

User Manual for “ProteinDecision”

[I. Before using this tool](#)

[1. Pre-regulating the data](#)

[2. System requirements](#)

[II. Installation guide](#)

[III. How to use this tool](#)

[1. Input the file](#)

[2. Select the peaks](#)

[3. Set Parameters](#)

[4. Run “ProteinDecision”](#)

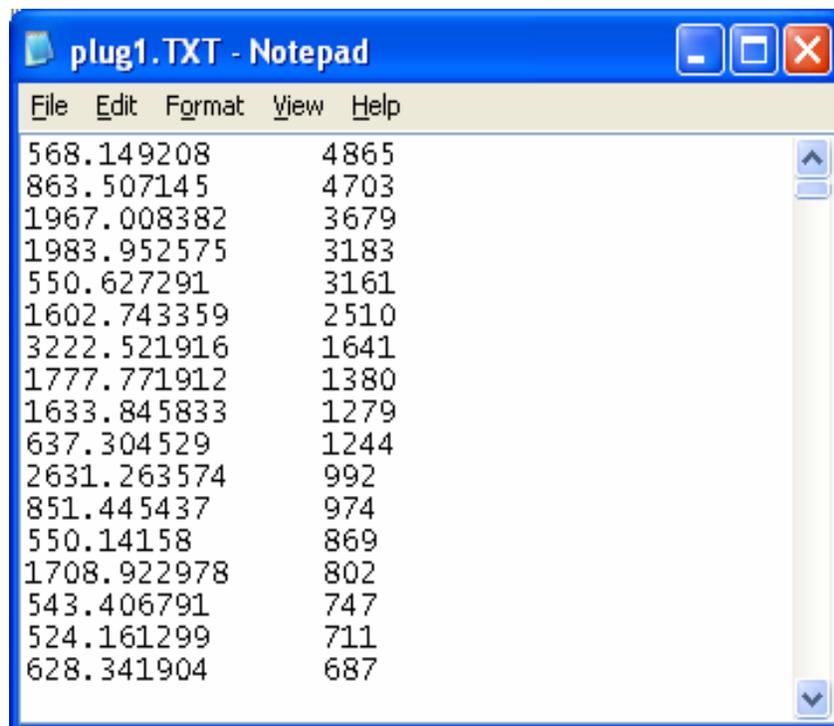
[5. Display of the result](#)

[6. Other Tools of the software](#)

I. Before using this tool

1. Pre-regulating the data

The input data of the “ProteinDecision” are the PMF spectral peaks derived from the experiments. The X axis is the spectra of the mass-to-charge ratios for the digested peptides and the Y axis is the peak intensity related to the abundance of the peptide. The regular input data file is in text format. In each row of the data file, the first column is the mass-charge ratio and the second column is the peak intensity. Before running the “ProteinDecision”, we need to sort the peaks of intensity in descending sequence.



File	Edit	Format	View	Help
568.149208			4865	
863.507145			4703	
1967.008382			3679	
1983.952575			3183	
550.627291			3161	
1602.743359			2510	
3222.521916			1641	
1777.771912			1380	
1633.845833			1279	
637.304529			1244	
2631.263574			992	
851.445437			974	
550.14158			869	
1708.922978			802	
543.406791			747	
524.161299			711	
628.341904			687	

2. System requirements

Please download and install the JRE 5.0, before running the “ProteinDecision”.

