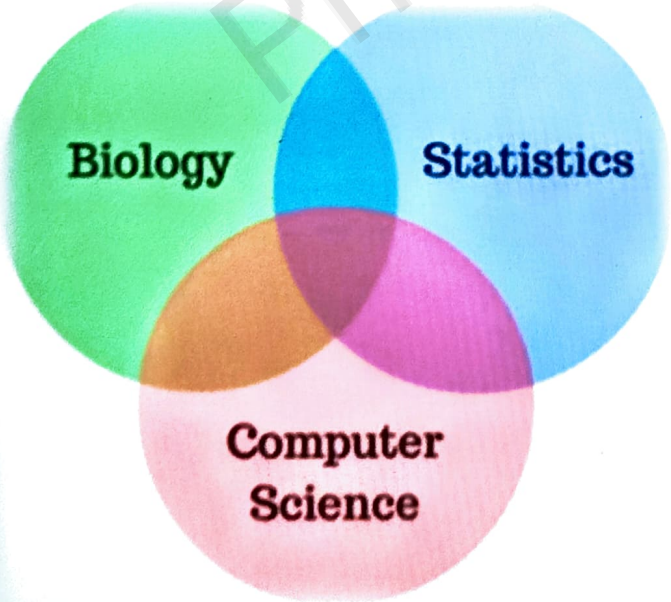
UNIT-IV

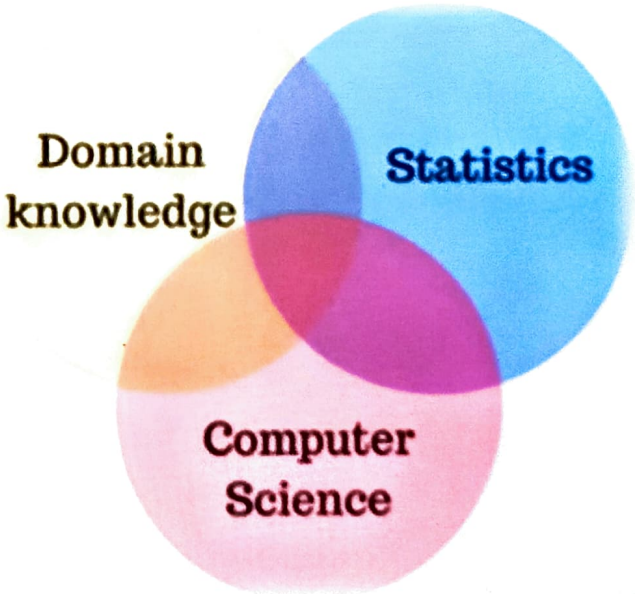
Informatics & methods in drug design

Points to be covered in this topic

- ❑ INTRODUCTION TO BIOINFORMATICS
- ❑ CHEMOINFORMATICS
- ❑ ADME DATABASE
- ❑ CHEMICAL DATABASE
- ❑ BIOCHEMICAL DATABASE
- ❑ PHARMACEUTICAL DATABASE

