Tutorial

# **Using AnalyzerPro® for Metabolomics**



A streamlined untargeted metabolomics data processing workflow to understand fungal infection in frogs

# **SpectralWorks**

# Contents

Introduction					
Example data					
AnalyzerPro workflow					
1. lı	mporting and loading data files	3			
2. S	etting up a processing method	4			
3. C	Calculating a retention index	6			
3.1. N	/Janual RI ladder	6			
3.2. S	ingle data file	6			
4. B	Building a target component library	7			
5. P	Processing a sequence	10			
6. V	/iewing sequence results	11			
6.1. P	CA and data visualisation	11			
6.2. F	old change and one-way ANOVA	14			
Checklist for processing a sequence					

### Introduction

The following tutorial has been designed to help you use AnalyzerPro in your metabolomics workflow.

The example that we present here is a metabolomics analysis of the Alpine Tree Frog where the skin has been exposed to the Chytrid fungus.

Chytridiomycosis is an infectious disease that affects amphibians worldwide. It is caused by the chytrid pathogen *Batrachochytrium dendrobatidis* (Bd), a fungus capable of causing sporadic deaths in some amphibian populations and 100 per cent mortality in others. The disease has been implicated in the mass die-offs and species extinctions of frogs since the 1990s. However, its origin and true impact on frog populations remains uncertain. Using an untargeted GC-MS based metabolomics approach, it is hoped that an improved understanding of the molecular basis underlying the pathophysiology of chytridiomycosis can be determined. The findings will be fundamental for i) explaining differences in population susceptibility, and ii) identifying targets for intervention and improved management. In this tutorial we will only be working with a small data set to teach you the basic steps necessary to use AnalyzerPro. A robust experimental design requires as many sample replicates as possible to give the resulting data sufficient statistical power. Here we demonstrate the AnalyzerPro workflow using a subset of a greater sample set. We will be using 3 controls (Frog 1-3) and 3 infected samples (Frog 4-6). For more information about the metabolomics workflow and best practice approach, you can download our metabolomics guide from:

#### http://bit.ly/2c13n86

The samples we are using were homogenised and metabolites extracted with methanol and water containing  ${}^{13}C_{6}$ -sorbitol as the internal standard. Extracts were freezedried and derivatized for GC-MS analysis and were acquired on a Shimadzu QP Ultra GC-MS operated in full-scan mode.

### **Example data**

Before working through this tutorial, you will need to download the data set from the following link:

#### http://bit.ly/2bHGumn

The data set contains the 3 controls and 3 infected samples as well as 12 data files containing spectra for authentic metabolite standards (Beta lactose, DL ornithine, Dopamine, Glutamine, Inosine, Isovaleric acid, L serine, Pantothenic acid, Putrescine, Spermidine, Spermine, Tryptophan) which are of interest to this data set. A common approach in metabolomics is to build an in-house library of spectra from available standards and then append the library with components from your sequence of samples. Additionally, components from samples can be matched with external libraries such as the NIST mass spectral database. Following these identification processes you will have a final library of components which includes identified compounds, putatively identified compounds and unknowns.

You can also watch a video tutorial of this online at:

http://bit.ly/2cBaRjw

# AnalyzerPro workflow

There are 6 key steps to follow to maximize your productivity with AnalyzerPro, these are:

- Importing and loading data files
  Set up a processing method
  - Create a RI method
  - Build target component library
  - Process your sequence
  - Interrogate your data

# 1. Importing and loading data files

Go to Home | Vendor's Files. Using the ... button, select the data files to convert and the folder you would like them to appear once converted. You can choose to either Convert Data Only, or Convert Data and Auto Load to generate a sequence table. If you selected convert files only, go to Home | Data File to load the converted data files into a sequence table.

If you have downloaded the frog data files which are already in the .swx format, go to Home | Data File to load the data files.

Once data is loaded, you will be able to view the chromatograms and mass spectra for each file as shown:





# 2. Setting up a processing method

Setting up a method is where you determine your peak detection parameters. Your method won't be complete until you've created an RI calculation file (if required) and created a library to match against but you will need to process a data file or two to set these up so follow the instructions below to process a file and then follow the instructions to calculate a retention index (section 3) and automatically build a target library (section 4) to complete your untargeted metabolomics data processing method.

