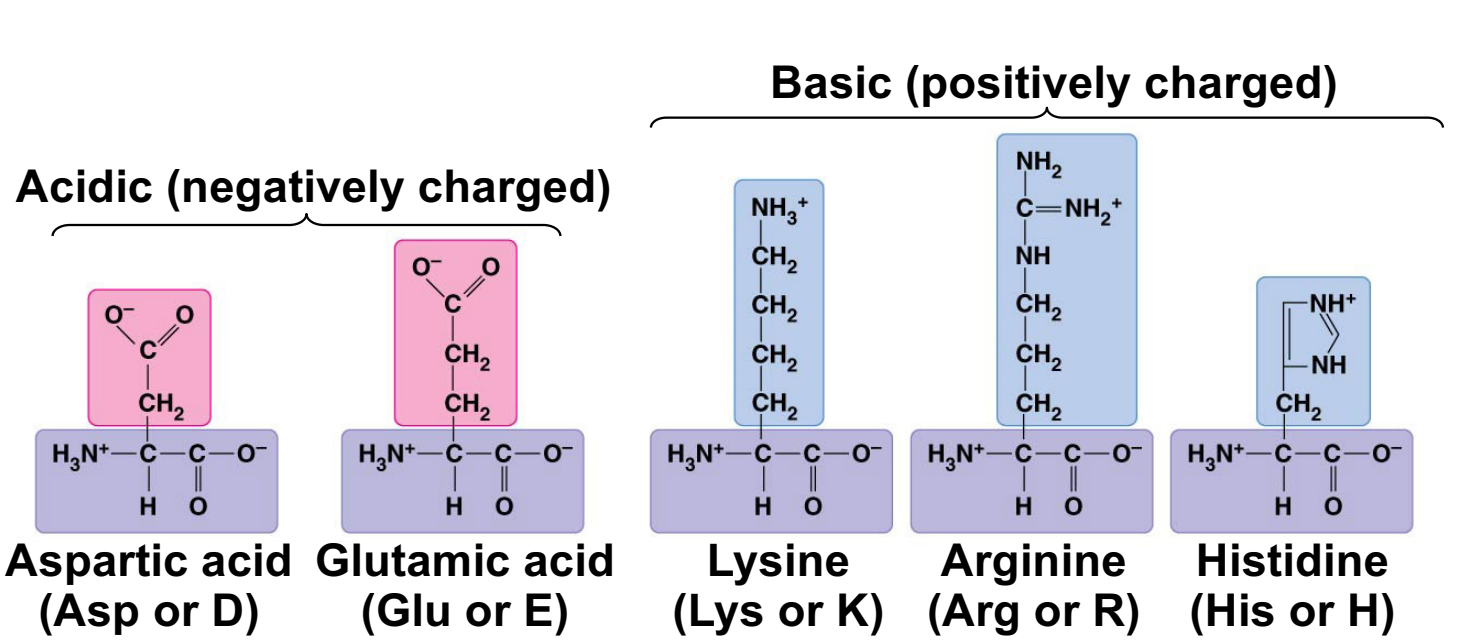


Asparagine
(Asn or N)

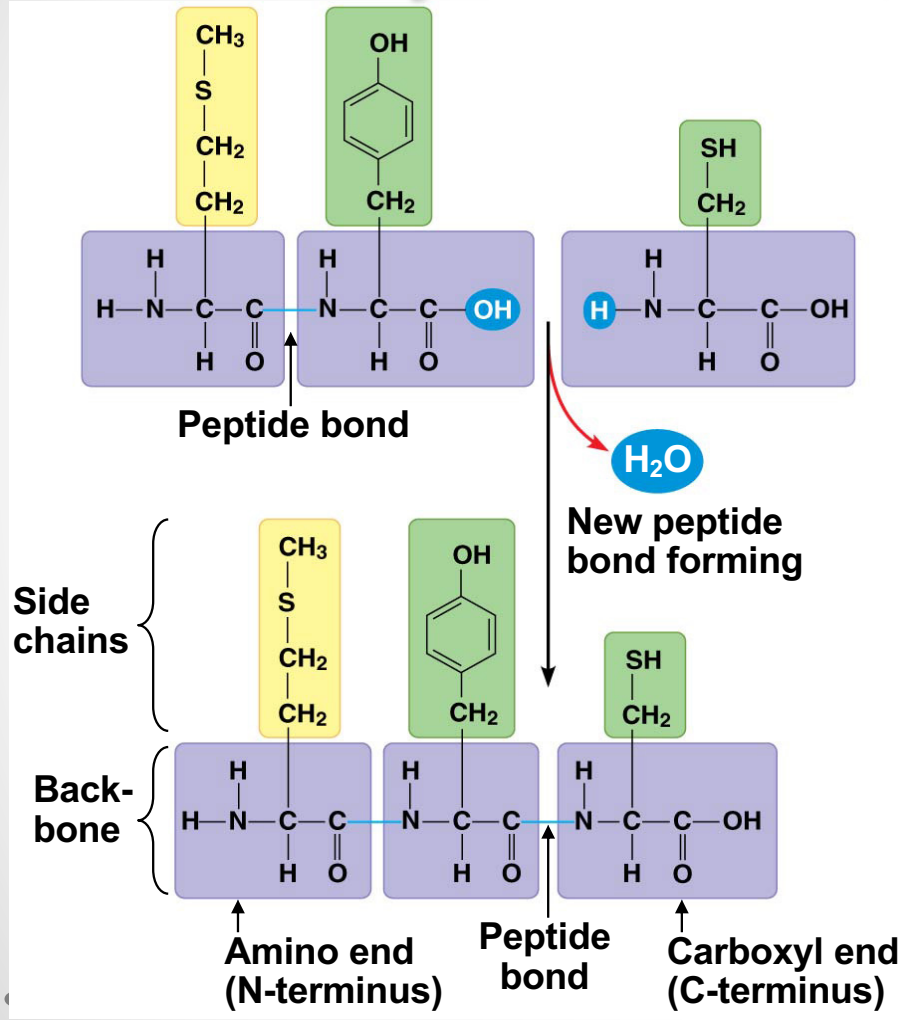


Glutamine
(Gln or Q)

Electrically charged side chains 带电



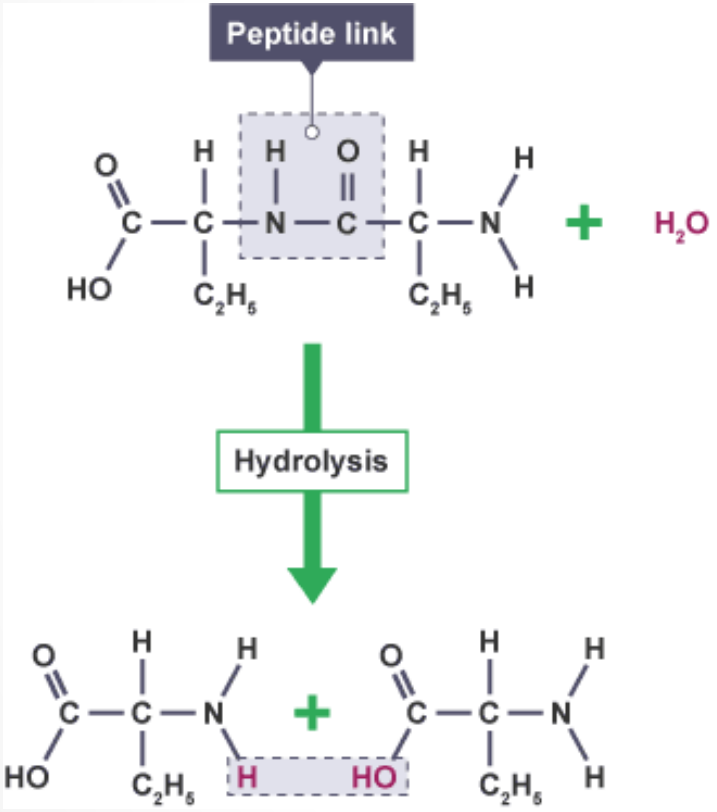
Dehydration (Polymerization)



- Dehydration 脱水（合成）
- **Polarity** of elongation 合成的方向性: **N -> C**

  The **peptide bonds**（肽键）are formed one at a time, starting with the amino acid at the amino end (N-terminus).
- A polypeptide（多肽）has a single amino end (N-terminus) and a single carboxyl end (C-terminus)

Hydrolysis (breakdown)



Hydrolysis 水解:

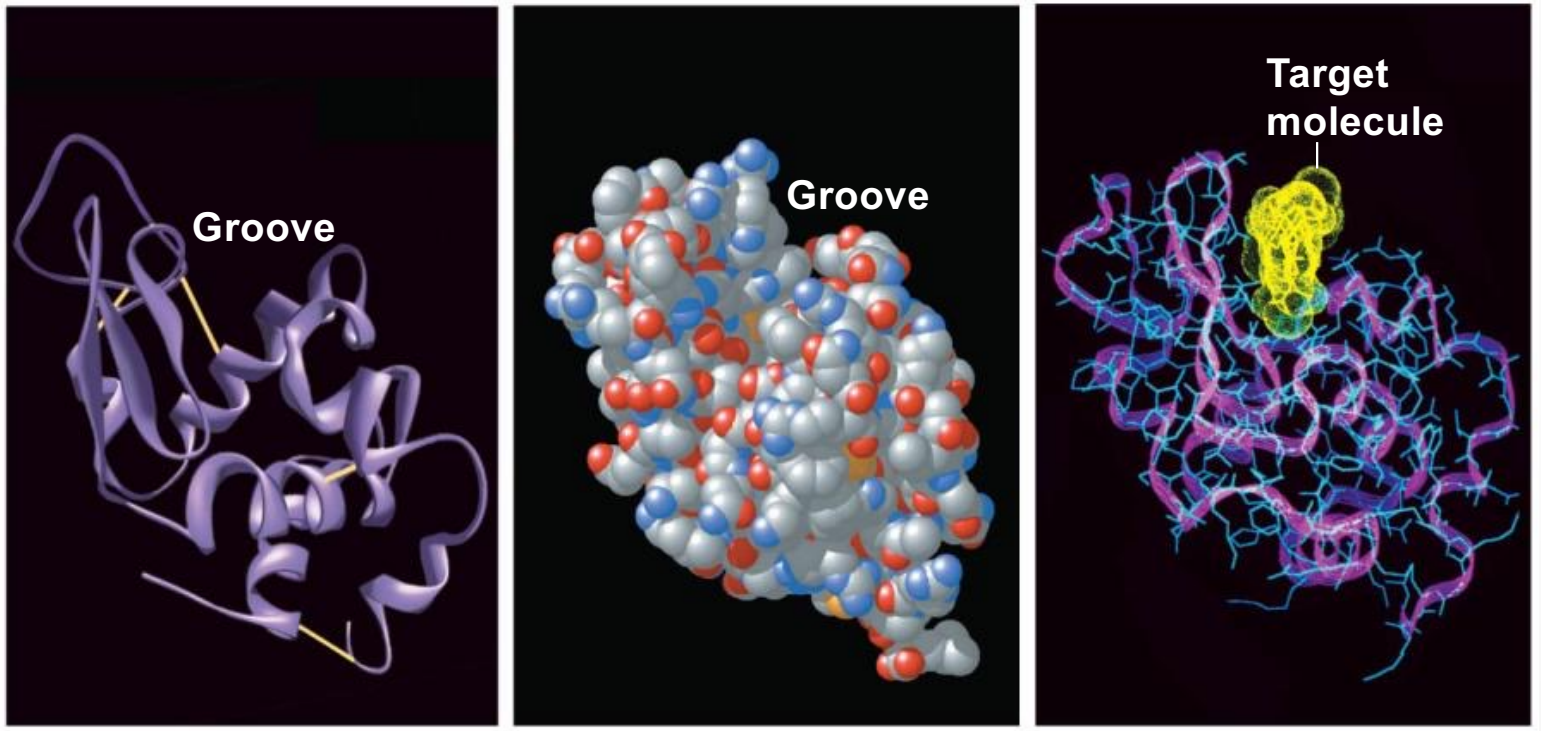
- Breakdown of protein or peptide into smaller peptides and free amino acids
- Directional: depending on enzyme (Proteases蛋白酶) properties
  - Aminopeptidases氨肽酶: starting from N-terminal
  - Carboxypeptidases羧肽酶: starting from C-terminal
  - Endopeptidases: specific sites

