APPLICATION NOTE AN53037

Food protein powders classification and discrimination by FTIR spectroscopy and principal component analysis

Author

Ron Rubinovitz, Ph.D.
Thermo Fisher Scientific

Key Words

FTIR, ATR, proteins, principal component analysis

Thermo Fisher Scientific solutions

Nicolet iS50 FTIR spectrometer, OMNIC software, TQ Analyst software

Abstract

Food protein powders are effectively classified and discriminated using FTIR ATR measurements and principal component analysis (PCA). Successful grouping by product type was achieved by PCA models based on general composition as well as by focusing on protein secondary structure differences found in the amide I region.

Application benefits

The combination of FTIR spectroscopy and principal component analysis offers a facile means to classify and discriminate protein powders based on product type as well as supplier source and repeatability. The experiments are simple and straightforward and require no sample preparation. The methodology can be readily adopted by many food manufacturers for incoming material inspections as well as QA/QC.

Introduction

Food protein powders are complex mixtures that contain proteins, carbohydrates and fats. Depending on the origin and the extraction/isolation process, even the same type of protein powders can differ in nutritional and processing characteristics. The capability to rapidly classify protein powders of particular type, to discern nominally similar proteins between suppliers, and to assess lot-to-lot variation is of great importance for many food manufacturers to achieve consistent product quality.

One of the most important characteristics of any protein is its secondary structure, defined by local structural conformations dependent on the patterns of hydrogen bonding between amine hydrogen and carbonyl oxygen atoms in the backbone peptide bonds. FTIR has long been established as a viable analytical technique for protein secondary structure characterization. Through the deconvolution or curve-fitting of the amide I band (~1650 cm^{-1}), originating from the C=O stretching vibration of the protein's amide group[1,2], contributions of different secondary structures can be estimated to provide important protein structure characteristics, including conformation and stability[1,2,3,4,5]. It is noted, however, that the curve-fitting approach relies on band assignments of the secondary structures (α-helix and β-sheet) from pure proteins. Therefore, while successful in characterizing isolated single proteins, this approach is less than optimal for the analysis of complex mixtures such as protein powders, since interactions between multiple proteins and non-protein materials affect the spectral features in the amide I region (1700-1600 cm^{-1})[6].

In this note, the feasibility of classifying and discriminating food protein powders based on the combination of FTIR spectroscopy and principal component analysis (PCA) is presented. By selecting appropriate spectral ranges for PCA, similarities/differences between different types of protein powders, the same protein powders from different vendors, and/or different lots, can be successfully assessed with respect to their overall composition as well as protein secondary structure.

Experimental

Samples of milk, pea, rice, and whey protein powder from a variety of vendor sources were made available for analysis. As summarized in Table 1, the number of vendor sources varied from as many as five (whey protein) to one (milk protein), while the number of lots from a single vendor varied from three to one.

Protein powder	Vendor	Number of Lots	Samples
Milk	A	3	A1, A2, A3
Pea	B	2	B1, B2
	C	2	C1, C2
	D	1	D1
	E	1	E1
Rice	F	2	F1, F2
	G	1	G1
	H	1	H1
Whey	I	2	I1, I2
	J	3	J1, J2, J3
	K	2	K1, K2
	L	2	L1, L2
	M	2	M1, M2

Table 1: Protein powder samples measured by FTIR spectroscopy.

Spectra of the protein powders were measured in attenuated total reflectance (ATR) mode using the built-in ATR accessory of the Thermo Scientific™ Nicolet™ iS50 FTIR Spectrometer, which utilizes a monolithic diamond crystal. For each measurement, a small amount of protein powder was placed on the diamond ATR crystal. The pressure tower of the accessory was used to assure good contact between the powder and the diamond crystal. Three sub-samples from each sample were measured at 4 cm^{-1} resolution and 512 scans. The Nicolet iS50 spectrometer was purged with nitrogen to eliminate the influence of water vapor on the spectra. After data collection, the advanced ATR-correction feature of Thermo Scientific™ OMNIC™ Software was applied to all spectra. Results reported here are averages over the three sub-samples. Spectra evaluation for characterization and classification was carried out by principal component analysis using the Thermo Scientific™ TQ Analyst™ Software.