To create a new method, click Methods | New, then click Methods | Edit to edit your method.

Define processing constraints such as the minimum masses required to define a peak, the mass range to process, the retention time (RT) window to process and any masses to reject.

Define the detection parameters. This includes area threshold, height threshold, width threshold, signal to noise, scan windows, chromatographic resolution and Gaussian smoothing.

If you have accurate mass data, check the accurate mass processing check box and enter the actual precision (ppm) of your data and select the linearity of that precision. You also need to set the precision in your preferences so that displayed spectra reports m/z to the desired number of decimal places. To view your preferences select **Preferences** from the application button to the left of the **Home** tab.

Preferences

Processing		Results	
Mass Accuracy:	0 ~	Allow result	s files to be overwritten?
Library Searching			
Show top:	5 libra	ry hits	
Update the 'Spec	List' contents in NIST M	S Search by?	Appending $\sim$
Remove comp	oonents that fail confiden	ce threshold after lib	rary searching?
Default Directories	-		
Library Files:	C:\Users\HayleyA\De	sktop\Frog Data Tuto	orial
Reports:	C:\Users\Public\Docu	ments\SpectralWork	s\Reports
Plate Size		Quantitation	
Rows: 8	Columns: 12	Display Window:	0.5 ± mins
		ОК	Cancel Help

For the frog data, the mass accuracy processing option should not be checked and the number of decimal places can remain as the default 0.

Select the Enable target component searching? check box. Any one of the available target component libraries can be selected from the list box. Found components will be searched against the selected library and the results will be available in the Target Component Report. Select the Enable library searching? check box if you want to enable this function. Found components will be searched against libraries such as NIST if you have them installed.

Select the type, if any, of retention index processing that is required from the drop down list. Retention indicies is covered in more detail in the following section, as is building a target library. For initial processing, no RI method is defined and no target library is checked.

For frog data, the following processing method was used:

Search Oual Method Editor

Minimum Masses:	4		Reject Mass(es):	73
Mass range:			RT: Process	full retention time
0 -	1000			
etection Parameters				
Area Threshold:	5000		Signal to Noise:	10
Height Threshold:	1	%	Scan Window:	5
Width Threshold:	0.01	mins	Resolution:	Minimum
Fronting:	0	%	Smoothing:	7
Tailing:	0	%		
Data Channel:	All Channels			
curate Mass				
Enable accurate n	nass processing?	,		
Precision	5	0.000	Linearity:	Low

To save these details to the method file select **Method | Save** or **Method | Save As** and complete the dialog. Unless these changes are saved to a method file, you may lose them.

# 3 Calculating a retention index

Retention indicies such as Kovats for organic compounds are a standard used to calculate a retention index for a compound relative to, in the case of Kovats, hydrocarbons. For example, n-dodecane ( $C_{12}$ ) has a retention index of 1200 and  $C_{15}$  has a retention index of 1500. Compounds eluting between  $C_{12}$ -  $C_{15}$  therefore have retention indices of 1200 - 1500. Retention indicies are useful because they are constant regardless of temperature program and column properties such as length and diameter so compounds analysed under different conditions can be matched to a library using RI rather than retention time.

Using AnalyzerPro, there are three ways to calculate RI. 1) Manually; 2) using a method and single data file; and 3) for each data file using a method.

For small-scale studies where the work is performed within a maintenance cycle of the instrument, the manual or single data file methods can be used. In large-scale studies where maintenance is performed and/or retention time shifts occur, using the 'calculate a retention index for each file using a method' is the best option.

## 3.1 Manual RI ladder

To build an RI method manually, you can use a data file containing only the retention index compounds or a sample file whether that's a QC sample, experimental sample or extraction blank. As long as the retention times of the RI components are accurate for the method used to acquire the data set, any file can be used.

Open a data file and find and record the retention time of each RI component in the data file. Once the retention times are known, populate the RI ladder table found in **Methods | Qual | RI Ladder**. For the frog data, n-alkanes have been used. The following table has been populated using n-alkane retention times from the DL ornithine standard file.

	Name	RT Minutes	RI
1	C12	14.8133	1200
2	C15	22.2917	1500
3	C19	30.2550	1900
4	C22	35.2867	2200
5	C28	43.6367	2800
6	C32	48.3050	3200
7	C36	52.4250	3600
*			

Once the RI method is complete and saved, you can edit your processing method to include your retention index calculation method.