Protein structure

- **Primary structure:**
  - **Linear chain** of amino acids
- **Secondary structure:**
  - Regions stabilized by hydrogen bonds between atoms of the **polypeptide backbone**
- **Tertiary structure:**
  - Three-dimensional shape stabilized by **interactions between side chains**
- **Quaternary structure**
  - Association of two or more polypeptides (some proteins only)

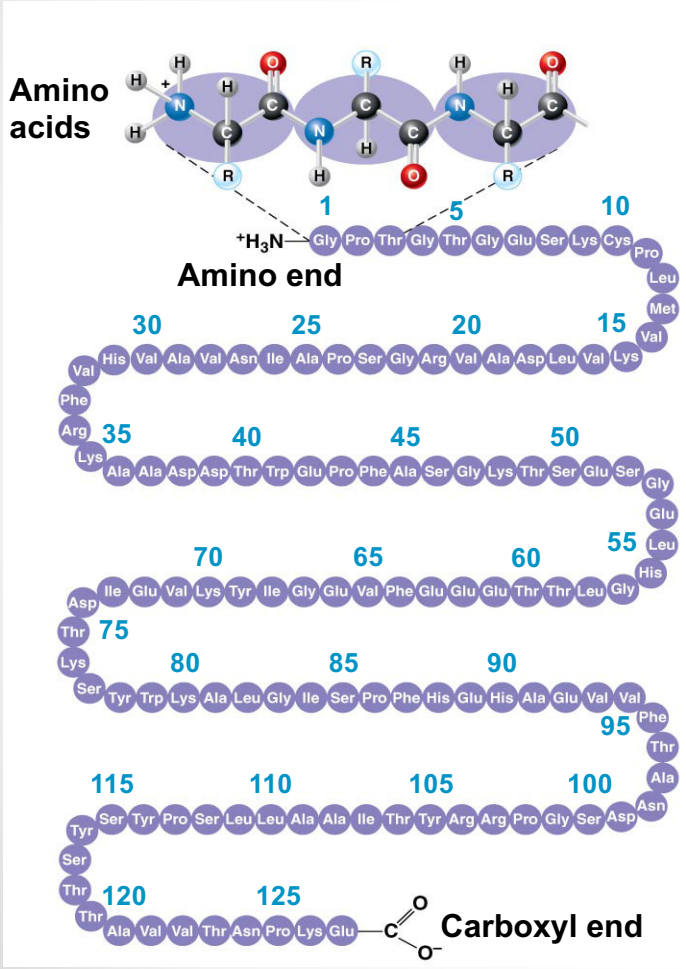
Protein structure

Structure of **Lysozyme** 溶菌酶

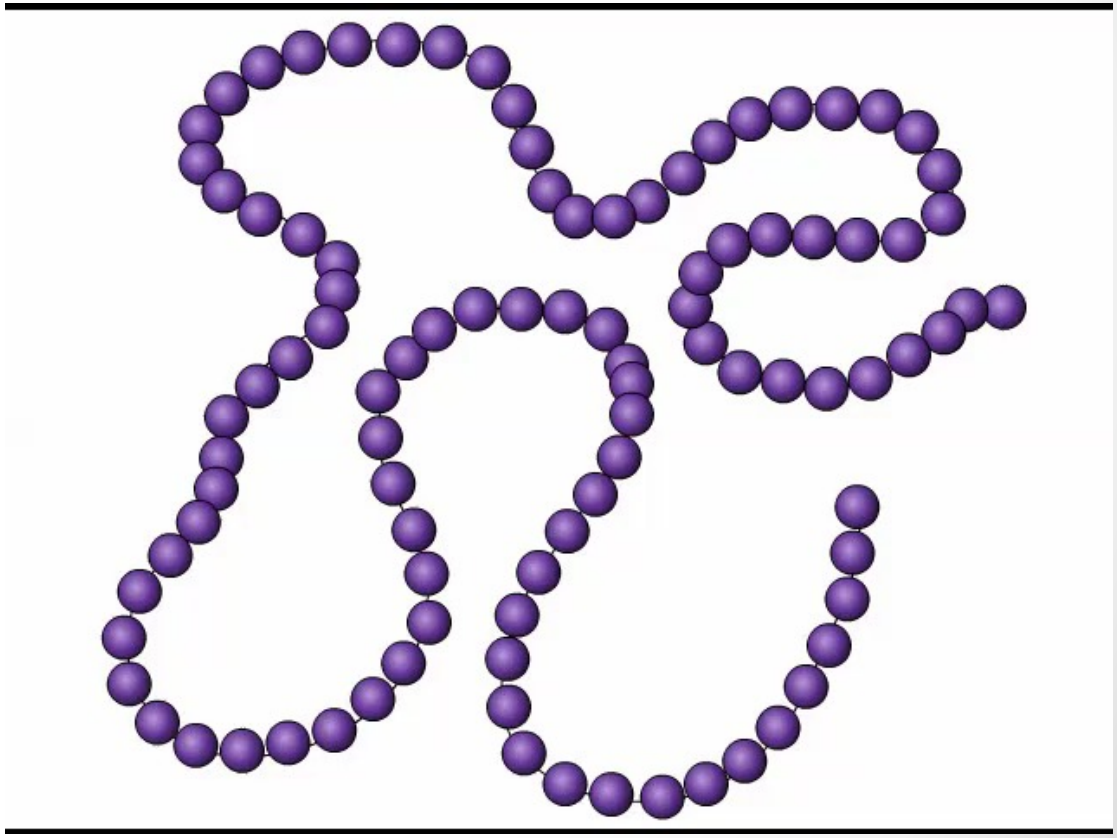


(a) A ribbon model (b) A space-filling model (c) A wireframe model

Primary Structure
蛋白质一级结构：氨基酸序列

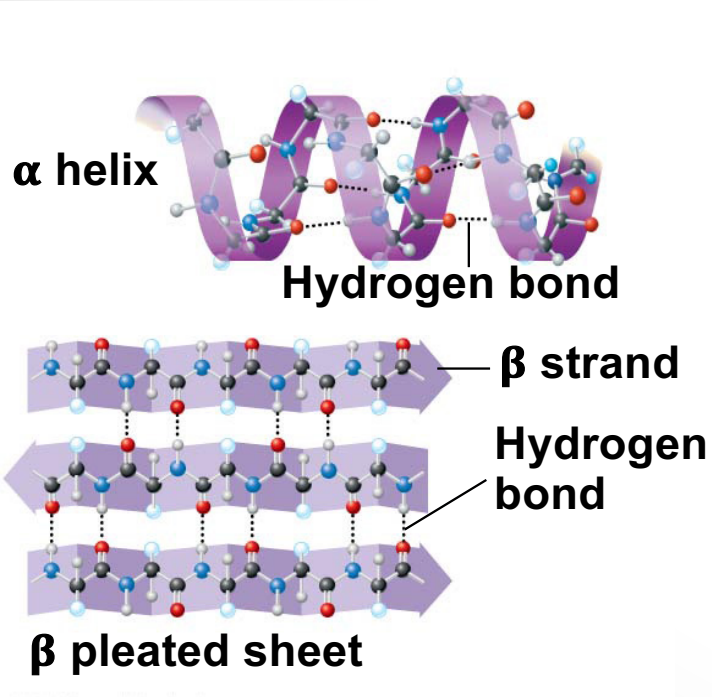


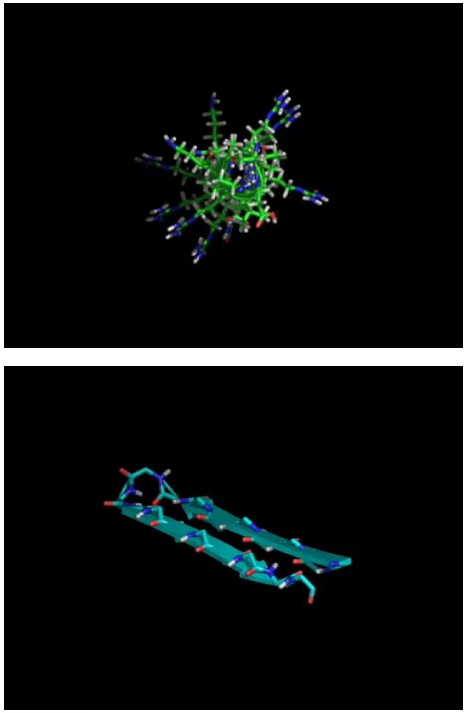
- The **primary structure** of a protein is its sequence of amino acids

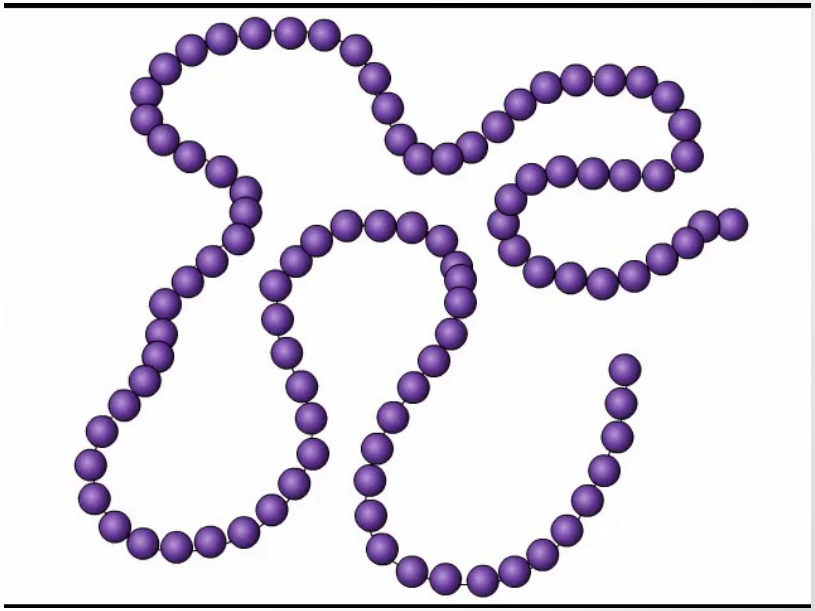


Secondary Structure

- **Hydrogen bonds氢键** between repeating constituents of the **polypeptide backbone多肽链骨架**
- A coil called an α **helix** （α 螺旋）and a folded structure called a β **pleated sheet** （ β折叠）

