

Emergence of Life

Lecture 1&2

Acknowledgements:

Alberts - Molecular Biology of the Cell

Scitable by Nature Education

Nature Resources

Internet resources

Outline of the Lecture

Objective

- ❖ Retrace the assembly of matter to form life

Topics

- ❖ Introduction and background
 - ❖ Some scientific/philosophic questions
- ❖ Concept of order
 - ❖ Physical, Chemical and Biological Parameters
- ❖ Information encoding in biomolecules
- ❖ The “creation” of the cell

Was life supposed to evolve?

- What we normally think of as life is based on chains of carbon atoms with a few other primary atoms such as oxygen, hydrogen, nitrogen and phosphorous.
- What about silicon-based life?
 - Yes, maybe! but carbon-based seems the most favorable because it has the richest chemistry

The Anthropic Principle:

For a given universe, it is possible that the values of the physical constants, will allow the existence of objects like carbon atoms that can act as the building blocks of living systems



<https://www.youtube.com/watch?v=KtRcAuunEMg>

Why did life occur?

- ❖ We don't have an answer! Its like asking – Does God Exist?
- ❖ The *anthropic principle* tells us that LIFE would have happened if not on planet earth, then on some other planet
- ❖ We do know, by observation, what are the properties that define life
 - ❖ Have a set of instructions that tell the system how to sustain and reproduce itself
 - ❖ A mechanism to carry out the instructions
- ❖ By the second law of thermodynamics, to do this, we need to have order

What is the role of “Order” in the creation of the first cell?

■ How do you determine if a system has order - in the sense of structured-ness and organization ?

- Self Organization
- Complexity
- Emergence of new properties

- One of the most fundamental problems in biology concerns the origin of forms and their associated functions
- It has been an important question of developmental biology

Third law of Thermodynamics states that the entropy (disorder) of the universe is continuously increasing with time.

Therefore, decrease of entropy, as in a cell, is permitted if the corresponding increase in entropy in the environment is greater.

The history of life

- ❖ And it took 2.5 billion years for life to evolve from the earliest cells to multi cell animals
 - ❖ Another 1 billion years to evolve through fish and reptiles to mammals
 - ❖ Then evolution seemed to have speeded up
 - ❖ took about a hundred million years to develop from the early mammals to us

We will examine how “Order” occurs

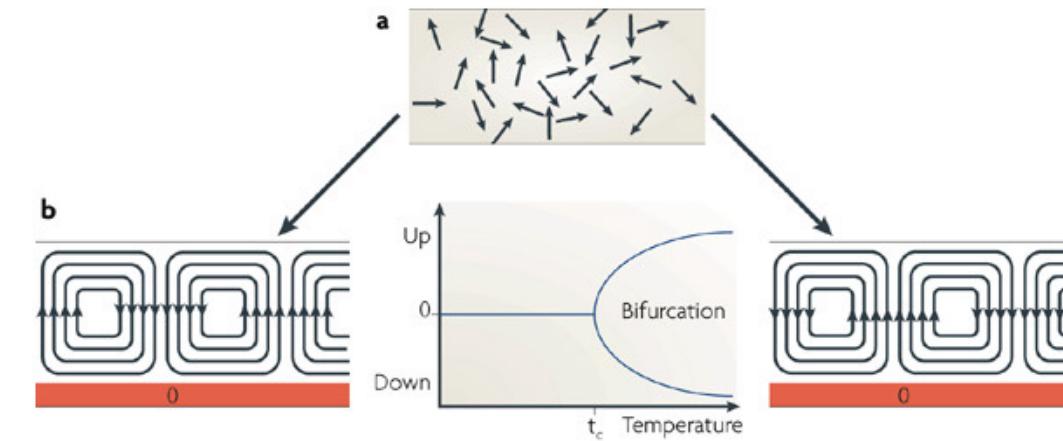
In-animate Systems

- ❖ Physical systems
- ❖ Chemical systems
- ❖ Biological self-assembled systems

Living Systems

- ❖ The cell

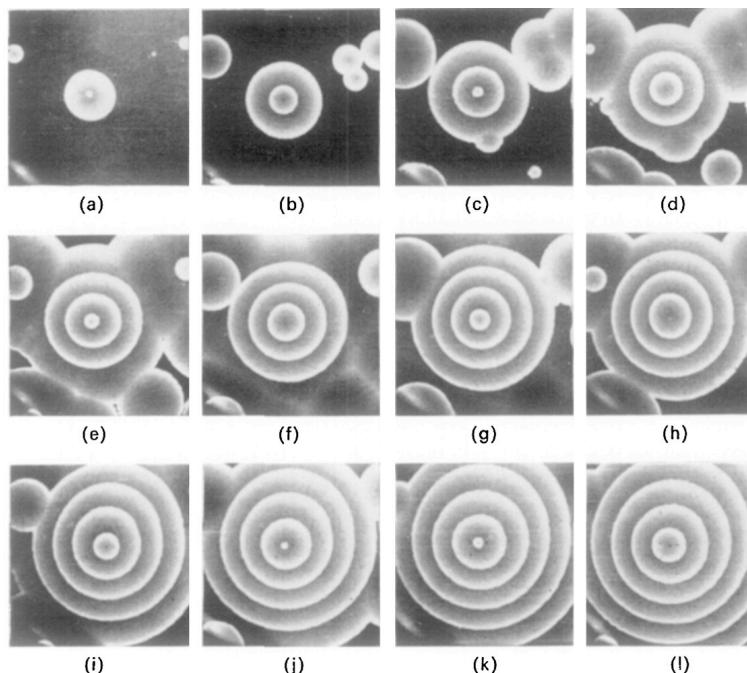
Order in inanimate systems: Physical Systems



Nature Reviews | Molecular Cell Biology

- a) In a liquid layer, molecules are agitated by thermal motion.
- b) The molecules in the liquid layer are heated from below (red zone) and self-organize into rolls (drawn in cross-section) when the temperature reaches a critical value (t_c). At this value, the molecules start to move collectively either up or down at point 0, which determines the alternative orientation of the rotation of the rolls throughout the layer. The orientation of the rotation choice is unpredictable and determined by local fluctuations at t_c

Order (Self Organization) in Chemical Reactions

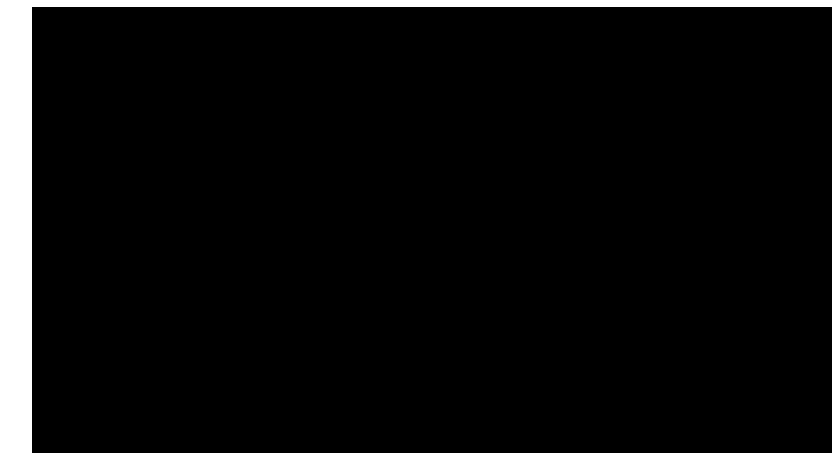


Photographs taken at 30 s intervals

In the system, propagation of a single wave was observed under following conditions:

$\text{NaBrO}_3 = 0.23 \text{ M}$, bromomalonic acid = 0.16 M, (Ferroin)
 $\text{Fe}(\text{phen})_3 = 0.003 \text{ M}$, $\text{H}_2\text{SO}_4 = 0.26 \text{ M}$, $T = 14^\circ\text{C}$.

The wave velocity equalled approximately 0.01 cm/s.



In class demonstration of the B-Z reaction

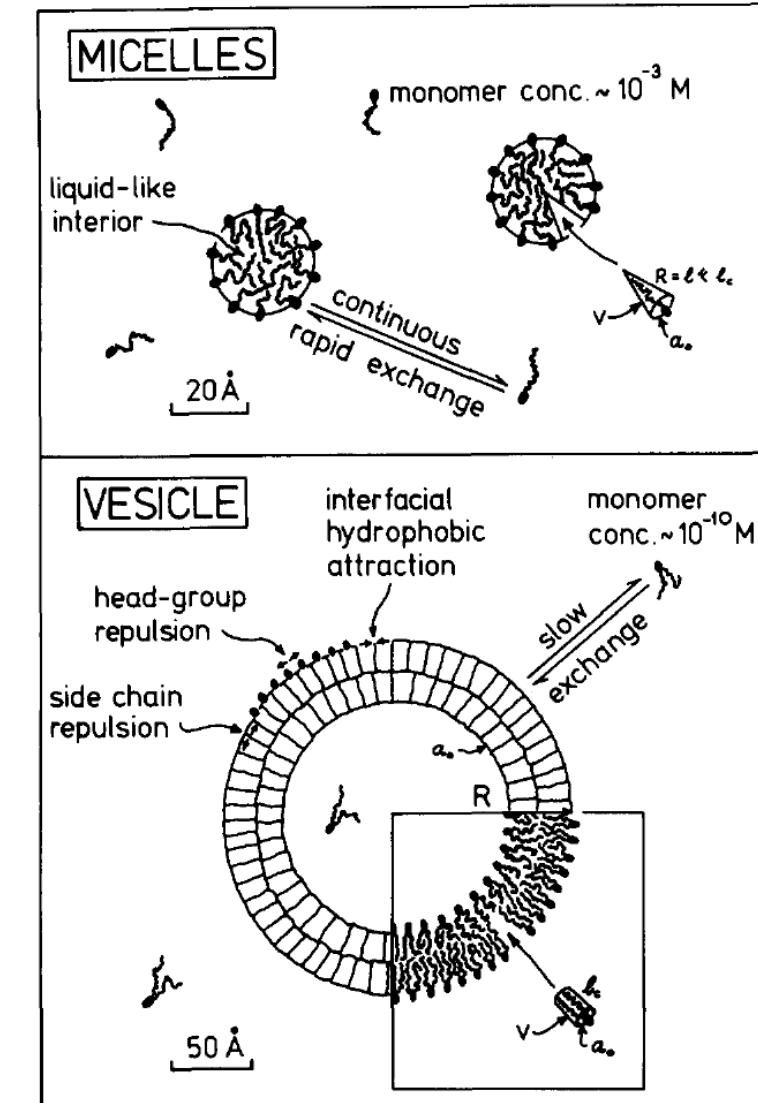
<https://www.youtube.com/watch?v=o72GGxQqWt8>

Belousov–Zhabotinsky reaction J. Theor. Biol.
(1973) 40, 45-61

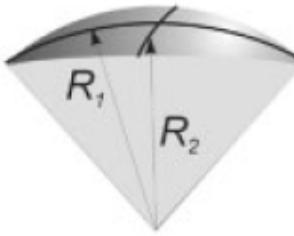
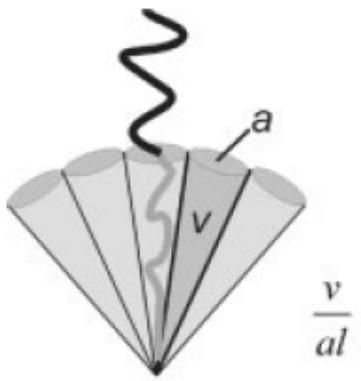
Self-assembled systems that occur in Biological world

The boundary: inanimate to animate

- ❖ The concepts of interaction free energies, molecular geometry and entropy, when taken together, furnish a framework for a theory of self-assembly or self-organization
- ❖ Micelles and bilayer vesicles assembly are driven by hydrophobic and hydrophilic interaction - of two 'opposing forces'
- ❖ Size of the micelle is determined by "optimal surface area" per head group at which the total interaction free energy per lipid molecule is a minimum.



Some basic rules for micelle shapes



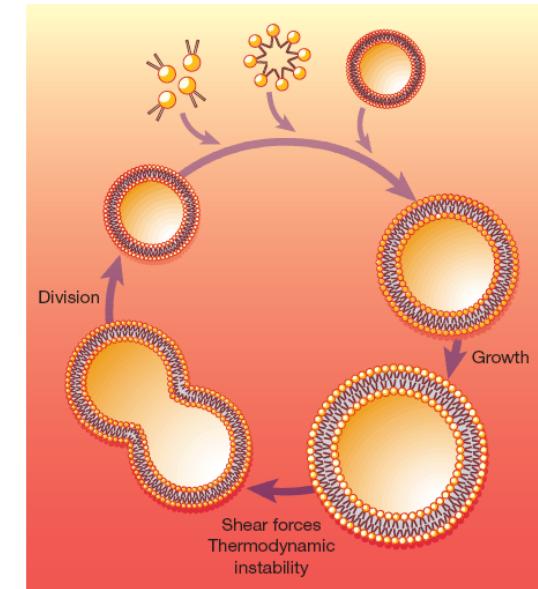
$$H = \frac{1}{2} \left(\frac{1}{R_1} + \frac{1}{R_2} \right)$$
$$K = \frac{1}{R_1 R_2}$$

Shape	$v/(al)$	H	K
Sphere	1/3	1/R	1/R ²
Cylinder	1/2	1/(2R)	0
Bilayer	1	0	0

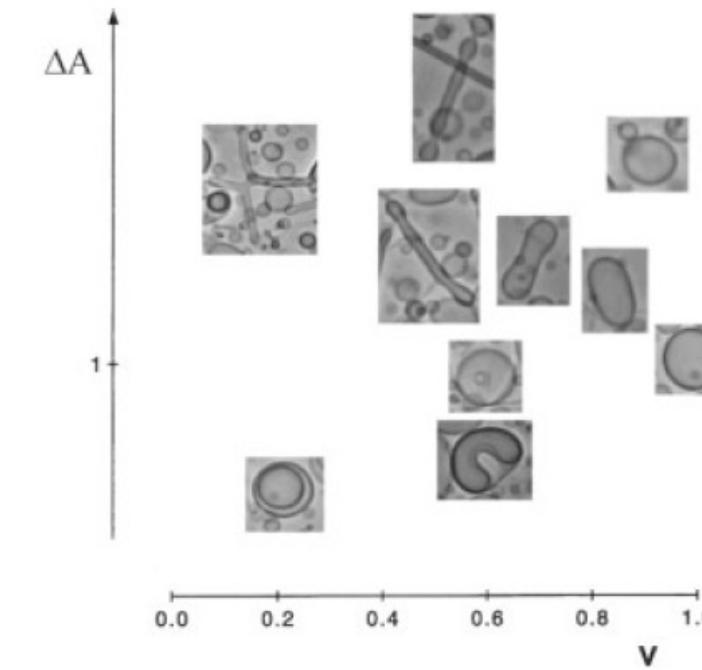
Adv Mater 2003 15 (16)

- Surfactant first forms a bilayer and then closes to form a vesicle
- The ratio of the hydrophobic to hydrophilic portion of the molecule determines the radius of curvature at the interface

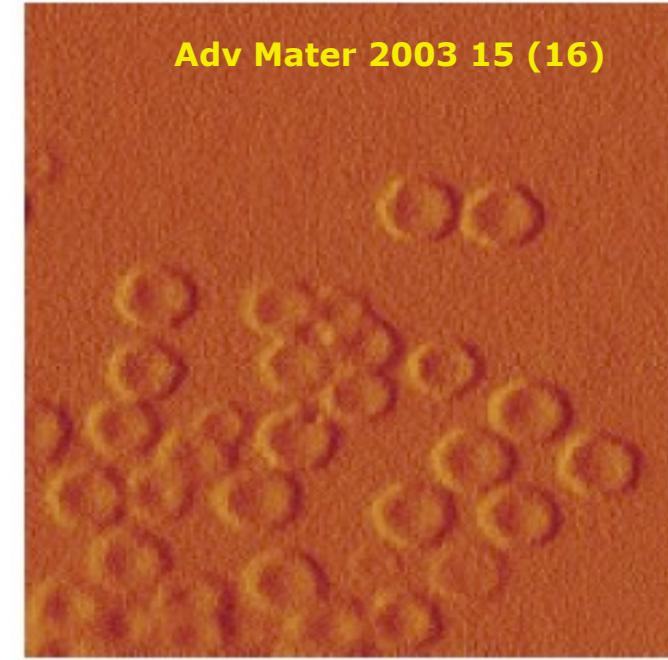
$$\frac{v}{al} = 1 + Hl + \frac{Kl^2}{3}$$



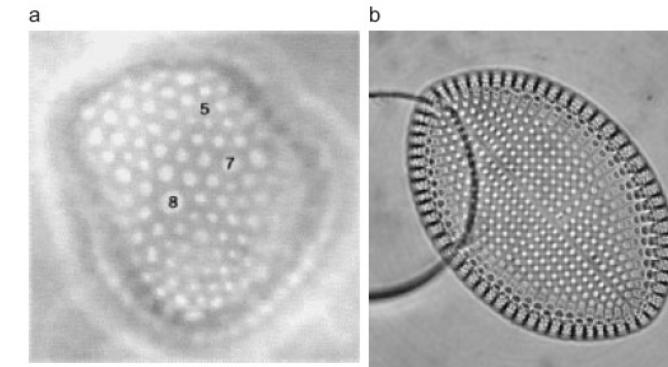
Vesicle Size and Shape



- ❖ Depends on mixing entropies (pull towards many assemblies) and molar bending energies (tends towards a smaller number of vesicles)
- ❖ ΔA is the difference in area, v is the dimensionless volume to area ratio

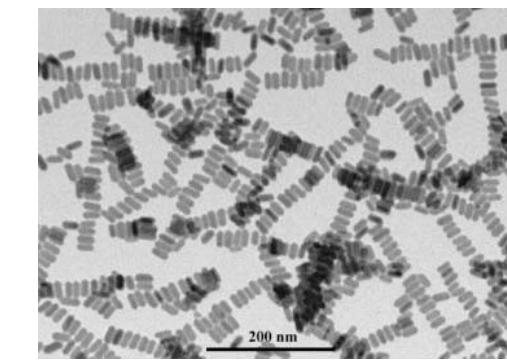


Similarity in structure of self-organized biological and non-biological entities

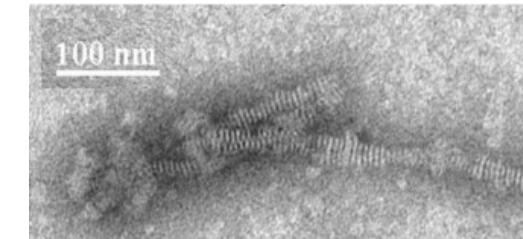


(a) High-genus block copolymer vesicle (b) structure of a diatom
[Adv Mater 2003 15 (16)]

❖ Genus order $g > 100$

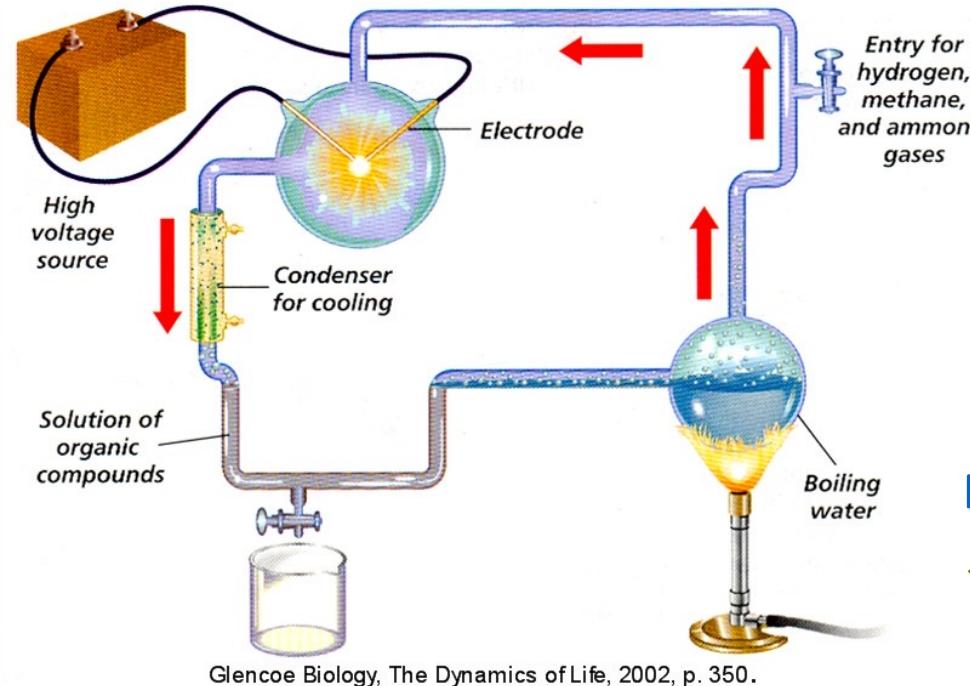


TEM image showing ordered chains of prismatic BaCrO₄ nanoparticles prepared in AOT microemulsions
Angew. Chem. Int. Ed. 2003, 42, 2350 – 2365



DMPC-apoE CT domain complexes; reconstituted lipoprotein particles were stained with 2% phosphotungstate for visualization [Biochem. J. (2005) 387 (747–754)]

Urey Miller Experiment



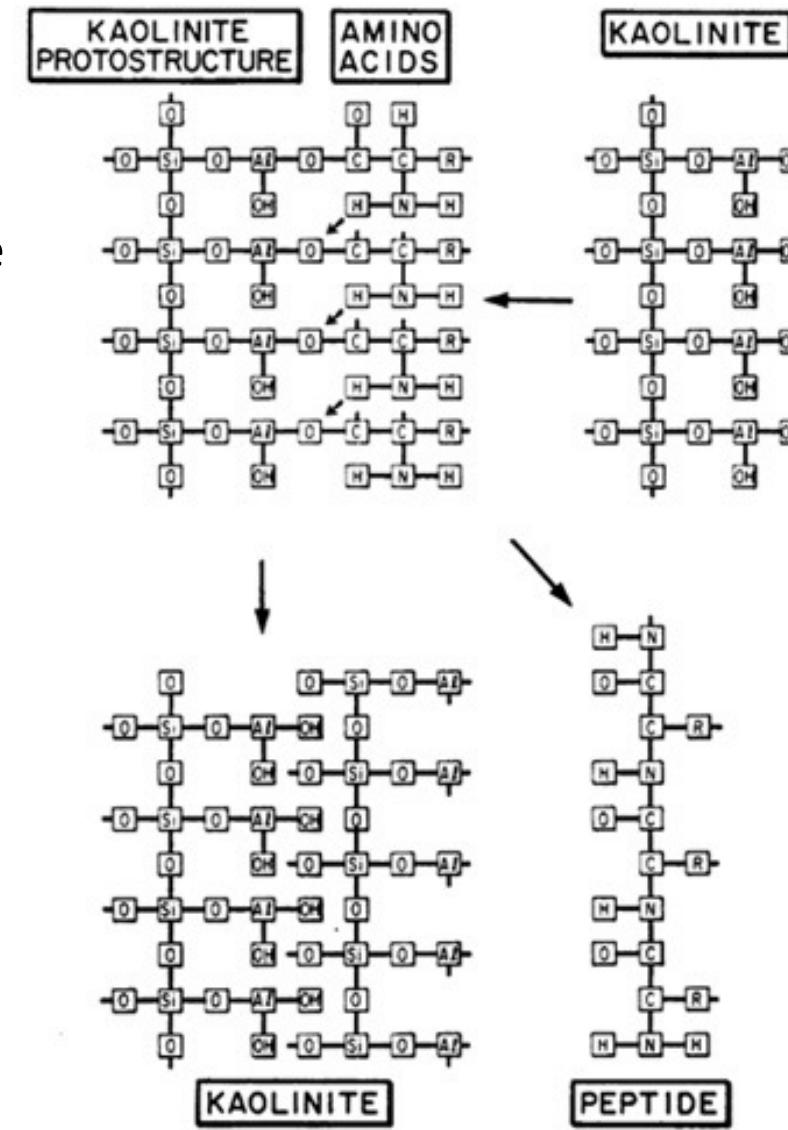
❖ In 1953 Stanley Miller and Harold Urey performed the first experiment that produced amino acids in what was assumed to be a pre-life atmosphere. They passed a mixture of water vapor, methane, hydrogen and ammonia gases through an electric arc to simulate what would happen if these gases were subjected to lightning.

Result:

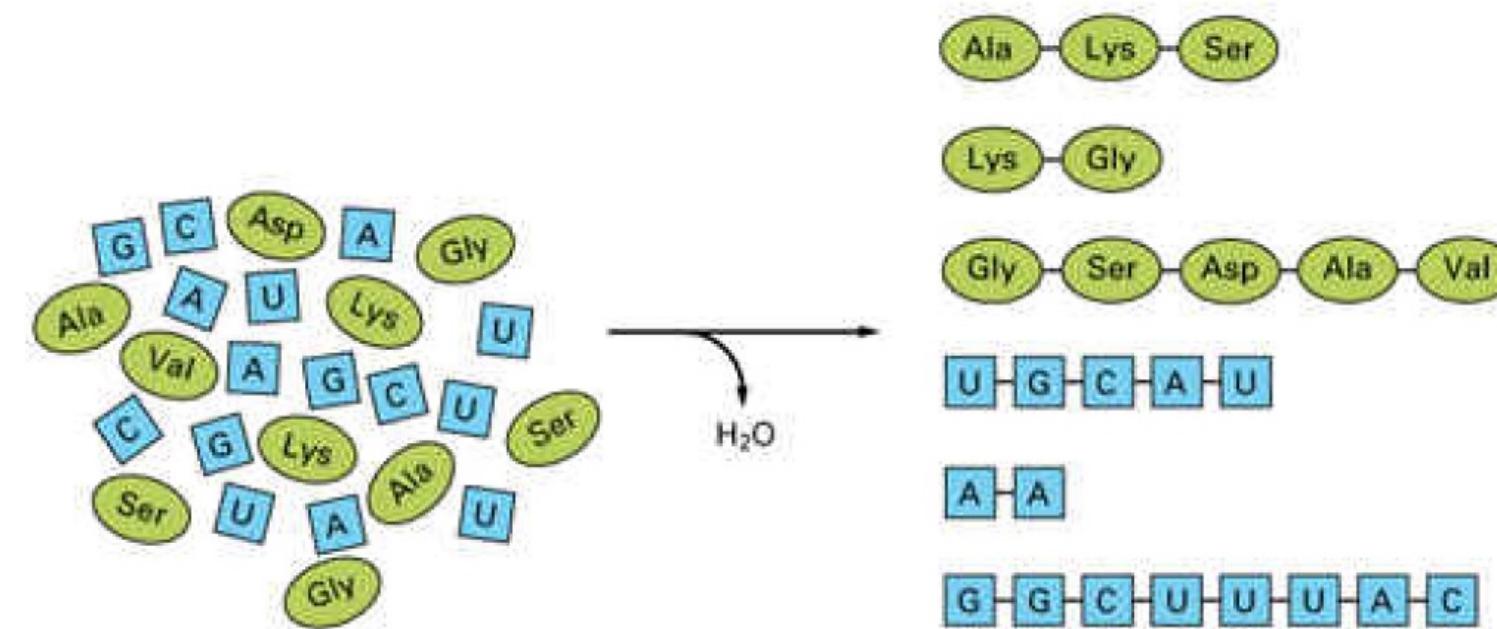
❖ **10 biologic amino acid types**
25 non-biologic amino acid types
Formaldehyde
Sugars

First proteins

- ❖ In the presence of kaolinite, amino acids are picked up from an aqueous solvent and brought into solid solution
 - ❖ Amino groups become hydrogen bonded to structural oxygen
 - ❖ In water, amino acids cannot polymerize because of dipole-dipole interactions
 - ❖ In solid solution, however, amino acids will polymerize, because the solvent medium does not interfere
 - ❖ about 1000 times more amino acids were polymerized to peptides
- ❖ Kaolinite is also instrumental in preferentially synthesizing pentoses and hexoses from formaldehyde and transforming them into polysaccharides
- ❖ Preferential polymerization of L-amino acids on kaolinite can be attributed to the inherent enantiomorphism of the edges of the octahedral layer of kaolinite

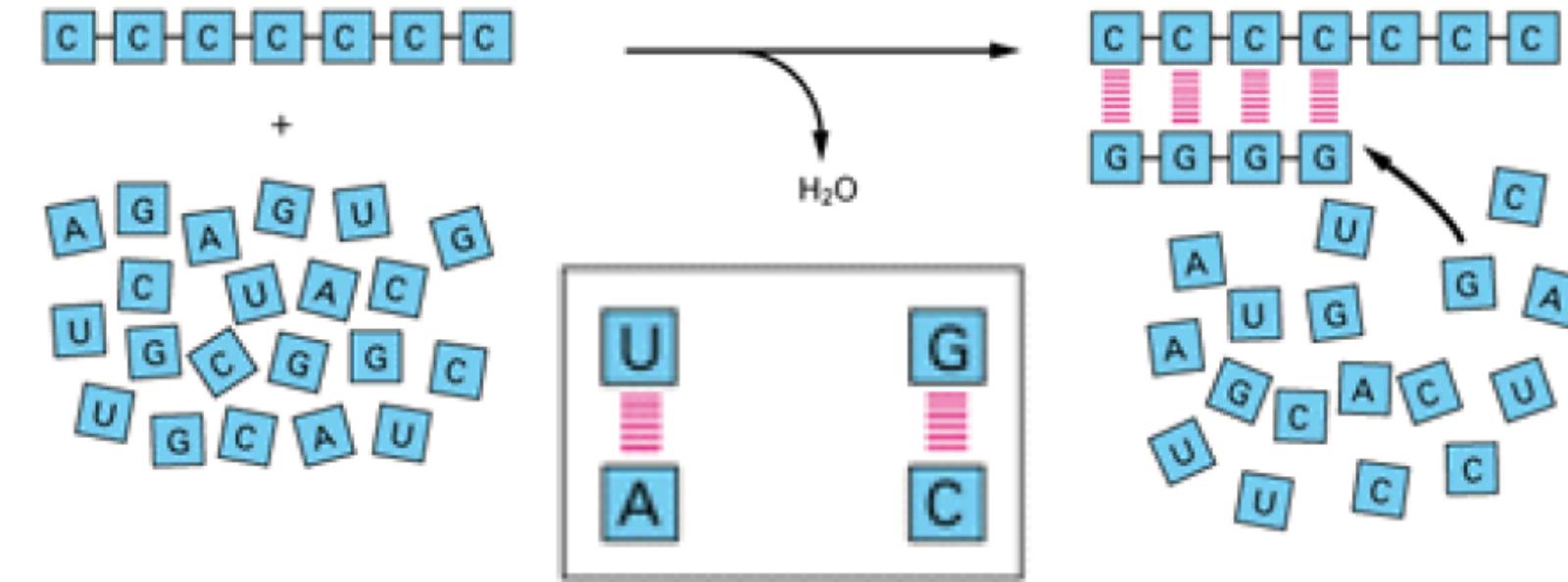


Formation of polynucleotides and polypeptides



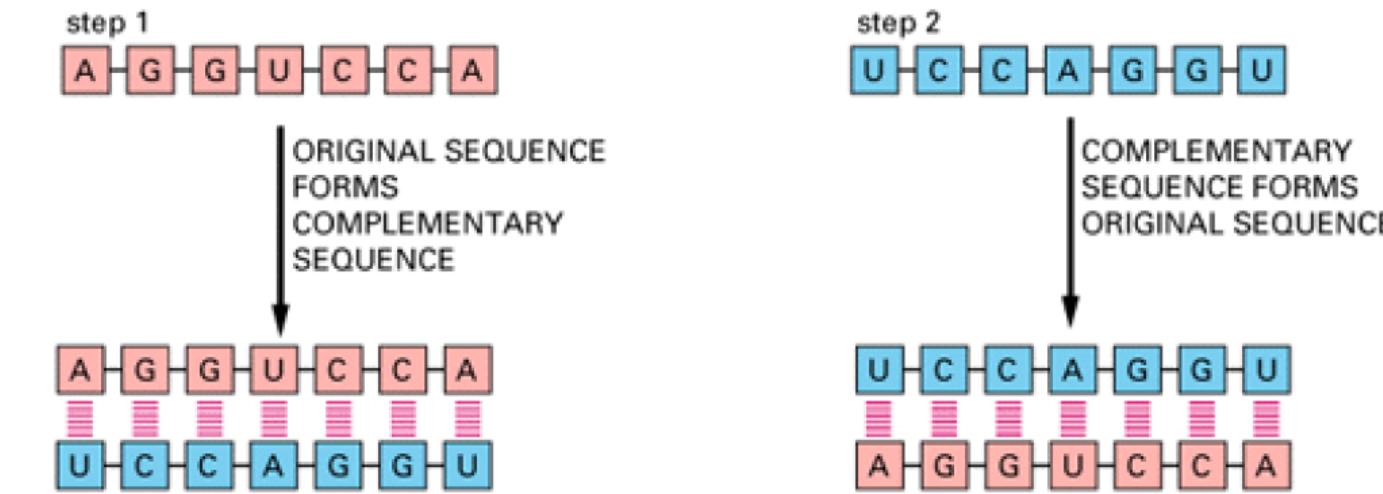
- ❖ Nucleotides of four kinds (here represented by the single letters A, U, G, and C) can undergo spontaneous polymerization with the loss of water. The product is a mixture of polynucleotides that are random in length and sequence.
- ❖ Similarly, amino acids of different types, symbolized here by three-letter abbreviated names, can polymerize with one another to form polypeptides. Present-day proteins are built from a standard set of 20 types of amino acids.

Polynucleotides as templates



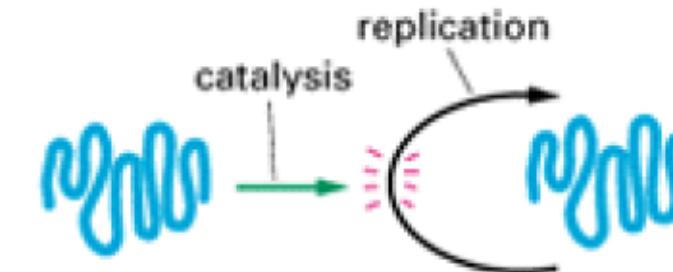
- ❖ Preferential binding occurs between pairs of nucleotides (G with C and U with A) by relatively weak chemical bonds

Replication of a polynucleotide sequence

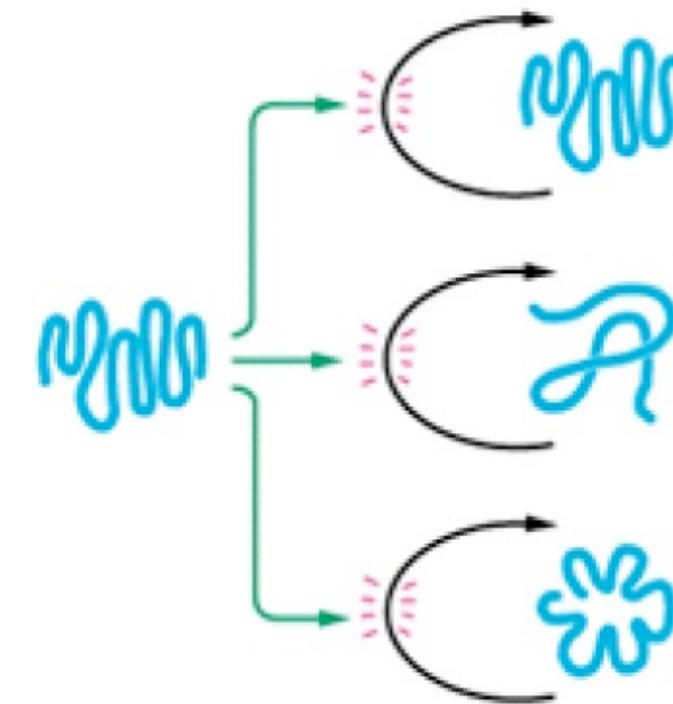


- ❖ The original RNA molecule acts as a template to form an RNA molecule of complementary sequence.
- ❖ This complementary RNA molecule itself acts as a template, forming RNA molecules of the original sequence. Since each templating molecule can produce many copies of the complementary strand, these reactions can result in the "multiplication" of the original sequence

Conformation of an RNA molecule

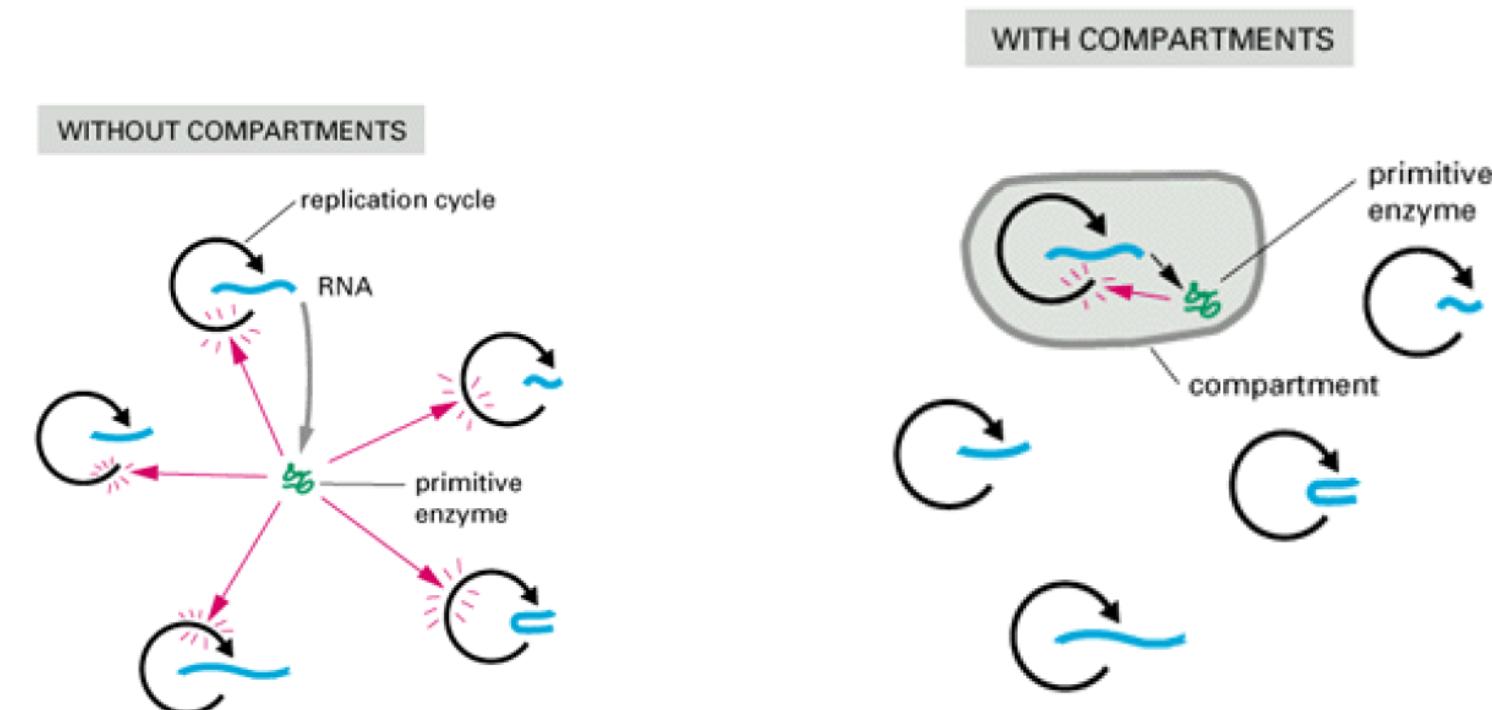


catalytic RNA molecule that joins together nucleotides to reproduce its own nucleotide sequence and therefore its shape



family of mutually supportive catalytic RNA molecules, one catalyzing the reproduction of the others

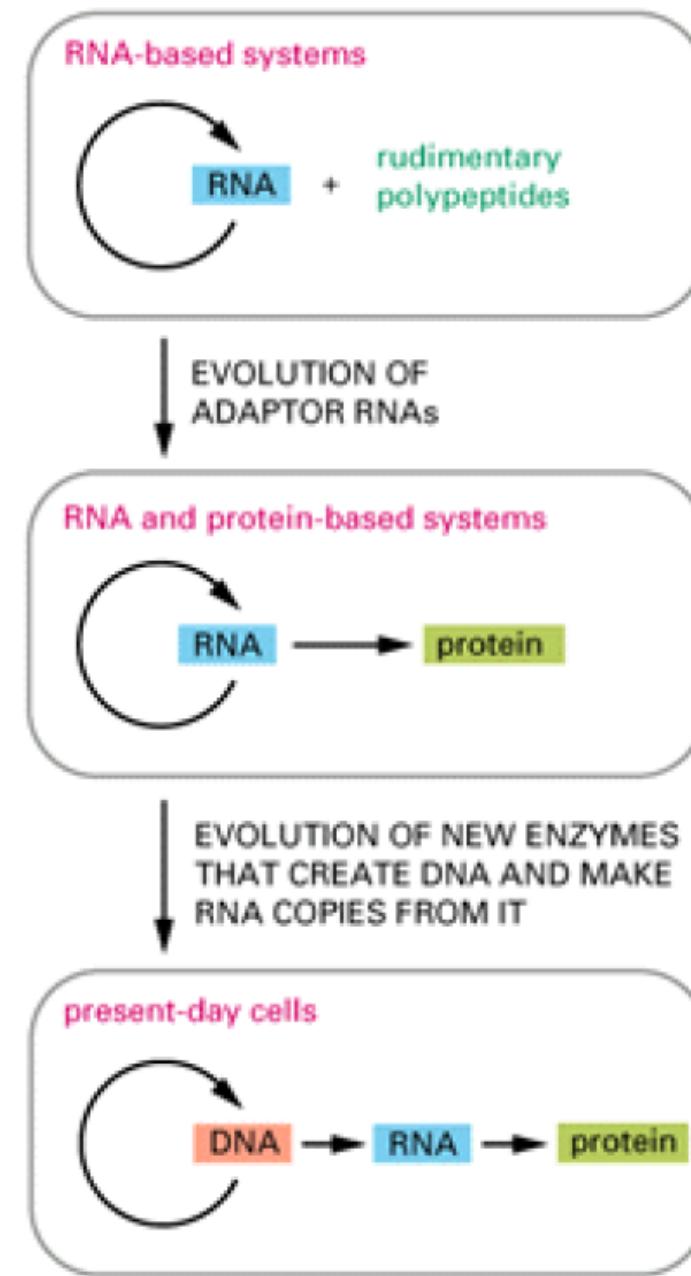
Evolutionary significance of cell-like compartments



- ❖ Any improved form of RNA that is able to promote formation of a more useful protein must share this protein with its neighboring competitors.
- ❖ If the RNA is enclosed within a compartment, such as a lipid membrane, then any protein the RNA causes to be made is retained for its own use; the RNA can therefore be selected on the basis of its guiding production of a better protein.

Central Dogma

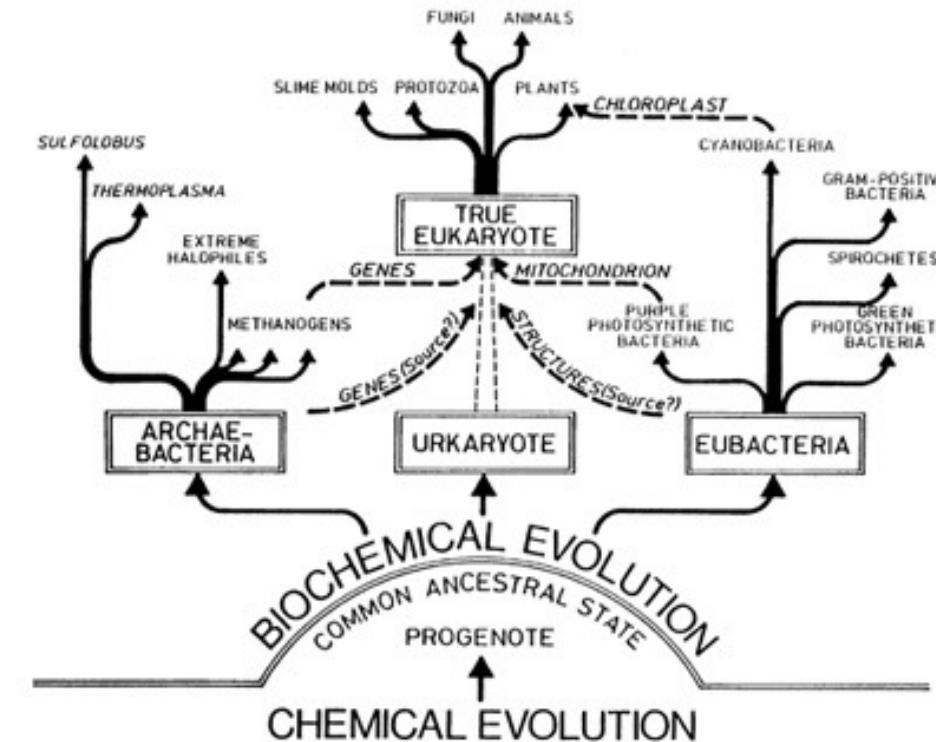
Suggested stages of evolution from simple self-replicating systems of RNA molecules to present-day cells. Today, DNA is the repository of genetic information and RNA acts largely as a go-between to direct protein synthesis



Last Universal Common Ancestor

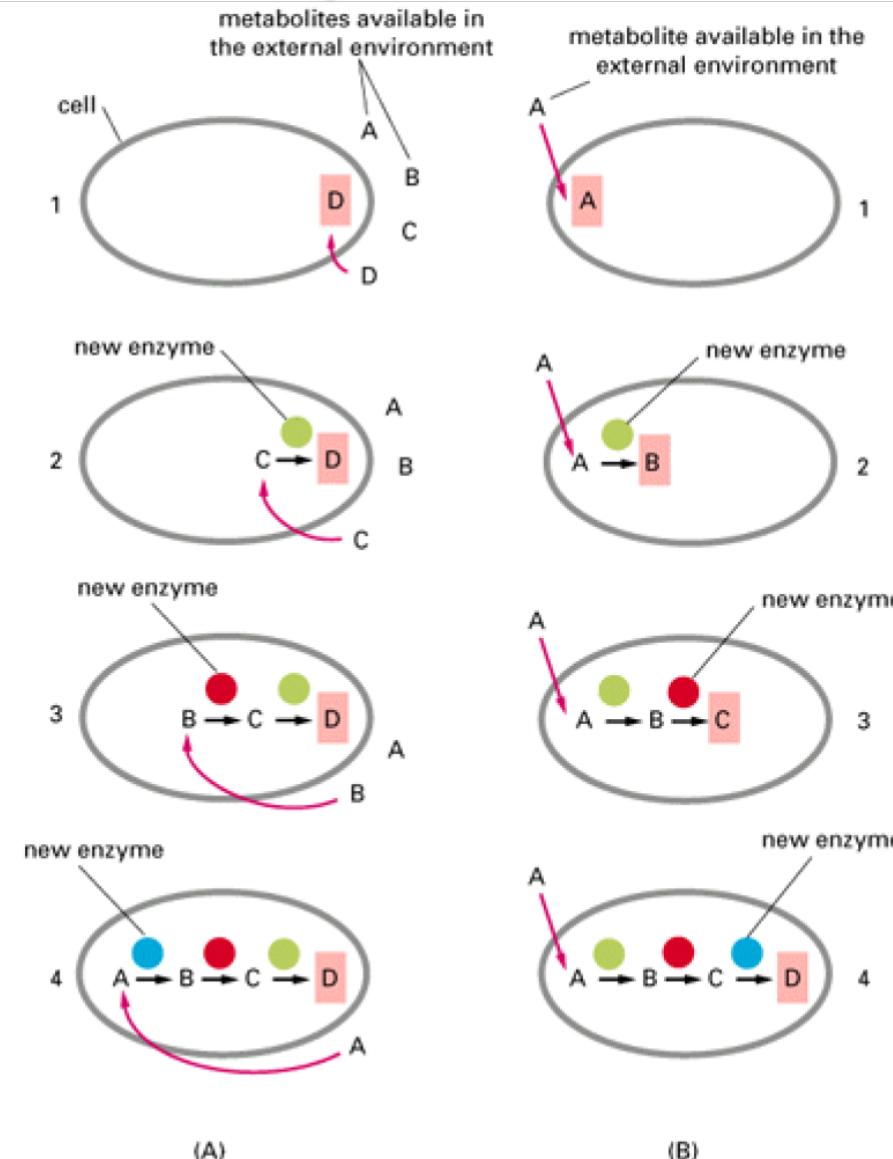
- ❖ The evolution of the translation apparatus occurred in a series of increasingly complex stages, rather than all at once,
- ❖ The stages subsequent to the establishment of the basic mechanism were concerned by and large with increasing the mechanism's accuracy, and possibly speed as well

J. Mol Evol 10, 1-6, 1977

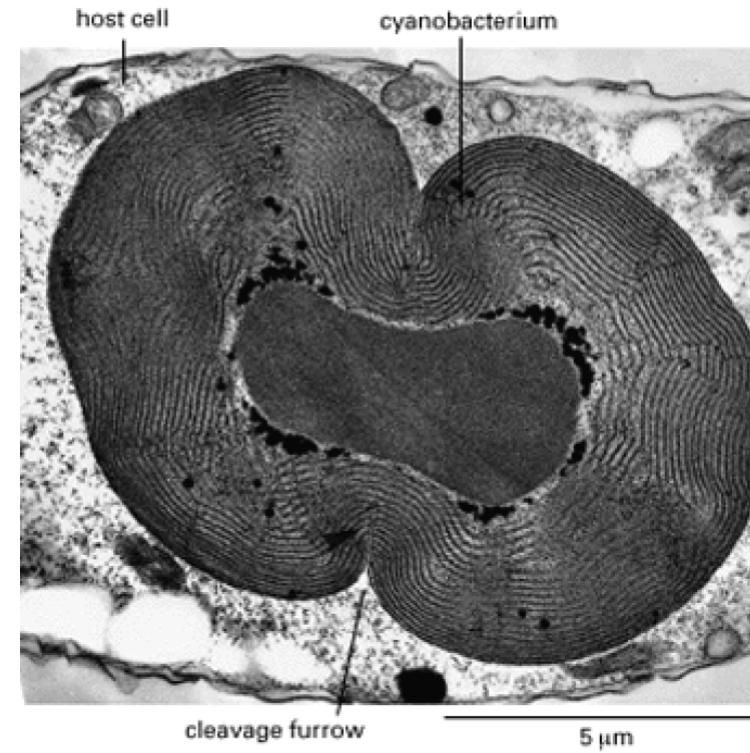


Evolution of Metabolic Pathways

- The cell on the left is provided with a supply of related substances (A, B, C, and D) produced by prebiotic synthesis. One of these, substance D, is metabolically useful. As the cell exhausts the available supply of D, a selective advantage is obtained by the evolution of a new enzyme that is able to produce D from the closely related substance C. Fundamentally important metabolic pathways may have evolved by a series of similar steps.
- On the right, a metabolically useful compound A is available in abundance. An enzyme appears in the course of evolution that, by chance, has the ability to convert substance A to substance B. Other changes then occur within the cell that enable it to make use of the new substance. The appearance of further enzymes can build up a long chain of reactions.

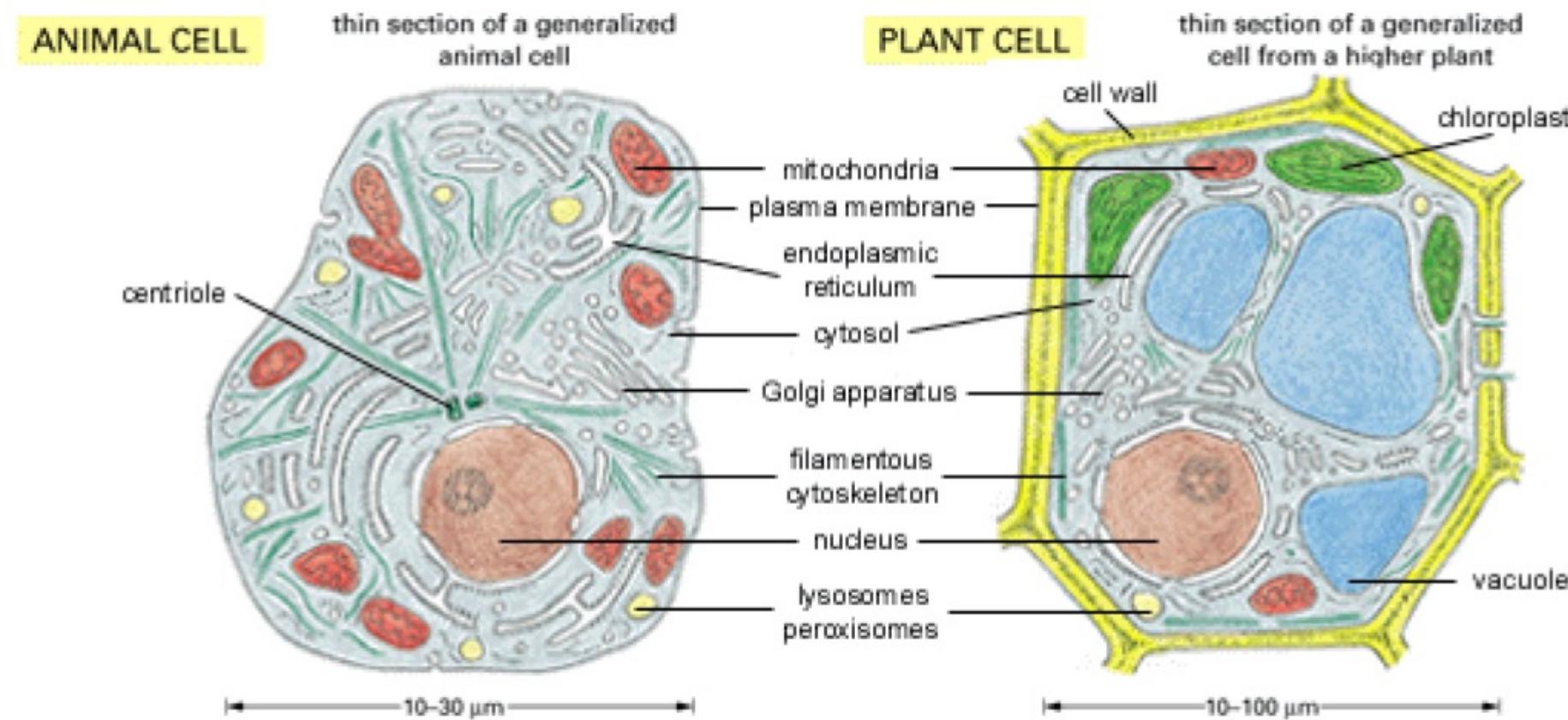


Predation to new cellular structures



- ❖ A close relative of present-day cyanobacteria that lives in a permanent symbiotic relationship inside another cell. The two organisms are known jointly as *Cyanophora paradoxa*
- ❖ The "cyano-bacterium" is in the process of dividing

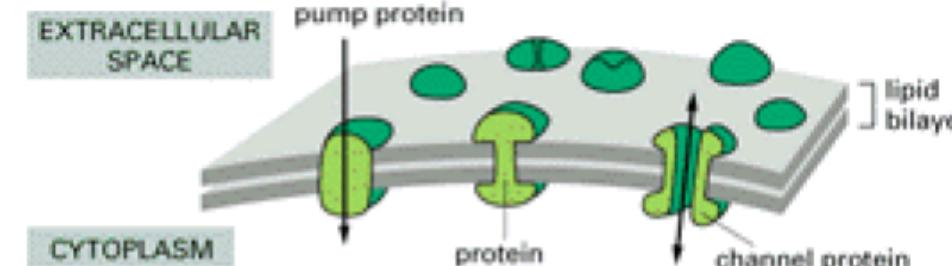
Animal and Plant Cell structures



Cellular Membrane System

PLASMA MEMBRANE

The outer boundary of the cell is the plasma membrane, a continuous sheet of phospholipid molecules about 4–5 nm thick in which various proteins are embedded.

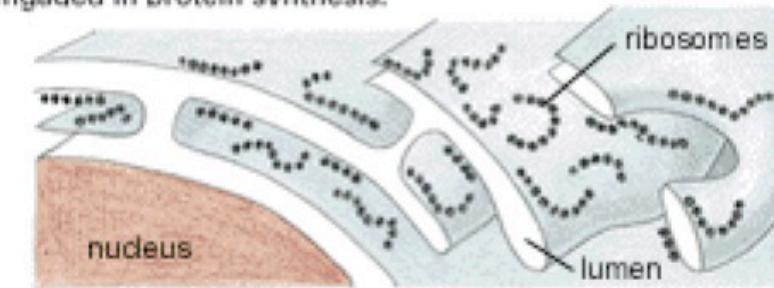


Some of these proteins serve as pumps and channels for transporting specific molecules into and out of the cell.

ENDOPLASMIC RETICULUM

Flattened sheets, sacs, and tubes of membrane extend throughout the cytoplasm of eucaryotic cells, enclosing a large intracellular space. The ER membrane is in structural continuity with the outer membrane of the nuclear envelope, and it specializes in the synthesis and transport of lipids and membrane proteins.

The **rough endoplasmic reticulum (rough ER)** generally occurs as flattened sheets and is studded on its outer face with ribosomes engaged in protein synthesis.



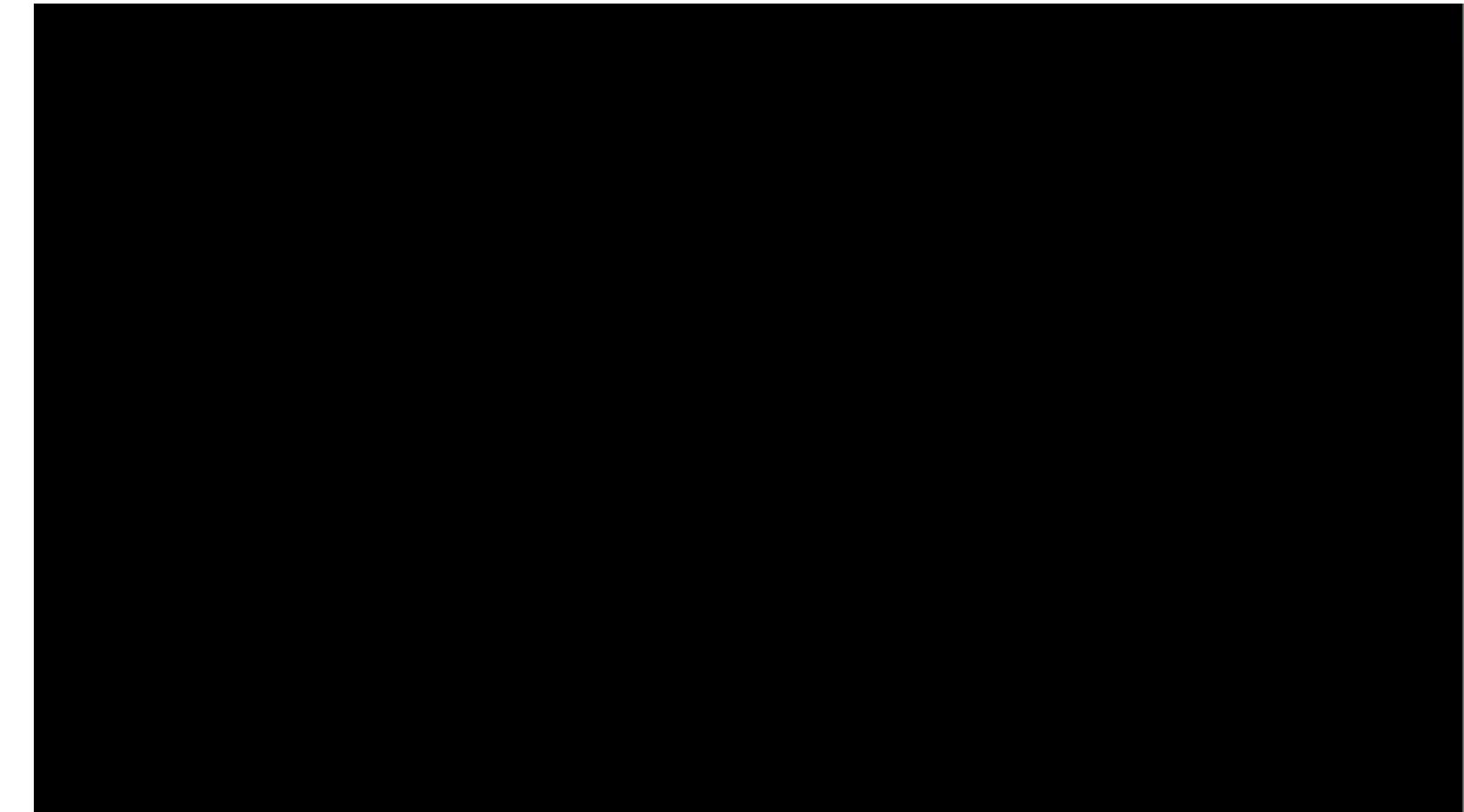
The **smooth endoplasmic reticulum (smooth ER)** is generally more tubular and lacks attached ribosomes. It has a major function in lipid metabolism.



Summary

- ❖ Autocatalytic mechanisms fundamental to living systems began with the evolution of families of RNA molecules that could catalyze their own replication.
- ❖ Families of cooperating RNA catalysts developed the ability to direct synthesis of polypeptides.
- ❖ Accumulation of additional protein catalysts allowed more efficient and complex cells to evolve, the DNA double helix replaced RNA as a more stable molecule for storing the increased amounts of genetic information required by such cells
- ❖ Present-day living cells are classified as prokaryotic (bacteria and their close relatives) or eukaryotic.

Inner Life of a Cell

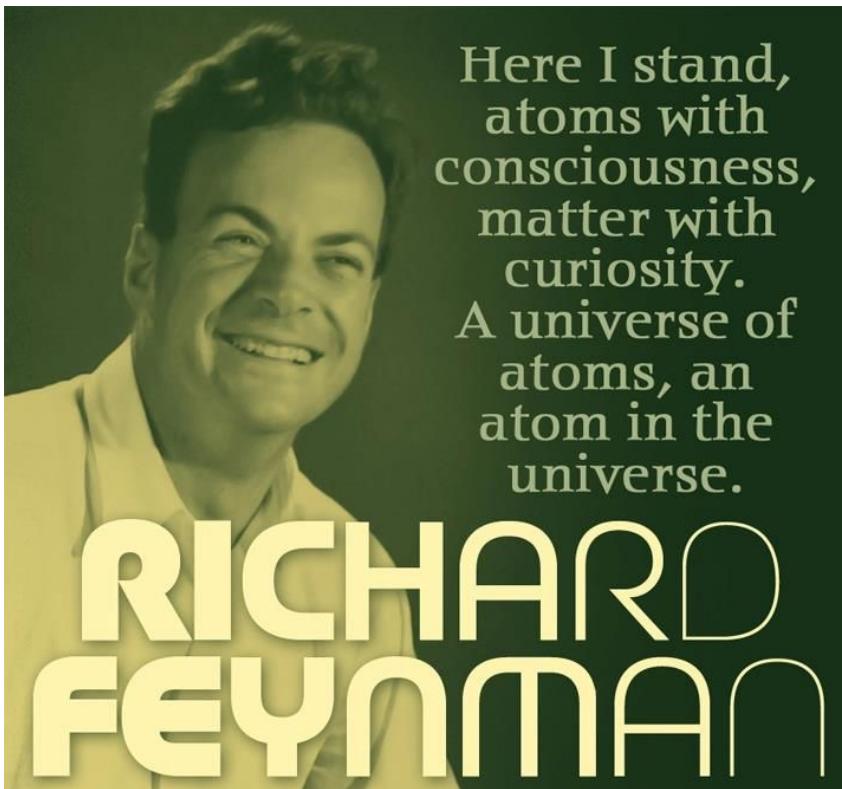


Full version

<https://www.youtube.com/watch?v=FzcTgrxMzZk>

Fundamental Units of Life

Lecture 3



Here I stand,
atoms with
consciousness,
matter with
curiosity.
A universe of
atoms, an
atom in the
universe.

**RICHARD
FEYNMAN**

**SURELY YOU MUST BE
JOKING MR FEYNMAN**

“Life is an emergent, rather than an inherent, property of matter. Although it arises from the material world, it cannot be reduced to it”

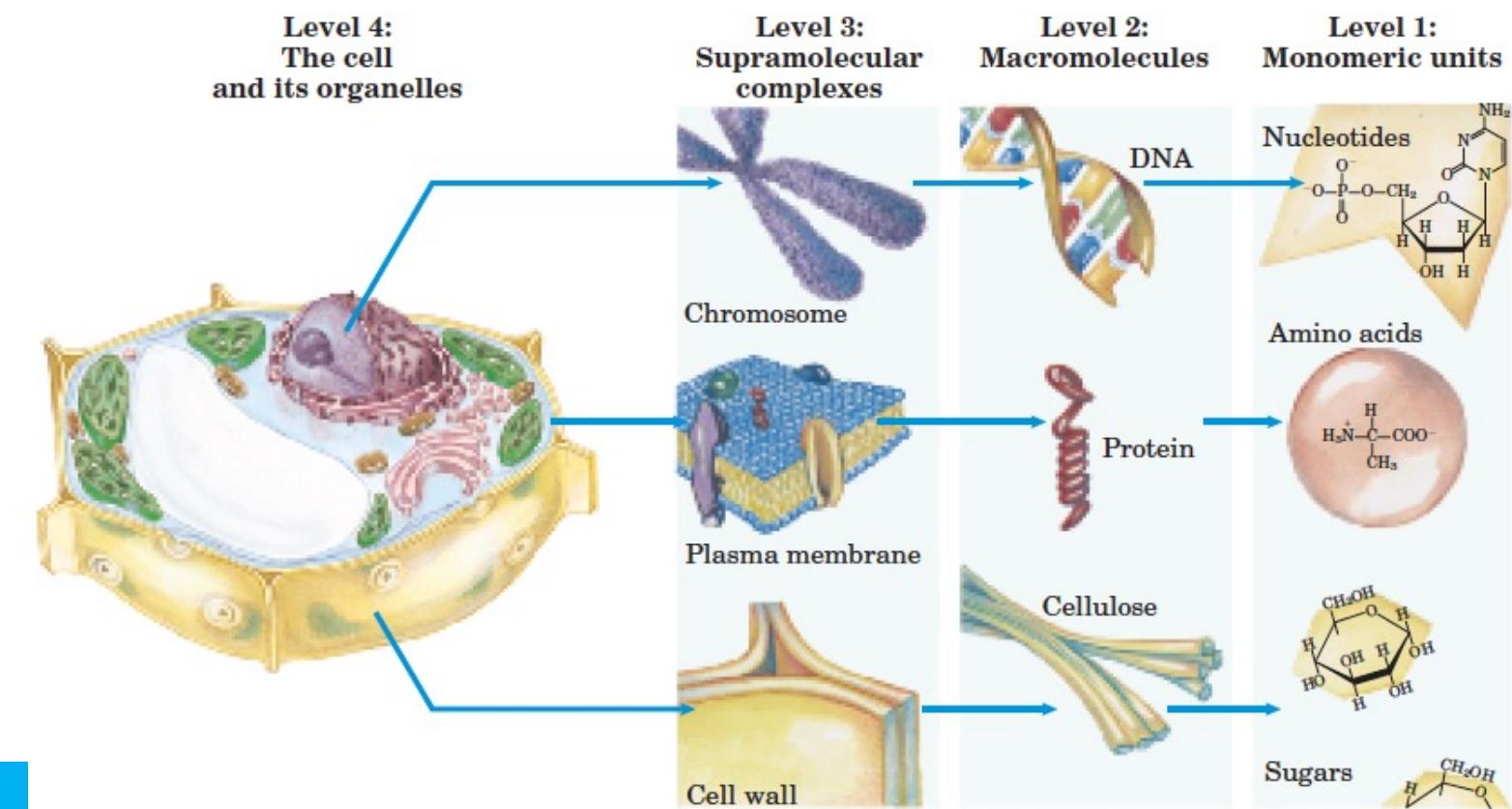
All organisms are like a machine; while a machine implies a machine maker, an organism is a self-organizing entity

Outline of the lecture

Objective: to learn about the main biomolecules that are part of the building blocks of cells

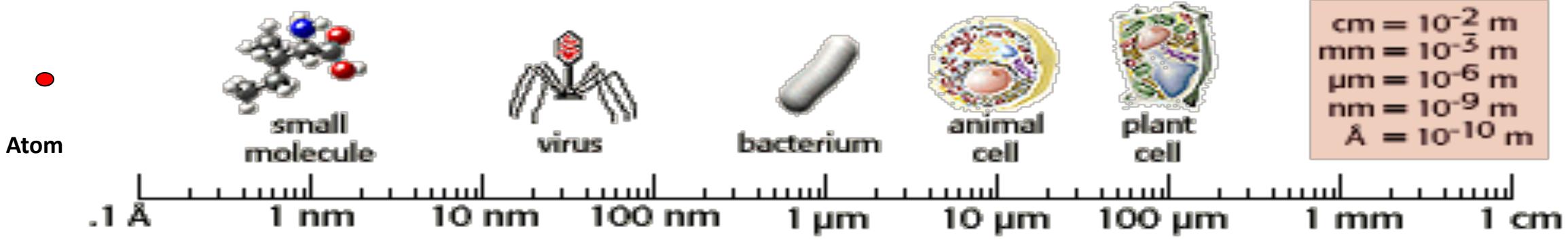
- ❖ Carbohydrates and Lipids – cell structure
- ❖ Amino acids – proteins
- ❖ Nucleotides – DNA and RNA

Look at some applications



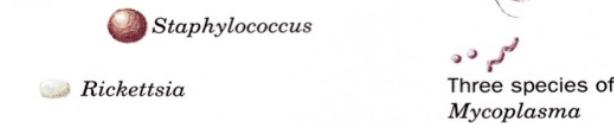
Sizes of Things

Relative sizes of cells and their components



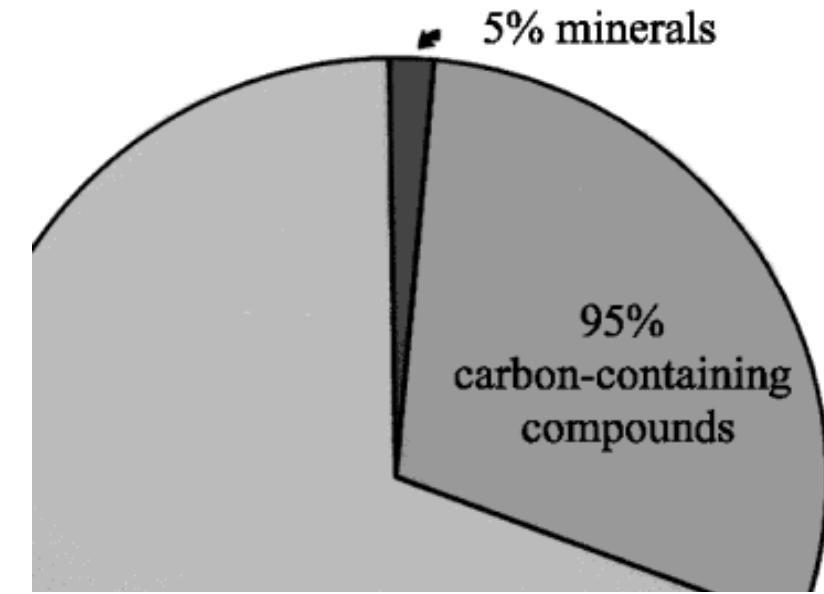
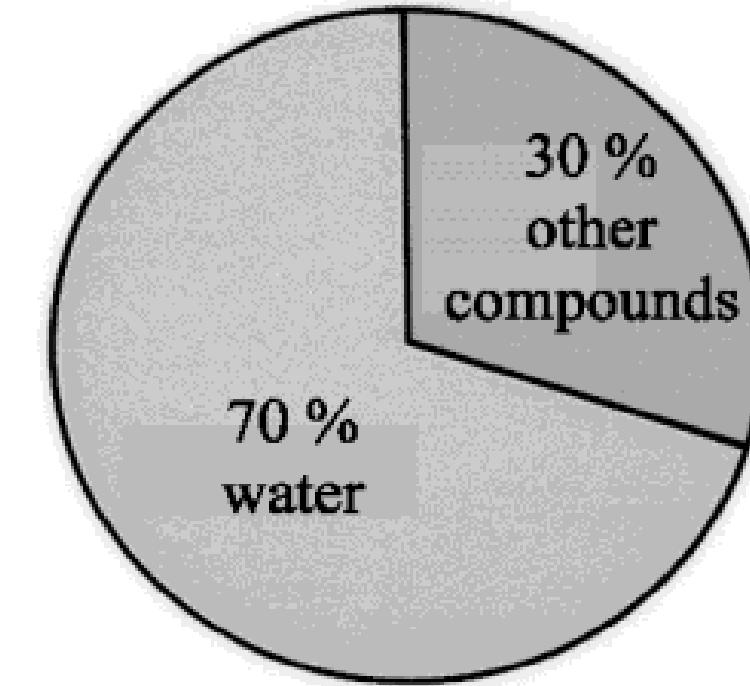
electron microscope

light microscope



10 μm

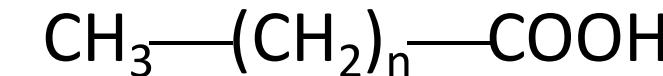
Chemical composition of the cell, the main biogenic elements.



