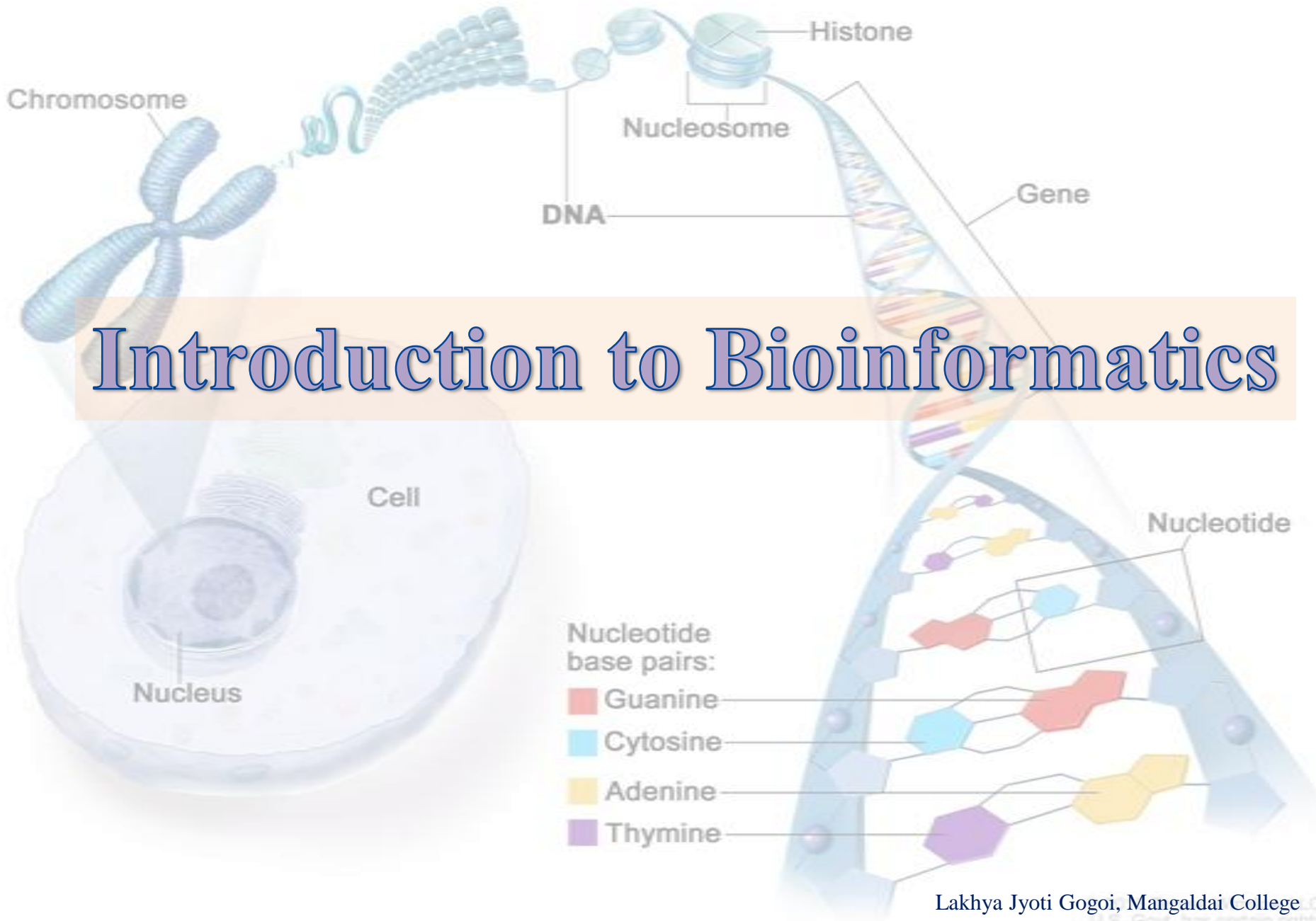


## DNA Structure



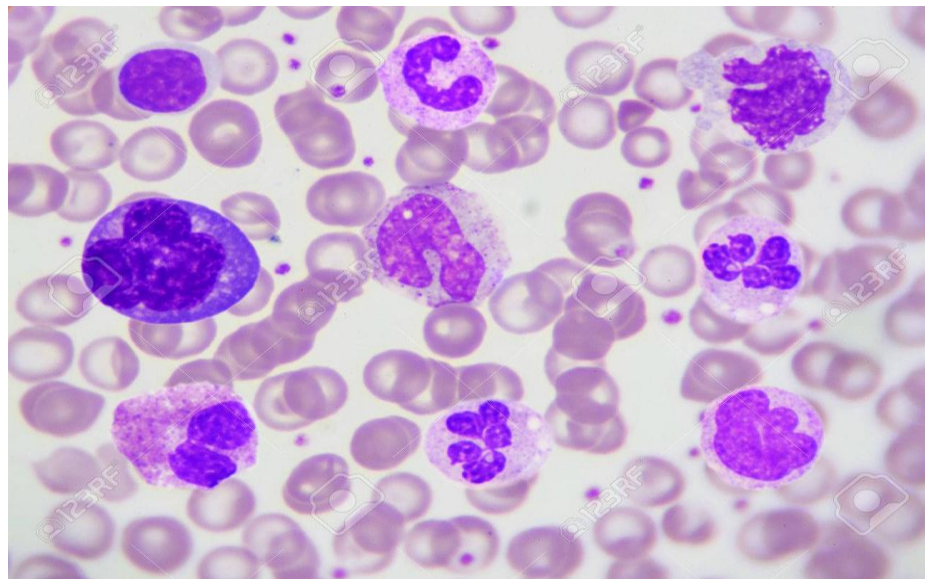
# Introduction to Bioinformatics

# Cell and Nucleus

The human body is made up of billions of cells, each with a nucleus in the center. The nucleus is like a “central office” for the cell, where the most important instructions come from for

- How the cell will operate
- What it will manufacture
- How it will communicate with other cells

Most of a cell’s genetic material is stored in the nucleus.



