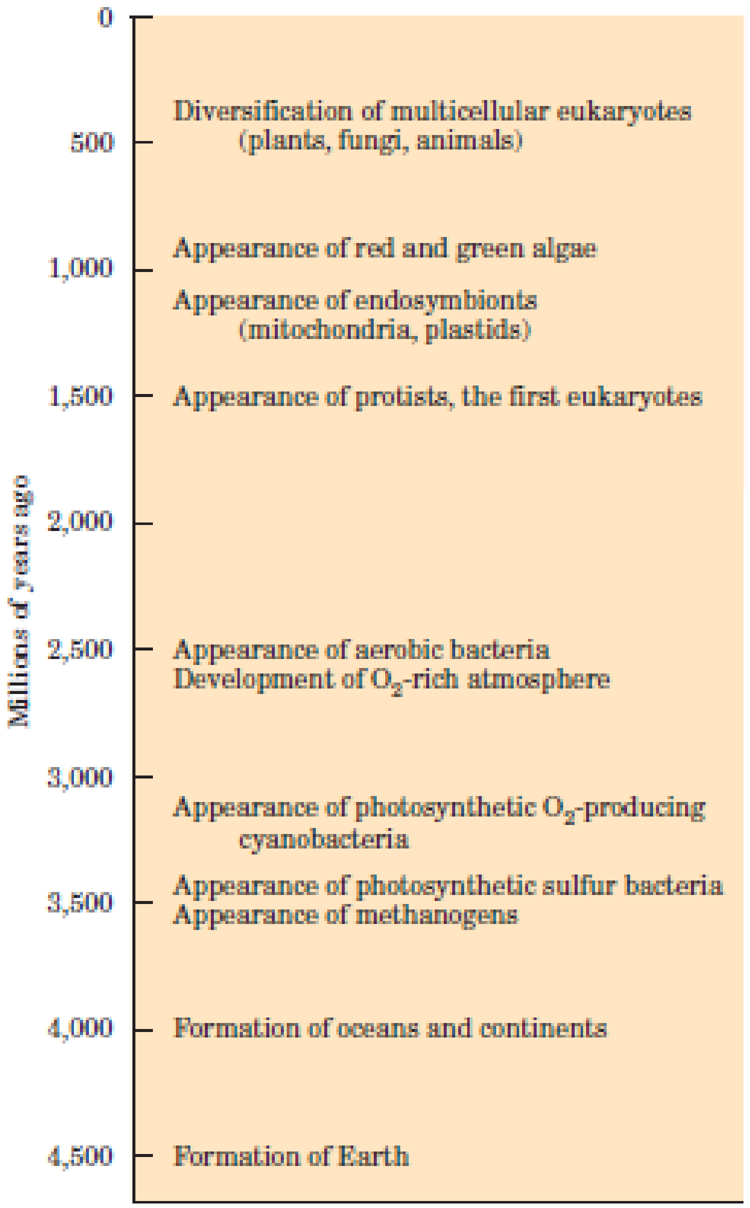
What is Life?

What are the principles of Biology ?

Landmarks in the evolution of life on Earth



Life's Origins

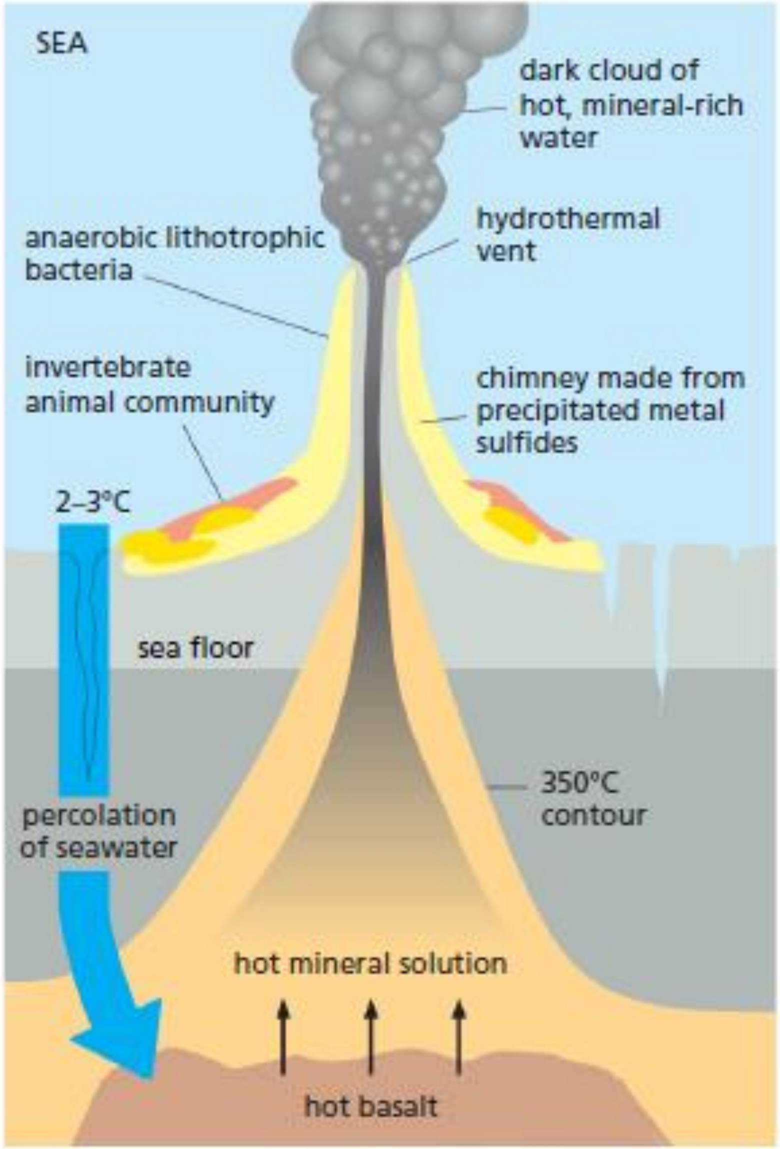
Today, research into the origin of life is interdisciplinary with workers trying to answer four main questions:

1. What was the Earth's physical environment like when life first evolved?

2.What sorts of chemical reactions could produce the building blocks of life and could these occur naturally in the early Earth's environment?

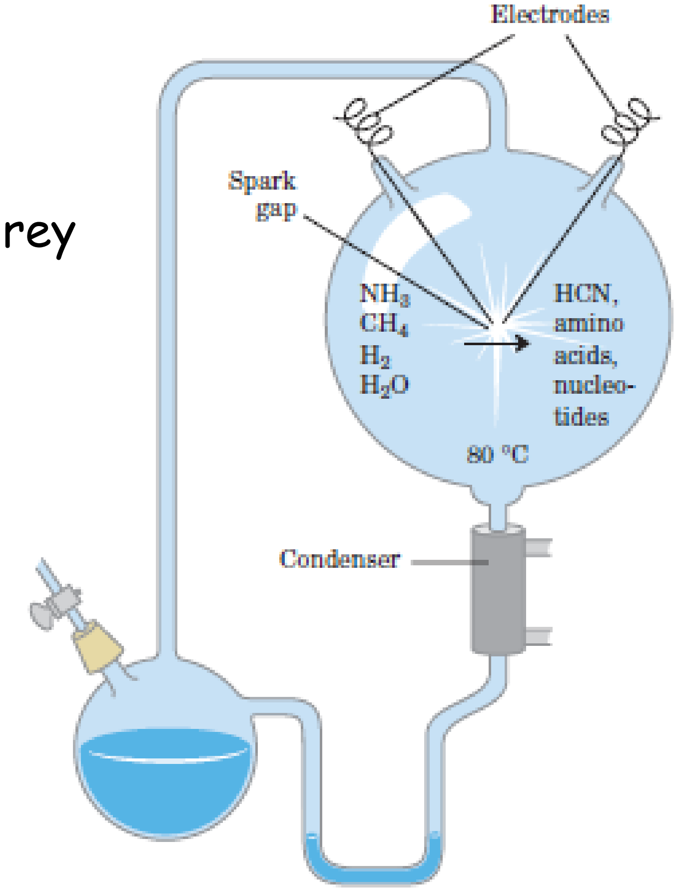
3. How could the complex organic molecules be compartmentalized into a contained unit?

4. How did the genetic code evolve?


SEA
dark cloud of hot, mineral-rich water
hydrothermal vent
anaerobic lithotrophic bacteria
invertebrate animal community
chimney made from precipitated metal sulfides
2–3°C
sea floor
350°C contour
percolation of seawater
hot mineral solution
hot basalt

How did the first living organisms acquire their characteristic organic building blocks?

Miller and Urey Experiment



Hypothesis:

Effects of powerful atmospheric forces—ultraviolet irradiation, lightning, or volcanic eruptions—on the gases in the prebiotic Earth's atmosphere, and on inorganic solutes in superheated thermal vents deep in the ocean.

The appearance of the first living cell???

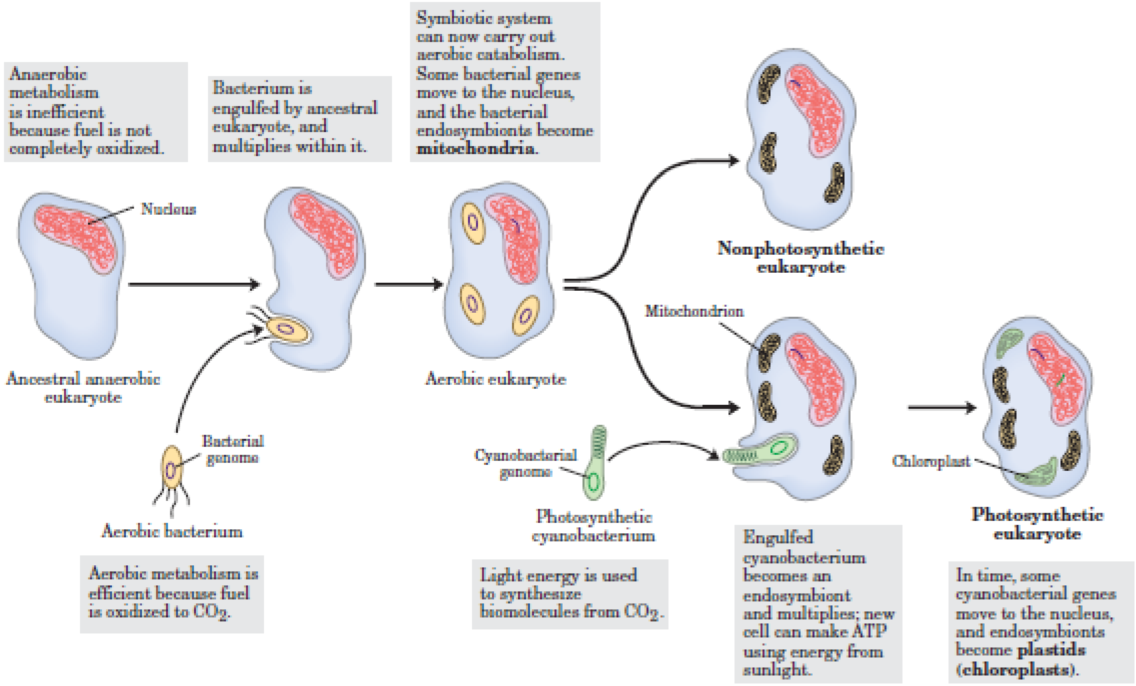
The First Cell Used Inorganic Fuels???

Lynn Margulis and the Theory of Endosymbiosis

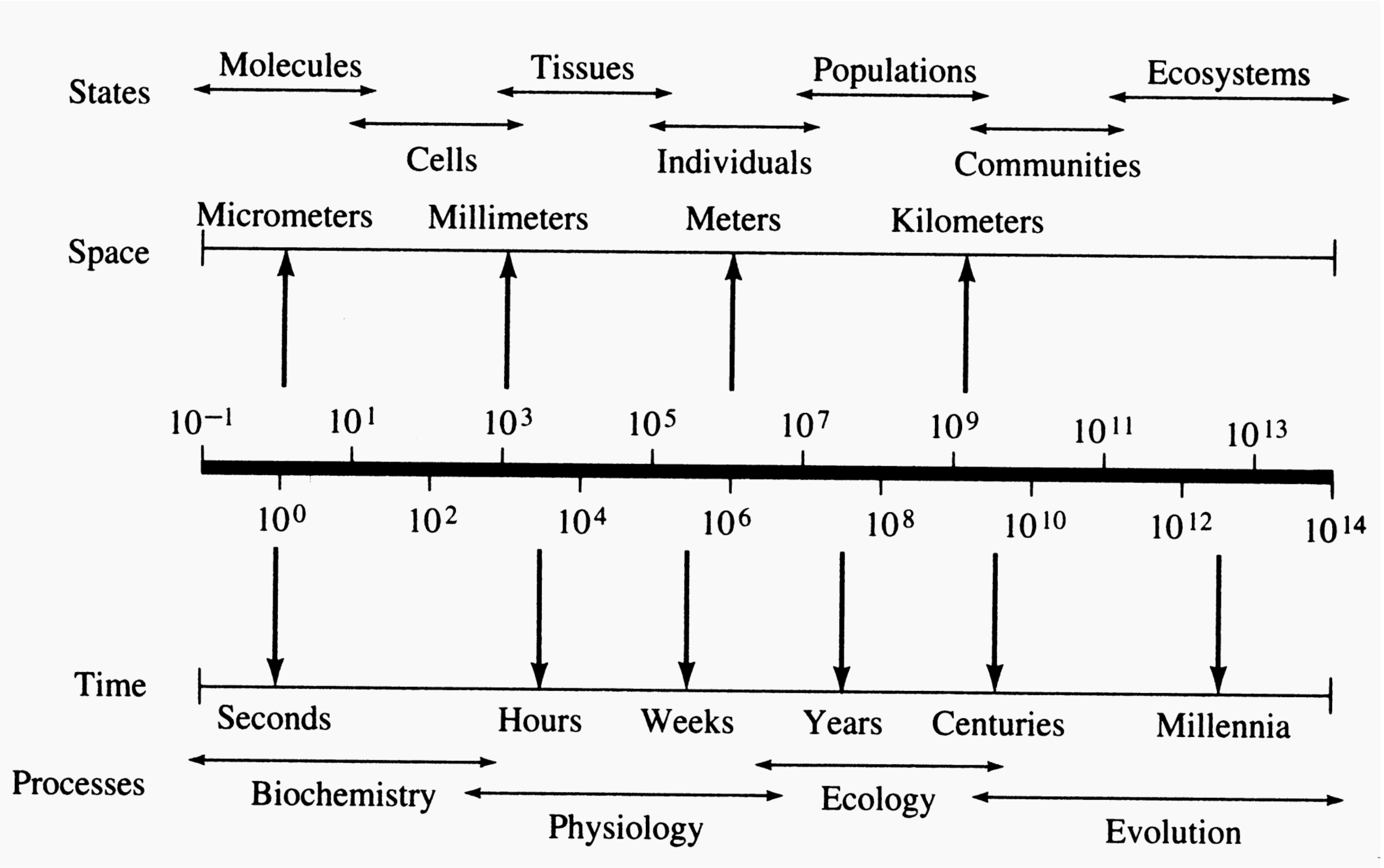
- Eukaryotic cells originated from a series of endosymbiotic events involving multiple prokaryotes

- This idea was considered outrageous at the time (1967), but many of Margulis's ideas have since become widely accepted

- Mitochondria are the descendants of oxygen-respiring bacteria and chloroplasts were originally photosynthetic bacteria, are almost universally accepted

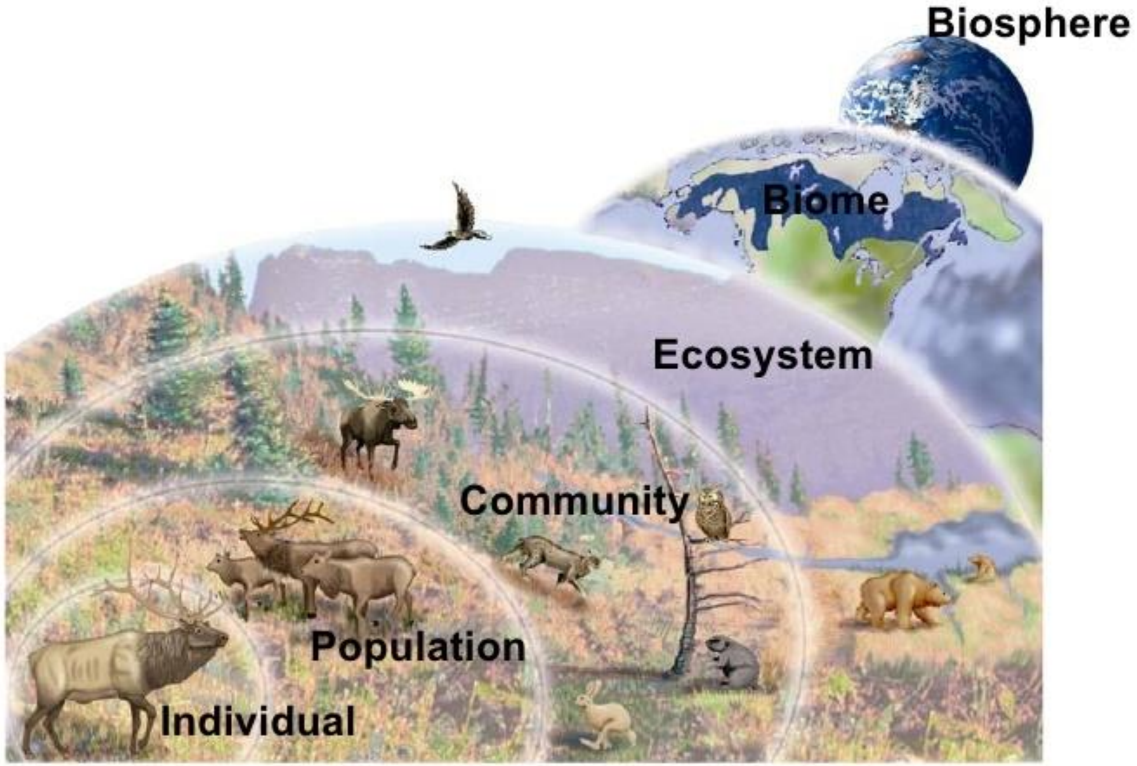


Biology is a vast discipline



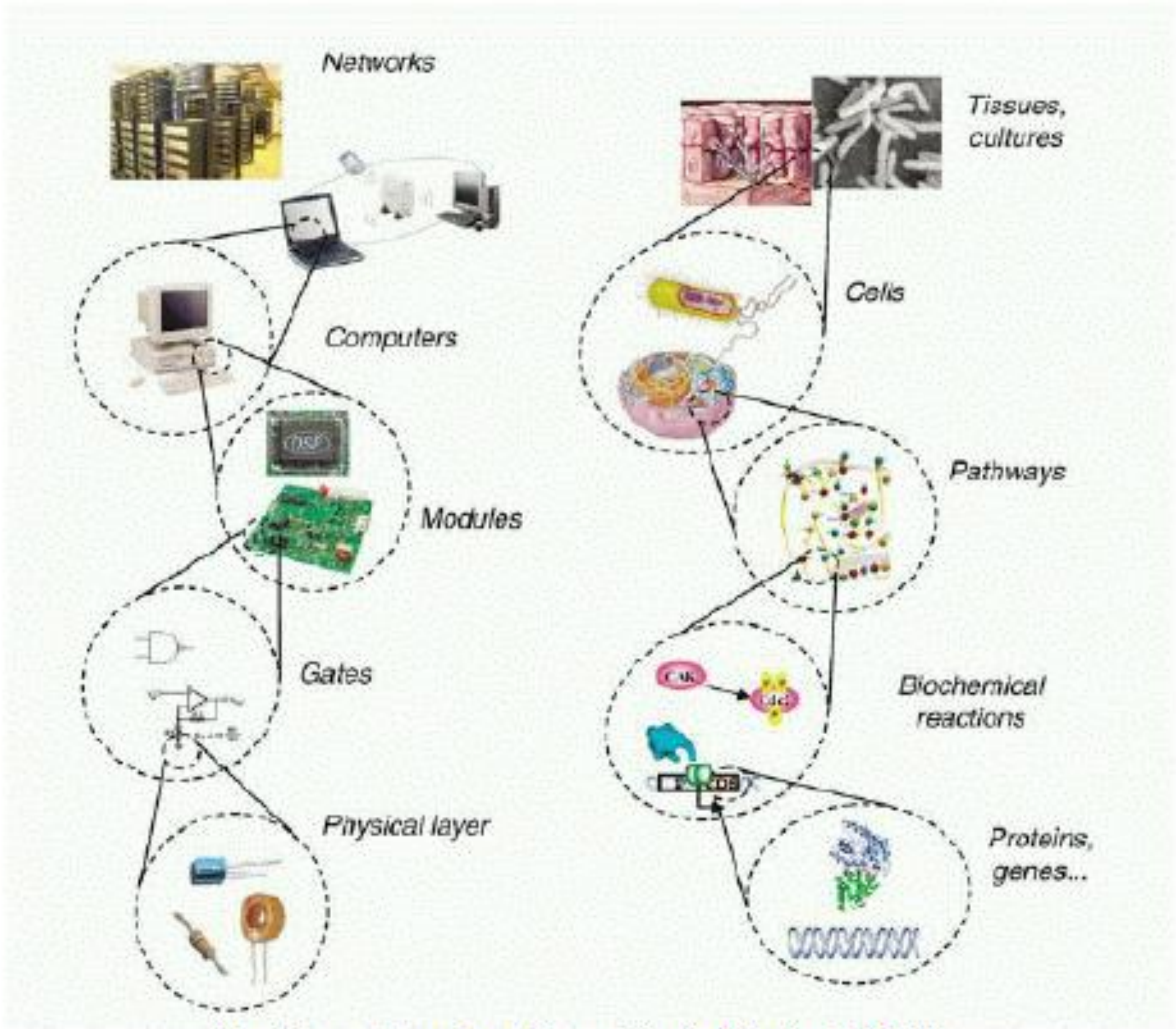
The Hierarchical Structure of Life

Levels of Organization



Individual = 1 Species
Population = Many of the same species
Community = Different populations (Biotic Factor = living)
Ecosystem = Various populations along with abiotic factors (non-living) coexisting.
Biome = Many ecosystems (Tundra, Tropical Rain Forest, Desert, etc...)
Biosphere = Many biomes

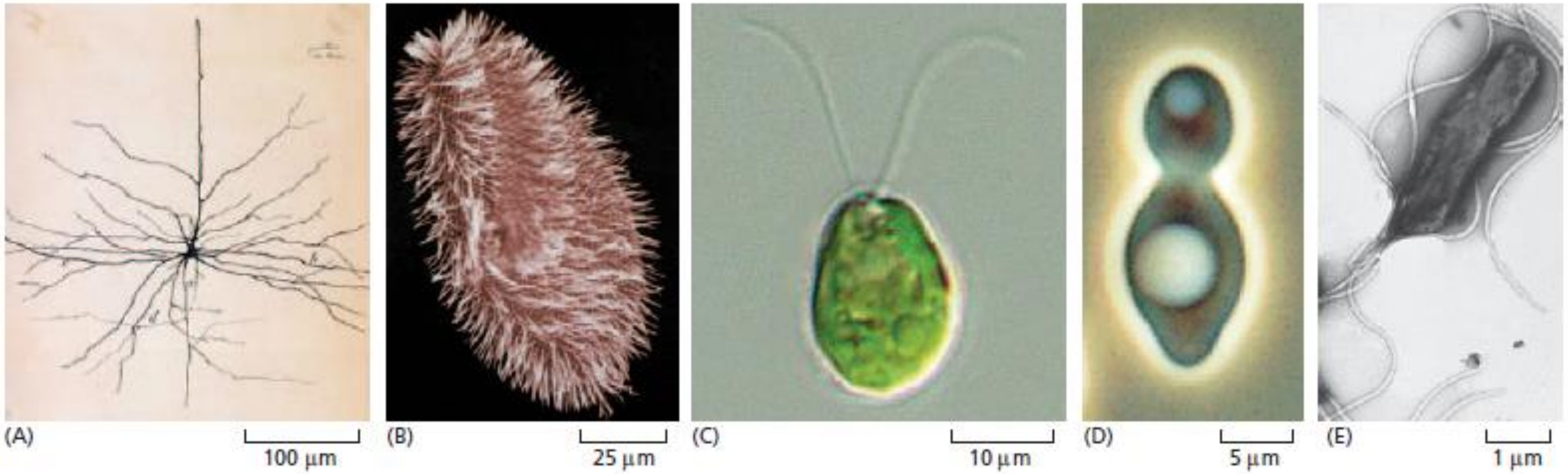
Cells compute?



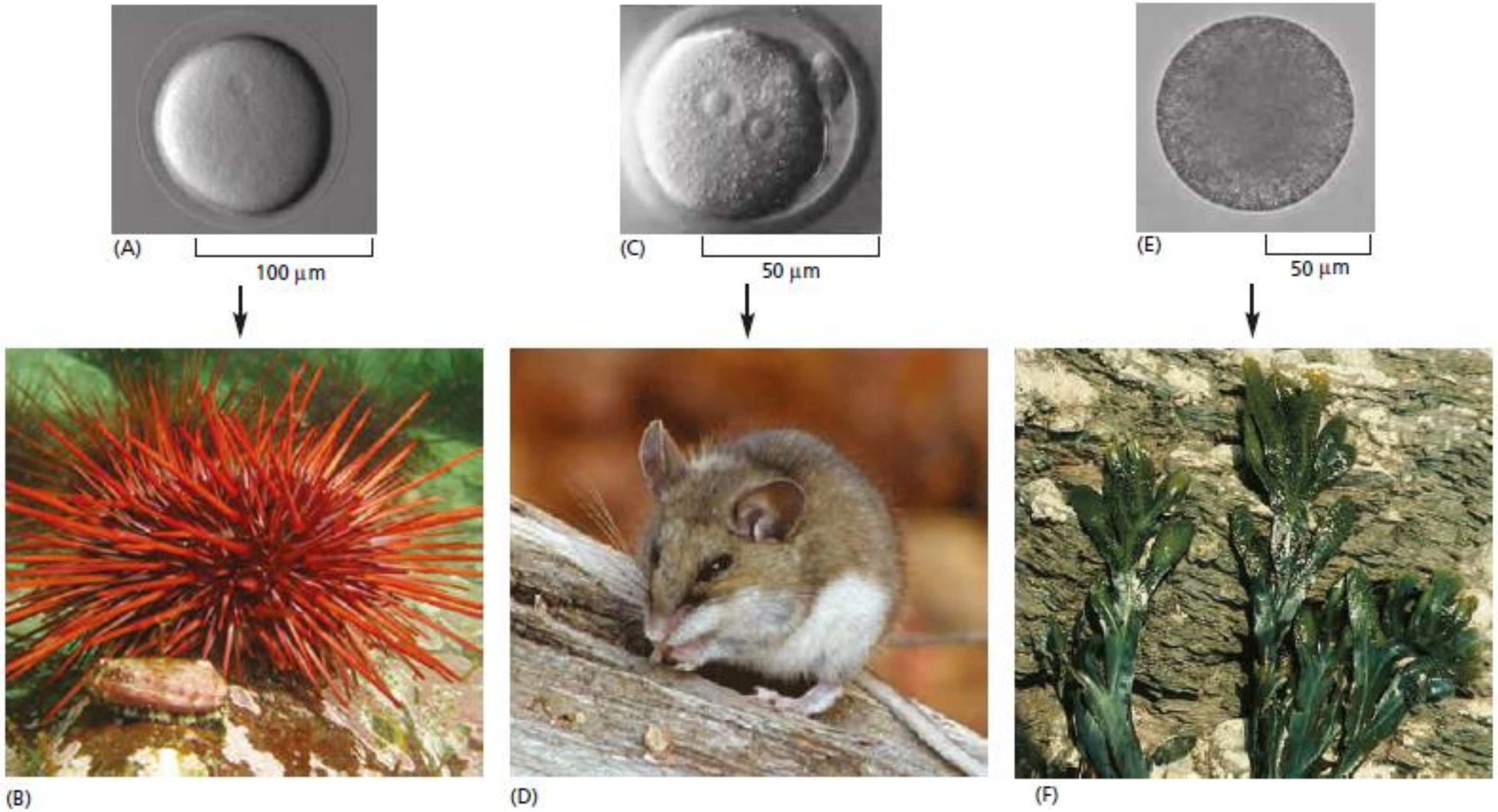
(Andrianantoandro, Basu, Karig, Weiss (2006))

Can we decipher the biological hardware and software?

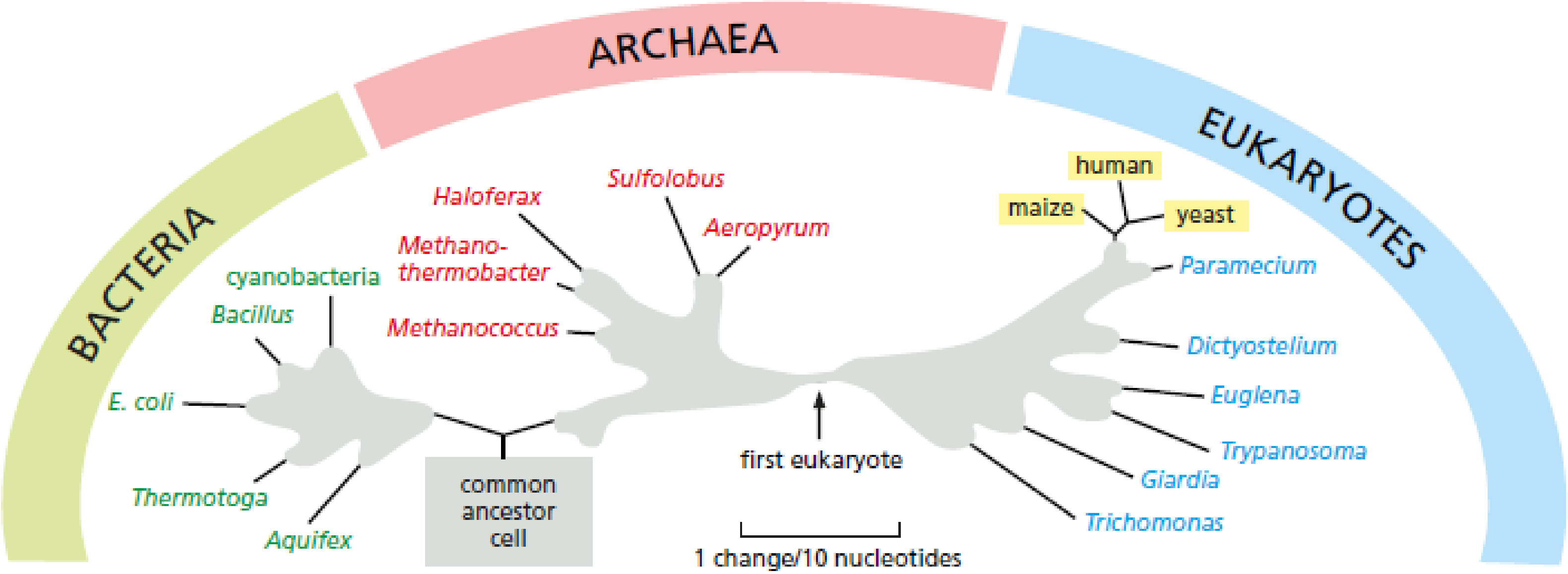
Cells come in variety of shapes and sizes



Origin from single cell

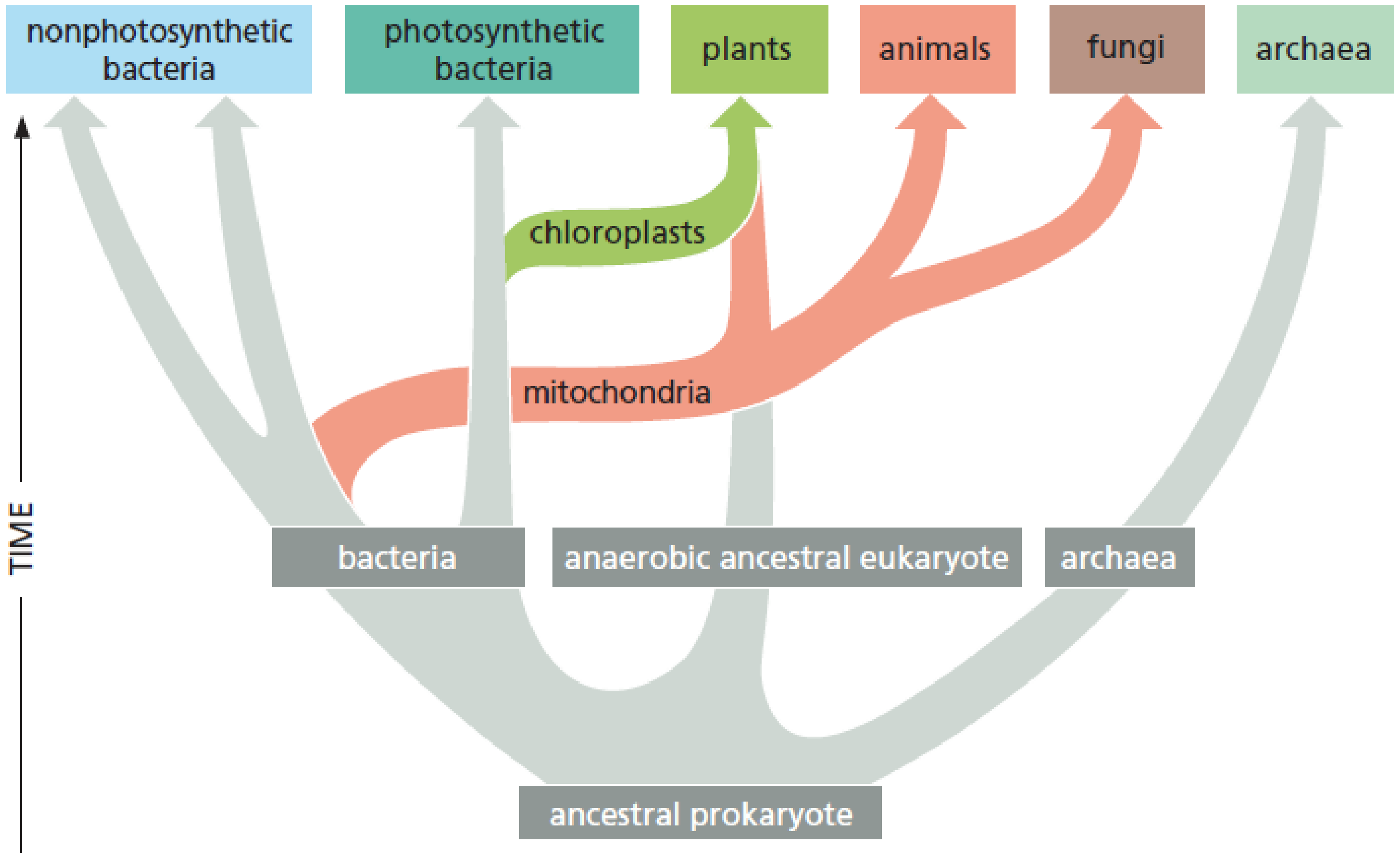


The Tree of Life



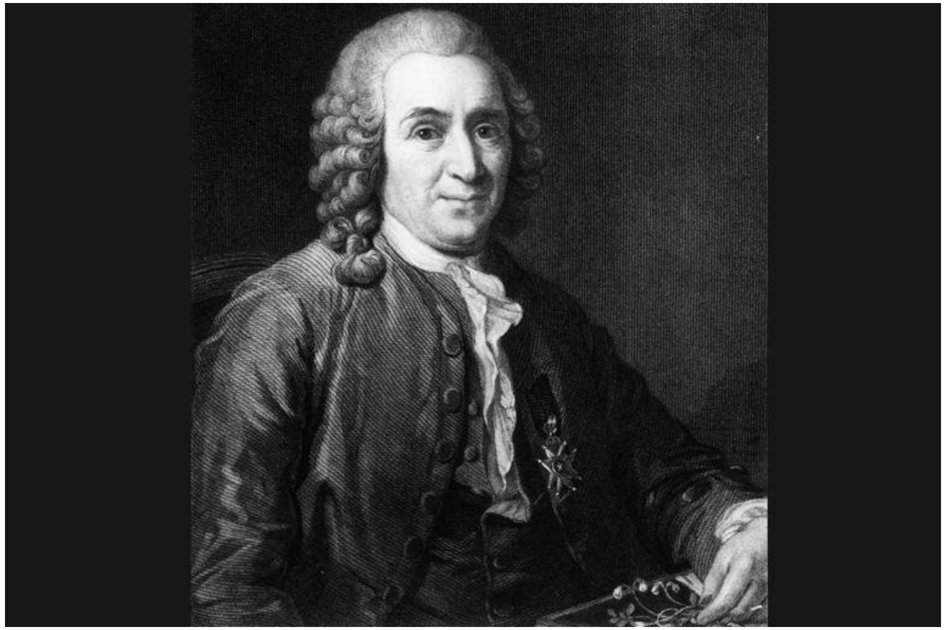
The domain system of classification was developed by Carl Woese and places organisms under three domains: **Archaea**, **Bacteria**, and **Eukarya**.

Under the domain system, organisms are further grouped into six Kingdoms. The Kingdoms include: **Archaebacteria** (ancient bacteria), **Eubacteria** (true bacteria), **Protista**, **Fungi**, **Plantae**, and **Animalia**.



nonphotosynthetic bacteria
photosynthetic bacteria
plants
animals
fungi
archaea
chloroplasts
mitochondria
TIME
bacteria
anaerobic ancestral eukaryote
archaea
ancestral prokaryote

Taxonomy and Organism Classification



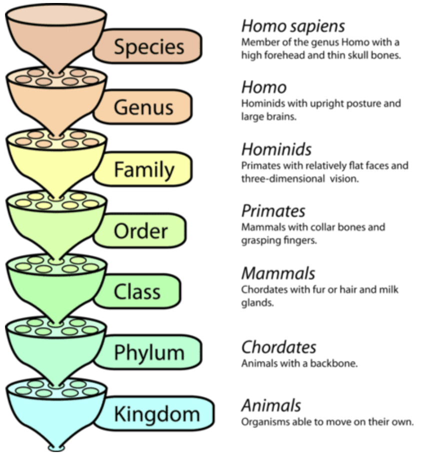
Botanist Carl von Linnaeus (1707-1778), founder of the modern system of binomial nomenclature for plants

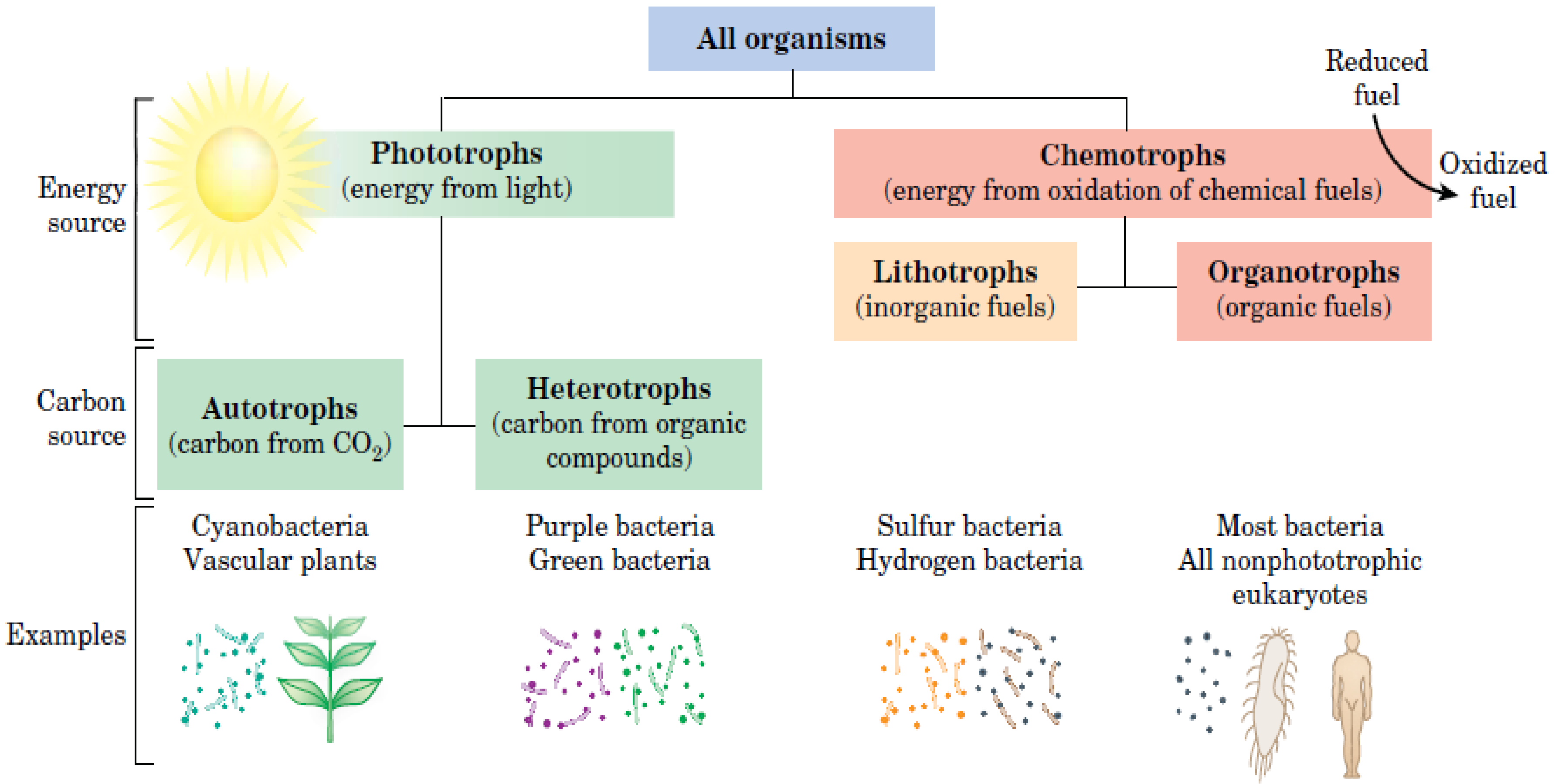
Kingdom, **Phylum**, **Class**, **Order**, **Family**, **Genus**, and **Species**.

A helpful aid for remembering the taxonomic categories:

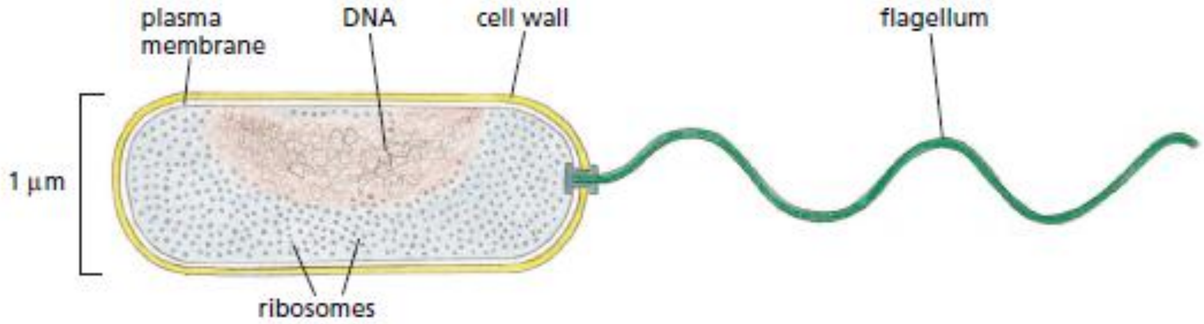
Domain, **Kingdom**, **Phylum**, **Class**, **Order**, **Family**, **Genus**, and **Species** is the mnemonic device:

Do **K**eep **P**lates **C**lean **O**r **F**amily **G**ets **S**ick.

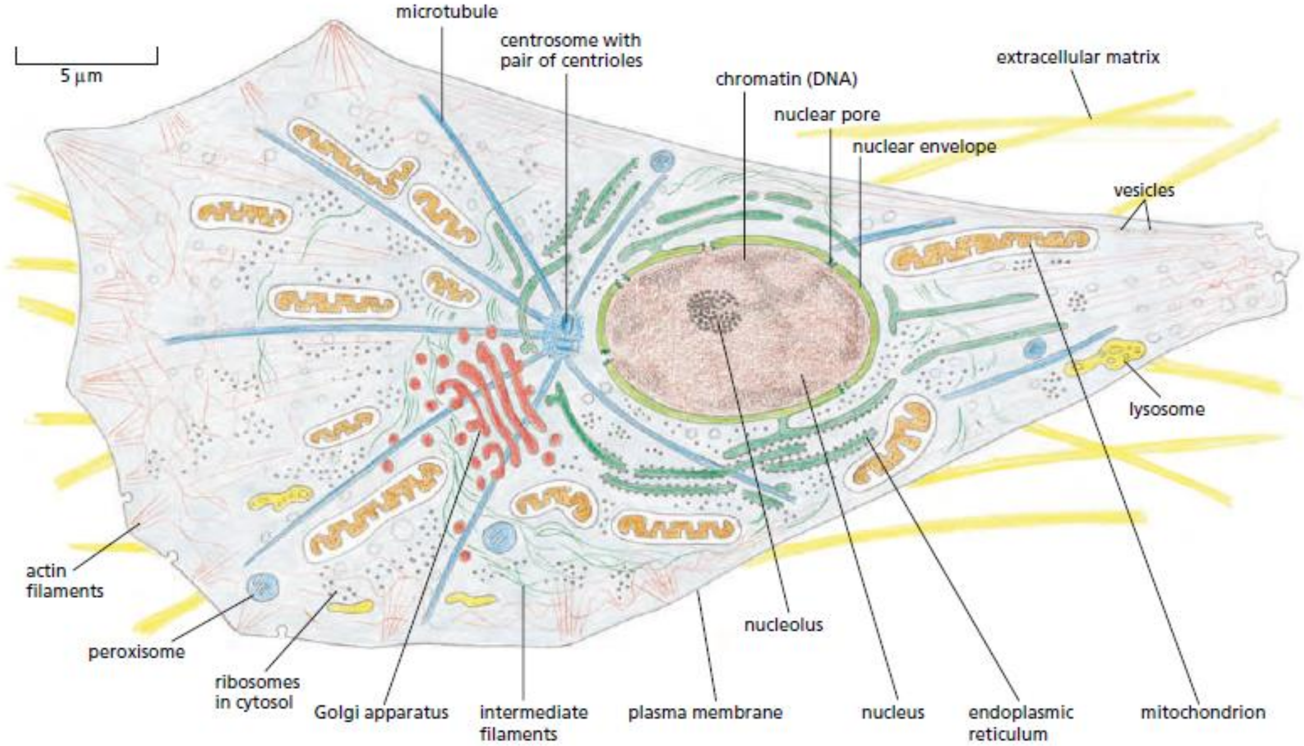



All organisms
Reduced fuel
Oxidized fuel
Energy source
Phototrophs
(energy from light)
Chemotrophs
(energy from oxidation of chemical fuels)
Lithotrophs
(inorganic fuels)
Organotrophs
(organic fuels)
Carbon source
Autotrophs
(carbon from CO_2)
Heterotrophs
(carbon from organic compounds)
Examples
Cyanobacteria
Vascular plants
Purple bacteria
Green bacteria
Sulfur bacteria
Hydrogen bacteria
Most bacteria
All nonphototrophic eukaryotes

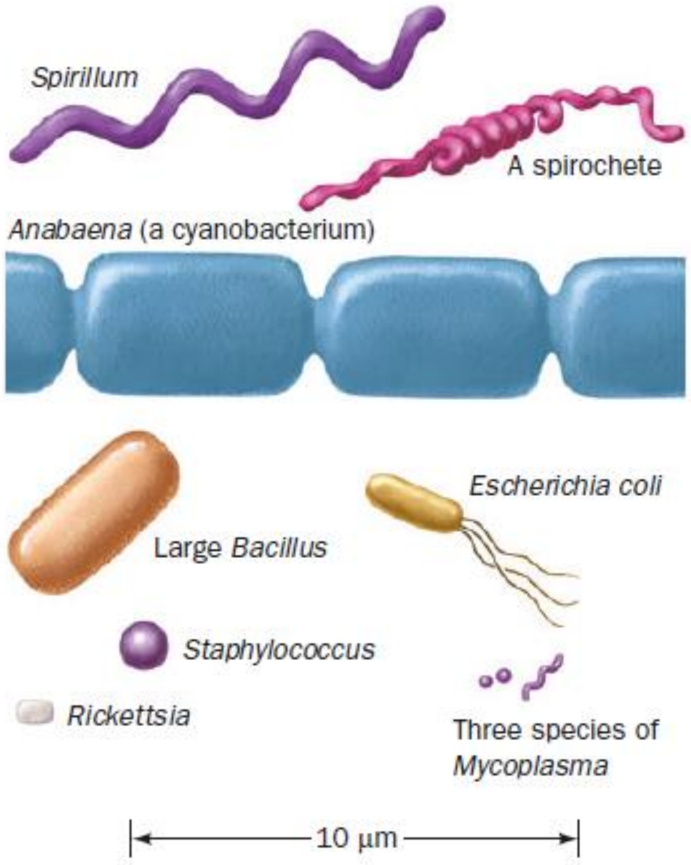
Bacterial cell



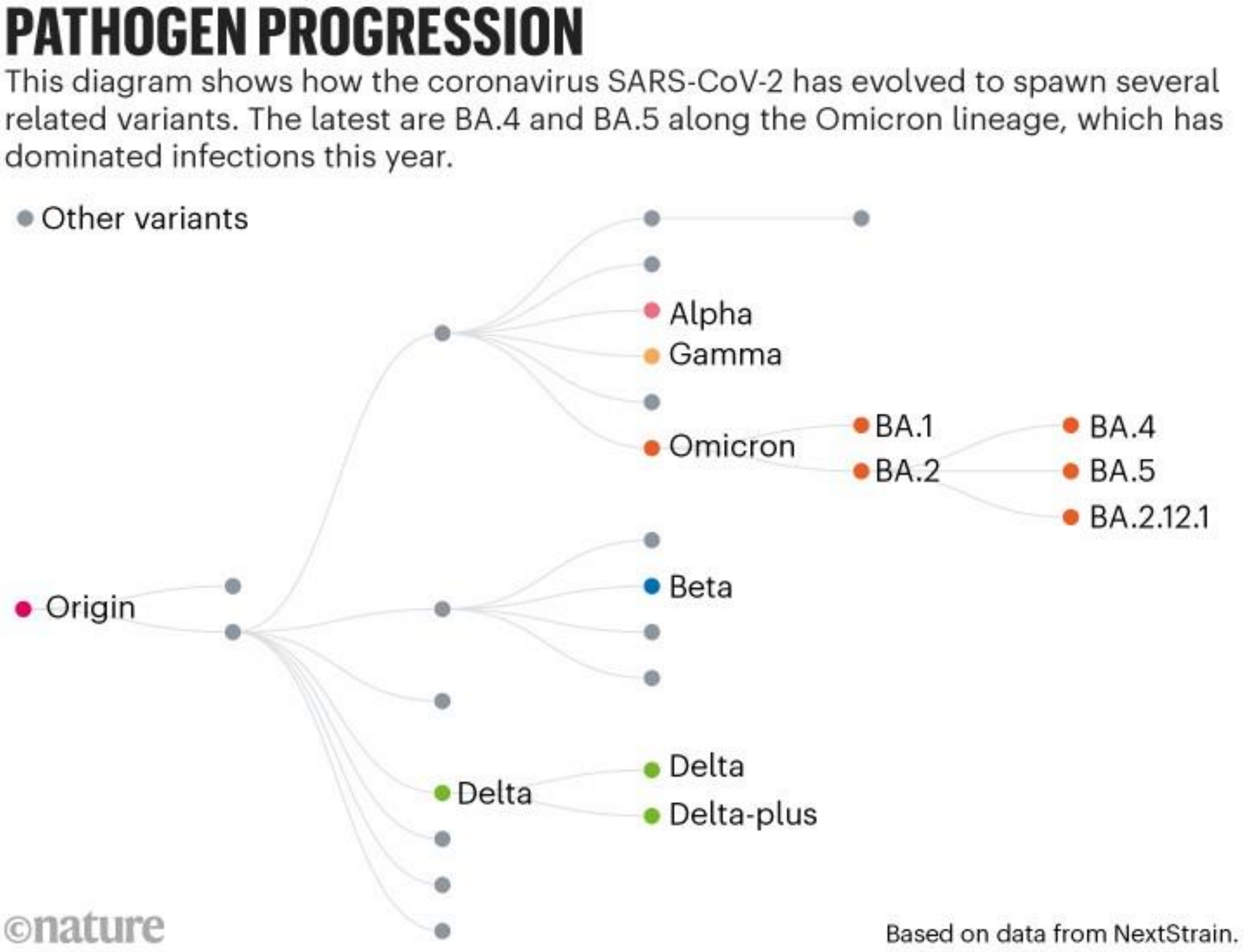
Animal cell



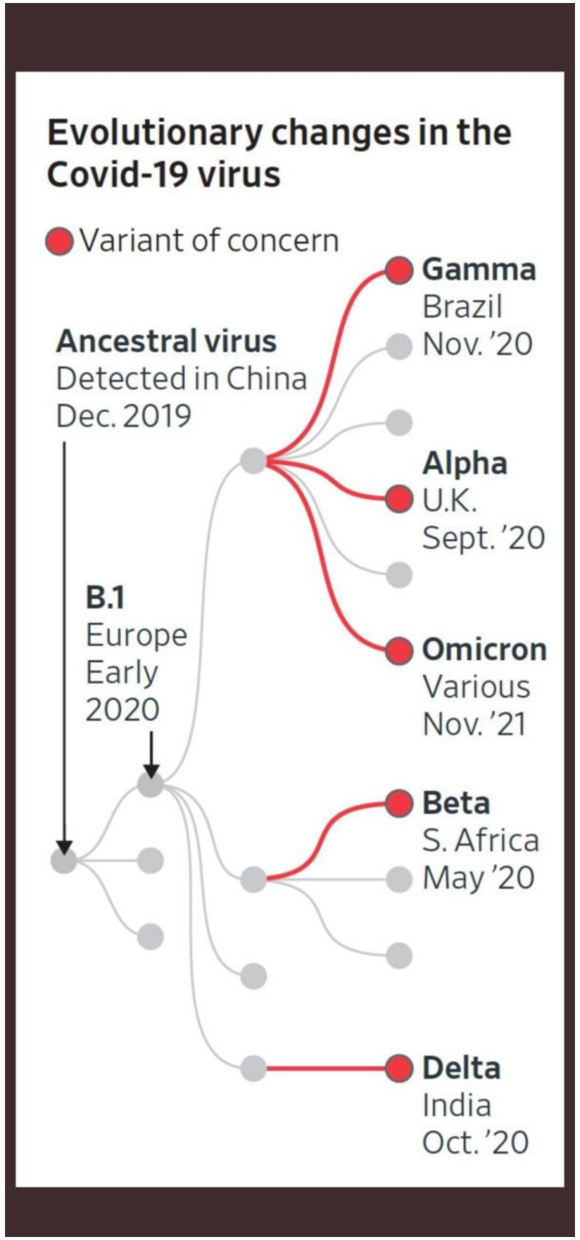
Scale drawings of some prokaryotic cells



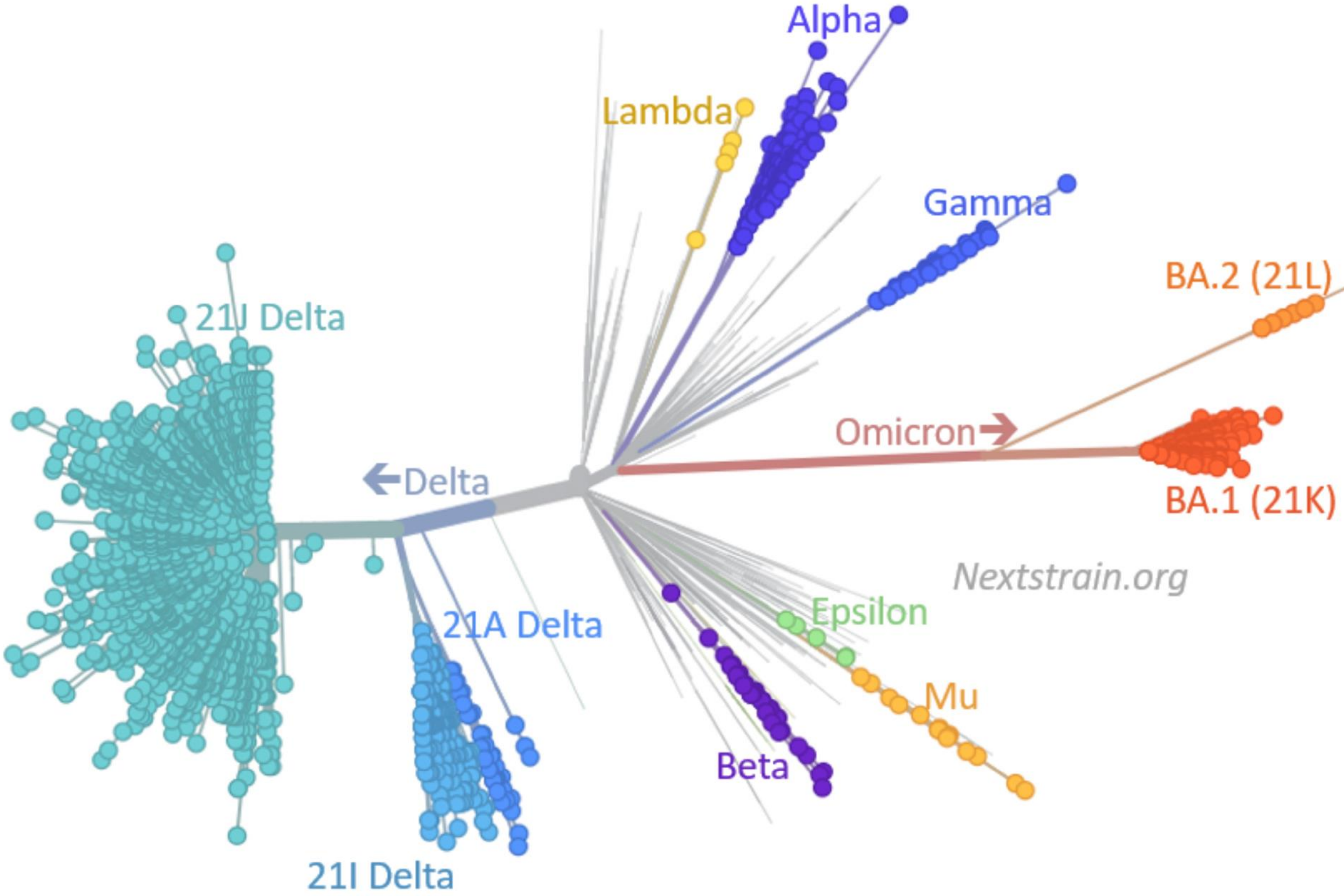
Three basic structures: spheroidal (cocci), rodlike (bacilli), and helically coiled (spirilla)



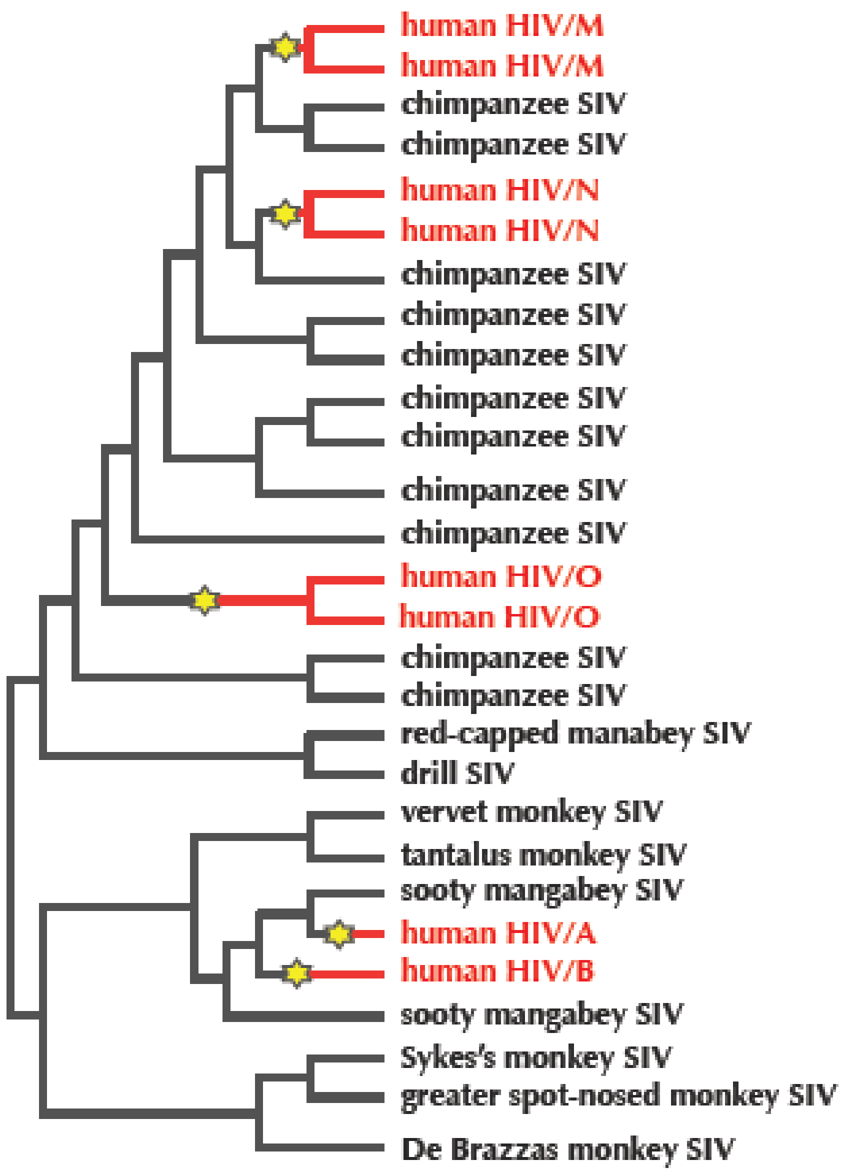
PATHOGEN PROGRESSION
This diagram shows how the coronavirus SARS-CoV-2 has evolved to spawn several related variants. The latest are BA.4 and BA.5 along the Omicron lineage, which has dominated infections this year.
Other variants
Alpha
Gamma
Omicron
BA.1
BA.2
BA.4
BA.5
BA.2.12.1
Beta
Origin
Delta
Delta
Delta-plus
©nature
Based on data from NextStrain.

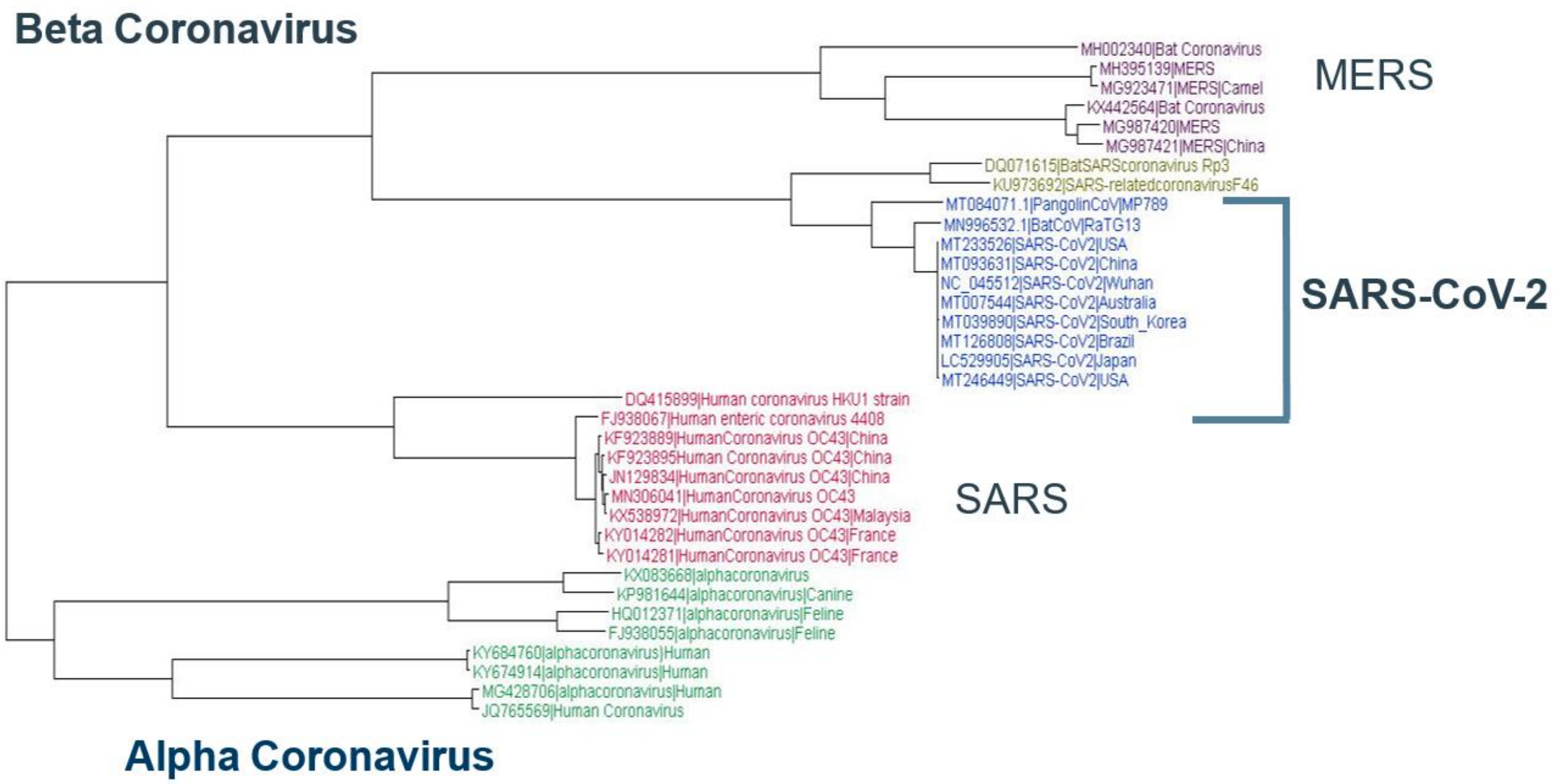


Evolutionary changes in the Covid-19 virus
Variant of concern
Gamma
Brazil
Nov. '20
Ancestral virus
Detected in China
Dec. 2019
Alpha
U.K.
Sept. '20
B.1
Europe
Early
2020
Omicron
Various
Nov. '21
Beta
S. Africa
May '20
Delta
India
Oct. '20



Alpha
Lambda
Gamma
BA.2 (21L)
21J Delta
Omicron
Delta
BA.1 (21K)
Nextstrain.org
Epsilon
21A Delta
Mu
Beta
21I Delta


human HIV/M
human HIV/M
chimpanzee SIV
chimpanzee SIV
human HIV/N
human HIV/N
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
human HIV/O
human HIV/O
chimpanzee SIV
chimpanzee SIV
red-capped manabey SIV
drill SIV
vervet monkey SIV
tantalus monkey SIV
sooty mangabey SIV
human HIV/A
human HIV/B
sooty mangabey SIV
Sykes's monkey SIV
greater spot-nosed monkey SIV
De Brazzas monkey SIV



Beta Coronavirus
MH002340|Bat Coronavirus
MH395139|MERS
MG923471|MERS|Camel
KX442564|Bat Coronavirus
MG987420|MERS
MG987421|MERS|China
MERS
DQ071615|BatSARScoronavirus Rp3
KU973692|SARS-relatedcoronavirusF46
MT084071.1|PangolinCoV|MP789
MN996532.1|BatCoV|RaTG13
MT233526|SARS-CoV2|USA
MT093631|SARS-CoV2|China
NC_045512|SARS-CoV2|Wuhan
MT007544|SARS-CoV2|Australia
MT039890|SARS-CoV2|South_Korea
MT126808|SARS-CoV2|Brazil
LC529905|SARS-CoV2|Japan
MT246449|SARS-CoV2|USA
SARS-CoV-2
DQ415899|Human coronavirus HKU1 strain
FJ938067|Human enteric coronavirus 4408
KF923889|HumanCoronavirus OC43|China
KF923895Human Coronavirus OC43|China
JN129834|HumanCoronavirus OC43|China
MN306041|HumanCoronavirus OC43
KX538972|HumanCoronavirus OC43|Malaysia
KY014282|HumanCoronavirus OC43|France
KY014281|HumanCoronavirus OC43|France
SARS
KX083668|alphacoronavirus
KP981644|alphacoronavirus|Canine
HQ012371|alphacoronavirus|Feline
FJ938055|alphacoronavirus|Feline
KY684760|alphacoronavirus|Human
KY674914|alphacoronavirus|Human
MG428706|alphacoronavirus|Human
JQ765569|Human Coronavirus
Alpha Coronavirus

Let us play the sequence alignment game!

Strings ATGCATGC and TGCATGCA !

```
ATGCATGC
TGCATGCA

ATGCATGC-
-TGCATGCA
```

We postulate a notion of a good alignment as one that matches as many symbols as possible.

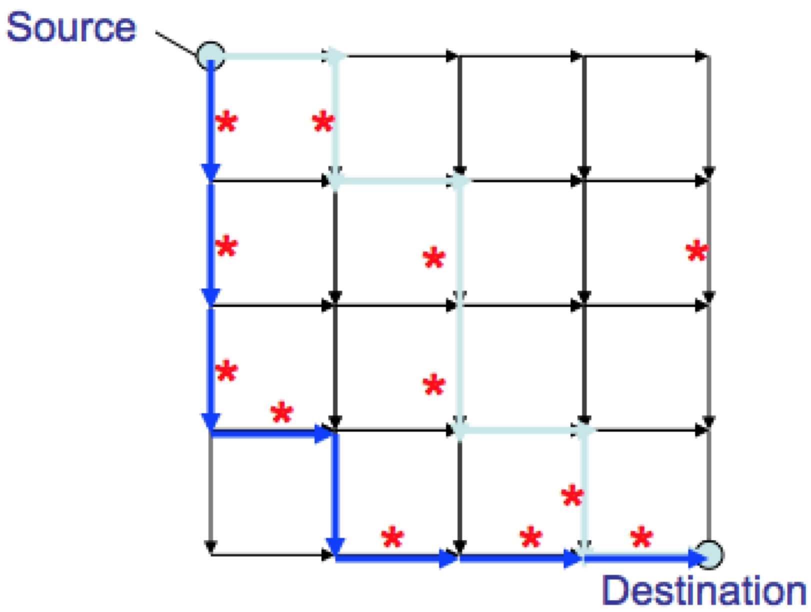
Strings ATGCTTA and TGCATTAA

```
ATGC-TTA-
-TGCATTAA
```

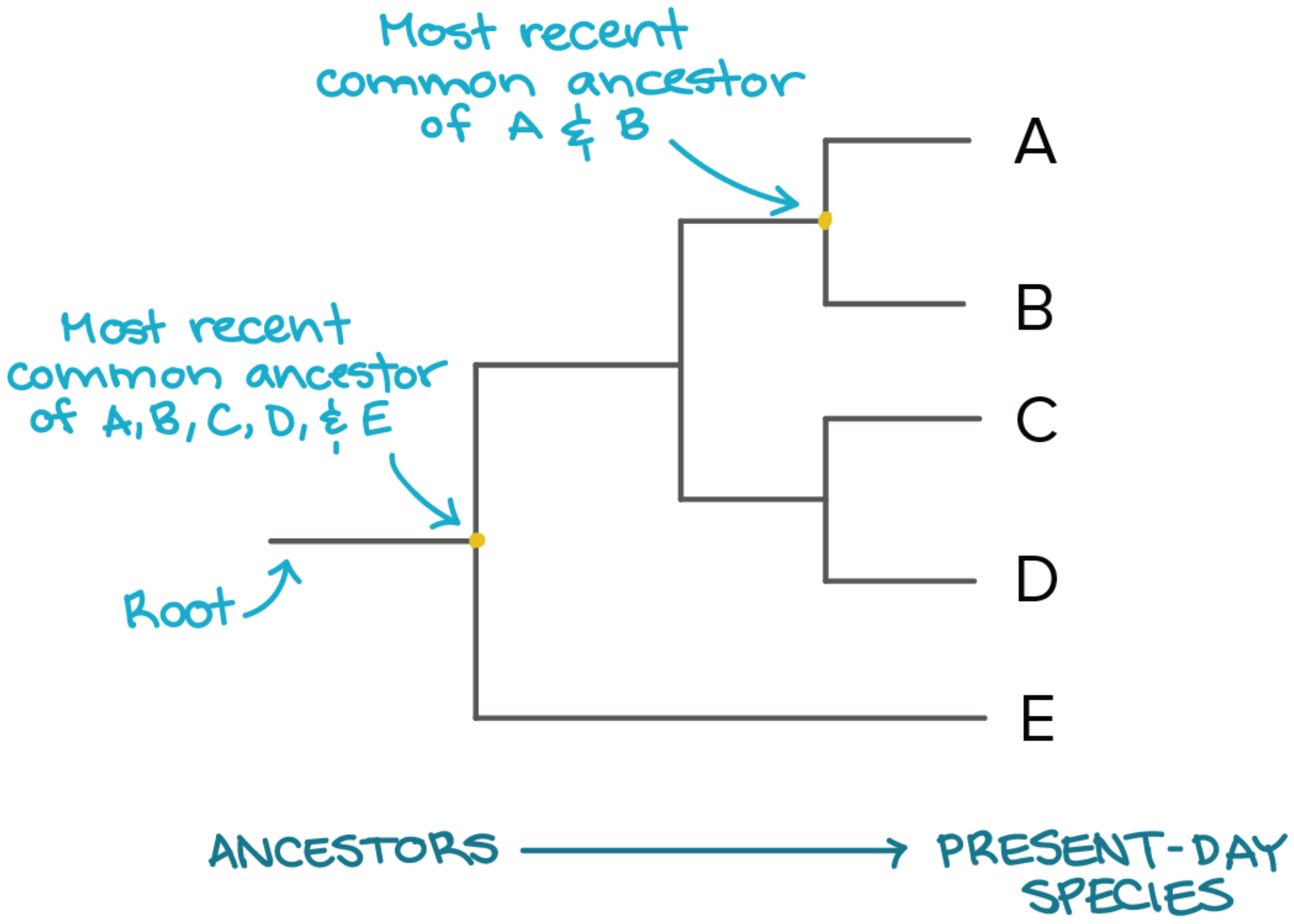
Longest Common Subsequence Problem:
Find a longest common subsequence of two strings.

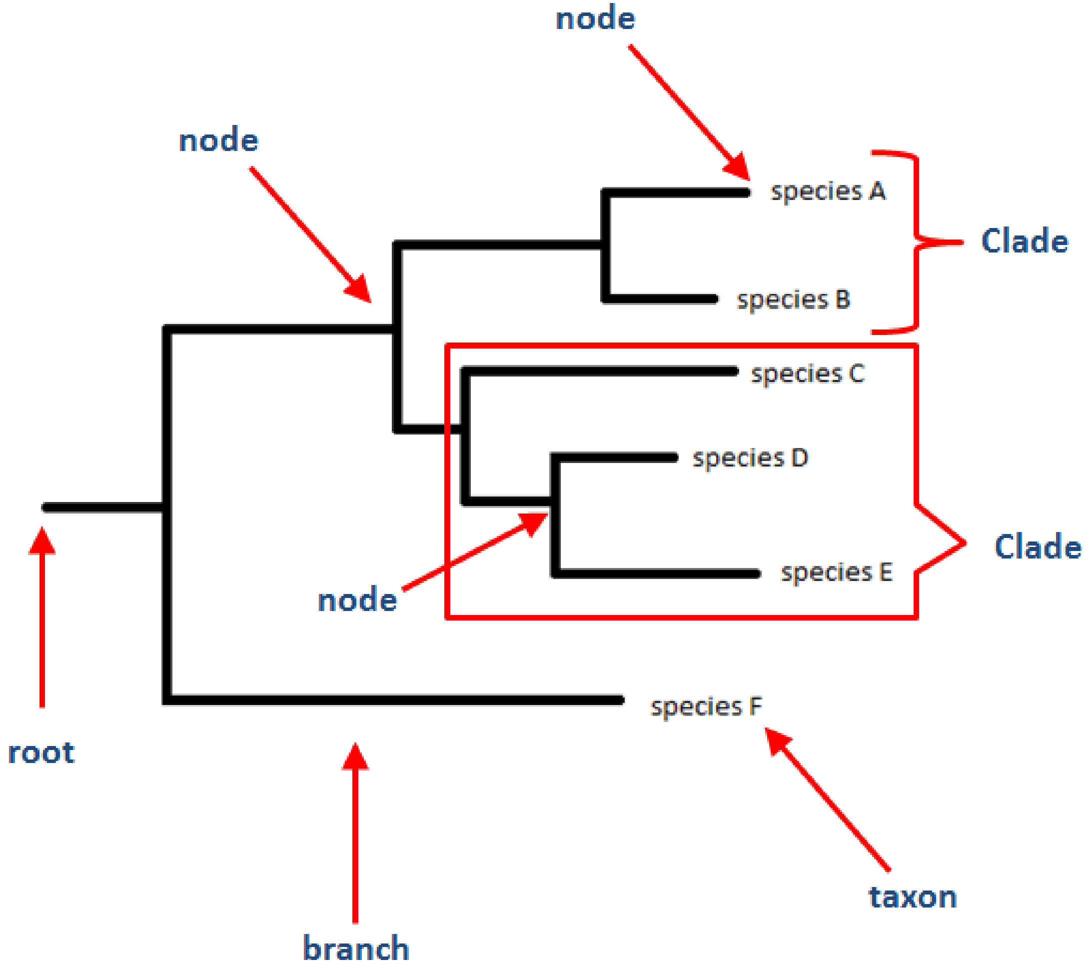
Input: Two strings.
Output: A longest common subsequence of these strings.

The Manhattan tourist problem!



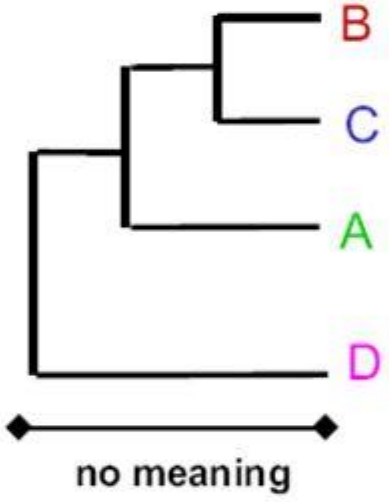
Parts of a phylogenetic tree





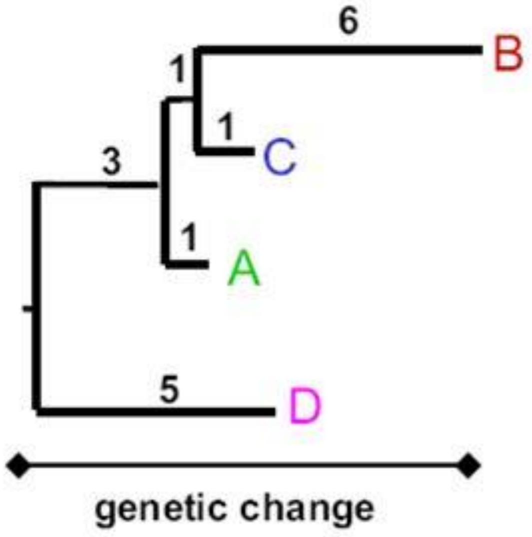
Types of trees

Cladogram



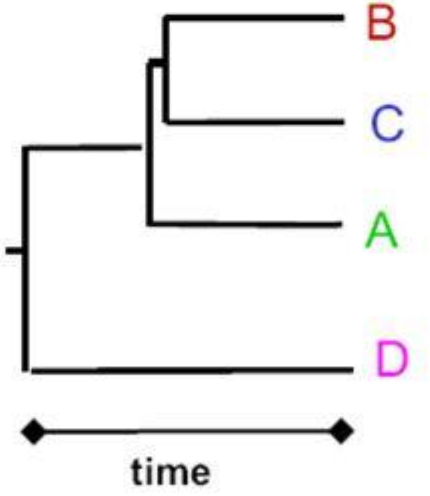
Simply shows relative recency of common ancestor

Phylogram

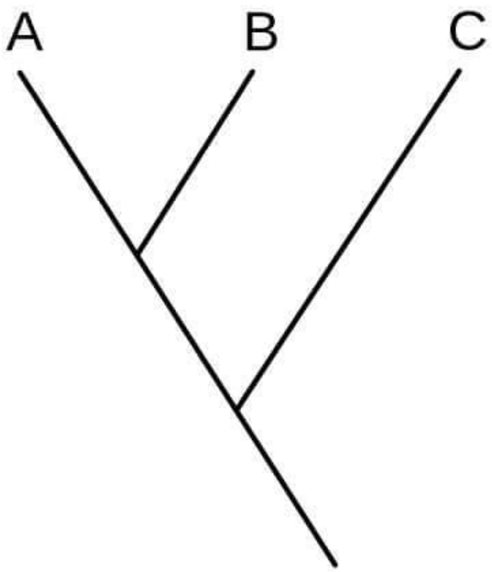


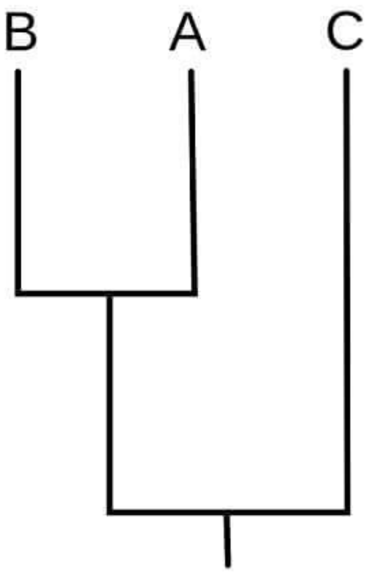
A cladogram with branch lengths

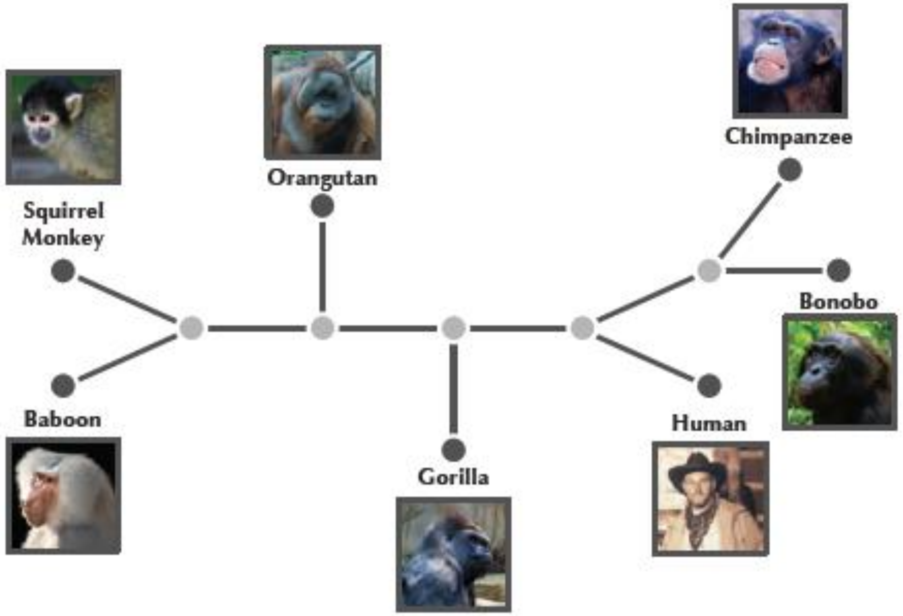
Ultrametric tree

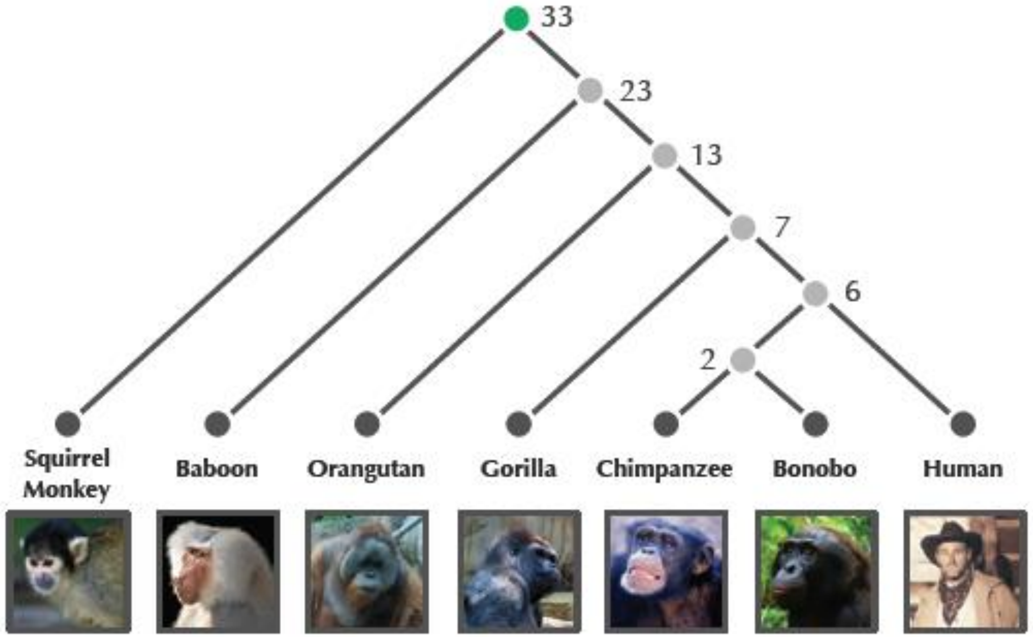


A dendogram having all tips equidistant from root


A
B
C


B
A
C

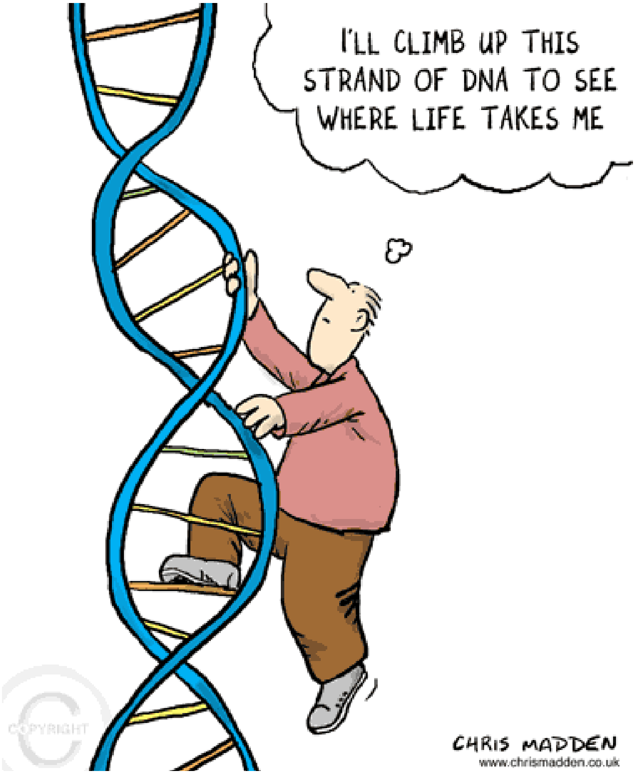

Squirrel Monkey
Orangutan
Chimpanzee
Bonobo
Baboon
Gorilla
Human


33
23
13
7
6
2
Squirrel Monkey
Baboon
Orangutan
Gorilla
Chimpanzee
Bonobo
Human

Some definitions

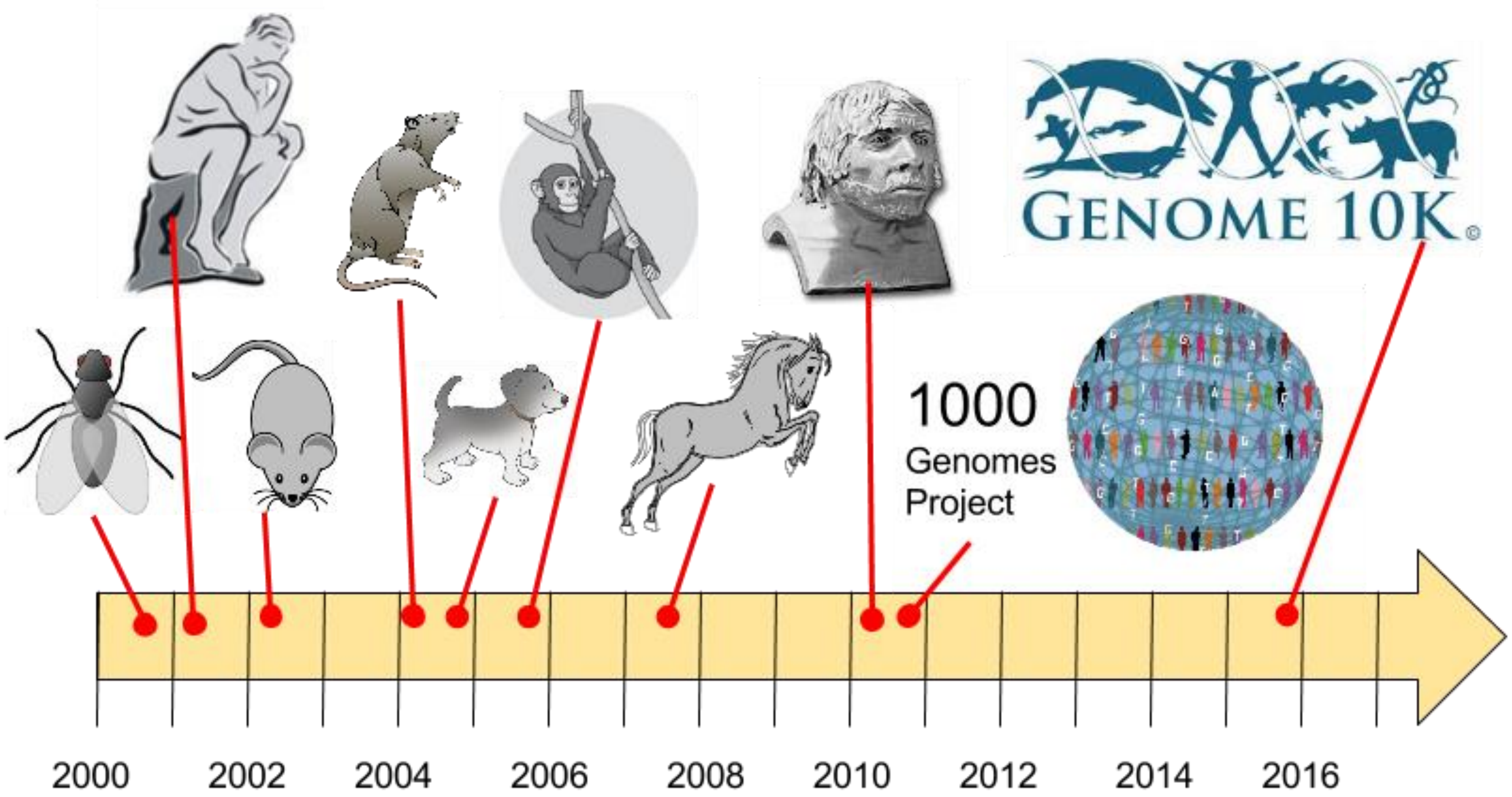
- Classification - Grouping of things into classes.
  - e.g. Kingdom, phylum, class, order, family, genus, species.
- Taxonomy - Giving names to things.
  - e.g. Taxons -Mammalia, *Homo sapiens*, *Felis*
- Systematics- Understanding the relationships between things.
  - e.g. The daughter of my aunt is my cousin. The sister of my grandmother is my great-aunt.
- Phylogeny - The evolutionary history of a species.
  - If you don't know your history, you don't know anything!
- Cladistics - Phylogenetic systematics. Using the evolutionary history of species to understand their relationships, classify them, and give them names!

Universal feature-Chemical code



Organism	Genome size* (nucleotide pairs)	Approximate number of genes
Escherichia coli (bacterium)	4.6×10^6	4300
Saccharomyces cerevisiae (yeast)	13×10^6	6600
Caenorhabditis elegans (roundworm)	130×10^6	21,000
Arabidopsis thaliana (plant)	220×10^6	29,000
Drosophila melanogaster (fruit fly)	200×10^6	15,000
Danio rerio (zebrafish)	1400×10^6	32,000
Mus musculus (mouse)	2800×10^6	30,000
Homo sapiens (human)	3200×10^6	30,000

*Genome size includes an estimate for the amount of highly repeated DNA sequence not in genome databases.


GENOME 10K
1000
Genomes
Project
2000
2002
2004
2006
2008
2010
2012
2014
2016

TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATATTCTACTTTTAC
CATTGAAAATATTGAGGAAGTTATTTATATTTCTATTTTTTATATATTATATATTTTATGTATTT
TAATATTACTATTACACATAATTATTTTTTATATATATGAAGTACCAATGACTTCCTTTTCCAG
AGCAATAATGAAATTTCACAGTATGAAAATGGAAGAAATCAATAAAATTATACGTGACCT
GTGGCGAAGTACCTATCGTGGACAAGGTGAGTACCATGGTGTATCACAAATGCTCTTTCC
AAAGCCCTCTCCGCAGCTCTTCCCCTTATGACCTCTCATCATGCCAGCATTACCTCCCTGG
ACCCCTTTCTAAGCATGTCTTTGAGATTTTCTAAGAATTCTTATCTTGGCAACATCTTGTAG
CAAGAAAATGTAAAGTTTTCTGTTCCAGAGCCTAACAGGACTTACATATTTGACTGCAGT
AGGCATTATATTTAGCTGATGACATAATAGGTTCTGTCATAGTGTAGATAGGGATAAGCCA
AAATGCAATAAGAAAAACCATCCAGAGGAAACTCTTTTTTTTTCTTTTTCTTTTTTTTT
TTCCAGATGGAGTCTCGCACTTCTCTGTCACCCGGGCTGGAGCGCAGTGGTGCAATCTT
GGCTCACTGCAACCTCCACCTCCTGGGTTCAGGTGATTCTCCCACCTCAGCCTCCCGAGT
AGTAGCTGGAATTACAGGTGCGCGCTCCCACACCTGGCTAATTTTTTGTATTCTTAGTAG
AGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGCCCTCAGGTGATCTG
CCCACCTTGGCCTCCCAGTGTTGGGTTTACAGGCGTGAGCCACCGCGCCTGGCCTGGA
GGAAACTCTTAACAGGGAAACTAAGAAAGAGTTGAGGCTGAGGAACTGGGGCATCTG
GGTTGCTTCTGGCCAGACCACCAGGCTCTTGAATCCTCCCAGCCAGAGAAAGAGTTTCC
ACACCAGCCATTGTTTTCCTCTGGTAATGTCAGCCTCATCTGTTGTTCCTAGGCTTACTTG
ATATGTTTGTAAATGACAAAAGGCTACAGAGCATAGGTTCCTCTAAAATATTCTTCTTCCT
GTGTCAGATATTGAATACATAGAAATACGGTCTGATGCCGATGAAAATGTATCAGCTTCTG
ATAAAAGGCGGAATTATAACTACCGAGTGGTGATGCTGAAGGGAGACACAGCCTTGGA
TATGCGAGGACGATGCAGTGCTGGACAAAAGGCAGGTATCTCAAAAGCCTGGGGAGCC
AACTCACCCAAGTAACTGAAAGAGAGAAACAAACATCAGTGCAGTGGAAGCACCCAAG
GCTACACCTGAATGGTGGGAAGCTCTTTGCTGCTATATAAAATGAATCAGGCTCAGCTAC
TATTATT …………

The Human Genome

2003

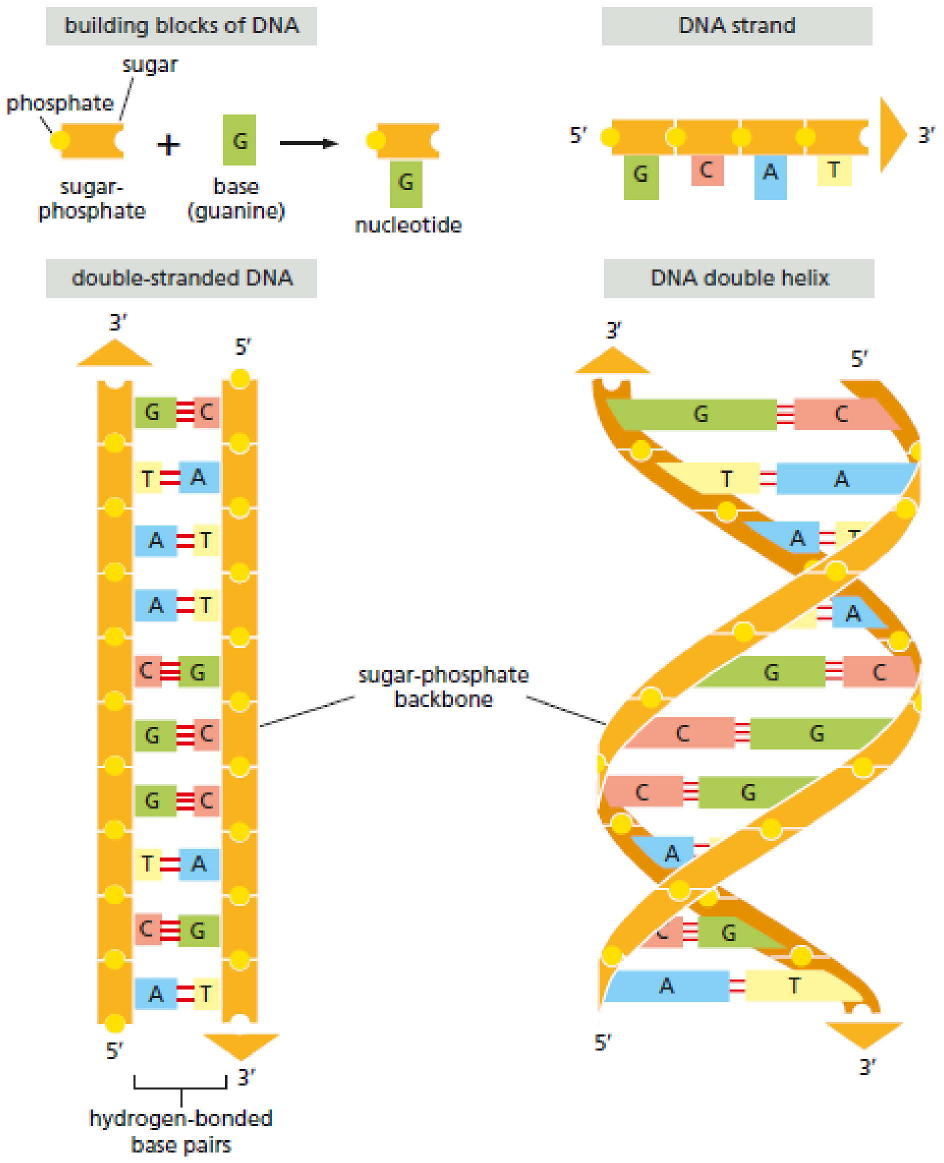
“Fifteen years, six countries, twenty centers. Three billion dollars, three billion letters. One dollar per letter – such a deal!”

Eric Lander's seven-word " Nano Lecture"

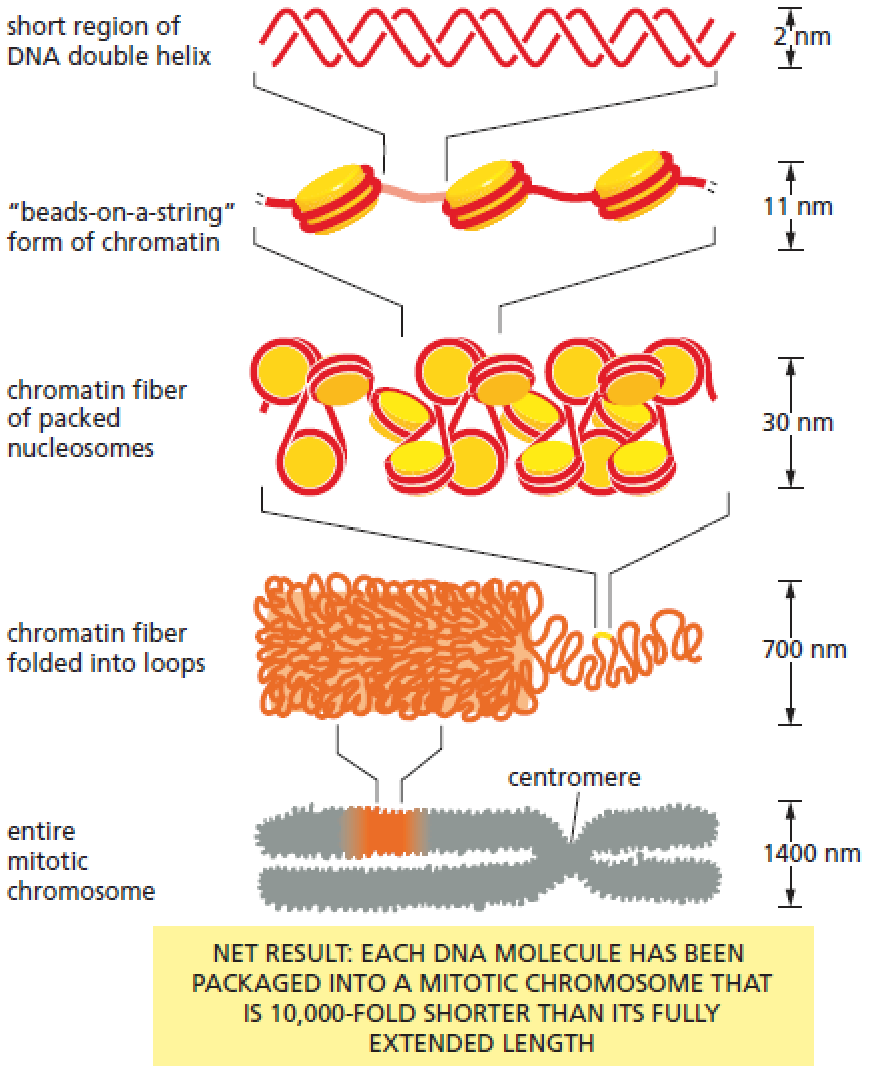
"Genome: bought the book; hard to read"

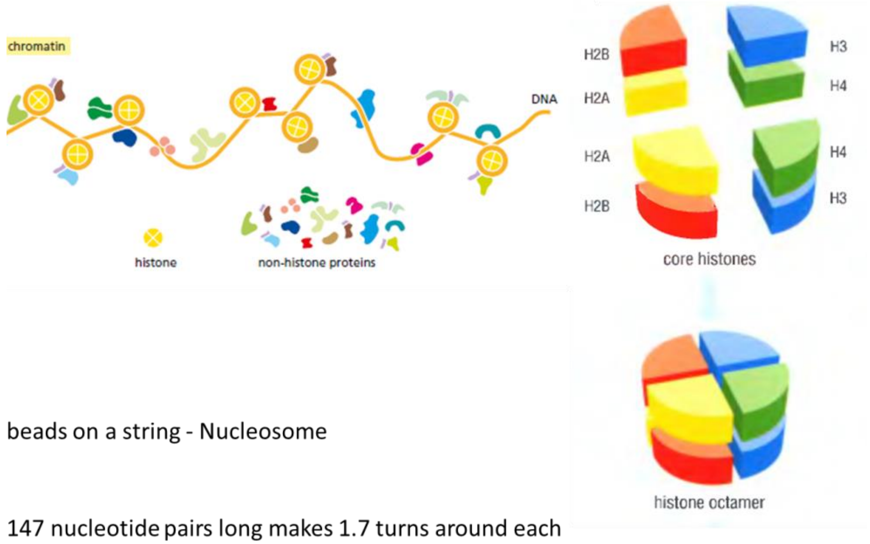
TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATATTCTACTTTTACCATTGAAAATATTGAGGAAGTTATTTATATTTCTATTTTTTATA
TATTATATATTTTATGTATTTTAATATTACTATTACACATAATTATTTTTTATATATATGAAGTACCAATGACTTCCTTTTCCAGAGCAATAATGAAATTTCACAGTA
TGAAAATGGAAGAAATCAATAAAATTATACGTGACCTGTGGCGAAGTACCTATCGTGGACAAGGTGAGTACCATGGTGTATCACAAATGCTCTTTCCAAAG
CCCTCTCCGCAGCTCTTCCCCTTATGACCTCTCATCATGCCAGCATTACCTCCCTGGACCCCTTTCTAAGCATGTCTTTGAGATTTTCTAAGAATTCTTATCTTG
GCAACATCTTGTAGCAAGAAAATGTAAAGTTTTCTGTTCCAGAGCCTAACAGGACTTACATATTTGACTGCAGTAGGCATTATATTTAGCTGATGACATAATA
GGTTCTGTCATAGTGTAGATAGGGATAAGCCAAAATGCAATAAGAAAAACCATCCAGAGGAAACTCTTTTTTTTTTCTTTTTCTTTTTTTTTTTTTCCAGATG
GAGTCTCGCACTTCTCTGTCACCCGGGCTGGAGCGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCTGGGTTCAGGTGATTCTCCCACCTCAG
CCTCCCGAGTAGTAGCTGGAATTACAGGTGCGCGCTCCCACACCTGGCTAATTTTTTGTATTCTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGG
TCTCAAACTCCTGCCCTCAGGTGATCTGCCCACCTTGGCCTCCCAGTGTTGGGTTTACAGGCGTGAGCCACCGCGCCTGGCCTGGAGGAAACTCTTAACA
GGGAAACTAAGAAAGAGTTGAGGCTGAGGAACTGGGGCATCTGGGTTGCTTCTGGCCAGACCACCAGGCTCTTGAATCCTCCCAGCCAGAGAAAGAG
TTTCCACACCAGCCATTGTTTTCCTCTGGTAATGTCAGCCTCATCTGTTGTTCCTAGGCTTACTTGATATGTTTGTAAATGACAAAAGGCTACAGAGCATAGG
TTCCTCTAAAATATTCTTCTTCCTGTGTCAGATATTGAATACATAGAAATACGGTCTGATGCCGATGAAAATGTATCAGCTTCTGATAAAAGGCGGAATTATAA
CTACCGAGTGGTGATGCTGAAGGGAGACACAGCCTTGGATATGCGAGGACGATGCAGTGCTGGACAAAA

DNA and its building blocks



The organization of Chromosomes

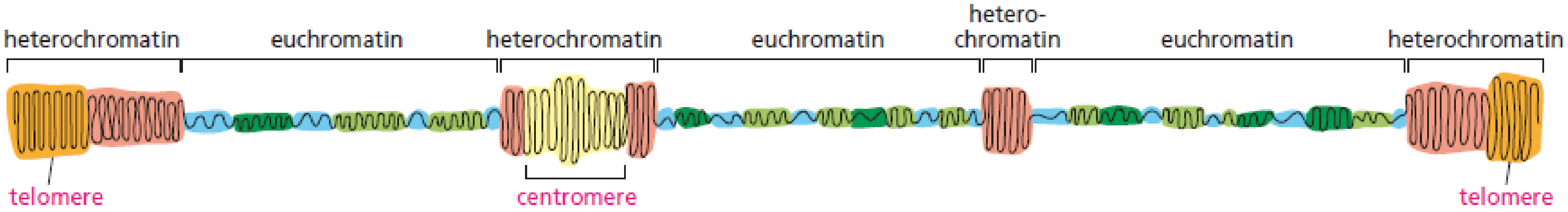




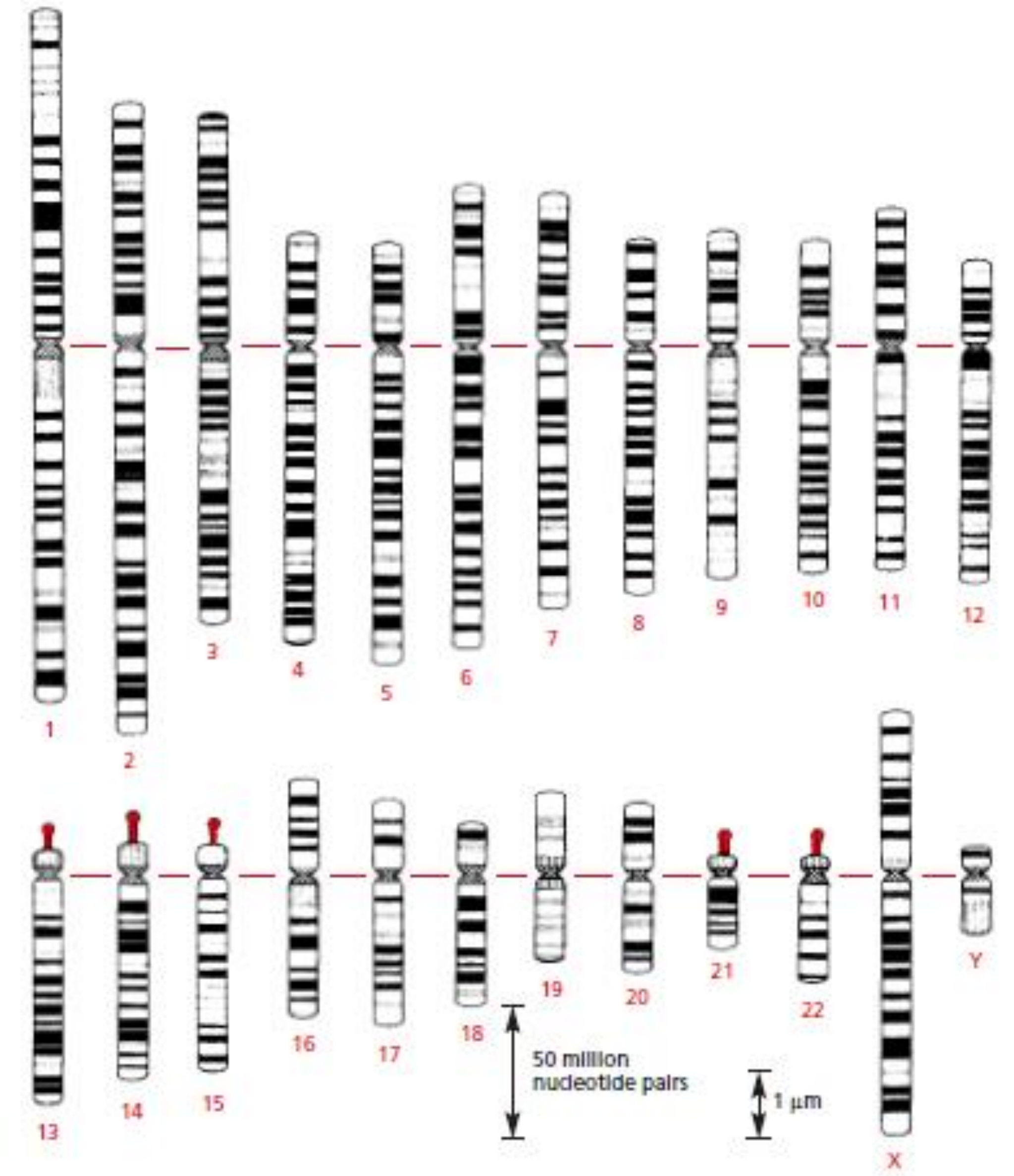
beads on a string - Nucleosome

147 nucleotide pairs long makes 1.7 turns around each protein core

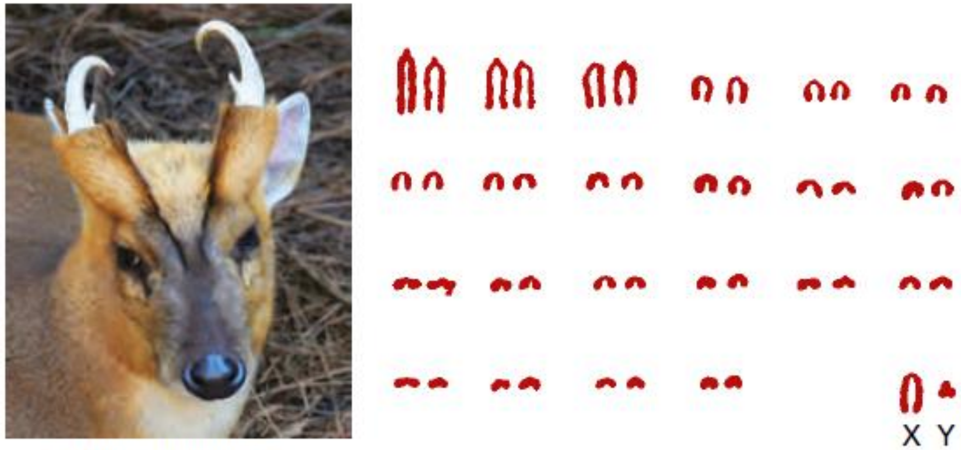
The structure of chromatin varies



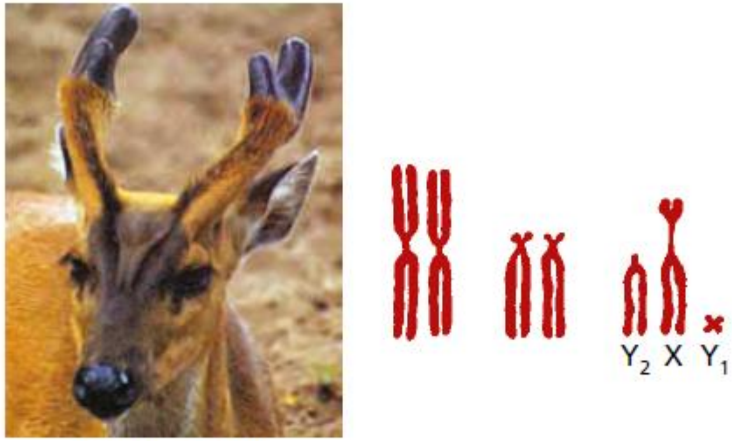
The banding patterns of human chromosomes



Giemsa Staining



Chinese muntjac



Indian muntjac

Interesting observations from Beatrice !

- In some insects like grasshoppers, crickets, and roaches, there is only one kind of sex chromosome, X. If you get 2 of them, you’re female, and if you just get one, you’re a male.

