

Hands-On SRB-6

Serpentine Receptor Class Beta-6

Summary of Hands-On Activity:

In this hands-on exercise we will explore some current bioinformatics tools. We will use different databases and other Internet resources to learn about the SRB-6 gene and protein in different organisms. We will use multiple sequence alignment tools to find conserved regions in a number of SRB proteins. Finally, we will learn about protein families and sub-families and how to find them on the Internet.

Objective

The objective of this hands-on exercise is to introduce students to:

- Some of the biological databases on the Internet
- Basic bioinformatics tools for
 - Pairwise sequence alignment of nucleotide or protein sequences
 - Multiple sequence alignment of nucleotide or protein sequences
 - Exploring protein/nucleotide interaction networks
- Conserved regions in proteins

A) The SRB-6 Gene and Protein

We are going to retrieve the *C.elegans* srb-6 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 srb-6 gene by typing in “srb-6” in the search window and clicking on the blue “Search”
- From the result page, choose the *C.elegans* srb-6 (with Gene ID: 191788) by clicking on “[srb-6](#)”

1) When was the record last updated? _____.

Upon clicking at the “WormBase” link to the right of “Primary source” and reading “Summary”, we learn that the protein encoded by this gene is a seven-transmembrane protein predicted to function as a chemosensory receptor.

We are going to retrieve the *C.elegans* SRB-6 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 srb-6 protein by typing in “srb-6” in the search window
- Click on the blue “Search”
- From the result page, choose the *C.elegans* srb-6 (with accession number P54141.1) by clicking on the link above the accession number

2) a) When was the record last updated? _____.

b) What is the size of the SRB-6 protein? _____.

We know that the protein is a 7-transmembrane protein. Under “Features” there are seven regions labeled as “Transmembrane region”.

3) Record in the following table, the starting and ending locations of the seven transmembrane regions.

Transmembrane Region	Strating Position	Ending Position
First		
Second		
Third		
Fourth		
Fifth		
Sixth		
Seventh		

B) Using BLAST at NCBI for finding orthologs

We are going to use BLAST at NCBI to find orthologs of the *C.elegans* SRB-6 protein.

- Go to the main page of NCBI: <http://www.ncbi.nlm.nih.gov>
- Click on “BLAST” under “Popular Resources” on the right hand side of the page
- In the new page, click on “protein BLAST” rectangle
- Enter “P54141.1” in the “Enter Query Sequence” window.

The SRB-6 sequence can now be compared to various datasets of protein sequences in various databases.

- Scroll down and choose “UniProtKB/Swiss-Prot(swissprot)” from the dropdown window next to “Database”
- Make sure that “blastp (protein-protein BLAST)” is chosen under “Program Selection”
- Choose “Show results in a new window”
- Click on the blue “BLAST” button to start the search

The graphical view (under “Color key for alignment scores”) shows an overview of the results where the sequences detected in SwissProt by the BLAST search (the “hits”) are aligned with the *C.elegans* SRB-6 protein. The “Color key for alignment scores” shows the degree of similarity between the Query sequence and the results. Below the graphical overview, the detailed list of the “sequences producing significant alignments” is given.

In the "Descriptions" section (under “Sequences producing significant alignments:”), you can examine the database matches in more details. Each database sequence has an identifier string, an accession number shown as a blue link.

On each row under “Sequences producing significant alignments” you will find a one-line, short description of the protein, and the Total score that shows the level of similarity to the QUERY sequence and the E value assigned to each “hit”. The Total score and E value are special statistics that measure the degree of similarity between two sequences. Basically, the higher the Total score, the greater the similarity between the two sequences. The lower the E value, or the closer it is to zero, the more "significant" the match is.

Below the list of hits, the individual “Alignments” for each hit are shown. For each alignment, the query sequence (“Query” – the *C.elegans* protein) is shown at the top and the hit (“Sbjct” – the human protein returned by BLAST) underneath it, with the position of the amino acids indicated on the right and left of the alignment.

Go to the second alignment (with SRB-15) and answer the following questions:

- 4) a) What is the species of that second best hit? _____.
- b) How long is this protein? _____.

- 5) What is the degree of similarity between the query and the hit?
- 6) What is the probability that the similarity between the query and the hit occurs only by chance?
- 7) What do you think the four stretches/gaps “---“ represent?
- 8) Look at the alignment with the third most relevant hit. How similar is it to the *C.elegans* SRB-6 protein sequence?