Reference URL: <http://java.sun.com/j2se/1.5.0/jre/install.html>

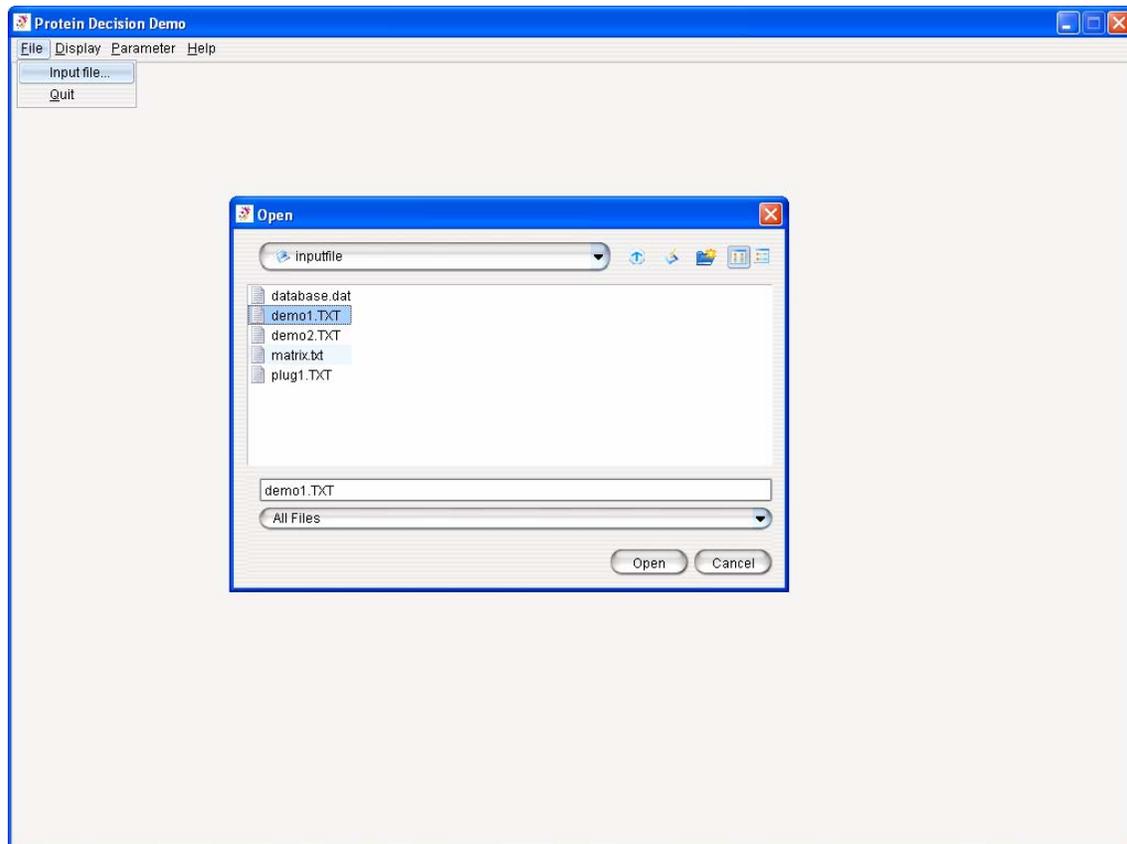
II. Installation guide

The software package can be unzipped and installed at any place on your computers. With JRE installed, you can run the “ProteinDecision” by clicking “ProteinDecision.exe”

