BCHEM 365
Lecture 3
September 12, 2018

Complexity of DNA –Cot curves

$C_o t$ = initial DNA concentration (C_o)
in moles of nucleotides per liter

× time (t) in seconds

If all other factors (concentration, pH, ionic concentration) are equal, extent of renaturation of complementary strands depends on the $C_o t$ (i.e. complexity)

One particular value that is useful is $C_o t_{½}$, the *Cot* value where half of the DNA has reannealed

There is a linear relationship between DNA complexity and $C_o t_{½}$

$C_o t$ curves

There is a linear relationship between DNA complexity and $C_o t_{½}$

Compare calf genome size - 2×10^9 bp of DNA

to T4 bacteriophage genome size - 2×10^5 bp of DNA

10,000 times as many base pairs in calf DNA

Implication: it takes 10,000 times as long to reanneal denatured calf DNA and

$C_o t$ value is 10,000 times larger

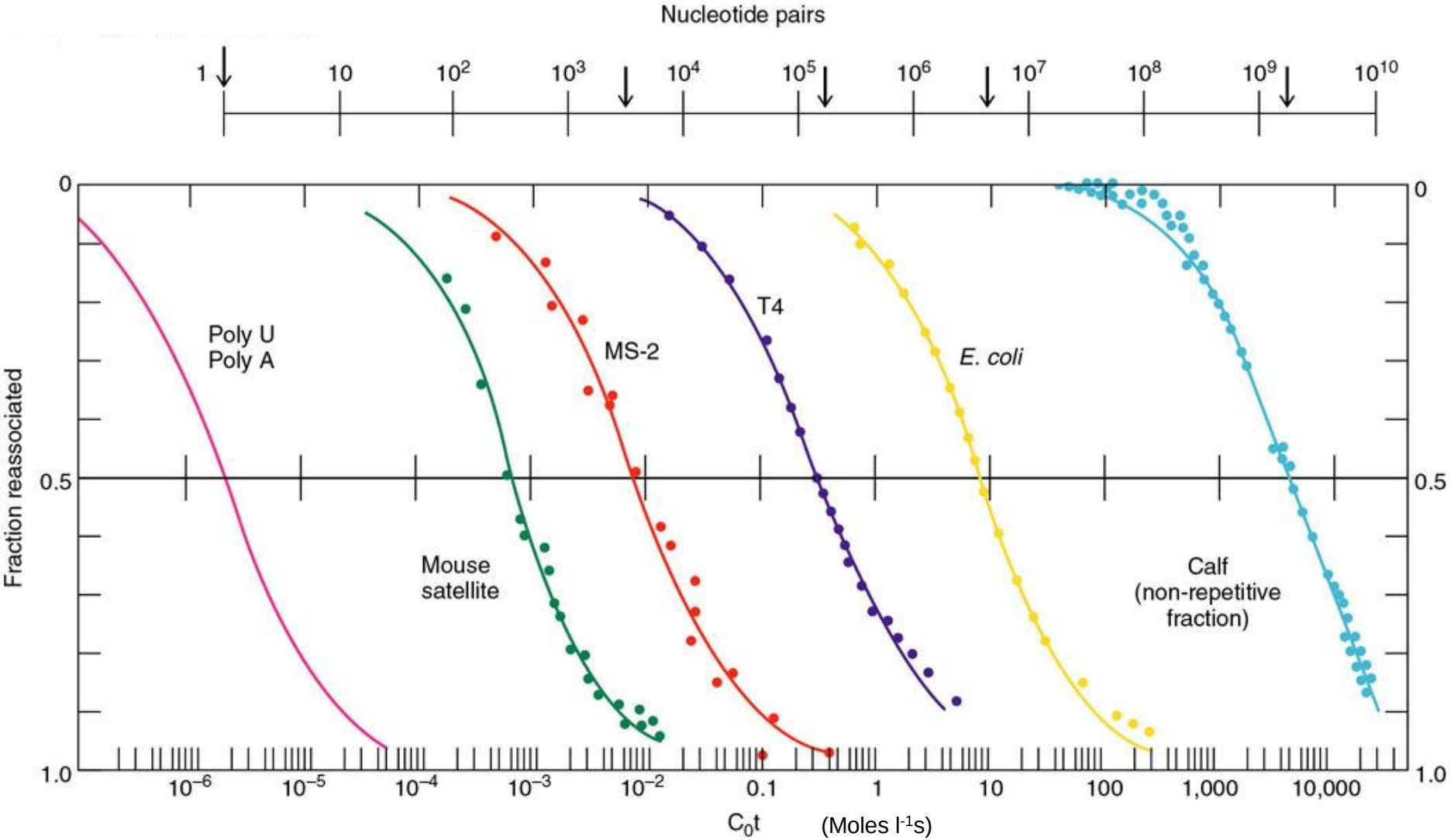
$C_o t$ curves

- Construction of a $C_o t$ curve involves measurement of extent of reannealing of DNA strands
- $C_o t$ curve is drawn with x-axis having DNA concentration in moles/L multiplied by time ($C_o t$) on a logarithmic scale. Since the initial concentration is considered as C_o, multiplying by time t, gives "$C_o t$" and the graph is known as $C_o t$ curve
- The y-axis is % reanealled DNA (fraction reassociated)

$C_o t$ curves

- The Y-axis may also be expressed as percent of the DNA that remains single stranded or not annealed
- This is expressed as a ratio of the concentration of single-stranded (fraction not annealed) DNA (*C*) to the total concentration of the initial DNA (C_o), that is, C/C_o
- The higher the ratio (that is more single-stranded DNA), the less DNA that is reannealed
- As can be seen the curve is rather smooth which indicates that reannealing occurs gradually over a period of time

C_0t curves



It takes DNA toward right side of graphs much longer time to reanneal. The converse is true. Each dot represents a variable repeat: repetitive or nonrepetitive DNA. It takes nonrepetitive DNA much longer time to reanneal. Mouse satellite DNA consists of highly repetitive tandem sequences located in the centromeric region and forms the main structure of heterochromatin. It is noncoding.

