Analysis of biofluids in aqueous environment based on mid-infrared spectroscopy

Heinz Fabian
Peter Lasch
Dieter Naumann
Robert Koch-Institute
P13, Nordufer 20
13353 Berlin, Germany
E-mail: naumannd@rki.de

Abstract. In this study we describe a semiautomatic Fourier transform infrared spectroscopic methodology for the analysis of liquid serum samples, which combines simple sample introduction with high sample throughput. The applicability of this new infrared technology to the analysis of liquid serum samples from a cohort of cattle naturally infected with bovine spongiform encephalopathy and from controls was explored in comparison to the conventional approach based on transmission infrared spectroscopy of dried serum films. Artifical neural network analysis of the infrared data was performed to differentiate between bovine spongiform encephalopathy-negative controls and animals in the late stage of the disease. After training of artifical neural network classifiers, infrared spectra of sera from an independent external validation data set were analyzed. In this way, sensitivities between 90 and 96% and specificities between 84 and 92% were achieved, respectively, depending upon the strategy of data collection and data analysis. Based on these results, the advantages and limitations of the liquid sample technique and the dried film approach for routine analysis of biofluids are discussed. [DOI: 10.1117/1.1917844]

Keywords: infrared spectroscopy; biofluids; blood serum; bovine spongiform encephalopathy; artifical neural networks.

Paper 04172 received Aug. 31, 2004; revised manuscript received Oct. 1, 2004; accepted for publication Oct. 1, 2004; published online May 24, 2005.

1 Introduction

Biological fluids (blood, serum, synovial fluid, amniotic fluid) are ideal candidates for diagnostic analyses, because biochemical changes associated with numerous kinds of diseases influence their composition and since they can easily be isolated from the body. Therefore body fluids are routinely analyzed by clinical chemists to attain information on pathological processes. For many pathological conditions, however, there is no single analyte which alone is predictive for the presence or the stage of a disease. Instead, complex changes in fluid composition involving many chemical species are the more general case. Diagnostic approaches that are able to detect many or even all biochemical species at the same time are vibrational spectroscopic techniques, infrared (IR) and Raman. The vibrational spectrum of a biological fluid is the superposition of all vibrationally active components weighted with respect to their concentration. As a consequence of the complex spectra obtained by IR or Raman methodologies, sophisticated techniques of spectral data analysis are often required to extract diagnostic relevant information. Moreover, biological fluids contain mostly water. This is of little consequences for other techniques, but important for IR spectroscopy due to the high water absorptivities in the mid-IR region. Thus, IR spectra of biological fluids are dominated by water absorptions, with weak features from the dissolved species superimposed on the strong water absorption bands.

Drying the sample to form a film disk on an infrared-transparent sample carrier material suitable for absorbance/transmission measurements aids to overcome this problem, an approach often used in infrared clinical chemistry.[1–4] This procedure preserves the relative abundance of each IR active constituent in the fluid, but the sample integrity is lost. The latter prevents, e.g., to probe biochemical events ongoing in the original sample. In addition, the distribution of the constituents over the sample area of the dried film may vary. Moreover, control of the drying process and of stable environmental conditions is mandatory because the spectral features of many IR absorption bands may be affected by changes in the relative humidity of the film. Another sampling technique used to obtain IR spectra of biofluids is attenuated total reflection (ATR) spectroscopy.[5] For ATR measurements, the sample is prepared on the surface of an IR-transparent crystal and the IR beam is guided through the crystal in such a way that a few total reflections take place at the surface. Since the IR beam provides an evanescent wave entering into the medium of lower refractive index (i.e., the biofluid), the deposition of IR absorbing matter on the crystal surface causes the IR light to be partially absorbed. The penetration depth of the IR radiation in this arrangement is dependent on the wavelength and may be up to a few micrometers. The measured IR

Address all correspondence to Dieter Naumann, Robert Koch-Institute, P13, Nordufer 20, Berlin 13353 Germany. Tel: 49-30-45472259, Fax: 49-30-45472606, E-mail: naumannd@rki.de

1083-3668/2005/$22.00 © 2005 SPIE

spectrum thus contains only information on a very thin layer of the sample that is in close proximity to the surface of the crystal. This allows a spectrum of a sample in water to be obtained relatively easily, without much interference from IR absorption of bulk water. Detrimental to practical use, surface adsorption of proteins to the ATR crystal, however, may significantly change structure and spectra of analytes. The amount of the adsorbed molecules varies widely depending on the specific protein.[6] Since the ATR technique samples preferentially the molecules close to the crystal surface, the contribution of those molecules to the absorbance measured may not be negligible. Moreover, data interpretation is more complex in ATR (e.g., due to the wavelength dependence of the penetration depth) than in transmission spectroscopy. Beyond that, ATR crystals are expensive and not easy to clean optical materials.

An alternative approach, that can be used to bring the strong water absorption into an absorbance range, which is low enough in the mid-infrared region that solute absorptions may be recovered by subtracting the spectrum of the aqueous medium, is to use transmission cells with optical path lengths of only 5–8 μm.[7,8] However, such short path lengths limit the intensities of the IR bands of interest. Consequently, relatively high sample concentrations are required for the measurement. Blood serum samples represent highly concentrated solutions of biomolecules and, thus, are well suited for IR transmission spectroscopy. Here, we present a semiautomatic analytical technique, which allows to carry out transmission IR spectroscopy of serum (and other biofluids) in aqueous environment. To demonstrate the performance of this approach, we recorded IR spectra of liquid bovine serum samples from a cohort of bovine spongiform encephalopathy (BSE)-positive cattle and from BSE-negative control cattle. In parallel, spectra of the identical sera were collected by using the dry film technique. The infrared data were analyzed by artifical neural network (ANN) methodologies in order to differentiate between samples originating from BSE-positive and BSE-negative cattle. The advantages and limitations of the liquid and the dried film approach are discussed in the context of a mid-IR spectroscopy-based routine method for biofluid analysis.

