BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics

Course Code : BM35S1BI

Unit-II

TOPIC: Structuren: PDB and NDB

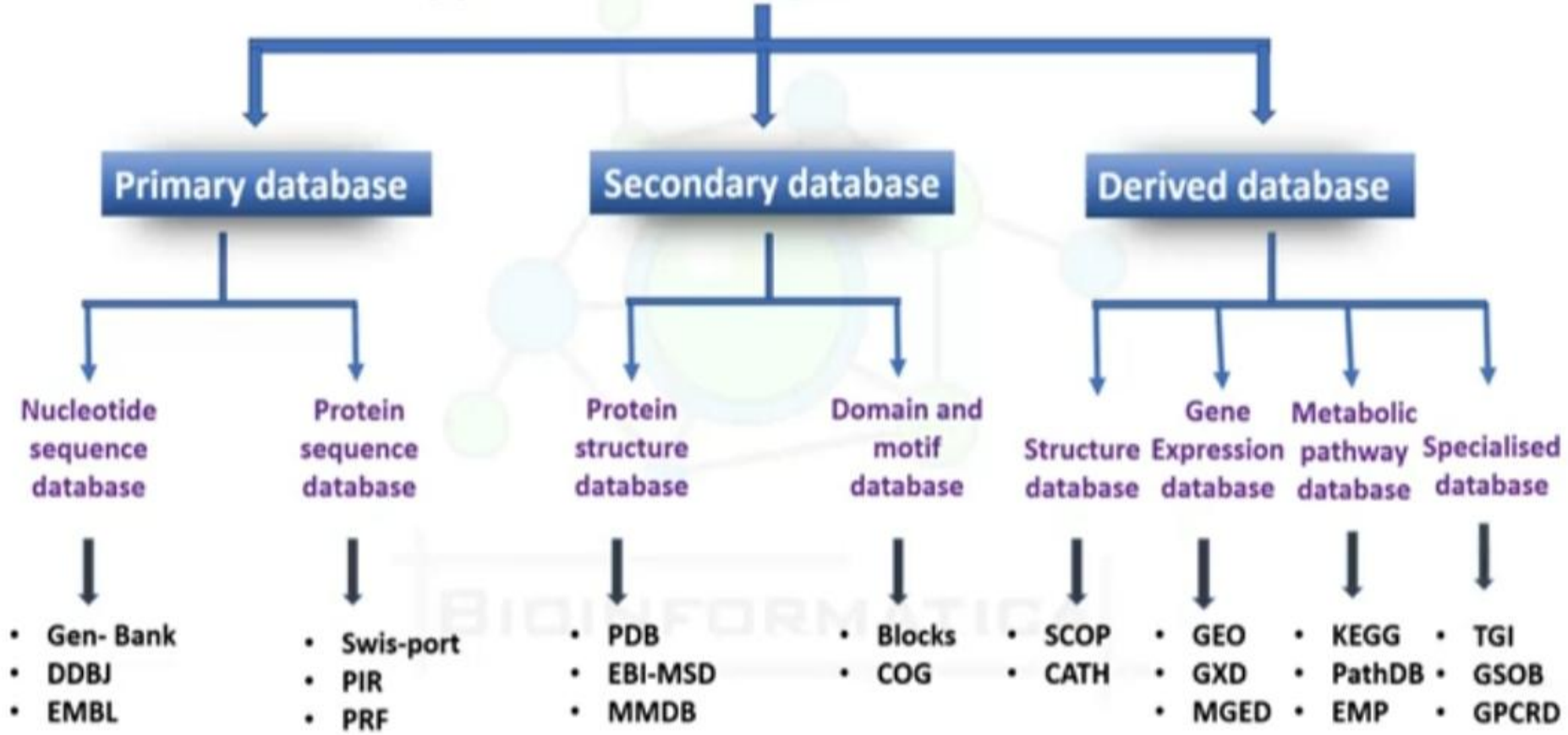
Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

Structuren: PDB and NDB

BIOLOGICAL DATABASES

- This main purpose of a biological databases is to Store and manage biological Data and information Like biological sequence , structures, binding sites, Metabolic interactions , Protein families etc.,(in computer related forms).
- Data are arranged by set of rules Which are programmed into Software that manages the data called Database Management System Or DBMS.

