Application note

A guide to raw material analysis using Fourier transform near-infrared spectroscopy

Author

Jeffrey Hirsch, Thermo Fisher Scientific
Madison, WI, USA

Keywords

Antaris, FT-NIR, raw material identification, raw material qualification, spectral libraries

In this document, we discuss the principles behind the planning, development, and implementation of Fourier transform near-infrared (FT-NIR) spectroscopic libraries for raw material identification and qualification. This document considers not only the mechanics of library construction but also the steps in establishing a raw material program. This includes the planning stages as well as the implementation and maintenance of the library. The considerations for raw material library construction are based on precedents that appear in the literature and upon the experience of our customers.

Introduction

FT-NIR spectroscopy has found increased use in the pharmaceutical, polymer, and chemical industries in recent years. Although FT-NIR has been used for quantitative analysis in many circumstances, the most frequent use has been for the unambiguous confirmation of raw material identity.

While FT-NIR raw material testing has been implemented in many pharmaceutical companies, polymer and chemical companies have also been attracted to this application. Organizations that supply materials to pharmaceutical companies experience the same regulatory driving force to implement a scrupulous raw material testing program. Such a testing regimen can, in addition, help with zero-defect programs. As a result, confirming the identity of materials before they are used for a particular process is a sensible approach to production.

Benefits of using FT-NIR include:

- Reduced use of technician time for raw material testing
- Elimination of solvent consumption and disposal
- Elimination of worker exposure to noxious chemicals used for wet-chemical raw material testing
- Facilitation of conformance to regulatory requirements
- Faster raw material approval in the receiving area with the resulting increase in material throughput and reduction in quarantine time

FT-NIR has been used to confirm the process suitability of materials. This type of analysis is often called “qualification” and involves the careful and rigorous comparison of the spectral signature of each lot with an extensive library of the material in question. If the lot of the material being tested closely matches the spectral properties of library standards, then the subject lot should also be suitable. The extension of FT-NIR for qualification can save a substantial amount of money by preventing the production of an out-of-specification product.

We have written this document as an aid to those who intend to utilize a Thermo Scientific™ Antaris™ FT-NIR Analyzer for raw material testing. We describe the principles behind the use of qualitative FT-NIR spectroscopy for raw material testing and the steps in planning, constructing, and implementing customized libraries.

Experimental

Instrumentation

For all of the experimental work reported in this document, a Thermo Scientific Antaris Method Development Sampling (MDS) System was used. The Antaris MDS System has all of the sampling devices needed for raw material analysis (Figure 1) and also provides the sampling techniques needed to analyze any sample type in any environment. Included with the analyzer is a Thermo Scientific™ SabIR™ Fiber Optic Probe for remote sampling of liquids and solids, an Integrating Sphere for reflectance analyses, and a compartment for transmission analyses of liquids, semi-solids, or transparent solids. An optional, solid transmission module can be added, which may be advantageous for certain raw material analyses. For example, the quickest and easiest way to confirm the identity of a raw material, such as a polymer film, is with the solid transmission device.

Figure 1: Antaris FT-NIR Method Development Sampling System.

Data collection

All FT-NIR data were collected using Thermo Scientific™ RESULT™ Software, which provides the tools needed to comply with 21 CFR Part 11. The software also comes with the necessary documentation for end-user validation. The Thermo Scientific™ ValPro™ System Qualification Software was used for all equipment qualification and performance verification testing using NIST-traceable standards in accordance with USP Chapter <1119>.

Because all of the samples in our example library were solid, the data reported in this document were collected by diffuse reflectance using the Integrating Sphere module. The powders were placed in standard, 2-dram glass vials. The vials were then placed directly on the Integrating Sphere, allowing the spectra of the materials to be collected through the bottoms of the vials. Data were collected using the 16 cm^{-1} resolution and five co-averaged scans.

Data analysis

The data were analyzed after collection using Thermo Scientific™ TQ Analyst™ Software. For the example library created for this report, no spectral pre-treatments were used. However, it is often desirable to use these pre-treatments for more difficult circumstances. TQ Analyst provides the capability to allow pre-treatments, such as Savitzky-Golay and Norris smoothing, derivatives (first and second order), Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) with De-trending.

Results and discussion

Spectral libraries

We should first consider the definition of a spectral library. A spectral library will be defined as a collection of spectra used for qualitative analysis. Any number of algorithms can be applied to such a library to produce a model that can be used as a qualitative analysis method.