III. How to use this tool

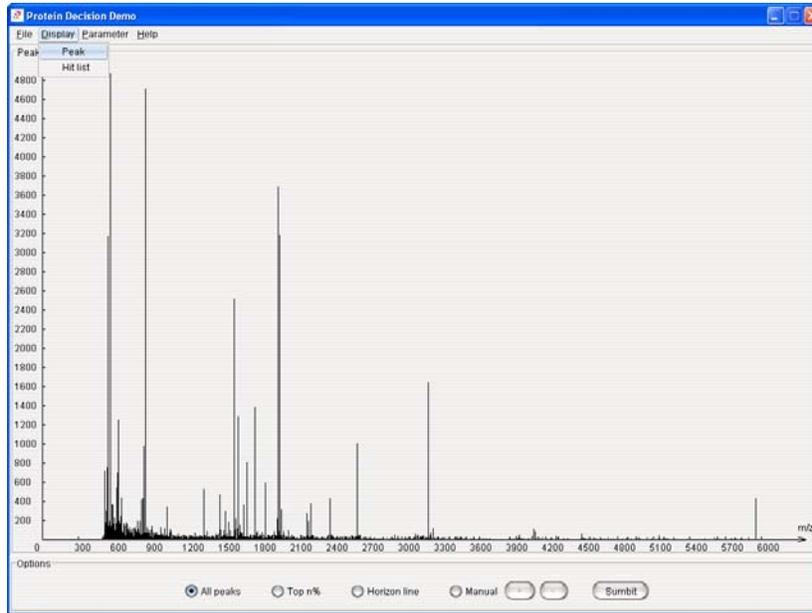
1. Input the file

Using menu “File->Input file...” to input the data of the PMF spectral peaks, the file type should be the text file. If loading data file successfully or finding some format errors in the data file, it will be shown in the content panel.



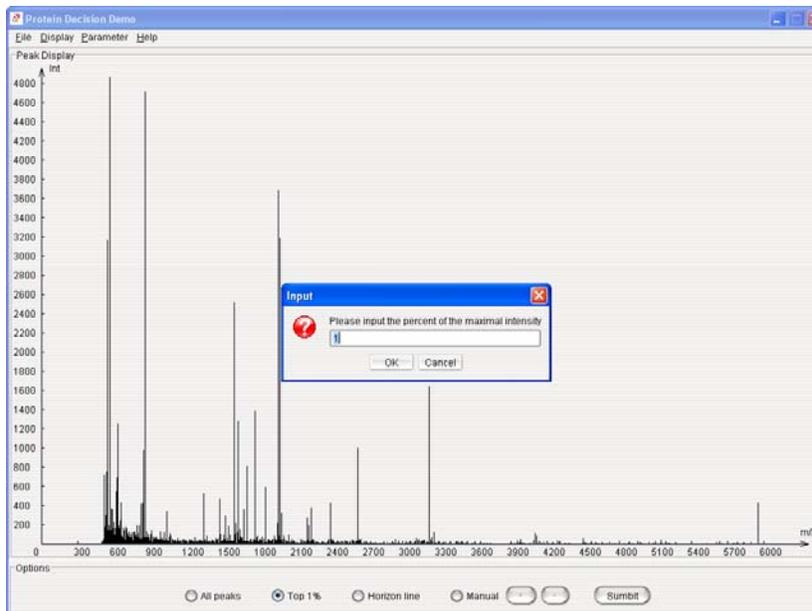
2. Select the peaks

Using menu “Display->Peak” to show all the input peaks in a 2D coordinate. You can use different methods to select the peaks that you want to submit to the scoring schema as the real signal. The un-selected peaks will be discarded.

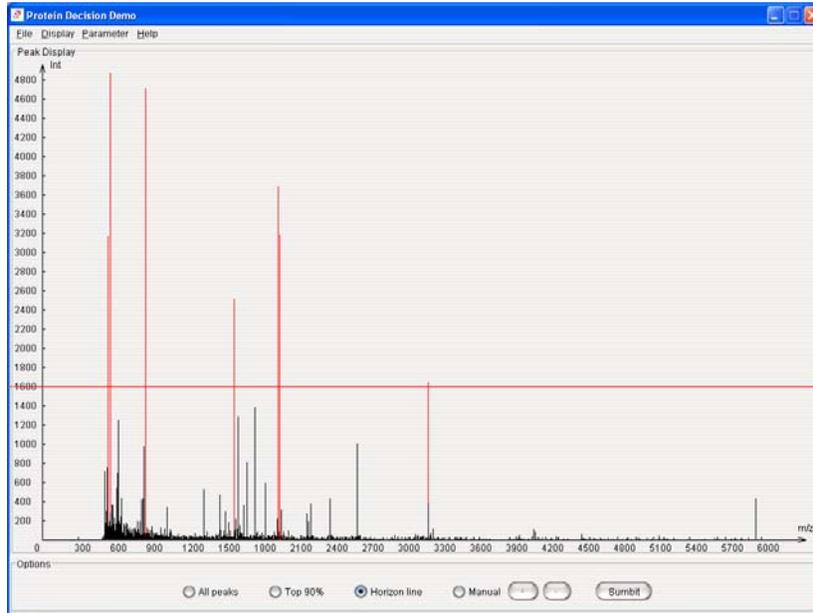


“All peaks”: select all of the input peaks as the valid data.