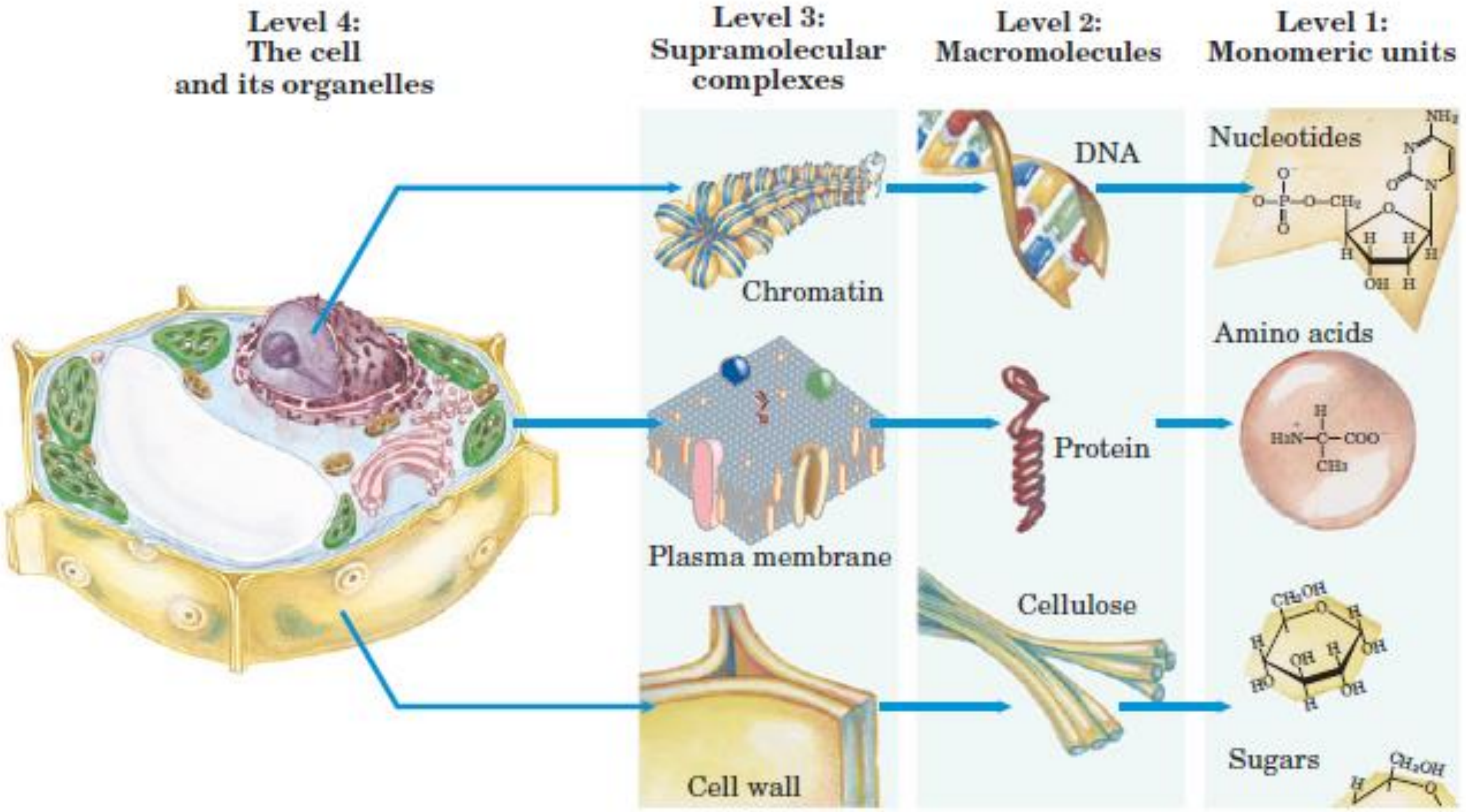
- In birds (and even some insects and fishes) it’s the females that have different sex chromosomes. Females are ZW and males are ZZ. Scientists changed the letters to differentiate between the XX-female, XY-male scenario.

- In ants and bees, there are no sex chromosomes. Instead, sex is determined by whether or not an egg was fertilized. If the egg isn’t fertilized, the offspring is male. If the egg is fertilized, it’s female. So male ants have no fathers, and they have half as many chromosomes as females. Poor little things!

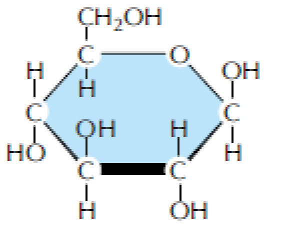
- A species of deer called the Indian muntjac (as opposed to the Chinese muntjac) has three different kinds of sex chromosomes–X and two versions of Y, called Y1 and Y2. Females are XX, and males are XY1Y2. Ha, those crazy muntjacs.

- May we never forget that biology is a complicated, messy, random process that has found several ways for dealing with sex and reproduction (and just about every other bit of life business). There is really no rhyme or reason for any of it. If it works, it works. If it doesn’t, you die. That’s pretty much the only rule in nature that applies all the time.

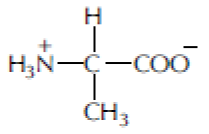
Structural hierarchy in the molecular organization of cells



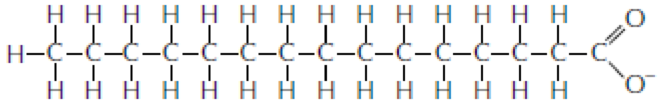
Cells function as biochemical factories



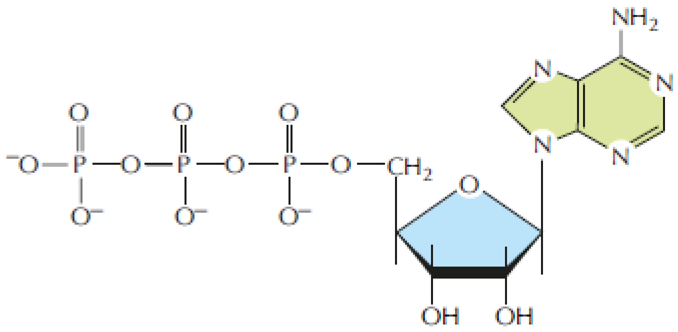
A SUGAR



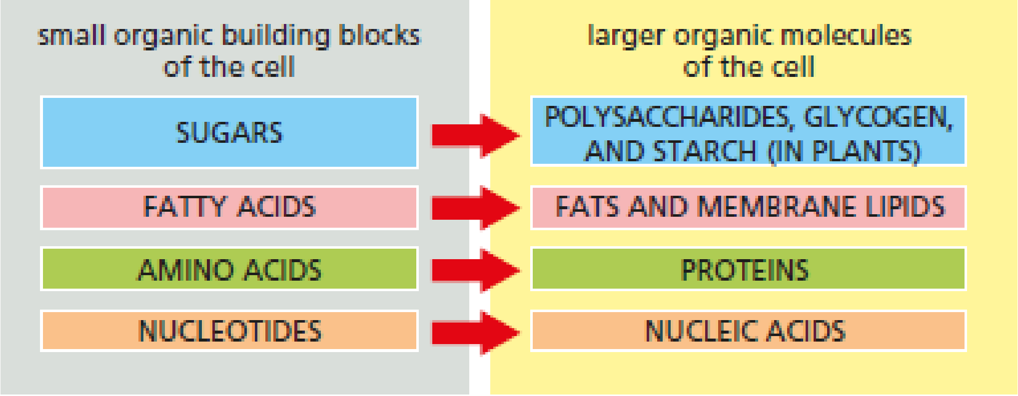
AN AMINO ACID

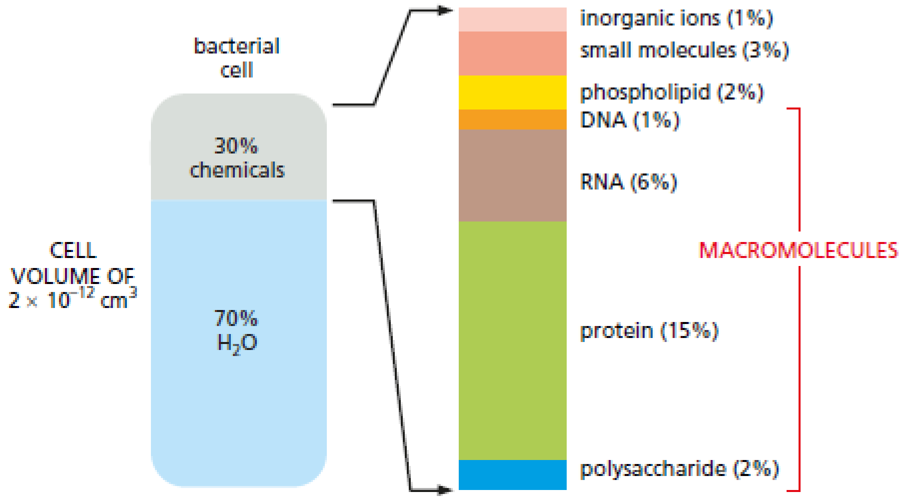


A FATTY ACID

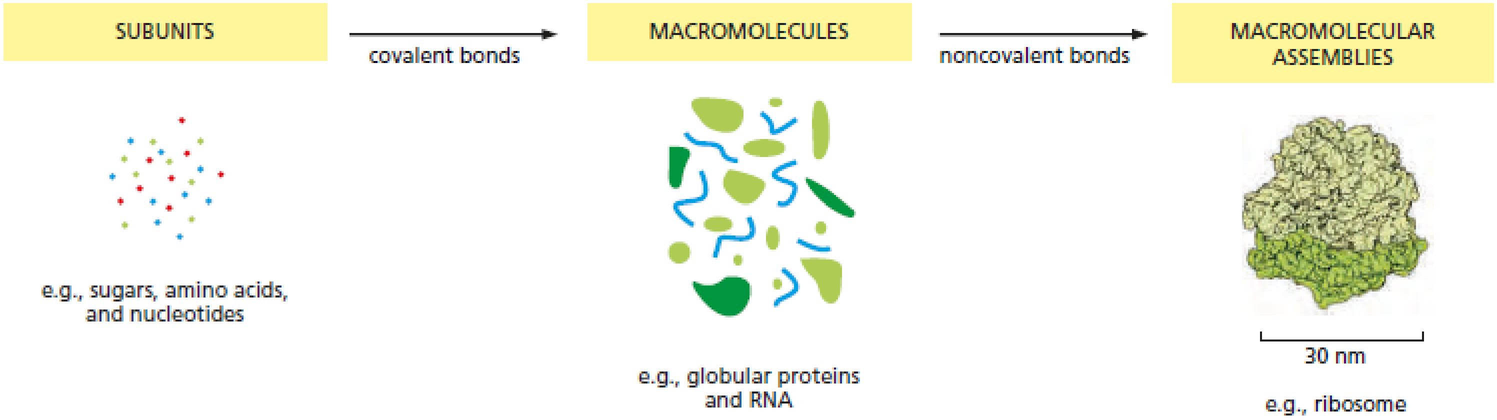


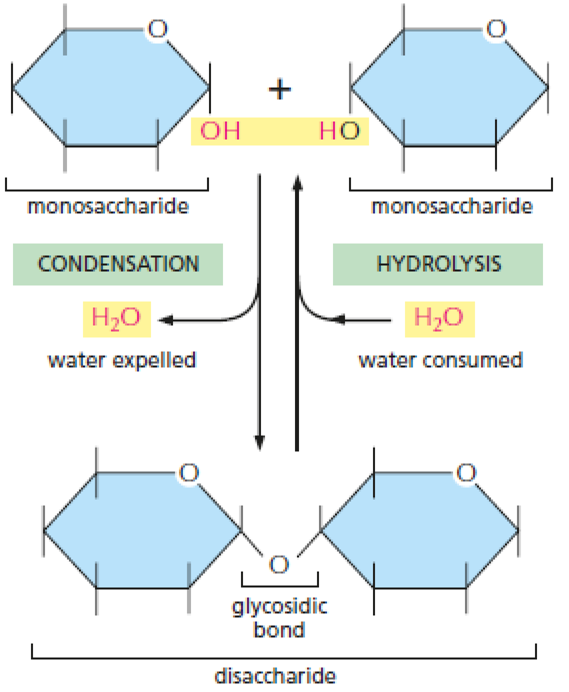
A NUCLEOTIDE



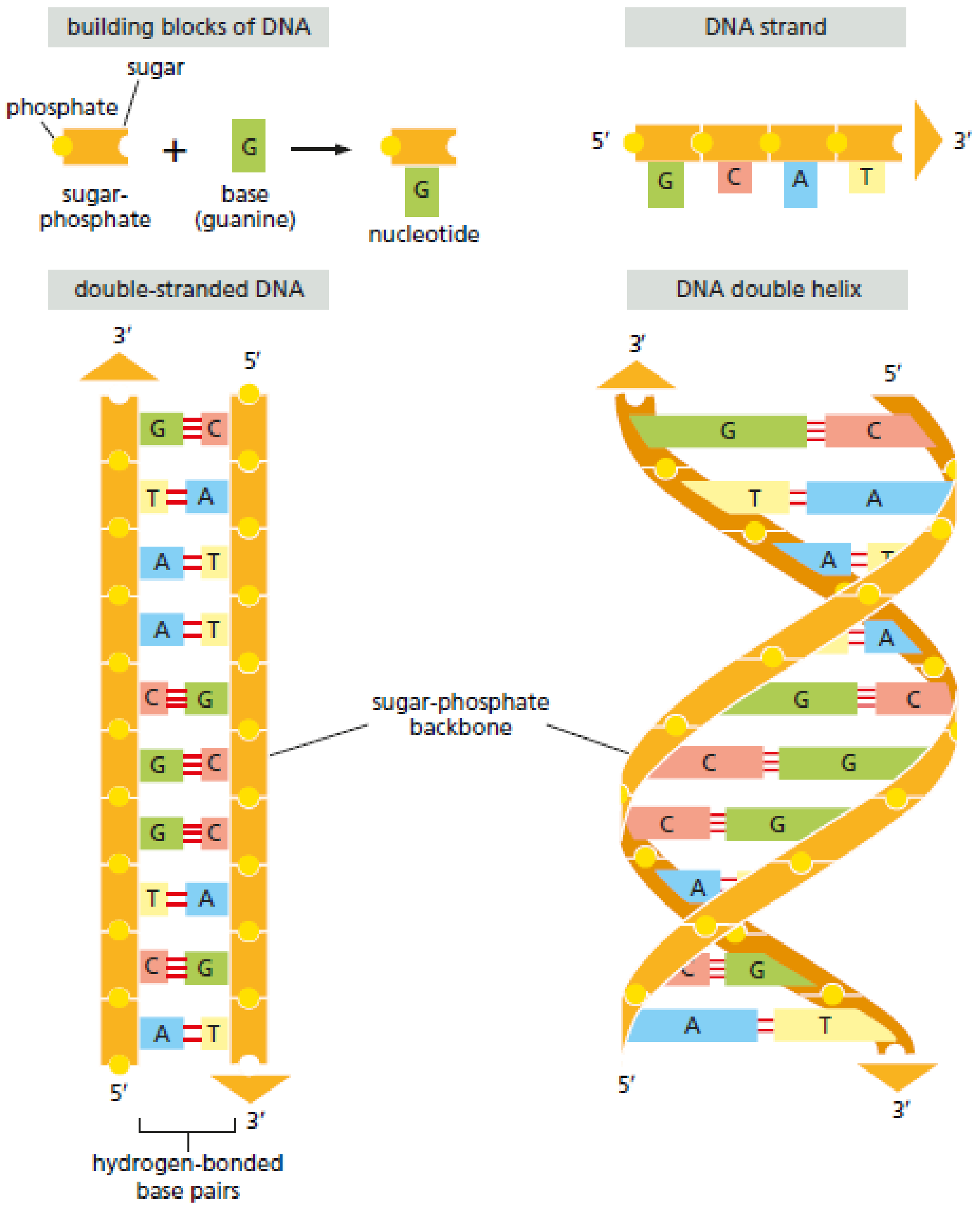


Biochemical bond formation



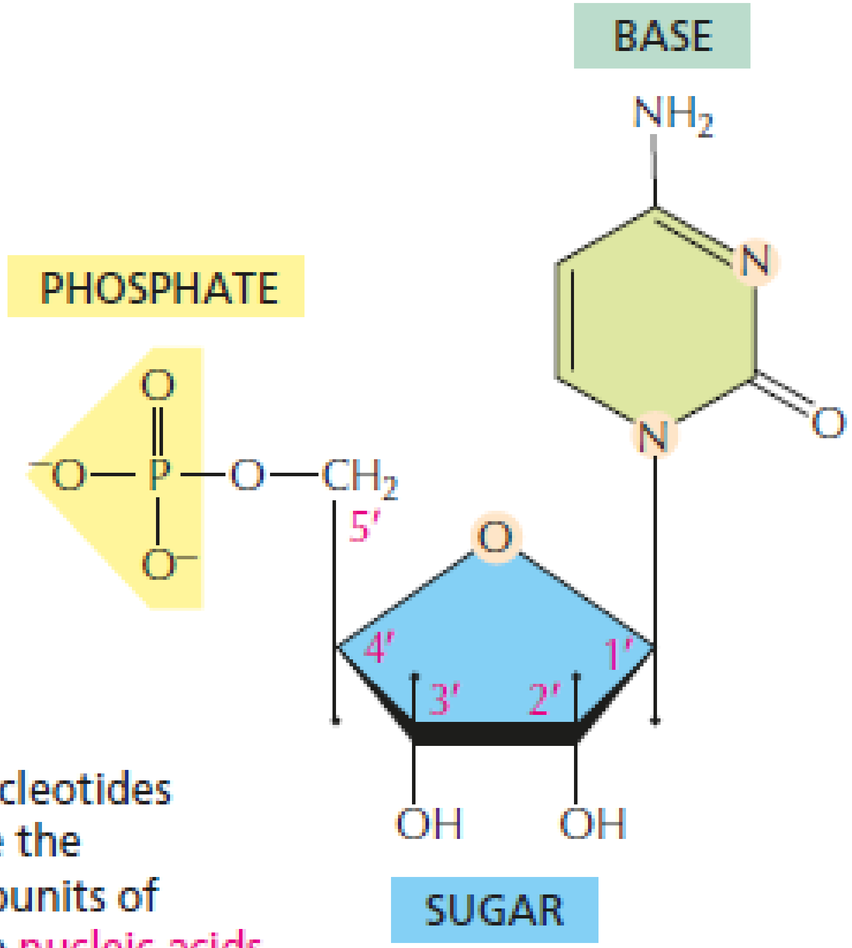


DNA and its building blocks



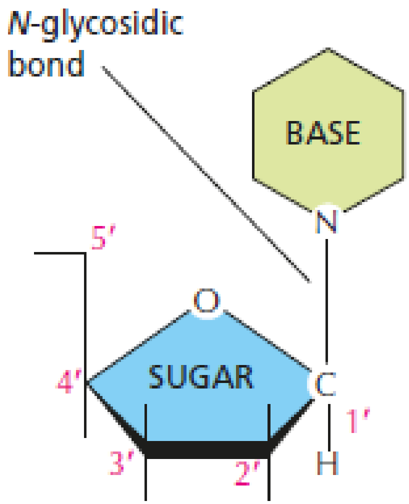
Nucleotides

A nucleotide consists of a nitrogen-containing base, a five-carbon sugar, and one or more phosphate groups.



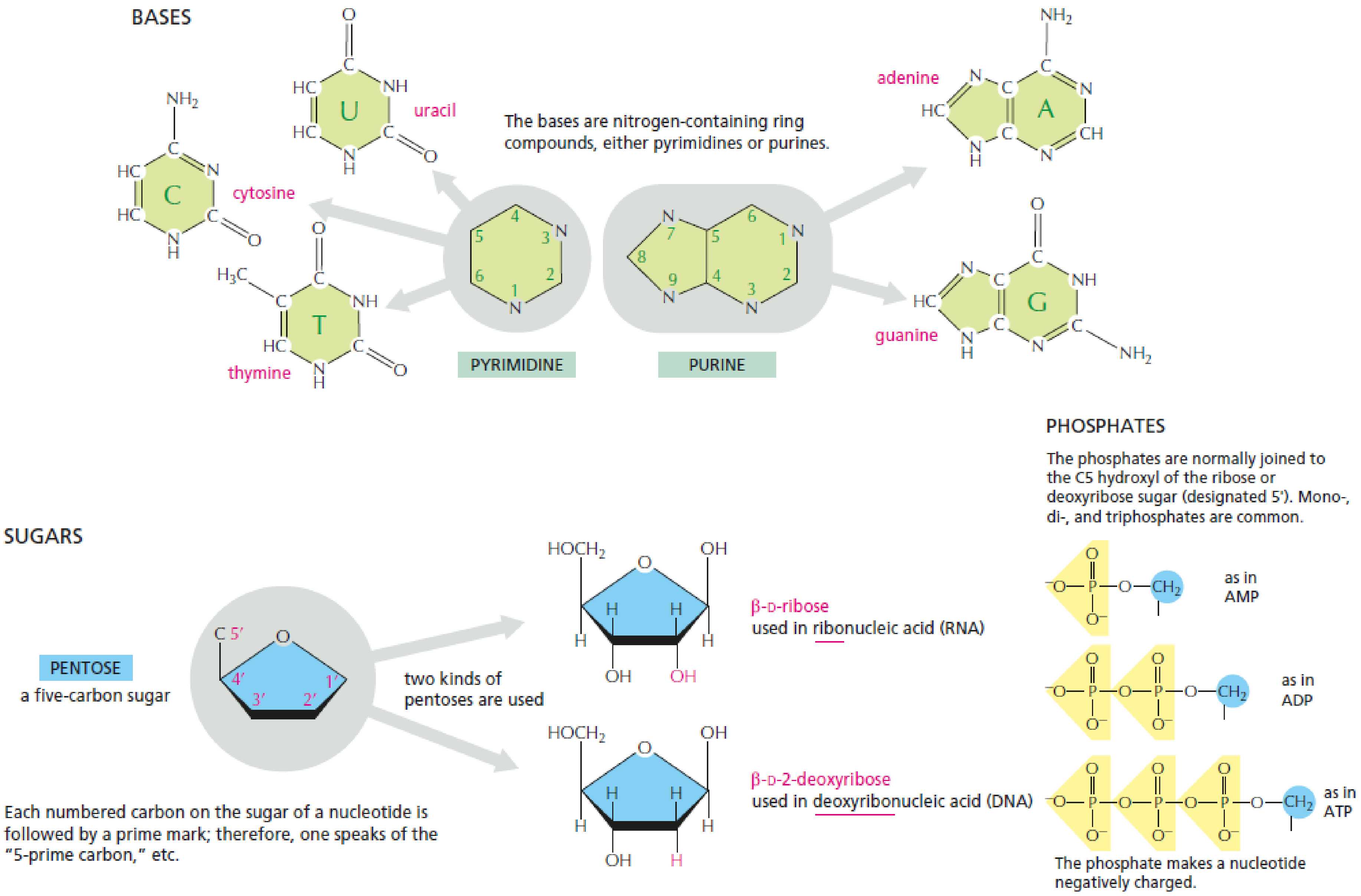
Nucleotides are the subunits of the nucleic acids.

BASE–SUGAR LINKAGE



The base is linked to the same carbon (C1) used in sugar–sugar bonds.

Nucleotides



Nucleotides vs Nucleoside

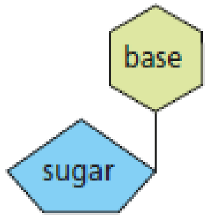
NOMENCLATURE

The names can be confusing, but the abbreviations are clear.

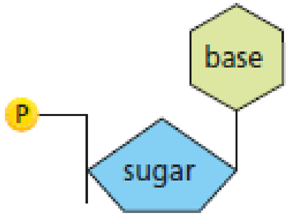
BASE	NUCLEOSIDE	ABBR.
adenine	adenosine	A
guanine	guanosine	G
cytosine	cytidine	C
uracil	uridine	U
thymine	thymidine	T

Nucleotides are abbreviated by three capital letters. Some examples follow:

AMP = adenosine monophosphate
dAMP = deoxyadenosine monophosphate
UDP = uridine diphosphate
ATP = adenosine triphosphate

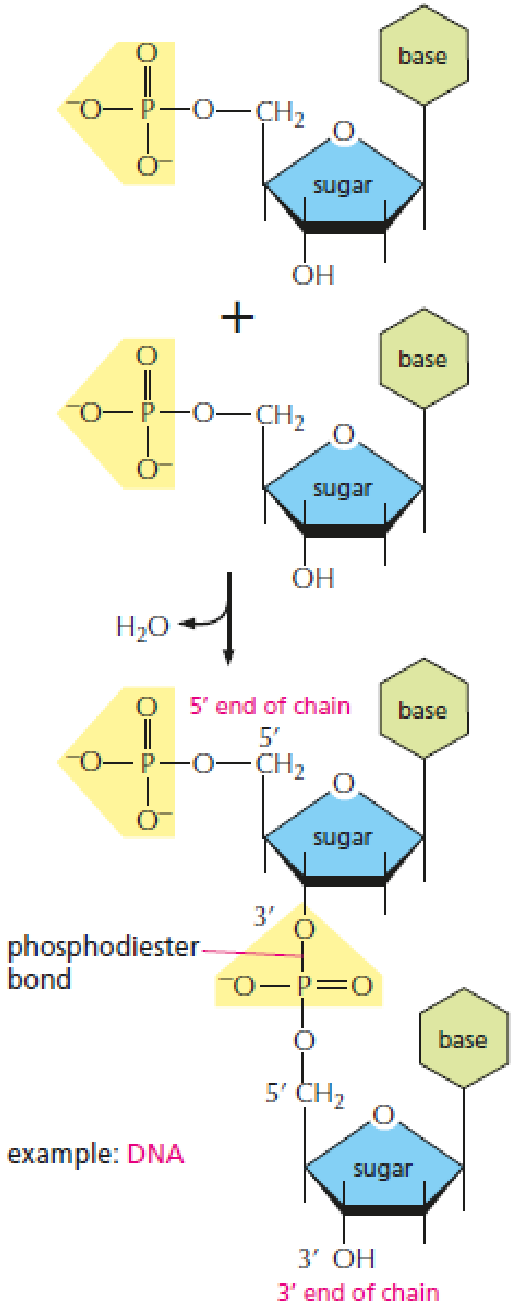


BASE + SUGAR = NUCLEOSIDE



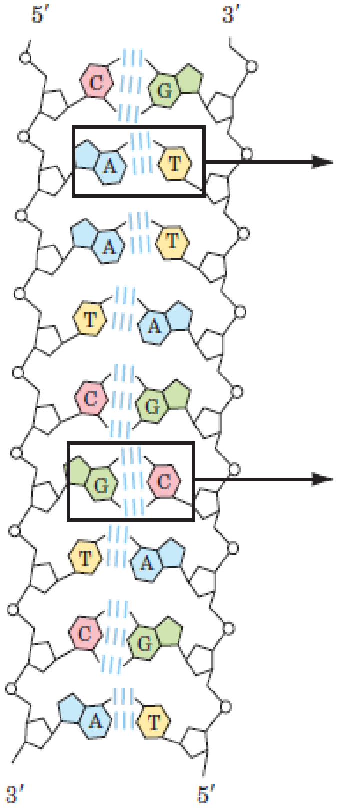
BASE + SUGAR + PHOSPHATE = NUCLEOTIDE

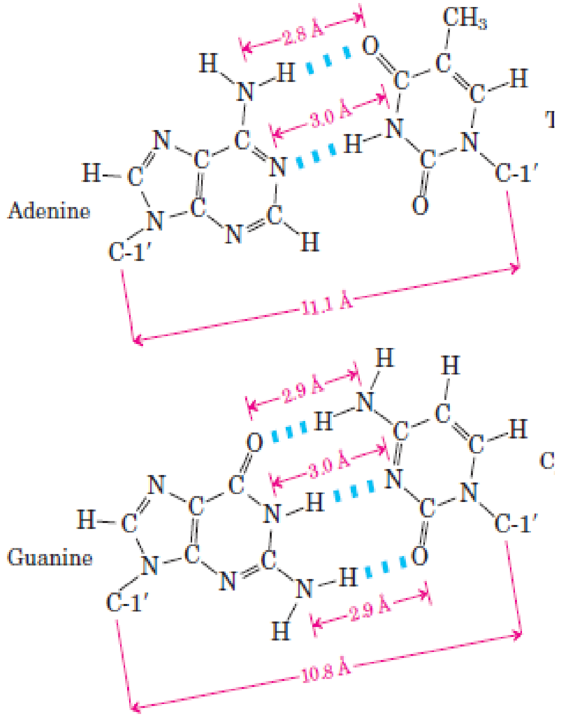
Nucleic Acids

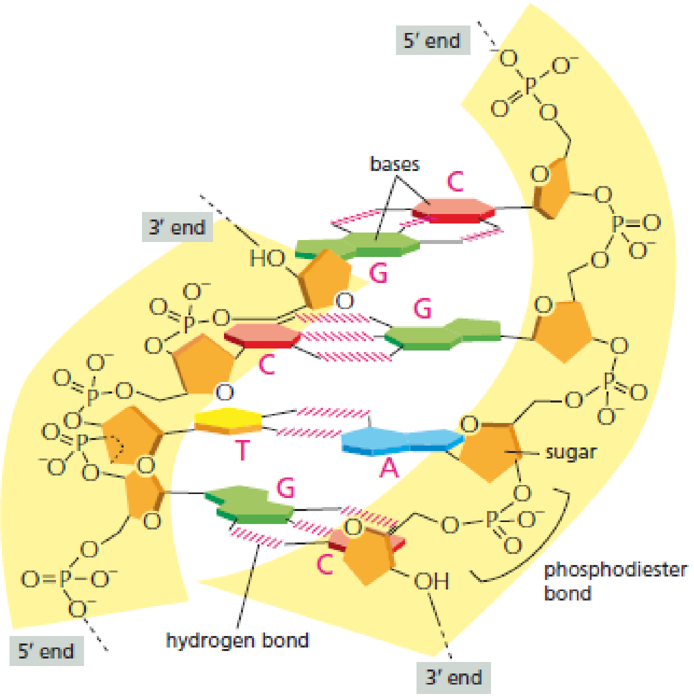


Nucleotides are joined together by phosphodiester bonds between 5’ and 3’ carbon atoms of the sugar ring, via a phosphate group, to form nucleic acids. The linear sequence of nucleotides in a nucleic acid chain is commonly abbreviated by a one-letter code, such as AGCTTACA, with the 5’ end of the chain at the left.

Hydrogen-Bonding Patterns



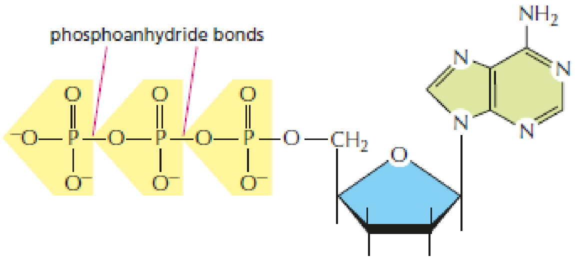




Nucleotides

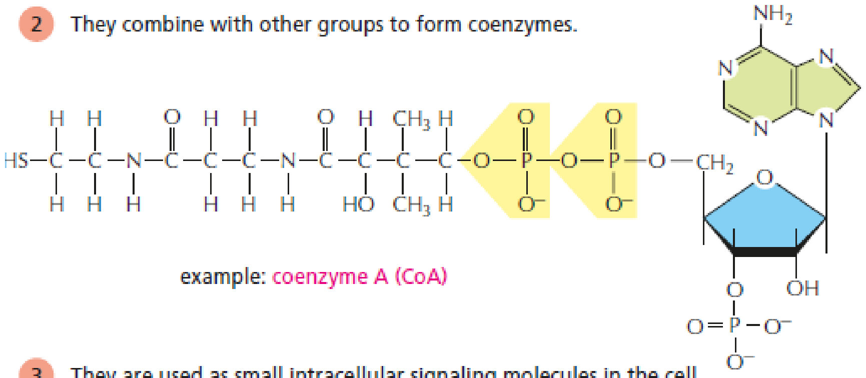
NUCLEOTIDES HAVE MANY OTHER FUNCTIONS

1. They carry chemical energy in their easily hydrolyzed phosphoanhydride bonds.



example: ATP (or ATP)

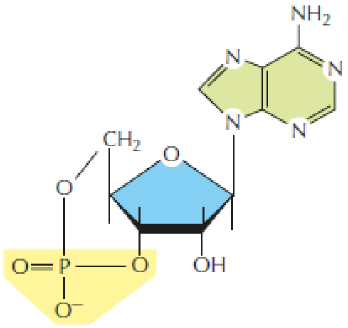
2. They combine with other groups to form coenzymes.



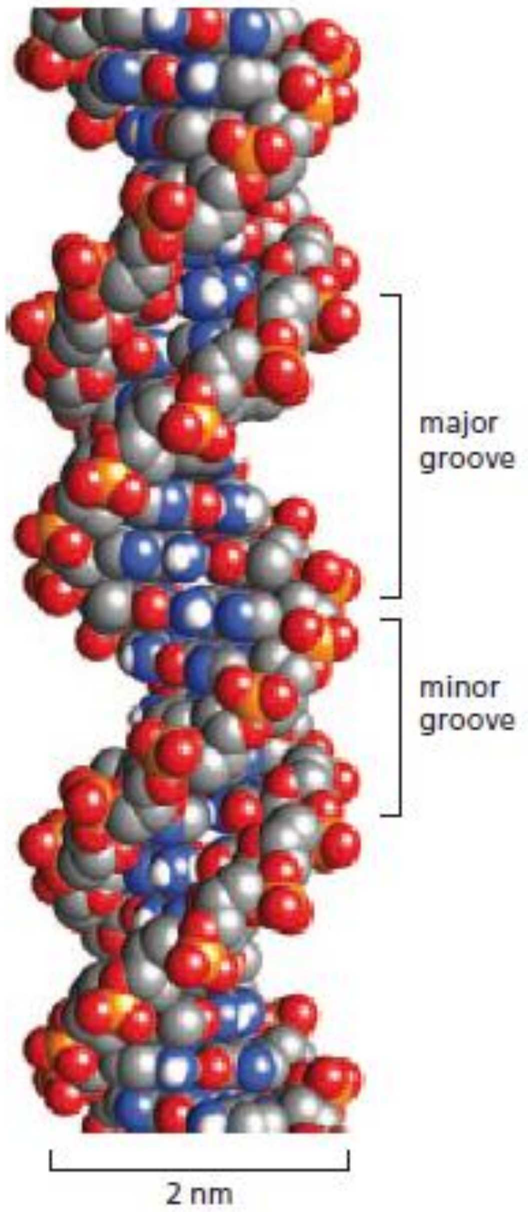
example: coenzyme A (CoA)

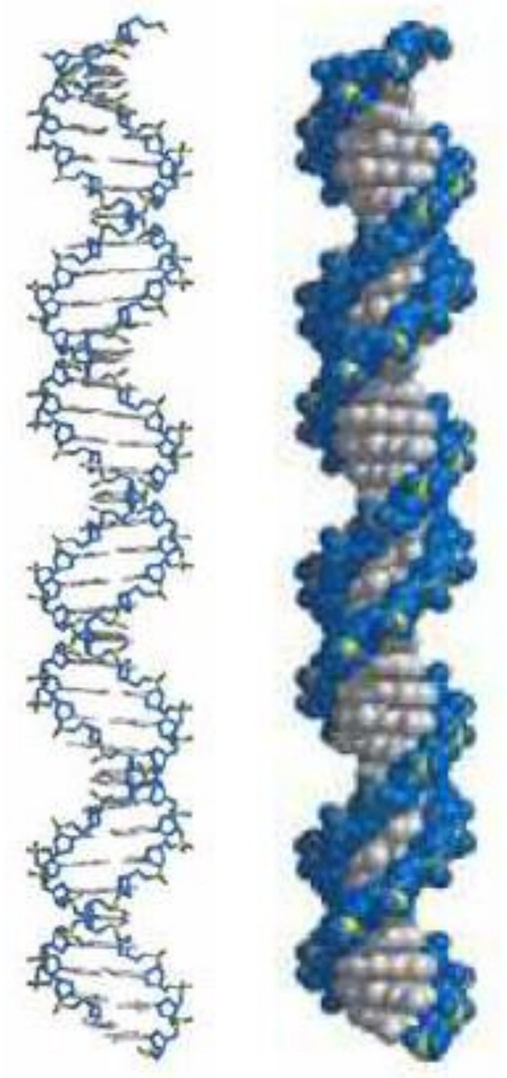
3. They are used as small intracellular signaling molecules in the cell.

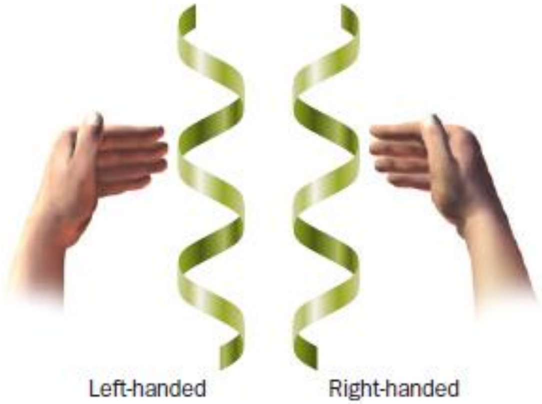
example: cyclic AMP



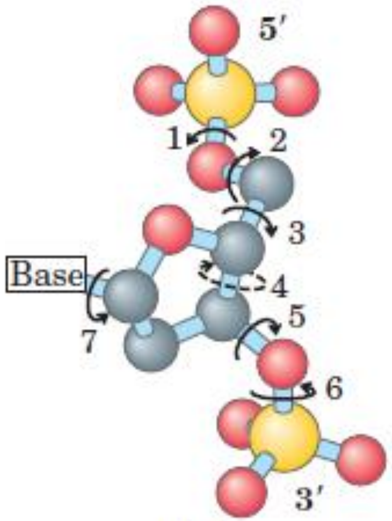
Structure of DNA

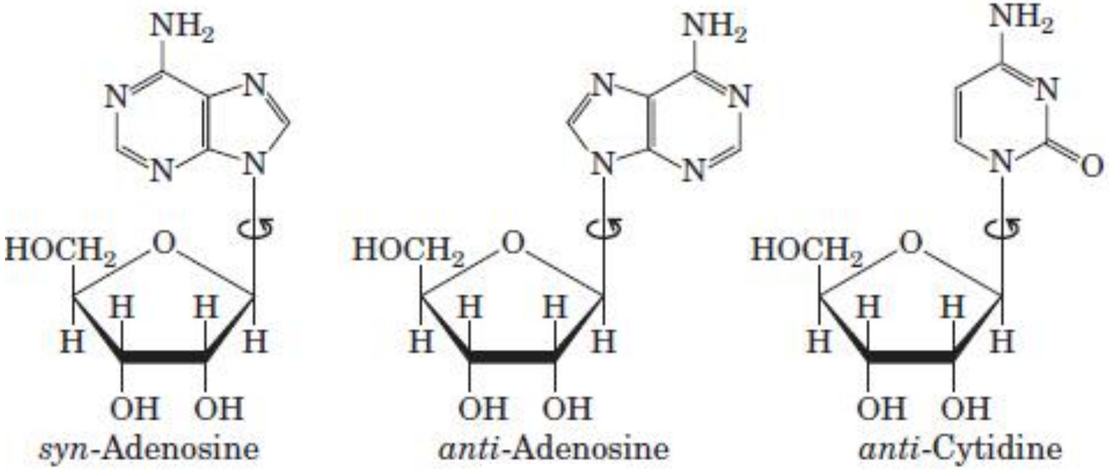




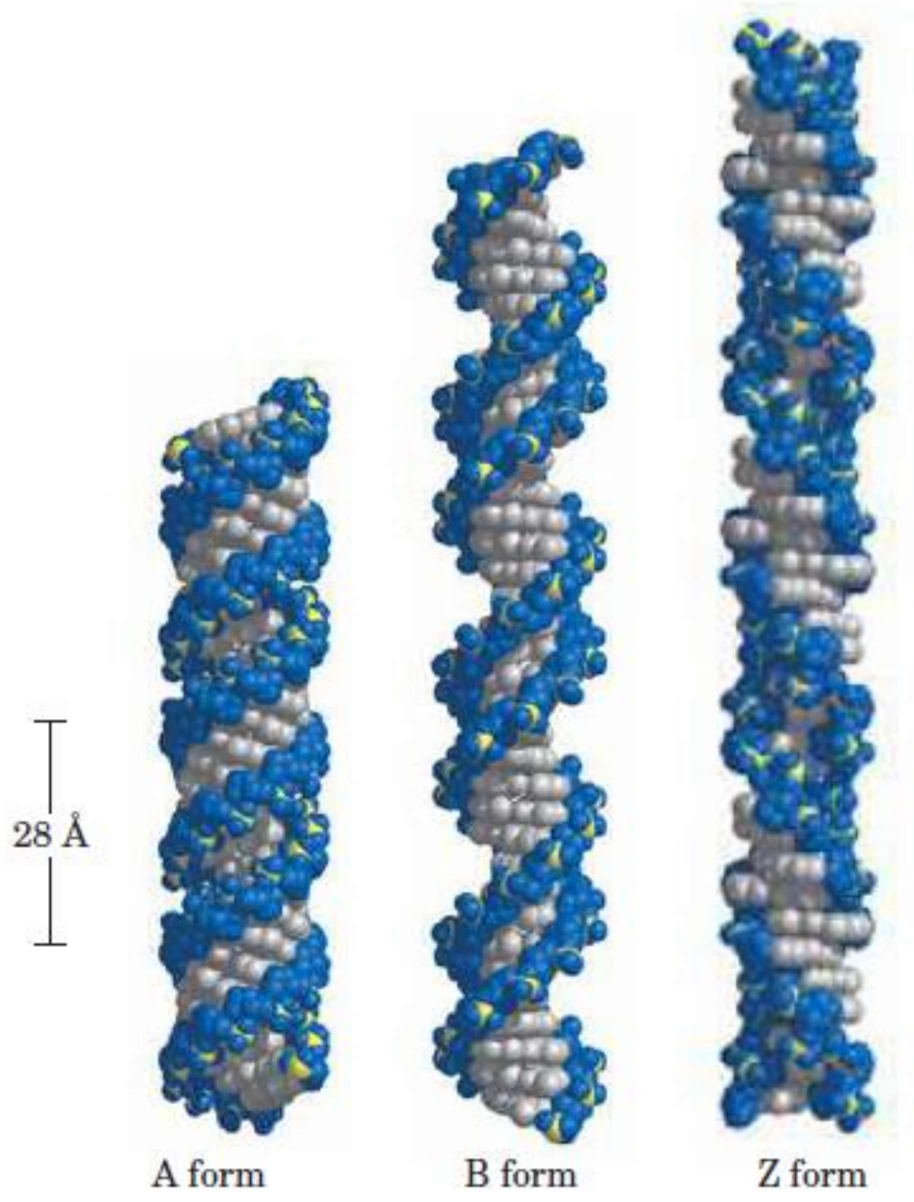


Structural variation in DNA

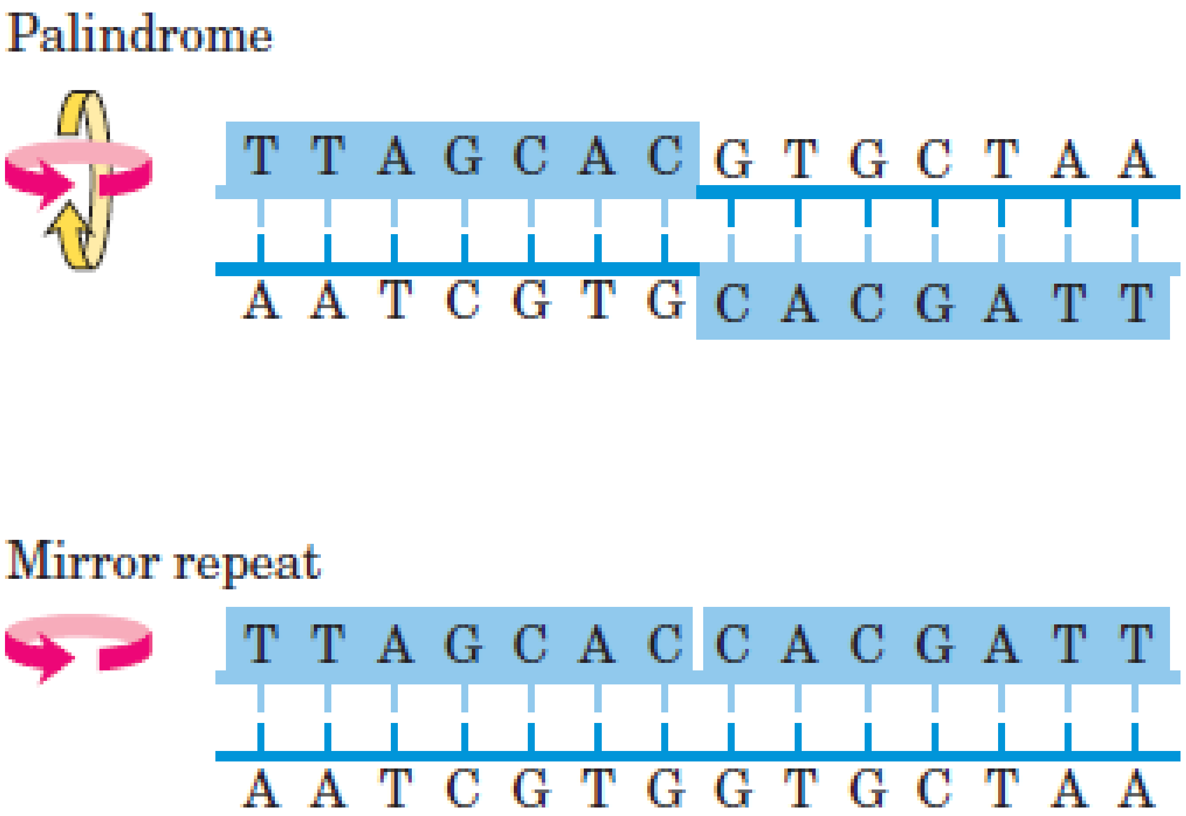




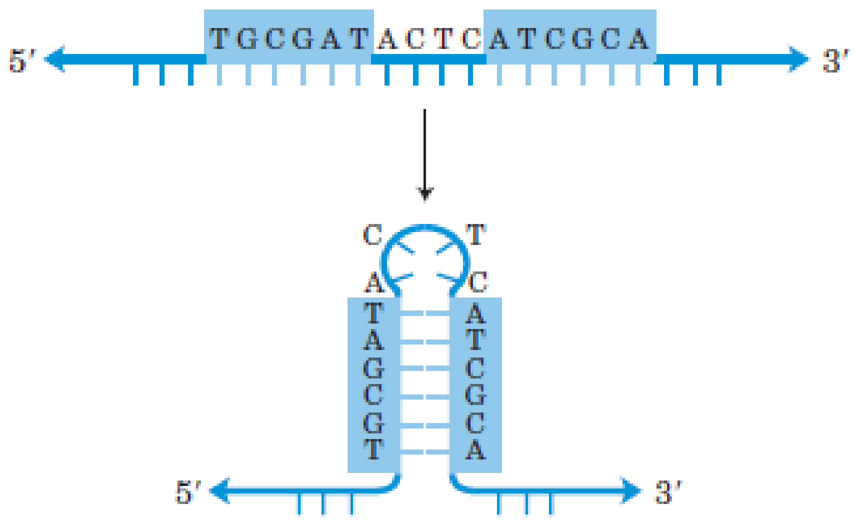
Structural variation in DNA



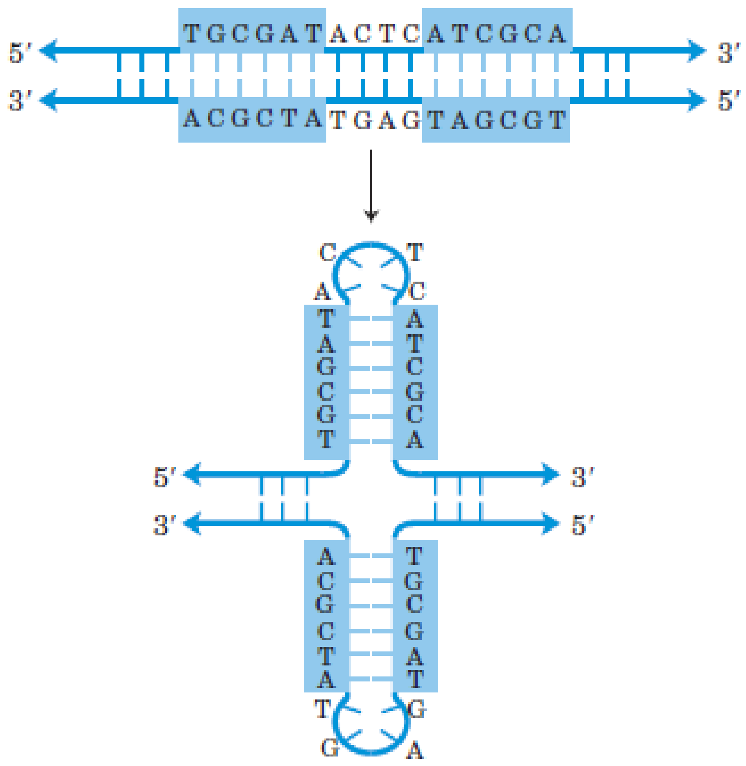
Certain DNA Sequences Adopt Unusual Structures



Hairpins and cruciforms

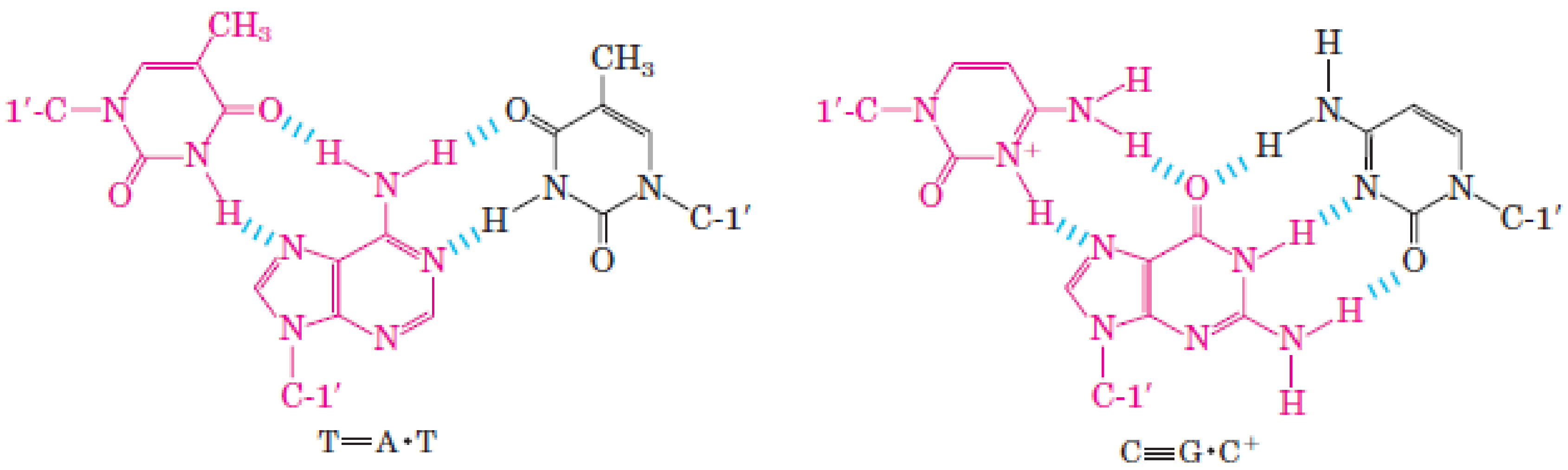


Hairpin

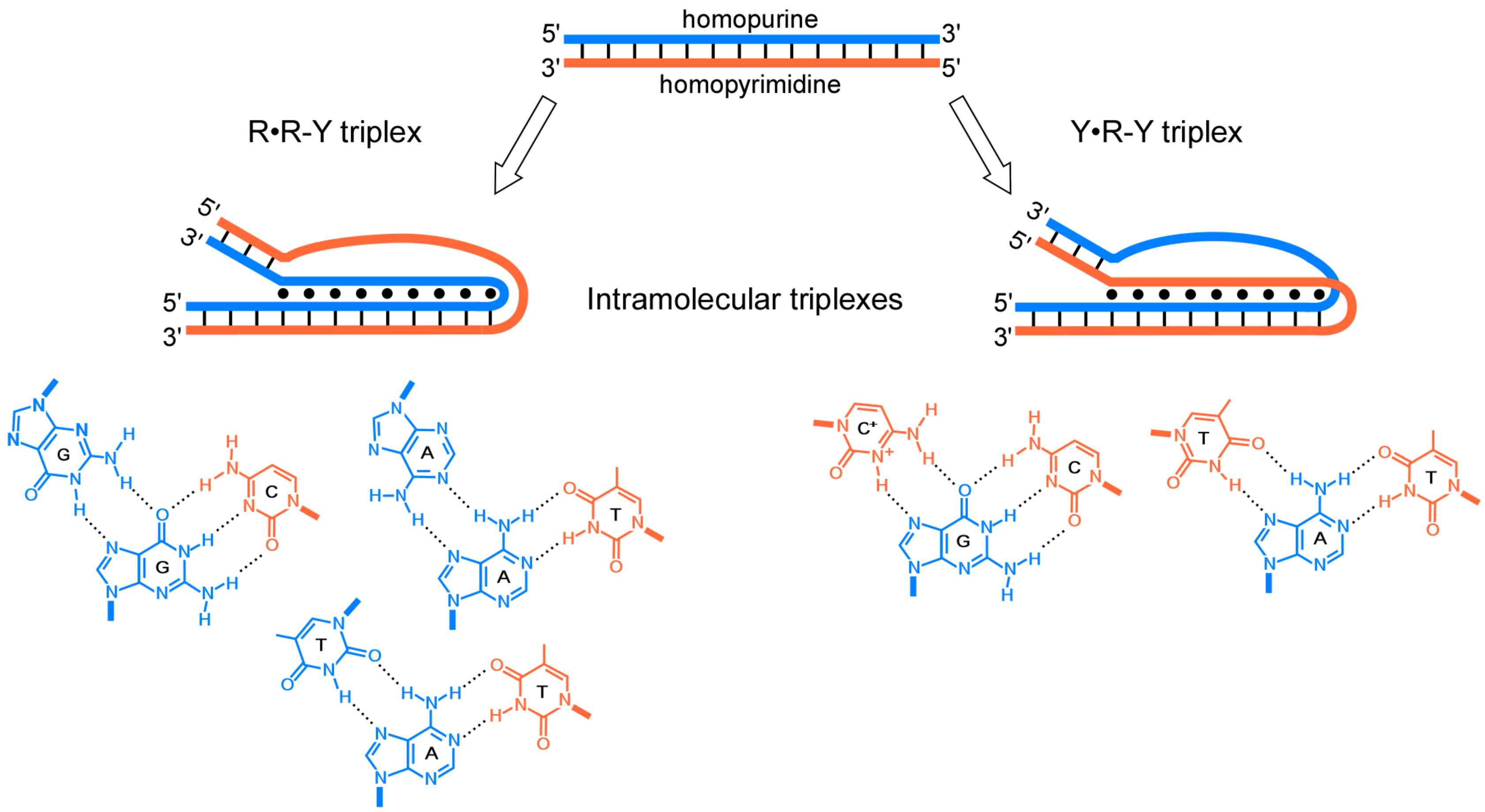


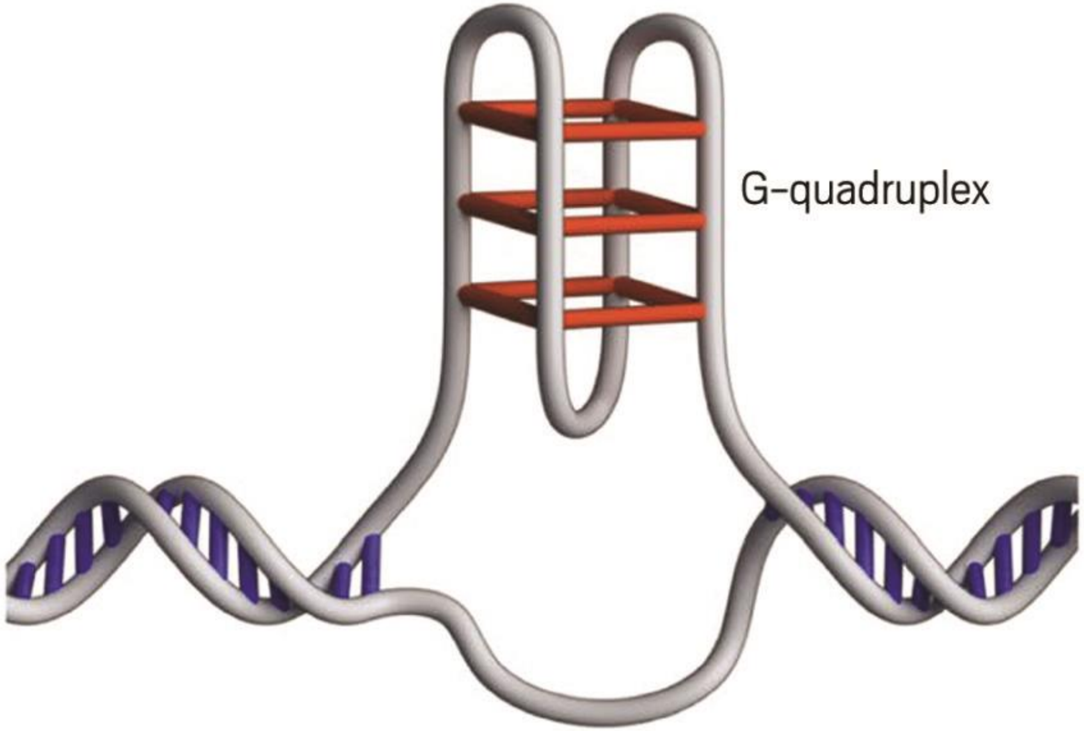
Cruciform

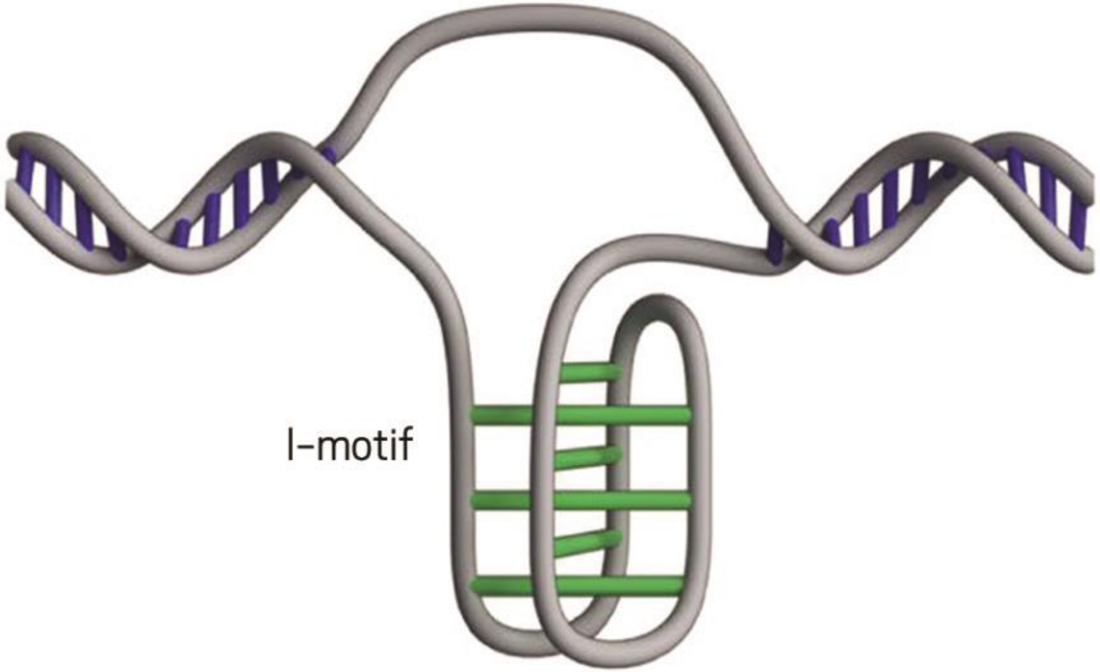
Several unusual DNA structures involve three or even four DNA strands



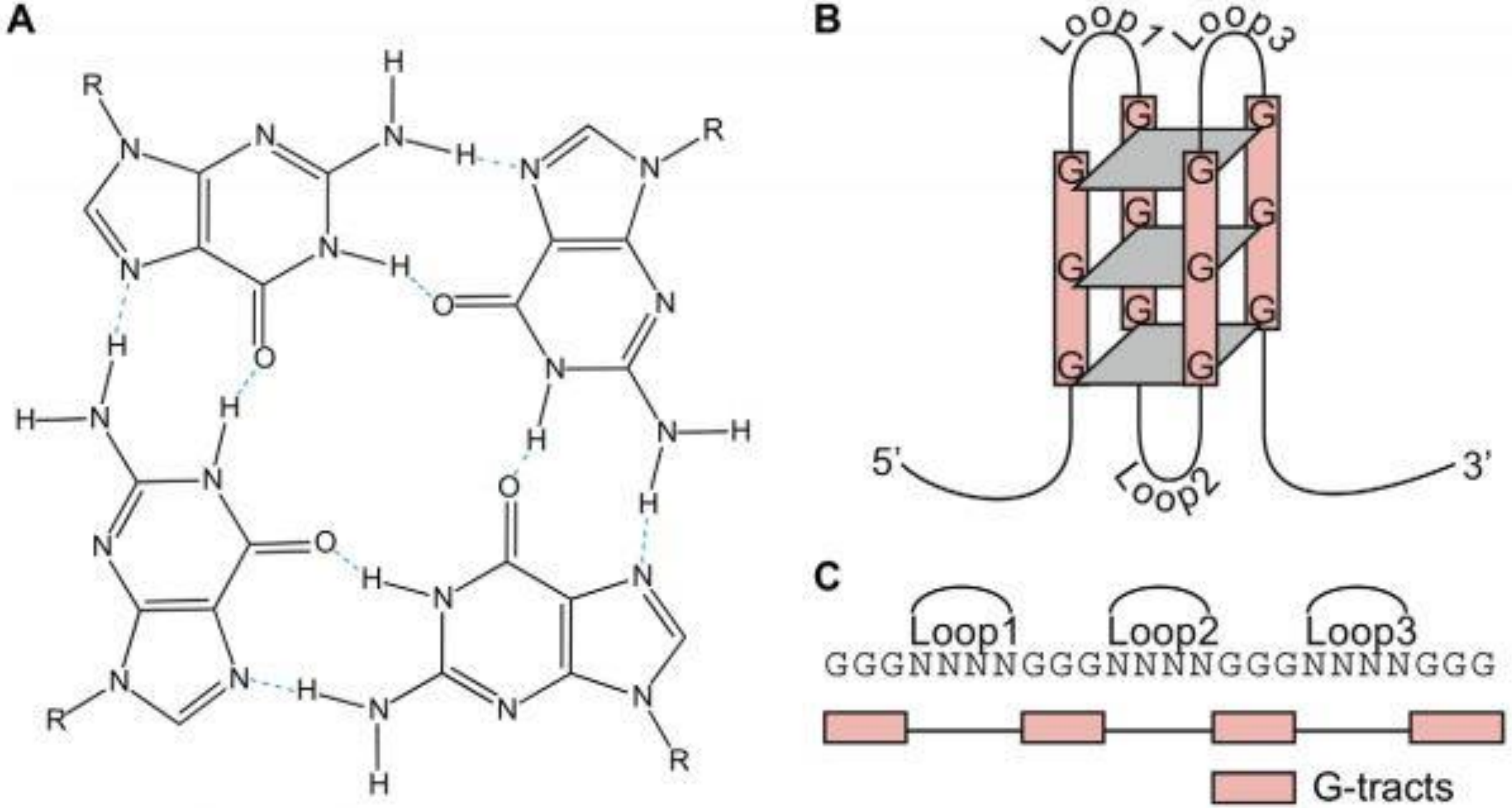
The atoms that participate in the hydrogen bonding of **triplex DNA**, are often referred to as **Hoogsteen positions**, and the non-Watson-Crick pairing is called **Hoogsteen pairing**, after Karst Hoogsteen, who in 1963 first recognized the potential for these unusual pairings


homopurine
homopyrimidine
5'
3'
R•R-Y triplex
Y•R-Y triplex
Intramolecular triplexes
G
C
A
T
C+

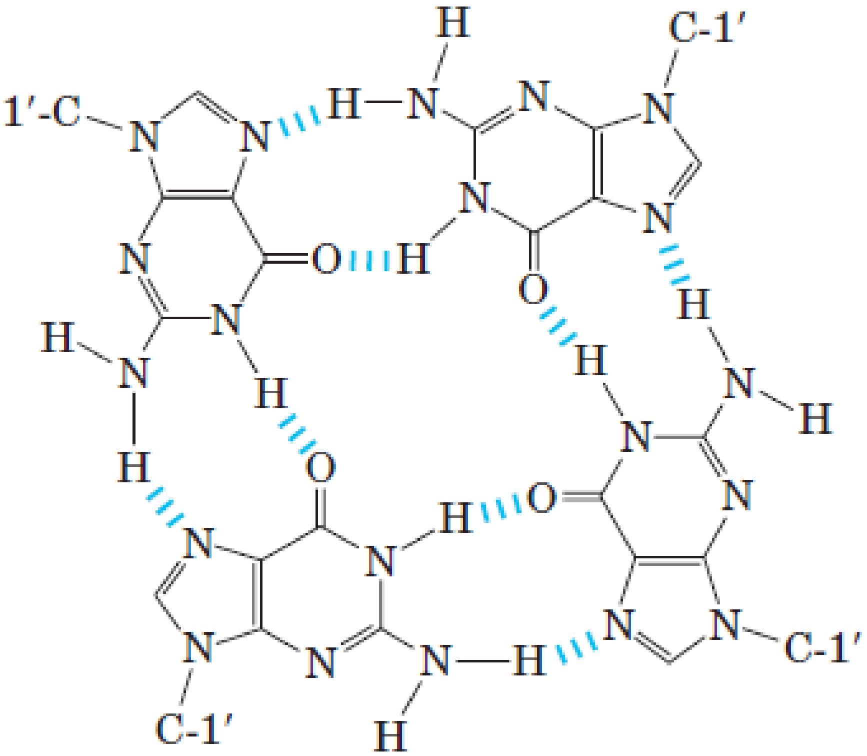

G-quadruplex

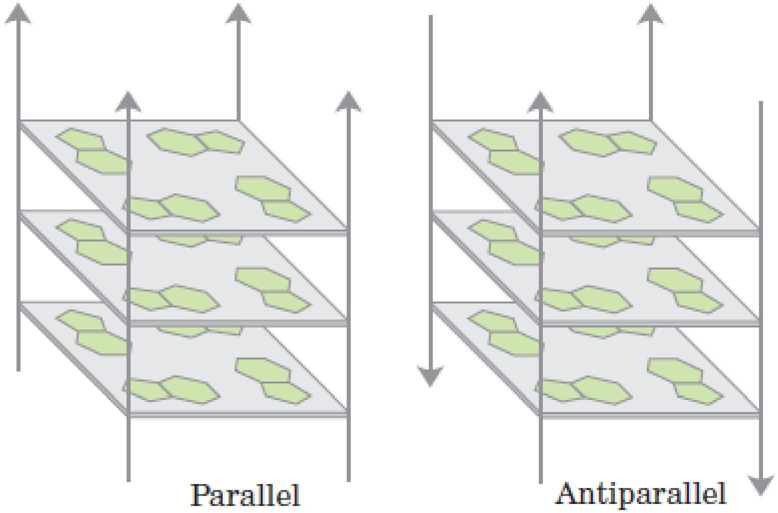

I-motif

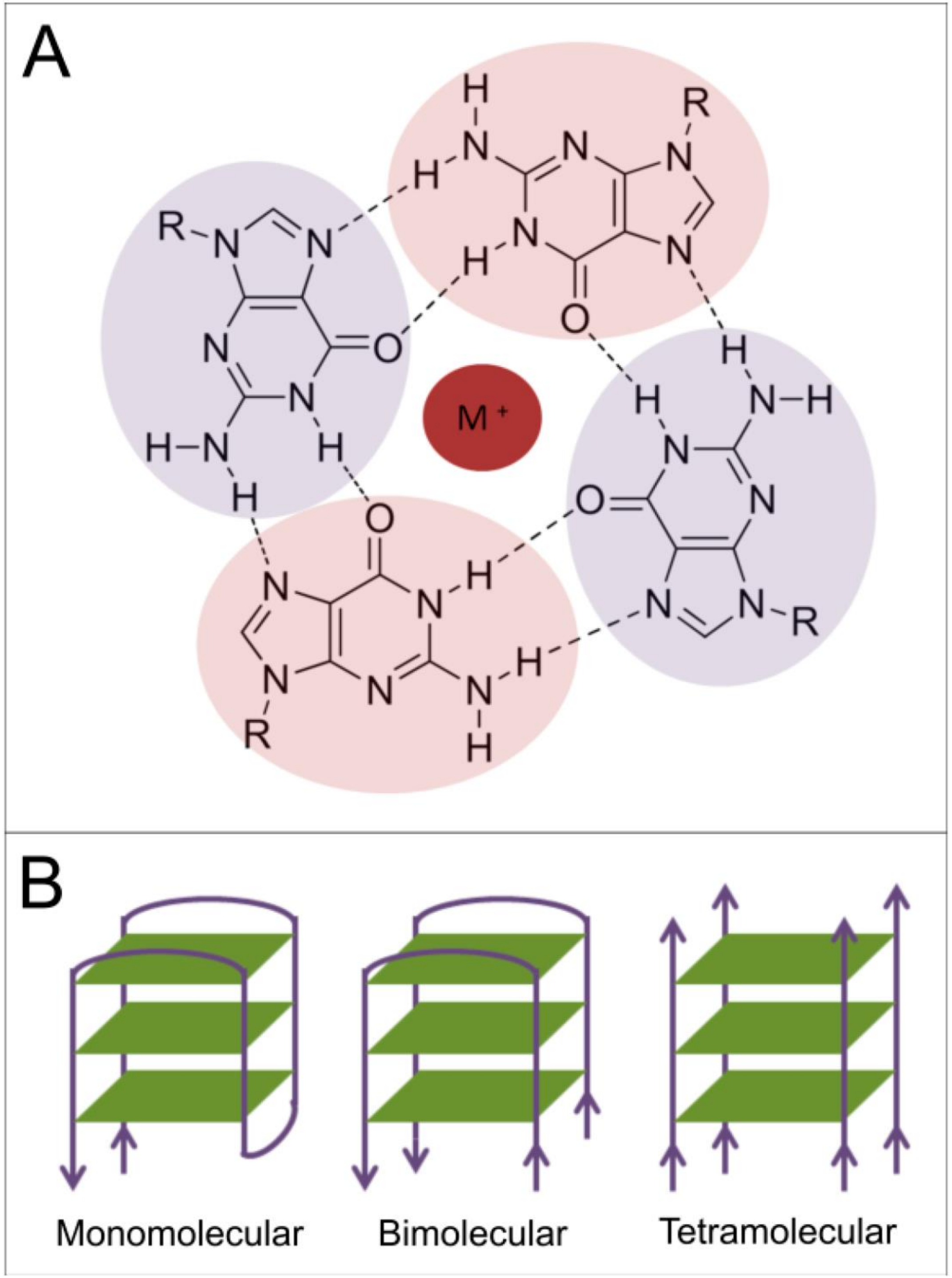
G-quadruplexes



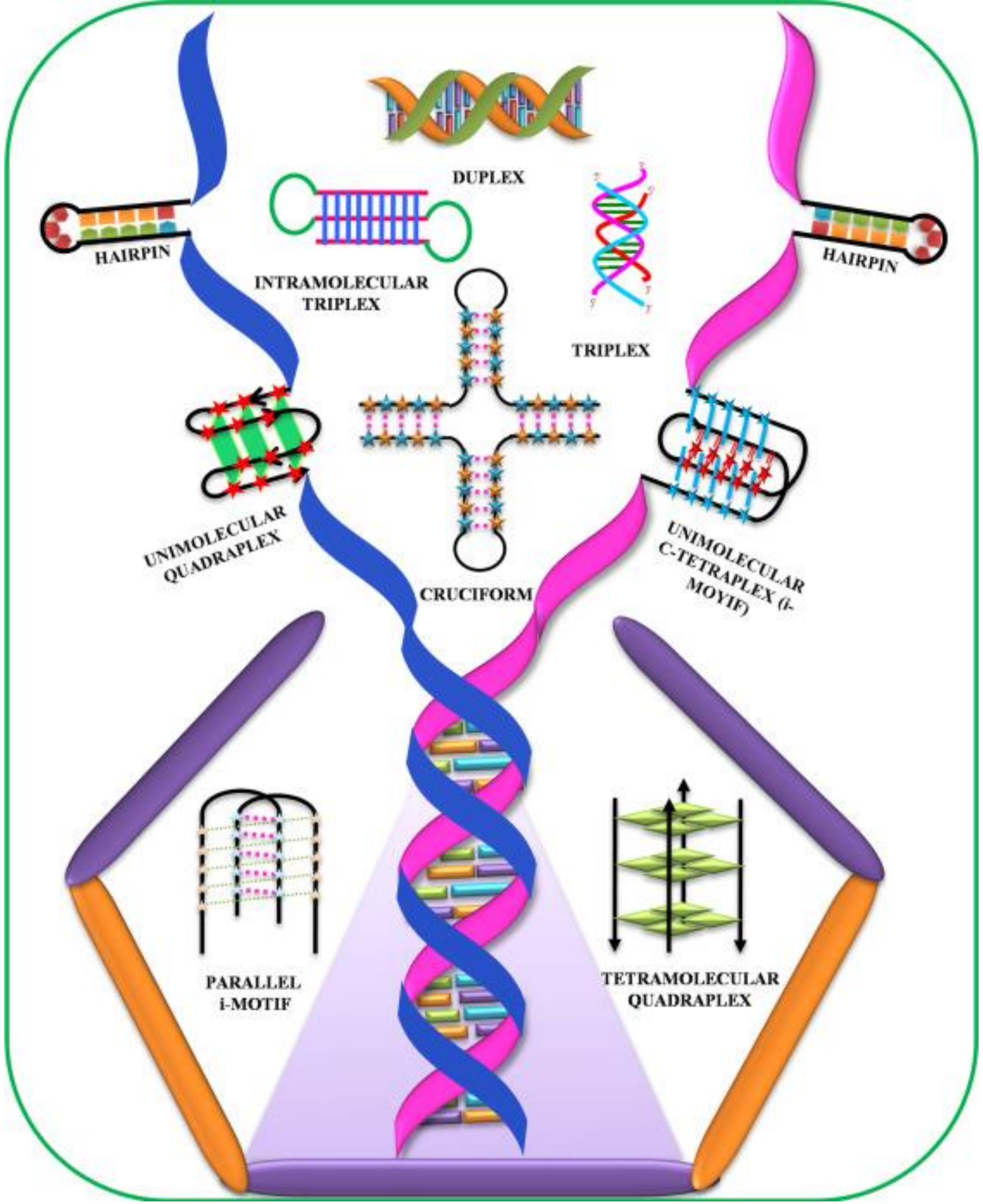
G-quadruplexes







A
M+
B
Monomolecular
Bimolecular
Tetramolecular

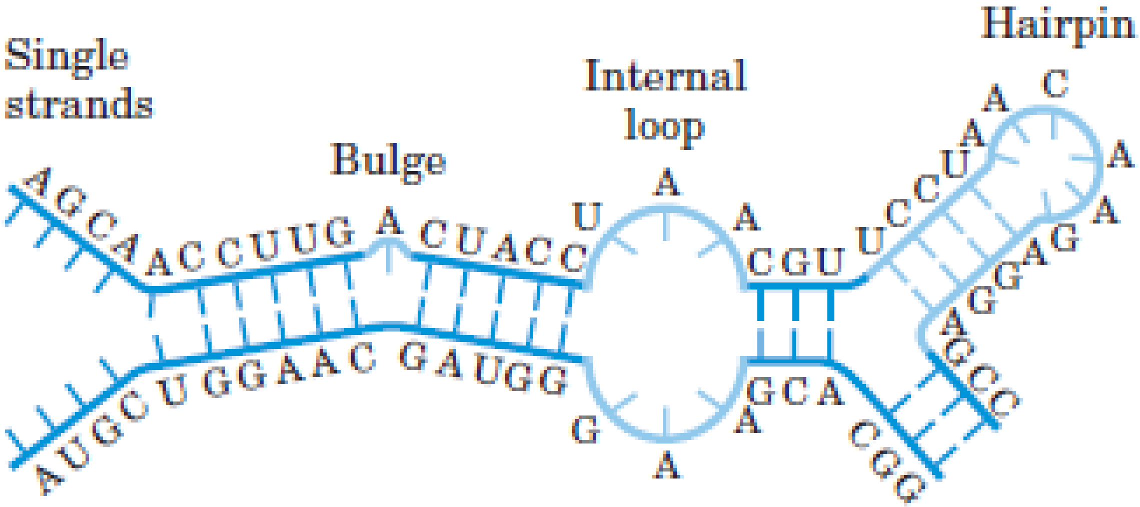


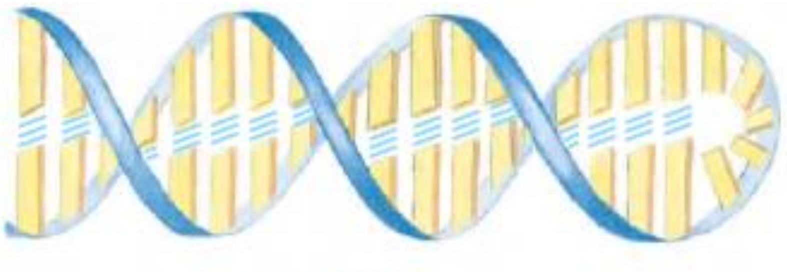
Intramolecular

Vs

Intermolecular

Structure of RNA

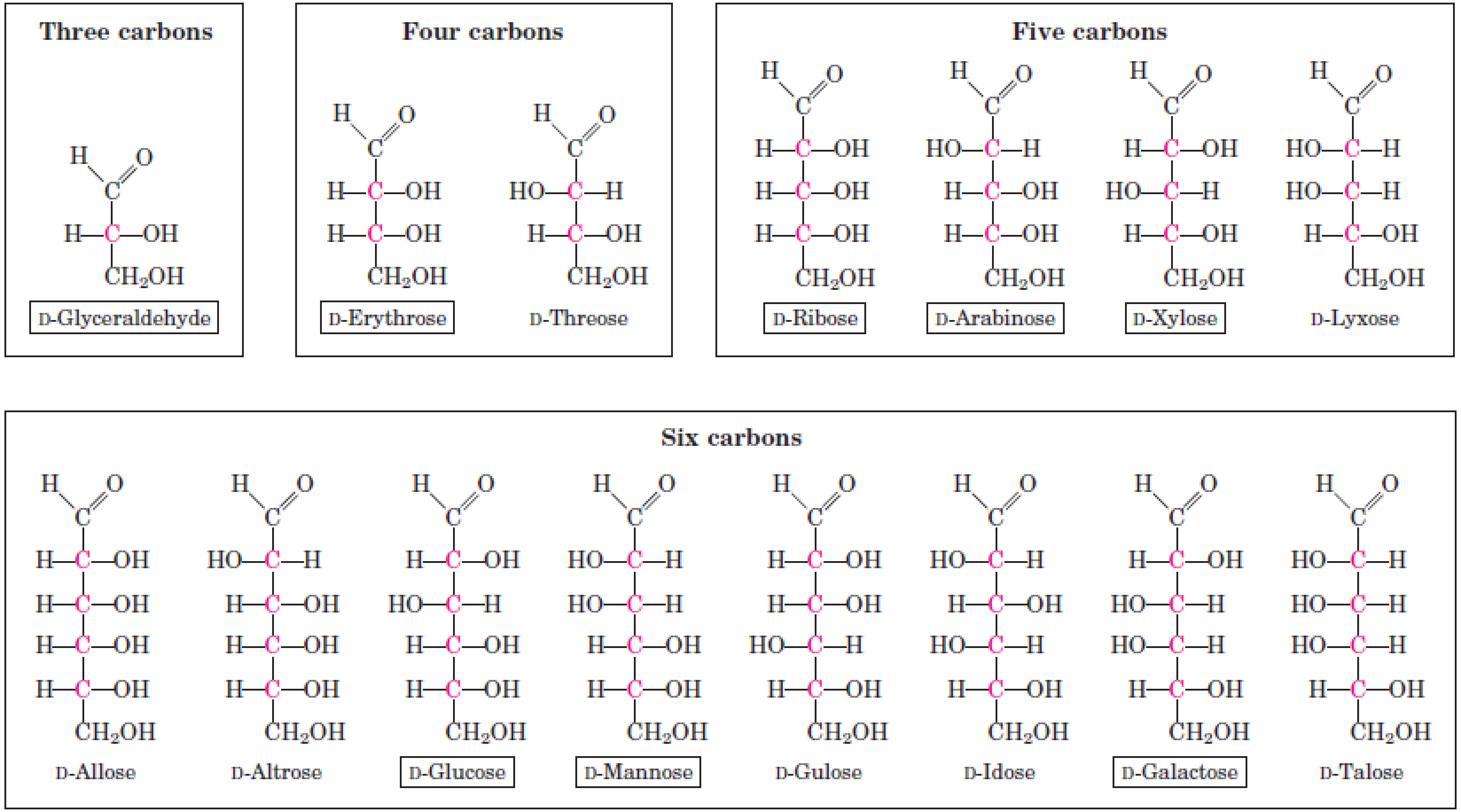




Hairpin double helix

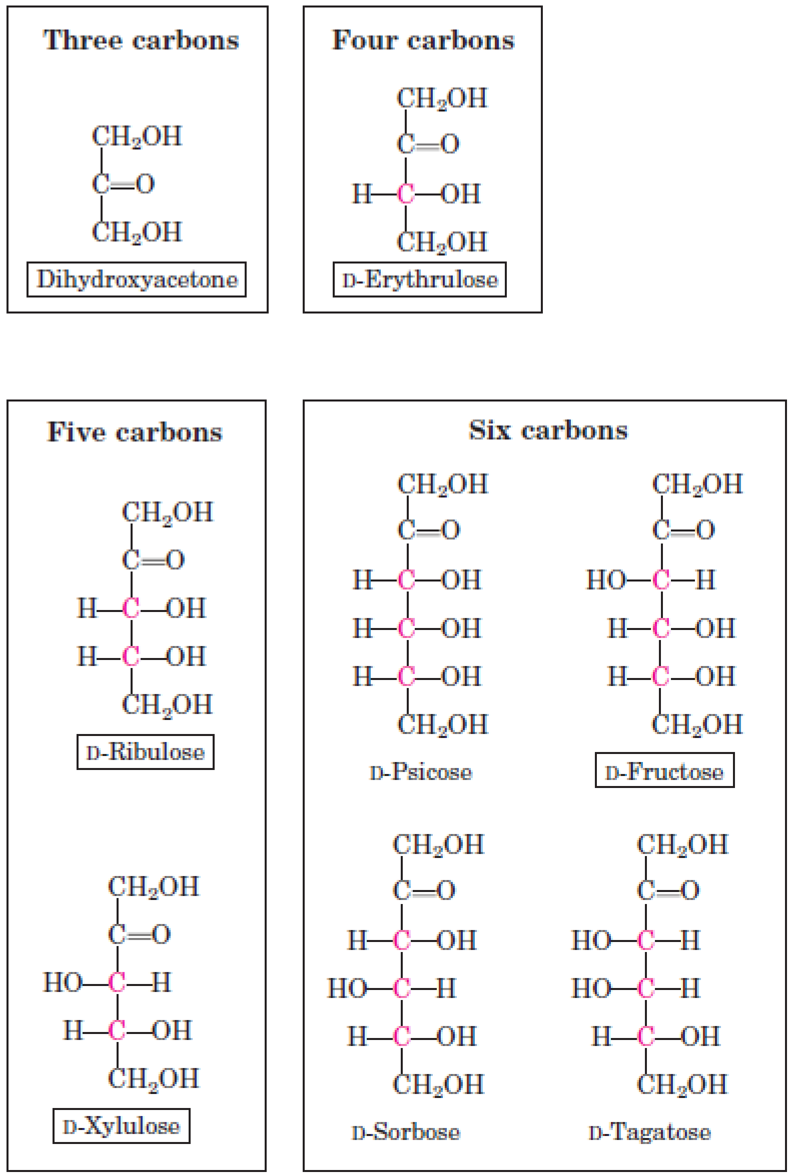
The paired regions generally have an A-form right-handed helix, as shown for a hairpin

Monosaccharides



D-Aldoses

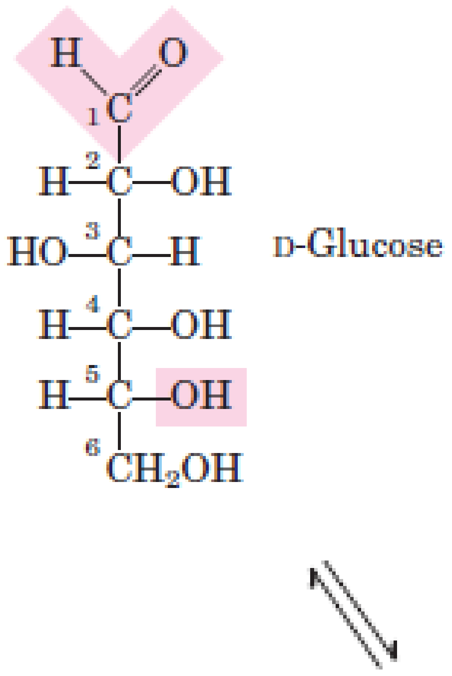
Monosaccharides

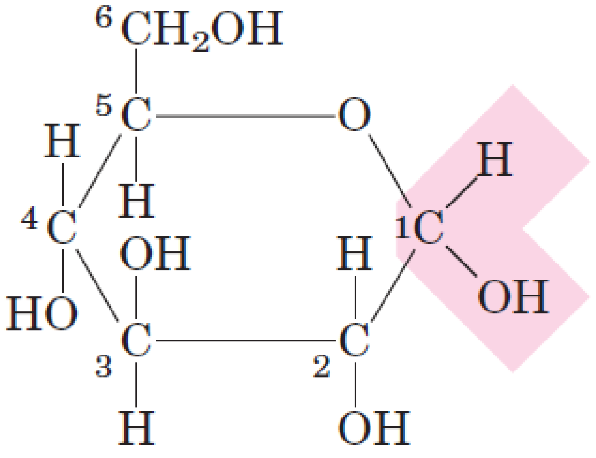


D-Ketoses

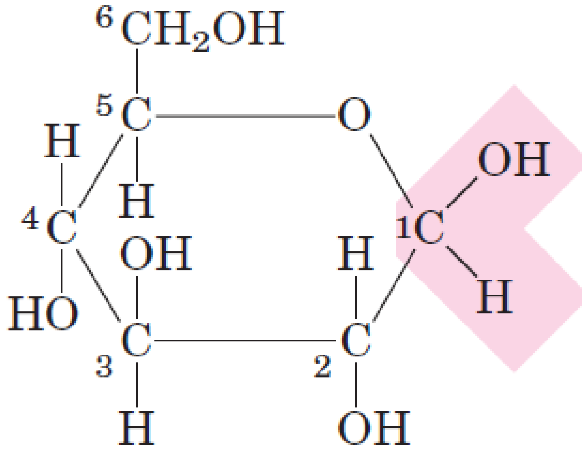
Monosaccharides

Formation of the two cyclic forms of D-glucose

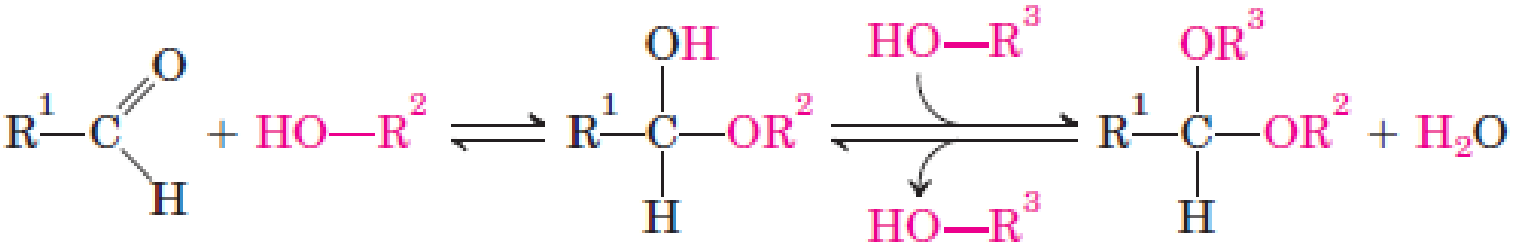




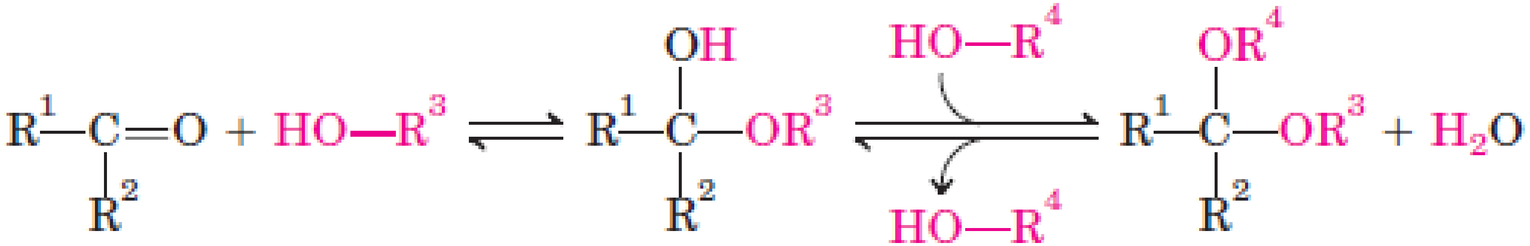
α-D-Glucopyranose



β-D-Glucopyranose



Aldehyde Alcohol Hemiacetal Acetal

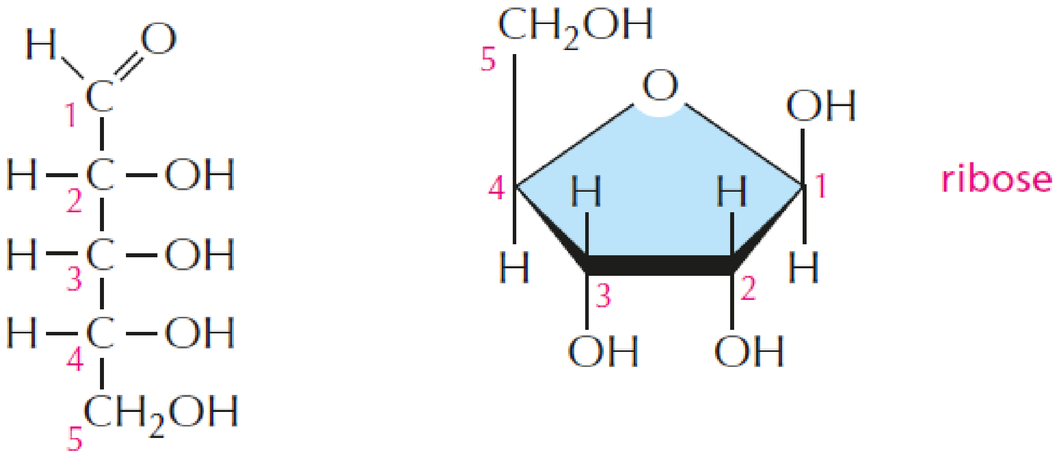


Ketone Alcohol Hemiketal Ketal

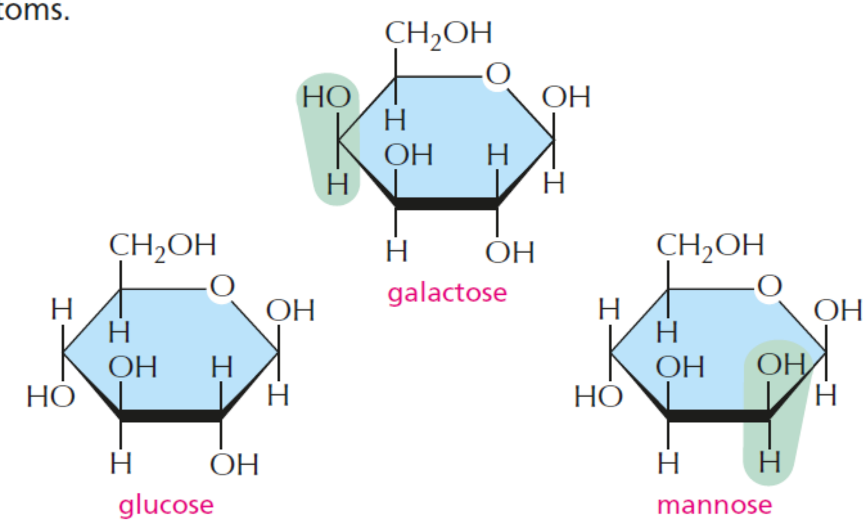
Formation of these ring structures is the result of a general reaction between alcohols and aldehydes or ketones to form derivatives called **hemiacetals** or **hemiketals**

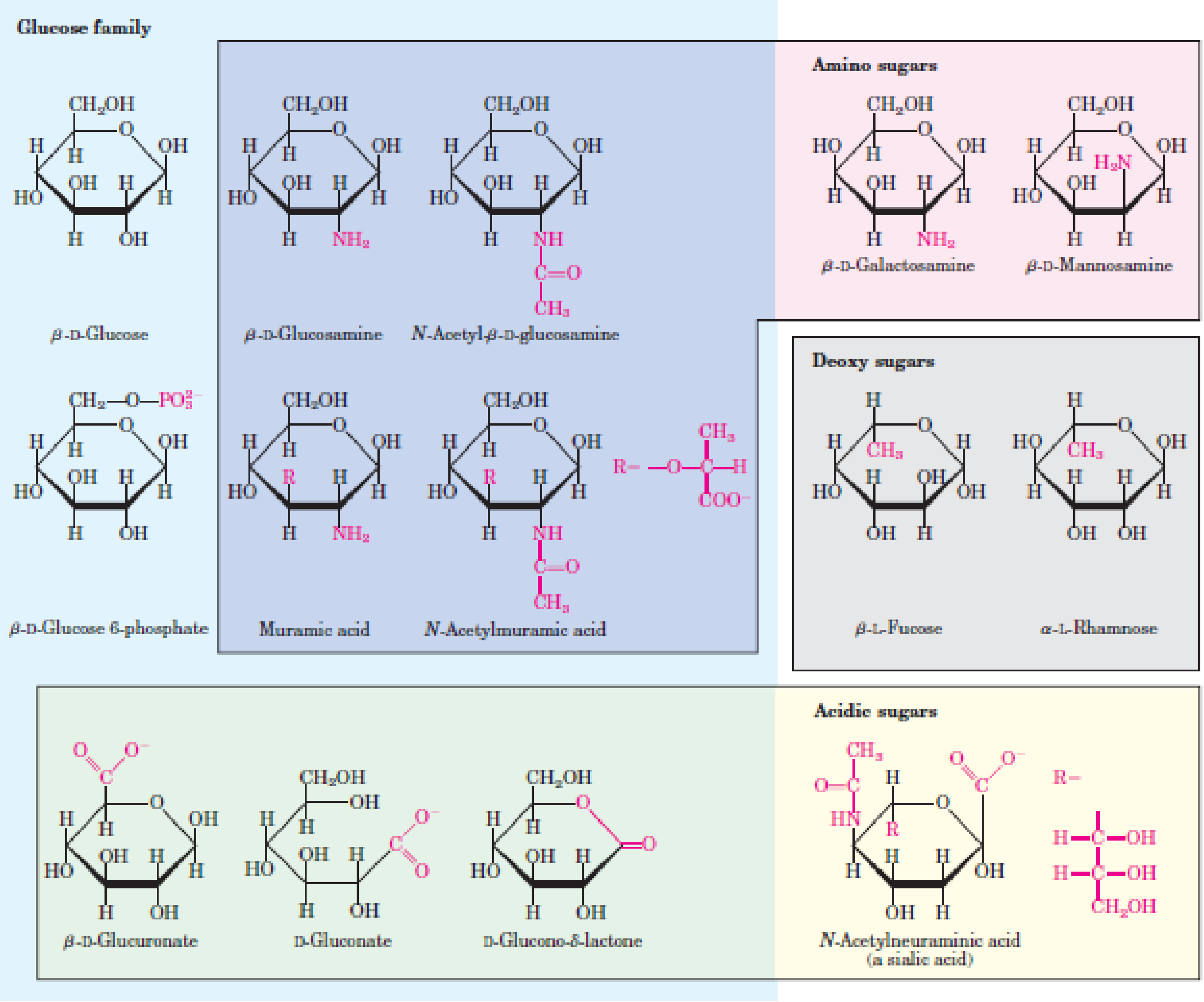
The hemiacetal (or carbonyl) carbon atom is called the **anomeric carbon**.

Monosaccharides



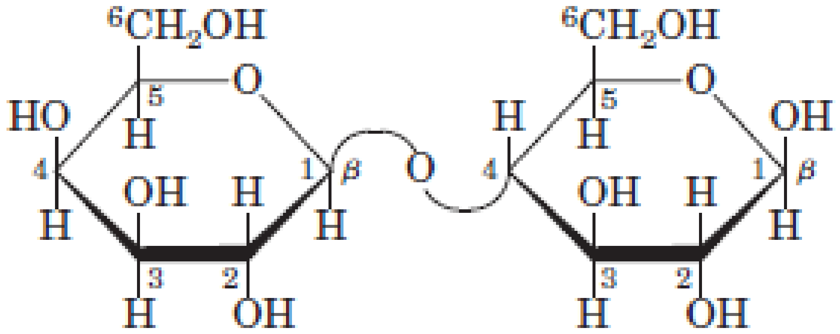
Many monosaccharides differ only in the spatial arrangement of atoms—that is, they are isomers. For example, glucose, galactose, and mannose have the same formula ($C_6H_{12}O_6$) but differ in the arrangement of groups around one or two carbon atoms.



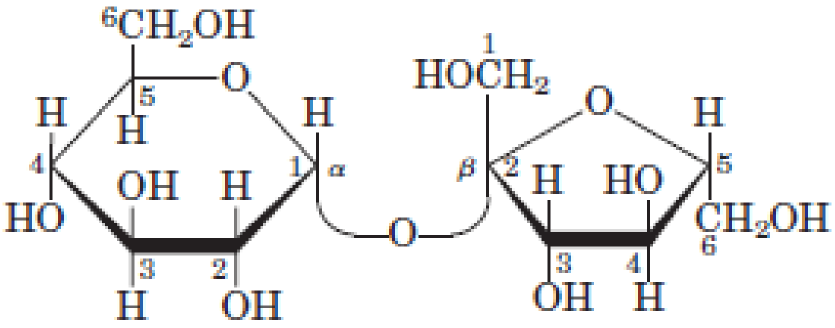


Glucose family
Amino sugars
Deoxy sugars
Acidic sugars
β-D-Glucose
β-D-Glucosamine
N-Acetyl-β-D-glucosamine
β-D-Galactosamine
β-D-Mannosamine
β-D-Glucose 6-phosphate
Muramic acid
N-Acetylmuramic acid
β-L-Fucose
α-L-Rhamnose
β-D-Glucuronate
D-Gluconate
D-Glucono-δ-lactone
N-Acetylneuraminic acid
(a sialic acid)

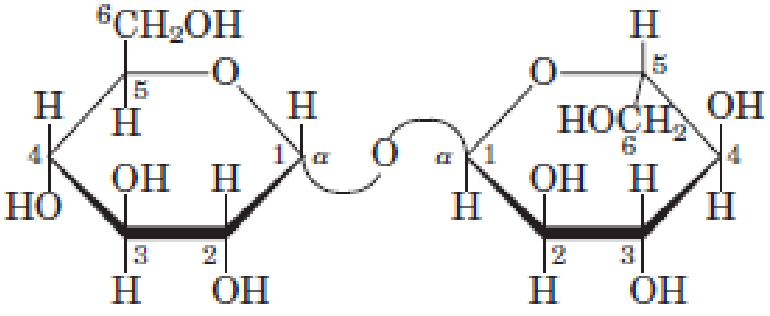
Disaccharides



Lactose (β form)



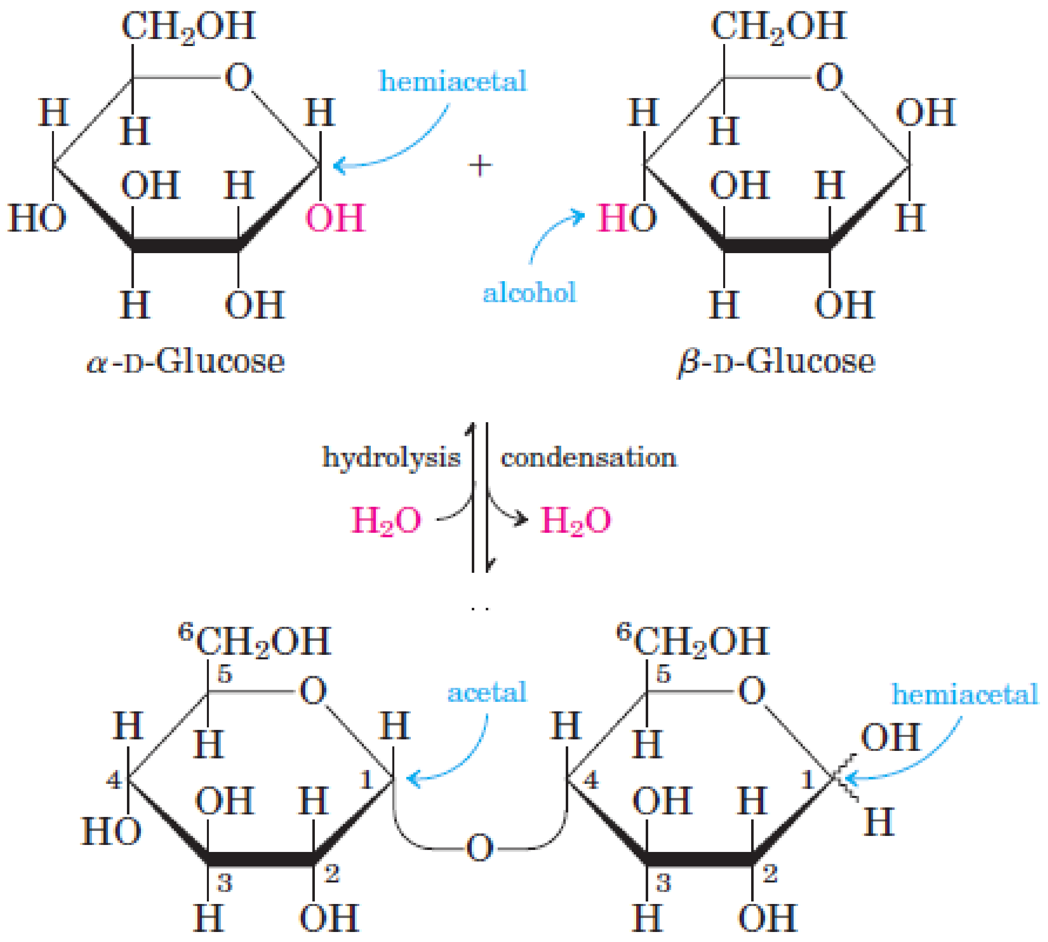
Sucrose



Trehalose

Three common disaccharides are

maltose (glucose + glucose)
lactose (galactose + glucose)
sucrose (glucose + fructose)

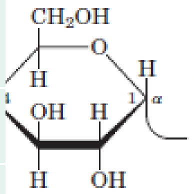


Maltose
α-D-glucopyranosyl-(1→4)-D-glucopyranose

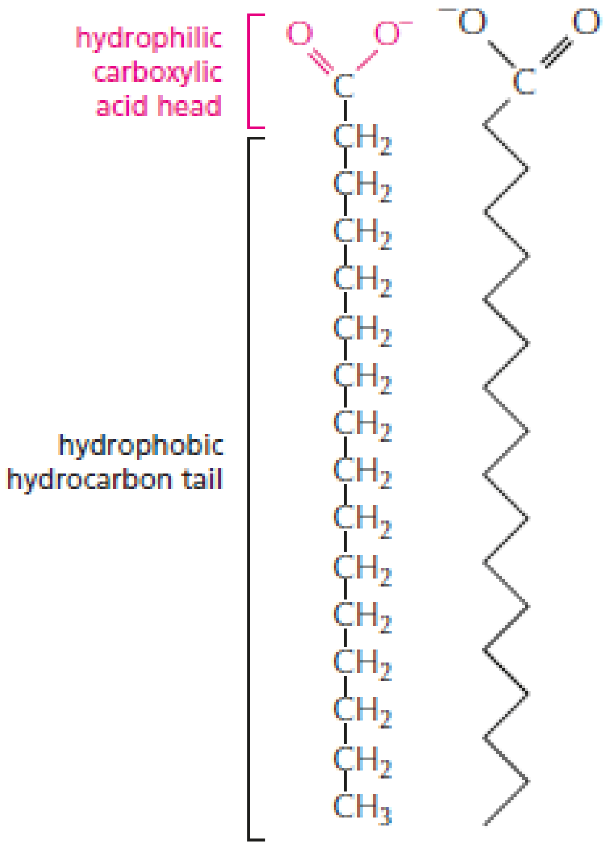
Polysaccharides

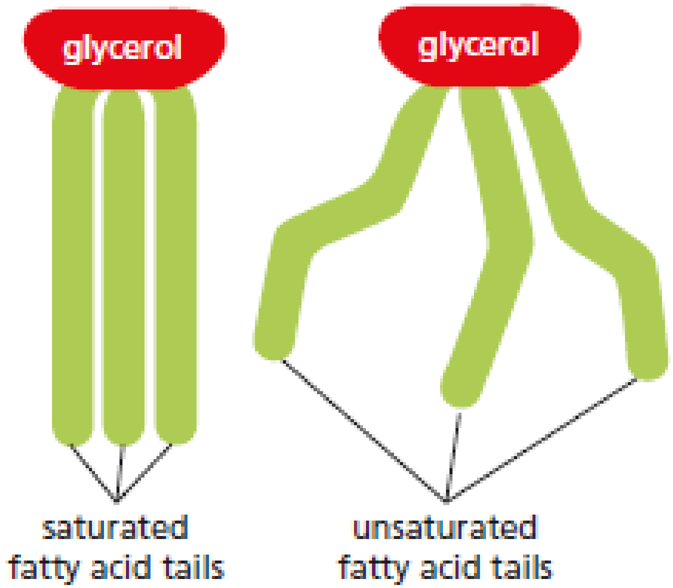
Homopo
Unbranc

Polymer	Type*	Repeating unit†	Size (number of monosaccharide units)	Roles/significance
Starch				Energy storage: in plants
Amylose	Homo-	(α1→4)Glc, linear	50–5,000	
Amylopectin	Homo-	(α1→4)Glc, with (α1→6)Glc branches every 24–30 residues	Up to 10^6	
Glycogen	Homo-	(α1→4)Glc, with (α1→6)Glc branches every 8–12 residues	Up to 50,000	Energy storage: in bacteria and animal cells
Cellulose	Homo-	(β1→4)Glc	Up to 15,000	Structural: in plants, gives rigidity and strength to cell walls
Chitin	Homo-	(β1→4)GlcNAc	Very large	Structural: in insects, spiders, crustaceans, gives rigidity and strength to exoskeletons
Dextran	Homo-	(α1→6)Glc, with (α1→3) branches	Wide range	Structural: in bacteria, extracellular adhesive
Peptidoglycan	Hetero-; peptides attached	4)Mur2Ac(β1→4) GlcNAc(β1	Very large	Structural: in bacteria, gives rigidity and strength to cell envelope
Agarose	Hetero-	3)D-Gal(β1→4)3,6-anhydro-L-Gal(α1	1,000	Structural: in algae, cell wall material
Hyaluronan (a glycosaminoglycan)	Hetero-; acidic	4)GlcA(β1→3) GlcNAc(β1	Up to 100,000	Structural: in vertebrates, extracellular matrix of skin and connective tissue; viscosity and lubrication in joints



Fats - depend on the length and saturation of the fatty acid chains they carry



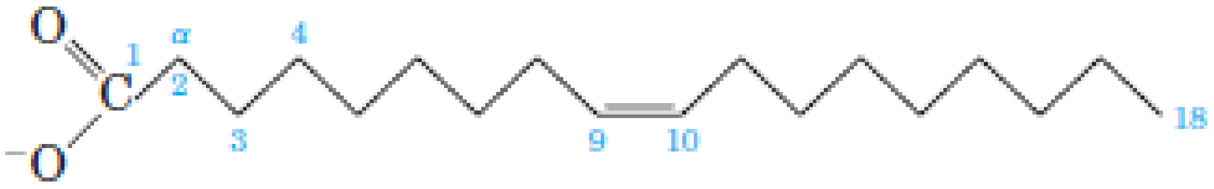


Fatty acids are stored in the cytoplasm of many cells in the form of droplets of *triacylglycerol* molecules

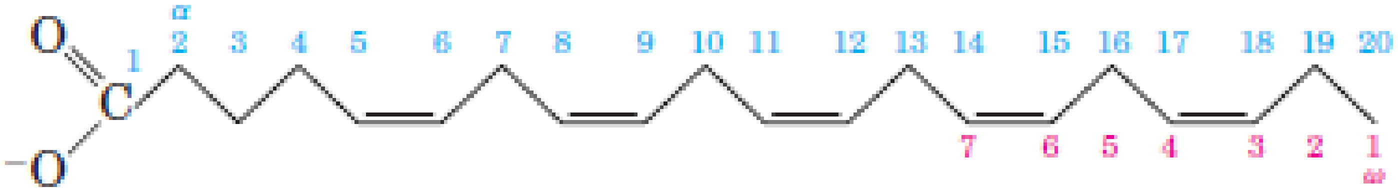
Saturated fats are found in meat and dairy products

Plant oils, such as corn oil, contain unsaturated fatty acids

Fatty acids are carboxylic acids with hydrocarbon chains ranging from 4 to 36 carbons long (C4 to C36).



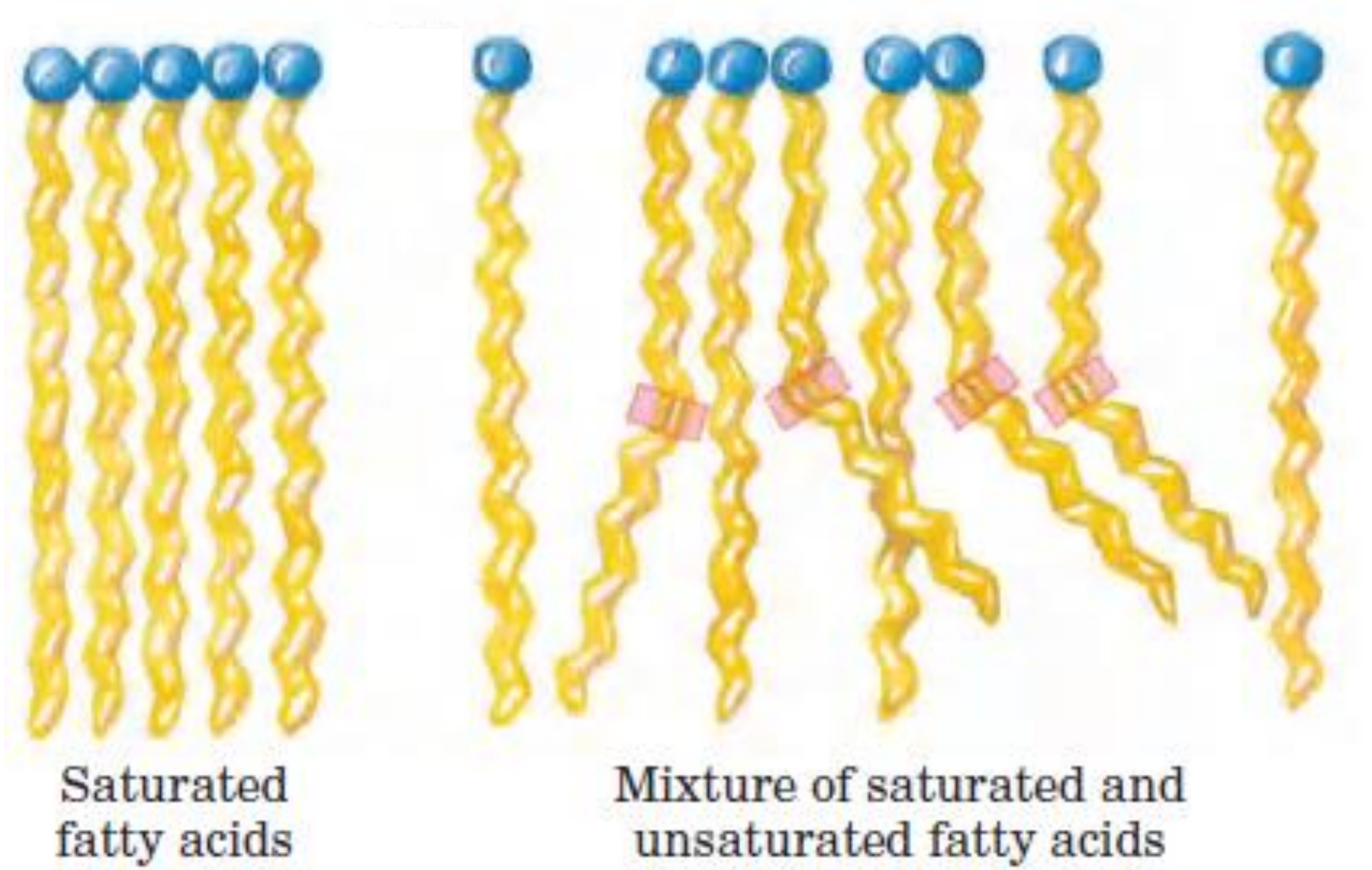
(a) 18:1(Δ^9) *cis*-9-Octadecenoic acid



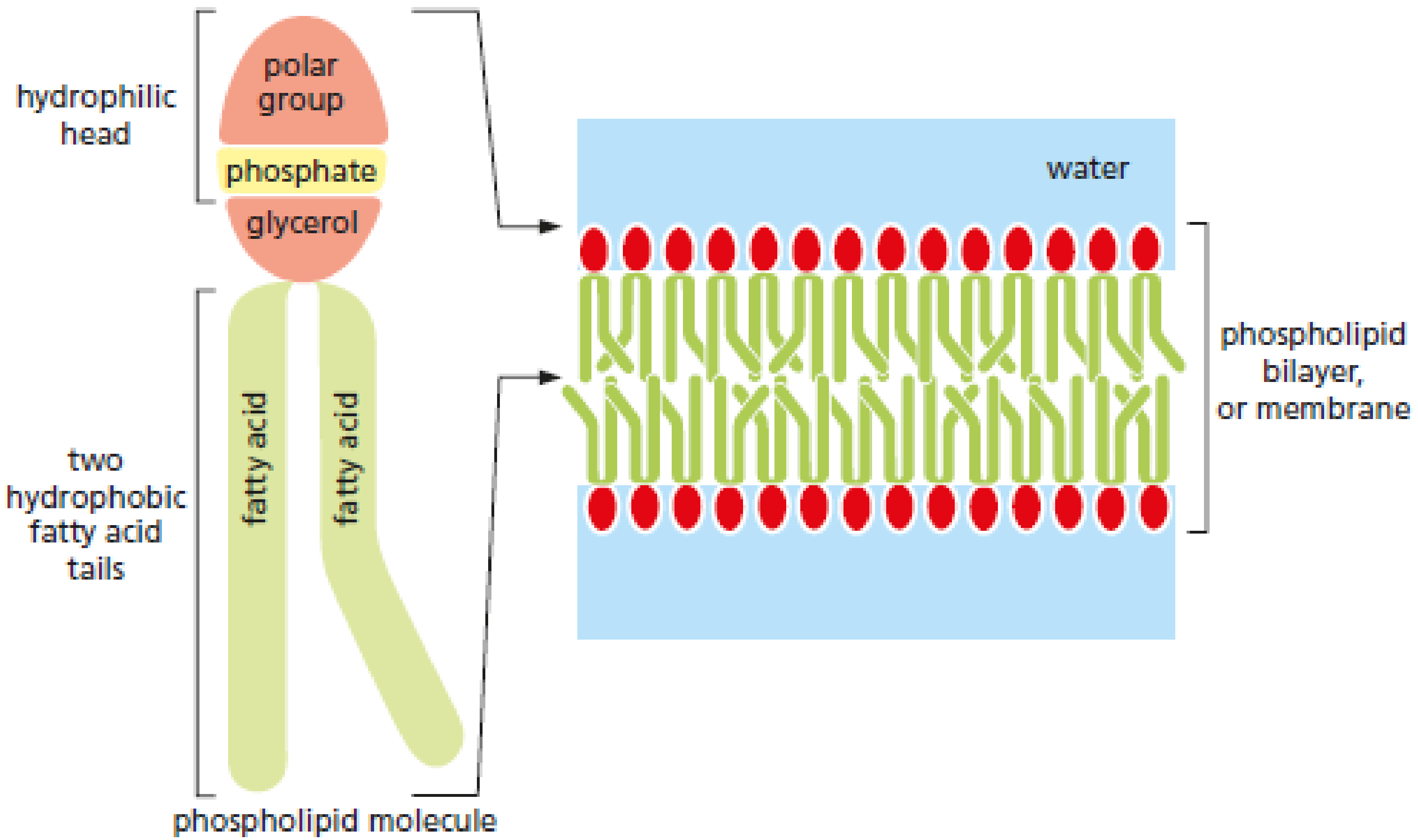
(b) 20:5($\Delta^{5,8,11,14,17}$) Eicosapentaenoic acid (EPA), an omega-3 fatty acid

Carbon skeleton	Structure*	Systematic name[†]	Common name (derivation)
12:0	$CH_3(CH_2)_{10}COOH$	*n*-Dodecanoic acid	Lauric acid (Latin *laurus*, "laurel plant")
14:0	$CH_3(CH_2)_{12}COOH$	*n*-Tetradecanoic acid	Myristic acid (Latin *Myristica*, nutmeg genus)
16:0	$CH_3(CH_2)_{14}COOH$	*n*-Hexadecanoic acid	Palmitic acid (Latin *palma*, "palm tree")
18:0	$CH_3(CH_2)_{16}COOH$	*n*-Octadecanoic acid	Stearic acid (Greek *stear*, "hard fat")
20:0	$CH_3(CH_2)_{18}COOH$	*n*-Eicosanoic acid	Arachidic acid (Latin *Arachis*, legume genus)
24:0	$CH_3(CH_2)_{22}COOH$	*n*-Tetracosanoic acid	Lignoceric acid (Latin *lignum*, "wood" + *cera*, "wax")
16:1(Δ^9)	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COOH$	*cis*-9-Hexadecenoic acid	Palmitoleic acid
18:1(Δ^9)	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$	*cis*-9-Octadecenoic acid	Oleic acid (Latin *oleum*, "oil")
18:2($\Delta^{9,12}$)	$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$	*cis*-,*cis*-9,12-Octadecadienoic acid	Linoleic acid (Greek *linon*, "flax")
18:3($\Delta^{9,12,15}$)	$CH_3CH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$	*cis*-,*cis*-,*cis*-9,12,15-Octadecatrienoic acid	α-Linolenic acid
20:4($\Delta^{5,8,11,14}$)	$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CH(CH_2)_3COOH$	*cis*-,*cis*-,*cis*-,*cis*-5, 8,11, 14- Icosatetraenoic acid	Arachidonic acid

The packing of fatty acids into stable aggregates



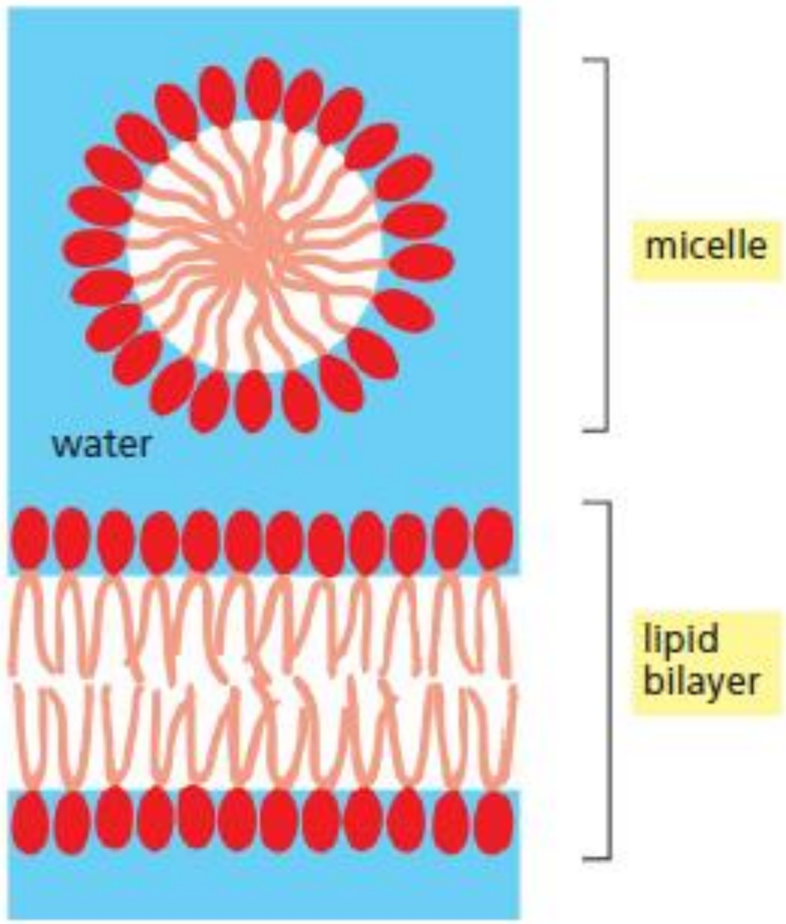
Phospholipid - Lipid Bilayer

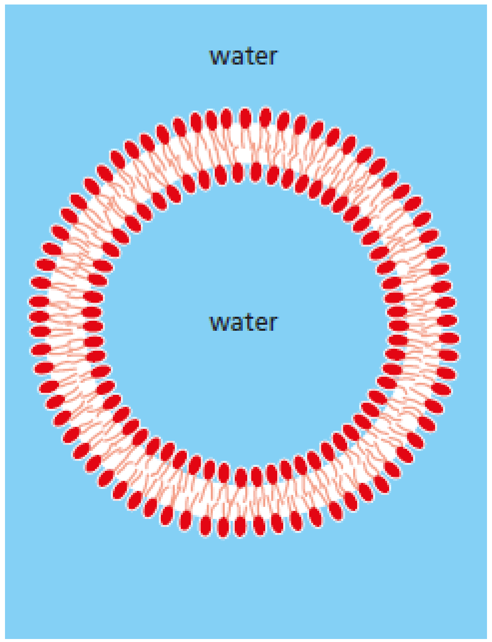


Glycerol is joined to two fatty acid chains.