Libraries can be standardized or customized. A standardized library is generally available for purchase and is constructed over time using spectra from multiple sources. In such commercial libraries, the sources of the materials used for the data are not discriminated with respect to grade or material manufacturer. These libraries tend to be very large. In the realm of vibrational spectroscopy, such libraries are often used for mid-infrared or Raman applications quite successfully. Alternatively, a customized library contains spectra that represent materials of interest for a particular application. The spectra in customized libraries typically represent the specific grade of materials used by a particular organization and are generally vendor-specific. For example, a specific application requiring a customized library might be the identification of incoming raw materials for a particular facility. To accomplish this, the library for this application will contain only the materials received by that facility. The library would then be tailored to the individual grades of material and to the vendor who manufactured each material.

Strategies for library construction

There are several steps and considerations with respect to library development and construction. The 12-step program shown in Figure 2 is essential to the development of a successful raw material program. A noteworthy observation is that most of the work done when developing an FT-NIR raw material analysis program involves planning and implementation; only a relatively small amount of time is required in the laboratory.

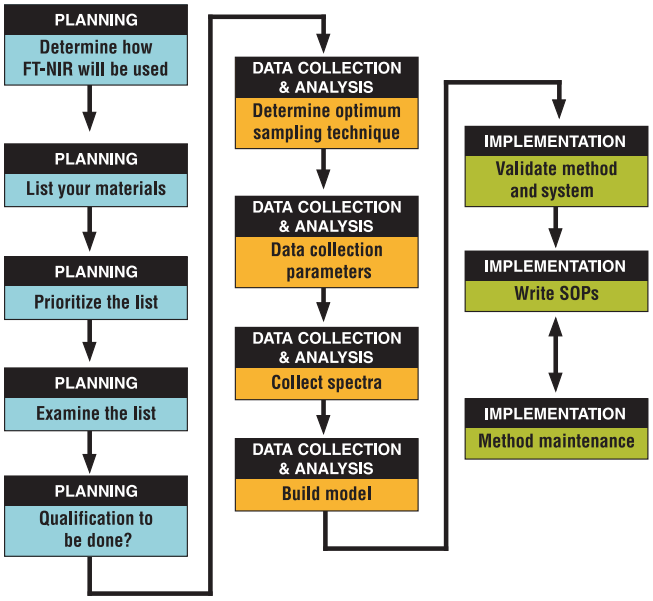

Figure 2: Flow diagram showing the steps in planning, developing and implementing a raw material program.

Step 1—Logistics: Carefully consider how the logistics of a raw material analysis program will operate in your facility. Most questions about the construction of an FT-NIR library can be answered by considering how it will be employed by its users.

The method developer must decide on the following questions:

- *Will confirmation of raw material identities be the only analysis, or will material qualification also be done?*
- *Will analysis be done in the receiving room, laboratory, or a different location?*
- *What sampling scheme will be used to accommodate the organizational logistics?*
- *Will FT-NIR be used for testing hazardous or toxic materials?*
- *Will analysis of difficult material types be done?*
- *Will flexibility toward expansion be needed?*
- *Will the number of spectra and material classes be increased?*
- *Will multiple shifts use the instrument?*
- *Will stringent security requirements be employed?*

The answers to these questions will determine how the library will be used and, in turn, determine how it should be developed to meet the particular needs of any program. The Antaris FT-NIR Analyzer provides the proper tools for raw material analysis in a diverse range of environments.

Step 2—List your raw materials: It is important to make a comprehensive list of raw materials for the library. The list will vary in size and scope for different projects. In general, pharmaceutical companies will want to identify every raw material used. Smaller companies may limit their list to twenty or fewer materials. On the other hand, larger companies may include several hundred materials. Cosmetic companies may have two to three thousand. Making this list prior to development will help in planning the library's overall development much more effectively. The amount of work required will be directly proportional to the number of materials included.

Step 3—Prioritize the list: If the list of raw materials is large, prioritize the list and plan on dates for implementing the library in stages. For companies with large lists of materials (i.e., 50 or more), tackling the entire array at once is a daunting task. For example, a twenty-material library may require only three to four weeks to implement. A 200-material library may need eight to ten weeks to complete. Waiting to implement the entire library at once delays the benefits of using FT-NIR. A temporary, partial library, in this case, becomes a reasonable alternative.

