Science of Living System

BS20001

Arindam Mondal
School of Bioscience
Email: ***arindam.mondal@iitkgp.ac.in***
Tel: 03222-260516

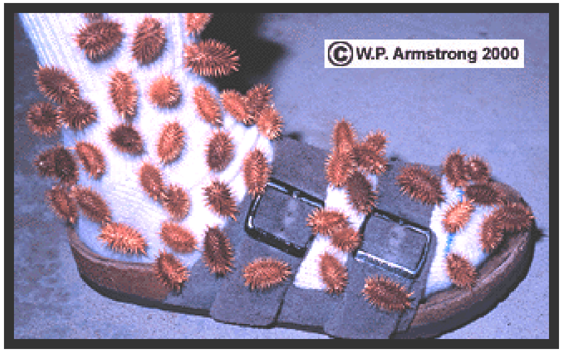
Why study Living System?

Shinkansen Bullet Train in Japan: inspired by shape of Kingfisher head

Eiji Nakatsu

Invention of Velcro®: inspired by Cockleburs

Cockleburs



George de Mestral invented Velcro®: Unique, two-sided fastener, one side with stiff **hooks** like the burs and the other side with soft **loops** like the fabric. a combination of the words velour and crochet.

Why study Living System?

STORAGE LIMITS

Estimates based on bacterial genetics suggest that digital DNA could one day rival or exceed today's storage technology.

	Hard disk	Flash memory	Bacterial DNA
Read–write speed (µs per bit)	~3,000–5,000	~100	<100
Data retention (years)	>10	>10	>100
Power usage (watts per gigabyte)	~0.04	~0.01–0.04	$<10^{-10}$
Data density (bits per cm³)	$\sim10^{13}$	$\sim10^{16}$	$\sim10^{19}$

WEIGHT OF DNA NEEDED TO STORE WORLD'S DATA

©nature

Microsoft Corporation – DNA Storage Research

A DNA-Based Archival Storage System Bornholt J, et. al. ASPLOS 2016 (International Conference on Architectural Support for Programming Languages and Operating Systems)

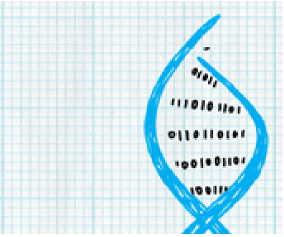
Microsoft Store Products Support

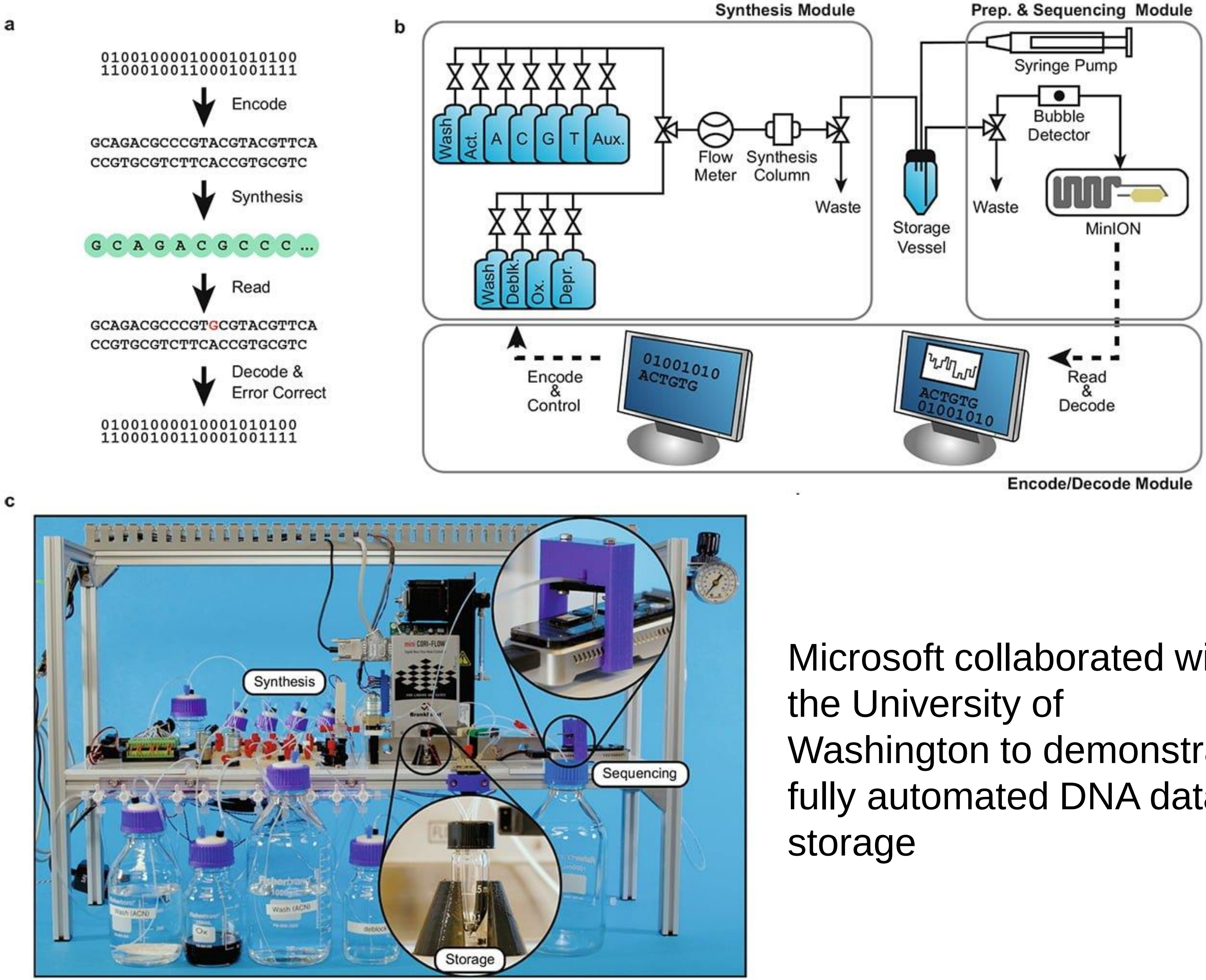
Research Research areas Products & Downloads Programs & Events People Careers About

DNA Storage

Established: January 1, 2015

The amount of digital data produced has long been outpacing the amount of storage available. This project enables molecular-level data storage into DNA molecules by leveraging biotechnology advances in synthesizing, manipulating and sequencing DNA to develop archival storage. Microsoft and University of Washington researchers are collaborating to use DNA as a high density, durable and easy-to-manipulate storage medium.





Microsoft collaborated with the University of Washington to demonstrate fully automated DNA data storage

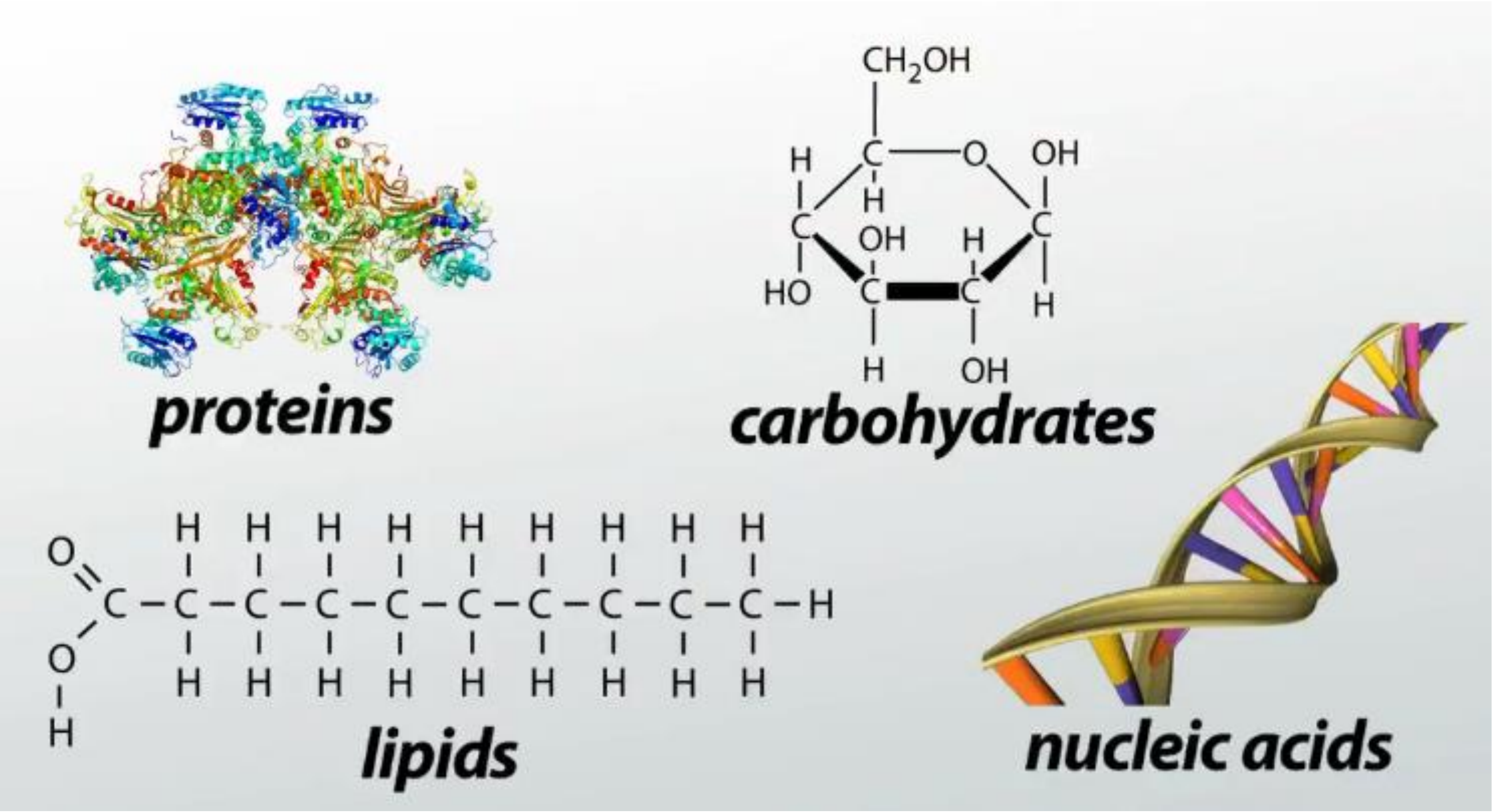
What are the characteristics of a Living System?

- High degree of **complexity**
- Mechanisms for **sensing** and **responding** to alterations in surroundings
- Systems for extracting, transforming and using **energy** from the environment
- Ability to **adapt** and **evolve**
- Ability to develop and **grow**
- Capacity for precise **self-replication** and **self-assembly**, known as reproduction

Components of Living Systems
C-H-O-N-S-P

- All life forms on Earth are composed of building blocks that are made of combinations of Carbon and other elements:
- Hydrogen
- Oxygen
- Nitrogen
- Sulphur
- Phosphorus

Molecules of Living Systems: **Biological Macromolecules**



Science of Living System (BS20001)

- **SOUMYA DE**
- **Office:** Rm# 219, Life Sciences Building
- **Phone:** 03222-284552
- **E-mail:** somde@iitkgp.ac.in
- **Website:** http://www.iitkgp.ac.in/department/BS/faculty/bs-somde

- **Research interests:**

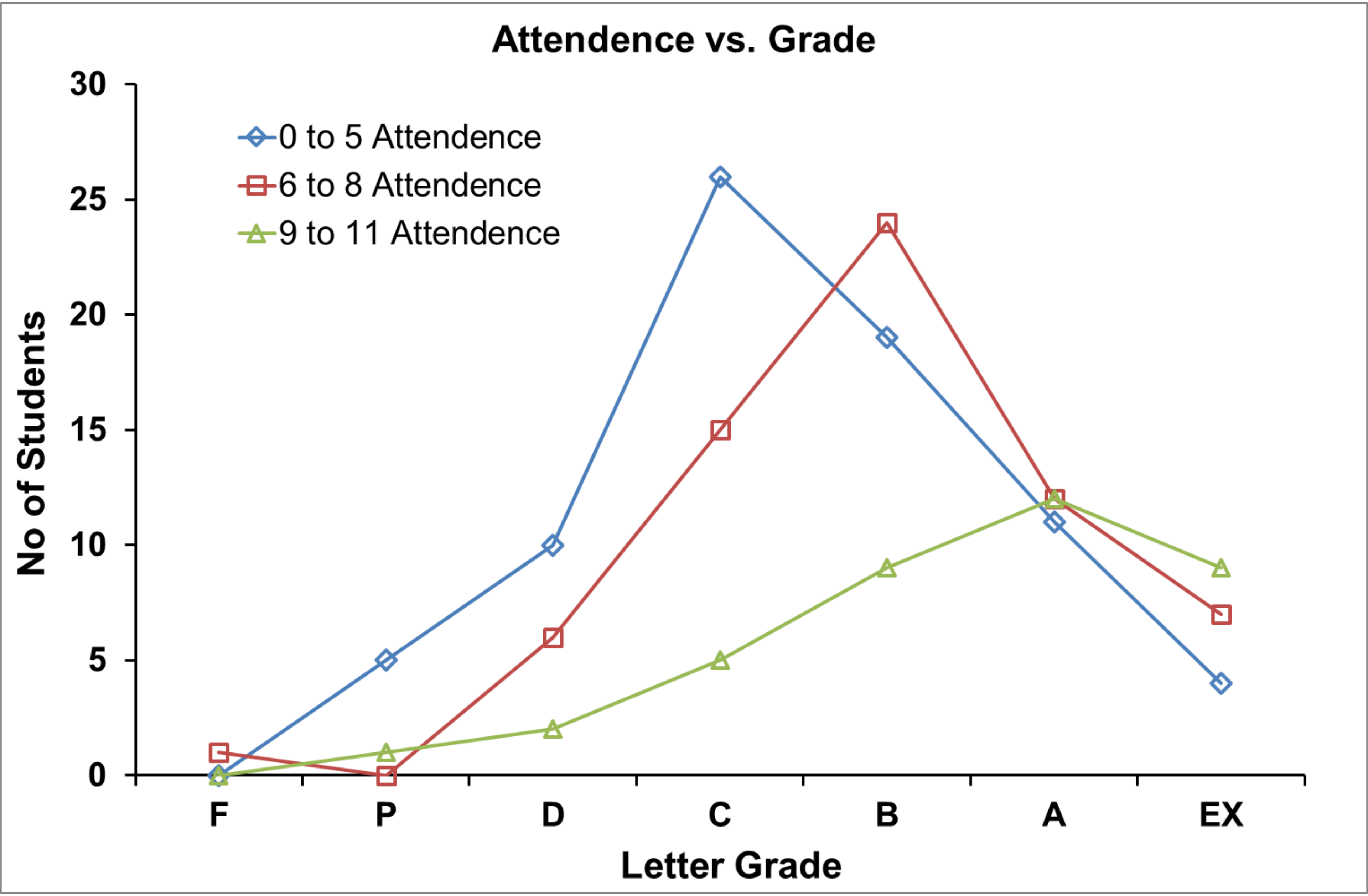
  Biophysics

  Nuclear Magnetic Resonance (NMR) Spectroscopy

  Protein Engineering

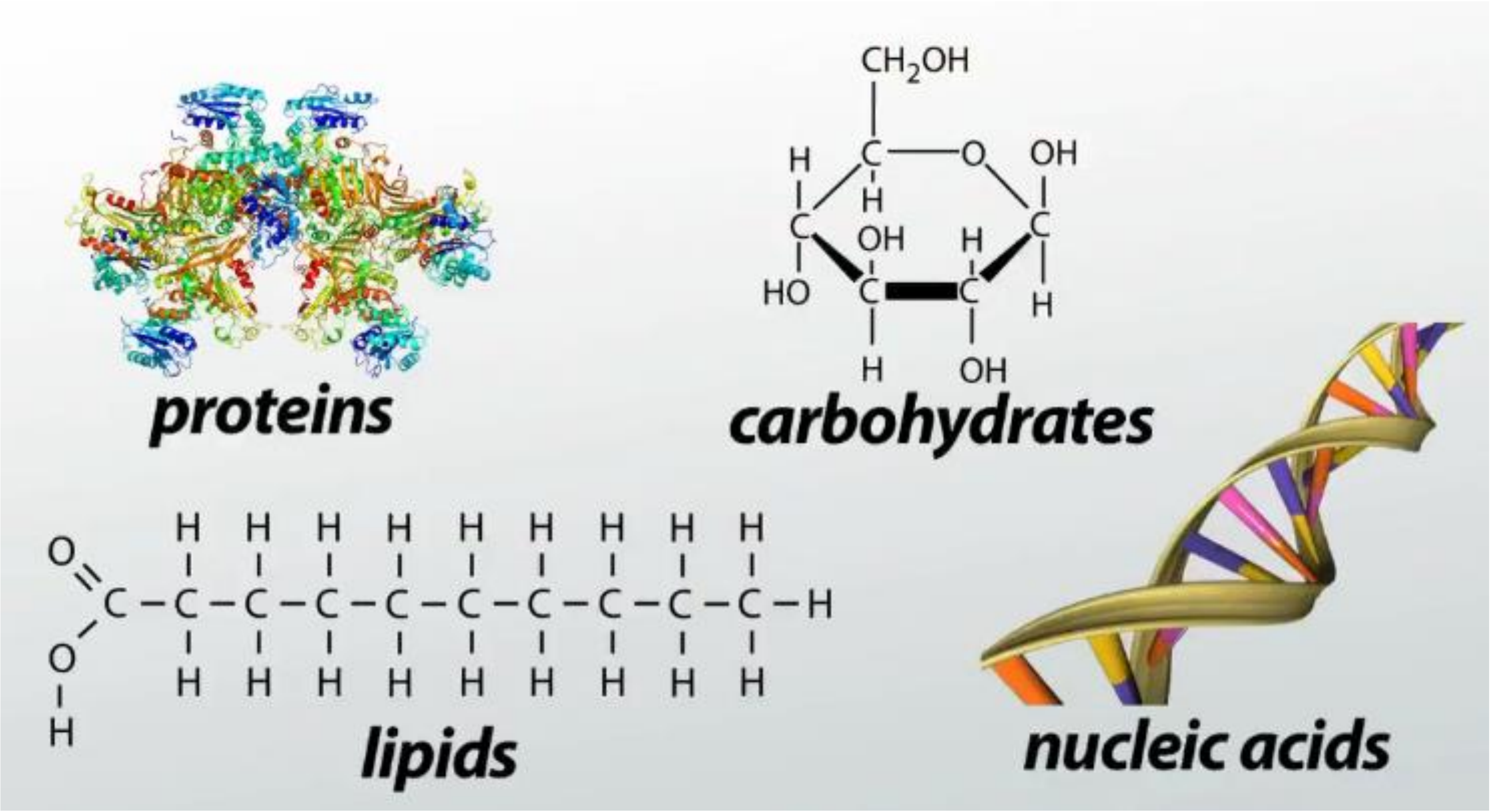
  Signal Transduction and Gene Expression

  Enzymology



Attendence vs. Grade
0 to 5 Attendence
6 to 8 Attendence
9 to 11 Attendence
No of Students
0
5
10
15
20
25
30
F
P
D
C
B
A
EX
Letter Grade

Molecules of Life
Biological Macromolecules



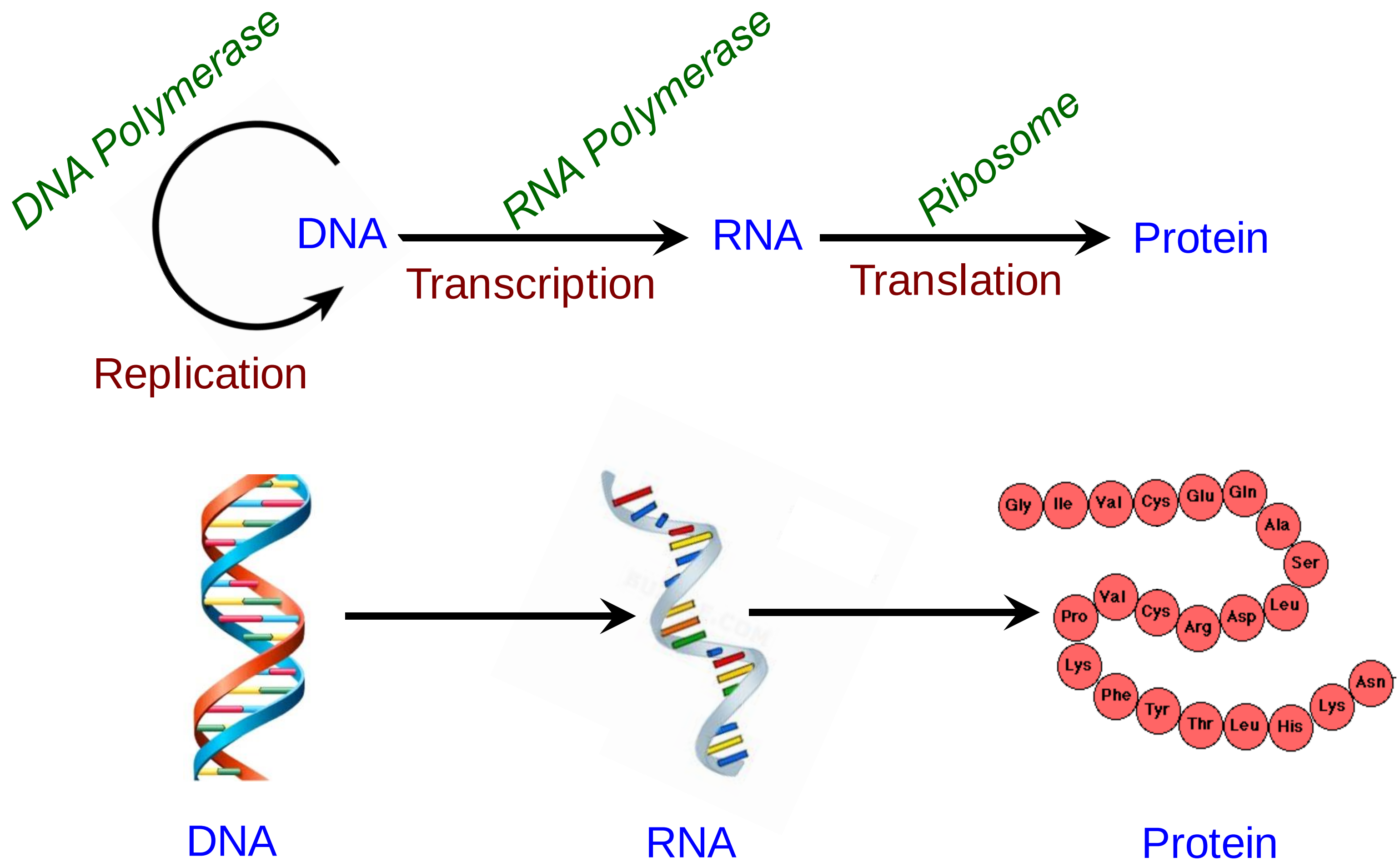

Nucleic Acid

Nucleic Acid

RNA: Ribonucleic Acid

DNA: Deoxyribonucleic Acid

Flow of Genetic Information: The Central Dogma of Molecular Biology

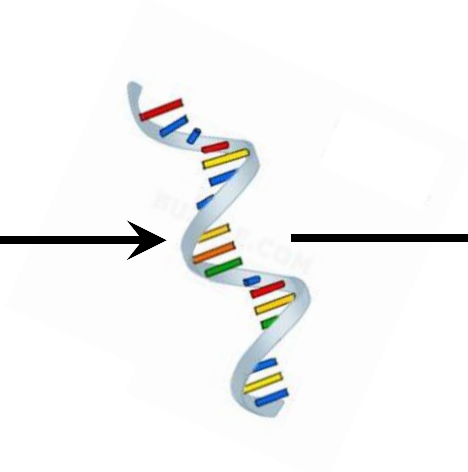


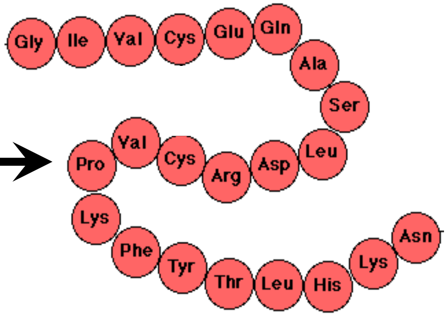
Polymer of nucleotides

Polymer of nucleotides

Polymer of amino acids

DNA is a long thread that stores information





DNA

Polymer of
nucleotides

RNA

Polymer of
nucleotides

Protein

Polymer of
amino acids

Cassette Tape

Why study Living System?

STORAGE LIMITS

Estimates based on bacterial genetics suggest that digital DNA could one day rival or exceed today's storage technology.

	Hard disk	Flash memory	Bacterial DNA
Read–write speed (μs per bit)	~3,000–5,000	~100	<100
Data retention (years)	>10	>10	>100
Power usage (watts per gigabyte)	~0.04	~0.01–0.04	$<10^{-10}$
Data density (bits per cm³)	$\sim10^{13}$	$\sim10^{16}$	$\sim10^{19}$

WEIGHT OF DNA NEEDED TO STORE WORLD'S DATA

455 Exabyte/gm

Microsoft Corporation – DNA Storage Research

A DNA-Based Archival Storage System Bornholt J, et. al. ASPLOS 2016 (International Conference on Architectural Support for Programming Languages and Operating Systems)

Microsoft Store Products Support

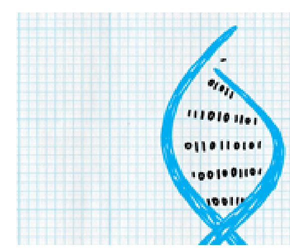
Research | Research areas | Products & Downloads | Programs & Events | People | Careers | About

DNA Storage

Established: January 1, 2015

https://ieeexplore.ieee.org/document/7948677

The amount of digital data produced has long been outpacing the amount of storage available. This project enables molecular-level data storage into DNA molecules by leveraging biotechnology advances in synthesizing, manipulating and sequencing DNA to develop archival storage. Microsoft and University of Washington researchers are collaborating to use DNA as a high density, durable and easy-to-manipulate storage medium.



PCMag India | News

Microsoft Is Using DNA to Solve Our Impending Data Storage Crisis

BY MATTHEW HUMPHRIES MARCH 22, 2019, 7:02 P.M.

The data stored in a warehouse-sized datacenter today would fit into 'a space roughly the siz few board game dice.'

Microsoft device stores digital info as DNA

It translated "HELLO" into DNA and back again -- and it only took 21 hours.

istine Fisher, @cfisherwrites
21.19 in Internet

Comments

1399
Shares

Microsoft and University of Washington demonstrate automated DNA data storage

BY ALAN BOYLE on March 21, 2019 at 6:00 am

MUST READ: Twitter bots and trolls promote conspiracy theories about Australian bushfires

Microsoft's DNA storage breakthrough could pave way for exabyte drives

Scientists develop a more efficient way to find and selectively retrieve files stored on DNA.

DNA Storage is robust and has high storage density

Tweet

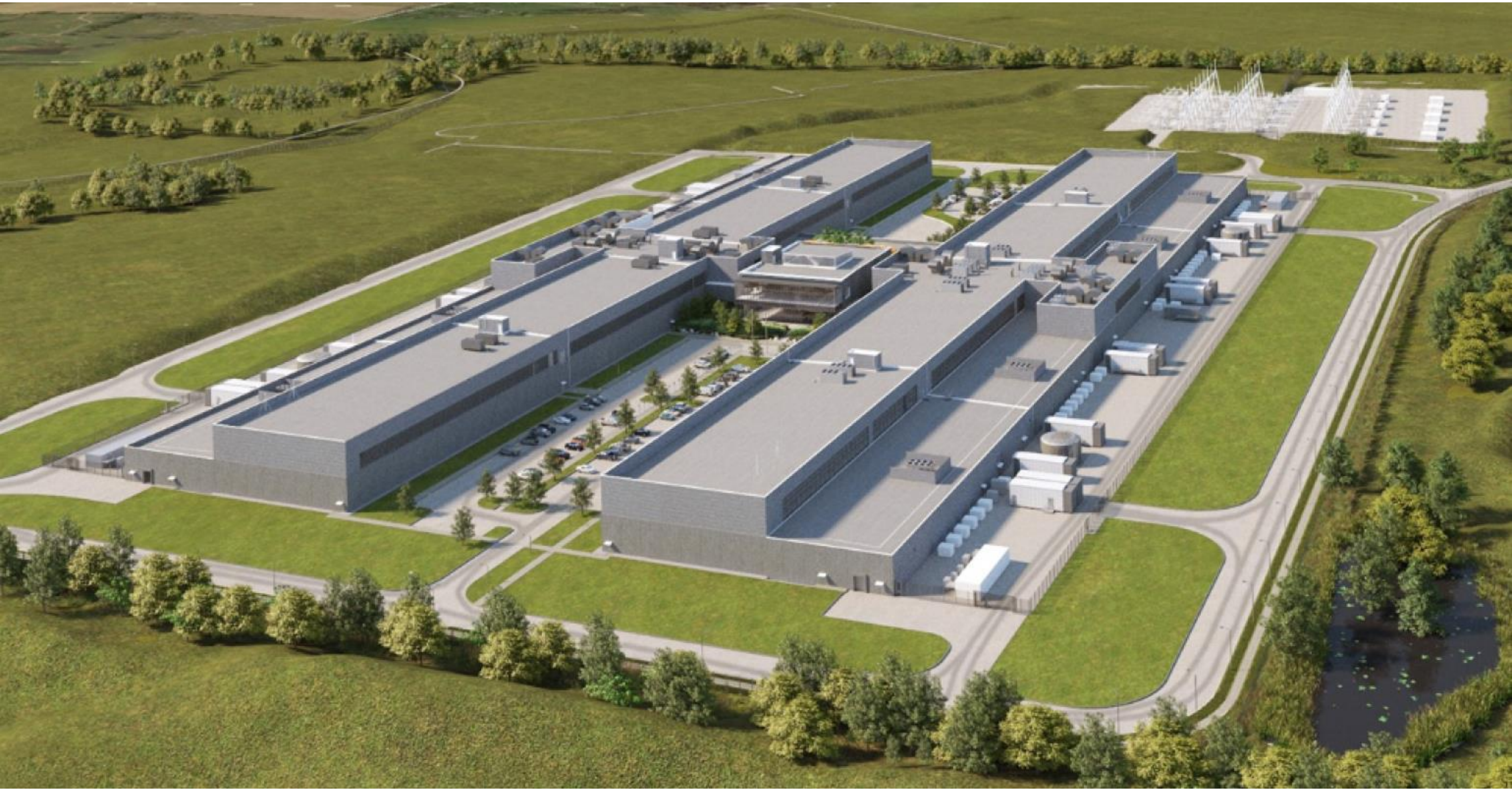
ETH Zurich
@ETH_en

The ETH spin-off Turbobeads has stored the @NetflixDE series #Biohackers in #DNA. In the Making of, the researchers explain how the digital record was translated into a sequence of the four DNA building blocks adenine, guanine, cytosine and thymine.

youtube.com
Biohackers | First Original Series stored in DNA | Netflix
Biohackers is a gripping science thriller about biotechnologies. We have turned the series itself into a ...

2:43 PM · Aug 17, 2020 · Falcon Social Media Management

DNA Storage has high storage density



$800 million data center in Gallatin, a small town just outside of Nashville, Tennessee. Facebook acquired the 809 acres of land for this data center in 2020. One cricket ground is ~ 4 acres. Storage capacity is several Exabyte of data. 1 gm of DNA can store 455 Exabyte of data.

DNA Storage is robust

https://en.wikipedia.org/wiki/Otzi

https://www.bbc.com/news/science-environment-58191123

- Otzi the mummy was discovered in 1991 in the Ötztal Alps on the border between Austria and Italy.
- He lived from 3275 BC to 3230 BC i.e. 5000 years ago.
- Otzi's full genome has been sequenced and published on February 2012.
- DNA analysis also showed him at high risk of atherosclerosis and lactose intolerance.
- DNA of over 3,700 Tyrolean male blood donors revealed that Otzi has living relatives in Austria today.
- Oldest DNA sequenced: DNA extracted from the tooth of 1.2 million year old mammoth. (Feb 2021) https://www.nature.com/articles/d41586-021-00436-x
- In 2016, researchers reported protein sequences from 3.8-million-year-old ostrich eggshells from Tanzania.

Nucleic Acid

- **Nucleic acids are polymers**

- **Monomer---nucleotides**
  - **Nitrogenous bases**
    - **Purines**
    - **Pyrimidines**
  - **Sugar**
    - **Ribose**
    - **Deoxyribose**
  - **Phosphates**

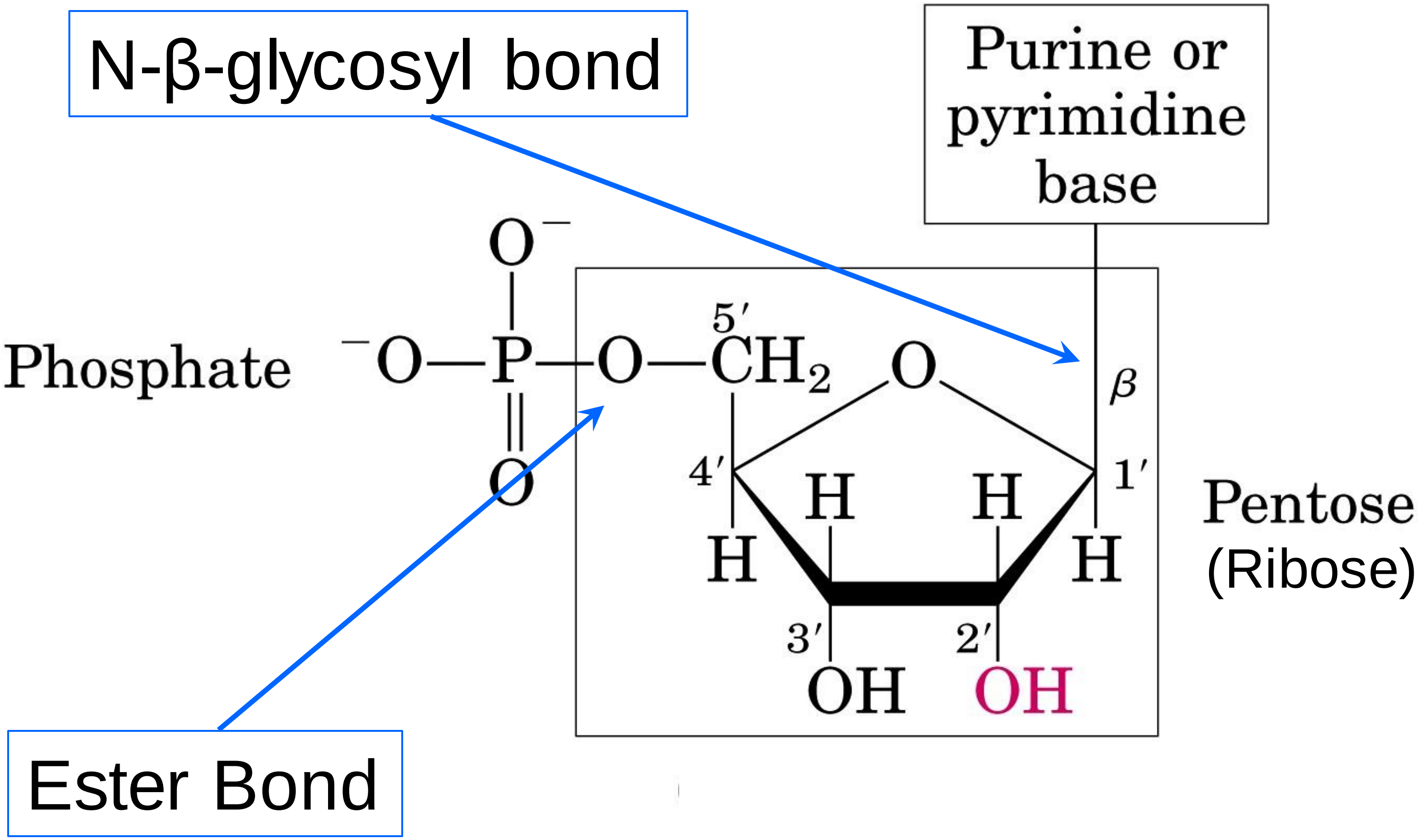
Nucleosides

Nucleotides

Nucleic Acid monomer structure

RNA - Ribonucleic Acid **(OH)**

DNA - Deoxyribonucleic Acid **(H)**



Two Types of Nitrogenous Bases

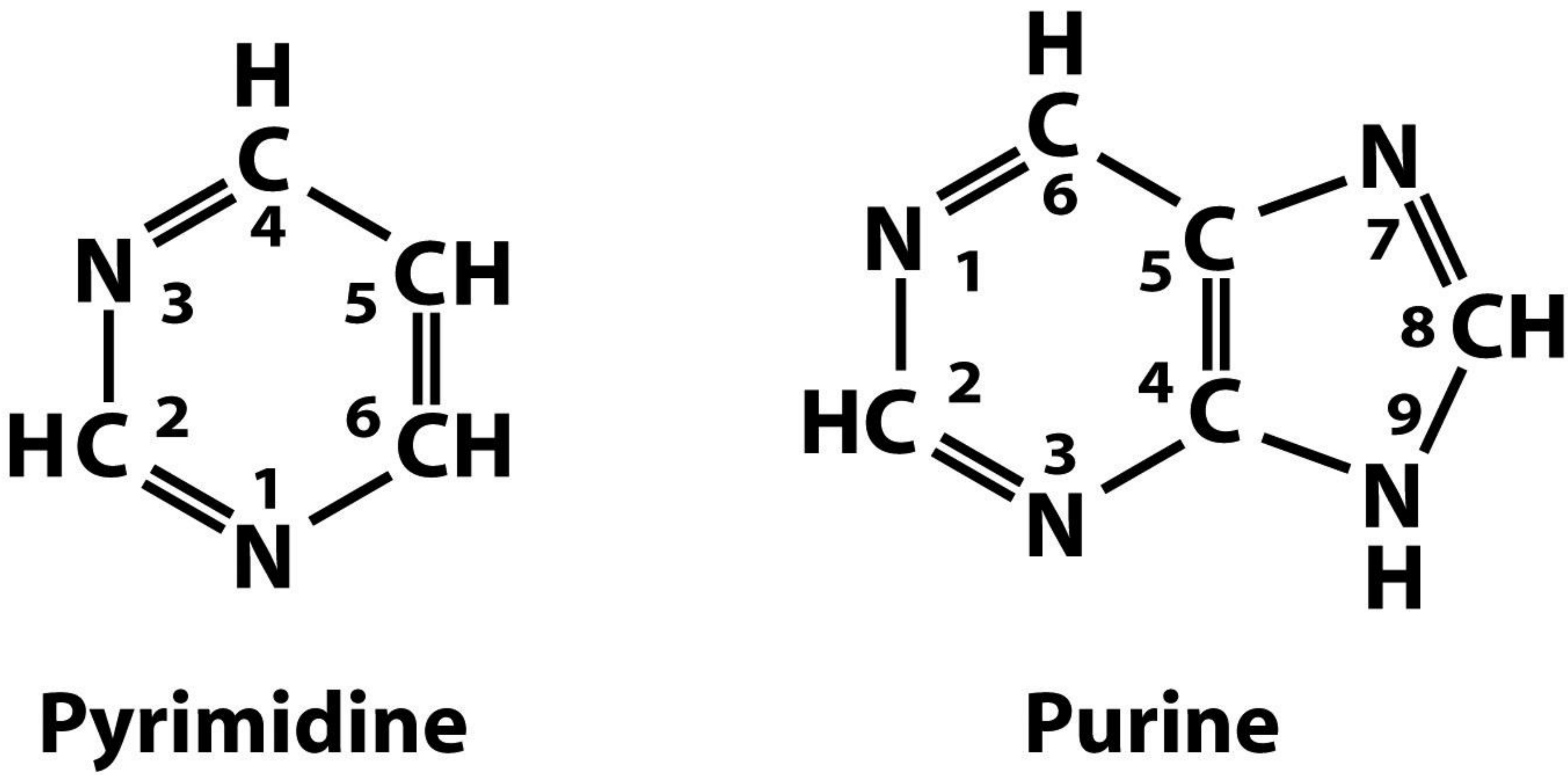
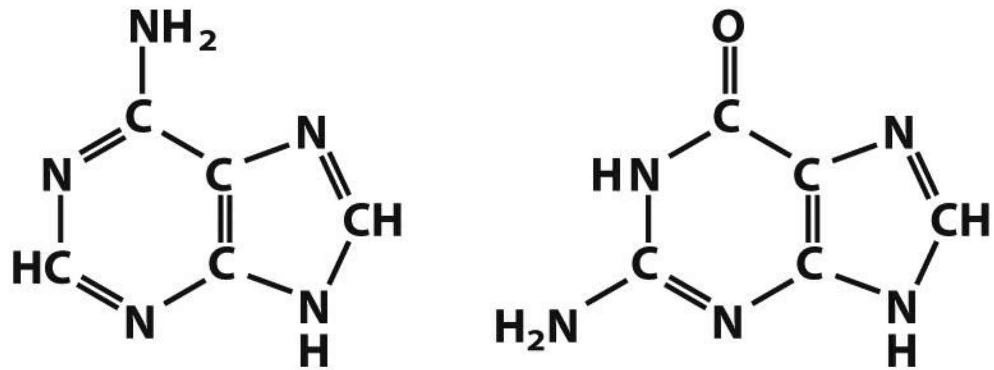


Figure 8-1b
Lehninger Principles of Biochemistry, Fifth Edition

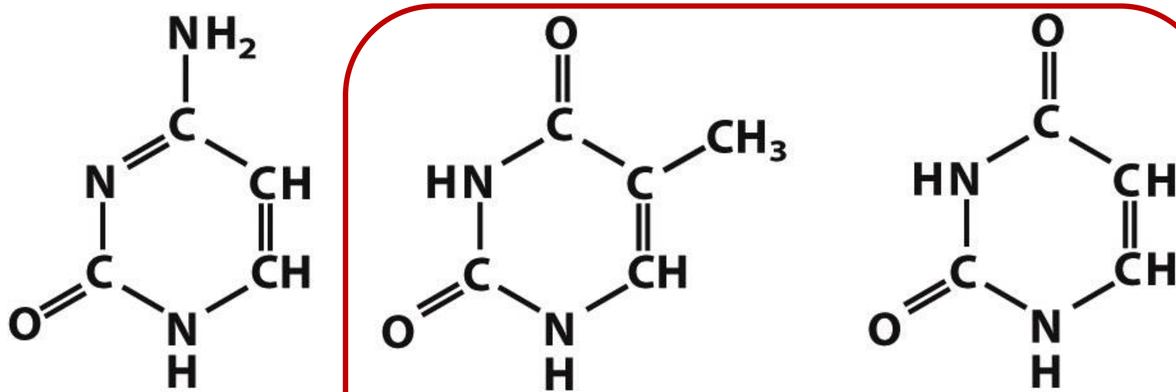
Purine and Pyrimidine Bases



Adenine

Guanine

Purines



Cytosine

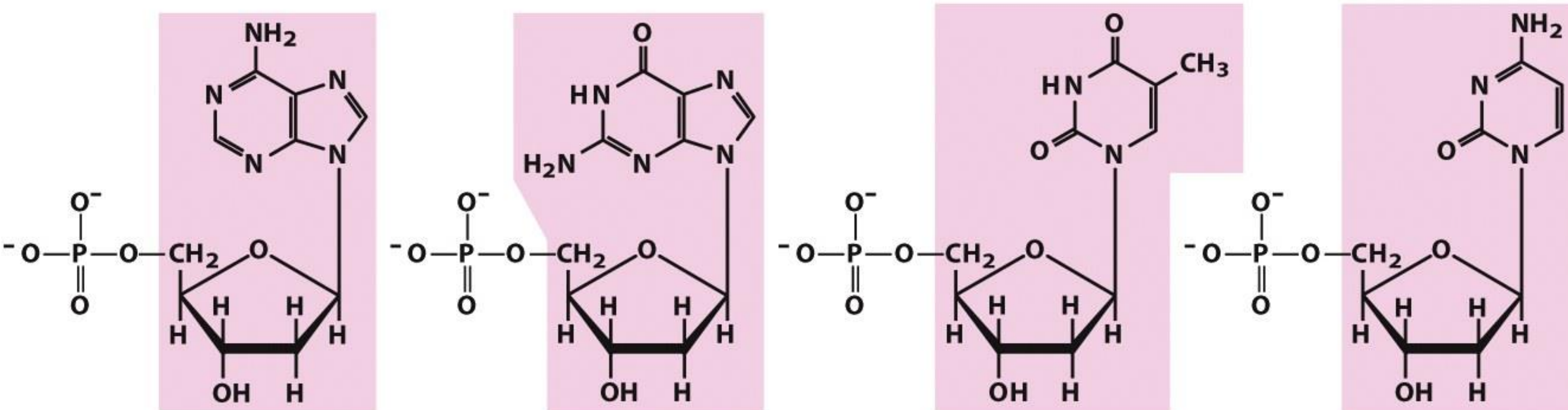
Thymine (DNA)

Uracil (RNA)

Pyrimidines

Nucleotide = Nucleoside + Phosphate

(Nucleoside = Sugar + Base)



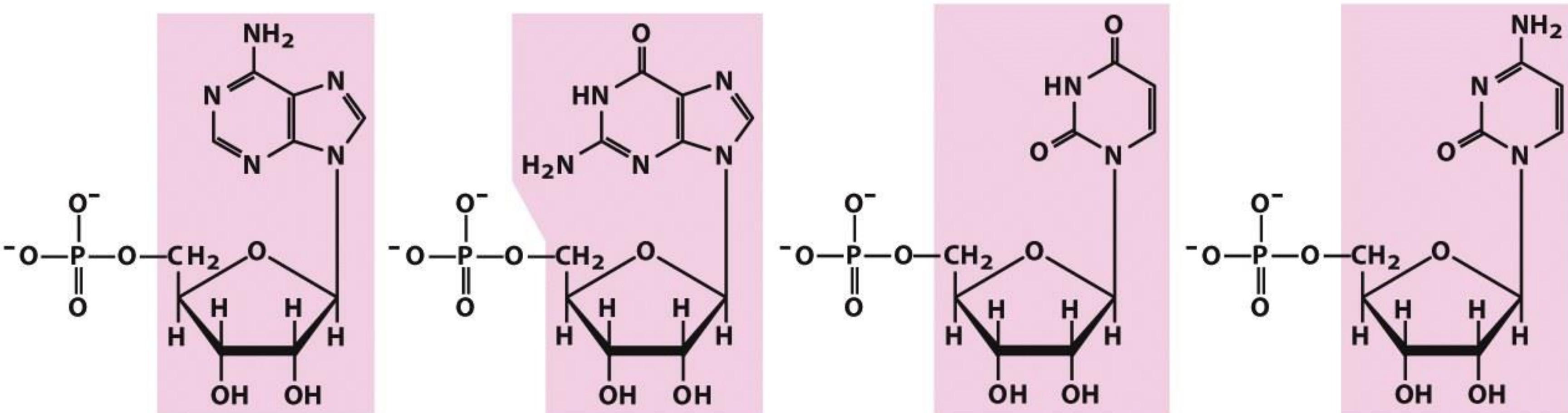
Nucleotide:	Deoxyadenylate (deoxyadenosine 5′-monophosphate)	Deoxyguanylate (deoxyguanosine 5′-monophosphate)	Deoxythymidylate (deoxythymidine 5′-monophosphate)	Deoxycytidylate (deoxycytidine 5′-monophosphate)
Symbols:	A, dA, dAMP	G, dG, dGMP	T, dT, dTMP	C, dC, dCMP
Nucleoside:	Deoxyadenosine	Deoxyguanosine	Deoxythymidine	Deoxycytidine

Deoxyribonucleotides

Figure 8-4a
Lehninger Principles of Biochemistry, Fifth Edition

Nucleotide = Nucleoside + Phosphate

(Nucleoside = Sugar + Base)



Nucleotide:	Adenylate (adenosine 5′-monophosphate)	Guanylate (guanosine 5′-monophosphate)	Uridylate (uridine 5′-monophosphate)	Cytidylate (cytidine 5′-monophosphate)
Symbols:	A, AMP	G, GMP	U, UMP	C, CMP
Nucleoside:	Adenosine	Guanosine	Uridine	Cytidine

Ribonucleotides

Figure 8-4b
Lehninger Principles of Biochemistry, Fifth Edition

Phosphodiester Linkages in the Covalent Backbone of Nucleic Acid

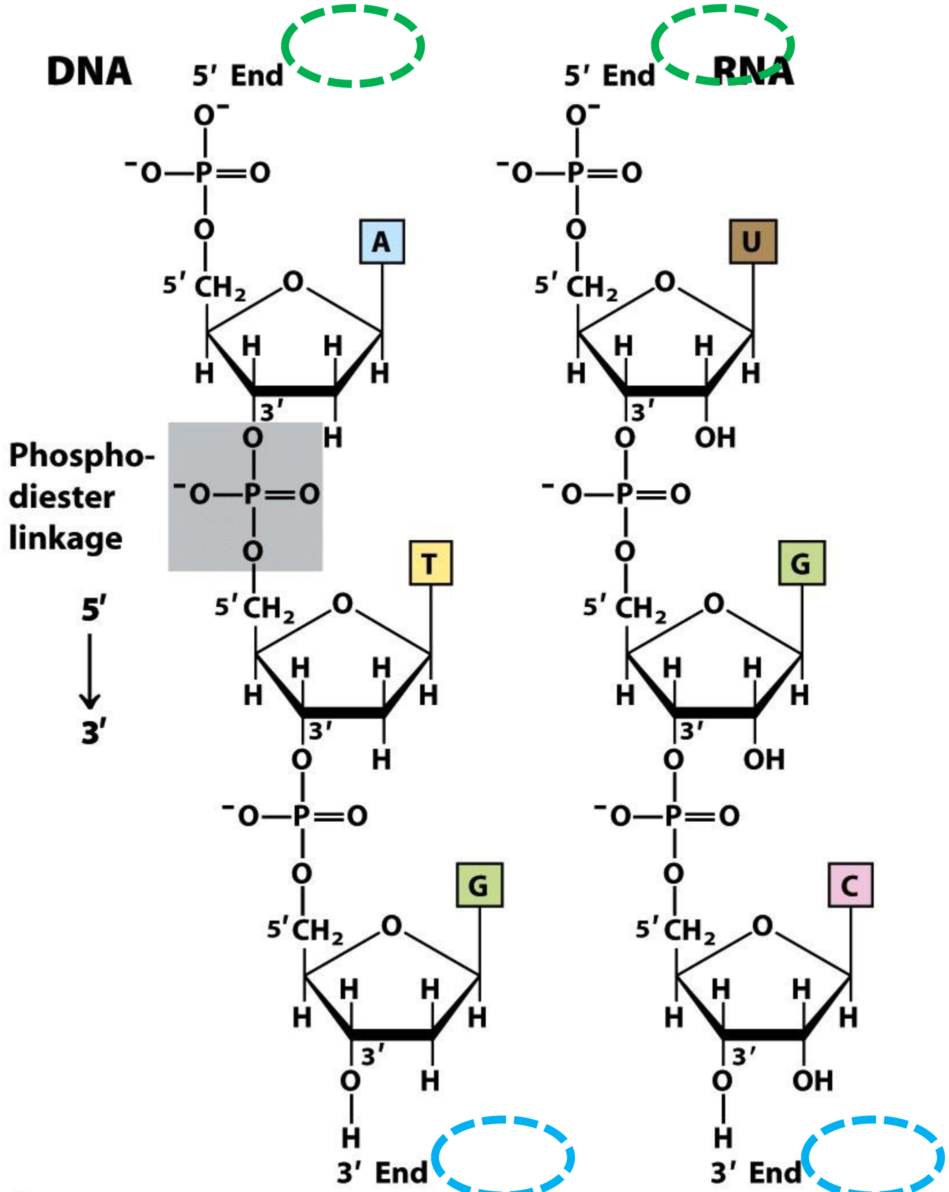
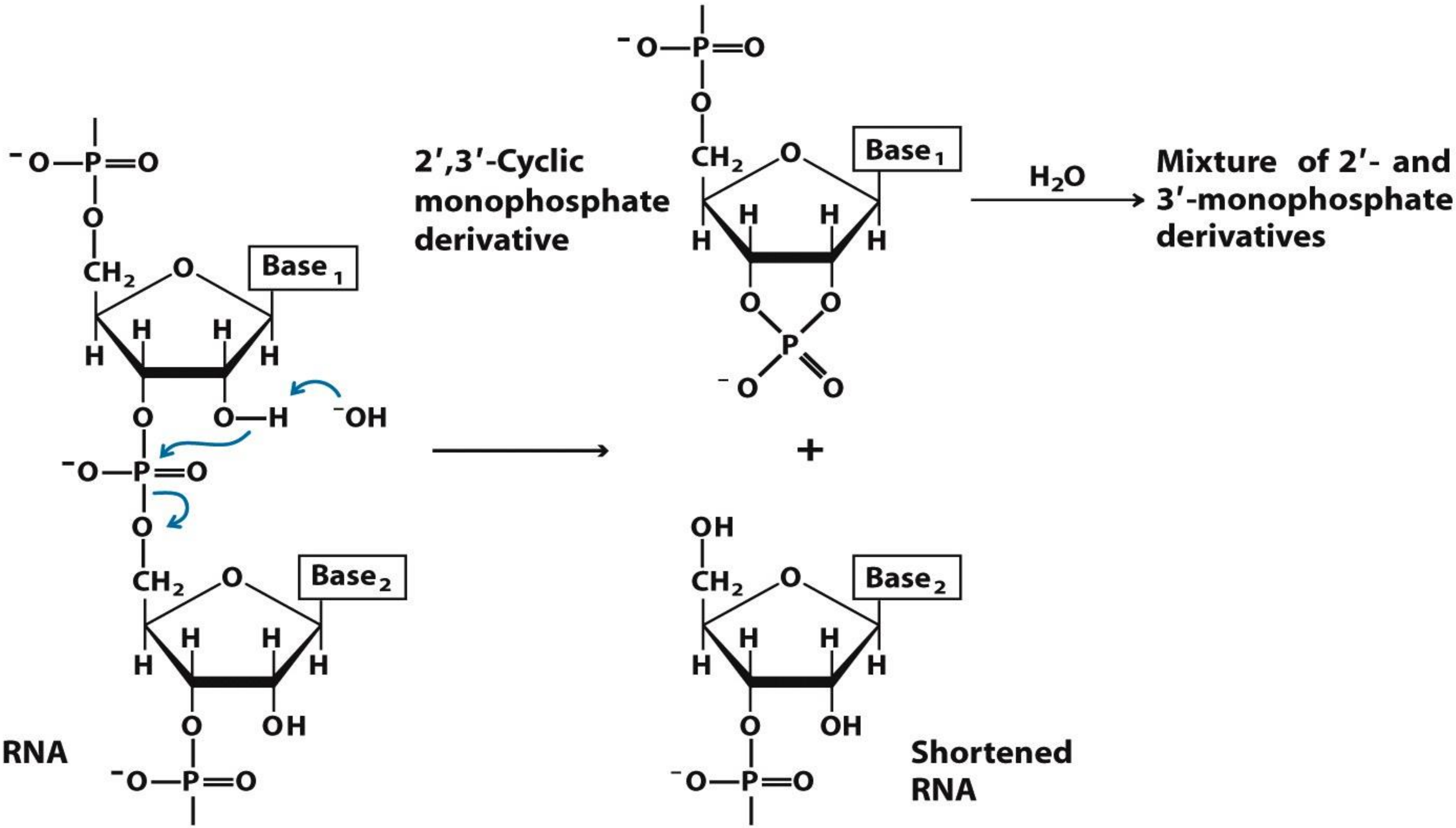


Figure 8-7
Lehninger Principles of Biochemistry, Fifth Edition

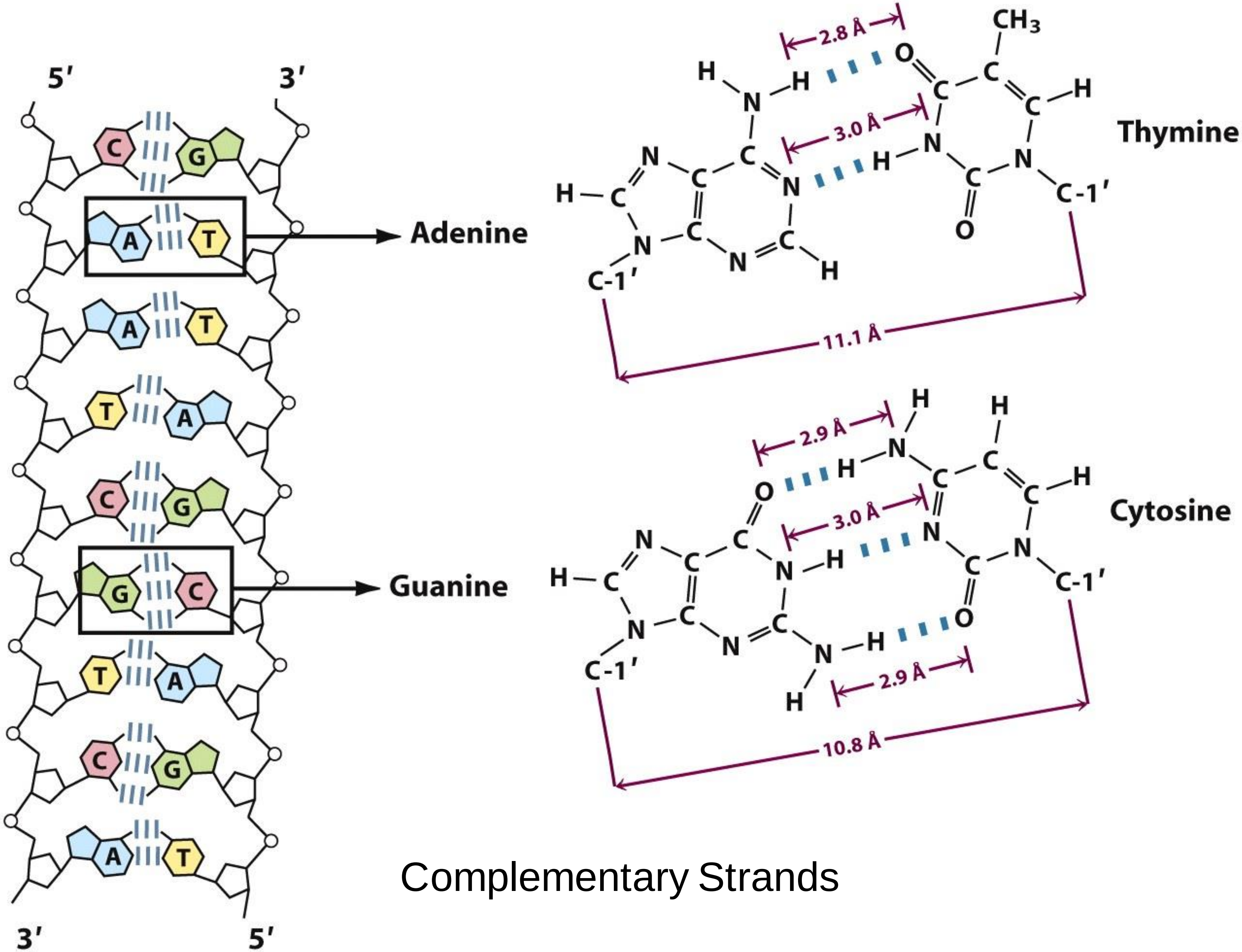
How information is stored in DNA or RNA?

