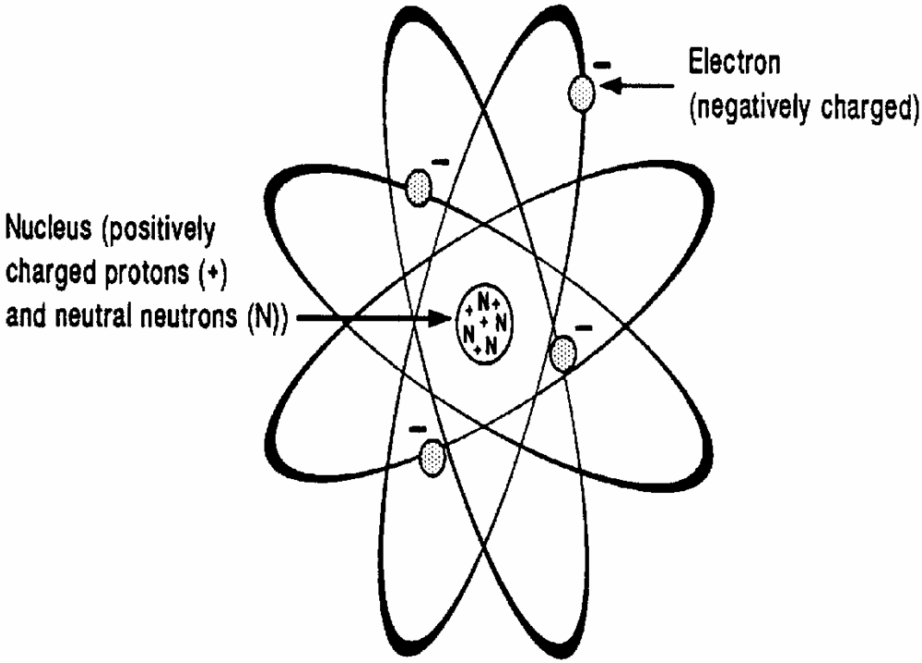
UNIT II

- Structure and functions of :
  - carbohydrates
  - lipids
  - proteins
    - enzymes
    - hormones
  - DNA
  - RNA
- The human genome project
- Genomics
- Sequence databases
- BLAST tool

Chemistry of life: chemical bonds

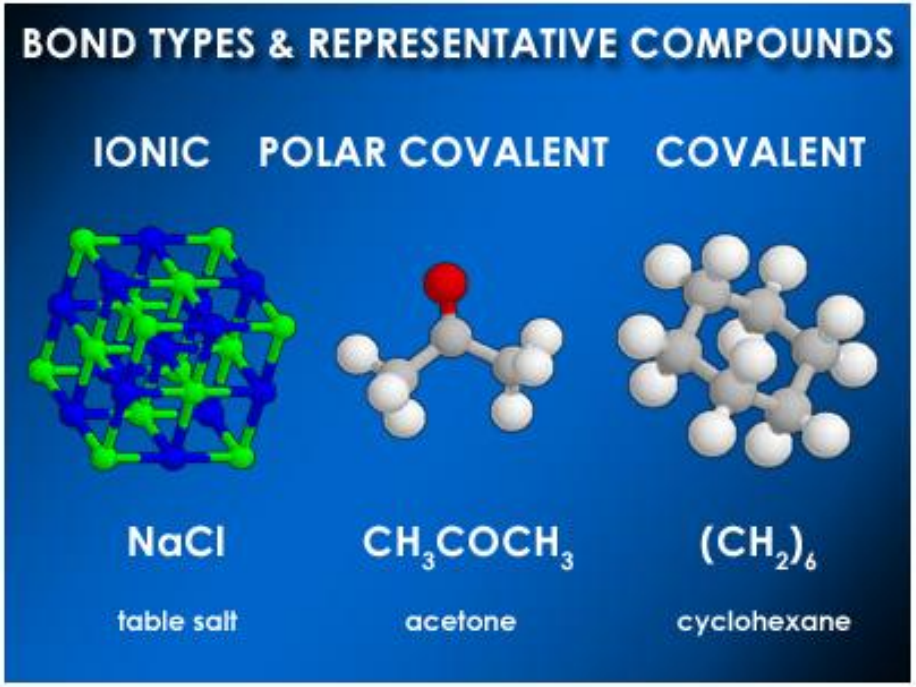
Chemistry of Life

- All matter is built up of simple units called atoms.
- Although the word atom means something that cannot be cut (a = ‘‘without,’’ tom = ‘‘cut’’), these elementary particles are actually made up of many smaller parts, which are themselves further divisible.
- Elements are substances that consist of the same kinds of atoms.
- Compounds consist of units called molecules, which are intimate associations of atoms (in the case of compounds, different atoms) joined in precise arrangements.

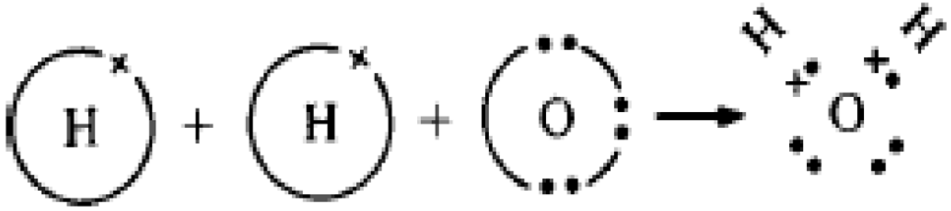


- **Atoms interact with one another to form chemical communities. The tightly knit atoms making up the communal molecules are held together by chemical bonding.**

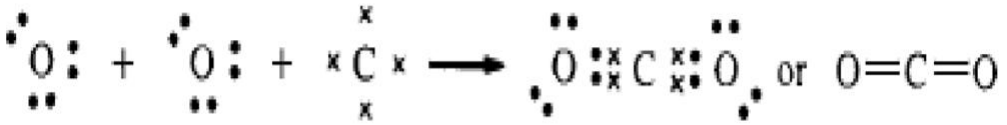
- **One way of achieving this more stable state is for an atom with very few electrons in its outer shell to donate them to an atom with an outer shell that is almost complete.**
  - **The atom that donates the electrons will then have more protons than electrons and assume a positive charge; it is called a cation. The atom receiving the electrons assumes a negative charge and is called an anion.**
  - **These two oppositely charged ions are electrostatically attracted to each other and are said to have an ionic, or polar, bond.**



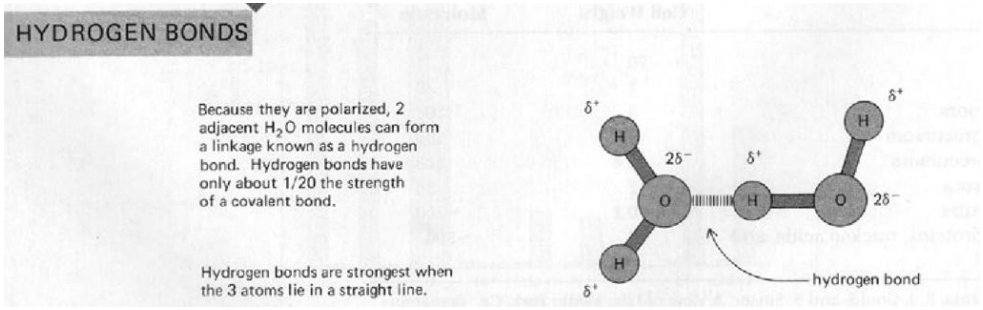
- A second way in which atoms may join with one another to bring about a filling of their outermost shells is by sharing a pair of electrons.
  - The two bonding atoms provide one electron each in creating the shared pair. This pair of electrons forms a covalent bond that holds the two atoms together. It is represented by a solid line in the formula of a compound.

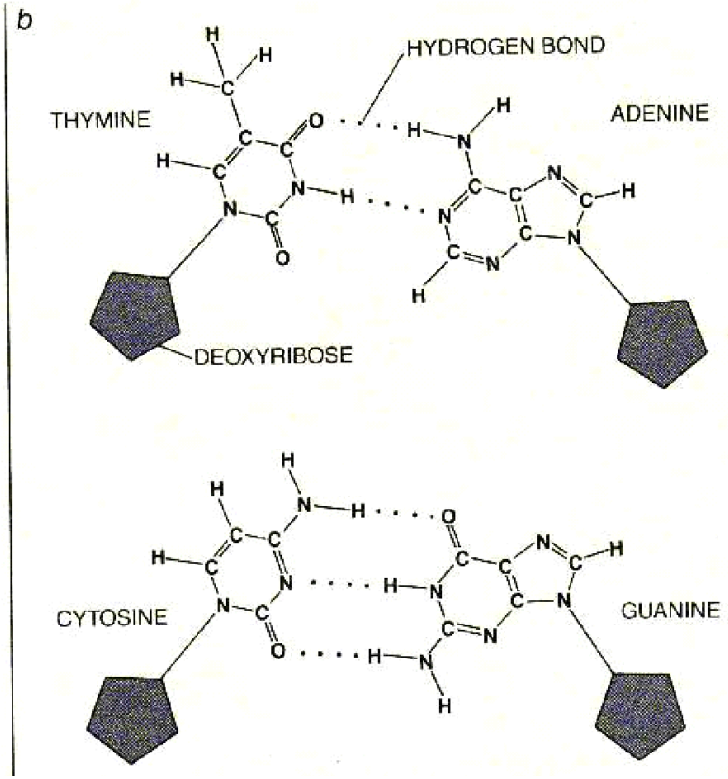


- In many molecules, covalent bonding may occur not just singly (sharing a single pair of electrons), but may involve the formation of double or triple bonds in which two and even three pairs of electrons are shared.
  - These double and triple bonds tend to fix the position of the participating atoms in a rigid manner.



- **Non-covalent bonds (ionic, hydrogen)** are much weaker than **covalent bonds** (electron sharing) and so protein shape can be disrupted especially by temperature, pH , ions (salt).
- It involves more dispersed variations of electromagnetic interactions.
- Critical in maintaining the three-dimensional structure of large molecules, such as proteins and nucleic acids
- There are four commonly mentioned types of non-covalent interactions: hydrogen bonds, ionic bonds, van der Waals forces, and hydrophobic interactions.
  - The noncovalent interactions hold together the two strands DNA in the double helix, stabilize secondary and tertiary structures of proteins, and enable enzyme-substrate binding and antibody-antigen association.





Biochemistry and Human biology

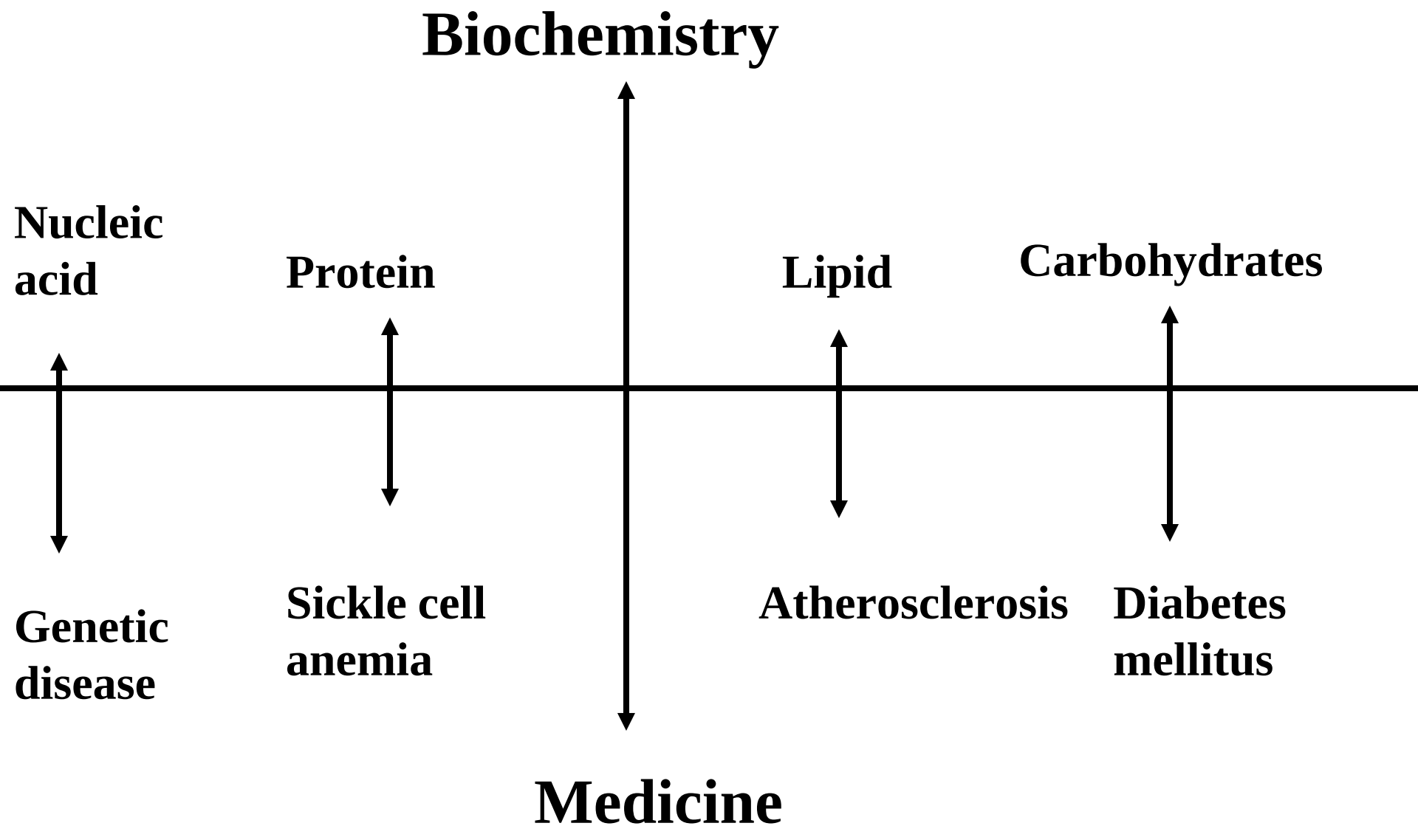
Biochemistry:
Where Chemistry & Biology Meet

- **Living things require millions of chemical reactions just to survive.**
- **Metabolism = all the chemical reactions occurring in the body.**
- **Organic molecules:**
  - **usually associated with living things.**
  - **always contain CARBON.**
  - **are “large” molecules, with many atoms**
  - **always have covalent bonds (share electrons)**

Biochemistry and Human Biology

- **Biochemistry**: Science concerned with the chemical constituents of living cells and with the reaction and process that they undergo.
  - Complete understanding at the molecular level of all the chemical processes associated with living cells
  - An appreciation of the biochemistry of less complex form of life is often direct relevance to human biochemistry

- **Reciprocal relationship between biochemistry and medicine has stimulated mutual advance**
  - Biochemistry studies have illuminated many aspects of health & disease


Biochemistry
Nucleic acid
Protein
Lipid
Carbohydrates
Genetic disease
Sickle cell anemia
Atherosclerosis
Diabetes mellitus
Medicine

S. No.	Disease	Causes
1	Scurvy rickets	deficiencies of vitamins C and D respectively
2	Atherosclerosis	genetic, dietary, environmental factors
3	Cystic fibrosis	mutation in the gene coding the CFTR protein (Cystic fibrosis transmembrane conductance regulator, a protein involved in the transport of chloride ions across cell membranes)
4	Cholera	exotoxin of vibrio cholera
5	Diabetes mellitus type I	genetic and environmental factors resulting in deficiency of insulin
6	Phenylketonuria	mainly mutation in the gene coding phenylalanine hydroxylase

Carbon-based Molecules

- Although a cell is mostly water, the rest of the cell consists mostly of carbon-based molecules

