Analyze IQ SmartSearch Version 3.0

User Manual

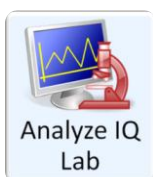# Table of Contents

1   Introduction

Analyze IQ SmartSearch allows you to search quickly and accurately through a library of spectra, to identify the best matches for a spectrum that you wish to identify, enabling you to answer "what is this?" questions easily and reliably.

Analyze IQ SmartSearch is the ideal complement for **Analyze IQ Lab**, the world's most advanced Chemometric modeling software, which allows you to perform detailed quantitative and qualitative analyses of complex mixtures. **Analyze IQ SmartSearch** also integrates with **Analyze IQ Spectra Manager**, with which you can store and organize your valuable chemical spectra and associated information.

This manual provides detailed instructions on the use of **Analyze IQ SmartSearch**'s rapid and accurate spectrum matching algorithms to allow you to identify a ranked set of spectra that are the closest matches for a spectrum of interest, using the variety of methods provided.

1.1   Overview of the Analyze IQ Product Suite

SmartSearch is part of the Analyze IQ product suite.

Analyze IQ Lab: Provides rapid & accurate chemometric analysis methods to turn complex spectral data into decisions. Its model building wizard allows you to build, evaluate and refine sophisticated models, quickly and reliably.

Analyze IQ RealTime: Enables you to package chemometric models that have been constructed and validated by expert analysts and deploy them in fielded applications, where they can be applied automatically for real time predictions. Available in Portable, Desktop, or Cloud Editions.

Analyze IQ SmartSearch: Allows you to search rapidly and accurately through a library of spectra, to identify the best matches for unknown spectra, enabling you to answer "what is this?" questions easily and reliably, and providing you with useful supplementary data.

Analyze IQ Spectra Manager: Makes it easy for you to **store** and **organize** your valuable spectra and related meta-data. An ideal alternative to the high cost & effort of implementing a LIMS.
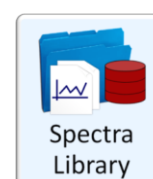
Analyze IQ Spectra Library: A carefully-curated collection of 1868 Raman spectra of common organic materials, including laboratory chemicals, hazardous materials, research chemicals, and solvent mixtures.

SmartSearch can perform searches using the **Analyze IQ Spectra Library** or third-party libraries (See Section 7). You can use **Analyze IQ Spectra Manager** to create user libraries for searching with **SmartSearch**, and **Spectra Manager** can also add spectra to a library or remove some from it, or to update details in a library.

While **SmartSearch** is designed to answer high-level "what is this?" questions, **Analyze IQ Lab** is designed to perform specific and detailed analyses, by providing a range of quantitative and qualitative chemometric methods that can be used to identify specific constituents in complex mixtures and relate spectra to quantities of interest.

1.2   Overview of this Manual

Section 2 describes the **Analyze IQ SmartSearch** installation process.

Section 3 briefly describes the data stored in **Analyze IQ** spectra libraries and explains key terms used.

Section 4 provides a quick overview of the basics of how to perform a search.

Section 5 explains in greater detail how to use **SmartSearch** to identify the best matches for unknown spectra, by searching through a library of known spectra.

Section 6 demonstrates how to examine supplementary details after performing a search, including details taken from sources on the web.

Section 7 describes how to install, select, and activate **Analyze IQ** spectra libraries and third party libraries.

Section 8 provides details of the search algorithms available in **SmartSearch**.

Section 9 contains some information on technical support and sales, and finally **Section 10** provides a list of references.

2   Installation

Analyze IQ SmartSearch can be installed and run under all recent versions of the Windows desktop operating system, including Windows XP, Windows Vista, Windows 7, and Windows 8. It works fully on both 32-bit and 64-bit operating systems.

As part of installing **SmartSearch**, Microsoft SQL Server *Express Edition* is installed as it used by **SmartSearch** as the underlying database software. Depending on your system's configuration, it may also require installation of the Microsoft.NET Framework and Windows Installer. If the same components, or later versions of them, are already installed on your system, they will not be re-installed or you may manually cancel their installation. In particular, if an appropriate database version is already installed, either as part of a previous installation of **SmartSearch** or as part of an installation of **Analyze IQ Spectra Manager**, the installer will give you the option to skip reinstalling it.

Please do not hesitate to contact Technical Support with any concerns, or if you require assistance with installation of the software. Please refer to Section 9 for contact information.

2.1   USB Key Installation

To install from USB key, insert the USB key into a USB slot and follow the instructions that appear after you open it.

If the installation screen does not appear automatically when the USB key is inserted, navigate to the top level of the drive and run the **StartHere.exe** program to start the Analyze IQ software installation.

2.2   Website

If you wish to download **Analyze IQ SmartSearch** from the Analyze IQ website, www.AnalyzeIQ.com, you must first register on it, unless you have previously registered to download other products in the **Analyze IQ** product suite, for example **Analyze IQ Lab** or **Spectra Manager**. To do so, go to the website and click on *Register* under the login form. A confirmation email is sent to the address that you entered into the registration form. This confirmation email includes a link for activating your new Analyze IQ account.

After you have registered as a user, you can download the **Analyze IQ SmartSearch** software package at the following address: http://www.AnalyzeIQ.com/Download.html. This page allows you to download the software setup file. After downloading the setup file, run it to start the installation process.

2.3   Getting a License

At the end of the installation process, when you first run **Analyze IQ SmartSearch**, you will be prompted to 'Get a License', which requires an internet connection. This final step must be carried out in order to be able to run **Analyze IQ SmartSearch**.

If you are installing an evaluation copy of the software, you will enter your registered username to get a license after you install the software. Otherwise, you will enter a Commercial License Code that is given to you by Analyze IQ Ltd, or your software supplier, when you purchase the software.

Using the *Help* menu in **SmartSearch** (shown below), choose the *About* menu command to view your license information and the *Get Support* menu command to get links to contact information. You can use the *Update License* command to update your license, for example when switching from an evaluation license to a commercial license. In addition, the *Open User Manual* command opens this document.
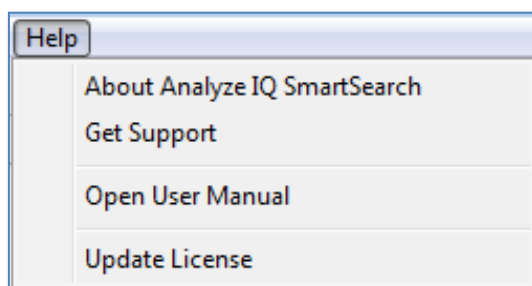

Figure 1: The SmartSearch Help menu

3   Data stored in Analyze IQ Libraries

Analyze IQ spectral data libraries, which are used by **SmartSearch**, **Spectra Manager** and **Analyze IQ Lab**, stores a list of spectra, where each spectrum is assigned a unique ID within the library. In addition to storing the frequency and spectral intensity for each data point in the spectrum, the library maintains information related to the sample for which the spectrum was recorded:

- It stores a list of *substances,* where a substance is a compound that is defined by a CAS number and an IUPAC name.
- It stores a list of *constituents,* where a constituent refers to a particular specimen of a substance and is described by the following properties: *Substance, Manufacturer, Lot Number, Catalog Number, Purity, Location* and *Date Opened.* Note that apart from substance, the other properties of a constituent may be unspecified. For example, a substance with IUPAC name Acetonitrile and CAS Number 000075-05-8 is stored in library. A sample of Acetonitrile manufactured by Aldrich with 99% purity and Lot Number 3725 is an example of one constituent. For each substance recorded in the library, one or more constituents are also recorded.
- A spectrum represents either a *pure* sample, which comprises a single constituent at 100% concentration, or a *mixture* of two or more constituents, where the sum of the constituent concentrations is equal to 100%.

If you purchase the **Analyze IQ Raman Library**, all of its spectra and associated information are included in **SmartSearch** and/or **Spectra Manager** when you install them.

Third party spectra libraries are also available; contact us for details (see Section 9).

If you have **Spectra Manager**, you can use it to create your own *user libraries.* With **Spectra Manager**, you can add and remove spectra to/from your libraries, and you can edit associated details. (Please refer to the ***Spectra Manager User Manual*** for further information.) You can then use **SmartSearch** to perform searches against your libraries.

4   Quick Start: Performing a Basic Search

*The key purpose of **Analyze IQ SmartSearch** is as follows: given an unknown spectrum that you are viewing (referred to as the **query**) and a set of known spectra (stored in a **library**), compare the query spectrum with every spectrum in the library, in order to produce a ranked list of the best matches.*

Performing a basic search is simply a matter of loading a query spectrum, pressing the **Search** button, and examining the results!

4.1   Opening a Spectrum File

A spectrum file can be opened either via the menu or the shortcut (see Figure 2). In either case, you are presented with an *Open File* window, in which you browse to the appropriate directory, select a spectrum file and click on *Open*.

Sample spectrum files can be found in a sub-folder of the main **Analyze IQ SmartSearch** installation folder. For the purposes of this example, we will open a file called "1-Bromobutane.aiq", which is in the AIQ spectrum format and stores a spectrum of 1-Bromobutune.
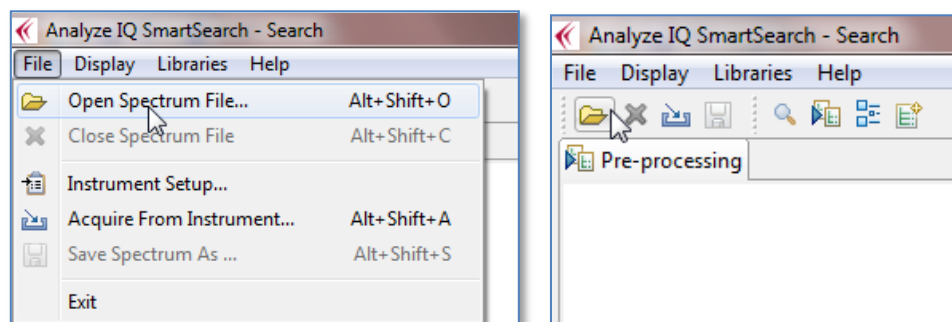

Figure 2: Open Spectrum file using: (left) the menu; (right) the shortcut

When the spectrum file in any format is opened, details of the file contents are shown in the right-hand pane and a graph of the spectrum is shown in the middle of the **Analyze IQ SmartSearch** screen (see Figure 3 below).

4.2   Performing a Search

Now that a query spectrum is loaded, the *Search* button appears to the *Search* pane, as shown in Figure 3; **Analyze IQ SmartSearch** is now ready to perform a search.

To perform the search, pick your *search settings* (See Section 5.6.1) and then press on the *Search* button.
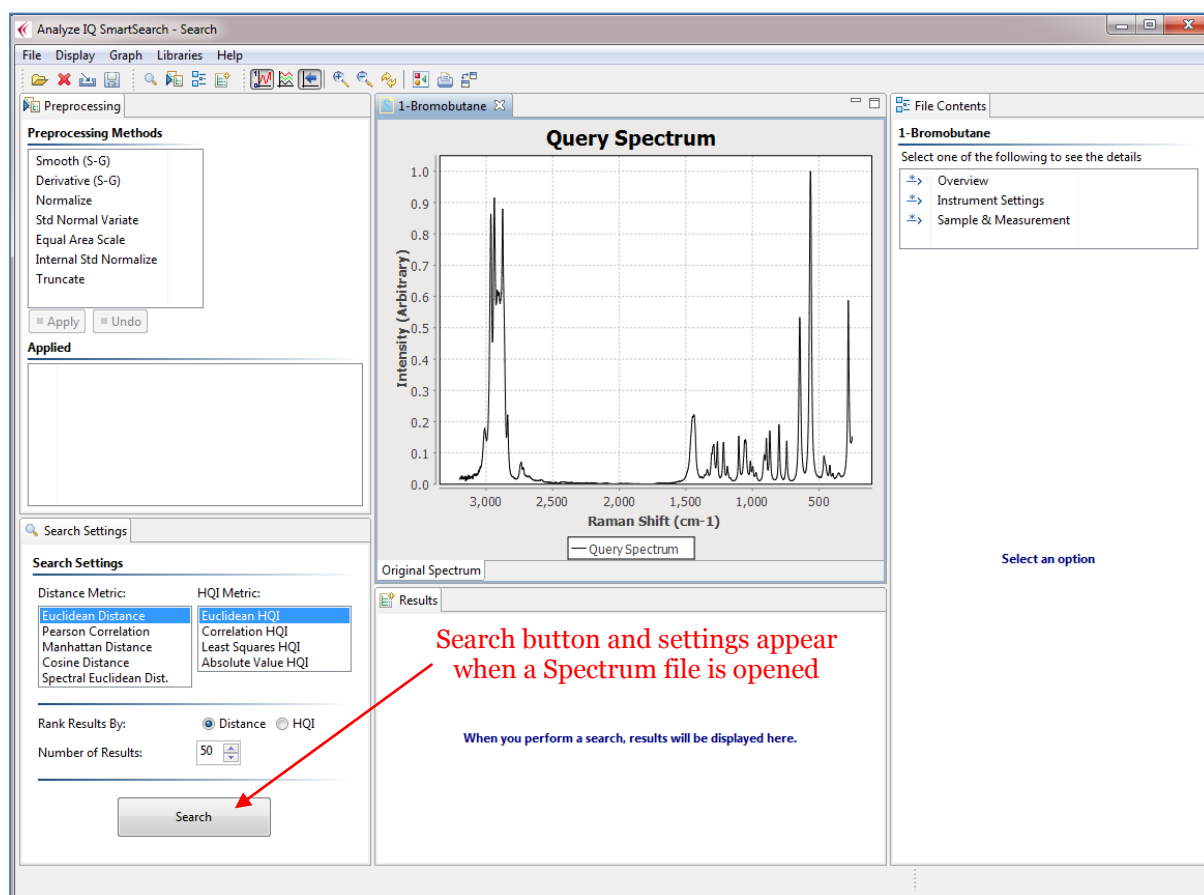
Figure 3: Analyze IQ SmartSearch page after Spectrum file is opened

4.3 View Search Results

When a search has been performed, the results are displayed in the bottom-middle area of the Analyze IQ SmartSearch screen. Figure 4 shows the results of applying a search on the spectrum with no pre-processing and using the Euclidian Distance Comparison Metric. There are two aspects to the display of results.

Firstly, the *Results* pane in the lower middle area of the page lists the set of spectra that have been identified and ranked as the closest matches for a spectrum of an unknown substance of interest as shown in Figure 4.

Figure 4: Lists the set of spectra that have been identified.

Secondly, if a result is selected by ticking the box as shown in Figure 5, this new spectrum is now shown in the spectrum viewing area of the page as well as the original spectrum.
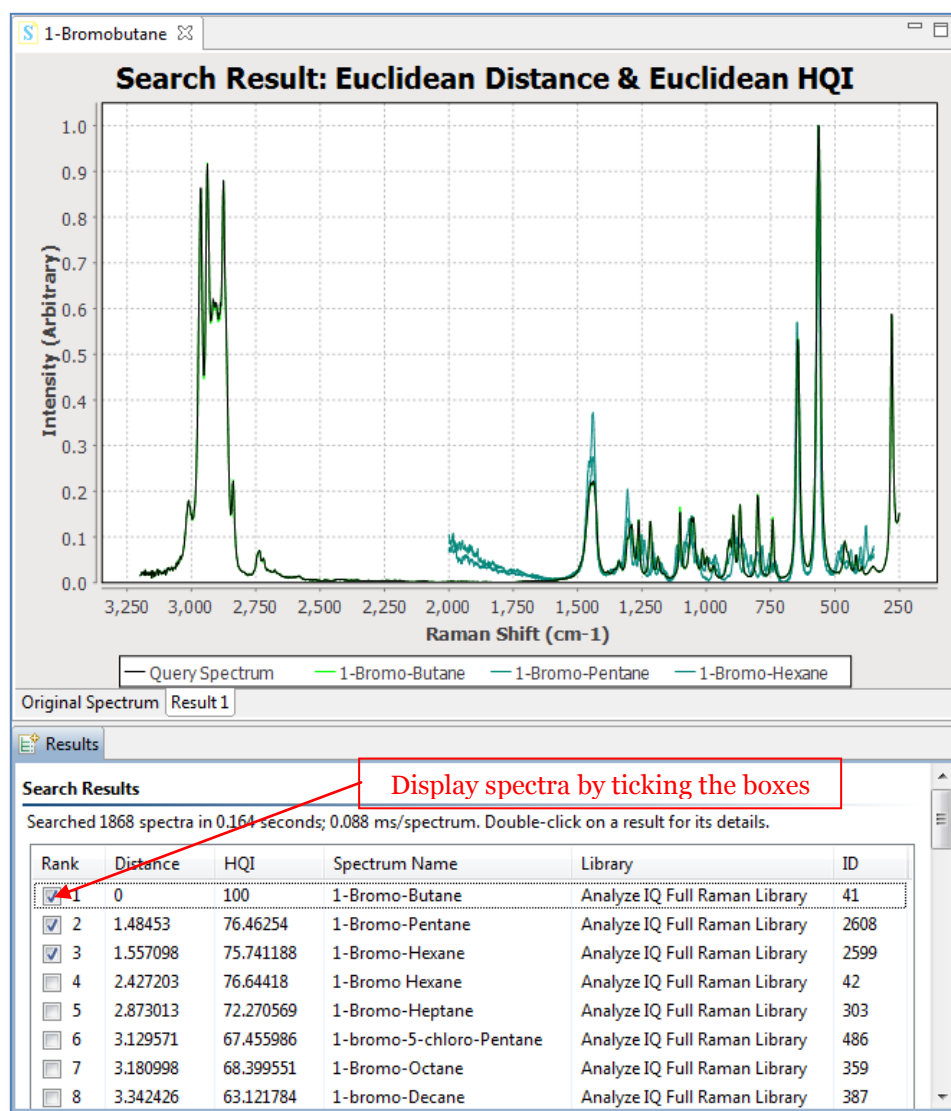

Figure 5: Select Spectra in results pane to see in the Spectrum viewing area.