2 Materials and Methods

2.1 *Sample Preparation*

Serum samples from cattle were obtained from the Veterinary Laboratory Agency (VLA), Weybridge, U.K., and from BSE-negative cattle from the Federal Animal Research Center, Braunschweig, Germany. All BSE-positive samples were obtained by the VLA and originated from cattle that had been tested to be BSE-positive by histopathological examination or immunocytochemistry methods. Serum was stored at −80 °C. Sample storage and handling was done in a L3[xx] laboratory. After thawing, the serum samples were warmed to room temperature, and filtered through disposable high performance liquid chromatography (HPLC) sample microfilters of 0.5 μm porosity to remove particles prior to injection into the flow-through IR cuvette. In parallel, a small portion of each filtered serum sample was stored at −80 °C for subsequent measurements as dried film samples. For this purpose, after rethawing ~2.6 μl of each sample were pipetted onto each of three different ZnSe windows of a multisample cuvette and allowed to air dry for 15 min at 37 °C. The cuvette was hermetically sealed with a KBr window and then transferred to the Fourier transform infrared (FTIR) spectrometer. Each multisample cuvette carried up to 15 samples.[9]

2.2 *FTIR Spectroscopy*

Infrared spectra were recorded with an IFS 28/B FTIR spectrometer (Bruker Optik GmbH, Ettlingen, Germany). For measurements of the liquid samples, a liquid-nitrogen cooled broadband mercury cadmium telluride (MCT) detector was utilized that allowed, in combination with an optical cutoff filter, to obtain IR spectra with very high signal-to-noise ratio in the region between 1000 and 3100 cm^{-1}. The measurements of the dried film spectra were performed in transmission using a deuterated triglycine sulfate (DTGS) detector, which in contrast to a MCT detector can operate without liquid nitrogen cooling. Moreover, DTGS detectors have the advantage of an extended linear range (i.e., are linear at higher absorbance values than MCT detectors) and thus may be the detector of choice for dried film measurements under certain conditions. For each sample, 128 interferograms were coadded and Fourier transformed to yield spectra with a nominal resolution of 4 cm^{-1}. Typically, three replicate samples were measured for each serum specimen. The sample chambers and optics were purged with dry air in order to keep the water vapor level constantly low.

2.3 *Data Pre-Processing and Analysis*

Rough data pre-processing of the IR spectra (calculation of derivatives or mean spectra) was performed applying routines of the OPUS IR software package (Bruker Optik GmbH). Data pre-processing for ANN analysis, network training, and classification was performed by using Synthon's NeuroDeveloper software package specifically designed for ANN analysis of infrared spectral data (Synthon GmbH, Heidelberg, Germany). Data pre-processing included tests for spectral quality, calculation of second derivatives (Savitzky-Golay algorithm, five smoothing points), vector normalization in the spectral region 2820–2985 cm^{-1}, and averaging three adjacent data points. After defining spectral windows, the spectral information from these regions was used for feature selection (for details see Ref. 10). For ANN analysis the data were split into subsets for teaching, internal and external validation. Spectra of the three independent measurements on each sample were always grouped such that they appeared in only one of the corresponding subsets. The calculation of the so-called "differentiation indices" (D values) and the signal-to-noise (S/N) ratios was carried out by using MATLAB-based software developed in-house (The MathWorks, Natick, MA, USA) on the basis of the second derivatives (Savitzky-Golay algorithm, five smoothing points) of the IR spectra.

2.4 *Mass Spectrometry*

For matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, the sera were diluted 100× by water, 1:1 mixed with matrix solution (50 mM sinapinic acid in acetonitrile/trifluoroacetic acid mixture), and then spotted on a stainless steel target plate supplied by Bruker Daltonics. The mass spectra were acquired using an

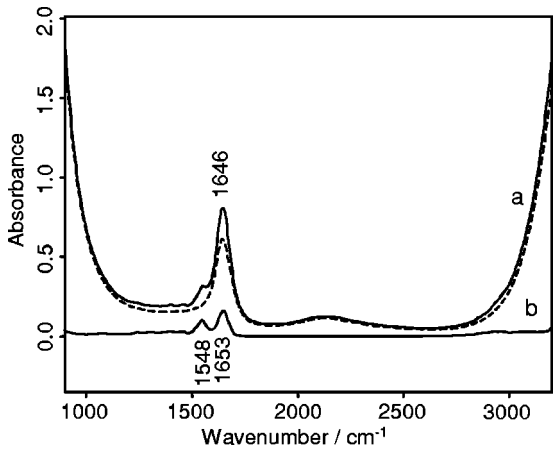

Fig. 1 (a) IR spectrum of bovine blood serum (solid line) and of water (dashed line). (b) Difference spectrum generated by subtraction of the spectrum of water from the spectrum of serum.

Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany). Typically, each spectrum represents an average of 100 laser shots.

3 Results and Discussion

3.1 *An Apparatus for Measurements of Liquid Samples with Simple Sample Introduction and High Throughput*

Flow-through cells with Luer-lock fittings, CaF_2 windows, and spacers covering path lengths of 6–10 μm are often used for IR transmission measurements of aqueous samples. These cells are commercially available, but have some disadvantages in their practical use. They are typically difficult to clean and to fill without accidentally trapping air bubbles at the windows. To circumvent these problems, we used custom-made IR cells of a different design, which consisted of a flat cover disk (made of CaF_2) and a second disk of the same material, with the center deepened to form a recessed parallel surface surrounded by a trough.[7,8] To seal the cell, the cover disk was then pressed onto the sample disk which contained the solution or suspension. Depending upon the diameter and the depth of the recessed surface of the window (i.e., the path length of the cell), only a few microliters were required to fill the cell. This type of cell can very easily be filled with a sample solution, assembled and disassembled, and cleaned between measurements. Figure 1 shows a representative IR spectrum of bovine serum measured in such a custom-made IR cell with a path length of ~6 μm. Comparison with the spectrum of water measured in a matched second cell of slightly lower path length, which takes into account the slightly lower water concentration in the serum sample, demonstrates that water contributes significantly to the IR spectrum of serum. The major water absorptions in the mid-IR region are due to the O-H stretching vibration centered near 3400 cm^{-1} and the H-O-H bending vibration near 1645 cm^{-1}. The intense absorption below 1000 cm^{-1} arises primarily from the CaF_2 windows of the IR cell. The IR spectrum of the sample after digital subtraction of water reveals the characteristic infrared bands of the serum sample. The major absorption features of the spectrum are attributed to proteins (e.g., amide I centered at 1653 cm^{-1} and amide II at 1548 cm^{-1}). Because IR spectroscopy is an averaging technique, the amide absorptions in the serum spectra contain contributions from all the proteins of the serum. As such, it is very difficult, if not impossible, to assign specific amide I absorptions to characteristic secondary structures of any particular protein within the mixture. However, from the frequency maxima of the amide I and II bands it is possible to infer that the serum spectrum shown in Fig. 1 is dominated by proteins with mostly α-helical conformation.[7,11]

