Bioinformatics Methods

Iosif Vaisman

Email: ivaisman@gmu.edu

Secondary Structure: Computational Problems

Secondary structure characterization
Secondary structure assignment
Secondary structure prediction

Protein structure classification

Structural classes of proteins

- Myoglobin: all α
- Rubredoxin: all β
- Lactalbumin: α/β
- Hemoglobin: all α
- Prealbumin: all β
- Pyruvate Kinase: all α
- Trypsin: all α
- Coagulation Factor: all α
- Lysozyme: all α

Protein Structure Classification

SCOP - Structural Classification of Proteins
FSSP - Fold classification based on Structure-Structure alignment of Proteins
CATH - Class, architecture, topology and homologous superfamily

SCOP: Structural Classification of Proteins

Current release: 1.75
38221 PDB Entries (June 2009), 110800 Domains.

http://scop.mrc-lmb.cam.ac.uk/scop/

The SCOP database aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known. Proteins are classified to reflect both structural and evolutionary relatedness. Many levels exist in the hierarchy; the principal levels are family, superfamily and fold.

**Family**: Clear evolutionarily relationship
**Superfamily**: Probable common evolutionary origin
**Fold**: Major structural similarity

SCOP: Structural Classification of Proteins

**Family**: Clear evolutionarily relationship

Proteins clustered together into families are clearly evolutionarily related. Generally, this means that pairwise residue identities between the proteins are 30% and greater. However, in some cases similar functions and structures provide definitive evidence of common descent in the absence of high sequence identity; for example, many globins form a family though some members have sequence identities of only 15%.
SCOP: Structural Classification of Proteins

**Superfamily:** Probable common evolutionary origin

Proteins that have low sequence identities, but whose structural and functional features suggest that a common evolutionary origin is probable are placed together in superfamilies. For example, actin, the ATPase domain of the heat shock protein, and hexokinase together form a superfamily.

SCOP: Structural Classification of Proteins

**Fold:** Major structural similarity

Proteins are defined as having a common fold if they have the same major secondary structures in the same arrangement and with the same topological connections. Different proteins with the same fold often have peripheral elements of secondary structure and turn regions that differ in size and conformation. In some cases, these differing peripheral regions may comprise half the structure. Proteins placed together in the same fold category may not have a common evolutionary origin: the structural similarities could arise just from the physics and chemistry of proteins favoring certain packing arrangements and chain topologies.

### SCOP Statistics

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of folds</th>
<th>Number of superfamilies</th>
<th>Number of families</th>
</tr>
</thead>
<tbody>
<tr>
<td>All alpha proteins</td>
<td>284</td>
<td>507</td>
<td>871</td>
</tr>
<tr>
<td>All beta proteins</td>
<td>174</td>
<td>354</td>
<td>742</td>
</tr>
<tr>
<td>Alpha/beta proteins (a/b)</td>
<td>147</td>
<td>244</td>
<td>803</td>
</tr>
<tr>
<td>Alpha+beta proteins (a+b)</td>
<td>376</td>
<td>552</td>
<td>1055</td>
</tr>
<tr>
<td>Multi-domain proteins</td>
<td>66</td>
<td>66</td>
<td>89</td>
</tr>
<tr>
<td>Membrane and cell surface proteins</td>
<td>58</td>
<td>110</td>
<td>123</td>
</tr>
<tr>
<td>Small proteins</td>
<td>90</td>
<td>129</td>
<td>219</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1195</strong></td>
<td><strong>1962</strong></td>
<td><strong>3902</strong></td>
</tr>
</tbody>
</table>