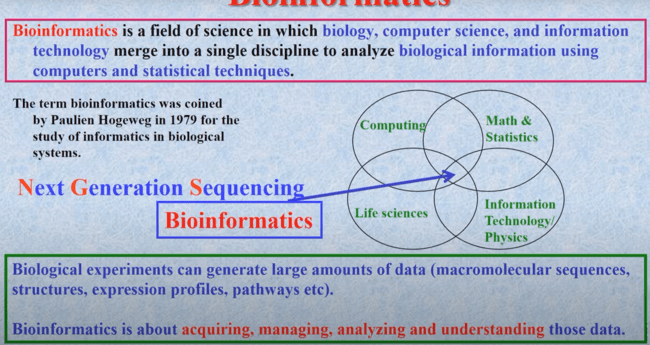
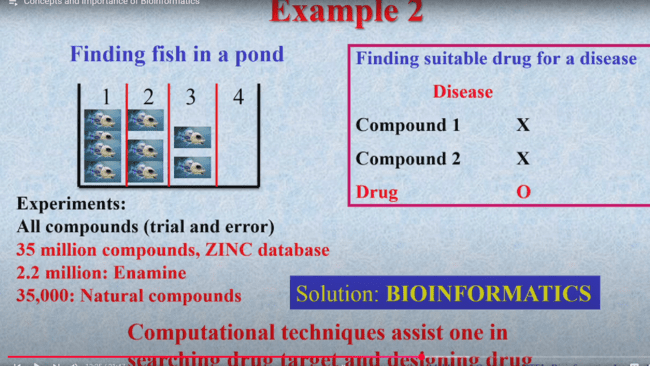
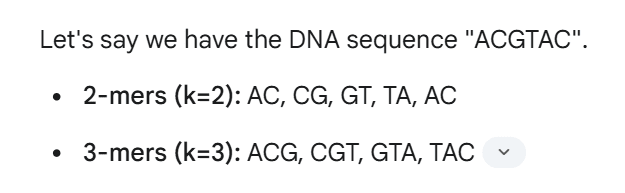
BIOINFORMATICS🡪



In bioinformatics, **k-mers** are subsequences of **length *k* derived** from a longer sequence (typically a DNA, RNA, or protein sequence)

**Length k:** The "k" in k-mer refers to the fixed length of each subsequence. For example, if k = 3, we're dealing with 3-mers (also called triplets). If k = 4, we're dealing with 4-mers, and so on



A screenshot of a white background

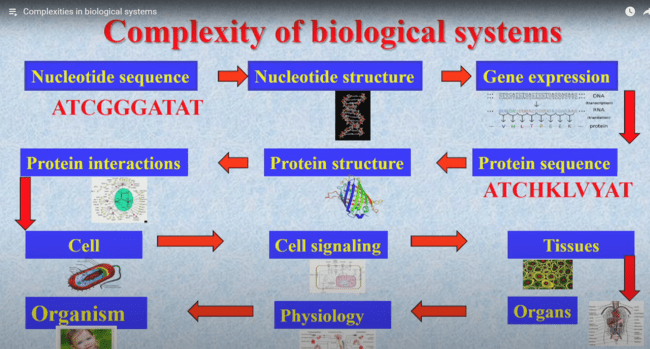
AI-generated content may be incorrect.

**FASTA🡪 FAST ALL**

The FASTA format is a widely used text-based format for representing nucleotide sequences (DNA or RNA) or protein sequences

**Two parts:** Each sequence entry in a FASTA file typically consists of two parts:

* **Header line (definition line):** This line begins with a greater-than symbol (">") and contains a description for the sequence.
* **Sequence data:** This part contains the actual sequence of nucleotides or amino acids.