Step 4—Examine the list for potential conflicts: The larger the list of materials, the more potential there will be for conflicting pairs of products. Conflicts are sets of one or more materials that can potentially be indistinguishable or poorly distinguishable from one another.

The materials with the greatest potential for conflict are those that are chemically similar but of a different grade. For example, lactose is a material that can be purchased in many different forms (i.e., different grades). Many times, the differences between these grades are the differences in particle size distribution, as shown in Figures 3 and 4. Other examples of potential conflicts include:

- Long-chain fatty acids (stearic acid vs. palmitic acid)
- Organic materials with different counterions (sodium stearate vs. magnesium stearate)
- Oils from different sources (different natural oils)

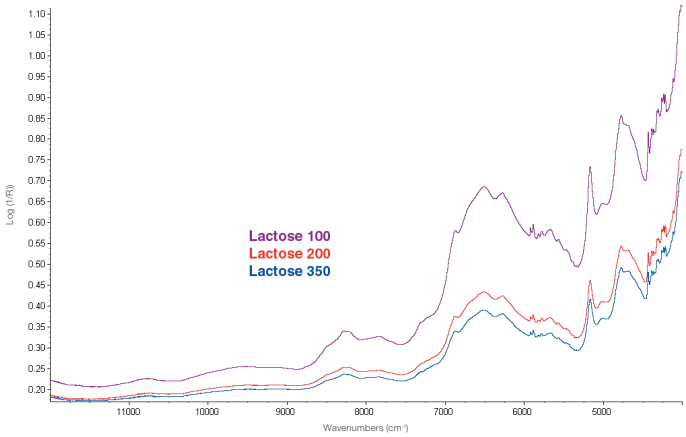

Figure 3: Untreated spectra for three grades of lactose, which differ by particle size distributions.

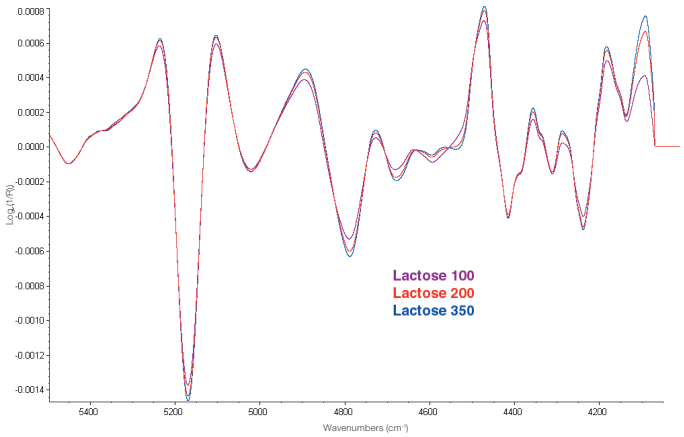

Figure 4: Expanded 2nd-derivative spectra for different grades of lactose.

Step 5—Decide whether the material qualification is desired and/or practical: At this point, it should be decided if material qualification is worth the investment of extra time. Qualification goes beyond the relatively straightforward procedure of identification to determine the appropriateness of a material for its intended purpose. The implementation of this type of analysis increases the likelihood that problems with a raw material will be detected prior to the production of an unacceptable product. For production support personnel who have struggled consistently with faulty raw materials that have caused their final product batches to be scrapped, such methods will become a useful analytical procedure.

The main difference between an identification method and a qualification method, in terms of the experimental work, is the number of lots of material put into the library. While three to five lots of a material are usually adequate for an identity confirmation, twenty or more are generally required to effectively accomplish material qualification. Qualification requires an algorithm that assumes a parametric distribution of variance in the library data. Hence, the software needs to have an adequate representation of material to discern the difference between the "good" and the "bad." If the development of qualification methods is desired for some of the more prominent materials, then it should be planned appropriately. TQ Analyst allows a method that starts as an identification method to be converted to a qualification method through the addition of more samples to the library.

Step 6—Decide on the sampling technique(s): This decision will be largely based on the information from the first five steps. The sampling scheme is a critical consideration. With the Antaris MDS system, all sampling modes are accommodated on one instrument. Modules for 1) transmission analyses of liquids, 2) reflectance analyses of solids and semi-solids, 3) transflectance analyses of liquids and semi-solids, and 4) transmission of solid samples can be accommodated without system modification. This presents significant logistical advantages for the user in terms of avoiding tedious hardware changes, and it obviates the temptation to use one type of sampling for all materials. System re-qualification becomes unnecessary when a change in sampling mode is employed.

