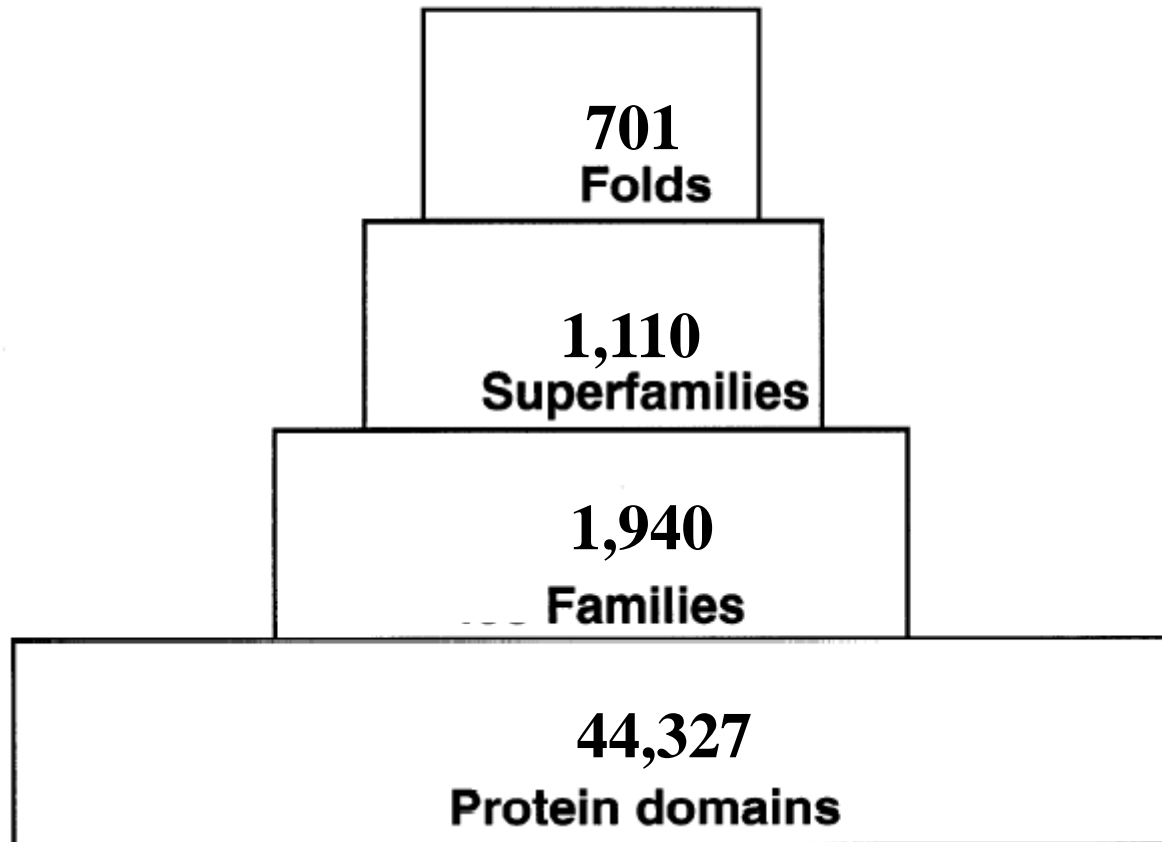


Tutorial No. 8

Protein Domains

Problem 1: *Evolution of structures.* Offer two competing hypotheses that explain why two superfamilies may share the same fold.

SCOP – a structural classification of proteins



A fold is a topology of the folded protein backbone.