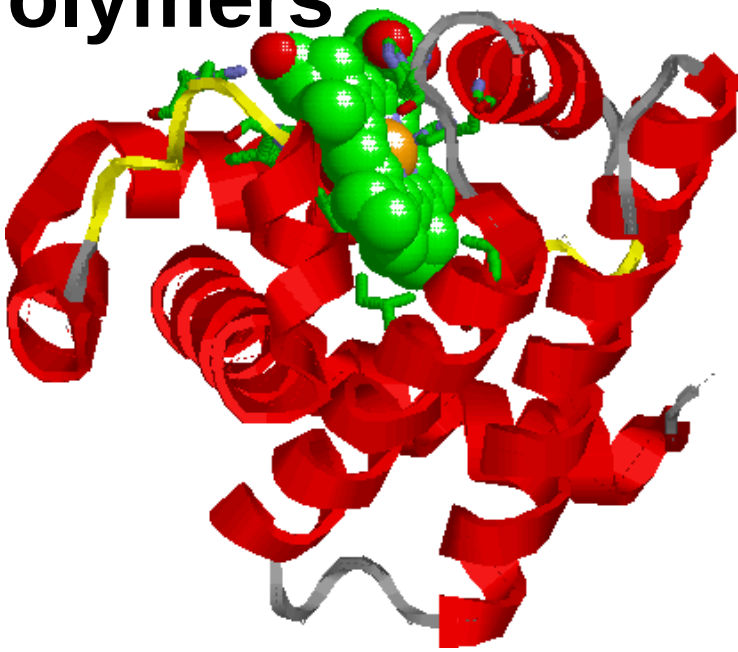
Organic chemistry is the study of carbon compounds

Carbon is a Versatile Atom

- **It has four electrons in an outer shell that holds eight**

Carbon can share its electrons with other atoms to form up to four covalent bonds

Giant Molecules - Polymers



- Large molecules are called polymers
- Polymers are built from smaller molecules called monomers
- Biologists call them macromolecules

Macromolecules in Organisms

- There are four categories of large molecules in cells:

Carbohydrates

Lipids

Proteins

Nucleic Acids

Examples of Polymers

- Proteins

Lipids

Carbohydrates

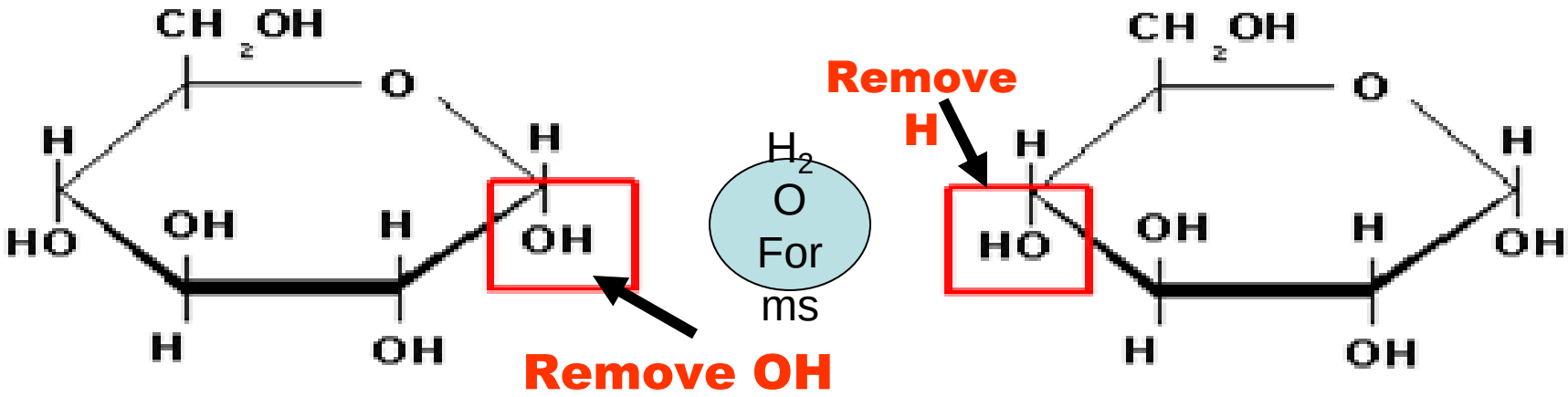
Nucleic Acids

Carbohydrates

- **Carbohydrates include:**
  - **Small sugar molecules in soft drinks**
  - **Long starch molecules in rice, wheat, pasta and potatoes**

Linking Monomers

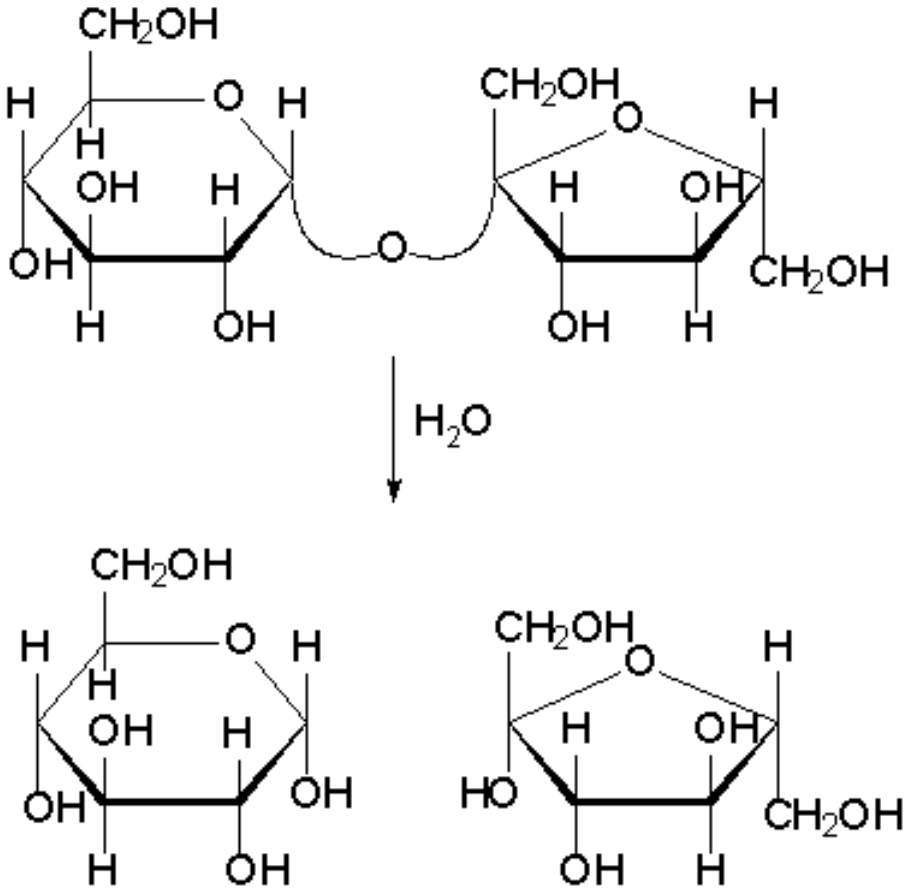
Cells link monomers by a process called condensation or dehydration synthesis (removing a molecule of water)



This process joins two sugar monomers to make a double sugar

Breaking Down Polymers

- **Cells break down macromolecules by a process called hydrolysis (adding a molecule of water)**



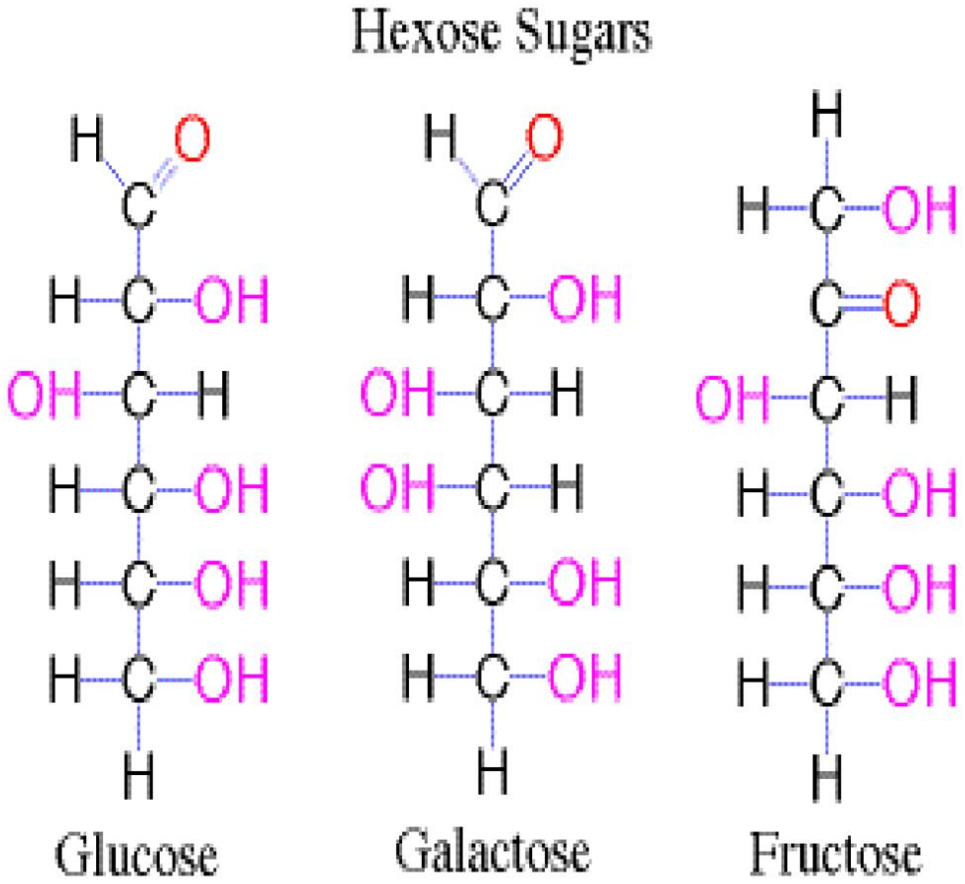
Water added to split a double sugar

Monosaccharides

- Called simple sugars

Include glucose, fructose, & galactose

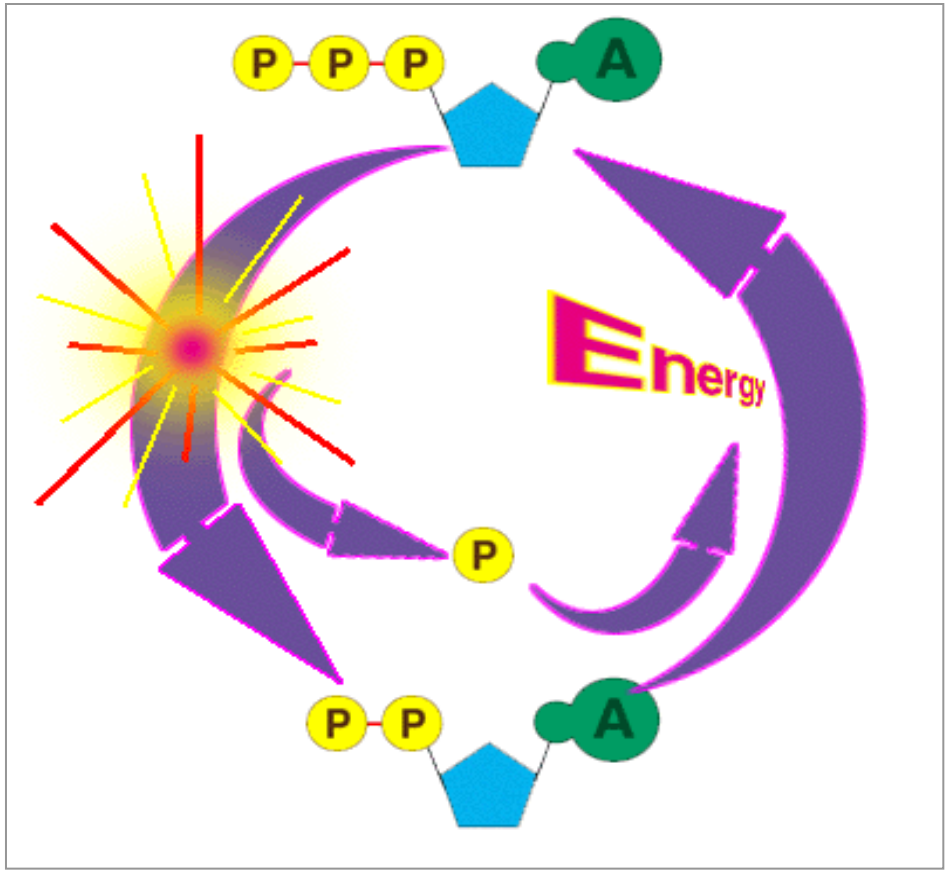
Have the same chemical, but different structural formulas



$C_6H_{12}O_6$

Cellular Fuel

- **Monosaccharides are the main fuel that cells use for cellular work**



ATP

Disaccharides

- **A disaccharide is a double sugar.**

They're made by joining two monosaccharides

Involves removing a water molecule (condensation)