- ❖ The bulk of cell's mass is made up by water
- ❖ Carbon containing compounds are degraded to CO_2 and H_2O by combustion, mineral compounds remain

Fatty Acids

Saturated fatty acids are relatively simple lipids with the general formula

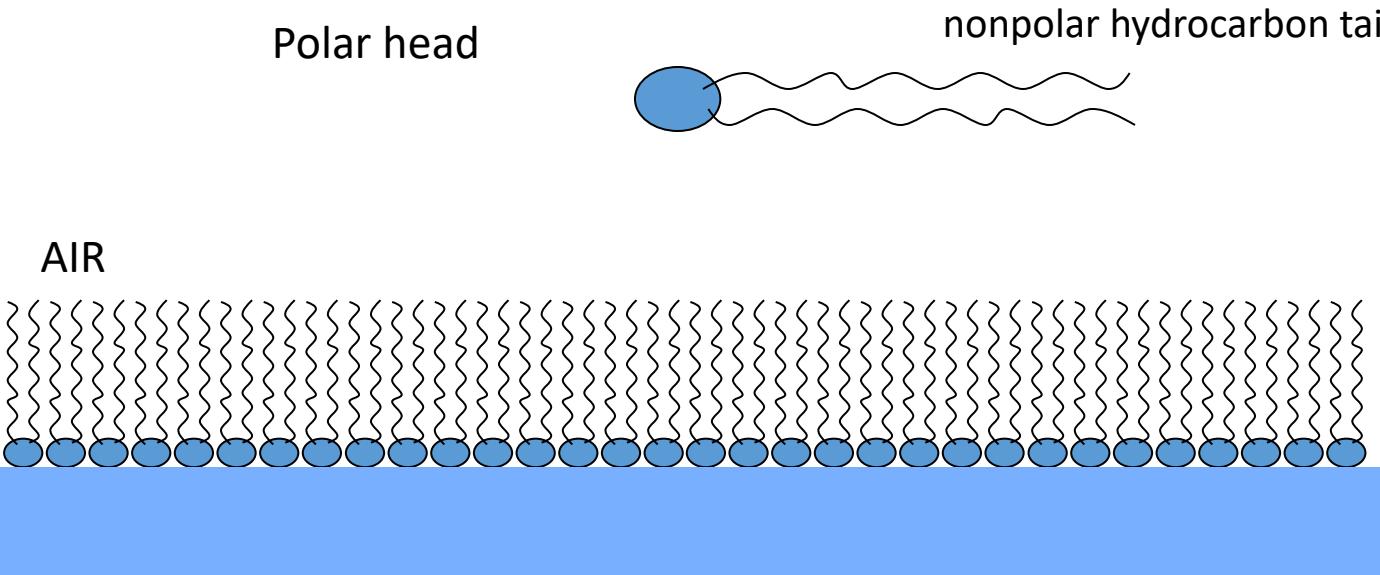


The value of n is typically between 12 and 20 (**Even Numbers in biological systems**)

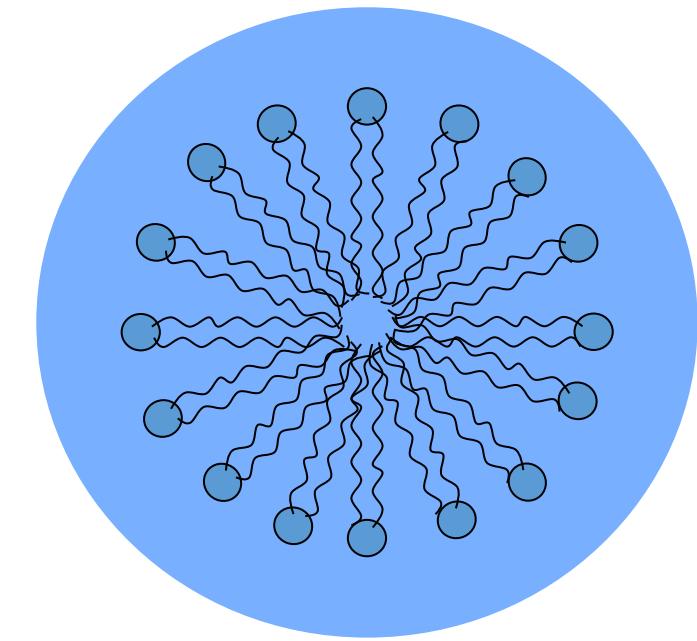
Stearic Acid (saturated) : $\text{CH}_3\text{—}(\text{CH}_2)_{16}\text{—COOH}$

Oleic acid (unsaturated): $\text{CH}_3\text{—}(\text{CH}_2)_7\text{—HC=CH—}(\text{CH}_2)_7\text{—COOH}$

Configurations of Fatty Acids in Water



Lipid monolayer at the air water interface

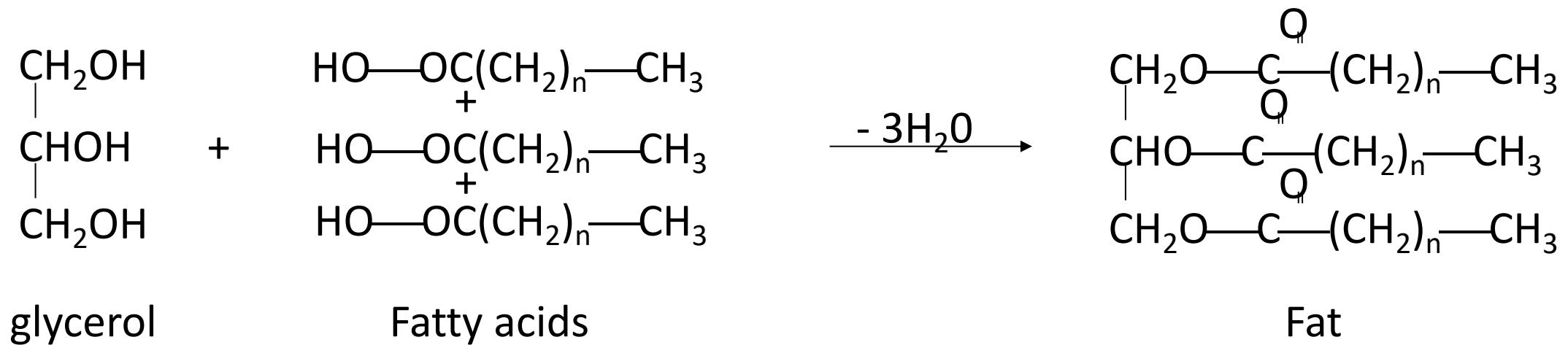


Lipid micelle in water

Lipid molecules have a very small solubility and elevation of the solution concentration above the monomolecular solubility results in the condensation of the excess solute into larger ordered structures called *micelles* ($\Delta G < 0$)

Fats

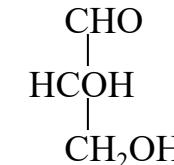
Fats are condensation products of fatty acids and glycerols (esters)



Monosaccharides

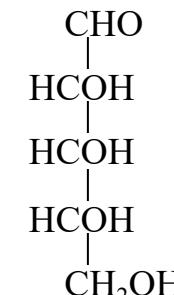
TRIOSE

ALDOSES



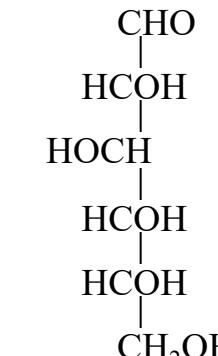
D-glyceraldehyde

PENTOSE

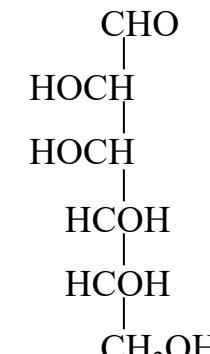


D-ribose

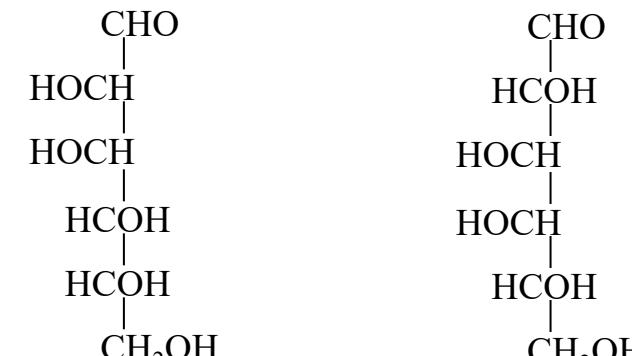
HEXOSE



D-glucose

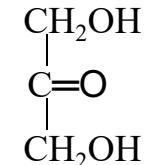


D-mannose

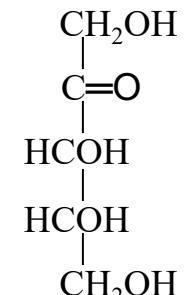


D-galactose

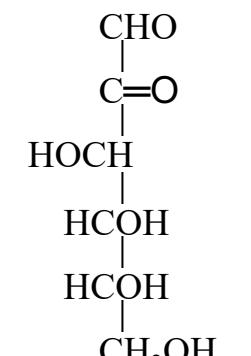
KETOSES



Dihydroxyacetone



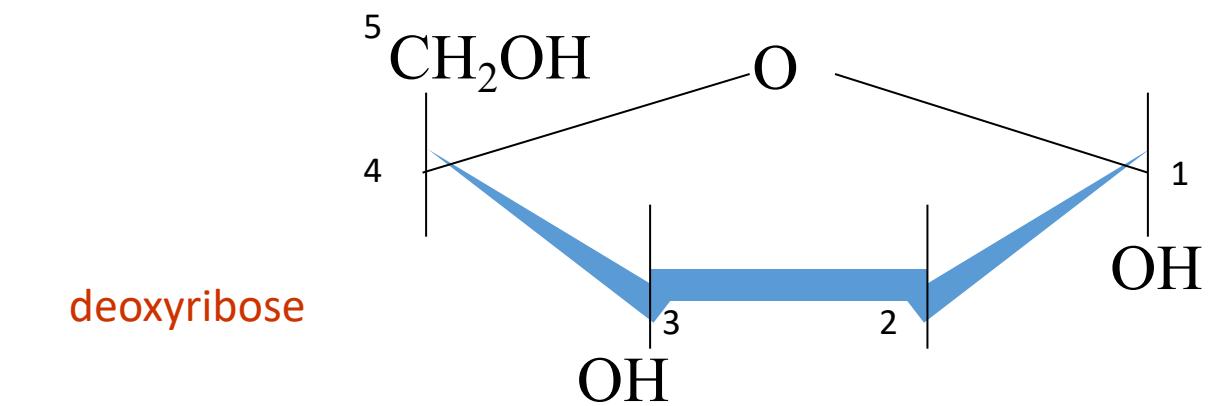
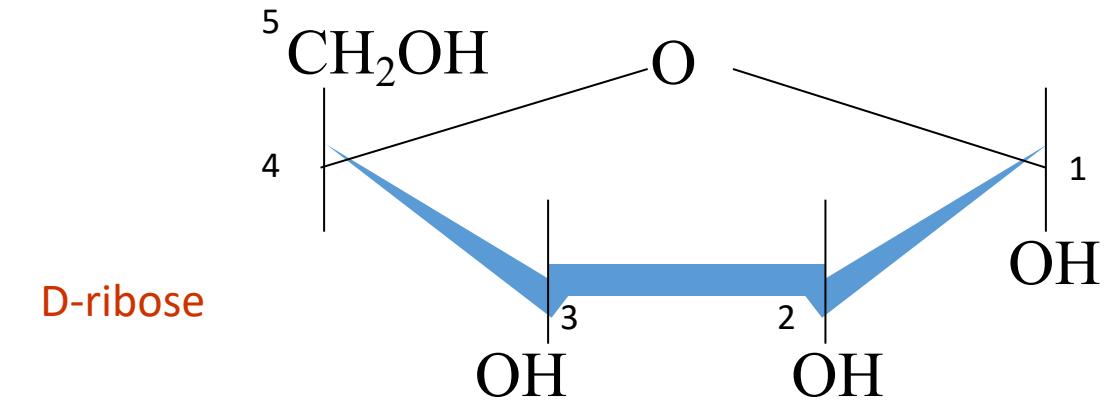
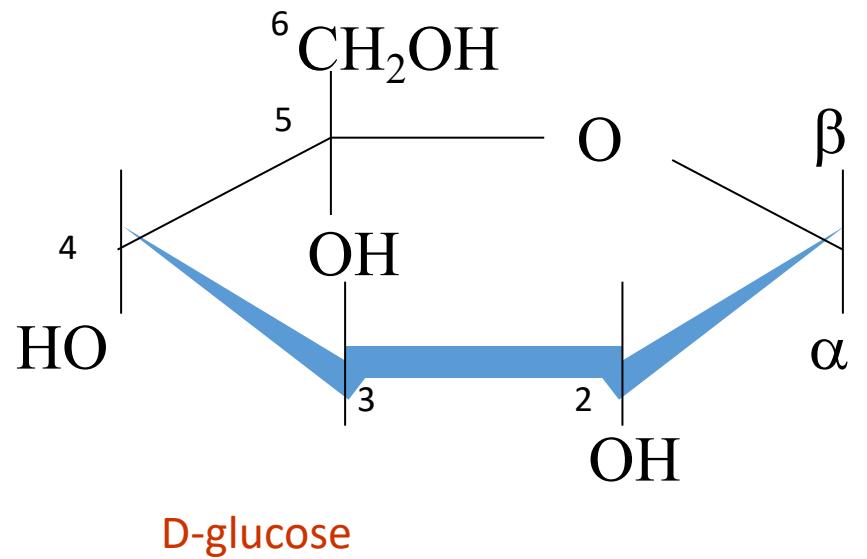
D-ribulose



D-fructose

Monosaccharides

- ❖ D-glucose is the most common monosaccharide found in living organisms
- ❖ Simple sugars are found either in the aldehyde or keto form: For example, glucose is an aldohexose
- ❖ D- is the optical isomer almost exclusively found in living organisms



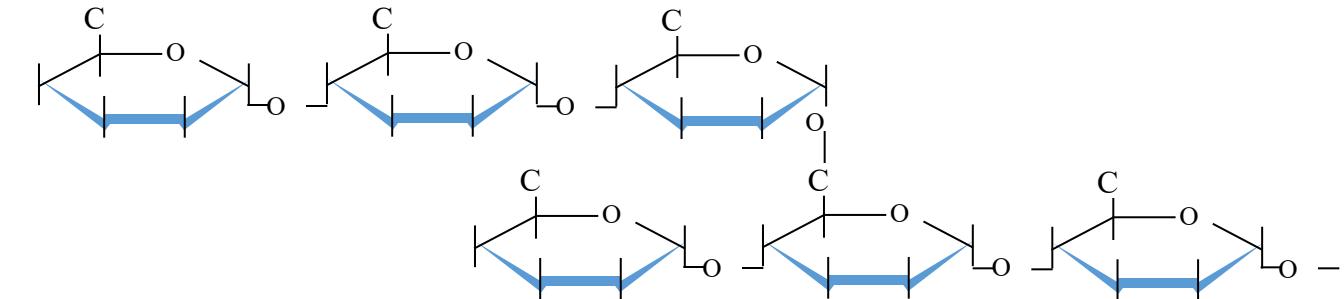
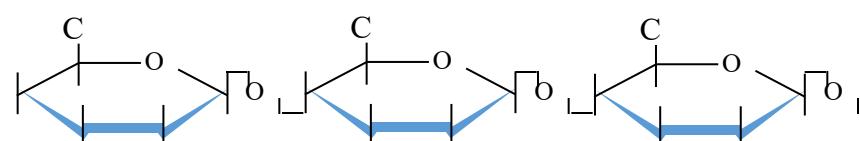
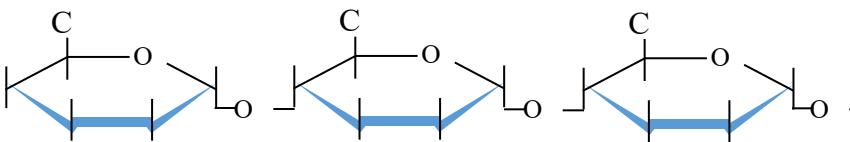
- The five membered rings, D-ribose and deoxyribose are the primary components of the nucleic acid monomers DNA and RNA

Polysaccharides – Starch and Cellulose

- ❖ Amylose – continuous α -1,4-glucosidic bonds
 - ❖ Storage material
 - ❖ Amylose is straight water insoluble
 - ❖ Average molecular weight of amylose is $0.5\text{--}1 \times 10^6$
- ❖ Amylopectin is crosslinked polymer
 - ❖ with branches occurring every 25 glucose units by condensation with the C₆ –OH group
 - ❖ Average molecular weight of amylopectin is $1\text{--}2 \times 10^6$
- ❖ Glycogen – storage material in animals (livers)
 - ❖ Granules having amylopectin structure but more cross-linked; almost every 12 glucose units
- ❖ Cellulose – continuous β -1,6-glucosidic bonds
 - ❖ Structural material

Polysaccharides

- ❖ Draw the structure of amylose
- ❖ Draw the structure of amylopectin
- ❖ Draw the structure of cellulose



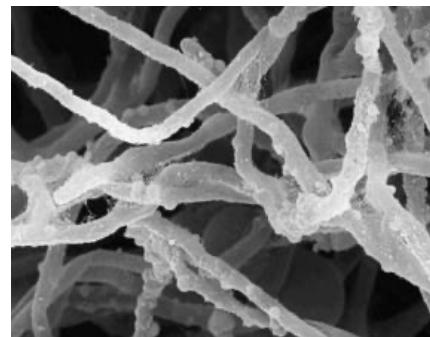
Brown Rot Versus White Rot fungi



Cellulose is eaten away

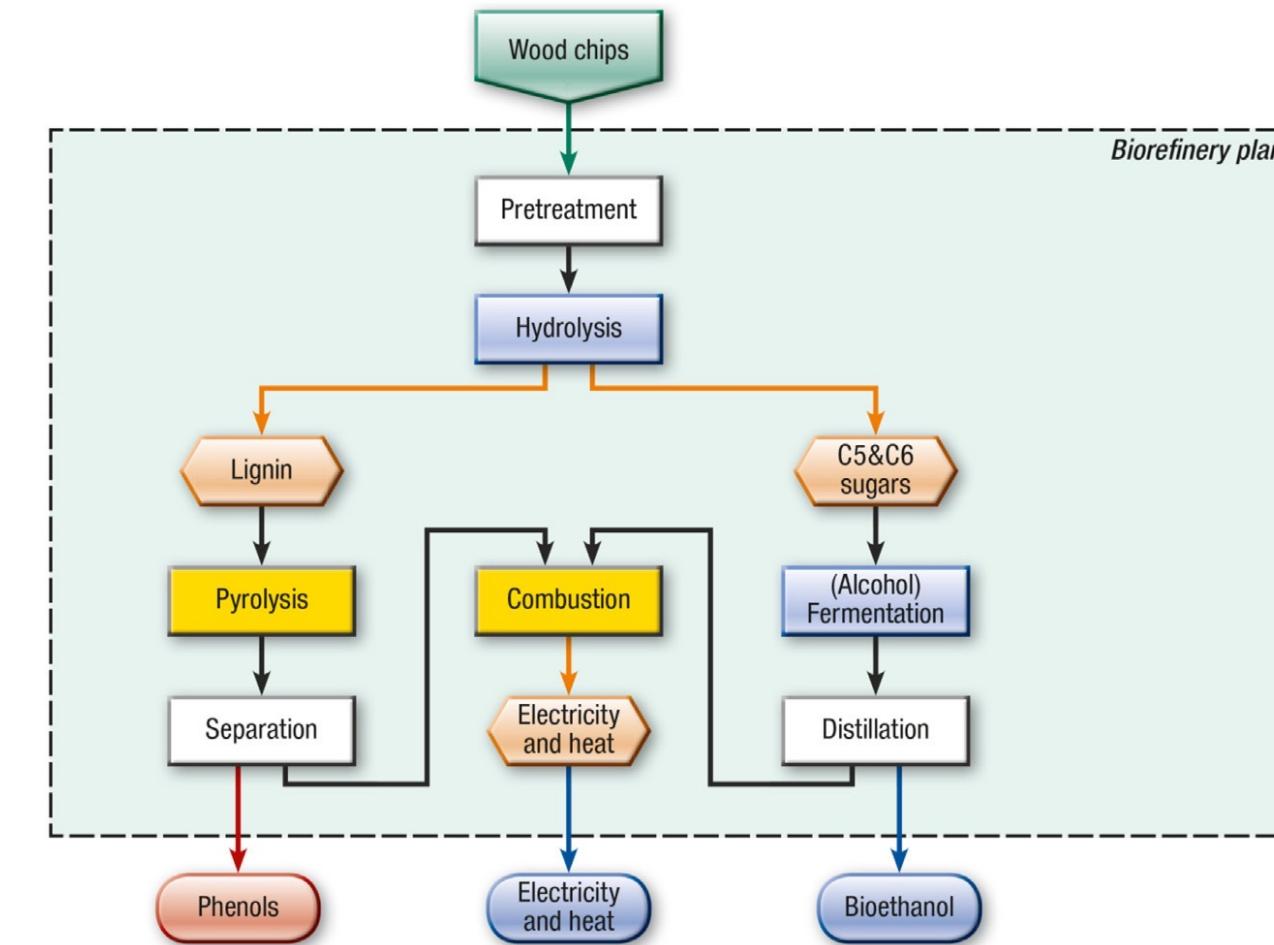


Lignin is eaten away



*Phanerochaete
chrysosporium*

Biorefinery using wood as feed-stock



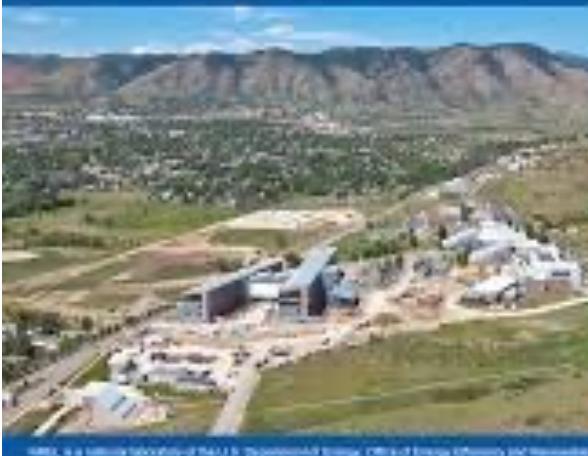
<https://doi.org/10.1016/B978-0-12-813056-8.00005-4>

Production of Bioethanol

N
Q
L
A
C
H
M
T
D
G
K
N
Q
L
P
Y
R
E
I
F
W
L
P
Y
N
Q
L
P
Y

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

**Biochemical Refining of Lignocellulose
to Biofuels: Status and Prospects**



AIChE Annual Meeting,
Salt Lake City, Utah

Sustainable Biorefineries
Plenary, Paper 431b

James D. McMillan
jim.mcmillan@nrel.gov

November 10, 2010

NREL/PR-5100-51132

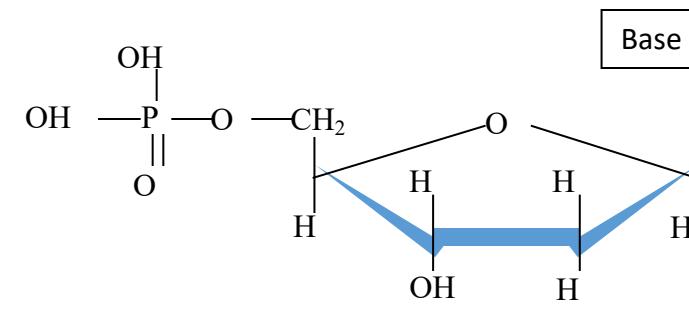
NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.



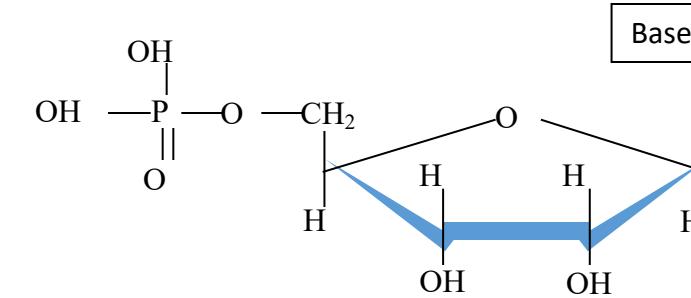
Nucleotides, RNA and DNA

Nucleotides

- ❖ Present in nucleic acids
- ❖ Made up of three components
 - ❖ Phosphoric acid
 - ❖ Ribose or deoxyribose 5-C sugars
 - ❖ Nitrogenous base either purine or pyrimidine



Deoxyribonucleotide

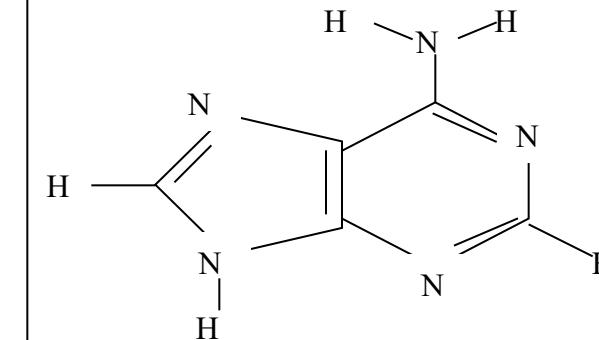


ribonucleotide

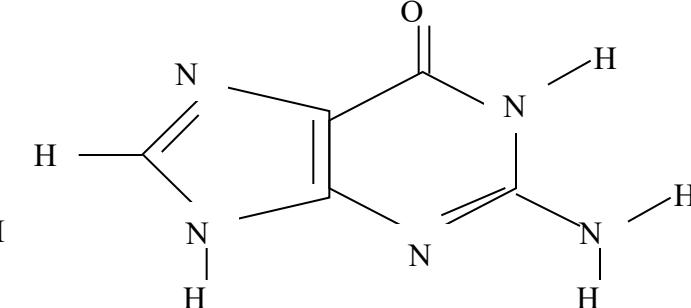
Purines and Pyrimidines

Purines

DNA only



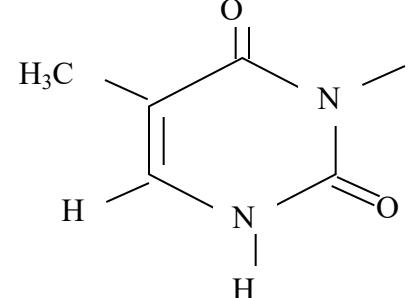
DNA & RNA



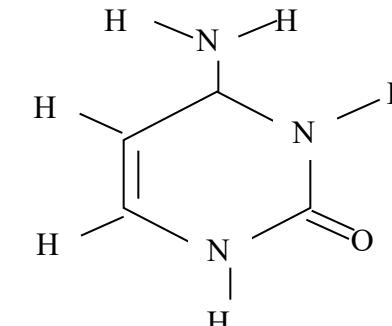
Adenine

Guanine

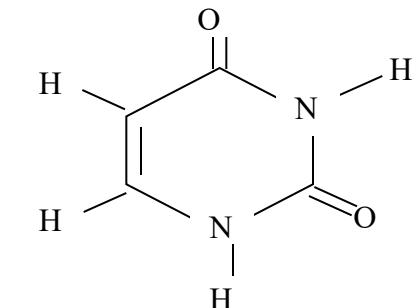
RNA only



Thymine



Cytosine

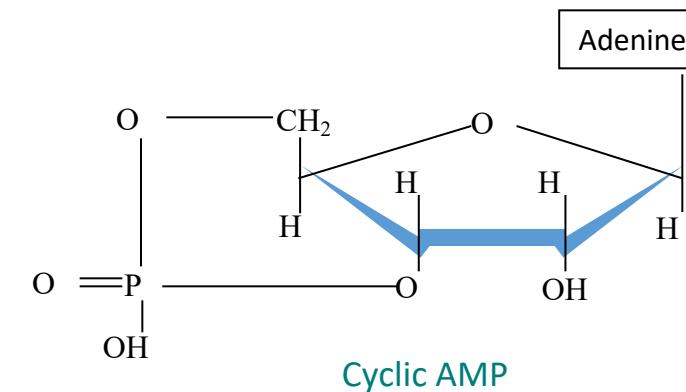
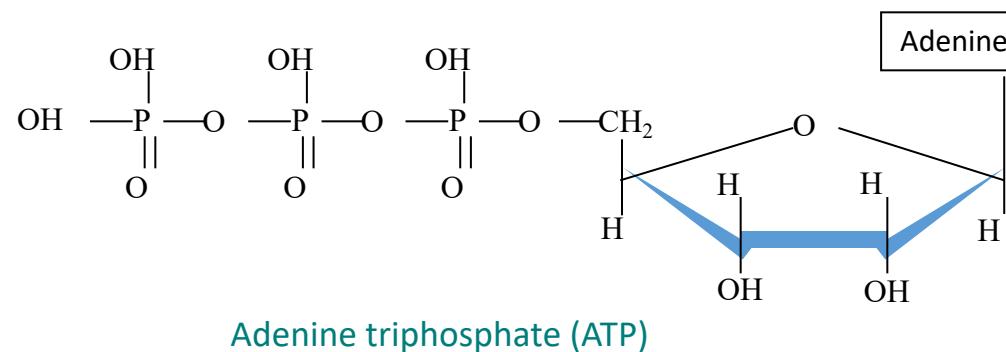
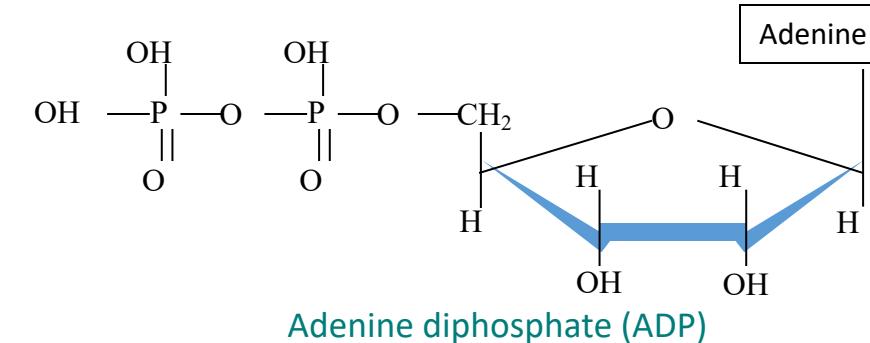
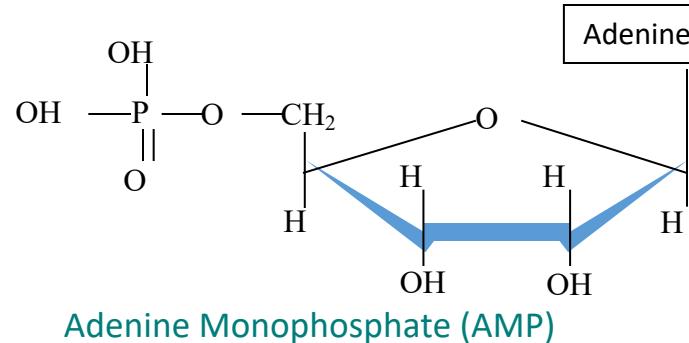


Uracil

Nomenclature of Nucleosides and Nucleotides

Base	Nucleoside	Nucleotide
Adenine (A)	Adenosine	Adenylylate (AMP)
Cytosine (C)	Cytosine	Cytidylate (CMP)
Guanine (G)	Guanosine	Guanulate (GMP)
Uracil (U)	Uridine	Uridylate (UMP)
Thymine (T)	Deoxythymidine	Deoxythymidylate dTMP)

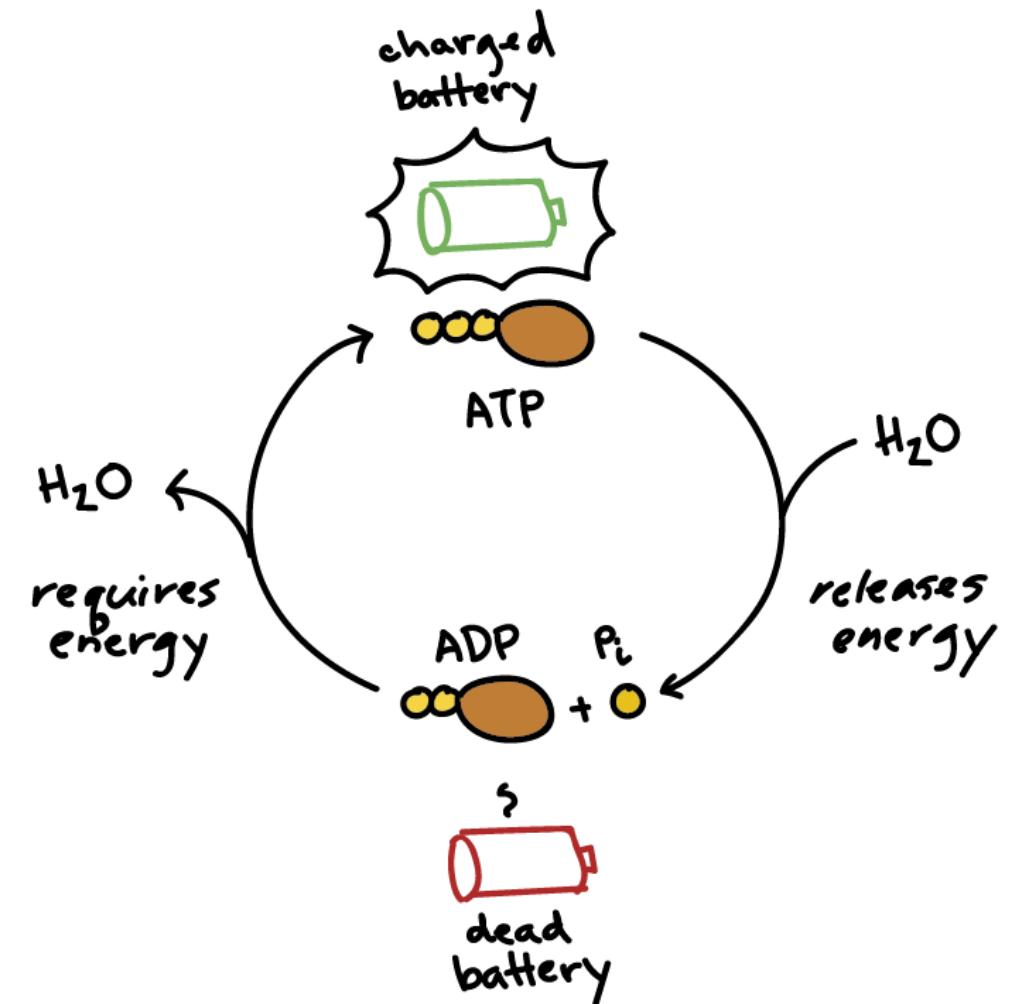
The Phosphates of Adenosine



- ❖ AMP, ADP and ATP are important in cellular energy transfer processes
- ❖ Cyclic AMP serves in regulatory functions

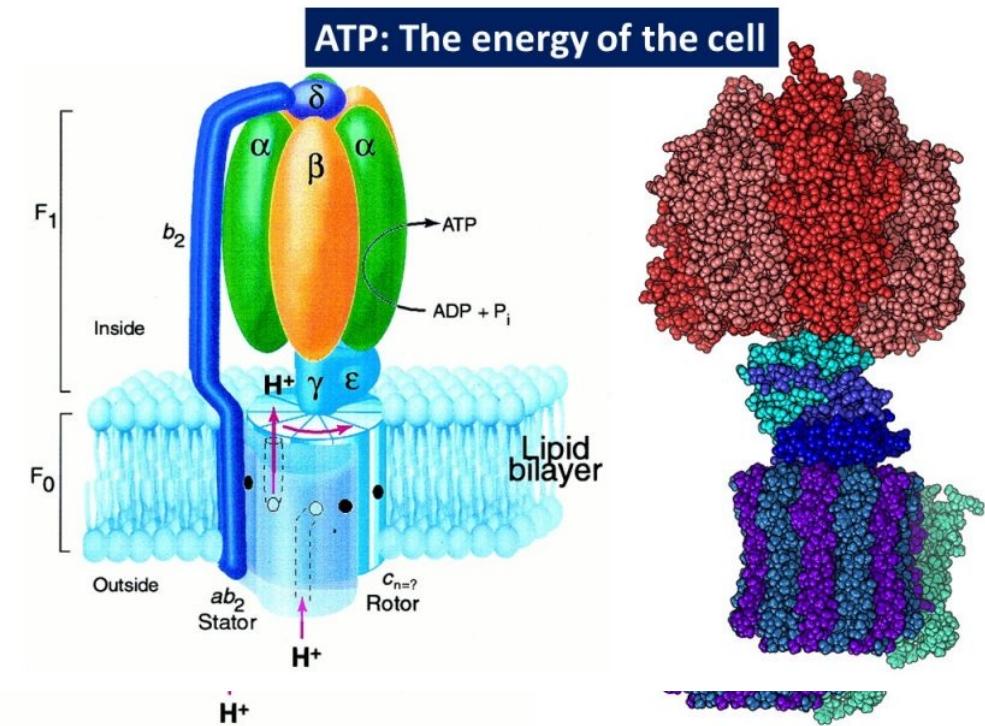
ATP the Energy Currency

- ❖ The conversion of ATP to ADP and phosphate is accompanied by a standard free energy of -7.3 kcal/mol at 37°F and pH 7
- ❖ Energy derived from nutrients or sunlight is stored as ATP in cells
- ❖ Cyclic AMP serves as a regulator in many cellular reactions

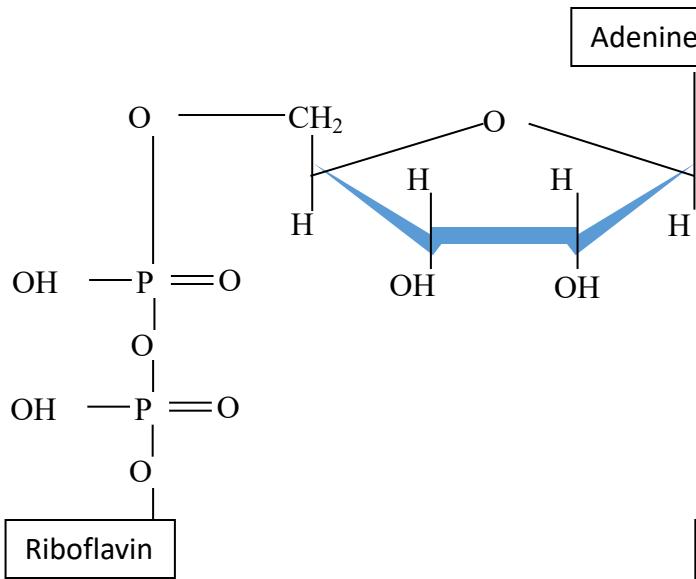


The ATP-DNA Paradox

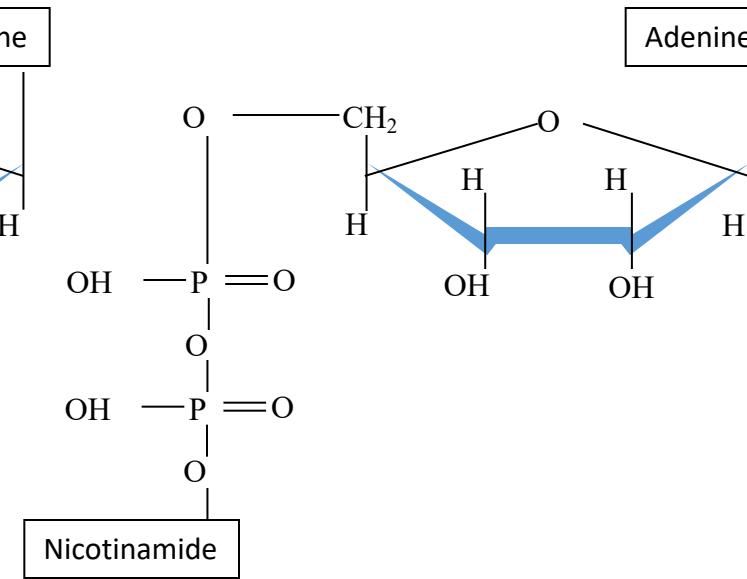
ATP is synthesized by a molecular machine called the ***ATP Synthase***. This machine has to be built from information contained in DNA. But converting this information and stitching amino-acids together require ATP as energy currency.



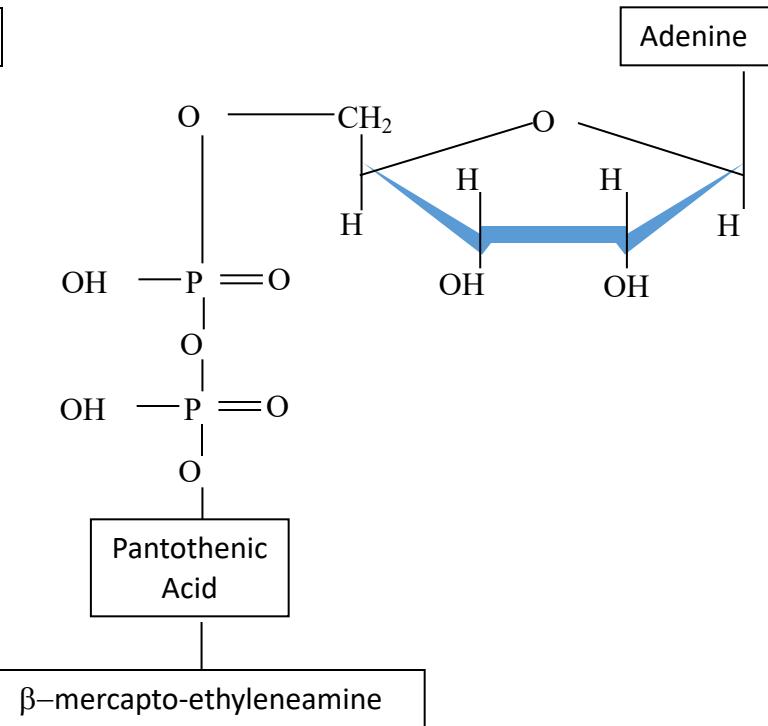
Co-enzymes Derived from Nucleotides



Flavin adenine dinucleotide (FAD)
Oxidized form



Nicotinamide adenine dinucleotide (NAD)

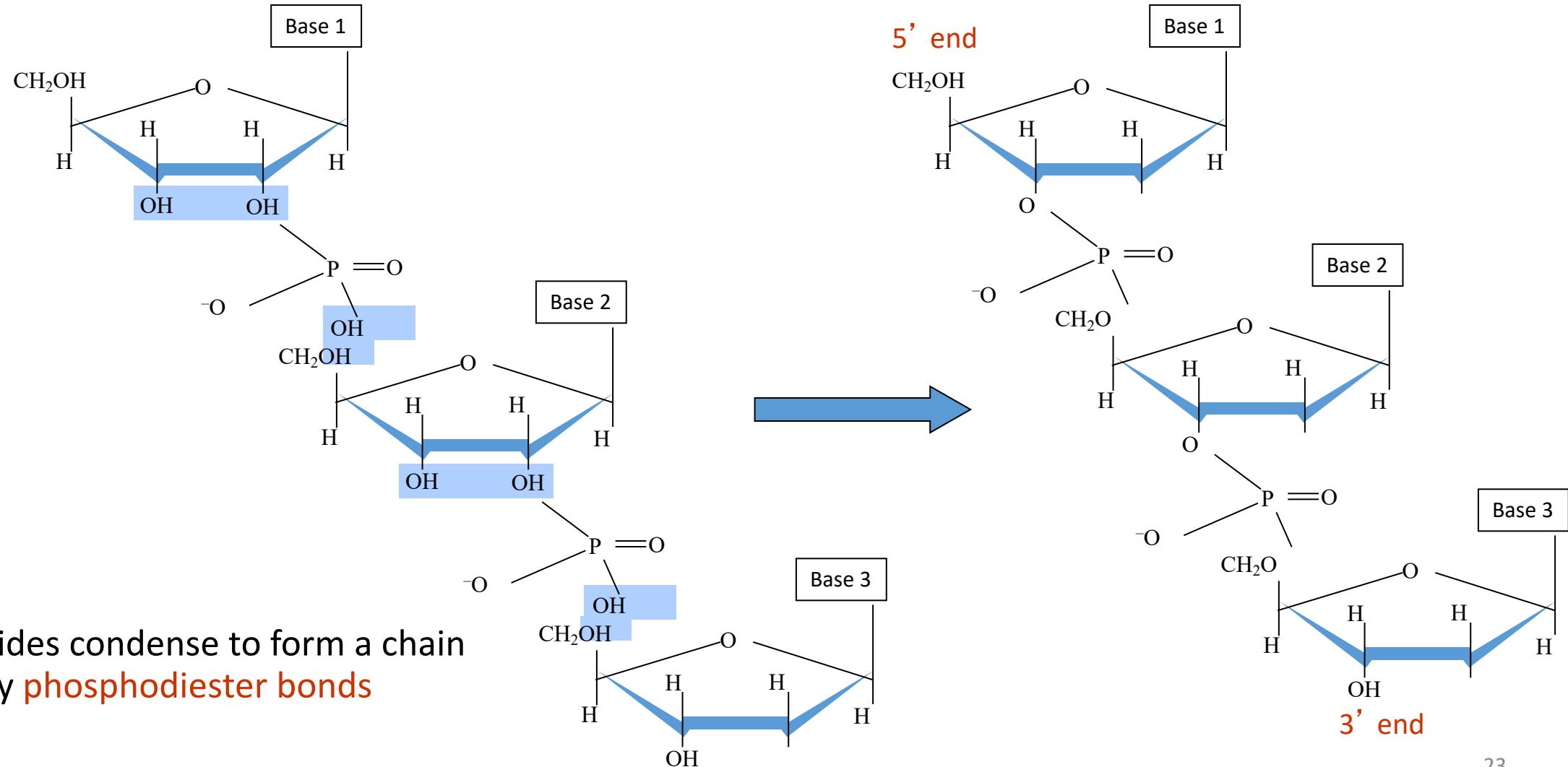


Co-enzyme A

Three important co-enzymes derived from nucleotides

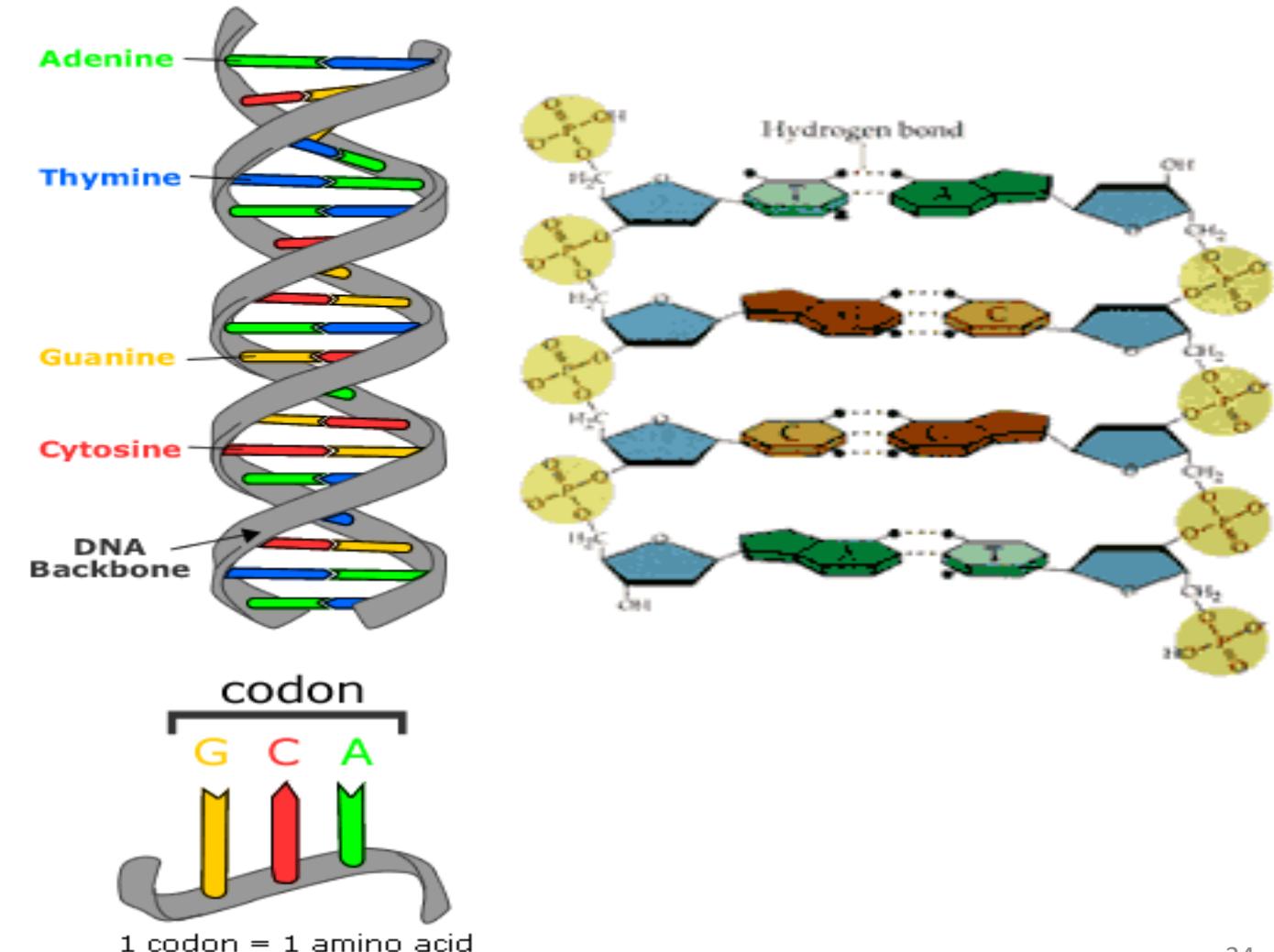
DAO-FAD

Biological Information Storage

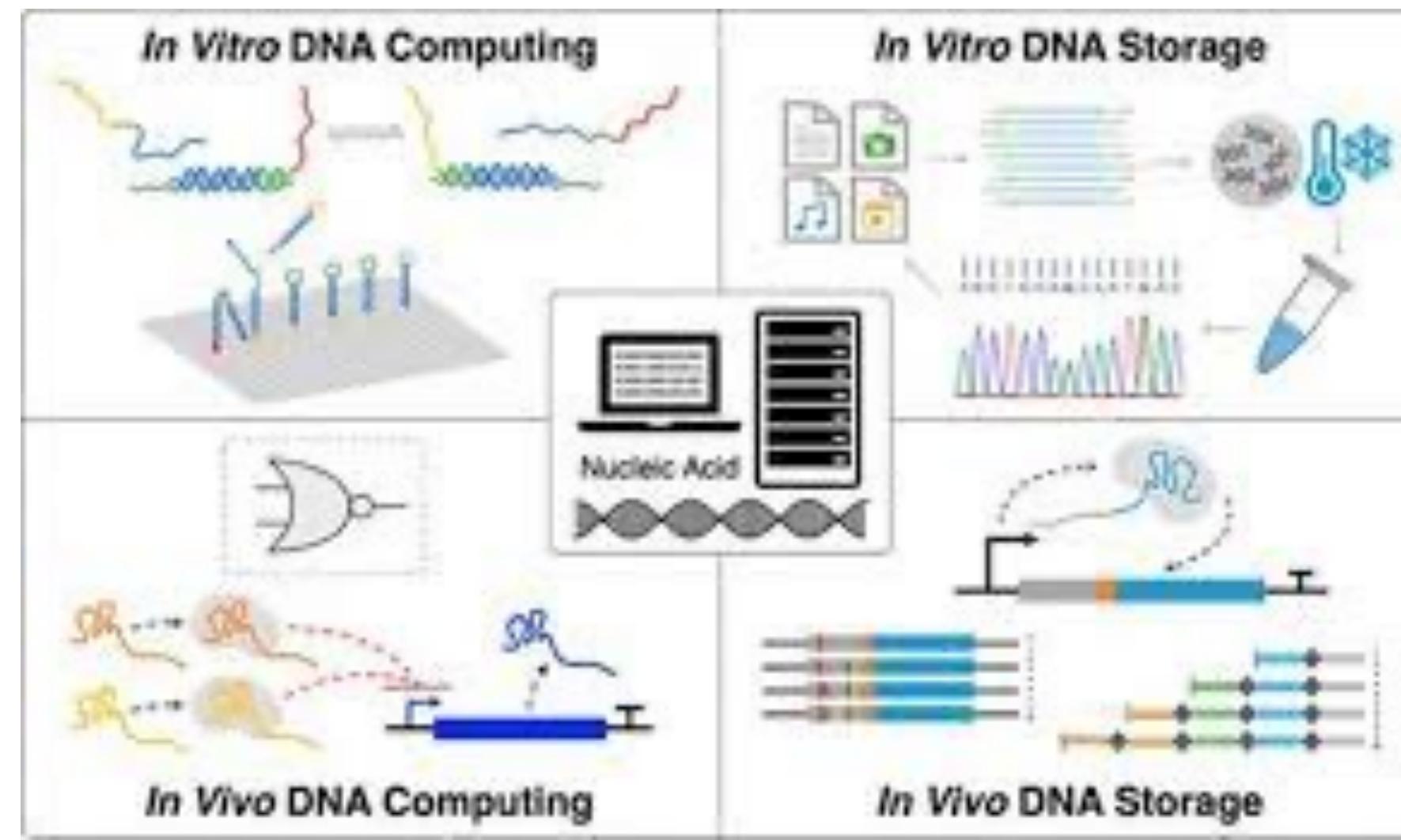


Double Helix Structure of DNA

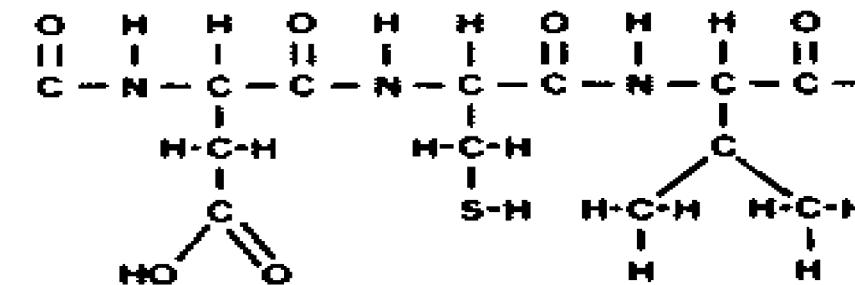
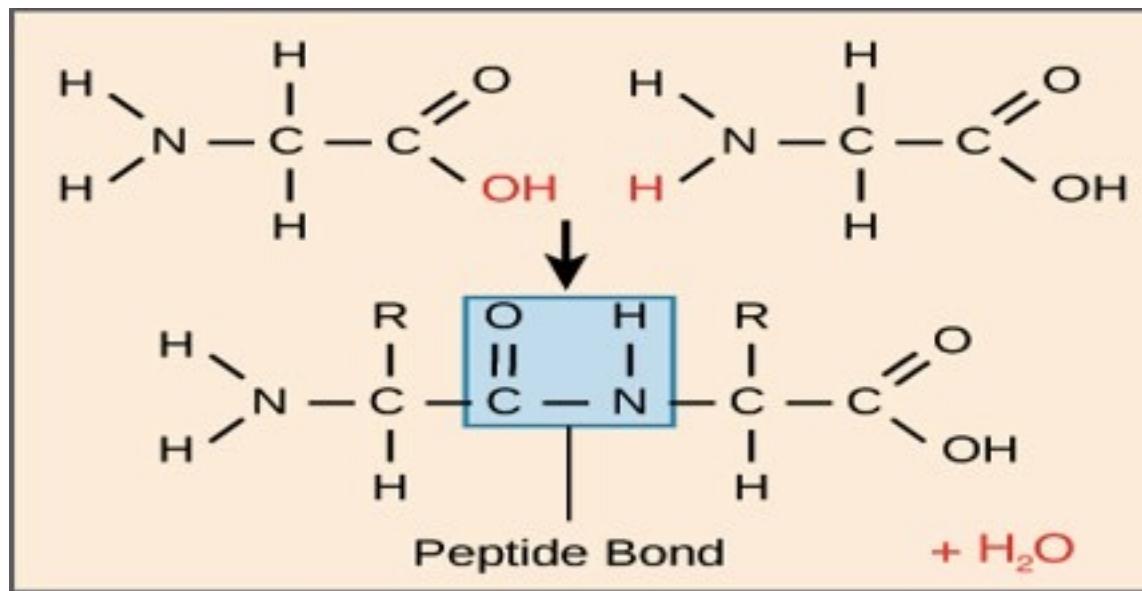
- ❖ James Watson and Francis Crick deduced in 1953, the DNA molecule consists of two polynucleotide chains coiled into a double helix
- ❖ The regular backbone of the molecule is composed of sugar and phosphate units
- ❖ In the interior of the double helix are the purine and pyrimidine bases



DNA Computing



Chemical composition of the cell ...

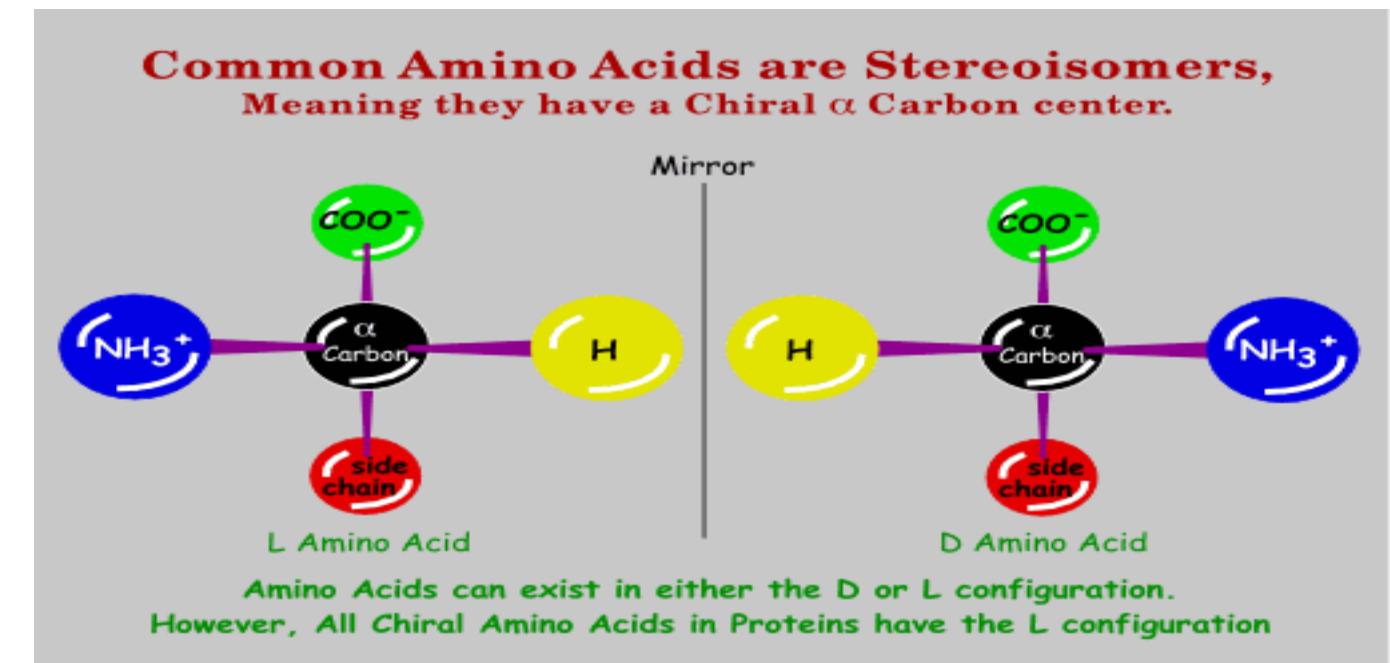


Proteins

- Amino acids are the building blocks of proteins
- There are 20 amino acids commonly found in proteins
- Simple proteins are polymers formed by the condensation of amino acids by forming the peptide bond

L and D Forms of Amino Acids

- ❖ The L and D forms are not superimposable
- ❖ Glycine where R=H is the only exception
- ❖ Living organisms only have the L-form with very rare exception in the cell wall of some bacteria



Individual Properties of Amino Acids

amino acid			mass	surface ^b	volume ^c	pK _a ^d	pl ^e	solubility ^e	density ^e
<u>Alanine</u>	<u>ALA</u>	<u>A</u>	71.09	115	88.6	-	6.107	16.65	1.401
<u>Arginine</u>	<u>ARG</u>	<u>R</u>	156.19	225	173.4	~12	10.76	15	1.1
<u>Aspartic Acid</u>	<u>ASP</u>	<u>D</u>	115.09	150	111.1	4.5	2.98	0.778	1.66
<u>Asparagine</u>	<u>ASN</u>	<u>N</u>	114.11	160	114.1	-	-	3.53	1.54
<u>Cysteine</u>	<u>CYS</u>	<u>C</u>	103.15	135	108.5	9.1-9.5	5.02	very high	-
<u>Glutamic Acid</u>	<u>GLU</u>	<u>E</u>	129.12	190	138.4	4.6	3.08	0.864	1.460
<u>Glutamine</u>	<u>GLN</u>	<u>Q</u>	128.14	180	143.8	-	-	2.5	-
<u>Glycine</u>	<u>GLY</u>	<u>G</u>	57.05	75	60.1	-	6.064	24.99	1.607
<u>Histidine</u>	<u>HIS</u>	<u>H</u>	137.14	195	153.2	6.2	7.64	4.19	-
<u>Isoleucine</u>	<u>ILE</u>	<u>I</u>	113.16	175	166.7	-	6.038	4.117	-
<u>Leucine</u>	<u>LEU</u>	<u>L</u>	113.16	170	166.7	-	6.036	2.426	1.191
<u>Lysine</u>	<u>LYS</u>	<u>K</u>	128.17	200	168.6	10.4	9.47	very high	-
<u>Methionine</u>	<u>MET</u>	<u>M</u>	131.19	185	162.9	-	5.74	3.381	1.340
<u>Phenylalanine</u>	<u>PHE</u>	<u>F</u>	147.18	210	189.9	-	5.91	2.965	-
<u>Proline</u>	<u>PRO</u>	<u>P</u>	97.12	145	112.7	-	6.3	162.3	-
<u>Serine</u>	<u>SER</u>	<u>S</u>	87.08	115	89.0	-	5.68	5.023	1.537
<u>Threonine</u>	<u>THR</u>	<u>T</u>	101.11	140	116.1	-	-	very high	-
<u>Tryptophan</u>	<u>TRP</u>	<u>W</u>	186.12	255	227.8	-	5.88	1.136	-
<u>Tyrosine</u>	<u>TYR</u>	<u>Y</u>	163.18	230	193.6	9.7	5.63	0.0453	1.456
<u>Valine</u>	<u>VAL</u>	<u>V</u>	99.14	155	140.0	-	6.002	8.85	1.230

^a mass [dalton], surface [\AA^2], volume [\AA^3], pK_a [side chain], pl [at 25°C], solubility [g/100g, 25°C], density [crystal density, g/ml].

name: information from [NIST Chemistry WebBook](#), three letter code: GIF, one letter code: VRML

Amino Acid Table

- The R groups are ionizable
- Hydrophobic, hydrophilic, acidic and basic
- Condensation reaction between the amino group of one acid and the carboxyl group of another results in the formation of the peptide bonds

Amino Acid Table									
Non-polar					Polar				
Alanine A	Valine V	Leucine L	Isoleucine I	Proline P	Methionine M	Phenylalanine F	Tryptophan W	Glycine G	Serine S
Threonine T	Cysteine C	Asparagine N	Glutamine Q	Tyrosine Y	Aspartic Acid D	Glutamic Acid E	Lysine K	Arginine R	Histidine H
Acidic					Basic				

Proteins – Food Industry



The hamburger was made from 20,000 muscle fibres grown from stem cells.

Photograph: David Parry/EPA

Dr Post's team at Maastricht University. These fibres were extracted from individual culture wells and then painstakingly pressed together to form the hamburger that was eaten. The objective is to create meat that is biologically identical to beef but grown in a lab rather than in a field as part of a cow.

Chemical composition of the cell ...

Elements that make up life

- ❖ 92 elements are found in nature (118 total elements)
- ❖ Living objects are composed of 25 – 26 elements.

Elements of life...

Elements	Symbols	%
Oxygen	O	65
Carbon	C	18
Hydrogen	H	10
Nitrogen	N	3
Calcium	Ca	1,5
Phosphorus	P	1,0
Sulphur	S	0,25
Potassium	K	0,2
Sodium	Na	0,15
Chlorine	Cl	0,15
Magnesium	Mg	0,05
Others		0,75

❖ The concentration (%) of most important elements (macro-elements) in human body

Elements of Life ...

Approximate amounts of important microelements within 70 kg mass human body

Element	Symbol	Amount	Role in the organism
Iron	Fe	4 – 5 g	Red-ox reactions; oxygen transportation within erythrocytes
Zinc	Zn	1,4 – 2,3 g	Regulation of the growth and development, synthesis of hormones and proteins (hair, skin).
Copper	Cu	75 – 150 mg	Oxidation reactions, biosynthesis of the skin pigment melanin
Manganese	Mn	12 – 20 mg	Formation of skin and mucous layers, development of blood cells
Molybdenum	Mo	5 – 9 mg	Red-ox reactions at respiration
Cobalt	Co	1 – 1,5 mg	Metabolic processes, the component of vitamin B ₁₂
Chromium	Cr	0,6 – 1,4 mg	Sugar turnover, action of insulin

Elements of life...

❖ Other important microelements:

- Lithium (Li) – regulation of nerve functions;
- Selenium (Se) – protein biosynthesis, hair;
- Fluorine (F) – development of bones and teeth;
- Iodine (I) – hormone biosynthesis, neural regulation.

Ultramicroelements:

Arsenic (As) and Gold (Au) – regulation of growth and metabolism

Living organisms operate within laws of physics and chemistry

- ❖ Conservation of Mass, Energy
- ❖ Laws of Chemical Kinetics
- ❖ Principles of Chemical Reactions

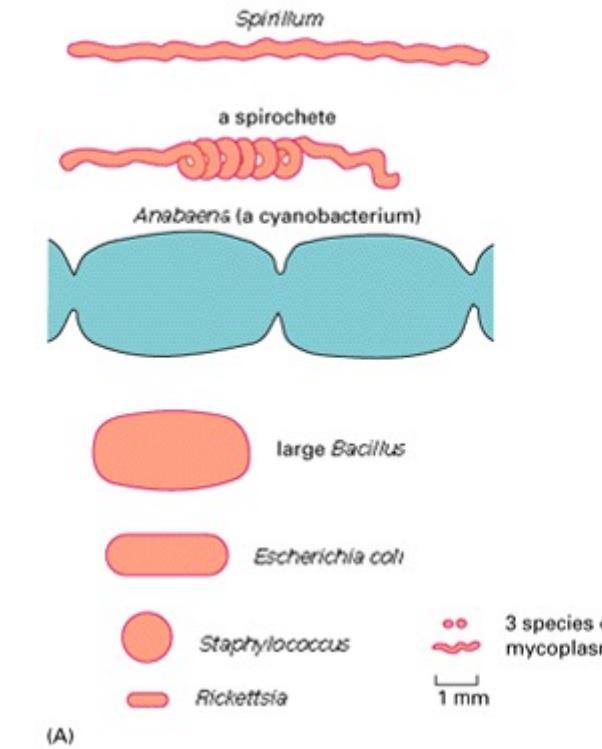
- ❖ Converts energy to work
- ❖ Catalyzes chemical transformations
- ❖ Assembles complex molecules from simple subunits.
- ❖ Complex molecules combine to form supra molecular components, organelles and finally assemble into a cell
- ❖ Cells store and pass on instructions for the assembly of all future generations from simple non-living precursors

From Prokaryotes to Eukaryotes

Lecture 4

Prokaryotic Cells Are Structurally Simple but Biochemically Diverse

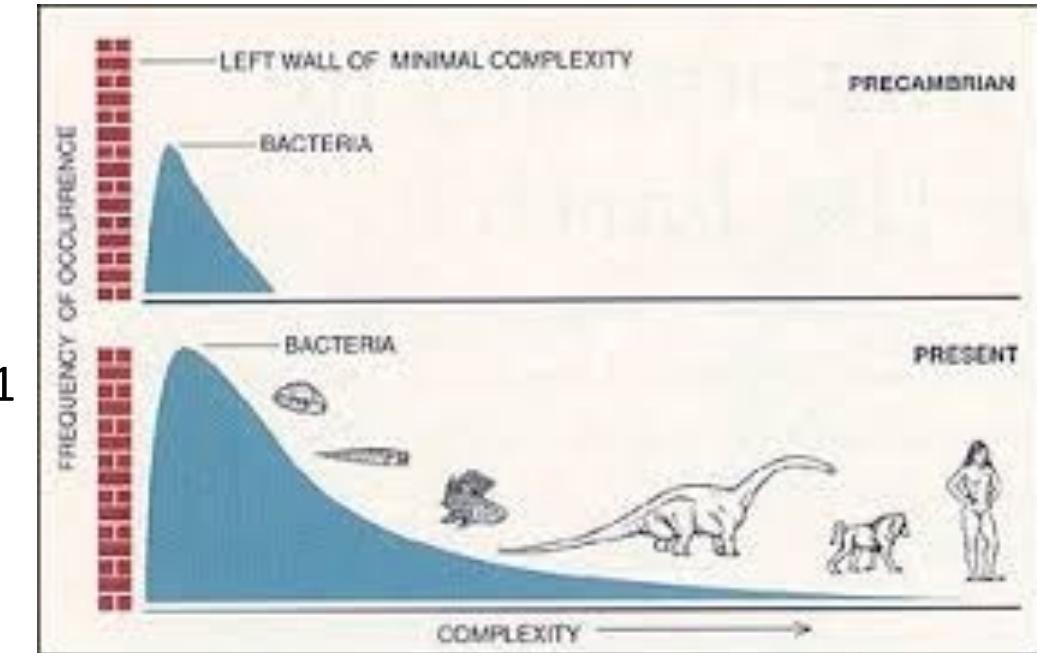
- ❖ Bacteria are the simplest organisms found in most natural environments.
- ❖ Spherical or rod-shaped cells, commonly several micrometers in linear dimension Possess a tough protective coat, cell wall, beneath which a plasma membrane encloses a single cytoplasmic compartment containing DNA, RNA, proteins, and small molecules.



Prokaryote sizes and structures. (A) Some prokaryotic cells drawn to scale. (B) Electron micrograph of a longitudinal section through a bacterium (*Escherichia coli*); the cell's DNA is concentrated in the palely stained region. (Courtesy of E. Kellenberger.)

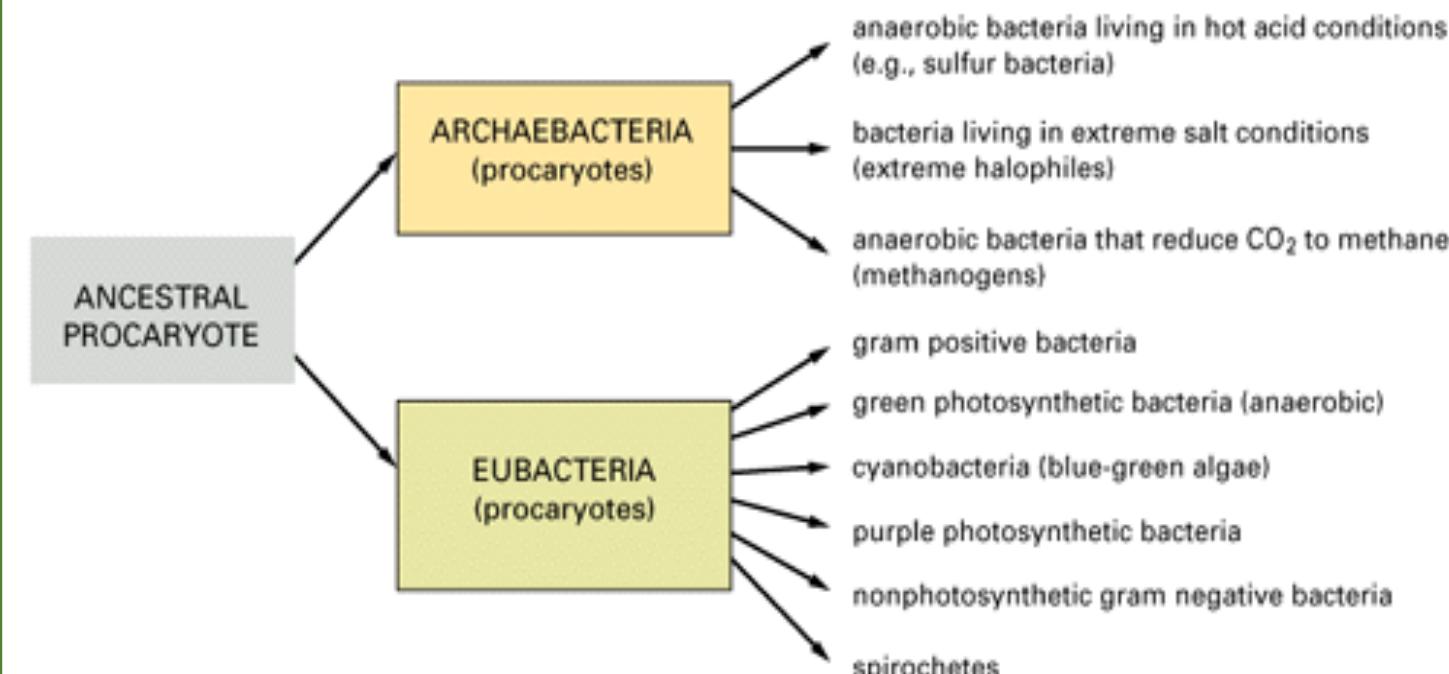
The power of the lowly bacteria

- ❖ Bacteria replicate quickly, dividing in two by binary fission.
- ❖ When food is plentiful, "survival of the fittest" means survival of those that can divide the fastest.
- ❖ Under optimal conditions a single prokaryotic cell can divide every 20 minutes and thereby give rise to 5 billion cells (approximately equal to the present human population on earth) in less than 11 hours.
- ❖ Ability to divide quickly enables bacteria to adapt rapidly to changes in their environment.

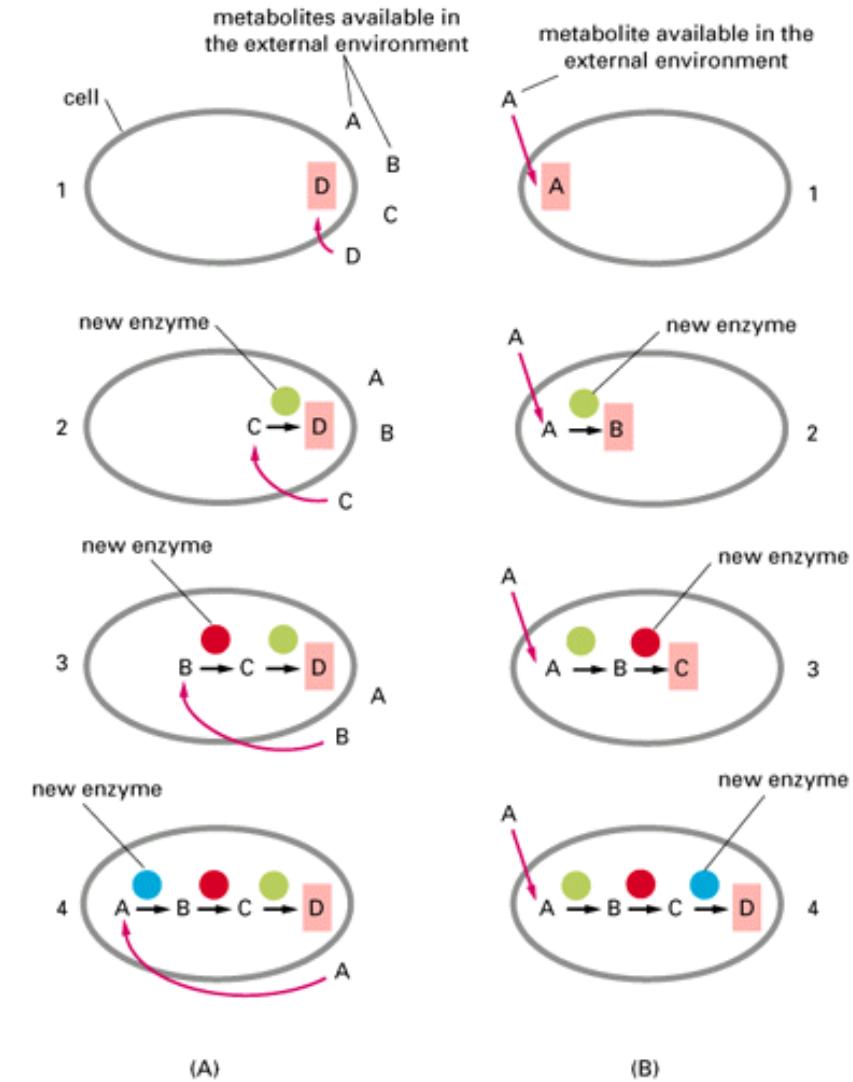


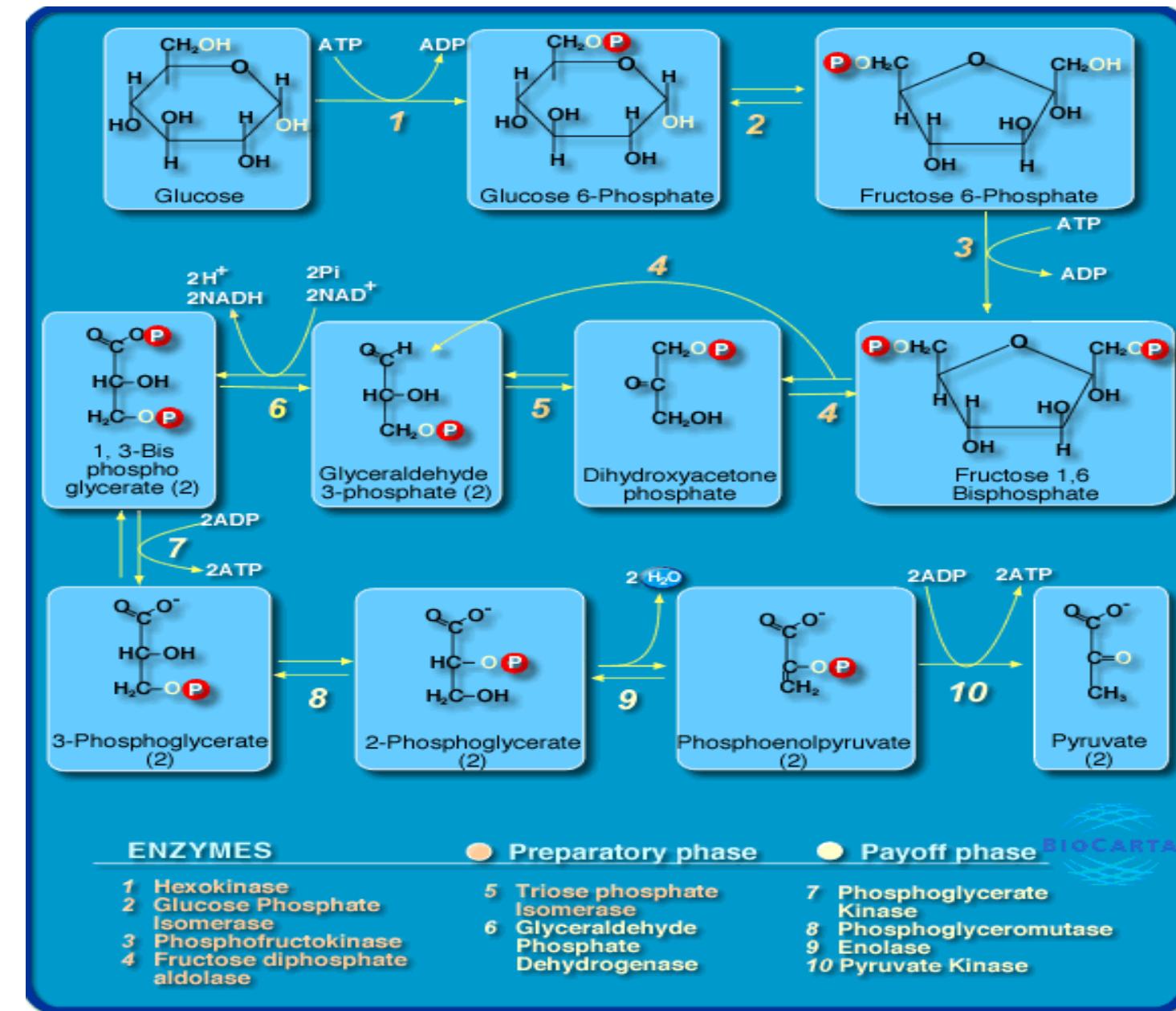
Family relationships between present-day bacteria

- ❖ Bacteria live in an enormous variety of ecological niches, and show a corresponding richness in their underlying biochemical composition.
- ❖ Two distantly related groups can be recognized: eubacteria, inhabit soil, water, and larger living organisms; and archaebacteria, found in ocean depths, salt brines, and hot acid springs.

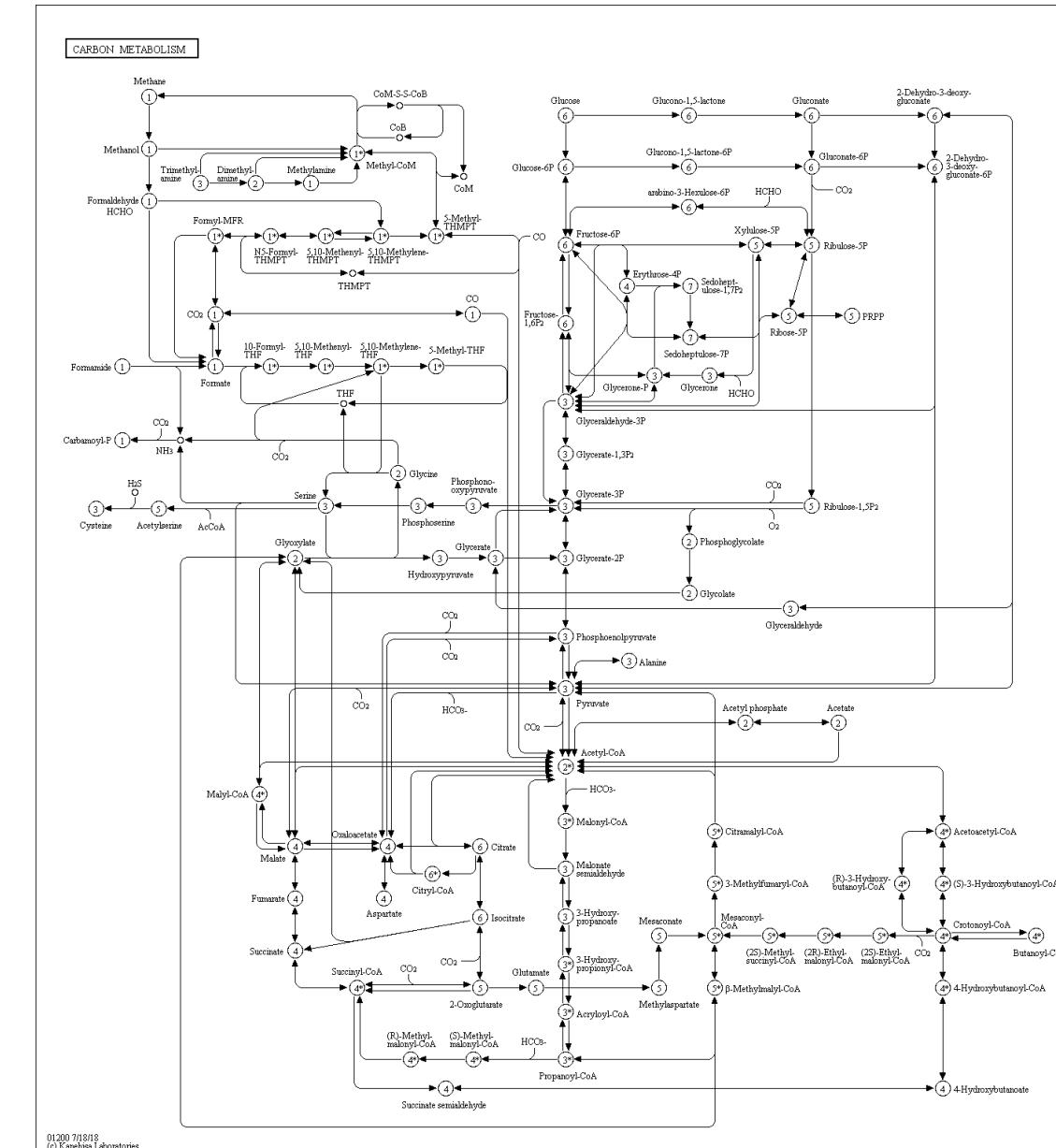


More about metabolic reactions





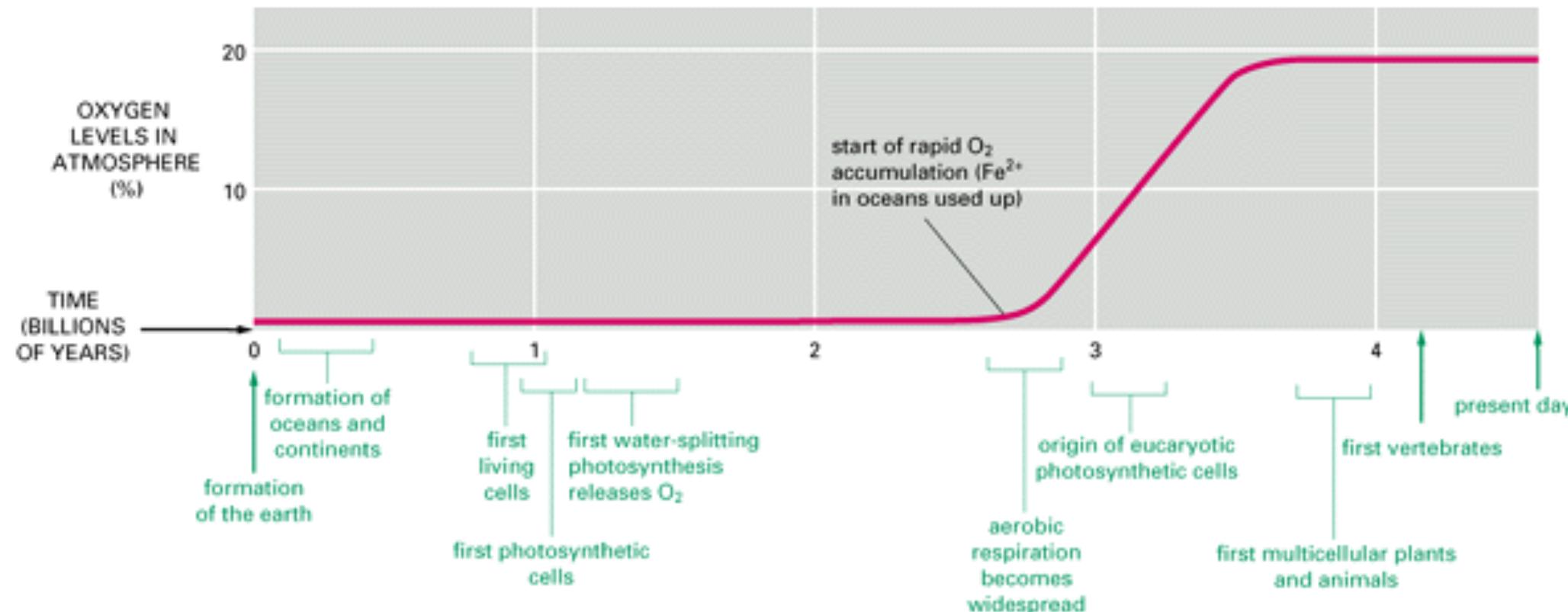
Carbon Metabolism



Metabolic reactions are similar in all organisms

- ❖ Similarity in all kinds of organisms, suggesting an extremely ancient origin
- ❖ Linked to core reactions of glycolysis are hundreds of other chemical processes
 - ❖ Generation of energy in ATP-ADP currency
 - ❖ Synthesis of small molecules
 - ❖ Make large polymers specific to the organism
 - ❖ Degrade complex molecules, taken in as food, into simpler chemical units

The synthesis and release of oxygen into the atmosphere



Cyanobacteria were the first organisms to synthesize oxygen

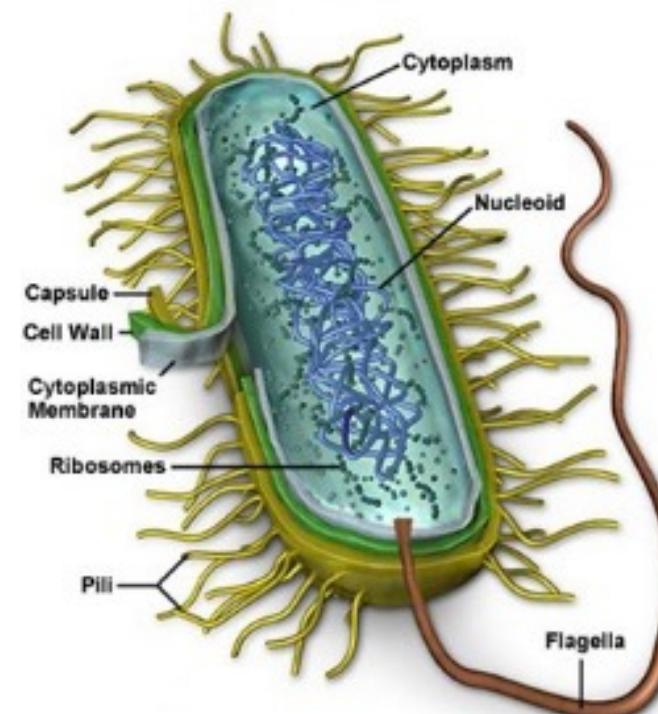
- ❖ As competition for raw materials for organic syntheses intensified, a selective advantage gained by organisms able to utilize carbon and nitrogen atoms (in the form of CO₂ and N₂) directly from the atmosphere
- ❖ While they are abundantly available, it required a large amount of energy to convert CO₂ and N₂ to a usable organic form like simple sugars
- ❖ The major mechanism that evolved to achieve this was photosynthesis: radiant energy captured from the sun converted CO₂ into organic compounds
- ❖ Interaction of sunlight with chlorophyll, excites an electron to a more highly energized state. As the electron drops back to a lower energy level, the energy it gives up drives chemical reactions that are facilitated and directed by protein molecules



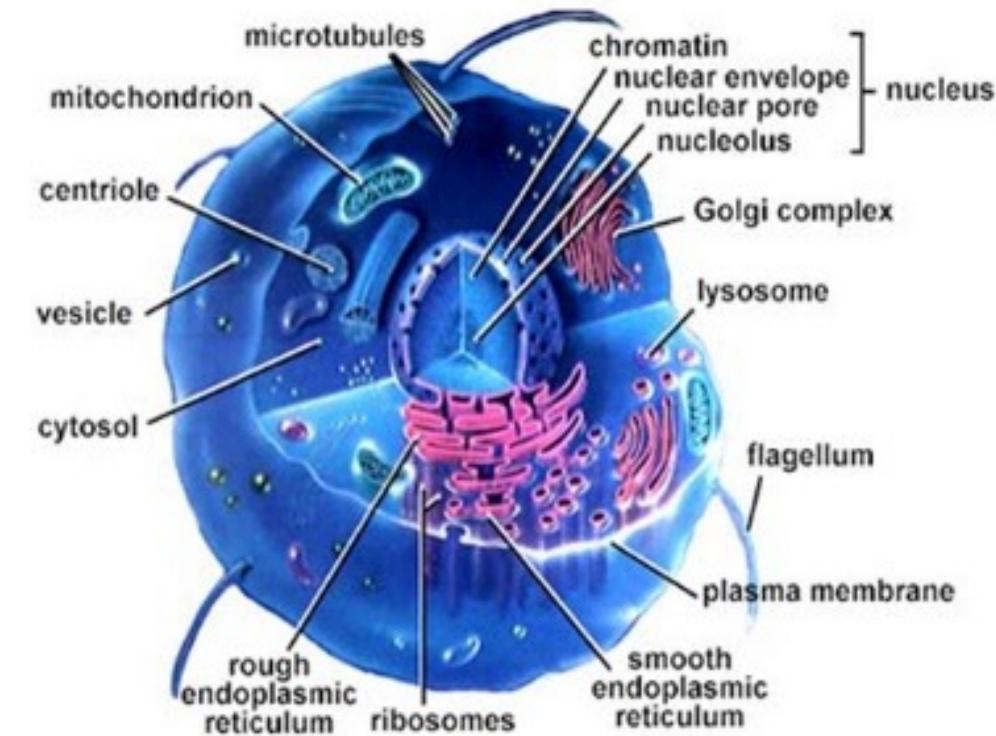
The utilization of oxygen by organisms

- ❖ Extremely reactive chemical that can interact with most cytoplasmic constituents; must have been toxic to many early organisms
- ❖ The simplest of carbon molecules used by organisms is glucose
- ❖ In the absence of oxygen glucose broken down only to lactic acid or ethanol, the end products of anaerobic glycolysis.
- ❖ In the presence of oxygen glucose completely degraded to CO₂ and H₂O; much more energy can be derived from each gram of glucose

Prokaryotes and Eucaryotes



prokaryotic cell
(bacteria)

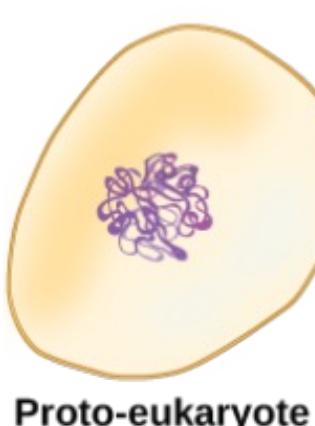


eukaryotic cell
(protists, fungi, animals, plants)

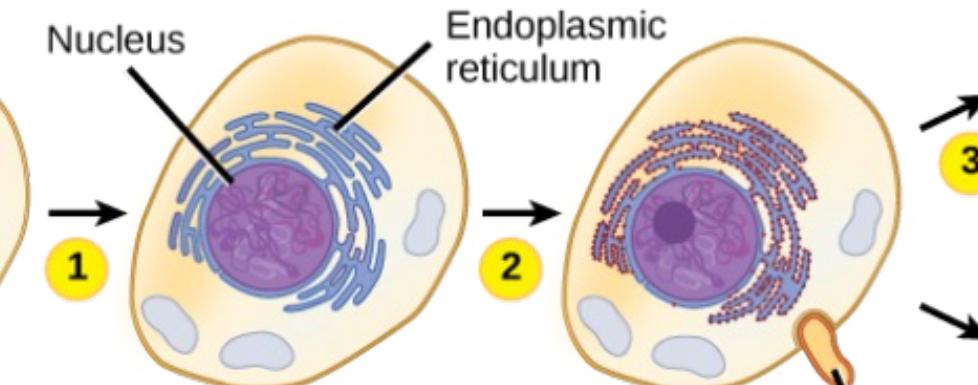
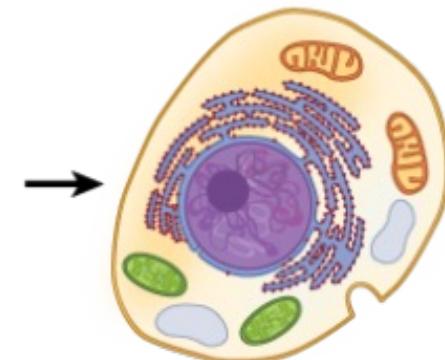
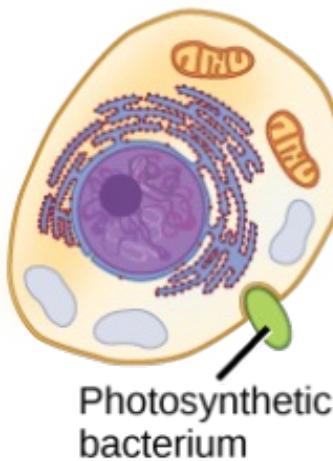
The creation of the eukaryote cell

The ENDOSYMBIOTIC THEORY

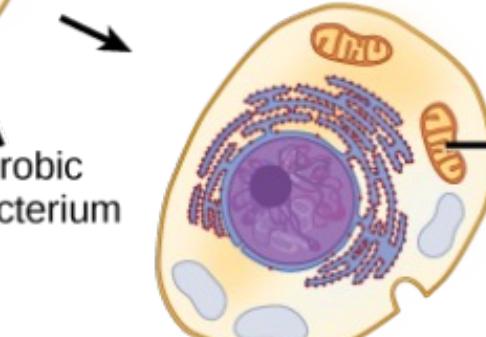
1 Infoldings in the plasma membrane of an ancestral prokaryote gave rise to endomembrane components, including a nucleus and endoplasmic reticulum.



