ELM Basics
MSE 493

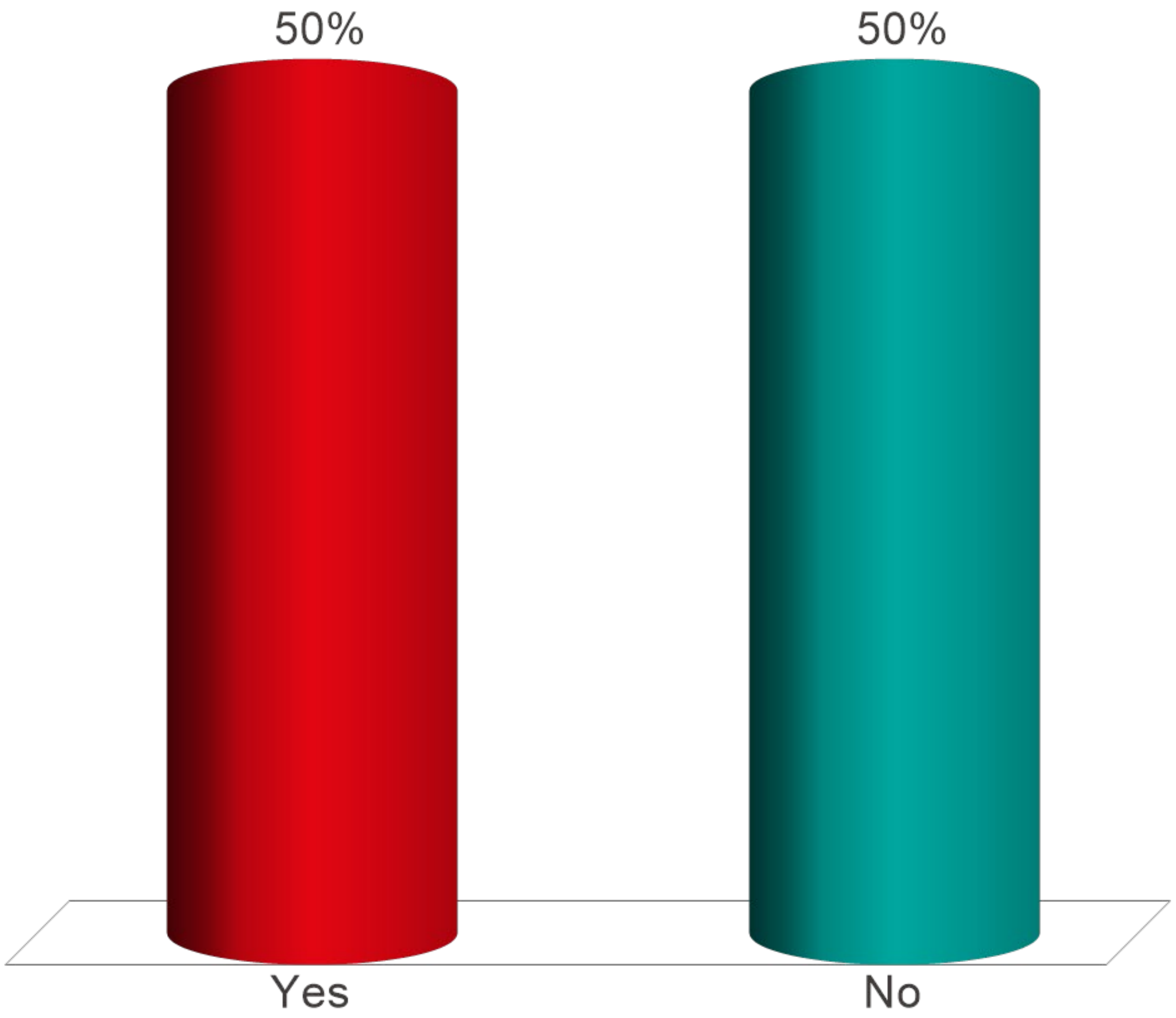
Prof. Tiffany Abitbol

2024

How many of you have taken a biology course at university level? go to: ttpoll.eu Session ID: MSE493 (no space)

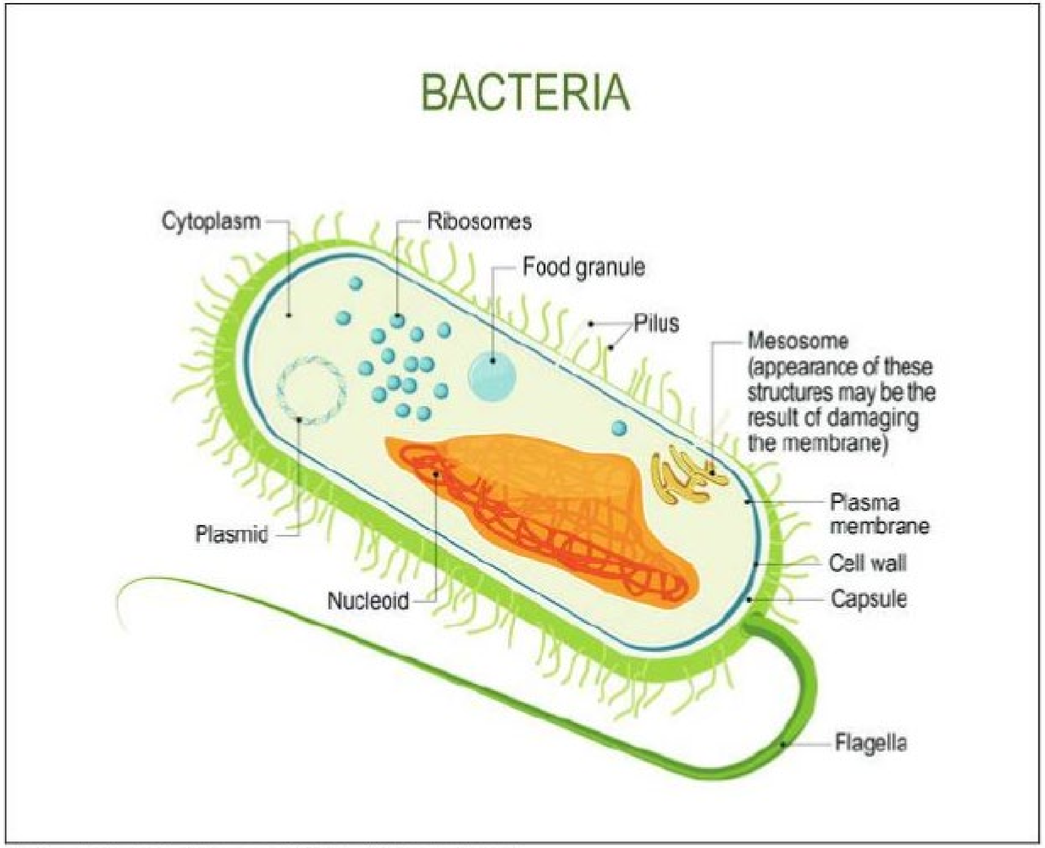
A. Yes

B. No



Cells

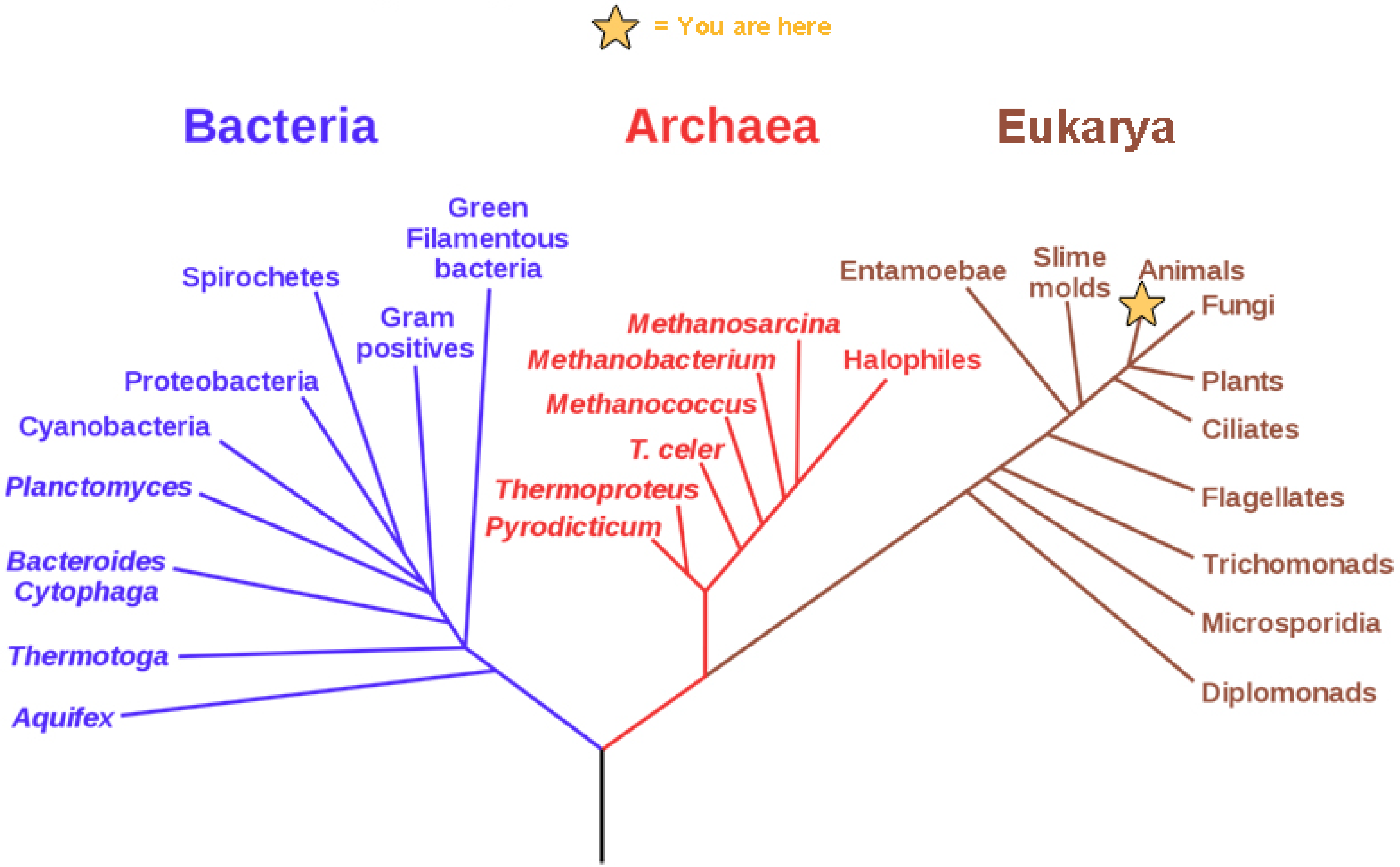
- Basic unit of organization in biological systems
- Many different types, but some common features
- Cell membrane
- Sometimes a cell wall (bacteria, yeasts, plant cells)
- Contains genetic material that provides instructions for function



Structure of a bacterial cell. Image via iStock/ttsz. [Used under license.]

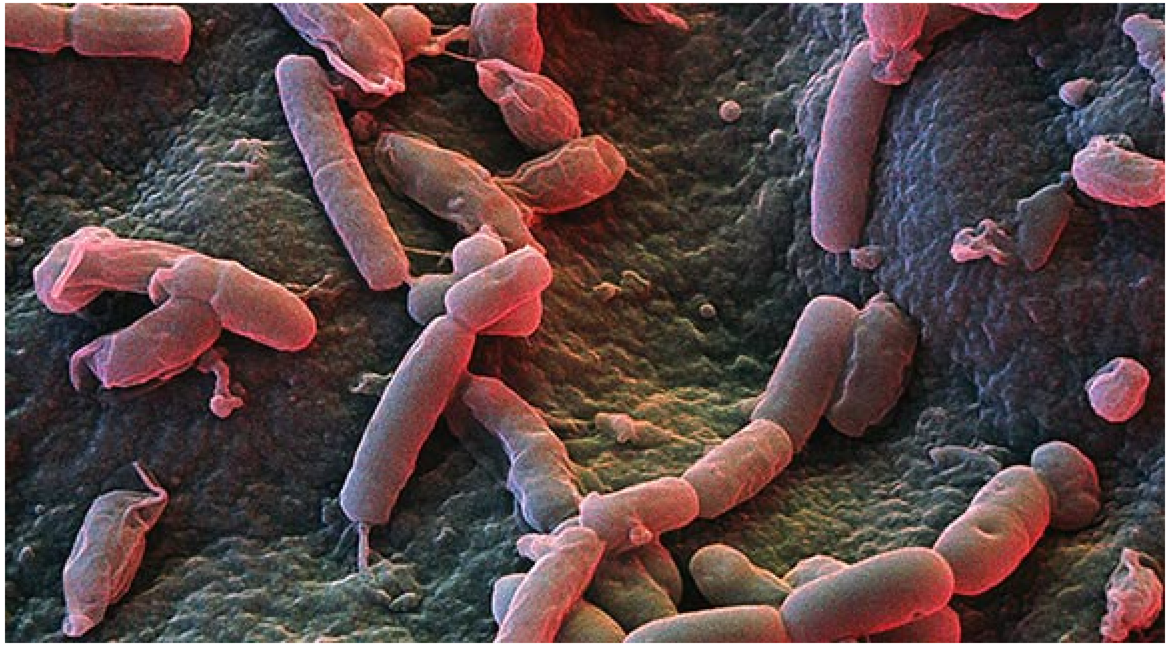
https://doi.org/10.1017/CBO9780511800986.004

Phylogenetic Tree of Life



[Three domains]

Bacteria and Archaea



Prokaryotes:
- No nucleus, no organelles, single cell
- Genetic information (DNA) is not compartmentalized
- Bacterial cell 1-10 μm

Eukarya

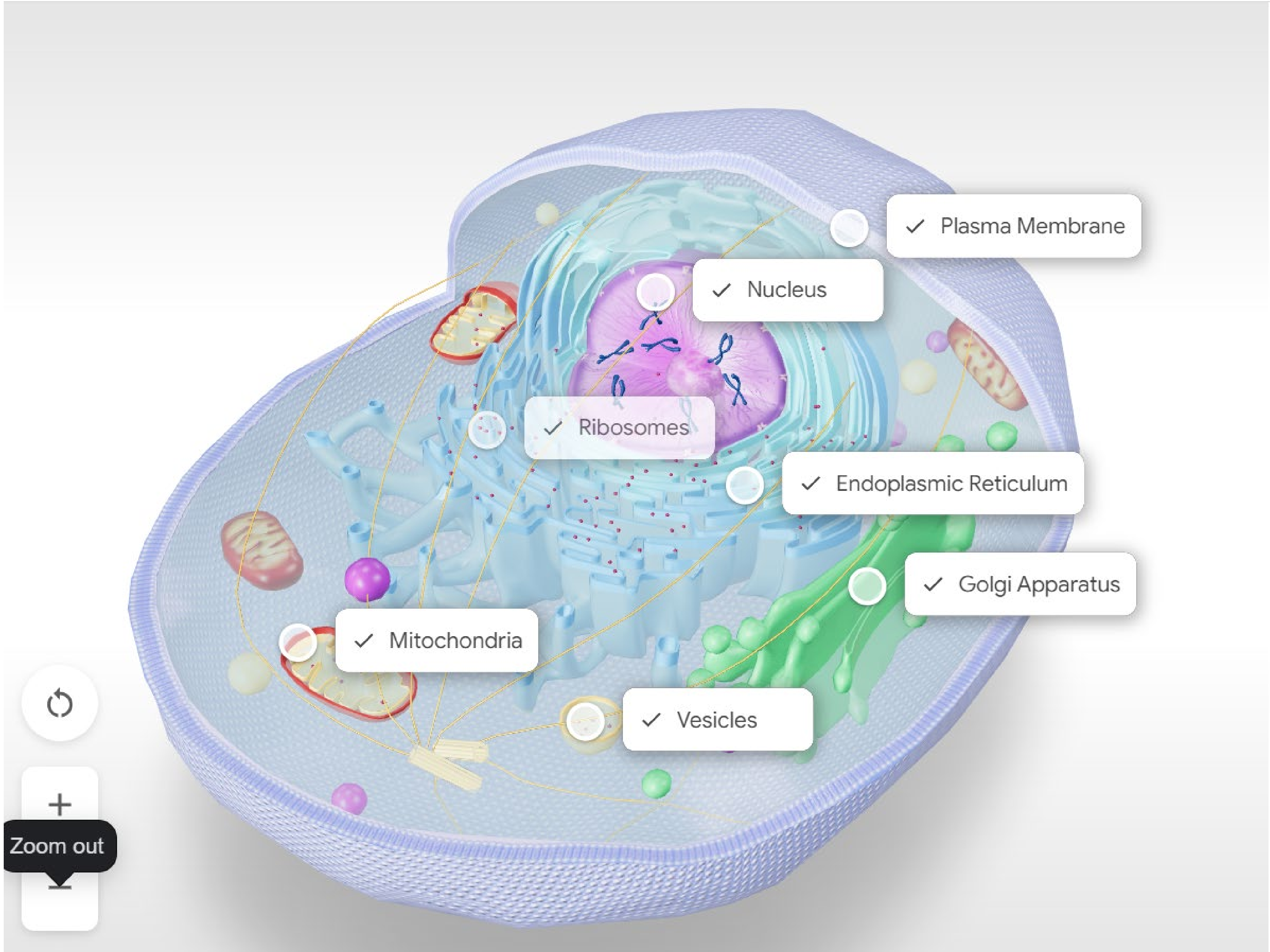
Eukaryotes:
- Nucleus, organelles, multicellular, more complex
- DNA located in membrane-bound nucleus
- Plant/animal cells 10-100 μm

Three domains 2

EPFL

Eukaryote

From Visible Body



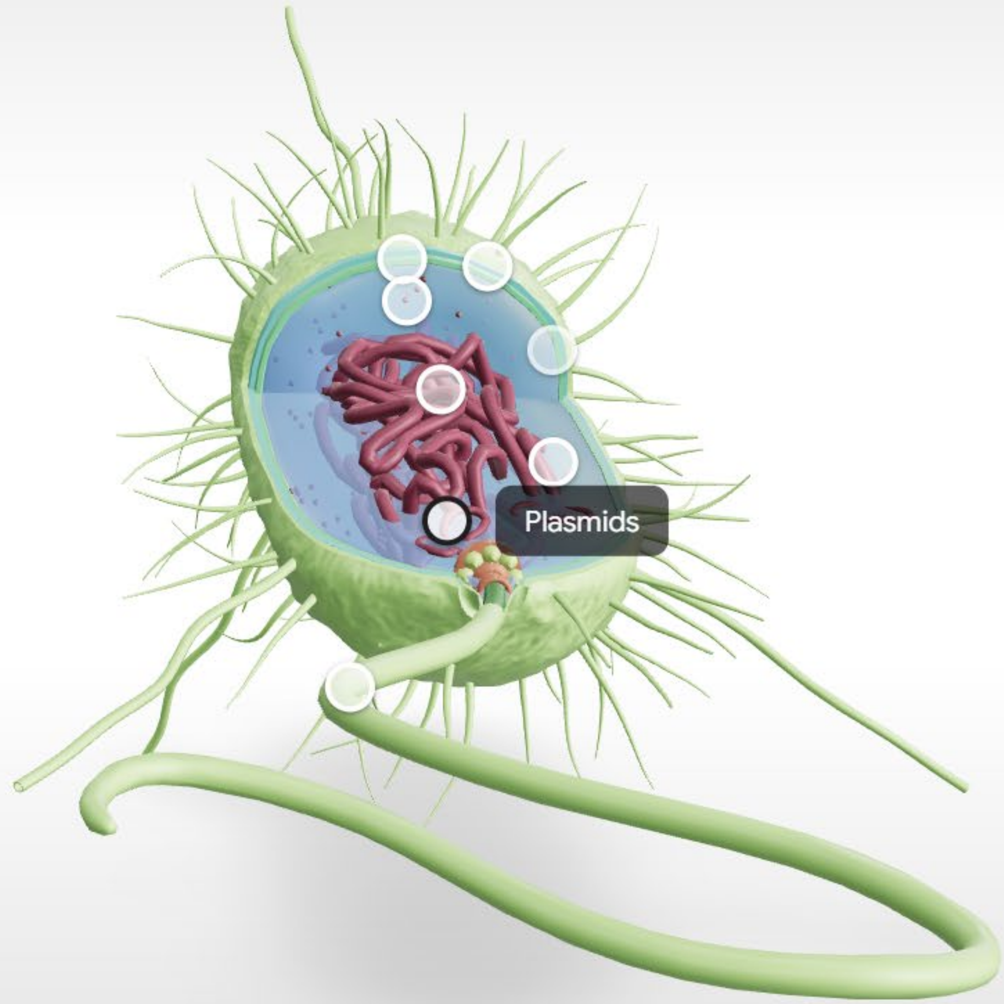
Notable mentions:

- Nucleus (chromosome, DNA)
- Ribosomes (protein factory)
- Mitochondria (powerhouse of cell)

Prokaryote

From Visible Body

Read Play



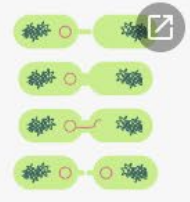
Fimbria

Flagellum

Nucleoid

Plasma Membrane

Plasmids



In addition to the DNA located in the nucleoid, bacterial cells have small, circular DNA molecules called plasmids.