Results and discussion

Protein powder spectra

Figure 1A shows representative spectra of each protein type. Although milk protein powder was available from only one vendor, the lot-to-lot reproducibility of this product can be seen in Figure 1B. For the remaining protein types, representative spectra of the samples from different vendors are grouped in Figures 1C-E. There are noticeable variations amongst different protein types in both the amide I region (1700-1600 cm^{-1}) and the amide II region (1580-1510 cm^{-1}), resulting from the differences in their secondary structure. The whey protein spectra group (Figure 1E) has the largest vendor-to-vendor variation, whereas the pea protein spectra group (Figure 1C) has the least. In addition, variation resulting from the non-protein components is also observed. For example, the carbohydrate peak at ~1080 cm^{-1} varies substantially within each protein group. The lipid peak at ~1743 cm^{-1}, as another example, also varies across the samples. While extremely weak in the milk protein samples (Figure 1B), the lipid peak feature is evident in the pea protein spectra (Figure 1C), the rice protein spectra (Figure 1D) and the whey protein spectra (Figure 1E) with varying intensities.

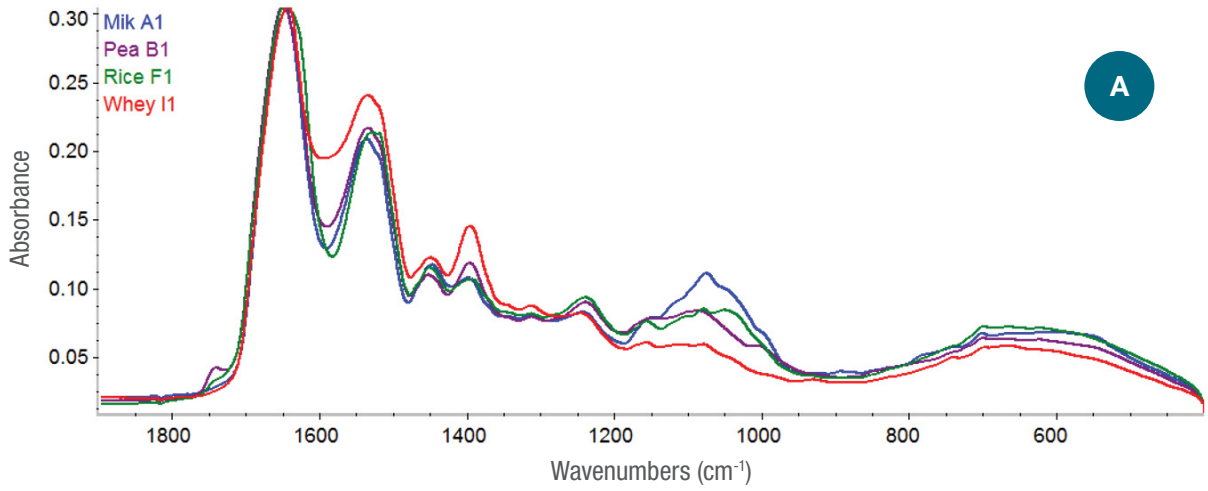

Figure 1: (A) Full-scale ATR-corrected spectra of protein powders.

