

## CHEM 2300H PROTEIN BIOINFORMATICS ASSIGNMENT, Instructions and Guide

Bioinformatics refers to the databases of information on biomolecules and the tools used to analyze this information. These databases are often linked, so that one can access the others easily from many different points of entry. Tools are likewise numerous, as can be seen by a visit to the ExpASy site <http://www.expasy.org/>

### Materials

- A device with internet access.
- The Material in the Bioinformatics Seminars Folder on Blackboard
- **The UniProtKB Accession number for your query protein. This will be provided to you by email to your Trent email address by Friday March 20.**

### Advice

- Much information can be found by keeping your eyes open on the UniProtKB, ExpASy and PDB web sites you visit.
- The BLAST search and alignment may take several minutes. If it stalls or gives you an error, try later.
- When in a jam, ask me for help, but please don't wait until the last day to do so.

**Procedure and Write-up.** You will be emailed the UniProtKB accession number of your query protein by March 20. Use UniProtKB, ExpASy, and the Protein Data Bank to provide the following information on the target:

#### 1) Protein parameters (6 marks). Refer to the seminar on UniProt and ExpASy

Provide the key parameters of your protein (the 'query'):

- The UniProt Accession Number of your query. This is the one that was emailed to you
- Its name and species
- Its function
- Its molecular weight
- Its isoelectric point (pI)
- Its amino acid sequence (**Use Courier font, 8-point - this is the best font for displaying sequences**)

#### 2) BLAST and ALIGN (5 marks). Refer to the seminar on BLAST and MSAs

Perform a BLAST search on your query sequence.

- Use the **structural database** to do your search (Target Database = "with 3D Structure (PDB)")
- From the BLAST search results, take the query sequence, and the next three sequences and align them.
- Download the alignment, and copy it into Word. **Use Courier font, 8-point.**
- Comment on the degree of sequence similarity (or lack of it) among the four proteins in the alignment.

**3) 3D Structure (5 marks). Refer to the seminars on viewing structures on the PDB and searching the PDB**

- From the UniProtKB page with your query, scroll down to Structure
- From the 3D Structure Databases, select the link destination “**RCSB PDB**”, then select the first of the PDB files. This will take you to the PDB Web site and the structural file you have selected.
- State the name of this PDB file.
- Use the 3D View tool within the browser to prepare an image of your protein with the following features:
  - a) The protein chain is shown in **cartoon** style
  - b) The protein chain is coloured by **secondary structure**
  - c) Save a screenshot of your image as a .png file or similar. Paste the downloaded file into your report.

Prepare your report as a single pdf. Make sure it is viewable before you submit it. A sample report of what we expect is shown on the next page.

***Upload your report to the Bioinformatics Dropbox on the CHEM 2300H site by Wednesday April 8 at 11:59pm***

(The blue text contains tips & instructions. With the exceptions noted below for the sequence and the alignment, font size and font type is up to you. Save your document as a pdf)

### 1. Protein Parameters

Accession Code (AC): **P00167**

Name and species: Cytochrome *b<sub>5</sub>*, *Homo sapiens* (Remember to italicize species names)

Function: Cytochrome b5 is an electron carrier for several membrane-bound oxygenases.

Molecular Weight: 15,330 Daltons

Isoelectric Point (pI): 4.86

Sequence: (Use Courier Font, 8-point for the sequence; it is a monospaced font that is ideal for presenting sequences and sequence alignments)

```

      10      20      30      40      50
MAEQSDEAVK YTTLEEIQKH NHSKSTWLIL HHKVYDLTKF LEEHPGGEEV
      60      70      80      90     100
LREQAGGDAT ENFEDVGHST DAREMSKTFI IGELHPDDRP KLNKPPETLI
      110     120     130
TTIDSSSSWW TNWVIPAISA VAVALMYRLY MAED

```

**2. BLAST and ALIGN** (This is obtained by clicking on Text Output in the Align Results page. After copying into your document, use Courier Font, 8-point for the alignment. You may have to adjust the margins to get a proper alignment. Depending on how long your sequences are you may have to use additional pages)

```

sp|P00167|CYB5_HUMAN      MAEQSDEAVKYYTTLEEIQKHNSKSTWLI LHHKVYDLTKFLEEHPGGEEVLREQAGGDAT 60
sp|P00167-3|CYB5_HUMAN   MAEQSDEAVKYYTTLEEIQKHNSKSTWLI LHHKVYDLTKFLEEHPGGEEVLREQAGGDAT 60
sp|P00169|CYB5_RABIT     MAAQSDKDKVYYTTLEEI KKHNSKSTWLI LHHKVYDLTKFLEEHPGGEEVLREQAGGDAT 60
sp|P00173|CYB5_RAT       MAEQSDKDKVYYTTLEEI QKHKDSKSTWVI LHHKVYDLTKFLEEHPGGEEVLREQAGGDAT 60
**  ***:  *****: ** : .*****: *****: *****: *****: *****: *****

sp|P00167|CYB5_HUMAN      ENFEDVGHSTDAREMSKTFI IGELHPDDRPKLNKPPETLITIDSSSSWWTNWVIPAISA 120
sp|P00167-3|CYB5_HUMAN   ENFEDVGHSTDAREMSKTFI IGELHP-----ETLITIDSSSSWWTNWVIPAISA 110
sp|P00169|CYB5_RABIT     ENFEDVGHSTDARELSKTFI IGELHPDDRSKLSKPMETLITTVDSNSSWWTNWVIPAISA 120
sp|P00173|CYB5_RAT       ENFEDVGHSTDARELSKTYI IGELHPDDRSKI AKPSETLITTVESNSSWWTNWVIPAISA 120
*****: *****: *****: *****: *****: *****: *****: *****: *****: *****

sp|P00167|CYB5_HUMAN      VAVALMYRLYMAED 134
sp|P00167-3|CYB5_HUMAN   VAVALMYRLYMAED 124
sp|P00169|CYB5_RABIT     LIVALMYRLYMADD 134
sp|P00173|CYB5_RAT       LVVALMYRLYMAED 134
: *****: *

```

These three sequences are 88 to 100% identical, respectively with the query sequence (This information is obtained by clicking on Percent Identity Matrix in the Align Results page).

### 3. 3D Structure: 2i96.pdb