**Uniprot** 🡪 is the database that contains **protein sequences**

**PIR and SWISSPROT are both protein sequence databases. PIR is maintained by the Protein Information Resource, while SWISSPROT is maintained by the Swiss Institute of Bioinformatics. Both databases contain information about protein sequences, including their functions, structures, and modifications.**

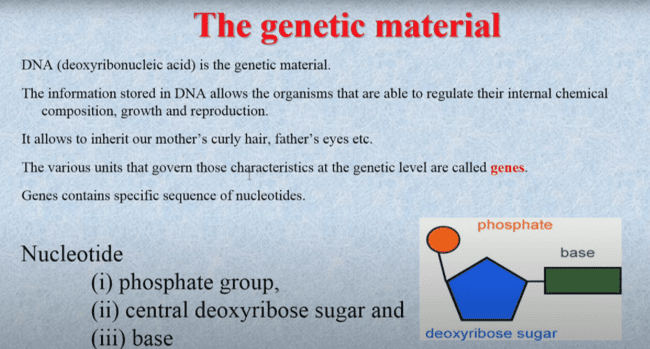
**PIR and SWISS-PROT merged to form the UniProt database.**

* **From protein sequence we will get to know about their Composition**
* **Example 🡪 if a protein sequence having a patch of hydrophobhic amino acids between 15-20 amino acids ,then it can form Alpha helical structure,which usually present in the plasma membrane**
* **+vely charged amino acids(Arginine,histidine,Lysine) binds to the DNA**

**PDB🡪 PROTEIN DATABANK(PROTEIN STRUCTURE)**

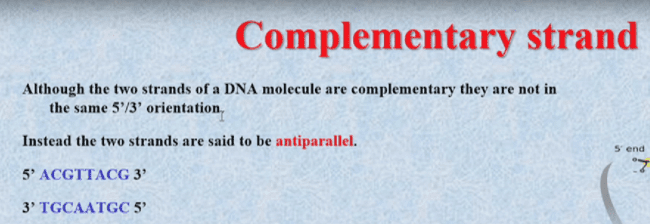
**Techniques: X-ray crystallography, NMR spectroscopy**

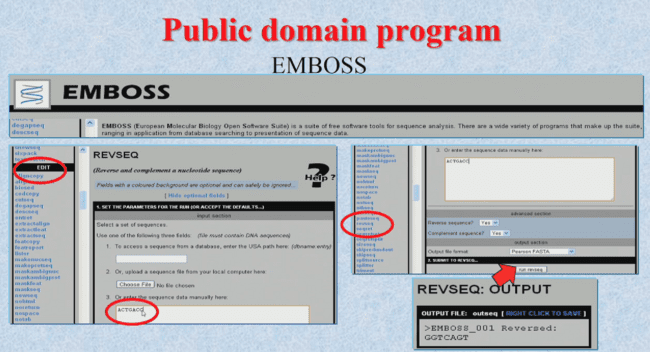
**By comparing the structures of proteins from different organisms, scientists can learn how proteins have evolved over time.**