Although the use of the custom-made IR cells allowed us to obtain high-quality IR spectra of biofluids, this approach is not very efficient for large-scale measurements of samples. Thus, we have designed an apparatus that combines simple sample introduction with high sample throughput for large scale series measurements of different liquid samples. A schematic diagram of the apparatus is shown in Fig. 2. The commercially available principle elements of the system are a HPLC pump model 1100 (Agilent Technologies, Palo Alto, CA, USA), an Agilent 1100 series manual injector with a Rheodyne 7-port sample injection valve (model 7725, Rheodyne LP, Cotati, CA, USA), a cheminert valve model C2 1/16″ (Valco International, Switzerland), and a flow-through cell AquaSpec AS 1100M (micro-biolytics, Freiburg, Germany). In front of the IR cell a bulk-head filter with a removable 2 μm screen (model ZBUFR1C 1/16″, Valco International) is mounted, which allowed us to remove possible residual particulate matter from the samples and, thus, to minimize the danger of damage of the expensive hardware (AquaSpec cell). The AquaSpec cell is a specially designed microfabricated flow-through cell with a spacer from a sealant material of defined elasticity, which ensures high-pressure stability and fast relaxation for a constant effective sample thickness.[12,13] Both, the form and the thickness of the spacer can be customized to fit the experimental requirements. Typical path lengths for our experiments were in the order of 6–8.5 μm. Valve switching, triggering of sample injection, and control of HPLC pump were synchronized by means of the microcomputer of the pump. To start a measurement, rinsing of the flow cell with distilled water is stopped, and a background spectrum of the cell filled with water is recorded. At the same time, the sample loop of the injection port is filled manually with sample [Fig. 2(A)]. To bring the sample solution into the flow, the injection port is switched manually into the injection position [Fig. 2(B)]. After a defined time, a trigger signal is sent to the stop/flow valve to stop the flow through the IR cell after proper filling of the IR cell with the sample [Fig. 2(C)]. After releasing the pressure, the spectrum of the sample solution is measured. During the measurement, the sample injection port can be switched back to its initial position and refilled for the next measurement. In order to rinse the system with water, the stop/flow valve is switched back to its initial position after the measurement.

3.2 *IR Spectroscopy of Blood Serum Samples*

Typical mid-IR spectra of two serum samples measured as a liquid (traces a and a′) and as a film dried onto a ZnSe window (traces b and b′), respectively, are shown in Fig. 3. Serum, derived from plasma with clotting factors removed, contains 60–80 mg of protein/ml in addition to small molecules including lipids, amino acids, sugars, and salts.[14] As a conse-

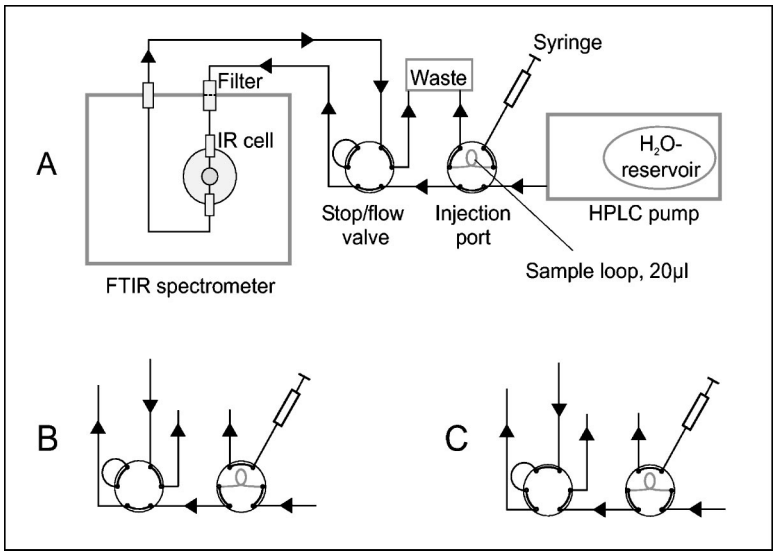

Fig. 2 Schematic diagram of the apparatus developed for series measurements of biofluids. (A) Injection port in fill position, stop/flow valve in flow position. (B) Injection port in inject position, sample is transported into the flow cell. (C) Stop/flow valve in stop position, sample is measured.

quence, the IR spectra of sera are dominated by the absorptions of the protein constituents, with spectral features of other molecular species superimposed upon the protein absorption profile. The prominent band in the range 1600–1700 cm^{-1} (amide I) arises primarily from the C═O stretching vibrations of all the amide groups. This band is a sensitive indicator of protein secondary structure.[7,11] The amide II (N–H bending and C–N stretching vibration, which occurs between 1520 and 1560 cm^{-1}) and amide III (C–N stretching and N–H bending, which occurs in the range 1220–1320 cm^{-1}) may also provide conformational information. In addition to these bands arising from the protein backbone, bands due to amino acid side chains are also seen, for example at 1400 cm^{-1} (symmetric stretching vibration of carboxylate groups). The range 1000–1200 cm^{-1} contains infrared bands associated with stretching vibrations of carbon–carbon and carbon–oxygen single bonds, such as those present in sugars or carbohydrates. Lipids also give rise to a number of absorptions in IR spectra. The most intense of these absorptions are found in the range 2800–3000 cm^{-1}, attributed to asymmetric and symmetric stretching vibrations of CH_3 and CH_2 groups of the acyl chains. The latter region also contains the corresponding stretching vibrations from amino acid side chains of the proteins.

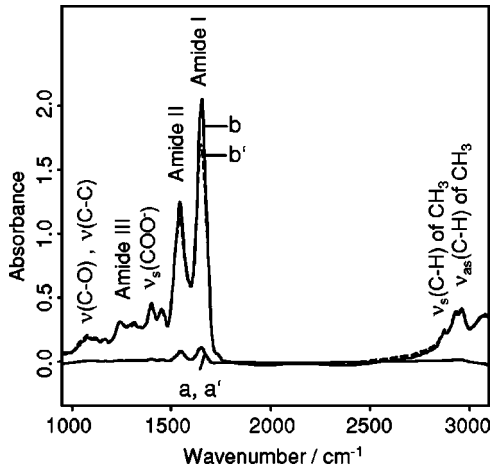

Fig. 3 Comparison of IR spectra of two blood serum samples measured as a liquid (a and a′, bottom traces) and as a film dried onto a ZnSe window (b and b′, top traces).

