

TUTORIAL REVIEW

Fourier-transform infrared spectroscopy of biofluids: A practical approach

Ashton G. Theakstone¹  | Christopher Rinaldi¹ | Holly J. Butler² |
James M. Cameron² | Lily Rose Confield^{1,3} | Samantha H. Rutherford¹ |
Alexandra Sala^{1,2}  | Sayali Sangamnerkar¹ | Matthew J. Baker^{1,2} 

¹WestCHEM, Department of Pure and Applied Chemistry, Technology and Innovation Centre, Glasgow, UK

²ClinSpec Diagnostics Ltd, Royal College Building, Glasgow, UK

³CDT Medical Devices, Department of Biomedical Engineering, Wolfson Centre, Glasgow, UK

Correspondence

Matthew J. Baker, ClinSpec Diagnostics Ltd, Royal College Building, Glasgow, UK.
Email: matthew.baker@strath.ac.uk

Funding information

Cancer Research UK, Grant/Award Number: A28345; Engineering and Physical Sciences Research Council, Grant/Award Numbers: EP/L015595/1, EP/L505080/1

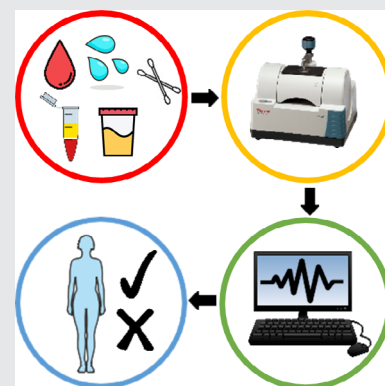
Abstract

Biofluid spectroscopy is an emerging technology in the field of clinical investigation, providing a simple way to extract diagnostic and observational information from easy to acquire samples. Infrared spectroscopy is well suited to analyse a large range of biofluid samples, including blood and its derivatives, due to flexible sampling modes and high sensitivity to subtle biological changes. As the technology advances towards the clinic, factors influencing successful clinical translation are becoming apparent. Here, we

provide a tutorial for effective biofluid spectroscopy study design, discussing sample and instrument parameters, as well as clinical considerations. The aim is to present the current understanding of clinical translation in the field of biofluid spectroscopy, and to facilitate other clinical applications to advance to the clinic.

KEYWORDS

biofluids, chemometrics, clinical translation, infrared, vibrational spectroscopy



1 | INTRODUCTION

In recent years, applications of Fourier-transform infrared (FTIR) spectroscopy have been rapidly expanding beyond simple structural characterisation of molecules, regardless of their chemical or biological contexts [1–3]. As molecular

vibrations are represented in the mid-IR range, the advantages offered by FTIR are not solely attributed to its intrinsic fundamental principles but also operational simplicity and analytical sophistication [4]. The technique allows for rapid, label-free and non-destructive analysis of biofluid samples with none-to-minimal sample preparation and produces results that are reproducible both qualitatively and quantitatively [5]. Furthermore, low analysis cost and minimal reagent use during analysis makes this technique

Ashton G. Theakstone and Christopher Rinaldi are co-first authors and contributed equally to this paper.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Translational Biophotonics* published by Wiley-VCH GmbH.

cost-effective and economically sustainable for clinical biofluid investigations. Additionally, FTIR combined with advanced chemometrics permits interpretation of complex biological spectra of metabolomic, forensic and clinical samples [6, 7].

FTIR spectroscopy may be performed with different sampling modalities for biofluid applications, either utilising attenuated total reflection (ATR-FTIR), transmission, or transmittance mode approaches. Principally, FTIR spectrometers irradiate mid-IR light on to a sample deposit where resultant absorption by sample constituents at specific frequencies enables determination of their molecular composition [8]. Absorbed frequencies appear at fundamental resonant vibrational frequencies of the atoms within the molecule and reflect the vibrational transitions from a lower/ground vibrational energy state to a higher vibrational state. These transitions appear on the spectrum as peaks which can be interpreted qualitatively and/or quantitatively using their position, shape and intensity [9].

