

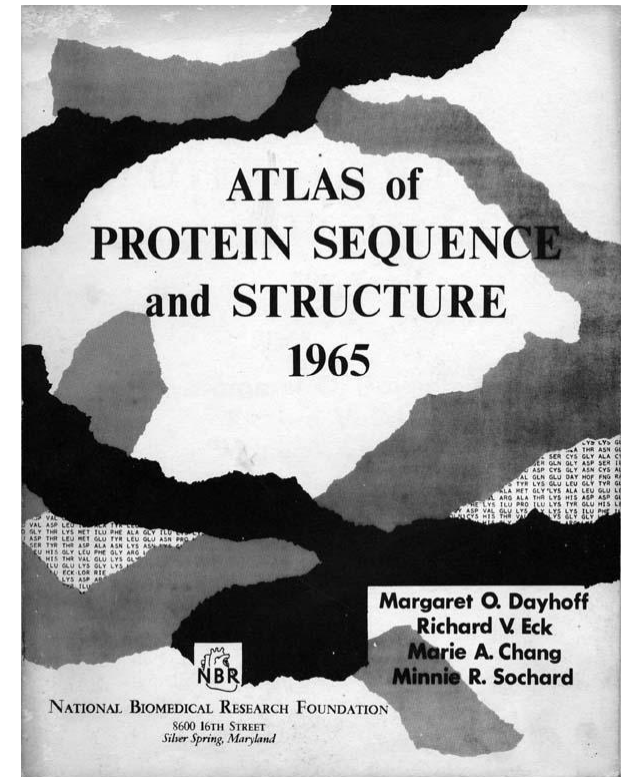
# Bioinformatics

Discipline of quantitative analysis of information relating to biological macromolecules with the aid of computers

Developed to due advances in both molecular biology and computer science over the past 30-40 years

# Milestones

- **1960's:**
  - Earliest bioinformatics efforts with Margaret Dayhoff and her development of computer aids to help determine protein sequences (1961)
  - She also developed 1<sup>st</sup> protein sequence database called 'Atlas of Protein Sequence and Structure' (1965)



- **1970's:**

- Protein Data Bank (PDB) established for archiving 3-D protein structures

- Had less than 12 structure at start; now has more

J Mol Biol. 1970 Mar;48(3):443-53.

**A general method applicable to the search for similarities in the amino acid sequence of two proteins.**

Needleman SB, Wunsch CD.

by Needleman and Wunsch

- Paved way for routine sequence comparison and database searching

- First protein structure prediction by Chou and Fasman

- Pioneered series of developments in protein structure prediction



A MEMBER OF THE PDB An Information Portal to Biological Macromolecular Structures As of Tuesday Apr 26, 2011 at 5 PM PDT there are 72717 Structures | PDB Statistics

Contact Us | Print

PDB ID or Text PDB ID lookup or Text search of the complete structure file Search Advanced Search

MyPDB Hide

Login to your Account Register a New Account

Home Hide

- News & Publications Usage/Reference Policies Deposition Policies Website FAQ Deposition FAQ Contact Us About Us Careers External Links Sitemap New Website Features

Deposition Hide

All Deposit Services Electron Microscopy X-ray | NMR

# 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.

Hide Welcome Message

Featured Molecules Hide

Structural View of Biology List View of Archive By: Title | Date | Category Health & Disease

Molecule of the Month:

Customize This Page

New Features Hide

PDB-101: Edu. Resources Structural View of Biology Latest features released: Website Release Archive:

RCSB PDB News Hide

Weekly | Quarterly | Yearly

wwPDB News Hide

Statement on Retraction of PDB Entries

- **1980's:**

- Genbank:

- Database of publicly available genetic (DNA) sequences (and their protein translations) at the NIH (National Institute of Health)

- NCBI:

- Established in 1988 as a national resource for molecular biology information
    - Division of the National Library of Medicine (NLM) at the NIH
    - Creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information
    - Aim is to improve understanding of molecular processes affecting human health and disease