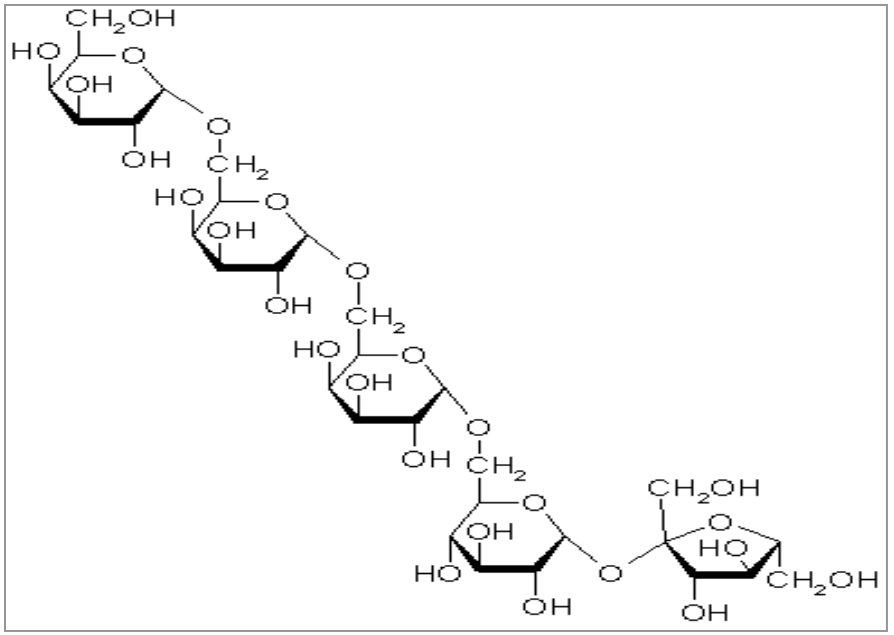
Bond called a GLYCOSIDIC bond

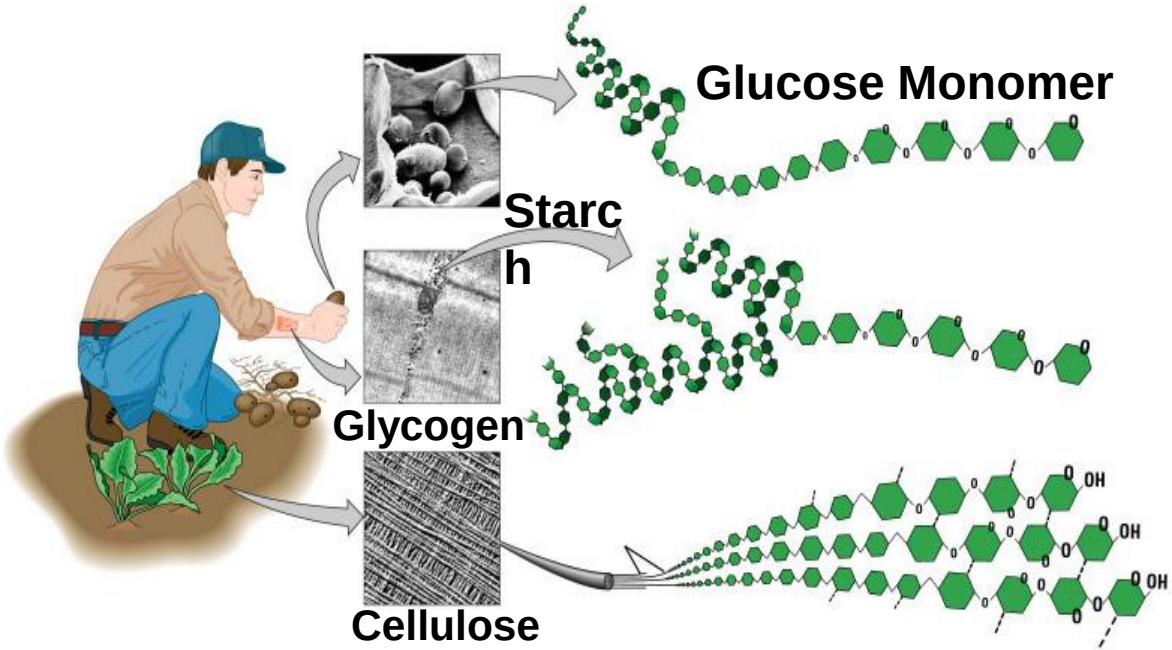
Polysaccharides

- **Complex carbohydrates**

Composed of many sugar monomers linked together

Polymers of monosaccharide chains

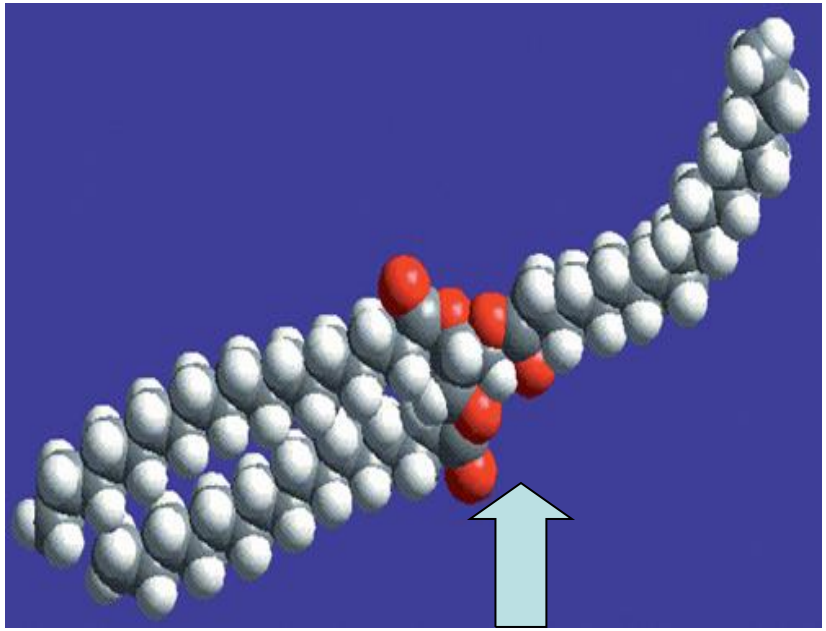




Lipids

- **Lipids are hydrophobic –"water fearing"**
- **Do NOT mix with water**
- **Includes fats, waxes, steroids, & oils**

- **Fats store energy, help to insulate the body, and cushion and protect organs**



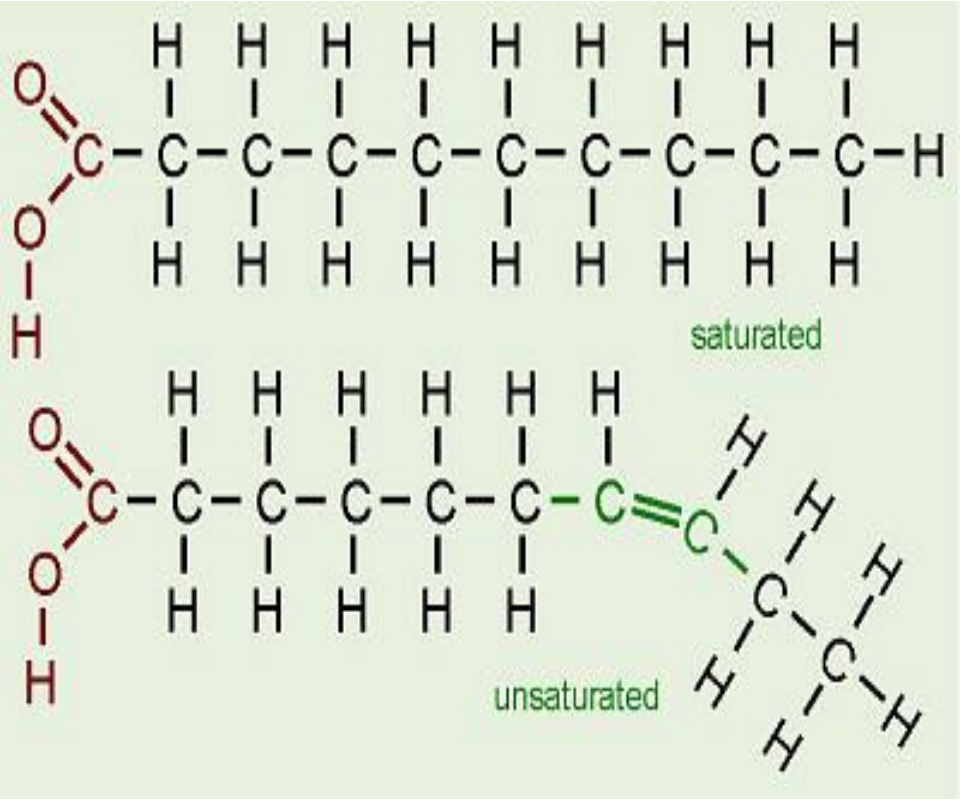
FAT MOLECULE

Types of Fatty Acids

Saturated fatty acids have the maximum number of hydrogens bonded to the carbons (all single bonds between carbons)

Unsaturated fatty acids have less than the maximum number of hydrogens bonded to the carbons (a double bond between carbons)

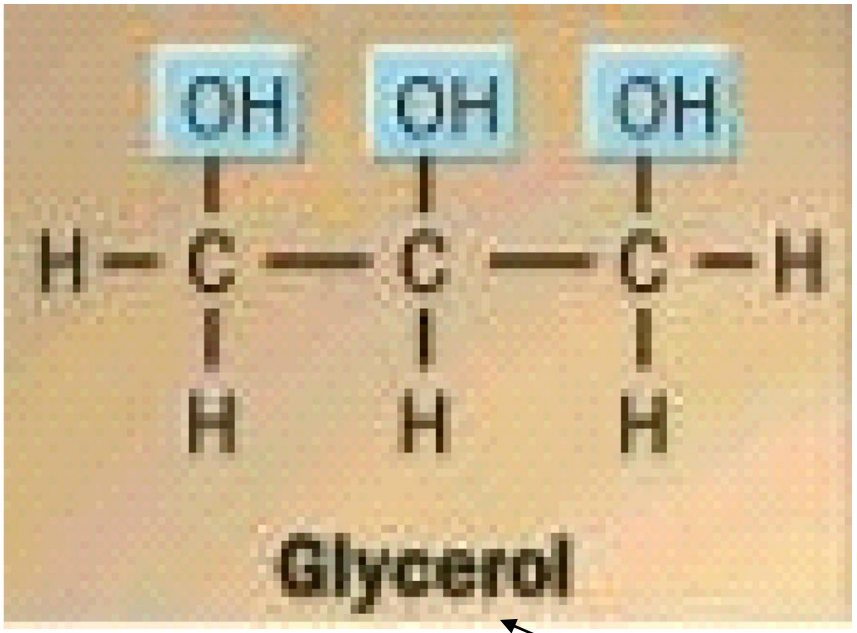
Single Bonds in Carbon chain



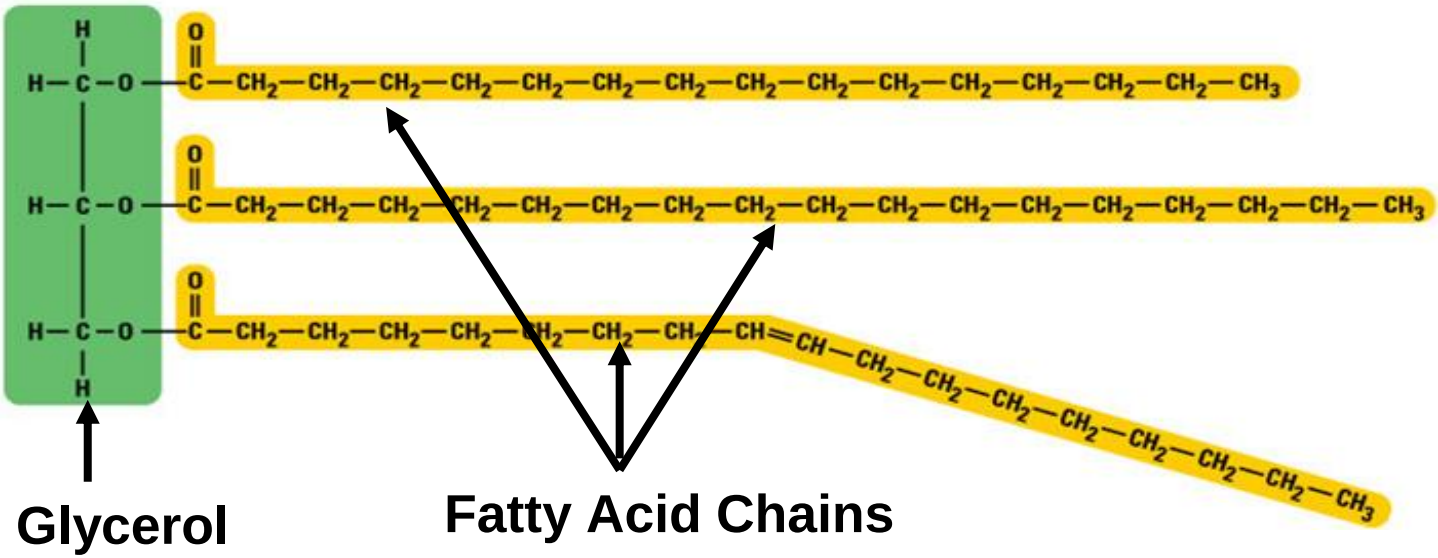
Double bond in carbon chain

Triglyceride

- Monomer of lipids
- **Composed of Glycerol & 3 fatty acid chains**
- **Glycerol forms the "backbone" of the fat**

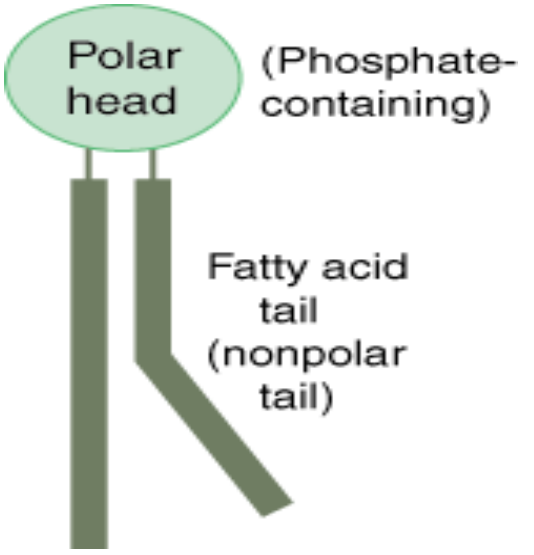


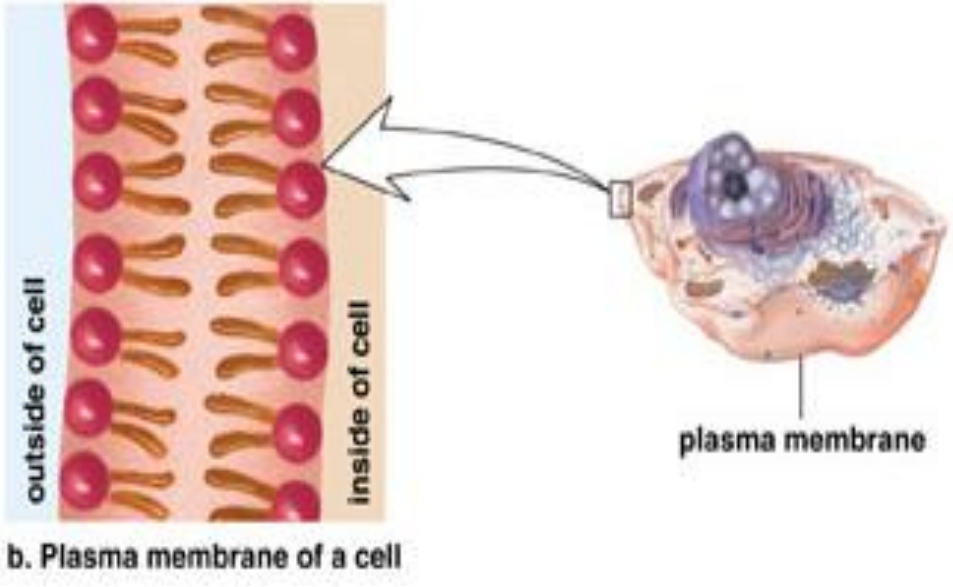
Organic Alcohol
(-OL ending)

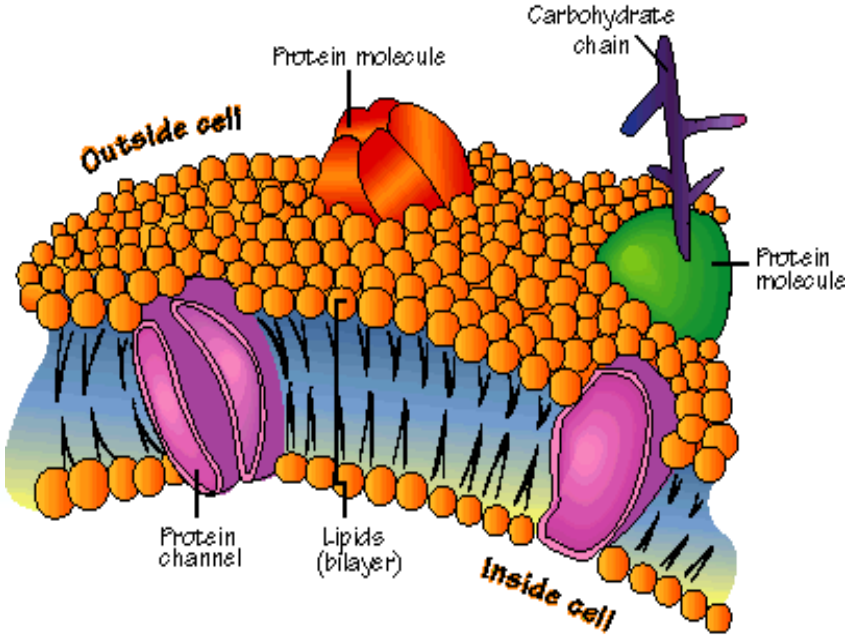


Lipids & Cell Membranes

- Cell membranes are made of lipids called **phospholipids**
- Phospholipids have a **head** that is polar & attract water **(hydrophilic)**
- Phospholipids also have **2 tails** that are nonpolar and do not attract water **(hydrophobic)**