$C_o t$ curves

- The shape of a "$C_o t$" curve for a given species is a function of two factors:
  - the size or complexity of the genome
  - the amount of repetitive DNA within the genome

Conclusions from $C_o t$ curves

- As genome size increases, a higher concentration is required for the denatured DNA to anneal
- As genome size increases and nonrepetive sequences increase, a longer time is required for denatured DNA to anneal

DNA Complexity

$C_0t_{1/2}$ indicates the total length of different sequences present = complexity (usually in bp)

$C_0t_{1/2}$ of any genome is proportional to its complexity

Complexity of any DNA is determined by comparing its $C_0t_{1/2}$ with a standard DNA of known complexity (usually *E.coli*).

$$\frac{C_0t_{1/2}\text{ (DNA of any genome)}}{C_0t_{1/2}\text{ (}E.\ coli\text{ DNA)}} = \frac{\text{Complexity of any genome}}{4.2 \times 10^6\text{ bp}}$$

$$\text{Cot}_{1/2}\text{ E. coli DNA} = 4.2\times10^6\text{ bp}$$

Therefore $\text{Cot}_{1/2}$ of DNA of any genome = what complexity of the genome?

$$\text{complexity of unknown genome} = \frac{\text{Cot}_{1/2}\text{ of DNA of any genome}}{\text{Cot}_{1/2}\text{ E. coli DNA}} \times 4.2\times10^6\text{ bp}$$

DNA Complexity: bacterial vs. eukaryotic

Eukaryotic genomes have $C_o t$ values
up to 8 orders of magnitude –
much broader than in prokaryotes

Eukaryotic genome includes several components
each reassociating with characteristic kinetics

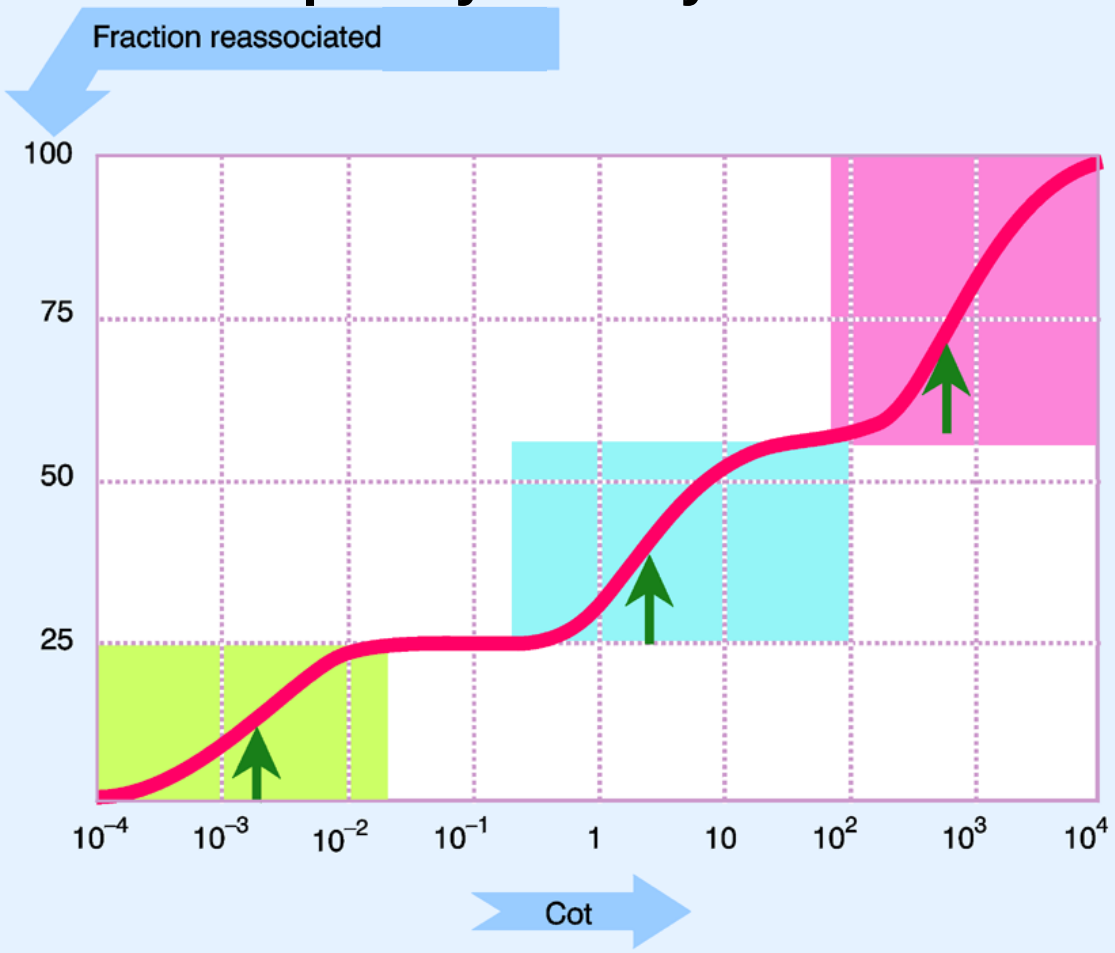
$C_o t$ curve reveals a crucial difference between bacteria and eukaryotic genomes:

bacteria genomes - a single kinetic component

eukaryotic genomes – varying kinetic components (much more complex)

Complexity: eukaryotic

The reassociation kinetics of eukaryotic DNA show three types of component (indicated by the shaded areas). The arrows identify the $Cot_{1/2}$ values for each component.