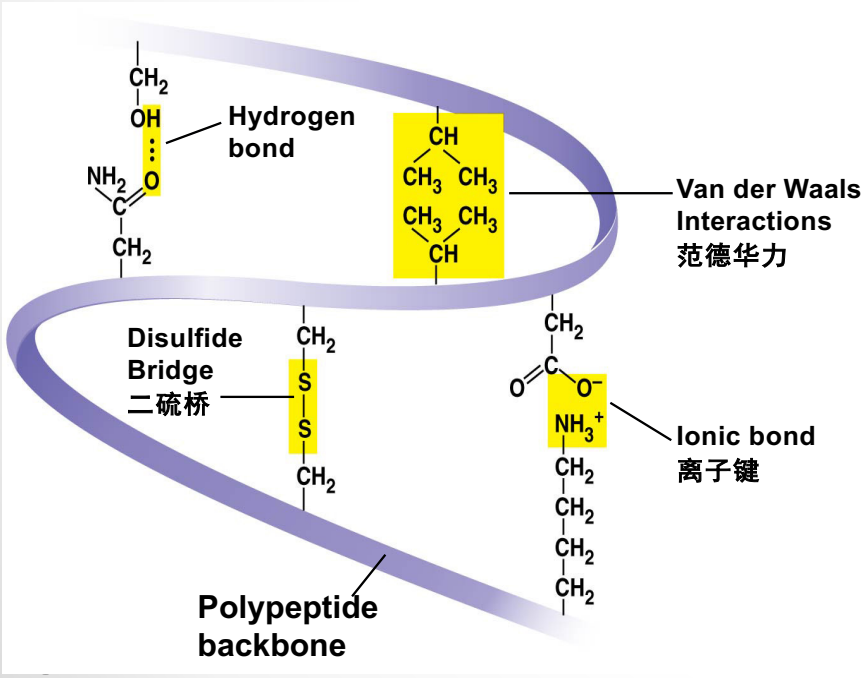


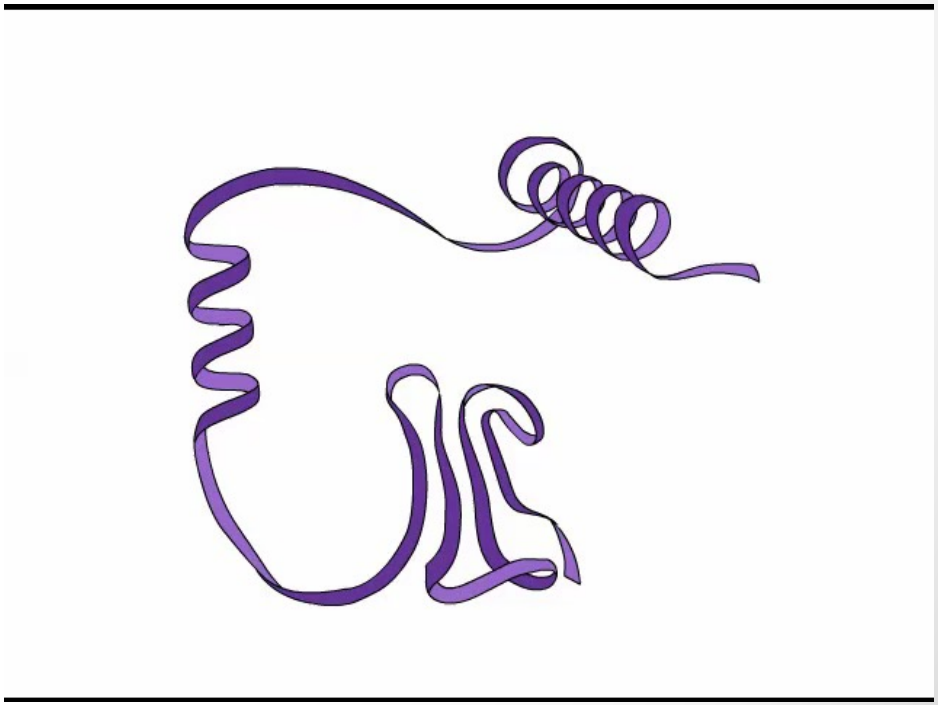




Tertiary Structure

- **Interactions between R groups侧链之间的相互作用**
- These interactions include **hydrogen bonds氢键, ionic bonds离子键, and van der Waals interactions范德华力**
- Strong covalent bonds 共价键 called **disulfide bridges (S-S)二硫桥** may reinforce the protein structure

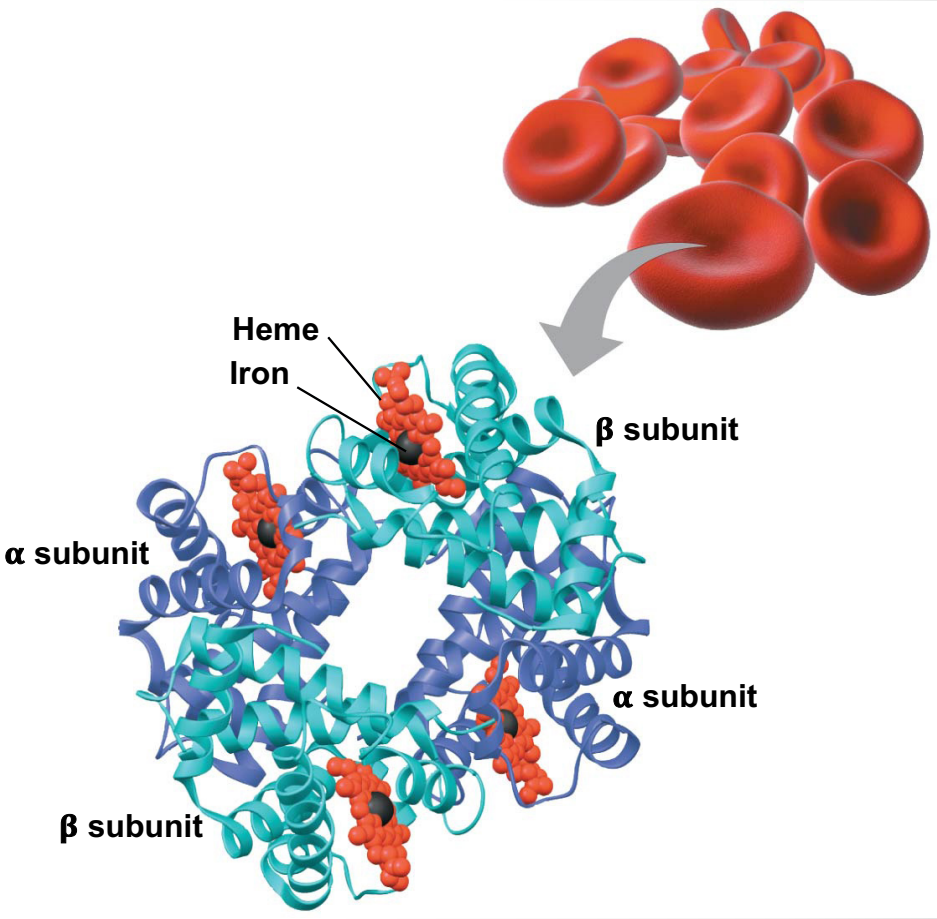




Quaternary Structure

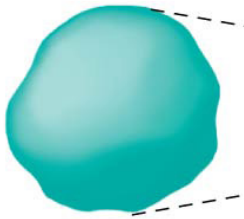
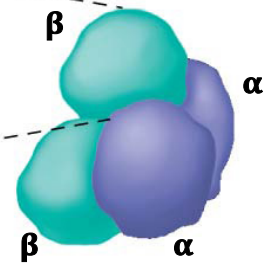
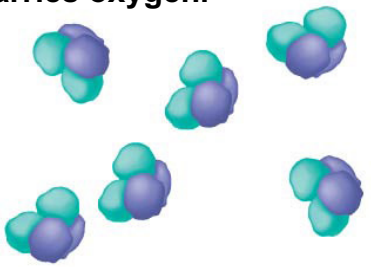
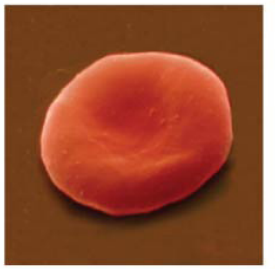
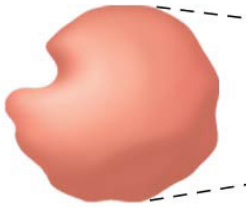
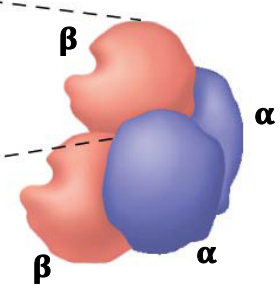
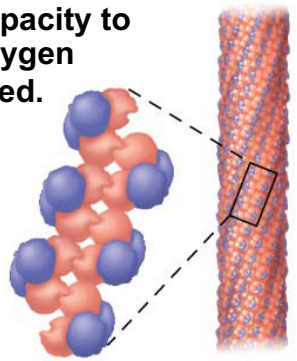
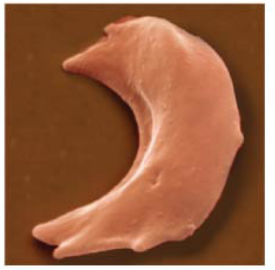
- **Quaternary structure：** two or more polypeptide chains form one functioning macromolecule 多个肽链复合成一个具有功能的蛋白分子

- **Hemoglobin血红蛋白** is a protein consisting of four polypeptides: two alpha and two beta chains

