



## Description

Biological databases emerged as a response to the huge data generated by [low-cost DNA sequencing technologies](#). One of the first databases to emerge was GenBank, which is a collection of all available protein and DNA sequences. It is maintained by the National Institutes of Health (NIH) and the National Center for Biotechnology Information (NCBI). GenBank paved the way for the Human Genome Project (HGP). The HGP allowed complete sequencing and reading of the genetic blueprint. The data stored in biological databases is organized for optimal analysis and consists of two types: raw and curated (or annotated). Biological databases are complex, heterogeneous, dynamic, and yet inconsistent. The inconsistency is due to [the lack of standards at the ontological level](#).

## Why are these Important?

Earlier, databases and databanks were considered quite different. However, over the time, database became a preferable term. [Data is submitted directly](#) to biological databases for indexing, organization, and data optimization. They help researchers find relevant biological data by making it available in a format that is readable on a computer. All biological information is readily accessible through data mining tools that save time and resources. Biological databases can be broadly classified as [sequence and structure databases](#). Structure databases are for protein structures, while sequence databases are for nucleic acid and protein sequences.

## Kinds of Biological Databases

Biological databases can be further classified as primary, secondary, and composite databases.

Primary databases contain information for sequence or structure only. Examples of primary biological databases include:

- Swiss-Prot and PIR for protein sequences
- GenBank and DDBJ for genome sequences
- Protein Databank for protein structures

Secondary databases contain information derived from primary databases. Secondary databases store information such as conserved sequences, active site residues, and signature sequences. Protein

---

Databank data is stored in secondary databases. Examples include:

- SCOP at Cambridge University
- CATH at the University College of London
- PROSITE of the Swiss Institute of Bioinformatics
- eMOTIF at Stanford

Composite databases contain a variety of primary databases, which eliminates the need to search each one separately. Each composite database has different search algorithms and data structures. The NCBI hosts these databases, where links to the Online Mendelian Inheritance in Man (OMIM) is found.

## The Future

Because of high-performance computational platforms, these databases have become important in providing the infrastructure needed for biological research, from data preparation to data extraction. The simulation of biological systems also requires computational platforms, which further underscores the need for biological databases. The future of biological databases looks bright, in part due to the digital world.

In terms of research, bioinformatics tools should be streamlined for analyzing the growing amount of data generated from genomics, metabolomics, proteomics, and metagenomics. Another future trend will be the annotation of existing data and better integration of databases.

With a large number of biological databases available, the need for integration, advancements, and improvements in bioinformatics is paramount. Bioinformatics will steadily advance when problems about nomenclature and standardization are addressed. The growth of biological databases will pave the way for further studies on proteins and nucleic acids, impacting therapeutics, biomedical, and related fields. If you use biological databases and would like to share any insights, comment in the section below!

### Category

1. Career Corner
2. PhDs & Postdocs

### Date Created

2017/08/11

### Author

daveishan