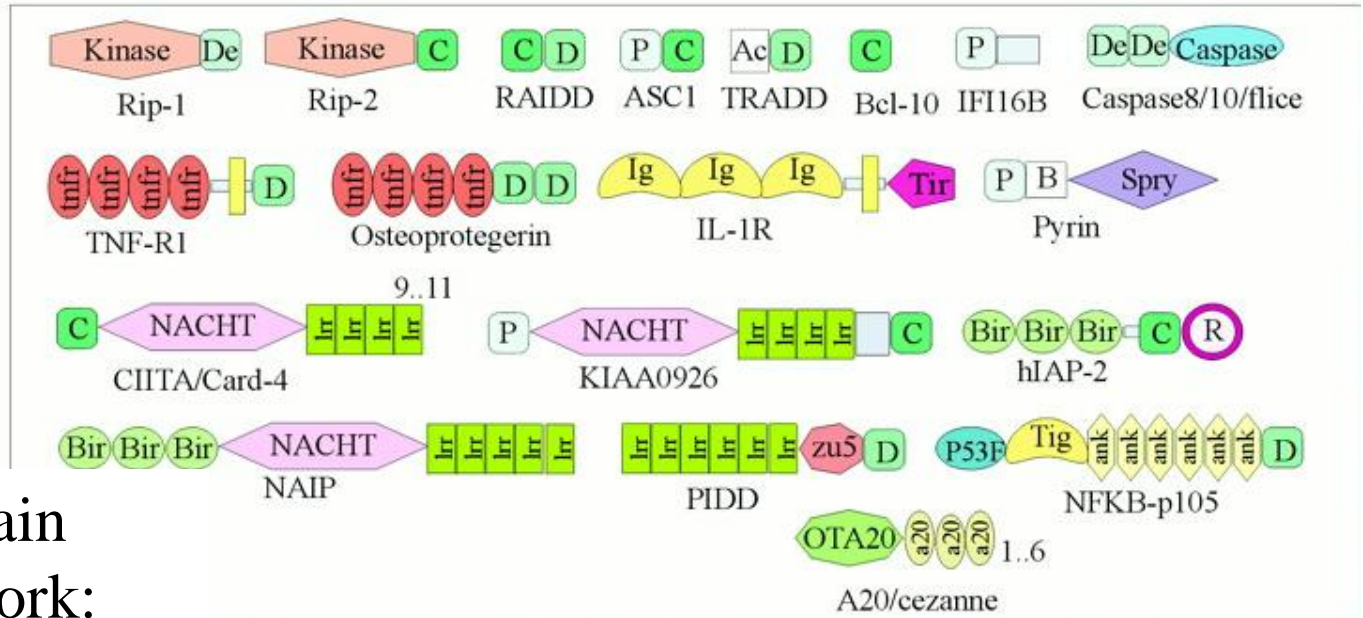
Unknown whether superfamilies of the same fold are monophyletic

Problem 2: *Domain architecture network.* Identify the network for the proteins domains S_TKc, SH3, PDZ, PH, and GED. Each domain is a node and each edge represents an instance(s) where the two domains are found in the same protein (without regard to the organism in which it is present). Use the SMART database (<http://smart.embl-heidelberg.de/>). Go to “Genomic mode”. Enter your queries on the right side of the site, labeled ‘Architecture Analysis’ (for example enter query: ‘SH3 AND S_TKc’ and then click “Architecture query”. To display the domain architecture graphically then click ‘ALL (52)’). The network should be submitted as a network figure as shown in the lecture. For each edge note one example protein and the organism from which it comes (for example “[AGAP001683-PA](#) - *Anopheles gambiae*”).

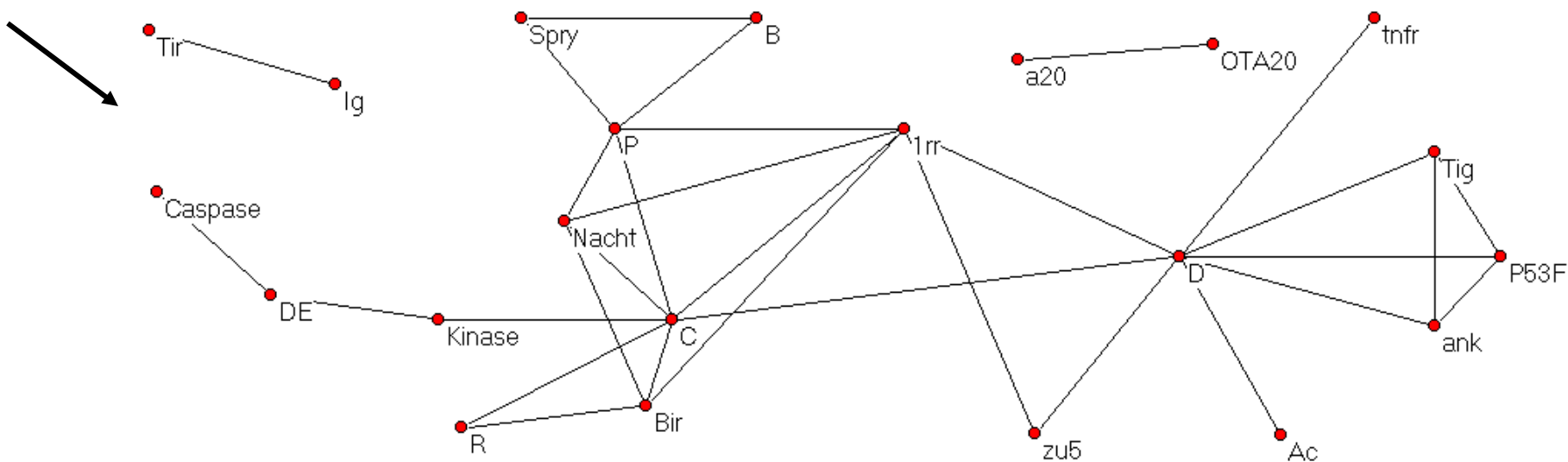
Domain Architecture Networks:

A pair of domains are linked if there is at least one instance of a protein containing both of the domains.