3 In a second endosymbiotic event, the early eukaryote consumed photosynthetic bacteria that evolved into chloroplasts.



2 In a first endosymbiotic event, the ancestral eukaryote consumed aerobic bacteria that evolved into mitochondria.



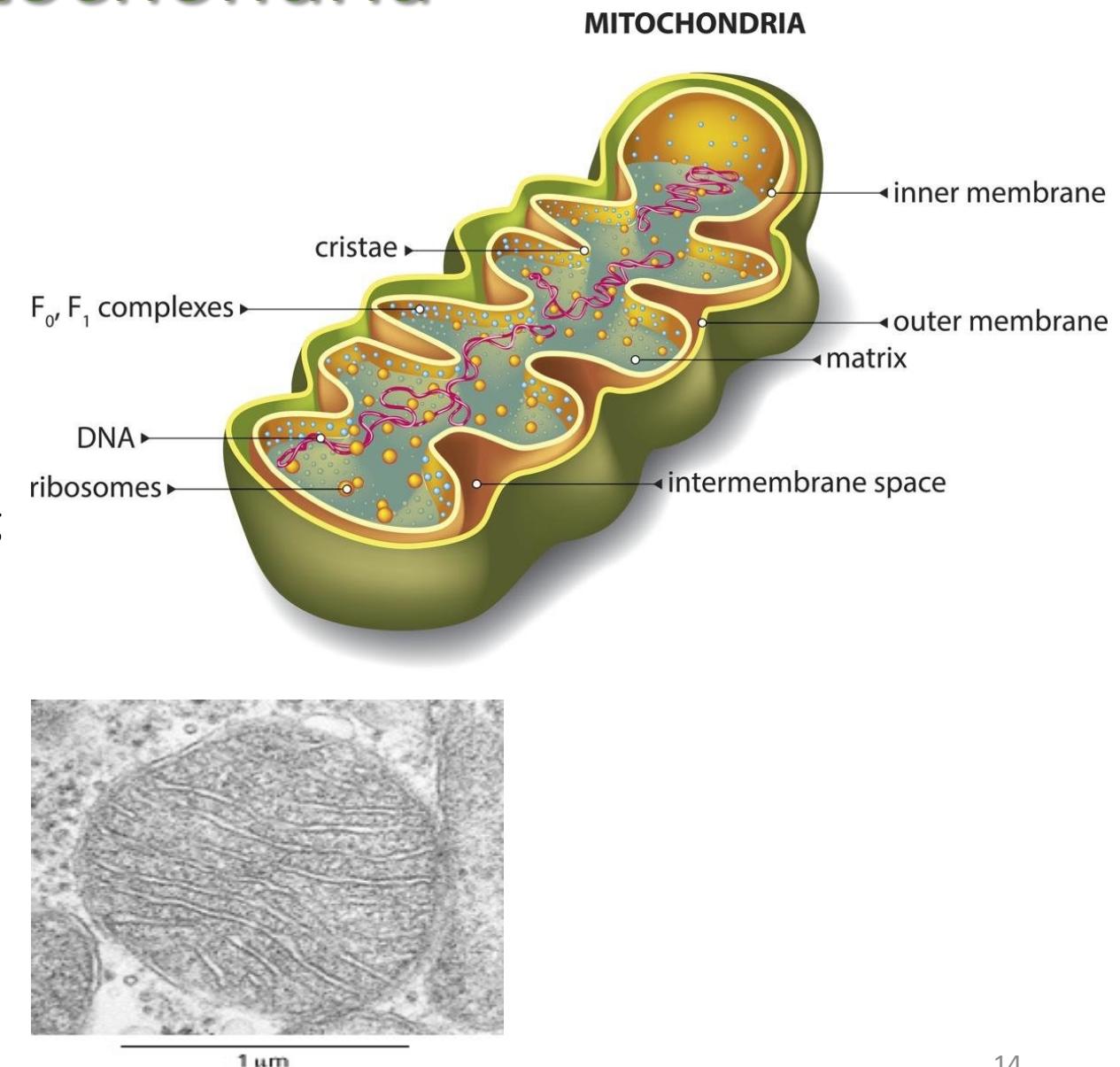
Modern heterotrophic eukaryote

Mitochondrion

Modern photosynthetic eukaryote

The mitochondria

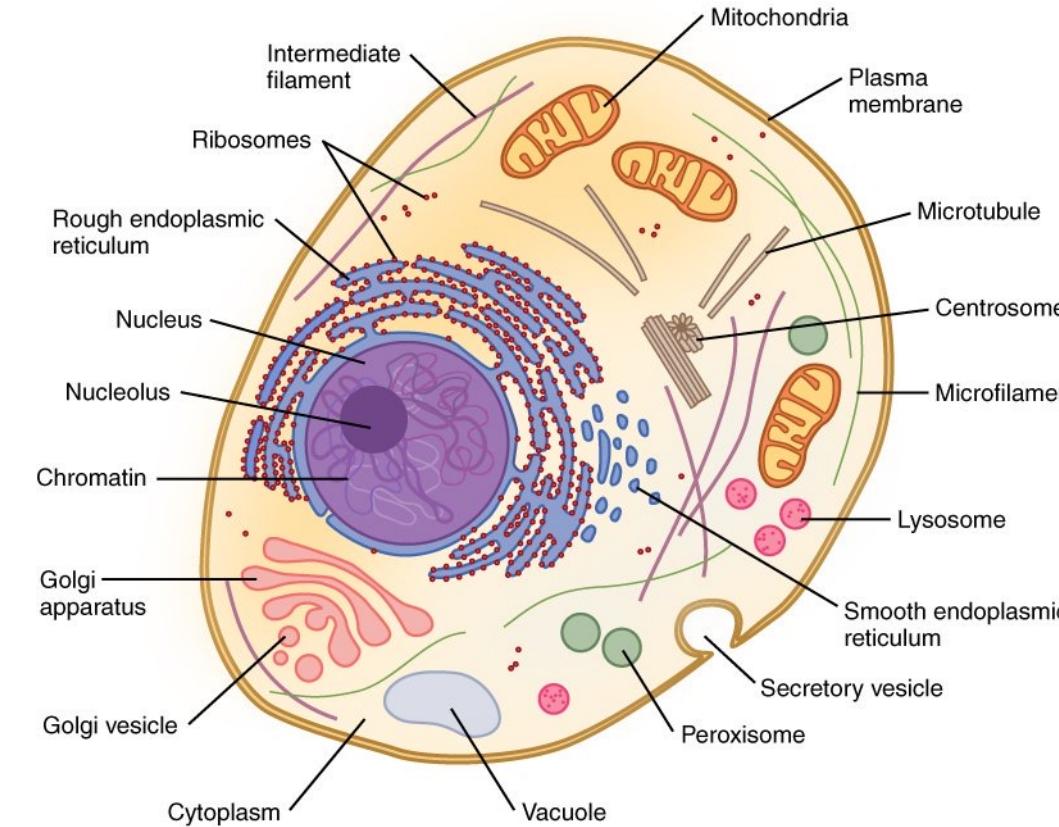
- ❖ Often resemble bacteria in size and shape,
- ❖ Contain DNA, make protein, reproduce by dividing in two and is responsible for respiration.
- ❖ Many present-day bacteria respire like mitochondria
- ❖ The amoeba *Pelomyxa palustris*, while lacking mitochondria, nevertheless carries out *oxidative* metabolism by harboring aerobic bacteria in its cytoplasm in a permanent symbiotic relationship



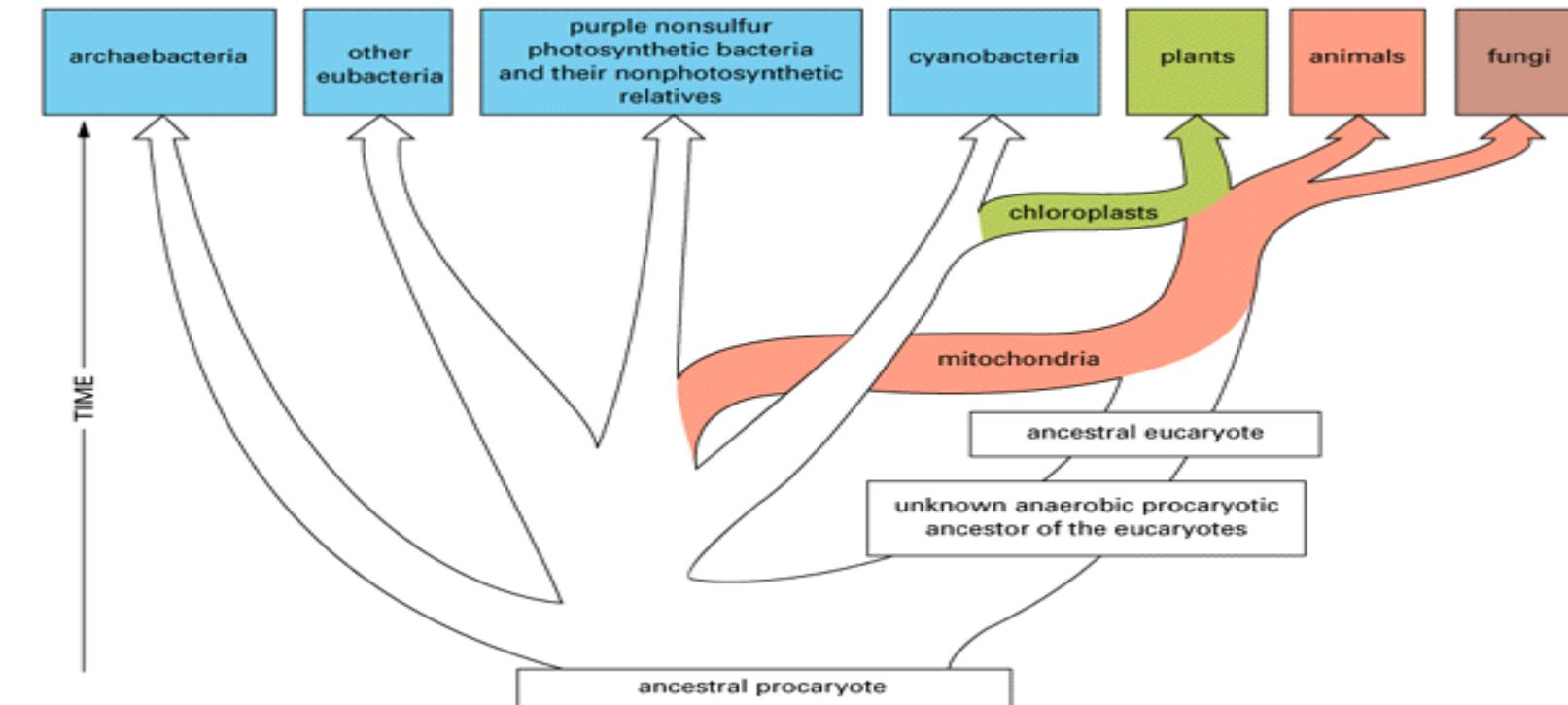
Implications of the creation of the mitochondria

- ❖ Acquisition of mitochondria must have had many repercussions!
- ❖ Plasma membrane is heavily committed to energy metabolism in prokaryotic cells but not in eukaryotic cells, where this crucial function has been relegated to the mitochondria
 - ❖ Separation of functions left the eukaryotic plasma membrane free to evolve important new features.
- ❖ Because eukaryotic cells need not maintain a large H⁺ gradient across their plasma membrane, as required for ATP production in prokaryotes, it became possible to control changes in ion permeability of the plasma membrane for cell-signaling.
 - ❖ A variety of ion channels appeared in the eukaryotic plasma membrane.
 - ❖ Today, these channels mediate elaborate electrical signaling processes in higher organisms - notably in the nervous system - and they control much of the behavior of single-celled free-living eukaryotes such as protozoa.

More organelles of the eukaryote cell

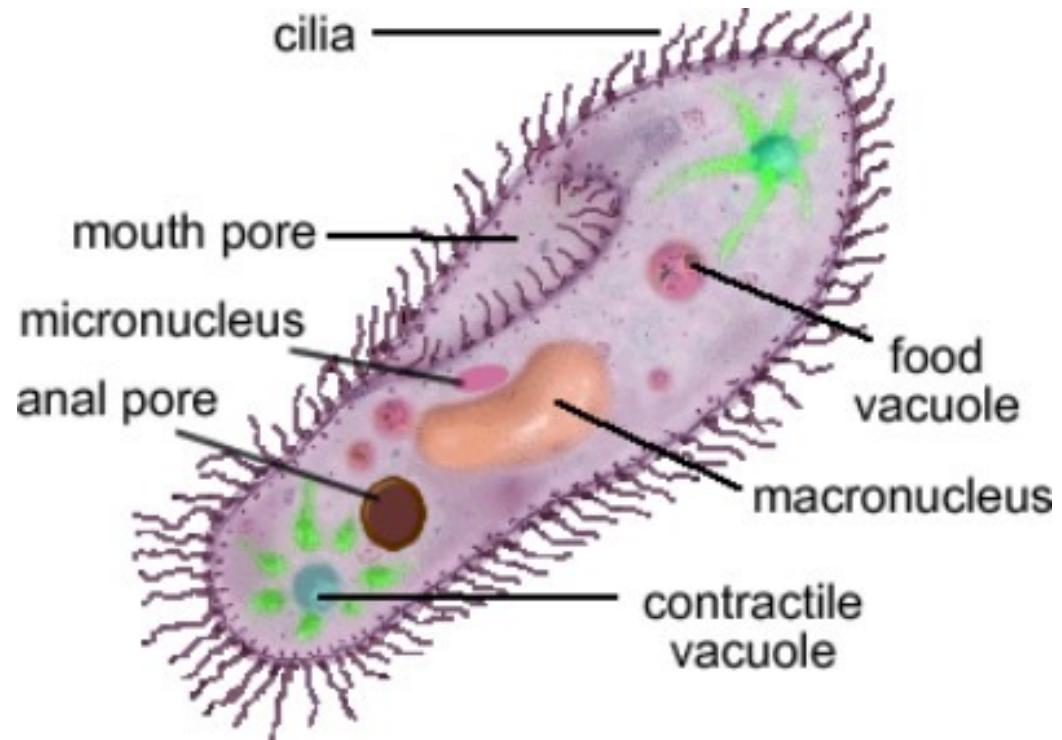


Postulated origin of the eukaryotic cell



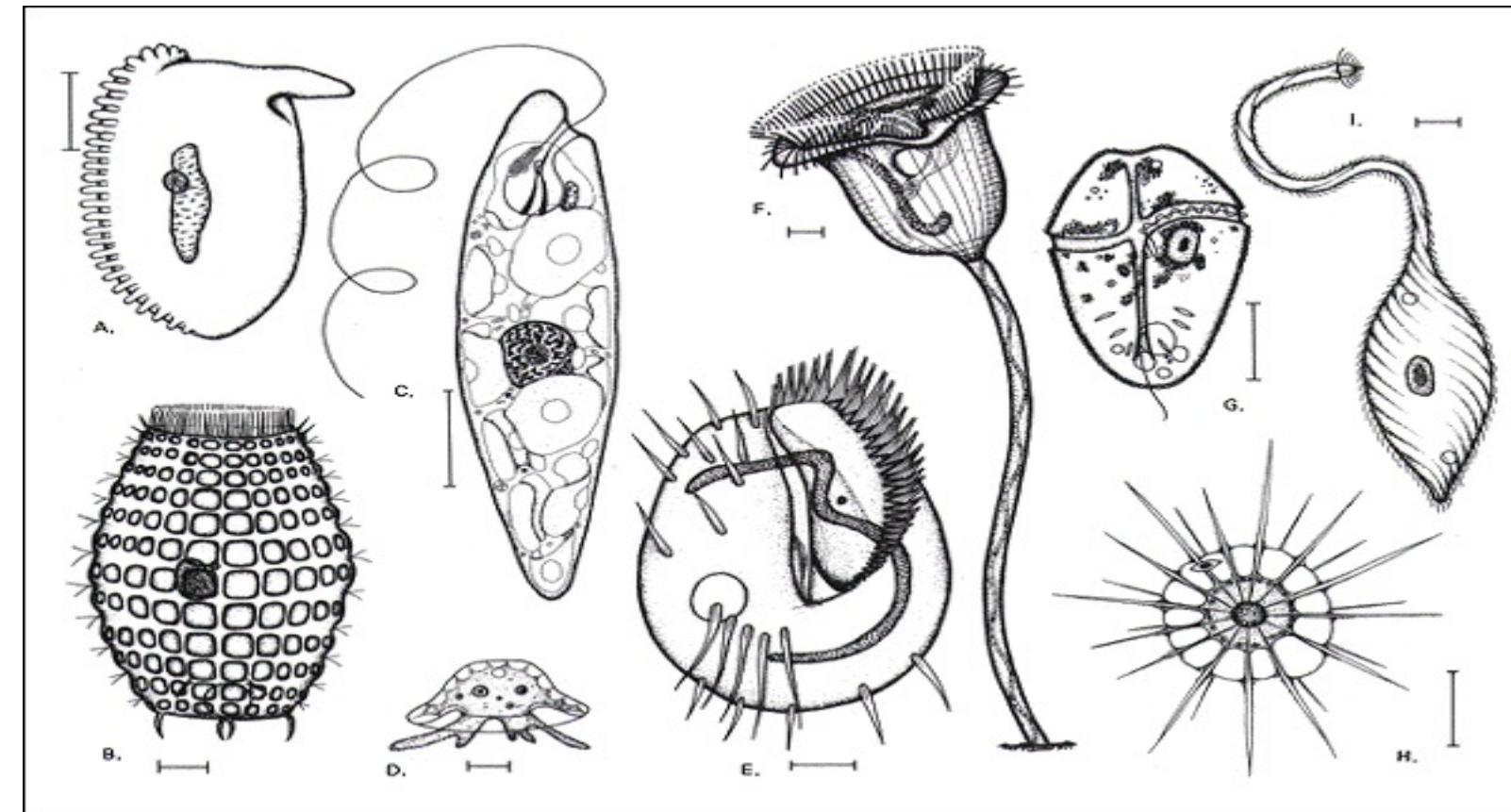
Complexity of the protest – a single-celled creature

- ❖ The complexity that can be achieved by a single eucaryotic cell is nowhere better illustrated than in the free-living, single-celled eucaryotes known as protists
 - ❖ Evolutionarily diverse
 - ❖ Exhibit a bewildering variety of different forms and
 - ❖ Behaviors: photosynthetic, carnivorous, motile, sedentary
 - ❖ Anatomy: complex and includes structures as: sensory bristles
 - ❖ photoreceptors, flagella, leg-like appendages, mouth parts, stinging darts, muscle like contractile bundles



<https://www.youtube.com/watch?v=0-6dzU4gOJo>

The length-scale of protists



These drawings are done to different scales, but in each case the bar denotes 10 mm. The organisms in (A), (B), (E), (F), and (I) are ciliates; (C) is an euglenoid; (D) is an amoeba; (G) is a dinoflagellate; (H) is a heliozoan.

From M.A. Sleigh, *The Biology of Protozoa*. London: Edward Arnold, 1973.

Summary

1. Order and self-organization
 - a) Physical, chemical, small molecules with structural information
2. Association certain molecules according
 - a) RNA and proteins, later DNA
3. Isolation of these specialized molecules from the environment
 - a) LUCA
 - b) The central dogma DNA → RNA → Proteins
4. Development of metabolic reactions to utilize specific nutrient molecules
5. Oxygen synthesizing cells to oxygen utilizing cells
6. Prokaryotes to eukaryotes
7. Unicellular organisms

Acknowledgements:

Molecular Biology of the Cell. Fifth Edition. 2007. Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter

Cellular Assemblies

Lecture 5

Acknowledgements:

Alberts - Molecular Biology of the Cell

Scitable by Nature Education

Nature Resources

Internet resources



OBJECTIVE OF THE LECTURE

1. Understand the meaning of self-assembly
 - a) Some examples of artificially created self assembled structures
2. Self assembled structures in cells
3. PDB structure
4. Understand the principles Protein folding

Definitions of Self Assembly & Self Organization

- *Self organization is a process in which pattern at the global level of the system emerges solely from the numerous interactions among the lower-level components of the system. Moreover, the rules specifying interactions among the system's components are executed using only local information without reference to the global pattern ... Camazine et al, 2003*

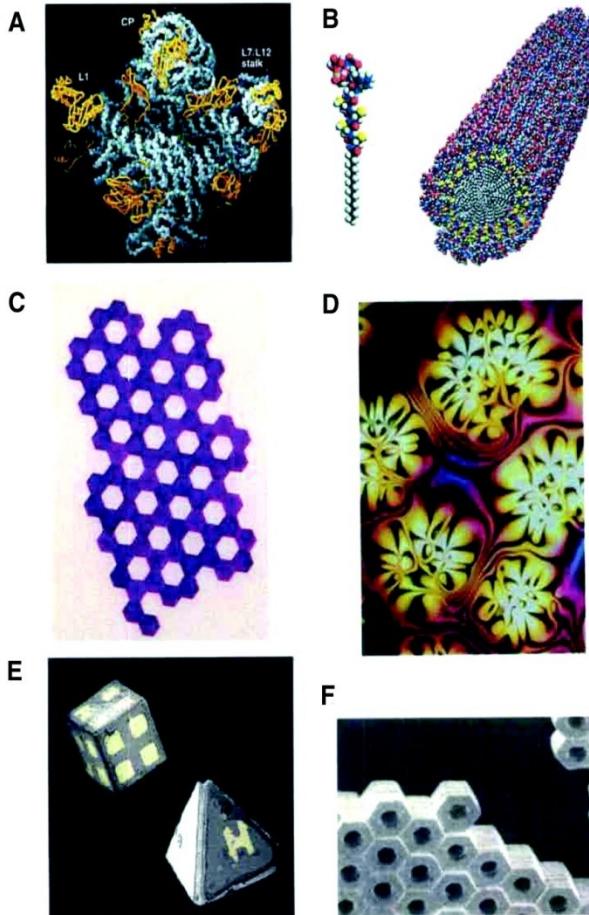
ON THE OTHER HAND

- Self-assembly is the fundamental principle which generates structural organization on all scales from molecules to galaxies. It is defined as reversible processes in which pre-existing parts or disordered components of a preexisting system, form structures or patterns



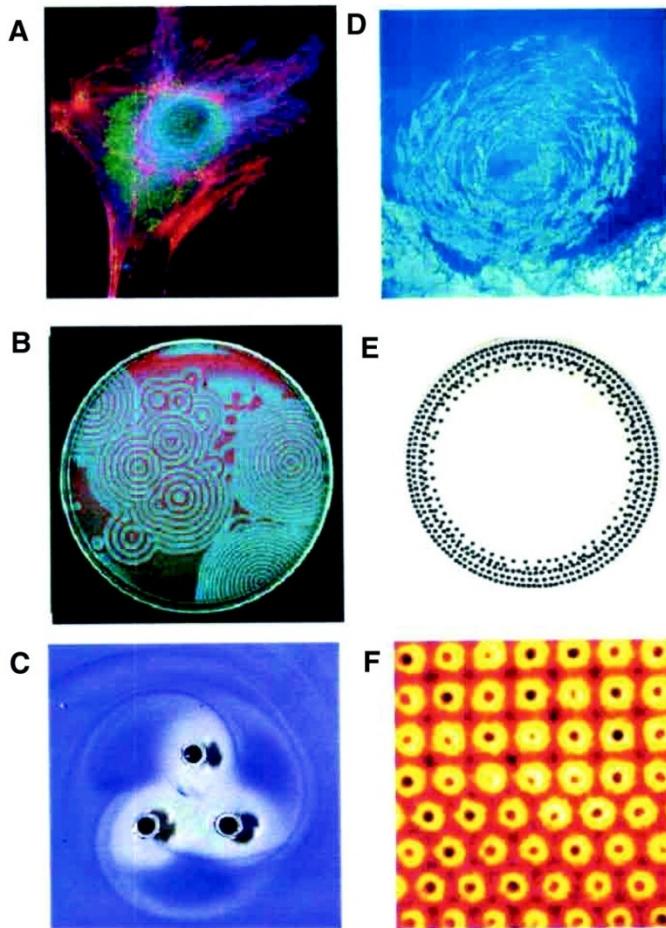
Flocks of Starlings

Examples of Self assembly



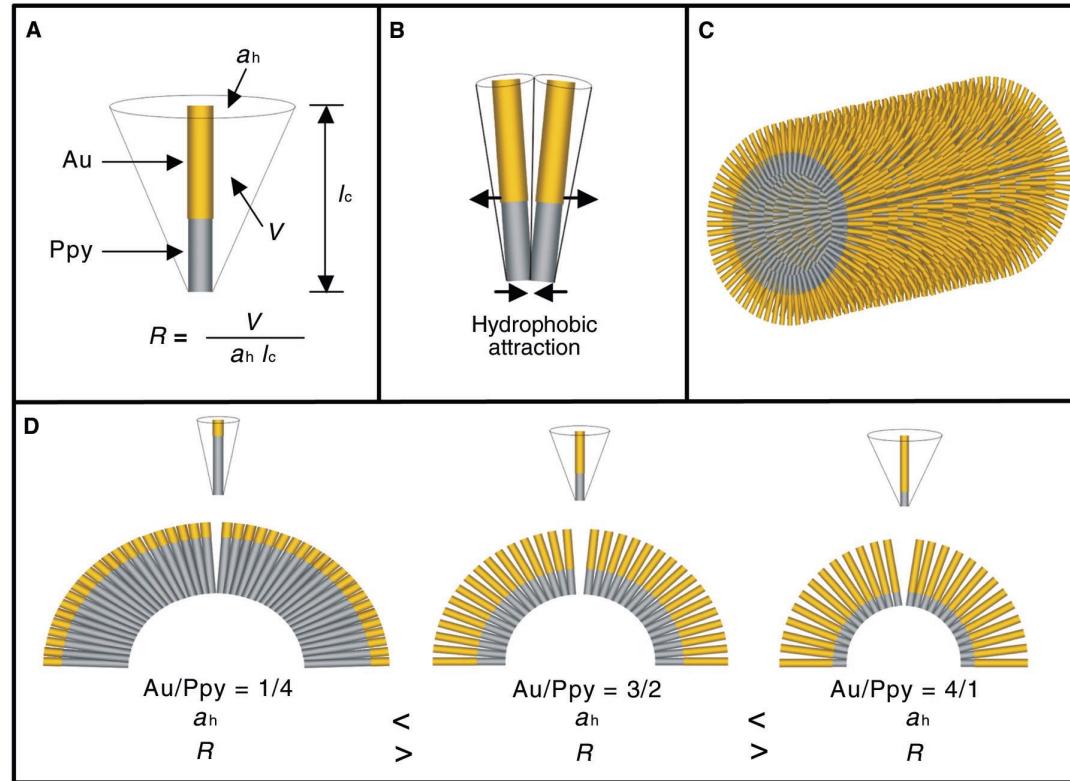
- A. Crystal structure of a ribosome.
- B. Self-assembled peptide-amphiphile nanofibers.
- C. An array of millimeter-sized polymeric plates assembled at a water/perfluorodecalin interface by capillary interactions.
- D. Thin film of a nematic liquid crystal on an isotropic substrate.
- E. Micrometer-sized metallic polyhedra folded from planar substrates.
- F. A three-dimensional aggregate of micrometer plates assembled by capillary forces.

Examples of Self Organization



- A. An optical micrograph of a cell with fluorescently labeled cytoskeleton and nucleus; microtubules (~24 nm in diameter) are colored red.
- B. Reaction-diffusion waves in a Belousov-Zabatinski reaction in a 3.5-inch Petri dish.
- C. A simple aggregate of three millimeter-sized, rotating, magnetized disks interacting with one another via vortex-vortex interactions.
- D. A school of fish.
- E. Concentric rings formed by charged metallic beads 1 mm in diameter rolling in circular paths on a dielectric support.
- F. Convection cells formed above a micropatterned metallic support. The distance between the centers of the cells is ~2 mm.

Self-Assembly of Mesoscopic Metal-Polymer Amphiphiles



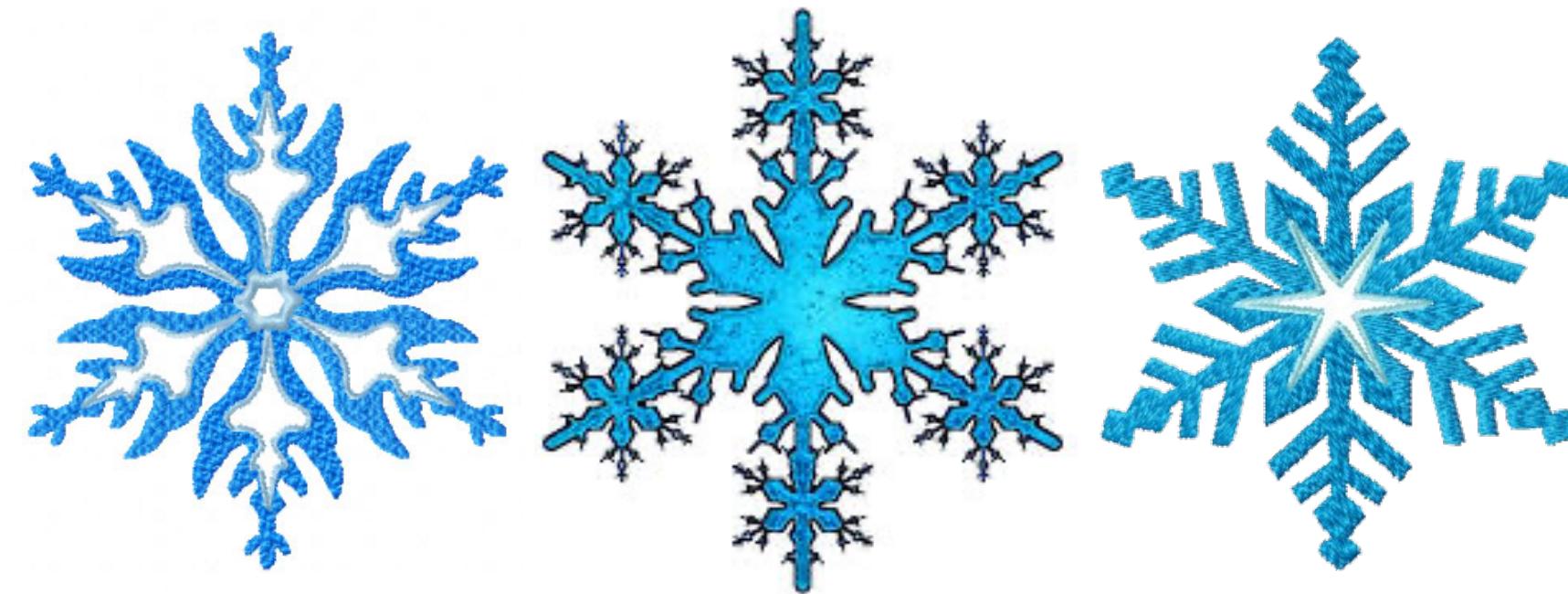
Hard hydrophilic domain is an inorganic material such as gold, and the soft domain is a hydrophobic conducting polymer such as oxidized polypyrrole, which can be electrochemically polymerized within the confines of an alumina template

Self assembly of gold-polypyrrole rods

❖ Au block diameter was 400 (30) nm and the polypyrrole block diameter was 360 (25) nm. These structures self-organize into mesoscopic architectures with unusual structures, including bundles, tubes of varying diameters, and sheets

Science 303, 348 (2004);
Sungho Park et al.⁶

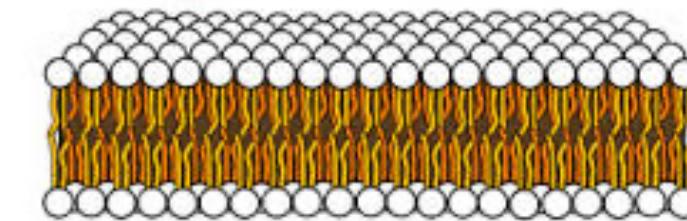
Self assembly of snowflakes



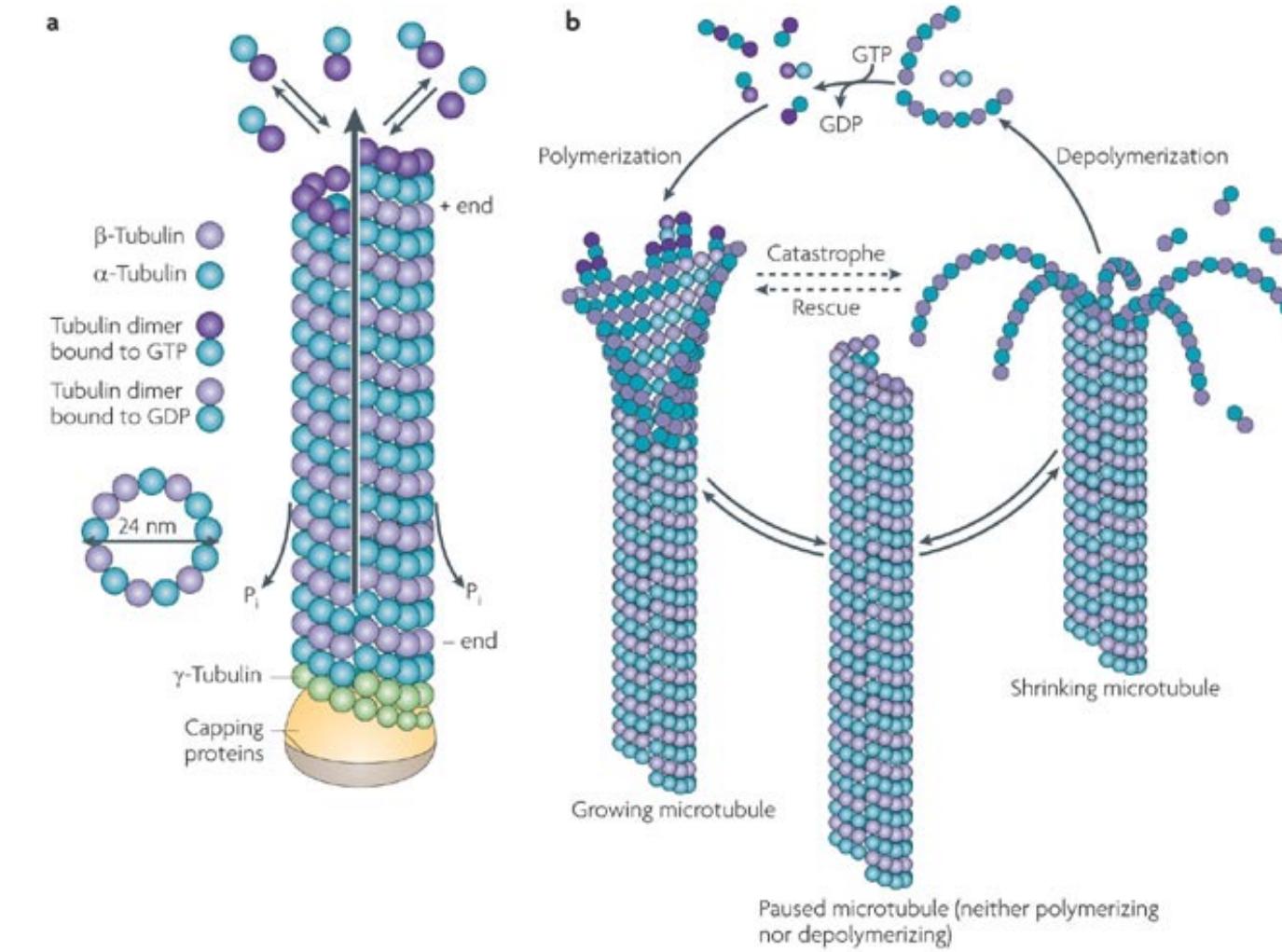
[Snowflakes BBC](#)

Bilayer Self Assembly

[Bilayer sheet self assembly video](#)

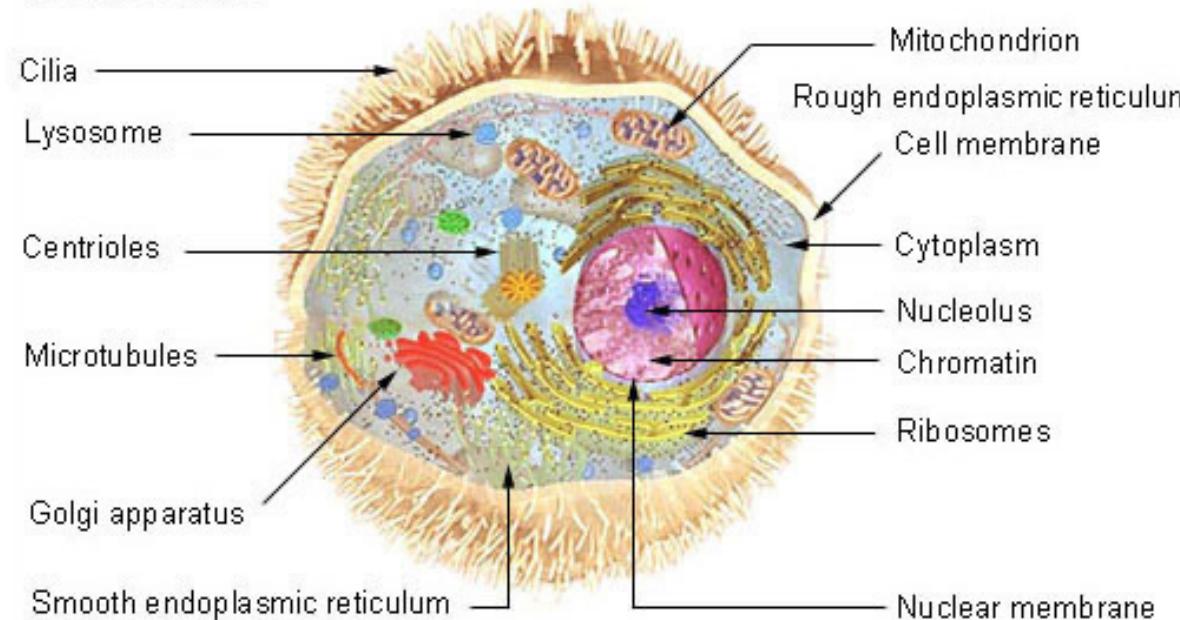


Tubulin and Microtubules



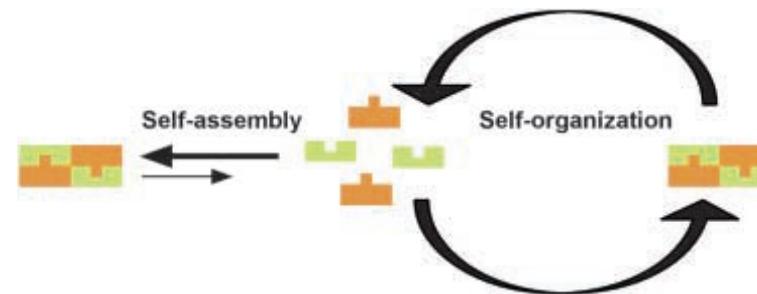
Self Assembly and self organization of internal cellular structures

Cell Structure



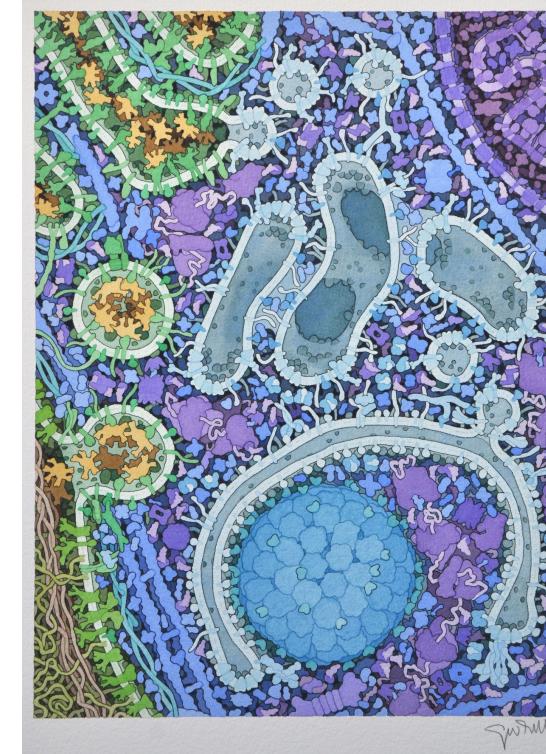
Cellular Components

- i. Nucleus
- ii. Golgi complex
- iii. Actin filaments for cellular structure
- iv. Autophagosomes
- v. Vesicles

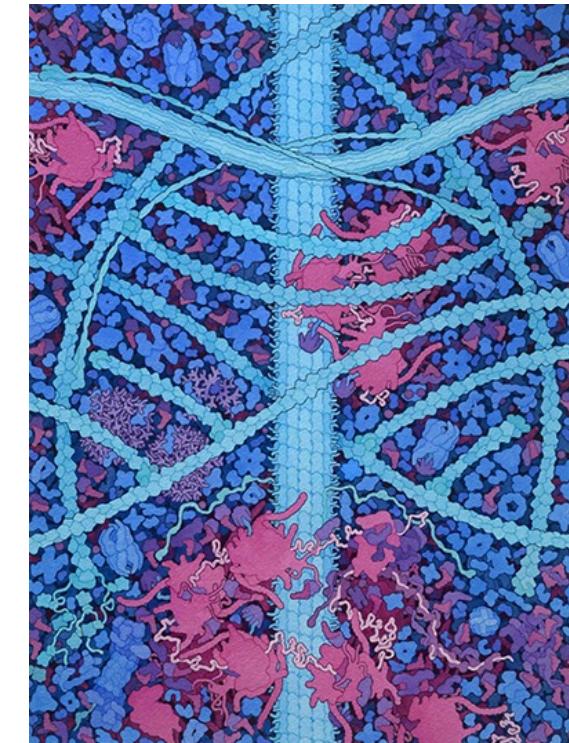


Images from the RCSB site

Autophagy



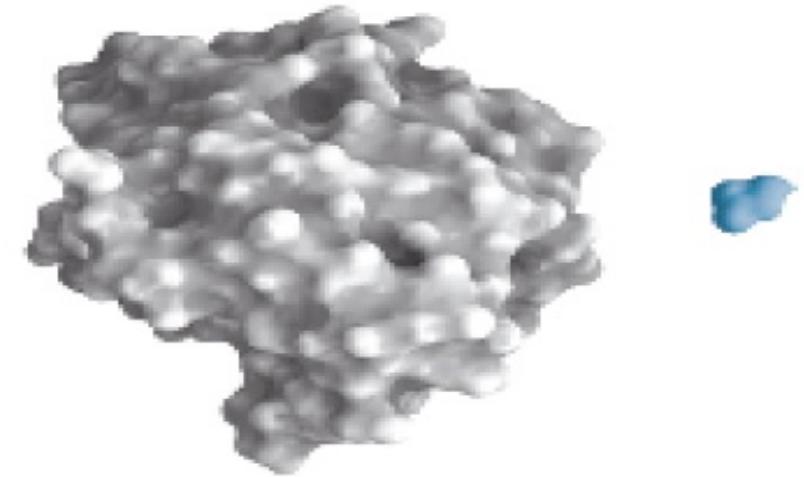
Cytoskeleton



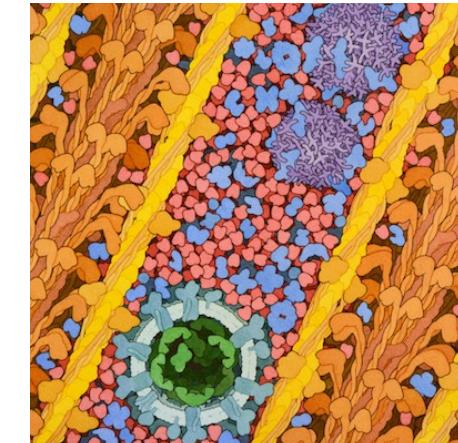
<https://ccsb.scripps.edu/goodsell/>

Protein structure and function

- ❖ Proteins can assume an unlimited number of configurations and yet possess a specific chemical and structural function
- ❖ The known three-dimensional structures of proteins are archived in the Protein Data Bank, or PDB (www.rcsb.org/pdb)
- ❖ The data from the PDB files provide only a series of coordinates detailing the location of atoms and their connectivity



Structure of the enzyme chymotrypsin



Myoglobin in a Whale Muscle Cell

Protein Folding – Secondary Structures

Lecture 6

Acknowledgements:

Alberts - Molecular Biology of the Cell

Scitable by Nature Education

Nature Resources

Internet resources

OBJECTIVE OF THE LECTURE

1. Understand the principles Protein folding
2. Phi-psi angles
3. Ramachandran plots
4. Secondary Protein structures

3-D structure of Proteins

1. the three-dimensional structure of a protein is determined by its amino acid sequence
2. the function of a protein depends on its structure
3. an isolated protein usually exists in one or a small number of stable structural forms
4. the most important forces stabilizing the specific structures maintained by a given protein are noncovalent interactions

Stability of protein structures

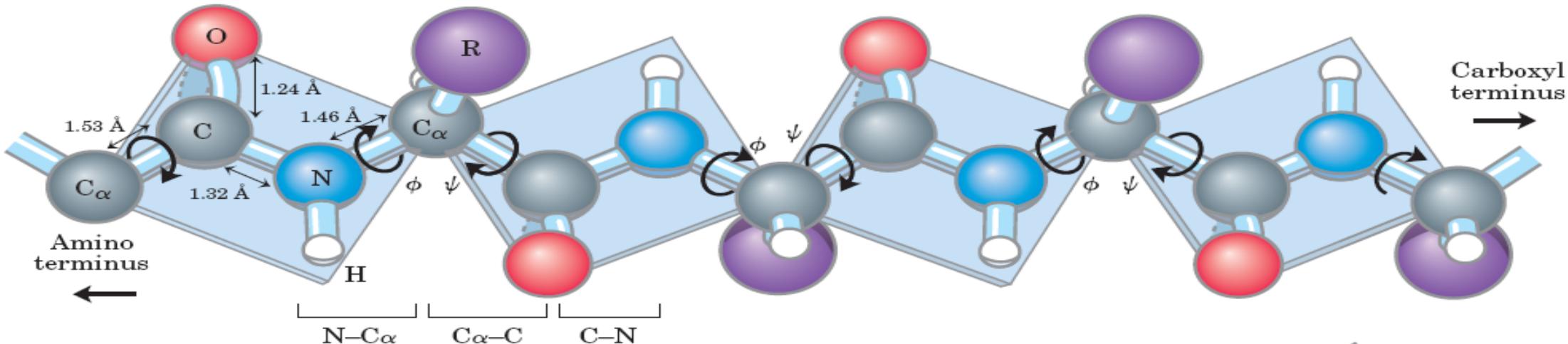
- ❖ The stability of protein structures depends on weak interactions
- ❖ It requires 200-460 kJ/mol to break a single covalent bond, compared to 4-30kJ/mol for weak interactions
 - ❖ The weak interactions predominate because they are numerous
- ❖ The free energies of the folded and un-folded states are similar

Governing Equations

$$\Delta G = \Delta H - \Delta TS$$

$$\Delta G = -RT\ln k$$

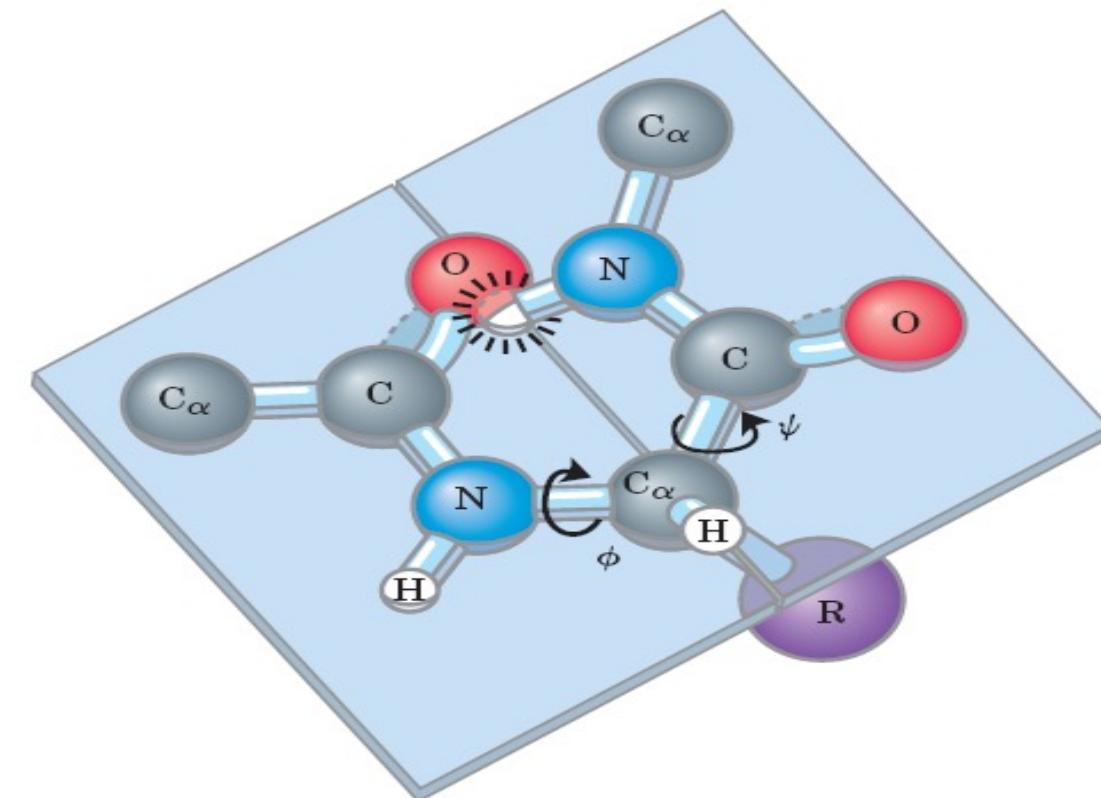
The phi and psi angles



- ❖ Three bonds separate sequential α carbons in a polypeptide chain. The $N-C_{\alpha}$ and $C-C_{\alpha}$ bonds can rotate, with bond angles designated ϕ and ψ , respectively
- ❖ The peptide $C-N$ bond is not free to rotate
- ❖ Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. In the conformation shown, ϕ and ψ are 180 deg (or -180 deg).
- ❖ As one looks out from the α -carbon, the ϕ and ψ angles increase as the carbonyl or amide nitrogens (respectively) rotate clockwise

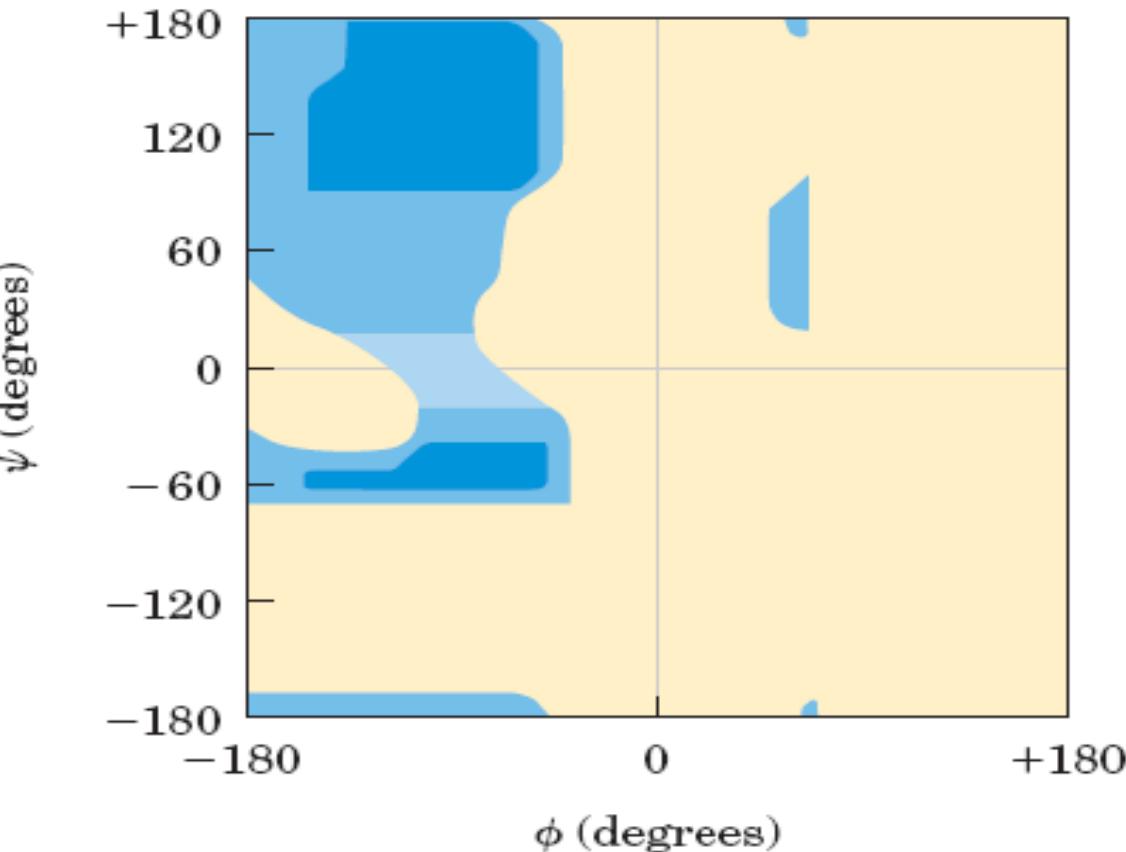
The zero phi & psi angles

- ❖ By convention, both ϕ and ψ are defined as 0 deg when the two peptide bonds flanking that carbon are in the same plane and positioned as shown.
- ❖ In a protein this conformation is prohibited by steric overlap between an – carbonyl oxygen and an – amino hydrogen atom
- ❖ To illustrate the bonds between atoms, the balls representing each atom are smaller than the van der Waals radii for this scale. $1 \text{ \AA} = 0.1 \text{ nm}$.



The Ramachandran Plot for L-Ala

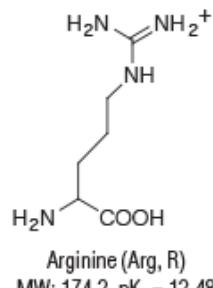
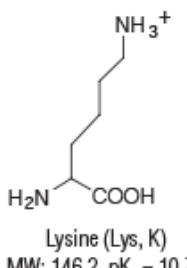
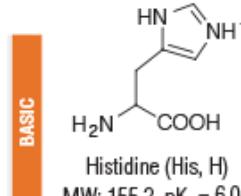
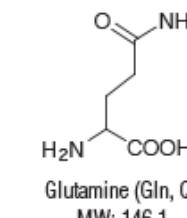
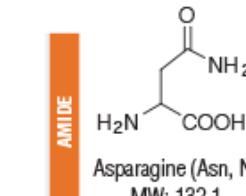
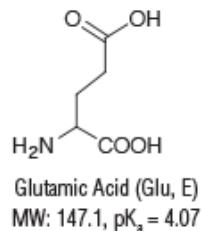
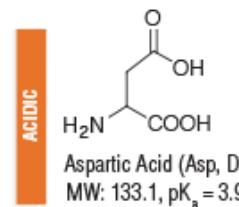
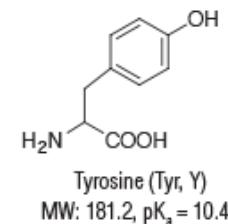
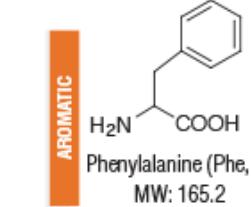
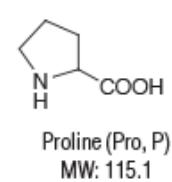
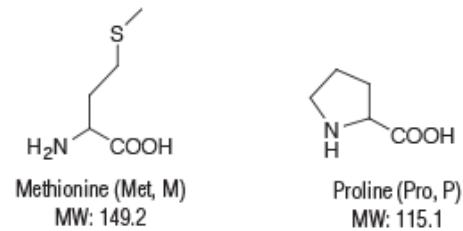
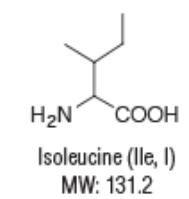
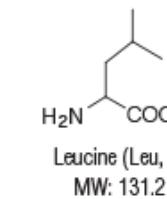
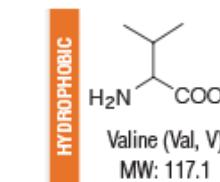
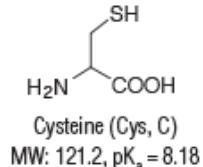
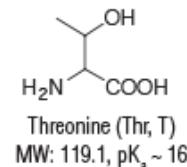
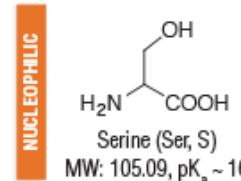
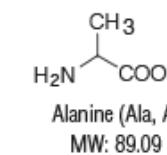
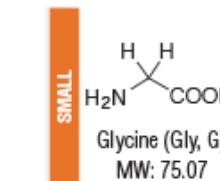
- ❖ Conformations deemed possible are those that involve little or no steric interference, based on calculations using known van der Waals radii and bond angles.
 - ❖ The areas shaded dark blue reflect conformations that involve no steric overlap and thus are fully allowed
 - ❖ medium blue indicates conformations allowed at the extreme limits for unfavorable atomic contacts
 - ❖ the lightest blue area reflects conformations that are permissible if a little flexibility is allowed in the bond angles
- ❖ The asymmetry of the plot results from the L stereochemistry of the amino acid residues



About Ramachandran Plots

1. What kind of plot do you expect for other un-branched amino-acids?
2. What kind of plot do you expect for branched amino-acids (eg Ile)?
3. What kind of plot do you expect for glycine?
4. What kind of plot do you expect for Proline?

Structure of Amino Acids



What are the possible applications of the RAMACHANDRAN PLOT?

PROTEIN STRUCTURE PREDICTION

ALPHA FOLD

<https://alphafold.ebi.ac.uk/>

SCFBio IIT Delhi

<http://www.scfbio-iitd.res.in>

PROTEIN STRUCTURE

Overview of Protein Structure

- ❖ Every protein has a three-dimensional structure that reflects its function.
- ❖ Protein structure is stabilized by multiple weak interactions. Hydrophobic interactions are the major contributors to stabilizing the globular form of most soluble proteins; hydrogen bonds and ionic interactions are optimized in the specific structures that are thermodynamically most stable.
- ❖ The nature of the covalent bonds in the polypeptide backbone places constraints on structure. The peptide bond has a partial double bond character that keeps the entire six-atom peptide group in a rigid planar configuration. The N–C_α and C_α–C bonds can rotate to assume bond angles of ϕ and ψ , respectively.

Classification of Structures

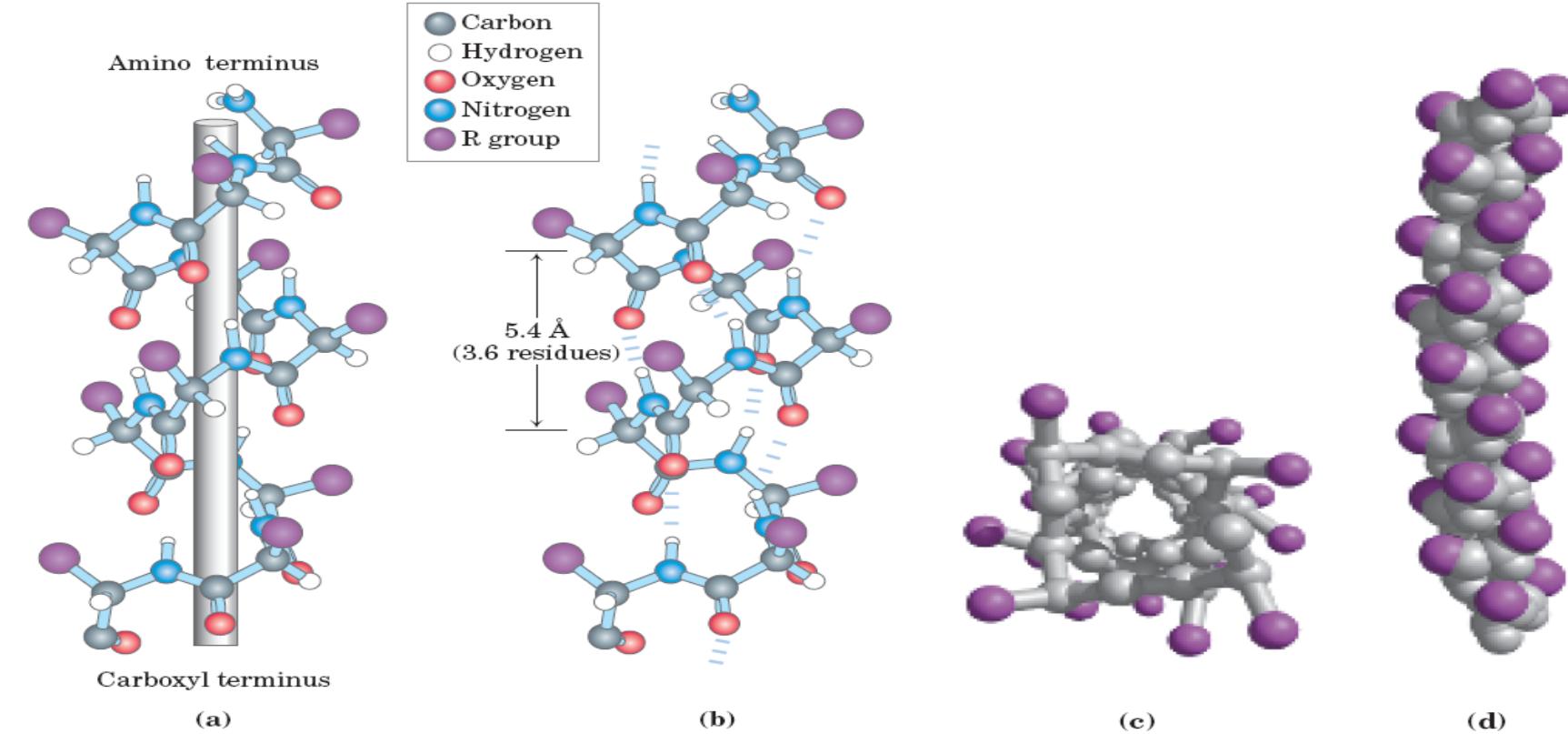
- ❖ Secondary
 - ❖ Alpha Helix
 - ❖ Beta sheets
- ❖ 3-D conformations
 - ❖ Structure and function
- ❖ Tertiary Structures
 - ❖ Globular proteins
 - ❖ Multimerics
 - ❖ Homo
 - ❖ hetero

Protein Secondary Structure

❖ The α -helix architecture

- ❖ Linus Pauling, Robert Corey
- ❖ X-ray results of William Astbury (1930) of proteins that make up porcupine quills (α -keratin)
 - ❖ Regular structure that repeats every 5.16 – 5.2 Å
 - ❖ polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues protrude outward from the helical backbone
 - ❖ The amino acid residues in an
 - ❖ helix have conformations with $\psi = -45$ to -50 deg and $\phi = -60$ deg

N
D
G
K
L
A
M
T
S
C
H
I
F
W
L
P
D
G
K
D
G

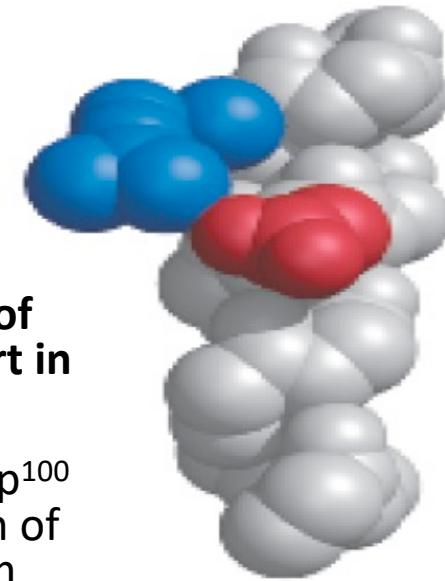


- Formation of a right-handed α -helix. The planes of the rigid peptide bonds are parallel to the long axis of the helix, depicted here as a vertical rod
- Ball-and-stick model of a right-handed α -helix, showing the intrachain hydrogen bonds. The repeat unit is a single turn of the helix, 3.6 residues
- The α -helix as viewed from one end, looking down the longitudinal axis
- Atoms in the center of the α -helix are in very close contact

Amino Acid Sequence Affects α -Helix Stability

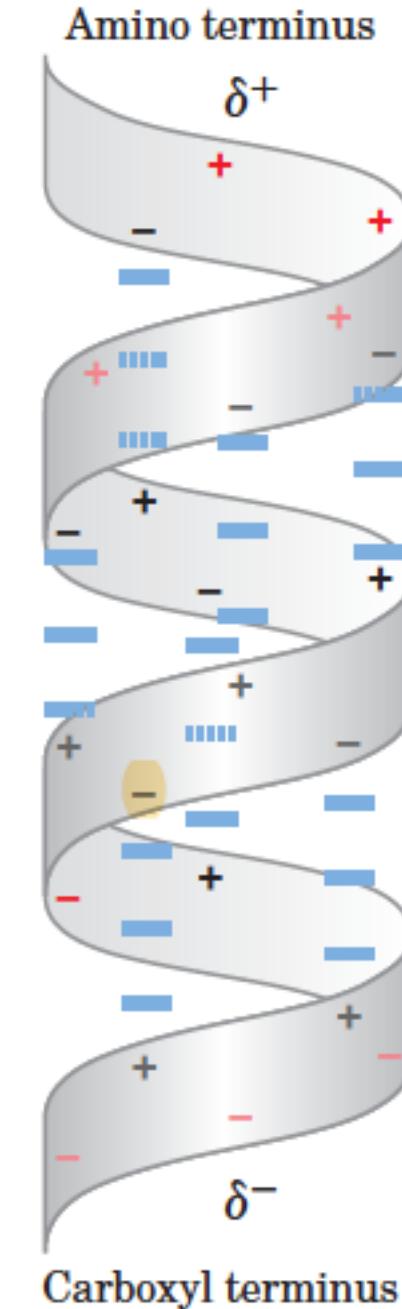
- ❖ If a polypeptide chain has a long block of Glu residues, this segment of the chain will not form an α -helix at pH 7.0.
 - ❖ The negatively charged carboxyl groups of adjacent Glu residues repel each other so strongly that they prevent formation of the α -helix
 - ❖ If there are many adjacent Lys and/or Arg residues, which have positively charged R groups at pH 7.0, they will also repel each other and prevent formation of the α -helix
 - ❖ The bulk and shape of Asn, Ser, Thr, and Cys residues can also destabilize an α -helix if they are close together in the chain
 - ❖ The twist of an α -helix ensures that critical interactions occur between an amino acid side chain and the side chain three (and sometimes four) residues away on either side of it. Positively charged amino acids are often found three residues away from negatively charged amino acids, permitting the formation of an ion pair

- ❖ Interactions between R groups of amino acids three residues apart in an α -helix
- ❖ An ionic interaction between Asp¹⁰⁰ and Arg¹⁰³ in an α -helical region of the protein troponin C, a calcium binding protein associated with muscle
- ❖ Polypeptide backbone (carbons, α -amino nitrogens, and α -carbonyl oxygens) is shown in gray for a helix segment 13 residues long
- ❖ The interacting Asp (red) and Arg (blue) side chains



Helix Dipole

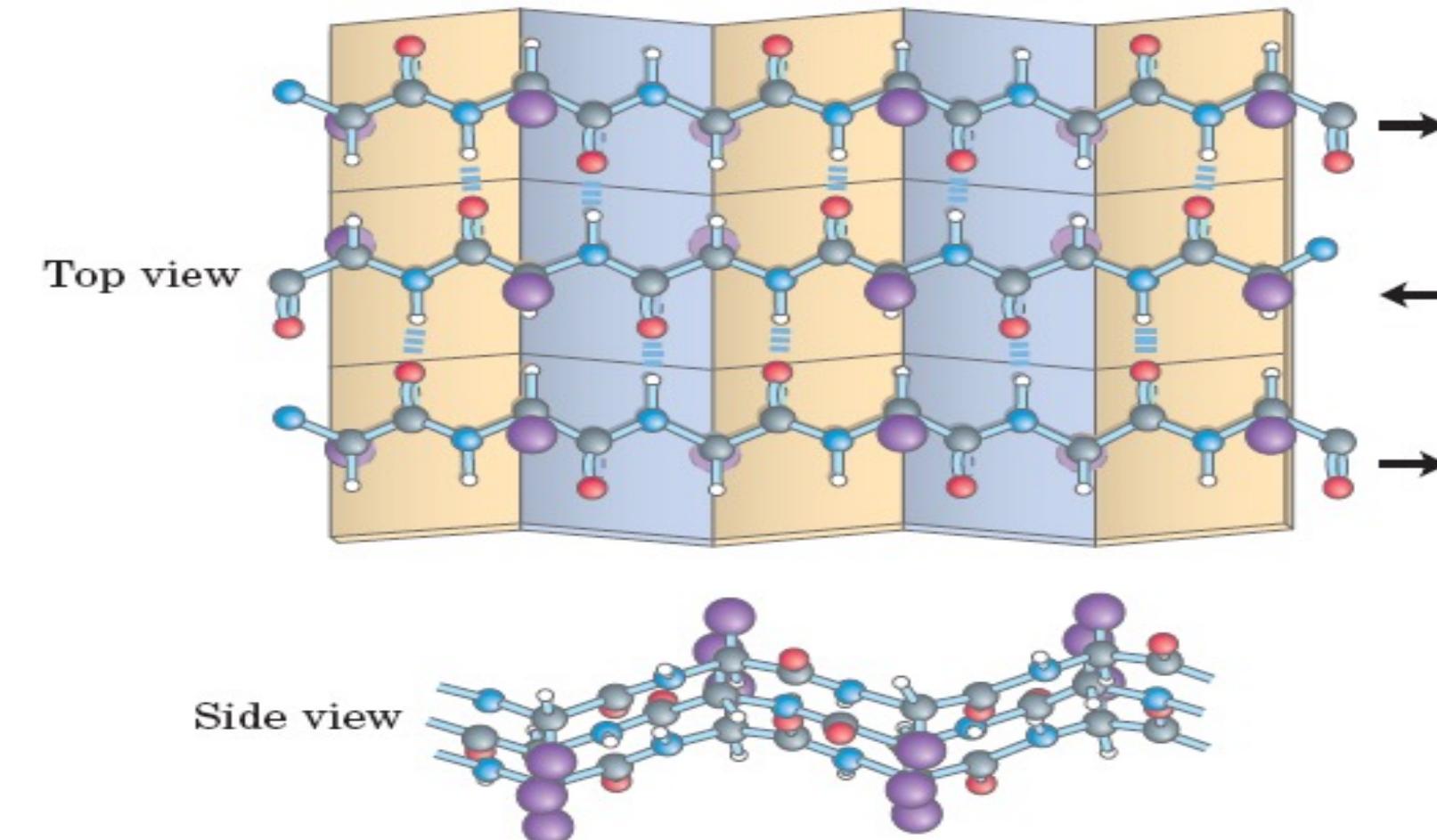
- ❖ The electric dipole of a peptide bond is transmitted along an α -helical segment through the intra-chain hydrogen bonds, resulting in an overall helix dipole
- ❖ The amino and carbonyl constituents of each peptide bond are indicated by + and - symbols
- ❖ Non-hydrogen bonded amino and carbonyl constituents in the peptide bonds near each end of the α -helical region are shown in red.