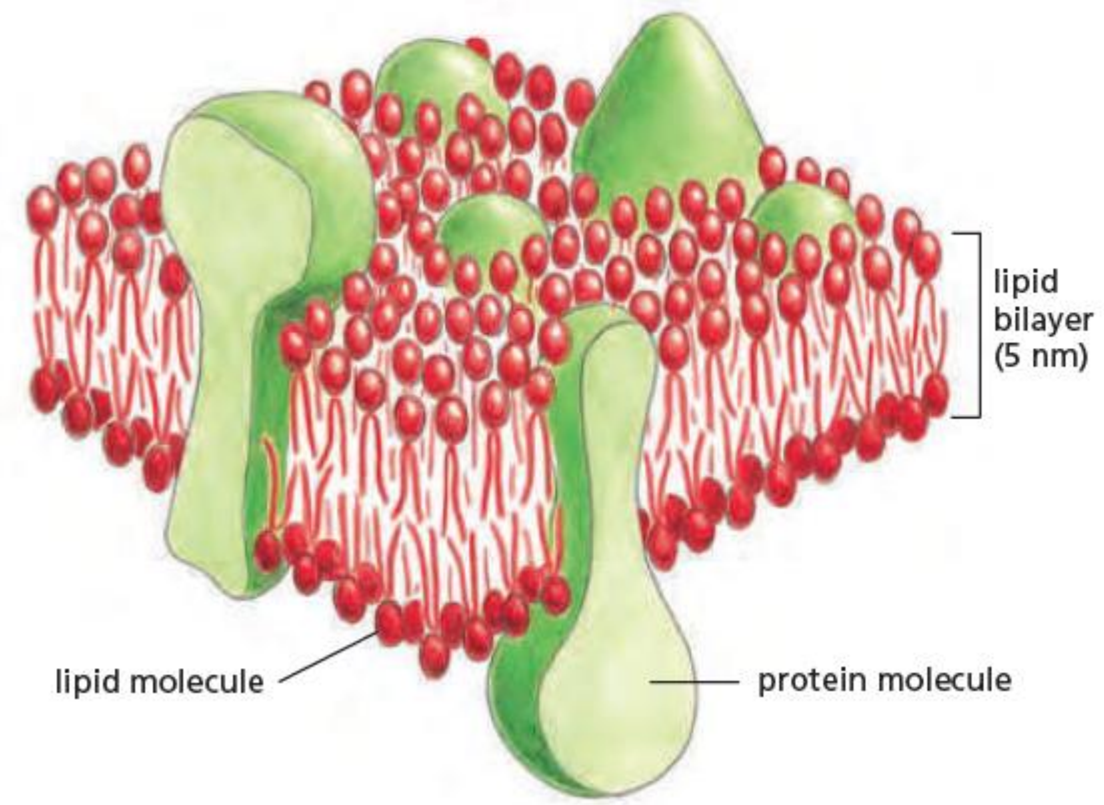
The remaining –OH group on the glycerol is linked to a hydrophilic phosphate group, which in turn is attached to a small hydrophilic compound such as choline.

Phospholipids Spontaneously Form Bilayers

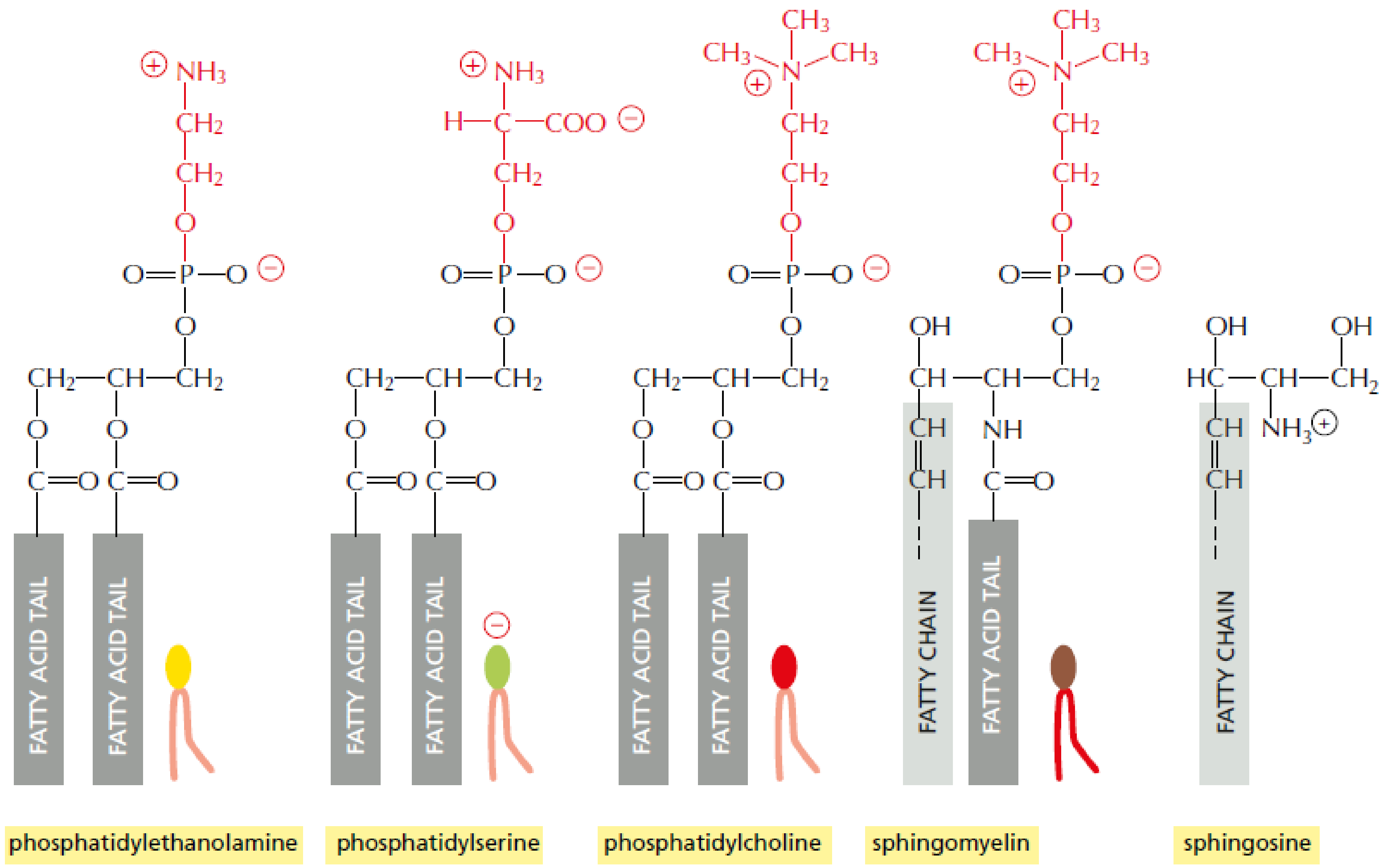




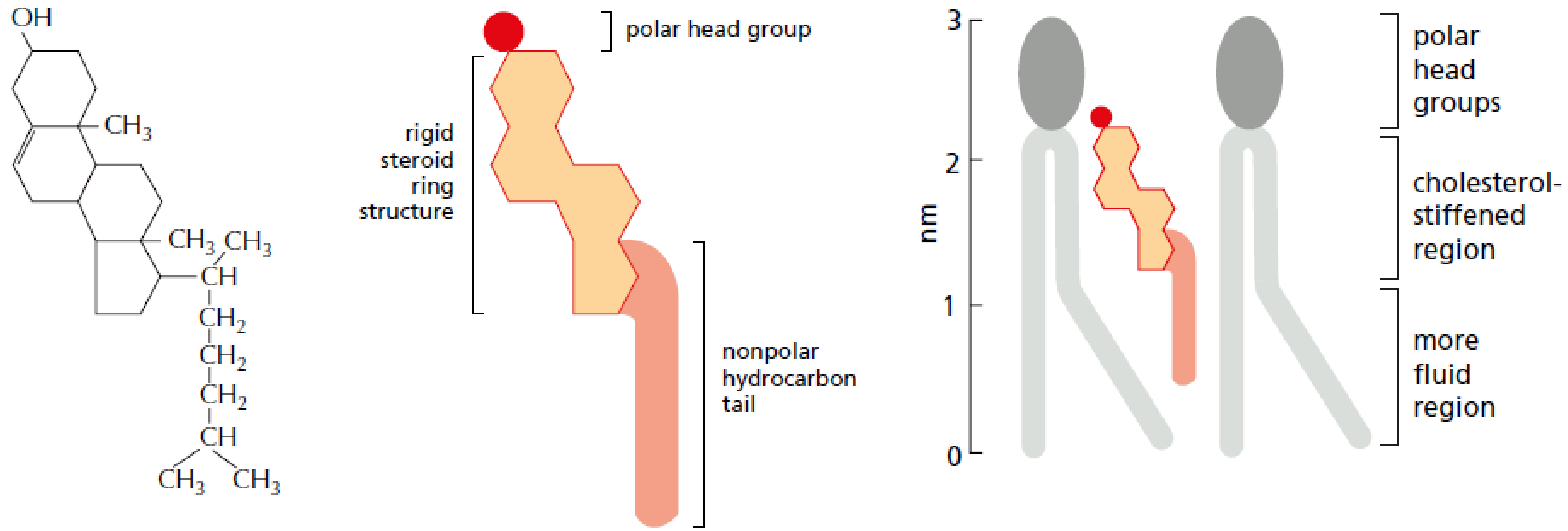
A three-dimensional schematic view of a cell membrane



Four major phospholipids in mammalian plasma membranes



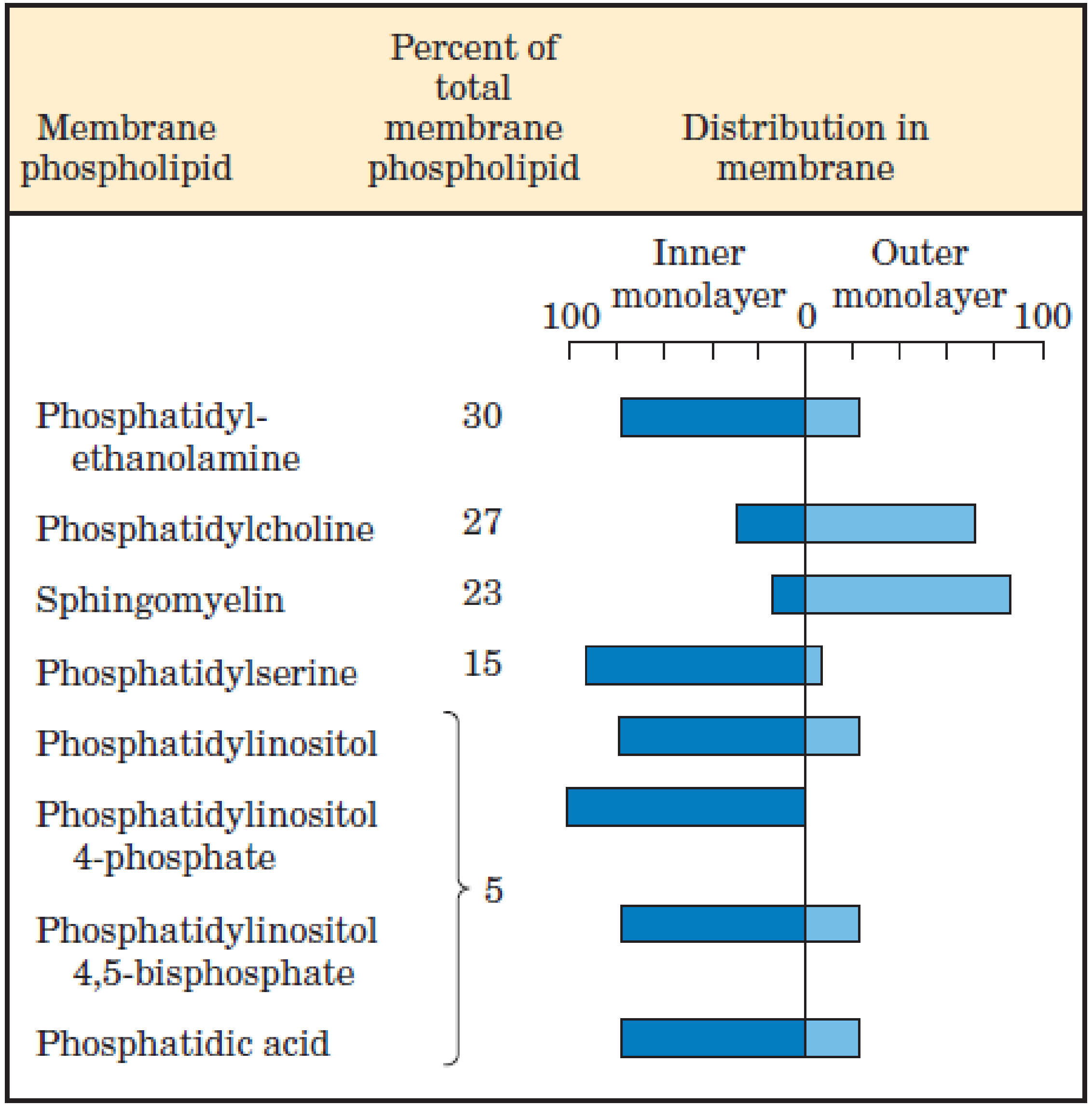
The structure of cholesterol

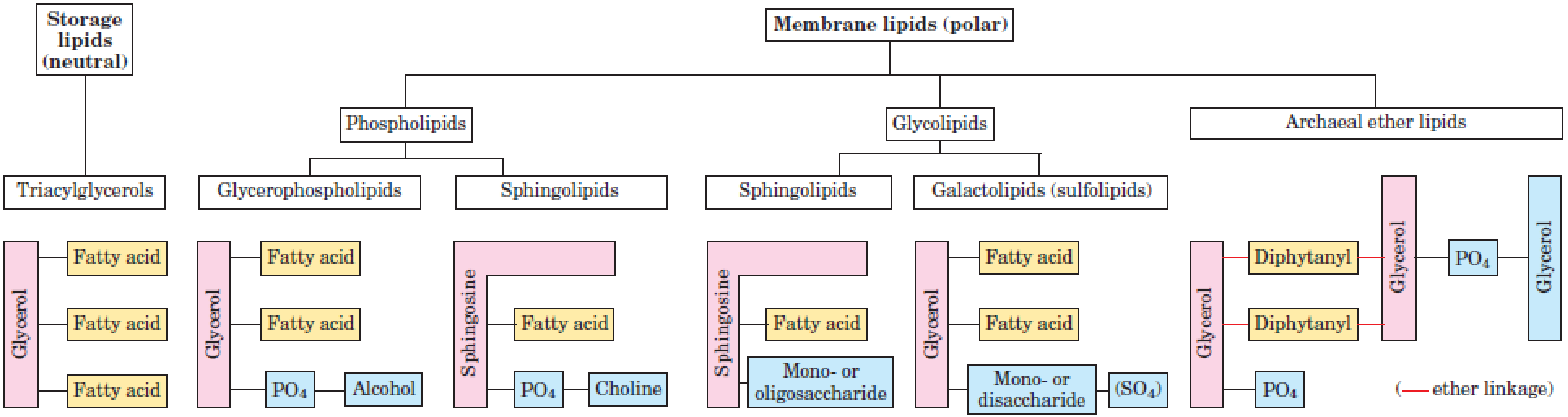


Lipid composition of the plasma membrane and organelle membranes of a hepatocyte

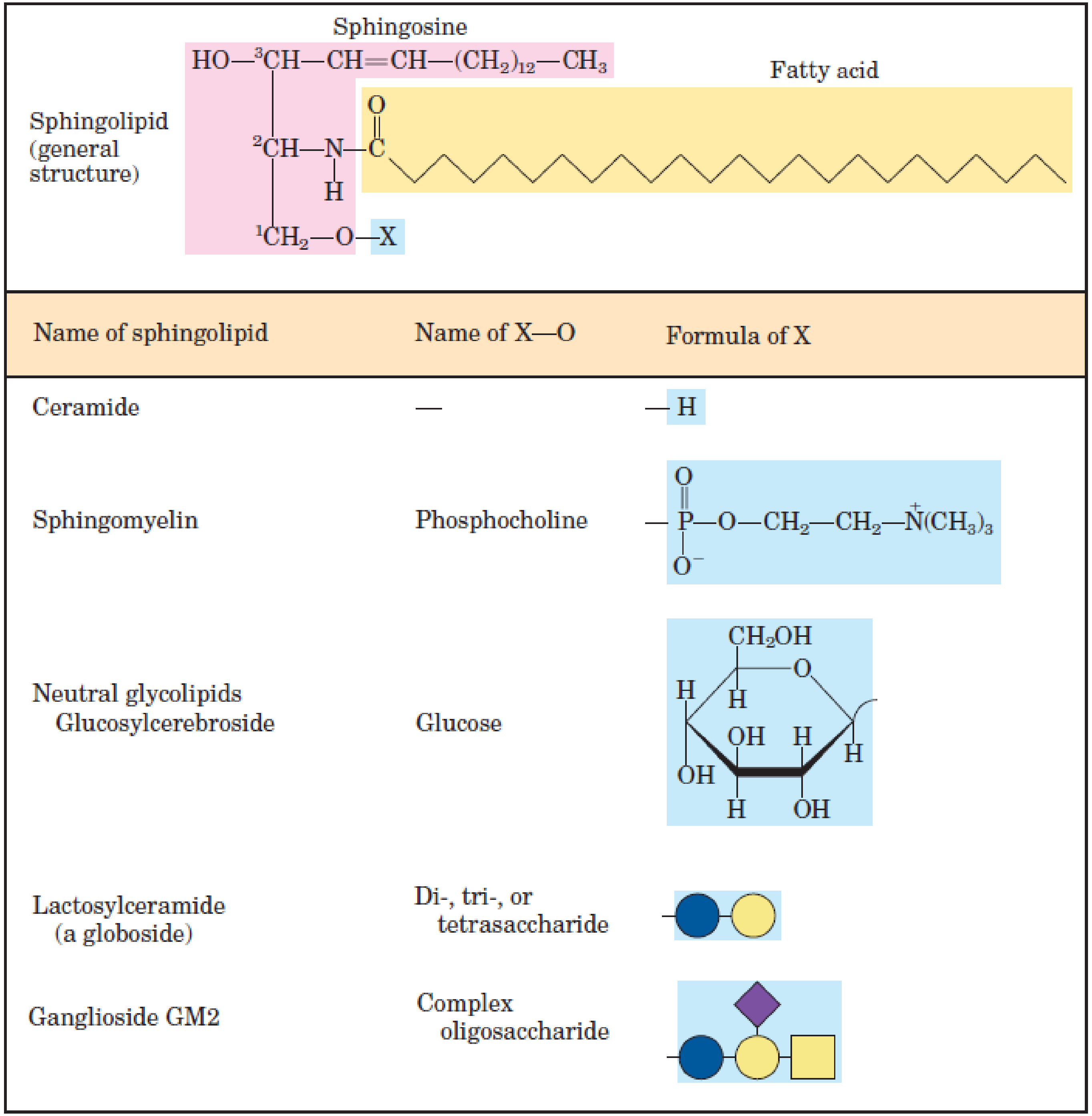
Lipid	Percentage of total lipid by weight					
	Liver cell plasma membrane	Red blood cell plasma membrane	Myelin	Mitochondrion (inner and outer membranes)	Endoplasmic reticulum	*E. coli* bacterium
Cholesterol	17	23	22	3	6	0
Phosphatidylethanolamine	7	18	15	28	17	70
Phosphatidylserine	4	7	9	2	5	trace
Phosphatidylcholine	24	17	10	44	40	0
Sphingomyelin	19	18	8	0	5	0
Glycolipids	7	3	28	trace	trace	0
Others	22	14	8	23	27	30

Distribution of phospholipids between the inner and outer monolayers



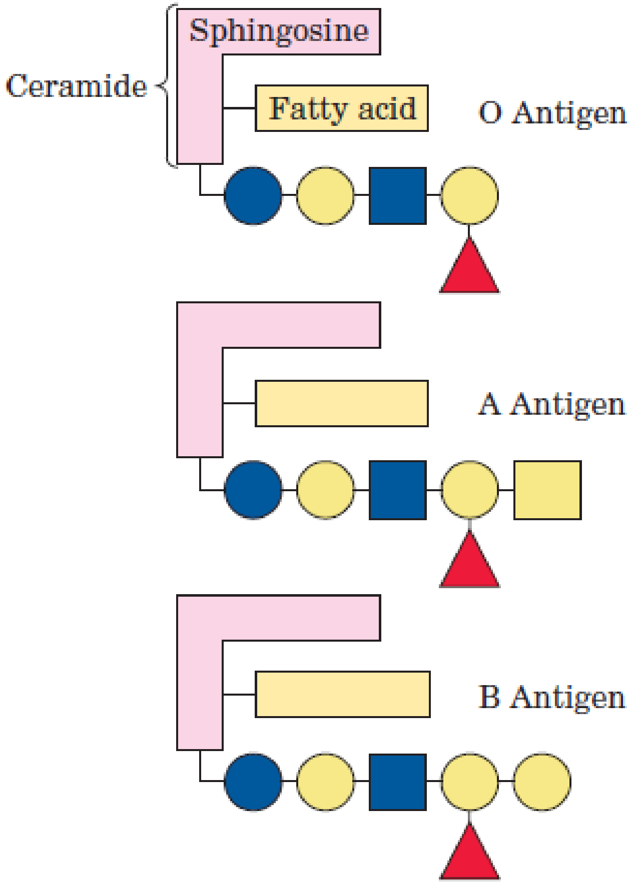


Storage lipids (neutral)
Membrane lipids (polar)
Phospholipids
Glycolipids
Archaeal ether lipids
Triacylglycerols
Glycerophospholipids
Sphingolipids
Sphingolipids
Galactolipids (sulfolipids)
Glycerol
Fatty acid
Fatty acid
Fatty acid
Glycerol
Fatty acid
Fatty acid
PO4
Alcohol
Sphingosine
Fatty acid
PO4
Choline
Sphingosine
Fatty acid
Mono- or oligosaccharide
Glycerol
Fatty acid
Fatty acid
Mono- or disaccharide
(SO4)
Glycerol
Diphytanyl
Diphytanyl
Glycerol
PO4
Glycerol
PO4
(ether linkage)



Sphingolipids at Cell Surfaces Are Sites of Biological Recognition

- The carbohydrate moieties of certain sphingolipids define the human blood groups and therefore determine the type of blood that individuals can safely receive in blood transfusions



- Lipids as Signals,Cofactors,and Pigments

Proteins

ENZYMES

function: Catalyze covalent bond breakage or formation.

examples: Living cells contain thousands of different enzymes, each of which catalyzes (speeds up) one particular reaction. Examples include: *tryptophan synthetase*—makes the amino acid tryptophan; *pepsin*—degrades dietary proteins in the stomach; *ribulose bisphosphate carboxylase*—helps convert carbon dioxide into sugars in plants; *DNA polymerase*—copies DNA; *protein kinase*—adds a phosphate group to a protein molecule.

STRUCTURAL PROTEINS

function: Provide mechanical support to cells and tissues.

examples: Outside cells, *collagen* and *elastin* are common constituents of extracellular matrix and form fibers in tendons and ligaments. Inside cells, *tubulin* forms long, stiff microtubules, and *actin* forms filaments that underlie and support the plasma membrane; *keratin* forms fibers that reinforce epithelial cells and is the major protein in hair and horn.

TRANSPORT PROTEINS

function: Carry small molecules or ions.

examples: In the bloodstream, *serum albumin* carries lipids, *hemoglobin* carries oxygen, and *transferrin* carries iron. Many proteins embedded in cell membranes transport ions or small molecules across the membrane. For example, the bacterial protein *bacteriorhodopsin* is a light-activated proton pump that transports H^+ ions out of the cell; *glucose carriers* shuttle glucose into and out of cells; and a *Ca^{2+} pump* clears Ca^{2+} from a muscle cell's cytosol after the ions have triggered a contraction.

MOTOR PROTEINS

function: Generate movement in cells and tissues.

examples: *Myosin* in skeletal muscle cells provides the motive force for humans to move; *kinesin* interacts with microtubules to move organelles around the cell; *dynein* enables eukaryotic cilia and flagella to beat.

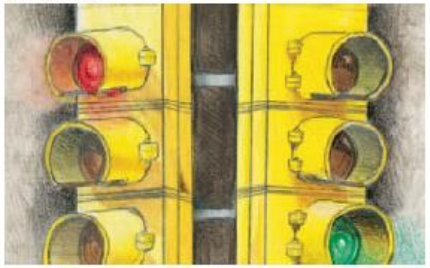
STORAGE PROTEINS

function: Store amino acids or ions.

examples: Iron is stored in the liver by binding to the small protein *ferritin*; *ovalbumin* in egg white is used as a source of amino acids for the developing bird embryo; *casein* in milk is a source of amino acids for baby mammals.

SIGNAL PROTEINS

function: Carry extracellular signals from cell to cell.



examples: Many of the hormones and growth factors that coordinate physiological functions in animals are proteins; *insulin*, for example, is a small protein that controls glucose levels in the blood; *netrin* attracts growing nerve cell axons to specific locations in the developing spinal cord; *nerve growth factor* (*NGF*) stimulates some types of nerve cells to grow axons; *epidermal growth factor* (*EGF*) stimulates the growth and division of epithelial cells.

RECEPTOR PROTEINS

function: Detect signals and transmit them to the cell's response machinery.

examples: *Rhodopsin* in the retina detects light; the *acetylcholine receptor* in the membrane of a muscle cell is activated by acetylcholine released from a nerve ending; the *insulin receptor* allows a cell to respond to the hormone insulin by taking up glucose; the *adrenergic receptor* on heart muscle increases the rate of the heartbeat when it binds to adrenaline.

GENE REGULATORY PROTEINS

function: Bind to DNA to switch genes on or off.

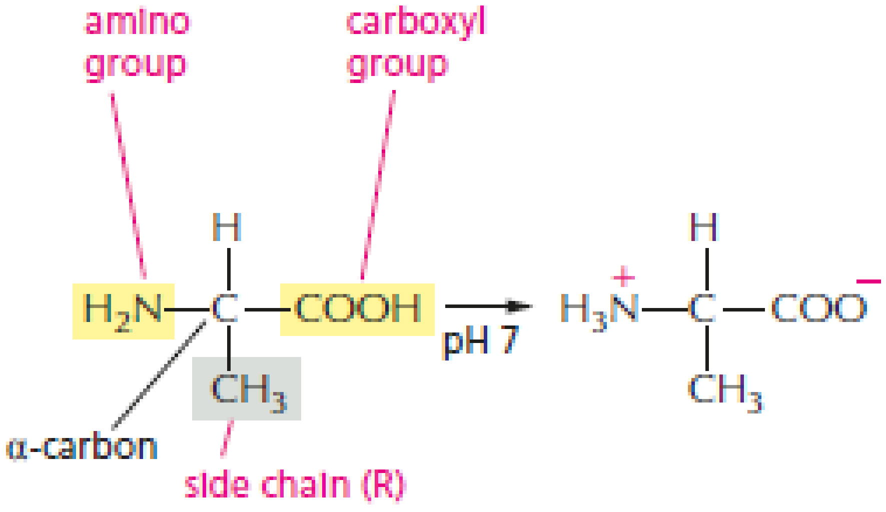
examples: The *lactose repressor* in bacteria silences the genes for the enzymes that degrade the sugar lactose; many different *homeodomain proteins* act as genetic switches to control development in multicellular organisms, including humans.

SPECIAL-PURPOSE PROTEINS

function: Highly variable.

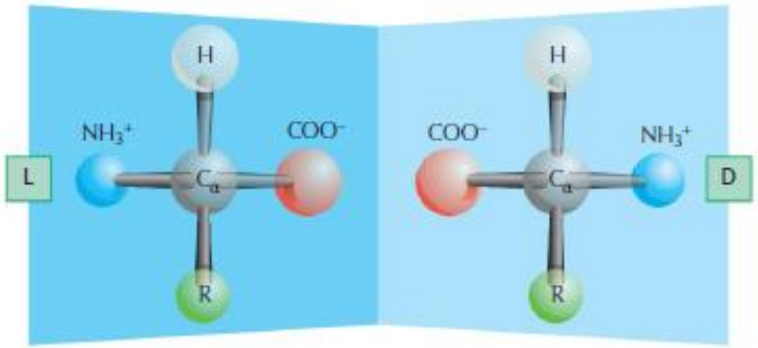
examples: Organisms make many proteins with highly specialized properties. These molecules illustrate the amazing range of functions that proteins can perform. The *antifreeze proteins* of Arctic and Antarctic fishes protect their blood against freezing; *green fluorescent protein* from jellyfish emits a green light; *monellin*, a protein found in an African plant, has an intensely sweet taste; mussels and other marine organisms secrete *glue proteins* that attach them firmly to rocks, even when immersed in seawater.

Amino acids



nonionized form ionized form

OPTICAL ISOMERS The α-carbon atom is asymmetric, which allows for two mirror images (or stereo-) isomers, L and D.



Proteins consist exclusively of L-amino acids.

FAMILIES OF AMINO ACIDS

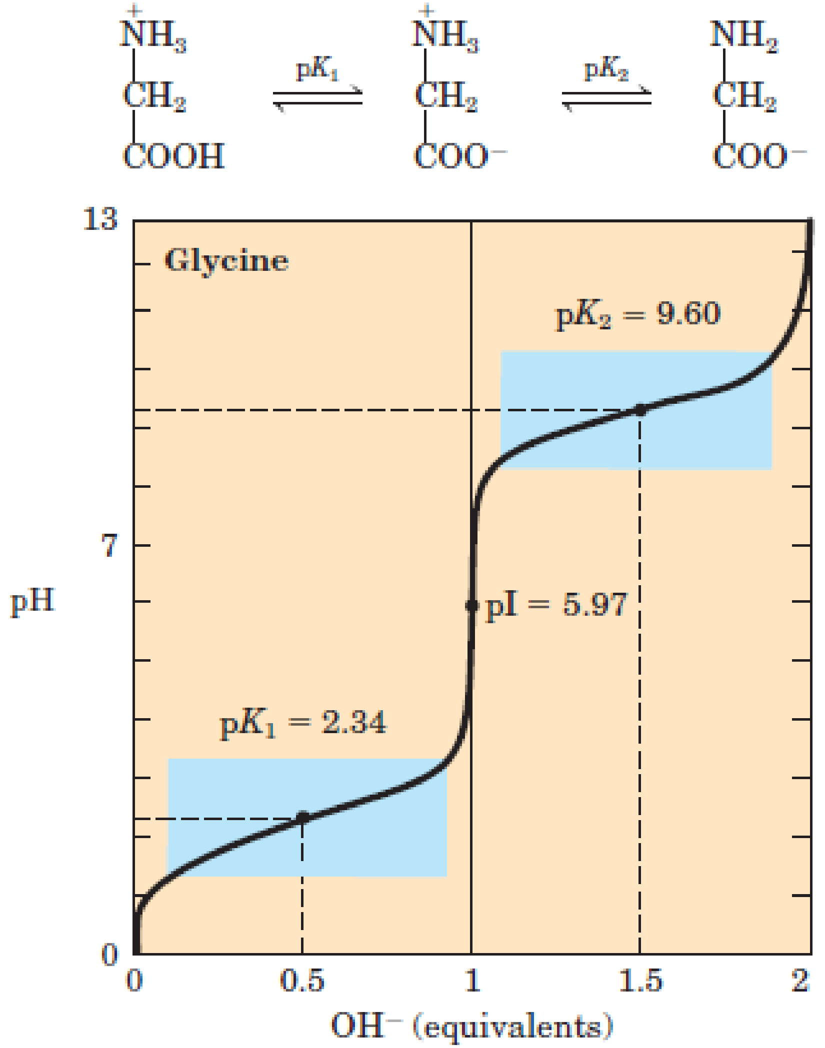
The common amino acids are grouped according to whether their side chains are

- acidic
- basic
- uncharged polar
- nonpolar

These 20 amino acids are given both three-letter and one-letter abbreviations.

Thus: alanine = Ala = A

Amino Acids Have Characteristic Titration Curves



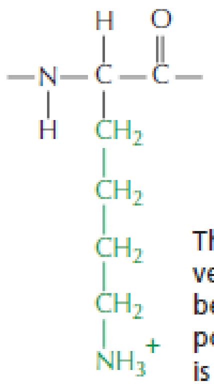
Amino Acids Can Act as Acids and Bases

Amino acids

BASIC SIDE CHAINS

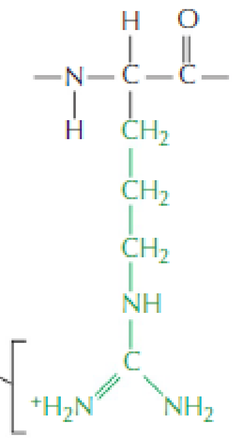
lysine

(Lys, or K)



arginine

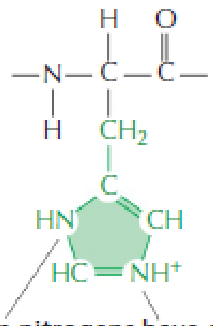
(Arg, or R)



This group is very basic because its positive charge is stabilized by resonance (see Panel 2–1).

histidine

(His, or H)

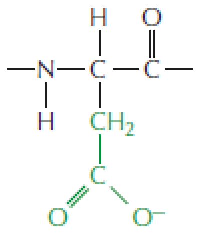


These nitrogens have a relatively weak affinity for an H^+ and are only partly positive at neutral pH.

ACIDIC SIDE CHAINS

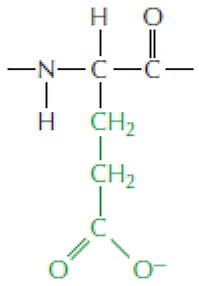
aspartic acid

(Asp, or D)

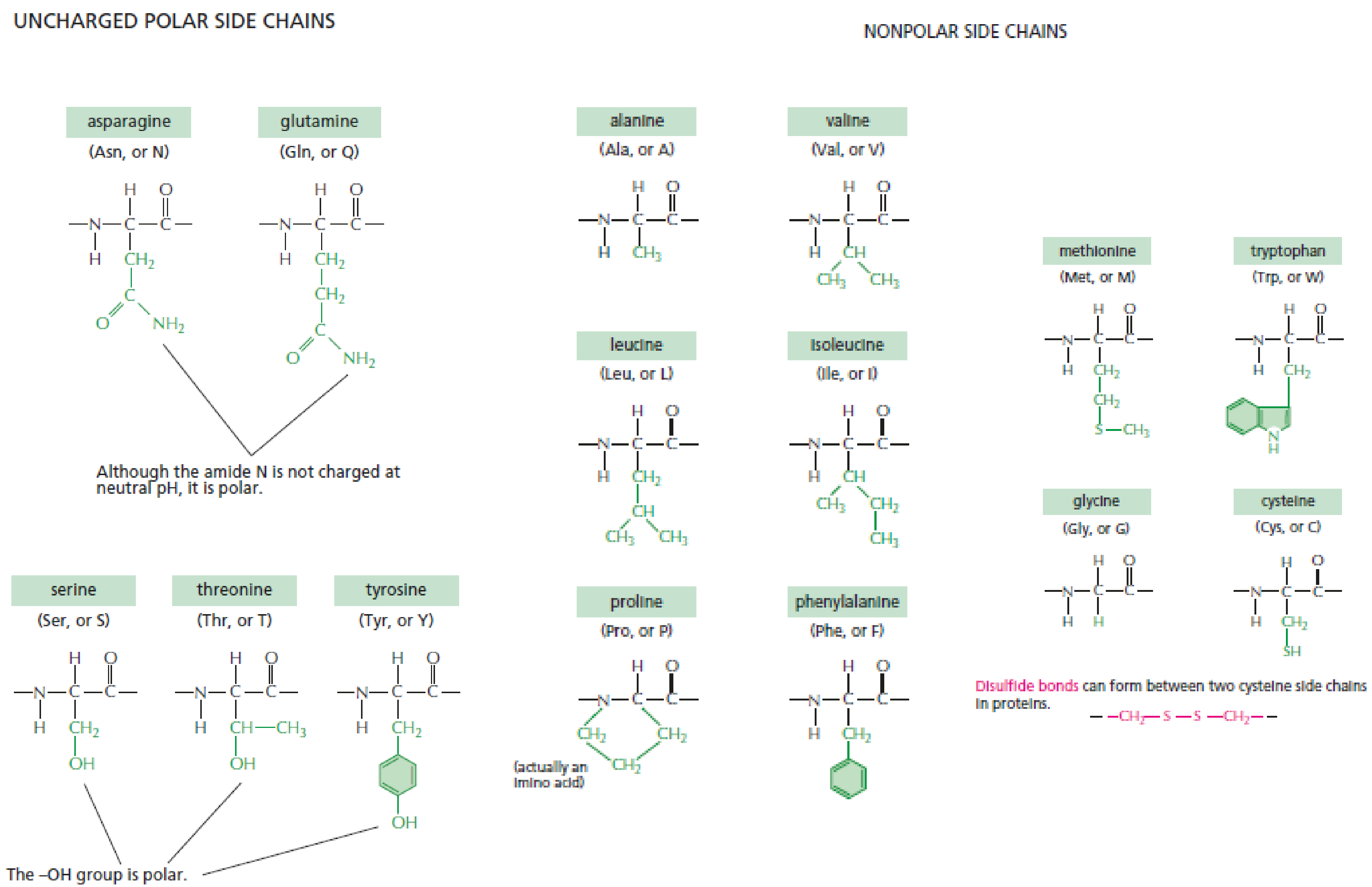


glutamic acid

(Glu, or E)



Amino acids



UNCHARGED POLAR SIDE CHAINS
asparagine
(Asn, or N)
glutamine
(Gln, or Q)
Although the amide N is not charged at neutral pH, it is polar.
serine
(Ser, or S)
threonine
(Thr, or T)
tyrosine
(Tyr, or Y)
The –OH group is polar.
NONPOLAR SIDE CHAINS
alanine
(Ala, or A)
valine
(Val, or V)
leucine
(Leu, or L)
isoleucine
(Ile, or I)
proline
(Pro, or P)
(actually an imino acid)
phenylalanine
(Phe, or F)
methionine
(Met, or M)
tryptophan
(Trp, or W)
glycine
(Gly, or G)
cysteine
(Cys, or C)
Disulfide bonds can form between two cysteine side chains in proteins.

Amino acids

AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negative
Glutamic acid	Glu	E	negative
Arginine	Arg	R	positive
Lysine	Lys	K	positive
Histidine	His	H	positive
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

POLAR AMINO ACIDS

AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

NONPOLAR AMINO ACIDS

Six amino acids (CHIMSV), the first letter of the amino acid name is unique

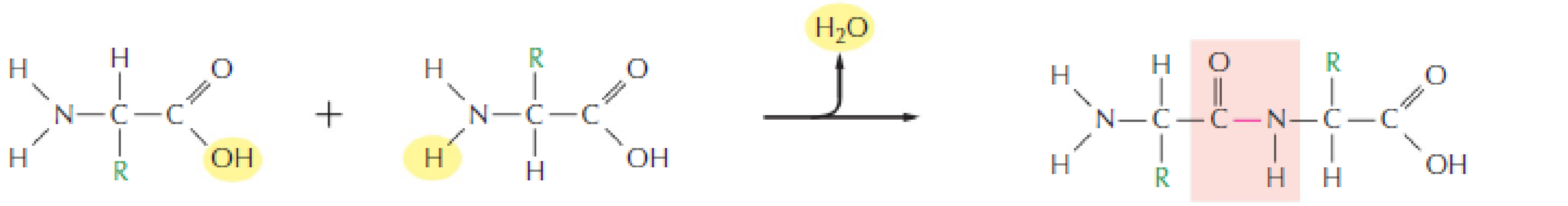
Phonetically suggestive - RFYW: aRginine, Fenylalanine, tYrosine, tWiptophan.

Proteins

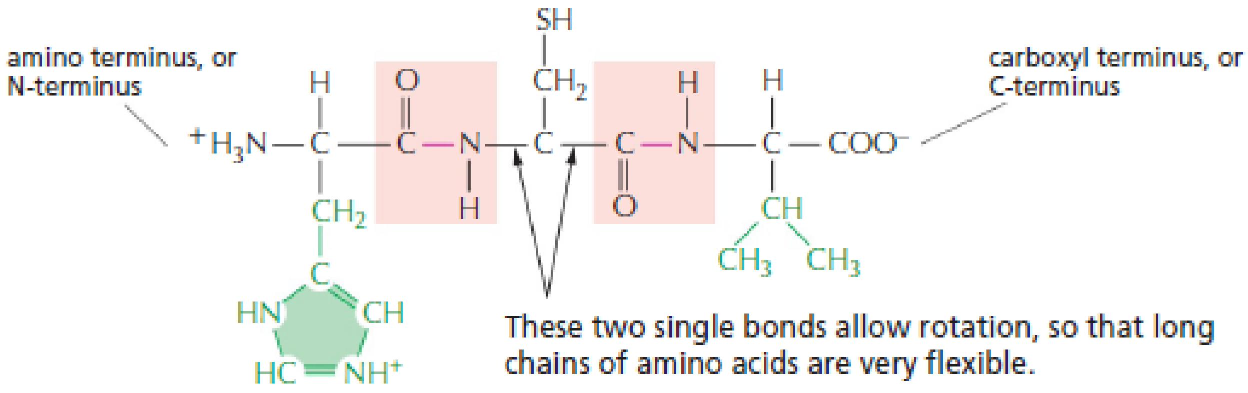
PEPTIDE BONDS

In proteins, amino acids are commonly joined together by an amide linkage, called a peptide bond.

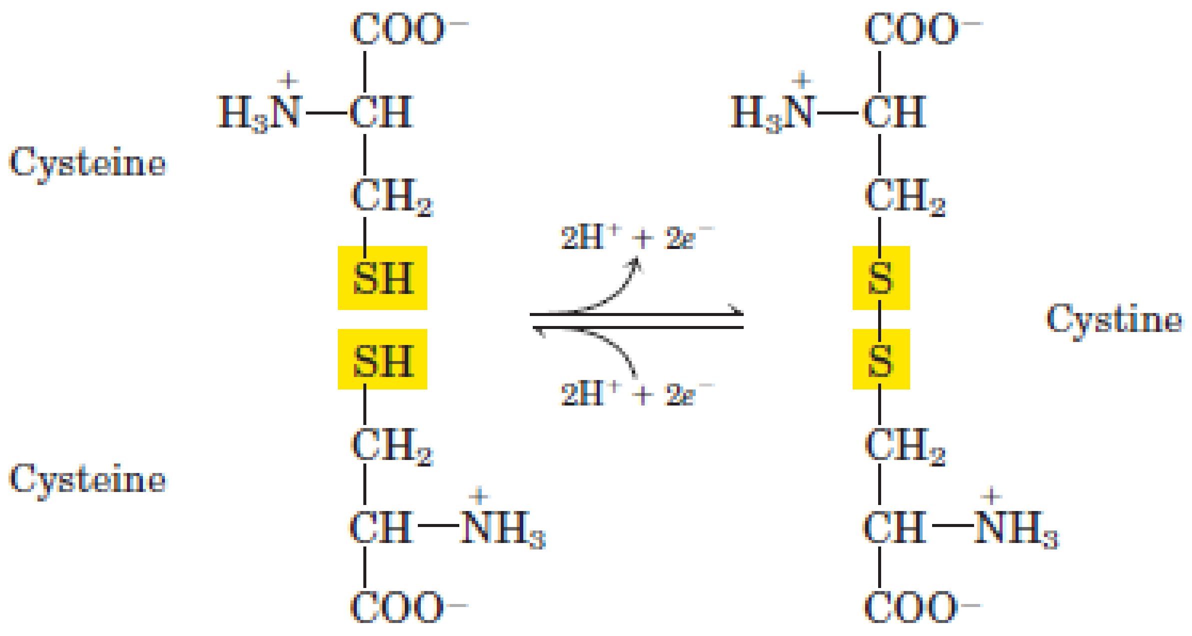
The four atoms in each peptide bond *(red box)* form a rigid planar unit. There is no rotation around the C–N bond.



Proteins are long polymers of amino acids linked by peptide bonds, and they are always written with the N-terminus toward the left. Peptides are shorter, usually fewer than 50 amino acids long. The sequence of this tripeptide is histidine-cysteine-valine.



Reversible formation of a disulfide bond by the oxidation of two molecules of cysteine



Molecular Data on Some Proteins

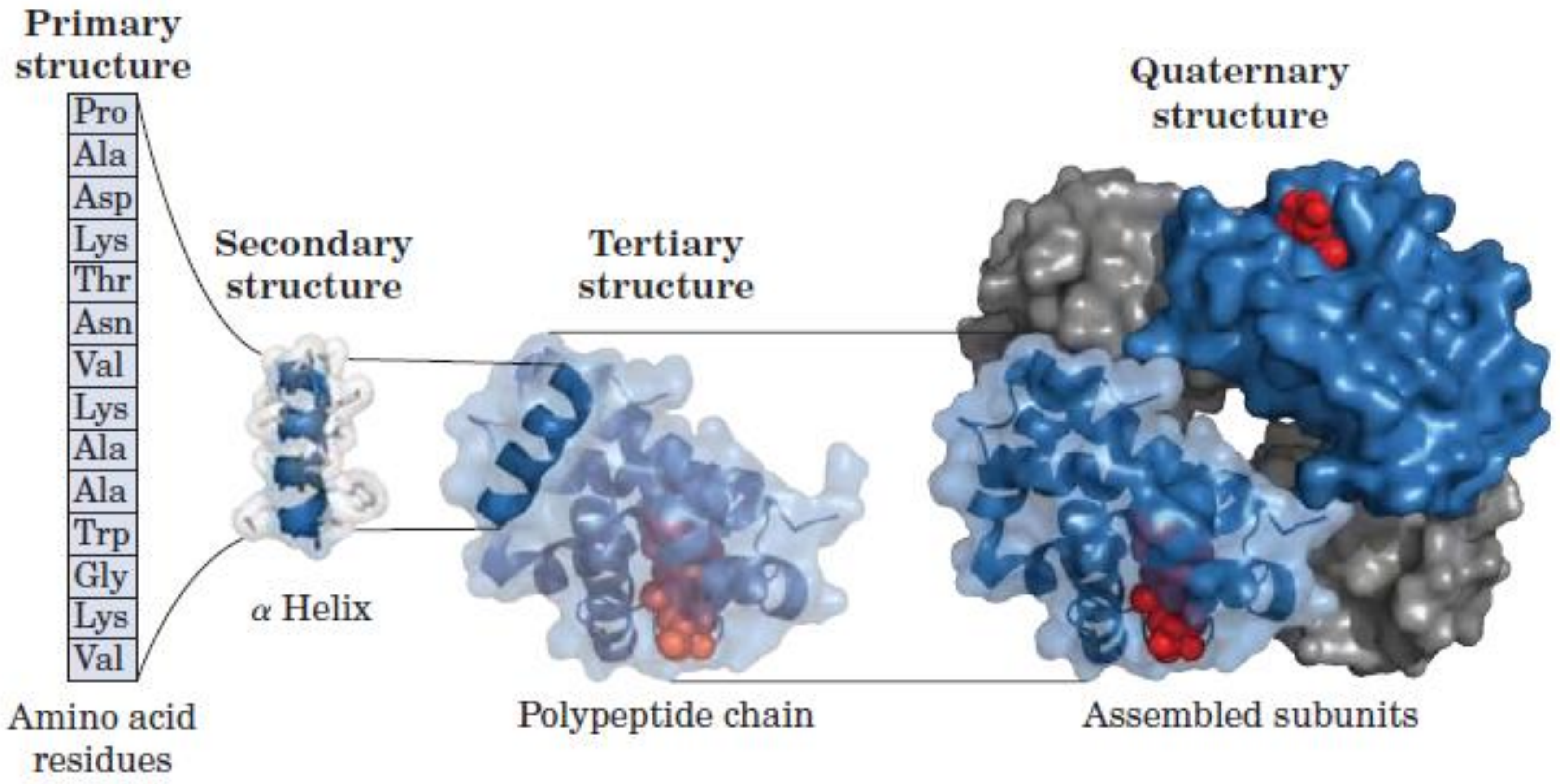
	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	12,400	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (chicken egg white)	14,300	129	1
Myoglobin (equine heart)	16,700	153	1
Chymotrypsin (bovine pancreas)	25,200	241	3
Chymotrypsinogen (bovine)	25,700	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	66,000	609	1
Hexokinase (yeast)	107,900	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

Some Proteins Contain Chemical Groups Other Than Amino Acids

Class	Prosthetic group	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Conjugated Proteins

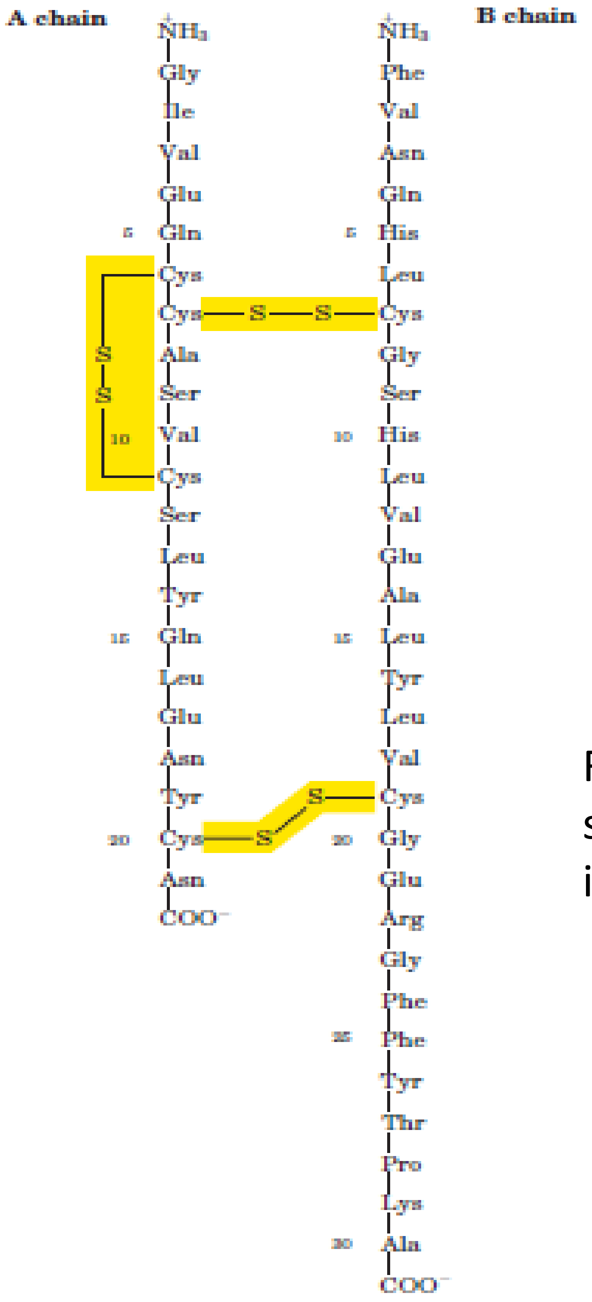
The Structure of Proteins



Aligning protein sequences

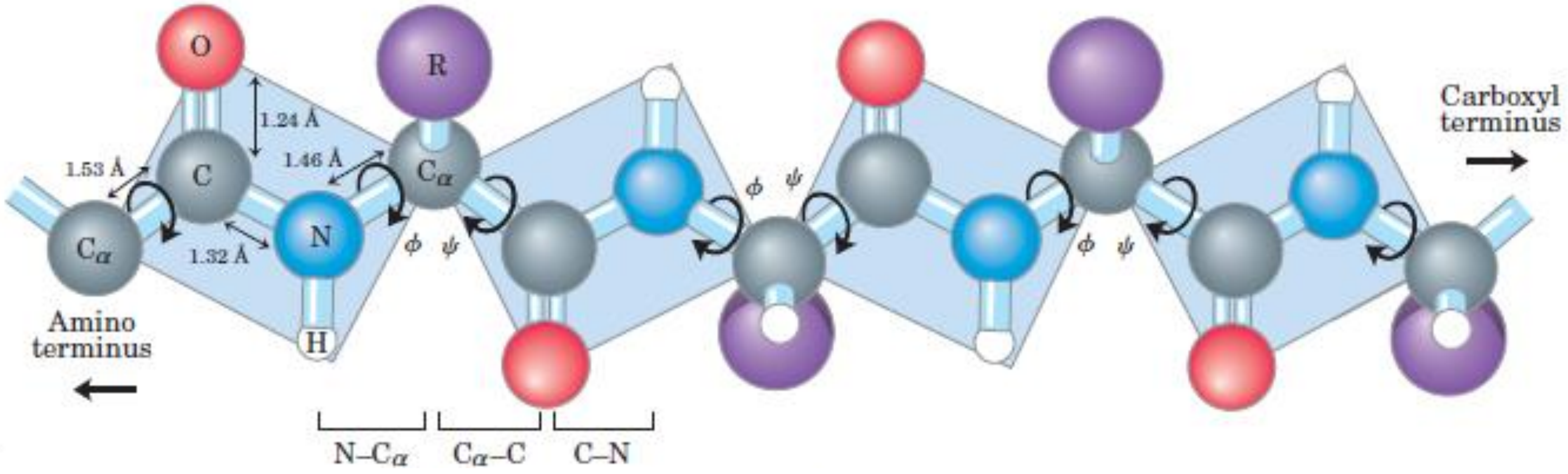
```
E. coli      TGNRTIAVYDLGGGTFDISIIEIDEVDGEKTFEVLATNGDTHLGGEDFDSRLIHYL
B. subtilis  DEDQTILLYDLGGGTFDVSILELGDG----TFEVRSTAGDNRLGGDDFDQVIIDHL
                                       Gap
```

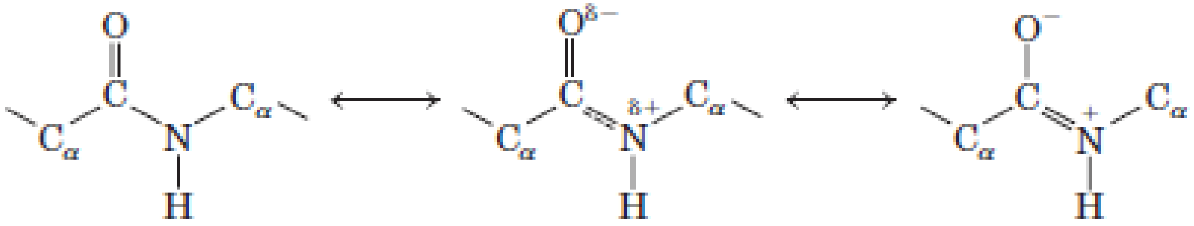
Amino acid sequence of bovine insulin



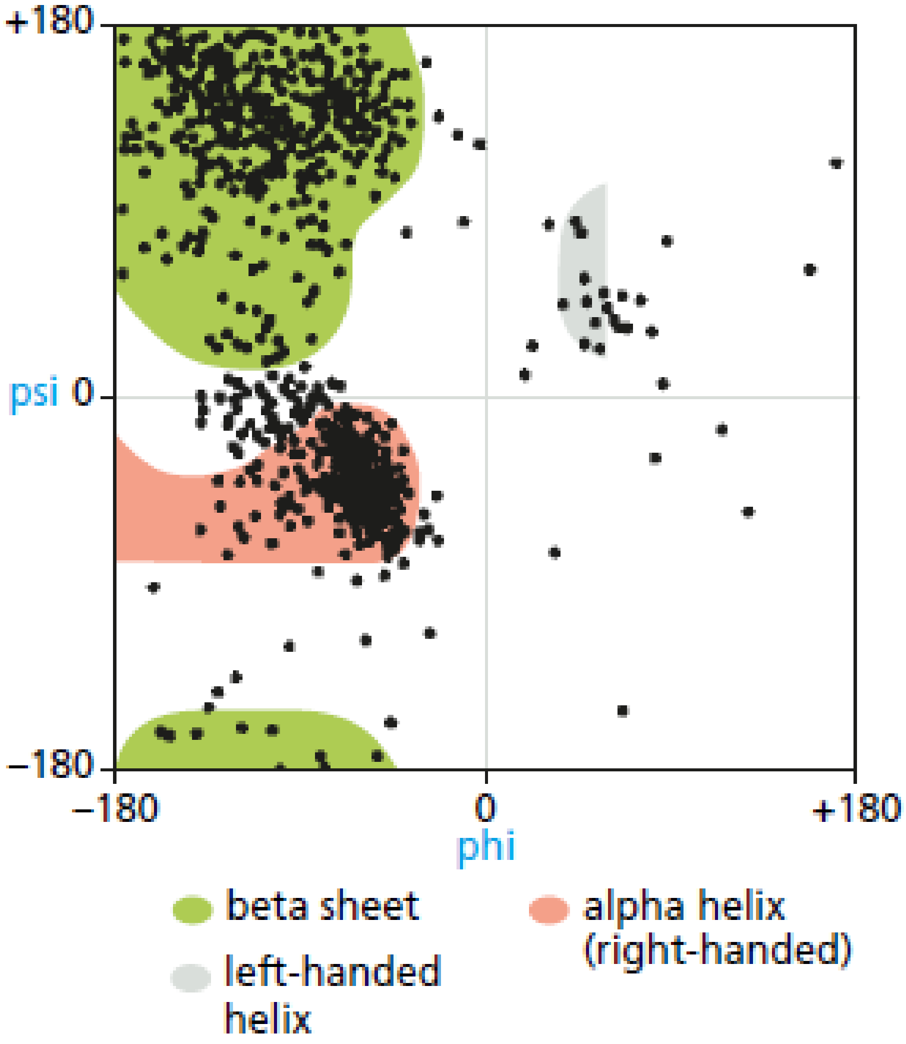
Frederick Sanger worked out the sequence of amino acid residues in the polypeptide chains of the hormone insulin

The Peptide Bond Is Rigid and Planar

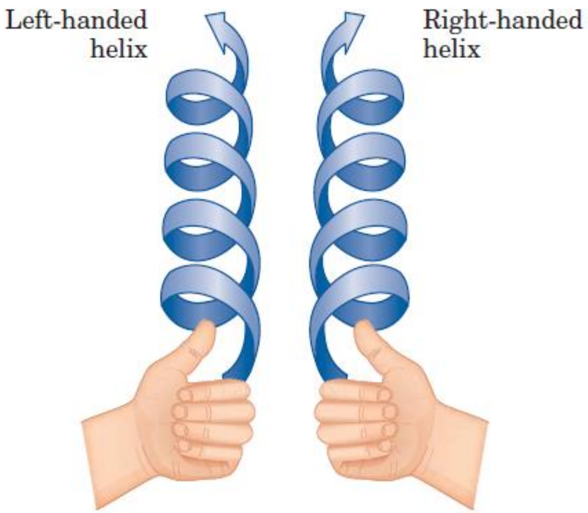




Steric limitations on the bond angles in a polypeptide chain

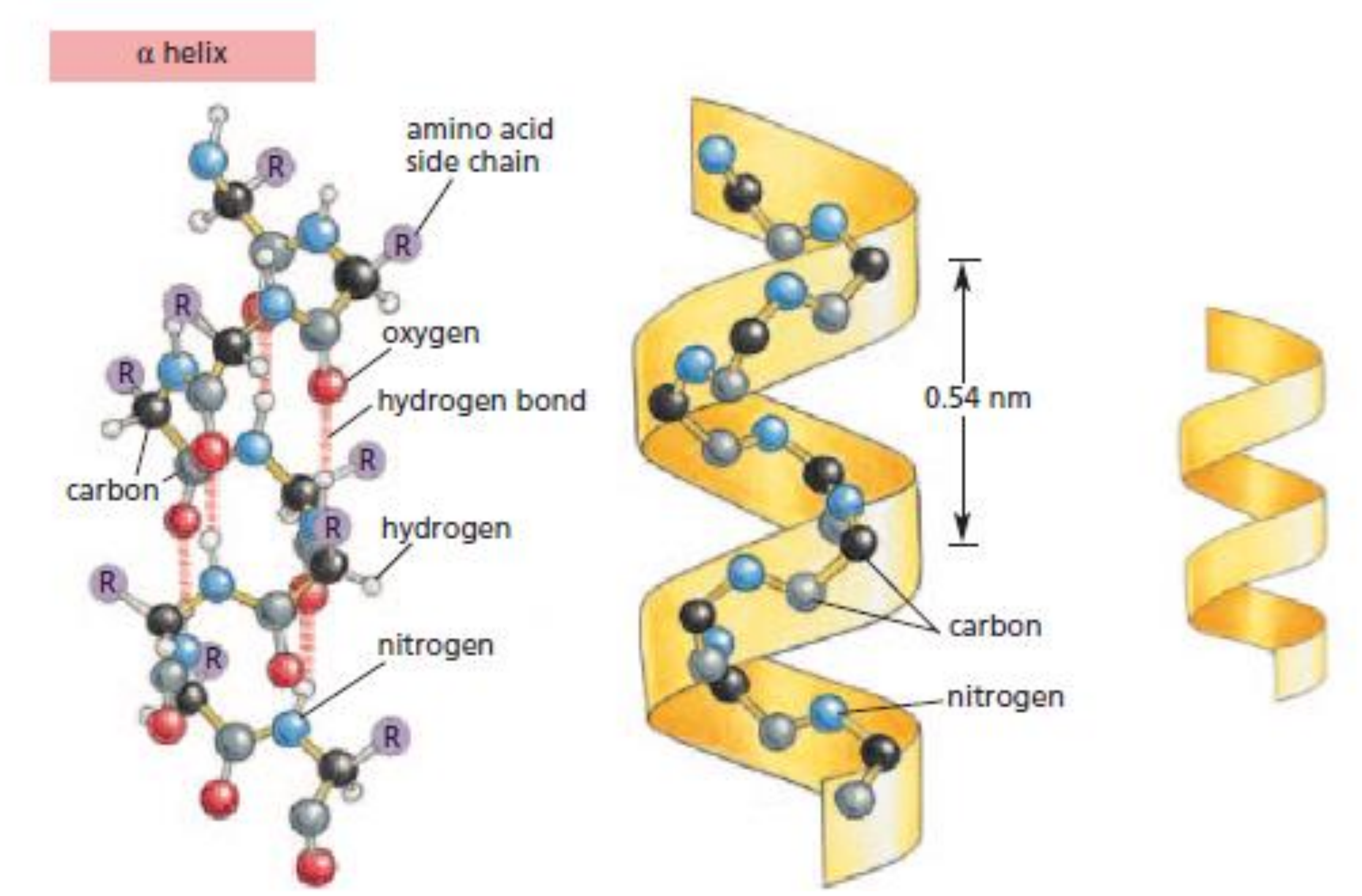


Ramachandran plot



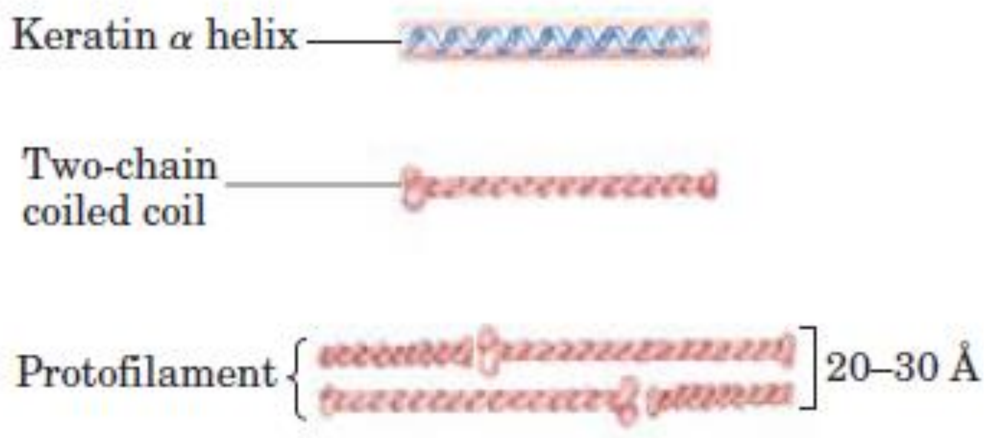
Each dot represents an observed pair of angles in a protein.

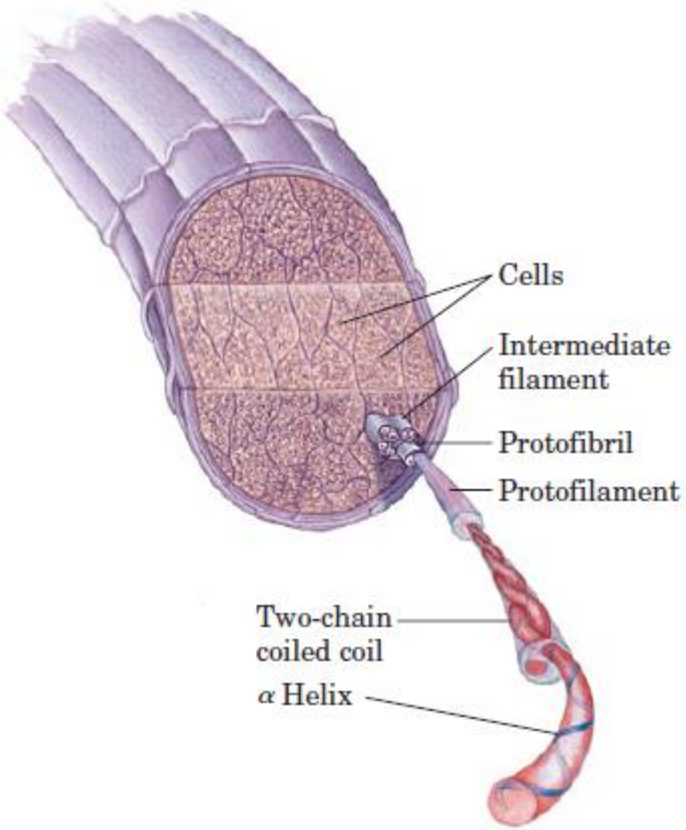
The regular conformation of the polypeptide backbone in the α helix



The N–H of every peptide bond is hydrogen-bonded to the C=O of a neighboring peptide bond located four peptide bonds away in the same chain

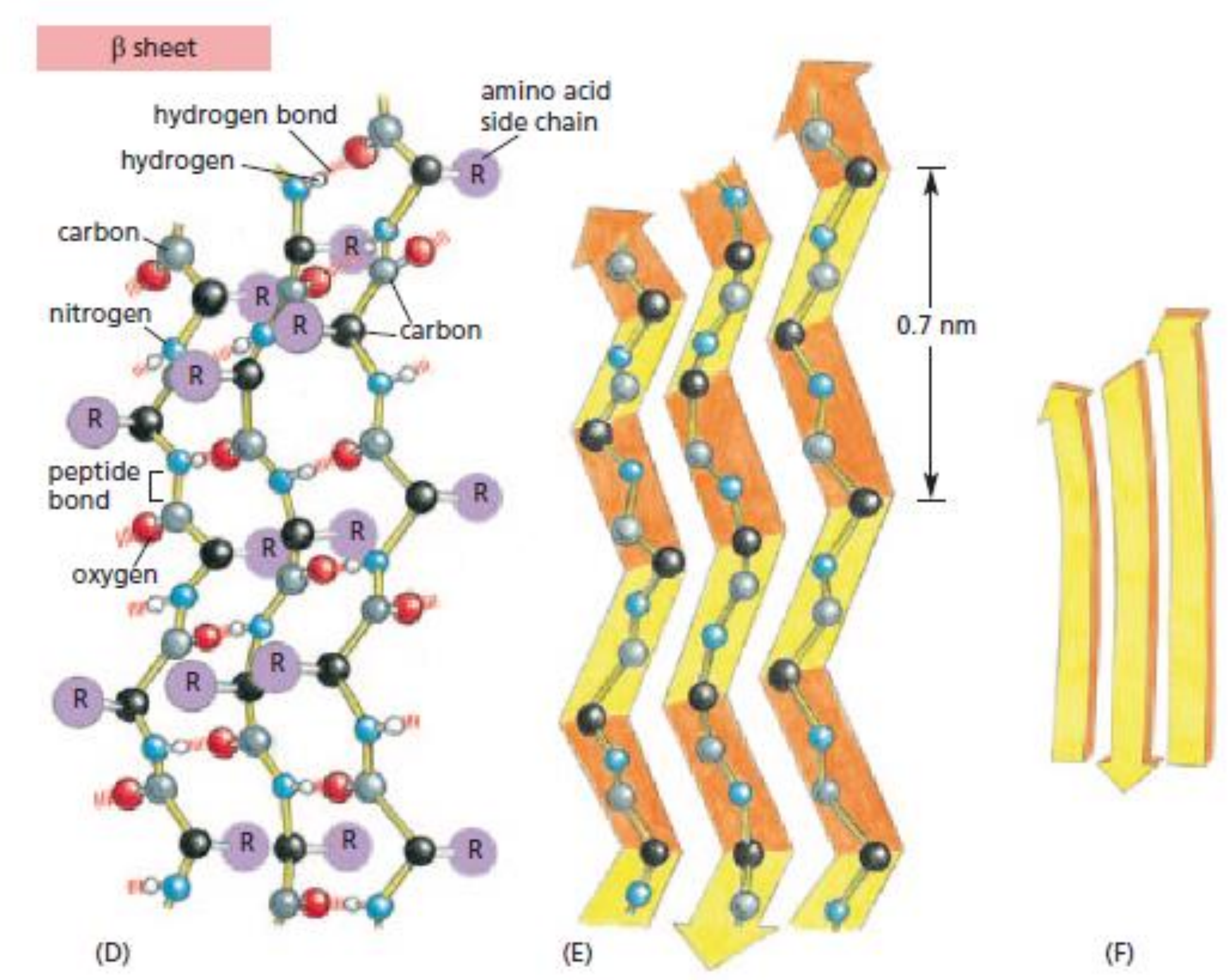
Fibrous Protein



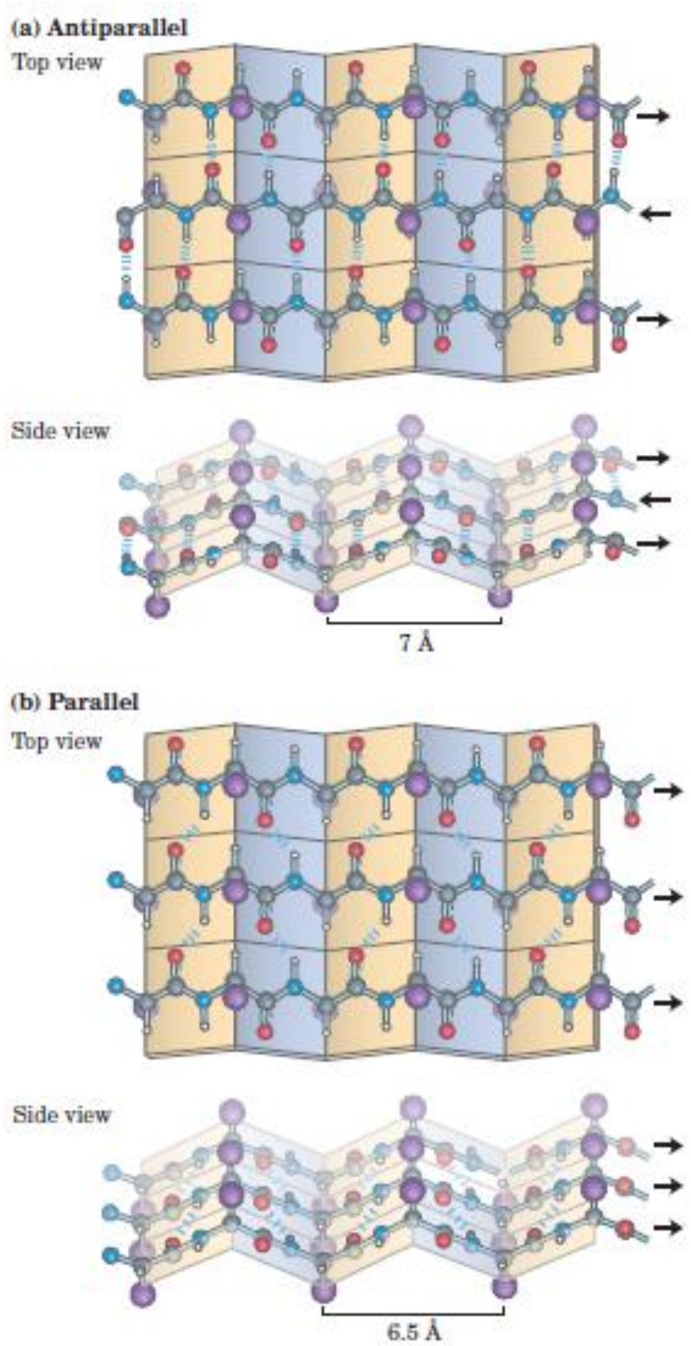


Cross section of a hair

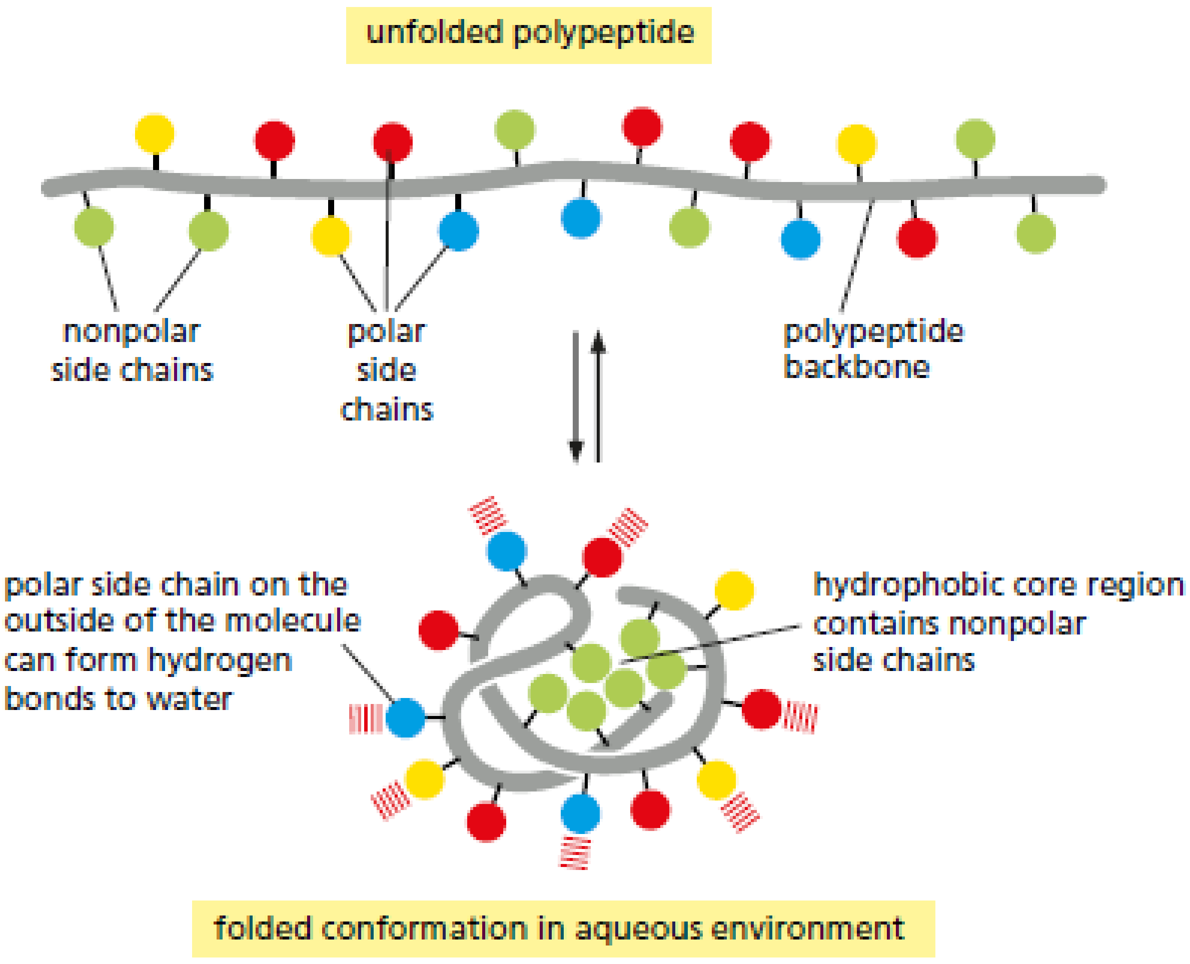
The regular conformation of the polypeptide backbone in the β sheet



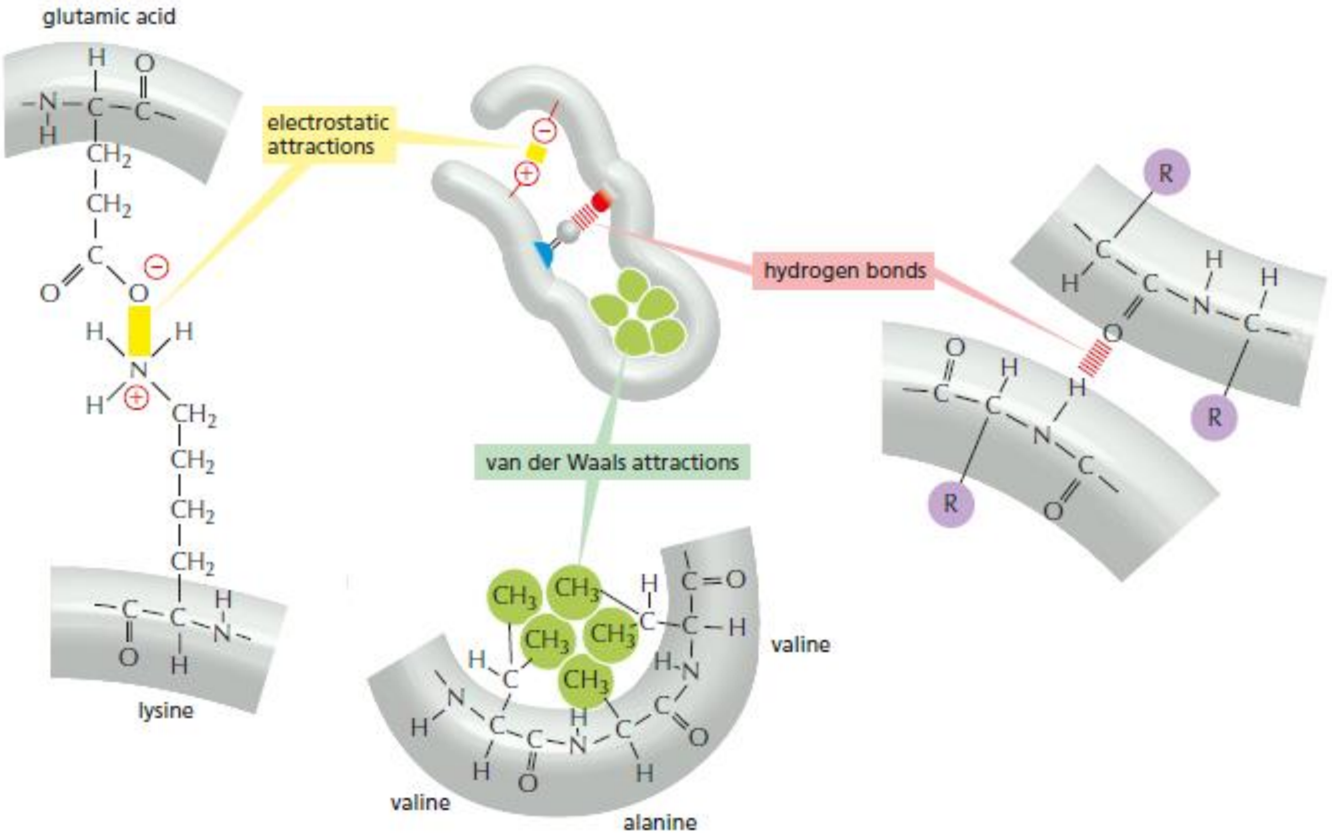
β sheet


(a) Antiparallel
Top view
Side view
7 Å
(b) Parallel
Top view
Side view
6.5 Å

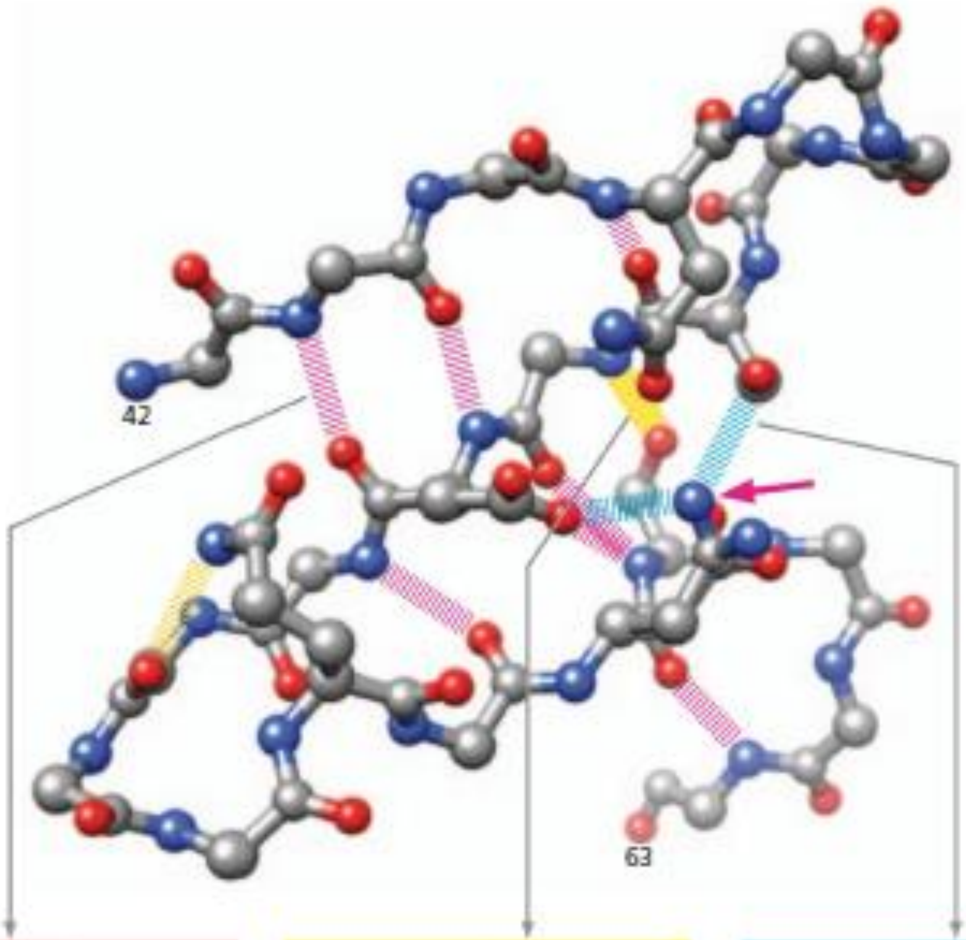
How a protein folds into a compact conformation



Three types of noncovalent bonds help proteins fold

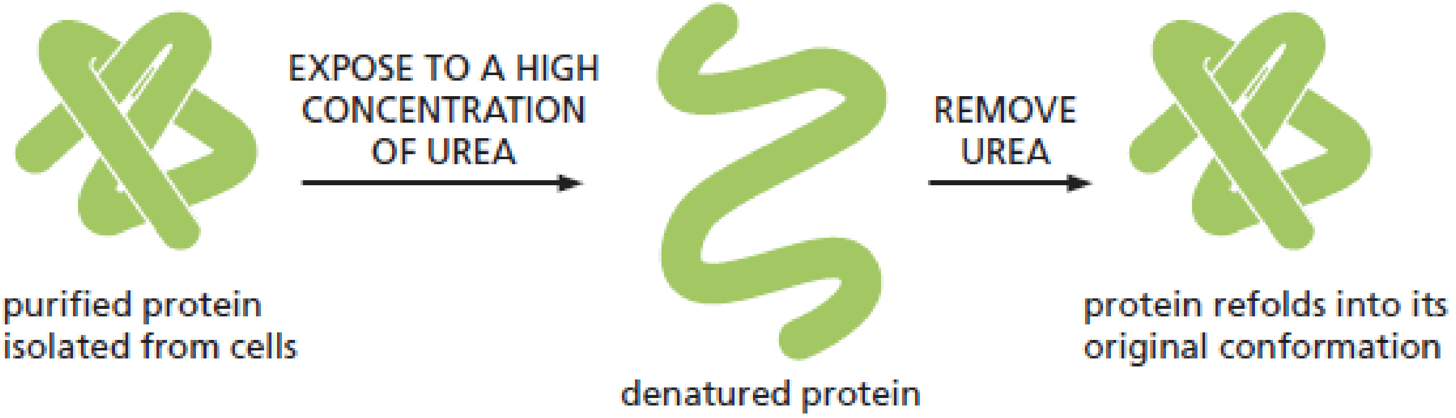


Hydrogen bonds within a protein molecule



backbone to backbone	backbone to side chain	side chain to side chain
hydrogen bond between atoms of two peptide bonds	hydrogen bond between atoms of a peptide bond and an amino acid side chain	hydrogen bond between two amino acid side chains

Denatured proteins can often recover their natural shapes

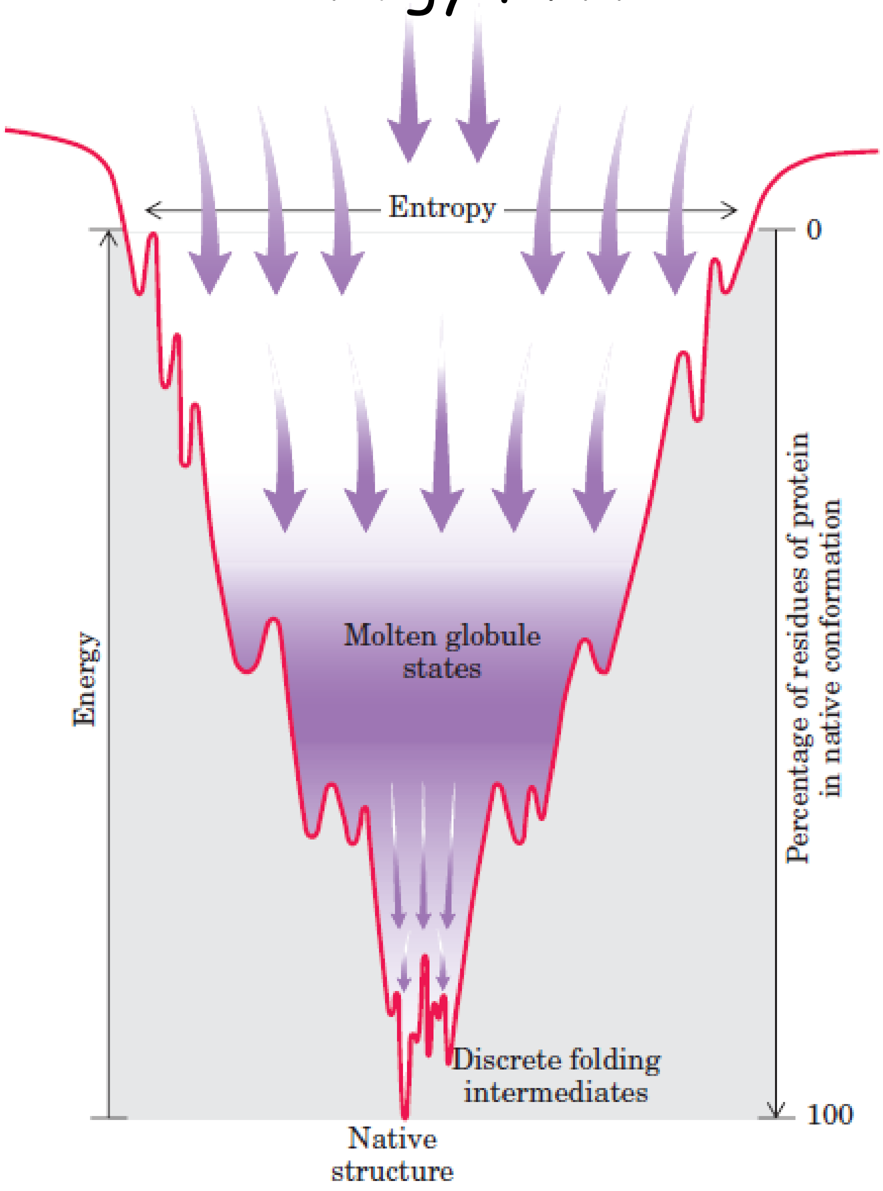


Levinthal's paradox

"Proteins have an incredible number of possible conformational states, yet they are able to fold very quickly"

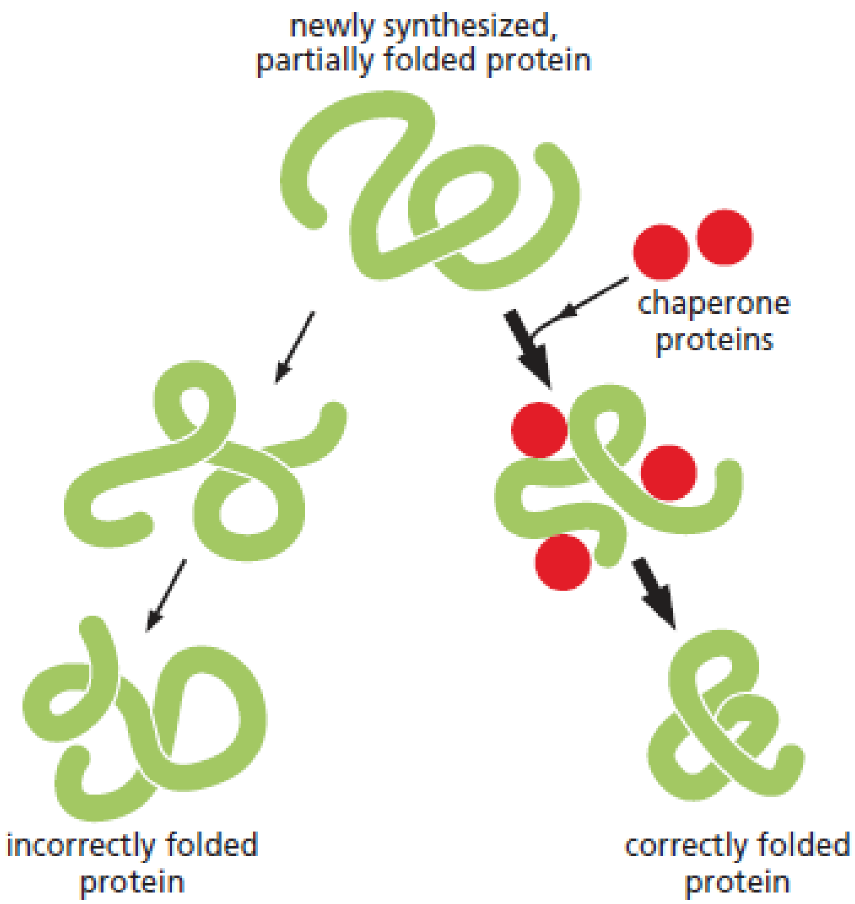
150 AA domain
10 different conformations per side chain
-------> 10^{150} possible conformations.

The thermodynamics of protein folding depicted as a free-energy funnel

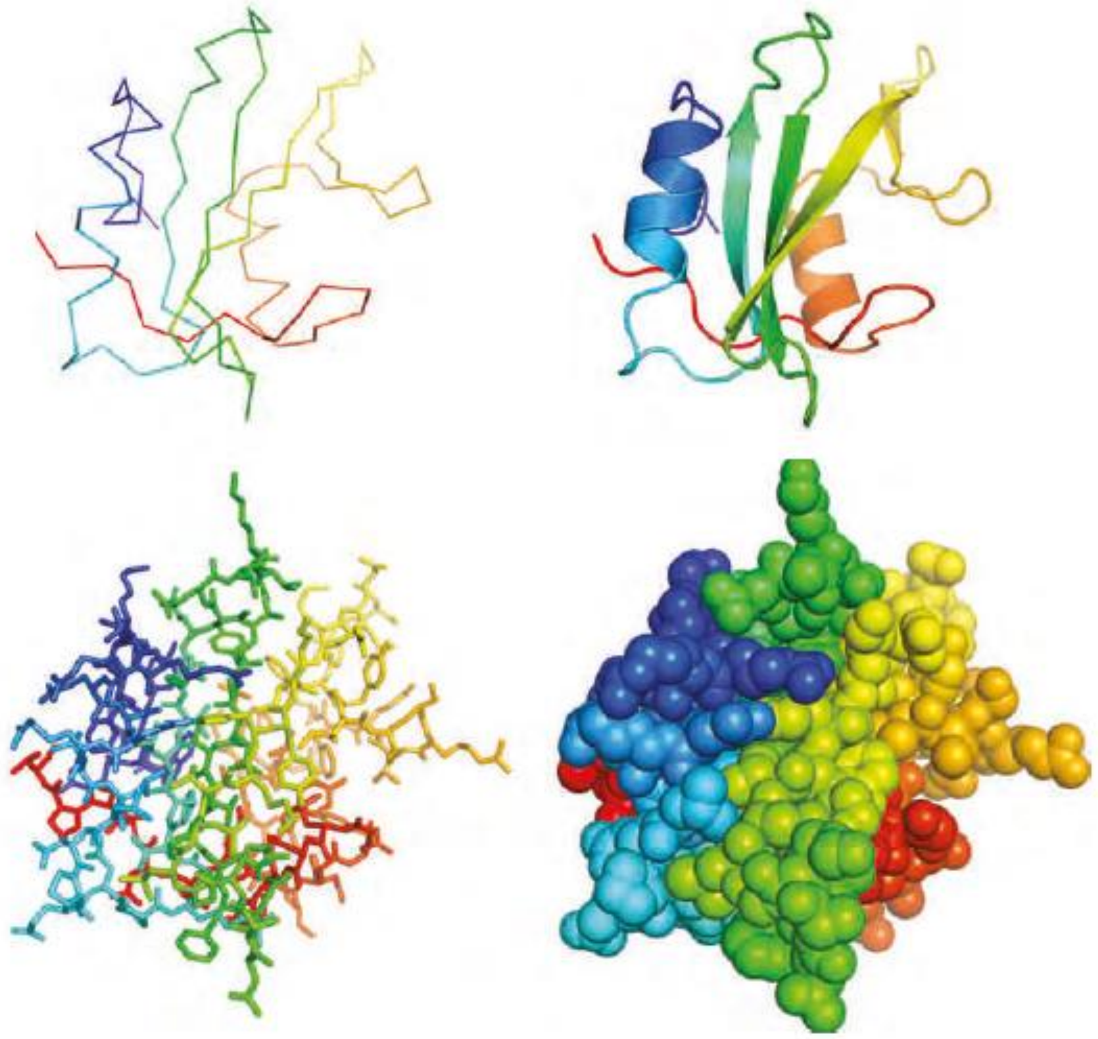


Polypeptides Fold Rapidly by a Stepwise Process

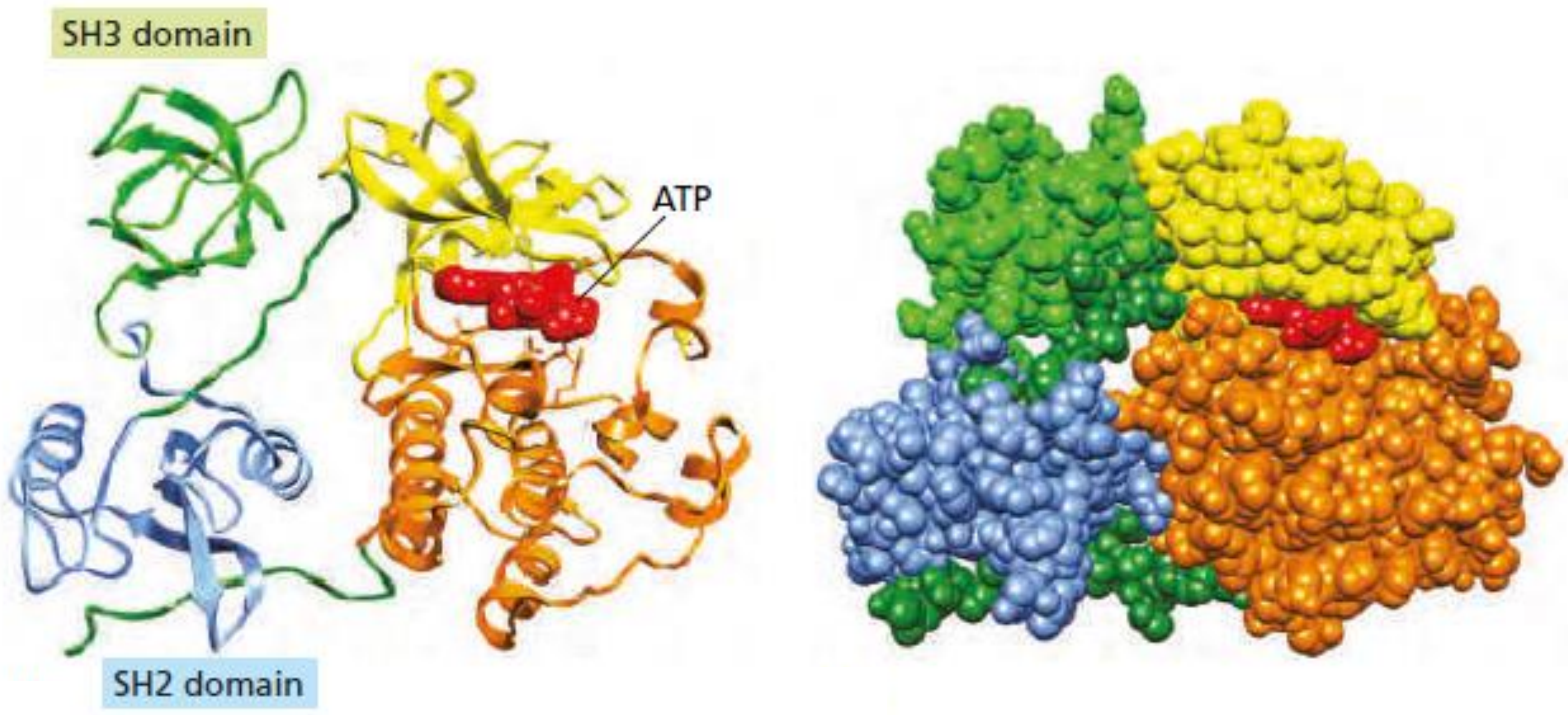
Chaperone proteins can guide the folding of a newly synthesized polypeptide chain



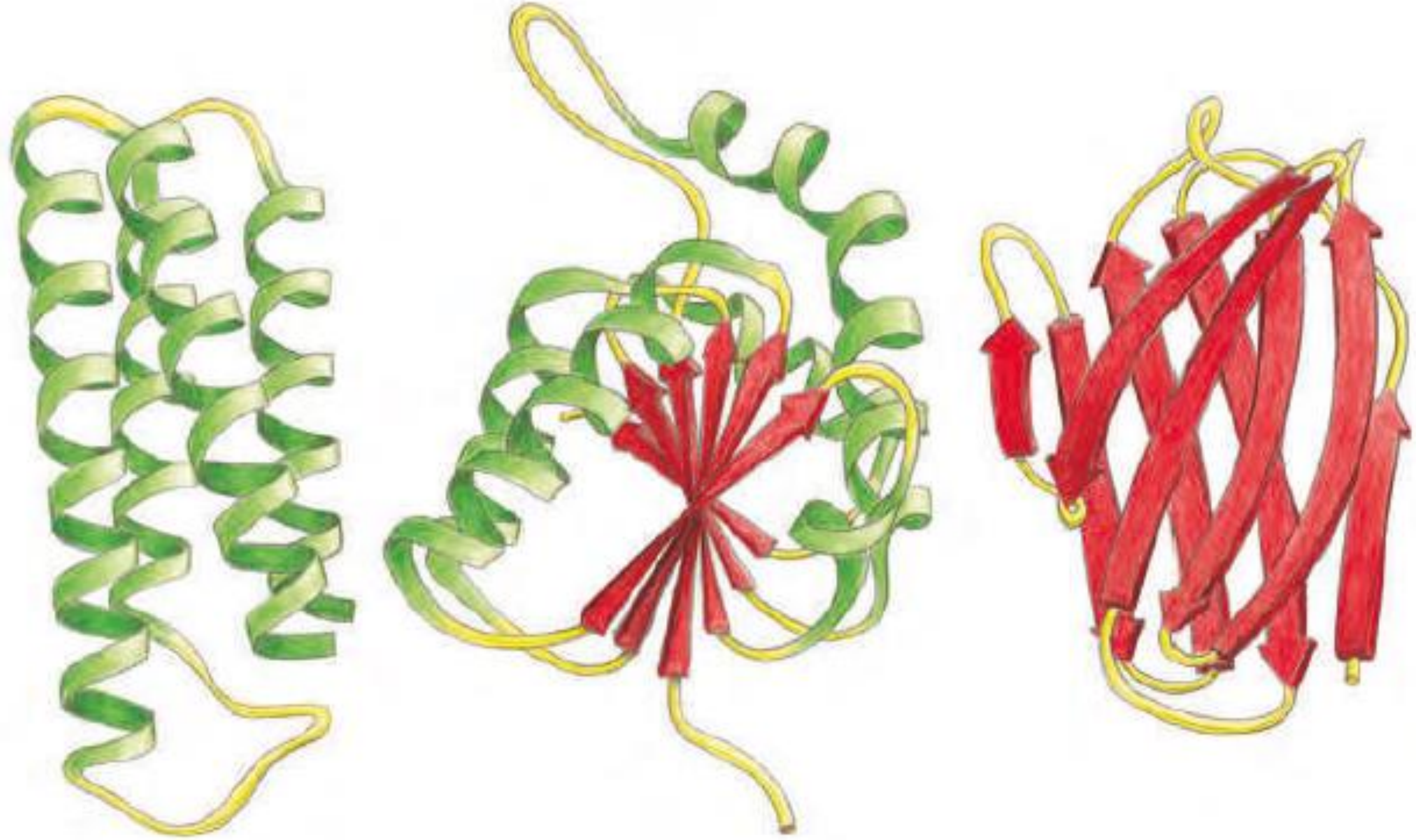
Four representations describing the structure of a small protein domain



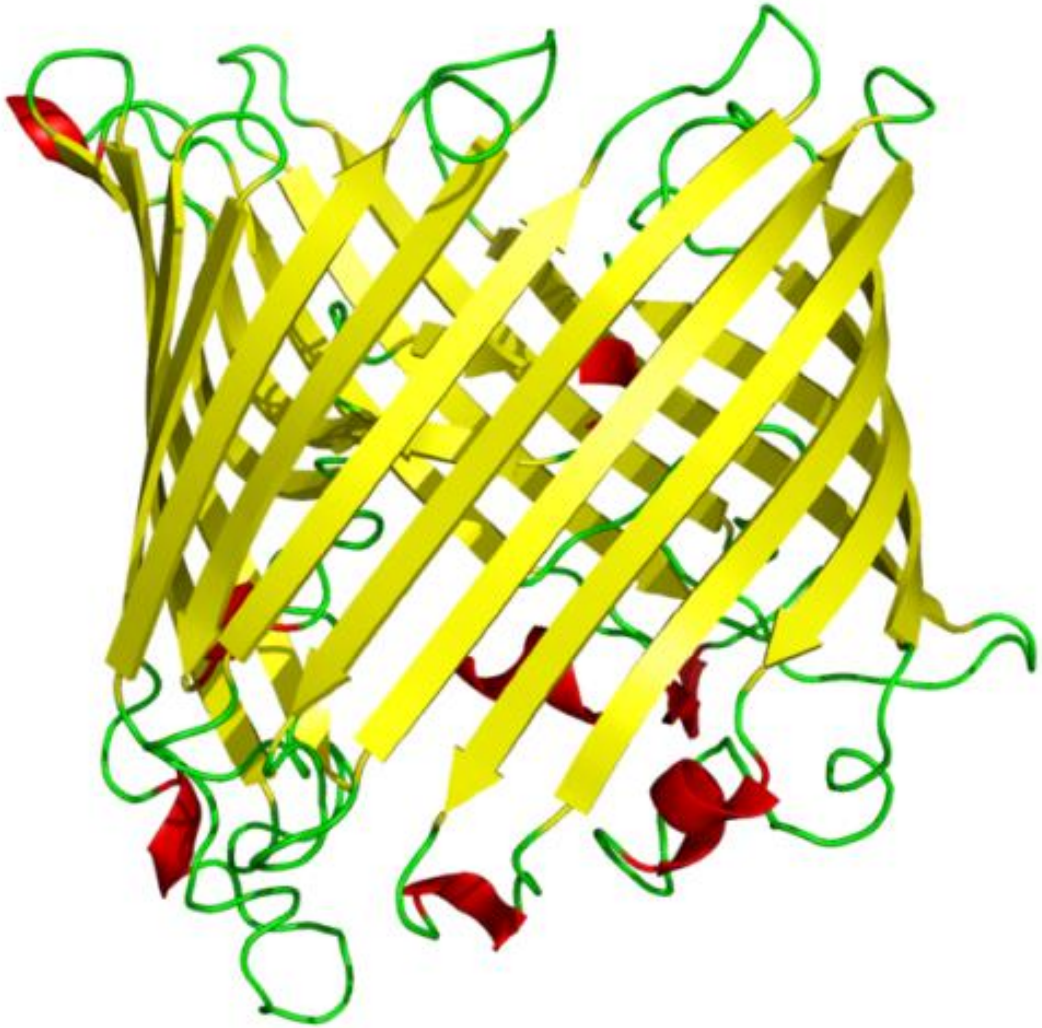
Protein Domains Are Modular Units from Which Larger Proteins Are Built



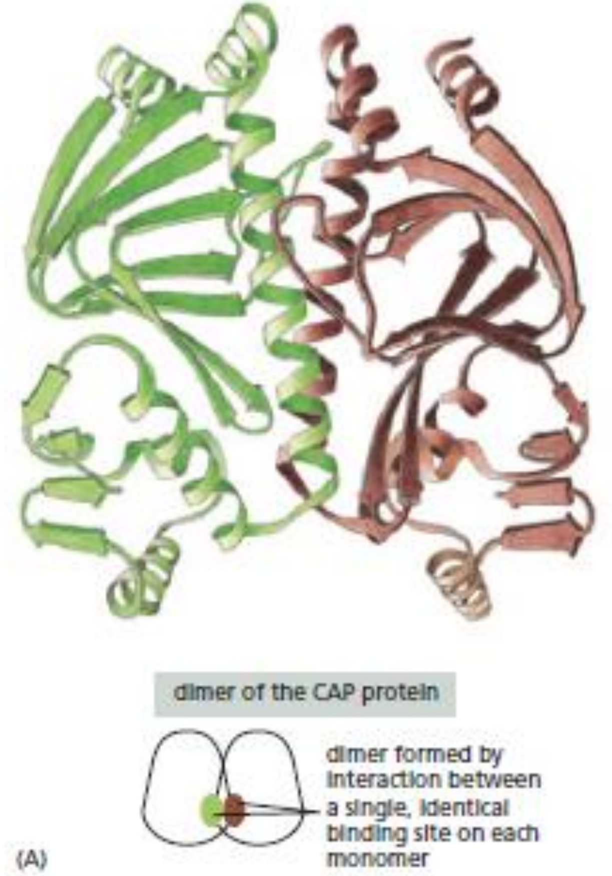
Three differently organized protein domains