	Fast component	Intermediate component	Slow component
Percent of genome	25	30	45
$Cot_{1/2}$	0.0013	1.9	630
Complexity, bp	340	6.0×10^5	3.0×10^8
Repetition frequency	500,000	350	1

Intermediate and fast components

Intermediate and fast components are made up of moderately repetitive and highly repetitive DNA, respectively

	Fast component	Intermediate component	Slow component
Percent of genome	25	30	45
$Cot_{1/2}$	0.0013	1.9	630
Complexity, bp	340	6.0×10^5	3.0×10^8
Repetition frequency	500,000	350	1

Intermediate component

Not a single length of DNA

but a variety of individual sequences, whose total length together comes to 6×10^5 bp

Dispersed throughout the genome and their average repetition is 350

	Fast component	Intermediate component	Slow component
Percent of genome	25	30	45
$Cot_{1/2}$	0.0013	1.9	630
Complexity, bp	340	6.0×10^5	3.0×10^8
Repetition frequency	500,000	350	1

Repetitive DNA

If repetitive DNA is heated, the double strands melt gradually over a wide temperature range

If allowed to renature, the reassociated DNA do not consist of exactly paired molecules, some mispairing occur

The more mispairing, the lower the T_m because mismatches unwind easily

Moderately repetitive and highly repetitive DNA

Moderately repetitive DNA is dispersed throughout genome

are relatively short individual sequences

Many are ribosomal genes

Highly repetitive DNA often forms discrete clusters

Found at centromeres

Usually not transcribed nor translated into protein

Mouse

40% of mouse DNA is repetitive

100,000 -1 million copies per genome

Highly repetitive is about 10% of the total DNA

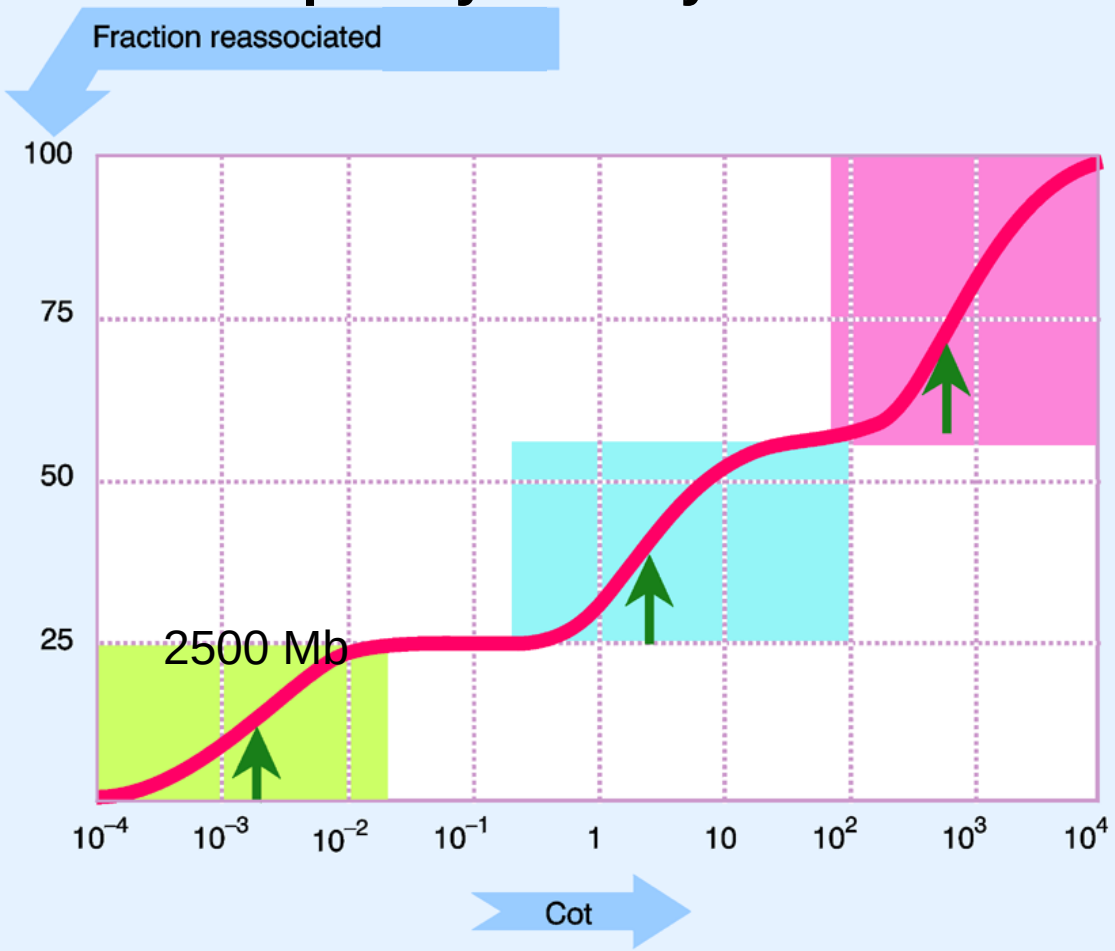
Certain short sequences are repeated 1 million times in genome

Moderately repetitive is about 30% of the total DNA

Single copy DNA

- Forms the slow component
- Are relatively longer sequences
- Are not repetitive
- Make up genes, therefore are the coding DNA sequence (transcribed and translated)
- Takes higher concentration and longer time to reanneal

Complexity: eukaryotic



	Fast component	Intermediate component	Slow component
Percent of genome	25	30	45
$Cot_{1/2}$	0.0013	1.9	630
Complexity, bp	340	6.0×10^5	3.0×10^8
Repetition frequency	500,000	350	1

Figure 3.6 The reassociation kinetics of eukaryotic DNA show three types of component (indicated by the shaded areas). The arrows identify the $Cot_{1/2}$ values for each component.

Single copy DNA

- Maize genome size is 2.5 billion bp. Number of genes (protein-coding genes) is 32,000 spread across 10 chromosomes

- The human genome (23 chromosomes) is estimated to be about 3.2 billion bp long, contains 20,000–25,000 distinct protein-coding genes

Conclusions

The double helix is denatured at high temperatures

$C_o t$ is the product of
the initial DNA concentration (C_o)
in moles of nucleotides per liter

and time (t) in seconds

and is a measure of the **complexity** of DNA
Complexity = size of the DNA; type of sequence
(highly repetitive, mod. repetitive, or single copy)

DNA sizes and shapes

- DNA sizes expressed in three ways
  - Molecular weight
  - Base pairs
  - Length
    - All are related

