

Address Code Exercise Worksheet

Name: _____

Instruction on how to perform are on the back.

- 1) The number of the unknown protein to use is circled.
Attach the generated NLS mapper result for the unknown protein.

#1

#2

#3

#4

- 2) Write the highest scoring predicted monopartite NLS amino acid sequence. If no result was listed, write 'none'.

Predicted Monopartite NLS		
Pos	Sequence	Score

- 3) 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)

Unknown Protein	
-----------------	--

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

- 5) Explain how the generated NLS mapper supports the identity of the protein obtained from the BLAST search.

Cell Biology (BIOL380) Exercise

Tasks

- 1) Use the amino acid sequence of an unknown protein to map nuclear localization signals within its primary structure
- 2) Use a BLAST search to determine the name of the unknown protein **and** use knowledge of the protein to assess whether the nuclear location makes sense for the location of the determined protein

Materials

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

Procedures

Predicting nuclear localization signals in a query sequence

* For ease (basically so you don't need to type in all of the letters of the protein sequence), this document has been posted to Blackboard so the digital amino acid sequence for the unknown protein can be copied and pasted.

- 1) Open the word document with the unknown digital amino acid sequences. Highlight the amino acid sequence assigned to you (either #1, #2, #3 or #4).
- 2) Copy the digital amino acid sequence for unknown proteins. Make sure to get the entire sequence from beginning to end.
EXAMPLE:
MTAKMETTFYDDALNASFLPSESGPYGYSNPKILKQSMTLNLADPVGSLKPHLRAKNSDLLTSPDVGLLKLASPELERLIQSSNGHITTTPTPTQ
FLCPKNVTDEQEGFAEGFVRLAELHSQNTLPSVTSAAQPVN
- 3) Open the nuclear localization signal mapper program: http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi
- 4) Paste the sequence into the submission box in the NLS Mapper program.
- 5) **Scroll to the middle of the program's homepage and set the "select a cutoff score" to 7**
- 6) Scroll to the bottom of the program's homepage and click "predict NLS".
- 7) Wait. The NLS mapper result will be generated.
- 8) Print out the NLS mapper results page and attach it to the worksheet.
- 9) Write the position, amino acid sequence and score of the highest scorer on the worksheet.

BLAST protein search

- 1) Open the BLAST-protein program:
http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
- 2) Copy the same digital amino acid sequence for unknown proteins. Make sure to get the entire sequence from beginning to end.
- 3) Paste the sequence into the submission box in the BLAST program
- 4) Scroll to the middle of the program's homepage to the "choose search set" area and type in the organism (optional) box the following: Homo sapiens (taxid:9606)
- 5) Scroll to the bottom of the page and Hit the blue "BLAST" button
- 6) Wait while a list of putative proteins is generated. This may take a few minutes and the screen is automatically updated.
- 7) Record the description (title) of the top putative protein on the worksheet.
- 8) 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. nucleus, cytosol).
- 9) On your worksheet, explain whether the NLS locator supports the BLAST results of the unknown protein.

Locating Address Codes, an *In Silico* exercise

Unknown Protein #1 digital amino acid sequence

MTAKMETTFYDDALNASFLPSESGPYGYSNPKILKQSMTLNLADPVGSLKPHLRKNSDLLTSPDVGLLKL
ASPELERLIQSSNGHITTTPTPTQFLCPKNVTDEQEGFAEGFVRLAELHSQNTLPSVTSAAQPVNGAGM
VAPAVASVAGGSGSGGFSASLHSEPPVYANLSNFNPGALSSGGGAPSYGAAGLAFPAQPQQQQPPHH
LPQQMPVQHPRLQALKEEPQTVPEMPGETPPLSPIDMESQERIKAEKRMRNRRIAASKCRKRKLERIARL
EEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF

Unknown Protein #2 digital amino acid sequence

MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGKHVPRAVFVDL
EPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLDRIKRLADQCTGLQGFLVFHSG
GGTSGSFTSLLMERLSVDYGGKSKLEFSIYPAPQVSTAVVEPYNSILTTHTTLEHSDCAFMDNEAIYDICR
RNLDIERPTYTNLNLISQIVSSITASLRFDGALNVDLTFQTNLVPYPRIHFLATYAPVISAEEKAYHEQLSV
AEITNACFEPANQMVKCDPRHGKYMACECLLYRGDVVPKDVNAAIATIKTKRSIQFVDWCPTGFKVGINYQP
PTVVPGGDLAKVQRAVCMLSNTTAAIEAWARLDHKFDLMYAKRAVHWYVGEEMEEGEFSEAREDMAA
LEKDYEEVGVDSVEGEGEEEGEEY

Unknown Protein #3 digital amino acid sequence

METPSQRRATRSGAQASSTPLSPTRITRLQEKEDLQELNDRLAVYIDRVRSLETENAGLRLRITESEEVVS
REVSGIKAAYEAELGDARKTLDSVAKERARLQLELSKVREEFKELKARNTKKEGDLIAAQARLKDLEALLN
SKEAALSTALSEKRTLEGELHDLRGQVAKLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNIYS
EELRETKRRHETRLVEIDNGKQREFESRLADALQELRAQHEDQVEQYKKELEKTYSAKLDNARQSAERNS
NLVGAAHEELQQSRIRIDSLSAQLSQLQKQLAAKEAKLRDLEDSLARERDTSRLLAEKEREMAEMRARM
QQQLDEYQELLDIKLALDMEIHAYRKLLEGEEERLRLSPSPTSQRSRGRASSHSSQTQGGGSVTKKRLE
STESRSSFSQHARTSGRVAVEEVDEEGKFVRLRNKSNEDQSMGNWQIKRQNGDDPLLTYRFPPKFTLKA
GQVVTIWAAGAGATHSPPTDLVWKAQNTWGCNSLRTALINSTGEEVAMRKLVRVSVTVVEDEDEDGD
DLLHHHHVSGSRR

Unknown Protein #4 digital amino acid sequence

MVLADLGRKITSALRSLSNATIINEEVLNAMLKEVCTALLEADVNIKLVKQLRENVKSAIDLEEMASGLNKRK
MIQHAVFKELVKLVDPGVKAWTPTKKGKQNVIMFVGLQGSGKTTTCSKLAYYYQRKGWKTCLICADTFRAG
AFDQLKQNAKARIPFYGSYTEM DPVIIASEGVEKFKNENFEIIIVDTSGRHKQEDSLFEEMLQVANAIQPDN
IVYVMDASIGQACEAQAKAFKDKVDVASVIVTKLDGHAKGGGALSAAVATKSPIIFIGTGEHIDDFEPFKTQP
FISKLLGMGDIEGLIDKVNELKDDNEALIEKLGKQFTLRDMYEQFQNMKMGPFQSILGMIPGFGTDFMS
KGNEQESMARLKKLMTIMDSMNDQELDSTDGAKVFSKQPGRIQRVARGSGVSTRDVQELLTQYTKFAQM
VKKMGGIKGLFKGGDMSKNVSQSQMAKLNQQMAKMMDPRVLHMGGMAGLQSMRQFQQGAAGNM
KGMMGFNNM