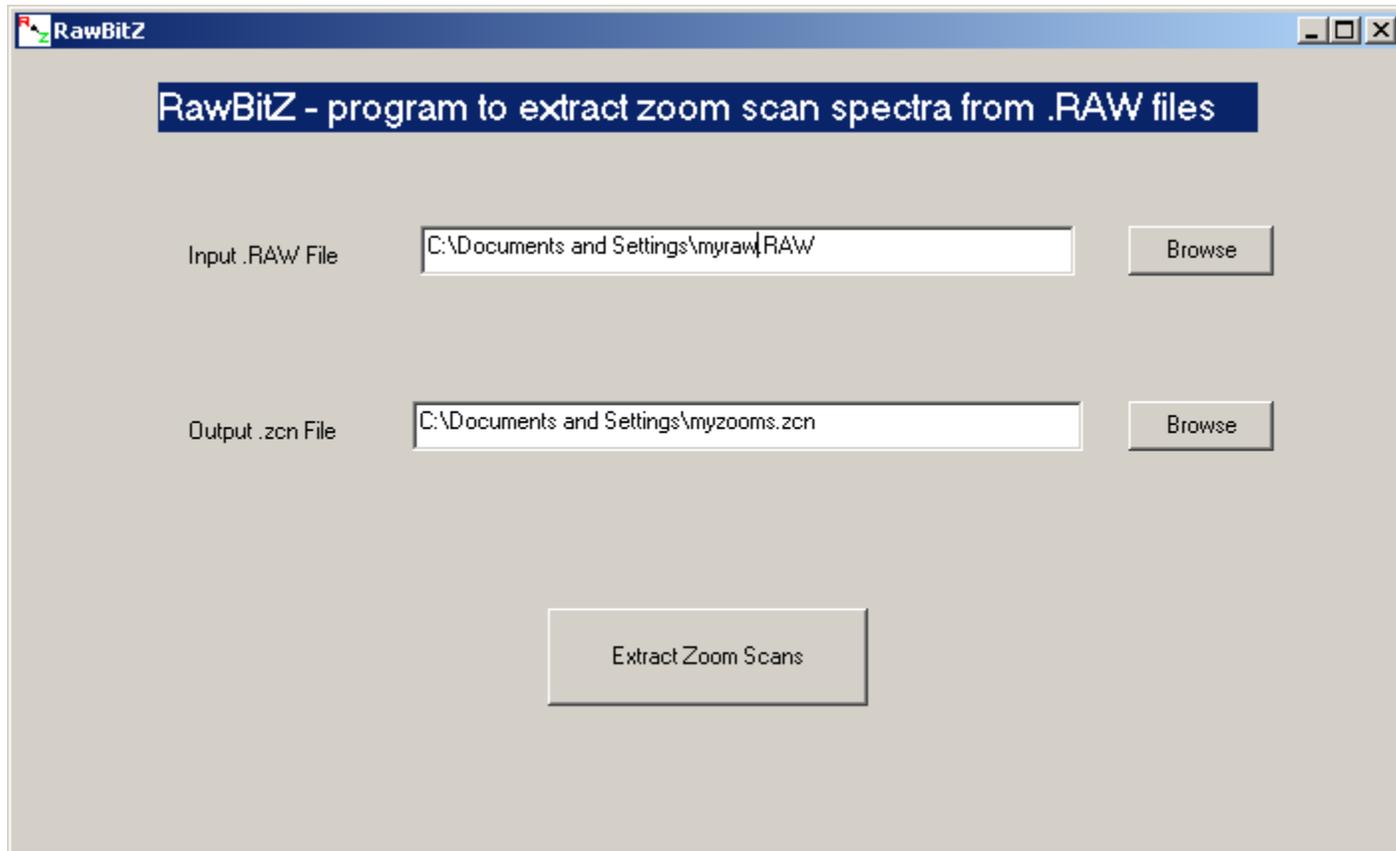# ZoomQuant Tutorial

Overview

The ZoomQuant application is part of a suite of programs that help to automate and simplify the process of analyzing experiments using ^{18}O labeling of peptides. The data analysis process requires four steps:
1. Use Sequest to search the .RAW file for peptide matches in the database of your choice. Make sure to place the output .OUT files in a separate directory, i.e. don't mix .OUT files from different .RAW files (For information on using Sequest, please see the documentation distributed by Thermo).
2. Use Epitwrapper to filter and format the results from the .OUT files to create the .colon file. Epitwrapper is a GUI wrapper to the Epitomize program.
3. Use the RawBitZ program to extract the zoom scan spectra from the .RAW file to produce the .zcn file.
4. Load the .colon file and .zcn file into the ZoomQuant application and carry out the analysis. Note: The .colon file and .zcn file MUST BE FROM THE SAME .RAW file.

RawBitZ

The RawBitZ program is a simple utility to take the Thermo binary .RAW file and extract the zoom scan spectra information. It requires that the XDK toolkit provided with Thermo's BioWorks be installed on the computer and runs only under Windows. RawBitZ opens the binary file, identifies the zoom scans, and saves all the zoom scan spectra in an ASCII format. A screenshot is shown below.

RawBitZ

RawBitZ – program to extract zoom scan spectra from .RAW files

Input .RAW File   C:\Documents and Settings\myraw.RAW   Browse

Output .zcn File   C:\Documents and Settings\myzooms.zcn   Browse

Extract Zoom Scans

The user browses to find the .RAW file and to choose the output filename and directory.
The 'Run' button starts the process, which can take several minutes depending on the size of the .RAW file and the .
When the process is complete, the program indicates the number of scans processed.

The .zcn file is now ready to be used with ZoomQuant. Note that this process needs to be done only once per .RAW file. If the .RAW file is searched with different Sequest parameters, the same .zcn file can be used.

Epitomize

The Epitwraper program provides the graphical user interface to the Epitomize program. Epitomize reads all of the Sequest .OUT output files in a directory, filters and groups the results and the produces the .colon file used by ZoomQuant.

Create a summary file from your Sequest .out files

Select data directory.

| top\N15_data\11-15-04_N15_zoom | ... |

Select summary file name.

| 11-15-04_N15_zoom | ... |

Ignore filter and include all top hits

☐ Ignore Filter

(Optional) Load a filter.

| | ... |

Fill in filter parameters.

| ☑ use charge state 1		☑ use charge state 2		☑ use charge state 3		☐ use charge state 4+	
1.8	min xcorr	2.4	min xcorr	3	min xcorr	3	min xcorr
1	max mass	1	max mass	1	max mass	1	max mass
50	min initial rank	50	min initial rank	50	min initial rank	50	min initial rank
0.6	min ion match	0.6	min ion match	0.6	min ion match	0.6	min ion match
5000	min TIC	5000	min TIC	5000	min TIC	5000	min TIC
2	# tryptic termini	2	# tryptic termini	2	# tryptic termini	2	# tryptic termini

(Optional) Save filter to file.

| | ... |

(Optional) Load Name Keep Filter.

| | ... |

(Optional) Load Name Reject Filter.

The user uses standard Windows dialog boxes to select the directory containing the .OUT files and the location and filename of the output file. A standard set of filter setting is provided, and the user can alter these using the interface. The user can choose which charge states to use and select individual filter values for each charge state. The user can save these new filter values to a file using the Save filter to file option. These saved filter values can then be recalled with the Load a filter option. Several sample filters are provided. Additionally, there is an option to ignore the filter values and include all of the best matches from all of the .OUT files. Matches can also be additionally filtered based on keywords in the best match in the .OUT file. If the user provides a file of keywords in a text file (one keyword per line), using the Name Keep Filter option, all matches that contained the keywords. For example, a file containing the words kinase and phosphatase would cause all hits matching those keywords to appear in the output, regardless of the filter values. Similarly, the Name Reject Filter, will exclude all hits that match the keywords. This can be used to exclude common contaminants such as keratin or trypsin. The program can produce output in several different formats. Colon is the format used by the ZoomQuant application and is the default option. XML and SXML produce an XML like data file. Tab produces a tab delimited text file and Storable produces a binary representation of all of the data from all of the .OUT files. HTML produces a summary of the .OUT files suitable for viewing with a browser.   After the parameters have been entered, the user can click the 'Run summary' button to produce the results.
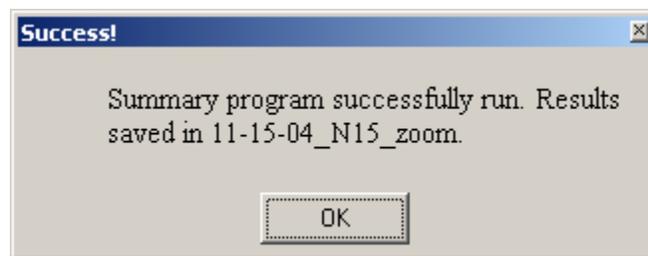
The program then indicates success and the filename of the output file. This file is then ready to be used with ZoomQuant.

ZoomQuant

   After displaying a 'splash screen', the ZoomQuant application opens the main window. This window allows the user to load and analyze data.

ZoomQuant v1.1 Academic

Select a zoom scan file and summary file

Zoom scan file: Zoom file ...

Summary file: Summary file ...

Load files

No peptide selected

Use in analysis

Get ratios

Use averagine

ratio labeled:unlabeled (Method 1)
ratio labeled:unlabeled (Method 2)
ratio labeled:unlabeled (Method 3)
efficiency (Method 1 data)

Report Format

HTML   Text   Tab seperated   Save report   Select report file   ...

The user loads the .zcn and .colon files using standard Windows dialog boxes. IT IS CRITICAL THAT BOTH THE .ZCN AND .COLON FILES ARE DEVRIVED FROM THE SAME .RAW FILE. The program will identify file mismatches and warn the user. Once both files have been identified, the user clicks the 'Load Files' button and the data are loaded to memory for better performance.

ZoomQuant v1.1 Academic

Select a zoom scan file and summary file

Zoom scan file: `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` [...]

Summary file: `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` [...]

[Load files]

sp|P02188|MYG_HORSE Myoglobin - Equus caball
sp|P46406|G3P_RABIT|std| Glyceraldehyde 3-phos
sp|P18031|PTN1_HUMAN|std| Protein-tyrosine pho
sp_vs|P16612-2|P16612|std| Splice isoform VEGF-A

No peptide selected

☐ Use in analysis

[Get ratios]   ratio labeled:unlabeled (Method 1)
                ratio labeled:unlabeled (Method 2)
☐ Use averagine ratio labeled:unlabeled (Method 3)
                efficiency (Method 1 data)

Report Format

◉ HTML ○ Text ○ Tab seperated   Save report `Select report file` [...]

ZoomQuant then lists the proteins have matches that passed the Epitomize filter in the left listbox. Clicking on one of the proteins brings up the list of scans matching that protein.

ZoomQuant v1.1 Academic

Select a zoom scan file and summary file

Zoom scan file `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` `...`

Summary file `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` `...`

Load files

---

sp|P02188|MYG_HORSE Myoglobin - Equus cabal
sp|P46406|G3P_RABIT|std| Glyceraldehyde 3-phos
sp|P18031|PTN1_HUMAN|std| Protein-tyrosine pho
sp_vs|P16612-2|P16612|std| Splice isoform VEGF-A

1530:1382.4076:2:5.2461:HGTVVLTALGGILK[:63:
1533:1382.5377:2:4.6260:HGTVVLTALGGILK[:63:
1340:1506.3445:3:4.5074:HGTVVLTALGGILKK:6S
1346:1506.6784:3:4.3326:HGTVVLTALGGILKK:6S
799:1606.6007:2:4.2850:VEADIAGHGQEVLIR:69:
1043:1360.3198:2:3.4834:ALELFRNDIAAK:61:101
790:1610.3702:3:3.4538:VEADIAGHGQEVLIR[:69
1040:1360.3305:2:3.3979:ALELFRNDIAAK:61:101
559:1271.2642:2:3.3021:LFTGHPETLEK:58:90:14
1540:1378.0494:3:3.1315:HGTVVLTALGGILK:63:
793:1610.2763:3:3.1262:VEADIAGHGQEVLIR[:69
638:941.0273:2:2.7302:YKELGFQG:44:64:10:13:0
1046:1360.3515:3:2.6875:ALELFRNDIAAK:61:101
635:940.9886:2:2.5931:YKELGFQG:44:64:10:13:0

No peptide selected

☐ Use in analysis

Get ratios

☐ Use averagine

ratio labeled:unlabeled (Method 1)
ratio labeled:unlabeled (Method 2)
ratio labeled:unlabeled (Method 3)
efficiency (Method 1 data)

Report Format

◉ HTML ○ Text ○ Tab seperated Save report `Select report file` `...`

The scan data in the center column consists of compressed data containing the scan number, mass, charge, xcorr sequence and atomic formula, sorted by xcorr. Clicking on one of the scans displays the scan data and prepares the scan for analysis.

ZoomQuant v1.1 Academic

Select a zoom scan file and summary file

Zoom scan file: `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` [...]

Summary file: `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` [...]

[Load files]

Left panel:
```
sp|P02188|MYG_HORSE Myoglobin - Equus caball
sp|P46406|G3P_RABIT|std| Glyceraldehyde 3-phos
sp|P18031|PTN1_HUMAN|std| Protein-tyrosine pho
sp_vs|P16612-2|P16612|std| Splice isoform VEGF-A
```

Middle panel:
```
1530:1382.4076:2:5.2461:HGTVVLTALGGILK[:63:
1533:1382.5377:2:4.6260:HGTVVLTALGGILK[:63:
1340:1506.3445:3:4.5074:HGTVVLTALGGILKK:69
1346:1506.6784:3:4.3326:HGTVVLTALGGILKK:69
799:1606.6007:2:4.2850:VEADIAGHGQEVLIR:69
1043:1360.3198:2:3.4834:ALELFRNDIAAK:61:101
790:1610.3702:3:3.4538:VEADIAGHGQEVLIR[:69
1040:1360.3305:2:3.3979:ALELFRNDIAAK:61:101
559:1271.2642:2:3.3021:LFTGHPETLEK:58:90:14
1540:1378.0494:3:3.1315:HGTVVLTALGGILK:63:1
793:1610.2763:3:3.1262:VEADIAGHGQEVLIR[:69
638:941.0273:2:2.7302:YKELGFQG:44:64:10:13:0
1046:1360.3515:3:2.6875:ALELFRNDIAAK:61:101
635:940.9886:2:2.5931:YKELGFQG:44:64:10:13:0
```

Right panel:

sp|P02188|MYG_HORSE Myoglobin - Equus caballus (Horse), and Equus burchell
HGTVVLTALGGILK[

scan number: 1530
mass: 1382.4076
charge: 2
xcorr: 5.2461
formula: C63 H112 N17 O17 S0

[ ] Use in analysis

[Get ratios]   ratio labeled:unlabeled (Method 1)
                  ratio labeled:unlabeled (Method 2)
[ ] Use averagine ratio labeled:unlabeled (Method 3)
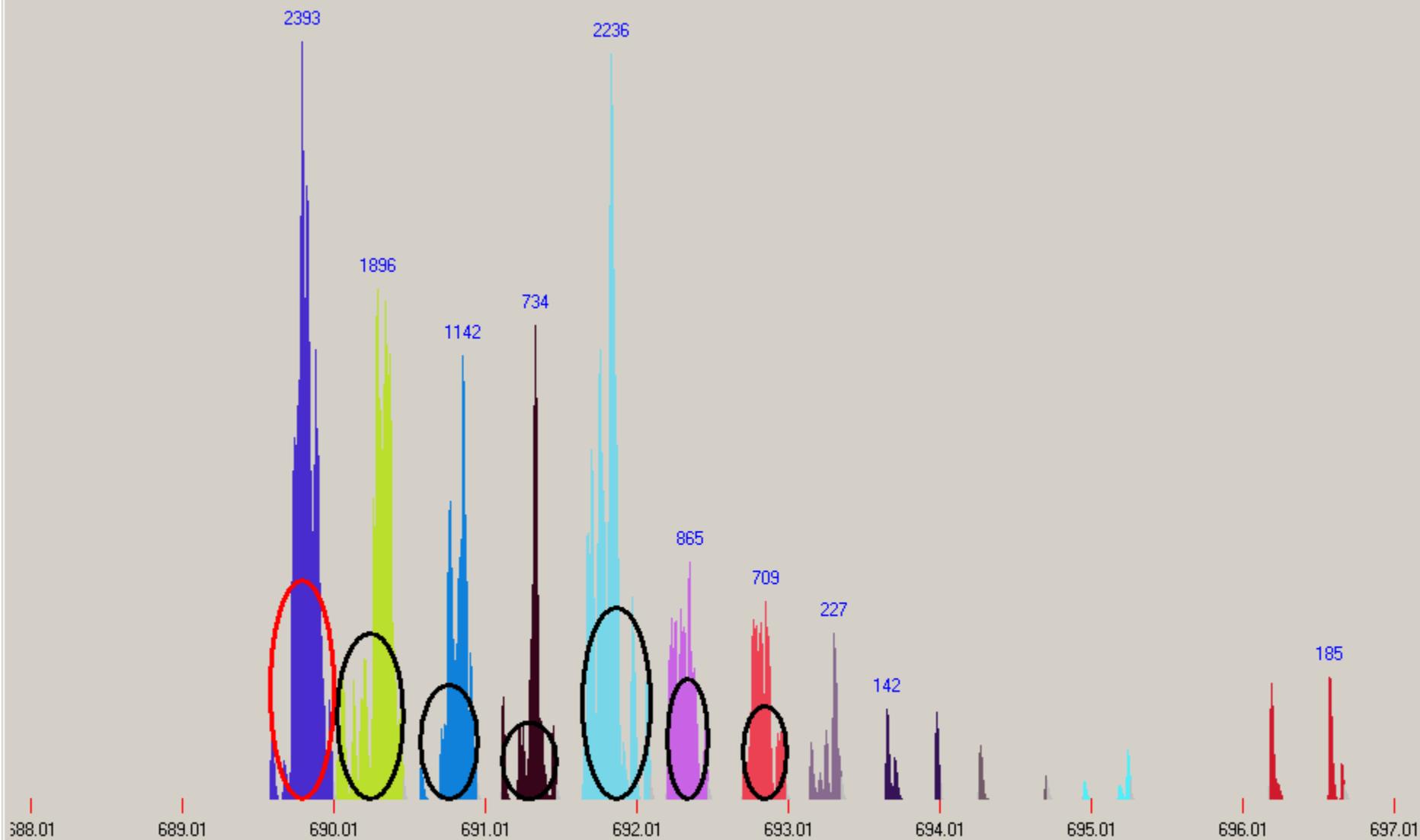   efficiency (Method 1 data)

Report Format

(•) HTML   ( ) Text   ( ) Tab seperated   Save report   Select report file   [...]

Clicking the 'Peaks' button opens a window that displays the spectra and peak identifications.

If the user accepts the peak assignments, they can click the 'Get ratios' button in the main window. This will calculate the ratio of labeled to unlabeled using three different methods. Method 1, is the method described in the ZoomQuant paper. Method 2 is the peak ratio method described by the Catherine Fenselau laboratory, modified to use peak areas in place of intensities. Method 3 is a similarly modified version of the 'averagine' method described by the David Muddiman laboratory. The 'efficiency' value is the ratio of $^{18}O_2$ to ($^{18}O_2 + {}^{16}O_1{}^{18}O_1$) and is a measure of how complete the exchange reaction was. The user then has the option of checking the 'use in analysis' box to have this scan included in the final report.

Select a zoom scan file and summary file

Zoom scan file `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` ...

Summary file `nd Settings/BHalligan/Desktop/Dist 6_02_05/Sample Data/4pro` ...

Load files

---

sp|P02188|MYG_HORSE Myoglobin - Equus caballu
sp|P46406|G3P_RABIT|std| Glyceraldehyde 3-phosp
sp|P18031|PTN1_HUMAN|std| Protein-tyrosine phos
sp_vs|P16612-2|P16612|std| Splice isoform VEGF-A

1530:1382.4076:2:5.2461:HGTVVLTALGGILK[:63:
1533:1382.5377:2:4.6260:HGTVVLTALGGILK[:63:
1340:1506.3445:3:4.5074:HGTVVLTALGGILKK:69
1346:1506.6784:3:4.3326:HGTVVLTALGGILKK:69
799:1606.6007:2:4.2850:VEADIAGHGQEVLIR:69:1
1043:1360.3198:2:3.4834:ALELFRNDIAAK:61:101:
790:1610.3702:3:3.4538:VEADIAGHGQEVLIR[:69:
1040:1360.3305:2:3.3979:ALELFRNDIAAK:61:101:
559:1271.2642:2:3.3021:LFTGHPETLEK:58:90:14:
1540:1378.0494:3:3.1315:HGTVVLTALGGILK:63:1
793:1610.2763:3:3.1262:VEADIAGHGQEVLIR[:69:
638:941.0273:2:2.7302:YKELGFQG:44:64:10:13:0:
1046:1360.3515:3:2.6875:ALELFRNDIAAK:61:101:
635:940.9886:2:2.5931:YKELGFQG:44:64:10:13:0:
964:752.2334:1:1.3081:ALELFR[:35:57:9:9:0:0:0

---

sp|P02188|MYG_HORSE Myoglobin - Equus caballus (Horse), and Equus burchell
HGTVVLTALGGILK[

scan number:	1530
mass:	1382.4076
charge:	2
xcorr:	5.2461
formula:	C63 H112 N17 O17 S0

☑ Use in analysis

Get ratios

☐ Use averagine

|	
ratio labeled:unlabeled (Method 1)	93.62%
ratio labeled:unlabeled (Method 2)	98.52%
ratio labeled:unlabeled (Method 3)	104.89%
efficiency (Method 1 data)	80.02%

Report Format

◉ HTML ○ Text ○ Tab seperated Save report Select report file ...

Some spectra are not accurately automatically identified. The Peaks window allows the user to assist in the identification and allow the spectra to be used in the final report. Two examples are shown below.
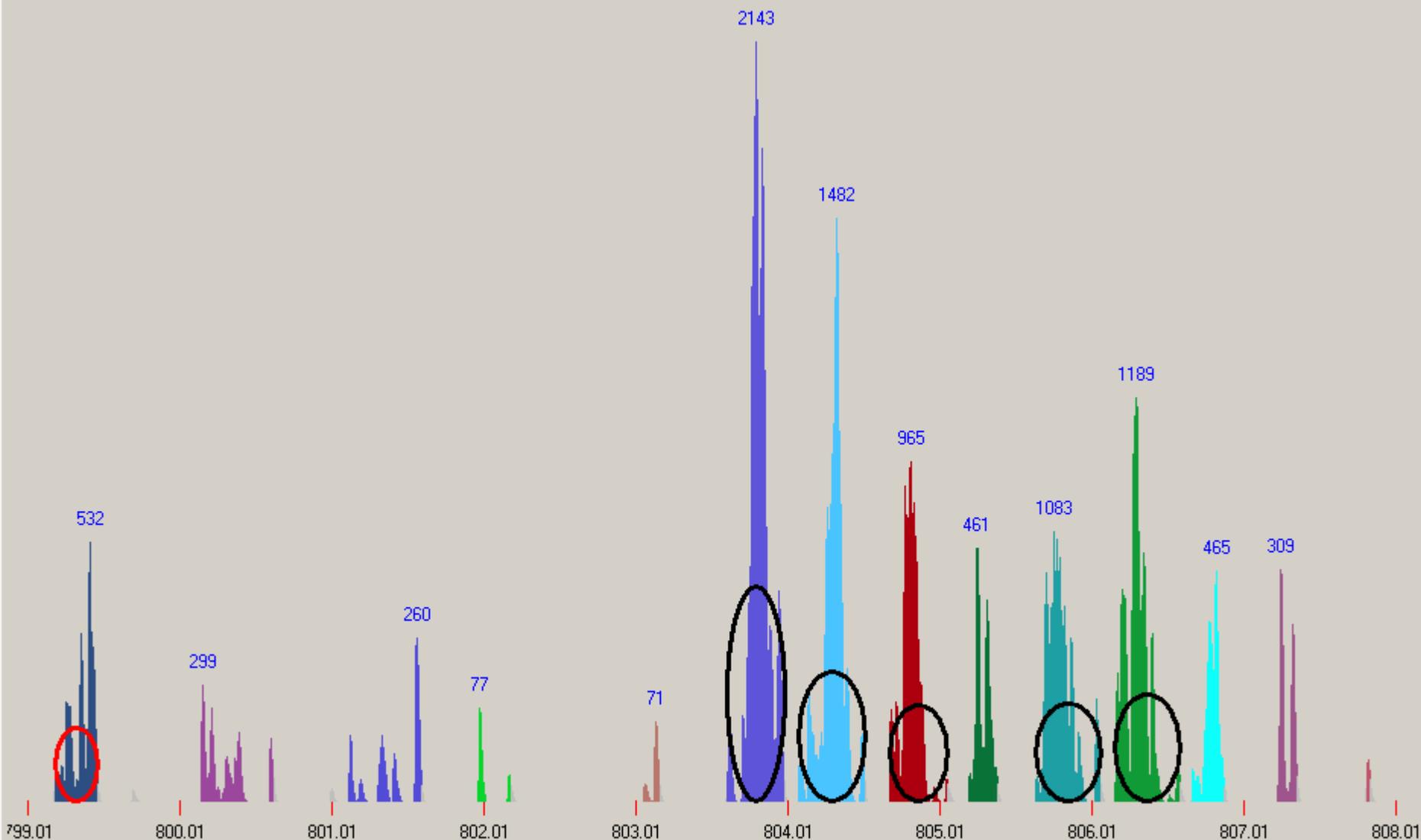
. Each of the six peaks used in the calculation are identified with an oval and these six peaks are referred to as the 'comb'. The number above each peak is the area of the peak. The area in each peak is color coded. In this spectrum, the program has misidentified the start of the isotopic envelope. There are two methods to correct this problem. One is to delete the misidentified peak and recalculate the comb. The radio buttons control the actions that take place when a peak is clicked on. Clicking on 'delete' and then the peak on the left with the red oval, grays it out and 'deletes' it from consideration.
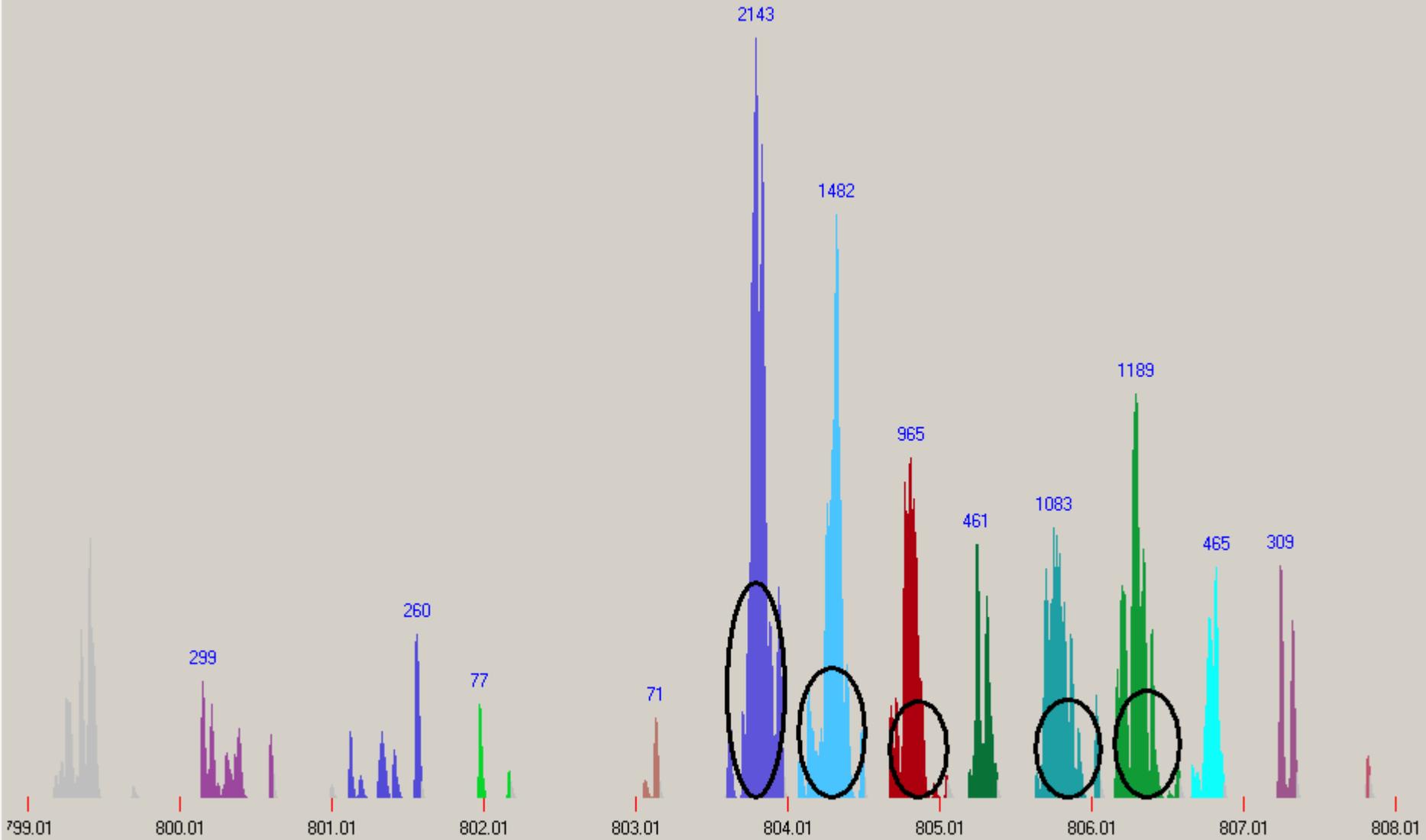
By then clicking on the 'Recalculate comb' button, chooses new peaks for the comb, ignoring that first peak.
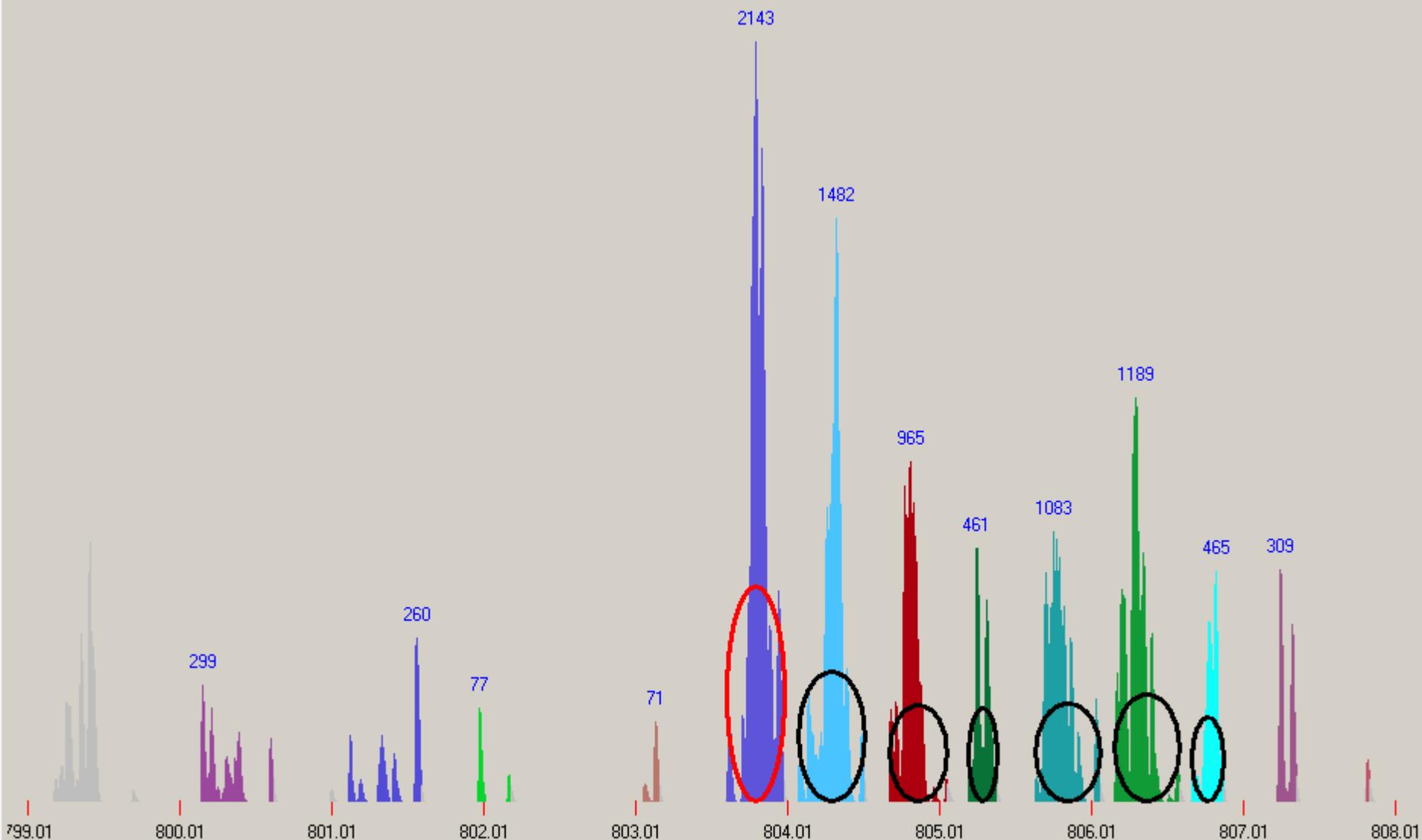
The numbered radio buttons on the bottom correspond to the peak position in the comb. The red oval indicates the selected position. Selecting a different comb number shows which peak is associated with each position. By selecting the 'comb' radio button and clicking on a peak, that peak will be assigned to the comb position selected on the bottom. Additionally, right clicking on any position in the spectra displays a ruler with spacing corresponding to the expected peak spacing based on the charge.
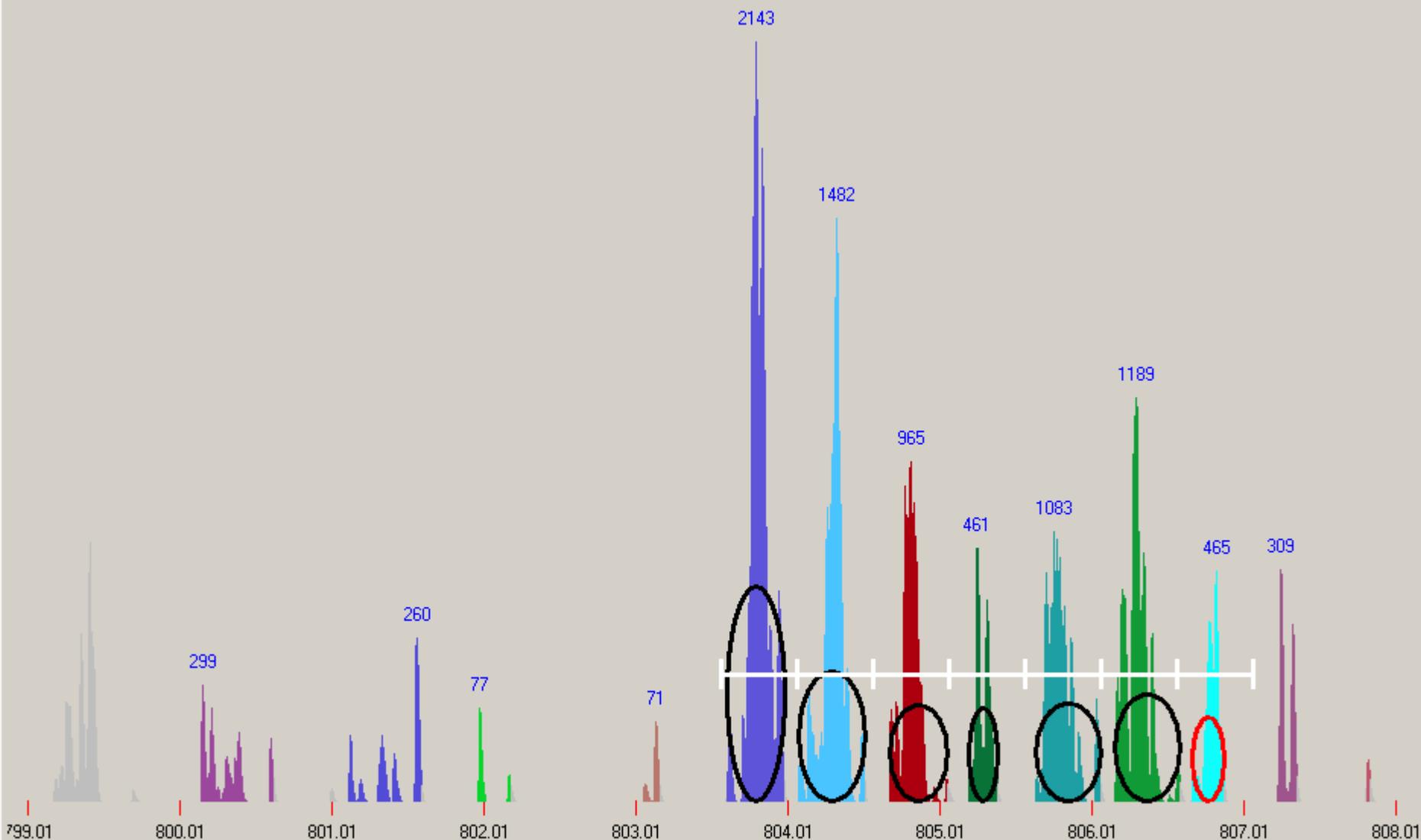
The 'split' and 'combine' radio buttons allow area to be removed or added to peaks. The 'nothing' radio button has no action associated with it and is the default. Clicking on 'nothing' provides a safe 'parking place' to avoid any change to the spectra. Some spectra cannot be quantitated. The 'Restore peaks' button removes the changes and restores the original peaks and comb. The 'Recalculate comb' button causes the program to re-identify the comb peaks based on the changes made by the user. The 'Delete small peaks' button can be used to automatically remove small background peaks from the spetrum.. The 'Save spectrum' button saves a high resolution postscript version of the spectrum suitable for use in publications.

The example below represents the carboxyl terminal peptide of the protein, and as expected, there is no label incorporated. Other problems that arise are multiple overlapping peptides and excessive noise. These are indicated by a ratio value of -100 and are automatically excluded from the reports.   A ratio of -100 can also be reported to indicate that the ratio of the first and second peak is too far out of agreement with the expected ^{13}C isotopic ratio based on mass and often indicates a misidentification by Sequest of the charge.
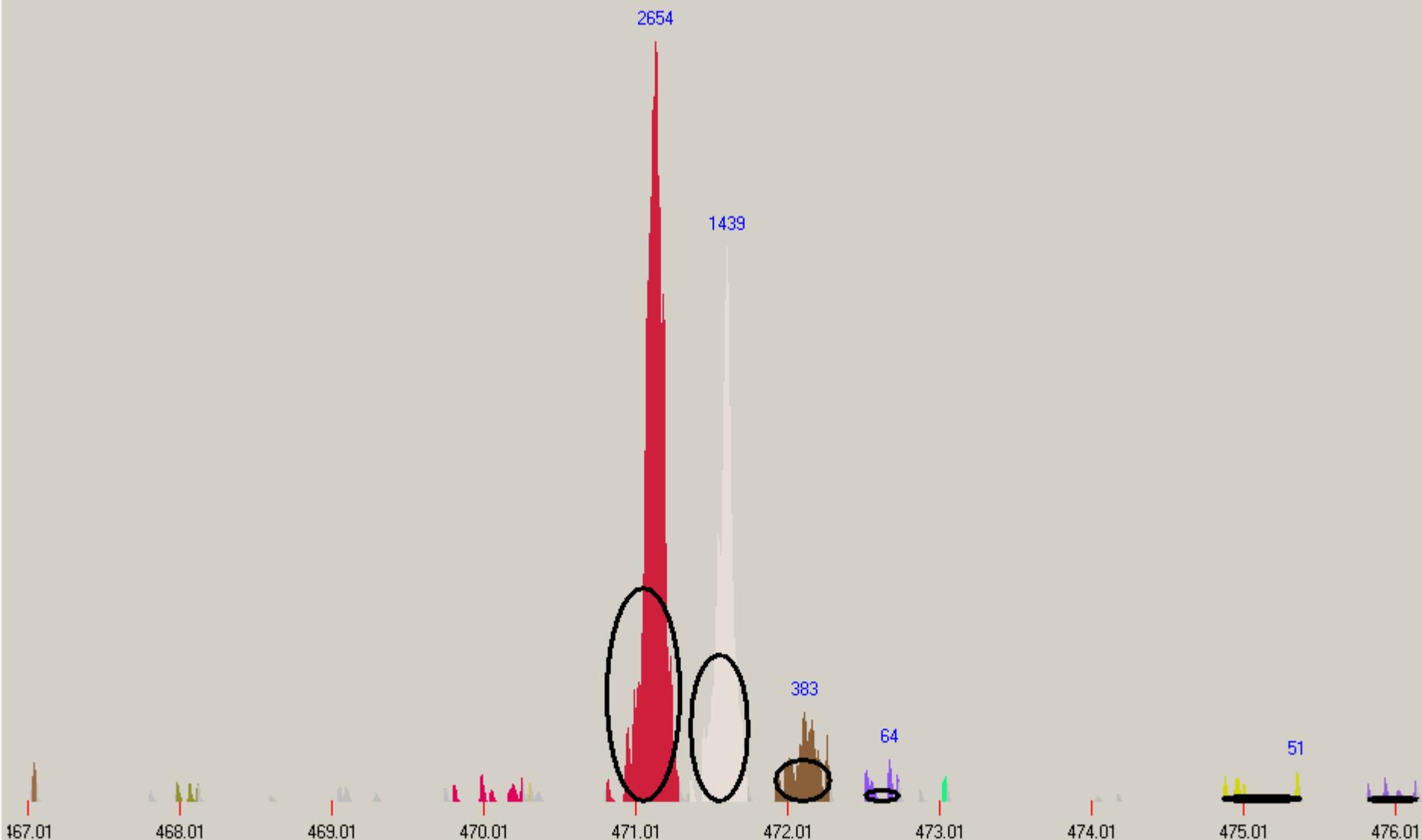
After each of the spectra has been reviewed and either rejected or chosen for inclusion in the report, a report can be generated from the main window by selecting the report format and using the file dialog box to enter a file name. The report can be formatted in HTML for viewing with a browser, a text format that can be incorporated in a text or Word document, or a tab separated text file (.tsv) that can be opened by a spreadsheet program such as Excel.

