IR Spectroscopic Mixture Deconvolution

Michelle D'Souza, Ph. D., Marie Scandone, and Gregory Banik, Ph. D.
Bio-Rad Laboratories, Inc., Informatics Division, Philadelphia, PA, USA

Spectroscopy 95934

Abstract

Deconvolution of mixtures using various spectroscopic techniques is a challenging task. Manual separation of spectral components, even when they are known in advance, is a tedious job. To assist in this process, Bio-Rad has created a Mixture Analysis application as part of it KnowItAll® software. This tool deconvolutes components of a mixture by comparing a sample spectrum against Bio-Rad's KnowItAll databases or user's own proprietary spectra. The result is a series of composite spectra, each accompanied by the individual component spectra that comprise the composite spectrum as well as the residual spectrum (the difference between the query spectrum of the actual mixture and the composite spectrum). The composite spectra are ranked by how closely they resemble the query spectrum. Through a series of cases studies, we will demonstrate the efficiency and accuracy of this approach and clear benefits it has over traditional methodolgies.

Introduction

Bio-Rad has been researching the automation of spectroscopic analysis utilizing computer technologies. Spectral search against a large data collection is an obvious method. However, the spectrum of an unknown mixture rarely exists in a collection of reference spectra. A component can be missed as a candidate when its reference spectrum is not one of the top hits. Spectral deduction is often used for separating known from unknown, which depends as much on the availability of reference spectra as it does on a user's skill.

Alternatively, one can associate a functional group with spectral peaks. This is an additive approach whose success depends on a knowledge base that may not provide a complete picture of the unknown.

User experience is an important factor to the reliability of this kind of analysis outcome. Bio-Rad has successfully applied its patented Overlap Density Heatmap technology to highlight the similarities and dissimilarities in a group of spectra. This method provides researchers with clues about whether or not an unknown belongs to a particular, well-characterized group of compounds. Again, the user experience level is critical.

Similarly, a user can perform a multivariate analysis to categorize spectra. Usually, expert knowledge is needed to perform this type of analysis, yet the end result is to categorize an unknown, not to identify components. These methods are often used in conjunction to piece together a more complete picture of a mixture sample.

To streamline this process, the Informatics Division of Bio-Rad Laboratories created a Mixture Analysis application for its award-winning KnowItAll software that, unlike traditional methods, rapidly and reliably deconvolutes mixtures. This application can be used to compare a sample spectrum against licensed KnowItAll reference databases, such as the renowned Sadtler and Wiley IR databases — the world's largest collection of infrared spectra. In addition, it can use KnowItAll databases created in-house. The application attempts to find all combinations of spectra in the libraries that, when combined in the right proportions, result in a minimal difference between the query and composite spectra. The user can search for two, three, or four components, although the application will return fewer components if the match is better.

The result is a series of composite spectra, each accompanied by the individual component spectra that comprise the composite spectrum as well as the residual spectrum (the difference between the query spectrum of the actual mixture and the composite spectrum). The composite spectra are ranked by how closely they resemble the query spectrum. The speed of this process has been highly optimized, and numerous tests have confirmed its accuracy. Several case studies are provided below as examples.

Case Studies

Mixtures of Polymers (KBr Pellet)

The ability to use spectroscopic techniques to analyze polymer mixtures has been a repeated request from the IR user community. In the past, besides providing the largest collection of IR spectra for searching, Bio-Rad developed complementary IR and Raman functional group analysis applications. Recently, a similar application was created for polymer functional group analysis by FT-IR.

The sample in this case study was a mixture of three different polymers. Simply searching the unknown mixture spectrum against the Sadtler Basic Monomers and Polymers Volume I and II databases yields a hit list with twenty-nine polyvinyl chlorides at the top. Polyvinyl chloride is a dominant component, but not the only one. Polymer functional group analysis provides the associations of structural features and polymer types to spectral peaks, but does not identify the exact chemical components.

The Mixture Analysis application was run using the Sadtler Basic Monomers and Polymers Volume I and II databases with the FTI-IR spectrum of the mixture of three polymers (KBr pellet). The first analysis result is shown in Figure 1. The Mixture Analysis identified the top two components, polyvinyl chloride and polycarbonate, correctly. The reference databases selected did not contain a spectrum of polymethylmethacrylate, the third component, however the Mixture Analysis application did recommend a co-polymer containing 30% of the polymethylmethacrylate as the third component. While not an exact match, the co-polymer does provide useful information about the nature of the third component.

The HQI (Hit Quality Index) column shows the quality of each composite spectrum (999 is a perfect match). The first composite spectrum (red) closely matched the query spectrum (black); the residual (blue), the difference between the composite spectrum and the query spectrum, was relatively small.