Types of Biological database



STRUCTURE DATABASE

- MSD: The macromolecular structure database – A relational database representation of clean protein data bank(PDB).
- 3DSEQ: 3D sequence alignment server annotation of the alignment between sequence database and the PDB.
- FSSP: based on exhaustive all-against- all 3D structure comparision of protein structure currently in the protein data bank(PDB).
- DALI: Fold classification based on structure assignments.
- 3DEE: Database of protein domain definitions wherein, the domains have been clustered on sequence structural similarity
- NDB: Nucleic acid Structure Database.

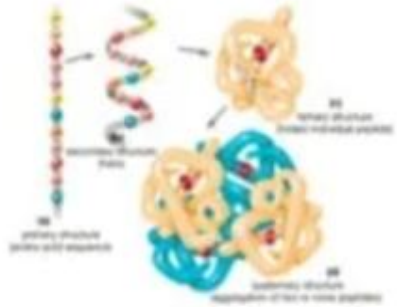
Protein Data Bank(PDB)

- What?
- **A primary Databases for 3D structural information of large biomolecules Such as proteins and Nucleic acids.**
- How?
- **X- ray crystallography**
- **NMR spectroscopy**
- **Cryo electron microscopy(rare)**
- **When & where?**
- **Started 1971 at Brookhaven National lab.**

Introduction

- Archive of experimentally determined 3D structures of biological macromolecules.

