MACQUARIE University

Amino Acids 2: Physical Properties and Making Proteins from Genes

Shoba Ranganathan

Applied Biosciences

T: 02 9850 6262; E: shoba.ranganathan@mq.edu.au

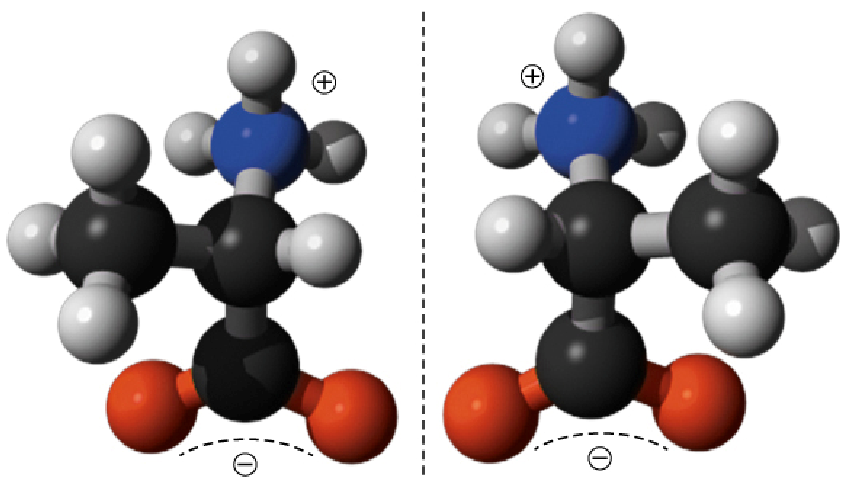
Objectives

- Chirality (‘handedness’) in amino acids
- Absorption of light
- Covalent modification of amino acid side chains
- Biological function of amino acids or their derivatives
- How are proteins assembled from amino acids?

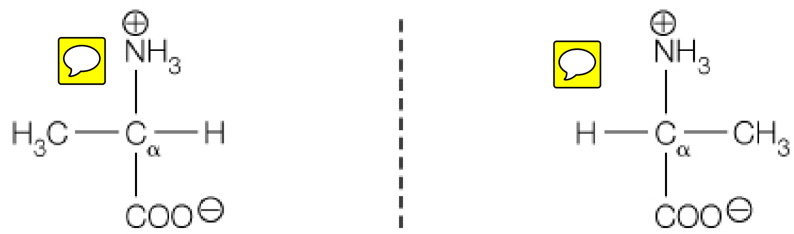
AAM: Chapters 5, 6, 4 and 1

α-Amino Acid Stereochemisty - 2

- All α-amino acids (20), except glycine, contain an asymmetric α-carbon
- When R = $-CH_3$ (methyl group), we have the amino acid alanine
- L-alanine is the mirror image of D-alanine; they are enantiomers



(a) L-Alanine and its enantiomer D-alanine are shown as ball-and-stick models. The alanine side chain is $-CH_3$. The two models are nonsuperimposable mirror images. The mirror plane is represented by the vertical dashed line.

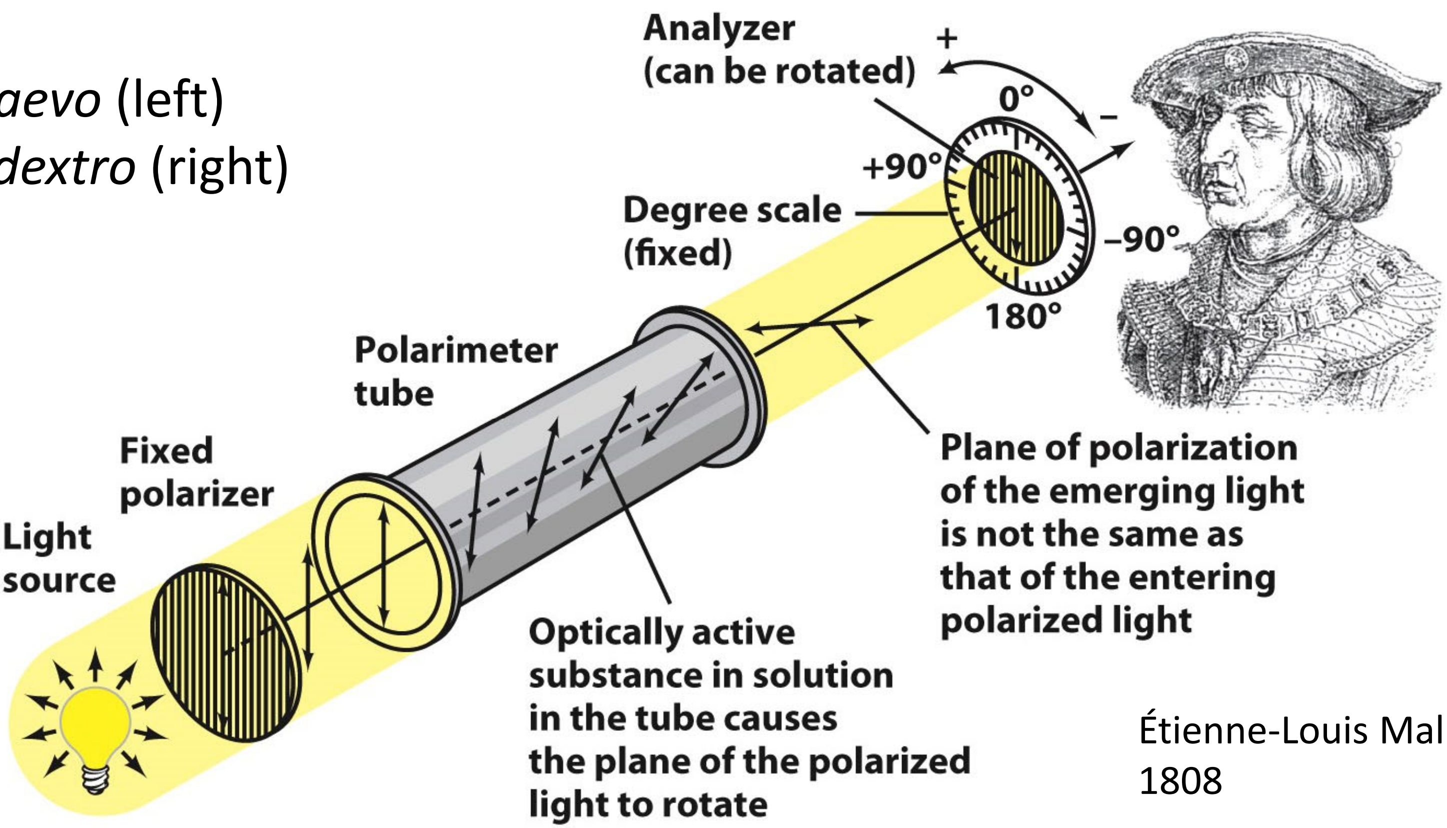


(b) The same two enantiomers in Fischer projections.

FIGURE 5.5 Stereoisomers of *α*-amino acids.

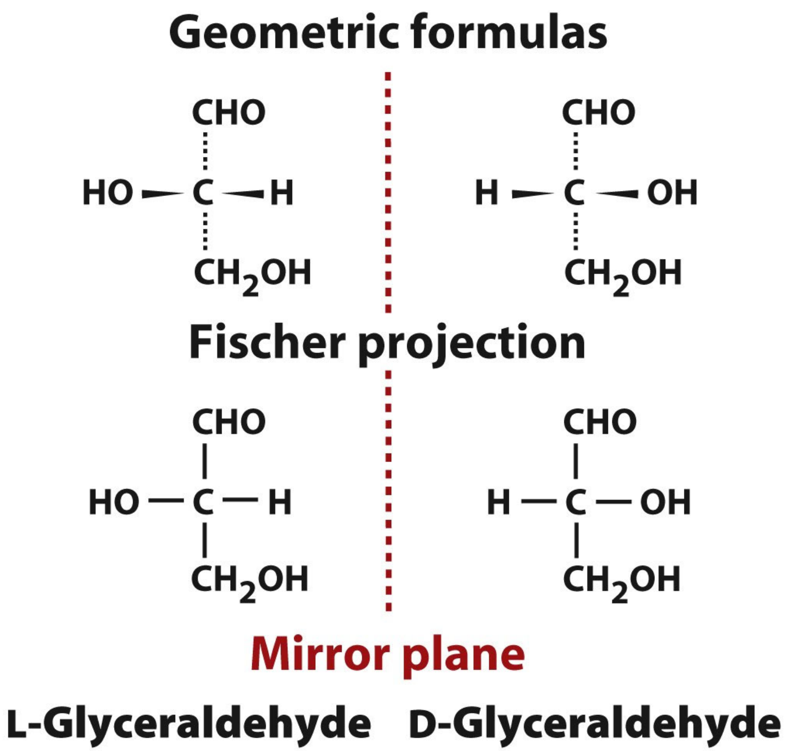
Chiral molecules rotate polarized light

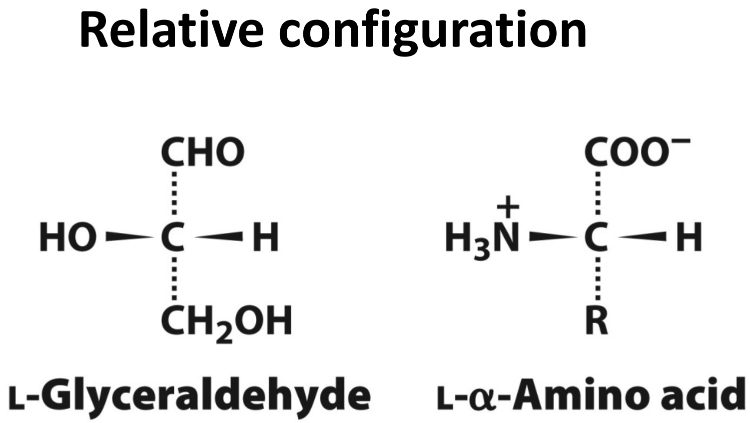
L: *laevo* (left)
D: *dextro* (right)