****

**A diagram of different bases

Description automatically generated**

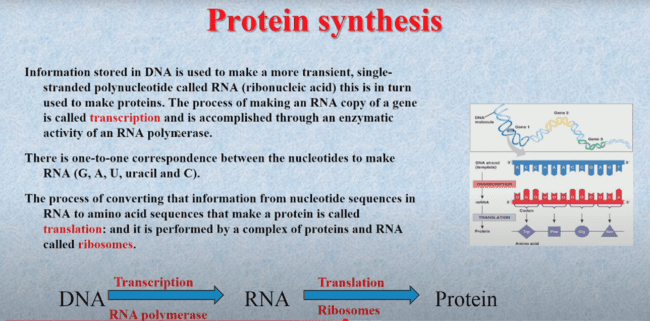
****

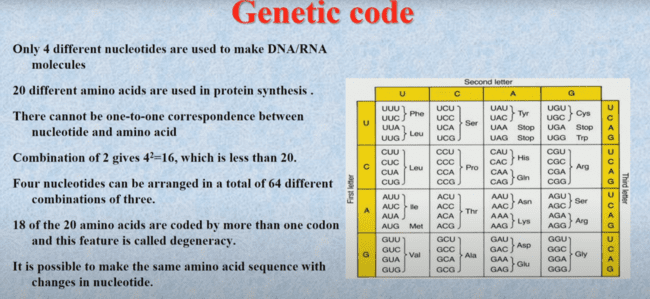
****

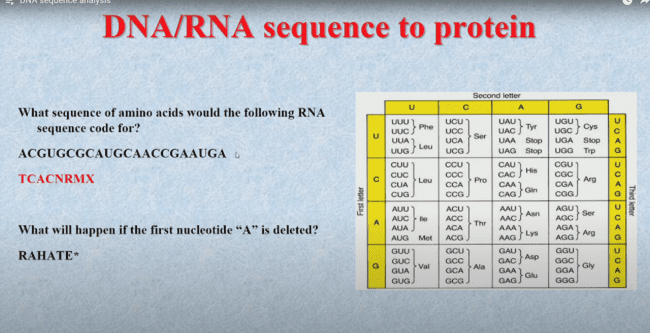
**EMBOSS stands for European Molecular Biology Open Software Suite**

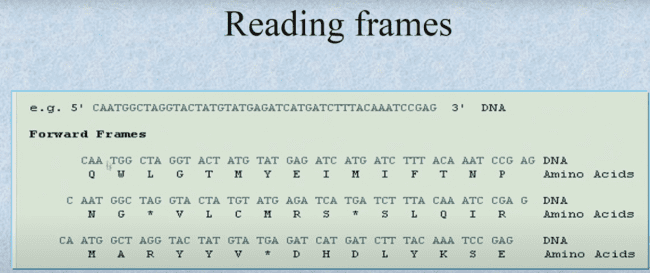
**Use🡪**

1. **To get a complementary strand**
2. **For translation and transcription**

****

****

****

****

**A close-up of a reverse frames

Description automatically generated**

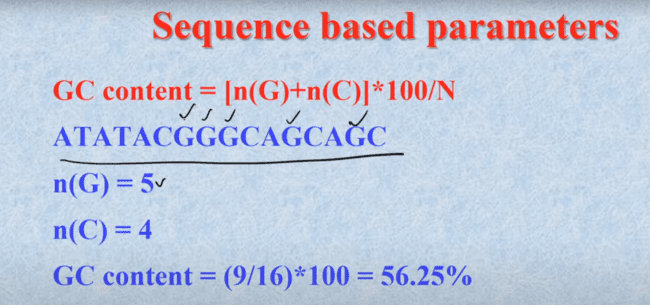
**GC/AT CONTENT🡪**

**Stability: GC pairs have three hydrogen bonds, while AT pairs have two. Higher GC content generally indicates a more stable DNA molecule, requiring higher temperatures to separate the strands.**

**Melting Temperature: This is crucial in PCR, where primers with higher GC content have higher melting temperatures, ensuring efficient amplification.**

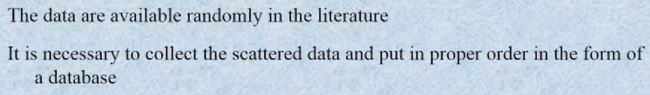
**GC CONTENT MORE🡪 Structure is more regid**