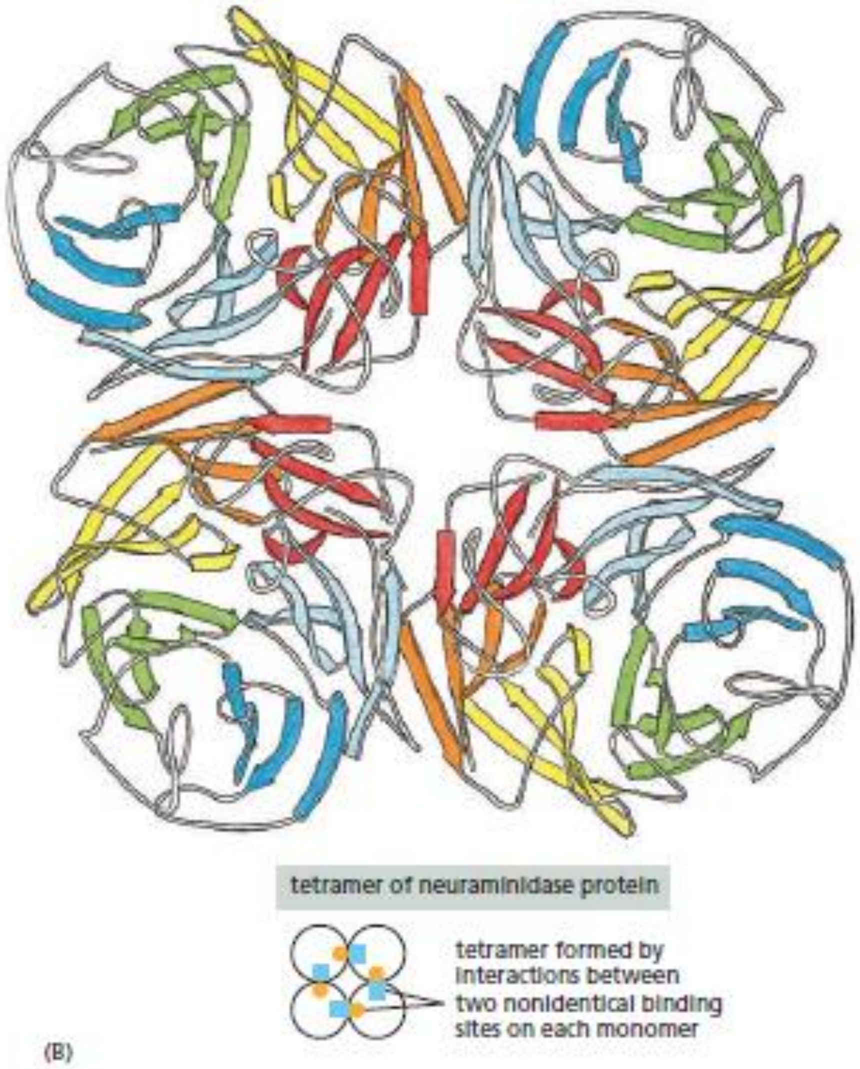


β Barrel

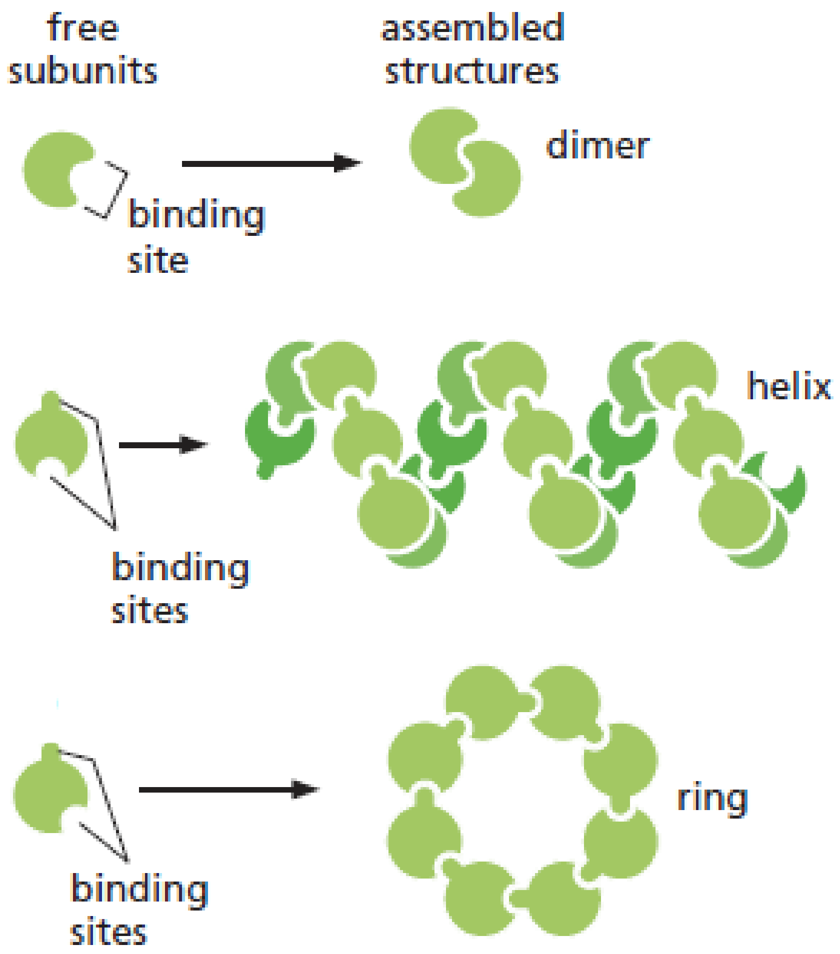


Large Protein Molecules Often Contain More Than One Polypeptide Chain

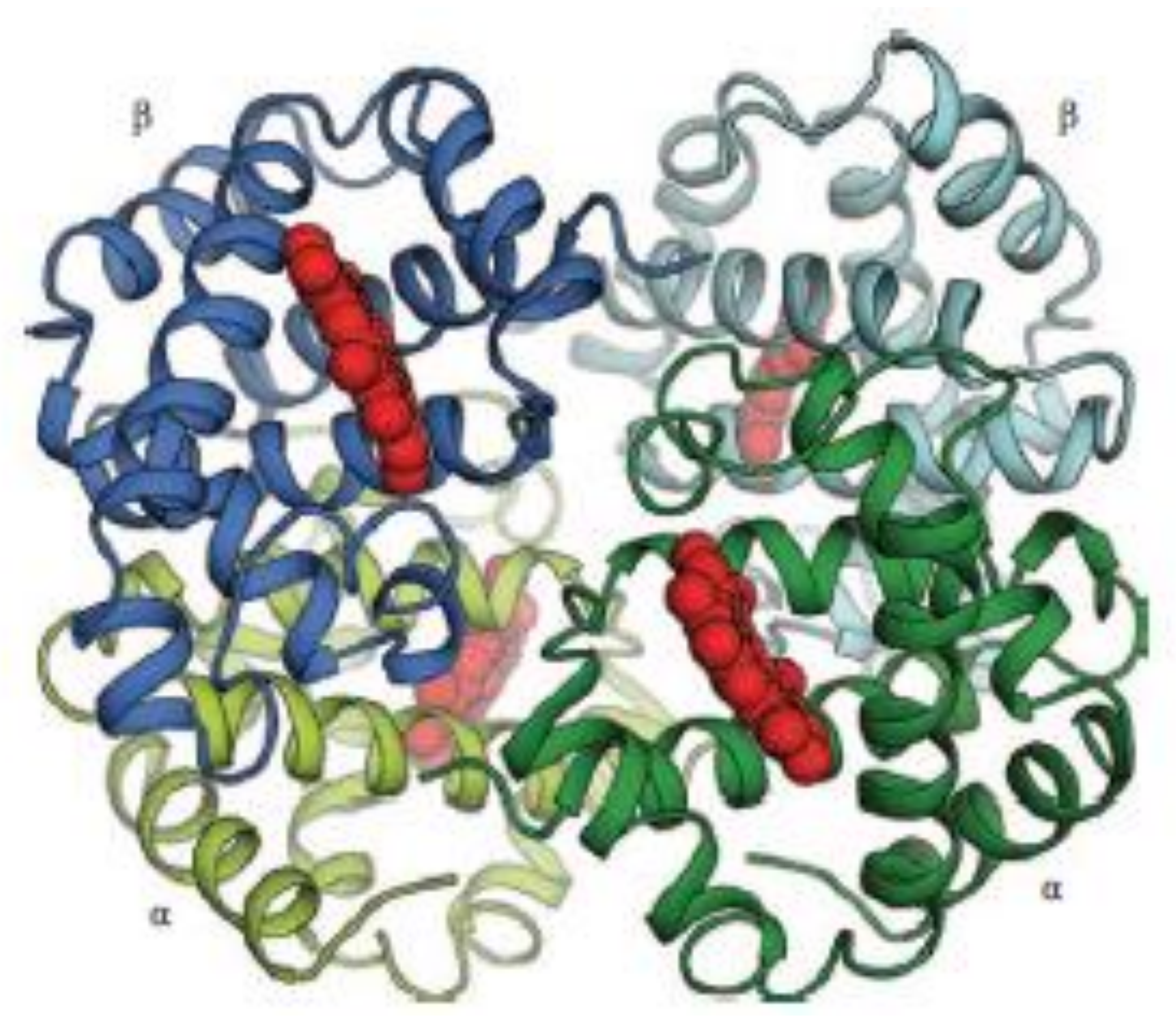




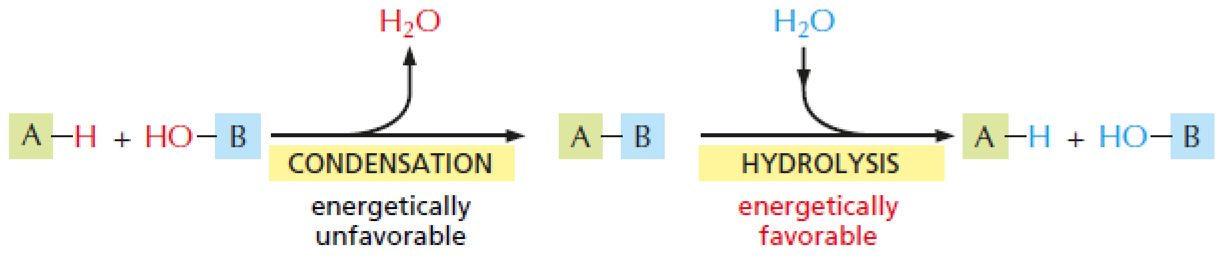
Identical protein subunits can assemble into complex structures

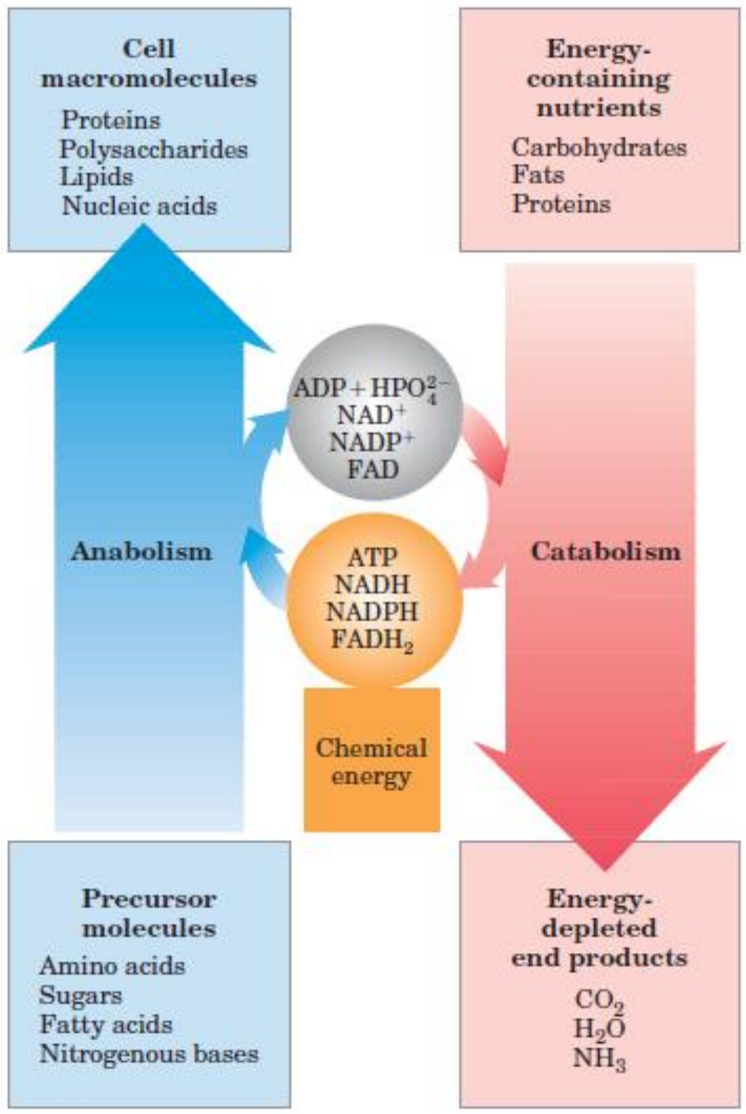


Some proteins are formed as a symmetrical assembly of two different subunits



Catalysis and use of energy by cells

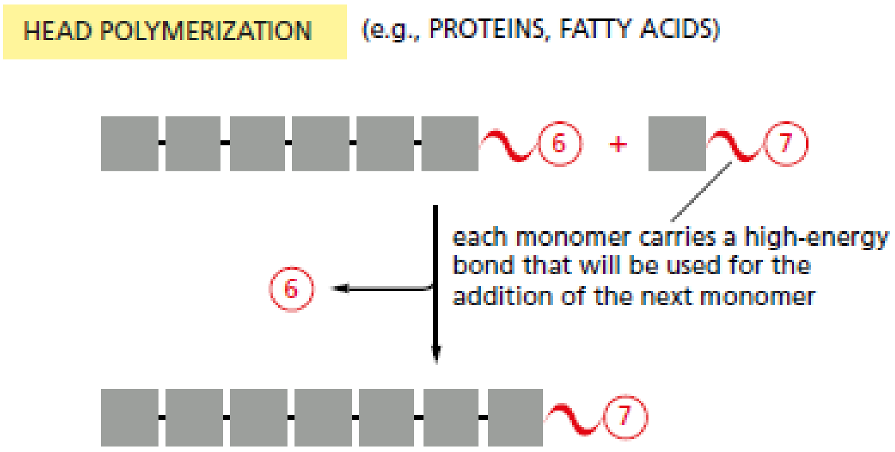


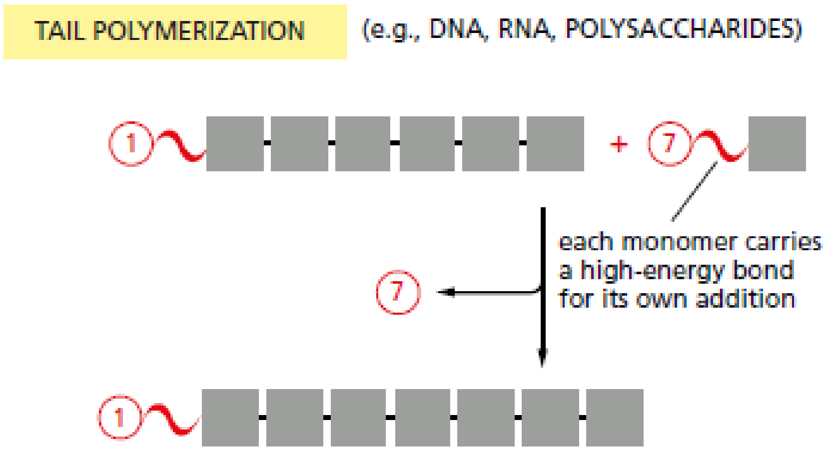


Some Activated Carrier Molecules Widely Used in Metabolism

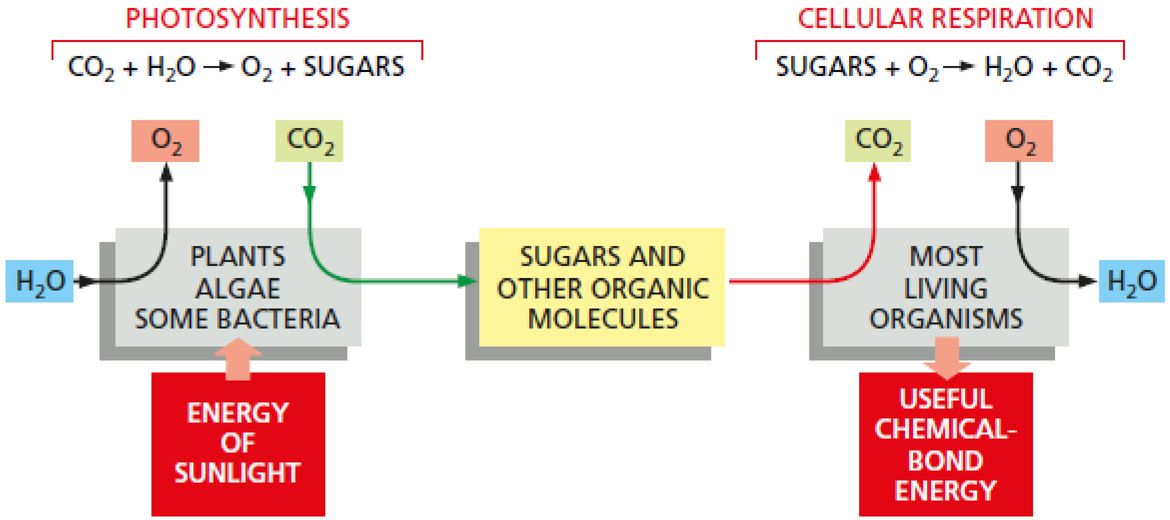
Activated carrier	Group carried in high-energy linkage
ATP	Phosphate
NADH, NADPH, $FADH_2$	Electrons and hydrogens
Acetyl CoA	Acetyl group
Carboxylated biotin	Carboxyl group
S-Adenosylmethionine	Methyl group
Uridine diphosphate glucose	Glucose

Anabolism - The orientation of the active intermediates

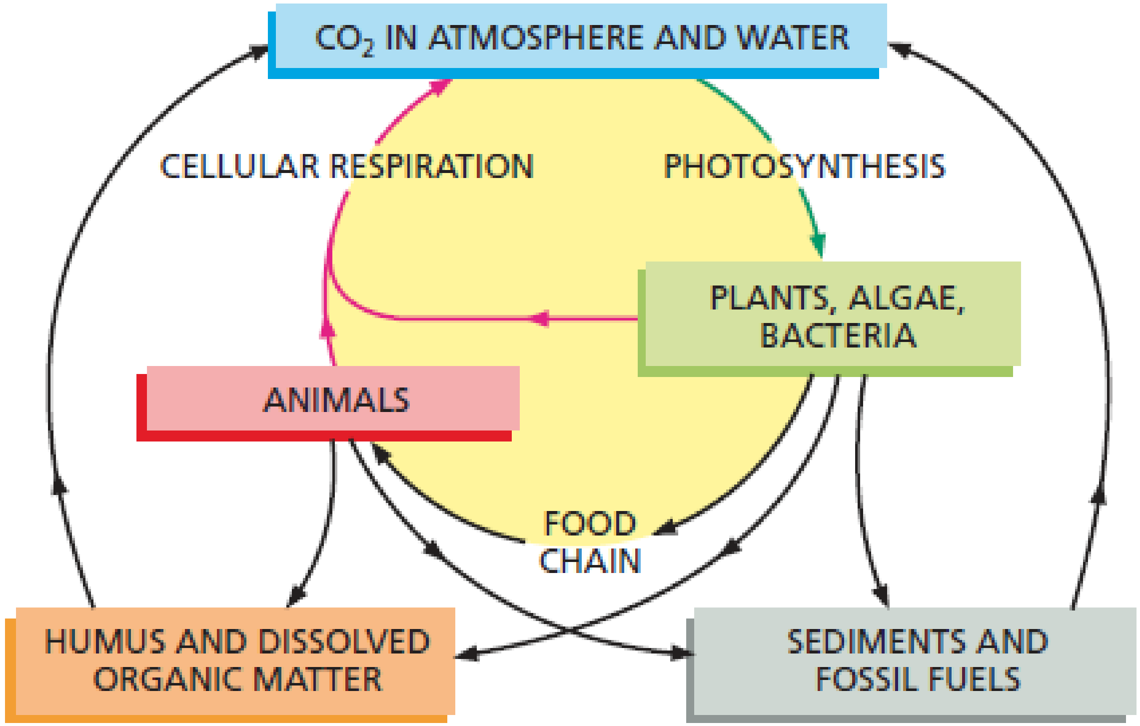




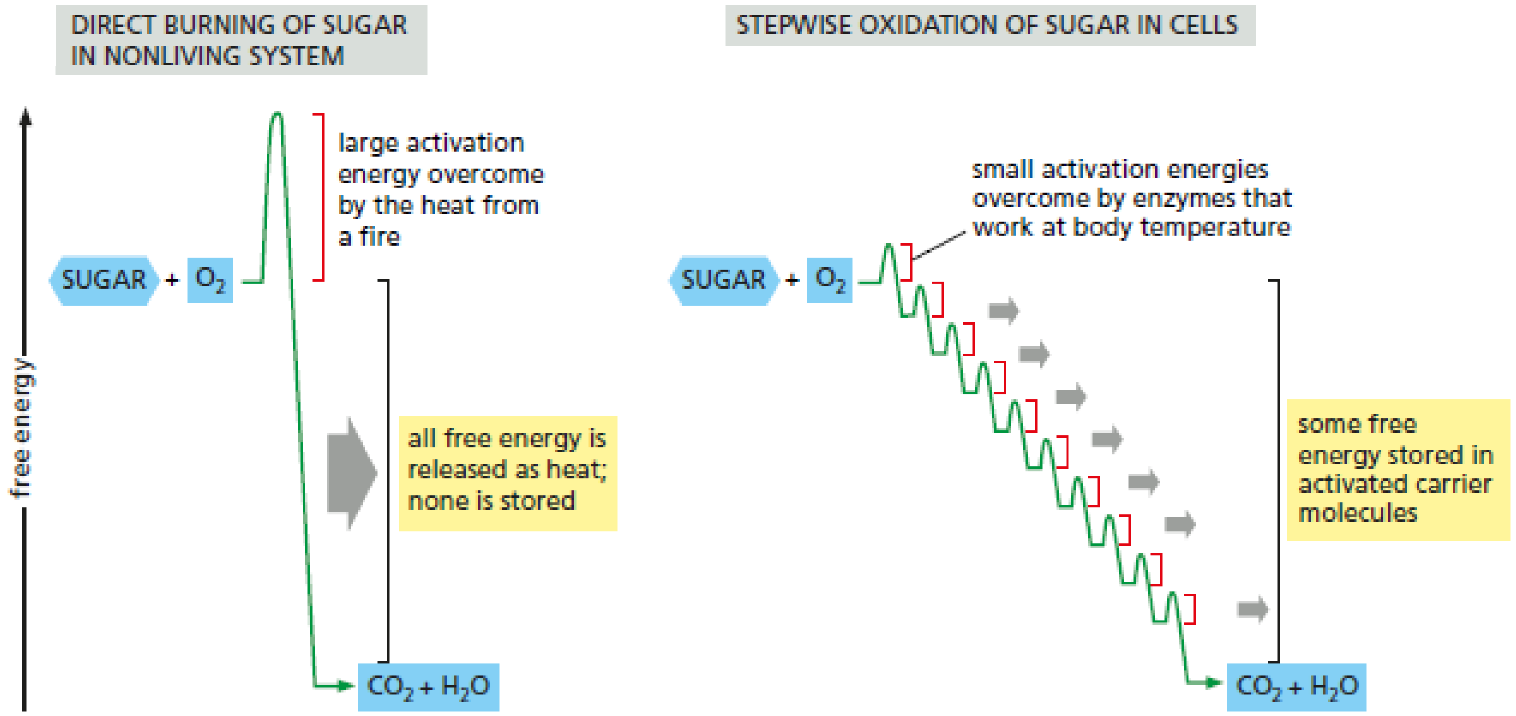
Photosynthesis and respiration as complementary processes in the living world.



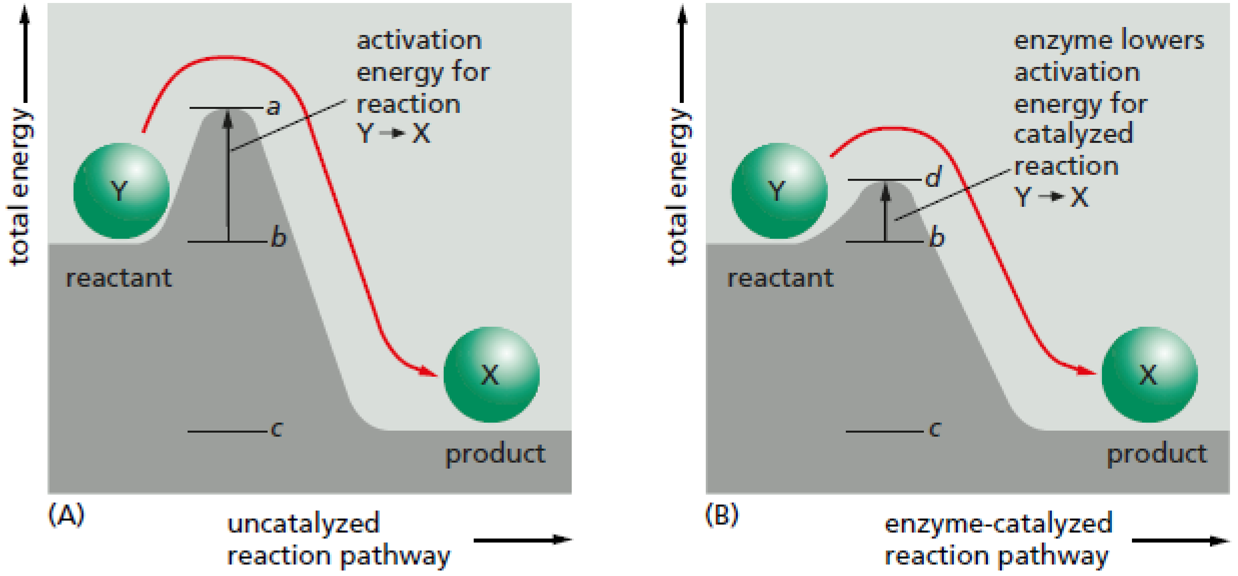
Carbon atoms cycle continuously through the biosphere



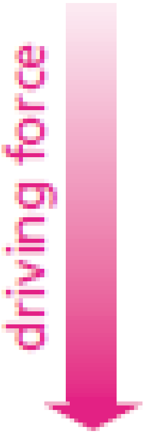
Catabolism - HOW CELLS OBTAIN ENERGY FROM FOOD ??



Principle of activation energy

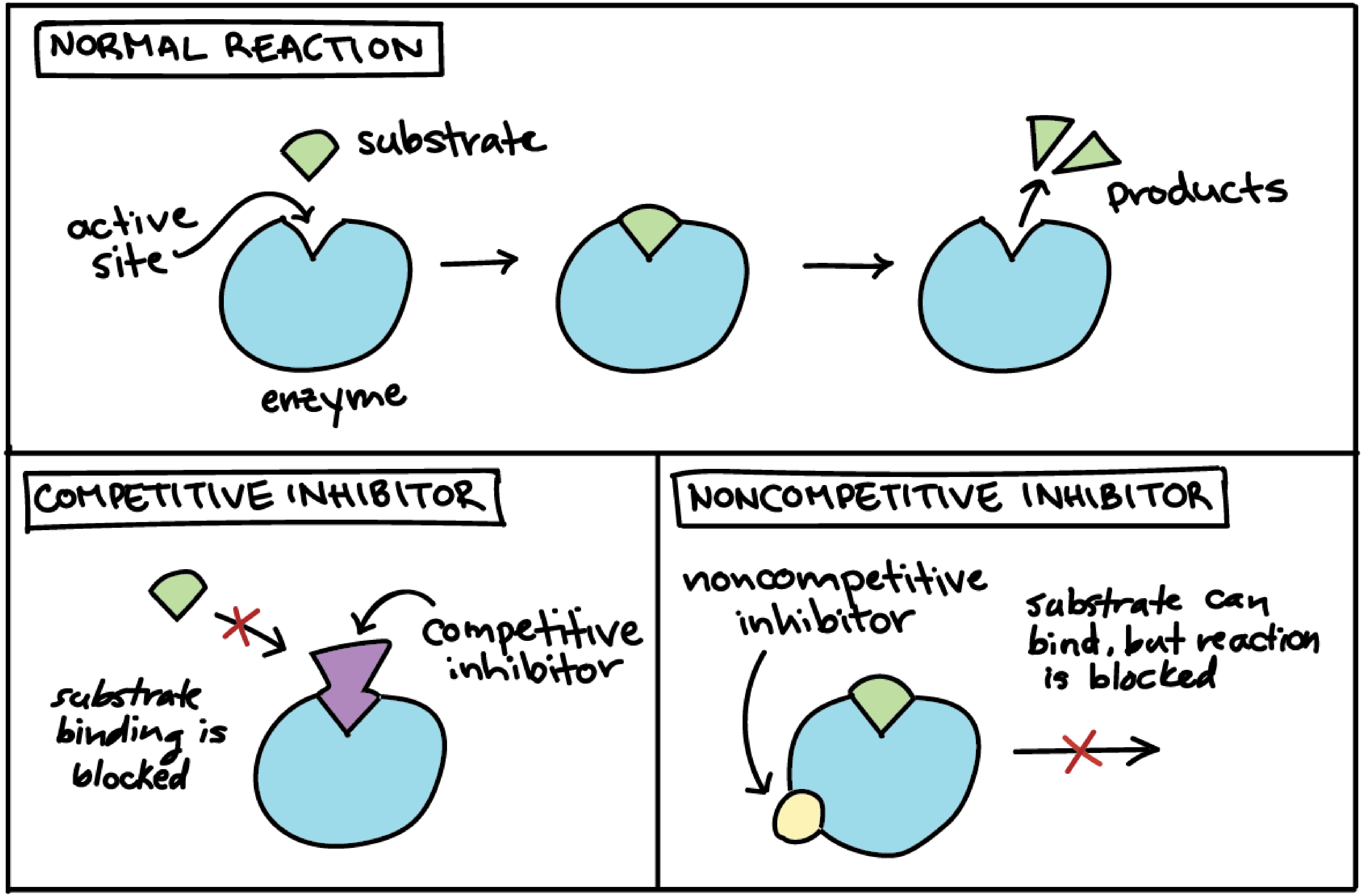


Predicting Reactions



$\Delta G°$ for some reactions

glucose-1-P $\rightarrow$ glucose-6-P	–1.7 kcal/mole
sucrose $\rightarrow$ glucose + fructose	–5.5 kcal/mole
ATP $\rightarrow$ ADP + P_i	–7.3 kcal/mole
glucose + $6O_2$ $\rightarrow$ $6CO_2$ + $6H_2O$	–686 kcal/mole


NORMAL REACTION
substrate
active site
enzyme
products
COMPETITIVE INHIBITOR
competitive inhibitor
substrate binding is blocked
NONCOMPETITIVE INHIBITOR
noncompetitive inhibitor
substrate can bind, but reaction is blocked

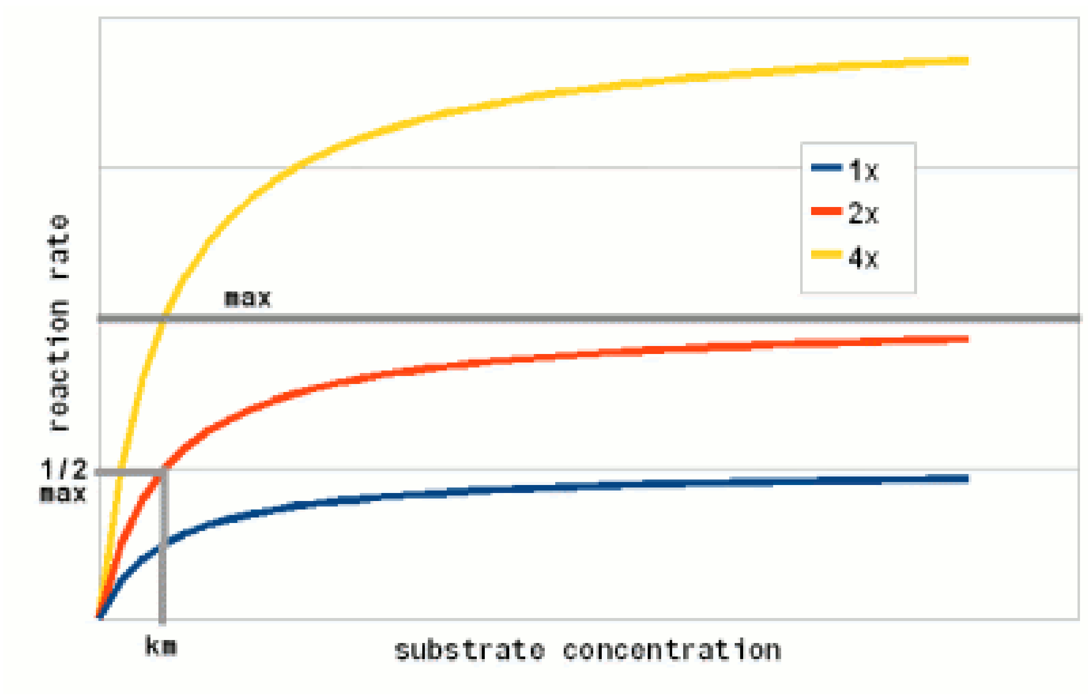

1x
2x
4x
max
1/2
max
km
reaction rate
substrate concentration

TABLE 6–3 International Classification of Enzymes

Class no.	Class name	Type of reaction catalyzed
1	Oxidoreductases	Transfer of electrons (hydride ions or H atoms)
2	Transferases	Group transfer reactions
3	Hydrolases	Hydrolysis reactions (transfer of functional groups to water)
4	Lyases	Addition of groups to double bonds, or formation of double bonds by removal of groups
5	Isomerases	Transfer of groups within molecules to yield isomeric forms
6	Ligases	Formation of C—C, C—S, C—O, and C—N bonds by condensation reactions coupled to cleavage of ATP or similar cofactor

Enzyme Commission number (E.C. number)

Enzymes

- All enzymes are assigned a number (EC number) which defines exactly which reaction is catalyzed by the enzyme.
- Example trypsin is **EC 3.4.21.4**

EC 3.4.21.4

3 describes the enzyme class (hydrolase)- hydrolysis reaction

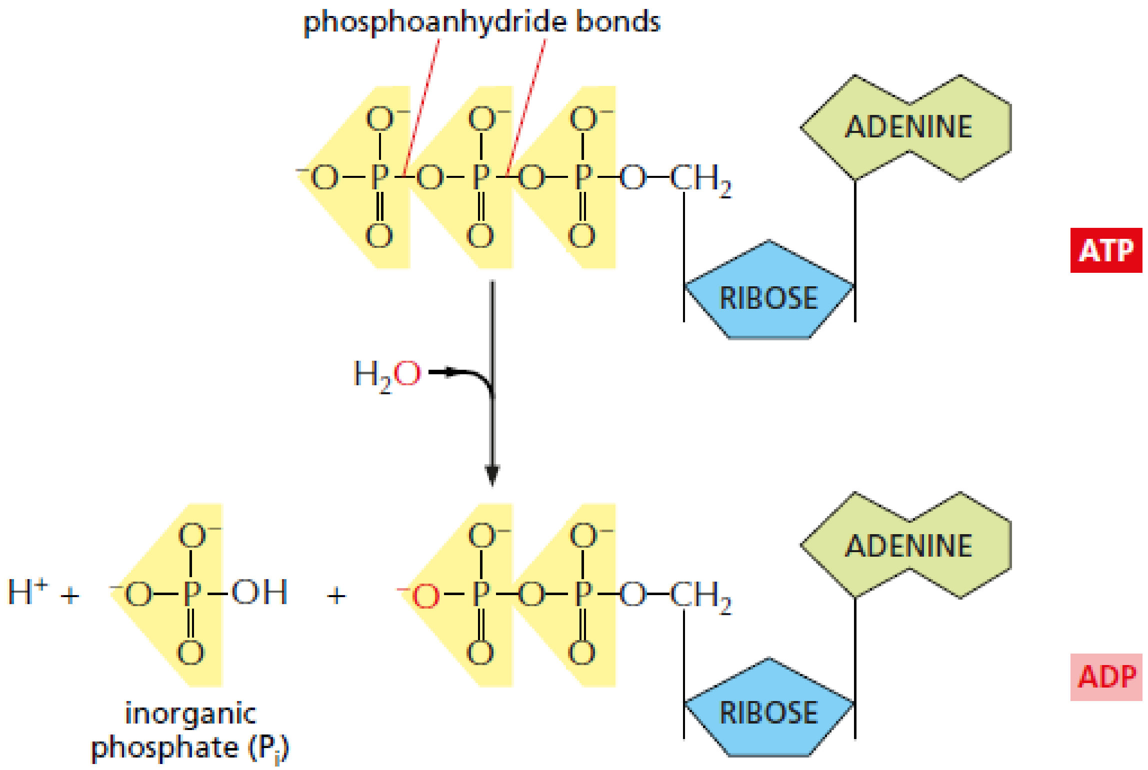
4 is for the subclass (acts on peptide bonds, so it is a peptide hydrolase)

21 denotes its sub-subclass as a serine peptidase

4 because it is the 4th entry in this subclass

- Enzyme Nomenclature Database - http://expasy.org/enzyme

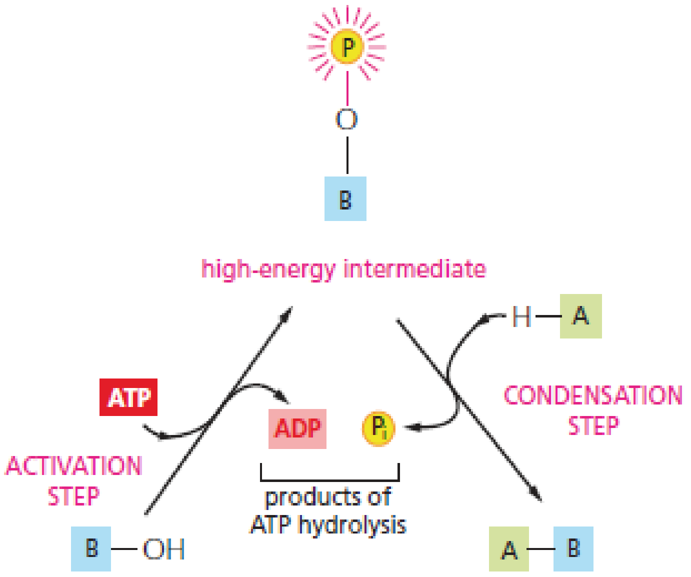
ATP Is the Most Widely Used Activated Carrier Molecule

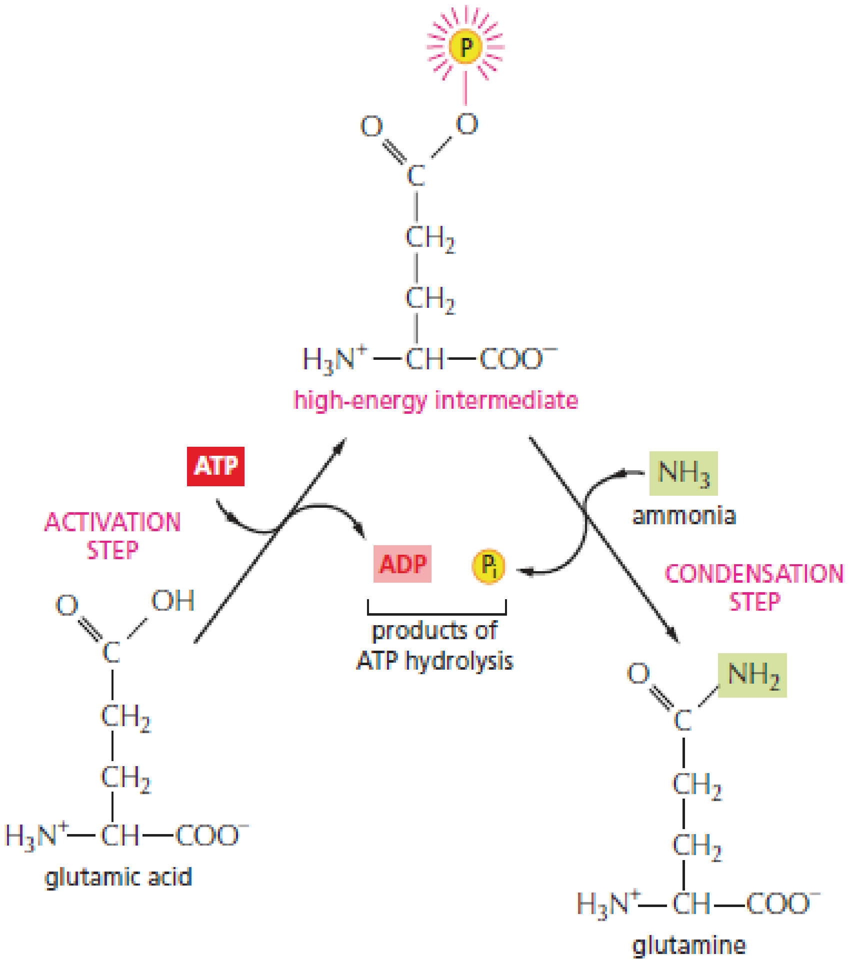


Hydrolysis of the terminal phosphate of ATP yields between 46 and 54 kJ/mole of usable energy

Biosynthetic reaction driven by ATP hydrolysis

$$A\text{–}H + B\text{–}OH \rightarrow A\text{–}B + H_2O$$

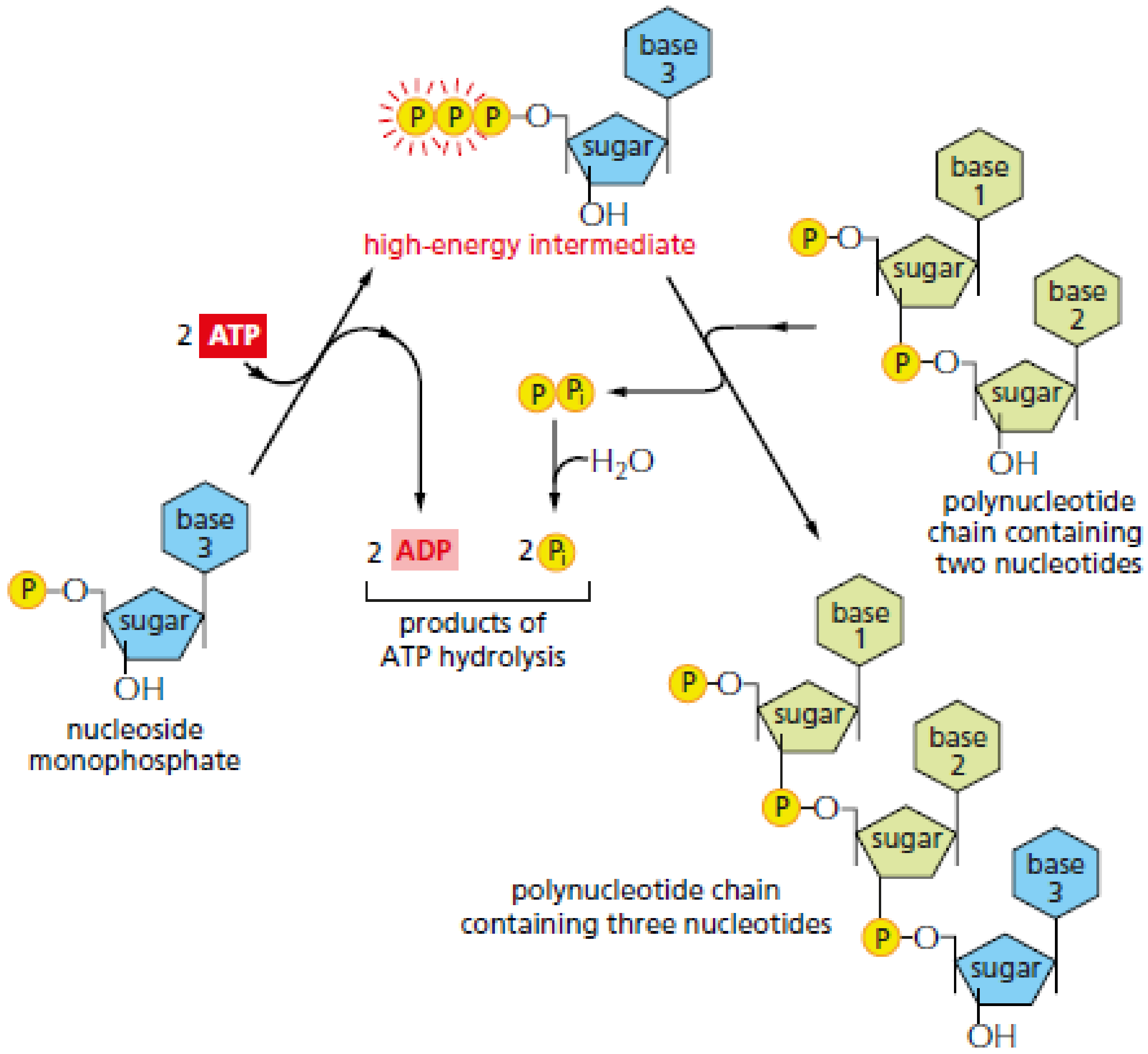




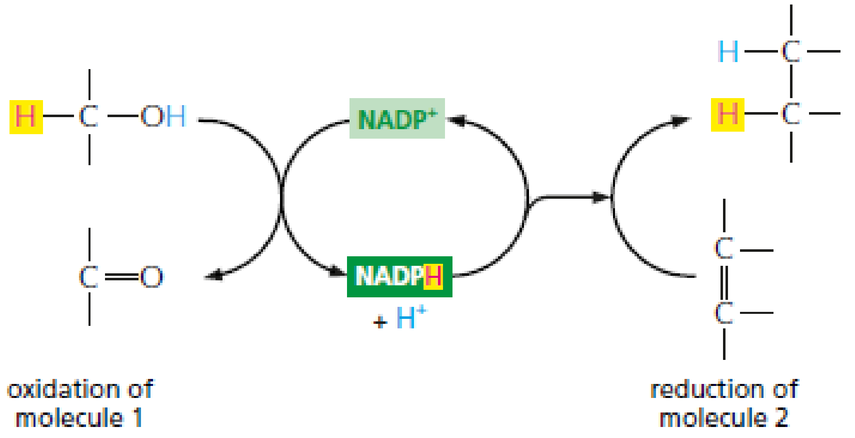
1. $B\text{–}OH + ATP \longrightarrow B\text{–}O\text{–}PO_3 + ADP$
2. $A\text{–}H + B\text{–}O\text{–}PO_3 \longrightarrow A\text{–}B + P_i$

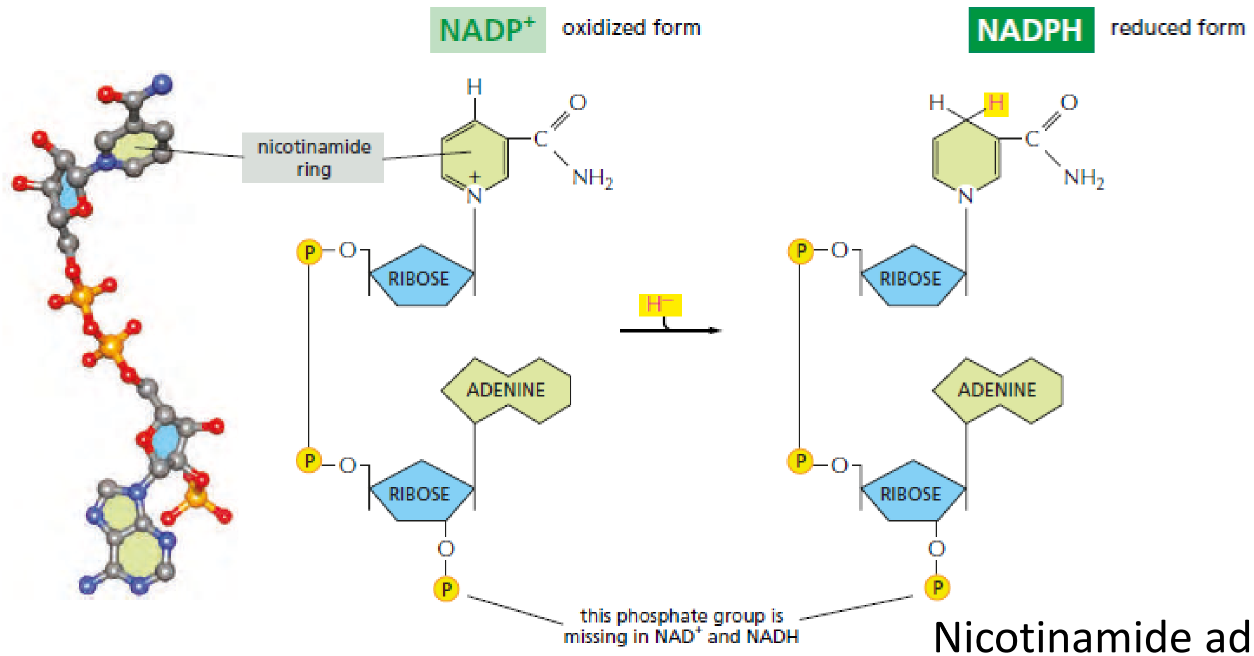
Net result: $B\text{–}OH + ATP + A\text{–}H \longrightarrow A\text{–}B + ADP + P_i$

Synthesis of a polynucleotide, RNA or DNA, is a multistep process driven by ATP hydrolysis



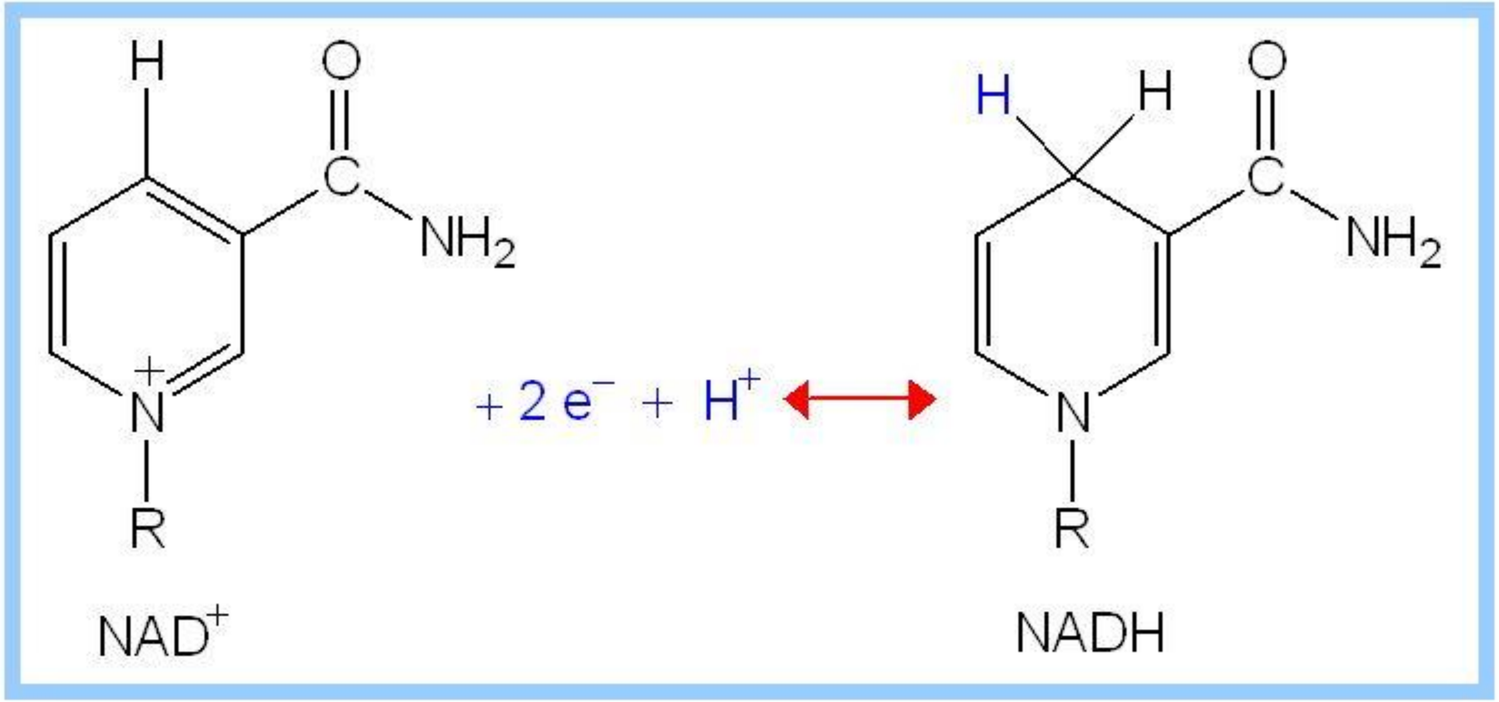
NADH and NADPH Are Important Electron Carriers





Nicotinamide adenine dinucleotide

$NAD^+/NADH$



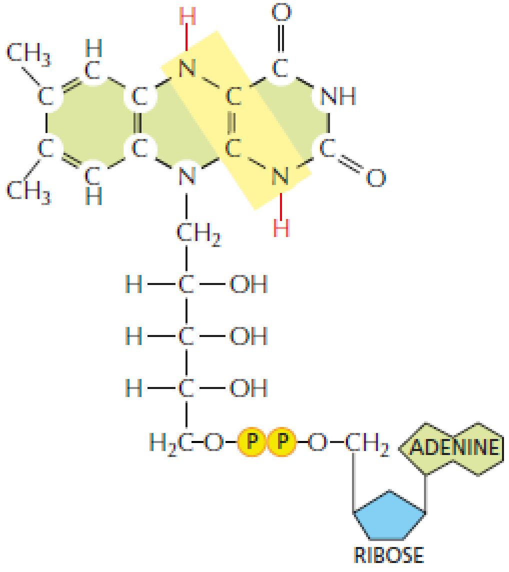
The electron transfer reaction may be summarized as :

$NAD^+ + 2e^- + H^+ \leftrightarrow NADH.$

It may also be written as:

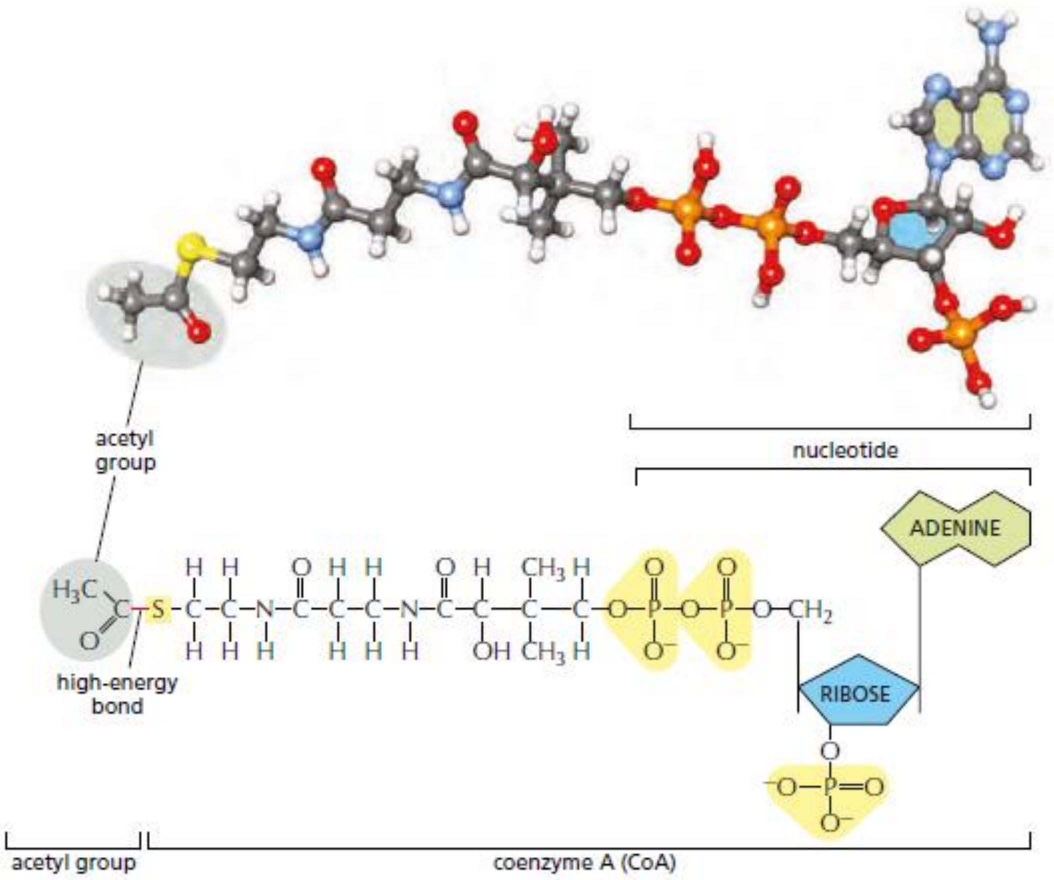
$NAD^+ + 2e^- + 2H^+ \leftrightarrow NADH + H^+$

$FADH_2$



Flavin adenine dinucleotide

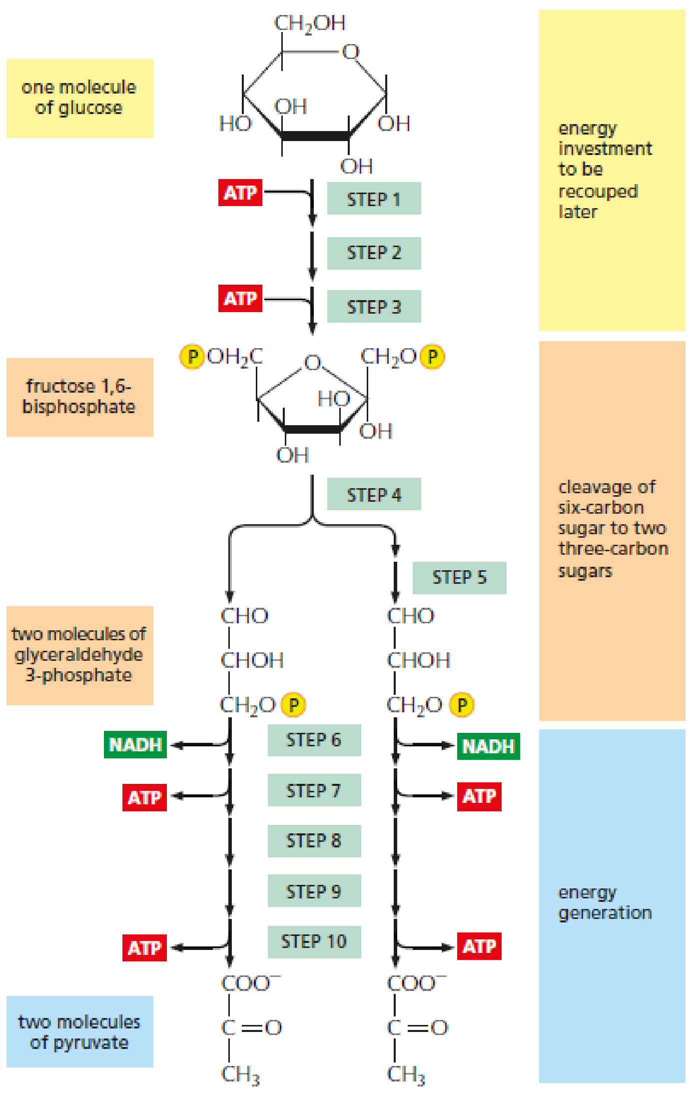
acetyl CoA



acetyl CoA transfers two-carbon acetyl groups

used to add two carbon units in the biosynthesis of larger molecules

Glycolysis is a Central ATP-Producing Pathway



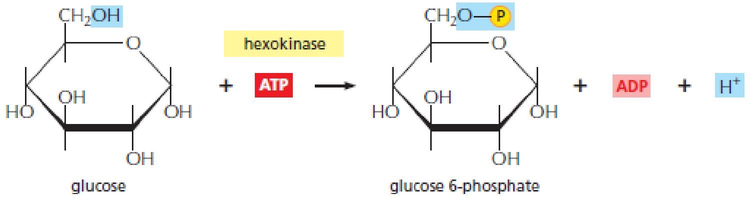
The overall equation for glycolysis

$$\text{Glucose} + 2\text{NAD}^+ + 2\text{ADP} + 2\text{P}_i \longrightarrow 2\text{ pyruvate} + 2\text{NADH} + 2\text{H}^+ + 2\text{ATP} + 2\text{H}_2\text{O}$$

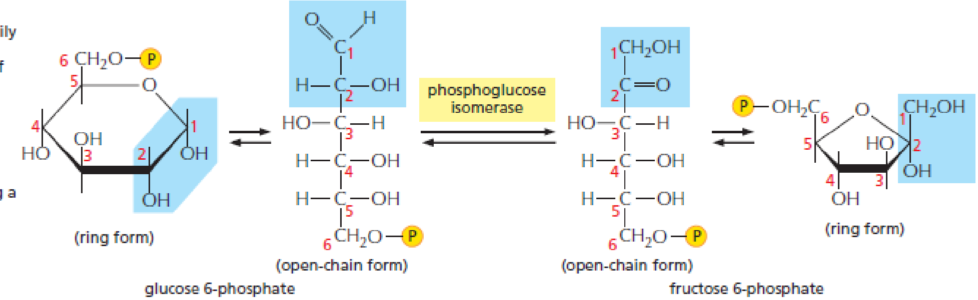
Some types of enzymes involved in glycolysis

Enzyme type	General function	Role in glycolysis
Kinase	catalyzes the addition of a phosphate group to molecules	a kinase transfers a phosphate group from ATP to a substrate in steps 1 and 3; other kinases transfer a phosphate to ADP to form ATP in steps 7 and 10
Isomerase	catalyzes the rearrangement of bonds within a single molecule	isomerases in steps 2 and 5 prepare molecules for the chemical alterations to come
Dehydrogenase	catalyzes the oxidation of a molecule by removing a hydrogen atom plus an electron (a hydride ion, H^-)	the enzyme glyceraldehyde 3-phosphate dehydrogenase generates NADH in step 6
Mutase	catalyzes the shifting of a chemical group from one position to another within a molecule	the movement of a phosphate by phosphoglycerate mutase in step 8 helps prepare the substrate to transfer this group to ADP to make ATP in step 10

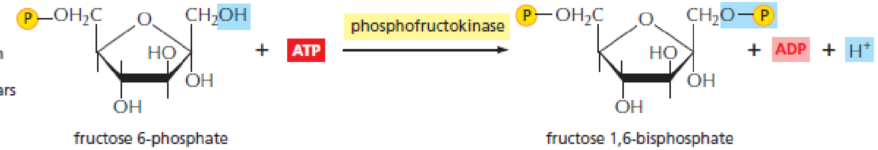
Step 1 Glucose is phosphorylated by ATP to form a sugar phosphate. The negative charge of the phosphate prevents passage of the sugar phosphate through the plasma membrane, trapping glucose inside the cell.



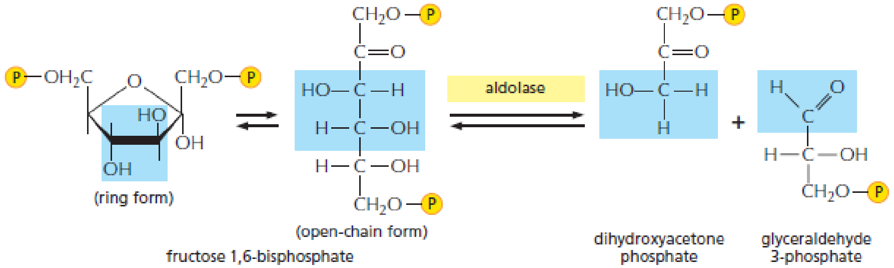
Step 2 A readily reversible rearrangement of the chemical structure (isomerization) moves the carbonyl oxygen from carbon 1 to carbon 2, forming a ketose from an aldose sugar. (See Panel 2–3, pp. 70–71.)



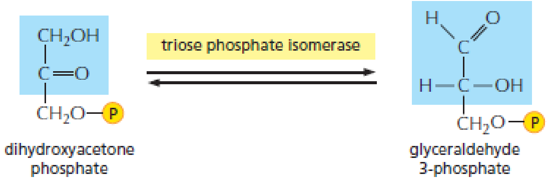
Step 3 The new hydroxyl group on carbon 1 is phosphorylated by ATP, in preparation for the formation of two three-carbon sugar phosphates. The entry of sugars into glycolysis is controlled at this step, through regulation of the enzyme *phosphofructokinase*.



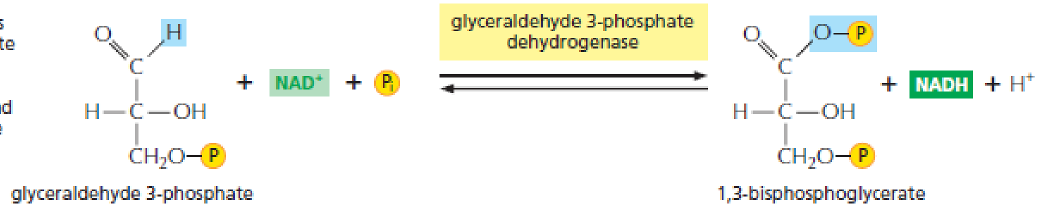
Step 4 The six-carbon sugar is cleaved to produce two three-carbon molecules. Only the glyceraldehyde 3-phosphate can proceed immediately through glycolysis.



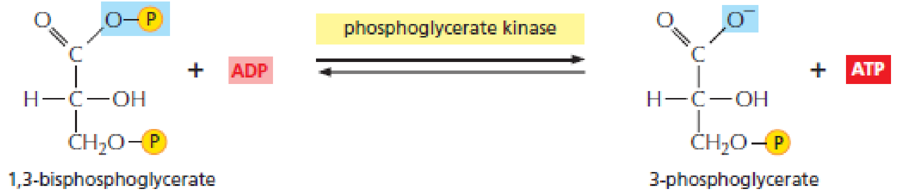
Step 5 The other product of step 4, dihydroxyacetone phosphate, is isomerized to form glyceraldehyde 3-phosphate.



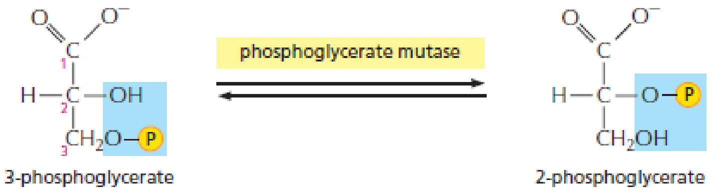
Step 6 The two molecules of glyceraldehyde 3-phosphate are oxidized. The energy-generation phase of glycolysis begins, as NADH and a new high-energy anhydride linkage to phosphate are formed (see Figure 13–5).



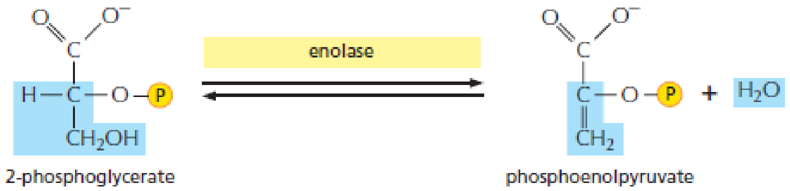
Step 7 The transfer to ADP of the high-energy phosphate group that was generated in step 6 forms ATP.



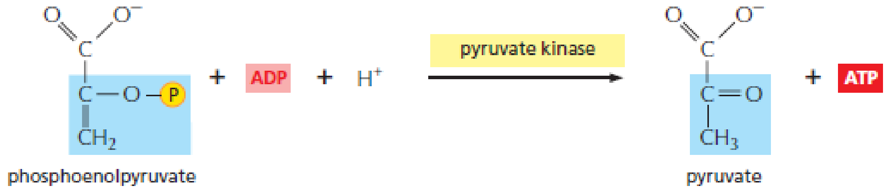
Step 8 The remaining phosphate ester linkage in 3-phosphoglycerate, which has a relatively low free energy of hydrolysis, is moved from carbon 3 to carbon 2 to form 2-phosphoglycerate.



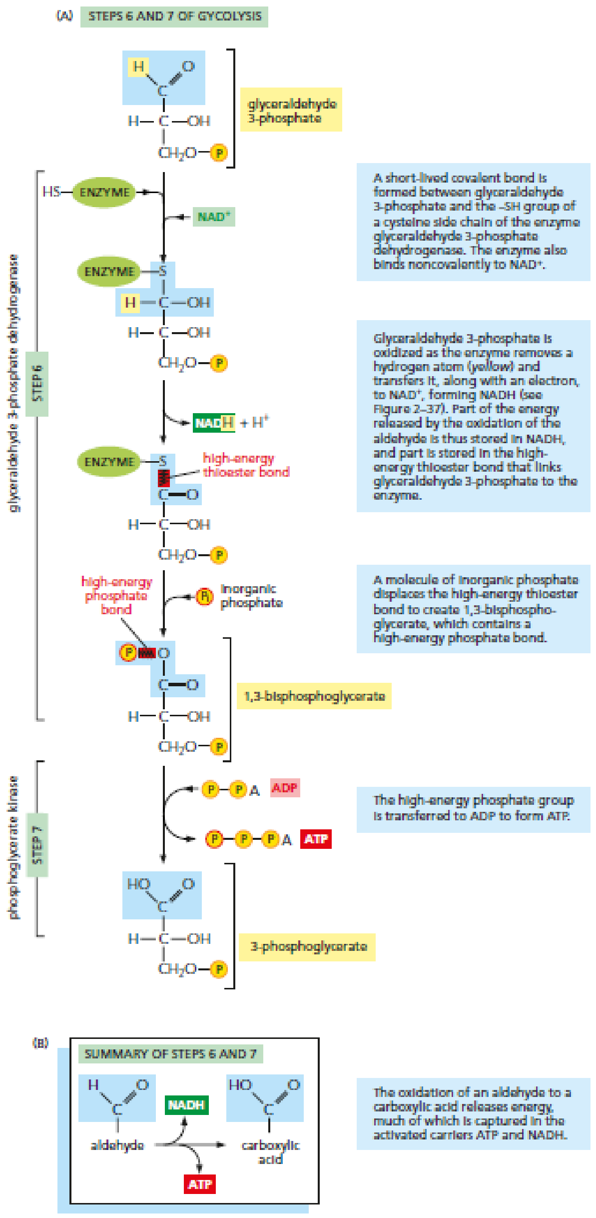
Step 9 The removal of water from 2-phosphoglycerate creates a high-energy enol phosphate linkage.



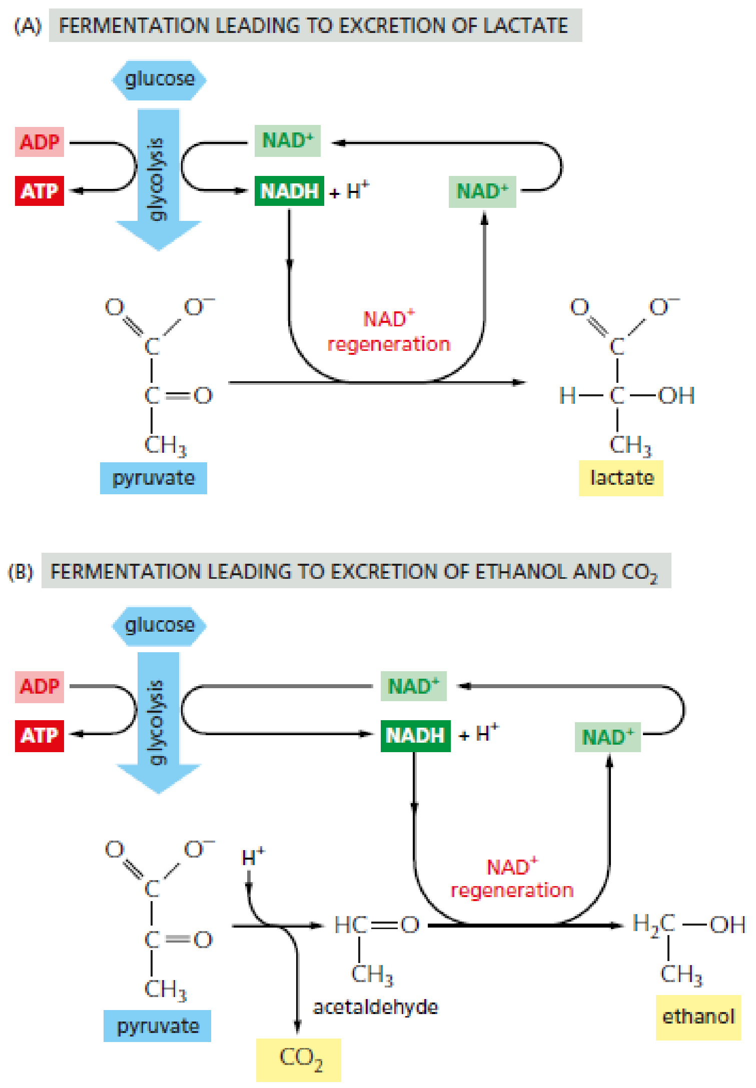
Step 10 The transfer to ADP of the high-energy phosphate group that was generated in step 9 forms ATP, completing glycolysis.



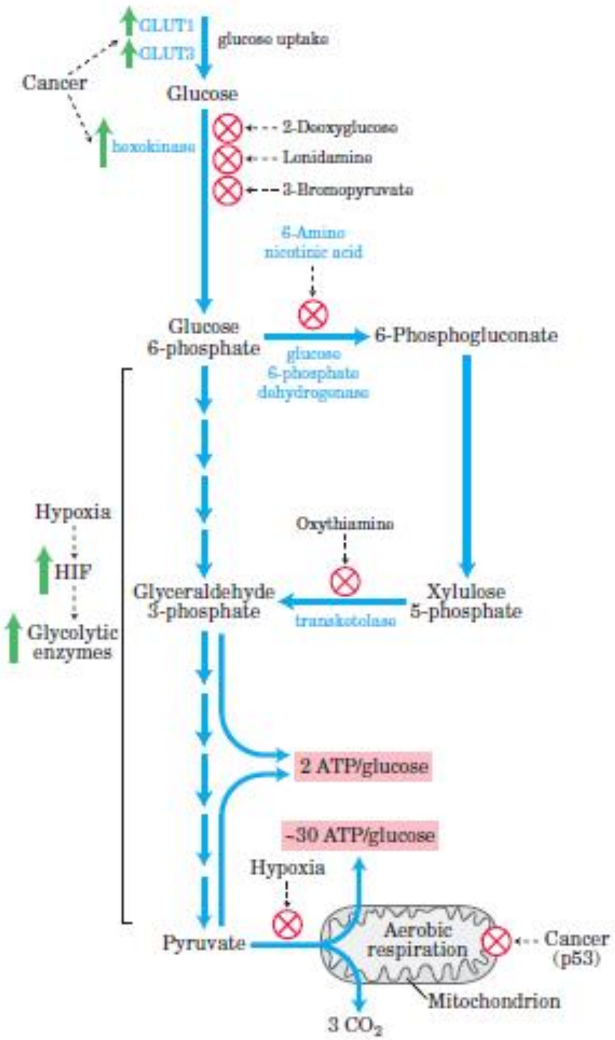
Energy storage in steps 6 and 7 of glycolysis



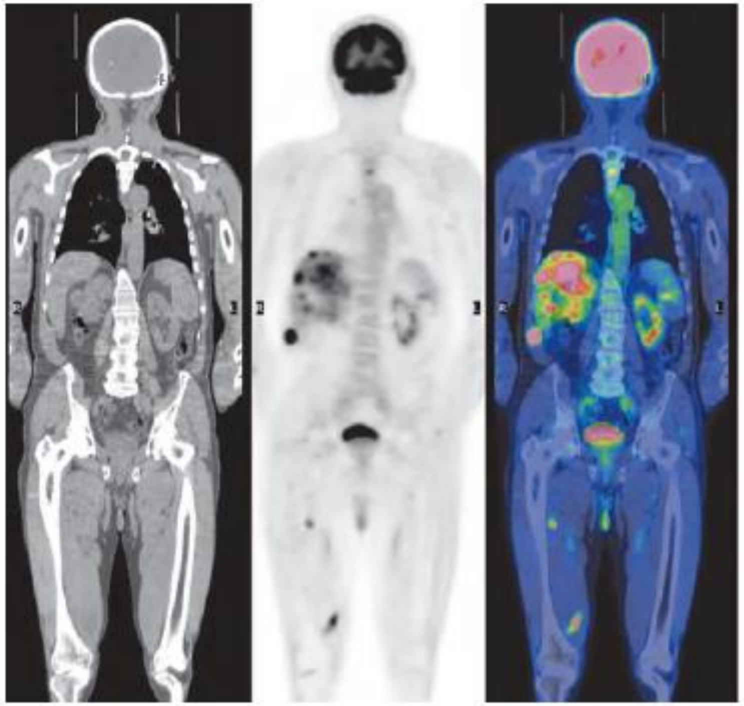
Anaerobic breakdown of pyruvate



The anaerobic metabolism of glucose in tumor cells

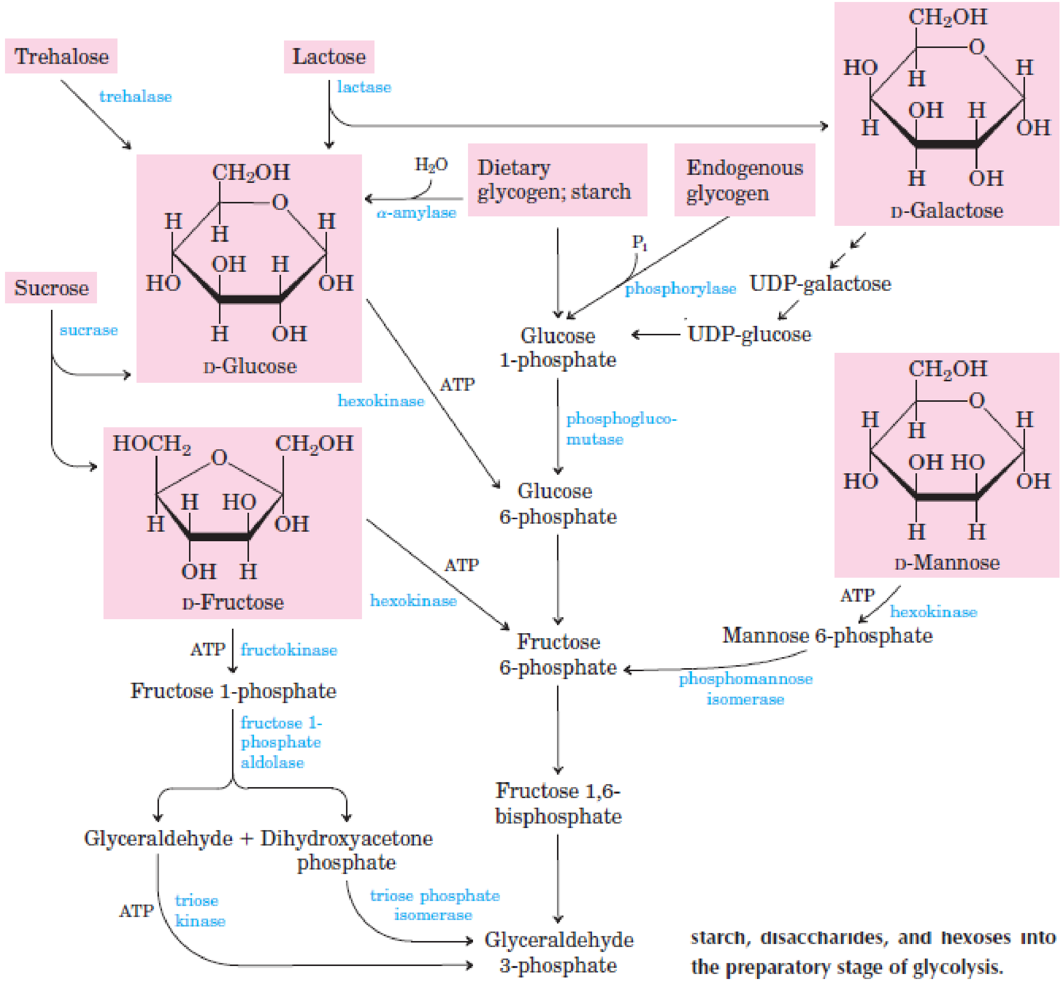


High glycolytic rate in tumor cells also has diagnostic usefulness.



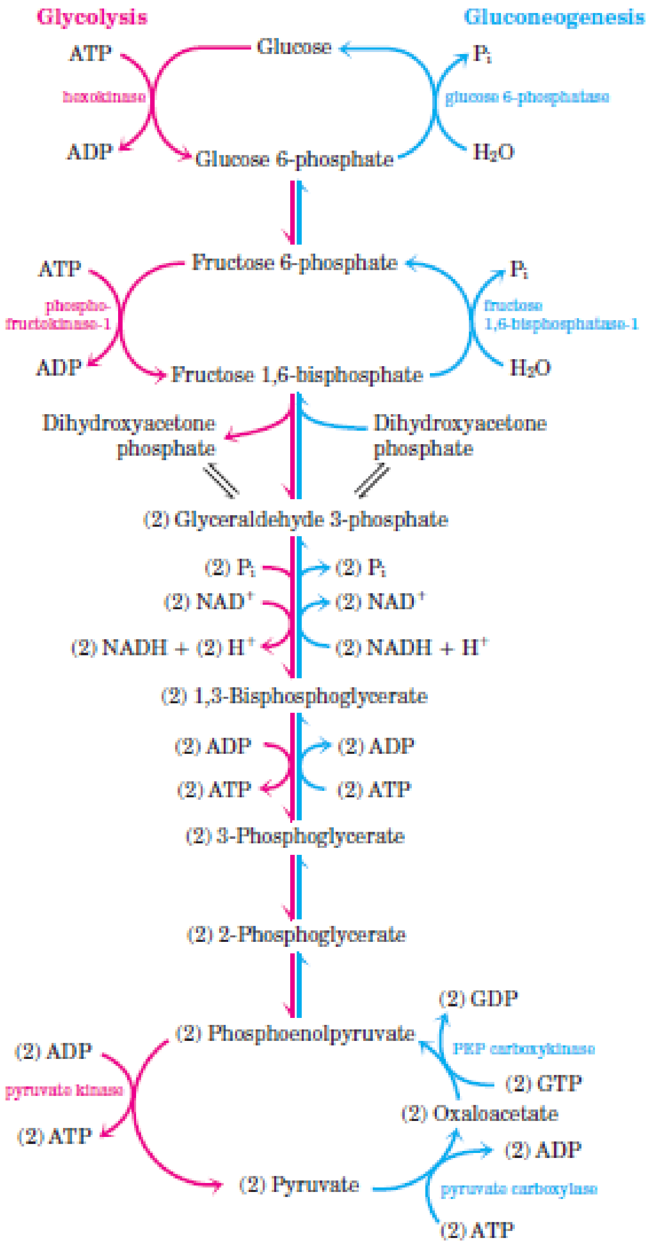
Positron emission tomography (PET) using isotopically Labeled glucose analog

Entry of glycogen, starch, disaccharides, and hexoses into the preparatory stage of glycolysis



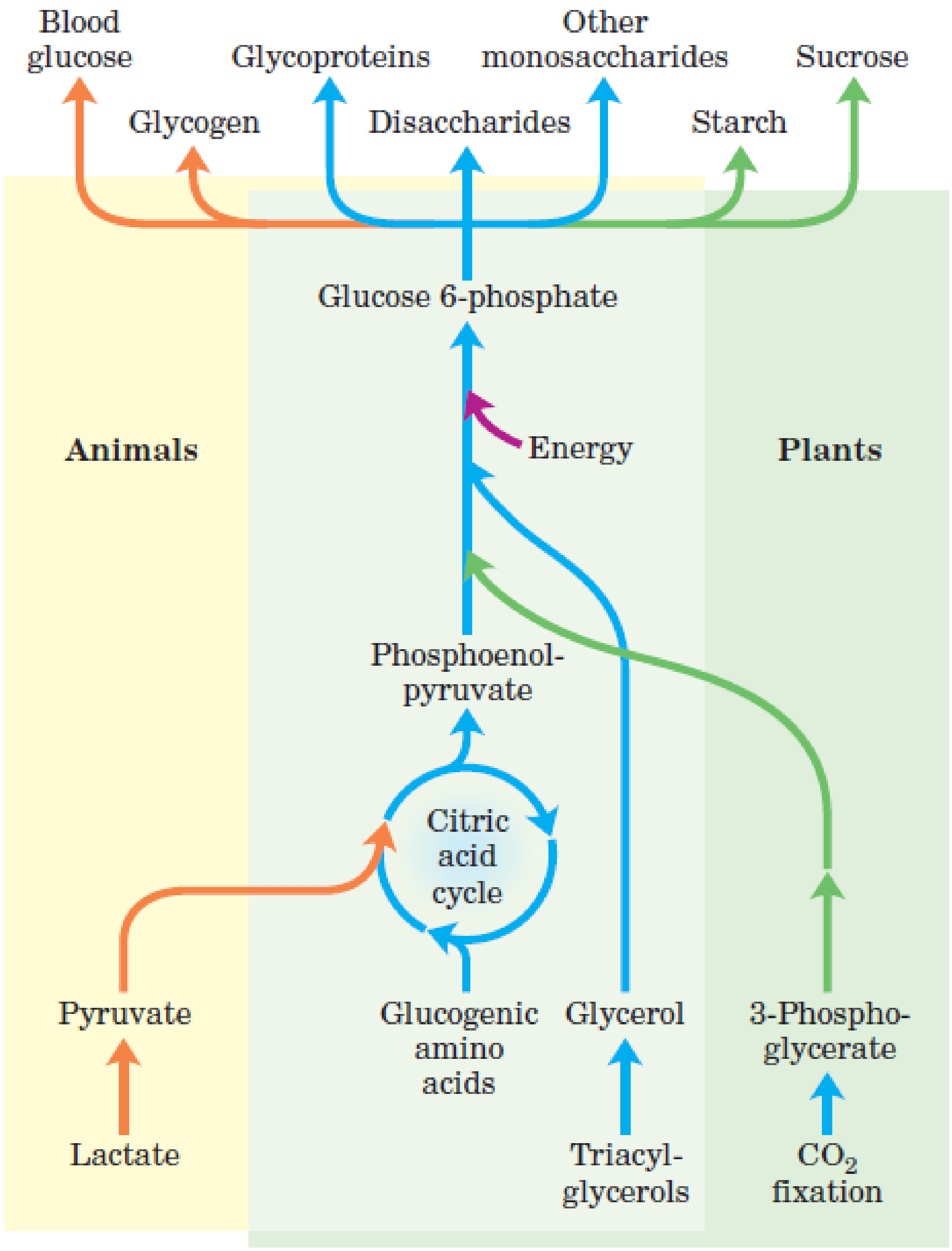
starch, disaccharides, and hexoses into the preparatory stage of glycolysis.

Opposing pathways of glycolysis and gluconeogenesis



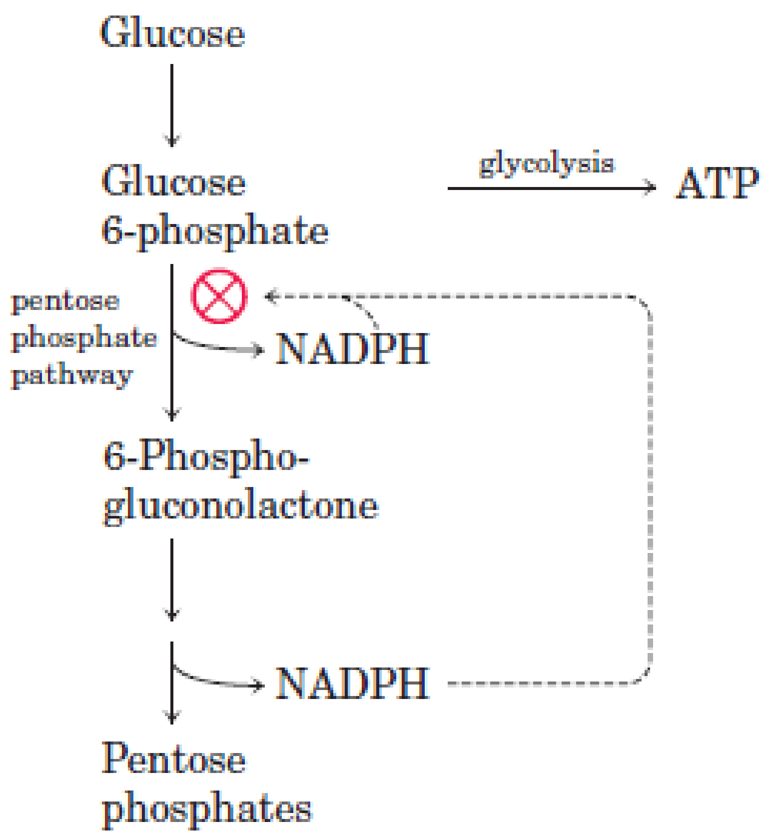
Gluconeogenesis is Energetically Expensive, but Essential

$$2\ \text{Pyruvate} + 4\text{ATP} + 2\text{GTP} + 2\text{NADH} + 2\text{H}^+ + 4\text{H}_2\text{O} \longrightarrow \text{glucose} + 4\text{ADP} + 2\text{GDP} + 6\text{P}_i + 2\text{NAD}^+$$



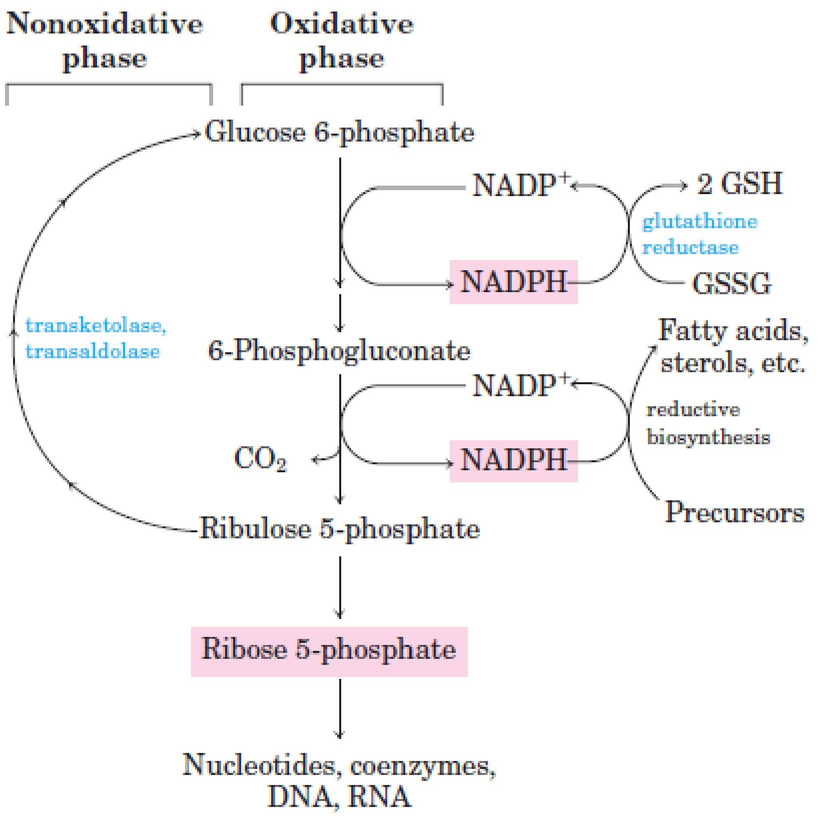
Blood glucose
Glycogen
Glycoproteins
Disaccharides
Other monosaccharides
Starch
Sucrose
Glucose 6-phosphate
Animals
Energy
Plants
Phosphoenol-pyruvate
Citric acid cycle
Pyruvate
Glucogenic amino acids
Glycerol
3-Phospho-glycerate
Lactate
Triacyl-glycerols
CO_2 fixation

Glucose 6-Phosphate Is Partitioned between Glycolysis and the Pentose Phosphate Pathway



Whether glucose 6-phosphate enters glycolysis or the pentose phosphate pathway depends on the current needs of the cell and on the concentration of $NADP^+$ in the cytosol.

Pentose Phosphate Pathway of Glucose Oxidation



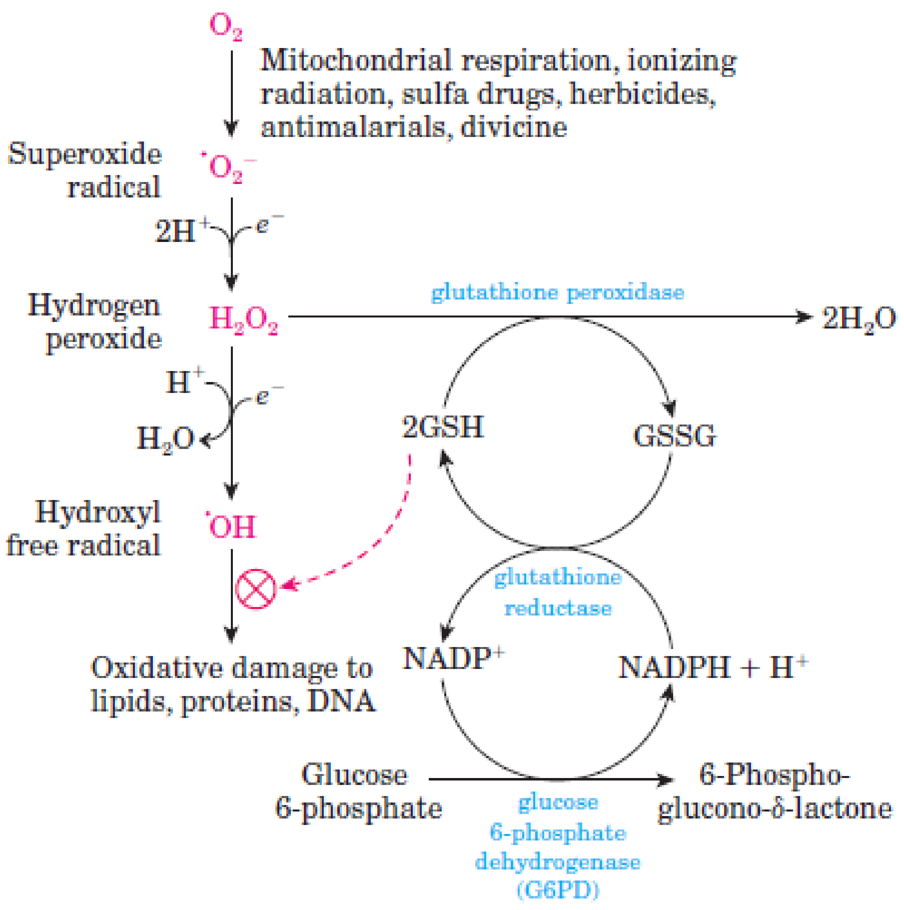
Rapidly dividing cells, such as those of bone marrow, skin, and intestinal mucosa, and those of tumors, use the pentose ribose 5-phosphate to make RNA, DNA, and such coenzymes as ATP, NADH, FADH2, and coenzyme A.

electron donor NADPH, needed for reductive biosynthesis or to counter the damaging effects of oxygen radicals

Tissues that carry out extensive fatty acid synthesis (liver, adipose, lactating mammary gland) or very active synthesis of cholesterol and steroid hormones (liver, adrenal glands, gonads) require the NADPH provided by this pathway.

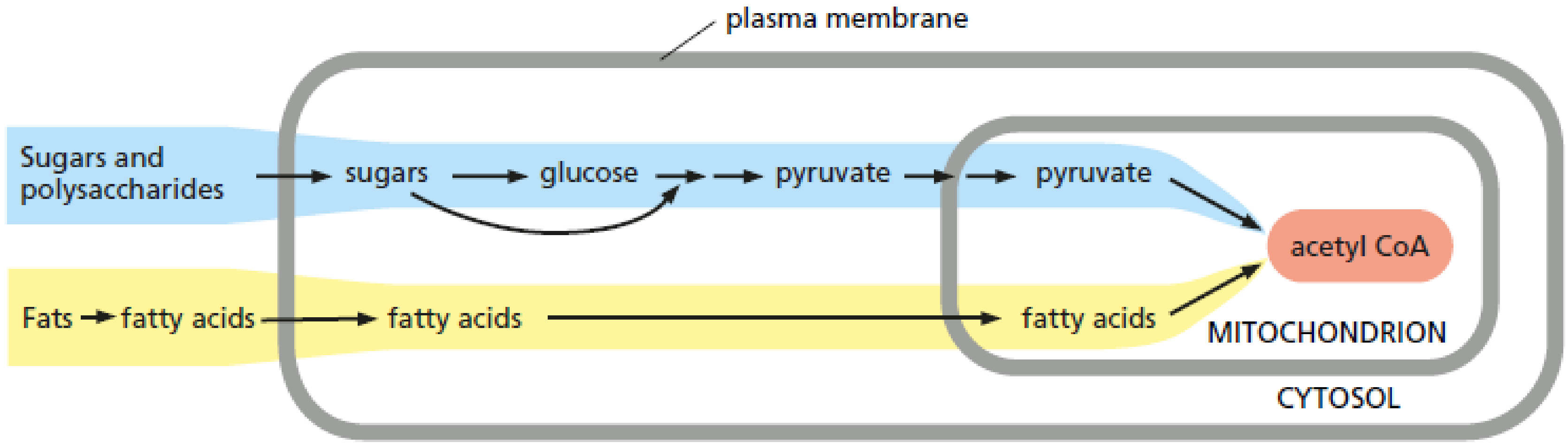
Also known as phosphogluconate pathway, or hexose monophosphate pathway

Role of NADPH and glutathione in protecting cells against highly reactive oxygen derivatives

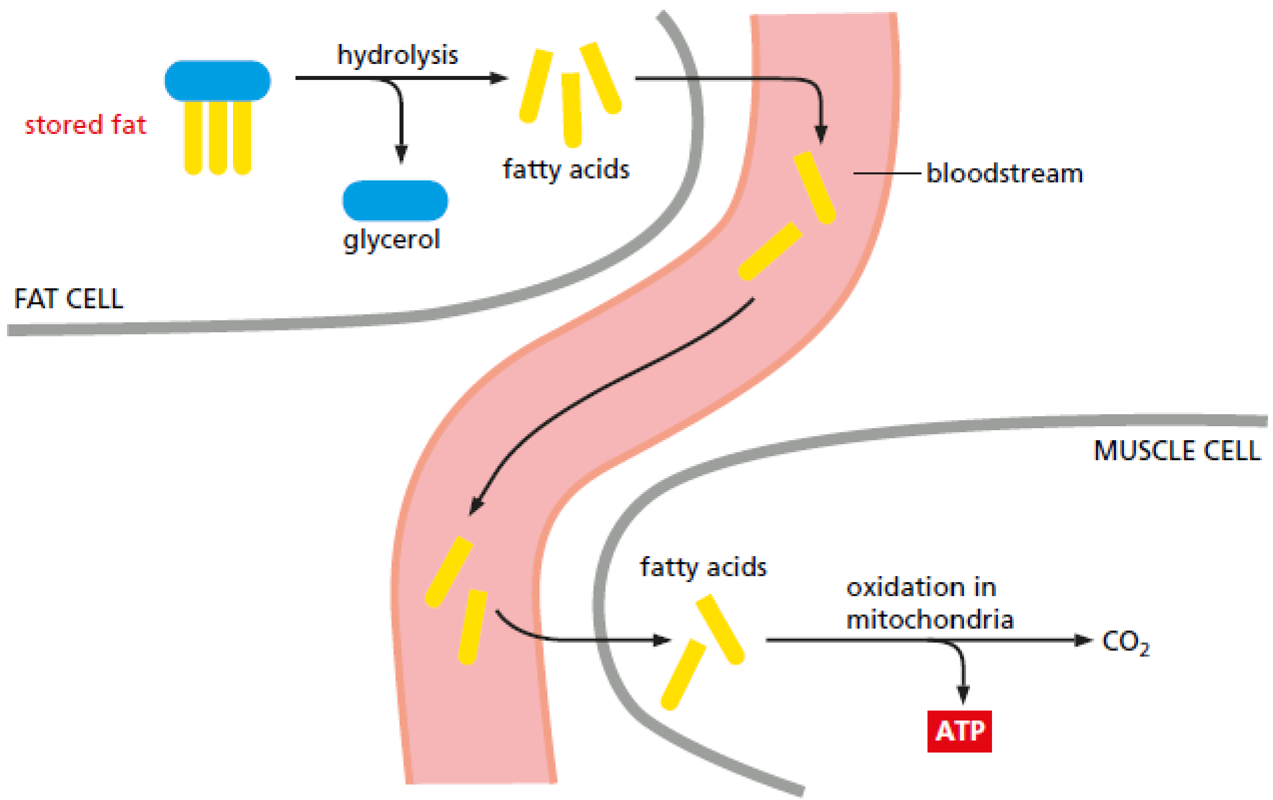


H_2O_2 is also broken down to H_2O and O_2 by catalase, which also requires NADPH

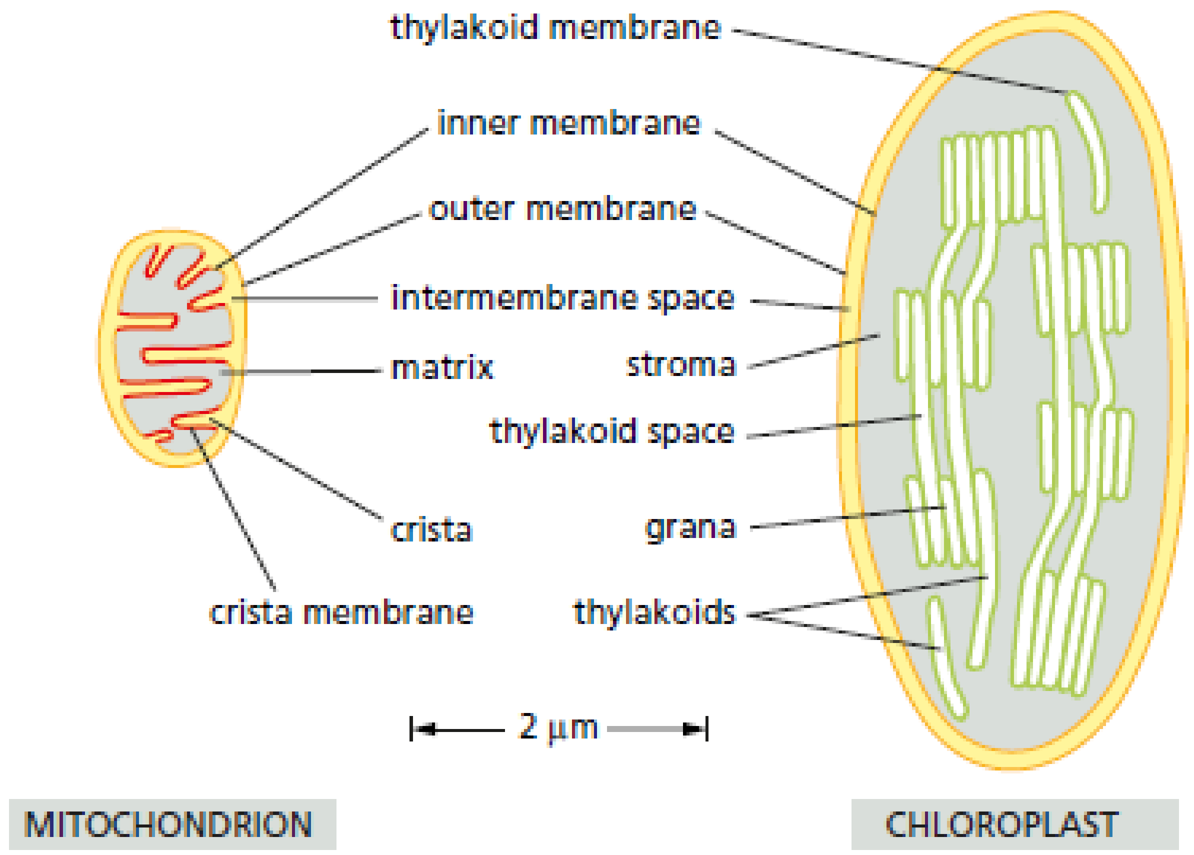
Sugars and Fats Are Both Degraded to Acetyl CoA in Mitochondria



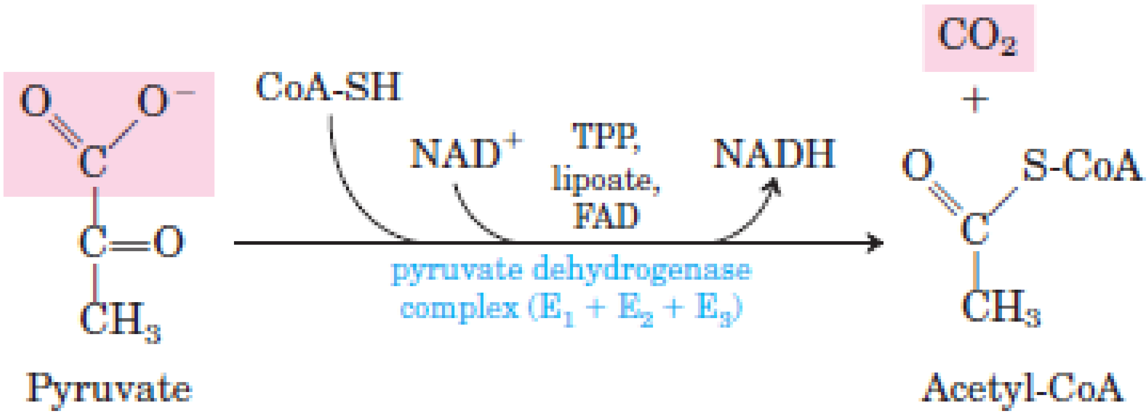
How stored fats are mobilized for energy production in animals



CHLOROPLASTS AND PHOTOSYNTHESIS



Pyruvate is Oxidized to Acetyl-CoA and CO_2



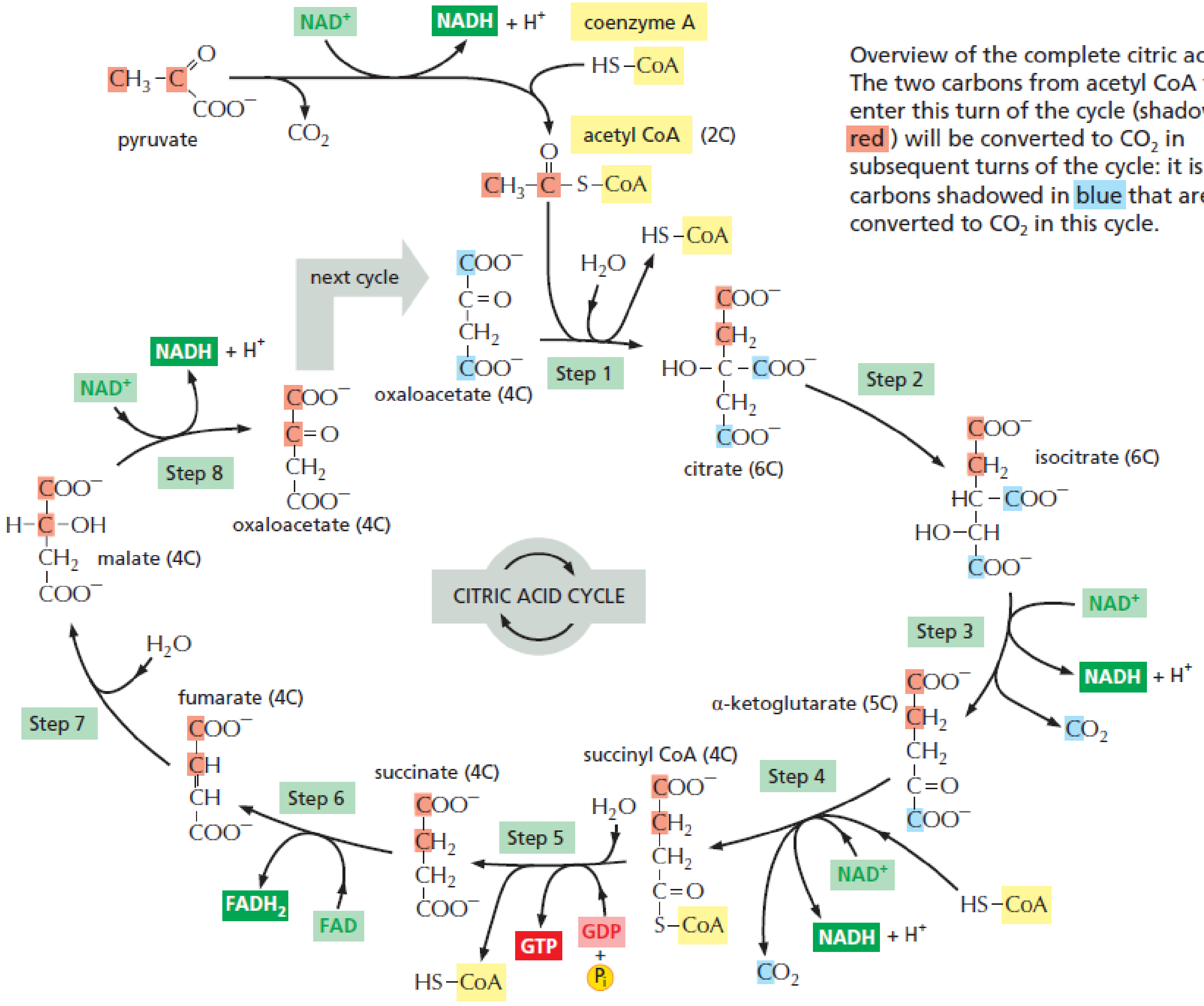
The PDH complex contains three enzymes—**pyruvate dehydrogenase** (E1), **dihydrolipoyl transacetylase** (E2), and **dihydrolipoyl dehydrogenase** (E3)—each present in multiple copies.

Thiamine pyrophosphate (TPP), flavin adenine dinucleotide (FAD), coenzyme A (CoA, sometimes denoted CoA-SH, to emphasize the role of the —SH group), nicotinamide adenine dinucleotide (NAD), and lipoate.

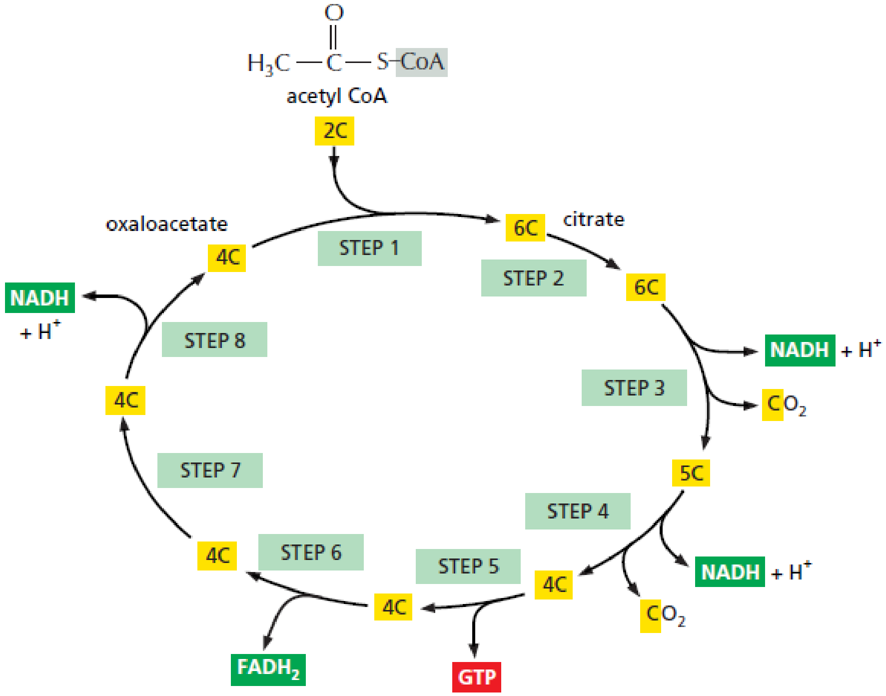
Four different vitamins required in human nutrition are vital components of this system:

Thiamine (in TPP),
riboflavin (in FAD),
niacin (in NAD), and
pantothenate
(in CoA).

citric acid cycle

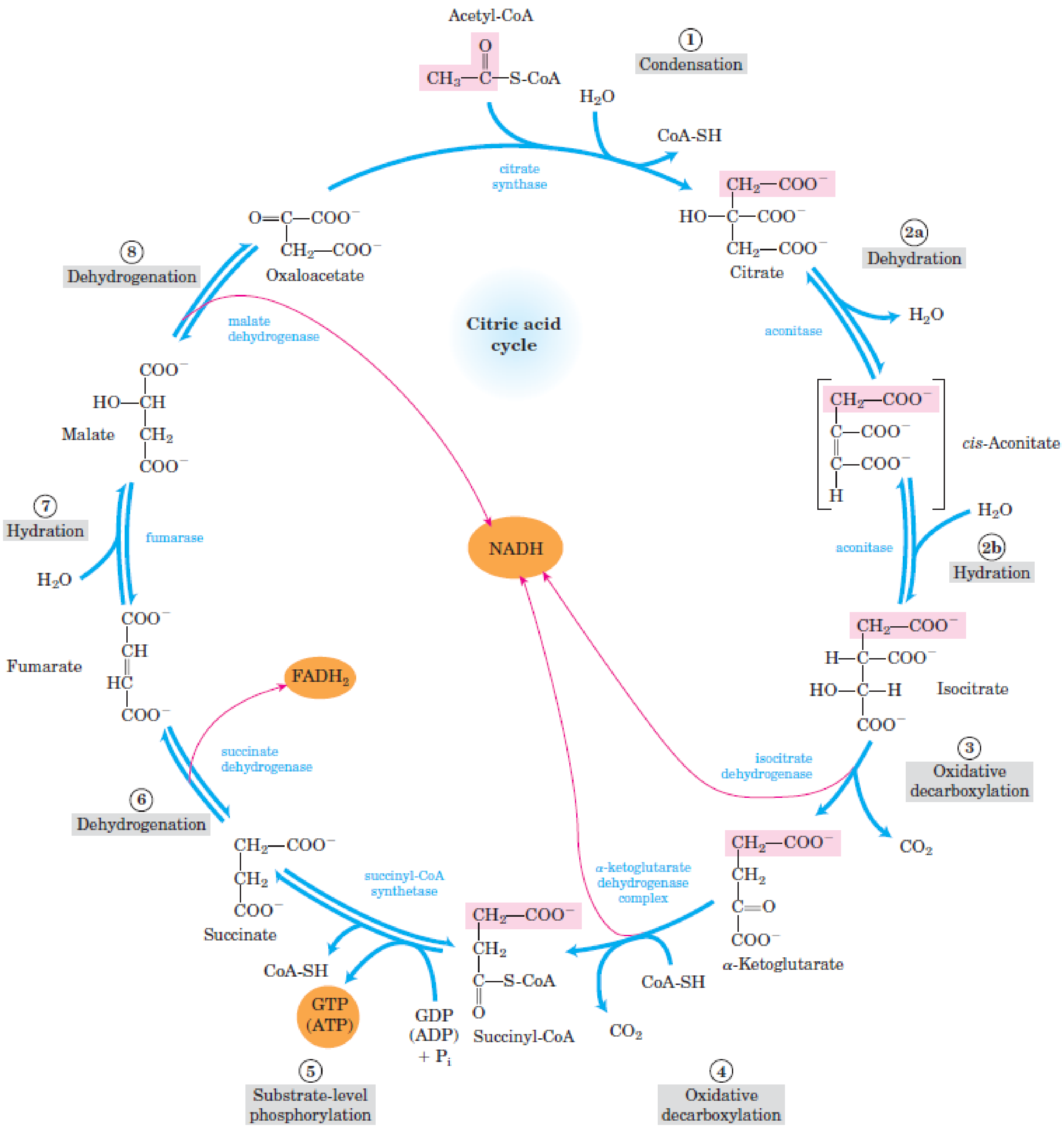


Overview of the complete citric acid cycle. The two carbons from acetyl CoA that enter this turn of the cycle (shadowed in red) will be converted to CO_2 in subsequent turns of the cycle: it is the two carbons shadowed in blue that are converted to CO_2 in this cycle.