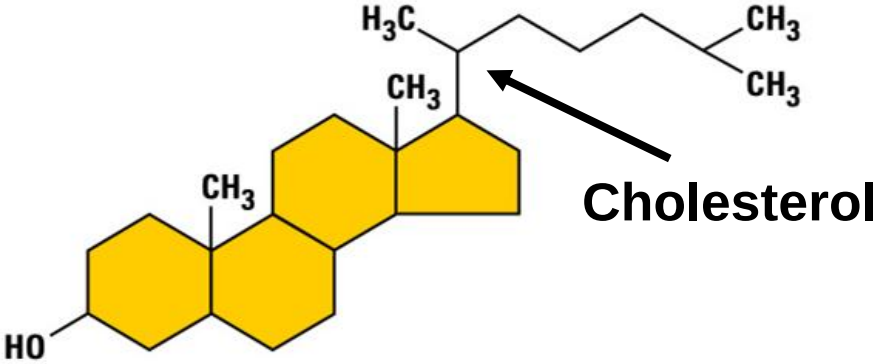
Cell membrane with proteins & phospholipids

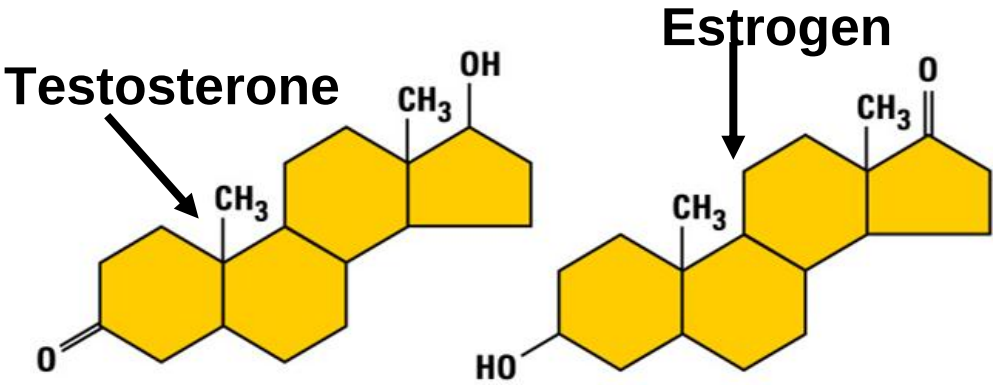
Steroids

- The carbon skeleton of steroids is bent to form 4 fused rings

- Cholesterol is the "base steroid" from which your body produces other steroids

- Estrogen & testosterone are also steroids



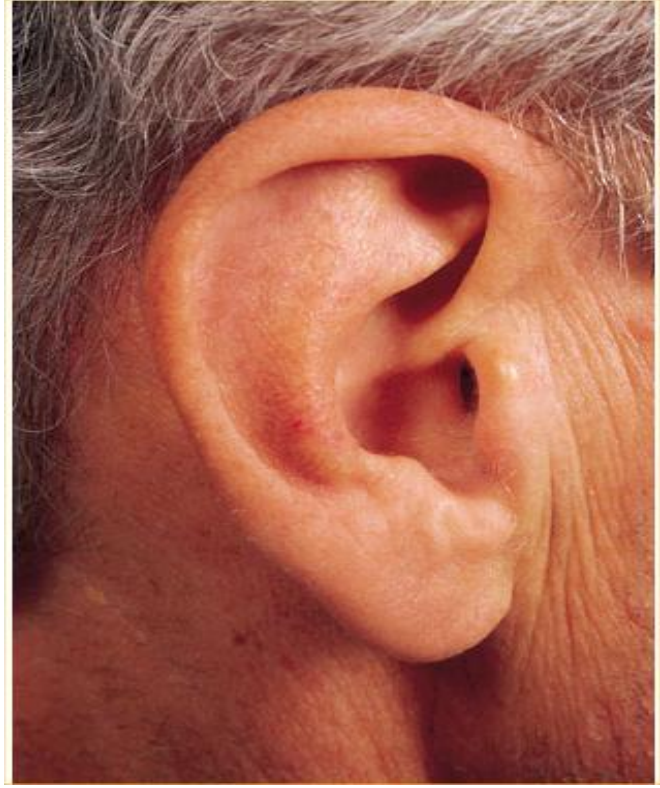


Synthetic Anabolic Steroids

- They are variants of testosterone
- Some athletes use them to build up their muscles quickly
- They can pose serious health risks

Waxes

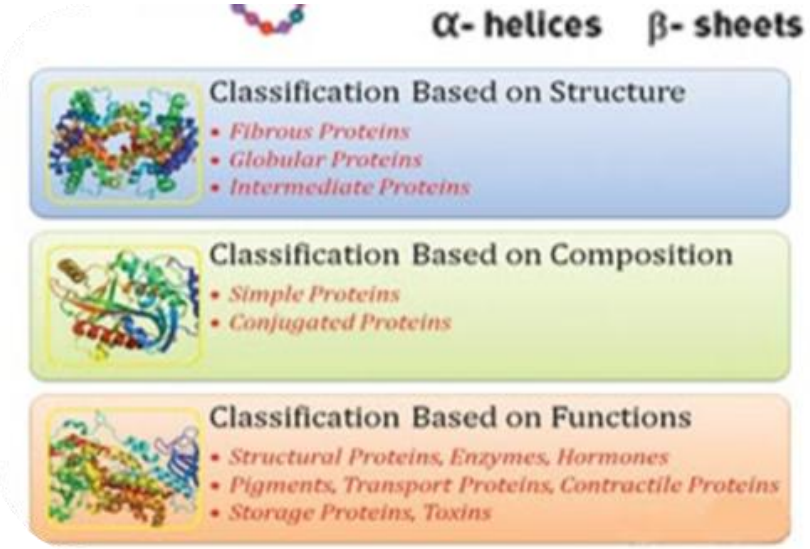
- A wax is a lipid because of its nonpolar solubility characteristics as well as its extremely hydrophobic (water-hating) properties.
- Waxes are composed of a single, highly complex alcohol joined to a longchain fatty acid in a typical ester linkage.
- Waxes are important structural lipids often found as protective coatings on the surfaces of leaves, stems, hair, skin, etc.
- They provide effective barriers against water loss and in some situations make up the rigid architecture of complex structures such as the honeycomb of the beehive.
- They serve a commercial use as well, in furniture polish, automobile coating compounds, and floor finishes.

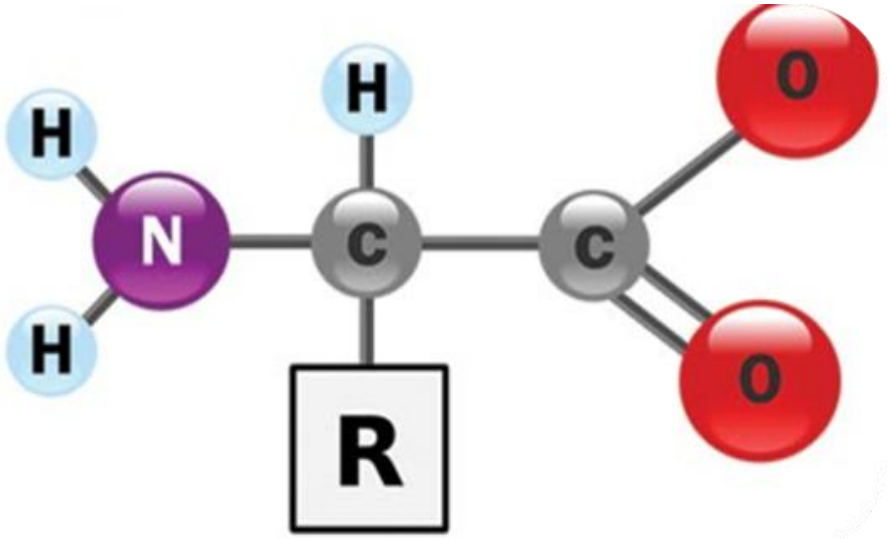


Proteins

Proteins —hydrolysis→ Peptides —hydrolysis→ Amino acids

- **Proteins are polymers made of monomers called amino acids**
- **All proteins are made of 20 different amino acids linked in different orders**
- **Proteins are used to build cells, act as hormones & enzymes, and do much of the work in a cell**





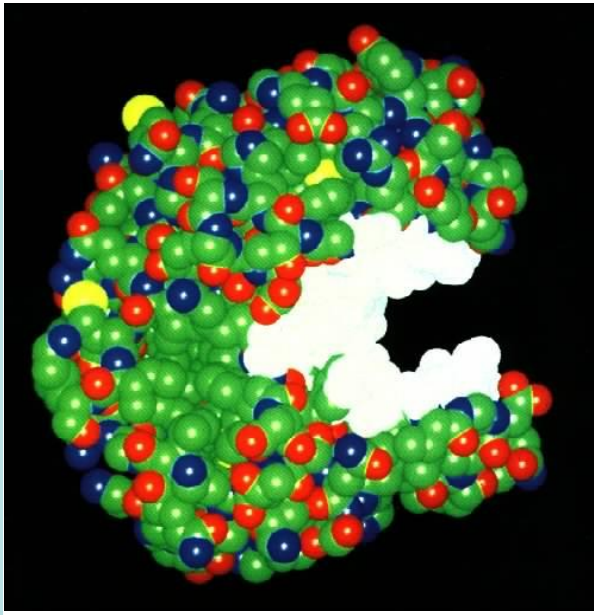
What Are Enzymes?

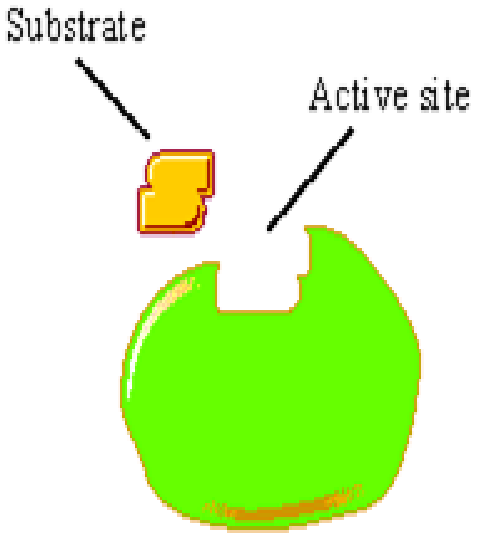
Most enzymes are

- **Proteins (tertiary and quaternary structures)**
- **Act as Catalyst to accelerates a reaction**
-
- **Not permanently changed in the process**
- **Are specific for what they will catalyze**
- **Are Reusable**
- **Name Ends in *–ase***
  - ***.Sucrase***
  - ***-Lactase***
  - ***-Maltase***

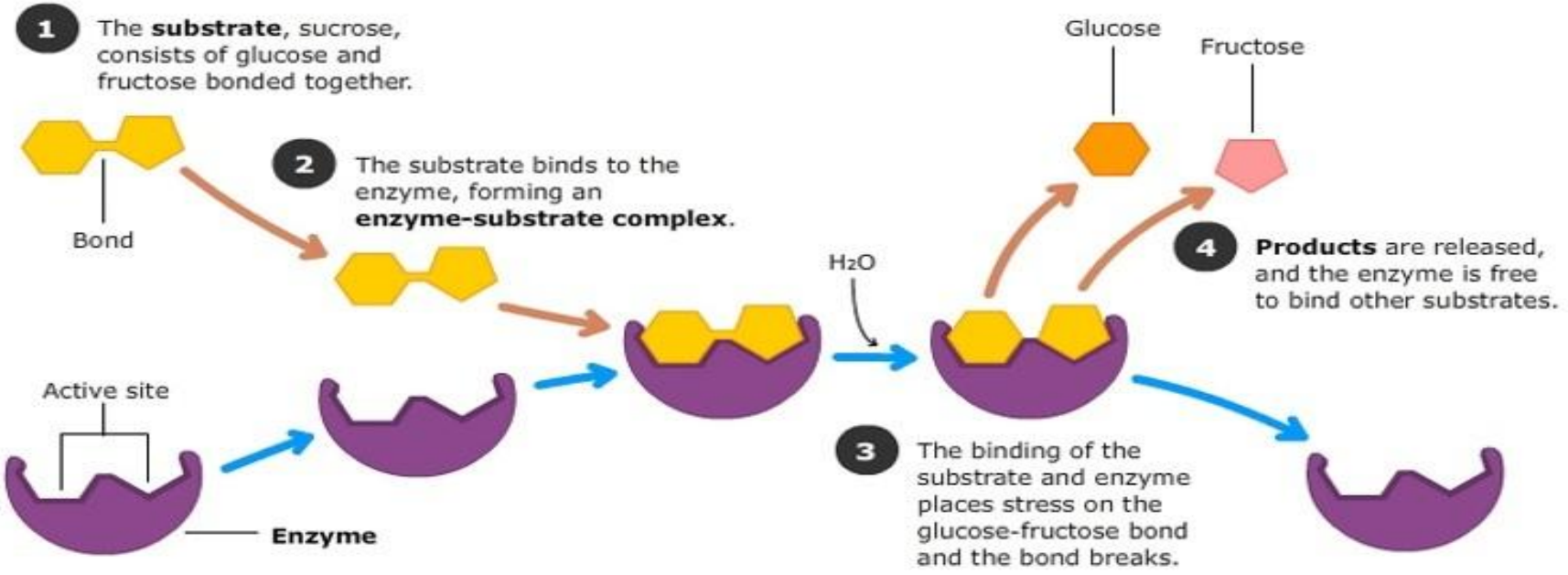
Why Enzymes?

- **Natural catalysts**
- **Speed: 10^{16} over un-catalyzed rates!**
- **Specificity: only the desired reaction occurs**
- **Permit reactions under mild conditions**





- Since most reactions in your body's cells need special enzymes, each cell contains thousands of different enzymes.
- Enzymes let chemical reactions in the body happen millions of times faster than without the enzyme. Because enzymes are not part of the product, they can be reused again and again.



© 2007-2011 The University of Waikato | www.sciencelearn.org.nz

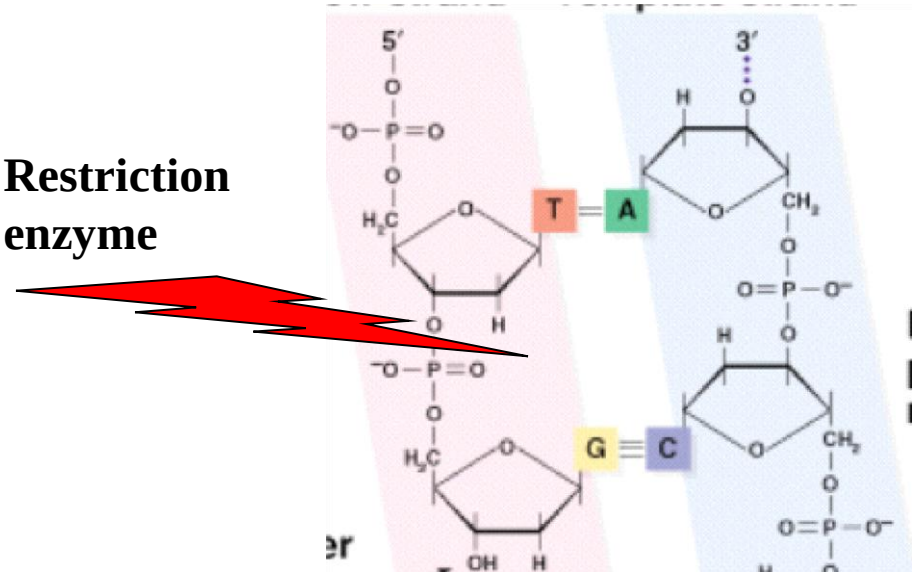
Example: Restriction enzymes

Recognizes specific base sequences in double-helical DNA and cleave, at specific places, both strands of a duplex containing the recognized sequences.

Restriction enzymes recognize specific bases pair sequences in DNA called restriction sites and cleave the DNA by hydrolyzing the phosphodiester bond.

