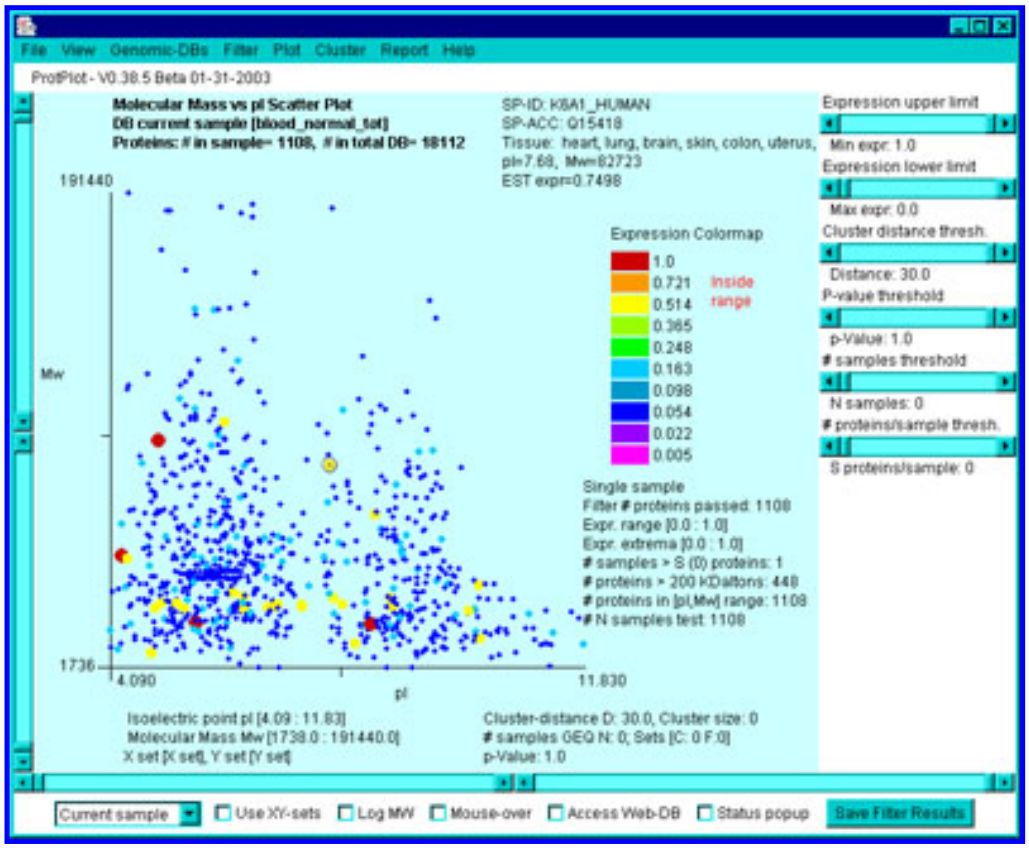


ProtPlot Data Mining Tool for Virtual Protein Expression Patterns

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | Revision history | PRP file format | PDF documents | History of ProtPlot | Latest version

Introduction

ProtPlot is a Java-based data-mining software tool for virtual 2D gels. It may be downloaded and run as a stand-alone application on your computer. Its exploratory data analysis environment provides tools for the data-mining of quantified virtual 2D gel (pIe, Mw, expression) data of estimated expression from the CGAP EST mRNA tissue expression database. This lets you look at the aggregated data in new ways: for example, which estimated "proteins" are in a specified range of (pI,Mw)? Or which sets of estimated "proteins" are up or down regulated or missing between cancer samples and normal samples? Which sets or "proteins" cluster together across different types of cancers or normals? Here, one may aggregate several different normal and several different cancers as well as specify other filtering criteria.

As is well known, mRNA expression generally does *not* correlate well with protein expression as seen in 2D-PAGE gels (Ideker et.al., *Science* 292: 929-934, 2001). However, some new insights may occur by viewing the transcription data in the protein domain. If actual protein expression data is available for some of these tissues, it might be useful to compare mRNA estimated expression and actual protein expression. This tool may helps find those proteins with similar expression and those that have quite different expression. This might be useful in thinking about new hypotheses for protein post-modifications or mRNA post-transcription processing.

ProtPlot generates an interactive virtual protein 2D-gel Map scatterplot based on a database of derived maximum EST expression over a variety of tissue types from data obtained from the NCI-NCBI CGAP EST database of human cancer, precancer and cancer mRNA expression (CGAP is the NCI's Cancer Genome Anatomy Project http://cgap.nci.nih.gov/. EST is the Expressed Sequence Tag of a mRNA found in particular tissues). The EST hit rate is a rough estimate of gene expression. These ESTs were mapped to SWISS-PROT (expasy.ch) accession numbers and Ids, the Mw and pI estimates were computed and used as estimates for corresponding proteins in a pseudo 2D-gel.

ProtPlot data is contained in a set of tissue- and histology-specific .prp (i.e., PRotPlot) files described in the data format documentation. These are kept in the PRP directory that comes with ProtPlot when you install it. You will be able to update these .prp files from the ProtPlot Web server http://www.lecb.ncifcrf.gov/TMAP.

The ProtPlot Web site and program Help menu provide additional information:

- History of ProtPlot
- Using ProtPlot
- Menus descriptions
- Estimating expression
- Data format documentation
- Downloading the program
- Disclaimer and License

- Screen Shots of ProtPlot (PDF)

NOTE: this software is undergoing Beta-testing so full functionality may not be available.

Contacts: Djamel Medjahed, LMT, SAIC-Frederick.
Peter Lemkin, LECB, NCI-Frederick.

Revised: 08-26-2004

Using ProtPlot for Data Mining Virtual Protein Expression Patterns

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | PRP file format | Revision history | PDF documents | History of ProtPlot | Latest version

Installing ProtPlot

First you need to download and install ProtPlot on your computer. The detailed steps are shown in the PDF Introduction to data-mining with ProtPlot: Screen Shots. This downloads the ProtPlot Java program and the CGAP derived data set of pseudo 2D-gels. If you download the version that includes the Java Virtual Machine (JVM), it will not interact with any other JVM you have installed.

Using ProtPlot

You start ProtPlot by clicking on the "ProtPlot Startup" icon if your computer supports that (Windows, MacOS-X, etc.) or type ProtPlot on the command line for Unix, Linux and other systems.

Once the ProtPlot program is started, it loads the set of PRP files that you downloaded with the ProtPlot program. The virtual protein data for each tissue is used to construct a Master Protein Index where proteins will be present for some tissues and not for others. The data is presented in a pseudo 2D-gel image with the estimated isoelectric point (pI) on the horizontal axis and the molecular mass (Mw) on the vertical axis. Sliders on each of the axes allow you control the minimum and maximum values of pI and Mw displayed and thus the Mw vs. pI scatterplot zoom region you want to select. By clicking on a spot in in the scatterplot, you will display information on that protein. You also define that protein as *the current protein*. The current protein is used in some of the clustering methods, protein specific reports (Expression Profile report), and the Expression Profile plot. If you have enabled the popup Genomic-ID Web browser and you are connected to the Internet, it will popup a Web page from the selected Genomic database for that protein.

You select various options from the pull-down menus. Some of the more commonly used options are replicated as check-boxes at the bottom of the window.

The Scatterplot Display Mode

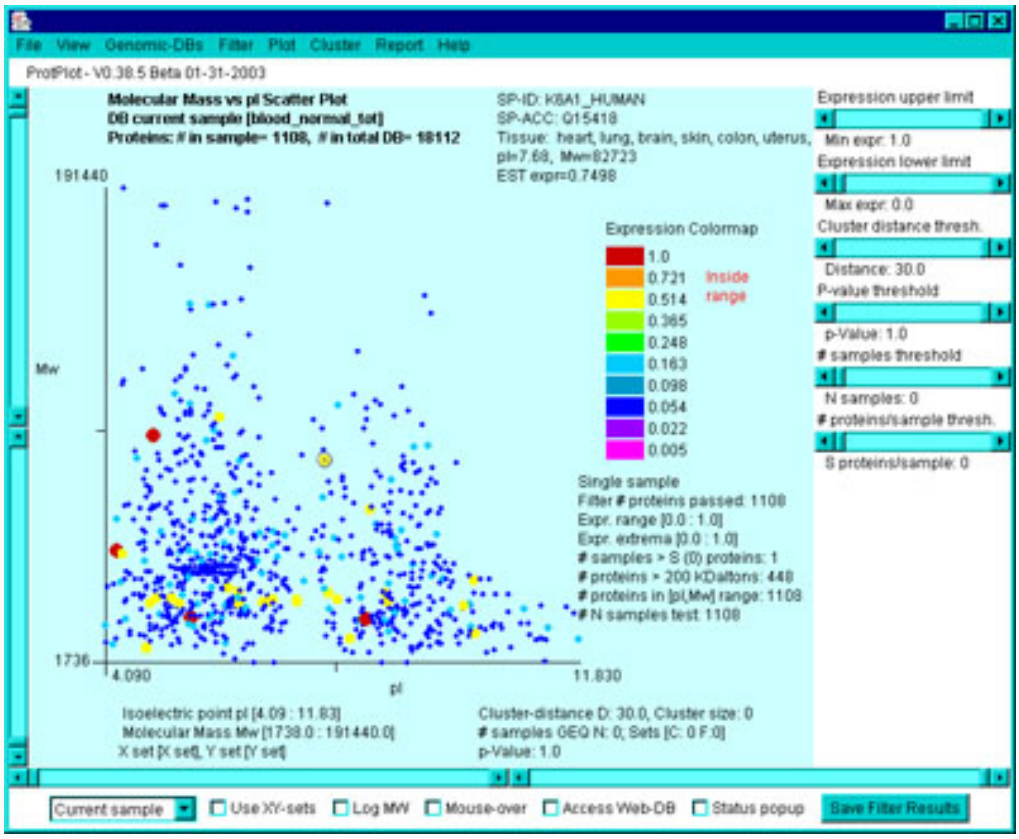
There are two primary types of pseudo 2D-gel (Mw vs pI) scatterplot display modes of this derived protein expression data: *expression mode* or *ratio mode*. The expression data may be for a single sample (*the current sample*) or the mean expression of a list of samples (called *the expression profile* or EP). The ratio data is compute as the ratio of two individual samples called X and Y. Ratio data may alternatively be computed from sets of X samples and sets of Y samples. Generally, one would group a set of samples with similar characteristis together having the same condition (e.g., cancer, normal, etc.). The ratio of X and Y may be single samples in which case the ratio is computed as:

```
ratio = (expression X / expression Y)
```