Figure 3 also illustrates that the infrared features of ~2.6 μl serum measured as a film dried onto an IR-transparent window are ~20× more intense than those of the same serum measured as liquid in a flow-through cuvette of ~8 μm path length. Moreover, varying absorbance values indicate variations in film thickness between the two dried samples (see traces b and b′ in Fig. 3). In case of the liquid technique these variations are much less pronounced, as indicated by almost identical traces a and a′ in Fig. 3. The liquid technique allowed us to obtain highly reproducible IR spectra of serum, such as that illustrated in Fig. 4(A) for three independent measurements of a given serum sample. Only minor differences between these measurements are indicated by the corresponding infrared difference spectra [Fig. 4(B)]. The negative features above 3000 cm^{-1} and in the amide I region of the serum spectra after subtraction of water [Fig. 4(A)] indicate a slight overcompensation of the water absorption bands. This is due to the identical path length of the IR cell used to obtain the spectra of water and serum, which prevents a perfect compensation of the water absorption bands. The corresponding negative spectral features are not seen when using matched IR cells of slightly different path lengths [see Fig. 1(b)].

For quantification of the reproducibility, the so-called ''differentiation indices'' D of pairs out of three independent measurements of a serum were calculated by the equation $D=(1-\alpha)1000$, where α is the so-called Pearson's product-momentum correlation coefficient.[15] The calculated D values for selected spectral regions of the IR spectra of the sera are shown in Table 1, which illustrates that the reproducibility varies as a function of the spectral region selected. The analysis of the amide I region revealed very low D values of 0.3

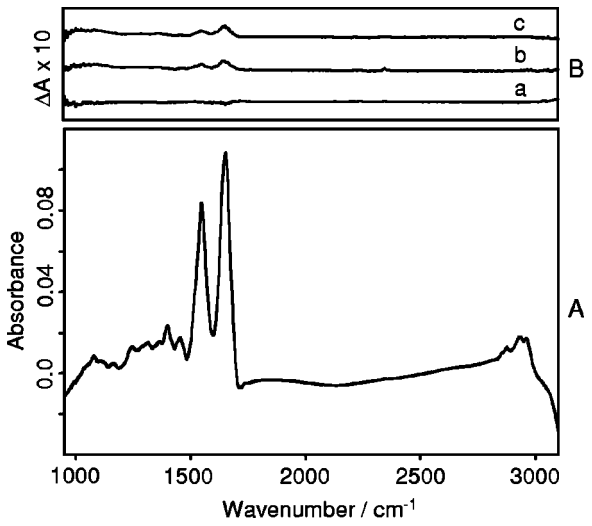

Fig. 4 IR spectra showing the reproducibility of the liquid sample technique. (A) Overlaid IR spectra of three independent measurements of a liquid serum sample in a 8 μm flow-through cell. (B) IR difference spectra: (a) spectrum 1 minus spectrum 2; (b) spectrum 1 minus spectrum 3; and (c) spectrum 2 minus spectrum 3. Note that the absorbance scale for the difference spectra in part B was expanded by a factor of 10 compared to the scale in part A.

(i.e., almost identical spectra) for the liquid spectra but almost 30× higher *D* values in the case of the spectra obtained by the dry film technique. This is most likely due to variations of the dry film thickness from sample to sample, which occasionally yield amide I absorbance values even above 2, which in turn may create problems in quantitative analysis due to nonlinearity of the DTGS detector at very high absorbance values. Just the opposite was observed for the spectral region 2800–3100 cm^{-1}. In this region, the reproducibility of the weak spectral features of the liquid serum samples is limited by the broad and very strong O–H water absorption centered near 3400 cm^{-1}. Comparable and very low *D* values for both sampling techniques were obtained when analyzing the information contained in the spectral range 1050–1500 cm^{-1}.

Table 1 Quantification of reproducibility: Differentiation indices (*D* values) of the IR spectra of the sera measured as liquid and as dry film, respectively. Average *D* values and the corresponding standard deviations have been calculated from the second derivatives of the spectra in the amide I region (1600–1700 cm^{-1}), the spectral region 1050–1500 cm^{-1}, and the CH stretching region (2800–3100 cm^{-1}).

Spectral window (cm^{-1})	Liquid samples (*D* values)*	Dry samples (*D* values)*
1600–1700	0.33±0.09	8.87±8.8
1050–1500	1.31±0.16	1.62±0.81
2800–3100	10.00±2.14	0.69±0.46

* $D=(1-\alpha)1000$, where α=Pearson's product-momentum correlation coefficient.

Another issue in this context is the S/N ratio of the spectra obtained by the two different approaches. In order to be able to estimate this quantity, we have calculated the S/N ratio of the spectra of sera obtained by the liquid and the dry film technique. The data illustrate that the S/N ratio for the dried film spectra was similar (in the amide I region) or approximately 4× higher (in the region 1350–1450 cm^{-1}) compared to those of the liquid spectra (see rows four and seven in Table 2), despite the fact that the peak intensities of the corresponding IR bands were more than one order of magnitude higher in the case of the film spectra. Similar S/N ratios of the spectra obtained by the two techniques indicate that the spectra of the liquid samples contained significantly less absolute noise than those of the dried films (see rows two and five in Table 2). The IR spectra of the liquid samples were recorded using a liquid-nitrogen cooled MCT detector which, together with optical filters, typically allowed us to obtain spectra of a much higher quality than those recorded with a DTGS detector.

Table 2 Quantification of noise and signal-to-noise ratio of the IR spectra of sera measured as liquid and as dry film, respectively. Mean values plus/minus standard deviations are given: (i) noise, (ii) amide I peak intensity, (iii) maximum absorption in the region 1050–1500 cm^{-1} ($\nu_s COO^-$ at 1400 cm^{-1}), and (iv) S/N ratio. The amide I peak intensities and the peak intensities at 1400 cm^{-1} were taken from the absorbance spectra. To estimate the peak-to-peak noise in the spectra of the liquid serum samples placed in the AquaSpec cell, we collected and processed the single beam "background" spectra and the single beam sample spectra from one and the same sample. The same procedure was applied for the dried samples deposited onto the ZnSe window of a multisample cuvette.

