

Validation of ATR Correction and Reverse ATR Correction Algorithms in the KnowItAll® Informatics System

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



Abstract

To address wavelength-dependent differences in intensity between ATR and transmission FTIR spectra, a mathematical ATR correction can be applied to an ATR spectrum to provide a better match when searching against a transmission FTIR reference database. Conversely, the reverse ATR correction can also be applied mathematically to a transmission FTIR spectrum to provide a better match when searching against an ATR reference database. This study aims to validate the ATR correction and the reverse ATR correction in Bio-Rad's KnowItAll Informatics System.

Introduction

Transmission FTIR versus ATR Spectra

Transmission Fourier transform infrared (FTIR) spectra are often acquired using a KBr pellet containing a sample. Alternatively, an Attenuated Total Reflectance (ATR) apparatus can be used to acquire the spectrum since it is usually simpler and faster than other methods and offers superior performance for some sample types. Because of differences in peak intensity caused by a wavelength-dependent difference in the depth of penetration in ATR spectra, a transmission FTIR spectrum of a sample will have a different appearance from an ATR spectrum of the same sample.

Database Searching

Reference databases such as the Sadtler FTIR collections and in-house databases have been created to allow spectroscopists to verify the identity of samples as well as to identify unknown samples. Because of the differences in intensity between a transmission FTIR spectrum of a sample and an ATR spectrum of the same sample, searching an ATR spectrum against a reference database of transmission FTIR spectra or vice versa can lead to differences in the matching factor between the query spectrum and the database hits retrieved by the search software.

Database searches are conducted using a Hit Quality Index (HQI) value on a 0 to 999 scale, with an HQI value of 999 indicating a perfect match between the query spectrum and a given reference database spectrum, an HQI value of 0 indicating no match whatsoever between the query spectrum

and a given reference database spectrum, and intermediate HQI values indicating an intermediate level of similarity between the query spectrum and a given database spectrum. Such HQI values can be calculated to compare a query spectrum to each record in a spectral reference database to generate a "hit list." The list is presented in descending order of HQI values to display those reference spectra that are most spectrally similar to the query spectrum first.

Because of subtle differences that can occur with sample preparation, contamination, etc., the first entry in a database search hit list may not be the "correct" hit. Nevertheless, the "correct" hit usually occurs within the top 50 hits if the query spectrum and reference database spectra have been collected properly.

ATR Correction

To address the wavelength-dependent differences in intensity between ATR and transmission FTIR spectra, a mathematical ATR correction can be applied to an ATR spectrum to provide a better match when searching against a transmission FTIR reference database. Conversely, the reverse ATR correction can also be applied mathematically to a transmission FTIR spectrum to provide a better match when searching against an ATR reference database. This study aims to validate the ATR correction (searching an ATR spectrum against a transmission FTIR reference database) and the reverse ATR correction (searching a transmission FTIR spectrum against an ATR reference database) in Bio-Rad's KnowItAll Informatics System.

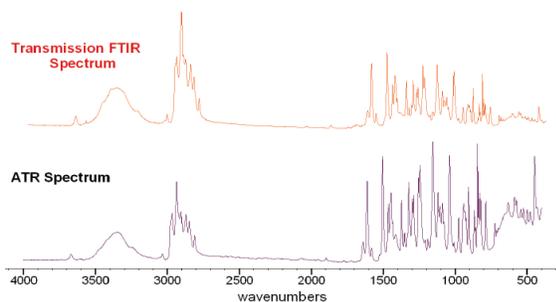


Fig. 1. Transmission FTIR spectrum and ATR spectrum for methylestradiol 3-methyl ether.

Methods

306 steroid samples were acquired and the spectra measured using a Bio-Rad Digilab FTS spectrometer. ATR spectra were measured by attaching a DuraSAMPLIR™ II ATR device from Smiths Scientific to the Bio-Rad instrument. Transmission FTIR spectra were measured either by creating a KBr pellet containing the sample or by dissolving the sample in dichloromethane and creating a film on a KBr crystal by evaporating the solvent from the sample solution on the KBr crystal. Of the 306 spectra of steroids that were obtained, 21 spectra contained obvious water contamination, two spectra had obvious baseline drift, and one additional spectrum containing a reactive chloroformate moiety that appeared to have decomposed. These 24 spectra were excluded from the validation study, leaving 282 spectra pairs measured by both transmission FTIR as well as ATR. Two databases of the 282 spectra were created in the KnowItAll Informatics System software (version 8.0.2): one database containing transmission FTIR spectra only and the other containing ATR spectra only.