5′-**ATG**-3′	5′-**ATC**-3′	5′-**ATA**-3′	5′-**ATT**-3′
5′-**TTG**-3′	5′-**TTC**-3′	5′-**TTA**-3′	5′-**TTT**-3′
5′-**GTG**-3′	5′-**GTC**-3′	5′-**GTA**-3′	5′-**GTT**-3′
5′-**CTG**-3′	5′-**CTC**-3′	5′-**CTA**-3′	5′-**CTT**-3′
5′-**AGG**-3′	5′-**AGC**-3′	5′-**AGA**-3′	5′-**AGT**-3′
5′-**TGG**-3′	5′-**TGC**-3′	5′-**TGA**-3′	5′-**TGT**-3′
5′-**GGG**-3′	5′-**GGC**-3′	5′-**GGA**-3′	5′-**GGT**-3′
5′-**CGG**-3′	5′-**CGC**-3′	5′-**CGA**-3′	5′-**CGT**-3′
5′-**ACG**-3′	5′-**ACC**-3′	5′-**ACA**-3′	5′-**ACT**-3′
5′-**TCG**-3′	5′-**TCC**-3′	5′-**TCA**-3′	5′-**TCT**-3′
5′-**GCG**-3′	5′-**GCC**-3′	5′-**GCA**-3′	5′-**GCT**-3′
5′-**CCG**-3′	5′-**CCC**-3′	5′-**CCA**-3′	5′-**CCT**-3′
5′-**AAG**-3′	5′-**AAC**-3′	5′-**AAA**-3′	5′-**AAT**-3′
5′-**TAG**-3′	5′-**TAC**-3′	5′-**TAA**-3′	5′-**TAT**-3′
5′-**GAG**-3′	5′-**GAC**-3′	5′-**GAA**-3′	5′-**GAT**-3′
5′-**CAG**-3′	5′-**CAC**-3′	5′-**CAA**-3′	5′-**CAT**-3′

RNA is Less Stable than DNA



DNA: Deoxyribonucleic Acid

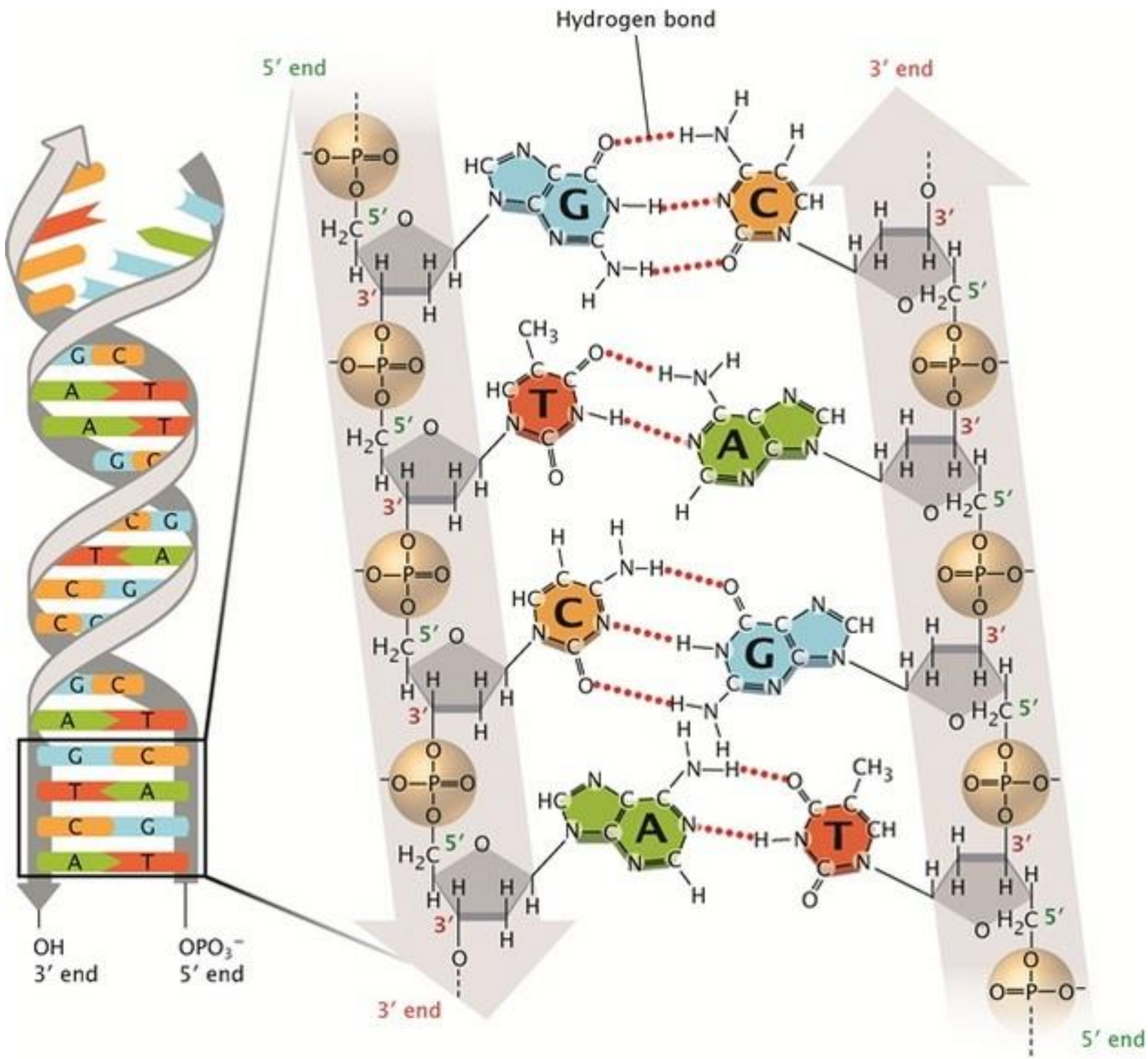


Complementary Strands

Some key features of DNA

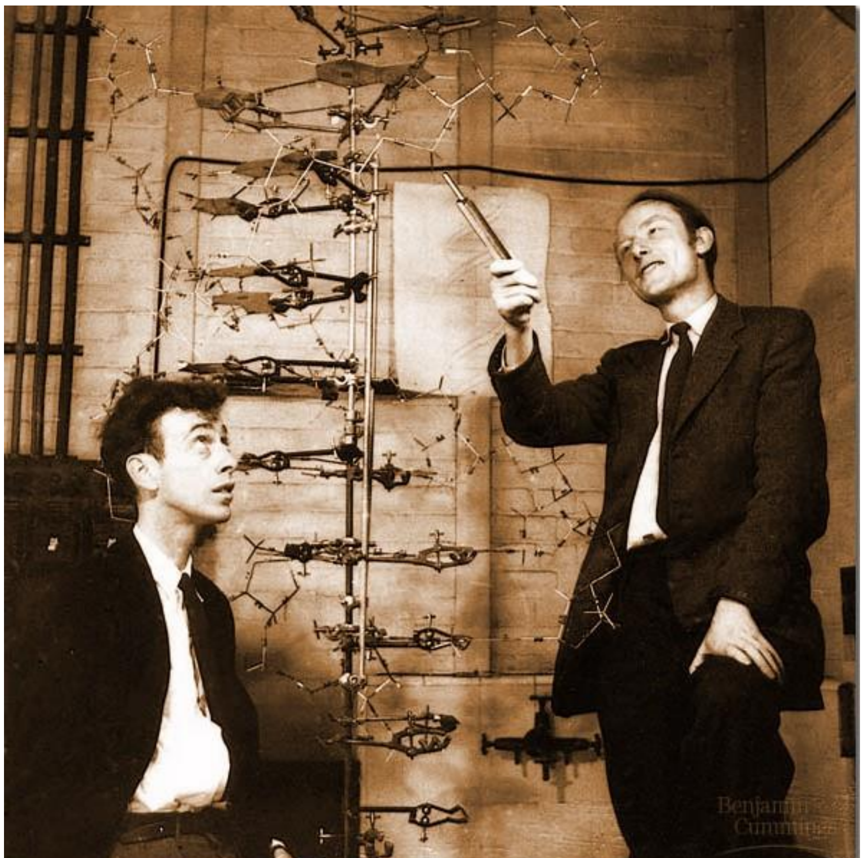
- In **DNA,** two nucleic acid strands anneal together through extensive **inter-strand H-bonding** between the bases**.** This **base pairing** follows the rule proposed by Watson and Crick.
- **Chargaff's rule: A always pairs with T and G pairs with C**
- Hence the two strands become **complementary** to each other
- **Directionality** of two strands is **opposite:** one is **5'-3'** and another **is 3'-5'**
- Hence complementary DNA strands are **antiparallel**

DNA: Deoxyribonucleic Acid



Discovery of the DNA Structure

- Structure was discovered in 1953 by James Watson and Francis Crick
- Awarded **Nobel Prize in 1962**



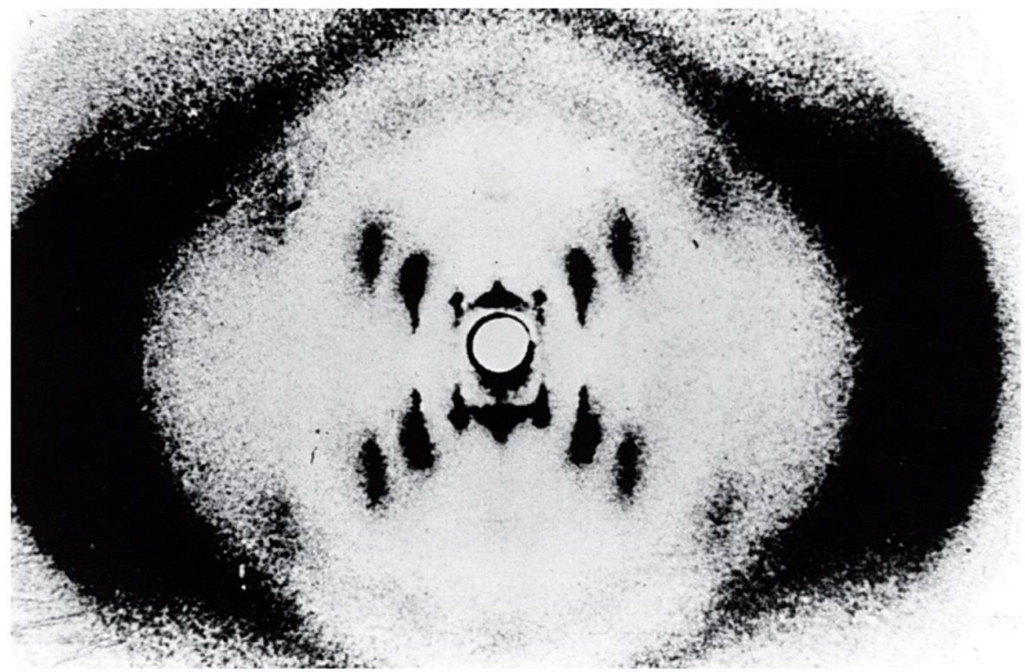
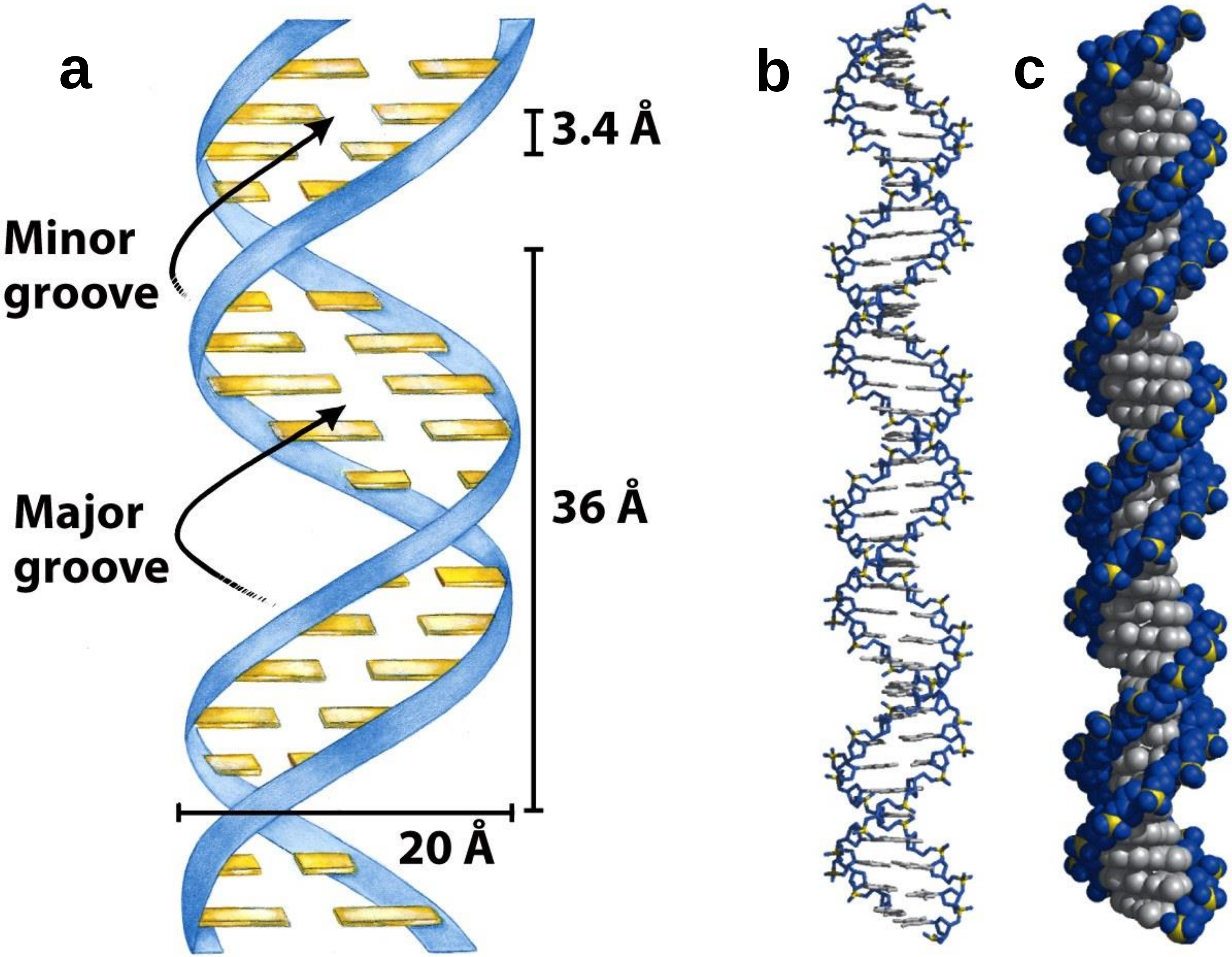


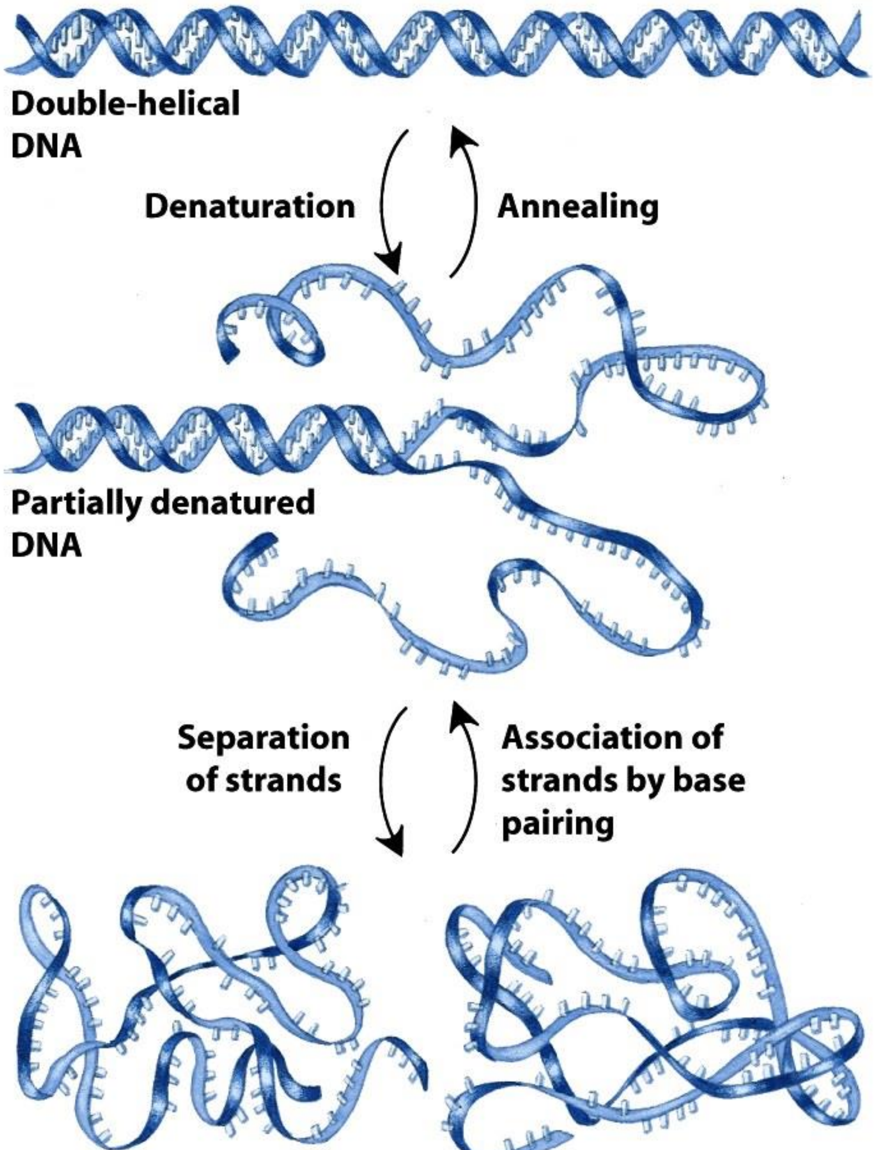
Figure 8-12
Lehninger Principles of Biochemistry, Fifth Edition

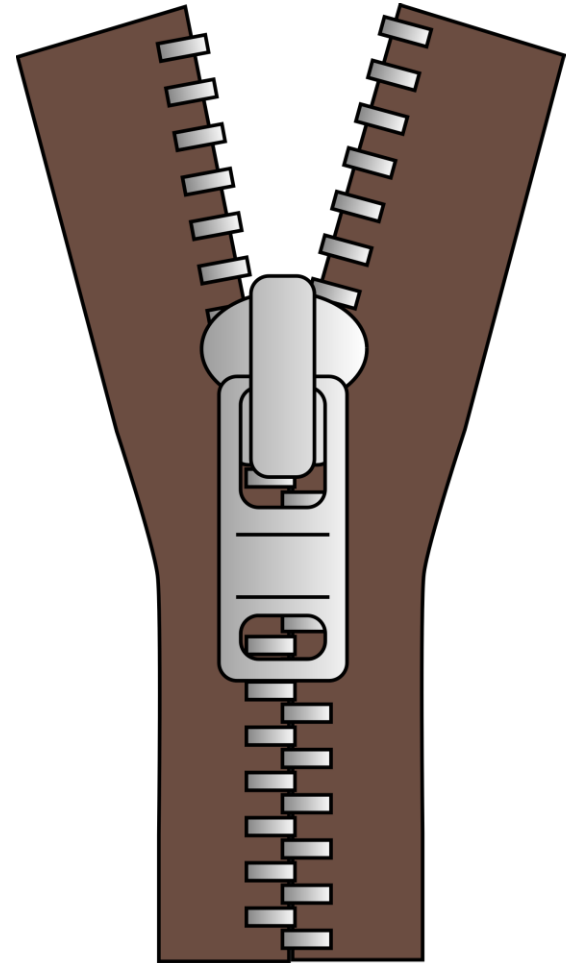
Rosalind Franklin

Watson-Crick Model for the Structure of DNA

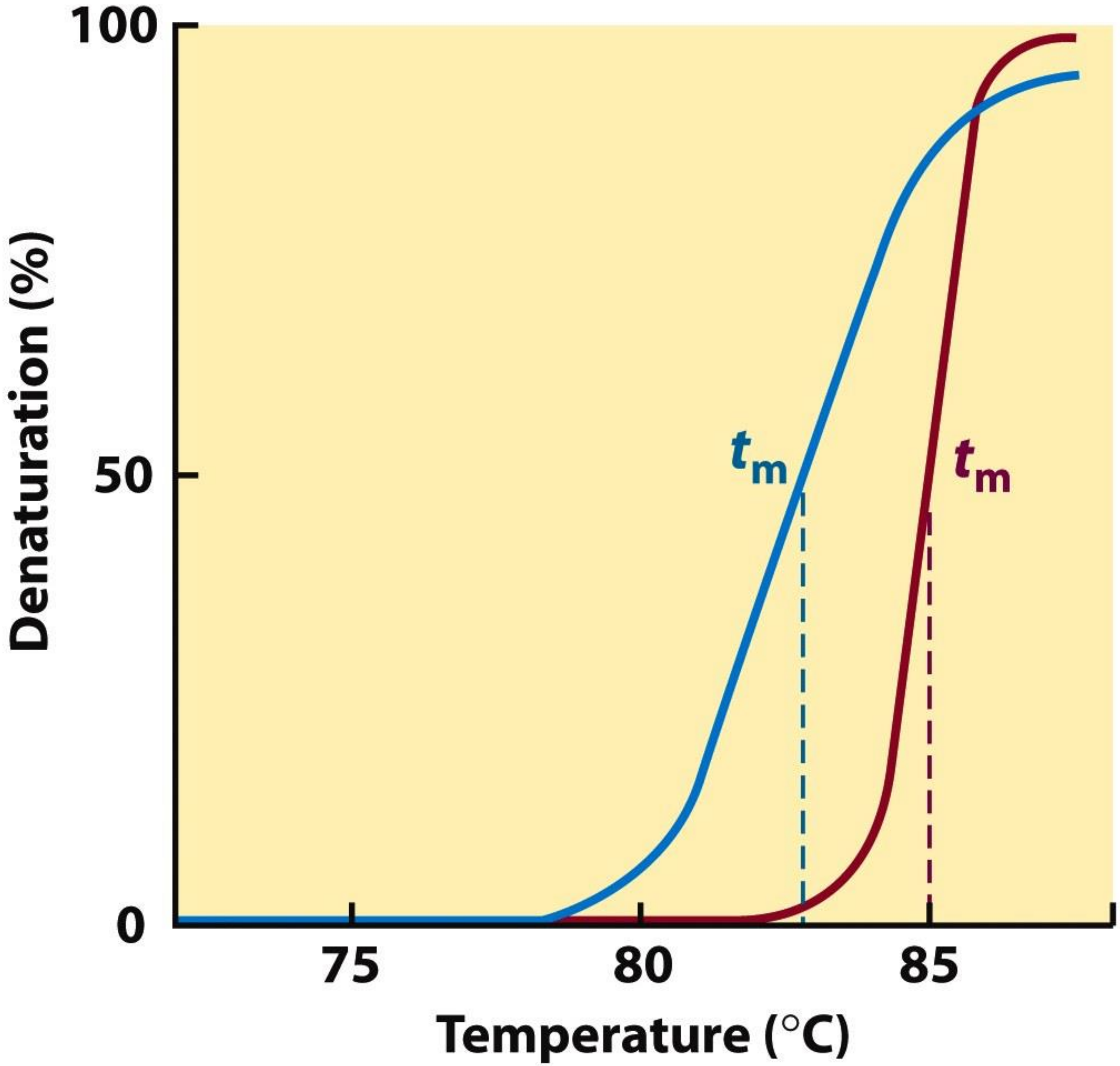


Reversible Denaturation and Annealing (Renaturation) of DNA

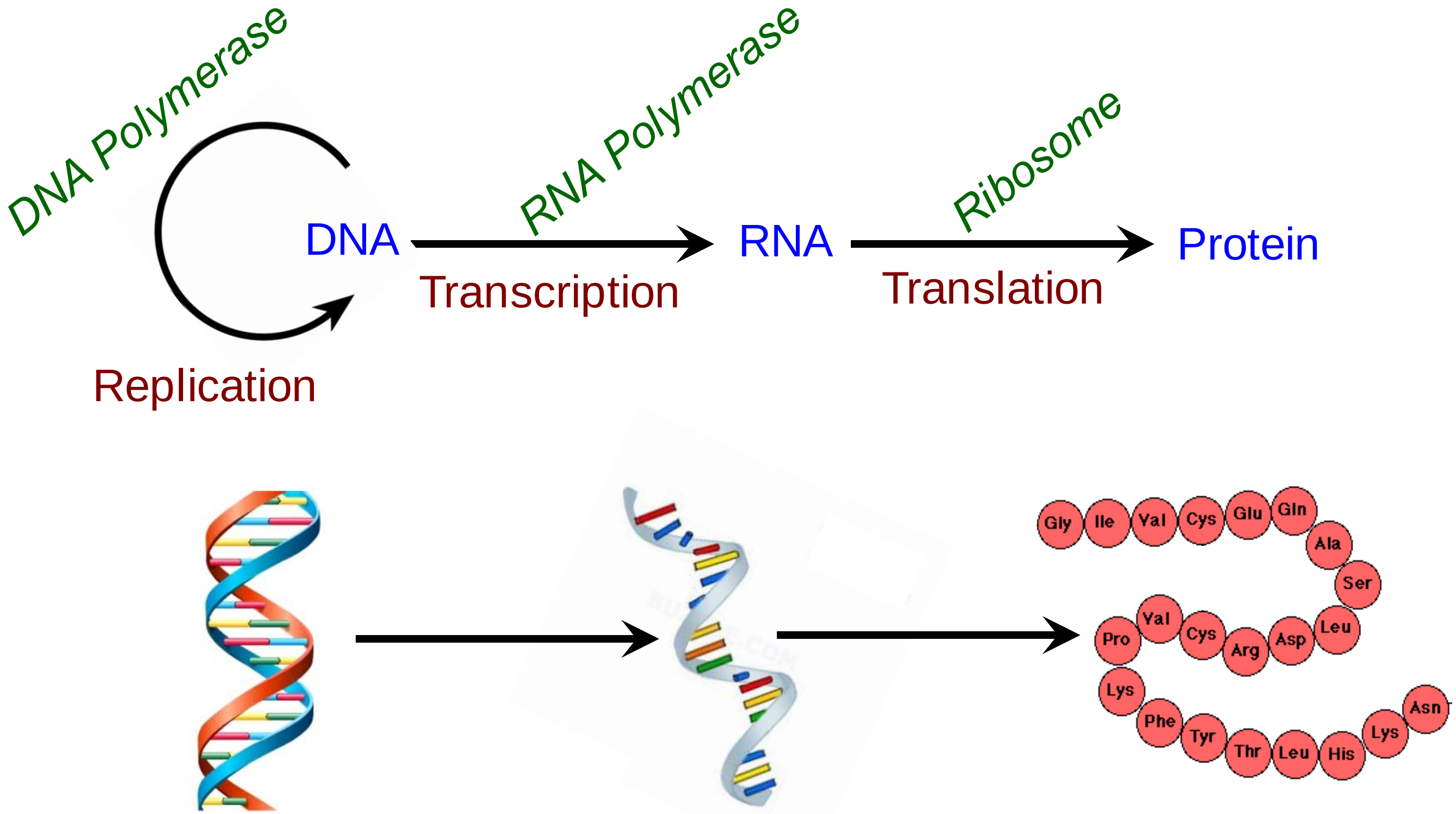




Heat Denaturation of DNA

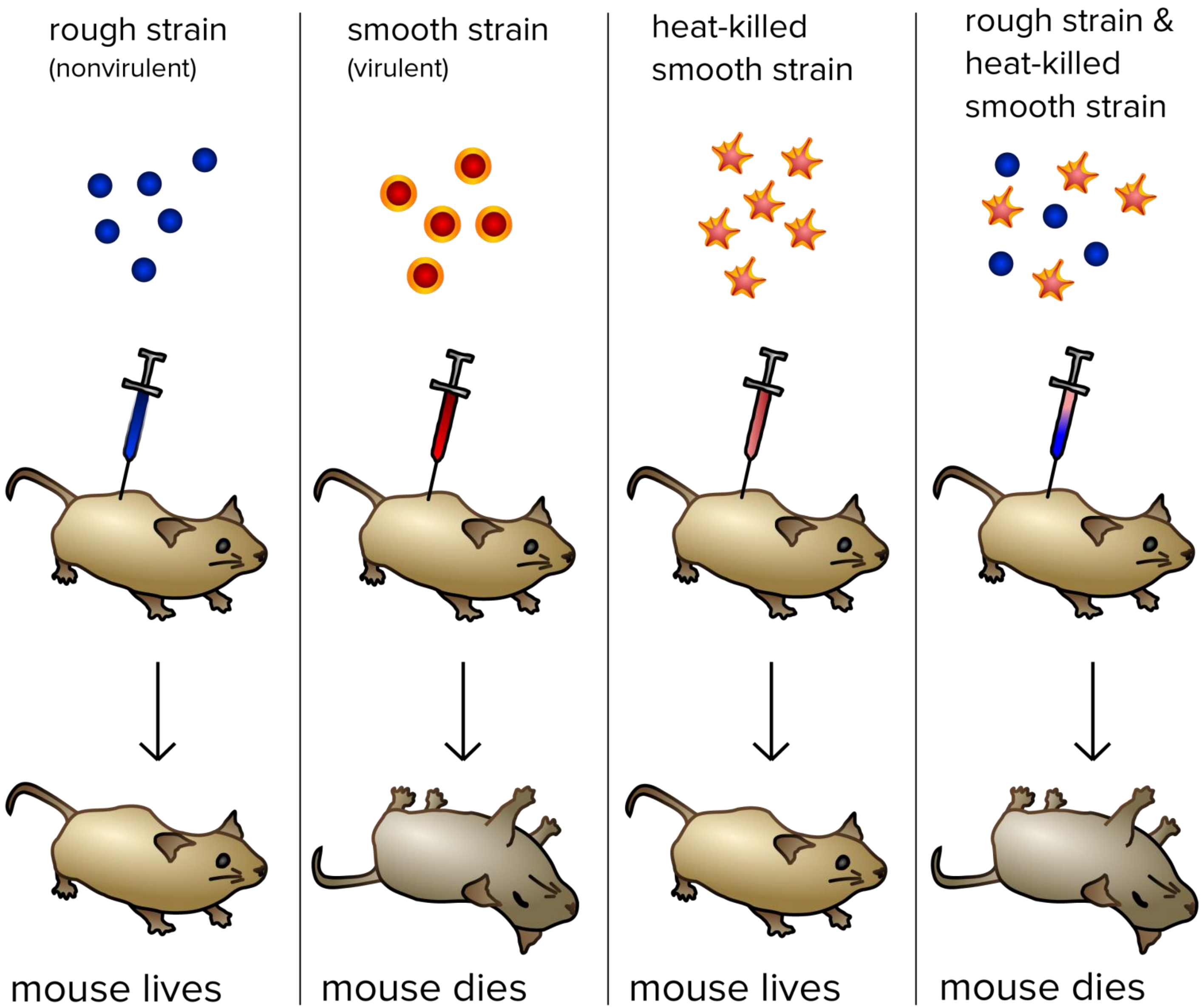


Flow of Genetic Information:
The Central Dogma of Molecular Biology



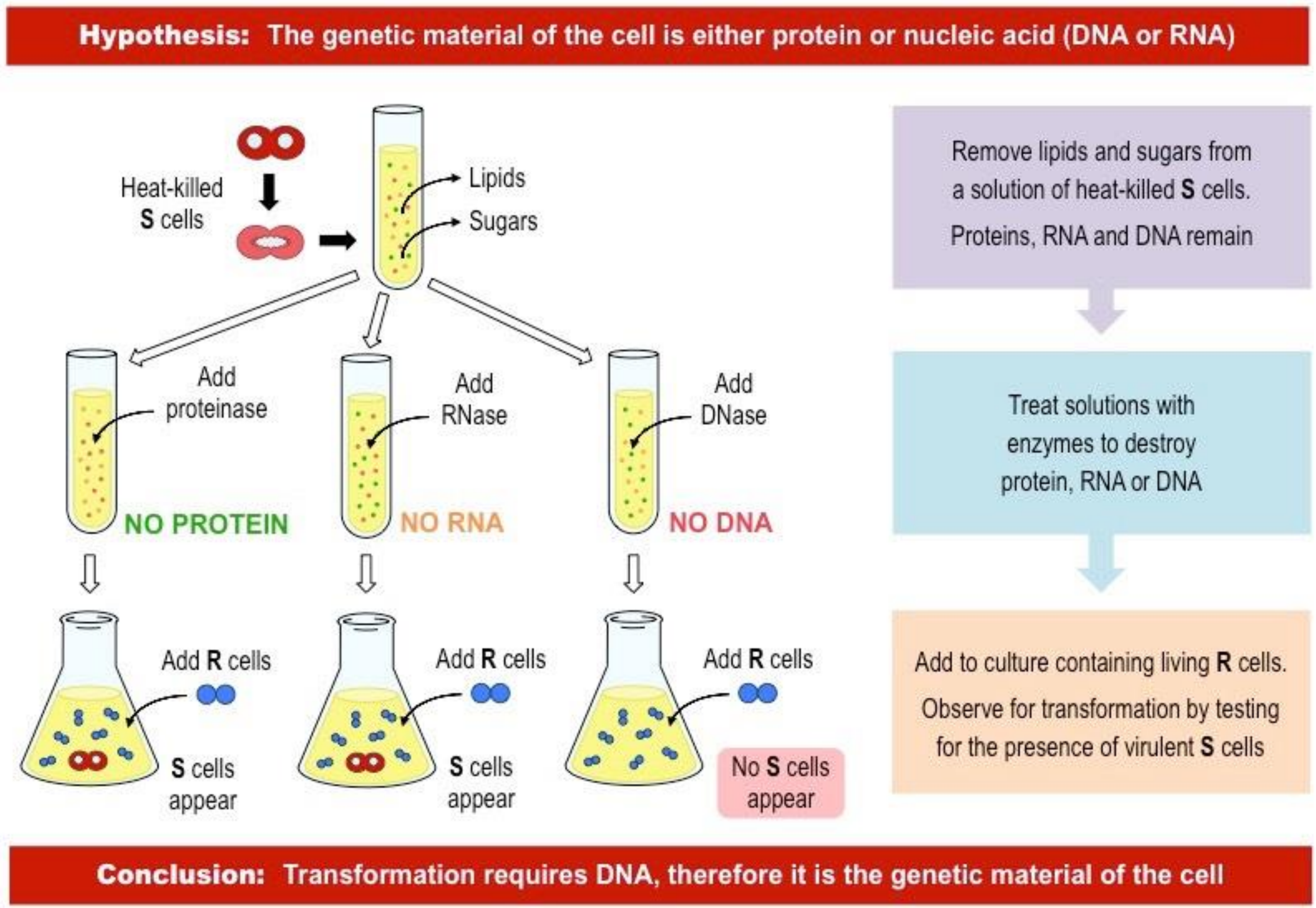
DNA	RNA	Protein
Polymer of nucleotides	Polymer of nucleotides	Polymer of amino acids

Experiment that Proves DNA is Our Genetic Material

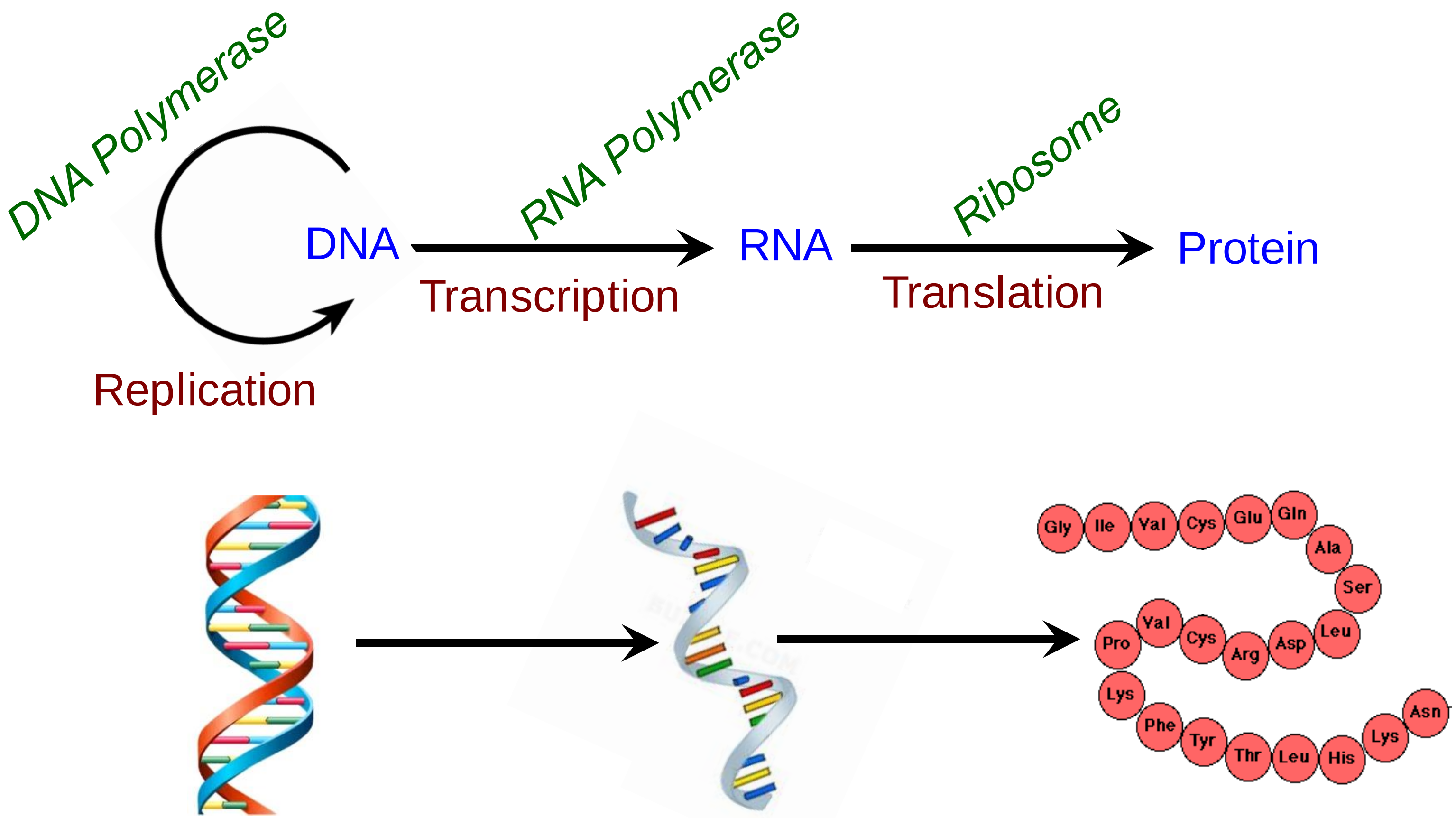


Frederick Griffith's experiment with bacteria (*Streptococcus pneumoniae*) (1928).

Experiment that Proves DNA is Our Genetic Material



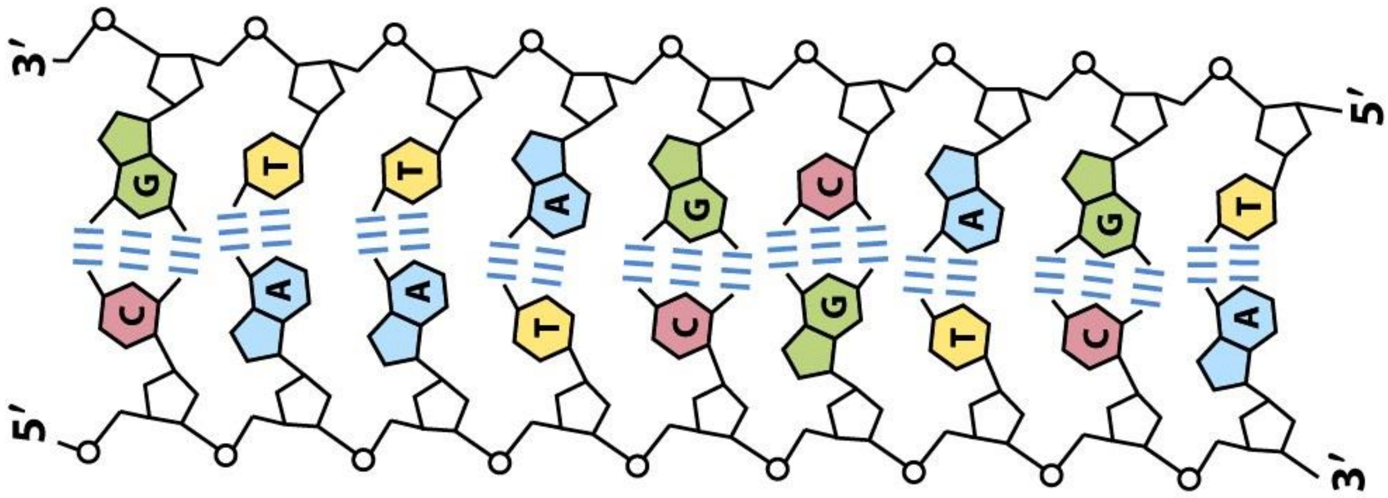
Flow of Genetic Information:
The Central Dogma of Molecular Biology

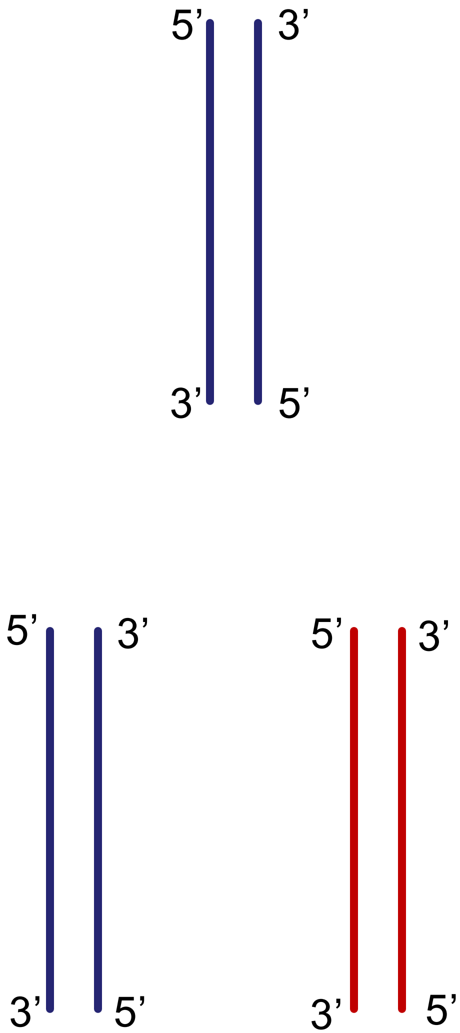


DNA	RNA	Protein
Polymer of nucleotides	Polymer of nucleotides	Polymer of amino acids

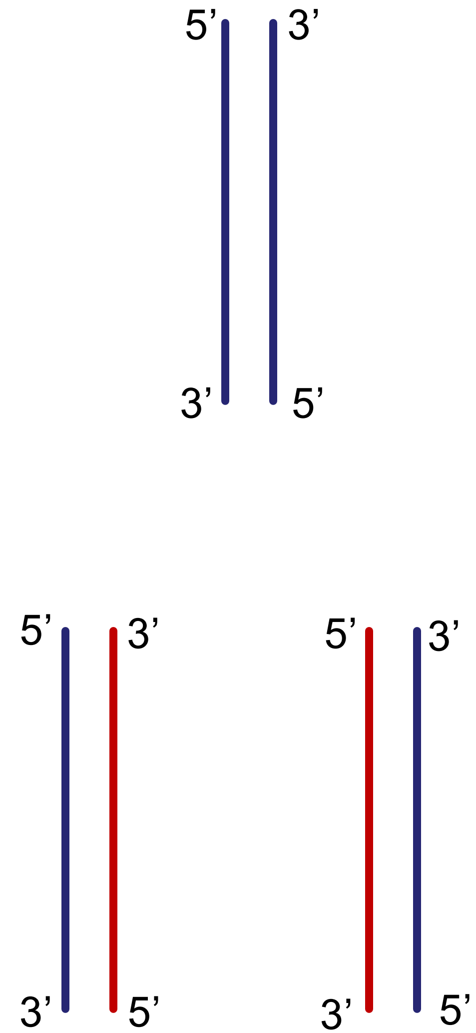
DNA Replication: An Overview

(Copying of the genetic information)



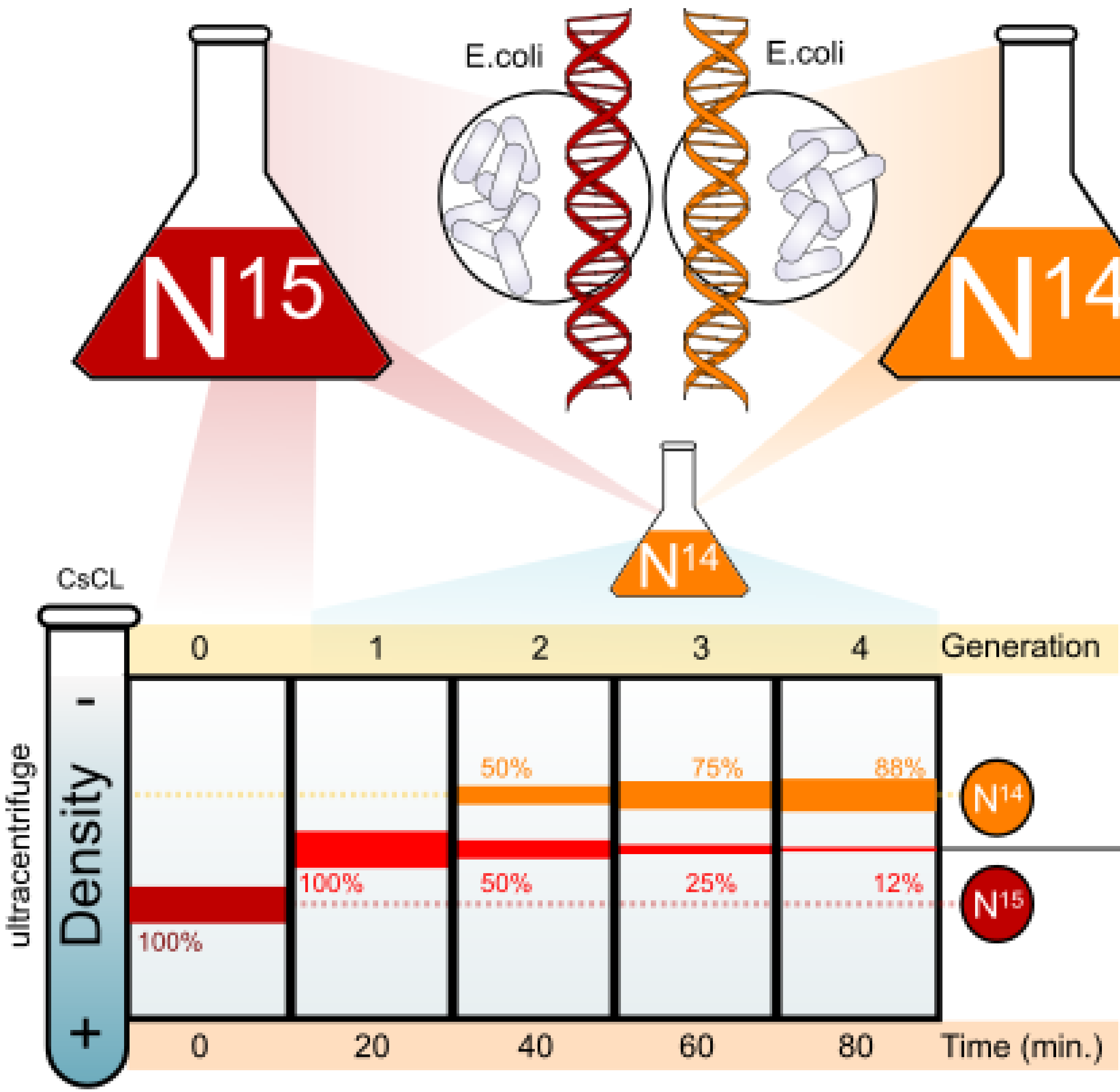


Conservative Replication

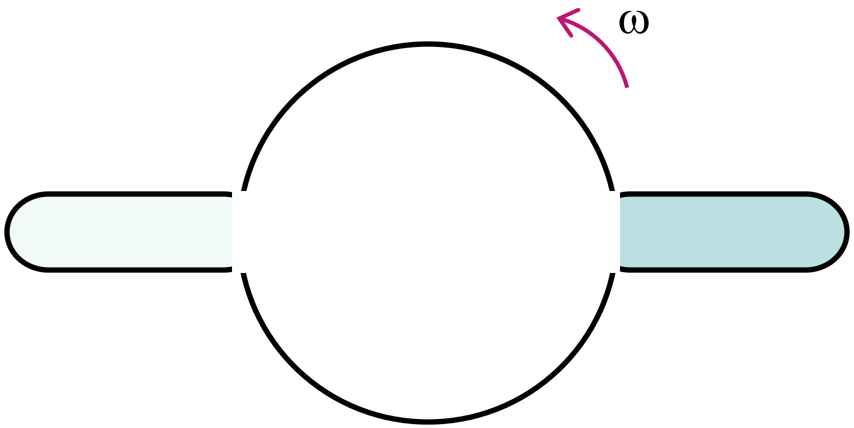


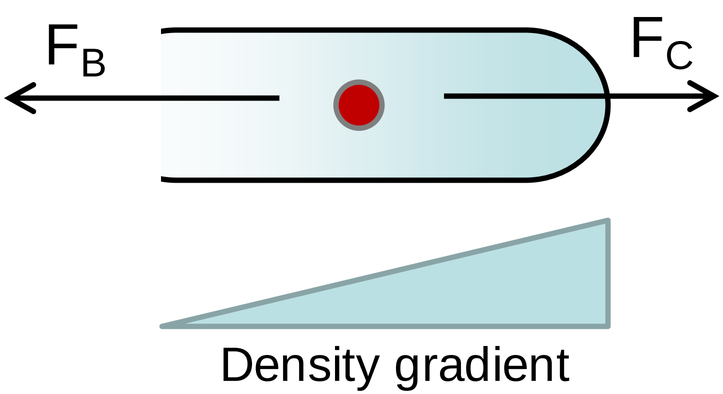
Semi-conservative Replication

The Meselson-Stahl Experiment (1958)



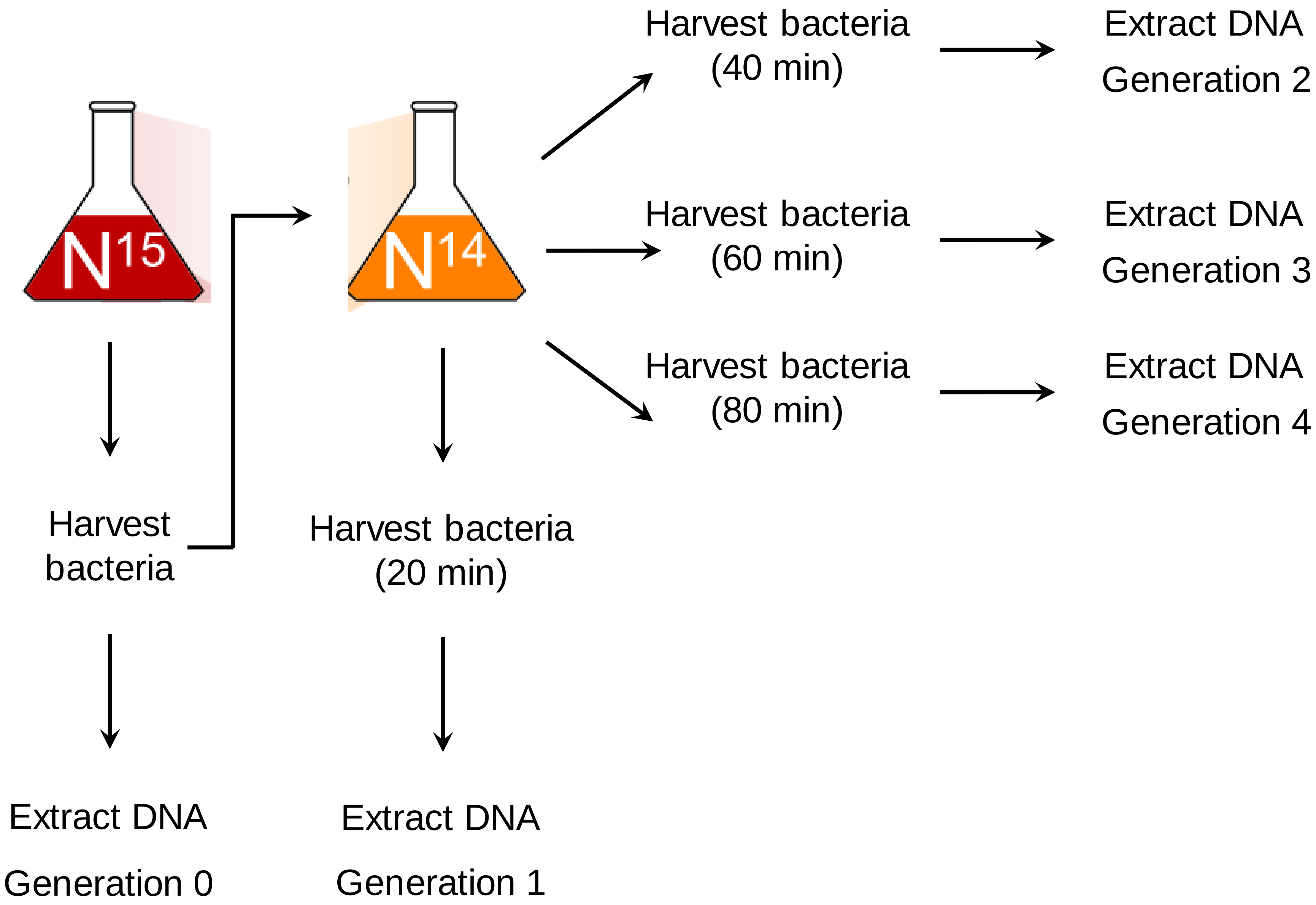
Centrifugation with density gradient



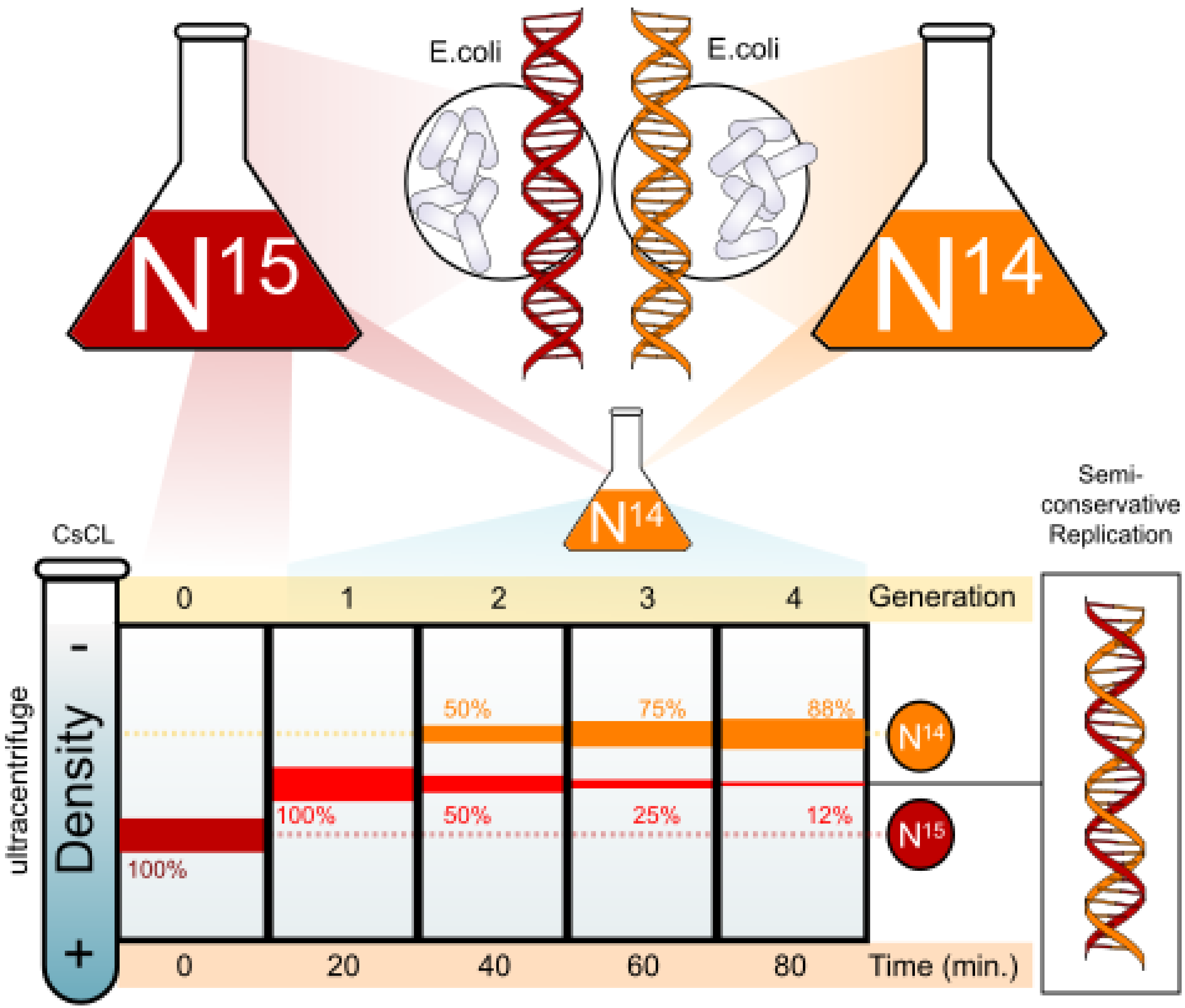


DNA with HIGHER density will move to the RIGHT i.e. BOTTOM of the tube.