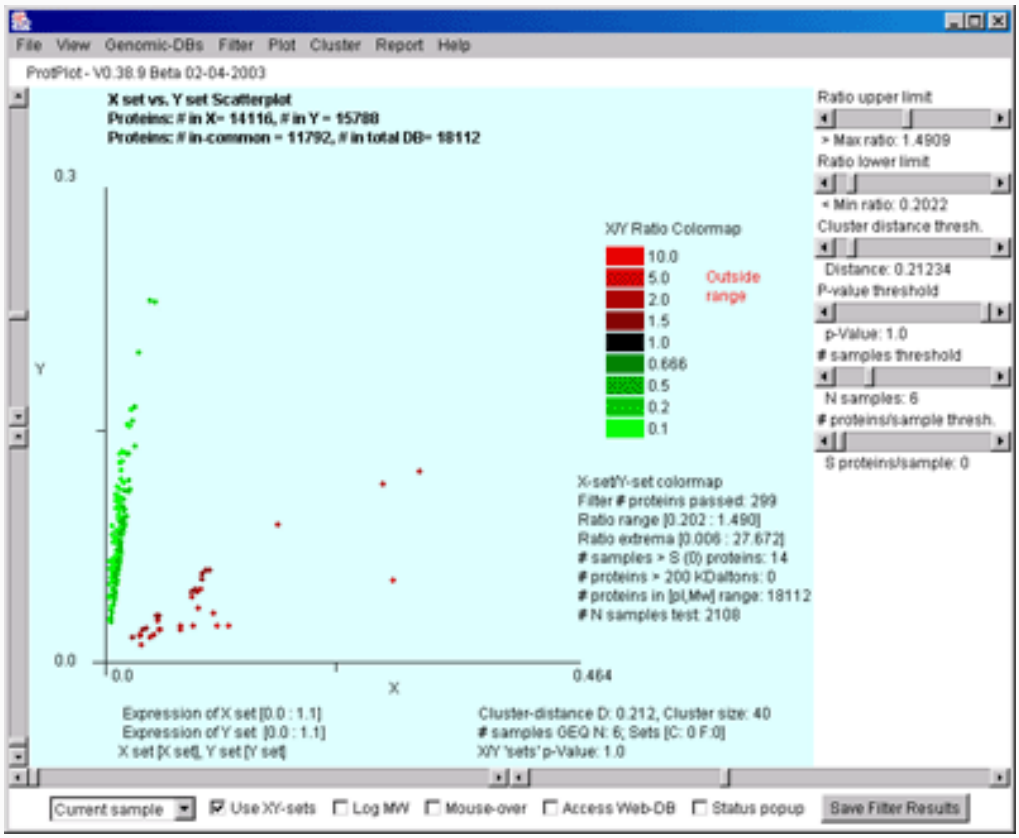
where expression X (expression Y) is the expression of corresponding proteins. Alternatively, you may compute the ratio of the mean expression of two different sets of samples (the X set and the Y set). The X and Y sets may be thought of as experimental conditions and the members of the sets being "replicates" in some sense. E.g., the X set could be cancer samples and the Y set could be normal samples. The ratio of the X/Y sets for each corresponding protein is computed as

```
ratio = (mean X-set expression / mean Y-set expression)
```

The following shows one of the (Mw vs. pI) scatterplots when the display mode was set to (X-set/Y-set) ratio mode:



It is also possible to create an (X vs Y) scatter plot or (Mean X-set vs. Mean Y-set) scatterplot when the corresponding ratio display mode is set. The following window shows the (Mean X-set vs. Mean Y-set) scatterplot:



The following table summaries the four types of display modes:

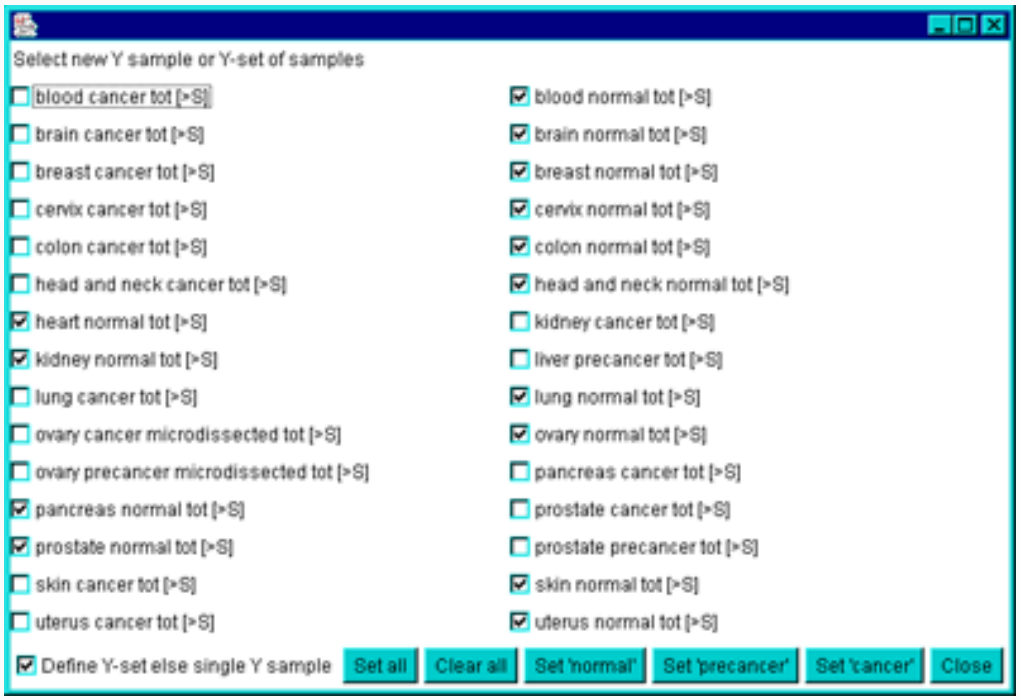
Display Mode	Current sample	Single X/Y	X-set/Y-set	EP-set
Expression	yes	no	no	no
Single samples ratio	no	yes	no	no
X-set and Y-set samples ratio	no	no	yes	no
Mean Expression	no	no	no	yes

Effect of display mode on filtering, clustering and reporting

You select the particular display mode using the Plot menu comands. When you select a particular display mode, it will enable and disable Filter, View, Cluster and Report options depending on the mode. For example, you may only use the t-Test or missing X Y set test if you are in XY-sets ratio mode. You may only perform clustering if you are in EP-set mode. You may change the display mode using the (Plot menu | Show *display mode*) commands. Alternatively, since it is used so often, there is a checkbox at the bottom of the main window "☑ Use XY-sets" that will toggle between the XY-sets ratio mode and whatever the previous mode you had set.

Selecting Samples

You select samples for the current sample, X sample, Y sample, X-set samples, Y-set samples, and EP-set samples using a popup checkbox list chooser of all samples.



This may be invoked either from the File menu or the pull-down sample selector at the lower-left corner of the main window.

For example, you invoke this chooser for a the specific tissue sample you want to view by using the (File menu | Select samples | Select Current PRP sample). For X (Y) data, you invoke the choosers using (File menu | Select samples | Select X (Y) PRP sample(s)). You may switch between single (X/Y) and (X set/Y set) mode using the (File menu | Select samples | Use Sample X and Y sets else single X and Y samples [CB]) command.

There is an alternative display called the 'Expression Profile' (EP) plot which display a list of a subset of PRP samples for the currently selected protein. You may also display the scatterplot on the mean EP data for all proteins. The EP samples are specified using the (File menu | Select samples | Select Expression List of samples) command.

Listing a report on sample assignments

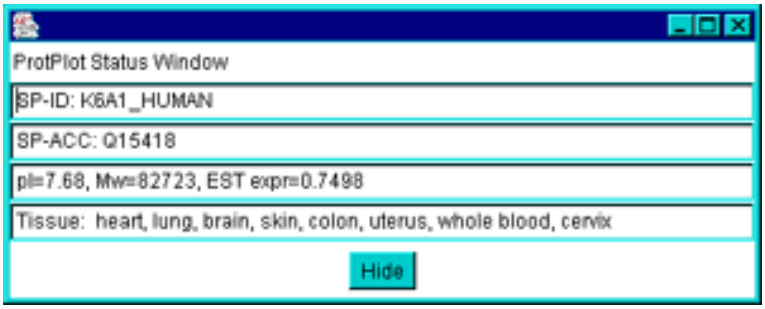
You may popup a report of the current sample assignments for the: current sample single X sample, single Y sample, X sample set, Y sample set, and EP sample set using the (File menu | Select samples | List sample assignments) command.

Assigning the X-set and Y-set condition names

The default experimental *condition* names for the X and Y sample sets are 'X set' and 'Y set'. You may change these by the (File menu | Select samples | Assign X (Y) set name) commands.

Status Reporting Window

There is a status popup window that first appears when the program is started and reports the progress while the data is loading. After the data is loaded, it will disappear. You may bring it back at any time by toggling the "☑ Status popup" checkbox at the bottom of the window. You may also press the "Hide" button on the status popup window to make it disappear.



Data Filtering

The pseudo-protein data is passed through a data filter consisting of the intersection of several tests including: pI range, MW range, sample expression range, expression ratio(X/Y) range (either inside or outside the range), t-Test comparing the X and Y sample sets, Kolmogorov-Smirnov test comparing the X

and Y sample sets, missing proteins test for X and Y sample sets, tissue type filter, protein family filter [Future], and clustering. The filtering options are selected in the Filter menu. If you are looking at the scatterplot in ratio mode, then you may filter by ratio of X/Y either inside or outside of the ratio range. The missing protein test defines missing as totally missing and present as having at least 'N' samples present. Note that the t-Test and the missing protein test are mutually exclusive in what they are looking for, so using both results in no proteins found.

Saving filtered proteins in sets for use in subsequent data filtering

You may save the set of proteins created by the current data filter settings by pressing the "Save Filter Results" button in the lower-right of the main window. This set of proteins is available for use in future data filtering using the (Filter menu | Filter by AND of Saved Filter proteins [CB]). When you save the state of the ProtPlot database (Filter menu | State | Save State), it will also write out the save protein sets (saved filtered proteins and saved clustered proteins) in the database "Set" folder with ".set" file name extensions.

In the (Filter menu | State | Protein Sets) submenu there are a number of commands to manipulate protein set files. You may individually save (or restore) any particular saved filtered set to (or from) a set file in the "Set" folder. There are also commands to compute the set intersection, union or difference between two protein set files and leave the resulting protein set in the saved Filter set.

Filter dependence on the display mode

Note that the particular filter options available at any time depend on what the current display mode is. The following table shows which options are available for which display modes.

indicate that the command

Filter Name	Current sample	Single X/Y	X-set/Y-set	EP-set
> 200K Daltons	yes	yes	yes	yes
Tissue type	yes	yes	yes	yes
Expression (Ratio) range	expression	ratio	ratio	expression
X/Y (inside/outside) range	no	yes	yes	no
(X-set, Y-set) t-Test	no	yes	yes	no
(X-set, Y-set) KS-Test	no	yes	yes	no
(X-set, Y-set) Missing data	no	yes	yes	no
At Most (Least) N samples	no	no	yes	yes
AND of saved cluster set	yes	yes	yes	yes
AND of saved filter set	yes	yes	yes	yes

The data-mining 'State'

The current data-mining settings of ProtPlot is called the 'state'. It may be saved in a named startup file called the 'startup state file' in the "State" folder. The "State" folder and other folders used by ProtPlot are found in the directory where you installed ProtPlot. Initially there is no startup state file. If you save the state it creates this file. You may create as many of these saved state files as you want. You may change the file and thus save various combinations of settings of samples for the current, X, Y and expression list of samples. The state also includes the the various filter, view and plot options as well as the pI, Mw, expression, ratio, cluster distance threshold, number samples threshold, p-Value threshold sliders, as well as other settings. The saved Filter and Cluster sets of proteins are also written out as .set files in the "Set" folder when you save the state.

Starting ProtPlot by clicking on the ProtPlot startup icon will not read the state file when it starts up. However, if you have saved a state, clicking on the state file or a shortcut to the state file will cause it to be read when ProtPlot starts up.

You may save the current state using either the (File | State | Save State) command to save it under the current name, or using either the (File | State | Save As State) command to save it under a new name you may specify. Then you may also change the current state using (File | State | Open Statefile) command.

The Molecular Mass vs pI Scatterplot : expression or ratio

There are to types of scatterplots: expression for a single sample or the ratio of 2 samples X and Y. The Plot menu lets you switch the display mode. Ratio mode itself has two types of displays: red(X) + green(Y), or a ratio scale ranging between <1/10 (green) and >10 (red). You may view a popup report of the expression or ratio values for the current protein. If 'Mouse-over' is enabled, then moving the mouse over a spot will show the name of the protein and its associated data. If mouse over is not enabled, then clicking on the spot will show its associated data.