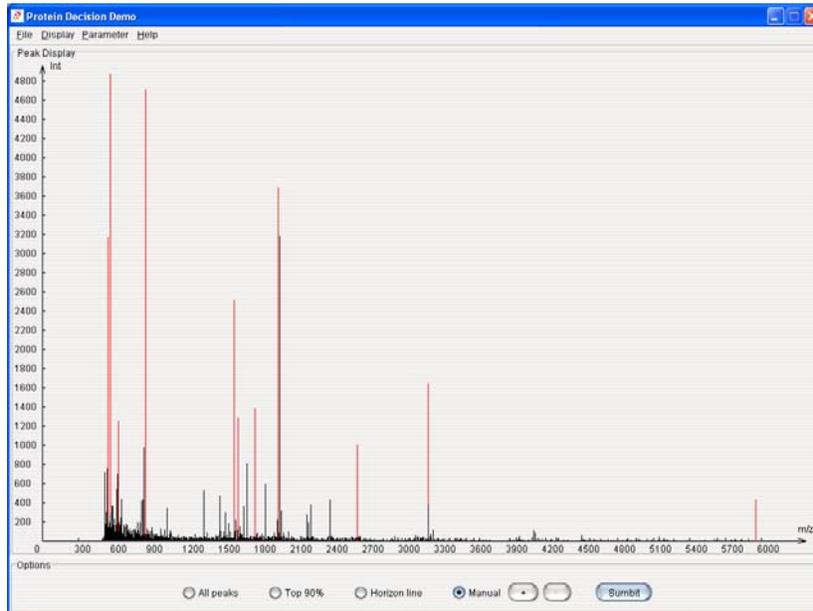
“Top n%”: select the peaks whose intensity are greater than or equal to n percent of the top peak.



“Horizon line”: show a horizon red line and drag this line to select the peaks whose intensity is above this line.



“Manual”: you can select each peak by manual and the selected peak will be shown in red. If you want to select one peak, click the top of this peak. You can use “+” to zoom in the curve and use “-” to zoom out the curve.