	Liquid samples	Dry samples
Noise, region 1600–1700 cm^{-1}	$(1.140\pm0.227)\times10^{-4}$	$(1.811\pm1.420)\times10^{-3}$
Amide I peak intensity	0.110±0.013	1.500±0.353
S/N ratio, amide I region	965±223	828±677
Noise, region 1350–1450 cm^{-1}	$(1.987\pm0.321)\times10^{-5}$	$(9.345\pm2.160)\times10^{-5}$
Peak intensity at 1400 cm^{-1}	0.0236±0.0028	0.378±0.092
S/N ratio at 1400 cm^{-1}	1192±238	4047±1355

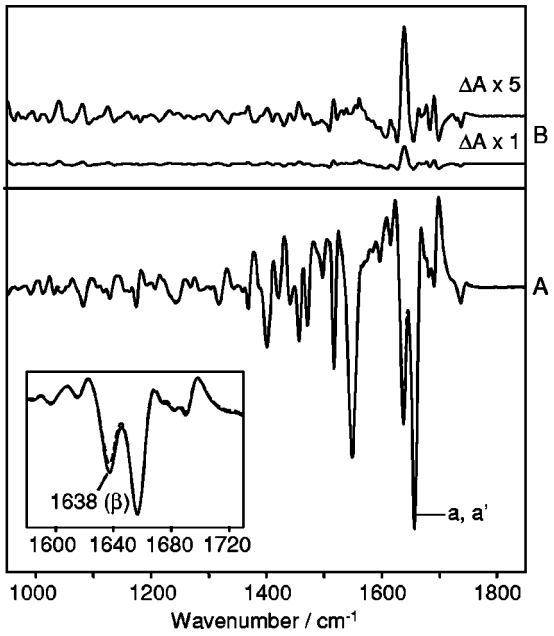

Fig. 5 (A) Comparison of mean IR spectra (second derivatives) of 115 BSE-positive (solid line, a) and 108 BSE-negative (dashed line, a′) liquid samples of sera, each sample measured in triplicate. Inset: comparison of the expanded amide I region. (B) IR difference spectra (BSE-negative minus BSE-positive) of the second derivatives shown in (A). In the top trace, the absorbance scale was expanded by a factor of 5.

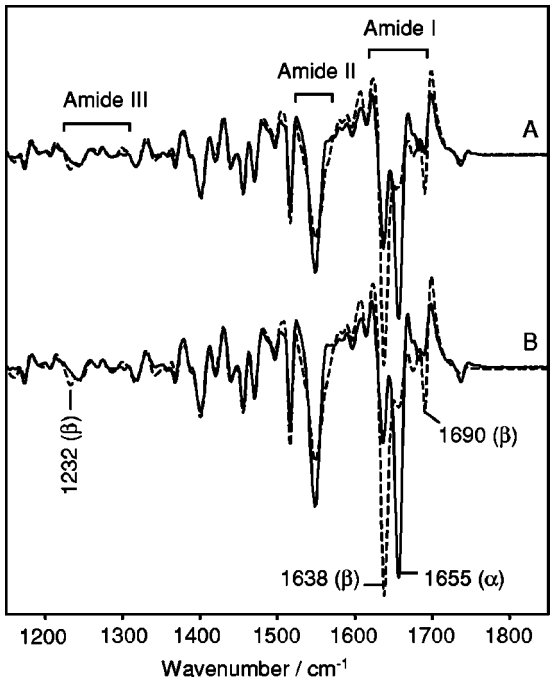

Fig. 6 Comparison of the IR spectra (second derivatives) of two sera (solid and dashed lines) originating from (A) BSE-negative and (B) BSE-positive cattle. Characteristic spectral features of the proteins are marked.

3.3 *IR Spectra of Bovine Serum from Cattle Infected with BSE and from Controls*

The mean second derivative IR spectra obtained from the liquid sera of 115 BSE-positive animals and of 108 BSE-negative controls, respectively, are shown in Fig. 5. We calculated the second derivatives to compensate for small baseline drifts within the spectra and for minor differences in the solute-to-water ratio of the samples [Fig. 5(A)]. Only small spectral differences were observed, most pronounced in the amide I region. A slightly more intense band centered at 1638 cm^{-1} is observed in the mean spectrum of the BSE-positive sera [see inset of Fig. 5(A)], which results in a positive feature in the difference spectrum (BSE-negative minus BSE-positive) [Fig. 5(B)]. Beside the amide I/II region, minor spectral differences were also observed in the range 1000–1500 cm^{-1}. An amide I band centered at 1638 cm^{-1} indicates the presence of β-sheet structures.[7,11] Thus, a slightly more intense band at 1638 cm^{-1} suggests that there should be more β-sheet proteins present in the sera of the BSE-positive animals.

The crucial step in transmission and manifestation of the prion diseases is the conversion of benign cellular prion protein exhibiting mainly an α-helical structure to the pathogenic β-sheet rich oligomeric isoform.[16] The IR feature at 1638 cm^{-1}, however, cannot be associated with the presence of misfolded prion proteins in serum, because the prion protein (which is not detectable in serum by the most sensitive western blot assays available to date) would be well below the detection limit of the IR technique. Blood serum is a complex body fluid that contains various proteins ranging in concentration over at least 9 orders of magnitude. The major protein constituents of serum include albumin ($c \sim 35-50$ mg/ml), immunoglobulins, transferrin, haptoglobulin, and lipoproteins.[14] In addition to these major constituents, serum also contains many other proteins that are synthesized and secreted, shed, or lost from cells and tissues throughout the body. Up to 10^4 different proteins may be commonly present in serum, most of which present at fairly low relative abundance. The prion protein, if present at all in serum, would certainly be one of those very low abundance proteins. As IR spectroscopy is an averaging technique, the amide absorptions in serum represent the mean of the amide absorption characteristics arising from all proteins in the serum, weighted according to the concentration of each protein. It is therefore impossible to attribute particular amide absorptions to individual low-abundant proteins in serum. Thus, the small spectral differences between the mean IR spectra of sera of BSE-positive and BSE-negative are certainly not directly associated with the presence of the prion protein. Rather, it seems that a majority of serum constituents is slightly changed as suggested by the complex spectral difference pattern shown in Fig. 5(B). Moreover, a detailed inspection of the spectra of each individual serum revealed large variations in the amide I/II region from sample to sample. Extreme cases are shown in Fig. 6. For example, we obtained spectra of sera dominated by β-sheet proteins (band components at 1638 and 1690 cm^{-1}, dashed traces in A and B). At the same time, we detected sera whose spectra were dominated by α-helical proteins (band component at 1655 cm^{-1}, solid traces in A and B).