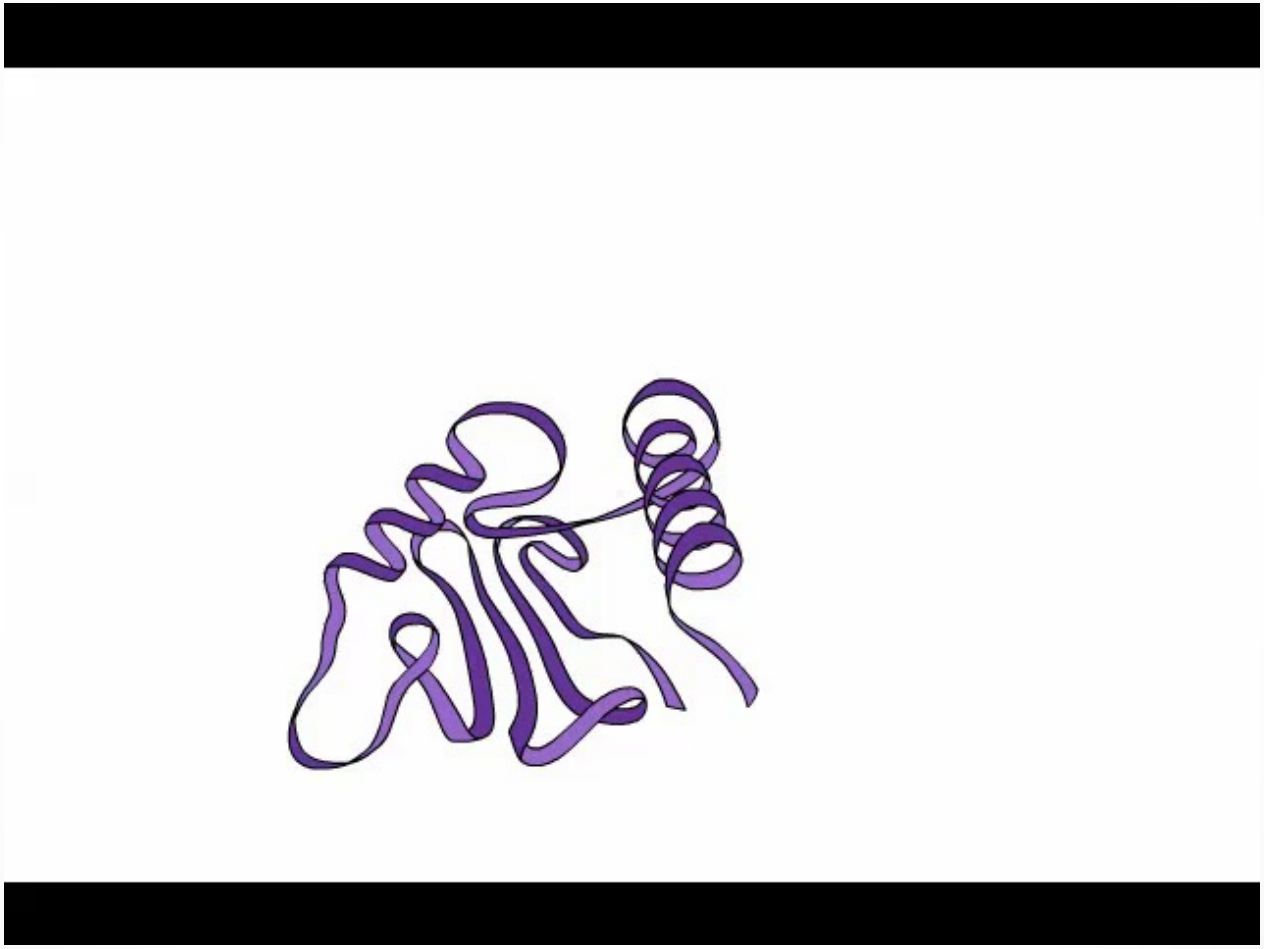


Sickle-cell disease（镰状细胞贫血症）

谷氨酸(Glu)被缬氨酸(Val)代替

	Primary Structure	Secondary and Tertiary Structures	Quaternary Structure	Function	Red Blood Cell Shape
Normal	1 Val 2 His 3 Leu 4 Thr 5 Pro 6 Glu 7 Glu	Normal β subunit 	Normal hemoglobin 	Proteins do not associate with one another; each carries oxygen. 	
Sickle-cell	1 Val 2 His 3 Leu 4 Thr 5 Pro 6 Val 7 Glu	Sickle-cell β subunit 	Sickle-cell hemoglobin 	Proteins aggregate into a fiber; capacity to carry oxygen is reduced. 	

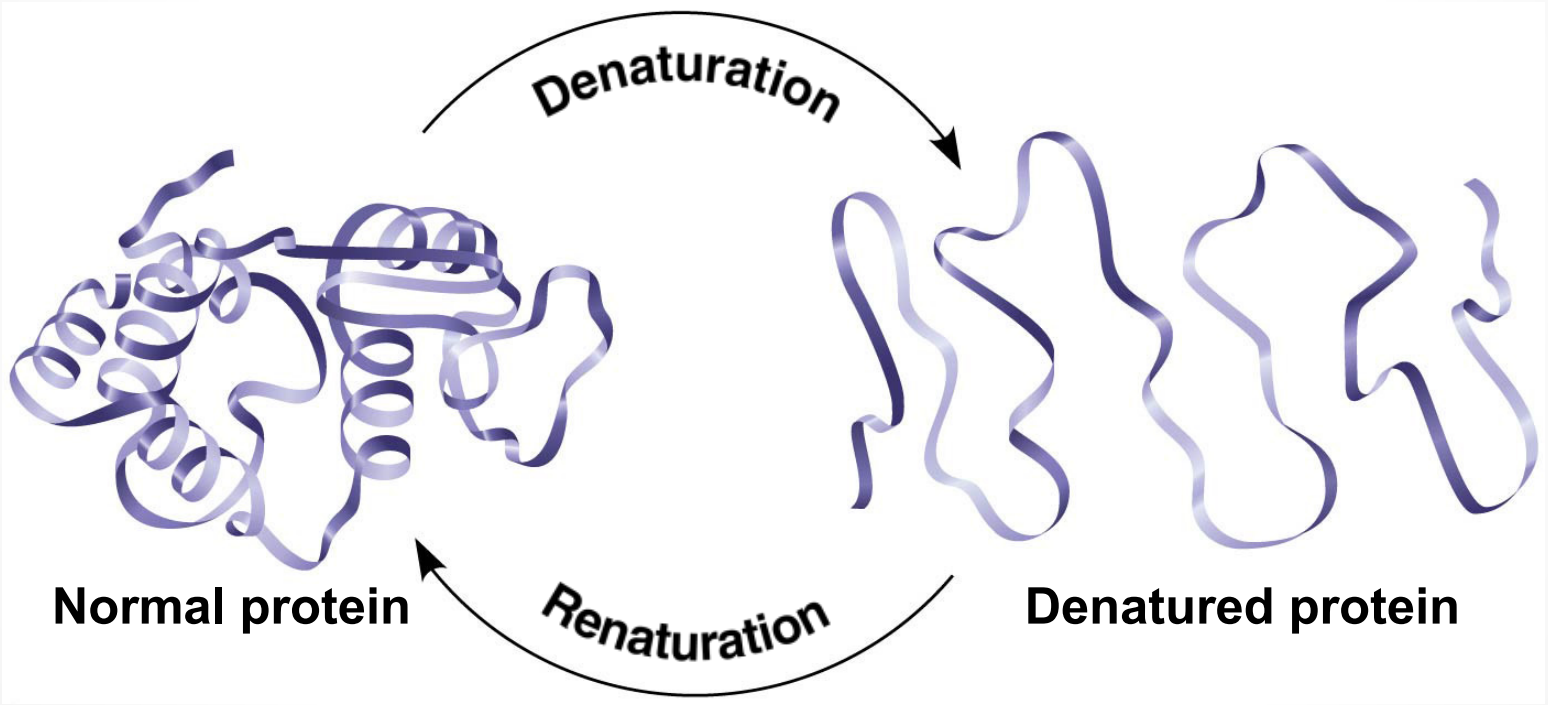
Quaternary Protein Structure



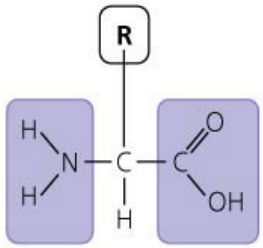
Protein function

- **Enzymes 酶**: catalysts for almost all biochemical reactions in the cell
- **Cytoskeletons 细胞骨架**: create biochemical structures
  - Microtubule and actin
- **Receptor proteins 受体蛋白**
  - Cellular communication
- **Other functions…**

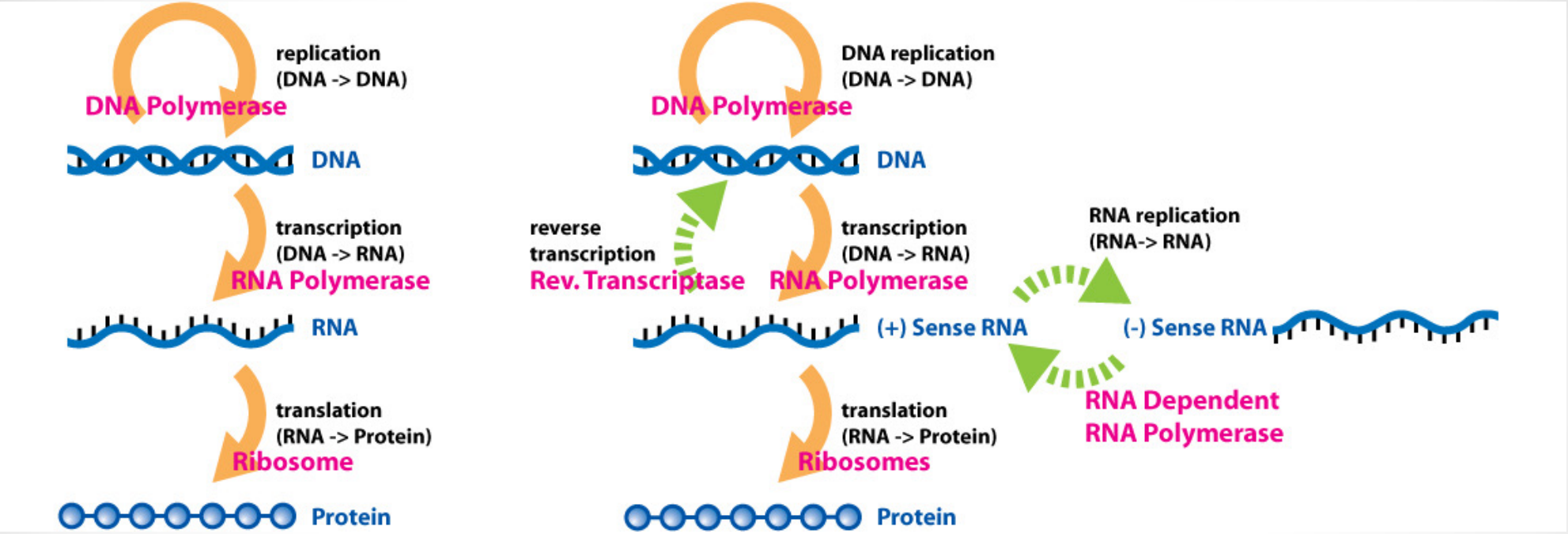
Protein denaturation 变性 and renaturation 复性



Summary: proteins

Large Biological Molecules	Components	Examples	Functions
Proteins include a diversity of structures, resulting in a wide range of functions (pp. 75–84)	 Amino acid monomer (20 types)	• Enzymes • Structural proteins • Hormones • Receptor proteins • Motor proteins • Defensive proteins	• Catalyze chemical reactions • Provide structural support • Coordinate organismal responses • Receive signals from outside cell • Function in cell movement • Protect against disease

