

Bioinformatics: Life Sciences on Your Computer

Week 3 Practice Problems - Part 2

These practice problems cover predictions of molecular weight, isoelectric point, localization and secondary structure.

1. Use this mitochondrial protein from yeast.

```
>yeast mitochondrial protein
MVLLHKSTHIFPTDFASVSRAFFNRYPNPYSPHVLSDITISRNVDQEGNLRTRLLKKSGKLPWVKPFL
RGITETWIIIEVSVVNPNANSTMKTYTRNLDHTGIMKVEEYTTYQFDSATSSTIADSRVKFSSGFNMGIKSK
VEDWSRTKFDENVKKS RMGMAFVIQKLEEARNPQF
```

Try the following localization programs. List the organelle prediction (if any), and probability (if any).

- TargetP (<http://www.cbs.dtu.dk/services/TargetP/>)
- MitoProt II (<http://ihg.gsf.de/ihg/mitoprot.html>) –Raw format only--do not paste in top descriptive line of sequence
- Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>)
- CELLO (<http://cello.life.nctu.edu.tw/>)

2. Take the sequence below and enter into the PHD server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html) in RAW format. Count the number of helices and sheets and find the largest stretch for each structure type.

```
>SARS coronavirus protein E
MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPTVYVYSRVKNLNSSEG
VPDLLV
```

- What percent of the protein is alpha helix, extended strand and random coil?
- Run HNN (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html) and compare and contrast the results.

3. Go to a site with a Chou-Fasman prediction (http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=misc1). Switch the pull-down program window to Chou-Fasman and use the protein below.

```
>Seed allergenic protein RAG1
MASNKVVFVSVLLLVLSVLAAAMATMADHHQVYSPGEQCRPGISYPTYSLPQCRTLVRQQCVGRGAASAA
DEQVWQDCCRQLAAVDDGWCRCGALDHMLSGIYRELGATEAGHPMAEVFPGCRRGDLERAAASLPAFCNV
DIPNGPGGVVCYWLGYPRTPRTGH
```

- How many alpha-helices (arrows)?
- How many beta-sheets (Es)?
- What do the Ts represent?
- What indication suggests that the prediction may not be correct in some places?

4. Enter the following sequence into the Compute pI/MW tool. Use RAW format.

```
> MBP [Homo sapiens]
MASQKRPSQRHGSKYLATASTMDHARHGFLPRHRDTGILDSIGRFFGGDRGAPKRSGKDSHHPARTAHY
GSLPQKSHGRQTQDENPVVHFFKNIVTPRTPPPSQGKGAEGQRPFGYGGGRASDYKSAHKGFKGVDAQGTL
SKIFKLGGDRSRSGSTMARR
```

- a. What is the molecular weight?
- b. What is the isoelectric point?
- c. At neutral pH (7.0), would you expect this protein to have a net negative charge or net positive charge? Briefly explain.

Answers

- 1.a. mitochondrial (.500 probability)
 - b. 0.1926
 - c. possibly mitochondrial (0.20), elsewhere 0.79
 - d. Mitochondrial - total score (which is addition of five probabilities) = 2.005
-
- 2.a. 43.42% alpha helix; 31.58 extended strand; 25.0% random coil
 - b. With PHD, long alpha helix goes from 15-44 (and possibly 15-47). With HNN, the long alpha helix goes from 10-44.
-
- 3.a. five alpha helices
 - b. six beta strands
 - c. turns
 - d. Alpha helix and beta sheet is predicted in the same location. Both cannot be correct.
- Note--the dots that appear above some of the letters are there simply to mark every 10th amino acid.***
-
- 4. a. 17347.34
 - b. 11.14
 - c. Positive charge. Because the pI is greater than 7, this is a basic protein meaning it has a net positive charge at neutral pH.

Note: MBP = Myelin Basic Protein