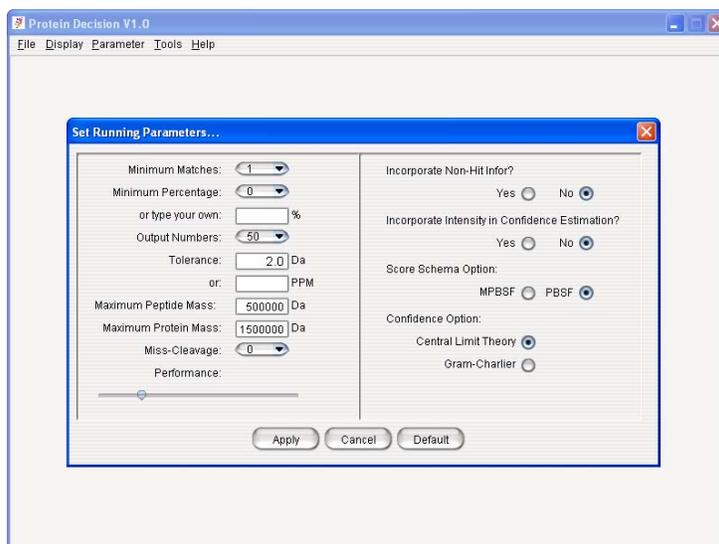
3. Set Parameters

3.1 Set Running Parameters

The software has an interface for you to set the running parameters of your own. The default parameters are shown in the following figure; however, you can change them in specific application. The “Minimum Matches” and “Minimum Percentage” gives the lower bounds of the minimum number of peptide and minimum percentage of peptide that get matched between mass-spectrum and database proteins which can be considered as a candidate. The “Maximum Peptide Mass” and “Maximum Protein Mass” give the upper bounds of the peptide mass and protein mass that can be considered as a candidate. The “Output Numbers” defines the size of output set for candidate proteins. The “Tolerance” defines the “matching area” between mass-spectrum and digested peptide. The “Miss-Cleavage” supports uncompleted digestion with the specified number of miss cleavage. The miss-cleaved peptides are possibly being considered when matching with the experiment data.

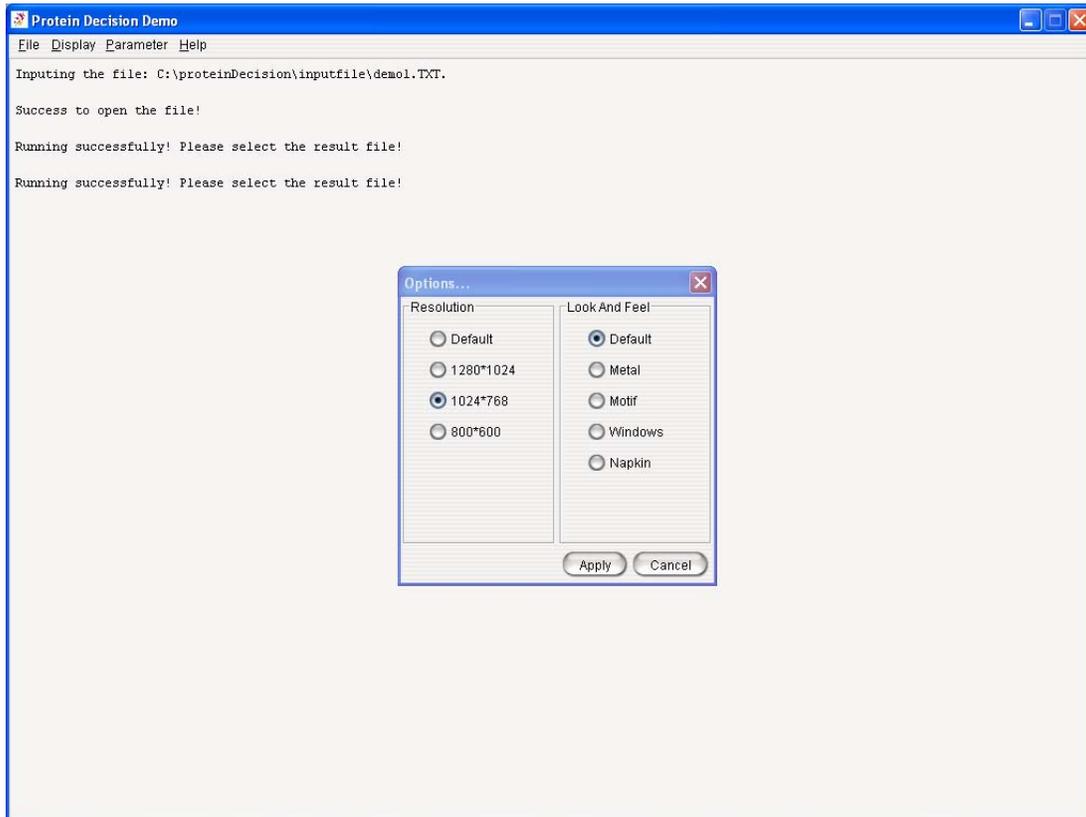
You can also set performance value according to their computer’s configuration. When the slider’s position is at the right most, the program reaches the highest performance. However, it requires the largest computer memory at meanwhile.