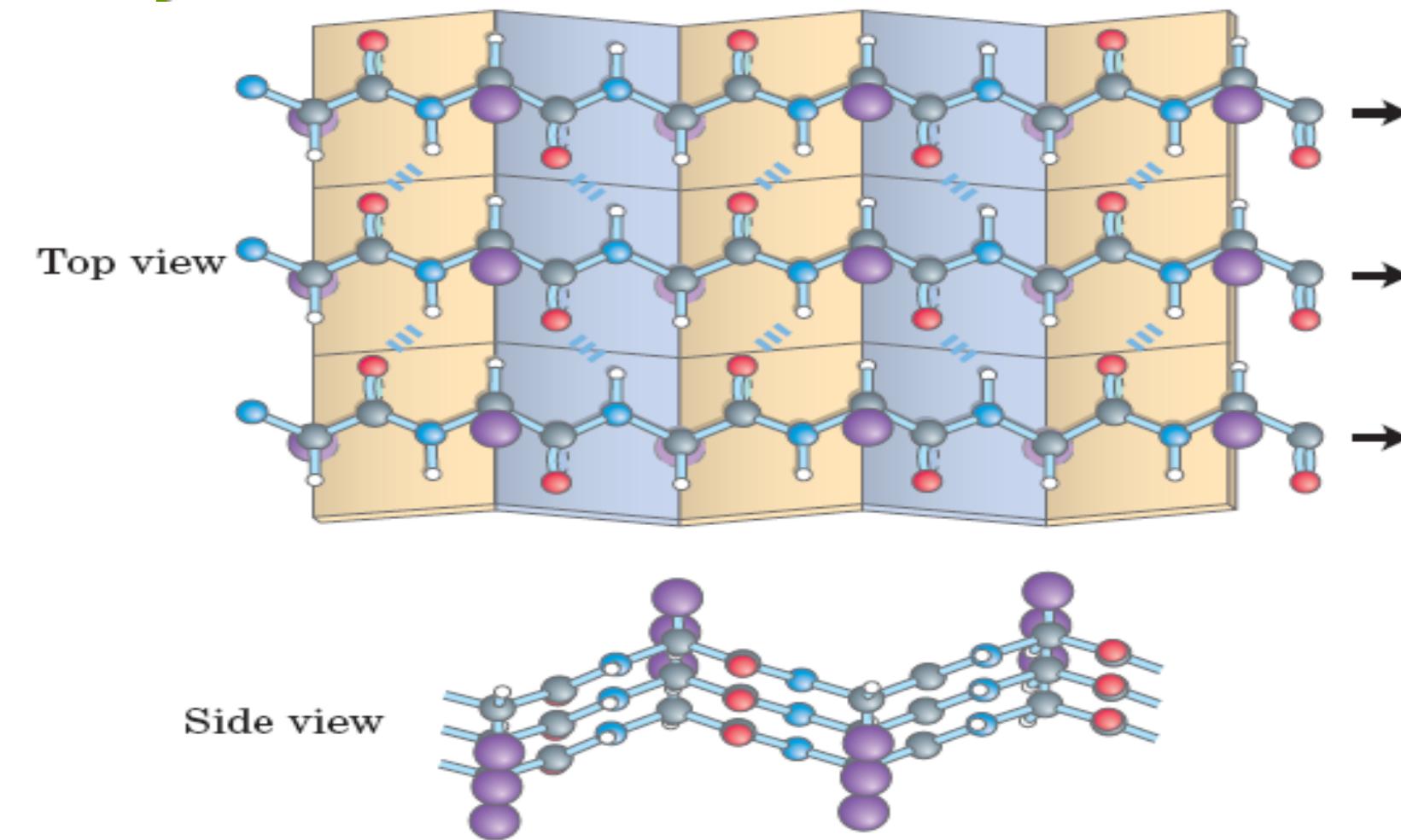
Constraints affecting stability of α -helix

1. The electrostatic repulsion (or attraction) between successive amino acid residues with charged R groups
2. The bulkiness of adjacent R groups
3. The interactions between R groups spaced three (or four) residues apart
4. The occurrence of Pro and Gly residues
5. The interaction between amino acid residues at the ends of the helical segment and the electric dipole inherent to the helix

Antiparallel β -Sheet Conformation



Parallel β -Sheet Conformation

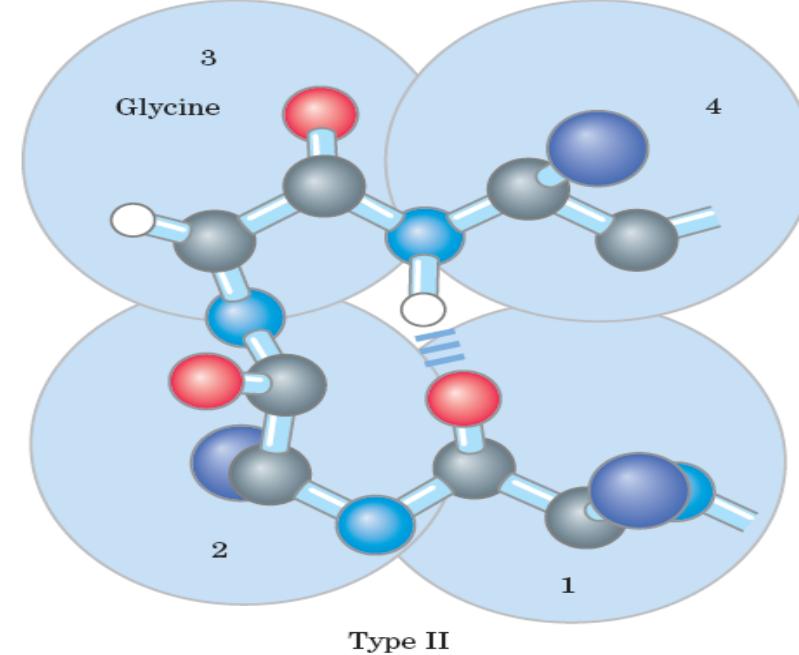
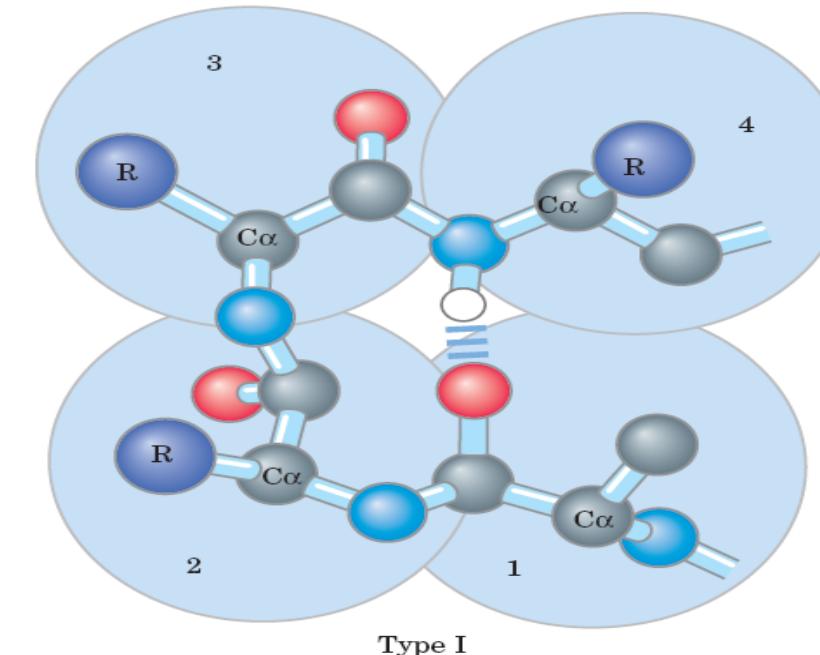


β -Sheet Conformation

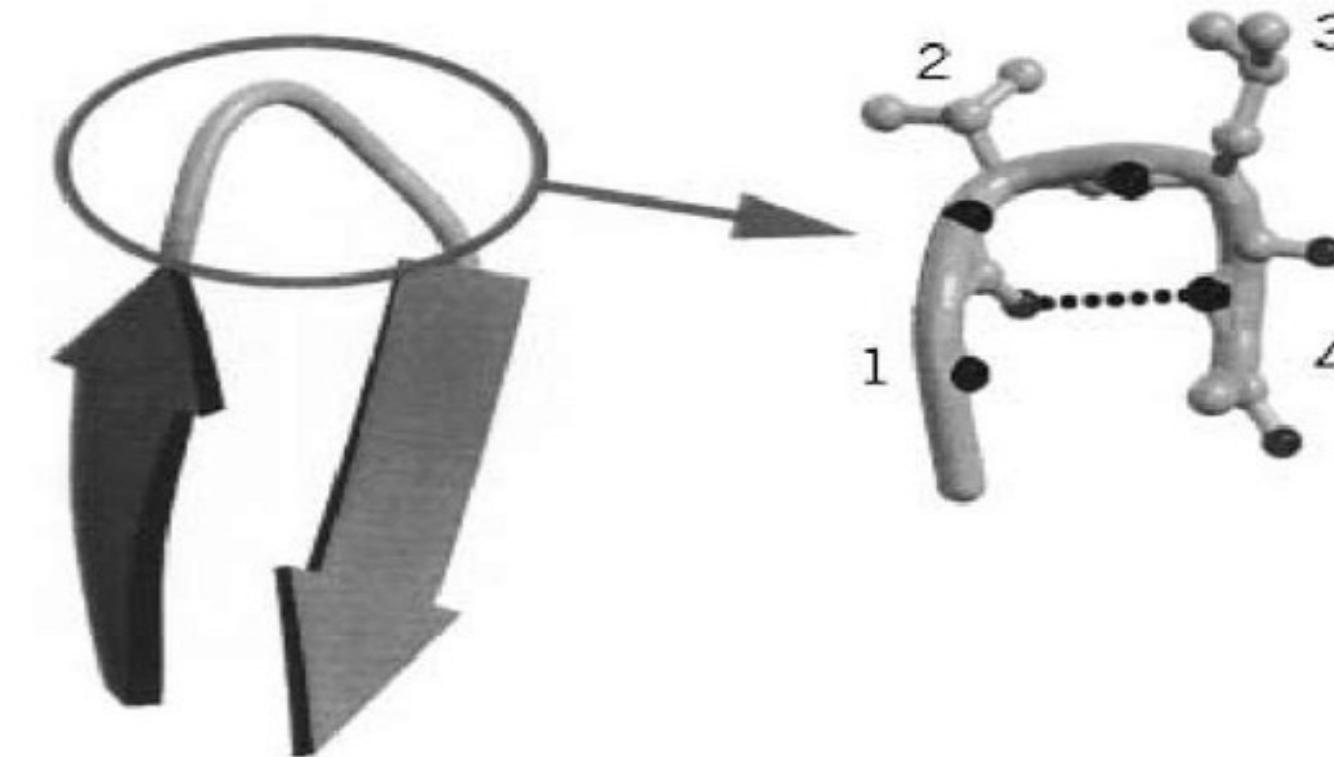
- ❖ The zigzag polypeptide chains can be arranged side by side to form a structure resembling a series of pleats
- ❖ Hydrogen bonds are formed between adjacent segments of polypeptide chain
- ❖ The adjacent polypeptide chains in a sheet can be either parallel or antiparallel (having the same or opposite amino-to-carboxyl orientations)
- ❖ The repeat period is shorter for the parallel conformation 6.5 Å, versus 7 Å for antiparallel

β -turns in protein architecture

- ❖ Turns that connect the ends of two adjacent segments of an antiparallel sheet. The structure is a 180 deg turn involving four amino acid residues, with the carbonyl oxygen of the first residue forming a hydrogen bond with the amino-group hydrogen of the fourth



Beta Turn



Bonding between the 1st and 3rd amino acid. The 3rd position is usually occupied by glycine or proline. In such positions, proline takes on a cis-orientation (which has only about 6% occurrence in proteins)

Protein Folding – Tertiary and Quarternary Structures

Lecture 7

Acknowledgements:

Alberts - Molecular Biology of the Cell
Scitable by Nature Education
Nature Resources
Internet resources



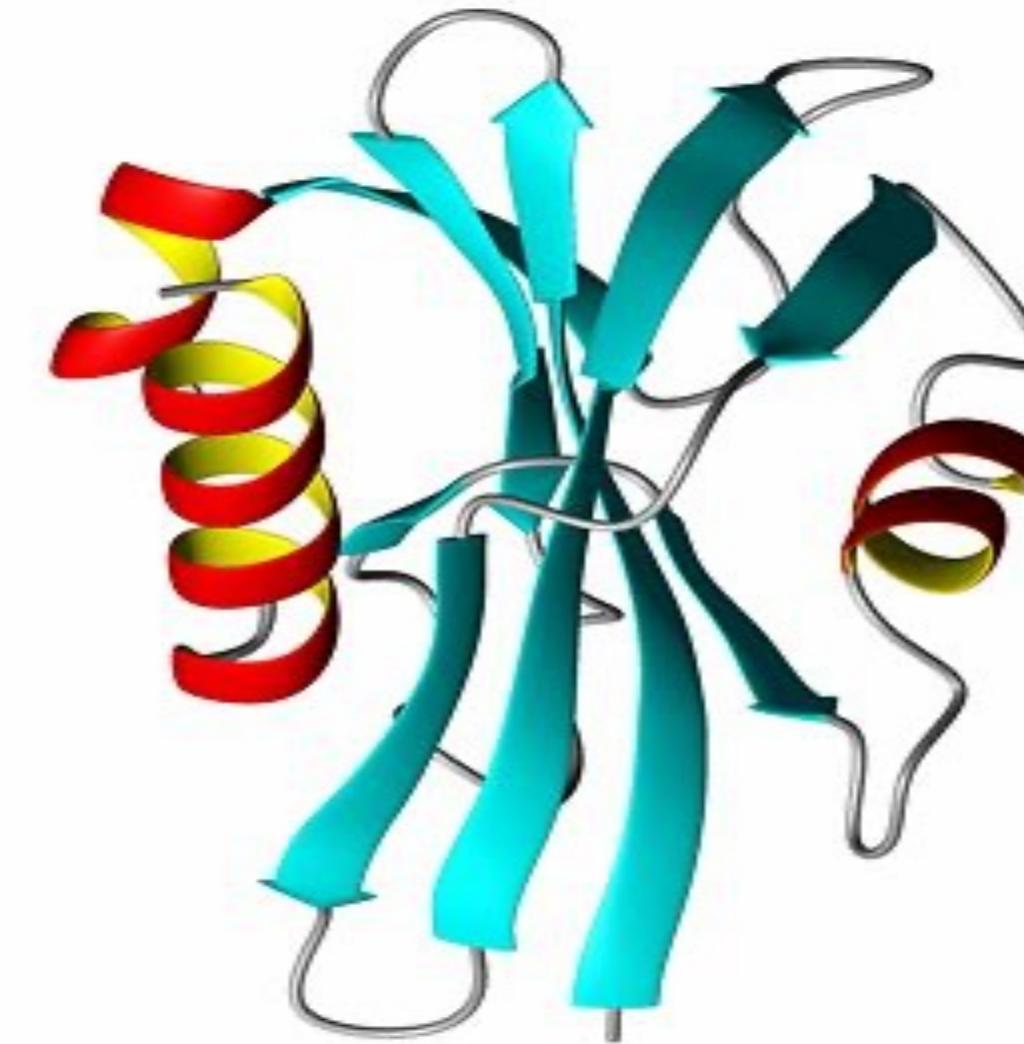
OBJECTIVE OF THE LECTURE

1. 3-D conformation – structure and function
2. Tertiary structures and fold symmetry
3. Protein Denaturation and re-folding
4. Thermodynamic of Protein Folding
5. Chaperone Proteins

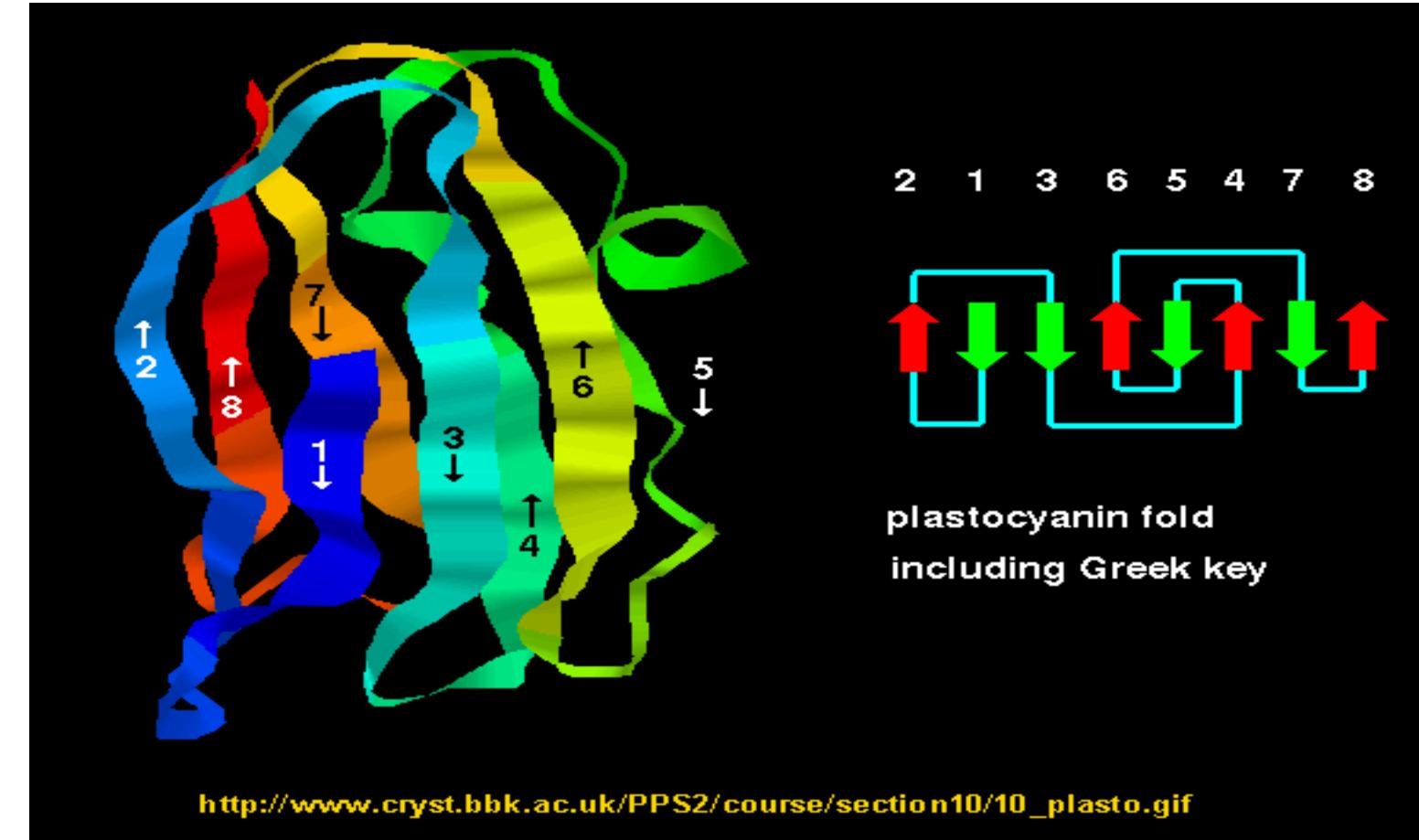
3-D structure of Proteins

1. the three-dimensional structure of a protein is determined by its amino acid sequence
2. the function of a protein depends on its structure
3. an isolated protein usually exists in one or a small number of stable structural forms
4. the most important forces stabilizing the specific structures maintained by a given protein are noncovalent interactions

Protein Structure of Beta Sheets and Alpha Helices

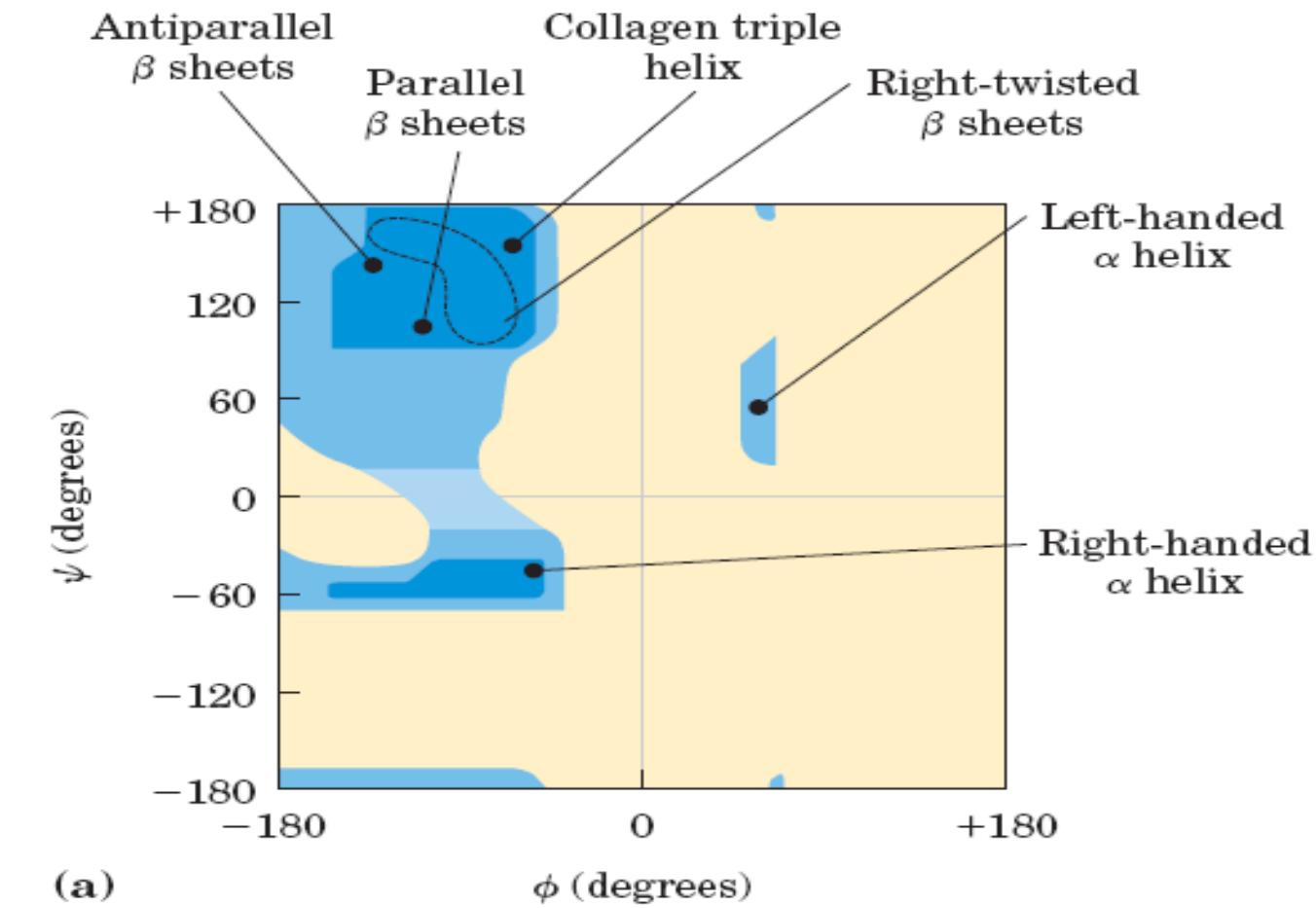


Greek Key Motif

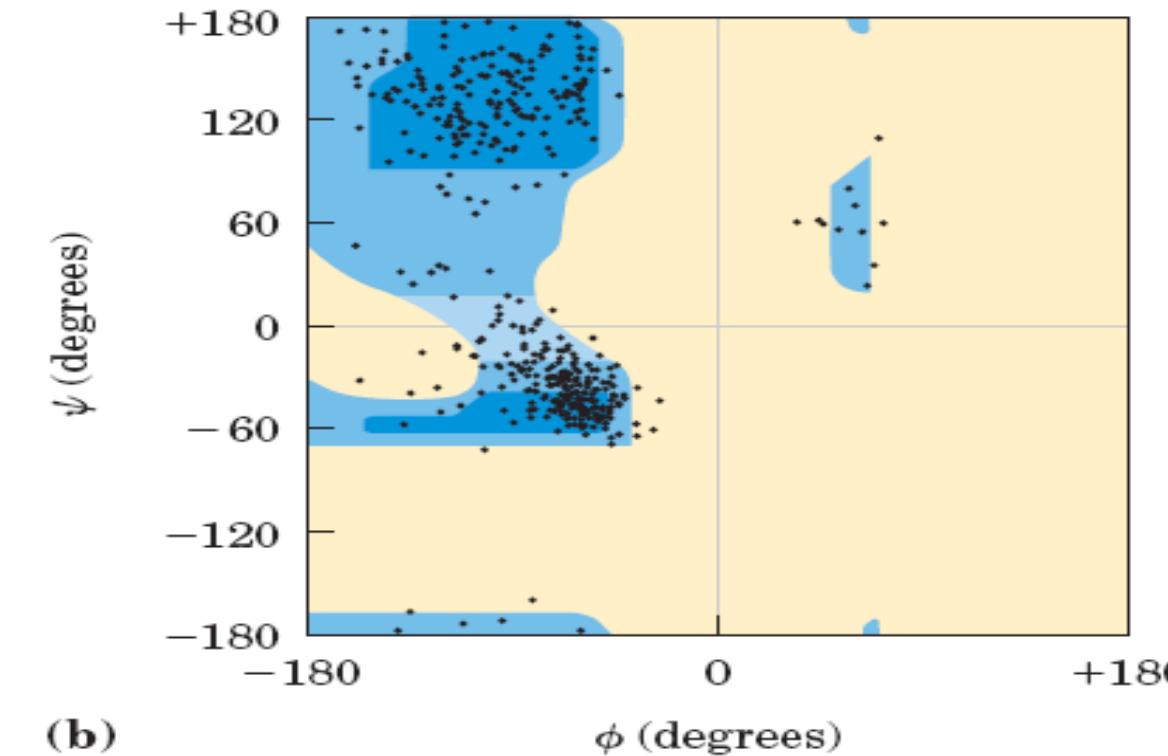


Secondary Protein Structure

- ❖ Secondary structure is the regular arrangement of amino acid residues in a segment of a polypeptide chain, in which each residue is spatially related to its neighbors in the same way
- ❖ The most common secondary structures are the alpha helix, the beta conformation, and beta turns
- ❖ The secondary structure of a polypeptide segment can be completely defined if the phi and psi angles are known for all amino acid residues in that segment



- ❖ The values of ϕ and ψ for various allowed secondary structures are overlaid
- ❖ Although left-handed helices extending over several amino acid residues are theoretically possible, they have not been observed in proteins



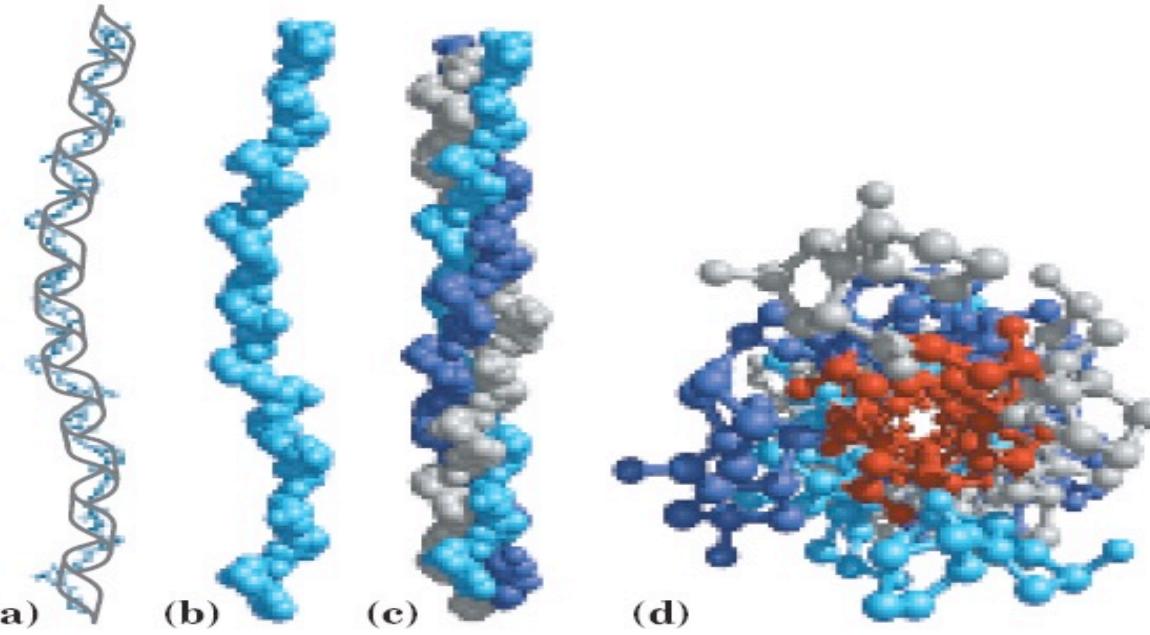
- ❖ All the amino acid residues except Gly in the enzyme pyruvate kinase (isolated from rabbit) are overlaid on the plot of theoretically allowed conformations. The small, flexible Gly residues were excluded because they frequently fall outside the expected ranges

Tertiary and Quaternary Structures

- ❖ The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's **tertiary structure**
- ❖ Some proteins contain two or more separate polypeptide chains, or subunits, which may be identical or different. The arrangement of these protein subunits in three-dimensional complexes constitutes **quaternary Structure**
 - ❖ In considering these higher levels of structure, it is useful to classify proteins into two major groups:
 - ❖ **fibrous proteins**, having polypeptide chains arranged in long strands or sheets, and
 - ❖ **globular proteins**, having polypeptide chains folded into a spherical or globular shape.

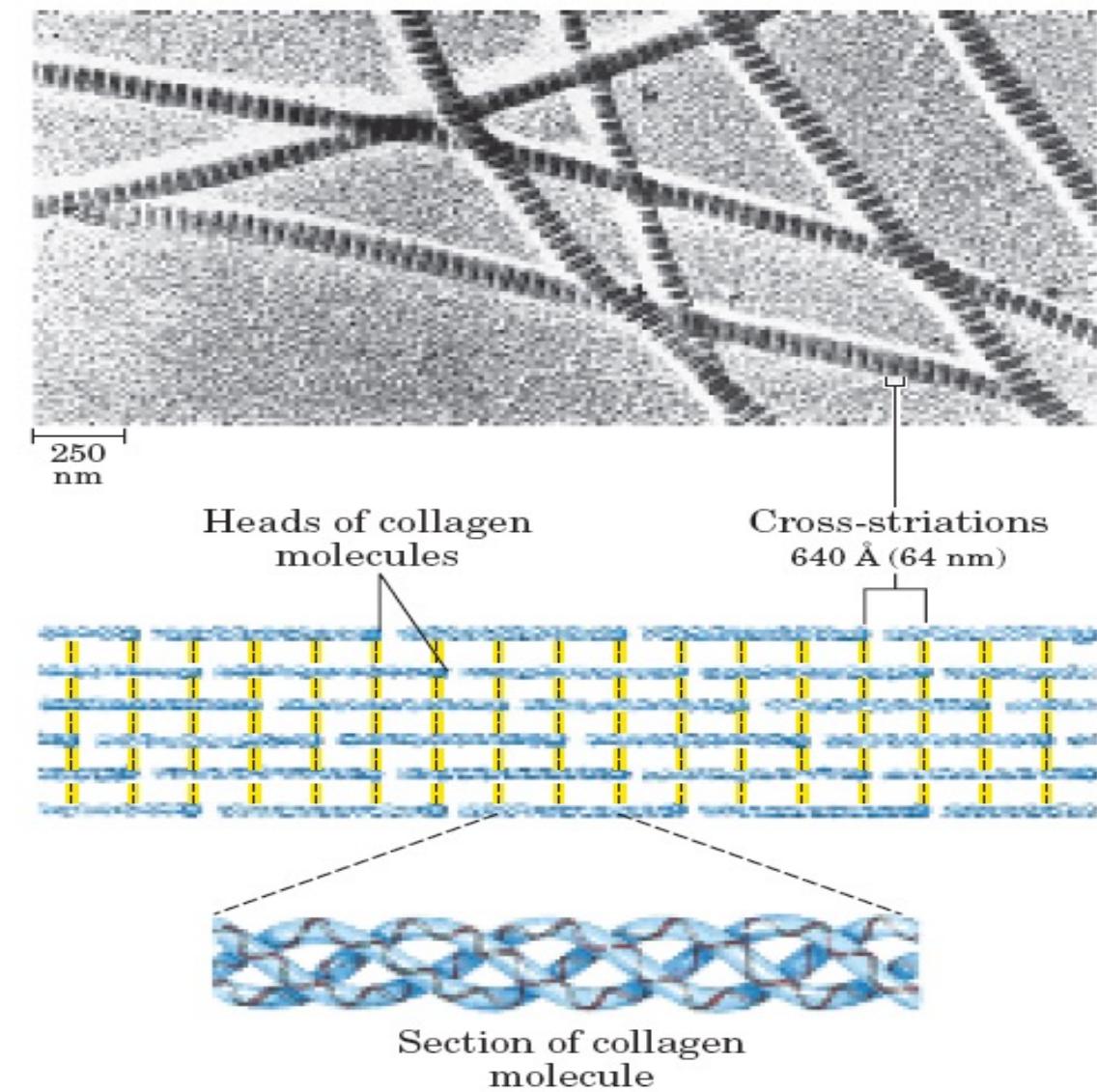
Fibrous Proteins: α -keratin, collagen

- ❖ Collagen: has evolved to provide strength. It is found in connective tissue such as tendons, cartilage, the organic matrix of bone, and the cornea of the eye
- ❖ Collagen is also a coiled coil, but one with distinct tertiary and quaternary structures: three separate polypeptides, called α -chains (not to be confused with α -helices), are supertwisted about each other
- ❖ The tight wrapping of the chains in the collagen triple α -helix provides tensile strength greater than that of a steel wire of equal cross-section



- a) The repeating tripeptide sequence Gly–X–Pro or Gly–X–4-Hyp adopts a left-handed helical structure with three residues per turn. The repeating sequence used to generate this model is Gly–Pro–4-Hyp
- b) Space-filling model of the same chain
- c) Three of these helices (shown here in gray, blue, and purple) wrap around one another with a right-handed twist
- d) The three-stranded collagen superhelix shown from one end, in a ball-and-stick representation. Gly residues are shown in red

Structure of Collagen



Collagen arrangement in a fish scale

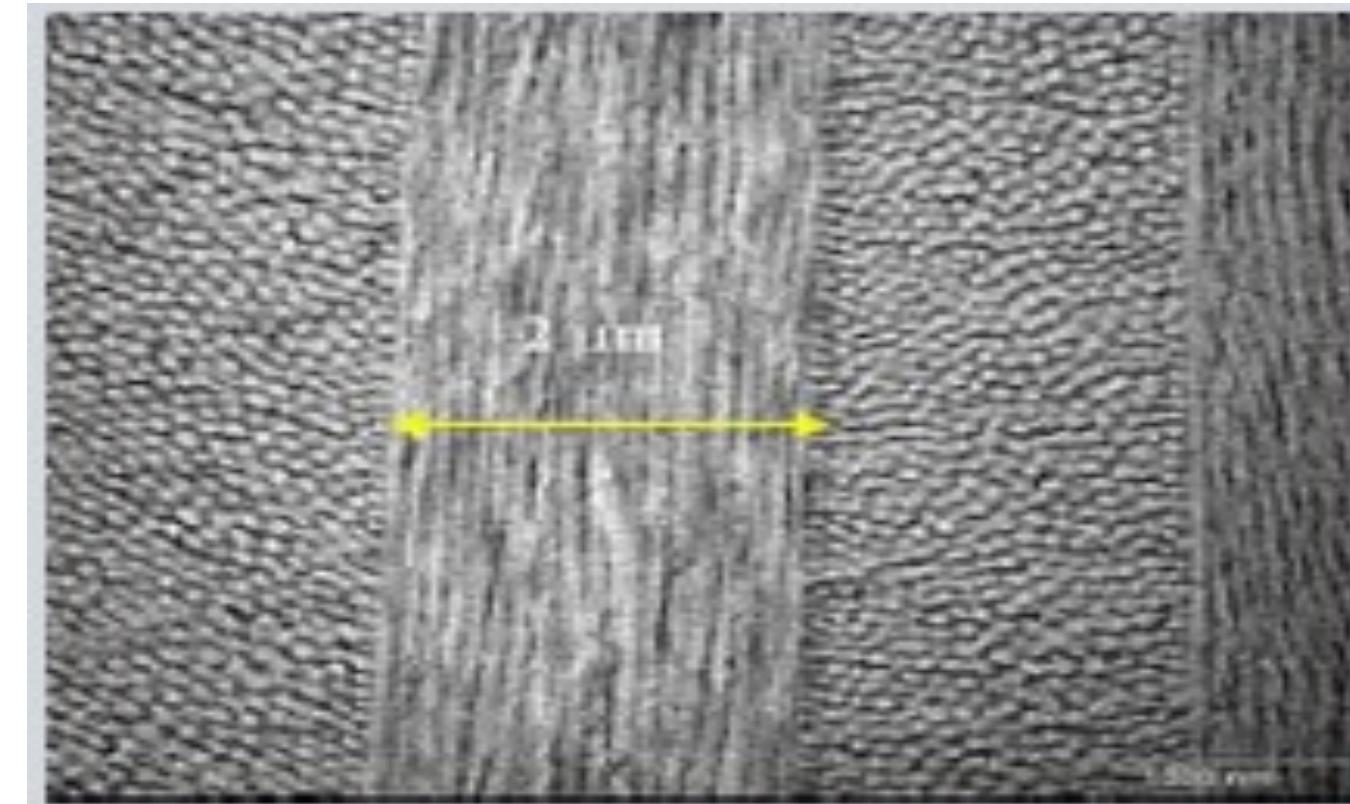


Fig.1: Transmission electron microscopic image of structure of the inside of fish scale.
The lining-up structure of collagen-fibril sheets with alternate rotation of 90 degrees was observed.

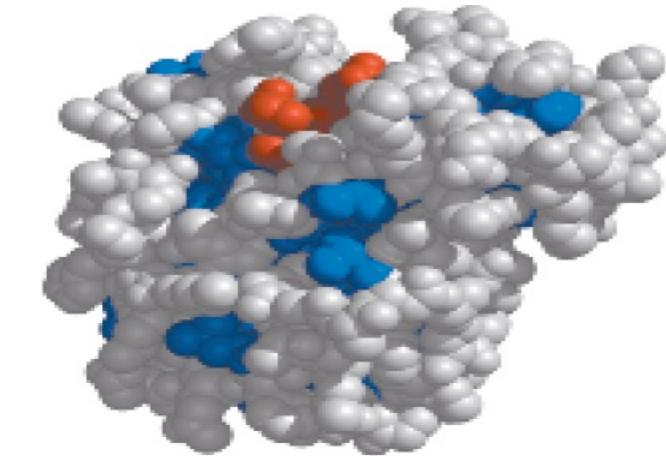
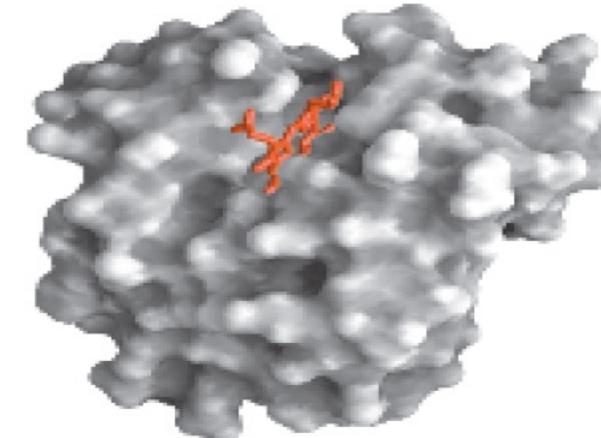
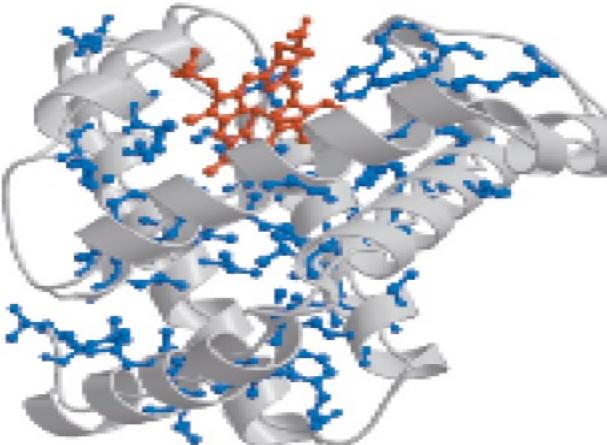
Globular Proteins

- ❖ Folding generates a compact form relative to polypeptides in a fully extended conformation
- ❖ The folding also provides the structural diversity necessary for proteins to carry out a wide array of biological functions
- ❖ Globular proteins include enzymes, transport proteins, motor proteins, regulatory proteins, immunoglobulins, and proteins with many other functions.

Globular Protein Structure - Myoglobin

- ❖ The first breakthrough in understanding the three-dimensional structure of a globular protein came from x-ray diffraction studies of myoglobin carried out by John Kendrew and his colleagues in the 1950s
- ❖ Myoglobin contains a single polypeptide chain of 153 amino acid residues of known sequence and a single iron protoporphyrin, or heme, group
 - ❖ The same heme group is found in hemoglobin, the oxygen-binding protein of erythrocytes, and is responsible for the deep red-brown color of both myoglobin and hemoglobin.

Different visualizations of Myoglobin

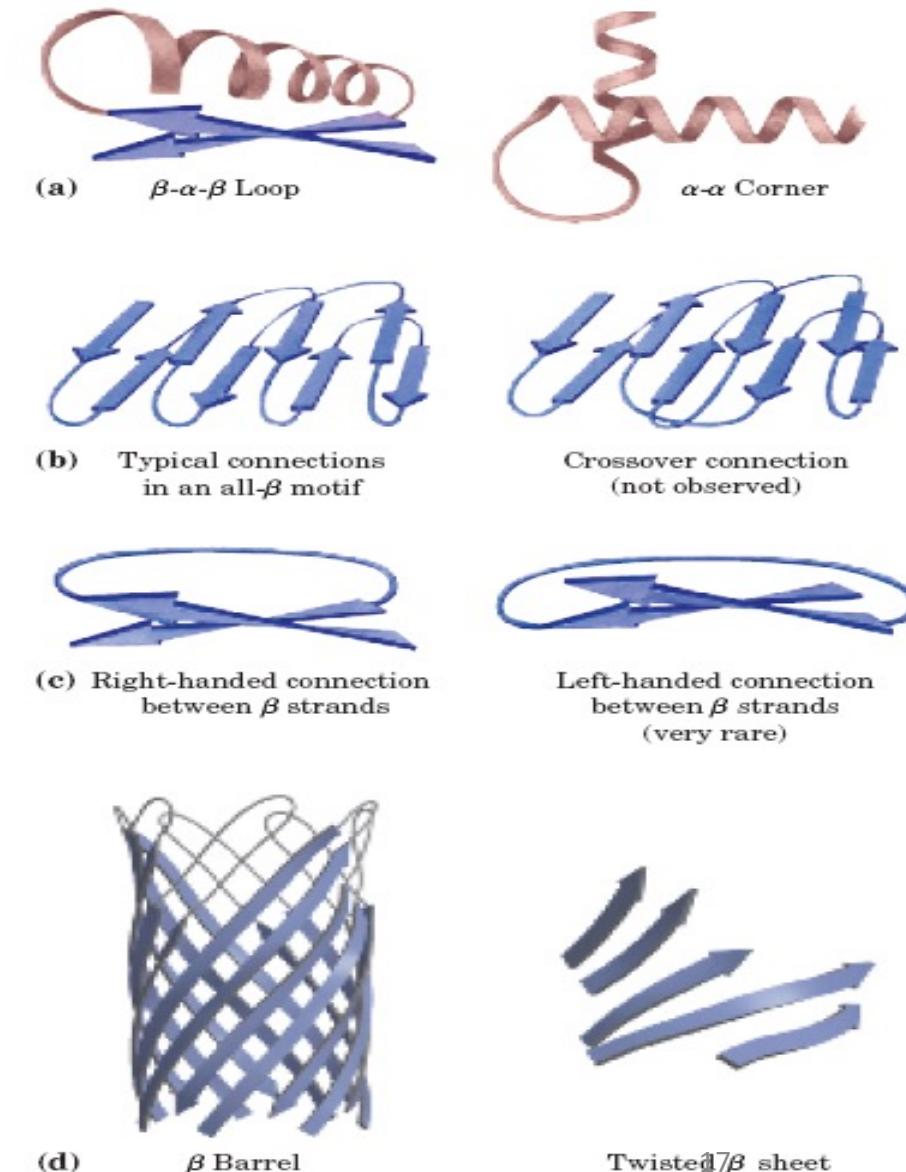


Tertiary structure of sperm whale myoglobin

- ❖ A ribbon representation, including side chains (blue) for the hydrophobic residues Leu, Ile, Val and Phe
- ❖ A surface contour image is useful for visualizing pockets in the protein where other molecules might bind
- ❖ A space-filling model with all amino acid side chains. Each atom is represented by a sphere encompassing its van der Waals radius. The hydrophobic residues are again shown in blue; most are not visible, because they are buried in the interior of the protein

Common Structural Patterns

- ❖ Supersecondary structures, also called motifs or simply folds, are particularly stable arrangements of several elements of secondary structure and the connections between them
- ❖ A single large motif may comprise the entire protein
 - ❖ coil of –keratin
- ❖ Motifs of different types occur based on structural constraints



Rules for Folding

1. Hydrophobic interactions make a large contribution to the stability of protein structures. Burial of hydrophobic amino acid R groups so as to exclude water requires at least two layers of secondary structure. Two simple motifs, the β - α - β **loop** and the α - α **corner**, create two layers.
2. Where they occur together in proteins, α -helices and β -sheets generally are found in different structural layers. This is because the backbone of a polypeptide segment in the β conformation cannot readily hydrogen-bond to an α -helix aligned with it

- 
3. Polypeptide segments adjacent to each other in the primary sequence are usually stacked adjacent to each other in the folded structure. Although distant segments of a polypeptide may come together in the tertiary structure, this is not the norm.
 4. Connections between elements of secondary structure cannot cross or form knots
 5. The β conformation is most stable when the individual segments are twisted slightly in a right handed sense. This influences both the arrangement of β sheets relative to one another and the path of the polypeptide connection between them.

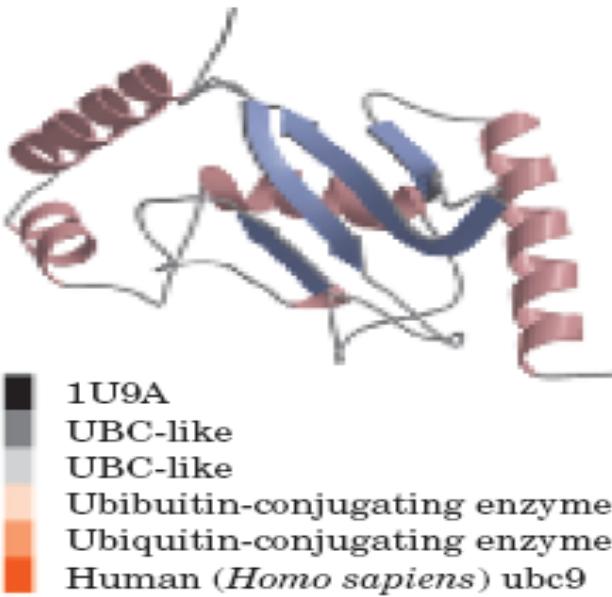
α -Helix and β -Sheet content in different proteins vary

Protein (total residues)	Residues (%) [*]	
	α Helix	β Conformation
Chymotrypsin (247)	14	45
Ribonuclease (124)	26	35
Carboxypeptidase (307)	38	17
Cytochrome c (104)	39	0
Lysozyme (129)	40	12
Myoglobin (153)	78	0

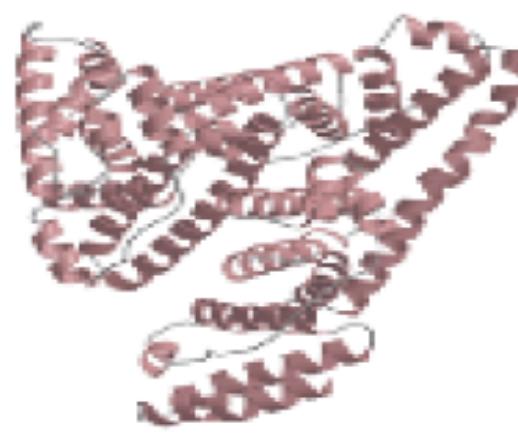
Structural Classification of Proteins

- ❖ The Structural Classification of Proteins (SCOP) database offers a good example of this very important trend in biochemistry
 - ❖ At the highest level of classification, the SCOP database (<http://scop.mrc-lmb.cam.ac.uk/scop>) borrows a scheme already in common use
 - ❖ all α
 - ❖ all β
 - ❖ α/β (in which the α and β segments are interspersed or alternate)
 - ❖ α and β (in which the regions are somewhat segregated)

N Q L A D V A C H M T
D G K N Q L P Y R E I F W L P D G K D G

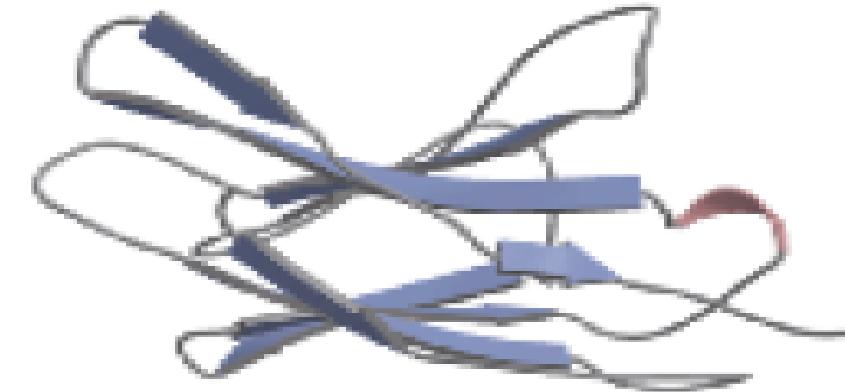


1AO6
Serum albumin
Serum albumin
Serum albumin
Serum albumin
Serum albumin
Human (*Homo sapiens*)

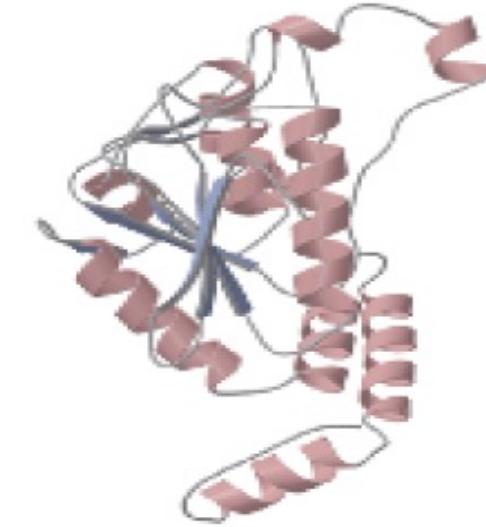


PDB identifier
Fold
Superfamily
Family
Protein
Species

1CD8
Immunoglobulin-like β sandwich
Immunoglobulin
V set domains (antibody variable domain-like)
CD8
Human (*Homo sapiens*)



1DUB
ClpP/crotonase
ClpP/crotonase
Crotonase-like
Enoyl-CoA hydratase (crotonase)
Rat (*Rattus norvegicus*)

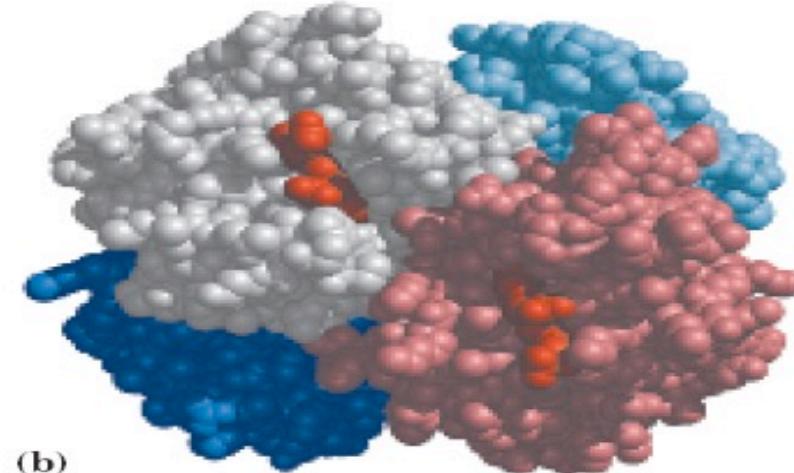
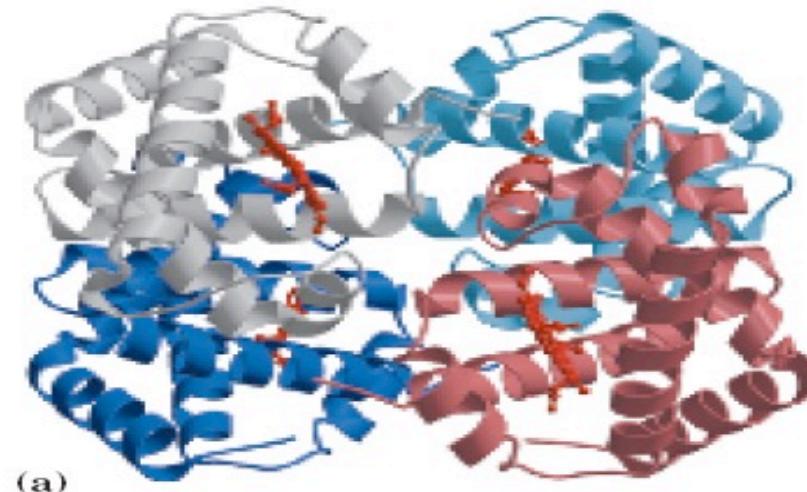


Quaternary Protein Structures

- ❖ Many proteins have multiple polypeptide subunits. The association of polypeptide chains can serve a variety of functions.
 - ❖ Many multisubunit proteins have regulatory roles
 - ❖ The binding of small molecules may affect the interaction between subunits, causing large changes in the protein's activity in response to small changes in the concentration of substrate or regulatory molecules
- ❖ A multisubunit protein is also referred to as a **multimer**
- ❖ **Few subunits – Oligomers**
- ❖ **Repeating subunits - protomers**

Hemoglobin

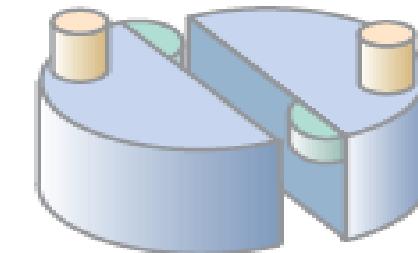
- ❖ X-ray diffraction analysis of deoxyhemoglobin (hemoglobin without oxygen molecules bound to the heme groups) shows how the four polypeptide subunits are packed together
- ❖ A ribbon representation.
- ❖ A space-filling model. The α subunits are shown in gray and light blue; the β subunits in pink and dark blue
- ❖ the heme groups (red) are relatively far apart



- ❖ Identical subunits of multimeric proteins are generally arranged in one or a limited set of symmetric patterns. A description of the structure of these proteins requires an understanding of conventions used to define symmetries.
- ❖ Oligomers can have either **rotational symmetry** or **helical symmetry**

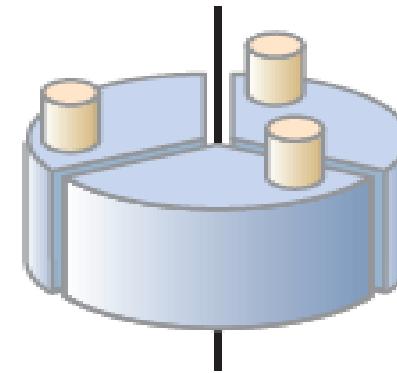
Cyclic Symmetry

Twofold



C_2

Threefold

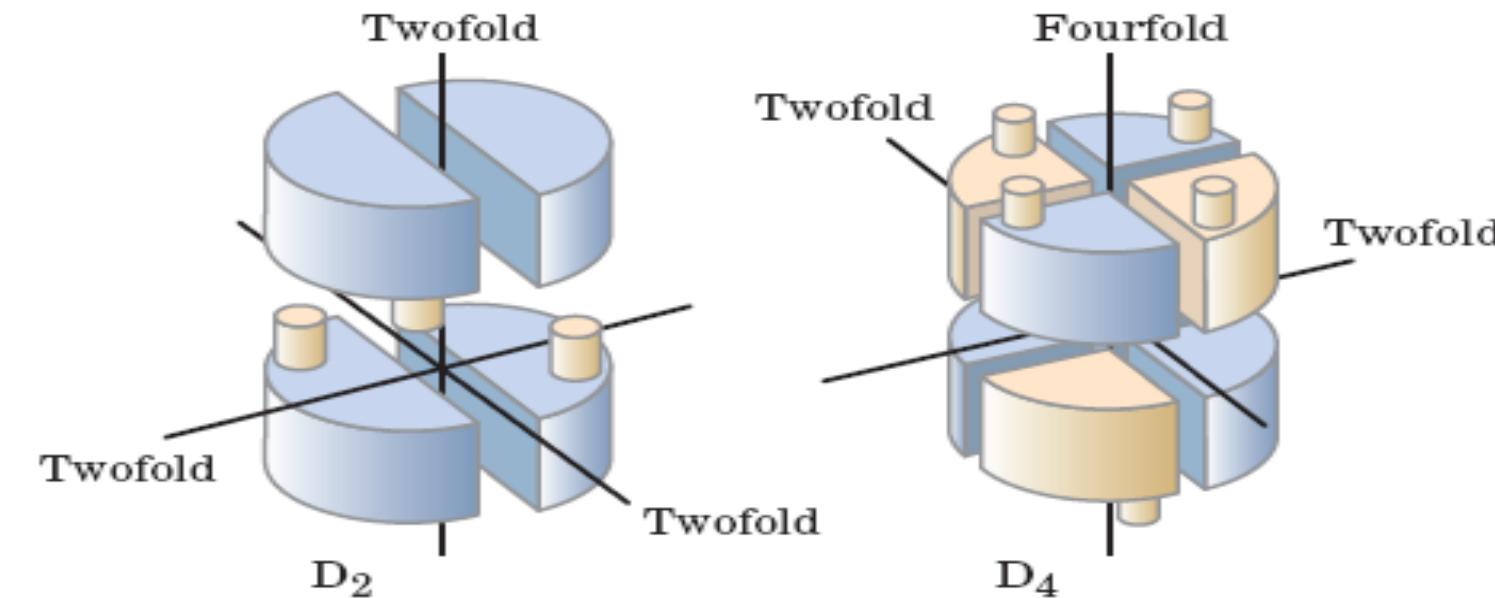


C_3

Two types of cyclic symmetry

- ❖ In cyclic symmetry, subunits are related by rotation about a single n -fold axis, where n is the number of subunits so related. The axes are shown as black lines; the numbers are values of n . Only two of many possible C_n arrangements are shown

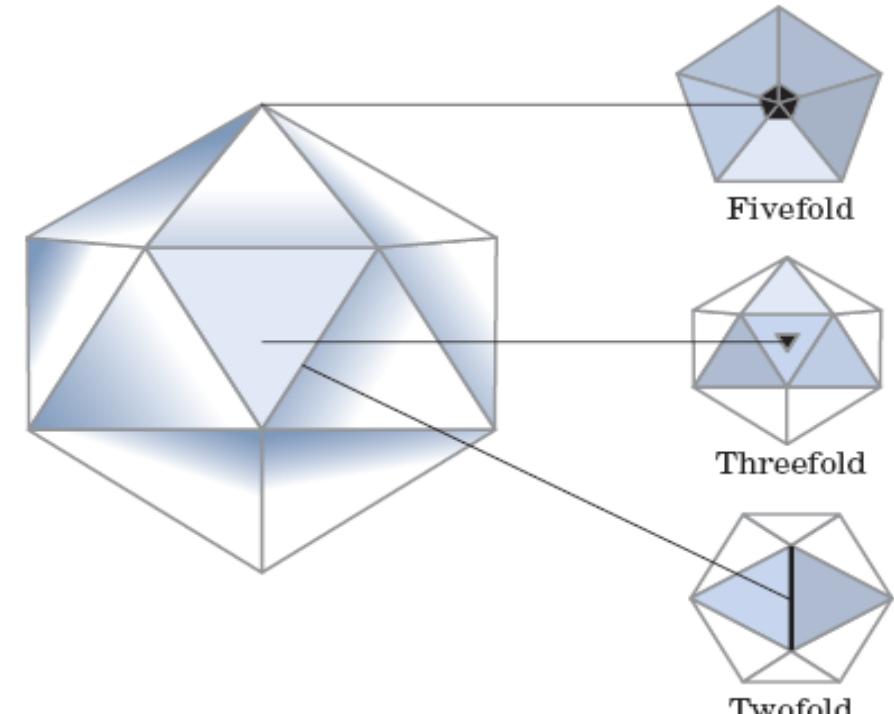
Dihedral Symmetry



- ❖ In dihedral symmetry, all subunits can be related by rotation about one or both of two axes, one of which is twofold. D_2 symmetry is most common.

Icosahedral Symmetry

- ❖ Icosahedral symmetry. Relating all 20 triangular faces of an icosahedron requires rotation about one or more of three separate rotational axes: twofold, threefold, and fivefold.



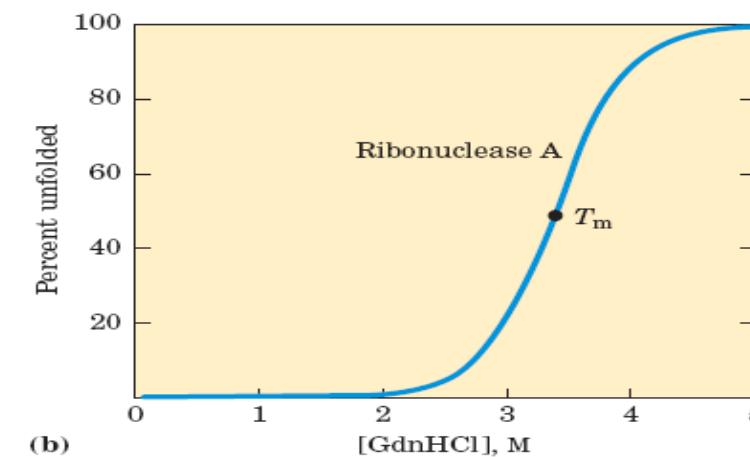
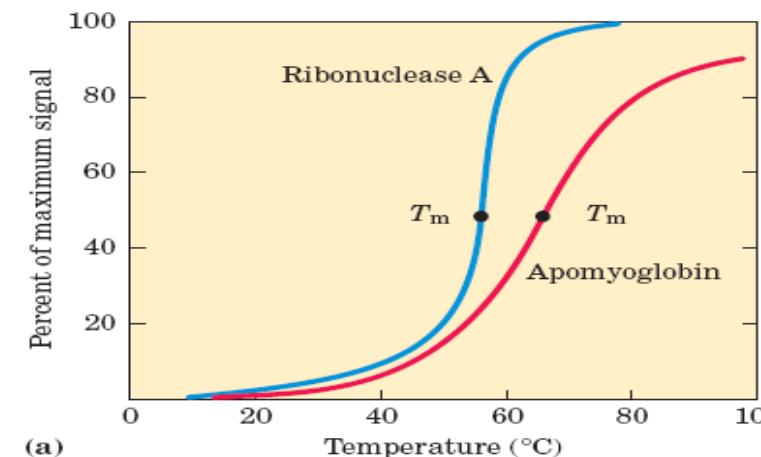
Tertiary and Quaternary Structure

- ❖ Tertiary structure is the complete 3-D structure of a polypeptide chain. There are two general classes of proteins based on tertiary structure: fibrous and globular.
- ❖ Fibrous proteins, which serve mainly structural roles, have simple repeating elements of secondary structure.
- ❖ Globular proteins have more complicated tertiary structures, often containing several types of secondary structure in the same polypeptide chain. The first globular protein structure to be determined, using x-ray diffraction methods, was that of myoglobin.
- ❖ The complex structures of globular proteins can be analyzed by examining stable substructures called supersecondary structures motifs, or folds. The thousands of known protein structures are generally assembled from a repertoire of only a few hundred motifs. Regions of a polypeptide chain that can fold stably and independently are called domains.
- ❖ Quaternary structure results from interactions between the subunits of multisubunit (multimeric) proteins or large protein assemblies. Some multimeric proteins have a repeated unit consisting of a single subunit or a group of subunits referred to as a protomer. Protomers are usually related by rotational or helical symmetry

Denaturation and folding

- ❖ Proteins are marginally stable
 - ❖ Changes in the environmental conditions affects structure and function
- ❖ Denaturation is the process in which change in the 3-D structure of the protein is sufficient to cause a loss of function
- ❖ Factors:
 - ❖ Temperature
 - ❖ pH
 - ❖ Solvents (acetone, alcohol)
 - ❖ Solutes such as urea, guanidine hydrochloride
 - ❖ Detergents

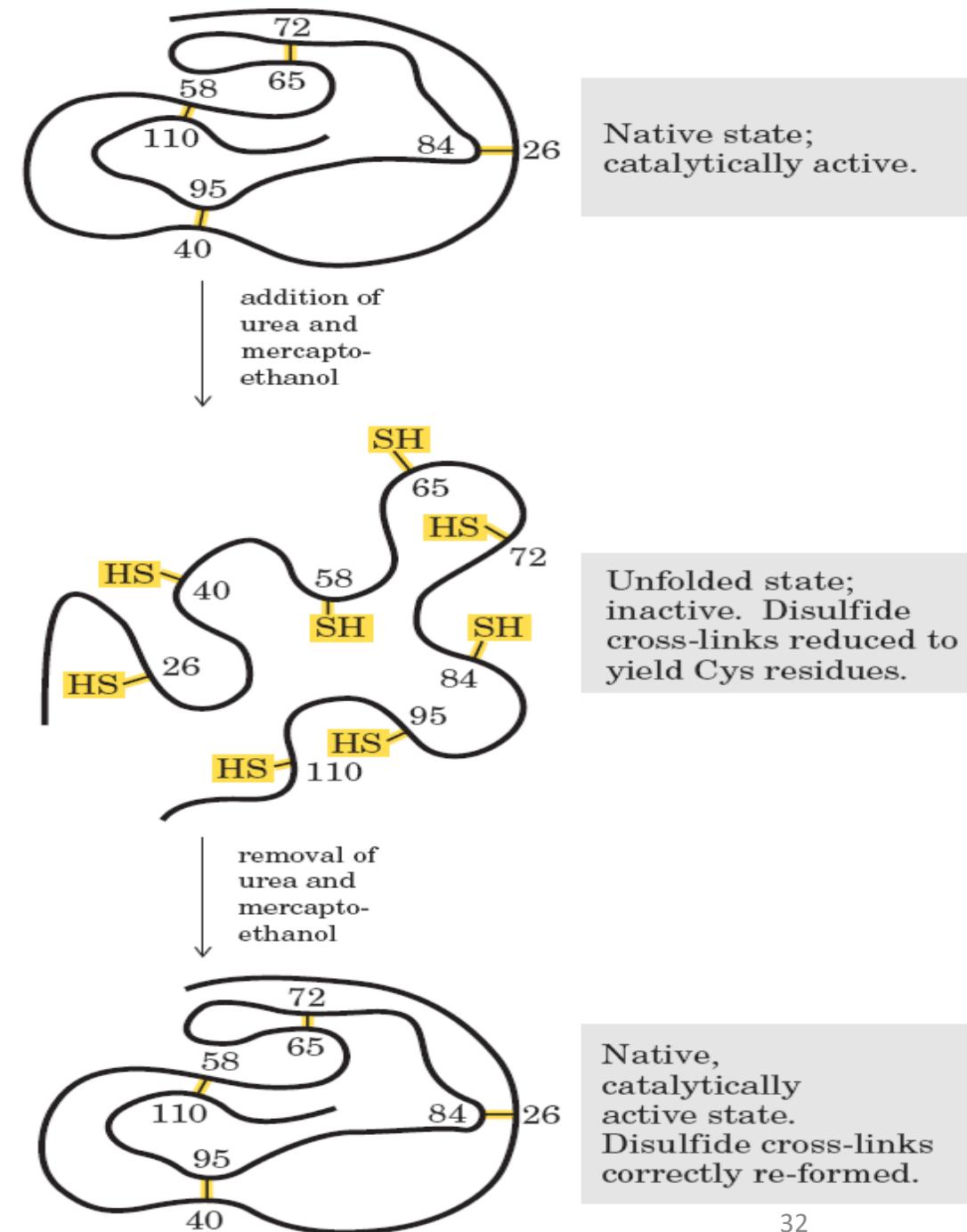
Protein Denaturation



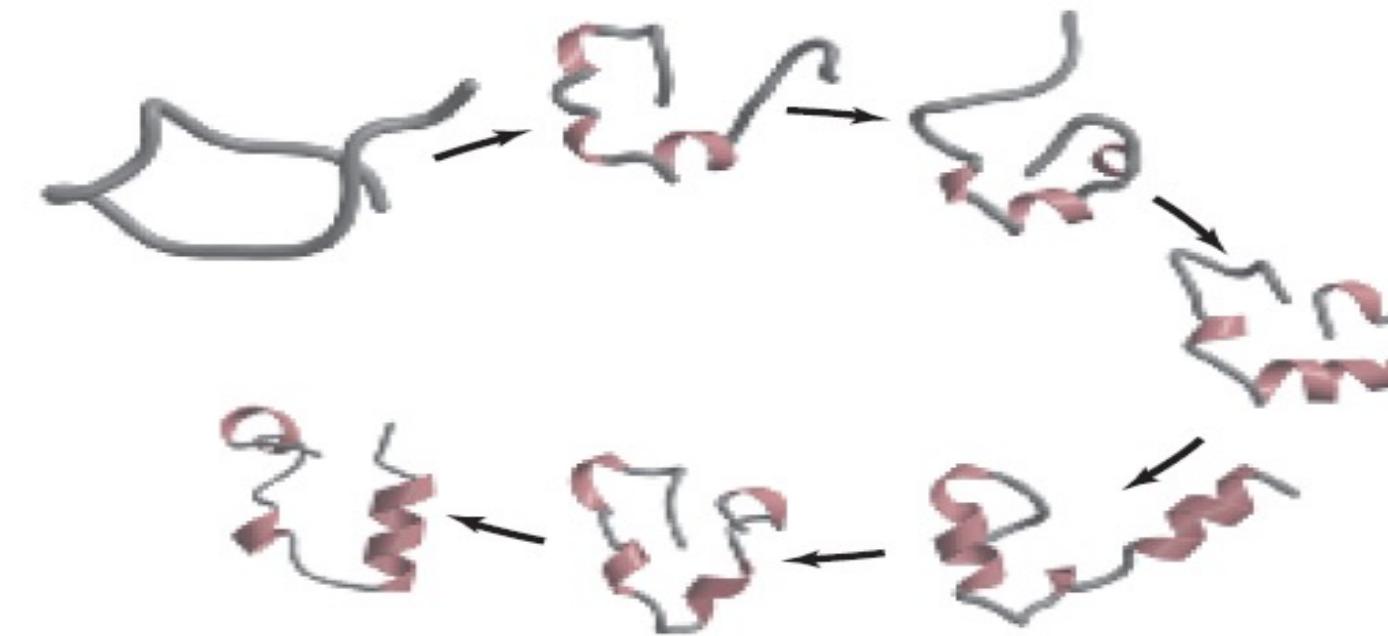
- Denaturation (a) by temperature change (b) by addition of guanidine HCl monitored using Circular Dichroism

Renaturation of unfolded Protein

- ❖ Urea is used to denature ribonuclease, and mercaptoethanol ($\text{HOCH}_2\text{CH}_2\text{SH}$) to reduce and thus cleave the disulfide bonds to yield eight Cys residues. Renaturation involves reestablishment of the correct disulfide cross-links.



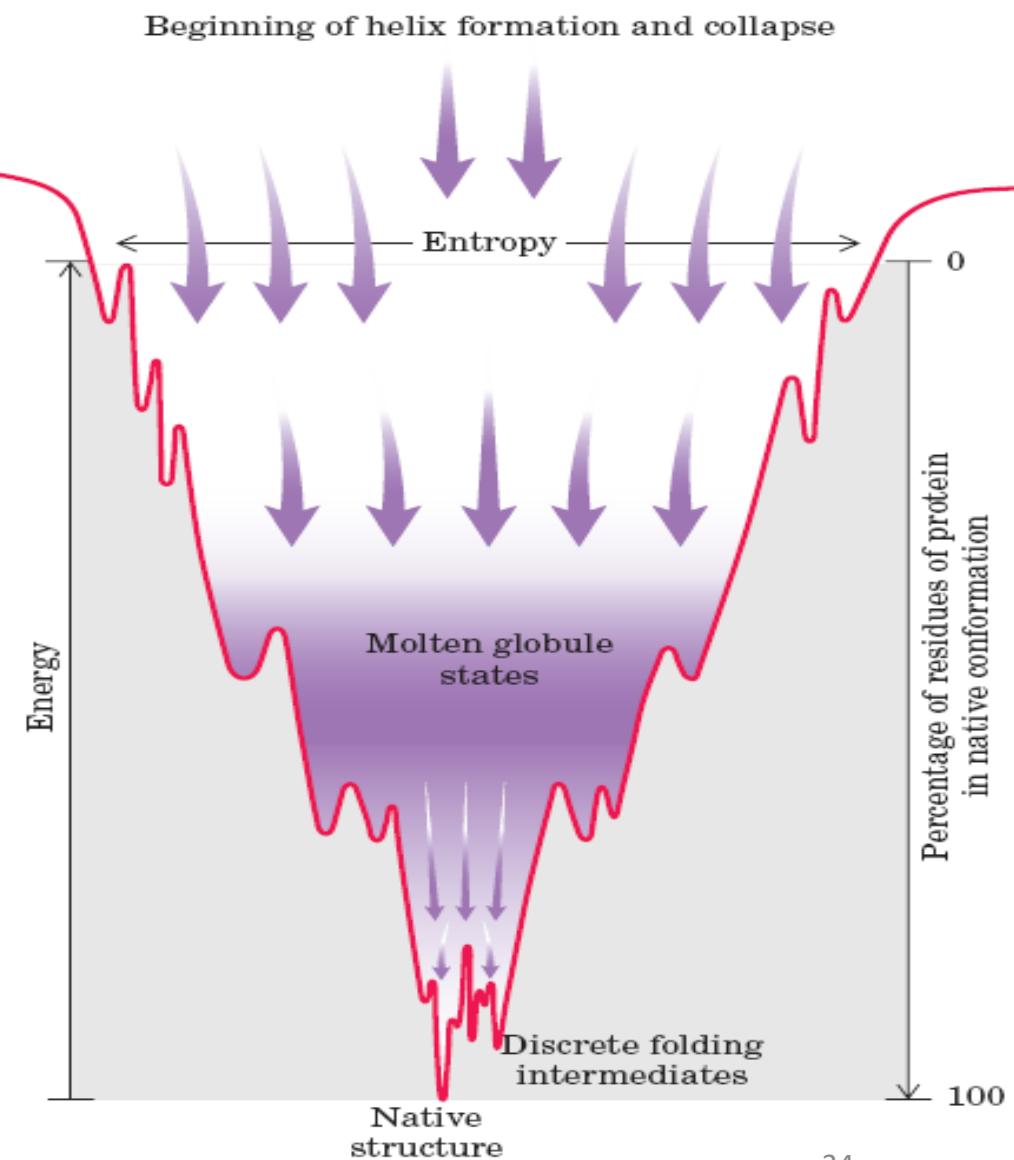
Protein folding simulation



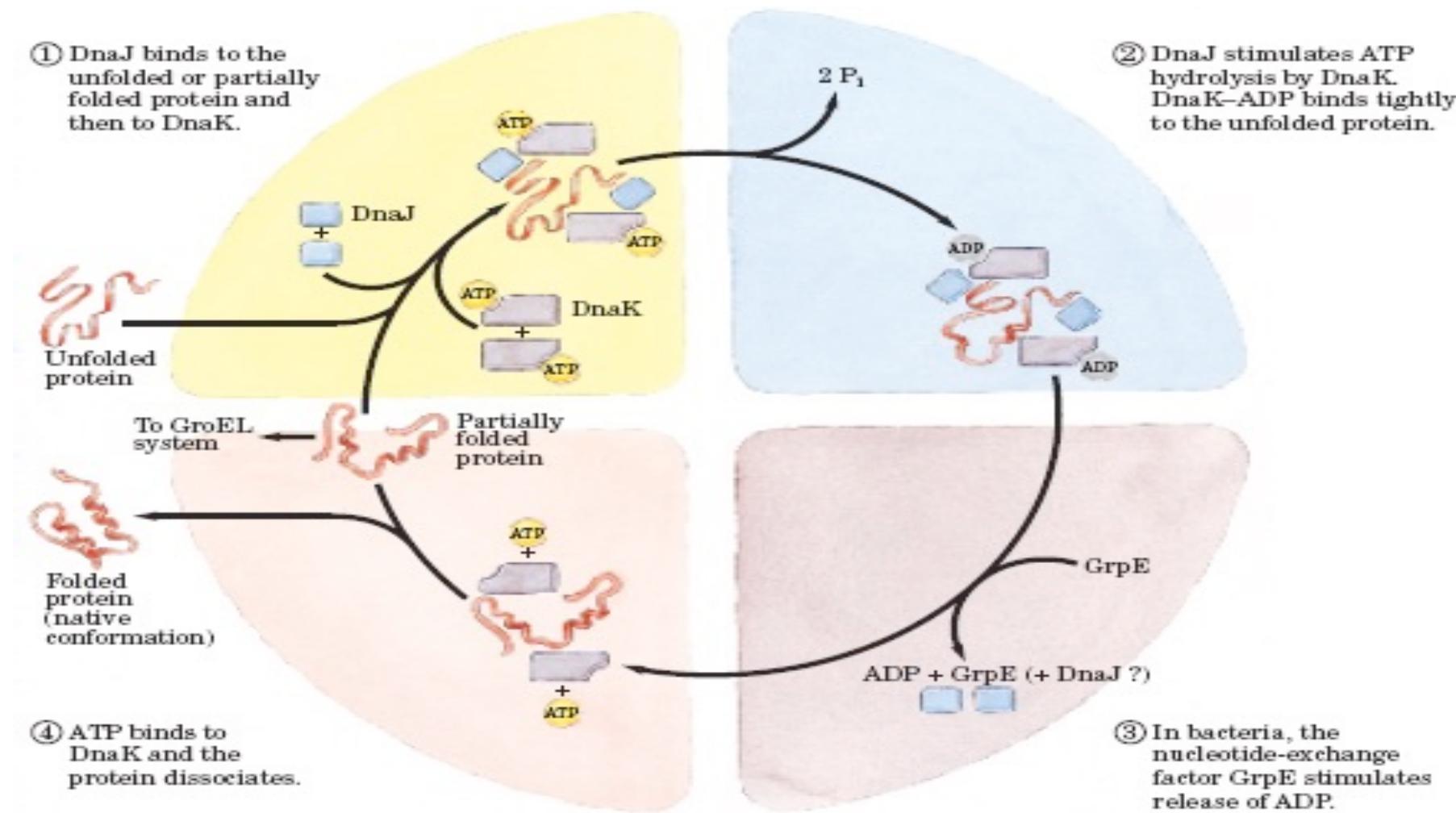
- ❖ The process started with the randomly coiled peptide and 3,000 surrounding water molecules in a virtual “water box.” The molecular motions of the peptide and the effects of the water molecules were taken into account in mapping the most likely paths to the final structure among the countless alternatives.

Thermodynamics of Protein Folding

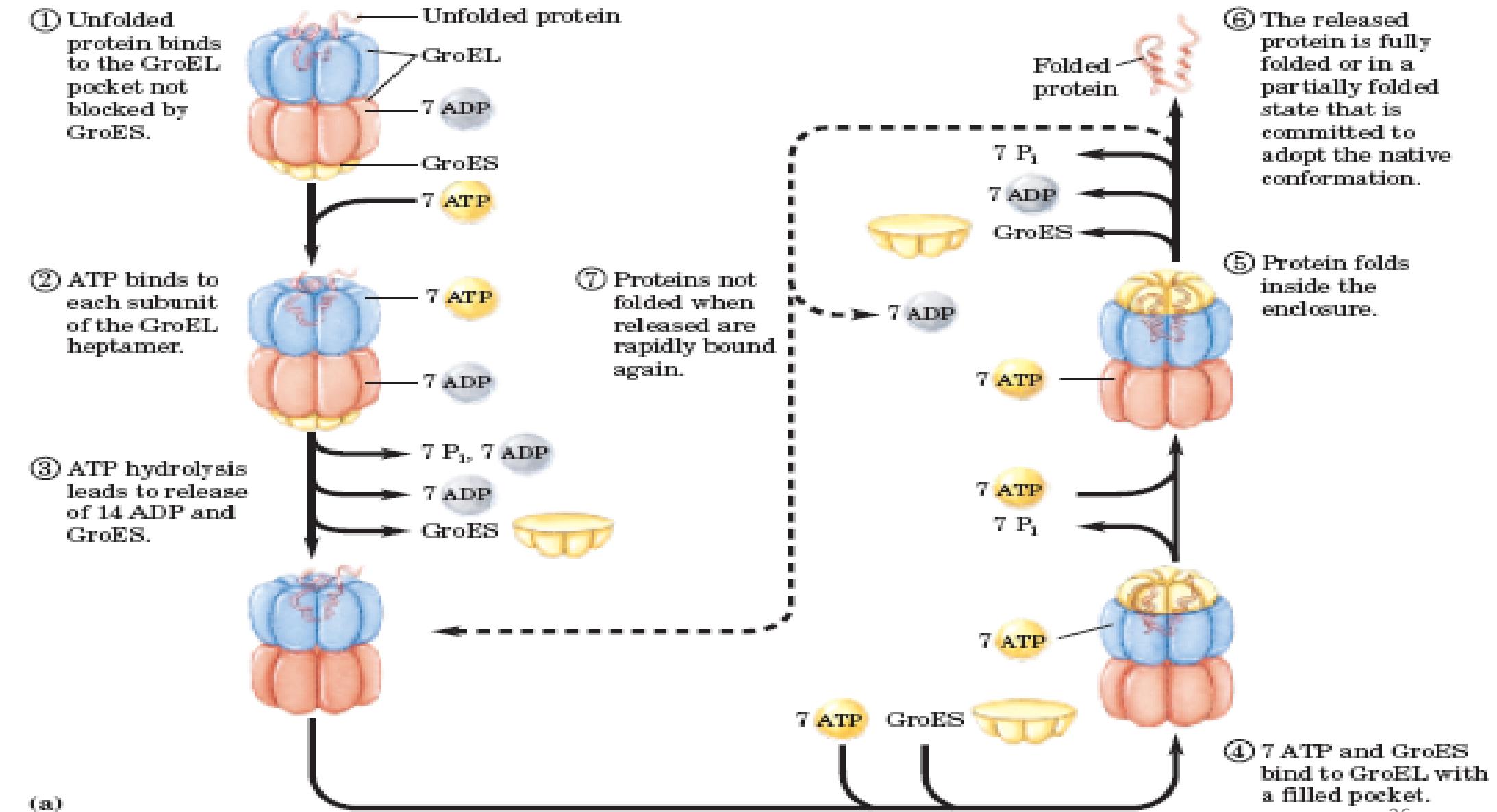
- ❖ The number of conformations, and hence the conformational entropy, is large. Only a small fraction of the intramolecular interactions that will exist in the native conformation are present.
- ❖ As folding progresses, the thermodynamic path down the funnel reduces the number of states present (decreases entropy), increases the amount of protein in the native conformation, and decreases the free energy.
- ❖ Depressions on the sides of the funnel represent semistable folding intermediates, which may, in some cases, slow the folding process.



Chaperones in Protein Folding



GroEL-GroES – member of HSP 60 family





Lectures 9

The Genetic Code

Acknowledgements:
Leninger Chap 27
Scitable
Internet Resources

Objectives of the lecture

1. Learn about the background research that led to the deciphering of the Genetic Code
2. The Experiment of Nirenberg
3. The establishment of the genetic code

The protein synthesis process is complex

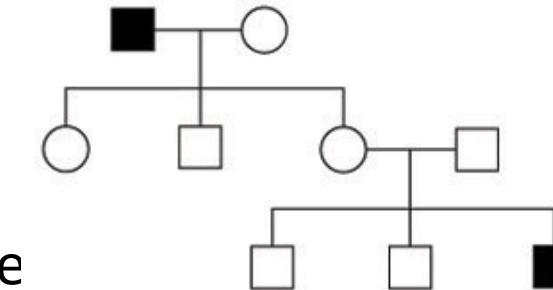
1. Eukaryotic protein synthesis involves more than
 - a) 70 different ribosomal proteins
 - b) 20 or more enzymes to activate the amino acid precursors
 - c) a dozen or more auxiliary enzymes and other protein factors for the initiation, elongation, and termination of polypeptides
 - d) 100 additional enzymes for the final processing of different protein
 - e) 40 or more kinds of transfer and ribosomal RNAs
2. Overall, almost 300 different macromolecules cooperate to synthesize polypeptides
3. Every prokaryote or eukaryote cell has thousands of copies of different RNAs and proteins which constitutes about 35% of the cell dry weight
4. Protein synthesis utilizes about 90% of the chemical energy of the cell

Important Contributions that Lead to the Deciphering of the Genetic Code



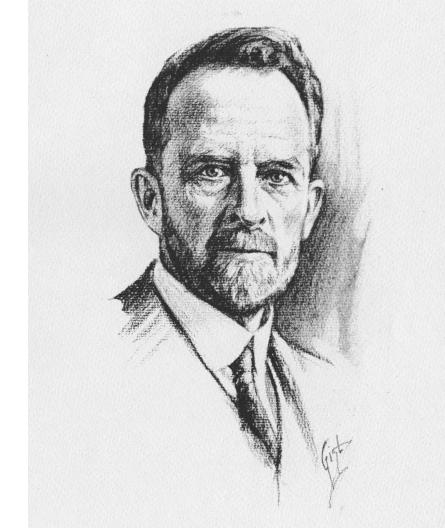
Gregor Mendel

- Elementen
- External resemblance
- Internal nature

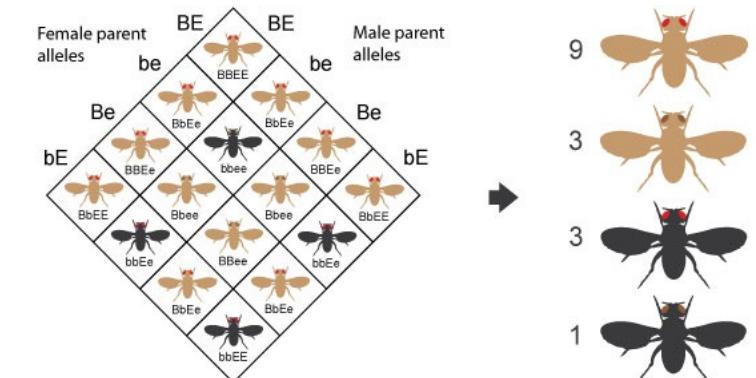


Walter Sutton

Existence of
Chromosome
s in pairs



Thomas Hunt Morgan



Deciphering the Genetic Code

- ❖ 1865 –Mendel defined the basic unit of inheritance as the gene
- ❖ 1900 –Mendel’s forgotten work resurfaces; nature of gene is still unknown
- ❖ 1944 –it is established that a gene is made of DNA
- ❖ 1953 –Watson-Crick’s double helix structure for DNA

DNA: L = {A, C, G, T}

RNA: L = {A, C, G, U}

Double Stranded DNA

5' A T T G C C C A T 3'

 | | | | | | | |

3' T A A C G G G T A 5'

One big question remained unanswered: how is the information in the DNA strand translated to protein?

George Gamow and the “RNA tie Club”

- ❖ Brotherhood consisted of 20 regular members, one for each amino acid
 - ❖ Watson was PRO (proline)
- ❖ Four honorary members, one for each nucleotide
- ❖ Eight of these members were or became Nobel Laureates

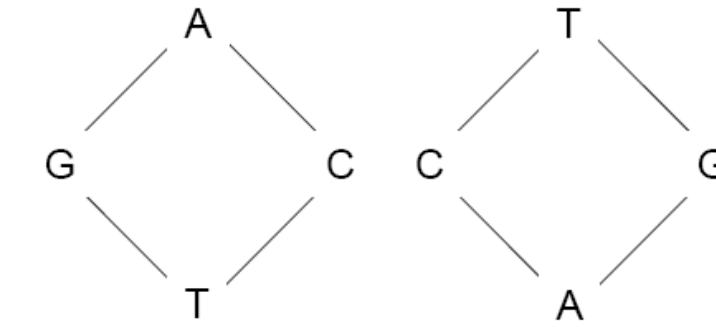


Georgiy Antonovich Gamov
March 4, 1904- August 19, 1968
Big Bang Theory
Formation of stars

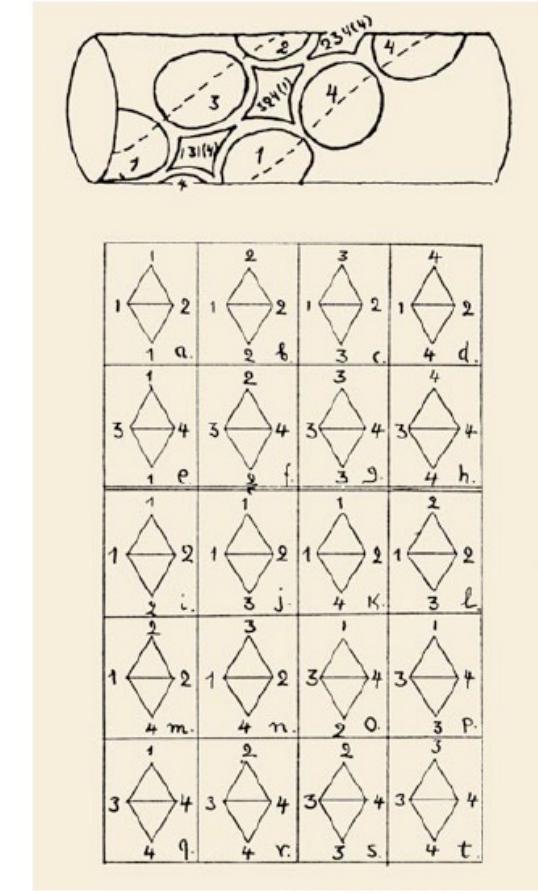
Some of the Ideas Proposed by the RNA Club

- ❖ The Adapter Hypothesis by Francis Crick – some unknown biological entity carried the amino acids and put them in the sequence order
- ❖ Gamow proposed that a three-letter code would be sufficient to define all 20 amino acids

Combinatorial Figures of Gamow's proposition



- ❖ Number of diamonds where top and bottom are identical
- ❖ ${}^4C_1 \times 2 = 8$
- ❖ Number of diamonds where top and bottom are different
- ❖ ${}^4C_2 \times 2 = 12$



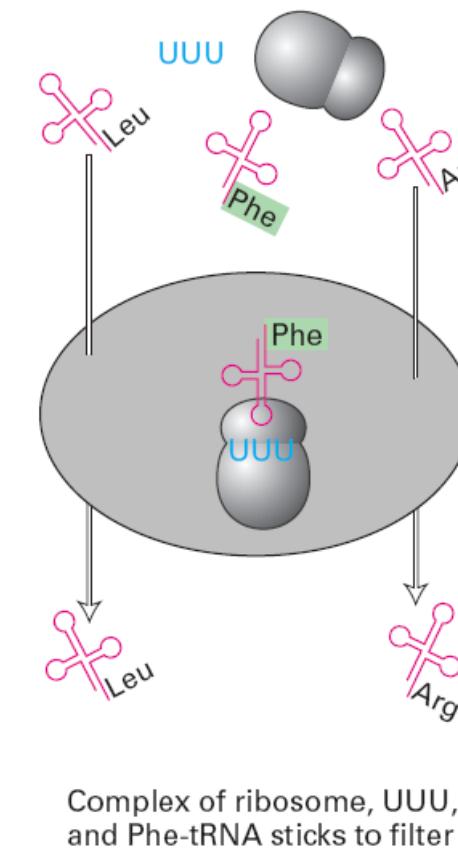
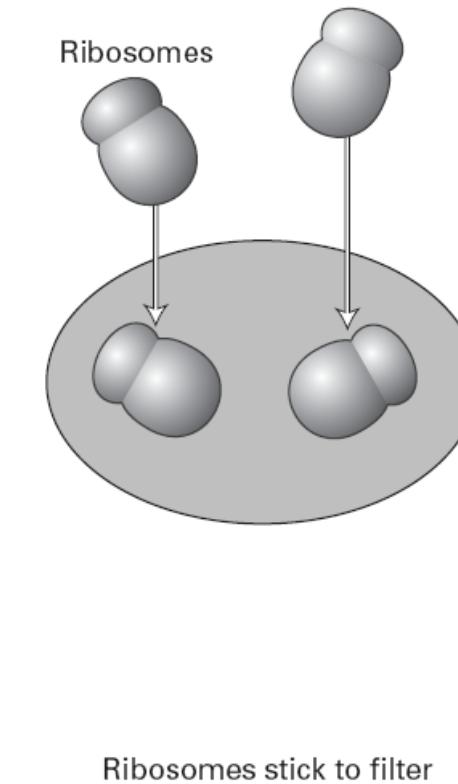
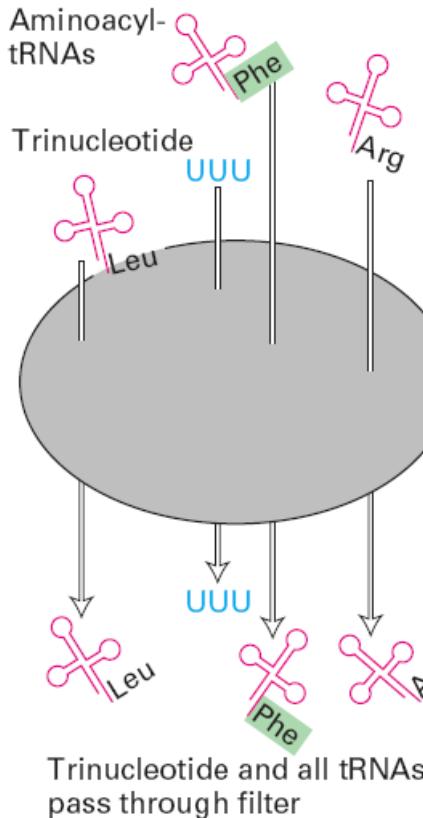
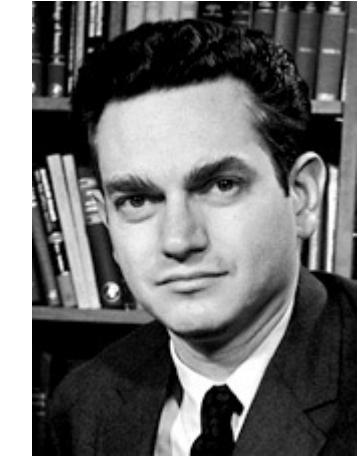
Analysis Presented by Crick

- ❖ Genetic code is in triplets (codon)
 - ❖ There are 20 amino acids
 - ❖ There are 4 alphabets A, T, G, C
 - ❖ $4 & 4^2 < 20, \therefore 4^3 = 64$ (but with redundancies ?)
- ❖ The genetic code should be comma free
- ❖ Only one valid reading frame
 - ❖ [abc][def][ghi][jkl]
 - ❖ NOT a [bcd][efg]hij]...
 - ❖ NOT ab[cde][fgh][ijk]...

