

## *Proteomics Data Informatics: Scaffold (self-paced tutorial)*

### Purpose of this tutorial:

This lab exercise is a self-paced tutorial designed to introduce you to the statistical analysis of large proteomics datasets as well as web-based bioinformatics tools. It is intended to simulate the process you might go through when evaluating the search results from a large proteomics project.

### About your sample:

An organelle fractionation was performed to isolate mitochondria from rat cells. The mitochondrial proteins were further fractionated by reverse phase HPLC. The resulting fractions were reduced with DTT, alkylated with iodoacetamide, and digested with trypsin. The digested fractions were then analyzed by LC-MS/MS utilizing a quadrupole-time-of-flight mass spectrometer. The resulting spectra were searched against rat proteins in the Swiss-Prot database using the Mascot, Sequest, and X! Tandem database search engines. The database search results from the three search engines were combined in Scaffold and a statistical analysis was performed to validate the protein identifications.

It is your job to evaluate the dataset in Scaffold and determine which of the protein identifications are valid. Once you have confidence in your protein identifications, you will need to begin to explore the biological significance of the proteins in your list.

## **Tutorial**

### **Statistics View:**



Click on the Statistics icon on the left side of the screen.

Highlight the sample called “Mitochondria Pooled Fractions”. The search results from all of the fractions have been pooled together in this dataset.

*How many spectra are in this dataset? What percentage of the spectra were actually identified? (hint: the boxes can be resized to see all of the columns)*

Click the Mascot tab (+2 charge state) in the score histogram pane.

*Is there enough data to utilize the peptide and protein prophet statistical analysis? (hint: is there enough data to fit to a bimodal distribution?)*

## Samples View:



Click on the Samples icon on the left side of the screen.

Examine your list of identified proteins.

*How many proteins were identified (with the default threshold settings) in the pooled dataset? How many are in each fraction? (hint: if you double click on a column it will sort all the proteins in the list by what is found in that sample)*

Change the probability threshold settings and the minimum # of unique peptides and observe how this changes the number of identified proteins.

Many of the proteins were found in all of the fractions. One example is the voltage-dependent anion-selection channel protein. Find this protein in the list and highlight it. Scaffold has extracted the GO terms for this protein.

*(hint: Type the protein name in the search box to help you locate the protein in the list)*

*What are the cellular and physiological processes associated with this protein?*

*What is the cellular location of this protein?*

## Proteins View:



Double click on the serum albumin precursor protein that was found in the pooled dataset and fraction 3 (*again, you can type this protein name in the search box to find it*).

*What is the molecular weight of this protein?*

*What percent sequence coverage was identified in the experiment?*

Find and highlight the peptide with sequence (R)FPNAEFAEITK(L).

*Was this peptide identified by both search engines? What are the search engine scores? How many times was this peptide found in the sample?*

Examine the fragmentation spectrum for this peptide.

*Is there good signal-to-noise? Is there good coverage of y and b ions?*

*With your manual evaluation, would you say this is a good identification?*

## Samples View:



Click the Samples icon again on the left side of the screen.

Now that you are familiar with how to navigate in Scaffold you can evaluate the data and determine the appropriate probability threshold settings. To do this, vary the peptide, protein, and minimum # of unique peptide threshold settings and then look at the actual spectra in the Proteins view to evaluate the quality of the identifications at that threshold. You can focus your evaluations on proteins with < 95% confidence (color coded as yellow). (*hint: Double click on a protein in the list to see the peptide spectra.*) Raise the thresholds to a level at which you are confident in the results.

When you are confident of the proteins identified in your list it is time to begin exploring the biological significance/relevance of the proteins.

Find and highlight the voltage-dependent anion-selection channel protein (VDAC1). At the bottom left of the screen choose to look up the accession number in Swiss-Prot Database. Click on the accession number in the box to directly link out to the Swiss-Prot entry for this protein.

ExPASy Home page | Site Map | Search ExPASy | Contact us | Swiss-Prot

Search:  for

### UniProtKB/Swiss-Prot entry [Q9Z2L0](#)

[\[Entry info\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#) [\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information	
Entry name	VDAC1_RAT
Primary accession number	Q9Z2L0
Secondary accession numbers	A1L125 Q3MHT8 Q5M944 Q5M972 Q6IN28 Q6P9W9
Integrated into Swiss-Prot on	December 1, 2000
Sequence was last modified on	January 23, 2007 (Sequence version 4)
Annotations were last modified on	May 1, 2007 (Entry version 46)

Name and origin of the protein	
Protein name	Voltage-dependent anion-selective channel protein 1
Synonyms	VDAC-1 VDAC1 Outer mitochondrial membrane protein porin 1
Gene name	Name: Vdac1
From	Rattus norvegicus (Rat) [TaxID: 10116]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

References	
[1] NUCLEOTIDE SEQUENCE [MRNA]	
TISSUE=Heart;	
DOI=10.1016/S0167-4781(98)00088-8; PubMed=9714728 [NCBI, ExPASy, EBI, Israel, Japan]	
Artifous K, Blondel O, Bernard A, Khrestchatsky M, Ventura-Clapier R.	
"Characterization of rat porin isoforms: cloning of a cardiac type-3 variant encoding an additional methionine at its putative N-terminal region.;"	
Biochim. Biophys. Acta 1399:47-50(1998).	
[2] NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.	
TISSUE=Acute tumor.	