Cut occurs between the 3’ carbon of the first nucleotide and the phosphate of the next nucleotide.

Restriction fragment ends have 5’ phosphates & 3’ hydroxyls.



Most restriction enzymes occur naturally in bacteria.

Protect bacteria against viruses by cutting up viral DNA.

Bacteria protects their DNA by modifying possible restriction sites (methylation).

More than 400 restriction enzymes have been isolated.

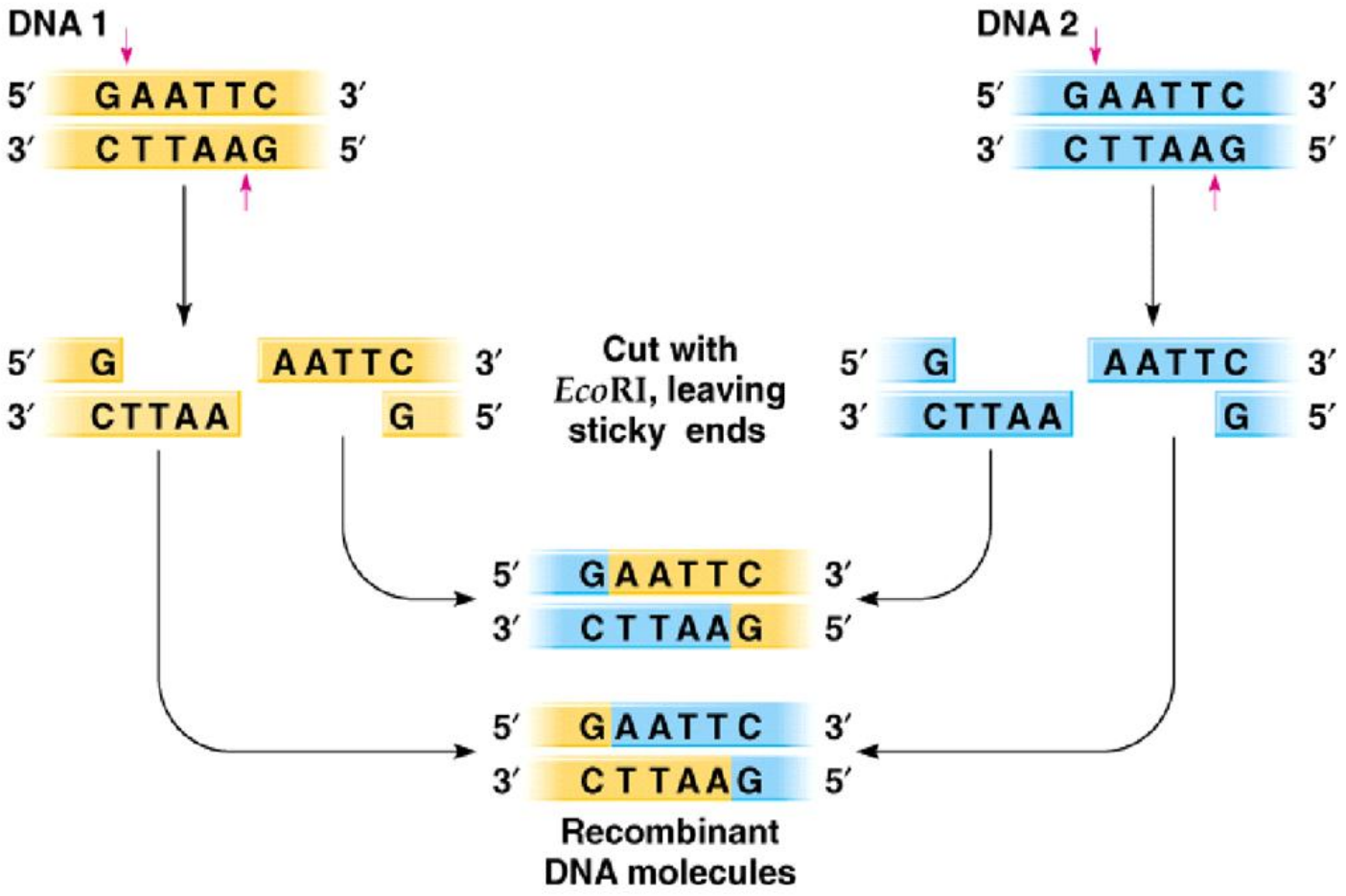
Names typically begin with 3 italicized letters.

Enzyme	Source
***Eco*RI**	***E. coli* RY13**
***Hin*dIII**	***Haemophilus influenzae* Rd**
***Bam*HI**	***Bacillus amyloliquefaciens* H**

Many restriction sites are palindromes of 4-, 6-, or 8-base pairs.

Short restriction site sequences occur more frequently in the genome than longer restriction site sequences, e.g., (1/4)n.

Cut and ligate 2 DNAs with *Eco*RI ---> recombinant DNA



Applications of Recombinant DNA technologies

- Pharmaceutical products
  - insulin – cheaper and safer compared to animal insulin
  - vaccine sub-unit (against hepatitis B) – safer since will not be infected by pathogens
  - DNA of vaccines against malaria, influenza etc.
- Gene therapy
  - replacing defective or missing gene with normal gene using adeno~ and retrovirus as vector
- Gene silencing
  - known as RNA interference (RNAi) using dsRNA called short interfering RNA (siRNA) that target specific gene (mRNA) and degrade it

Hormones

- Cells in multi-cellular organisms communicate with one another to coordinate their growth and metabolism;
- Cell to cell communicate is mainly via Extracellular signaling molecules or Hormones;
- Hormones carry information from Sensor Cells, that sense changes in the environment, to Target Cells that respond to the changes;
- Hormones tend to coordinate various metabolic processes in the body;

Examples

- **INSULIN:**
  - Insulin is a Protein Hormone secreted by Beta cells in Islets of Langerhans in Pancreas,
  - Insulin is a major hormone that regulates Blood Glucose level,
  - Insulin is an Hydrophilic (Lipophobic) hormone, thus it acts via membrane receptors on target cells; • Main target cells: Skeletal Muscle & Adipose tissue

Lack of insulin causes increase in blood sugar level called diabetes

GLUCAGON:

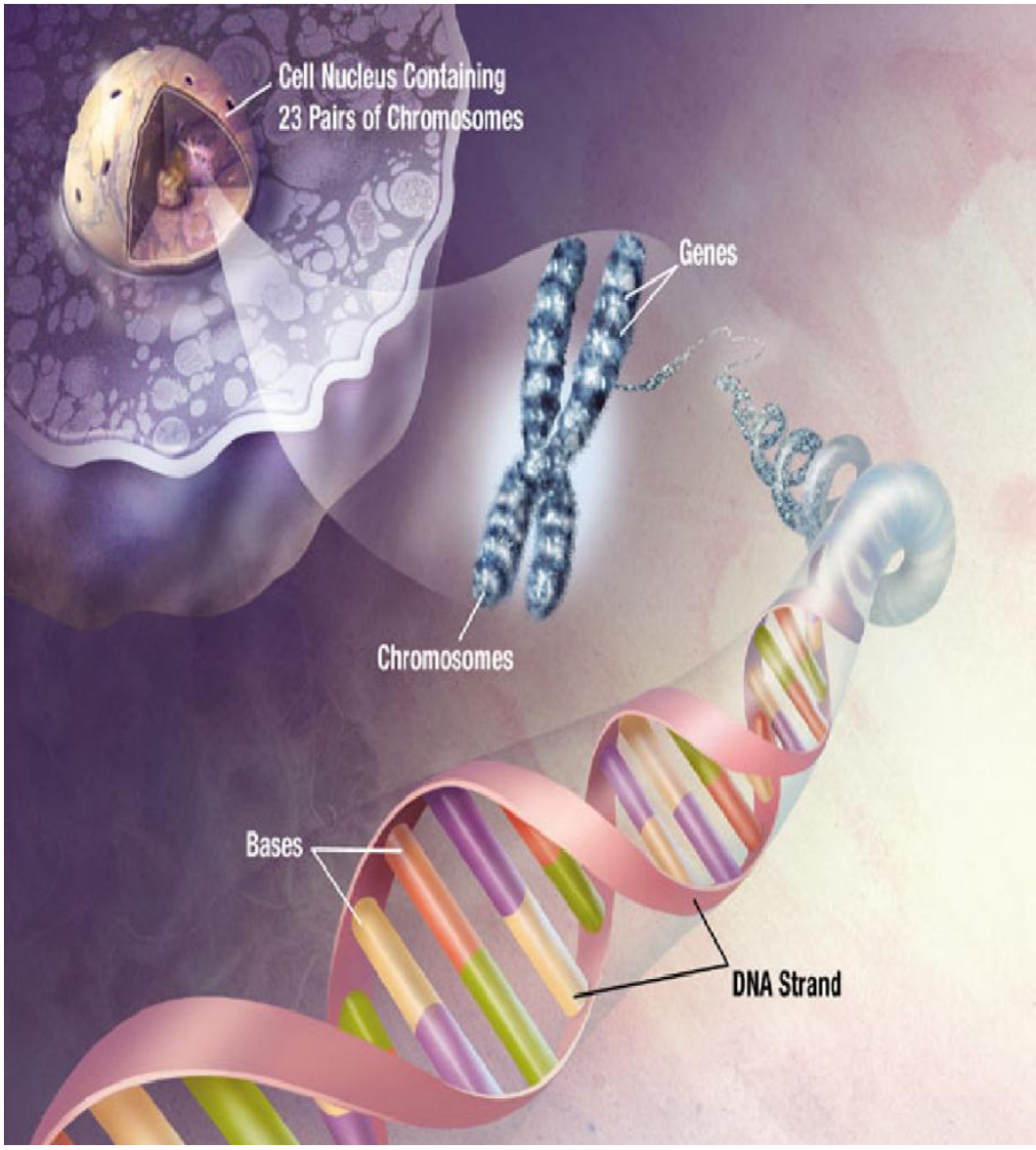
- Glucagon is a hormone produced by Alpha cells in the Pancreas;
- Glucagon is an Insulin Counter-Regulatory Hormone,
- Action of Glucagon is to increase Blood Glucose Level from Low to Normal,
- Glucagon acts mainly in the Liver to stimulate the breakdown of Glycogen to Glucose, which is then released into the blood;

Production of Glucagon is stimulated by:

- Hypoglycemia (Low Glucose level in blood) • Increase absorption of Amino Acids in the blood (as occurs after a protein-rich meal),
- High Blood Glucose Level Inhibits the production and release of Glucagon,

Nucleic Acids

- Store hereditary information
  - **Contain information for making all the body's proteins**
  - **Two types exist --- DNA & RNA**

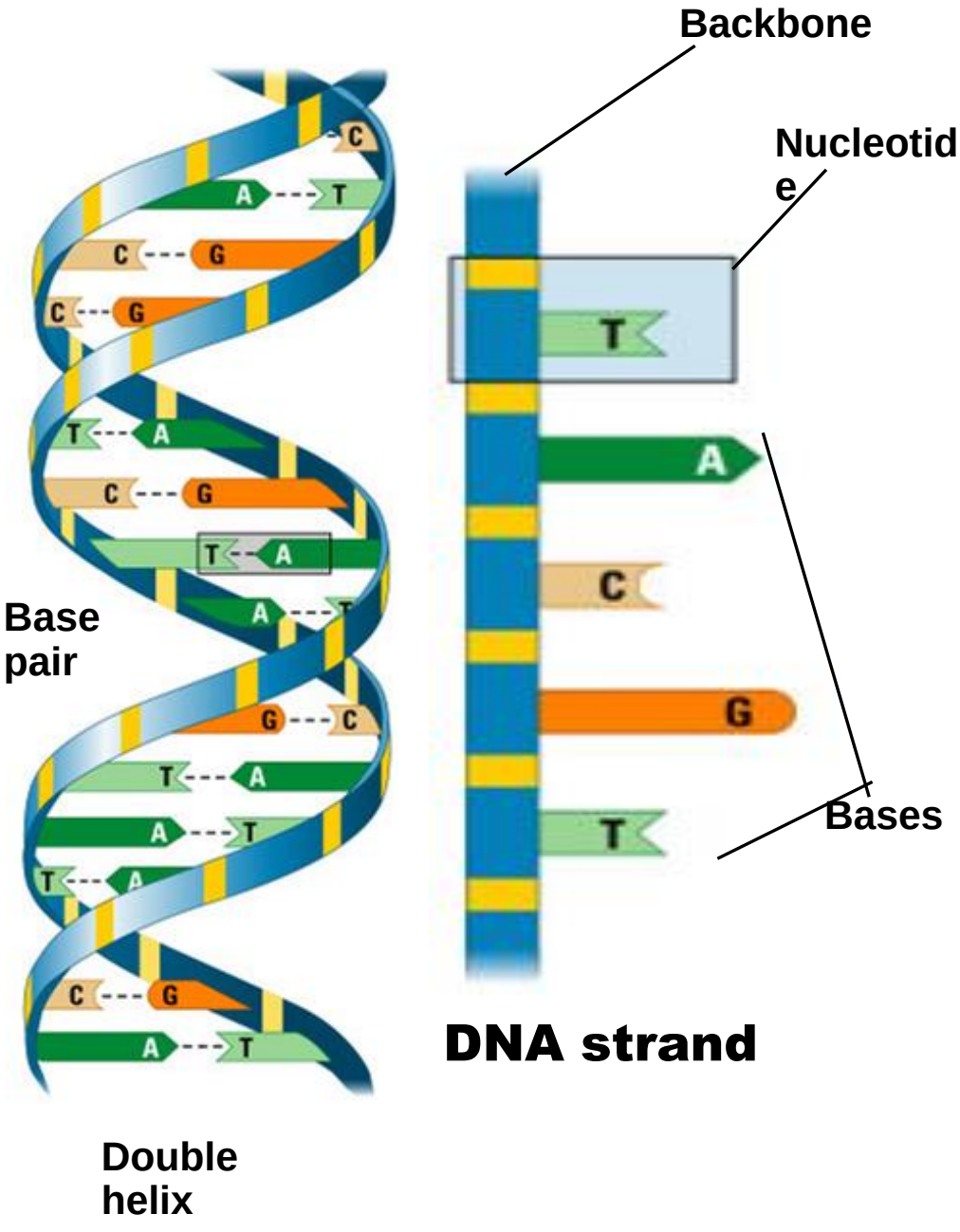


DNA-Deoxyribonucleic acid

- **Two strands of DNA join together to form a double helix**

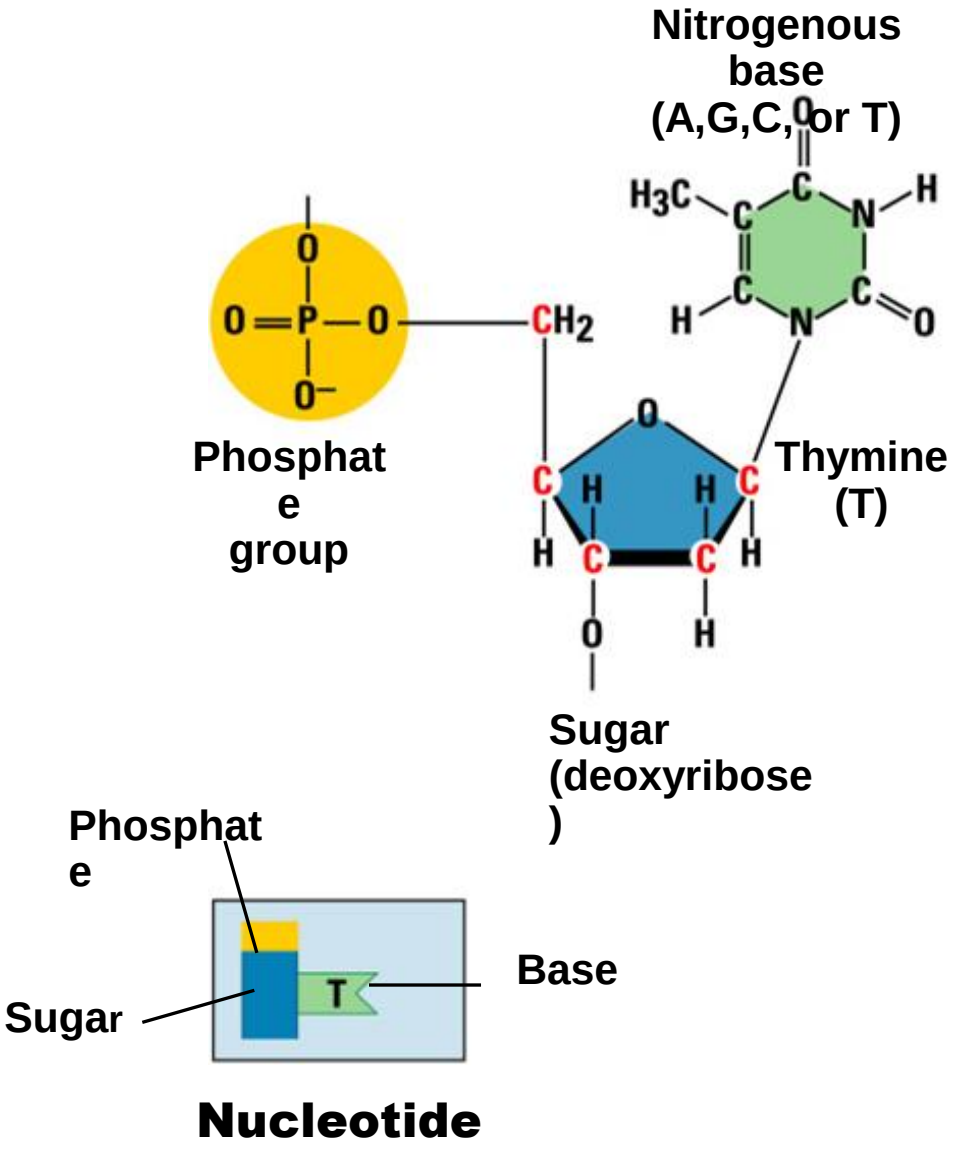
- **Nucleotides form long chains called DNA**

