DALI and SCOP: Structure databases

Course- SOS In Microbiology

Semester- M.Sc 2nd Semester

Paper- 204 (Bioinformatics)

Unit -3

DALI

- Comparative analyses of protein sequences and structures play a fundamental role in understanding proteins and their functions.
- The Dali server (http://ekhidna2.biocenter.helsinki.fi/ dali) is a network service for comparing protein structures in 3D.
- The Dali program optimizes a structural alignment, that is, a sequential set of one-to-one correspondences between C-alpha atoms.
- The Dali server performs three types of structure comparisons: PDB search, pairwise comparison and all against all structure comparison.

- The all against all structure comparison is a new option. The user inputs a set of N structures and the server computes the N × N matrix of pairwise similarities (Dali Z-scores).

User interface

Inputs- The input to the server is one or more protein structures in PDB format.

- The query structure can be specified as a PDB identifier plus chain identifier, or a PDB file uploaded by the user. All backbone atoms (N, CA, C, O) are required and the minimum chain length is 30 amino acids.

- If only the amino sequence is known, it can be mapped to the closest known structure by a sequence similarity search against PDB using e.g. the SANS parallel server.
- Outputs- PDB search, pairwise comparison and all against-all comparison produce summaries of structural neighbours in a common output format and share interactive analysis tools.
- The all-against-all comparison additionally generates a dendrogram and correspondence analysis plot of the similarity matrix.

- The summaries consist of-

(i) a list of structural neighbours, ranked by Z-score,

(ii) The alignment data. The results are presented as plain text for downloading by downstream applications, and as hypertext for interactive analysis

SCOP

- A Structural Classification of Proteins Database for the Investigation of Sequencesand Structures.
- **Website:** http://scop.mrc lmb.cam.ac.uk/scop/
- Manual classification of protein structural domains based on similarities of their structures and amino acid sequences.
- Created in 1994.
- Maintained by Alexei G. Murzin and his colleagues at the Laboratory of Molecular Biology in Cambridge, England.

- It provides a detailed and comprehensive description of the relationships of known Protein structures.
- Classification is on hierarchical levels.
- It is freely accessible.
- Current version of SCOP is 2.03(October 2013)- http://scop.berkeley.edu/

- Classification

Family (identical structure and function)

- All proteins that have residue identities of 30%and greater or Pairwise sequence similarity >25%.
- The proteins that have lower sequence identities but whose functions and structures are very similar.
- Clear evolutionary relationship.

Superfamily (common structure and function)

- Probable common ancestry
- Families whose proteins have low sequence identities but whose structures and, in many cases, functional features suggest that a common evolutionary origin is probable.

Fold (common core structure)

- Major structural similarity
- SSE's in similar arrangement
- Superfamilies and families are defined as having a common fold if their proteins have the same major secondary structures in the same arrangement and with the same topological connections.

Class

- Similar secondary structure content
- Folds have been grouped into five structural classes:

1. all- α, those whose structure is essentially formed by α -helices;
2. all- β, those whose structure is essentially formed by β -sheets;

3. α/β, those with α -helices and β -strands;
4. α+β, those in which α helices and β -strands are largely segregated;

SCOP IN BIOINFORMATICS

SCOP classification is used:

1. As reference set of data to develop automatic classification methods used in analyzing families, superfamilies, and folds
2. For integrative structural data mining to develop predictive methods and structure-comparison tools

- Understanding evolution of protein enzymatic functions, evolutionary change of protein folds, hierarchical structural evolution.
- To study distantly related proteins with the same fold
- To study sequence and structure variability and its dependence in homologous proteins.

- To derive amino acid similarity matrices and substitution tables useful for sequence comparison and fold recognition studies.
- To study the structural anatomy of folds and domains, to extract structural principles for use in protein design experiments
- SCOP domains have been used to study combinations of different domains and their decomposition in multi domain proteins
- Recent structural genome projects have been using SCOP extensively in identifying new targets