5   Details of Using SmartSearch

This section explains in greater detail how to use **Analyze IQ SmartSearch** to search rapidly and accurately through a library of spectra, in order to identify the best matches for an unknown query spectrum.

5.1   *The SmartSearch User Interface*

When **SmartSearch** is first started, you are presented with the page shown in Figure 6. This figure highlights the five areas of this page, which are as follows:

1. ***Pre-processing****:* shows the pre-processing techniques that can be applied to the currently opened spectrum and the library. It also shows a history of what pre-processing steps have already been applied, so that they can be removed.
2. ***Search Settings****:* Shows a set of spectrum comparison metrics.
3. ***View Spectrum****:* displays a plot of the spectrum of the currently opened file.
4. ***View Results****:* after performing a search, presents a ranked list of spectra that are the closest matches for the query spectrum.
5. ***Spectrum Details****:* shows details of the currently opened spectrum file, such as date recorded, instrument manufacturer, etc.

When **Analyze IQ SmartSearch** is started, all areas of the page are empty, until you load a spectrum.
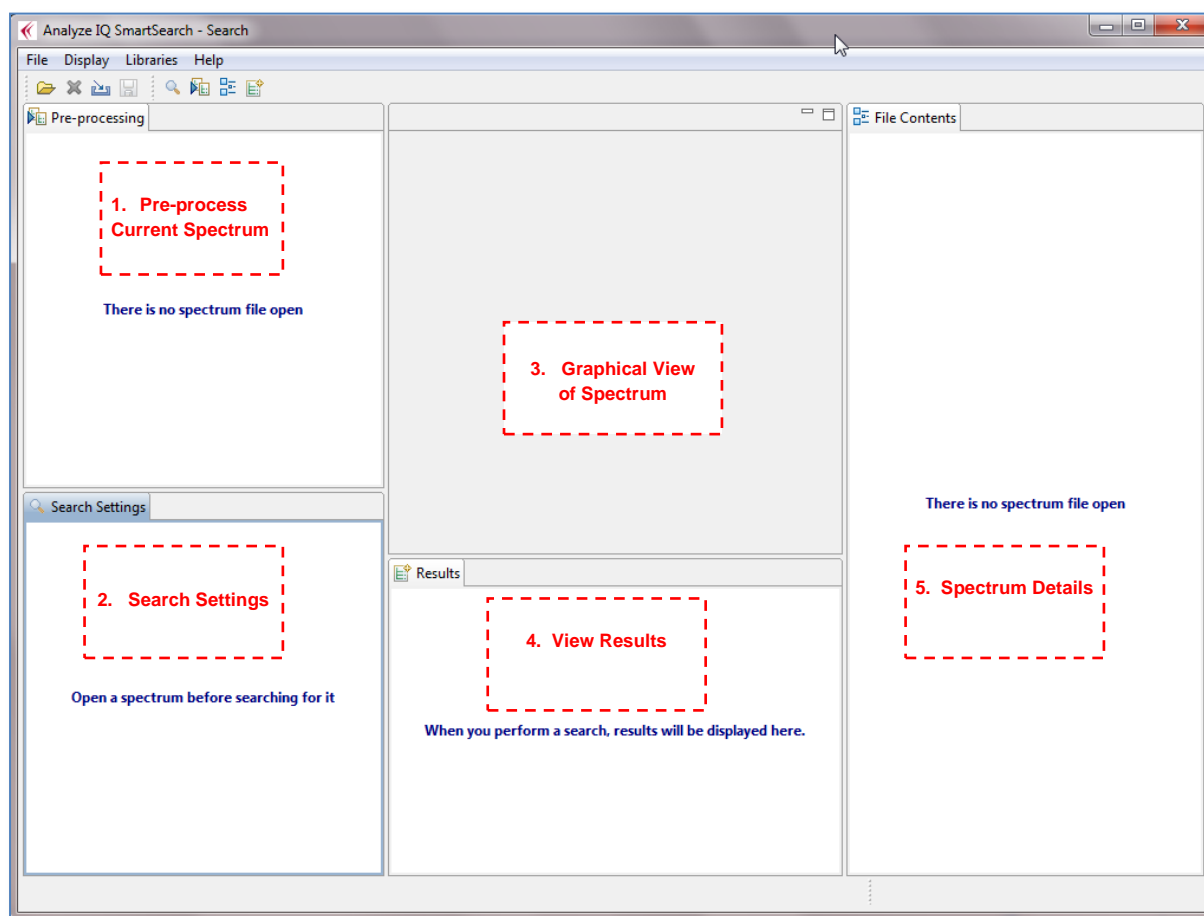
Figure 6: Analyze IQ SmartSearch Screen

5.2   Details of Steps to Perform a Search

The purpose of performing a search with SmartSearch is to produce a ranked list of the spectra in a library that are the best matches for an unknown spectrum that you wish to identify, and that is referred to as the *query spectrum*.

The full set of steps is as follows:

1. Load the query spectrum by either opening a spectrum file (see Section 5.3), or acquiring a spectrum from an instrument (see Section 5.3.2).
2. Examine the spectrum graph (see Section 5.4) and associated data (see Section 5.5).
3. Select a library to search against (see Section 7).
4. Optionally, select and apply pre-processing (see Section 5.6).
5. Select search settings: a Distance Metric, a Hit Quality Index (HQI) Metric, how many results to display, and whether to rank results by Distance or HQI (see Section 5.7.1 for a description of the Search Settings pane and Section 8 for full details of search algorithms).
6. Press the *Search* button.

7. Examine the search results (see Section 5.7.2).
8. Click on a search result to view associated information, including details from the Spectra Manager library and relevant information from Wikipedia (see Section 6).

The following sections explain these steps.

5.3   Open a Spectrum File

As was described in Section 4.1, a spectrum file can be opened either via the menu or the shortcut (see the figure below). In either case, you are presented with an *Open File* dialog, in which you browse to the appropriate directory, select a spectrum file and click on *Open*.

Sample spectrum files can be found in a sub-folder of the main **Analyze IQ SmartSearch** installation folder. For the purposes of this example, we will open a file called "1-Bromobutane.aiq", which is in the AIQ spectrum format and stores a spectrum of 1-Bromobutune.
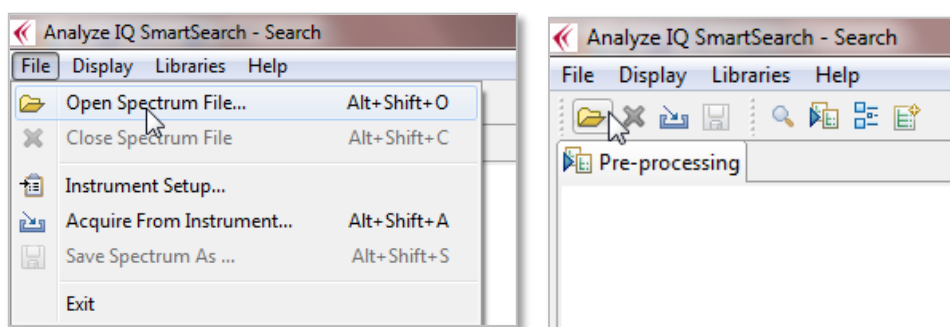
Figure 7: Open Spectrum file using: (left) the menu; (right) the shortcut

5.3.1 Supported File Formats

Analyze IQ SmartSearch supports several popular file formats; see the figure below:

1. **SPC:** Thermo-Fisher's GRAMS SPC file format.
2. **SpectroML:** NIST's XML-based format for spectroscopy and chromatography data.
3. **AIQ:** The .AIQ file format is used to store data in a suitable format for Analyze IQ. It is also used in Instrument Interfaces. AIQ files use an XML format that is almost a subset of the NIST SpectroML format, **except** that it has an added property, instrumentSetting.excitationLine, that does not occur in SpectroML.
4. **JCAMP-DX:** IUPAC's popular JCAMP-DX file format. (Note that there are many variations of JCAMP-DX format; the Analyze IQ software supports most but may not support all. Contact Technical Support if you have problems.)
5. **XY Text Files:** These are simple files with either two rows or two columns of data, the first with X values and the second with Y values. See below for further details.
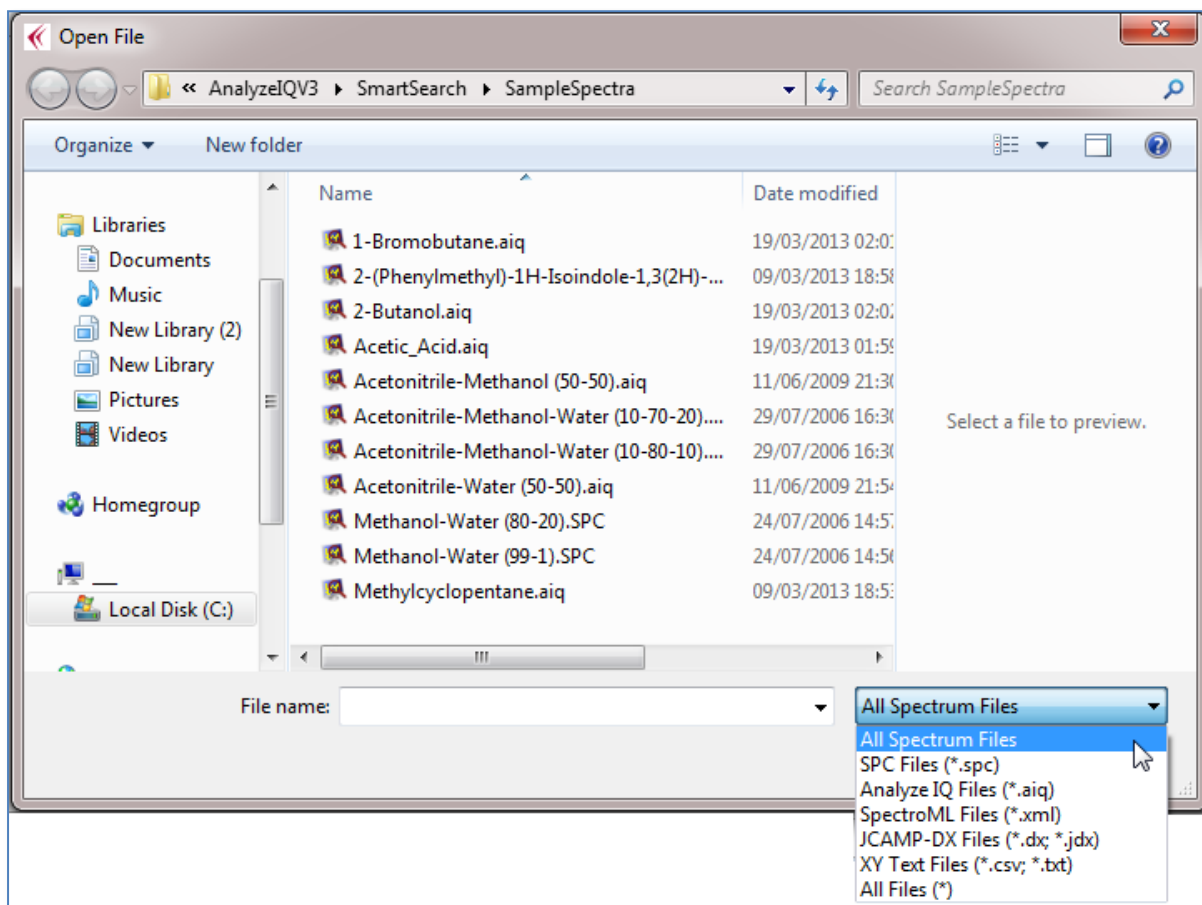
Figure 8: File Formats

As noted above, an *XY Text File* format simply consists of two rows or two columns of data, the first with X values and the second with Y values. Values may be separated by tabs or commas; the latter are known as Comma Separated Values (CSV) files. You may optionally include X and Y axis labels by putting them at the start of the two rows/columns.

The XY Text File format is particularly useful when you need to import data into the **Analyze IQ** software suite from a software product that does not support any of our standard formats. You can easily generate files in this format with a spreadsheet program (such as Microsoft Excel or OpenOffice Calc) or a text editor.

An example of the start of an XY Text File is shown below; this one is in two rows and includes axis labels:

Wavenumber (cm-1)	250.0	252.0	254.0	256.0	...
Intensity	2454.0	2439.0	2307.0	2202.0	...

Figure 9: Example of the start of an XY Text File in row format with axis labels

When the spectrum file in any format is opened, details of the file contents are shown in the right-hand pane and a graph of the spectrum is shown in the middle of the *Home Screen* page (see Figure 10 below).

Note also that the *Search* button and search settings appear in the *Search* pane; **Analyze IQ SmartSearch** is now ready to search through a library to find best matches for the query spectrum.
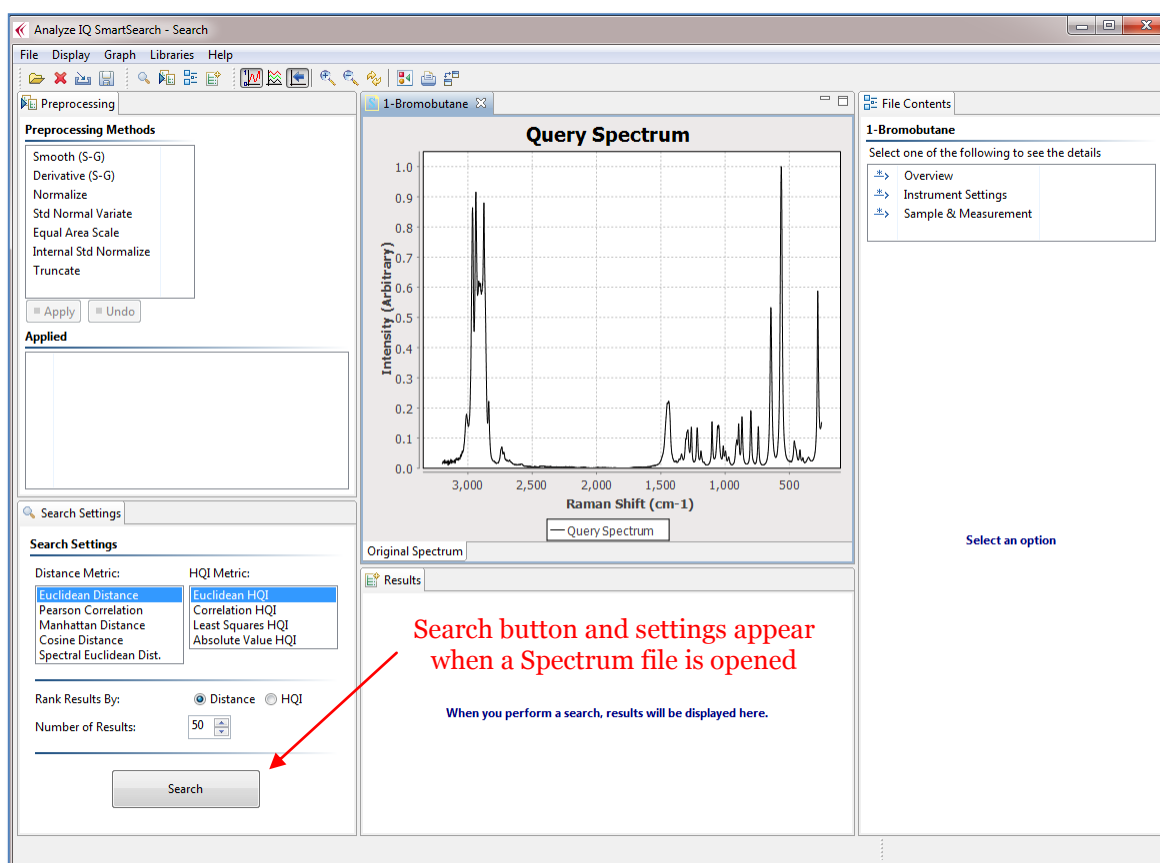

Figure 10: Analyze IQ SmartSearch screen after Spectrum file is opened

5.3.2 Acquire from Instrument

With **Analyze IQ SmartSearch** and other products in the **Analyze IQ** software suite, you can acquire a spectrum directly from an attached spectrometer, provided that appropriate supporting software has been supplied with your installation.

You can select which instrument to use by using the *Instrument Setup* menu command in the *File* menu, as seen in Figure 11 below. This opens an Instrument Setup dialog box (not shown) which shows the status of all installed instruments and allows you to select one.

After you have selected the instrument, you then can acquire a spectrum. To do this, select the *Acquire From Instrument...* command from the *File* menu, as shown in the figure below. An Instrument dialog box will appear, allowing you to specify settings

appropriate to the specific instrument. (This dialog box is not shown, and its appearance varies from instrument to instrument.)
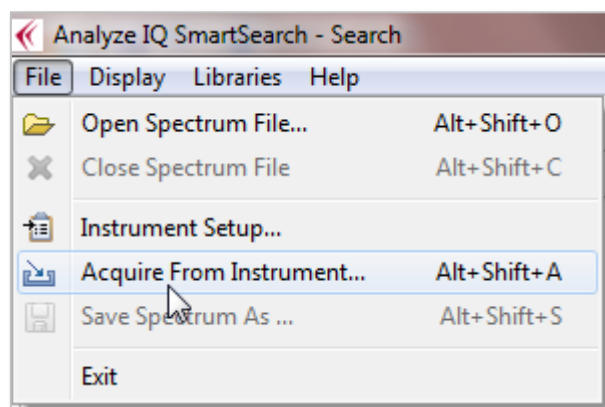

Figure 11: Menu options to select an instrument and acquire a spectrum directly from the attached instrument

Press the *Acquire* button on the Instrument dialog box to begin data acquisition. When data acquisition is finished, the Instrument dialog box will close and the data will be displayed, just as if you had opened a file.