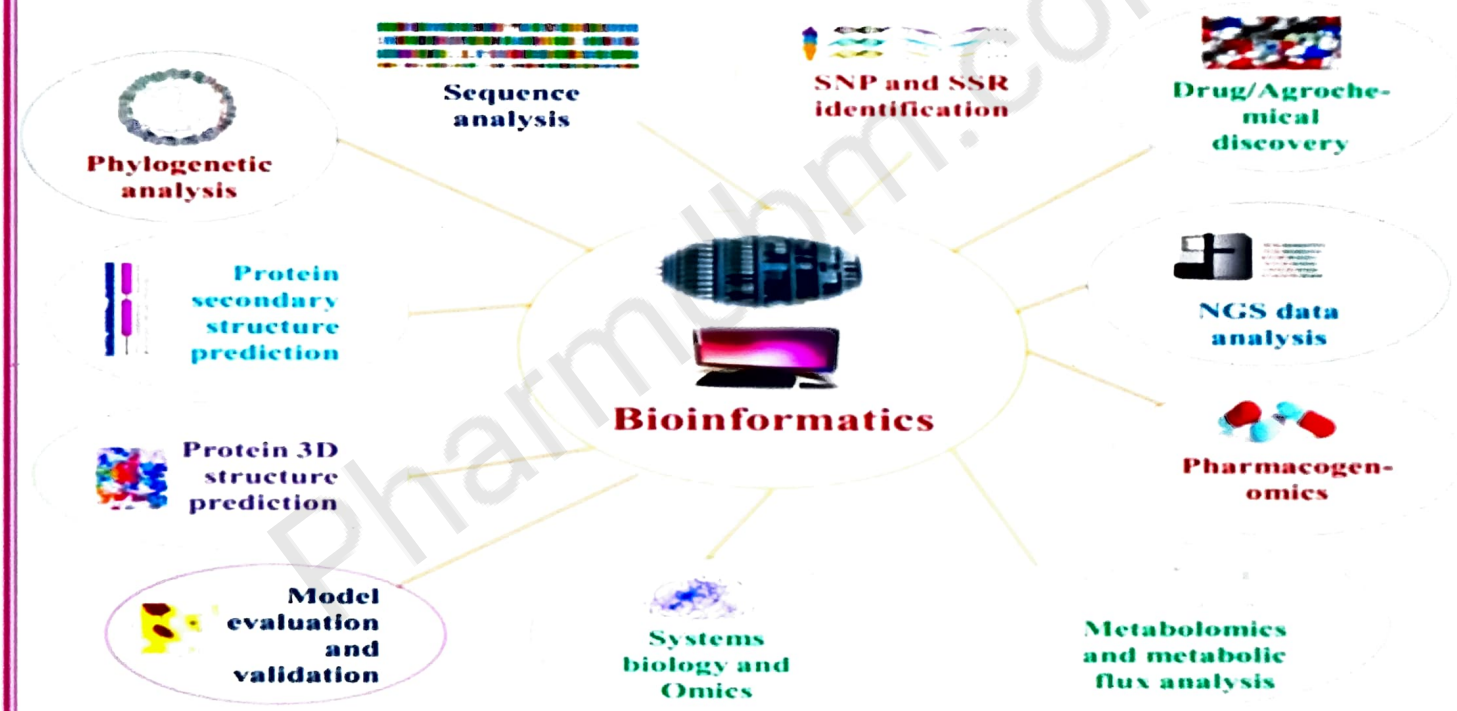
Pharmdbm.com





❑ INTRODUCTION TO BIOINFORMATICS

- The processes of **designing a new drug** using bioinformatics tools have open a new area of research.
- Bioinformatics Supports computer-aided drug design Research where computational methods are **used to simulate drug-receptor interactions.**
- One search method is virtual high-throughput screening. In VHTS, protein targets are screened against databases of small-molecule compounds to see which molecules **bind strongly to the target.**
- If there is a **"hit"** with a particular compound, it can be extracted from the database for further testing.



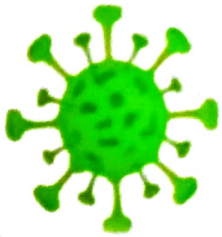
- Most drug candidates fail in Phase III clinical trials after many years of research and millions of dollars have been spent on them, and most fail **because of toxicity or problems with metabolism.**
- The key characteristics for drugs are: **Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) and efficacy-in other words bioavailability and bioactivity.**
- Although these properties are usually measured in the lab, they can also be predicted in advance with bioinformatics software to save cost.

❖ Bioinformatics Tools

- The processes of designing a new drug using bioinformatics tools have open a new area of research. However, computational techniques assist one in searching **drug target and in designing drug in silico, but it takes long time and money.**

➢ In order to design a new drug one need to follow the following path

1. Identify target disease:-



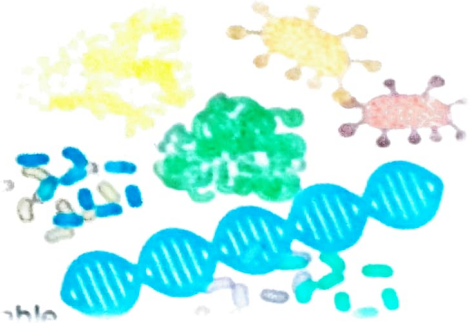
- One needs to know all about **the disease and existing or traditional remedies.** It is also important to look at very similar afflictions and their known treatments.
- Target identification alone is not sufficient in order to achieve a successful treatment of a disease. **A real drug needs to be developed.**
- This drug must influence the target protein in such a way that **it does not interfere with normal metabolism.**
- Bioinformatics methods have been developed to virtually screen the target for compounds that **bind and inhibit the protein.**

2. Study Interesting Compounds

- One needs to **identify and study the lead compounds** that have some activity against a disease.
- These may be only marginally useful and may have severe side effects.
- These compounds provide a **starting point** for refinement of the chemical structures.

3. Detection the Molecular Bases for Disease

- If it is known that a **drug must bind to a particular spot** on a particular protein or nucleotide then a drug can be tailor made to bind at that site.