# Animal Cell Structure

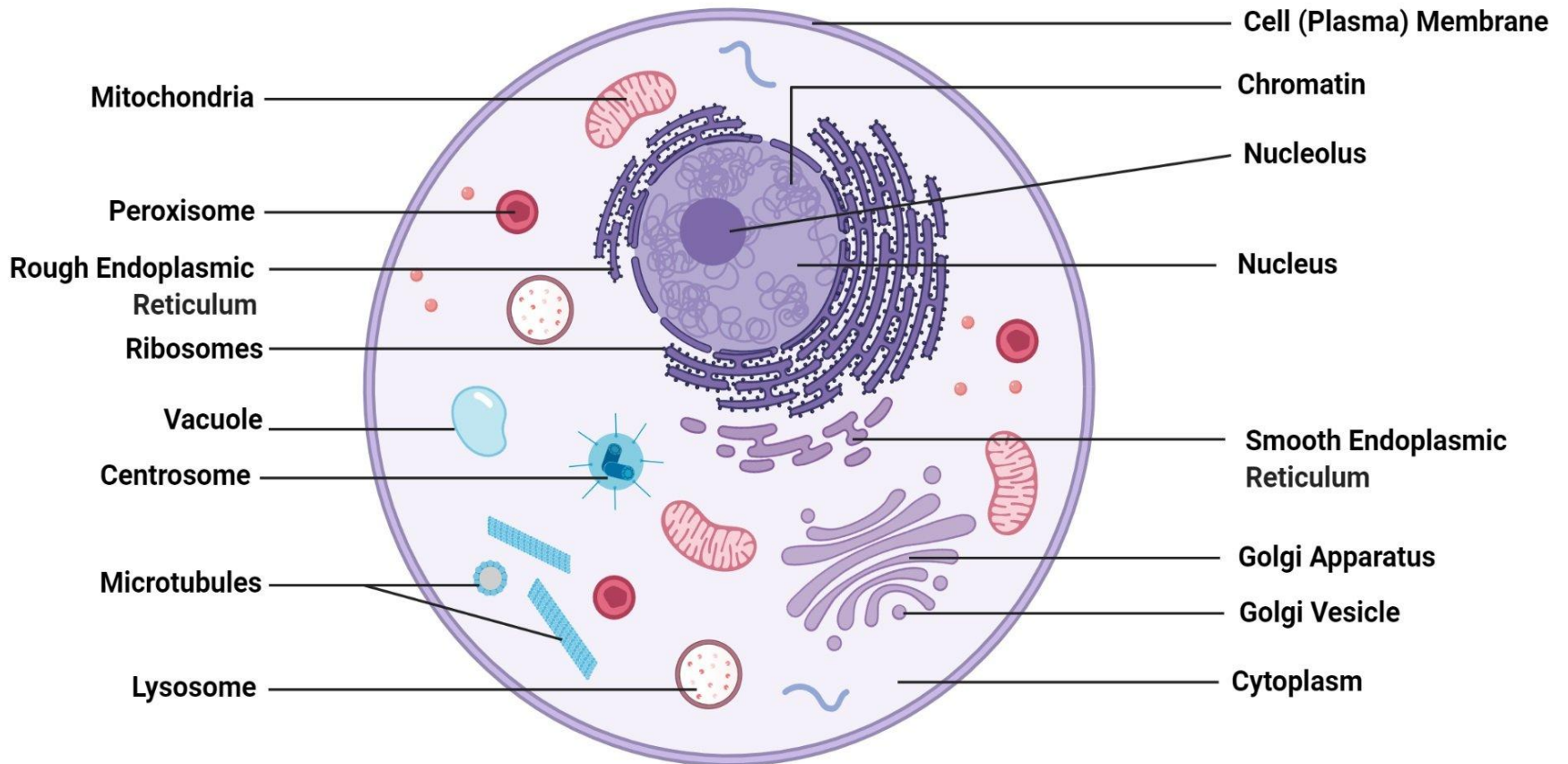
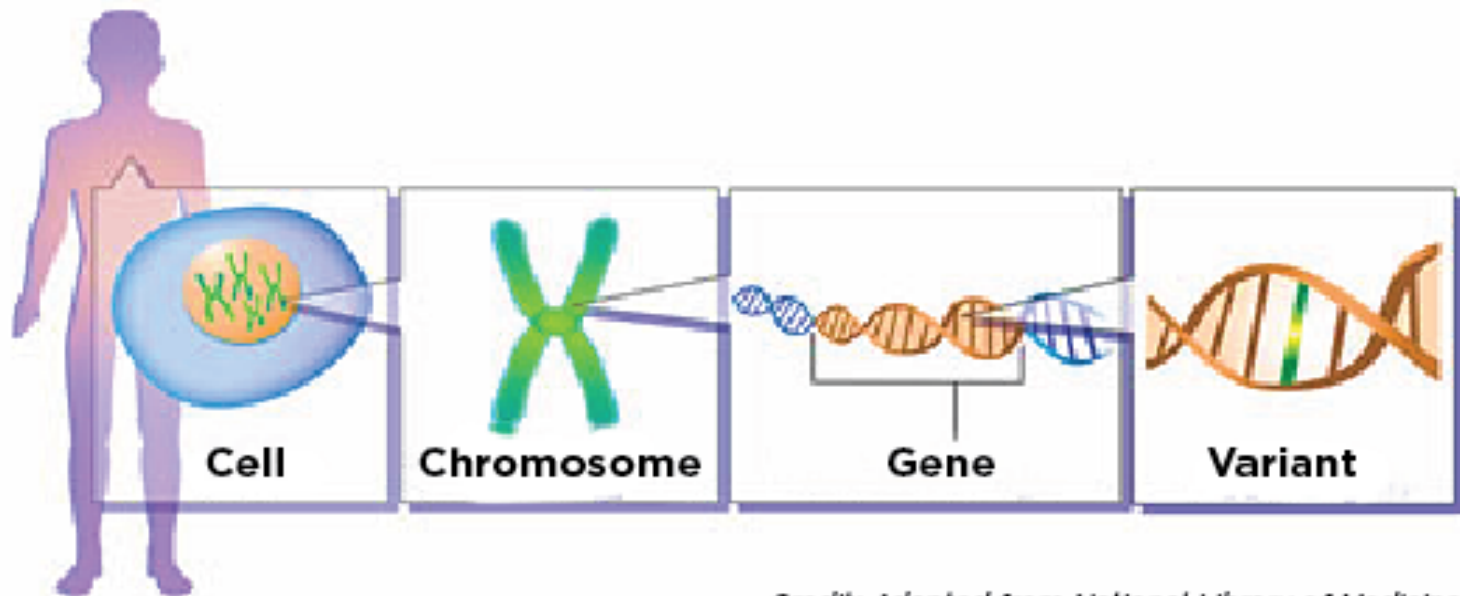


Figure: Animal Cell Structure, Image Copyright © Sagar Aryal, [www.microbenotes.com](http://www.microbenotes.com)

## Genes:

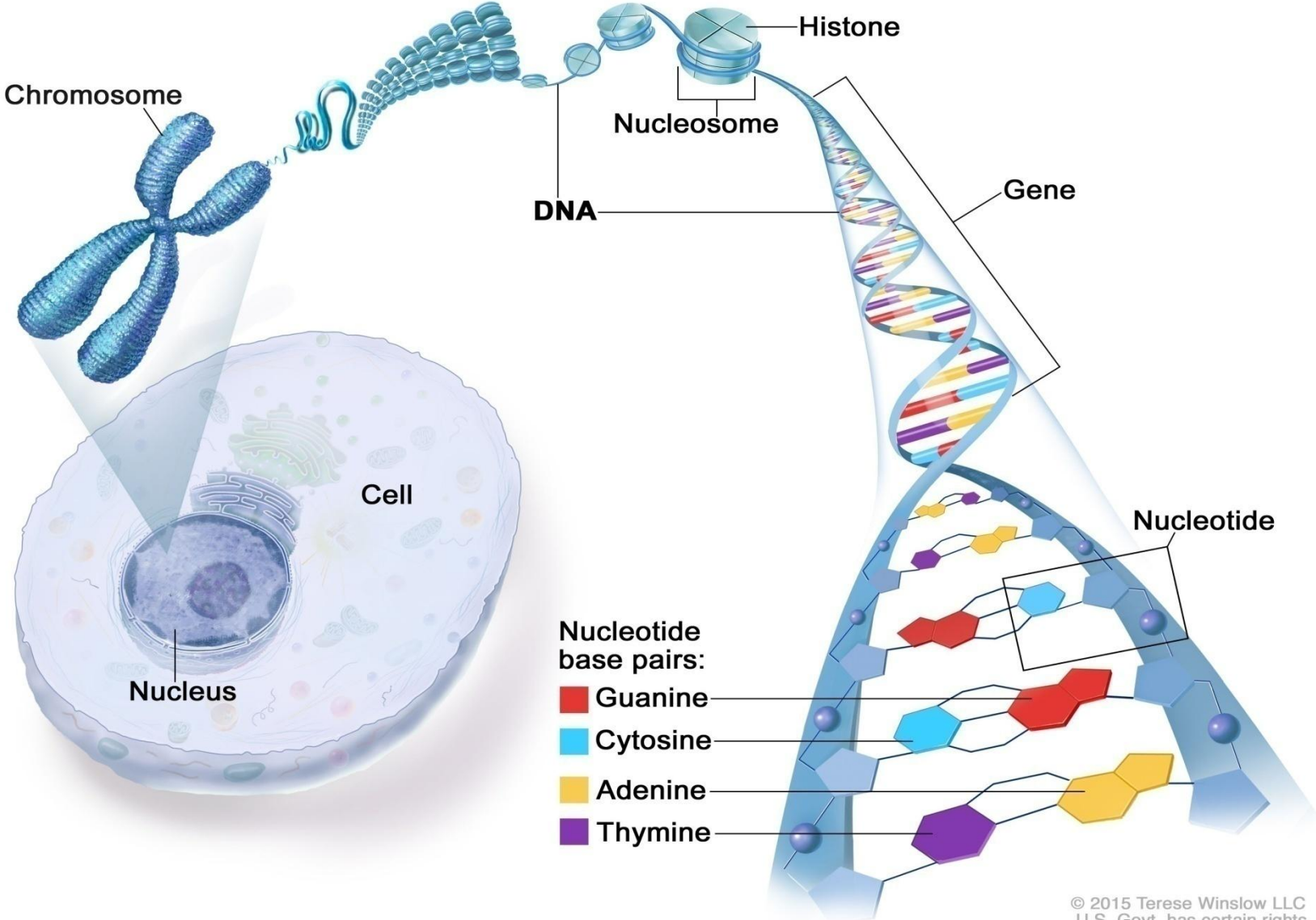
- ✓ Each chromosome contains thousands of genes. A gene is a portion of DNA that codes for a protein.
- ✓ Each gene is made up of billions of letters, which our body reads in three-letter units called **codons**.
- ✓ Each three-letter codon determines a specific amino acid.
- ✓ The amino acids are strung together to make proteins, such as hormones, enzymes, and antibodies.
- ✓ The many genes that provide the instructions for the proteins in our bodies determine a wide range of features, including outwardly appearing physical traits such as height, eye color and hair color, to inner functioning, such as how each organ system works.
- ✓ Some variation in genes may not change amino acids or proteins.

