Secondary Structures

• Does everyone know what the backbone and residue/side chains are?
• Clear about 1°, 2° 3° structures?

Heteropolymer

Mostly in regular secondary structure

• Secondary structure can be defined by phi and psi angles or by hydrogen bond patterns.
• A connected sequence of amino acids must have the right phi-psi angles to make the secondary structure.
What's the pattern?
Ci>Ni+?

Each side can have different properties
All of the amino acids are on the outside
1f3c 31-50

move around not quite 120°

The boxes show amino acids
Motifs

- Scop: Mruzin et al JMB 1995 (Cyrus Chothia)
- Cath: Orengo et al Structure 1997 (Janet Thornton)
- Each starts with domains

Classifications are descriptive; Structural diversity comes from evolution - we use classification to deduce evolutionary relationships.

- Homolog - Phylogenetically related - derived from a common ancestor gene. (PFAM-genes; Scope/Cath-structures).
  - Ortholog - retain the same function
  - Paralog - function has diverged
- Family: related proteins with similar function
- Superfamily: related proteins with different functions

Proteins are made of domains.
- A domain is a structural and an evolutionary unit.
- They have 50-200 residues.
- Compact folded unit, quasi-independent structurally and functionally
- Domains that are families or superfamilies come from a common ancestor.
  - similar sequence - family
  - diverged sequence but similar fold and function - superfamily

Chothia and Gough (Biochem J (2009) 419, 15-28)
Multidomain proteins have more domains than their prokaryote counterparts, but the remainder comes from families of known structure. The distribution of the number of combinations made by the different families is again high for both prokaryotes and eukaryotes. For example, the β-barrel-like architecture (A→C) (A→D) is not found in bacteria, but is found in a large number of different bacterial families. The β-barrel-like architecture (A→D) has an outer β-barrel and an inner β-barrel and an outer layer of α-helices (Figure 2). Alternatively, the three-layer β-sandwich architecture (A→B) consists of a central β sheet which is covered by a layer of α-helices on both sides of the sheet (Figure 2).

**Figure 2.** An example of a superdomain. The P-loop-containing NTP hydrolase domain and the translation Protein domain (3) occur in prokaryotic and eukaryotic translation factors that hydrolyze guanosine triphosphate (GTP). GTP hydrolysis in the P-loop domain drives the conformational changes in the Translation Proteins domain, which is then transmitted onto the ribosome. The superdomains occur in 70 different domain architectures, and 6 of these are shown here. The inset at left shows a protein of known structure, which contains the superdomain. IF, initiation factor; EF, elongation factor; RF, release factor; RNA, transfer RNA.
CATH version 3.3 (class, architecture, topology, homology) contains 128 688 domains, 2386 homologous superfamilies and 1233 fold groups

Alexis J. Cuff, Ian Sillitoe, Tony Lewis, Andrew B Clegg, Robert Rentzsch, Nicholas Furnham, Malalasekera Pellegrini-Calleja, David Jones, Janet Thornton, and Christine A Orengo

**Scop Classification Statistics**

*SCOP: Structural Classification of Proteins.*

<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>567</td>
<td>871</td>
</tr>
<tr>
<td>All beta proteins</td>
<td>178</td>
<td>344</td>
<td>442</td>
</tr>
<tr>
<td>All alpha and beta proteins (α+β)</td>
<td>147</td>
<td>264</td>
<td>403</td>
</tr>
<tr>
<td>All alpha and beta proteins (α/β)</td>
<td>978</td>
<td>902</td>
<td>1203</td>
</tr>
<tr>
<td>Multi-domain proteins</td>
<td>68</td>
<td>60</td>
<td>89</td>
</tr>
<tr>
<td>All topologies and all superfamilies</td>
<td>899</td>
<td>2052</td>
<td>3942</td>
</tr>
<tr>
<td>TIM barrel</td>
<td>119</td>
<td>192</td>
<td>202</td>
</tr>
</tbody>
</table>

Nucleic Acids Research 2001

Orengo et al: Structure 1997 5:1093

**Figure S2.**

Log Plot showing the growth of CATH from the beginning of the PDB to 2007 indicating the increase in the number of domains, architectures, topologies, superfamilies and CATH classes in the database.

Orengo et al: Structure 1997 5:1093
The recurrence of common motifs within many of the available complex architectures possessing more diffuse loci to each other. This mechanism would not be available, layer-based architectures can further accommodate because, in addition to the ability to expand by adding more profound disturbances to the architecture. This is more useful to subclassify these regions of fold space prediction, from sequence methodsd. It would therefore show that only 7dOp of these domains occur also as interestingly an analysis of the distribution of domain superfolds and major architectures gives rise to an overlap hydrophobic core.
Atpase

http://scop.mrc-lmb.cam.ac.uk/scop/search.cgi?ver=1.75&key=1bmf