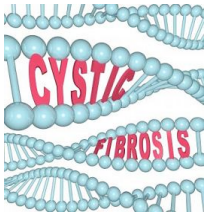
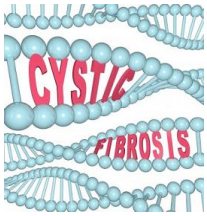
Hands-On Eighteen



Cystic Fibrosis



Transmembrane Conductance Regulator

A) The CFTR Gene and Protein

We are going to retrieve the human CFTR gene from NCBI:

- Go to NCBI: http://www.ncbi.nlm.nih.gov/
- Click on "Gene" under "Popular Resources" (right-hand side of page)
- Get the CFTR human gene by typing in "cystic fibrosis" in the search window and clicking on the blue "Search"
- From the result page, choose the human CFTR (with Gene ID: 1080) by clicking on "CFTR"

1) When was the record last updated? _____.

Upon reading "Summary", we learn that the protein encoded by this gene (the CFTR protein) is a member of the ATP-binding cassette (ABC) transporter superfamily.

2) What do ABC proteins transport?

_____.

3) What does the CFTR protein function as?

_____.

4) What diseases are associated with mutations of the CFTR gene?

_____.

We are going to retrieve the CFTR protein from NCBI:

- Go to NCBI: http://www.ncbi.nlm.nih.gov/
- Click on "Protein" under "Popular Resources" (right-hand side of page)
- Get the CFTR protein by typing in its accession number "NP_000483"
- Click on the blue "Search"

5) When was the record last updated? _____.

6) What is the size of the CFTR protein? _____.

Many proteins have conserved domains. Let us view the conserved domains of the human CFTR protein.

- Click on "Identify Conserved Domains" under "Analyze this sequence" on the right hand-side of the NP_000483 record.

©2016 Sami Khuri

7) Why do you think the CFTR protein has a region in the middle that is not similar to other ABC transporters? What do you suppose is the function of this unique region?

- Hover (place the mouse without clicking) over the "ABC_membrane" domain to learn more about it. Note there are two of these domains in the protein.

8) Are they membrane-spanning proteins? _____. Explain.
_____.

9) Record in the following table, the starting and ending locations of the two ABC_membrane domains.

ABC Membrane	Strating Position	Ending Position
First		
Second		

- Hover over the "ABCC_CFTR1" and "ABCC_CFTR2" domains to learn more about them.

10) Are they membrane-spanning proteins? _____. Explain.
_____.

11) Record in the following table, the starting and ending locations of the two conserved domains.

ABCC CFTR	Strating Position	Ending Position
ABCC_CFTR1		
ABCC_CFTR2		

- Go back to the page that contains the NP_000483 record.

To find sequences that are similar to the human CFTR protein, one could blast NP_000483 against the protein sequences of GenBank. Alternatively, we could look at precomputed blast results.

- Scroll down to "Related information" on the right-hand side of the page, and hover over "BLink".

12) What does BLink represent?
_____.

- Click on "BLink" to go to the "Pre-computed BLAST results" page.

13) How many proteins are reported to be similar (with a score of at least a 100) to the human CFTR protein (NP_000483)? _____.

14) Are all the reported hits from the same species? _____. Explain.
_____.

©2016 Sami Khuri

B) Transmembrane Protein Structure Predictors

We are going to compare several transmembrane protein predictors by running them with one of the conserved domains of the human CFTR protein.

- Obtain the human CFTR protein sequence in Fasta format: "Human_CFTR_protein_wildtype.txt". Alternatively, we can get it from GenBank at NCBI. Its accession number is NP_000483 (and P13569).

Since this protein is big, we are going to extract from "Human_CFTR_protein_wildtype.txt" a region that includes the first ABC_Membrane conserved domain (recall Question 9 on page 2).

- Go to the Sequence Manipulation Suite at: http://www.bioinformatics.org/sms2/
- Click on "Range Extractor Protein" under "Format Conversion" on the left-hand side of the page.
- Paste the "human_CFTR_protein_wildtype.txt" sequence in the window (replacing the "sample sequence")
- Enter "50..400" in the extractor window and click on "Submit".
- Save the resulting sequence under "human_CFTR_protein_extract.txt".