You may scroll the scatterplot in both the pI and Mw axes by adjusting the end-point scrollbars on the corresponding axes. You may display the scatterplot with a log transform of MW by toggling the log MW switch.

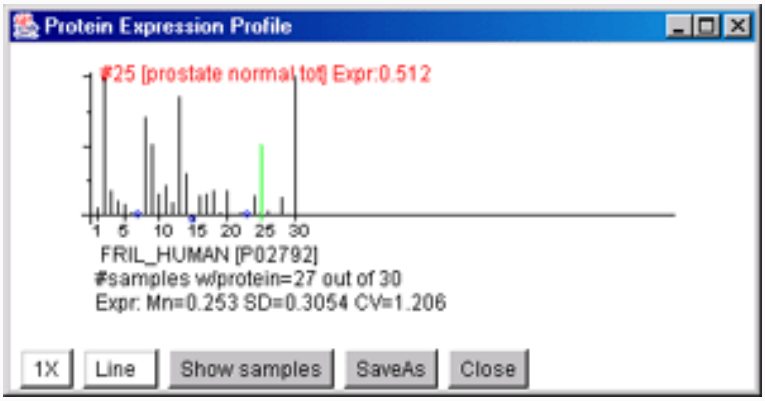
The popup plots and scatterplot may be saved as .gif image files which are put into the project's "Report" folder. Similarly, reports are saved as tab-delimited .txt text files in the "Report" folder. Because it prompts you for a file name, you may browse your file system and save the file in another disk location.

X sample(s) vs Y samples scatterplot

If you are in X/Y ratio mode (single X/Y samples or X-set/Y-set samples), you may view a scatterplot of the X vs Y expression data. Enable the XY scatterplot using the (Plot menu | Display (X vs Y) else (Mw vs pI) scatterplot - if ratio mode [CB]). You may zoom the scatterplot just as you do for the (Mw vs pI) scatterplot. The proteins displayed are those passing the data filter that have both X and Y data (i.e., expression is > 0.0).

Expression Profile plot of a specific protein

An expression profile (EP) shows the expression for a particular protein for all samples that have that protein. The (Plot menu | Enable expression profile plot) pops up a EP plot window and displays the EP plot for any protein you select by clicking on it. The relative expression is on the vertical axis and the sample number on the horizontal axis. Pressing on the "Show samples" button pops up a list showing the samples and their order in the plot. Pressing on the "nX" button will toggle through a range of magnifications from 1X through 50X that may be useful in visualizing low values of expression. Clicking on a new spot in the (Mw vs. pI) scatterplot will change the protein being displayed in the EP plot. Within the EP plot display, you may display the sample and expression value for a plotted bar by clicking on the bar (which changes to green with the value in red at the top). You may save the EP plot as a GIF file. You may also click on the display to find out the value and sample. Note: since clustering uses the expression profile, you must be in 'mean EP-set display' mode.



Clustering of expression profiles

You may cluster proteins by the similarity of their expression profiles. First set the plot display mode to "Show mean EP-set samples expression data". The clustering method is selected from the Cluster menu. Currently there is one cluster method. Others are planned.

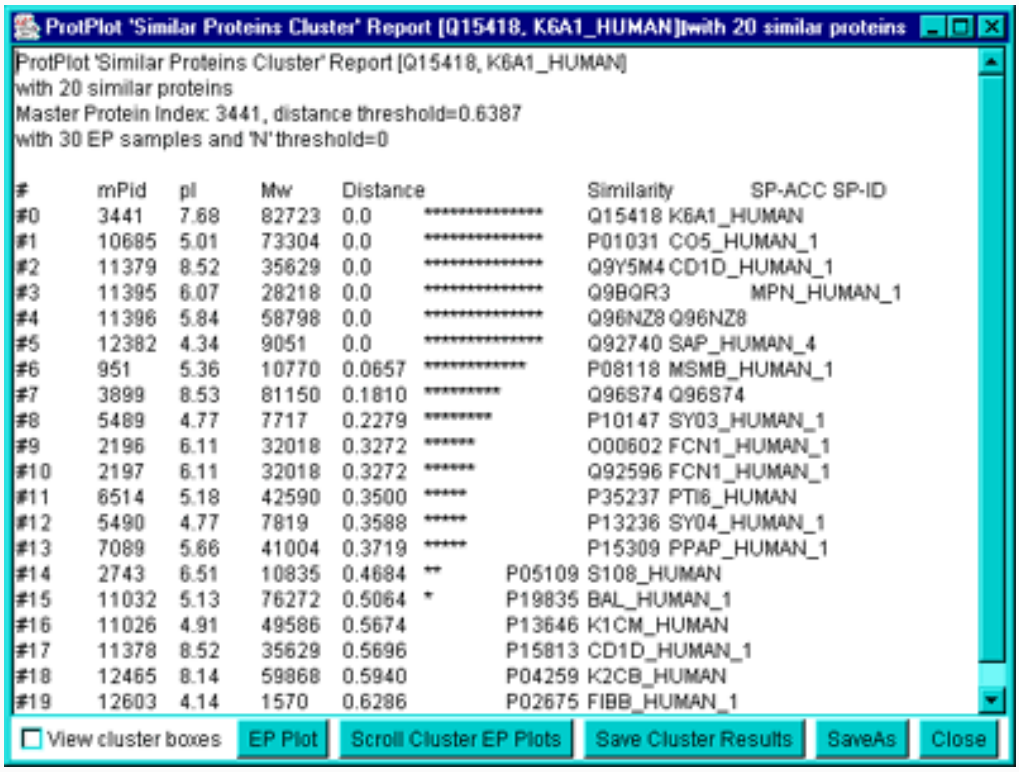
The cluster distance metric is the 'distance' between two proteins based on their expression profile. The metric may be selected in the Cluster Menu. Currently, there is one clustering method: cluster proteins most similar to the current protein (specified by clicking on a spot in the scatterplot or using the Find Protein by name in the Files menu). It requires you to specify a) the current protein, and b) the threshold distance cutoff. The threshold distance is specified interactively by the "Distance Threshold T" slider. The 'Similar Proteins Cluster' Report will be updated if you change either the current protein or the cluster distance.

The cluster distance metric must be computed in a way to take missing data into account since a simple Eucledian distance can not be used with the type of sparse data present in the ProtPlot database. ProtPlot has several ways to compute the distance metric using various models for handling missing data.

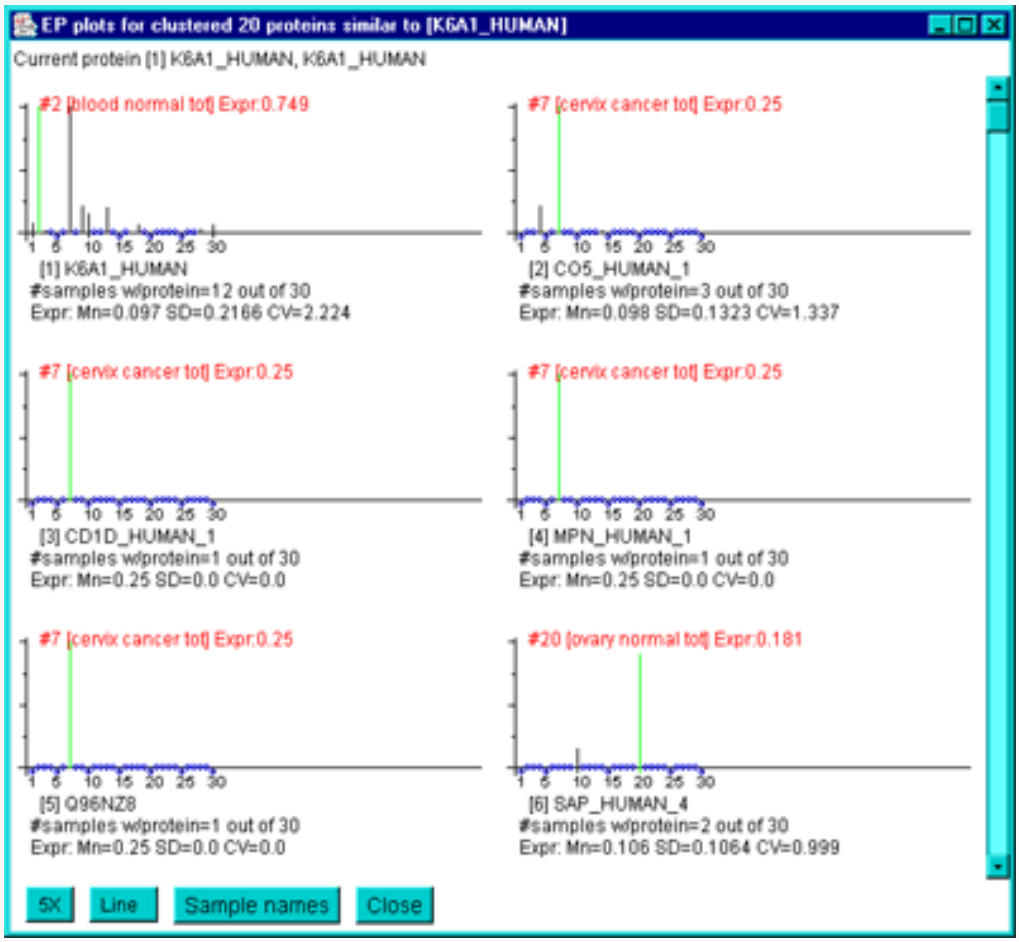
You may save the set of proteins created by the current clustering settings by pressing the "Save Cluster Results" button in the lower-right of the cluster report window. This set of proteins is available for use in

future data filtering using the (Filter menu | Filter by AND of Saved Clustered proteins [CB]). When you save the state of the ProtPlot database (Filter menu | State | Save State), it will also save the set of saved clustered proteins in the database "Set" folder. You may restore any particular saved clustered set file.

You may bring up the EP plot window by clicking on the "EP Plot" button and then click on any spot in the scatterplot to see its expression profile. Clicking on the "Scroll Cluster EP Plots" button brings up a scrollable list of expression profiles for just the clustered proteins sorted by similarity.



The following window illustrates the scrollable list of EP plots sorted by the current cluster report similarity.



You may mark the proteins belonging to the cluster in the scatterplot with black boxes by selecting the "☑ View cluster boxes" checkbox at the lower left of the cluster reportwindow. This is illustrated in the following window:

Reports

Various popup report summaries are availble depending on the display mode. All reports are tab-delimited and so may be cut & pasted into MS Excel or other analysis software. Reports also have a 'Save As' button so you can save the data into a tab-delimited file. The default /Report directory is in the directory where you installed ProtPlot. However, you may save it anywhere on your file system. The contents of some reports depends on the particular display mode. This is summarized in the table below.