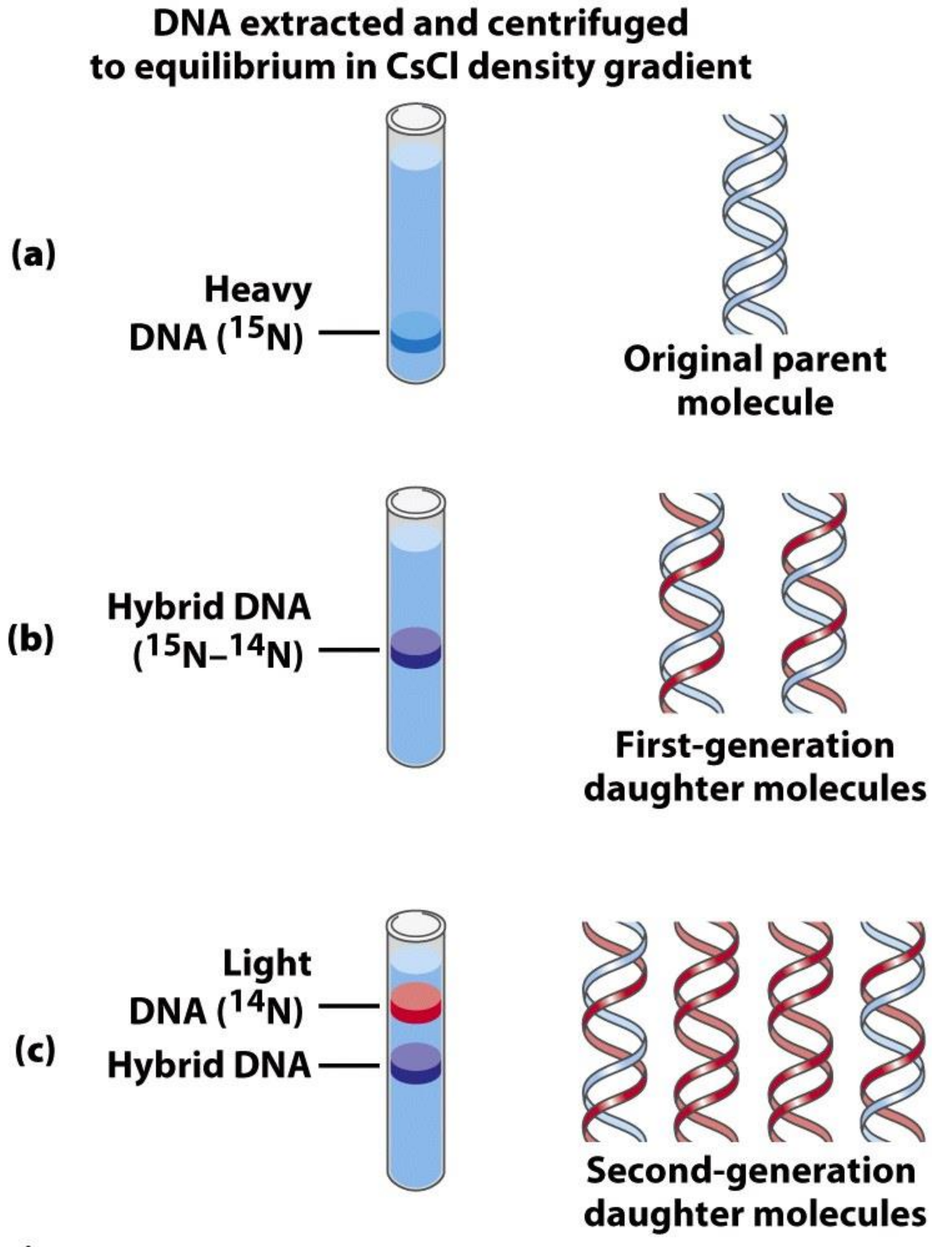
DNA with LOWER density will stay at the LEFT i.e. TOP of the tube.


N15
N14
Harvest bacteria
Harvest bacteria (20 min)
Harvest bacteria (40 min)
Harvest bacteria (60 min)
Harvest bacteria (80 min)
Extract DNA Generation 0
Extract DNA Generation 1
Extract DNA Generation 2
Extract DNA Generation 3
Extract DNA Generation 4

The Meselson-Stahl Experiment (1958)



The Meselson-Stahl Experiment Supports Semiconservative Replication

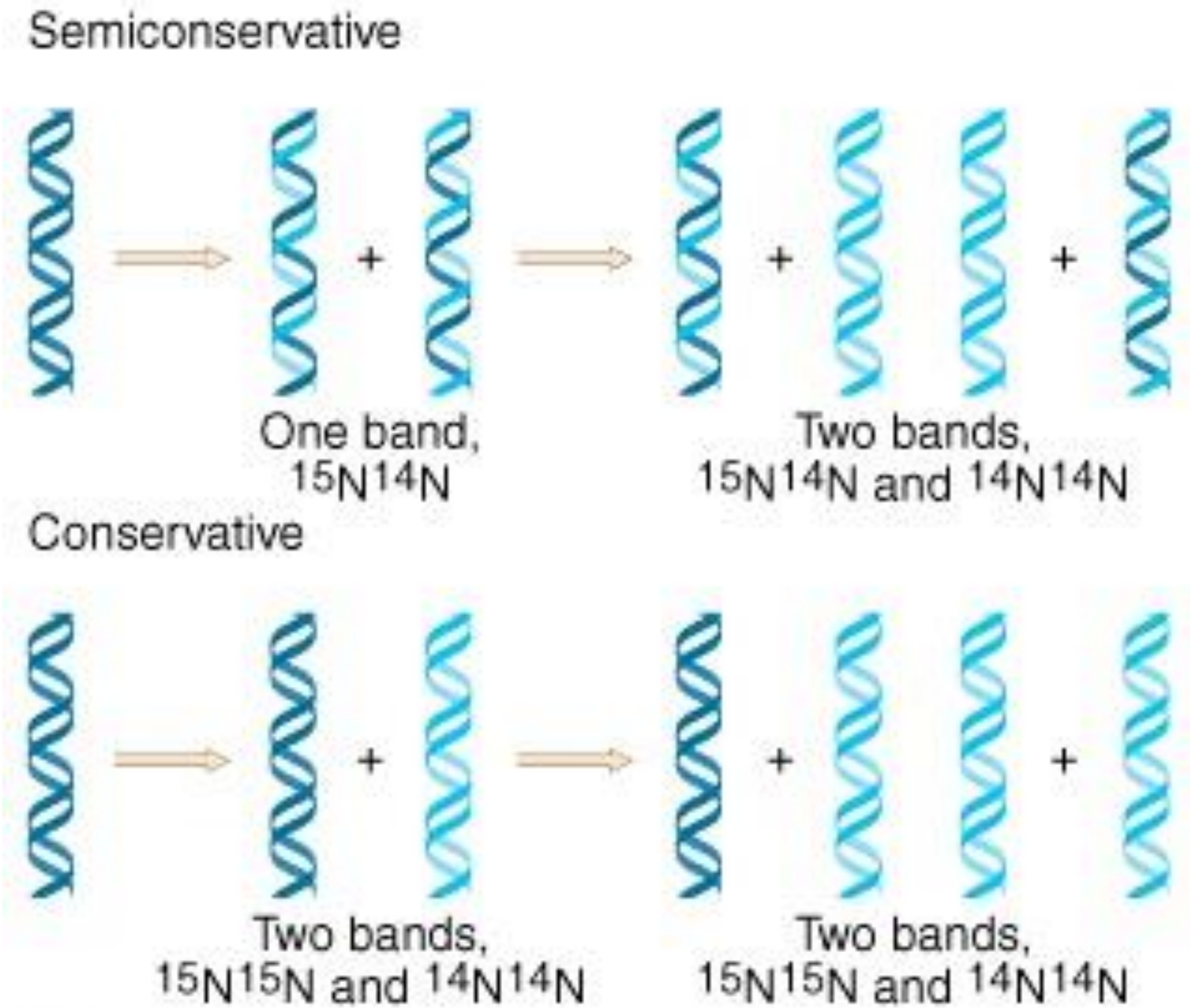


Meselson and Stahl grew a culture of *E.coli* for many generations in a medium that contained ^{15}N as the sole nitrogen source. ($^{15}NH4Cl$).
After many generations, all the *E.coli* cells had ^{15}N incorporated into the purine and pyrimidine bases of their DNA.

Meselson and Stahl took a sample of these bacteria and switched the rest of the bacteria to a medium that contained only ^{14}N (washed them before transferring to remove the medium containing ^{15}N).
They purified the DNA sample collected from the bacteria just before transfer to the ^{14}N containing medium.

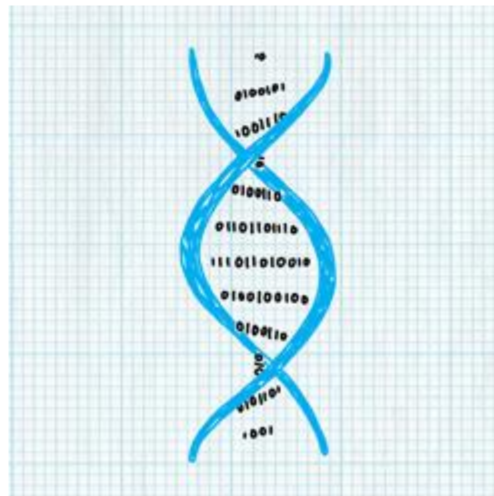
Meselson and Stahl collected some of the bacteria after each division and extracted DNA from the bacterial cells (Collected samples of bacteria over the next few cellular generations).
Under the conditions they used, *E.coli* replicates its DNA every 20 minutes. They took samples at an interval of 20 minutes.

After extracting DNA from bacteria they checked for the density of DNA


Semiconservative
One band,
15N14N
Two bands,
15N14N and 14N14N
Conservative
Two bands,
15N15N and 14N14N
Two bands,
15N15N and 14N14N

Science of Living System

BS20001 (2-0-0)



Dibyendu Samanta

School of Bio Science

Email: ***dibyendu.samanta@iitkgp.ac.in***

Tel: 03222-260295

DNA Replication: An Overview

The determination of DNA's double-helix structure

(Watson and Crick in 1953)

&

The discovery of high-temperature superconductors

(IBM researchers: Bendnorz and Muller in 1986)

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

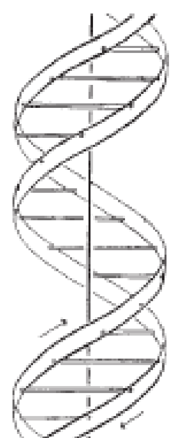
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each base is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

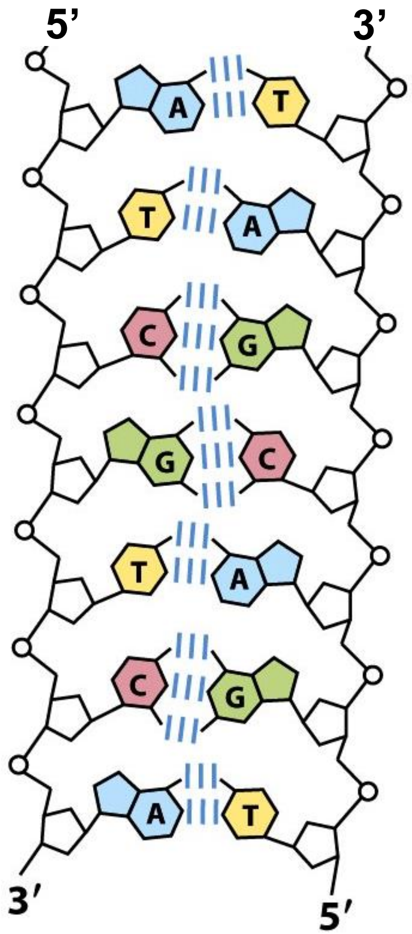
It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

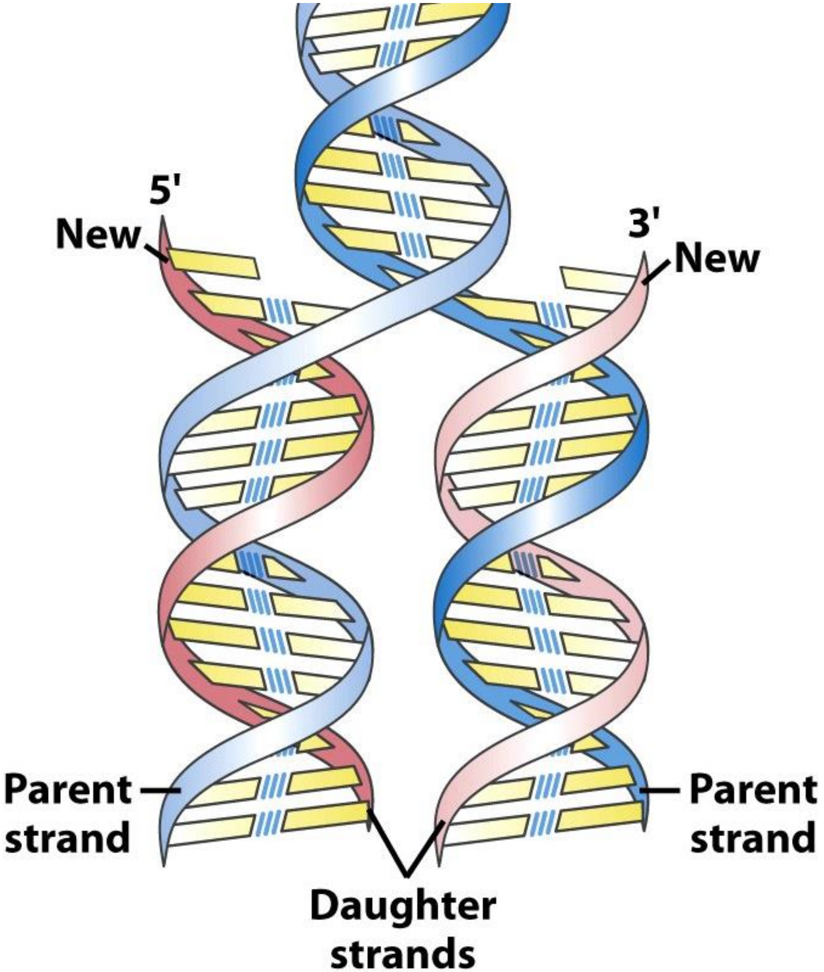
The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

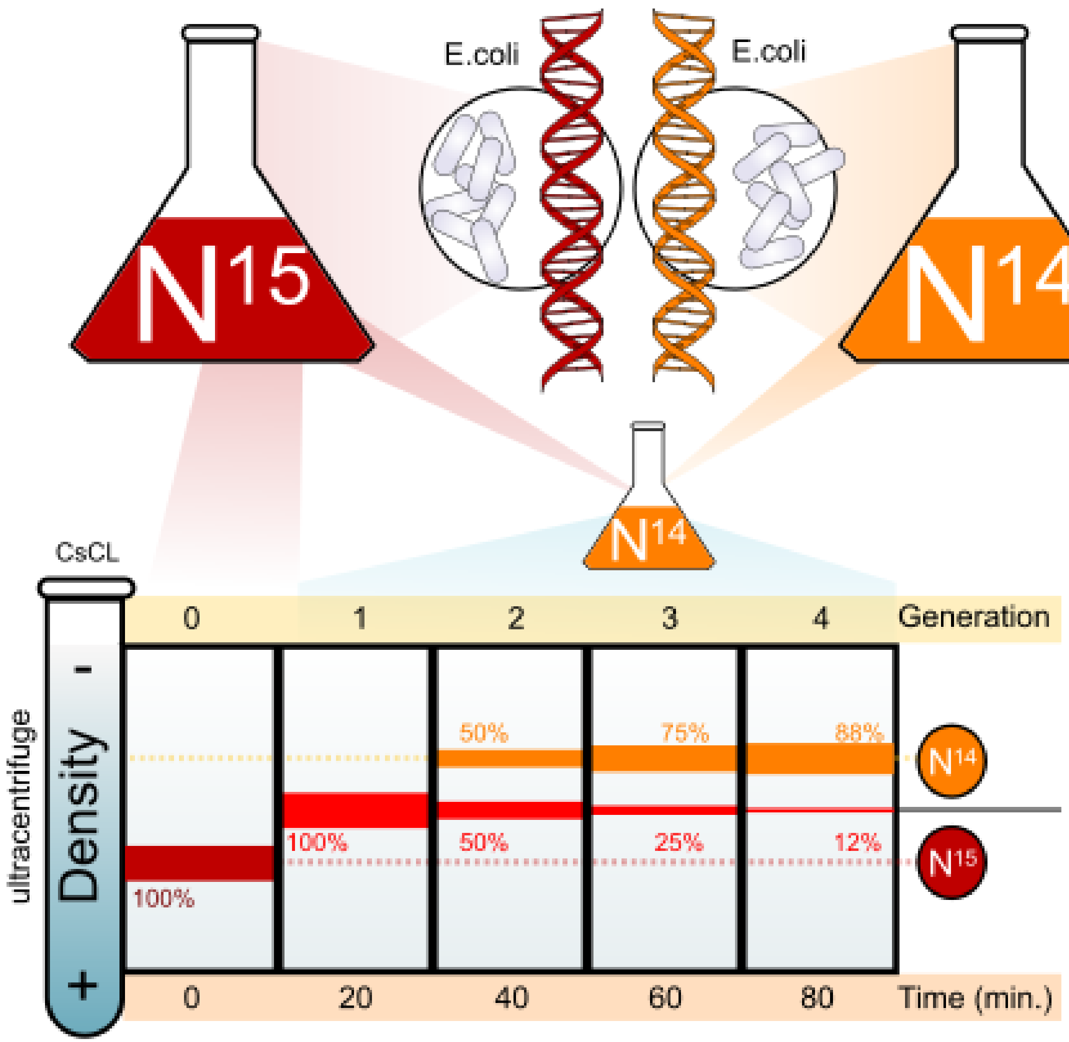
Semiconservative Replication of DNA as suggested by Watson and Crick

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

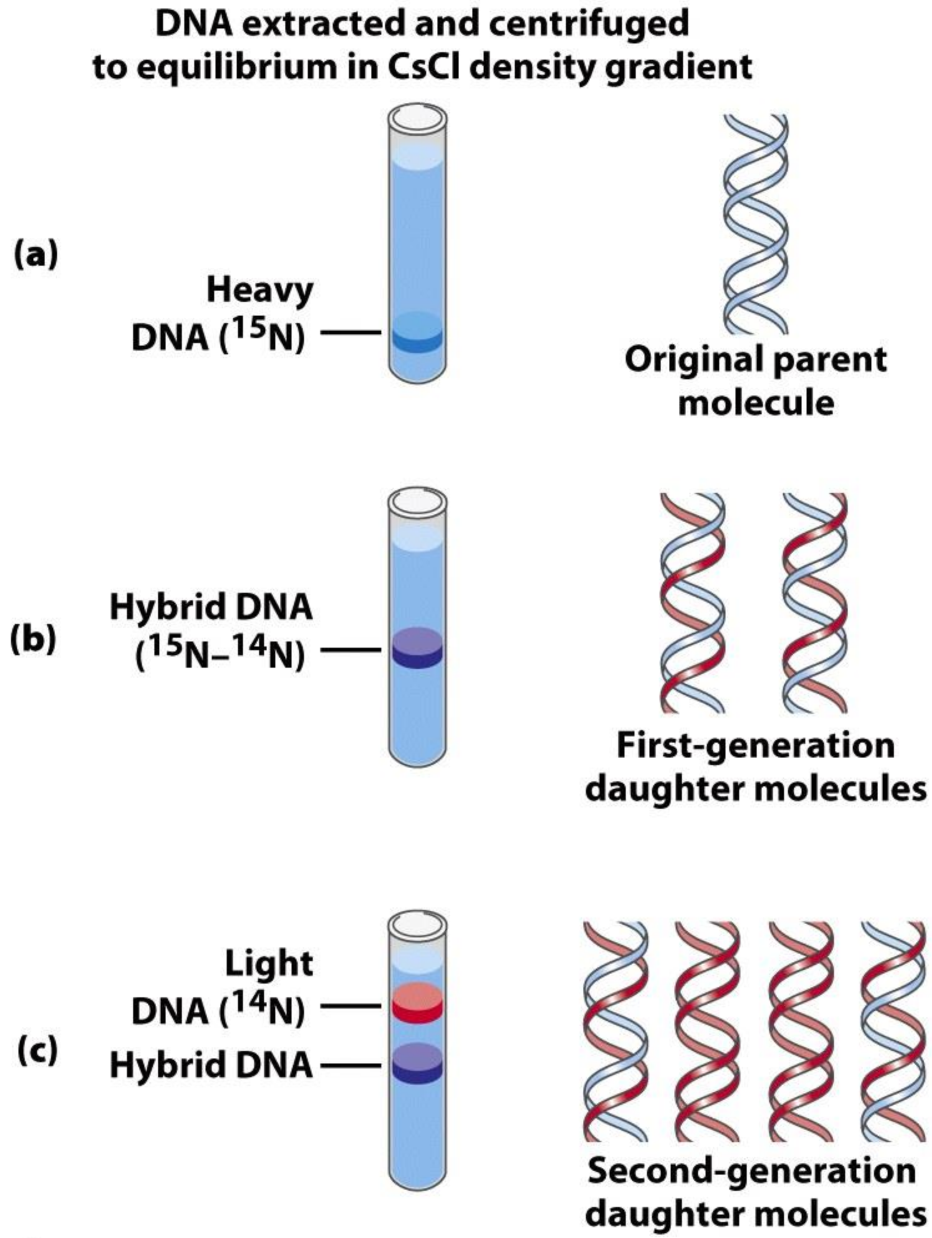




The Meselson-Stahl Experiment



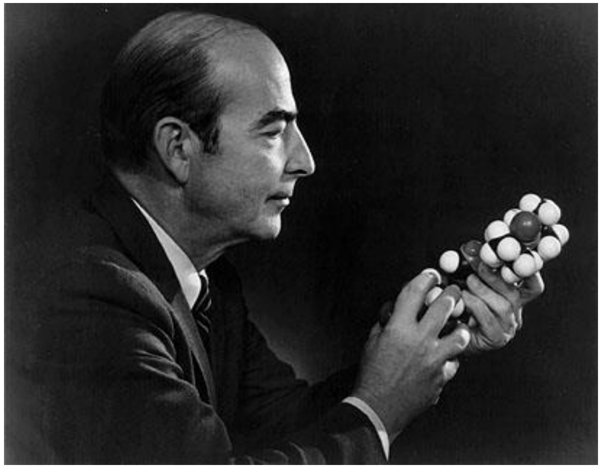
The Meselson-Stahl Experiment Supports Semiconservative Replication



Replication: The Basic Concept

Discovered the mechanisms of DNA synthesis

Worked with *E. coli*



Arthur Kornberg
Nobel Prize in 1959

Four components are required

1. dNTPs: dATP, dTTP, dGTP, dCTP

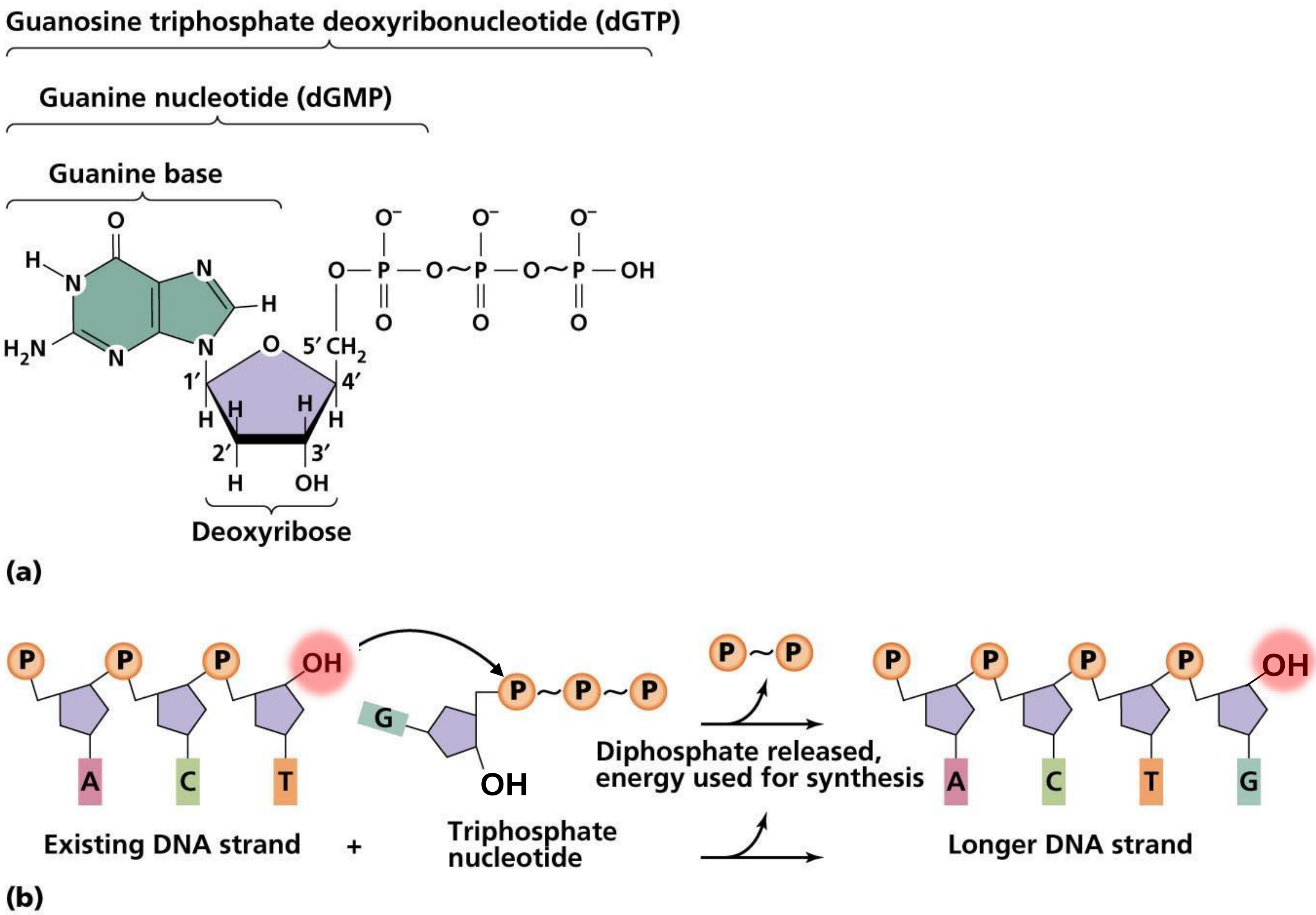
   (sugar + base + 3 phosphates)

2. DNA template

3. DNA polymerase (*Kornberg enzyme*)

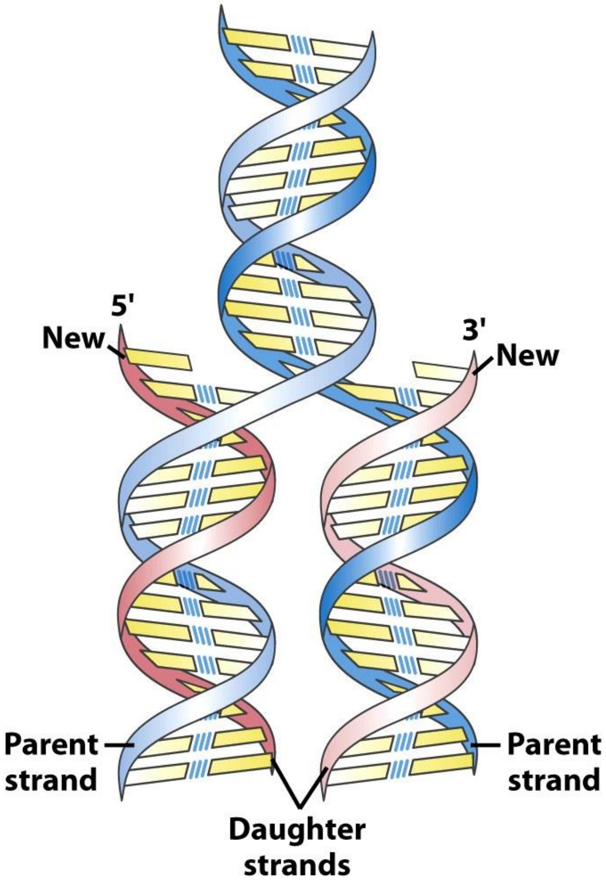
4. Mg^{2+} (optimizes DNA polymerase activity)

Polymerization of nucleotide triphosphates by DNA polymerase – needs free 3'OH



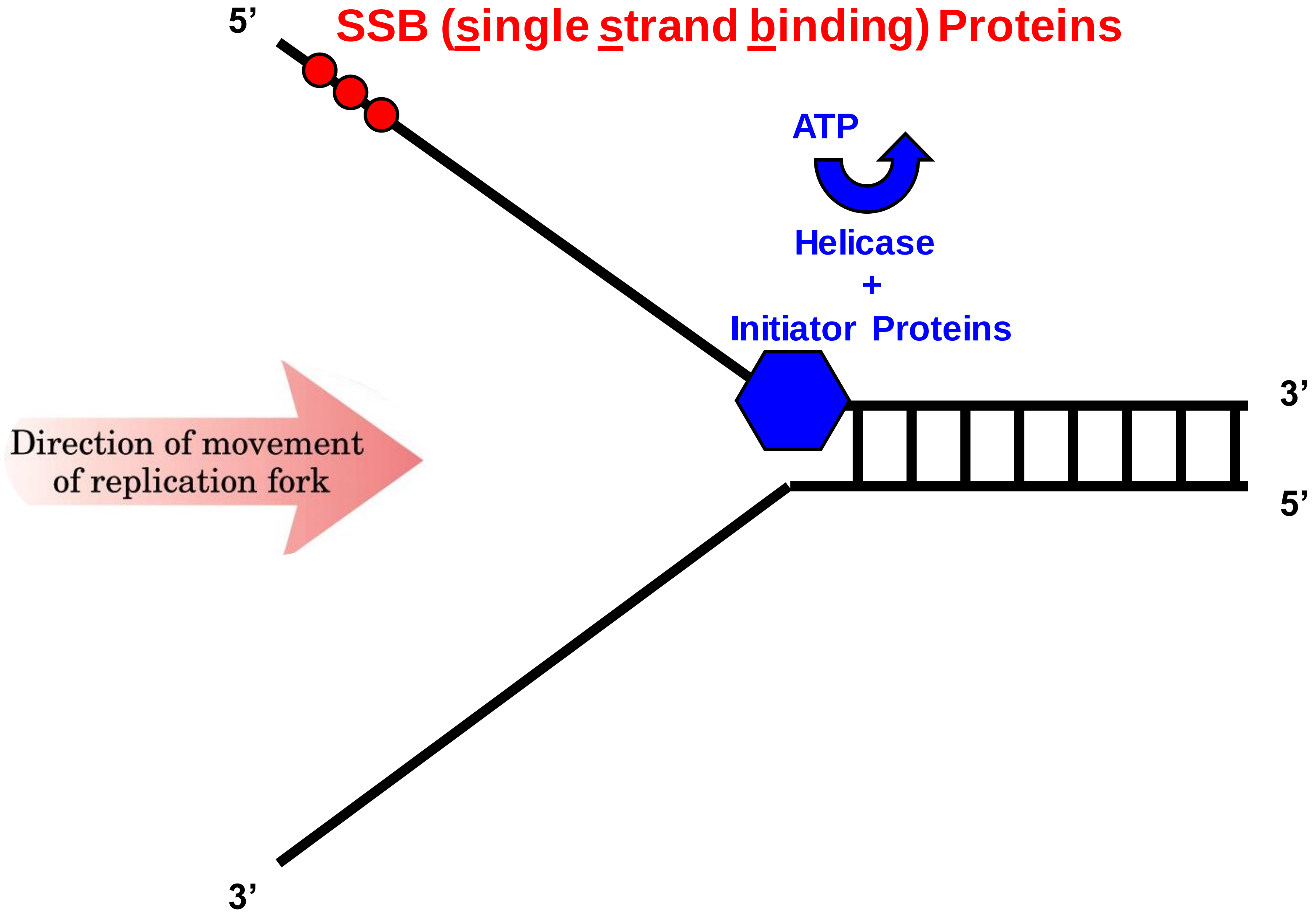
Problem -1

DNA is double helical. Hence the two strands has to be separated from each other



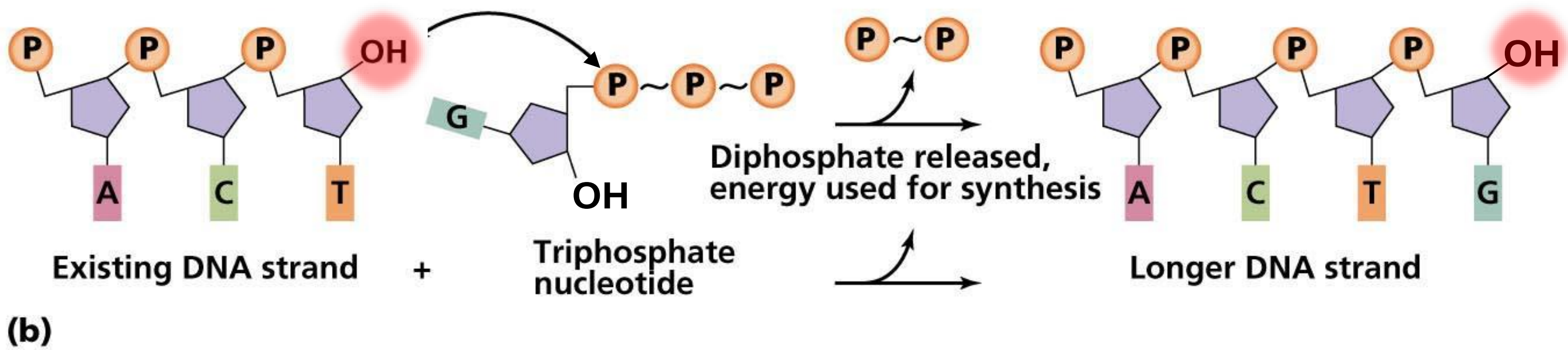
Replication: The Basic Concept

Supercoiled DNA relaxed by Topoisomerase & unwound by helicase + proteins



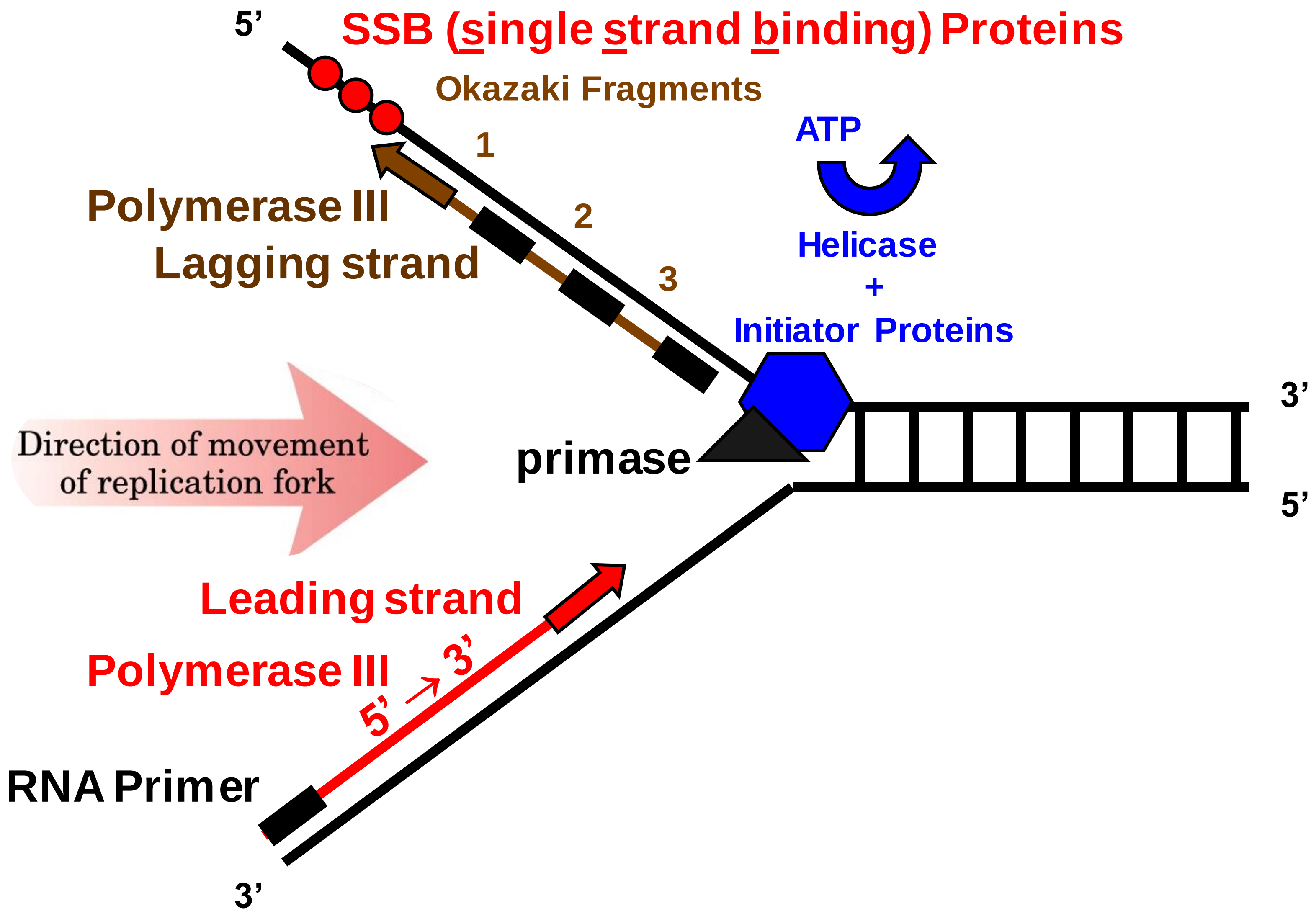
Problem -2

DNA polymerase needs a free 3'-OH of a pre-existing nucleotide to extend the chain



Replication: The Basic Concept

Supercoiled DNA relaxed by gyrase & unwound by helicase + proteins

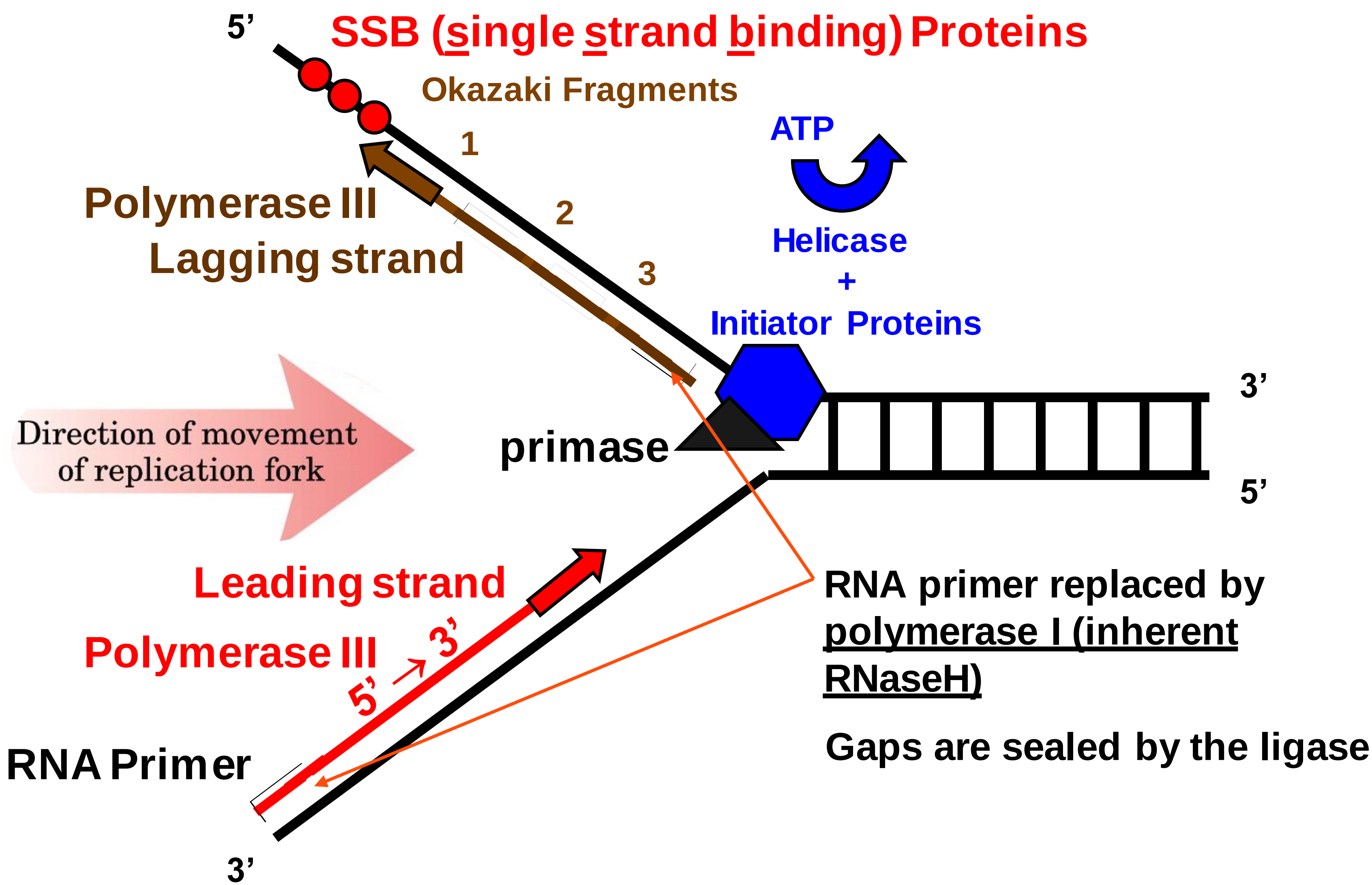


Problem -3

RNA primers needs to be removed

Replication: The Basic Concept

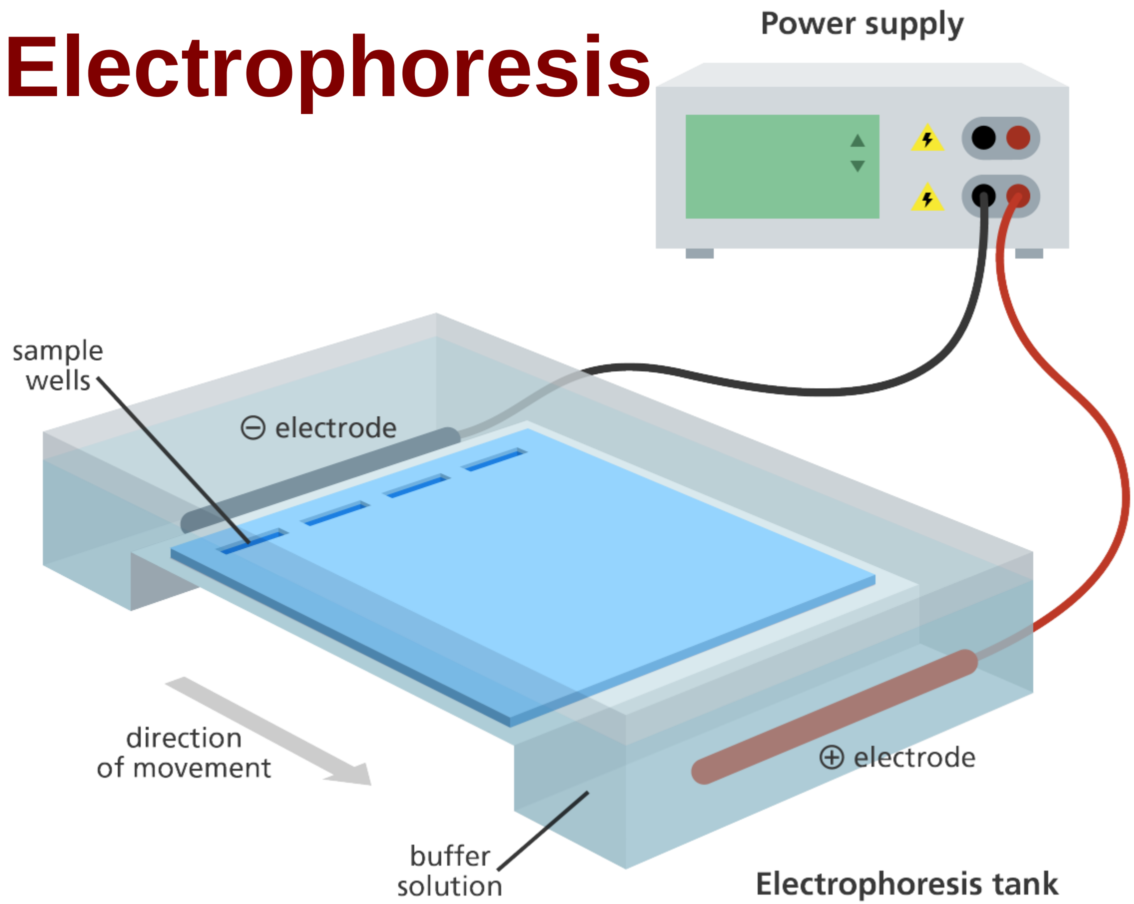
Supercoiled DNA relaxed by gyrase & unwound by helicase + proteins

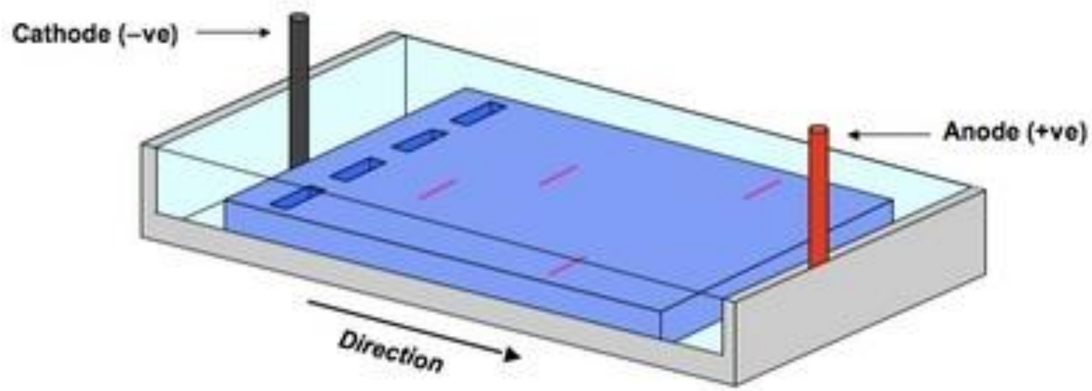


Laboratory Methods for the Analysis of DNA

Gel electrophoresis, PCR and DNA sequencing

DNA Gel Electrophoresis





DNA Gel Electrophoresis

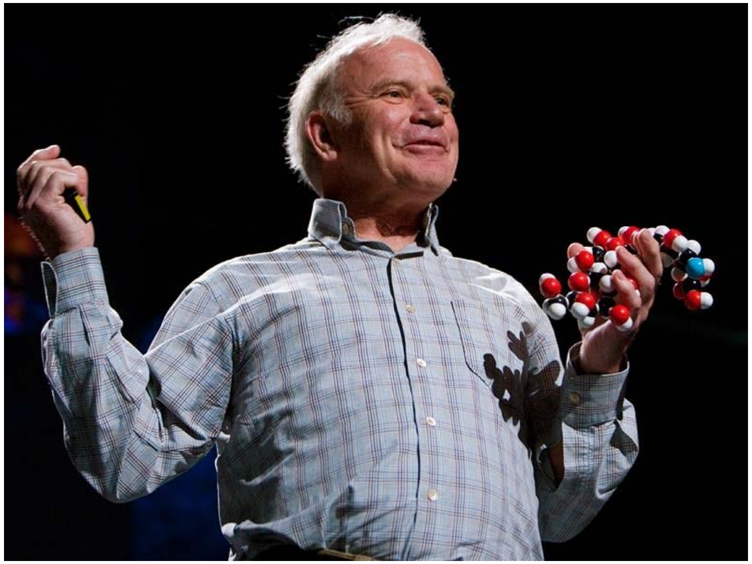
- Agarose gel is used to analyze a mixture of DNA.
- DNA is negatively charged and hence migrates towards the positive terminal in the applied electric field gradient.
- Different DNA molecules separate according to mass.
- Smaller molecules migrate faster.
- DNA is visualized in the gel by staining with ethidium bromide, which fluoresces under UV light.
- UV light at 302 nm or 365 nm wavelengths are used.

Polymerase Chain Reaction (PCR) and Its Applications

What is PCR?

PCR is an exponentially progressing synthesis of the defined target DNA sequences in vitro.

It was invented in 1983 by Dr. Kary Mullis



Nobel Prize in 1993

Polymerase Chain Reaction (PCR)

Why “Polymerase” ?

It is called “polymerase” because the only enzyme used in this reaction is DNA polymerase (Specifically Taq Polymerase).

Why “Chain” ?

It is called “chain” because the products of the first reaction become substrates of the following one, and so on.

Polymerase Chain Reaction (PCR)

The "Reaction" Components

1) Target DNA - contains the sequence to be amplified.

2) Pair of Primers - oligonucleotides that define the sequence to be amplified.

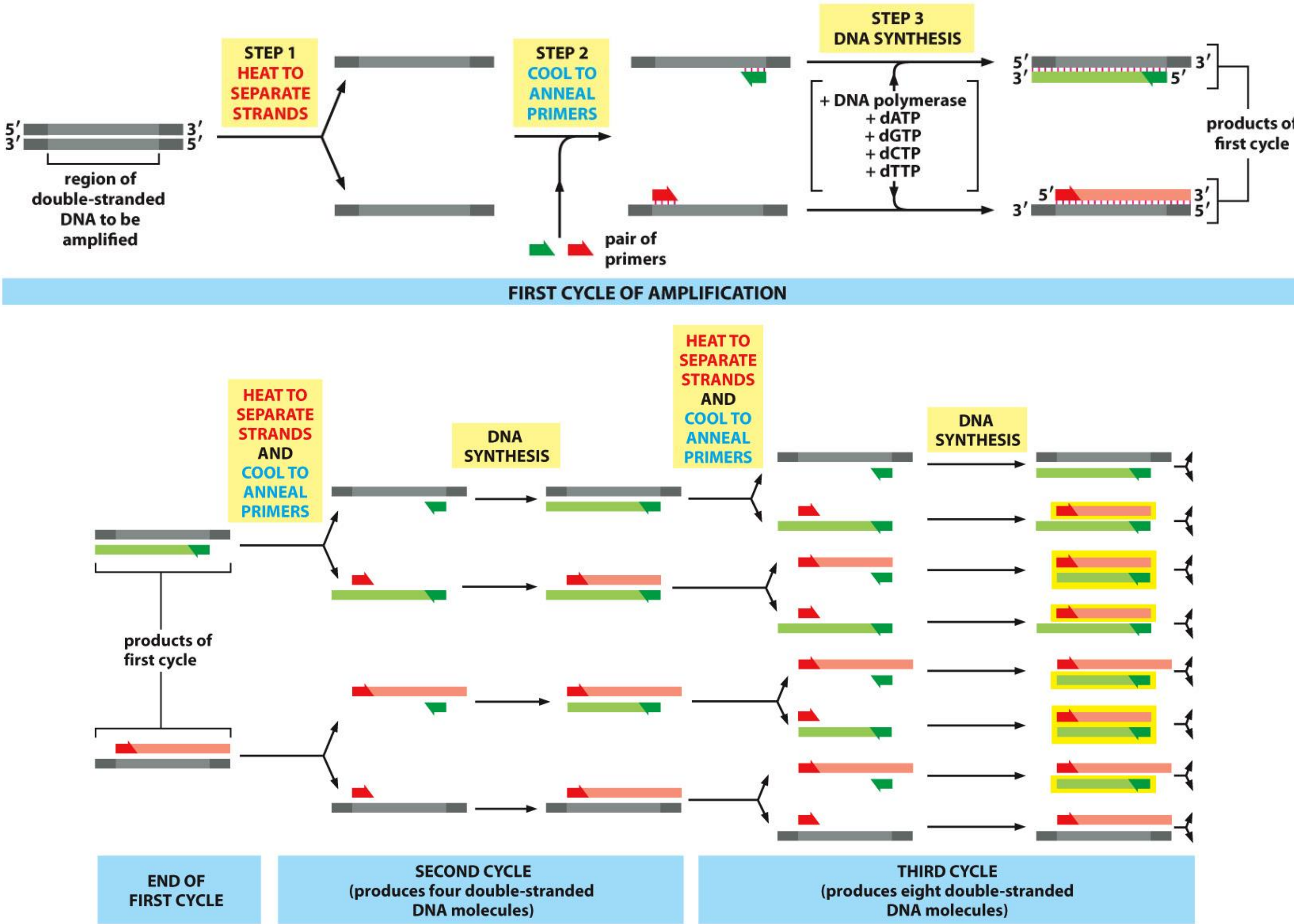
3) dNTPs - deoxynucleotidetriphosphates: DNA building blocks.

