

# PART 1 - The Pfam Database

When studying a protein, it is important to look for any already known domains in that protein. It is impossible to keep track of all the new domains in the literature. However, there are several databases that do this for you. For example: Pfam, SMART and InterPro offer comprehensive sets of domains, while databases such as *MEROPS* provides detailed information about a specific set of domains. Below are a set of exercises that are designed to give you an introduction to the sort of research/tasks that the Pfam and *MEROPS* database allows you to perform.

## 1.1 Introduction to the Pfam Database

Make a bookmark to

<http://www.sanger.ac.uk/Software/Pfam/>

There are three main ways to access Pfam.

1. **Browse by families** - Good if you know what domains your protein contains.
2. **Search by protein name or sequence** - Can give protein id or accession, this is fast as it uses precalculated results for known proteins. We can also give a novel sequence you know nothing about and find domains by searching against the Pfam HMM library. This will take a minute or two..
3. **Search by keyword** - This is good if you want to find domains involved in apoptosis generally.

We can look for the sequence MLTD\_ECOLI using the search page. You can get to this page by moving the mouse cursor over the 'Search by' menu button and then clicking on the Protein name or sequence option. Alternative, paste this url into your browser.

<http://www.sanger.ac.uk/Software/Pfam/search.shtml>

Now type in the sequence identifier in the appropriate text field and submit the search.

We are shown a schematic representation of the MLTD\_ECOLI protein. It contains an SLT domain and two other domains. If you put your cursor over the domain you will see the description of the domain. The protein is a transglycosylase, so it is no surprise to find an SLT domain.

We can also search sequences with the Pfam profile-HMMs but we don't want to flood the server and it is much slower than using the identifier.

We can look at Pfam annotation for the SLT domain by clicking on the domain picture. Note that the Pfam entry contains very little description but it does contain annotation from the InterPro resource. There are also useful links to the literature and other databases. Next to the annotation there is a picture of a structure containing the SLT domain. All Pfam domains that occur in the structure are coloured.

We can get alignments for each Pfam family.

**Click "Get alignment" button**

This gives you an alignment that is coloured according to the clustal colouring system. The alignment is also marked-up with secondary structure and active sites.

**Question: What residue is the important active site residue in this domain?**

**Pfam also provides many different alignment formats that can be selected.**

**Go back to the family annotation page. The SLT HMM use to detect the 480 sequences can be visually inspected using the HMM logo button.**

**Click "View HMM logo"**

**The height of the letters are proportional to the probability.**

**Question: What is the most likely residue at each of the following positions in a typical SLT domain: 24, 51, 114, 1115 ?**

**Now look at the structure link at CATH-PDBSUM by clicking on the structure image on the family page.**

<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=1531>

**We can get information such as secondary structure for a member of this family. See what other information is available on this page! Then return to SLT domain page in Pfam.**

**Question: Can you find out what organisms the SLT domain is present in?**

**Look at the section on species distribution.**

**Look at the domain organisation of all SLT domains.**

**Select radiobox for representative architectures(19) then click on "View graphic" button**

**You will see all the different domain architectures that the SLT domain is found in. There are also smaller 3-coloured boxes. These are Pfam-B domains.**

**Question: Have a look at some of the Pfam-B families. Do they have the same functionality as the normal Pfam-A families?**

---

**End of section 1. [Next section](#)**