In order to understand the rationale behind the variations in the protein composition from serum to serum, a number of BSE-negative serum samples were examined by classical clinical serum protein analysis and by MALDI-TOF mass

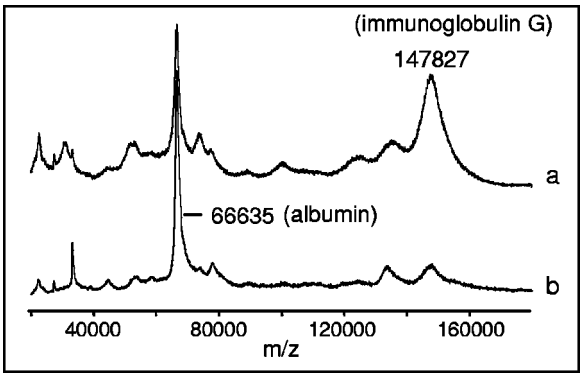

Fig. 7 MALDI-TOF mass spectra in the high mass range (m/z 20 000–180 000) recorded from two BSE-negative sera. (A) Spectrum of serum a, whose IR spectrum [dashed line in Fig. 6(A)] was dominated by β-sheet structure features. (B) Spectrum of serum b, which revealed pronounced α-helical features in its IR spectrum [solid line in Fig. 6(A)].

spectrometry. The MALDI-TOF mass spectra of two typical sera, to give an example, revealed numerous spectral differences, most pronounced in the high mass range around m/z 66 635 and 147 827 (Fig. 7). These two mass peaks have previously been assigned to serum albumin (m/z 66 635) and immunoglobulin G (m/z 147 827).[17] Albumin, whose structure is almost all α-helical, is the major serum protein, whereas immunoglobulins (β-sheet rich proteins) belong to the high abundance proteins in serum.[14] The intense mass peak around m/z 147 827 in spectrum a indicates that this serum sample contains much more immunoglobulins than the serum sample of spectrum b. In other words, the ratio of β-sheet to α-helical proteins is significantly different between the two sera. The increased amount of β-sheet proteins in serum a of Fig. 7 correlates with the pronounced β-sheet features in its IR spectrum [dashed trace in Fig. 6(A)]. This conclusion was confirmed by serum protein electrophoresis (data not shown). The differences observed in the amide I region between the mean spectra of the liquid serum samples of BSE-positive and BSE-negative animals, together with characteristic differences in the ranges 1000–1500 and 2800–3000 cm^{-1}, were also observed in the mean IR difference spectra of the dried samples of the corresponding sera [compare Figs. 8(A) and (B)].

3.4 *ANN Analysis of the Data*

Due to the complex pattern of both positive and negative features across a wide spectral range, ANN analysis, a supervised pattern recognition method, was used to differentiate between samples originating from BSE-positive and BSE-negative cattle based on their IR spectra. Supervised methods make use of the fact that clinical information concerning the samples from which the spectra were obtained is available. This information is used to teach the classifier in order to predict the class identity of unknown spectra afterwards. The quality of the teaching process can be assessed on-line by classifying data of a so-called internal validation subset. The spectra of those data sets are not used for teaching but are continuously analyzed during teaching. When the teaching process is finished, the classifier is challenged by an third independent data set for external validation. It is important to note that this latter subset of spectra (''unknown spectra'') is kept totally separate from the teaching and internal validation procedure.

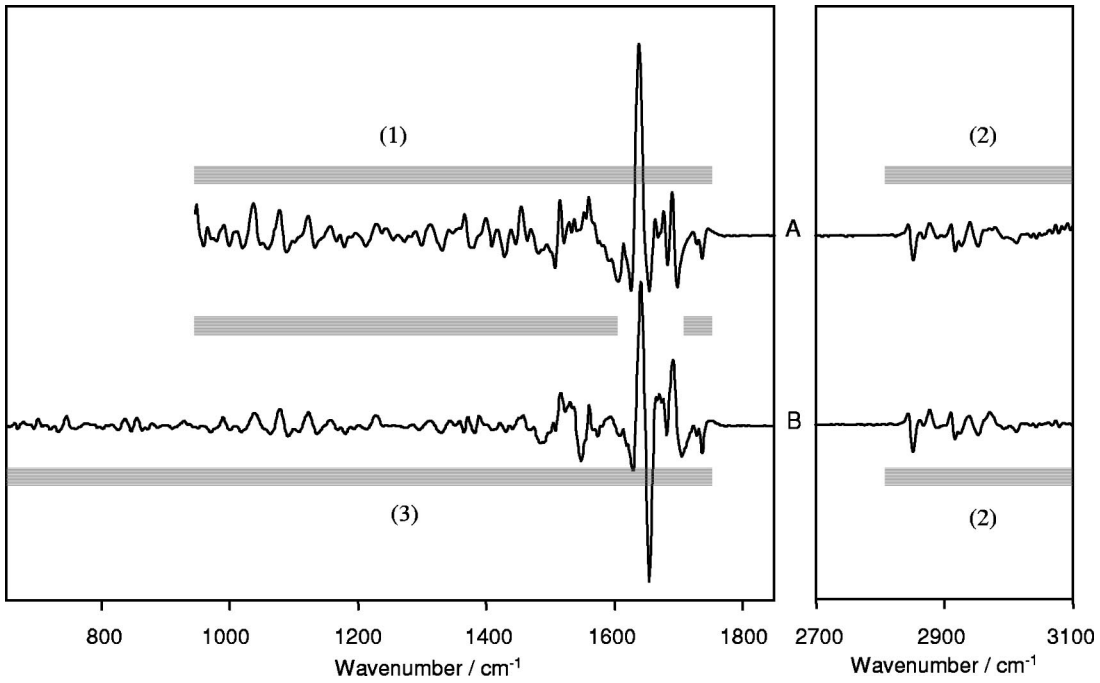

Fig. 8 Comparison of second derivative difference spectra (BSE-negative minus BSE-positive) of the mean spectra of BSE-positive and BSE-negative samples. (A) Based on spectra obtained from liquid samples; (B) Based on spectra of dried samples. The different spectral windows choosen for ANN classification are indicated in gray.

Table 3 Number of samples used for teaching and validating the ANN classifiers.