4) Thermostable DNA Polymerase – enzyme (Taq Polymerase) that catalyzes the reaction

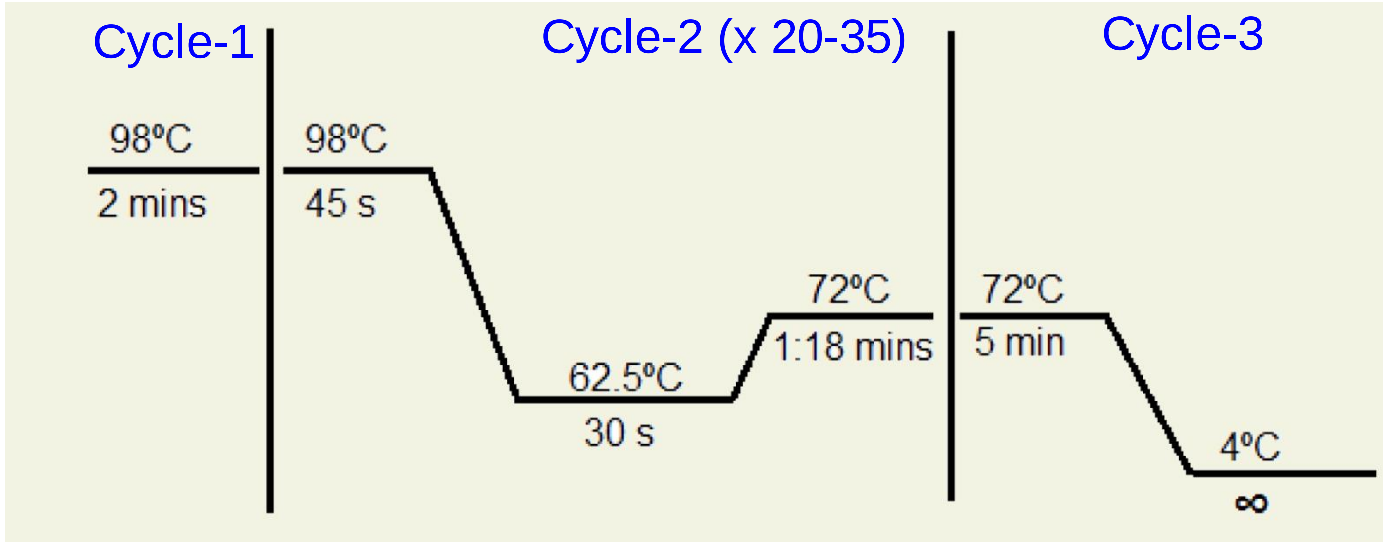
5) Mg^{++} ions - cofactor of the enzyme

6) Buffer solution - maintains pH and ionic strength of the reaction solution suitable for the activity of the enzyme

PCR Reaction Steps



Reaction Cycle and the Machine

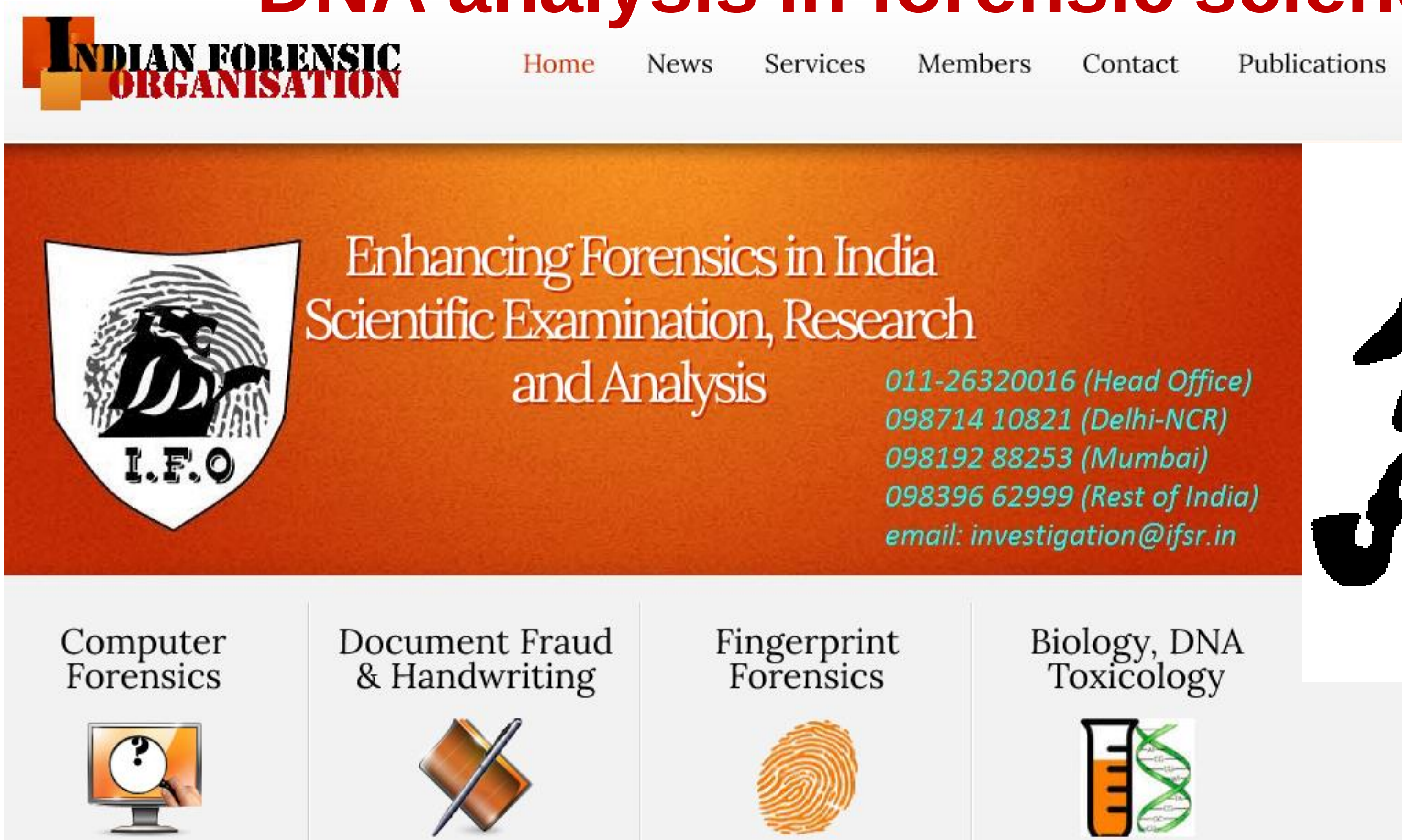


30 cycles = 2^{30} i.e. 1,073,741,824 copies

PCR tube

Thermocycler

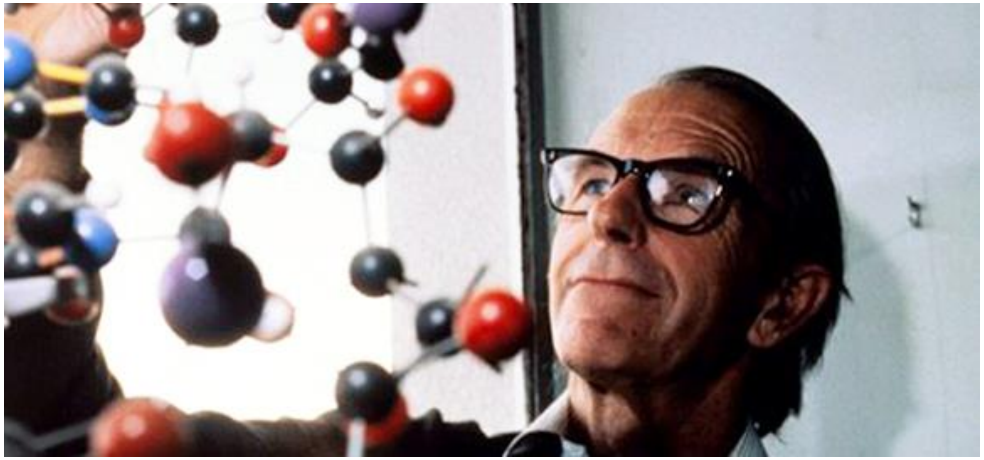
DNA analysis in forensic science



- DNA analysis has become an important tool in forensic science in the past decade.
- The innocence project (http://www.innocenceproject.org) in the USA has used DNA testing to free 343 wrongfully convicted people and finding of 147 real perpetrators.

DNA Sequencing and Its Applications

DNA Sequencing



Frederick Sanger
Nobel Prize in 1980

Sequencing: **The process by which you determine the exact order of the nucleotides in a given region of DNA**

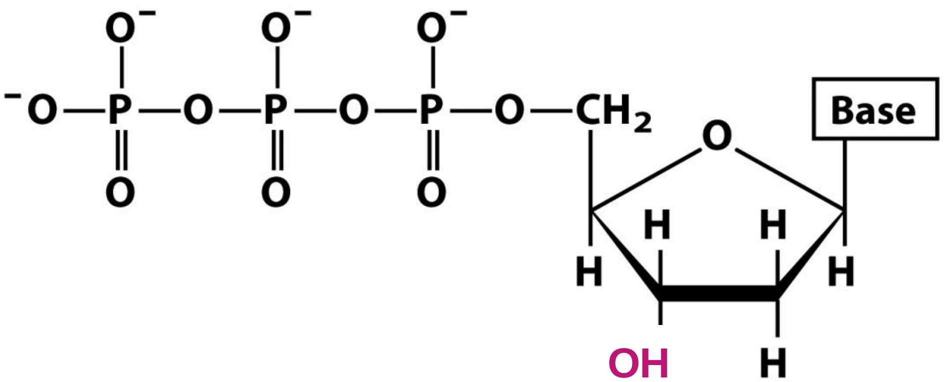
Why? **To decode the genetic information embedded in DNA**

How? **Through complementary chain synthesis and early termination**

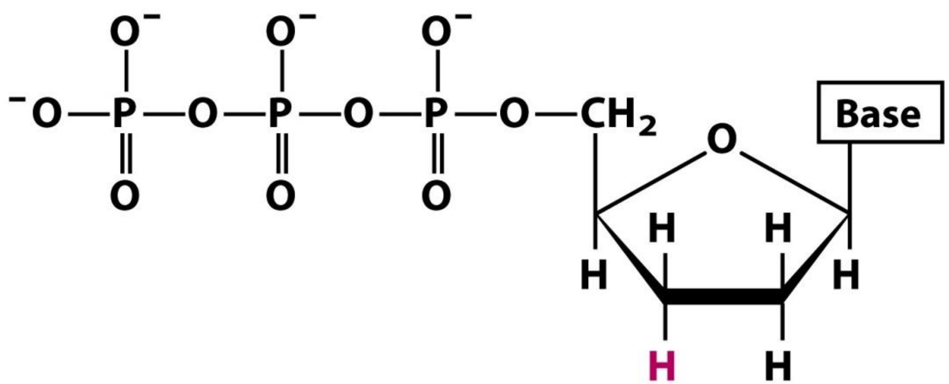
The synthesized chains are visualized by methods using:

- **Radioactive labels**
- **Nonradioactive labels**

Dideoxynucleotides (ddNTPs)



dNTP



ddNTP analog

The 3’ hydroxyl has been changed to a hydrogen in ddNTP’s, which terminates a DNA chain because a phosphodiester bond cannot form at this 3’ location

Requirements for Sanger-Coulson Sequencing

DNA to be sequenced must be in single strand form

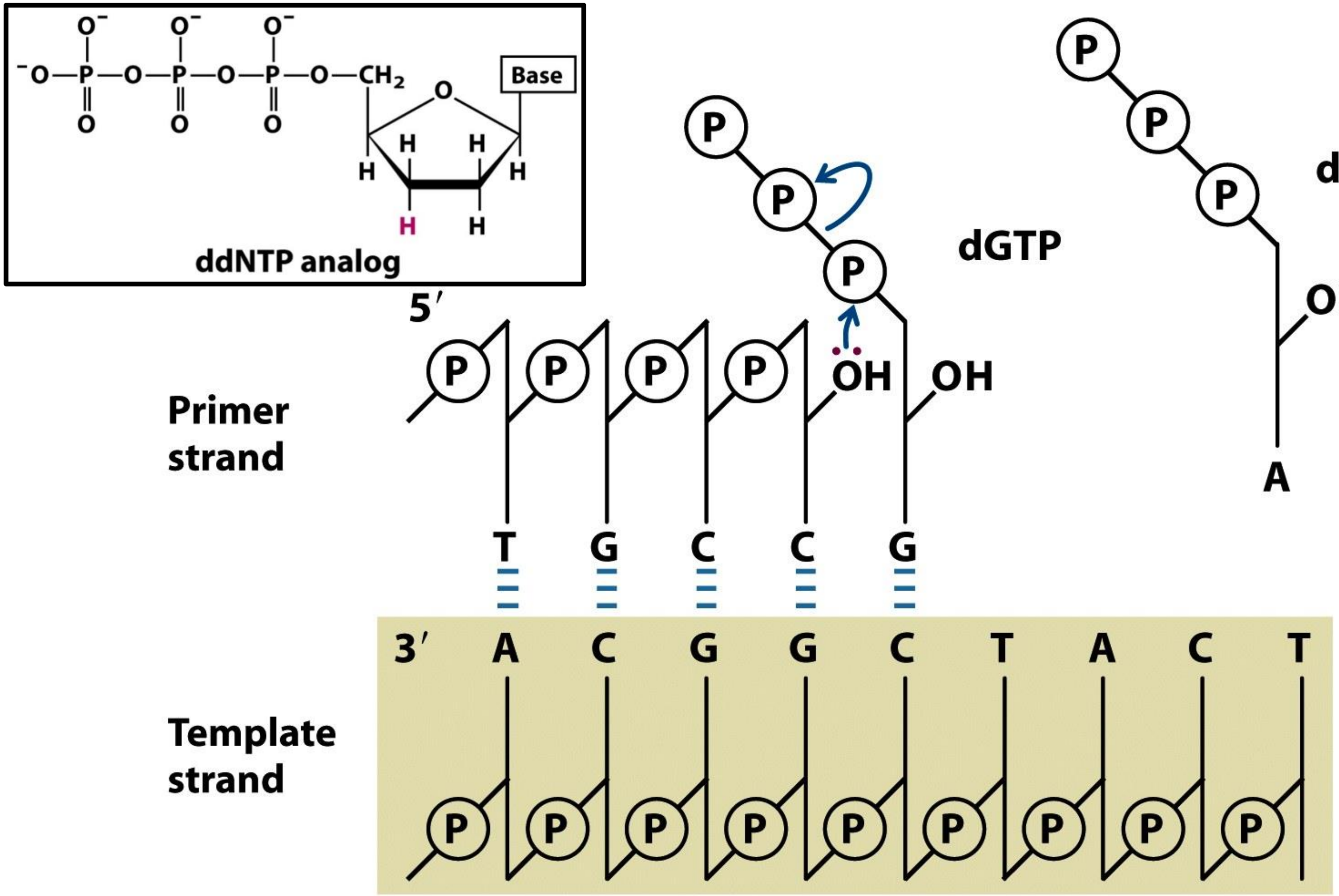
The region to be sequenced must be flanked by a stretch of 3' known sequence

Reagents needed are:

- **A primer complementary to the known region to direct chain synthesis.**
- **DNA polymerase.**
- **4 deoxynucleotide triphosphates (dNTPs).**
- **4 dideoxynucleotide triphosphates (ddNTPs)**

(one for each set)

Mechanism of DNA polymerization



Since the 3’ –OH is changed to a –H in ddNTPs, it is unable to form a phosphodiester bond by nucleophilic attack on the phosphate, and it will cause a termination in the DNA chain

Sanger's Dideoxy Sequencing Reaction

Template	3' – A T A T C T G T A C C T A G G T G A G T C A G T A C C -5'
Primer	5' – T A T A G A C
Differentially elongated fragments	5' – T A T A G A C ddA
	5' – T A T A G A C A T G G ddA
	5' – T A T A G A C A T G G A T C C ddA
	5' – T A T A G A C A T G G A T C C A C T C ddA
	5' – T A T A G A C A T G G A T C C A C T C A G T C ddA

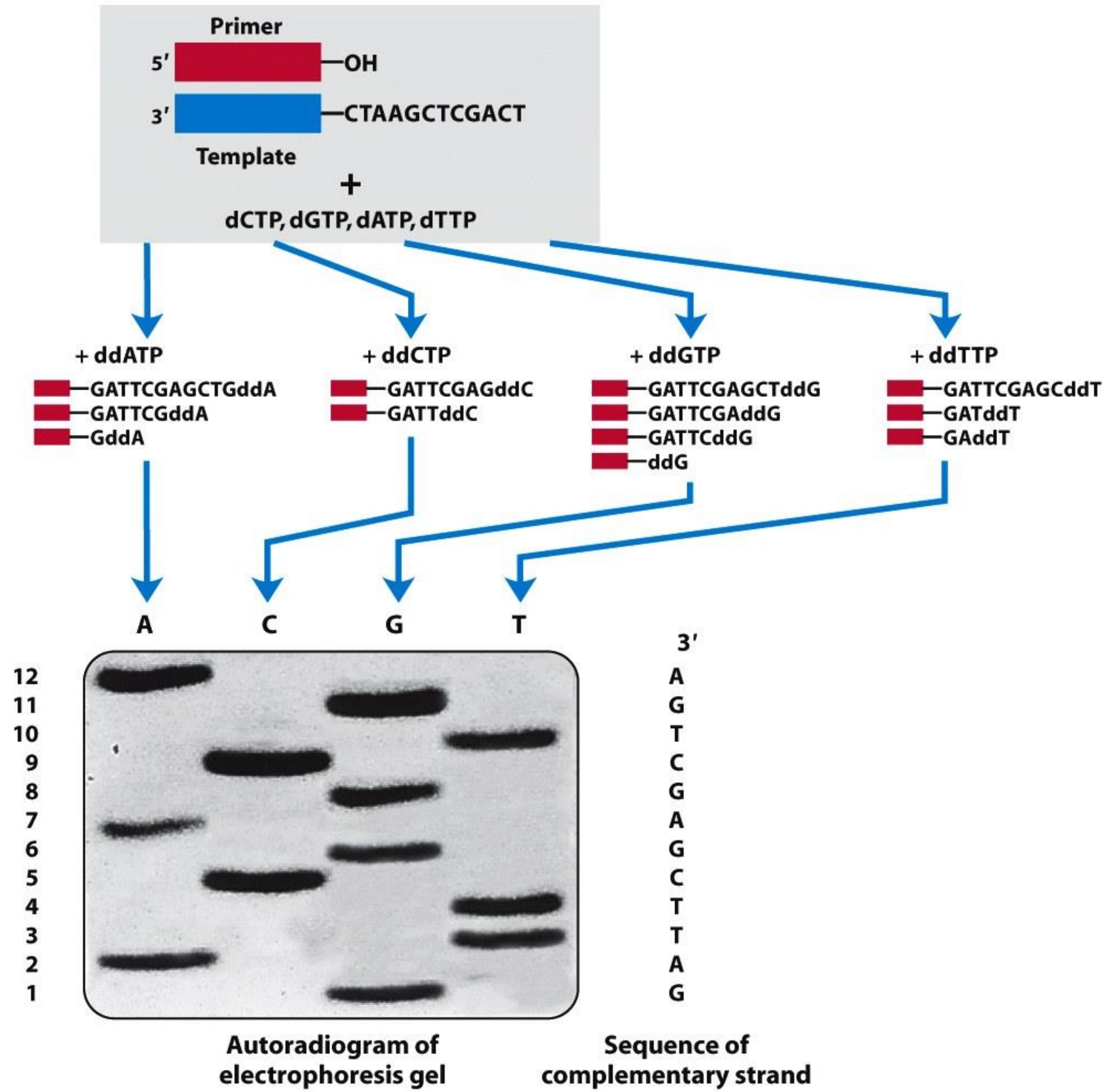
To each reaction along with the four regular dNTPs only one of the four dideoxynucleotides (ddATP in the above reaction) is added

The ddNTP is added to be approximately 100-fold lower in concentration than the corresponding dNTP

A series of differentially elongated DNA fragments (hence of different length) will be produced

Can be visualized by labelling the primers with radioactive phosphate (^{32}P or ^{33}P)

Sanger's Dideoxy Sequencing Reaction



Automated sequencing

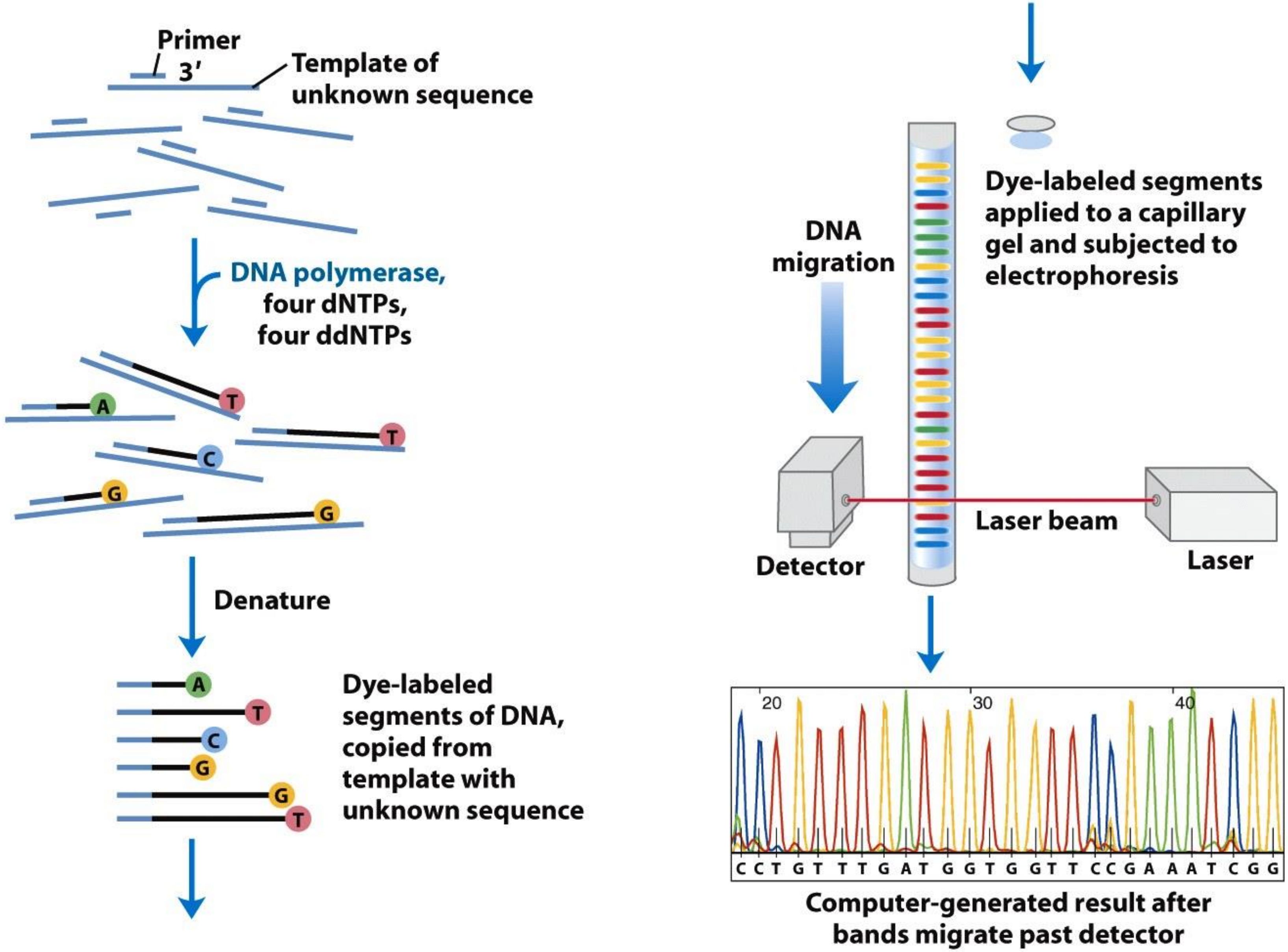


Figure 8-34
Lehninger Principles of Biochemistry, Fifth Edition

Automated sequencing

Each dideoxynucleotide used are linked to a fluorescent molecule that gives all the fragments terminating in that nucleotide a particular color.

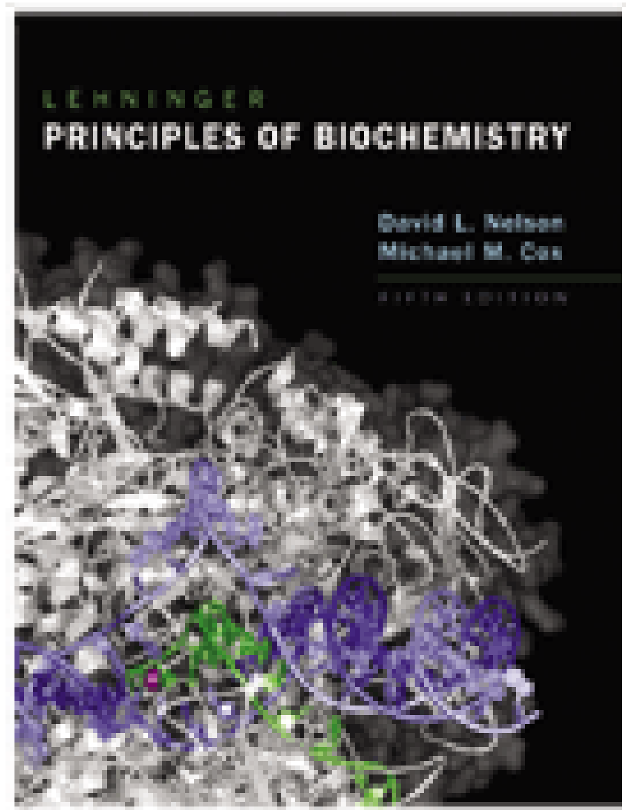
All four labelled ddNTPs are added to a single tube.

The resulting colored DNA fragments are then separated by capillary gel electrophoresis

The color associated with each peak is detected using a laser beam.

The DNA sequence is read by determining the sequence of colors in the peaks as they pass the detector.

This information is fed directly to a computer, which determines the sequence.



http://bcs.whfreeman.com/lehninger5e/

Extra resources

Book:

The Double Helix: A Personal Account of the Discovery of the Structure of DNA - By James D Watson

Videos:

DNA replication
https://www.youtube.com/watch?v=TNKWgcFPHqw

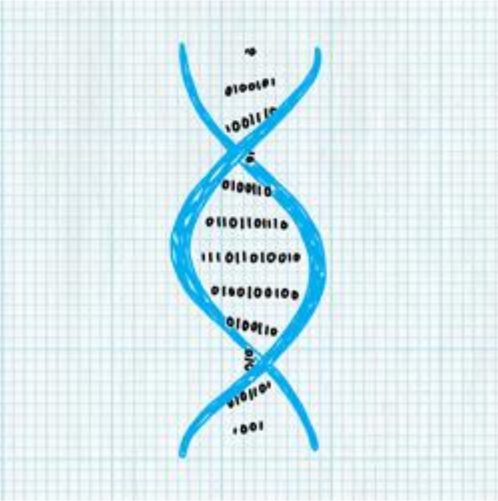
Gel electrophoresis
https://www.youtube.com/watch?v=vq759wKCCUQ

Polymerase Chain Reaction (PCR)
https://www.dnalc.org/view/15475-The-cycles-of-the-polymerase-chain-reaction-PCR-3D-animation.html

February, 2022

Science of Living System

BS20001 (2-0-0)



Dibyendu Samanta

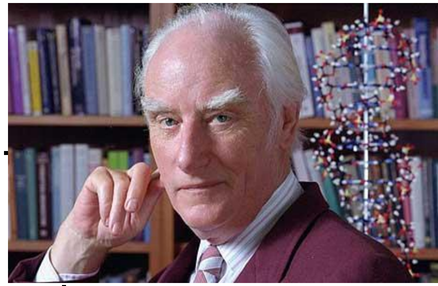
School of Bio Science

Email: dibyendu.samanta@iitkgp.ac.in

Tel: 03222-260295

Overview of Transcription and Translation

Flow of Genetic Information: The Central Dogma of Molecular Biology



Nobel Prize in 1962 for DNA Structure

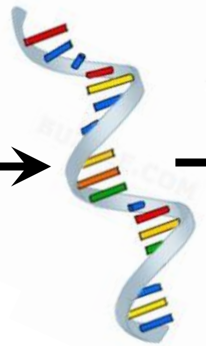
DNA Polymerase

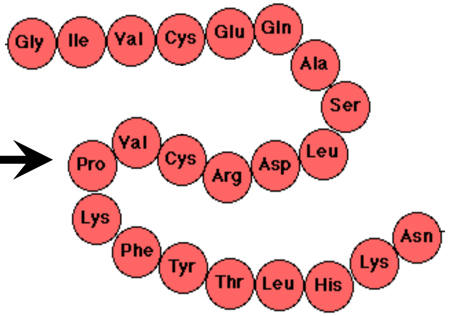
Replication

DNA $\xrightarrow{\text{RNA Polymerase}}$ RNA $\xrightarrow{\text{Ribosome}}$ Protein

Transcription

Translation





DNA

Polymer of nucleotides

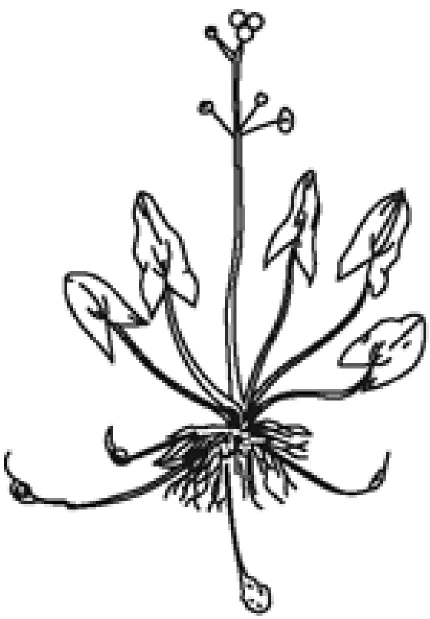
RNA

Polymer of nucleotides

Protein

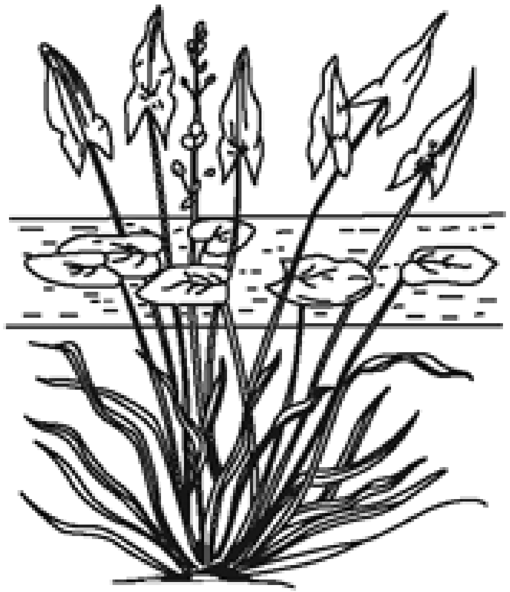
Polymer of amino acids

DNA: Contains the Instruction for Life



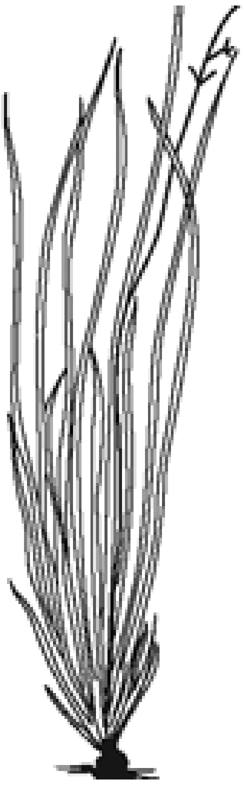
Plant-X

Completely terrestrial



Plant-Y

Partially submerged

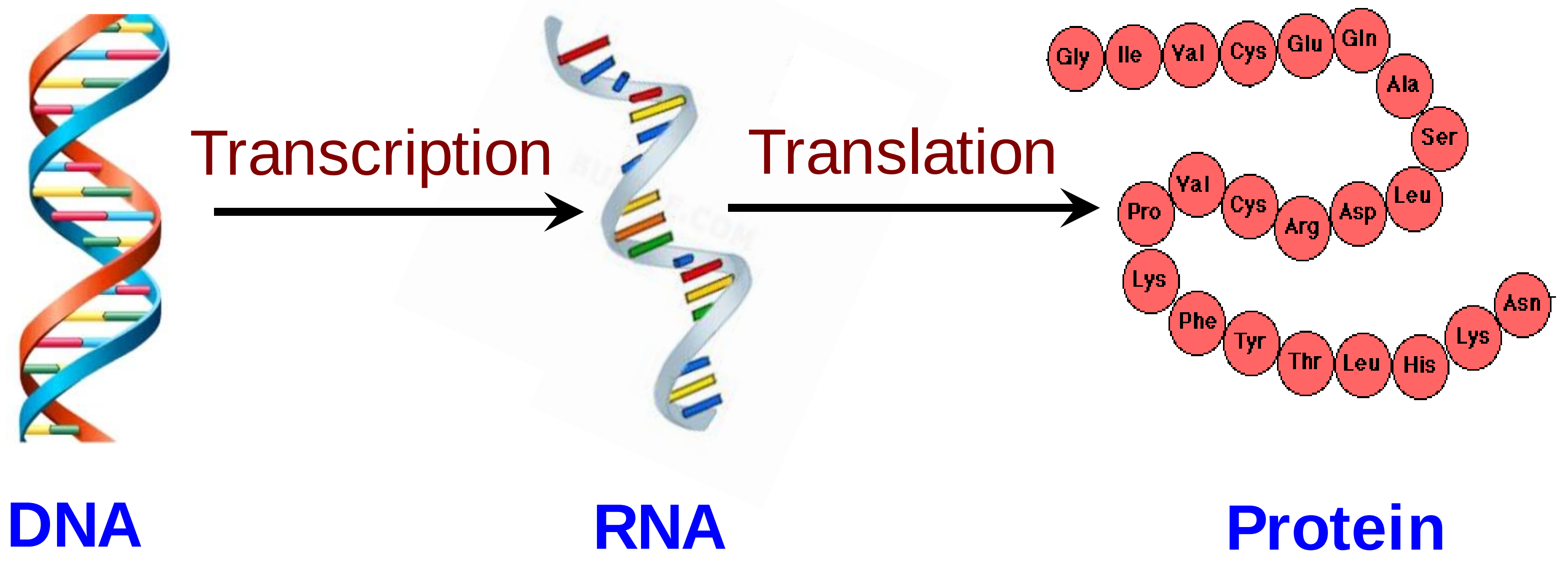


Plant-Z

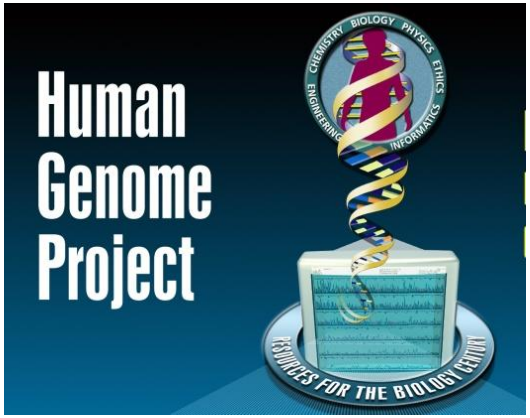
Completely submerged

The phenotype (visible configuration) of the marsh plant *Sagittaria sagittifolia* depends on its environment

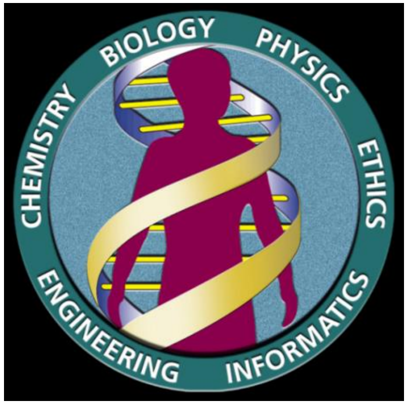
Correlations Between DNA Content and Its Down-stream Product

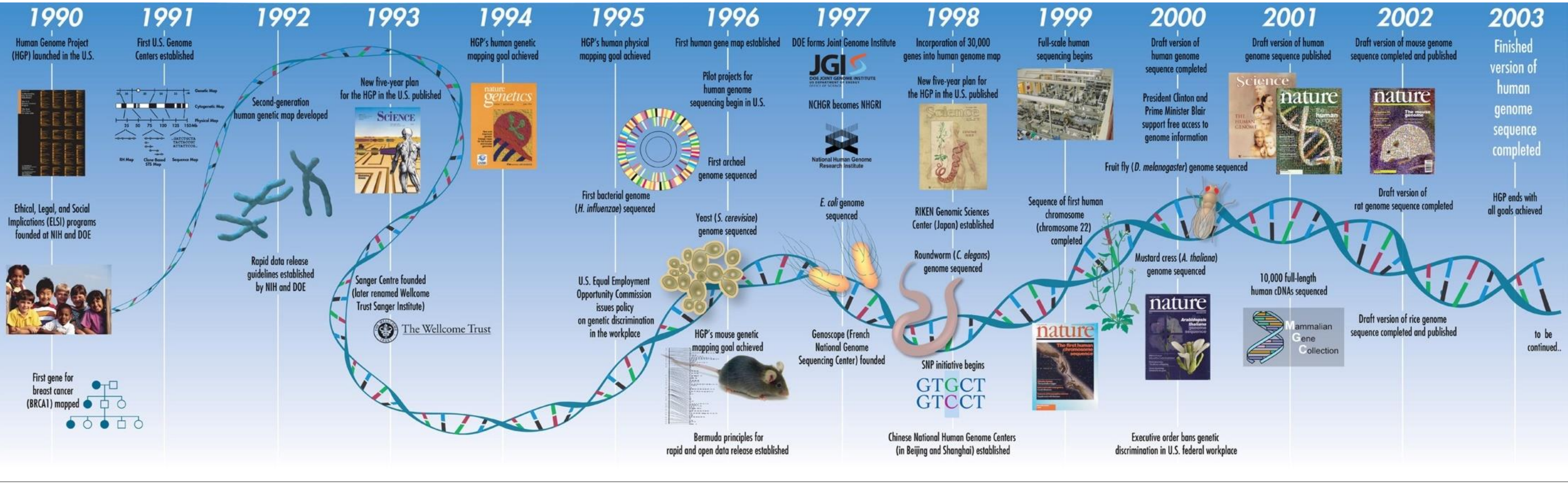


"Human Genome Project" Dramatically Enhanced Our Understanding on Gene Expression

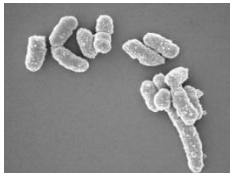
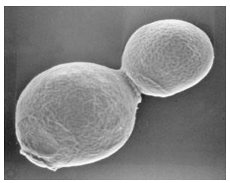


~21,000 human genes (appeared to be significantly fewer than previous estimates)





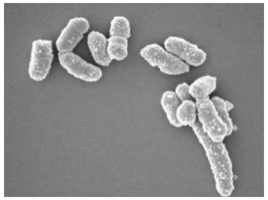
Genome Size, Gene Number, and Complexity of an Organism

	Organism	Genome size (bp)	Protein coding genes
	*E. coli*	4,600,000	4,250
	*S. cerevisiae*	12,160,000	5,616
	*C. elegans*	100,000,000	19,735
	Human	3,200,000,000	19,042
	Marbled lungfish	139,000,000,000	NA

Transcription

Genome size (bp)

Total DNA content vs transcribable content



4,600,000

3,200,000,000

►**Protein coding sequences is ~1.5% of total DNA content (human)**

Messenger RNA (mRNA)

►**Besides protein coding region, DNA can be transcribed into:**

Ribosomal RNA (rRNA)
Transfer RNA (tRNA)

►**Most of the DNA sequences are not transcribed**

What is a Machine?

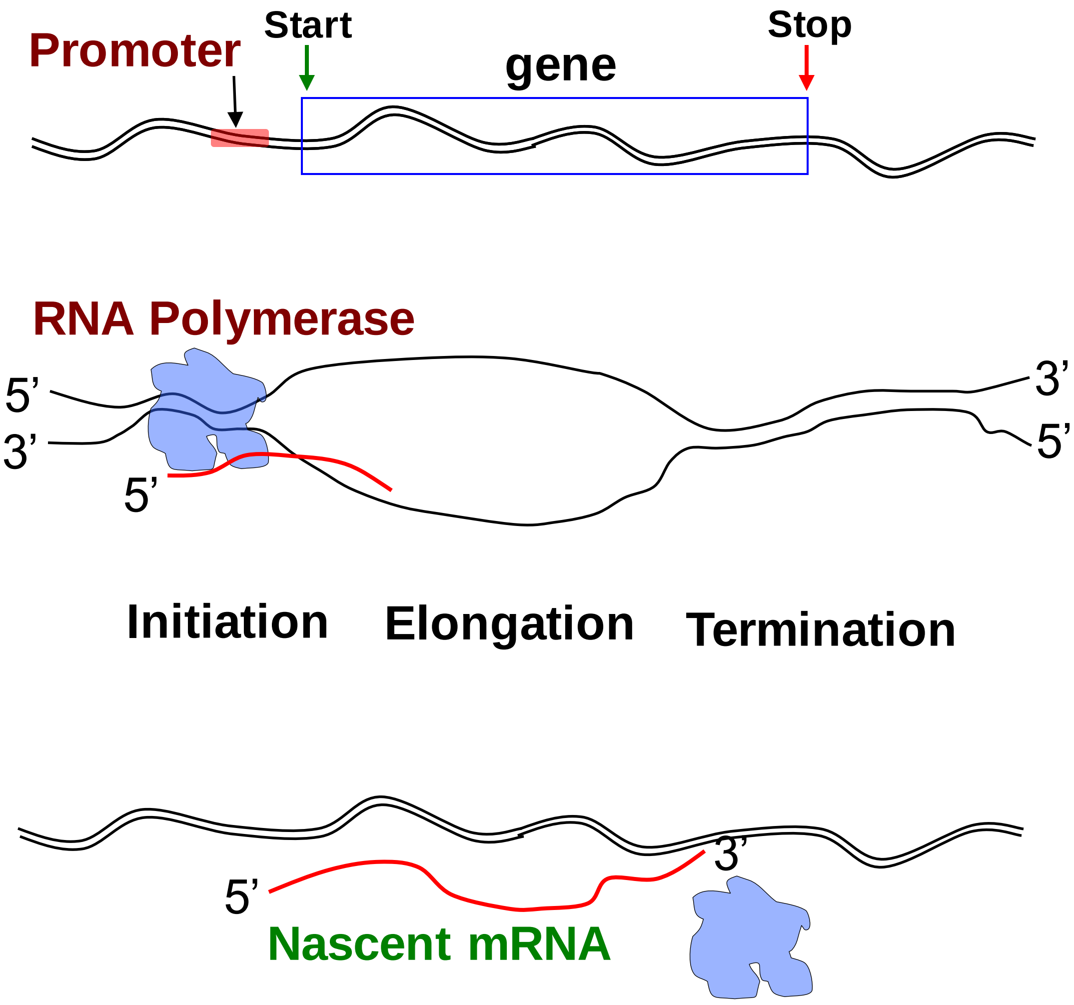
A piece of equipment with several moving parts that uses power to do a particular type of work.

- Cambridge dictionary

Biological machines:

- DNA polymerase
- RNA polymerase
- Ribosome

Transcription: Involved Machineries and Processes



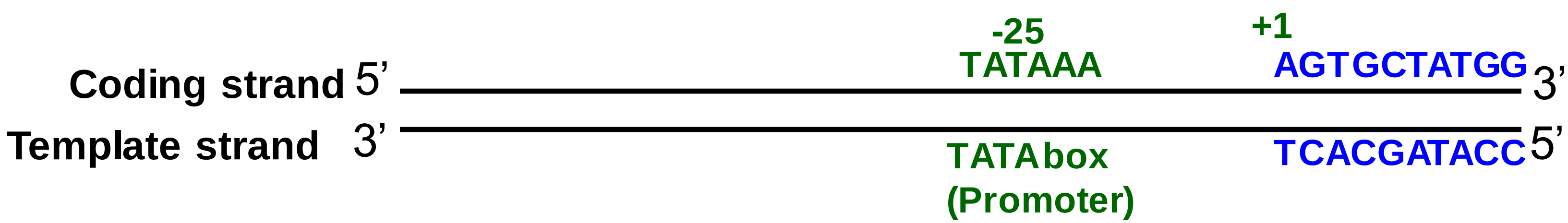
Key points to be discussed

1. **Promoter**
2. **RNA Polymerase**
3. **RNA synthesis**

Promoter for Transcription

Promoter is just like a "pointer" that points to the location of the information (gene) to be copied into mRNA

Prokaryotic promoter



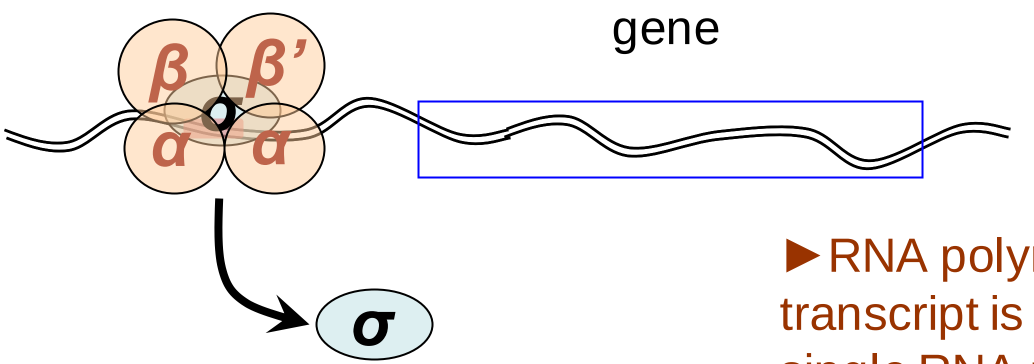
Eukaryotic promoter

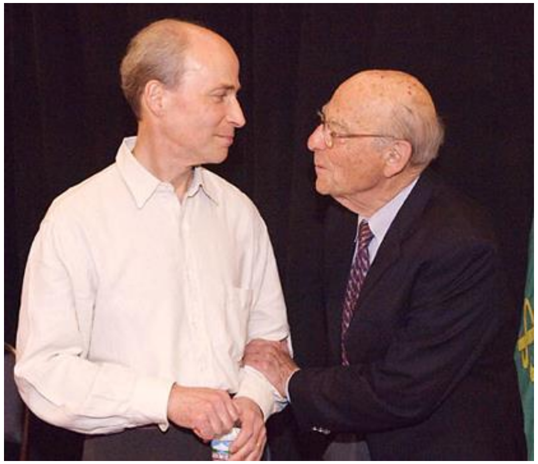
RNA Polymerase

Subunits of RNA Polymerase: ***α, α, β, β'* and *σ***

Holoenzyme: ***α, α, β, β'* and *σ***

Coreenzyme: ***α, α, β and β'***





Roger Kornberg
Nobel Prize in 2006

►RNA polymerase is completely Processive: A transcript is synthesized from start to end by a single RNA polymerase molecule.

►RNA polymerase can initiate the synthesis of RNA *de-novo* (No primer required)

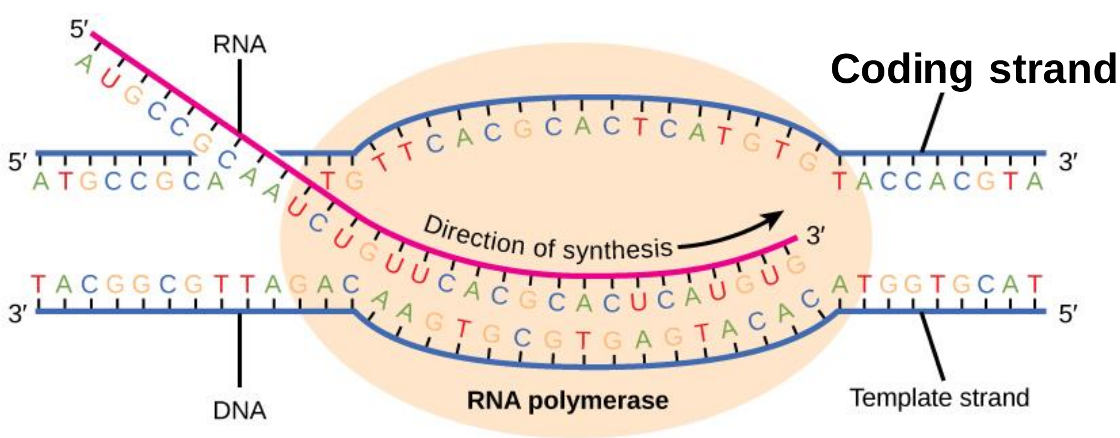
RNA Molecules in *E. coli*	
mRNA	5%
tRNA	15%
rRNA	80%

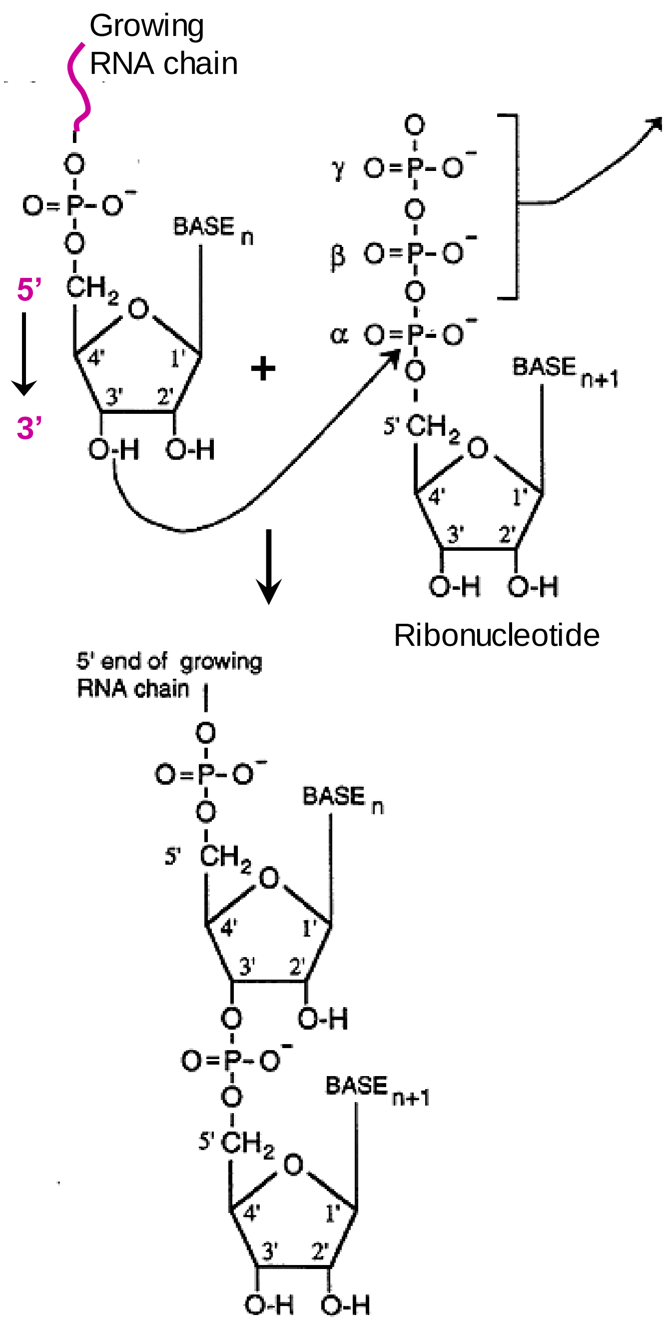
Who transcribes this huge pool of rRNA and tRNA?