The FTIR spectrum of biofluids contains a wealth of biological information and can be seen as a fingerprint-like biochemical snapshot of the condition of the sample/patient [5]. The most important regions in the IR spectrum for biological samples is the fingerprint region ($1800\text{--}900\text{ cm}^{-1}$), which contains the region where amide I and II bands are seen ($1700\text{--}1500\text{ cm}^{-1}$), as well as higher wavenumber regions ($3500\text{--}2550\text{ cm}^{-1}$) where stretching vibrations from O-H, S-H, C-H and N-H bonds can be seen [10]. In addition to this, the FTIR spectrum can also provide information on the secondary structures of proteins in biofluids; for instance, the position of Amide I and II band can be used to infer α -helical or β -sheet structures of proteins [11, 12]. The ability to qualitatively and quantitatively characterise biofluids is extremely valuable in a clinical context as samples contain several biomolecules, such as carbohydrates, lipids, nucleic acids and proteins, which interact with internal organs across the human body, which interact with internal organs across the human body [7, 13–15]. These biomolecules fundamentally share a structure and functional relationship influenced by their physiological environment and may be treated as biomarkers that can be used to diagnose pathologies and monitor disease progression and treatment therapeutics [7, 13–15]. Thus, FTIR spectroscopy represents an attractive analytical technique for clinical biofluid tests to aid patient diagnosis, prognosis, disease stratification and medical observation.

Infrared spectroscopy for biofluids is an ever-evolving field that has rapidly advanced over the past decades and demonstrated significant analytical capabilities for medical research; facilitating clinical investigations with a plethora of biofluids, including blood and blood

derivatives [16], sputum and saliva [17–19], urine [20], amniotic fluid [21], bile [22], cerebrospinal fluid [23] and pleural fluid [24]. Previously, Heise demonstrated the tremendous potential of mid-infrared spectroscopy for clinical biofluids, showing quantification of multiple blood chemistry biomarkers in blood plasma [25], and later whole blood and serum [26], utilising multivariate models and ATR, transmission and diffuse reflectance spectroscopy. At a similar period in time, Mansch's research, and later that of Petrich [27], also demonstrated quantification of specific blood serum constituents, such as total protein, albumin, triglycerides, cholesterol, glucose and urea, using transmission spectroscopy [28]. These seminal studies advanced the field by showing detection of clinically important biomarkers and, in the case of Heise, subsequently facilitated nanolitre quantification of glucose, firstly in aqueous solutions [29], and later in blood serum [30], at physiological concentrations towards a minimally-invasive detection strategy. Later, Heise's research demonstrated transmission FTIR detection of urea in blood plasma, by evaluating characteristic spectral peaks with multi-variate models, towards mid-infrared monitoring of patients during dialysis [31]. The pioneering work of Petrich further extended the potential of mid-infrared spectroscopy for detection of clinical disease, where spectroscopic serum analysis with multi-variate and advanced feature selection models showed sensitive and specific discrimination of rheumatoid arthritis patients [32]. Subsequently, Petrich demonstrated the far-reaching clinical utility of mid-infrared spectroscopy, with the ability to distinguish patients with myocardial infarction with high clinical accuracy within 6 hours of presentation of acute symptoms [33]. The promise of infrared spectroscopy for biofluids has further extended to oncology, with several important studies showing identification of numerous cancers of distinct biomolecular pathologies, including lung [18], brain [34, 35], breast [36–38], bladder [39], ovarian [4], cervical [40] and prostate [41] cancers, utilising transmission and ATR-FTIR spectroscopy. Additionally, spectral characterisation of biofluids has demonstrated quantification of biomarkers associated with infections [42] and metabolic disorders, and extended beyond disease diagnostics [7, 43–48], to treatment monitoring strategies [49–51], and understanding of molecular pathologies [52].

Recently, technological innovation, particularly regarding quantum cascade lasers (QCL's), has further advanced the field with the possibility to interrogate biofluids with a broadly tunable, coherent light source of high spectral density compared to traditional Globar infrared sources [53]. Analytically, QCL's have important implications for biofluid applications, allowing increased sensitivity and improved spectral characterisation of

liquid samples, as well as the ability to interrogate samples at discrete frequencies for multi-component quantification and reduced analysis time. Pioneering work of both Lendl and Petrich have shown adoption of QCL based spectroscopic systems for sensitive detection of amide I and amide II peaks [54, 55] and continuous monitoring of glucose [56], respectively, in aqueous samples, enabling qualitative and quantitative studies of proteins and other clinical biomarkers in liquid biofluid samples. Practically, QCL's are semi-conductor devices that operate at room temperature and may be thermal electrically cooled, and are, hence, ideally suited for miniaturisation and development of portable "point-of-care" instruments for spectroscopic biofluid applications in clinical settings.