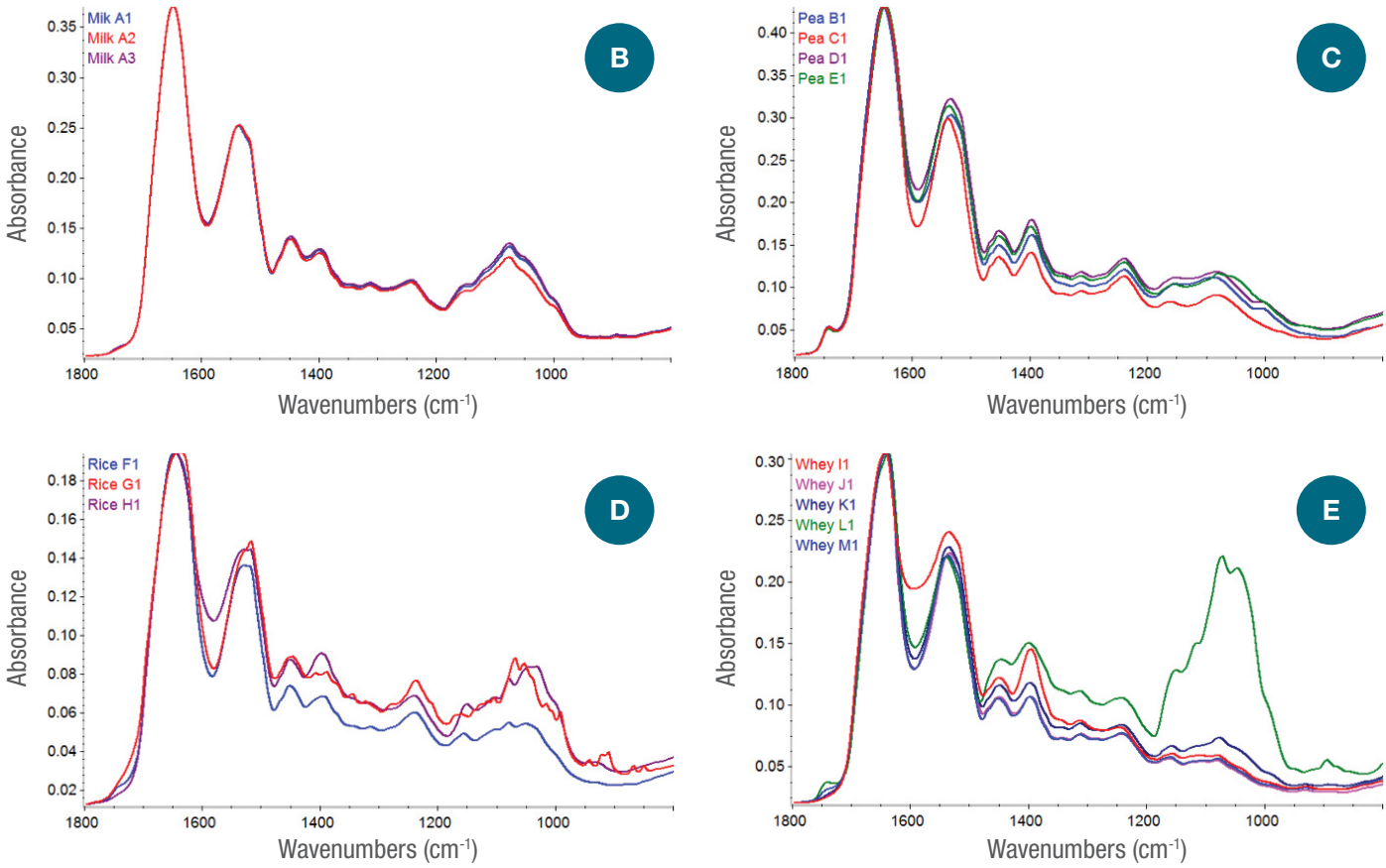

Figure 1: (B) Spectra of the three lots from the one milk protein vendor; (C) Spectra of pea protein powders from three different vendors; (D) Spectra of rice protein powders from three different vendors; and (E) Spectra of whey protein powders from five different vendors.

Classification by PCA using the overall mid-IR region

Principal component analysis is a statistical procedure often used to extract meaningful variance from a spectral calibration set. PCA uses the spectra to calculate factors that are modeled from spectral variance. The first factor has the largest variance in the data set, and each succeeding factor in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. Factors can be linearly combined to reconstruct each individual spectrum of the calibration set. The coefficients for each factor, also referred to as "scores," can be plotted in a PCA space to delineate similarities and/or differences between spectra. In so doing, a spectrum with thousands of wavelength values can be reduced to a single data point in a two- or three-dimensional space, if the overall variance can be effectively modeled using the first two or first three factors, respectively. Prior to principal component analysis, several pretreatments were applied to the ATR-corrected protein spectra. Spectral pre-processing using second derivatives, followed by a standard normal variate (SNV) correction were applied to the spectra in order to compensate for intensity variation caused by different packing densities on the ATR crystal. To obtain the most meaningful variation, only relevant spectral regions are used for the analysis. In this case, the full mid-IR spectral range of 4000-500 cm^{-1}, but excluding the region 2356-1900 cm^{-1} associated with diamond ATR measurements, was used for the PCA. The resulting scores plot is shown in Figure 2.

