Raman Library Matching
Using the ENLIGHTEN™ Plugin

Raman spectroscopy combines high specificity with tremendous ease of use to identify unknown compounds. The principle of operation is simple: point and shoot a laser at your sample, collect the light scattered back with a spectrometer, and you get a Raman spectrum which is as detailed and unique to the sample as a fingerprint – perfect for identification. And since no sample preparation is required, it is no wonder that Raman spectroscopy is becoming the go-to tool for instantaneous identification of unknowns. But how does that 'identification' actually occur? In this tech note, we'll introduce both the theory and best practices behind the most common method, 'library matching', and explain how a basic match can be performed easily using any of our Raman spectrometers in our ENLIGHTEN™ spectroscopy software. If you want a simple way to get started with Raman library matching and build a custom library perfectly tailored to your application, this is the resource for you!

To perform accurate Raman library matching, you must first start with a good-quality Raman spectrum. This calls for a high sensitivity spectrometer with good signal-to-noise ratio to reduce errors in matching. Adequate spectral resolution is also important, as it affects peak clarity and match accuracy. A spectrometer resolution of 8-15 cm^{-1} is sufficient for many applications, depending on the complexity of the sample and the quantity and similarity of the compounds to be differentiated. Choosing the appropriate excitation laser wavelength will ensure minimal fluorescence interference, and optimizing the laser interface with the sample will maximize signal. Learn more about choosing a portable Raman spectrometer and laser excitation wavelength.

WHAT IS LIBRARY MATCHING?

In library matching, an unknown compound is identified by comparing its characteristic Raman spectrum to a set of known reference spectra in a spectral library. If the unknown Raman spectrum is found to be sufficiently similar to one of the known spectra in the library, the unknown spectrum is said to be 'matched'. The entire process is therefore called 'library matching'.

But what really happens under the hood? First, the two spectra must be interpolated so that they have the same set of x-axis values to allow direct numerical comparison, as reported in wavenumbers (cm^{-1}). Then matching can begin. One simple matching algorithm involves determining the Pearson correlation coefficient, r, between the unknown spectrum and each of the library spectra across all spectral data points. The library spectrum with the highest correlation coefficient is reported as the 'best match'.

Perfect positive correlation
r = 1
Unknown compound data
Library compound data
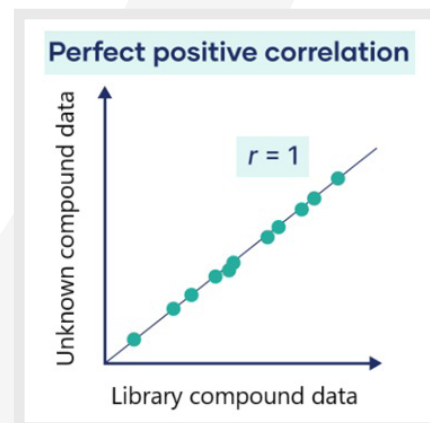
The matching score, which is often called the 'hit quality index' (HQI), is simply the correlation coefficient multiplied by 100. HQI scores above 95 are typically considered a confident match, while scores under 65 should be met with some skepticism. Depending on the application needs and consequences of a good or bad match, these lines might be redrawn.

HOW TO ACHIEVE GOOD QUALITY MATCHING

Library matching using the Pearson correlation coefficient is quick, yet still sensitive to small differences between two spectra. Since those spectral differences are most often due to chemical differences, this method has high chemical specificity, that is, a low rate of 'false positives' or incorrect identifications. This is very desirable in identification applications, where the consequences of an incorrect result can be far greater than those for an inconclusive one.

However, this sensitivity to small differences also applies to other factors which can interfere with accurate matching. These may include: 1) potential background fluorescence in the unknown sample, 2) significant differences in the spectral resolution or the spectral sensitivity between the test spectrometer and the spectrometer used to develop the library, and 3) any wavenumber offset between these two spectrometers.

Fortunately, many of these factors can be accounted for by following some best practices:

- Always perform a wavenumber correction against a known reference standard prior to library matching. This removes any x-axis offsets, which may occur due to temperature shifts or a small error in the assumed laser wavelength.

- If the sample shows evidence of background fluorescence, remove it mathematically with baseline correction in software or post-processing.

- For best matching accuracy, use the same or similar type of instrument to record both the library and the sample spectra – the closer the better.

- If a third-party library is being employed, use post-processing of the sample spectrum prior to matching to compensate for differences between the library and test spectrometer:

  - Use spectral deconvolution to account for resolution differences.

  - Apply an intensity calibration to account for instrument response function, or relative sensitivity differences.
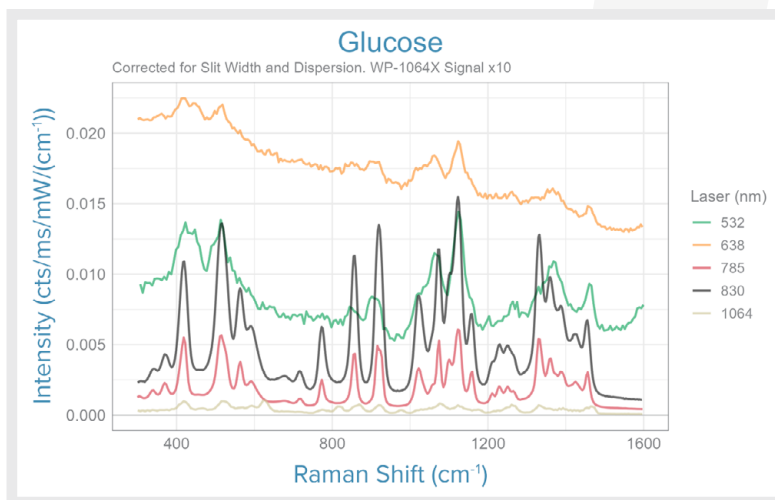
Background fluorescence can influence Raman spectral quality, and therefore matching accuracy. Choosing the optimal excitation wavelength for your sample can minimize background fluorescence.