Despite the potential of FTIR spectroscopy for the study of biofluids [43, 52, 57, 58], development of this technique for clinical translation is still in the early stages. Current processes performed in clinics and hospitals throughout the UK for biofluid analysis typically involve the use of wet chemistry and electrophoresis techniques. Immunoassays like the Biuret method is a common wet chemistry technique used for the detection of total protein in solution; however, like many other wet chemistries, this technique requires extensive sample preparation, requires large volumes (up to 1 mL) of the biofluid being assessed and can take up to several hours for results. Similarly, colorimetric and fluorometric assays, based on Enzyme-Linked Immunosorbent Assay (ELISA) platforms, are time-consuming multi-step processes, and further critically rely on availability of specific antibodies for proteins of interest [59–61]. Electrophoresis involves the separation of molecules present in a biofluid via migration through a matrix upon application of an electric field, separating the biofluid into various fractions based on their molecular weight and electric charges [62]. Electrophoresis techniques are also subject to interference with substances such as free lipids, drugs, haemoglobin and bilirubin which affect analytical results [63]. Furthermore, this technique can often take several

minutes to hours for final results and requires the accessibility of different support media depending on the type of test being performed. Nevertheless, current analytical techniques are widely employed in clinical biochemistry laboratories, since these technologies offer capabilities to perform highly automated, standardised testing, which is of critical importance for analysis of high volume clinical biofluid samples. Hence, development of rival analytical techniques must not only offer distinct practical and economic advantages, such as rapid, simple, label-free testing, but must also be conducive to automation and demonstrate a high degree of standardisation for high-throughput clinical testing. A comparison of FTIR spectroscopy and commonly employed biofluid techniques including the Biuret immunoassay method, ELISA assay and electrophoresis techniques is displayed within Table 1; which outlines the time required at each analysis step, sample volume size and limit of detection.

Recently, tremendous progress has been made regarding the development of high-throughput technologies for FTIR of biofluids [35, 69], with emergence of multi-well systems for both transmission and ATR-FTIR platforms. Nevertheless, the design, implementation and adoption of standard protocols is critically required to pave the way from the laboratory to the clinic and to make routine adoption of FTIR for biofluids a reality [16, 35, 70].

2 | EXPERIMENTAL DESIGN

2.1 | FTIR modes and sample considerations

Biofluid investigations may be conducted with FTIR spectroscopy utilising different sampling modalities; namely, attenuated total reflection (ATR-FTIR), transmission and transmittance mode approaches, where respective techniques have vitally important sample considerations.

ATR-FTIR is a spectroscopic sampling modality characterised by total internal reflection of infrared

TABLE 1 Comparison of analytical techniques for clinical translation in biofluid studies

Analysis step	Biuret method	ELISA	Electrophoresis	FTIR spectroscopy
Preparation time	2–4 h [64]	2.5 h ^a [65]	3–4 h [66]	10 min [10]
Sample volume	≤1 mL [60, 61]	<1 mL [65]	<1 mL [66]	<1 mL [10]
Drying time	N/A	N/A	N/A	<10 min [10]
Analysis time	<30 min [60, 61]	3.5 h [65]	45–90 min [66]	20 min [10]
Data analysis time	N/A	N/A	N/A	1 hour - 2 days ^b [10]
Limit of detection (LOD)	0.4 mg/mL [61]	39 pg/mL [65]	6–8 µg/mL [67]	7.5–150 µg/mL [68]

^aDepending on ELISA kit purchased. Overnight incubation may also be required.

