MsMetrix

GC-MS Data Analysis using MsXelerator™

This document describes a few applications of GC-MS data analysis using MsXelerator. In principle, all tools and algorithms available for LC-MS data analysis can be applied to GC-MS data as well. See www.msmetrix.com for an extended overview. Among the technique and tools available are:

Browser:

- Import of GC-MS data formats (NetCDF, ASCII, mzXML).
- Link with third party software to import a large variety of GC/MS data formats.
- For LC/MS the following data formats are supported: Thermo Xcalibur, Waters MassLynx, Bruker, Agilent MassHunter, Sciex Wiff, ASCII, NetCDF and mzXML.
- Data Pre-Processing: Base Line correction, Smoothing, De-Spiking, Alignment etc.
- Basic Data Mining on single samples.
- Interactively mine your data: view TIC/BPC/EIC and Mass Spectra, create EIC overlay plots, perform peak purity analysis, and create contour and 3D-plots.

MPeaks:

- Peak Picking; find all significant chromatographic peaks in your sample.
- Special for GC/MS: Clustering - convert peaks into components – grouping based on co-elution.
- Identification: based on the parent ion use existing lists to find degradation products or metabolites. Can be used at any resolution. Link detected peaks with user defined libraries.
- Accurate Mass Tools: determine accurate m/z values, peak heights and areas if needed. Calculate charge states, monoisotopic mass etc.
- Differential Analysis. Find differences between two samples. Includes alignment correction if chromatographic shifts between samples are significant.
- Perform accurate Peak Picking on one sample and use the results to check if peaks are present in other samples. Export Peak Picking results for multiple samples to Excel and MsCompare.

IPeaks

- Isotope Pattern Matching: find peaks in your data matching the requested isotopic pattern. Create your own isotope patterns. Find peaks containing Chlorine, Bromine or other natural or synthetically introduced isotopes in complex samples.
- Use MsX-Quant to perform relative quantitative analysis based on labeled peaks.
- Use any type of isotopic ratio analysis.

MsCompare:

- Handling of series of samples for BioMarker Discovery, Metabolomics, Stress Testing, Batch Comparison etc. Create groups of samples and find significant differences between those groups.
- Create TIC/BPC, Contour plots for many samples simultaneously in overlay or stacked mode. Interactively extract EICs for all samples simultaneously. Easy viewing of MS spectra and Mass Chromatograms at any resolution.
- MsCompare is linked to MPeaks and IPeaks. Import results from MPeaks Peak Picking into MsCompare.
- MsCompare can import and process 1-D data, e.g. single chromatograms or NMR spectra.
- Multivariate Analysis based on: TIC/BPC or EIC or a peak picking result table from MPeaks or from a Biomarker Map created in MsCompare. Includes Principal Component Analysis (PCA), Clustering, Correlation Maps and a number of statistical tests (t, p, ratio, uniqueness etc).
- View integrated peak results (area or peak height) for many samples in one view.

- MsCompare has five alignment algorithms to correct for shift differences between samples: Offset Shifting, Stretch & Shrink, Cross Correlation, Reference Peak Warping and Correlation Optimized Warping.
- MsCompare has 6 different normalization algorithms.
- Perform Biomarker Discovery and find differential peaks between groups of samples using 2D LC/MS (GC/MS) Search Techniques. Find unique peaks in one group not present or up-regulated in the other group.
- Export results to Excel, Text files or Matlab.

Browser:

Figure 1 shows the Browser (portal of MsXelerator) with a typical GC/MS sample loaded. The bottom window displays the TIC/BPC, the top window the MS spectrum at the selected retention time and the middle window shows the extracted ion current (EIC) for the m/z peak with the largest intensity, in this case m/z 175. Typically, for GC/MS many peaks are visible in the EIC (middle window). This is due to the fact that a lot peaks have the m/z 175 fragment in common.

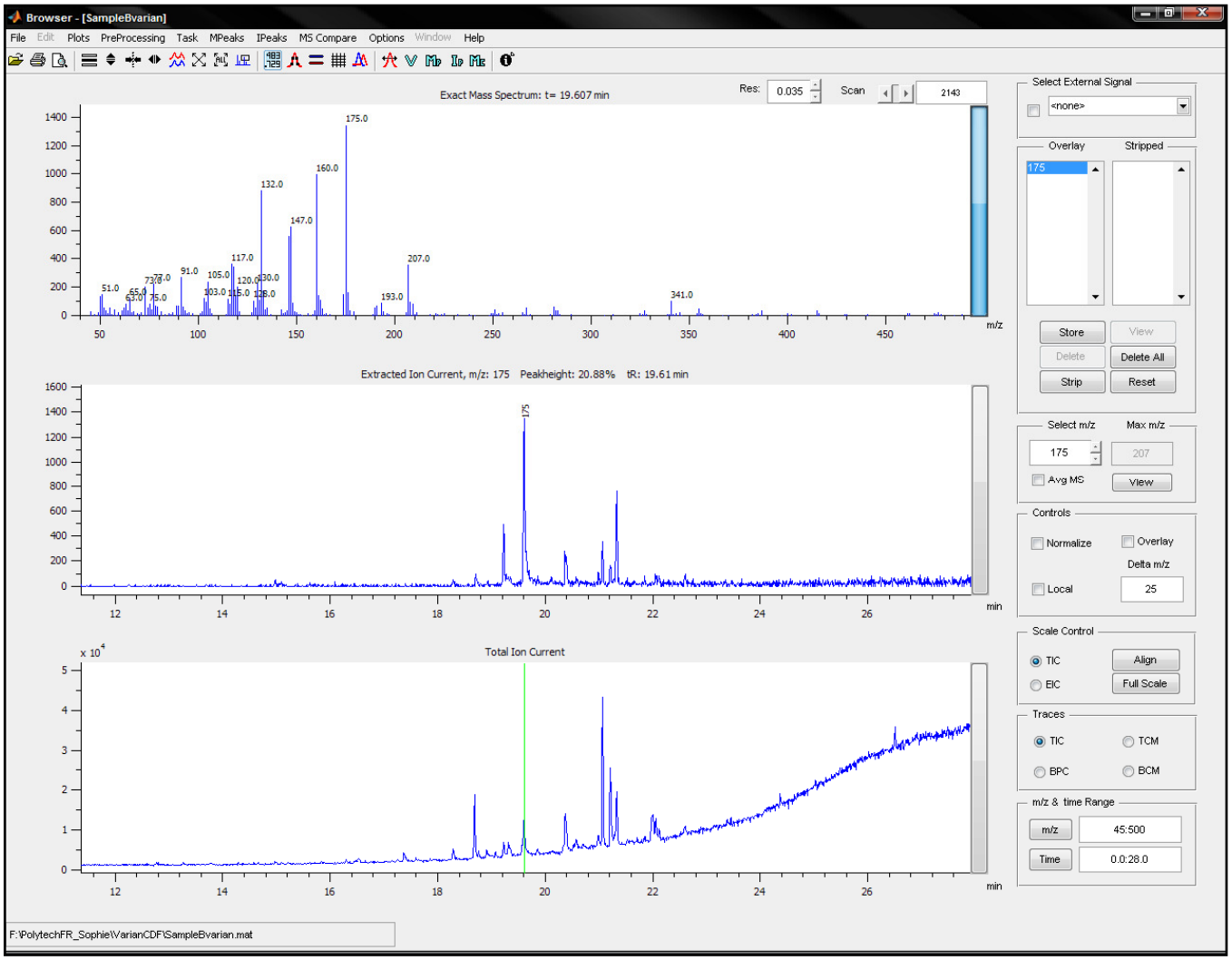


Figure 1: Browser overview: from bottom to top – TIC/BPC, EIC and MS Spectrum at selected retention time

Within the Browser you can overlay different EICs to see whether the peak is pure (Browser Viewer), subtract selected masses from the TIC (Stripping), correct for the large baseline present and determine basic peak picking parameters needed for MPeaks, e.g. estimate peak width, noise levels etc.

MPeaks:

To detect all significant peaks, run MPeaks. Define a few basis parameters like the estimated width of your peaks, the threshold (% compared to largest peak in the data set) and the signal to noise level. Figure 2 shows the MPeaks output. In total 861 peaks have been detected. MPeaks is very fast, typically a few seconds.
MPeaks will detect all significant peaks. This includes isotopes, fragments, adducts etc. In figure 2, the red trace is the original TIC, the blue trace is the reconstructed TIC based on MPeaks peak detection. Noise and background baseline have effectively been removed.
All detected peaks are presented in a so-called result table, shown in Figure 2. Clicking on an entry will plot the Extracted Ion Current for the selected m/z value. A total of 861 peaks were detected.

Clustering of fragment peaks into components:

For GS/MS data, MPeaks has a powerful Clustering algorithm. It will group co-eluting peaks into clusters based on Correlation Analysis. Basically this means that fragments will be grouped into clusters. A single cluster can be defined as a component.

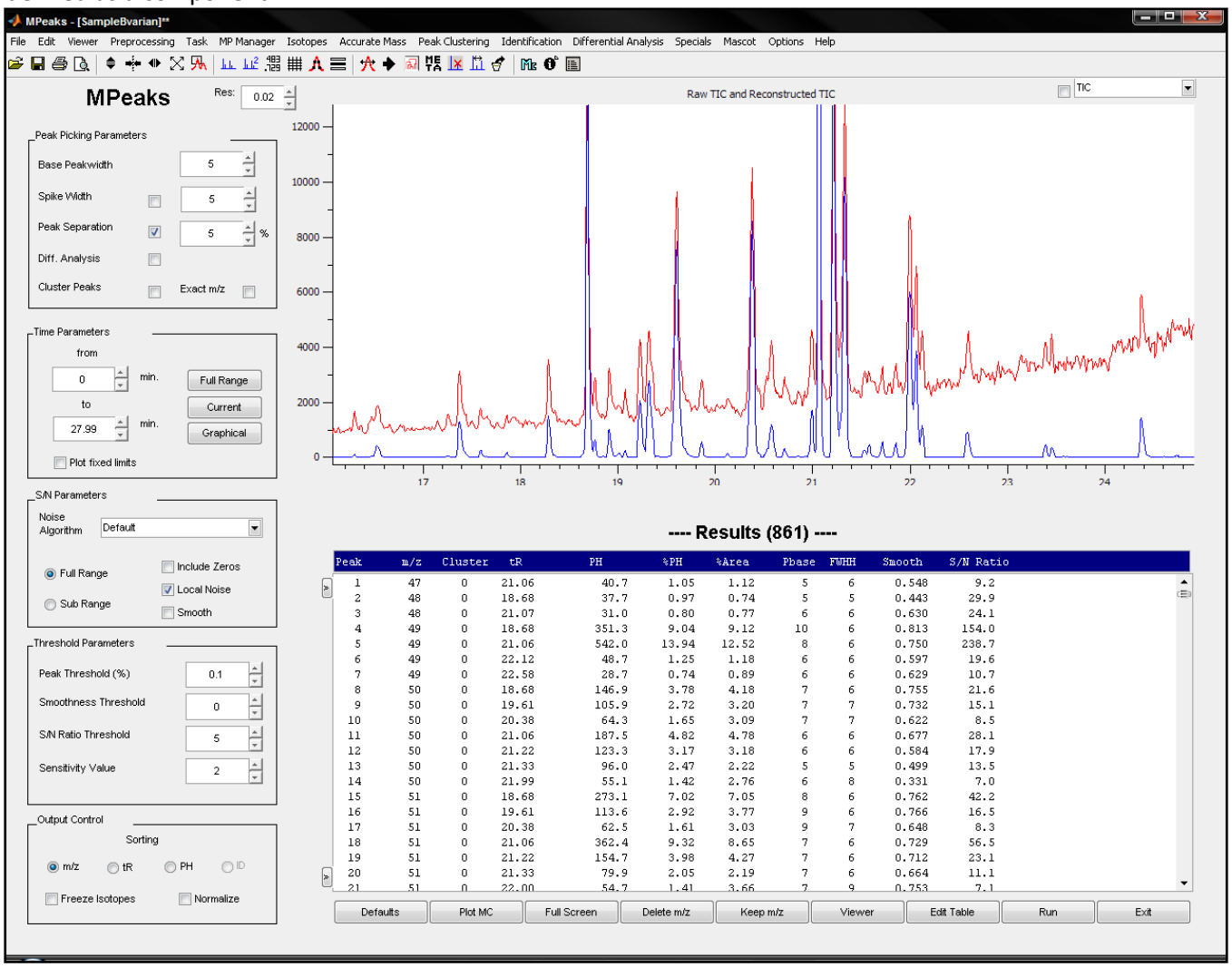


Figure 2: MPeaks output screen

In total 98 clusters were detected. An example of a number of peaks falling into the same cluster is shown in Figure 3. The overlay plot of selected EICs marked in the table (Cluster 62, all eluting at 21.06 minutes) is shown in the top window. It can be seen that all peaks perfectly co-elute.

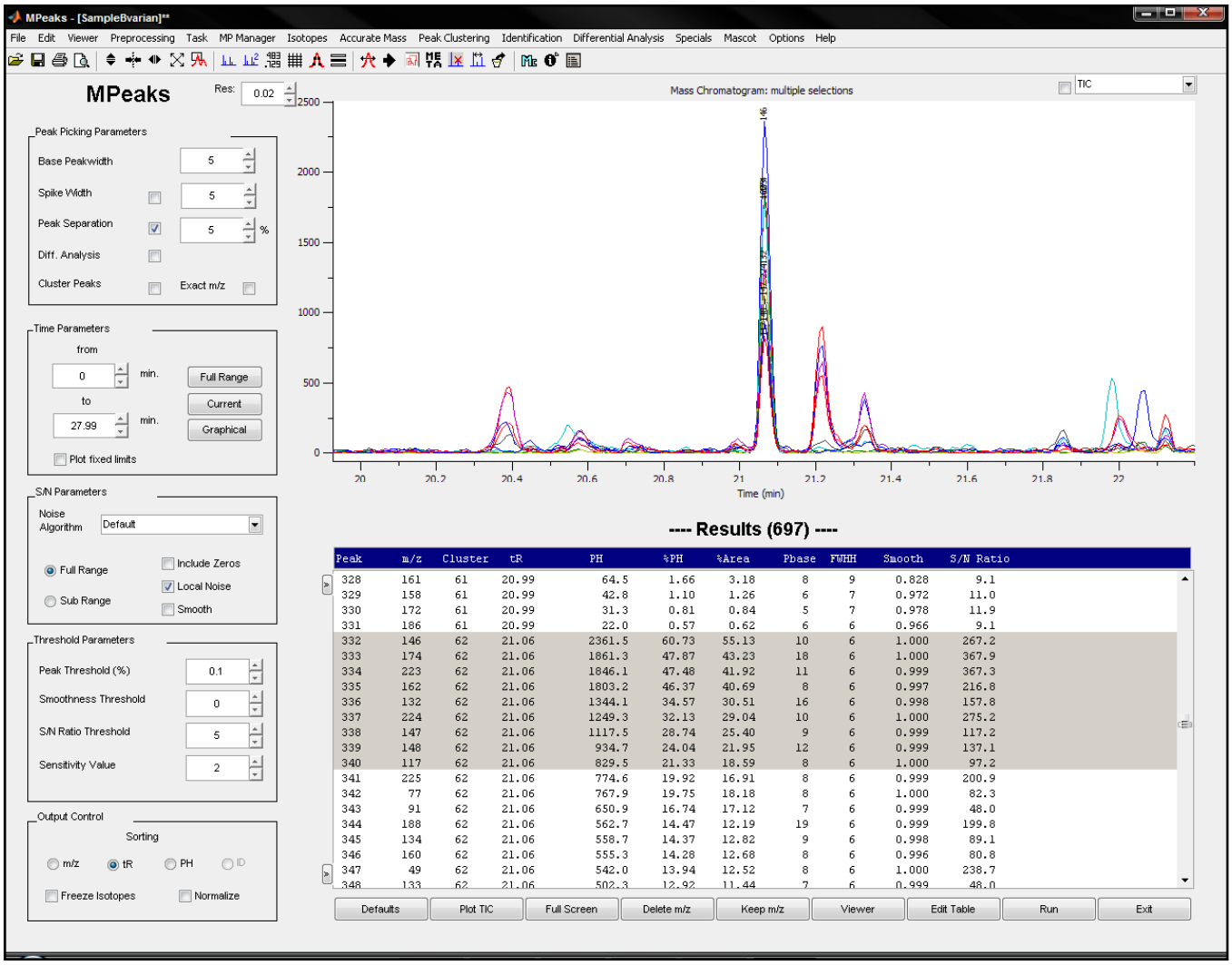


Figure 3: MPeaks result after clustering. Shown is the overlay of some fragments from cluster 62

MPeaks can remove all smaller fragments and only keep the largest from each cluster. Sometimes, this makes further data analysis simpler. From a selected peak you can easily switch to mass spec plotting. You can plot the original Mass Spectrum or the Reconstructed Mass Spectrum bases on the peaks belonging to the same cluster.

Peak Purity Analysis / Deconvolution:

Based on Clustering, MPeaks can determine if chromatographic peaks are pure or not. In some other software packages this is called deconvolution or peak purity analysis.

The red curve in Figure 4a shows the original TIC zoomed around a peak for which we want to check the purity. The blue trace is the reconstructed TIC, based on MPeaks Peak Picking. Peak Picking was performed on the selected retention time window. After Peak Picking, 3 clusters were detected, which means that the peak is definitely not pure. The results are shown in figure 4a and 4b. It is clear the peak from Figure 4a has 3 closely eluting peaks, of which the summation seems to resemble a pure peak.

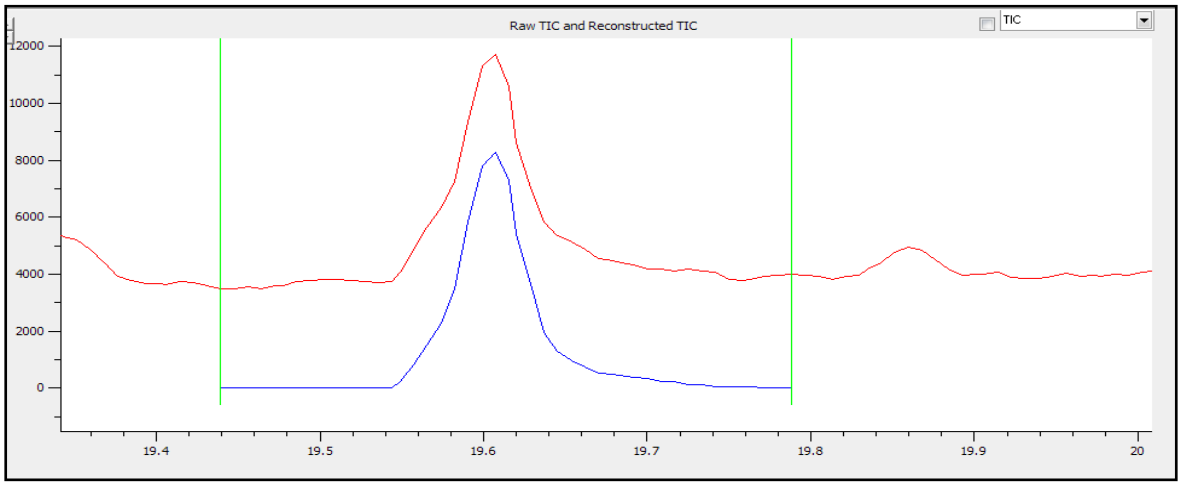
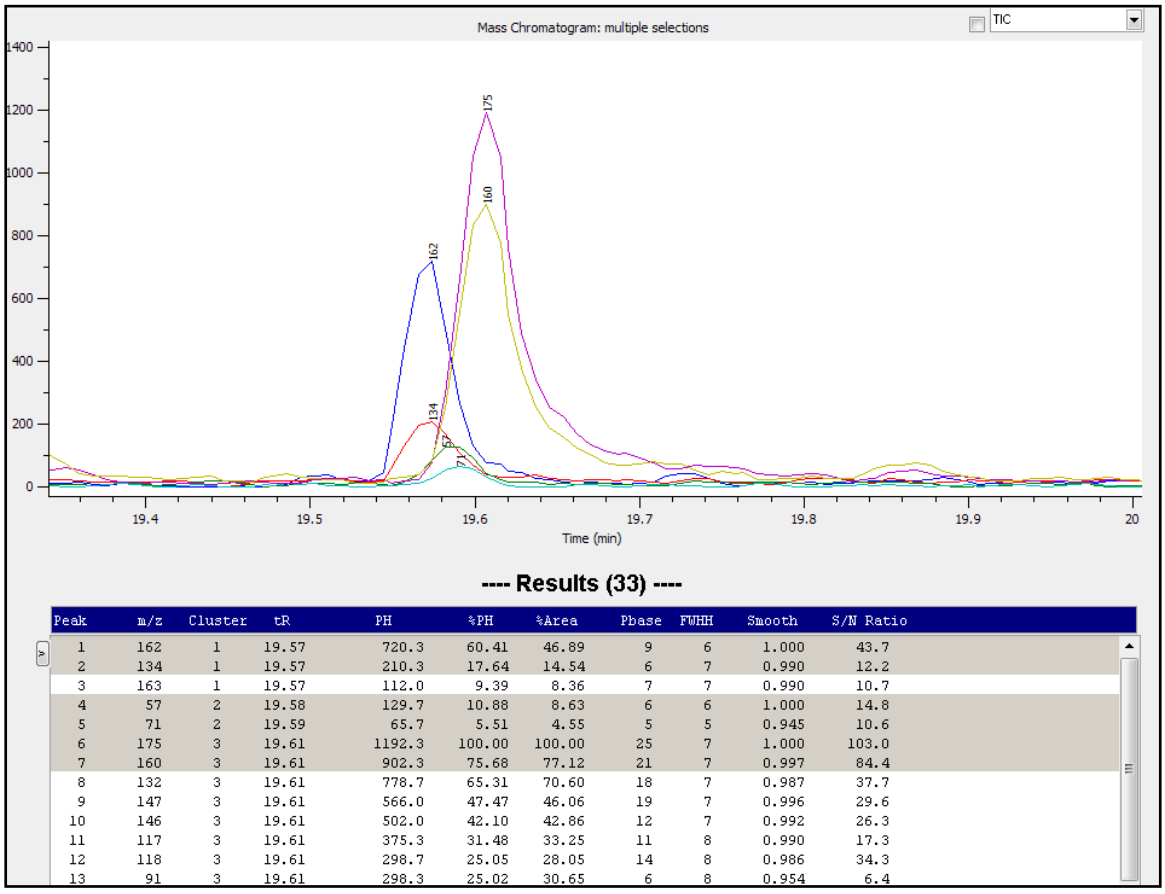


Figure 4a: TIC (red) and Reconstructed TIC (blue) of a selected peak used for Peak Purity Checking



---- Results (33) ----

Peak	m/z	Cluster	tR	PH	%PH	%Area	Pbase	FWHH	Smooth	S/N Ratio
1	162	1	19.57	720.3	60.41	46.89	9	6	1.000	43.7
2	134	1	19.57	210.3	17.64	14.54	6	7	0.990	12.2
3	163	1	19.57	112.0	9.39	8.36	7	7	0.990	10.7
4	57	2	19.58	129.7	10.88	8.63	6	6	1.000	14.8
5	71	2	19.59	65.7	5.51	4.55	5	5	0.945	10.6
6	175	3	19.61	1192.3	100.00	100.00	25	7	1.000	103.0
7	160	3	19.61	902.3	75.68	77.12	21	7	0.997	84.4
8	132	3	19.61	778.7	65.31	70.60	18	7	0.987	37.7
9	147	3	19.61	566.0	47.47	46.06	19	7	0.996	29.6
10	146	3	19.61	502.0	42.10	42.86	12	7	0.992	26.3
11	117	3	19.61	375.3	31.48	33.25	11	8	0.990	17.3
12	118	3	19.61	298.7	25.05	28.05	14	8	0.986	34.3
13	91	3	19.61	298.3	25.02	30.65	6	8	0.954	6.4

Figure 4B: results from peak picking in selected retention time window. Plotted are a few peaks from different clusters, demonstrating that the peak in not pure.

Differential Analysis:
To find differences between two samples, perform MPeaks Differential Analysis. The first step is to run peak picking on the sample, next select a reference or control and check which peaks are different or absent. The software will calculate all ratios (peak height and area) between peaks present in both samples. You can also run differential analysis on a series of control samples. The output of peak picking and ratio analysis can be imported into MsCompare for a better view of multiple samples.

Figure 5 shows the results of Differential Analysis. Plotted in overlay are the EICs of m/z 162 for the sample (blue) and control (red). About 10 peaks are detected in the range displayed.
Only one peak at 21.06 minutes is also present in the control sample. The calculated ratio is 1.35, and the peak is marked as being present in the reference/control sample. All other peaks for m/z 162 have large ratios, which means that they are absent in the control. This is confirmed by a visual inspection.

After differential analysis you can sort on ratios or decide to keep only those peaks that are really different. Differential Analysis can be run at different resolutions if necessary.

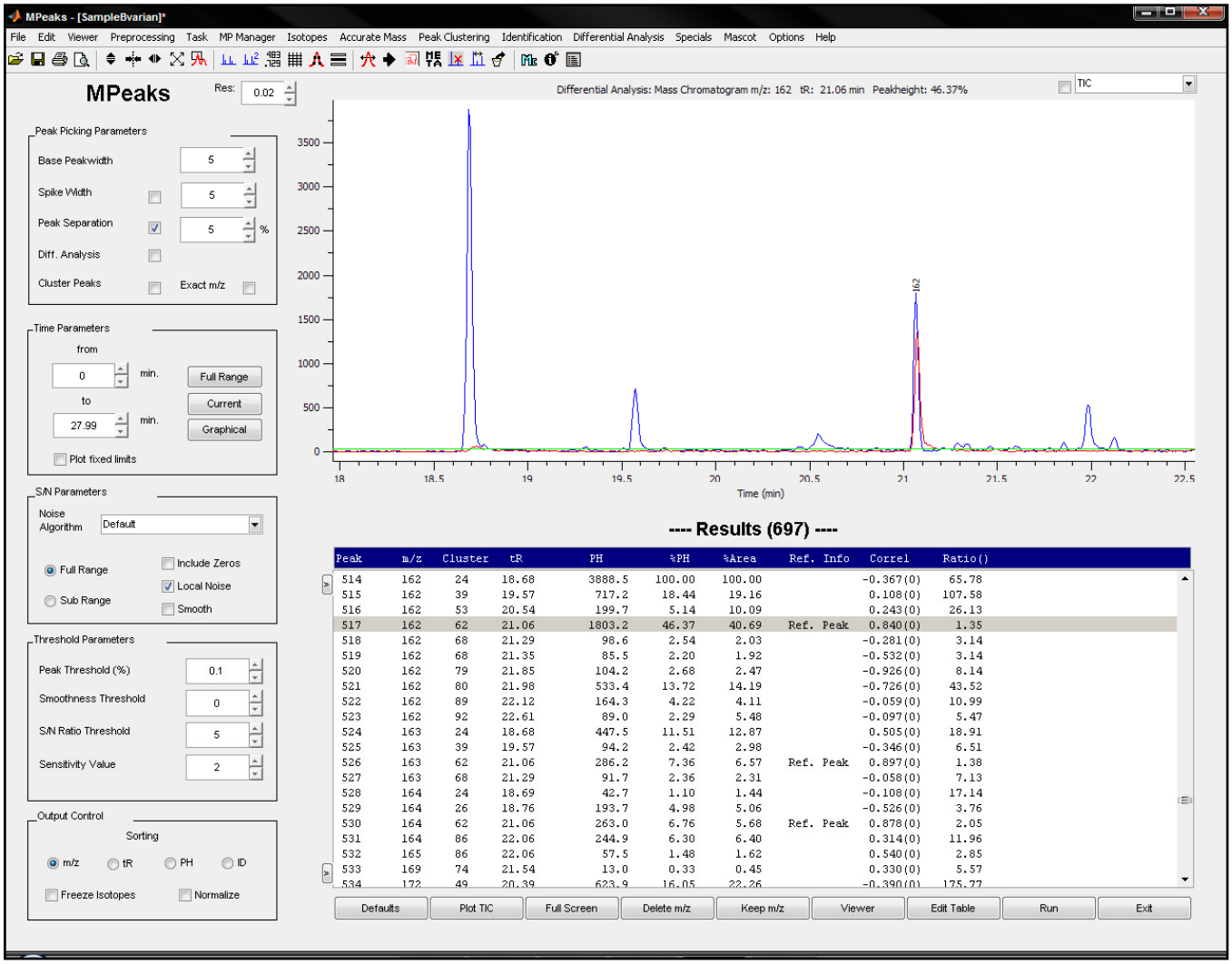


Figure 5: MPeaks Differential Analysis showing peaks that are different between sample and control

IPeaks: Isotope Pattern Matching

If your samples contain Chlorine or Bromine or peaks can be recognized by a specific isotopic pattern, you can use IPeaks to only find those peaks have the defined isotopic pattern. IPeaks is very powerful for detecting the few peaks having the specific pattern from all the background material.

IPeaks comes with a number of pre-defined isotopic pattern, but the user is free to create a pattern on the fly. Just define the mass difference (nominial or accurate) and the expected theoretical ratio between two isotope peaks.

Figure 6 shows the original TIC for the Sample and the Reconstructed TIC from IPeaks. The IPeaks search was performed to detect Chlorine containing peaks having a mass difference of 3 and a ratio of 3:1.
As can be seen only a few peaks seem to have the Chlorine pattern. After Isotopic Pattern Matching the user can run clustering in case of GC/MS data.

The detected peaks that have the correct pattern will be listed into a result table, similar to MPeaks. Tables can be edited, sorted, exported to text files or excel etc. etc.

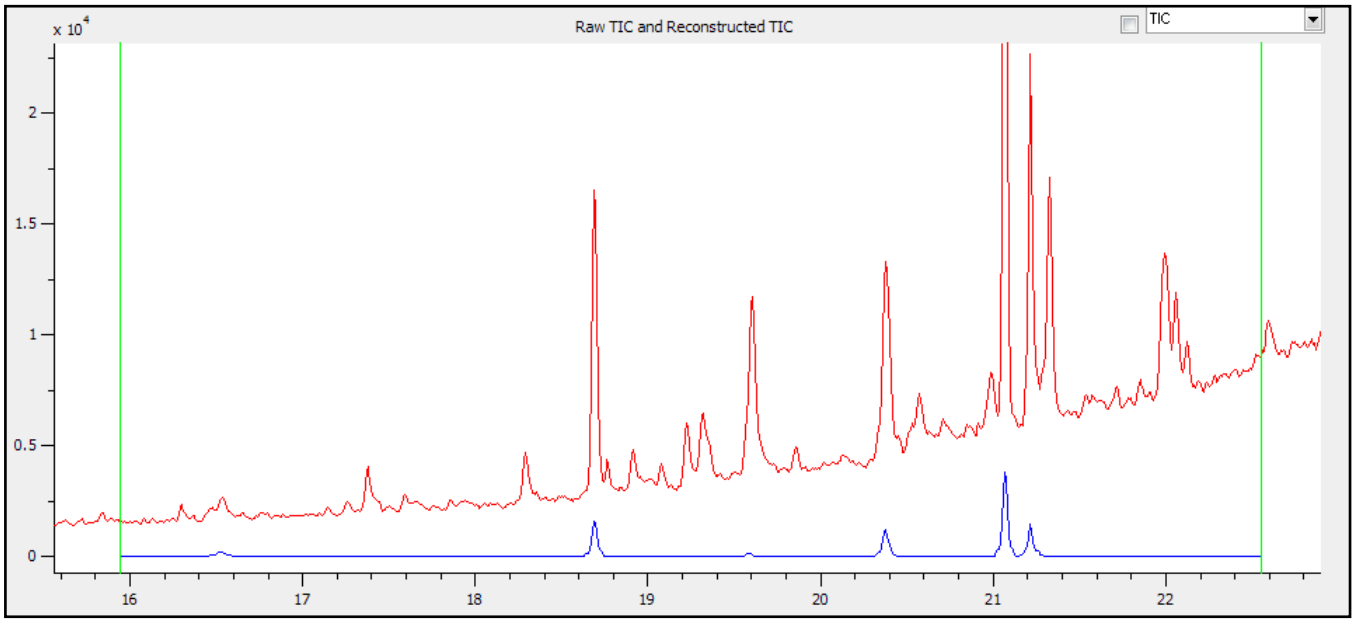


Figure 6: IPeaks Output: showing the original TIC in red and the Chlorine Reconstructed TIC from the IPeaks algorithm in blue.

MsCompare: BioMarker Discovery – Metabolomics – Batch Comparison

If you have more than 2 samples to compare or you have groups of samples, use MsCompare. MsCompare has many methods/algorithms and graphical tools to find difference between groups of samples. If you don't have groups, but just want to compare many samples at the same time, MsCompare is great.

MsCompare also has Multivariate Analysis like PCA, BioMarker Surface Maps, Clustering and Correlation Maps to detect differences between a large number of samples. All data, Mass Chromatograms or Mass Spectra can easily be observed in different viewing modes (overlay, stacked, contour, 3D).

Below some representative plots will be shown.

Overview Plot:
Figure 7 shows the overview plot of MsCompare. Loaded are 25 samples divided into two groups (5 Samples and 20 Controls). The Bottom view shows the Total Ion Currents in Stacked Plotting Mode. The top window shows extracted ion currents for m/z 471 for all samples simultaneously.

Clicking on a peak in the TIC window will automatically extract the EICs for all samples. MsCompare provides easy switching to Mass Spectrum Plotting. EIC can also be entered manually in the EIC edit box. You can provide nominal or accurate mass values.

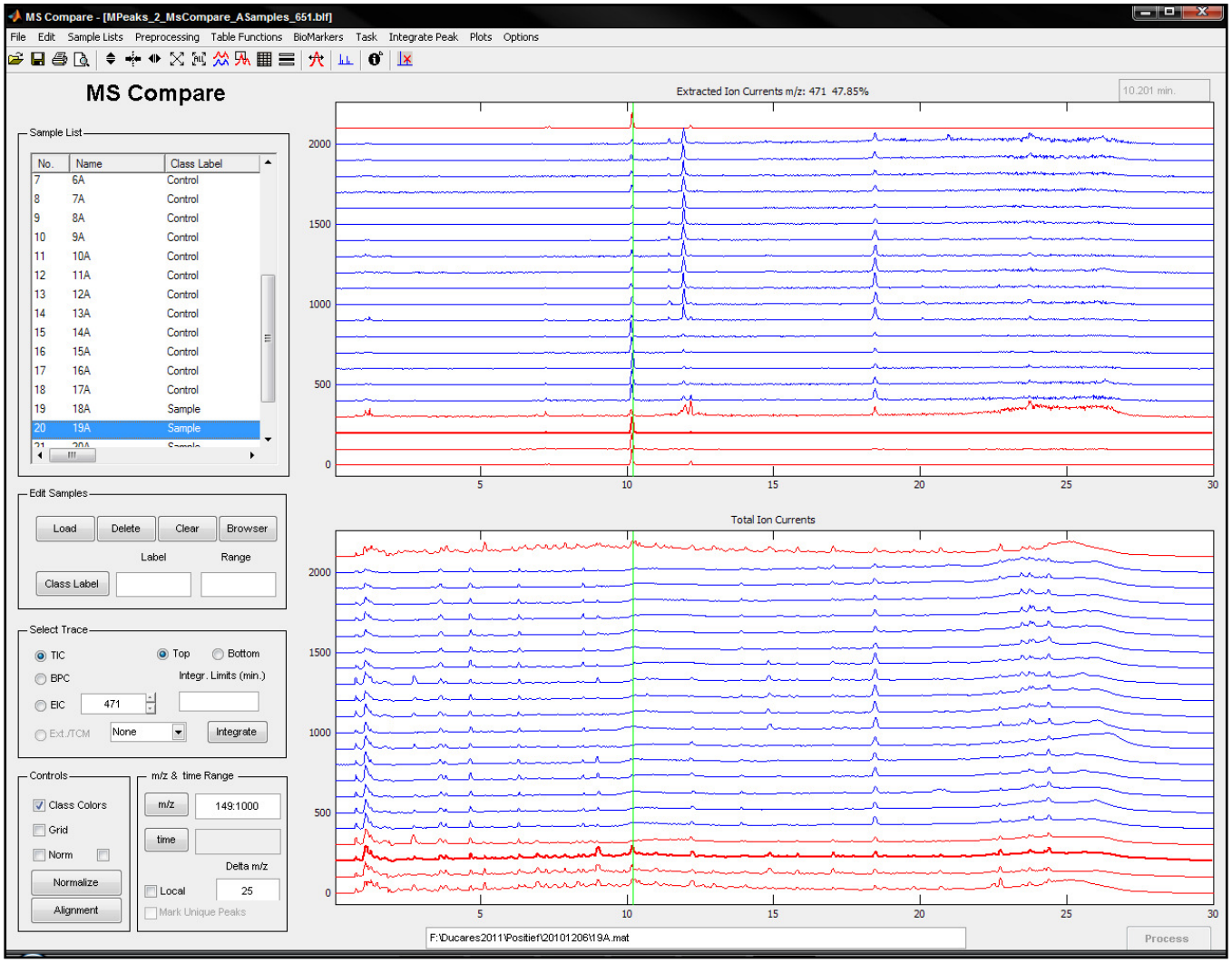


Figure 7: MsCompare overview Plot (TIC and extracted EICs) in stacked mode.

Loading and viewing MPeaks Peak Picking results in MsCompare:

You can perform Peak Picking for all samples simultaneously using MPeaks. Next, load the result into MsCompare. Figure 8 show the MPeaks result table (Peak ID, retention time, accurate m/z and all peak intensities based on accurate extracted EICs).

Clicking on any entry in the table will plot the EICs to the bottom window. Selected samples from the table, the plot and the sample list box are connected, to easily see what is what.

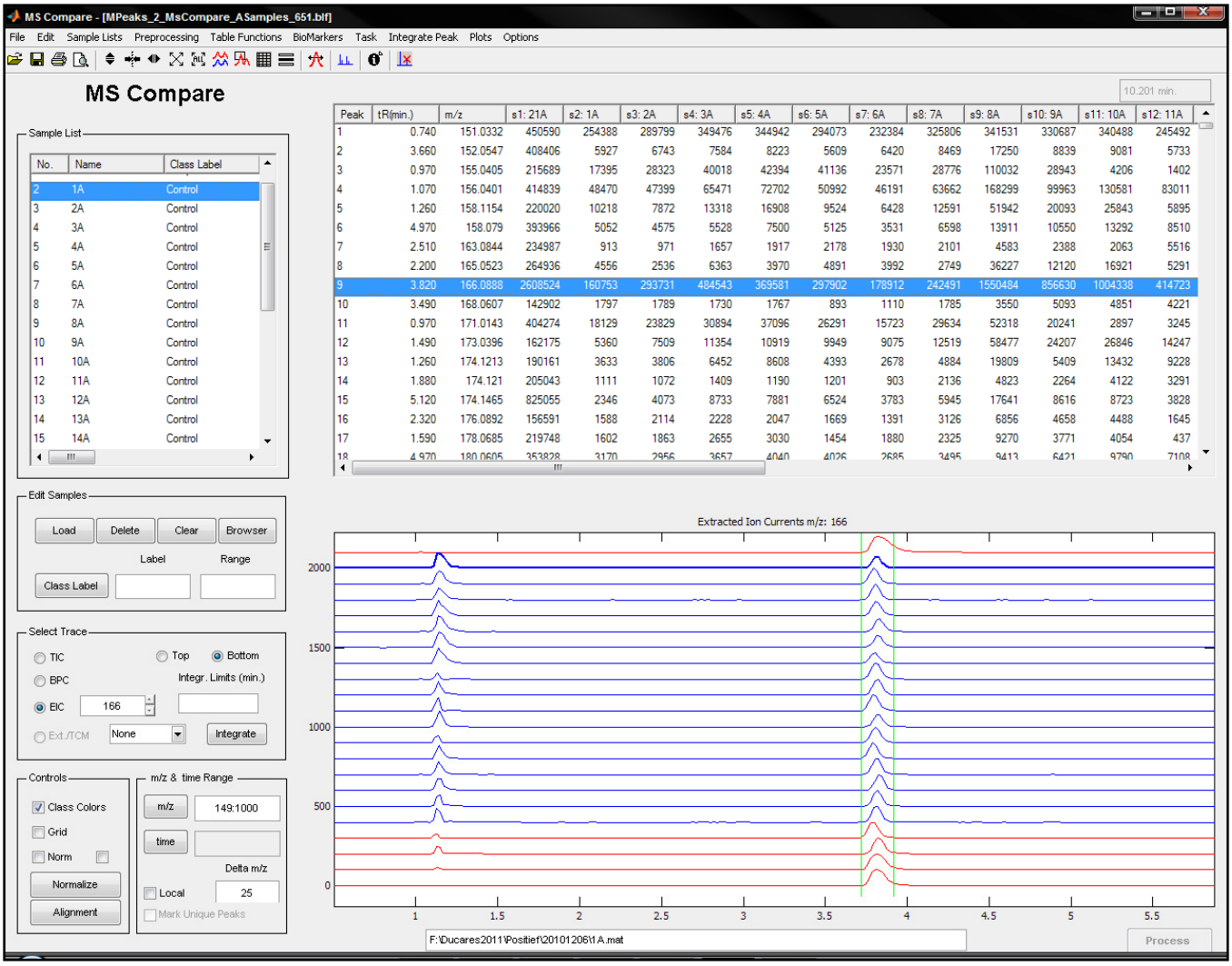


Figure 8: MPeaks result table loaded and viewed (EICs) in MsCompare

Extracting Peak Profiles for all samples:

Figure 9 shows a so-called peak profile plot for two selected peaks (m/z 166.08 eluting at 3.8 minutes and m/z 181.12 at 16.54 minutes). The Profile plots show the peak intensities as a function of sample number. For kinetic studies one can easily create different profile plots to see the behavior of different components at the same time. Profile plots can also be created directly in MsCompare, by selecting the m/z value and the retention time window for peak integration.

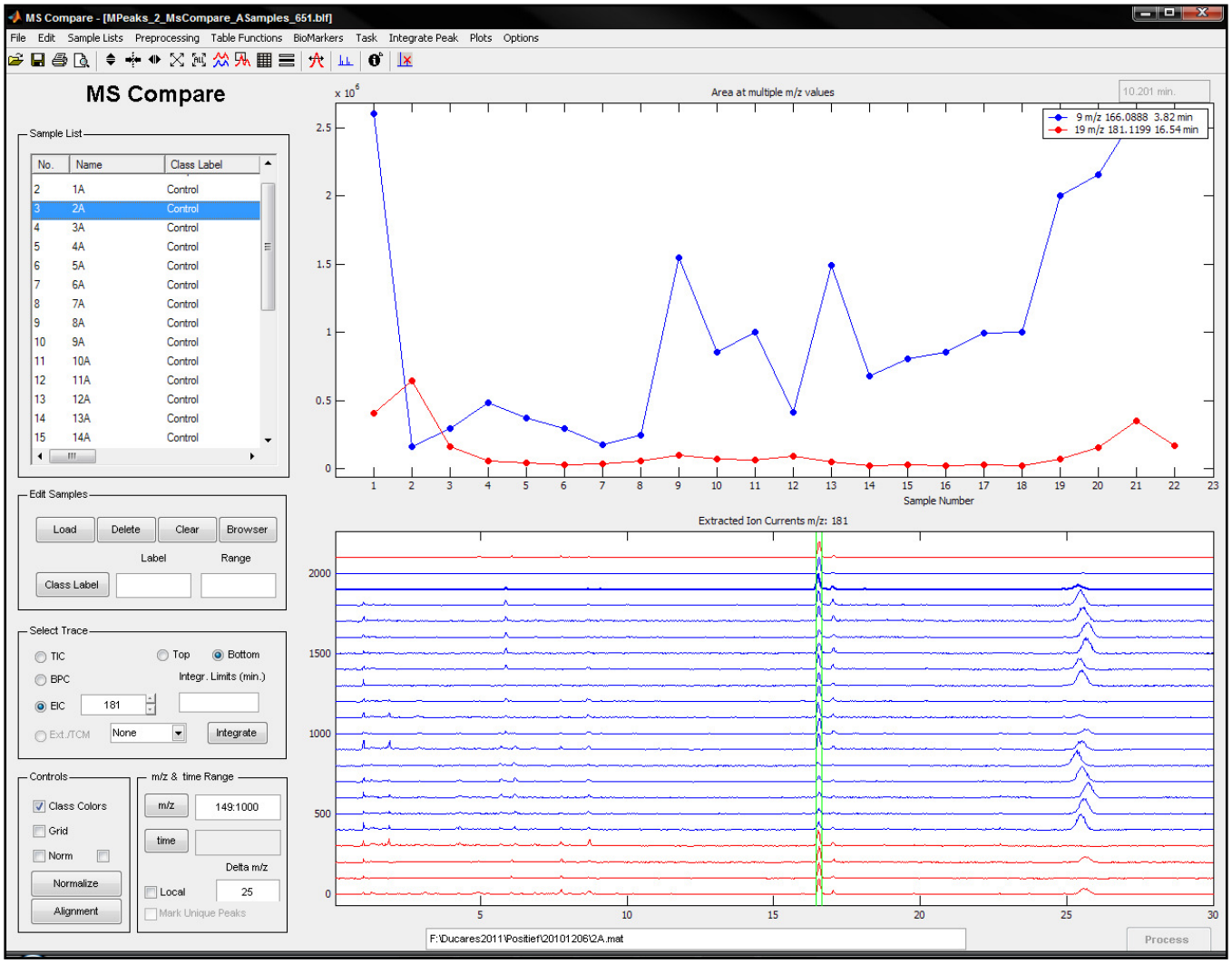


Figure 9: MsCompare Peak Profile Plot (top) and EIC of selected entry from the table (m/z 181 at 16.5 min.)

Multivariate Data Analysis:

MsCompare contains a number of Multivariate Analysis Techniques: PCA, Clustering, BioMarker Surface Maps and Correlation Maps. These algorithms can be used to detect differences and similarities between (groups of) samples. The algorithms can be applied to TIC/BPC/EIC Traces, Mass Spectra, loaded MPeaks result tables or result tables created from BioMarker Surface Maps.

Figure 10 shows PCA and Clustering applied to the full length TICs of all loaded samples. From the PCA score plot we can see that the samples can be discriminated from the control samples. In the dendrogram, shown at the bottom, a number of clusters can be observed. It appears that the samples all fall into one cluster. It also appears that a few other clusters are present. PCA and Clustering are unsupervised techniques, which mean that these algorithms don't use class information. Both techniques can be used to find natural clusters/grouping or outliers in the data.

To find differences between groups of samples or to link peaks with a continuous dependent variable, use BioMarker Surface Maps, PLS or LDA.