(AND THE FOLLOWING THOUGH NOT QUITE CORRECT)

- ❖ AAA, TTT, GGG, CCC are not possible because for example AAAAAA the reading frame is ambiguous
 - ❖ That leaves $64 - 4 = 60$
- ❖ ATGATG must be read unambiguously, So, whenever ATG is a codon, TGA or GAT is not
- ❖ That gives $(1/3) * 60 = 20$

Marshall Nirenberg Deciphered the Genetic Code in 1961



Marshall Warren Nirenberg
April 10, 1927 – January 15, 2010; Jewish American
biochemist and geneticist

M. W. Nirenberg and P. Leder, 1964, *Science* 145:1399

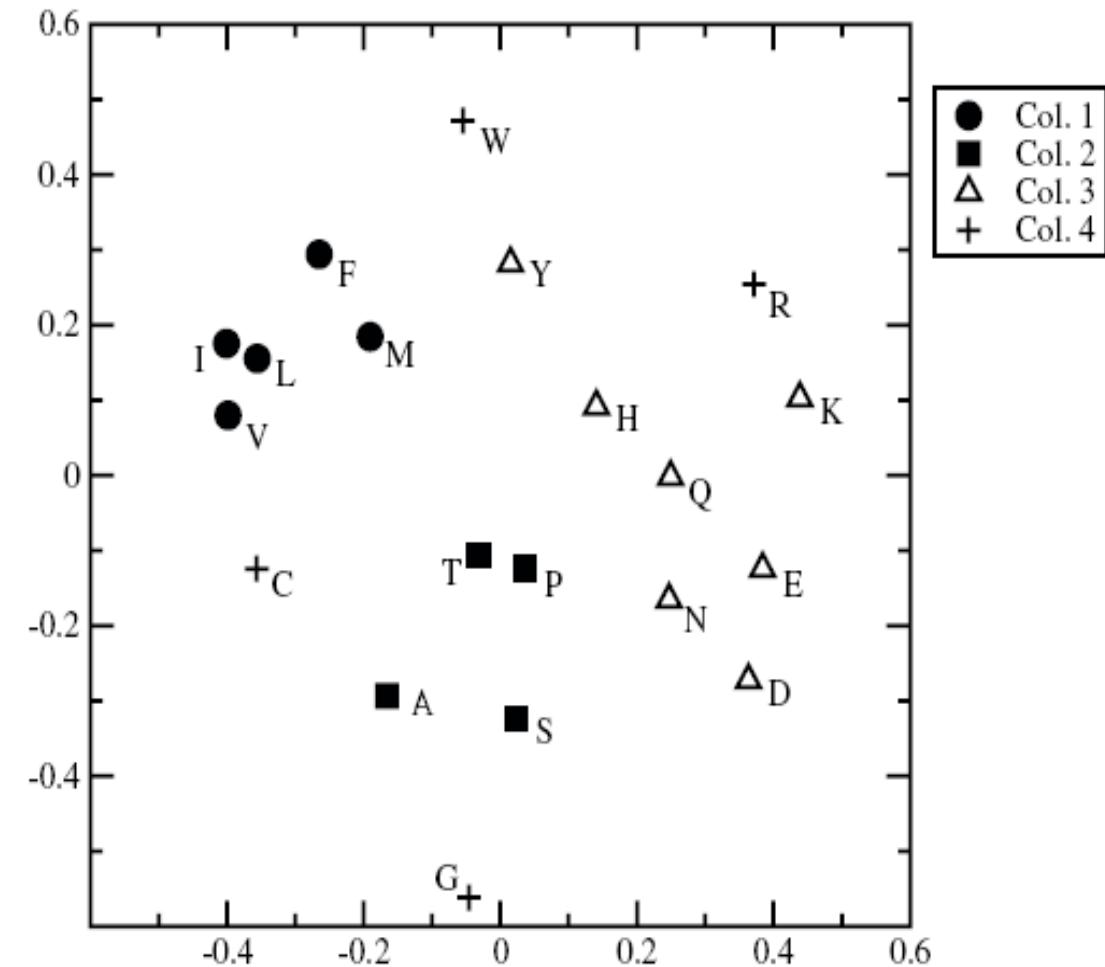
The Genetic Code

The genetic code is a map of Codons “C” to Amino Acids “A”
g: C → A

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

Grouping by Physical Properties of Amino Acids Best Explains the Genetic Code Table

		2nd base					
		U	C	A	G		
1st base	U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Stop UGG Trp	U C A G	3rd base
	C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg GGC Arg CGA Arg GGG Arg	U C A G	
	A	AUU Ae AUC Ae AUA Ae AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G	
	G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gy GGC Gy GGA Gy GGG Gy	U C A G	



Important points related to translation

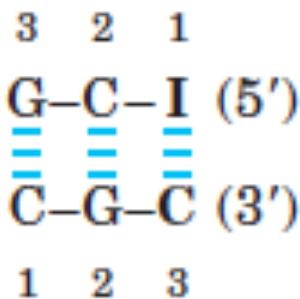
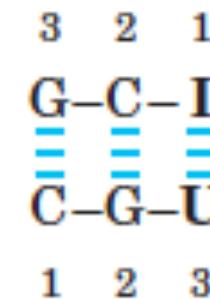
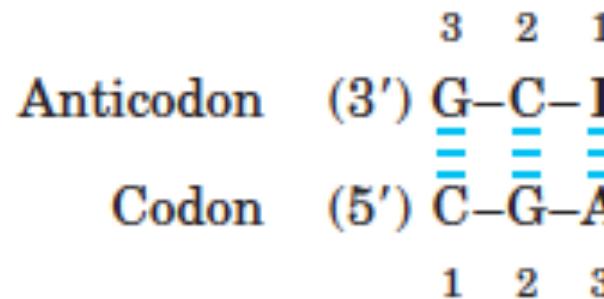
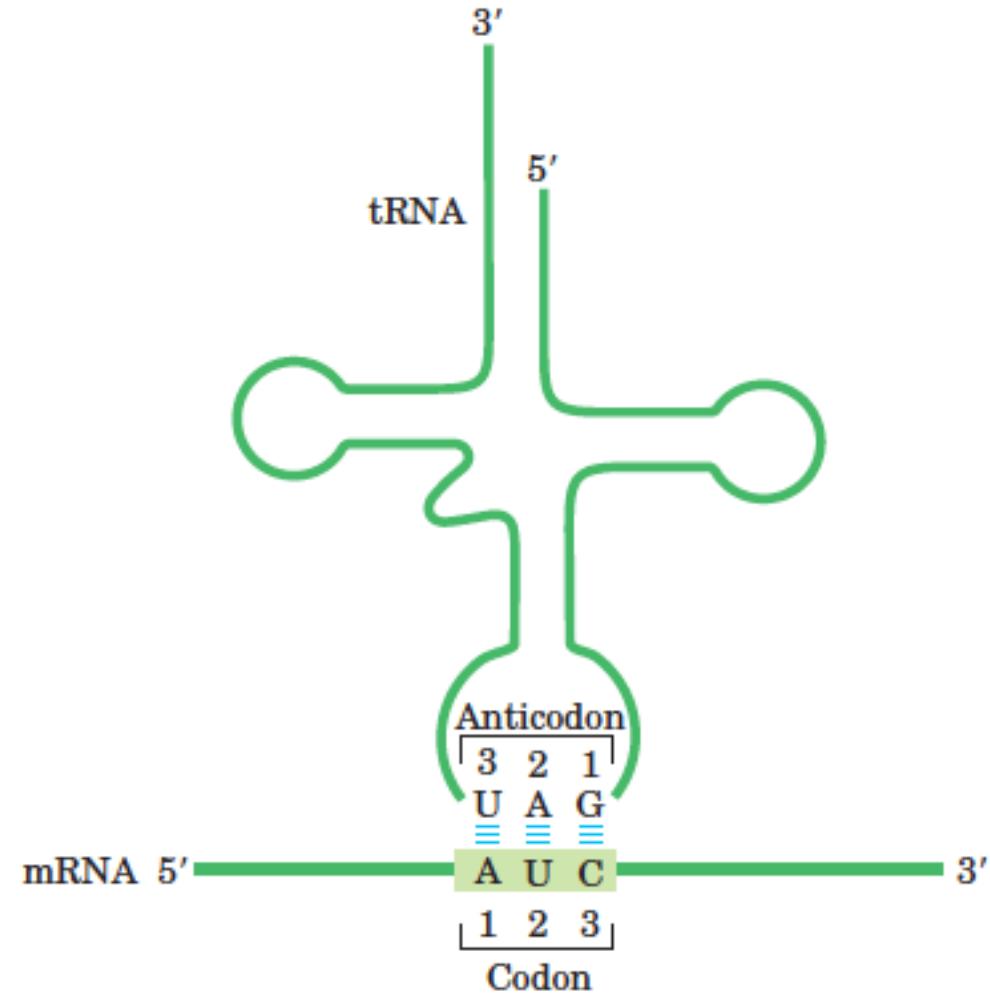
- ❖ The particular amino acid sequence of a protein is constructed through the translation of information encoded in mRNA. This process is carried out by ribosomes.
- ❖ Amino acids are specified by mRNA codons consisting of nucleotide triplets. Translation requires adaptor molecules, the tRNAs, that recognize codons and insert amino acids into their appropriate sequential positions in the polypeptide.
- ❖ The base sequences of the codons were deduced from experiments using synthetic mRNAs of known composition and sequence.
- ❖ The codon AUG signals initiation of translation. The triplets UAA, UAG, and UGA are signals for termination.

Degeneracy of the Genetic Code

<i>Amino acid</i>	<i>Number of codons</i>	<i>Amino acid</i>	<i>Number of codons</i>
Met	1	Tyr	2
Trp	1	Ile	3
Asn	2	Ala	4
Asp	2	Gly	4
Cys	2	Pro	4
Gln	2	Thr	4
Glu	2	Val	4
His	2	Arg	6
Lys	2	Leu	6
Phe	2	Ser	6

The Wobble Hypothesis

- ❖ Alignment of the two RNAs is antiparallel. The tRNA is shown in the traditional cloverleaf configuration
- ❖ Three different codon pairing relationships are possible when the tRNA anticodon contains inosinate.



1. One codon recognized:

1. Anticodon

(3') X-Y-C (5')

(3') X-Y-A (5')

Codon

(5') Y-X-G (3')

(5') Y-X-U (3')

2. Two codons recognized:

1. Anticodon

(3') X-Y-U (5')

(3') X-Y-G (5')

Codon

(5') Y-X-A (3')

(5') Y-X-C (3')

3. Three codons recognized:

1. Anticodon

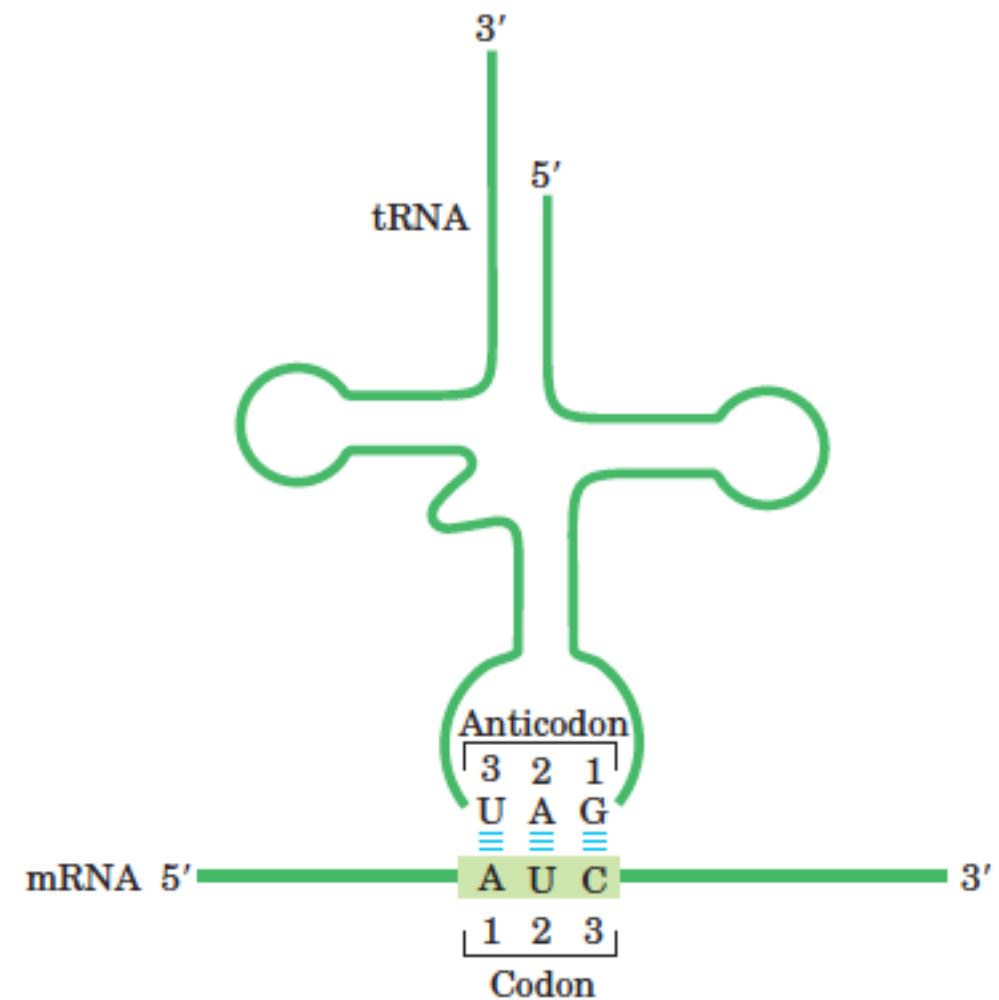
(3') X-Y-I (5')

Codon

(5') Y-X-A (3')

(5') Y-X-U (3')

(5') Y-X-C (3')



G	GUU GUC GUA GUG	Val	GCU GCC GCA GCG	Ala	GAU GAC GAA GAG	Asp	GGU GGC GGA GGG	Gly	U C A G
---	--------------------------	-----	--------------------------	-----	--------------------------	-----	--------------------------	-----	------------------

C	CUU CUC CUA CUG	Leu	CCU CCC CCA CCG	Pro	CAU CAC CAA CAG	His	CGU CGC CGA CGG	Arg	U C A G
---	--------------------------	-----	--------------------------	-----	--------------------------	-----	--------------------------	-----	------------------

Reading Frames

Reading frame 1 5'---[U U C] [U C G] [G A C] [C U G] [G A G] [A U U] [C A C] [A G U] --- 3'

Reading frame 2 --- **U** **U C U** **C G G** **A C C** **U G G** **A G A** **U U C** **A C A** **G U** ---

Reading frame 3 ---U U C U C G G A C C U G G A G A U U C A C A G U---

Nonoverlapping code A U A C G A G U C _____
 1 2 3

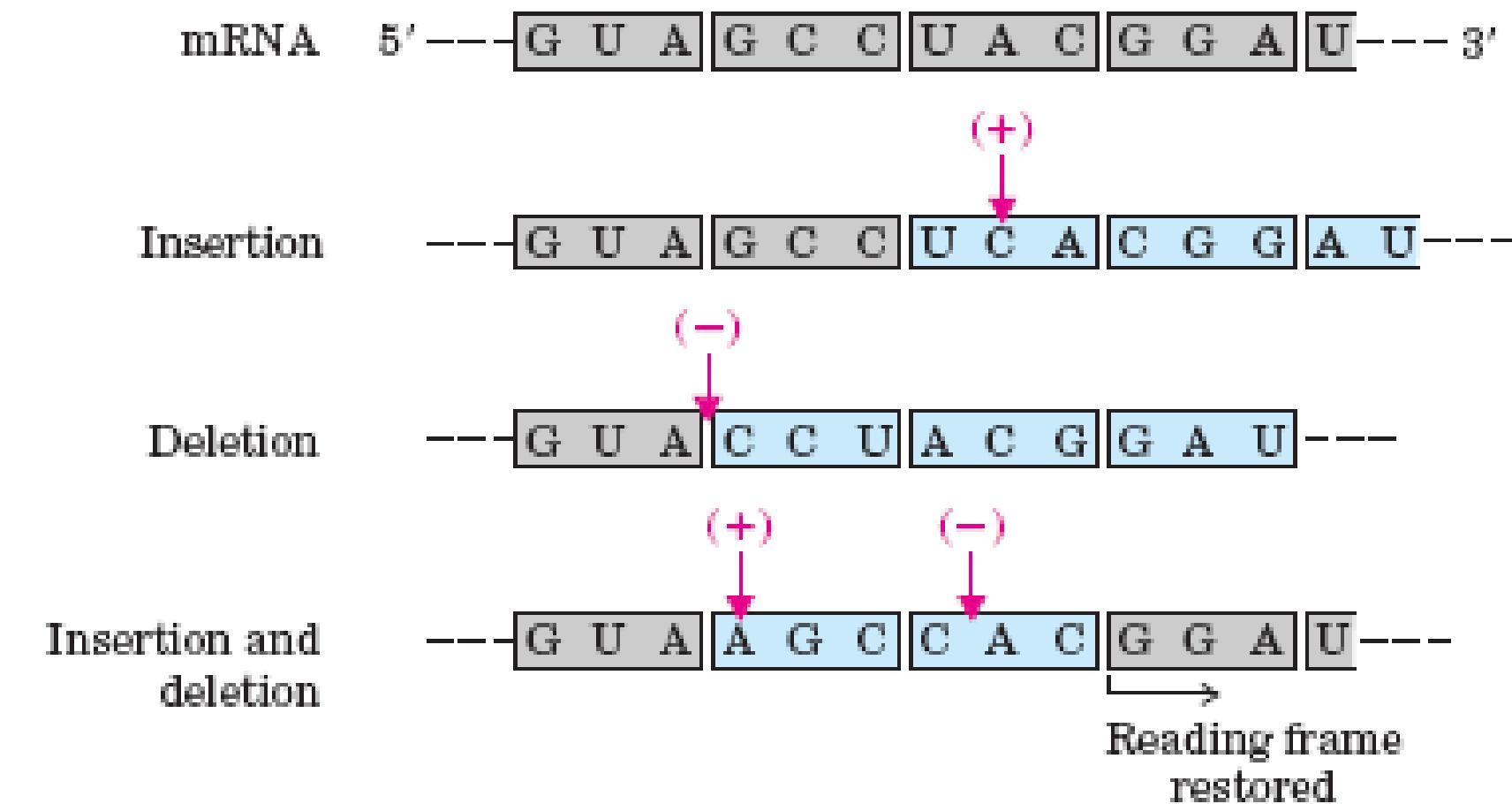
Overlapping code A U A C G A G U C
 1

3

Reading frame 1 5' --- **G U A** **A G U** **A A G** **U A A** **G U A** **A G U** **A A --- 3'**

Reading frame 2 --- G **U A A** **G U A** **A G U** **A A G** **U A A** **G U A** ---

The Triplet Non-overlapping Code



Lecture 10

Protein Synthesis

Acknowledgements:
Leninger Chap 27
Scitable
Internet Resources

Objectives of this Lecture

1. Stages of Protein Synthesis
2. The role of ribosomes
3. The role of tRNA
4. Details of each of the stages of protein synthesis

Five Stage of Protein Synthesis in *E. coli*

1. Activation of Amino Acids
2. Initiation
3. Elongation
4. Termination
5. Folding and post translational processing

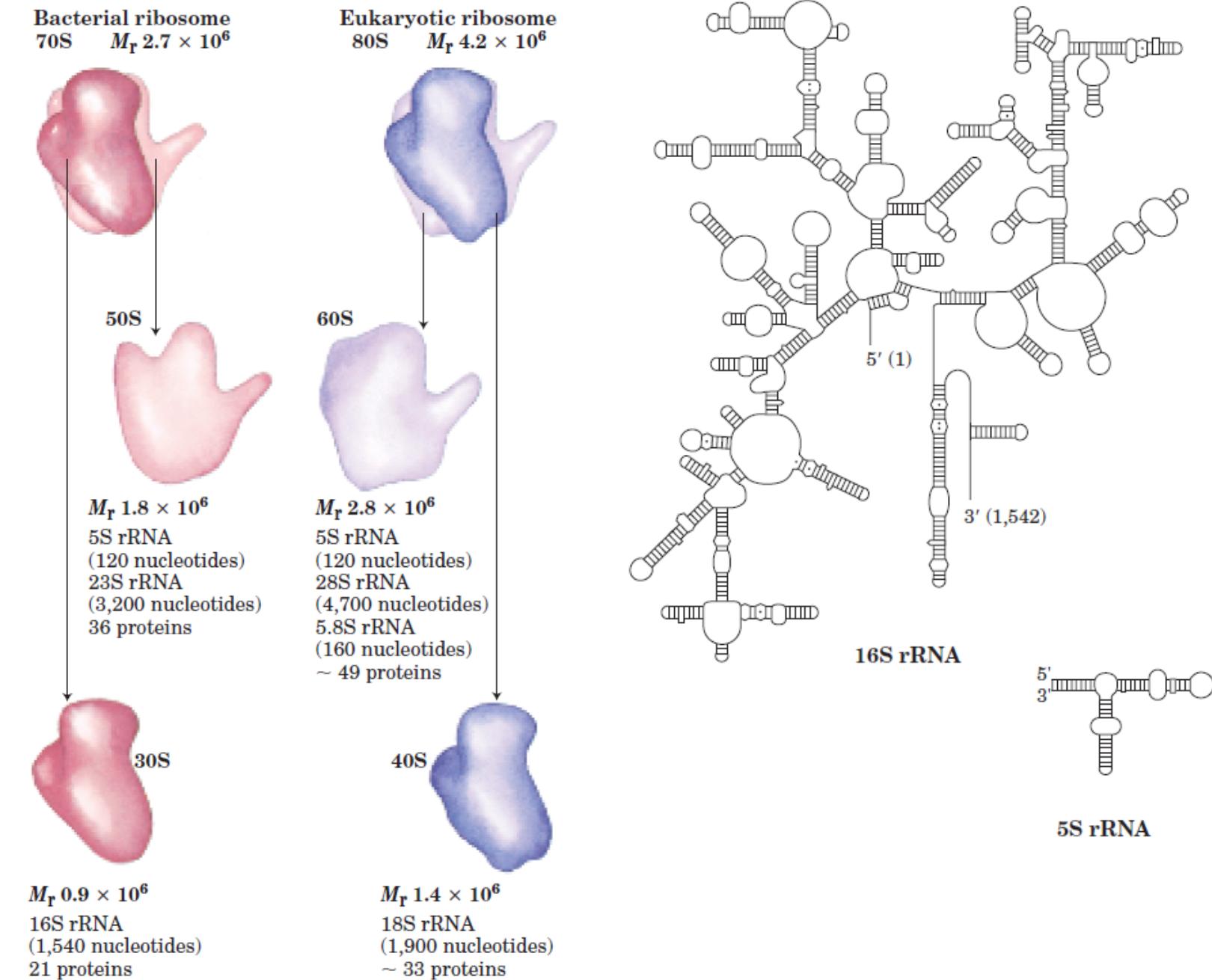
Ribosome is a Complex Supramolecular Machine

1. Each *E. coli* contains 15000 or more ribosomes (1/4 cell weight)
2. 18 nm is size
3. Two subunits (i) 30S (ii) 50S combined 70S (S is the sedimentation coefficient)
4. Subunits are made of many proteins and at least one large rRNA
5. Bacterial ribosomes have 55 proteins with molecular weights varying from 6000 to 75000

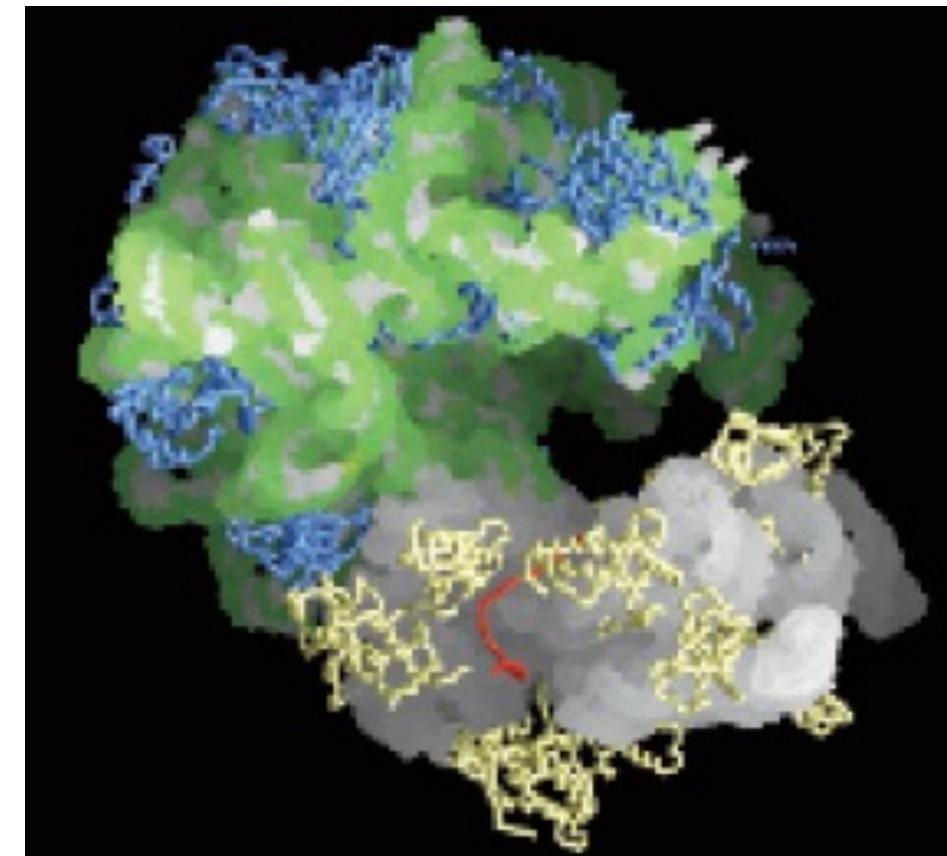
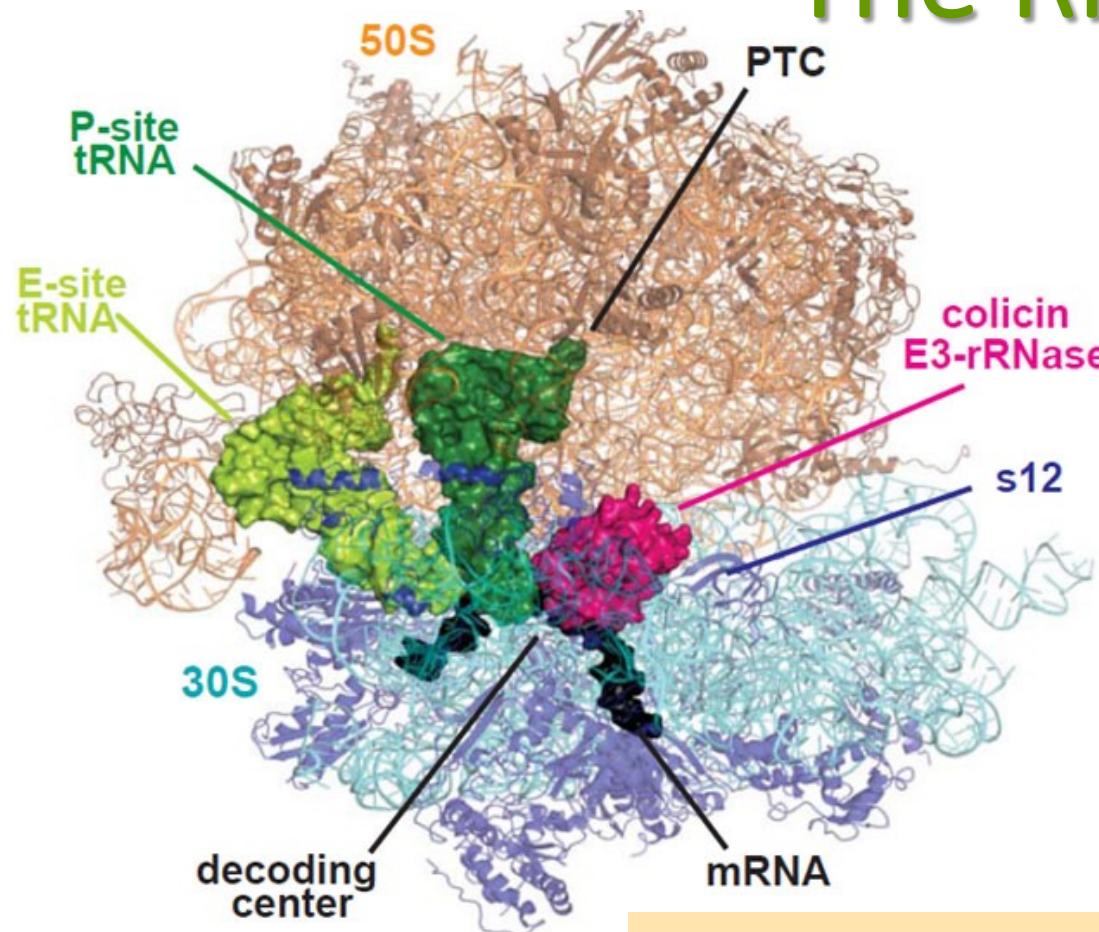
[Look up: PDB ID 1JJ2 and 1GIY](#)

Bacterial rRNA

Secondary structure of *E. coli* 16S and 5S rRNAs. The first (5' end) and final (3' end) ribonucleotide residues of the 16S rRNA are numbered.



The Ribosome



RNA and Protein Components of the *E. coli* Ribosome

Subunit	Number of different proteins	Total number of proteins	Protein designations	Number and type of rRNAs
30S	21	21	S1-S21	1 (16S rRNA)
50S	33	36	L1-L36*	2 (5S and 23S rRNAs)

Nobel Prize in Chemistry 2009



The Nobel Prize in Chemistry 2009 was awarded jointly to Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath "for studies of the structure and function of the ribosome"

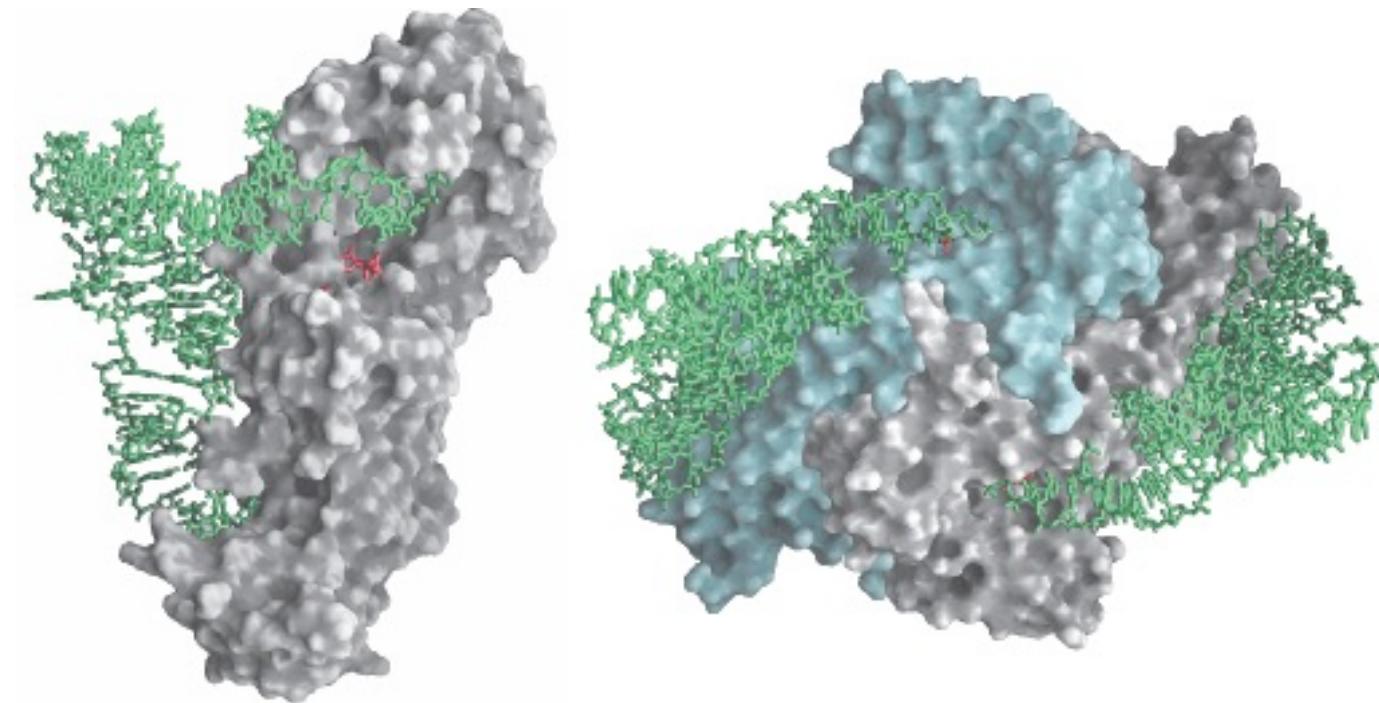
Visit to KSBS Aug 2020

<https://www.nobelprize.org/prizes/chemistry/2009/summary/>

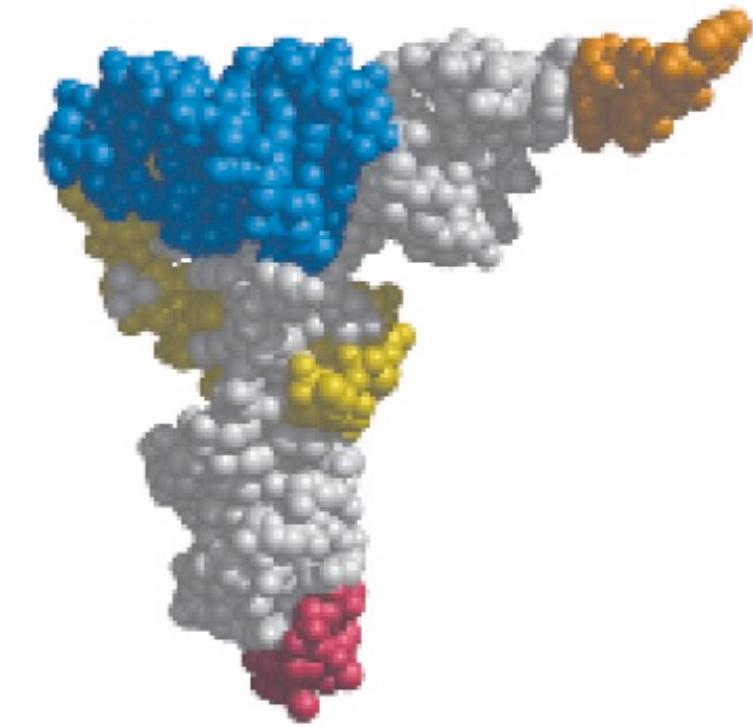
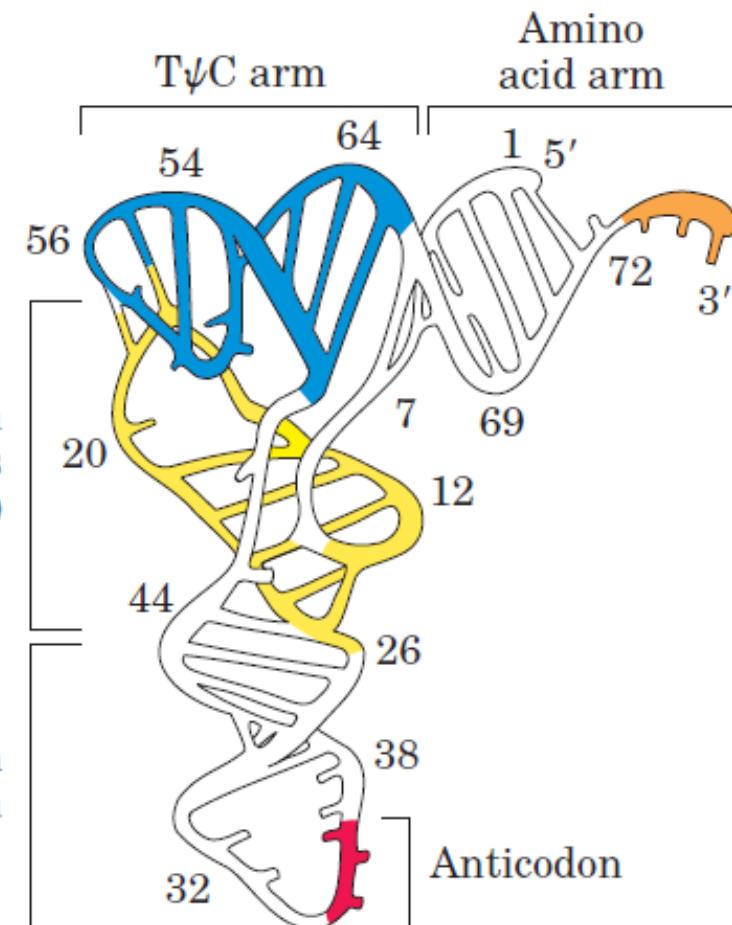
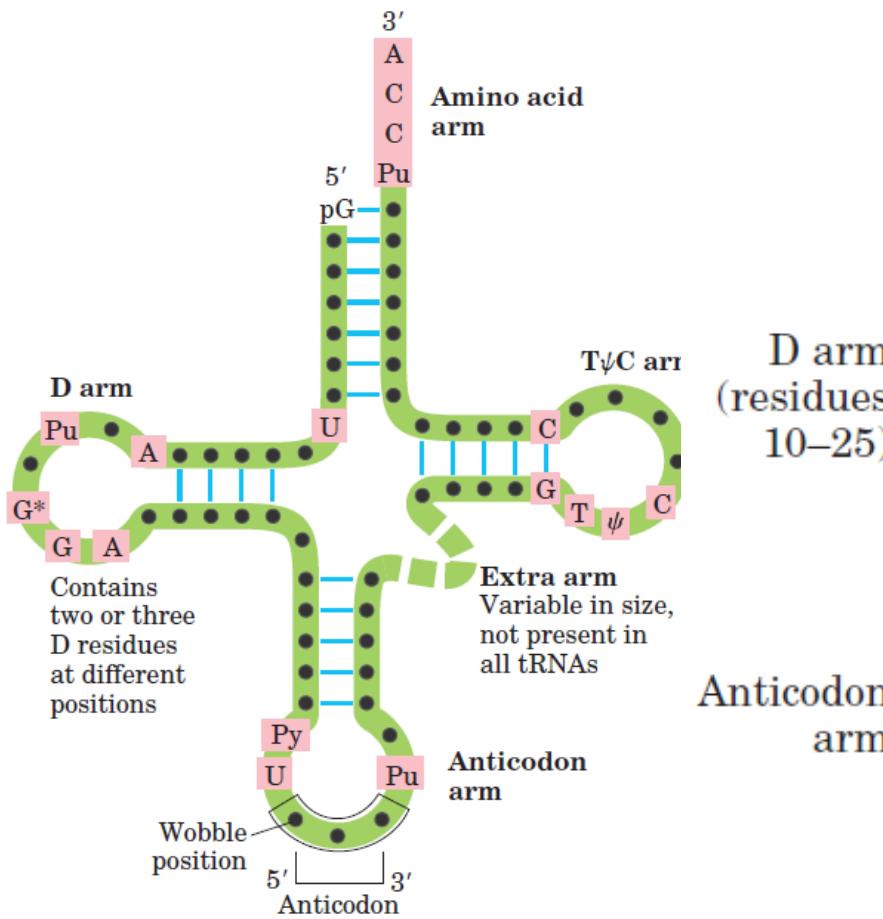
Aminoacyl t-RNA Synthetase

**Aminoacyl-tRNA Synthetases
Attach the Correct Amino Acids to
Their tRNAs**

- ❖ aminoacyl-tRNA synthetases esterify the 20 amino acids to their corresponding tRNAs. Each enzyme is specific for one amino acid and one or more corresponding tRNAs
- ❖ Proofreading by Aminoacyl-tRNA Synthetases
- ❖ Interaction between an Aminoacyl-tRNA Synthetase and a tRNA



2-D and 3-D structure of tRNA

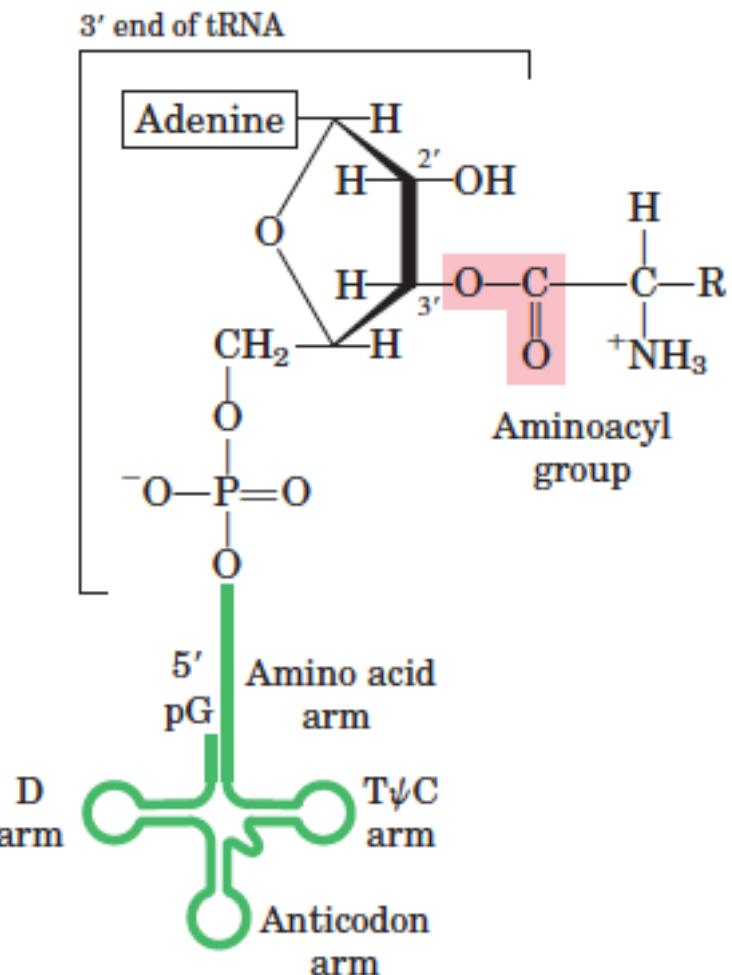
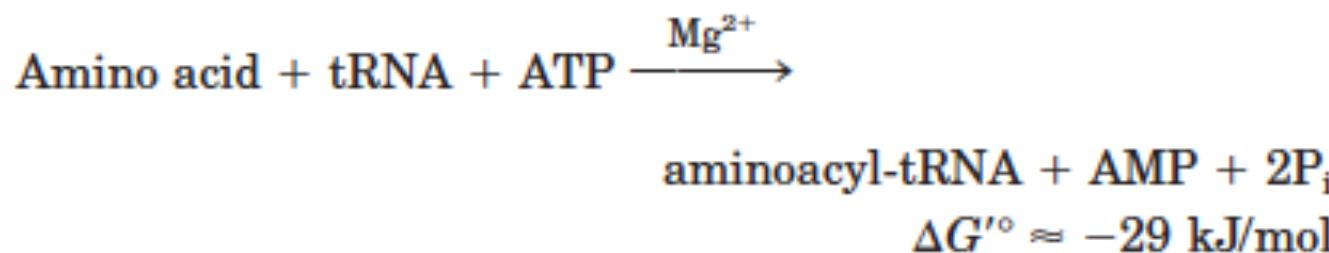


TWISTED "L" STRUCTURE

Stage 1: Attaching the correct amino acid to the correct tRNA

in the cytosol, aminoacyl-tRNA synthetases esterify the 20 amino acids to their corresponding tRNAs.

- ❖ Each enzyme is specific for one amino acid and one or more corresponding tRNAs
- ❖ Most organisms have one aminoacyl-tRNA synthetase for each amino acid
- ❖ For amino acids with two or more corresponding tRNAs, the same enzyme usually aminoacylates all of them



Proof Reading by Aminoacyl tRNA Synthetases

Aminoacylation of tRNA accomplishes

1. Activation of amino acid for peptide bond formation
2. attachment of amino acid to an adaptor tRNA for placement of amino acid

The amino acid attached is not checked on the ribosome!

How is the fidelity assured?

Consider Valine and Isoleucine - different by only – CH₂ –

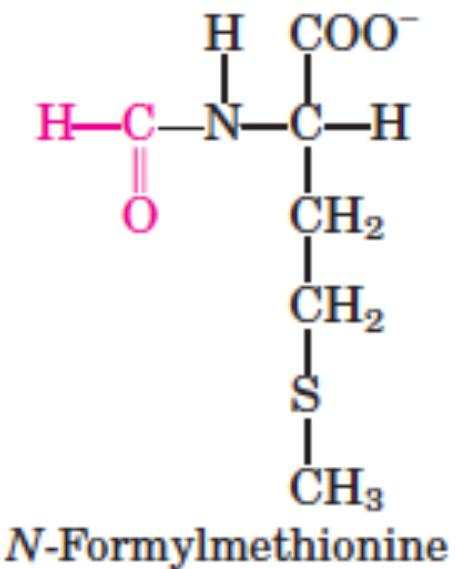
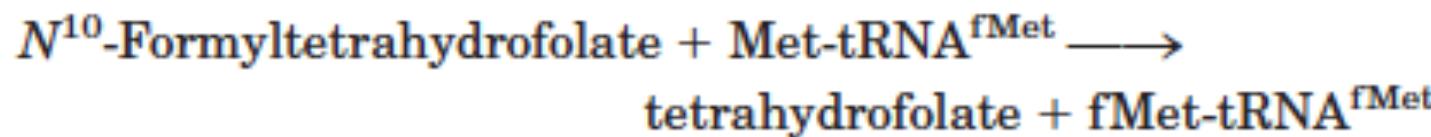
In the case of Ile-tRNA synthetase

1. Activation of Ile is favored by a factor of 200
2. Binding is carried out in 2 steps (acts as filter)
3. Incorrect binding occurs at a second site that has a higher hydrolytic rate
4. In this case, overall process is 1:3000 in favor of the correct amino acid Ile

Stage 2: A Specific Amino Acid Initiates Protein Synthesis

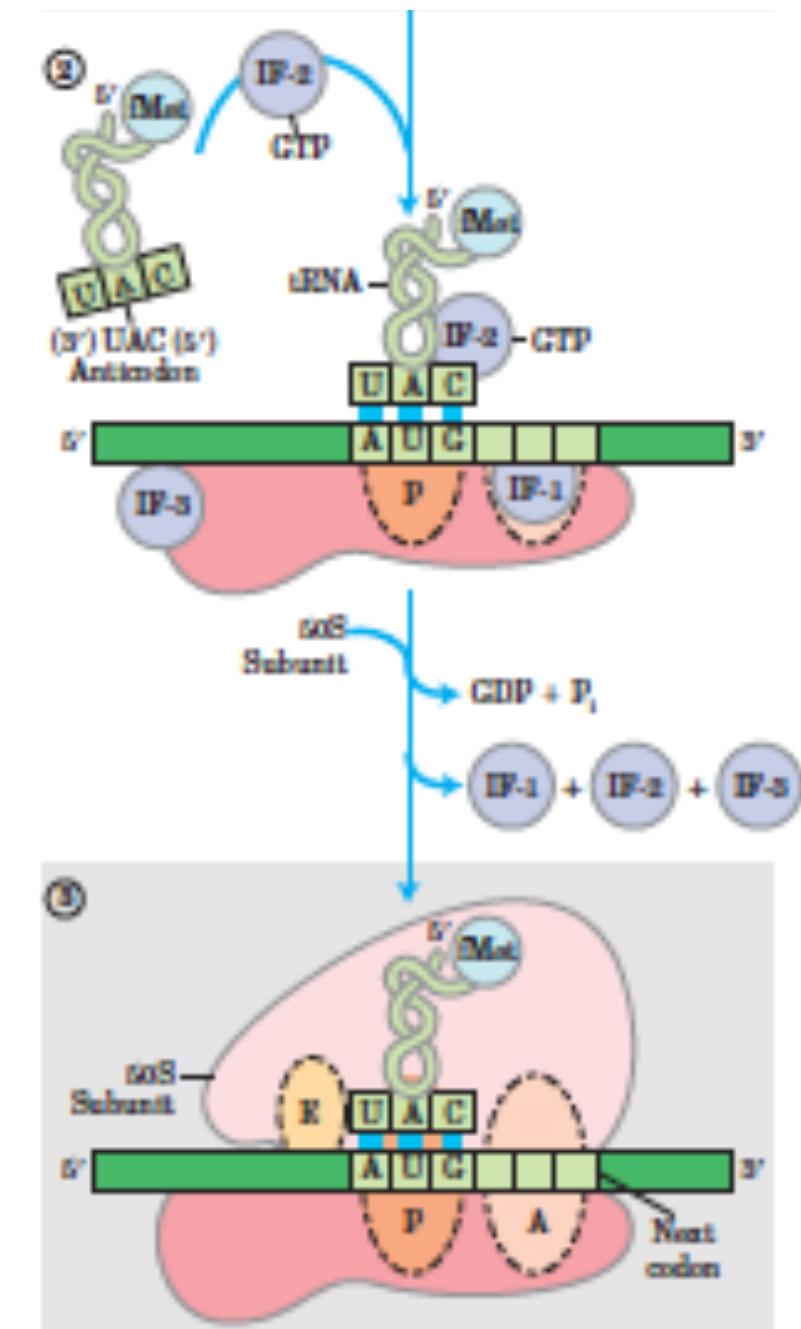
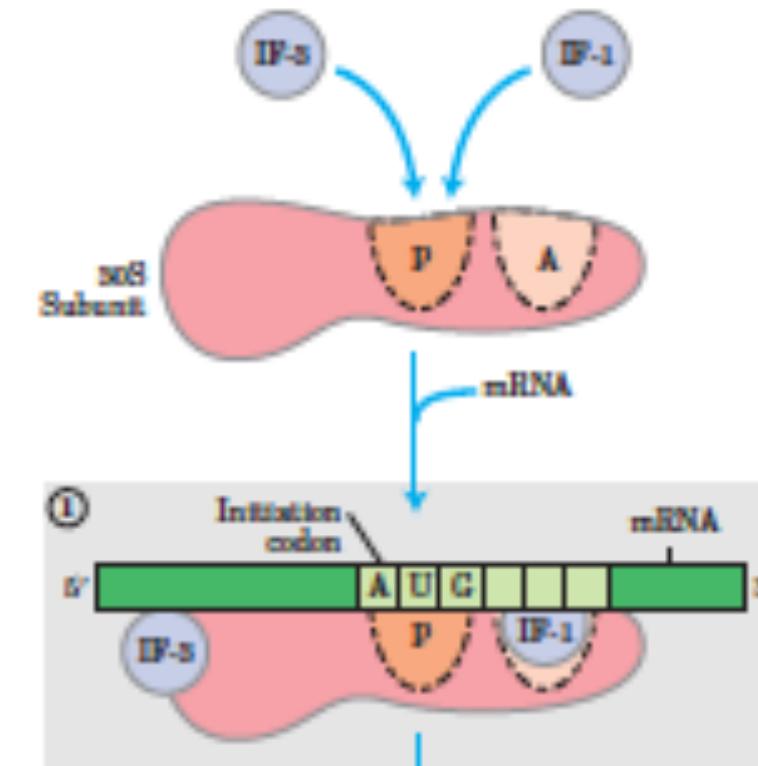
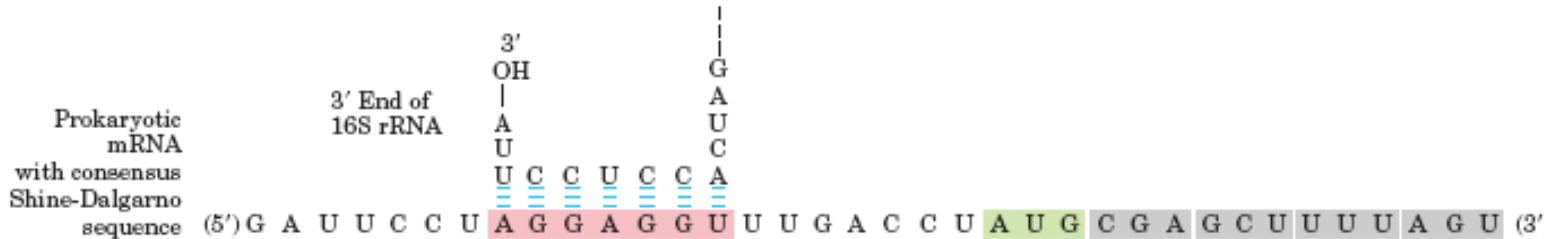
Although methionine has only one codon, (5')AUG, all organisms have two tRNAs for methionine

- ❖ One is used exclusively when (5')AUG is the initiation codon for protein synthesis
- ❖ The other is used to code for a Met residue in an internal position in a polypeptide
- ❖ The amino acid incorporated in response to the (5')AUG initiation codon is N-formylmethionine (fMet)



Formation of the Initiation Complex

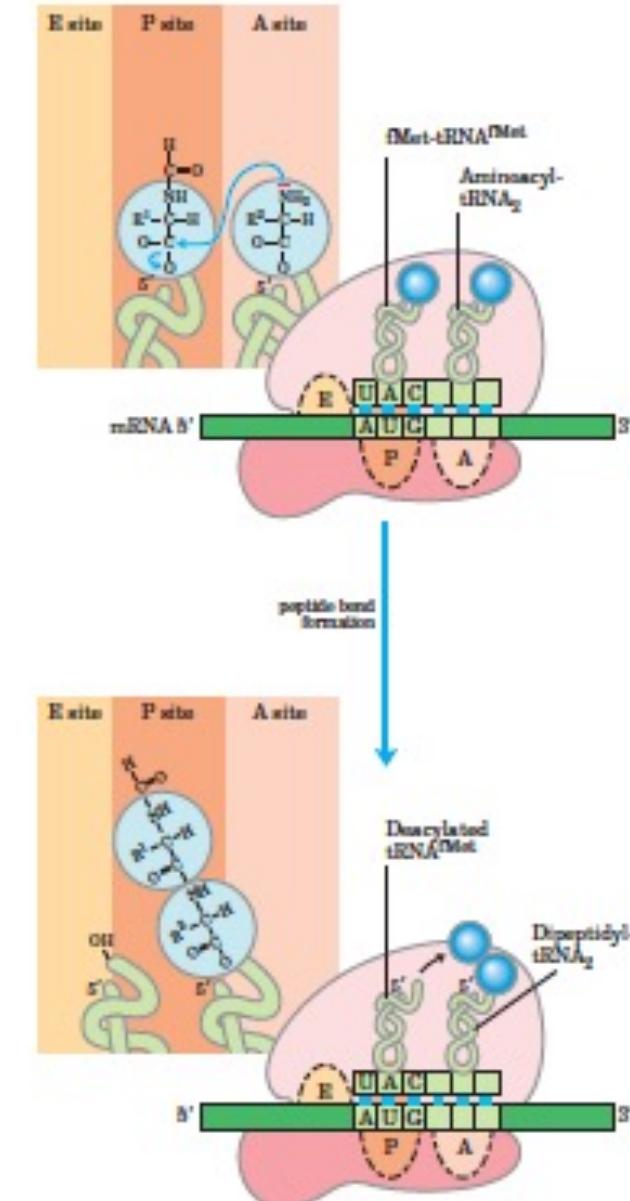
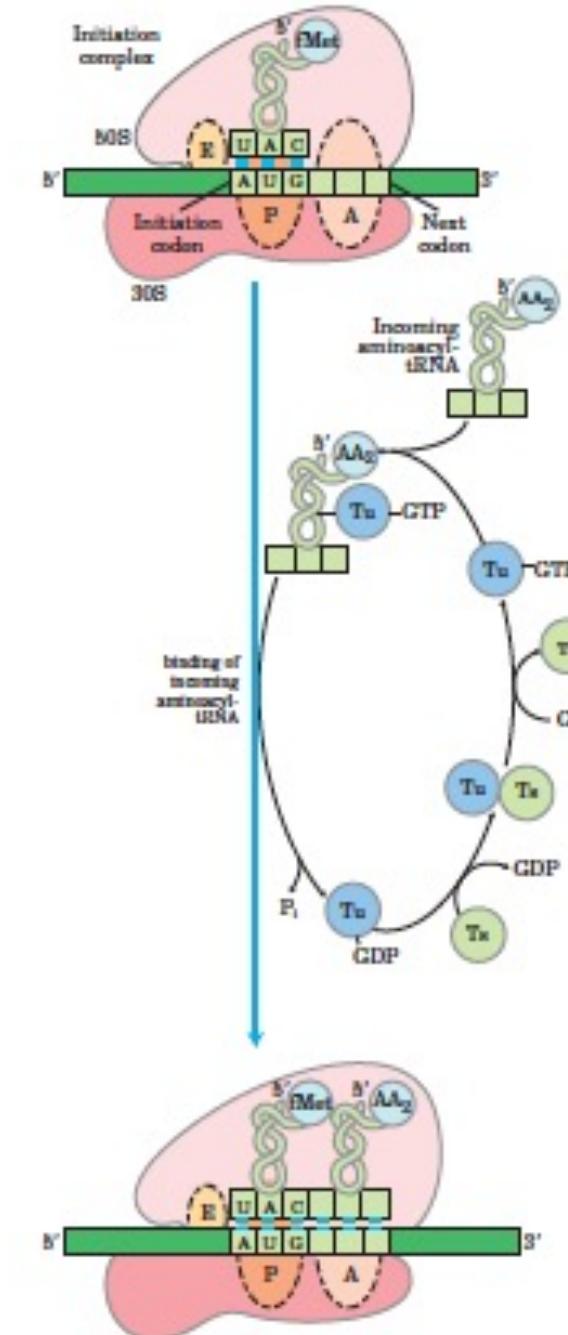
1. 30S ribosomal subunit
 2. mRNA coding for the polypeptide to be made
 3. Initiating fMet-tRN_fAfMet
 4. A set of three proteins called initiation factor (IF-1, IF-2, and IF-3)
 5. GTP
 6. 50S ribosomal subunit
 7. Mg²⁺.



Stage 3: Elongation of the Peptide Chain

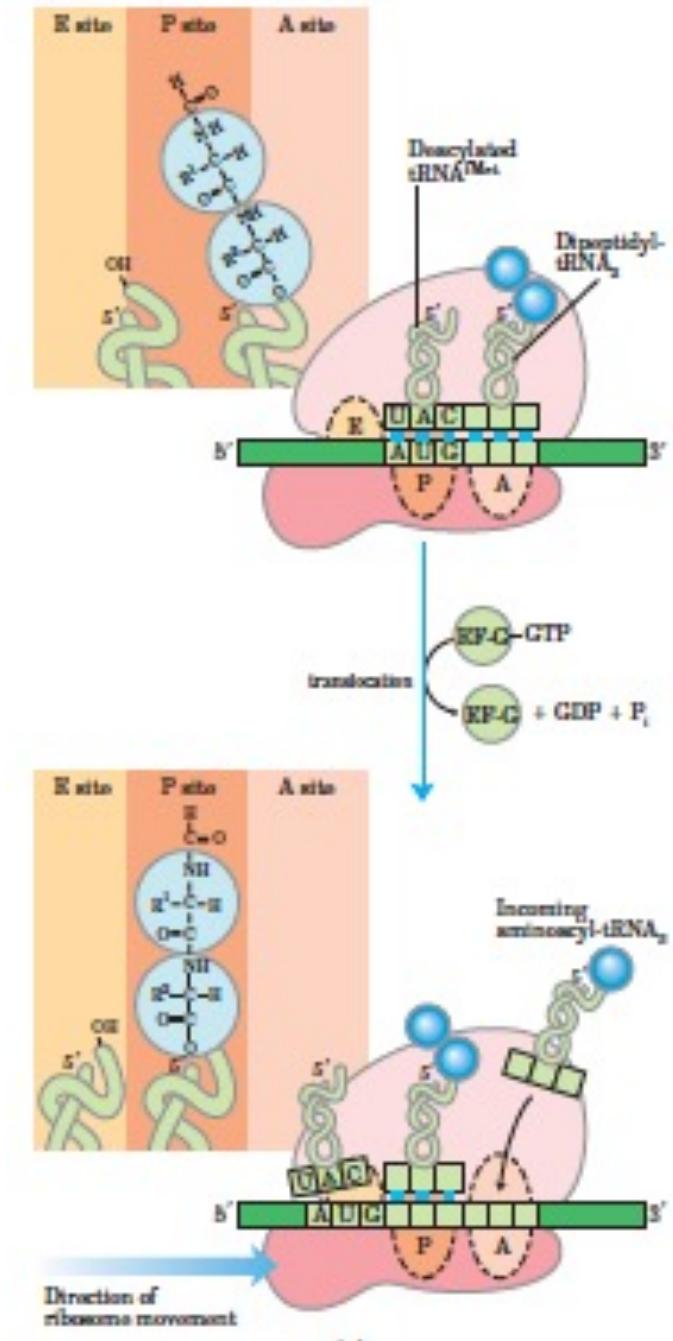
Elongation requires

1. the initiation complex
2. aminoacyl-tRNAs
3. a set of three soluble cytosolic proteins called elongation factors (EF-Tu, EF-Ts, and in bacteria)
4. GTP



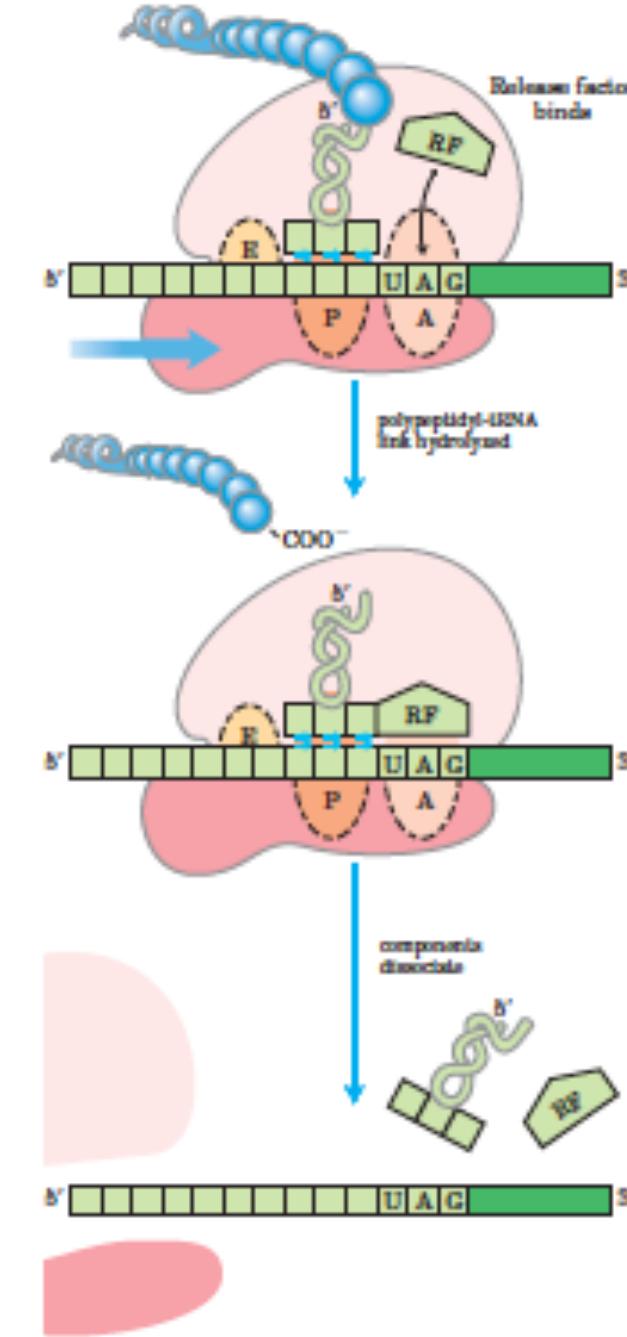
Stage 3 : Translocation

1. The ribosome moves one codon toward the 3' end of the mRNA
This shifts the anticodon of the tRNA attached to the 2nd codon from the A→P site and the deacylated tRNA from P→ E site
2. This movement required EF-G (translocase) and the energy is provided by the hydrolysis GTP→GDP+Pi
3. The uncharged tRNA is dislocated from the E-site and a peptide bond is formed between the growing chain and the new amino acid carried by the tRNA at the A-site



Stage 4: Termination of Synthesis

1. Elongation continues until the last amino acid in the sequence
2. Termination is signaled by the presence of one of the 3 stop codons UAA, UAG or UGA immediately following the final coded amino acid
3. Once the terminal codon occupies A-site, three termination or release factors RF1, RF2 and RF3 contribute to
 - i. Hydrolysis of terminal peptidyl tRNA bond
 - ii. Release of polypeptide from P-site
 - iii. Dissociation of the 70S ribosome into the 30S and 50S subunits



Summary of the 5 stages of protein synthesis

Stage	Essential components
1. Activation of amino acids	20 amino acids 20 aminoacyl-tRNA synthetases 32 or more tRNAs ATP Mg^{2+}
2. Initiation	mRNA <i>N</i> -Formylmethionyl-tRNA ^{fmet} Initiation codon in mRNA (AUG) 30S ribosomal subunit 50S ribosomal subunit Initiation factors (IF-1, IF-2, IF-3) GTP Mg^{2+}
3. Elongation	Functional 70S ribosome (initiation complex) Aminoacyl-tRNAs specified by codons Elongation factors (EF-Tu, EF-Ts, EF-G) GTP Mg^{2+}
4. Termination and release	Termination codon in mRNA Release factors (RF-1, RF-2, RF-3)
5. Folding and posttranslational processing	Specific enzymes, cofactors, and other components for removal of initiating residues and signal sequences, additional proteolytic processing, modification of terminal residues, and attachment of phosphate, methyl, carboxyl, carbohydrate, or prosthetic groups



Review

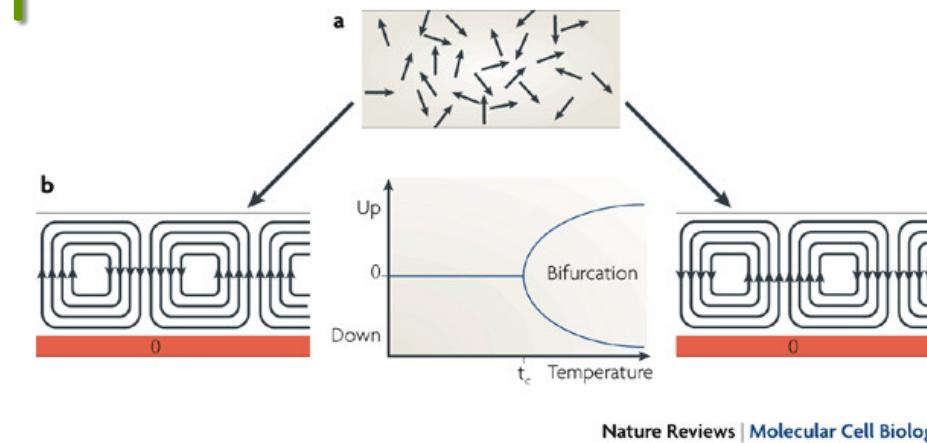
HIGHLIGHTS OF THE IMPORTANT TOPICS OF THE COURSE

Lecture 11

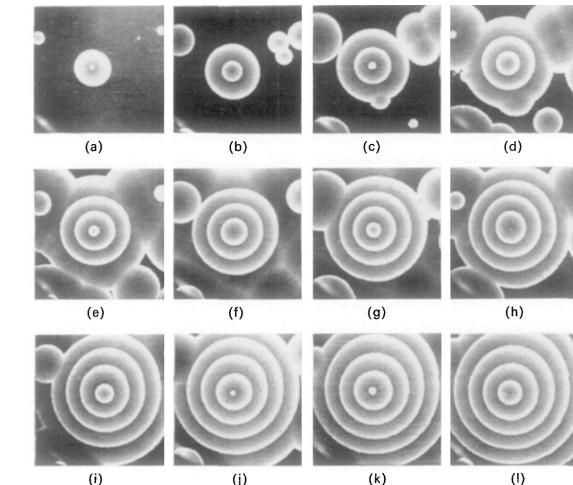
We discussed how principles of self-assembly leads to the notion that it is possible that a living cell could be created under the conditions existing in primordial earth

In-animate Systems

- ❖ Physical systems
- ❖ Chemical systems
- ❖ Biological self-assembled systems

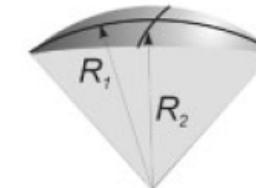
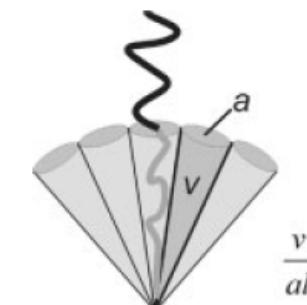


Nature Reviews | Molecular Cell Biology

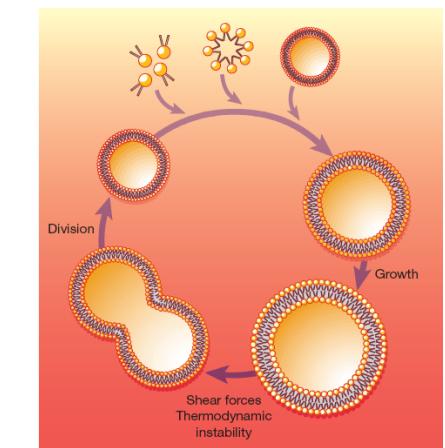


Living Systems

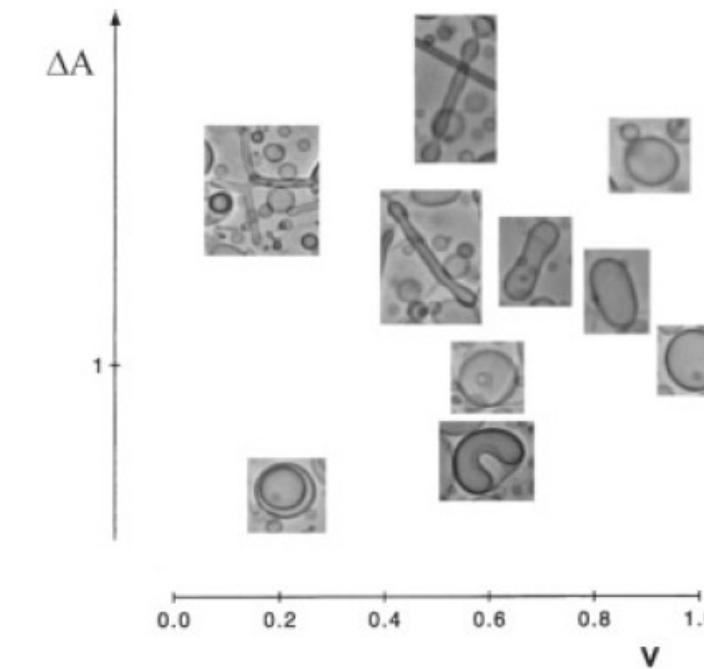
- ❖ The cell



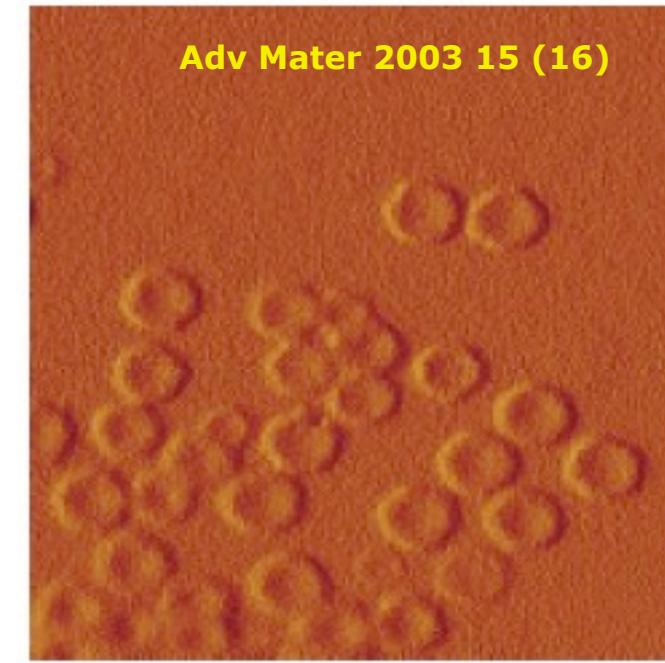
$$K = \frac{1}{R_1 R_2}$$



Vesicle of any desired size and shape can be synthesized under lab conditions



- ❖ Depends on mixing entropies (pull towards many assemblies) and molar bending energies (tends towards a smaller number of vesicles)
- ❖ ΔA is the difference in area, v is the dimensionless volume to area ratio

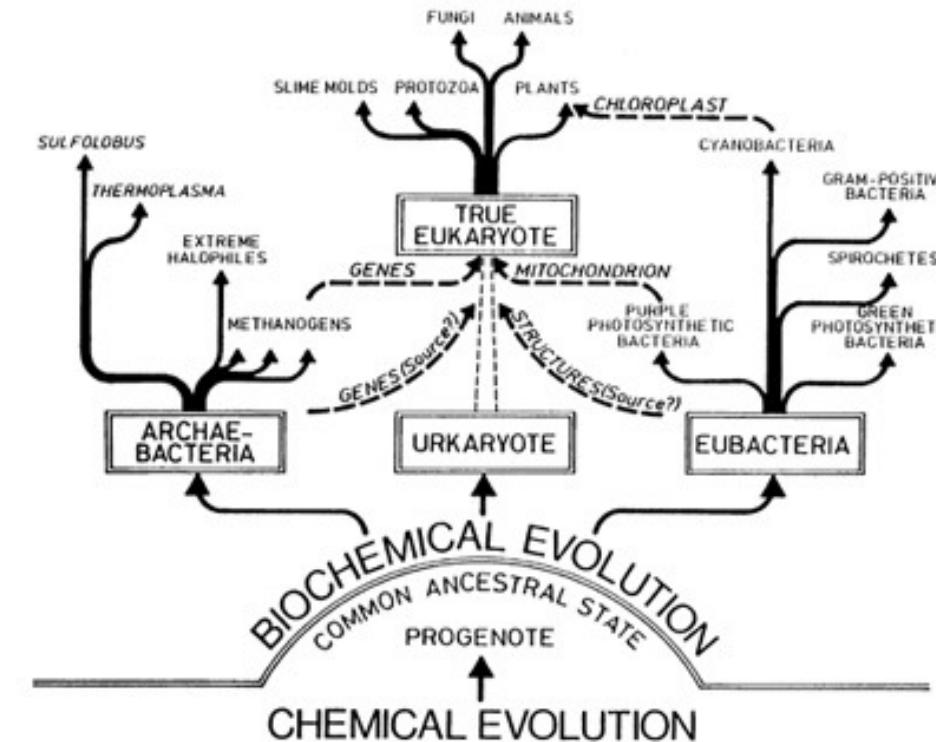


AFM image of a polymer vesicle. The erythrocyte-like shape arises from the evaporation of water from the vesicle interior, leaving the prominent rim

Last Universal Common Ancestor

- ❖ The evolution of the translation apparatus occurred in a series of increasingly complex stages, rather than all at once,
- ❖ The stages subsequent to the establishment of the basic mechanism were concerned by and large with increasing the mechanism's accuracy, and possibly speed as well

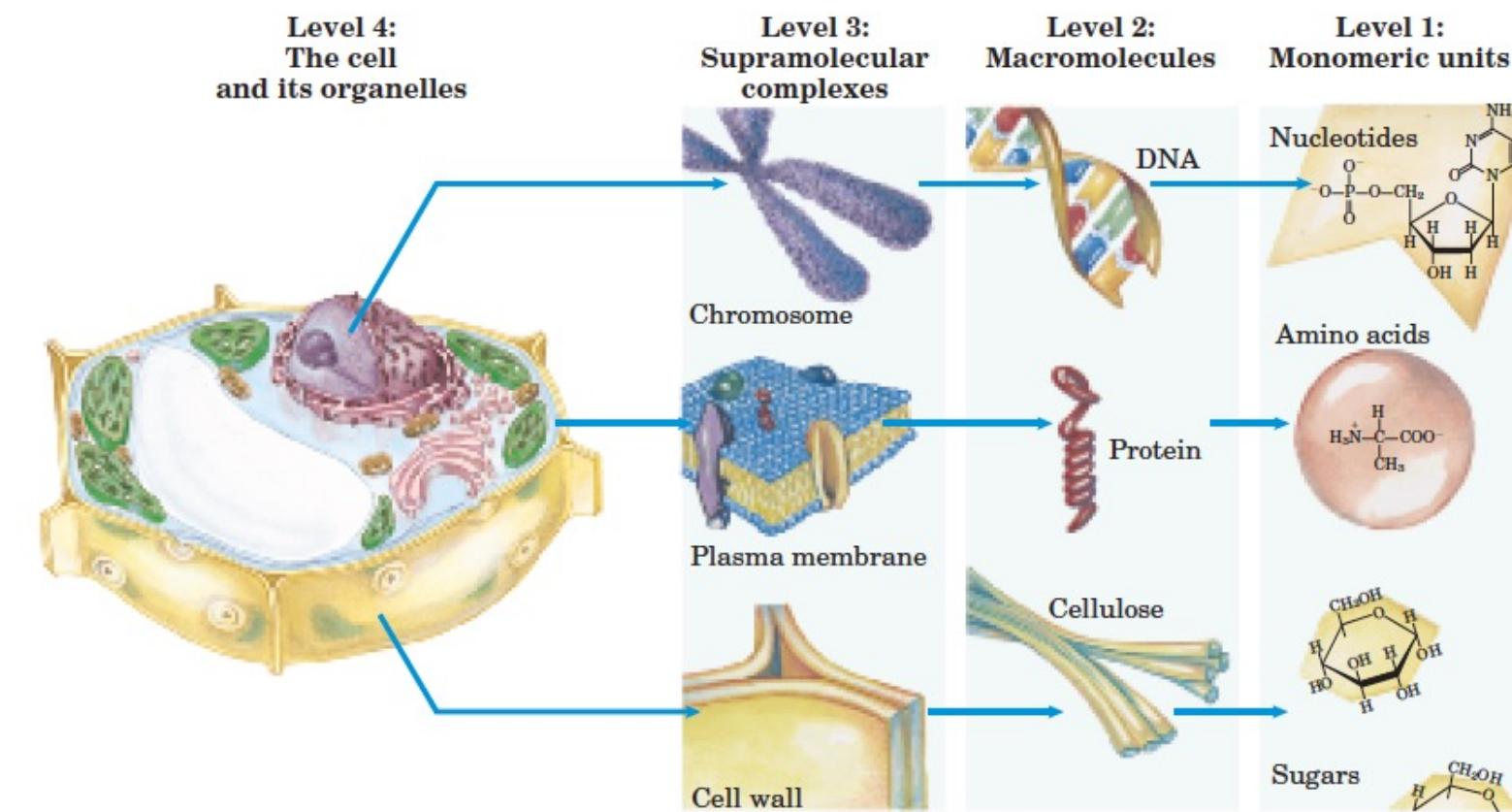
J. Mol Evol 10, 1-6, 1977



Fundamental Units of Life

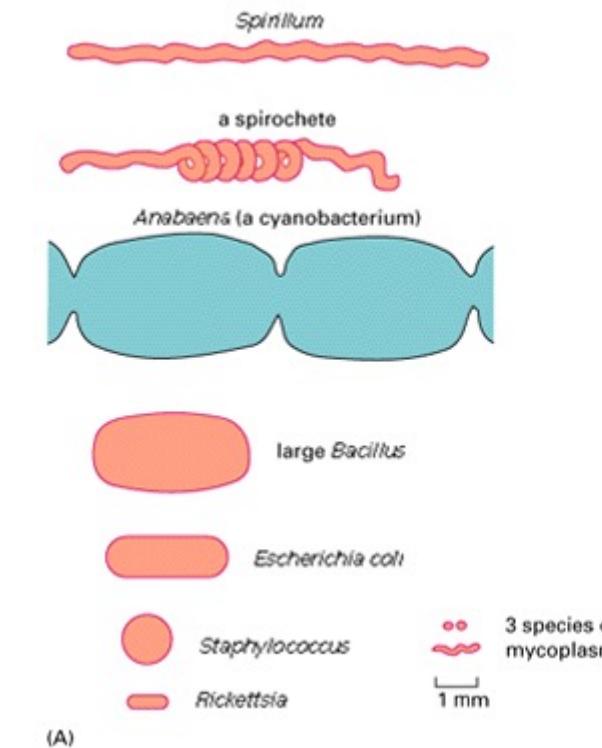
The main biomolecules that are part of the building blocks of cells

- ❖ Carbohydrates and Lipids – cell structure
- ❖ Amino acids – proteins
- ❖ Nucleotides – DNA and RNA



Prokaryotic Cells Are Structurally Simple but Biochemically Diverse

- ❖ Bacteria are the simplest organisms found in most natural environments.
- ❖ Spherical or rod-shaped cells, commonly several micrometers in linear dimension Possess a tough protective coat, cell wall, beneath which a plasma membrane encloses a single cytoplasmic compartment containing DNA, RNA, proteins, and small molecules.