^bDepending on size of data set and analysis required.

radiation incident on an internal reflection element (IRE) where evanescent waves penetrate sample components to identify their molecular composition. The IRE comprises an infrared transparent material where the refractive index of the material must exceed the refractive index of the sample to satisfy total internal reflection of the incident infrared beam. Materials with a high refractive index are chosen to minimise the critical angle since the angle of incidence must exceed the critical angle to instigate total internal reflection [71]. The critical angle, θ_c , may be calculated as a function of the refractive index of the sample, n_1 , and the refractive index of the IRE, n_2 .

$$\theta_c = \sin^{-1} \frac{n_2}{n_1}$$

Total internal reflection of infrared radiation at the interior surface of the IRE causes propagation of evanescent waves parallel with and confined to the surface of the IRE with a typical depth of penetration of ~ 1.0 – $2.0 \mu\text{m}$ in the fingerprint region [10, 72]. The penetration depth is defined as the distance from the external surface of the IRE where the electric field amplitude decreases by $1/e$ of its initial value [73]. The penetration depth of the evanescent wave, d_p , may be calculated given the wavelength of applied light, λ , the angle of incidence with respect to normal of the IRE, θ , the refractive index of the sample, n_1 , and the refractive index of the IRE, n_2 .

$$d_p = \frac{\lambda}{2\pi(n_1^2 \sin^2 \theta - n_2^2)^{1/2}}$$

Consideration of the optical properties of ATR crystals is imperative for successful spectroscopic interrogation of biofluid samples, since the angle of incidence and refractive index of the IRE have significant influence on the resultant amplitude of evanescent waves. In particular, the angle of incidence of infrared radiation should be appropriately aligned with respect to the calculated critical angle for a given IRE to maximise spectroscopic signals for biofluid experiments, since the depth of penetration of evanescent waves decreases exponentially further from the critical angle.

Careful consideration should further be given to selection of the appropriate IRE for a given spectroscopic biofluid application since component design has a profound influence on acquired infrared spectra and capabilities for high-throughput clinical testing. Currently, IREs are predominantly manufactured from diamond, zinc selenide and germanium materials due to both their high refractive index and superior optical transparency over mid-infrared wavelengths. Arguably, diamond is

considered the gold-standard IRE for biofluid experiments since the substrate provides access to the entire biologically relevant spectral window, while boasting significant robustness and chemical inertness over the full, 1–14, pH range, albeit at an increased overall cost. Conversely, zinc selenide IREs are not suitable for analysis of biofluid samples of pH <5 and >9 due to surface sample interactions, while germanium IREs give rise to a lower wavenumber cut-off of 780 cm^{-1} that may obscure biological information and produces a reduced depth of penetration for a fixed angle of incidence relative to other IRE substrates. While conventional IRE materials have demonstrated excellent performance for biofluid diagnostics, these substrates have limited sample capacity and are cost-prohibitive for high-throughput testing given the fixed point of analysis [35]. Furthermore, implementation of standard IREs into fast paced clinical environments may be challenging since the fixed nature of substrates pose risks of biological contamination between sample analyses and would require thorough and time-consuming sterilisation procedures. Therefore, careful consideration should be given to prospective biofluid applications prior to selection of an appropriate IRE material to establish whether the substrate is conducive to a particular clinical context, whether that be basic research or high-volume clinical testing.

Recently, silicon wafers have been successfully utilised as low-cost ATR crystals, in both single and multi-reflectance formats, where the reduced optical pathlength permits access to the biologically important fingerprint region previously obscured with standard silicon hemisphere IREs [35, 74, 75]. Traditional silicon IREs have a long optical pathlength that promote significant silicon multi phonon and interstitial oxygen absorptions and result in poor transmission of infrared radiation $<1500 \text{ cm}^{-1}$ [35, 74, 76, 77]. Consequently, pertinent vibrational modes in the fingerprint region will undoubtedly be obscured from spectroscopic analysis with standard silicon IREs that have previously been shown to provide vital biological information on biofluid components related to disease pathophysiology [35]. Similarly, conventional multi-reflectance silicon ATR crystals exhibit poor transmittance of infrared light $<1500 \text{ cm}^{-1}$ since multiple reflections increase the optical path length and accentuate intrinsic silicon lattice vibrations [75]. Hence, micro-fabricated silicon IREs based on silicon wafer technology should be selected over standard silicon IREs for biofluid experiments, and may provide a low-cost, high-throughput alternative to other previously mentioned IRE substrates.