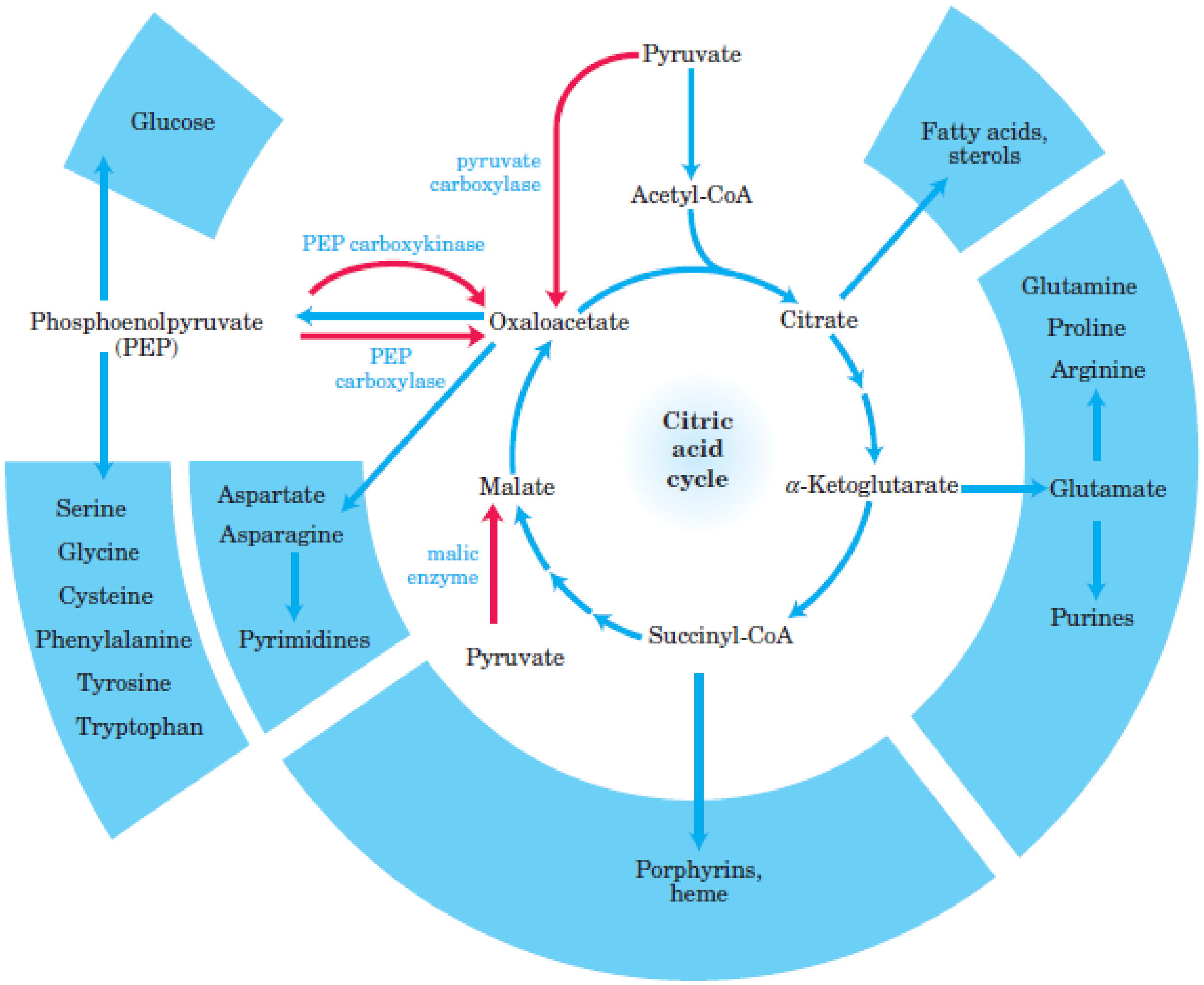
NET RESULT: ONE TURN OF THE CYCLE PRODUCES THREE NADH, ONE GTP, AND ONE $FADH_2$ MOLECULE, AND RELEASES TWO MOLECULES OF CO_2

Hans Krebs, 1900–1981



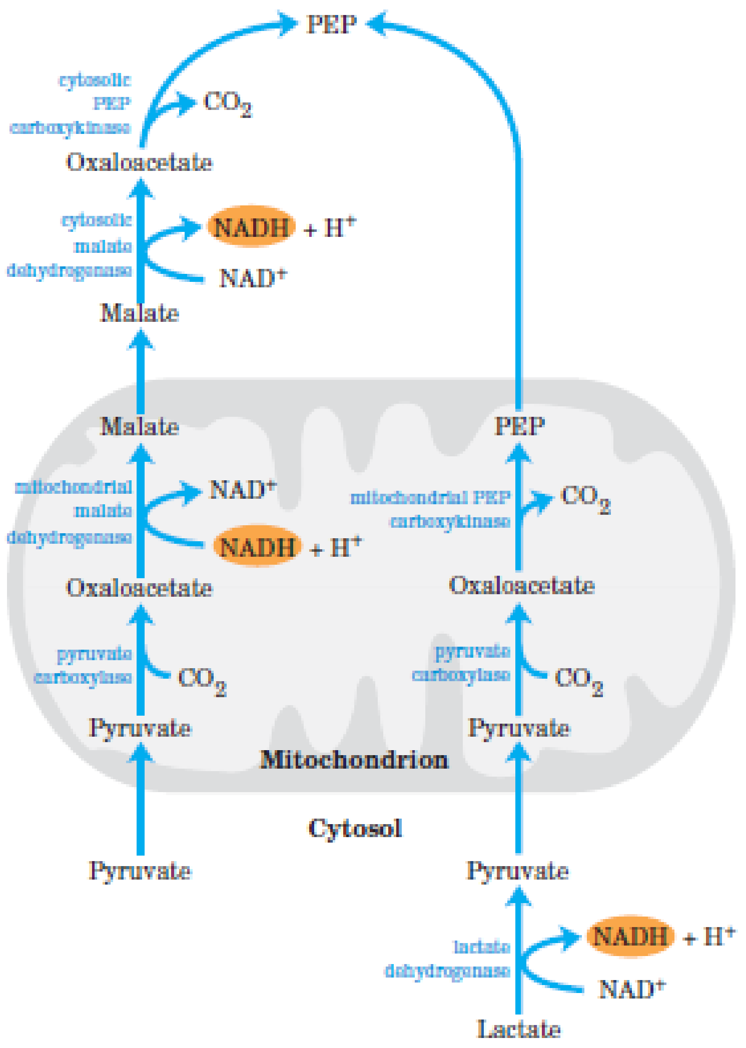
Acetyl-CoA
Condensation
citrate synthase
CoA-SH
Citrate
Dehydration
aconitase
cis-Aconitate
Hydration
Isocitrate
isocitrate dehydrogenase
Oxidative decarboxylation
α-Ketoglutarate
α-ketoglutarate dehydrogenase complex
Oxidative decarboxylation
Succinyl-CoA
succinyl-CoA synthetase
GDP (ADP) + Pi
GTP (ATP)
Substrate-level phosphorylation
Succinate
succinate dehydrogenase
FADH2
Dehydrogenation
Fumarate
fumarase
Hydration
Malate
malate dehydrogenase
Dehydrogenation
Oxaloacetate
Citric acid cycle
NADH

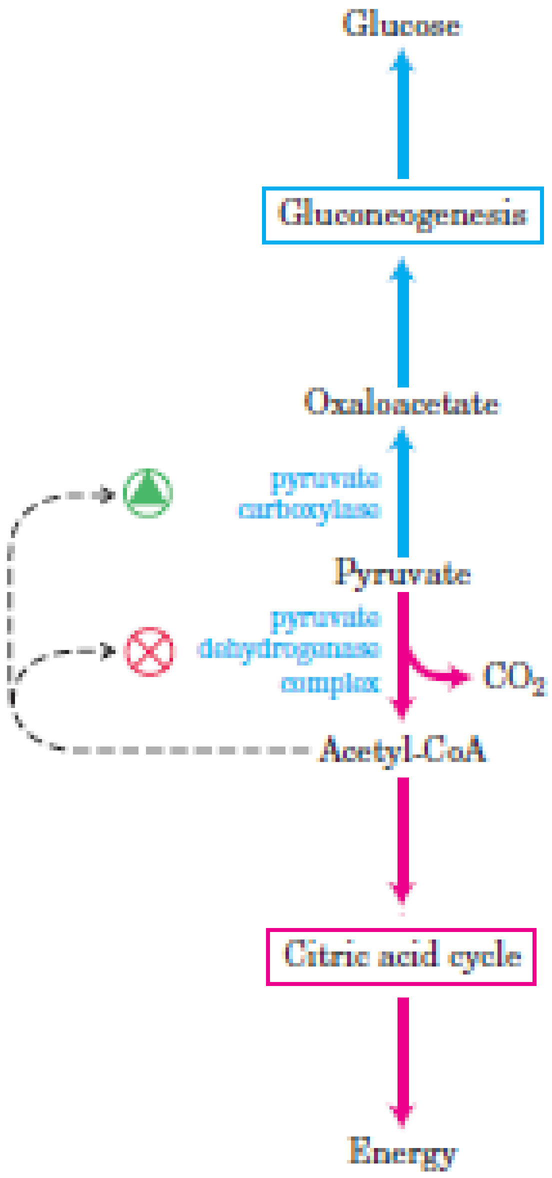
Role of the citric acid cycle in anabolism



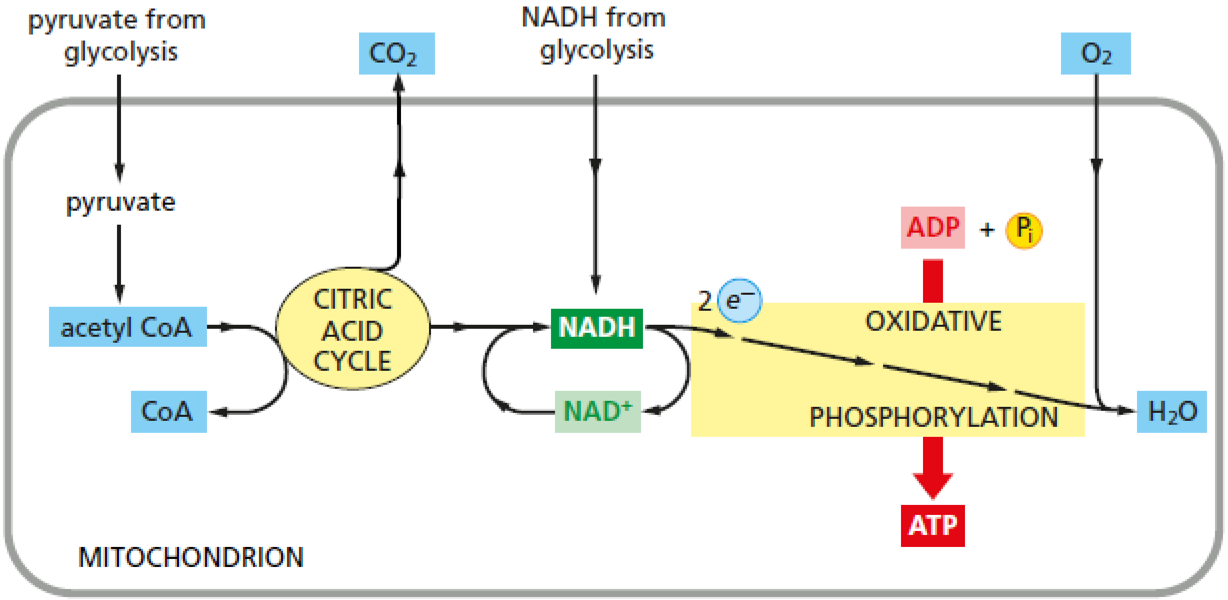
Anaplerotic Reactions Replenish Citric Acid

Alternative paths from pyruvate to phosphoenolpyruvate




Glucose
Gluconeogenesis
Oxaloacetate
pyruvate carboxylase
Pyruvate
pyruvate dehydrogenase complex
CO2
Acetyl-CoA
Citric acid cycle
Energy

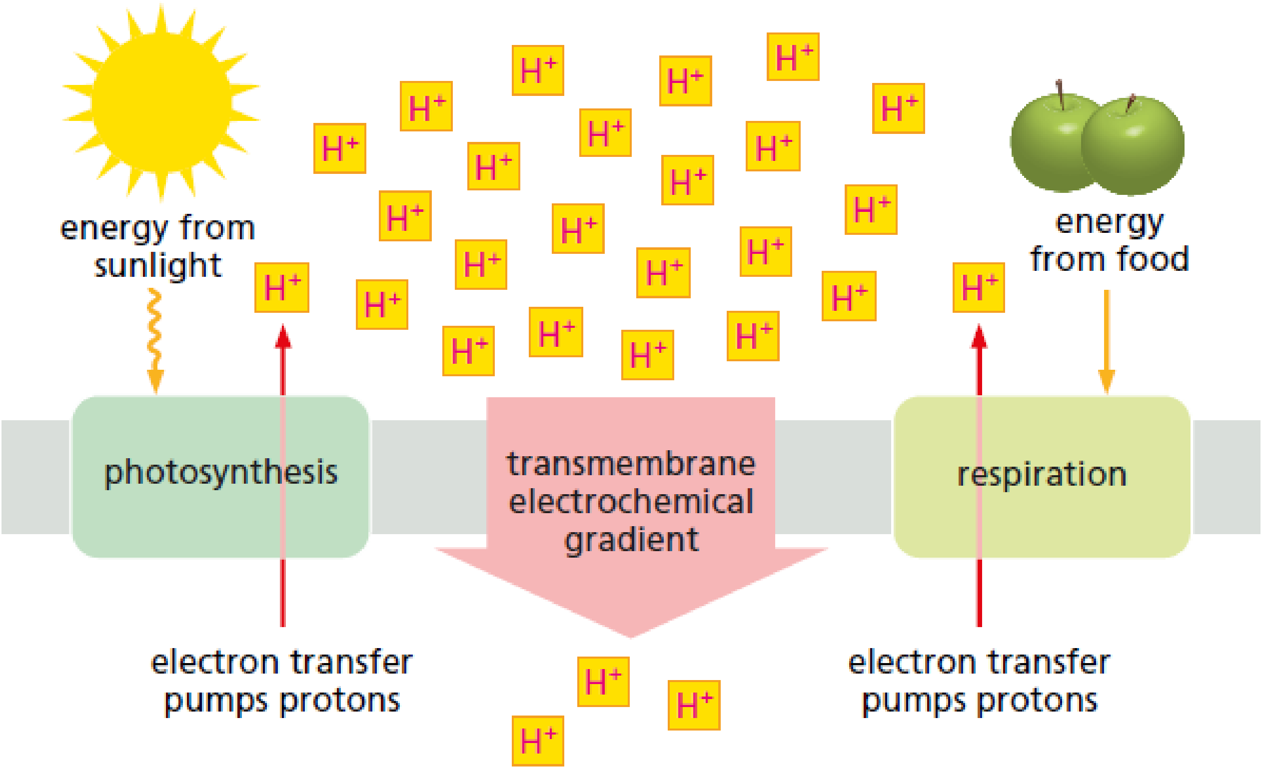
The final stages of oxidation of food molecules



Electron Transport Drives the Synthesis of the Majority of the ATP in Most Cells

High-energy electrons from NADH are passed to a membrane-bound electron-transport chain

Membrane-based mechanisms use the energy provided by food or sunlight to generate ATP

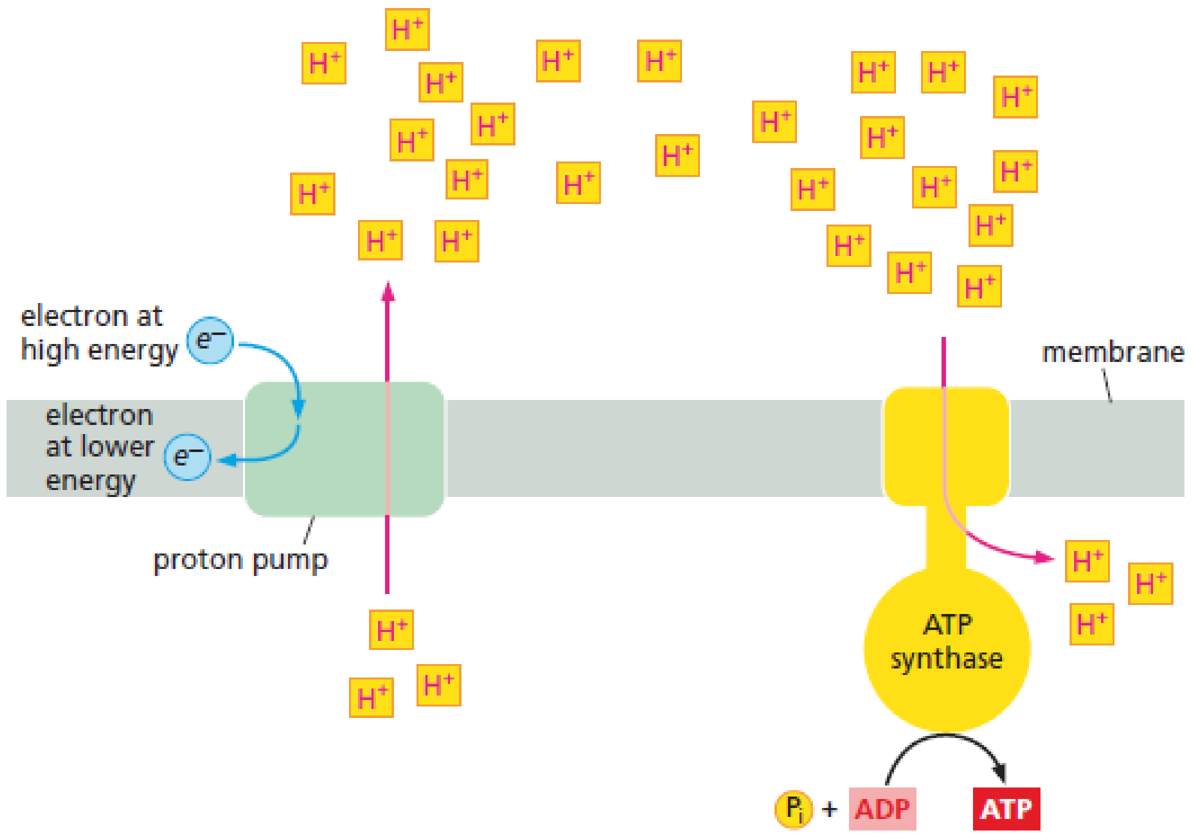


Electron transfers release energy that is used to pump protons

Generate an electrochemical proton gradient

An ion gradient across a membrane is a form of stored energy that can be harnessed to do useful work

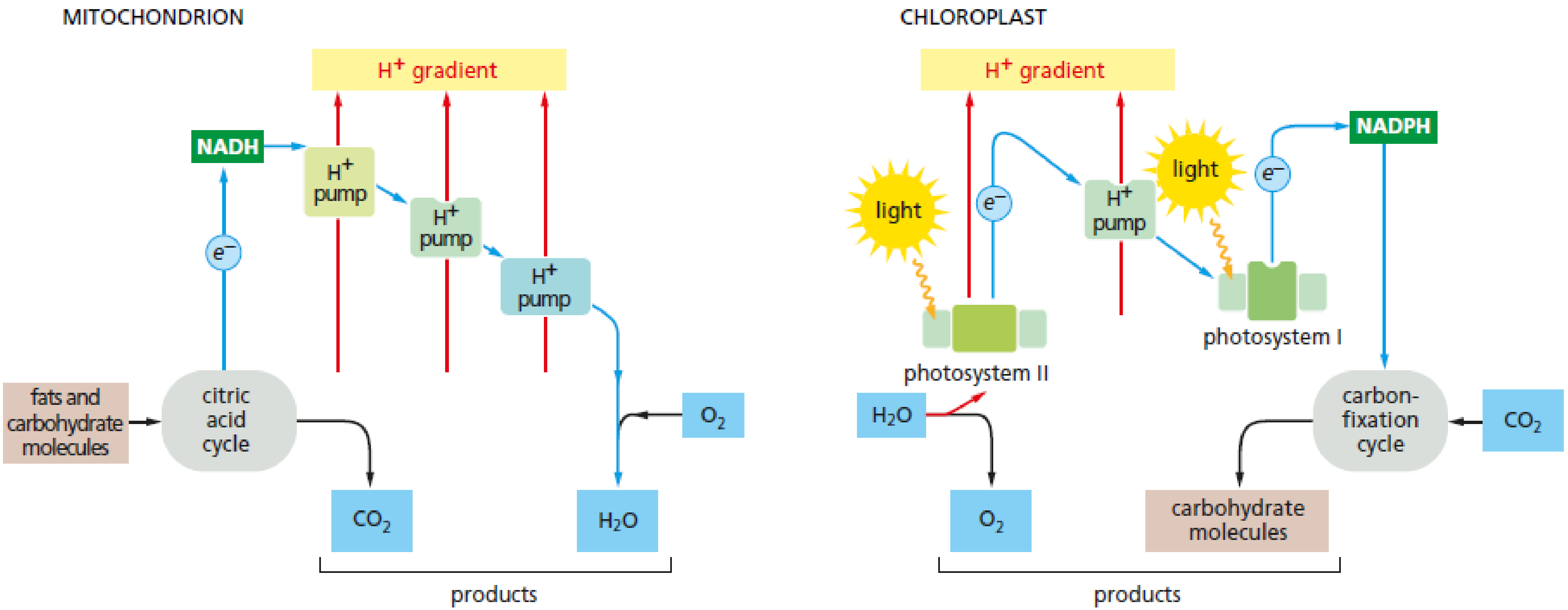
Membrane-based systems use the energy stored in an electrochemical proton gradient to synthesize ATP



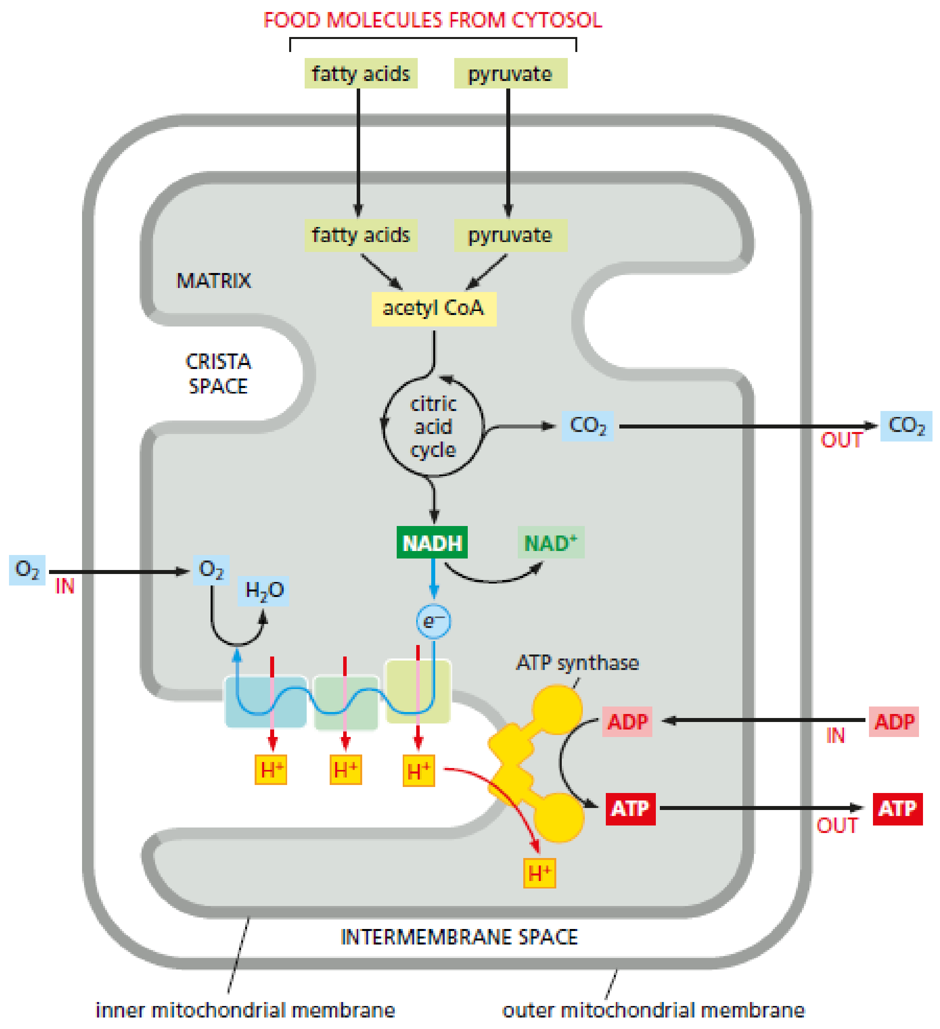
STAGE 1: ENERGY OF ELECTRON TRANSPORT IS USED TO PUMP PROTONS ACROSS MEMBRANE

STAGE 2: ENERGY IN THE PROTON GRADIENT IS HARNESSED BY ATP SYNTHASE TO MAKE ATP

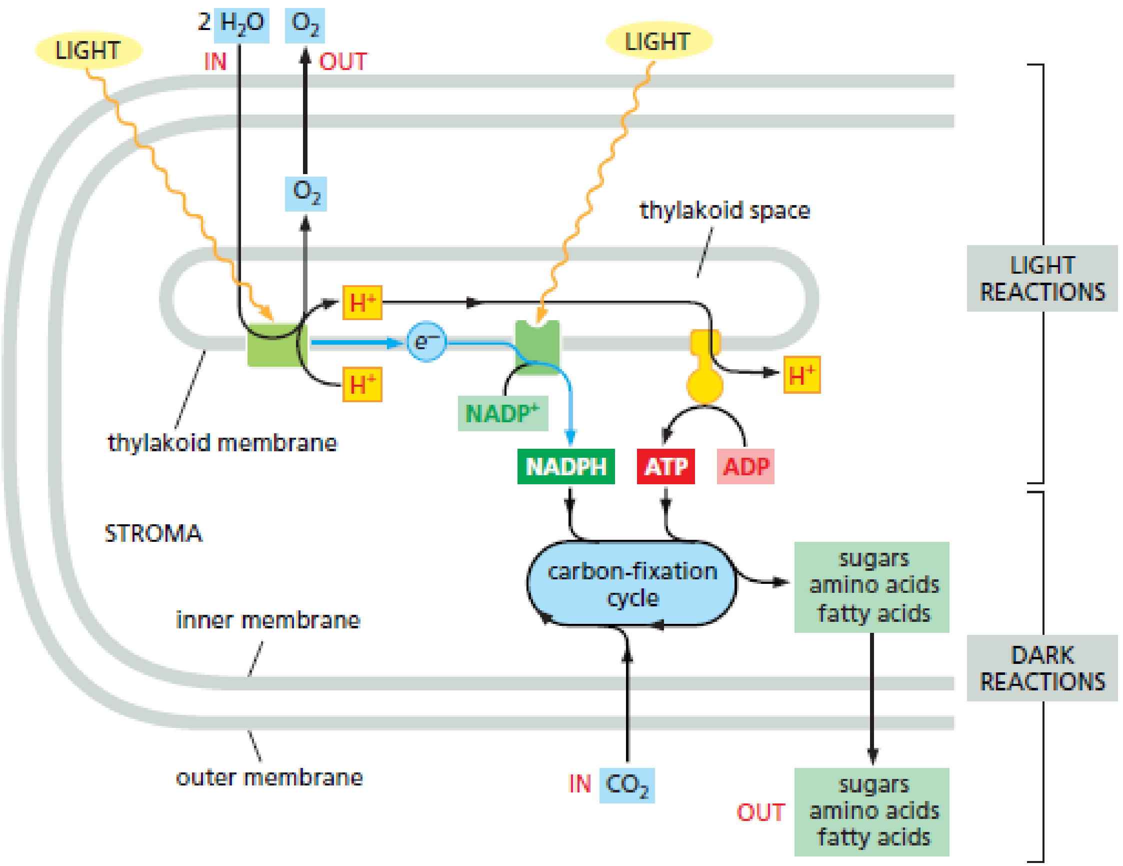
Electron-transport processes



The final stages of oxidation of food molecules

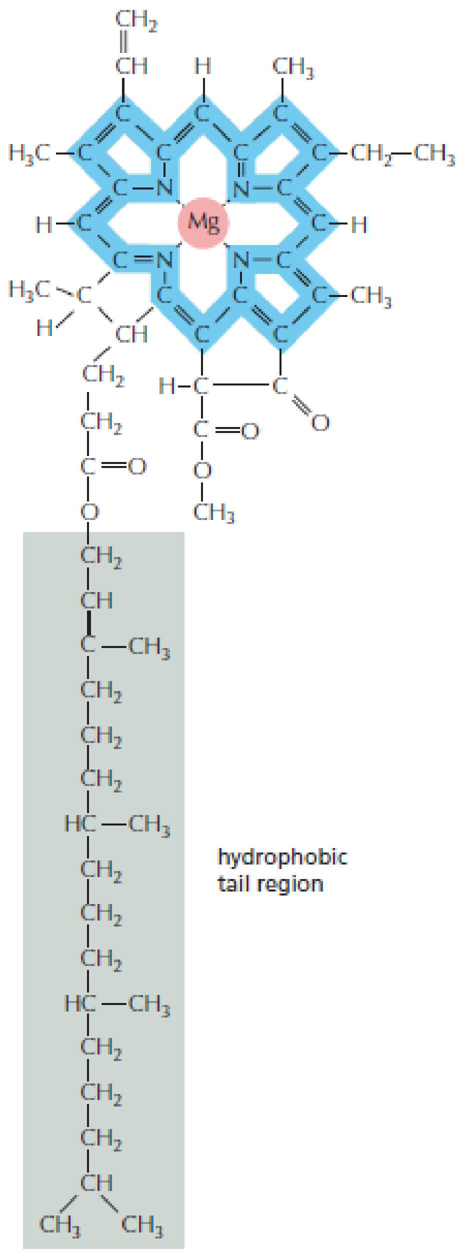


Both stages of photosynthesis depend on the chloroplast.

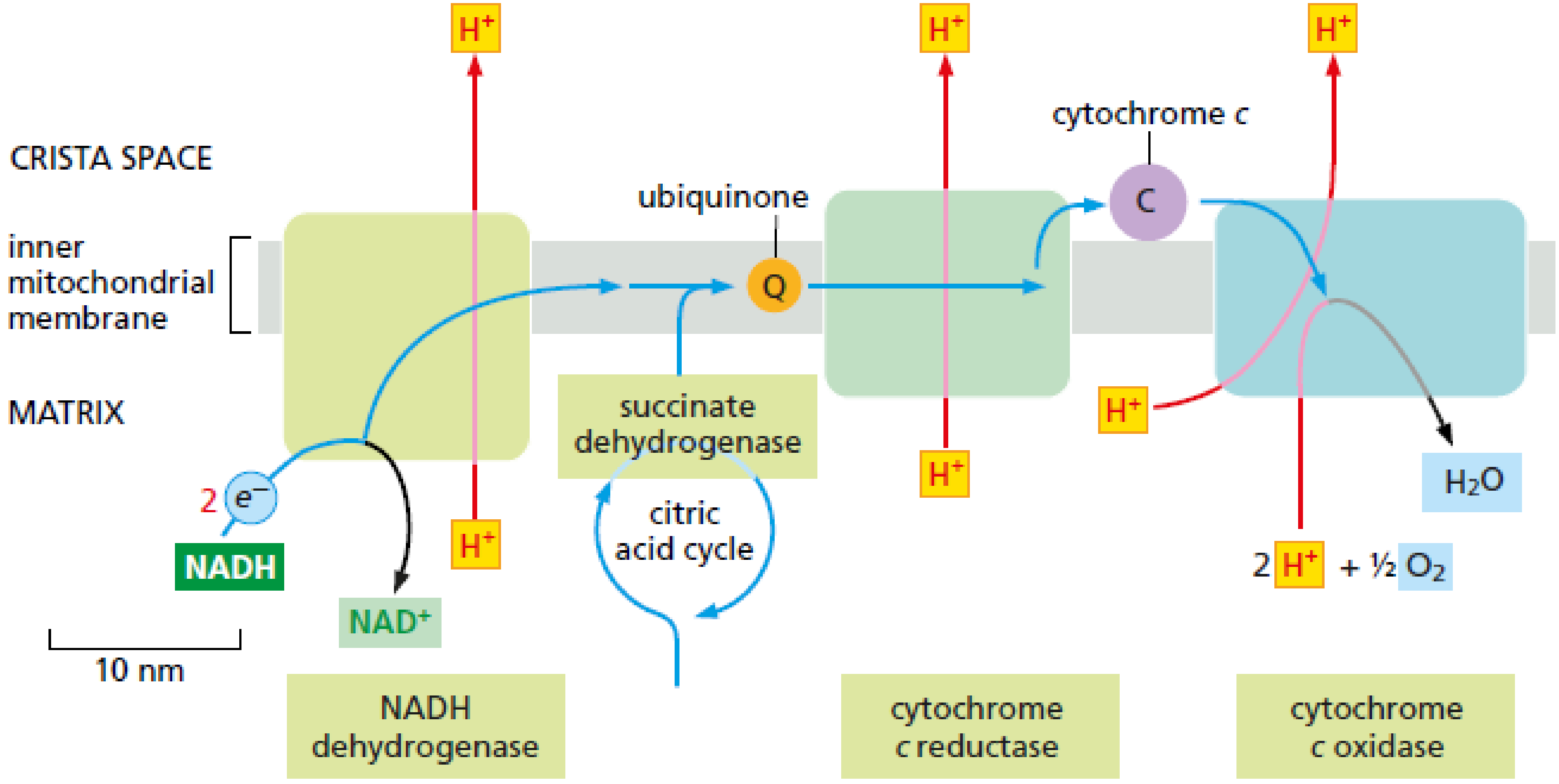


1. Chlorophyll's structure allows it to absorb energy from light. Each chlorophyll molecule contains a porphyrin ring with a magnesium atom (*pink*) at its center. This porphyrin ring is structurally similar to the one that binds iron in heme. Light is absorbed by electrons within the bond network shown in *blue*, while the long, hydrophobic tail (*gray*) helps hold the chlorophyll in the thylakoid membrane.

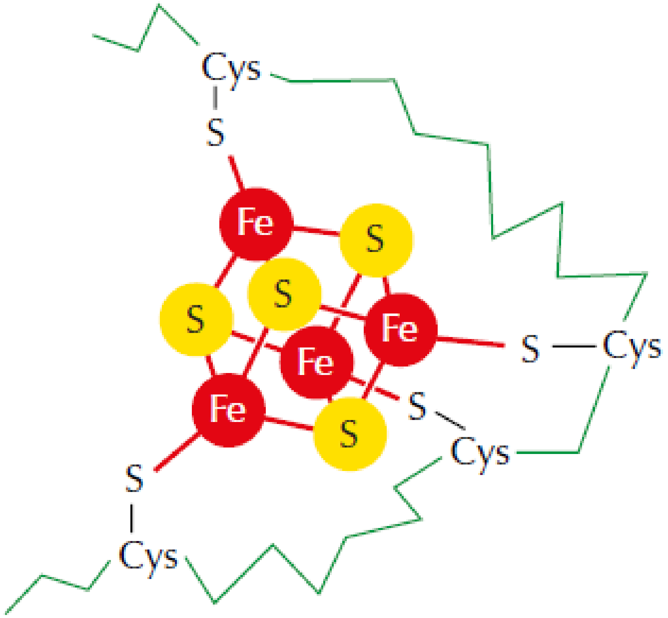
2. A photon (a quantum of light) knocks an electron out of the green pigment molecule *chlorophyll* in the first reaction center, creating a positively charged chlorophyll ion. This electron then moves along an electron-transport chain and through a second reaction center in much the same way that an electron moves along the respiratory chain in mitochondria.



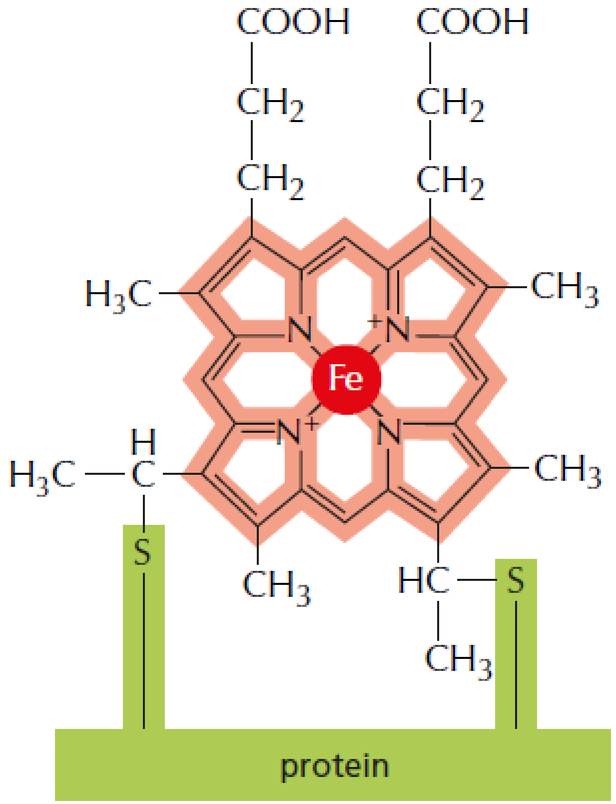
High-energy electrons are transferred through three respiratory enzyme complexes in the inner mitochondrial membrane



Transition Metal Ions and Quinones Accept and Release Electrons Readily

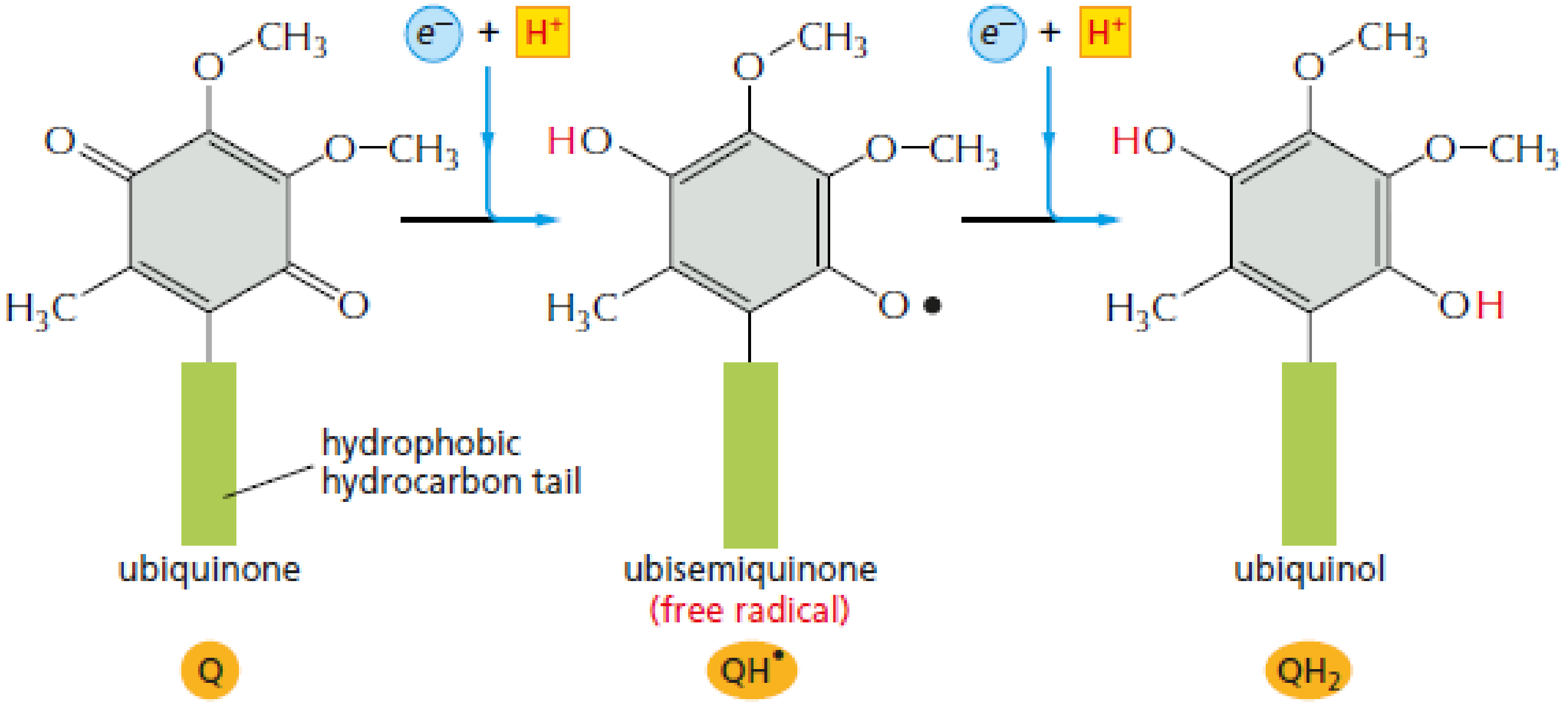


The structure of an iron–sulfur cluster

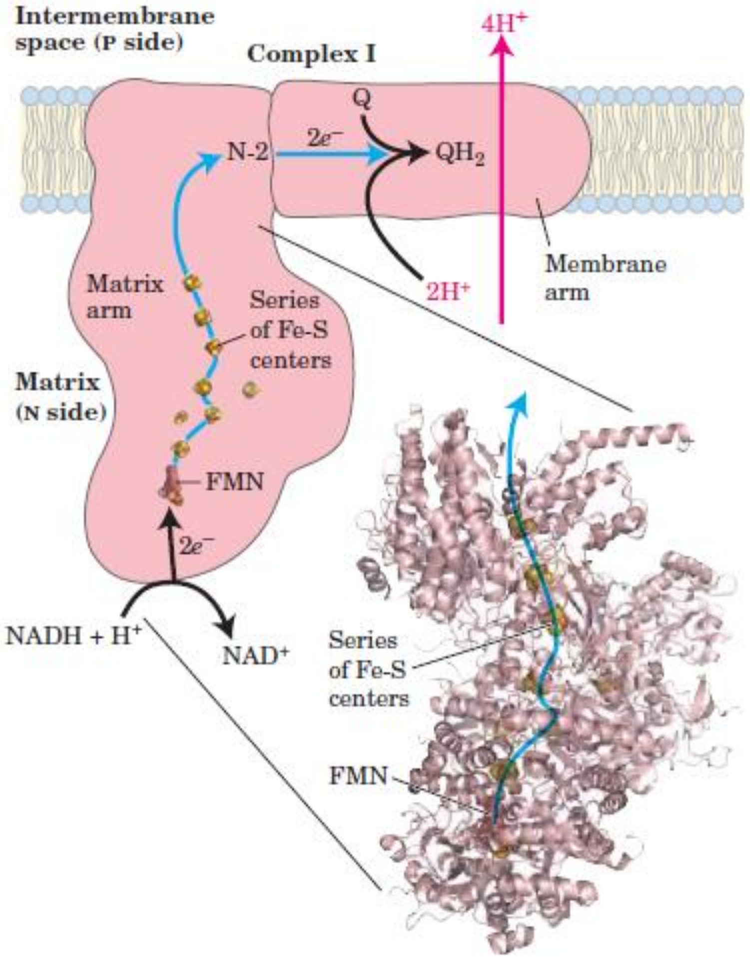


Heme group, in which an iron atom is tightly held by four nitrogen atoms at the corners of a square in a *porphyrin ring*

Quinone electron carriers



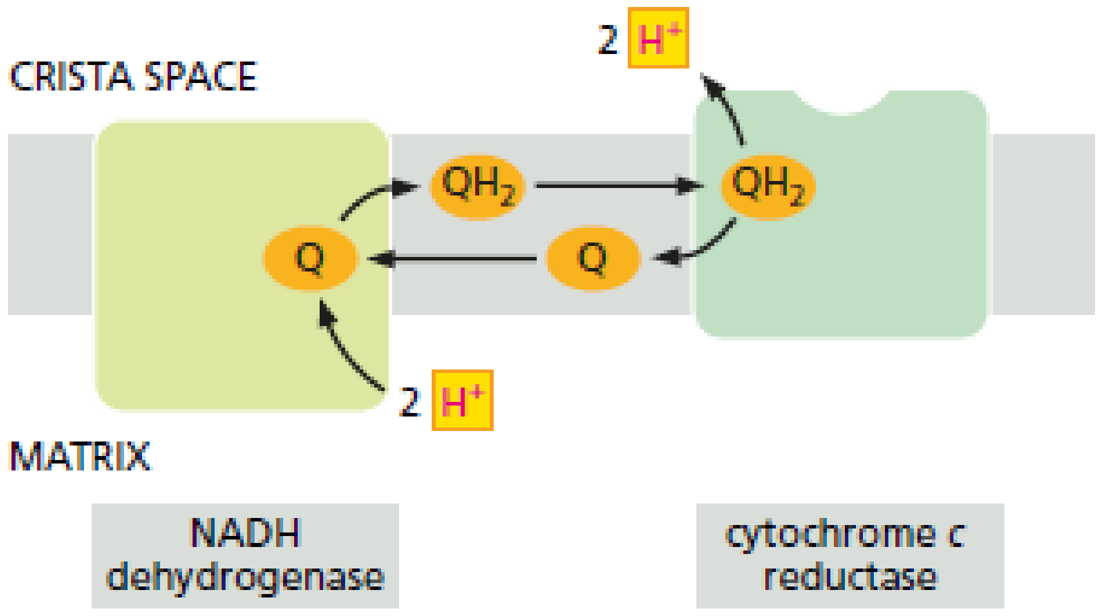
Electron-transport in NADH dehydrogenase (complex 1)



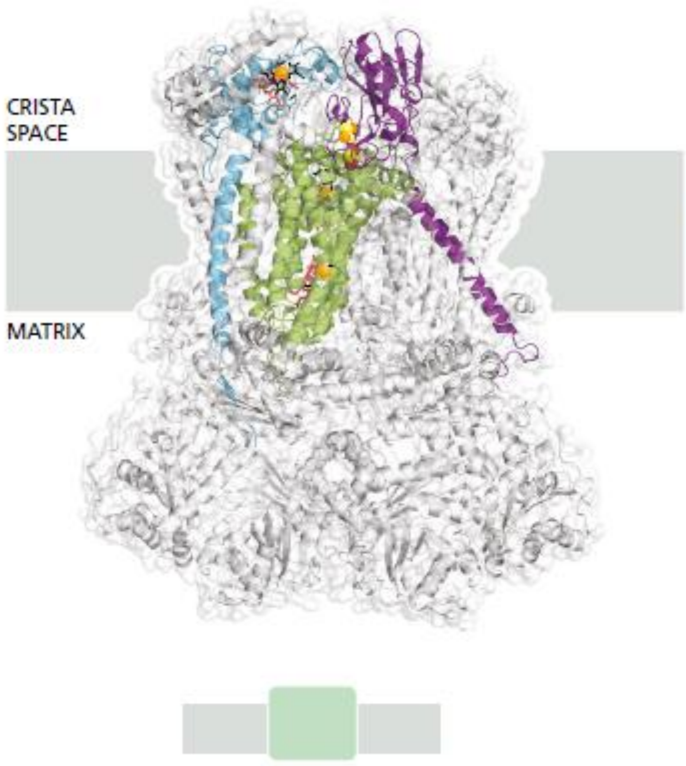
$$NADH + 5H_N^+ + Q \longrightarrow NAD^+ + QH_2 + 4H_P^+$$

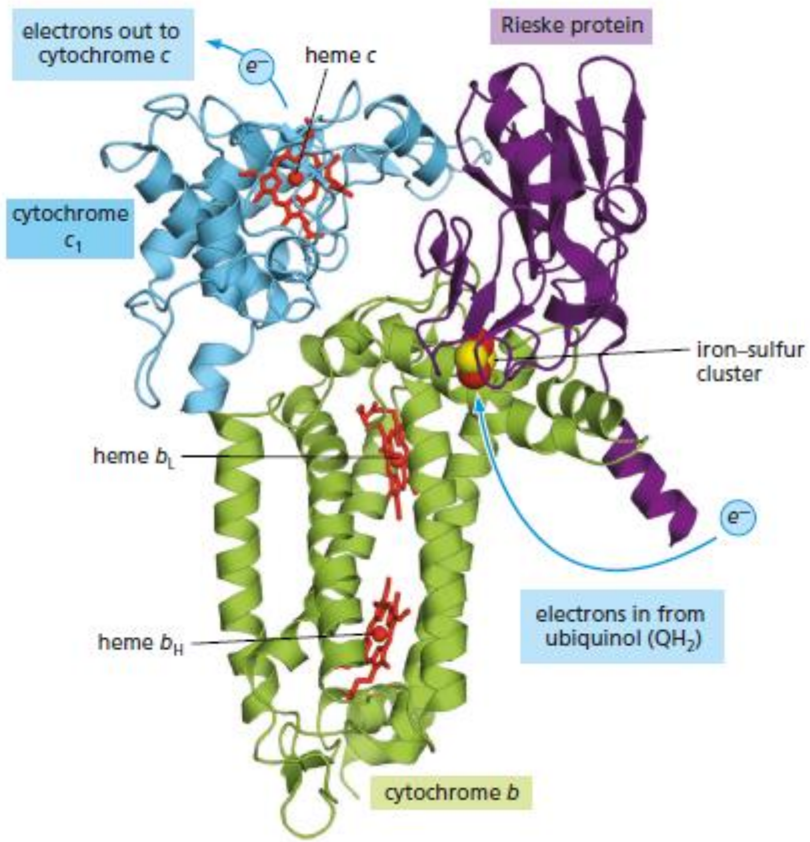
NADH donates two electrons, via a bound flavin mononucleotide (FMN)

How a directional release and uptake of protons by a quinone pumps protons across a membrane

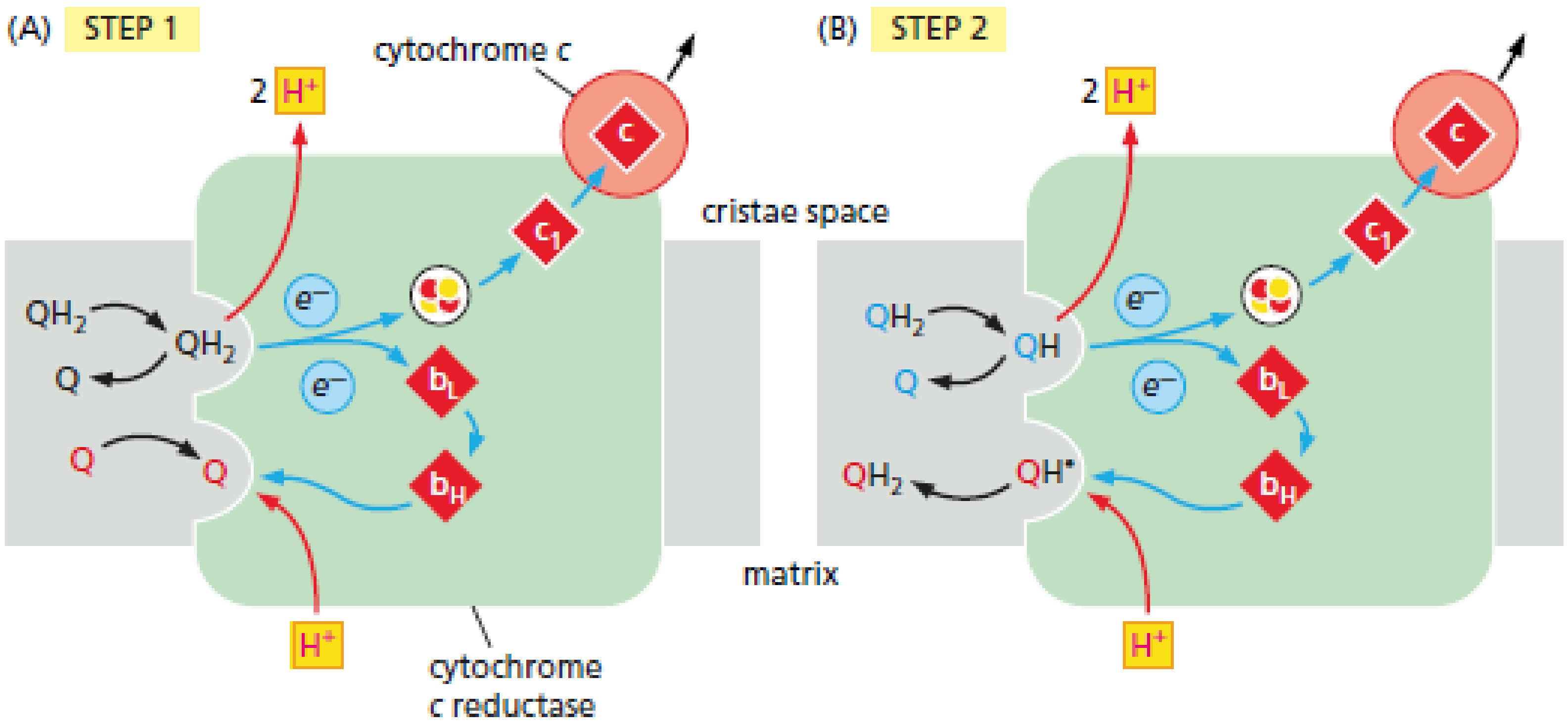


The structure of cytochrome c reductase





The two-step mechanism of the cytochrome c reductase Q-cycle

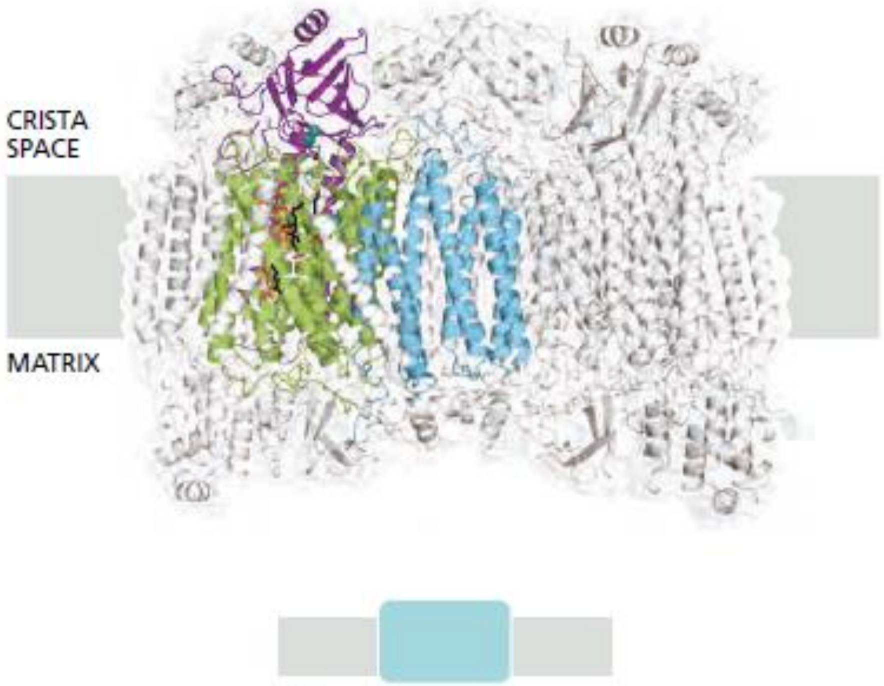


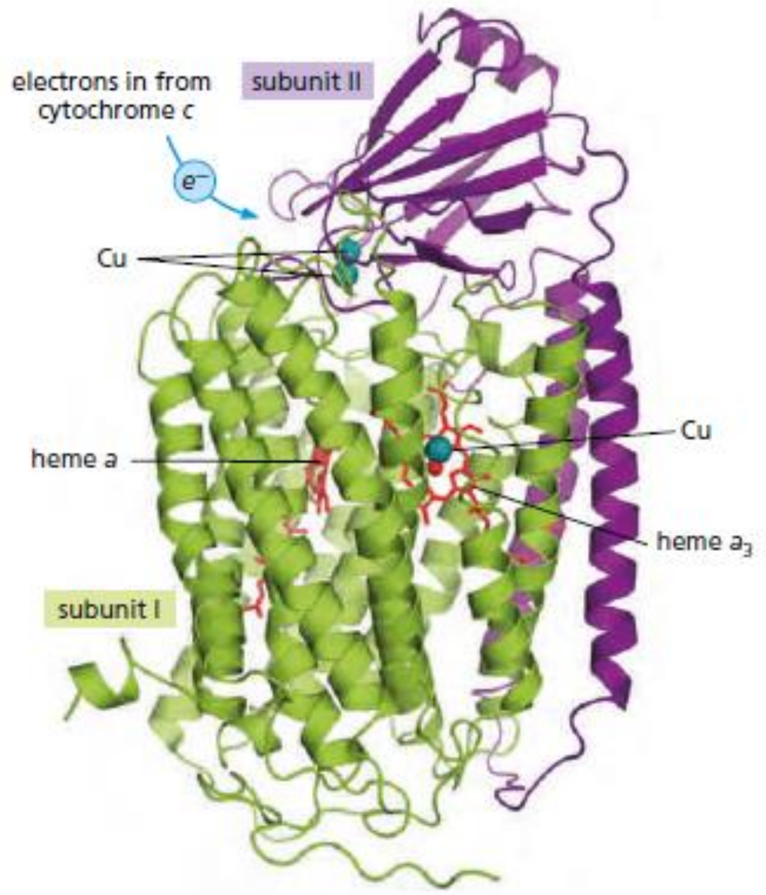
$$QH_2 + Q + \text{cyt } c_1\,(\text{oxidized}) \longrightarrow {}^{\bullet}Q^- + Q + 2H^+_P + \text{cyt } c_1\,(\text{reduced})$$

$$QH_2 + {}^{\bullet}Q^- + 2H^+_N + \text{cyt } c_1\,(\text{oxidized}) \longrightarrow QH_2 + 2H^+_P + Q + \text{cyt } c_1\,(\text{reduced})$$

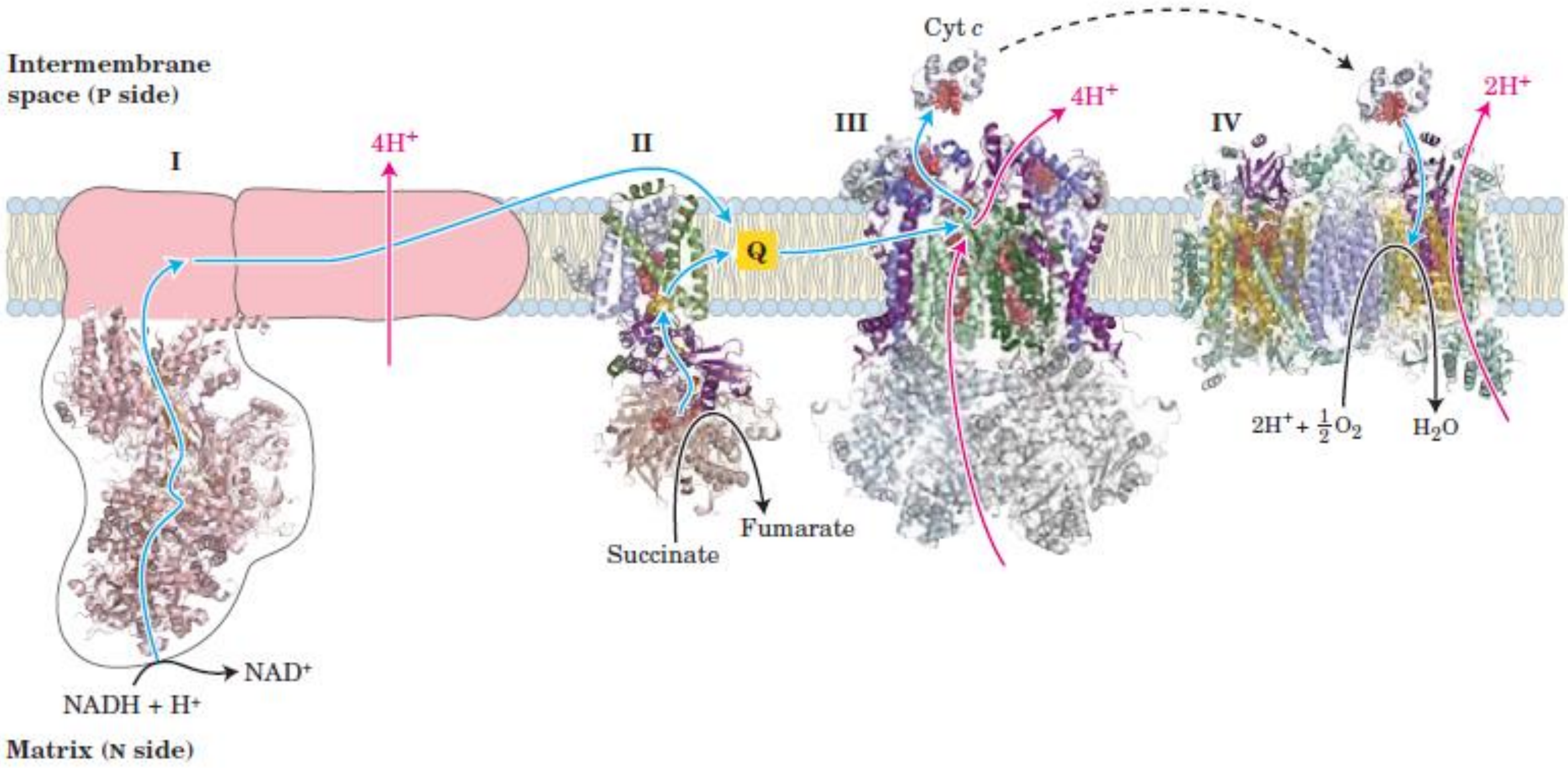
$$\text{Net equation: } QH_2 + 2\text{ cyt } c_1\,(\text{oxidized}) + 2H^+_N \longrightarrow Q + 2\text{ cyt } c_1\,(\text{reduced}) + 4H^+_P$$

The structure of cytochrome c oxidase



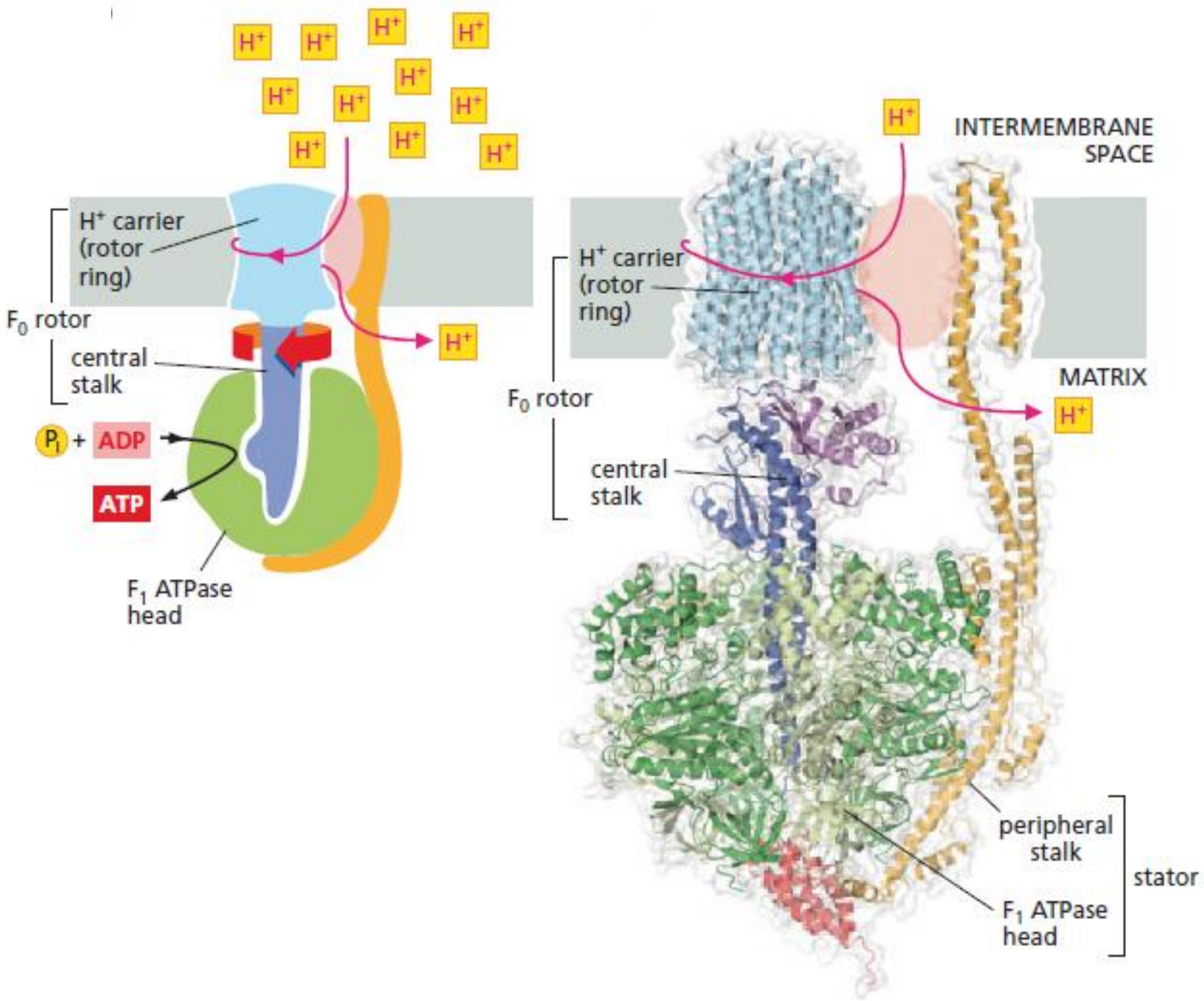


Summary of the flow of electrons and protons through the four complexes of the respiratory chain

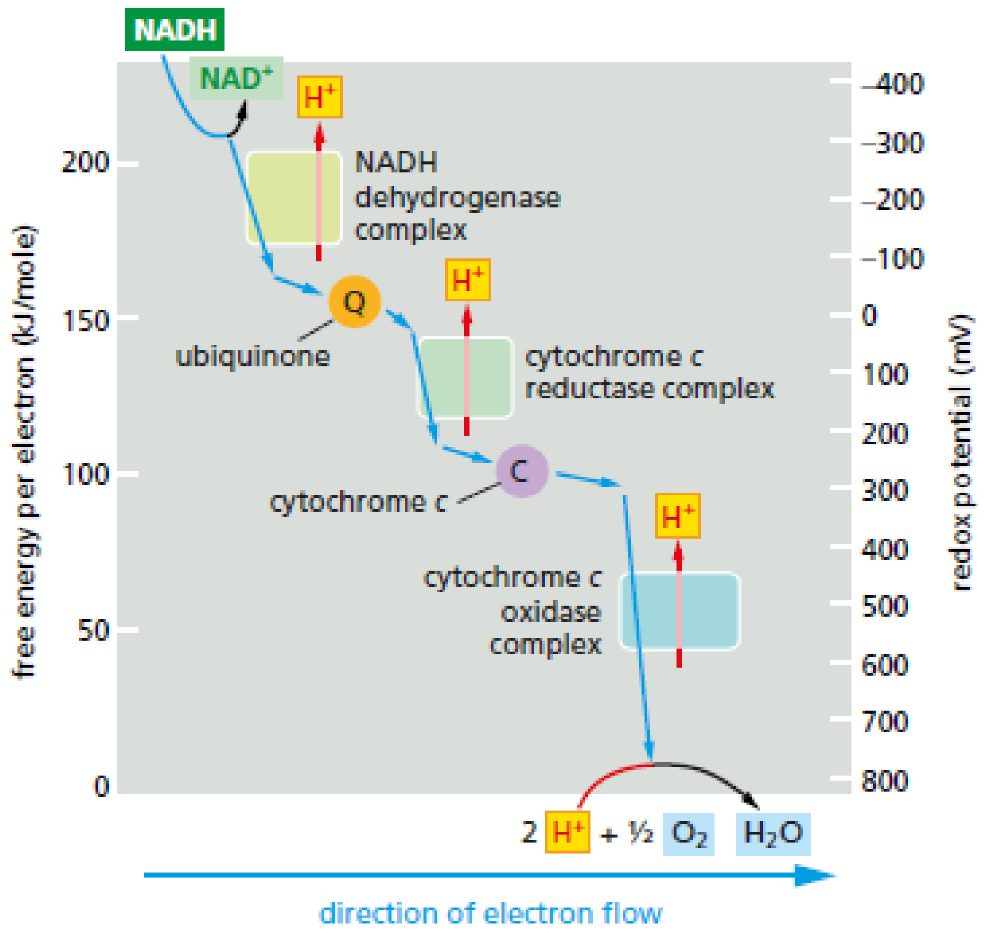


$$NADH + 11H_N^+ + \frac{1}{2}O_2 \longrightarrow NAD^+ + 10H_P^+ + H_2O$$

ATP synthase acts like a motor to convert the energy of protons flowing down their electrochemical gradient to chemical-bond energy in ATP



Redox potential changes along the mitochondrial electron transport chain



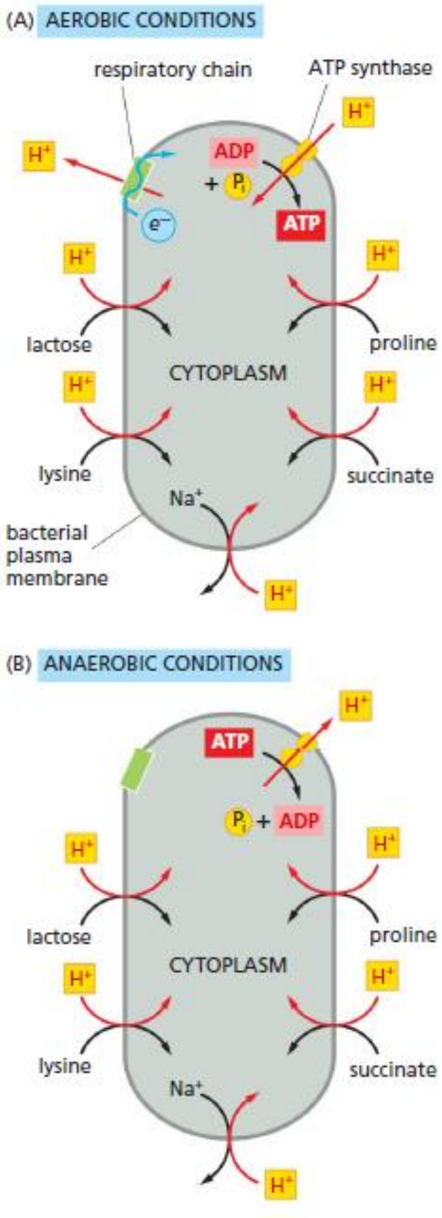
Product Yields from the Oxidation of Sugars and Fats

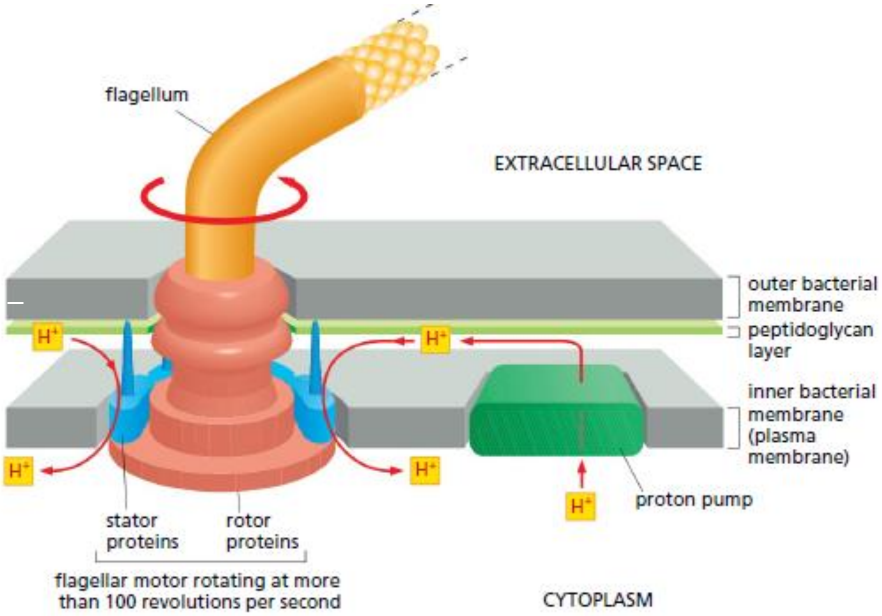
A. Net products from oxidation of one molecule of glucose
In cytosol (glycolysis) 1 glucose → 2 pyruvate + 2 NADH + 2 ATP In mitochondrion (pyruvate dehydrogenase and citric acid cycle) 2 pyruvate → 2 acetyl CoA + 2 NADH 2 acetyl CoA → 6 NADH + 2 $FADH_2$ + 2 GTP
Net result in mitochondrion 2 pyruvate → 8 NADH + 2 $FADH_2$ + 2 GTP
B. Net products from oxidation of one molecule of palmitoyl CoA (activated form of palmitate, a fatty acid)
In mitochondrion (fatty acid oxidation and citric acid cycle) 1 palmitoyl CoA → 8 acetyl CoA + 7 NADH + 7 $FADH_2$ 8 acetyl CoA → 24 NADH + 8 $FADH_2$ + 8 GTP
Net result in mitochondrion 1 palmitoyl CoA → 31 NADH + 15 $FADH_2$ + 8 GTP

NADH produced in mitochondria can provide energy for the formation of about 2.5 molecules of ATP

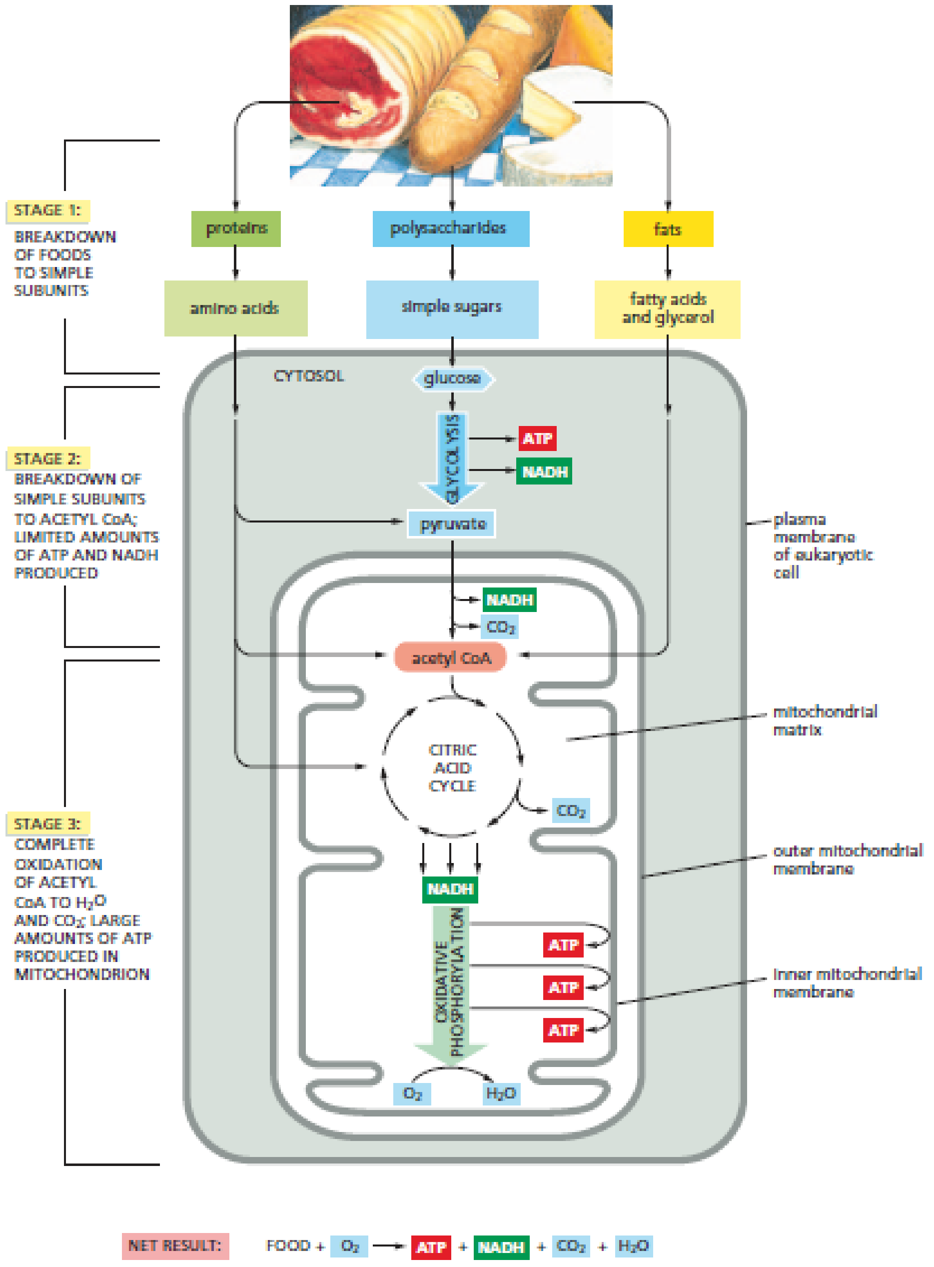
$FADH_2$ produced can provide 1.5 molecules of ATP

The importance of H^+-driven transport in bacteria

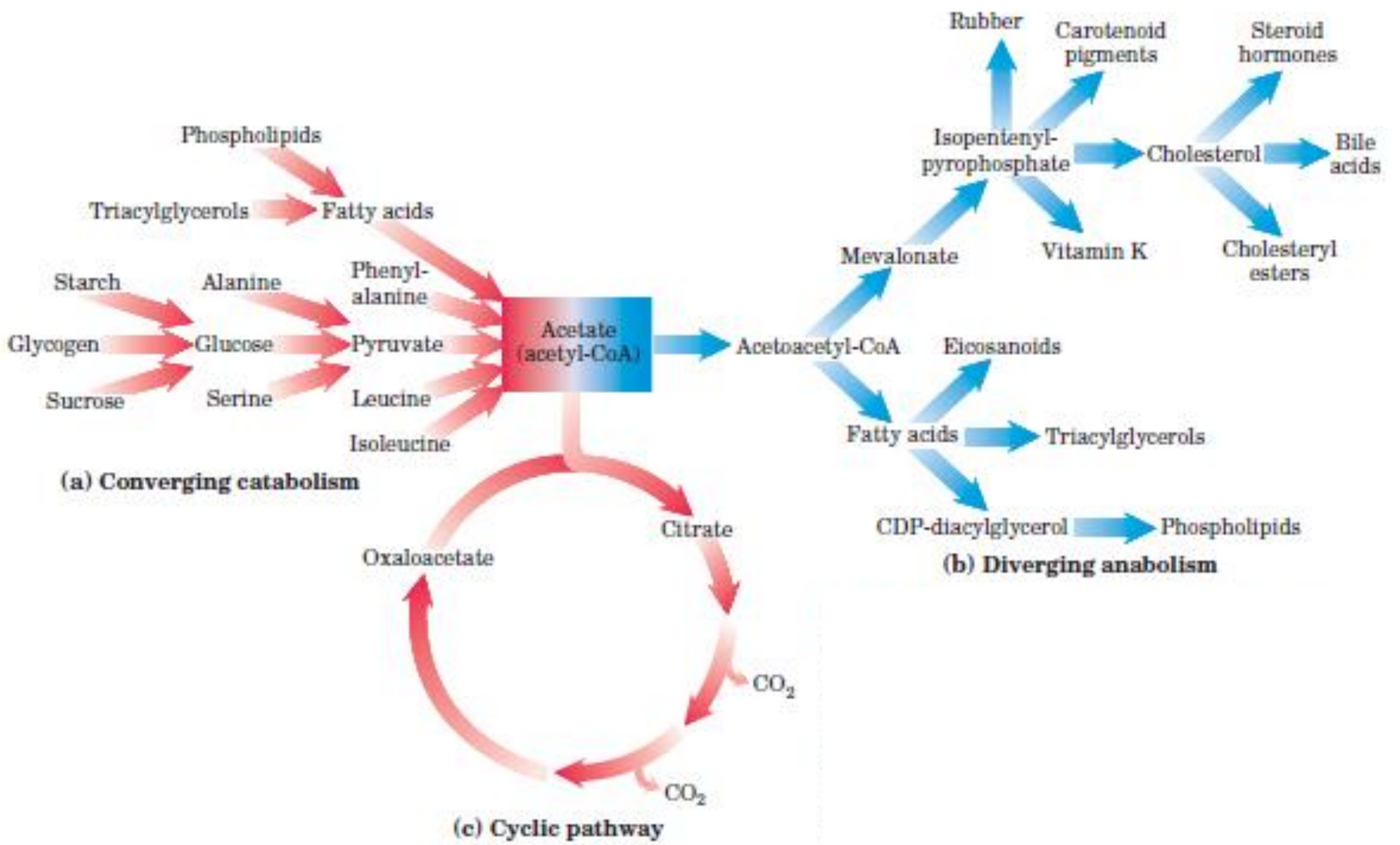




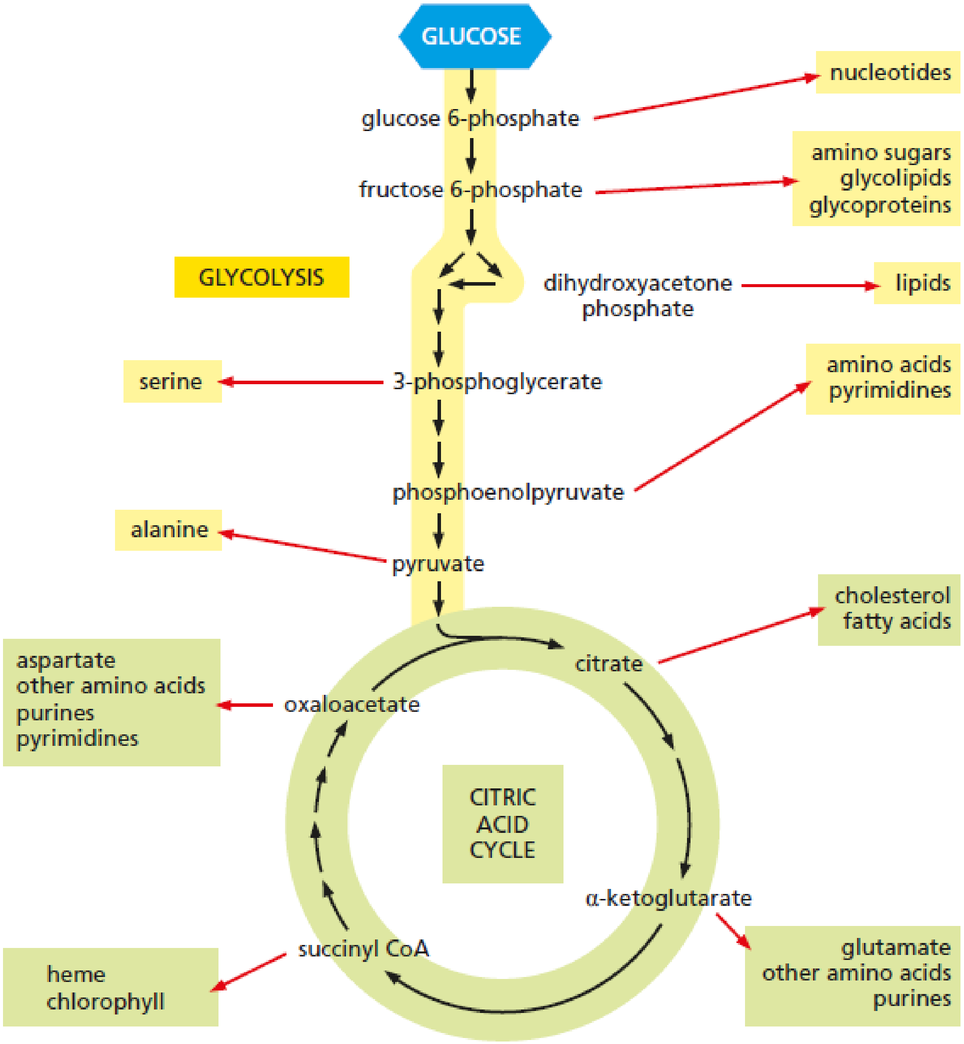
The breakdown of food molecules occurs in three stages



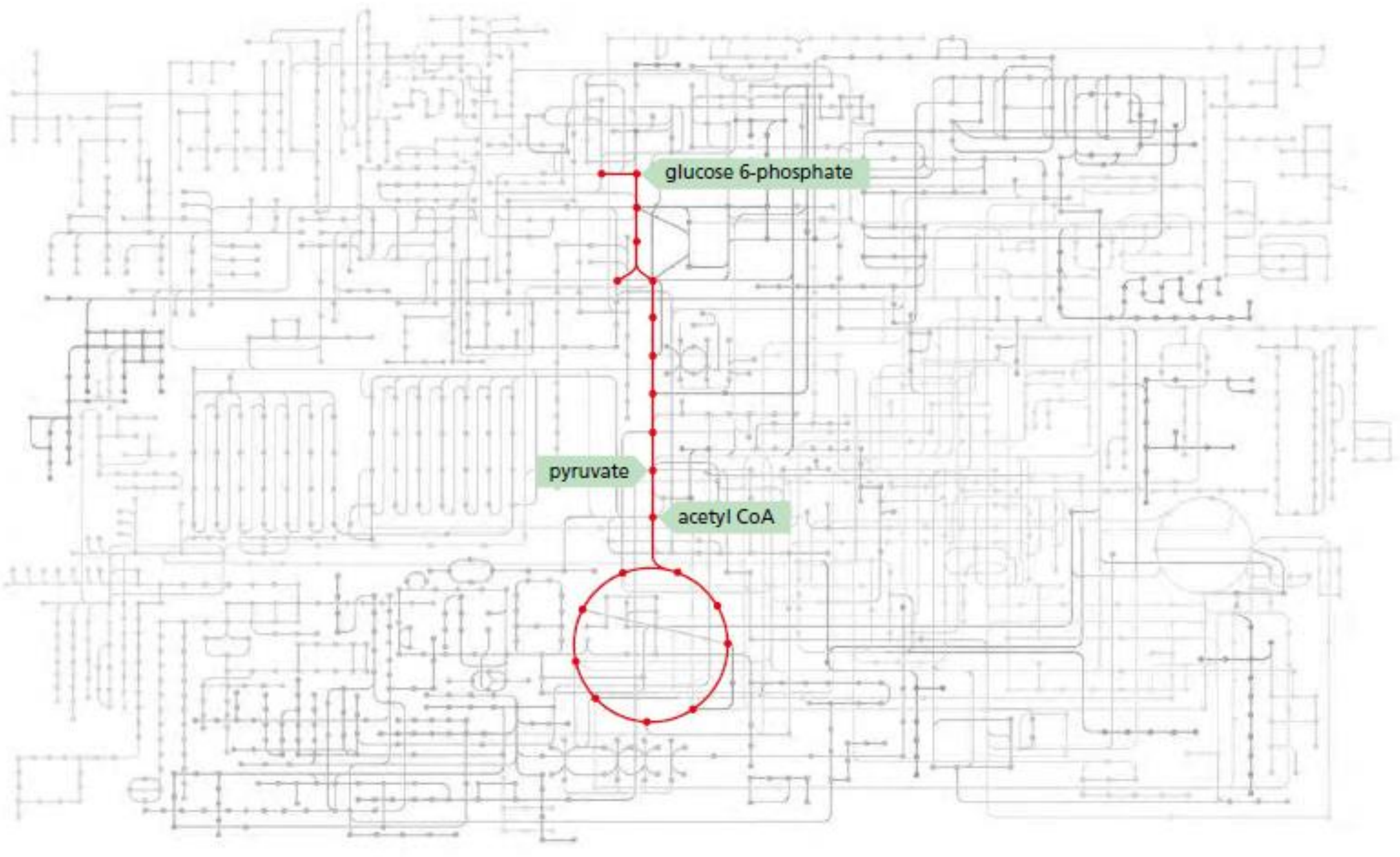
Three types of metabolic pathways



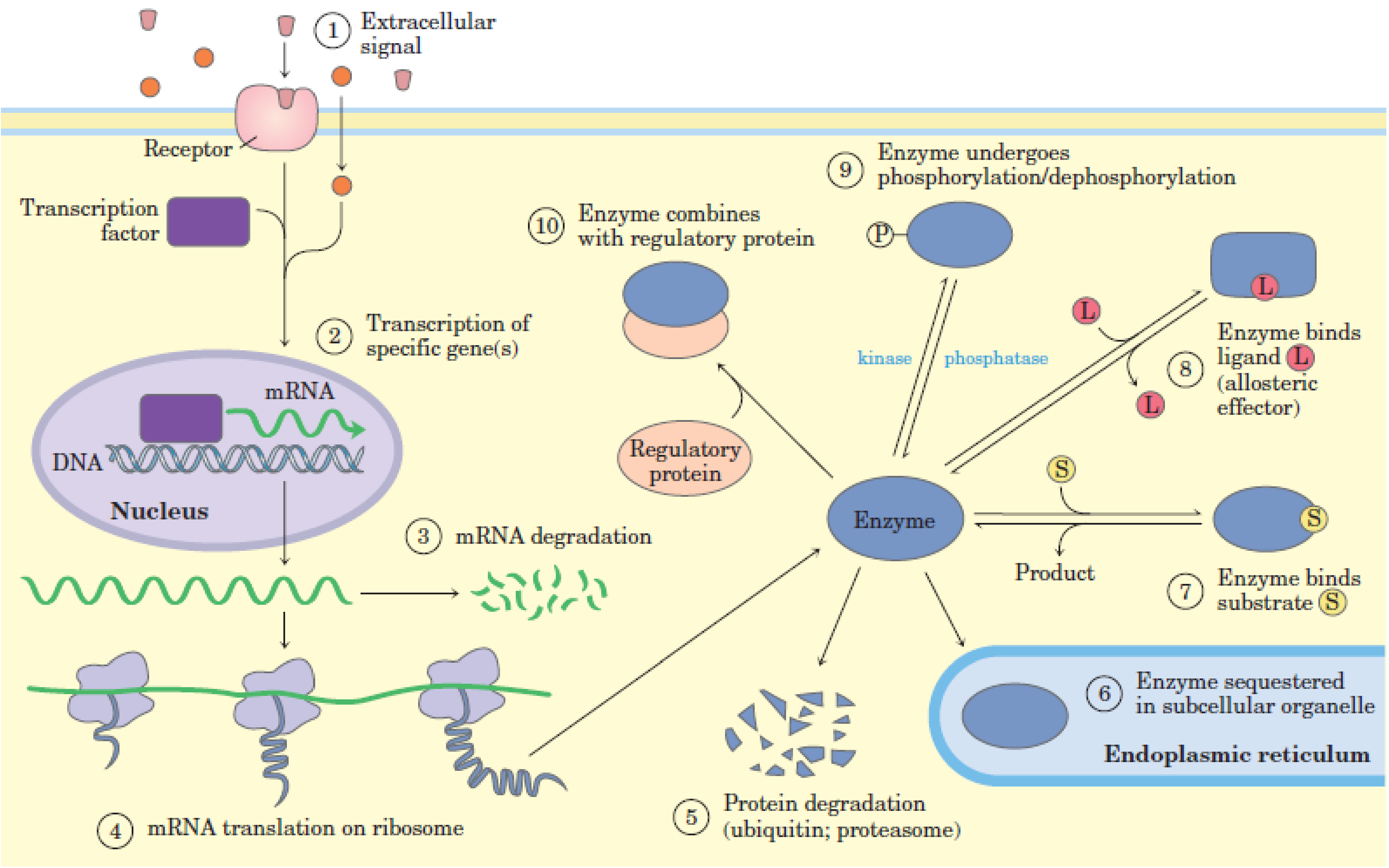
Glycolysis and the citric acid cycle provide the precursors needed to synthesize many important biological molecules



Glycolysis and the citric acid cycle are at the center of an elaborate set of metabolic pathways in human cells

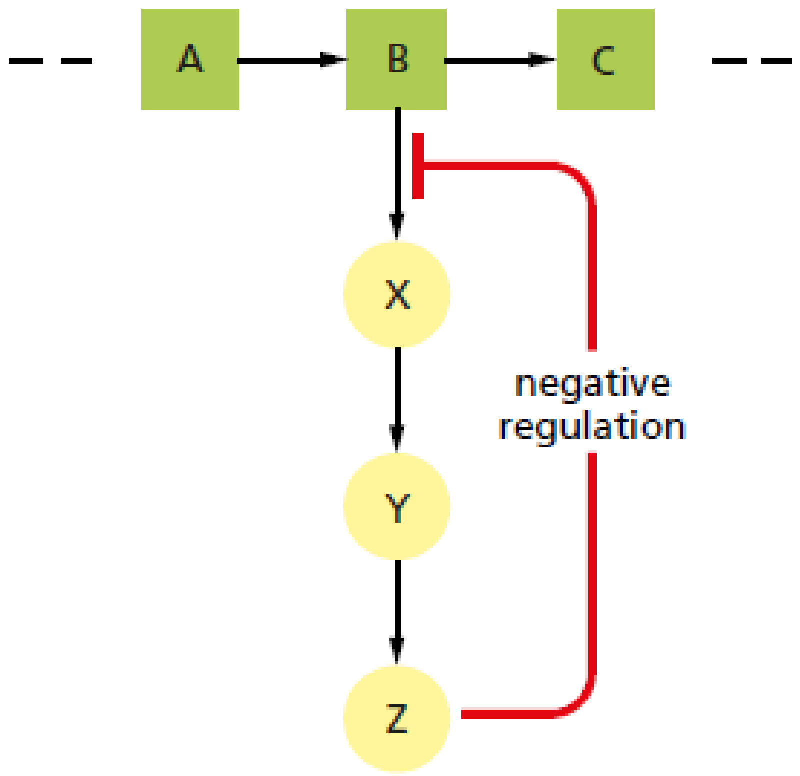


Factors affecting the activity of enzymes

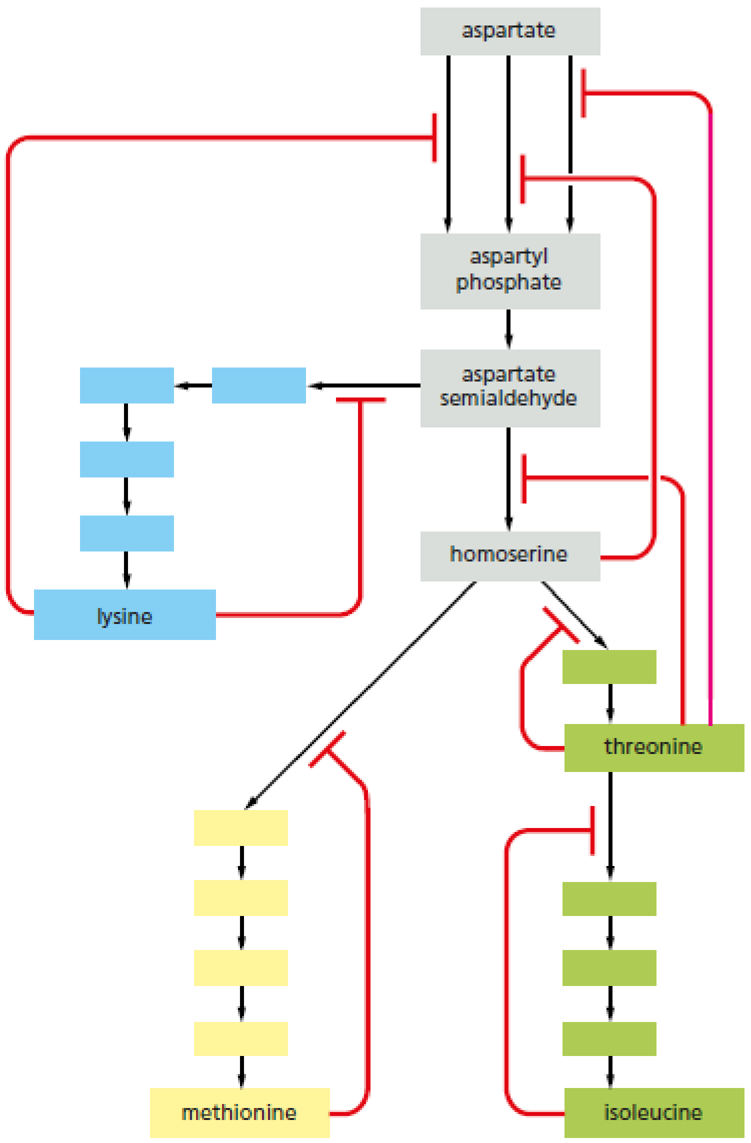


Feedback inhibition

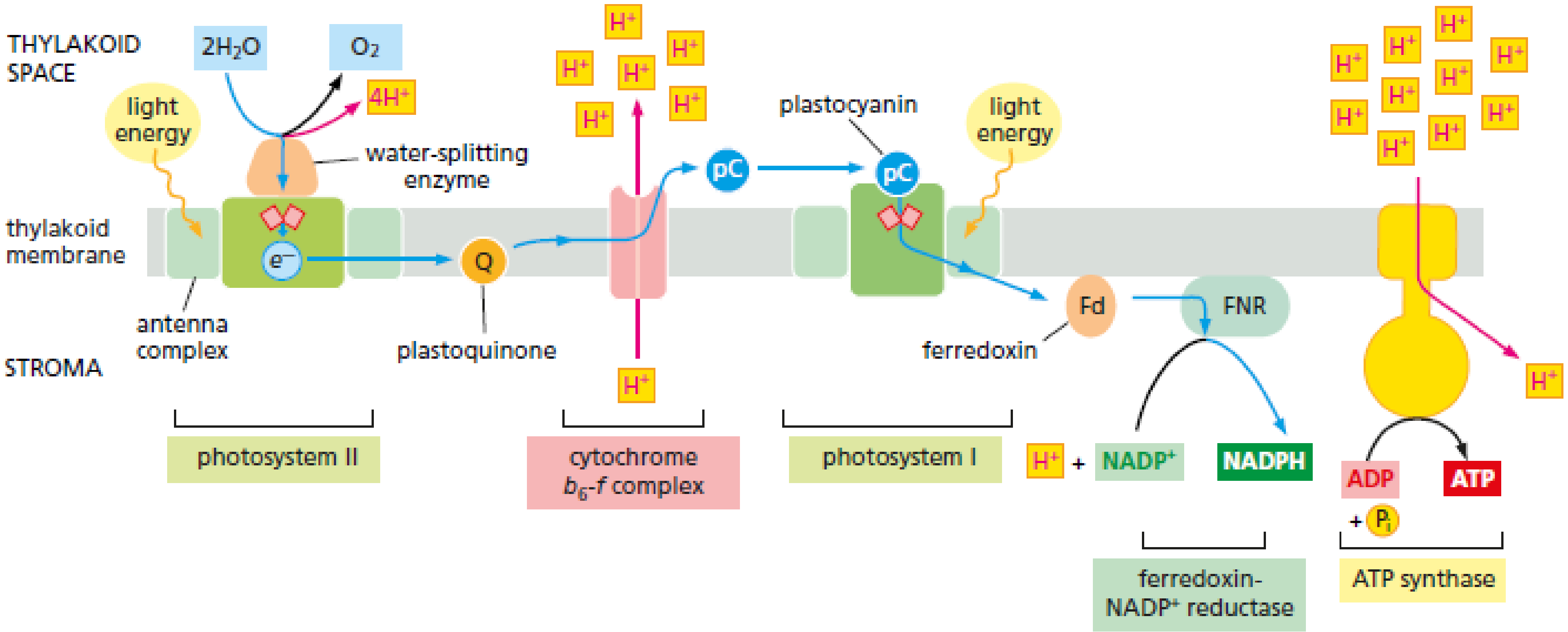
Feedback inhibition of a single biosynthetic pathway



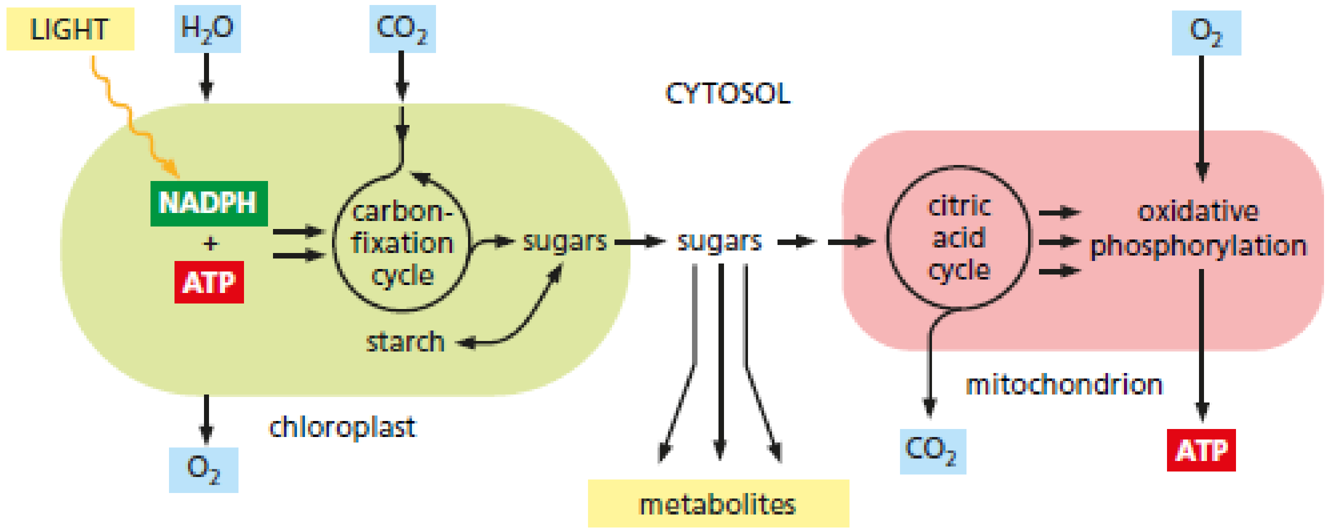
Multiple feedback inhibition



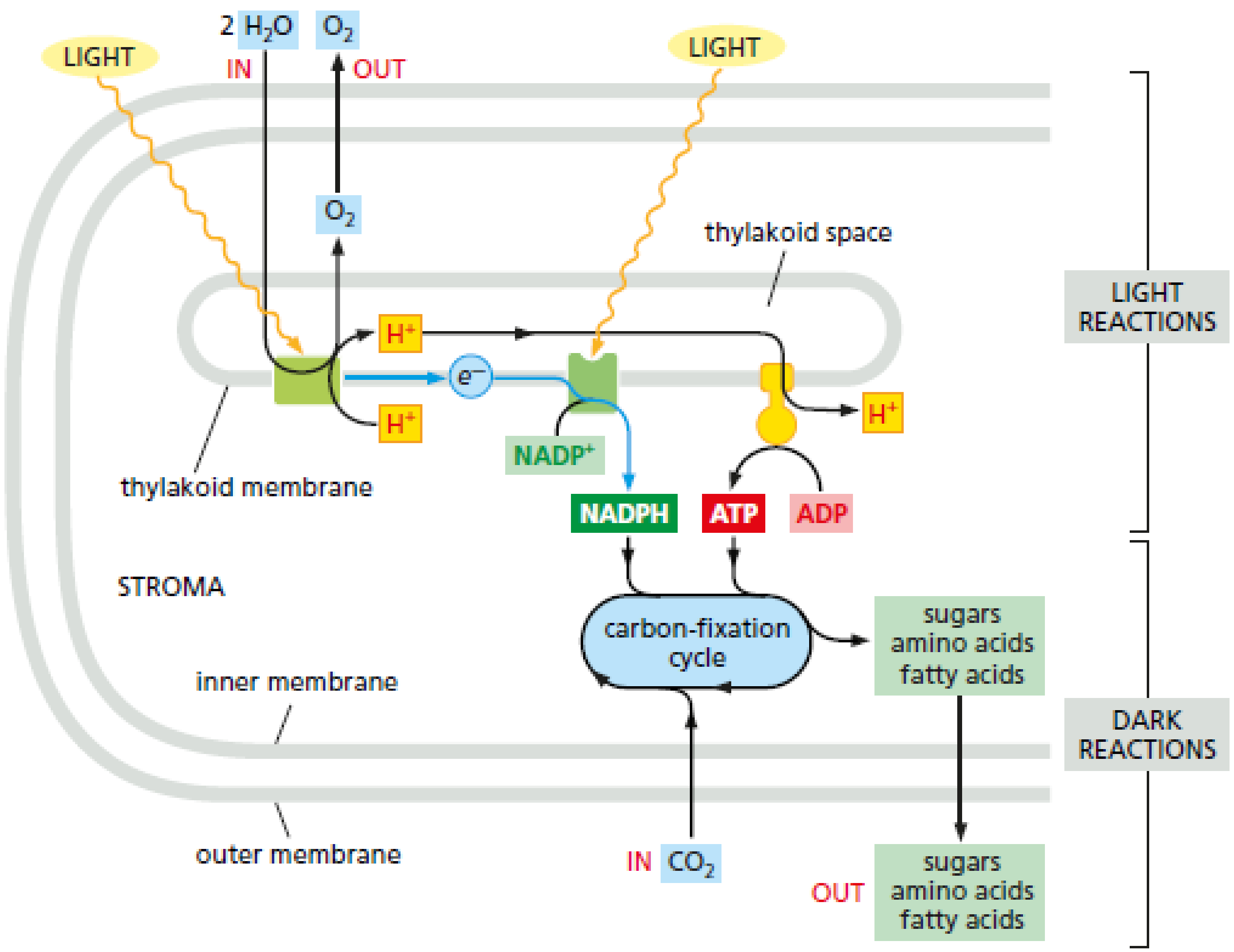
CHLOROPLASTS AND PHOTOSYNTHESIS



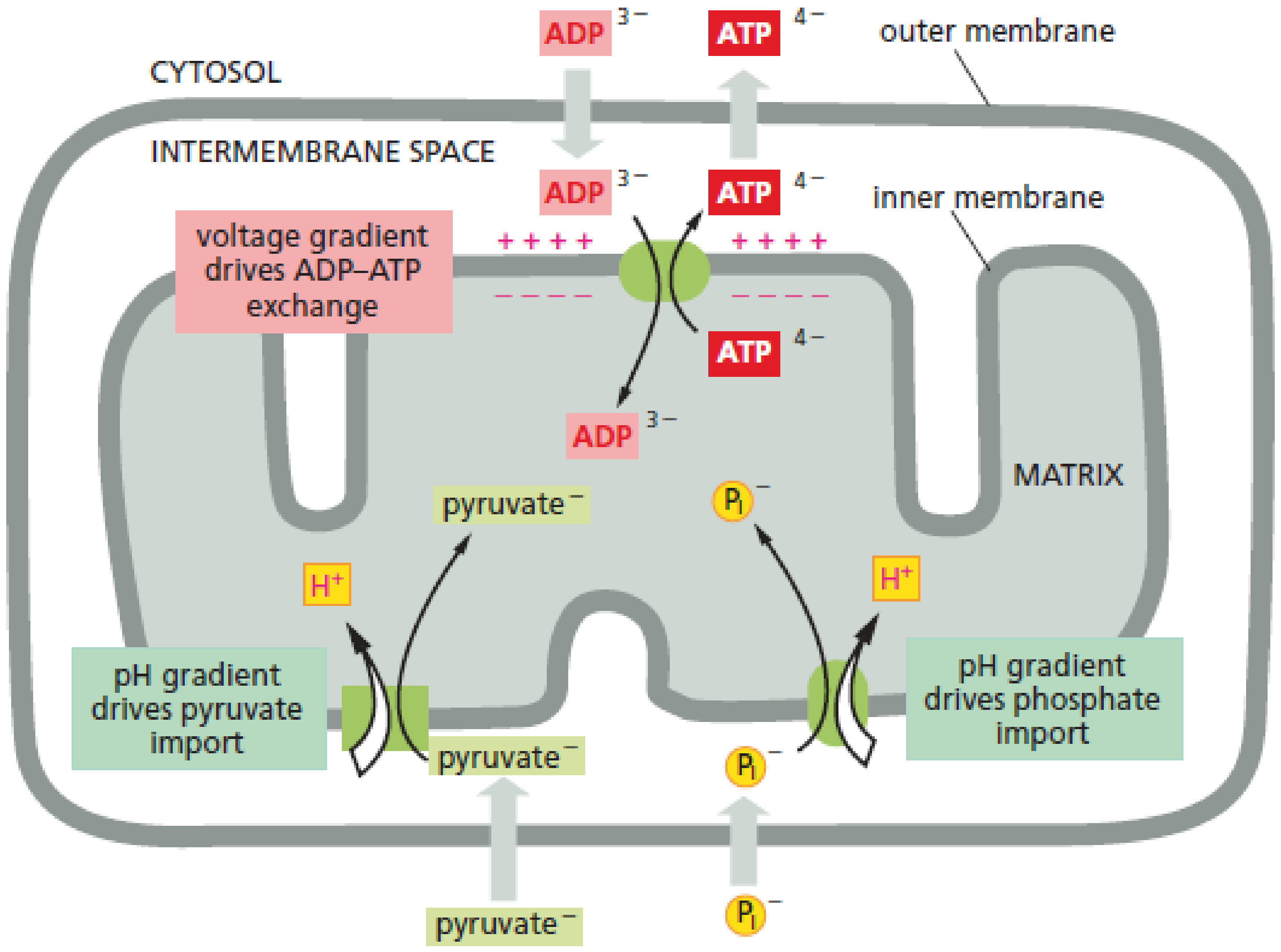
How chloroplasts and mitochondria collaborate to supply cells with both metabolites and ATP



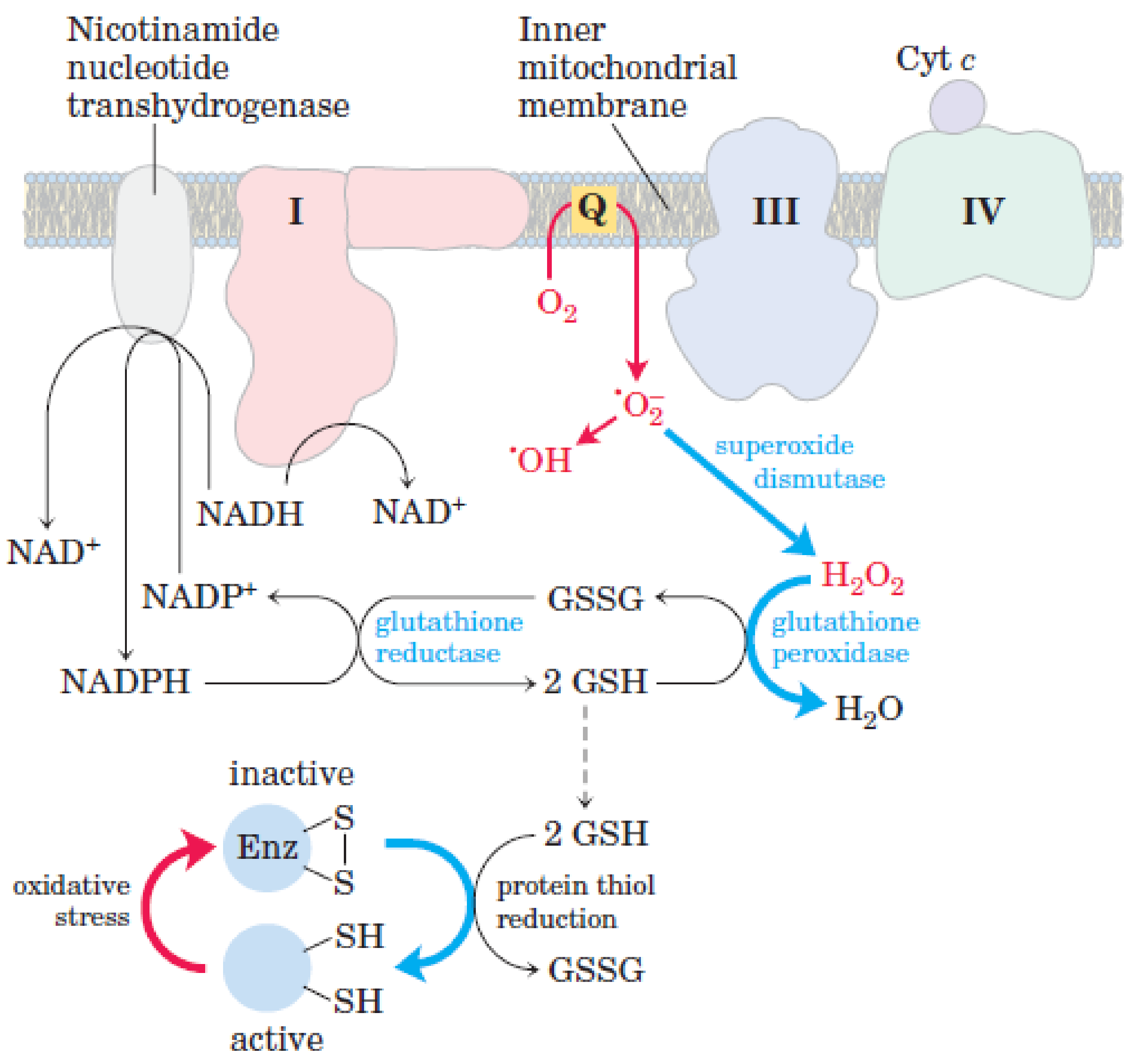
CHLOROPLASTS AND PHOTOSYNTHESIS



Coupled transport processes



ROS formation in mitochondria and mitochondrial defenses

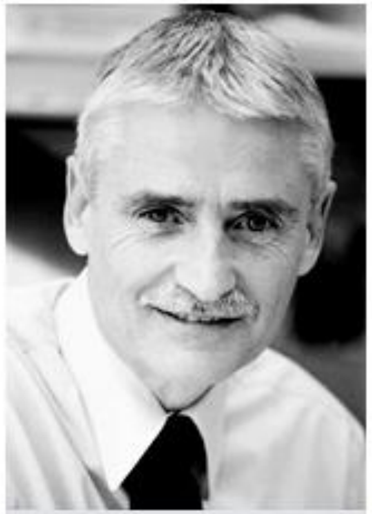


Factors that slow the flow of electrons through the respiratory chain increase the formation of superoxide

Medicine

The Nobel Prize in Physiology or Medicine 2001

"for their discoveries of key regulators of the cell cycle"



Leland H. Hartwell

1/3 of the prize

USA

Fred Hutchinson Cancer Research Center
Seattle, WA, USA

b. 1939

Tim Hunt

1/3 of the prize

United Kingdom

Imperial Cancer Research Fund
London, United Kingdom

b. 1943

Sir Paul M. Nurse

1/3 of the prize

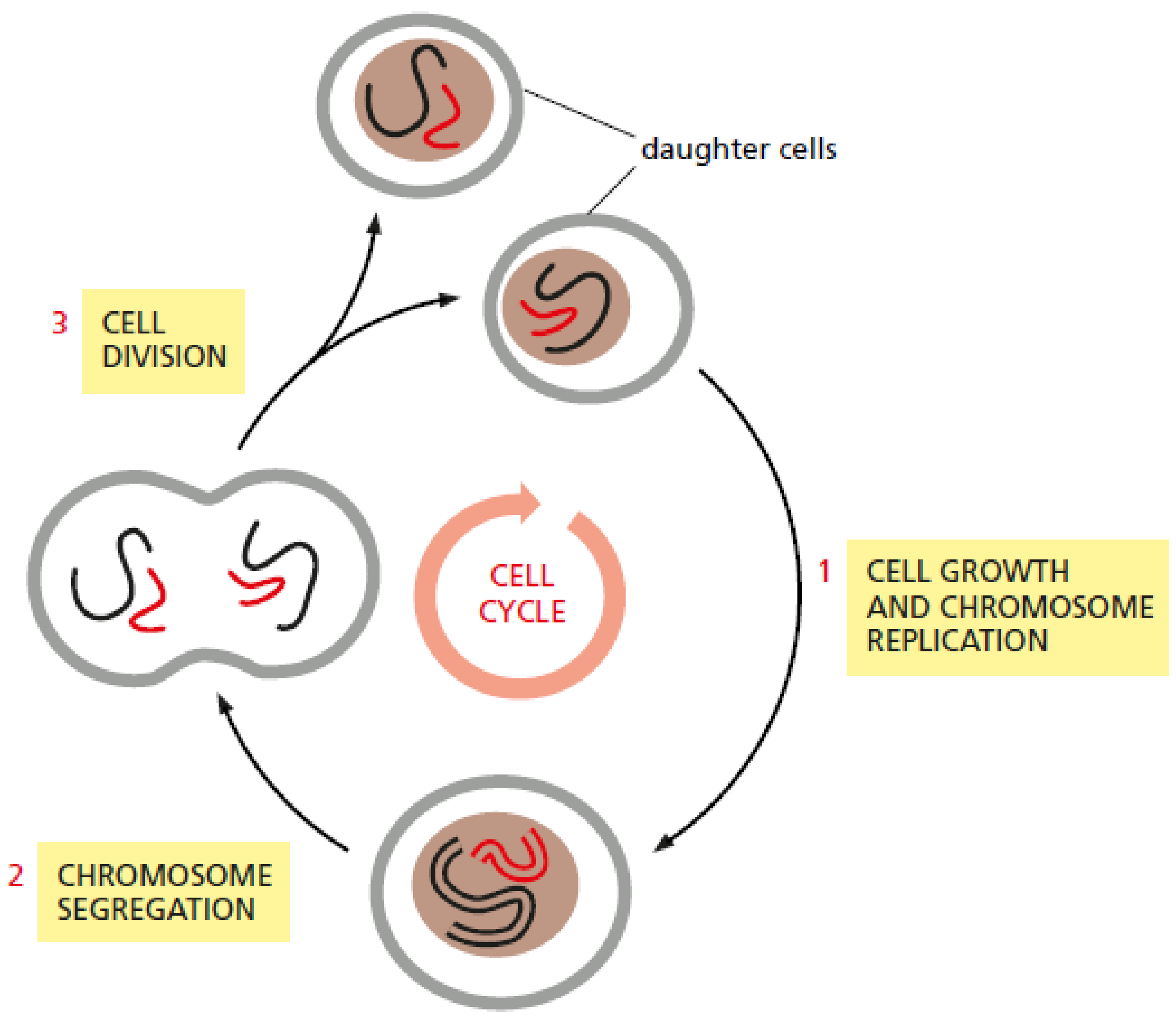
United Kingdom

Imperial Cancer Research Fund
London, United Kingdom

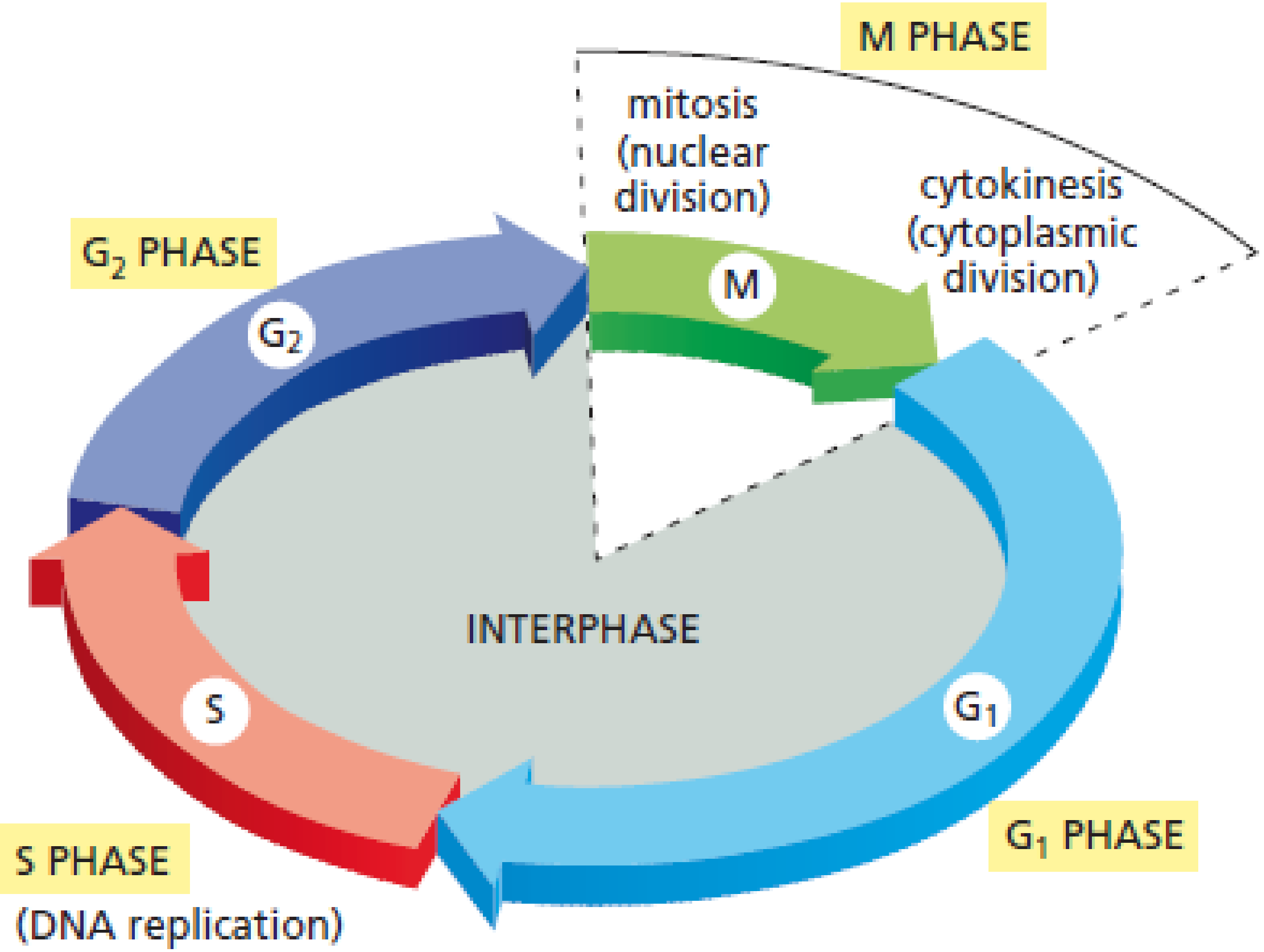
b. 1949

Titles, data and places given above refer to the time of the award.

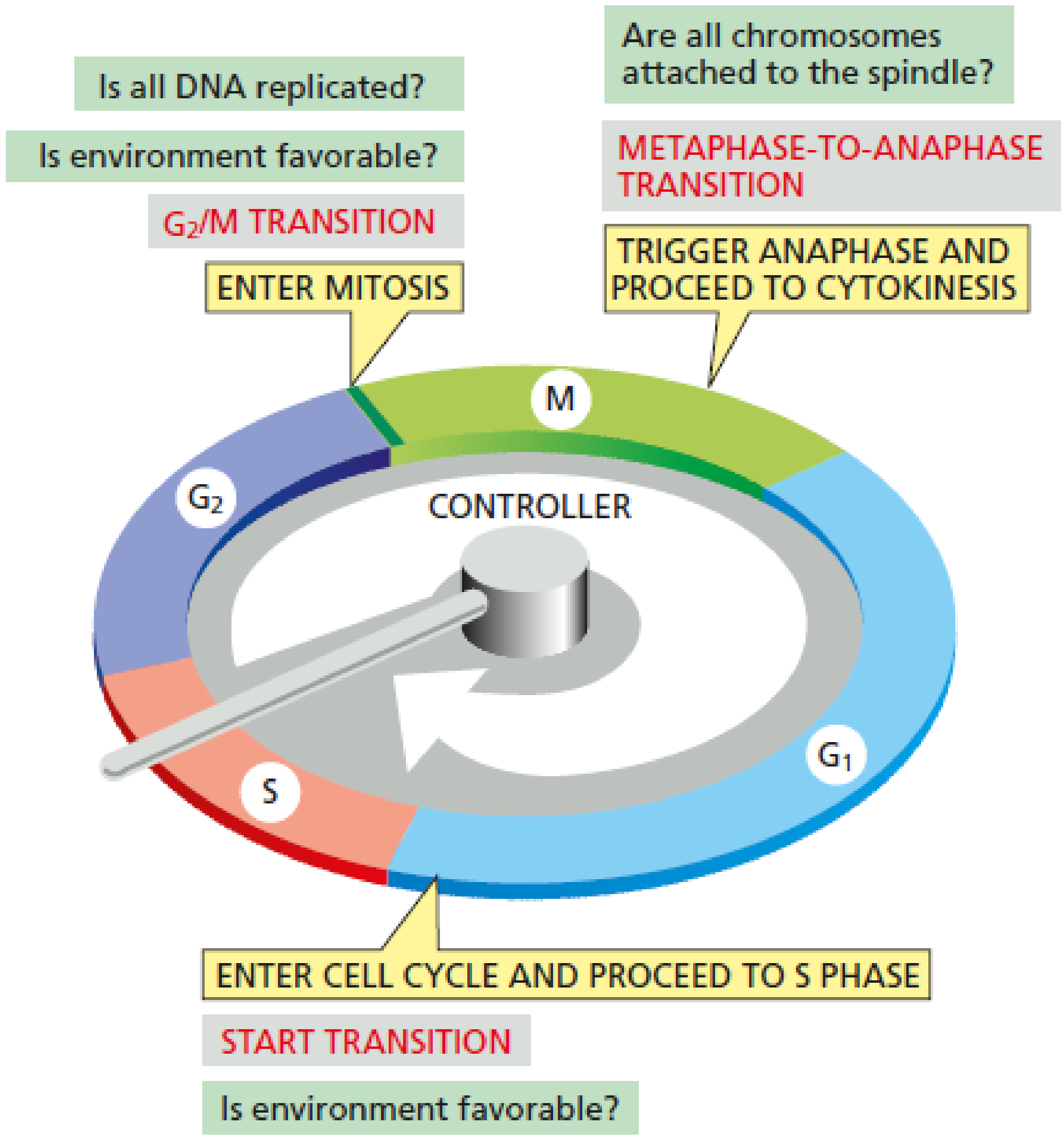
The cell cycle



Eukaryotic Cell Cycle Usually Consists of Four Phases

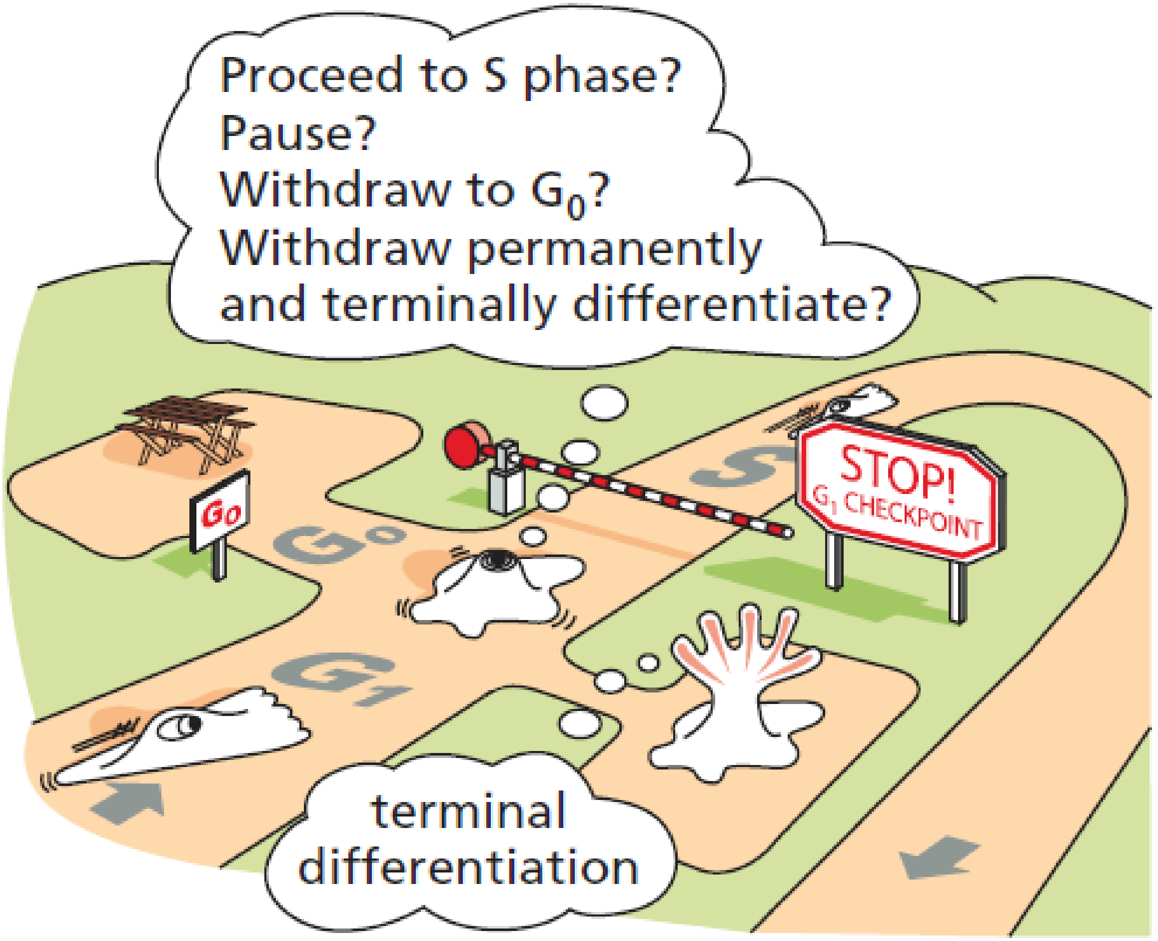


Controller

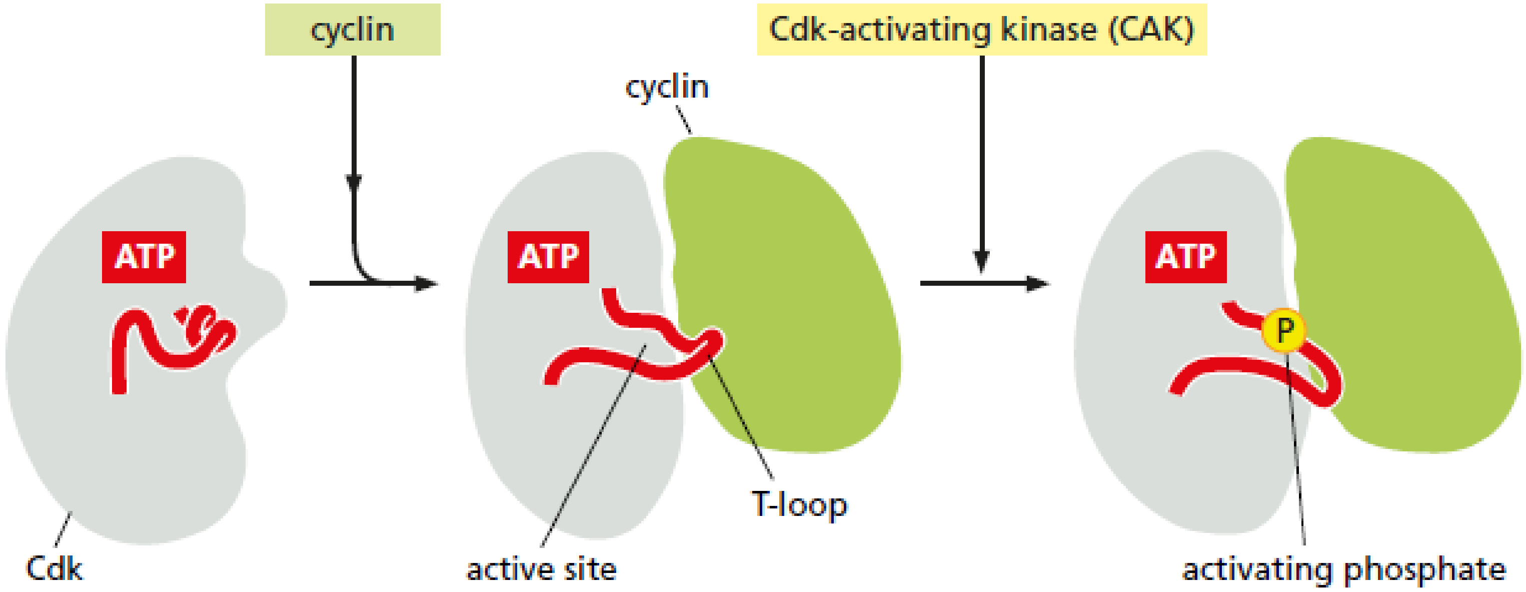


Is all DNA replicated?
Is environment favorable?
G2/M TRANSITION
ENTER MITOSIS
Are all chromosomes attached to the spindle?
METAPHASE-TO-ANAPHASE TRANSITION
TRIGGER ANAPHASE AND PROCEED TO CYTOKINESIS
M
G2
CONTROLLER
G1
S
ENTER CELL CYCLE AND PROCEED TO S PHASE
START TRANSITION
Is environment favorable?

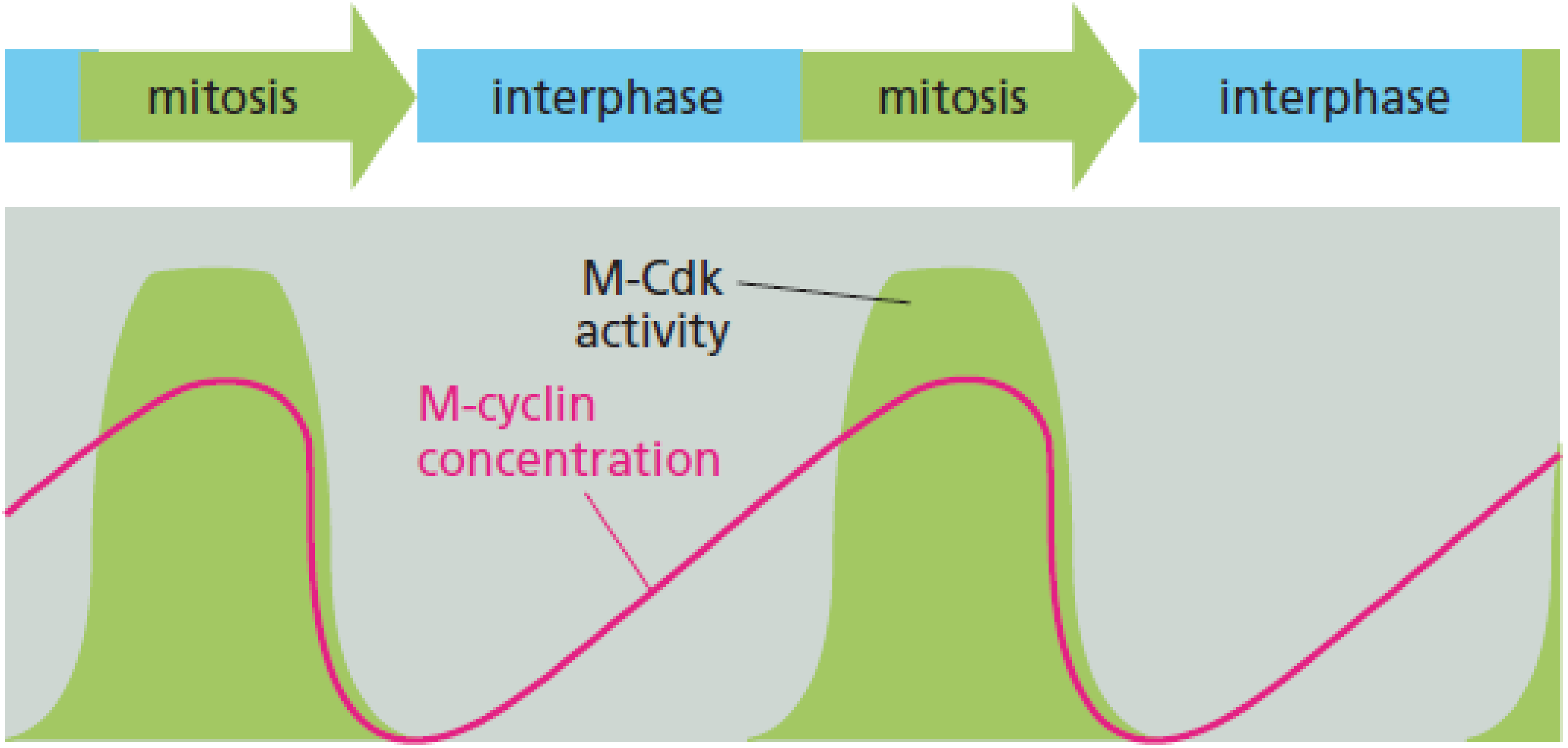
The transition from G1 to S phase offers the cell a crossroad



Cyclin–Cdk complexes of the cell-cycle control system



The accumulation of cyclins helps regulate the activity of Cdks



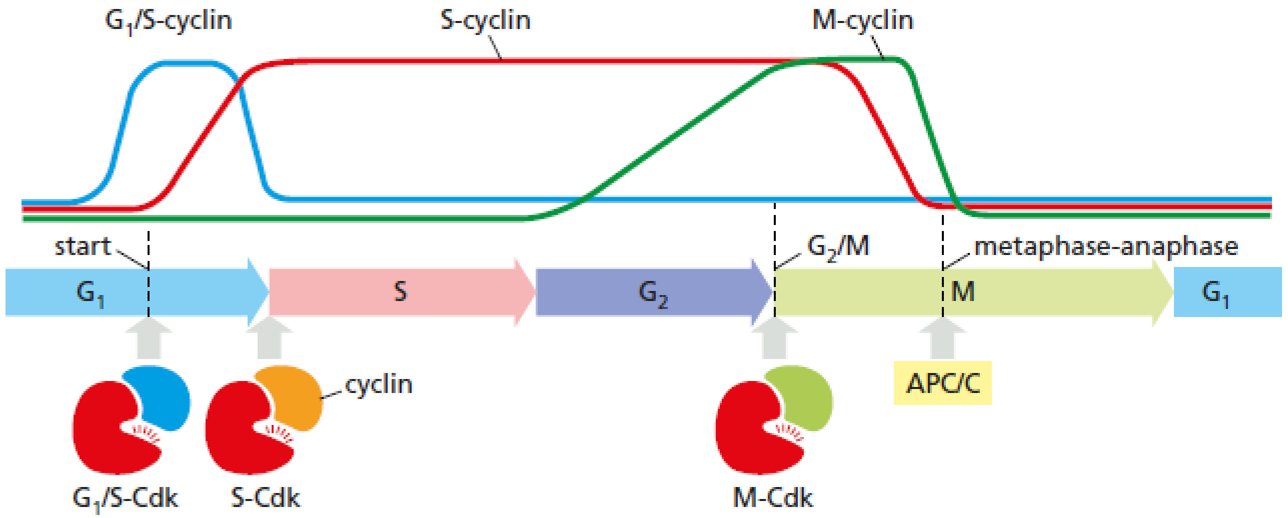
The Major Cyclins and Cdks of Vertebrates

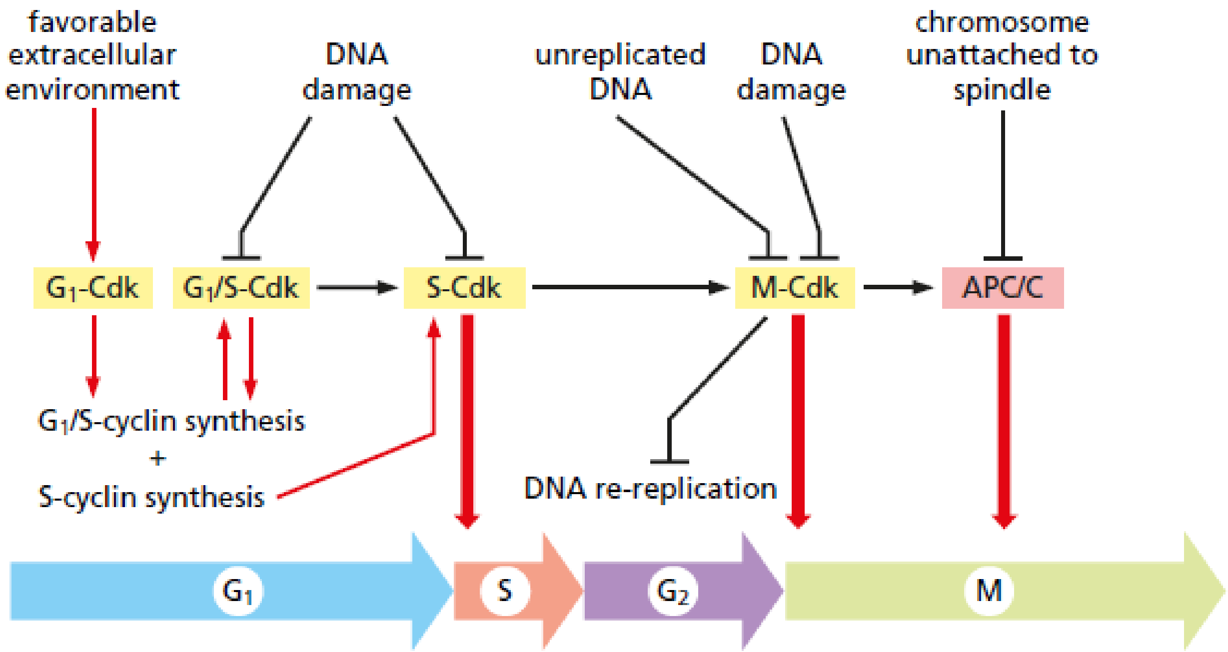
Cyclin–Cdk Complex	Cyclin	Cdk Partner
G_1-Cdk	cyclin D*	Cdk4, Cdk6
G_1/S-Cdk	cyclin E	Cdk2
S-Cdk	cyclin A	Cdk2
M-Cdk	cyclin B	Cdk1

*There are three D cyclins in mammals (cyclins D1, D2, and D3).

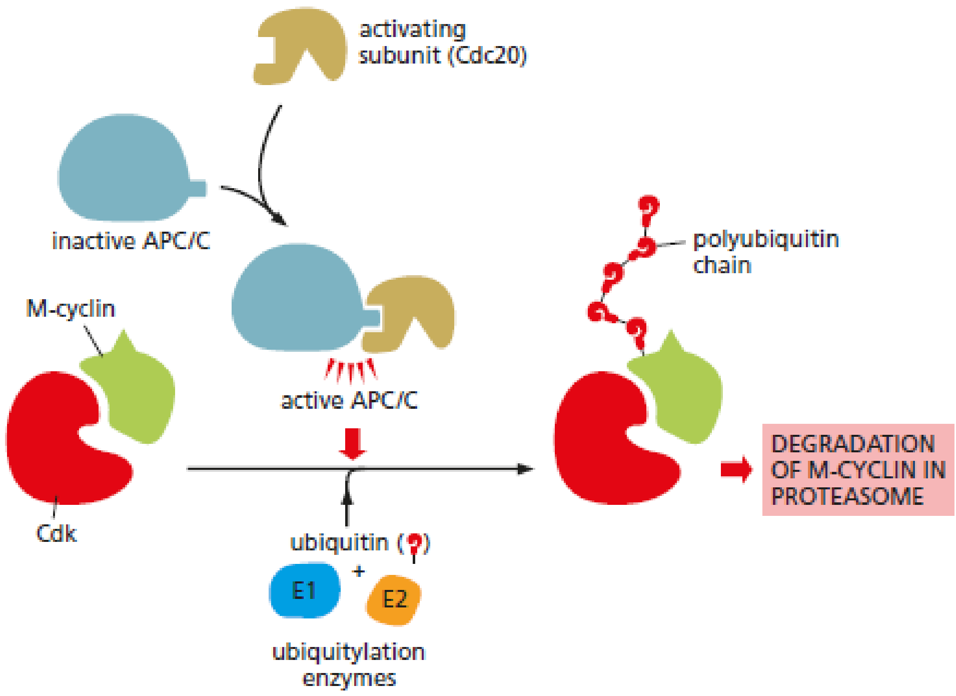
Cyclin Concentrations are Regulated by Transcription and by Proteolysis (degradation)

Cyclin–Cdk complexes of the cell-cycle control system

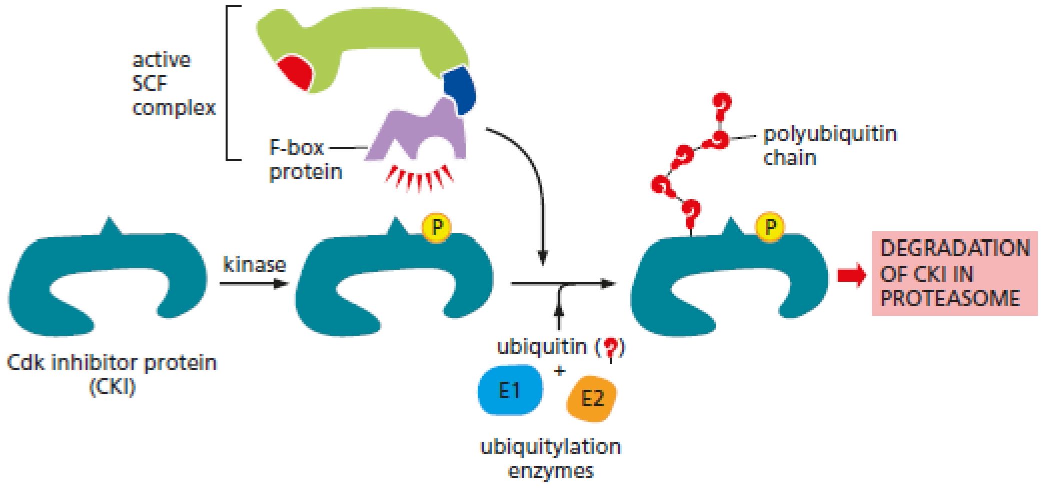




Protein Degradation

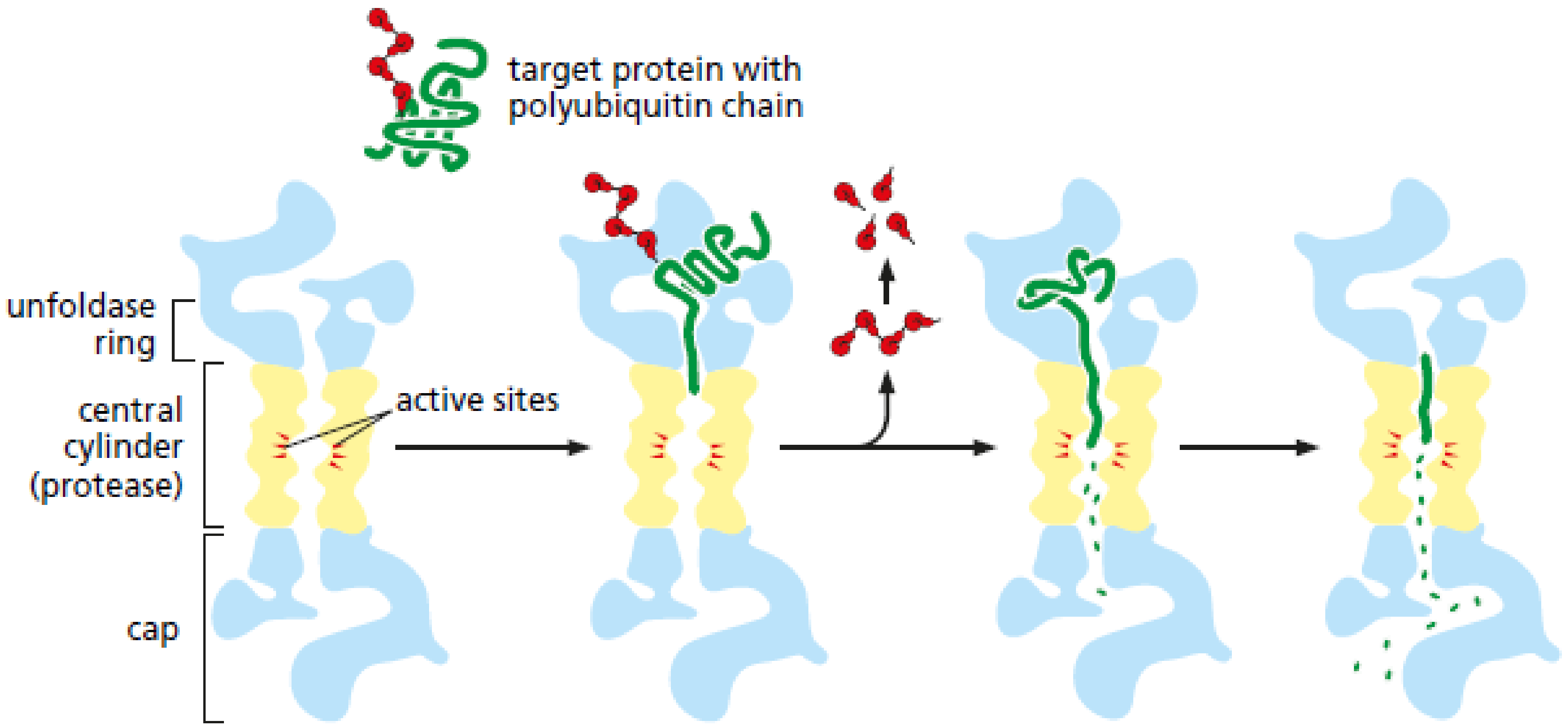


Anaphase promoting complex (APC)

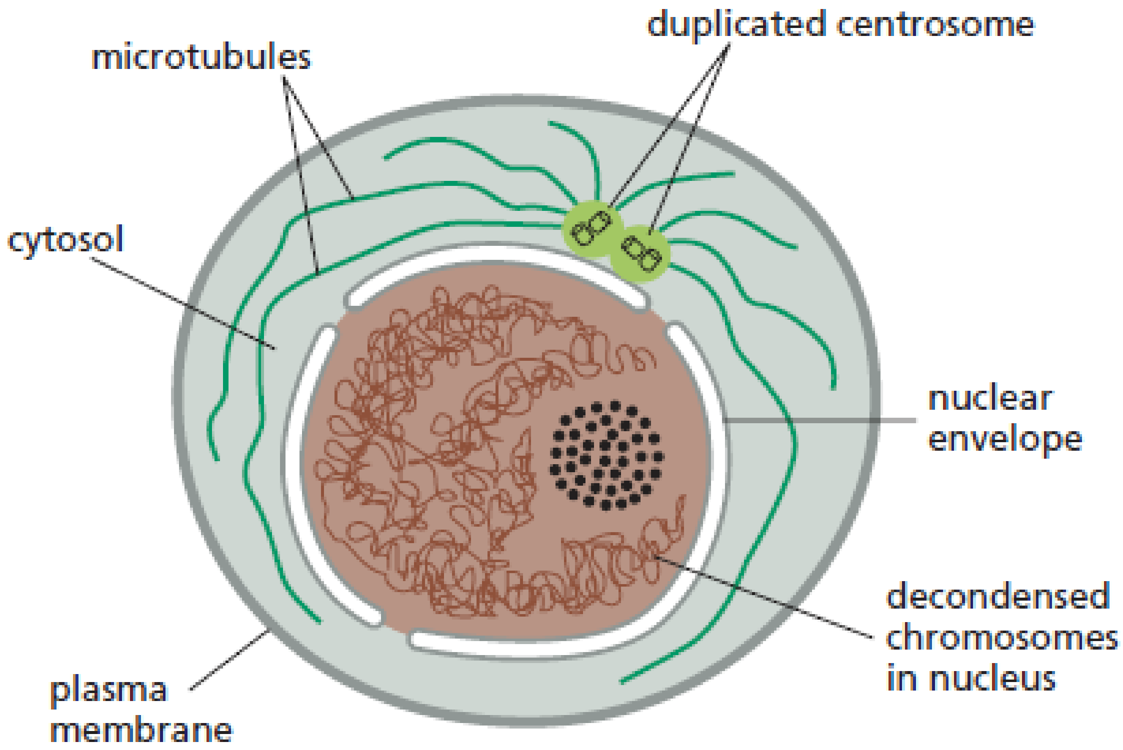


Control of proteolysis
by SCF (Skp, Cullin, F-box
containing complex)

Processive protein digestion by the proteasome

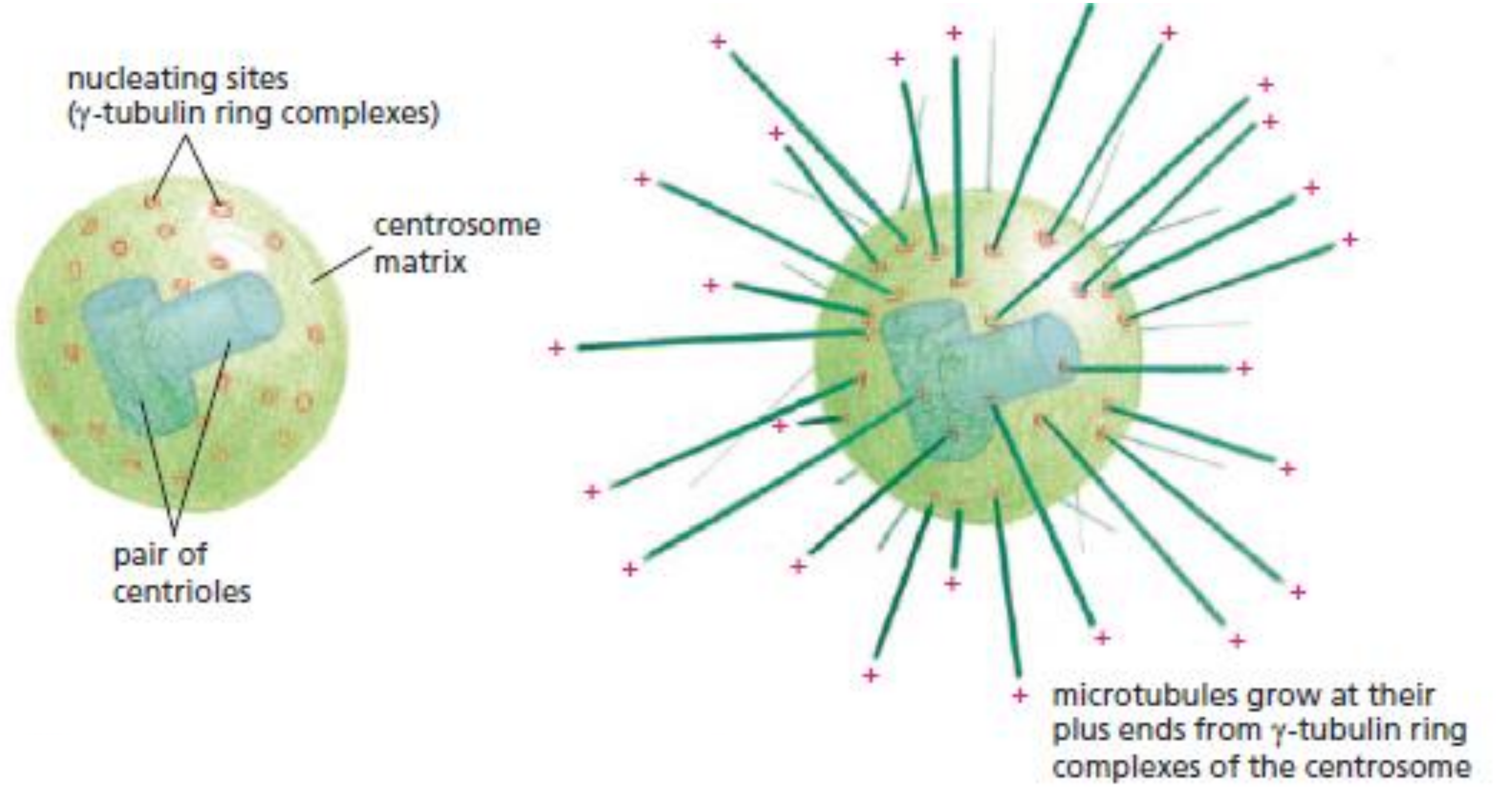


Interphase

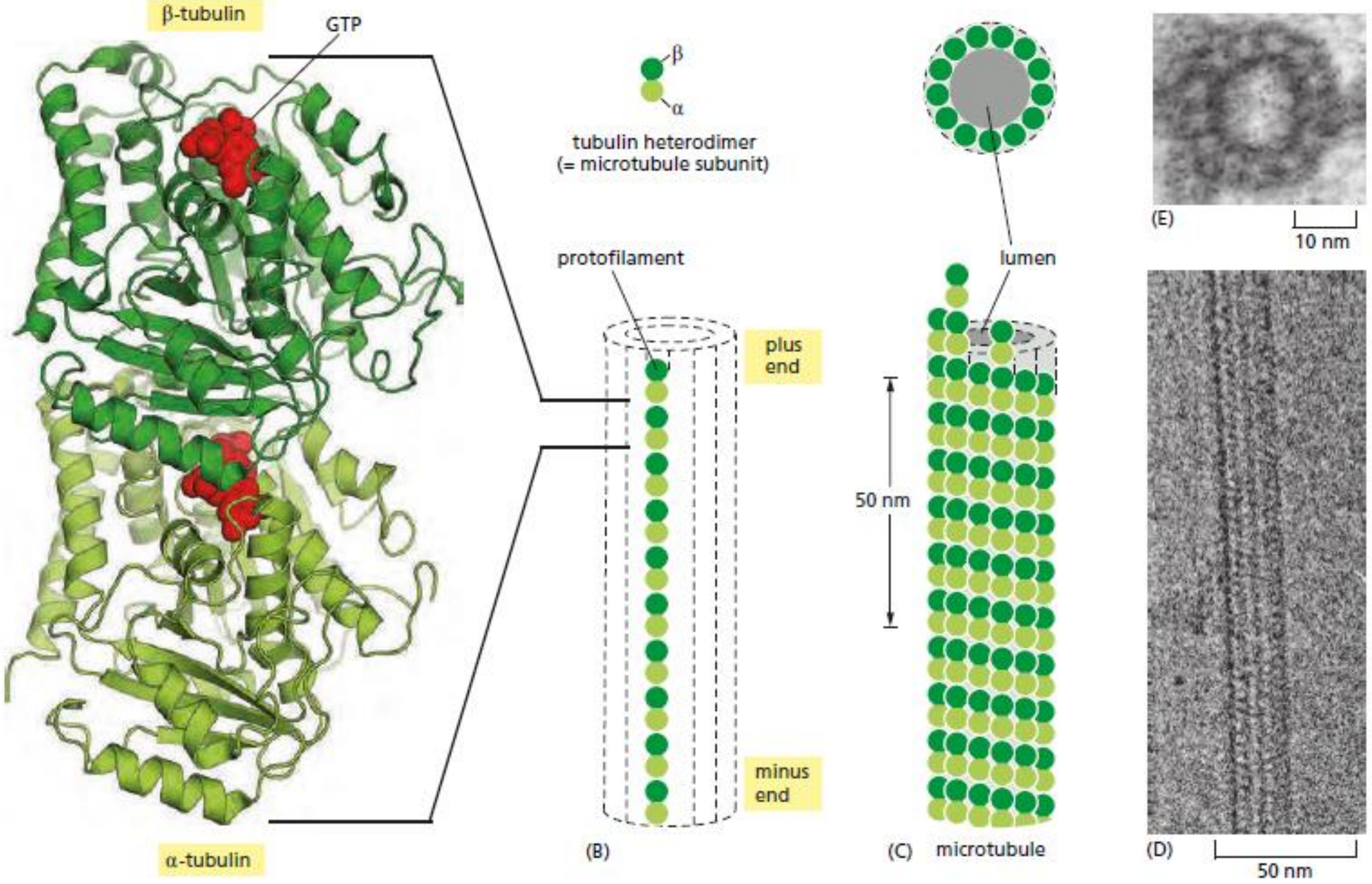


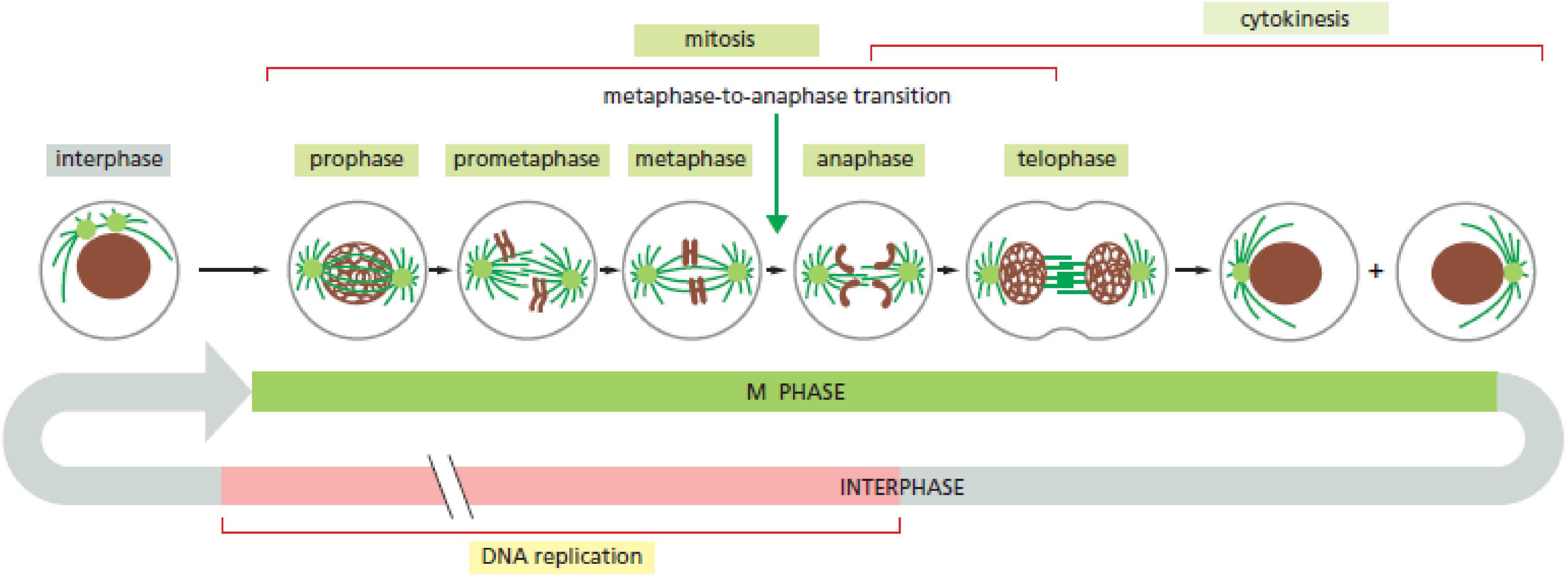
During interphase, the cell increases in size. The DNA of the chromosomes is replicated, and the centrosome is duplicated.

Tubulin polymerizes from nucleation sites on a centrosome



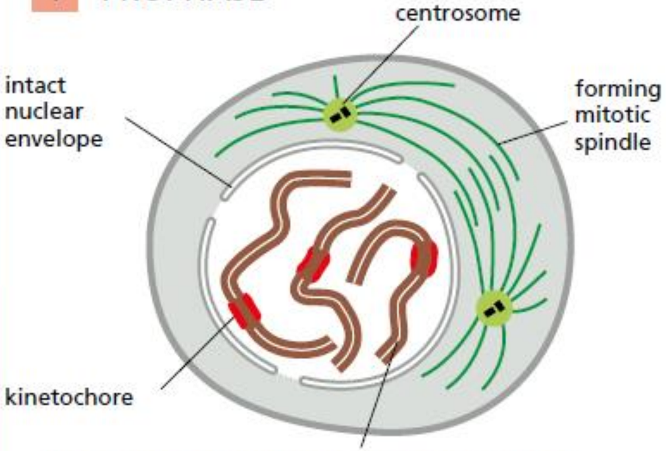
Microtubule Structure





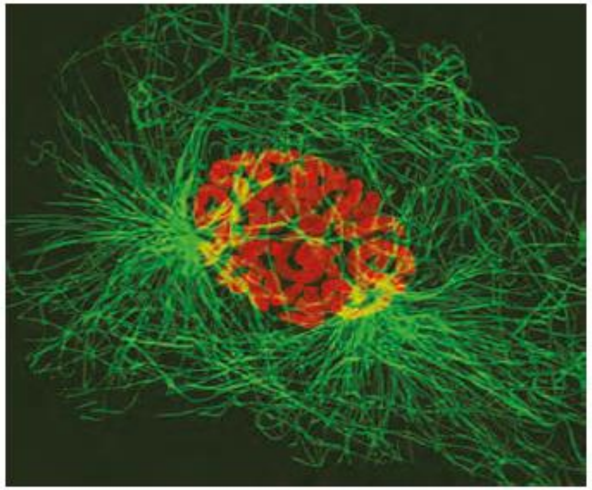
mitosis
cytokinesis
metaphase-to-anaphase transition
interphase
prophase
prometaphase
metaphase
anaphase
telophase
+
M PHASE
INTERPHASE
DNA replication

1 PROPHASE

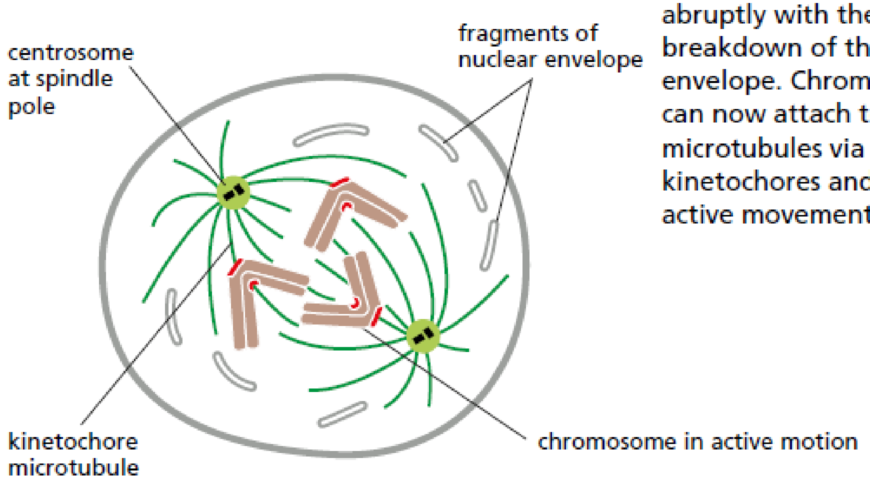


condensing replicated chromosome, consisting of two sister chromatids held together along their length

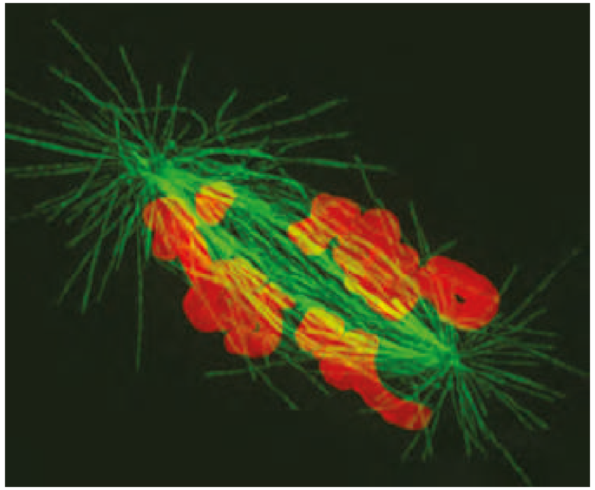
At prophase, the replicated chromosomes, each consisting of two closely associated sister chromatids, condense. Outside the nucleus, the mitotic spindle assembles between the two centrosomes, which have replicated and moved apart. For simplicity, only three chromosomes are shown. In diploid cells, there would be two copies of each chromosome present. In the fluorescence micrograph, chromosomes are stained *orange* and microtubules are *green*.



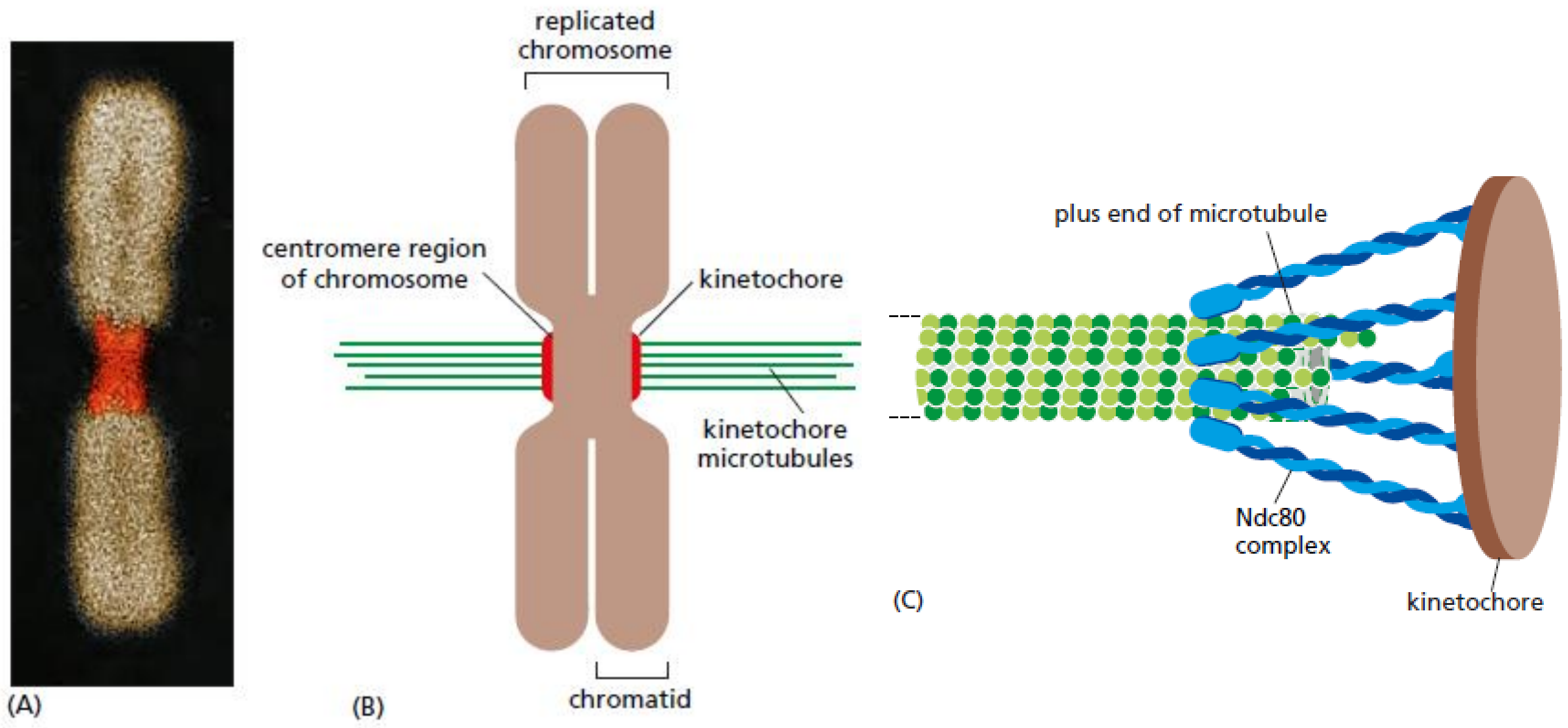
2 PROMETAPHASE



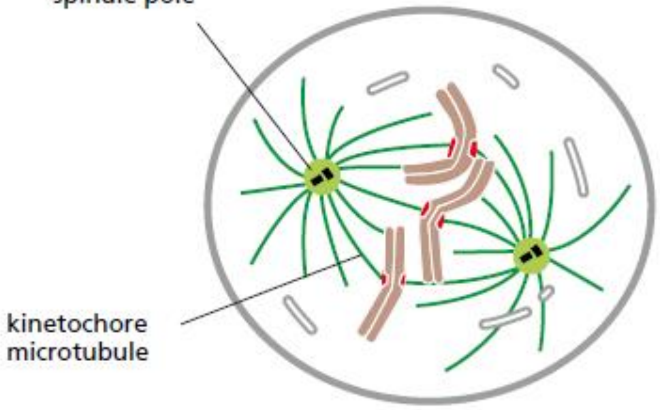
Prometaphase starts abruptly with the breakdown of the nuclear envelope. Chromosomes can now attach to spindle microtubules via their kinetochores and undergo active movement.



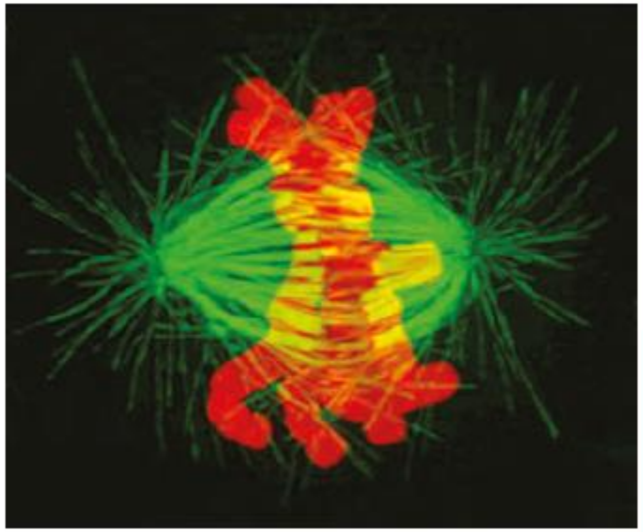
The kinetochore



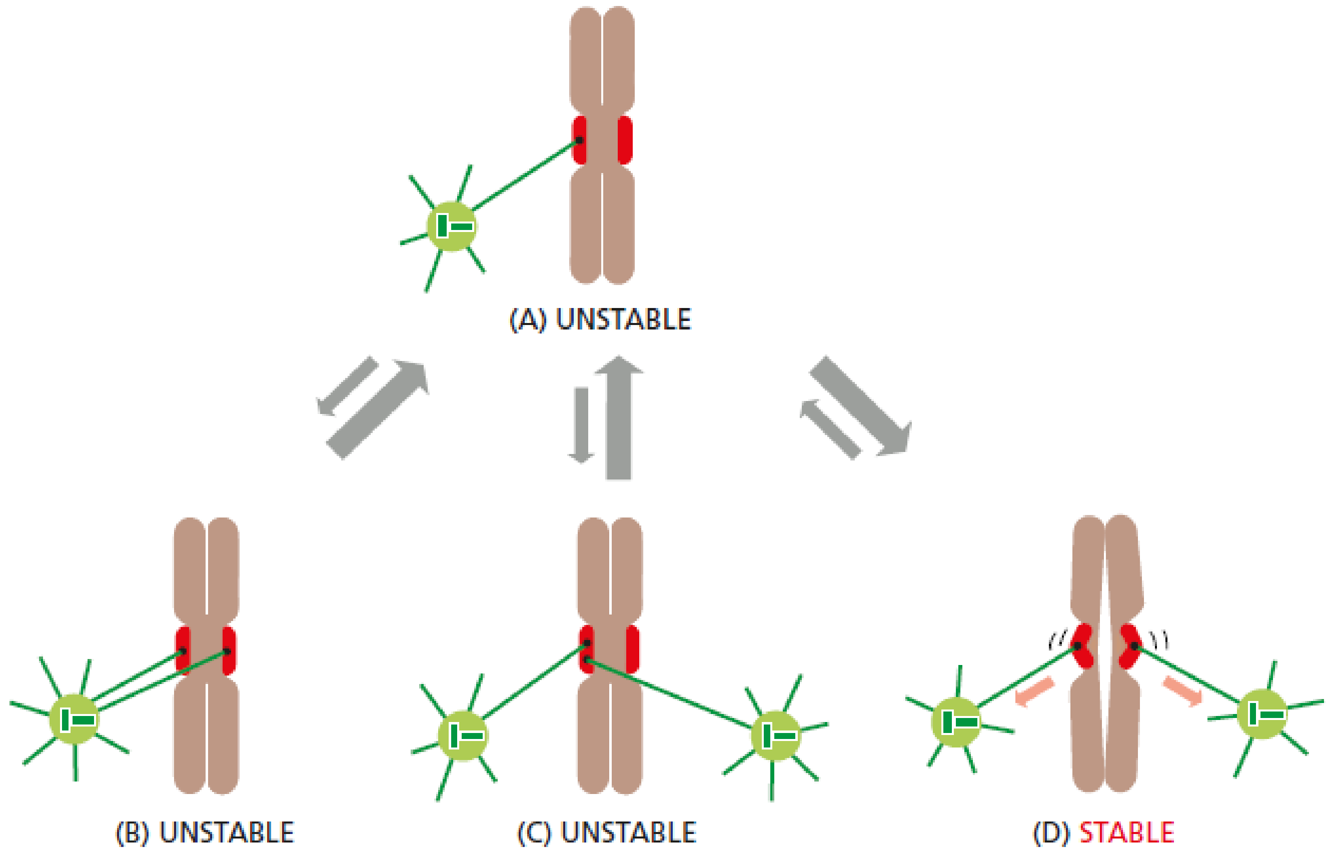
3 METAPHASE



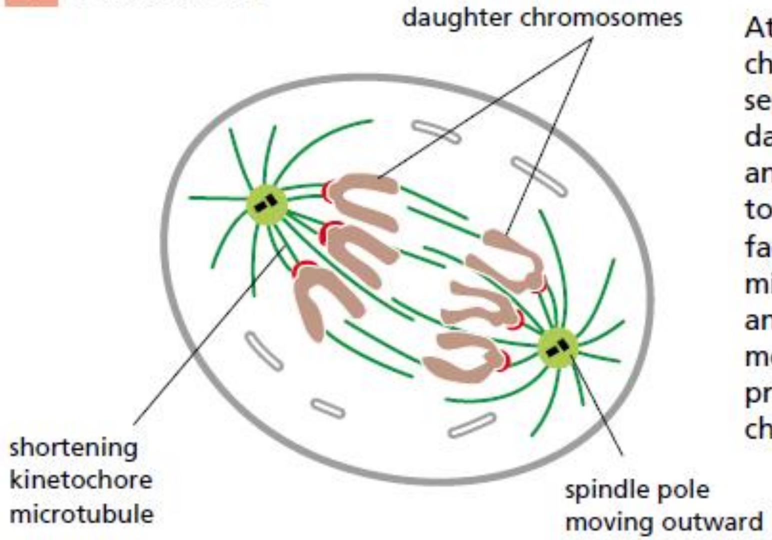
At metaphase, the chromosomes are aligned at the equator of the spindle, midway between the spindle poles. The kinetochore microtubules attach sister chromatids to opposite poles of the spindle.



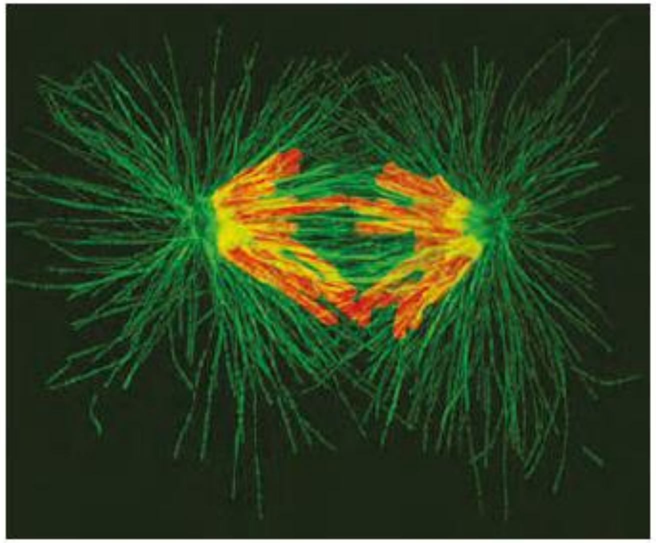
Alternative forms of kinetochore attachment to the spindle poles



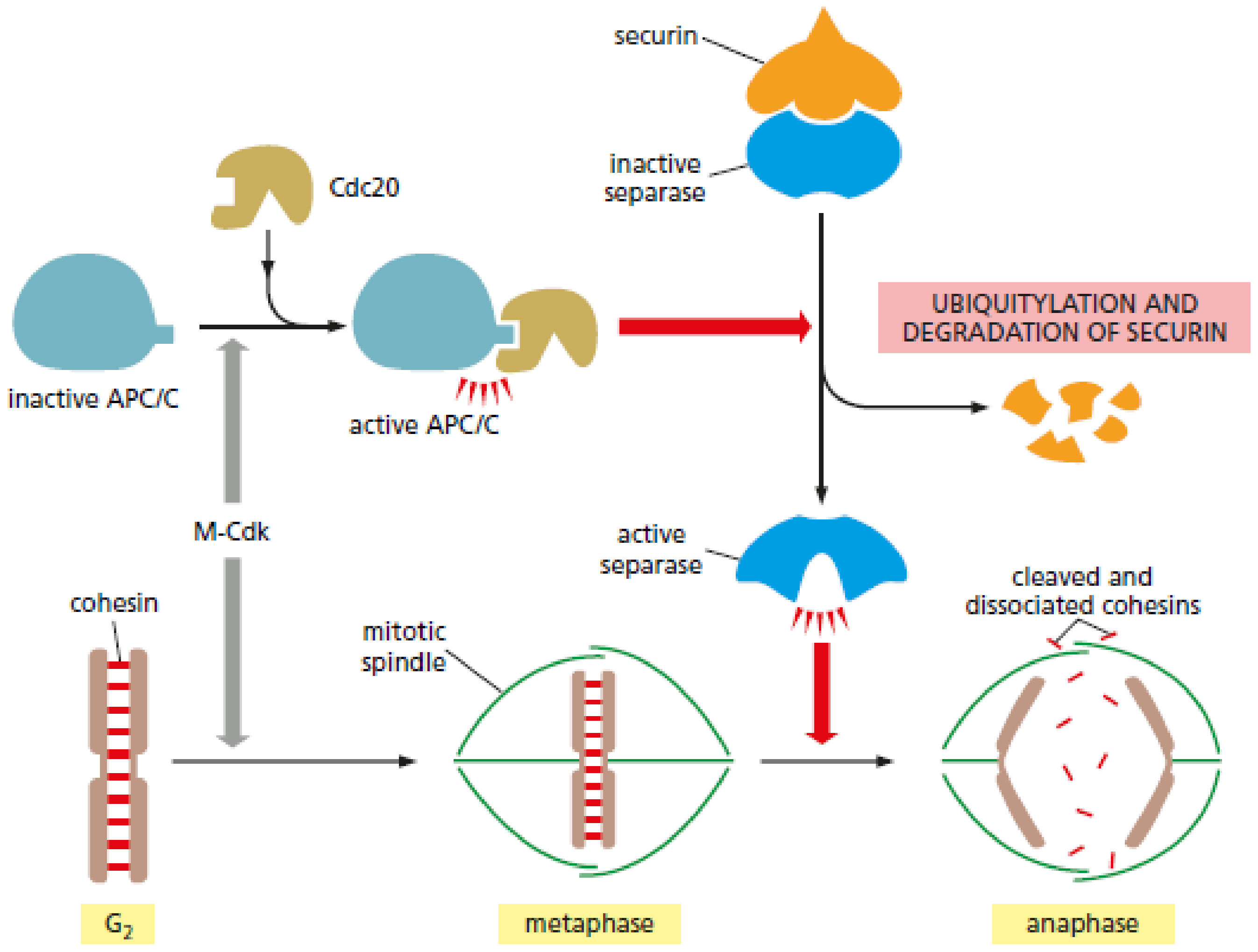
4 ANAPHASE



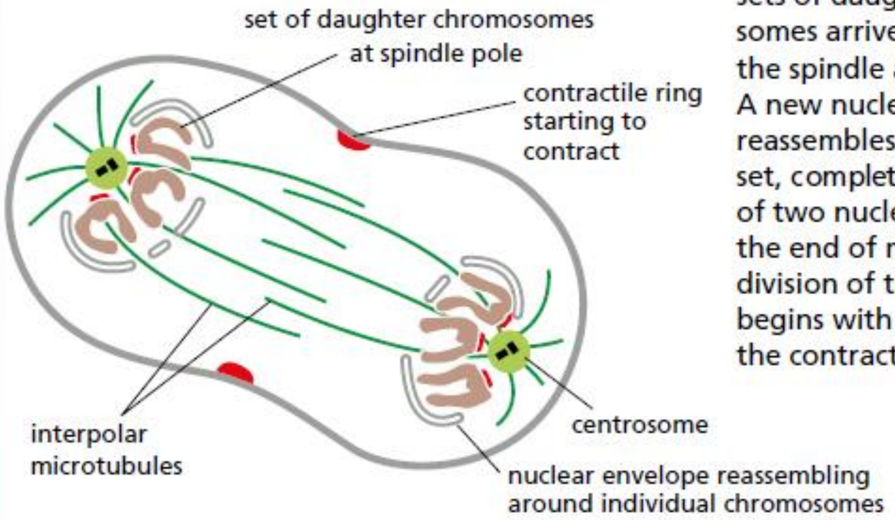
At anaphase, the sister chromatids synchronously separate to form two daughter chromosomes, and each is pulled slowly toward the spindle pole it faces. The kinetochore microtubules get shorter, and the spindle poles also move apart; both processes contribute to chromosome segregation.



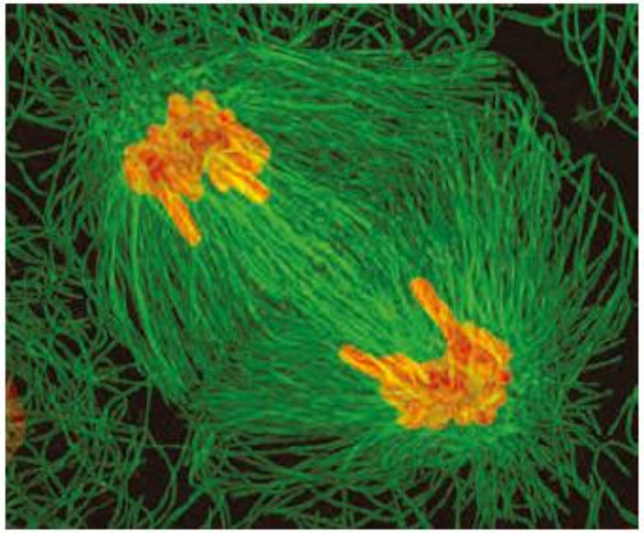
The initiation of sister chromatid separation by the APC/C



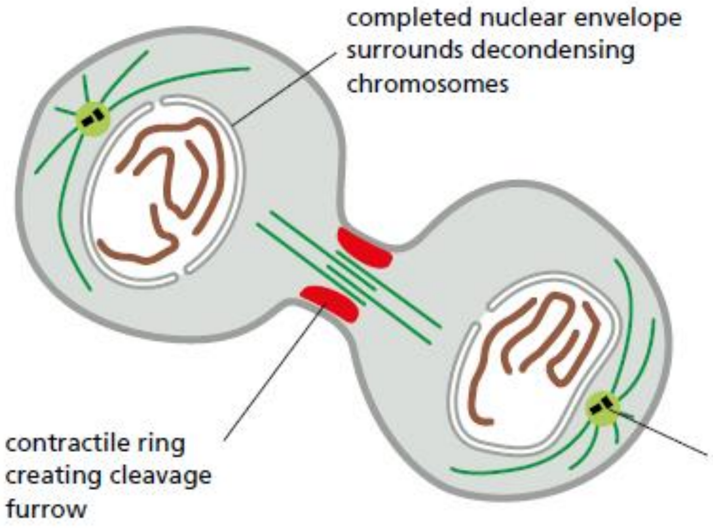
5 TELOPHASE



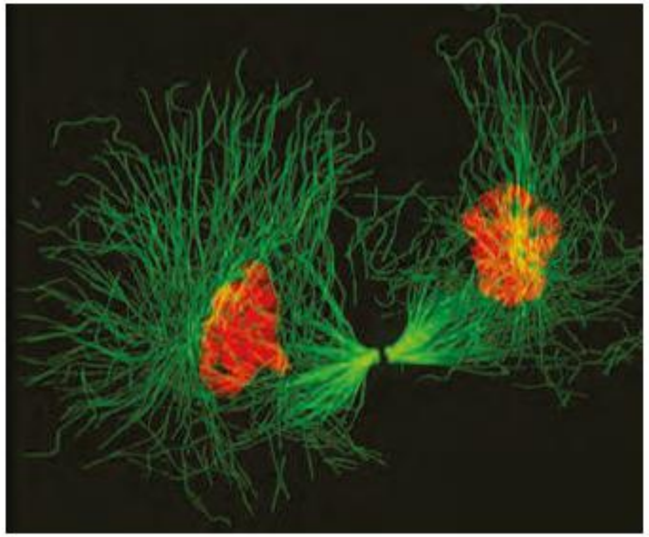
During telophase, the two sets of daughter chromosomes arrive at the poles of the spindle and decondense. A new nuclear envelope reassembles around each set, completing the formation of two nuclei and marking the end of mitosis. The division of the cytoplasm begins with contraction of the contractile ring.



6 CYTOKINESIS

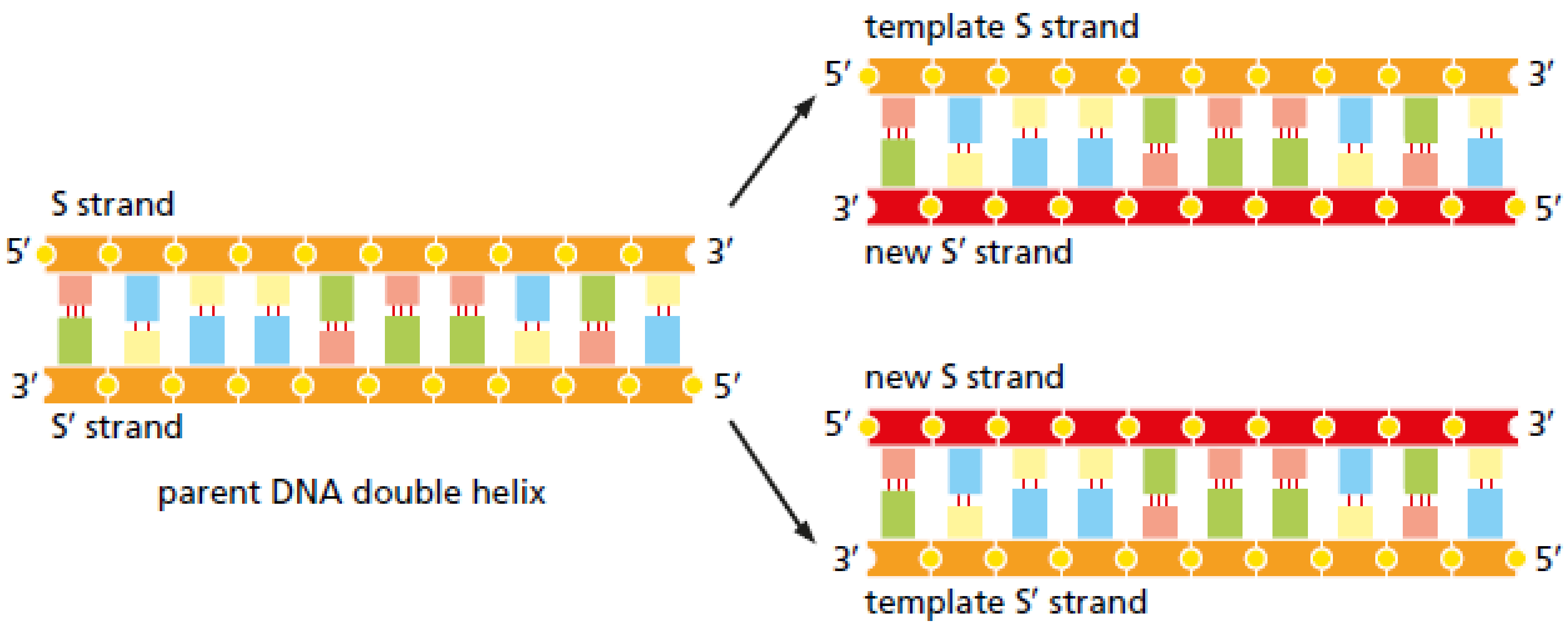


During cytokinesis, the cytoplasm is divided in two by a contractile ring of actin and myosin filaments, which pinches the cell in two to create two daughters, each with one nucleus.



(Micrographs courtesy of Julie Canman and Ted Salmon.)

DNA replication

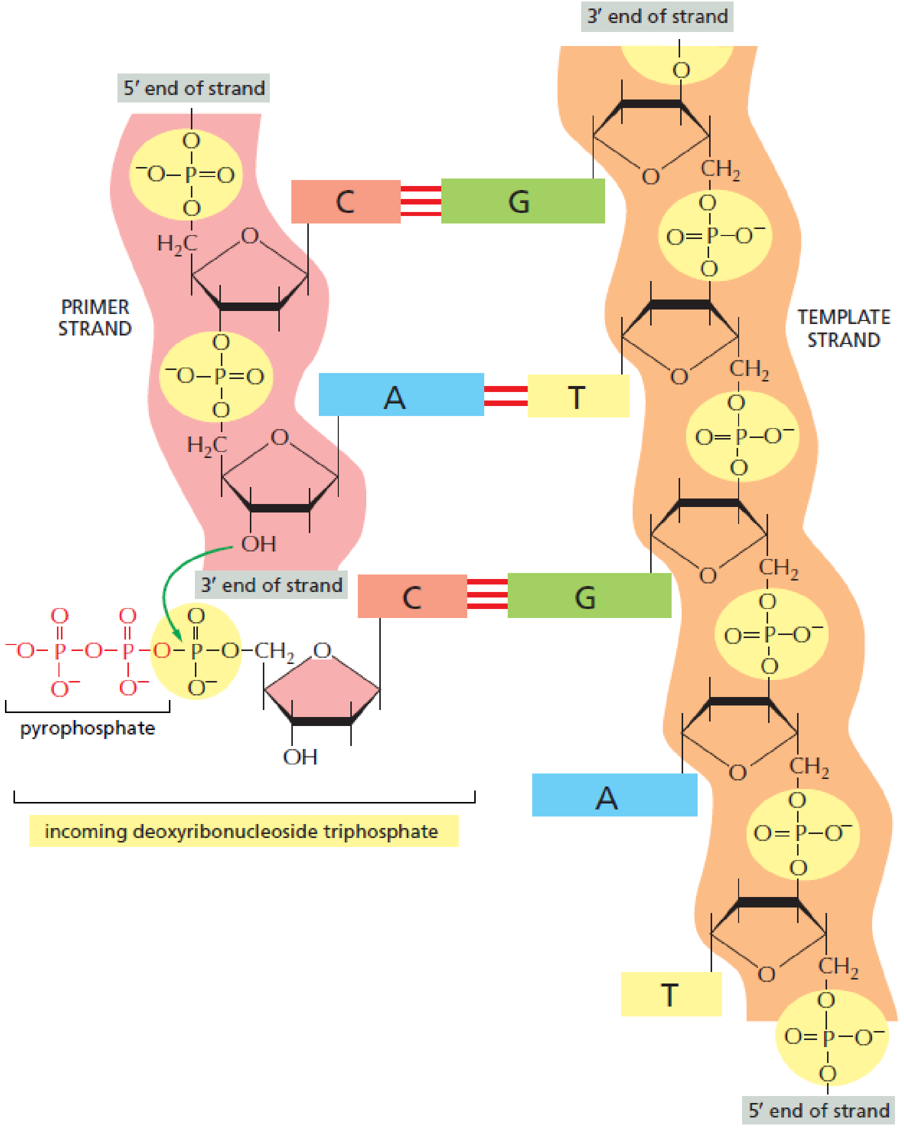


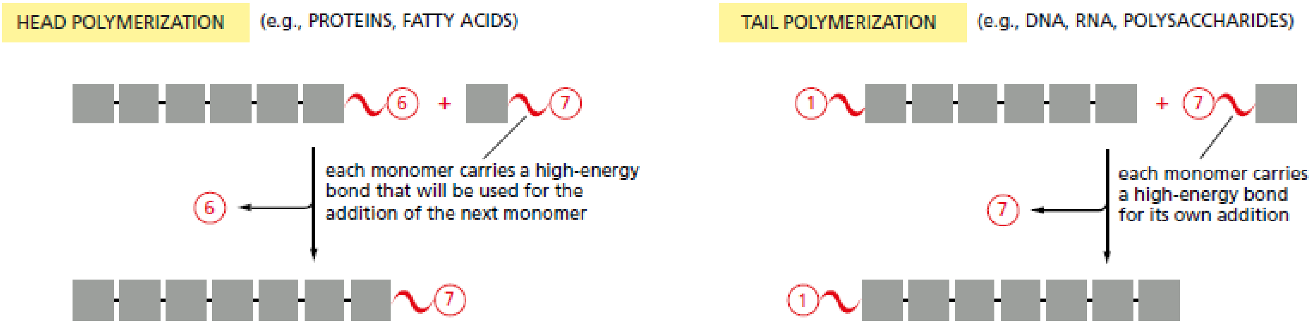
Base-Pairing Underlies DNA Replication and DNA Repair

Duplicating DNA at rates of 100 nucleotides per second (in bacteria 500-1000)

Human genome has approximately 10,000 such origins—an average of 220 origins per chromosome (bacteria – 1)

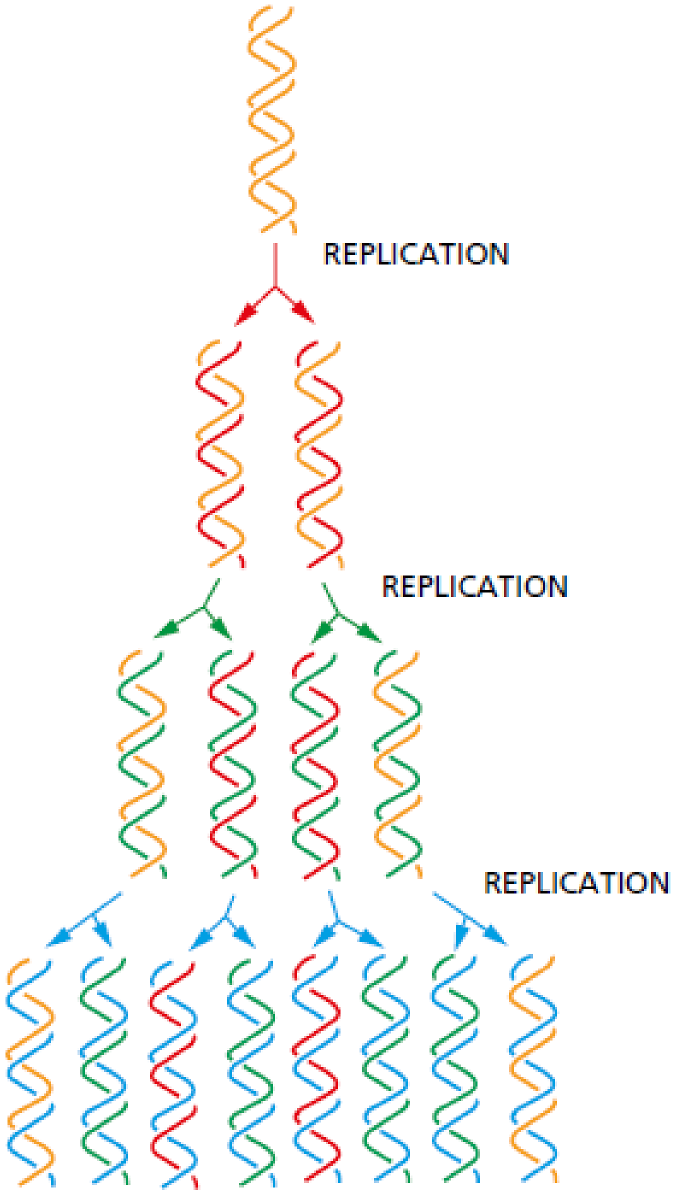
A-T-rich stretches of DNA are typically found at replication origins


3′ end of strand
5′ end of strand
PRIMER STRAND
TEMPLATE STRAND
C
G
A
T
C
G
A
T
OH
3′ end of strand
pyrophosphate
OH
incoming deoxyribonucleoside triphosphate
5′ end of strand

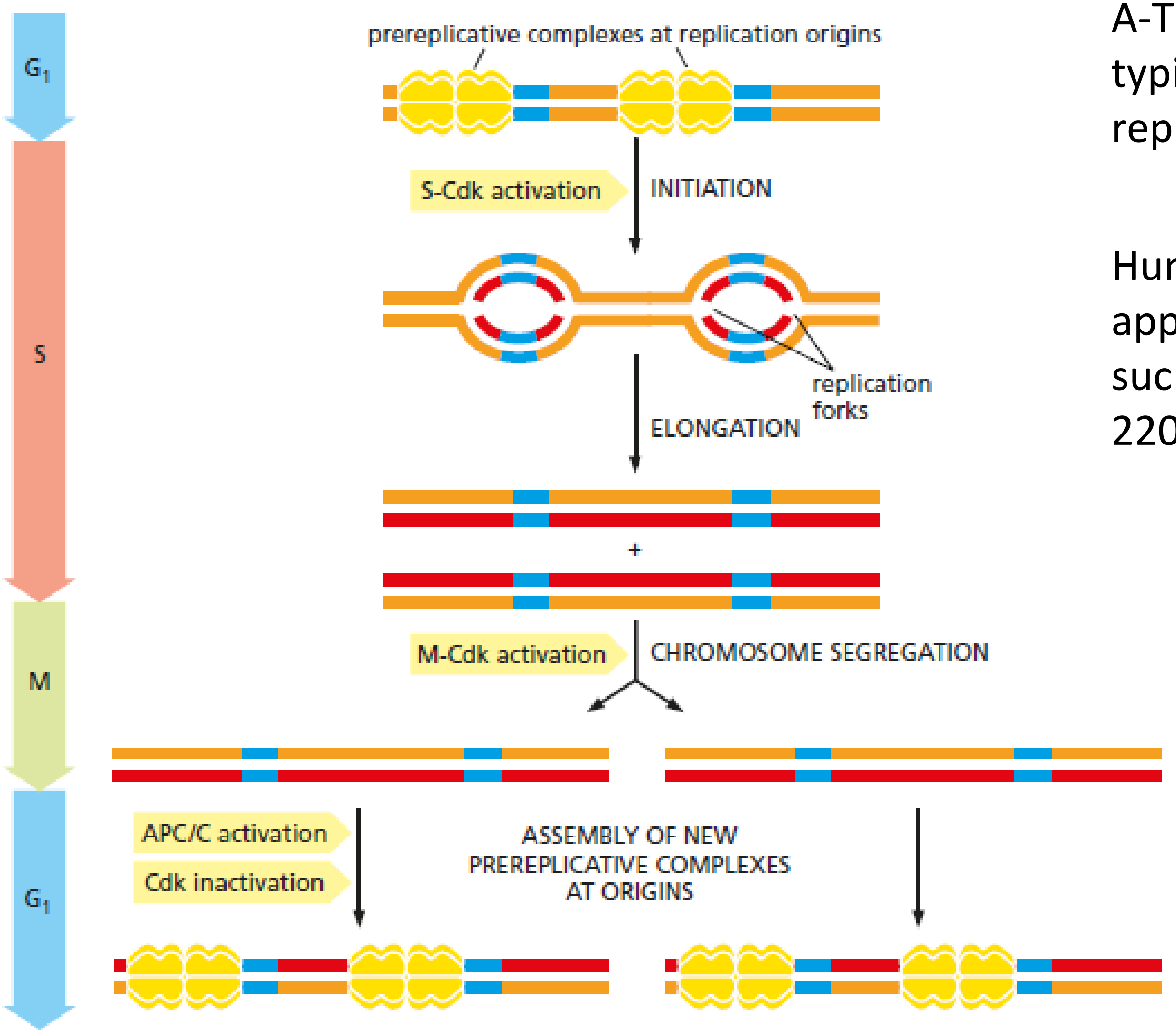

HEAD POLYMERIZATION
(e.g., PROTEINS, FATTY ACIDS)
each monomer carries a high-energy bond that will be used for the addition of the next monomer
TAIL POLYMERIZATION
(e.g., DNA, RNA, POLYSACCHARIDES)
each monomer carries a high-energy bond for its own addition

DNA double helix is said to be replicated "semiconservatively"

The original strands remain intact through many cell generations



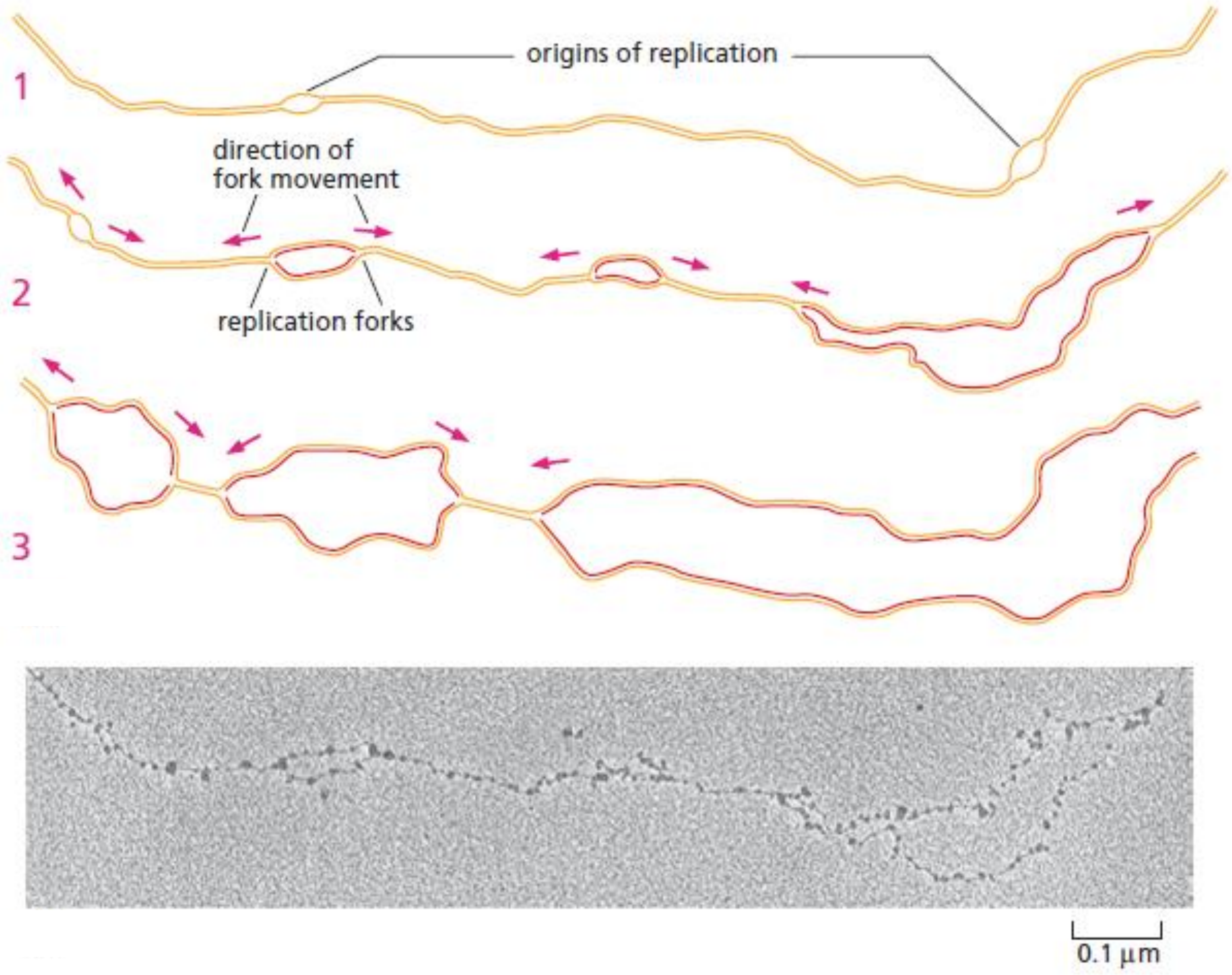
DNA Synthesis Begins at Replication Origins



A-T-rich stretches of DNA are typically found at replication origins

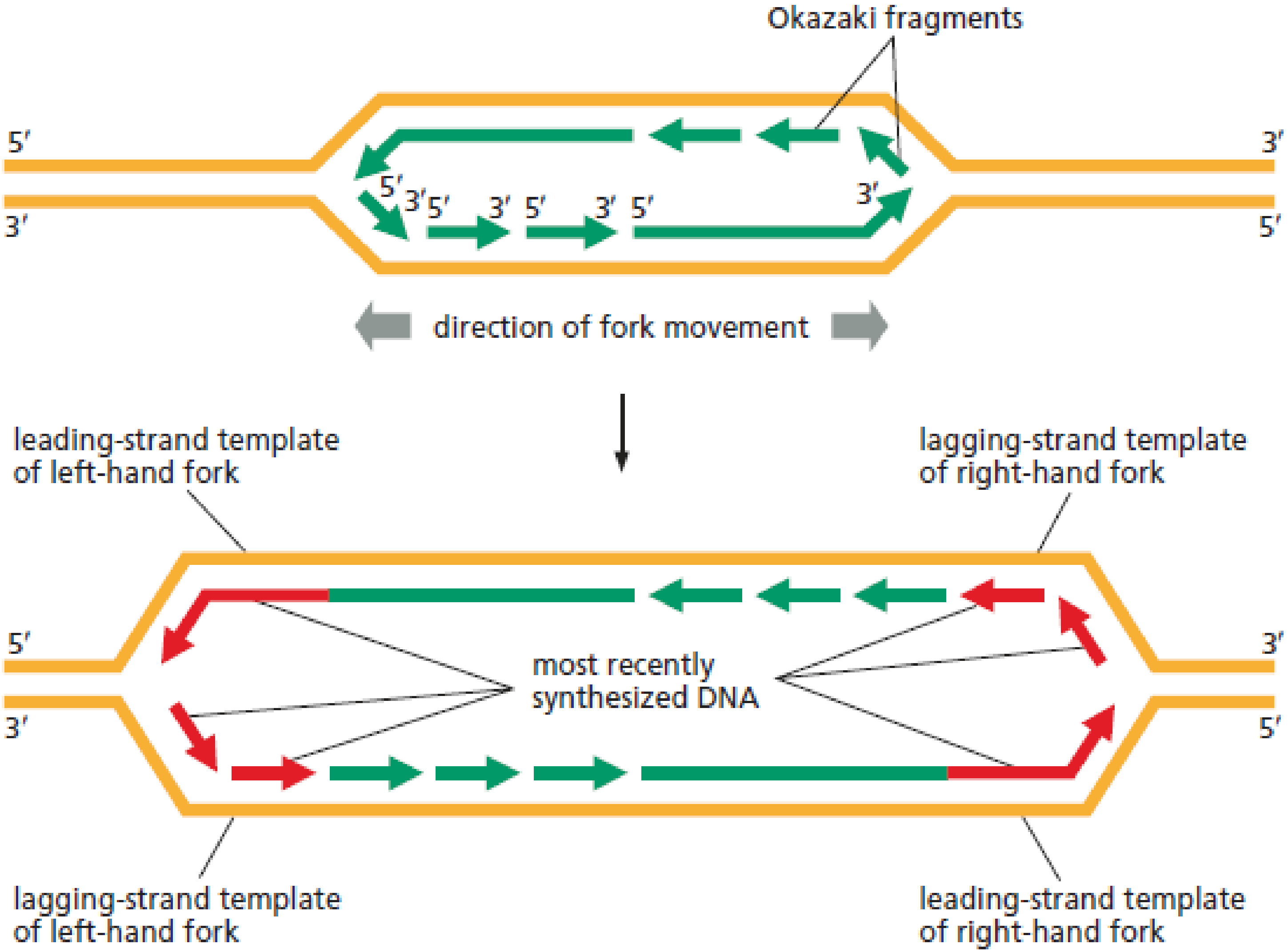
Human genome has approximately 10,000 such origins—an average of 220 origins per chromosome

Replication forks

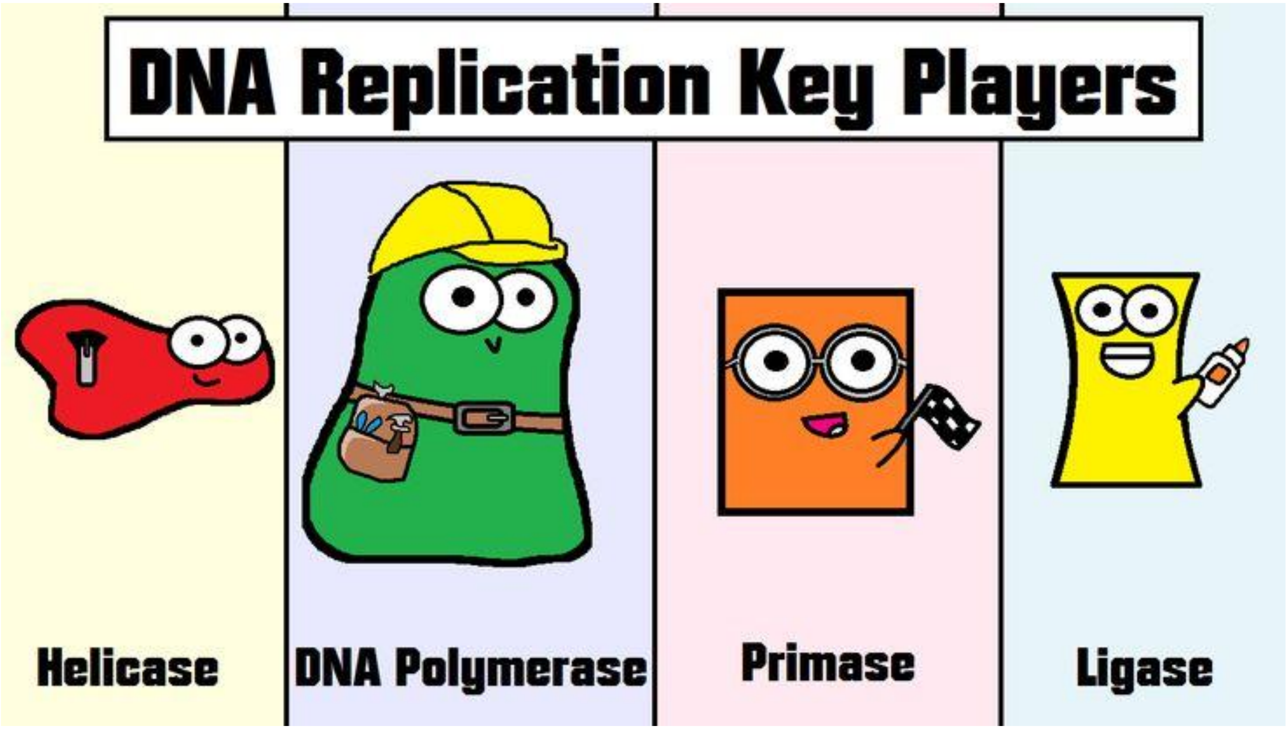


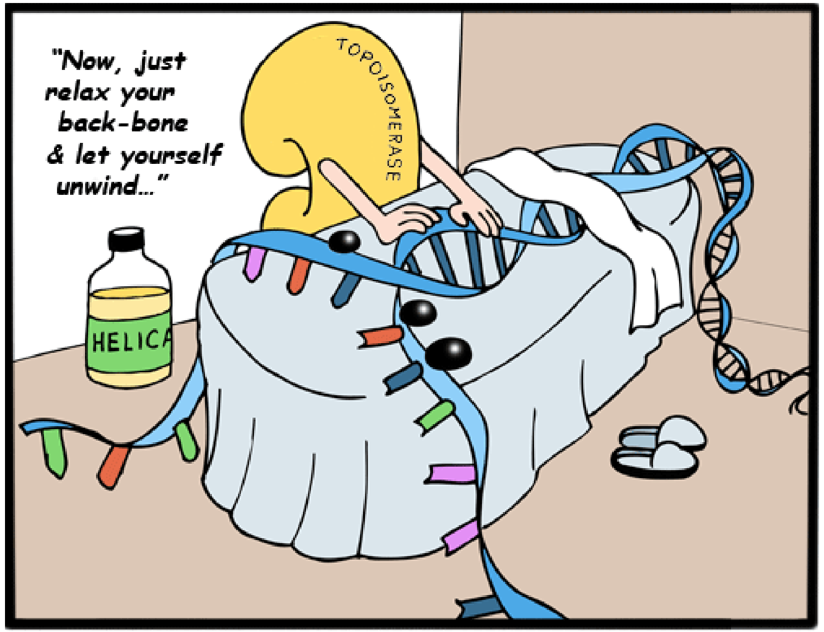
Two Replication Forks Form at each Replication Origin and move away from the origin in opposite directions-**bidirectional**

The Replication Fork Is Asymmetrical

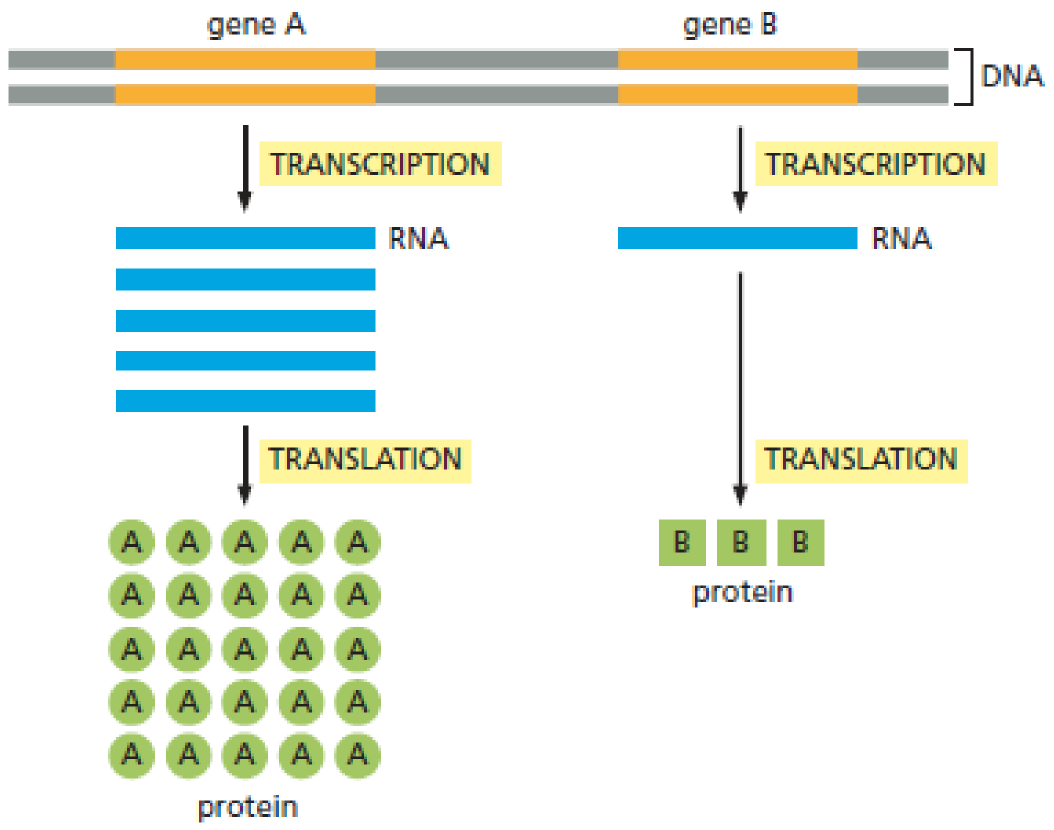


DNA polymerase uses a **backstitching mechanism**: it synthesizes short pieces of DNA (called Okazaki fragments) in the 5′-to-3′ direction and then moves back along the template strand (toward the fork) before synthesizing the next fragment.

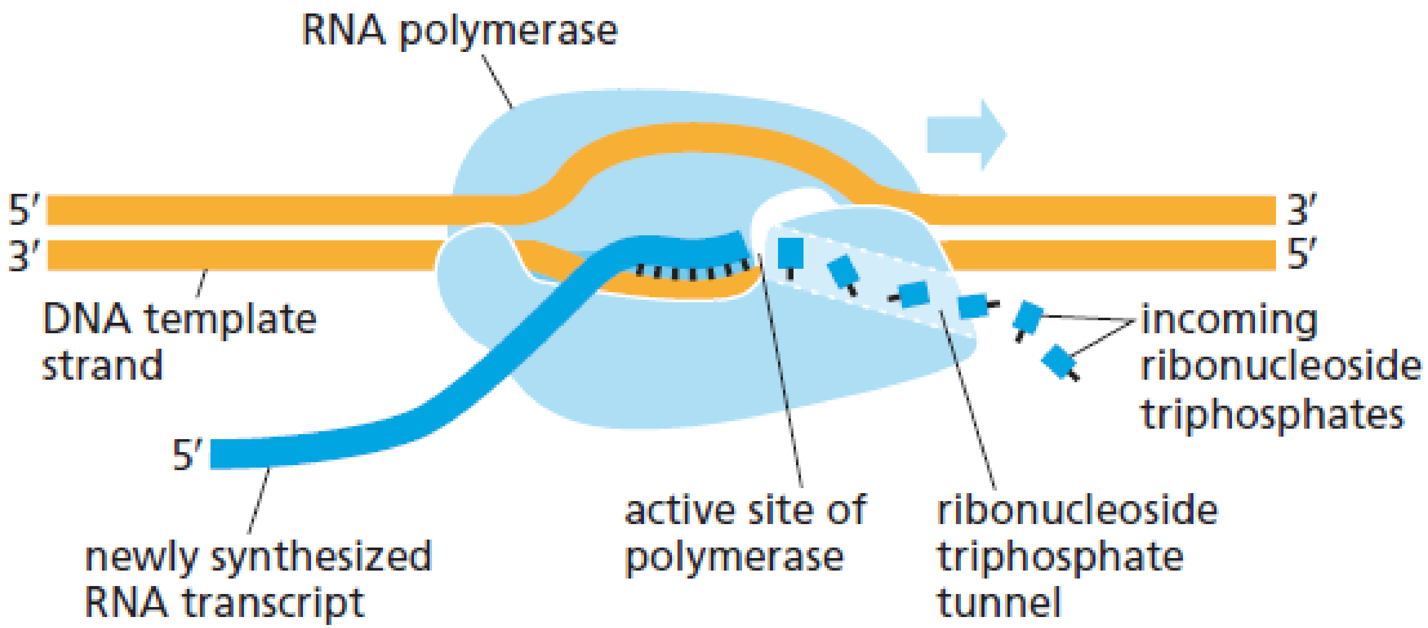

DNA Replication Key Players
Helicase
DNA Polymerase
Primase
Ligase


"Now, just relax your back-bone & let yourself unwind..."
TOPOISOMERASE
HELICA

How Cells Read the Genome



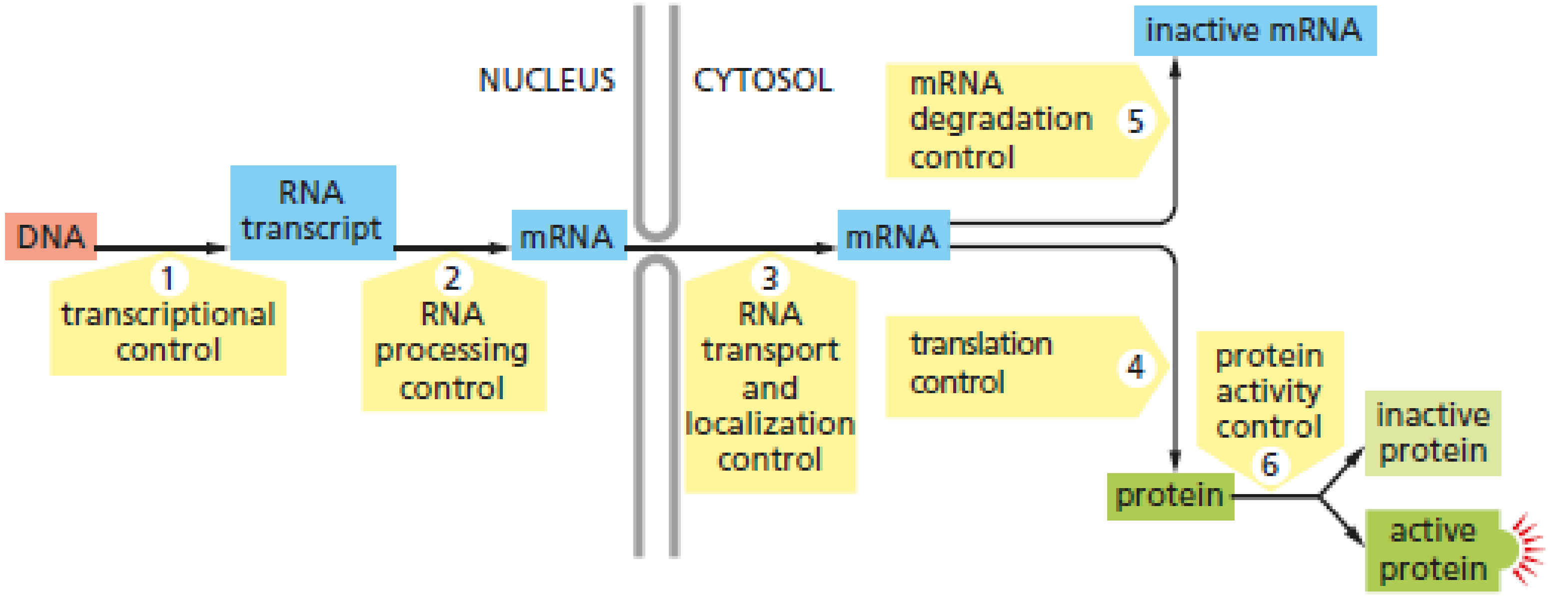
DNA is transcribed by the enzyme RNA polymerase



RNA polymerases make about one mistake for every 10^4 nucleotides copied into RNA

Moves at the rate of 50 nucleotides per second, over a thousand transcripts can be synthesized in an hour from a single gene.

Six steps at which eukaryotic gene expression can be controlled



Principal Types of RNAs Produced in Cells

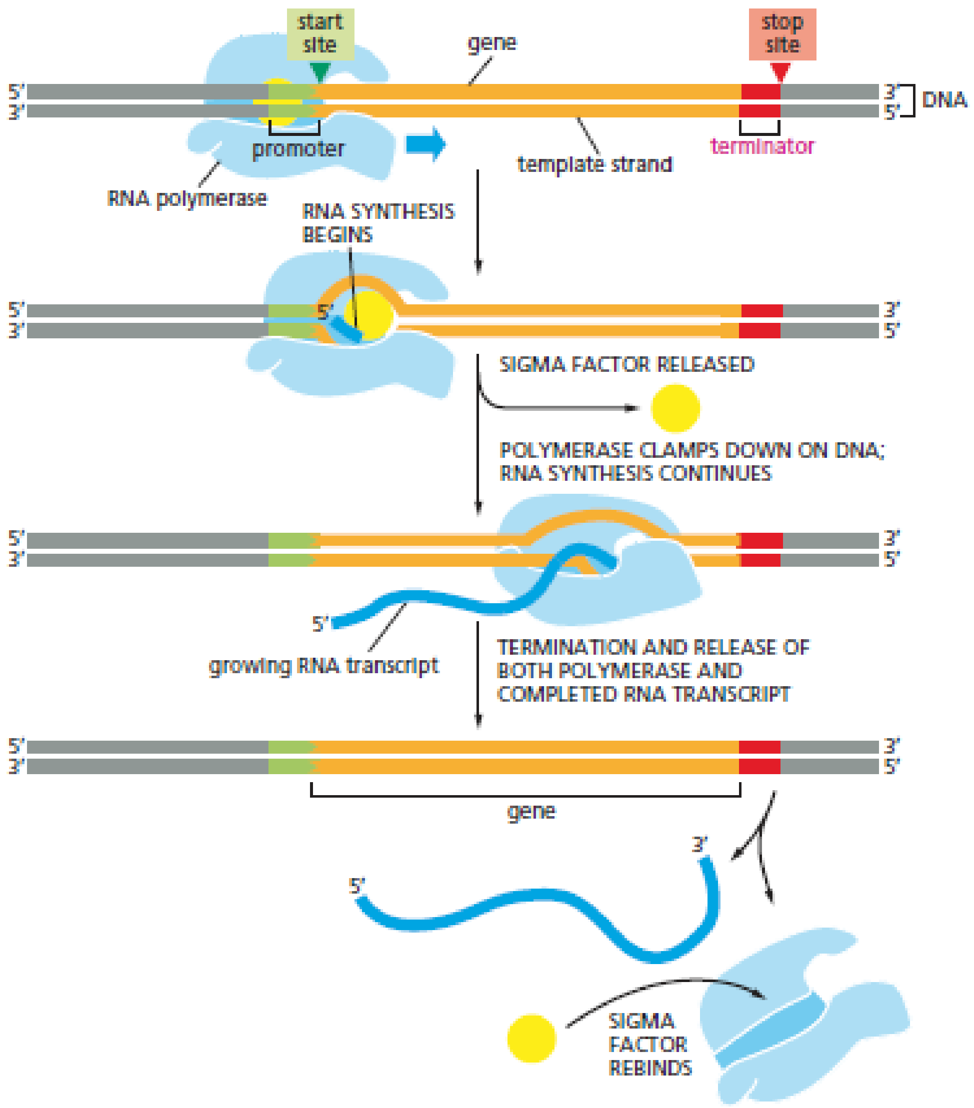
Type of RNA	Function
mRNAs	Messenger RNAs, code for proteins
rRNAs	Ribosomal RNAs, form the basic structure of the ribosome and catalyze protein synthesis
tRNAs	Transfer RNAs, central to protein synthesis as adaptors between mRNA and amino acids
snRNAs	Small nuclear RNAs, function in a variety of nuclear processes, including the splicing of pre-mRNA
snoRNAs	Small nucleolar RNAs, help to process and chemically modify rRNAs
miRNAs	MicroRNAs, regulate gene expression by blocking translation of specific mRNAs and cause their degradation
siRNAs	Small interfering RNAs, turn off gene expression by directing the degradation of selective mRNAs and the establishment of compact chromatin structures
piRNAs	Piwi-interacting RNAs, bind to piwi proteins and protect the germ line from transposable elements
lncRNAs	Long noncoding RNAs, many of which serve as scaffolds; they regulate diverse cell processes, including X-chromosome inactivation

Most of the RNA in cells is rRNA

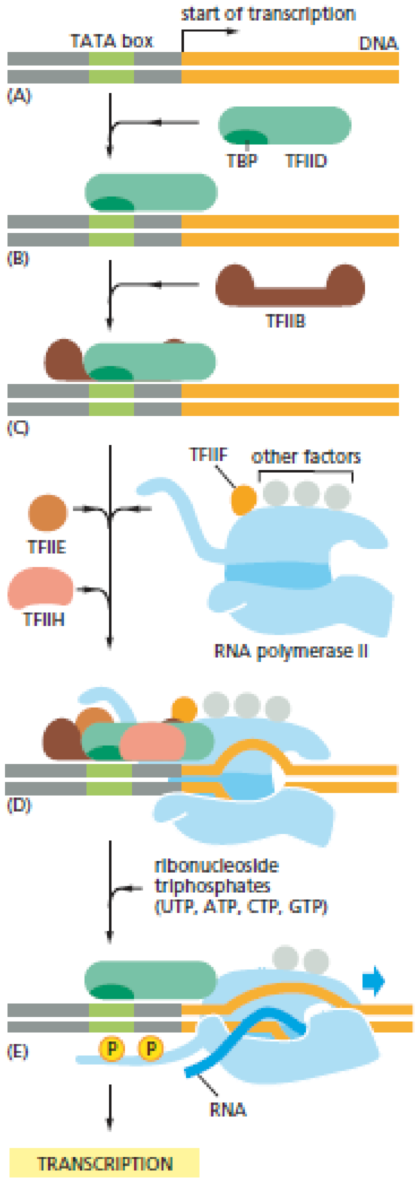
mRNA comprises only 3–5% of the total RNA in a typical mammalian cell

On average only 10–15 molecules of each species of mRNA present in each cell

Signals in DNA Tell RNA Polymerase Where to Start and Finish Transcription



Initiation of Eukaryotic Gene Transcription Is a Complex Process

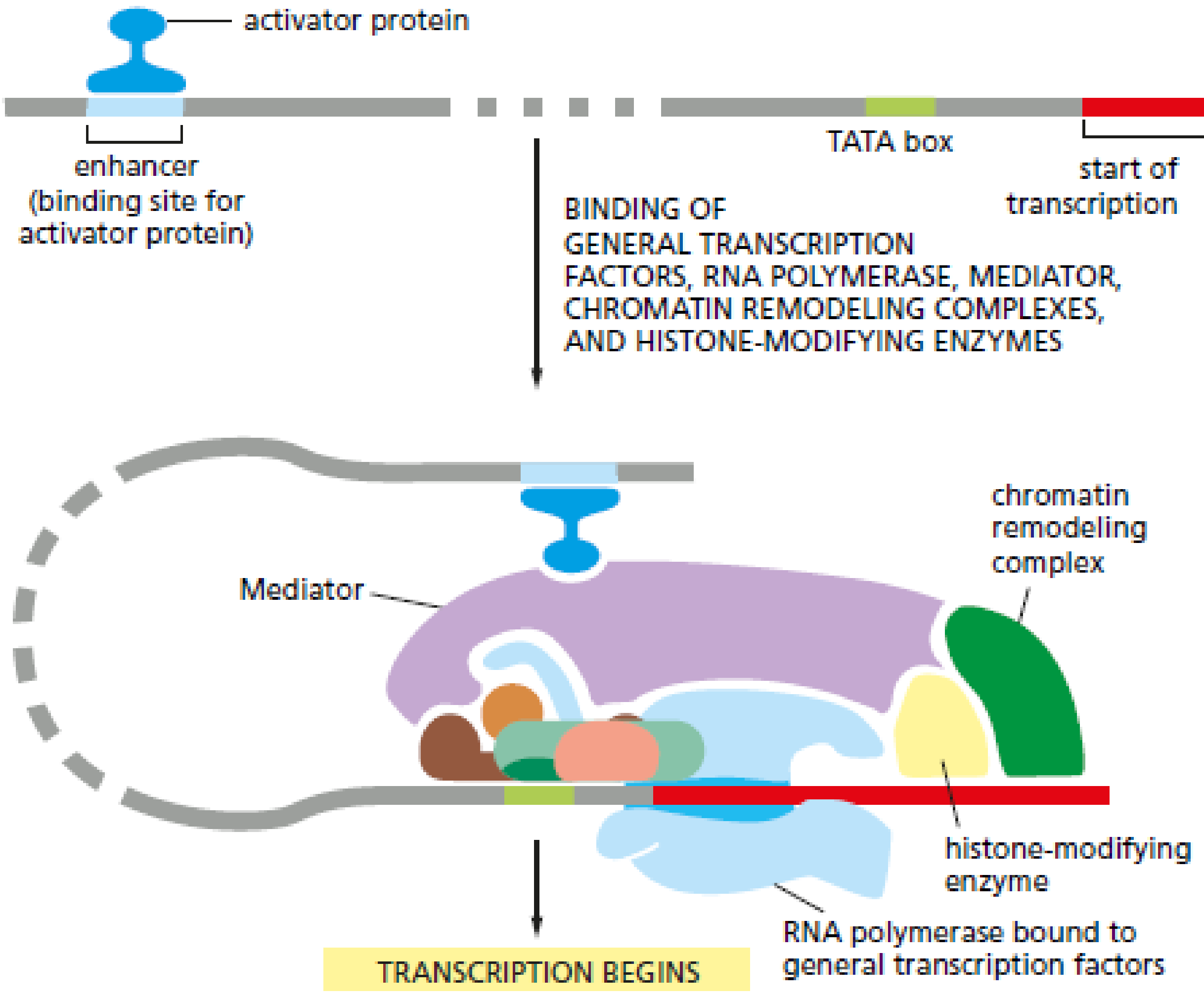


TBP is TATA binding protein

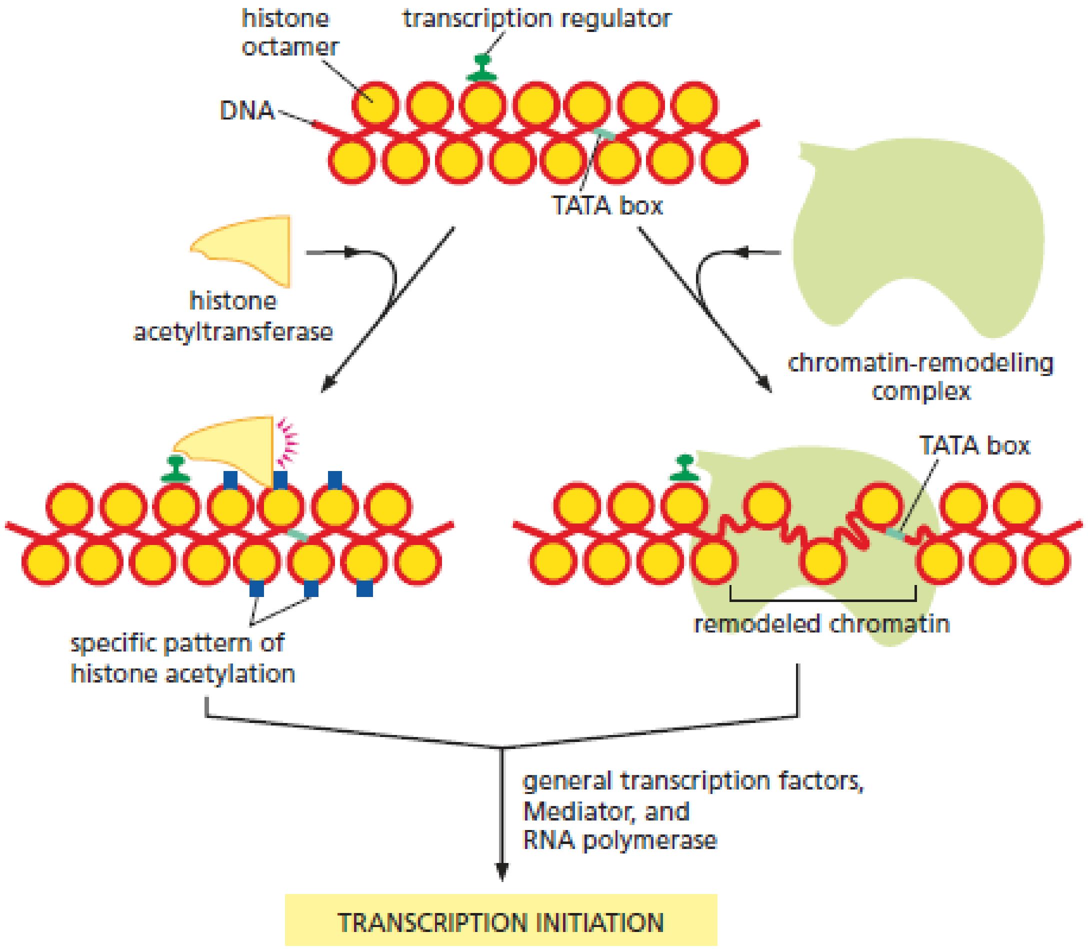
The general transcription factors help to :
(1) position eukaryotic RNA polymerase correctly at the promoter
(2) aid in pulling apart the two strands of DNA to allow transcription to begin and
(3) release RNA polymerase from the promoter to start its elongation mode

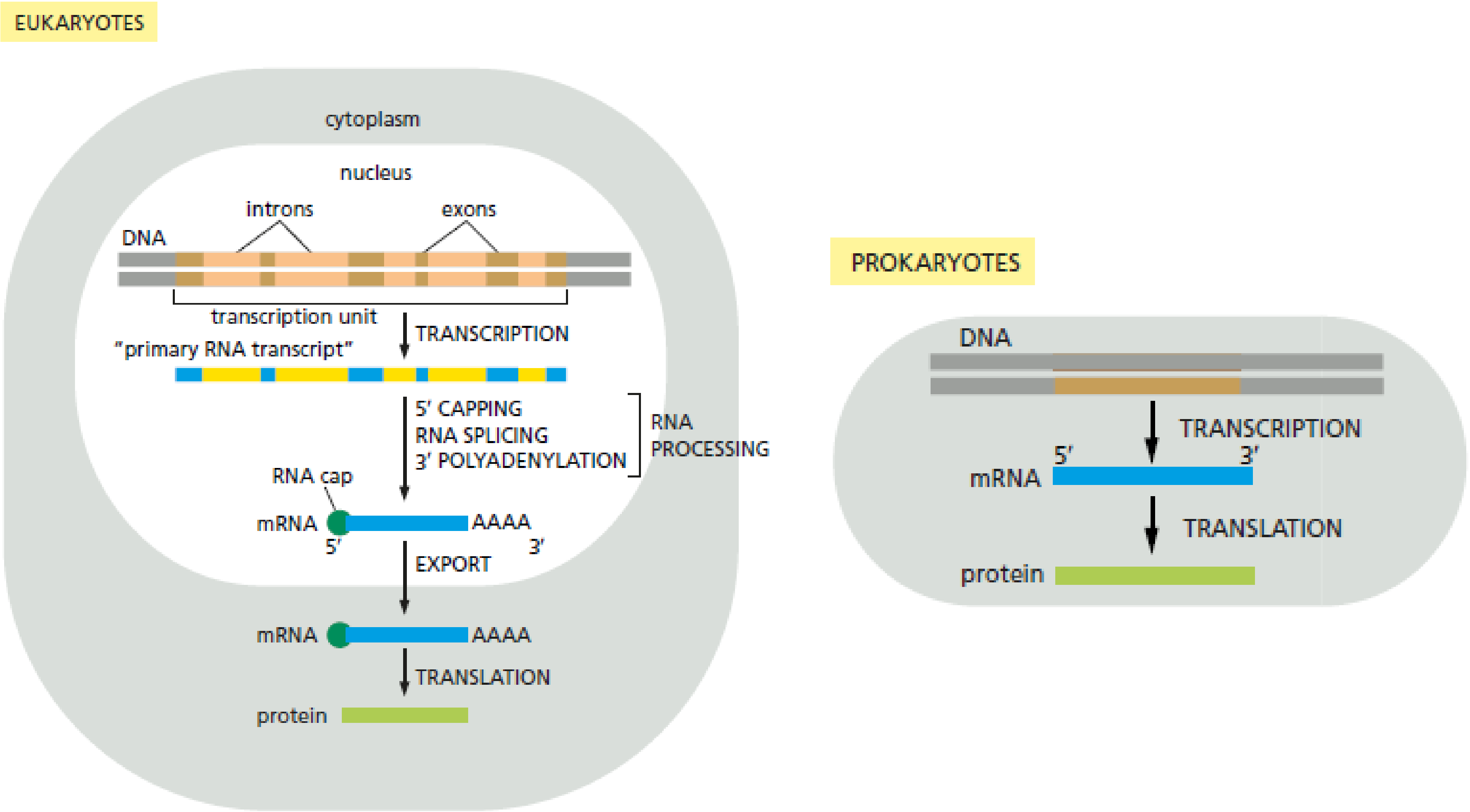
The TATA box is typically located 25 nucleotides upstream from the transcription start site.