If you are satisfied with the spectrum that you have acquired from the instrument, you will want to save it, as described next.

5.3.3 Save a Spectrum

The *File – Save Spectrum As...* menu function to save your data in either AIQ or SpectroML format. All ancillary information, as displayed in the File Contents pane, is also saved.

Note that the spectrum data is saved in its 'raw' form, **without** any pre-processing you may have performed, such as normalisation or computing a derivative (see Section 5.6 for a description of preprocessing operations). This is to avoid problems subsequently arising with spectra being treated as 'raw' when they may have been pre-processed in a way that greatly changes their appearance, such as with a first derivative transform.

Note also that **Analyze IQ SmartSearch** does not currently support saving SPC files; they are "read-only".

5.4   Examine the Spectrum Graph

When a query spectrum is open, a graphical view of it is displayed in the top centre of the **Analyze IQ SmartSearch** screen as shown in Figure 12 below.
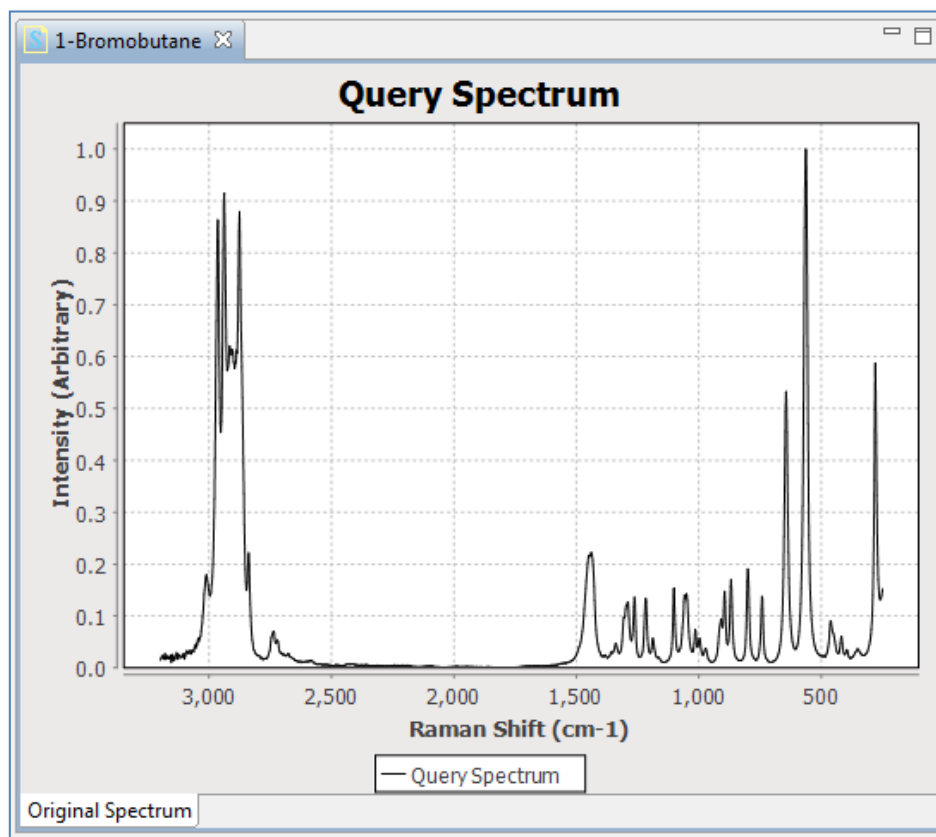
Figure 12: Graphical view of query spectrum

5.4.1   Adjusting the Spectrum Plot

The menu bar buttons shown below can be used to adjust to the graph's appearance, and to save and print it, as described in the sub-sections below. The same functions are available on the *Graph* menu, as shown in Figure 14.
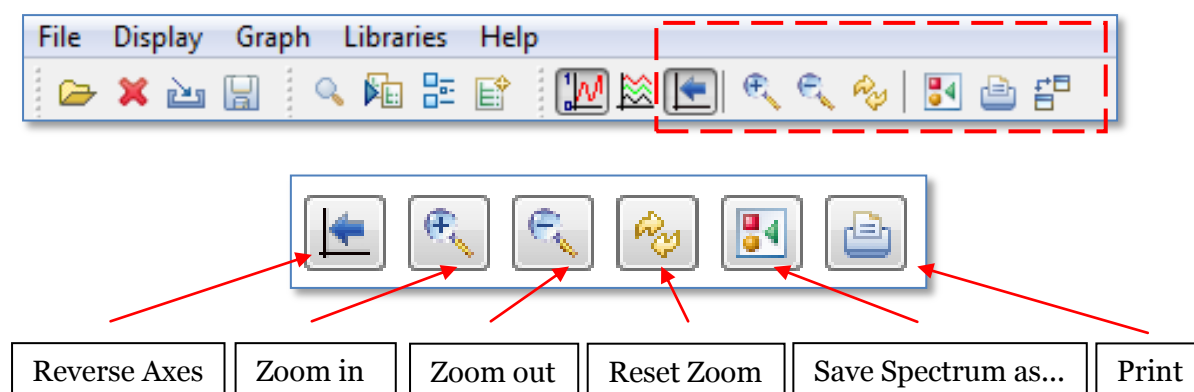
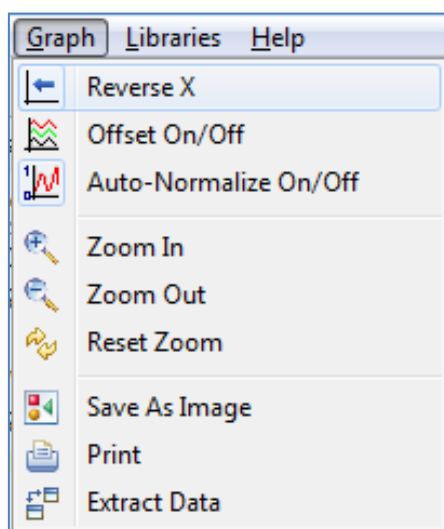**Figure 13: Menu bar buttons to adjust the spectrum plot**

Figure 14: Graph menu commands corresponding to the menu bar buttons

Note that two of the commands, *Offset On/Off* and *Auto-Normalize On/Off*, are used when viewing the results of a search, so they are described later on, in Section 5.7.2.

5.4.1.1 Reverse X Axis

Some types of spectra, such as those where wavenumbers or with units such as cm⁻¹, are commonly plotted with the X axis decreasing. When you open a spectrum, **SmartSearch** checks the axis labels and attempts to automatically decide what the axis direction should be. However, you can override this by pressing the Reverse X Axis button, as shown in Figure 15.
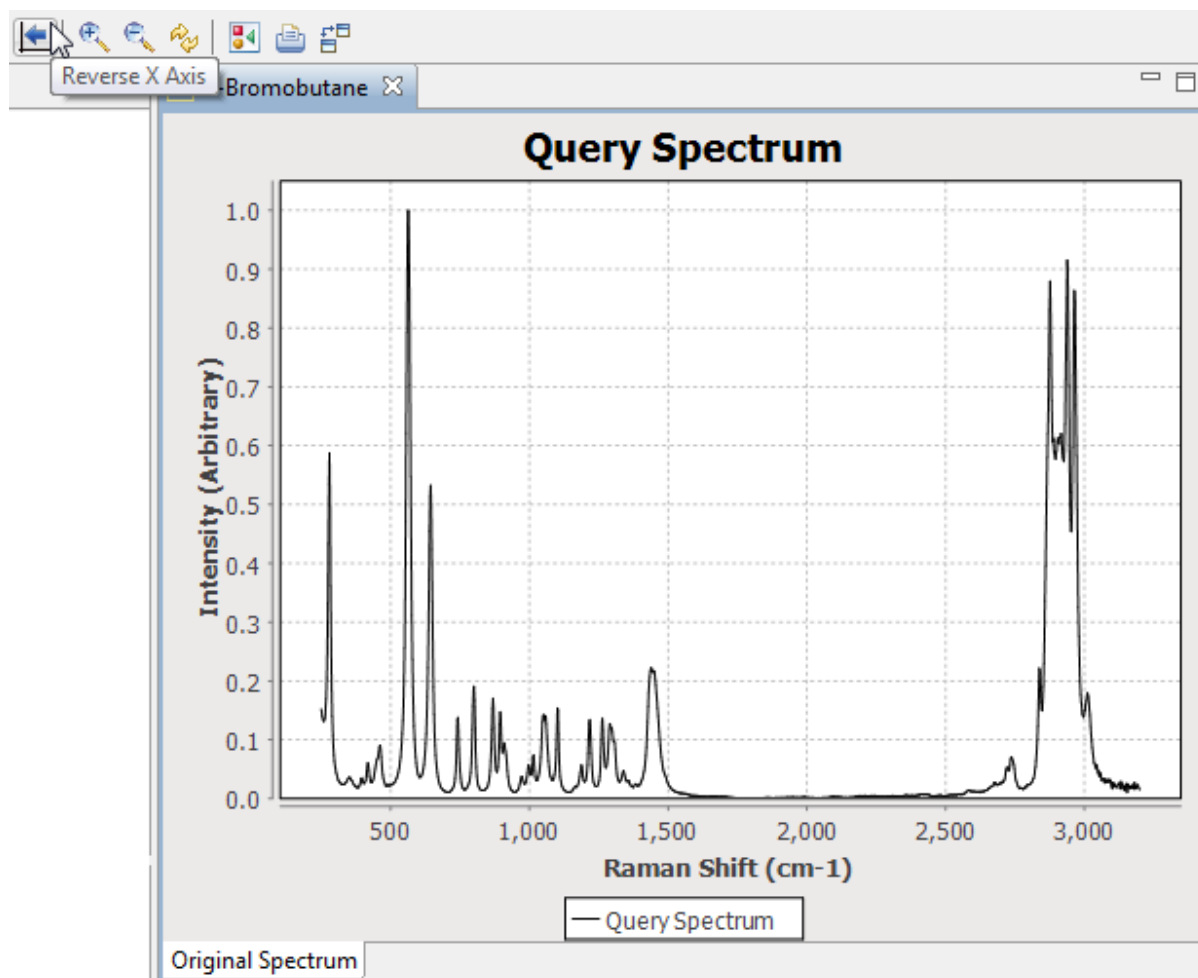
Figure 15: Reverse X Axis

5.4.1.2 Zoom In and Out on Spectrum Plot

You can zoom in and zoom out of specific regions of a plot using the *Zoom In* and *Zoom Out* buttons shown in Figure 13, and also by selecting regions directly in the plot area.

To zoom in on a specific area (e.g. a peak) of the plot, click on the plot area and draw a rectangle around the section of interest. For example, selecting the area shown in Figure 16 results in the plot displayed Figure 17. Click on the *Reset* button indicated in Figure 13 to bring the plot back to its original settings.
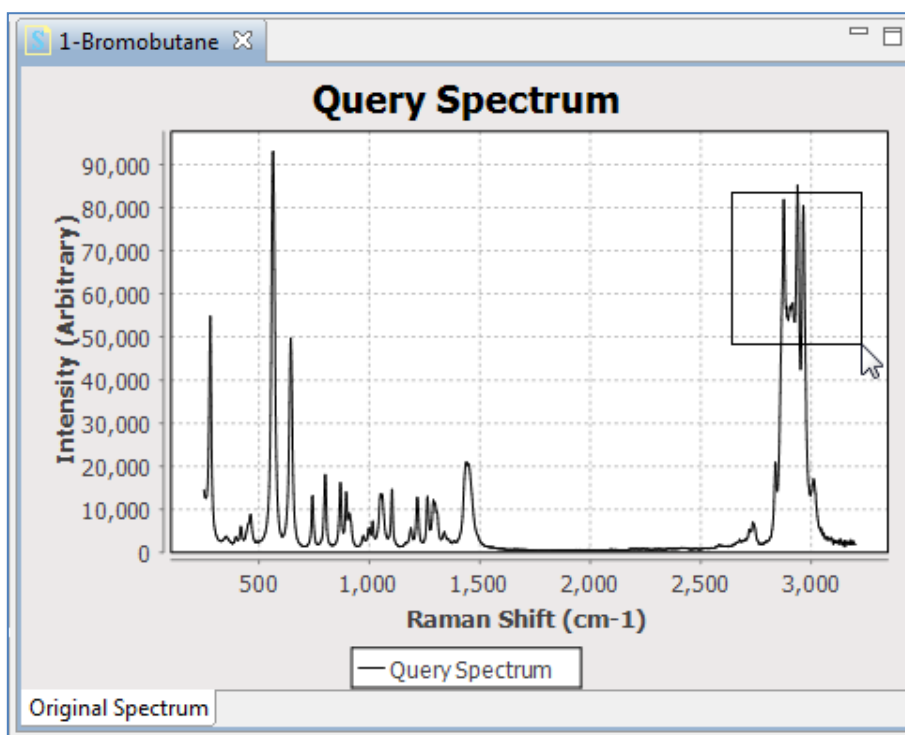
Figure 16: Zooming in on Spectrum Plot
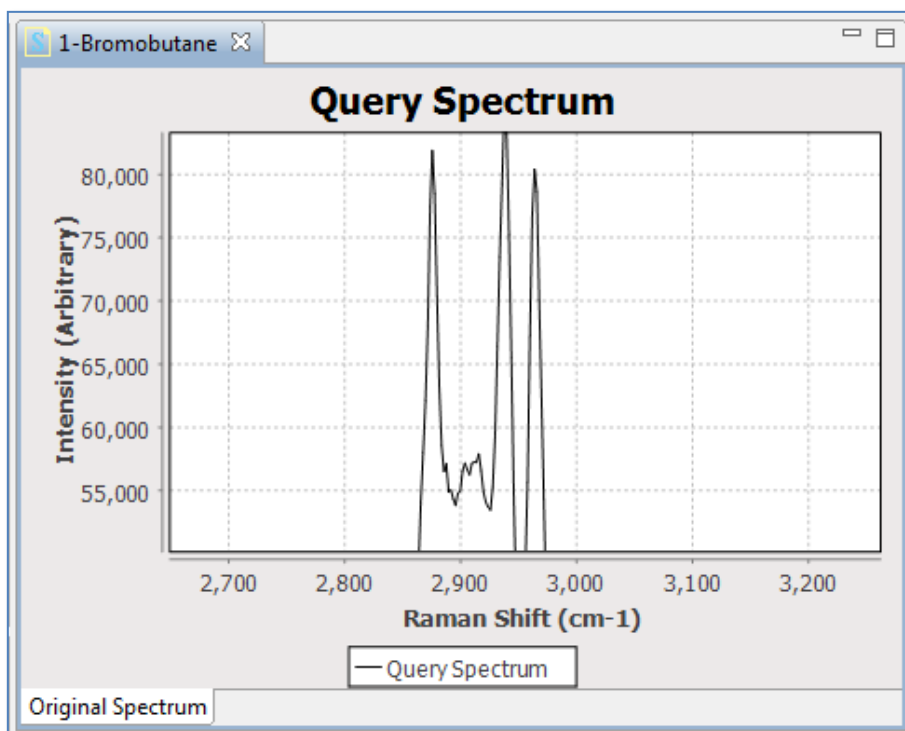

Figure 17: Zooming in on Spectrum Plot

5.4.1.3 Saving and Printing a Plot Image

As shown in Figure 13 and Figure 14, there are also menu bar buttons and menu commands to allow the user to save the spectrum as an image or print it.

5.5   Viewing Associated Details

5.5.1 Viewing Spectrum Details: SPC Format

When a spectrum file is opened, you can view details of the file, by selecting one of the categories listed in the *Spectrum File Contents* pane on the right-hand side of the *Analyze IQ SmartSearch* page. The categories that are available depend on what kind of file you have opened.

Figure 18 demonstrates the details that are available when an SPC is open. In the figure, the *Log Data* section of the currently opened spectrum file is displayed. *Log Data* contains details such as the date the spectrum was recorded on, the detector temperature, whether background was acquired, etc. Clicking on *Header*, *Sub-Header* and *Log Data Settings* displays other details of the file.
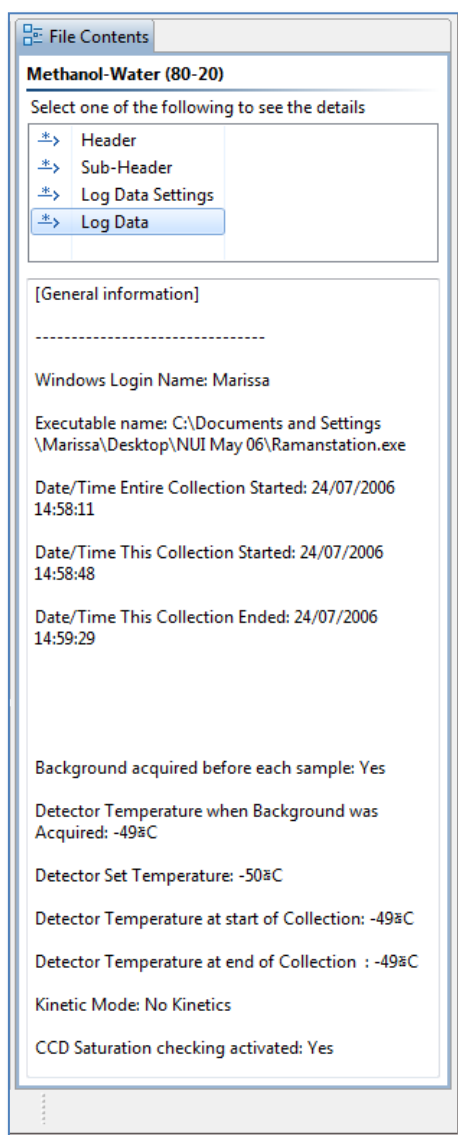

Figure 18: View SPC file contents: Log Data section (Right side of screen)

5.5.2 Viewing Spectrum Details: AIQ or SpectroML Format

The data contained in AIQ or SpectroML files is in general more accessible to interpretation than that in SPC files.