ATR-FTIR spectroscopy represents an attractive analytical tool for biofluid diagnostics since the technique offers a rapid, economical and non-destructive platform

that requires minimal sample preparation and negates the need for costly labels and reagents [10]. ATR-FTIR spectroscopy has previously demonstrated considerable promise as a diagnostic platform for the clinical arena, with successful proof-of-concept studies involving a wide range of biofluids, such as blood serum [4, 38, 78], blood plasma [4, 19], saliva [17, 19] and urine [19, 79]. ATR-FTIR spectroscopy is particularly suited for biofluid applications since the penetration depth of the evanescent wave is well-defined and sample pathlength remains constant between measurements irrespective of small deviations in sample thickness, in contrast to transmission approaches [80]. This is a particularly important consideration for dried samples where coffeeing, gelation and cracking patterns together contribute to formation of inhomogeneous biofluid films of non-uniform thickness [44]. To this extent, careful consideration should be given to temperature, humidity and the consequent evaporation rate when performing sample drying to maximise spectral reproducibility, since such parameters profoundly influence drying patterns [44, 81, 82]. Furthermore, sample volumes should be carefully considered during sample preparation protocols, such that deposited biofluids should cover the entire surface of the IRE and provide a minimum sample thickness of three to four times the penetration depth of evanescent waves, both to prevent scattering artefacts and promote sufficient spectral signal to noise, respectively [10].

ATR-FTIR spectroscopy should be preferentially considered for liquid biofluid analysis over other FTIR methods since the reduced sample pathlength and exponentially decaying evanescent waves reduces interaction with the strong dipole moment of water molecules [44]. Hence, adoption of ATR-FTIR approaches for wet biofluid analysis should permit reduced absorption of infrared light by water molecules, which often saturates infrared spectra in transmission FTIR studies [47, 83]. While water molecules still obscure pertinent biological signatures in ATR-FTIR spectra, Sala *et al.* have recently successfully demonstrated digital drying of liquid blood serum sample spectra to remove spectral contributions from water, enabling discrimination of brain cancer from non-cancer patients with sensitivities and specificities greater than 93% and 83%, respectively [84]. Nevertheless, dry sample analysis is still conventionally employed for biofluid ATR-FTIR studies, given the substantial increases in absorption signals of biological spectral features [80]. Overall, ATR-FTIR spectroscopy is an effective and widely recognised spectroscopic technique for biofluid diagnostics, and considering recent advances in IRE technology, may now be on the cusp of clinical translation.

Transmission FTIR is a spectroscopic sampling modality characterised by the projection of a spectrum of infrared radiation on to an optically transparent cell where the wavelength and intensity of transmitted infrared light permits molecular classification of sample constituents. Infrared light absorbed by the sample, A , corresponds to the intensity of incident, I_0 , and transmitted light, I , and is a function of molar absorptivity, ϵ , pathlength, l , and concentration, c , of the sample, in accordance with the Beer-Lambert law [26].

$$A = \log_{10} \left(\frac{I_0}{I} \right) = \epsilon l c$$

The sample pathlength in transmission FTIR spectroscopy is not wavelength dependent and instead permeates the sample producing spectra indicative of bulk sample components in contrast to ATR-FTIR spectroscopy [71]. Sample path length is imperative to acquisition of quality spectra with transmission FTIR spectroscopy and should be specified at 1–20 μm to prevent signal saturation and non-linearity of the Beer-Lambert law, both of which are detrimental to clinical spectroscopic analyses [71]. Furthermore, liquid interrogation of aqueous biological media should necessitate implementation of a significantly reduced sample path length, $\sim 6 \mu\text{m}$, to account for the strong infrared absorption of water molecules [83]. Practically, transmission FTIR spectroscopic analysis of wet biofluid samples may be difficult to implement in the context of high-volume clinical testing, given reproducibility issues with spacer thicknesses, surface interactions, presence of air bubbles in samples, and the laborious and time-consuming configuration of current liquid cells [83]. Furthermore, transmission FTIR spectroscopy is notoriously challenging for wet biofluid analysis with regards to sample reproducibility and the ability to accurately maintain a consistent level of sample wetness. Alternatively, discrete frequency infrared spectroscopy that utilises a quantum cascade laser (QCL) source may be considered for transmission mode experiments of wet biofluids, since the increased brilliance and emission power of the source facilitates interrogation of liquid samples at increased optical pathlengths [85]. Adoption of transmission FTIR spectroscopy is widespread for dried biofluid films, and the recent emergence of a silicon high-throughput transmission FTIR accessory (HTS-XT) has enabled high-throughput clinical biofluid testing on a 384 well silicon plate [5, 86]. Sample dilution protocols should be carefully considered prior to clinical biofluid testing with HTS-XT platforms, since 3-fold dilutions were found to promote reproducible drying patterns of serum samples resulting in improved spectral acquisition [5, 86]. However, the ratio of amide I and II peaks and