# 3.2 'Using a method and single data file' and 'for each file using a method'

Choose an appropriate data file to build your RI method (.rim file). This can be any file from the experiment acquisition which contains the retention index compounds. A mid-run pooled QC sample would be a good choice if available. Process the file using appropriate peak detection parameters and for speed, turn off library searching. For the frog data, DL ornithine containing n-alkanes has been used to create the .rim file.

Once the file is processed, find your retention index compounds and record their retention times. For ease, extract masses of interest (right click chromatogram and choose **Options | Trace Type | Mass Chromatogram**) i.e. for n-alkanes extract mass 57, 71, 85 etc. Then go to **Home | Results | Retention Index** button to create your method (.rim file).

Delete components you do not wish to use for RI calculation. Using the retention time you previously recorded, and spectrum, identify your RI compounds and give them the appropriate name and retention index i.e. C12 or n-Dodecane, RI: 1200. You will need to choose a retention time window based on your data. For large metabolomics studies, there is the possibility of RT shift/ drift even for GCMS. Try to quantify the RT shift before choosing the window. You will also need to include a relative intensity window (Spectrum button with >%). By selecting a relative intensity of 10, only those ions which have a relative intensity of 10 or greater will be used for matching. If there is any co-elution or difficulty deconvoluting RI components, by using a relative intensity of 10, ions which may be 'impurities' will not be included in the matching process allowing a better match with each file.

ri	📉 RI Method Editor											
File H	File Help											
🗃 🗎 🖳 🍸 🍾 🖍 💼												
		Name	RT	RT Window	Retention Index	Coelution Window	Relative Window	Spectrum				
1	$\checkmark$	C12	14.8133	0.2	1200	5	10.0	(41,57);(42,1				
2	$\checkmark$	C15	22.2783	0.2	1500	5	10.0	(41,52);(42,1)				
3	$\checkmark$	C19	30.2550	0.2	1900	5	10.0	(41,42);(42,1				
4	$\checkmark$	C22	35.2817	0.2	2200	5	10.0	(41,37);(42,1				
5	$\checkmark$	C28	43.6483	0.2	2800	5	10.0	(41,29);(43,8)				
6	$\checkmark$	C32	48.3133	0.2	3200	5	10.0	(41,23);(43,74				
7	$\checkmark$	C36	52.4300	0.2	3600	5	10.0	(41,20);(43,6				

It's a good idea to re-analyse the sample or a couple of samples to make sure that your RIs are correct. Since you may like to use a method with RI to match spectra rather than RT, it's important to get it right or you won't get any library matches. To do this, view the summary report (**Report | Summary**). Once the RI method is complete and saved, you can edit your processing method to include your retention index calculation method. If you select using a method and data file, the data file you wish to use as the reference (DL ornithine for frog data) will need to be included in your method.

# 4 Building a target component library

Your data set can be completely defined by the component library generated during the automated library building stage.

Following deconvolution, each component found can be assigned an identity. This identity does not have to be absolute and subsequent analysis can be performed on identified components or by treating each component as an unknown. Using AnalyzerPro, components can be identified automatically by matching to public libraries or from a specific in-house library of authentic standards.

Libraries are best built using QC samples but they can be built using your entire data set, too. If you have used an internal standard (IS) in your analysis, you can find it in the file you processed to create your RI method (.rim) file and add it to a library before processing a sequence. Alternatively, process a file which you know contains your IS and add it to a new library by right clicking the deconvoluted spectrum and selecting 'add as target component'. Lastly, you can let the internal standard be found during the automated library building stage of sequence processing. You can:

- a) Create library from a single file
- b) Automatically build library from a sequence of files (preferred)

You may like to find your internal standard in your data and add it to a new library first. You can then append this library with the new components from a) your single file or b) your sequence. If you have QC files it's not necessary to do this step because you will need to re-process anyway but if you don't have QC files, and want to process your sequence at the same time you build your library, find your IS first so that you can normalise your data immediately.

When you are ready to build your library, go to **Home | Sequence | Analyze**. Specify the name and output location of your sequence and then check the 'automatically build target library' button. Be sure to have the library you want to append selected in your method.