Filter Name	Current sample	Single X/Y	X-set/Y-set	EP-set
Statistics or proteins passing filter	SP-ACC/ID, pI, Mw, expression	SP-ACC/ID, pI, Mw, X/Y, X, Y expr, Tissues	SP-ACC/ID, pI, Mw, mnX/mnY, (mn,sd,cv,n) expr for X- & Y-sets, Tissues. If using t-test then (dF, t-stat, F-stat). If using KS-test then (dF, D-stat)	SP-ACC/ID, pI, Mw, (mn,sd,cv,n) exprfor EP-set, Tissues

Expression profiles of proteins passing filter	SP-ACC/ID, expr data EP-set	SP-ACC/ID, expr data EP-set	SP-ACC/ID, expr data EP-set	SP-ACC/ID, expr data EP-set
X &Y sets of missing proteins pasing filter	no	no	SP-ACC/ID, (mn,sd,cv,n)for X- & Y-sets	no
EP set statistics of proteins passing filter	no	no	no	SP-ACC/ID, (mn,sd,cv,n) for EP-set
List of samples in current EP profile	{Nbr, sample-name, expression)	{Nbr, sample-name, expression)	{Nbr, sample-name, expression)	{Nbr, sample-name, expression)
List of all sample assignments	Current, X, Y, X-set, Y-set, EP-set	Current, X, Y, X-set, Y-set, EP-set	Current, X, Y, X-set, Y-set, EP-set	Current, X, Y, X-set, Y-set, EP-set
List of # proteins/sample	{Sample-name, # proteins in sample}	{Sample-name, # proteins in sample}	{Sample-name, # proteins in sample}	{Sample-name, # proteins in sample}
ProtPlot state	State	State	State	State

Genomic Databases

If you are connected to the Internet and have enabled ProtPlot to 'Access Web-DB', then clicking on a protein will popup a genomic database entry for that protein. The particular genomic database to use is selected in the Genomic-DB menu.

Djamel Medjahed, LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

ProtPlot Menus

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | Revision history | PRP file format | PDF documents | History of ProtPlot | Latest version

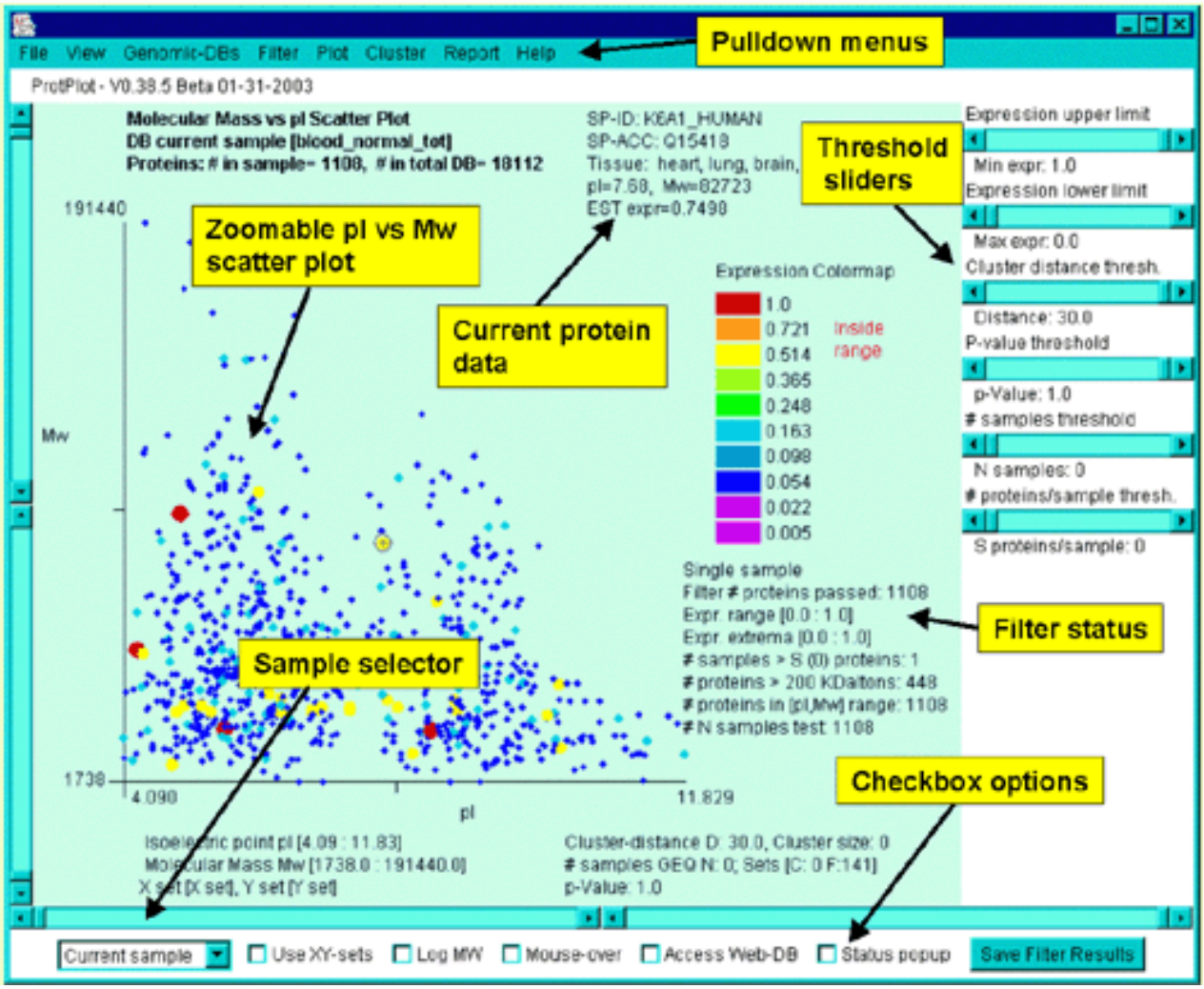
Menu Descriptions for ProtPlot

The user interacts with ProtPlot via various direct manipulation controls including:

- Clicking on a protein spot in the scatterplot
- Adjusting the Mw and pI sliders for the scatterplot
- Selecting commands from the pull-down menu bar at the top of the window

- Adjusting threshold sliders on the right
- Selecting samples in the pull-down menu in the lower left menu bar
- Toggling frequently used checkbox options in the lower menu bar

These controls are shown in the following figure:



Interrogating the database by clicking on a spot in the scatterplot

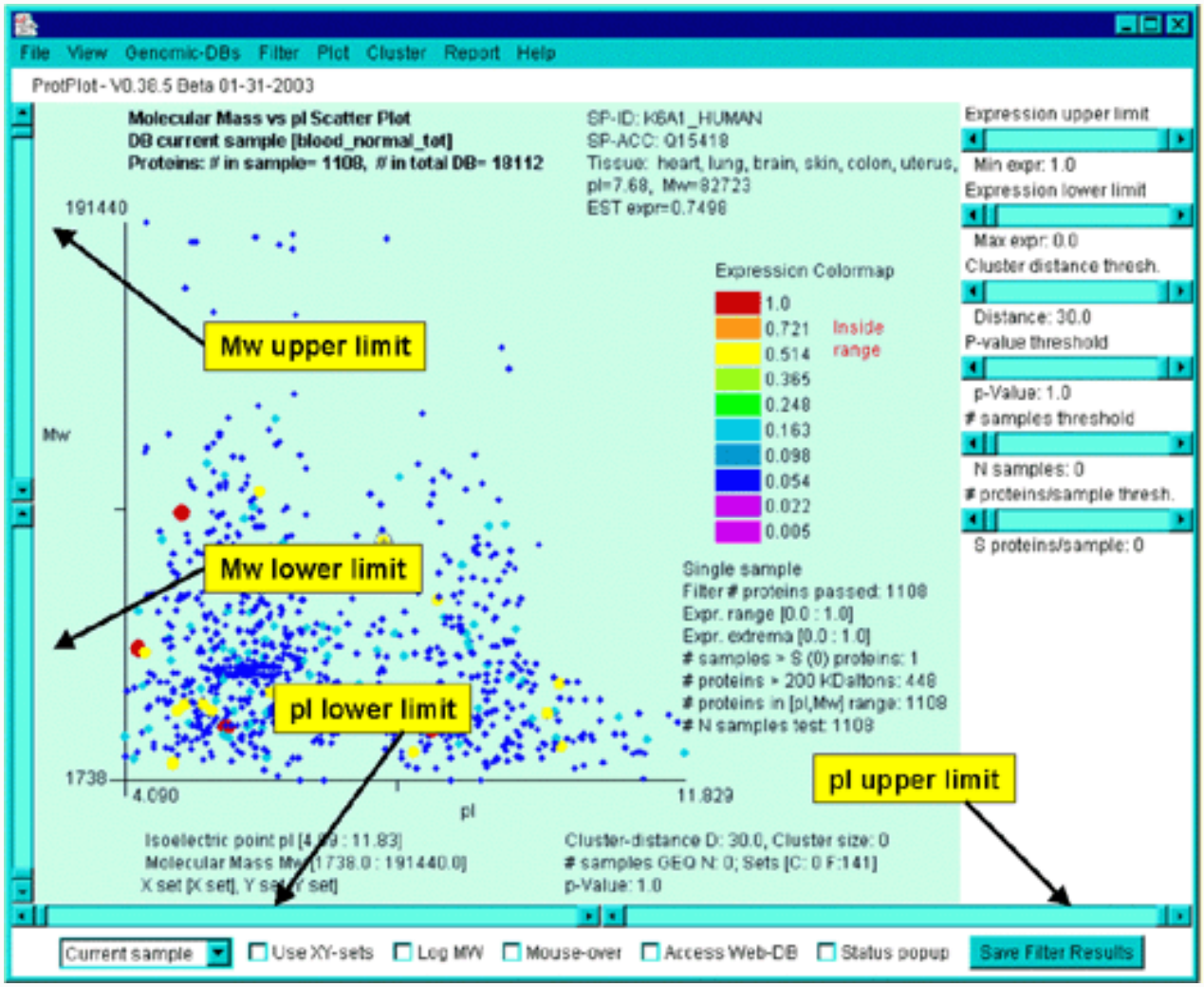
You may interrogate the database by clicking on a spot in the scatter plot. This will report data for that protein in the upper right part of the scatterplot window. The actual data reported depends on which display mode you have selected. All entries show the SP-ACC and SP-ID. Additional data is added depending on the display mode:

- Single sample: it will report expression data on only that sample.
- Mean samples (EP set): mean expression, std-dev expression, Coefficient of Variation (CV) of expression, number of samples.
- (X sample)/(Y sample) ratio of 2 samples: for both the X and Y sets it reports the expression and the ratio (X sample)/(Y sample).
- (Mean X-set)/(Mean Y-set) ratio of samples: for both the X and Y sets it reports the mean

expression, std-dev expression, CV of expression, number of samples. In addition, it reports the ratio computed as (Mean X-set)/(Mean Y-set).

Mw and pI upper and lower limits sliders

The upper and lower limits of Mw and pI may be controlled directly thorugh sliders on the left and right respectively as shown in the following figure.



