

## LAB 2

Below are 4 exercises to become more familiar with some of the tools we discussed in class. Work at your own pace, but spending about 15 minutes on each of those exercises is probably a good idea. At the end of each exercise, some suggested tools are listed (small fonts).

1. Explore the proteins with PDB identifiers 1BD0, 1PTM, 2SBL, and 2FCR. In particular, verify the unusual helices at the locations described below:
  - (a) Residues 40-44 on the A chain of 1BD0,
  - (b) Residues 211-216 on the A chain of 1PTM,
  - (c) Residues 496-504 of 2SBL,
  - (d) Residues 40-45 of 2FCR.

Obtain the sequence of amino acids in one-letter format. Which secondary structure elements does DSSP delineate at those positions?

PDB, DSSPcont, DeepView.

2. Compare the proteins 1SBP and 1POT, considering sequence, structure, and evolution (i.e., try to assess the sequence similarity, structural similarity, and see if you can find out something about common origins and function).

BLAST, PDB, CATH, SCOP, DALI, VAST.

3. Consider the following three amino acid sequences (they are also on the course web page, to copy and paste it into a field on a web server):

```
VTMYKLVIKGNNVKGETTTKALDAQTAEQVFKEYRNDKGVWTFDDSDQYQVME
STSYNYIFGARNIKGHTSMACVGADTGFNIFIGVENGNNKWMHPHAYVVYQVVL
VRHCQGVFSNIELKKTQPSPAPIIEWEWLGYNTQRQHGIEREFDPHAMWMLKV
```

They were derived from a single domain with known structure, by making mutations in the amino acid sequence more and more likely (using a BLOSUM substitution matrix). For each of those sequences, try to find the original domain using sequence alignment tools. Find proteins similar in structure to the domain you identified. Looking at those, how good are the secondary structure predictions on the above sequences?

BLAST, PDB, CATH, SCOP, DALI, PSIPRED, JPRED.

4. Consider the sequence of this hypothetical protein (also available on the course webpage):

```
RTLGGQLVDAMQFICGDRGFYFNRSTDYGMSTGILDHCQFRGCDMRNLTMYPPLAGSSGSSGSGEKPFQCECGKRFTQN
SHLHSHQRVHTGKPSGSSGLTPAITTYKLVINGKSLRGQTTTKAVDSDTSTKFCFCQYCNDRGVDGVWYDDATSTYTSTE
```

It is made out of mutated sequences from three separate domains. Try to infer the domain boundaries by

- (a) eye-balling the sequence,
- (b) carrying out some sequence alignments,
- (c) parsing it into domains using some domain-related webservers.

BLAST, PFAM, PRODOM, CDD, PDB.