Step 7—Data collection parameters: Decide on the parameters to be used for data collection. For some materials, the library can be started with as few as one or two lots of material. Duplicate data collection for each sample is always desirable. We suggest at least five co-averaged scans for each data collection event. If there are potentially conflicting materials, then a resolution of 2 or 4 cm^{-1} should be employed. Otherwise, 8 or 16 cm^{-1} resolution should be adequate. The higher resolution available on an Antaris allows potentially conflicting materials to be more easily distinguished. You should remember that there is a trade-off between resolution and analysis time. However, the difference in time may be insignificant, depending on the number of replicate scans per analysis.

If the library is to be employed for material qualification, then at least twenty lots should be used to represent each material. Data should be collected in triplicate, and one-minute collection events should be employed. Since this analysis is more demanding, it is worth using 2 or 4 cm^{-1} resolution from the outset.

Step 8—Data collection and reference analyses: After the planning and up-front decision-making are completed, it will finally be time to begin data collection. The extent of this process will vary with the number of materials in the library and the library's purpose.

Reference analyses performed in support of the library construction process are very important. If the library is simply to be used for identity confirmation, the reference analysis should be the USP, AOAC, or some other standard method for each material. This should be done for the purposes of library validation and to safeguard against an unlikely error in material identity.

If the library is to be used for material qualification, then a reference analysis for the property of concern should be done in parallel if such a method exists. However, the "reference analysis" may be the result of the production lot for which the material is used. In such cases, a successful production run is a de facto indication of a good raw material.

Step 9—Create the model: The model can be created after the data are collected. The model is the chemometric procedure that actually allows the mathematical distinction between the materials in the library. There are four qualitative (classification) algorithms in TQ Analyst. "QC Compare" is a correlation algorithm that employs a 1-Nearest-Neighbor (1-NN) approach and is used for simple cases. "Distance Match" (DM) is a method for more difficult distinctions that uses the average spectrum and standard deviation for each material. DM works well for distinguishing materials with different particle sizes. DM is a good tool for material qualification methods.

Two techniques based on Principal Component Analysis (PCA) are also available. These tools become more useful as the library's complexity increases. "Discriminant Analysis" (DA), which uses the Mahalanobis Distance as the metric for matching, is the more common of the two because it is less restrictive with respect to the number of samples per class of material. The other technique, Soft Independent Modeling of Class Analogies (SIMCA), is excellent in situations where the developer has multiple batches (15 or more) per class of sample. As such, the method is typically applied for material qualification as often as identity confirmation.

Both PCA techniques are accessed through Discriminant Analysis in the TQ Analyst Software. DA, however, uses a single distribution for all products, while SIMCA uses a unique distribution for each class. The TQ Analyst Software offers Wizards to help the user choose the proper technique.

Step 10—Validate the library: To validate identification methods, reasonable challenges should be used to verify the model's specificity. Ruggedness and method reproducibility should also be demonstrated.

Positive challenges should be used for specificity to verify that the library can be used appropriately to identify known materials. Positive challenge samples are specifically those compounds represented in the library. However, these samples should be batches of materials not actually used to construct the library. The response anticipated for positive challenge samples is that they will be identified as the labeled material in each case. By contrast, negative challenge samples should also be used for validation and, when tested, should not be identified as part of the library. Negative challenge samples are materials not represented in the library but could be logically confused with substances that are represented among the library inventory. The model should fail to recognize these materials.

In order for the method to be valid, you must use qualified equipment. The Antaris is equipped with the proper tools to allow equipment qualification and validation. One tool is the validation wheel that is provided with the ValPro system qualification option (see Figure 5). The validation wheel includes NIST-traceable and serialized standards required to run the USP-recommended performance tests. Automated tests using the validation wheel allow for ongoing qualification of the Antaris unit as per the current USP guidelines. The ValPro package also provides the tools for user qualification (DQ, IQ, OQ, and PQ). RESULT Software conforms to all relevant regulatory requirements, including cGMP, ISO, and 21 CFR Part 11.

Figure 5: Validation wheel for the Antaris

Step 11—Standard operating procedures: This part of the program is not unfamiliar to those in the pharmaceutical industry. Standard Operating Procedures (SOPs) must be written to govern the ongoing performance of the testing program. The exact repertoire of procedures will vary slightly from organization to organization, but procedures governing instrument qualification and maintenance, remedial action upon instrument qualification failure, routine method performance, sample processing and documentation, sample test failures, library updates (see Step 12) and data archival should be included among the SOP list. RESULT Software is able to display electronic SOPs specific to the operation of the Antaris FT-NIR Analyzer and qualification of instrument performance giving the user convenient access to operating procedures.