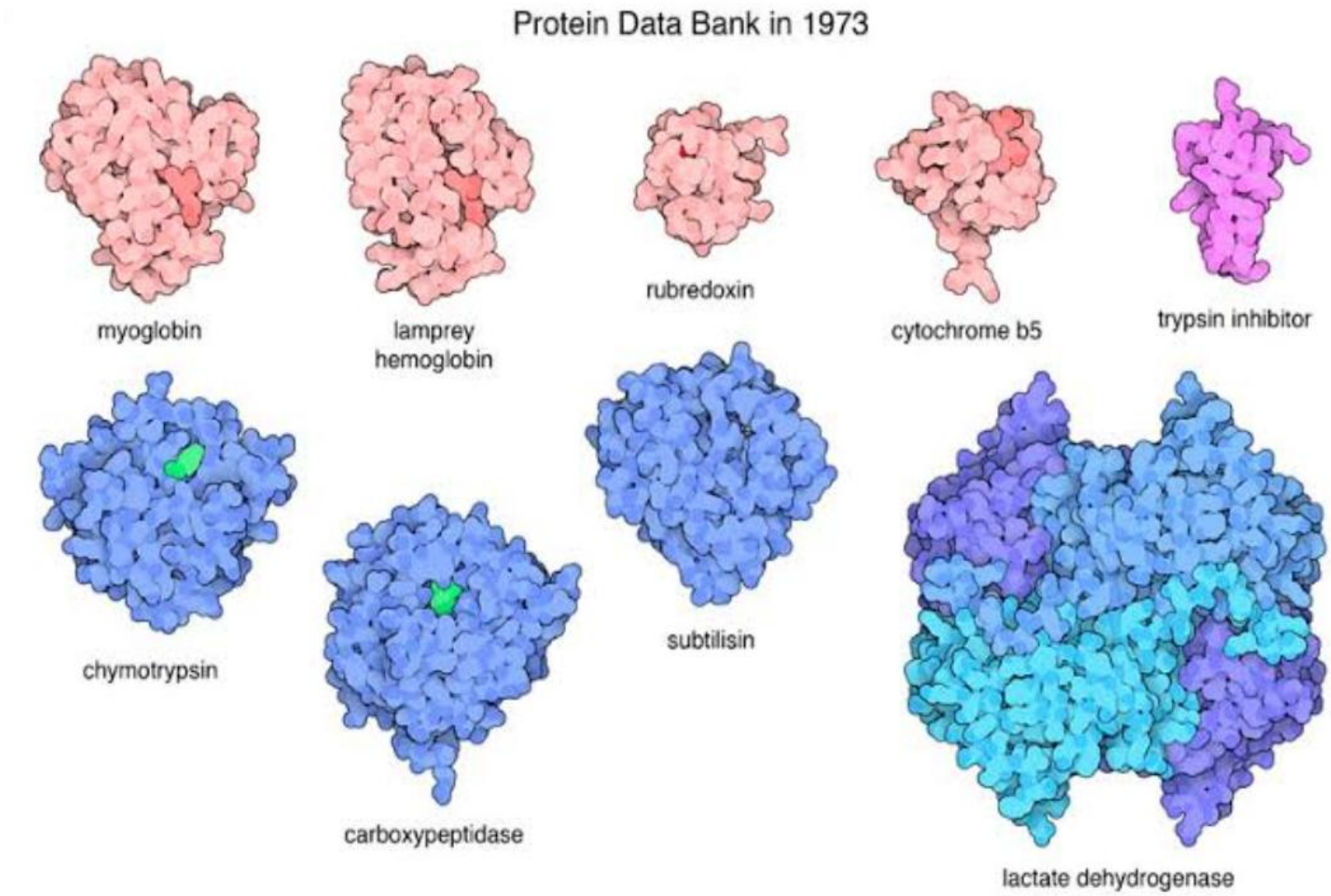
Archive



3D structure of biological macromolecules

Experiments

- Established in 1971, by Research Collaboratory for Structural Bioinformatics (RCSB), Brookhaven National Laboratories, USA.
- Archive contain atomic coordinates, bibliographic citations, primary and secondary structure information, crystallographic structure factors, NMR experimental data.


Protein Data Bank in 1973
myoglobin
lamprey hemoglobin
rubredoxin
cytochrome b5
trypsin inhibitor
chymotrypsin
carboxypeptidase
subtilisin
lactate dehydrogenase

gateways to access PDB files

Swiss-Prot, NCBI, EMBL

CATH, Dali, SCOP, FSSP

databases that interpret PDB files

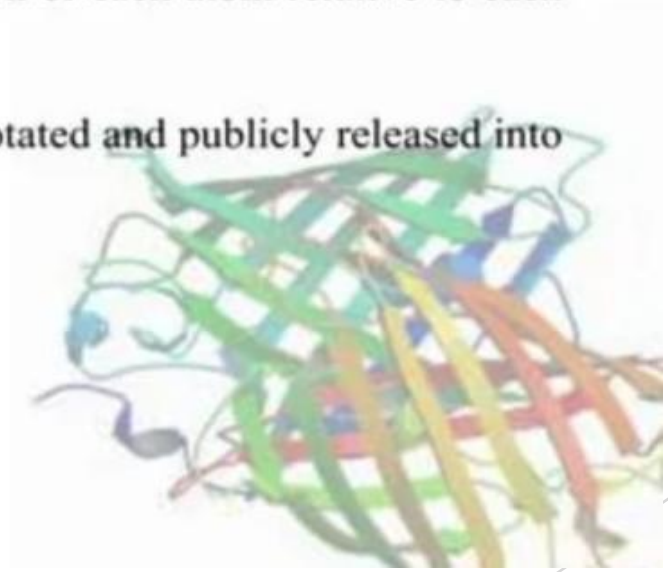
WORLDWIDE PDB
PROTEIN DATA BANK

Single worldwide archive of macromolecular structural data

- **Ensures that the PDB remains a single & uniform archive publicly available to the worldwide community**
- **3 founding members: RCSB PDB, PDBj, MSD-EBI**

How data is collected?

- The PDB is a repository of atomic coordinates and other information describing proteins and other important biological macromolecules
- Structural biologists use methods such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy to determine the location of each atom relative to each other in the molecule
- They then deposit this information, which is then annotated and publicly released into the archive by the wwPDB

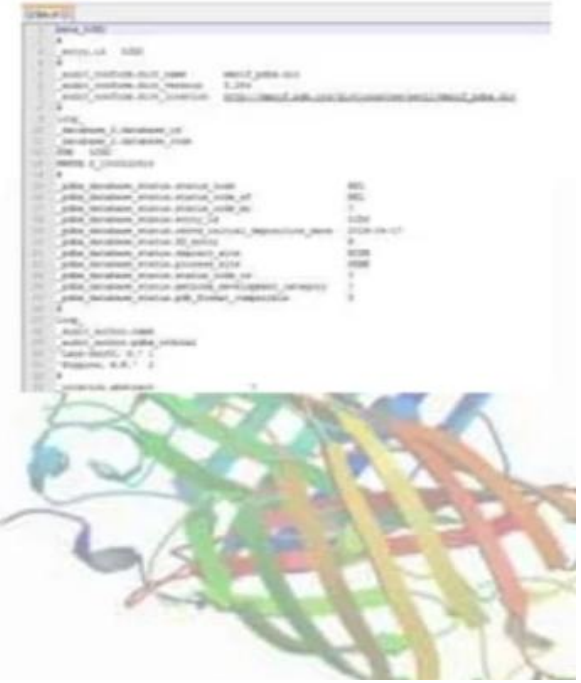


How to search ?