Figure 19 shows an example of what is displayed when an AIQ or SpectroML formatted file is opened. There are three headings, *Overview, Instrument Settings,* and *Sample & Measurement*. Click on a heading to display the relevant details. If a specific detail has not been stored, the text "(none)" is displayed. In the figure, the *Instrument Settings* heading has been selected, and details such as the experiment title, instrument manufacturer, model and resolution are visible.
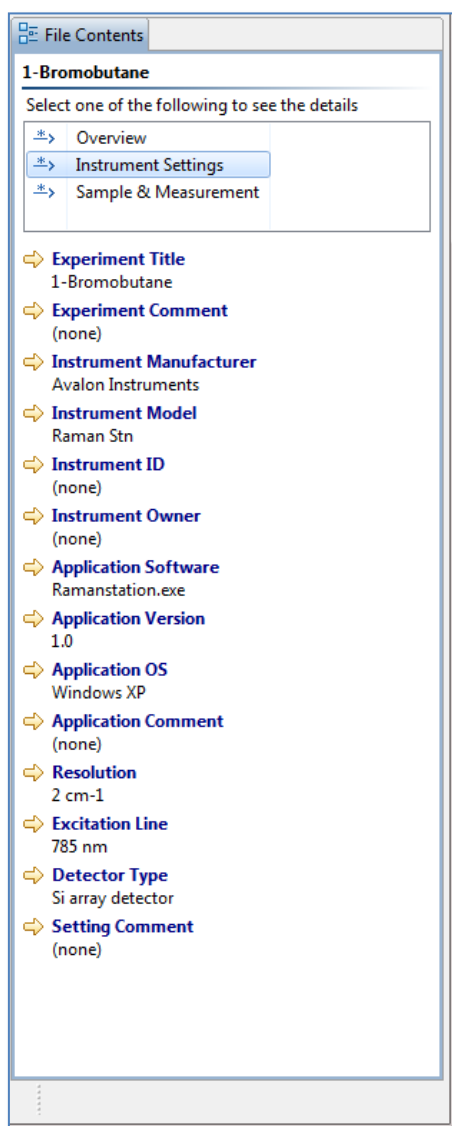

Figure 19: View AIQ file contents: Instrument Settings section (right side of screen)

Note that files in other formats, such as XY Text Files or JCAMP-DX, are converted internally to AIQ format. They are therefore displayed in the same way as the above, although many of the details may be unspecified.

5.6   Preprocessing

5.6.1 Preprocessing Methods Available

The preprocessing methods listed in the top left-hand top corner of the **Analyze IQ SmartSearch** screen, as shown in Figure 20, can be applied to the data.

Note: when you apply a preprocessing method, it is applied not only to the query spectrum shown on screen but to the currently loaded library, so that any searches are performed on library of spectra that are comparable with the query spectrum. For this reason, if you have a large library selected, there may be a slight delay in applying the more computationally intensive preprocessing methods (Smoothing and Derivative).

Different methods have different settings associated with them; when you select a method, the associated settings are displayed beside it. For example, Figure 20 shows the settings associated with the *Smoothing* preprocessing method.
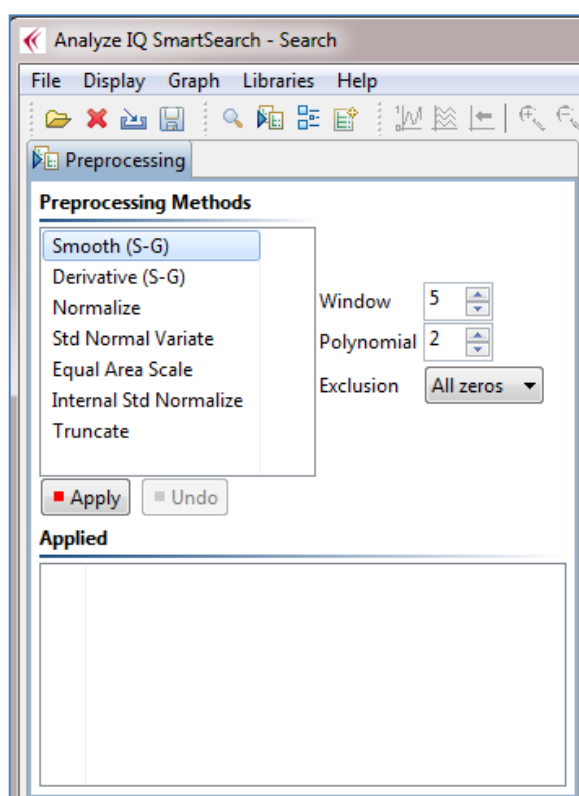

Figure 20: Preprocessing methods in Analyze IQ SmartSearch

The details of the preprocessing methods are as follows:

1. ***Smooth (Savitzky-Golay):*** This is a method for smoothing curves, to remove small, high-frequency fluctuations. It involves performing a local polynomial regression (of degree *k*) over a sliding window of at least *k*+1

equally spaced points, to compute a moving average. See *Savitzky & Golay* [2] for more details on this preprocessing technique.

2. ***Derivative (Savitzky-Golay):*** This is a method to compute a derivative of a curve. It is strongly related to the Savitzky-Golay Smoothing method; it involves performing a local polynomial regression (of degree k) over a sliding window of at least k+1 equally spaced points to determine smoothed value for each point, and then fitting computing the derivative based on the local curvature of the polynomial. Note that selecting a zero-order derivative is equivalent to smoothing. See *Savitzky & Golay* [2] for more details on this pre-processing technique.

3. ***Normalization:*** This performs 0/1 Normalization of spectra. Each sample is re-scaled, so that the each one's lowest point corresponds to 0 and its highest point corresponds to 1.

4. ***Standard Normal Variate:*** The SNV transform re-scales each spectrum, so that the points on the spectrum end up having an average of 0 and a standard deviation of 1. This is known to produce results that are very similar to **Multiplicative Scatter Correction** (MSC), and in practice results in spectra with similar properties for chemometric modeling [1].

5. ***Equal Area Scale:*** This transformation rescales all spectra so that they all have equal area under their curves. You can specify a **scale factor**, which defaults to 1000. The larger the scale factor, the larger the overall height of the scaled spectra.

6. ***Internal Standard Normalize:*** This is like the standard Normalization, except that instead of the spectrum being scaled such that its overall highest point is rescaled to 1, with Internal Standard Normalization, a point at the specified X value is rescaled to 1. This is typically used where a specific peak in the spectrum corresponds to a component that is known and constant (the internal standard), and all other spectra are to be normalized relative to it. In using this, you must specify the **X Location** of the internal standard. To allow for minor variations in measurements between spectra, you can also specify a **Plus/Minus** value (which defaults to 0), and the spectrum is then normalized relative to the highest peak found in the range of the X ± this value.

7. ***Truncate:*** There are two forms of truncation supported, **Crop** and **Notch**:
   a. ***Crop:*** The start and end of the spectrum are cropped; the sections of the spectrum before the specified starting position and after the specified ending position on the X axis are replaced with the replacement value (which defaults to 0)
   b. ***Notch:*** A notch is removed within the spectrum; a section of the spectrum going from the specified starting position to the specified ending position on the X axis is replaced with the replacement value (which defaults to 0).

5.6.1.1 Settings for Smooth and Derivative Pre-Processing

The following is a description of the parameter settings used in the *Smooth (S-G)* and *Derivative (S-G)* methods:

1. ***Window:*** the length, specified in number of spectral data points, of the window over which smoothing is applied. (Default is 5.)
2. ***Polynomial Order:*** the order of the polynomial fitted to the spectrum for smoothing. (Default is 2, i.e. second order.)
3. ***Derivative:*** The order of the derivative: applies to *Derivative* only. (Default is 1, i.e. first derivative.)
4. ***Exclusion:*** determines whether wrapping is used (*Wrap* option) in the smoothing calculation for the beginning and the end of the spectrum or not (*All zeros* option). For example, if *Wrap* is selected, a window of length 5 centered on the last data point includes the previous two spectral data points and the first two data points of the spectrum. Similarly, a window of length 5 centered on the first data point of the spectrum includes the second and third data points and the last two data points of the spectrum. If *All Zeros* is selected with a window length of 5, the first two data points and the last two data points of the spectrum are set to zero as a result of the pre-processing.

5.6.2 Example of Applying Pre-Processing

For example, to apply a first order derivative to the spectra (see Figure 21):

1. Select *Derivative (S-G)* in the list under the *Pre-processing Methods* title.
2. Set the appropriate parameters. In this example, first-derivative (*Derivative* = 1, the default) using a second-order polynomial (*Polynomial* = 2) with a window size of 5 (*Window* = 5, the default) is being applied.
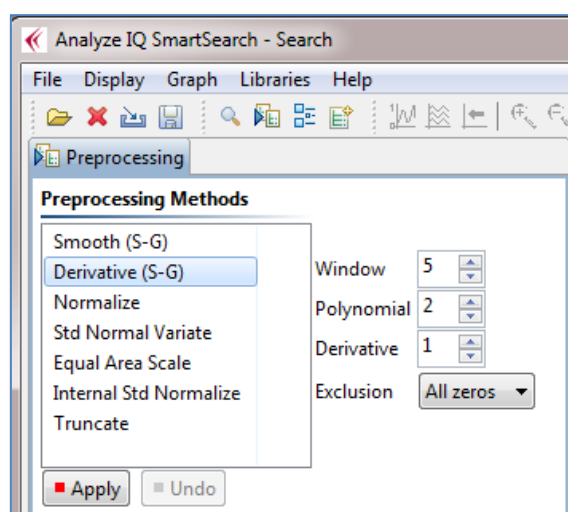3. Click on the *Apply* button.

Figure 21: Applying a Derivative (Savitzky-Golay) to the spectra

After applying the derivative, the spectrum displayed in the middle of the **Analyze IQ SmartSearch** screen page is modified accordingly as shown in below.
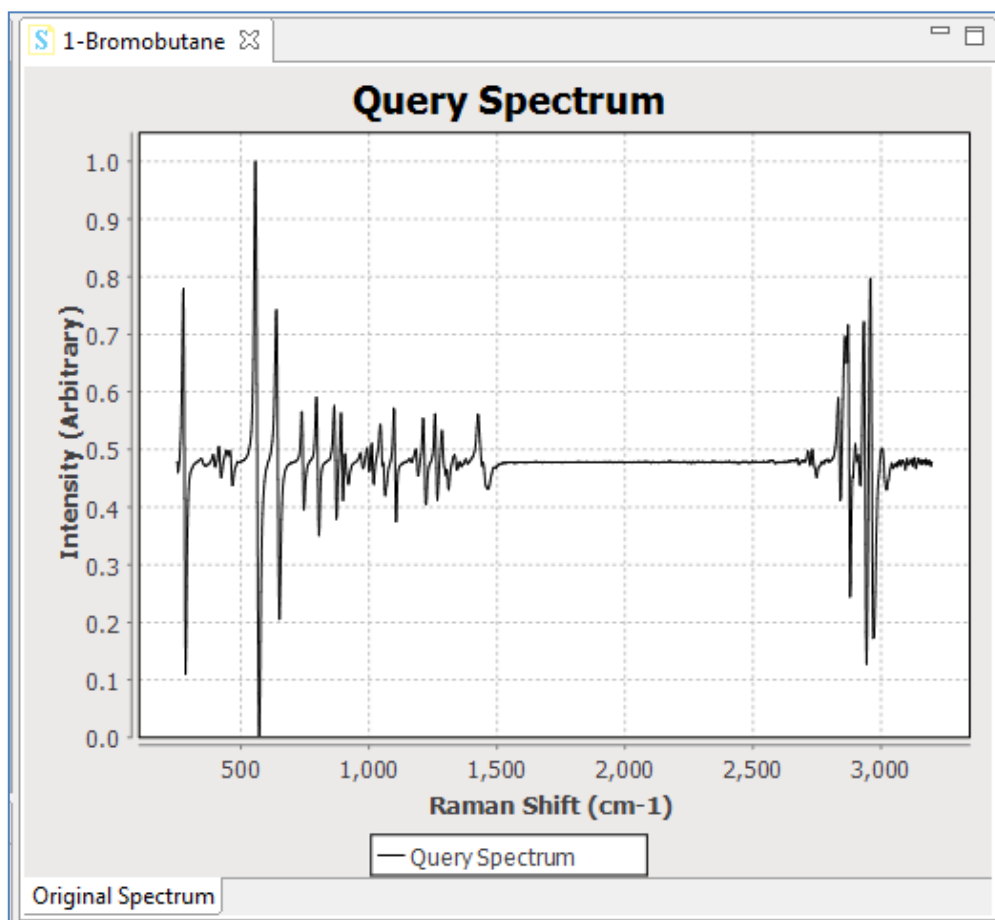

Figure 22: After applying a Derivative (Savitzky-Golay) to current spectrum.

Note that a history of all previously applied pre-processing steps is recorded under the *Applied* section in the *Pre-processing* pane. Each of these previously applied pre-processing steps can be undone by successive clicks of the *Undo* button.

5.7   Perform a Search

To search the library for results that best match the current query spectrum, firstly apply *pre-processing* if you wish (see Section 5.6), then pick your *search settings* (See Section 5.7.1 below) and click on the *Search* button. You can then examine the search results, as will be described in Section 5.7.2.

5.7.1 Search Settings

The Search Settings pane is shown below.

One of the distinctive features of **SmartSearch** is that when you are about to perform a search, you can choose both a distance metric and a HQI metric:

- **A distance metrics** is expressed as a number in the range 0 to infinity, with 0 being the distance between two identical spectra and the distance increasing as the spectra become less similar.
- **A HQI metric** is expressed as a number between 100 and 0, measuring how strong the "hit" is when comparing two spectra. HQI of 100 is the highest quality hit, when the two spectra are identical, and the HQI reduces as spectra become less similar.

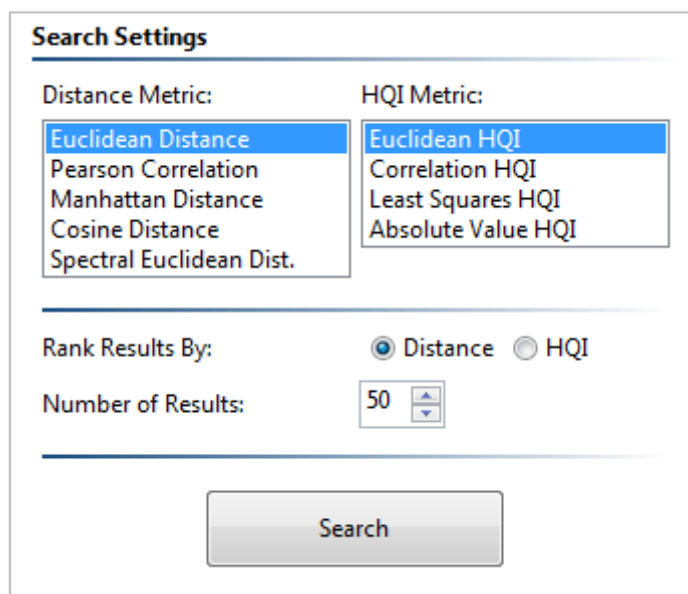Please refer to Section 8 for technical details of the distance and HQI metrics.

Figure 23: Search Settings pane

Selecting the distance and HQI metrics is done by clicking on entries in the list boxes in the upper part of the pane. Below that, you can choose to rank results by distance or HQI, and adjust the number of results that will be displayed.

Note that after performing a search, you can click on the headings in the Results list to re-sort the results by distance, HQI, or name. However, there is one difference in choosing to rank results by distance or HQI before you perform the search: since a finite number of results (for example, 50) are displayed, the top 50 results when ranked by HQI might not exactly be the same as the top 50 results when ranked by distance.

5.7.2 View Search Results

When a search has been applied to a spectrum, the results are displayed in the bottom-middle area of the **Analyze IQ SmartSearch** screen. Figure 23 shows the results of applying a search on the "1-Bromobutane.aiq" file opened previously, with no pre-processing and using the Euclidian Distance and Euclidean HQI metrics. There are two parts to the displaying of results.

Firstly, the *Results* pane in the lower-middle area of the page lists the set of spectra that have been identified and ranked as the closest matches for the query spectrum, as shown in Figure 23.

Search Results

Searched 1868 spectra in 0.164 seconds; 0.088 ms/spectrum. Double-click on a result for its details.