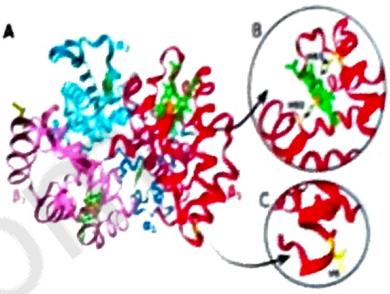


Pharmdbm.com

- This is often model computationally using any of several different techniques.
- Traditionally, the primary way of **determining what compounds would be tested computationally** was provided by the researchers' understanding of molecular interactions.
- A second method is the **brute force testing** of large numbers of compounds from a database of available structures.

4. Rational Drug Design Techniques

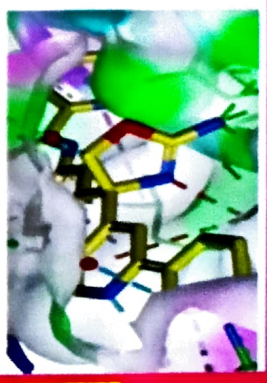
- These techniques attempt to reproduce the researchers' understanding of **how to choose likely compounds built** into a software package that is capable of modeling a very large number of compounds in an automated way.



- Many different algorithms have been used for this **type of testing**, many of which were adapted from artificial intelligence applications.
- The complexity of biological systems makes it **very difficult to determine the structures of large biomolecules.**
- Ideally experimentally **determined (x-ray or NMR) structure is desired,** but **biomolecules are very difficult to crystallize.**

5. Refinement of Compounds

- Once you got a number of lead compounds have been found, computational and laboratory techniques have been very successful in refining the molecular structures to give a greater drug activity and fewer side effects.
- Done both in the laboratory and computationally by examining the molecular structures to **determine which aspects are responsible for both the drug activity and the side effects.**



6. Quantitative Structure Activity Relationships (QSAR)

- Computational technique should be **used to detect the functional group** in your compound in order to refine your drug.

Pharmdbm.com

7. Solubility of Molecule

- One need to check whether the **target molecule is water soluble** or readily soluble in fatty tissue will affect what part of the body it becomes concentrated in.
- The **ability to get a drug** to the correct part of the body is an important factor in its potency.

8. Drug Testing

- Once a drug has been shown **to be effective** by an initial assay technique, much more testing must be done before it can be given to human patients.
- Animal testing is the primary type of testing at this stage. Eventually, the compounds, **which are deemed suitable at this stage, are sent on to clinical trials.**
- In the clinical trials, **additional side effects may be found and human dosages are determined.**

❖ OBJECTIVES OF BIOINFORMATICS

- At its simplest and basic level, bioinformatics organizes data in a way that **allows researchers to access existing information and to submit new entries,** as produced (e.g.) the protein data Bank for 3D macromolecular structures.
- The second key objective is **to develop tools and resources** that aid in the analysis of data. **For example**, having sequenced a particular protein, it is of interest to compare it with previously characterized sequences.
- The third objective is **to use these tools to analysis the data and interpret the results** in a biologically meaningful manner. Traditionally, biological studies examined individual systems in detail, and frequently compared them with a few that are related.

❖ APPLICATION

- Bioinformatics is used in **primer design.**
- Bioinformatics is **used to attempt to predict the function** of actual gene products.

Pharmdbm.com

- Molecular modeling/structural biology is a growing field which can be considered part of bioinformatics.
- There are other fields- **for example,** Image analysis, that might be considered part of bioinformatics.
- There is also a whole other discipline of **biologically inspired computation: Genetic algorithms, AI, Neural networks etc.**
- Sequence analysis has an **importance** in bioinformatics.
- Micro array data provided a great challenge for computational techniques, because of **their large dimensionality and their small sample sizes.**
- Mass spectrometry (MS) technique is emerging as a new and attractive tool for **disease diagnosis and protein-based biomarker profiling.**

❑ CHEMOINFORMATICS

- The set of computer algorithms and tools to store and analyse chemical data in the context of **drug discovery and design projects etc.**
- The mixing of information resources to **transform data into information and information** into knowledge, for the intended purpose of making better decisions faster in the arena of **drug lead identification and optimization.**
- Chemoinformatics encompasses the **design, creation, organisation, management, retrieval, analysis, dissemination, visualization and use of chemical information.**
- A fundamental assumption of any chemoinformatics study is the correctness of the input data generated by experimental scientists and available in various data sets.
- The issue of **chemical data curation in a systematic way** pursues the following major goals:

a) To alert the chemoinformatics and molecular modelling community **towards the significance of chemical and bioactivity data.**

b) To develop a set of data curation procedures that would **process the input data and correct structural errors whenever possible.**

- To share organized protocols for data curation with the scientific community by providing sample case studies and explicit pointers to the sources.
- To illustrate, with at least a **few examples,** that carefully developed QSAR models using well curated primary data may be employed not only for predicting new structures but also to spot and correct errors in biological data reported in databases used either for model development or model validation.

