Bioinformatics Methods

Iosif Vaisman
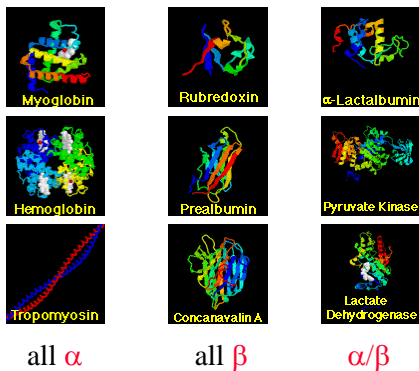
Email: ivaisman@gmu.edu

Secondary Structure: Computational Problems

Secondary structure characterization
Secondary structure assignment
Secondary structure prediction
Protein structure classification

Structural classes of proteins

Myoglobin   Rubredoxin   α-Lactalbumin
Hemoglobin   Prealbumin   Pyruvate Kinase
Tropomyosin   Concanavalin A   Lactate Dehydrogenase

all α   all β   α/β

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 absense 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 hexakinase 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

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