- **Nucleotides are joined by sugars & phosphates on the side**



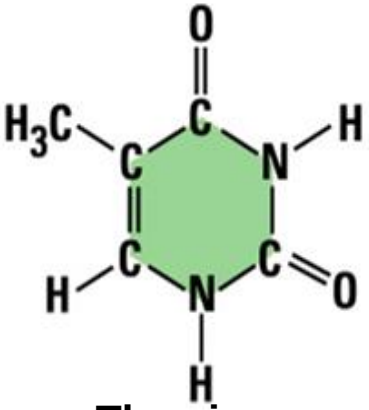
Nucleic Acids

Nucleic acids are polymers of nucleotides

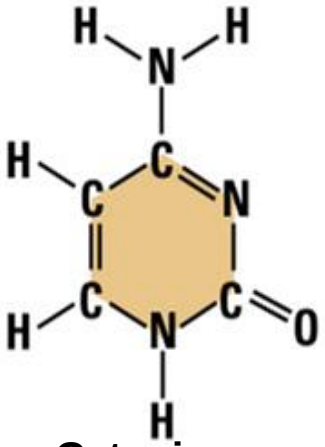


Bases

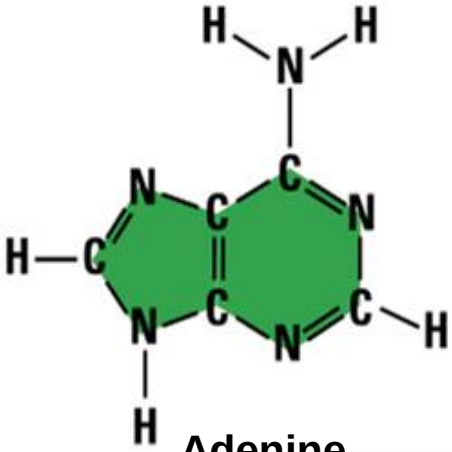
- **Each DNA nucleotide has one of the following bases:**
  - **Adenine (A)**
  - **Guanine (G)**
  - **Thymine (T)**
  - **Cytosine (C)**



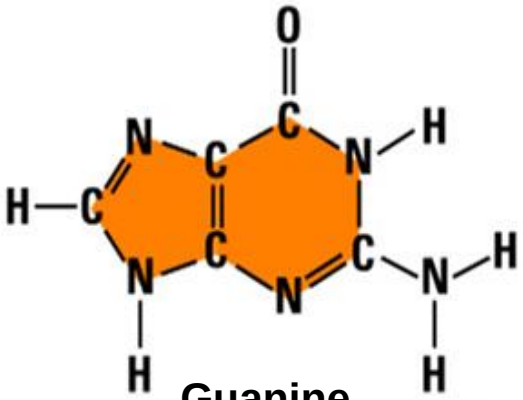
Thymine
(T)



Cytosine
(C)



Adenine
(A)

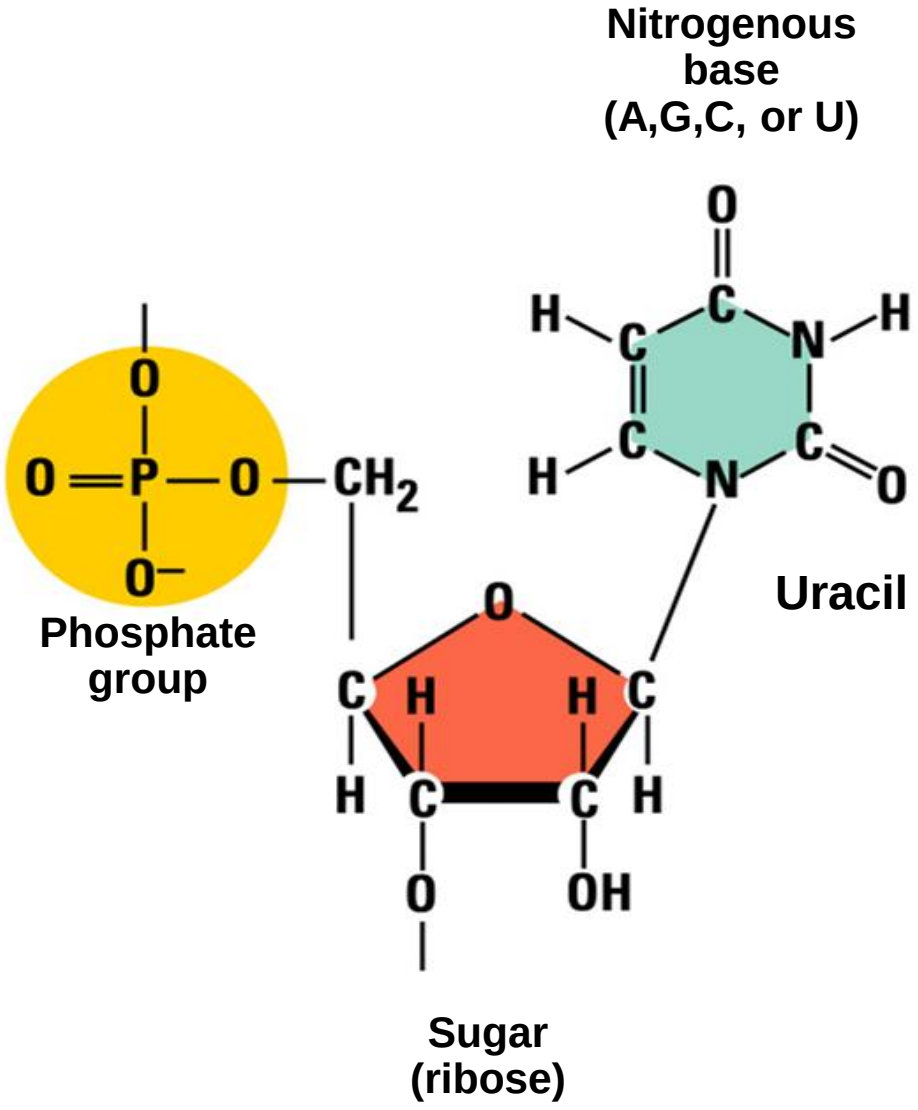


Guanine
(G)

RNA – Ribonucleic Acid

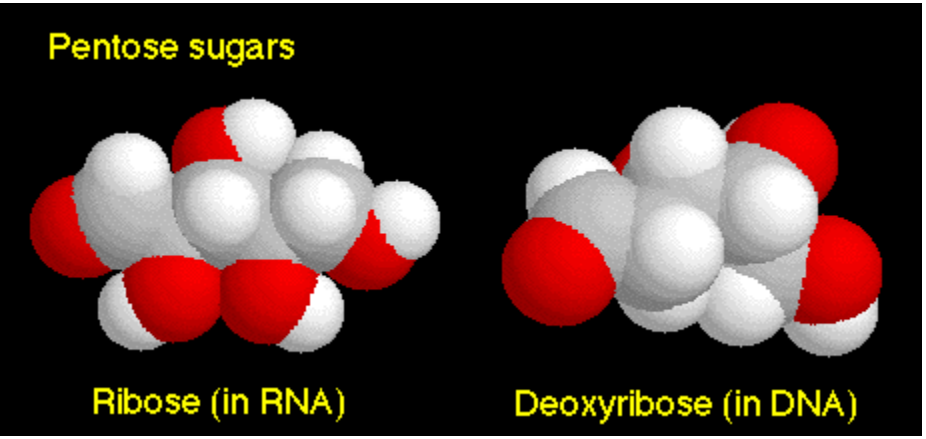
- **Ribose sugar has an extra –OH or hydroxyl group**

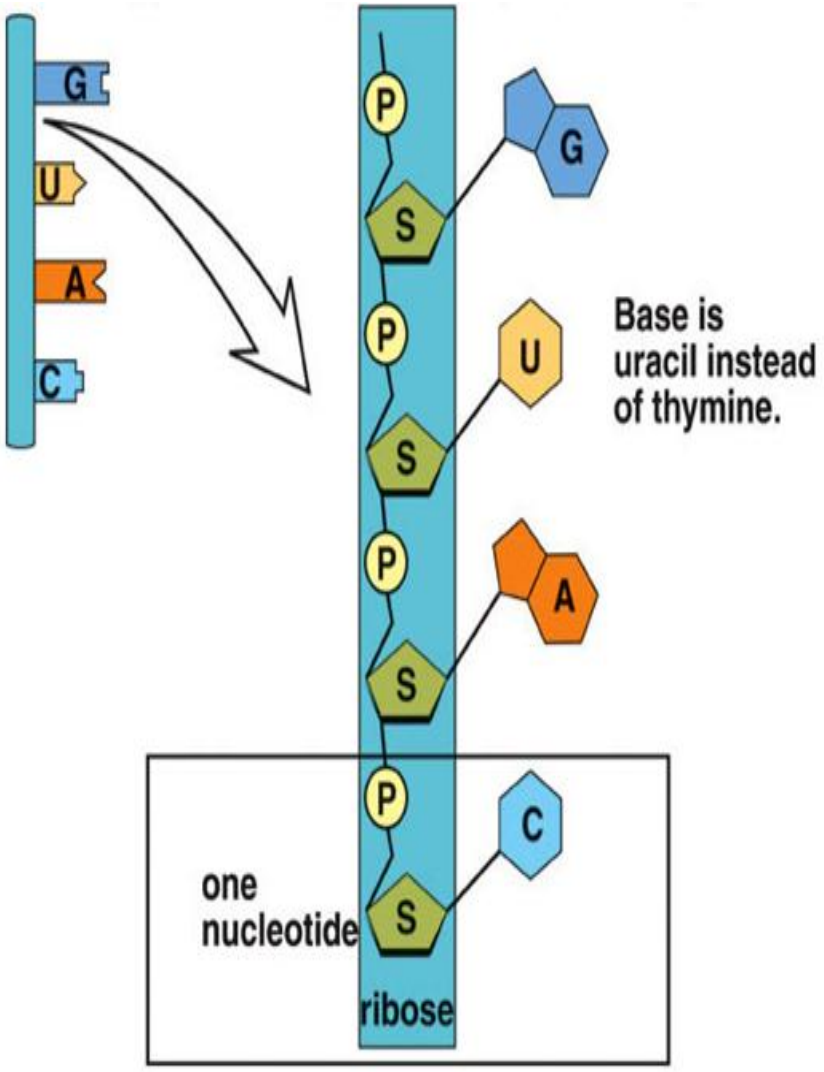
- **It has the base uracil (U) instead of thymine (T)**



RNA Differs from DNA

1. **RNA has a sugar ribose**
   DNA has a sugar deoxyribose
2. **RNA contains the base uracil (U)**
   DNA has thymine (T)
3. **RNA molecule is single-stranded**
   DNA is double-stranded





Three Types of RNA

- **Messenger RNA (mRNA) carries genetic information to the ribosomes**

  (blueprint for the construction of a protein)

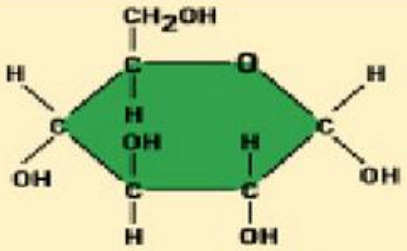
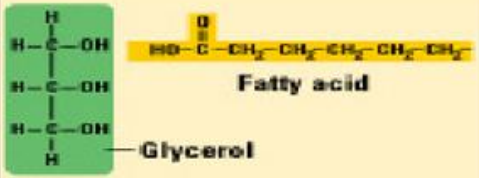
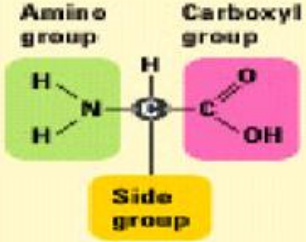
- **Ribosomal RNA (rRNA), along with protein, makes up the ribosomes**

  (construction site where the protein is made)

- **Transfer RNA (tRNA) transfers amino acids to the ribosomes where proteins are synthesized**

  (truck delivering the proper amino acid to the site at the right time)