❖ Steps for chemical data curation

1. Removal of Inorganics and Mixtures.

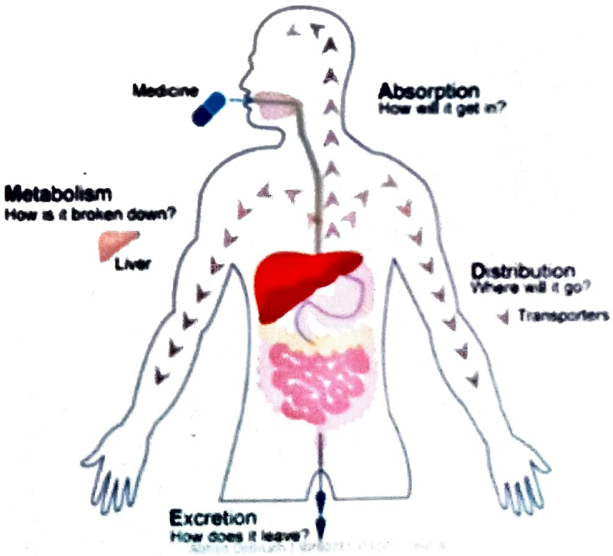
- The majority of molecular descriptors can be computed for organic compounds only, not inorganic molecules.
- The limitation of conventional chemoinformatics software is its inability to model inorganics.
- There is a need to develop ideal chemical descriptors for these molecules and include them in descriptor developing software.
- At present, all inorganic compounds must be removed before the descriptors are calculated.

2. Structural Conversion and Cleaning.

- In this step the SMILES string is converted into 2D molecular graphs. Many programs use this method like **ChemAxon, MOE, 34Sybyl, 37 Open Babel etc.**

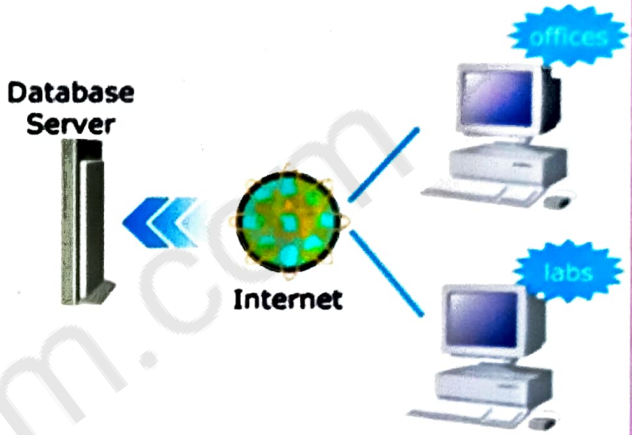
❑ ADME DATABASE

- It provides the latest and most comprehensive data on specific interactions of substances with Human Drug Metabolizing Enzymes and Human Transporters.
- It is designed for use in drug research and development,



- Including drug-drug interactions and ADME **(Absorption, Distribution, Metabolism and Excretion)** studies.
- There is considerable interest in developing either experimental or computational methods that can **identify problematic drug leads at the earliest stages in their development.**
- Most pharmaceutical scientists refer to this **biological processing as ADME (Absorption, Distribution, Metabolism and Excretion).**
- It may also be called ADMET where the **'T' refer to toxicology.**

❖ Advantages



- **No additional software needed.**
- **No need to setup a dedicated server within a company.**
- **Quick login to the system using only a user ID and password.**
- **Fujitsu's tight security measures being implemented**

❖ List of ADMET database:-

- **PACT-F: Preclinical And Clinical Trials Knowledge Base on Bioavailability (PACT-F).**
- **TOXNET: Databases on toxicology, hazardous chemicals, environmental health, and toxic releases** that can be accessed using a common search interface, provided by the United States NLM.
- **Leadscope Toxicity Database:** Database of 160,000 chemical structures with toxicity data, Distributed by Leadscope.
- **WOMBAT-PK:** Database for Clinical Pharmacokinetics and Drug Target Information. **WOMBAT-PK contains 1260 entries (1260 unique SMILES),** totaling over 9,450 clinical pharmacokinetic measurements.
- **SIDER: Contains information on marketed medicines and their recorded adverse drug reactions.** The information is extracted from public documents and package inserts.

Pharmdbm.com

- **AdmetSAR:** AdmetSAR provides the manually curated data for diverse chemicals associated with **known Absorption, Distribution, Metabolism, Excretion and Toxicity profiles.**
- **The ADME databases:** Databases for benchmarking the **results of experiments, validating the accuracy of existing ADME predictive models, and building new predictive models.**
- **The ADME database:** **Provides comprehensive data for structurally diverse compounds associated with known ADME properties, including human oral bioavailability, enzymes metabolism, inhibition and induction, transport, plasma protein binding and blood brain barrier,**
- **UCSF-FDA Transportal:** The purpose of this database is to be a **useful repository of information on transporters important in the drug discovery process** as a part of the US Food and Drug Administration-led Critical Path Initiative.