You can use a bunch of radio button options for different search. If the “Incorporate Non-Hit Infor” option is “Yes”, the scoring schema used in the program integrates the "non-hit" spectrum fragment information as part of the score. It's an option for users, however, not recommended for most cases. If the “Incorporate Intensity in Confidence Estimation” is “Yes”, the program uses the peak intensity in confidence evaluation; If the “Score Schema Option” is “PBSF”, the program uses PBSF score schema and if the option is “MPBSF”, it uses the modified “PBSF”. The detailed information of these parameters can be found in the paper: *Zhao Song, Luonan Chen, Ashwin Ganapathy, Xiu-Feng Wan, Laurent Brechenmacher, Nengbing Tao, David Emerich, Gary Stacey, Dong Xu. Development and Assessment of Scoring Functions for Protein Identification Using Peptide Mass Fingerprinting Data. Electrophoresis. 28:864-870, 2007.*



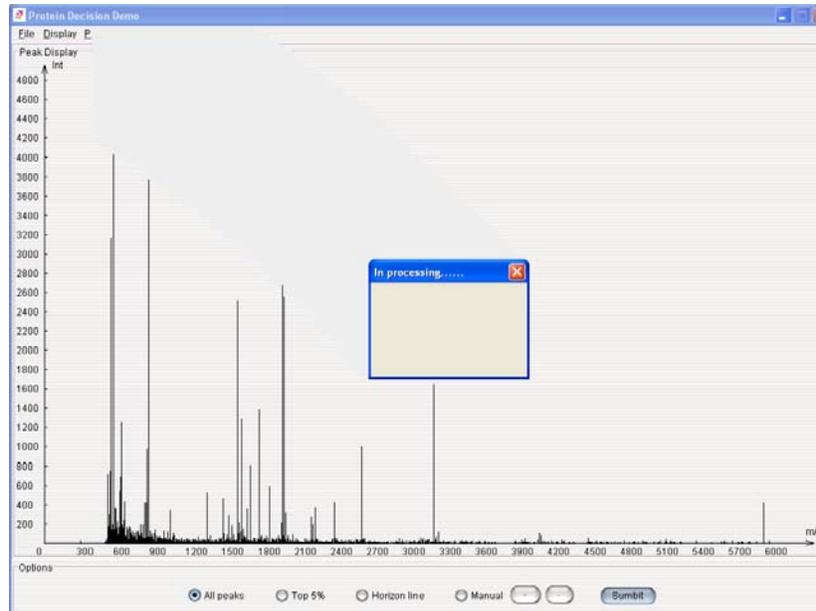
3.2 Set Display Options

Using menu “Parameter->Options”, you can select the fit resolution and style of software. The default resolution is the maximum resolution of your screen. You can select other solution according to the monitor type. You can also change the program skin for the characteristic at the Look and Feel panel.



4. Run “ProteinDecision”

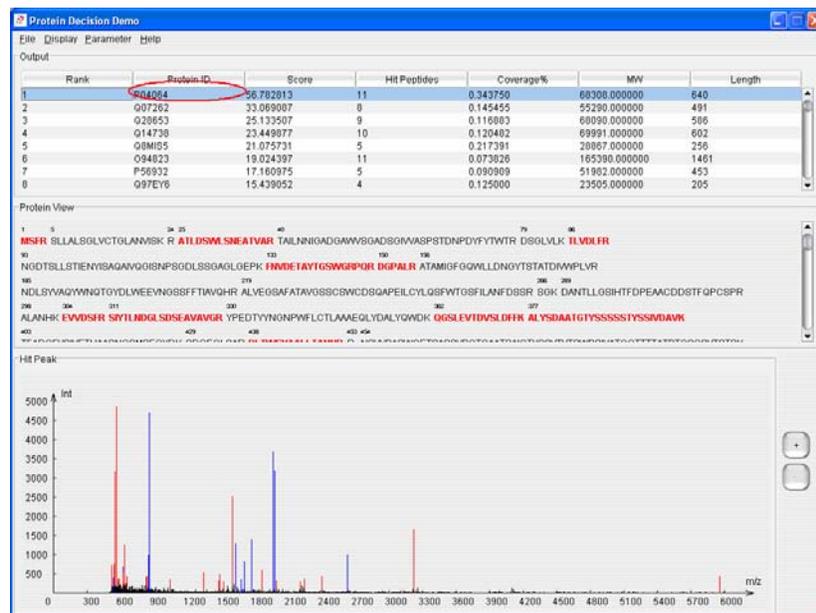
In the peak displaying panel, after selecting the peaks, click “submit” to run the “ProteinDecision”. It will take few seconds to get the result.