Étienne-Louis Malus, 1808

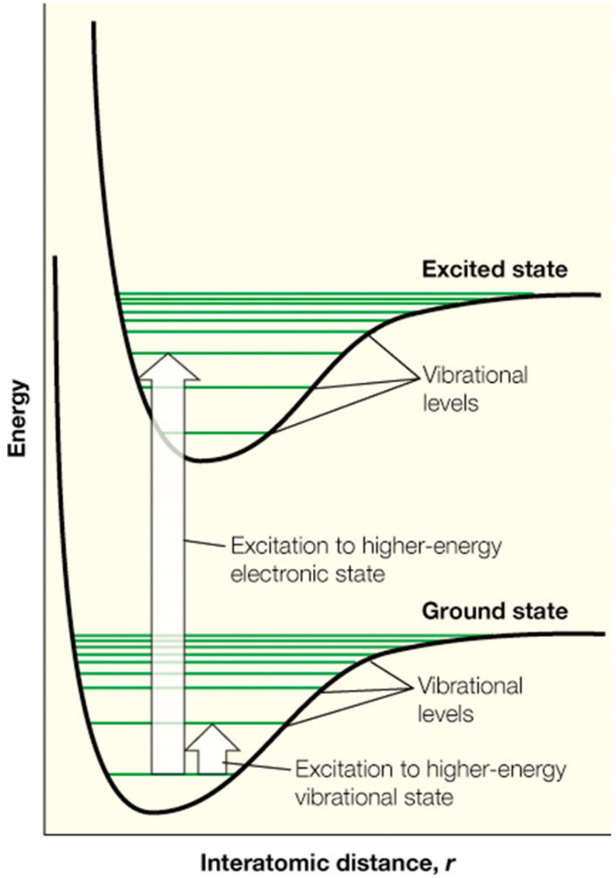
Humans have all L-amino acids



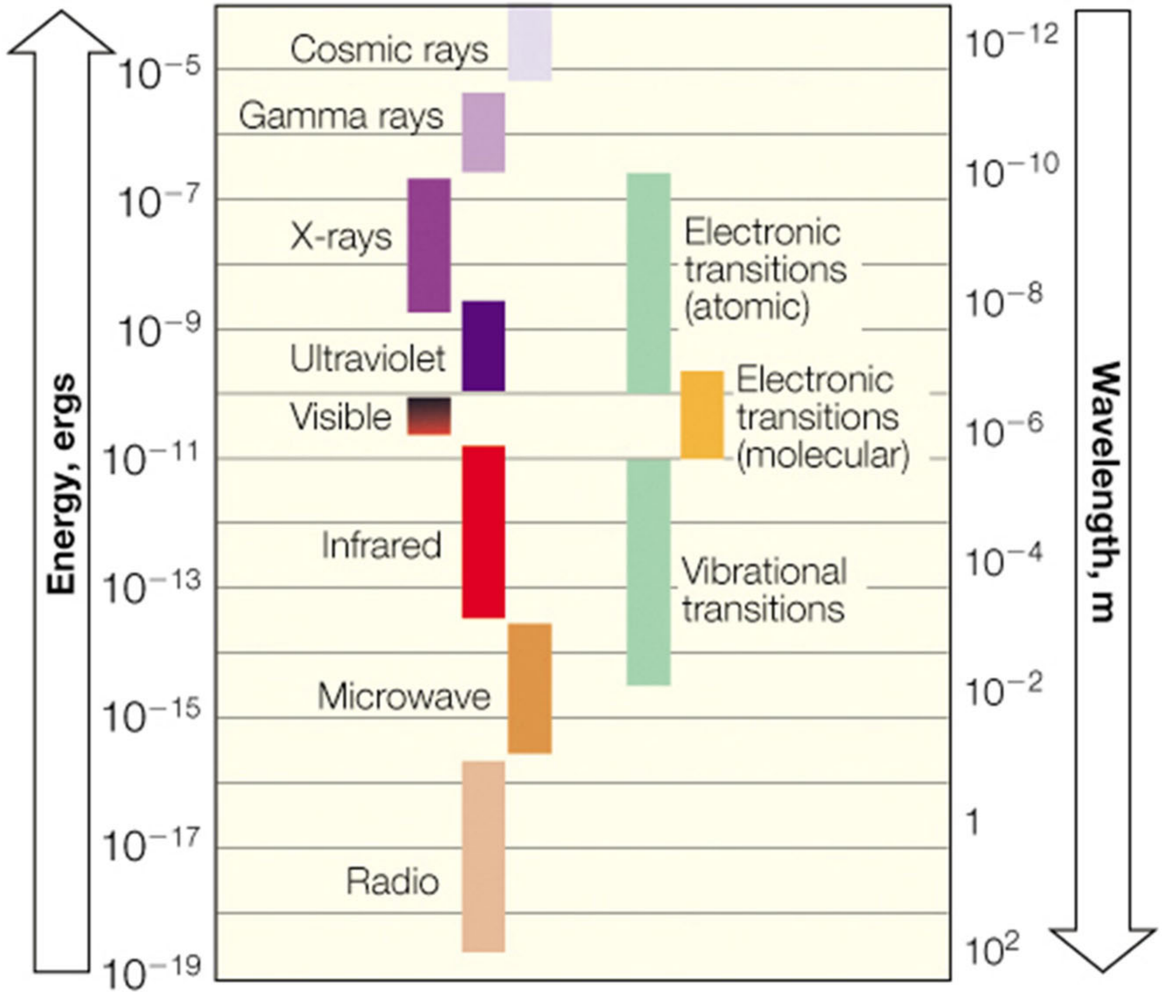


- Organic synthesis invariably leads to a mixture of D- and L- isomers (racemic mixture)
  - ➢ Ibuprofen is marketed as a mixture, although the D-isomer (*S*-) has greater activity!
- However, biochemical reactions result in 100% of one stereo isomer!

Amino acids absorb light



(a) Electronic and vibrational transitions between allowed states in a diatomic molecule.



(b) The electromagnetic spectrum.

Figure 6A. The principles of absorption spectroscopy

UV absorption spectra of tyrosine and tryptophan

- Proteins can be quantified by examining the UV absorbance at 280 nm, which is dominated by tyrosine (Tyr) and tryptophan (Trp)
- In comparison, nucleic acids absorb most strongly at 260 nm

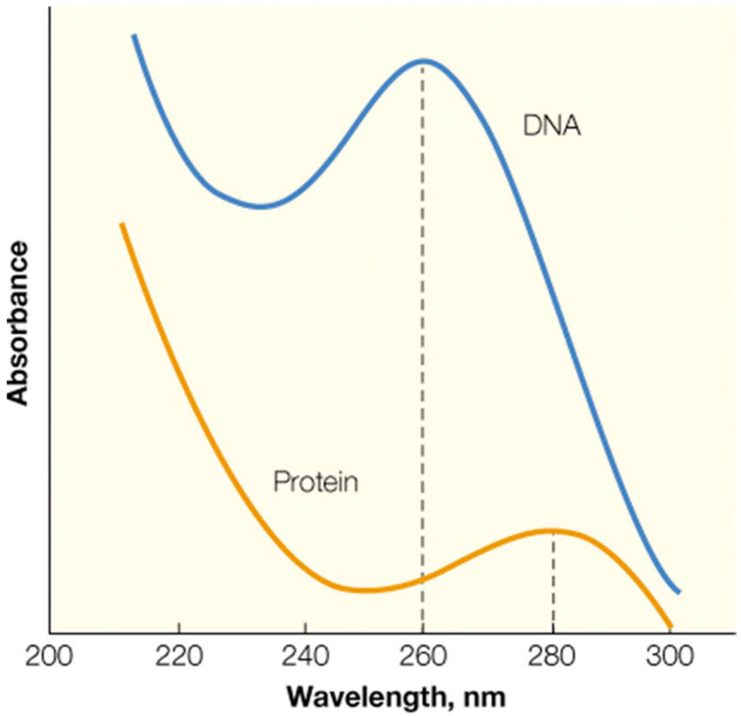
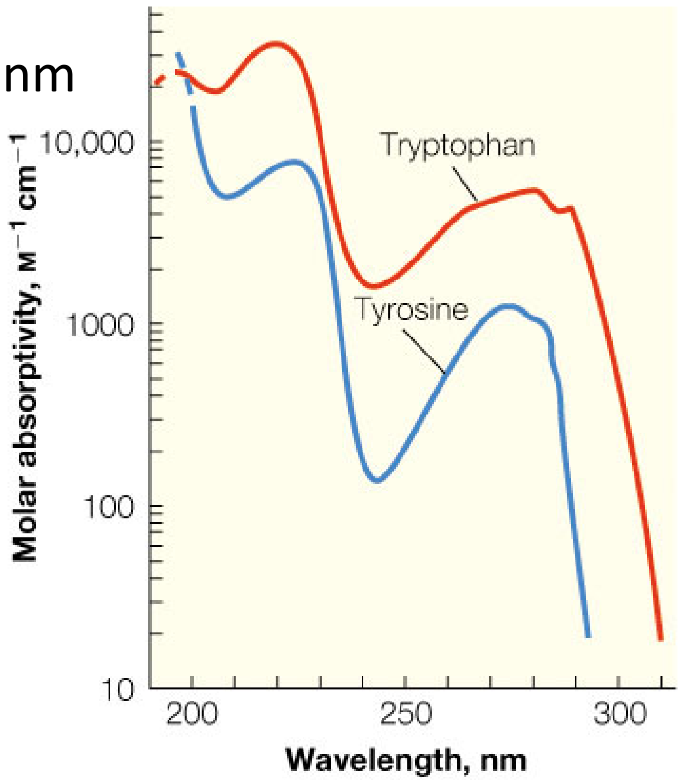


FIGURE 5.6 Absorption spectra of two aromatic amino acids in the near-ultraviolet region.

Figure 6A.2, The principles of absorption spectroscopy.