- One can search for their protein of interest by using the search bar in the RCSB PDB website
- It allows one to search either by typing the PDB ID, name of the author (who has deposited the structure), or the sequence of the protein or any particular ligand of interest

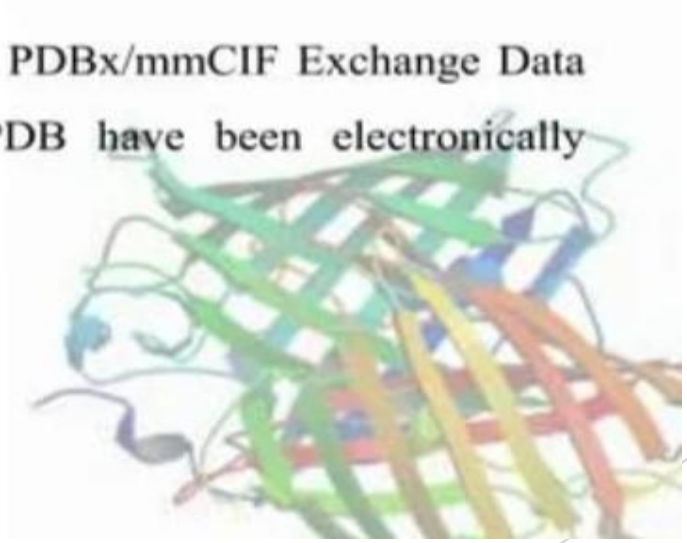
File formats

- The data in PDB is usually stored in 3 different file formats
  - PDB file format
  - mmCIF format
  - PDBML



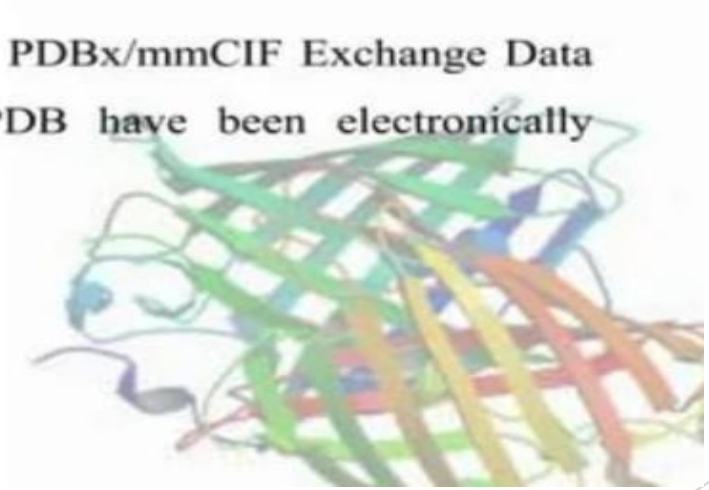
PDB file format

- The Protein Data Bank Markup Language (PDBML) provides a representation of PDB data in XML format
- The description of this format is provided in XML schema of the PDB Exchange Data Dictionary
- This schema is produced by direct translation of the PDBx/mmCIF Exchange Data Dictionary Other data dictionaries used by the PDB have been electronically translated into XML/XSD schemas



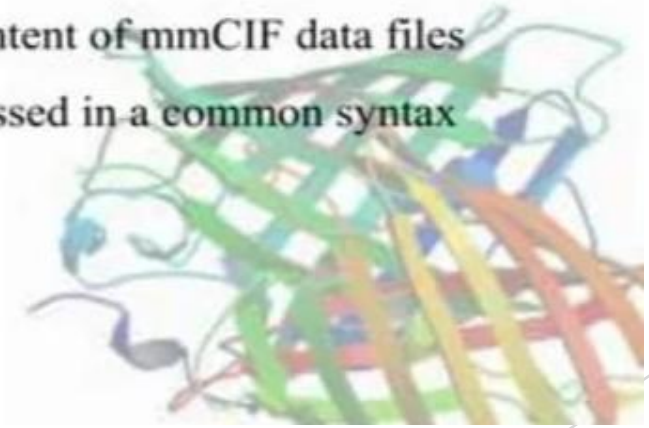
PDBML

- The Protein Data Bank Markup Language (PDBML) provides a representation of PDB data in XML format
- The description of this format is provided in XML schema of the PDB Exchange Data Dictionary
- This schema is produced by direct translation of the PDBx/mmCIF Exchange Data Dictionary Other data dictionaries used by the PDB have been electronically translated into XML/XSD schemas