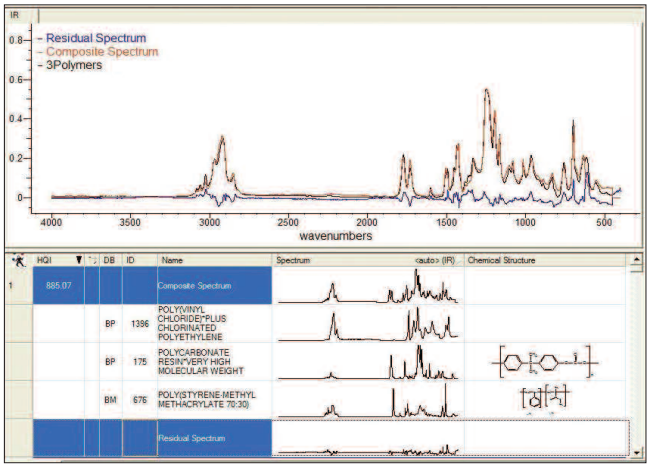

Fig. 1. Analysis of a three component polymer mixture (FT-IR).

Bio-Rad's databases contain chemical fragments or structures when they are available. The last hit is a copolymer, and two structure fragments are displayed

Mixtures of Pharmaceuticals (KBr Pellet)

Drugs are often a mixture of active compounds and excipients, and on rare occasion, contaminants. Pharmaceutical mixture deconvolution is thus an attractive application area for spectrum analysis. In this case study, a simple database search of an IR spectrum (KBr pellet) of a mixture of three steroids was performed against the complete collection of 220,000 IR spectra resulting in prednisolone (an actual component) at the top of the hit list followed by ethisterone (another component) as the fourth hit. Epiandrosterone (3-β-Hydroxyl-5-α-adrostan-17-one), the third component, was at the ninth position. This result showcases the common confusion and mistakes that can occur in component identification by simple spectrum searching. IR functional group analysis for this mixture provided only fragmented chemical group information.

The same collection of 220,000 IR spectra was selected to conduct a mixture analysis for the mixture of three steroids. The result is shown in Figure 2.

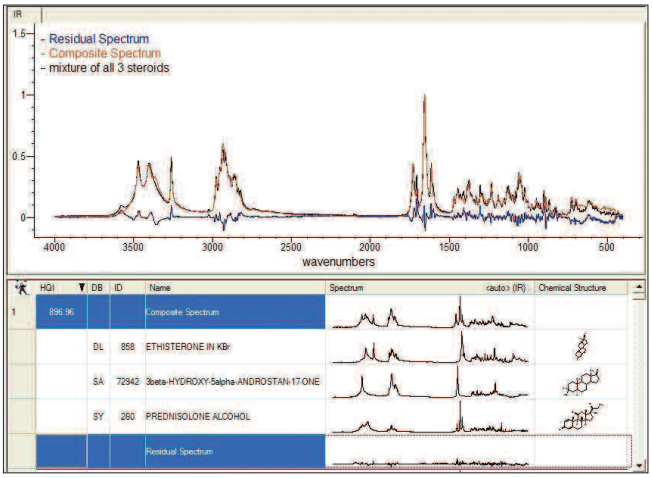

Fig. 2. Analysis of a three component steroid mixture (FT-IR).

All three components—epiandrosterone, ethisterone, prednisolone alcohol—were positively identified by the Mixture Analysis application in KnowItAll as the most highly probable components of the mixture. Since the chemical structures of these compounds are stored in the databases, they are displayed to confirm the identities for these chemicals in the right column.

Mixtures of Pharmaceuticals (ATR)

Attenuated Total Reflectance (ATR) is a special IR sampling technique. Its sample preparation is much simpler than that of traditional FT-IR. In recent years, it has been applied in many

sample analysis areas. It is important that Bio-Rad's mixture analysis tool work well with this type of spectrum. When using an ATR spectrum of a mixture of two steroids (ethisterone and epiandrosterone) to search against Sadtler ATR of Steroids, Androgens, Progestins & Estrogens database, ethisterone was obtained as the first hit; however, epiandrosterone was seen as the thirtieth hit. This again highlights the problem of component identification by traditional methods.

The Sadtler ATR of Steroids, Androgens, Progestins & Estrogens database was then used to analyze the ATR spectrum of the steroid mixture using the Mixture Analysis application. The result is shown in Figure 3.

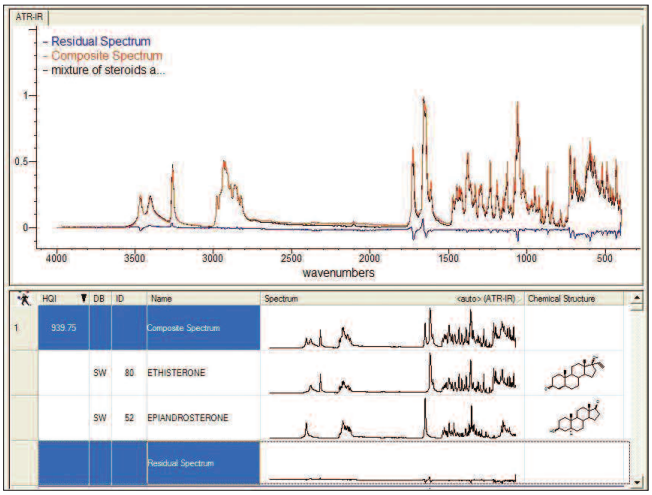

Fig. 3. Analysis of a two-component steroid mixture (ATR).