The central dogma

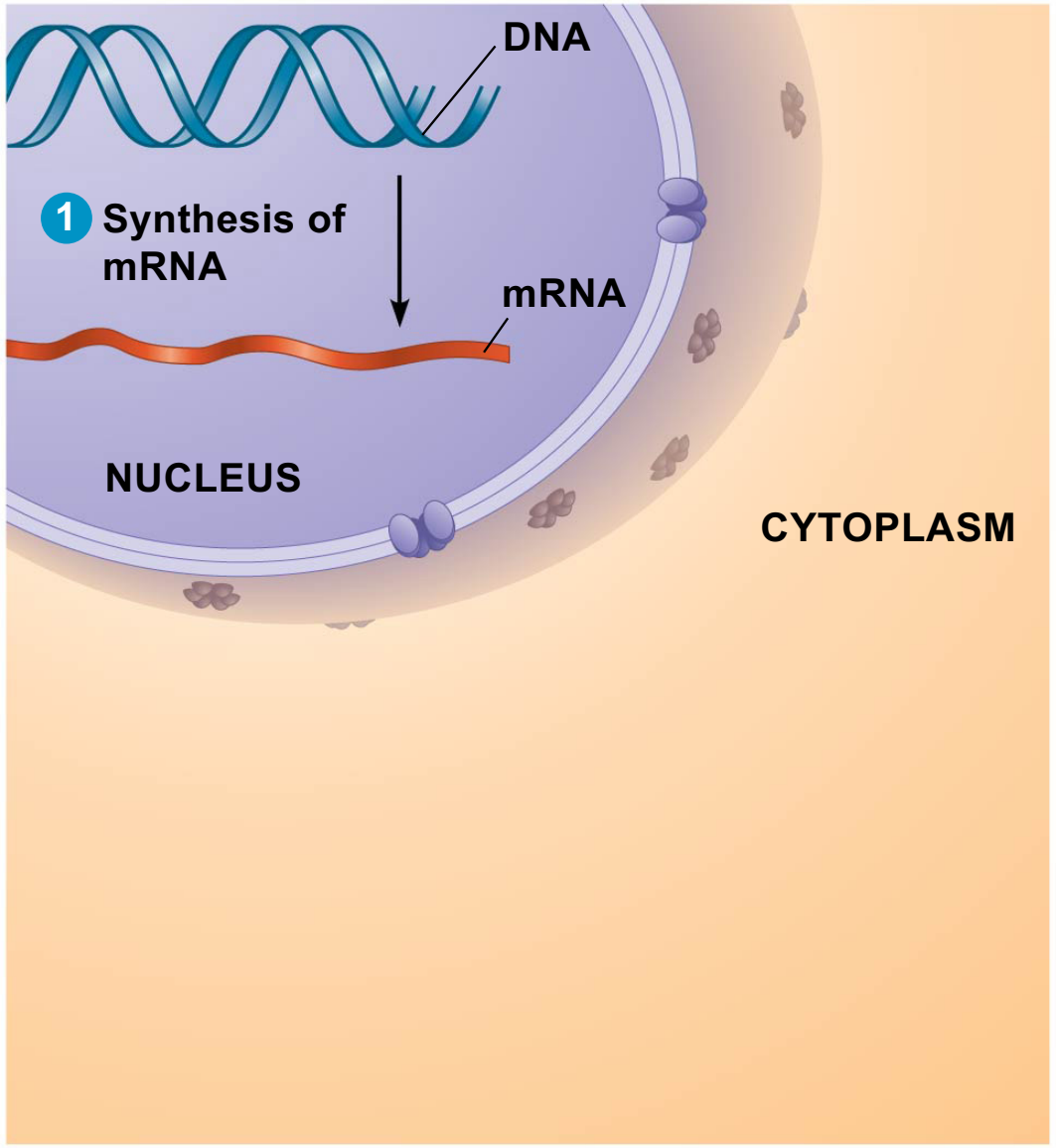


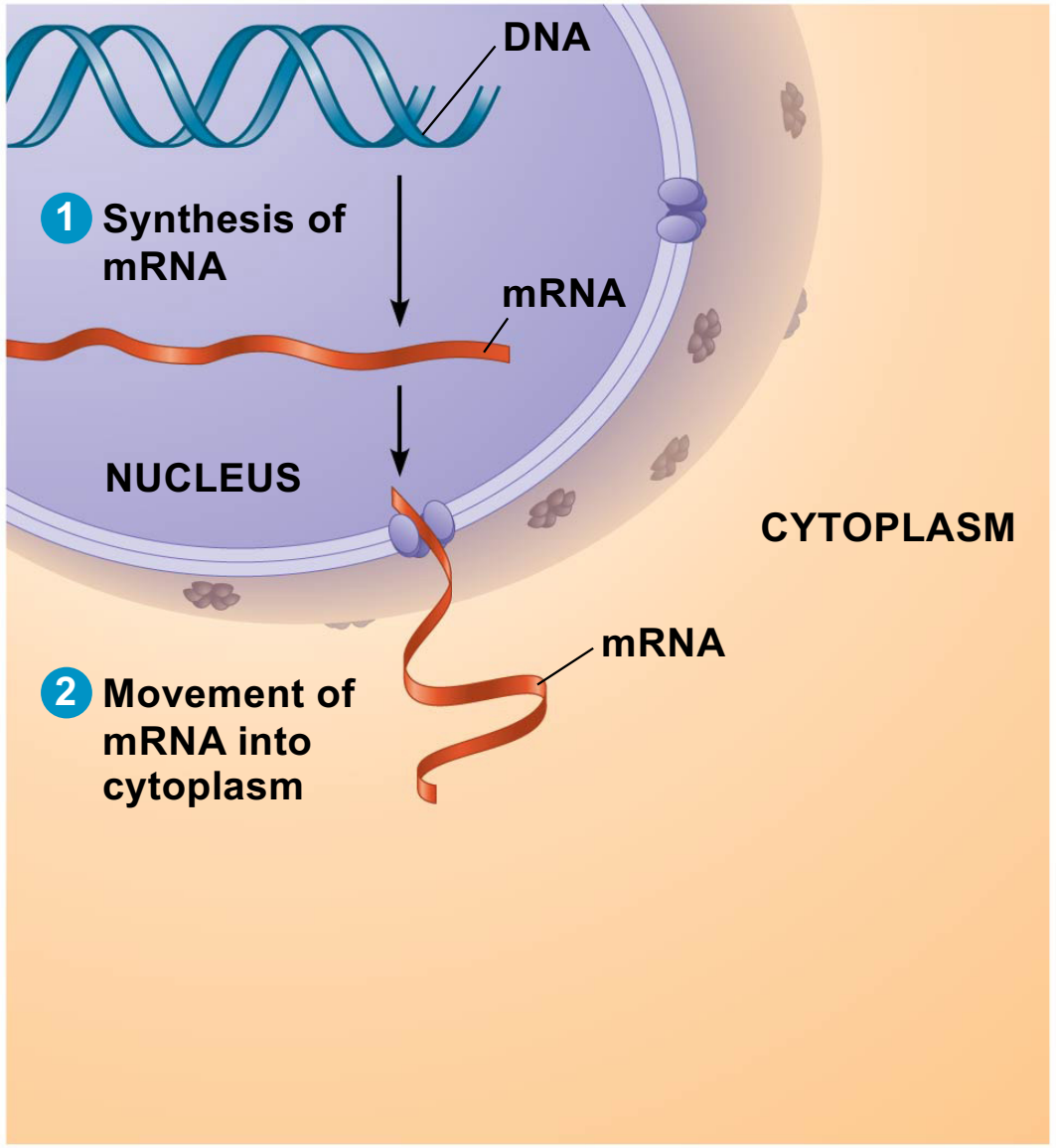
The original dogma

The modified dogma

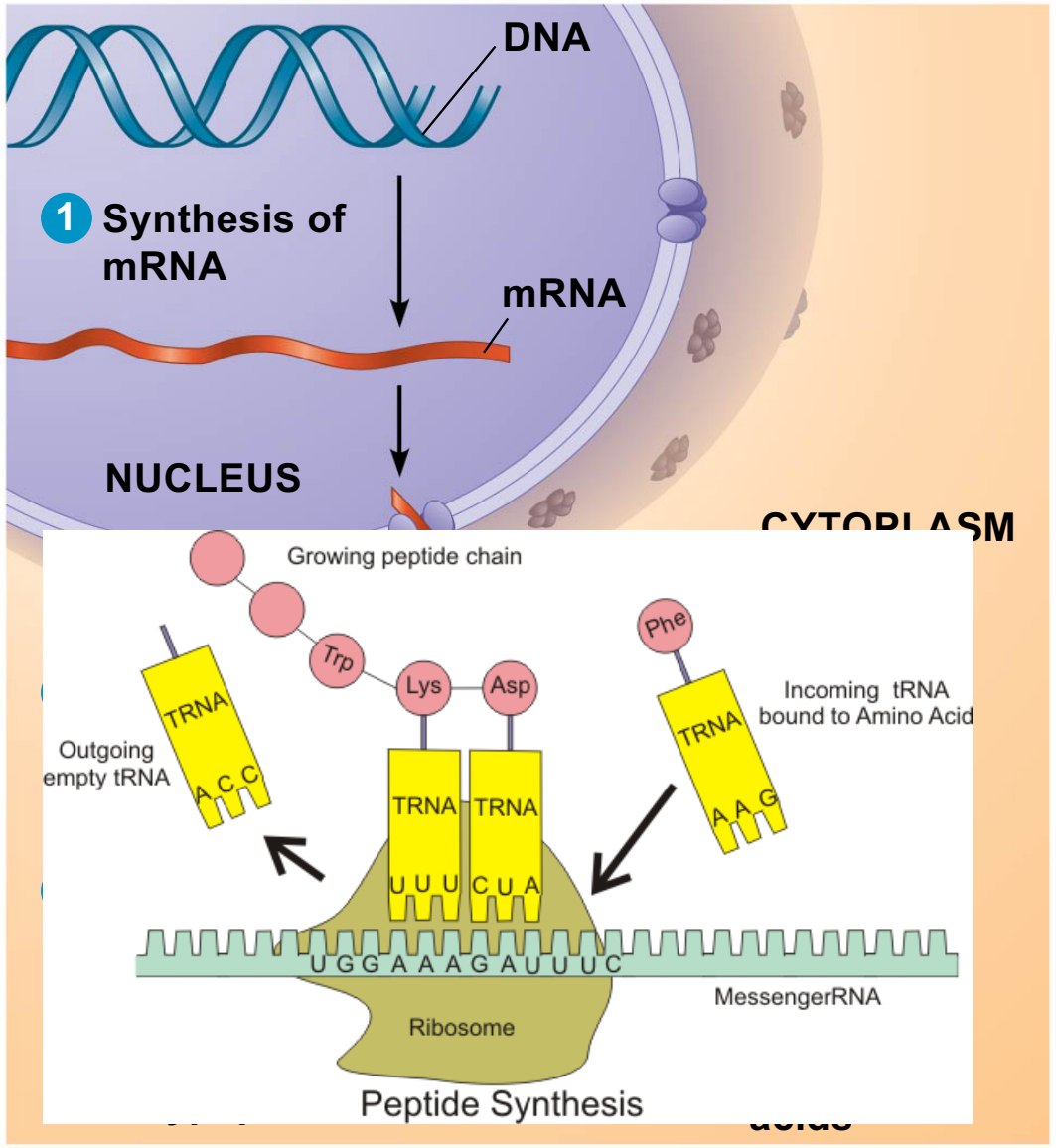
F.H. Crick, 1958: **The transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible.**

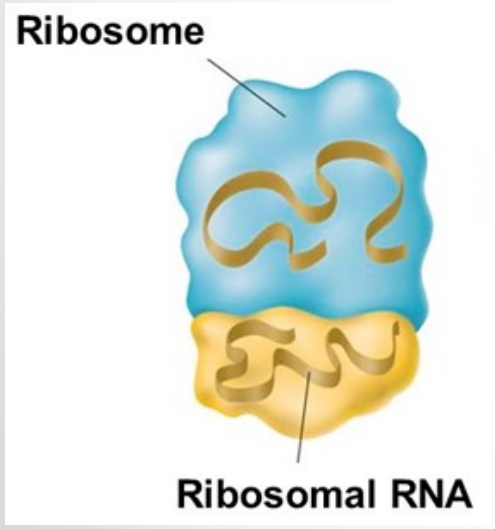
Transcription转录: Each **gene** directs synthesis of a messenger RNA (**mRNA**)