Rank	Distance	HQI	Spectrum Name	Library	ID
1	0	100	1-Bromo-Butane	Analyze IQ Full Raman Library	41
2	1.48453	76.46254	1-Bromo-Pentane	Analyze IQ Full Raman Library	2608
3	1.557098	75.741188	1-Bromo-Hexane	Analyze IQ Full Raman Library	2599
4	2.427203	76.64418	1-Bromo Hexane	Analyze IQ Full Raman Library	42
5	2.873013	72.270569	1-Bromo-Heptane	Analyze IQ Full Raman Library	303
6	3.129571	67.455986	1-bromo-5-chloro-Pentane	Analyze IQ Full Raman Library	486
7	3.180998	68.399551	1-Bromo-Octane	Analyze IQ Full Raman Library	359
8	3.342426	63.121784	1-bromo-Decane	Analyze IQ Full Raman Library	387

Figure 24: Lists the set of spectra that have been identified.

Secondly, when a spectrum in the results section is selected by ticking the box as shown in Figure 24, this new spectrum is now shown in the spectrum viewing area of the page as well as the original spectrum.
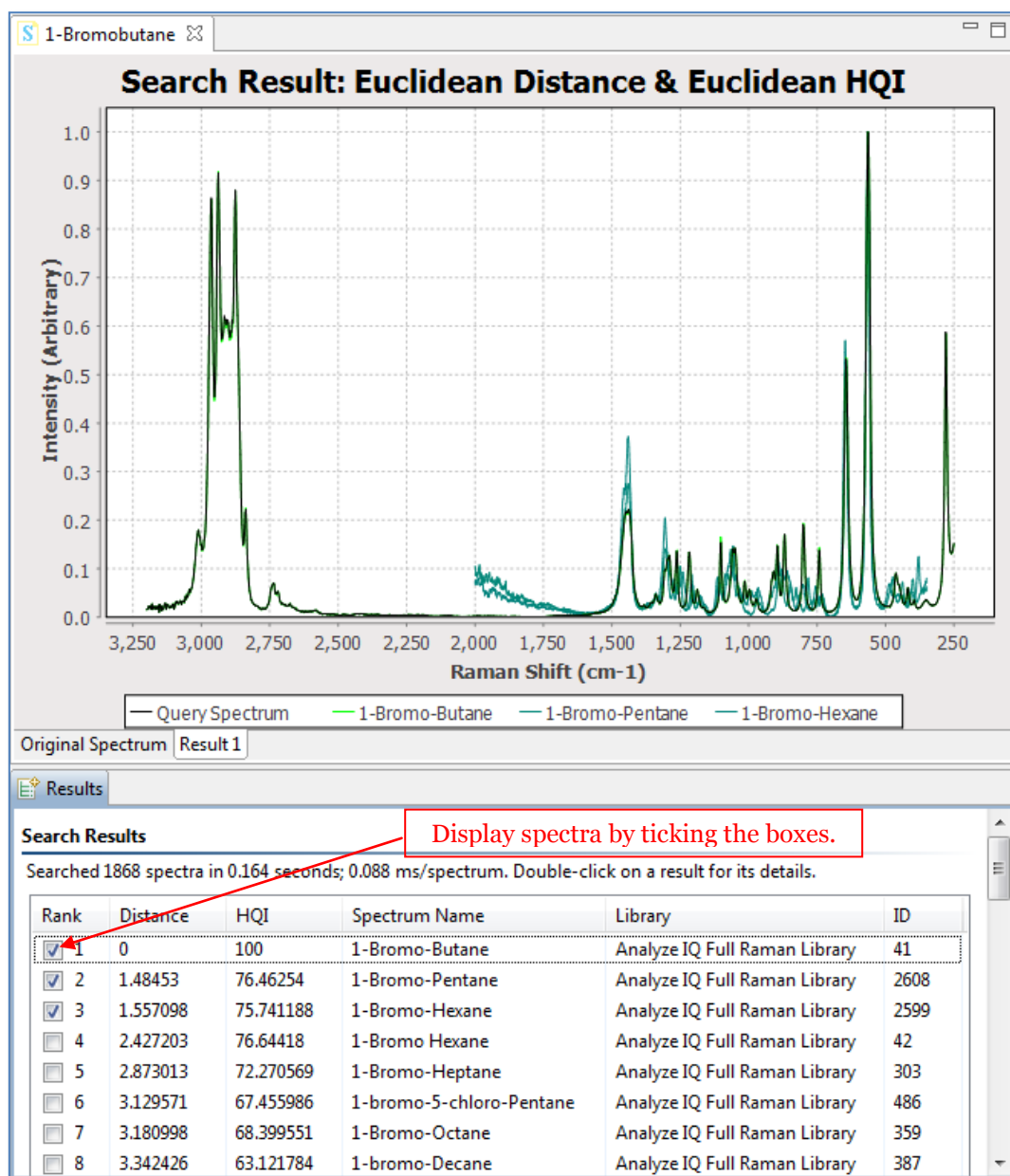
Figure 25: Select Spectra in results pane to see in the Spectrum viewing area

To apply an offset to the spectrum so they do not overlap directly, press the *Offset* button, as shown in Figure 26: Turning Offset On.

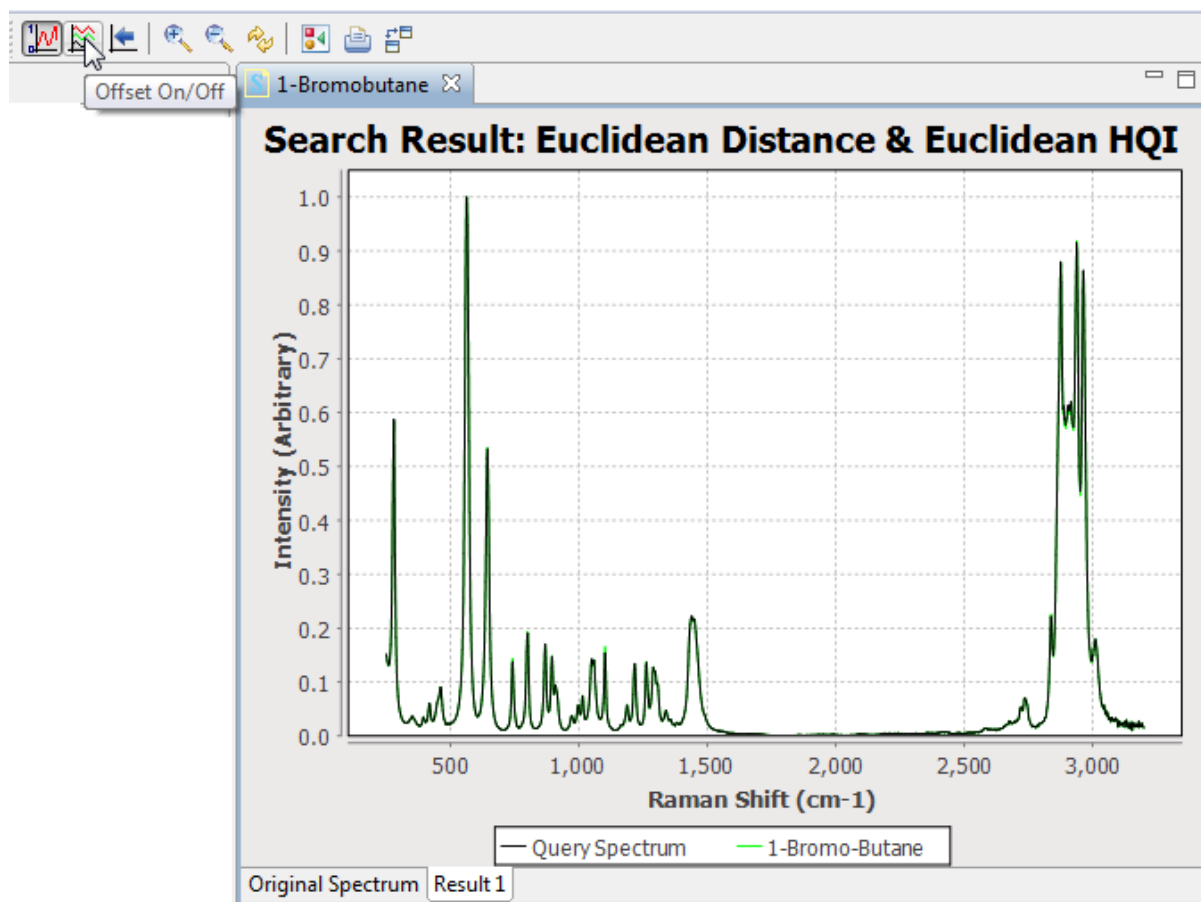Note that the Graphical Viewing pane must be active for the *Offset* button to be selected.

Figure 26: Turning Offset On

Once the *Offset* button is turned to On, the spectra that have been selected will appear in the Spectrum Viewing pane offset from the query spectrum as shown in Figure 27.

The search results are identified and ranked in order of closest matches for the unknown spectra. They are also color coded, the closest match plotted in *Green* as shown in Figure 27 when the distance is low or HQI is close to 100, through to *Blue* (when HQI is around 50) and finally to *Red* for very bad matches (when HQI is close to 0).
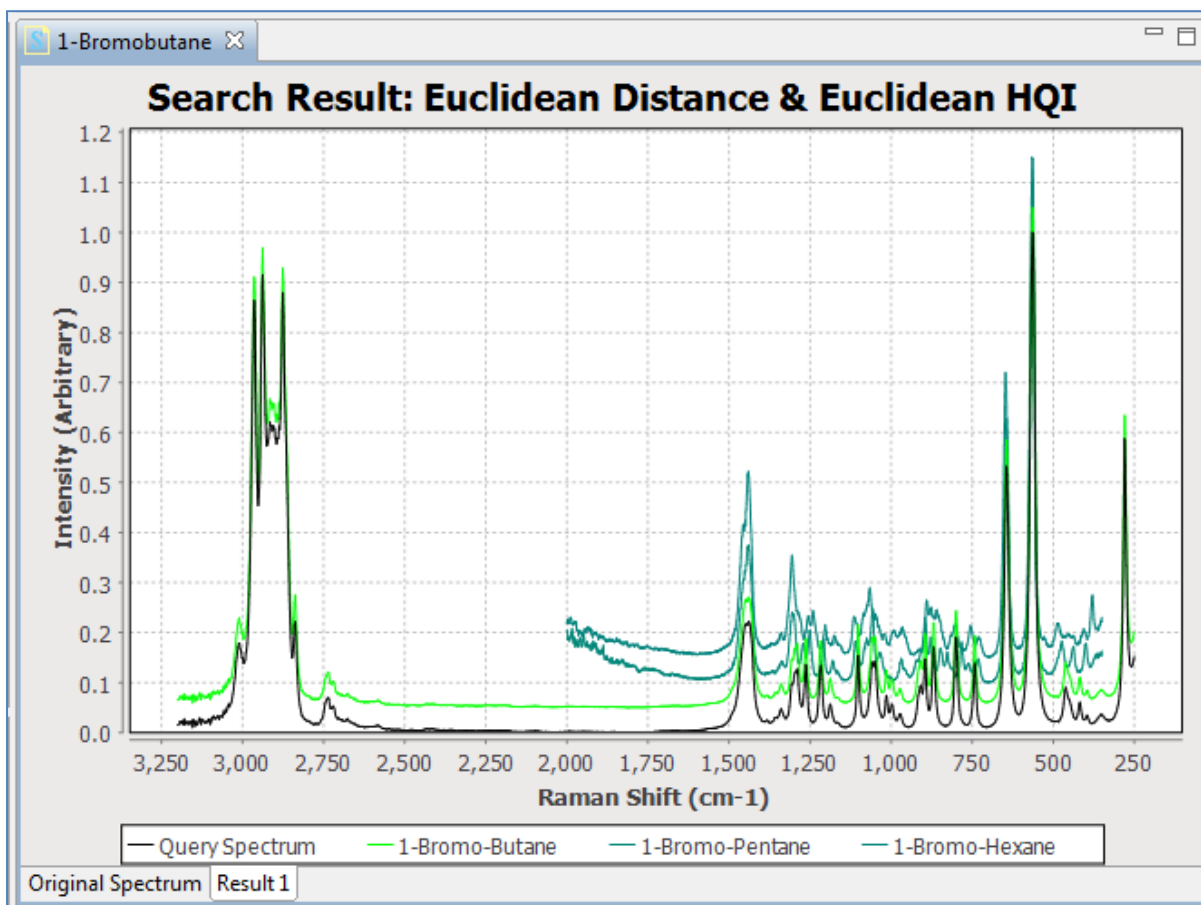
Figure 27: Selected results plotted with query spectrum, when Offset is turned on

6 Viewing Detailed Information about a Search Result
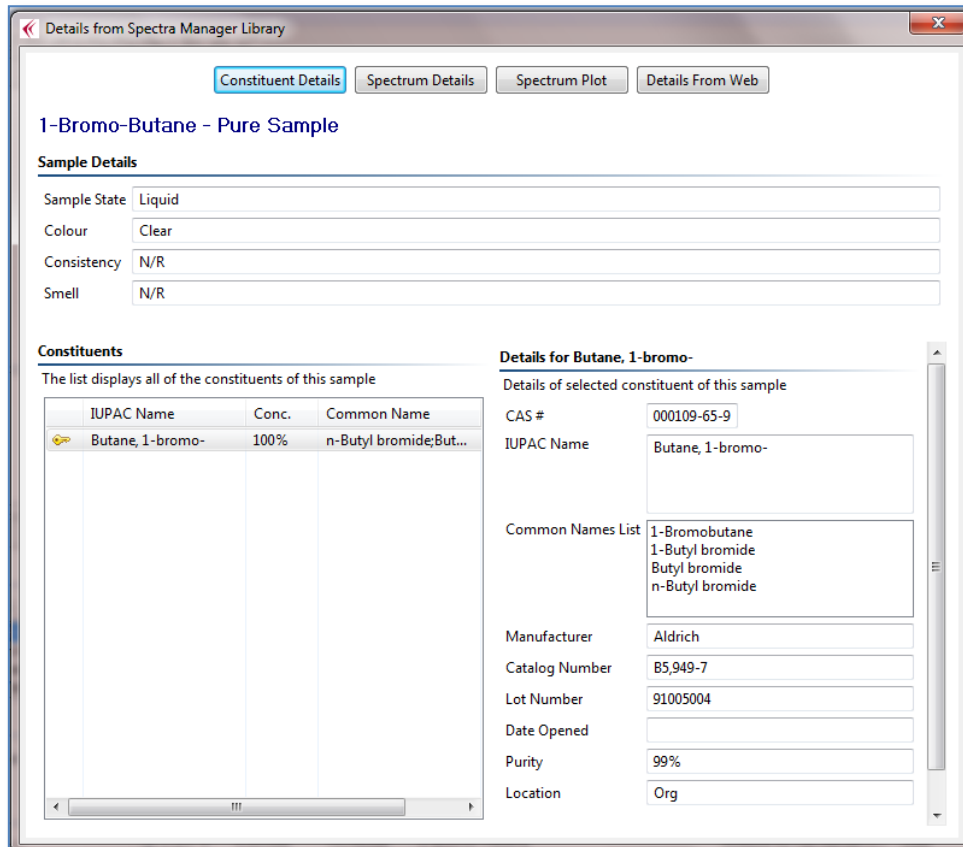
After performing a search, you can double-click on any of the results in the results pane in order to view additional details:

1. Details about the constituent (the physical sample from which the spectrum was collected)
2. Details about the spectrum (how it was collected and by whom, etc.)
3. A plot of the spectrum
4. Information about the substance from Wikipedia, if available.

These details are the same as those that can be seen in **Analyze IQ Spectra Manager**, except that in **Spectra Manager**, many details can be edited, whereas in **SmartSearch** they can only be viewed.

6.1 Constituent Details

When you double-click on a result, the *Constituent Details* are displayed by default, as shown in Figure 27. This figure shows the constituent details for a 1-Bromobutane. The *Sample Details* sections lists information about the sample as a whole: the sample state (liquid, gas or solid), colour, consistency and smell. The notation "N/R" entered for consistency and smell in indicates "Not Recorded".