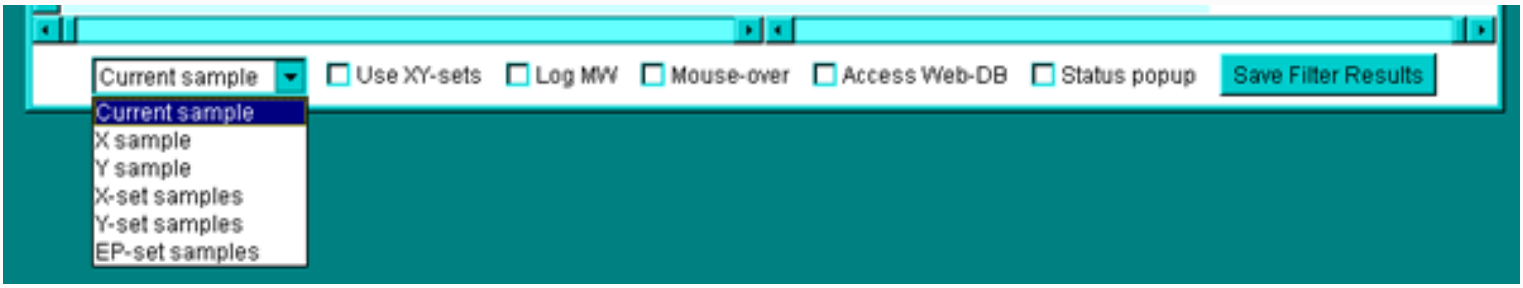
The continuous parameter threshold slider controls

Parameters may be defined interactively using direct manipulation sliders on the right side of the main window. These right slider controls include:

- Expression upper limit - this adjusts the upper limit of the expression data filter (while in expression mode)
- Expression lower limit - this adjusts the lower limit of the expression data filter (while in expression mode)
- Ratio upper limit - this adjusts the upper limit of the ratio (X/Y) data filter (while in ratio mode)
- Ratio lower limit - this adjusts the upper lower of the ratio (X/Y) data filter (while in ratio mode)

- Cluster distance threshold - this adjusts the D cluster distance parameter if clustering is active in "Mean EP display mode"
- P-value threshold - this adjusts the p-value parameter if the t-Test or other similar filtering is enabled
- # samplesthreshold - this adjusts the N samples parameter if the minimum or maximum number of samples filtering is enabled
- # proteins/sample threshold - this adjusts the S proteins/sample parameter

Samples menu and checkboxes in the lower command bar



These lower controls include:

- A **pull-down sample selector** to let you redefine the current sample, X sample, Y sample, X-set of samples, Y-set of samples, and EP (expression profile) list of samples.
- ☐ Use XY-sets - this toggles between the ratio (Mean X-set / Mean Y-set) of samples mode and the previous display mode
- ☑ Expression - this toggles the displaying the spots in the pseudo 2D-gel as colored expression values or simply location
- ☐ Low Mw - this toggles displaying the Mw scale as a log(Mw)
- ☐ Mouse-over - this toggles the mouse-over option. Holding the mouse over a spot will display its annotations
- ☐ Status popup - this toggles the popup status window on and off

Pull-down menu commands in the upper menu bar

The top pull-down menu selections include commands which invoke an action, checkbox options [CB] which are independent of one another, and radio button options [RB] which may only have one member active at a time. Menu entries followed by the ▷ symbol, indicate that the menu entry has a submenu. Selections prefaced with a '☑' and indicate '☐' indicate that the command is a checkbox that is enabled and disabled respectively. Checkbox menu items have a "[CB]" at the end of the command. Selections prefaced with a '◉' and indicate '○' indicate that the command is a multiple choice "radio button" that is enabled and disabled respectively, and that only one member of the group is allowed to be on at a time. Radio button menu items have a "[RB]" at the end of the command.

There are a number of pull-down menus in ProtPlot including:

1. File - select samples, save the state and quit
2. View - select viewing options
3. Genomic-DBs - enable access to popup Web genomic databases
4. Filter - select protein data filter options
5. Plot - select primary plot mode and plotting options
6. Cluster - select cluster distance metrics and perform clustering
7. Report - generate popup reports
8. Help - the help menu - local and server based

File menu

This provides commands to select samples for the current, X, Y, X- and Y-sets, and EP-set of samples. It invokes a popup sample chooser. Find a protein by SwissProt Accession number (SP-ACC) or SwissProt ID (SP-ID). Save and restore the data-mining state. Update ProtPlot and its data from the Web server.

- Select samples
  - Select Current PRP sample - select sample for expression mode
  - Select X PRP sample - select X sample for use in X/Y ratio mode. If use sample sets mode is on, then define the X set.
  - Select Y PRP sample - select Y sample for use in X/Y ratio mode. If use sample sets mode is on, then define the Y set.
  - Select Expression List of samples - the subset of samples is used in Expression Profile Plots and in clustering
    ---
  - List sample assignments - for current, X, Y, and X set, Y set and EP list
  - Assign X set name - to assign a new condition name to the X samples set
  - Assign Y set name - to assign a new condition name to the Y samples set
- Find proteins
  - Find protein by SwissProt Acc - if the SP-ACC is in the database
  - Find protein by SwissProt ID - if the SP-ID is in the database
- State
  - Reset State - set the state to the defaults
  - Protein Sets
    - Open Cluster protein set file from .set file - change the saved Cluster set to one you specify contained in a Set/*.set file.
    - Open Filter protein set file from .set file - change the saved Filter set to one you specify contained in a Set/*.set file
    - Save Cluster protein set file to .set file - save the saved Cluster protein set to one you specify in a Set/*.set file.
    - Save Filter protein set file to .set file - save the saved Filter protein set to one you specify in a Set/*.set file
      ---
    - Set Filter set to Intersection of two .set files - set the saved Filter protein set to the

intersection of previously saved protein set files.
        - Set Filter set to Union of two .set files - set the saved Filter protein set to the Union of previously saved protein set files.
        - Set Filter set to Difference of two .set files - set the saved Filter protein set to the difference of previously saved protein set files (proteins in first file but not in the second).
    - Open State file - change the state by reading the previously saved startup file and the saved cluster and filter sets.
    - Save State - save the state and saved protein sets as State/*.prpstate and Set/*.set files and continue the program
    - Save As State - same as "Save State", but prompt for a different state file name to use
    - Save State and Close - save the state and exit the program
- Update
    - Update ProtPlot program from TMAP ProtPlot server - get latest version
    - Update PRP data files from TMAP ProtPlot server - get latest version
      ---
    - Use new PRP data file with this working database - lets you use other in the current running database. The samples may derived from CGAP or from other sources (see .prp data file format)
    - Add new .prp file to your local PRP directory - lets you add other samples to your local PRP database. The sample may be derived from CGAP or from other sources (see .prp data file format).
- Close - exit the program, do not save state

View menu

Specify various scatterplot viewing options.

- Show status window
- Hide status window

- ☑ Use pseudocolor for expression else grayscale [CB]
- ☐ Use log of MW [CB]
- ☐ Enable mouse-over to see data [CB]
- ☑ Show expression data else just position [CB]
- ☑ Use radius proportional to expression else constant [CB]
- ☐ Show clustered proteins by black boxes [CB]
- ☐ Show all X and Y samples if doing X/Y filtering [CB] - this works with either single samples or sets of sample. Proteins missing in the X sample are shown as a gray '+' and th0se missing from the Y sample are shown as gray boxes.
- ☑ Auto-update reports if state changes [CB]

Genomic-DBs menu

This connects ProtPlot with various Genomic databases. If enabled, it will pop up a Web browser for the selected protein on the specified Genomic database.

- ☐ Access a genomic Web server if you click on a spot [CB]

- ◉ Use Swiss-Prot Web server [RB] - data base to use when select a spot
- ○ Use PIR ProClass Web server [RB] - data base to use when select a spot

Filter menu

Select the data filter options. If you change an option or adjust a parameter slider, it will re-run the data filter. See table of filters as a function of display mode.

- ☑ Filter proteins > 200K Daltons [CB] - enable to use all proteins, otherwise just use those below 200K Daltons
- ☐ Filter data by protein families [CB] [FUTURE]
- ☐ Filter data by tissue types [CB] - to filter by a subset of tissues
- ☑ Filter data by expression (or ratio) range [CB]
- ☑ Filter by (X/Y) outside [minRatio:MaxRatio] sliders [CB] - only for ratio mode else filter by inside ratio range
- ☐ Filter by X & Y sets t-test results [CB] - do t-test if using X and Y sets
- ☐ Filter by X & Y sets Kolmogorov-Smirnov test results [CB] - do test if using X and Y sets
- ☐ Filter by X & Y sets missing and > N samples [CB] - present in X or Y set with > N samples using the slider 'N' and is completely missing in the other set
  ----
- ◉ Filter by requiring at least N samples [RB] - using the slider 'N'
- ○ Filter by requiring ayt most N samples [RB] - using the slider 'N'

- ☐ Filter by Saved Clustered proteins [CB] - the clustered proteins set may be created when doing clustering by pressing the "Save Cluster Results" when clustering
- ☐ Filter by Saved Filtered proteins [CB] - the filter proteins set may be created by pressing the "Save Filtered Results"

Plot menu

Specify the scatterplot display. Switching from expression data to ratio data changes the expression range sliders to ratio range sliders (and vice-versa). The Expression Profile plot is enabled from this menu. See table of filters for a discussion on display modes.

- ◉ Show current sample expression data [RB]
- ○ Show mean EP-set samples expression data [RB]
- ○ Show X-sample/Y-sample ratio (Red=X + Green=Y) data [RB]
- ○ Show X-sample/Y-sample ratio range color map [RB]
- ○ Show X-Set-samples/Y-Set-samples ratio range color map [RB]

- ☑ Display (X vs Y) else (Mw vs pI) scatterplot - if ratio mode [CB] - if in X/Y or X-set/Y-set ratio display mode, then display the X-Y scatterplot instead of the Mw vs. pI scatterplot.
- Save scatterplot as GIF file - saves it in the /Report folder

---

- ☐ Popup Expression Profile plots [CB] - you may adjust the zoom by clicking on the 1X button. Clicking on a sample bar shows what the sample is and its value. Selecting another protein in the scatterplot will change the EP plot.
- Popup scrollable list of EP plots for proteins passing the Filter

Cluster menu

Initiate clustering. Specify the distance metric. When clustering is enabled, the Cluster Distance Slider (D) is active. After you have defined an initial cluster, you may popup up a scrollable list of EP plots for the clustered proteins. These are sorted by the similarity order in the cluster report

- Cluster similar proteins to current protein by EP profile - you must be in 'mean EP-set samples' display mode

---

- ◉ Use weighted distance metric [RB] - this adjusts the distance taking missing sample values into account (Jain and Dubes)
- ○ Use average all data distance metric [RB] [FUTURE]
- ○ Use Fisher-clustering low values distance metric [RB] [FUTURE]

Report menu

Generate various tab-delimited reports that may be saved or cut-and-pasted into MS Excel. See table of reports as a function of display mode.

- Report statistics of proteins passing the filter - the report generated depends on the plot mode and filter options selected
- Report Expression Profiles of proteins passing the filter
- Report X&Y sets statistics of proteins passing missing protein filter
- Report EP set statistics of proteins passing missing protein filter

---

- Report list of samples in the current the expression profile
- Report list of all sample assignments
- Report list of # proteins/sample
- Report ProtPlot state

- ☑ Sort report assending order [CB]

Help menu

Help is available on the Web site and make be invoked from ProtPlot when it is running by selecting a topic.