If using the tutorial data, process one of the frog samples and find  ${}^{13}C_6$ -sorbitol at 30.6 min. Use m/z 323 as your target ion in the library.





Target Component Library										
	Enabled	Name	RT	Window	Туре	Forward	Reverse	Target Ion		
109	<b>~</b>	Frog 1 Control_0109	1912.0200	5.0000	RI	500	500			
110		Frog 1 Control_0110	1919.0800	5.0000	RI	500	500			
111		13C6-Sorbitol	1920.9700	5.0000	RI	500	500	323		
112	<b>\</b>	Frog 1 Control_0112	1932.3000	5.0000	RI	500	500			
113	<b>v</b>	Frog 1 Control_0113	1938.3600	5.0000	RI	500	500			

If you would like to include authentic metabolite standards in your processing, first process these as a sequence and automatically build a library. Alternatively, since there is a small number of files, add the major derivatization product to the library by finding the spectrum, right clicking and selecting add to target library. Using the 12 standards supplied, create a sequence by opening the data files and process the sequence using the automated library build function.

Once processed, data files will look like this:



The green lollipops represent found components. Information for each component can be found in the summary report, accessible by clicking **Home | Results | Qual**. Clicking each green lollipop will display the deconvoluted peak and mass spectrum as shown below:

Frog 4 Infected[MS], Time 22.7810 mins, Scan# 8869, Intensity 8.94e+005, Relative Intensity 30.60% All Channels Time 22.7567 mins, Scan# 8854, NL 2.9111E+006, m/z 109.383, Rei Int 57.79% All Channels



### 5. Processing a sequence

The time it takes to process a sequence depends on how large your data files are. For the six-sample frog data set, the processing time is approximately 10 minutes.

Load the processing method you previously created. Load the library to match against and ensure library searching is checked in your processing method with the appropriate library selected. Include calculate RI manually or using a method (.rim file) in your method.

Define categories: Home | Sequence | Category Information. Once this is done your sequence table should resemble:

Sec	Sequence - Untitled										
	Filename	Status	Found	Targets	Category	Sample Weight	Method				
1	Frog 1 Control.swx	Waiting			Control	1.0					
2	Frog 2 Control.swx	Waiting			Control	1.0					
3	Frog 3 Control.swx	Waiting			Control	1.0					
- 4	Frog 4 Infected.swx	Waiting			Infected	1.0					
5	Frog 5 Infected.swx	Waiting			Infected	1.0					
6	Frog 6 Infected.swx	Waiting			Infected	1.0					

Check the **Enabled** checkbox in the **Automatically Build Target Library** group on the **Sequence Processing Options** dialog, having first ensured the library to append with new components is selected in the processing method. Refer to the end of this document for checklist for processing a sequence.

#### Sequence Processing Options

#### Save sequence file as:

C:\Users\HayleyA\Desktop\Frog Data Tutorial\Frog tutorial v2.sws	
Quantitation	Print Reports
Export Calibration Curve	Custom Reports: None v
Use External Calibration Curve:	Standard Reports: Analysis Component Summary Target Component
Export Results	Automatically Build Target Library
Copy data files and results to network drive or folder location?	Enabled
	Forward: 500 Reverse 500 Window: RI V
Email results To:	Intensity: 0 %
Export results as specified by the settings in the 'Export Options'?	
	Process Cancel Help

#### 6 Viewing sequence results

From **Home | Results**, once your sequence is processed, click **Matrix Analyzer**. Large metabolomics data sets can be modelled using principal component analysis to visualise which metabolites [components] contribute most to the variance observed. It also allows the identification of outliers that are likely to contribute to the skewing of the variance of a particular component. PCA points you in the direction needed for interpretation, but is only one part of the data analysis workflow. Fold-changes and p-values can be calculated in AnalyzerPro to aid in determining key metabolites of interest that may help to answer the biological question.

#### 6.1 Principal component analysis and data visualisation

Once your sequence has finished processing, click **Home | Results | Matrix Analyzer**. Here you will be able to view your PCA (if desired; with or without log transformation), results and normalised results. You can edit the category information post processing to generate a PCA (**Home | Sequence | Category Information**).