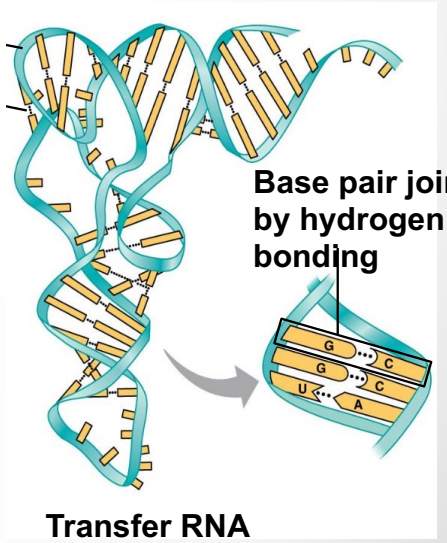



DNA
1 Synthesis of mRNA
mRNA
NUCLEUS
CYTOPLASM
mRNA
2 Movement of mRNA into cytoplasm

Translation 翻译:
mRNA-> Protein







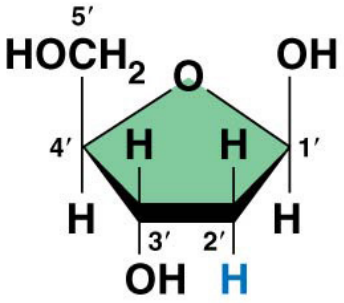
Nucleic acids

- Central dogma: DNA->mRNA->protein
- DNA & RNA: right-handed helix 右旋
- Structure:
  - Components
  - Steps to form DNA and RNA
  - Base pairing rule
  - 3D structure

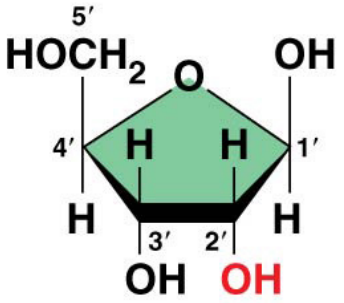
The Components of Nucleic Acids

- Nucleic acids are polymers called **polynucleotides 多聚核苷酸**
- Each polynucleotide is made of monomers called **nucleotides 核苷酸**
- Each nucleotide consists of a **nitrogenous base**碱基, a **pentose sugar**五碳糖, and one or more **phosphate groups**磷酸基团
- The portion of a nucleotide without the phosphate group is called a **nucleoside 核苷**