**AT CONTENT MORE🡪 Structure is much more flexible**

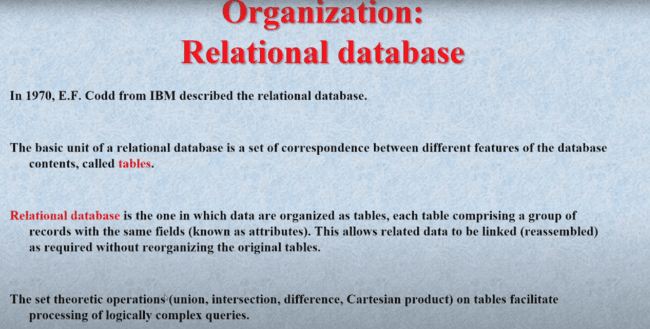
****

**A red letter on a white surface

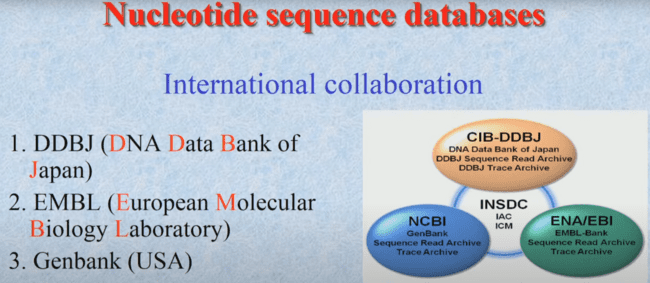
Description automatically generated**

****

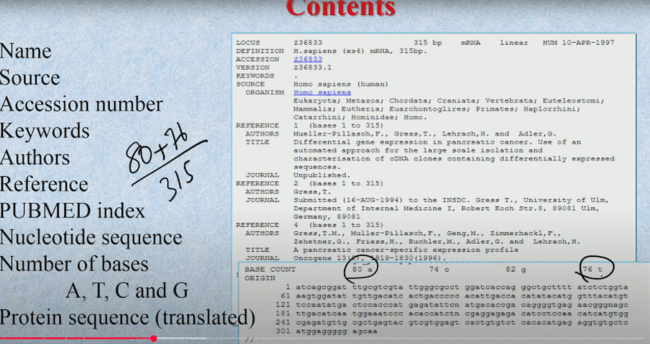
****

****

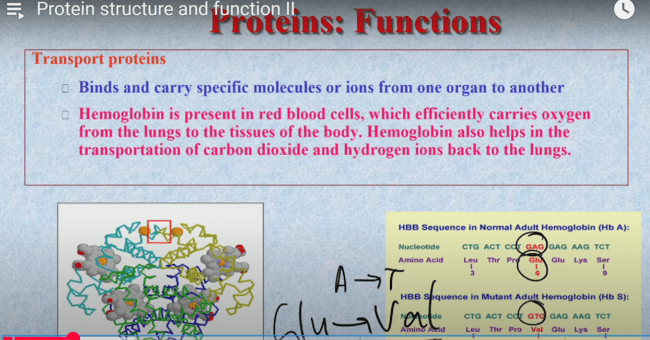
**The International Nucleotide Sequence Database Collaboration (INSDC) is a global collaboration between three organizations that maintain the world's most comprehensive databases of DNA and RNA sequences.**