- ✓ There are over 20,000 genes in the human genome.
- ✓ Although genes and chromosomes are mostly similar from person to person, there is variation among people. Most of the time, this variation does not impact health or development. Genetic variation explains some of the wonderful differences that we see among humans.



*Credit: Adapted from National Library of Medicine*

# DNA Structure

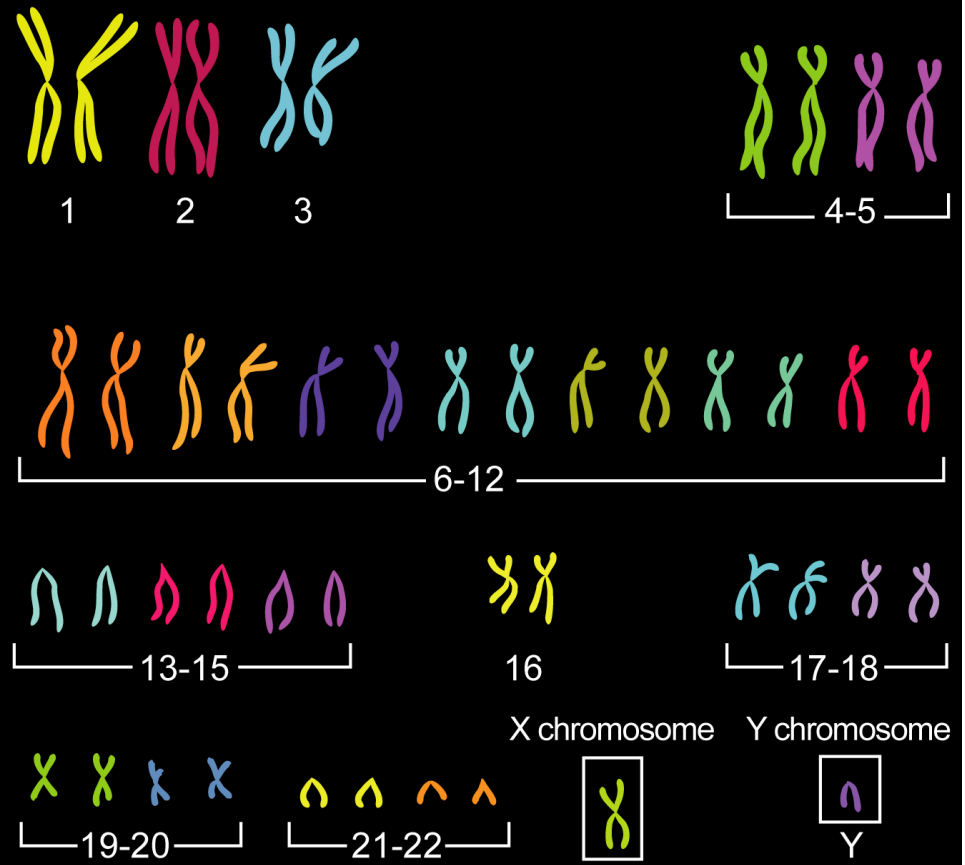


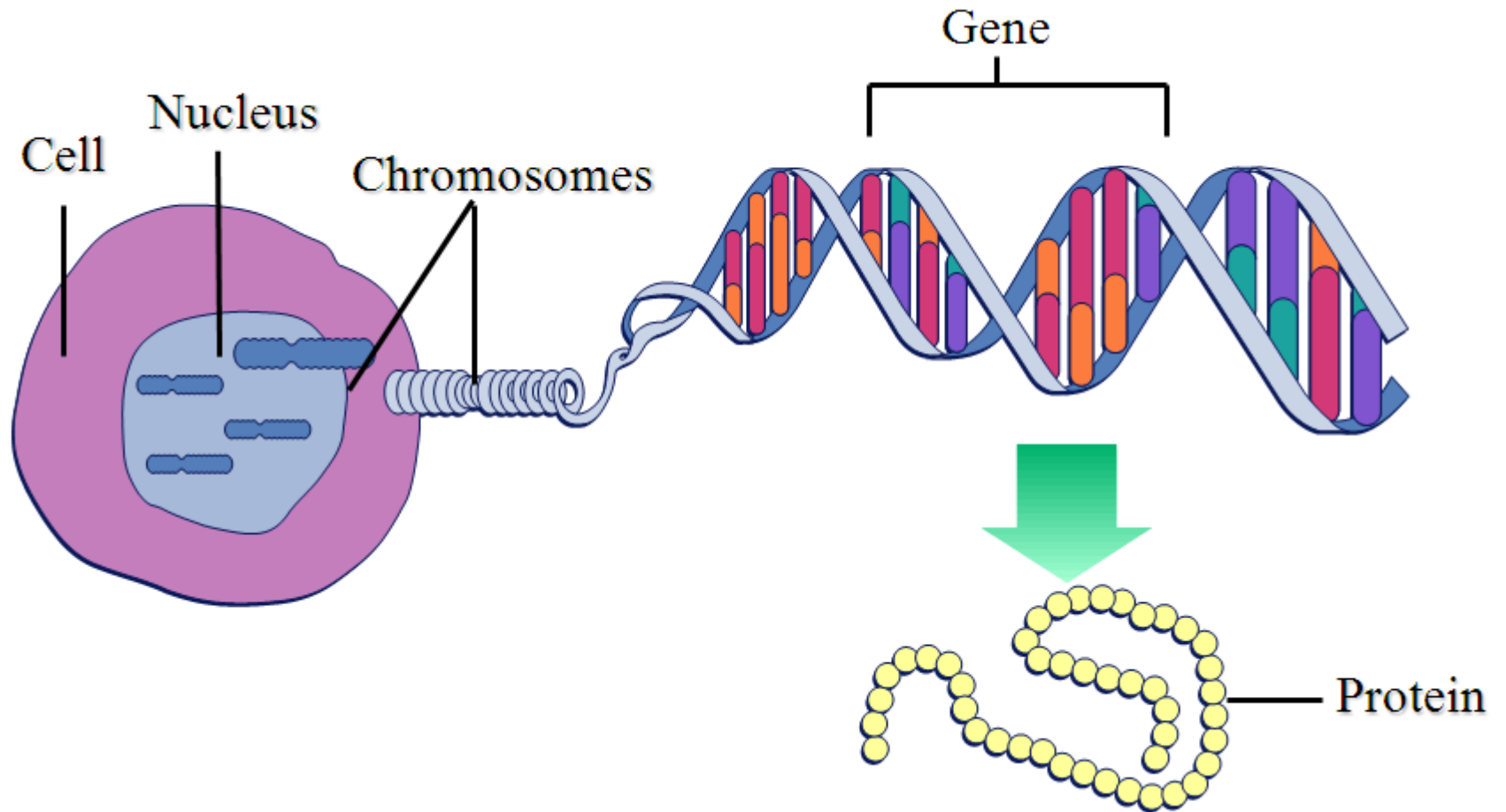
© 2015 Terese Winslow LLC  
U.S. Govt. has certain rights