Sizes of DNA from various sources

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

Table 2.4 Sizes of various DNAs

Source	Molecular weight	Base pairs	Length
Subcellular Genetic Systems:			
SV40 (mammalian tumor virus)	3.5×10^6	5226	1.7 μm
Bacteriophage ϕX174 (double-stranded form)	3.2×10^6	5386	1.8 μm
Bacteriophage λ	3.3×10^7	5×10^4	13 μm
Bacteriophage T2 or T4	1.3×10^8	2×10^5	50 μm
Human mitochondria	9.5×10^6	16,596	5 μm
Prokaryotes:			
Haemophilus influenzae	1.2×10^9	1.83×10^6	620 μm
Escherichia coli	3.1×10^9	4.65×10^6	1.6 mm
Salmonella typhimurium	8×10^9	1.1×10^7	3.8 mm
Eukaryotes (content per haploid nucleus):			
Saccharomyces cerevisiae (yeast)	7.9×10^9	1.2×10^7	4.1 mm
Neurospora crassa (pink bread mold)	$\approx 1.9 \times 10^{10}$	$\approx 2.7 \times 10^7$	≈9.2 mm
Drosophila melanogaster (fruit fly)	$\approx 1.2 \times 10^{11}$	$\approx 1.8 \times 10^8$	≈6.0 cm
Rana pipiens (frog)	$\approx 1.4 \times 10^{13}$	$\approx 2.3 \times 10^{10}$	≈7.7 m
Mus musculus (mouse)	$\approx 1.5 \times 10^{12}$	$\approx 2.2 \times 10^9$	≈75 cm
Homo sapiens (human)	$\approx 2.3 \times 10^{12}$	$\approx 3.5 \times 10^9$	≈120 cm
Zea mays (corn, or maize)	$\approx 4.4 \times 10^{12}$	$\approx 6.6 \times 10^9$	≈2.2 m
Lilium longiflorum (lily)	$\approx 2 \times 10^{14}$	$\approx 3 \times 10^{11}$	≈100 m

DNA sizes and shapes

- To convert bp to length
  -About 10.5 bp equal one helical turn of 34 Å long

Question

Phage P1 has a double stranded DNA with 91,500 bp (91.5 kb). How many full double-helical turns are present in this DNA?

-10.5 bp make one turn, therefore 91,500 bp?

$$\frac{91{,}500\ \text{bp}}{10.5\ \text{bp}} \times 1\ \text{helical turn} = 8{,}710\ \text{helical turns}$$

DNA sizes and shapes

- How long is this DNA in microns?

  -The spacing between one base pair is about 3.4 Å or 3.4 × 10^{-4} μm along the helical axis.

Therefore 91,500 bp ?

$$\frac{91{,}500 \text{ bp}}{1 \text{ bp}} \text{ x } 3.4\text{x}10^{-4}\mu\text{m} = 31.11 \ \mu\text{m}$$

DNA sizes and shapes

- To convert bp to molecular weight

  - Multiply bp by 660 Daltons, the approximate molecular weight of one nucleotide pair

- What is the molecular mass of the phage P1 DNA?

  -One bp has a molecular weight of about 660 Daltons, therefore 91, 500 bp?

$$\frac{91{,}500\ \text{bp}}{1\ \text{bp}} \times 660\ \text{D} = 6.039 \times 10^{7}\ \text{D}$$

DNA sizes and shapes

- Circular DNA
- Supercoiled DNA

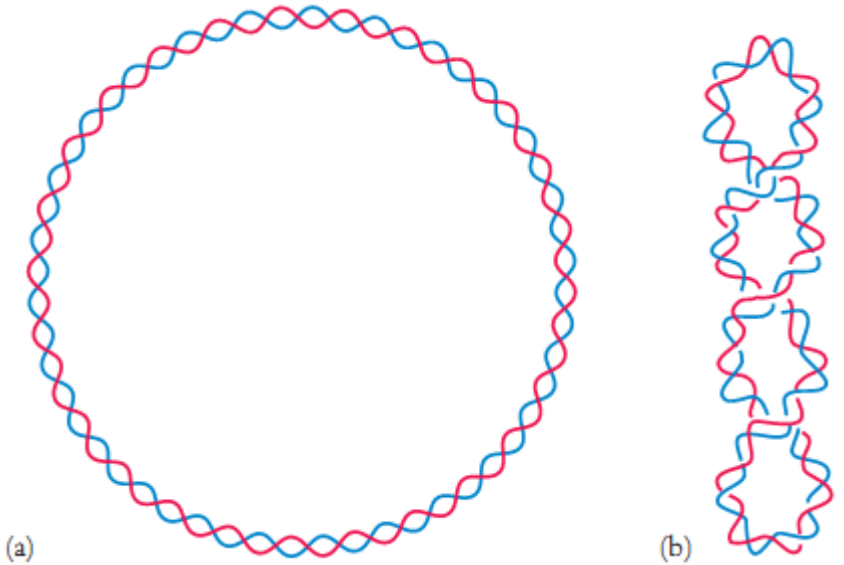
  DNA coils around itself like a twisted rubber band

DNA Forms

(a) The DNA double helix of almost all bacteria is in the shape of a closed circle.

(b) The circular DNA strands, already coiled in a double helix, are twisted a second time to produce supercoils. There are a few exceptions to the above picture.