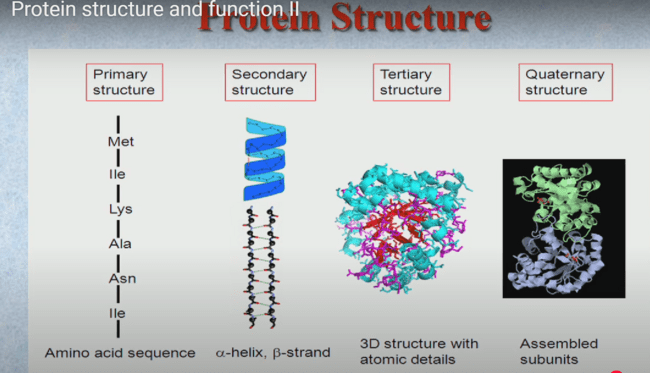
****

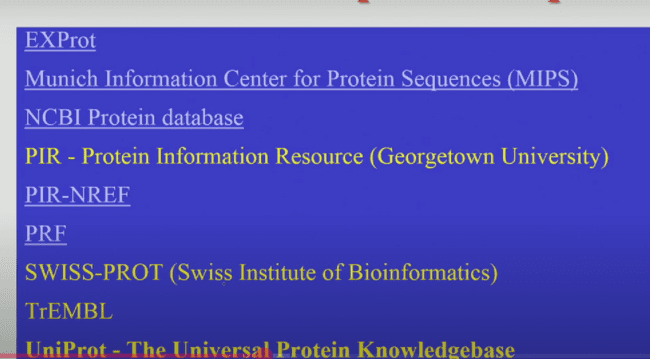
**DDBJ🡪**

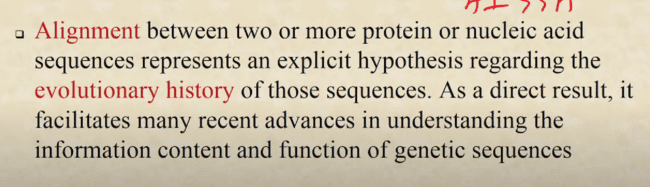
****

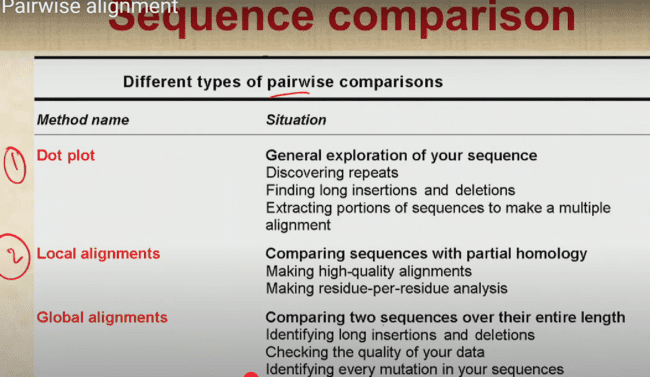
****

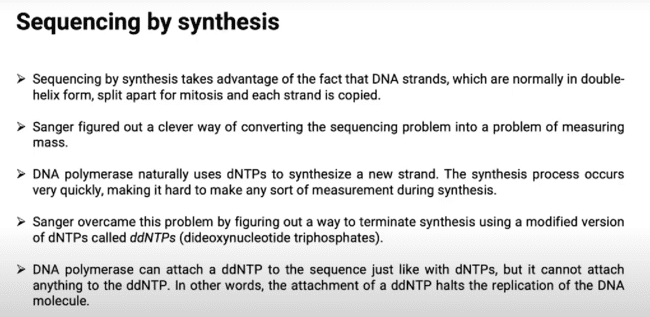
****

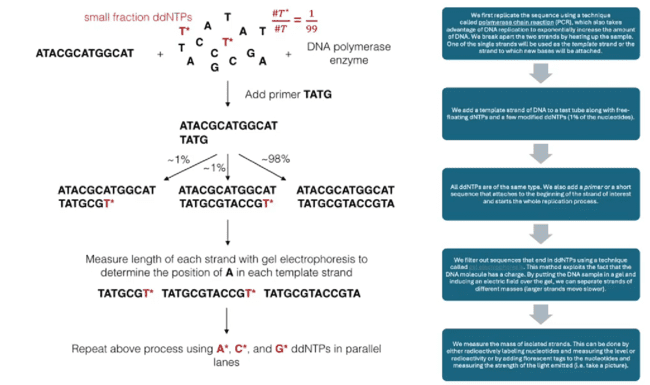
****

****

****

****

****

****

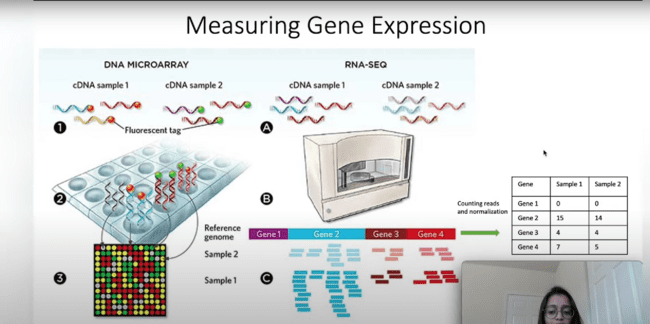
**A close up of a sign

AI-generated content may be incorrect.**

**A diagram of a cell

AI-generated content may be incorrect.**

**Whole concept of dna microarray based upon dna dna hybridization**

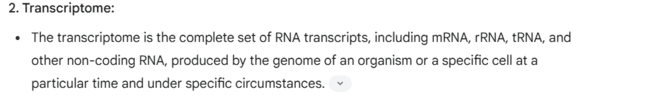
****

**A white text on a white background

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

****

**A screenshot of a computer

AI-generated content may be incorrect.**

**PHENOL CHLOROFORM** METHOD IS USED TO EXTRACT RNA .

**BLAST🡪 Tool to identify similarity of a query sequence to existing sequences available in databanks.**

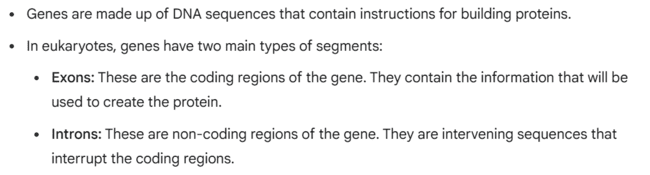