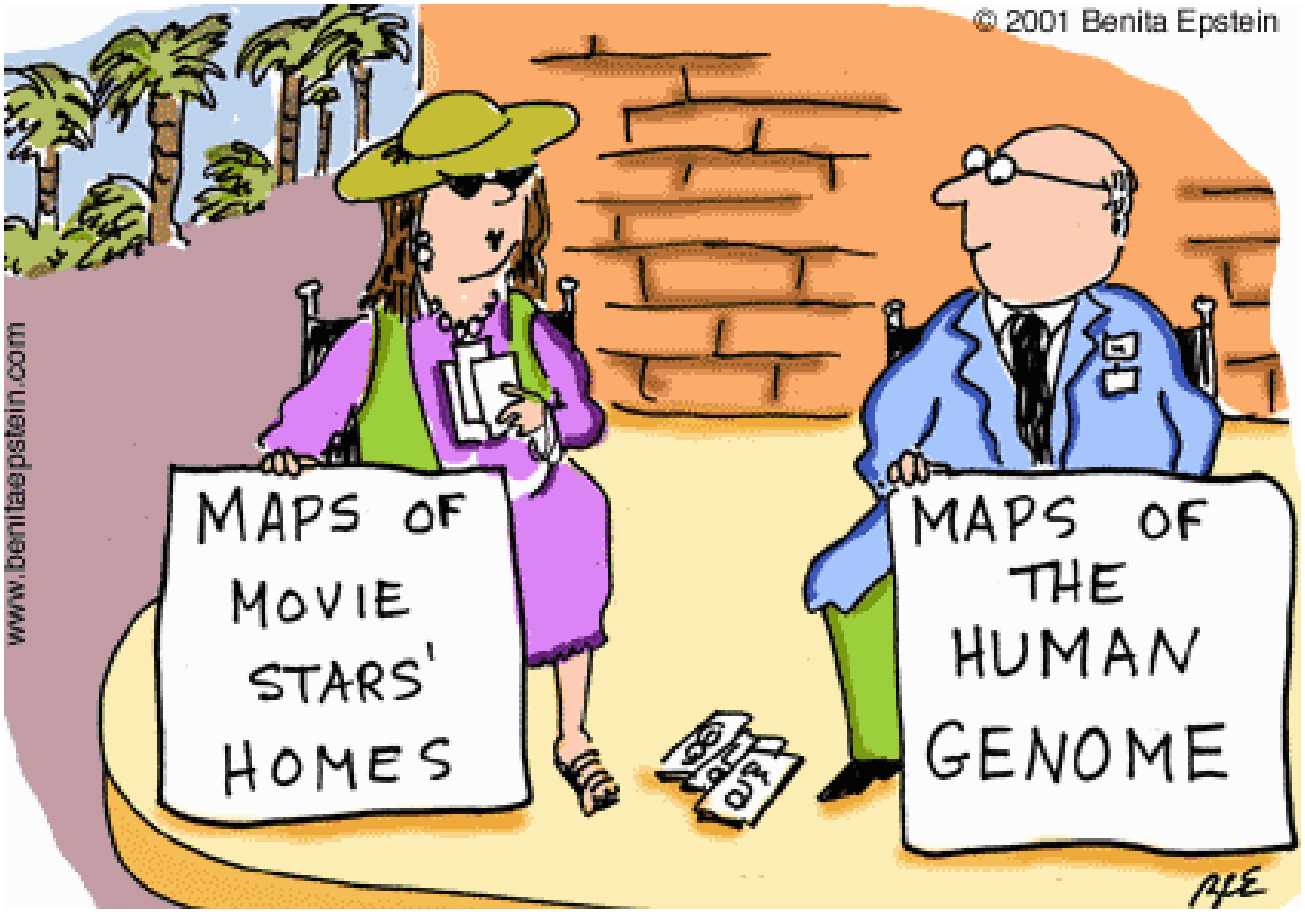
Macromolecules

Biological macromolecule	Function	Monomer	Examples
Carbohydrates	Dietary energy; storage; plant structure	 Monosaccharide	Monosaccharides: glucose, fructose. dissaccharides: lactose, sucrose. Polysaccharides: starch, cellulose.
Lipids	Long-term energy storage (for fats); hormones (for steroids)	 Components of a fat molecule	Fats, oils, steroids
Proteins	Enzymes, structure, storage, contraction, transport, etc.	 Amino acid	Lactase (an enzyme), hemoglobin
Nucleic acids	Information storage	Nucleotide	DNA, RNA

- The human genome project
- Genomics
- Sequence databases
- BLAST tool

HUMAN GENOME PROJECT [HGP]

- **HGP aim:** sequence the entire human genome and provide the data free to the world.



- Global collaboration which was the largest biological research project ever undertaken, involving thousands of staff in institutes across the globe.
- 13 years of work before a rough draft of the human genome was published in 2003
- It provided information of 3 trillion base pairs- and sequences of 30,000 genes.
- the data is provided as free and open access to everyone in the scientific community and the public domain through freely available, online public databases.

APPLICATIONS

Molecular Medicine

- Through genetic research, medicine will look more into the fundamental causes of diseases rather than concentrating on treating symptoms.
- Genetic screening will enable rapid and specific diagnostic tests making it possible to treat countless maladies.
- DNA-based tests clarify diagnosis quickly and enable geneticists to detect carriers within families.
- Genomic information can indicate the future likelihood of some diseases.

Waste Control and Environmental Cleanup

- microbes that live under extreme temperature and pressure conditions have been sequenced. By learning the unique protein structure of these microbes, researchers may be able to use the organisms and their enzymes for such practical purposes as waste control and environmental cleanup .

Energy Sources

- Having the genomic sequence of the methane-producing microorganism will allow researchers to explore the process of methanogenesis in more detail and could lead to cheaper production of fuel-grade methane .

Risk Assessment

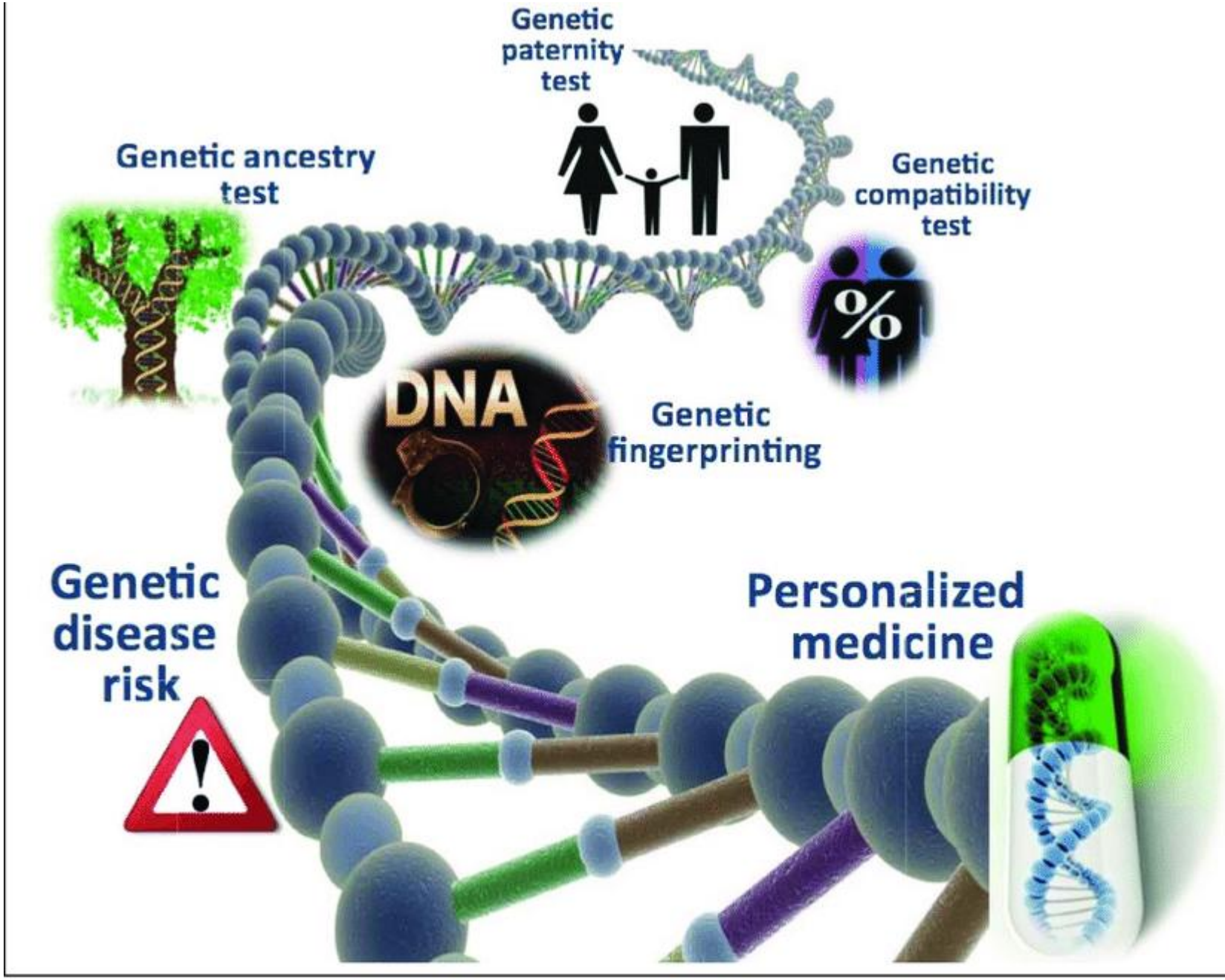
- ability to assess risks posed to individuals by environmental exposure to toxic agents. Scientists know that genetic differences cause some people to be more susceptible than others to such agents. to understand the effects of low-level exposures to radiation and other energy-related agents, especially in terms of cancer risk.

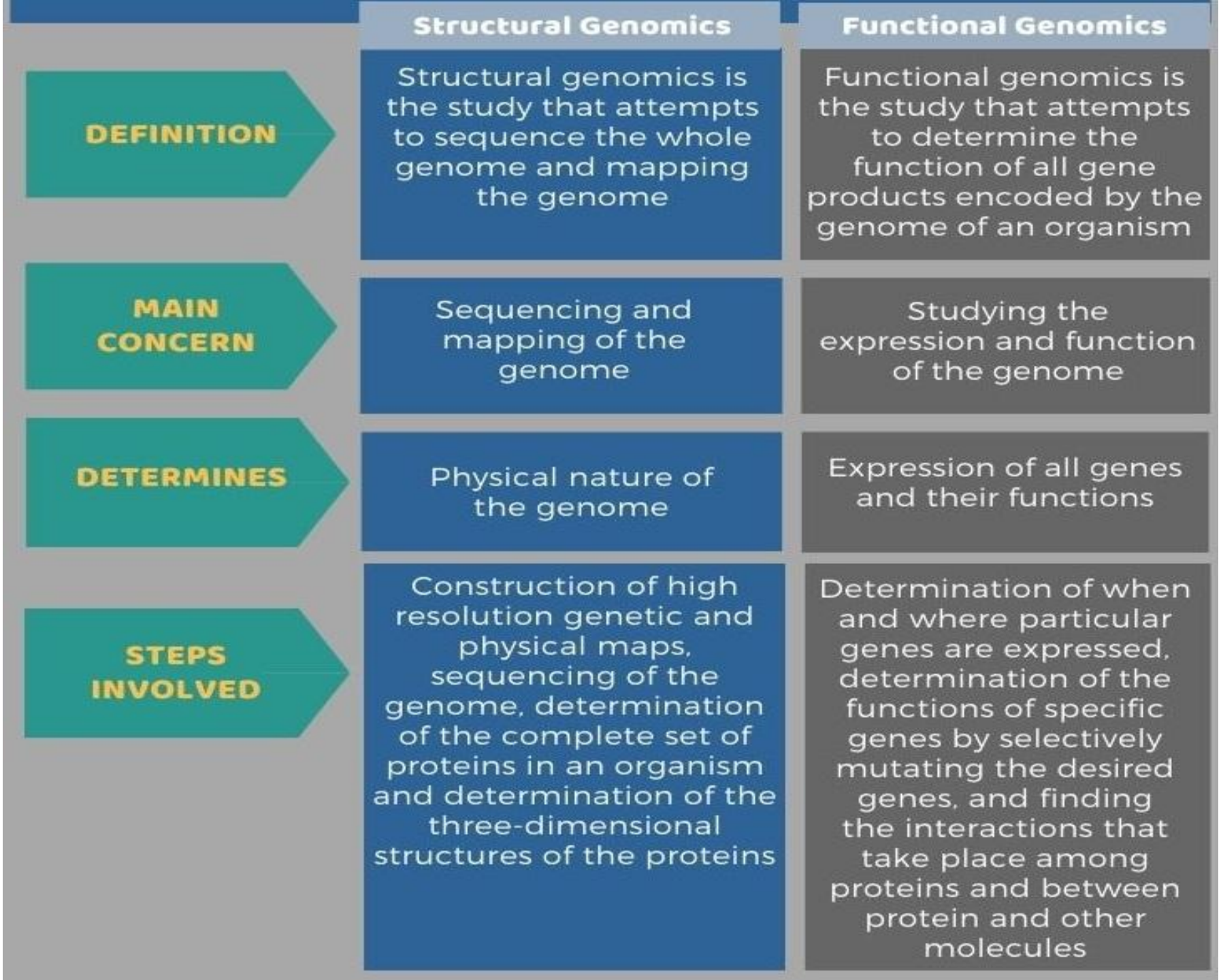
Genomics

- Genomics is the study of whole genomes of organisms, and incorporates elements from genetics.
- Types:
- **Structural Genomics** is the initial phase of genome sequencing which give the structure of every protein coded by the genome
- **Functional genomics** is the study of how genes and intergenic regions of the genome contribute to different biological processes.
- **Comparative genomics** Comparison of whole genomes from different organisms

Main points related to genomics are

- i. It is a computer aided study of structure and function of entire genome of an organism.
- ii. It deals with mapping and sequencing of genes on the chromosomes.
- iii. use of genomic techniques has become indispensible in plant breeding and genetics.


Genetic paternity test
Genetic ancestry test
Genetic compatibility test
%
DNA
Genetic fingerprinting
Genetic disease risk
!
Personalized medicine


Structural Genomics
Functional Genomics
DEFINITION
Structural genomics is the study that attempts to sequence the whole genome and mapping the genome
Functional genomics is the study that attempts to determine the function of all gene products encoded by the genome of an organism
MAIN CONCERN
Sequencing and mapping of the genome
Studying the expression and function of the genome
DETERMINES
Physical nature of the genome
Expression of all genes and their functions
STEPS INVOLVED
Construction of high resolution genetic and physical maps, sequencing of the genome, determination of the complete set of proteins in an organism and determination of the three-dimensional structures of the proteins
Determination of when and where particular genes are expressed, determination of the functions of specific genes by selectively mutating the desired genes, and finding the interactions that take place among proteins and between protein and other molecules

Comparative Genomics

NHGRI FACT SHEETS
genome.gov

Researchers choose the appropriate time-scale of evolutionary conservation for the question being addressed.

Common features of different organisms such as humans and fish are often encoded within the DNA evolutionarily conserved between them.

Looking at **closely related species** such as humans and chimpanzees shows which genomic elements are unique to each.

Genetic differences **within one species** such as our own can reveal variants with a role in disease.

NIH National Human Genome Research Institute

Genome databases/ Biological Databases

- One of the hallmarks of modern genomic research is the generation of enormous amounts of raw sequence data.
- As the volume of genomic data grows, sophisticated computational methodologies are required to manage the data deluge.
- Thus, the very first challenge in the genomics era is to store and handle the staggering volume of information through the establishment and use of computer databases.
- A biological database is a large, organized body of persistent data, usually associated with computerized software designed to update, query, and retrieve components of the data stored within the system.
- The chief objective of the development of a database is to organize data in a set of structured records to enable easy retrieval of information.

Uses of biological Databases :

- It helps the researchers to study the available data and use in their research hypothesis
- It helps scientists to understand the concepts of biological phenomena.
- The database acts as a storage of information.
- It helps remove the redundancy of data.

