Title
Procedure for Extracting a Data Matrix from Acquired GCMS Data using Analyzer-Pro

Scope

This Operating Procedure is intended to provide chemical analysts, research assistants and data analysts with a method for extracting a data matrix from a batch of empirical, GCMS data files.

Since no standard description of a data matrix is available, the reader should use this document together with any instructions as to the format of the data matrix as required by the Informatics Group.

The process involves

- Converting ChemStation *.D folders in to *.swx files native to Analyzer Pro
- Performing an untargeted alignment of QC standards against a suitable component library
- Generating a Target Component List
- Saving the Target Component List as a readable Analyzer Pro library
- Configuring the Matrix Analyzer Plug-in in Analyzer Pro
- Compiling a Processing Sequence
- Running a Processing Sequence
- Extracting the Data Matrix

This Operating Procedure does not include

- Configuring instrumentation and the computer interfacing required to acquire the raw data
- Analyzing and quality control standards protocols for sample analysis
- Demonstrating statistical control
- Application of descriptive statistics protocols

Definitions

Empirical data file:	Acquired by ChemStation and saved in *.D format
Converted data file:	Empirical data file saved in *.swx format
QC Standard:	Component mix where all constituents are known and recorded
Preferences:	Establishes rudimentary locations, file paths and search settings
Processing method:	Sets parameters and identities for peak picking and library searching
Component Library:	Compendium of compounds that include retention time and mass spectral data
Library generator:	Sets target includes, matching parameters, component list name and location
Component editor:	Allows setting and adjustment of component attributes to refine library matching
Matrix analyzer:	Embedded procedure for extracting a data matrix from a sequence of data files
Processing sequence:	List of data files to processed with a common set of preferences and processing method
Data matrix:	Data array listing sample component peak areas or heights (rows) against data file (columns)

Compatibilities

The data matrix as described in the “Definitions” subsection (above) is not consistent with the prescribed data matrix structure suggested by the Metabolomics Australia, Bioinformtics documentation. As such it will not be compatible with R-scripts provided by the Bioinformatics group to provide post processing support. Please export the data matrix out of Analyzer-Pro in *.xls format and perform a “Transpose” operation in Excel, then save the transposed data matrix in *.csv format.

Exporting the data matrix out of Analyzer-Pro in *.csv format poses a problem because the component names are formatted text that contain commas. All data matrices should be exported in *.xls format and modified from Excel.

Converting ChemStation *.D folders in to *.swx files native to Analyzer Pro

ChemStation *.D folders are converted into Analyzer Pro *.swx files using the “Import Data File” option in the “Files” menu.

This is described in detail in the Help pages and can be accessed by typing “Import” into the “Keyword” field under the “Index” tab.

Checking the “Preferences” Settings

System “Preferences” should be checked by selecting the “Preferences” option in the “Files” menu. Of primary importance here is to the parameters for library searching and assign data paths.

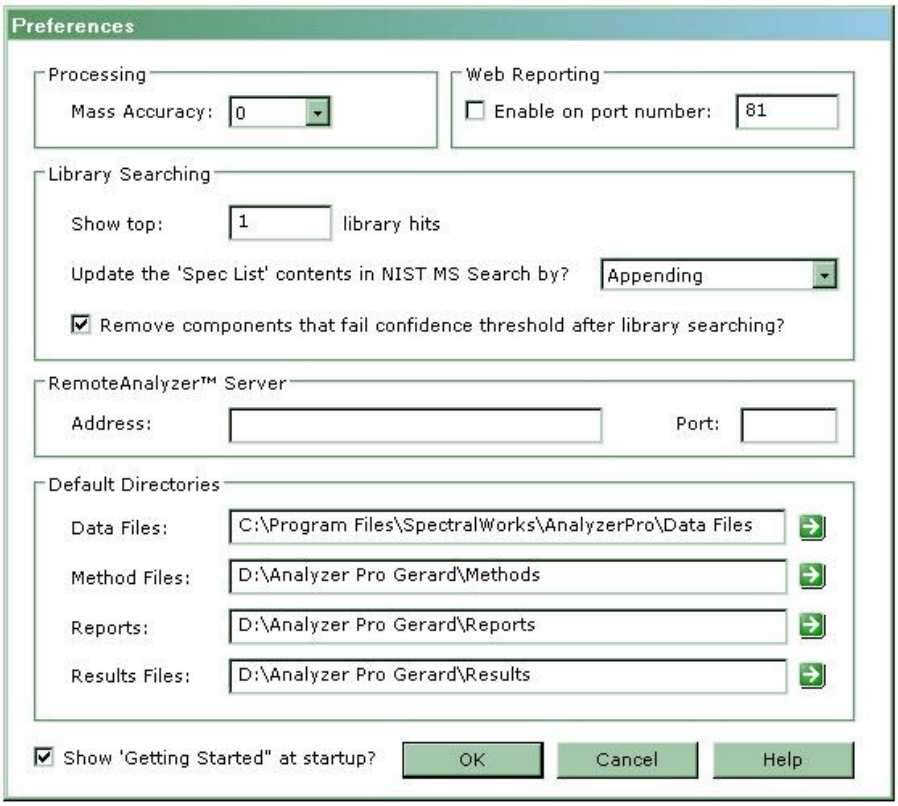

Figure 1: Shows the “Preferences” form with the required settings for Library Searching and assigning Default Directories.

Library Searching
As shown in figure 1, only show the library hit with the highest match factor. Select “Append” to update the “Spec List”. Check the tick box to “remove components that fail the confidence threshold after library searching”.

Default Directories
Ensure the file paths for all input and output files are known and point to folders on storage media (drives) with sufficient resources to store processing results, especially if these are the results from processing large batches of files.

Editing the “Processing Method”

Component mass rejection parameters, Detection windows and thresholds, search repository selection and retention index parameters are all accessible from this form.

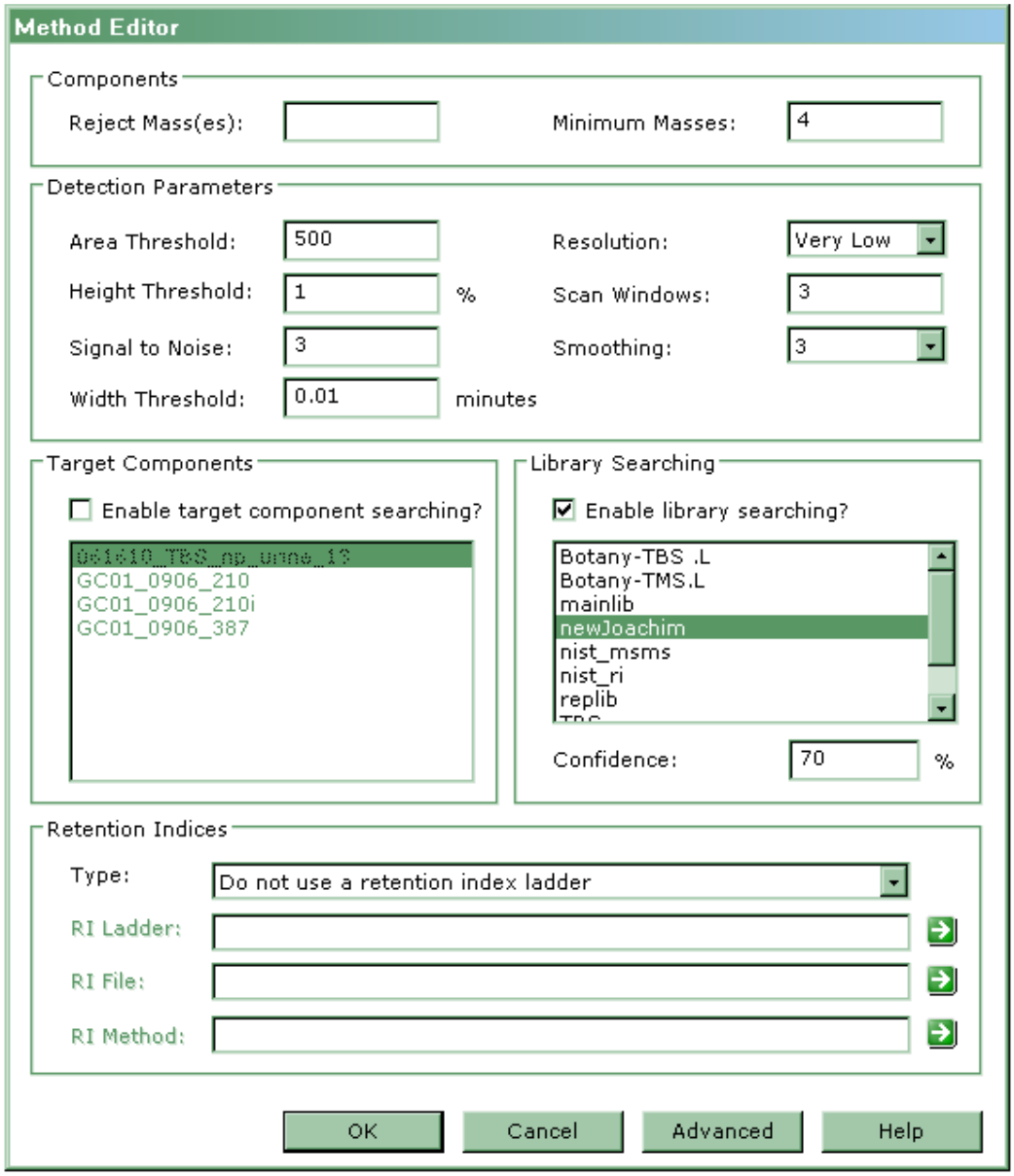

Figure 2: The "Method Editor" form with default "detection Parameters" and enabled "Library Searching" selected.

To be able to generate a "Target Component List", the "Enable Library Searching" check box must be ticked.

It should be noted that the "Detection Parameters" will need to be optimized. Although the default settings will work to yield the target component list, care is required when the "Scan Window" is set to avoid data loss due to an insufficient sampling frequency. Click the "OK" button to save settings.

Performing an untargeted alignment of QC standard components

Opening a Data File to Generate a Target Component List

After the "Open Data File" option is selected from the "File" menu, and open file dialogue box is displayed.

Typically a QC standard is opened for processing. This so, ideally, all sample components are known and exhibit good system response. This helps ensure that the maximum number of compounds of interest will be matched after the library search.

Analyzing an open data File

Once a processing method has been created and saved, selecting "Run Analysis" either from the Toolbar or the "Analysis" drop down menu will start peak detect processing cycle. This cycle includes smoothing the baseline, deconvoluting and searching a specified library or list of target components.

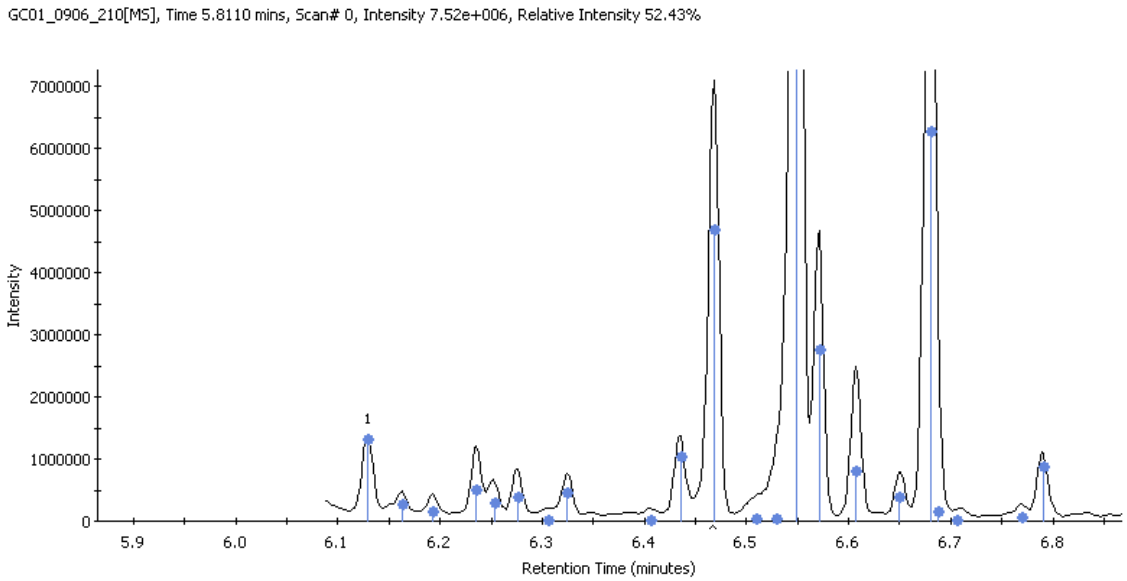

Figure 3: A sub-domain of a TIC illustrating the deconvolution of peaks, with the separated peak centroids indicated by the placement of the blue dots.

Generating a target component list from the analysis results of a processed data file

After the peak find and untargeted alignment against a selected library is complete, selecting the "Generate Target Component List" option in the "Results" drop down menu displays the "Target Library Generator" form.

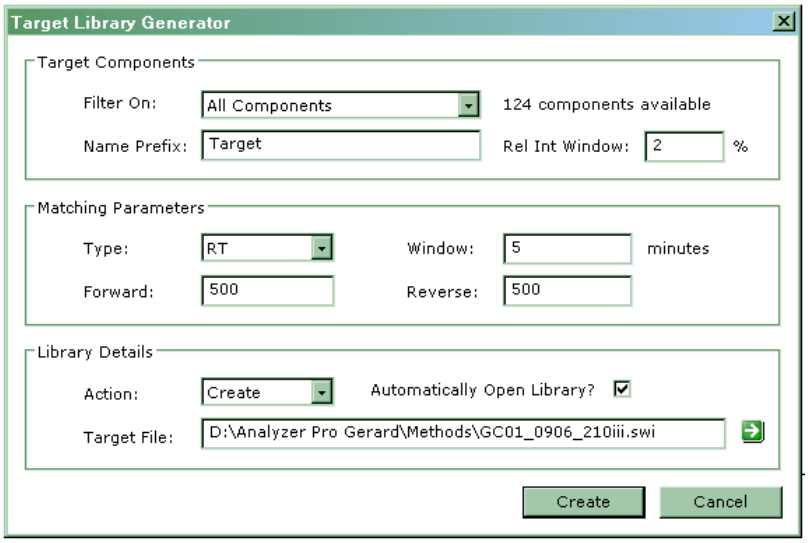

Figure 4: Shows typical default Matching Parameter settings in the Target Library Generator. The number of available components is shown next to "Filter on:" drop down list.

This process of generating a suitable component list should be considered an iterative process. The initial default settings will provide the seed from which optimization to generate a suitable target component list can begin.

Ensure that “Action” field is set to “Create”. The “Append” option will add the results of the previous analysis to any open library. Be sure the “Target File” is unique to prevent overwriting any previous saved lists.

Clicking the “Create” button creates and opens the “Target Component List” for editing to remove unwanted entries or modify matching parameters.

Editing the Target Component List

The “Target Component Form” provides a convenient interface through which the performance of the library search can be improved.

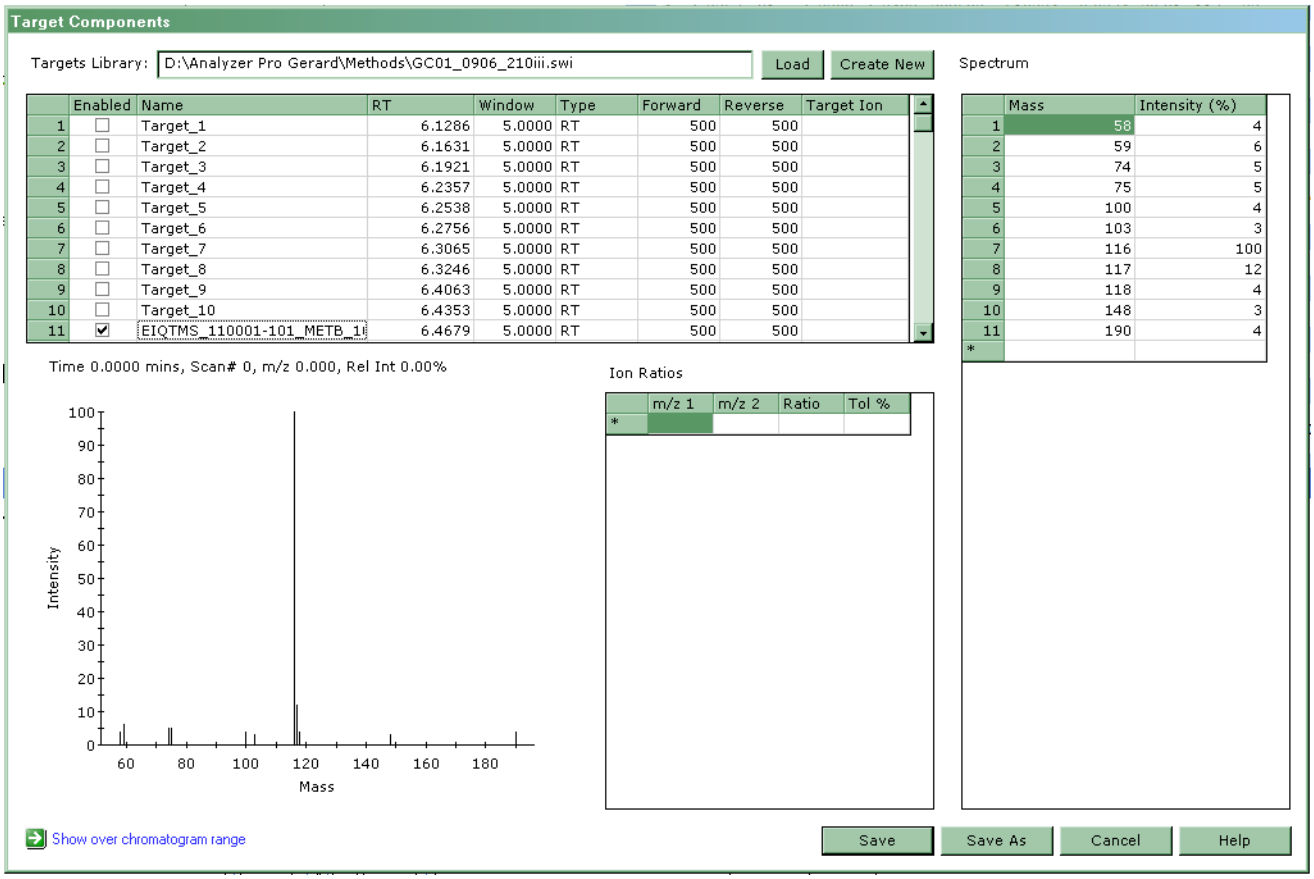

Figure 5: The “Target Components” form displays the file path and identity of the library file being edited. A list of identified and unidentified components correlates peak number, inclusion (“Enabled”), component name, retention time, window, window type, forward sweep, reverse sweep and target ion. Peak 11 is named and enabled. The “Spectrum” table shows all relevant masses for peak 11 together with their corresponding intensities. The Spectrum for peak 11is provided under the component list. Next to the relevant spectrum is a table where pair wise mass charge ratios can be selected.

Un-checking the “Enabled” tick box ensures any unidentified components (designated “Target #”) won’t be listed in any subsequent search with this saved target component list.

Reducing the time “Window” reduces the likelihood that a peak’s neighbors will be mistakenly be identified as the peak of interest. This window is not moving but rather, symmetrically specific and centered on the peaks retention time.

Reducing the “Forward” and “Reverse” parameters are used to determine if any components found match library spectra using the NIST matching algorithm.

Enhancing matching selectivity using retention time characteristics alone is considered a purely qualitative approach.

By providing a “Target Ion”, the extraction of semi-quantitative data from all ions or a single ion, in component spectra can be achieved.

To further confirm a component match, “Ion Ratios” together with a “Tolerance” can be assigned. The mass charge ratios can be selected from the spectral intensities in the “Spectrum” table.

Saving the Component List as a Readable Analyzer Pro Library

Once the target component list of interest has been edited and the “Save” button clicked then the list is saved as a component library in a *.swi format in the data path specified in the “Targets Library” field as depicted in figure 5.

Configuring the Matrix Analyzer Plug-in in Analyzer Pro

The “Matrix Analyzer” is a plug-in method outside the conventional peak find, deconvolute and align process in Analyzer Pro. To start the process of data matrix generation, the Matrix Analyzer Plug-in must be enabled and configured.

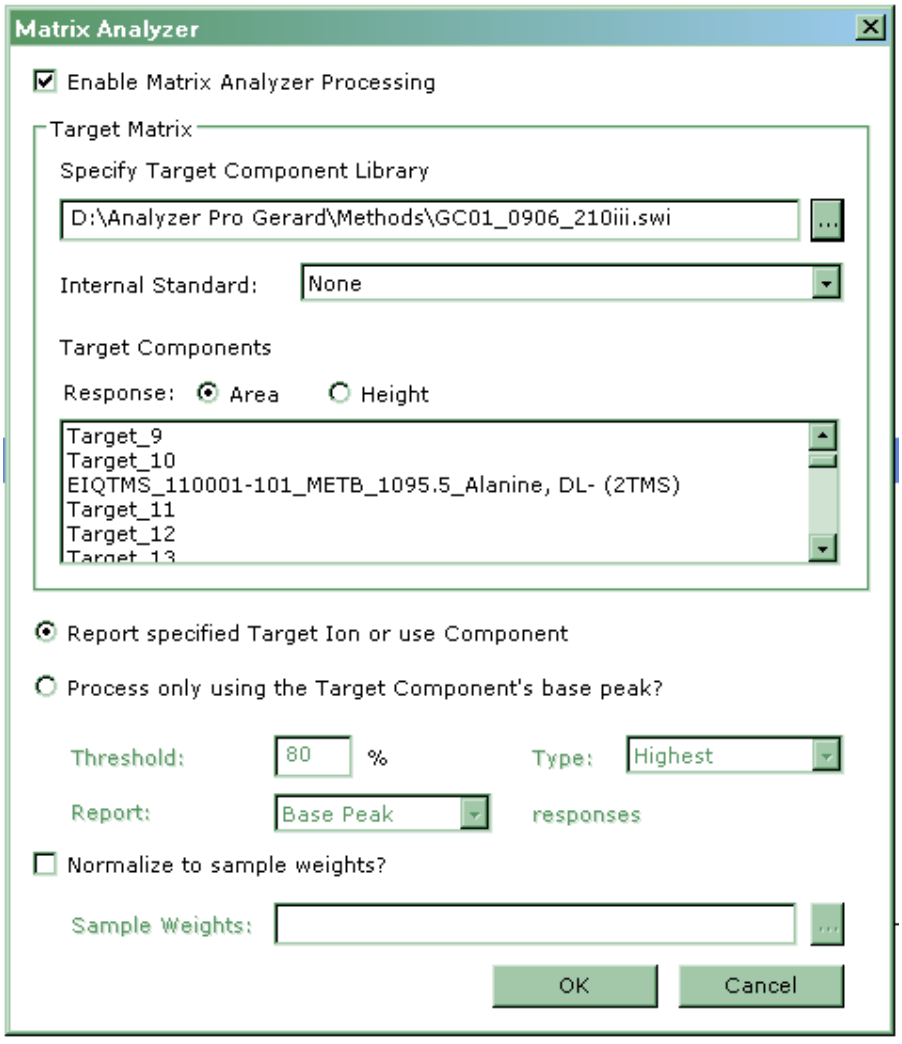

Figure 6: The Matrix Analyzer must be enabled and a target component library specified before the analysis process is started.

The “Matrix Analyzer” form is accessed from the “Methods” drop down menu by selecting it from the “Plug-in Methods” list. By checking the “Enable” tick box and specifying a “Target Component Library” a data matrix can be generated after processing a sequence of data files.

Compiling a Processing Sequence

A processing sequence can be compiled by opening more than one data file in the “Open File” dialogue form.

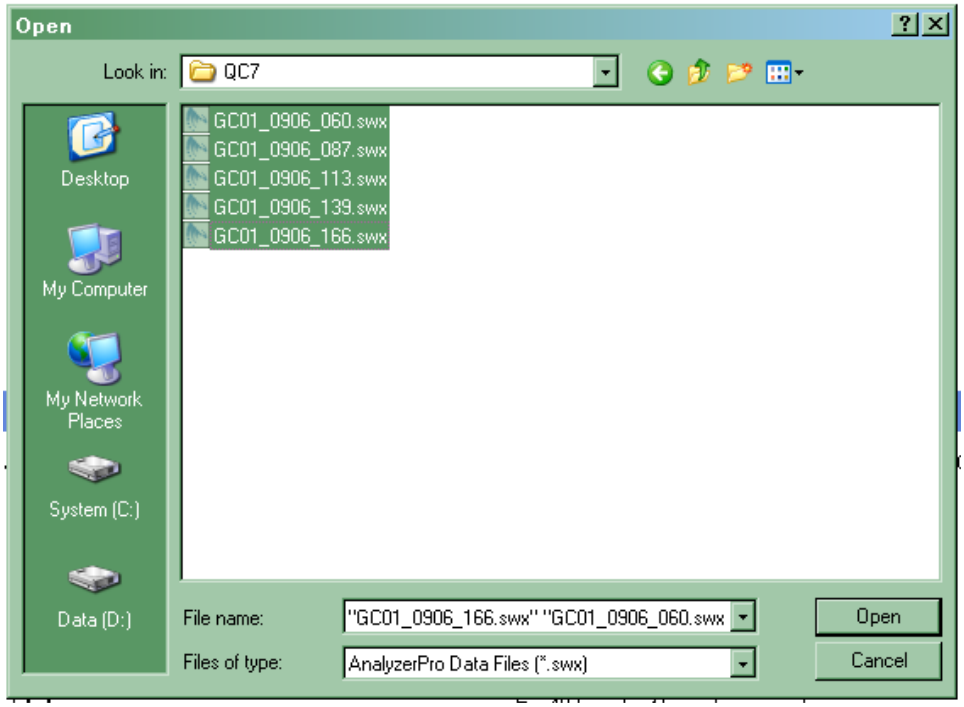

Figure 7: Shows the “Open File” dialogue form to set up a batch process sequence of five data files.

Clicking the “Open” button with all five files in this folder listed in the “File name” field automatically compiles and displays this sequence.

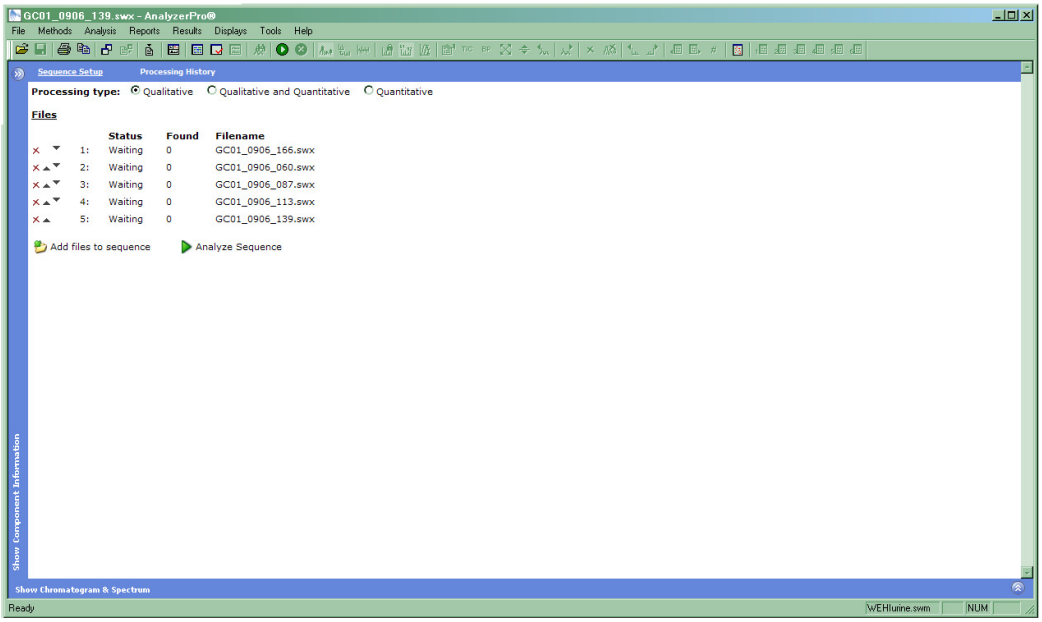

Figure 8: Is the “Sequence Setup” form. The files queued in the sequence are listed. Without an internal standard specified, the processing type can only be “Qualitative”

The Sequence queue can be appended to by clicking on the “Add files to sequence” icon. This will reopen the “Open File” dialogue and files can be selected by conventional browsing of file paths.

Clicking the “Analyze Sequence” icon will start the batch process and open the “Sequence Processing Options” form. The “Save sequence file as” field allows the file path and name the sequence will be stored as. The “Copy” data check box must be ticked before a file path for the metadata and result files can be specified. To export formatted results summaries the “Export” check box must be ticked before the “Edit Options” icon is clicked.

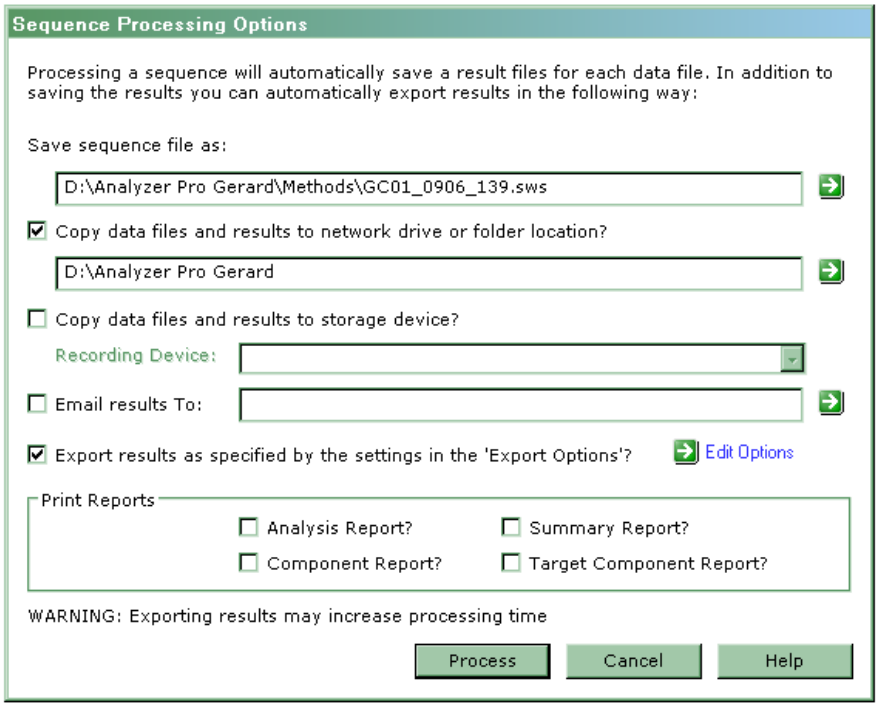

Figure 9: Shows a configured Options form for the sequence of five files depicted in figure 8.

Clicking the “Edit Options” icon opens the “Export Options” form. Result summaries can be output in a number of specific summary formats and file formats.

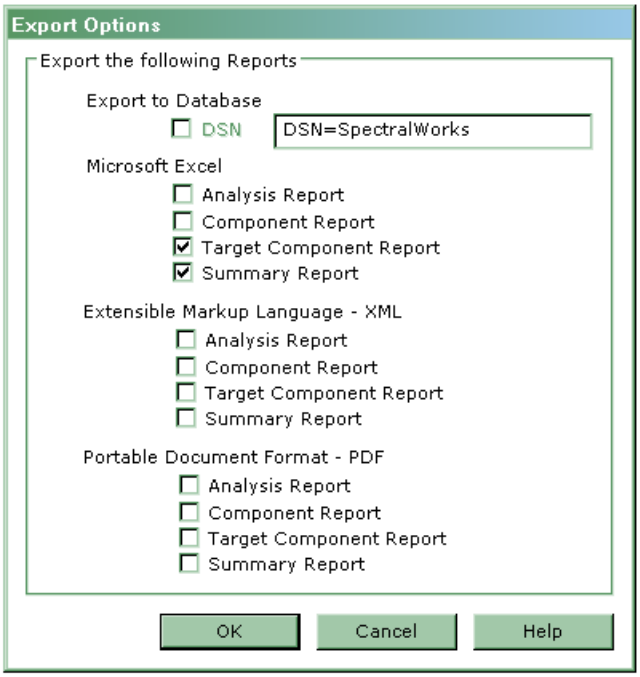

Figure 10: Is the “Export Options” form configured to output a “Target Component” and “Summary” report in Excel format.