The two steroids in the mixture were correctly identified from the single spectrum of the mixture, accompanied by their corresponding chemical structures. The residual spectrum is very small, indicating that not much signal remains when subtracting the composite spectrum from the sample spectrum.

Mixtures of Organic Compounds (KBr Pellet)

An important quality of a mixture analysis tool is the ability to distinguish very similar compounds. This case study involves such a sample: a mixture of three very structurally and spectrally similar cyclohexylamines (Figure 4).

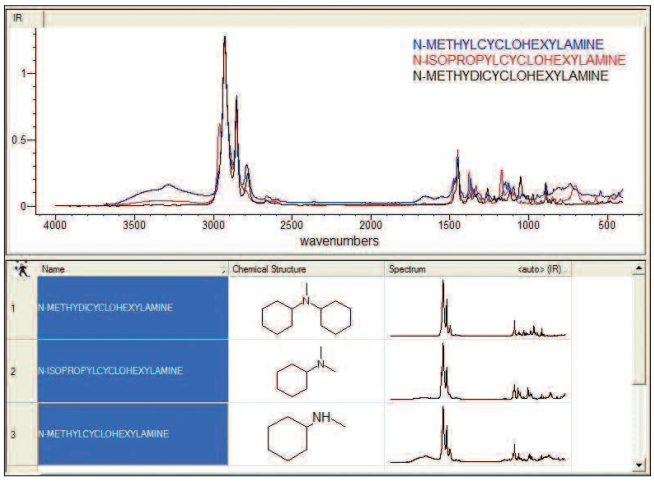

Fig. 4. Cyclohexylamines structures and FT-IR spectra.

These three components of the mixture all contain a cyclohexylamine moiety, and their spectral peaks overlap heavily. A simple spectral search of this mixture resulted in a list where N-methylcyclohexylamine was the top hit, and dicyclohexylamine, a chemical structurally close to N-methyldicyclohexylamine, was the third hit. The actual spectra of N-methyldicyclohexylamine and the third component, N-isopropylcyclohexylamine (the dominant ingredient of the commercial product Polycat 21) were not present in the top twenty hits. IR functional group analysis of this sample highlighted their chemical structure commonalities but could not tell them apart.

For deconvolution of this complex mixture, all 220,000 IR spectra were used to analyze the spectrum using a platform-independent web-based version of the Mixture Analysis application with KnowItAll AnyWare™. The results are shown in Figure 5.

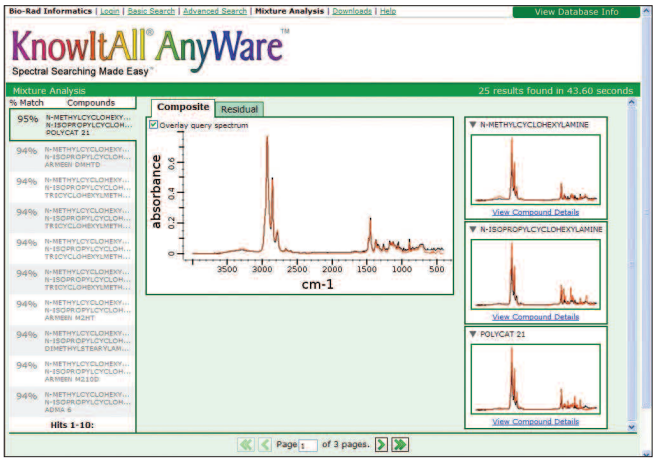

Fig. 5. Analysis of three cyclohexylamines (FT-IR).

In the web version of Mixture Analysis, the percentage match (% Match) shows the quality of each composite spectrum (100% is a perfect match). This example shows that the first composite spectrum (red) closely matched the query spectrum (black). The individual component is presented in the smaller boxes at the right. The residual, the difference between the composite spectrum and the query spectrum, is available for review in a separate tab. A user can drill down from the link of the individual compound to see the complete structure and property information.

Conclusion

These case studies show the advantages of KnowItAll Mixture Analysis over traditional methodologies. Bio-Rad's extensive tests have proven it to be an accurate and effective technique for rapid and reliable spectroscopic component deconvolution.

Acknowledgement

Bio-Rad offers great thanks to Mr. Lewis Smith of Forensic Spectral Research for providing the IR spectra of the mixtures used in these case studies.