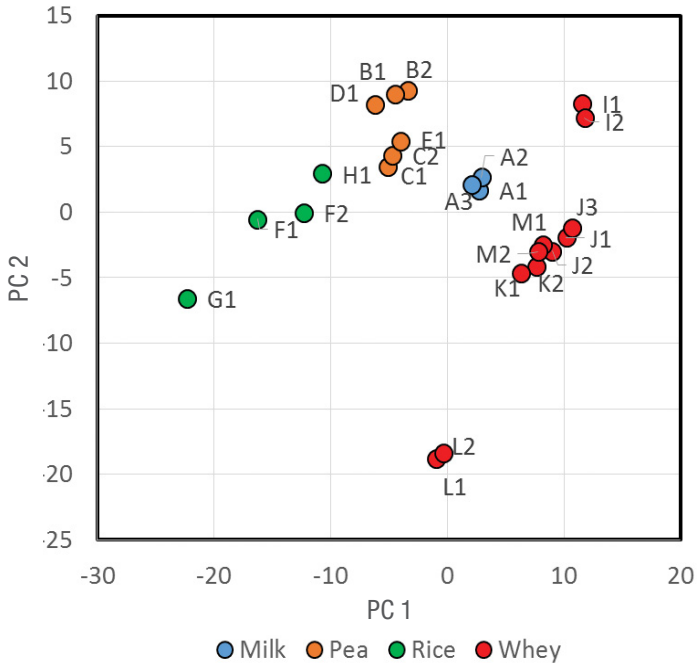

Figure 2: Principal component scores plot from protein powder samples using most of mid-IR region.

As seen in Figure 2, an effective classification of different protein types is achieved with only two principal components (PCs). Each protein type congregates into its own domain in the PC space. Closer inspection of the clusters in Figure 2 also reveals the variation within each protein group. Data points from the same vendor, such as I1 and I2 of the whey protein and A1, A2 and A3 for the milk protein, are closely clustered, indicating a reproducible process for each vendor that yields products of minimal variation. The variation amongst different vendors, however, is generally more pronounced. For example, while vendors J, K, and M appear to make similar whey protein powders, the whey protein products from vendor L and I are distinctly different. The spectrum of the whey protein from vendor L (the green trace in Figure 1E) has a significantly larger absorbance in the 1080 cm^{-1} region than the rest of the group, suggesting a higher carbohydrate content in this product. For the rice protein powders, the data points from different vendors are relatively scattered. For the pea protein powders, while product B is similar to D and product C is similar to E, the two clusters are distant from each other. It is important to note that the PCA described above is based on the overall mid-IR spectral range, the variance, therefore, includes the contributions from both protein and non-protein components.

Analysis of the amide I region

In order to directly compare the proteins in the products, a PCA based only on the amide I spectral region was performed. The amide I spectral region was chosen because it is specific to the protein secondary structure. The amide I region of the spectra for all products are shown in Figure 3, where the shape of the amide I band varies from product to product and from vendor to vendor.