Step 12—Library updates and method maintenance: Long-term database management is a key issue as the library will need to be updated and adjusted over time. Before implementation, you should decide how this will be done and how an updated library will be validated. In step 3, we discussed the issue of stepwise library development for those companies with large raw material inventories. If you take this approach, you need to decide how the library will be validated at each stage of expansion. The most critical issues to address are the validity of the new data for the classes of materials they represent and the assurance that there are no conflicting materials in the overall library following the update. There will also be a need to update the library in a similar manner as new raw materials are introduced into the organization. A procedure must be in place to accommodate these additions.

The initial data in a library will not necessarily be representative of incoming raw materials indefinitely. This presents another long-term issue with respect to library updating, often termed “method maintenance.” The nature of all materials is that they will “drift” over time. This infers that the physical and also possibly the chemical status of incoming materials can, due to many factors, change slightly as time passes compared to the materials in the library. This means that monitoring of the library status is essential. The library will need to be updated by discarding old samples no longer representative of incoming material and replacing them with newer samples.

When a raw material library is managed properly, it will continue to effectively serve its intended purpose. But the management technique must be carefully planned and implemented at the outset. With each step, the program manager should decide how the changes will be validated. With each change or expansion, new positive and negative challenge samples should be tested, and the library re-examined for conflicting pairs.

Example library

For the purpose of illustrating some of the principles described above, we have constructed a small library, in which we have included ten common materials. The list is given below:

- D-glucose
- D-fructose
- Sucrose
- D-mannitol
- D-sorbitol
- α-D-lactose monohydrate
- Acetylsalicylic acid
- Acetaminophen
- L-ascorbic acid
- Citric acid

Included are six sugars (four monosaccharides and two disaccharides) used frequently as excipients. Acetylsalicylic acid (aspirin) and acetaminophen (also known as paracetamol) are two common active ingredients for analgesics. Ascorbic acid (antioxidant) and citric acid (flavoring and pH control) also are often used as excipients in pharmaceutical formulations. With a small library such as this, we would not need to consider implementation in stages. We will only use this library for identity confirmation. All of the analyses were done in the laboratory, so the integrating sphere was the sampling method of choice. Because the integrating sphere is not subject to variations due to sample presentation and placement, it provides more reproducible results than the fiber optic probe.

The untreated spectra for these materials are shown in Figure 6. It is apparent that all of the materials are spectrally unique. This is generally the case with qualitative material identification. Even for materials that are in the same chemical class, such as sugars, the FT-NIR spectra are different. We would expect a simple QC Compare method to work well for this case. Indeed, when these spectra were analyzed using the QC Compare algorithm, no conflicting compounds were found. This is commensurate with our visual inspection of the spectra.

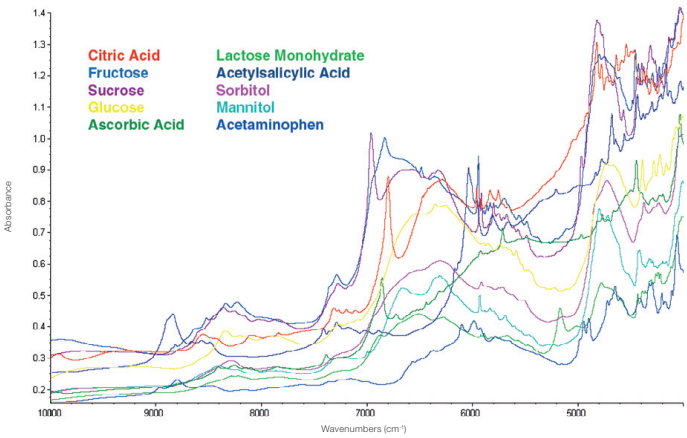

Figure 6: Untreated spectra used to construct an example library.