	Teaching data set	Internal validation data set	External validation data set
BSE negative	77	11	21
BSE positive	76	13	26

For ANN analysis, spectra of the data sets (see Table 3; triplicate measurements of both liquid serum samples and of dried films of the sera) were pre-processed as already outlined: second derivative spectra were vector normalized in the region of 2820–2985 cm^{-1}. Three combinations of spectral windows, indicated in Fig. 8, were then chosen for further network analysis. (i) The range 950–1750 cm^{-1} (1) in combination with the range 2800–3100 cm^{-1} (2). (ii) Regions (1) and (2), but excluding the amide I range (i.e., 1600–1700 cm^{-1}). (iii) An extended wave number range 650–1750 cm^{-1} (region 3), in combination with the range 2800–3100 cm^{-1}. The latter combination of the spectral windows was only available for dried samples due to the optical window material used (ZnSe instead of CaF_2 for liquid samples). The spectral information from these three combinations of spectral windows was then subjected to feature selection. Feature selection was based on the calculation of the covariance of the spectral data points.[10,18] After averaging three adjacent data points, the best 80 discriminative spectral features were obtained for the teaching data sets. Although obtained from the teaching subsets only, these features were used also for the data sets of internal and external validation. The features selected as input for classification between BSE-positive and BSE-negative samples are illustrated in Fig. 9, suggesting that a large number of spectral features is suitable for classification. To teach the ANNs, spectra of 76 BSE-positive and 77 BSE-negative animals were used (Table 3). The most accurate classification results were obtained for feed-forward network architectures of 80 input neurons, 3 neurons in the hidden layer, and 2 output neurons. The quality of teaching was assessed by the classification results of spectra from the internal validation data set originating from 13 BSE-positive and 11 BSE-negative animals (see Table 3). After teaching, the data of the external classification subsets were analyzed by ANN classifiers. As shown in Table 3 the external validation data set consisted of spectra from 26 BSE-positive and 21 BSE-negative animals. The sensitivities (number of correctly classified BSE-positive samples divided by the total number of BSE-positive samples) and the specificities (number of correctly classified BSE-negative samples divided by the total number of BSE-negative samples) are given in Table 4. They indicate, that more than 90% of the BSE-positive samples and ~90% of the BSE-negative samples were identified correctly. Moreover, the similar classification results suggest that the specific methodology of data collection has no major impact on the differentiation between samples originating from BSE-positive and BSE-negative animals. It is also interesting to note that the serum samples, which were mis-classified by the liquid and the dry film approach, were mostly the same. This result suggests the existence of specific biochemical characteristics of those serum samples, so far not represented by our spectral data set used for network teaching.

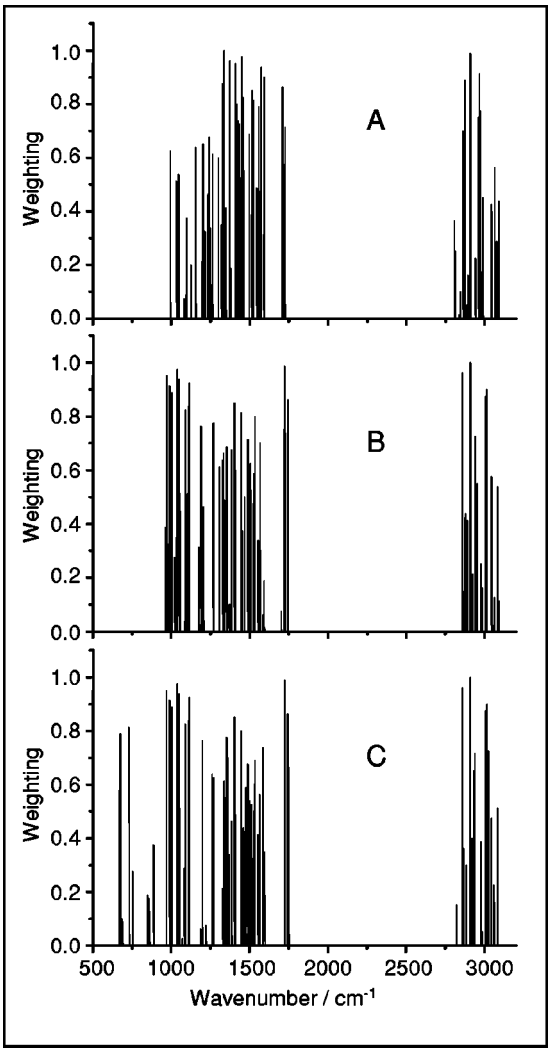

Fig. 9 Spectral features selected for ANN classification of BSE-positive and BSE-negative samples, excluding the amide I region. (A) spectra of liquid samples; (B) spectra of dried samples; (C) spectra of dried samples with an extended low wave number range. The higher the weighting value, the more discriminative is the corresponding spectral feature.

These results correlate well with our data recently obtained by multivariate analysis of a larger data set of infrared spectra of dried films of unfiltered bovine sera.[10] The present results also suggest, that the evaluation of the spectral information contained in the amide I region does not improve the classification results. This is most likely due to the strong impact of concentration variations of the high-abundant serum proteins (albumin and immunoglobulins) in this spectral range. Typically, an increased amount of immunoglobulins in serum can be taken as an indicator of ongoing inflammatory processes, e.g., as the result of some kind of infectious diseases. Indeed, a number of sera in our study included sera from animals suffering from a variety of ''classical'' viral or bacterial infectious diseases.[10] Finally, it is interesting to note that the best sensitivity (see column 10 in Table 4) was obtained by ana-

Table 4 Sensitivity and specificity determined by external validation with 26 BSE-positive and 21 BSE-negative serum samples. Results of three neuronal net approaches (ANN I-III) are shown. The last row gives the average of the corresponding results. The spectral regions considered for analysis are: without amide I region (950–1600, 1700–1750, and 2800–3100 cm^{-1}); amide I region included (950–1750 and 2800–3100 cm^{-1}); fingerprint region included (650–1600, 1700–1750, and 2800–3100 cm^{-1}).

	Liquid technique				Dry film technique					
	Without amide I		With amide I		Without amide I		With amide I		With fingerprint	
ANN	SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC
I	96.2	85.7	92.3	95.8	92.3	90.5	88.5	95.8	95.8	92.3
II	92.3	81.0	92.3	85.7	92.3	95.8	88.5	90.5	96.2	90.5
III	88.5	85.7	84.6	95.8	92.3	90.5	96.2	90.5	96.2	90.5
Average	92.3	84.1	89.7	92.4	92.3	92.3	91.1	92.3	96.1	91.1

SENS=sensitivity (%) and SPEC=specificity (%).

lyzing the spectral information contained in the extended wave number range (3) between 650 and 1750 cm^{-1}, in combination with the CH stretching region (2800–3100 cm^{-1}). This suggests, that the so-called fingerprint region (below 1000 cm^{-1}) of the serum spectra also contained some information of diagnostic relevance [see also Fig. 9(C)].

4 Conclusions

We have shown that a semiautomatic experimental approach for IR measurements of liquid serum samples, which combines simple sample introduction with high sample throughput, allowed us to obtain high-quality infrared spectra of liquid serum samples with very high reproducibility. Artifical neural network analysis was employed to analyze serum samples from a cohort of cattle infected with bovine spongiform encephalopathy and from controls based on their infrared spectra. In parallel, spectra of dried serum samples were recorded and were analyzed in the same way. This comparative study yielded sensitivities between 90 and 96% and specificities between 84 and 92%, respectively.