- BLAST: Basic local alignment search tool

- **1990's:**

- Start of human genome project
- Development of internet increased access, exchange and dissemination of biological data

# Overview of bioinformatics advances

1962	Pauling's theory of molecular evolution
1965	<a href="#">Margaret Dayhoff's</a> Atlas of Protein Sequences
1970	<a href="#">Needleman-Wunsch</a> algorithm
1982	GenBank Release 3 made public
1982	Phage lambda genome sequenced
1988	<a href="#">National Center for Biotechnology Information</a> (NCBI) created at NIH/NLM
1990	<a href="#">BLAST</a> : fast sequence similarity searching
1995	First <a href="#">bacterial genomes</a> completely sequenced
1996	<a href="#">Yeast genome completely sequenced</a>
1998	<a href="#">Worm (multicellular) genome completely sequenced</a>
1999	Fly genome completely sequenced

Human genome sequenced: 2003

- Explosion of genomic sequence information (via genome projects) generated demand for efficient computational tools to manage and analyse data
  - Development depended on knowledge generated from *mathematics, statistics, computer science, information technology and molecular biology*
  - Merger of these disciplines created an information-orientated field in biology now known as *bioinformatics*



## Bioinformatics:

# Storage, retrieval, manipulation & analysis of biological information

- Union of biology and informatics
- **Computers** are used for storage, retrieval, manipulation and distribution of information related to biological macromolecules such as **RNA, DNA and proteins.**

# Applications of bioinformatics

- Basic genomic & molecular biology research
- Biotechnology and biomolecular science
  - Knowledge-based drug-design
    - Computational protein-ligand interaction allow rapid identification of novel leads for synthetic drugs
  - Forensic DNA analysis
  - Pharmacogenomics
    - Identification of mutations in genome could allow early diagnosis and/or determine type of treatment

# Summary

- Biological information:
  - sequences of DNA
  - sequences of proteins
  - sequences of genomes
  - references
- Because of this large amount of Biological information it is necessary to have computerized databases to:
  - Store, organise and index all the information
  - Update, retrieve, view, analyse and evaluate data
  - Pubmed is an example of a reference database (housed by NCBI) and we will use this as our first example.

# Bioc 315 course

- **Databases and Resources studied**

- NCBI (USA; <http://www.ncbi.nlm.nih.gov/>)

- **NCBI = National Centre for Biotechnology Information**
- Pubmed, Books, Nucleotide, Protein, OMIM, Structure etc
- BLAST

- Swiss Institute for Bioinformatics

- EXPASY (<http://ca.expasy.org/>)
- **EXPASY = Expert Protein Analysis System**

- \*[EMBL (European Molecular Biology Lab;  
<http://www.embl.org/>)

- EMBL-EBI (DNA/RNA), UniProt (proteins), EMSD (3D)]

\* **Not for test/exam**

## On NCBI you can find....

**Pubmed:** Life sciences articles for biomedical journals

**Bookshelf:** Collection of biomedical books

**OMIM:** Catalog of human genes and genetic disorders  
( = Online Mendelian Inheritance in Man)

**Taxonomy:** Organisms represented in the genetic database with at least one nucleotide or protein sequence

**Structure:** Macromolecular 3D structures as well as tools for their visualization and comparative analysis

**Nucleotide:** Nucleotide sequences

**Protein:** Protein sequences

**BLAST:** Basic Local Alignment Search Tool

**ENTREZ:** Integrated, text-based search and retrieval system for the major databases

# On EXPASY you can find....

## **UniProt Knowledgebase (Swiss-Prot and TrEMBL):**

Protein knowledgebase consists of completed protein sequences and translated DNA sequences

**ProtParam:** prediction of protein size, PI, half-life

**PROSITE:** protein domains, families & functional sites

## **SWISS-MODEL Repository:**

Theoretically calculated (automatically generated) protein models.

## **SWISS-2DPAGE:**

Two-dimensional polyacrylamide gel electrophoresis

**...and more**