FAQ

Epitomize

Q.   After running epitomize, I got an empty file. What happened?

A.   Most likely there were no .OUT files in the directory you specified or the filter was too stringent for your data. Check the directory and rerun the program with the' ignore filter' box clicked. If the output file is not blank, repeat with different filter settings.

Q.   What is the Storable output option do?

A.   This option is reserved for future use. It saves the entire data from all .OUT files in a binary format.

RawBitZ

Q.   It seems to take RawBitZ a long time to run? What makes it so slow?

A.   The program has to extract each of the zoom scan spectra in the file. Each spectrum contains approximately 20,.000 data points. This is also why the .zcn files can become quite large.

Q.   Why won't RawBitZ run on computers that do not have the BioWorks software installed?

A.   RawBitZ was written using the Xcalibur Developer's Kit (XDK). A .dll file provided by the XDK is required to be installed for RawBitZ to run.

Q.   Will I be able to use mzXML files with ZoomQuant so that I can analyze data from different instruments?

A.   We are currently developing a replacement for RawBitZ that will read zoom scan information from mzXML. In theory, this will allow data from other sorts of instruments to be used with ZoomQuant.

ZoomQuant

Q.   What happens if I use a .zcn and .colon file from different .RAW files?
A.   Bad things – don't do it.

Q.   I have a blank name for a protein. What happened?
A.   We have occasionally observed this. It is usually due to some special characters in the description line of the protein in the fasta database used for the Sequest search. We have tried to take into account all possibilities, but if it does reappear, remove the special characters from the database and repeat the Sequest search and re-filter with epitomize.

Q.   The protein description seems truncated. How can I fix this?
A.   The truncation is done by Sequest.

Q.   The color of the peaks in the spectrum view change color when refreshed or recalculated. Why is this?
A.   The peak colors are assigned new random values each time they are refreshed. This is a 'feature' not a bug. Contact rslyper@umich.edu for more information.

Q.   I see too few peaks in the spectrum. What happened?
A.   First, keep in mind that the data are the data. The ZoomQuant program just displays the spectrum. That being said, there are two reasons to see too few peaks. One is that you are analyzing the C-terminal peptide in a protein. The trypsin labeling method does **NOT** label this peptide. A second possibility is that one of the two species (labeled or unlabeled) is below the limit of detection. This can be a result, but is incompatible with generating a ratio.

Q.   I see too many peaks in the spectrum. What happened?
A.   The spectrum can be noisy or there can be two or more co-eluting peptides. Sometimes it is possible to analyze these if they do not overlap, but it does require some judgment by the user.

Q. I am using 95% ^{18}O enriched water. Do I have to alter my calculations?
A. ZoomQuant assumes 95% enrichment and adjusts the labeled values accordingly. In reality, the differences in ratios between 95% and 99% enrichment is much smaller than experimental variation.

Q. I see reproducible differences in the ratios for different peptides from the same protein. Why is this?
A. We have also observed this and visual inspection of the data shows that this is not a calculation artifact but a real experimental result. It has been observed that partial digest peptides are often 'under' labeled. This may show a correlation between trypsin cleavage and trypsin labeling. When enough peptides have been analyzed in this way, we will attempt to apply the usual bioinformatics bag 'o tricks to understand this phenomena better.

Q. I would like to use ZoomQuant with data from different types of instruments. How can I do this?
A. We are currently developing and testing a replacement for RawBitZ that will extract zoom scan information from mzXML files. mzXML is a format developed by the Institute for System Biology and they have developed translators for several different instruments. Check back for more details.

Q. I would like to use ZoomQuant for a different type of isotopic labeling. Are there other versions of ZoomQuant in the works?
A. Yes, a version to analyze ^{15}N metabolically labeled peptides is currently being tested.

Q. I think that ZoomQuant is the neatest thing since sliced bread. Who can I contact for more information?
A. Brian Halligan, halligan@mcw.edu.

Q. I would like to distribute ZoomQuant as part of a commercial product. Who should I contact?
A. Contact me, Brian Halligan, halligan@mcw.edu, and I will put you in contact with the appropriate institutional officials.