Sugars



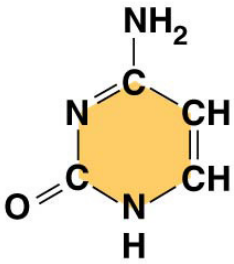
Deoxyribose (in DNA)
脱氧核糖



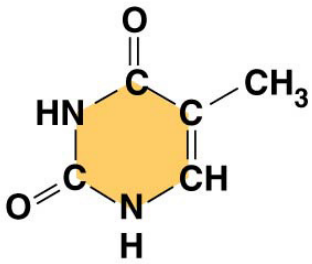
Ribose (in RNA)
核糖

Nitrogenous Bases

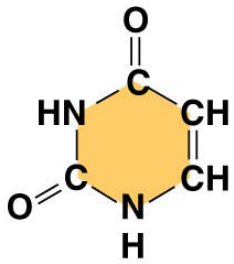
Pyrimidines



Cytosine
(C)
胞嘧啶

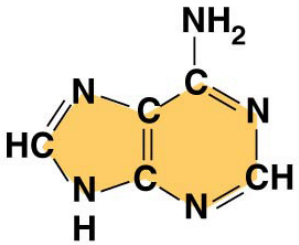


Thymine
(T, in DNA)
胸腺嘧啶

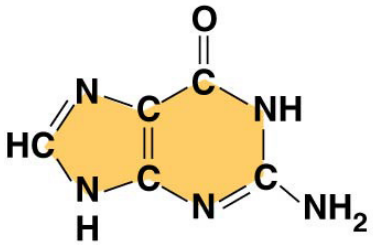


Uracil
(U, in RNA)
尿嘧啶

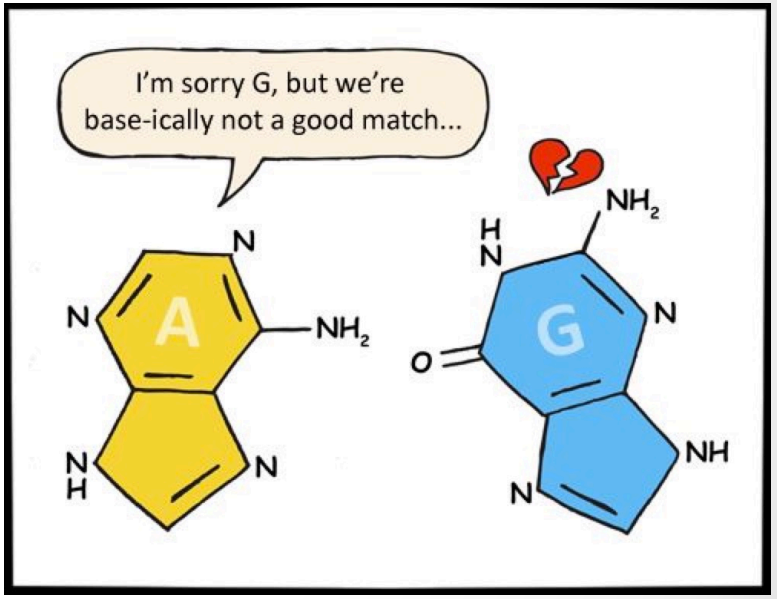
Purines



Adenine (A)
腺嘌呤

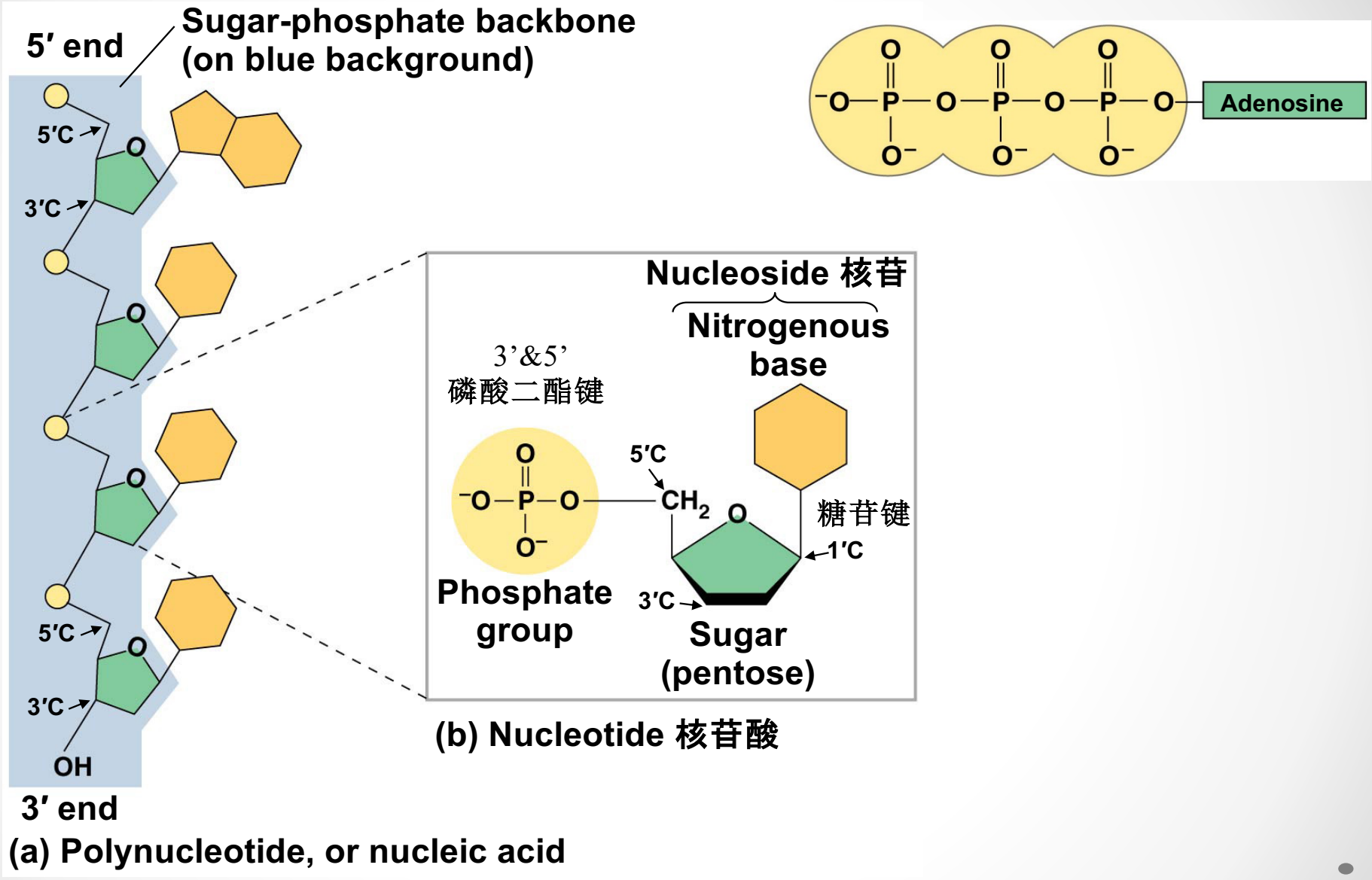


Guanine (G)
胸腺嘧啶



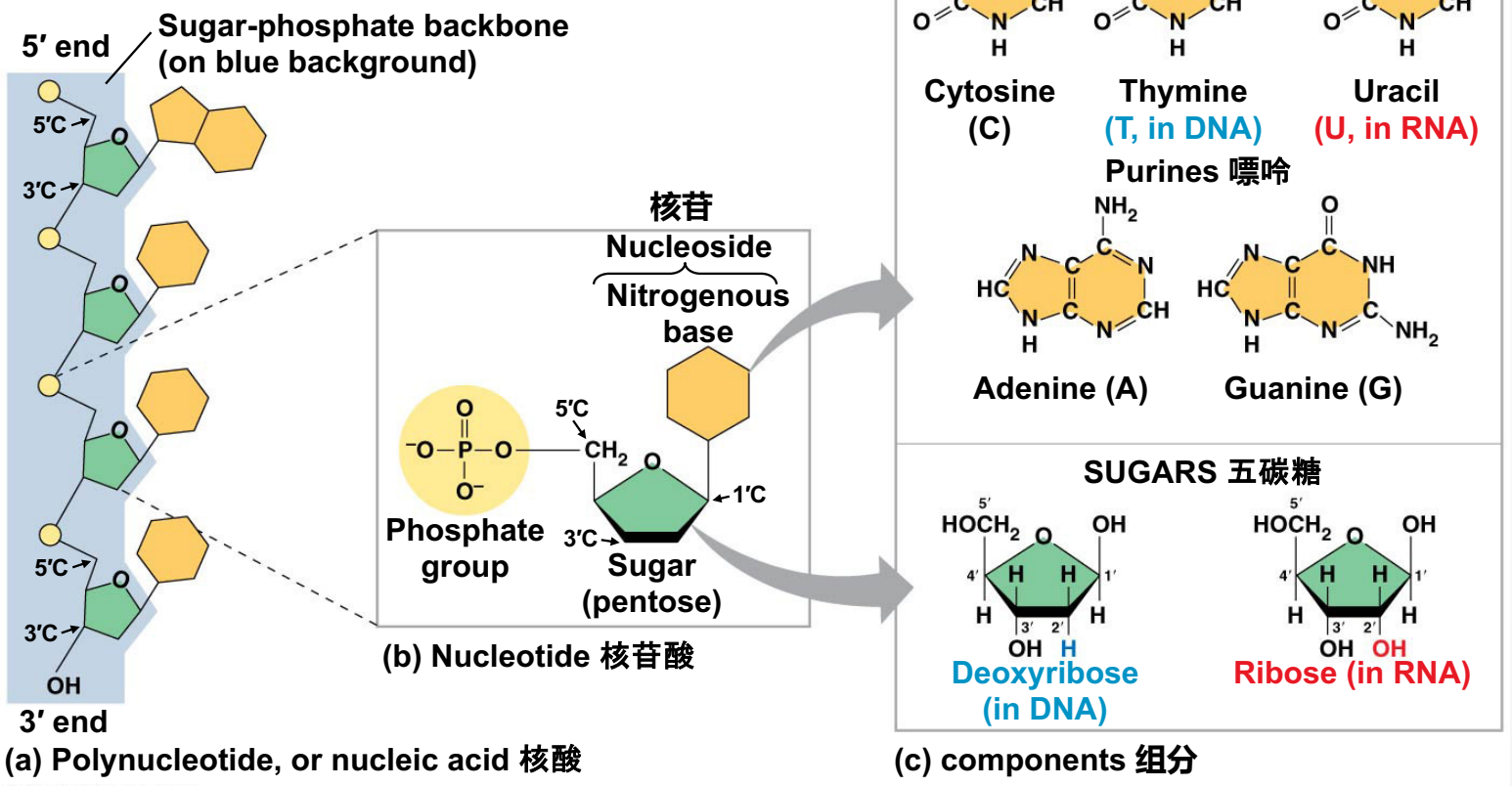
Chargaff's Rule: [A]=[T], [C]=[G]

- Erwin Chargaff, 1940s
- DNA from any cell of all organisms should have a 1:1 ratio of purine (嘌呤） and pyrimidine （嘧啶） bases
- The amount of Adenine（腺嘌呤）= Thymine（胸腺嘧啶）, Cytosine（胞嘧啶） = Guanine（鸟嘌呤）
- "That...such giant shadows are cast by such pygmies only shows how late in the day it has become" (on Watson and Crick)


5′ end
Sugar-phosphate backbone
(on blue background)
5′C
3′C
5′C
3′C
OH
3′ end
(a) Polynucleotide, or nucleic acid
Adenosine
Nucleoside 核苷
Nitrogenous base
3’&5’
磷酸二酯键
5′C
糖苷键
1′C
3′C
Phosphate group
Sugar (pentose)
(b) Nucleotide 核苷酸

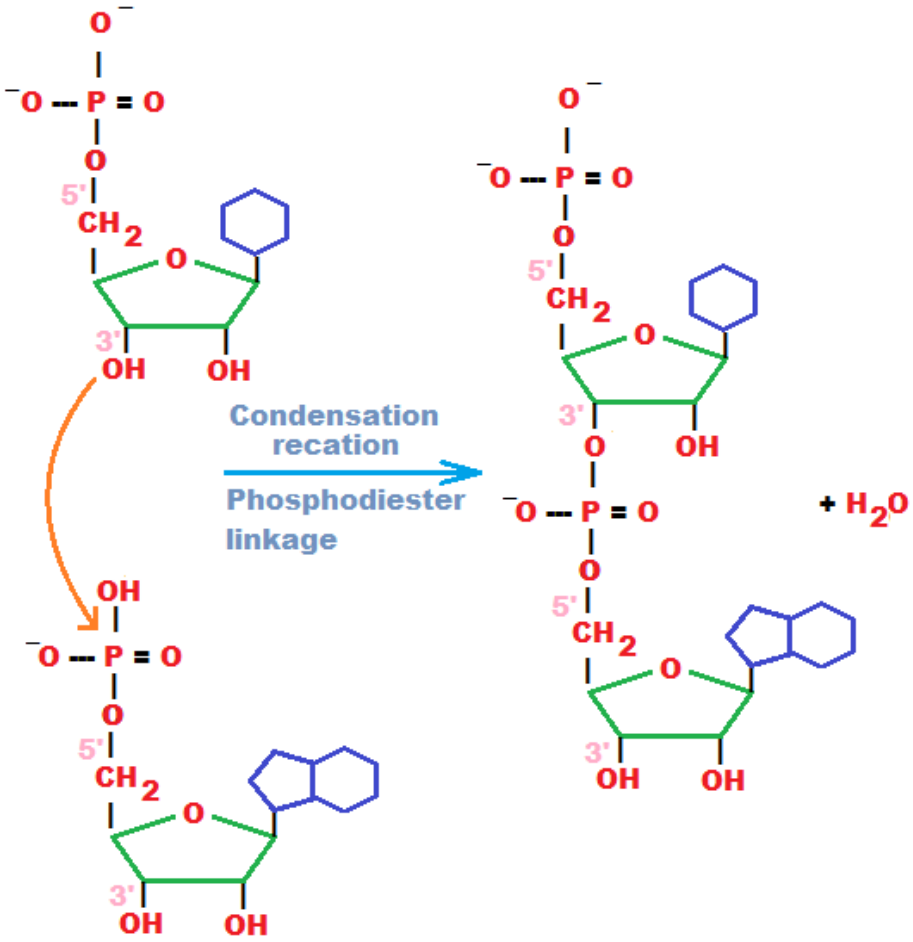
The Components of Nucleic Acids

1. Nucleoside = nitrogenous base + sugar



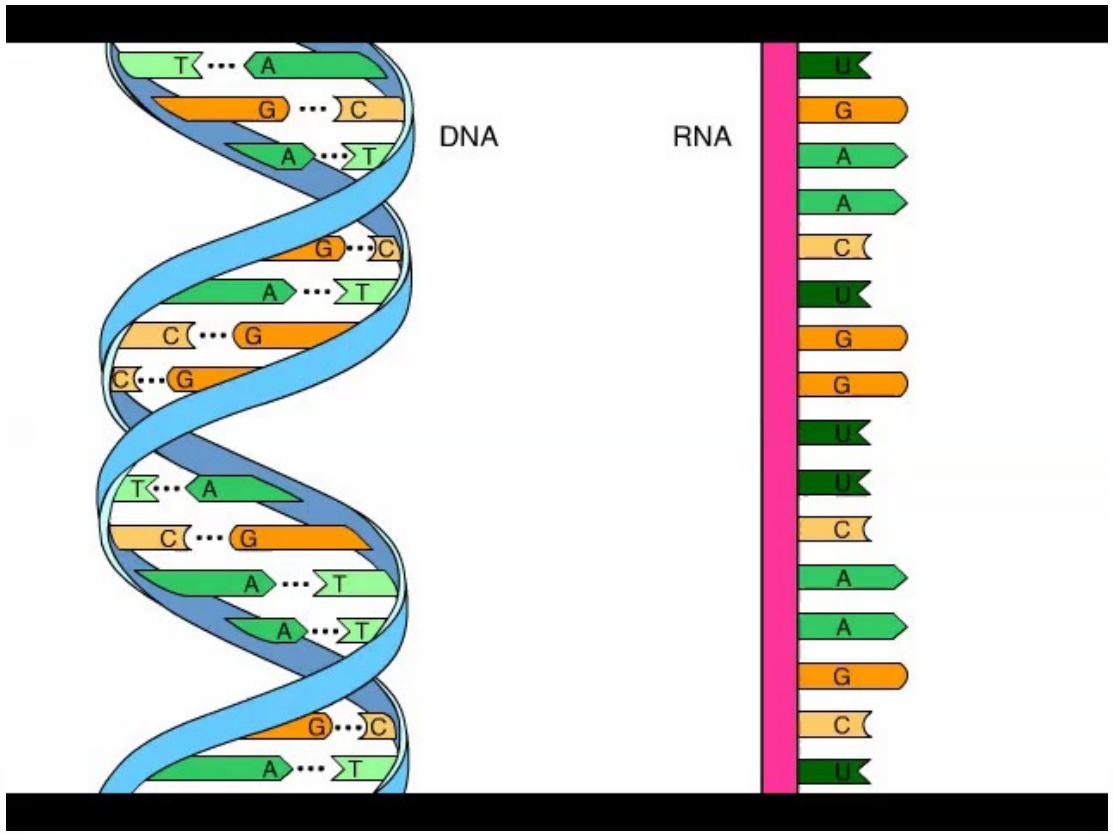
2. Nucleotide = nucleoside + phosphate group

Nucleotide Polymers

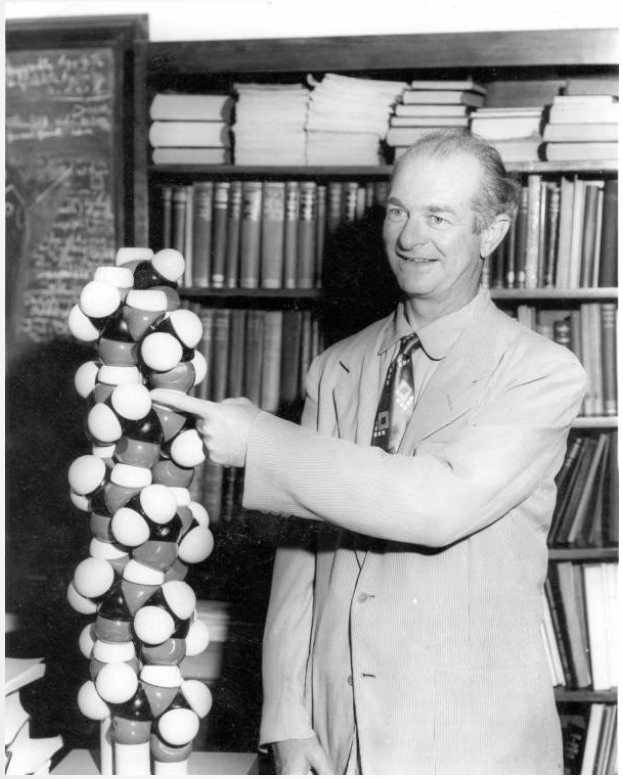


- Adjacent nucleotides are joined by a **phosphodiester linkage**（磷酸二酯键）, which consists of a phosphate group that links the sugars of two nucleotides
- **Direction** of synthesis: 5'->3'
- The sequence of bases along a DNA or mRNA polymer is unique for each gene
- One DNA molecule includes many genes

DNA and RNA Structure



Linus Pauling: triple helix model



Linus C. Pauling
1901 -1994

- Two-time Nobel Laureate: chemistry (nature of chemical bond, 1954) and peace (against nuclear weapon test, 1962)
- Discovered alpha-helical structure in some proteins（蛋白质二级结构：alpha 螺旋）
- Made a crazy triple-helix model of DNA