In bacteria same RNA polymerase transcribe all these three types of RNA

In eukaryotes different RNA polymerases are involved in transcription of mRNA, rRNA and tRNA

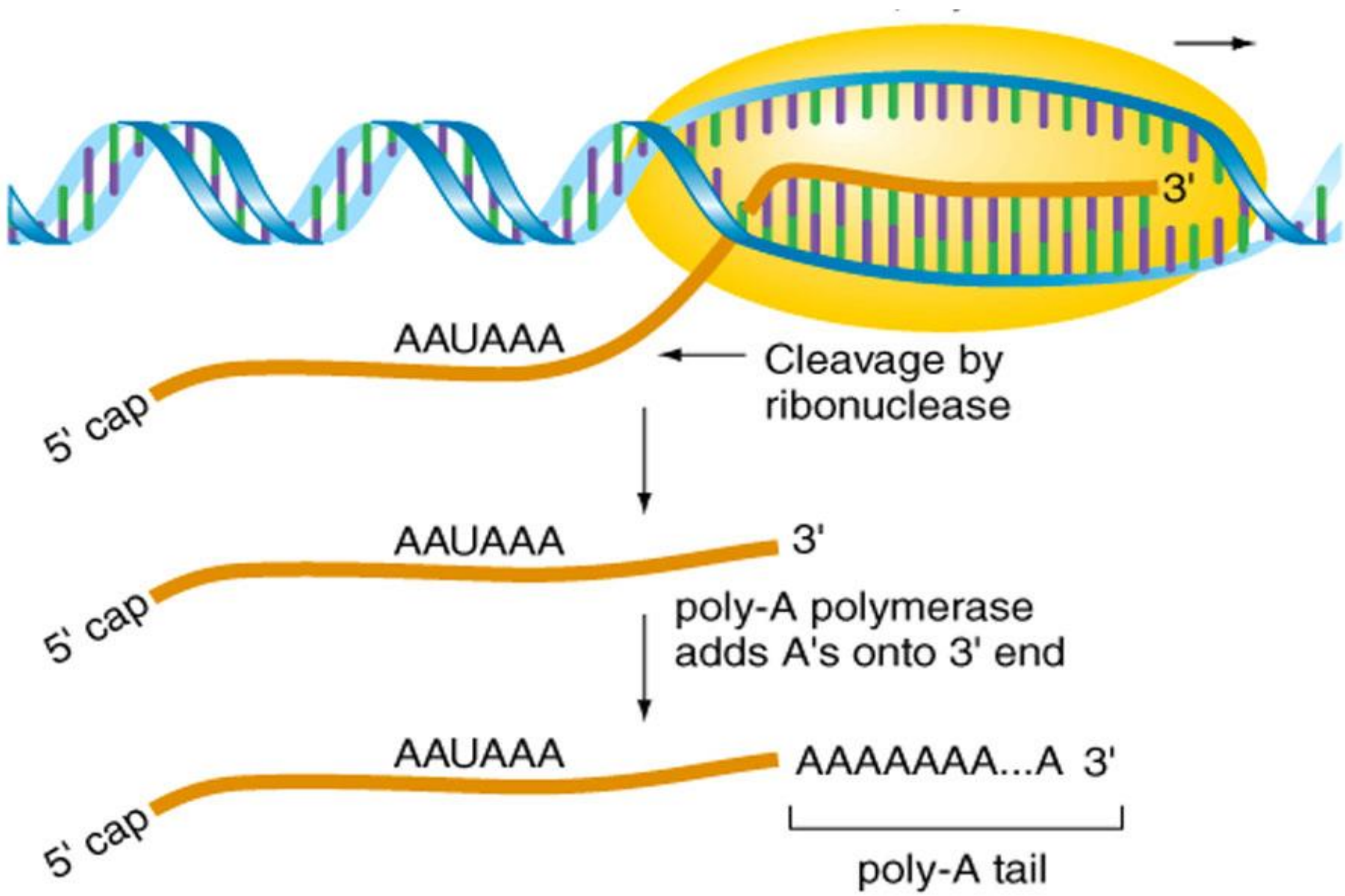
RNA Synthesis





Eukaryotic Transcripts Need to be Processed

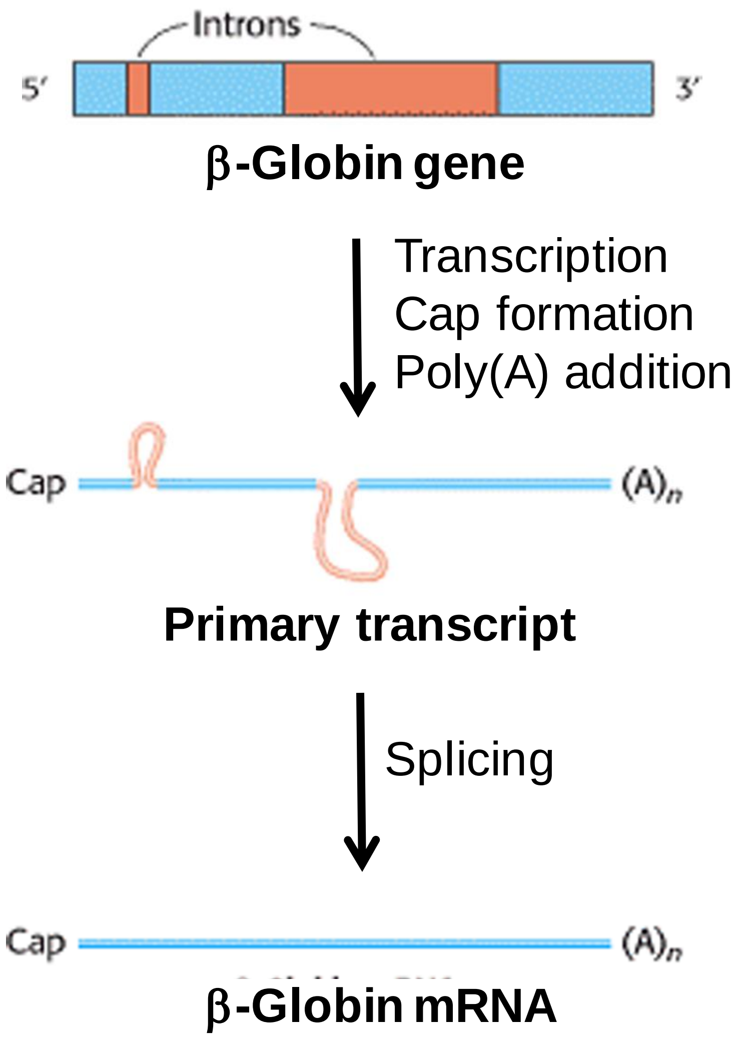
►**Ends of a nascent mRNA acquire a 5' cap and a 3' poly A tail**



- **Increase stability of mRNA**
- **More effective template for translation**

Eukaryotic Transcripts Need to be Processed

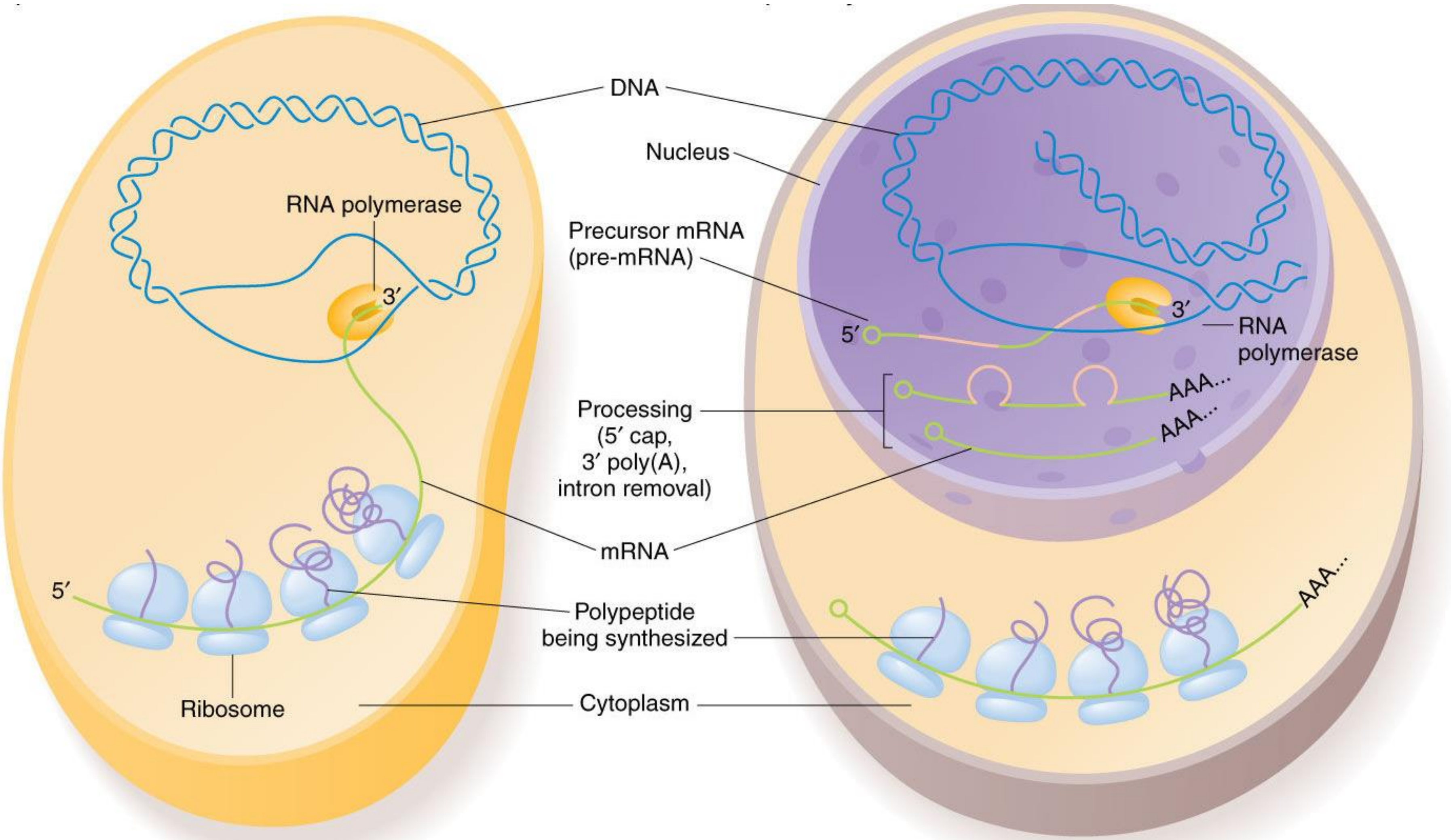
►Splicing (mediated by specialized enzymatic machineries consisting of snRNAs and proteins) removes introns from nascent mRNA



Thomas Cech
[Nobel prize in 1989](#)

****Splicing generates more variation**

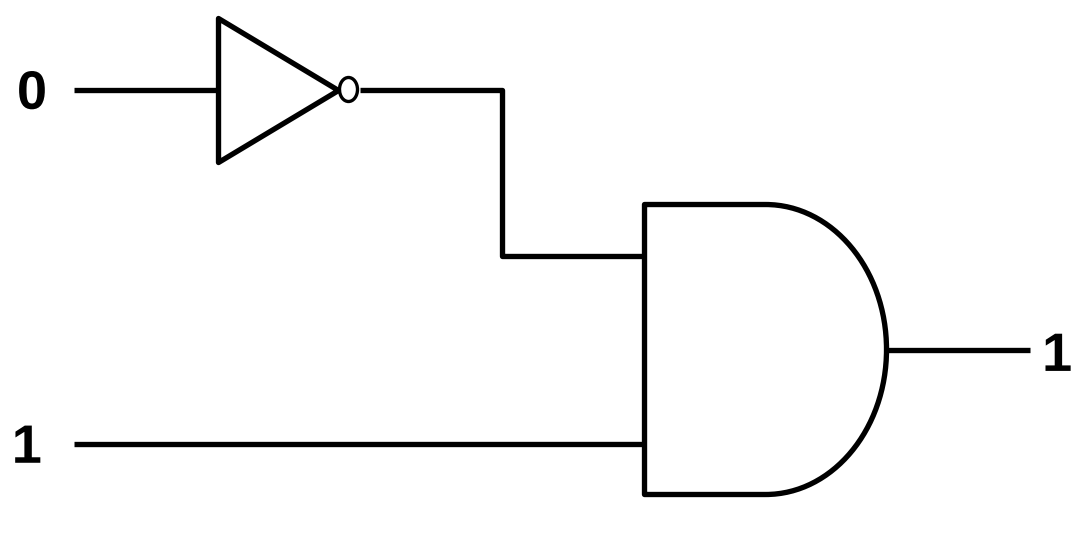
Transcription: At a Glance



Bacteria

Eukaryote

Regulation of Gene Expression
(Biological circuits)



Regulation of Gene Expression

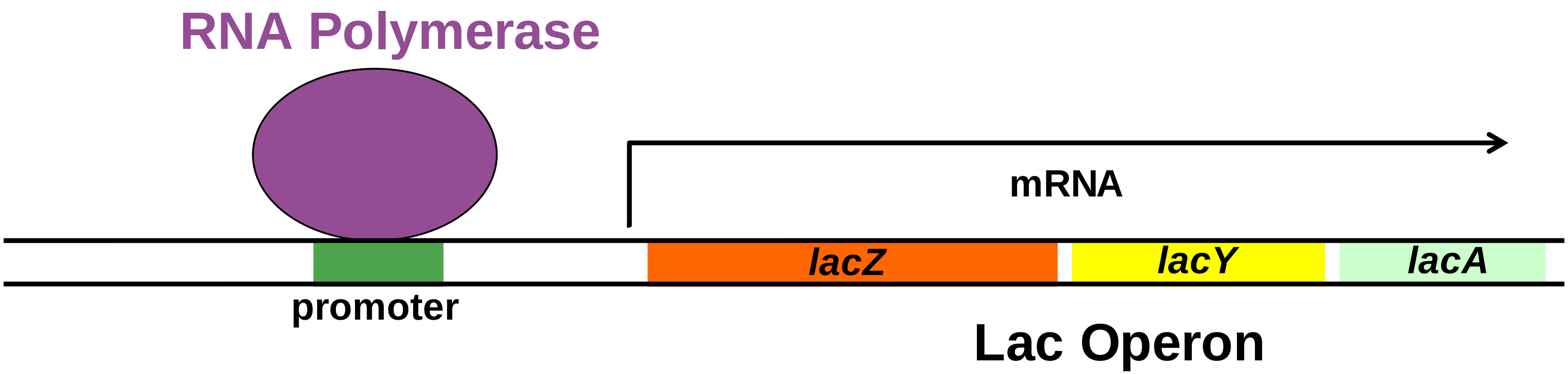
Each cell contains all the genetic material for growth and development

Some of these genes are expressed all the time

Other genes are not expressed all the time. They are switched on an off at need

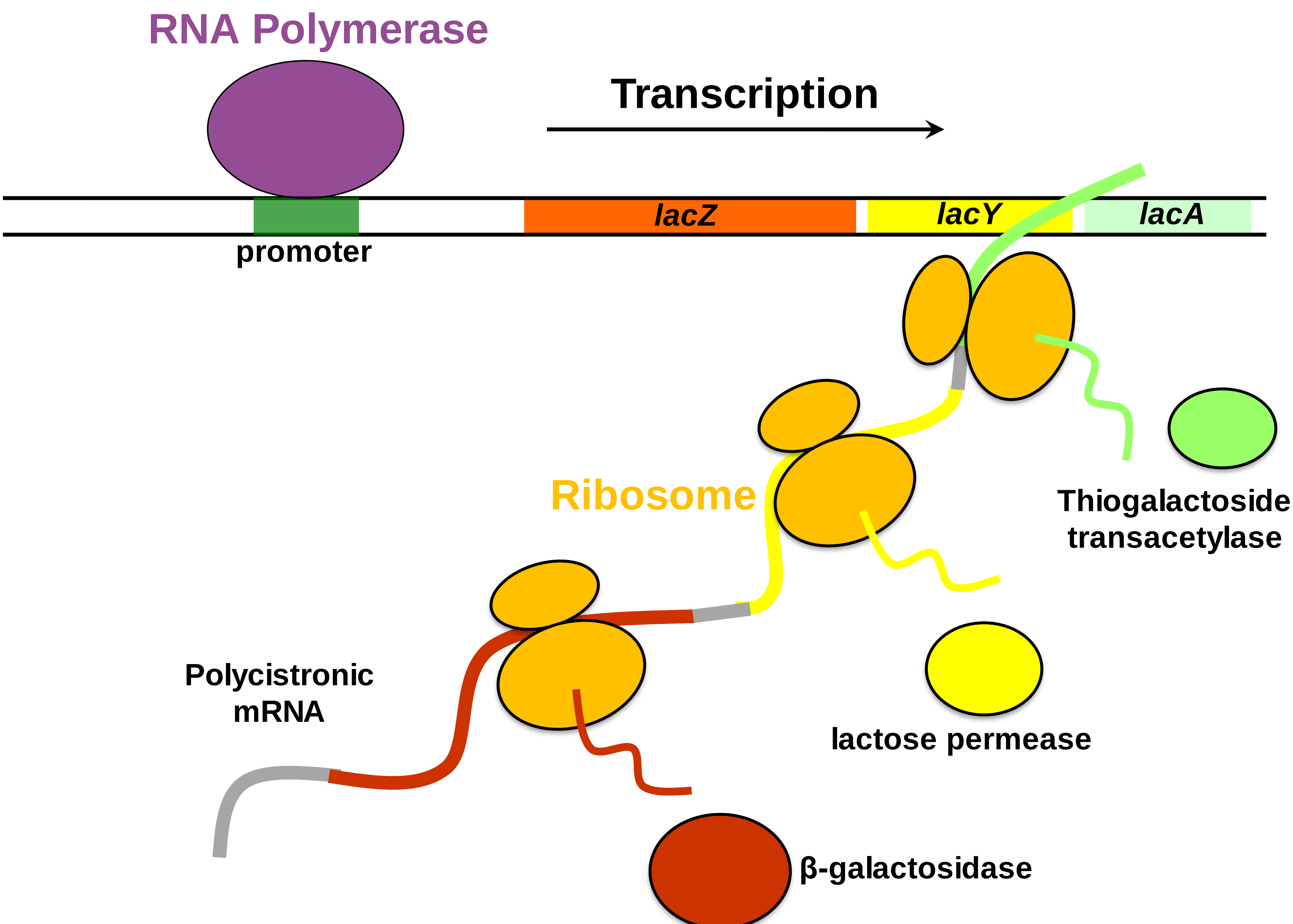
Lac Operon: A Classic Example of Bacterial Gene Expression Control

Operon: Cluster of genes, related by function, regulated by a single promoter and transcribed into one mRNA (polycistronic).

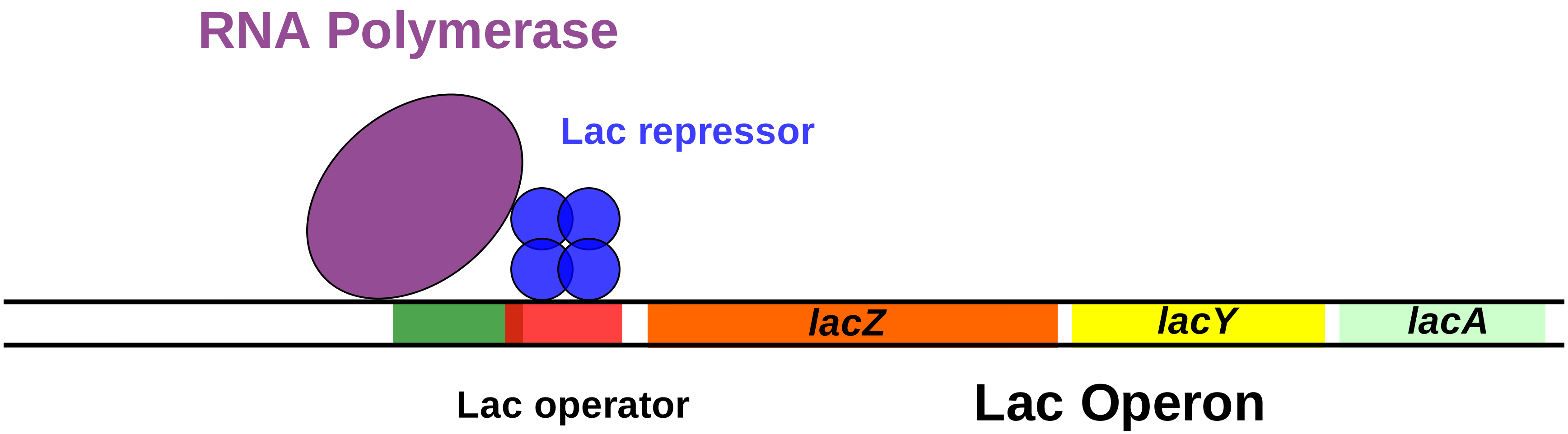


lacZ	β-galactosidase	Breaks lactose into galactose and glucose.
lacY	lactose permease	Imports lactose into the bacterial cell.
lacA	thiogalactoside transacetylase	Cell detoxification.

Functional Outcome of Lac Operon

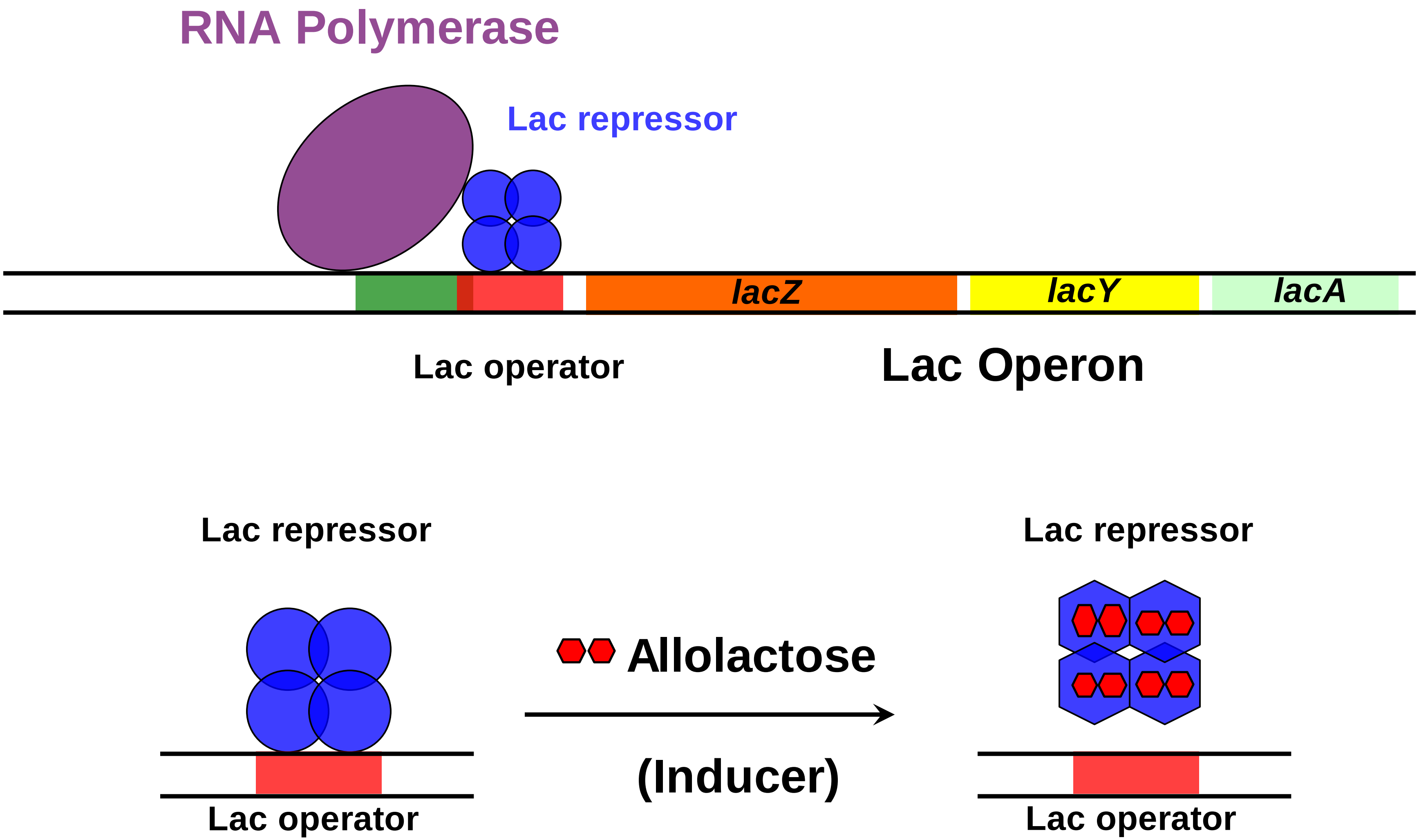


Lac repressor is a negative regulator of the Lac operon



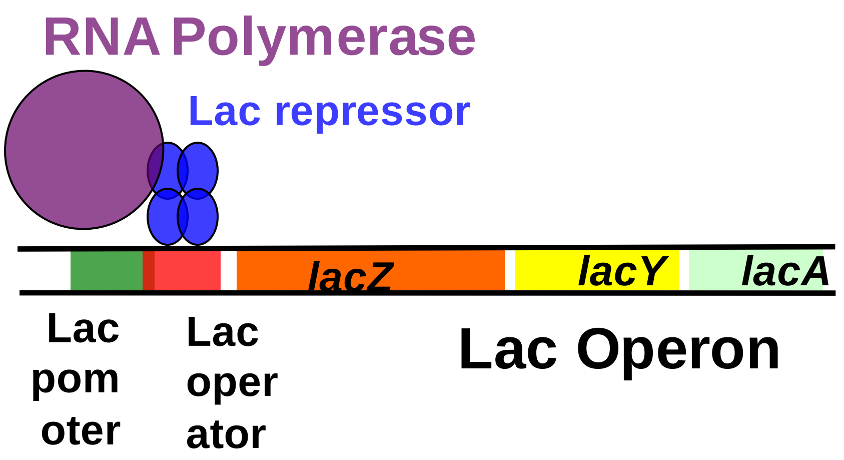
lacZ	β-galactosidase	Breaks lactose into galactose and glucose.
lacY	lactose permease	Imports lactose into the bacterial cell.
lacA	thiogalactoside transacetylase	Cell detoxification.

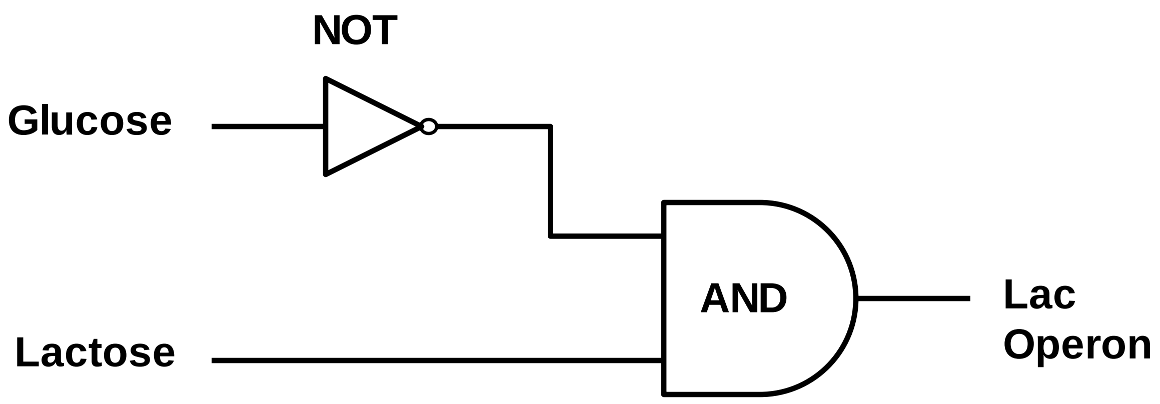
Lactose (Allolactose) Can Displace Lac Repressor From the Operator Site



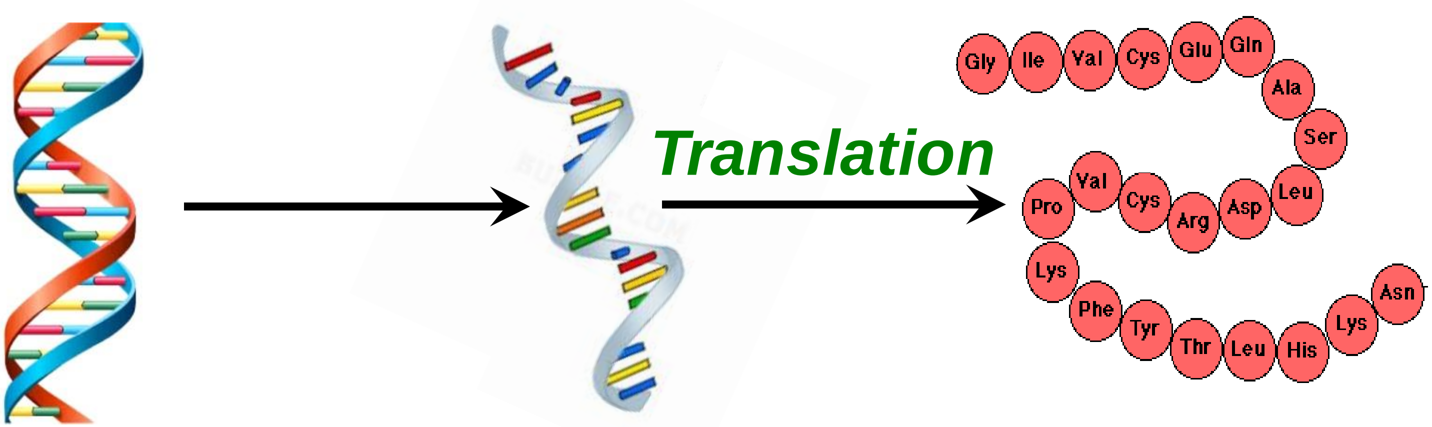
Four Possible Situations

Glucose	Lactose	Lac repressor bound	Lac operon
1	0	YES	OFF (0)
1	1	YES	OFF (0)
0	1	NO	ON (1)
0	0	YES	OFF (0)



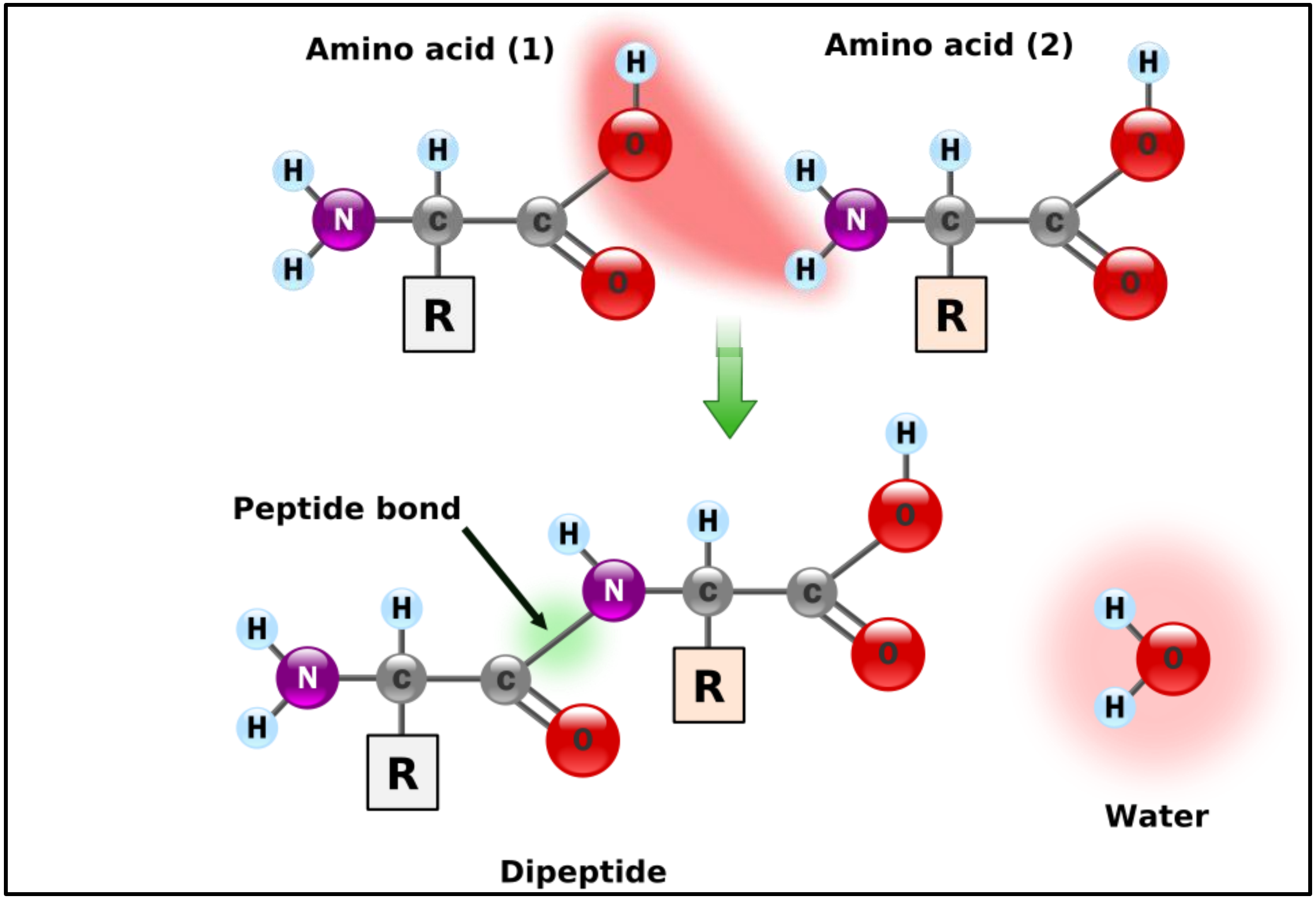


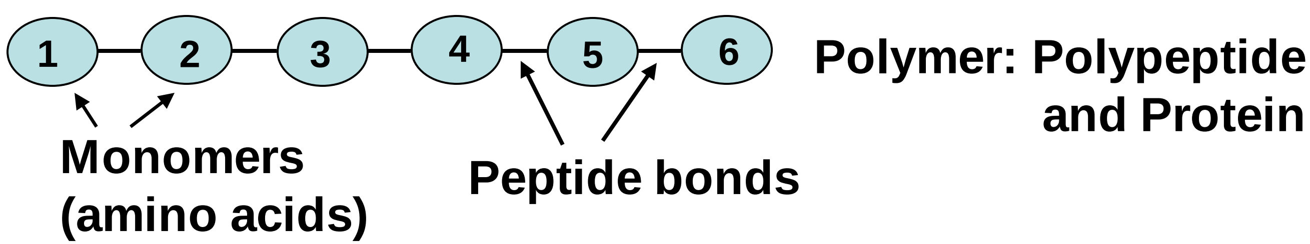
Translation



DNA	RNA	Protein
Polymer of nucleotides	Polymer of nucleotides	Polymer of amino acids

How Amino Acids are Linked Together





Translation

Template for protein synthesis

5’ —— 3’ mRNA

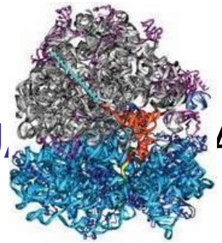
5’UAAGGAGA AUCGUCAUGAAGAGGCCC…….UAAUUA 3’
(RBS)

Start codon

Stop codon

Met–Lys–Arg–Pro……
Polypeptide

Nascent polypeptide



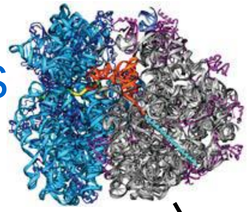
5’U… A AUCGUCAUGAAGAGGCCC…….UAAUUA 3’

► **In Eukaryotes, 5’ 7mG cap is recognized by ribosome**

Protein synthesizing machinery

<u>Ribosome</u>: (rRNA + proteins)

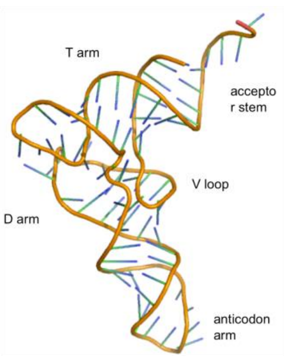
30S



50S

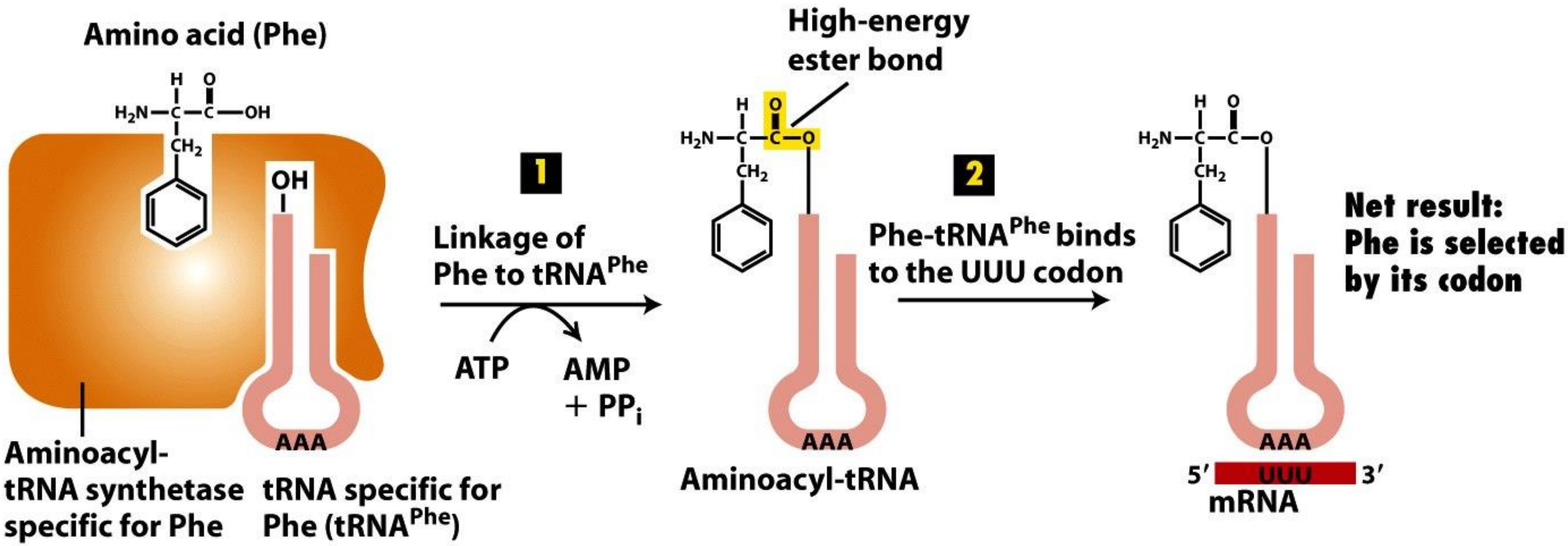
16S rRNA
21 proteins

23S rRNA
5S rRNA
34 proteins



tRNA

How Correct Amino Acids are Selected During Protein Synthesis



Genetic code

Genetic code is the relation between the sequence of bases in DNA (or its RNA transcripts) and the sequence of amino acids in proteins

A codon is a set of 3 nucleotides that specifies a particular amino acid

Why three nucleotides?

64 Codons present. Three of them (UAA, UAG, UGA) can't code any amino acids, called STOP codons

AUG serves as the "initiator" or "start codon, which starts the synthesis of a protein

We have 61 codons that code for amino acids, and we have 20 amino acids. So, multiple codons may specify a single amino acid

Khorana, Nirenberg, Holley

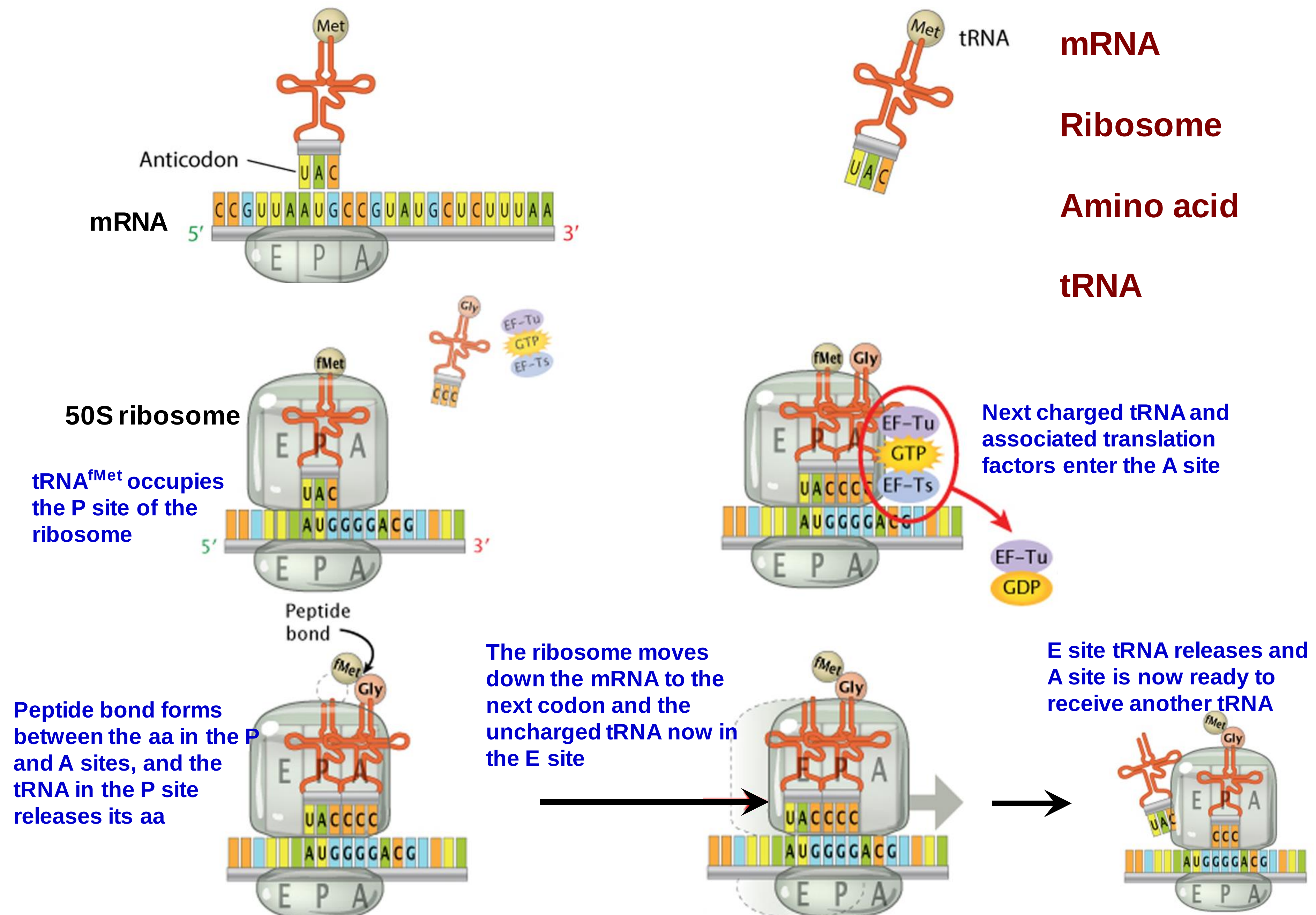
Nobel Prize in 1968

Genetic code

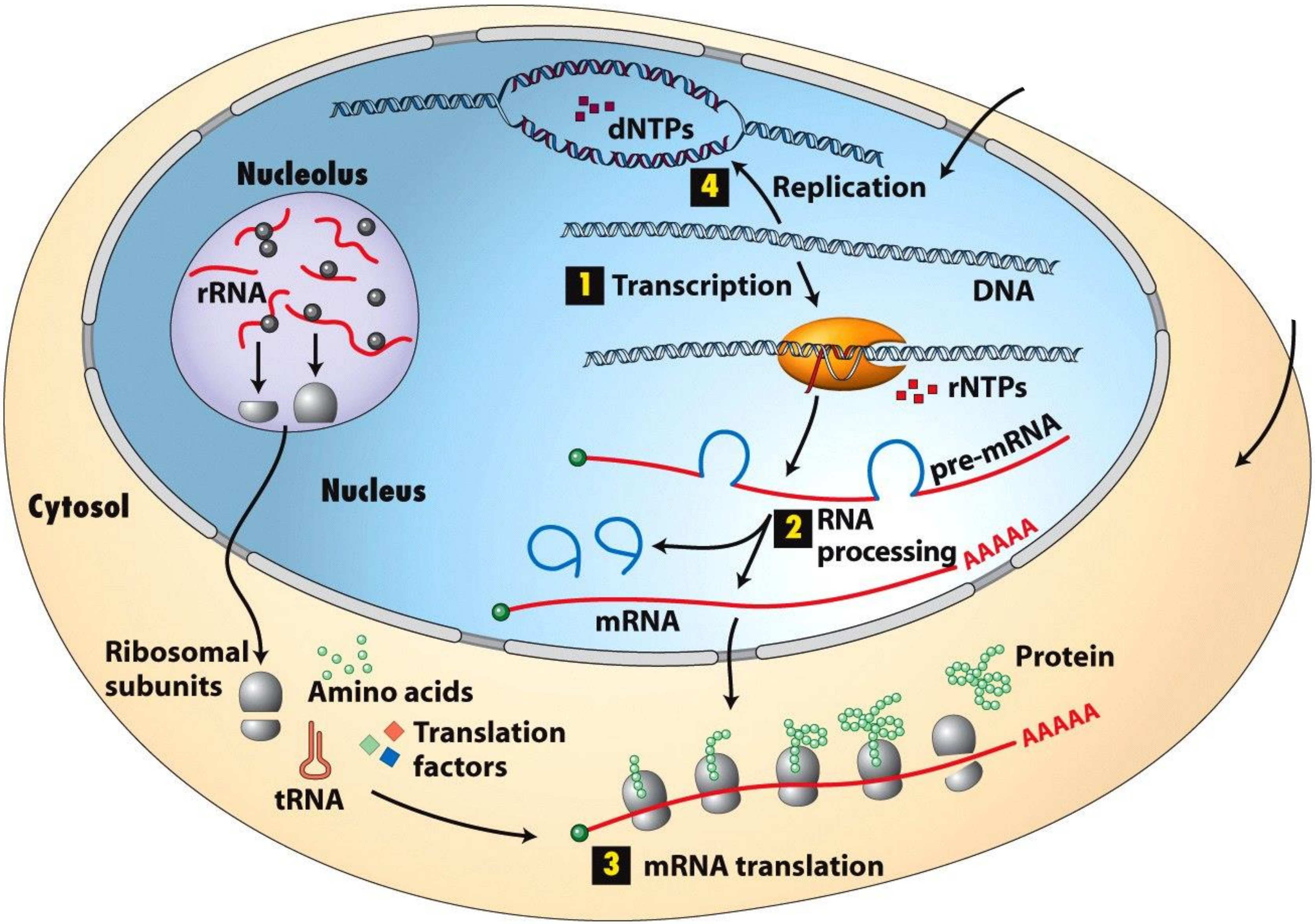
Second Letter

1st letter	U	C	A	G	3rd letter
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
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 Ile AUC Ile AUA Ile 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 Gly GGC Gly GGA Gly GGG Gly	U C A G

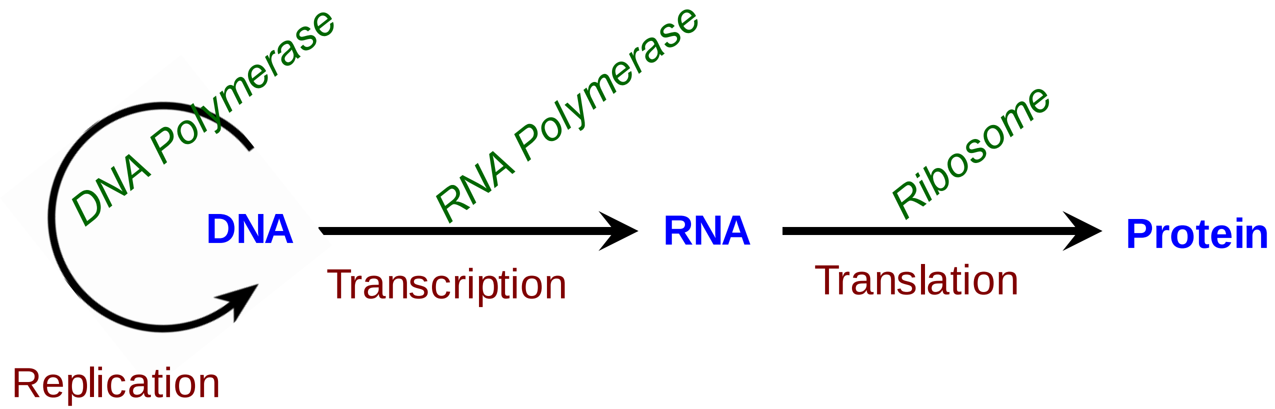
Translation: Involved Machineries and Processes

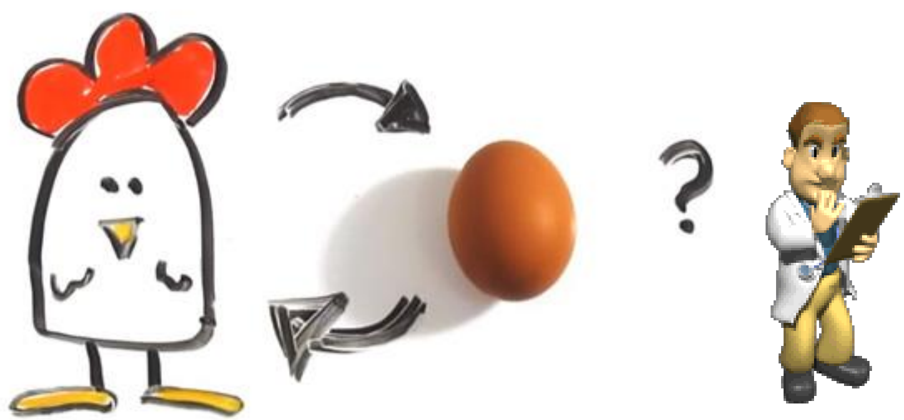


Translation: At a Glance



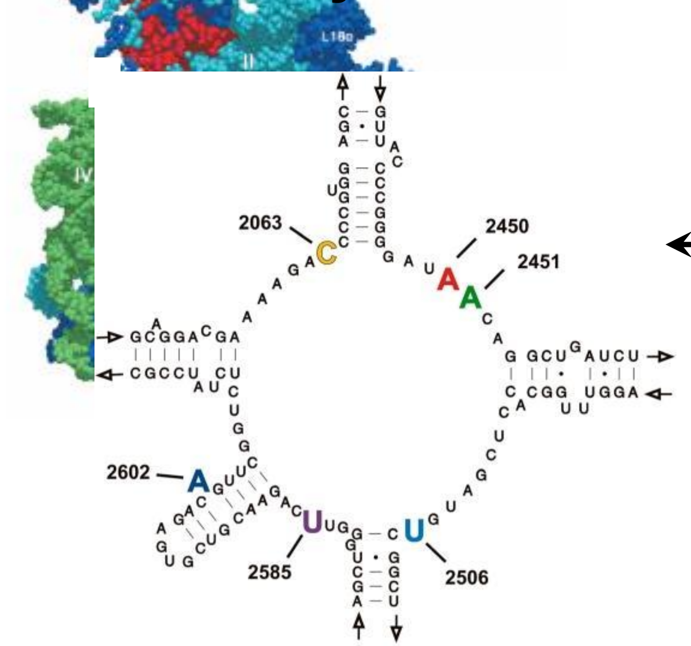
Which Came First? Nucleic acids or Proteins

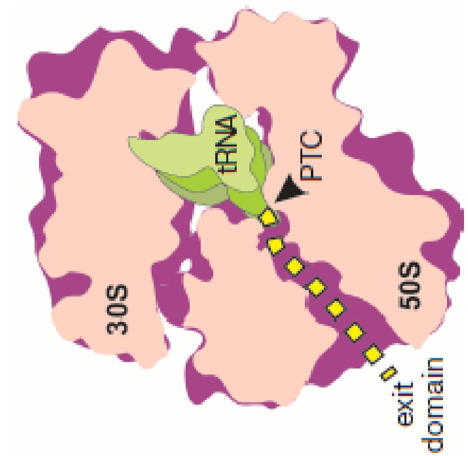


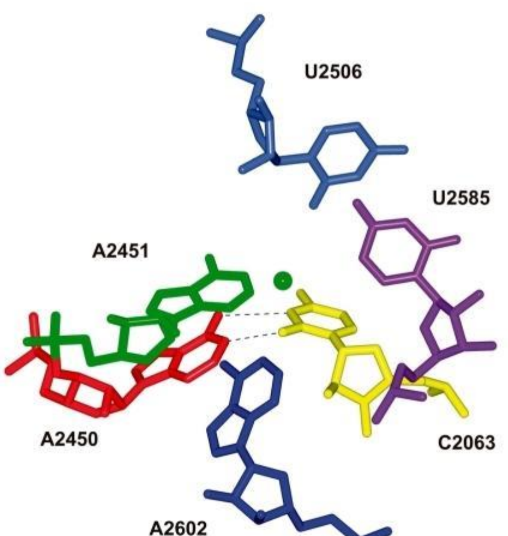


► **RNA has enzymatic activity**

What Happens Inside the Ribosome? Chemical and Physical Consideration of Protein Synthesis

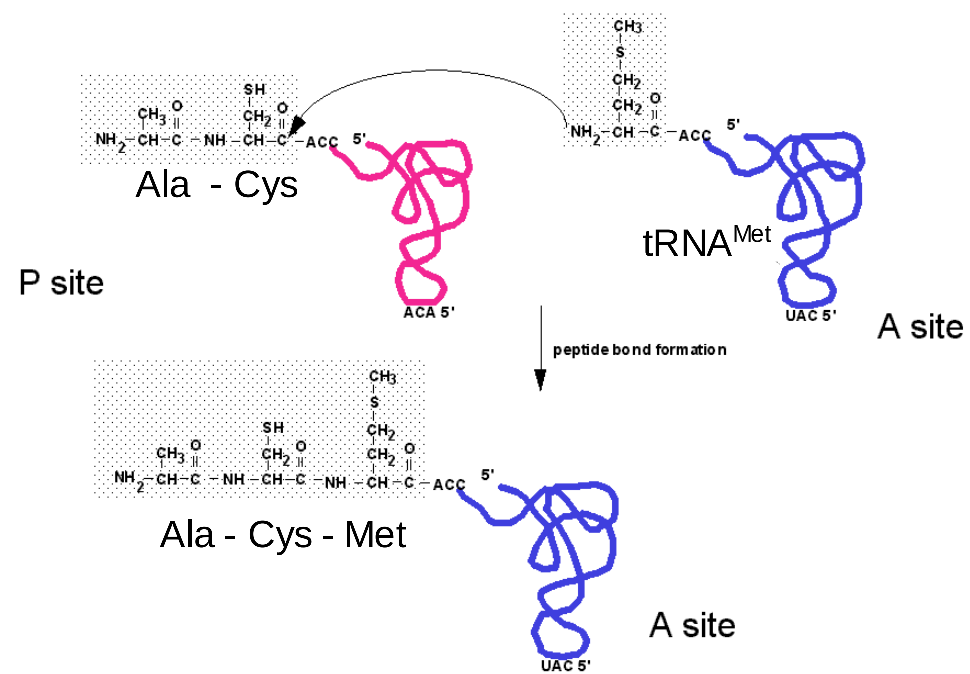




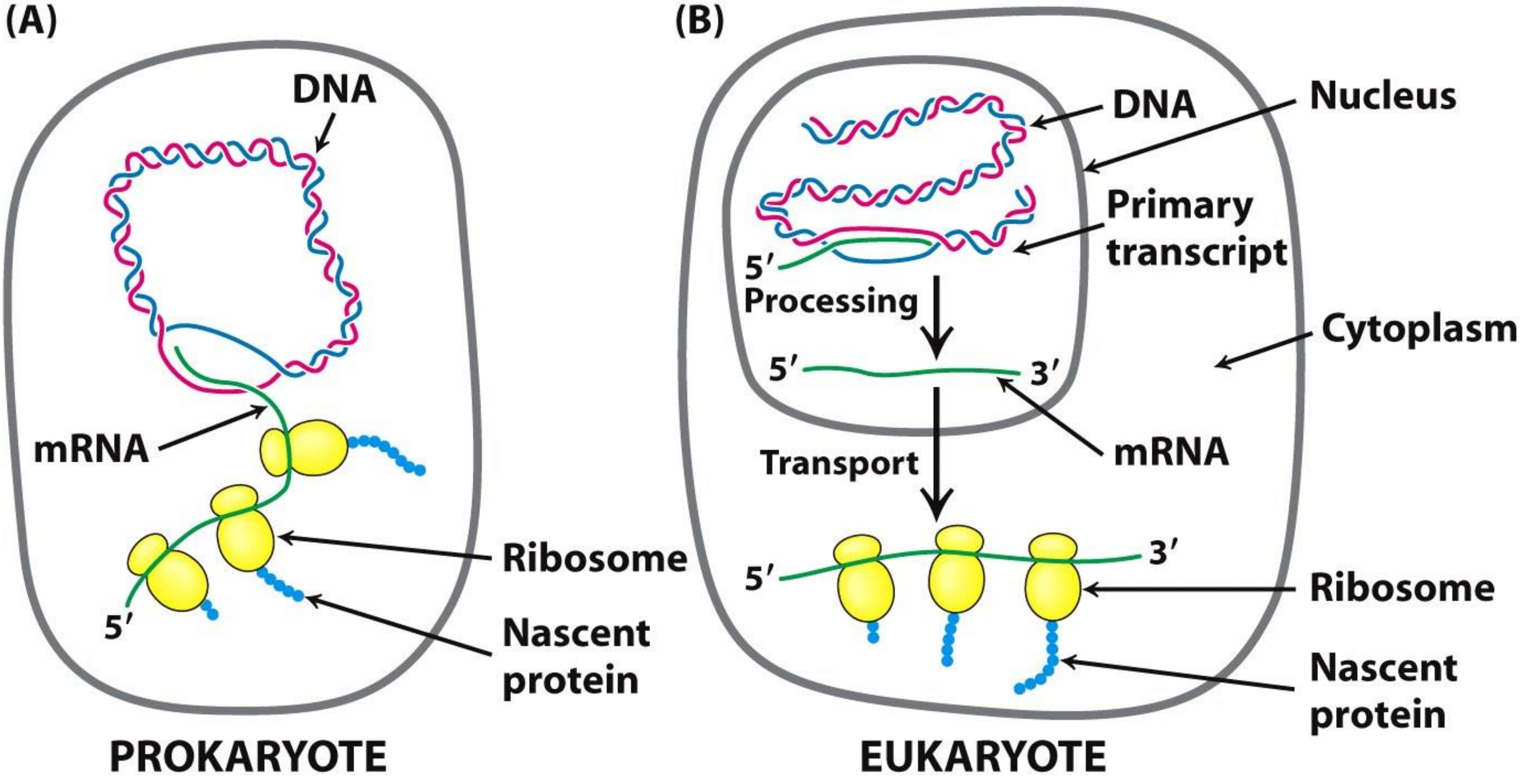


Peptidyl transferase center

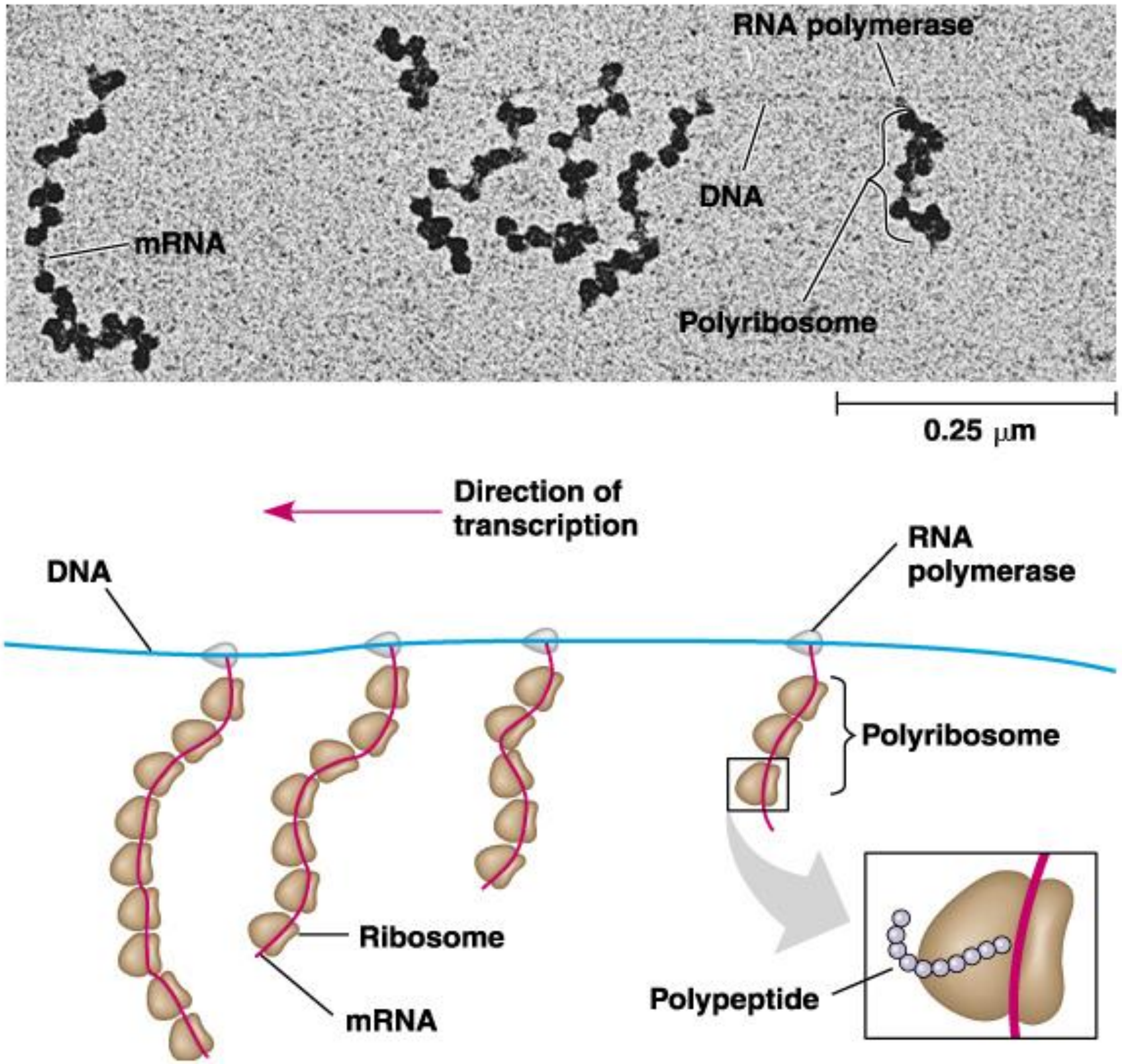
Peptidyl transferase reaction



Time, Space and Correlation between Transcription and Translation



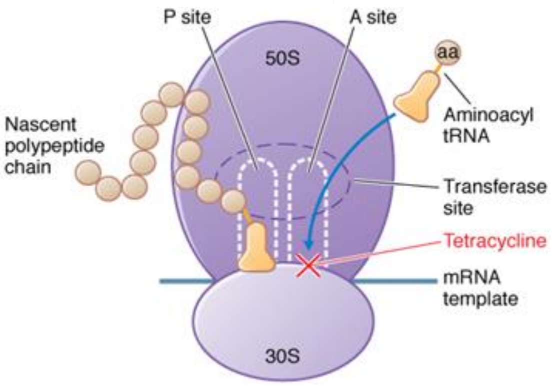
Time, Space and Correlation between Transcription and Translation



Translation Machineries: Attractive Targets For Therapeutics

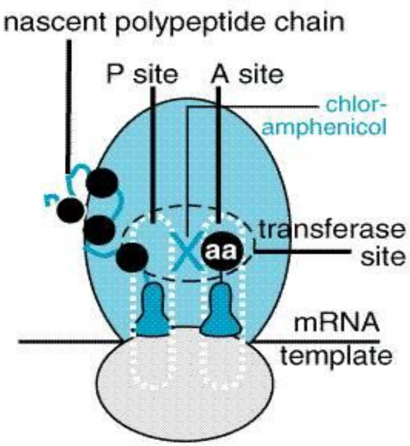
Tetracycline

Binds to the 30S ribosome, and blocks binding of aminoacyl-tRNA to the A-site



Chloramphenicol

Blocks the peptidyl transferase reaction on 50S ribosomes

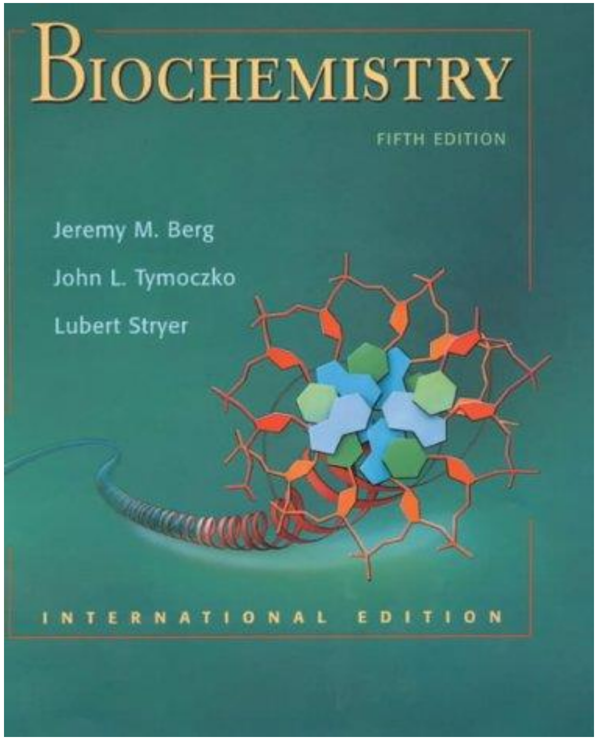


Streptomycin Binds to the 30S ribosome, prevents the transition from initiation to chain-elongation

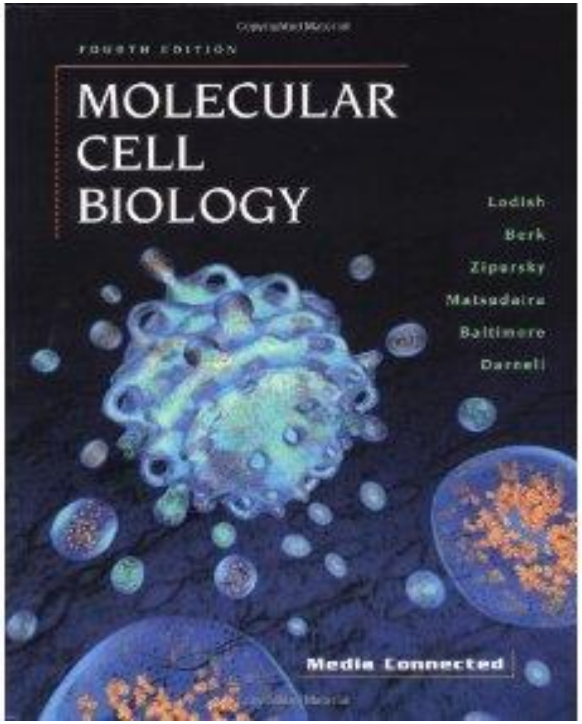
Erythromycin Binds to the 50S ribosome, and blocks the translocation

Why ribosome is an attractive target for the development of antibiotics?