(a)

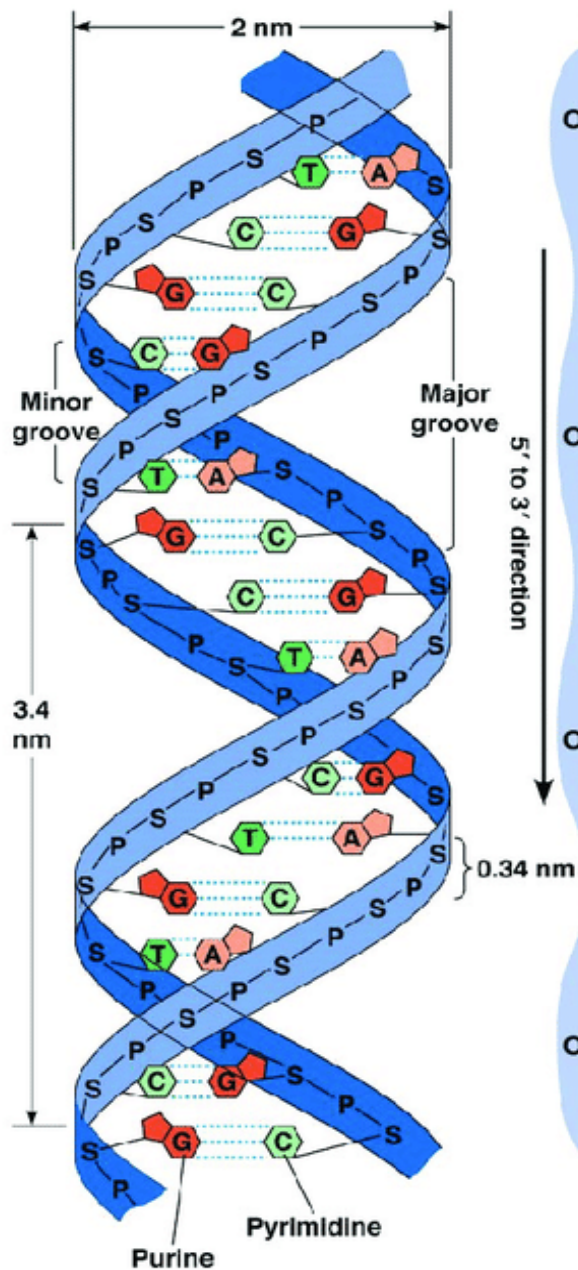


(b)