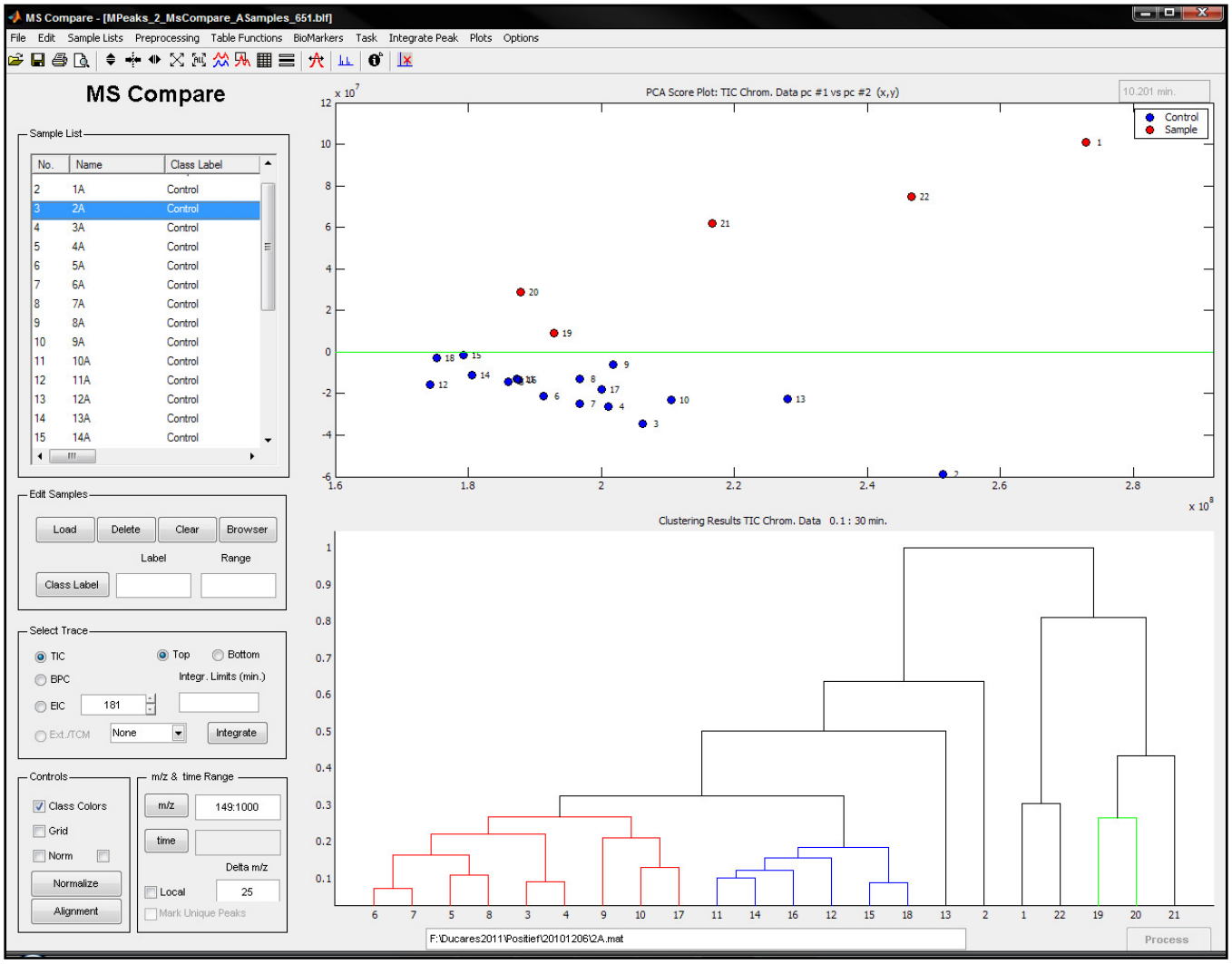


Figure 10: MsCompare PCA and Clustering

BioMarker Surface Maps:

To search for real differences between the two groups, use the BioMarker Surface Map algorithm, PLS or LDA. BioMarker Surface Maps are used to search the raw 2D LC/MS or GC/MS data for significant differences. PLS and LDA can be applied to MPeaks result tables.

Figure 11 shows a small part of the BioMarker Surface Map. The contour plot (time versus m/z) show regions where the active samples are significantly higher in intensity compared to the controls samples. The top window shows one of the significant features (EIC) of a peak that discriminates between both groups.

The map is easily converted to a result table. This table can be further examined in MsCompare or exported to Excel, Matlab, Unscrambler or SIMCA.

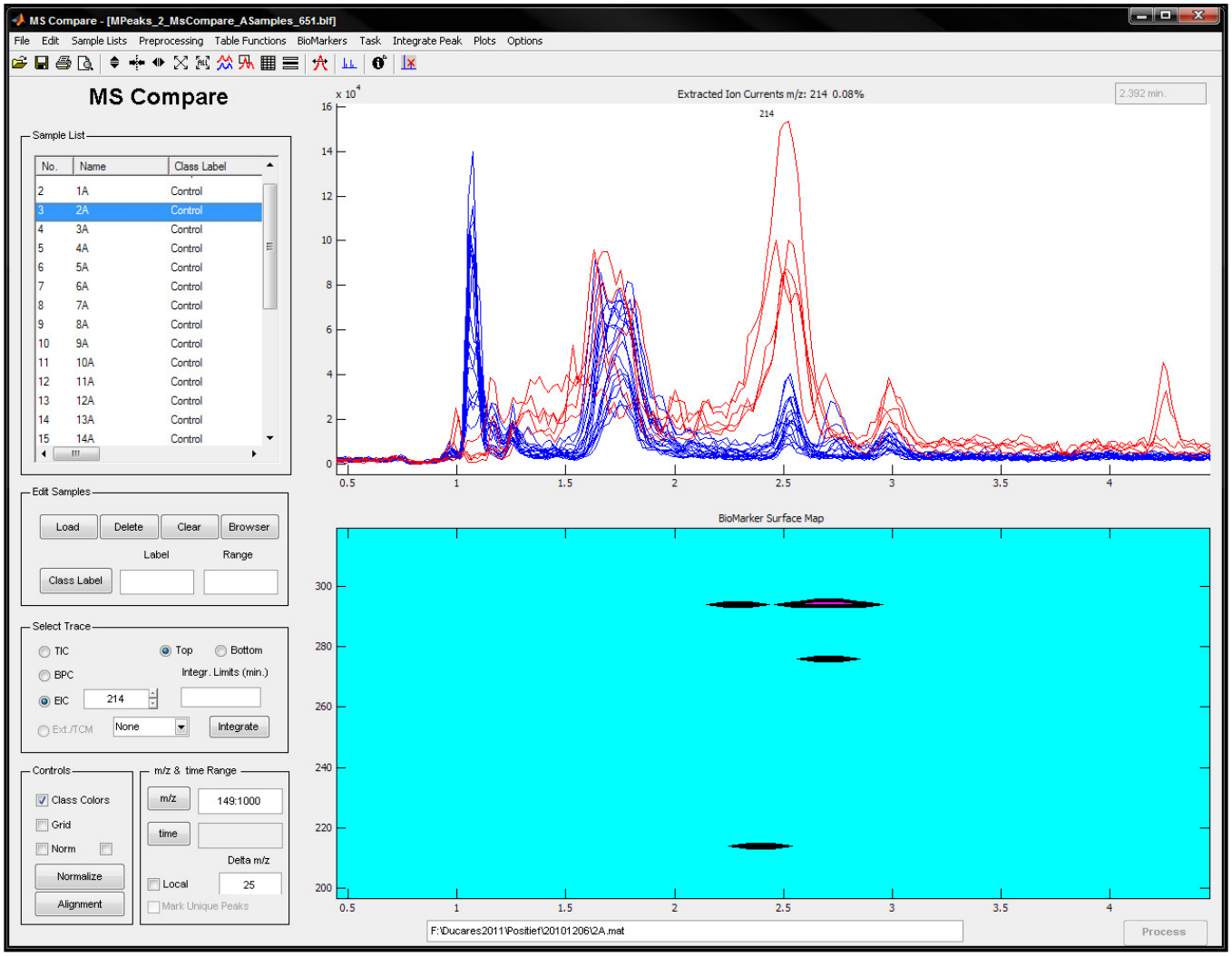


Figure 11: MsCompare – Biomarker Surface Map and extracted ion current for a discriminating feature m/z 214 at 2.4 minutes (red=sample, blue=control).

==========