

Bioinformatics: Life Sciences on Your Computer

Week 3 Practice Problems - Part 1

The first problem uses Prosite. The second uses Interpro. See the PDF supplemental reading on motifs/Prosite for more on Interpro. You can use both to solve problem 3.

1. Use prosite (<http://prosite.expasy.org/>) to find motifs in gi|[347834361](#). Unclick the checkbox that would exclude patterns that match frequently. Scan Prosite.
 - a. How many distinct patterns are above the line that reads “hits by patterns with a high probability of occurrence or by user-defined patterns?”
 - b. Looking at the pattern that is above that line, what types of molecules do proteins with this pattern transport?
 - c. Click on the prosite number (starts with PS). Scroll down for the consensus pattern. At what position in the pattern is an “any amino acid except threonine or aspartic acid?”
2. Go to Interpro (<http://www.ebi.ac.uk/interpro/>) and enter the sequence with gi|[14548146](#). Use the default applications.
 - a. How many results refer to “transcription factor 2b” (TFIIB)?
 - b. How many results for cyclin or cyclin-like?
 - c. Click the transcription factor link by the result marked “PR....” (not the one starting with IPR). IIB associates with what other two transcription factors bound to DNA?
 - d. Back at the InterPro result, click SSF47954. What are listed for SCOP class and SCOP fold?
 - e. Based on the SCOP class, what secondary structure would you expect to be heavily found in this protein?
 - f. Back at InterPro, click the SM... link. What three proteins contain this helical domain?
3. Analyze the following sequence:

```
>hypothetical protein [Comamonadaceae bacterium CR]
MSLKERIQEEMKAAMRAKDTARLGAIRLLLLAAIKQKEVDERVMLDDAAIIAVVDKLIKQRRDSVTAYQQA
QRSDLADKEAAEITVLEAYLPQRWSRAEVEEAVSRVVAETAATGPGDMGKVMAAVKAQCQ GKADMAVVSQ
VVKAAALAAARGG
```

The protein sequence is from a bacterial genome. The genome was released on October 21, 2013. Which programs did you use and what did you find out? What is the protein and what does that protein do? You can use BLAST but only to help verify your results from searching for domains and motifs.

Answers

1. a. Just one pattern above the line.
b. sugars
c. [LIVMSTAG]-[LIVMFSAG]-{SH}-{RDE}-[LIVMSA]-[DE]-**{TD}**-[LIVMFYWA]-G-R-[RK]-x(4,6)-[GSTA]
7th position (see bold above)
2. a. 3
b. 3 (but within a single grouping)

c. TFIID and TFIIA (those are transcription factor 2D and transcription factor 2A -- the II is a Roman numeral notation)

d. SCOP class: All alpha proteins; SCOP fold: cyclin-like

e. alpha helix

f. cyclins, TFIIB, retinoblastoma

3. There is not much from Prosite, other than possible phosphorylation sites. From Interpro databases, it looks to be a subunit of the enzyme aspartyl/glutamyl-tRNA amidotransferase. That is an enzyme involved in linking amino acids to tRNA molecules so that translation can occur. Check some of the database documentation for more info on this enzyme and discuss in the forums.