Exercise Worksheet   Name:_____
Instruction on how to perform are on the back.

1) **The number of the unknown protein to use is circled. Attach a picture of the generated hydropathy plot for the unknown protein. (note: plots will look weird if window size was not set to 19)**

                    #1   **#2**   **#3**

2) **Use BLAST to determine the identity of the unknown protein and record the name of the first protein from the BLAST generated list (note: result will be weird if "9606" is not chosen)**

3) **Based on what we learned about in class, where is this protein located?**

4) **Based on what we learned about in class, describe the structure of this protein, particularly if it is an integral or peripheral membrane protein?**

5) **Using the information from questions #3 and 4, explain why the generated hydropathy plot supports the identity of the protein obtained from the BLAST search.**

Cell Biology (BIOL380) Exercise

Tasks

1) Use the amino acid sequence of the unknown protein assigned to you to generate a hydropathy plot for the protein using a hydropathy plot generating program

2) Use a BLAST search to determine the name of the unknown protein **and** research the structure of the protein to assess whether the hydropathy plot determination makes sense for the structure of the determined protein

Materials

This exercise is conducted *in silico* so the only material needed is a computer that has internet access.

Procedures

Generating Hydropathy Plots

* For ease (basically so you do not need to type in all of the letters of the protein sequence), the digital amino acid sequences for unknown proteins has been posted to Blackboard as a word document.
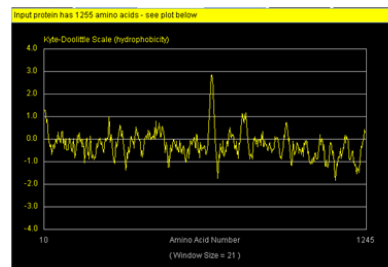
1) Open the word document with the unknown digital amino acid sequences. Choose the amino acid sequence assigned to you (either #1, #2 or #3)

2) Open the hydropathy plot generating program: http://web.expasy.org/cgi-bin/protscale/protscale.pl

3) Copy the digital amino acid sequence for chosen unknown protein. Make sure to get the entire sequence from beginning to end.

    EXAMPLE:
    MADVEKGKKIFIMKCSLCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKK
    YIPGTKMIFVGIKKKEERADLIAYLKKAKKNE

4) Paste the sequence into the submission box in the hydropathy plot generating program

5) Scroll to the bottom of the program's homepage and set the window size to 19

6) Hit submit.

7) Wait. A hydropathy plot will be generated.

8) Print out the hydropathy plot and attach it to the worksheet. Circle the number of the unknown protein used on the worksheet.

    Hydropathy plots (see example to right) have quite a bit of 'noise'. To determine the true transmembrane domains from background scatter, draw an imaginary and arbitrary line at 1.0 on the y-axis, where substantial regions above 1.0 would be considered 'true' hydrophobicity. The generated Kyle-Doolittle hydropathy plot for the example unknown protein indicates that this protein is an integral membrane protein because it has a region of ~20 amino acids that reached into the +ΔG region indicating a transmembrane domain

BLAST protein search

1) Copy the same unknown digital amino acid sequence again from the word document

2) Open the BLAST-protein program:
http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

3) Copy the digital amino acid sequence for the chosen unknown protein. Make sure to get the entire sequence from beginning to end.

4) Paste the sequence into the submission box in the BLAST program

5) Scroll to the middle of the program's homepage to the "choose search set" area and type in the organism (optional) box the following: 9606. This is the taxonomic identification (taxid) number for *homo sapiens*.

6) Scroll to the bottom of the page and Hit the blue "BLAST" button

7) Wait while a list of putative proteins is generated. This may take a few minutes and the screen is automatically updated.

8) Record the description of the top putative protein on the worksheet.

9) Using your textbook, notes from the lecture and/or internet resources, research the unknown protein, now identified protein. Record on the worksheet the protein's location (e.g. is it in the cytosol or associated with a membrane) and one structural feature (in particular its association with a membrane – is it peripheral or is it integral, how many transmembrane domains if integral?).