Database types

1. Primary databases :

- It can also be called an archival database since it archives the experimental results submitted by the scientists. The primary database is populated with experimentally derived data like genome sequence, The data entered here remains un curated
- It obtains unique data obtained from the laboratory and these data are made accessible to normal users without any change.
- Examples of Primary database-

Nucleic Acid Database

- GenBank -

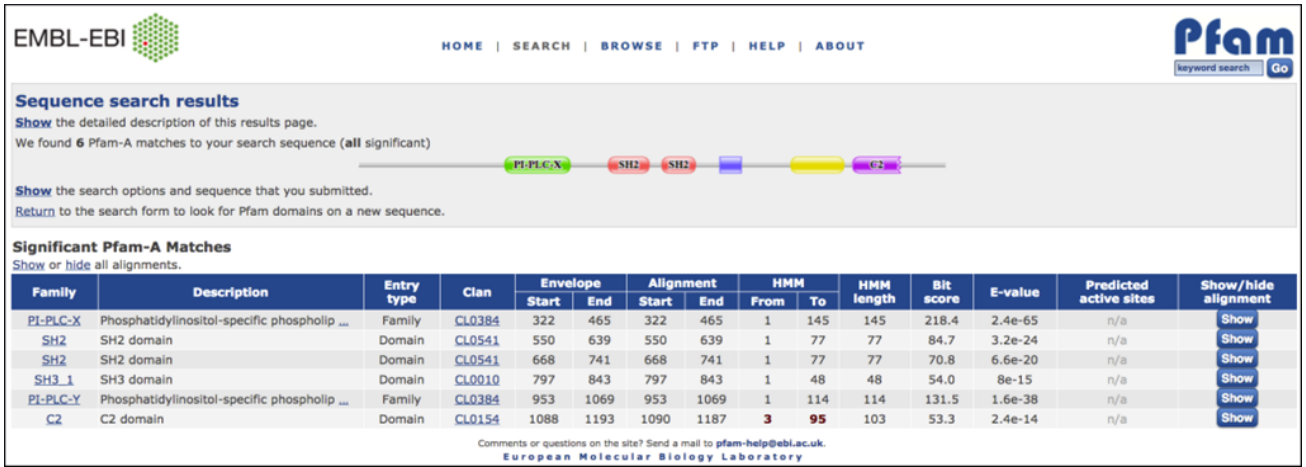
https://www.ncbi.nlm.nih.gov/nucleotide/

Protein Databases are

- PDB- Protein structure database https://www.rcsb.org/
- UniprotKB- Protein sequence database Database https://www.uniprot.org/uniprotkb?query=*

Secondary databases

- Secondary databases comprise data derived from the results of analyzing primary data.
- Secondary databases often draw upon information from numerous sources, including other databases (primary and secondary), controlled vocabularies and the scientific literature.
- They are highly curated, often using a complex combination of computational algorithms and manual analysis and interpretation to derive new knowledge from the public record of science.
- Examples
- InterPro (protein families, motifs and domains)- https://www.ebi.ac.uk/interpro/
- Pfam
- Prosite



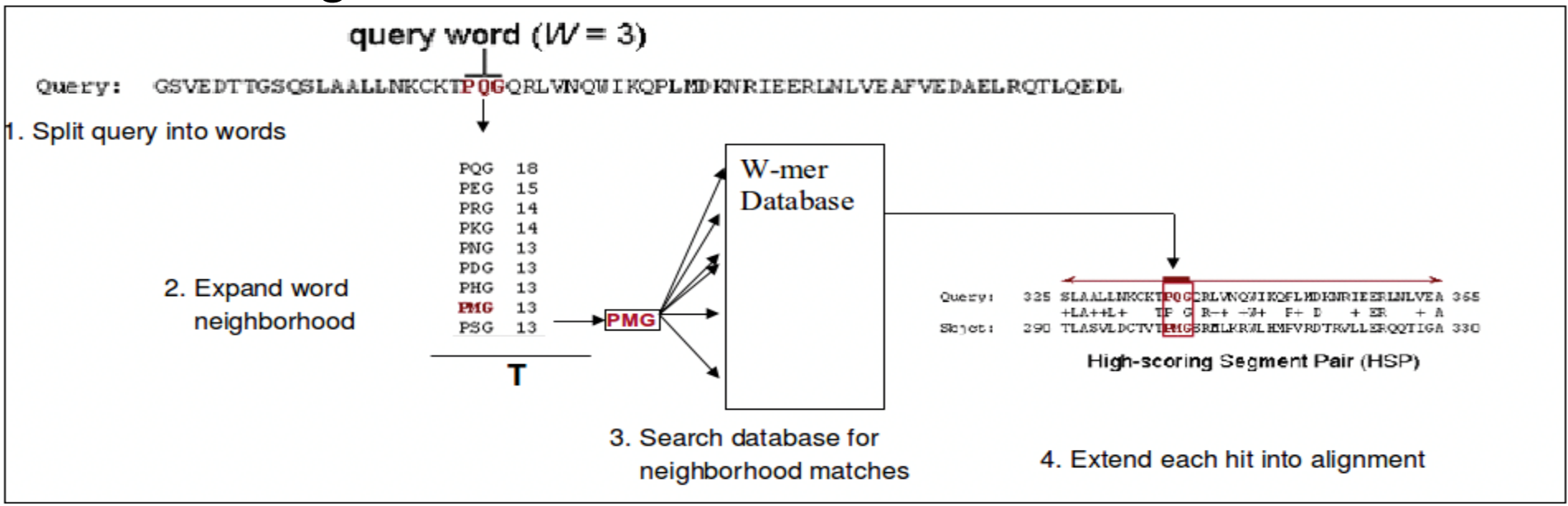
Searching and querying the databases

- Database searching for similar sequences is ubiquitous in bioinformatics.
- Databases are large and getting larger
- Need fast methods
- Search Tools types
- Similarity Search Tools – Smith-Waterman Searching
- Heuristic Search Tools – FASTA – BLAST

BLAST

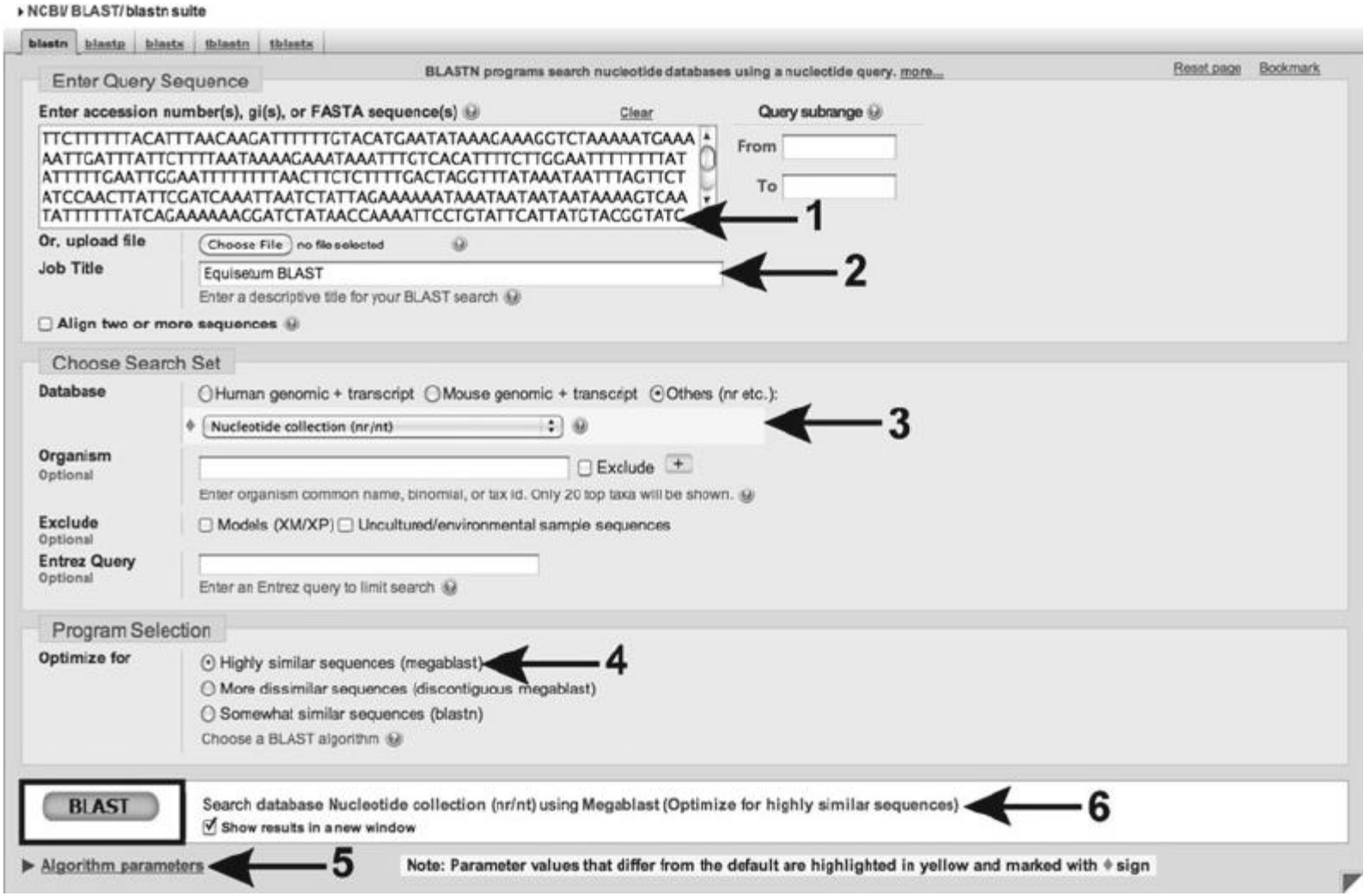
- BLAST stands for Basic Local Alignment Search Tool.
- It is a local alignment algorithm-based tool that is used for aligning multiple sequences and to find similarity or dissimilarity among various species.
- BLAST is a heuristic method which means that it is a dynamic programming algorithm that is faster, efficient but relatively less sensitive.
- Finds regions of similarity between sequences.
- The program compares nucleotide or protein sequences and calculates the statistical significance of matches.

BLAST Algorithm

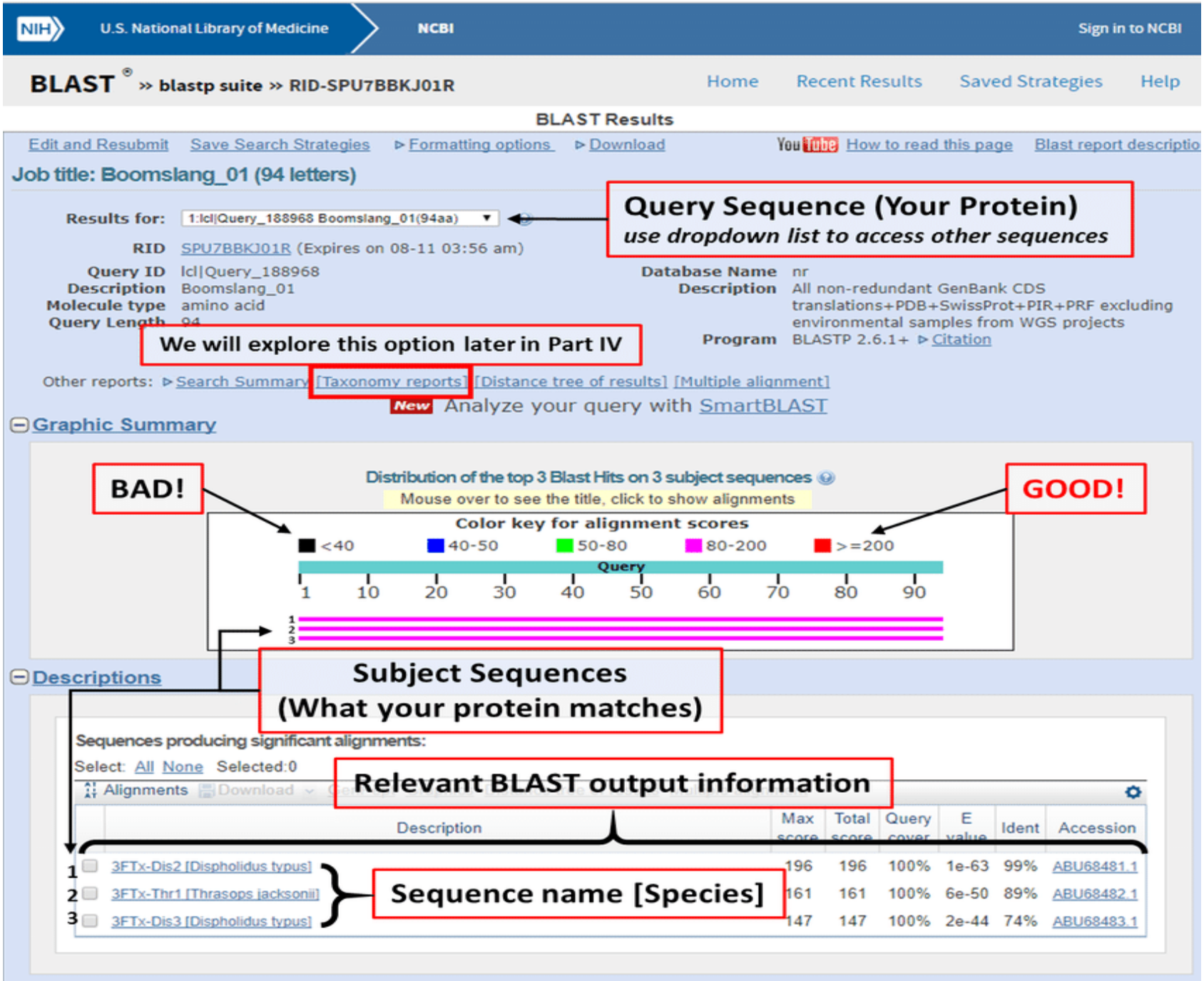


The steps are as follows:

- Split query into overlapping words of length W (the W-mers)
- Find a “neighborhood” of similar words for each word
- Lookup each word in the neighborhood in a hash table to find the location in the database where each word occurs.
- Call these the *seeds*, and let *S* be the collection of seeds.
- Extend the seeds in *S* until the score of the alignment drops off below some threshold *X*.
- Report matches with overall highest scores



(1) Enter the query sequence; (2) Select a job title; (3) Select the database to search; (4) Select the BLAST algorithm to use; (5) Adjust the algorithm as necessary; (6) Start the BLAST.


U.S. National Library of Medicine
NCBI
Sign in to NCBI
BLAST » blastp suite » RID-SPU7BBKJ01R
Home
Recent Results
Saved Strategies
Help
BLAST Results
Edit and Resubmit
Save Search Strategies
Formatting options
Download
How to read this page
Job title: Boomslang_01 (94 letters)
Results for: 1:lcl|Query_188968 Boomslang_01(94aa)
Query Sequence (Your Protein)
use dropdown list to access other sequences
RID SPU7BBKJ01R (Expires on 08-11 03:56 am)
Query ID lcl|Query_188968
Description Boomslang_01
Molecule type amino acid
Query Length 94
Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.6.1+ Citation
We will explore this option later in Part IV
Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment]
New Analyze your query with SmartBLAST
Graphic Summary
BAD!
GOOD!
Distribution of the top 3 Blast Hits on 3 subject sequences
Mouse over to see the title, click to show alignments
Color key for alignment scores
<40
40-50
50-80
80-200
>=200
Query
1 10 20 30 40 50 60 70 80 90
Descriptions
Subject Sequences
(What your protein matches)
Sequences producing significant alignments:
Select: All None Selected:0
Alignments
Download
Relevant BLAST output information
Description
Max score
Total score
Query cover
E value
Ident
Accession
1 3FTx-Dis2 [Dispholidus typus] 196 196 100% 1e-63 99% ABU68481.1
2 3FTx-Thr1 [Thrasops jacksonii] 161 161 100% 6e-50 89% ABU68482.1
3 3FTx-Dis3 [Dispholidus typus] 147 147 100% 2e-44 74% ABU68483.1
Sequence name [Species]

- Blast returns the output in the form of hit tables that are arranged in decreasing order of matched accession number along with their titles, query coverage, sequence identity, score, and an e-value in separate columns. The reliability of the matched sequences is assessed by e-value.
- E value-The Expectation value or Expect value represents the number of different alignments with scores equivalent to or better than **S** that is expected to occur in a database search by chance. The lower the E value, the more significant the score and the alignment.

BLAST Program Types	Further details
nucleotide blast or blastn	Compares a nucleotide query sequence against a nucleotide sequence database.
protein blast or blastp	Compares an amino acid query sequence against a protein sequence database.
blastx	Compares a nucleotide query sequence translated in all reading frames against a protein sequence database. You could use this option to find potential translation products of an unknown nucleotide sequence.
tblastn	Compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.
tblastx	Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. Please note that the tblastx program cannot be used with the nr database on the BLAST Web page because it is computationally intensive

THANK YOU