Figure 28: Constituent Details

The *Constituents* section lists each constituent in a mixture and the *Details for…* section on the right shows the details for the currently-selected constituent, as shown in Figure 30. The following details for a constituent are displayed:

- CAS Number
- IUPAC Name
- Common Names List
- Manufacturer of the constituent material
- Catalog number
- Lot Number
- Date Opened
- Purity
- Location

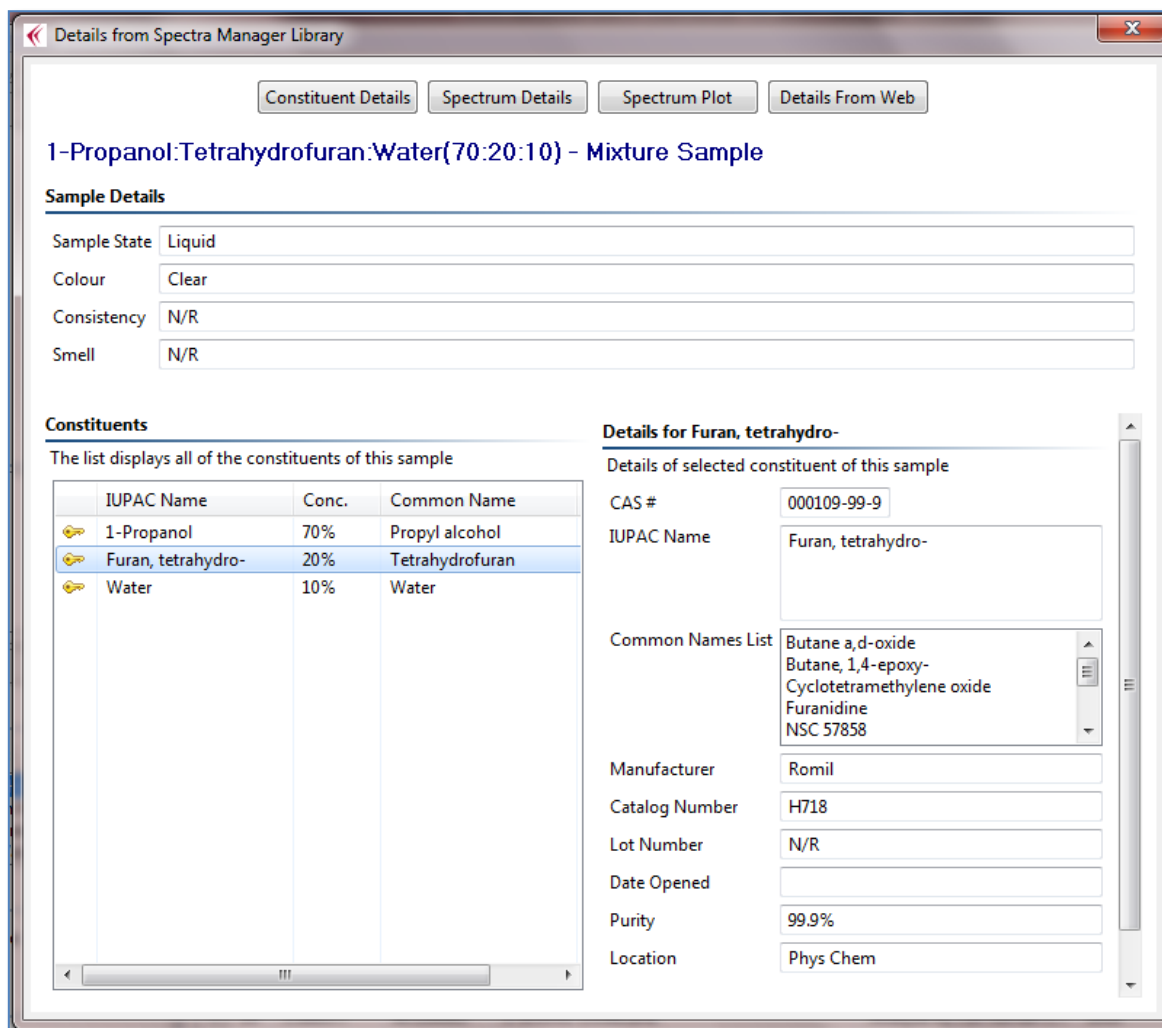**Figure 29: View Constituent Details**

Figure 30: Selecting Constituent in Constituent Details

To view the details of a different constituent in a mixture, select that constituent in the *Constituents* section, as shown in Figure 30, where the Furan, tetrahydro- constituent has been selected.

6.2   *Spectrum Details*

Press the *Spectrum Details* button to view this information. Figure 31 shows the details of the spectrum of a mixture.
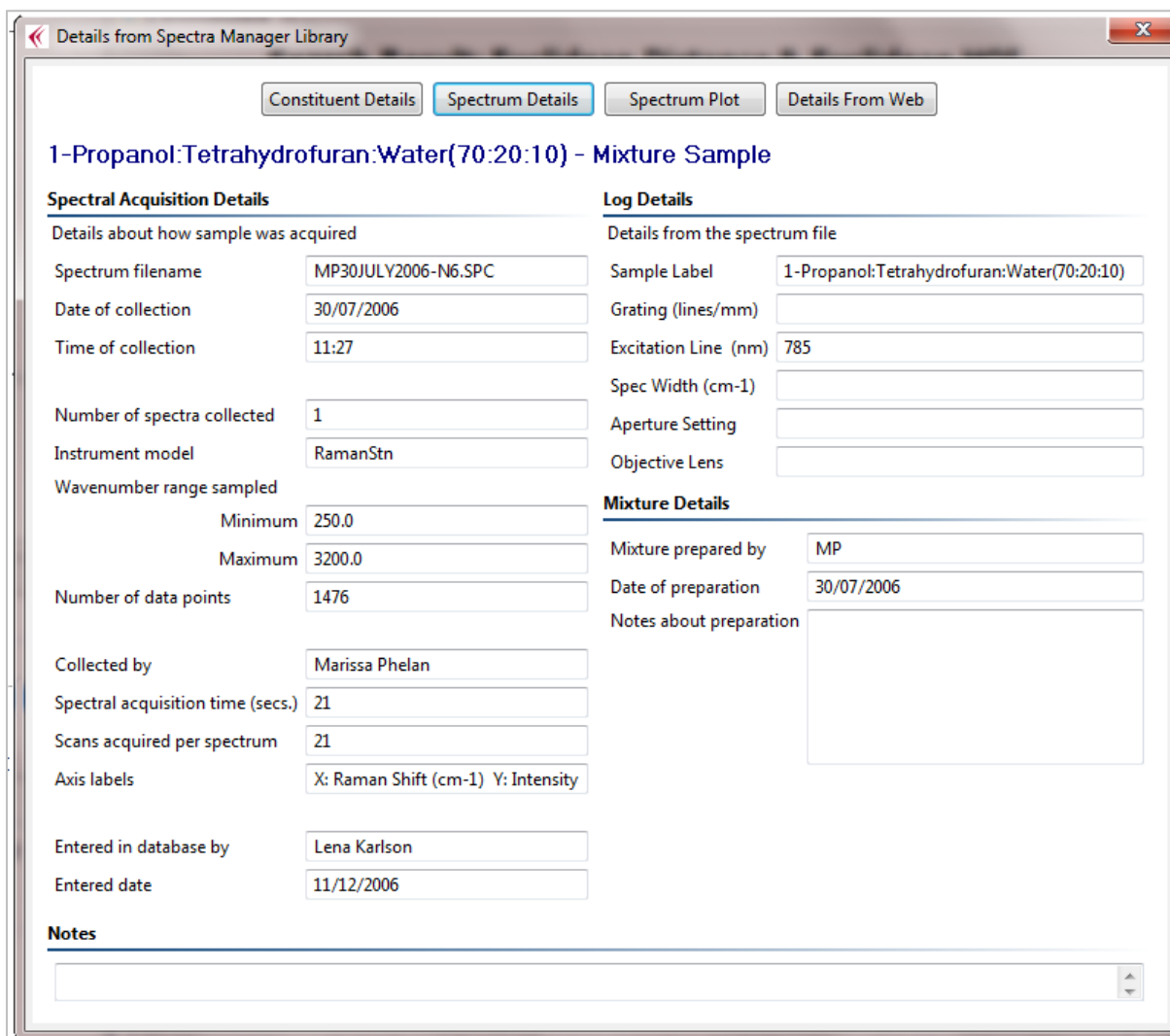
Figure 31: View Spectrum Details

The Spectrum Details view has four main sections, as listed below.

1. **Spectral Acquisition Details**

   This includes data that was automatically entered into the database from the Spectrum file (this cannot be edited) and information recorded during the entry of the spectral sample into **Spectra Manager**:

   - *Spectrum filename:* The spectrum file name from which the spectrum was imported.
   - *Date of collection:* The date the spectrum was recorded.
   - *Time of collection:* The time the spectrum was recorded.
   - *Number of spectra collected:* The number of spectra collected.
   - *Instrument:* Name assigned to the instrument used to record the spectrum.

- *Wavenumber range sampled:* The minimum and maximum wavenumber over which the spectrum was recorded.
- *Number of data points:* Number of data points in the spectrum.
- *Collected by:* Name of person who recorded the spectrum.
- *Spectral acquisition time (secs):* Duration of spectral acquisition in seconds.
- *Scans acquired per spectrum:* Number of scans acquired for this spectrum.
- *Axis Labels:* X and Y axis labels.
- *Entered in database by:* Name of person who entered this spectrum into **Spectra Manager**.
- *Entered date:* Date of entry of this spectrum into **Spectra Manager**.

2. Log Details

This data is extracted from the *Log* section of the spectrum file from which the spectrum was imported:

- *Sample Label:* Name given to this sample.
- *Grating:* For dispersive spectrometers, the choice of diffraction
- Grating used during analysis, expressed in lines/millimetre, e.g. 50 lines/mm and 1800 lines/mm.
- *Excitation Line:* The wavelength, in nanometres, of the spectrometer, e.g. 785.
- *Spec. Width:* For dispersive spectrometers, the width of the slit that the dispersed scattered signal is passed through. This determines the frequency resolution of the spectrum. For example, a spectrum recorded at an interval of 350–2000 cm^{-1} (1650 channels) with a confocal aperture setting of 200 micrometres, using a 950 lines per mm grating, gives a spectral resolution of ~1 cm^{-1}.
- *Aperture Setting:* The confocal aperture setting, expressed in micrometres.
- *Objective Lens:* The choice of objective lens used to magnify and illuminate the sample with laser light, e.g. 10x, 40x, 50x or 100x.

3. Mixture Details

This section only applies to the spectra of mixtures:

- *Mixture prepared by:* Name of person who prepared this mixture.
- *Date of preparation:* Date on which the sample was prepared.
- *Notes about preparation:* Additional notes about the mixture preparation.

4. Notes

This section is used to record additional information that may be of benefit to an analyst. An example entry is "The material had a yellow colour that might indicate decomposition of the material."

6.3   Spectrum Plot

To view a spectrum plot, select the spectrum and click on the *Spectrum Plot* button, as shown in Figure 32.

This graph area works in the same way as the main **SmartSearch** graph area, which was described previously in Section 5.4. As shown in below in Figure 33, a set of buttons are provided for reversing the X axis, zooming in and out, resetting the zoom, saving the image, and printing the image.
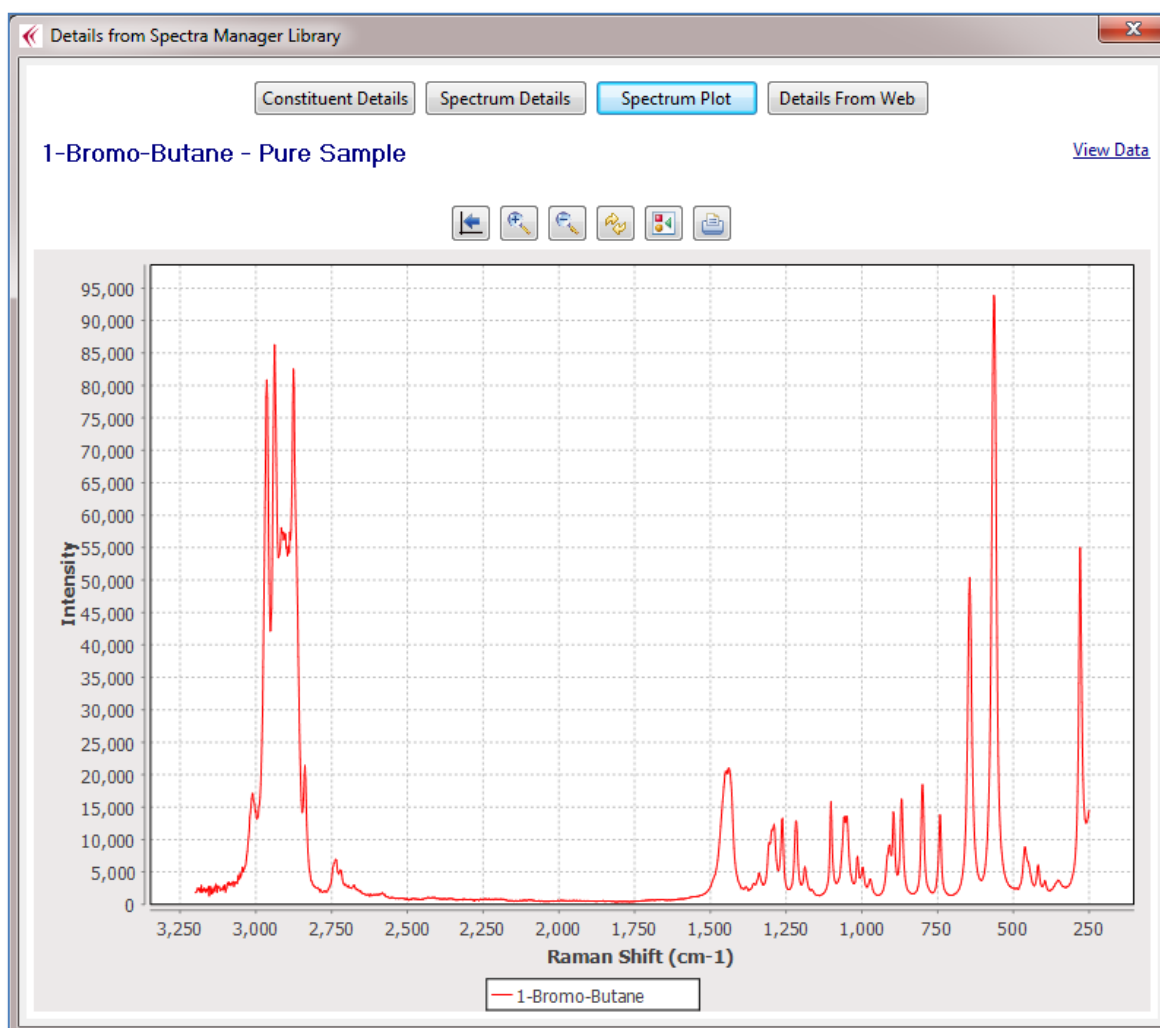

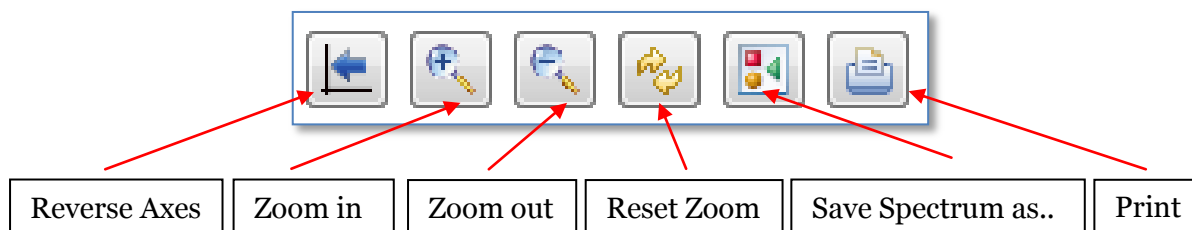**Figure 32: View Spectrum Plot**

Figure 33: View Spectrum Plot

in the same manner that was described in Section 5.4, you can zoom in on a specific area (e.g. a peak) of the plot by clicking on the plot area and drawing a rectangle around the section of interest. For example, selecting the area shown in Figure 35 results in the plot displayed in Figure 36 . Click on the Reset button indicated in to bring the plot back to its original settings.
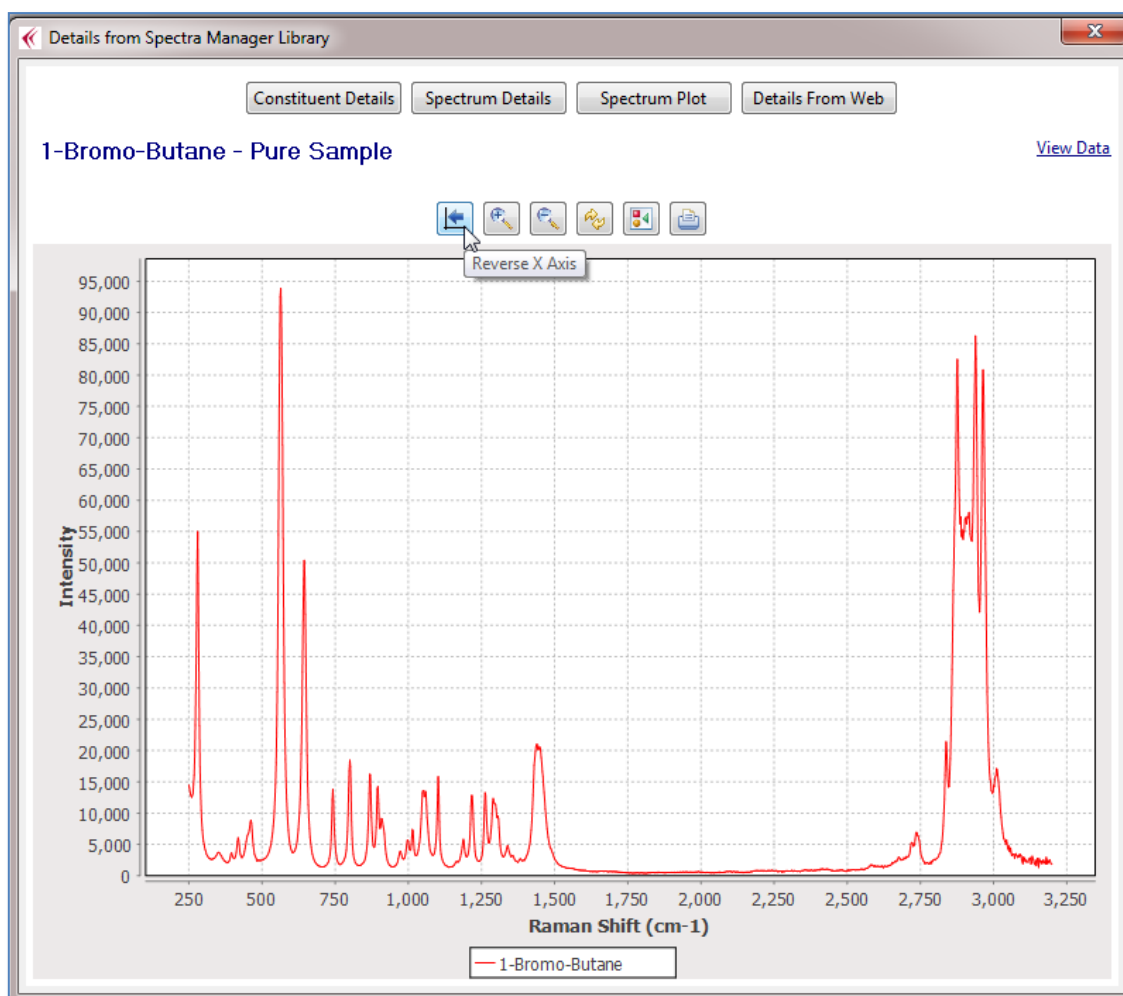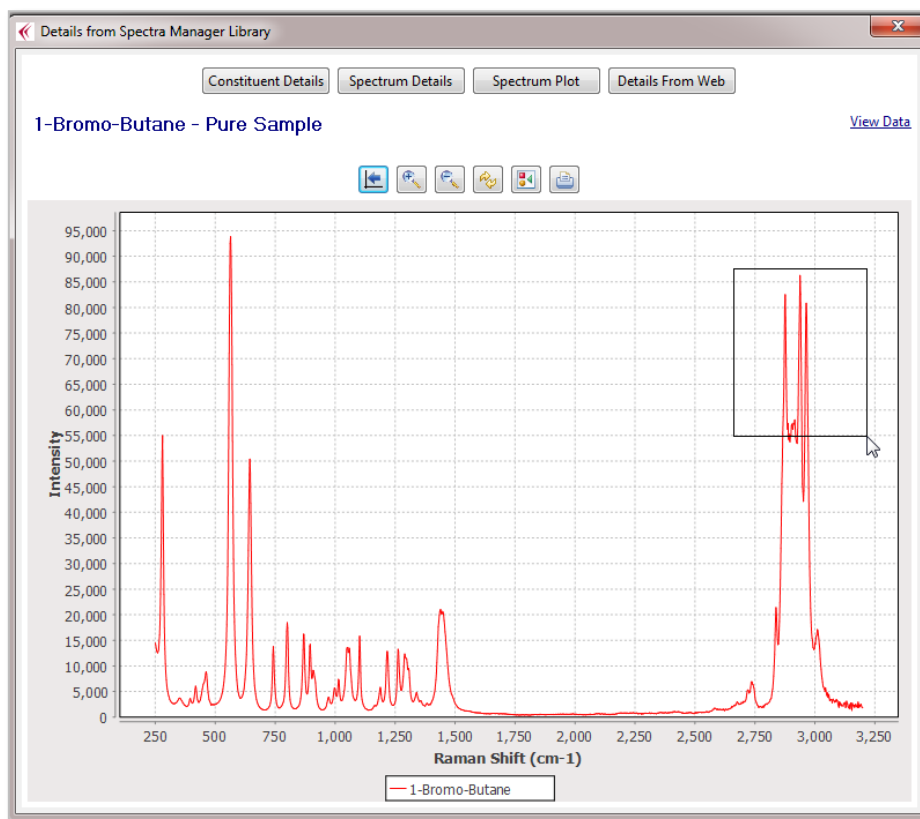