❖ There are three types of databases in bioinformatics:

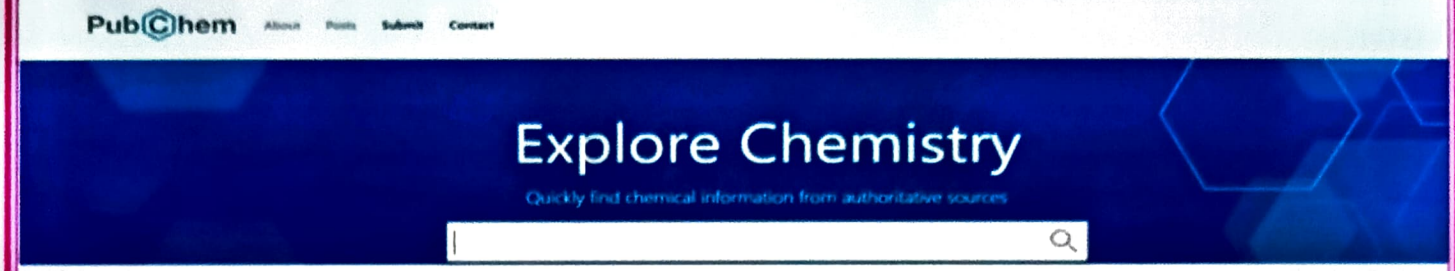
1. **Chemical,**
2. **Biochemical and**
3. **Pharmaceutical Data bases.**

- In the field of bioinformatics, these databases **collect and organize data around a single class of macromolecules (e.g.GPCRs),** or around a particular topic of interest **(e.g. cancer),** whereas in the field of Chemoinformatics, they often collect target-focused small molecules **(e.g. kinase inhibitors).**
- Many bioinformatics databases **cover the 'Target' space** in an integrated manner.
- The primary objective of this database is **proteins.**
- **Whereas, Swiss-Prot is most integrated target database** as it is cross-referenced with more than 60 different databases.

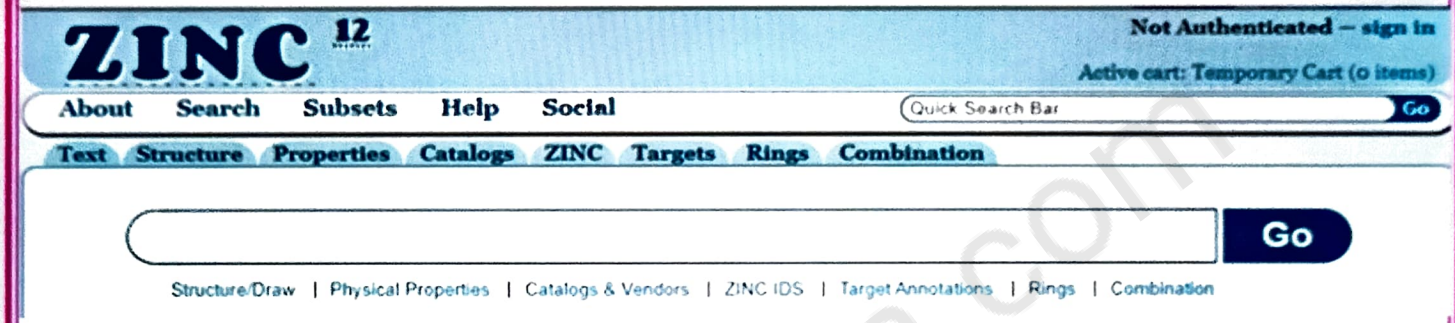
Pharmdbm.com

❑ CHEMICAL DATABASE

- **PubChem** - Maintains three types of information namely, **substance, compound and Bioassay**