mmCIF

- mmCIF is the acronym for the macromolecular Crystallographic Information File
- mmCIF is based on a subset of the syntax rules for the Self Defining Text Archive (STAR) file
- A Dictionary Description Language (DDL) defines the structure of mmCIF dictionaries
- Dictionaries provide the metadata which define the content of mmCIF data files
- mmCIF data files, dictionaries and DDLs all are expressed in a common syntax



PDB HEADER PLANT SEED PROTEIN 11-OCT-91 1CBN

mmCIF

```
_struct.entry_id '1CBN'
_struct.title 'PLANT SEED PROTEIN'
_struct_keywords.entry_id '1CBN'
_struct_keywords.text 'plant seed protein'
_database_2.database_id 'PDB'
_database_2.database_code '1CBN'
_database_PDB_rev.rev_num 1
_database_PDB_rev.date_original '1991-10-11'
```

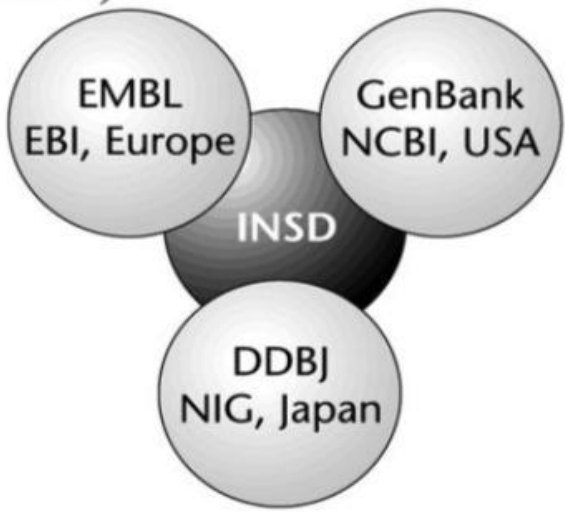
Sequence databases

Nucleotide databases

- **Protein databases.**
- **The biological information Of nucleic acid is available as sequence while the data of protein are Available as Sequences and structures.sequences are represented in a Single dimension Whereas the structure contains the three dimensional Data of sequences.**
- **The database is complEmented with generalized Software for processing,archiving, Quering and distributIng data.**
- **Such databases consisting Of nucleotide sequences are called Nucleic acid sequence databases.**

Nucleotide Sequence Databases

- International Nucleotide Sequence Database (INSD)



1. GenBank at NCBI
2. European Molecular Biology Laboratories (EMBL) Nucleotide Sequence Database at European Bioinformatics Institute (EBI)
3. DNA database of Japan (DDBI)

- Features
  - All published nucleotide sequences are requested to be deposited in the one of these three databases;
  - Data are exchanged among these three databases on daily basis;
- Sequences stored in the GenBank at NCBI can be downloaded by anonymous ftp at ftp://ftp.ncbi.nih.gov.

Nucleic acid structure databases

- NDB Nucleic acid-containing structures http://ndbserver.rutgers.edu/
- NTDB Thermodynamic data for nucleic acids http://ntdb.chem.cuhk.edu.hk/
- RNABase RNA-containing structures from PDB and NDB http://www.rnabase.org/
- SCOR Structural classification of RNA: RNA motifs by structure, function and tertiary interactions
- http://scor.lbl.gov/


DDBJ
DNA Data Bank of Japan
National Institute of Genetics
INSDC
GenBank
NCBI
ENA
European Nucleotide Archive
EMBL-EBI

Nucleotide databases

GENE BANK: it is a Us primary sequences Databses established in 1998 which is maintained by NCBI(National centre for biotechnology information).It is very comprehensive Biological databases and provides 42 different resources.It Provides a simple And easy to Use wen interface.