Let us further study the *C.elegans* SRB-6 protein by reading more information available about it in other relevant biological databases.

C) The SwissProt Database

By going to <http://www.uniprot.org/>, you would have accessed the “Swiss-Prot Protein knowledgebase” database hosted by the “Swiss Institute of Bioinformatics”. Note that the Protein Knowledgebase database is one of several UniProt (Universal Protein Resource) databases.

- 9) What is the mission of UniProt?
 - Type the accession number “P54141” in the “Query” field at the top of and click on “**Search**”.

The result of your search is a page for the *C.elegans* SRB-6 protein.

The page contains information grouped in categories [Function], [Names & Taxonomy], [Subcellular Location], [PTM/Processing], [Interaction], [Structure], [Family & Domains], [Sequence], [Cross-references], [Entry information], [Miscellaneous], and [Similar proteins] easily identified by golden, underlined headers.

Scroll up and down the page to study the different categories of information available.

Proteins usually interact with other proteins.

- Scroll to “Interaction”

Click on “[6239.R05H5.6](#).” next to STRING

- 10) What do the nodes in the network represent?
- 11) What can be inferred from the size of the nodes?
- 12) What do the edges/connections in the network represent?
- 13) How were the proteins associated with SRB-6 determined (example: known, predicted, etc....)?
 - Go back to the UniProtKB page of SRB-6.
 - Scroll to “Family & Domains”
 - Go to “Phylogenomic databases”
 - Click on “[ENSGT00390000002942](#).” next to GeneTree.

14) From the new page: “Gene Tree Image”, does the *srb-6* gene form a close cluster with other *srb* genes? Explain.

- Go back to the UniProtKB page of SRB-6.
- Scroll to “Subcellular location” to view the locations of the 7 transmembrane spanning helices.

Note that to the right of each transmembrane region, we have the possibility of using BLAST to find other short sequences of proteins that are very similar to the SRB-6 transmembrane region.

15) Record in the following table, the starting and ending locations of the seven transmembrane regions.

Transmembrane Region	Starting Position	Ending Position
First		
Second		
Third		
Fourth		
Fifth		
Sixth		
Seventh		

16) Compare the table above with the one you obtained from NCBI (question 3). Explain where they agree and where they disagree.

D) Comparing SRB proteins of different species

It is often useful to compare multiple sequences, for example, SRB proteins from different sequences. In general, similar protein sequences yield similar protein structures and functions. To find similarity between multiple proteins, bioinformaticians build multiple sequence alignments.

The file “SRB_seqs.doc” can be found at <http://www.cs.sjsu.edu/~khuri/Biology115/>.

It contains three SRB-6 protein sequences, one SRB-1 protein sequence, and one SRB-15 protein. All sequences are in FASTA format:

- **NP_496199.1** SRB-6 protein [Caenorhabditis elegans]
- **EGT33838.1** SRB-6 protein [Caenorhabditis brenneri]
- **CAP24158.2** CBR-SRB-1 protein [Caenorhabditis briggsae]
- **NP_494963.1** SRB-3 protein [Caenorhabditis brenneri]
- **NP_491969.2** SRB-15 protein [Caenorhabditis elegans]

To compare multiple sequences, we will use CLUSTAL Omega at the EMBL-EBI website in the United Kingdom.

- Go to <http://www.ebi.ac.uk/Tools/msa/clustalo/> and paste the five sequences (in FASTA format) from “SRB_seqs.doc” in the window.
- Use the default conditions provided by the program
- Click on "Submit" under STEP 3, at the bottom of the page to align the five sequences
- In the “Clustal Omega” result page, click on the "Show Colors" button.

Examine the alignment and answer the following questions.

17) What are the significances of the symbols “.” “:” and “*” ?

Hint: Consult: “<http://www.ebi.ac.uk/Tools/msa/clustalw2/help/faq.html#23>”.

18) Can you identify conserved regions in the five proteins?

Alternatively, we could perform the multiple sequence alignment with T-Coffee.

- Go to <http://tcoffee.crg.cat/apps/tcoffee/do:regular>
- Paste the 5 sequences in FASTA format
- Enter your email address and click on “Submit”

19) According to the “T-Coffee alignment result”, is this a good alignment or not?

20) Do you think that the first three sequences are orthologous? Explain.