The form of DNA influences their mobility on a gel

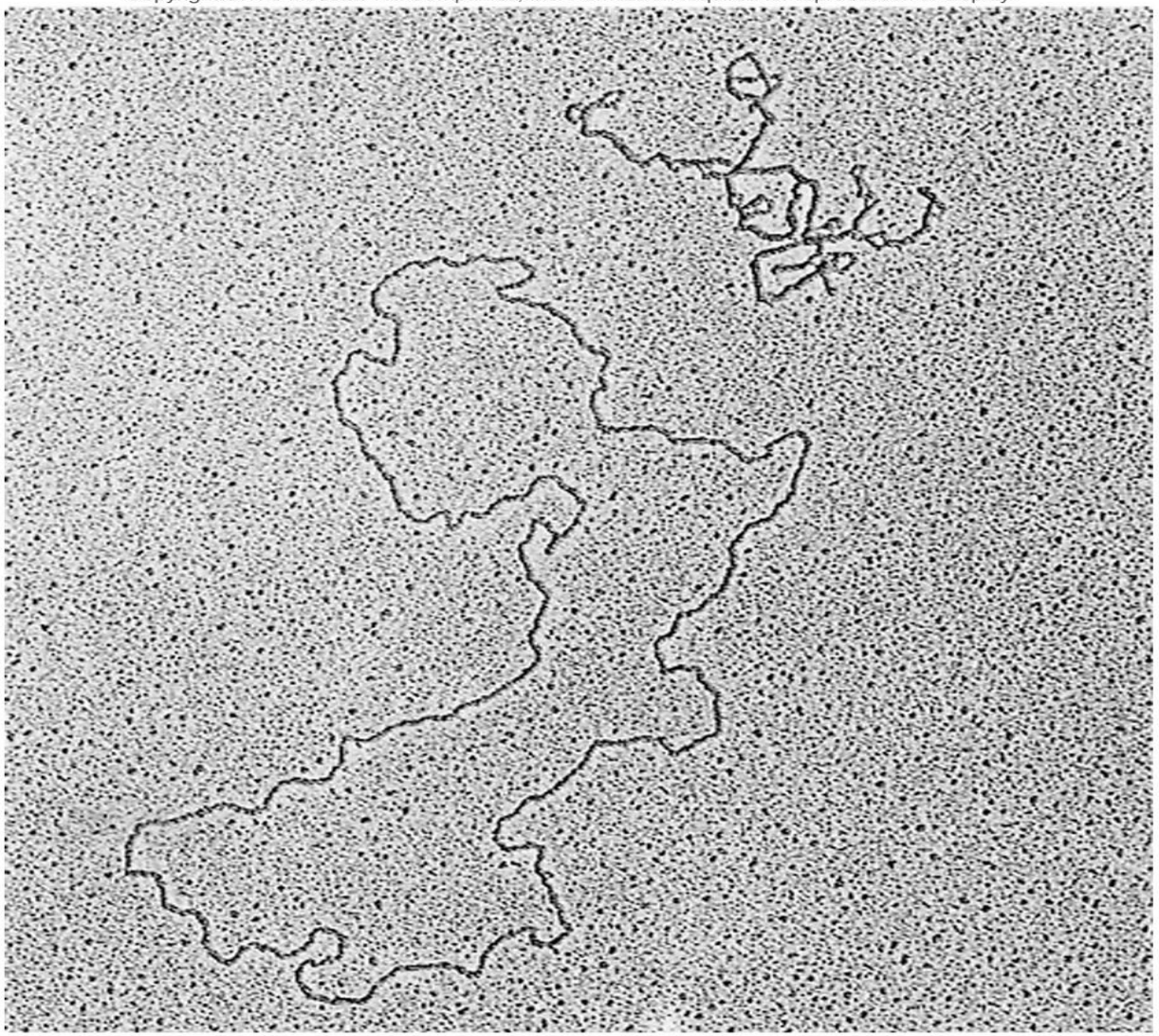


Supercoiling

- DNA can be twisted like a rope in a process called DNA supercoiling
- With DNA in its "relaxed" state, a strand usually circles the axis of the double helix once every 10.4 base pairs, but if the DNA is twisted the strands become more tightly or more loosely wound
- If the DNA is twisted in the direction of the helix, this is positive supercoiling, and the bases are held more tightly together
- If they are twisted in the opposite direction, this is negative supercoiling, and the bases come apart more easily.
- In nature, most DNA has slight negative supercoiling that is introduced by enzymes called **topoisomerases**
- These enzymes are also needed to relieve the twisting stresses introduced into DNA strands during processes such as transcription and DNA replication

Open circular DNA (left bottom); Supercolied form (upper right) of phage PM2

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



(Source: © Jack Griffith.)

Stability of DNA

- The structure of duplex DNA is governed by a balance of three forces:
- 1. **Noncovalent forces**
- 2. **Van der Waals dispersive forces**
- **3**. **Electrostatic effects**

Noncovalent forces

Two types: (1) stabilizing and (2) destabilizing

- Stabilizing (or favorable interactions): base stacking and hydrogen bonding

Stability of DNA- noncovalent forces

- Destabilizing: (1) Steric effects arising from

(a) electrostatic repulsion of one phosphate group from another along the same strand and repulsion of phosphates on different strands

(b) Repulsion of bases by size. Purine bases impose more repulsion than pyrimidine bases

- Variation in base pairing contribute to varying steric effects

Base stacking

- It is a noncovalent interaction controlled by several noncovalent forces.
- 1.“Aromatic stacking” refers to the geometry of face-to-face juxtaposition of two aromatic molecules such that the pi-electron cloud systems of the aromatic nucleus are in direct contact, though not directly aligned for purpose of maximizing surface area of contact
- In general, the distance between two aromatic planes is about. 3.4 Å in a stacked structure
- In the aqueous environment of the cell, the conjugated pi (π) bonds of nucleotide bases align perpendicular to the axis of the DNA molecule, minimizing their interaction with the aqueous environment. This character of DNA is spontaneous, hence the ΔG, Gibbs free energy, is small and negative