Function: Plasmids carry nonessential genes that help bacteria survive in extraneous conditions.

Found in: Bacterial cells

Source Visible Body · Natallia Yatskova

Plasmids

Biological molecules

Banana

Fruit ⋮

Nutrition facts | Recipes | Types

Bananas

Sources include: USDA

Amount Per 100 grams	
Calories 89	
	% Daily Value*
Total Fat 0.3 g	0%
Saturated fat 0.1 g	0%
Cholesterol 0 mg	0%
Sodium 1 mg	0%
Potassium 358 mg	10%
Total Carbohydrate 23 g	7%
Dietary fiber 2.6 g	10%
Sugar 12 g	
Protein 1.1 g	2%

Vitamin C	14%	Calcium	0%
Iron	1%	Vitamin D	0%
Vitamin B6	20%	Cobalamin	0%
Magnesium	6%		

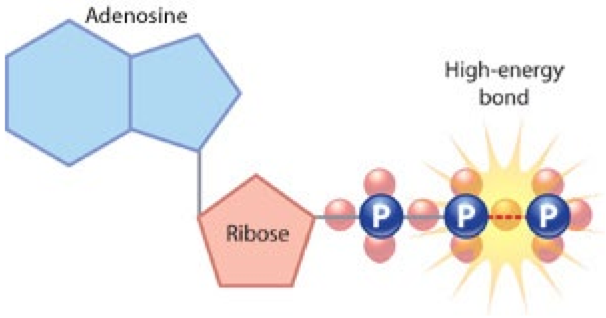
*Per cent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs.

- Lipids
- Carbohydrates
- Protein

digestion

Different cells metabolize the products of digestion to gain energy

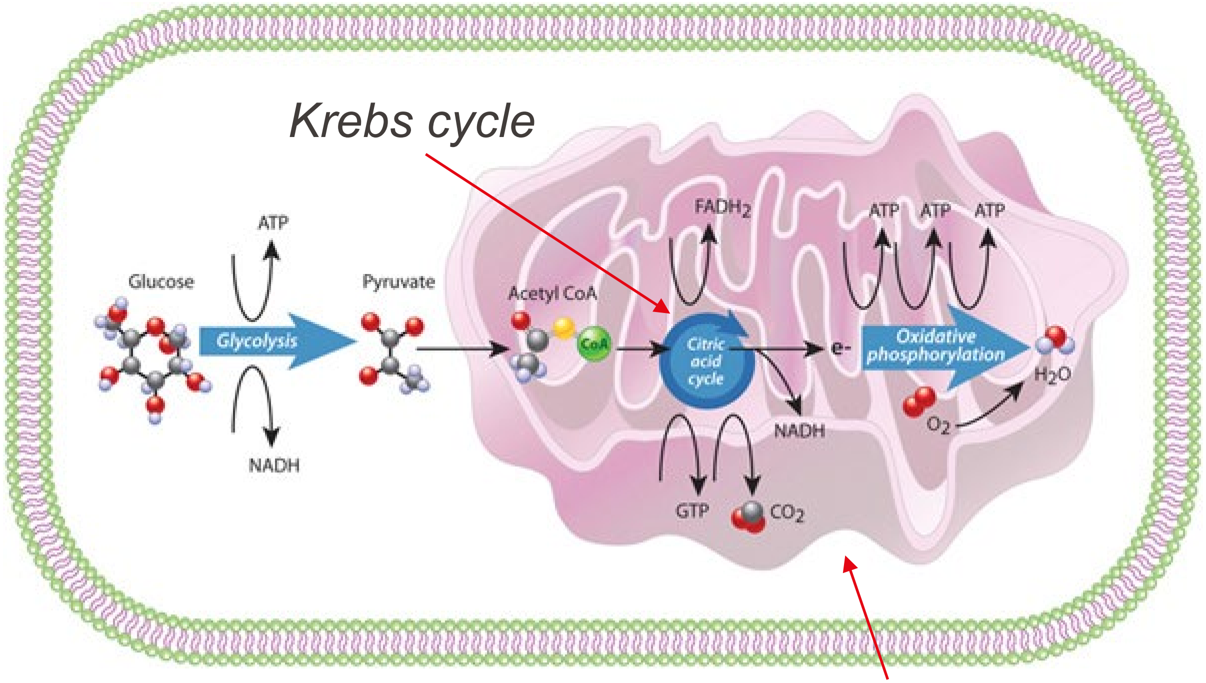
Glucose metabolism in eukaryotes



ATP: Adenosine 5'-triphosphate

- Cells transform the energy in the chemical bonds of food molecules to more readily usable forms – energy-rich carrier molecules
- ATP is the most abundant energy carrier molecule in cells

Glucose metabolism in a eukaryotic cell



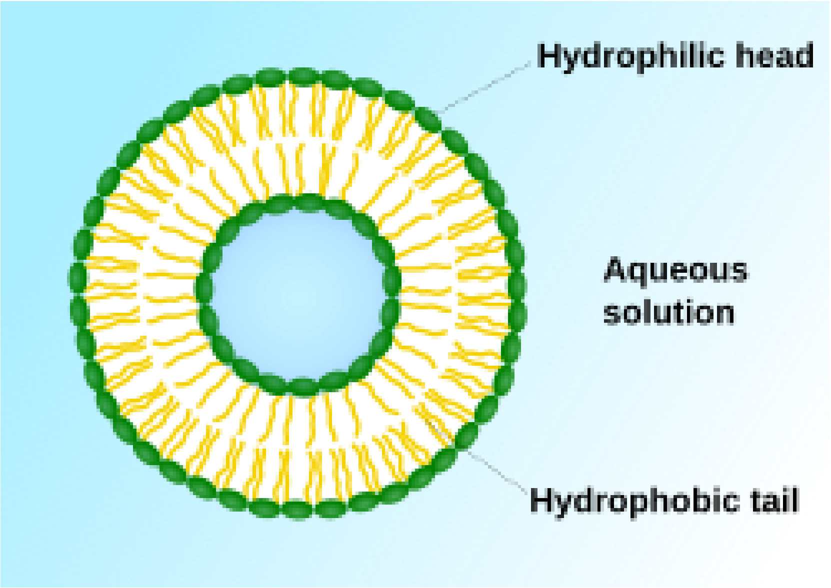
- *Cell's powerhouse*

1. Glycolysis
2. Citric acid cycle
3. Oxidative phosphorylation

1 glucose molecule = 36-38 ATP molecules

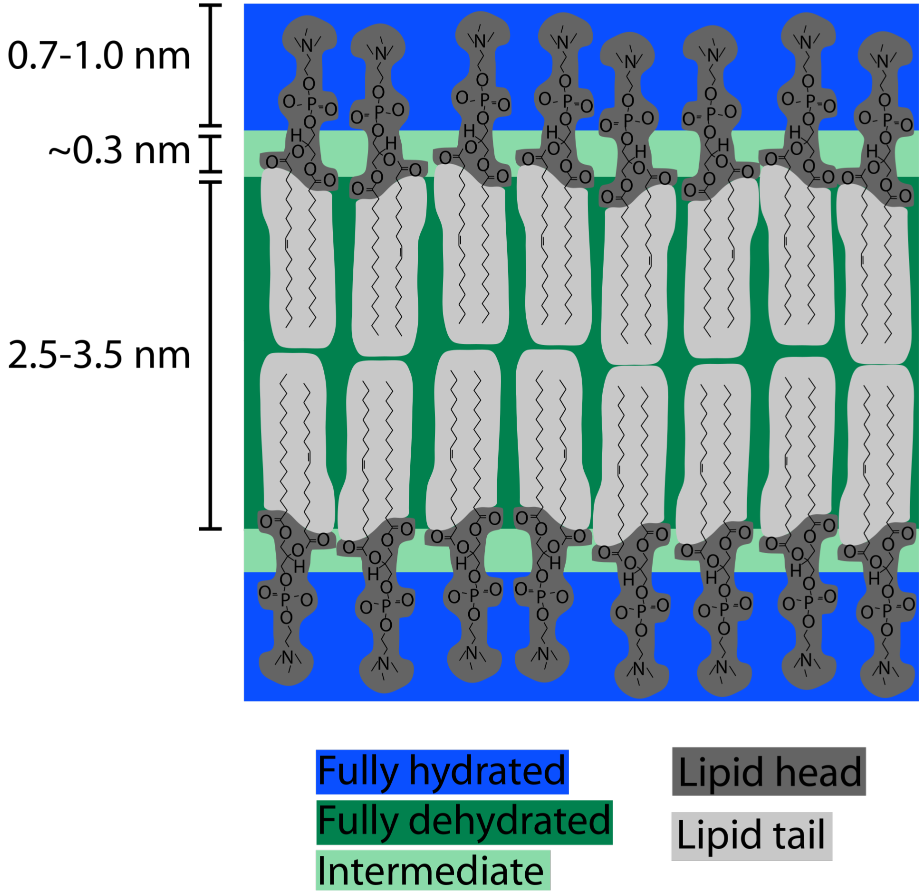
Cell energy

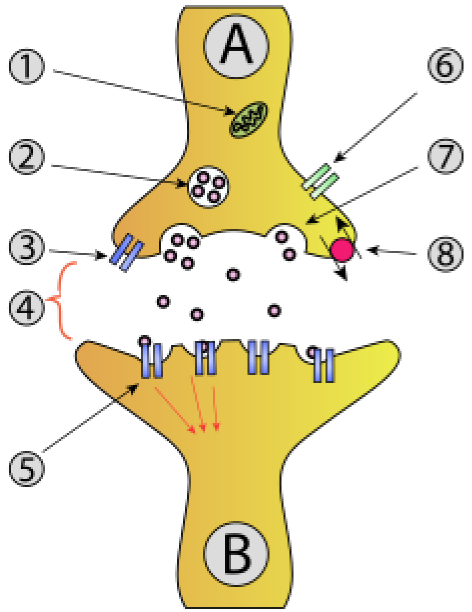
Lipids



Vesicle

Lipid bilayer

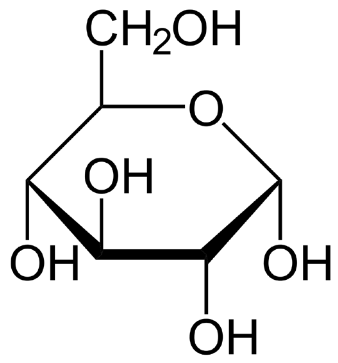




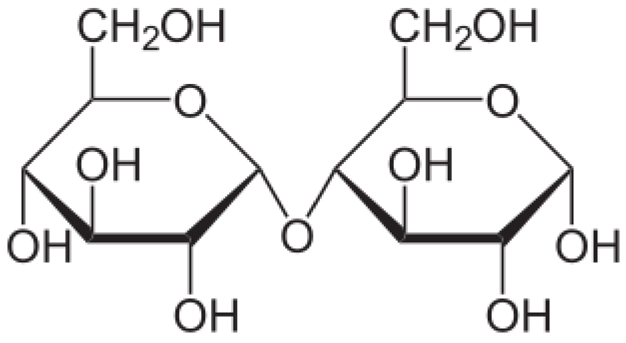
Synaptic vesicle

Carbohydrates

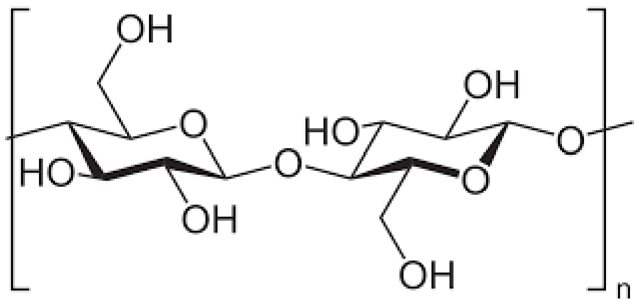
- $(CH_2O)_n$
- C:H:O = 1:2:1
- Subtypes: Monosaccharide, disaccharide, polysaccharide
- Link via dehydration



glucose



maltose



cellulose

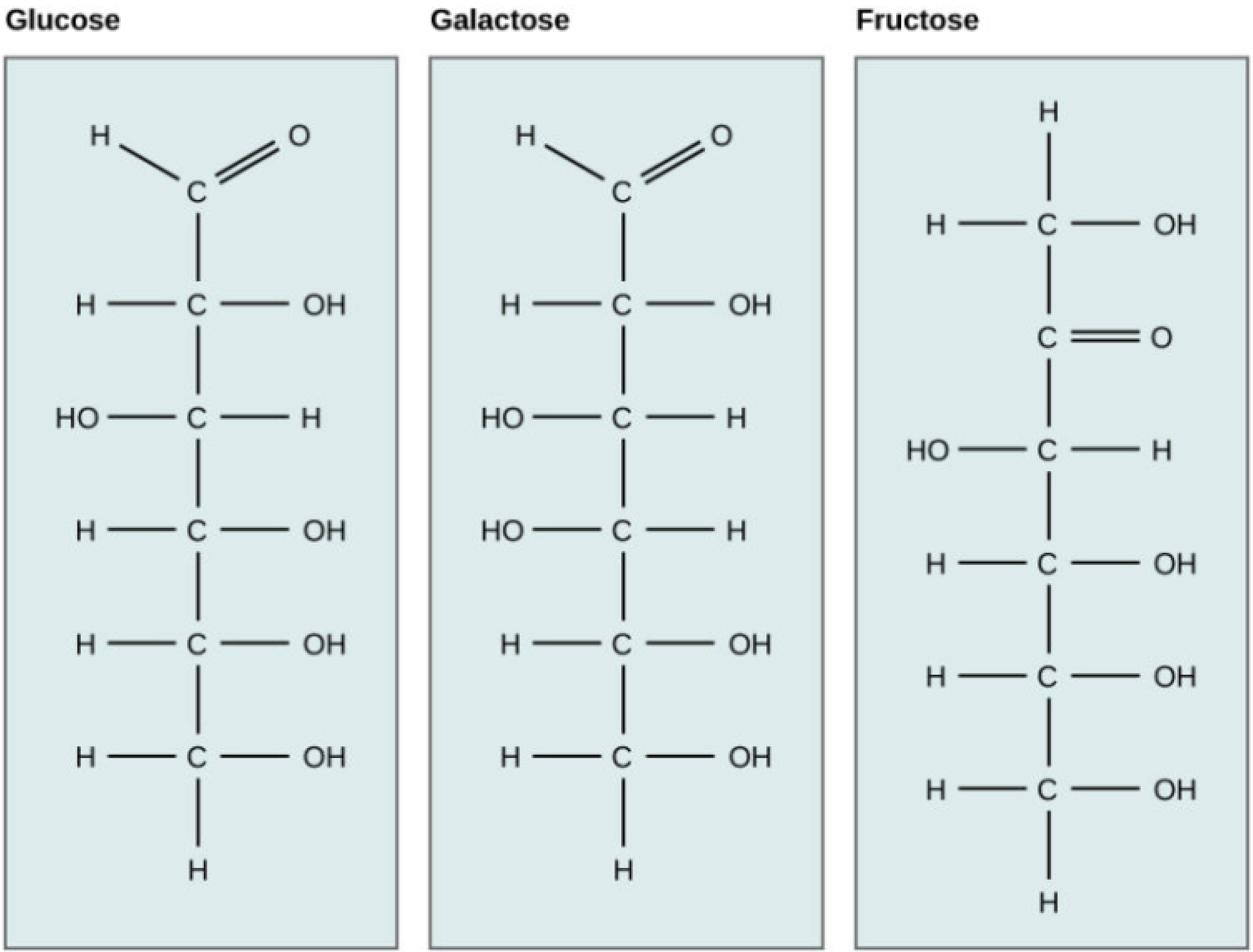
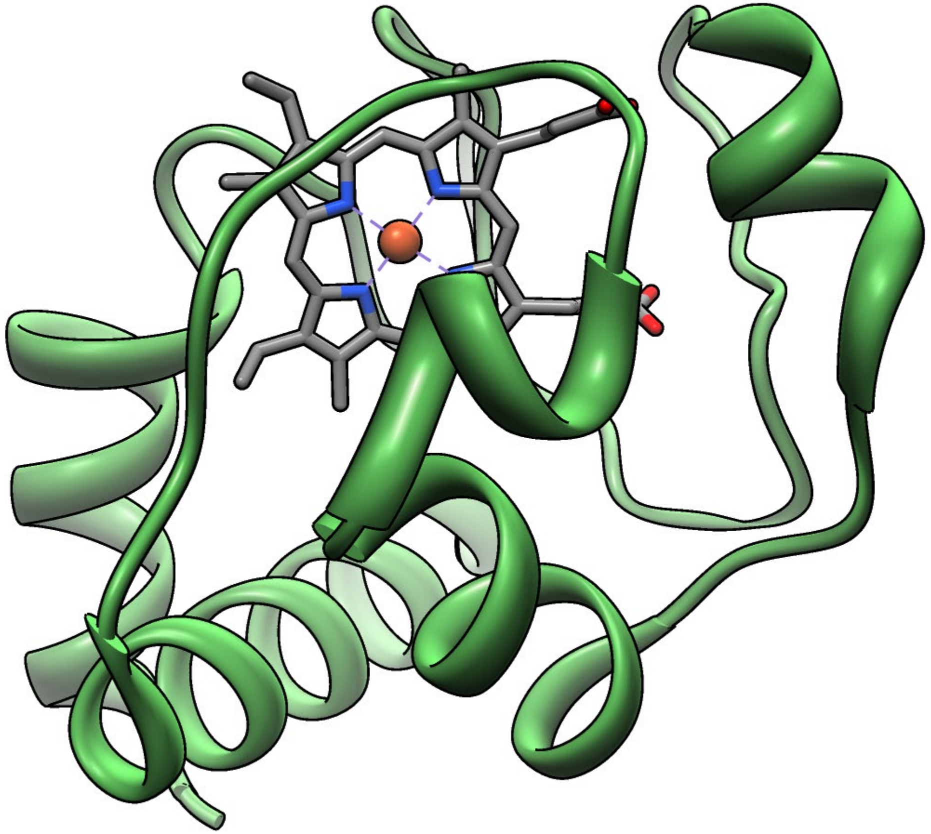


Figure 2.14 Glucose, galactose, and fructose are isomeric monosaccharides, meaning that they have the same chemical formula but slightly different structures.