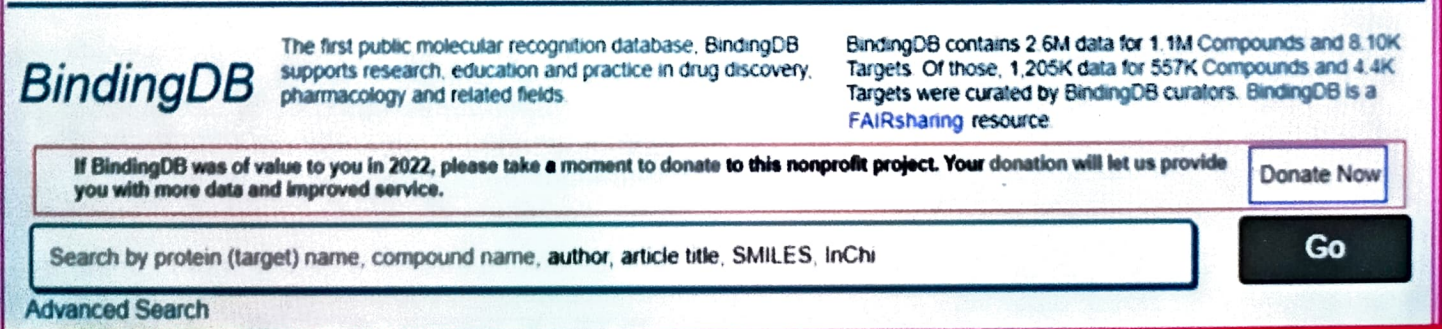
- **ZINC** - Contains **21 million compounds available for virtual screening**



- **DrugBank** - Detailed **drug (i.e. pharmacological and pharmaceutical) data with drug target**

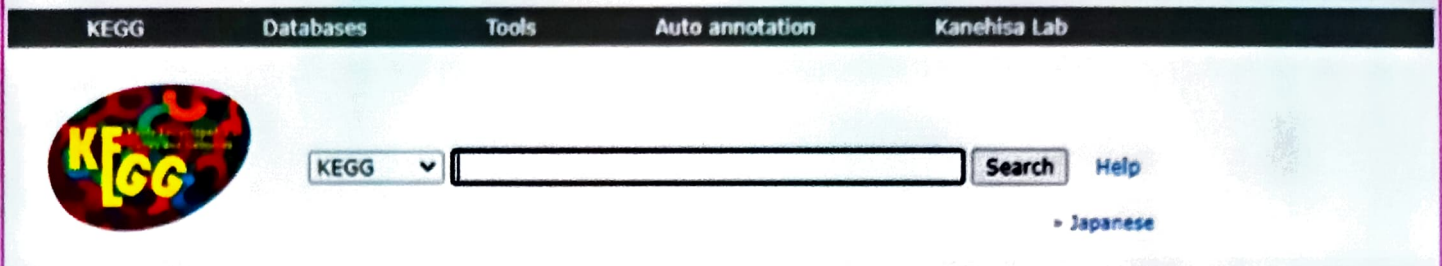
- **ChEMBL** - **1 million bioactive (small drug-like molecules) compounds with 8200 drug targets.**

- **BindingDB** - Contains **910,836 binding data, for 6,263 protein targets and 378,980 small molecules.**

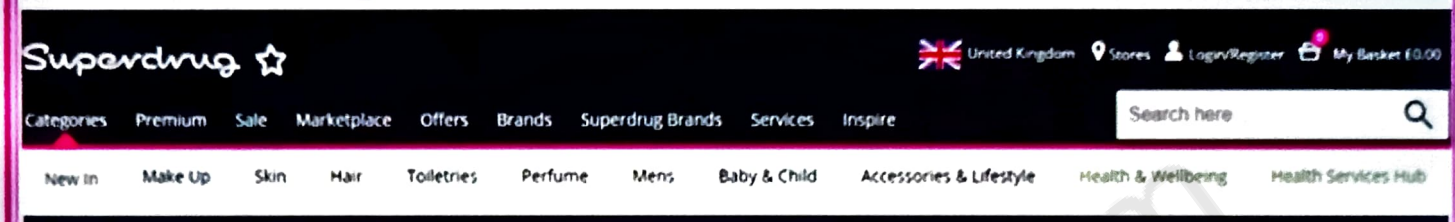


PharmDBM.com

- **KEGG-integrates- Genomic, Chemical and Systemic functional information.**



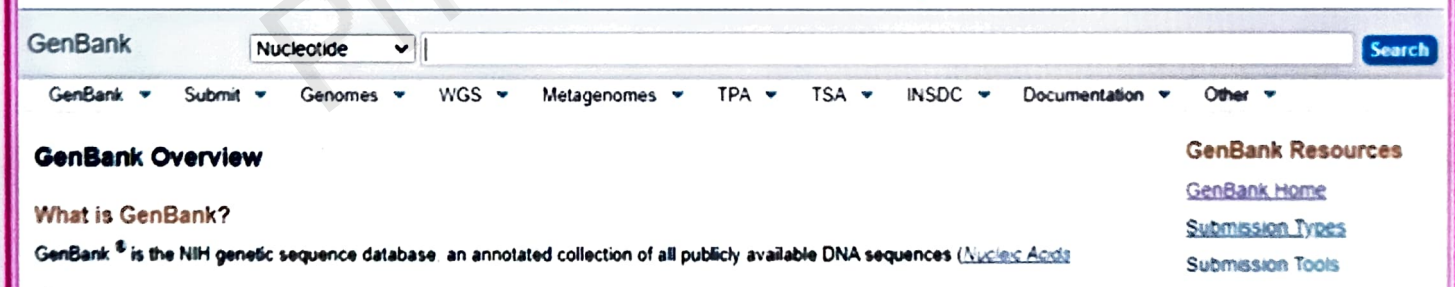
- **SuperDrug - 2500 3D-structures of active ingredients of essential marketed drugs.**