Stability of DNA

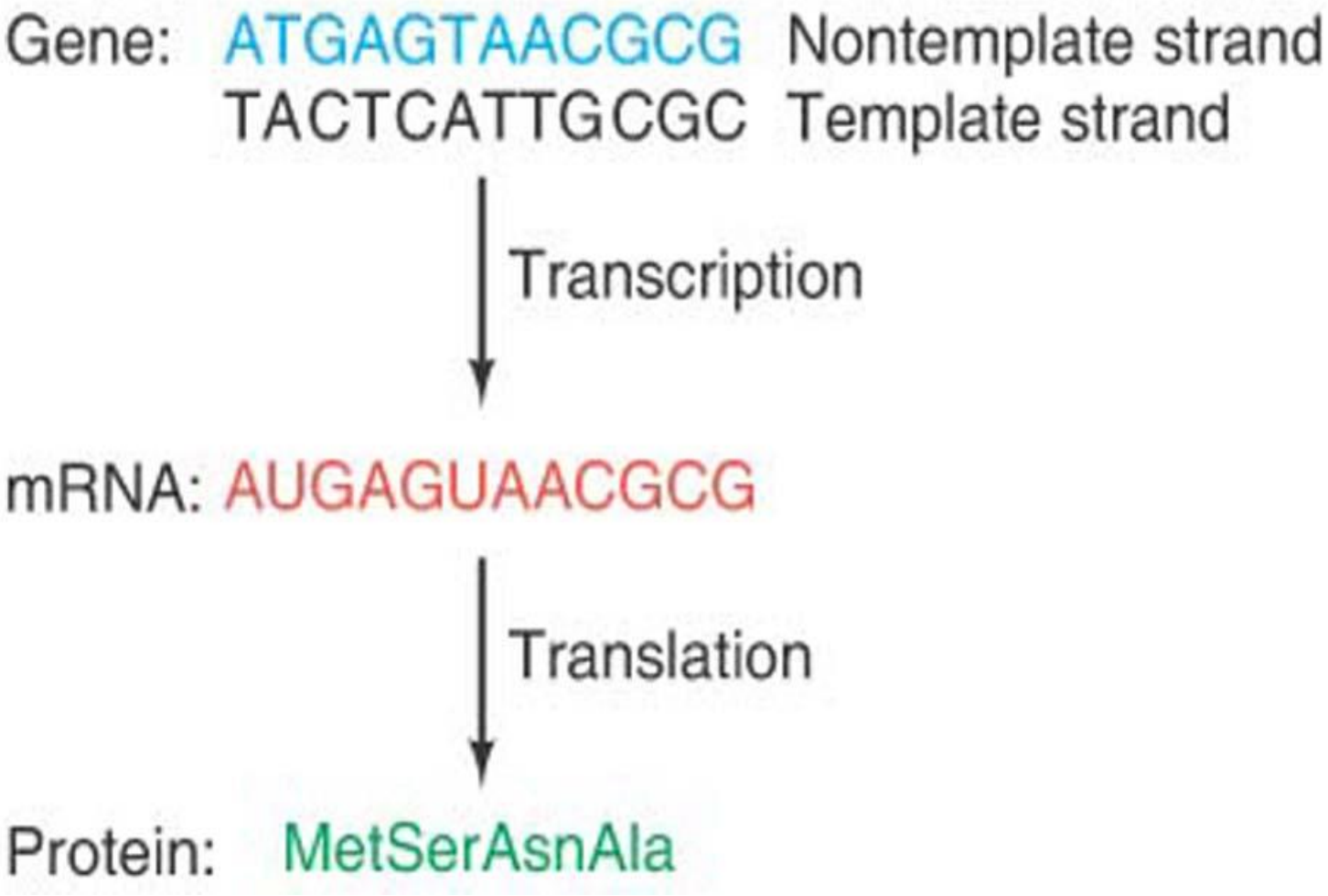
- 2.**Van der Waals dispersive forces**
- This is made up of dipole-induced dipole attractions which stabilize the stacking orientation
- 3. **Electrostatic effects** of interacting dipoles also influence stacking stability; this favorable or unfavorable contribution depends on the bond dipoles of the molecules

Protein structure

- Proteins are polymers, long, chain-like molecules
- The monomers are amino acids
- Informational relationship between DNA and protein: three nucleotides in a DNA gene stands for one amino acid in a protein – the triplet codon
- Proteins contain 20 different amino acids

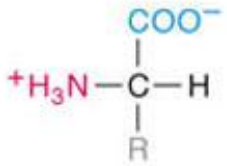
Outline of gene expression

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

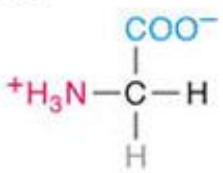


Structures of amino acids

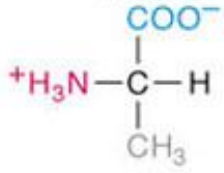
Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



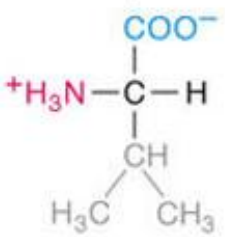
(a)



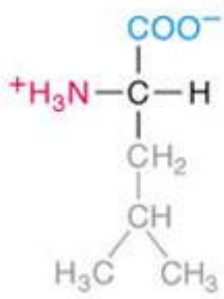
Glycine
(Gly; G)



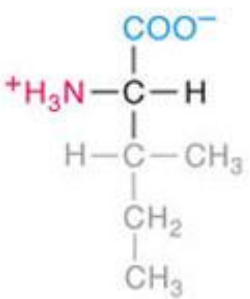
Alanine
(Ala; A)



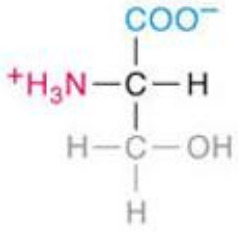
Valine
(Val; V)



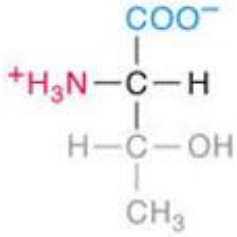
Leucine
(Leu; L)



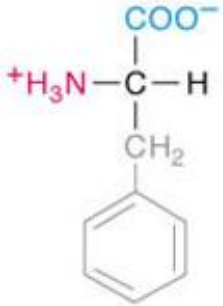
Isoleucine
(Ile; I)



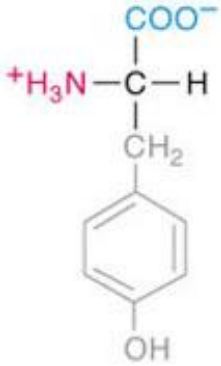
Serine
(Ser; S)



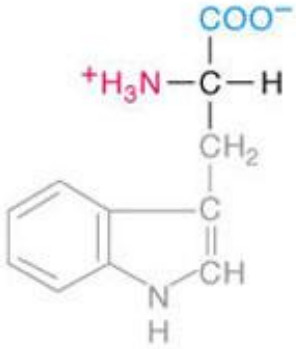
Threonine
(Thr; T)



Phenylalanine
(Phe; F)

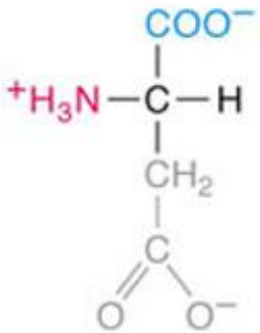


Tyrosine
(Tyr; Y)