Spectrophotometer for measuring light absorption

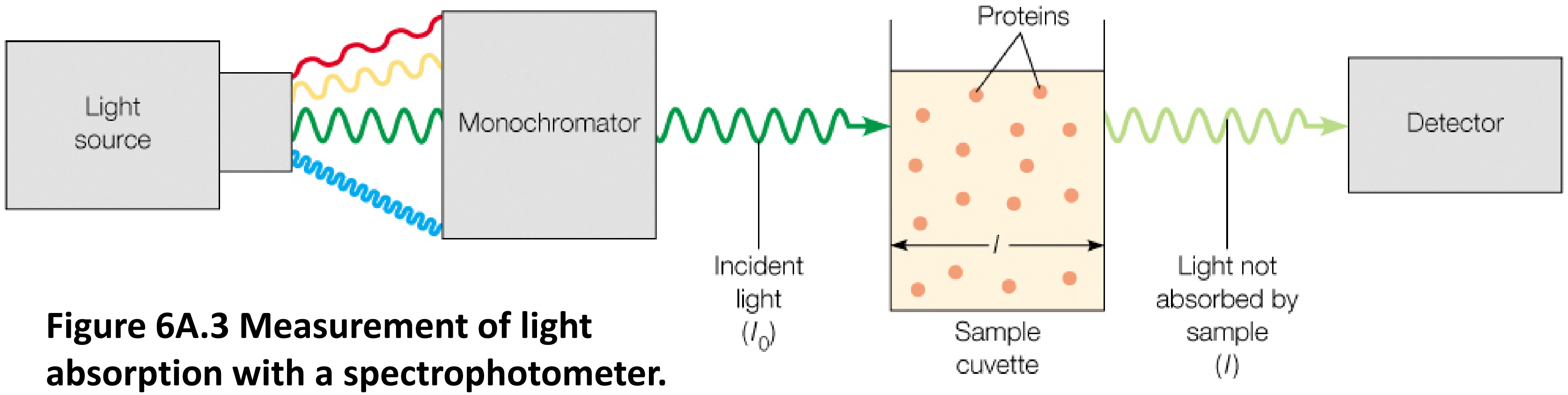
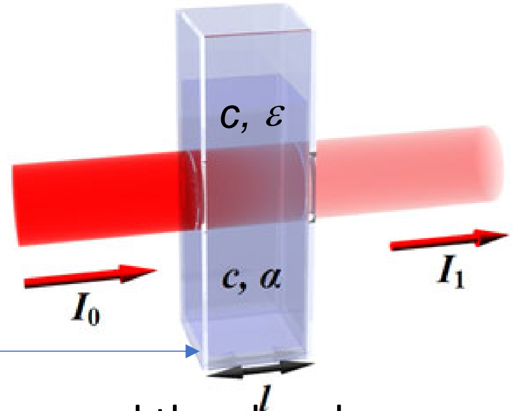


Figure 6A.3 Measurement of light absorption with a spectrophotometer.

- We will mainly be dealing with absorbance of light in the **visible region** (800 nm - 350 nm) and the **upper end of the ultraviolet region** (350 nm – 250 nm) of the spectrum
  - Many biological molecules absorb light in this range
  - It is thus a useful tool in examining biochemical processes
  - Instrumentation is relatively easy to use and the sample is placed in a cuvette
- Applications
  - identification by characteristic frequency of absorbance
  - amount of protein concentration present e.g. creatine (a protein) in urine.

Beer-Lambert's Law
(aka Lambert-Beer's or Beer's Law)



$\log I_o/I$ = $\varepsilon \circ c$ = A (what we measure, in a cuvette)

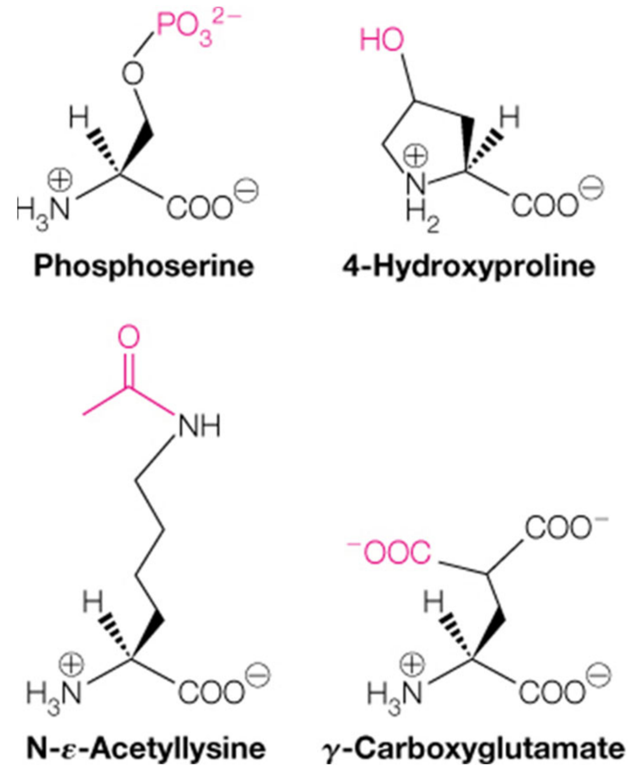
- where
  - **I/I_o** is the fraction of light absorbed - this is also called <u>transmittance</u> and the absorbance **A = -log T = - log (I/I_o)**
  - **e (Greek epsilon)** is the molar extinction coefficient, which is a measure of absorption in **M^{-1} cm^{-1}** or **L mol^{-1} cm^{-1}**
  - ○is the pathlength of light passing through the sample in cm (usually 1 cm)
  - **c** is the concentration in molar units: **M** or **mol L^{-1}**
  - **A** is called the absorbance (ratio – so no units)
  - Valid for dilute solutions

❖ Light absorption is used to **quantitatively** estimate the **amount of protein** in a biological sample (**Prac 2**).

❖ This property can also be used to **identify** amino acids (and proteins) and nucleotide bases in a biological sample (**Prac 4**).

Protein side chains can be modified

- After synthesis (i.e. post-translational), some protein side chains can be modified usually by simple chemical reactions:
  - ➢Hydroxylation, methylation, acetylation, phosphorylation, carboxylation, glycosylation
  - ➢N- and C- ends can also be modified or blocked.
- Some modifications are essential for biological function.
  - ➢signaling pathways, calcium binding, stabilizing structures such as collagen, or play roles in gene expression or suppression
- Many side chains can be modified in a protein.



N- and C-termini blocking and Disulfide bridges

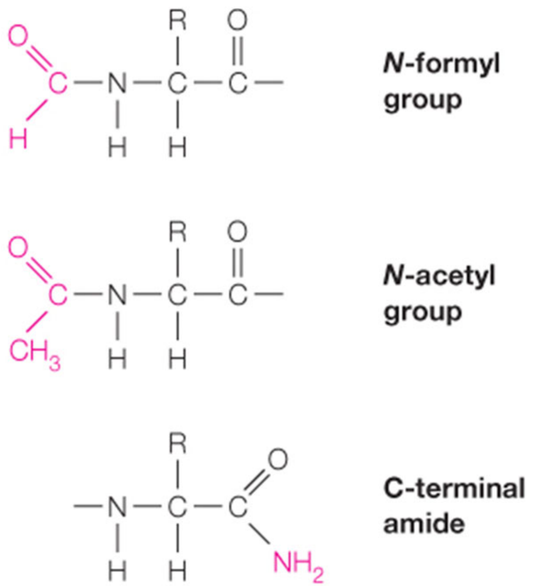
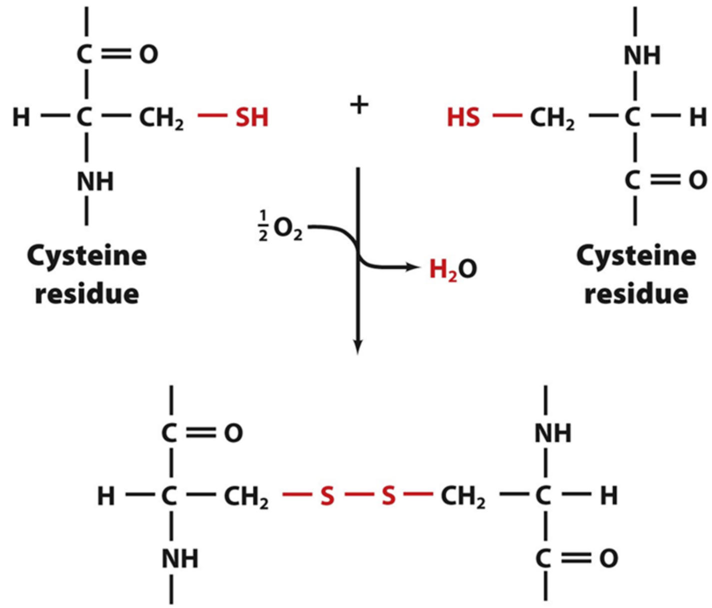


FIGURE 5.12 Groups that may block N- or C-termini in proteins.

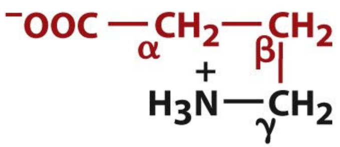
Side chains of **cysteine residues** can get oxidized to form a disulphide(-S-S-) bond.

- Link different parts of the same protein or even different protein chains.
- Increase protein stability.

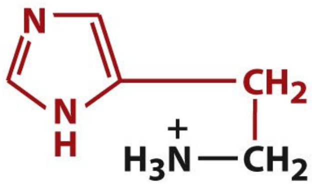


Some amino acids are biologically active

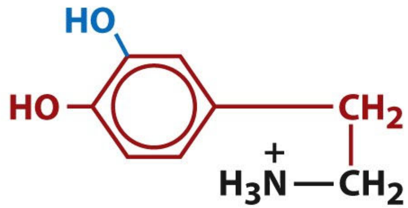
- Some amino acids transport nitrogen while others can provide energy as a fuel source (amino acid metabolism: later lectures)
- Some function as chemical messengers for cellular communication: e.g. **glycine and glutamate** are neurotransmitters
- Example **biologically active amino acid derivates** are shown below:



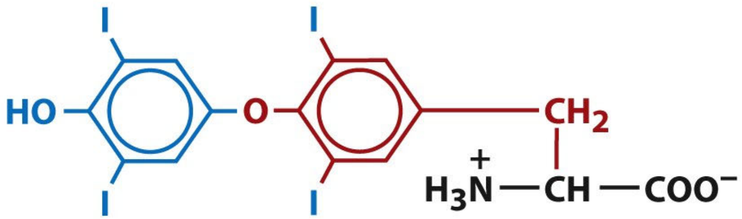
γ-Aminobutyric acid (GABA)
Neurotransmitter



Histamine
allergy response



Dopamine
neurotransmitter



Thyroxine thyroid hormone

Peptide Bond Formation between Amino Acids

- The **condensation** of two amino acids forms a **peptide bond** and releases ***water***
  - ➢The charges in the middle are lost!
  - ➢The ends are still charged.
- This reaction des not occur spontaneously, and is coupled to ATP hydrolysis during protein biosynthesis
  - ➢Can be extended from the carboxyl end…..

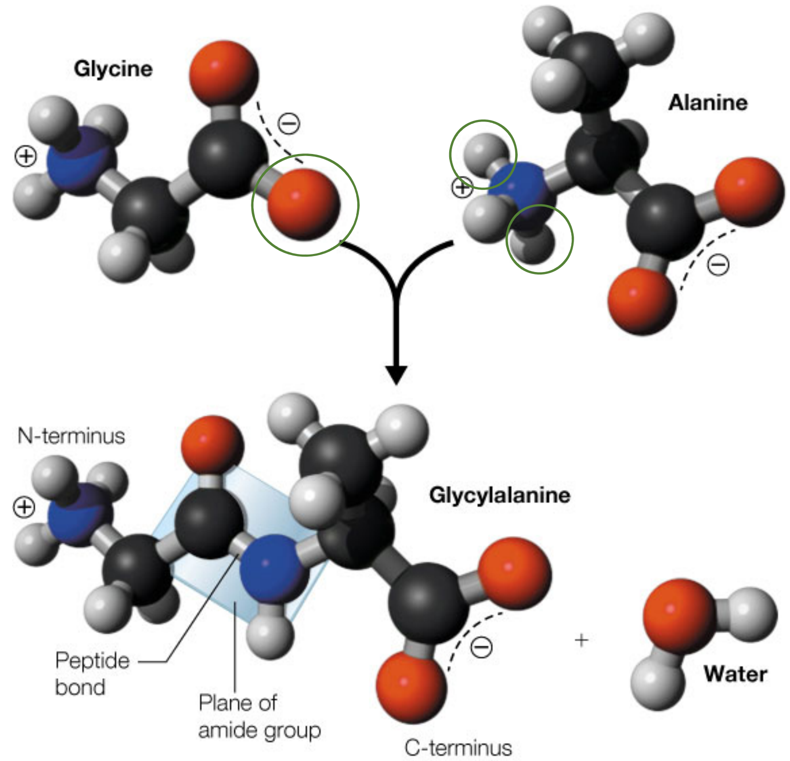
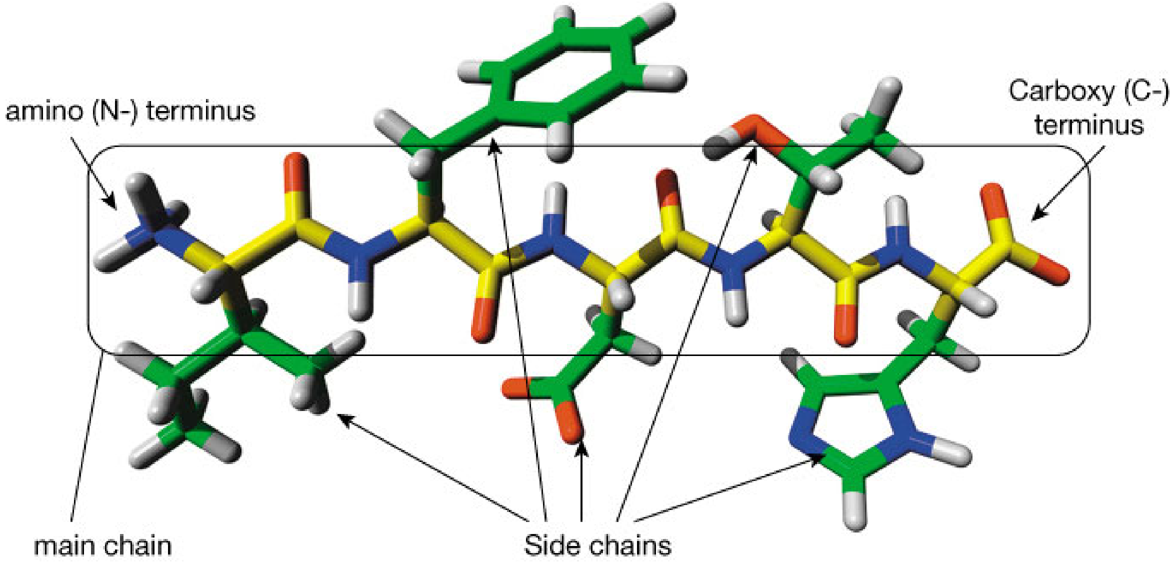


FIGURE 5.9 A peptide bond between amino acids.

Example peptides

- Important parts of a peptide, IFDTH
- The main chain is also called the backbone
- The ends are normally charged
- Amino acids in a peptide or protein are called **residues**



Direction of peptide sequence (N- to C-)

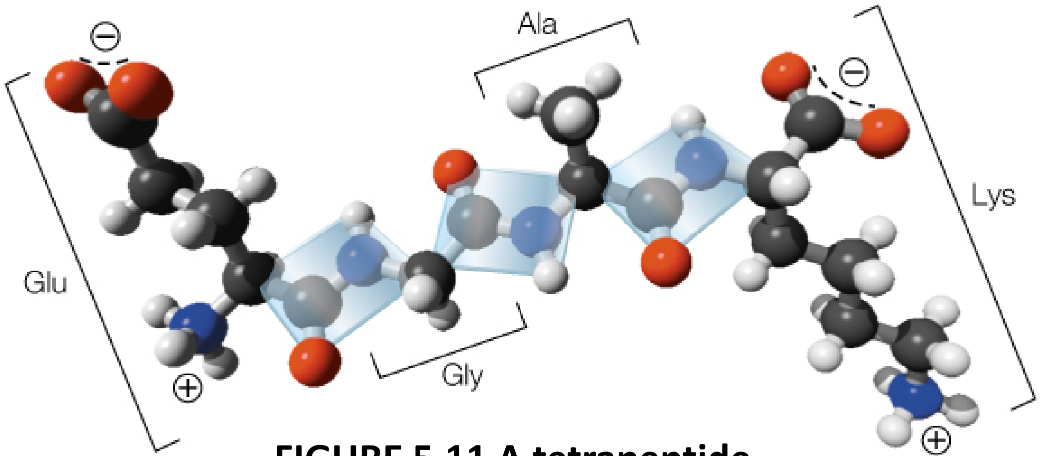
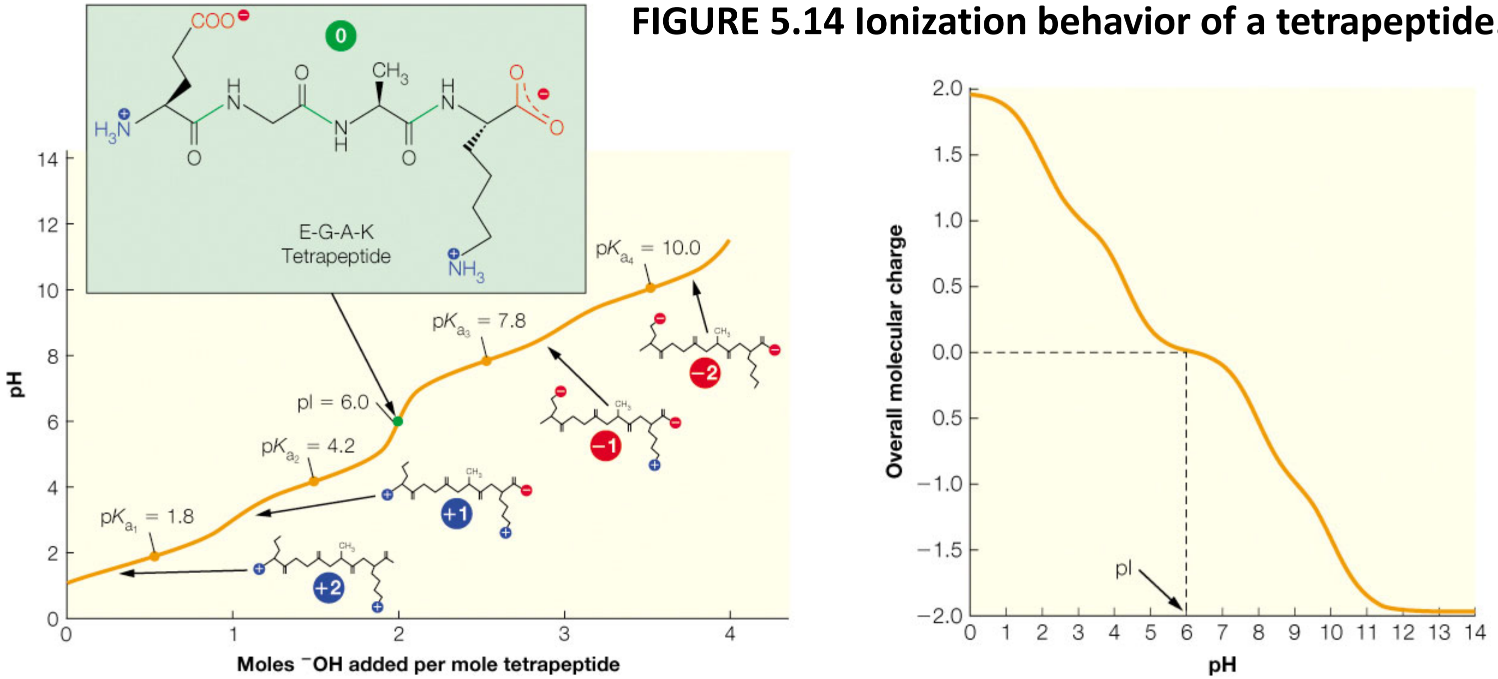


FIGURE 5.11 A tetrapeptide.

- Tetrapeptide:
- Glu-Gly-Ala-Lys
- EGAK

Peptides and Proteins as Polyampholytes

FIGURE 5.14 Ionization behavior of a tetrapeptide.



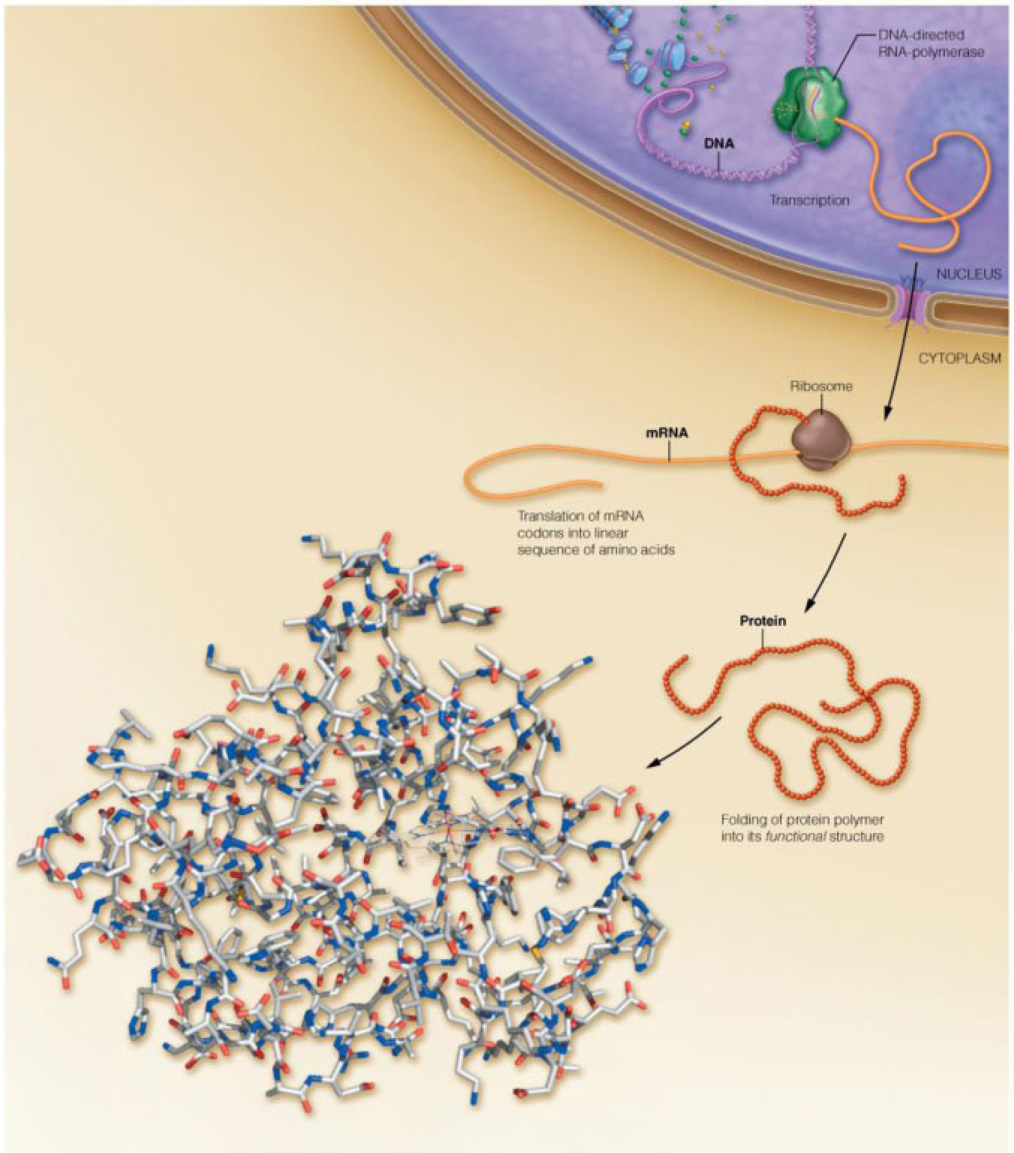
(a) This titration curve for the tetrapeptide Glu-Gly-Ala-Lys shows the major ionization states present as a function of pH. The tetrapeptide is shown schematically with (+) charges in blue and (−) charges in red. Net charges for the different ionization states are shown in solid circles.

(b) The net charge on Glu-Gly-Ala-Lys as a function of pH is not a step function. Rather, between pH 1 and 11, it changes gradually and smoothly from +2 to −2. At values of pH > pI, the overall charge on the tetrapeptide is negative. At values of pH < pI, the net charge is positive.

As pH increases the overall charge on a peptide will become more negative; as pH decreases it will become more positive.

Amino acids are the building blocks of proteins

- DNA is transcribed to form messenger RNA (mRNA) in the nucleus
- mRNA is exported to the cytoplasm, where it is bound to the ribosome
- mRNA is translated into a linear sequence of **amino acids** that folds into a 3D structure



From amino acid monomers to polymers

Peptides

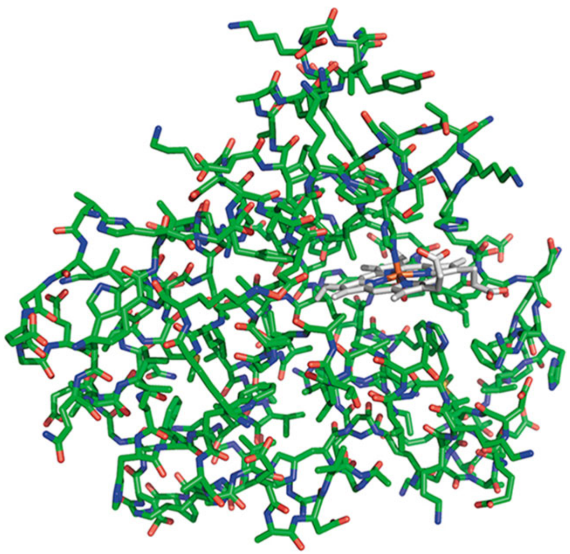
- Short polymers of amino acids (aa)
- Each unit is called a residue
- 2 residues - dipeptide
- 3 residues - tripeptide
- 12-20 residues - oligopeptide
- many - polypeptide

Proteins

- 1 or more chains of at least 40 aa
- One polypeptide chain - a monomeric protein
- More than one - multimeric protein
- Homomultimer - one kind of chain
- Heteromultimer - two or more different chains
- Hemoglobin, for example, is a heterotetramer, with α (alpha) chains and β (beta) chains

Proteins: polypeptides with Defined Sequences

- Every protein has a defined number and order of amino acids.
- This is called its "primary structure"
- The sequence of the sperm whale myoglobin determines its final biological form (folded) and its function



- Its similarity to the human myoglobin sequence is remarkable!
- Of 153 amino acids, 128 (84% are identical)
- 13 residues (green) have chemically similar side chains – adding these, 92% are similar!

Identical amino acids | Conservative substitutions | Nonconservative substitutions

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Human	G	L	S	D	G	E	W	Q	L	V	L	N	V	W	G
Whale	V	L	S	E	G	E	W	Q	L	V	L	H	V	W	A

Number	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Human	K	V	E	A	D	I	P	G	H	G	Q	E	V	L	I
Whale	K	V	E	A	D	V	A	G	H	G	Q	D	I	L	I

Number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Human	R	L	F	K	G	H	P	E	T	L	E	K	F	D	K
Whale	R	L	F	K	S	H	P	E	T	L	E	K	F	D	R

Number	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Human	F	K	H	L	K	S	E	D	E	M	K	A	S	E	D
Whale	F	K	H	L	K	T	E	A	E	M	K	A	S	E	D

Number	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Human	L	K	K	H	G	A	T	V	L	T	A	L	G	G	I
Whale	L	K	K	H	G	V	T	V	L	T	A	L	G	A	I

Number	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Human	L	K	K	K	G	H	H	E	A	E	I	K	P	L	A
Whale	L	K	K	K	G	H	H	E	A	E	L	K	P	L	A

Number	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
Human	Q	S	H	A	T	K	H	K	I	P	V	K	Y	L	E
Whale	Q	S	H	A	T	K	H	K	I	P	I	K	Y	L	E

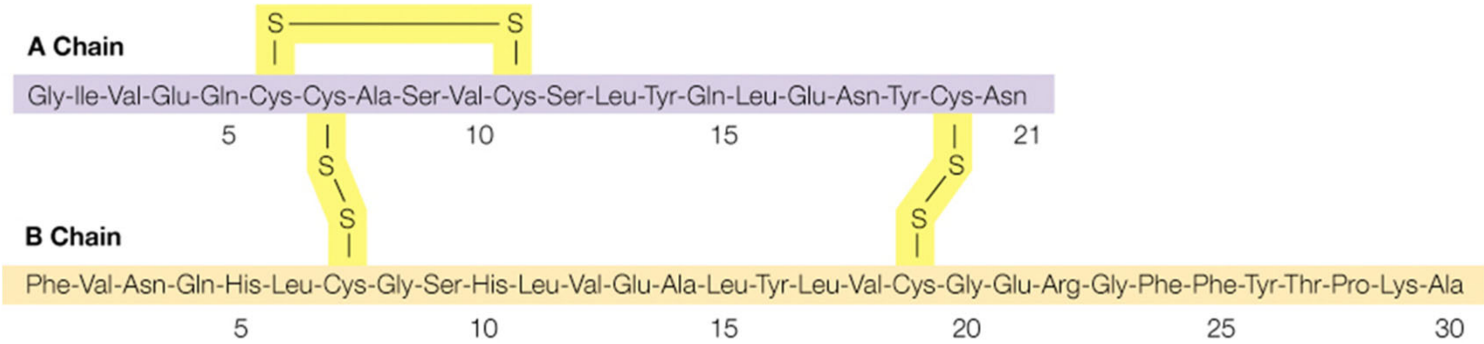
Number	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
Human	F	I	S	E	C	I	I	Q	V	L	Q	S	K	H	P
Whale	F	I	S	E	A	I	I	H	V	L	H	S	R	H	P

Number	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
Human	G	D	F	G	A	D	A	Q	G	A	M	N	K	A	L
Whale	G	N	F	G	A	D	A	Q	G	A	M	N	K	A	L

Number	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
Human	E	L	F	R	K	D	M	A	S	N	Y	K	E	L	G	F	Q	G
Whale	E	L	F	R	K	D	I	A	A	K	Y	K	E	L	G	Y	Q	G

How small or large are proteins?

- Insulin (bovine) – a small protein with an A chain of 21 residues and a B chain of 30 residues - total mol. wt. of 5,733



- Largest known protein is **titin** in muscle fibres: 34,350 residues (MW 3816 kD)
- Majority of proteins contain 100-1000 aa.
- Multisubunit proteins: many identical or non-identical polypeptide chains called subunits
  - ➢ Insulin has two chains.
  - ➢ There are two disulfide bonds holding the chains together

From Genes to Proteins

DNA
(recipe book at the library)

↓

mRNA
(quick scribbled-down copy of recipe)

↓

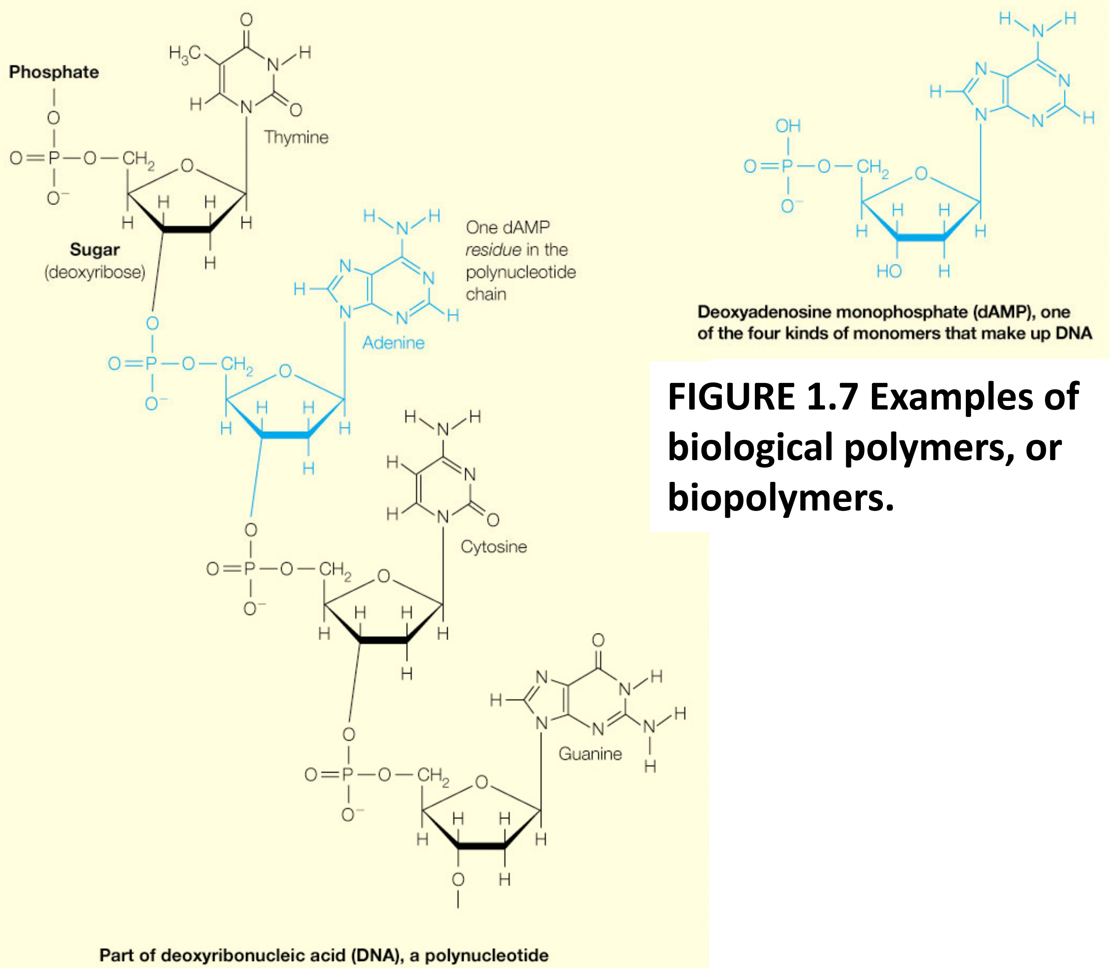
Protein
(result!)

Base in second position

Base in first position (5′ end)	U	C	A	G	Base in third position (3′ end)
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr UAA: Stop UAG: Stop	UGU, UGC: Cys UGA: Stop UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G
A	AUU, AUC, AUA: Ile AUG: Met/start	ACU, ACC, ACA, ACG: Thr	AAU, AAC: Asn AAA, AAG: Lys	AGU, AGC: Ser AGA, AGG: Arg	U C A G
G	GUU, GUC, GUA, GUG: Val	GCU, GCC, GCA, GCG: Ala	GAU, GAC: Asp GAA, GAG: Glu	GGU, GGC, GGA, GGG: Gly	U C A G

- The standard genetic code shows the **codons**, or **base triplets**, that correspond to each amino acid residue
  - A single amino acid can be encoded by 1-6 codons!
- The **AUG** start codon places a **Met** at the beginning of **eukaryotic proteins**,
- Three codons, **UAA**, **UGA**, and **UAG**, are stop codons, which signal termination of translation

Nucleic Acids make polymeric DNA



(b) A nucleic acid. The nucleic acids, DNA and RNA, are polymers of nucleotides. Part of a DNA molecule is shown, along with one of its monomers, deoxyadenosine monophosphate.

FIGURE 1.7 Examples of biological polymers, or biopolymers.

MACQUARIE University

Nucleotide bases in nucleic acids

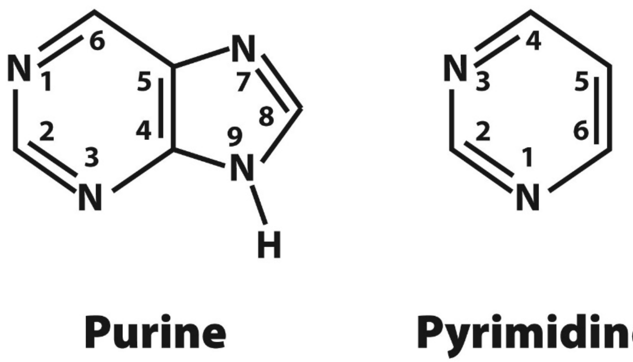
2 purines and 3 pyrimidines

Purines

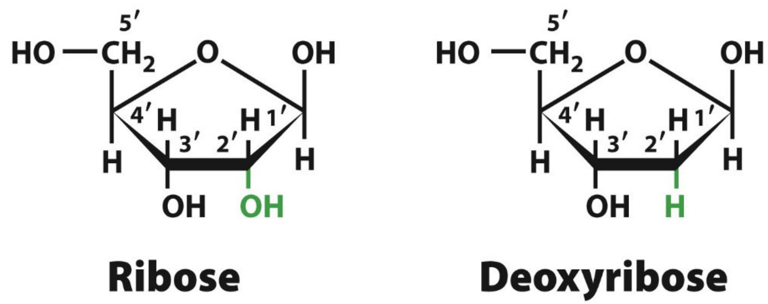
- A: Adenine
- G: Guanine

Pyrimidines

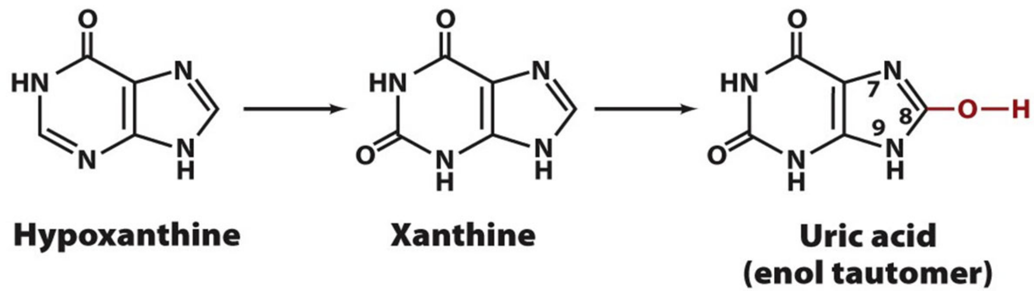
- C: Cytosine
- U: Uracil (RNA)
- T: Thymine (DNA)



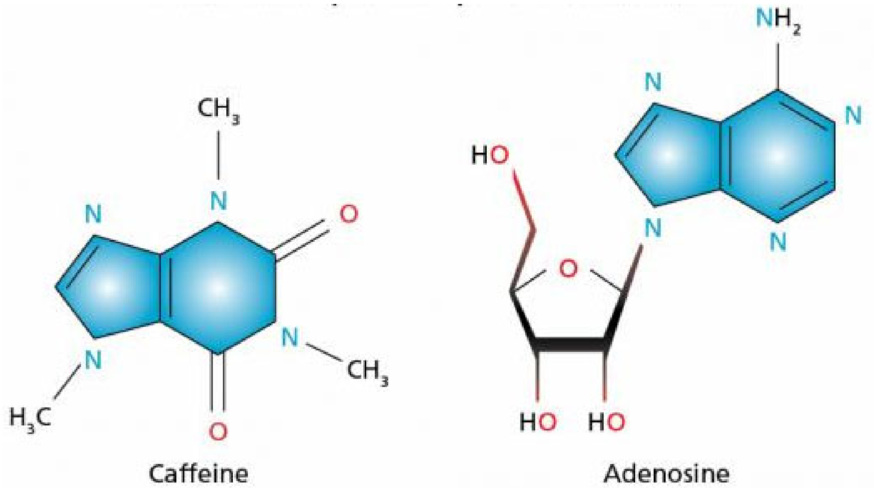
Pentose sugars



Other naturally occurring bases:



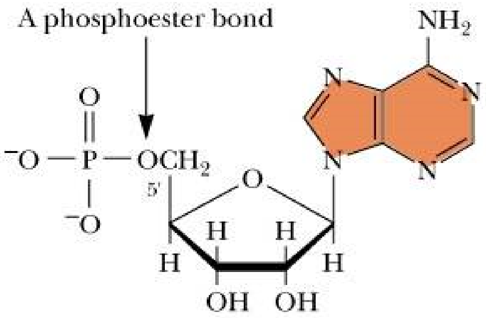
Adenosine:
(adenine + sugar) has physiological activity



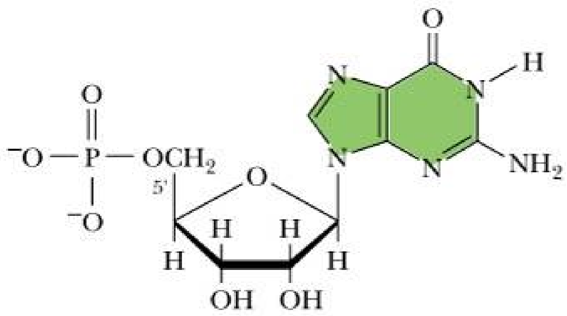
- is an **autacoid**, or local hormone, and neuromodulator.
- influences blood vessel dilation, smooth muscle contraction, neurotransmitter release, and fat metabolism.
- also a sleep regulator:
  - ➢Adenosine rises during wakefulness, promoting eventual sleepiness.
- **Caffeine** (structurally similar to adenosine) promotes wakefulness by blocking binding of adenosine to its neuronal receptors (inhibitory action).