MATCHING YOUR EXPECTATIONS TO YOUR LIBRARY

Finding the best library match using the Pearson correlation score assumes that the sample spectrum is pure or predominantly composed of a single compound, just like the library spectra. If the sample to be investigated is a mixture of two or more dominant compounds, each with their own significant contributions to the recorded spectrum, a mixtures analysis algorithm is needed instead, since no single 'pure' library spectrum will match the measured multicomponent sample spectrum.

Similarly, for many substances the spectrum of a compound in its pure state vs dissolved in water can also be quite different. This can be due to solvent effects, solvation and structural changes, changes in polarizability, and more. In those cases, a match of an aqueous solution sample spectrum against the library spectrum of the pure solid compound might be poor or fail entirely. It's therefore important to ensure that your library has the desired materials in the state in which they would be found as unknowns.

The ideal case is to use a library **you** have developed using the compounds in which you are interested, as well as any other compounds which may be present in the same environment. This helps you narrow the field to your specific application and improves the speed and accuracy of your results. If you require Raman library matching and have a want or need to develop your own library of unique materials, our ENLIGHTEN™ library matching plugin is for you!

ENLIGHTEN™ LIBRARY MATCHING PLUGIN

Wasatch Photonics offers a free, open-source spectroscopy operating software package called ENLIGHTEN™ for use with any of our Raman spectrometers or modules. [Learn more about ENLIGHTEN™ or download it now](#)

ENLIGHTEN™ is designed to be user-friendly and allow for quick setup and use of our spectrometers for Raman spectroscopy. Many functions are self-explanatory, with convenient 'mouse-over tooltips' to explain other buttons whose purpose may not be immediately clear. It even offers an 'Auto-Raman' feature, which offloads the optimization of settings to the software for both new and experienced users to allow push-button recording of high-quality Raman spectra. It also includes tools to apply the 'best practices' for Raman library matching described above: wavenumber correction, baseline correction, intensity correction, and spectral deconvolution.

ENLIGHTEN™ focuses on the data collection of spectra and 'outsources' the data analysis to a 'plugin' interface. This not only simplifies the user interface tremendously; it also allows every user to fine-tune their analysis to their specific needs. A collection of plugins is provided with ENLIGHTEN™ to make it easy to perform useful analyses right out of the box, and as a starting point for the user to adapt plugins to their own application needs.

ENLIGHTEN™ 4.1.22 (and later) makes use of this architecture and installs with a new 'Library Matching' plugin which offers spectral matching functionality based on a Pearson correlation algorithm. The plug-in continuously matches the latest spectrum shown in the graph window against the user's chosen library, quickly and easily reporting the top library compounds that best

match the unknown sample. This plugin includes a pre-installed library of approximately 50 common, benign compounds as quick-start examples, all taken with our WP Raman 785X spectrometer. See the full starter library in our ENLIGHTEN™ Github repository

In keeping with the ENLIGHTEN™ plugin philosophy of encouraging users to adapt the software to their specific needs, this mini library is meant as a starting point. The user can easily expand the library with specific samples that are relevant to their application by clicking on the 'Save to Library' button in the plugin. The spectrum will be saved to a separate folder and is immediately available for matching.

When considering your Raman library matching needs, it's important to note that a Pearson matching algorithm works best with a limited library. If the number of spectra in the library exceeds several hundred samples, the processing time tends to grow beyond a live display. Also, the probability of false positives grows with the number of spectra in the library. If a larger library is required, one solution is to sub-divide the sample spectra into easily identifiable groups. For example, create one sub-library for solvents, one for polymers, one for powders, etc. Prior to performing the library matching, simply select the appropriate sub-library depending on the matching application of interest.

That's what makes the Raman library matching capability in ENLIGHTEN™ so powerful – it allows you to develop your own custom, application-specific library at NO COST, no matter how unusual your samples might be. And by limiting it to only the compound spectra you need to identify, you can dramatically increase your matching speed and avoid potential false matches to irrelevant compounds. It also allows you to ensure that the custom library and sample spectra are taken with the same spectrometer, further increasing match accuracy and speed.

To learn more about our ENLIGHTEN™ library plugin, read the User Guide found on our website, or visit our YouTube channel for a video tutorial.

A QUICK AND EASY MATCH

Library matching using a simple Pearson correlation coefficient algorithm is a fast and effective way to identify unknown compounds, whether you are a beginner to using Raman for compound identification, or an advanced user developing your own custom material library. And much like Raman spectroscopy itself, our library matching capability in ENLIGHTEN™ places easy-to-use Raman specificity at your fingertips. Curious to learn more about the appropriate Raman spectrometers and analysis for your application? Contact us

User Note: *The ENLIGHTEN™ library matching plugin and associated example library is provided "as is" and is not approved for commercial use. Furthermore, ENLIGHTEN™ spectroscopy software and its associated plugins constitutes a 'product', and is therefore subject to our Terms & Conditions, which can be found at https://wasatchphotonics.com/wp-content/uploads/P-048_TermsConditions_RevB-21Sep17.pdf.*