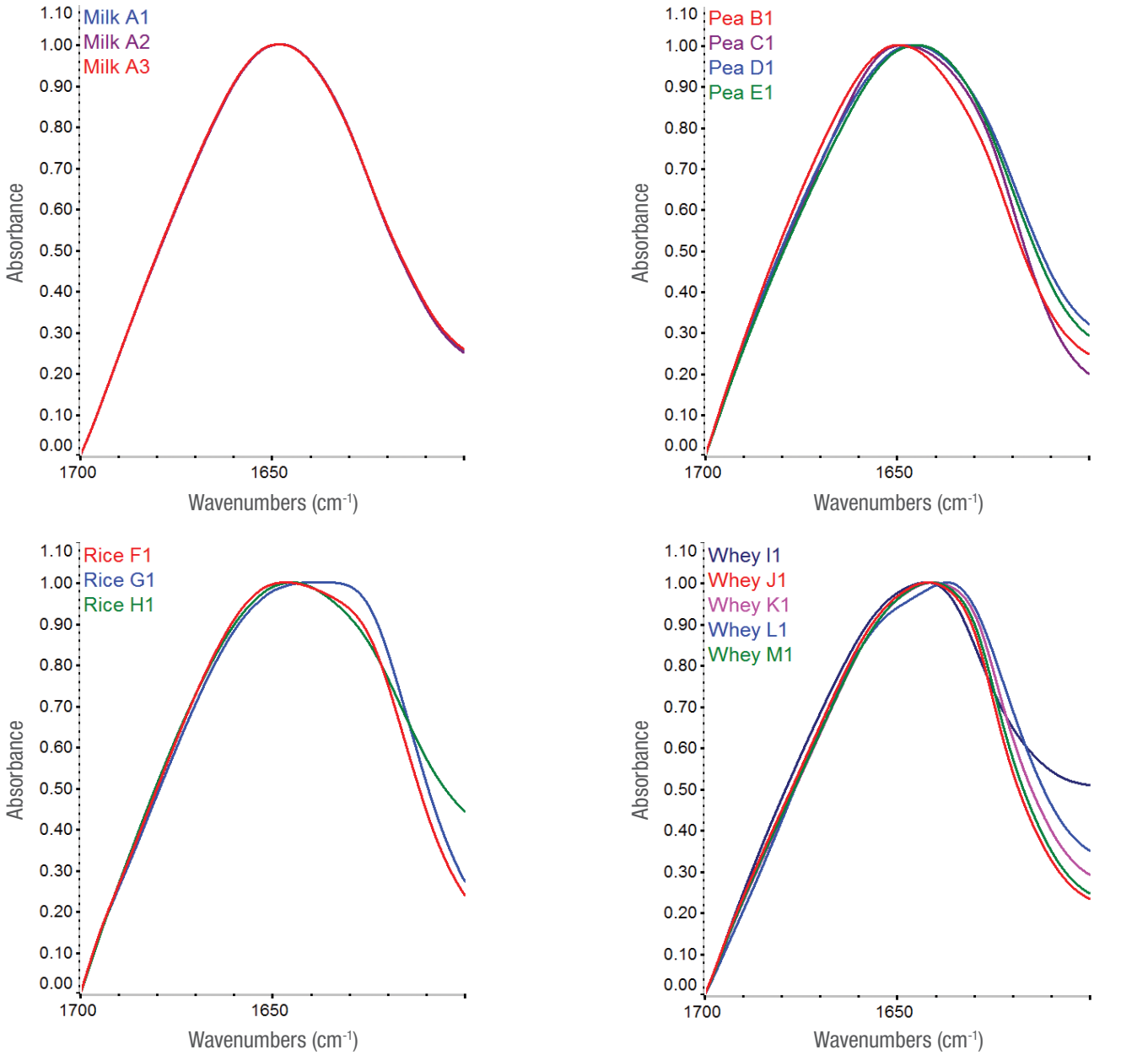

Figure 3: Full-scale, ATR-corrected spectra of protein powders in the amide I spectral region (1700-1600 cm^{-1}) showing each protein type with a spectrum from each vendor source. Since only one milk powder vendor sample was available, this plot includes 3 different lots.

The results of the corresponding PCA scores are shown in Figure 4. The general grouping of different protein types is similar to the full-range PCA scores plot of Figure 2. Each protein type congregates in its own domain, but the grouping is slightly tighter than in the full-range PCA model shown in Figure 2. This observation confirms that the variations manifested in Figure 2 indeed include both protein and non-protein contributions. A case in point is the whey proteins from vendor L. In the full-range PCA model (Figure 2), data points from vendor L (L1 and L2) are distant from the cluster that includes vendors J, K, and M, but much closer in the current model (Figure 4). It is reasonable to infer that the whey protein powders from vendor L differ from those from vendors J, K, and M primarily in the non-protein content. In contrast, products from vendor I remain distinctly different than the other whey protein powders in both PCA models, suggesting that the difference between sample I and the rest of whey proteins is, at least in part, due to the difference in protein conformation. This observation is corroborated by Figure 3, where the trace I1 is clearly different than the rest of the whey proteins.