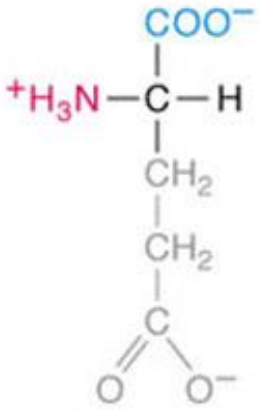


Tryptophan
(Trp; W)

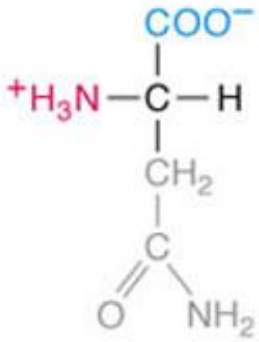
Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



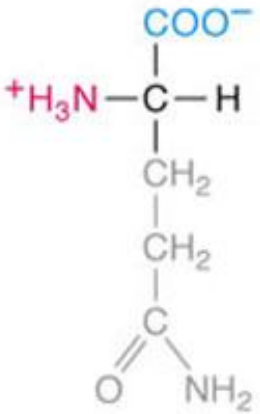
Aspartate
(Asp; D)



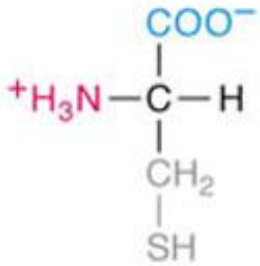
Glutamate
(Glu; E)



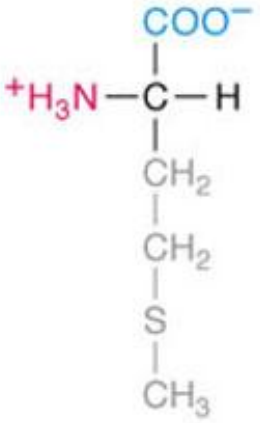
Asparagine
(Asn; N)



Glutamine
(Gln; Q)

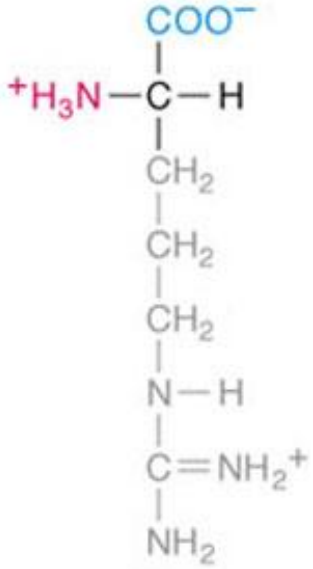


Cysteine
(Cys; C)

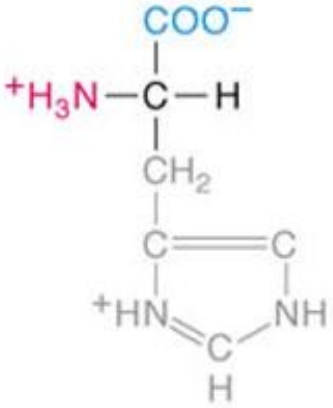


Methionine
(Met; M)

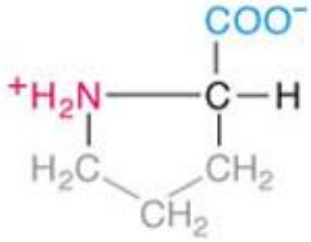
Lysine
(Lys: K)



Arginine
(Arg: R)



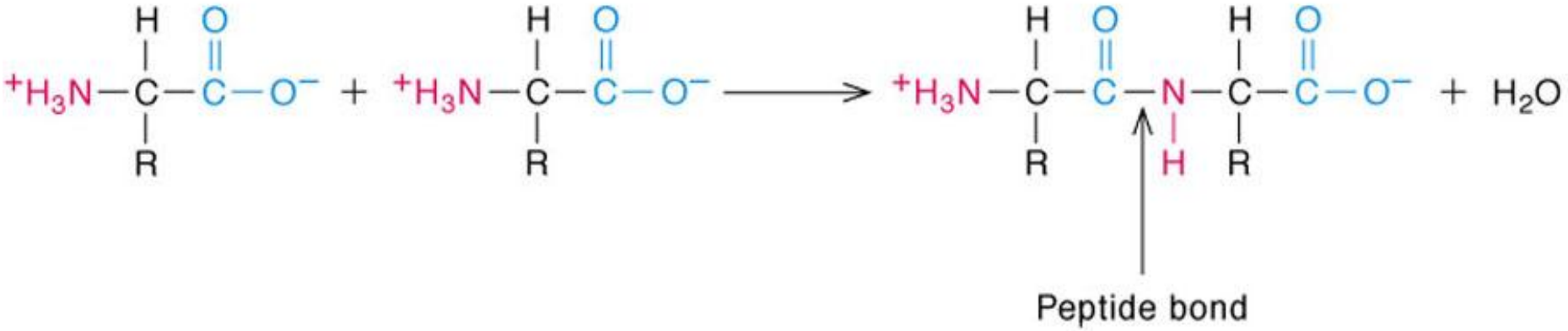
Histidine
(His; H)



Proline
(Pro; P)

(b)

Copyright © The MuGraw-Hill Companies, Inc. Permission required for reproduction or display.