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

EMBL: (European molecular biology laboratory) it is Europe's Primary nucleotide sequence Resource established by EMBl and maintained by EBI(European Bioinformatics institute) SRS(sequence retrieval Tool) is a retrieval Tool. https://www.embl.org/

DDBJ(DNA databank of Japan) mainly collets data from Japanese activities.

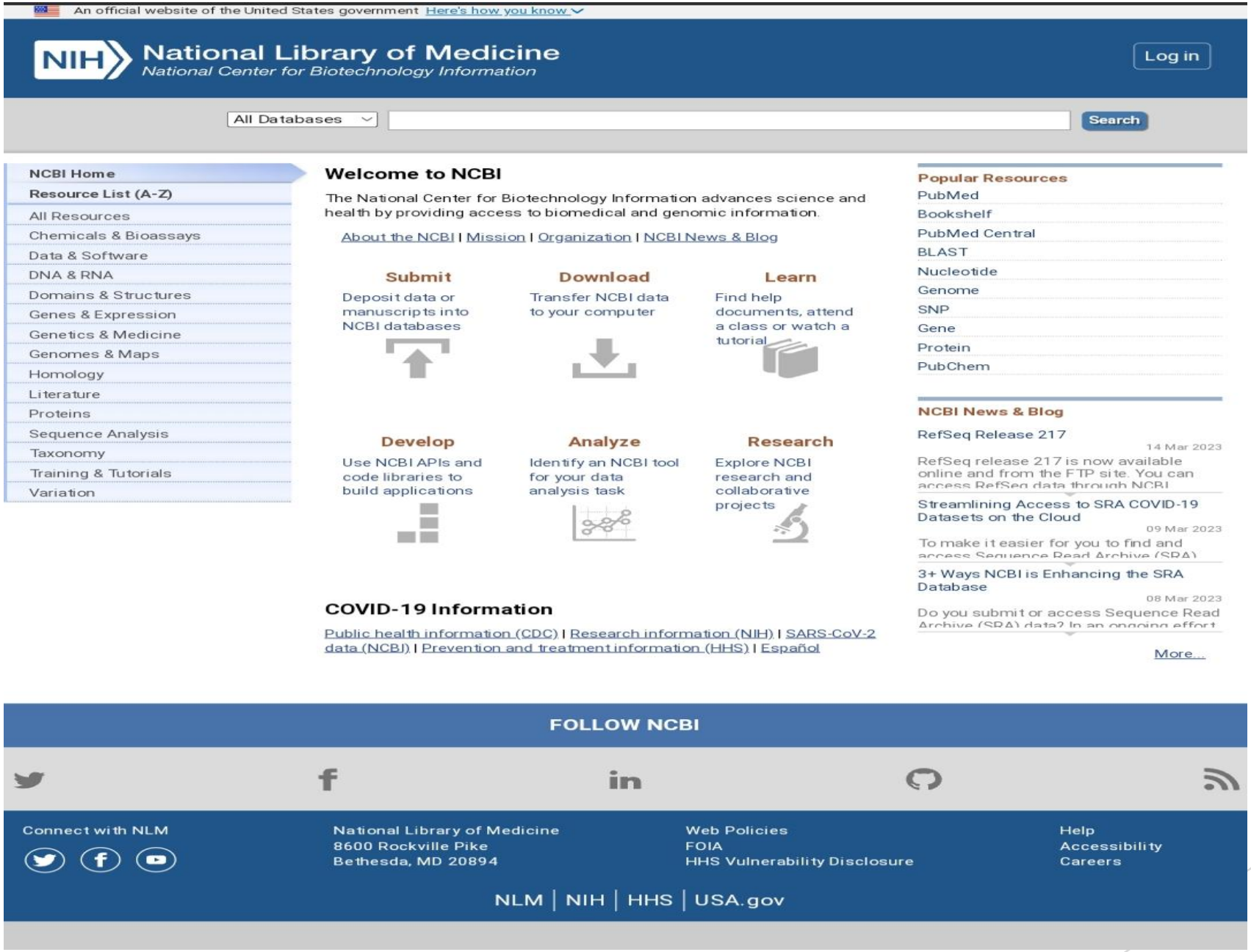
https://www.ddbj.nig.ac.jp/.