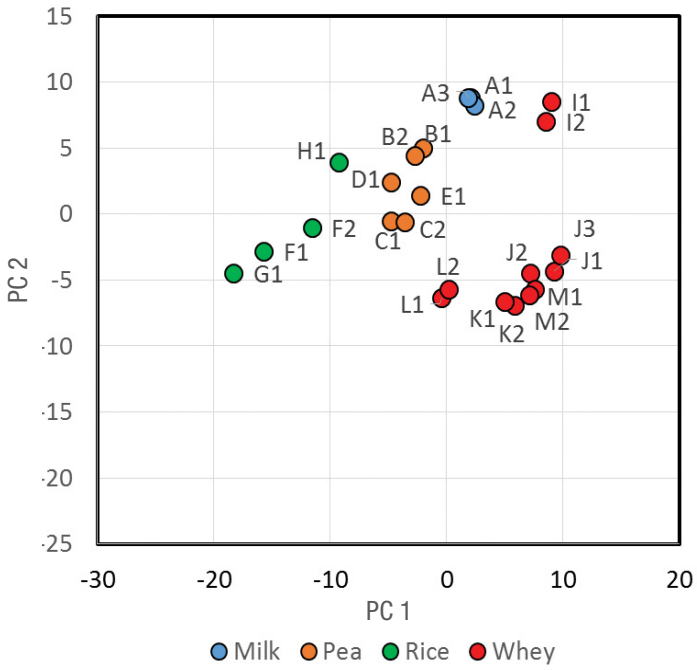

Figure 4: Principal component scores plot from protein powder samples using the amide I (1700-1600 cm^{-1}) region.

Conclusions

This application note demonstrates that FTIR spectroscopy combined with principal component analysis is an effective tool in the classification and discrimination of different food protein powders. Using the overall mid-IR spectral range, vendor formulation differences due to non-protein components can be readily seen and used as the basis for classification and discrimination. The scores plot of the PCA model based only on the amide I region keenly reflects the differences in protein secondary structure. While less susceptible to non-protein variation, the amide region-based PCA model still effectively classified each protein type and discriminated products from different vendors. Both models allow FTIR to classify and discriminate protein powders based on product type as well as supplier source and repeatability. The experiments are simple and straightforward and require no sample preparation. The methodology can be readily adopted by many food manufacturers for incoming material inspections as well as QA/QC.

References

1. Elliott, A., Ambrose, E. J. *Structure of synthetic polypeptides*, Nature (1950) 165, 921-922.
2. Byler, D.M., Susi, H. *Examination of the secondary structure of proteins by deconvolved FTIR spectra*, Biopolymers (1986) 25, 469-487.
3. Barth, A. *Infrared spectroscopy of proteins*, Biochim. Biophys. Acta (2007) 1767, 1073-1101.
4. Suja Sukumaran, *Protein secondary structure elucidation using FTIR spectroscopy*, Thermo Fisher Scientific Application Note AN52985.
5. Jackson, M., Mantsch, H.H. *The use and misuse of FTIR spectroscopy in the determination of protein structure*, Crit. Rev. Biochem. Mol. Biol. (1995) 30, 95-120.
6. Zeeshan, F., Ta2assum M., Jorgensen, L., and Medlicott, N. *Attenuated Total Reflection Fourier Transform Infrared (ATR FTIR) Spectroscopy as an Analytical Method to Investigate the Secondary Structure of a Model Protein Embedded in Solid Lipid Matrices*, Applied Spectroscopy (2018) Vol.72(2) 268-279.

Find out more at **thermofisher.com/is50**

ThermoFisher SCIENTIFIC

AN53037_E 08/18M