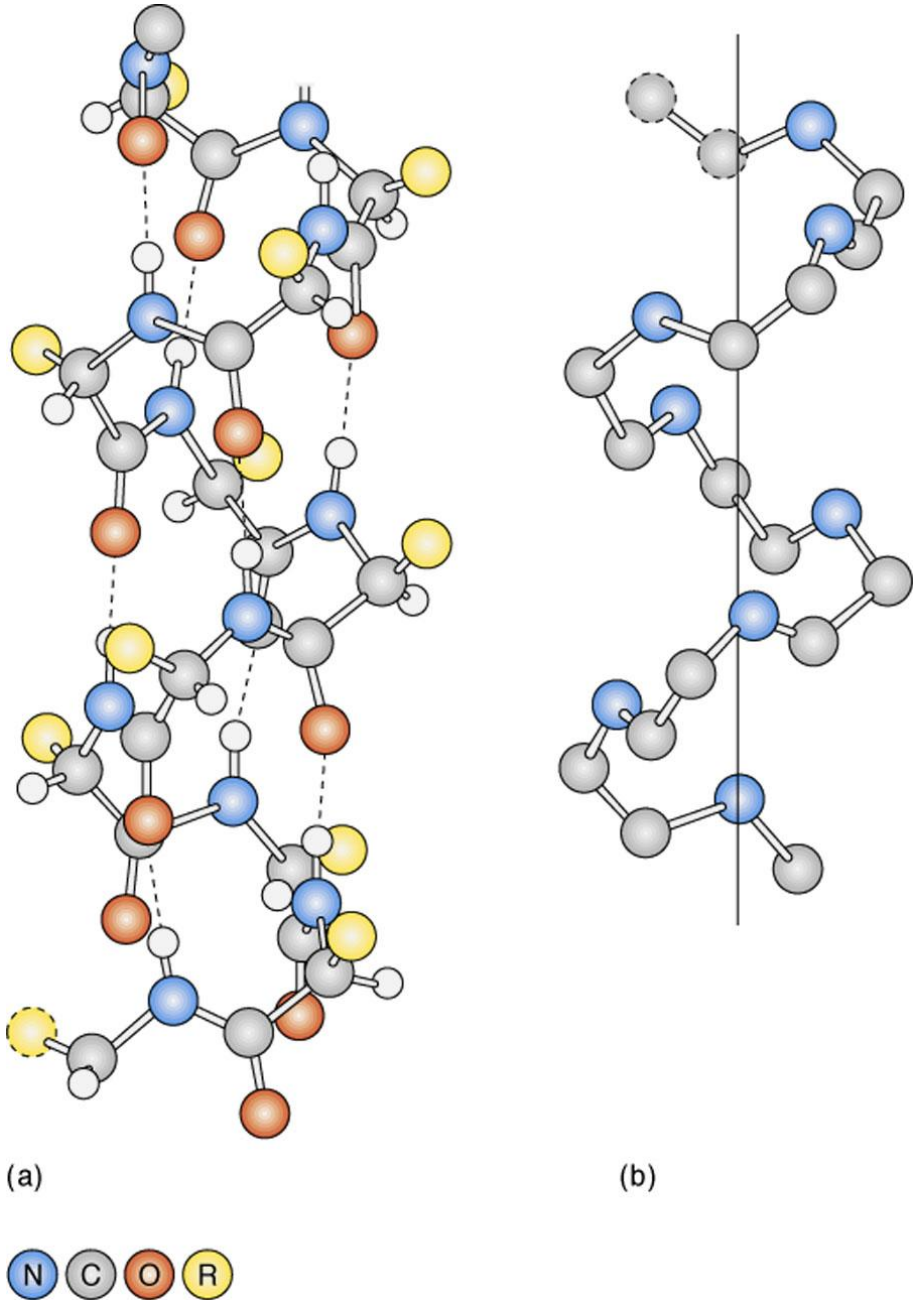
Protein structure

- The amino acid sequence of a protein is the primary structure
- Interaction of amino acids with their neighbours gives rise to secondary structure
  - α-helix: results from H- bonding among nearby amino acids
  - β-Pleated sheet: extended protein chains packed side by side and interact by H- bonding
- The three dimensional shape of a polypeptide is its tertiary structure: involves twisting and folding of the chains

Protein folding

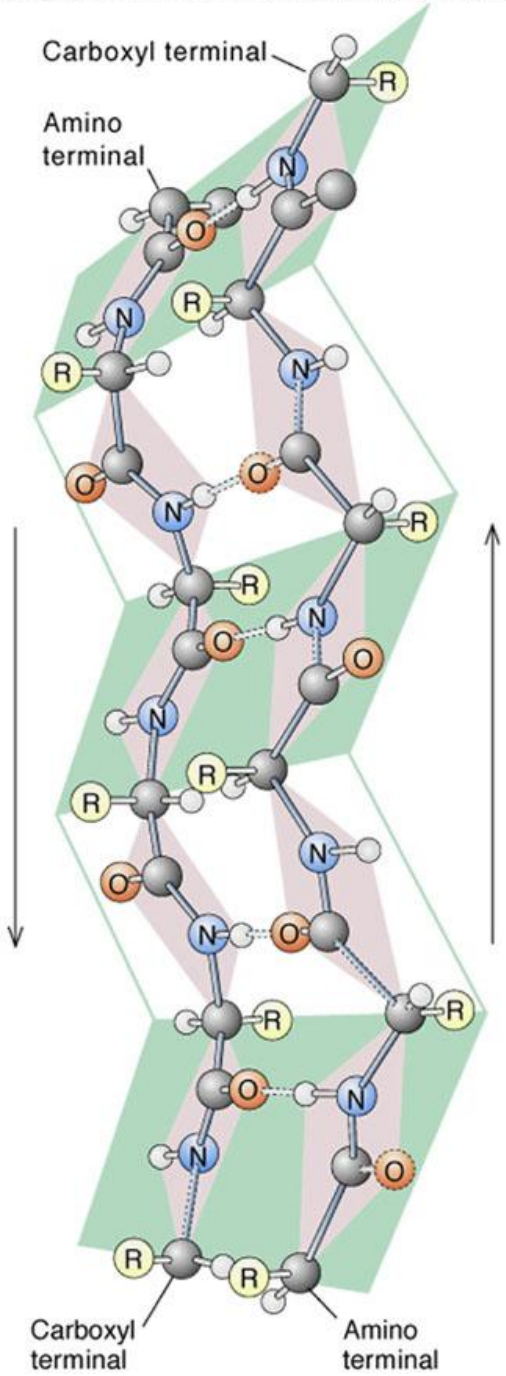
- Protein folding is governed by the distribution of its amino acid sequence: both polar and nonpolar
- There are three weak noncovalent bonding systems, whose combined effect creates a strong bonding within the polypeptide chain – hydrogen-bonding, ionic interactions, van der Waals interaction
- A hydrophobic interaction also determines shape of the protein

α-helix



β-pleated sheet

Copyright © The McGraw-Hill companies, Inc. Permission required for reproduction or display.



Copyright © The McGraw-Hill companies, Inc. Permission required for reproduction or display.

Tertiary structure