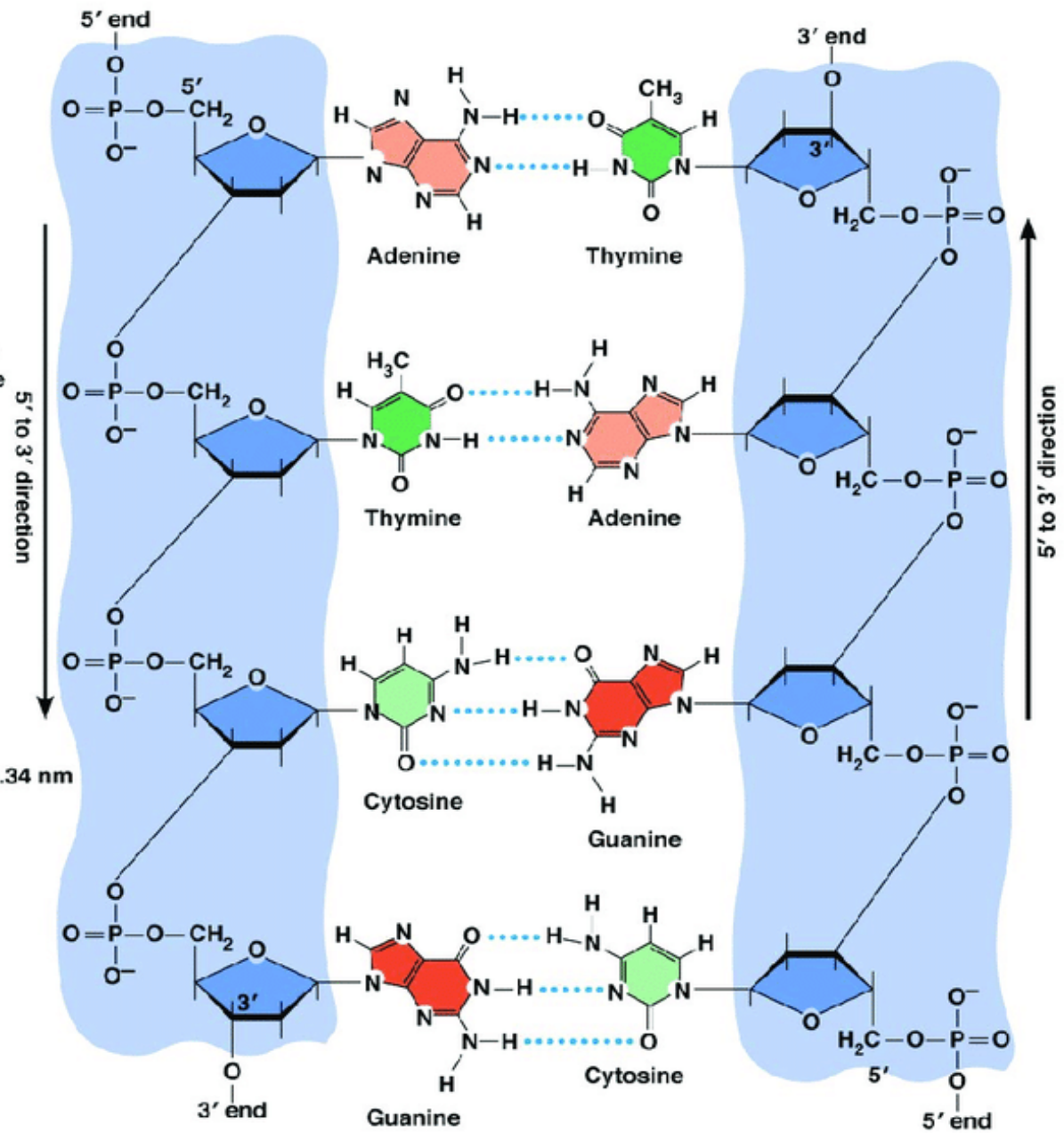








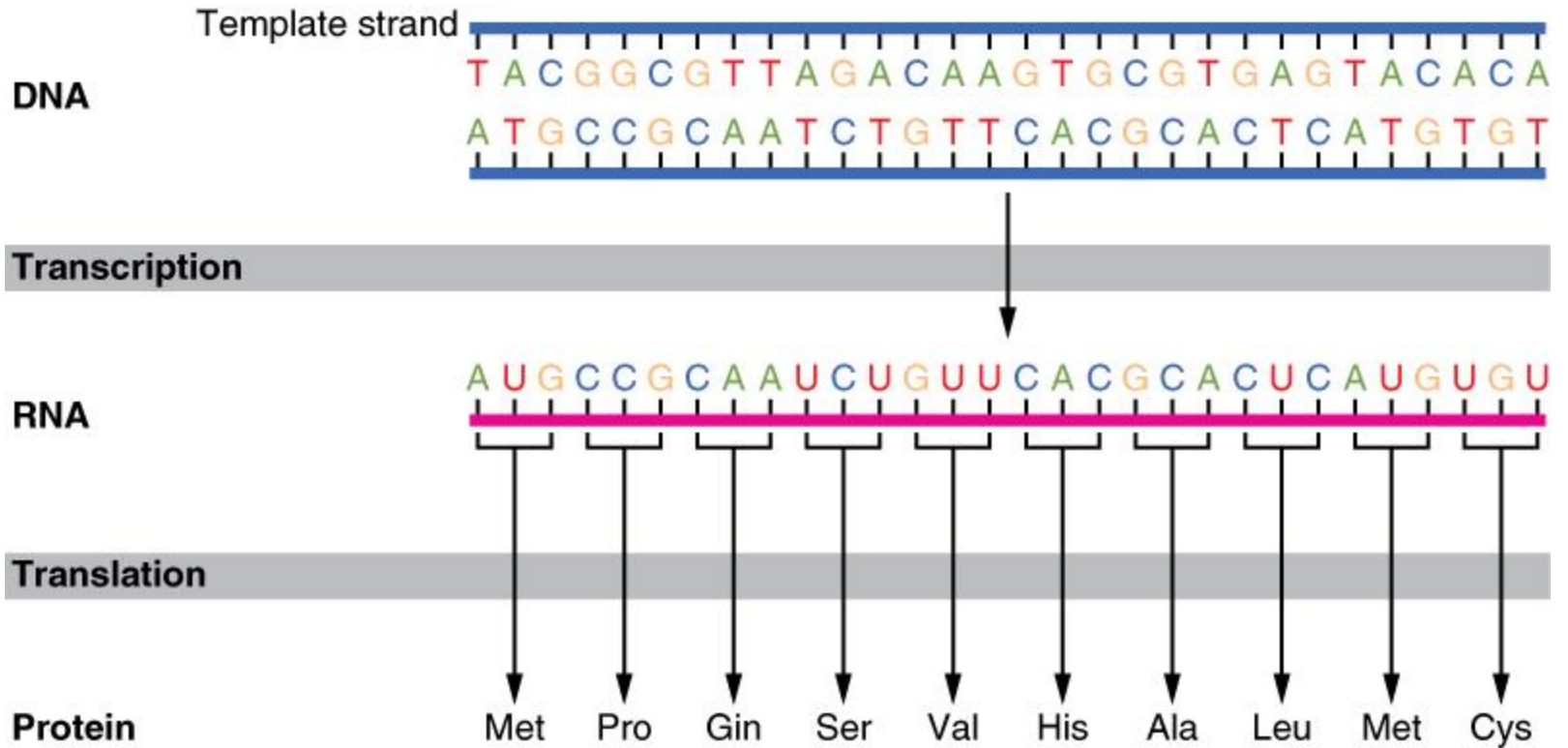
**(a) Double helix**



**(b) Antiparallel orientation of strands**

# Genetic Code :

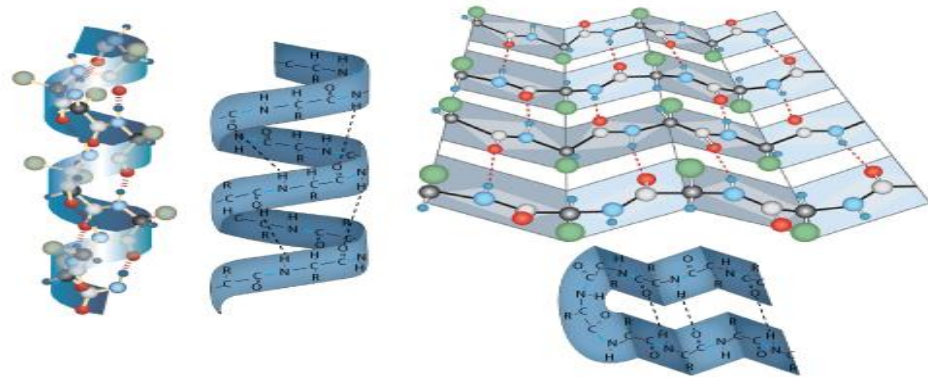
		Second Letter																			
		U		C		A		G													
1st letter	U	UUU   Phe	UCU   Ser	UAU   Tyr	UGU   Cys	U	UUC   Leu	UCC   Ser	UAC   Tyr	UGC   Cys	C	UUA   Leu	UCA   Ser	UAA   Stop	UGA   Stop	A	UUG   Leu	UCG   Ser	UAG   Stop	UGG   Trp	G
	C	CUU   Leu	CCU   Pro	CAU   His	CGU   Arg	U	CUC   Leu	CCC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