- TMAP Home page
- Introduction
- Using ProtPlot
- Menu descriptions
- ProtPlot data format
- ProtPlot history
- Estimating expression
- PDF documents
- Disclaimer

- Introduction to ProtPlot (PDF)
- About

Djamel Medjahed, LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

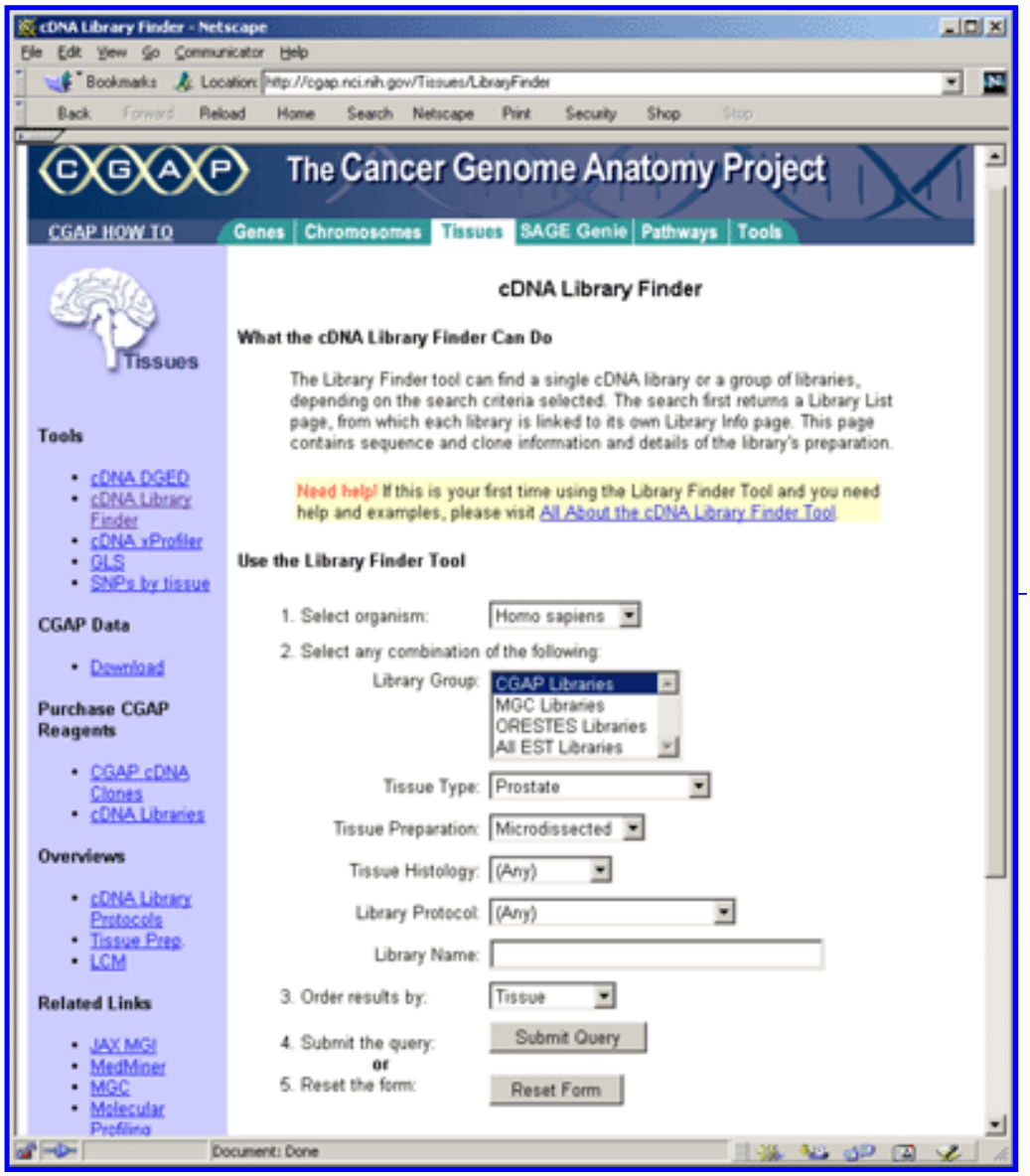
Computing the Estimated EST Expression

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | Revision history | PRP file format | PDF documents | History of ProtPlot | Latest version

**** This page is under construction *****

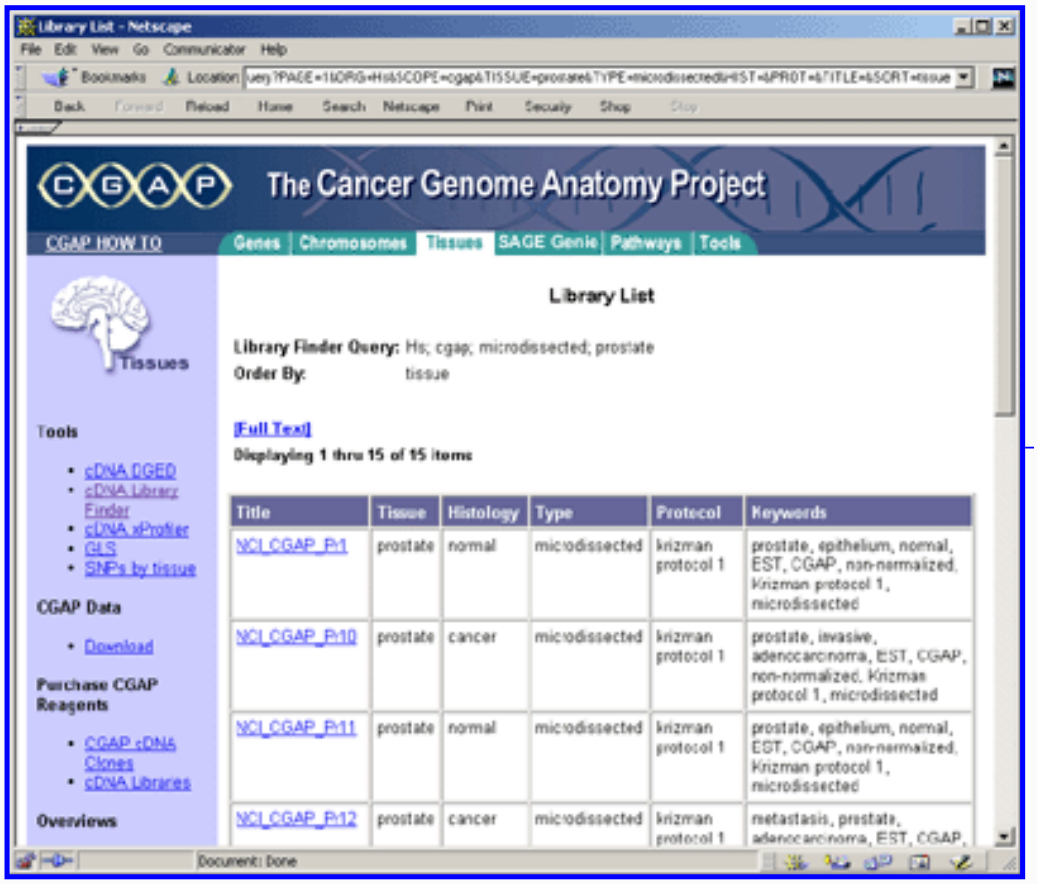
The CGAP tissue and histologic database

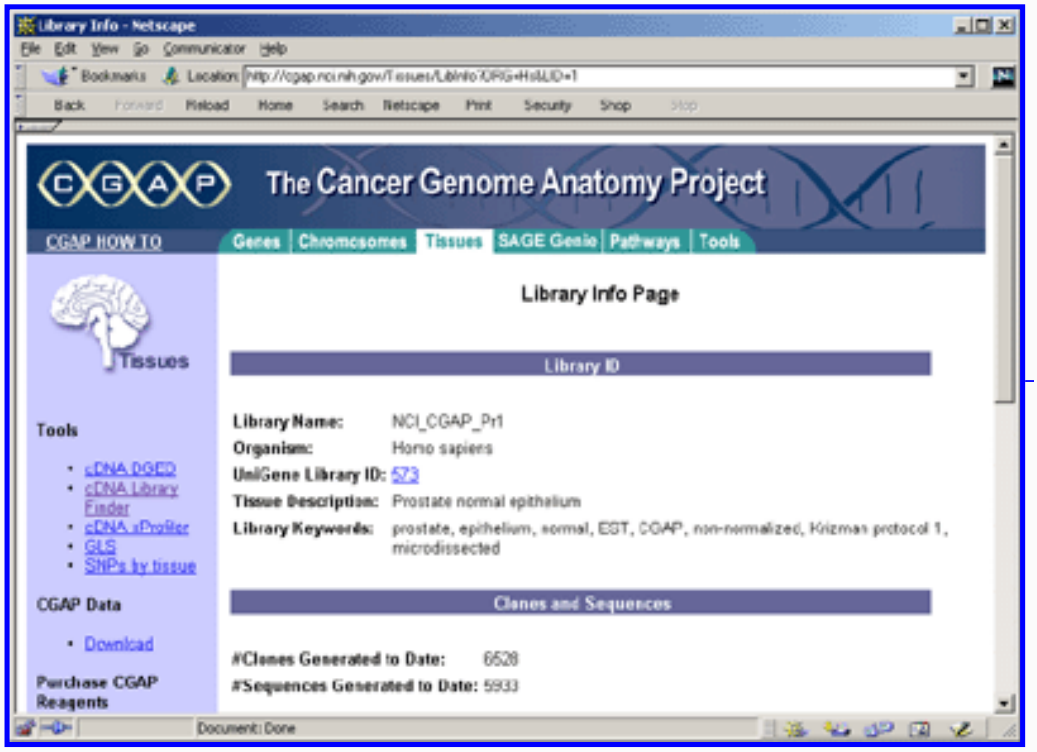
For each tissue, the NCI Cancer Genome Anatomy Program (CGAP database) may be queried by possible histological state, source, extraction and cloning method. In the initial query on the CGAP Web site, selecting the option "ANY" for all fields provides an initial overview of the available libraries. The more restrictive a search, the fewer the number of libraries that are selected. Within each library, transcripts are listed along with the number of times they were detected after a fixed number of PCR cycles.



As we were primarily interested in computing protein maps, using UniGene we extracted gene symbols associated with those CGAP EST's that were clustered to a gene of known function. To restate this: the CGAP Web site contains library specific expression data and the UniGene site contains the gene cluster symbol correspondence.

Finally, the Expasy SwissProt/trEMBL database contains gene symbols and protein sequence data. From this, one can compute the pI and Mw.


Library List - Netscape
The Cancer Genome Anatomy Project
CGAP HOW TO
Genes
Chromosomes
Tissues
SAGE Genie
Pathways
Tools
Tissues
Tools
cDNA DGED
cDNA Library Finder
cDNA xProfiler
GLS
SNPs by tissue
CGAP Data
Download
Purchase CGAP Reagents
CGAP cDNA Clones
cDNA Libraries
Overviews
Library List
Library Finder Query: Hs; cgap; microdissected; prostate
Order By: tissue
[Full Text]
Displaying 1 thru 15 of 15 items
Title
Tissue
Histology
Type
Protocol
Keywords
NCI_CGAP_Pr1
prostate
normal
microdissected
krizman protocol 1
prostate, epithelium, normal, EST, CGAP, non-normalized, Krizman protocol 1, microdissected
NCI_CGAP_Pr10
prostate
cancer
microdissected
krizman protocol 1
prostate, invasive, adenocarcinoma, EST, CGAP, non-normalized, Krizman protocol 1, microdissected
NCI_CGAP_Pr11
prostate
normal
microdissected
krizman protocol 1
prostate, epithelium, normal, EST, CGAP, non-normalized, Krizman protocol 1, microdissected
NCI_CGAP_Pr12
prostate
cancer
microdissected
krizman protocol 1
metastasis, prostate, adenocarcinoma, EST, CGAP,
Document: Done


Library Info - Netscape
The Cancer Genome Anatomy Project
CGAP HOW TO
Genes
Chromosomes
Tissues
SAGE Genie
Pathways
Tools
Tissues
Tools
cDNA DGED
cDNA Library Finder
cDNA xProfiler
GLS
SNPs by tissue
CGAP Data
Download
Purchase CGAP Reagents
Library Info Page
Library ID
Library Name: NCI_CGAP_Pr1
Organism: Homo sapiens
UniGene Library ID: 573
Tissue Description: Prostate normal epithelium
Library Keywords: prostate, epithelium, normal, EST, CGAP, non-normalized, Krizman protocol 1, microdissected
Clones and Sequences
#Clones Generated to Date: 6528
#Sequences Generated to Date: 5933
Document: Done

1. Mapping Gene symbols between CGAP, UniGene and SwissProt databases

A Perl script was outputs these gene symbols from the CGAP=UniGene derived data. This is cross=reference agains the Expasy SwissProt/trEMBL homosapiens data set to produce a list of corresponding SwissProt accession numbers (SP-ACC). This list can then be input to the Expasy pI/Mw tool server to produced tab-delimited data containing the pI (isoelectric focusing point), Molecular mass (Mw), and SwissProt ID (SP-ID) for the mature, unmodified proteins [Medjahed03a]. The following summarizes the steps in mapping the annotation mapping.