Figure 34 : Reverse Axes

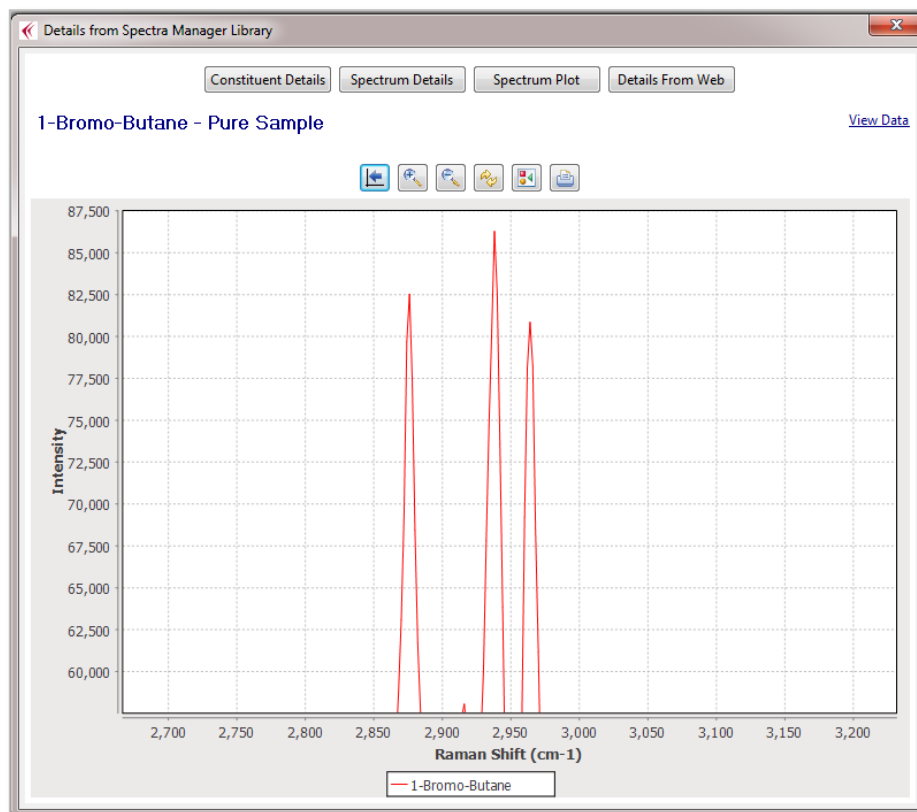**Figure 35: Zooming in on Spectrum Plot**

Figure 36: Zooming in on Spectrum Plot

6.4   Details from Web

A unique feature of **Analyze IQ SmartSearch** is that it can retrieve linked data from Wikipedia associated with your results and display them along with the results, as shown in Figure 37.

Figure 37: View details from Web

In the case where you are viewing details of a mixture, each of its constituents is listed in the box in the upper right of the view. Click on a constituent to see its details.

If you wish to view or change the Wikipedia source page for this information, you can use the three buttons in the upper left:

1.   Go To Source
2.   Refresh From Source
3.   Change Source

Their operation is described next.

6.4.1 Go To Source

When you select the *Go to Source* button as shown in Figure 38, your computer's default browser will open, displaying the Wikipedia web page for your selected spectrum, as shown in Figure 39.
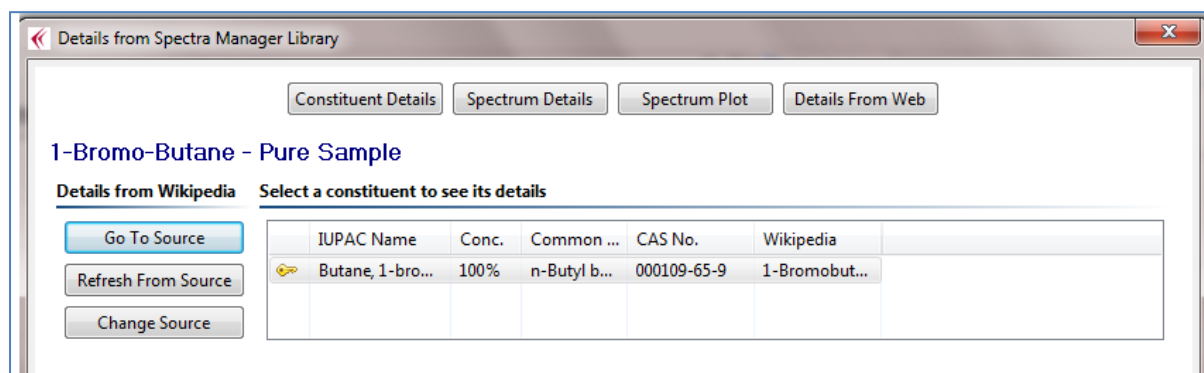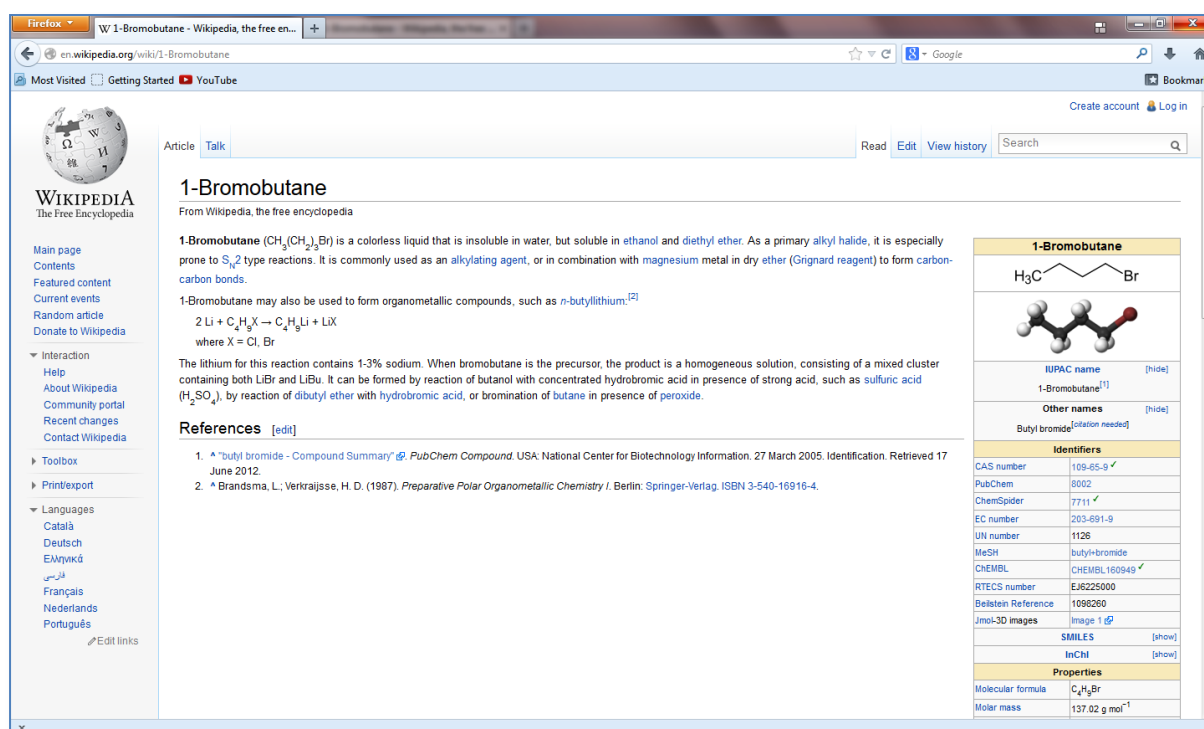

Figure 38: View Source

Figure 39: View Source

6.4.2 Refresh from Source

When this option is selected as shown in Figure 40, the data obtained from Wikipedia is automatically refreshed and updated. This can be useful if the Wikipedia page has changed very recently, to force **SmartSearch**'s internal cache to be updated to reflect the changes.

Figure 40: Refresh from Source.

6.4.3 Change Source

If there is no Wikipedia source page stored for a spectrum, or if you believe that the source page is incorrect for the spectrum selected, you can change the source of the information to the correct Wikipedia page. Once you press the *Change Source* button, you can enter the correct details as shown in Figure 41.

When entering a new source, you can enter its full URL or just the Wikipedia page name. So for example, you could enter either of the following for this sample:

- http://en.wikipedia.org/wiki/1-Bromobutane
- 1-Bromobutane

Note that **SmartSearch** cross-checks some it its internally stored information with the information on the new Wikipedia page that you enter, and will only accept the newly entered page if the checks are successful.

Figure 41: Changing the Wikipedia page used as a source of data

7   Working with Libraries

In **SmartSearch** and the other products in the **Analyze IQ** software suite, spectra and associated data are organized into libraries. Typically, a library of spectra is from one supplier, contains spectra of one type (e.g. Raman or FTIR), and may correspond to a specific category of substances (e.g. organics or minerals).

This section describes how to select a library for performing a search, how to enter an activation code after purchasing a license for a new library, how to install a purchased library, how to create a user library, and how to change connection settings on your computer.

7.1   Types of Library

Libraries are of three main types:

1. Analyze IQ libraries: supplied by Analyze IQ Limited
2. Third-party libraries: commercial libraries provided by other suppliers via Analyze IQ; contact us for details of supported libraries
3. User libraries: containing your own collections of spectra that you have built up and that may be proprietary to your organization.

Please contact us (see Section 9) for details of the specific commercial libraries that are available for you to purchase, or if you have difficulties installing newly purchased libraries.

7.2   Selecting a Library

Use the *Libraries – Select* menu command to open the dialog shown in Figure 42.
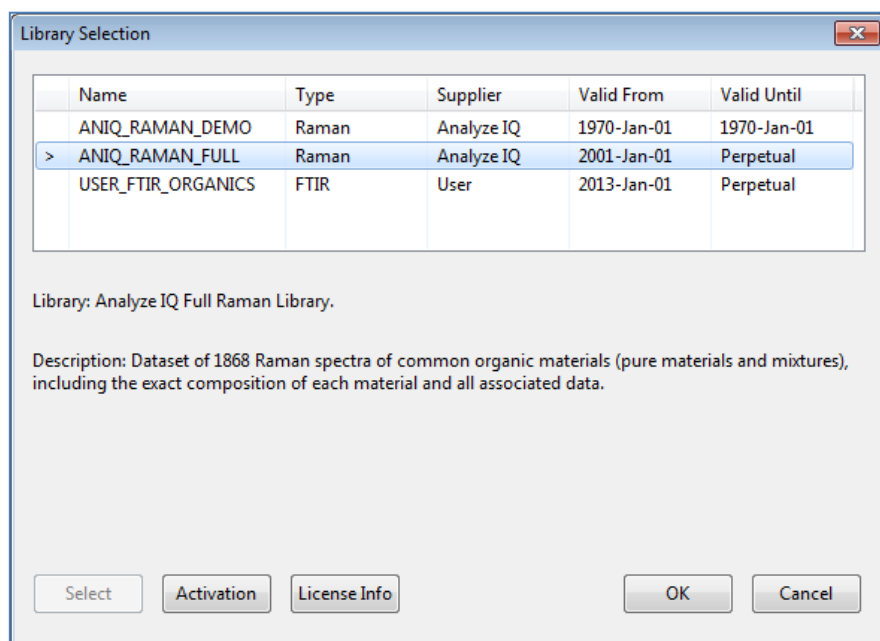
Figure 42: The Library Selection Dialog

The upper half of this dialog contains a list of available libraries, and the lower half contains more extensive information about the currently chosen library from the list.

If the library needs to be activated, or if the activation has expired, or if there are any other such problems, information is provided in the lower half of the window.

To select a library for subsequent searches, choose it in the list and then press the *Select* button, or double-click on it in the list.

7.3   License Information and Activation

Activating a library involves entering an activation code; Analyze IQ Limited will provide you with short-term activation codes to evaluate a library and long-term activation codes when you purchase a library.

Activation codes are provided via email and typically look like the following:

ANIQ_RAMAN_FULL|2001-Jan-01|2031-Jan-01|199|MCwFCDhp+FrB8KsDxEzjKCEZm64NrPoTAhQQQ8a2lpuqrkjTw+E9mPJ/m8LZ/w==

To activate a library, press the *Activate* button and enter the code when requested. Note that you may first be prompted to review and accept licensing terms.
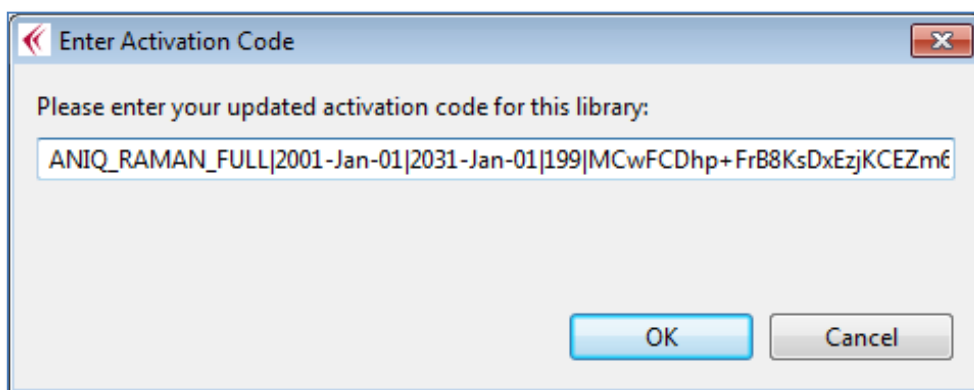
Figure 43: Entering a library activation code

To review the license terms associated with a library, press the *License Info* button on the Library Selection dialog.

7.4   Installing Purchased Libraries and Creating User Libraries

New libraries, including those produced by third party suppliers, may be purchased via Analyze IQ; contact us for information.

When you purchase a library, you will be provided with installation instructions for it. In general, the procedure is to use the *Library – Install* menu command, select

the library file that you were provided with after your purchase, and follow the instructions on screen.

You can use **Analyze IQ Spectra Manager** to create your own user libraries, add new spectra to them, edit the information associated with spectra, and delete entries. Whenever you create a library in **Spectra Manager**, it becomes available automatically in **SmartSearch**. Likewise, changes that you make to libraries in **Spectra Manager** are automatically seen in **SmartSearch**.

8 Search Algorithms

This section describes the Distance metrics and Hit Quality Index (HQI) metrics that are available in **Analyze IQ SmartSearch**. It begins discussing definitions and principles, and then how to interpret search results, before moving on to describe the algorithms available.

