

Secondary Databases of Proteins

Protein databases have become a crucial part of modern biology

Searching databases is the first step in the study of a new protein

These databases reorganize or annotate the data and provide predictions

Comparison between proteins or protein families provide information about the relationship between proteins within a genome or across different species

The availability of multiple databases helps researchers understand the structure and function of a protein as well as to study the interactions among different proteins

Protein Data Bank

- Established at Brookhaven National laboratories in 1971 as an archive for biological macromolecular crystal structures
- The mode of access to PDB has changed over the years with the advent of modern technologies
- Initially PDB had been limited to a small group of experts involved in structural research but the inclusion of a diverse group of researchers from biology, chemistry and computer science has expanded its horizons
- The management of PDB became the responsibility of the Research Collaboratory for Structural Bioinformatics (RCSB) in October 1998
- RCSB works with the vision of creating a resource based on the most modern technology that facilitates the use and analysis of structural data thus creating an enabling resource for biological research
- PDB adopts robust technology for data deposition, data processing and data distribution and therefore manages a large influx of data

Cambridge Structural Database (CSD)

- The Cambridge structural database covers the crystal structures of organic and metal-organic compounds as the demand for the study of metal organic compounds has increased in the past few years
- The database currently contains 146272 entries corresponding to 137268 structures and the difference between the numbers reflect that there are multiple studies for certain compounds
- The CSD contains more than 1000 entries for peptides of 2-25 aminoacids which can be searched in terms of acyclic/cyclic structures, internal sequences or terminal residues, simple or substituted aminoacids, exact or wild card residues

Molecular Modelling Database (MMDB)

Useful features of the molecular modelling database:

- I. Facilitate computation on 3D structure data.
- II. Analysis of individual structures and relationships among the proteins
- III. Interactive views of sequence-structure relationships
- IV. Establishes a relationship between 3D structure records and associated literature, molecular and chemical data

SCOP: Structural Classification of Protein database

- SCOP provides a detailed and comprehensive description of the relationships of known protein structures
- The classification is based on hierarchical levels: the first two levels, family and superfamily describe near and distant evolutionary relationships and the third fold describes geometrical relationships. The different folds have been grouped into classes.
- The distinction between evolutionary relationships and those that arise from the physics and chemistry of proteins is a feature that is unique to this database

contd.....

SCOP not only includes the proteins in the current version of the PDB but also those proteins for which there are published descriptions but whose coordinates are not available

The unit of classification is usually the protein domain but smaller proteins and proteins having medium size with one domain are treated as a whole but domains in larger proteins are usually classified individually

CATH

CATH provides hierarchical classification of protein domains based on their folding patterns

The classification of protein domains is based not only on sequence information but also on the structural and functional properties of proteins

The domains are classified in a hierarchical scheme with four main levels called class (C), architecture (A), topology (T) and homologous superfamily (H), hence the name CATH

CATH offers an important tool to researchers as proteins with very little sequence similarity are often both structurally and functionally related