Each spectrum in the ATR database was searched against all spectra in the transmission FTIR database using the ATR correction algorithm in the KnowItAll software. The HQI value of the top hit was noted, as was the ordinal position and HQI of the “correct” hit in the hit list, that is, the compound from the transmission FTIR database that is identical to the query ATR spectrum sample. Ideally, the ordinal position of the “correct” hit resulting from searching an ATR spectrum against a transmission FTIR database after applying mathematical ATR correction would be 1. The Euclidean Distance algorithm was used for database searching, and automatic baseline corrections were not performed prior to searching. Similarly, each spectrum in the transmission FTIR database was searched against all spectra in the ATR database using the reverse ATR correction algorithm in the KnowItAll software. The HQI value of the top hit was noted, as was the ordinal position

and HQI of the “correct” hit in the hit list, that is, the compound from the ATR database that is identical to the query transmission FTIR spectrum sample. The Euclidean Distance algorithm was used for database searching, and automatic baseline corrections were not performed prior to searching.

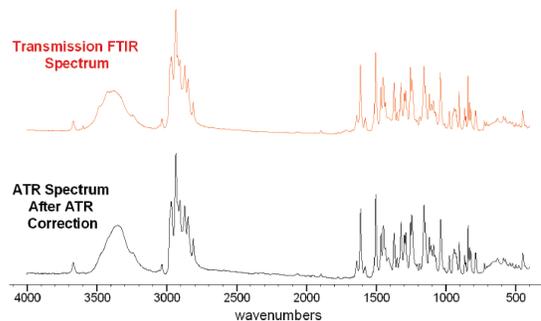


Fig. 2. Transmission FTIR spectrum and ATR spectrum to which ATR correction algorithm has been applied for methylestradiol 3-methyl ether.

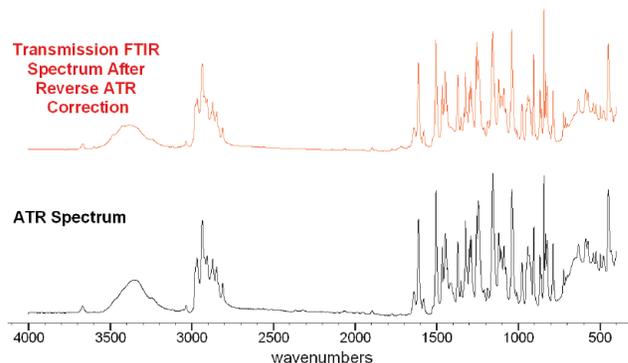


Fig. 3. Transmission FTIR spectrum to which reverse ATR correction algorithm has been applied and ATR spectrum for methylestradiol 3-methyl ether.

Results

ATR Correction Validation

Searching the 282 ATR spectra against the transmission FTIR database resulted in 282 individual hit lists. The “correct” hit occurred in the first position in the hit list in 92.9% of the searches, in the first or second positions in 98.6% of the searches, and in the top ten hits in 100% of the searches. The ordinal rankings of the four hits not in the top two positions were 3, 3, 4, and 6.

Reverse ATR Correction Validation

Searching the 282 transmission FTIR spectra against the ATR database resulted in 282 individual hit lists. The “correct” hit occurred in the first position in the hit list in 95.4% of the searches, in the first or second positions in 99.6% of the searches, and in the top ten hits in 100% of the searches. The ordinal ranking of the single hit not in the top two positions was 3.

Conclusion

Steroids all share a common framework, making this validation study particularly challenging. Nevertheless, the ATR correction algorithm in the KnowItAll Informatics System appears to be very effective at correctly identifying a compound whose spectrum is measured using the ATR method and searched against a database of transmission FTIR spectra. Similarly, the reverse ATR correction algorithm appears to be equally effective at correctly identifying a compound whose spectrum is measured via transmission FTIR and searched against a database of ATR spectra. The correct compound was identified using both algorithms in the vast majority of all searches performed, resulting in “correct” hit being in the top few hits of a hit list almost 100% of the time.



Bio-Rad
Laboratories, Inc.