Suggested Textbook…



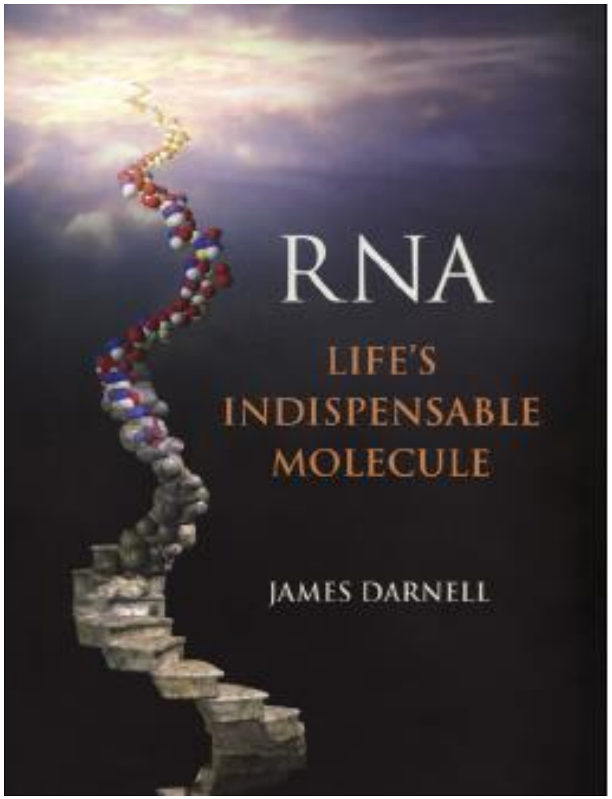
Stryer…



Baltimore, Lodish..

Extra Resources

Further Reading...



James Darnell

Videos…

mRNA synthesis (Transcription)
https://www.youtube.com/watch?v=_C9Un4dIpR4

Protein synthesis (Translation)
https://www.youtube.com/watch?v=Ikq9AcBcohA

Overview
https://www.youtube.com/watch?v=gG7uCskUOrA

Science of Living System (BS20001)

- **SOUMYA DE**
- **Office:** 2nd Floor, Diamond Jubilee Building
- **Phone:** 03222-2284552
- **E-mail:** somde@iitkgp.ac.in
- **Website:** http://iitkgpbioscience.weebly.com/soumya-de.html

- **Research interests:**

  Biophysics
  Nuclear Magnetic Resonance (NMR) Spectroscopy
  Protein Engineering
  Signal Transduction and Gene Expression
  Enzymology

Central Dogma of molecular biology

DNA: Storage Medium

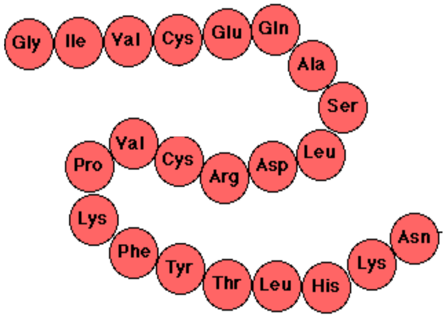
Polymer of nucleotides

CCTGAGCCAACTATTGATGAA

RNA: Transmission Medium

Polymer of nucleotides

CCUGAGCCAACUAUUGAUGAA



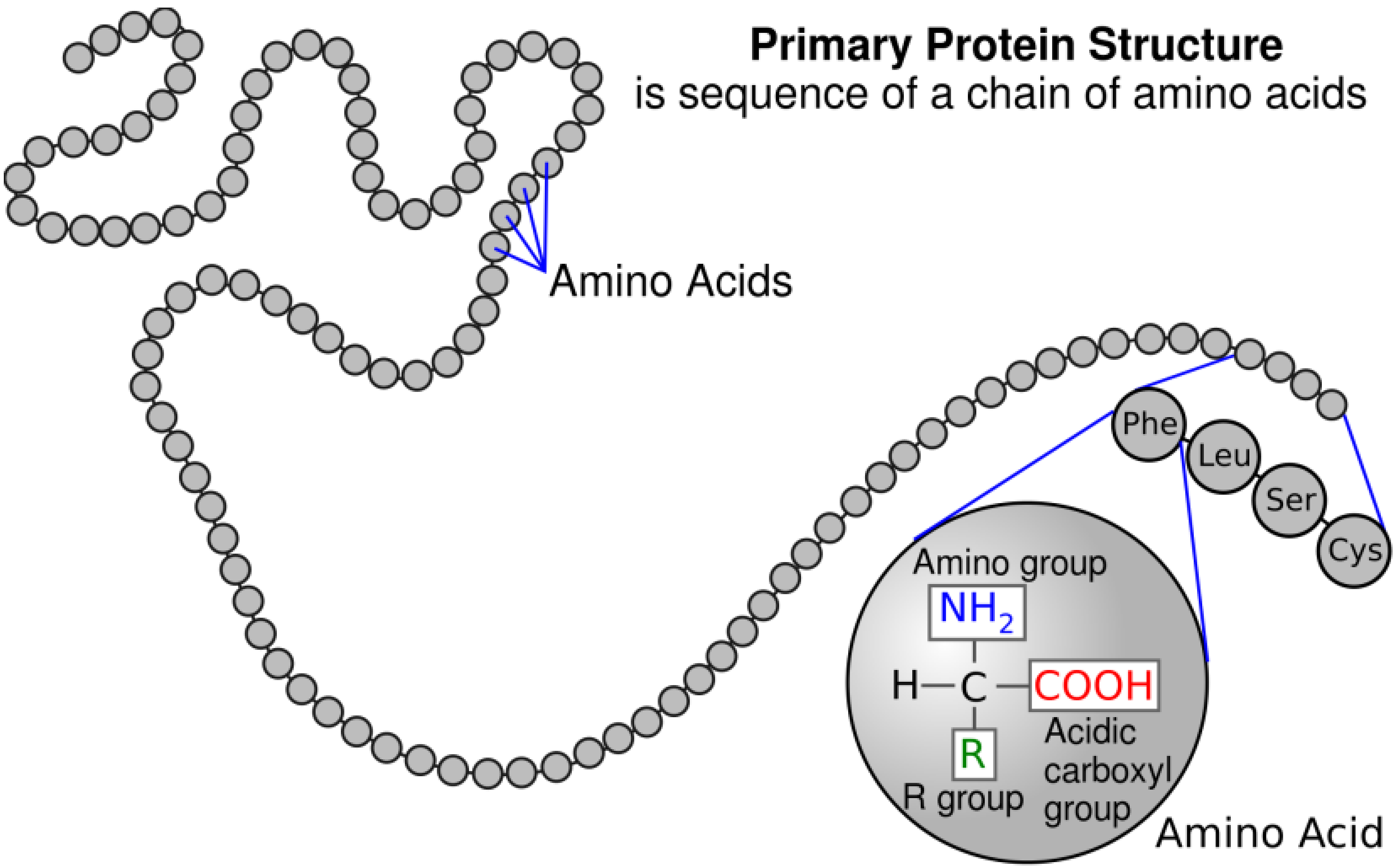
Protein: Molecular Machines

Polymer of amino acids

PEPTIDE

Primary Structure of Proteins

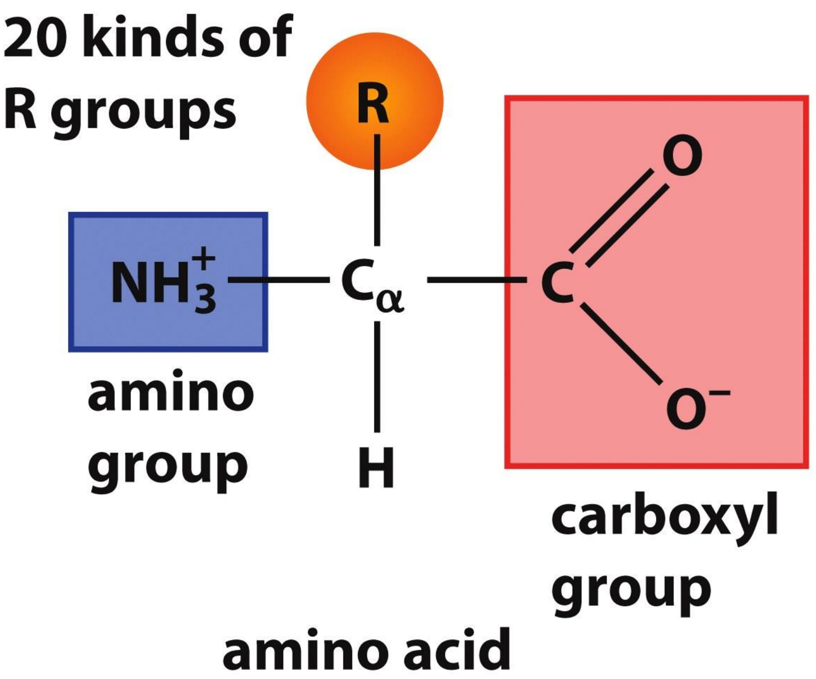
The primary structure of a protein is its amino acid sequence



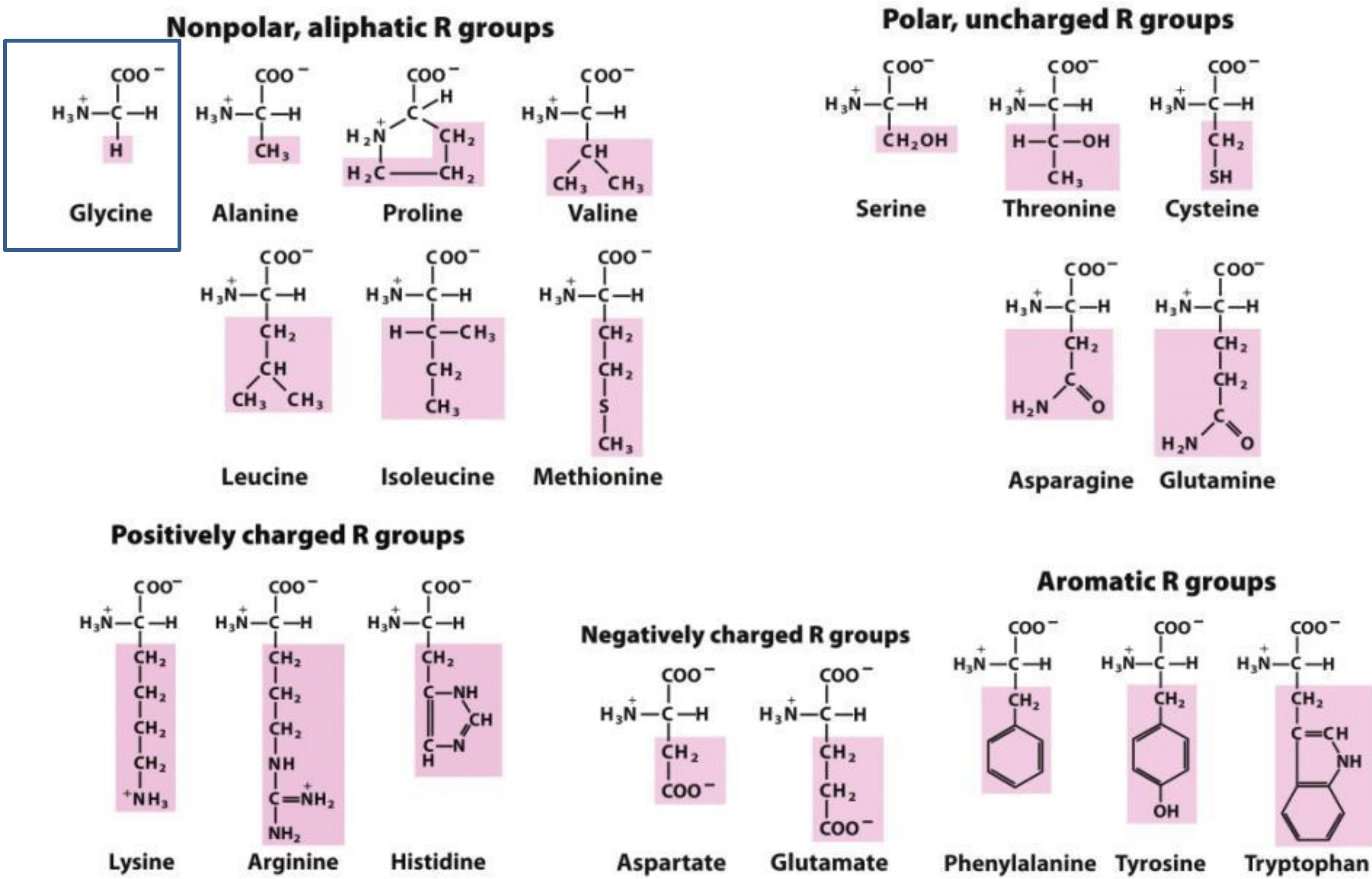
Amino acids: Building blocks of Proteins

- Protein is a polymer of amino acids.
- There are 20 common amino acids.
- Amino acids have a common chemical structure - A tetrahedral sp^3 carbon (C_α) with four different functional groups:

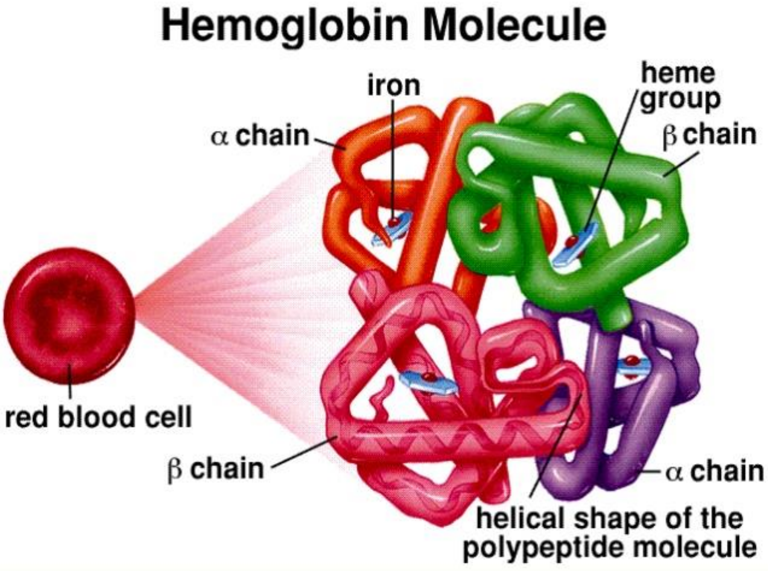
1. Amino group
2. Carboxyl group
3. H-atom
4. Side chain (R) with distinct chemical property

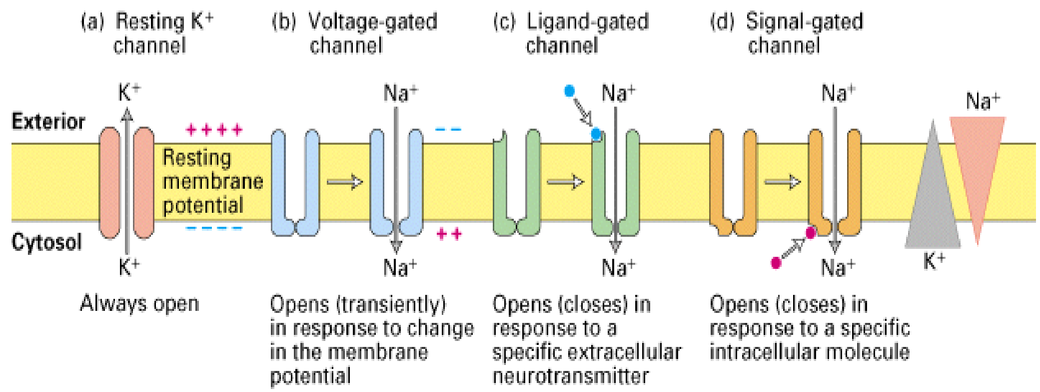


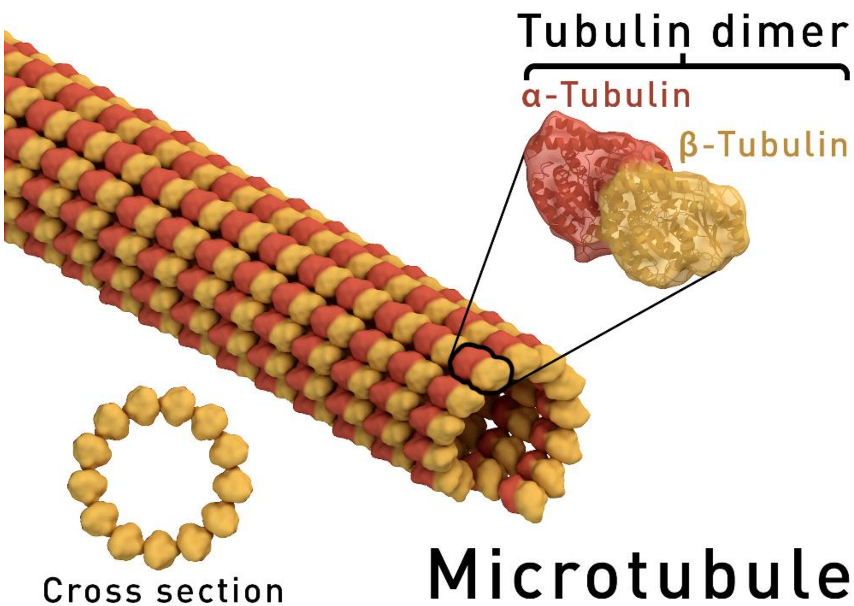
The 20 Common Amino Acids of Proteins

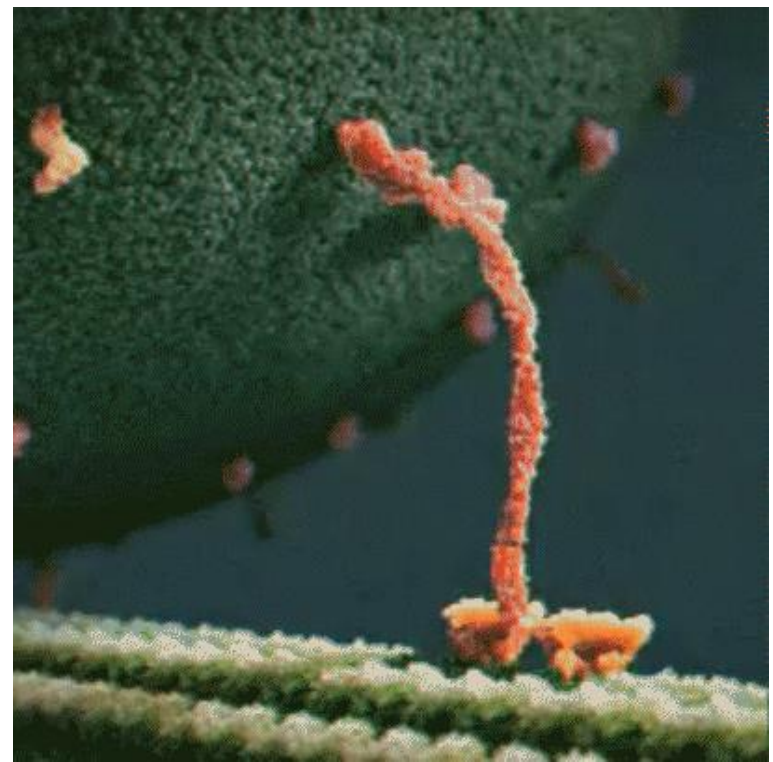


Proteins come in various shapes and sizes



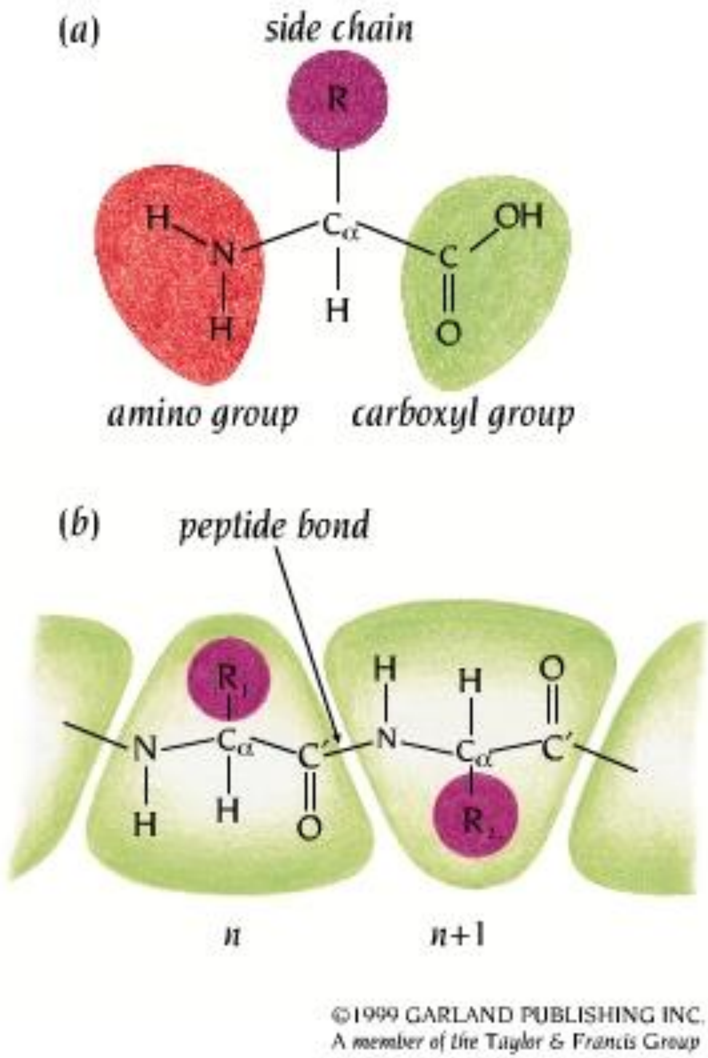




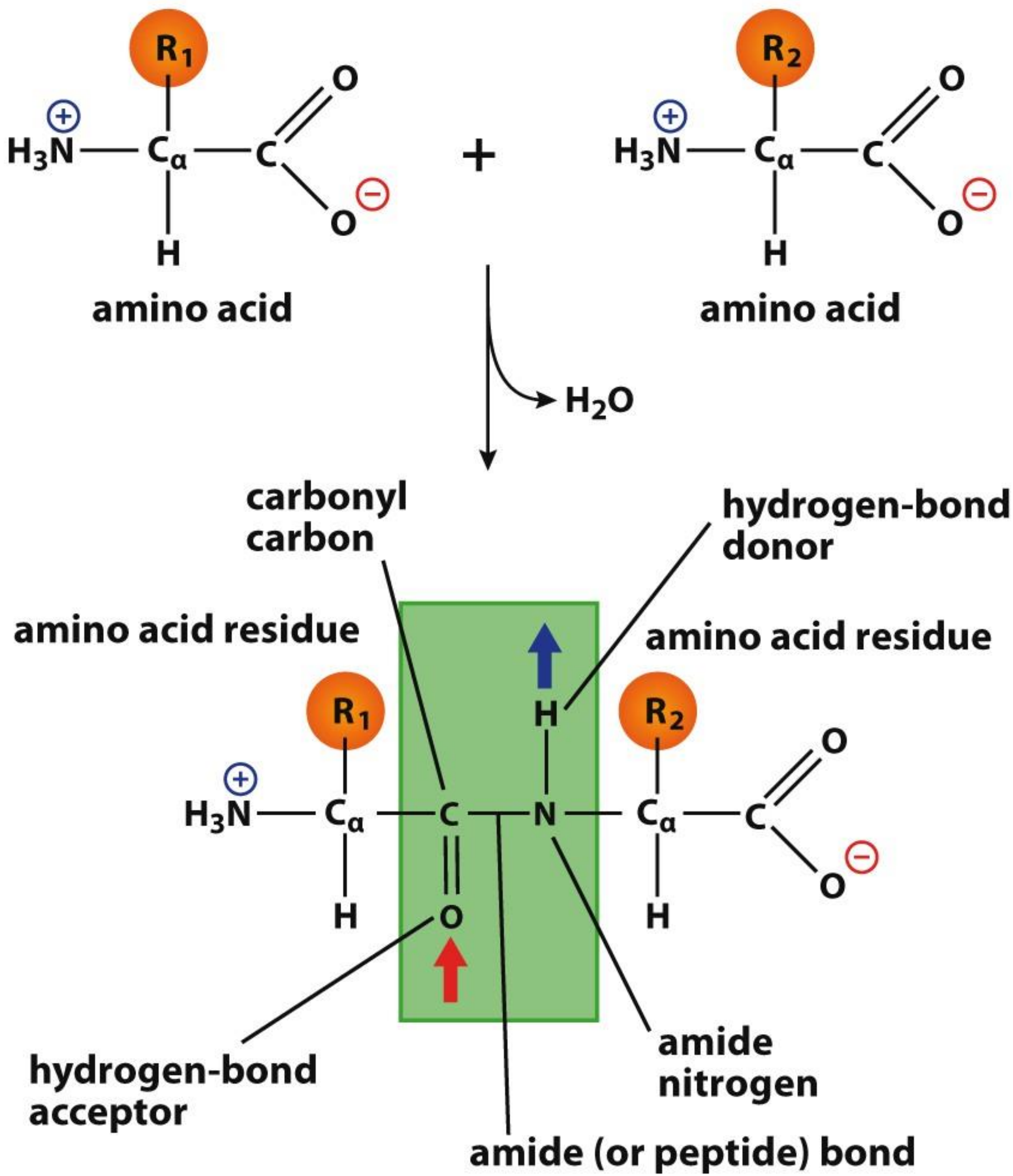


Proteins are polypeptide chains

Successive polypeptide bonds: main chain or backbone



Formation of the peptide bond



The amide plane: partial double bond character of the peptide bond

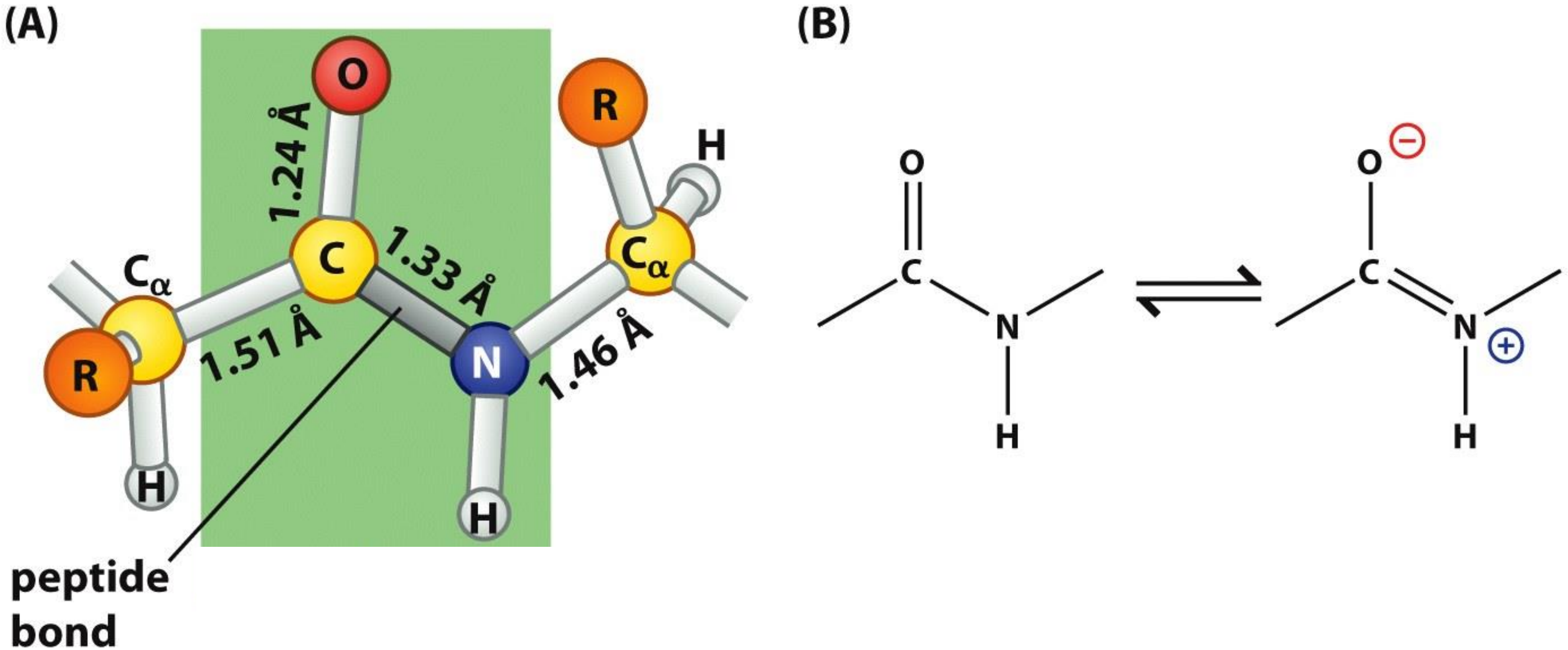
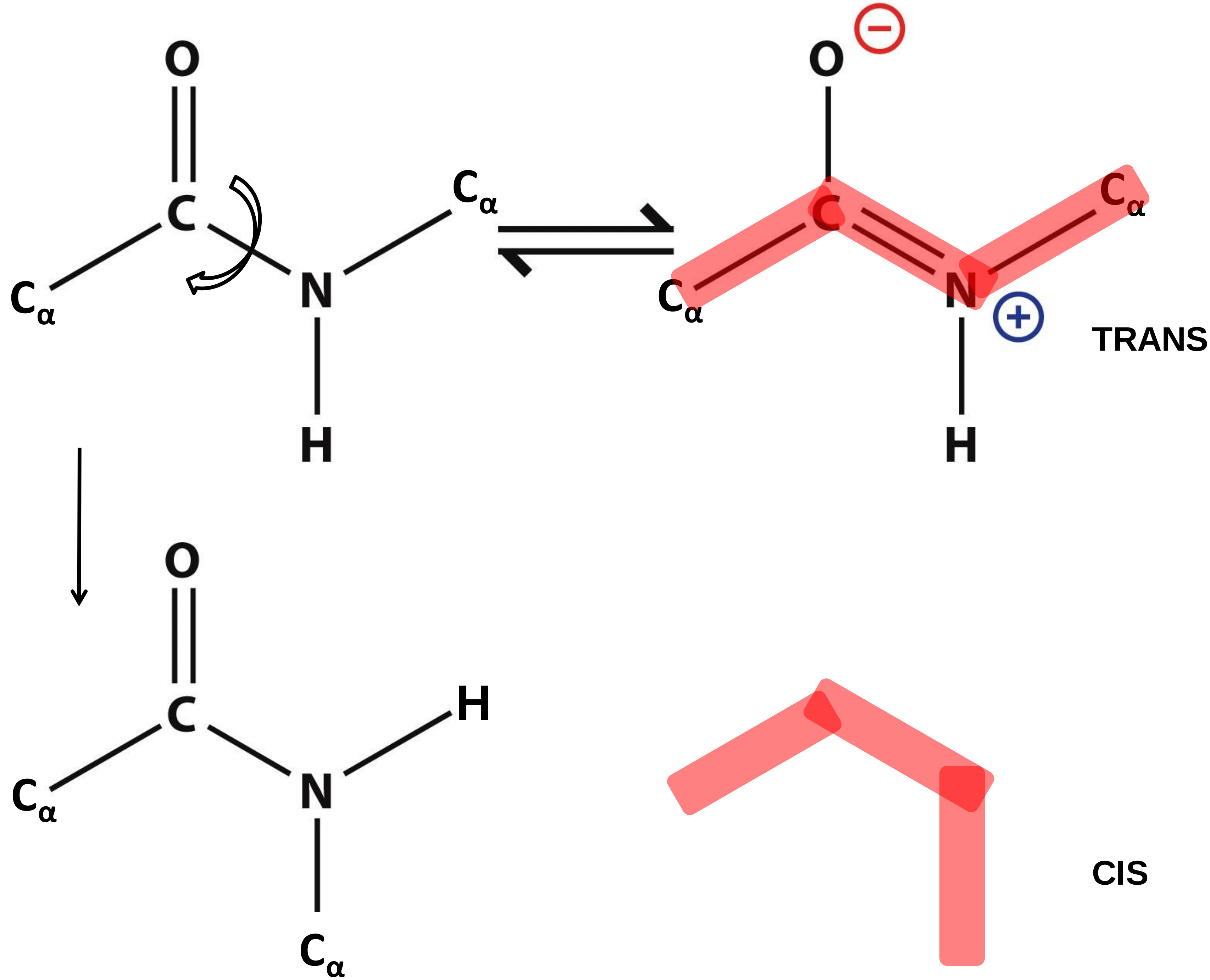


Figure 4.14 The Molecules of Life (© Garland Science 2013)


TRANS
CIS

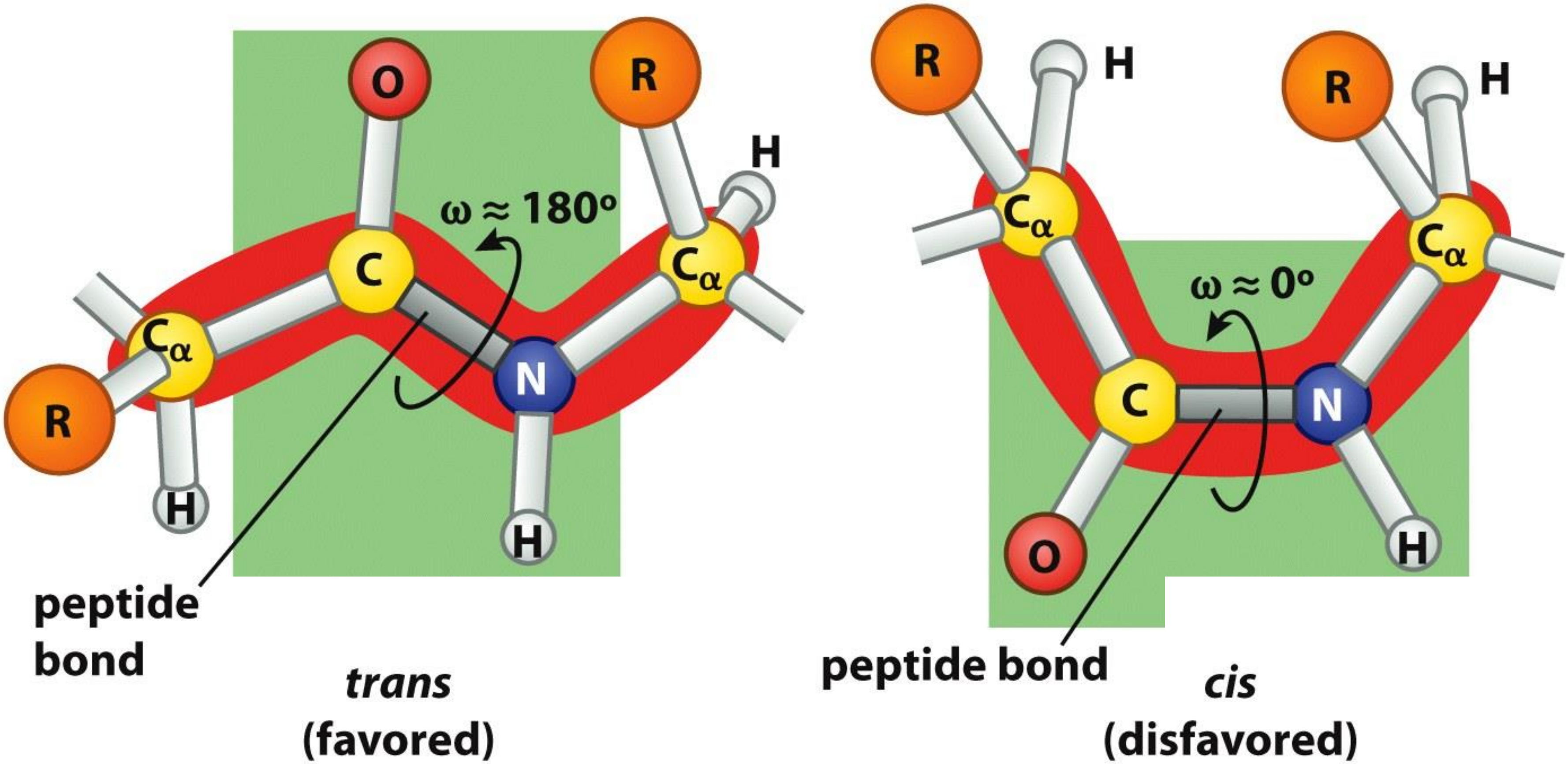
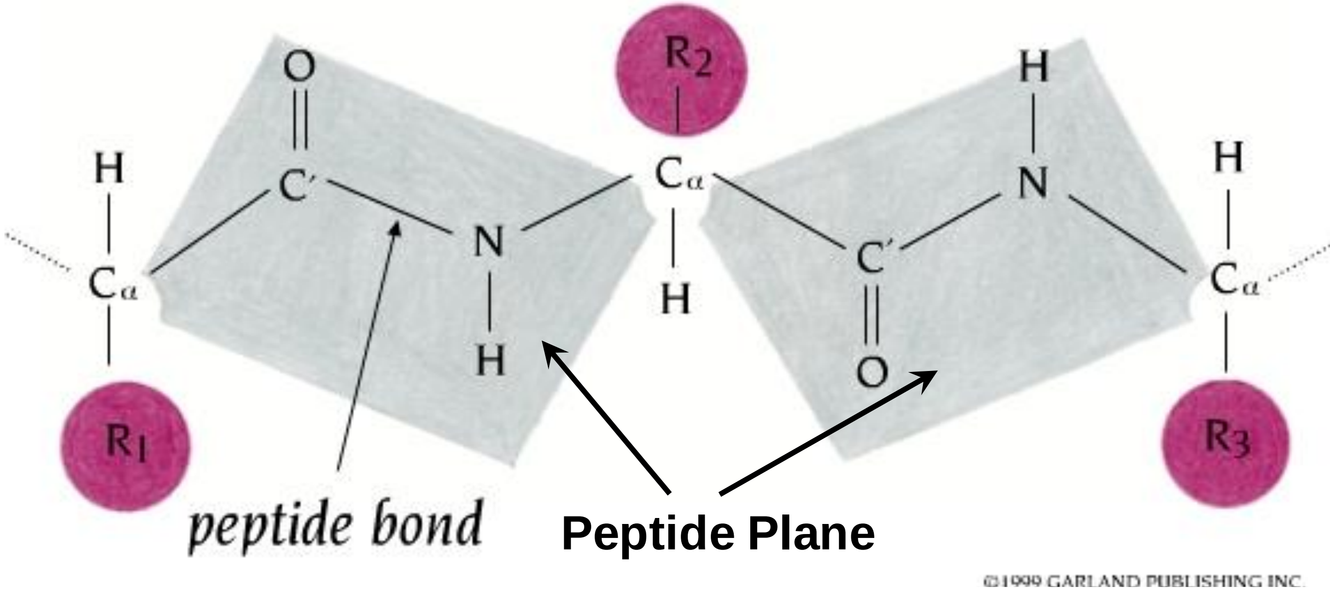
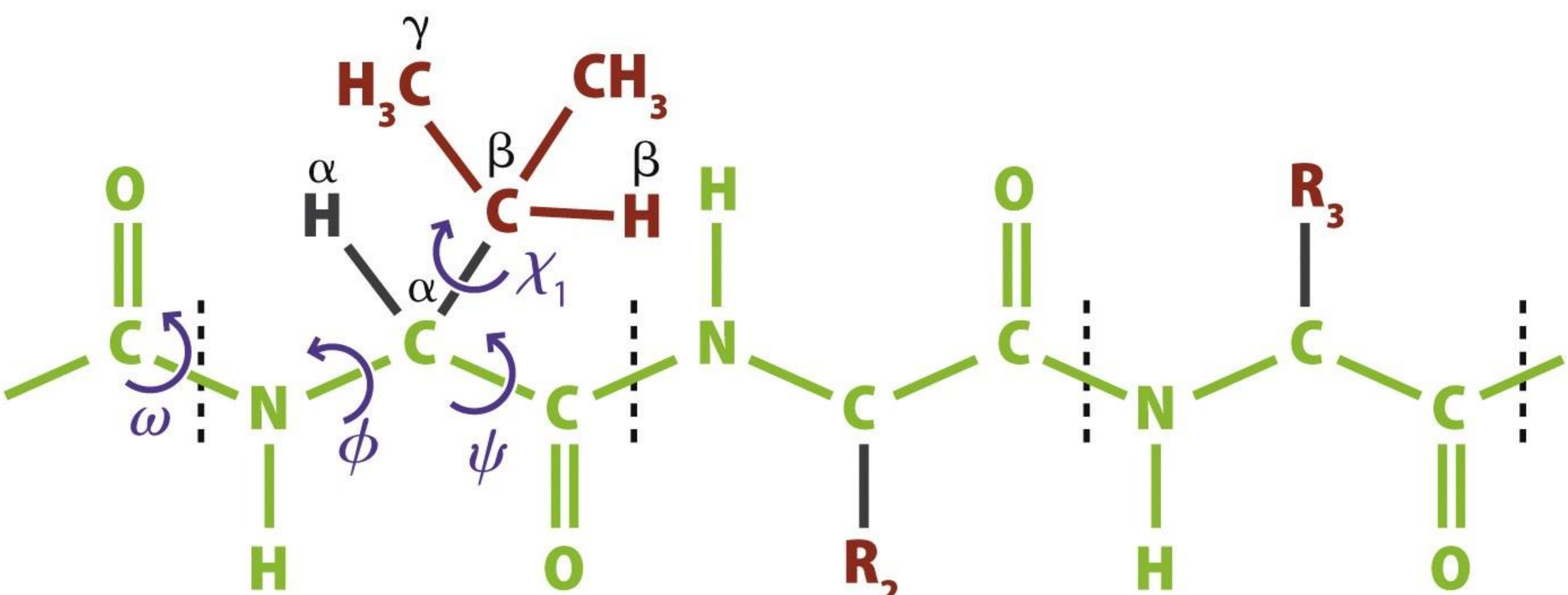


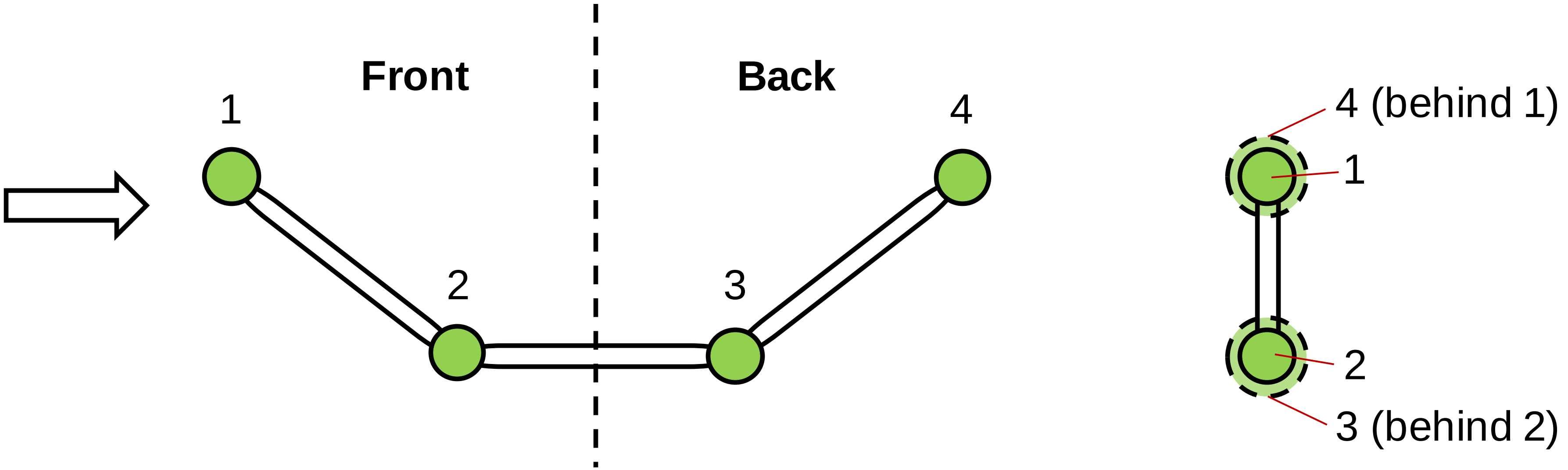
Figure 4.15 The Molecules of Life (© Garland Science 2013)

Torsion angles: Φ (phi) and Ψ (psi)