1. Remove all EST and empty gene symbol entries
2. Sort by the Hs. UniGene identifiers
3. Lookup the gene symbols in the sorted UniGene data
4. Using the gene symbols, lookup (SP-ACC,SP-ID,pI,Mw) on the Expasy.org Web site

2. Computing the estimated EST expression

We then needed to compute the estimated EST expression. In the case of a single library, this information was computed from the expression-detection counts. The number of hits for each CGAP EST was first divided by the sum total of sequences within that library to provide a relative expression for each transcript.

Then, the results were renormalized by dividing relative expression levels by the maximum relative expression level so that the maximum expression was normalized to 1.0. Expression values are > 0.0 (least abundent) and less than or equal to 1.0 (most abundent).

A tissue search may find several libraries fulfilling the requirements of the initial query. Therefore, to improve the signal-to-noise ratio, the search results were pooled to generate a non-redundant list of entries. This leads to a more comprehensive expression map for that tissue corresponding to that histological state.

1. Pool and add CGAP libraries coresponding to the same tissue and histological state
2. Compute the relative EST frequencies from this pooled data
3. Compute the maximum relative EST frequency (i.e., MaxESTexpr)
4. Merge the (SP-ACC,SP-ID,pI,Mw) with the (SP-ID,MaxESTexpr) data to generate (SP-ACC,SP-ID,pI,Mw,MaxESTexpr) data used to for the ProtPlot master protein index.

3. Generating the ProtPlot .prp files

The resulting data is a tab-delimited '.prp' formatted file that contains expression levels ranging from 0.0 (undetected) to 1.0 (most abundant). The following sequence of operations is performed on each (tissue, histological state) to create a ProtPlot sample. Each ProtPlot sample is saved in a tab-delimited ".prp" formated file containing the (SP-ACC,SP-ID,pI,Mw,MaxESTexpr) data.

Additional details on these methods are available in ([Medjahed02], [Medjahed03a], [Medjahed03b]).

Djamel Medjahed, LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

Downloads for The MicroArray Explorer Project

Program downloads | Update programs | Update jar file | Installation hints

The table below lists the various types of downloads: program installer, source code file, jar file, and information on installing the programs.

Types of download files available

You may download program installers for your particular computer for both ProtPlot. Click on the entries to download the installer or files.

Access of ProtPlot from the Web server

Program	Installer Version	Update Program Jar Version	Program installers	Jar file(s)	Source	PRP data
ProtPlot	0.39.6	0.39.6	ProtPlot	ProtPlot.jar	N.A. yet	PRP data

Upgrading the ProtPlot JAR program file after the initial installation

If you want to upgrade your installation to the latest JAR files, simply download the JAR files and save them wherever you have installed the programs replacing the previous jar files. For example, in a typical Windows OS installation, the **ProtPlot.jar** is installed in **C:\Program Files\ProtPlot** folder. Alternatively, an easier way is to update the Jar file when running ProtPlot as described in the next paragraph.

Updating the ProtPlot JAR file from the running programs

You can use the new "Update ProtPlot" command in the Files menu to quickly download and install just the JAR file. This first prompts you to verify that you want to update your program. Then it will: (1) backup the current ProtPlot.jar file as ProtPlot.jar.bkup; (2) copy the latest ProtPlot.jar file from the TMAP Web site and replace your ProtPlot.jar file in your installation directory. Then when you restart ProtPlot, it will use the new version of the program.

Hints on downloading the stand-alone ProtPlot program

You may freely download and install the current stable release of the stand-alone version of the MAExplorer program. You are free to use or redistribute ProtPlot (see disclaimer). We also include a subset of CGAP derived data. in the PRP subdirectory which is loaded by ProtPlot when you run it.

After you have first installed ProtPlot, run it and do an update of the latest jar file by going into the (File

menu | Update | Update ProtPlot from TMAP ProtPlot server). This will get the latest version of the software. Note: you may do this anytime so you do not have to do the full re-installation to update the program data.

If you want to update your copy of the PRP database, then go into the (File menu | Update | Update PRP data files from TMAP ProtPlot server). This will get the latest version of the database. Note: you may do this anytime. [This will not be updated for a while, but we expect to automate its update and will then make it available.]

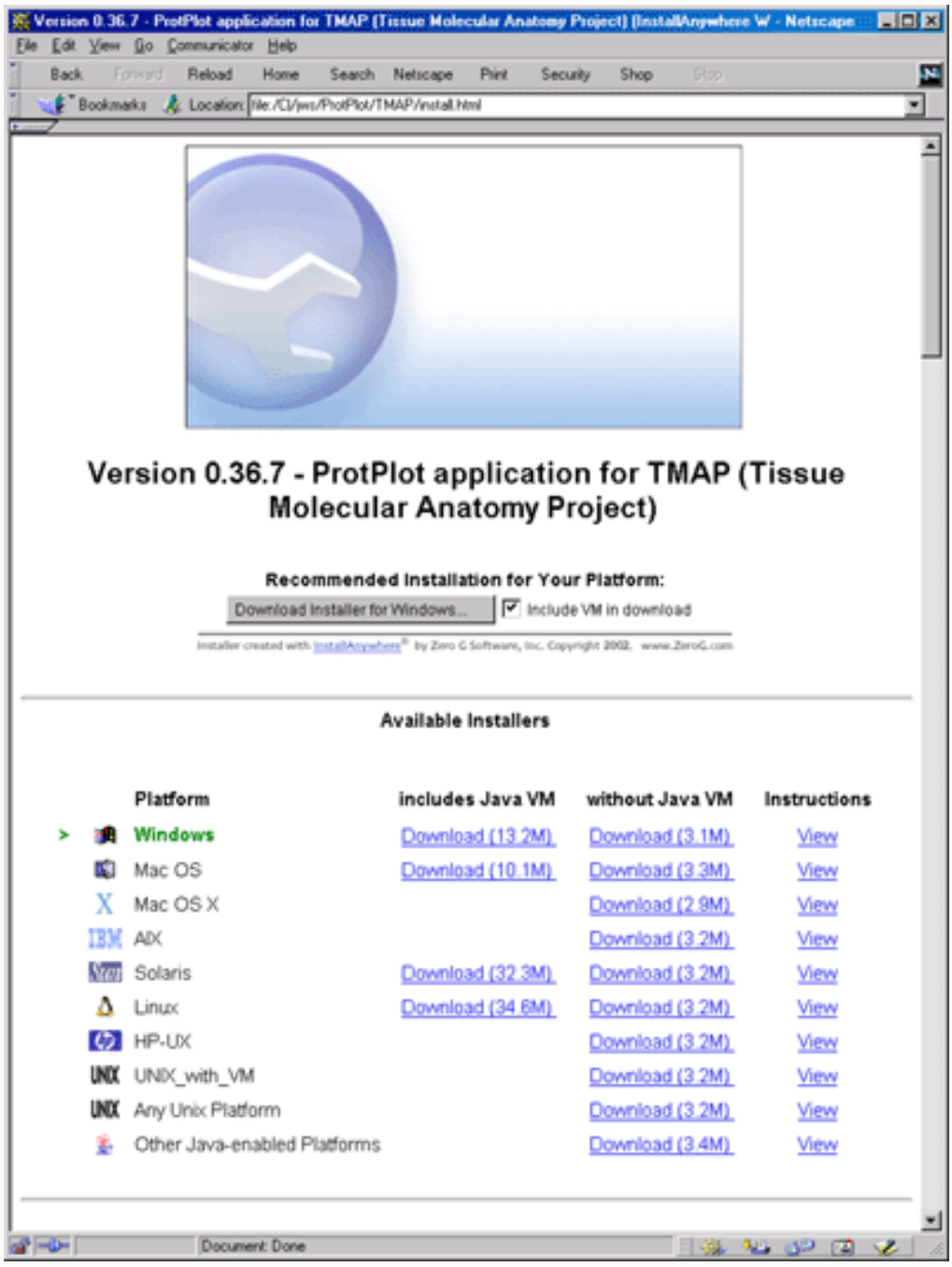


Figure. Web page showing options for installing ProtPlot as a stand-alone application. Installers are available for Windows95/98/NT/2000/XP, MacOS-8/9, MacOS-X, Solaris, HP-UX, Linux, Unix, and other Java enabled platforms. [Click on the figure to see a high resolution version.]

Distribution contents

1. We recommend **including** the Java Virtual Machine (JVM) for a more robust installation.
   This will not affect any of your other Java applications or Web browsers as it is used only with ProtPlot.
2. The distribution includes:
   - The ProtPlot Java stand-alone application,
   - A set of PRP samples data derived from http://cgap.nci.nih.gov/ DB.
   - Support files for your operating system - possibly including the JVM which you may optionally download.

1. Procedure for downloading and installing ProtPlot on your computer

1. Click here to select the current installer for your operating system. This Web page allows you to select the operating system you are using. If you have problems downloading the installer with Netscape 4.7x or later, then try Internet Explorer 5.0. It could be a Mime/type problem with your browser setup.

2. You start the download process when you click on the installer for your computer platform. (You may alternatively use the default installer discussed below.) Follow the directions it provides as you download the installer. It also provides instructions in the "View" hyperlink adjacent to the operating system you selected that tells you what to do after you finished the download. Part of the installation consists of telling the installer where you want to 1) put the executable installer (a temporary directory where you have lots of room is a good choice), and 2) the "installation" directory where you will typically leave the distribution after the installer unpacks it.

We use the commercial InstallAnywhere(TM) program to create the installers. It provides installers for:

- Windows 95/98/NT/2000/XP
- Mac OS (OS8 and OS9, OS-X)
- Solaris
- HP-UX
- Linux
- Unix
- Other Java enabled platforms

Other systems will be added as installers become available through InstallAnywhere (www.ZeroG.com).

1.1 The Default Installer

Alternatively, you can use the default installer that is selected for your computer. ***If you want to control where the files are saved on your computer, then use the explicit installer for your particular platform described above***.

The default installer will put the installer executable in a fixed directory and the installed ProtPlot files in another fixed directory.

- For Windows, the installer will go into `C:\InstallAnywhere_Installers\` and the program files into `C:\Program Files\ProtPlot\`.

- For Unix systems, it will put them at `$HOME/InstallAnywhere_Installers` and the program files into

`$HOME/ProtPlot/`.

- For MacOS, it will put them on the desktop.

1.2 Installation Notes

Currently, the Windows, Linux, MacOS-X installers are robust. We have had mixed success with MacOS-8/9 Solaris.

Note that the installers (where possible) will include a copy of a recent Java Virtual Machine (JVM) from InstallAnywhere(TM) to make running ProtPlot on your computer more robust. This is used locally and *only* affects the running of ProtPlot. It will *not* affect any other Java applications on your computer. In the case of Mac OS, if you have an older version of the MRJ JVM, it will ask you if you want to upgrade to the newer version (MRJ-2.4.5) - however you do not have to unless you want to.

1.3 Downloading just the ProtPlot.jar file after initial install

If you have previously done an installation. you may avoid a complete re-installation download by getting just the latest Java ProtPlot.jar file. You should replace the old version of this file on your system with the one you are downloading. This will work if the new ProtPlot.jar file does not depend on any new changes in the data files files (which generally the case - try it and see what happens). If necessary, try doing a complete re-installation where you uninstall the old version first.

1.4 Starting ProtPlot Using a ".prpstate" startup state file

If you are on Windows 95/98/NT/2000/XP system, simply click on the particular .prpstate file you want to use. These files reside in the state folder where you have installed ProtPlot.Hint: you might put a short-cut to the *installation-directory* on your desk-top to make it more convenient to find the .prpstate files.

If you are on a Macintosh system, then start ProtPlot and then run the startup .prpstate file you want by going to the (File menu | State | Change state) command. Then browse your disk and then open up the startup state file of interest.

