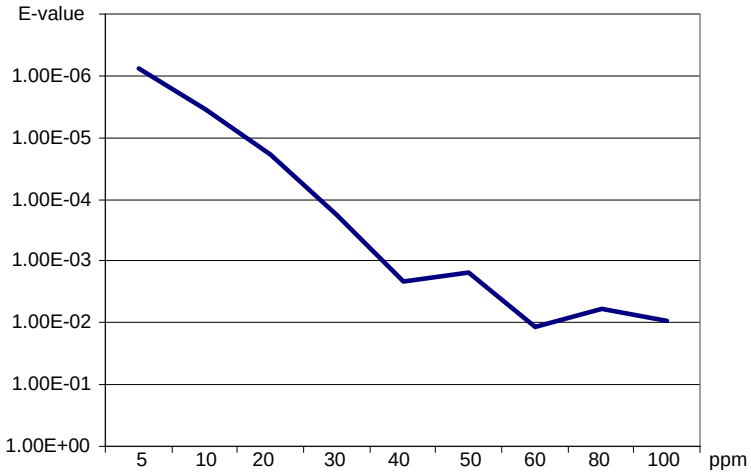
PeakErazor

Introduction

When performing a peptide mass fingerprinting experiment you want as many mass values as possible in order to maximize your chance of finding a protein in the database (a so called hit). However, it is not just the number of mass values that are important in the search, but other parameters play just as important a role:

Database: The larger the database, the smaller the score – other things being equal. It thus pays to select the smallest database that contains all the proteins of interest. PeakErazor has no influence on this parameter.

Precision: The more precise your data are, the better the chance of finding only relevant proteins. Particularly as the precision comes below 30 ppm, the score (E-value) increases (in this case it is randomized precision searched on Mascot):



One of the main functions of PeakErazor is to maximize the accuracy of the input data. Compared to a normal 2-point or external calibration, PeakErazor will normally increase the accuracy by 30-50%.

Input data: Including non-specific data into a dataset will decrease the score of the hit. If the non-specific data originates in keratin, i.e. really are specific for a contaminant protein, it gets worse, as you will find non-correct hits. PeakErazor clears up your input data by removing known contaminants.

In addition PeakErazor will give you a good estimate of your accuracy and help you identify potential deamidation, wrong isotope choice, peaks that are close to contaminant values but are ‘real’ data etc.

Furthermore, you get a number of additional tools for handling peak lists – identify contaminants in your sample preparation, project specific contaminants, analyze isoforms etc.

Program layout.

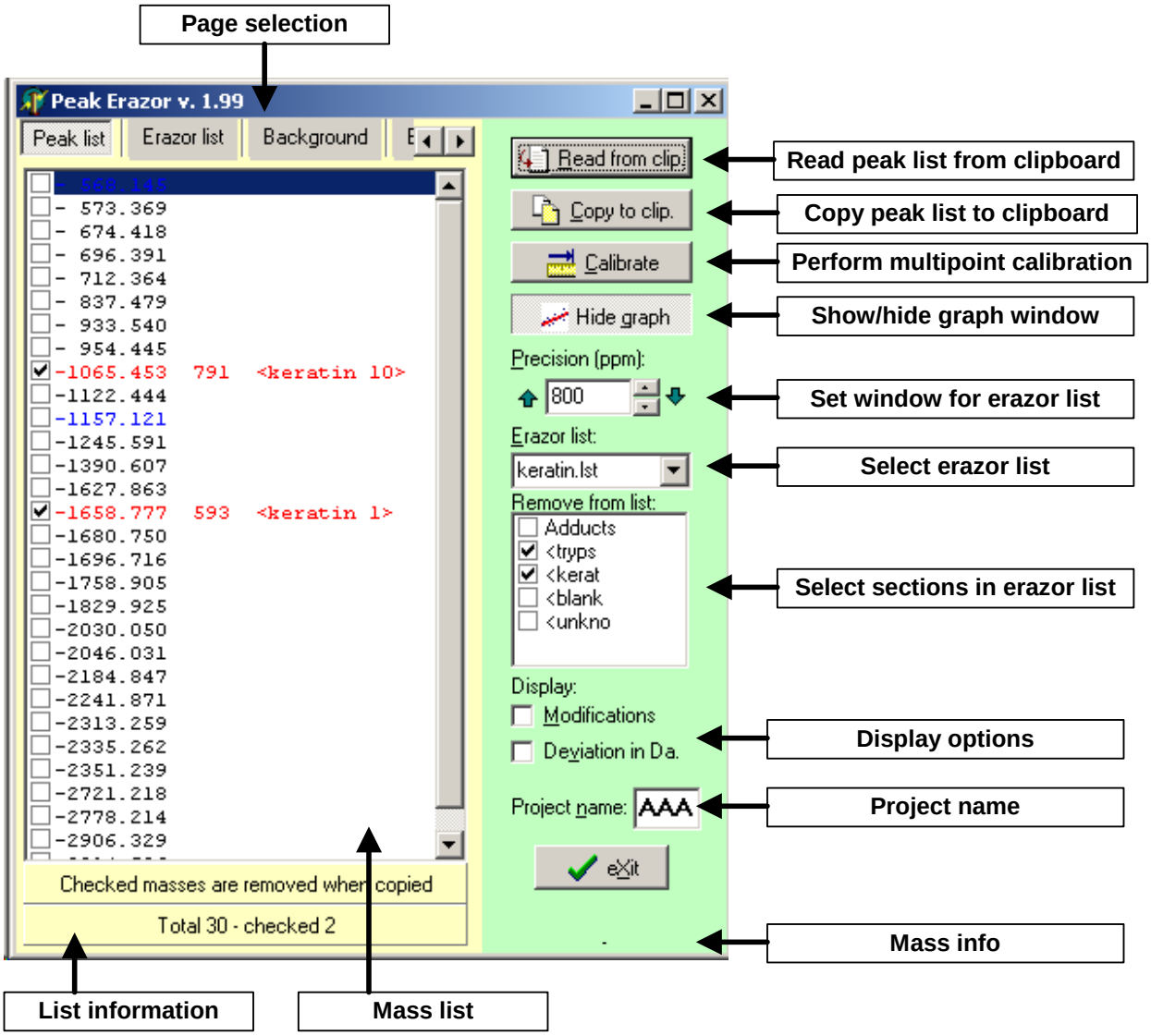
The program consists of two windows:

The main window which takes care of data input and output and controls the display of the second window.

The graph window which shows the calibration points for your data.

Both windows can be sized individually. The program ‘remembers’ the layout on the screen so you have the same position and size when you reopen the program.

Main window:



Page selection: Lets you change the display between ‘Peak list’, ‘Erazor list’ (list of contaminants), ‘Background’ (compare mass values from a project), ‘Evaluate’ (find new contaminant and manage erazor lists).

Read peak list: Copies a peak list from the clipboard into the peak list table. Any prior peak list will be cleared.

Copy peak list: The (calibrated) peak list will be copied to the clipboard, ready for pasting into the PMF search program. If a project name is specified, the list will also be copied to the ‘All mass’ file on disk, used for the ‘Evaluation’ page. If the button is pressed multiple times, only one copy of the data will be copied to ‘All mass’.

Calibrate: When pressed, a linear multipoint calibration will be performed on all checked mass values in the mass list. At least two values have to be selected.

Graph window: Toggles the display of the graph window (Show/Hide).

Precision (ppm): Determines the level below which the mass list is compared to the Erazor list. The left and right button (up and down arrows) sets the high and low values (set in Setup). The field is synchronized with the display graph, so the y-scale of the graph corresponds to the setting of the ppm precision.

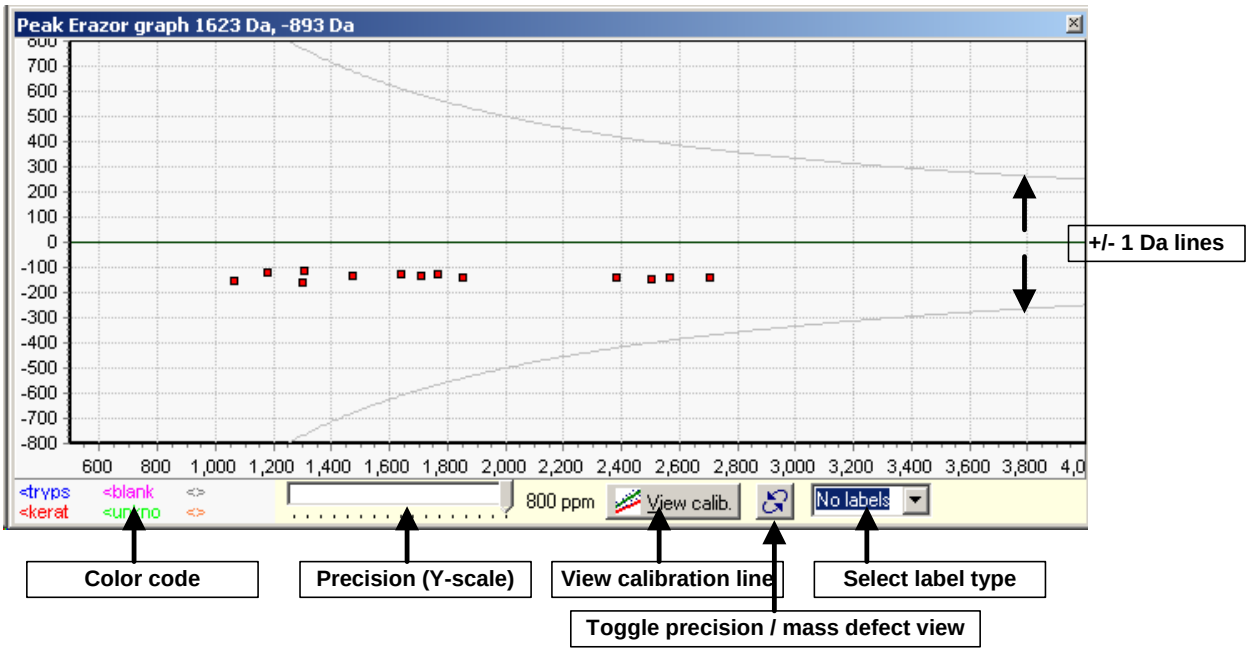
Sections: Each Erazor list can be divided into sections (max. 6) that can be toggled on and off in this box. Only the first six characters in the section name is important.

Display options: Lets you toggle display of modifications and display of deviation in Da. on and off. Display of deviation in ppm is always on.

Project name: If a project name is entered, the peak list is copied to the main mass list on disk. Later versions of PeakErazor will include functions to analyze individual projects (i.e. files saved under the same project name).

Mass list: The mass list shows all masses that are pasted from the clipboard. If any match is found to a mass in the 'Erazor list', the mass will be shown in red and the line will be checked (can be un-checked manually). Mass deviation in ppm values will be shown along with the name of the identified contamination. If the line is selected, the mass and the corresponding value from the Erazor list will be shown in the 'Mass info' line.
Checked mass values are removed from the list when it is copied onto the clipboard.
You may edit the value of a given number in the mass list by double clicking it.
Mass values in blue indicate a deviation of more than 125 ppm from the average mass defect.

Graph window:



The Graph window show all the matches between the mass list and the corresponding mass values in the 'Erazor list'. Each section in the Erazor list will be color coded in order to tell the difference in the graph.

Color code: Color of each section in the Erazor list.

Precision: Sets the y-scale with the maximum precision. The scale is synchronized with the main display window.

View Calib. Toggles the multipoint calibration line display on and off.

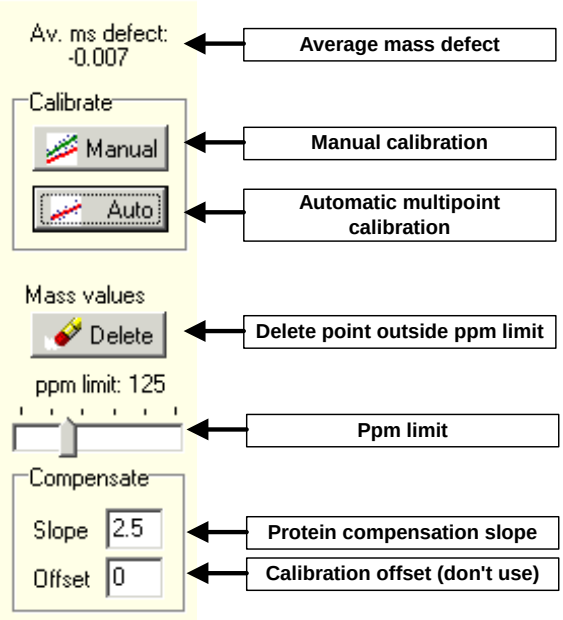
Toggle: Toggles the display between normal (deviation of mass values from Erazor list) and mass defect (deviation of all mass values from the average mass defect) view.

Labels: Selects the kind of label to display for each spot in the graph. Choices: “No label” , “Deviation”, “Mass” and “Mass and Deviation”.

The curving gray lines show the + and – 1 Da from the current calibration. When mass values fall on this line, it can indicate that we have selected a wrong isotope or deamidation (+1 Da) has occurred.

The mass defect panel.

When the graph window is switched into ‘Mass defect’ mode, a right-hand panel appears with functions specific for the mass defect mode:



Average mass defect: Just what the label says.

Manual calibration: Click the mouse button first in the lower end of the calibration line, then on the uppper end.

Automatic calibration: A linear multipoint calibration is performed.

Delete: Deletes points with a mass defect higher than defined in ”ppm limit” below.

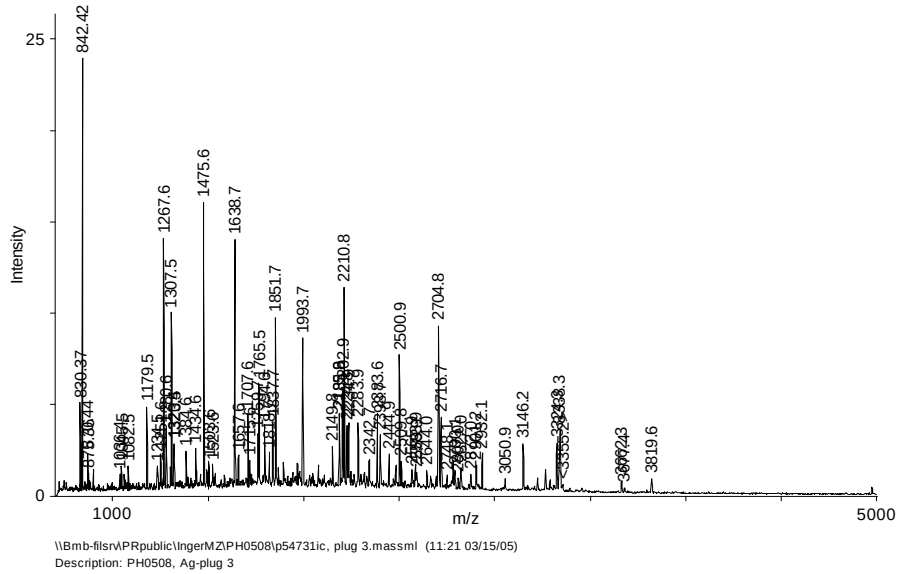
ppm limit: Slider that defines the ppm limit, shown as gray lines in the graph.

Slope: Compensation for variation in amino acid composition. A value of 2.5 fits most proteins.

Offset: Mass offset - should not be used at present.

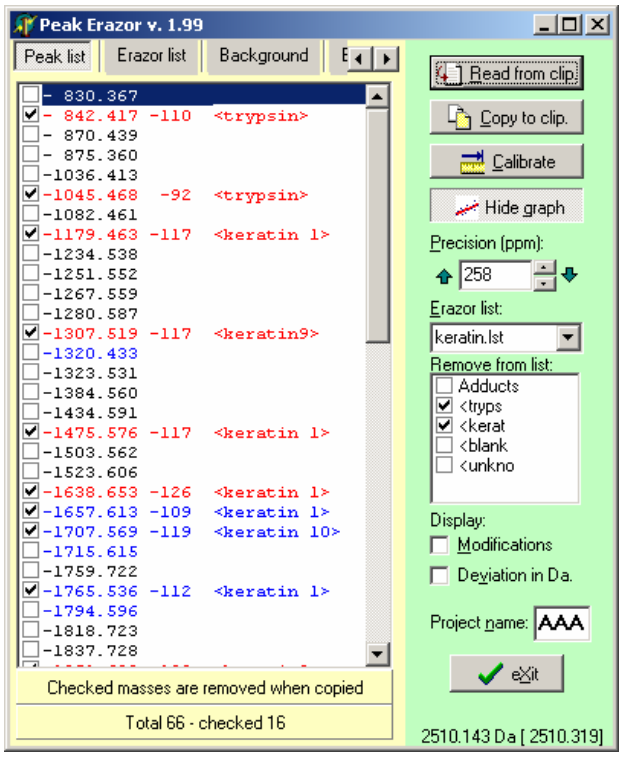
Calibrating in praxis.

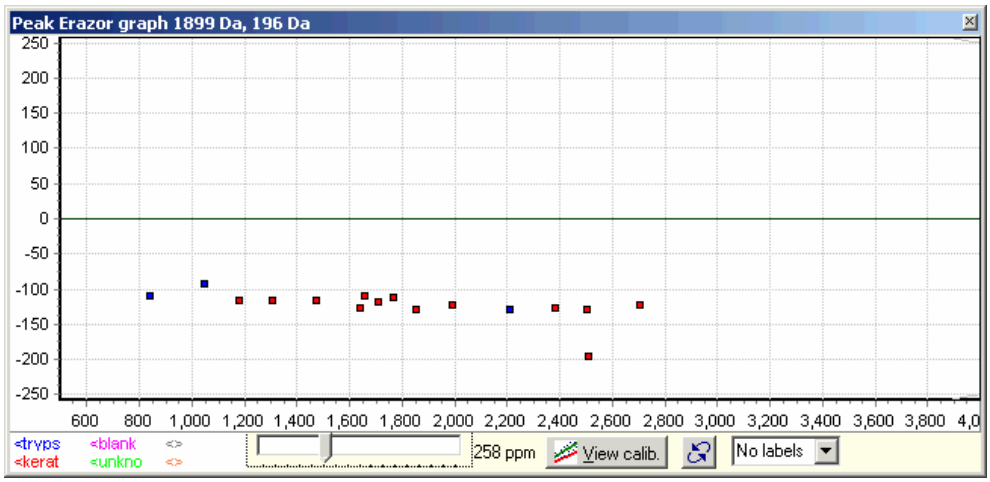
A mass spectrum of a protein digest shows no autolytic tryptic peaks on which to calibrate:



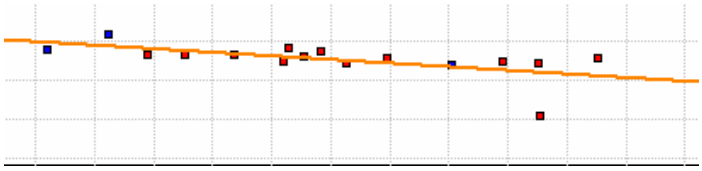
Copy the mass list to clipboard ( ) and paste it into PeakErazor:

The program immediately recognizes a number of peaks as originating from trypsin, keratin 1, 9 and 10. Deviation is off by ~110 ppm.



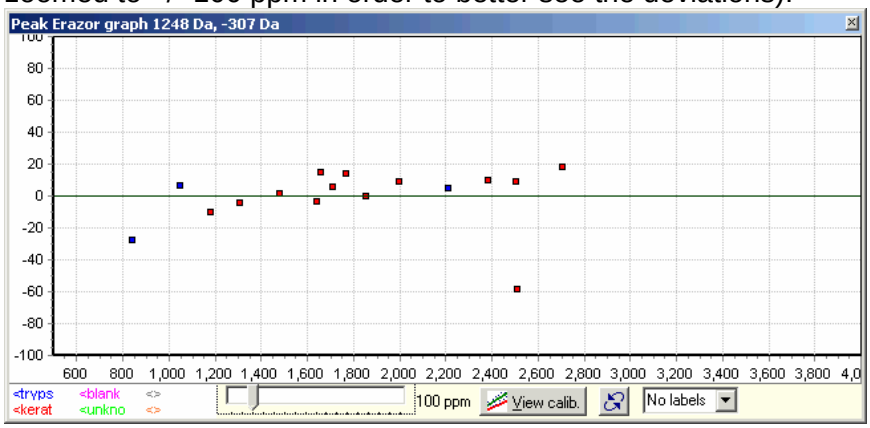


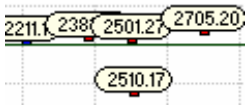
The found trypsin and keratin peaks can in the graph clearly be seen as lying on a straight line. This line can be visualized by turning on the calibration line (“View calib.” button).



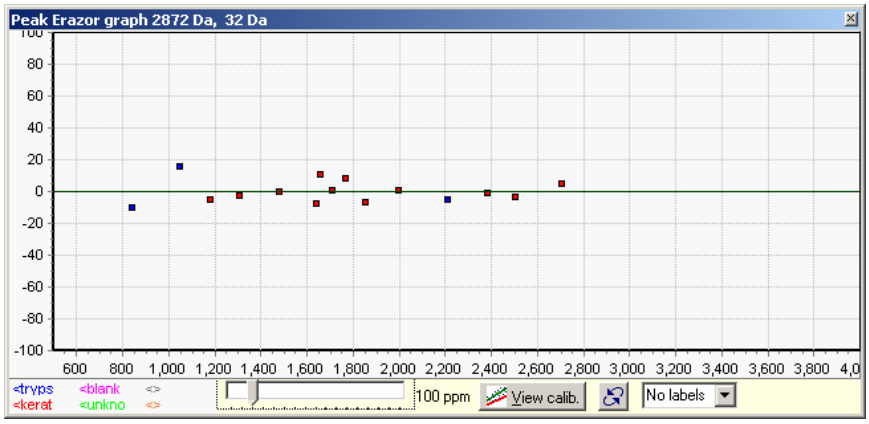
If you have high mass values (e.g. 4000- 6000 Da) you can expand the x-axis of the graph by right-click and select “Extended view” from the pop-up menu.

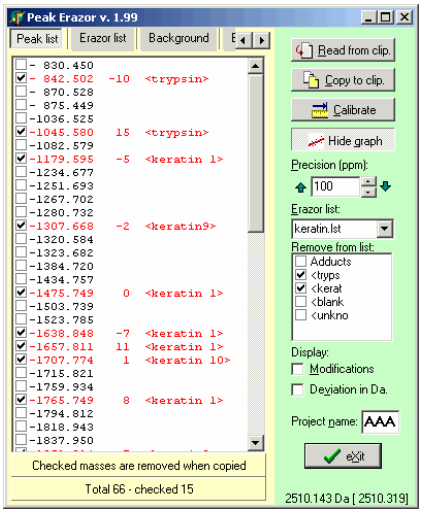
Press the “Calibrate” button in the main window and the peak list is calibrated (here the window is zoomed to +/- 100 ppm in order to better see the deviations).





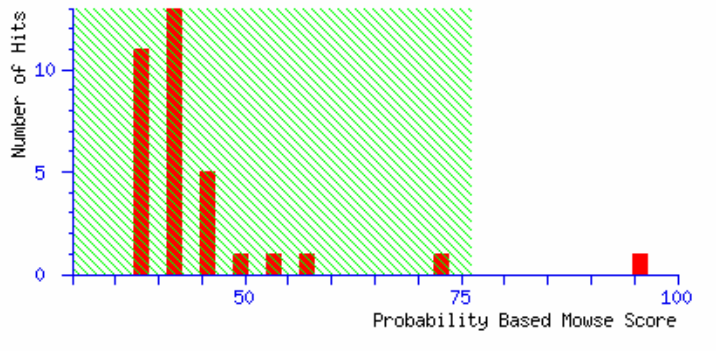
One ‘hit’ at m/z 2510 can be seen to deviating considerably from the other ‘hits’ and can be checked off in the main list followed by another click on the calibration button, resulting in a much improved fit.





In the main PeakErazor window it can be seen that the deviation is from –10 ppm to +15 ppm, indicating that setting the PMF search precision to 25 ppm should catch all hits.

A search using the Mascot search engine on the NCBI nr database yields a nice hit on a heat chock protein from the organism under investigation.



What to look out for:

If deviations of a single (or a few) peak(s) is larger than the rest, uncheck it from the list of contaminants used for calibration as it is not likely to be a known contaminant but could be a target peptide which accidentally have the same nominal mass.

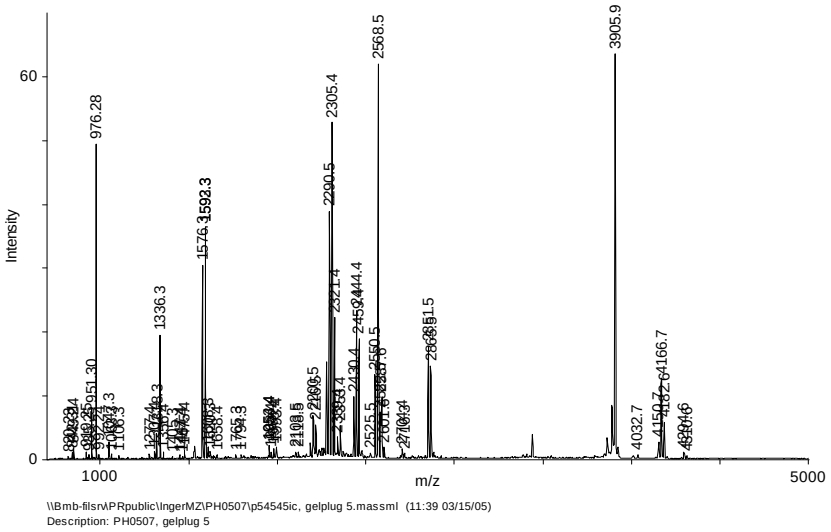
Alternatively, check the peak shape in the mass spectrum, if the peak shape is abnormal, delete the mass value from the list.

If a dot in the calibrated spectrum falls on the upper gray line, it may be the result of a deamidation (amide to acid) or wrong isotope selection (check the spectrum). If it falls on the lower gray line, it may be wrong isotope.

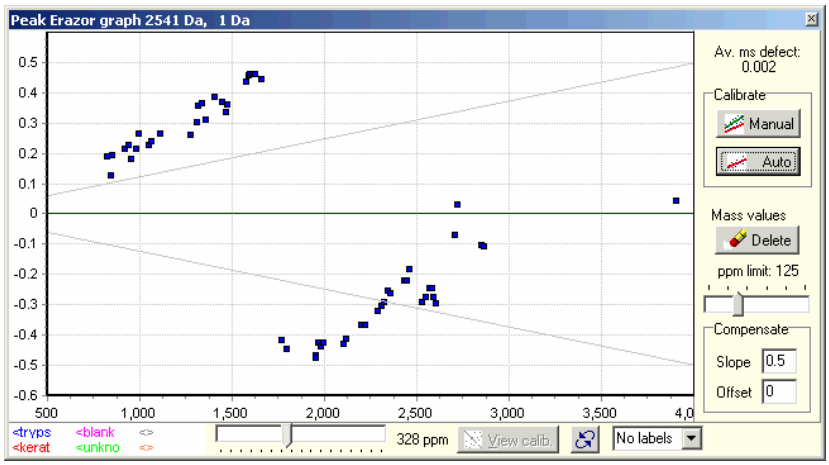
Unusual ‘shape’ of the ‘calibrants’, i.e. if the contaminant peaks do not fall on a straight line, it may indicate that the mass spectrometer is defect or the parameters during data acquisition wrong. This can checked further by analyzing a known protein with a large number of identified peptides where you load the theoretical peak list into the <blank> section.

What – no calibrants??

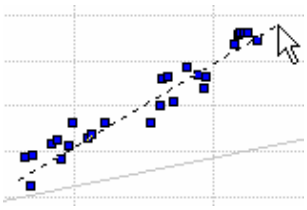
It regularly happens that you get a mass spectrum with no peaks on which to calibrate, particularly if you digest in solution or have coomassie stained spots (e.g. have a lot of material).



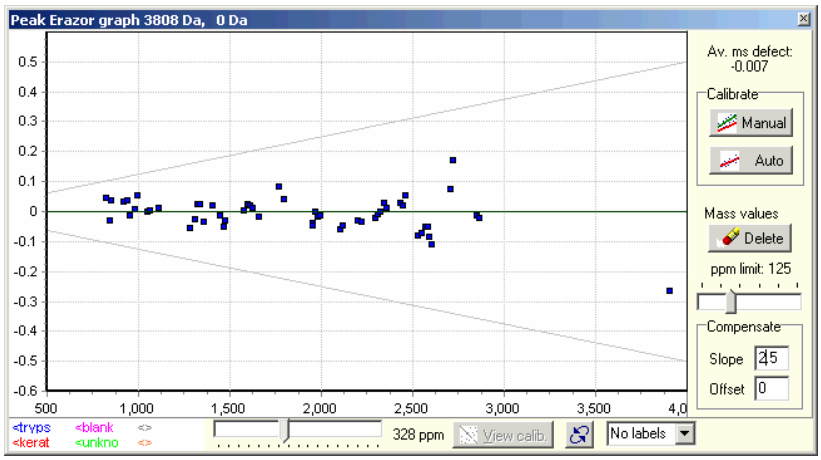
After pasting the peak list into PeakErazor, you find that no calibrants (a.k.a. contaminants) can be located. In these cases we can use the mass defect feature of PeakErazor. Click on the mass defect toggle button  and you suddenly see a dot for each mass value in your peak list:



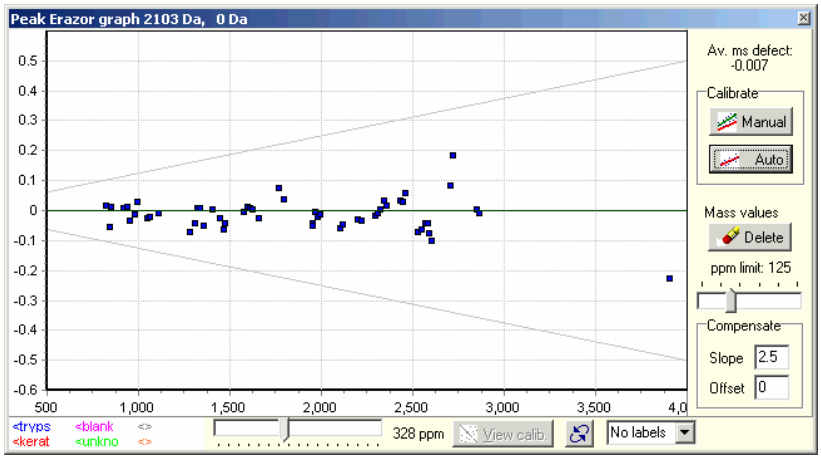
In this case the mass defect dots are divided into two sections, and we need to correct this in order to perform a linear calibration. Start by selecting “Manual”, click on the bottom part of the low mass group (i.e. around m/z 800, 0.1 ppm), drag a line through the low mass group of dots and click in the upper part of the group of dots (i.e. around m/z 1700, 0.45 ppm).



The peak list will now be partially calibrated (all the dots will be in a single group):

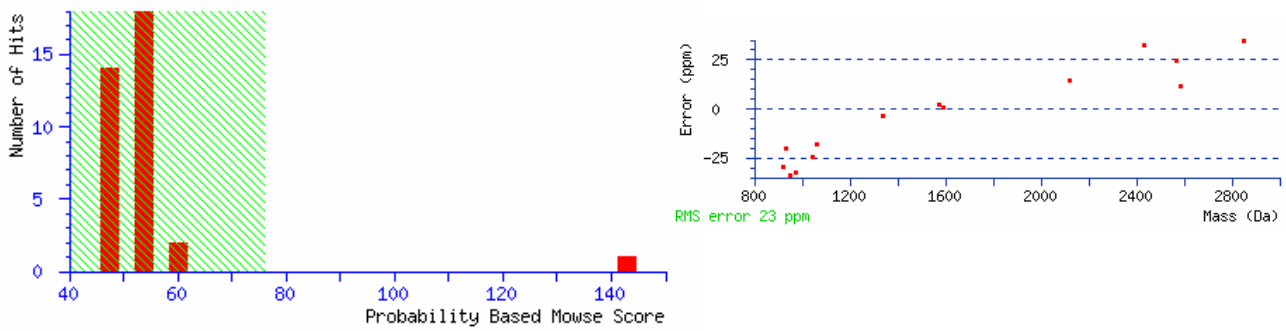


Now press the “Auto” button to perform a multipoint calibration.



If you have any points outside of the gray lines (default is 125 ppm, set in the right-hand slider) they are likely to arise from non peptide peaks (typically matrix ions) or from adduct ions (Na, K) and can be deleted through the “Delete” button. If you delete any dots, you have to perform another auto calibration. The “Slope” values vary a bit between proteins, but a value of 2.5 fits most proteins. The “Offset” is currently not used in PeakErazer and should always be zero.

As the precision of the mass defect is not as good as a standard internal calibration it is safer to set a rather wide band for PMF, in this case 50 ppm. In this case the result showed the MBL protein with all 14 peptide hits within +/- 30 ppm.



If you are analyzing a large number of similar proteins, it can be rewarding to optimize the ‘slope’ value for each dataset. Looking at the errors of the hits (above right) a clear trend of the deviations can be seen. You may be able to correct this by adjusting the ‘slope’ value.

What to look out for:

If the mass defect peaks are divided into two or three sections, you always have to draw a manual calibration line first. Draw it on the first section, as this is values closest to the ‘correct’ calibration (unless you have more than 1 Da error).

After initial ‘auto’ calibration you should delete values outside of the 125 ppm limit (the gray lines). Above m/z 1000 there is less than 1% chance of a peptide deviating more than 125 ppm from the ‘average’ mass defect used for calibration. After removal of ‘outside’ values you have to repeat the ‘auto’ calibration.

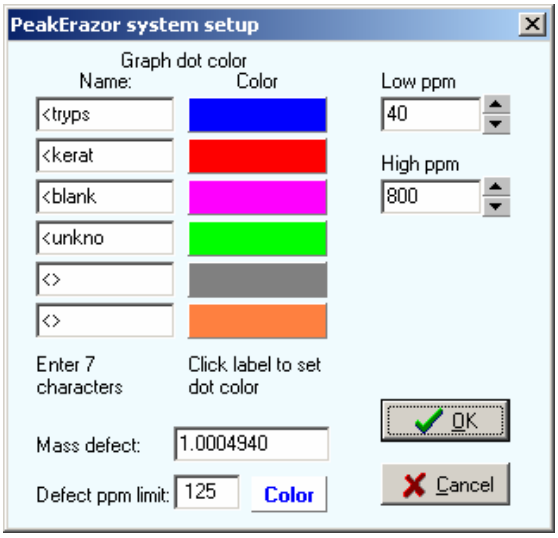
If you have a lot of adduct ions in your spectrum (e.g. Na and K, not O) you are likely to get a less accurate result as the mass defect of adducts deviate considerably more from the average peptide.

You may also perform a mass defect calibration on purified peptides if you can find enough contaminating peptides in the spectrum ‘grass’. Even if the signal to noise is quite bad (~2-3) it will be sufficient in many cases if you find enough peaks (e.g. at least 10).

Setup

Several settings in PeakErazer can be set by the user through the Setup menu.

Right-click in the main window to get the pop-up menu and select the “Setup” option. This opens the setup dialog box:



The left-hand part of the dialog enables you to set the color of each section in the current Erazor list. The section names are read into the Name boxes and the corresponding colors are shown in the boxes to the right of each name. Click on a colored box to open the ‘Edit color’ box and select a new color for the relevant box.

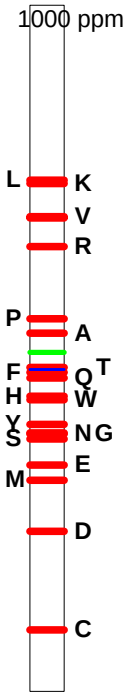
The right-hand edit boxes lets you set the low and high values for the ppm precision (the up and down arrow buttons in the main window).

Finally, at the bottom of the list you can enter alternate values for the average mass defect (default is 1.000494) and the default ppm limit in the mass defect view.

The “Color” buttons lets you set the color of the dots in the mass defect window.

The mass defect

The mass defect comes from the fact that proteins are made from a limited number of atoms, as carbon, oxygen, nitrogen and hydrogen (+ a little sulphur) make up the entire molecules and in almost the same proportions.

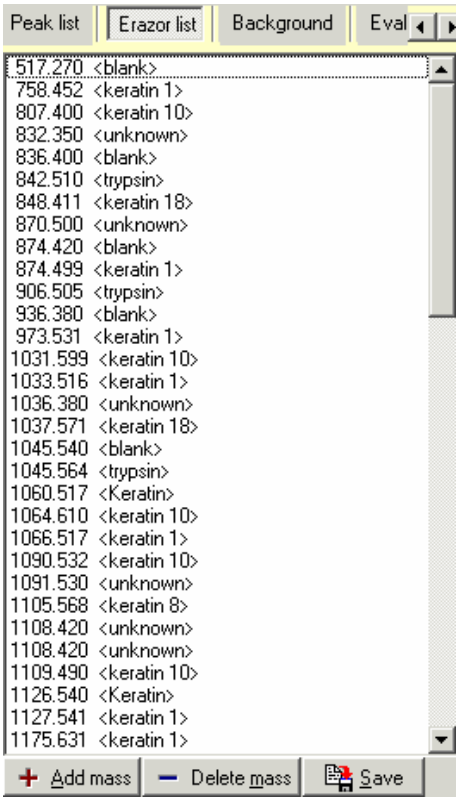


This results is that a peptide weighs more than the corresponding integer values, ie. Alanine weighs 71.037 Da instead of 71.0 Da. The value 0.037 is called the mass defect. The mass defect varies from residue to residue (figure at right) but is on average 0.0454 %, which results in the fact that peptides that have been correctly calibrated have a decimal value around .5 at m/z 1000 and .0 around m/z 2000.

Using this feature, it is possible to calibrate on the mass defect. However, as we are usually looking at tryptic peptides, which always (except for the C-terminal peptide) terminate in a lysine or arginine, we get a skewed dataset. Lys and Arg are two of the residues that have the highest mass defect of all 20 residues, which results in a higher relative mass defect at low m/z values and a lower one at high m/z values. This is compensated in PeakErazer and is further modified by the “Slope” parameter.

Furthermore, analysis on the entire protein database shows that only ~0.5% of all tryptic peptides deviate more than 125 ppm from the calculated average, it is relatively safe to delete all m/z values that are outside of this limit in the calibrated mass list.

The Erazor list



The Erazor list is the backbone of the workings of PeakErazer. It consists of a list of m/z values representing typical contaminants in your mass spectra.

Each value is presented with three decimals and a name in parenthesis. The name can be up to 12 characters long, and has to start with '<' and end with '>', leaving 10 characters for the name. Up to 6 names can be in use at any given time, but only the first 6 characters are used in each group, thus enabling having names like 'keratin 1' and 'keratin 9' that are grouped together but listed with different names.

Through the pop-up menu (right-click in the window) you can change the order of the list (by mass or by name).

Through the button at the bottom of the window you can add and delete values. When adding a mass, the default group is the one currently selected in the Erazor list. If you change the group name, a new group will be added unless it already exists. A value can be edited by double-clicking on the relevant number.

Lists are saved through the 'Save' button. You have to exit and reload the program in order to use a newly saved erazor list. You can 'Save as' a new name through the pop-up menu, where you can also save the list as a straight text file.

Much of the maintenance of the list can easily be done through the 'Evaluation page', see below.

The <blank>

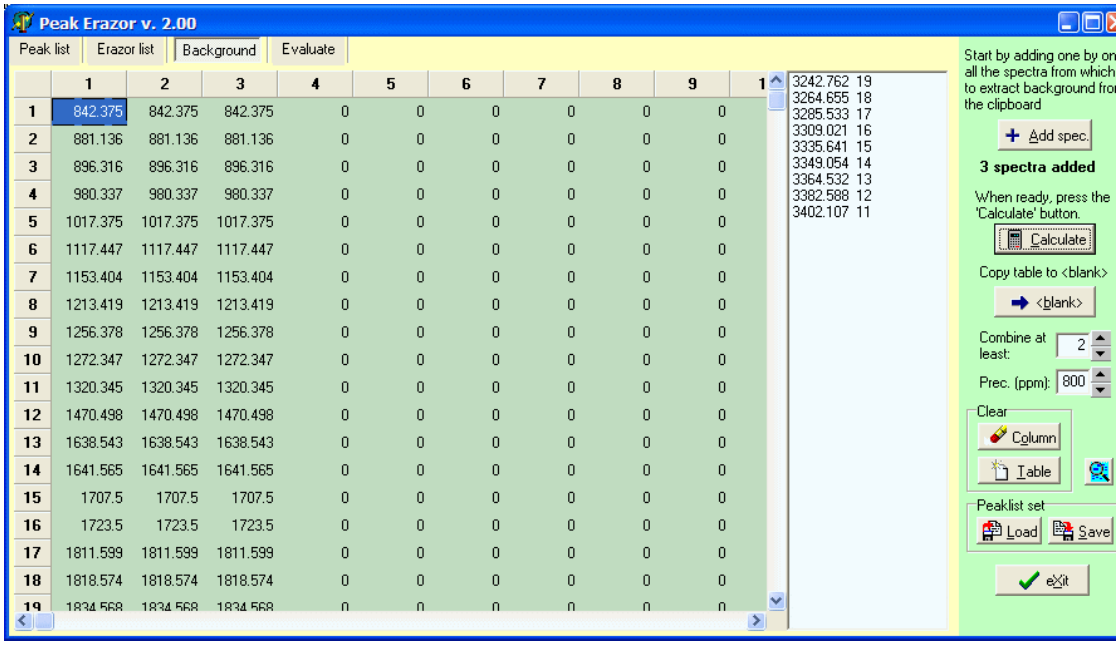
The section called <blank> has a special meaning in PeakErazor. From the main window you can copy a mass list into the <blank> section by selecting the appropriate menu option in the pop-up menu (this will delete any <blank> section already existing.

This option can be useful if you want to compare a theoretical mass list to an experimental list – e.g. first copy the *in silico* generated peak list (from GPMAW) into PeakErazor, copy this list to the <blank> section and finally copy the experimental peak list into the PeakErazor. In the graph you can then see the hits between the experimental and theoretical by turning the appropriate sections on and off.

Results from the 'Background' analysis (see below) can also copy the into the <blank> section.

Background

When the ‘Background’ section is selected, the PeakErazor window widens to show a grid.



Into this grid you can paste a number of mass lists by pressing the ‘Add spec.’ button for each mass list. When the desired number of mass lists has been entered, you can calculate the mass values that are common in the lists by pressing the ‘Calculate’ button. The criteria for combining mass values are set by the field “Combine at least”, e.g. at least this number of values has to be within the precision defined by the “Prec.” field (in ppm).

The calculated “common” list can be transferred to the <blank> section of the current erazor list by pressing the <blank> button (existing <blank> section is deleted).

The column which hold the currently selected cell can be cleared through the “column” button, and the entire table through the “Table” button.

A peaklist set (in .pkl format) can be saved or loaded through the “Peaklist set” buttons.

If you press the “Graph” button  you toggle a graphical picture of your data.

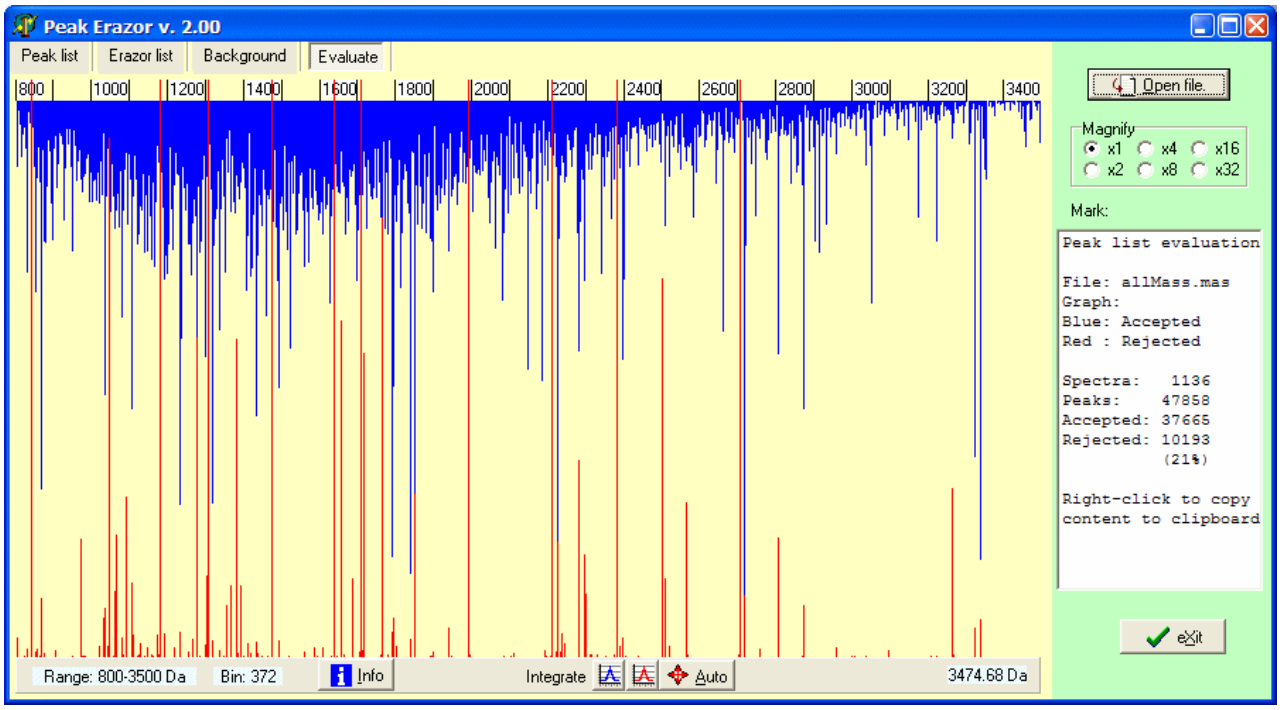
In this graph each peak list is presented as a horizontal line with each m/z value as a dot. The top graph shows the combined number of dots in each mass slot, the height representing the number of values. Dot size, horizontal and vertical grid can be set in the right-hand panel.

The “Dot click” options determines: “Row” when clicked, a given row is highlighted; “Column” when a column (in the top graph) is clicked, the corresponding dot values in the graph below is highlighted; “All col.” p.t. same as “column”

Evaluation

The last section in PeakErazor is the “Evaluation” page. If the ‘Project name’ is defined on the “Peak list” page, the results of every ‘Copy to clipboard’ operation is saved in a file on disk called “Allmass.mas” – only mass values and the date is saved, so it is not possible to identify samples.

When you select the “Evaluate” page, you are asked to select a file to open. You will normally select the ‘Allmass.mas’, but you may transfer the file from other computers. Alternatively, you can press the “Open file” button and select an appropriate file. When the file is loaded you will see a representation of all the peak lists saved by PeakErazor.