10) Analyze the hydropathy plot, in particular noting how many, if any, transmembrane domains. On your worksheet, write one way in which the hydropathy plot supports the BLAST results of the unknown protein.

Hydropathy Plots, an *In Silico* exercise

Unknown Protein #1 digital amino acid sequence

MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDT
LMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE

Unknown Protein #2 digital amino acid sequence

MIASQFLSALTLVLLIKESGAWSYNTSTEAMTYDEASAYCQQRYTHLVAIQNKEEIEYLNSILSYSP
SYYWIGIRKVNNVWVWVGTQKPLTEEAKNWAPGEPNNRQKDEDCVEIYIKREKDVGMWNDER
CSKKKLALCYTAACTNTSCSGHGECVETINNYTCKCDPGFSGLKCEQIVNCTALESPEHGSLVC
SHPLGNFSYNSSCSISCDRGYLPSSMETMQCMSSGEWSAPIPACNVVECDAVTNPANGFVECF
QNPGSFPWNTTCTFDCEEGFELMGAQSLQCTSSGNWDNEKPTCKAVTCRAVRQPQNGSVRC
SHSPAGEFTFKSSCNFTCEEGFMLQGPAQVECTTQGQWTQQIPVCEAFQCTALSNPERGYMN
CLPSASGSFRYGSSCEFSCEQGFVLKGSKRLQCGPTGEWDNEKPTCEAVRCDAVHQPPKGLV
RCAHSPIGEFTYKSSCAFSCEEGFELHGSTQLECTSQGQWTEEVPSCQVVKCSSLAVPGKINMS
CSGEPVFGTVCKFACPEGWTLNGSAARTCGATGHWSGLLPTCEAPTESNIPLVAGLSAAGLSLL
TLAPFLLWLRKCLRKAKKFVPASSCQSLESDGSYQKPSYIL

Unknown Protein #3 digital amino acid sequence

MRARPRPRPLWATVLALGALAGVGVGGPNICTTRGVSSCQQCLAVSPMCAWCSDEALPLGSP
RCDLKENLLKDNCAPESIEFPVSEARVLEDRPLSDKGSGDSSQVTQVSPQRIALRLRPDDSKNFS
IQVRQVEDYPVDIYYLMDLSYSMKDDLWSIQNLGTKLATQMRKLTSNLRIGFGAFVDKPVSPYMY
ISPPEALENPCYDMKTTCLPMFGYKHVLTLTDQVTRFNEEVKKQSVSRNRDAPEGGFDAIMQAT
VCDEKIGWRNDASHLLVFTTDAKTHIALDGRLAGIVQPNDGQCHVGSDNHYSASTTMDYPSLGL
MTEKLSQKNINLIFAVTENVVNLYQNYSELIPGTTVGVLSMDSSNVLQLIVDAYGKIRSKVELEVRD
LPEELSLSFNATCLNNEVIPGLKSCMGLKIGDTVSFSIEAKVRGCPQEKEKSFTIKPVGFKDSLIVQ
VTFDCDCACQAQAEPNSHRCNNGNGTFECGVCRCGPGWLGSQCECSEEDYRPSQQDECSPR
EGQPVCSQRGECLCGQCVCHSSDFGKITGKYCECDDFSCVRYKGEMCSGHGQCSCGDCLCD
SDWTGYYCNCTTRTDTCMSSNGLLCSGRGKCECGSCVCIQPGSYGDTCEKCPTCPDACTFKK
ECVECKKFDRGALHDENTCNRYCRDEIESVKELKDTGKDAVNCTYKNEDDCVVRFQYYEDSSG
KSILYVVEEPECPKGPDILVVLLSVMGAILLIGLAALLIWKLLITIHDRKEFAKFEEERARAKWDTAN
NPLYKEATSTFTNITYRGT