Polymerase II Also Requires Activator, Mediator, and Chromatin- Modifying Proteins

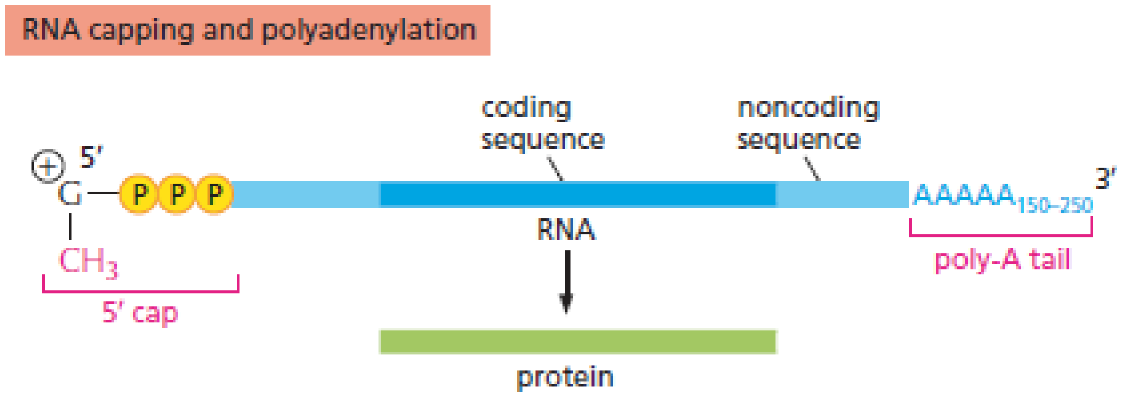


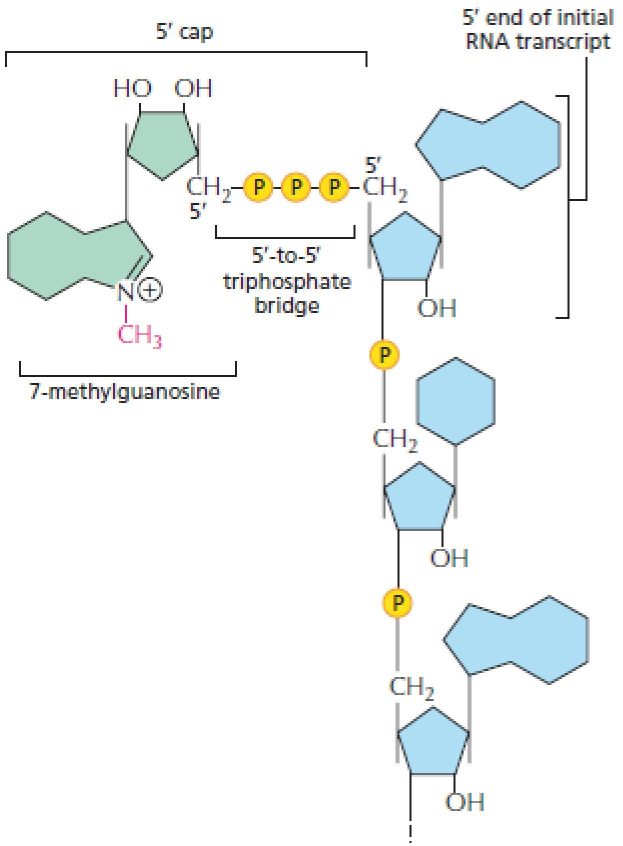
Eukaryotic transcriptional activators can recruit chromatin modifying proteins to help initiate gene transcription



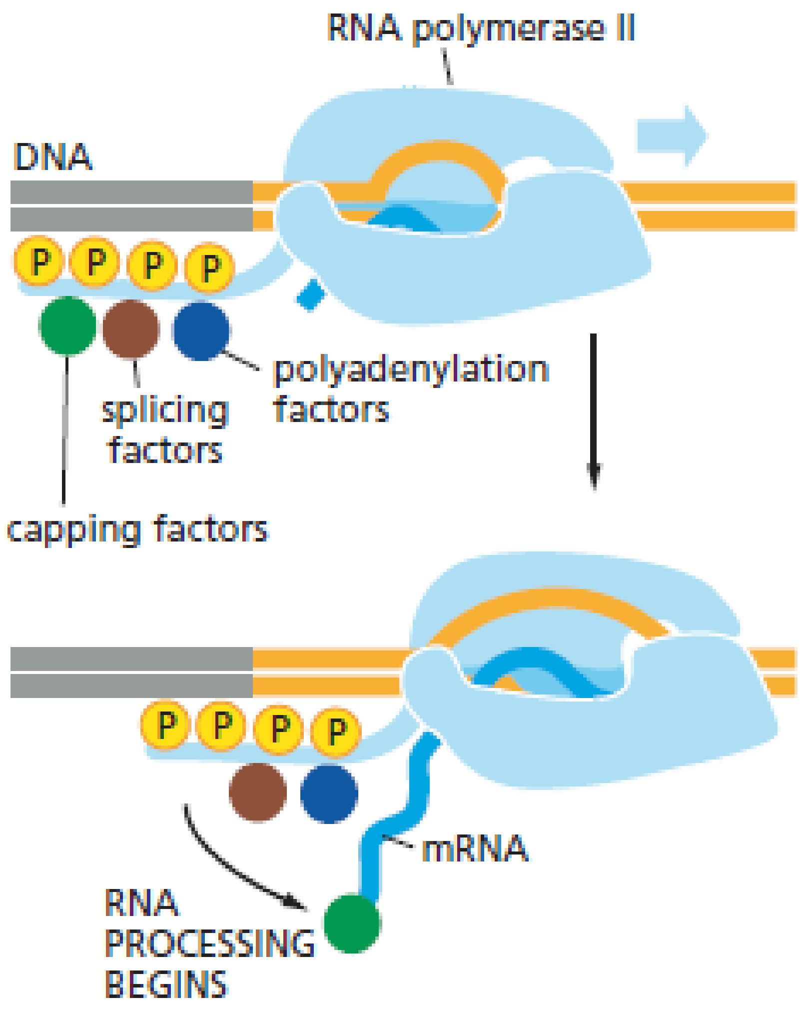

EUKARYOTES
cytoplasm
nucleus
introns
exons
DNA
transcription unit
TRANSCRIPTION
"primary RNA transcript"
5′ CAPPING
RNA SPLICING
3′ POLYADENYLATION
RNA PROCESSING
RNA cap
mRNA
AAAA
5′
3′
EXPORT
mRNA
AAAA
TRANSLATION
protein
PROKARYOTES
DNA
TRANSCRIPTION
5′
3′
mRNA
TRANSLATION
protein

Transcription Elongation in Eukaryotes Is Tightly Coupled to RNA Processing





Phosphorylation of the tail of RNA polymerase II allows RNA-processing proteins to assemble there



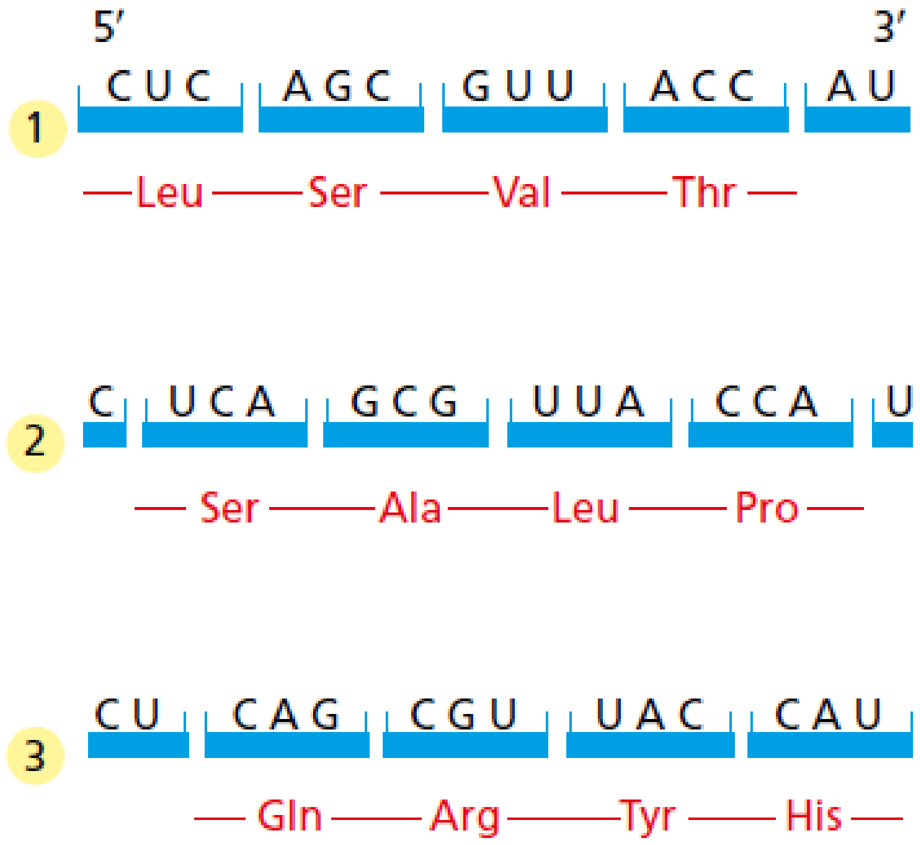
From RNA to protein

	AGA									UUA					AGC					
	AGG									UUG					AGU					
GCA	CGA						GGA			CUA				CCA	UCA	ACA			GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC			GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		UUC	CCG	UCG	ACG		UAC	GUG	UAG
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

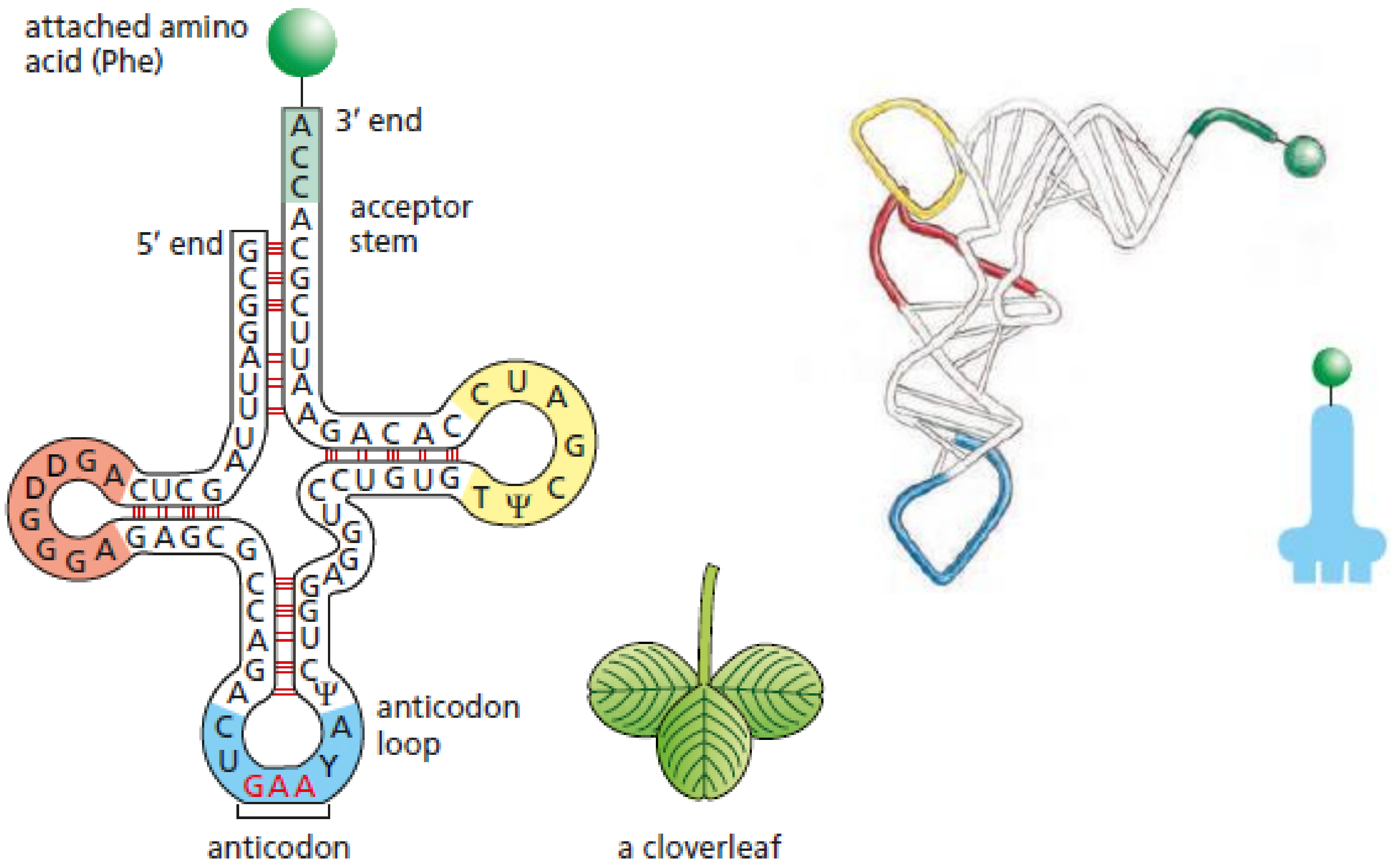
The nucleotide sequence of an mRNA is translated into the amino acid sequence of a protein via the **genetic code**

Each group of three consecutive nucleotides in RNA is called a **codon**

The three possible reading frames in protein synthesis



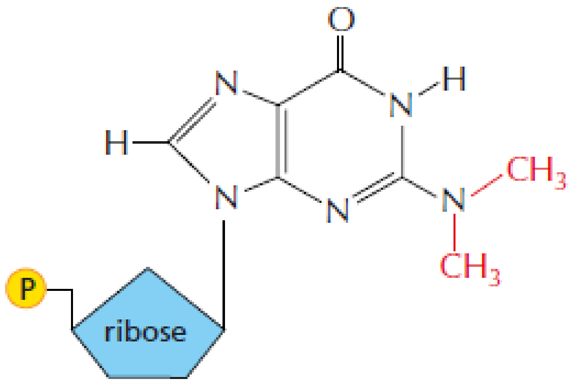
tRNA Molecules Match Amino Acids to Codons in mRNA



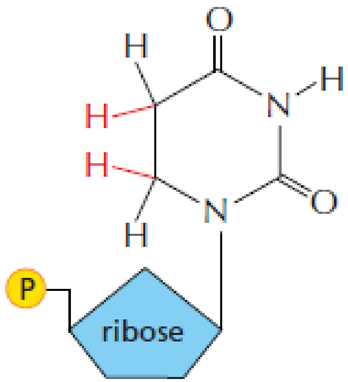
80 nucleotides in length

The bases denoted ψ (pseudouridine) and D (dihydrouridine) are derived from uracil

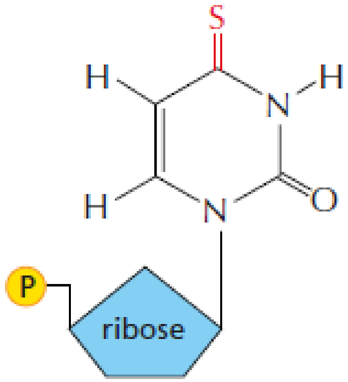
A few of the unusual nucleotides found in tRNA molecules



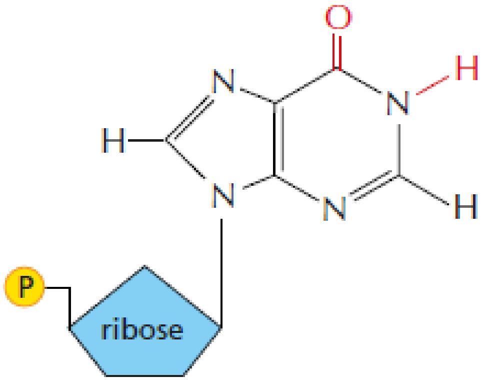
two methyl groups added to G
(*N*,*N*-dimethyl G)



two hydrogens added to U
(dihydro U)

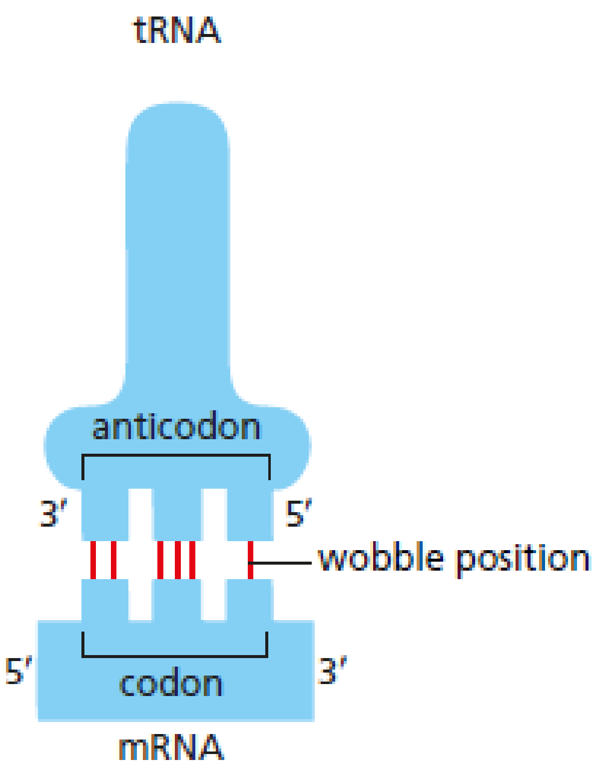


sulfur replaces oxygen in U
(4-thiouridine)



deamination of A
(inosine)

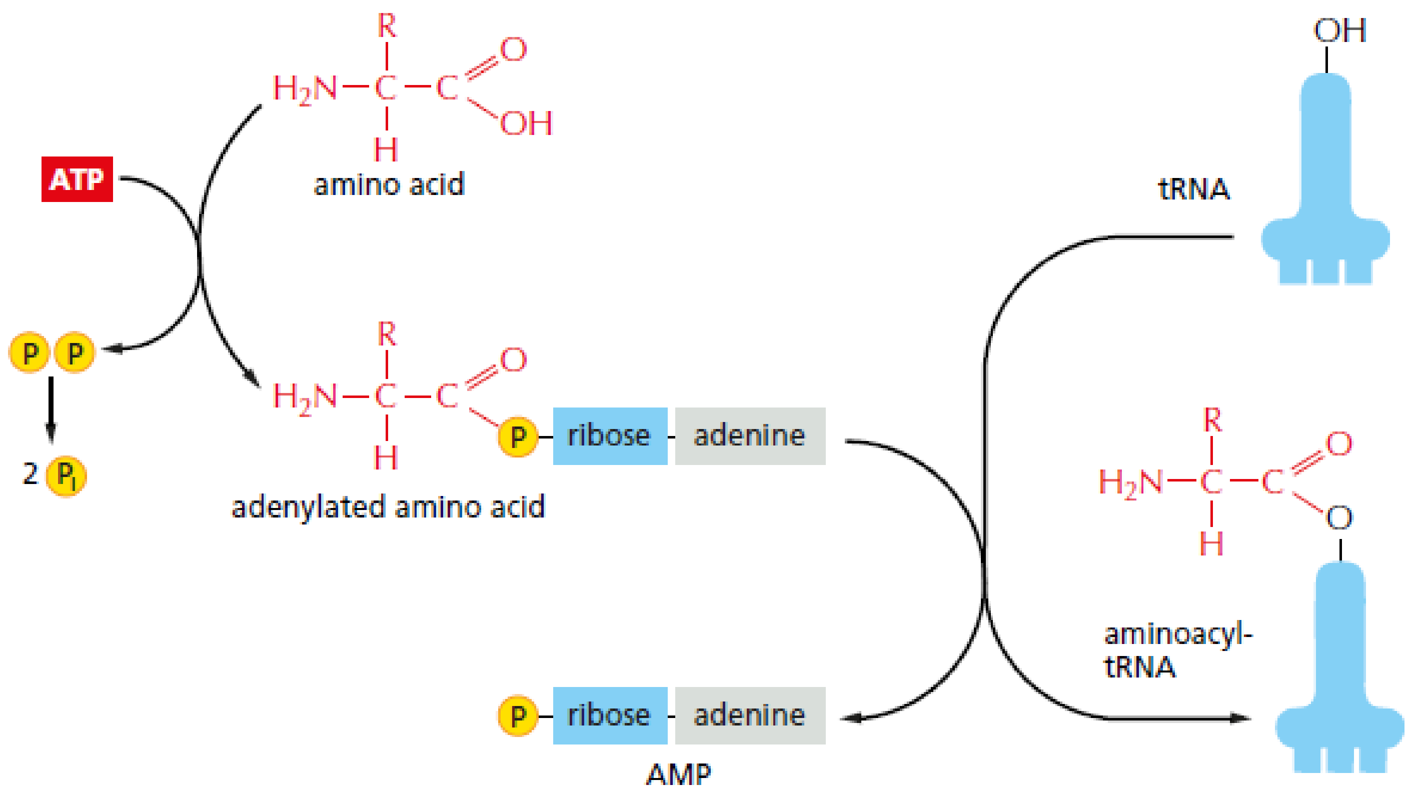
Wobble base-pairing between codons and anticodons



eukaryotes

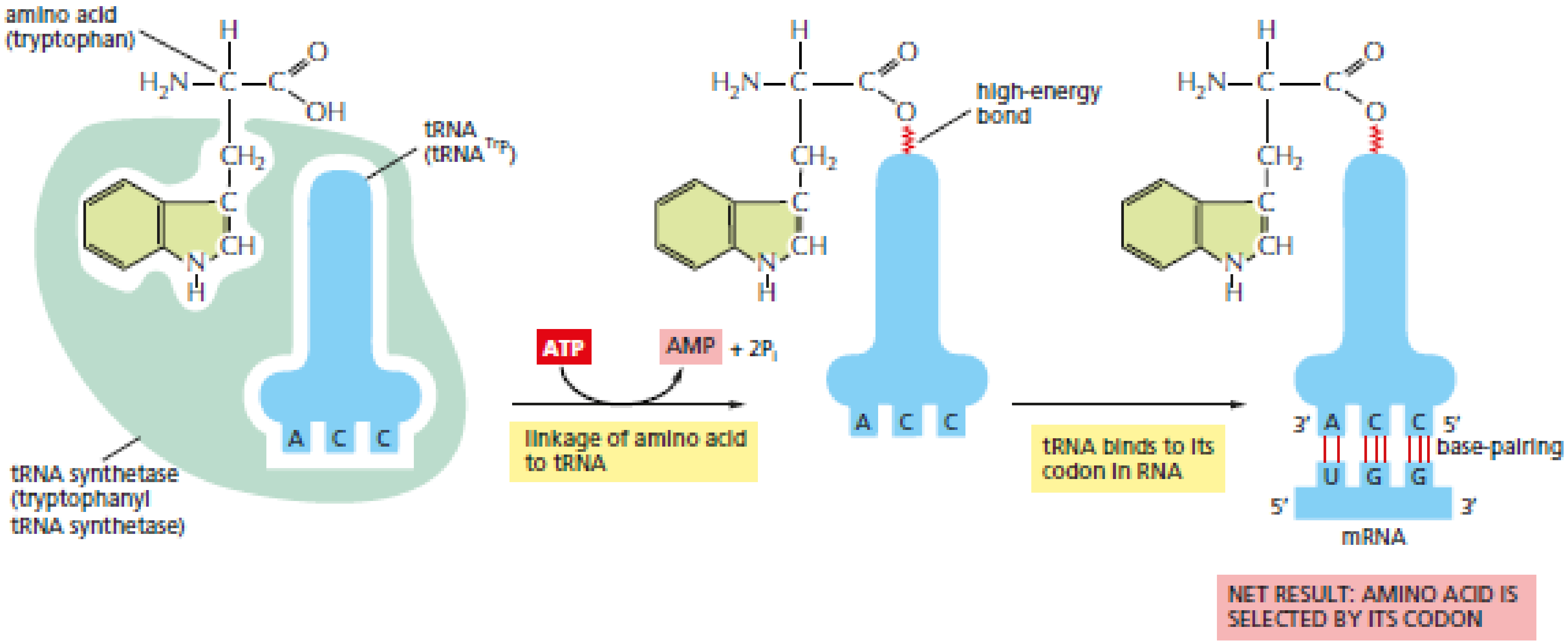
wobble codon base	possible anticodon bases
U	A, G, or I
C	G or I
A	U
G	C

Specific Enzymes Couple tRNAs to the Correct Amino Acid



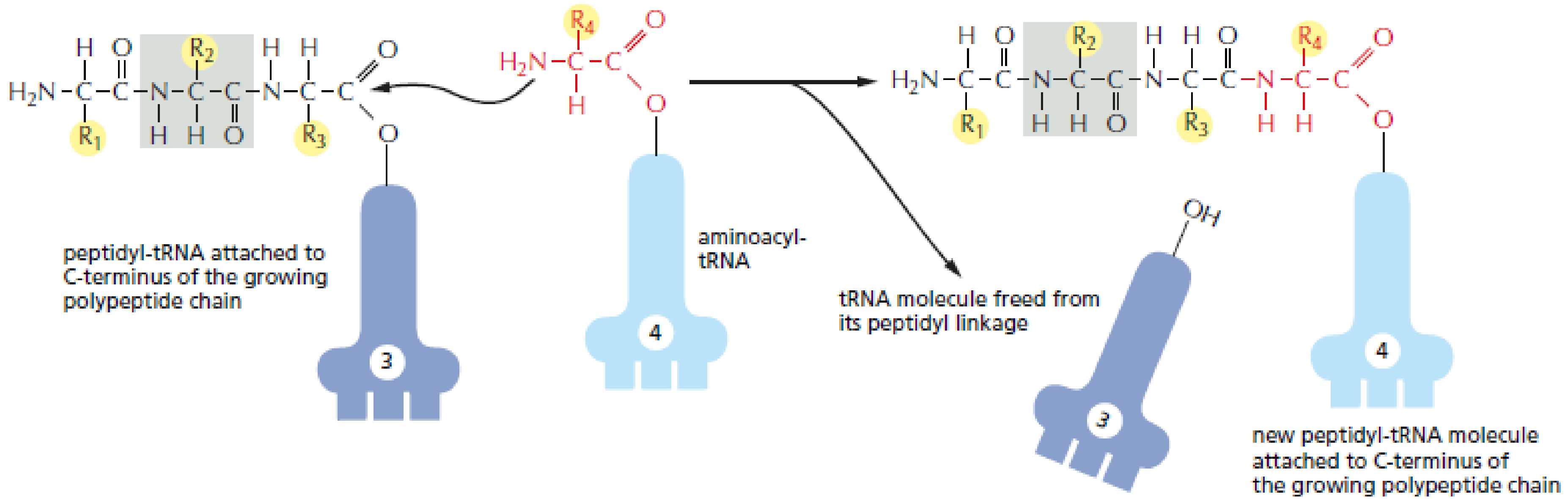
Recognition and attachment of the correct amino acid depends on enzymes called **aminoacyl-tRNA synthetases**, which covalently couple each amino acid to its appropriate set of tRNA molecules

Specific Enzymes Couple tRNAs to the Correct Amino Acid

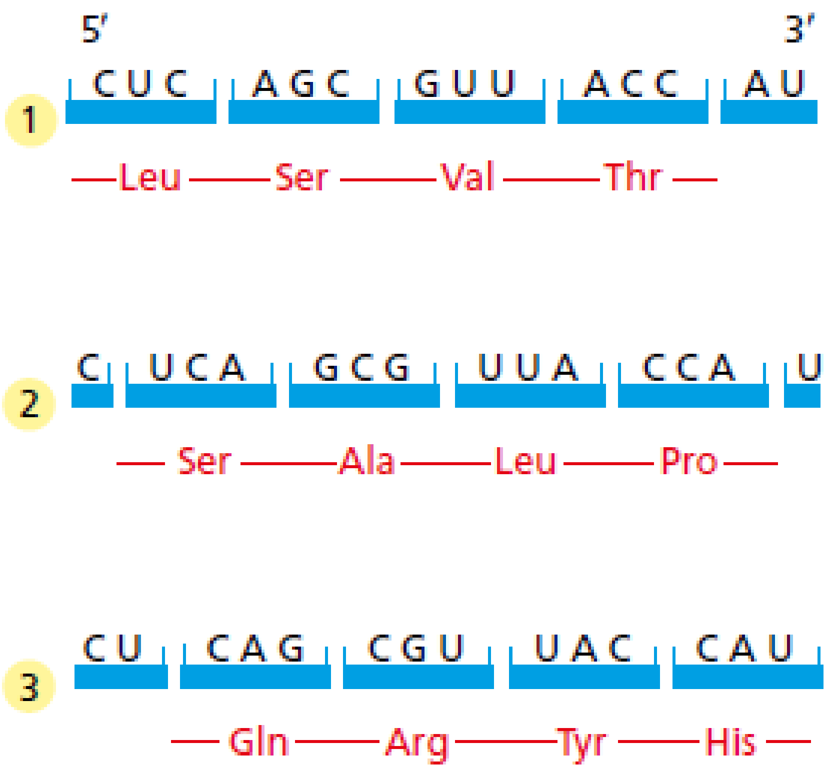


Recognition and attachment of the correct amino acid depends on enzymes called **aminoacyl-tRNA synthetases**, which covalently couple each amino acid to its appropriate set of tRNA molecules

The incorporation of an amino acid into a protein



The RNA Message Is Decoded in Ribosomes

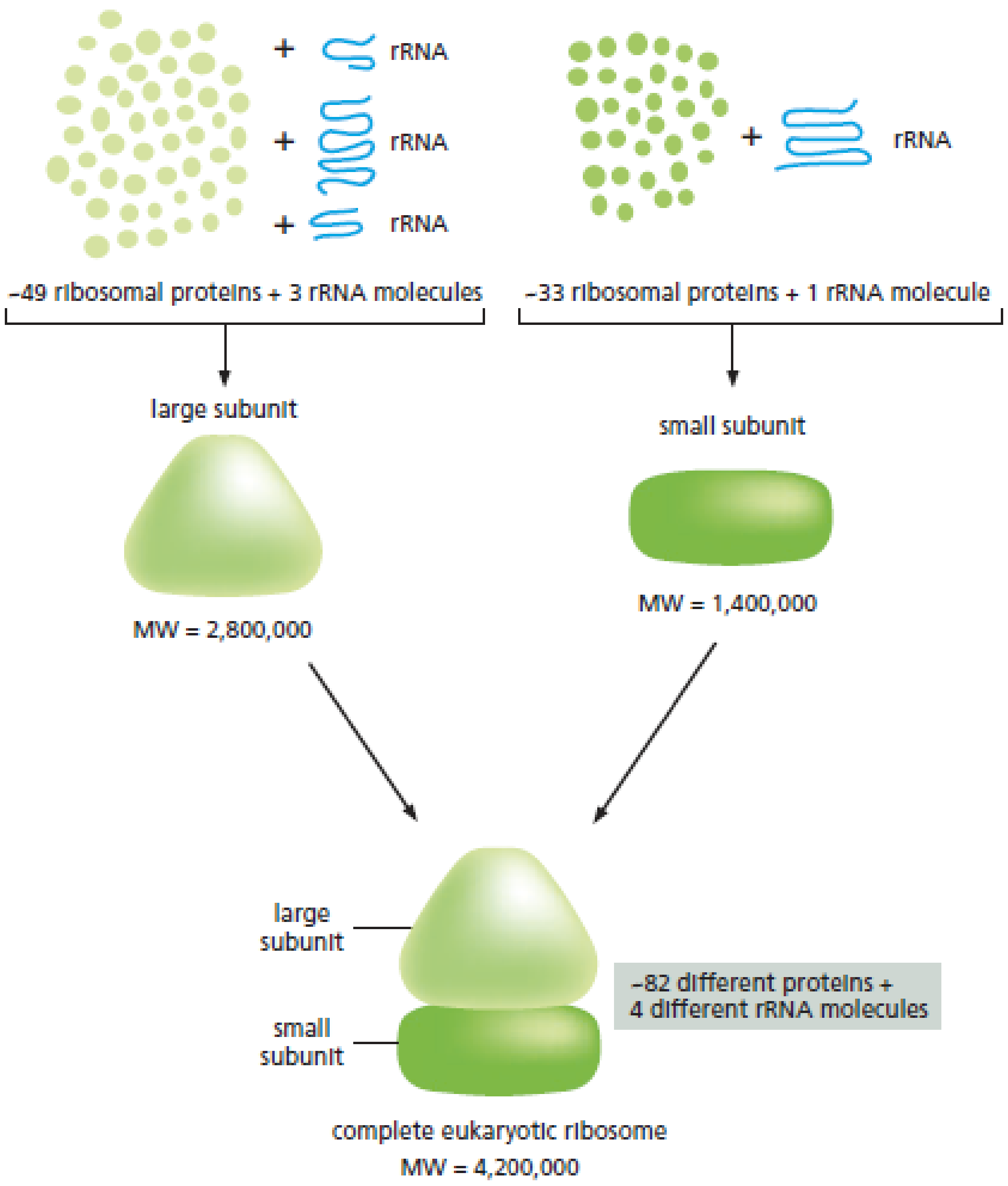


The three possible reading frames in protein synthesis

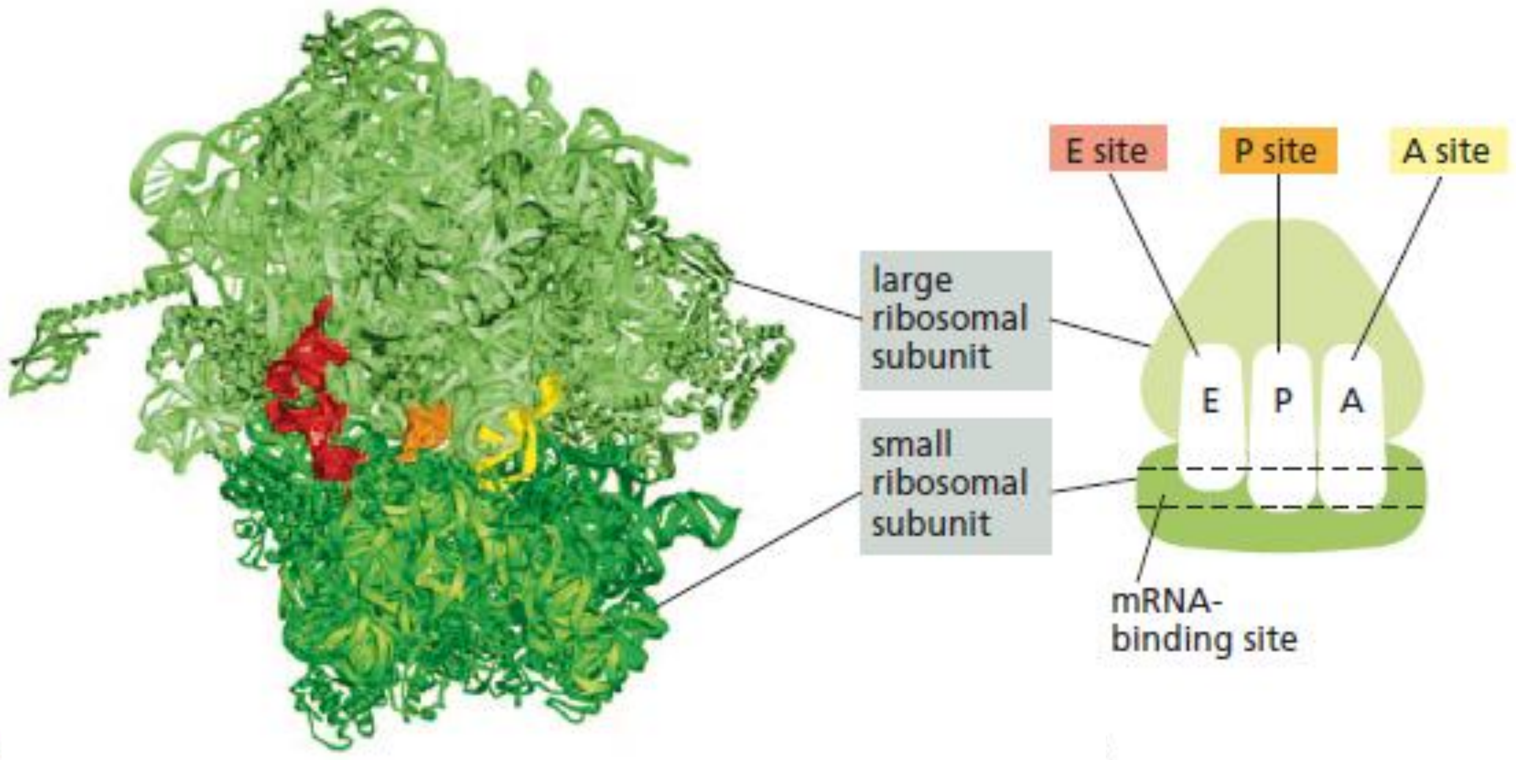
However, only one of the three possible reading frames in an mRNA encodes the required protein.

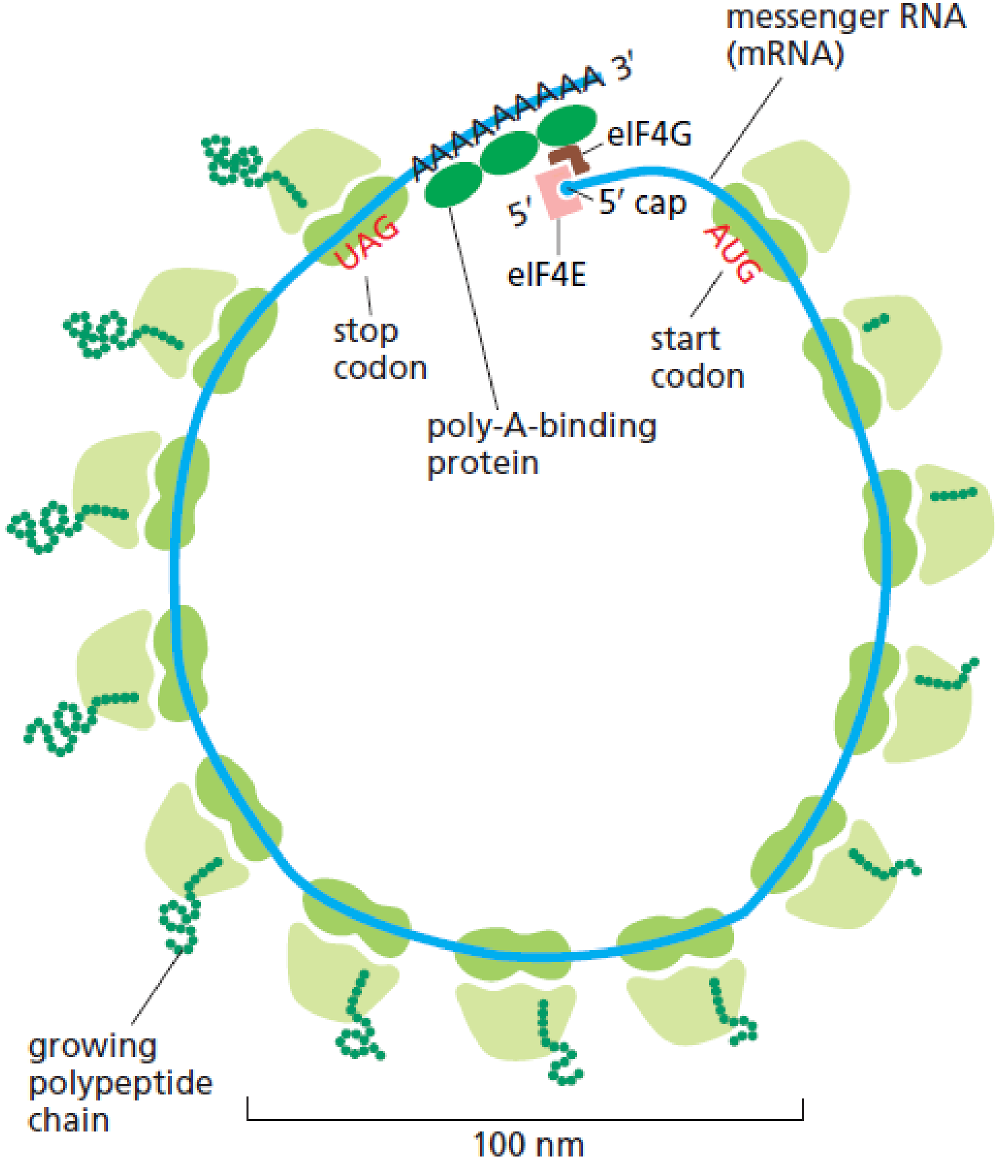
To maintain the correct reading frame and to ensure accuracy (about 1 mistake every 10,000 amino acids), protein synthesis is performed in the ribosome

Eukaryotic ribosomes

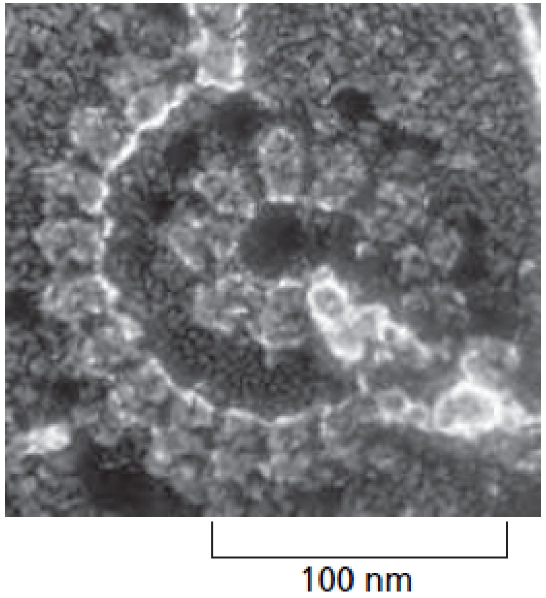


Each ribosome has a binding site for mRNA and three binding sites for tRNA



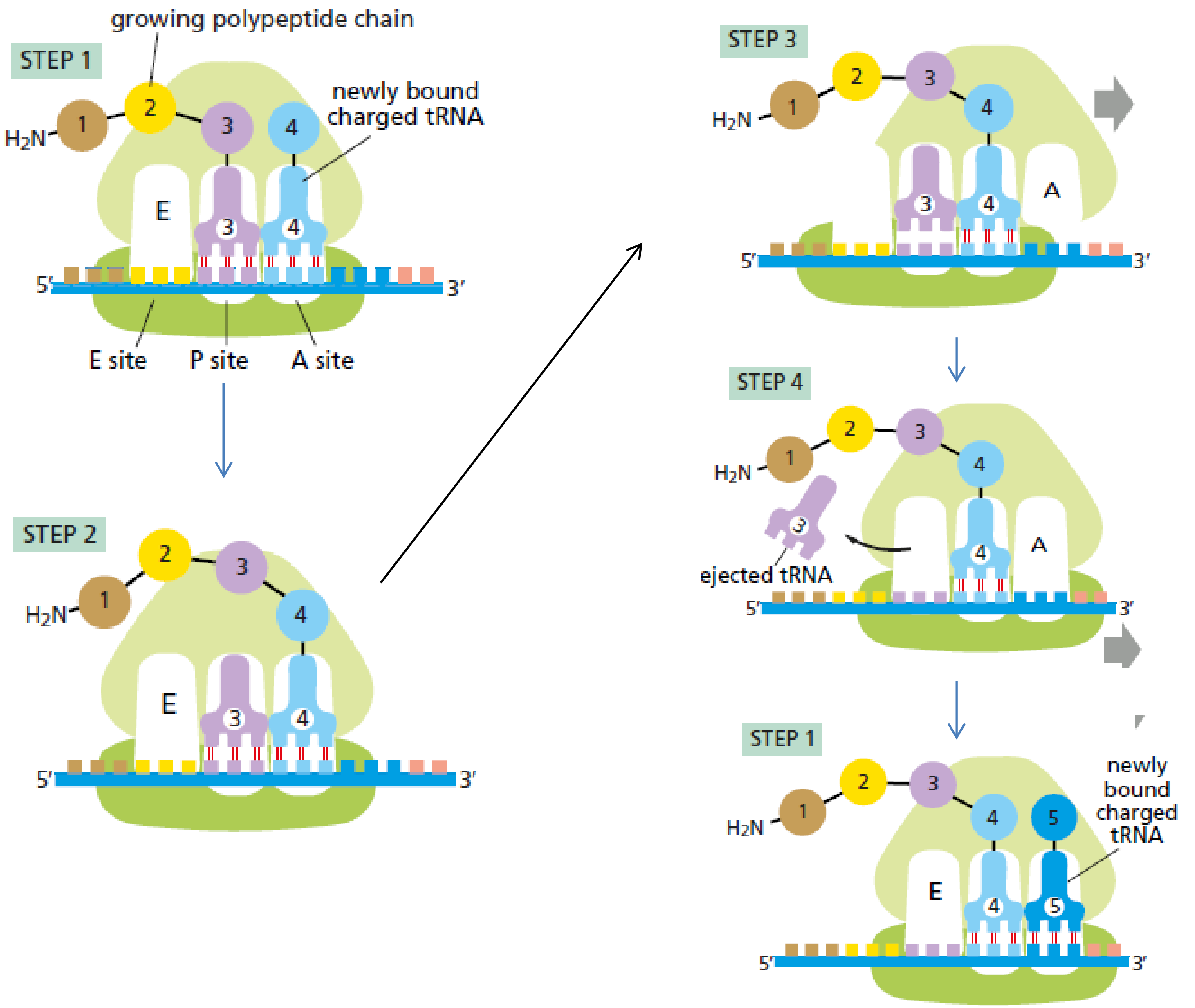

messenger RNA
(mRNA)
AAAAAAAAAA 3′
eIF4G
5′
5′ cap
eIF4E
UAG
AUG
stop
codon
start
codon
poly-A-binding
protein
growing
polypeptide
chain
100 nm

(A)

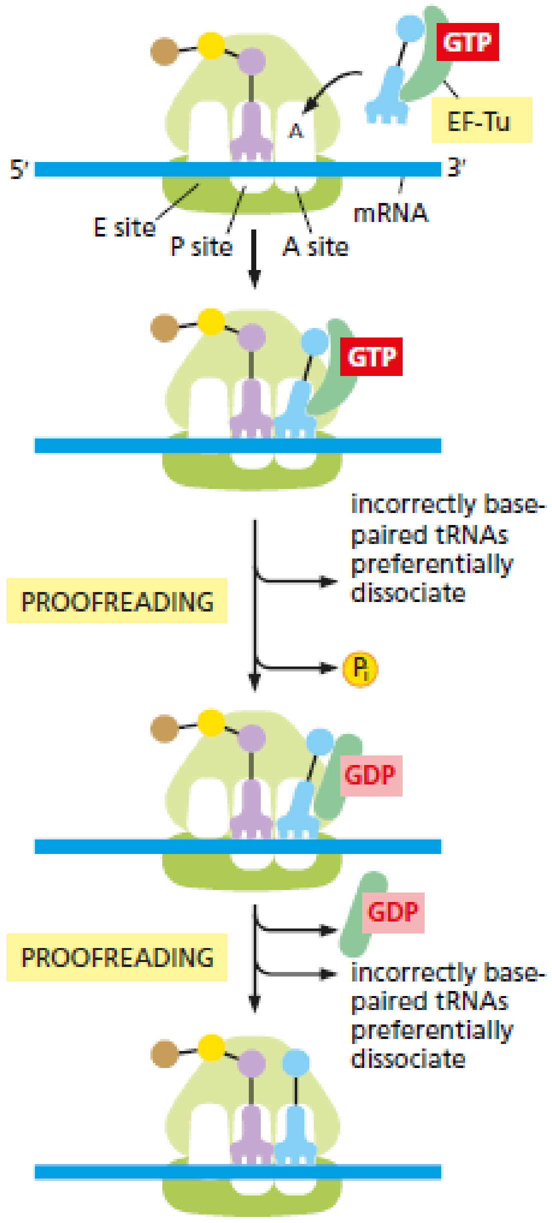

100 nm

(B)

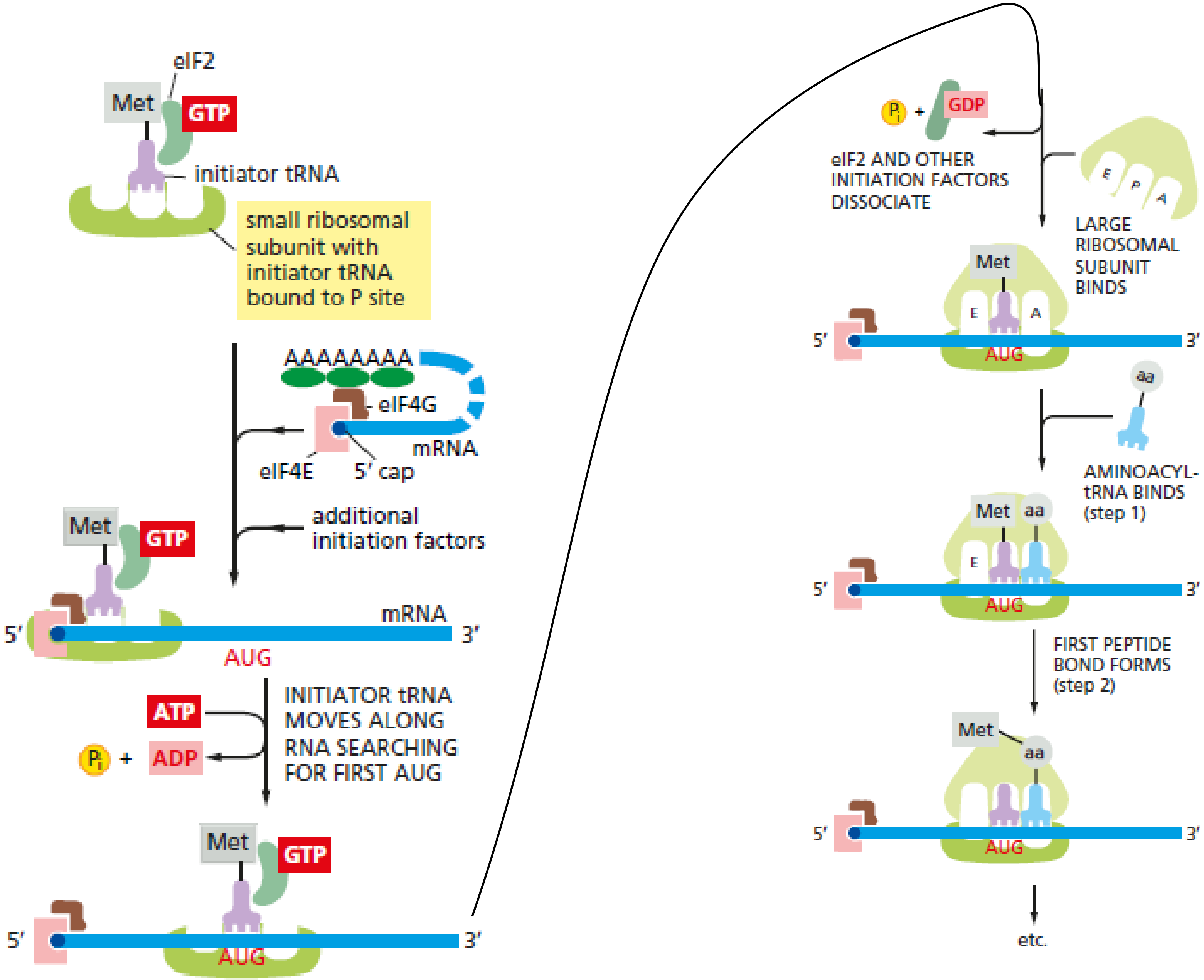
Translating an mRNA molecule



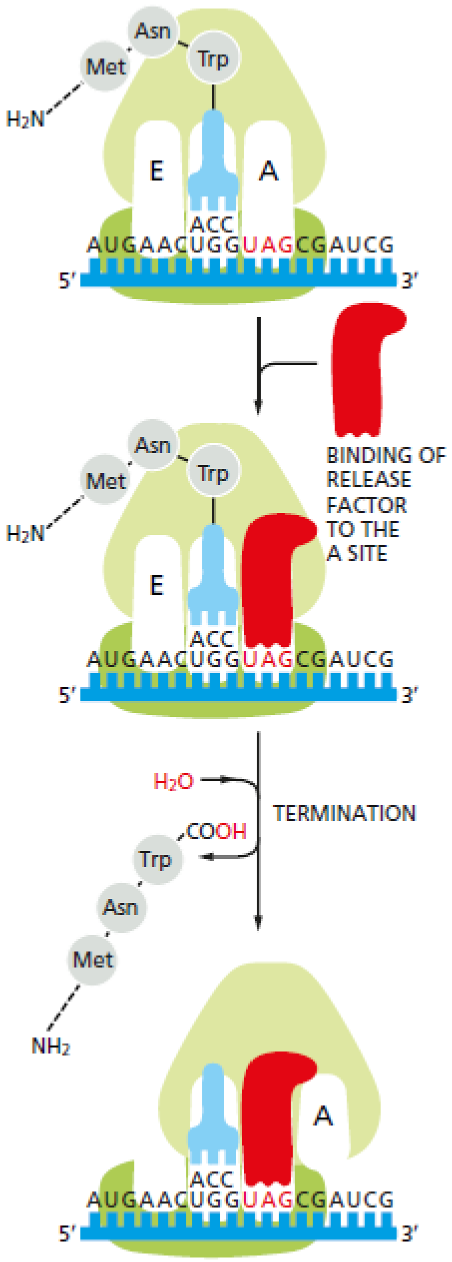
Proofreading



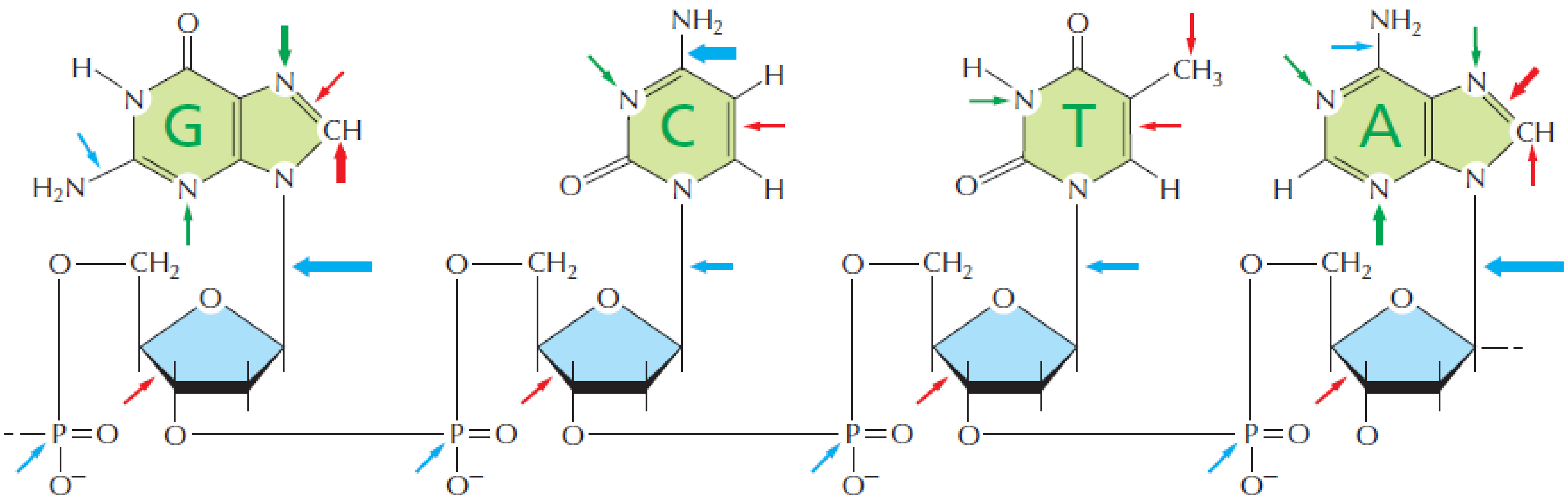
The initiation of protein synthesis



The final phase of protein synthesis

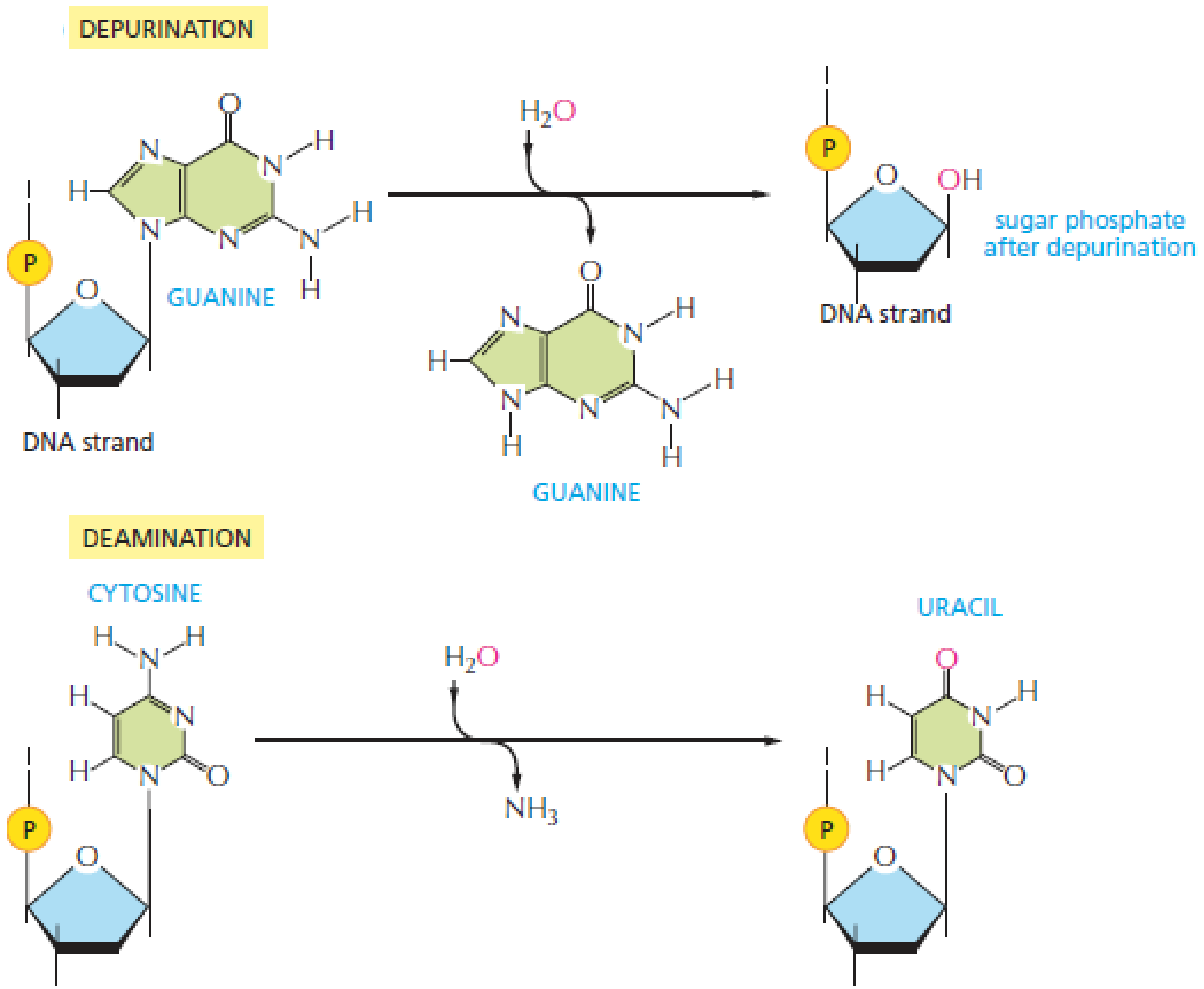


DNA Damage Occurs Continually in Cells



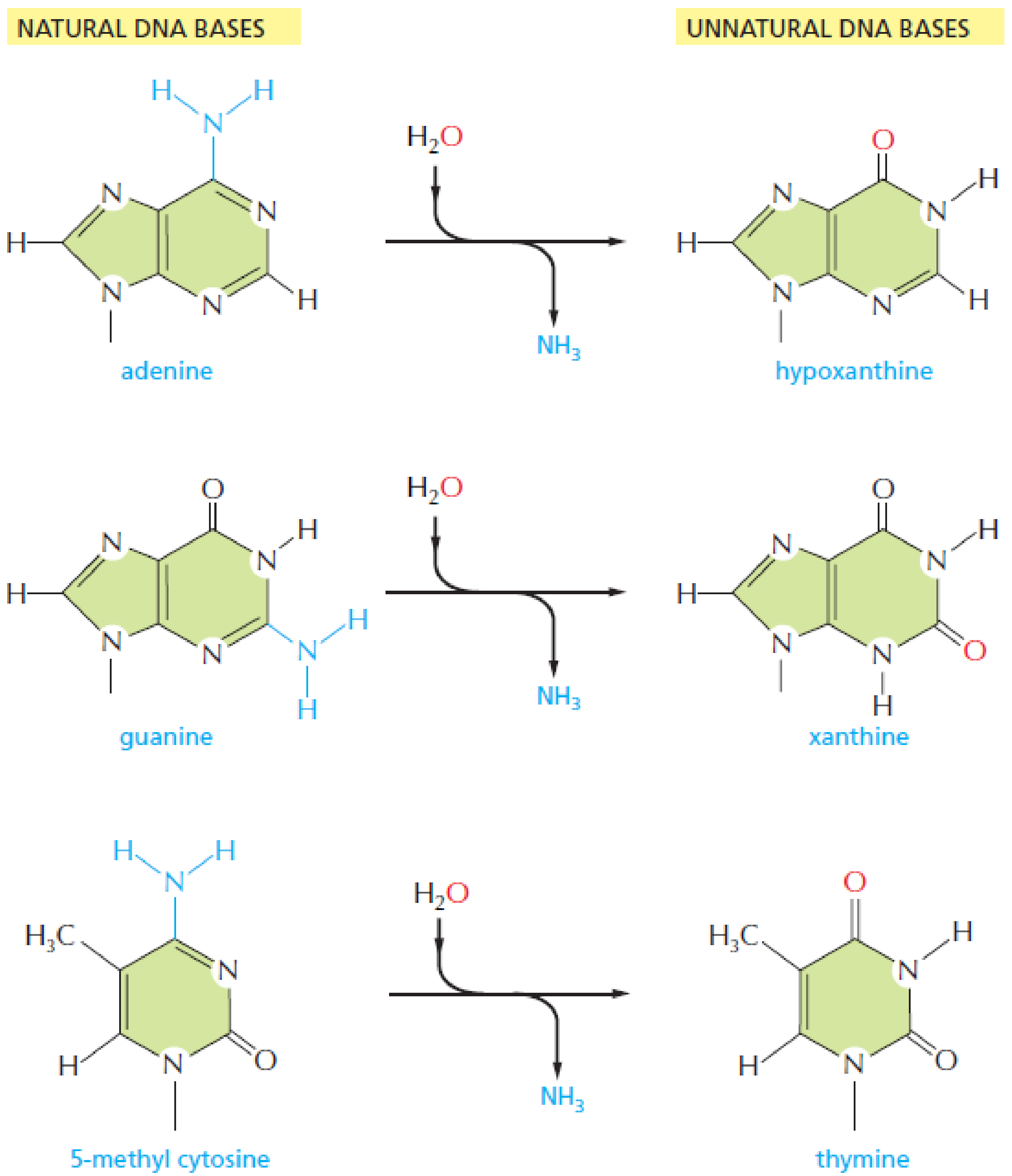
The sites on each nucleotide modified by spontaneous oxidative damage *(red arrows)*, hydrolytic attack *(blue arrows)*, and methylation *(green arrows)* are shown, with the width of each arrow indicating the relative frequency of each event.

DNA Damage Occurs Continually in Cells



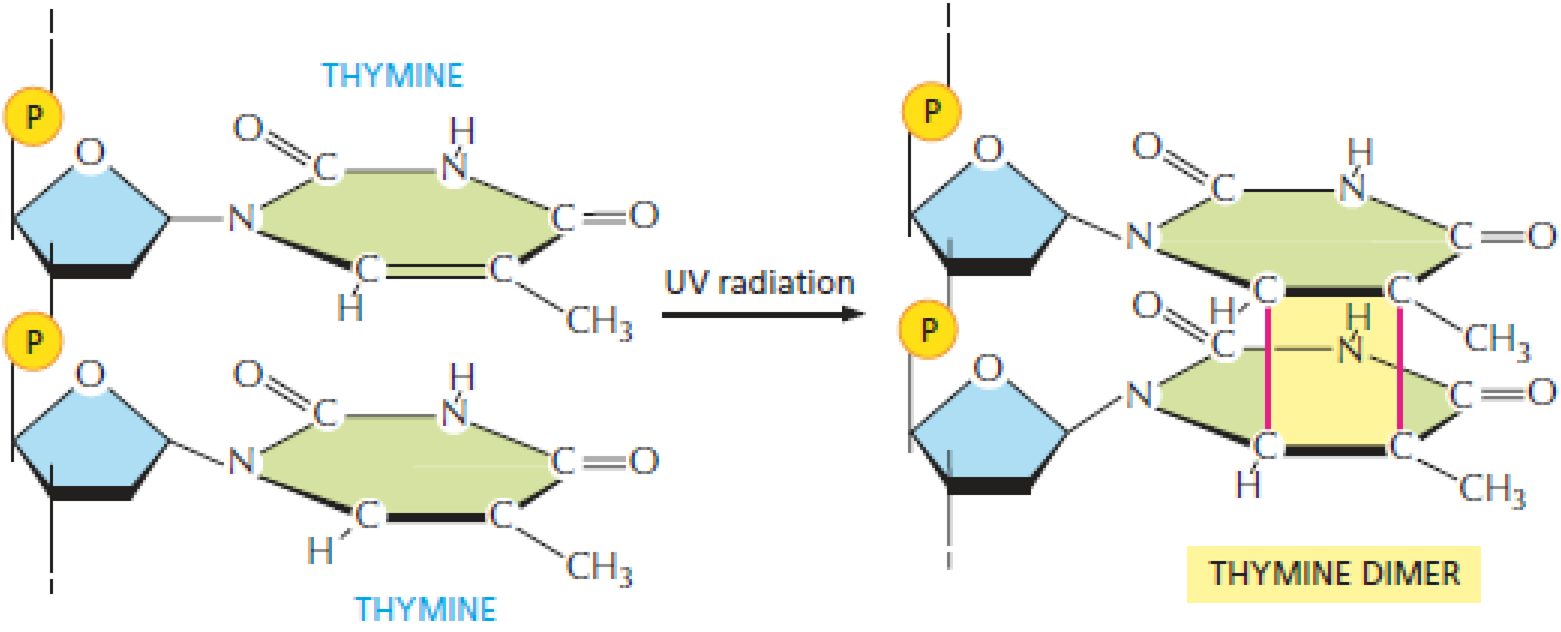
Depurination and deamination are the most frequent chemical reactions known to create serious DNA damage in cells

Thermal collisions or exposure to reactive metabolic by-products, DNA-damaging chemicals, or radiation

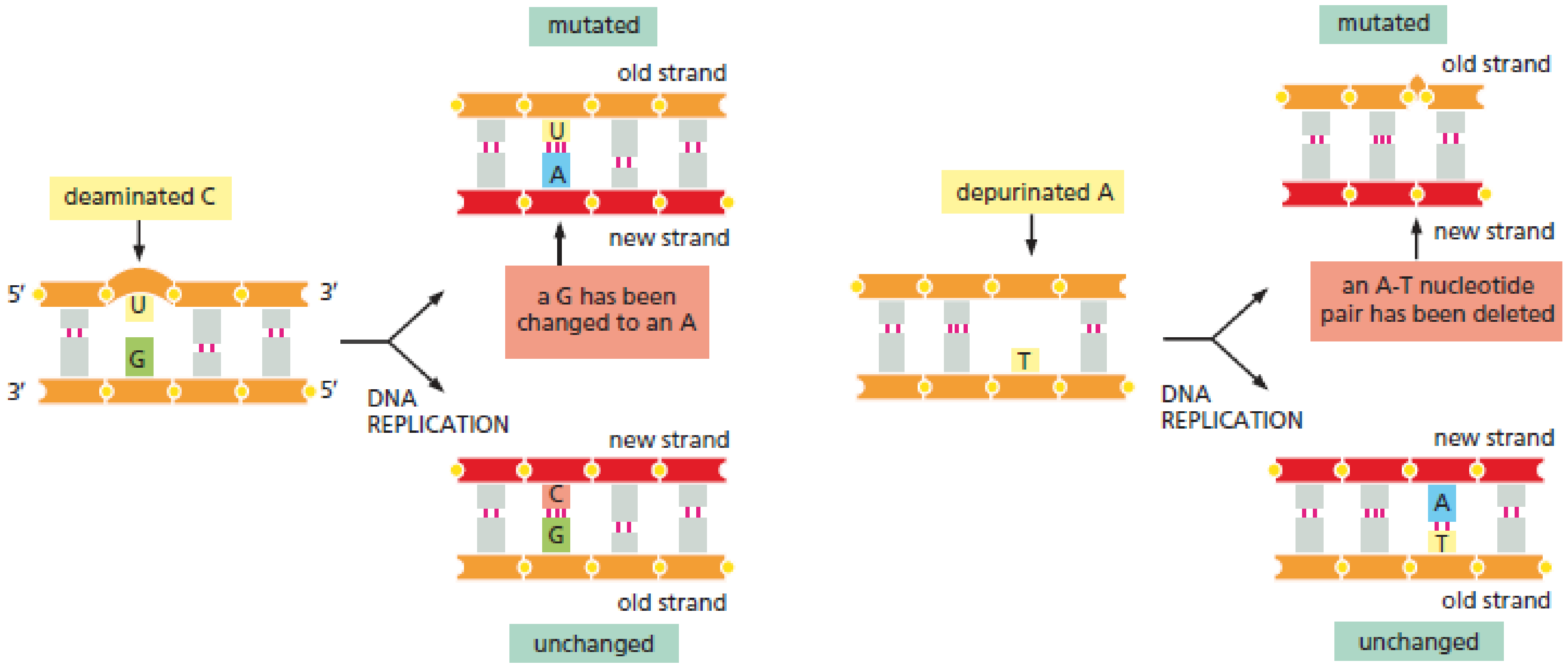


NATURAL DNA BASES
UNNATURAL DNA BASES
H₂O
NH₃
adenine
hypoxanthine
guanine
xanthine
5-methyl cytosine
thymine

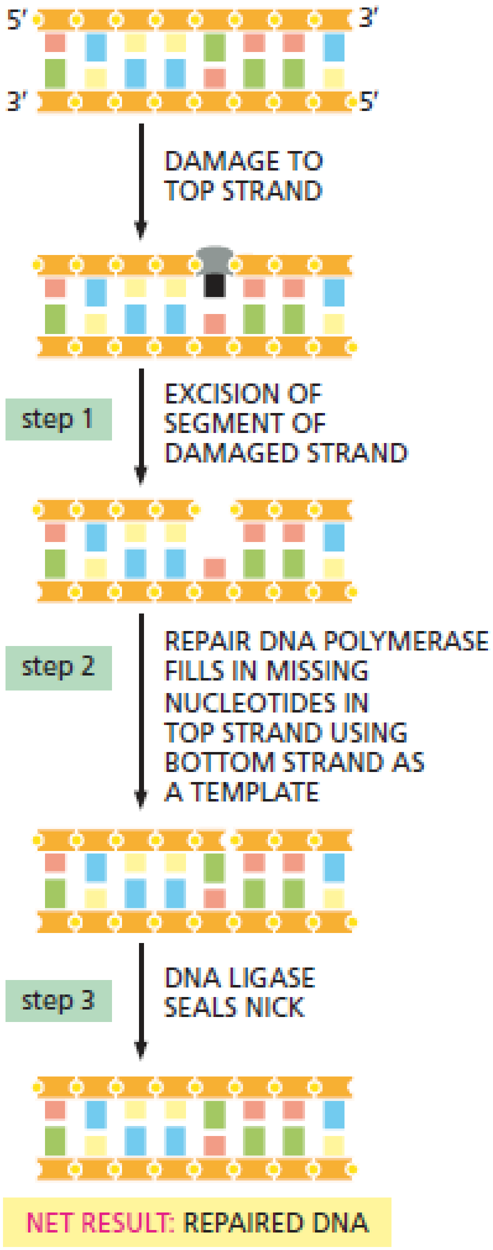
The ultraviolet radiation UV radiation in sunlight can cause the formation of thymine dimers



Chemical modifications of nucleotides, if left unrepaired, produce mutations

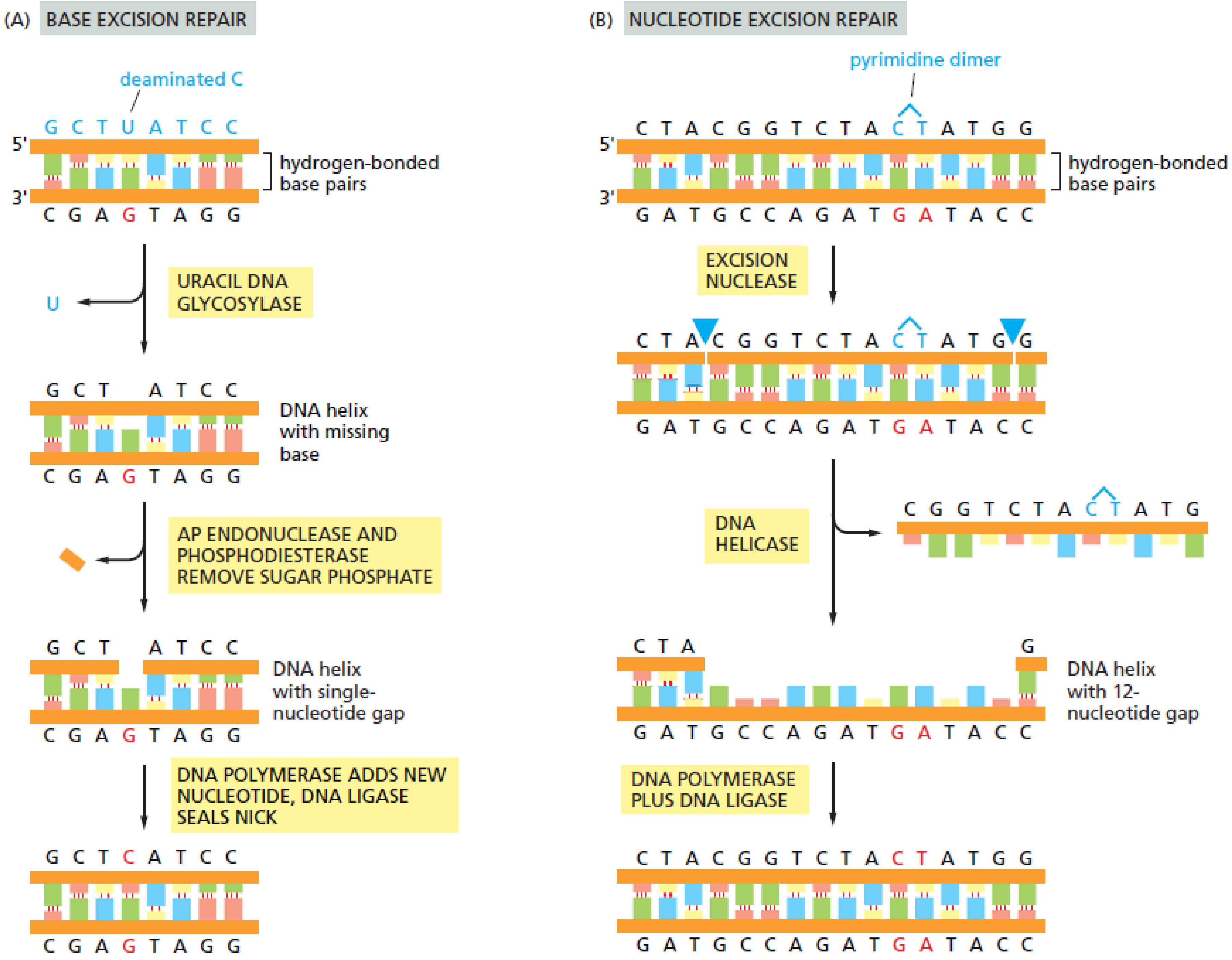


The basic mechanism of DNA repair involves three steps

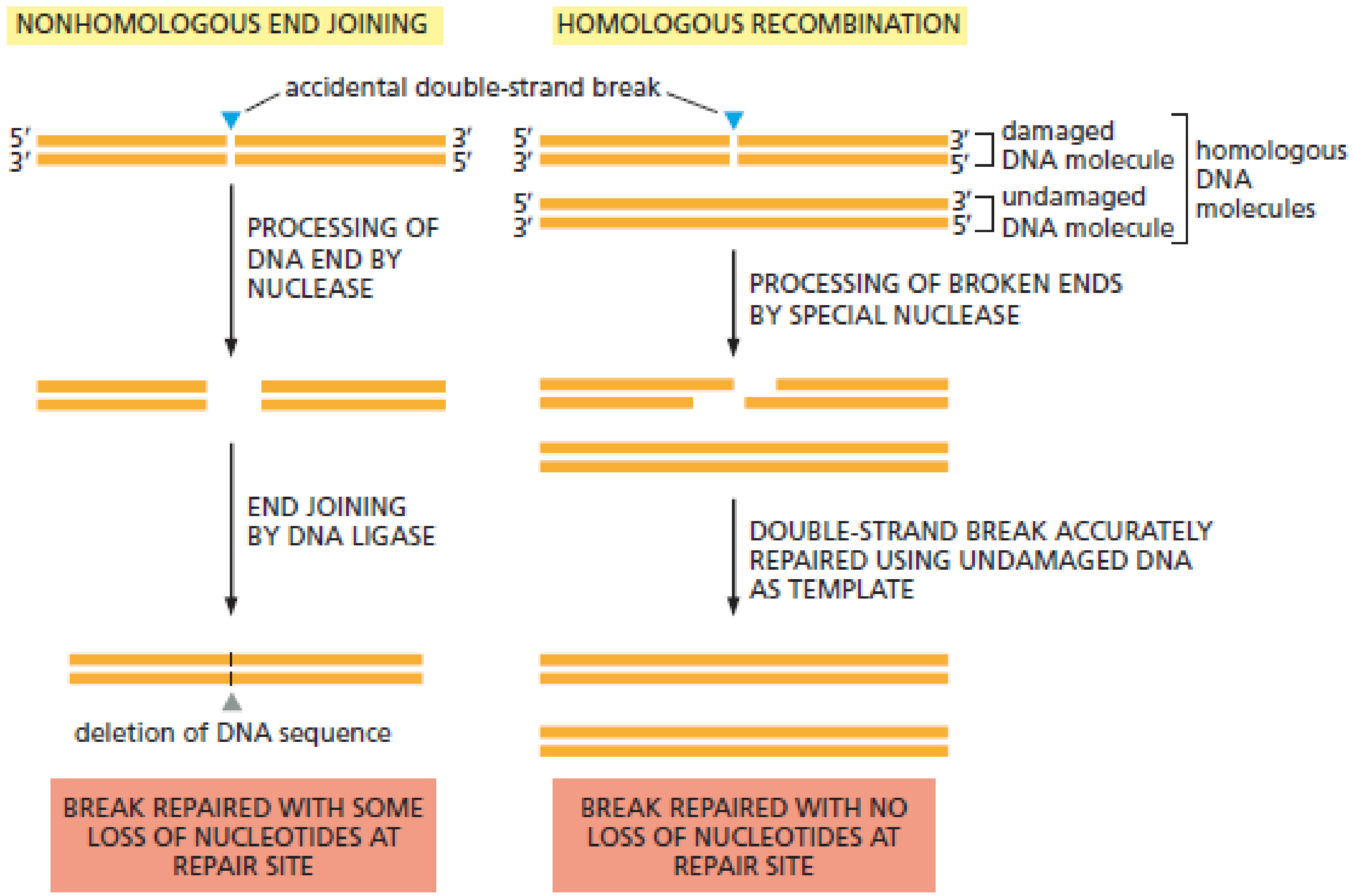


1. Base excision repair

2. Nucleotide excision repair

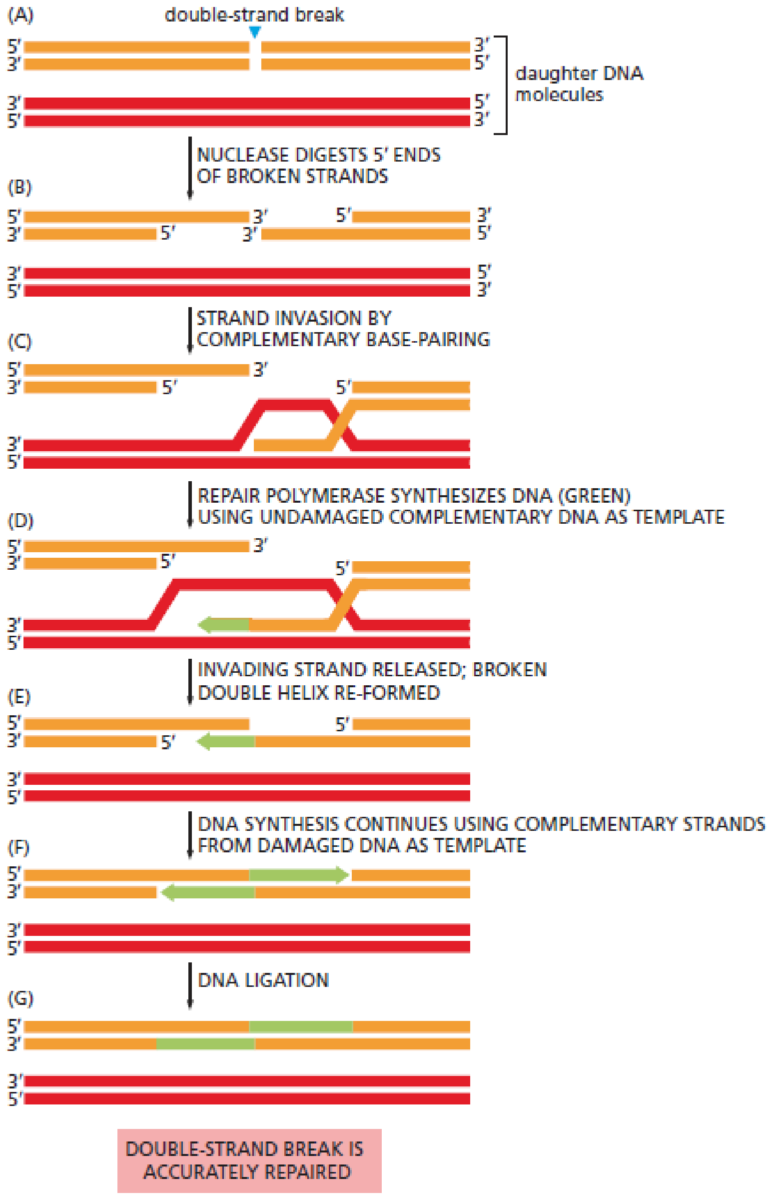

(A) BASE EXCISION REPAIR
deaminated C
G C T U A T C C
5'
3'
C G A G T A G G
hydrogen-bonded base pairs
U
URACIL DNA GLYCOSYLASE
G C T A T C C
C G A G T A G G
DNA helix with missing base
AP ENDONUCLEASE AND PHOSPHODIESTERASE REMOVE SUGAR PHOSPHATE
G C T A T C C
C G A G T A G G
DNA helix with single-nucleotide gap
DNA POLYMERASE ADDS NEW NUCLEOTIDE, DNA LIGASE SEALS NICK
G C T C A T C C
C G A G T A G G
(B) NUCLEOTIDE EXCISION REPAIR
pyrimidine dimer
C T A C G G T C T A C T A T G G
5'
3'
G A T G C C A G A T G A T A C C
hydrogen-bonded base pairs
EXCISION NUCLEASE
C T A C G G T C T A C T A T G G
G A T G C C A G A T G A T A C C
DNA HELICASE
C G G T C T A C T A T G
C T A
G
G A T G C C A G A T G A T A C C
DNA helix with 12-nucleotide gap
DNA POLYMERASE PLUS DNA LIGASE
C T A C G G T C T A C T A T G G
G A T G C C A G A T G A T A C C

What happens when both strands of the double helix are damaged at the same time?



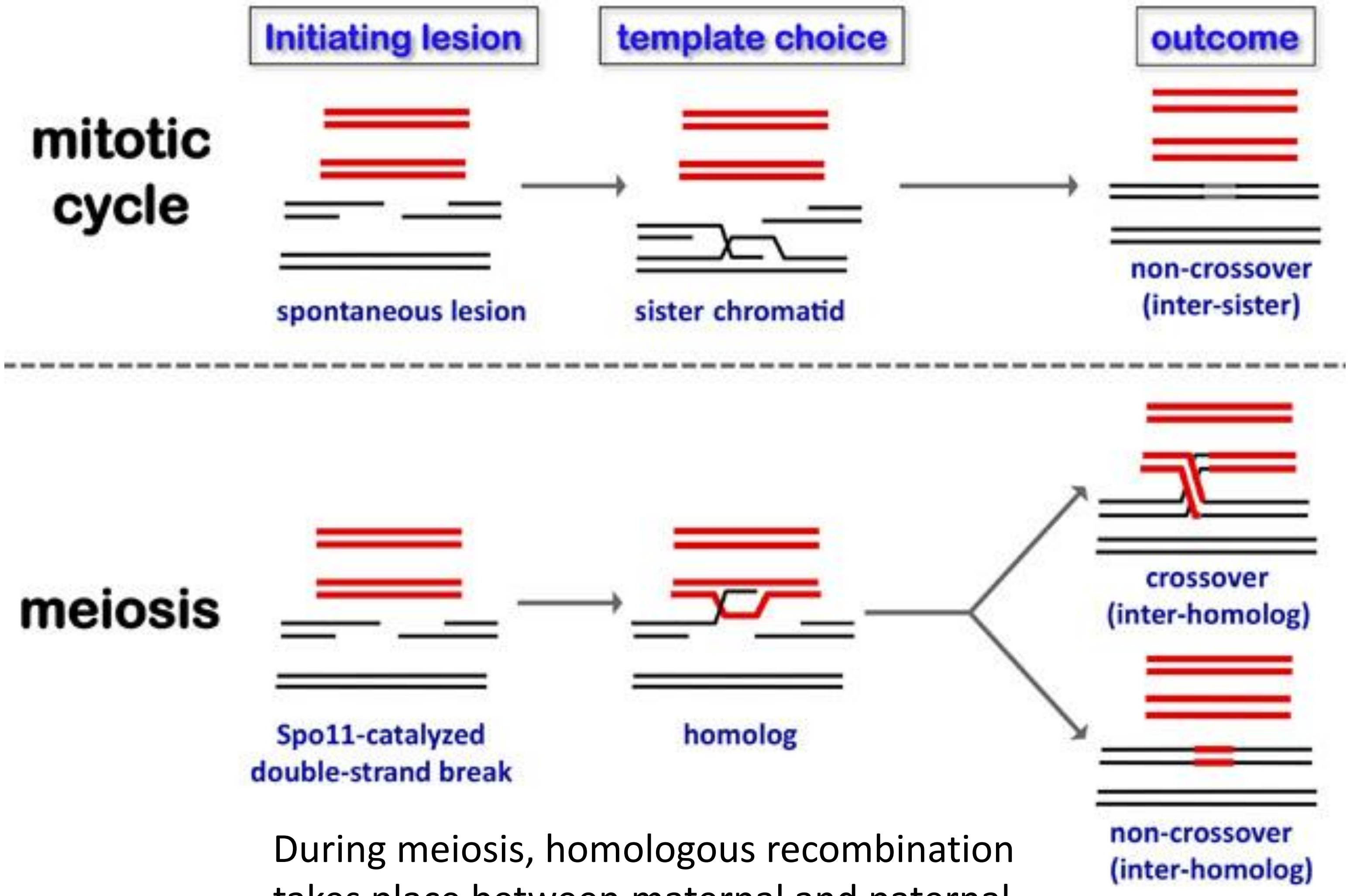
Homologous recombination most often occurs shortly after a cell's DNA has been replicated before cell division

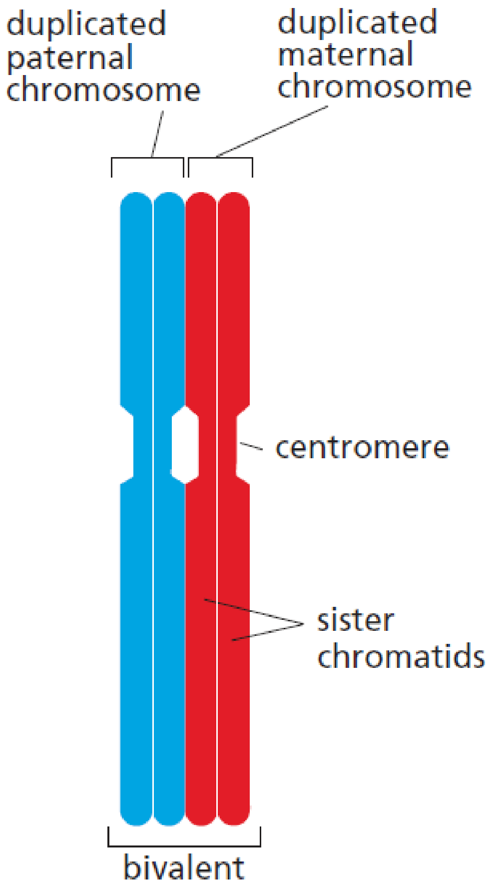
Homologous Recombination Can Flawlessly Repair DNA Double-Strand Breaks



Also called
Inter-sister repair or
Non-crossover

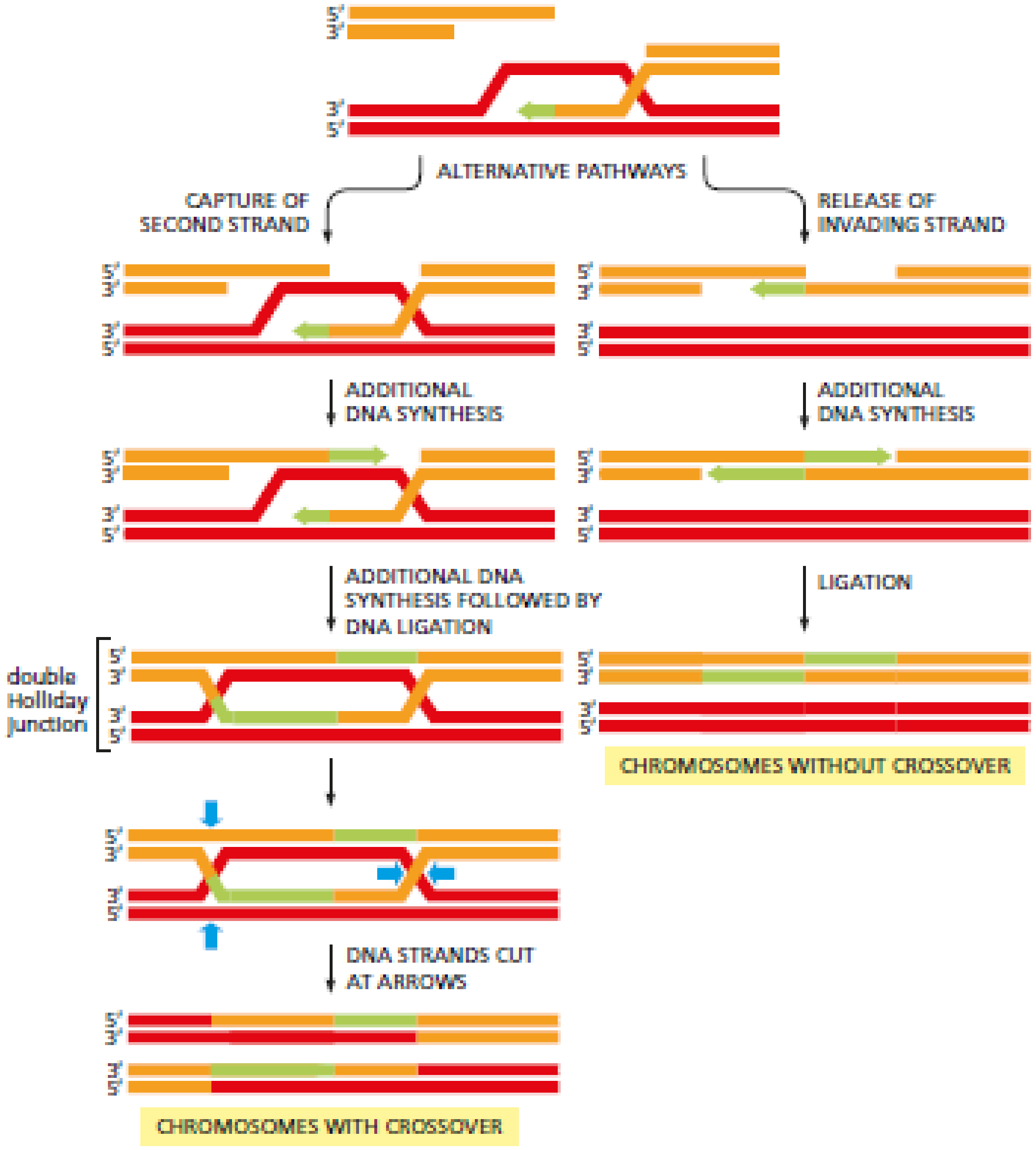
Contrasting regulation of recombination during the mitotic and meiotic programs





During meiosis, homologous recombination takes place between maternal and paternal chromosome homologs when they are held tightly together

Homologous recombination during meiosis



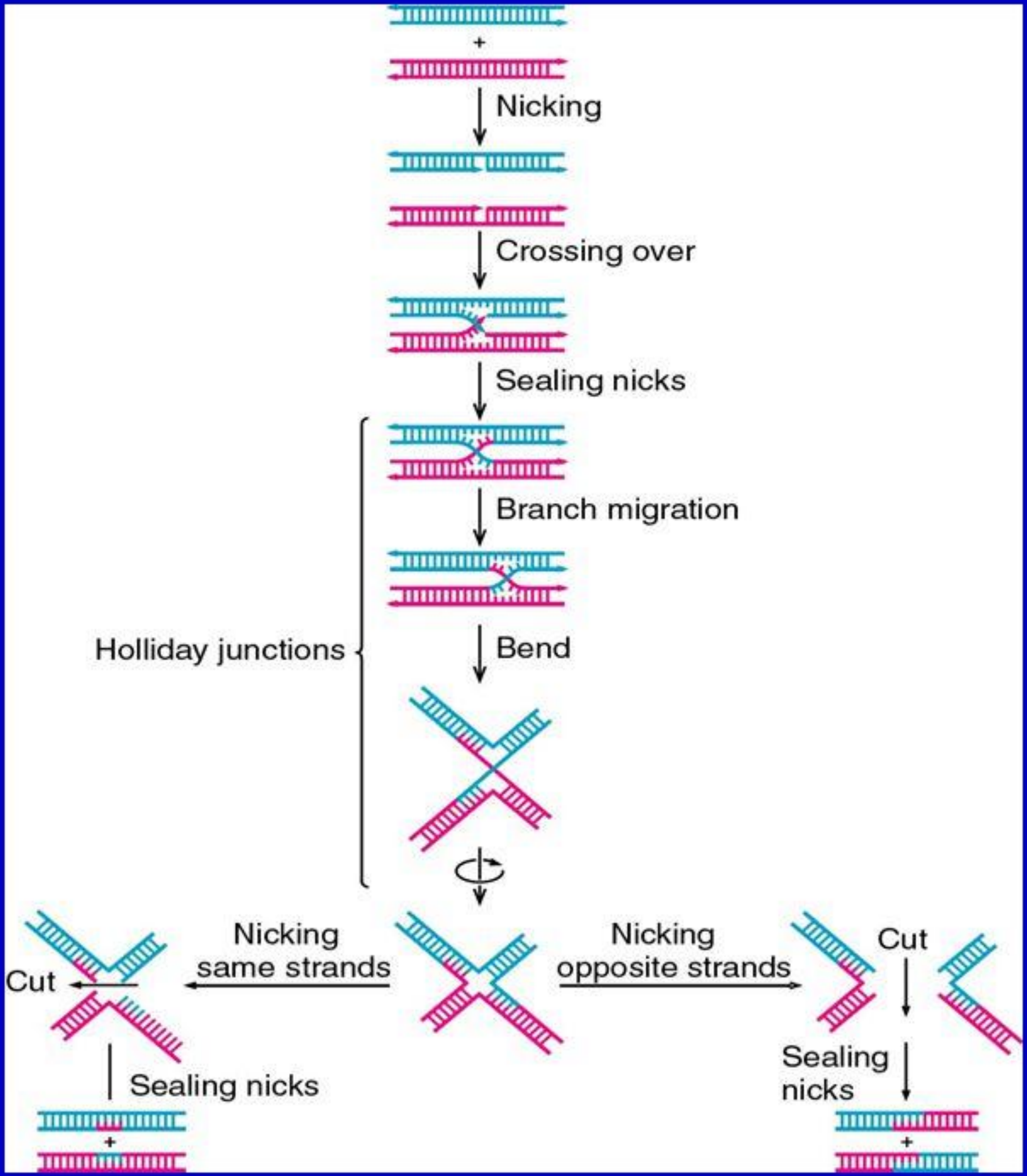
Holliday Model

R. Holliday (1964)

- Holliday Junctions form during recombination

- HJs can be resolved 2 ways, only one produces true recombinant molecules

patch



Manipulating and analysing DNA-
Recombinant DNA technology

1. **Cleavage of DNA** at specific sites by restriction nucleases, which greatly facilitates the isolation and manipulation of individual pieces of a genome.

2. **DNA ligation**, which makes it possible to seamlessly join together DNA molecules from widely different sources.

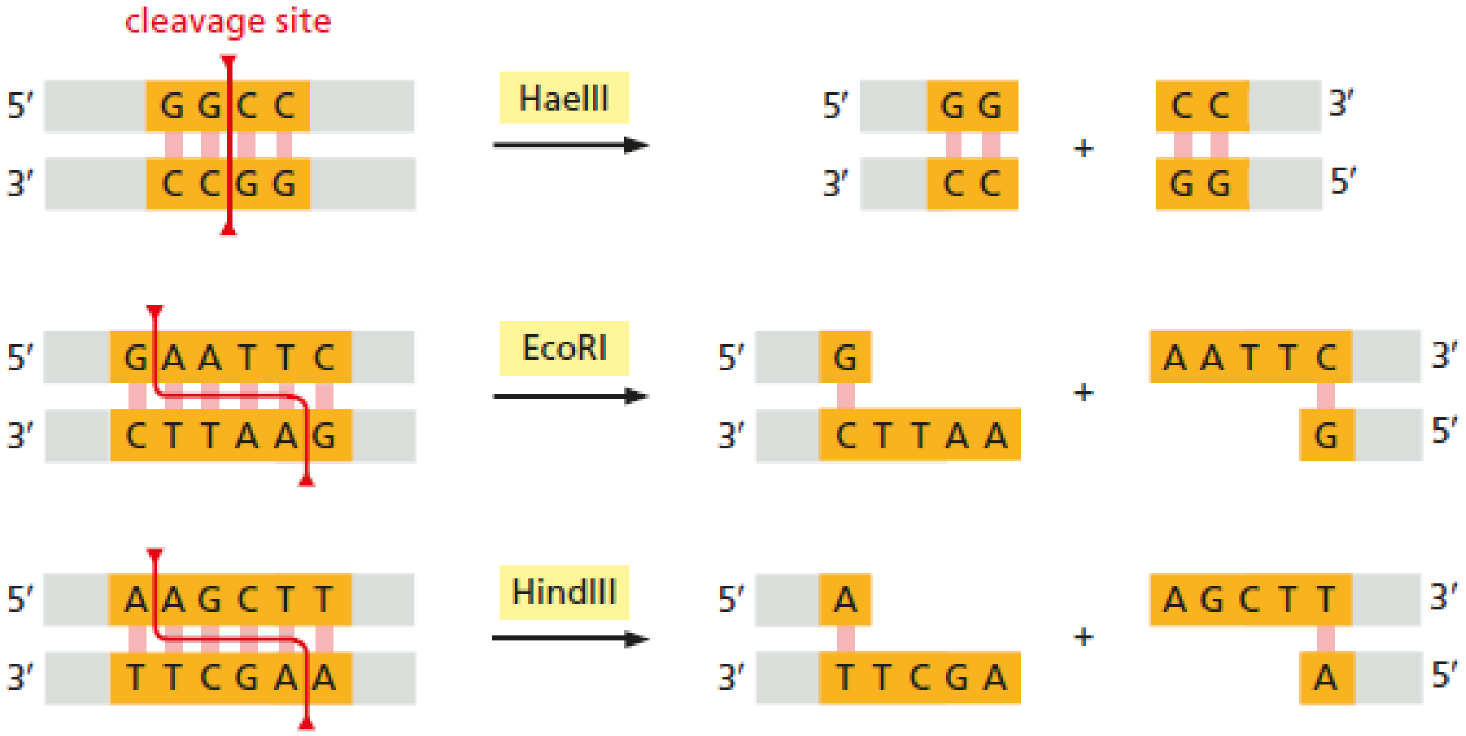
3. **DNA cloning** (through the use of either cloning vectors or the polymerase chain reaction) in which a portion of a genome (often an individual gene) is “purified” away from the remainder of the genome by repeatedly copying it to generate many billions of identical molecules.

4. Nucleic acid **hybridization**, which makes it possible to identify any specific sequence of DNA or RNA with great accuracy and sensitivity based on its ability to selectively bind a complementary nucleic acid sequence.

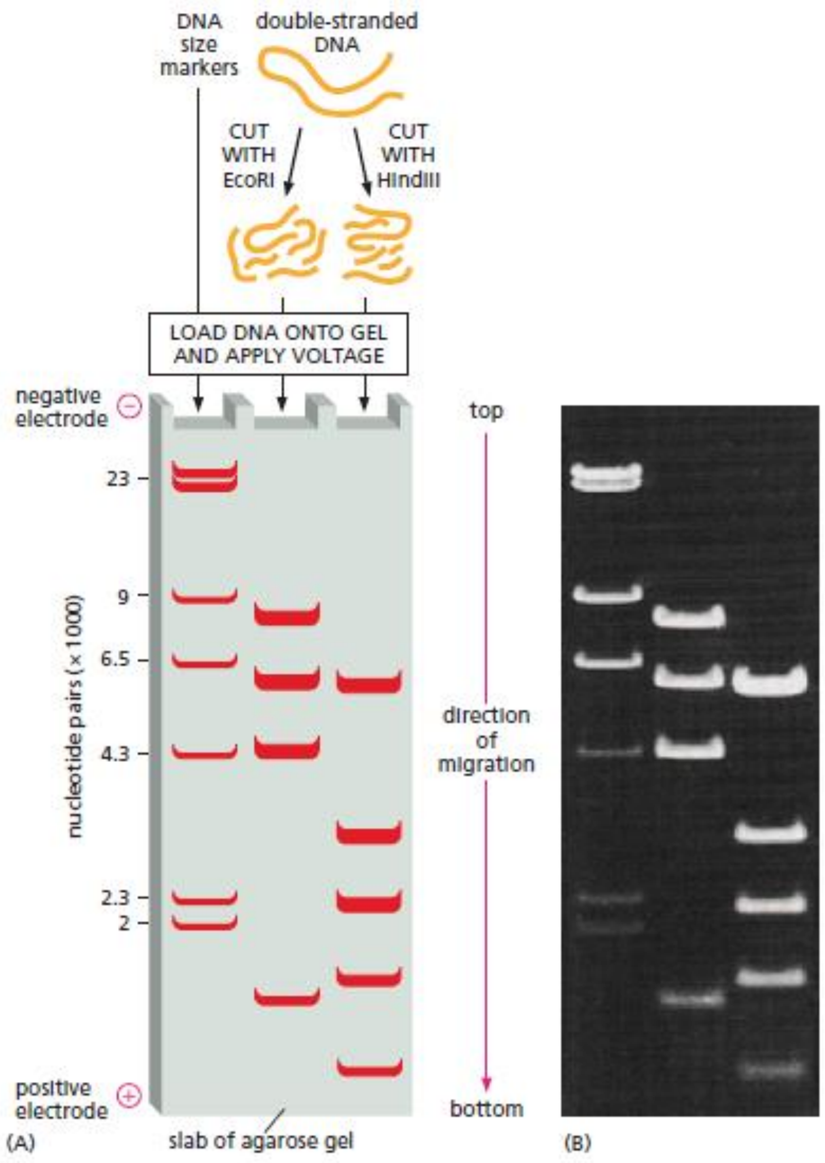
5. **DNA synthesis**, which makes it possible to chemically synthesize DNA molecules with any sequence of nucleotides, whether or not the sequence occurs in nature.