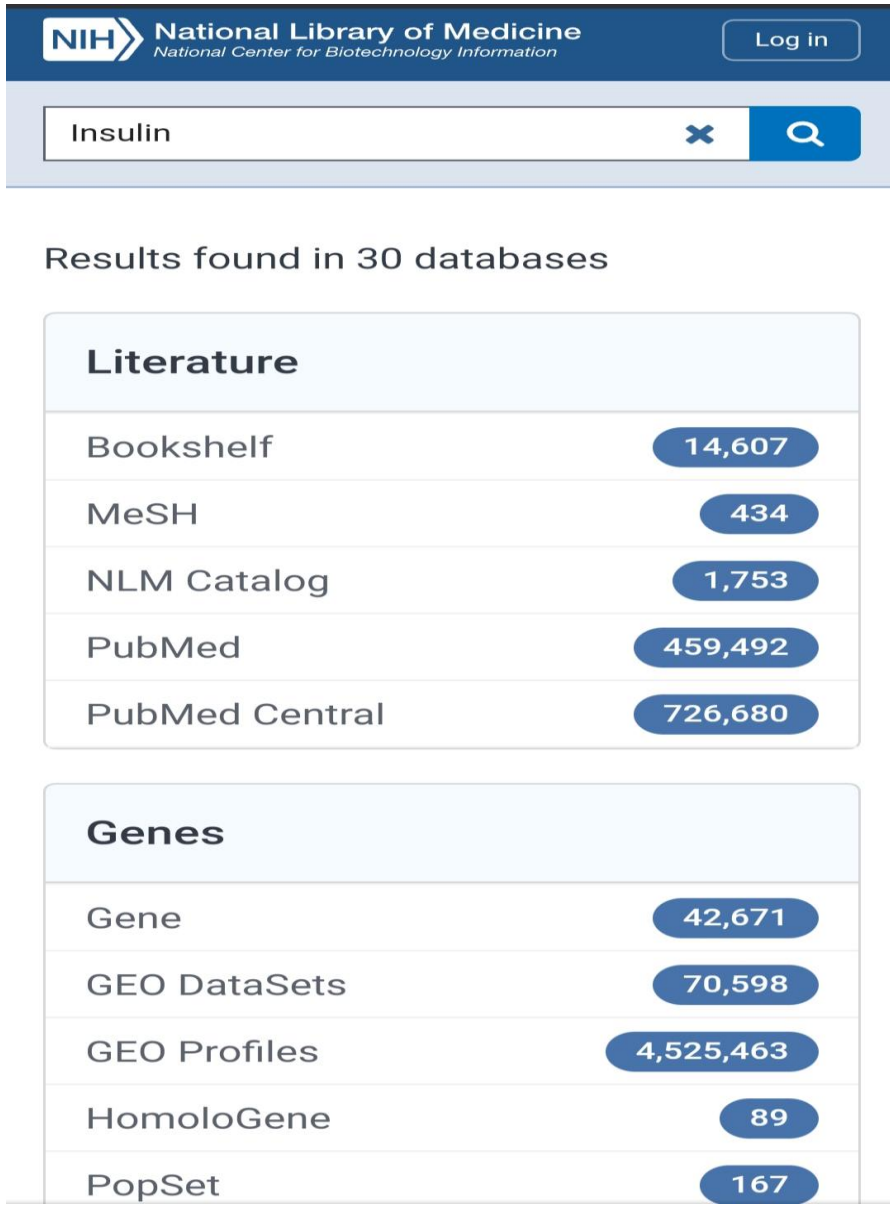
GENE BANK

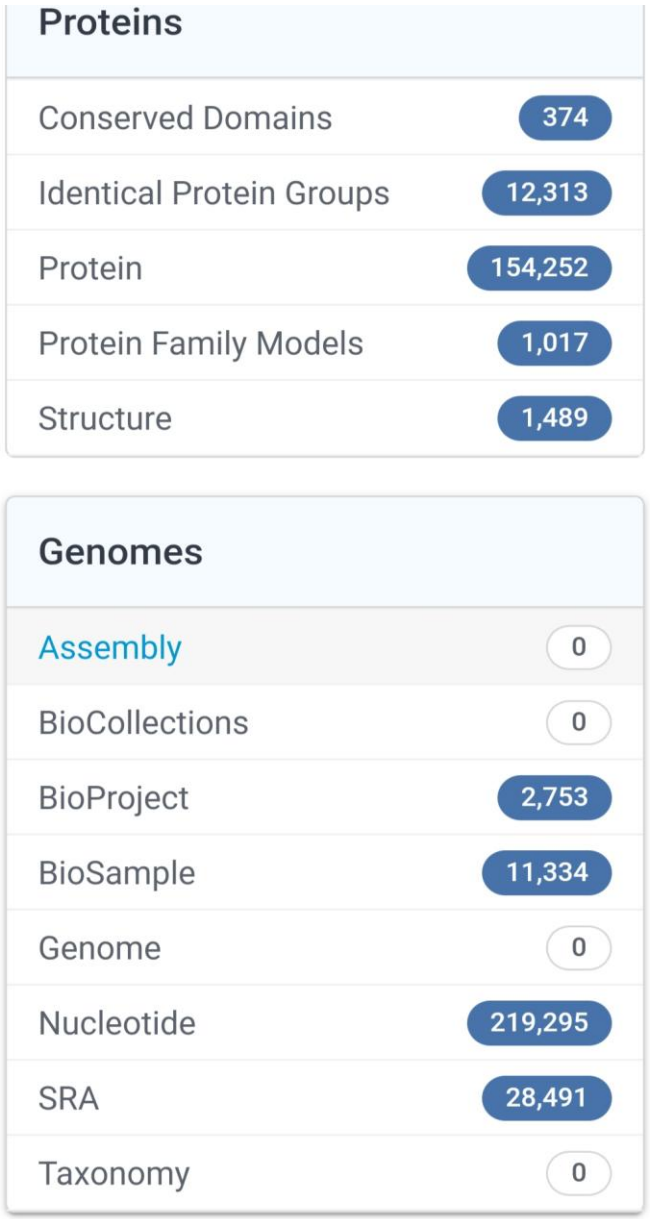
What?

The genebank database sequence is

- **An open access.**
- **Annotated collection of all.**
- **Publicly Available nucleotide sequences and their protein translations.**
- **It is produced and maintained by NCBI(National centre for biotechnology information) as the part of the international nucleotide sequences Databases collaboration (INSDC).**


An official website of the United States government Here's how you know
National Library of Medicine
National Center for Biotechnology Information
Log in
All Databases
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
COVID-19 Information
Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español
Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem
NCBI News & Blog
RefSeq Release 217
14 Mar 2023
RefSeq release 217 is now available online and from the FTP site. You can access RefSeq data through NCBI
Streamlining Access to SRA COVID-19 Datasets on the Cloud
09 Mar 2023
To make it easier for you to find and access Sequence Read Archive (SRA)