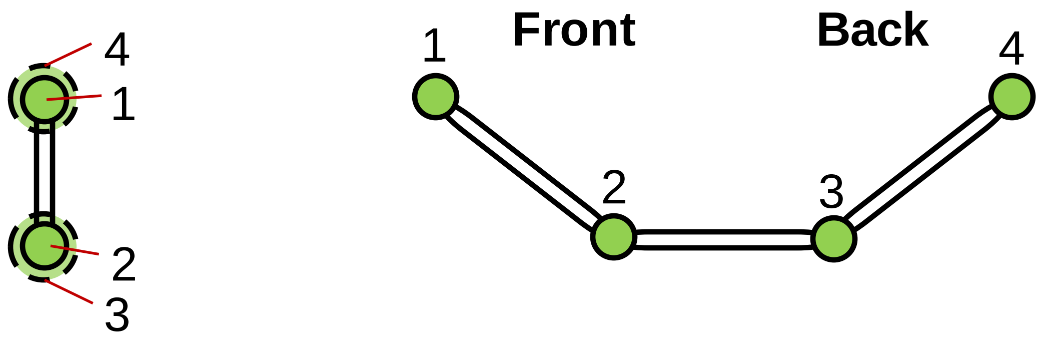


Visualizing a few torsion angles

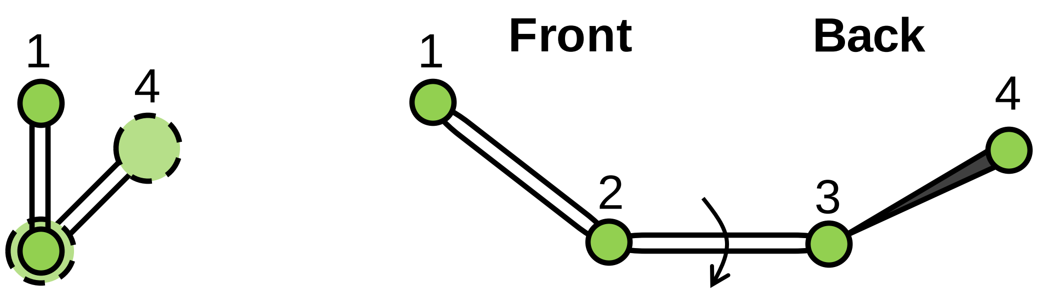


Visualizing a few torsion angles

0°



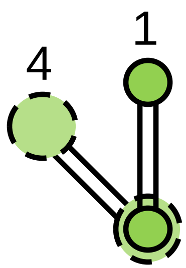
+45°

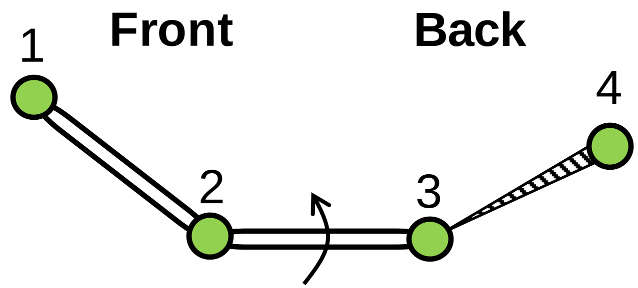


Atom 4 is above the plane of the board

Visualizing a few torsion angles

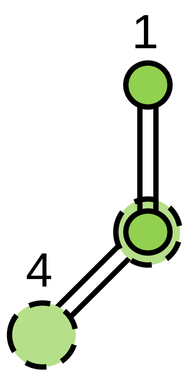
-45°

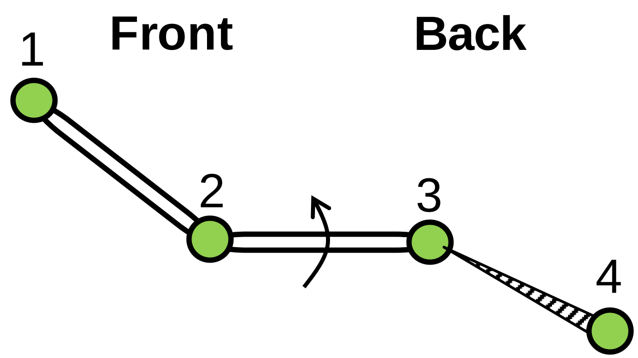




Atom 4 is below the plane of the board

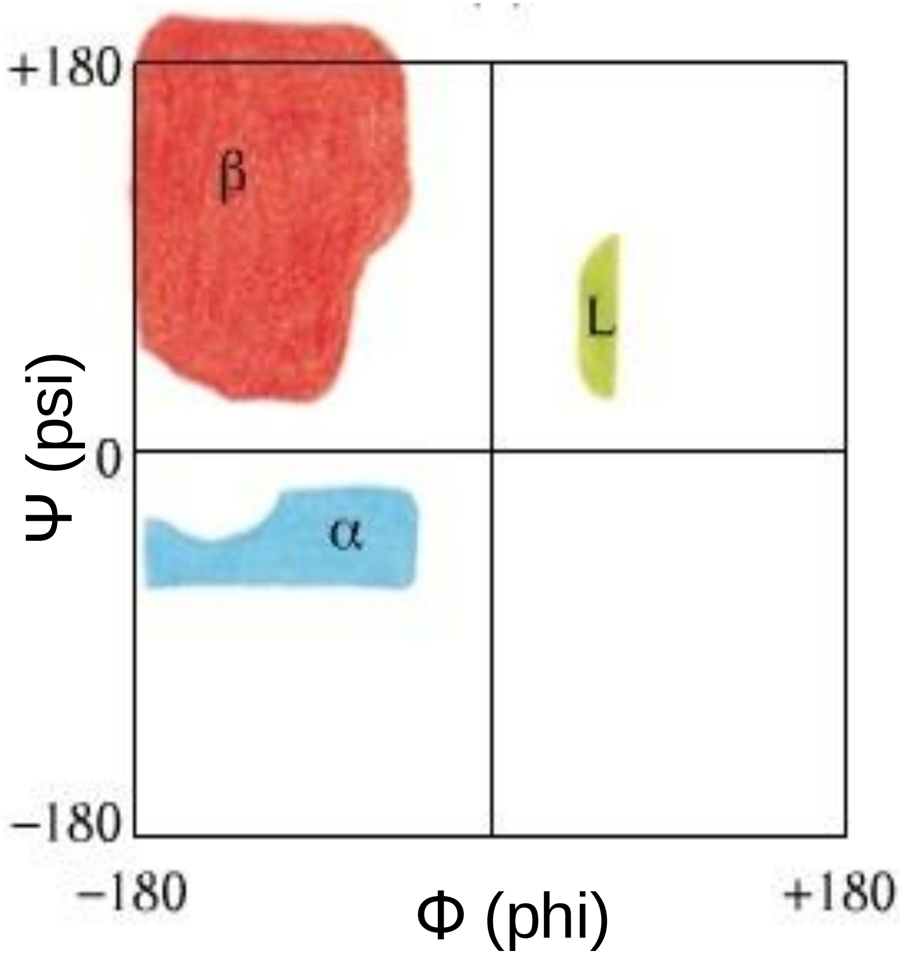
-135°



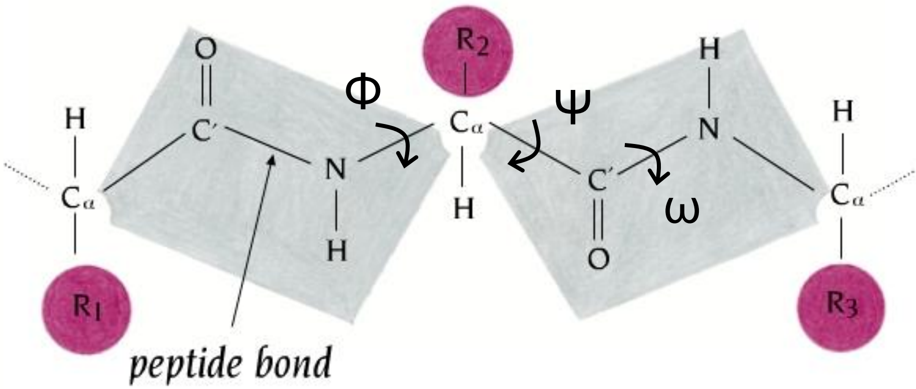


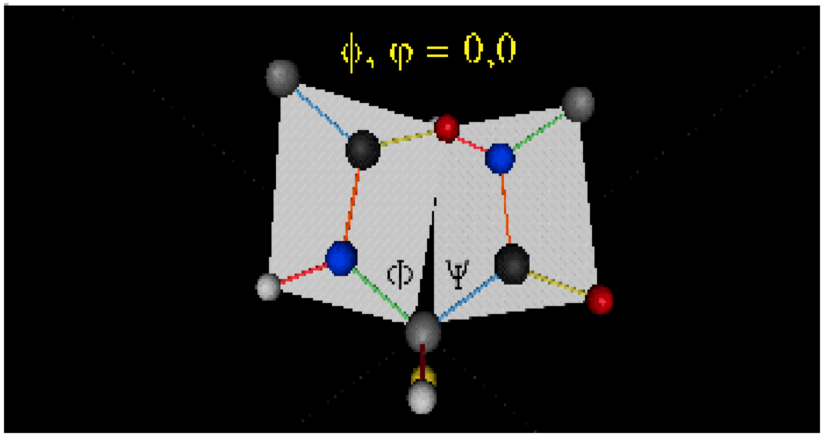
Atom 4 is below the plane of the board

Ramachandran Plot

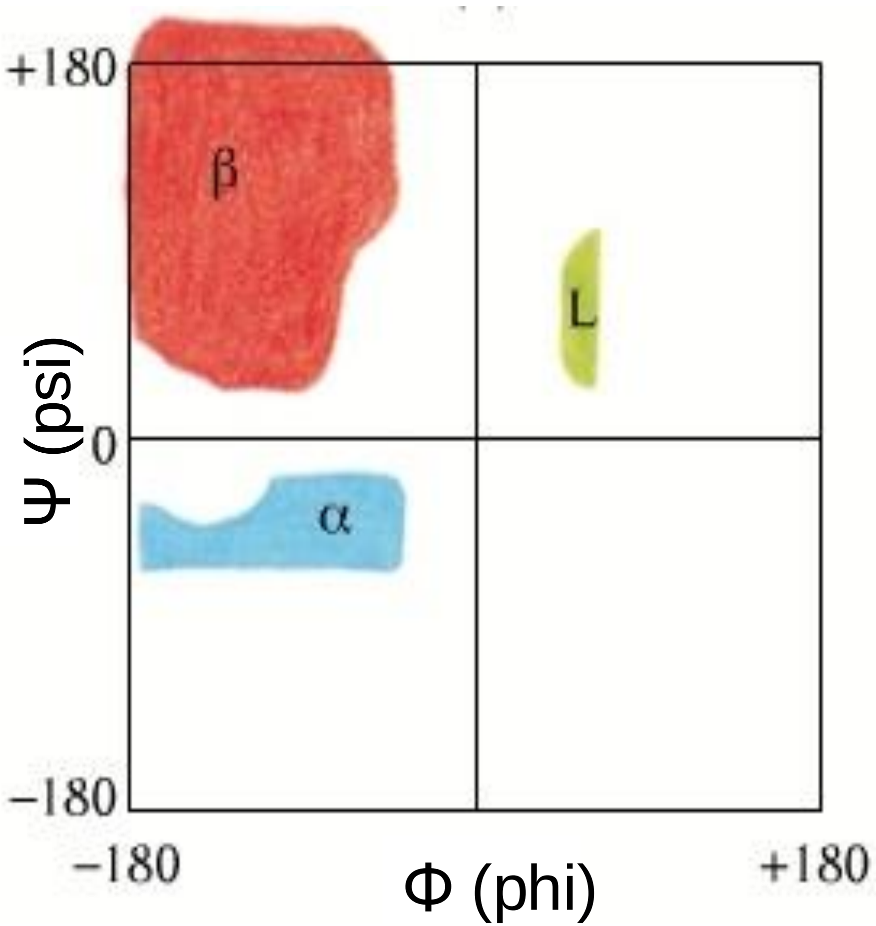


G. N. Ramachandran

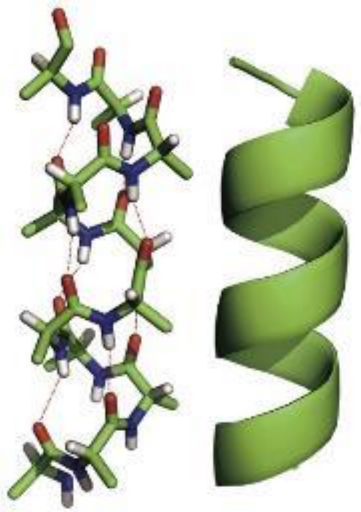




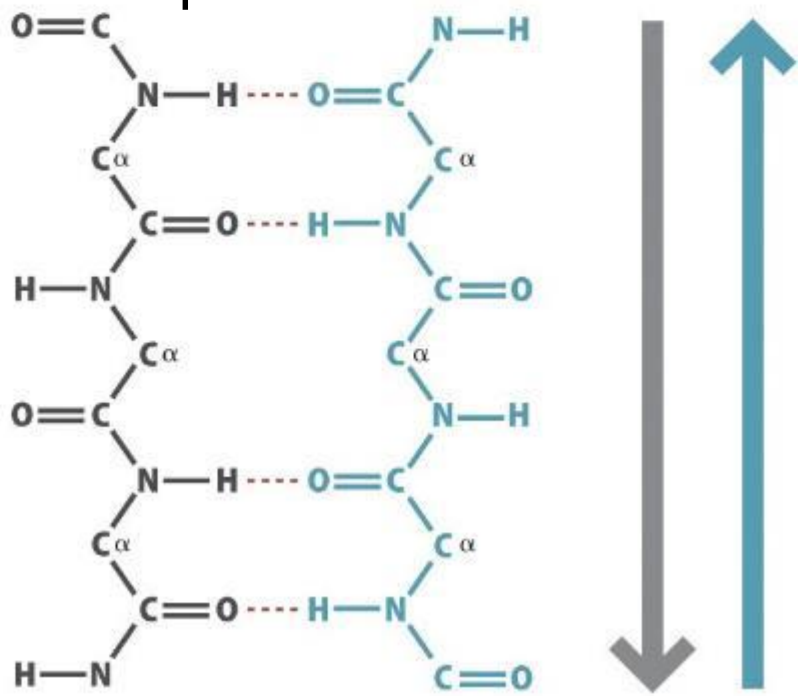
Ramachandran Plot



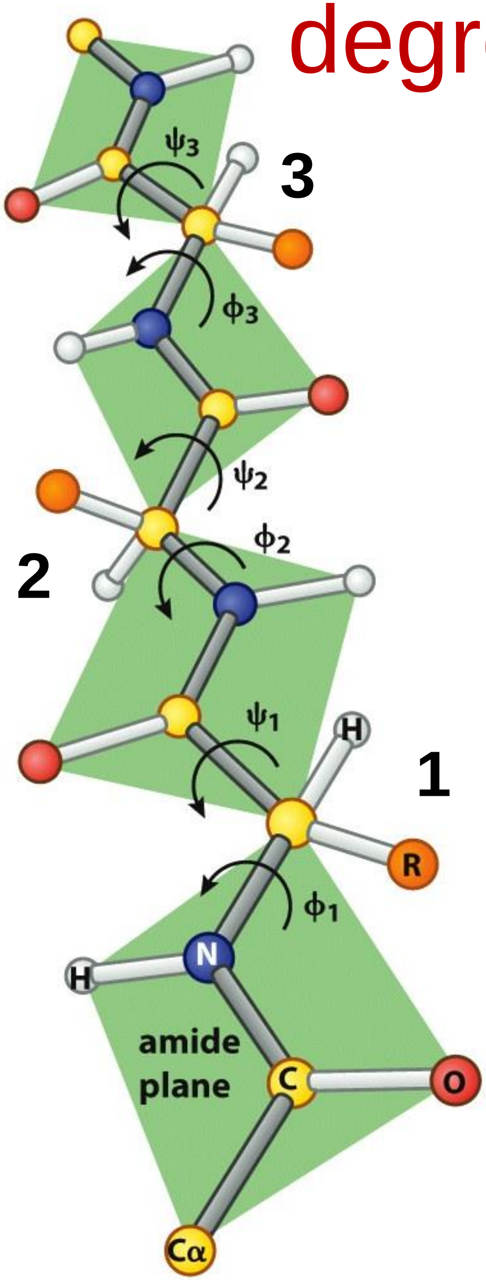
α-Helix



β-Strand



φ and ψ torsion angles are the only degrees of freedom for the backbone



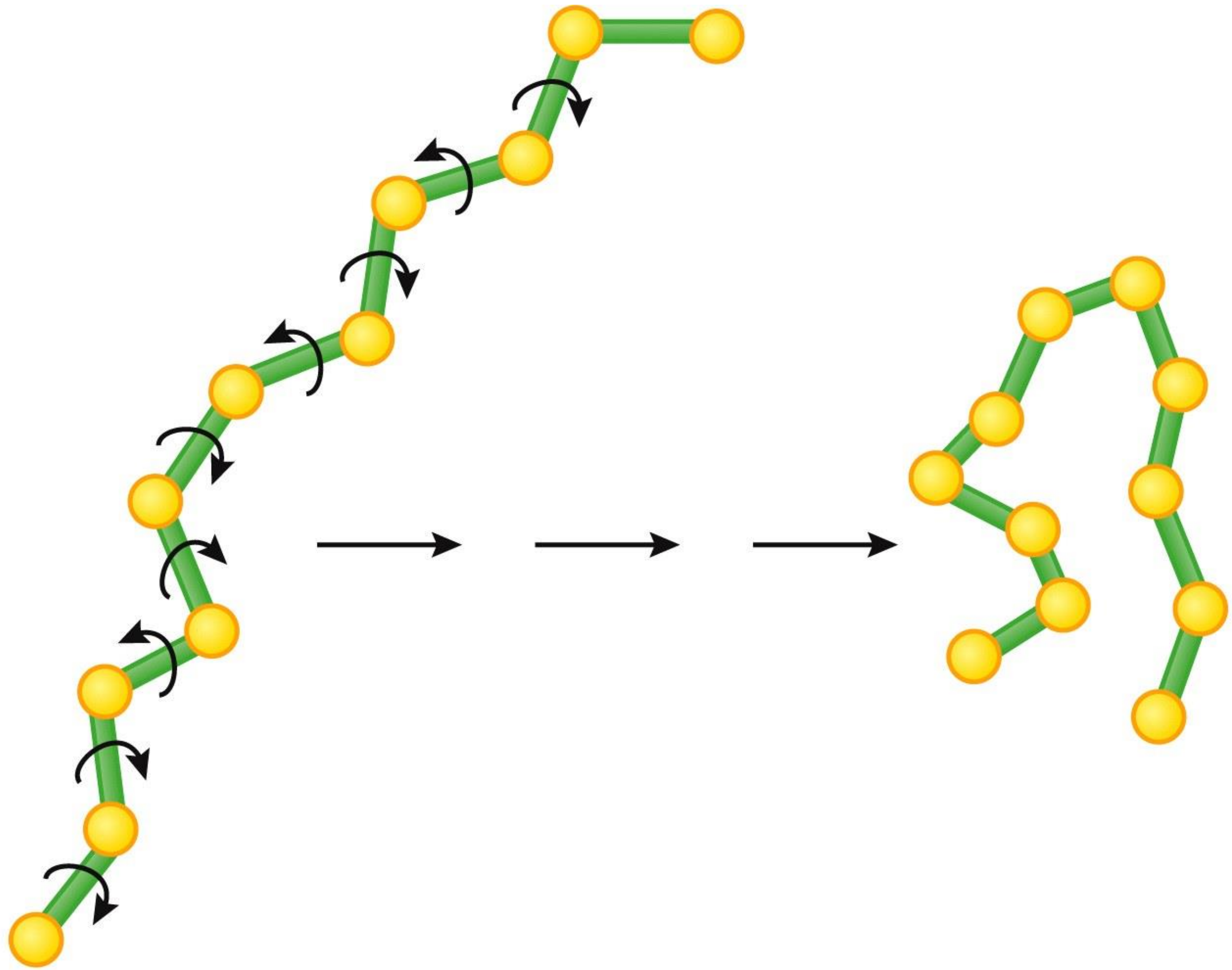
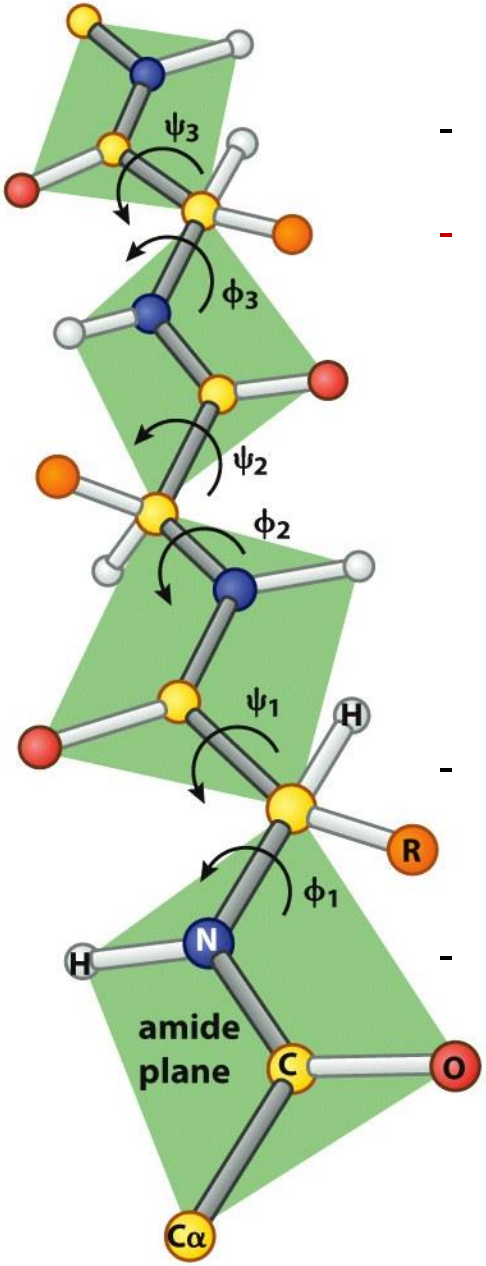


Figure 4.11b The Molecules of Life (© Garland Science 2013)

The protein folding problem

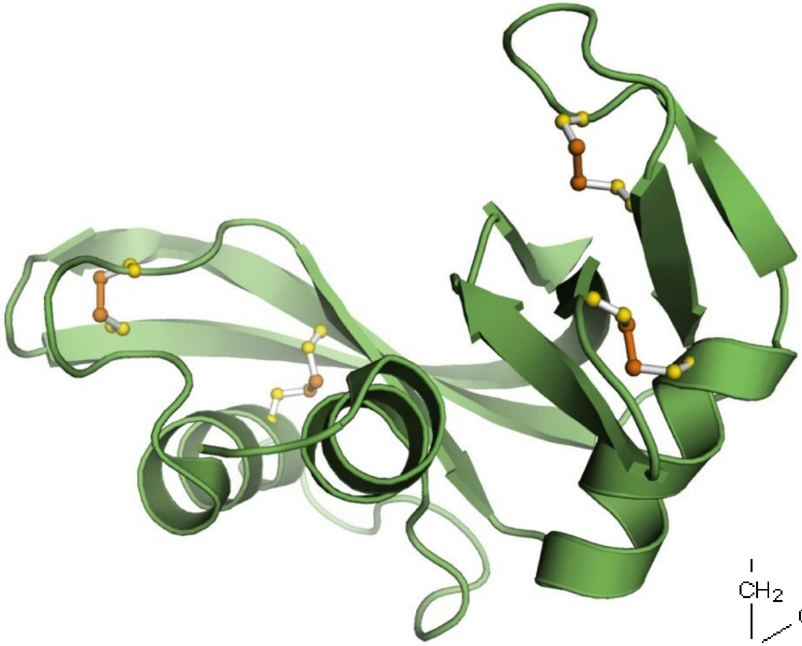


- Consider a small protein with 100 residues.
- Cyrus Levinthal calculated that, if each residue can assume three different conformations, the total number of structures would be $\mathbf{3^{100}}$, which is equal to 5×10^{47}. If it takes 10^{-13} s to convert one structure into another, the total search time would be $5 \times 10^{47} \times 10^{-13}$ s, which is equal to 5×10^{34} s, or **10^{27} years** i.e. longer than the age of the universe!
- Clearly, it would take much too long for even a small protein to fold properly by randomly trying out all possible conformations.
- The enormous difference between calculated and actual folding times is called *Levinthal's paradox.*

Important Questions on Protein Folding

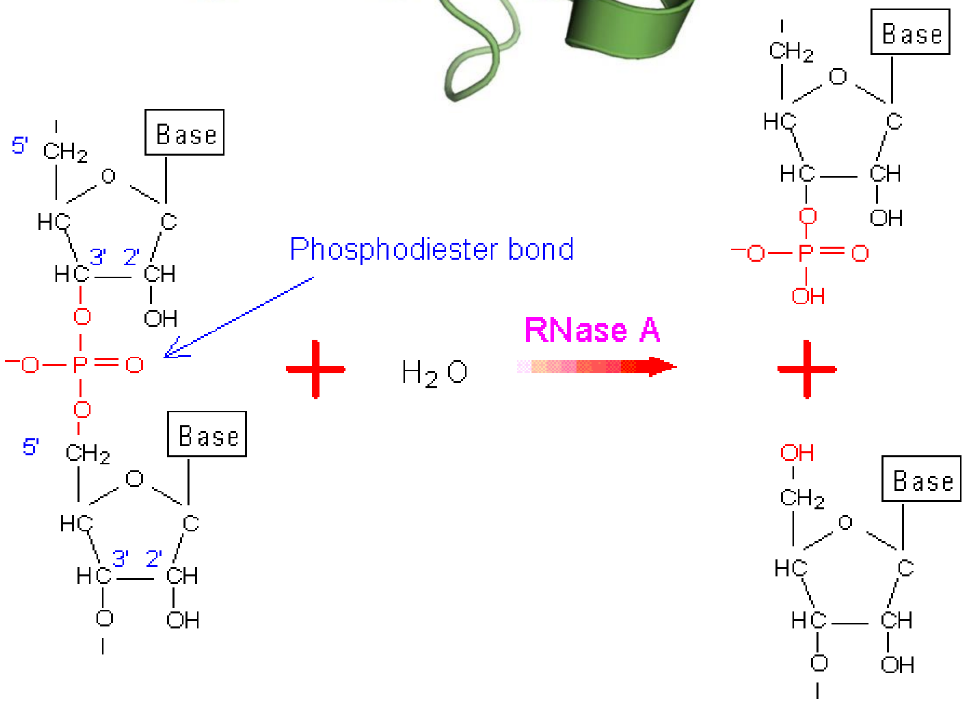
1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
2. How do proteins fold so fast? Most proteins fold within milliseconds.
3. Can we predict protein structures without experimentally solving them?
4. Can we design artificial proteins with unique functions to solve some of our problems?

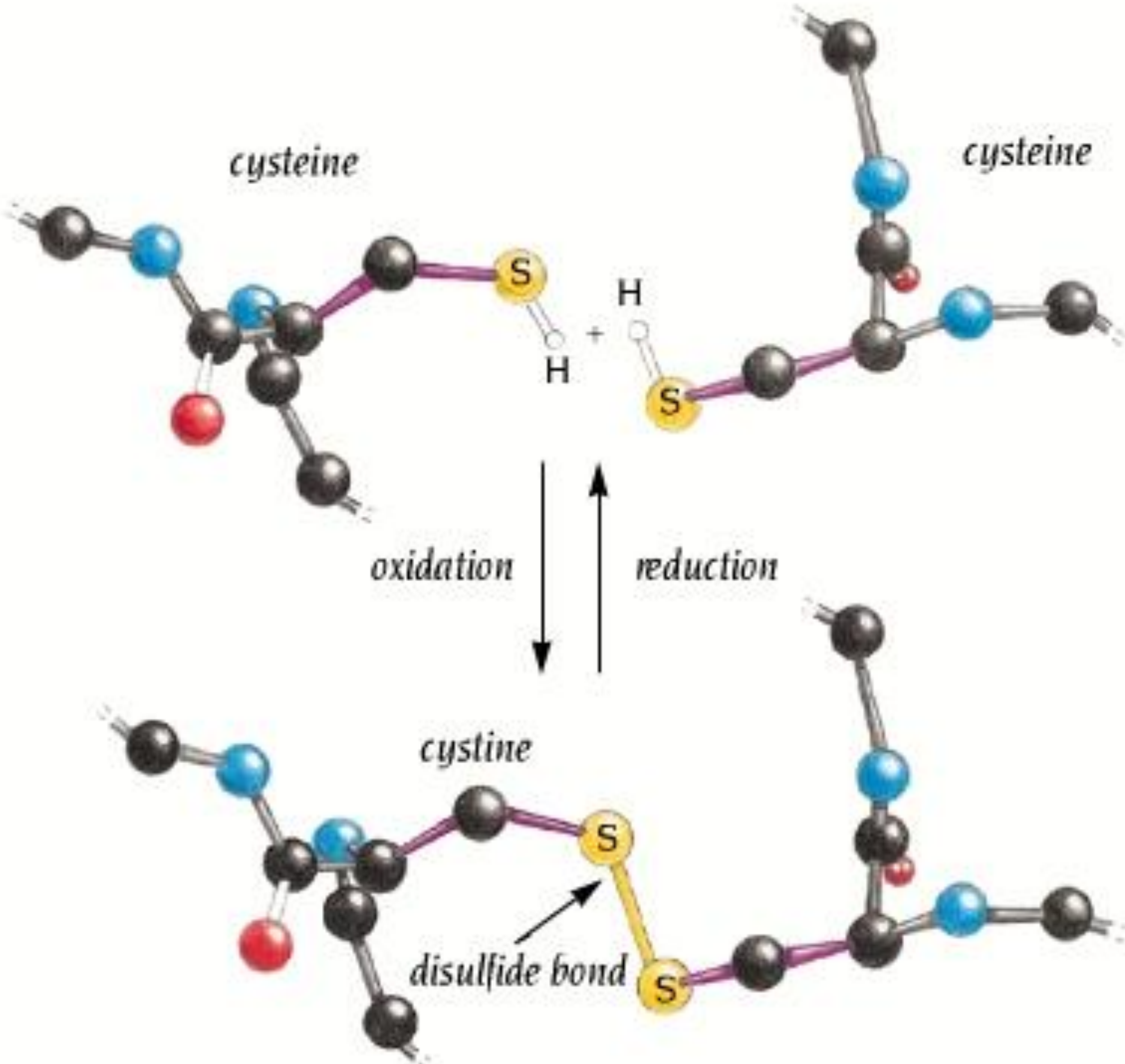
The 3D structure of a protein is encoded in its primary sequence: Anfinsen's Experiment



Thermodynamic hypothesis of Protein Folding: The interactions between the atoms in a protein control the folding of the protein molecule into a well-defined three-dimensional structure.

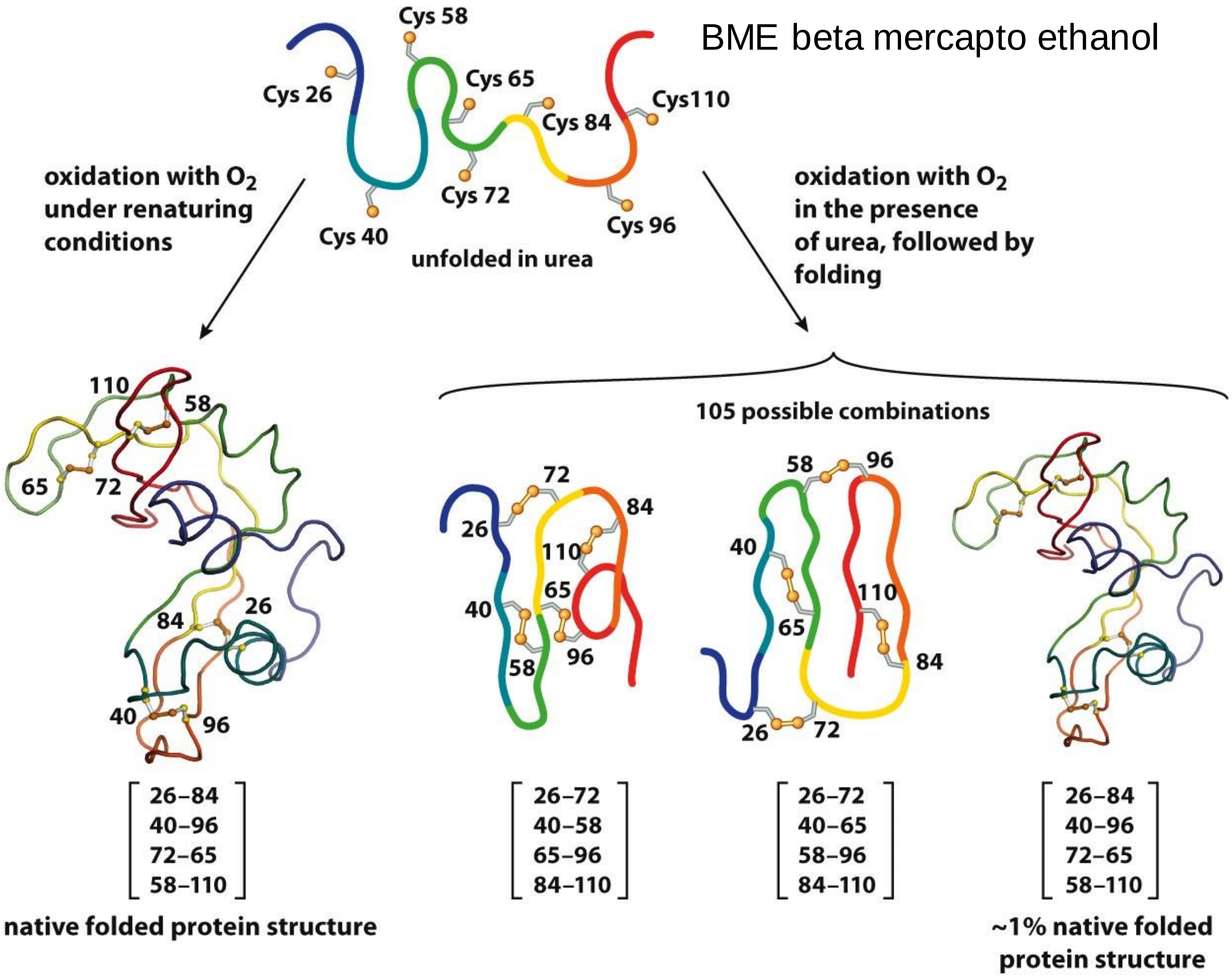
In other words, the protein sequence contains enough information required for the proper folding of the protein into its functional three-dimensional structure.



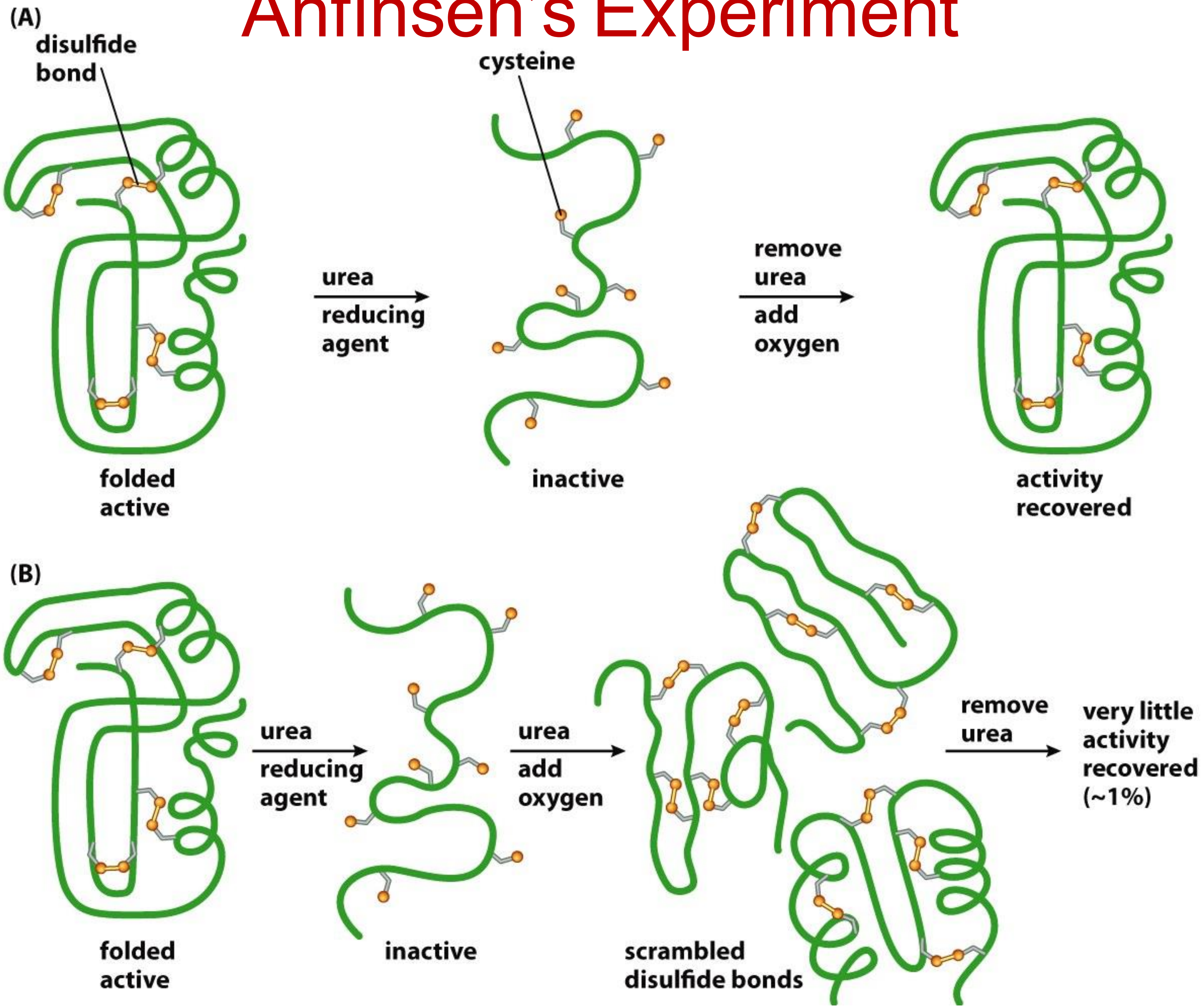

cysteine
cysteine
S
H
+
H
S
oxidation
reduction
cystine
S
S
disulfide bond

Anfinsen's Experiment

BME beta mercapto ethanol



Anfinsen's Experiment



If we understand HOW PROTEINS FOLD, we can predict their structure from sequence! Then we can design proteins with novel functions.

8 Cys

Select two at a time to form a disulphide bond

$$^{8}C_{2} * {}^{6}C_{2} * {}^{4}C_{2} * {}^{2}C_{2} / 4! = 2520/4! = 105$$

# of ways to choose the first disulfide bond

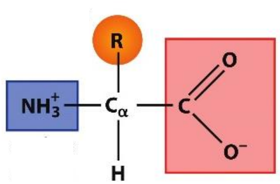
It does not matter in which order the FOUR disulphide bonds are formed. # of permutations of FOUR disulfide bonds.

Important questions on Protein Folding

3. Can we predict protein structure from sequence

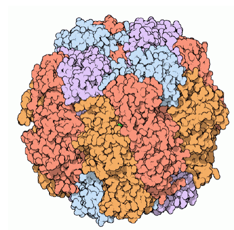
- Anfinsen's experiment demonstrates that a protein sequence encodes its structure.
- Can we decipher this code? I.e. can we predict the structure of a protein from its primary sequence?

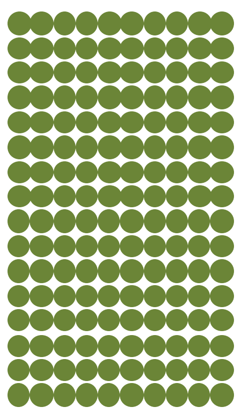
Sequence $\xrightarrow{\textbf{Prediction}}$ **Structure**



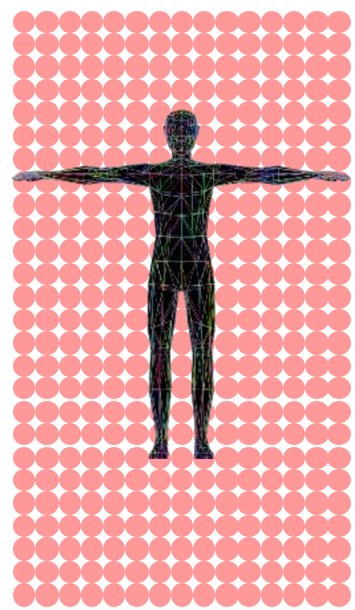
A	C	D	E
F	G	H	I
K	L	M	N
P	Q	R	S
T	V	W	Y

20 Amino Acids





300 AA/Protein



21,000 Proteins in Human body

200,000,000 Proteins in the world

Important questions on Protein Folding

3. Can we predict protein structure from sequence

- Anfinsen's experiment demonstrates that a protein sequence encodes its structure.
- Can we decipher this code? I.e. can we predict the structure of a protein from its primary sequence?

Sequence $\xrightarrow{\textbf{Prediction}}$ **Structure**

NEWS · 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Microsoft Azure

Overview Solutions Products Documentation Pricing Training Marketplace Partners

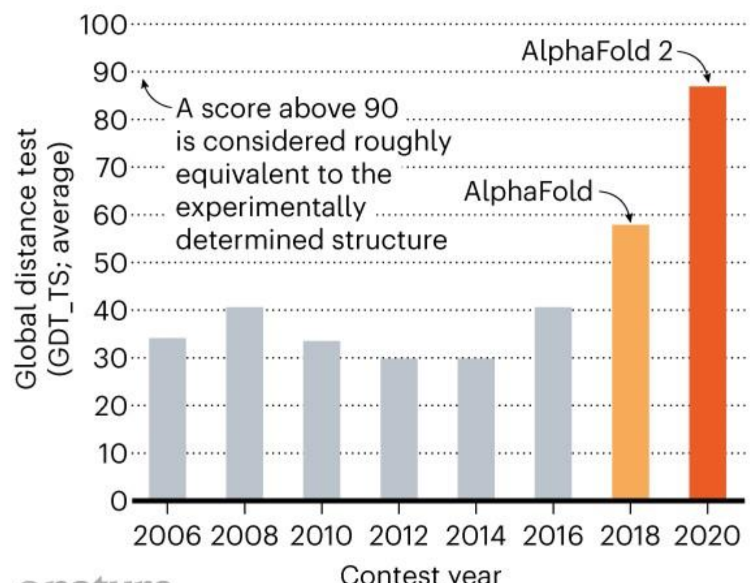
Blog / Cloud Strategy

Windows Azure Helps Scientists Unfold Protein Mystery and Fight Disease

Posted on 16 June, 2011

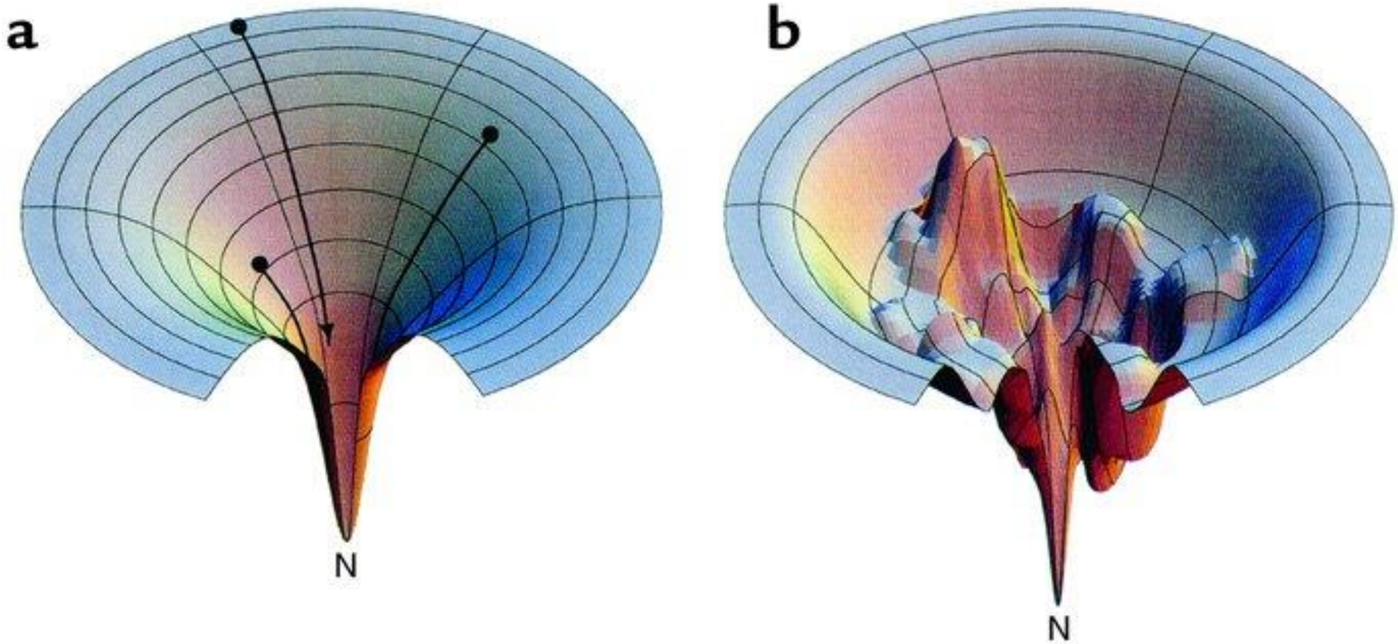
STRUCTURE SOLVER

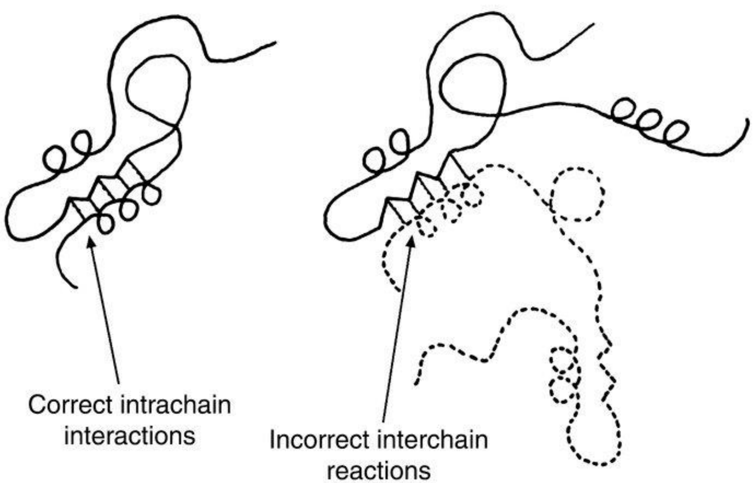
DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



©nature

How Proteins Fold





Important interactions between amino acids:

- Hydrophobic interactions
- Hydrogen bonding interactions
- Electrostatic interactions

How Proteins Fold

1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
2. How do proteins fold so fast? Most proteins fold within milliseconds.

- Protein Folding is a stochastic process i.e. not all conformations are sampled.
- Proteins fold in small segments (~20 amino acids) independent of the rest.

  These folded segments or FOLDONS collapse to give the final structure.

Molecular Dynamics (MD) Simulations use physical laws to study protein folding

Animation of Protein Folding Funnel
https://www.youtube.com/watch?v=YANAso8Jxrk

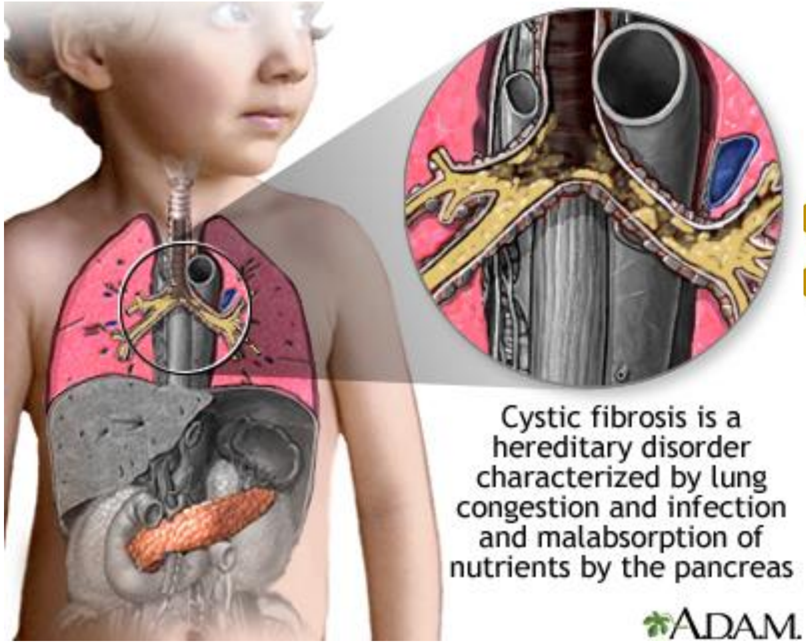
The protein folding game - Foldit

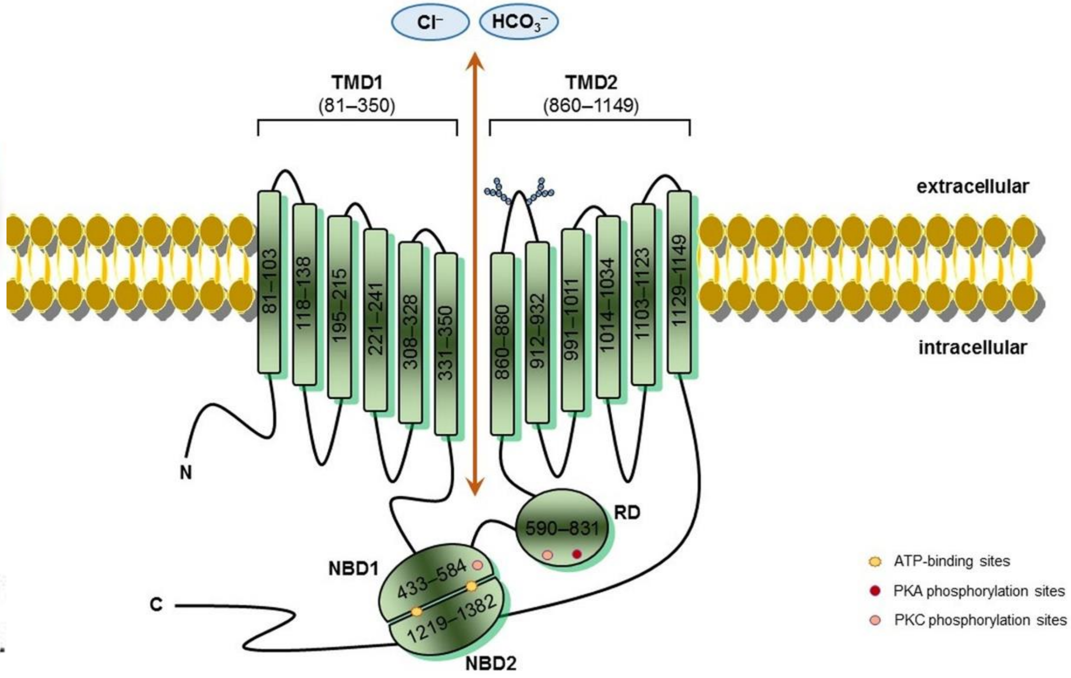
https://fold.it/portal/

Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

1) Cystic Fibrosis: It results from the misfolding of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), a protein functioning as a chloride (Cl^-) ion channel. The loss of CFTR function interferes with the body's ability to efficiently secrete fluids and salts. It damages the lungs and digestive systems.





https://medlineplus.gov/ency/article/000107.htm

Frontiers in Pharmacology 2016 v-7, p-275

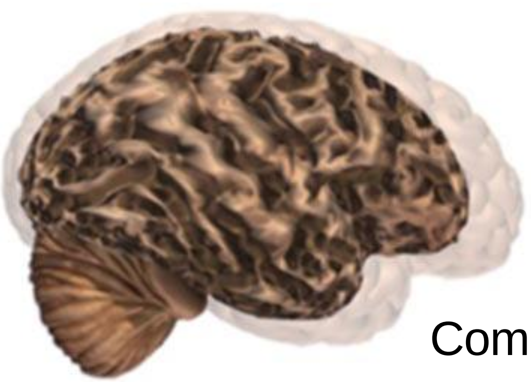
Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

2) **Alzheimer’s Disease (AD):** It results from the aggregation and precipitation of a peptide called amyloid-β (Aβ). AD is the most common form of progressive dementia in the elderly, and of neuro-degenerative diseases in general.

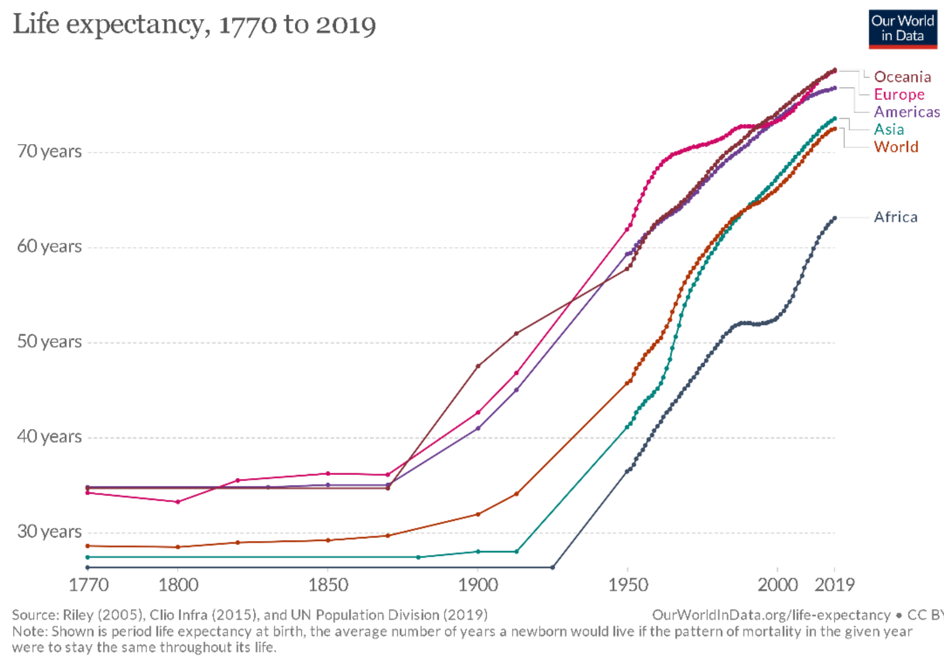
Normal Brain

Brain with AD



Comparison of the two Brains

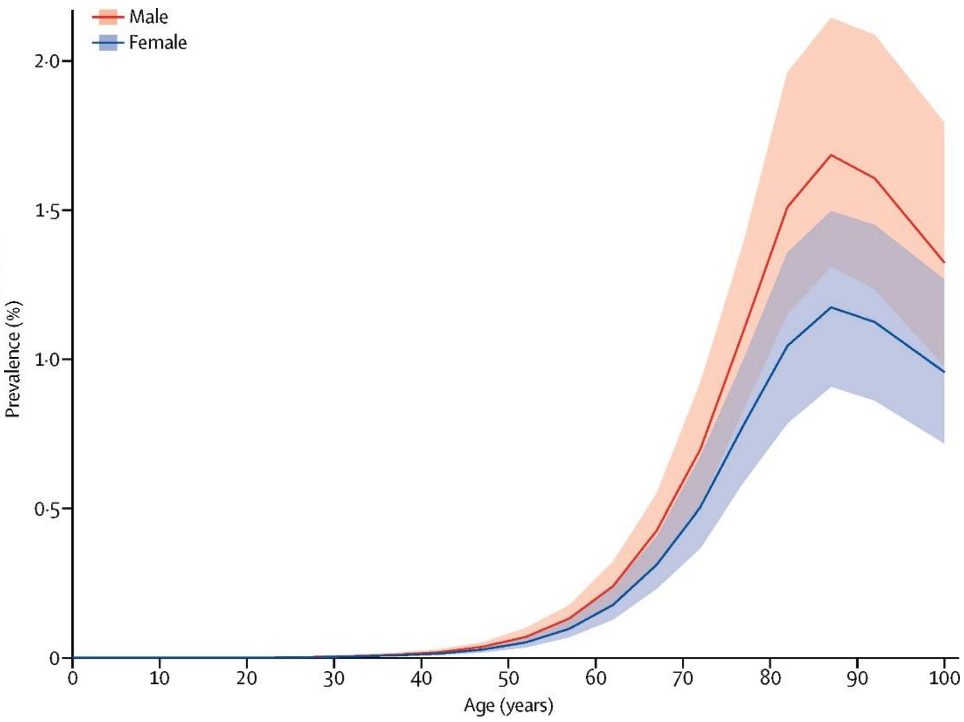
www.alz.org



Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

3) **Parkinson's Disease (PD):** It results from aggregation and precipitation of the protein α-synuclein. PD is a motor disorder common among the elderly (but can also hurt young people). It leads to shaking, stiffness, and difficulty with walking, balance, and coordination.



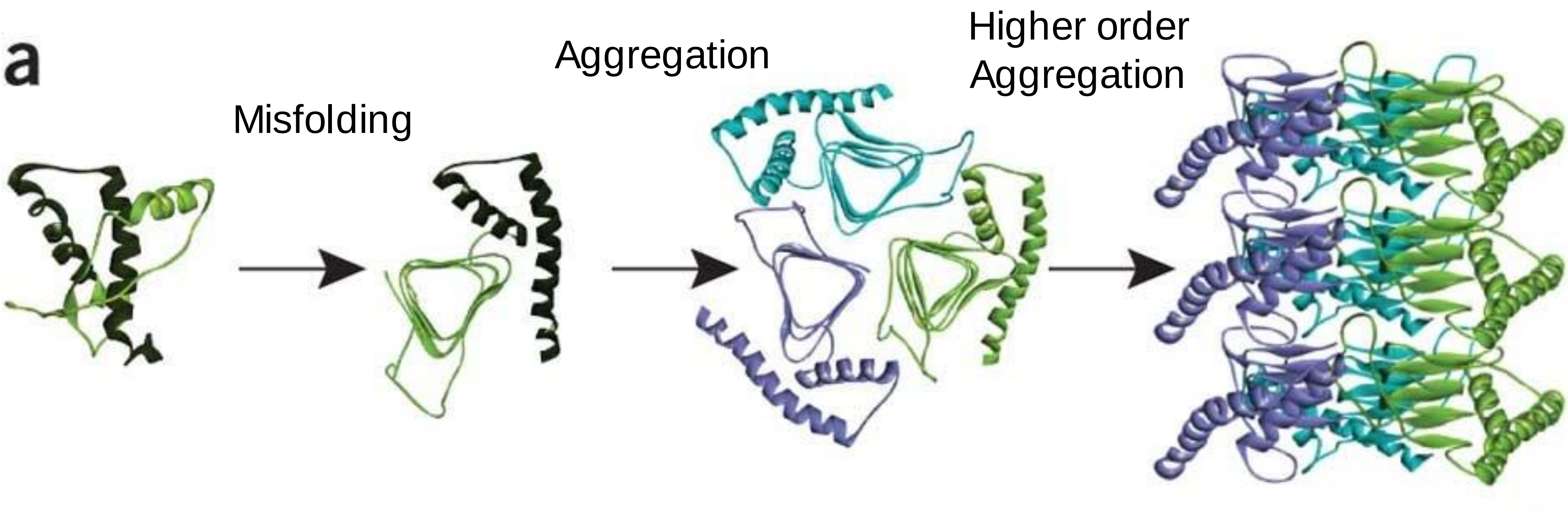
http://www.askdrray.com/parkinsons-disease-may-stopped/

Lancet Neurol 2018; 17: 939–53

Protein misfolding leads to several diseases

Several diseases occur due to misfolding of proteins. Few examples:

4) **Creutzfeldt-Jakob Disease (CJD):** It is caused by the aggregation and precipitation of the protein prion. It results in progressive motor dysfunction, cognitive impairment, and cerebral ataxia.