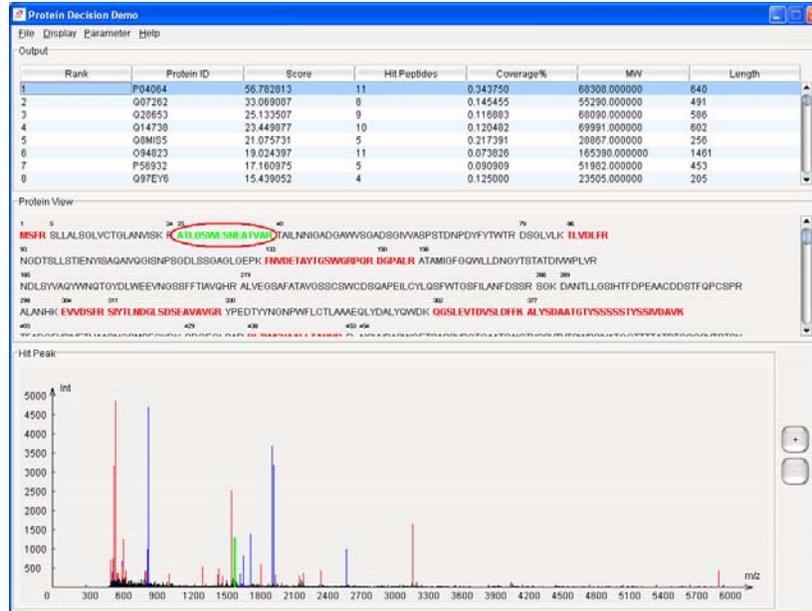
5. Display of the result

When the program is finished, the result will be shown automatically.

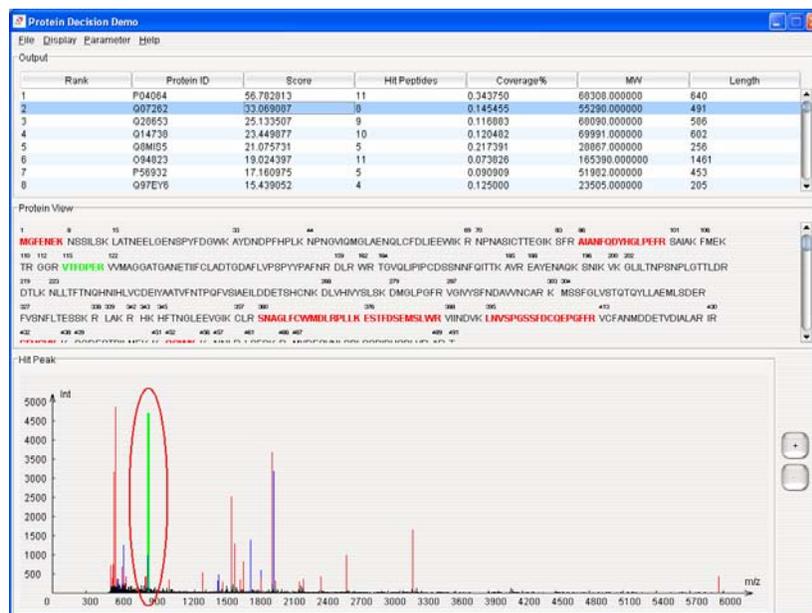
“Output” is the list of the hits proteins. When you select one protein, the whole sequence of this protein will show in “Protein view”. When you double-click the certain protein ID, you will see the corresponding web page in your browser.