If you are on a Unix system, then you supply the state startup file explicitly in the command line. You might consider adding the "installation" directory to your UNIX `$PATH` or `$path` variable to have UNIX automatically find the executable binary.

```
cd installation-directory/
ProtPlot.bin myLastSession.prpstate
```

1.5 Problems installing ProtPlot on some operating systems

1. The MacOS installer is available, but may not work with older versions of MacOS. In addition, there may be problems if file names are longer than 32 characters. For now, the solution is to use short file names. There may also be problems if your data files have embedded carriage returns in addition to line feeds. For now, the solution is to strip the CRs out of the data file.

2. On Solaris, and possibly other Unix systems, you may have problems with the stack limits. Do a "man limit" to read about the command for your particular Unix shell. We have found that the following seems to work. For the Unix C-shell (csh), add the following to your .cshrc startup file.

```
limit stacksize unlimited
```

In addition, we have set the default stack size that ProtPlot uses to 96Mbytes. If your computer has less physical memory, it will page. You may also increase this number as well if you have more memory and want to use it. The solution is to edit the ProtPlot.lax file found where you installed ProtPlot. Change the two instances of memory allocation from 96000000 to a smaller number that is less than your actual memory size.

3. On Solaris, if you download the version with the JVM, unless your Solaris system has been updated recently, it may not be able to find the libCrun.so.xxx version required by the JVM. Try downloading the non-JVM version or update your Solaris system.
4. If you have problems with the Sun installer, you may need to update your Solaris OS system patch set. It is not a single patch. It is the latest Recommended Patch Cluster from Sun. We **STRONGLY** recommend having your SysAdmin do this for you if you have not done this before. Point your Web browser to:

```
http://sunsolve.Sun.COM/pub-cgi/show.pl?target=patches/patch-access
```

and choose the appropriate patch set for the version of Solaris (2.6, 7, or 8) that you are running. Do not choose any of the x86 versions unless you are running Solaris x86. Click on either the Download HTTP option or Download FTP option, and click the GO button to download the patch set.

1.6 FAQ of problems using ProtPlot on Mac OS8/9 users

Q: How many characters can I use in array names for data read by ProtPlot?

A: For MacOS-X, with 256 character file names, this is not a problem. For MacOS 8 and 9 with 32 character file names it may be a problem. Because ProtPlot uses file extensions (eg. ".prpstate"), you are currently limited to 25 characters or less. We will be modifying the system to remove this limit.

Q: How do I start ProtPlot on my data automatically by double-clicking a **protplot.prpstate** file on my Mac.

A: There is no easy way to do this at this time. Use the (File menu | State | Change state) command to popup a browser to specify the new .prpstate file.

1.7 Sun Solaris (or other Unix system) Memory Problems

We have on occasion seen the following types of memory errors. This discusses how to handle them.

ProtPlot Stack size Memory Error on Sun Solaris

Running ProtPlot on a Solaris (or other Unix system) may produce this error:

```
% ProtPlot

Stack size of 97664 Kb exceeds current limit of 8192 Kb.
(Stack sizes are rounded up to a multiple of the system page size.)
See limit(1) to increase the stack size limit.
```

If the Sun (under Solaris) is slow in loading ProtPlot or has memory errors (shown above) one should first see what the memory limits are set to on your machine using the "limit" command. If they are too small they should be increased or set to "unlimited" (see in 1.5 above

1.8 ProtPlot LAX file

If the problems persist, one might have to edit the ProtPlot.lax file found in the ProtPlot directory (see example below). The default memory settings in the ProtPlot.lax file (found in the installation directory) should be no larger than the total memory of the machine or paging problems will occur. For instance, if you have 64Mb of memory in your Sun, edit the **"lax.nl.java.option.native.stack.size.max"** and **"lax.nl.java.option.java.heap.size.max"** options to be under 64Mb. You can use any text editor to do this. More memory may be needed to be installed on your Sun to run ProtPlot with very large datasets.

Default Lax settings
The Lax file is a startup file generated by InstallAnywhere when we packaged ProtPlot. It is used when ProtPlot starts up on your computer. We currently set the memory limits to 96Mbytes. If you have more memory, you can edit the Lax file to have it use more memory.

```
#   LAX.NL.JAVA.OPTION.JAVA.HEAP.SIZE.MAX
#   -------------------------------------
lax.nl.java.option.java.heap.size.max=96000000

#   LAX.NL.JAVA.OPTION.NATIVE.STACK.SIZE.MAX
#   ----------------------------------------
lax.nl.java.option.native.stack.size.max=96000000
```

NOTE: this software is undergoing Beta-testing.

Djamel Medjahed, in the LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

Format of .prp ProtPlot data files

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | Revision history | PRP file format | PDF documents | History of ProtPlot | Latest version

The ProtPlot data is contained in a set of tissue- and histologic-specific .prp files. The set of .prp files constituting the ProtPlot database is included when you download ProtPlot. A .prp file is named using the following convention:

```
   {tissue name}_{histologic state}_tot.prp
where:
```

```
{tissue name} could be: brain, prostate, or pancreas, etc. See the
         database file 'tissueNamesFile.txt' for a list of tissue names.
{histologic state} can be a disagnosis category: normal, precancer, or cancer.
```

The .prp format has the following tab-delimited format (without the quotes added here for clarity). The first row is the tab-delimited list of field names followed by the tab-delimited corresponding data. The order of the columns is not important. Additional columns may be included in the files, but are ignored if the key words are different from any of the keywords in the following list.

The data in the initial startup *must have* the following fields: (pI, Mw, SP-ID, SP-ACC, expression, tissue). On subsequent addition of data using the (File menu | Use new PRP data file with this working database), it only requires (SP-ID, expression) since it will get the rest of the missing data from Master Protein Index entry.

```
'pI'    'Molecular Mass' 'SP-ACC' 'SP-ID          'MaxESTexpr' 'Tissue' 'Family'
6.74     31544            O00108   AQP3_HUMAN      0.044334972  30       1
5.05     44106            P08727   K1CS_HUMAN      0.152709348  30       1
9.58      9330            P42677   RS27_HUMAN      0.004975124  30       1
                  . . .
```

where the following lists the files (case-independent) and their alternate names:

1. 'pI' is the estimated isoelectric point of the protein. It is a decimal number (e.g., 4.61). Alternate names: (pI, pIe)
2. 'Molecular Mass' is the molecular mass in Daltons (*not KiloDatons!*). Alternate names: (Mw, Molecular Mass)
3. 'SP-ACC' is the SwissProt Accession number. Alternate names: (SP_ACC, SP-ACC, SPACC, "SwissProt Acc")
4. 'SP-ID' is the SwissProt ID. Alternate names: (SP_ID, SP-ID, SPID, "SwissProt ID")
5. 'GB-ID' is the GenBank ID (optional field). Alternate names: (GB_ID, GB-ID, GBID, "GenBank ID")
6. 'MaxESTexpr' is the derived expression in the range of 0.0 to 1.0. Missing proteins are not entered. (NOTE: in the master protein index computed across all samples, 0.0 indicates there is no protein for a particular tissue when refering to its expression as part of the expression profile.) See discussion on how MaxESTexpr is computed for ProtPlot. Alternate names: (GB_ID, GB-ID, GBIDMaxESTexpr, "Max EST expr", estExpr, expr, expression).
7. 'Tissue' specifies the tissue(s) that constitute the sample. Alternate names: (Tissue, "Tissue Name"). It is either:
    1. a tissue number from the 'tissueNamesFile.txt' file, or
    2. a hexadecimal bit pattern of several tissue numbers (eg. a mixture of tissues), then it is represented by the sum of (2**tissue(i)) for a set of n tissues.
8. 'Family' is the protein families that the protein belongs to. It is a hexadecimal bit pattern of several protein family numbers from the 'familyNamesFile.txt' file. The mixture of families is sum of (2**family(i)) for a set of n families.. Alternate names: (Family, "Family Name", "Protein Family", "Protein Family Name"). [FUTURE] (This data is optional).

Djamel Medjahed, LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

PDF documents for ProtPlot

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | Revision history | PRP file format | PDF documents | History of ProtPlot | Latest version

This lists some PDF documents describing ProtPlot and its usage.

- TMAP web site documentation as a single PDF file.
- Medjahed D, Luke BT, Tontesh TS, Smythers GW, Munroe DJ, Lemkin PF, TMAP poster, Swiss Proteomics Meeting, Geneva, Dec, 2002. (PDF)
- Lemkin PF, Medjahed D (2003) ProtPlot - A Tissue Molecular Anatomy Program Java-based Data Mining Tool: Screen Shots. This is a series of screen shots illustrating the various data mining capabilities. (PDF)
- Medjahed D, Smythers GW, Powell DA, Stephens RM, Lemkin PF, Munroe DJ, VIRTUAL2D: A Web-accessible predictive database for proteomics analysis, Proteomics, 2003, **3**, 129-138. The Web site is http://www.lecb.ncifcrf.gov/~medjahed/VIRTUAL2D/.
- Medjahed D, Luke BT, Tontesh TS, Smythers GW, Munroe DJ, Lemkin PF, "TMAP" (Tissue Molecular Anatomy Project), an expression database for comparative cancer proteomics. Proteomics, 2003, (in press, June) (PDF).

Djamel Medjahed, LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

History ProtPlot

TMAP home | Introduction | Using ProtPlot | Menus | Estimating expression | Download | Revision history | PRP file format | PDF documents | History of ProtPlot | Latest version

The new version of ProtPlot as part of the Tissue Molecular Anatomy Project was derived from an earlier version of ProtPlot is discussed in a paper (Medjahed D, 2003a). A poster was given at the 2002 Swiss Proteomics Meeting in Geneva on TMAP. This has been submitted to the proceedings of the meeting (Medjahed D, 2003b).

This version of ProtPlot is run as a stand-alone Java program. After ProtPlot is downloaded and installed from the Web server at http://TMAP.sourceforge.net/, the user starts up the program the default startup icon or clicks on a specific .prpstate file to start it on a previous data mining session. You may have several copies running simultaneously if you have enough memory.

In the future, ProtPlot will be offered as open source software under the Mozilla Public License (version 1.1) subject to the limitations noted in the accompanying LEGAL file.

This work was derived from MAExplorer (http://maexplorer.sourceforge.net/) under the Mozilla 1.1 Open Source Public License by Peter Lemkin of the National Cancer Institute.

See disclaimer for more information on its usage.

Djamel Medjahed, LMT, SAIC-Frederick
Peter Lemkin, LECB, NCI-Frederick

Revised: 08-26-2004

TMAP/ProtPlot Revision History

This describes the revision history of the ProtPlot program of released or soon to be released versions. Generally, only the most recent versions are kept on the Web site (see Version).

- **V.0.39.7 08-26-2004:** Original release modified for the Sourceforge tmap.sourceforge.net server.
- **V.0.39.7 08-21-2003:** Original release for development server.

TMAPjarVersion.txt

```
There are two versions of the ProtPlot program is available on
the TMAP server server

     http://TMAP.sourceforge.net/

The first is the installer version that you download from the Web
server.  The second is the possibly latest version that you update by
having ProtPlot copy the ProtPlot.jar file from the Web server using
the (File | Update | Update ProtPlot program) menu command.
After you do the update, you need to restart ProtPlot to use it.

Until futher notice, TMAP is to be considered beta-level
code until it is officially released. This means that there may be
some functionality not fully implemented, that works incorrectly,
or that has changed. New commands and functionality are in the
```

process of being added. Please report problems and suggestions to us.

1. Full download installation TMAP version 0.40.1
Revised: 08-26-2004

2. Update TMAP program (TMAP.jar) version 0.40.1
Revised: 08-26-2004