6. Rapid **determination of the sequence** of nucleotides of any DNA or RNA molecule.

Restriction nucleases cleave DNA at specific nucleotide sequences



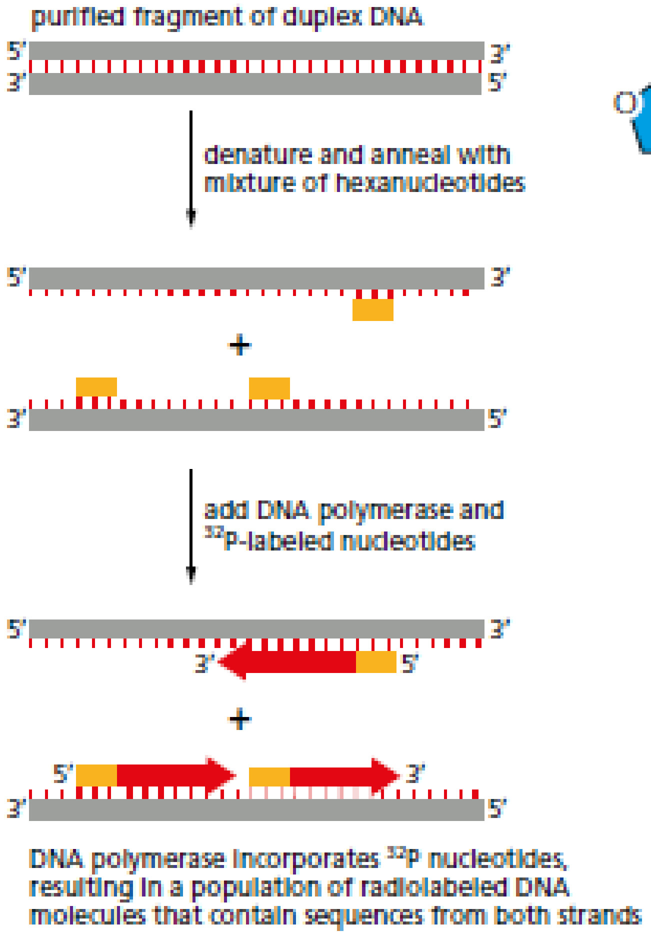
DNA molecules can be separated by size using gel electrophoresis

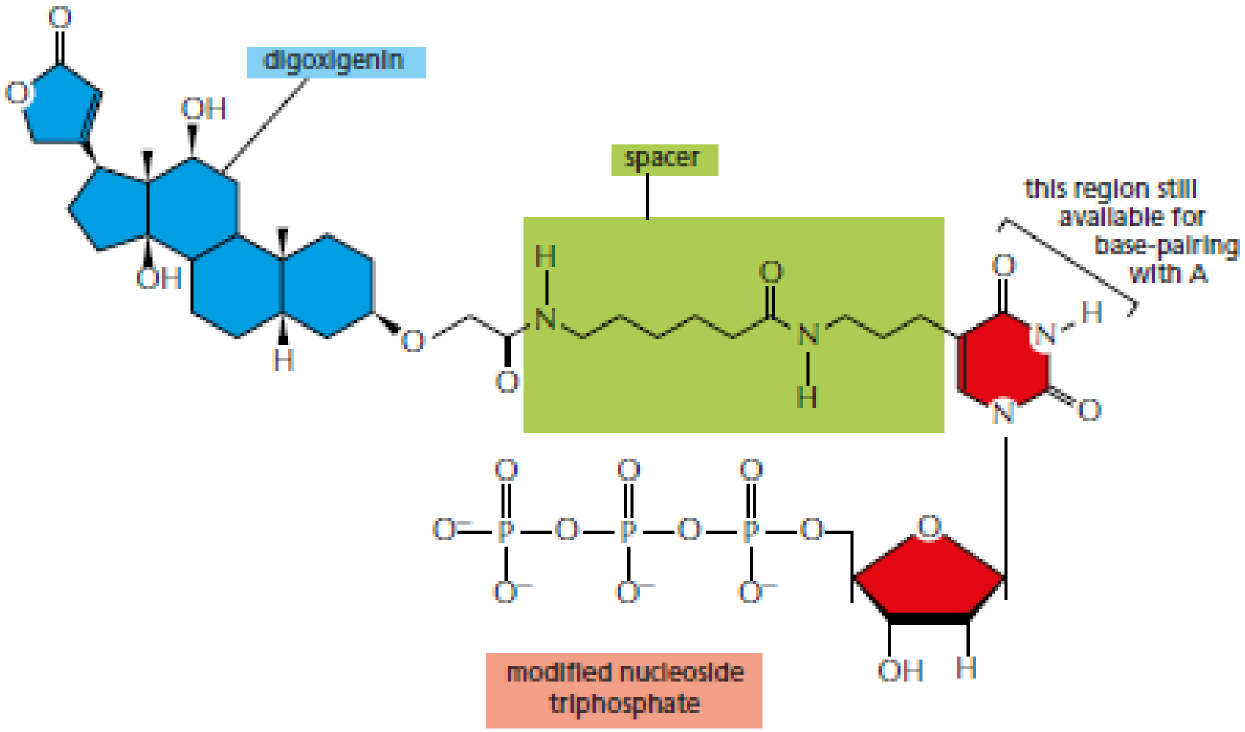


Agarose or polyacrylamide gels

Dye *ethidium bromide*, which fluoresces under ultraviolet light when it is bound to DNA is used to label DNA

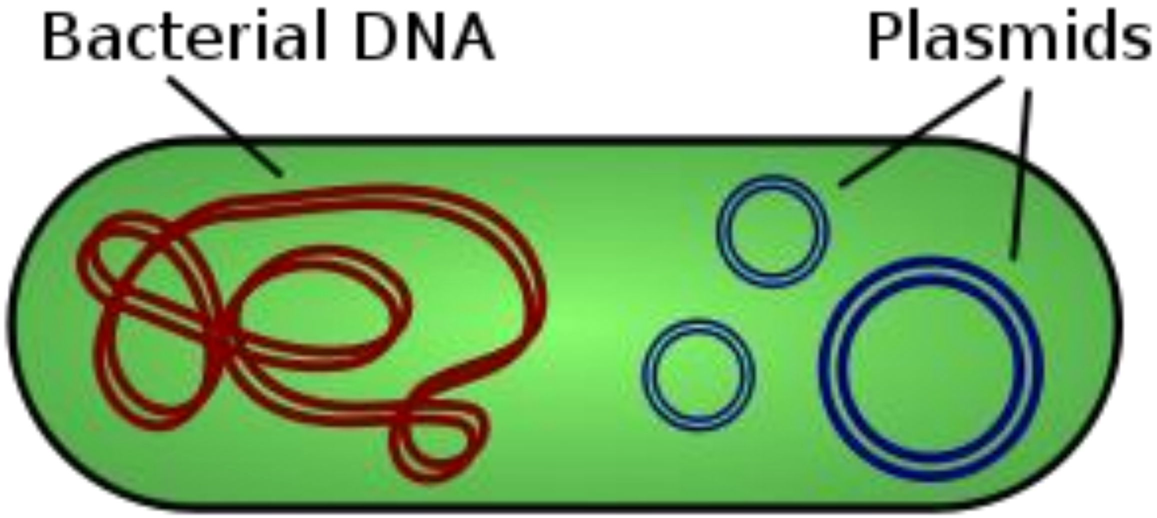
Methods for labeling DNA molecules *in vitro*

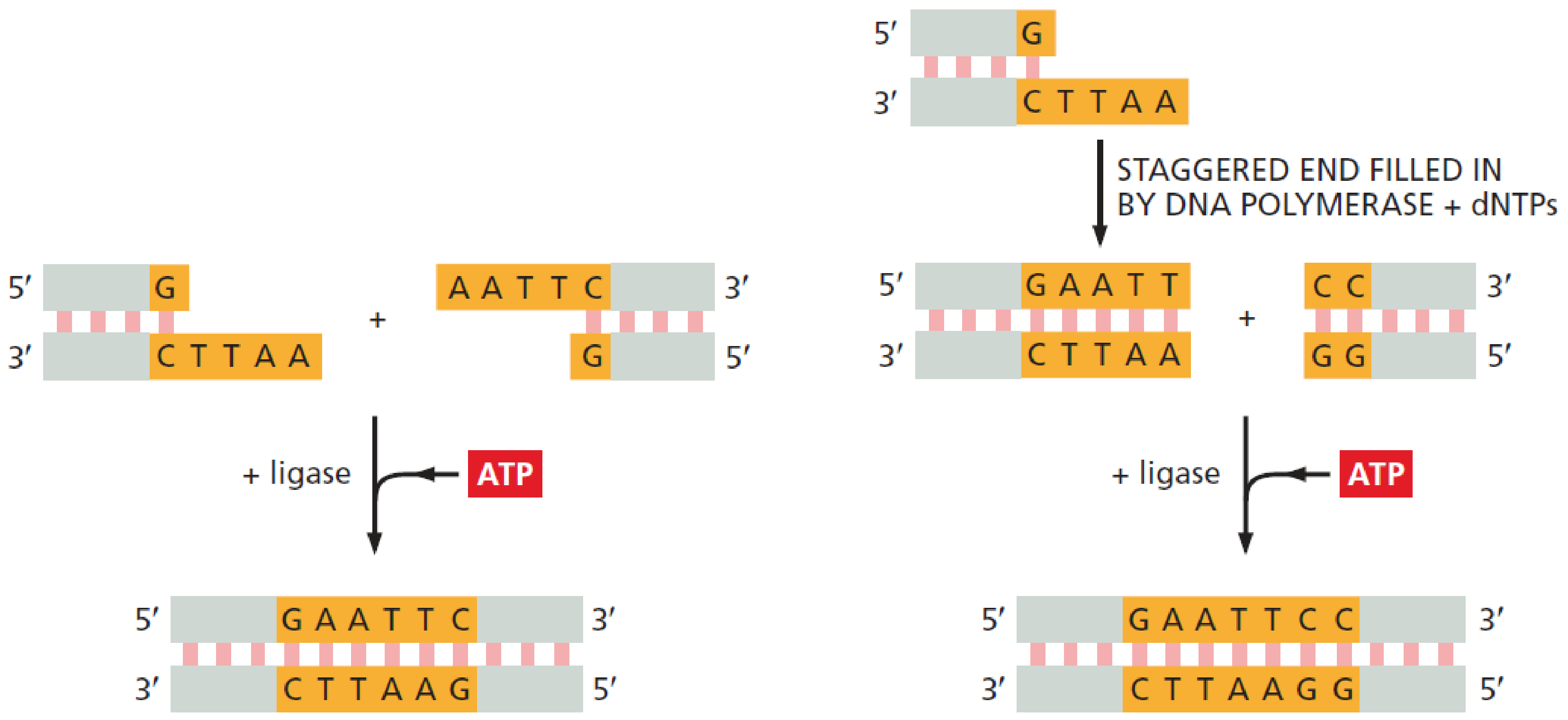




DNA labeled with the radioisotope ^{32}P can be detected following gel electrophoresis by placing the gel next to a piece of photographic film.

Illustration of a bacterium showing chromosomal DNA and plasmids

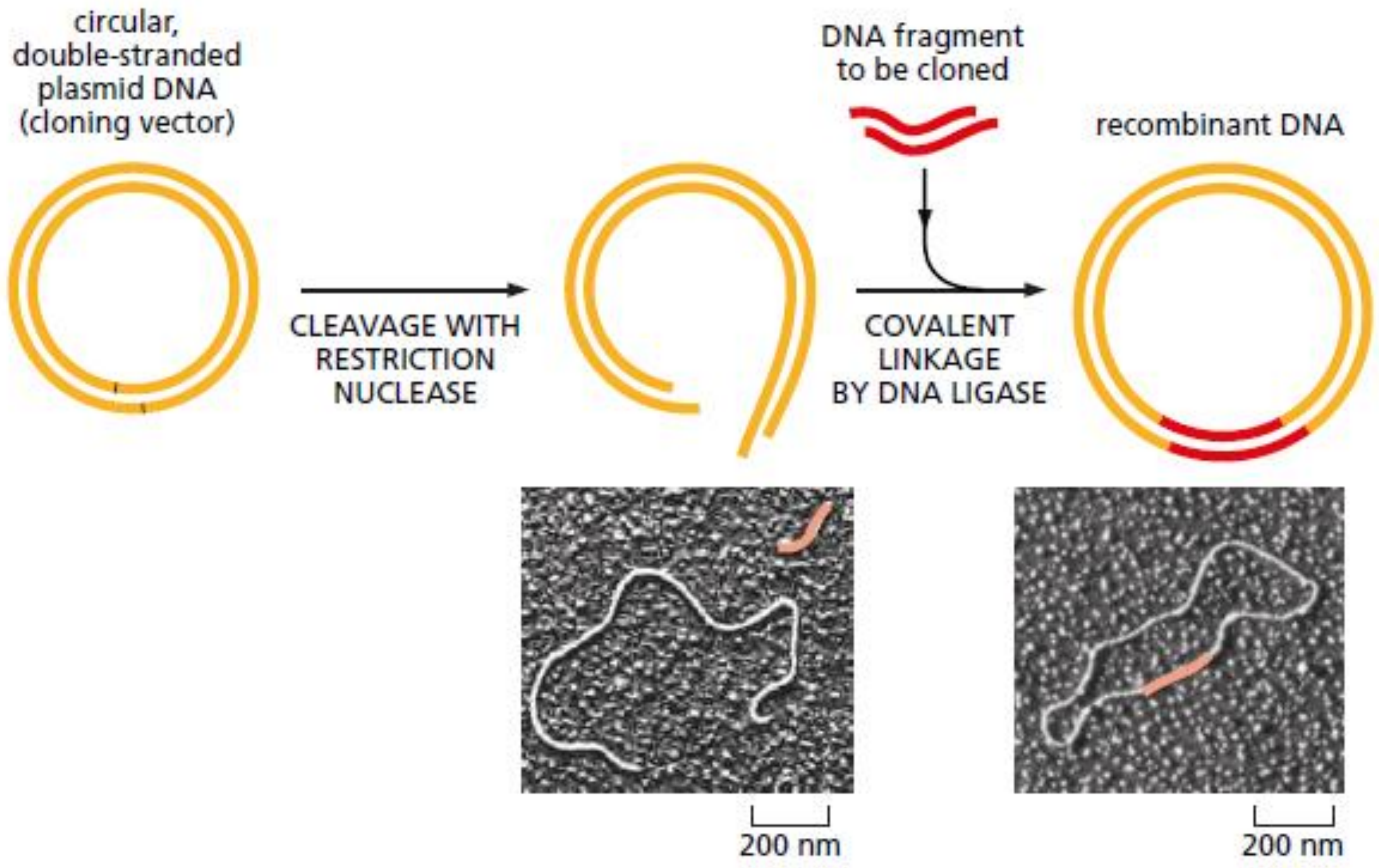


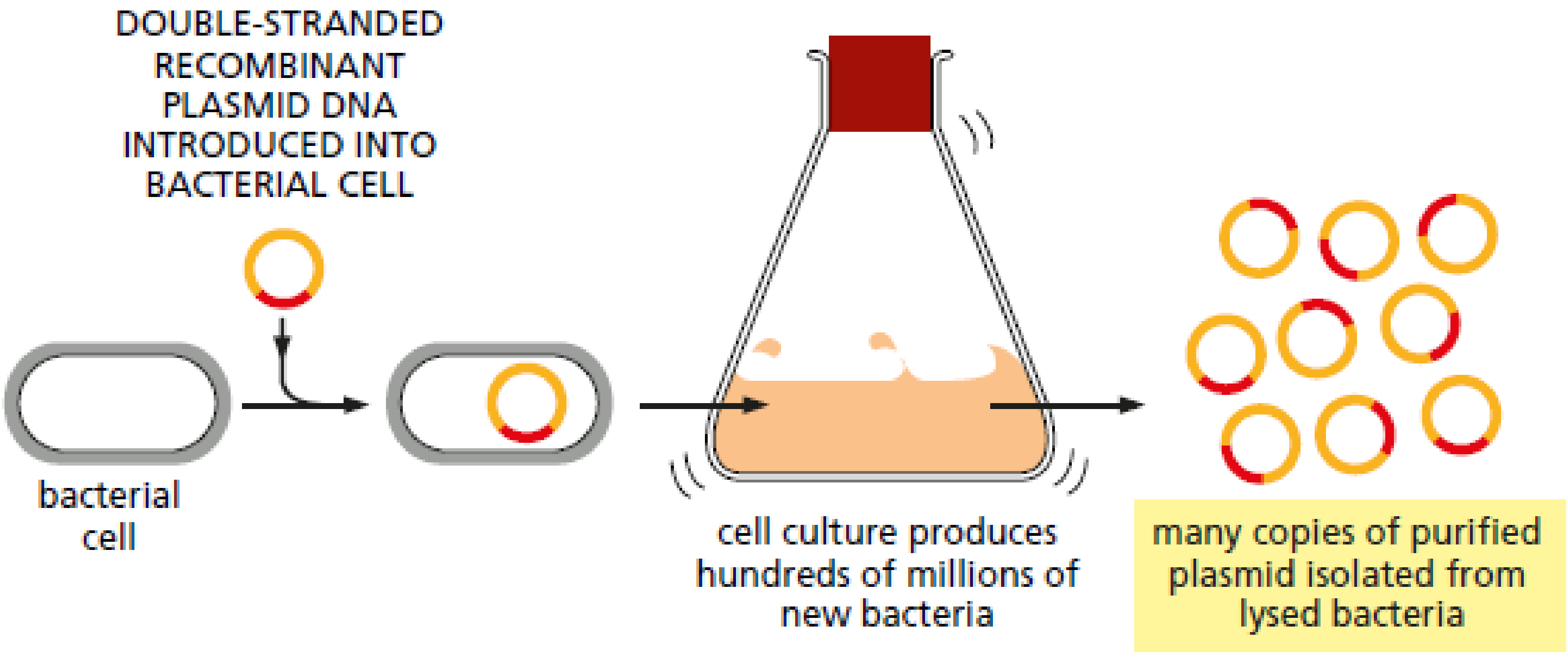


(A) JOINING TWO FRAGMENTS CUT BY THE SAME RESTRICTION NUCLEASE

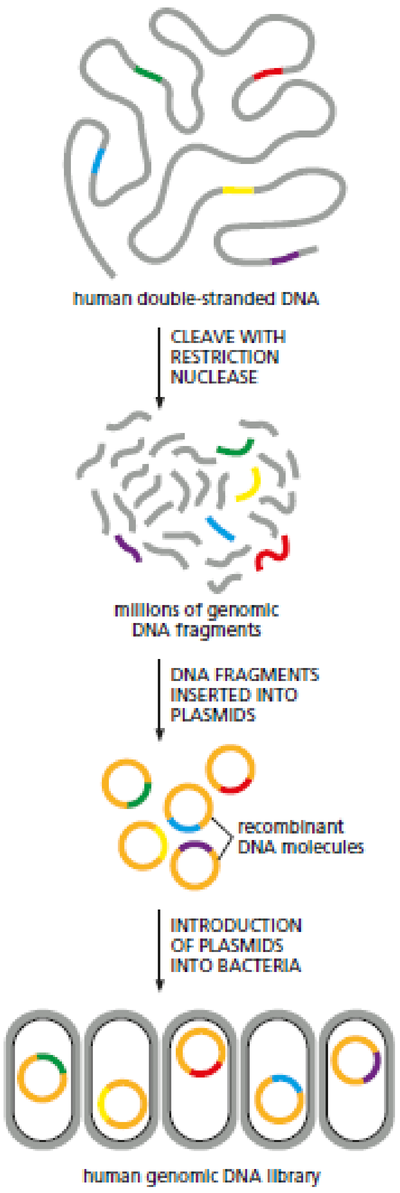
(B) JOINING TWO FRAGMENTS CUT BY DIFFERENT RESTRICTION NUCLEASES

The insertion of a DNA fragment into a bacterial plasmid with the enzyme DNA ligase



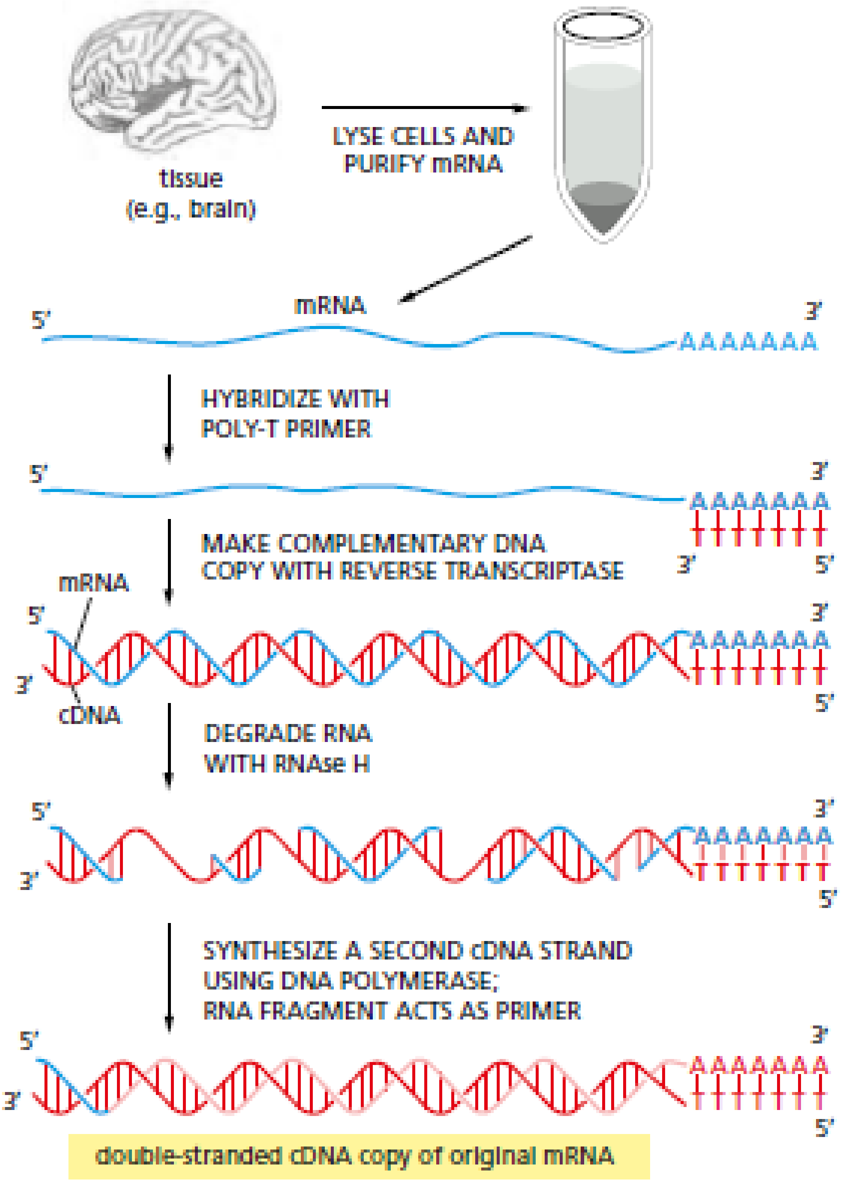

DOUBLE-STRANDED RECOMBINANT PLASMID DNA INTRODUCED INTO BACTERIAL CELL
bacterial cell
cell culture produces hundreds of millions of new bacteria
many copies of purified plasmid isolated from lysed bacteria

An Entire Genome Can Be Represented in a DNA Library



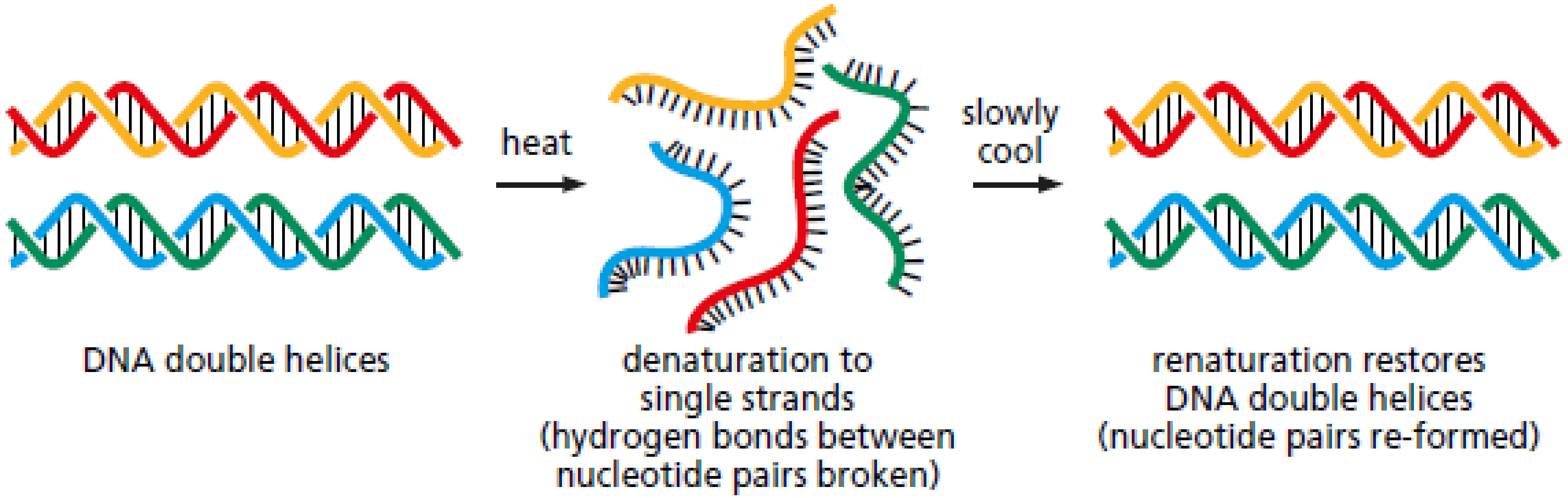
Genomic library consists of a set of bacteria, each carrying a different fragment of human DNA

cDNA library



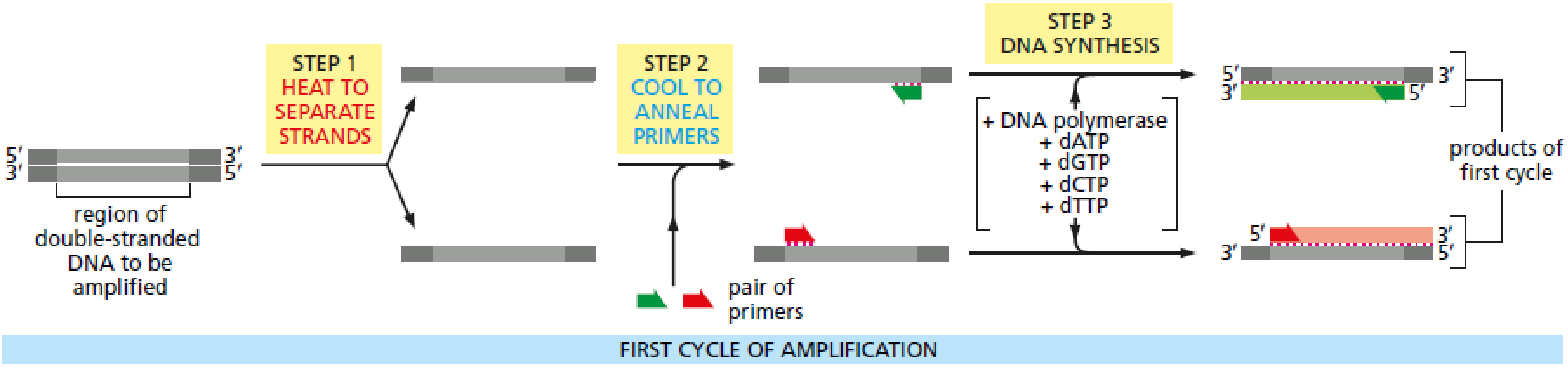
DNA-RNA hybrid

Hybridization Provides a Sensitive Way to Detect Specific Nucleotide Sequences



DNA probe typically 30 nucleotides in length can be synthesized to target certain sequences by hybridization.

Genes Can Be Cloned *in vitro* Using PCR



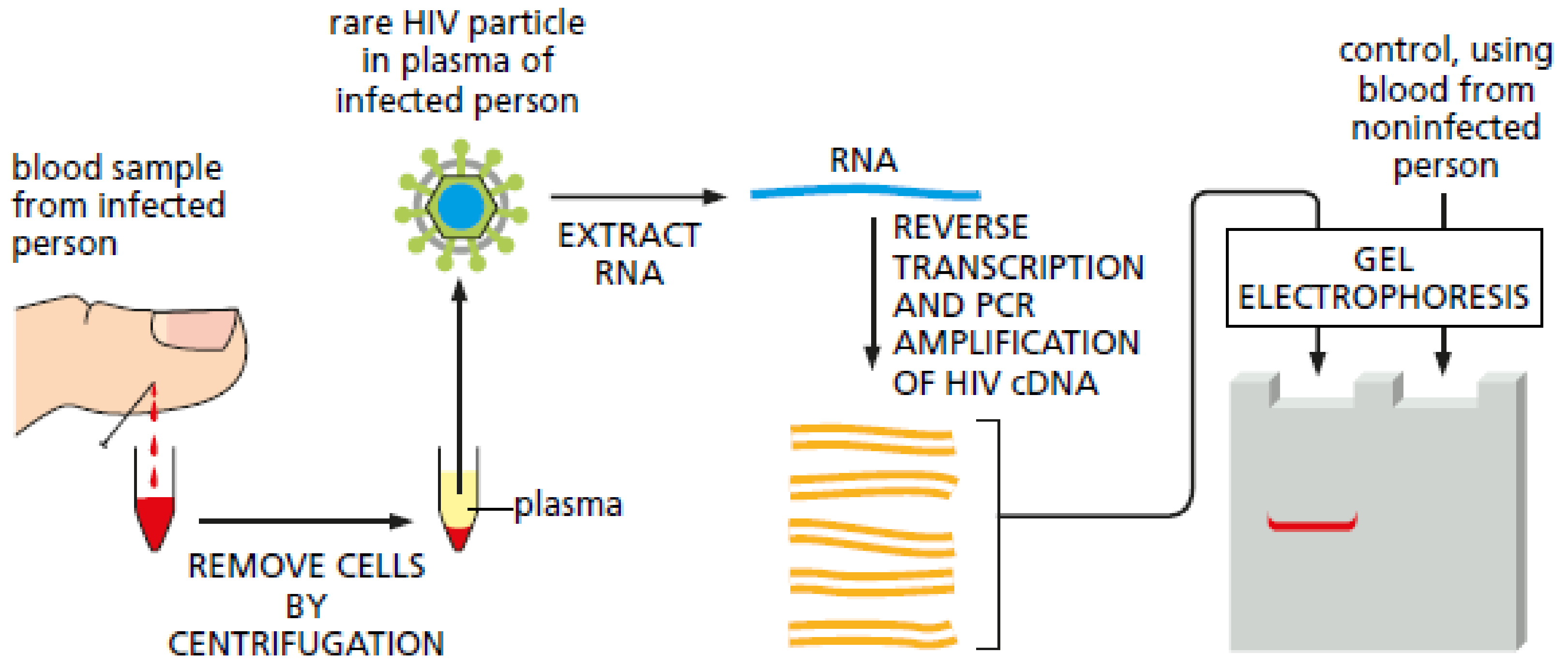
It uses go beyond simple cloning !!

PCR is also used for Diagnostic and Forensic Applications

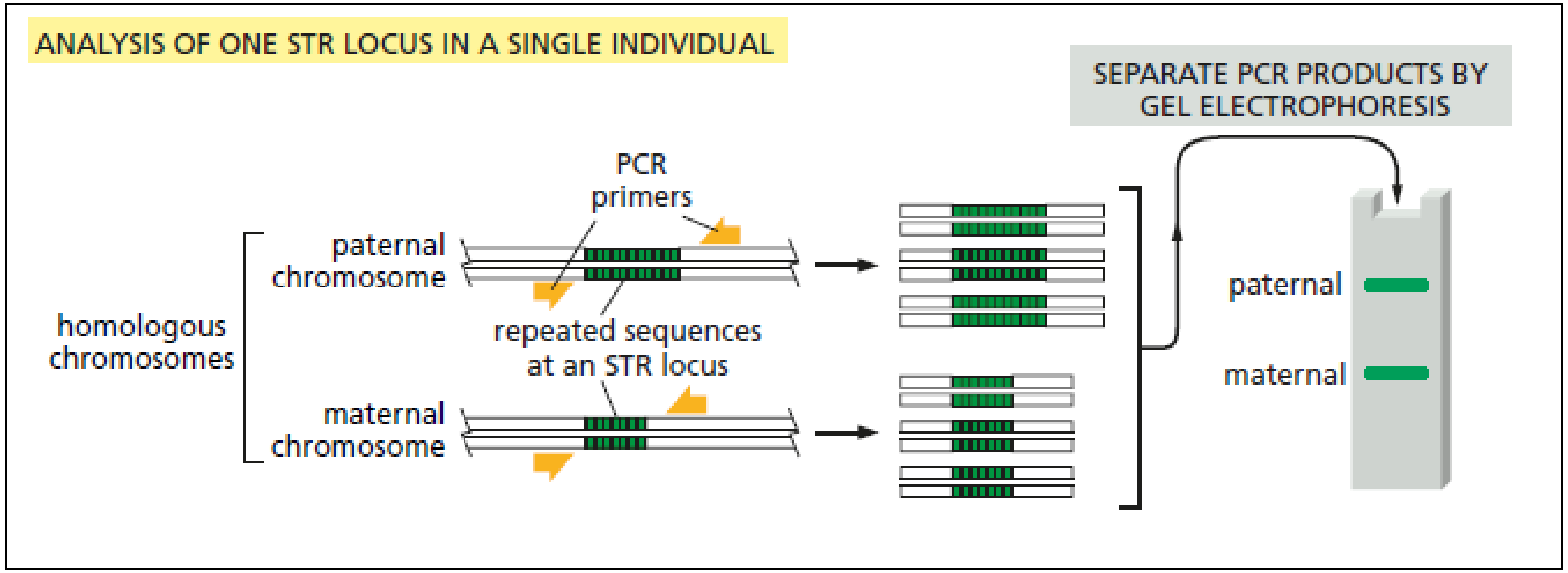
The PCR method is extraordinarily sensitive; it can detect a single DNA molecule in a sample if at least part of the sequence of that molecule is known.

It also used to verify the authenticity of a food source—for example, whether a sample of beef actually came from a cow (no politics please!)

PCR can be used to detect the presence of a viral genome in a sample of blood



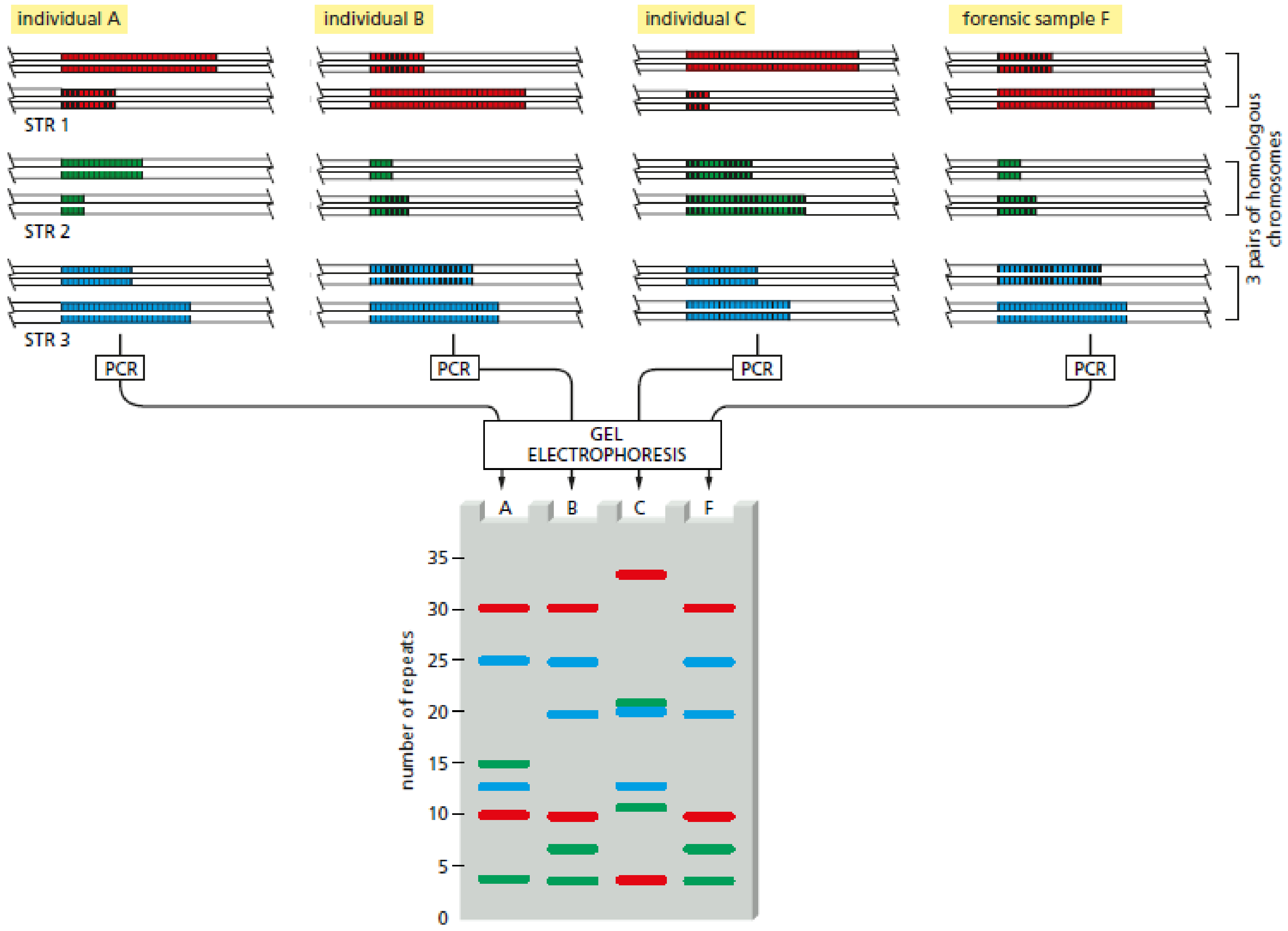
DNA fingerprinting



The DNA sequences analyzed are short tandem repeats (STRs) composed of sequences such as CACACA… or GTGTGT… STRs are found in various positions (loci) in the human genome.

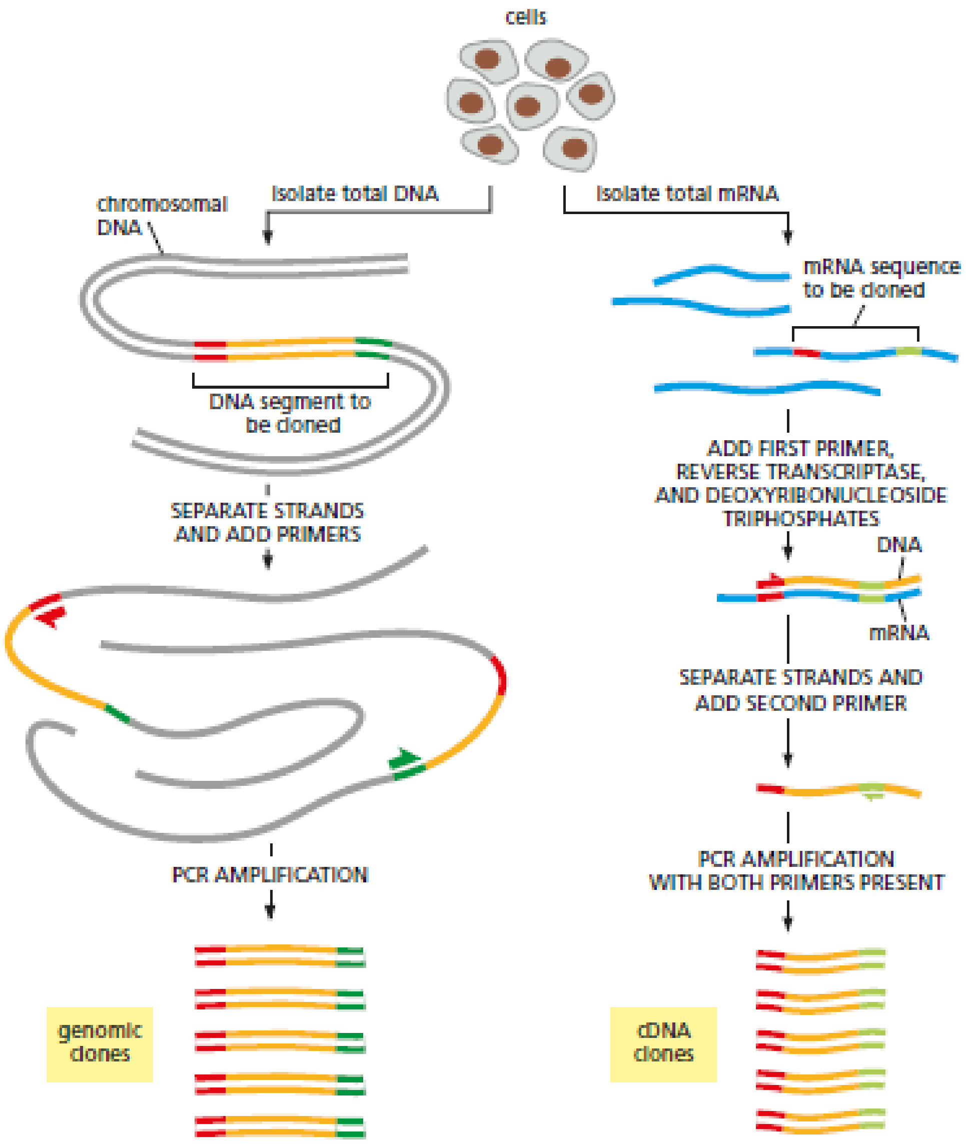
The number of repeats in each STR locus is highly variable in the population, ranging from 4 to 40 in different individuals.

Because of the variability in these sequences, individuals will usually inherit a different number of repeats at each STR locus from their mother and from their father.



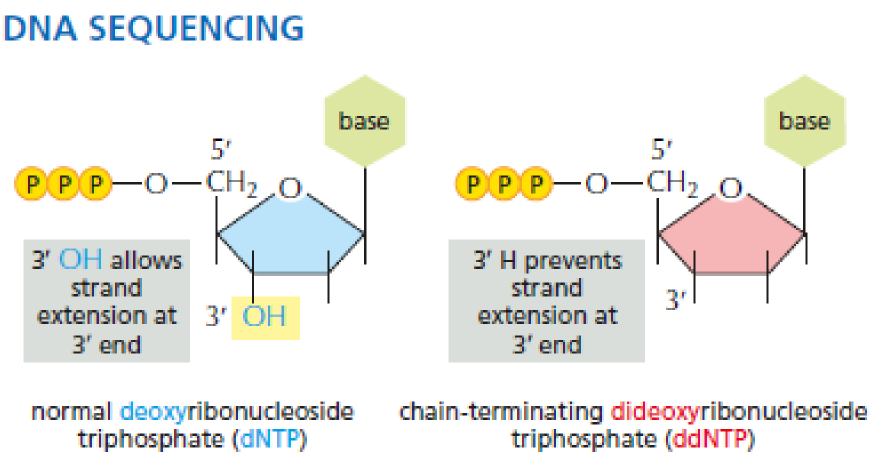
The more loci that are examined, the more confident one can be about the results. When examining the variability at 5–10 different STR loci, the odds that two random individuals would share the same fingerprint by chance are approximately one in 10 billion.

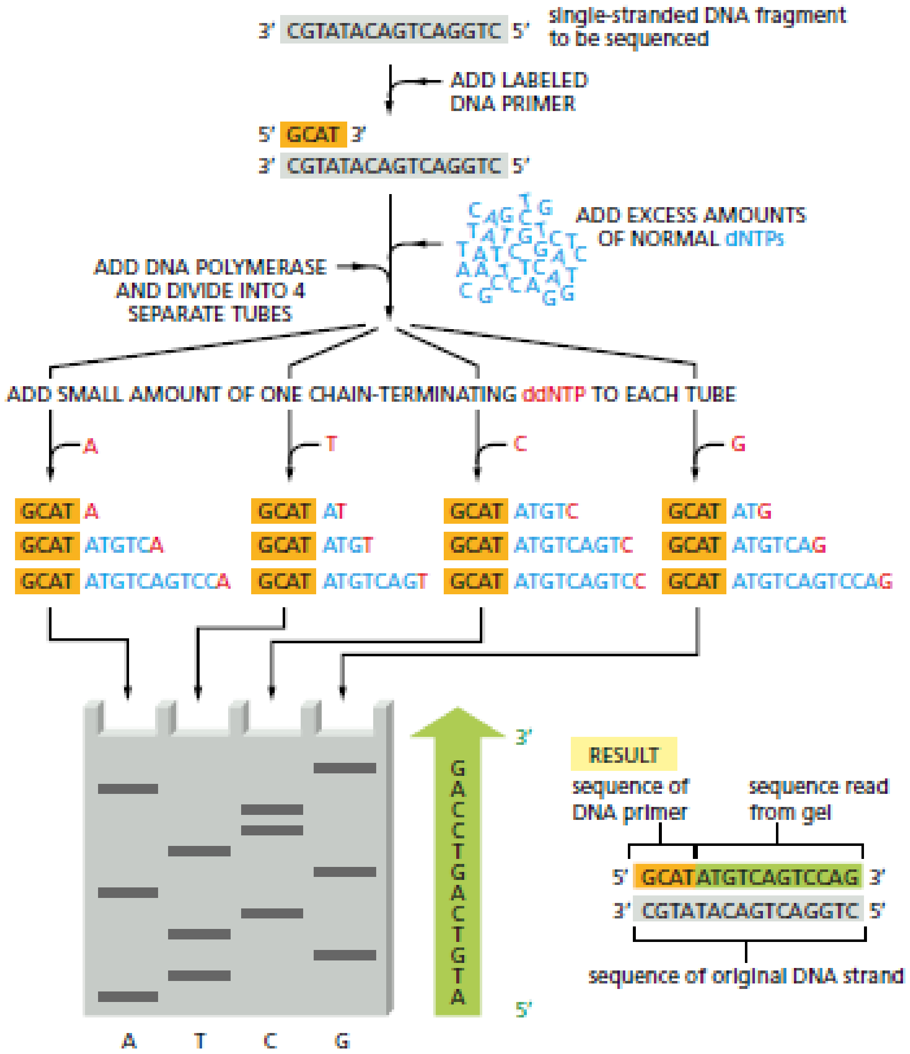
PCR can be used to obtain either genomic or cDNA clones



Both DNA and RNA Can Be Rapidly Sequenced

Dideoxy sequencing, or **Sanger sequencing**

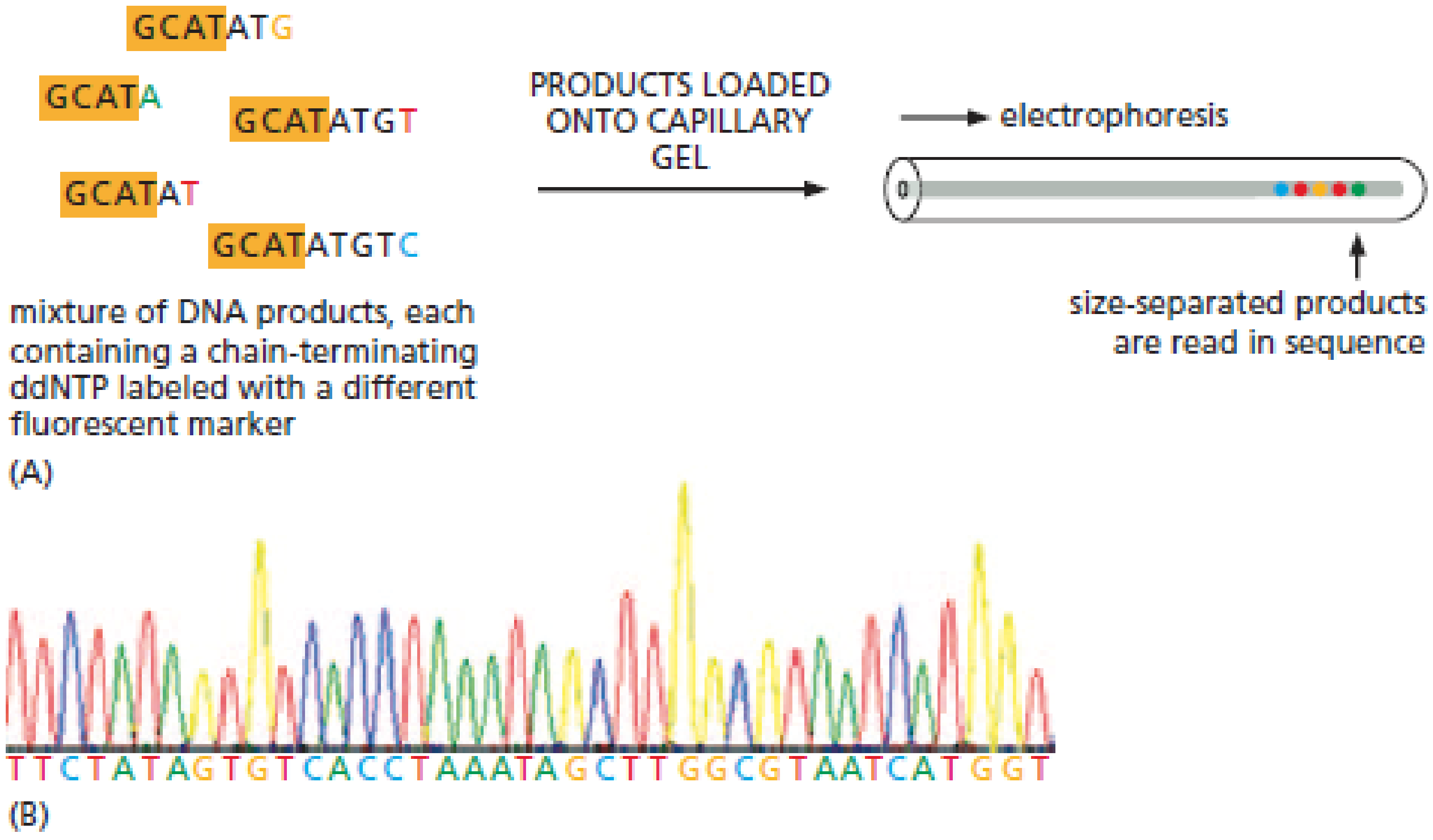




The DNA is first hybridized with a short DNA primer *(orange)* that is labeled with a fluorescent dye or radioisotope

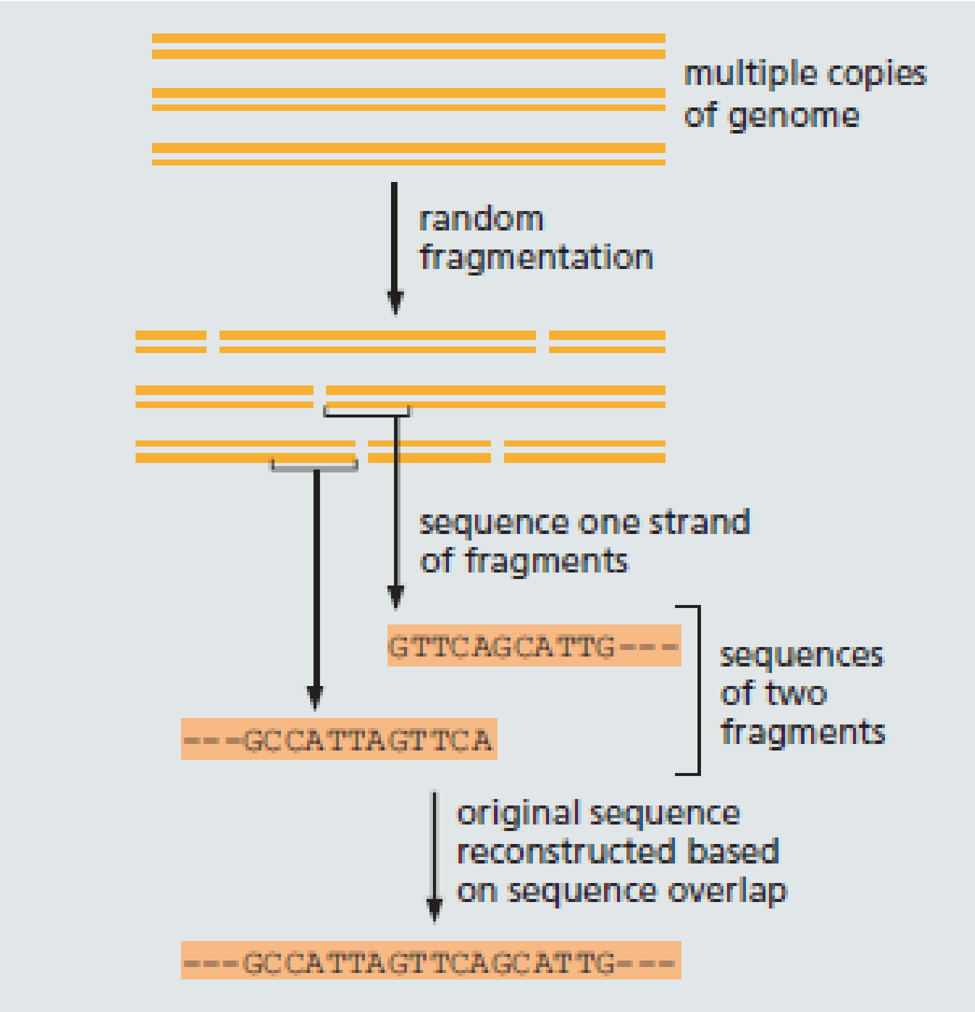
The Sanger method produces four sets of labeled DNA molecules.

Fully automated machines can set up and run Sanger sequencing reactions



The automated method uses an excess amount of normal dNTPs plus a mixture of four different chain-terminating ddNTPs, each of which is labeled with a fluorescent tag of a different colour

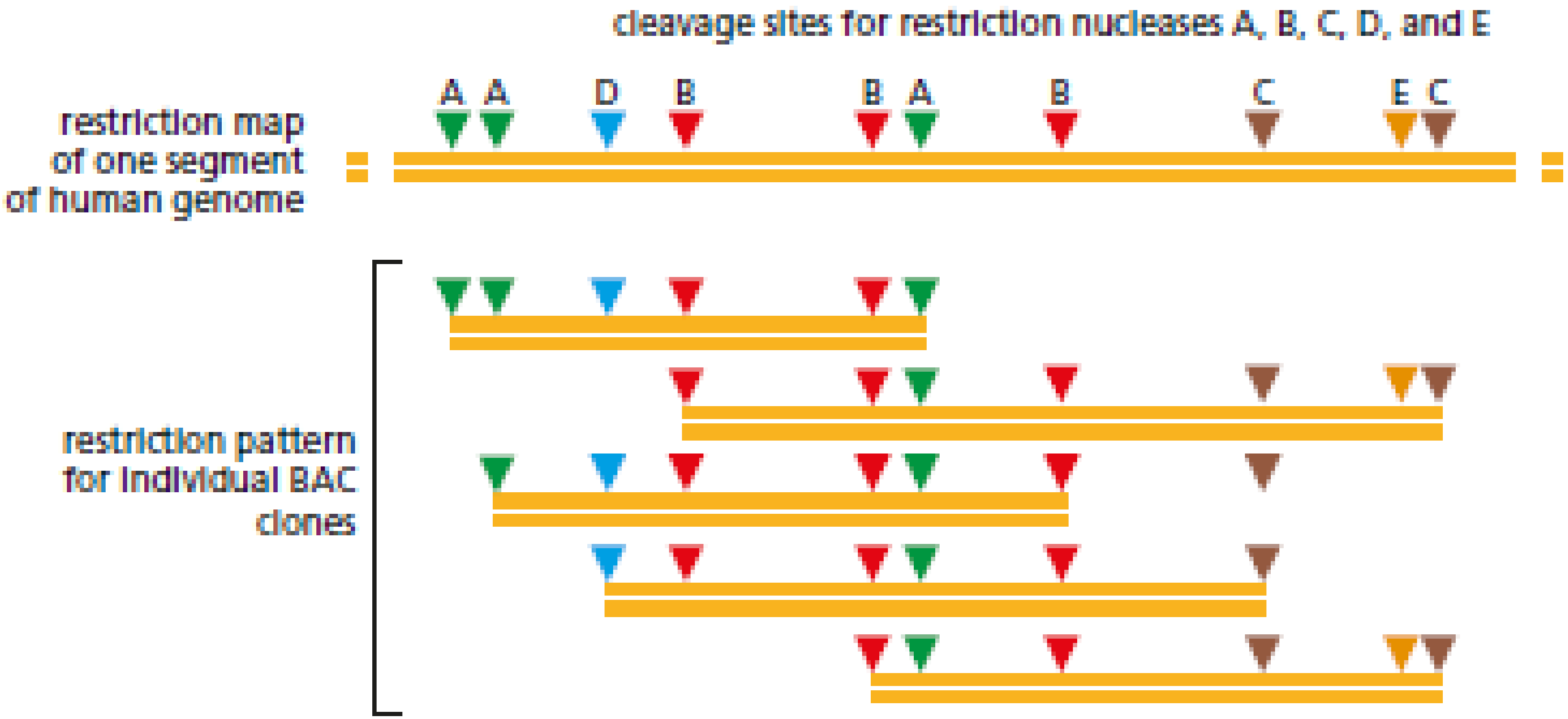
Shotgun sequencing is the method of choice for small genomes



Strand	Sequence
Original	AGCATGCTGCAGTCATGCTTAGGCTA
First shotgun sequence	AGCATGCTGCAGTCATGCT------- -------------------TAGGCTA
Second shotgun sequence	AGCATG-------------------- ------CTGCAGTCATGCTTAGGCTA
Reconstruction	AGCATGCTGCAGTCATGCTTAGGCTA

The shotgun method works well for small genomes (such as those of viruses and bacteria) that lack repetitive DNA.

BAC clones



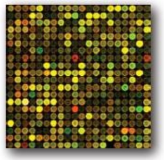
Most plant and animal genomes are large (often over 10^9 nucleotide pairs) and contain extensive amounts of repetitive DNA spread throughout the genome.

The order of the BACs along a chromosome was determined **by comparing the pattern of restriction enzyme cleavage sites** in a given BAC clone with that of the whole genome.

Sequencing technologies

Sanger DNA sequencing

1977-1990s



DNA Microarrays

Since mid-1990s

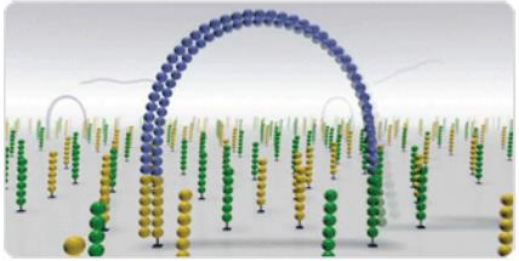
2nd-generation DNA sequencing

Since ~2007

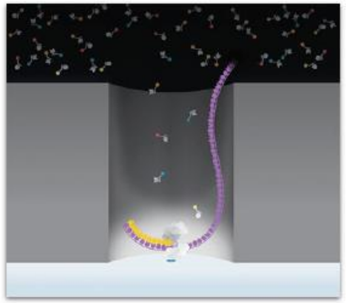
3rd-generation & single-molecule DNA sequencing

Since ~2010

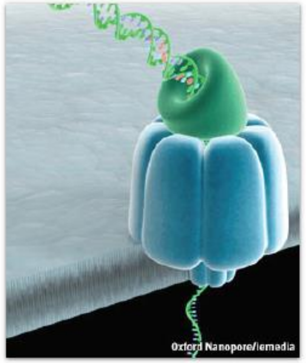
Since 2005, many DNA sequencing instruments have been described and released. They are based on a few different principles



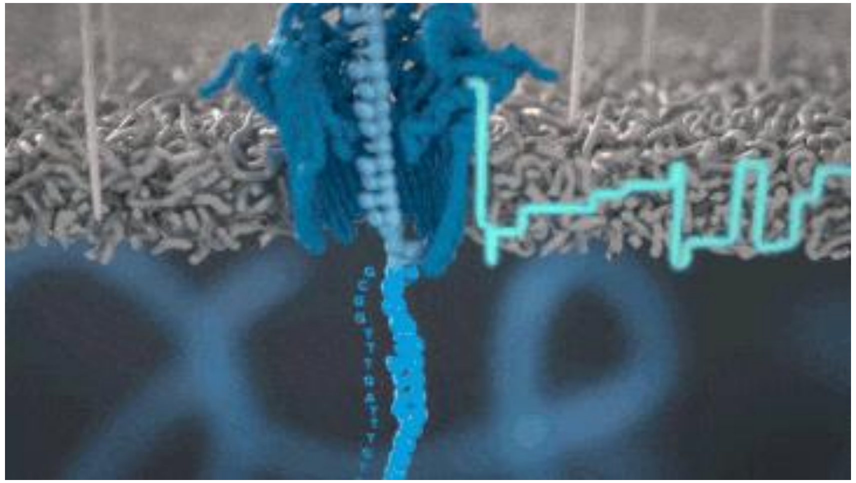
Synthesis / ligation



SMRT cell



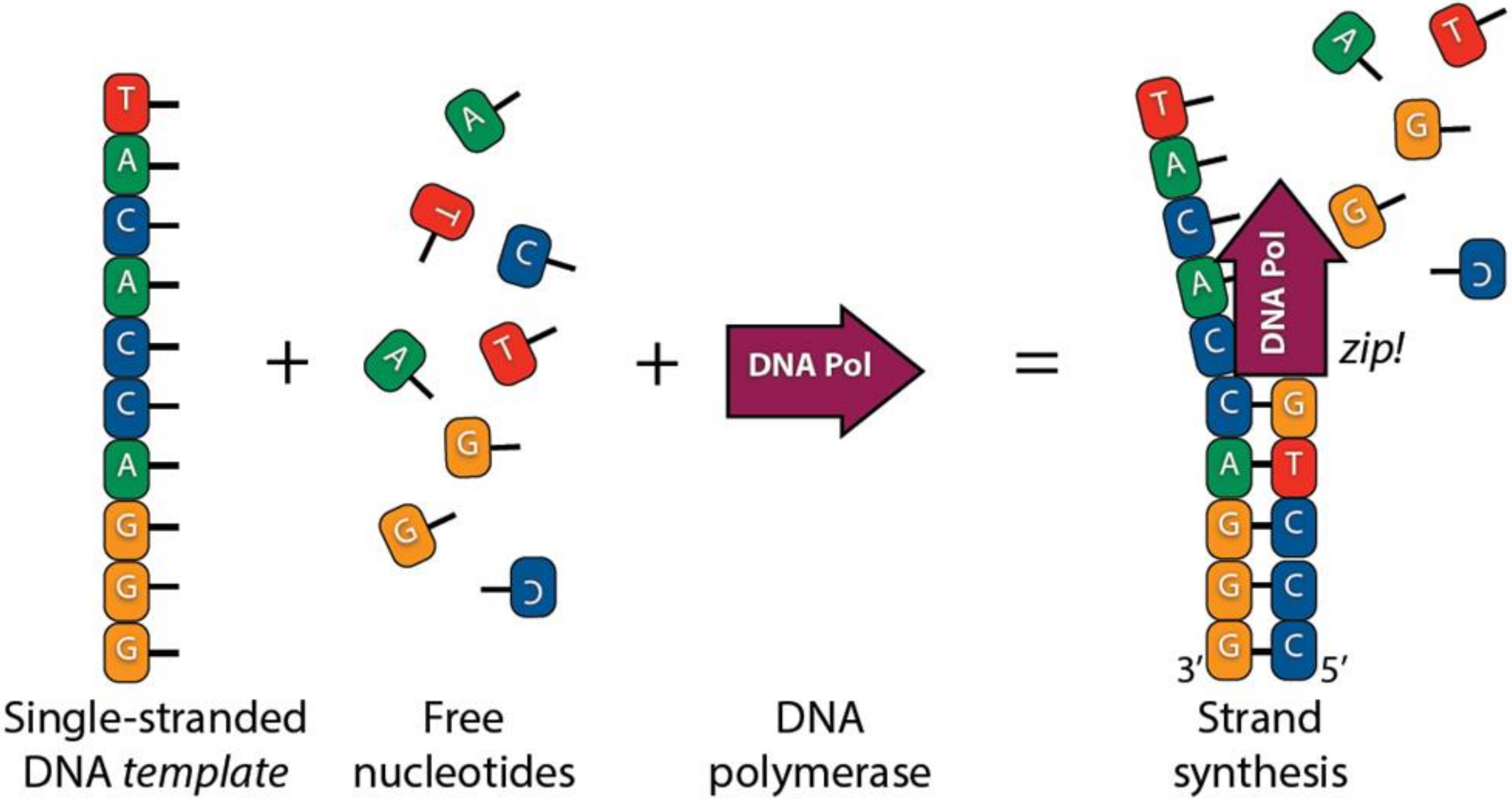
Nanopore



Sequencing by synthesis ("massively parallel sequencing") provides greatest throughput, and is the most prevalent today

Pictures: http://www.illumina.com/systems/miseq/technology.ilmn, http://www.genengnews.com/gen-articles/third-generation-sequencing-debuts/3257/

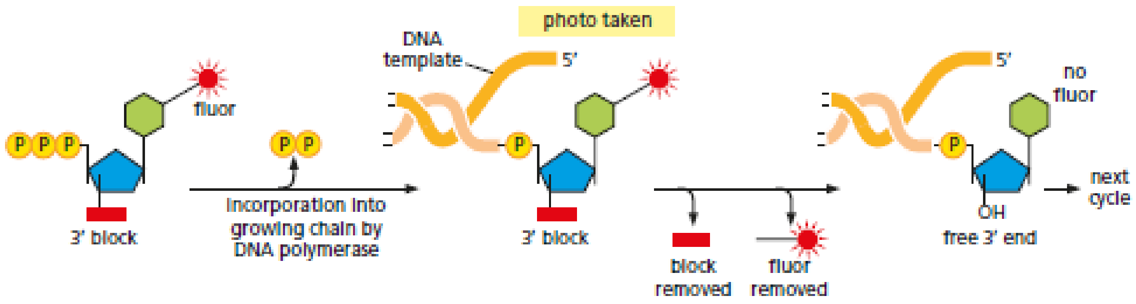
Sequencing technologies

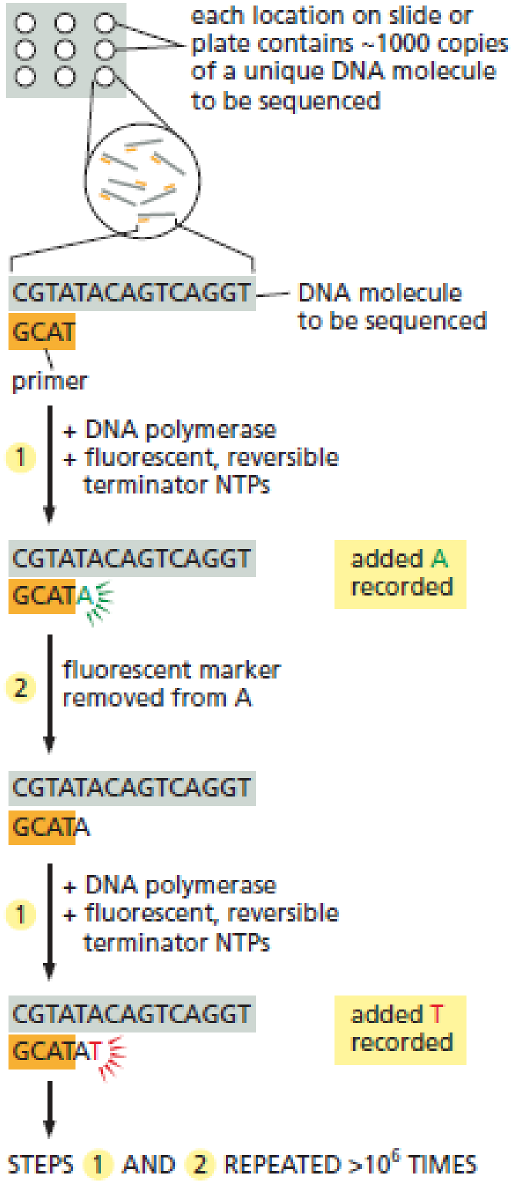


DNA polymerase moves along the template in one direction, integrating complementary nucleotides as it goes

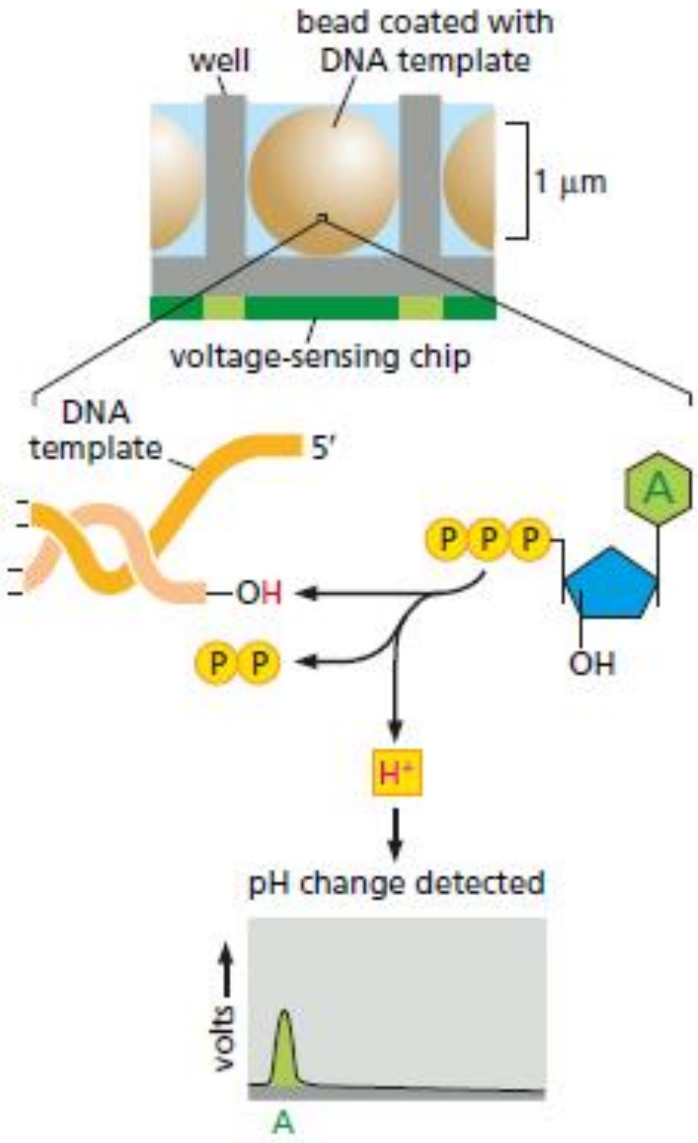
Second-generation sequencing methods rely on massively parallel sequencing reactions carried out on clusters of PCR-amplified DNA

This is the principle behind Illumina sequencing

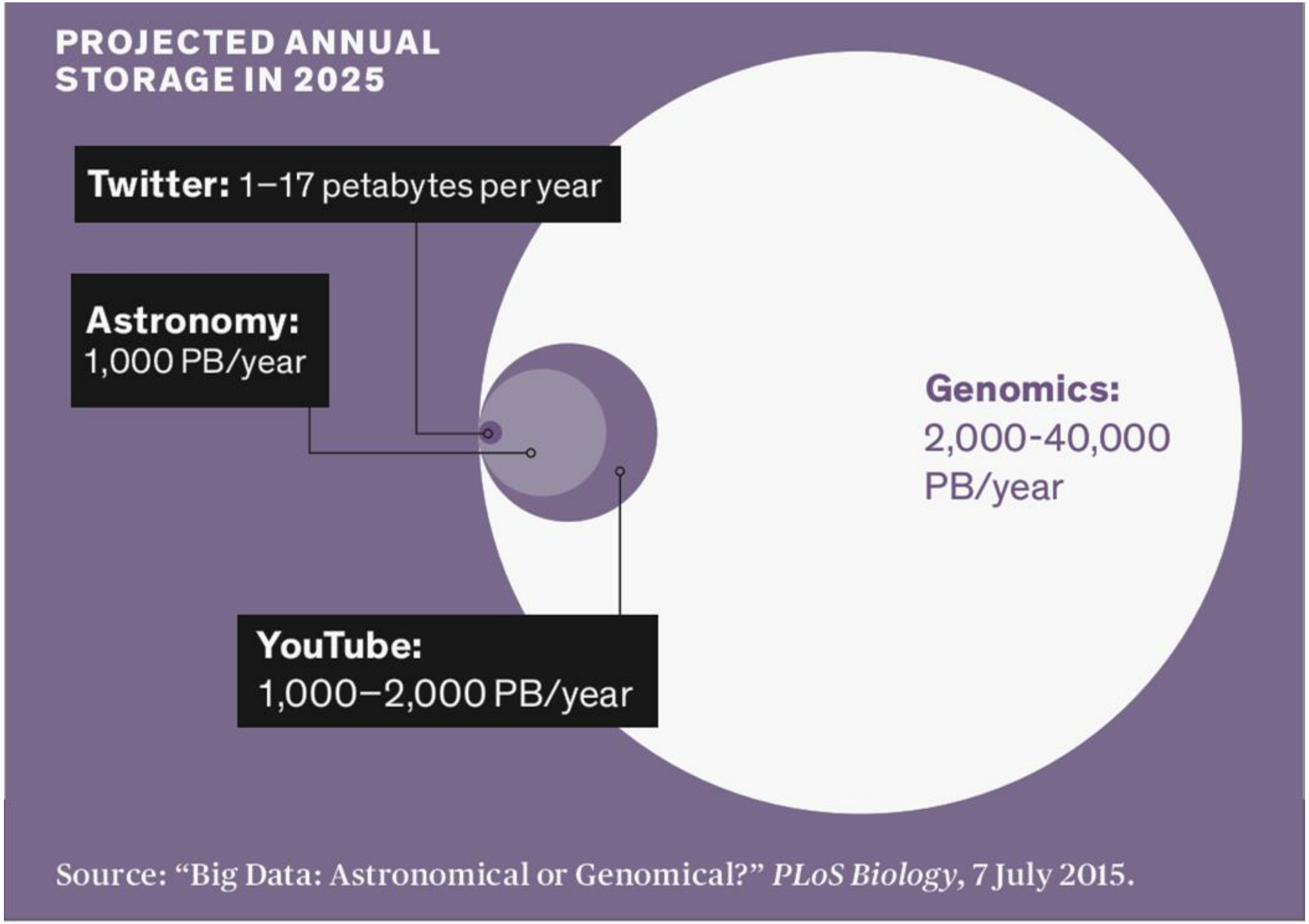




Ion torrent sequencing



Big Data – Biological sequences



Different Platforms

illumina

454 SEQUENCING

Roche

Applied Biosystems

Eric Lander’s seven-word “ Nano Lecture”

“Genome: bought the book; hard to read”

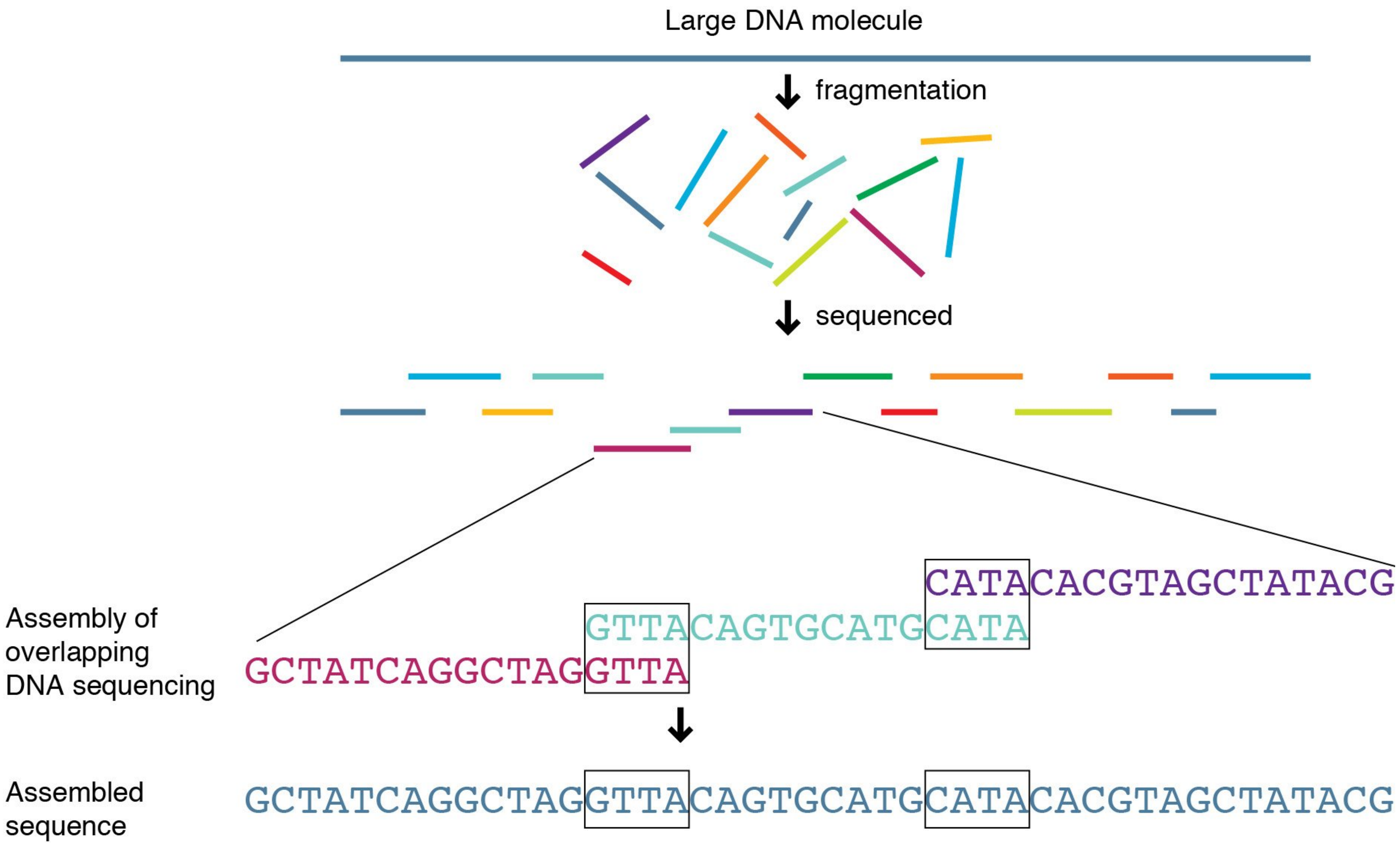
TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATATTCTACTTTTACCATTGAAAATATTGAGGAAGTTATTTATATTTCTATTTTTTATA
TATTATATATTTTATGTATTTTAATATTACTATTACACATAATTATTTTTTATATATATGAAGTACCAATGACTTCCTTTTCCAGAGCAATAATGAAATTTCACAGTA
TGAAAATGGAAGAAATCAATAAAATTATACGTGACCTGTGGCGAAGTACCTATCGTGGACAAGGTGAGTACCATGGTGTATCACAAATGCTCTTTCCAAAG
CCCTCTCCGCAGCTCTTCCCCTTATGACCTCTCATCATGCCAGCATTACCTCCCTGGACCCCTTTCTAAGCATGTCTTTGAGATTTTCTAAGAATTCTTATCTTG
GCAACATCTTGTAGCAAGAAAATGTAAAGTTTTCTGTTCCAGAGCCTAACAGGACTTACATATTTGACTGCAGTAGGCATTATATTTAGCTGATGACATAATA
GGTTCTGTCATAGTGTAGATAGGGATAAGCCAAAATGCAATAAGAAAAACCATCCAGAGGAAACTCTTTTTTTTTTCTTTTTCTTTTTTTTTTTTCCAGATG
GAGTCTCGCACTTCTCTGTCACCCGGGCTGGAGCGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCTGGGTTCAGGTGATTCTCCCACCTCAG
CCTCCCGAGTAGTAGCTGGAATTACAGGTGCGCGCTCCCACACCTGGCTAATTTTTTGTATTCTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGG
TCTCAAACTCCTGCCCTCAGGTGATCTGCCCACCTTGGCCTCCCAGTGTTGGGTTTACAGGCGTGAGCCACCGCGCCTGGCCTGGAGGAAACTCTTAACA
GGGAAACTAAGAAAGAGTTGAGGCTGAGGAACTGGGGCATCTGGGTTGCTTCTGGCCAGACCACCAGGCTCTTGAATCCTCCCAGCCAGAGAAAGAG
TTTCCACACCAGCCATTGTTTTCCTCTGGTAATGTCAGCCTCATCTGTTGTTCCTAGGCTTACTTGATATGTTTGTAAATGACAAAAGGCTACAGAGCATAGG
TTCCTCTAAAATATTCTTCTTCCTGTGTCAGATATTGAATACATAGAAATACGGTCTGATGCCGATGAAAATGTATCAGCTTCTGATAAAAGGCGGAATTATAA
CTACCGAGTGGTGATGCTGAAGGGAGACACAGCCTTGGATATGCGAGGACGATGCAGTGCTGGACAAAA

Can we predict genetic diseases from DNA alone?

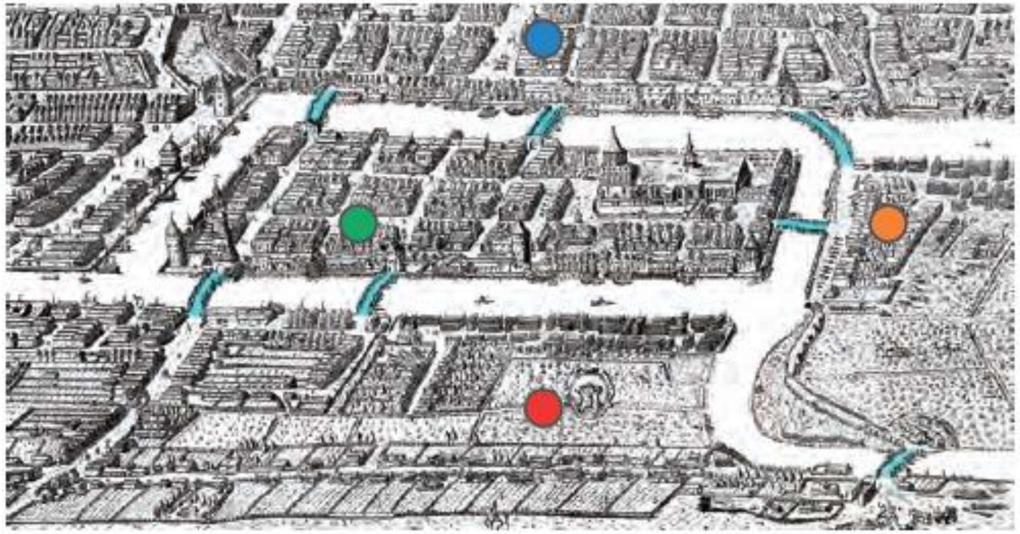
TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATATTCTACTT
TTACCATTGAAAATATTGAGGAAGTTATTTATATTTCTATTTTTTATATATTATATATTTT
ATGTATTTTAATATTACTATTACACATAATTATTTTTTATATATATGAAGTACCAATGACT
TCCTTTTCCAGAGCAATAATGAAATTTCACAGTATGAAAATGGAAGAAATCAATAAA
ATTATACGTGACCTGTGGCGAAGTACCTATCGTGGACAAGGTGAGTACCATGGTGT
ATCACAAATGCTCTTTCCAAAGCCCTCTCCGCAGCTCTTCCCCTTATGACCTCTCATC
ATGCCAGCATTACCTCCCTGGACCCCTTTCTAAGCATGTCTTTGAGATTTTCTAAGA
ATTCTTATCTTGGCAACATCTTGTAGCAAGAAAATGTAAAGTTTTCTGTTCCAGAGC
CTAACAGGACTTACATATTTGACTGCAGTAGGCATTATATTTAGCTGATGACATAATA
GGTTCTGTCATAGTGTAGATAGGGATAAGCCAAAATGCAATAAGAAAAACCATCCA
GAGGAAACTCTTTTTTTTTTCTTTTTCTTTTTTTTTTTTCCAGATGGAGTCTCGCACT
TCTCTGTCACCCGGGCTGGAGCGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCA
CCTCCTGGGTTCAGGTGATTCTCCCACCTCAGCCTCCCGAGTAGTAGCTGGAATTAC
AGGTGCGCGCTCCCACACCTGGCTAATTTTTTGTATTCTTAGTAGAGATGGGGTTTC
ACCATGTTGGCCAGGCTGGTCTCAAACTCCTGCCCTCAGGTGATCTGCCCACCTTG
GCCTCCCAGTGTTGGGTTTACAGGCGTGAGCCACCGCGCCTGGCCTGGAGGAAA
CTCTTAACAGGGAAACTAAGAAAGAGTTGAGGCTGAGGAACTGGGGCATCTGGG
TTGCTTCTGGCCAGACCACCAGGCTCTTGAATCCTCCCAGCCAGAGAAAGAGTTTC
CACACCAGCCATTGTTTTCCTCTGGTAATGTCAGCCTCATCTGTTGTTCCTAGGCTT

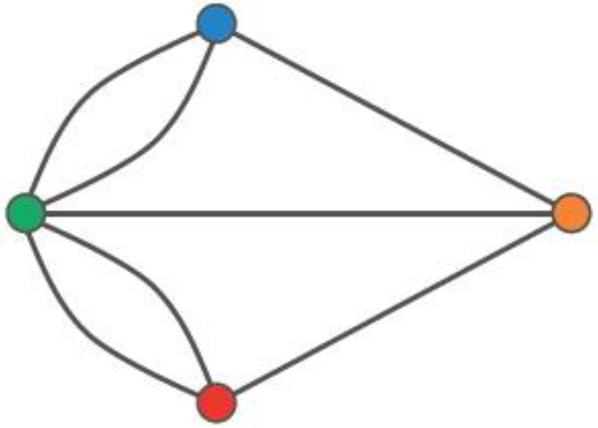
→ Probability of genetic diseases

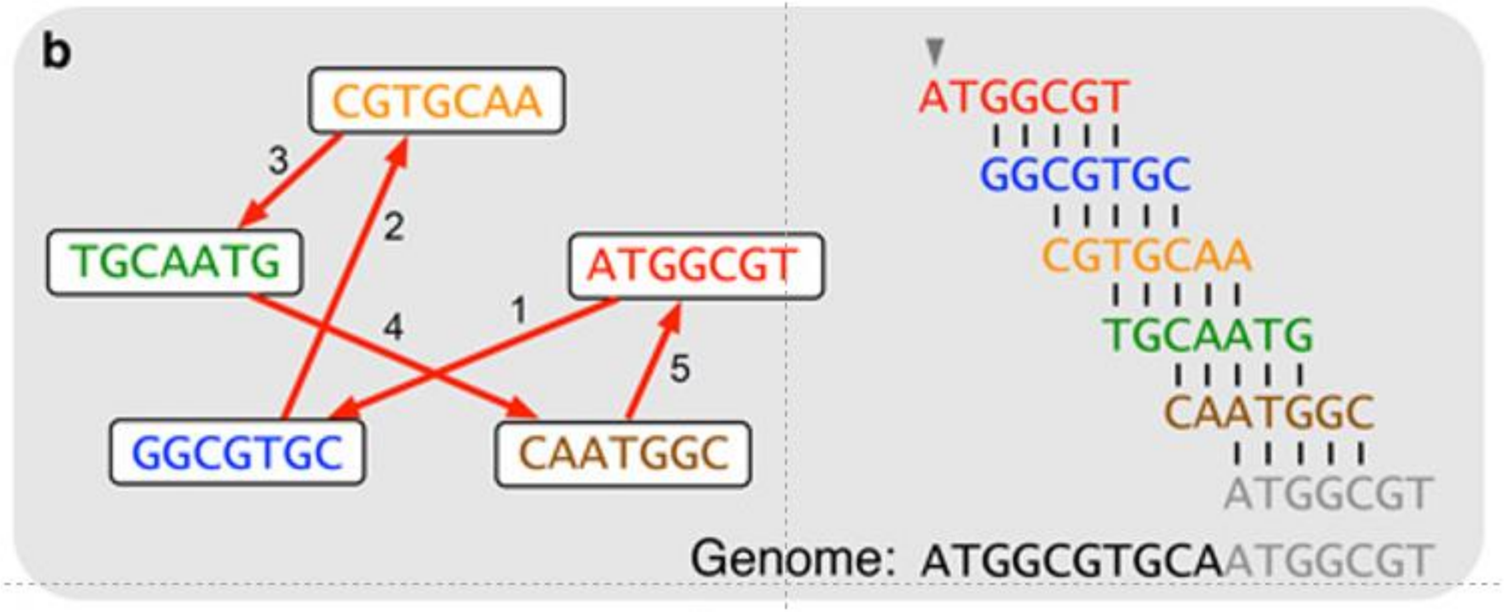
Read Mapping



Bridges of Königsberg Problem

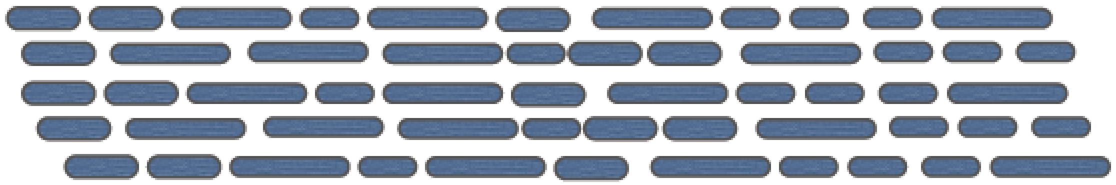




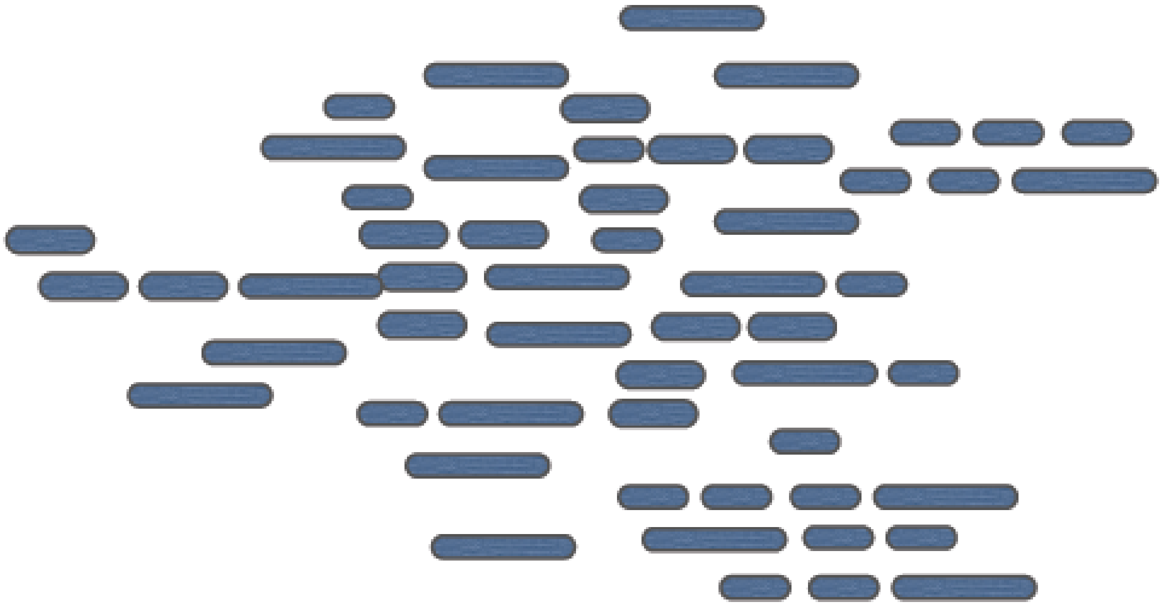

b
CGTGCAA
TGCAATG
ATGGCGT
GGCGTGC
CAATGGC
1
2
3
4
5
ATGGCGT
GGCGTGC
CGTGCAA
TGCAATG
CAATGGC
ATGGCGT
Genome: ATGGCGTGCAATGGCGT

Start with DNA
(millions of copies)

Break it



Put in sequencer



Reads

```
GTTGAGGCTTGCGTTTTTGGTACGCTGGACTTTGT
GTACTCGTCGCTGCGTTGAGGCTTGCGTTTTTGGT
ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
CTTGCGTTTATGGTACGCTGGACTTTGTAGGATAC
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
GCGTTTATGGTACGCTGGACTTTGTAGGATACCCT
GAGGCTTGCGTTTATGGTACGCTGGACTTTGTAGG
GCGTTGAGGCTTGCGTTTATGGTACGCTGGATTTT
CGTTTATGGTACGCTGGACTTTGTAGGATACCCTC
ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
GTTTATGGTACGCTGGACTTTGTAGGATACCCTCG
TCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTA
TGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTA
GCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTAC
TATGGTACGCTGGACTTTGTAGGATACCCTCGCTT
TCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTTTG
CGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCT
GTTGAGGCTTGCGTTTATGGTACGCTGGGCTTTTT
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
```

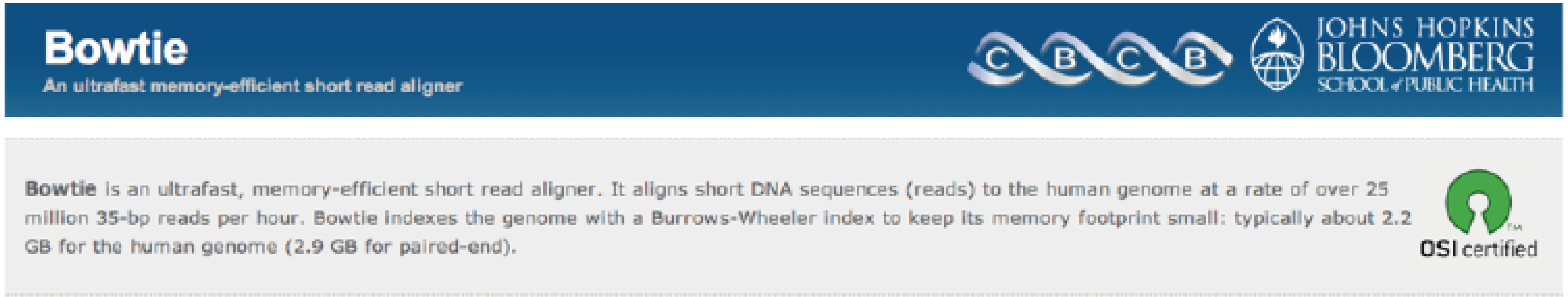
“sequence reads” (short bits of nucleotide sequence)

Start matching to reference

```
                   GTTGAGGCTTGCGTTTTTGGTACGCTGGACTTTGT
     GTACTCGTCGCTGCGTTGAGGCTTGCGTTTTTGGT
                                   ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
                           TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
                          CTTGCGTTTATGGTACGCTGGACTTTGTAGGATAC
                           TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
                             GCGTTTATGGTACGCTGGACTTTGTAGGATACCCT
                      GAGGCTTGCGTTTATGGTACGCTGGACTTTGTAGG
                 GCGTTGAGGCTTGCGTTTATGGTACGCTGGATTTT
                              CGTTTATGGTACGCTGGACTTTGTAGGATACCCTC
                                   ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
                               GTTTATGGTACGCTGGACTTTGTAGGATACCCTCG
 TCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTA
      TGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTA
       GCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTAC
                                  TATGGTACGCTGGACTTTGTAGGATACCCTCGCTT
   TCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTTTG
          CGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCT
                   GTTGAGGCTTGCGTTTATGGTACGCTGGGCTTTTT
                           TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
CTCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCTGGACTTTGTAGGATACCCTCGCTTTC
```

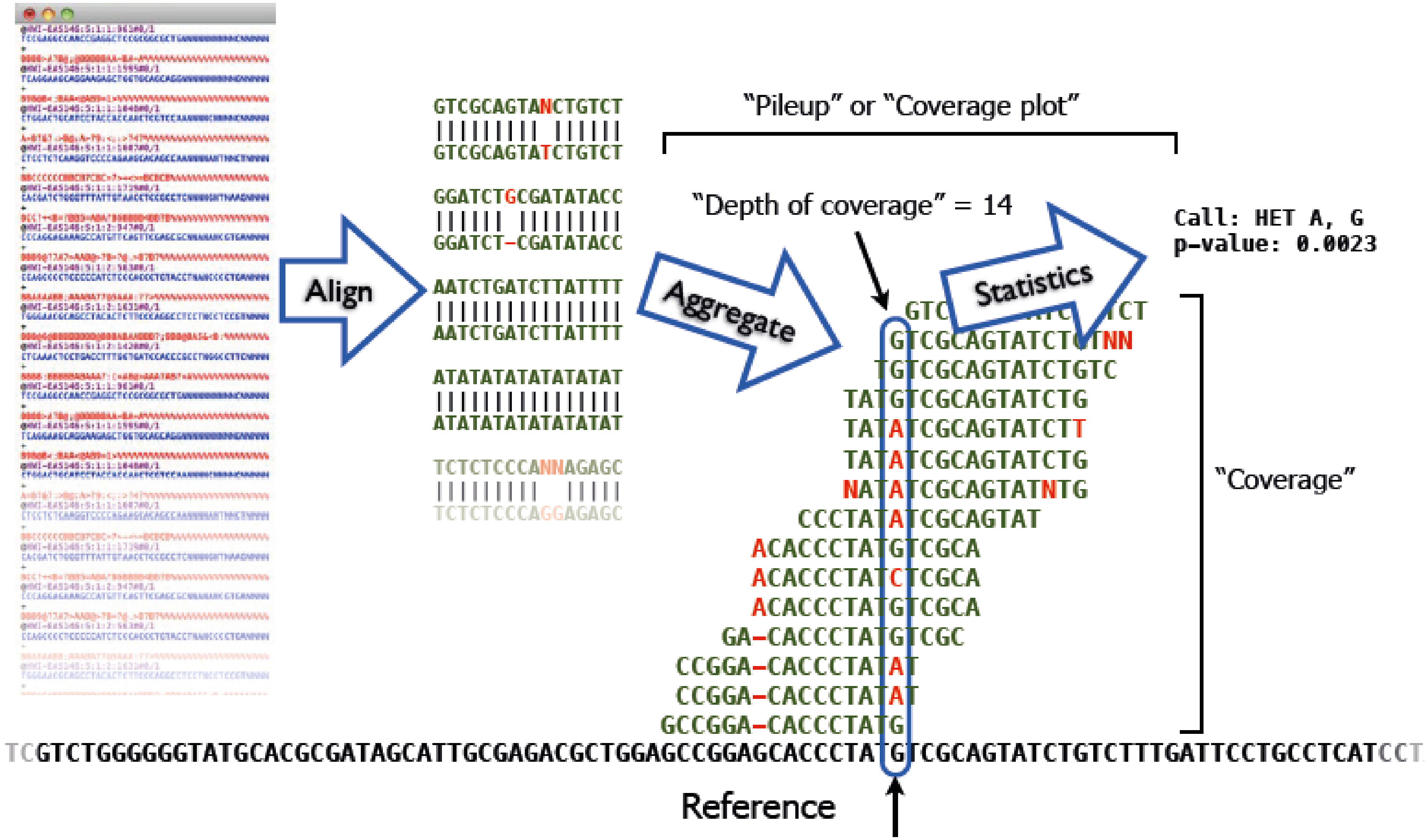
Matching

```
                    GTTGAGGCTTGCGTTTTTGGTACGCTGGACTTTGT
      GTACTCGTCGCTGCGTTGAGGCTTGCGTTTTTGGT
                                   ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
                         TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
                        CTTGCGTTTATGGTACGCTGGACTTTGTAGGATAC
```

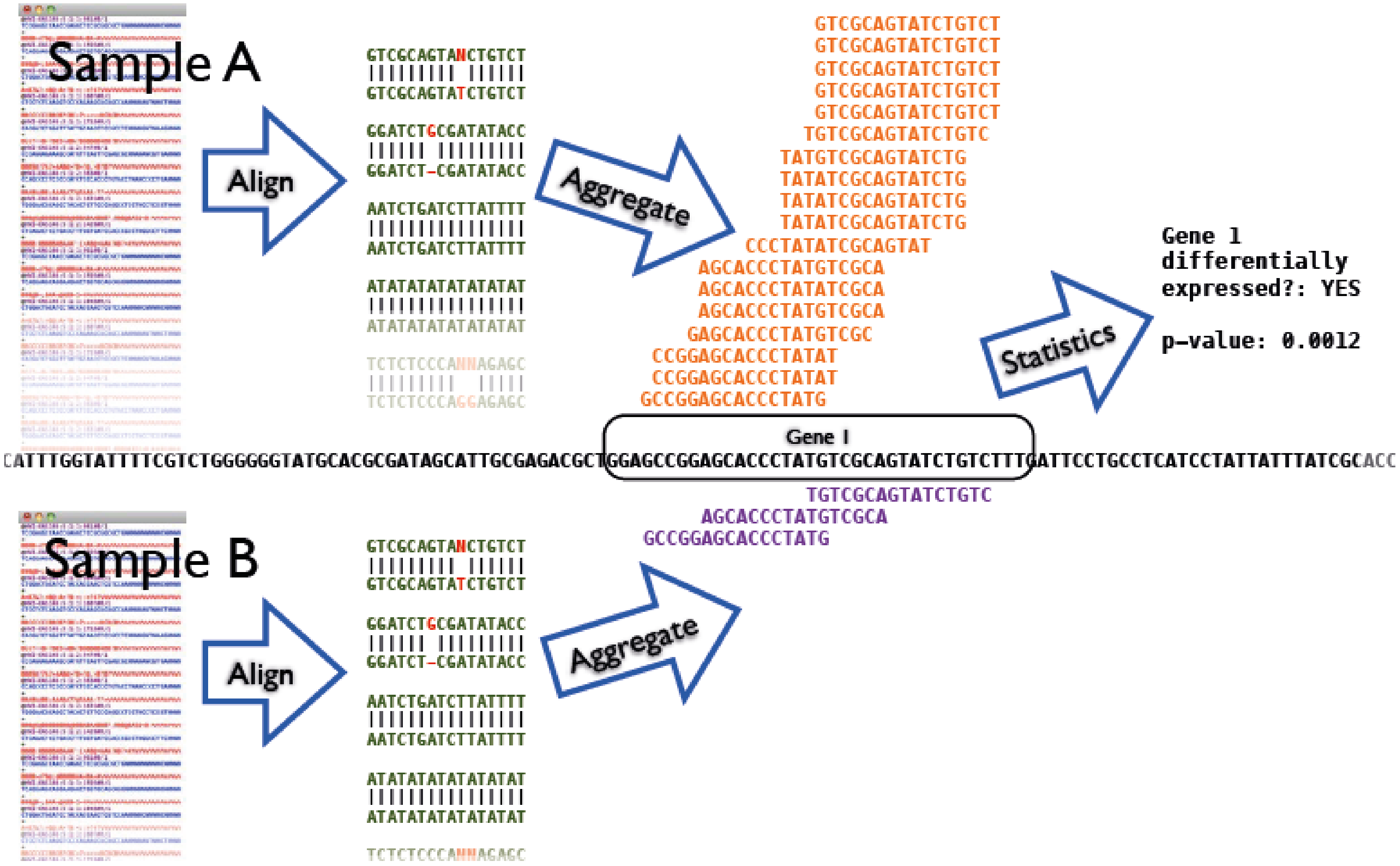



```
       TGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTA
        GCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTAC
                                  TATGGTACGCTGGACTTTGTAGGATACCCTCGCTT
    TCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTTTG
           CGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCT
                    GTTGAGGCTTGCGTTTATGGTACGCTGGGCTTTTT
                            TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
--------------------------------------------------------------------------
  CTCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCTGGACTTTGTAGGATACCCTCGCTTTC
```

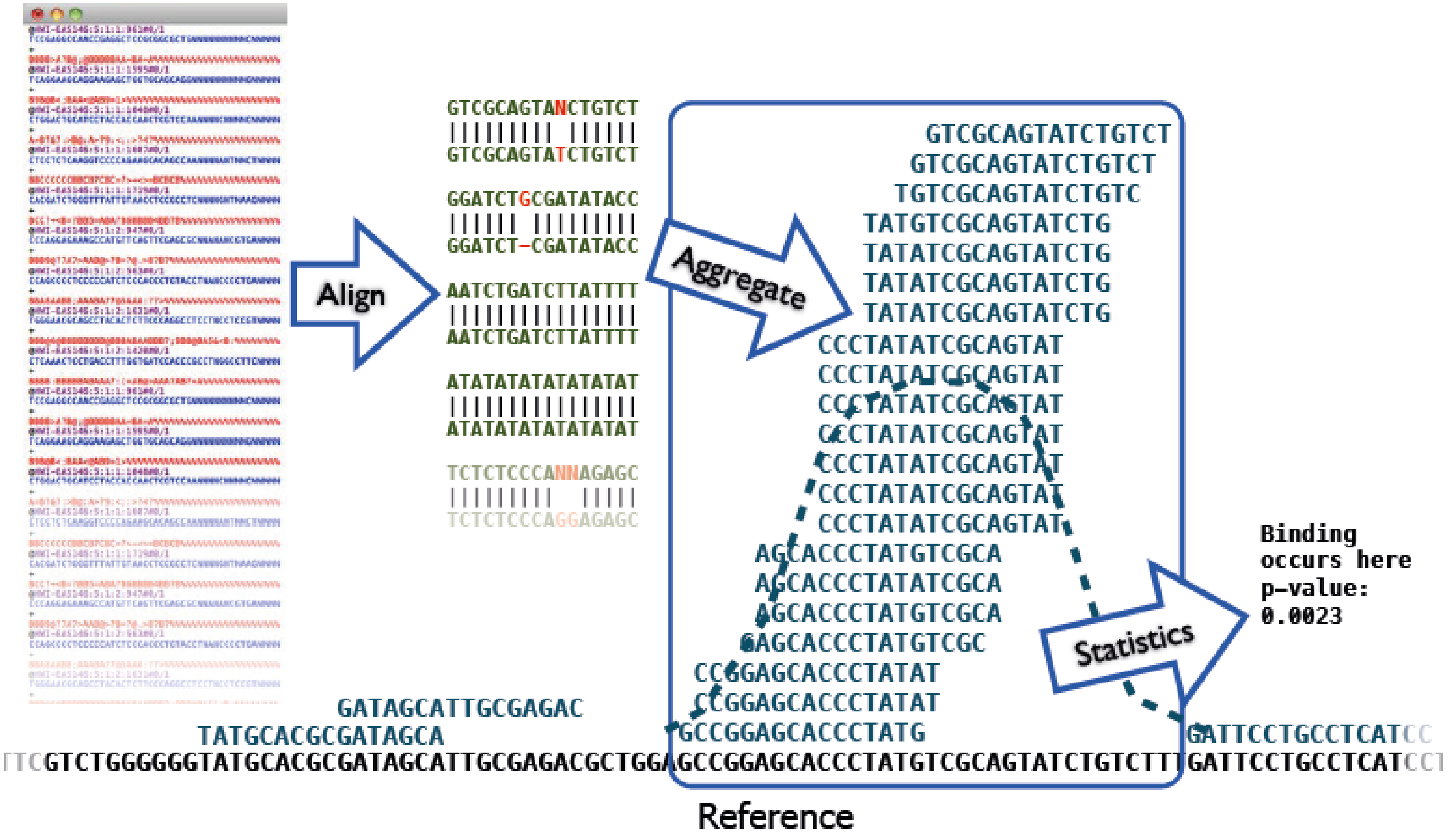
Variant Detection



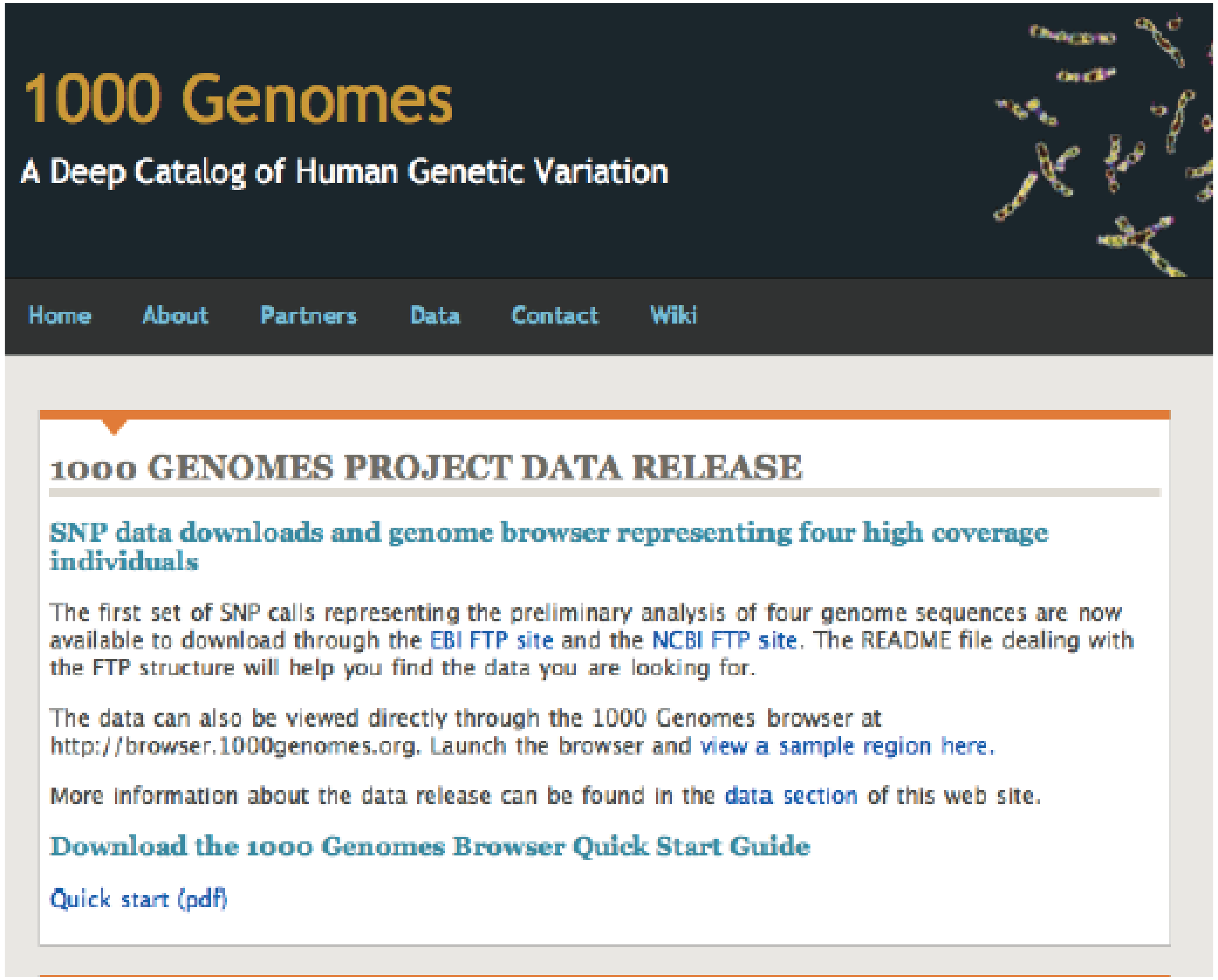
RNA-seq differential expression



Genome-wide Chromatin Immunoprecipitation Identifies Sites on the Genome Occupied by Transcription Regulators – ChIP-seq



1000 Genomes Project



1000 Genomes

A Deep Catalog of Human Genetic Variation

Home About Partners Data Contact Wiki

1000 GENOMES PROJECT DATA RELEASE

SNP data downloads and genome browser representing four high coverage individuals

The first set of SNP calls representing the preliminary analysis of four genome sequences are now available to download through the EBI FTP site and the NCBI FTP site. The README file dealing with the FTP structure will help you find the data you are looking for.

The data can also be viewed directly through the 1000 Genomes browser at http://browser.1000genomes.org. Launch the browser and view a sample region here.

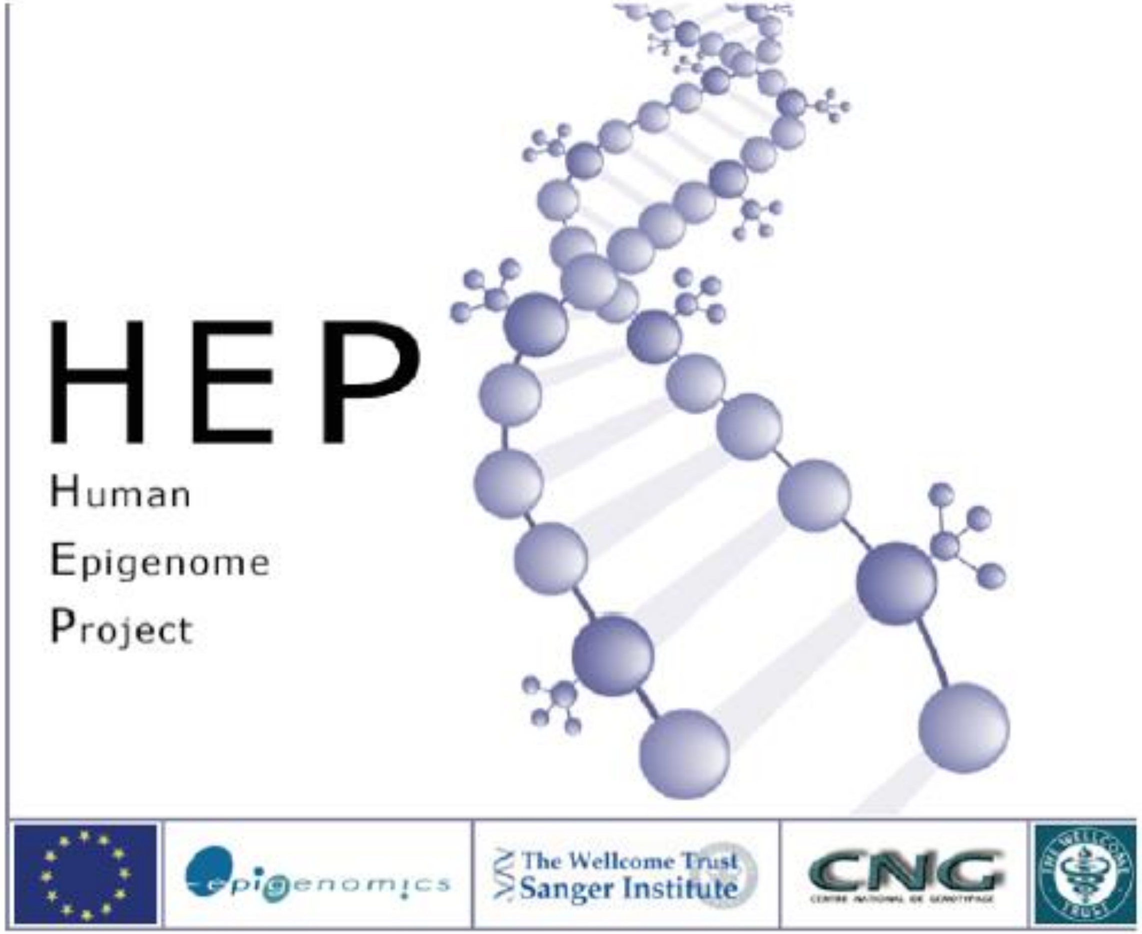
More information about the data release can be found in the data section of this web site.

Download the 1000 Genomes Browser Quick Start Guide

Quick start (pdf)

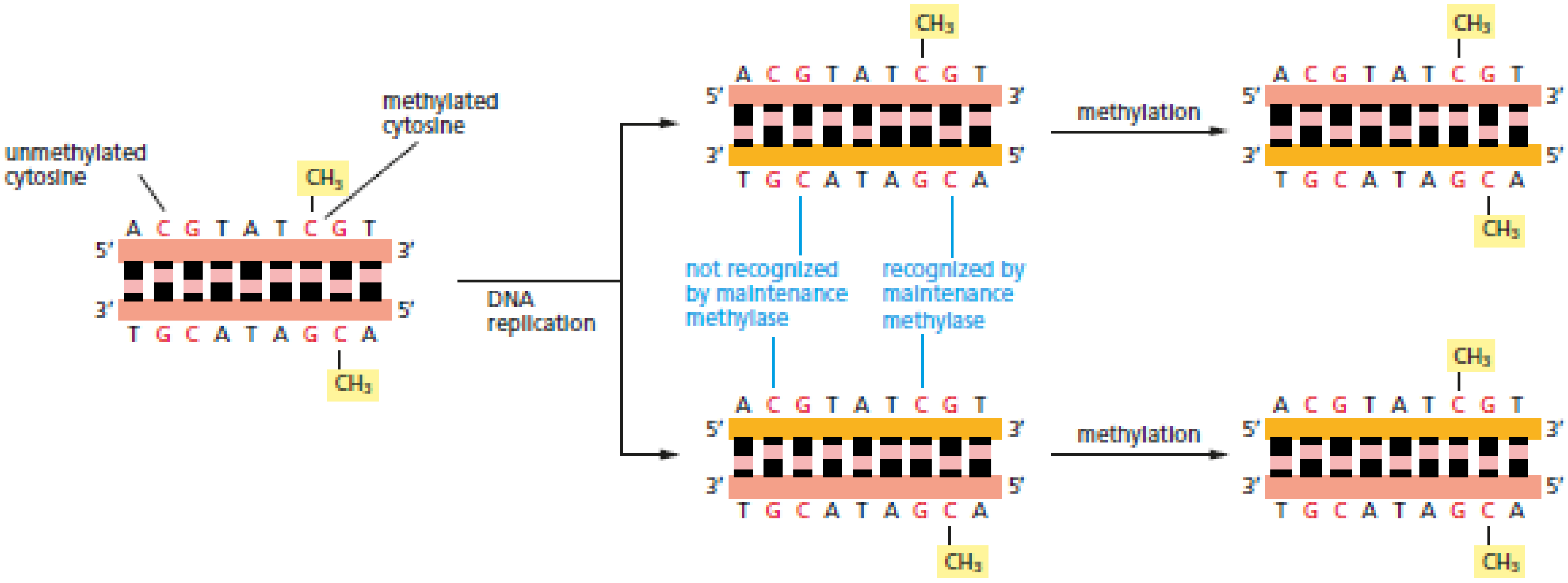
Genotyping

Human Epigenome Project



Methylation

Patterns of DNA Methylation Can Be Inherited When Cells Divide



DNA methylation in DNA occurs on cytosine (C) nucleotides largely in the sequence CG

DNA methylation mostly helps to repress transcription

High frequency of **CG** sites termed **CG islands** (CGIs)

Genomic Imprinting Is Based on DNA Methylation