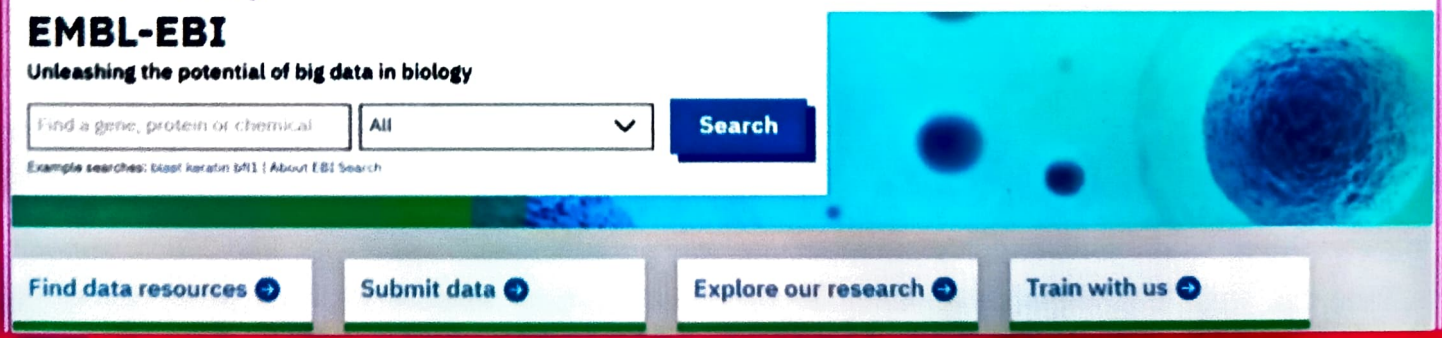
- **Vendor Databases - eMolecules, MolProt, Enamine**

❑ BIOCHEMICAL DATABASE

- **GenBank- GenBank is the genetic sequence database, an annoted collection of all publically available DNA sequences.**



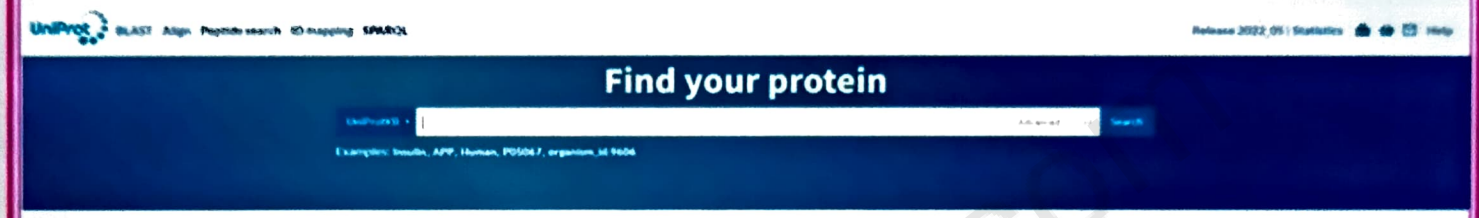
- **EMBL-EBI (European Molecular Biological Laboratory)- It is an Nucleic acid Database that comes under EBI (European Bioinformatics Institute).**