Primary



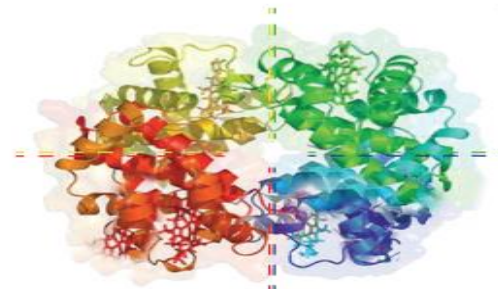
Secondary

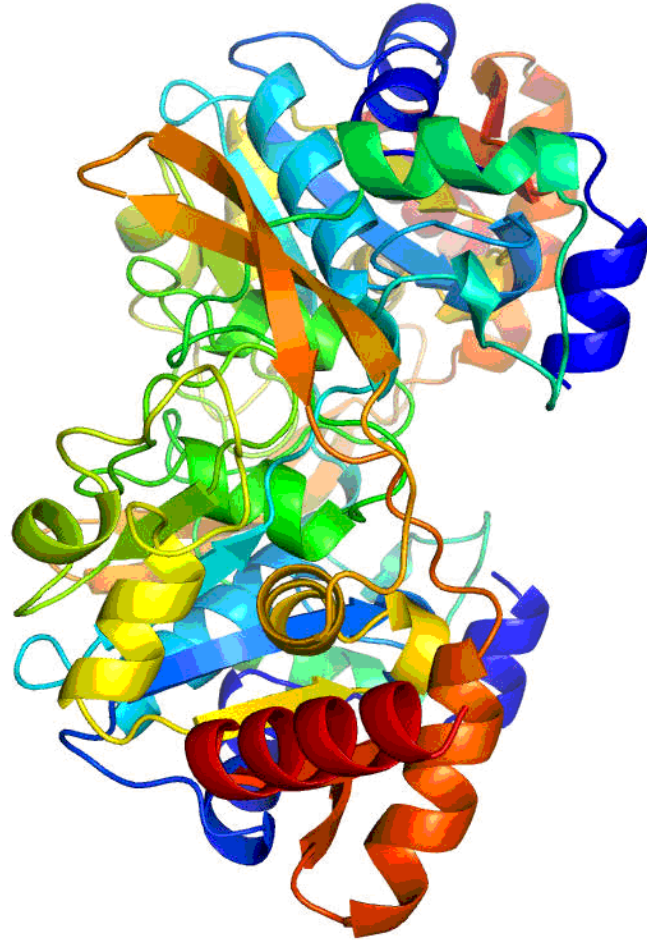


Tertiary

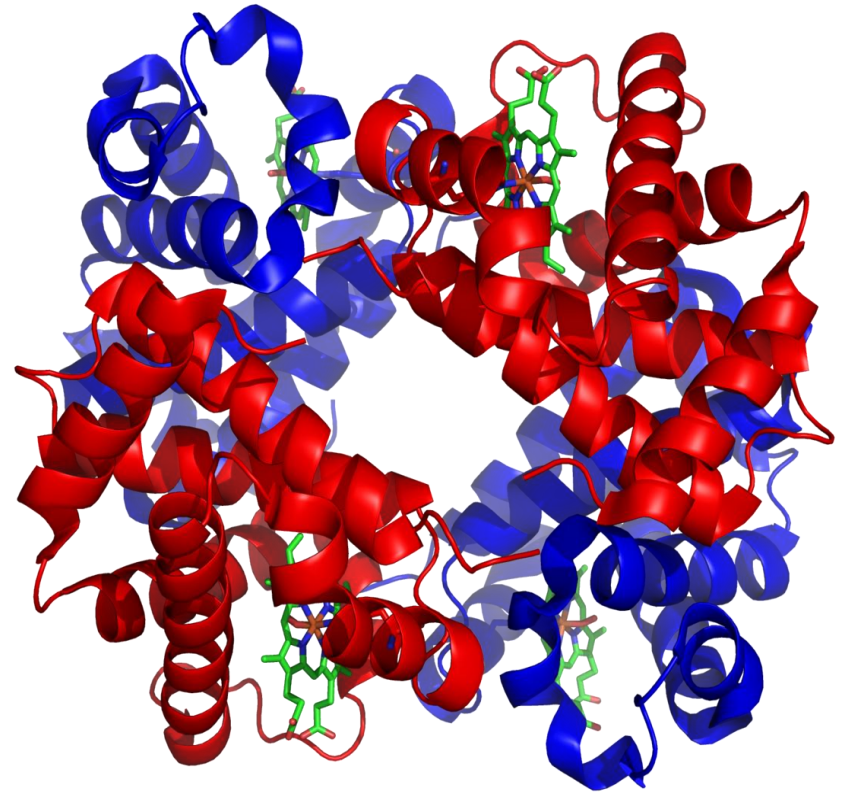


Quaternary





Anthrax Protein



Hemoglobin

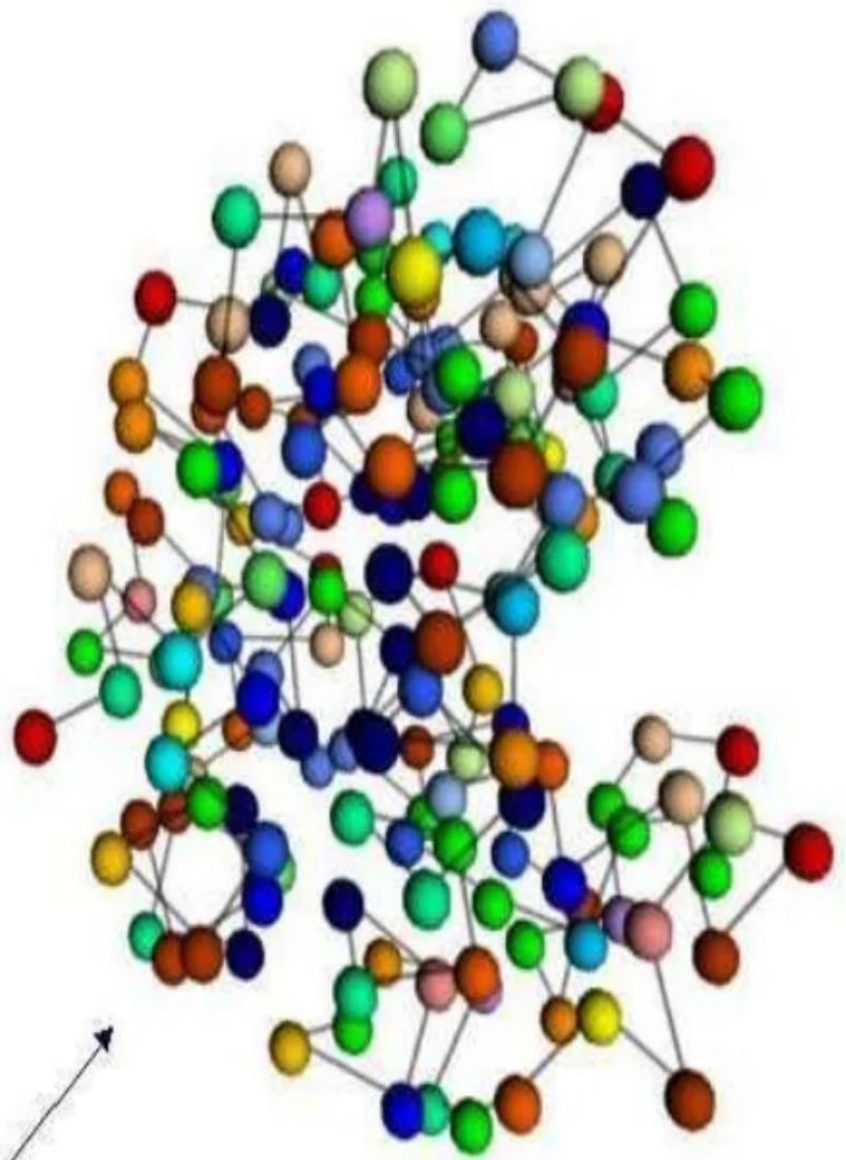
Bioinformatics is the application of information technology to store, organize and analyze the vast amount of biological data.

The stored data is available in the form of sequences and structures of proteins and nucleic acids (information carrier).

The biological information of nucleic acids are available as sequences while the data of proteins are available as sequences and structures.

Sequences are represented in single dimension where as the structure contains the three dimensional data of sequences.

RTDCYGNVNRIDTTG  
ASCKTAKPEGLSYCG  
VSASKKIAERDLQAM  
DRYKTI IKKVGKLC  
VEPAVIAAGIISRESH  
AGKVLKNGWDRGNG  
FGLMQVDKRS HKPQG  
TWNGEVHI TQGTTL  
INFIKTI QKKFPSWT  
KDQQLKGGI SAYNAG  
AGNVRSYARM DIGTT  
HDDYANDVVARAQYY  
KQHGY



Primary Sequence      3D Structure

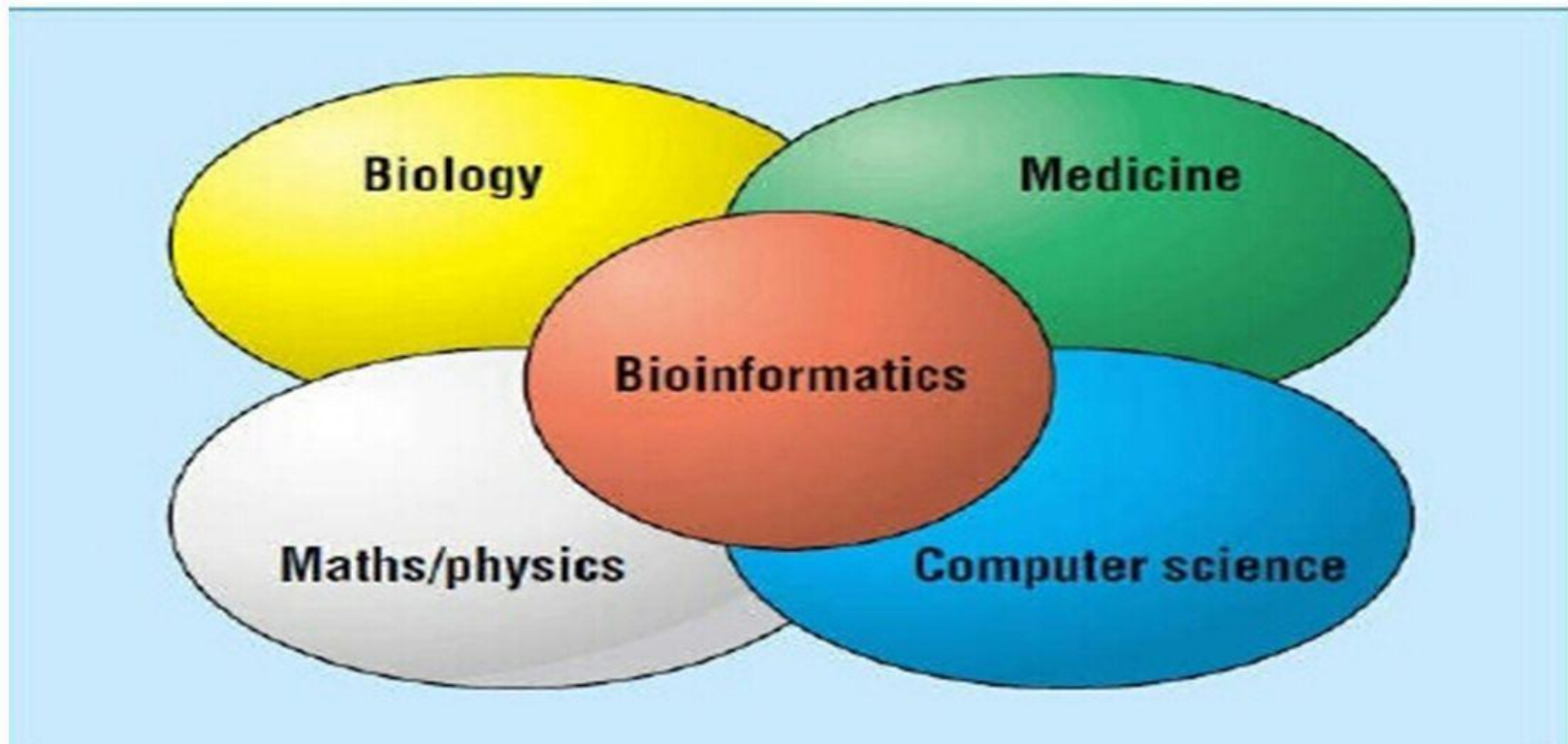
Bioinformatics is a field in which Biology, Mathematics, statistics, CS and IT are merged into a single discipline to process biological data.

Complex machines are used to read in biological data at a much faster rather than before.

The term “**Bioinformatics**” was invented by **Paulien Hogeweg** and **Ben Hesper** in **1970**.



Bioinformatics is a multifaceted discipline combining many scientific fields including computational biology, statistics, mathematics, molecular biology and genetics



## History of Bioinformatics:

**1859** : The 'on the origin of species' published by Darwin that introduced the theory of genetic evolution.

**1869** : The DNA from nuclei of white blood cells was first isolated by Friedrich Meischer.

**1953** : Watson and Crick proposed the double helix model for DNA.

**1955** : The sequence of the first protein to be analyzed, bovine insulin is announced by F. Sanger.

**1973** : The Brookhaven Protein Data Bank (PDB) is announced.

**1987** : Perl (Practical Extraction Report Language) is released by Larry Wall.

**1988** : National Centre for Biotechnology Information (NCBI) founded at NIH.

**1990** : Human Genome Project was launched.

**2001** : The Human Genome (3,000 Mbp) is published.

**2010** : Completion of the HGP.

# Why Bioinformatics is necessary?

The need of bioinformatics has arisen from the recent explosion of publicly available genomic information, such as resulting from the Human Genome Project.

Gain a better understanding of gene analysis, taxonomy and evolution.

To work efficiently on the rational drug designs and reduce the time taken for the development of drug manually.

# Goals of Bioinformatics

To uncover the wealth of Biological information hidden in the mass of sequences, structure, literature and biological data.

It is being used now and in the foreseeable future in the areas of molecular medicine.

It has environmental benefits in the identifying waste and clean up bacteria.

In agriculture, it can be used to produce high yield, low maintenance crops.

# Field of Bioinformatics

- Molecular medicine
- Gene therapy
- Drug development
- Molecular gene application
- Crop improvement
- Forensic analysis of microbes
- Biotechnology
- Evolutionary studies
- Bio weapon creation.....

# Applications of Bioinformatics

## Prediction of Protein Structure:

It is easy to determine the primary structure of proteins in the form of amino acids which are present on the DNA molecule but it is difficult to determine the secondary, tertiary and quaternary structures of protein. Tools of bioinformatics can be used to determine the complex protein structures.

## Genome annotation:

In genome annotation, genomes are marked to know the regulatory sequences and protein coding. It is a very important part of the human genome project as it determines the regulatory sequences.

## **Comparative genomics:**

Comparative Genomics is the branch of bioinformatics which determines the genomic structure and function relation between different biological species. For this purpose, intergenomic maps are constructed which enable the scientists to trace the processes of evolution that occurs in genomes of different species.

## **Health and Drug discovery:**

The tools of Bioinformatics are also helpful in drug discovery, diagnosis and disease management. Complete sequencing of human genes has enabled to scientists to make medicines and drugs which can target more than 500 genes.

## Preventative Medicine:

Gene identification by sequence inspection and prediction of splice site allows mutations to be corrected easily. This is much use in the analysis of mutations that cause cancer.

## Gene Therapy:

Mutation are easily detected and quantified through next generation sequencing technology in a heterogonous samples thus a cost effective precision medicine, '*right drug at right dose to the right patient at the right time*' can be administered.



# Biological Databases

# National Center for Biotechnology Information

NCBI Resources [x] How To [x] maglet My NCBI Sign Out

NCBI National Center for Biotechnology Information

All Databases [v]  Search

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News](#) | [Blog](#)

### Submit

Deposit data or manuscripts into NCBI databases



### Download

Transfer NCBI data to your computer



### Learn

Find help documents, attend a class or watch a tutorial



### Develop

Use NCBI APIs and code libraries to build applications



### Analyze

Identify an NCBI tool for your data analysis task



### Research

Explore NCBI research and collaborative projects



## Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[PubMed Health](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

## NCBI Announcements

New video on the NCBI YouTube channel: [Viral resources at NCBI](#)


11 Feb 2016

In the newest video on the NCBI YouTube channel: [Viral resources at NCBI](#)

NCBI to assist Louisiana State University in South and Southeast regional genomics hackathon

<https://www.ncbi.nlm.nih.gov>

# Protein Data Bank



PROTEIN DATA BANK

MyPDB Login

A MEMBER OF THE PDB

## An Information Portal to Biological Macromolecular Structures

As of Tuesday Feb 09, 2010 at 4 PM PST there are 63271 Structures | PDB Statistics

WHAT'S NEW | HELP | PRINT

PDB ID or keyword  Search | Advanced Search

### Home

News & Publications  
Policies  
FAQ  
Contact  
Feedback  
About Us

### Deposition

All Deposit Services  
Electron Microscopy  
NMR  
Validation Server  
BioSync Beamline  
Related Tools

### Search

Advanced Search  
Latest Release  
Latest Publications  
Sequence Search  
Ligand Search  
Unreleased Entries  
Browse Database  
Histograms

### Tools

File Downloads  
FTP Services  
File Formats  
Services: RESTful | SOAP  
Widgets  
Compare Structures

### Education

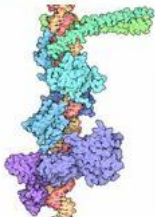
Looking at Structures  
Molecule of the Month  
Educational Resources

## A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

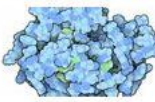
The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

### Molecule of the Month: Enhanceosome



Take a moment to ponder the form of your body: the shape of your face, the color of your eyes, the length of your fingers, the perfect articulation of your bones and muscles, the way your hair grows curly or straight. Now let your imagination travel inward, and think of the complex shapes and functions of your different cells, and the teeming molecular world inside each one. Remarkably, this amazing structure and form and function is specified by information in the genome, which encodes a mere 20,000-25,000 protein-coding genes. One of the great puzzles being pieced together by scientists is the mechanism by which these genes, and the methods used to control their expression, specify all of these different aspects of life. ■ [Read more ...](#) ■ [Previous Features](#)

### PSI Featured Molecule: Sugarcoating the surface: yeast Alg13



Many proteins in our cells are decorated with carbohydrate chains, which make the proteins more stable and assist with their function. Using NMR, PSI researchers now understand how this enzyme builds these essential carbohydrates. ■ [Read more from the Structural Genomics Knowledgebase](#) ■ [Previous Features](#)

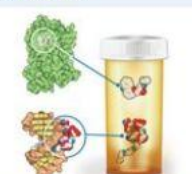
New user? Try the browser [compatibility check](#) and information on [Getting Started](#).