“Protein view” shows the whole sequence of selected protein. The red peptides are matched tides. When you click one peptide, it will become green and its corresponding peak (in “Hit Peak”) will become green also.



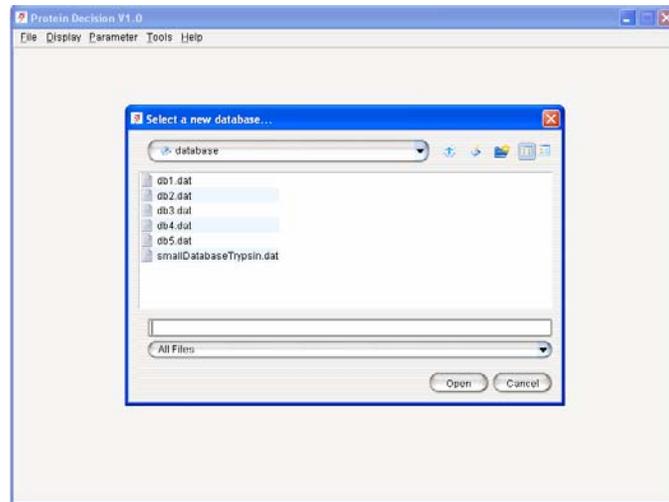
“Hit Peak” is a 2D coordinate in which the peaks will be shown in 3 different colors, blue means those peaks are hit-peaks, red means those peaks are selected by user and black means those peaks come from input file. You can click the top of the hit-peaks, it will become green and its corresponding peptide (in “Protein view Hit Peak”) will become green also. You can zoom out or zoom in the 2D-curve by using “+” or “-”.



6. Other Tools of the software

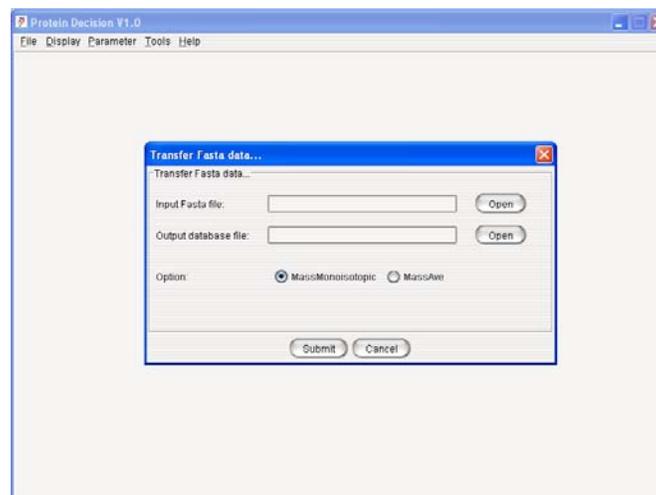
6.1 Select Searching Database

The software allows you using your own searching database. By clicking menu “Tools-> Change Database” the following dialog will pop up. You can browse the files and select the database (program defined format).



6.2 Transfer database

The program reads databases of program-defined format. However, these formatted files can be transferred from FASTA files and saved at any place on the computer. Using menu “Tools-> Transfer Data”, the following dialog will pop up. You need to select the input FASTA file and select an output file path (including file name) to save the result. The Radio Button “MassMonoisotopic” and “MassAve” are options for calculating the protein mass. “MassMonoisotopic” uses mono-isotopic mass for amino acid and “MassAve” uses average values.



6.3 Batch Submit (function paused : being improved)

When you have multi mass-spectrum input files, you may want to search the proteins in a batch. The “Batch Submit” dialog provides the function of running program in a batch mode. Using menu “Tools-> Batch Submit”, you need to fill the path information of the directory that contains the multi mass-spectrum input files and select the output path to submit the task as shown in the following:

