

## Hands-On PRINTS

This tutorial aims to illustrate some of the issues raised in the presentation regarding the use of PRINTS for sequence analysis, and how the resource complements the other databases that combine to make up InterPro.

### 1. Review our target sequence

First, we will locate our target sequence. Visit the UniProt web site (<http://www.uniprot.org/>) and search for Q9C929. Click on the entry and look at the protein name and associated keywords. What does the function of this protein appear to be?

### 2. Perform a BLAST search

Next we will perform a BLAST search to find proteins whose sequences are similar to our query protein, to see if the results may help us infer anything further about its function.

Click on 'fasta' at the top right hand side of the Q9C929 UniProt entry page to access the sequence in FASTA format. In a new browser tab or window, navigate to the EBI's NCBI-BLAST page (<http://www.ebi.ac.uk/Tools/sss/ncbiblast/>). Copy and paste the FASTA sequence into the form and perform a BLAST search against the UniProt Knowledgebase.

Examine the top BLAST hits. Do they support the protein function suggested in UniProt?

### 3. Searching the PRINTS database

Now we will search our sequence against the PRINTS database. Navigate to the FingerPRINTScan web site

(<http://www.bioinf.manchester.ac.uk/cgi-bin/dbbrowser/fingerPRINTScan/muppet/FPScan.cgi>). Type Q9C929 into the form and press submit.

What matches to the database are returned? How do they relate to each other (hint – click on 'Relations' tag next to your results). Clicking on the matches and reading the annotation, what does PRINTS suggest the function of the protein might be? How does this compare with the function suggested in UniProt?

### 4. Searching InterPro

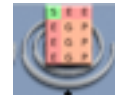
Next we will search our sequence against InterPro. Navigate to the InterPro web site (<http://www.ebi.ac.uk/interpro/> to use the current web site, or <http://wwwdev.ebi.ac.uk/interpro-5.2/> to try the experimental new interface). Enter Q9C929 into the search box and examine your results.

What database matches are returned? Do the different matches relate to each other? Do they support any of the previous functional assignments? Is there anything in the annotation to explain any discrepancies you might see between the protein function suggested in UniProt or by your BLAST results and PRINTS/InterPro matches?

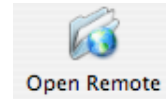
### 5. Examining sequence alignments

For the next task, we will align the sequence Q9C929 against some example LanC-like protein sequences and G-protein-coupled receptor sequences so that we can compare them directly. In order to perform this exercise you will need to use the CINEMA multiple sequence alignment editor, which is installed on your desktop; if you wish to use the software on your laptop, it is available for download at <http://utopia.cs.manchester.ac.uk/>

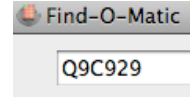
Click on the CINEMA icon: this may appear either on your desktop, on your toolbar or in your Applications menu.



Click on the Open Remote icon at the top of the CINEMA manager window – this invokes the Find-O-Matic tool.



Now let's locate the candidate sequence. In the Find-O-Matic search box, type Q9C929 and hit Return.



Drag-and-drop Q9C929\_ARATH into the CINEMA manager window: this fetches the sequence from UniProt (it may take a few of moments), and creates a new CINEMA editor window.



We now want to gather some LanC-like protein and G-protein-coupled receptor sequences, so that we can compare them to Q9C929. Return to the Find-o-Matic window, and type LANC1\_HUMAN into the query box and press return, then drag the resultant sequence into the window containing the Q9C929 sequence (*i.e.*, into the CINEMA editor window, not the CINEMA manager window). Repeat this process with LANC2\_MOUSE, OPSD\_BOVIN & HH1R\_BOVIN (the latter two being rhodopsin-like GPCRs) so that all 5 sequences are visible in the alignment editor.

To align the sequences, right click on a sequence, select the Alignment option from the context-sensitive menu, and choose ClustalW or Muscle – this will send all the sequences to the chosen Web service, returning a completed alignment in seconds.

Examine the sequence alignment. Does Q9C929 look more like the LanC-like protein sequences or the GPCR sequences?

## 5. Computationally predicting likely TM domains

Finally, within the sequence alignment, we want to show where any transmembrane (TM) domains may be predicted to be. We will start with a GPCR sequence (this type of protein characteristically has 7 TM domains). To do this, click on OPSD\_BOVIN to select it, right click to invoke the context-sensitive menu, select the Sequence OPSD\_BOVIN option and choose Annotate TM domains, using the TMHMM option. The locations of the predicted TM domains are highlighted within the sequence by means of blue helices beneath the relevant regions in the alignment. Repeat this process with the other sequences in the alignment. In terms of TM domains, does your sequence look more like a LanC-like protein or a GPCR?

## 6. Conclusion

The sequence we have been investigating was published in *Science* as the second ever GPCR identified in *Arabidopsis thaliana*. Based on the evidence you have seen during the above exercise, would you identify this sequence as a GPCR?

## 7. Additional sequences to investigate

If you have time, you might like to like to investigate other UniProt entries, such as A8T8V4 or A8UMV2. Repeating the process outlined above, what does UniProt report the function of these proteins to be? What do PRINTS, InterPro and associated sequence analysis resources suggest?

Note, some of the steps described in this hands-on session (running BLAST, searching the PRINTS database, *etc*) can be run directly from the CINEMA alignment editor, helping to streamline aspects of the sequence analysis process. For more information on CINEMA and its capabilities, the full spectrum of which we've not had time to describe, there are help documents online at <http://utopia.cs.man.ac.uk/utopia/documentation>.