I) TMHMM

Open a web browser and go to the TMHMM server at: www.cbs.dtu.dk/services/TMHMM/

- Paste the sequence (human_CFTR_protein_extract.txt) in the appropriate window, keep the default options, and press "Submit".
- From the new page, click on the "HELP" link at the top of the page to go to "TMHMM2.0 User's guide".
- Go over the guide, and then go back to the output page. Understand the results and complete the following table:

Starting Position	Ending Position	Location of Segment

15) Are you surprised by the result or did you expect it? Explain.

©2016 Sami Khuri

II) Open a web browser and go to the MINNOU server at:
http://minnou.cchmc.org/

- Paste the "human_CFTR_protein_extract.txt" in the appropriate window and edit it to remove the FASTA information line. In other words, remove the line that starts with the ">" symbol.
- Click on "Submit" and you will get the results in a new page. It might take a while though. Alternatively, you can leave your email address so as to be notified when the job is done.

16) Understand the results and complete the following table:

Starting Position	Ending Position	Location of Segment

III) We are going to use three more transmembrane protein predictors. All 3 packages can be accessed from http://expasy.org/tools/.

For the next 3 questions:

- Go to http://expasy.org/tools/ and scroll down to "Topology prediction".

a) Choose TMpred [TMpred - Prediction of transmembrane regions and protein orientation (EMBnet-CH)]

- In the new page, paste the "human_CFTR_protein_extract.txt" sequence in the appropriate window.
- Click on "Run TMpred".

17) Understand the results and complete the following table:

Starting Position	Ending Position	Location of Segment

b) Choose SOSUI - Prediction of transmembrane regions (Nagoya University, Japan) by clicking on SOSUI.

- Choose SOSUI (SOSUI engine ver. 1.11) from the new page
- Enter "human_CFTR_protein_extract.txt" in the window and click on "Exec"

18) Understand the results and complete the following table:

Starting Position	Ending Position	Location of Segment

c) Choose TopPred - Topology prediction of membrane proteins (France)

- Enter the sequence in the appropriate window and keep the default values.
- Click on "Run".
- Enter your email address.
- When prompted, type in the validation text.
- Choose "full screen view" under "Standard output".

©2016 Sami Khuri

19) Understand the results and complete the following table:

Starting Position	Ending Position	Location of Segment

We are now going to compare the results obtained by TMHMM, MINNOU, TMpred, SOSUI, and TopPred.

20) Explain where they agree and where they disagree.

21) How accurate are the results obtained by TMHMM, MINNOU, TMpred, SOSUI, and TopPred?
Clearly explain how you can answer this question.

C) Genetic Screening and the Inheritance of Cystic Fibrosis

Mary and her husband Tom would like to start a family. Mary's mother had cystic fibrosis (CF). Tom was adopted at a very young age and little is known of his biological parents. Mary and Tom are Caucasians and are aware of the fact that CF is the most common fatal genetic disorder among Caucasians: about 1 in 25 individuals carry the allele. Mary and Tom sought the assistance of a genetic counselor who has strongly recommended genetic screening for Tom.

22) Why hasn't the counselor recommended genetic screening for Mary?

23) If Tom is a carrier, what is the probability that their child has CF? Explain.

Mary and Tom followed the genetic counselor's advice. The file "CFTR_Screening_Mary_Tom.txt" contains four FASTA sequences representing the coding regions of each of Mary's two alleles and the coding regions of each of Tom's two alleles.

24) Discuss the mutations (if any) that Mary's alleles have and discuss how (if at all) they affect the CFTR protein. Can you explain which allele she must have inherited from each of her parents?

25) Discuss the mutations (if any) that Tom's alleles have and discuss how (if at all) they affect the CFTR protein.

[Part C adapted from Exploring Bioinformatics by C. Clair and J. Visick, 2015]

©2016 Sami Khuri