Our results suggest that both the dried film approach and the liquid sample technique are similarly useful methods for the analysis serum samples, in particular, and for biofluids in general. The dry film technique, however, has some advantages in practice, as it requires minimal sample preparation, tiny amounts of sample (~10× less than the liquid technique), and only moderate technical expertise. Moreover, FTIR spectrometers with multisample cuvette accessories for high sample throughput measurements of dried samples are commercially available, whereas the liquid sample technique requires, in addition to the FTIR spectrometer, expensive hardware (e.g., the apparatus for sample injection and the AquaSpec flow cell). A principal disadvantage of the dried film technique is, however, the fact that the spectral features of many IR bands are sensitively affected by changes in the relative humidity of the film. This requires very careful control of environmental conditions during the measurement in order to avoid misinterpretation of spectral changes or incompatibilities of spectral data. In addition, the amide I region is often not available for quantitative analysis due to considerable sample heterogeneity, nonuniformity in the thickness of the dried sample, and total absorption. Here is the major strength of the liquid sample technique, which allows IR spectra to be obtained in the conformation-sensitive amide I region with unprecedented reproducibility. This advantage, however, could not come into play within the framework of the analysis of the IR spectra of the serum samples due to the major non-BSE associated spectral changes in the amide I region described above. Finally, the liquid sample approach will be of great advantage in studies aimed at probing biochemical events ongoing in the sample induced by internal or external factors.

Acknowledgments

The authors acknowledge the technical assistance of R. Puttkamer and S. Wolgast, and the help of F. Sokolowski in preparing Fig. 2. They are grateful to M. Bardsley, D. Mathews, S. Hawkins, M. Parker, and A. Ives (VLA), and U. Meyer (Federal Animal Research Center, Braunschweig) for the supply of serum samples. This research is supported by the German Bundesministerium für Bildung und Forschung (BMBF Grant No. 0312727).

References

1. J. Wang, M. Sowa, H. H. Mantsch, A. Bittner, and H. M. Heise, ''Comparison of different infrared measurement techniques in the clinical analysis of biofluids,'' *Trends Analyt. Chem.* **15**, 286–295 (1996).
2. R. A. Shaw, S. Kotowich, M. Leroux, and H. H. Mantsch, ''Multianalyte serum analysis using mid-infrared spectroscopy,'' *Ann. Clin. Biochem.* **35**, 624–632 (1998).
3. K.-Z. Liu and H. H. Mantsch, ''Simultaneous quantification from infrared spectra of glucose concentrations, lactate concentrations, and lecithin/sphingomyelin ratios in amniotic fluid,'' *Am. J. Obstet. Gynecol.* **180**, 696–702 (1999).
4. K.-Z. Liu, K. S. Tsang, C. K. Li, R. A. Shaw, and H. H. Mantsch, ''Infrared spectroscopic identification of β-thalassemia,'' *Clin. Chem.* **49**, 1125–1132 (2003).
5. N. J. Harrick, *Internal Reflection Spectroscopy*, Harrick Scientific Corporation, New York (1987).
6. E. Goormaghtigh, V. Raussens, and J.-M. Ruysschaert, ''Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes,'' *Biochim. Biophys. Acta* **1422**, 105–185 (1999).
7. H. Fabian and W. Maentele, ''Infrared spectroscopy of proteins,'' Chap. 5 in *Handbook of Vibrational Spectroscopy*, J. M. Chalmers

and P. R. Griffiths, Eds., pp. 3399–3425, Wiley, Chichester, U.K. (2002).
8. S. Yu. Venyaminov and F. G. Prendergast, "Water (H_2O and D_2O) molar absorptivity in the 1000–4000 cm^{-1} range and quantitative infrared spectroscopy of aqueous solutions," *Anal. Biochem.* **248**, 234–245 (1997).
9. D. Naumann, "Infrared spectroscopy in microbiology," in *Encyclopedia of Analytical Chemistry*, R. A. Meyers, Ed., pp. 102–131, Wiley, Chichester, U.K. (2000).
10. P. Lasch, J. Schmitt, M. Beekes, T. Udelhoven, M. Eiden, H. Fabian, W. Petrich, and D. Naumann, "Antemortem identification of bovine spongiform encephalopathy from serum using infrared spectroscopy," *Anal. Chem.* **75**, 6673–6678 (2003).
11. W. K. Surewicz, H. H. Mantsch, and D. Chapman, "Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment," *Biochemistry* **32**, 389–394 (1993).
12. R. Masuch and D. A. Moss, "Stopped flow apparatus for time-resolved Fourier transform infrared difference spectroscopy of biological macromolecules in ^1H_2O," *Appl. Spectrosc.* **57**, 1407–1418 (2003).
13. S. Sokolowski, A. J. Modler, R. Masuch, D. Zirwer, M. Baier, G. Lutsch, D. A. Moss, K. Gast, and D. Naumann, "Formation of critical oligomers is a key event during conformational transition of recombinant Syrian hamster prion protein," *J. Biol. Chem.* **278**, 40481–40492 (2003).
14. J. N. Adkins, S. M. Varnum, K. J. Auberry, R. J. Moore, N. H. Angell, R. D. Smith, D. L. Springer, and J. G. Pounds, "Toward a human blood serum proteome," *Mol. Cell. Proteomics* **1**, 947–955 (2002).
15. D. Naumann, D. Helm, H. Labischinski, and P. Giesbrecht, "The characterization of microorganisms by Fourier-transform infrared spectroscopy (FT-IR)," Chap. 3 in *Modern Techniques for Rapid Microbiological Analysis*, W. H. Nelson, Ed., pp. 43–96, VCH, New York (1991).
16. K. M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, Z. Huang, R. J. Fletterick, F. E. Cohen, and S. B. Prusiner, "Conversion of α-helices into β-sheet features in the formation of the scrapie prion proteins," *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10962–10966 (1993).
17. L. Ferrari, R. Seraglia, C. R. Rossi, A. Bertazzo, M. Lise, G. Allegri, and P. Traldi, "Protein profiles in sera of patients with malignant cutaneous melanoma," *Rapid Commun. Mass Spectrom.* **14**, 1149–1154 (2000).
18. J. Schmitt, M. Beekes, A. Brauer, T. Udelhoven, P. Lasch, and D. Naumann, "Identification of scrapie infection from blood serum by Fourier transform infrared spectroscopy," *Anal. Chem.* **74**, 3865–3868 (2002).