

Problem Set 4

(Due Sept 23rd)

- As we discussed in class, there are several computational algorithms that allow biochemists to predict secondary structure. Let's see how good these algorithms hold up to real structural information.
 - Access the following two pdb files: 1DNK and 7ACN.
 - What are these proteins and what organism do they come from?
 - Exactly what position do they occupy in the genome (I showed you how to do this in class). Make sure to identify the chromosome and position.
 - Use the Chou-Fasman prediction program on ExPASy to predict the secondary structure. Compare this prediction with the actual secondary structure observed in the pdb file. You can do this visually using Chimera or by observing the text in the pbd file – either way, map the structure vs. prediction and determine how good the prediction is.
 - Based on the observed secondary structure, predict what the CD spectrum of each protein might look like. Please justify your sketch.
- Using the tools in ExPASy, determine the pI, MW, and molar absorptivity of each protein.
- Assuming that you start with a homogenous mixture of both of these proteins (and nothing else), predict what a 2D gel electrophoresis experiment would look like.
- For 1DNK, predict the MW of all peptides produced when it is treated with Cyanogen Bromide (CNBr). You are encouraged to use ExPASy to determine MW values, but you should manually determine where the peptide chain will be broken.
- Use the PeptideCutter tool in ExPASy to predict what fragments would be produced when 7ACN is digested with each of the following proteases. In this tool, make sure to select “only the following selection of enzymes and chemical” option. Also, it is quite helpful to choose the “Table of sites” display option.
 - Chymotrypsin
 - Arg-C
 - Thrombin
- A tandem MS experiment results in peaks at the following m/z ratios. Determine the sequence of this peptide.
128.2 185.2 313.3 460.5 623.7 736.9 851 980.1
A peak is also observed at m/z = 425.5. What is the source of this peak?
- Using the attached electropherogram:
 - Please describe how this data is generated.
 - Determine the sequence of nucleotides 50-100 (feel free to simply highlight the sequence on the image).
 - Discuss why there are regions that are not useful and highlight those regions.
- What is meant by pyrosequencing? What reactions are important in this process?

9. A team of researchers identify a human protein through 2D electrophoresis that they are interested in identifying. This protein is independently digested with Arg-C and Asp-N and sequenced by mass spectrometry. The resulting sequences are shown below. Identify this protein. Note that you may not be able to piece together the entire sequence, but you will be able to get enough to unambiguously determine what the protein is.

Asp-N	Arg-C
DSG	EIVR
DLT	LDLAGR
DIRK	AVFPSIVGR
DSYVG	GYSFVTTAER
DLAGR	DLTDYLMKILTER
DETTALVC	VAPEEHPTLLTEAPLNPKANR
DEAGPSIVHR	KDLYANNVMSGGTTMYPGIADR
DIKEKLCYVAL	HQGVMVGMGQKDSYVGDEAQSKR
DNGSGLVKAGFAG	MQKEITALAPSTMKIKIIPPER
DLYANNVMSGGTTMYPGIA	DEDETTALVCDNGSGLVKAGFAGDDAPR
DGVTHNVPIYEGYALPHAIMRL	TTGIVLDSGDGVTHNVPIYEGYALPHAIMR
DFENEMATAASSSSLEKSYELP	EKMTQIMFETFNPAMYVAIQAVLSLYASGR
DEAQSKRGILTLKYPIEHGIITNW	GILTLKYPIEHGIITNWDDMEKIWHHTFYNELR
DYLMKILTERGYSFVTTAEREIVR	CPETLFQPSFIGMESAGIHETTYNSIMKCDIDIR
DAPRAVFPSIVGRPRHQGVMVGMGQK	KYSVWIGGSILASLSTFQQMWITKQEYDEAGPSIVHR
DGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKC	DIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNER
DRMQKEITALAPSTMKIKIIPPERKYSVWIGGSILASLSTFQQM WITKQEY	
DMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREKMTQIM FETFNPAMYVAIQAV	
LSLYASGRTTGIVL	

10. With your protein:

- Determine the mRNA sequence that codes for your protein. You should approach this as you did in 1.a.i.
- Determine an endonuclease that will not cut your gene. I recommend using the NEBcutter tool (<http://tools.neb.com/NEBcutter2/index.php>)
- Design a set of primers that could be used to amplify your protein by PCR. These primers should be designed to include the endonuclease site (determined above) at the 5' and 3' ends.
- Predict the peptides that would result if you treat your protein with trypsin. Use the PeptideCutter tool in ExPASy to make this easier.

You'll be happy to know that we only have one more bioinformatics tool to learn about. The problem set next week will explore this.