The next step in our example case is the validation of the model. For this purpose, we have assembled duplicate lots of three of the materials (fructose, glucose, and acetaminophen) as positive challenge samples. Three materials were also available as negative challenge samples. Salicylic acid was used as a negative challenge for acetylsalicylic acid. This is a good choice because the two compounds are somewhat similar in appearance, are close in name, and could easily be manufactured in the same vendor facility. Also, salicylic acid is both a synthetic impurity (it is the penultimate material in the synthetic process) and a degradant of acetylsalicylic acid. 2-Acetamidophenol was used as a negative challenge sample for acetaminophen because it is a constitutional isomer and a possible synthetic impurity. α-D-lactose anhydrous was used as the negative challenge sample for the hydrated lactose. This is also a very good choice as the materials have similar names and are very likely to be produced in the same facility. They are also related because the monohydrate can be produced from the anhydrous material and vice versa. All of these challenge materials were tested against the model created for our identification library.

Figures 7 and 8 compare the spectra of two of the positive and negative challenge sample sets with their corresponding library materials. Table 1 shows the results of the validation experiment. For the average library using a QC Compare model, a match score of 90 or better would indicate a significant similarity to a library material and could thus be used as the passing threshold for a test material. The results for all of the challenge materials were as expected.

Conclusion

The information we have presented in this paper represents a practical guide for would-be developers of raw material libraries. We have outlined in detail the steps for developing a raw material program. In addition, an example library was presented. This information shows the potential benefits of raw material libraries both for identity confirmation and for material qualification. It also shows that, with proper planning, execution, development, and ongoing management, a raw material library can be used very effectively and productively in a variety of facilities, including those involved with the production of pharmaceuticals, chemicals, polymers, and other products. We have also described how a material qualification library can prevent the costly production of unsuitable products.

The data were collected using an older model instrument Antaris FT-NIR. Currently, Thermo Scientific offers an improved model, the Antaris II FT-NIR, which offers superior speed and performance over its predecessor model.

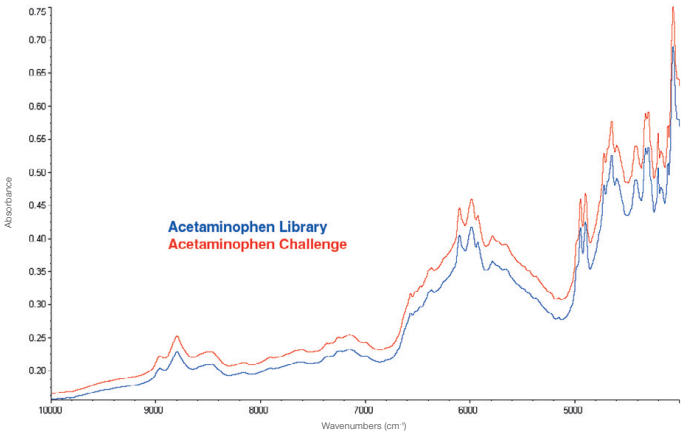

Figure 7: Untreated spectra for acetaminophen library material and the acetaminophen positive challenge sample.

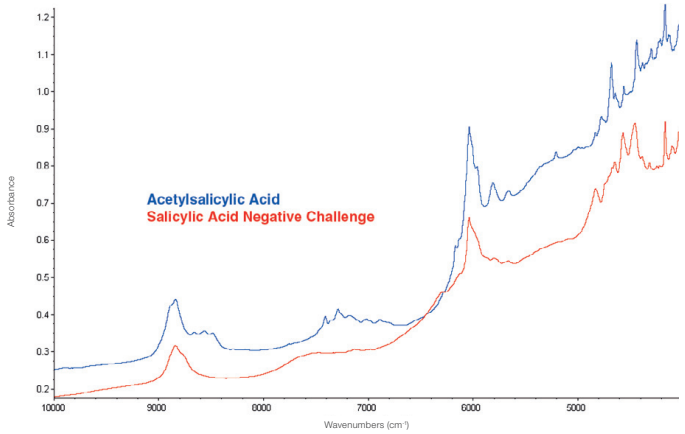

Figure 8: Untreated spectra for acetylsalicylic acid library material and the salicylic acid negative challenge sample.

Table 1. Results for library challenge samples

Challenge material	Type	Match score	Result
Glucose	Positive	99.9	Match with library glucose
Fructose	Positive	98.0	Match with library fructose
Acetaminophen	Positive	100.0	Match with library acetaminophen
Salicylic Acid	Negataive	55.2	No match
2-Acetamidophenol	Negataive	19.3	No match
α-D-lactose anhydrous	Negataive	68.9	No match

Typical threshold for a passing match score is 90.0

Learn more at **thermofisher.com/nir**