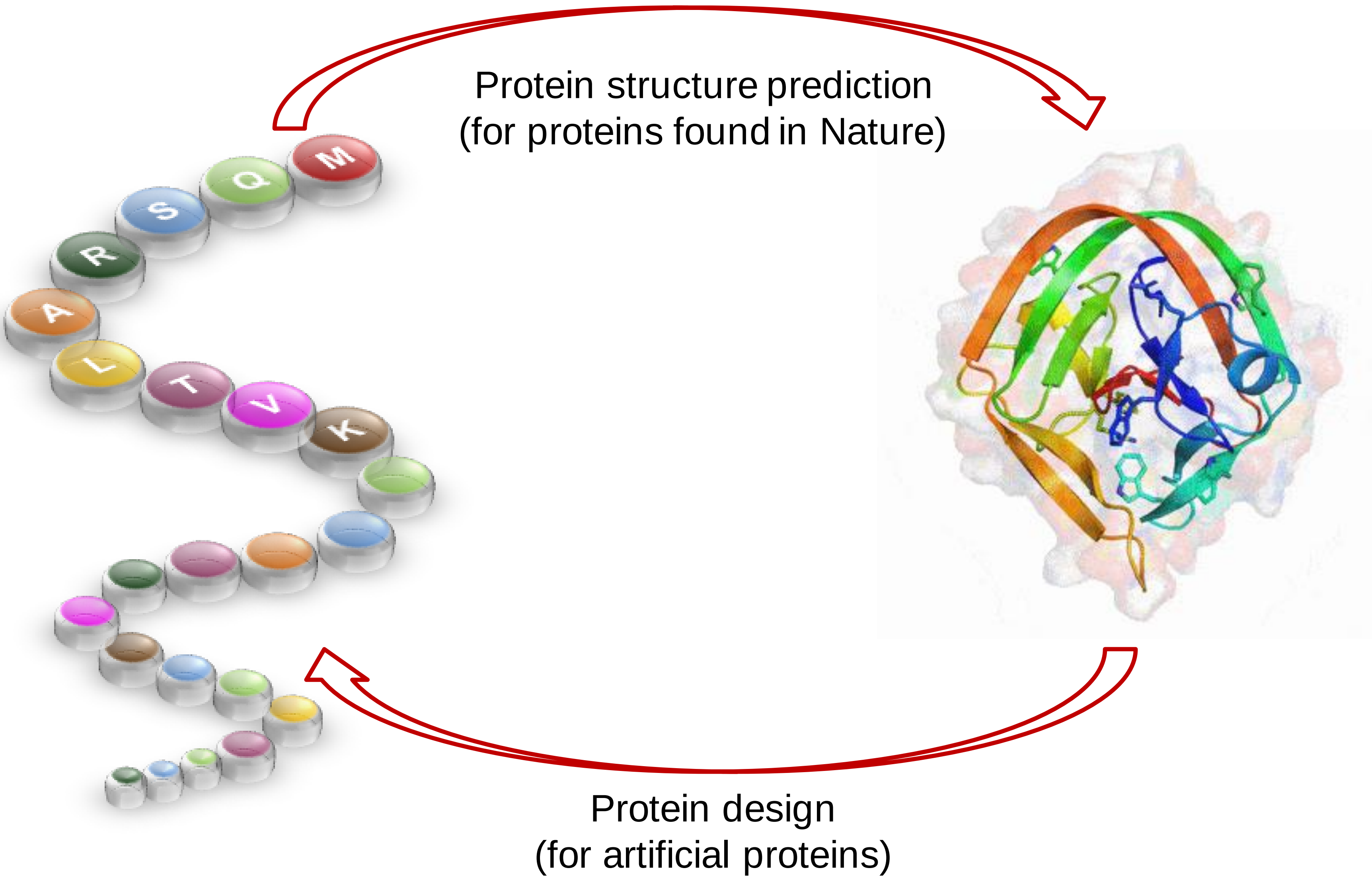


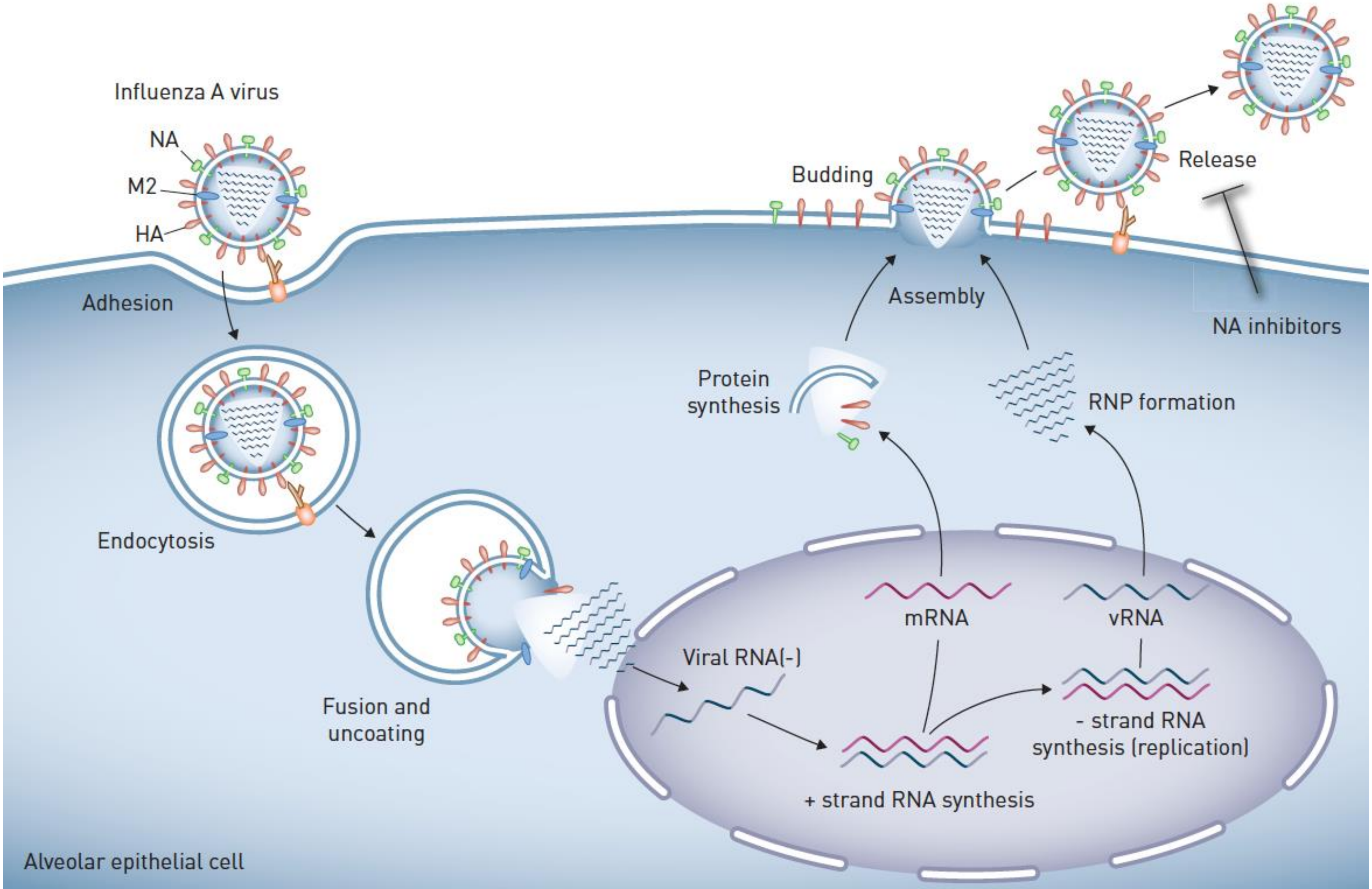
Nature Structural & Molecular Biology 19, 370-377 (2012)

Important Questions on Protein Folding

1. How do proteins fold? i.e. How do proteins achieve their final folded structure?
2. How do proteins fold so fast? Most proteins fold within milliseconds.
3. Can we predict protein structures without experimentally solving them?
4. Can we design artificial proteins with unique functions to solve some of our problems?

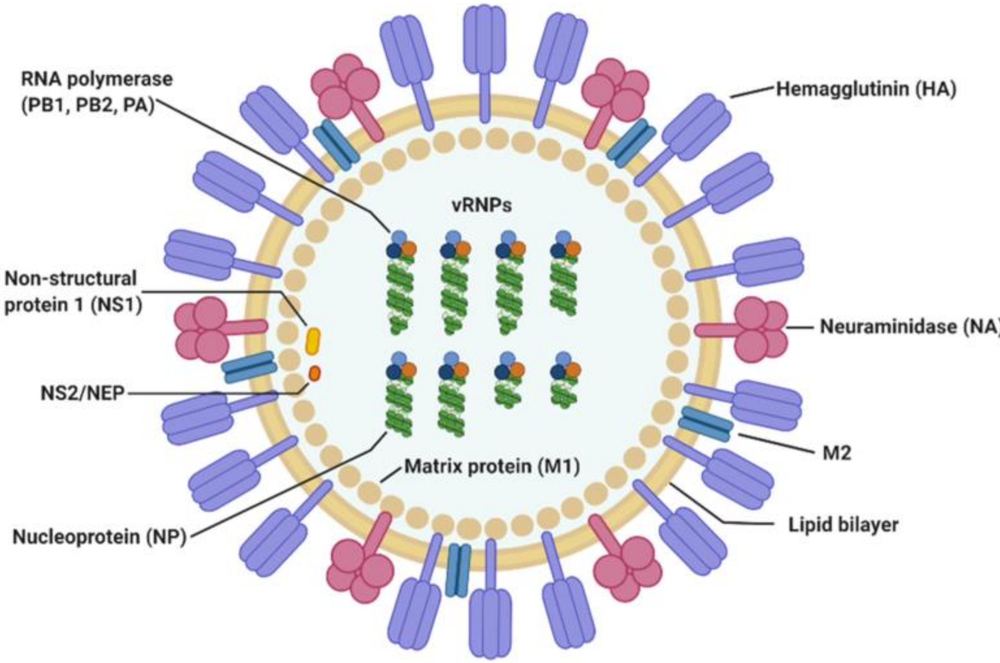
Protein Design Problem

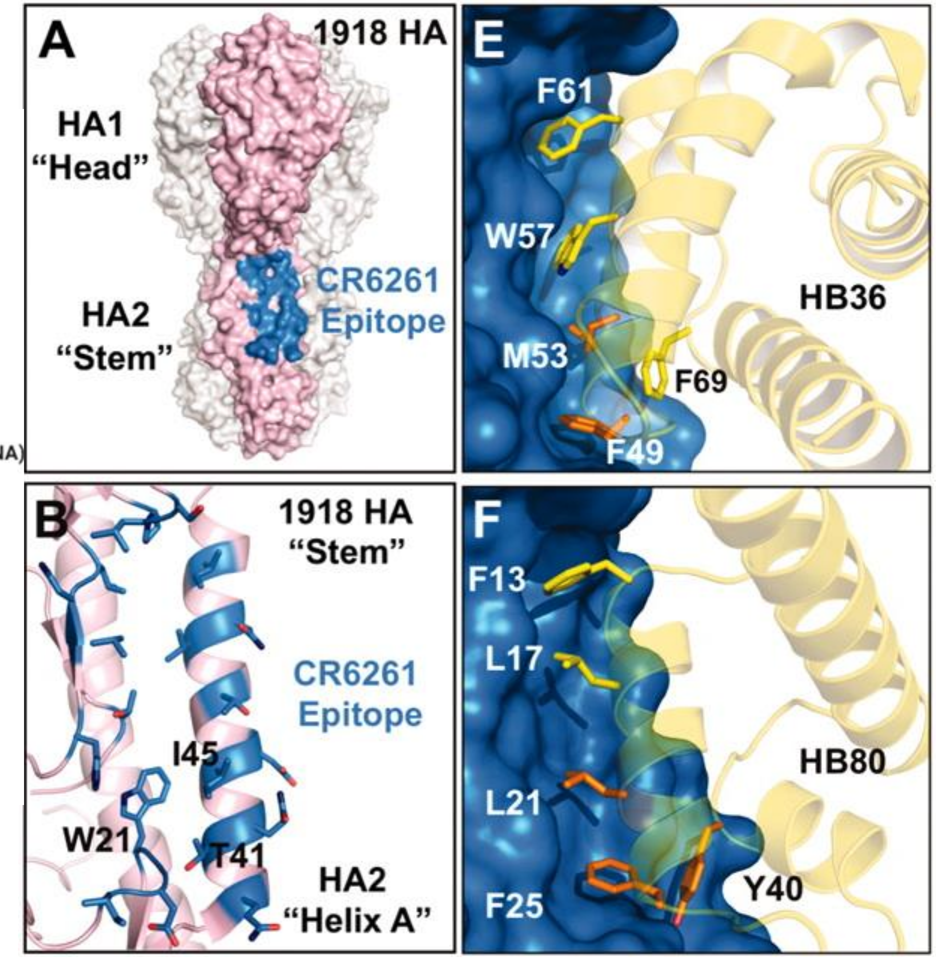




Influenza A virus
NA
M2
HA
Adhesion
Endocytosis
Fusion and uncoating
Viral RNA(-)
+ strand RNA synthesis
mRNA
vRNA
- strand RNA synthesis (replication)
Protein synthesis
RNP formation
Assembly
Budding
Release
NA inhibitors
Alveolar epithelial cell

Designed Protein Targets the Conserved Stem Region of Influenza Hemagglutinin

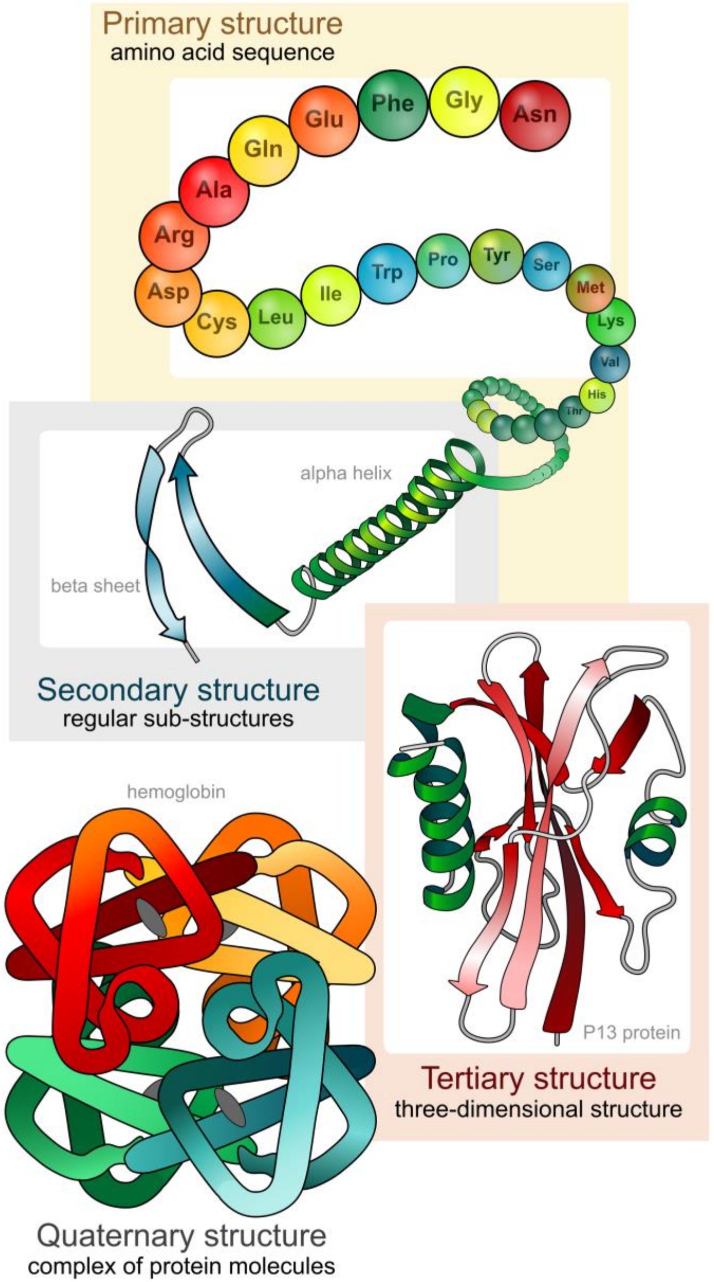




Science 2011 32:816

Hierarchy of Protein Structure

Protein Molecules are Organized in a Structural Hierarchy



Primary

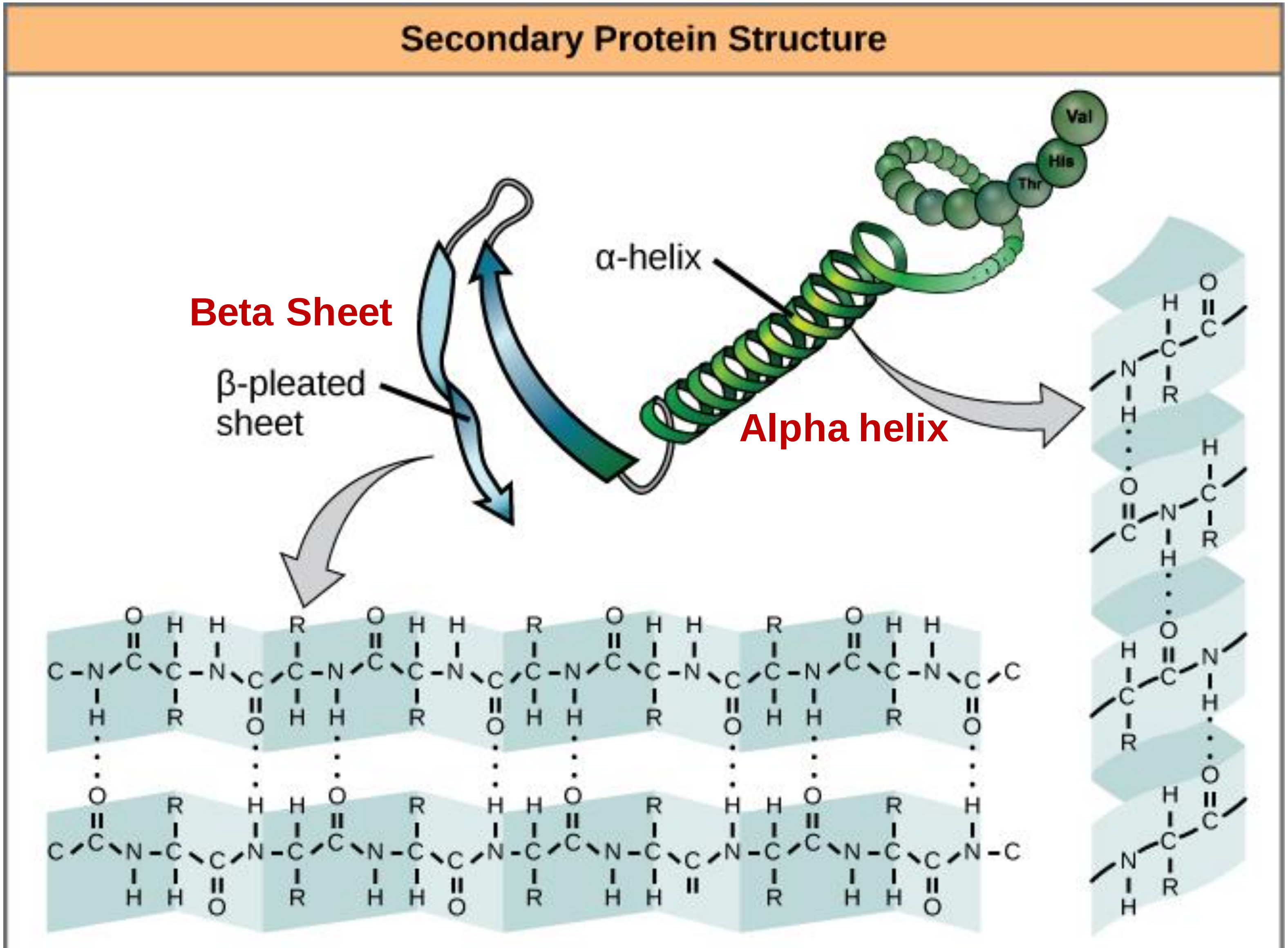
Secondary

Tertiary

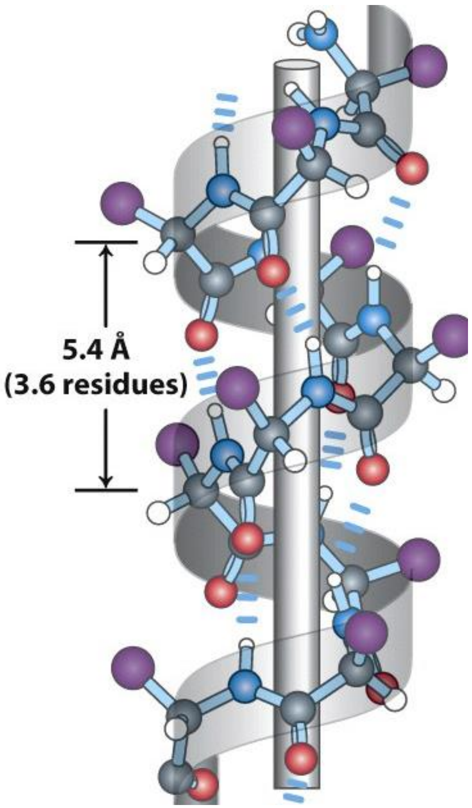
Quaternary

Secondary Protein Structure

Characterized by main chain NH and CO groups participating in H-bonds

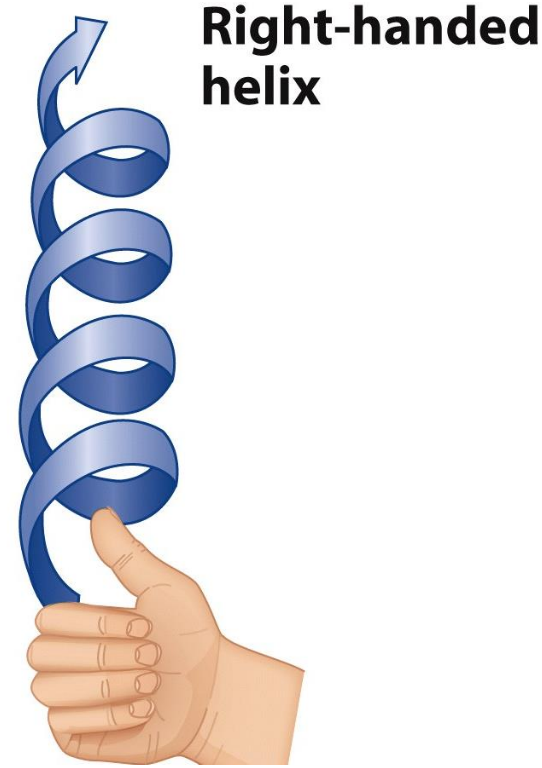


Alpha Helix

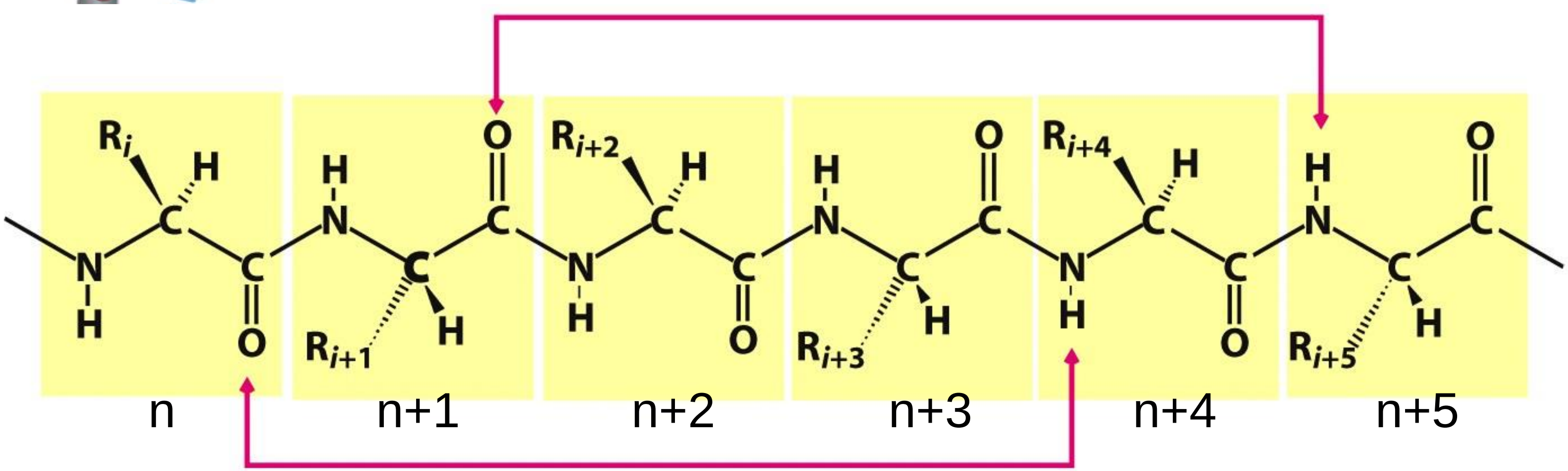


Every 3.6 residues make one turn

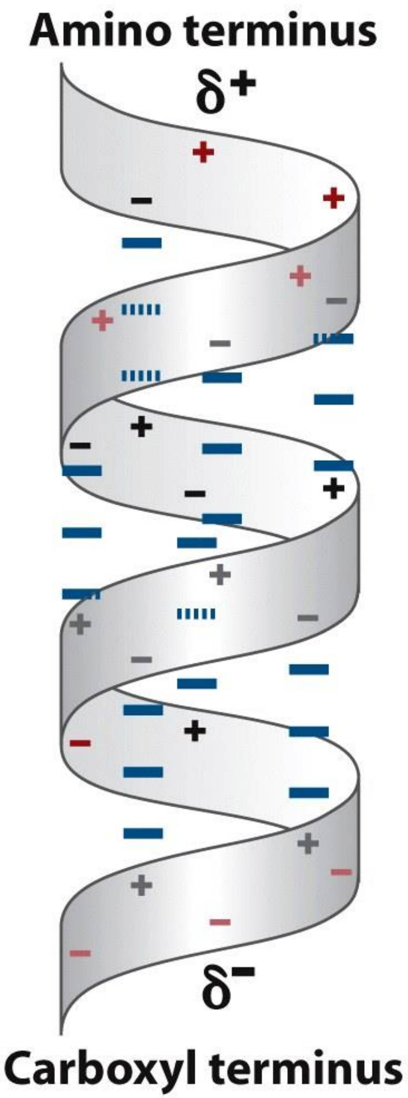
The distance (pitch of helix) between two turns is 5.4 Å

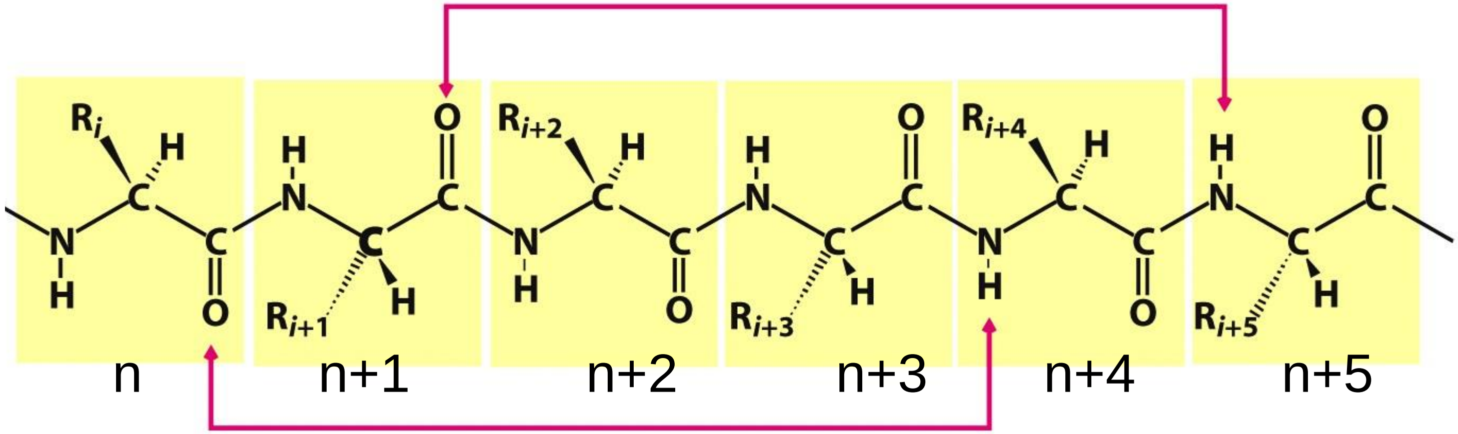


The C=O of residue ‘n’ is hydrogen bonded to N-H of residue ‘n+4’



The Alpha-Helix has a Dipole Moment





Macro dipole

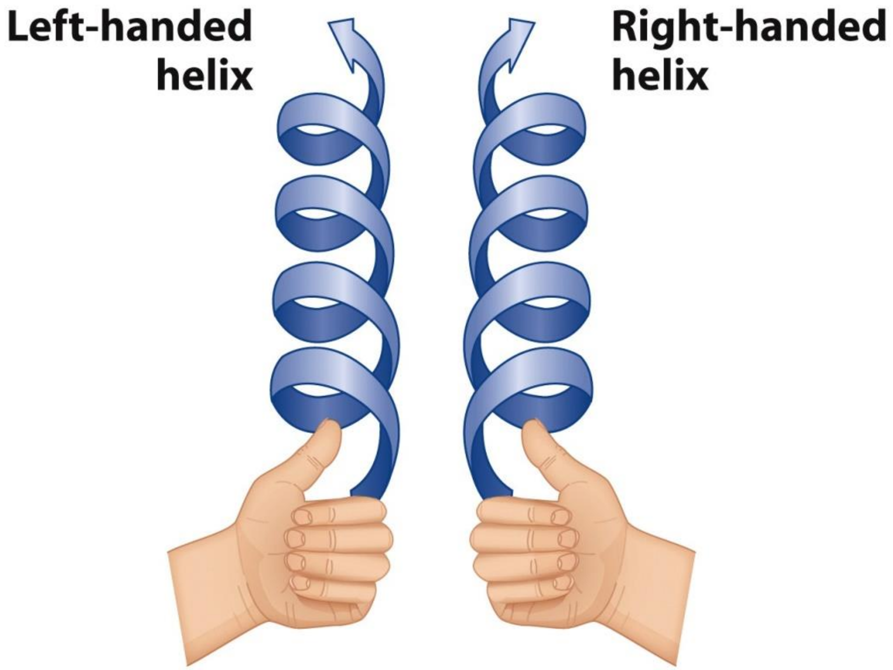
The dipoles of peptide units are aligned along the α helical axis

Alpha Helix: Right-handed or Left-handed?

Alpha helix can be – Right-handed or Left handed

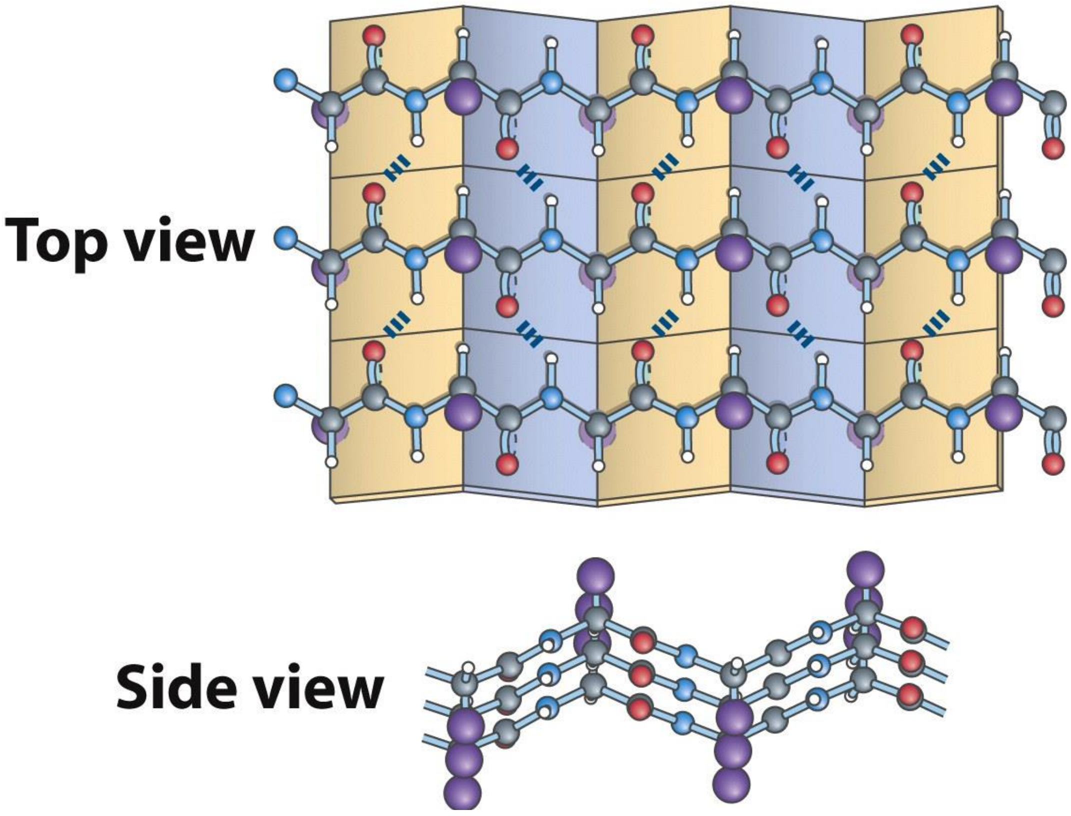
BUT, left handed helix is not possible for L-amino acids due to close approach of the side chains and CO group.

Right handed – most commonly observed in proteins.



β-sheet

(Number of β-Strands are Involved)

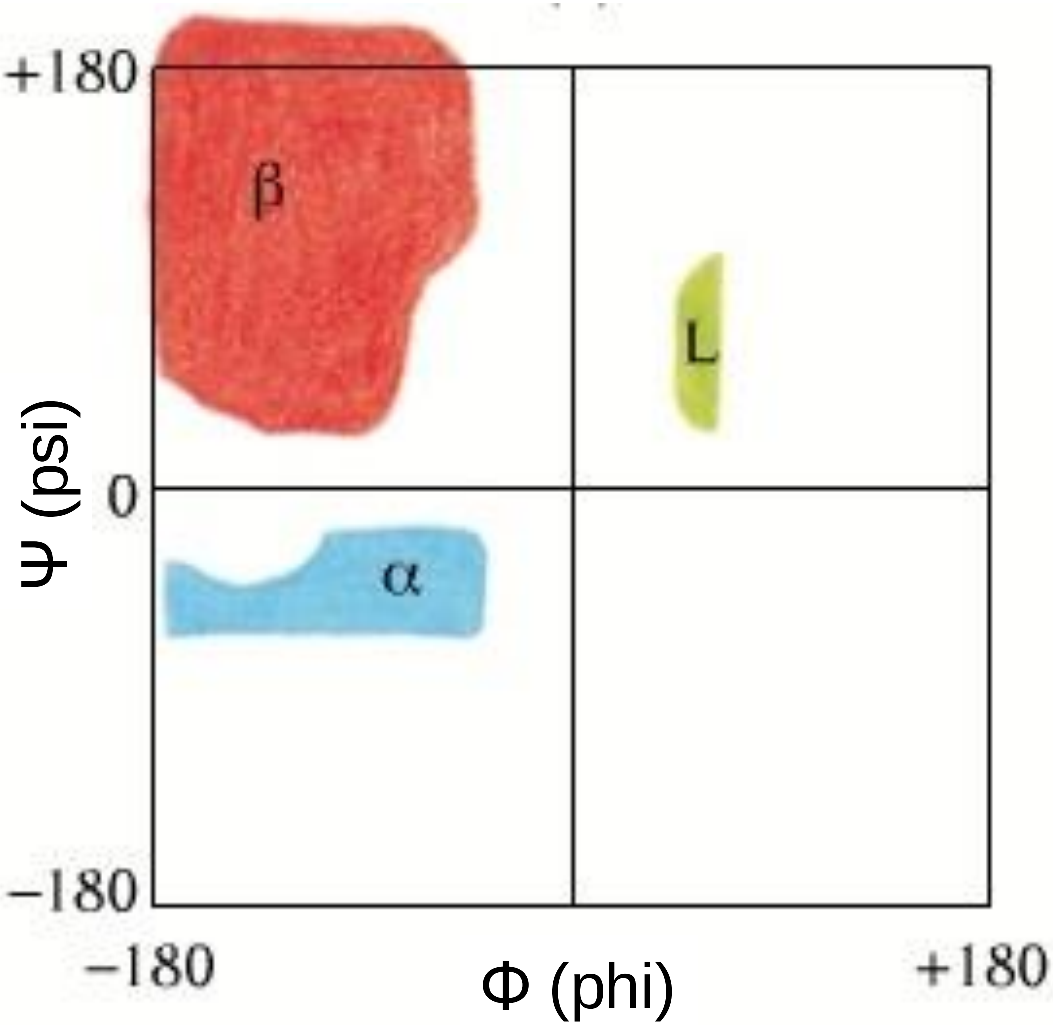


β-sheet from several regions of the chain; Each β-strand, typically 5-10 residues long

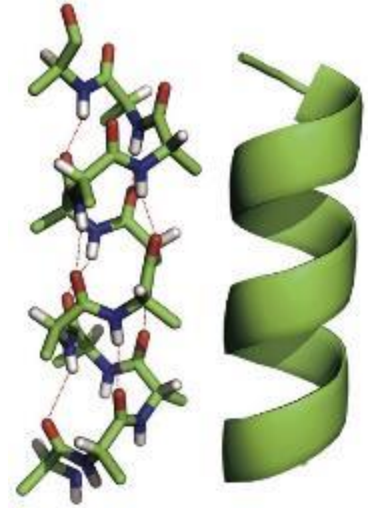
H-bonds are perpendicular to strands

α-helix: from one continuous region

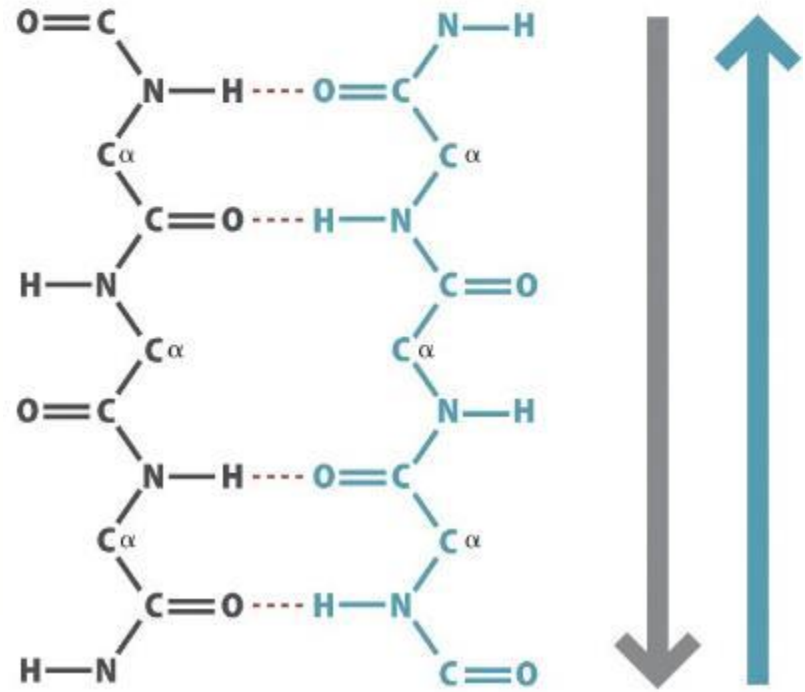
Ramachandran Plot



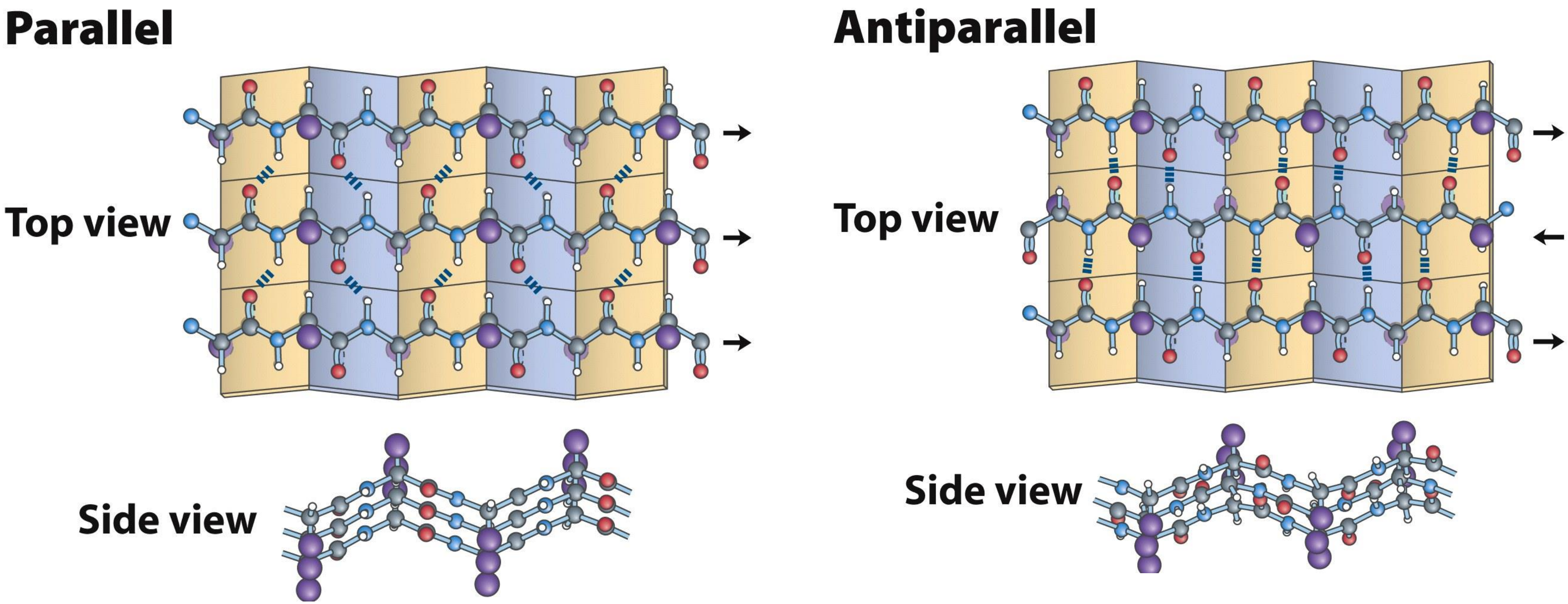
α-Helix

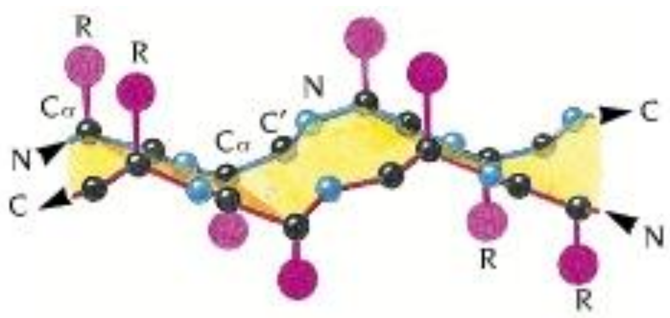


β-Strand



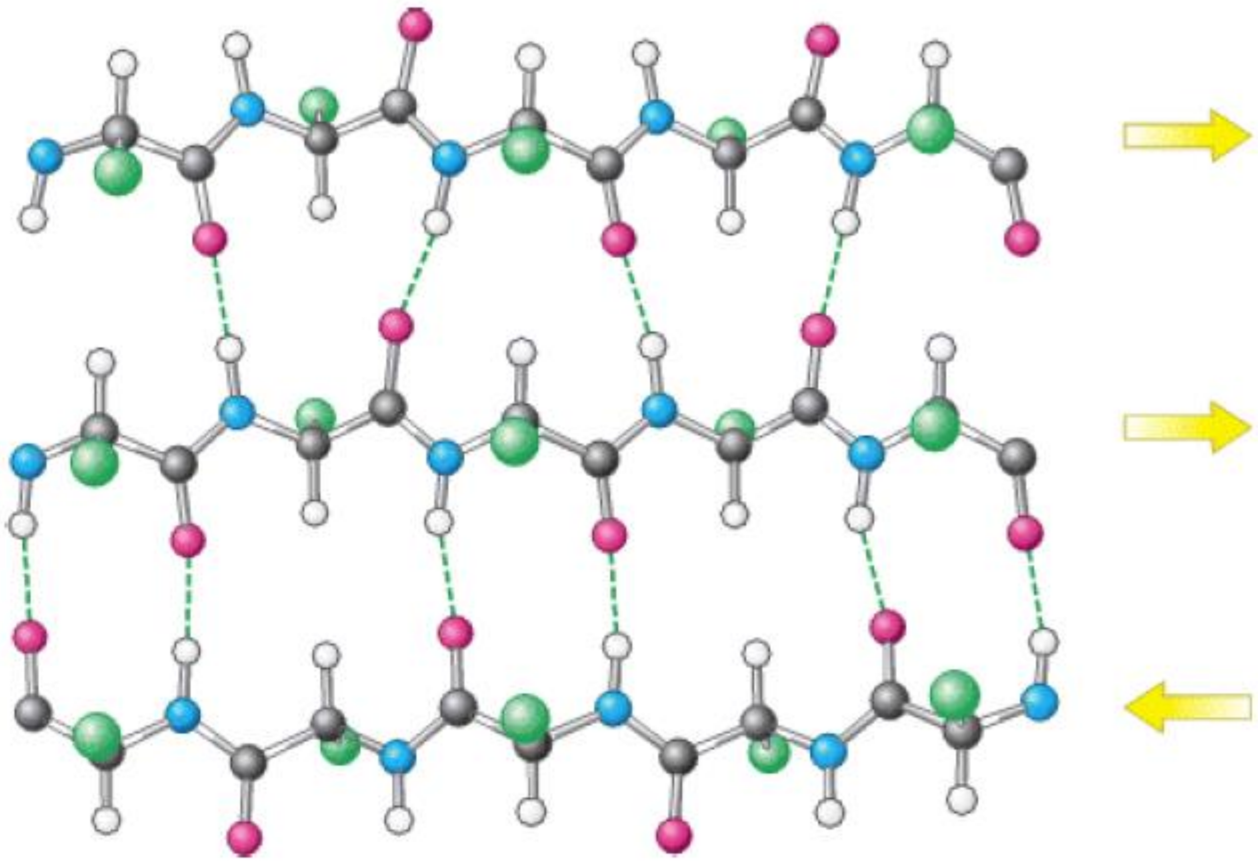
Parallel and Antiparallel β-sheet





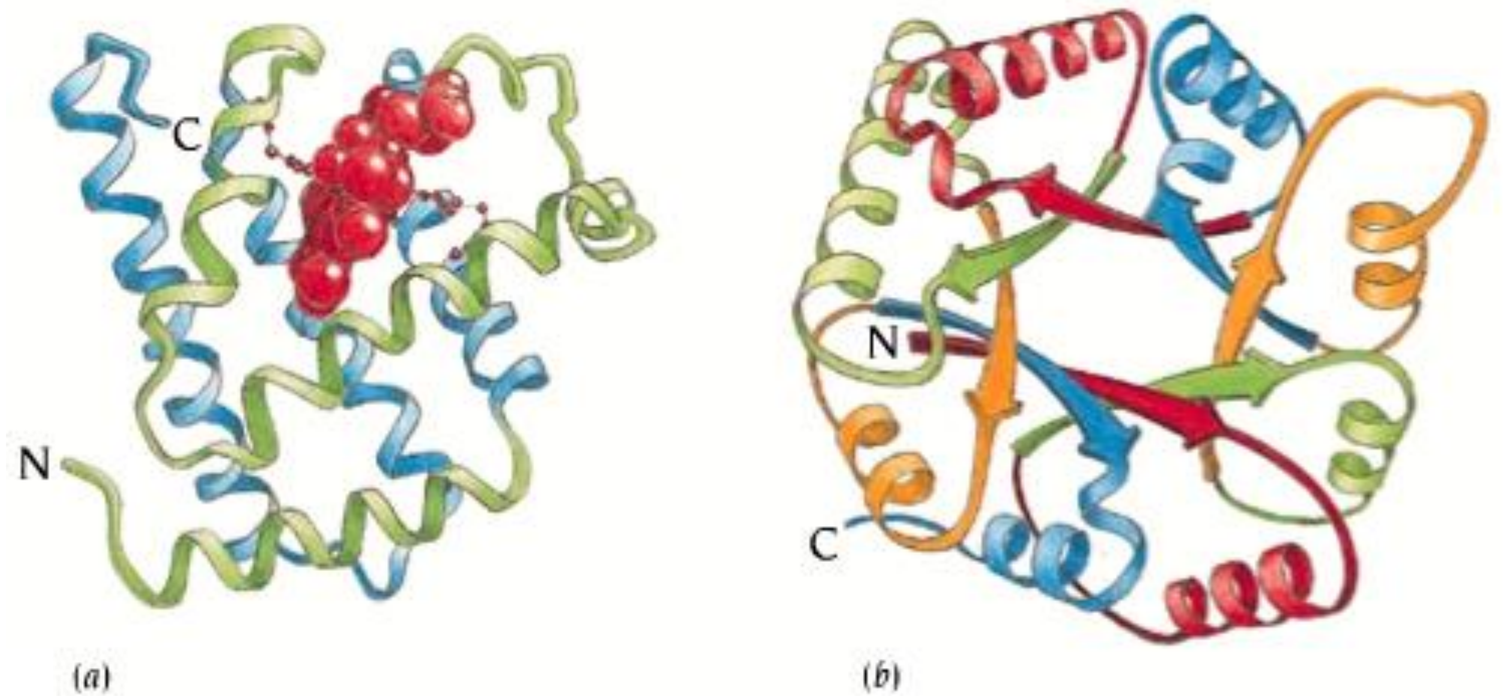
β-pleated sheet: ‘pleated’ because side chains point up and down alternatively

Mixed β-sheet

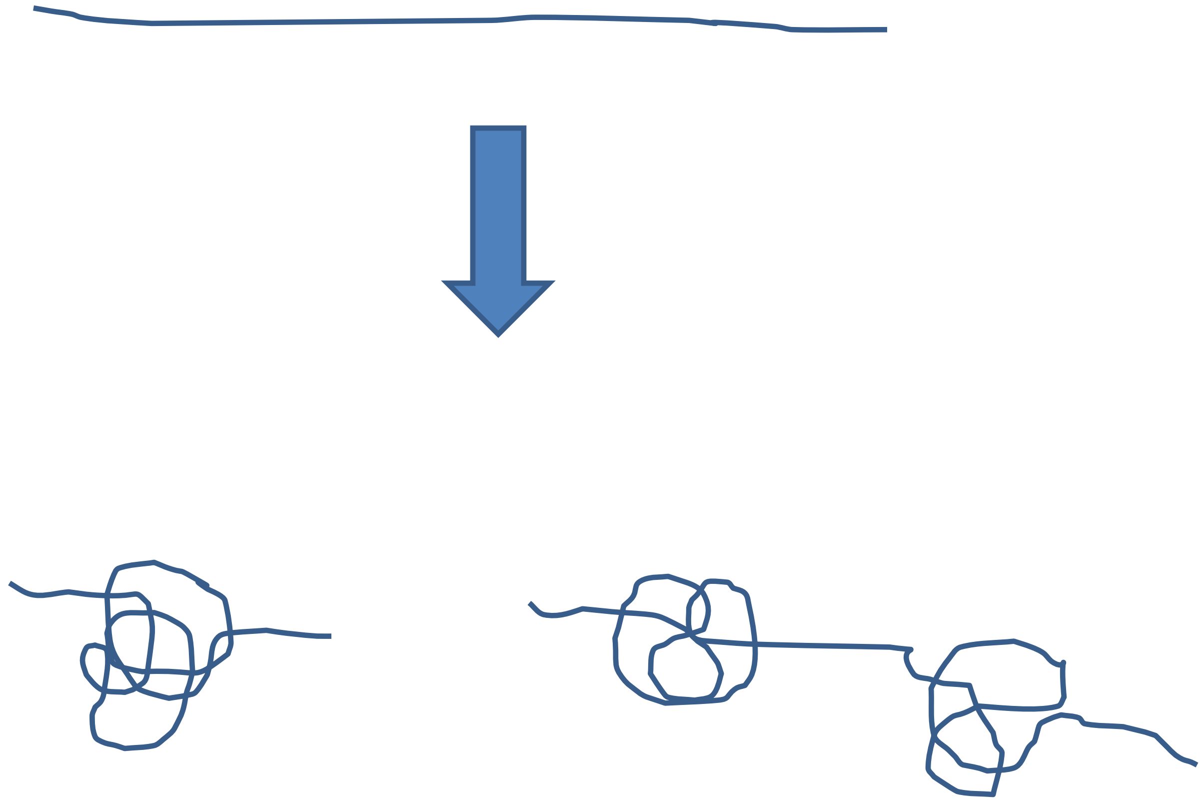


Polypeptide Chains Fold into Several Domains

•Fundamental unit of tertiary structure – DOMAIN
•Domain: polypeptide chain or a part of polypeptide chain that can independently fold into a stable tertiary structure
•Domains are also units of function



Tertiary structure refers to the spatial arrangement of amino acid residues that are far apart in the sequence and to the pattern of disulfide bonds.



Quaternary Structure

Proteins containing more than one polypeptide chain exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a subunit. Quaternary structure refers to the spatial arrangement of subunits and the nature of their interactions.

The simplest quaternary structure is a dimer, consisting of two identical subunits.

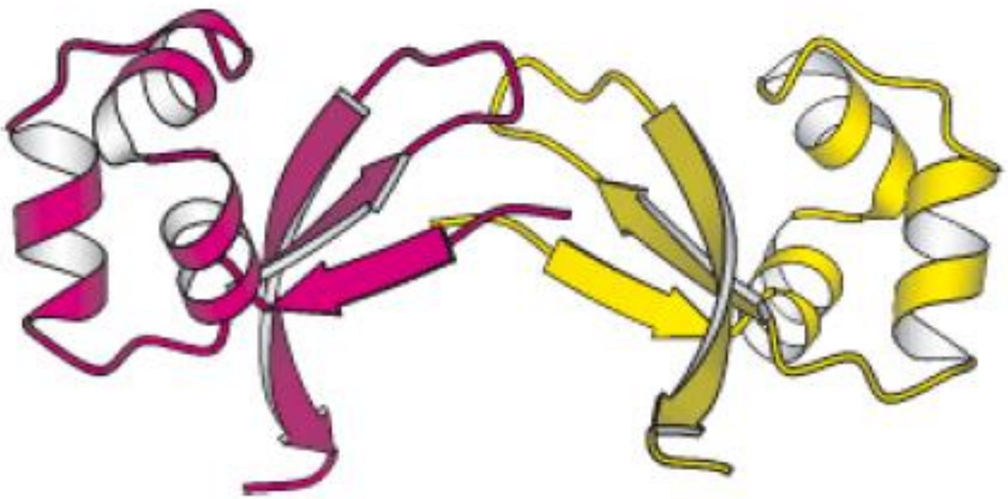
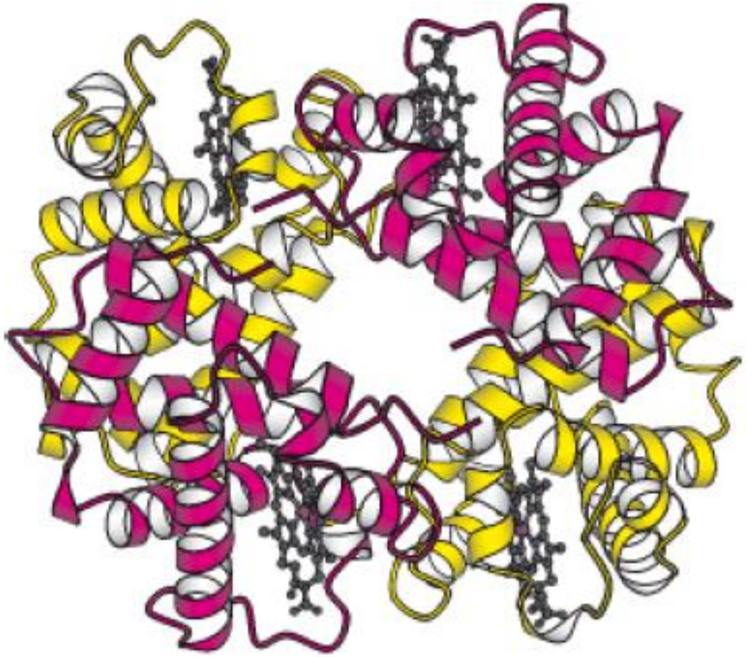
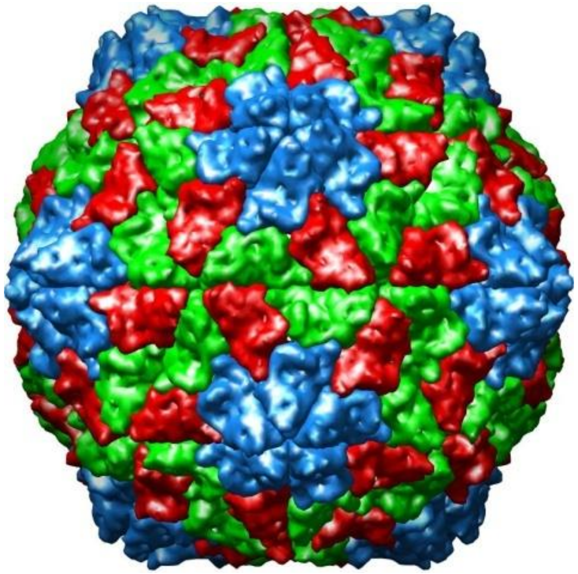


Figure 3.48. Quaternary Structure. The Cro protein of bacteriophage λ is a dimer of identical subunits.

Quaternary Structure (higher order)



The α2β2 tetramer of human haemoglobin. The structure of the two identical α subunits (red) is similar to but not identical with that of the two β subunits (yellow).



Complex Quaternary Structure. The coat of rhinovirus comprises 60 copies of each subunits

Methods to study protein structures

X-ray crystallography

- Prof. Amit K Das

  (http://www.iitkgp.ac.in/department/BT/faculty/bt-amitk)

Structure-guided protein engineering

- Prof. Dibyendu Samanta

  (http://iitkgp.ac.in/department/BS/faculty/bs-dibyendu.samanta)

NMR spectroscopy

- Nuclear magnetic resonance (NMR) spectroscopy
- Prof. Soumya De

  (http://iitkgp.ac.in/department/BS/faculty/bs-somde)

Protein Structure, Function, Kinetics and Energetics

Books Followed:

- How Proteins Work (Mike Williamson)
- Introduction to protein structure (Carl Branden & John Tooze)
- Biochemistry (Lubert Stryer)