Biological molecules

Proteins

- Expression of genetic information is achieved via proteins (especially enzymes)
- Living systems contain 1000s of different proteins, each with a unique function
- Transport, storage, membranes, toxins, enzymes, hormones
- Linear polymer of amino acids (AA's)
- Different molecular weights
- Different shapes, e.g., globular or fibrillar
- Shape is critical to function (determined by sequence and number of AA's)

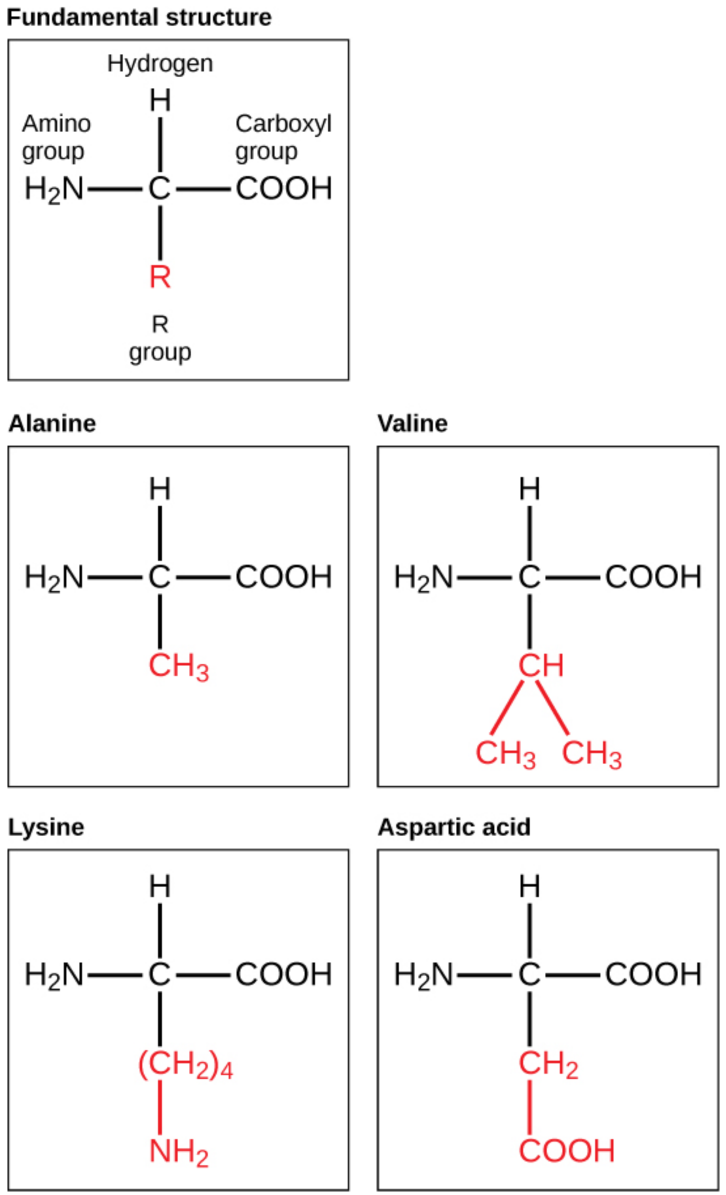


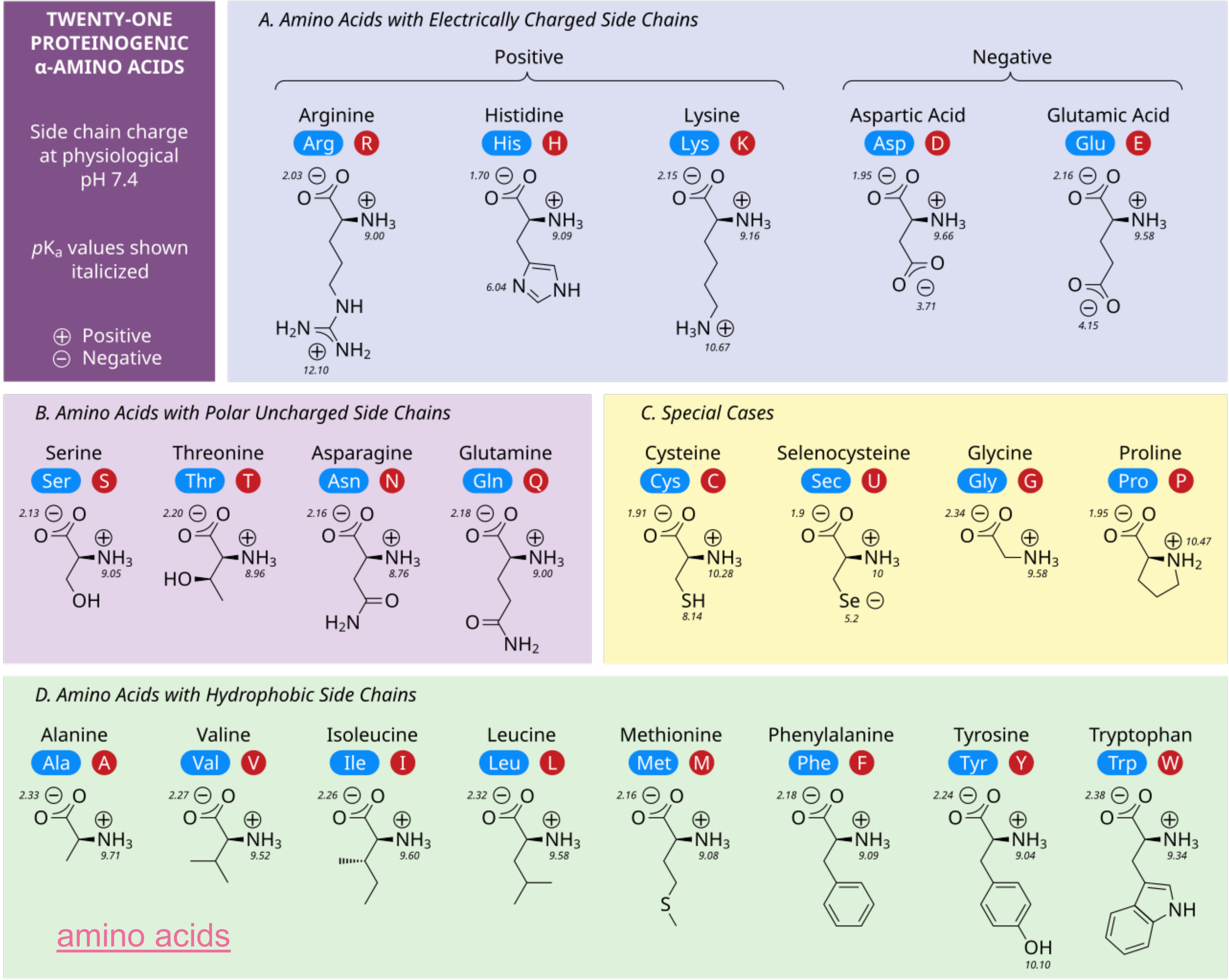
Cytochrome C

Biological molecules

Amino acids

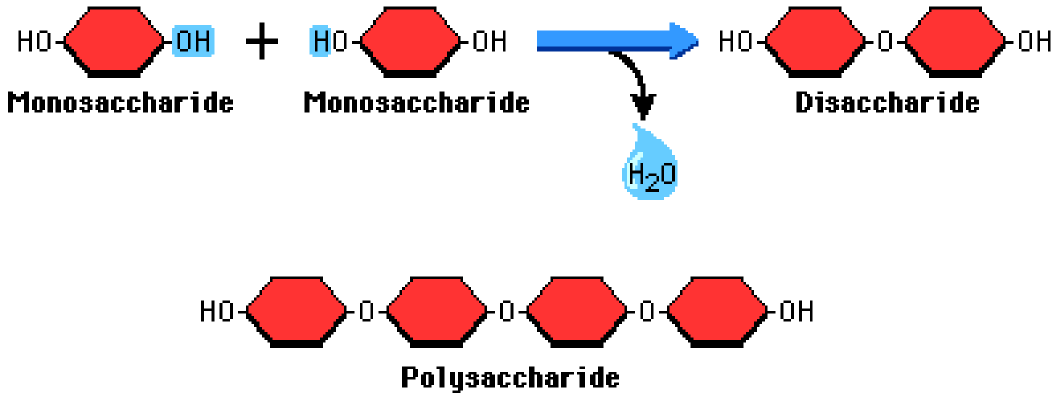
- 20 main AA’s
- Same fundamental structure
- Different R-groups: acidic, basic, polar, non-polar
- AA’s connect through peptide bonds via dehydration reactions
- A polypeptide is an AA polymer
- A protein is also an AA polymer, with a distinct shape and function
- Protein can consist of several polypeptides

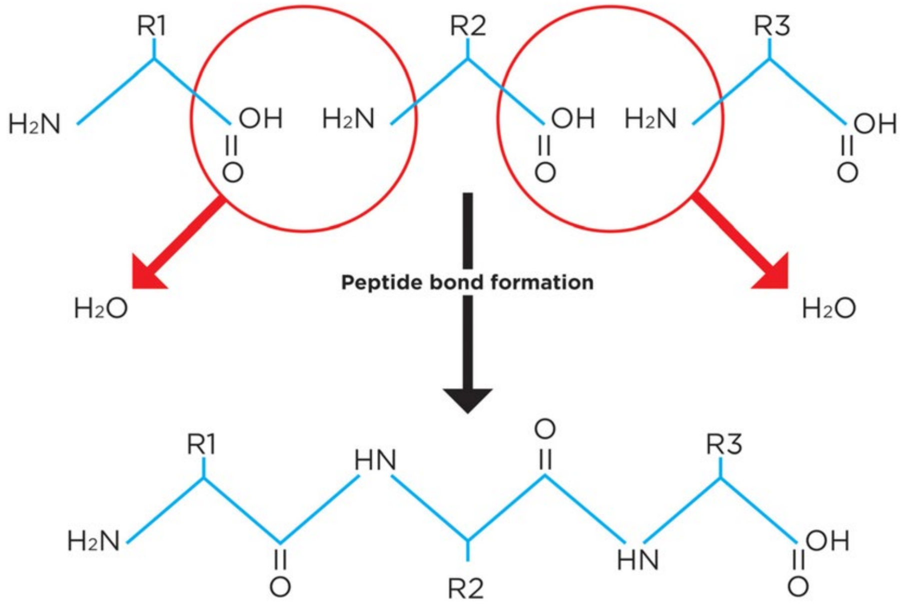



TWENTY-ONE PROTEINOGENIC α-AMINO ACIDS
Side chain charge at physiological pH 7.4
pKa values shown italicized
⊕ Positive
⊖ Negative
A. Amino Acids with Electrically Charged Side Chains
Positive
Negative
Arginine Arg R 2.03 9.00 12.10
Histidine His H 1.70 9.09 6.04
Lysine Lys K 2.15 9.16 10.67
Aspartic Acid Asp D 1.95 9.66 3.71
Glutamic Acid Glu E 2.16 9.58 4.15
B. Amino Acids with Polar Uncharged Side Chains
Serine Ser S 2.13 9.05
Threonine Thr T 2.20 8.96
Asparagine Asn N 2.16 8.76
Glutamine Gln Q 2.18 9.00
C. Special Cases
Cysteine Cys C 1.91 10.28 8.14
Selenocysteine Sec U 1.9 10 5.2
Glycine Gly G 2.34 9.58
Proline Pro P 1.95 10.47
D. Amino Acids with Hydrophobic Side Chains
Alanine Ala A 2.33 9.71
Valine Val V 2.27 9.52
Isoleucine Ile I 2.26 9.60
Leucine Leu L 2.32 9.58
Methionine Met M 2.16 9.08
Phenylalanine Phe F 2.18 9.09
Tyrosine Tyr Y 2.24 9.04 10.10
Tryptophan Trp W 2.38 9.34

amino acids

Dehydration reactions





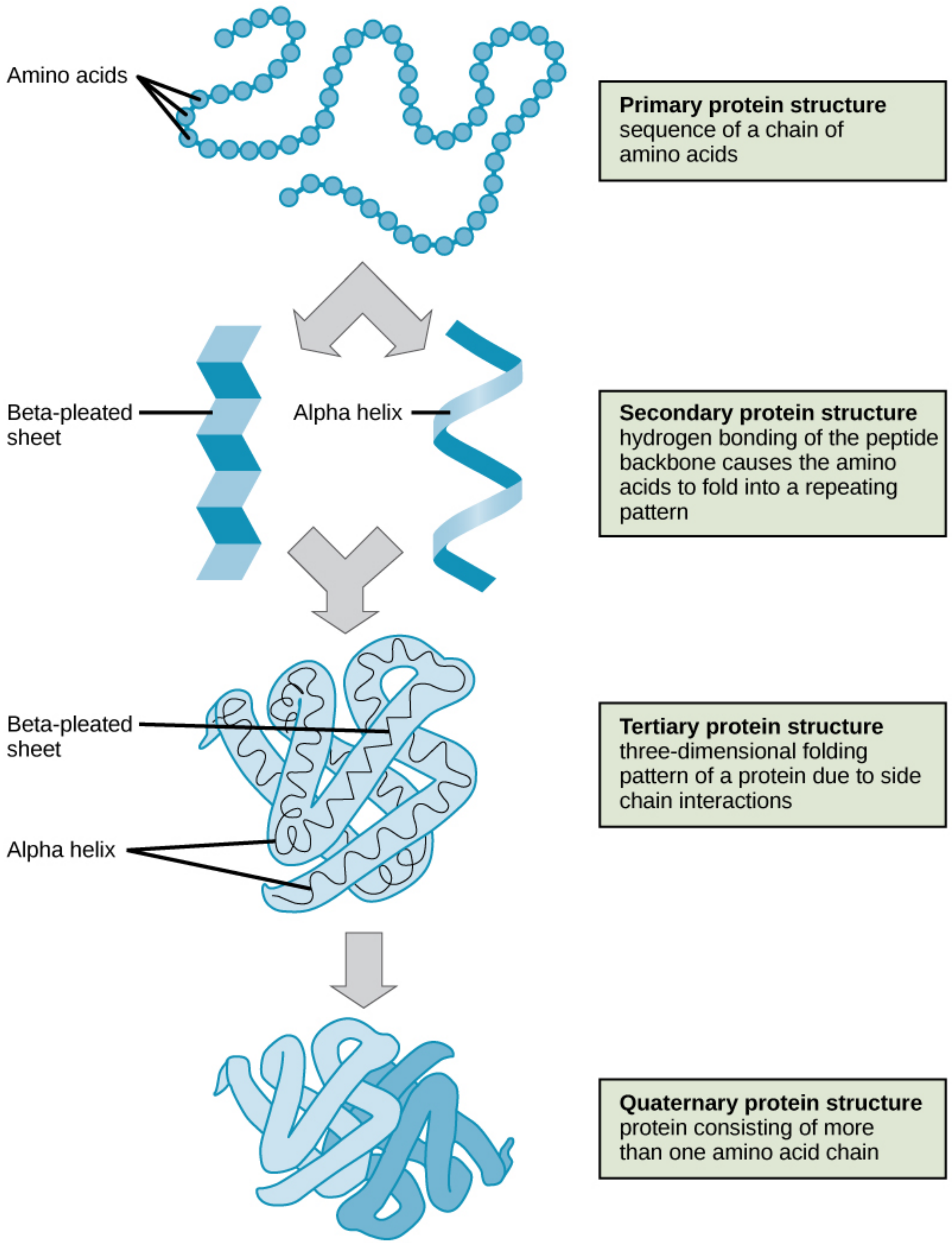
Glycosidic bonds

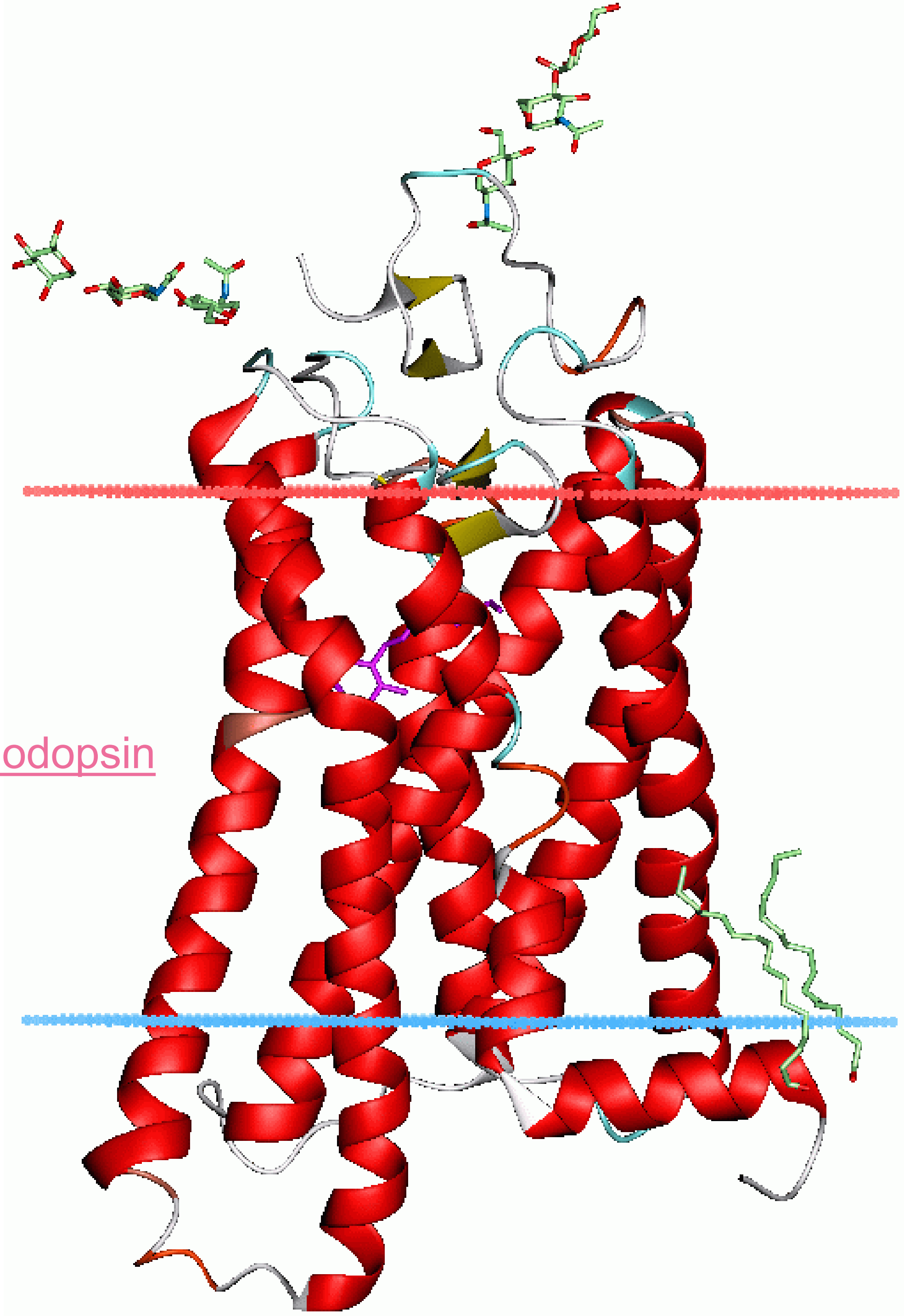
Peptide bonds

Protein structure

Biological molecules

- Primary – sequence and number of amino acids, determined by gene that encodes the protein
- Secondary – folding patterns from interactions between non-R groups, held together by H-bonds, e.g., α-helix and β-sheet
- Tertiary – 3D structure, mainly from interactions between R groups
- Quaternary – proteins formed from several polypeptides (subunits), functional form of many proteins