#### Export results for additional statistical analysis

In addition to the statistical tools built into AnalyzerPro, you can quickly and easily export results into csv or Microsoft Excel format so that additional statistical calculations can be performed by complimentary third-party software packages so get the most information from your data. Click the **Options** button (left button on the tool bar) to select which information to include in your matrix and which data transformation to use for PCA.

MatrixAnalyzer Options

Reporting Parar	meters						
Use	Target Ions or Comp	oonent ~	Report	●Area (	Height		
Normalize	to Target Component						
13C6-sor	rbitol	~	Normaliz	e to Sample W	/eights?		
Available Colun	nns						
Compo	nent Name	Expected RT		Delta R	т		
Results		Normalized F	Results				
Averag	e	Standard De	viation	RSD%	RSD%		
🗹 p-Value	e	Fold Change		☑ % Detected			
Principal Comp	onent Analysis Paramet	ters					
Variation:	90 %	Use log trans	formation	Use Nor	rmalized Results		
PCA Loadings P	Plot Style						
Color:	Default 🗸	Size: 7		Symbol:	Dot 🗸		
				ОК	Cancel Help		

As you scroll through components, you can inspect RT shift in your data and any main category effects for further interpretation:





PCA Loadings from normalized results with log transformation



By clicking on points in the loadings plot, the bar graph, which is colour coded by category, will update to show the relative concentration of the component in each sample.

In addition, the PCA data can be viewed using a volcano plot or bar chart. Generate a volcano plot to visualise all components plotted against their fold-change and p-values. This representation of data allows the user to determine those components with the highest fold-changes and lowest p-values (i.e. the most important results) with ease.

Volcano Plot



View each component by experimental group plotted as the mean (+/- SD). Group means are colour coded consistent with the PCA plot (according to class). This representation allows clear visualisation of a components relative concentration, further aiding the user's interpretation of results.



# 6.2 Fold change and one-way analysis of variance (ANOVA)

If categories have been assigned to the data files, fold changes can be calculated and displayed in the Matrix Analyzer results viewer. The fold change value represents the average peak area response of one class over another and is a useful measure to determine changes in compounds as a result of an experimental condition. Matrix results can be sorted according to fold change.

You can calculate a p-value using, if appropriate, a one-way ANOVA. The p-value, used in hypothesis testing, indicates that there is a statistically significant difference between one group mean when compared to another. A significant result (usually if the p-value is less than 0.05) indicates that the probability of the null hypothesis being accepted is very low (5%) and the alternative (test) hypothesis can be accepted. As with fold-change, matrix results can be sorted according to p-value. When sorted, DL-ornithine has the lowest p-value (highest significance).

In the example, the relative concentration of DL-ornithine is greater in the infected samples compared to the control. The average fold change is approximately 13.5 and p-value is <0.001.

	Components	Expected RT/RI	[N] Frog 1 Cont	[N] Frog 2 Con	[N] Frog 3 Cont	[N] Frog 4 Infec	[N] Frog 5 Infect	[N] Frog 6 Infe	p-Value	[N] p-Value	Fold Change	[N] Fold Change	% Detected
31	Frog 1 Control_0007	1082.8000	349.6259	418.8844	573.4193	743.5743	797.9068	786.6038	0.51937	0.00852	0.92971	1.73488	100.0
170	Frog 1 Control_0058	1427.3900	18.3178	19.8673	27.4165	31.3374	26.1762	52.5490	0.65133	0.15787	0.89913	1.67774	100.0
232	DL-ornithine 2	1755.0000	5.0401	4.6269	10.3076	133.7633	181.7217	174.3865	0.00000	0.00048	13.50774	24.52472	100.0
48	Frog 1 Control_0011	1109.0699	12.5273	11.3845	26.1007	27.1911	23.2355	38.0633	0.99274	0.11890	0.99803	1.76936	100.0
51	Frog 1 Control_0012	1131.6000	230.4215	154.8819	264.6666	194.4227	380.5426	380.6472	0.19670	0.21936	0.75515	1.47024	100.0

# Checklist for processing a sequence

Make a new folder each time you start a new project/analysis with sub folders for your data files and methods. The data folder should contain your AnalyzerPro .swx files. When your sequence is processed, results files (.swr) and any reports requested will also be stored in there.

Have you?