Notice the sections listed in purple font: **Entry info, Name and origin**, etc. When instructed, click on the purple font of each section to navigate to that section. Answer the questions in each of the sections below.

### **Entry Information:**

Recall that multiple entries can be found for a single protein. Similarly, multiple accession numbers may be associated with a single protein.

*How many accession numbers are there for your protein?* \_\_\_\_\_

### **Name and Origin of Protein:**

*What is the protein name?* \_\_\_\_\_

*The gene name?* \_\_\_\_\_

Knowing synonyms for your protein can help you mine more data than if you focus on a single name.

*Is your protein known by any other name?* \_\_\_\_\_

*What species is this particular protein from?* \_\_\_\_\_

### **References**

Here you will see publications with sequence information on VDAC1, which can be used to find structural, functional, and/or sequence information. Click on the **DOI** and **Pubmed** fields to be linked to abstracts or entire documents relevant to this protein. In some cases, you will have instant access to the publication. Experiment by clicking on publication number 1 (PubMed=9714728 [NCBI]).

### **Comments**

This section contains functional, structural, and biological information about your protein. Click on the **BOLD** words to obtain information about that field.

*What is the probable function of VDAC1?* \_\_\_\_\_

*Where would you expect to find VDAC1 in a cell?* \_\_\_\_\_

### **Cross reference**

This section contains information on the sequence and structure of the protein, as well as the origination of this data. There is a plethora of information in this section and you should feel free to click on anything that looks interesting. Here are some suggested items to review.

Notice the **Ontologies** section which describes from where the data originated. Click on “inferred from direct assay”.

*How confident would you be regarding the molecular function VDAC1 if this was the only evidence available?*

## Key words

This section can be used to gather additional information about other proteins that share the same or similar structural and functional features of your protein. Selecting any of these key words will provide a definition as well as a list of proteins that share the key word.

Scroll down to the very bottom of the page. You will notice several links. In **Sequence Analysis Tools** click **compute pI/MW**.

[View entry in original UniProtKB/Swiss-Prot format](#)

[View entry in raw text format \(no links\)](#)

[Report form for errors/updates in this UniProtKB/Swiss-Prot entry](#)

 [BLAST submission on ExPASy/SIB or at NCBI \(USA\)](#)



Sequence analysis tools: [ProtParam](#), [ProtScale](#), [Compute pI/MW](#), [PeptideMass](#), [PeptideCutter](#), [Dotlet \(Java\)](#)



[ScanProsite](#), [MotifScan](#)



[Submit a homology modeling request to SWISS-MODEL](#)



[NPSA Sequence analysis tools](#)

Scroll down on the next page and click **submit**.

*What is the estimated molecular weight \_\_\_\_\_ and pI \_\_\_\_\_?*

If this protein was isolated from a 1 or 2D gel, it would be important to compare these estimates with the actual migration in SDS PAGE gels.

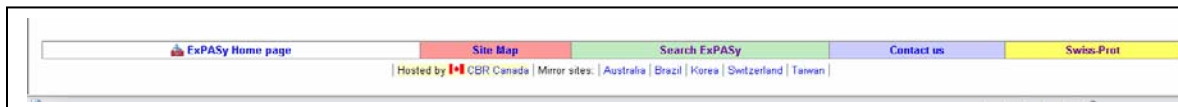
*If the information did not match what are some reasons why this would occur?*

- A) *You have a truncated form of the protein*
- B) *The wrong sequence is in the database*
- C) *You are working with a mutated form of the protein*
- D) *Any of the above may be correct*

Use your Back button to return to the SwissProt entry.

Highlight the amino acid sequence of VDAC1 and right click to copy.

## Other Useful Proteomics Tools:



Go to the very bottom of the page and click the yellow Swiss-Prot tab:



From here, click the orange tab Proteomics tools in the upper right corner:





This is the ExPASy Proteomics Tools website. It includes a large number of links to proteomics tools.

For example, scroll down to the bottom of the page to the section called Post-translational modification prediction. Click on the link NetPhos. Paste the sequence of VDAC1 and submit.

*What are the predicted phosphorylation sites for this protein?*

If time permits, investigate some of the other tools on the ExPASy page and use them to gather more information on VDAC1.