Once the “OK” button is clicked on the “Export Options” form, the screen will revert to the “Sequence Options Form” as depicted in figure 9.

Running a Processing Sequence

The batch processing of the sequence “GC01_0609_139.sws” will commence once the “Process” button is clicked. Processing progress will be displayed in the toolbar where the “Ready” status is displayed in Figure 8.

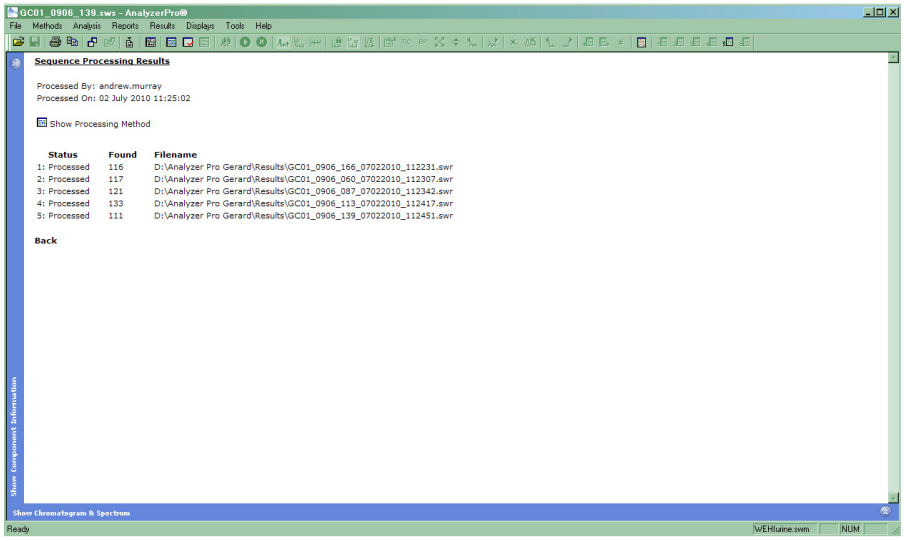

Figure 11: Displays the “Sequence Processing Results” form. The file “Status”, components “Found” and the storage location of the output “Filename” is tabulated.

Extracting the Data Matrix

With the status of all files in the sample queue listed as processed and as long as the “Matrix Analyzer” was enabled (see figure 6), the data matrix is extracted by accessing the “Matrix Analyzer Report” in the “Reports” menu, in the “Plug-In Reports” list. Clicking the “Matrix Analyzer” report option processes the acquired data from all sequence files and displays the data matrix on screen.

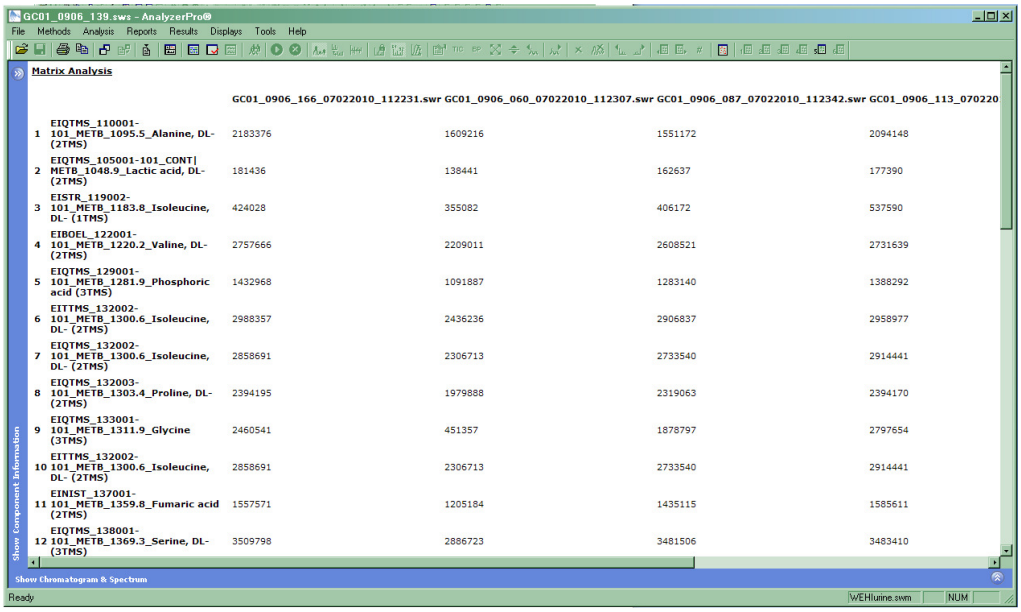

Figure 12: Depicts a portion of the data matrix. Sample components are numbered rows in the first column and all other columns are filename labels specific to each sample.

Links at the bottom of the screen display allow the data matrix to be exported to *.csv or *.xls file formats and saved for further informatics processing.