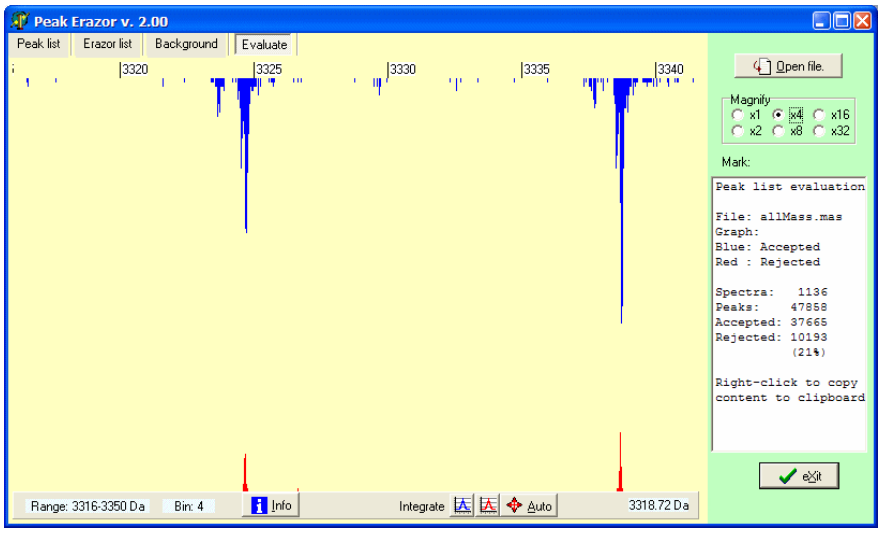


The mass values are presented as peaks divided into two sections: **Accepted** mass values (seen as blue peaks growing down from the top) and **rejected** mass values (seen as red peaks growing up).

The precision of the graph is 0.01 Da with the binning shown at bottom right of the window – in the above picture each pixel thus presents 3.72 Da. The right-hand box shows statistics on the currently loaded file (this information can at any point be recalled by pressing the ‘Info’ button).

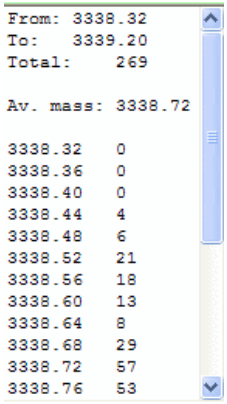
The graph is zoomed in the x-axis by clicking once at the low mass end (a green line will show the position) and click at the high mass end to finish the zoom. You zoom out by right-click in the graph. The y-axis is controlled by the ‘Magnify’ buttons to the right.

The main purpose of the ‘Evaluation’ page is to identify contaminant peaks in your mass spectra. In the picture above, two very strong peaks can be seen at m/z 3324 and 3338. In order to identify the peaks a zoom is performed (right).



To get the exact mass, you click the blue ‘Integrate’ button  followed by a click to the left and right of

the peak to identify in order to integrate the (blue) area between the two m/z points. The result of the integration is shown in the right-hand text box (right). From top to bottom is shown: **From**, **To** the integration limits; **Total** the number of values in the range (in this case 269 out of 1136 spectra approximately ¼ of all spectra – a typical contaminant!). **Av. mass** is the integrated value of the 'peak', not the average mass value. Below this is listed the number of values in each 0.01 Da slot.

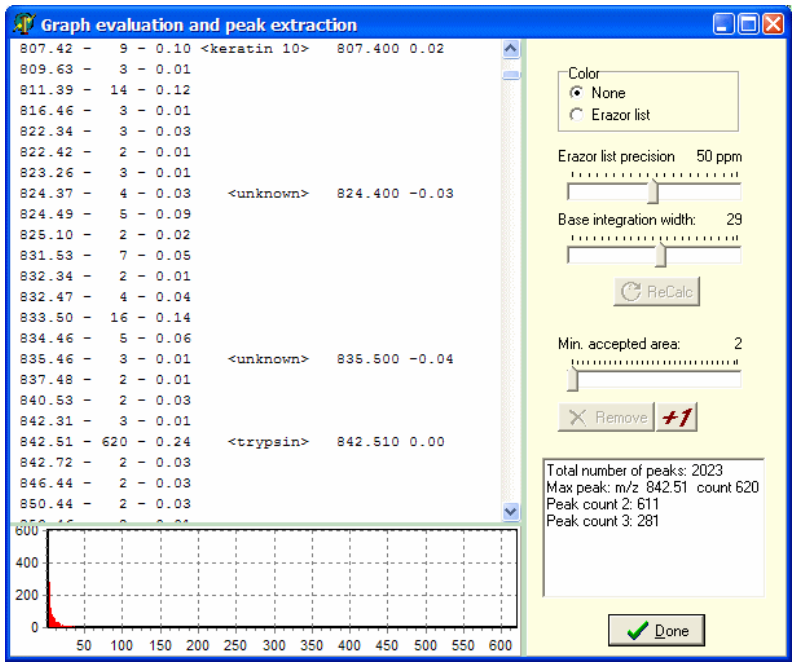


In the above case, the two mass values have been included in the Erazor list, which can be seen as two red peaks (rejected m/z values) are starting to 'grow up' below.

A slightly more automated evaluation can be done by selecting the 'Auto' button 

:

A multipanel window opens which shows a list of all extracted "peaks" (m/z values) from the graph. The way that the program calculates the "peaks" can be adjusted by the two sliders to the right followed by pressing the 'Recalc' button.

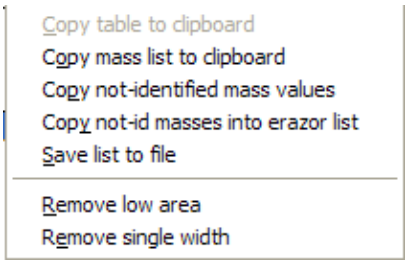


Mass values that are recognized in the current 'Erazor list' will be listed with the identified compound, their value in the list and the deviation. You may 'import' the color of the compounds from the peak list page by selecting 'Erazor list' from the color option to the right. This makes it easy to identify the different compounds when scrolling through the list.

When the list opens, every "peak" containing at least two values is listed. As this is unlikely to present contaminants (and the list gets long) you can use the 'Min. accepted area' slider to decrease the list to the more important m/z values (remember to press the 'Remove' button for every change). The '+1' button increases the minimum accepted area by one and recalculates the list. The information list below lists the number of entries in the peak list, the largest peak count and the number of entries in the list that have the minimum number of peaks (and the minimum + 1).

The graph at the lower left shows the distribution of counts from the minimum accepted area and up to 650.

The pop-up menu of the list gives you a number of choices for handling of the list. The most important command is the copy not identified m/z values into the current erazor list. When you select this option you are given the choice of either adding/replacing an existing section in the erazor list or to add a new section.



If you select to save the list to a file, it is saved as an erazor list, and you have to give a name to the values in the list without a name.

The 'remove low area' command is identical to the 'Remove' button, and the 'Remove single width' removes items in the list that has a peak width of 1.

This page is not intentionally left blank.

(I just ran out of words)