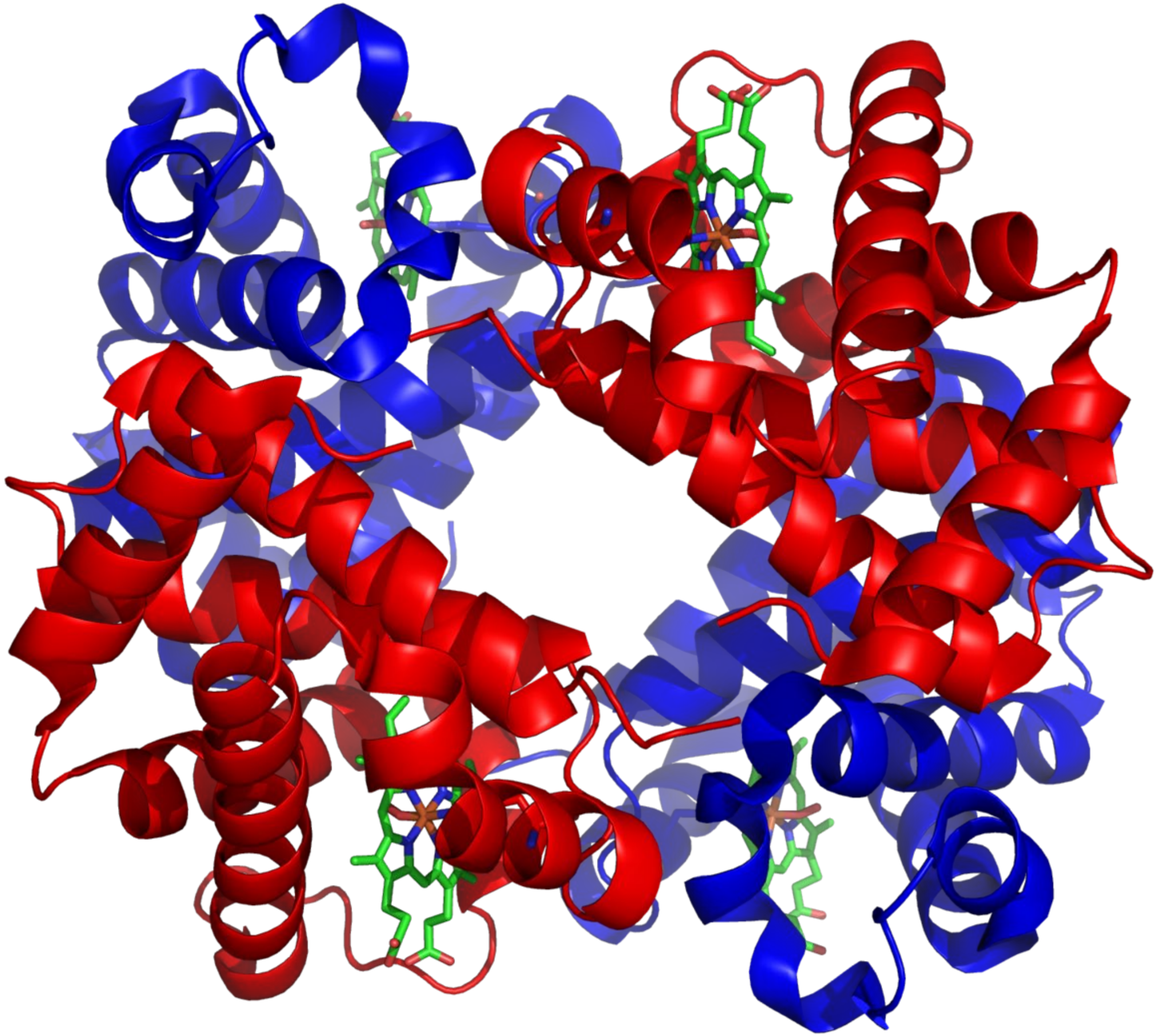




bovine rhodopsin

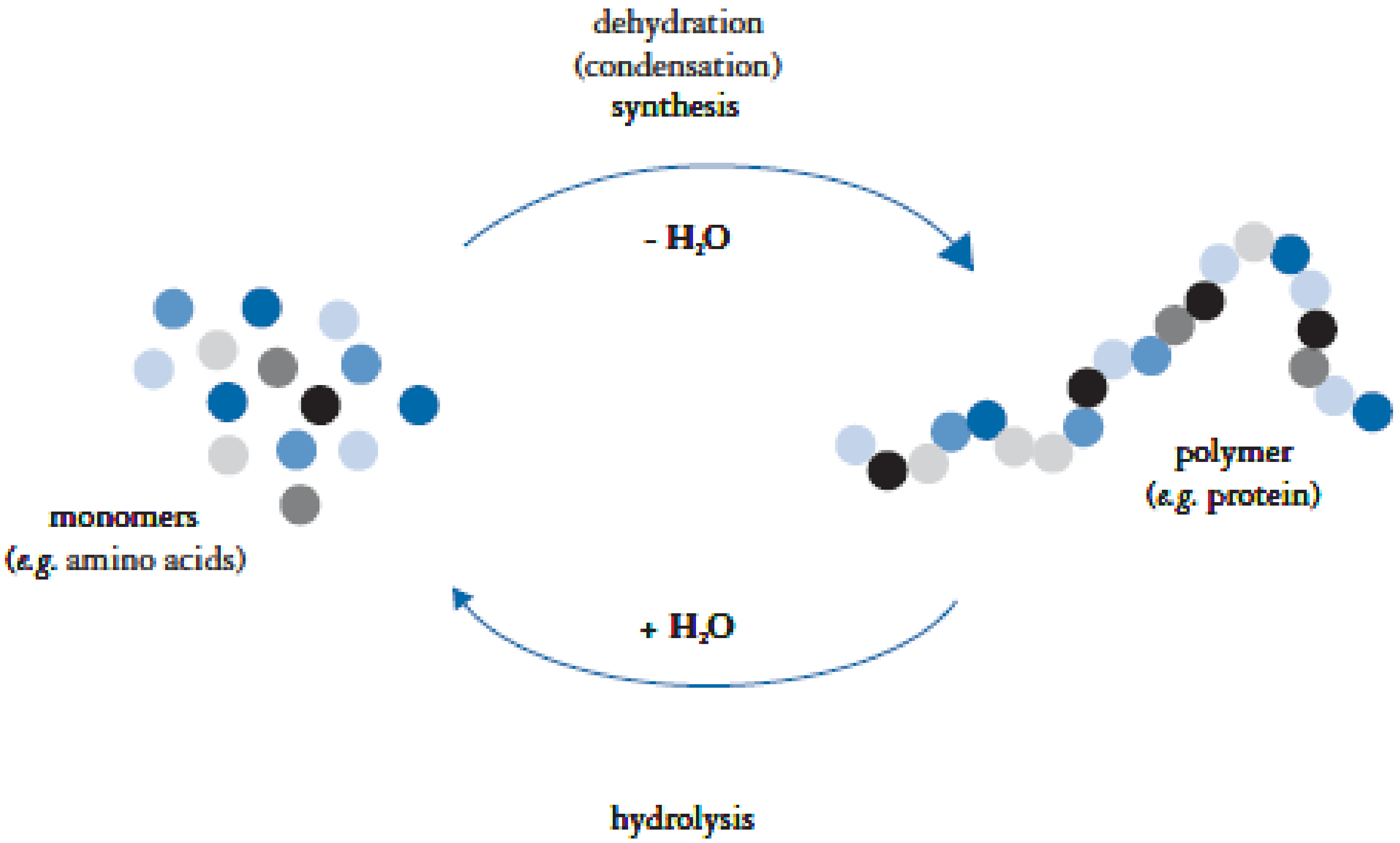
beta meander

Hemoglobin



- 4 subunits (quaternary structure)
- Oxygen transport in RBCs
- Contains 4 iron-containing heme groups = can bind up to 4 oxygen molecules
- Tense state: lower affinity to oxygen, promotes *oxygen release* in tissues where oxygen is low
- Relaxed state: upon binding 1 oxygen molecule, easier to bind other 3 (*cooperative binding*)

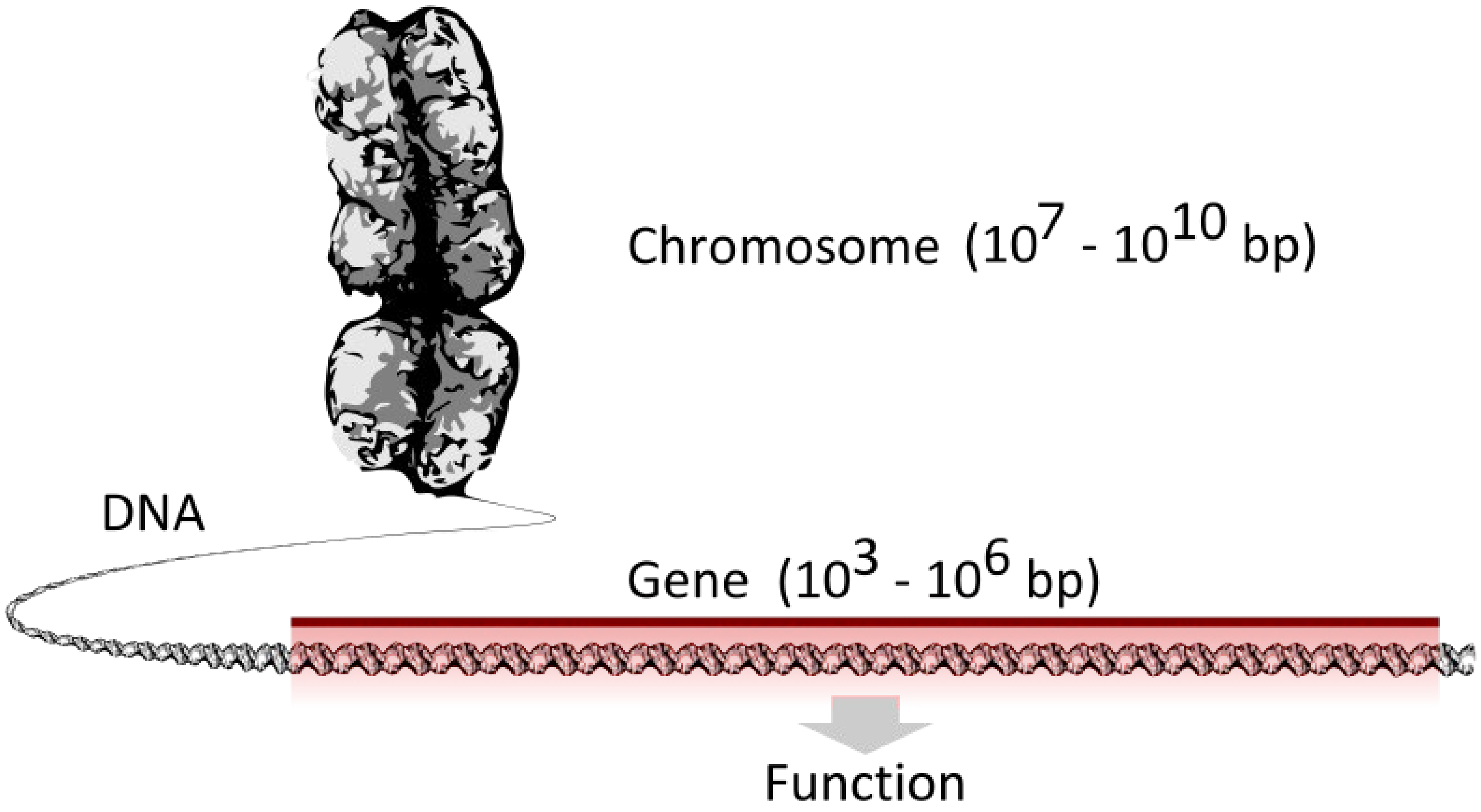
Monomer-polymer cycle



- Protein is a condensation heteropolymer (dehydration)
- Looks simple but protein synthesis requires many components whose functions are coordinated during the complex process of ***translation***

https://doi.org/10.1017/CBO9780511800986.004

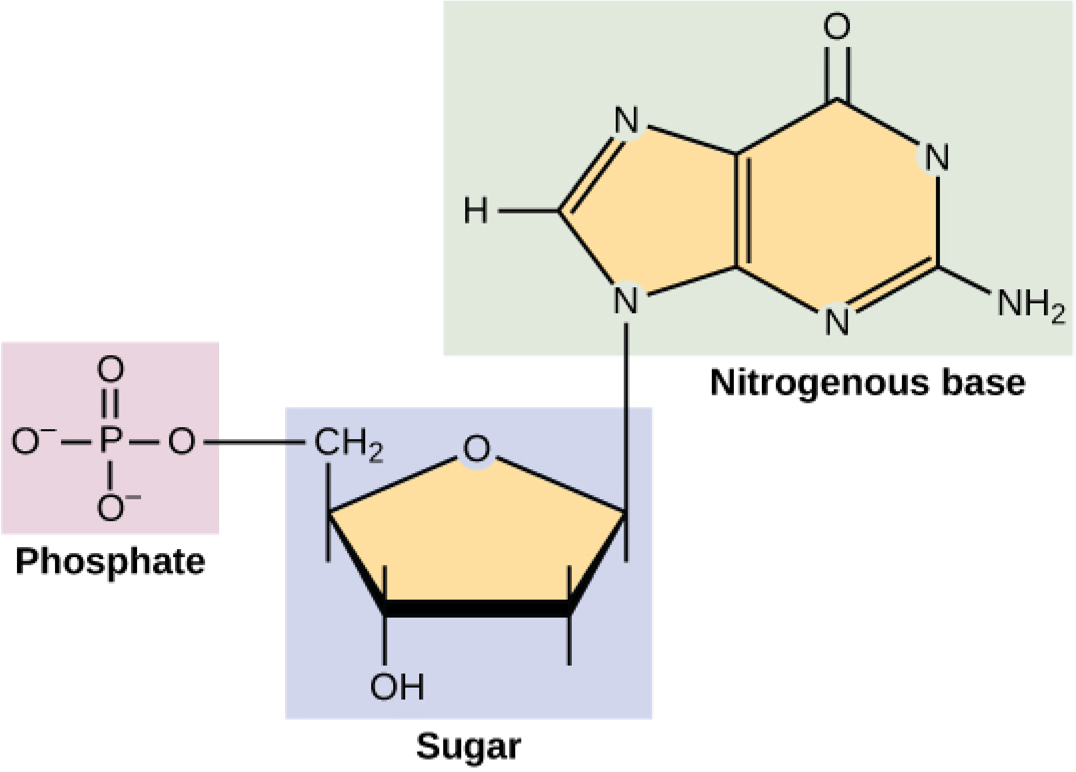
So how are proteins actually made?

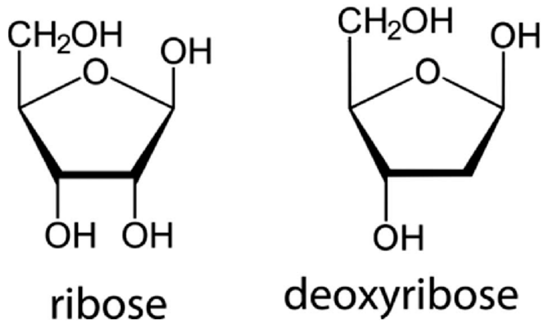


Chromosome

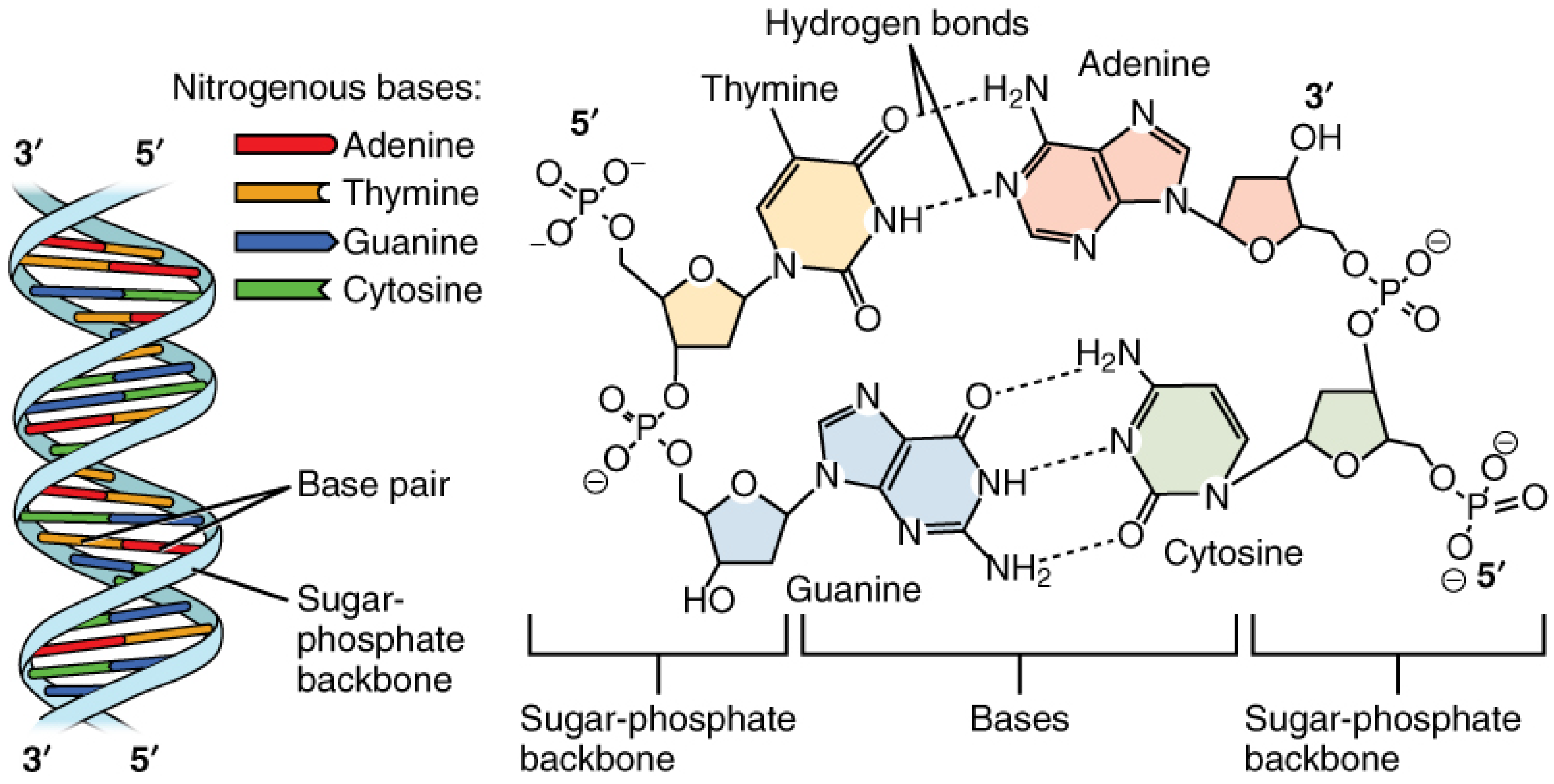
Nucleic acids (DNA/RNA)

- Carry genetic blueprint of life
- Two main types: DNA (deoxyribonucleic acid) and RNA (ribonucleic acid)
- DNA = genetic material of all living organisms
- RNA = mostly involved in protein synthesis
- Nucleotides are made of 3 components: nitrogenous base, pentose sugar, and a phosphate group
- Nitrogenous bases (**A**, **G**, **C**, **T**) have coding function in DNA



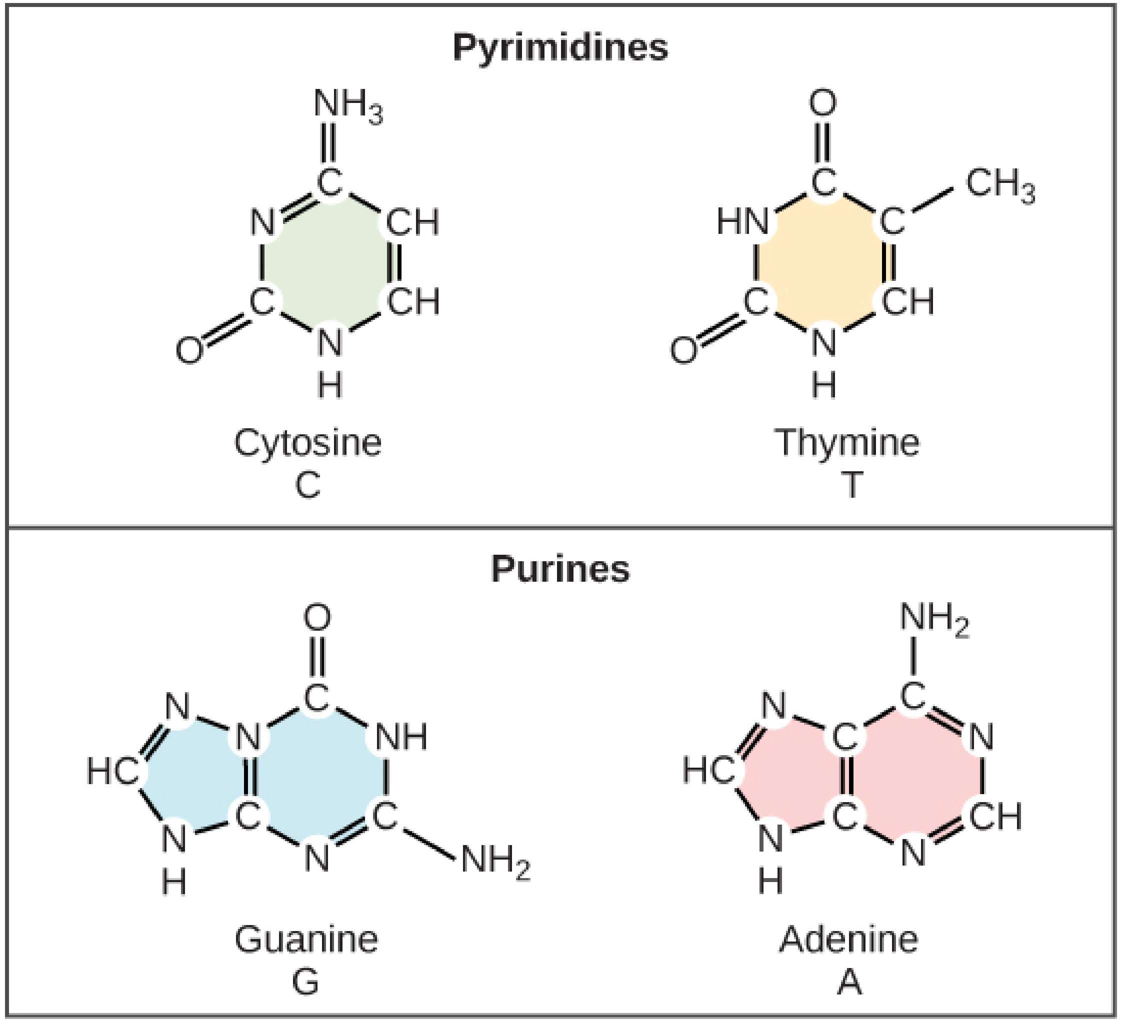


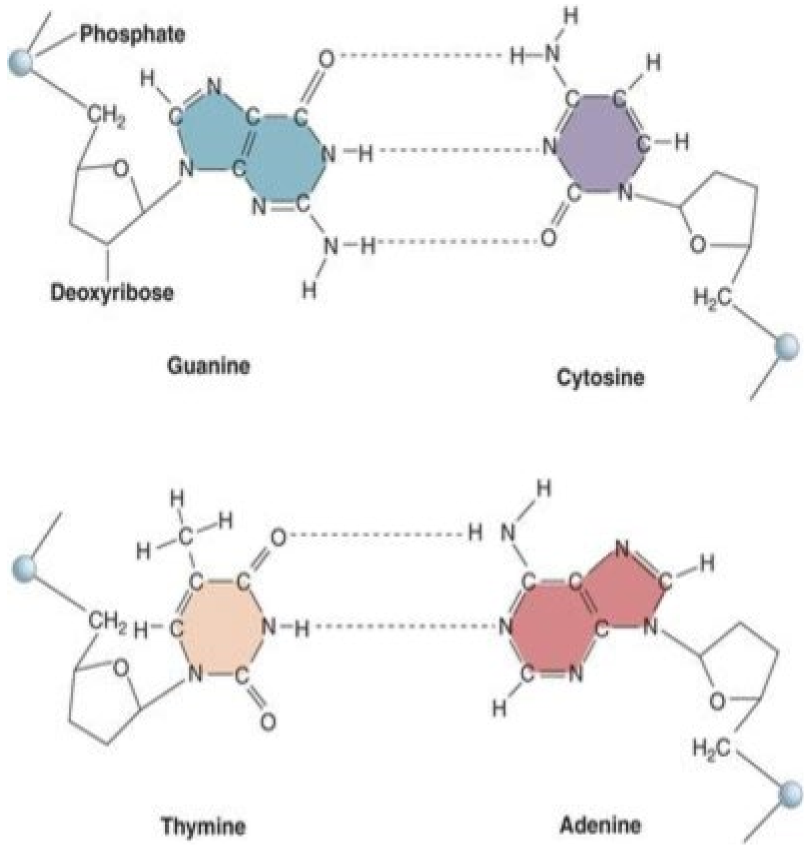
DNA



- Nucleotides join via 5’→3’ phosphodiester linkage that confers directionality
- 5’ end has a free phosphate, 3’ end has a free hydroxyl

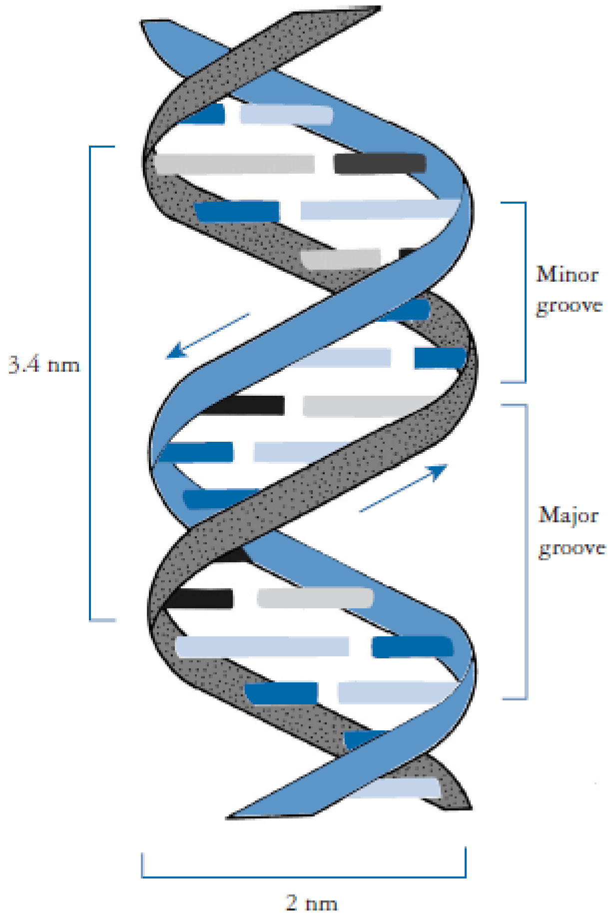
All genetic information – 4 letters only





DNA bases

DNA

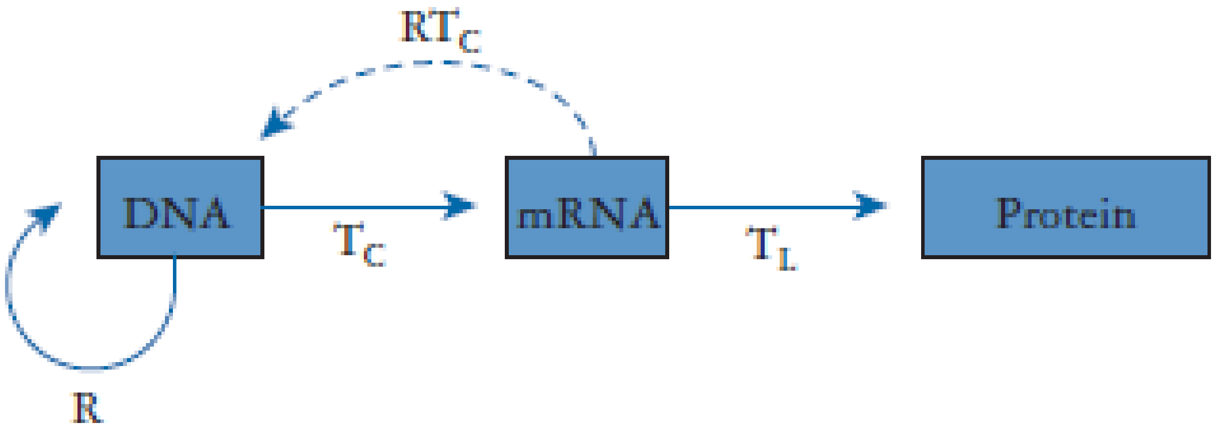


- Double-stranded DNA - primary genetic material of most organisms
- Must be stable (>100 years)
- Must be able to replicate during growth and development
- Limited potential for alteration of genetic materials (mutation) to enable evolutionary pressures to exert their effects

STABLE, REPLICABLE, MUTABLE

https://doi.org/10.1017/CBO9780511800986.004

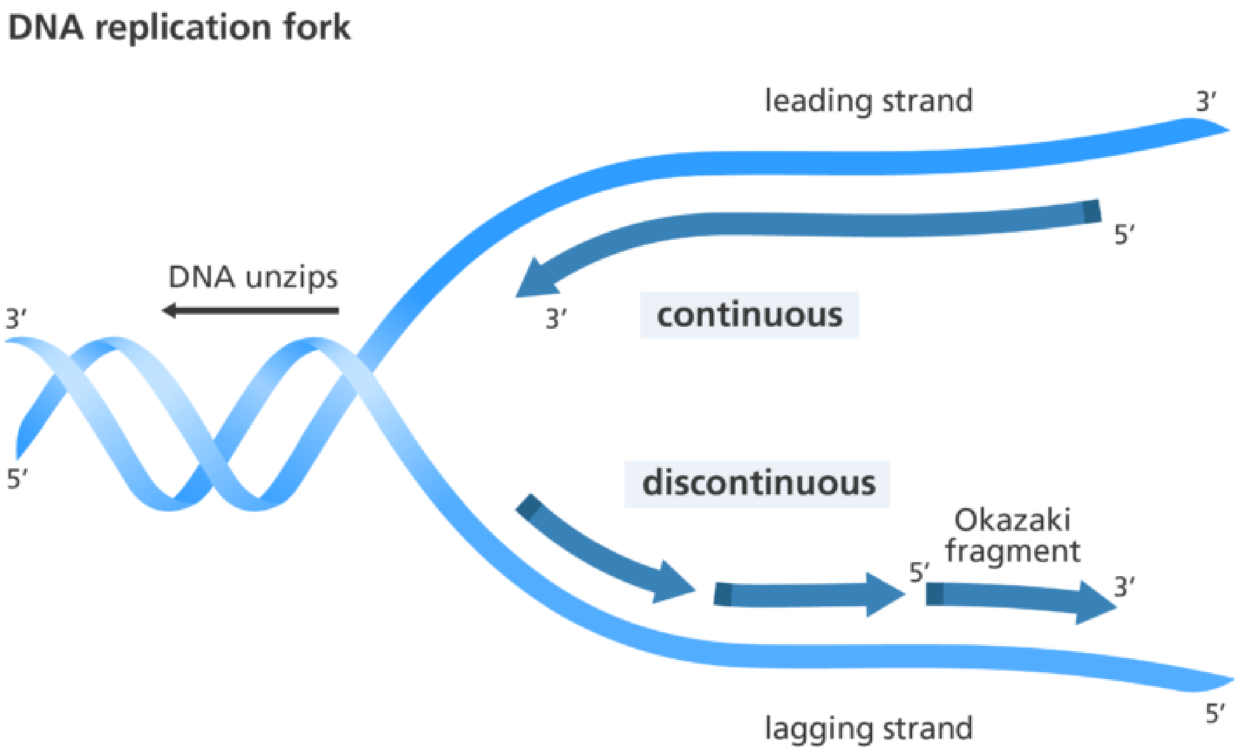
The central dogma



- Unidirectional transfer of information
- DNA's A,G, T, C codes are transferred to RNA's A,G.U, C codes
- Encoding proteins is done in **groups of 3** or ***codons***

Fig. 2.2 The Central Dogma states that information flow is unidirectional, from DNA to mRNA to protein. The processes of transcription (T_C), translation (T_L), and DNA replication (R) obey this rule. An exception is found in retroviruses (RNA viruses), which have an RNA genome and carry out a process known as reverse transcription (RT_C) to produce a DNA copy of the genome following infection of the host cell.

https://doi.org/10.1017/CBO9780511800986.004



Initiation – DNA is unwound by enzymes (helicase)

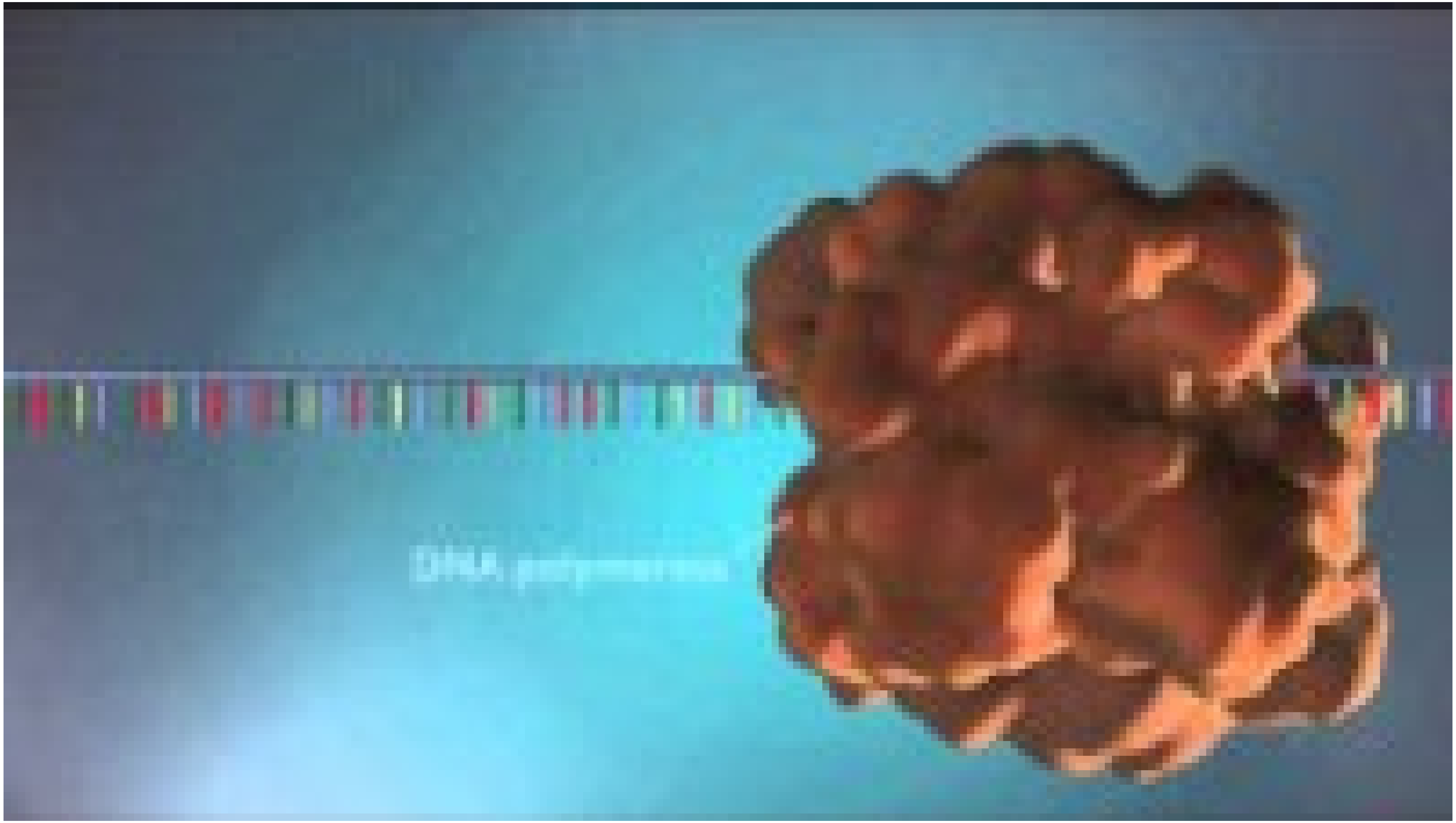
Elongation –

- Leading strand - short piece of DNA (primer) binds to 3’ end (primase); DNA polymerase binds to primer and processes toward the fork, adding complementary bases as it goes

- Lagging strand – DNA polymerize can only copy small lengths at a time, also uses DNA primer to start

Termination – exonuclease removes the primers, gaps filled by complementary bases, and DNA ligase joins the Ozaki fragments together

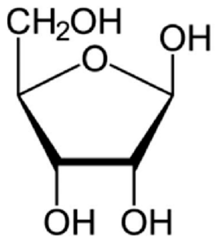
[DNA Replication]



[DNA Replication]

RNA

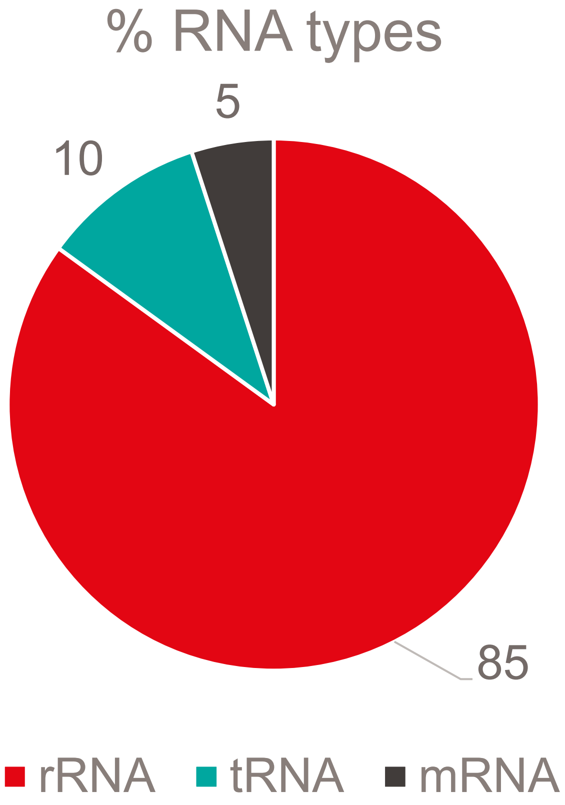
- Ribose instead of 2'-deoxyribose
- Uracil (U) instead of thymine (T)
- Usually single stranded instead of double stranded
- Three main types: messenger RNA (**mRNA**), ribosomal RNA (**rRNA**), and transfer RNA (**tRNA**)



ribose

https://doi.org/10.1017/CBO9780511800986.004

RNA

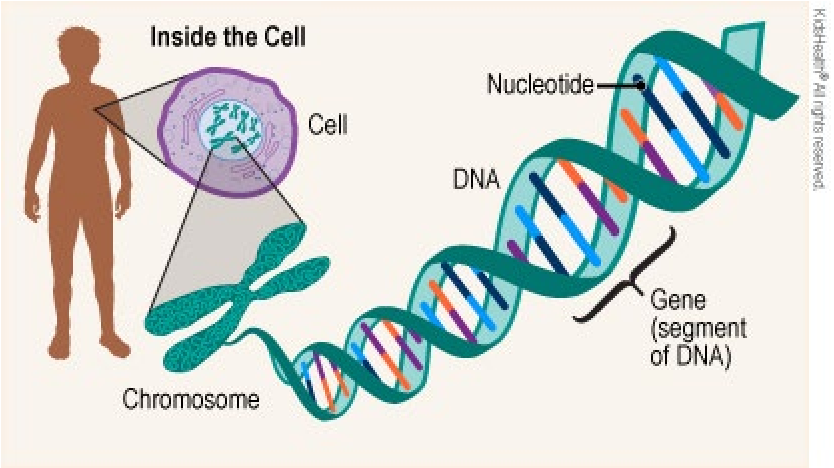


- **rRNA**: *associated with ribosomes, essential part of translational machinery*
- **tRNA**: *provides specificity that enables the insertion of the correct AA into the protein that is being synthesized*
- **mRNA**: *carries genetic material from DNA to translational machinery (ribosome)*

https://doi.org/10.1017/CBO9780511800986.004

Genes

- Basic units of genetic information
- Stretches of DNA that contain the instructions for building biological molecules, like protein
- Determine specific traits by encoding proteins, which carry out various functions in the body
- ‘Gene’ as a term usually represents the genetic information ***transcribed* into *one* molecule of RNA**, which is in turn ***translated* into *one* protein**



- Prokaryotes – in cytoplasm
- Eukaryotes – in nucleus

https://doi.org/10.1017/CBO9780511800986.004

The genetic code

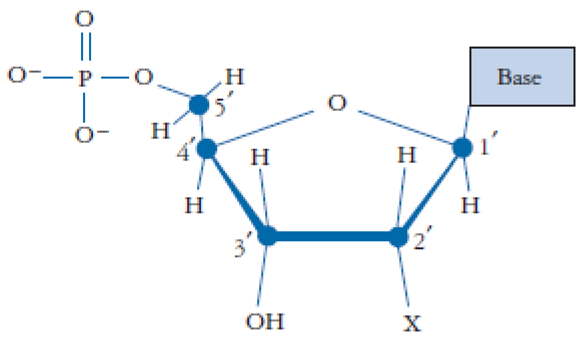
- 20 AA’s but only 4 letters (A,G, C, T)
- Triplet combinations of 4 letters: $4^3 = 64$ possibilities
- Codons are sets of 3 nucleotides (trinucleotides) that specify a given amino acid
- More codons than needed – 3 are ‘STOP’ codons, some amino acids are specified by more than 1 codon (redundancy in code)

https://doi.org/10.1017/CBO9780511800986.004

The genetic code

Table 2.1. The genetic code

First base (5′ end)	Second base U	C	A	G	Third base (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

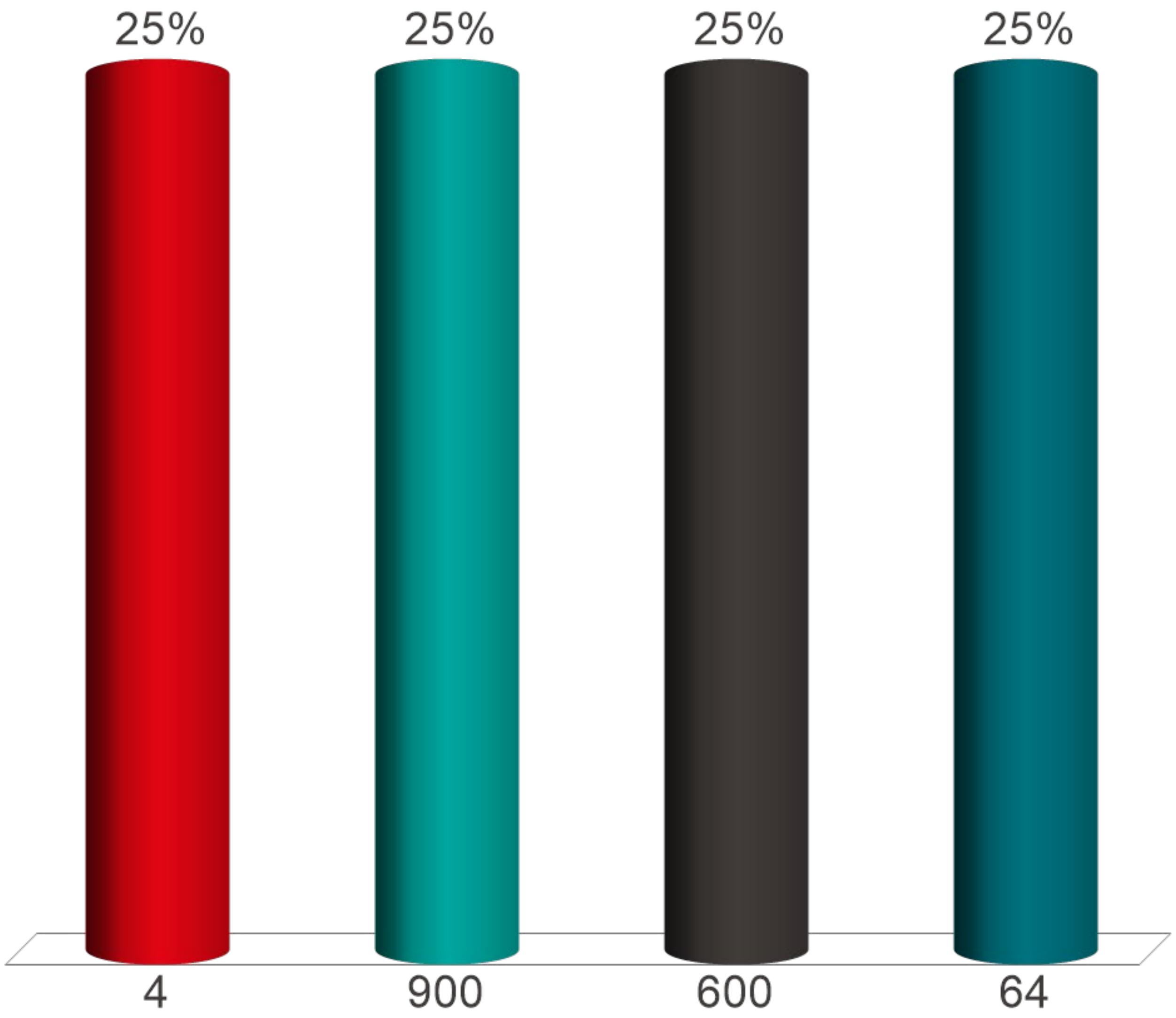


- Read 5’ → 3’
- RNA uses uracil (U) to replace thymine (T)
- Genes are made of sequences of trinucleotides that encode for a specific AA acid sequence, e.g. a protein

https://doi.org/10.1017/CBO9780511800986.004

A protein made up of 300 amino acids needs how many nucleotides at a minimum? go to: ttpoll.eu Session ID: MSE493 (no space)

A. 4
B. 900
C. 600
D. 64



Provide the codons for the ‘STOP’ command. go to: ttpoll.eu
Session ID: MSE493 (no space)

Gene structure (general)

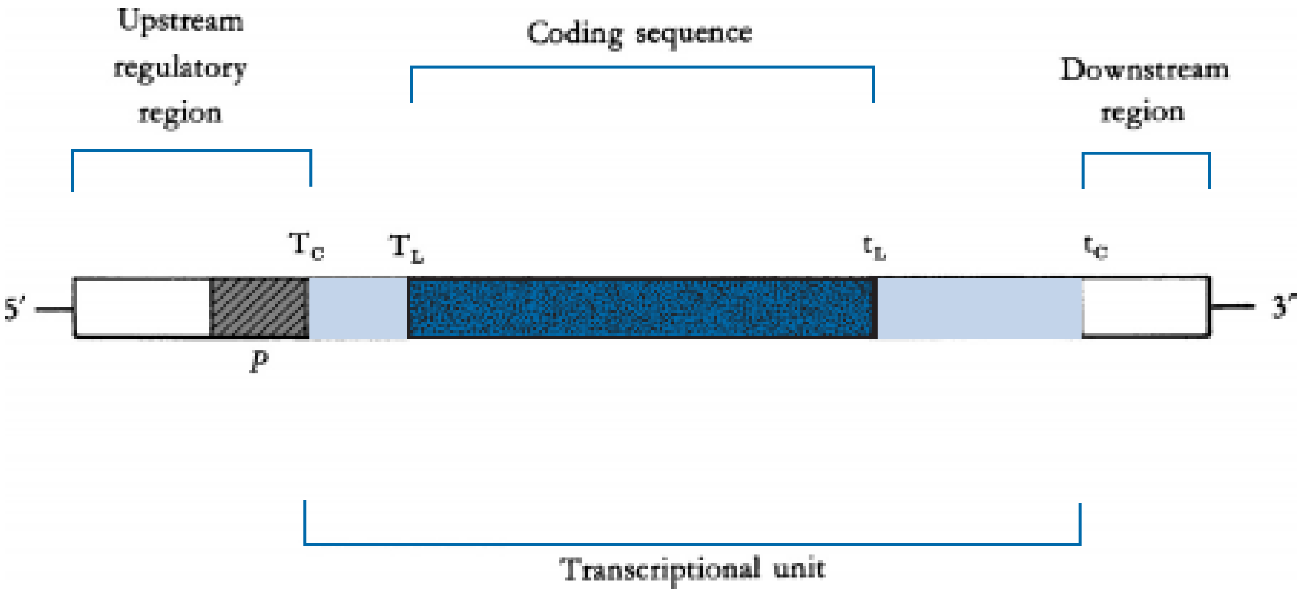


Fig. 2.6 Gene organisation. The transcriptional unit produces the RNA molecule and is defined by the transcription start site (T_C) and stop site (t_C). Within the transcriptional unit lies the coding sequence, from the translation start site (T_L) to the stop site (t_L). The upstream regulatory region may have controlling elements such as enhancers or operators in addition to the promoter (P), which is the RNA polymerase binding site.

- Sequence of bases that specify the codons in the protein coding gene
- Transcription starts at T_C, stops at t_C
- Preceded by a region that binds RNA polymerase, known as the **Promoter** (P)
- **RNA polymerase is the enzyme responsible for transcribing a DNA sequence into an RNA sequence**
- T_L and t_L, start and stop for translation

https://doi.org/10.1017/CBO9780511800986.004

Gene structure in prokaryotes (*lac* operon example)

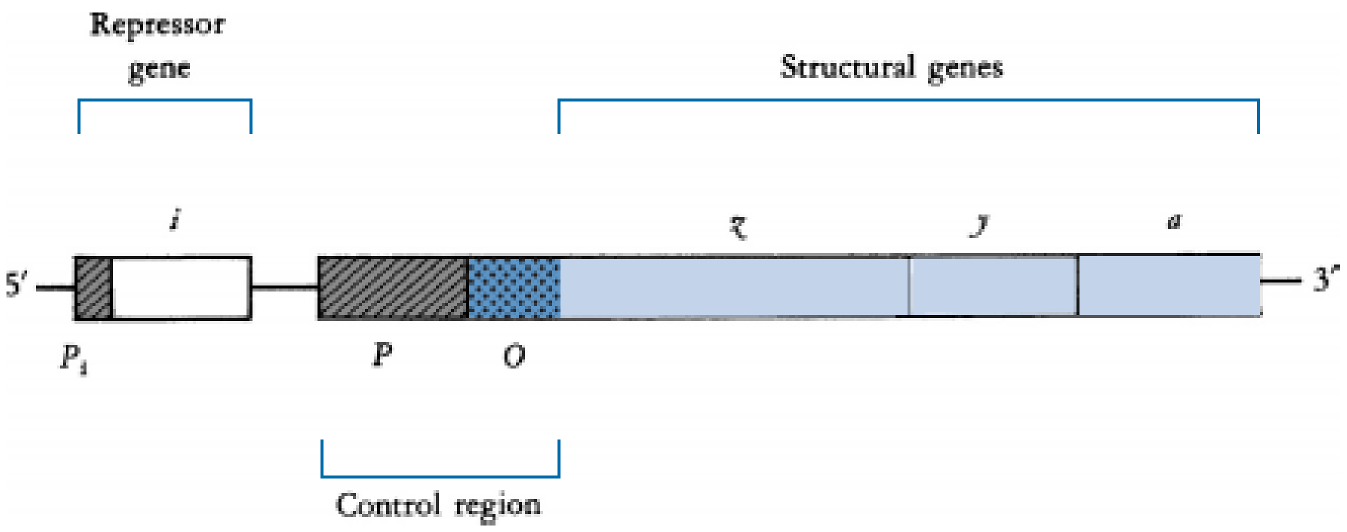
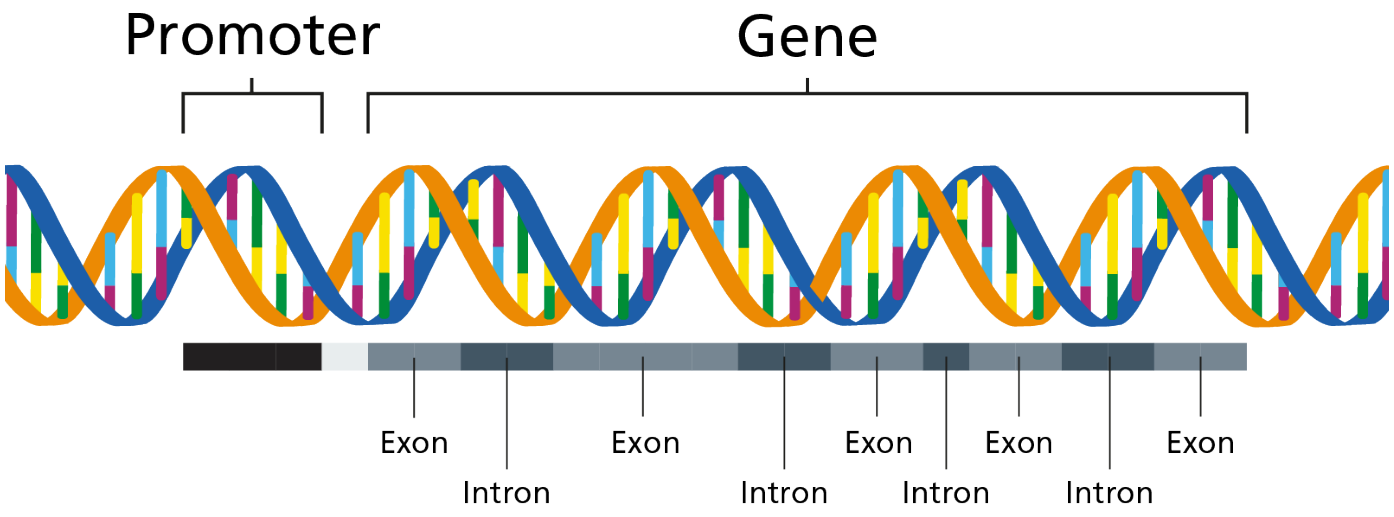


Fig. 2.7 The *lac* operon. The structural genes *lacZ*, *lacY*, and *lacA* (noted as *z*, *y*, and *a*) encode ß-galactosidase, galactoside permease, and a transacetylase, respectively. The cluster is controlled by a promoter (*P*) and an operator region (*O*). The operator is the binding site for the repressor protein, encoded by the *lacI* gene (*i*). The repressor gene lies outside the operon itself and is controlled by its own promoter, P_I.

- Operon is a cluster of genes usually found in prokaryotes, can code for more than 1 protein
- *Lac* operon - codes the enzymes for lactose catabolism
- 3 genes that code for proteins (structural genes)
- Control region with operator (O) and promoter (P) regulatory sites
- O is the binding site for the repressor protein – blocks the binding of RNA polymerase (off switch)
- Repressor gene with its own promoter is outside the operon

https://doi.org/10.1017/CBO9780511800986.004

Gene structure in eukaryotes



[Genomics Ed]

- Transcription takes place in nucleus
- Translation occurs in cytoplasm
- More complex in general than for prokaryotes (also contain genetic info in mitochondria and chloroplasts)
- Extra pieces of DNA that do not appear in mRNA encoded by gene – “introns”
- Sequences that make up mRNA are “exons” (contain protein-coding information)
- Introns contain important regulatory information but no protein-coding information

Gene structure in eukaryotes

Table 2.2. Size and structure of some human genes

Gene	Gene size (kbp)	Number of exons	% exon
Insulin	1.4	3	33
β-globin	1.6	3	38
Serum albumin	18	14	12
Blood clotting factor VIII	186	26	3
CFTR (cystic fibrosis)	230	27	2.4
Dystrophin (muscular dystrophy)	2400	79	0.6

Note: Gene sizes are given in kilobase pairs (kbp). The number of exons is shown, and the percentage of the gene that is represented by these exons is given in the final column.

- Introns need to be removed before mRNA can be translated
- Introns are spliced out in the nucleus
- After RNA processing in nucleus, fully functional mRNA is exported to the cytoplasm for translation

https://doi.org/10.1017/CBO9780511800986.004

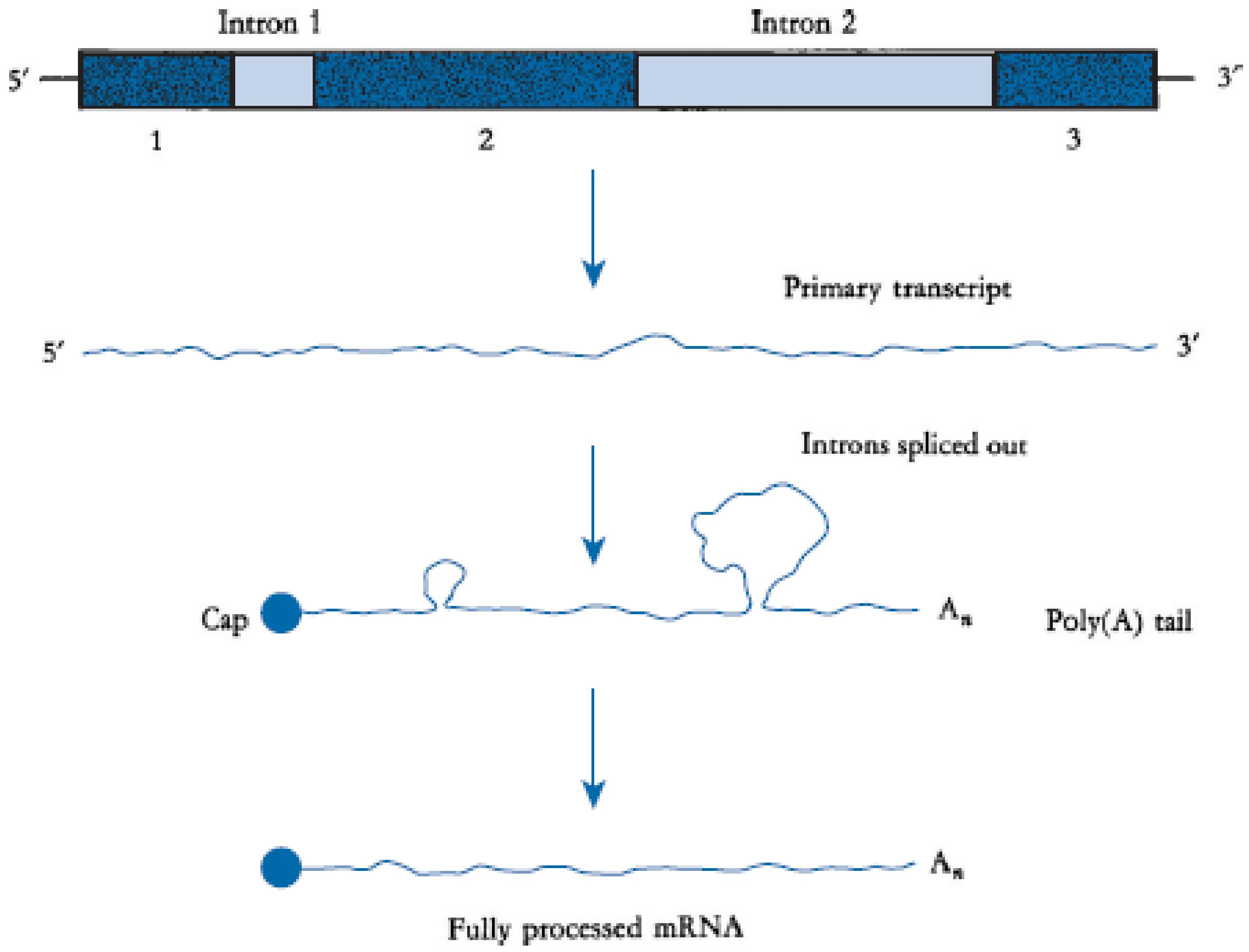
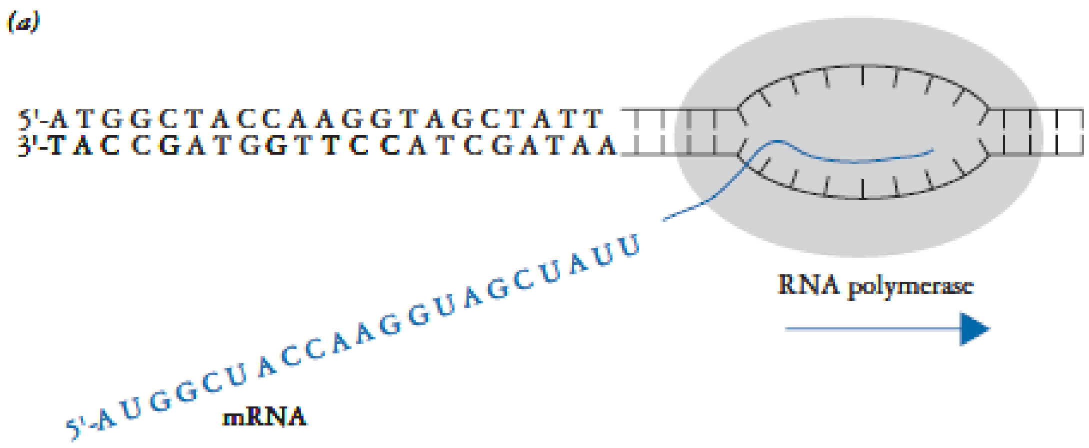


Fig. 2.8 Structure and expression of the mammalian ß-globin gene. The gene contains two intervening sequences or introns. The expressed sequences (exons) are shaded and numbered. The primary transcript is processed by capping, polyadenylation, and splicing to yield the fully functional mRNA.

https://doi.org/10.1017/CBO9780511800986.004

https://doi.org/10.1017/CBO9780511800986.004

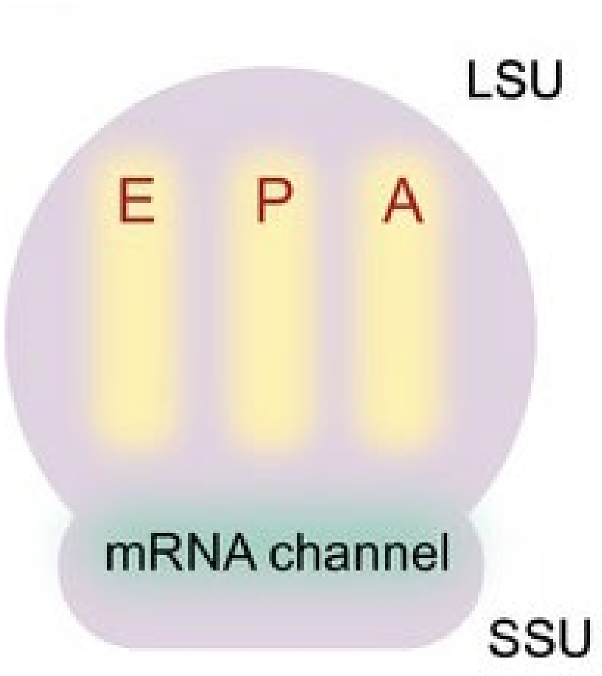


1. DNA has two complementary strands, one that is templating and one that is non-templating, the templating strand is used to code the mRNA. The templating strand is also referred to as the non-coding strand.

2. In the nucleus, RNA polymerase reads the nucleotides and transcribes them into mRNA

3. mRNA transported to cytoplasm for translation

https://doi.org/10.1017/CBO9780511800986.004

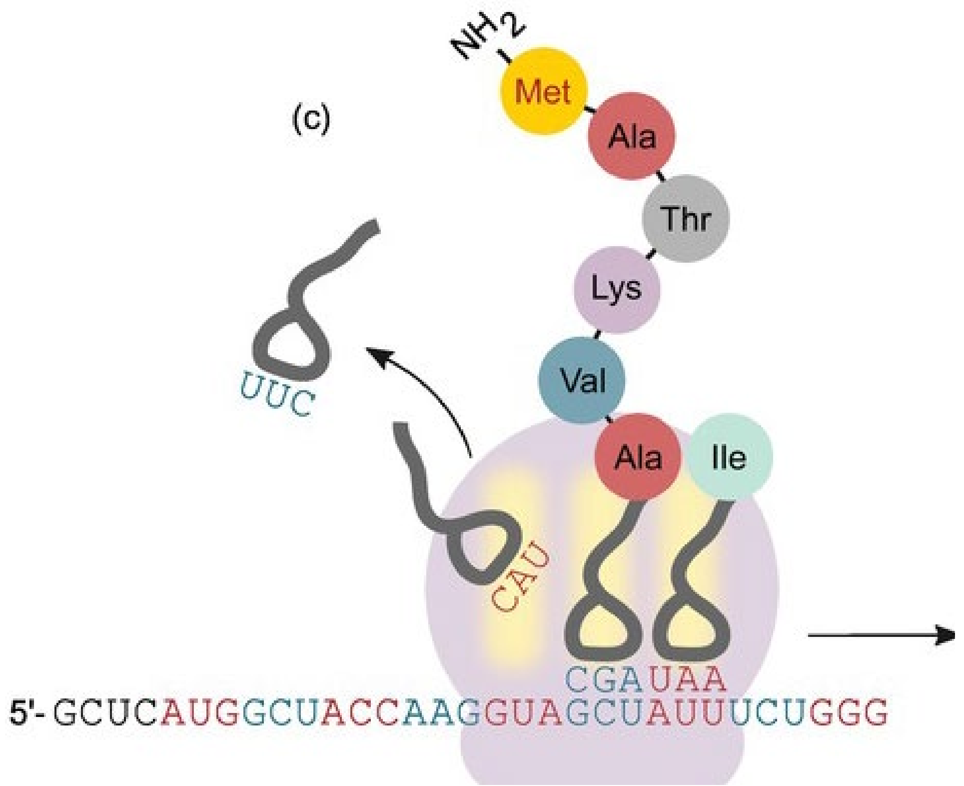


In the cytoplasm, the mRNA is read by the ribosomes in sets of codons. tRNA has anticodons that ensure that it brings the correct amino acid to the ribosome so it can be linked into a polypeptide chain

A and P sites are involved in the insertion of the correct tRNA-AA complex in the growing polypeptide chain (codon/anti-codon)

E site (exit) facilitates the removal of tRNA after peptide bond formation has removed it’s AA

https://doi.org/10.1017/CBO9780511800986.004



- mRNA translation: AA is inserted into the protein via a codon/ anti-codon event in ribosome
- 1st AA residue is encoded by AUG in mRNA (tRNA anti codon UAC) = methionine
- Translation occurs from 5' →3', growing from N-terminus

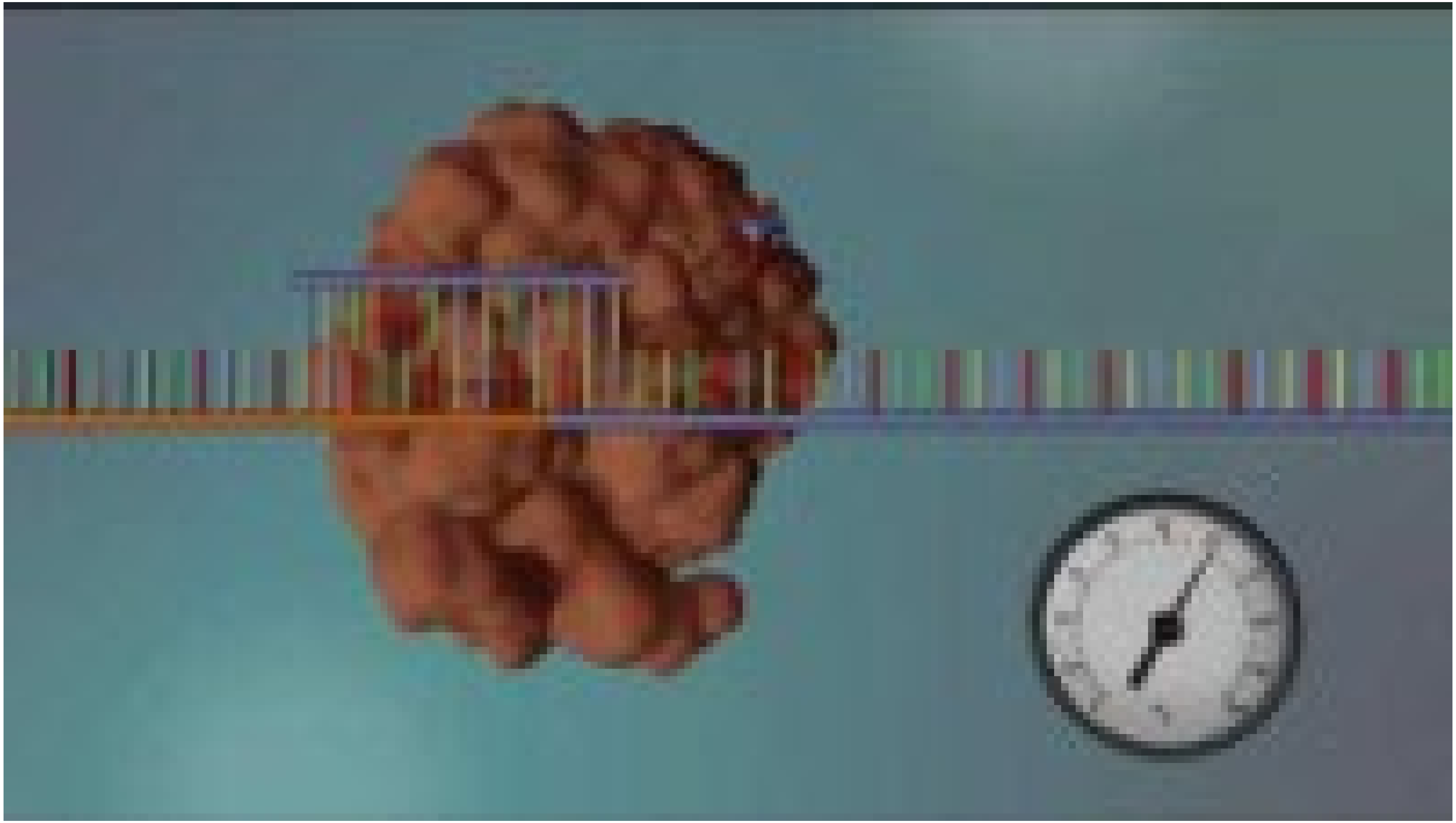
[From DNA to protein]

Gene expression - complex

Adaptive regulation	Developmental regulation	Developmental regulation
• Gene expression in response to environmental changes or stressors • Flexible and reversible • Survival in changing environments • Shorter time scales • Example: *lac* operon in bacteria, presence of lactose induces the expression of genes used in its metabolism	• Control of gene expression throughout the developmental stages. • Genes turned on or off at specific times during growth and differentiation to produce correct cell types (e.g., muscle cells, neurons, blood cells) • More rigid and permanent • Longer periods • Growth, development, maturation	"Housekeeping genes"; always needed, e.g., to encode proteins that are essential at all times

https://doi.org/10.1017/CBO9780511800986.004

DNA sequencing



Recombinant DNA technology

General idea:

- Used to alter the genetic make up of an organism
- Combines DNA from different sources to create sequences that do not naturally occur
- Can be used to get a host to produce a non-native trait or protein

Steps:

(1) Identify target gene

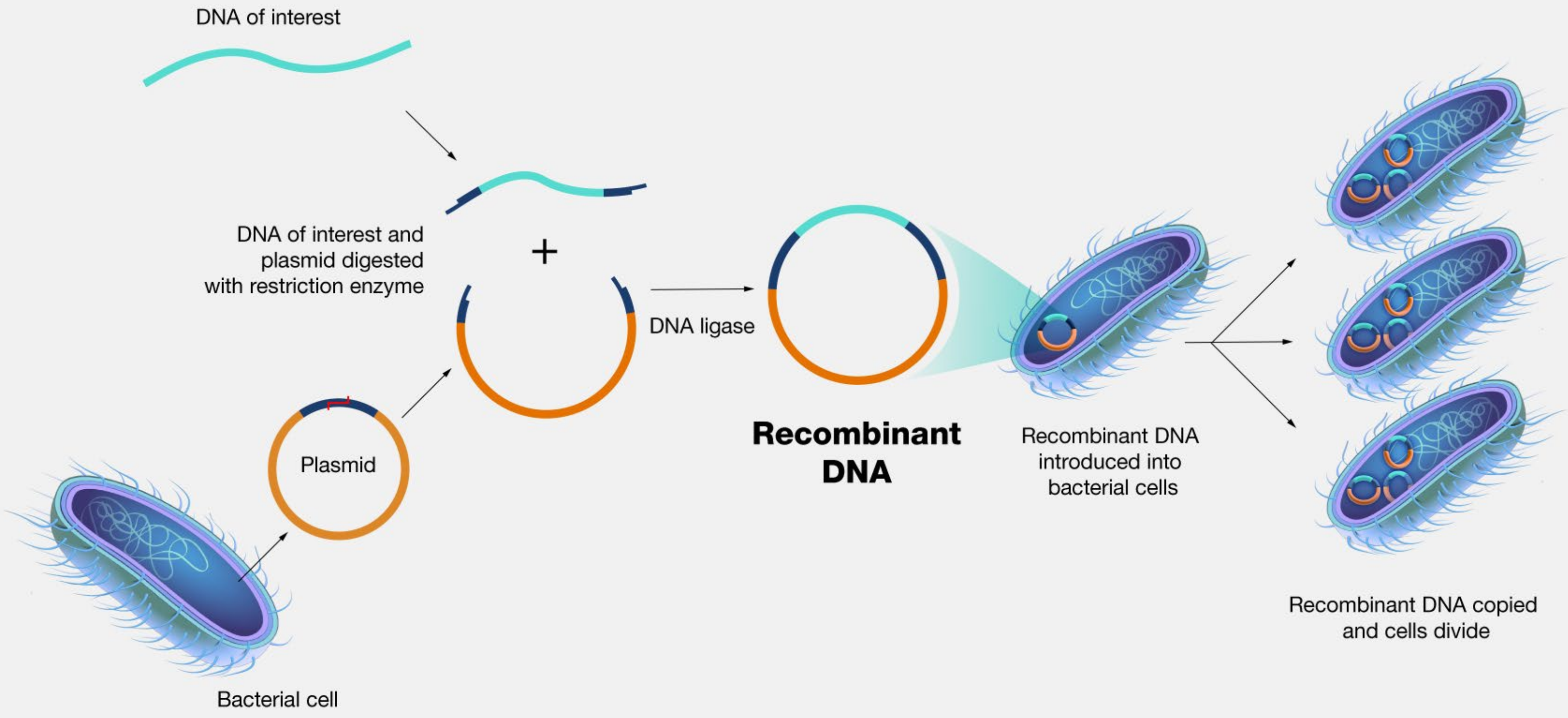
(2) Cut DNA with restriction enzymes

(3) Insert target gene into vector (carrier molecule, usually a plasmid), join with DNA ligase; now you have recombinant DNA

(4) Introduce rDNA into host organism

(5) Host organism can express new gene to produce, e.g., a protein of interest

Recombinant DNA technology



[Recombinant DNA technology]

Tools for gene editing - Recombinant DNA Technology

[Recombination]

What is CRISPR?

- CRISPR – “Clustered Regularly Interspersed Short Palindromic Repeats”
- DNA sequences in bacteria that are from past viral infection
- CRISPR DNA is coupled to a molecular scissor (Cas9)
- CRISPR-Cas9 defends bacteria against subsequent viral infection by recognizing the viral sequence, cutting the viral DNA, and deactivating the virus
- Part of the adaptive immune system of bacteria

https://www.nobelprize.org/uploads/2020/10/popular-chemistryprize2020.pdf

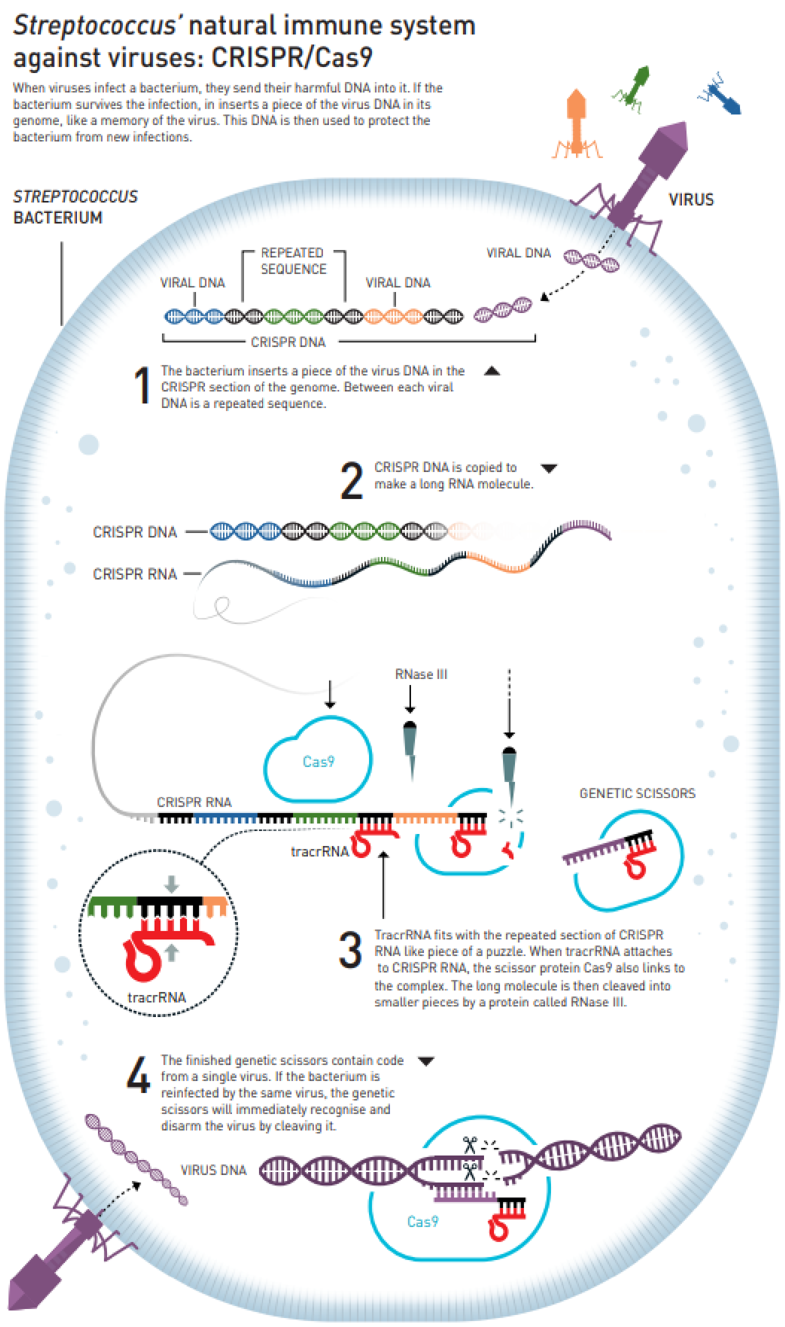
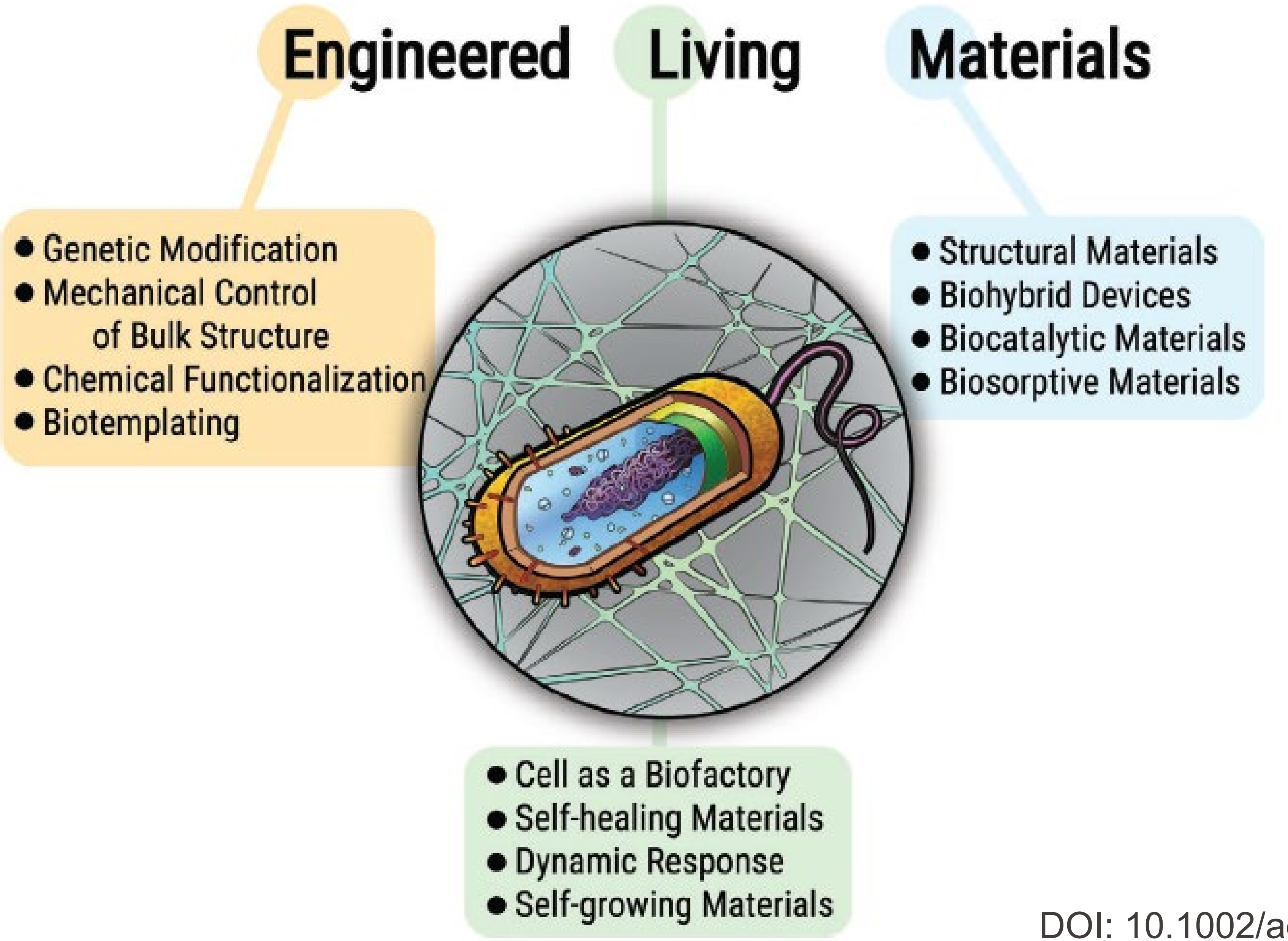


Figure 2

2020 Nobel Prize: CRISPR-Cas9, Emmanuelle Charpentier Jennifer A. Doudna

[CRISPR-Cas9 Story]

Tools for gene editing - CRISPR-CAS9



DOI: 10.1002/adma.201704847

OK, so what is an ELM?

- Directs the formation of the material itself, under complex conditions containing many components, including nutrient feedstocks and metabolic wastes
- Allows for stimuli-responsiveness over the course of the lifetime of the cells (adaptive gene expression)
- If desired, can be processed or programmed to kill the living cells at a given time, keeping the material intact (no concern for maintenance or potential biohazard threats)

DOI: 10.1002/adma.201704847

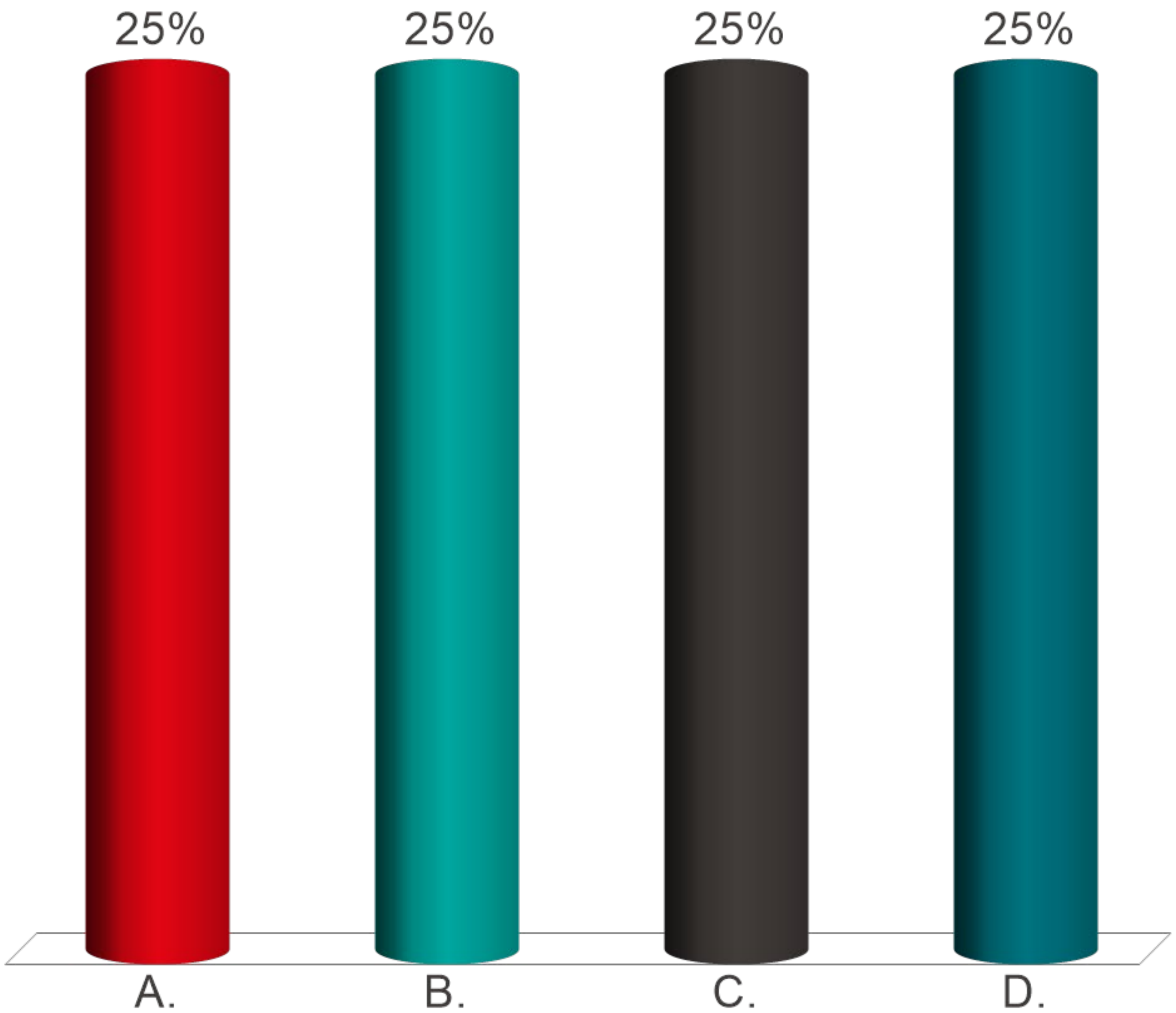
OK, so what is an ELM? (Strict)

- Composed of living cells that ***form or assemble*** the material or ***modulate the functional performance*** of the material
- The living cells in ELMs are ***materials factories***, *drawing energy from their environment* to create biopolymer building blocks and guide the formation and maintenance of the desired material
- "Engineered": genetic, spatial/mechanical engineering
- Composition: mainly cellular biomass or secrete the materials that form a large part of their structure (e.g. ***biofilm***), can integrate other components, such as polymers (organic, inorganic), particles, scaffolds as part of the assembly process

DOI: 10.1002/adma.201704847

What is a microbe? go to: ttpoll.eu
Session ID: MSE493 (no space)

A. Disease-causing organism
B. Single-celled organism
C. Microscopic, single or multi-cellular
D. Imaginary



The Microbes

In this section, we survey the representatives of microbes and the main methods to view them. Microbial diversity is the range of different kinds of unicellular organisms, including bacteria, archaea, protists, and fungi. Various microbes thrive throughout distinct strata of the biosphere. The resident microbes in any given biosphere often define the limits of life and create conditions conducive to the survival and evolution of other living beings. Entries include algae, eukaryotes and prokaryotes, and bacteria.

- Human genome project
  1990-2003

- Human microbiome project
  2007-2016

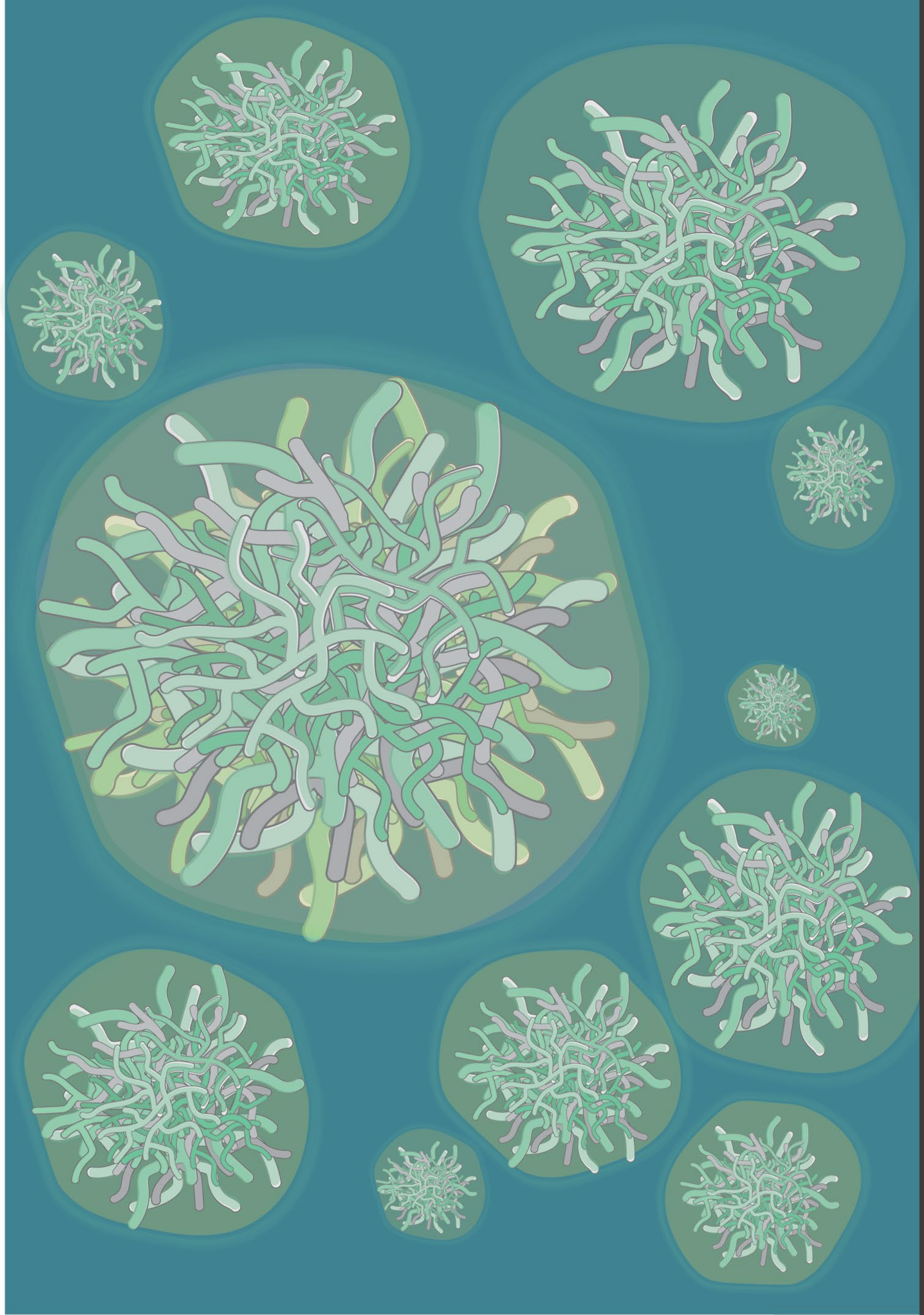
Genome project
Human microbiome project

Principles of Microbiology

EPFL

Microbes aka microorganisms (examples)

Microbe examples	Description	Strategies for obtaining nutrients
Bacteria	Single-celled organisms that lack a nucleus	Mostly heterotrophs
Fungi	Includes yeasts and molds, which can be single-celled or multicellular	Mostly heterotrophs
Algae	Simple, typically aquatic organisms that can conduct photosynthesis, including both single-celled and multicellular forms	Mostly autotrophs - photosynthesis



See you next week!