### News

- Complete News
- Newsletter
- Discussion Forum
- Job Listings

#### wwPDB Statement on Retraction of PDB Entries

09-February-2010  
**Poster Download: How Do Drugs Work?**



A new poster that explores different kinds of protein-drug structures found in the PDB archive is available for download. [More >>](#)

#### FTP Archive

The up-to-date PDB archive is available at: <ftp://wwpdb.org>

Time-stamped yearly snapshots are available at: <ftp://snapshots.wwpdb.org>

<https://www.rcsb.org>

# DNA Data Bank of Japan

The screenshot shows the DDBJ website interface. At the top left is the DDBJ logo with the text "DNA Data Bank of Japan". To the right is a search bar with "Google" カスタム検索 and a "Search" button. Below the logo is a navigation bar with links: "DDBJの紹介", "利用の手引き", "レポート・統計", "FAQ", and "お問い合わせ". The main content area shows a breadcrumb trail: "HOME > 検索・解析 > TXSearch". Below this is a dark banner with "TXSearch" and a "ヘルプ" link. Underneath is a "Taxonomy Browser" section with a search input field and "送信" and "clear" buttons. An "Example:" section lists a taxonomic hierarchy: "cellular organisms - Eukaryota - Opisthokonta - Metazoa - Eumetazoa - Bilateria - Deuterostomia - Chordata - Craniata - Vertebrata - Gnathostomata - Teleostomi - Euteleostomi - Sarcopterygii - Dipnotetrapodomorpha - Tetrapoda - Amniota - Mammalia - Theria - Eutheria - Boreoeutheria - Euarchontoglires - Primates - Haplorhini - Simiiformes - Catarrhini - Hominoidea - Hominidae - Homininae - Homo - Homo sapiens". Below the list is a table with columns: "Taxon (4)", "Taxonomy ID", "Rank", and "Common Name". The table contains three rows of data, with red circles and numbers (1, 2, 3) highlighting specific cells.

Taxon (4)	Taxonomy ID	Rank	Common Name
(3) Homo sapiens (1)	(2) 9606	species	man; human
(3) - Homo sapiens neanderthalensis	63221	subspecies	Neandertal man; Neanderthal; Neanderthal man; Neandertal
(1) - Homo sapiens ssp. Denisova	741158	subspecies	Denisovan; Denisovans; Denisova hominin

<https://www.ddbj.nig.ac.jp>

# European Molecular Biology Laboratory

The European Bioinformatics x

Secure | <https://www.ebi.ac.uk>

Services Research Training About us

EMBL-EBI Hinxton

## EMBL-EBI

The home for big data in biology

Open data

EMBL-EBI shares data from life science experiments.

More about EMBL-EBI's impact in our annual report [▶](#)  
Data from 2016

Our unique Search service helps you explore dozens of biological data resources.  
[More about EBI Search ▶](#)

Find a tool for your data analysis. [Find a tool ▶](#)

Share your scientific data with the world. [Deposit data ▶](#)

All  [🔍](#)

Example searches: [blast](#) [keratin](#) [bfl1](#)

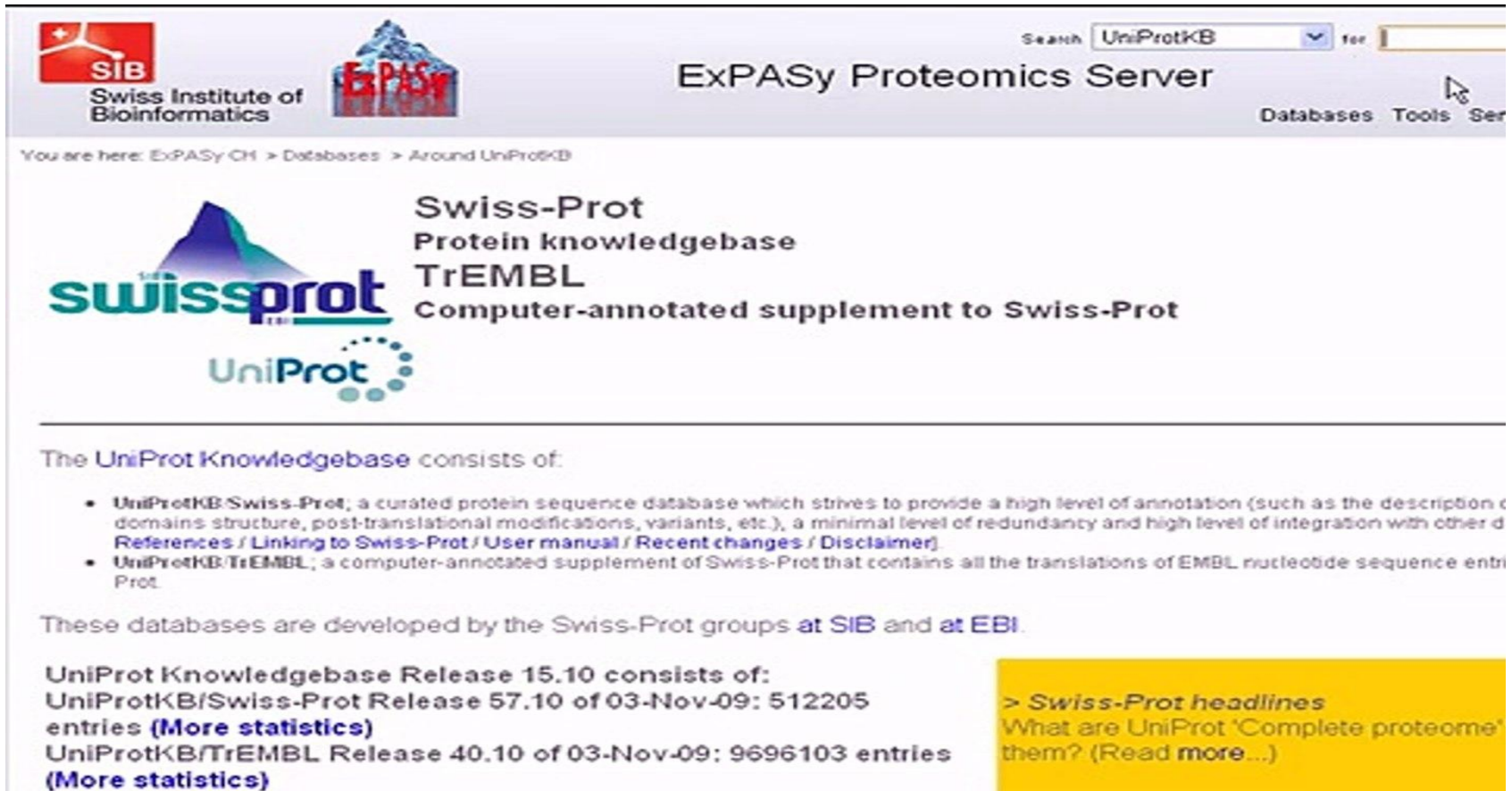
### Explore EMBL-EBI and our mission

The European Bioinformatics Institute (EMBL-EBI) shares [data from life science experiments](#), performs [basic research](#) in computational biology and offers an extensive [user training](#) programme, supporting researchers in academia and [industry](#). We are part of [EMBL](#), Europe's flagship laboratory for the life sciences.  
[More about EMBL-EBI and our impact ▶](#)

[Services](#) [Research](#) [Training](#) [Industry](#) [ELIXIR](#)

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

# SIB Swiss Institute of Bioinformatics SWISS-Port





Search UniProtKB for

ExPASy Proteomics Server

Databases Tools Ser

You are here: ExPASy CH > Databases > Around UniProtKB

 **Swiss-Prot**  
Protein knowledgebase

 **TrEMBL**  
Computer-annotated supplement to Swiss-Prot

The UniProt Knowledgebase consists of:

- **UniProtKB/Swiss-Prot**; a curated protein sequence database which strives to provide a high level of annotation (such as the description of domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases. [References](#) / [Linking to Swiss-Prot](#) / [User manual](#) / [Recent changes](#) / [Disclaimer](#)
- **UniProtKB/TrEMBL**; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries.

These databases are developed by the Swiss-Prot groups [at SIB](#) and [at EBI](#).

UniProt Knowledgebase Release 15.10 consists of:  
UniProtKB/Swiss-Prot Release 57.10 of 03-Nov-09: 512205 entries ([More statistics](#))  
UniProtKB/TrEMBL Release 40.10 of 03-Nov-09: 9696103 entries ([More statistics](#))

> *Swiss-Prot headlines*  
What are UniProt 'Complete proteome' them? ([Read more...](#))

<https://www.expasy.org>

<https://www.ncbi.nlm.nih.gov/genome/gdv/>