Pharmdbm.com

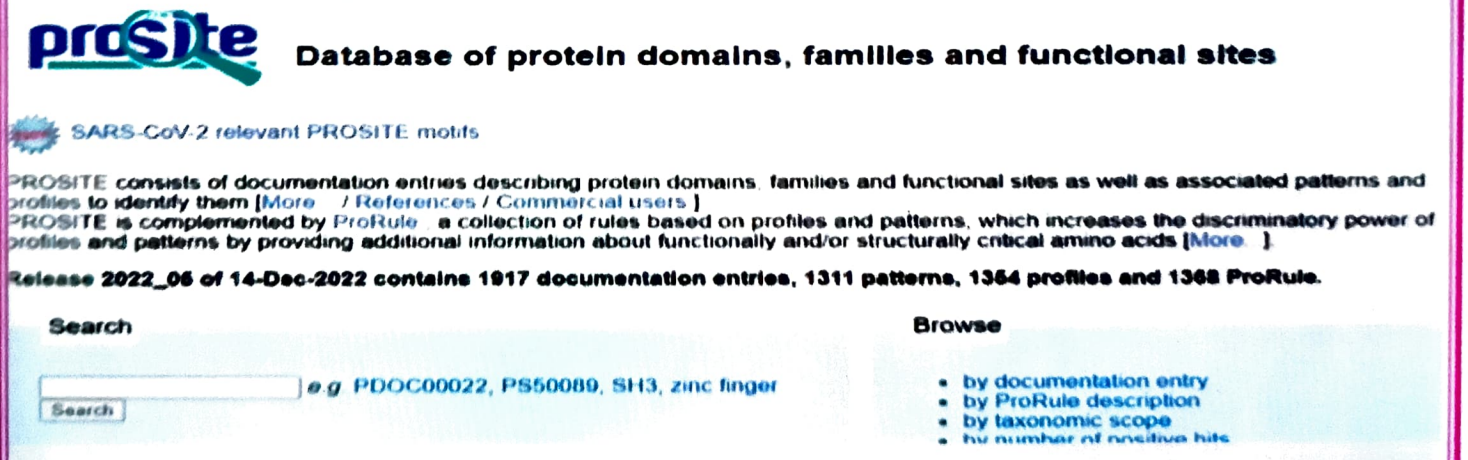
- **DDBJ- DNA Data Bank of Japan** is a biological database that **collects DNA Sequences.**
- **It is located at the National Institute of Genetics in the Shizuoka Prefecture of Japan.**

- **SWISS-PROT is a curated protein sequence database which strives to provide a high level of annotation such as description of the function of proteins, Its domains structure, post translational modifications, variants.**

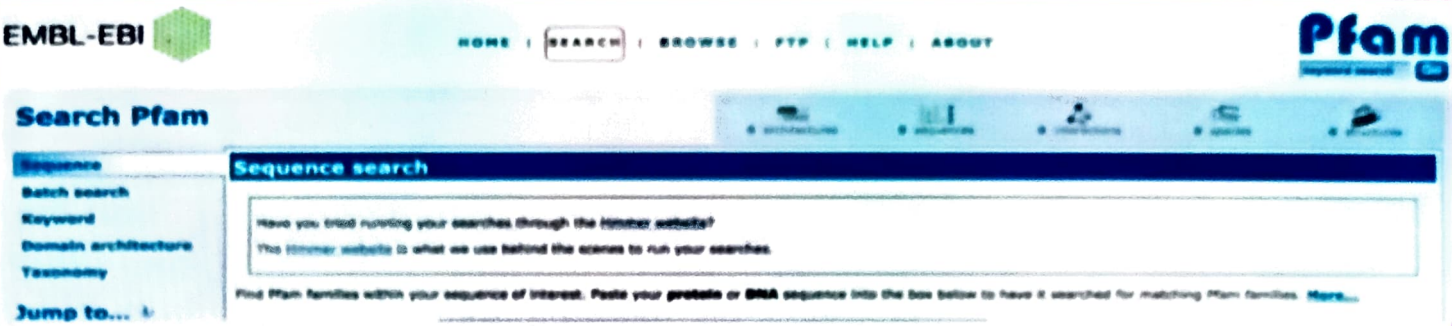


- **PIR- The Protein Information Resources (PIR)** is an integrated public resource of protein informatics that **supports genomic and proteomic research and scientific discovery.**

- **PROSITE is a protein database.** It consist of entries describing the **protein families, domains and functional sites as well as amino acid patterns and profiles in them.**



- **Pfam is a database of protein families that includes their annotations and multiple alignments generated using hidden markov models.**



❑ PHARMACEUTICAL DATABASE

- Drug databases and web resources play a **very important role** in the pharmaceutical field. Check out this lesson to learn all about these databases and resources and how they benefit a pharmacist.
- There are thousands of medications available, These medications range from **simple cough medications to more powerful medications such as chemotherapy drugs.** It would be impossible for any pharmacist to know about each and every drug available.
- There are drug databases and web resources that are **useful tools that can be used to learn more about a medication or drug** that may be somewhat unfamiliar to them.
- **FDA** drug database **includes most of the drugs** they have approved. Best of all, this database is extremely easy to use. To search this database, you simply need to go to the FDA drug databases website. Once you get to this website, you are able to **search the database by typing in the name of the drug or by typing in any active ingredient of a drug.**
- The FDA must APPROVE a drug before it is legally able to be **SOLD and USED. Therefore,** drug companies must formally submit an application to the FDA for the drug to be approved. The drugs that have not been submitted to the FDA but not yet approved can be found in this database.
- **The Orange Book (officially known as the Approved Drug Products with Therapeutic Equivalence Evaluations)** is a publication that provides information about drugs approved by the FDA.

Pharmdbm.com