Base + ribose sugar + phosphate = nucleotide

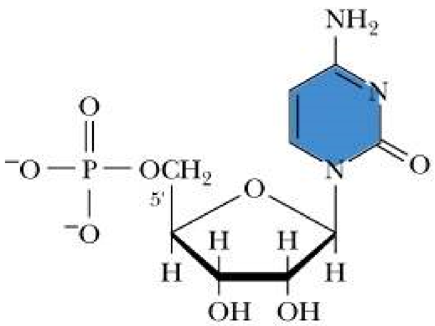
Adenylate, guanylate, cytidylate, uridylate formed



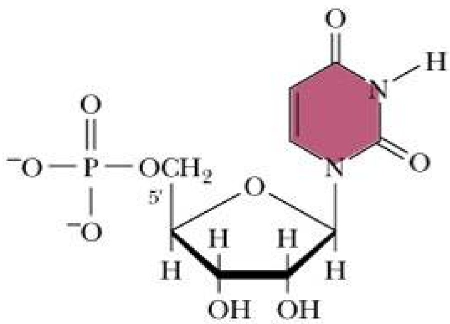
Adenosine 5'-monophosphate (or AMP or adenylic acid)



Guanosine 5'-monophosphate (or GMP or guanylic acid)



Cytidine 5'-monophosphate (or CMP or cytidylic acid)

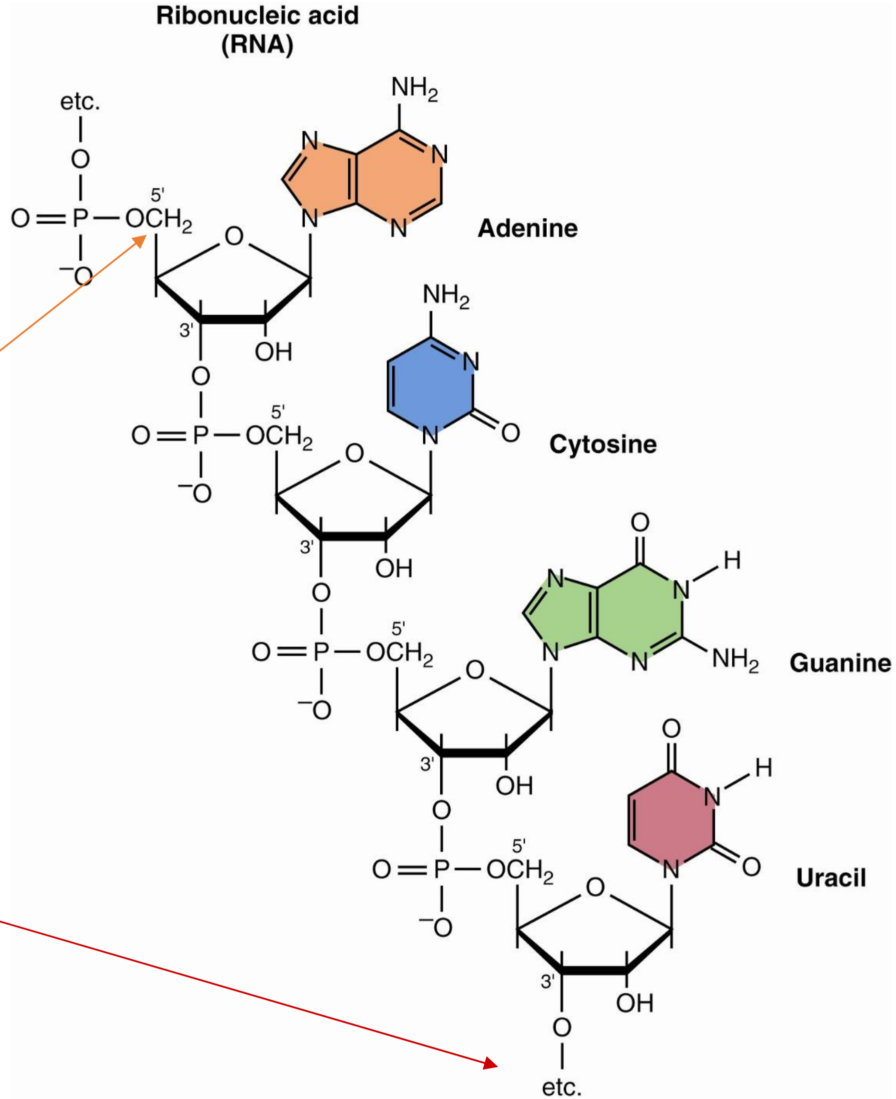


Uridine 5'-monophosphate (or UMP or uridylic acid)

- Energy carriers with 2 or 3 phosphate groups, e.g. ATP
- All tri-phosphates are important for metabolism
- Cyclized forms can be messengers

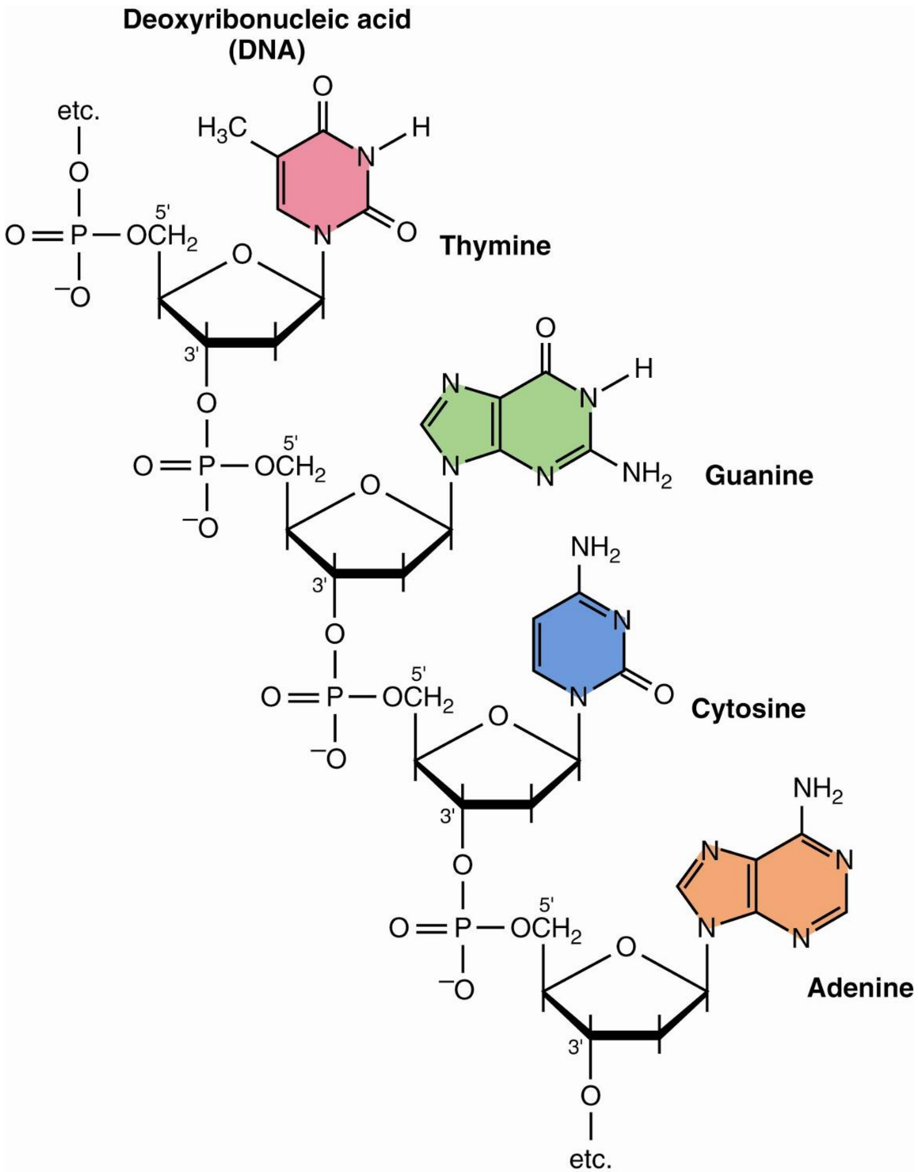
Making RNA from nucleotides

- Phosphodiester bridges link nucleotides together to form polynucleotide chains.
- The 5'-ends of the chains are at the top; the 3'-ends are at the bottom.
- Sequence is: ACGU



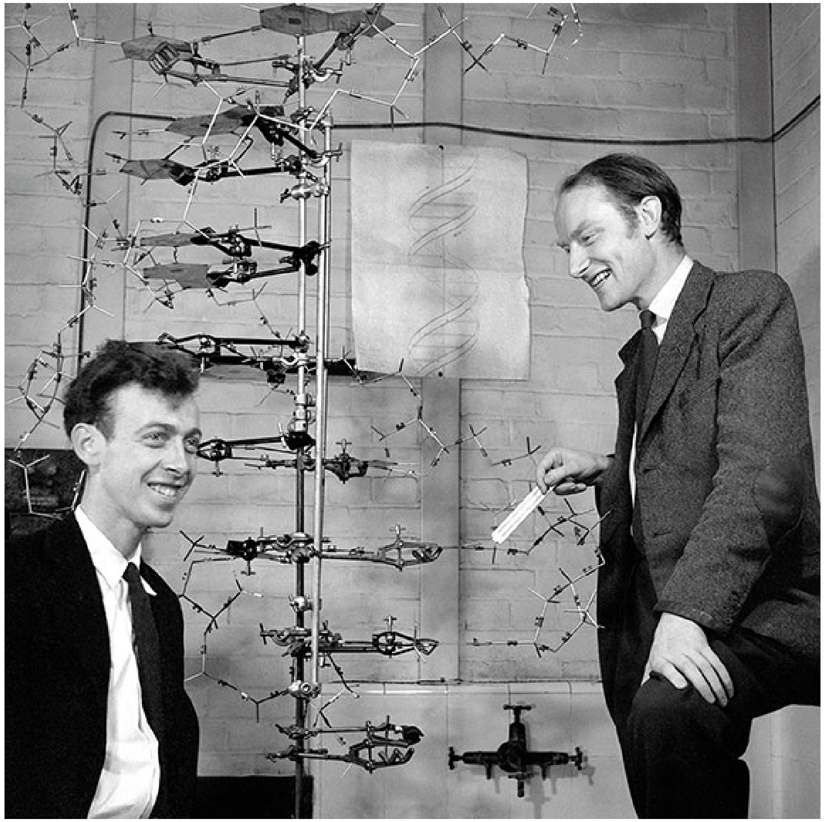
Making DNA from nucleotides

- Poly(deoxy)- nucleotide chains.
- The 5'-ends of the chains are at the top; the 3'-ends are at the bottom.
- Sequence is: TGCA
- DNA cannot be linked at 2’ (deoxy)
- **A** pairs with **T** and **G** with **C** to form an anti-parallel double helix.
- A%=T% and G%=C%



DNA structure solved in 1953

- The American biologist James Watson and the English physicist Francis Crick described the double-helical structure of DNA