- 1. Converted your data files to .swx?
- 2. Loaded the .swx files to process in the sequence?
- 3. Created AnalyzerPro method files (processing method, RI method, library)?
- 4. Check the preferences tab:

Be sure to use the correct path for library (method) files

5. Open target library:

Make sure it's correct and the correct file path is being used

- Open your processing method: Check parameters, RI method and target library
- 7. Define sample categories (i.e. control, infected). This step can be done post-processing, too.
- 8. Run sequence:

Choose name and output location (the project folder). Check the automatically build target library option

# **Acknowledgements**

Laura F. Grogan<sup>1,2</sup>, Joel P. A. Gummer<sup>3</sup>, Lee Berger<sup>1</sup>, Lee Skerratt<sup>1</sup>, Scott D. Cashins<sup>1</sup> and Robert D. Trengove<sup>3</sup> conducted the study. <sup>1</sup>One Health Research Group, College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland 4811, Australia. <sup>2</sup>Griffith Wildlife Disease Ecology Group, Environmental Futures Research Institute, School of Environment, Griffith University, Nathan, Queensland 4111, Australia. <sup>3</sup>Separation Science and Metabolomics Laboratory (Metabolomics Australia, Western Australia node), Murdoch University, Perth 6150 WA, Australia.

We thank P. Harlow, M. McFadden, D. Hunter and B. Scheele for assistance with logistics for the clinical experiment and species' insights. This study was conducted with approval by the James Cook University Animal Ethics Committee (Certificate no. A1589) and Scientific License number: S12848 (D. Hunter). This work was jointly funded by the US Fish and Wildlife Service - Wildlife Without Borders program and the IUCN Amphibian Specialist Group Seed grants program. Metabolomics Australia is a Bioplatforms Australia (BPA) funded initiative.

# **About SpectralWorks Limited**

SpectralWorks Limited is a leading UK based software development company.

We are dedicated to providing innovative solutions targeted for markets within the life sciences industry and have strong working relationships with the major instrument manufacturers. Coupled with our collaborations in academia and industry, we believe we have the right balance between scientific and software development expertise to provide the best scientific solutions for the end user.

Our vision at SpectralWorks is to improve the way software is integrated within the laboratory environment by providing the correct solutions to increase productivity and reduce overheads. We achieve this by maintaining focus on the end user, listening to their requirements and ensuring that they have the right tools to handle their day to day tasks.

In addition to the wide range of mass spectrometry software products, we provide highly respected consultancy services. These services are able to cover the complete software development cycle or specific steps within a project life cycle. These services are frequently utilized by instrument vendors and end users that demand the best for their laboratories.

We are a privately owned company, incorporated in the United Kingdom in January 2004 and are based in Runcorn, United Kingdom. Our offices are located on a 60 acre business park 25 miles from Manchester International Airport and 10 miles from Liverpool's John Lennon Airport.

Our facilities provide us with access to MS and other instrumentation as well as ensuring that we have the best infrastructure to develop and test software.

As well as direct sales, we have partner resellers. SpectralWorks offers Original Equipment Manufacturers (OEM) the ability to distribute our products and algorithms using their own branding. We also provide custom development for individual companies looking to automate and unify the software used within their laboratories using our extensive industry experience.

Our customer base is varied but predominately based within Australasia, Europe and North America. It encompasses academia, hospitals, government and industry.

Support is available via telephone, email or our web based support site. We always reply to communication as soon as possible and endeavour to reply within 24 hours with a statement of the problem and an estimated date for a solution.

# About AnalyzerPro

AnalyzerPro is a data deconvolution software application for LC-MS and GC-MS data. It utilizes proprietary algorithms to detect obscured components that existing software is unable to find without additional information.

For a free 15 day trial licence, visit our website:

http://www.spectralworks.com

# **SpectralWorks Limited**

The Heath Business and Technical Park Runcorn, Cheshire, WA7 4EB, United Kingdom T: +44 (0)161 327 2143 | E: info@spectralworks.com



Certificate Number 13452 ISO 9001:2008

This publication provides outline information only which (unless specifically agreed by SpectralWorks Limited in writing) may not be used, applied or reproduced for any purpose or form part of any order or contract or be regarded as a representation relating to the products or services concerned. SpectralWorks Limited reserves the rights to alter, without notice, the specification, design, price or conditions of supply of any product or services.