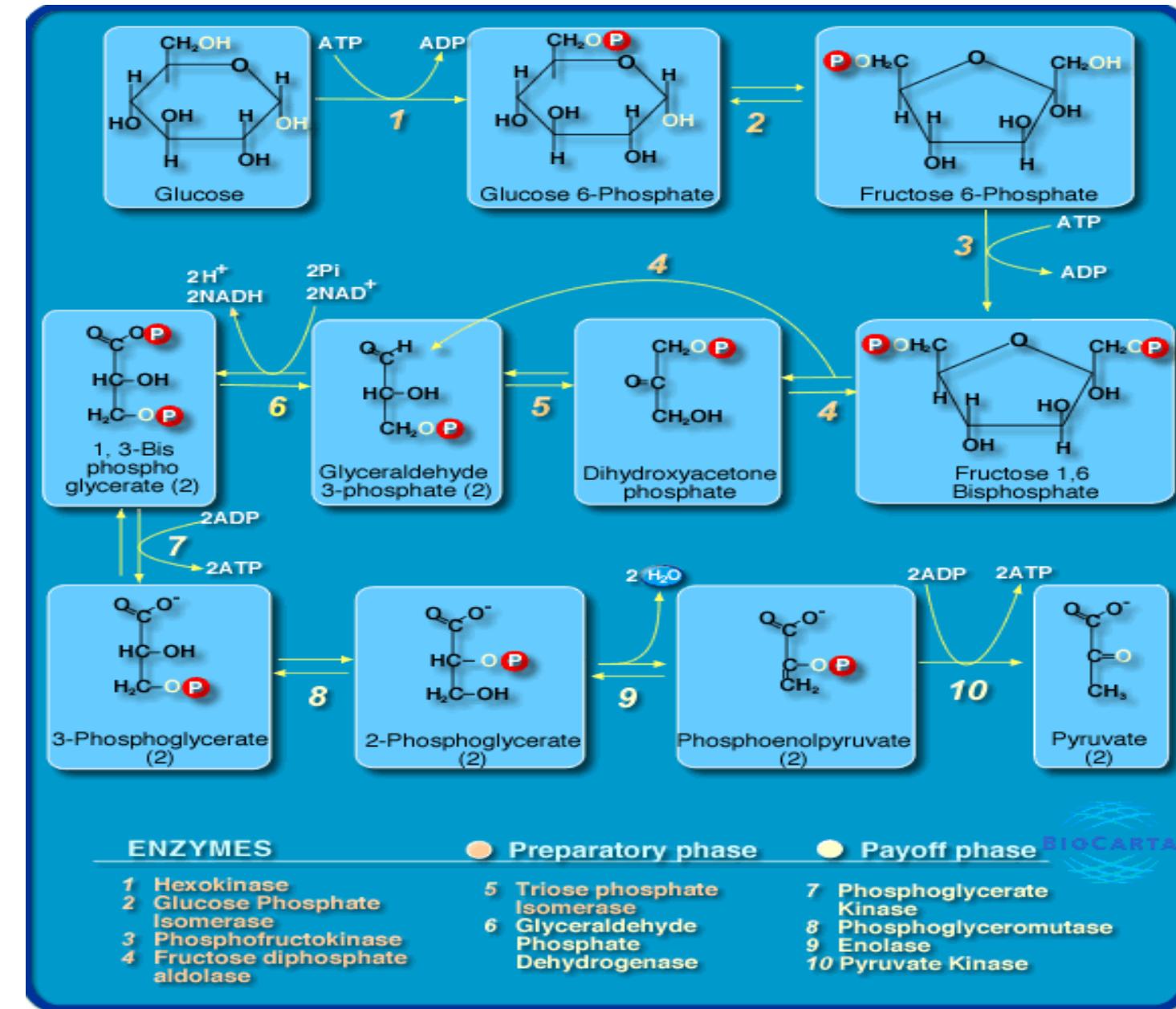
Prokaryote sizes and structures. (A) Some prokaryotic cells drawn to scale. (B) Electron micrograph of a longitudinal section through a bacterium (*Escherichia coli*); the cell's DNA is concentrated in the palely stained region. (Courtesy of E. Kellenberger.)

More about metabolic reactions

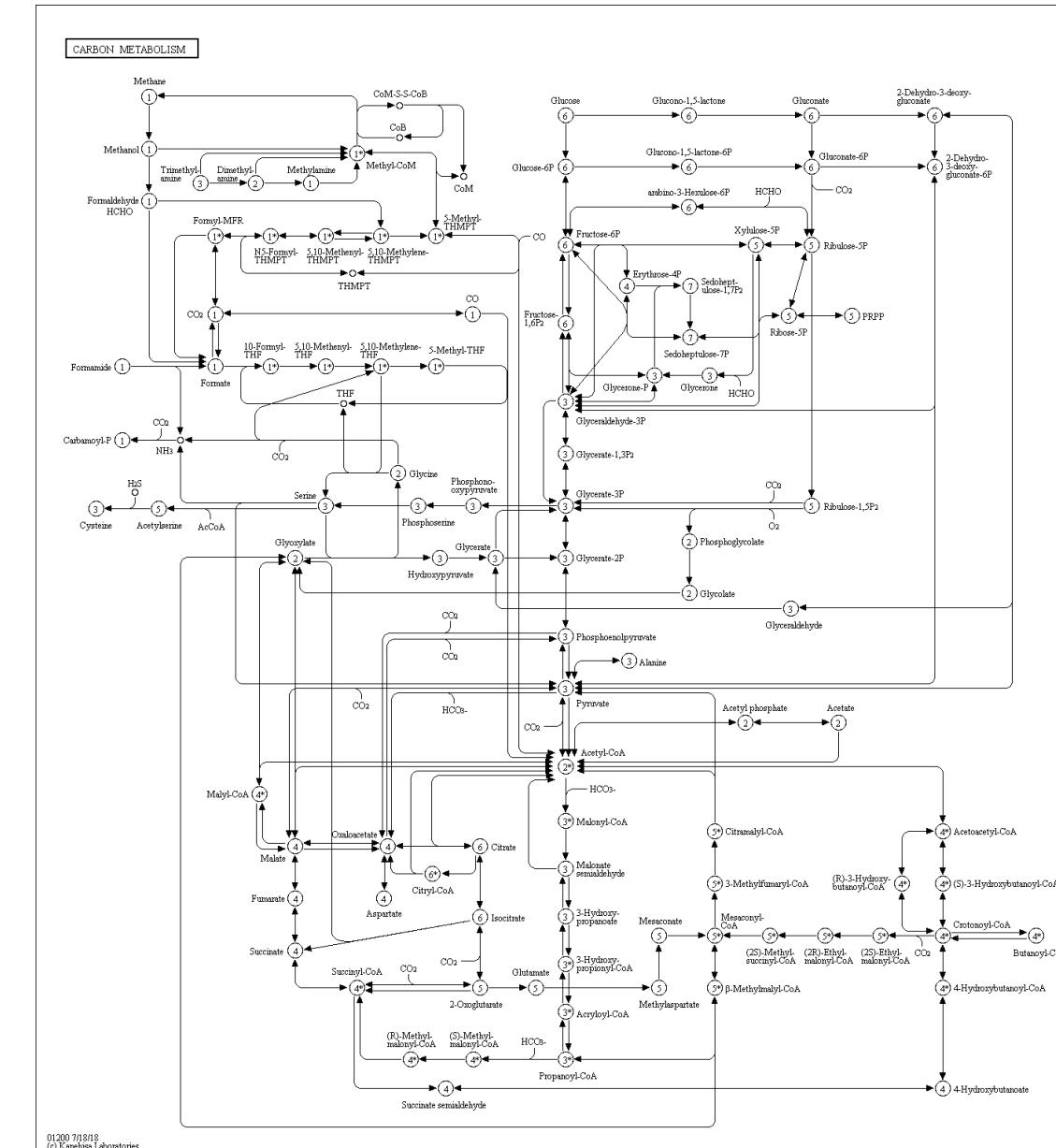
❖ KEGG



- ❖ Similarity in all kinds of organisms, suggesting an extremely ancient origin
- ❖ Linked to core reactions of glycolysis are hundreds of other chemical processes
 - ❖ Generation of energy in ATP-ADP currency
 - ❖ Synthesis of small molecules
 - ❖ Make large polymers specific to the organism
 - ❖ Degrade complex molecules, taken in as food, into simpler chemical units



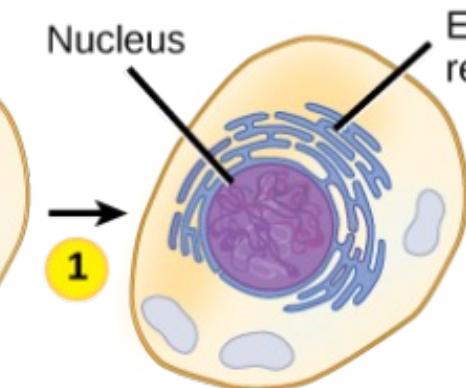
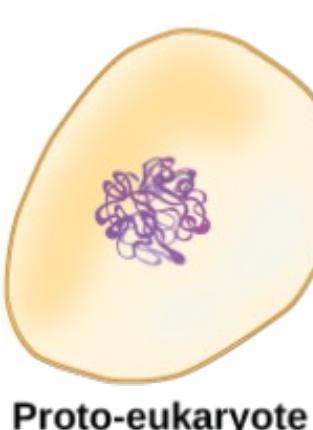
Carbon Metabolism



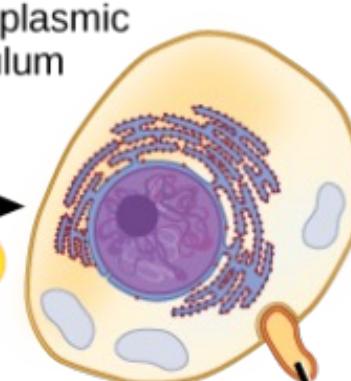
The creation of the eukaryote cell

The ENDOSYMBIOTIC THEORY

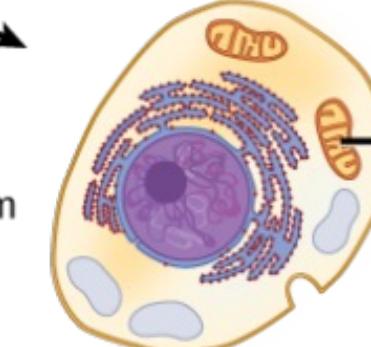
1 Infoldings in the plasma membrane of an ancestral prokaryote gave rise to endomembrane components, including a nucleus and endoplasmic reticulum.



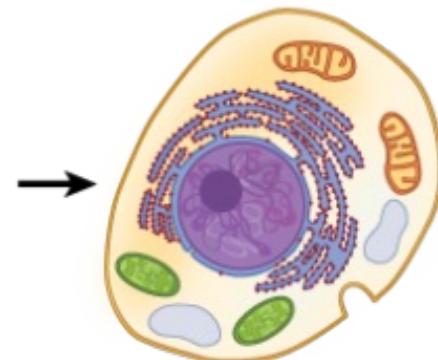
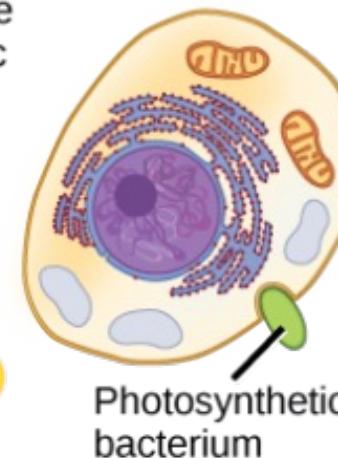
2 In a first endosymbiotic event, the ancestral eukaryote consumed aerobic bacteria that evolved into mitochondria.



3 In a second endosymbiotic event, the early eukaryote consumed photosynthetic bacteria that evolved into chloroplasts.

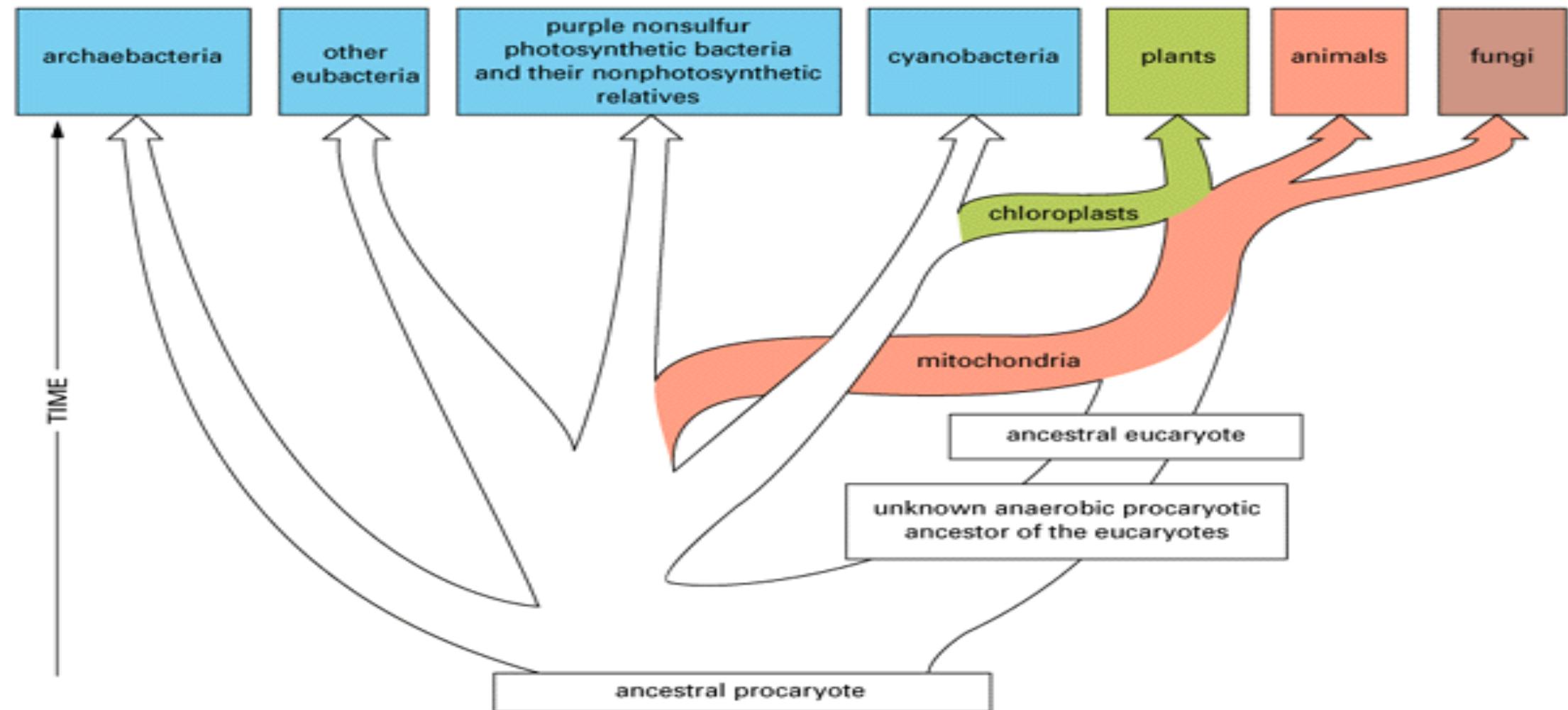


Modern heterotrophic eukaryote



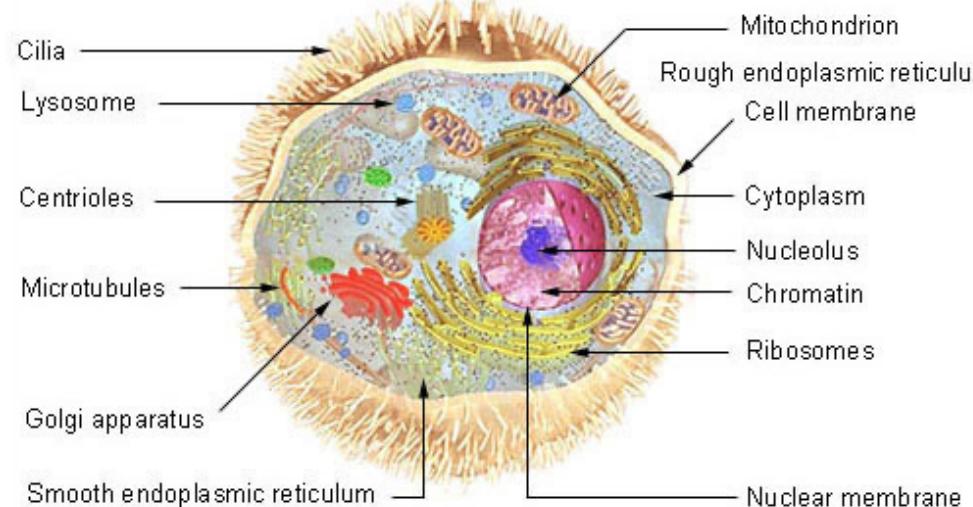
Modern photosynthetic eukaryote

Postulated origin of the eukaryotic cell



Self Assembly and self organization of internal cellular structures

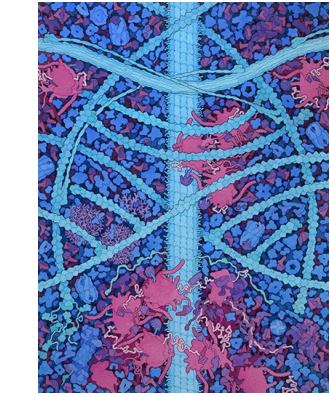
Cell Structure



Flocks of Starlings



Autophagy



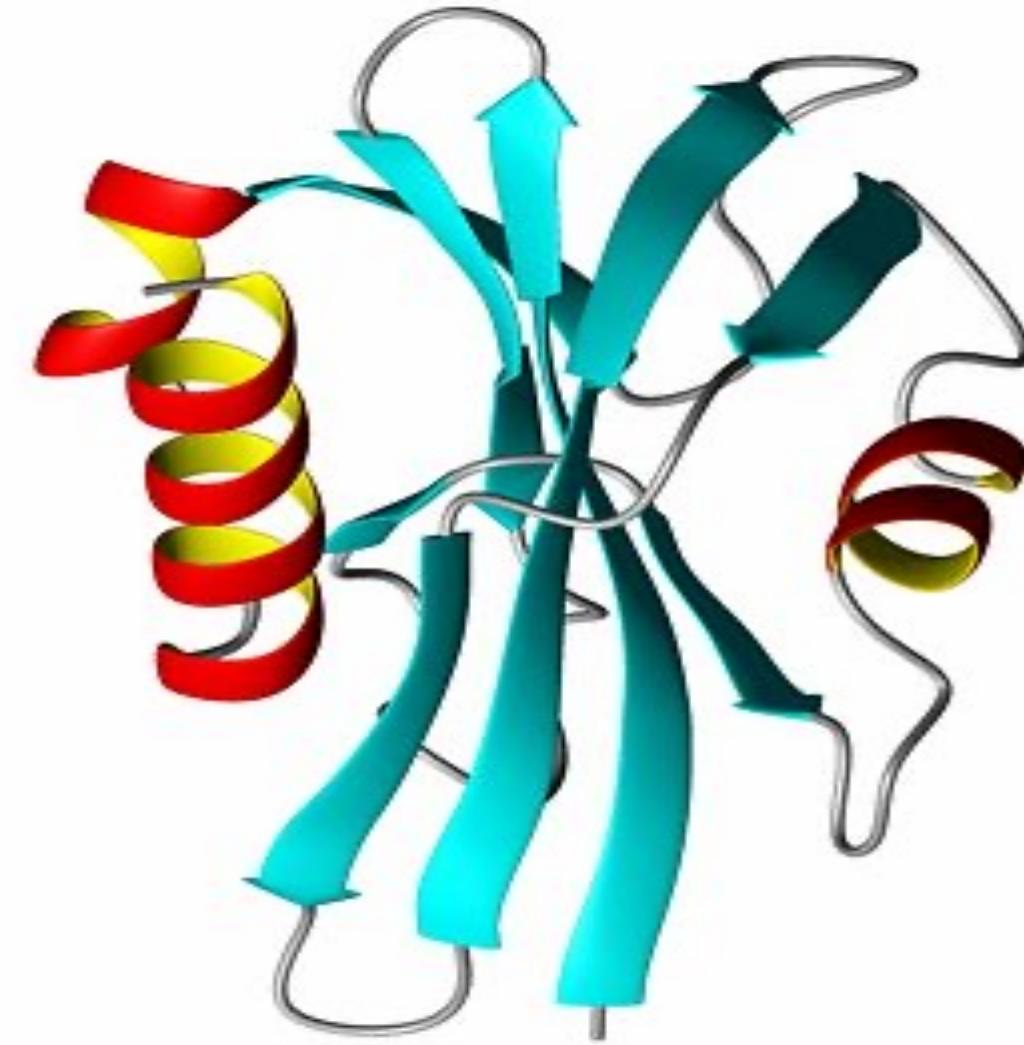
Cytoskeleton

Images from the RCSB site

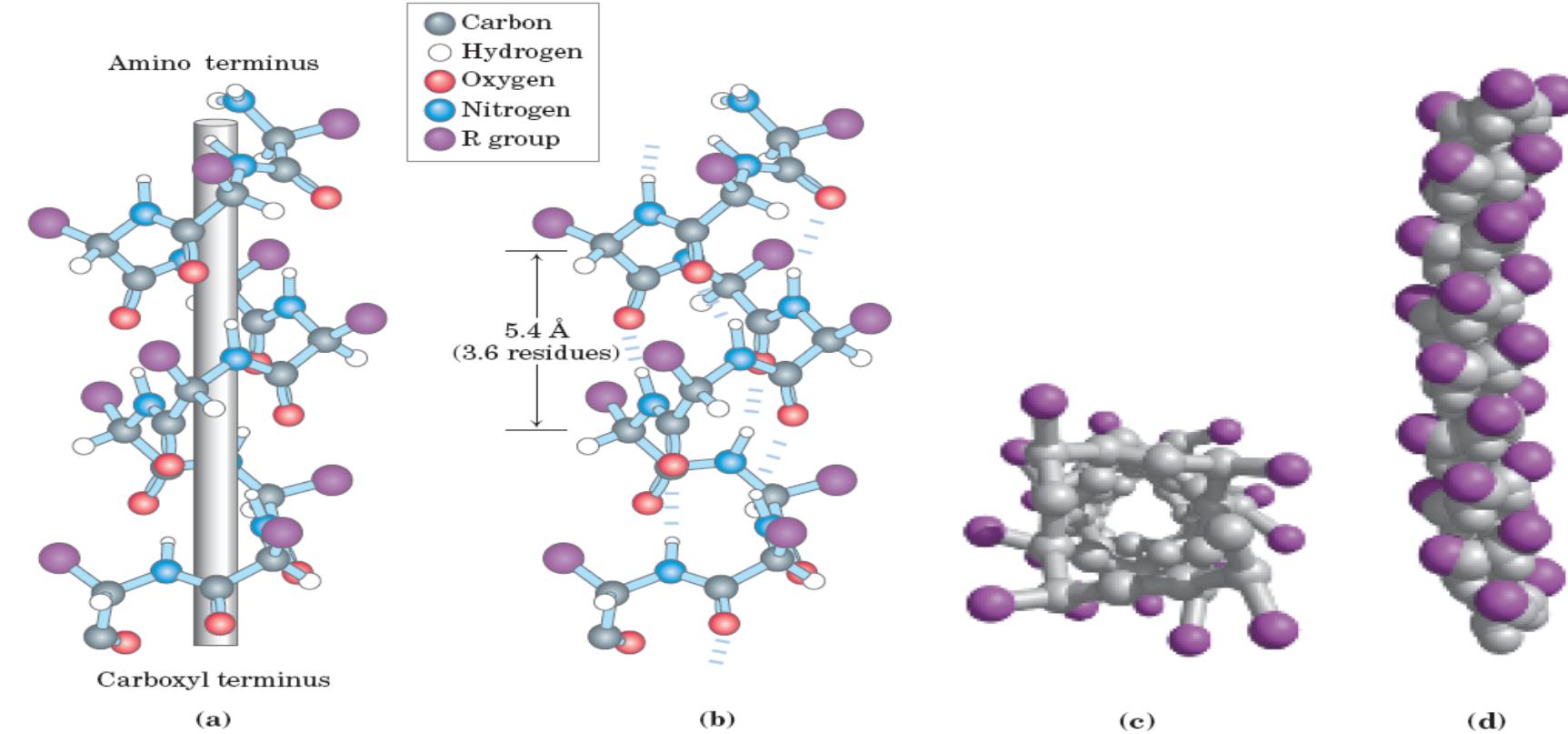
<https://ccsb.scripps.edu/goodsell/>

Protein Folding

- ❖ Secondary Structures
 - ❖ Alpha Helix
 - ❖ Beta Sheet
- ❖ Tertiary Structure (3-D conformation)
- ❖ Quarternary Structure

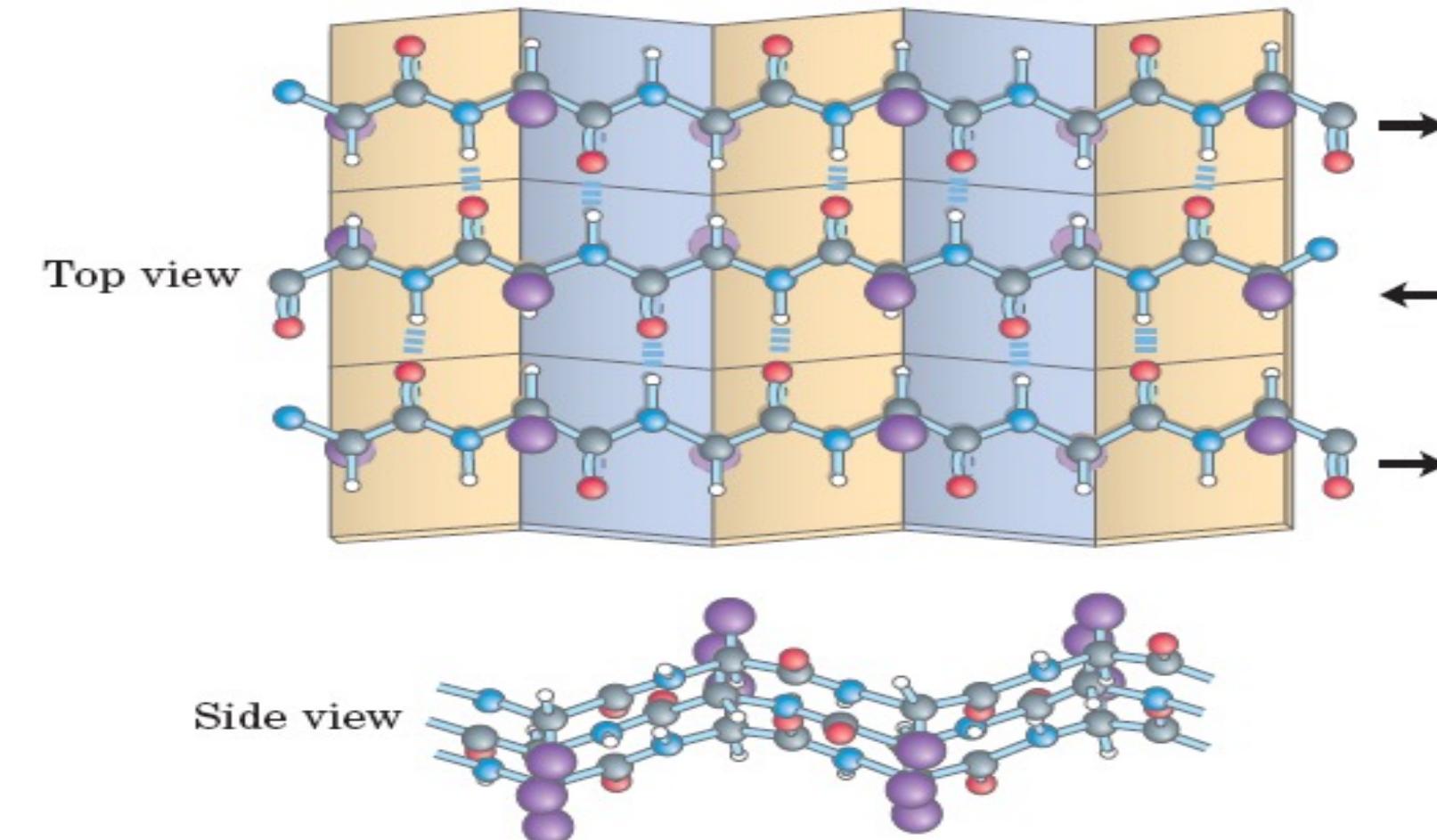


N
D
G
K
L
A
M
T
I
F
W
L
P
D
G
K
D
G

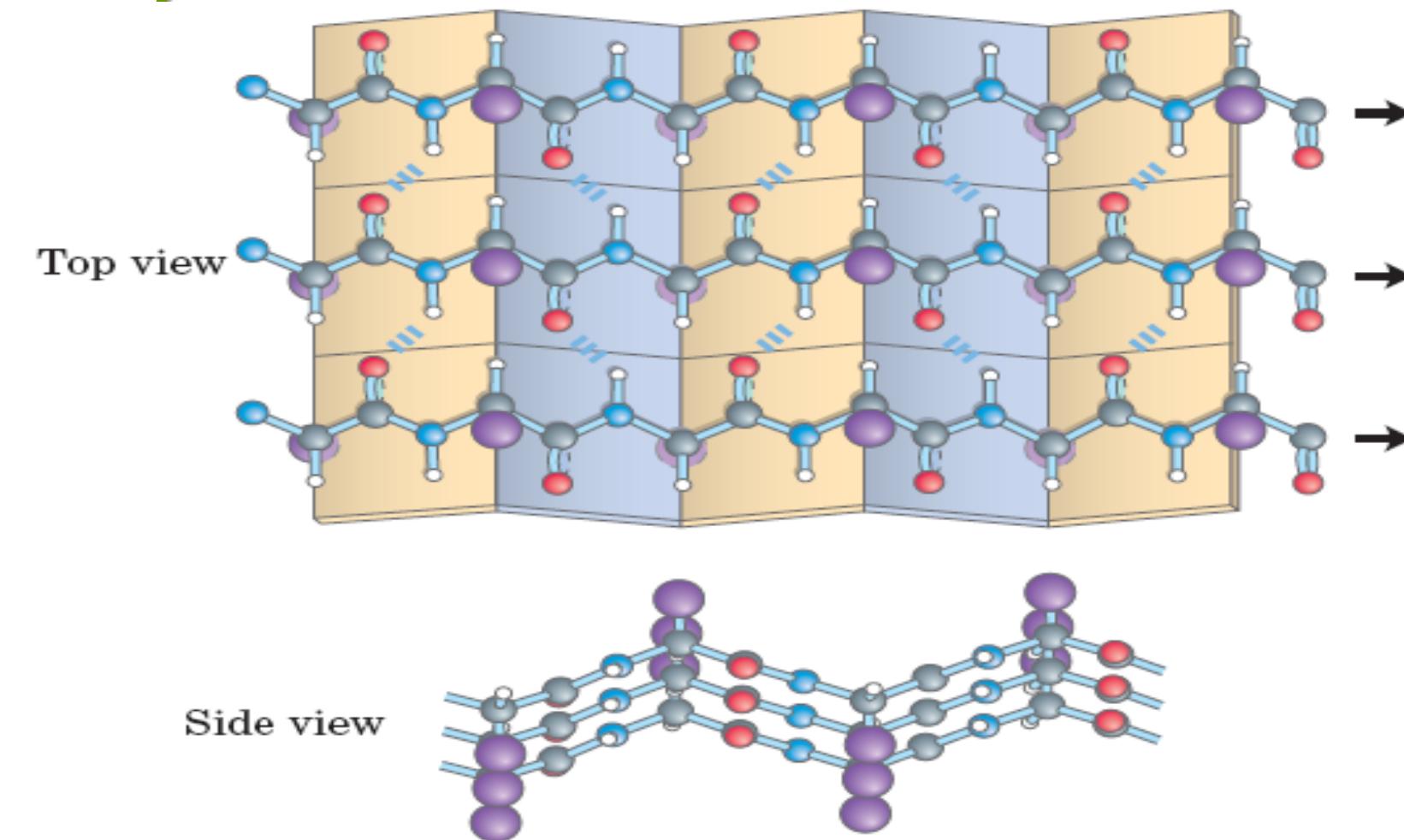


- a) Formation of a right-handed α -helix. The planes of the rigid peptide bonds are parallel to the long axis of the helix, depicted here as a vertical rod
- b) Ball-and-stick model of a right-handed α -helix, showing the intrachain hydrogen bonds. The repeat unit is a single turn of the helix, 3.6 residues
- c) The α -helix as viewed from one end, looking down the longitudinal axis
- d) Atoms in the center of the α -helix are in very close contact

Antiparallel β -Sheet Conformation



Parallel β -Sheet Conformation



Stability of protein structures

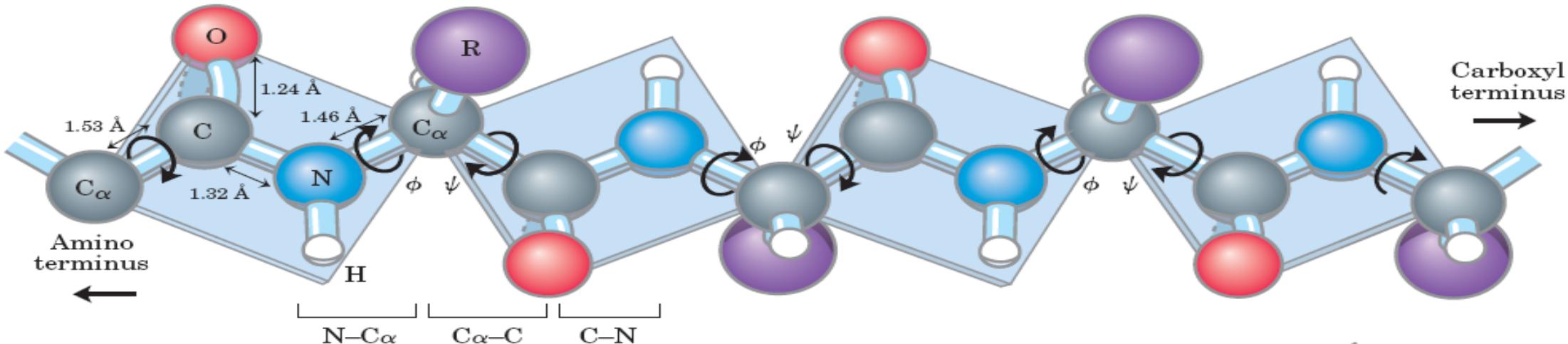
1. The stability of protein structures depends on weak interactions
2. It requires 200-460 kJ/mol to break a single covalent bond, compared to 4-30kJ/mol for weak interactions
 - a) The weak interactions predominate because they are numerous
3. The free energies of the folded and un-folded states are similar

Governing Equations

$$\Delta G = \Delta H - T \Delta S$$

$$\Delta G = -RT\ln k$$

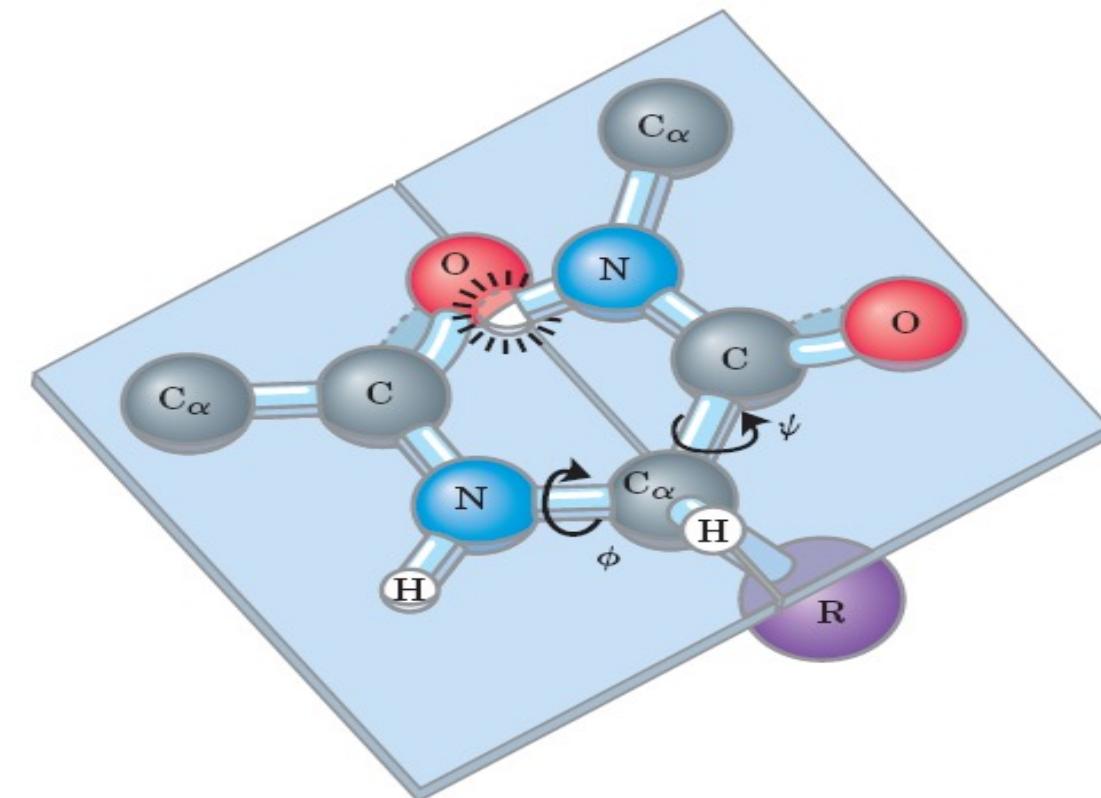
The phi and psi angles



- ❖ Three bonds separate sequential α carbons in a polypeptide chain. The $N-C_{\alpha}$ and $C-C_{\alpha}$ bonds can rotate, with bond angles designated ϕ and ψ , respectively
- ❖ The peptide $C-N$ bond is not free to rotate
- ❖ Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. In the conformation shown, ϕ and ψ are 180 deg (or -180 deg).
- ❖ As one looks out from the α -carbon, the ϕ and ψ angles increase as the carbonyl or amide nitrogens (respectively) rotate clockwise

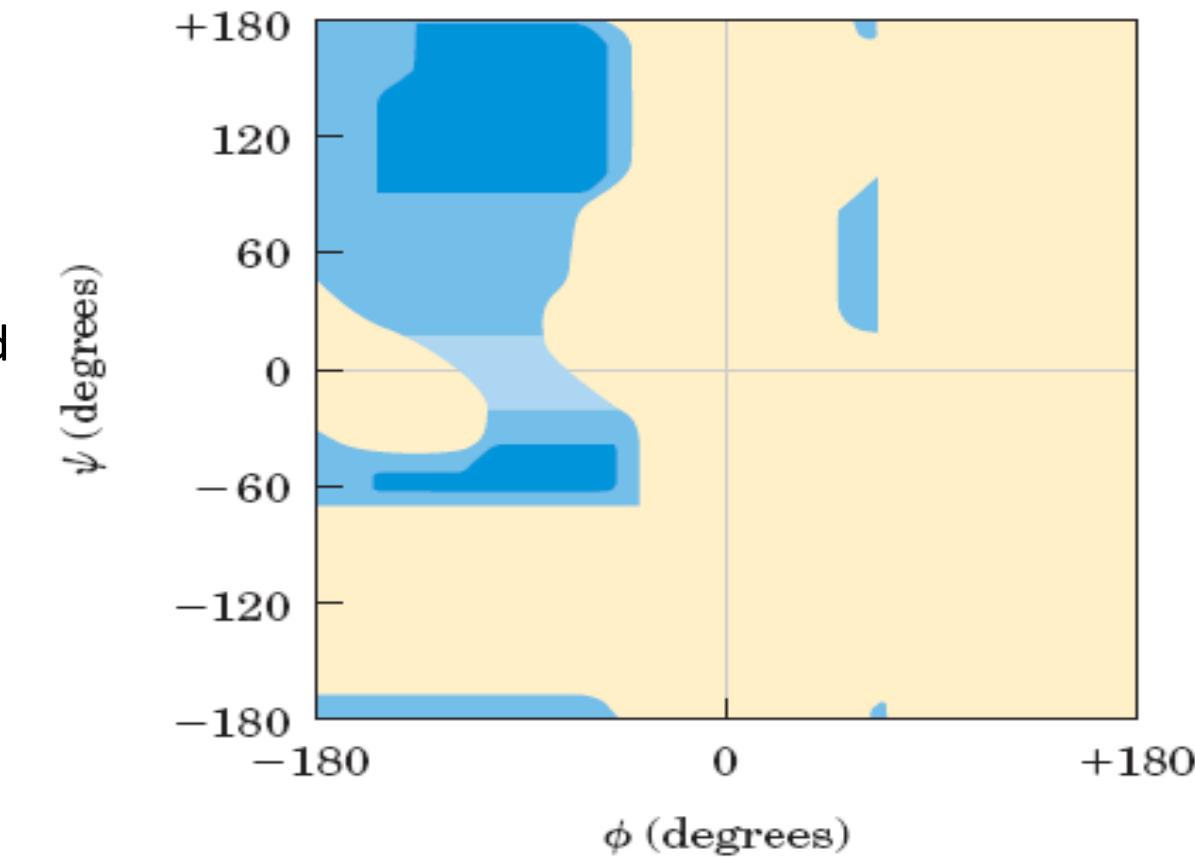
The zero phi & psi angles

- ❖ By convention, both ϕ and ψ are defined as 0 deg when the two peptide bonds flanking that carbon are in the same plane and positioned as shown.
- ❖ In a protein this conformation is prohibited by steric overlap between an – carbonyl oxygen and an – amino hydrogen atom
- ❖ To illustrate the bonds between atoms, the balls representing each atom are smaller than the van der Waals radii for this scale. $1 \text{ \AA} = 0.1 \text{ nm}$.



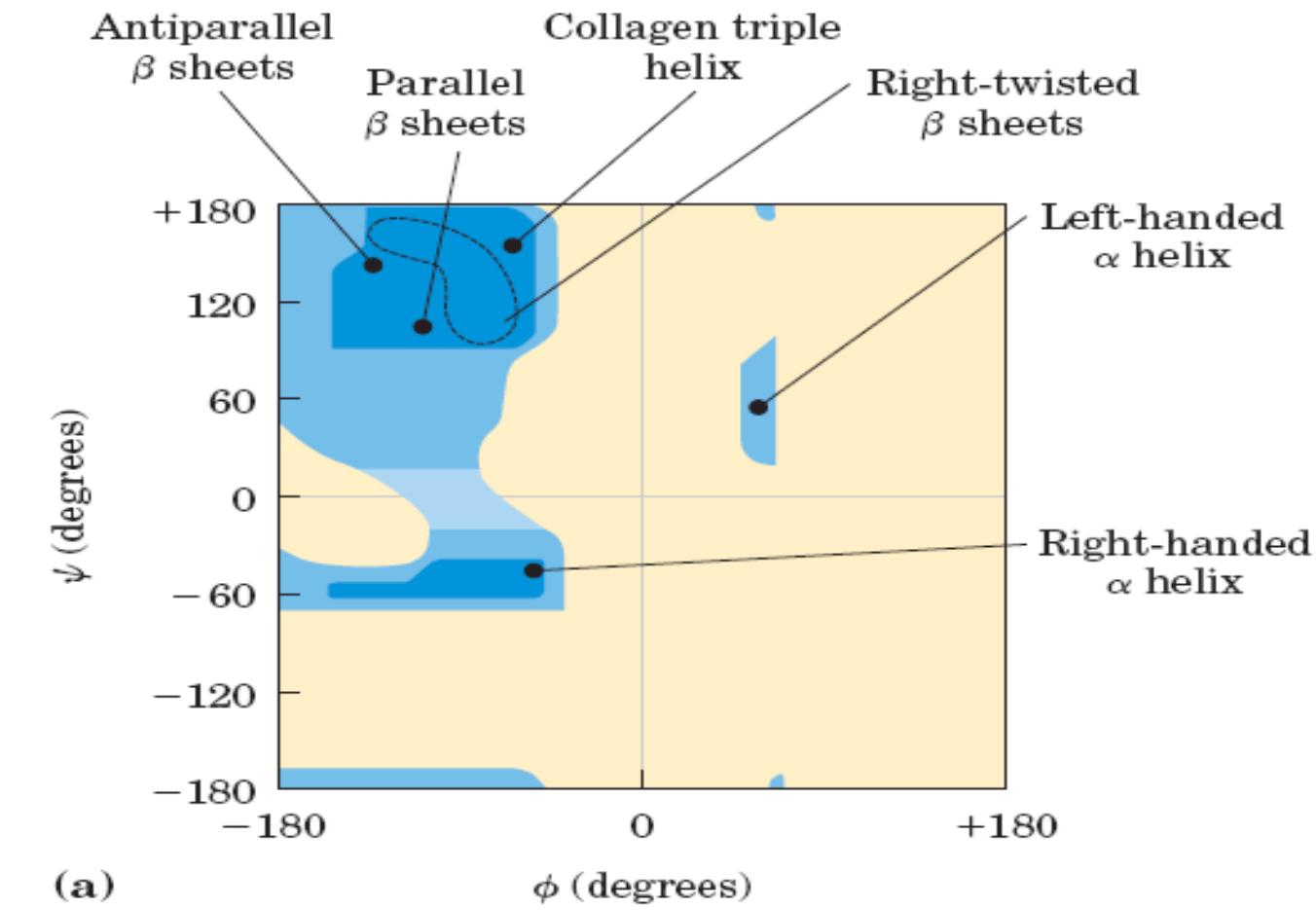
The Ramachandran Plot for L-Ala

- ❖ Conformations deemed possible are those that involve little or no steric interference, based on calculations using known van der Waals radii and bond angles.
 - ❖ The areas shaded dark blue reflect conformations that involve no steric overlap and thus are fully allowed
 - ❖ medium blue indicates conformations allowed at the extreme limits for unfavorable atomic contacts
 - ❖ the lightest blue area reflects conformations that are permissible if a little flexibility is allowed in the bond angles
- ❖ The asymmetry of the plot results from the L stereochemistry of the amino acid residues

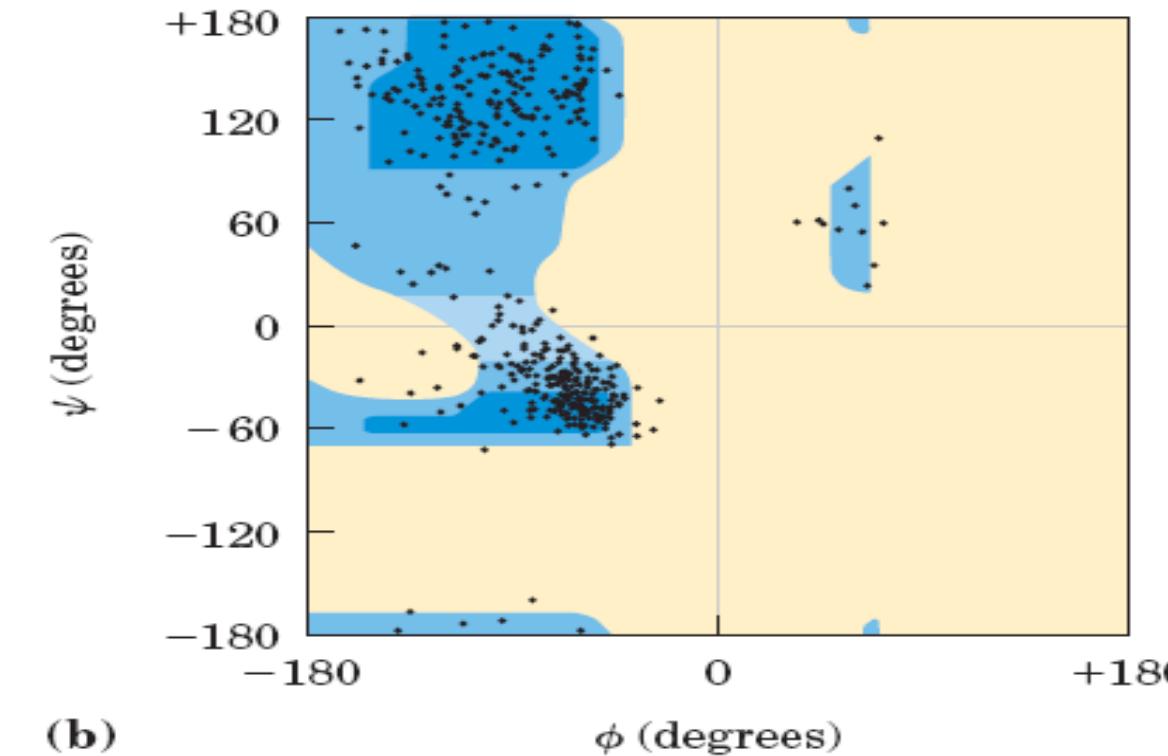


Secondary Protein Structure

- ❖ Secondary structure is the regular arrangement of amino acid residues in a segment of a polypeptide chain, in which each residue is spatially related to its neighbors in the same way
- ❖ The most common secondary structures are the alpha helix, the beta conformation, and beta turns
- ❖ The secondary structure of a polypeptide segment can be completely defined if the phi and psi angles are known for all amino acid residues in that segment



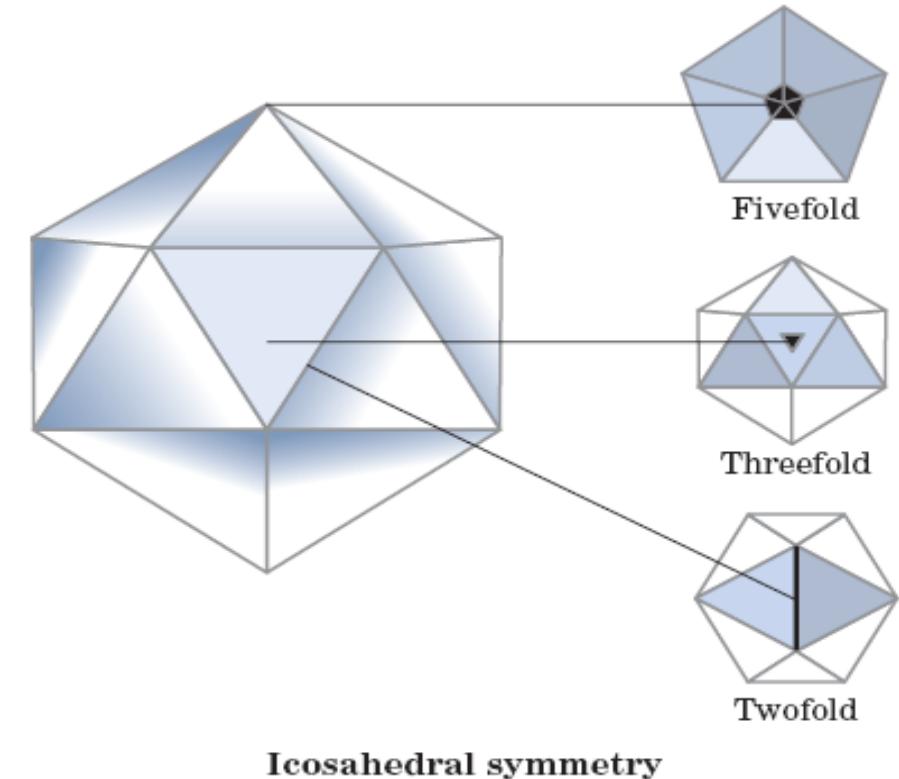
- ❖ The values of ϕ and ψ for various allowed secondary structures are overlaid
- ❖ Although left-handed helices extending over several amino acid residues are theoretically possible, they have not been observed in proteins



- ❖ All the amino acid residues except Gly in the enzyme pyruvate kinase (isolated from rabbit) are overlaid on the plot of theoretically allowed conformations. The small, flexible Gly residues were excluded because they frequently fall outside the expected ranges

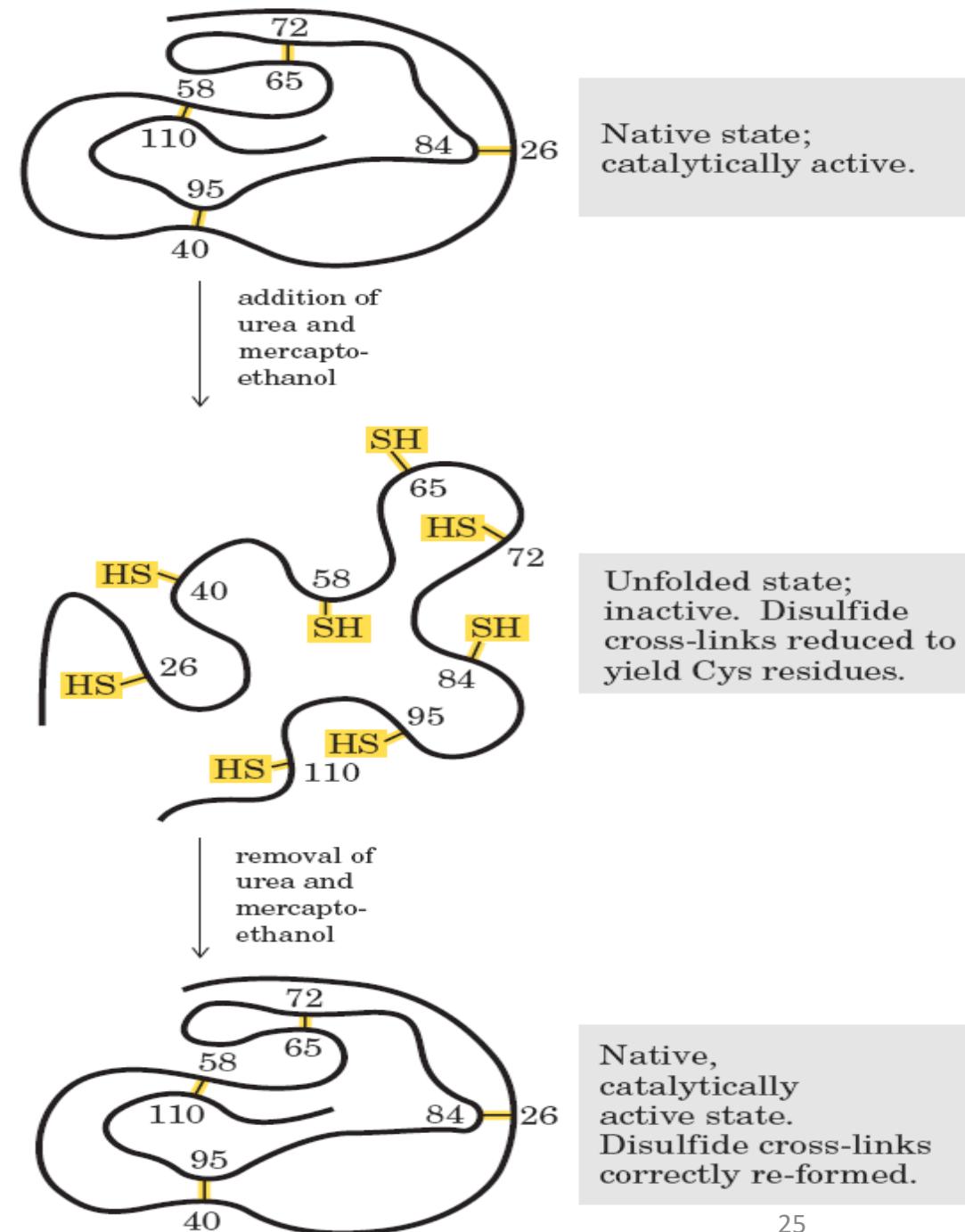
Symmetry in Tertiary Structures: Icosahedral Symmetry

- ❖ Icosahedral symmetry. Relating all 20 triangular faces of an icosahedron requires rotation about one or more of three separate rotational axes: twofold, threefold, and fivefold.

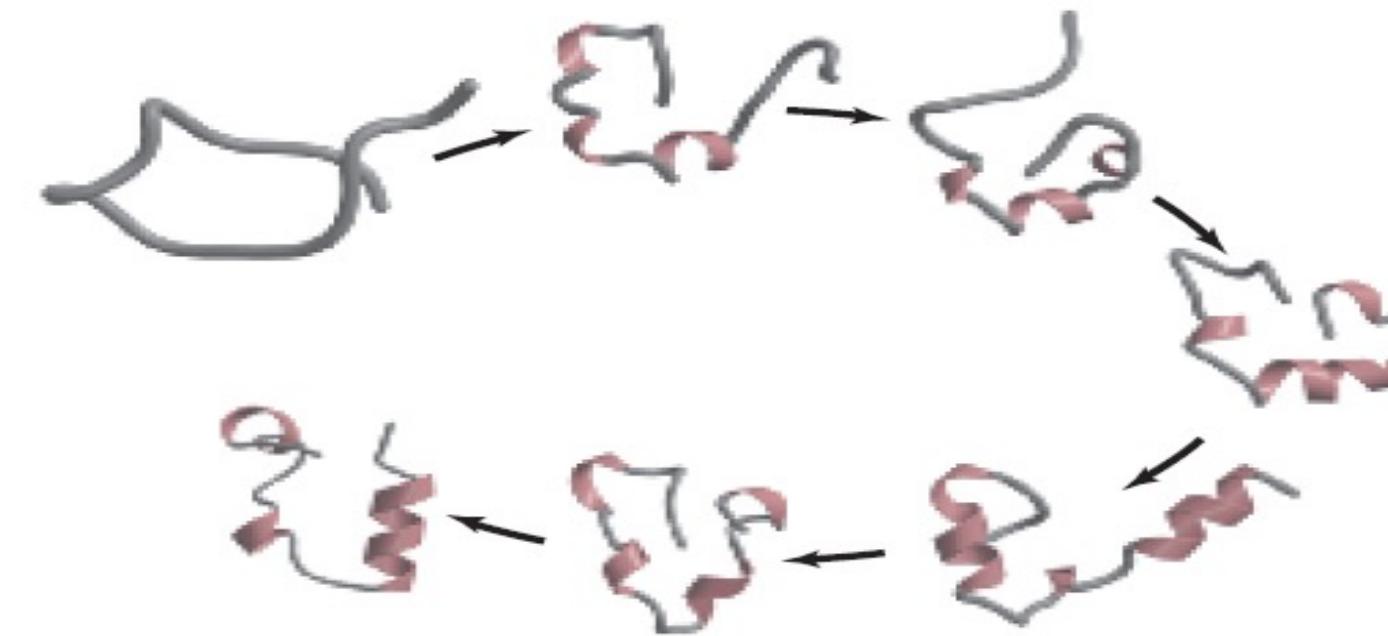


Renaturation of unfolded Protein

- ❖ Urea is used to denature ribonuclease, and mercaptoethanol ($\text{HOCH}_2\text{CH}_2\text{SH}$) to reduce and thus cleave the disulfide bonds to yield eight Cys residues. Renaturation involves reestablishment of the correct disulfide cross-links.



Protein folding simulation

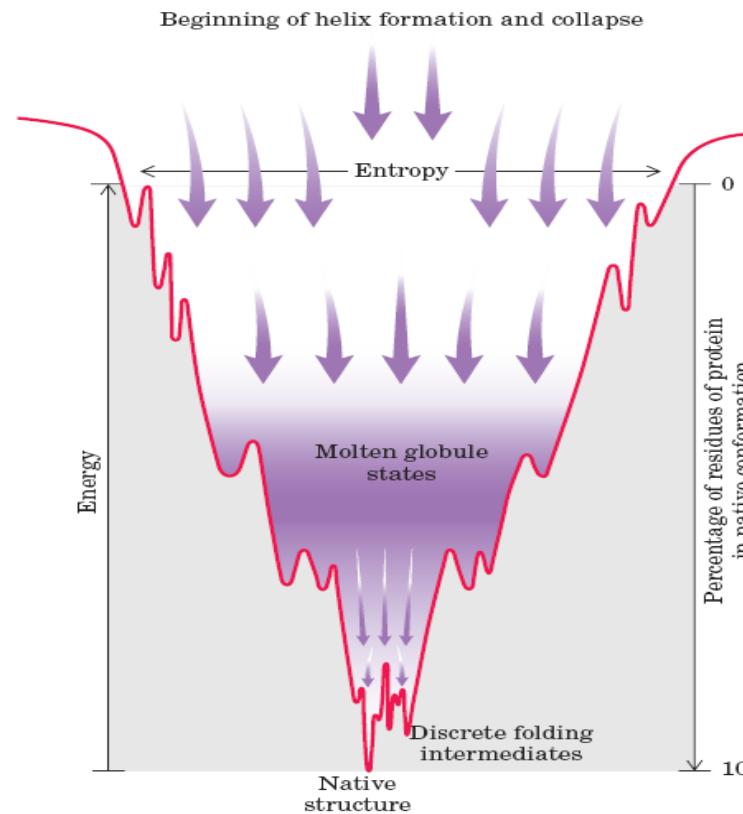


- ❖ The process started with the randomly coiled peptide and 3,000 surrounding water molecules in a virtual “water box.” The molecular motions of the peptide and the effects of the water molecules were taken into account in mapping the most likely paths to the final structure among the countless alternatives.
- ❖ 1 ms simulation time took half a billion integration steps on 2 Cray Computers

Thermodynamics of Protein Folding

WHAT IS THE LEVINTHAL'S PARADOX?

2 MODELS OF FOLDING



1. The number of conformations, and hence the conformational entropy, is large. Only a small fraction of the intramolecular interactions that will exist in the native conformation are present.
2. As folding progresses, the thermodynamic path down the funnel reduces the number of states present (decreases entropy), increases the amount of protein in the native conformation, and decreases the free energy.
3. Depressions on the sides of the funnel represent semistable folding intermediates, which may, in some cases, slow the folding process.

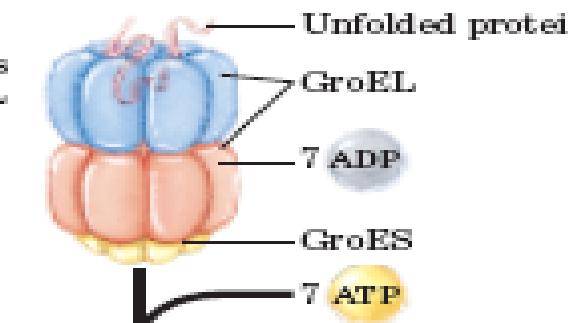
PROTEIN MISFOLDING IS THE CAUSE OF MANY FATAL DISEASES

- ❖ Spongiform encephalopathies – brain gets riddled with holes
 - ❖ Creutzfeldt-Jacob disease – fatal illness; symptoms – dementia, loss of coordination
- ❖ Stanley Prusiner traced the cause to PRION (*proteinaceous infectious only*) protein (PrP)
 - ❖ Two form – normal cellular form (PrP or PrP^C) and altered conformation "scrapie" form (PrP^{Sc})

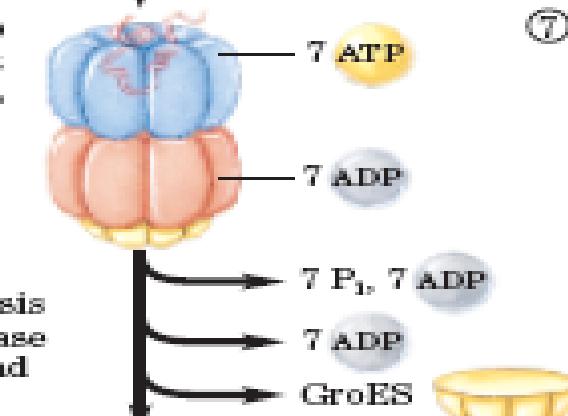
Cells have “Chaperone” proteins to assist proteins in folding correctly or refold proteins that have been denatured by heat or other conditions

GroEL-GroES – member of HSP 60 family

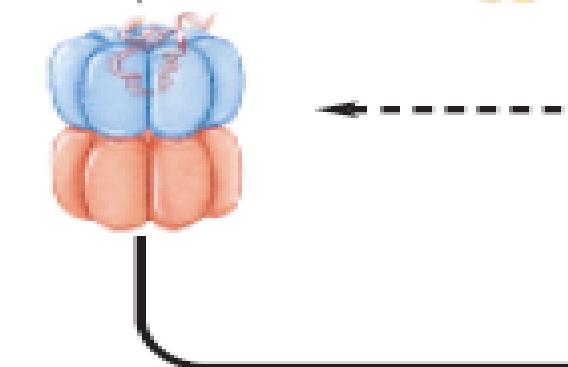
① Unfolded protein binds to the GroEL pocket not blocked by GroES.



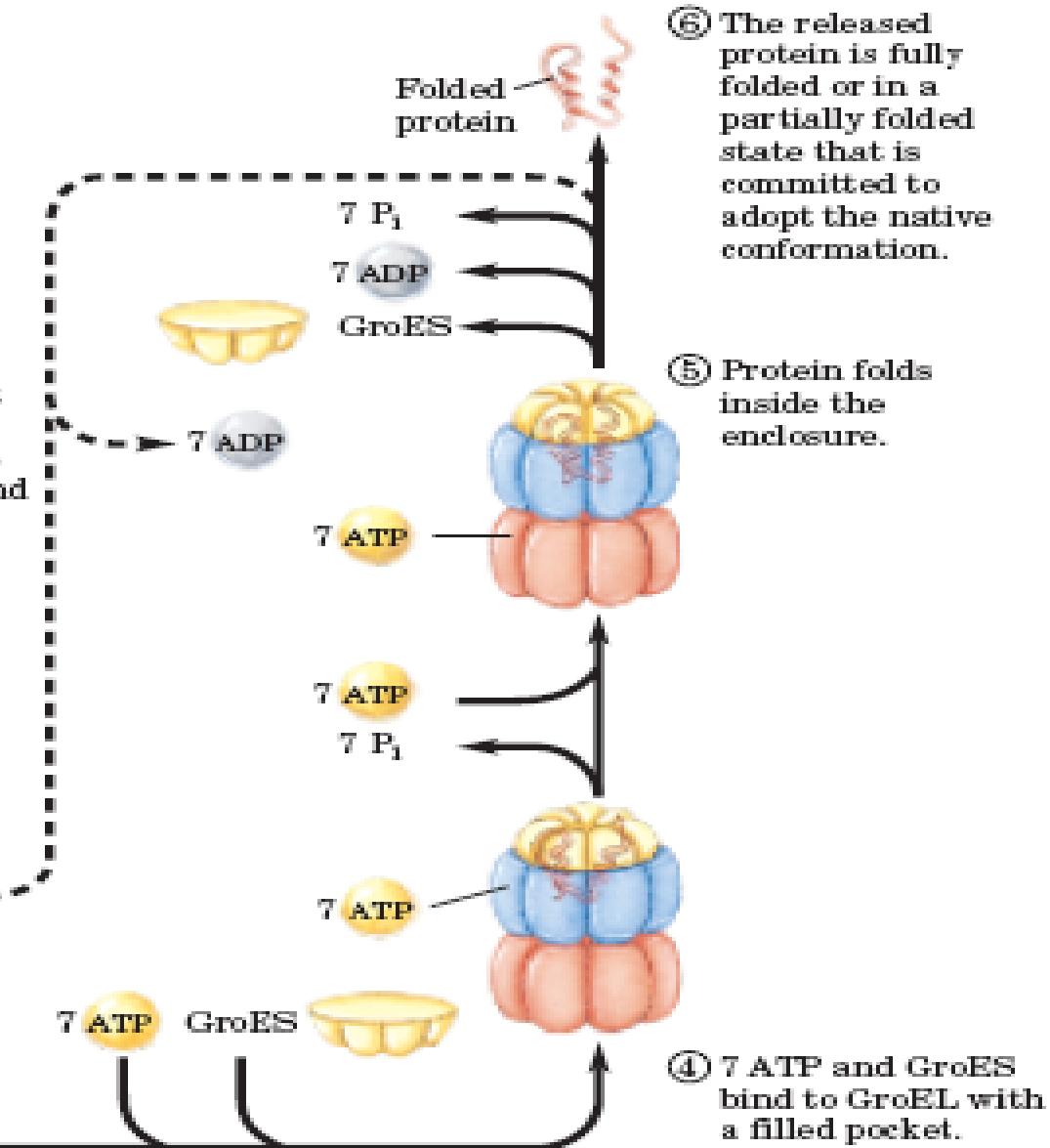
② ATP binds to each subunit of the GroEL heptamer.



③ ATP hydrolysis leads to release of 14 ADP and GroES.



⑦ Proteins not folded when released are rapidly bound again.



(a)

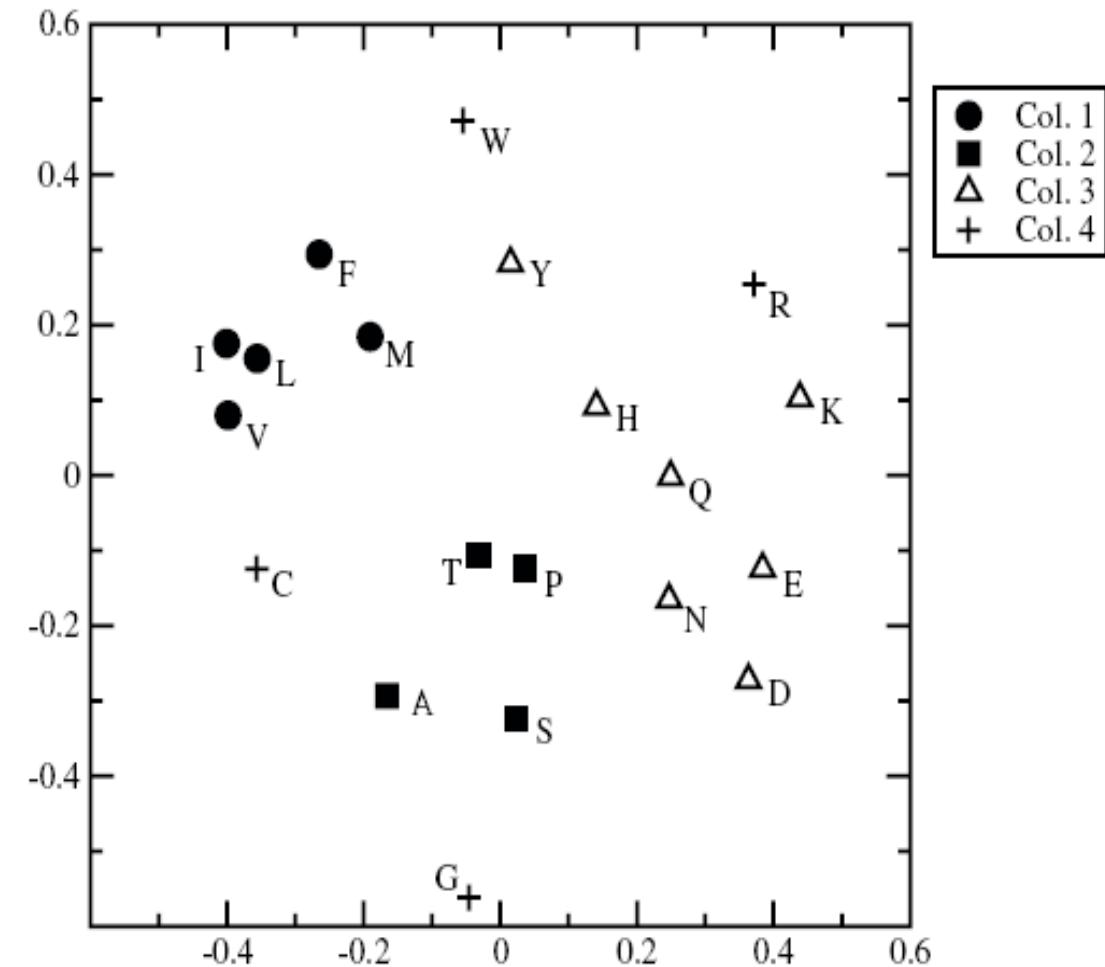
The Genetic Code

The genetic code is a map of Codons “C” to Amino Acids “A”
g: C → A

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

Grouping by Physical Properties of Amino Acids Best Explains the Genetic Code Table

		2nd base					
		U	C	A	G		
1st base	U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Stop UGG Trp	U C A G	3rd base
	C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G	
	A	AUU Ae AUC Ae AUA Ae AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G	
	G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gy GGC Gy GGA Gy GGG Gy	U C A G	



Important points related to translation

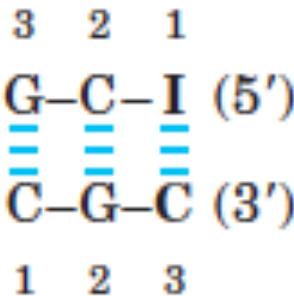
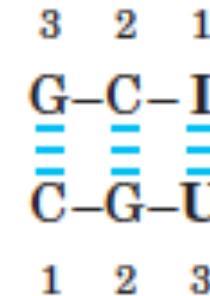
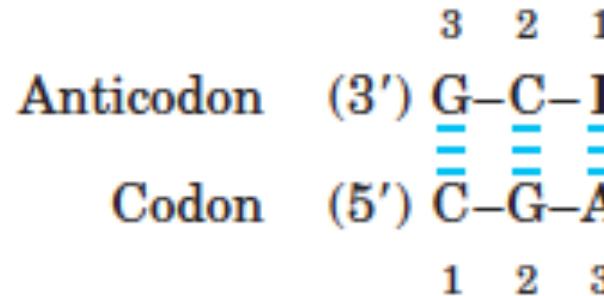
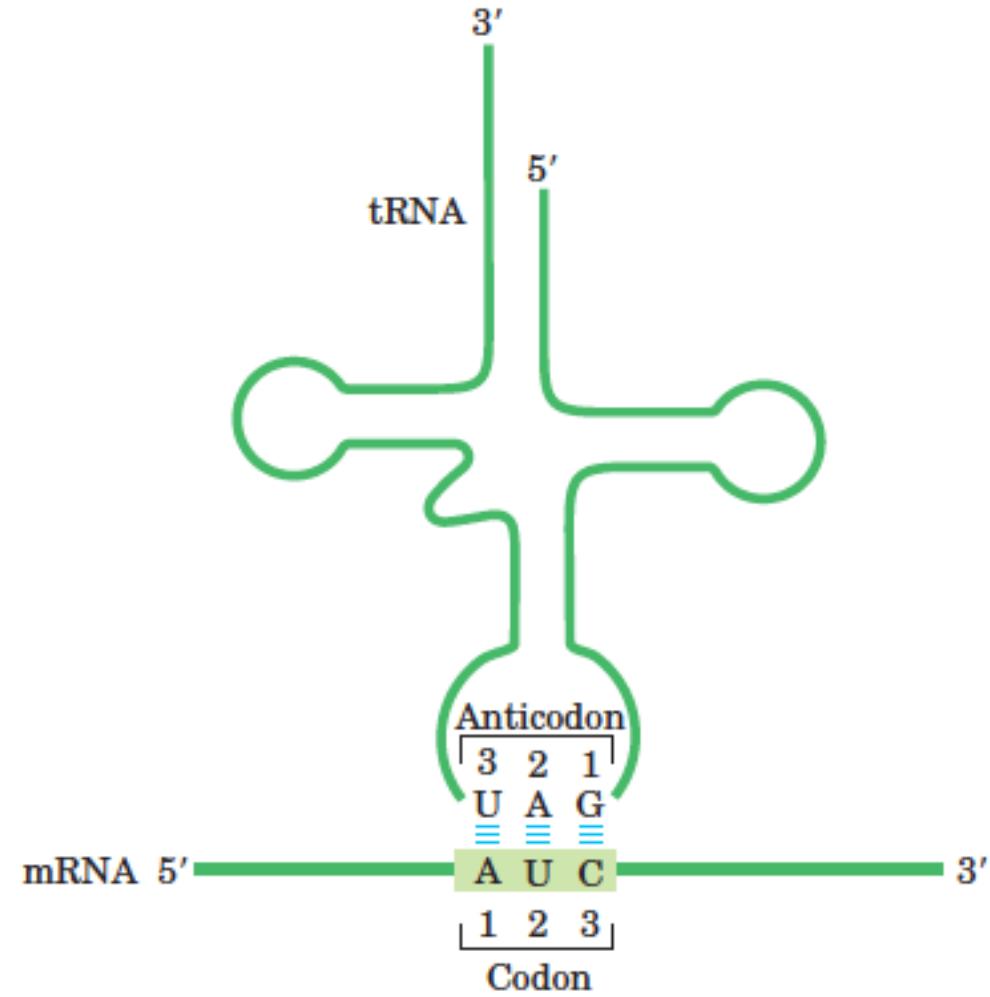
- ❖ The particular amino acid sequence of a protein is constructed through the translation of information encoded in mRNA. This process is carried out by ribosomes.
- ❖ Amino acids are specified by mRNA codons consisting of nucleotide triplets. Translation requires adaptor molecules, the tRNAs, that recognize codons and insert amino acids into their appropriate sequential positions in the polypeptide.
- ❖ The base sequences of the codons were deduced from experiments using synthetic mRNAs of known composition and sequence.
- ❖ The codon AUG signals initiation of translation. The triplets UAA, UAG, and UGA are signals for termination.

Degeneracy of the Genetic Code

<i>Amino acid</i>	<i>Number of codons</i>	<i>Amino acid</i>	<i>Number of codons</i>
Met	1	Tyr	2
Trp	1	Ile	3
Asn	2	Ala	4
Asp	2	Gly	4
Cys	2	Pro	4
Gln	2	Thr	4
Glu	2	Val	4
His	2	Arg	6
Lys	2	Leu	6
Phe	2	Ser	6

The Wobble Hypothesis

- ❖ Alignment of the two RNAs is antiparallel. The tRNA is shown in the traditional cloverleaf configuration
- ❖ Three different codon pairing relationships are possible when the tRNA anticodon contains inosinate.