Regions of DNA called <u>Genes</u> direct protein synthesis

Unnumbered 3 p50

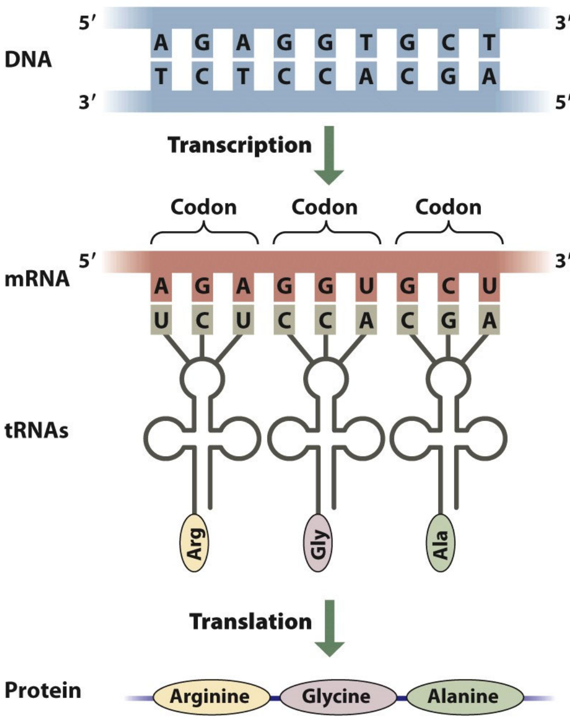


Figure 3-12

Translation occurs in a special organelle called the ribosome

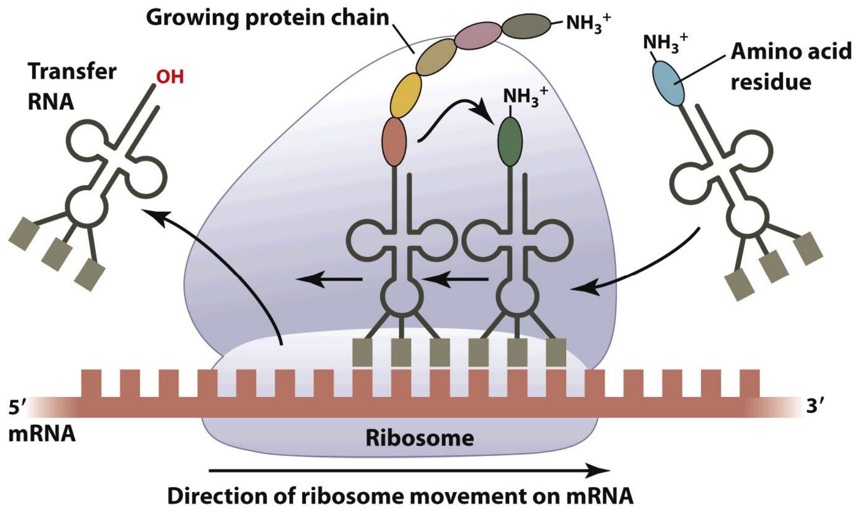
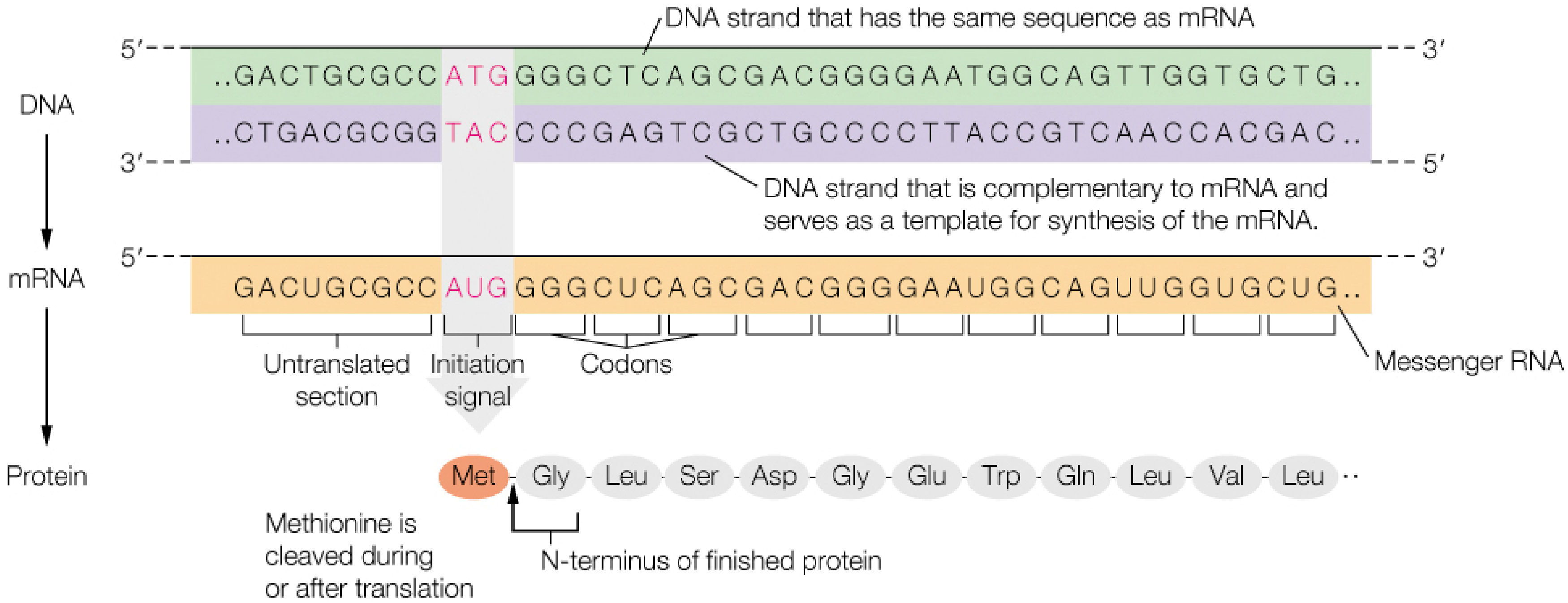


Figure 3-13

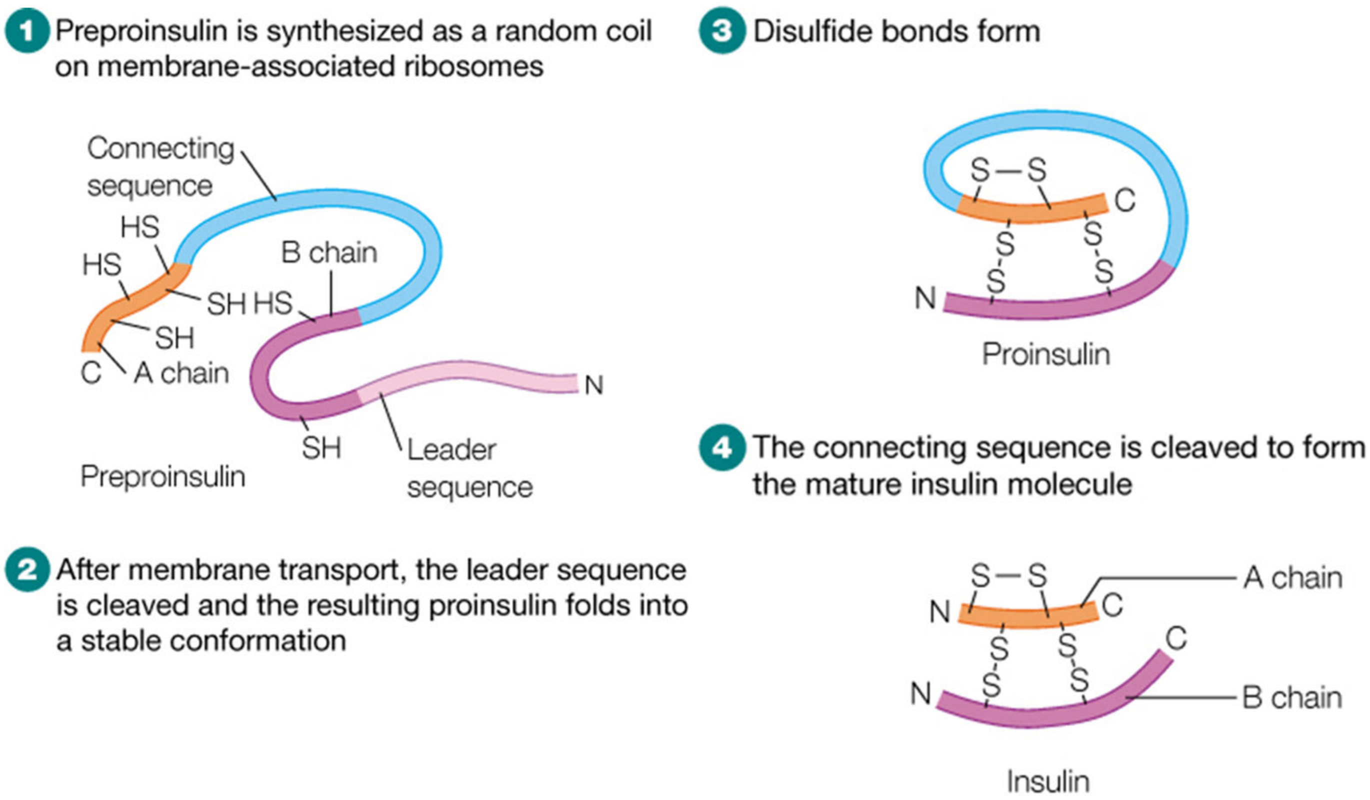
Making myoglobin

- The myoglobin gene is transcribed into an mRNA which is translated into protein



Some proteins have a more complex synthesis

- **Insulin** is synthesized as an inactive precursor, **preproinsulin**, which undergoes enzymatic cleavage to proinsulin, followed by folding and disulfide bond formation, followed by additional enzymatic proteolysis
- This allows insulin, like other peptide or protein hormones, to be synthesized and stored as an **inactive precursor**, available for rapid mobilization on demand



Amino acids – 2 summary

Stereochemistry and Light absorption

- Amino acids and many other biological compounds are chiral molecules whose configurations are shown by Fischer projections
- The amino acids in proteins all have the L stereochemical configuration
- Proteins absorb UV light at 280 nm

Amino acid derivatives

- The side chains of amino acid residues in proteins may be covalently modified.
- Some amino acids and amino acid derivatives function as hormones and regulatory molecules.
  - ➢Glycine and glutamate are themselves neurotransmitters

Amino acids to proteins

- The peptide bond between amino acids leads to peptides and proteins, with defined sequence and direction
- Proteins vary is size considerably and retain the properties of their constituent side chains.

Amino acids – 2 summary continued

Genes to proteins

- Proteins are encoded by genes: DNA to RNA to protein)
- DNA and RNA are composed of nucleic acids, ribose sugars and phosphate groups
- Nucleotides (base, ribose sugar and phosphate groups) carry energy (e.g. ATP) as well as information.
- DNA and RNA are both made up of **bases** (ATCG or AUCG) supported by a **sugar**, strung together by **phosphodiester** bonds.
    - RNA sugar is **ribose**; DNA sugar is **deoxyribose.**
- Regions of DNA called genes are transcribed into mRNA, which are then translated into proteins.
- After translation, proteins may be modified to generate the biochemically function form.

Reminders

- Pracs and Tutorials start this week
- Classes 1-3: Pracs this week
  - Pl. do your pre-lab
  - Pl. come to the lab on time, with your lab coat and fully enclosed shoes!
- Classes 4-6: Tutorials this week
  - Class 7: online tutorial
- Please sign up at the text book site from iLearn
  - ***Tutorial 1 Quiz (1%) for all classes on the textbook site.***