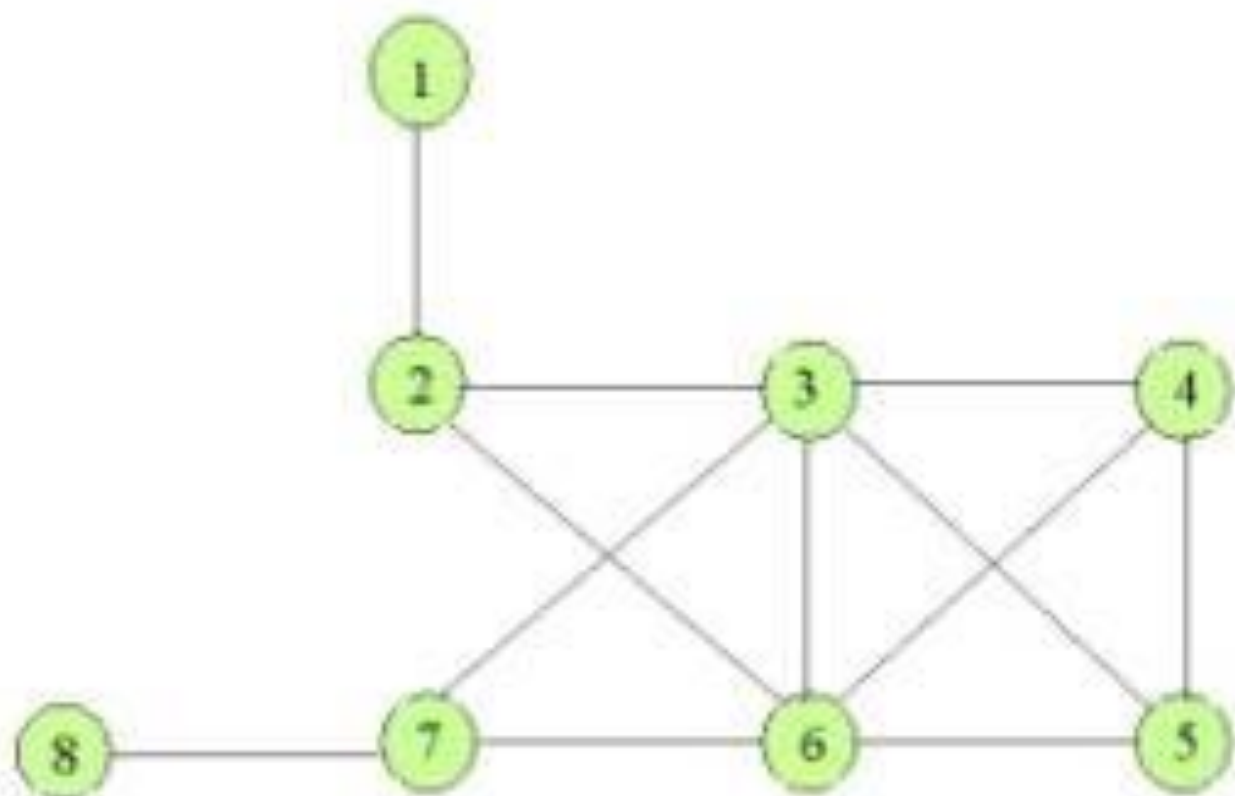
Apoptosis proteins:



The derived domain architecture network:



Problem 3: *Protein similarity networks.* In the network below the nodes correspond to protein sequences and edges correspond to sequence similarity between a pair of sequences. Assume that if a pair of proteins have a domain in common it can be detected by sequence similarity. Given that there are 5 domains altogether among the sequences (domain = recurrent sequence), what is the distribution of domains among the proteins? To answer this question, for each domain (domain A, domain B, domain C, domain D, and domain E) state in which proteins (1-8) it occurs.



Problem 4: Domain Age. Get to know the [Interpro database](#).

Using the phylogenetic distribution associated with each Interpro [domain](#) (in the *Taxonomy* section), find an example of each of the following distributions:

- A universal domain (present in all 3 kingdoms, not necessarily viral)
- A prokaryotic specific domain
- A eukaryotic specific
- A virus specific
- A metazoan specific

Problem 5: *Domain Rearrangements.* I have prepared a [list](#) of human-fly-worm orthologs with links to the InterPro database. Each line represents a set of human, fly and worm orthologous genes. Describe **two** instances where domain architecture has evolved in any of the three lineages for a given orthologous group. For example: Human gene [P24864](#) and Drosophila gene [O01501](#) have cyclin N and cyclin C domains while C. elegans [Q16595](#) has only the cyclin N domain.