1. One codon recognized:

1. Anticodon

(3') X-Y-C (5')

(3') X-Y-A (5')

Codon

(5') Y-X-G (3')

(5') Y-X-U (3')

2. Two codons recognized:

1. Anticodon

(3') X-Y-U (5')

(3') X-Y-G (5')

Codon

(5') Y-X-A (3')

(5') Y-X-C (3')

3. Three codons recognized:

1. Anticodon

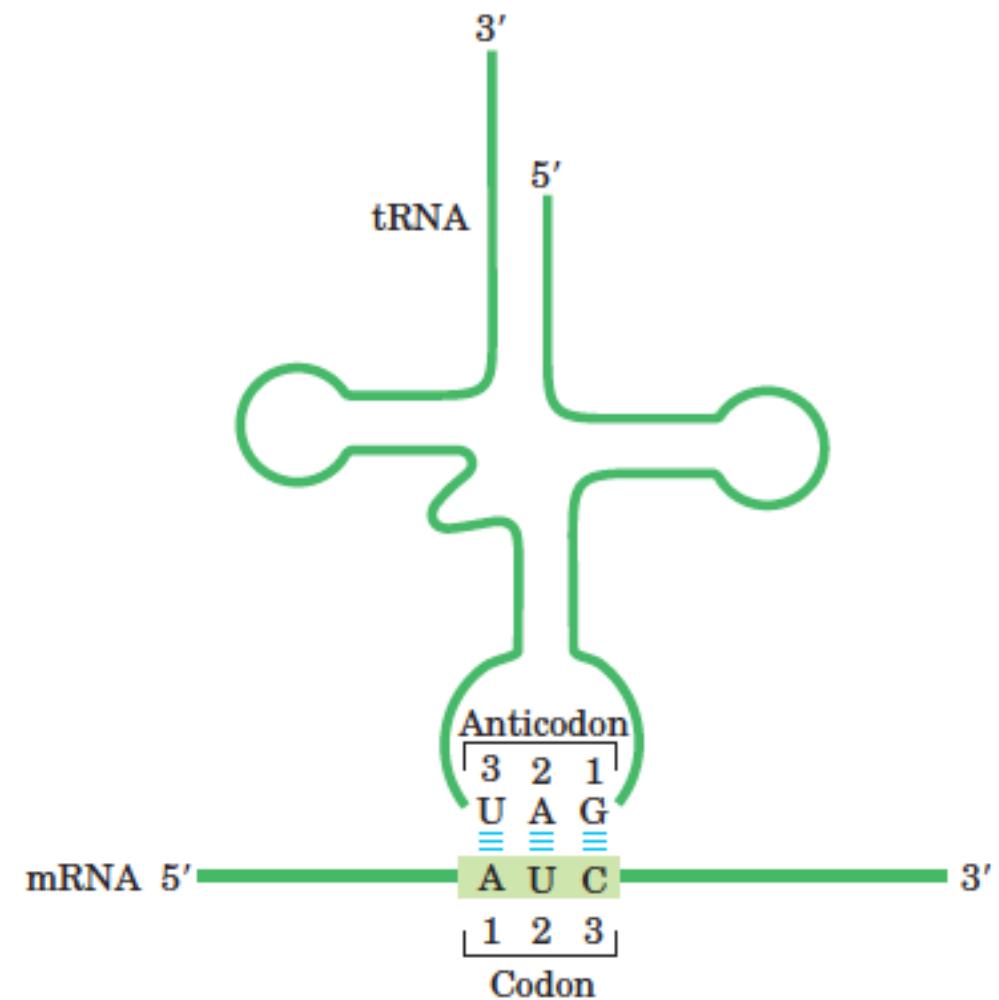
(3') X-Y-I (5')

Codon

(5') Y-X-A (3')

(5') Y-X-U (3')

(5') Y-X-C (3')



G	GUU GUC GUA GUG	Val	GCU GCC GCA GCG	Ala	GAU GAC GAA GAG	Asp	GGU GGC GGA GGG	Gly	U C A G
---	--------------------------	-----	--------------------------	-----	--------------------------	-----	--------------------------	-----	------------------

C	CUU CUC CUA CUG	Leu	CCU CCC CCA CCG	Pro	CAU CAC CAA CAG	His	CGU CGC CGA CGG	Arg	U C A G
---	--------------------------	-----	--------------------------	-----	--------------------------	-----	--------------------------	-----	------------------

Reading Frames

Reading frame 1 5'---[U U C] [U C G] [G A C] [C U G] [G A G] [A U U] [C A C] [A G U] --- 3'

Reading frame 2 --- **U** **U C U** **C G G** **A C C** **U G G** **A G A** **U U C** **A C A** **G U** ---

Reading frame 3 ---U U C U C G G A C C U G G A G A U U C A C A G U---

Nonoverlapping code A U A C G A G U C _____
 1 2 3

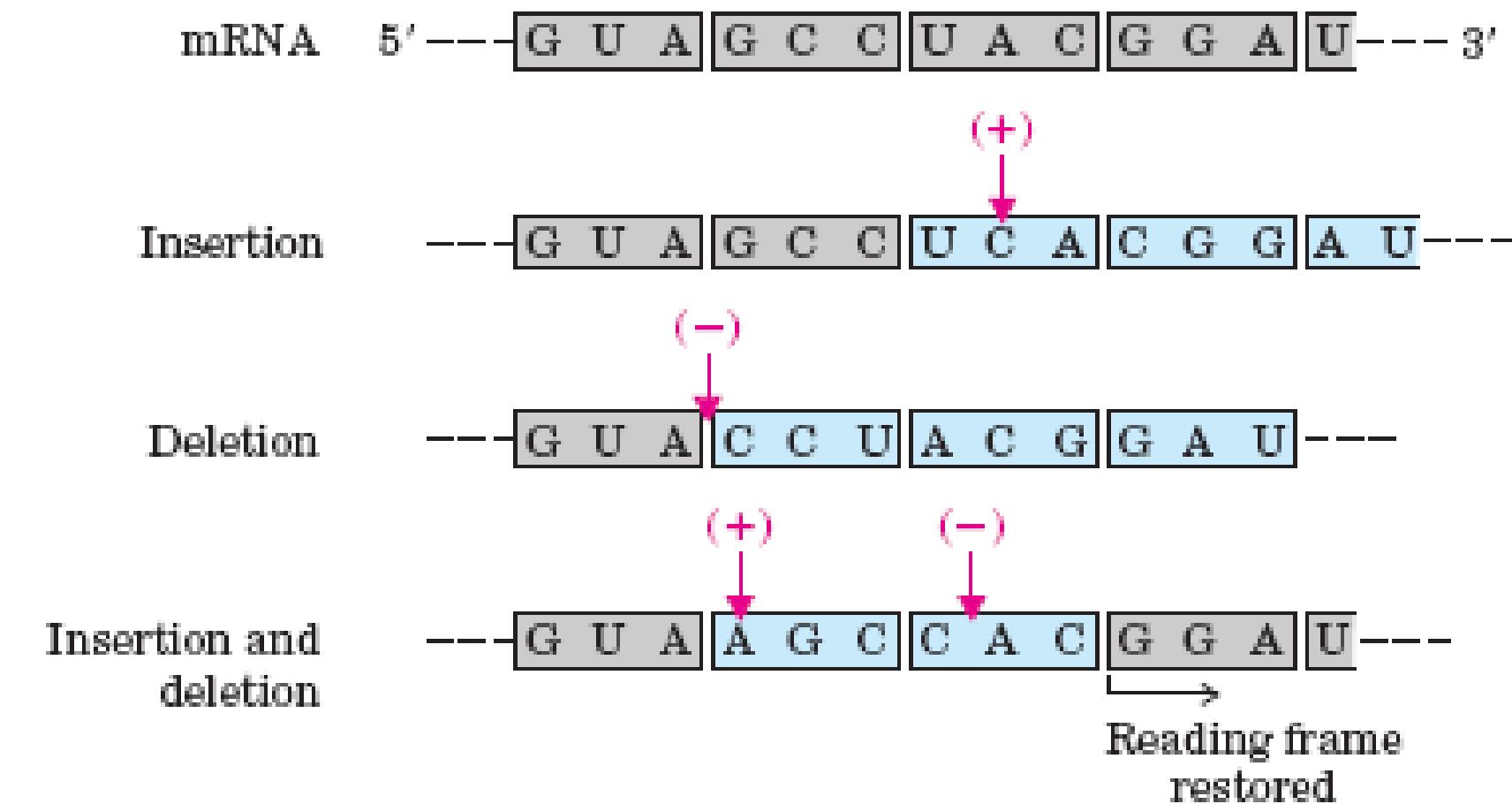
Overlapping code A U A C G A G U C
 1

3

Reading frame 1 5' --- **G U A** **A G U** **A A G** **U A A** **G U A** **A G U** **A A --- 3'**

Reading frame 2 --- G **U A A** **G U A** **A G U** **A A G** **U A A** **G U A** ---

The Triplet Non-overlapping Code



Five Stage of Protein Synthesis in *E. coli*

1. Activation of Amino Acids
2. Initiation
3. Elongation
4. Termination
5. Folding and post translational processing

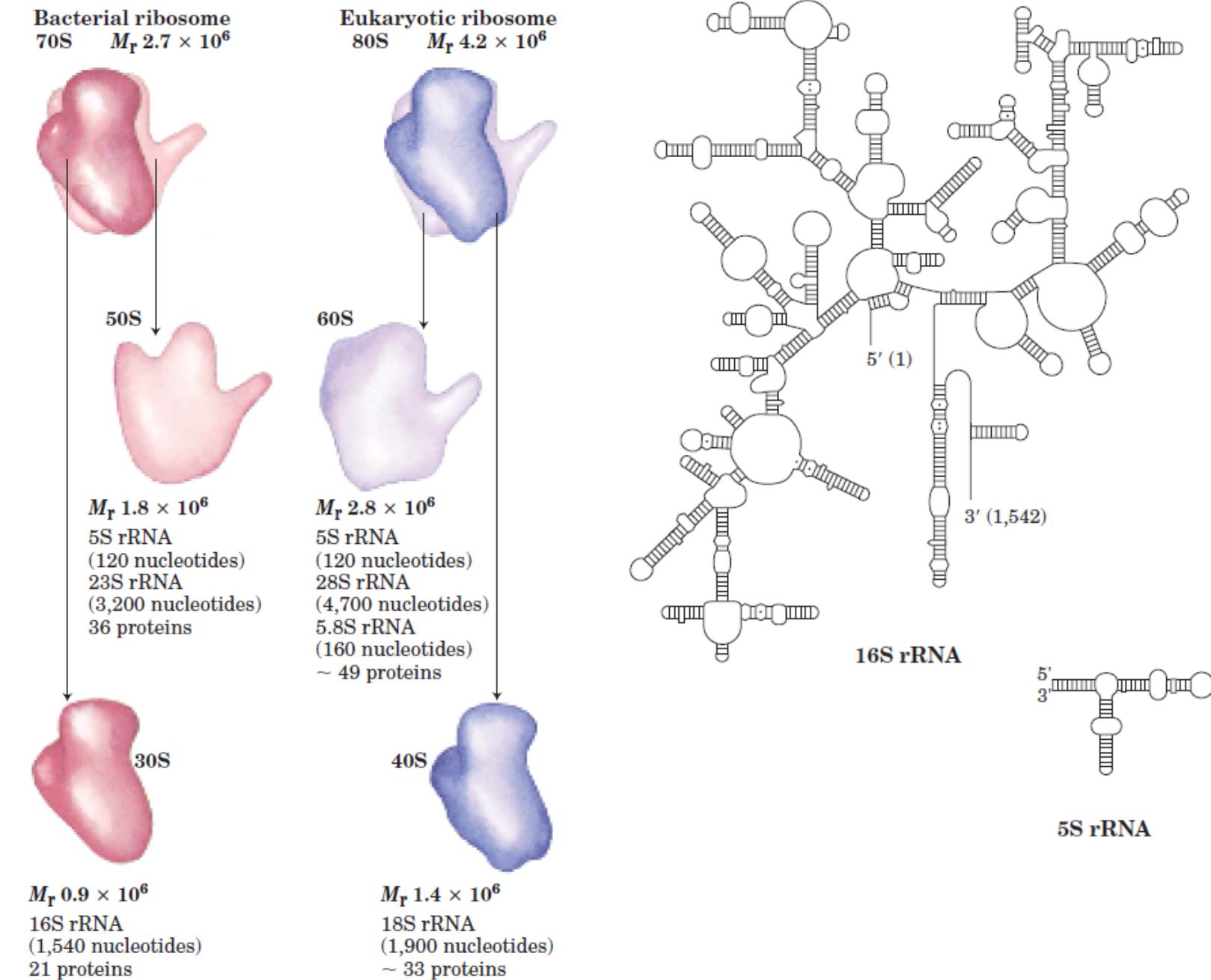
Ribosome is a Complex Supramolecular Machine

1. Each *E. coli* contains 15000 or more ribosomes (1/4 cell weight)
2. 18 nm is size
3. Two subunits (i) 30S (ii) 50S combined 70S (S is the sedimentation coefficient)
4. Subunits are made of many proteins and at least one large rRNA
5. Bacterial ribosomes have 55 proteins with molecular weights varying from 6000 to 75000

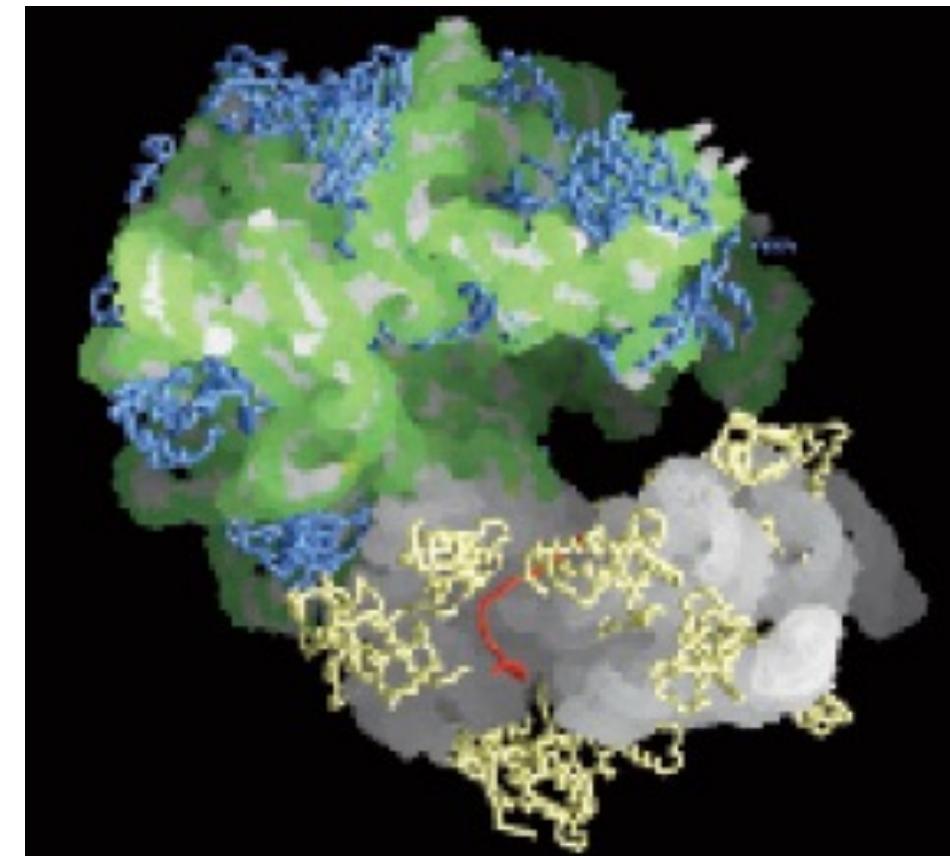
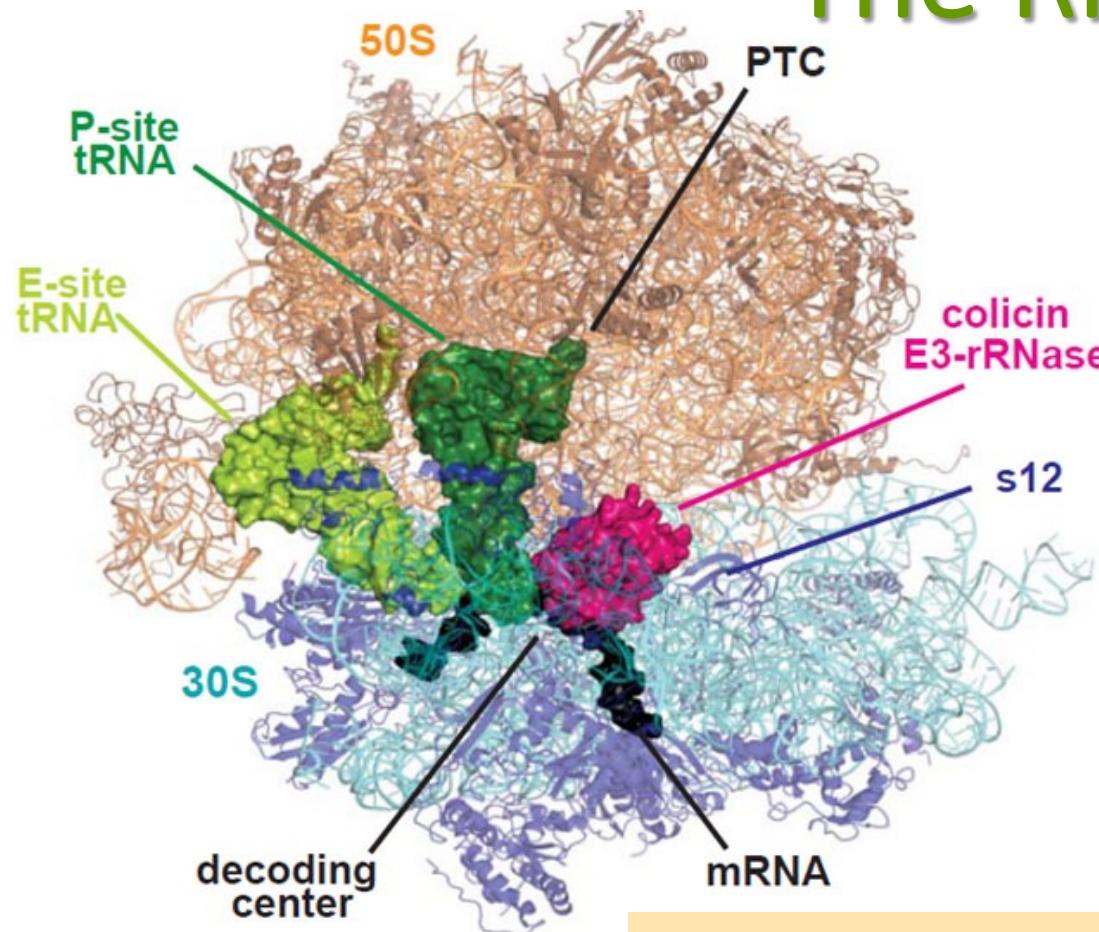
PDB ID **1JJ2** and **1GIY**

Bacterial rRNA

Secondary structure of *E. coli* 16S and 5S rRNAs. The first (5' end) and final (3' end) ribonucleotide residues of the 16S rRNA are numbered.



The Ribosome



RNA and Protein Components of the *E. coli* Ribosome

Subunit	Number of different proteins	Total number of proteins	Protein designations	Number and type of rRNAs
30S	21	21	S1-S21	1 (16S rRNA)
50S	33	36	L1-L36*	2 (5S and 23S rRNAs)

Nobel Prize in Chemistry 2009



The Nobel Prize in Chemistry 2009 was awarded jointly to Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath "for studies of the structure and function of the ribosome"

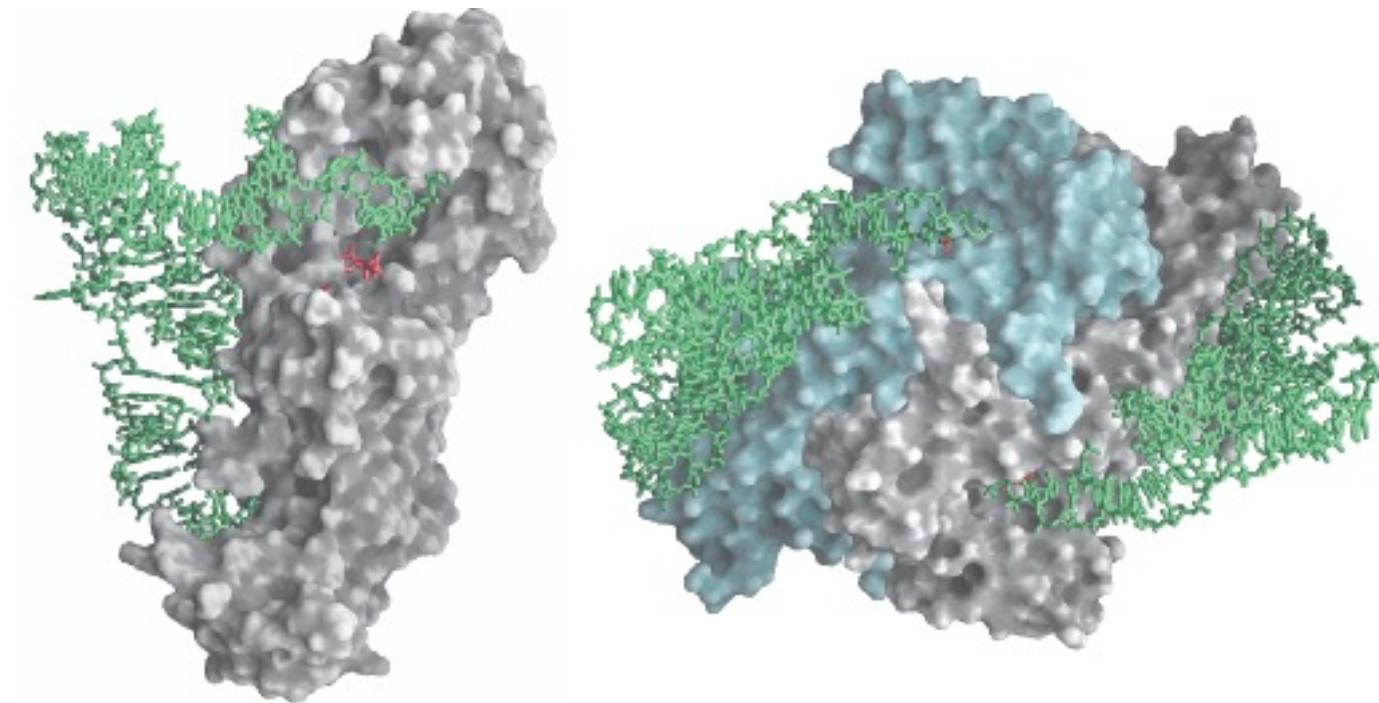
Visit to KSBS Aug 2020

<https://www.nobelprize.org/prizes/chemistry/2009/summary/>

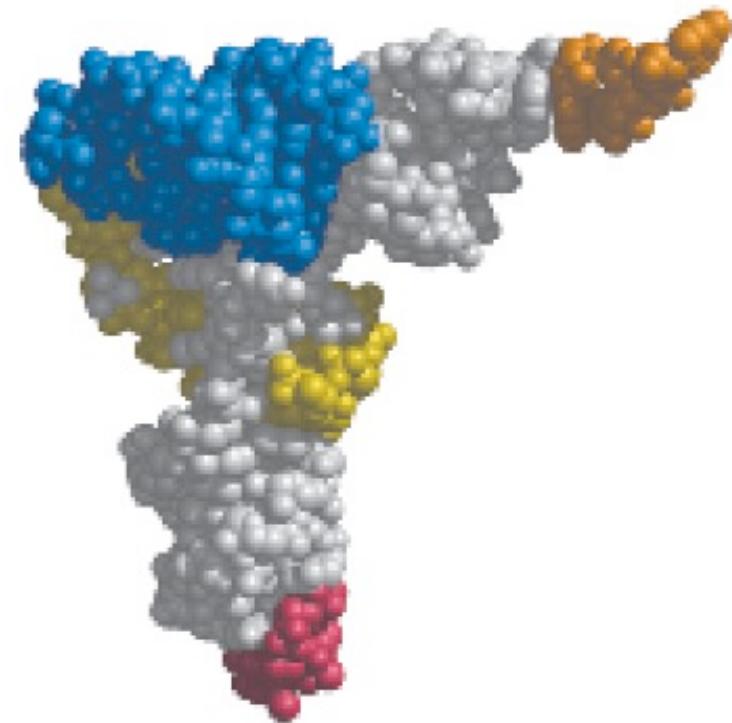
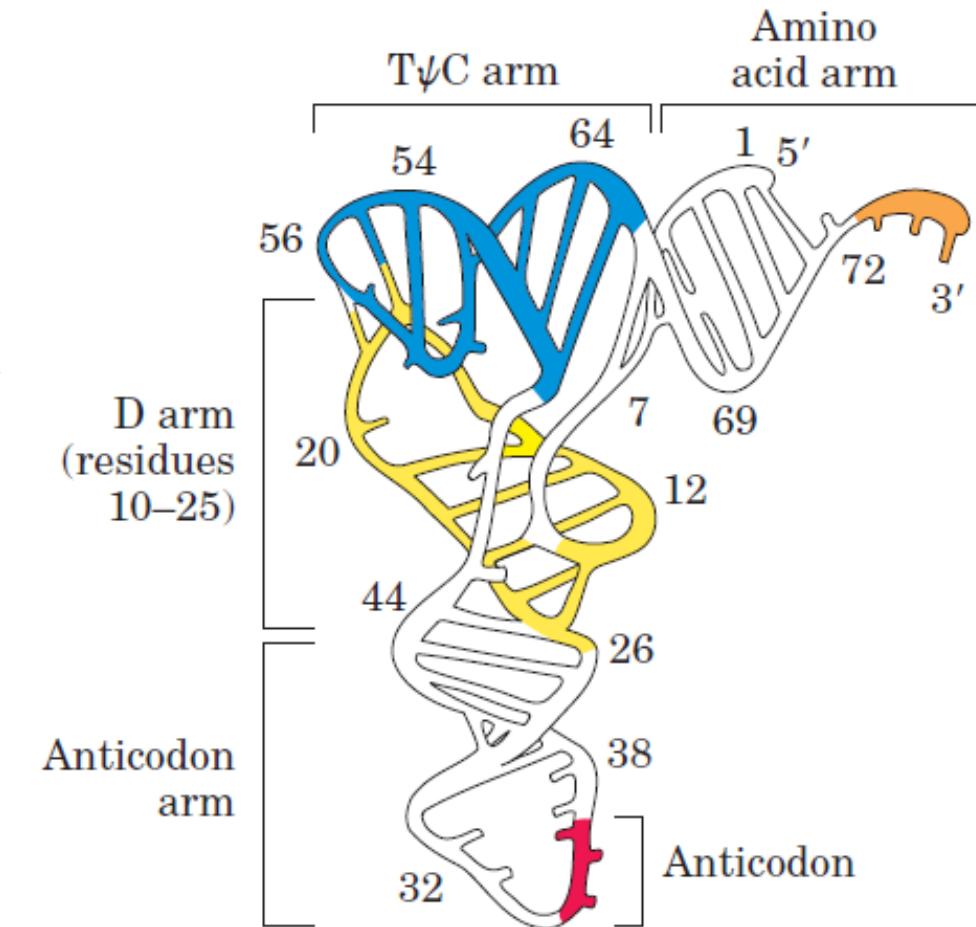
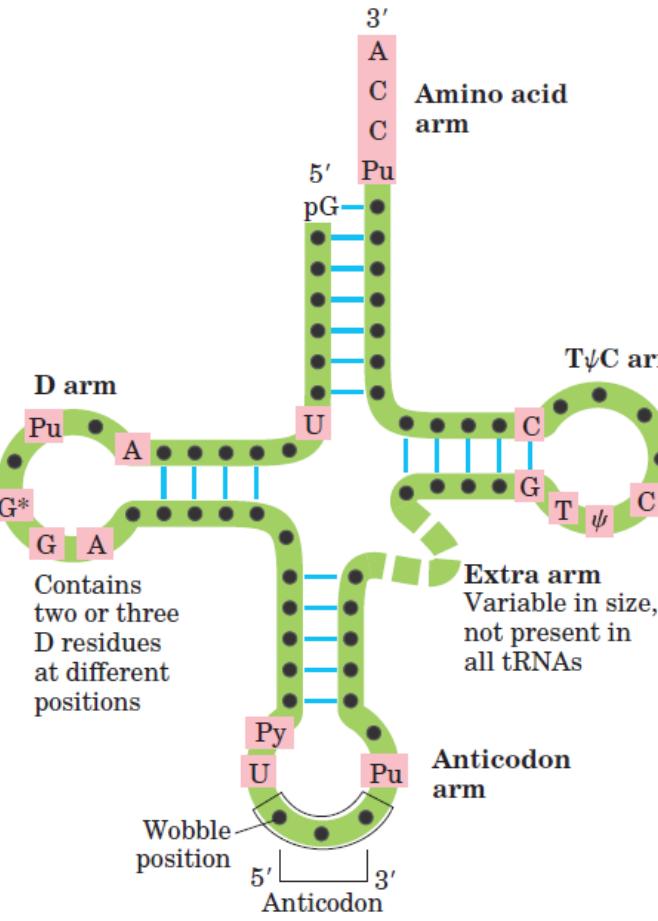
Aminoacyl t-RNA Synthetase

**Aminoacyl-tRNA Synthetases
Attach the Correct Amino Acids to
Their tRNAs**

- ❖ aminoacyl-tRNA synthetases esterify the 20 amino acids to their corresponding tRNAs. Each enzyme is specific for one amino acid and one or more corresponding tRNAs
- ❖ Proofreading by Aminoacyl-tRNA Synthetases
- ❖ Interaction between an Aminoacyl-tRNA Synthetase and a tRNA



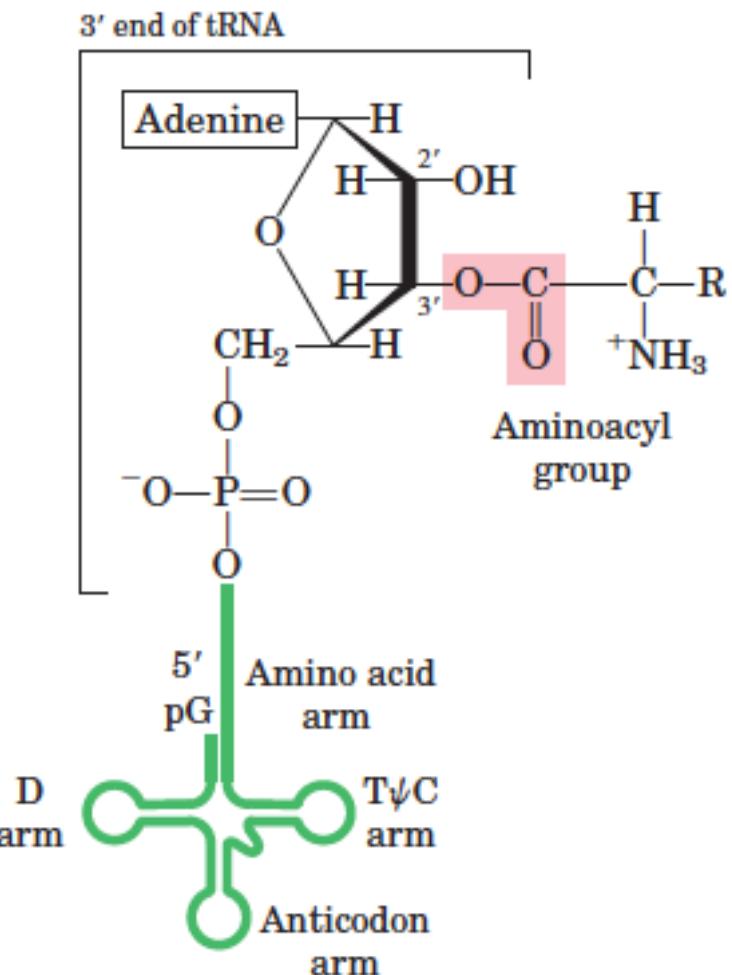
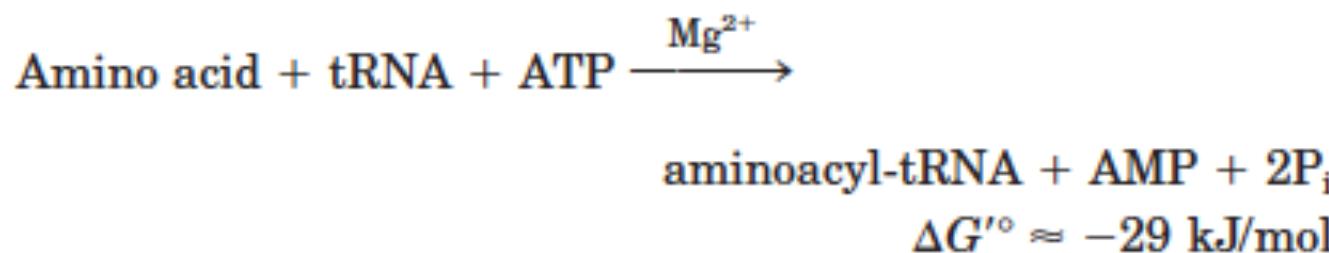
N
D G K N Q L P Y N Q L P Y D G K D G



Stage 1: Attaching the correct amino acid to the correct tRNA

in the cytosol, aminoacyl-tRNA synthetases esterify the 20 amino acids to their corresponding tRNAs.

- ❖ Each enzyme is specific for one amino acid and one or more corresponding tRNAs
- ❖ Most organisms have one aminoacyl-tRNA synthetase for each amino acid
- ❖ For amino acids with two or more corresponding tRNAs, the same enzyme usually aminoacylates all of them



Proof Reading by Aminoacyl tRNA Synthetases

Aminoacylation of tRNA accomplishes

1. Activation of amino acid for peptide bond formation
2. attachment of amino acid to an adaptor tRNA for placement of amino acid

The amino acid attached is not checked on the ribosome!

How is the fidelity assured?

Consider Valine and Isoleucine - different by only – CH₂ –

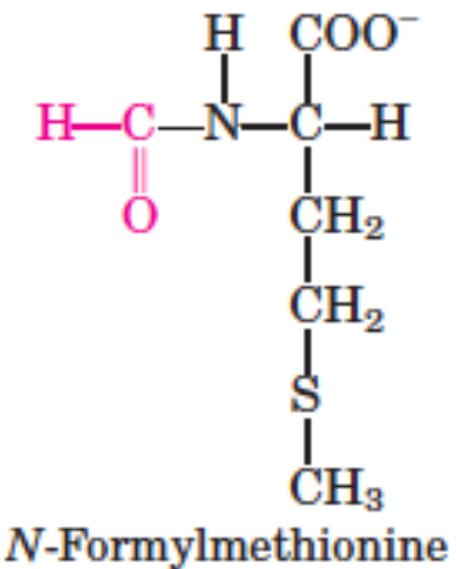
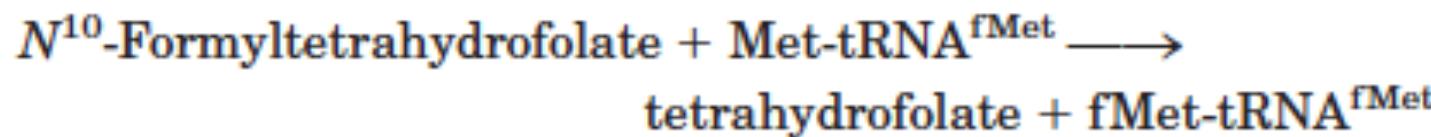
In the case of Ile-tRNA synthetase

1. Activation of Ile is favored by a factor of 200
2. Binding is carried out in 2 steps (acts as filter)
3. Incorrect binding occurs at a second site that has a higher hydrolytic rate
4. In this case, overall process is 1:3000 in favor of the correct amino acid Ile

Stage 2: A Specific Amino Acid Initiates Protein Synthesis

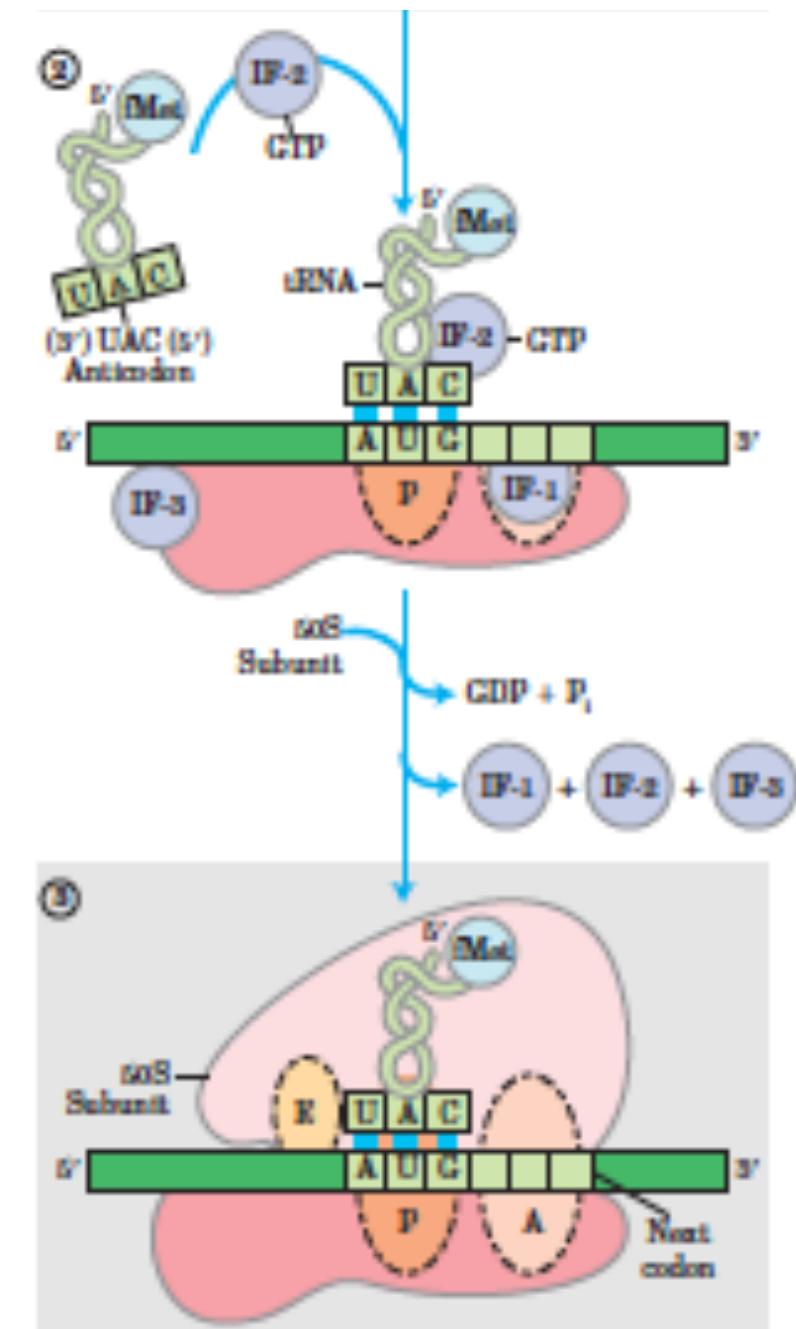
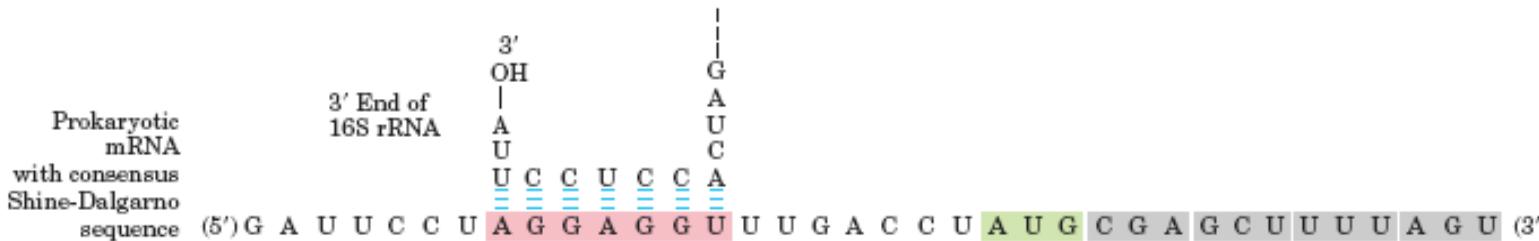
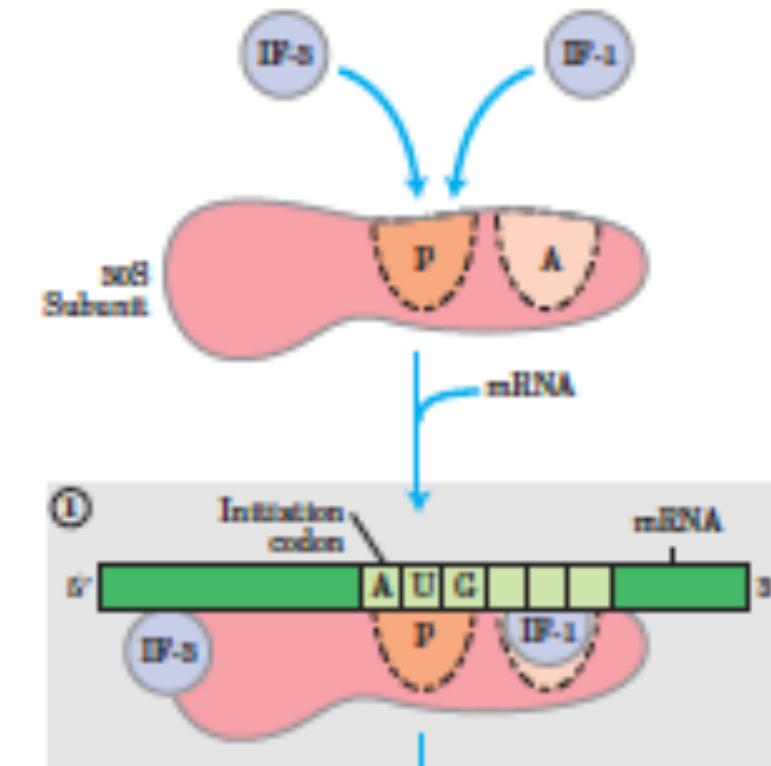
Although methionine has only one codon, (5')AUG, all organisms have two tRNAs for methionine

- ❖ One is used exclusively when (5')AUG is the initiation codon for protein synthesis
- ❖ The other is used to code for a Met residue in an internal position in a polypeptide
- ❖ The amino acid incorporated in response to the (5')AUG initiation codon is N-formylmethionine (fMet)



Formation of the Initiation Complex

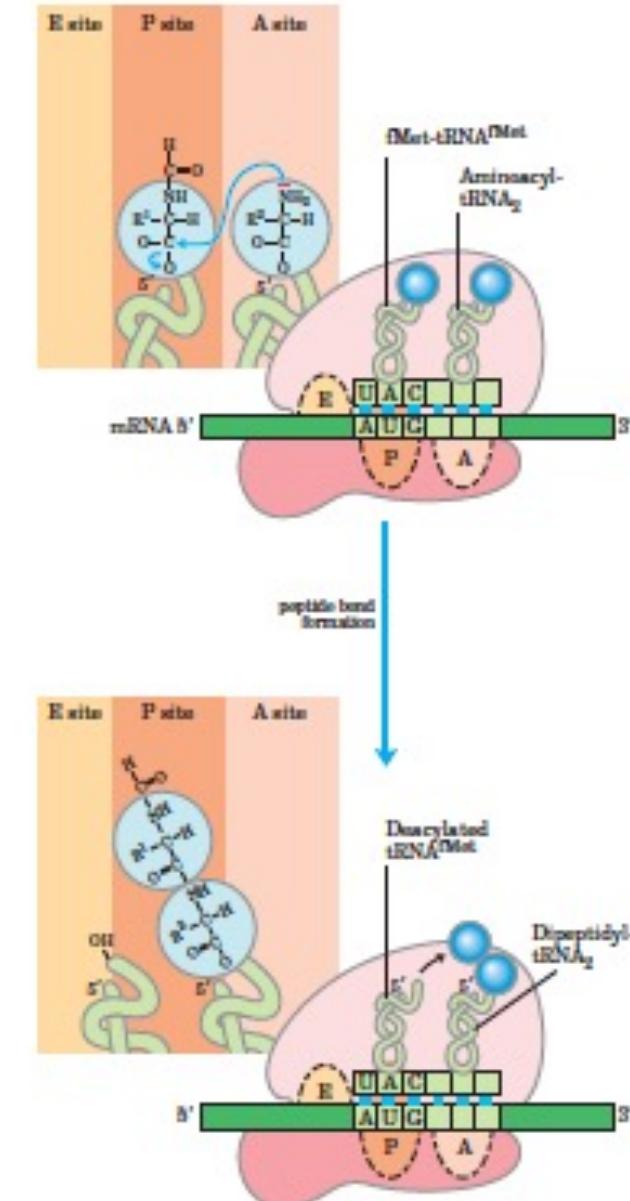
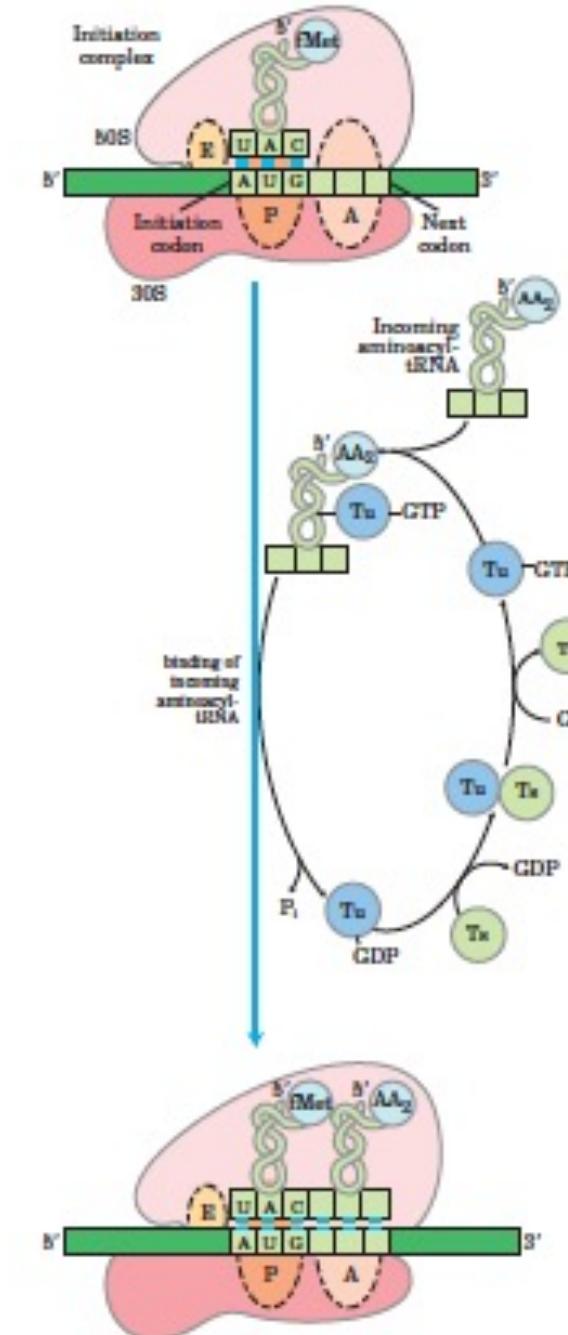
1. 30S ribosomal subunit
2. mRNA coding for the polypeptide to be made
3. Initiating fMet-tRNA_fMet
4. A set of three proteins called initiation factors (IF-1, IF-2, and IF-3)
5. GTP
6. 50S ribosomal subunit
7. Mg²⁺.



Stage 3: Elongation of the Peptide Chain

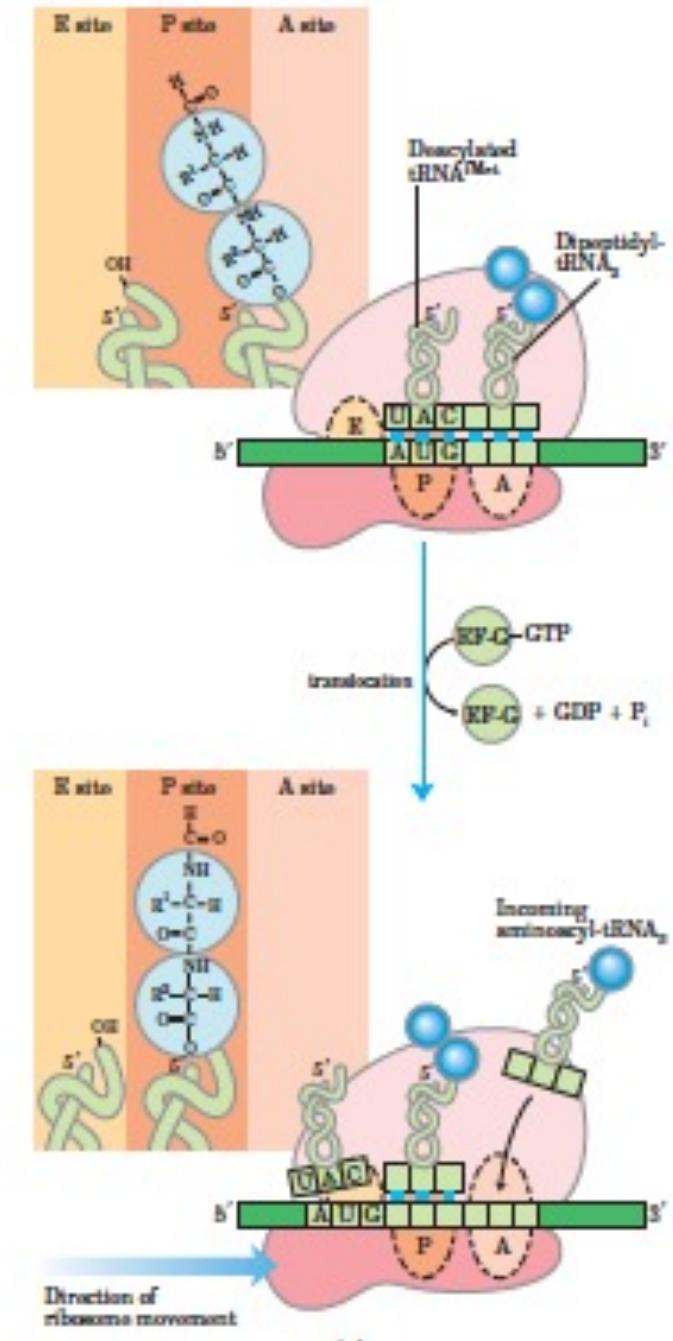
Elongation requires

1. the initiation complex
2. aminoacyl-tRNAs
3. a set of three soluble cytosolic proteins called elongation factors (EF-Tu, EF-Ts, and in bacteria)
4. GTP



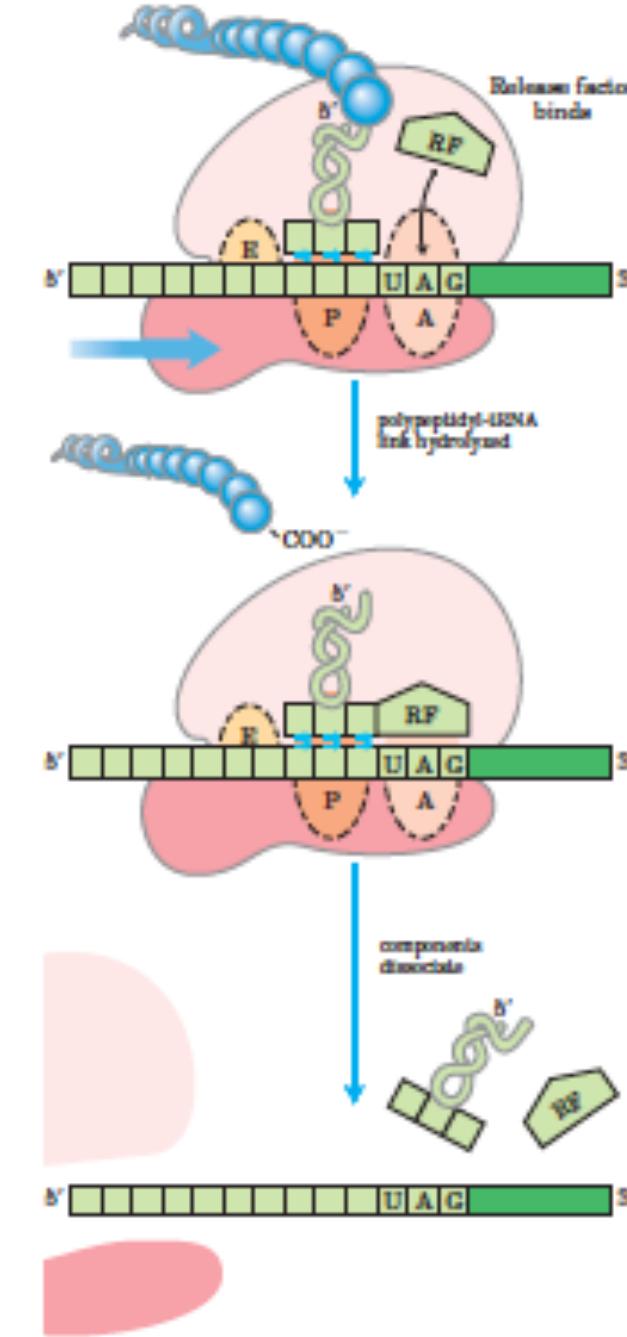
Stage 3 : Translocation

1. The ribosome moves one codon toward the 3' end of the mRNA
This shifts the anticodon of the tRNA attached to the 2nd codon from the A→P site and the deacylated tRNA from P→ E site
2. This movement required EF-G (translocase) and the energy is provided by the hydrolysis GTP→GDP+Pi
3. The uncharged tRNA is dislocated from the E-site and a peptide bond is formed between the growing chain and the new amino acid carried by the tRNA at the A-site



Stage 4: Termination of Synthesis

1. Elongation continues until the last amino acid in the sequence
2. Termination is signaled by the presence of one of the 3 stop codons UAA, UAG or UGA immediately following the final coded amino acid
3. Once the terminal codon occupies A-site, three termination or release factors RF1, RF2 and RF3 contribute to
 - i. Hydrolysis of terminal peptidyl tRNA bond
 - ii. Release of polypeptide from P-site
 - iii. Dissociation of the 70S ribosome into the 30S and 50S subunits



Summary of the 5 stages of protein synthesis

Stage	Essential components
1. Activation of amino acids	20 amino acids 20 aminoacyl-tRNA synthetases 32 or more tRNAs ATP Mg^{2+}
2. Initiation	mRNA <i>N</i> -Formylmethionyl-tRNA ^{fmet} Initiation codon in mRNA (AUG) 30S ribosomal subunit 50S ribosomal subunit Initiation factors (IF-1, IF-2, IF-3) GTP Mg^{2+}
3. Elongation	Functional 70S ribosome (initiation complex) Aminoacyl-tRNAs specified by codons Elongation factors (EF-Tu, EF-Ts, EF-G) GTP Mg^{2+}
4. Termination and release	Termination codon in mRNA Release factors (RF-1, RF-2, RF-3)
5. Folding and posttranslational processing	Specific enzymes, cofactors, and other components for removal of initiating residues and signal sequences, additional proteolytic processing, modification of terminal residues, and attachment of phosphate, methyl, carboxyl, carbohydrate, or prosthetic groups

Lecture 12

Regulation of Gene Expression
Negative regulation - Lac Operon
Attenuation - Tryptophan Operon

Acknowledgement: Leninger Chapter 28

Objectives

Emergence of Life

Fundamental units of life

Cellular assemblies

Protein Folding

Protein Synthesis

Gene Regulation

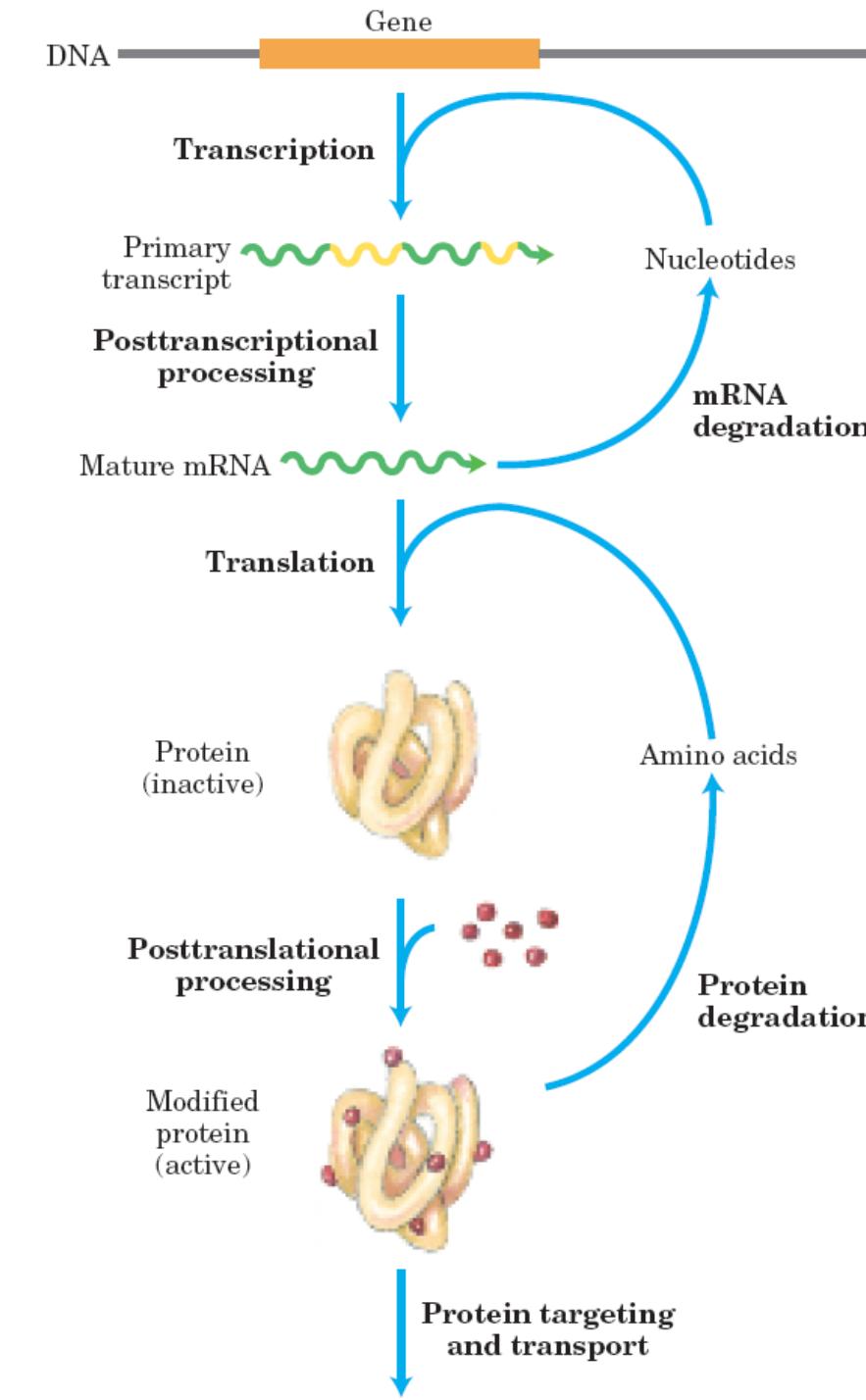
1. Understanding gene regulation
 - a) Operons and regulons
2. Negative and positive regulation
3. Lac operon
4. Attenuation regulation – Tryp operon

Genes are expressed when required

- ❖ Some proteins are expressed abundantly such as elongation factors and rubisco
- ❖ Others such as DNA repair enzymes are synthesized very few in number
- ❖ Requirements of gene products varies in the cell-type and in its life cycle
 - ❖ Ribosomes are synthesized rapidly during the exponential growth phase of the cell

What factors determine the cellular concentration of proteins

1. Synthesis of the primary RNA transcript (transcription)
2. Posttranscriptional modification of mRNA
3. Messenger RNA degradation
4. Protein synthesis (translation)
5. Posttranslational modification of proteins
6. Protein targeting and transport
7. Protein degradation



Gene regulation

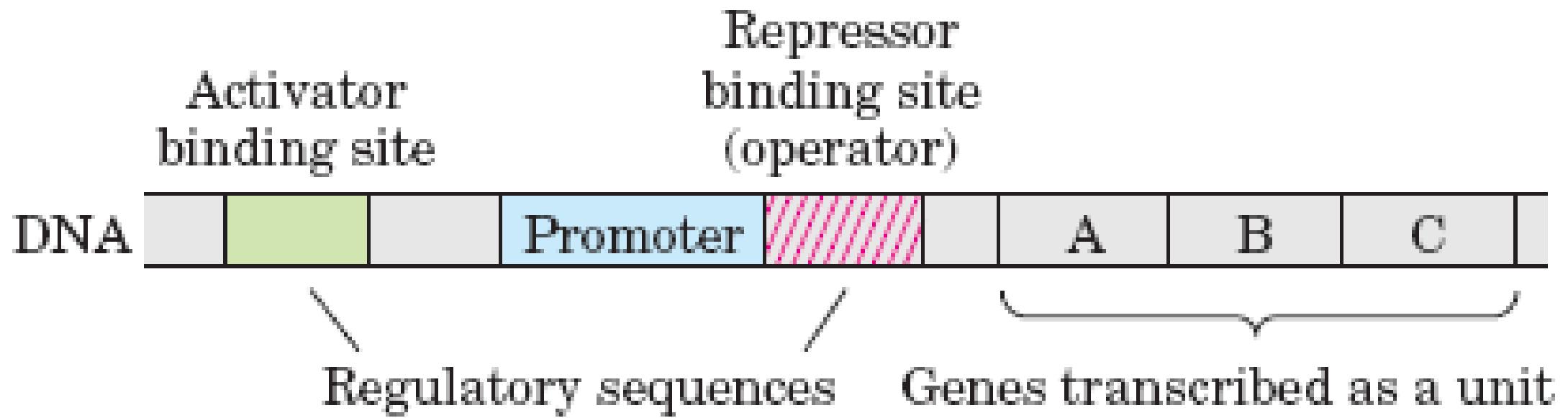
House Keeping Genes

- Constitutive gene expression

Regulated Genes

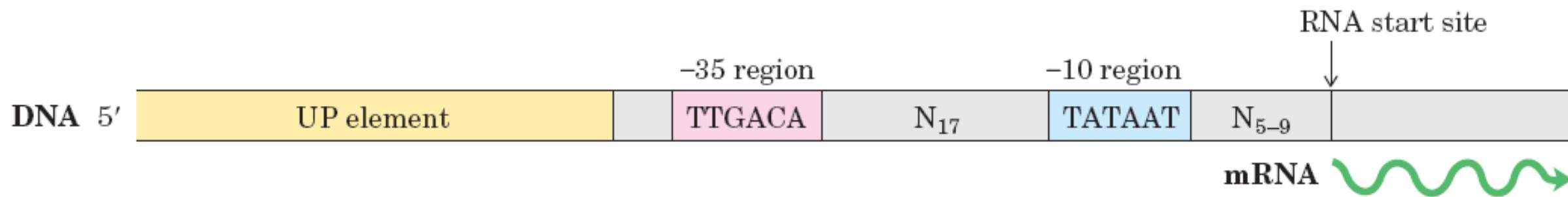
- Inducible gene expression
- Repressible gene expression

Representative Prokaryotic Operon



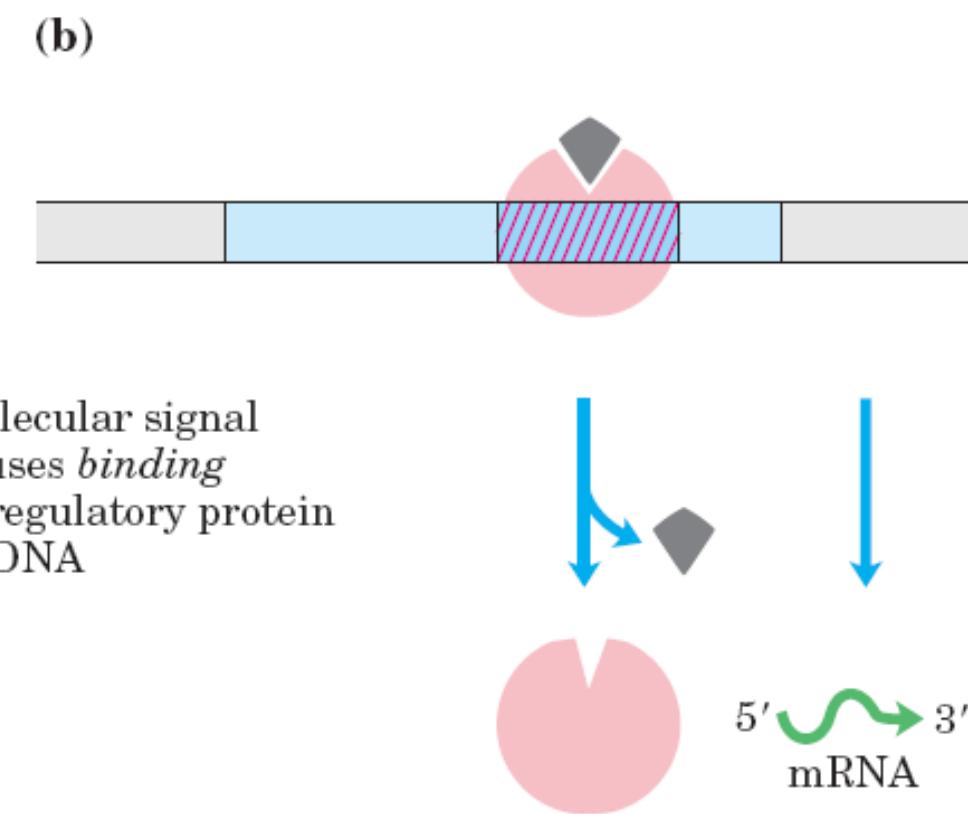
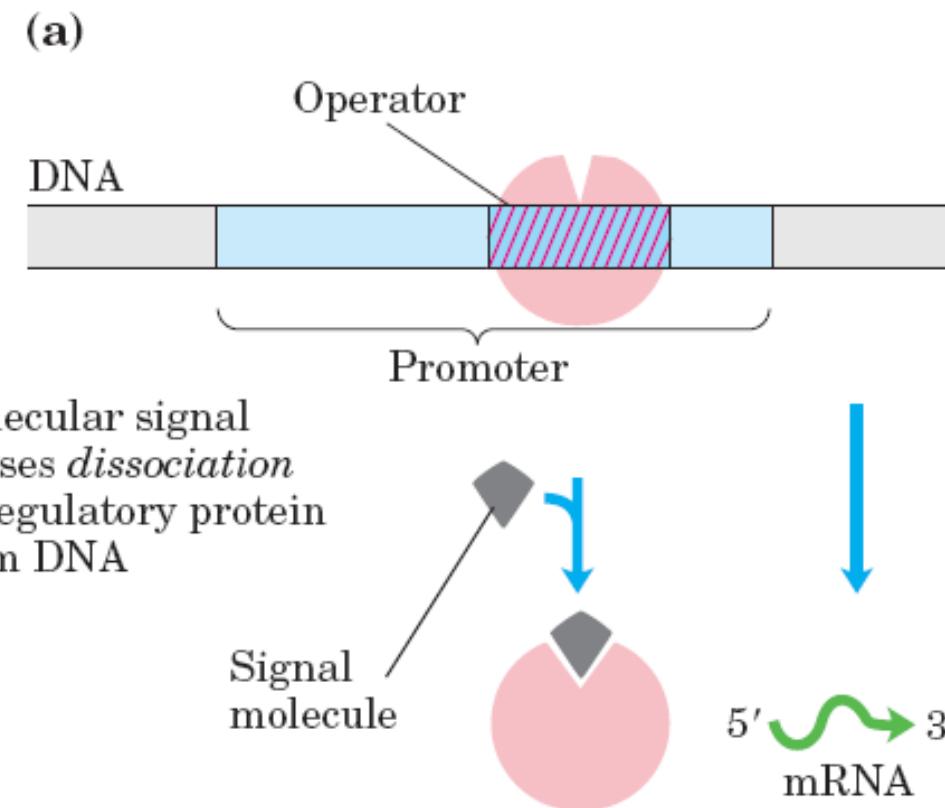
- ❖ Genes A, B, and C are transcribed on one polycistronic mRNA. Typical regulatory sequences include binding sites for proteins that either activate or repress transcription from the promoter

RNA polymerase

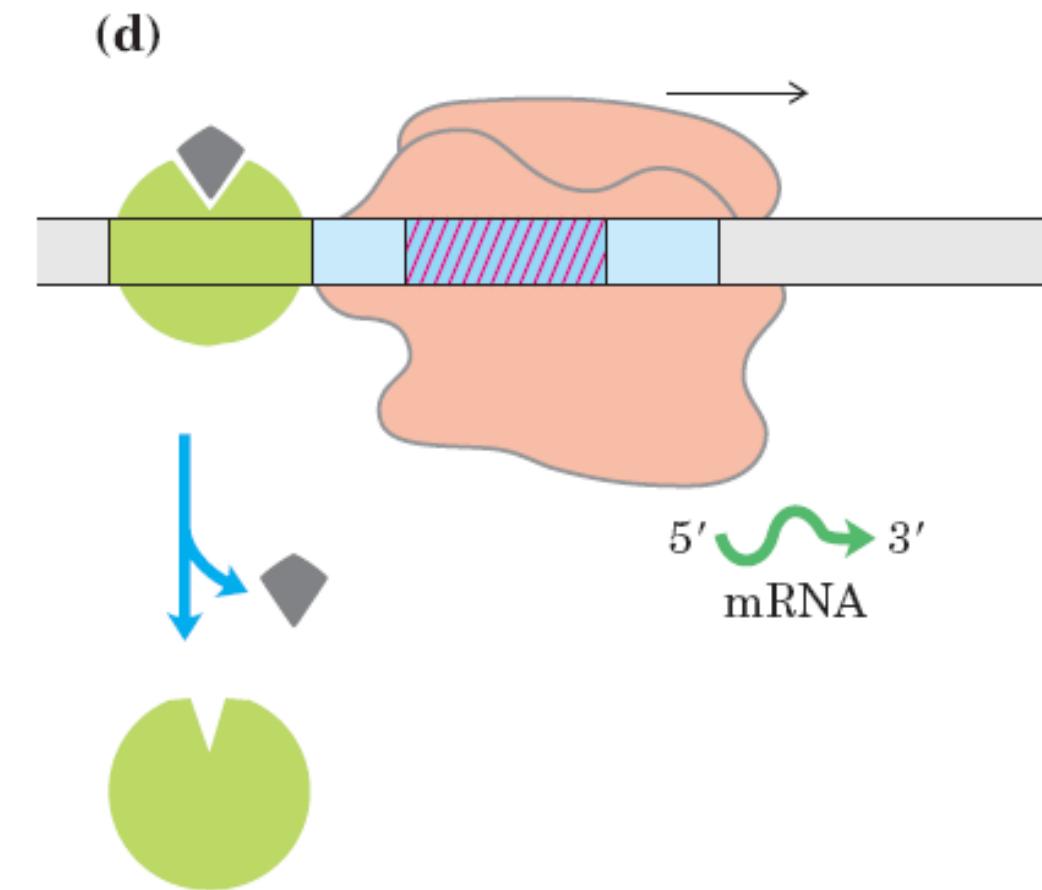
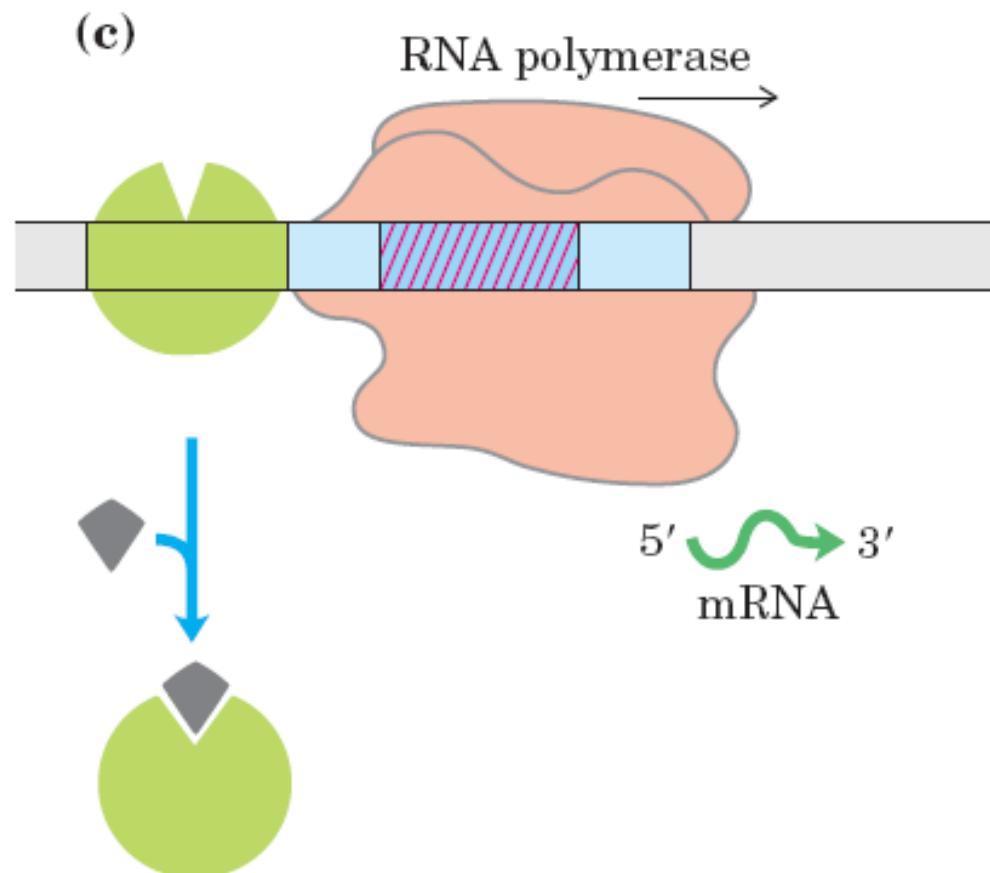


- ❖ RNA polymerases bind to DNA and initiate transcription at promoters, sites generally found near points at which RNA synthesis begins on the DNA template
- ❖ The regulation of transcription initiation often entails changes in how RNA polymerase interacts with a promoter

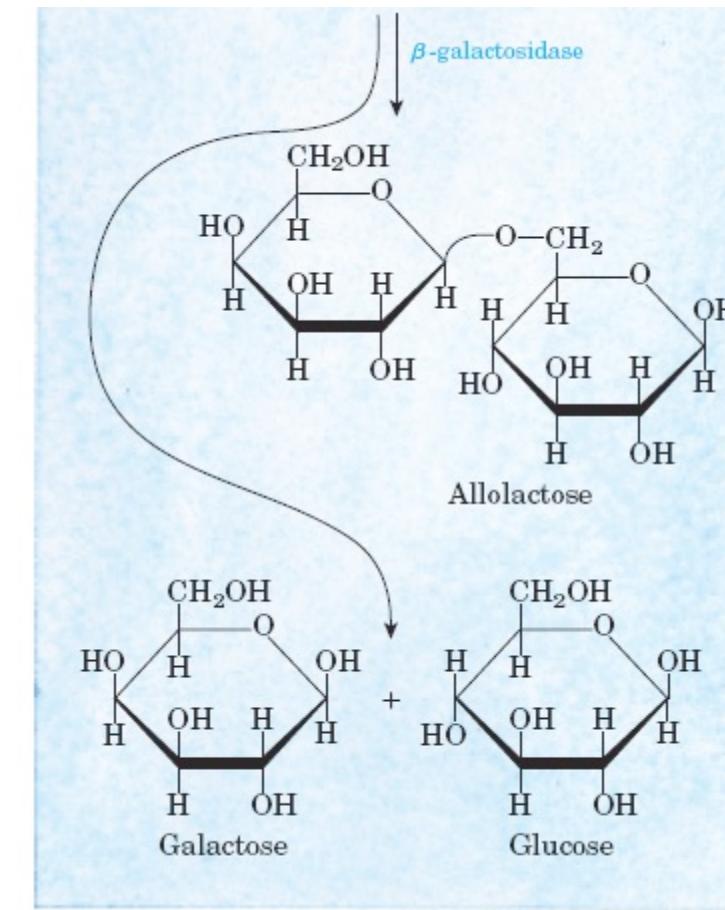
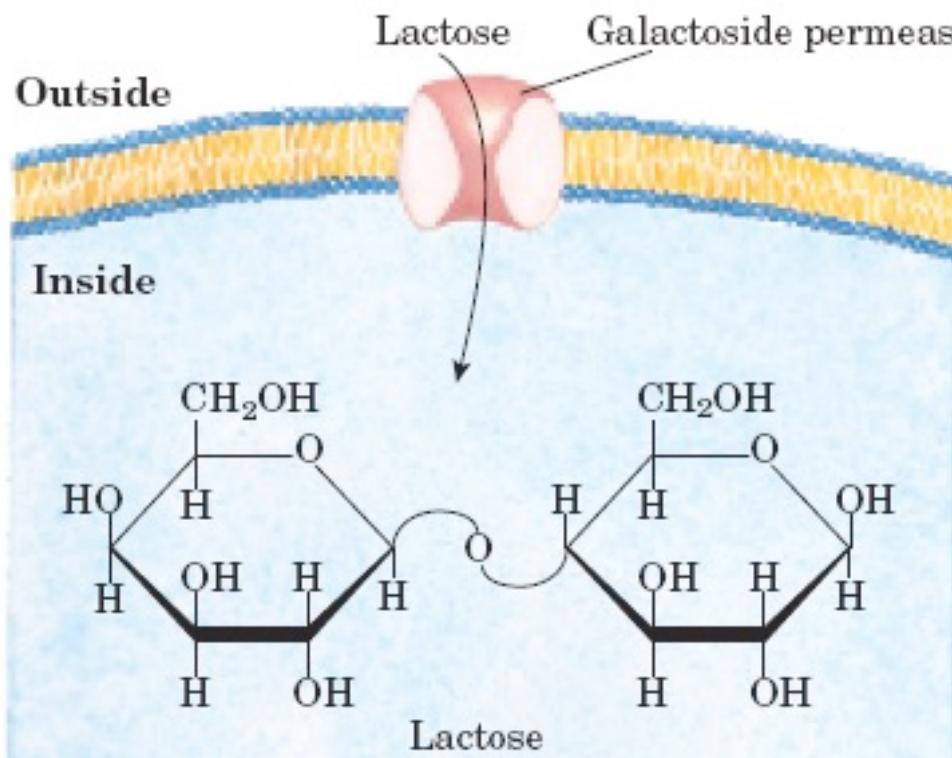
Negative Regulation of Gene Expression



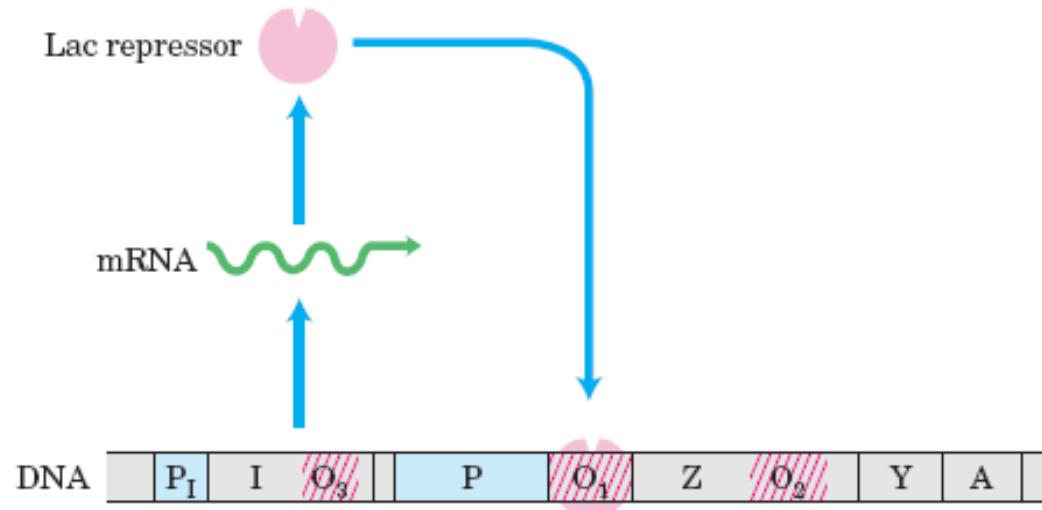
Positive Regulation of Gene Expression



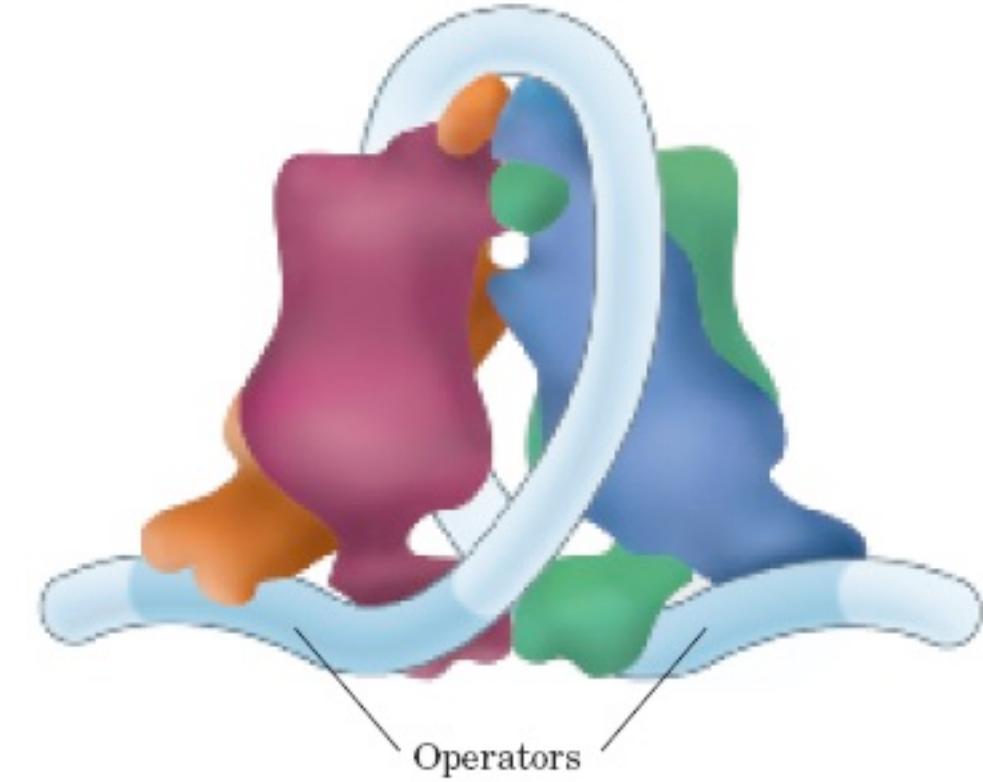
Lactose metabolism in *E. coli*



The Lac Operon

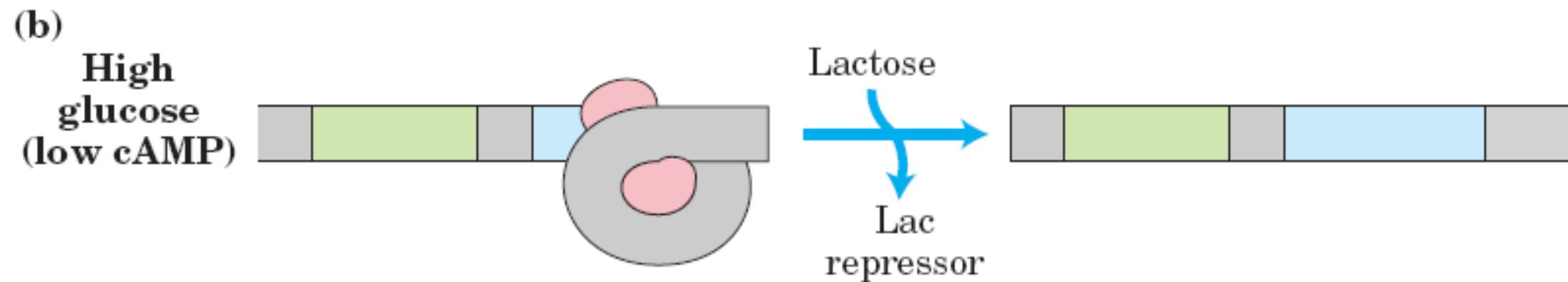
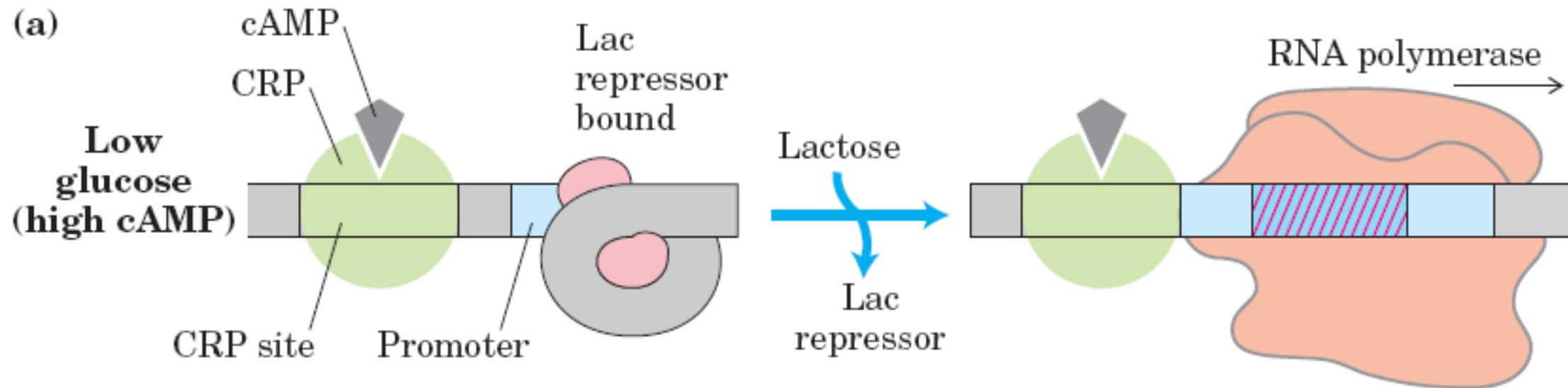


- ❖ The I gene encodes the Lac repressor. The lac Z, Y, and A genes encode beta-galactosidase, galactoside permease, and thiogalactoside transacetylase, respectively

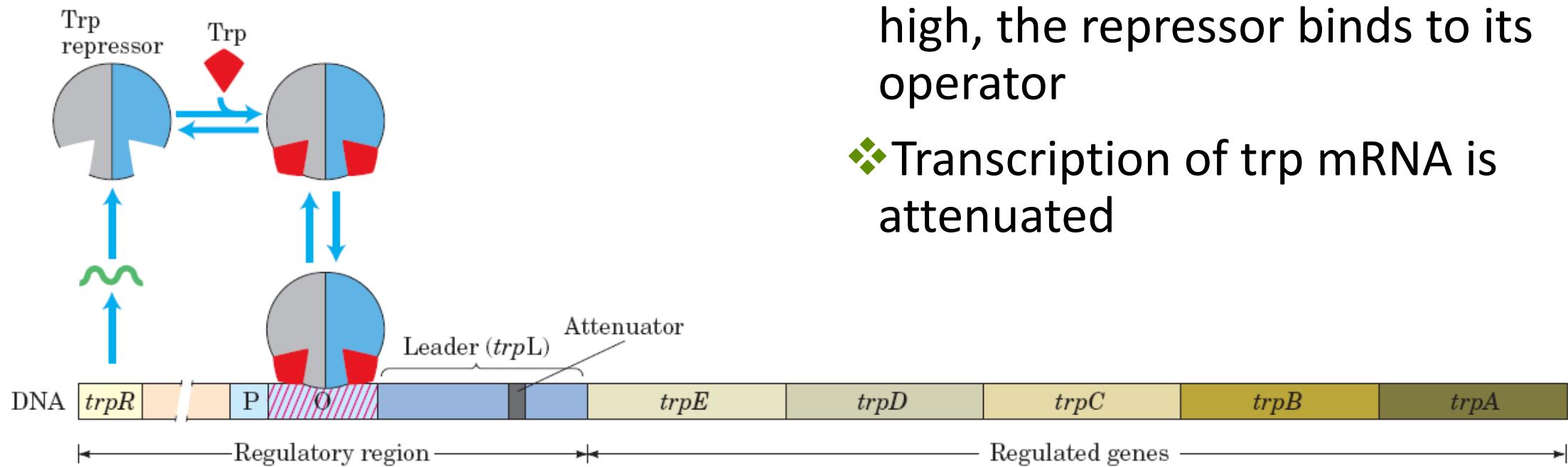


- ❖ O_1 is the main operator for the lac operon
- ❖ The Lac repressor binds to the main operator and O_2 or O_3 , apparently forming a loop in the DNA that might wrap around the repressor