**AlphaFold (DeepMind/Google): AI tool to predict protein structure**

**PDB🡪 Database for 3D protein structure**

**Bowtie : Tool to align short read DNA SEQUENCES obtained from NGS to a reference genome**

**In the replication process Nucleotides are comes in the form of triphosphates on the 3 prime end ,but they are incorporated in the Monophosphate form in the replication.**

**Beta and gamma phosphate(Pyrophosphate) got released.**

****

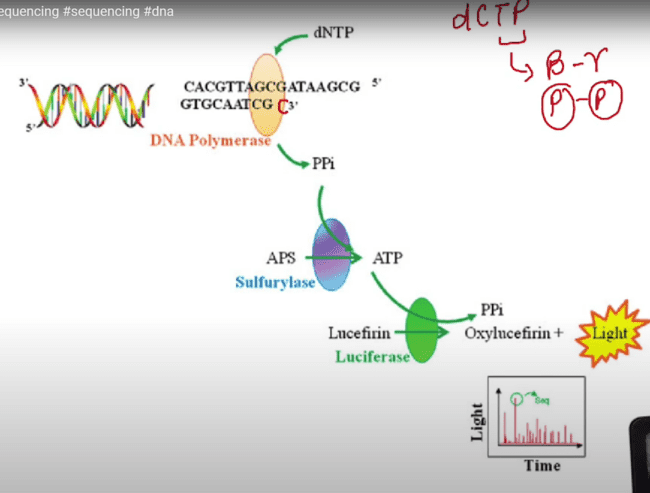
**A white text on a black background

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**PYROSEQUENCING🡪**

****

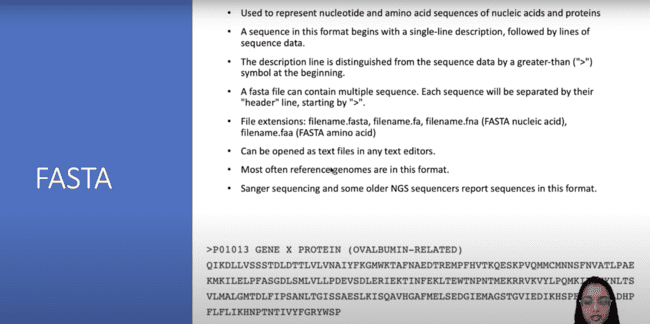
**A screenshot of a computer

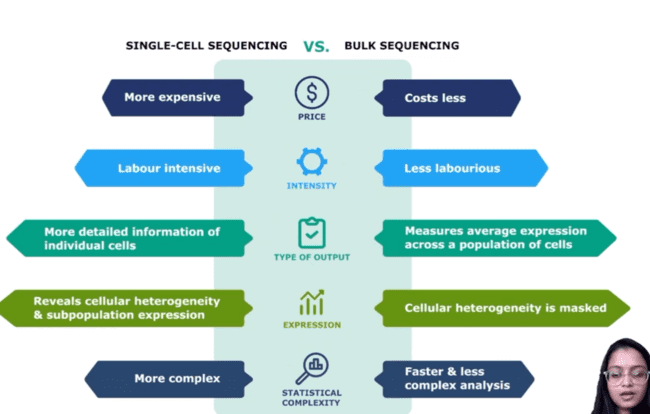
AI-generated content may be incorrect.**

****

**Several different machines with text

AI-generated content may be incorrect.**

****

****

**A screenshot of a cell phone

AI-generated content may be incorrect.**