8.1 Definitions and Principles

As noted earlier, the purpose of a spectral search algorithm is: given an unknown spectrum (referred to as the *query*) and a set of known spectra (referred to as the *library*), produce a ranked list of the best matches.

There are different ways of measuring how good a match is; these ways of measuring are called *metrics*. Some metrics will emphasize some differences more than others, and down-play others; which metric gives the best result depends entirely on your application, including exactly what kind of spectroscopy you are working with, what kinds of substances are in the library, and whether different library entries are quite different from each other or relatively similar.

Formally, a **metric** is a function that defines some measure of the similarity/difference between two spectra, where each spectrum is represented as a fixed-length vector with one element in the vector for each point on the spectrum. Similarity and difference are reciprocal concepts: high similarity corresponds to low distance, and low similarity corresponds to high distance.

As was noted in Section 5.7.1, one of the distinctive features of **SmartSearch** is that when you perform a search, you can choose both a distance metric and a HQI metric:

- **A distance metrics** is expressed as a number in the range 0 to infinity, with 0 being the distance between two identical spectra and the distance increasing as the spectra become less similar.
- **A HQI metric** is expressed as a number between 100 and 0, measuring how strong the "hit" is when comparing two spectra. HQI of 100 is the highest quality hit, when the two spectra are identical, and the HQI reduces as spectra become less similar.

SmartSearch includes a number of strategies to ensure that it gives reliable results when comparing queries and libraries that come from different instruments:

- Only the overlapping region between two spectra are compared, even though the full spectra are displayed on screen
- All spectra are internally normalized and re-sampled when being compared, to ensure comparability in searches
- Results are scaled to ensure that they are not distorted by the number of overlapping points between spectra.

8.2   Understanding Distance and HQI Values

Figure 44 shows the result of a search where the query is one of the sample spectra provided with **SmartSearch** (Acetic_Acid.aiq) and the library is the Analyze IQ Raman Library.

In this case, the top result is clearly a perfect match, as the distance is 0 and HQI is 100. However, this is because the exact same spectrum that we used as the query happens to be in the library. If the spectra are collected separately, matches will not be so good.

As the figure shows, the several next-best matches are also acetic acid, though mixed with increasing amounts of water, as the Analyze IQ Raman Library includes mixtures as well as pure substances.

Note that in general, it is not advisable to rely on searching software to perform detailed mixture analysis; chemometric software such as **Analyze IQ Lab** is much more suitable for that task.

Search Results

Searched 1868 spectra in 0.175 seconds; 0.094 ms/spectrum. Double-click on a result for its details.

Rank	Distance	HQI	Spectrum Name	Library	ID
☐ 1	0	100	Acetic acid	Analyze IQ Full Raman Library	318
☐ 2	0.314687	96.721321	Acetic Acid:Water (99:1)	Analyze IQ Full Raman Library	1238
☐ 3	0.456915	93.787621	Acetic Acid:Water (95:5)	Analyze IQ Full Raman Library	1240
☐ 4	0.766435	90.803848	Acetic Acid:Water (90:10)	Analyze IQ Full Raman Library	1242
☐ 5	1.18931	87.328217	Acetic Acid:Water (80:20)	Analyze IQ Full Raman Library	1244
☐ 6	1.62318	84.535774	Acetic Acid:Water (70:30)	Analyze IQ Full Raman Library	1246
☐ 7	2.024679	82.453789	Acetic Acid:Water (60:40)	Analyze IQ Full Raman Library	1247
☐ 8	2.60586	80.111786	Acetic Acid:Water (50:50)	Analyze IQ Full Raman Library	1249

Figure 44: Result of a search where the query spectrum is Acetic Acid

Figure 45 shows the result of a search where the query was another sample spectrum provided with the software, 1-Bromobutane.aiq. Again, the top match is the correct one. Subsequent matches have significantly lower scores, and while they correspond to chemically similar materials, they are not the same.

Search Results

Searched 1868 spectra in 0.132 seconds; 0.070 ms/spectrum. Double-click on a result for its details.

Rank	Distance	HQI	Spectrum Name	Library	ID
☐ 1	0	100	1-Bromo-Butane	Analyze IQ Full Raman L...	41
☐ 2	1.48453	76.46254	1-Bromo-Pentane	Analyze IQ Full Raman L...	2608
☐ 3	1.557098	75.741188	1-Bromo-Hexane	Analyze IQ Full Raman L...	2599
☐ 4	2.427203	76.64418	1-Bromo Hexane	Analyze IQ Full Raman L...	42
☐ 5	2.873013	72.270569	1-Bromo-Heptane	Analyze IQ Full Raman L...	303
☐ 6	3.129571	67.455986	1-bromo-5-chloro-Pentane	Analyze IQ Full Raman L...	486
☐ 7	3.180998	68.399551	1-Bromo-Octane	Analyze IQ Full Raman L...	359
☐ 8	3.342426	63.121784	1-bromo-Decane	Analyze IQ Full Raman L...	387

Figure 45: Result of a search where the query spectrum is 1-Bromobutane

Here are some general guidelines for applying search algorithms and interpreting the results:

- Bear in mind that different metrics provide different perspectives on how similar spectra are, so when working with a new dataset, experiment by searching for known spectra to get a feel for which distance metric and HQI metric work best for it.
- In evaluating results, examine both HQI and distance results, and look for good agreement between them.
- Search metric values are relative, not absolute; it is fine to compare results of searching for a specific query in a specific library (which is why we present them in a table), but beware of comparing results of searching in two different libraries or using two different search algorithms.
- There is no hard 'cut-off' point above which you can say that you have a match, and below which you can say that you don't have a match. Rather, it is entirely dependent on the data and the metrics.
- A good approach is to scan down through the list of results and look for a natural breakpoint or gap, such as the one between Result 1 and Result 2 in Figure 45.

8.3   List of Search Algorithms

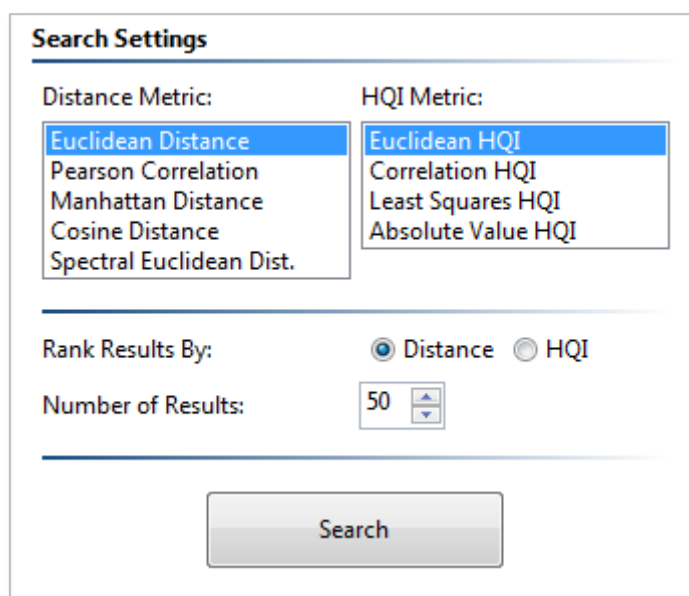The Search Settings pane (see below) presents a list of available algorithms, categorized as distance metrics and HQI metrics.

The distance metrics are:

1. Euclidean Distance
2. Pearson Correlation
3. Manhattan Distance
4. Cosine Distance
5. Spectral Euclidean Distance

The HQI metrics are:

1. Euclidean HQI
2. Correlation HQI
3. Least Squares HQI
4. Absolute Value HQI

Figure 46: Search Settings pane

8.4   Distance Metrics

Before describing distance metrics, we will define some notation. This is based in part on work published by Khan and Madden [1]. Note that even though that publication refers to Raman spectroscopy, the metrics are equally applicable to all forms of spectroscopy.

An unknown query spectrum is defined as a vector $X = [x_1, x_2, ..., x_N]$ where x_i denotes the response at the ith position on the spectrum (which may be typically a wavenumber or wavelength). This spectrum corresponds to a pure substance or a mixture of various substances in different proportions.

Similarly, a known spectrum from the spectral library is defined as a vector $Z = [z_1, z_2, ..., z_N]$, where z_i denotes the response at the ith position on the spectrum.

N is the total number of data points that are common to the two spectra. $\bar{X}$ and $\bar{Z}$ denote sample means and S_X and S_Z denote sample standard deviations.

The distance metric formulas are presented below, with additional remarks in the subsections following.

Distance Metric	Mathematical Formula
Euclidean Distance	$\sqrt{\sum_{i=1}^{N}(x_i - z_i)^2}$
Pearson Correlation Coefficient	$\dfrac{1}{n-1}\sum\left(\dfrac{x_i - \bar{X}}{S_X}\right)\left(\dfrac{x_i - \bar{Z}}{S_Z}\right)$
Manhattan Distance (Citiblock)	$\sum_{i=1}^{N}\left|x_i - z_i\right|$
Cosine Distance	$\dfrac{\sum_{i=1}^{N} x_i \times z_i}{\sqrt{\sum_{i=1}^{N}(x_i)^2} \times \sqrt{\sum_{i=1}^{N}(z_i)^2}}$
Euclidean Distance	$\sqrt{\sum_{i=1}^{N} w_i (x_i - z_i)^2}$ See main text for calculation of w_i

Figure 47: Distance metrics used in Analyze IQ SmartSearch

8.4.1 Euclidean Distance

This is the most widely used distance metric: if two spectra are represented as high-dimensional vectors, this corresponds to the straight-line distance between them.

8.4.2 Pearson Correlation

Pearson's correlation coefficient computes the covariance of between two spectra. It is defined as the covariance of the two spectra divided by the product of their standard deviations.

8.4.3 Manhattan Distance

This metric, also known as the citiblock distance metric, is strongly related to Euclidean distance, as can be seen by comparing their formulas in Figure 47. It measures distance using axis-aligned directions rather than the straight-line distance, by computing the sum of absolute differences.

8.4.4 Cosine Distance

Cosine distance, like other metrics such as Euclidean distance, considers the spectra to be vectors in high-dimensional space. However, rather than being based on the straight-line distance between the end-points of the vectors, it is based on the angle between the two vectors. Specifically, by using a formula based on the equation for a dot product of two vectors, it computes the cosine of the angle between them. This is particularly useful if we want to perform a comparison that is independent of the magnitude of the spectra, though it should be noted that **SmartSearch** normalizes spectra prior to calculating distances in any case.

8.4.5 Spectral Euclidean Distance

This is a unique new distance metric, only available in **Analyze IQ SmartSearch**, that was first proposed by Khan and Madden [1]. It is based on Euclidean distance, but modified to be better suited to the specific task of comparing spectra.

While calculating the squared distance between each individual points (and considering a baseline, b, below which peaks are ignored), it considers several cases and computes a weight factor w_i at that point:

1. $x_i = z_i$: distance is 0. Perfect match at this point: weight has no effect.
2. $x_i < z_i$ and $x_i > b$: significant peak in x_i and in z_i, but z_i one is higher, perhaps because x_i is diluted by the presence of another component. Such differences are less significant than typical, so downplay distance by making w_i small.
3. $x_i < b$ and $x_i > b$: significant peak in Z is missing in X: this is a significant difference that should be penalized, so make w_i large.

4. $x_i > z_i$ and $z_i > b$: peak present in both, but more strong in the query. This is not strong evidence either for or against the query being a component of the lib, so use neutral value for w_i.

5. $x_i > b$ and $x_i < b$: a significant peak in X is missing in Z. This may because the query spectrum contains components from multiple pure materials, so this is not evidence for or against a match, so use neutral value for w_i.

Having chosen a value for w_i, it is used in the formula shown previously in Figure 47.

Incidentally, the Spectral Euclidean Distance is not a metric in the according to the formal mathematical definition, as it is not symmetric: $D_{SpectralEuclid}(Z, X) \neq D_{SpectralEuclid}(X, Z)$. However, its lack of symmetry does not create any problems for its use in practice. In fact, it is a deliberate result of it being designed with the assumption that the query spectrum may include components not found in the library spectrum.

8.5   HQI Metrics

Hit Quality Index metrics are related to distance metrics, but reformulated to yield values in the range 100 to 0, where 100 is a perfect match. Therefore, a higher HQI value indicates a better match, whereas a lower distance value indicates a better match.

As in Section 0, an unknown query spectrum is defined as a vector $X = [x_1, x_2, ..., x_N]$ where x_i denotes the response at the ith position on the, and a known spectrum from the spectral library is similarly defined as a vector $Z = [z_1, z_2, ..., z_N]$.

In HQI calculations, we make use of vector dot products, denoted $A \bullet B$.

The HQI metric formulas are presented below, with additional remarks in the subsections following. Note that the formulas omit scale factors that are used to scale HQI values in the range 100-0.

HQI Metric	Mathematical Formula
Euclidean HQI	$\sqrt{1 - \dfrac{Z \bullet X}{\sqrt{Z \bullet Z}\sqrt{X \bullet X}}}$
Correlation HQI	$1 - \dfrac{(Z_m \bullet X_m)^2}{(Z_m \bullet Z_m)(X_m \bullet X_m)}$
Least Squares HQI	$\dfrac{\sum_{i=1}^{N}(x_i - z_i)^2}{N}$
Absolute Value HQI	$\dfrac{\sum_{i=1}^{N}|x_i - z_i|}{N}$

Figure 48: HQI metrics used in Analyze IQ SmartSearch

8.5.1 Euclidean HQI

The Euclidean HQI metric is widely used and is slightly faster than related metrics such as Correlation HQI because it does not have to perform mean-centering of the data, so that

8.5.2 Correlation HQI

Correlation HQI is strongly related to Euclidean HQI, as can be seen by comparing their formulas in Figure 48. The main difference is that it mean-centers the spectra before the computation:

$$Z_m = Z - \bar{Z}, \; X_m = X - \bar{X}$$

8.5.3 Least Squares HQI

This metric works like some of the distance metrics, performing a point-by-point comparison. It computes the mean squared difference between spectra. It places more emphasis on large peaks than some other metrics. However, values tend not to be as well separated as when using some of the other metrics; even relatively poor matches may get a high-looking score with this metric, but of course, as noted in Section 8.2, distance and HQI values are relative, not absolute.

8.5.4 Absolute Value HQI

The Absolute Value HQI is strongly related to the Least Squares HQI, as can be seen by comparing their formulas. The key difference is that it computes the absolute difference between points rather than the squared difference. Because of this, it puts somewhat less emphasis on differences in large peaks, but scores tend to be slightly better separated than those using Least Squares HQI.

8.6 Using Derivatives with Search Algorithms

Prior to performing a search in **SmartSearch**, you can apply any of the preprocessing methods supported in the Analyze IQ software suite, as shown in the figure below. When you apply any preprocessing method in SmartSearch, the operation is applied to the spectrum that is open *and* the currently selected library.
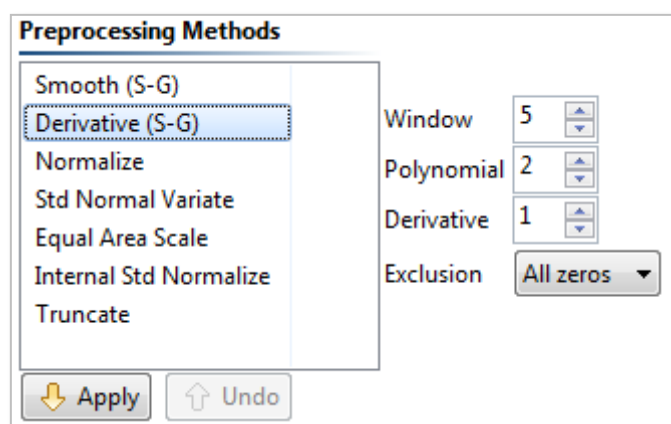

Figure 49: Preprocessing Options

Of these preprocessing methods, the most commonly used in SmartSearch is the Derivative, usually with the setting Derivative=1, to apply a first-order derivative using the Savitzky-Golay algorithm [2]. Its benefit is that it removes the effect of slow baseline drifts, and so is useful for spectra without a clear baseline. You can apply this derivative and then choose a search algorithm in order to achieve the same effect (more efficiently and with a more robust derivative algorithm) that can be achieved in other search software products that include algorithms with a first derivative, e.g. "First Derivative Absolute Value HQI".

9 Technical Support, Sales and Services

For assistance and technical support queries relating to **Analyze IQ SmartSearch**, please go to the **User Area** of the Analyze IQ website, http://www.AnalyzeIQ.com, or send an email message to support@AnalyzeIQ.com.

For information on extending an evaluation license or purchasing a full license for **Analyze IQ SmartSearch**, please contact our Sales Department by sending an email message to sales@AnalyzeIQ.com.

In addition to software sales and technical support, Analyze IQ Limited provides the following services for customers:

- Supply of third party spectra libraries
- Software training
- Data validation and analysis
- Chemometric model development
- Custom software development, including custom analysis solutions
- Integration of Analyze IQ software in third-party software
- Development of Instrument Interfaces
- OEM licensing

For information on these services, please contact our Sales Department by sending an email message to sales@AnalyzeIQ.com.

10   References

1. S.S. Khan and M.G. Madden. "New Similarity Metrics for Raman Spectroscopy", **Chemometrics and Intelligent Laboratory Systems**, Volume 114, 2012. DOI:10.1016/j.chemolab.2012.03.007.

2. Savitzky, A. & Golay, M.J.E. (1964). "Smoothing and differentiation of data by simplified least squares procedures." **Analytical Chemistry**, 36, 1627–1639.

11   Patent, Copyright & Trademark Notices