The Lac Operon



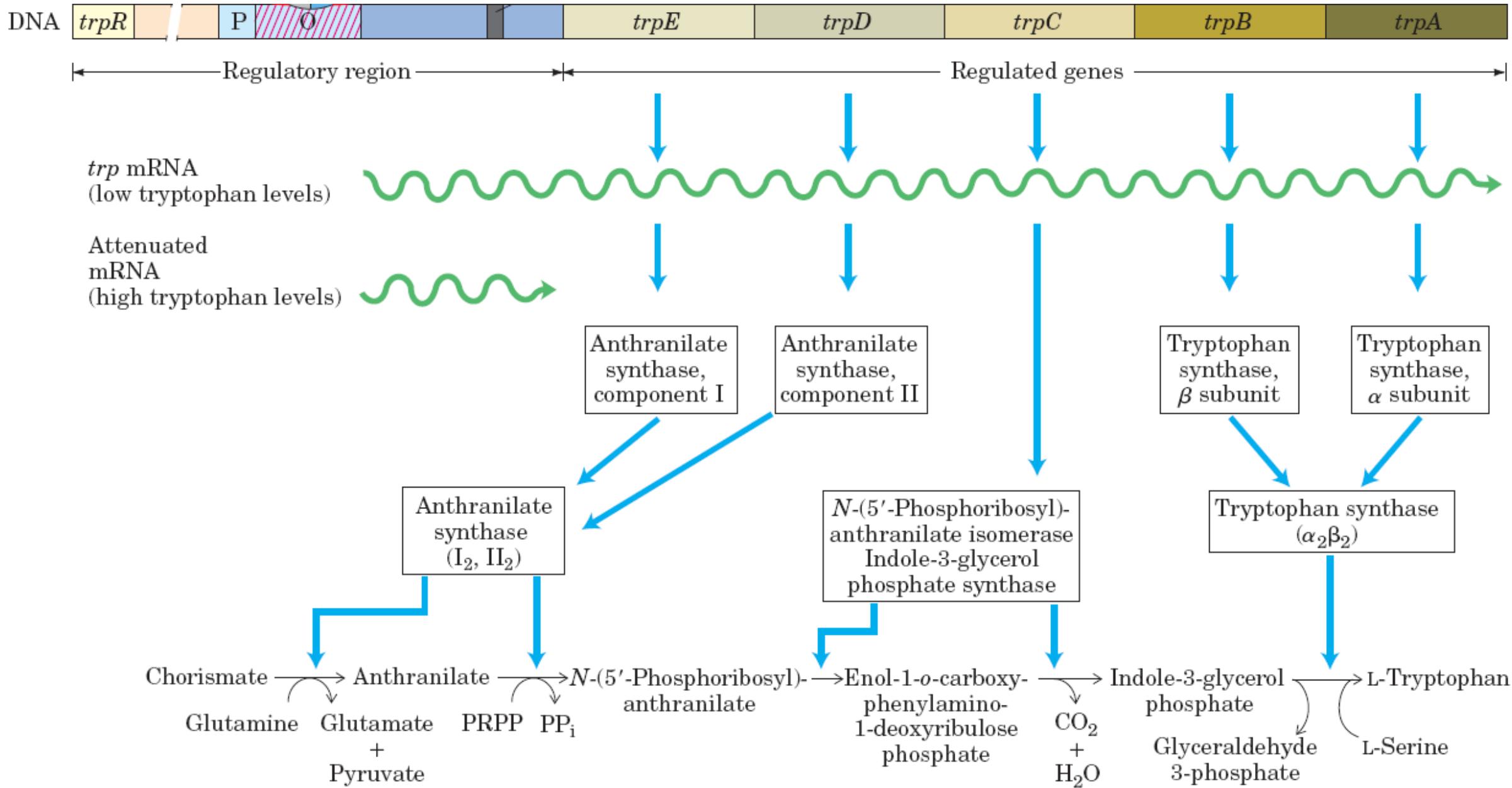
The Trp Operon



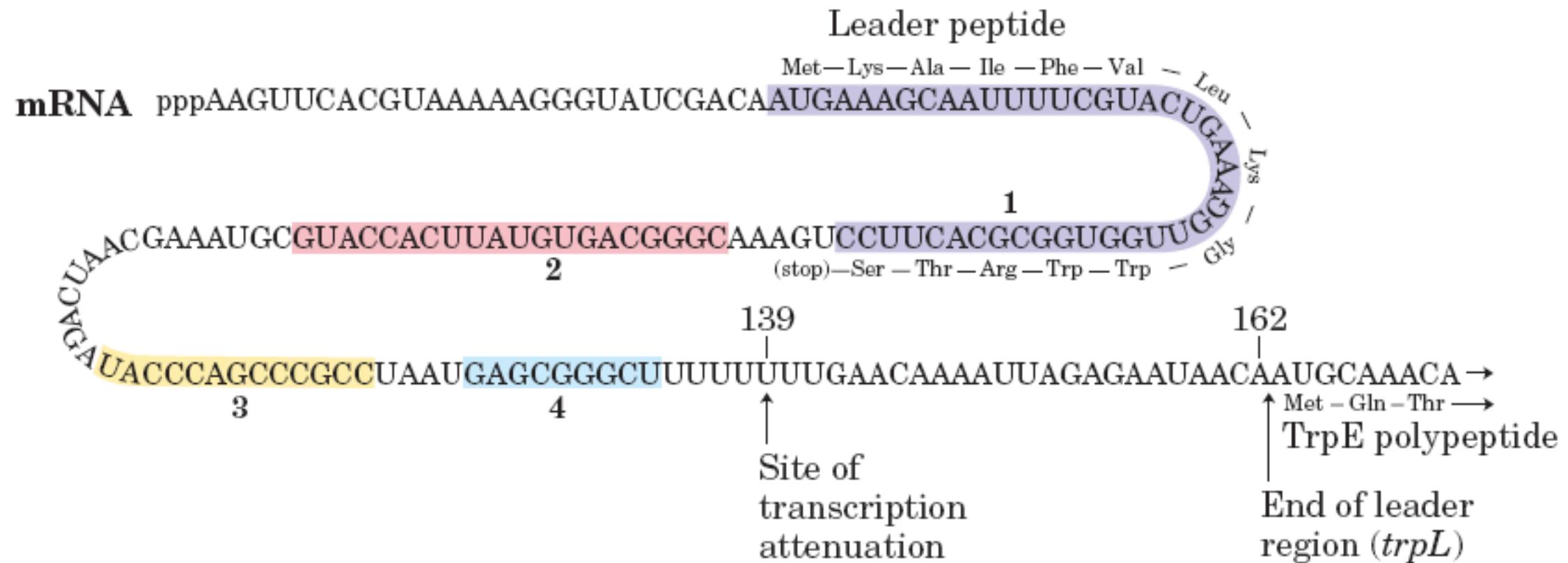
This operon is regulated by two mechanisms:

- ❖ When tryptophan levels are high, the repressor binds to its operator
- ❖ Transcription of trp mRNA is attenuated

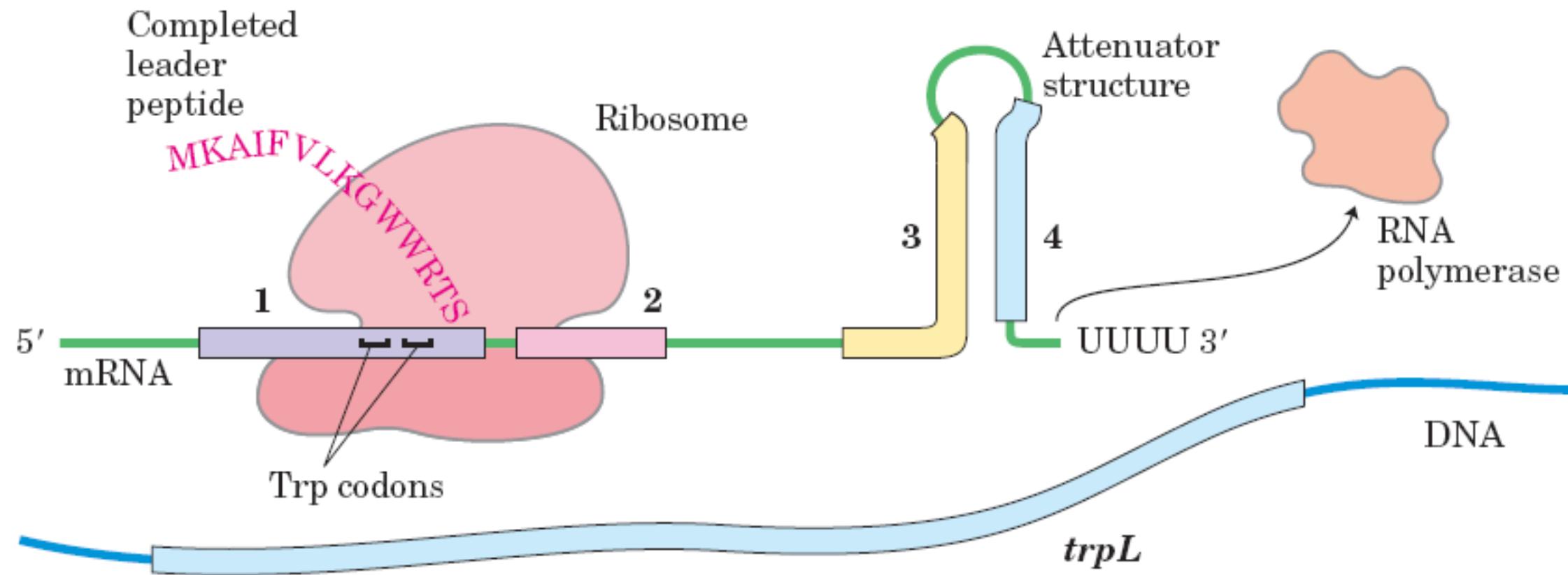
D N
C Q
K L
N Q
L P
Y R
E I
F W
L P
D G
K D
G



The Trp mRNA Sequence

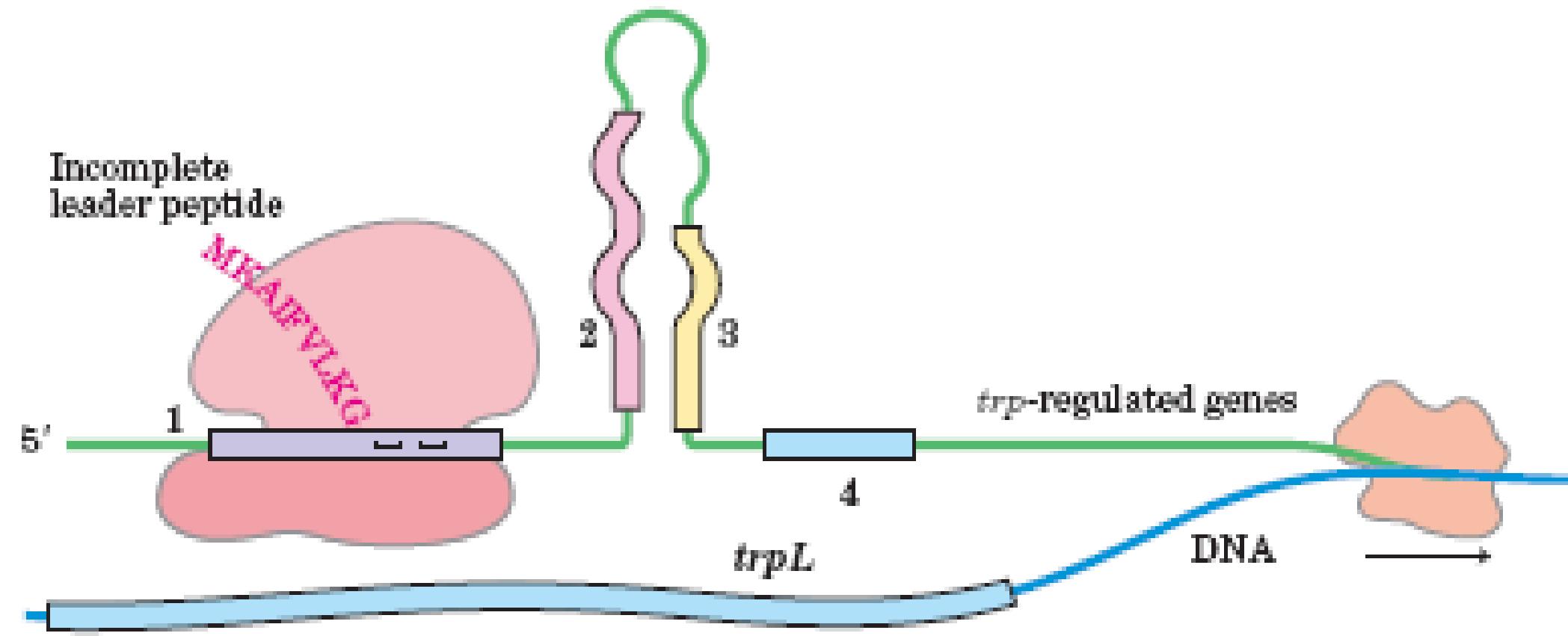


What happens at high Tryptophan levels



- ❖ When tryptophan levels are high, the ribosome quickly translates sequence 1 (open reading frame encoding leader peptide) and blocks sequence 2 before sequence 3 is transcribed. Continued transcription leads to attenuation at the terminator-like attenuator structure formed by sequences 3 and 4

What happens at low Tryptophan levels



- ❖ When tryptophan levels are low, the ribosome pauses at the Trp codons in sequence 1. Formation of the paired structure between sequences 2 and 3 prevents attenuation, because sequence 3 is no longer available to form the attenuator structure with sequence 4. The 2:3 structure, unlike the 3:4 attenuator, does not prevent transcription.

Computational Aspects of Gene Regulation

Lecture 13

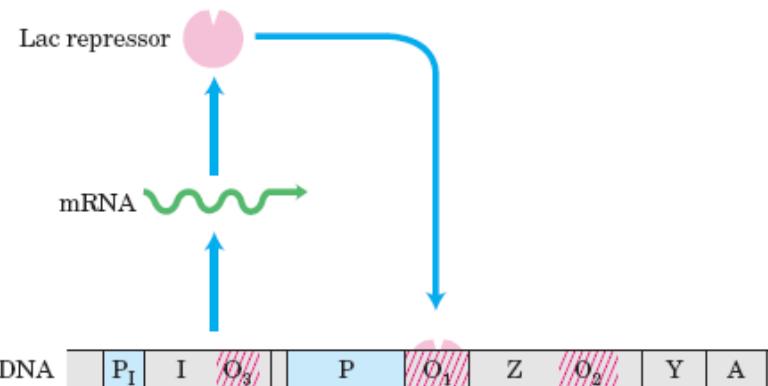
Objectives

In this lecture you will learn about:

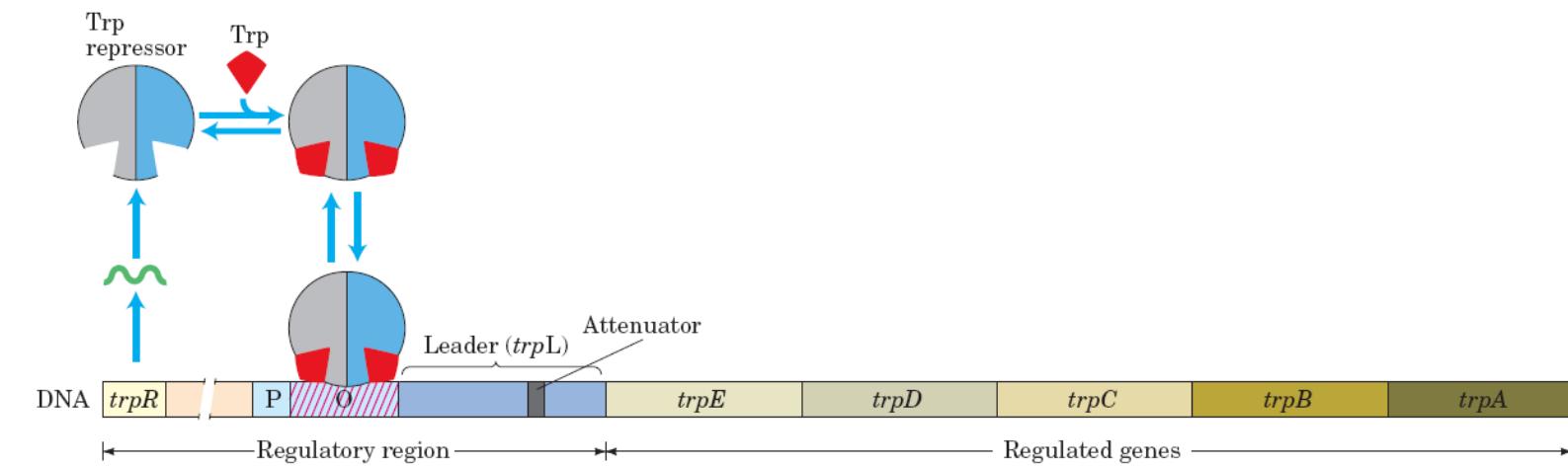
1. What are gene regulatory networks
2. Steps in building a gene regulatory network
3. Model of a transcription module
4. Network motifs and logic gates
5. Boolean representations

The Lac and Trp Operons

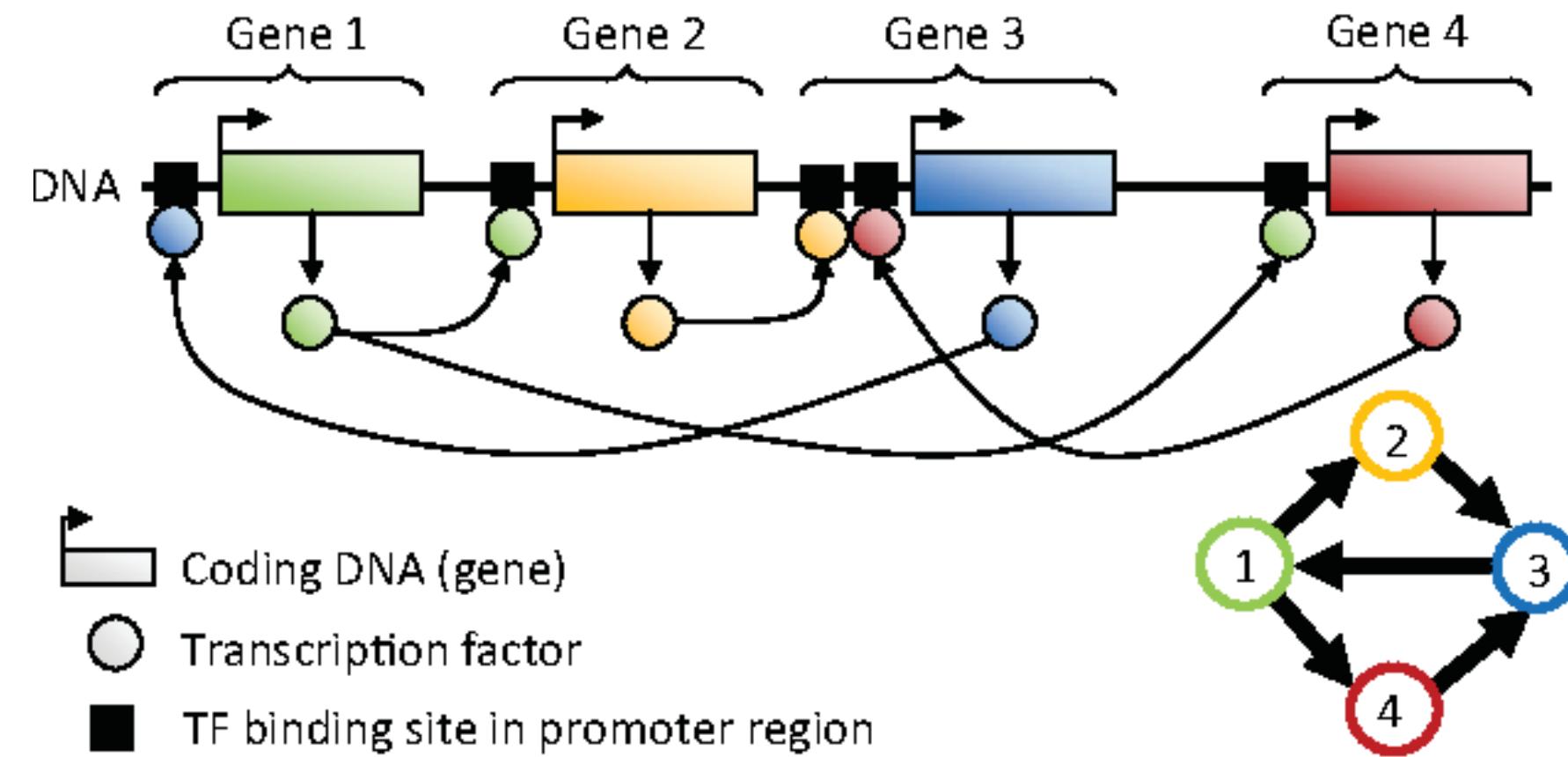
Lac Operon



Trp Operon



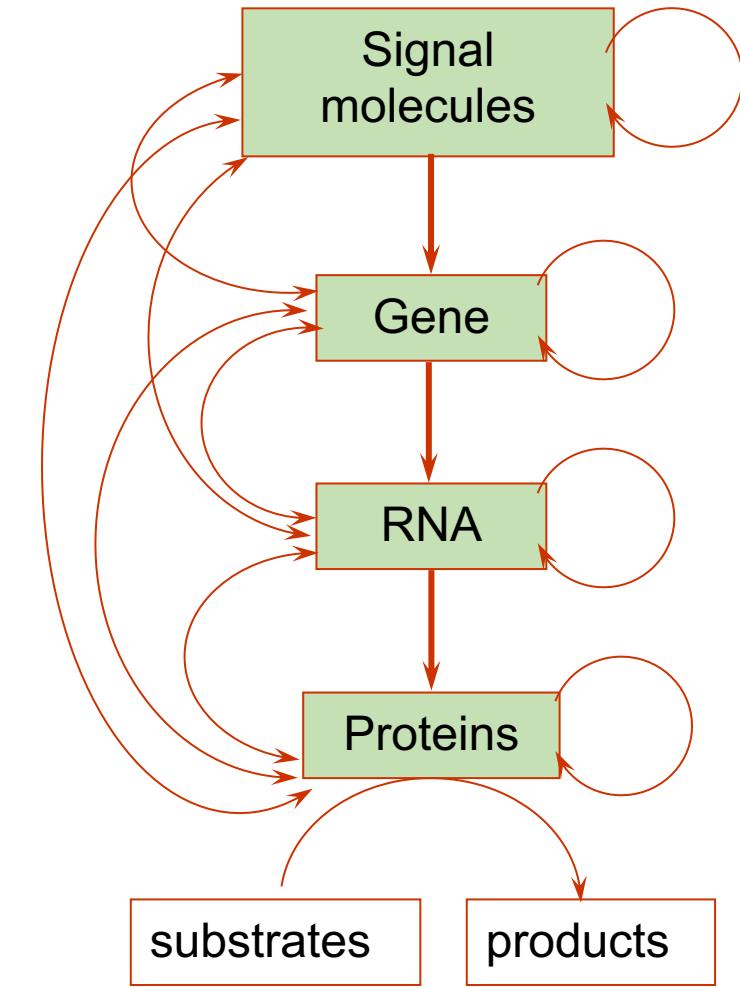
A typical gene regulatory network



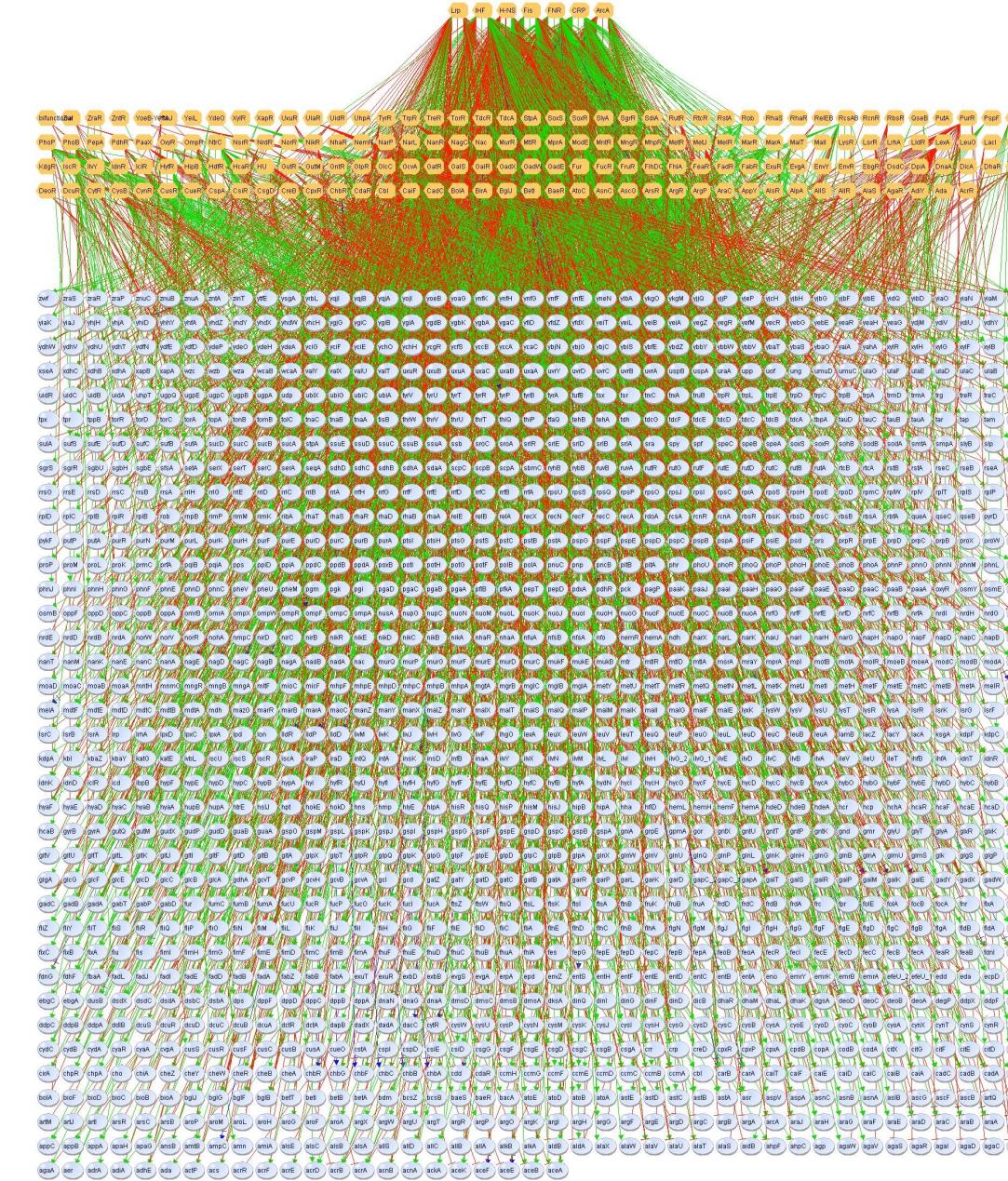
What are Gene Regulatory (Transcription) Networks?

This is one of the layers of information generation and transmission within a cell

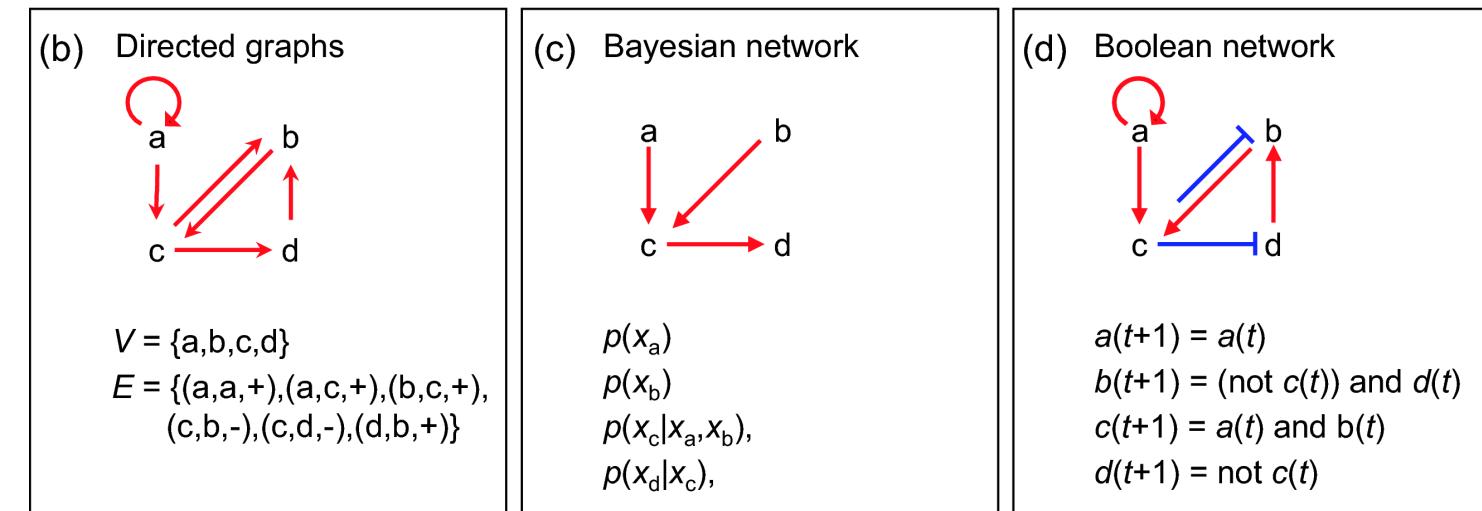
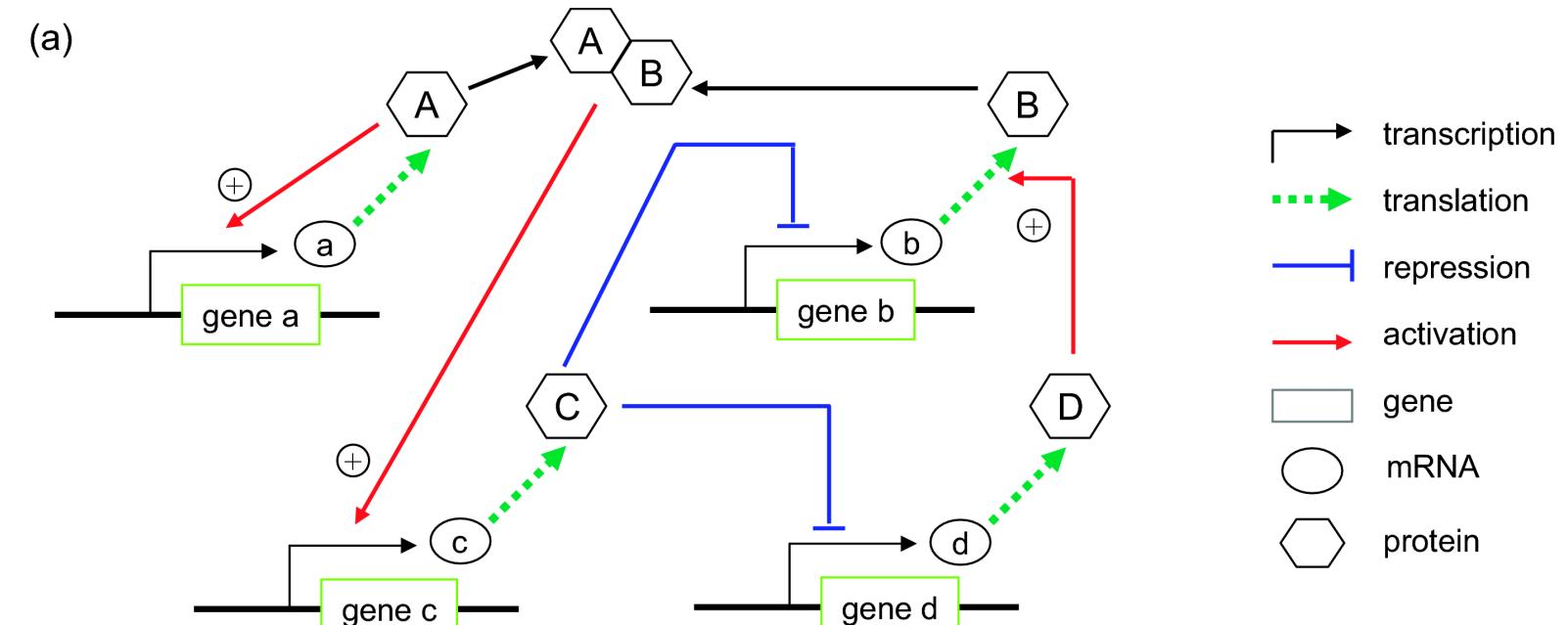
- ❖ It is the regulation of gene expression at the stage of transcription
- ❖ This is broadly the first step to gene expression and its control regulations the temporal programming in genes
- ❖ It is of interest because it mediates changes in cells and helps in understanding the onset and progression of disease



Regulatory Network of *E. coli* K12



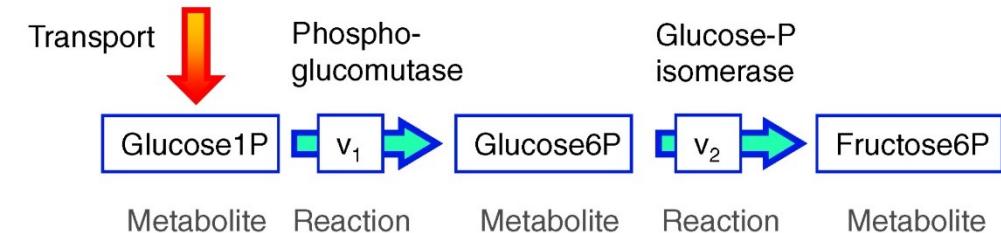
There are different methods of analysis



Steps to follow for building a regulatory network model

- ❖ Identify the elements of the model
- ❖ Characterize the kind of interaction/reaction
- ❖ Define the boundary of your observation (system)
- ❖ Identify the information/flow “into” and “out of” the boundary
- ❖ What are the intrinsic generation/degradation rates?
- ❖ Are all the parameters known?
- ❖ Assign the kinetics
- ❖ Code and Simulate

Basic Elements of Metabolic Networks

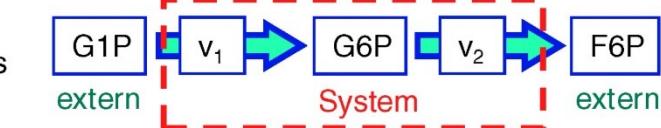


Design of Structured Dynamic Models

1. Setting system limits

2. Balancing

3. Assignment of Kinetics

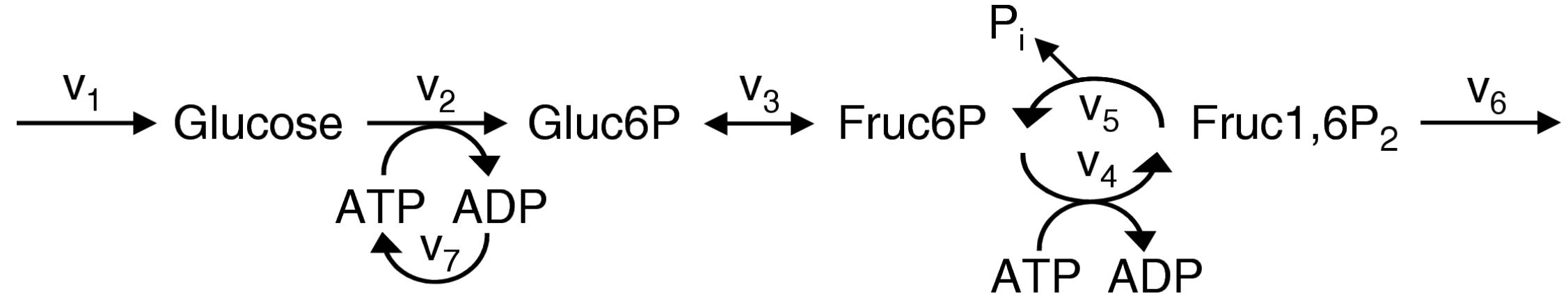


$$\frac{d}{dt}G6P = v_1 - v_2$$

$$v_1 = \frac{V_{\max,1} \cdot G1P}{K_{m,1} + G1P}$$

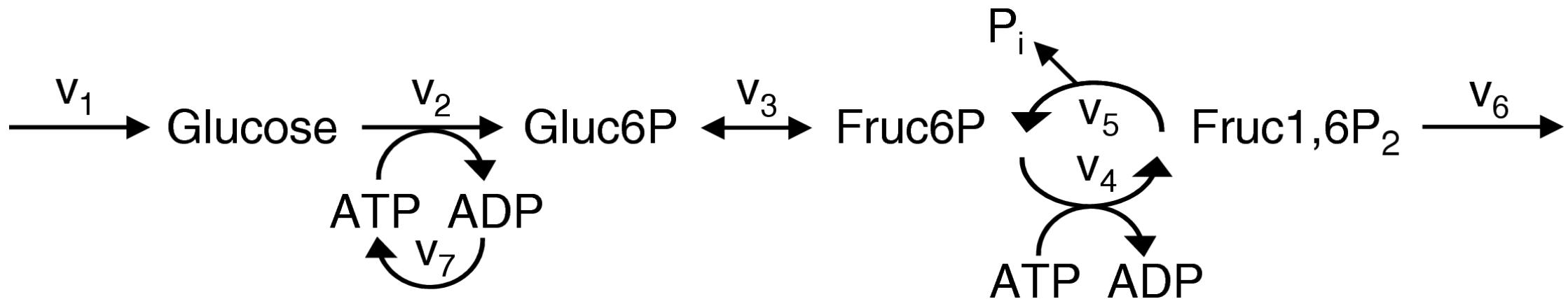
$$v_2 = \frac{V_{\max,2} \cdot G6P}{K_{m,2} + G6P}$$

Add the details of the important constituents



- ❖ Layout the reactions involved
- ❖ Are the rates balanced?

Define the rate equations



- ❖ This is the most important step
 - ❖ Go to the literature and ensure correctness of the reactions
 - ❖ Use various resources to determine the parameters that have been reported for the same or similar reactions (or cellular events)
 - ❖ Do your experiments results appear reasonable? Have the parameters been evaluated correctly?

$$\frac{d}{dt} Glu = v_1 - v_2$$

$$\frac{d}{dt} G6P = v_2 - v_3$$

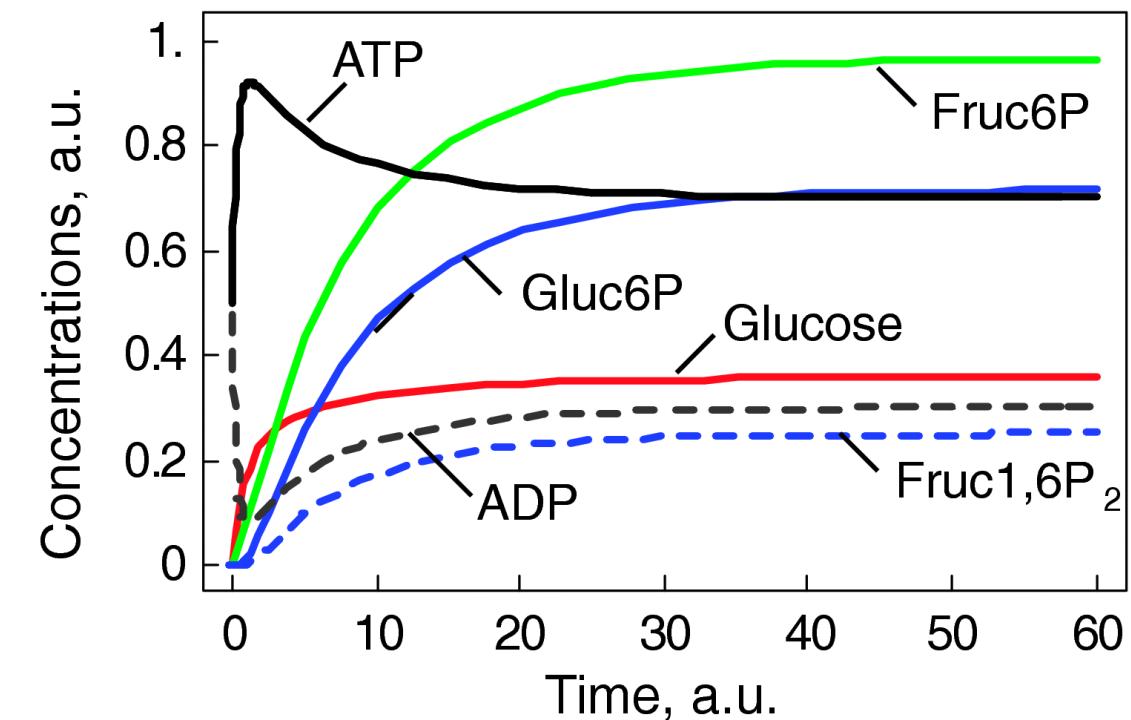
$$\frac{d}{dt} F6P = v_3 - v_4 + v_5$$

$$\frac{d}{dt} F1,6P2 = v_4 - v_5 + v_6$$

$$\frac{d}{dt} ATP = - \frac{d}{dt} ADP = -v_2 - v_4 + v_7$$

Performing simulation and analysis of results

- ❖ Cross check the results to make sure that it makes sense
 - ❖ Check the boundary results
 - ❖ Do the values at the boundaries satisfy the physical constraints?
 - ❖ Plot the data/results and analyze
- ❖ What is the insight you have gained?
- ❖ What is the hypothesis you can propose?



Transcription Modules

- In the general form, the ODE that models the output of the gene (Z) in response to a regulatory input S is given by the equation

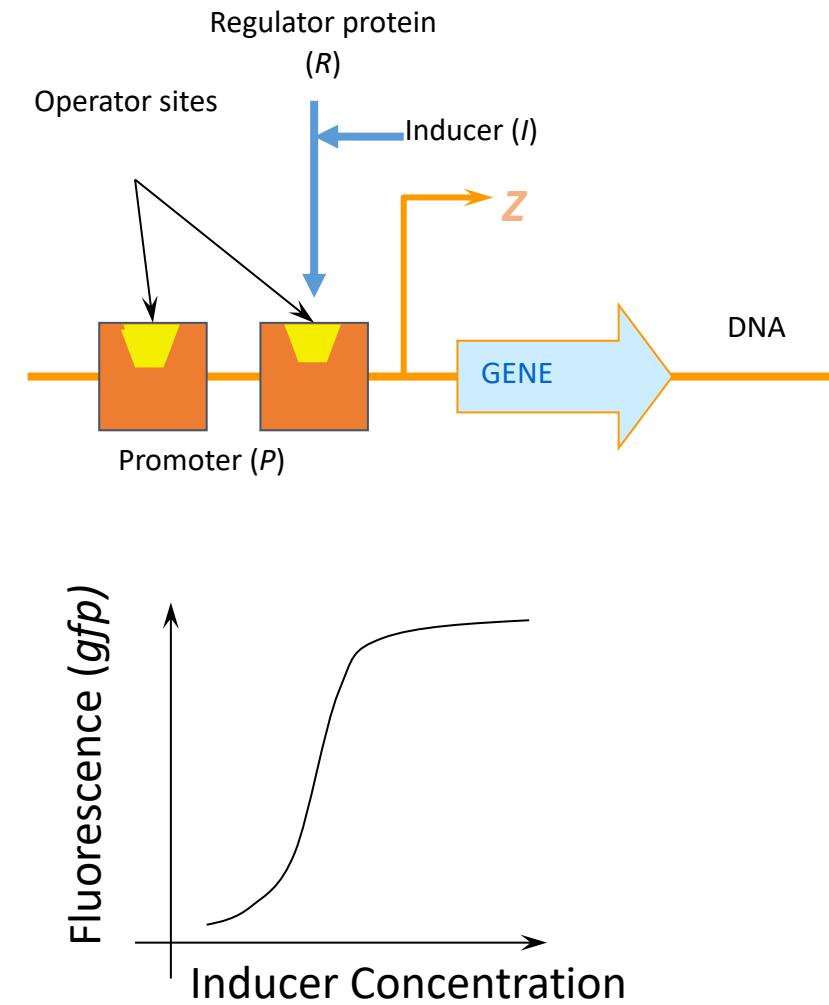
$$\frac{dZ}{dt} = k' + \frac{k \cdot (S^n / K^n)^\mu}{1 + (S^n / K^n)} - k_d \cdot Z$$

$$Z_{ss} = \frac{k}{k_d} \left(a + \frac{(S^n / K^n)^\mu}{1 + (S^n / K^n)} \right), \quad k' = a \cdot k$$

- Repression and Activation are taken care of by the parameter μ

$\mu = 0 \rightarrow$ repression; $\mu = 1 \rightarrow$ activation

- The parameters k' and k , represent the signal-independent and the signal-dependent gene expression.



Example of a cascade of genes

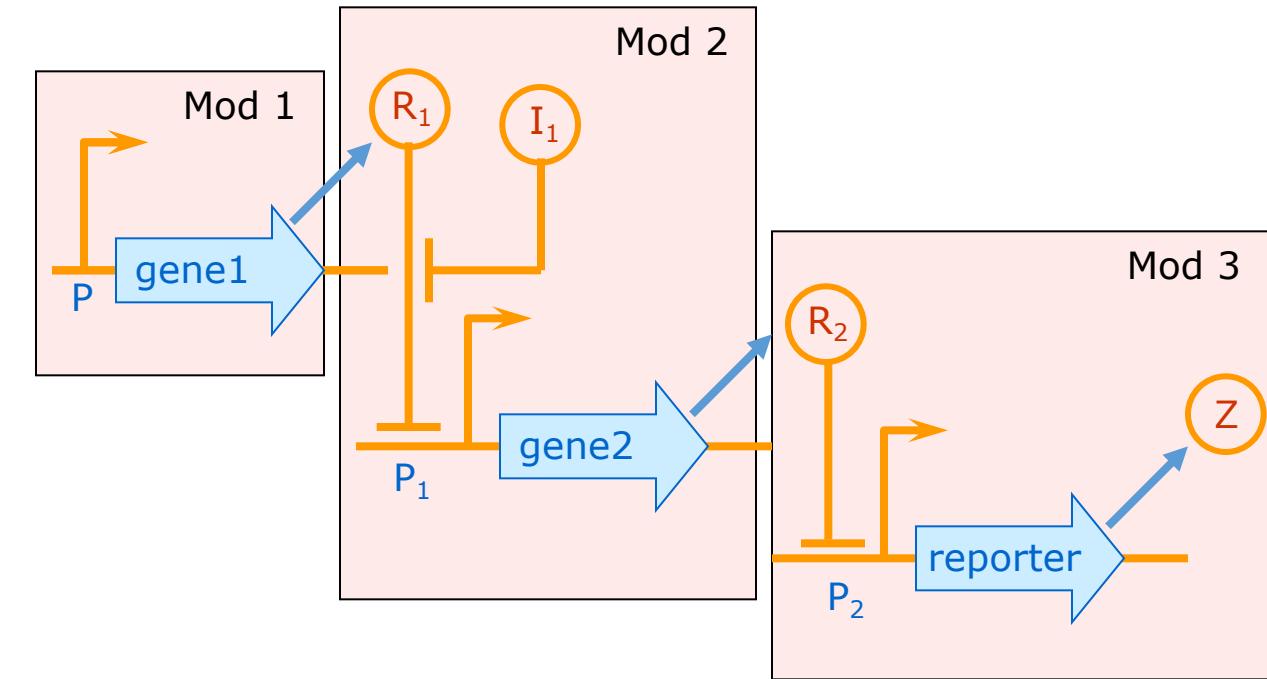
- ❖ Promoter P has no regulatory inputs

- ❖ P is constitutive and drives the expression of R_1 which in turn inhibits R_2

- ❖ The inducer I_1 regulates the signal R_2/P_2 by modulating the cellular abundance of R_2

$P1: n = 2.4, K = 5.5 \text{ nM}, k = 220 \text{ min}^{-1}$,

$P2: n = 1.7, K = 120 \text{ nM}, k = 255 \text{ min}^{-1}$,



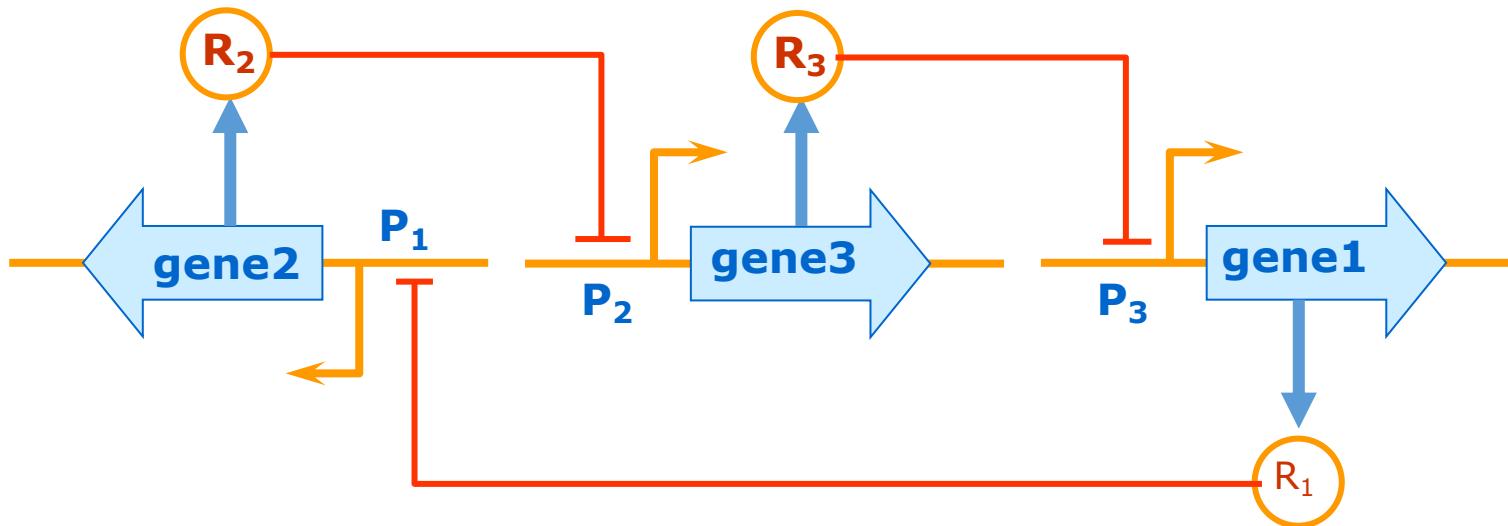
$$\frac{dR_2}{dt} = a_1 k_1 + \frac{k_1 \cdot (I_1 / K_1)^{n_1}}{1 + (I_1 / K_1)^{n_1}} - k_{d2} \cdot R_2$$

$$\frac{dZ}{dt} = a_2 k_2 + \frac{k_2}{1 + (R_2 / K_2)^{n_2}} - k_d \cdot Z$$

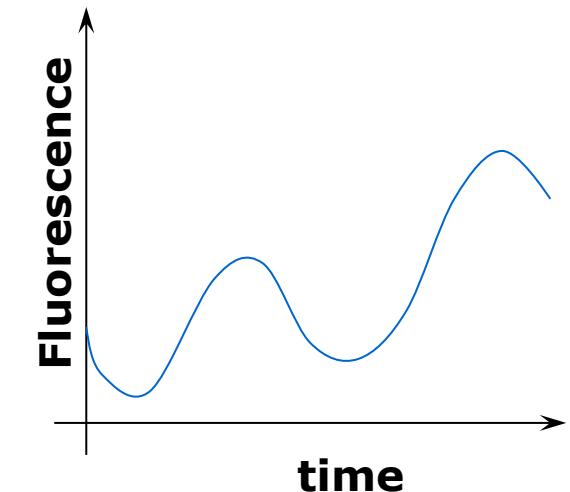
$$R_{2ss} = \frac{k_1}{k_{d2}} \left(a_1 + \frac{k_1 \cdot (I_1 / K_1)^{n_1}}{1 + (I_1 / K_1)^{n_1}} \right)$$

$$Z_{ss} = \frac{k_2}{k_d} \left(a_2 + \frac{k_2}{1 + (R_{2ss} / K_2)^{n_2}} \right)$$

Oscillatory Networks – The Repressilator



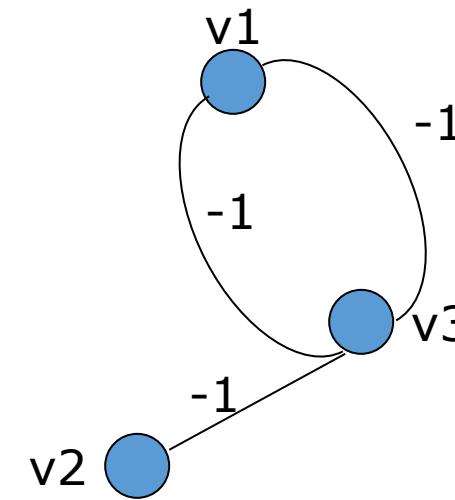
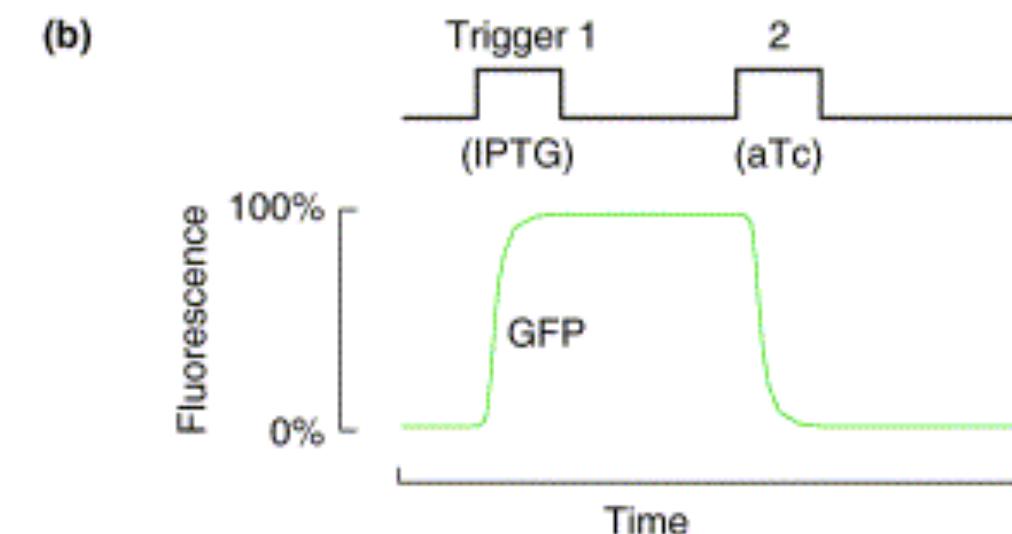
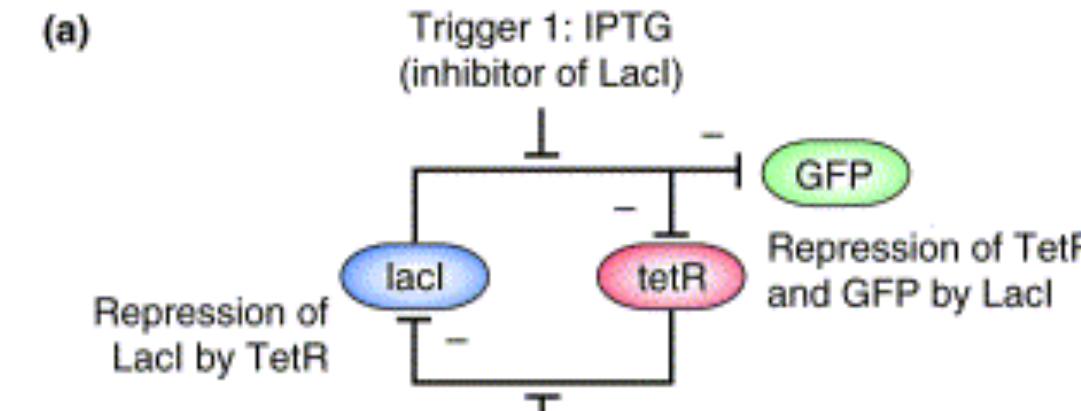
- ❖ Repressor R_1 inhibits the expression of repressor R_2 , repressor R_2 inhibits the expression of repressor R_3 , and repressor R_3 inhibits the expression of repressor R_1
- ❖ The separation of transcription and translation contributes to a response delay that results in the emergence of oscillations



$$\frac{dm_i}{d\tau} = \alpha\kappa + \frac{\kappa}{1 + r_j^n} - m_i$$

$$\frac{dr_i}{d\tau} = \varepsilon(m_i - r_i)$$

Regulatory circuits can also be represented as graphs

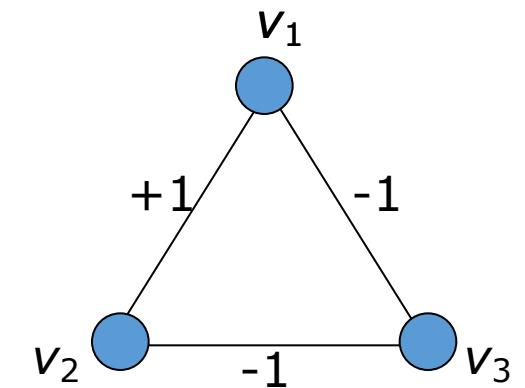


Signed Graphs

A *signed graph* S is an undirected network whose edges have functional values of $+1$ or -1 ; it is natural to refer to them as a positive edge or negative edge.

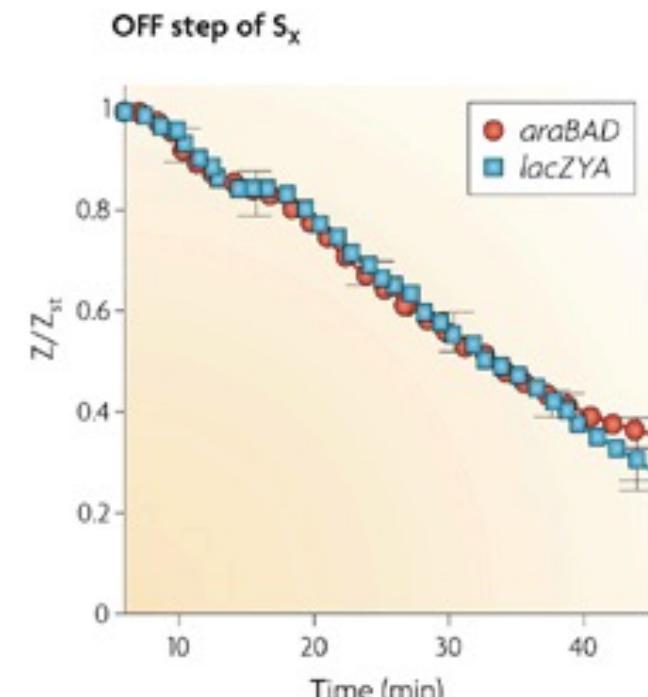
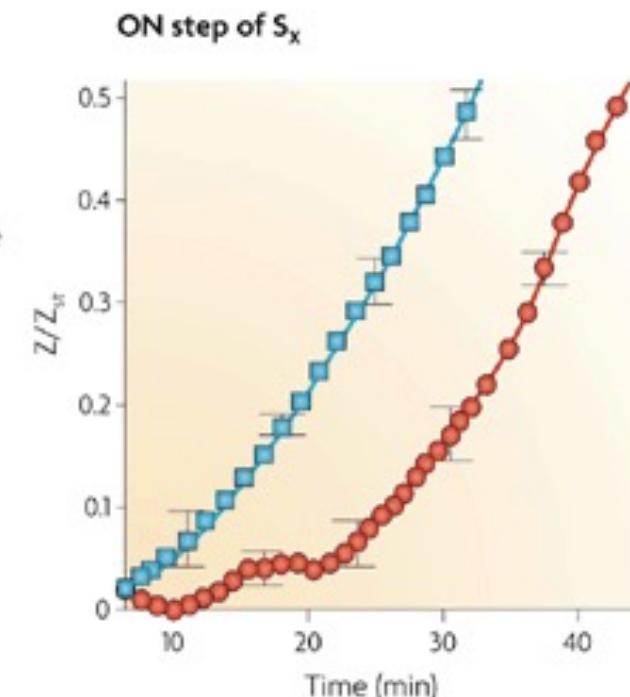
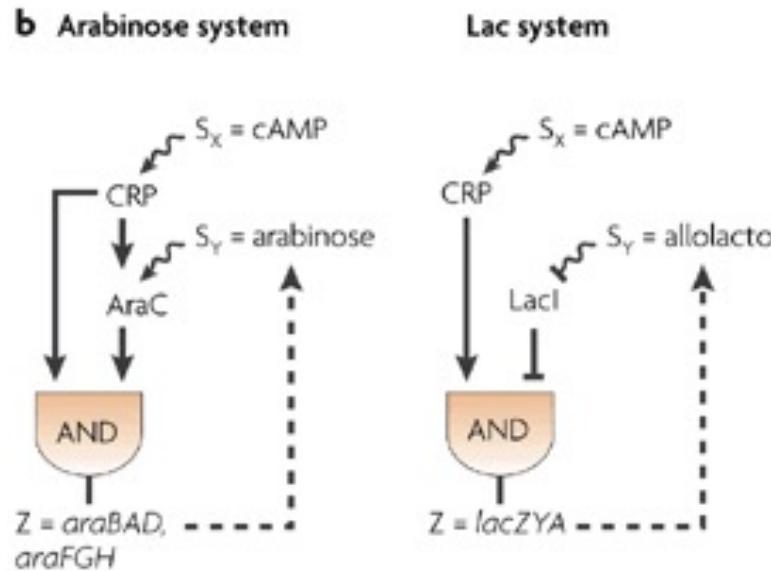
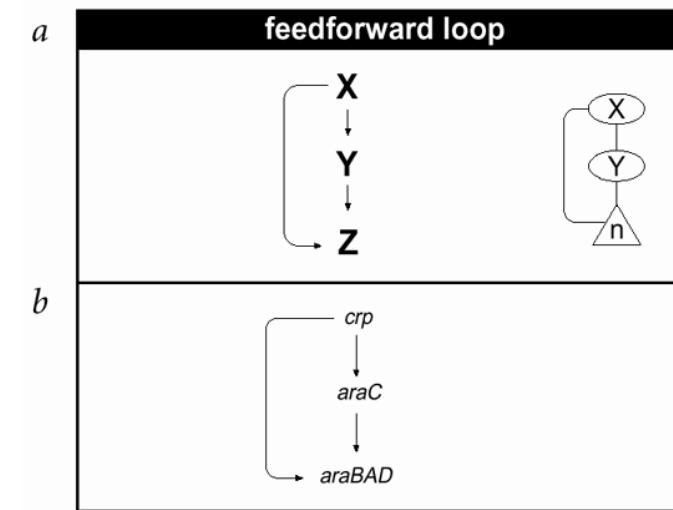
For example:

$$V = (v_1, v_2, v_3), E = (v_1v_2, v_2v_3, v_3v_1) \text{ and } f = \{(v_1v_2, +1) (v_2v_3, -1) (v_3v_1, -1)\}$$



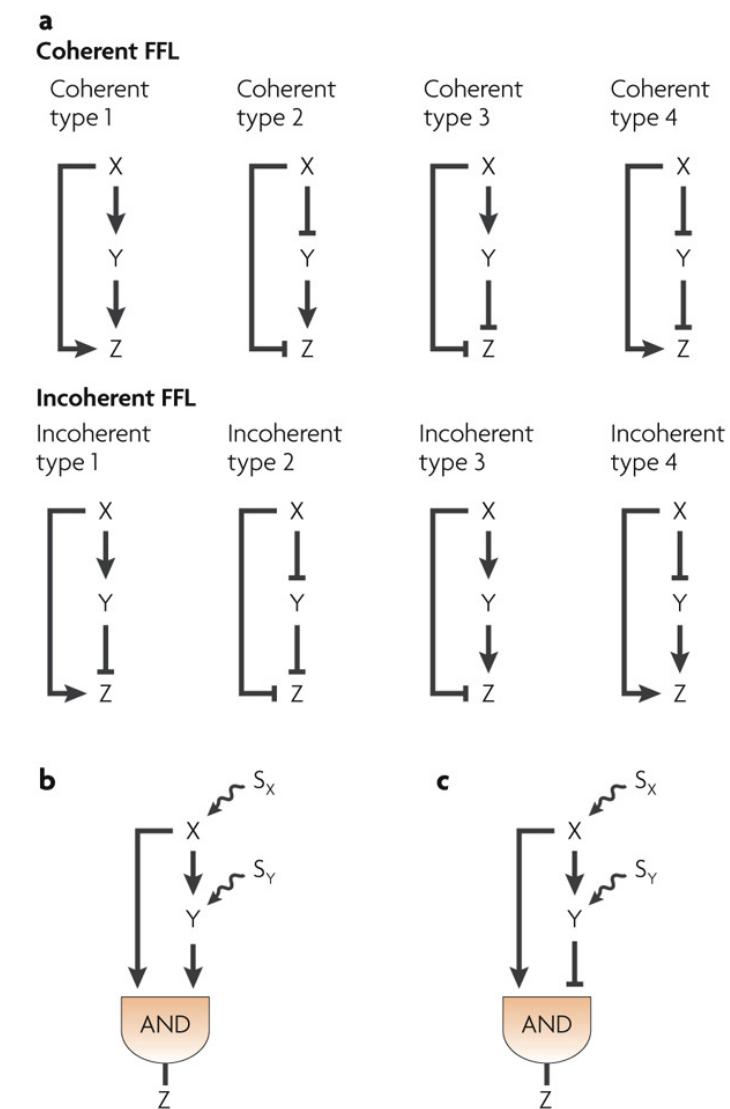
Gene Regulation can also be treated as logic gates

- ❖ Arabinose is only used if glucose is not present; proteins in this system are made only when condition arabinose “AND NOT” glucose is satisfied
 - ❖ The delay $T_{ON} \sim 20$ min
 - ❖ $X = CRP$, $S_x = cAMP$, $S_y = \text{arabinose}$, $Y = araC$,
- In the lactose operon, X does not regulate $Y = lacI$



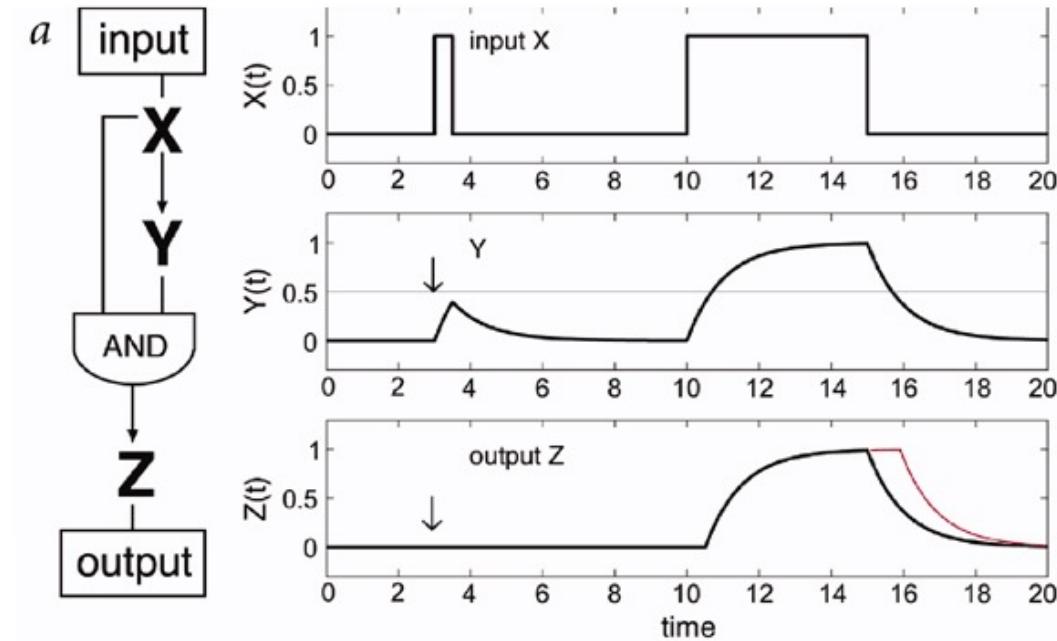
Structure of the feed-forward loop

- ❖ The feedforward loop in consideration
 - ❖ Has a direct path from X to Z
 - ❖ has an indirect path X to Y to Z
- ❖ Each edge can be an activation or repression; so there are $2^3=8$ FFLs
- ❖ These are classified into two groups
 - ❖ Coherent: the indirect path has the same overall sign as the direct path
 - ❖ Incoherent: the sign of the indirect path is opposite to that of the direct path
- ❖ There are two possible logic gates for the expression of Z : AND or OR



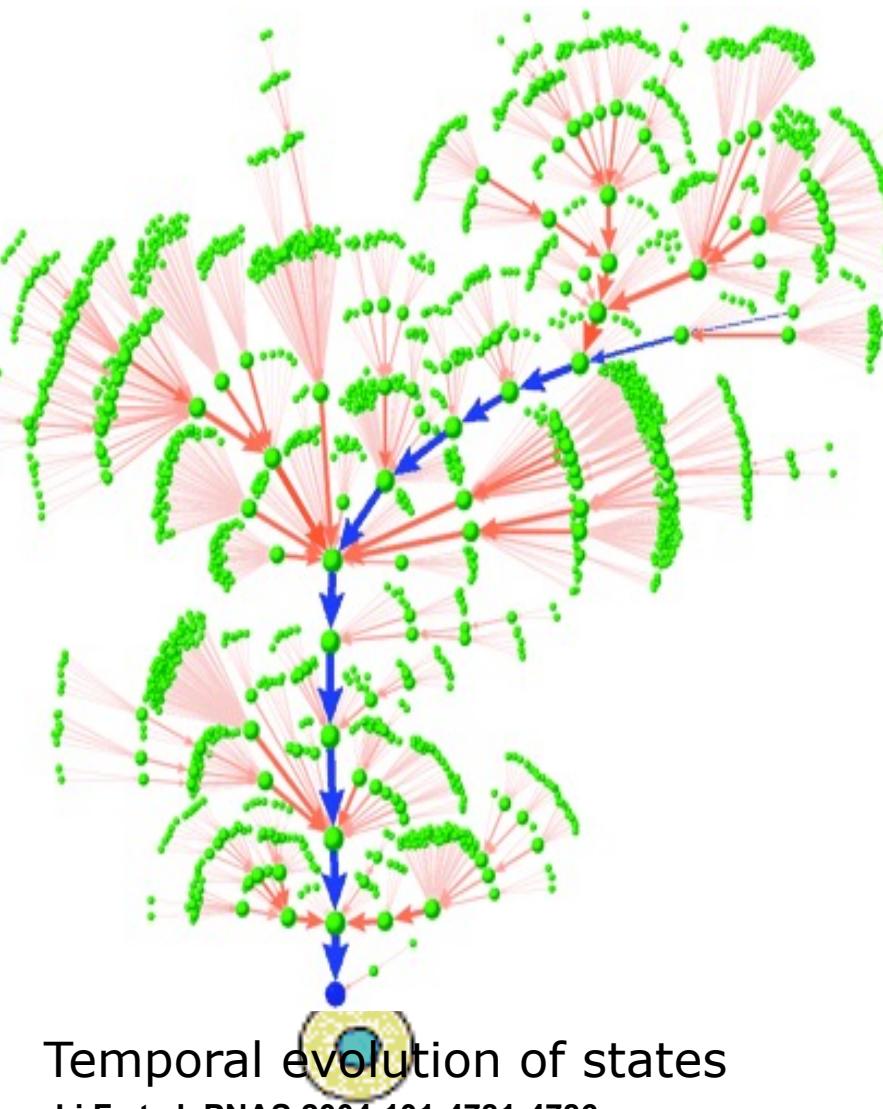
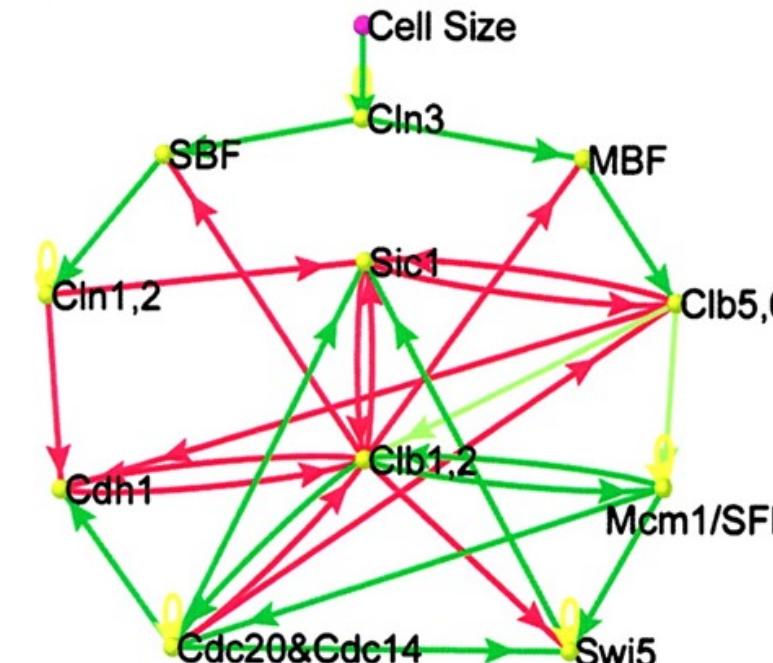
The FFL is a persistence detector

- ❖ A sign sensitive delay element can be considered as an asymmetric filter
- ❖ A brief pulse of X is results in a signal shorter than T_{ON}
- ❖ However, the motif responds immediately to a pulse OFF signal



Sign sensitive delays protect the gene circuit – the synthesis of Z is not initiated until the signal is confirmed – thereby energy is conserved

Yeast cell cycle



Boolean Attractors

- ❖ A boolean network is defined by $G(V, F)$

$$V = \{v_1, v_2, \dots, v_n\}$$

$$F = \{f_1, f_2, \dots, f_n\}$$

- ❖ Let $v_i(t)$ represent the state of v_i at time t . The overall expression level of all the genes in the network at time step t is given by the following vector:

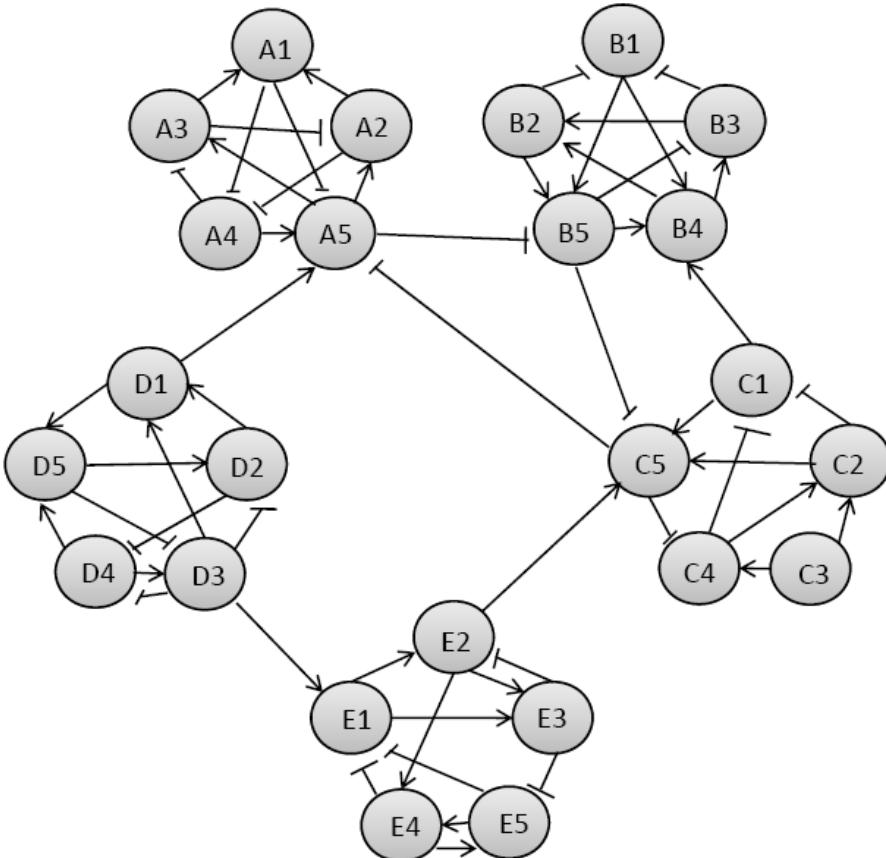
$$v(t) = [v_1(t), v_2(t), \dots, v_n(t)]$$

- ❖ There are 2^n possible states; the regulatory rules among the genes are given as follow

$$v_i(t+1) = f_i(v_{i_1}(t), v_{i_2}(t), \dots, v_{i_{k_i}}(t)), \quad i = 1, 2, \dots, n.$$

Developing New Methods of Analysis – An Example

- Re
- ne
- • Ne
- Ob
- Bo
- fin

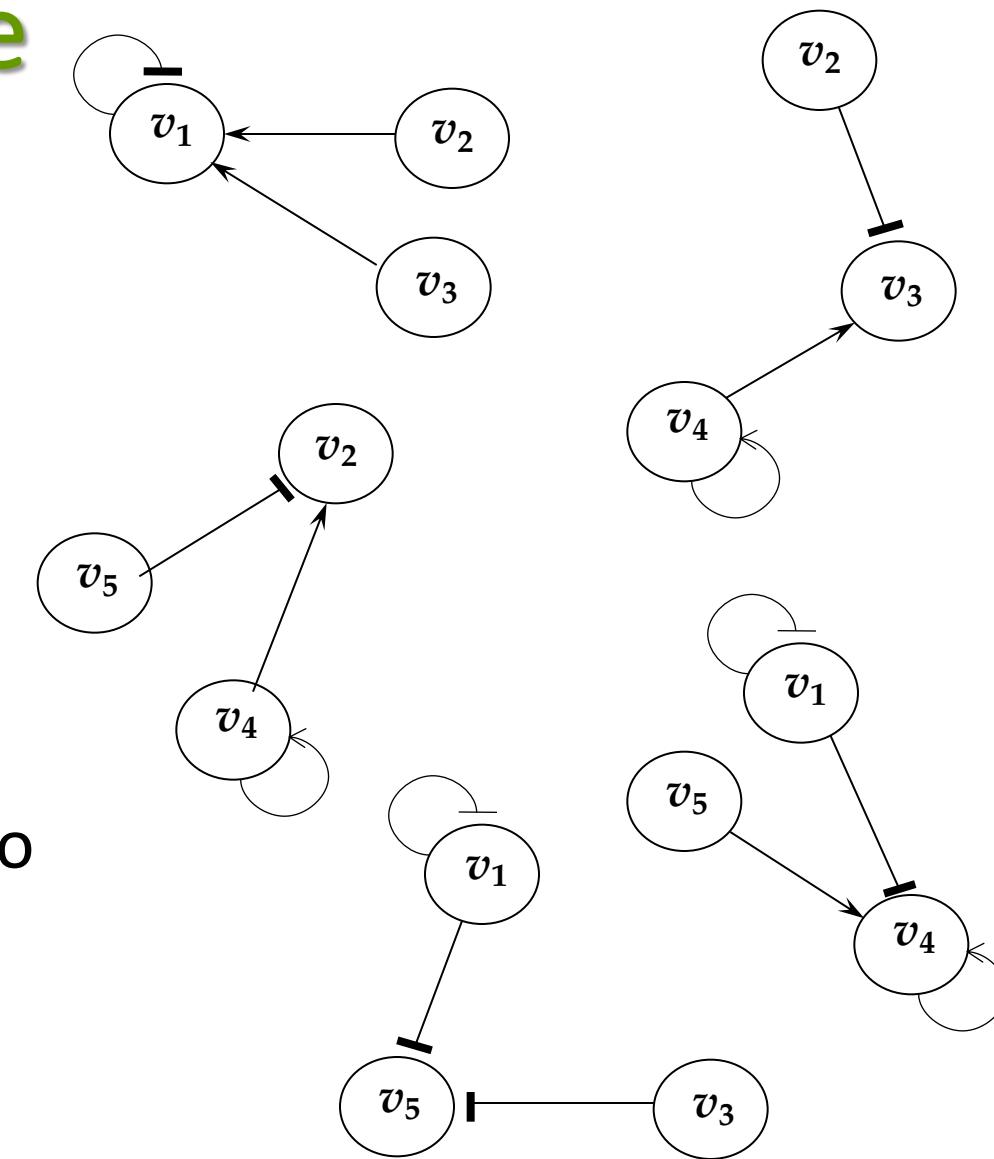
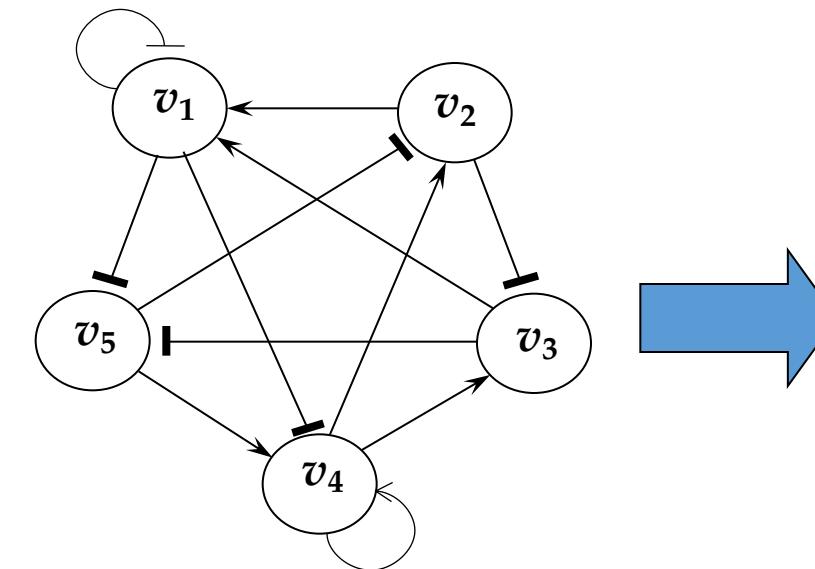


building the global

ysis
ere it is difficult to
 2^N
 $\sum_i 2^{K_i} < 2^N$

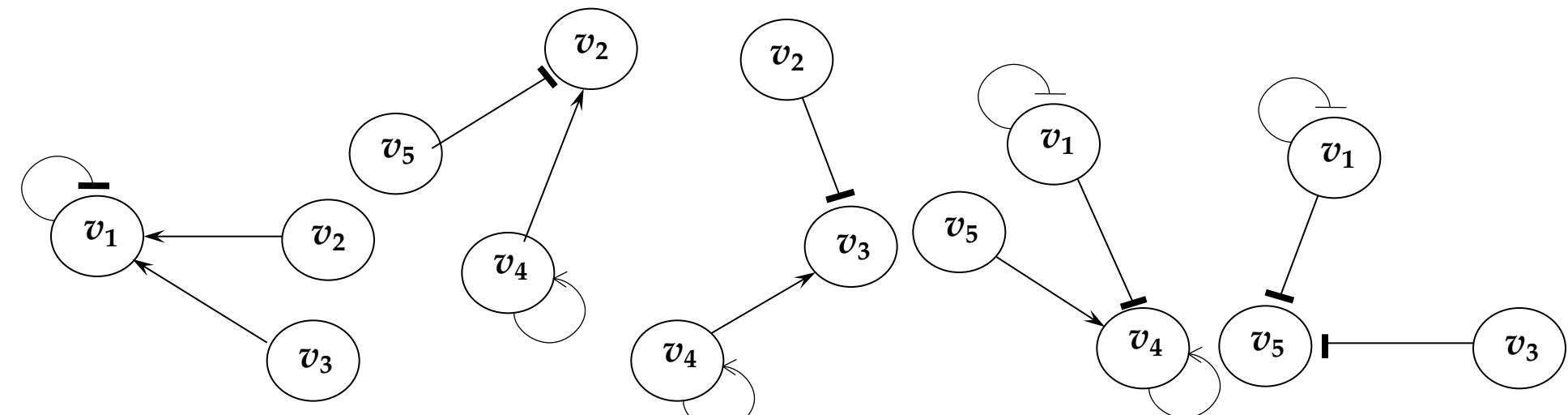
More than 100 times faster
for real networks

Toy Network Example



- ❖ Divide the network into subgraphs of individual nodes

Determine the steady states of subgraphs



V_1	V_2	V_3
0	0	0
1	0	1
1	1	0
1	1	1

V_2	V_4	V_5
0	0	0
0	1	0
0	1	1
1	0	0
1	0	1
1	1	1

V_3	V_2	V_4
0	0	0
0	1	0
0	1	1
1	0	0
1	0	1
1	1	1

V_4	V_1	V_5
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0

V_5	V_1	V_3
0	0	0
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1