3+ Ways NCBI is Enhancing the SRA Database
08 Mar 2023
Do you submit or access Sequence Read Archive (SRA) data? In an ongoing effort
More...
FOLLOW NCBI
Connect with NLM
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
Web Policies
FOIA
HHS Vulnerability Disclosure
Help
Accessibility
Careers
NLM | NIH | HHS | USA.gov


NIH National Library of Medicine
National Center for Biotechnology Information
Log in
Insulin
Results found in 30 databases
Literature
Bookshelf 14,607
MeSH 434
NLM Catalog 1,753
PubMed 459,492
PubMed Central 726,680
Genes
Gene 42,671
GEO DataSets 70,598
GEO Profiles 4,525,463
HomoloGene 89
PopSet 167


Proteins
Conserved Domains 374
Identical Protein Groups 12,313
Protein 154,252
Protein Family Models 1,017
Structure 1,489
Genomes
Assembly 0
BioCollections 0
BioProject 2,753
BioSample 11,334
Genome 0
Nucleotide 219,295
SRA 28,491
Taxonomy 0

REFERENCES

- HTTPS://WWW.GENOME.JP/TOOLS/MOTIF/
- HTTPS://WWW.GOOGLE.COM/URL?SA=T&SOURCE=WEB&RCT=J&URL=HTTPS://PROSITE.EXPASY.ORG/&VED=2AHUKEWIK7OG4H-Z9AHWXS2WGHWPWCAWQFNOECAUQAQ&USG=AOVVAW2YSYV7B25SZFM7D4PADBFC

THANK YOU!