The DNA lady

Rosalind E. Franklin
1920 -1958

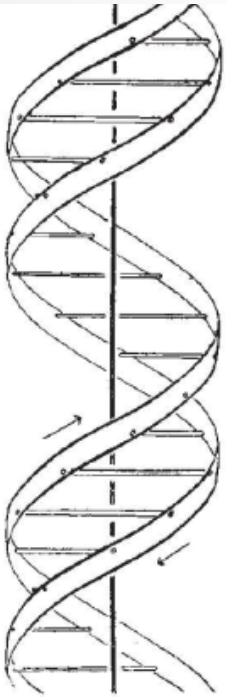
The key X-ray photograph involved in the elucidation of the DNA structure.
(Rosalind Franklin, winter of 1952–1953)

“Secret of life” uncovered: the double helix model

James D. Watson

Francis H. Crick

The most important paper in biology of the 20th century: only 1 page



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

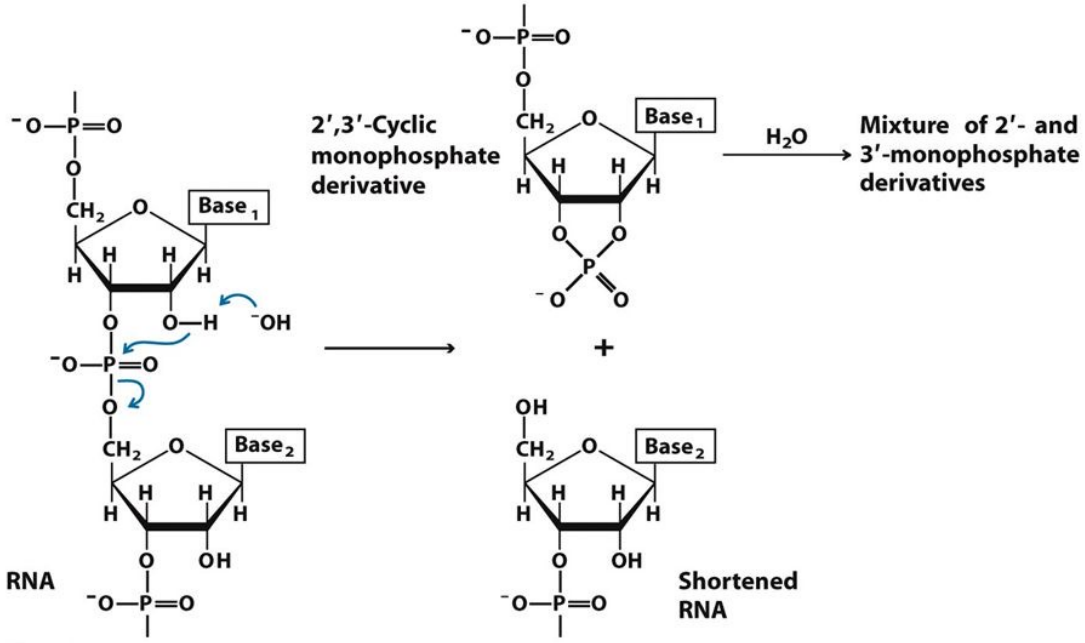
- **"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."**
- This paper marks the end of search for genetic materials, and the beginning of modern life science.

The Structure of DNA

- Two polynucleotides spiraling around an imaginary axis, forming a **double helix** 双螺旋
- The backbones run in opposite 5′ → 3′ directions "**antiparallel**" 反向平行
- **complementary base pairing** 碱基配对: A-T, C-G. Connect through **Hydrogen Bond** 氢键
- Two identical copies of each DNA molecule are generated in a dividing cell, with **10^{-15}** error rate during replication 极低错误率的半保留复制

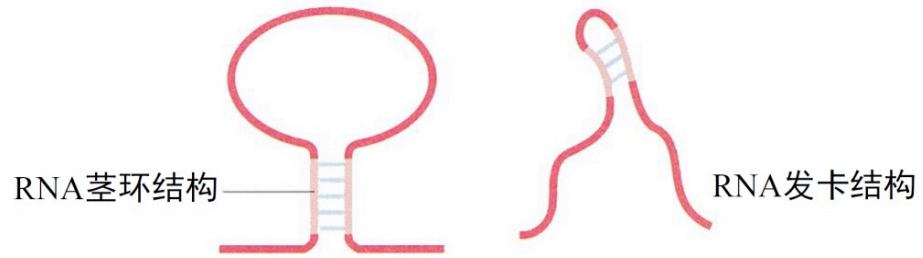
The Structure of RNA

- **Single strand**
- **Complementary pairing** occur between molecules or within one molecule
- In RNA, thymine (T) is replaced by **uracil (U, 尿嘧啶)** so A and U pair
- While DNA is always a double helix, RNA is more **variable** in form
- RNA **hydrolyze** in alkaline solutions 碱性溶液中，RNA水解

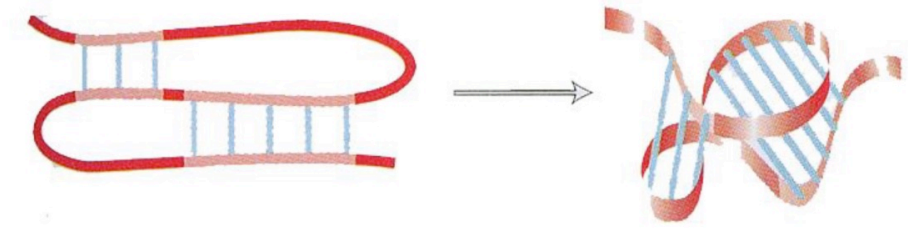


Complex structures of RNA

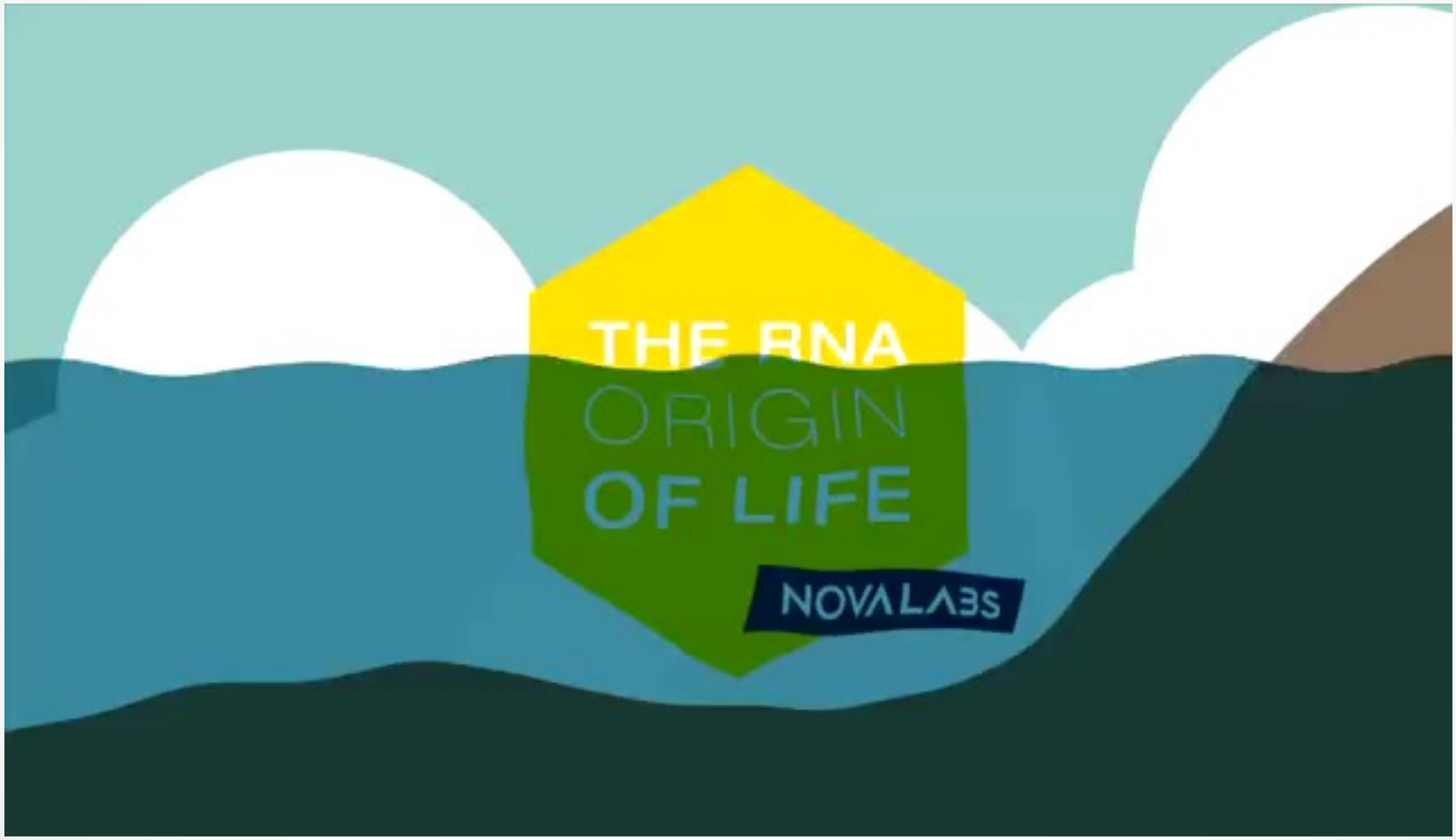
- **Secondary structure**



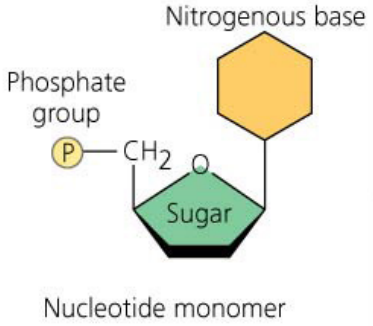
- **Tertiary structure**



RNA origin of life theory



Summary: nucleic acids

Large Biological Molecules	Components	Examples	Functions
Nucleic acids store, transmit, and help express hereditary information (pp. 84–87)	 Nucleotide monomer	**DNA:**  • Sugar = deoxyribose • Nitrogenous bases = C, G, A, T • Usually double-stranded	Stores hereditary information
		RNA:  • Sugar = ribose • Nitrogenous bases = C, G, A, U • Usually single-stranded	Various functions in gene expression, including carrying instructions from DNA to ribosomes

讨论题

1. 目前市面上有很多基因检测服务，右边表格是一份检测报告示例。此报告中的哪些项目有科学依据?

2. 有的人血型罕见，被称为“熊猫血”。什么是“熊猫血”？这些人往往会定期献血。献血对人体有害吗?

3. 有明星宣传通过“控糖”、“戒糖”来“抗衰老”和“护肤”。这件事情有科学依据吗?

4. 体内脂肪和胆固醇偏高往往被认为对健康有害，但也有“好脂肪”和“好胆固醇”的说法。它们“好”在哪里?

5. AlphaFold是什么，有什么用?

6. 你认为以下哪些是可遗传的特征，依据是什么?

——视力、身高、眼睛颜色、性格、智商、运动天赋、音乐天赋

酒精性脸红	喝酒可能脸红
苦味敏感度	对苦味敏感
耳垢类型	干燥的耳垢
吸烟倾向	吸烟不容易上瘾
血糖水平	空腹血糖水平较高
见光喷嚏	见光打喷嚏的几率正常
避错倾向	可能不善于避免错误
绝对音准	绝对音准能力不明显
ABO 血型	AB 型血
APOE分型	APOE基因型为ε3/ε3
深度睡眠	深度睡眠时间较短
基因身高	176.7 CM