defined band positions were still found to shift with transmission HTS-XT platforms compared to ATR-FTIR systems, attributable to dispersion effects and should be recognised prior to data interpretation [86]. Additionally, consideration should be given to implementation of transmission HTS-XT systems into clinical laboratories, since provided silicon multi-well plates do not offer the possibility of a disposable testing platform. In this regard, comprehensive cost benefit analyses should be conducted to determine the practical implications of introducing high-throughput transmission FTIR systems into clinical environments prior to biofluid diagnostic testing.

Transflection FTIR describes the projection of infrared light onto a sample deposited on a reflective coated slide where a small proportion of incident light is specularly reflected, with the majority transmitted to the underlying metal surface and then projected back through the sample. The reflected component of infrared light enables identification of characteristic frequencies of sample constituents and it should be noted that absorption bands on resultant infrared spectra are significantly larger than those from transmission and ATR-FTIR spectroscopy experiments, due to the increased sample path length [71]. Therefore, transflection FTIR spectroscopy is not applicable for wet biofluid diagnostics because the increased sample path length causes increased absorption of infrared light by water molecules relative to ATR-FTIR and transmission FTIR platforms. However, transflection FTIR spectroscopy boasts several practical and analytical advantages for dried biofluid analysis, including use of economical low emissivity slides which provide potential for high-volume, disposable testing, in contrast to conventional fixed mode ATR-FTIR platforms [70]. Furthermore, transflection FTIR spectroscopy provides increased absorbance of infrared light by dried biofluid components given the increased sample pathlength [70]. Nevertheless, transflection FTIR experiments are particularly susceptible to spectral artefacts for dried sample analysis given the increased interaction of infrared light with inhomogeneous biofilms and resultant resonant Mie scattering [10]. Hence, the signal-to-noise ratio (SNR) of infrared spectra is recognised to be poorer for transflection FTIR measurements in comparison to ATR-FTIR approaches [10]. Furthermore, transflection FTIR has been shown to suffer from the electric field standing wave effect where the intensity of absorption bands varies across infrared spectra, with the phenomena strongly influenced by sample thickness of dried biofluid films [1, 10, 71, 87]. For such reasons, it has previously been recognised that transflection FTIR may not be best suited to study biological materials in comparison to transmission FTIR and ATR-FTIR approaches [10]; we believe this is

particularly true for biofluid applications given practical difficulties in achieving reproducible biofilm drying patterns.

2.2 | FTIR microspectroscopy

FTIR microspectroscopy describes the coupling of conventional FTIR modalities with microscopy and facilitates the acquisition of spatially resolved information on chemical constituents within biological samples. To date, FTIR microspectroscopy has been successfully demonstrated on a plethora of cell and tissue samples [88–91] and provides a chemically rich alternative to traditional histopathological and cytological inspection of tissue architecture. Similarly, FTIR microspectroscopy has recently been successfully employed on biofluid samples [69, 92], where simultaneous analysis of sample arrays has enabled realisation of a high-throughput liquid biopsy screening strategy for clinical environments. The ability to spatially resolve dried biofluid films further permits comprehensive evaluation of heterogeneous samples influenced by drying patterns and the coffee-ring phenomenon.

FTIR microspectroscopy may be performed in either point or wide-field mode and spectral acquisition or imaging/mapping modes determined by the selection of either single element or linear array and focal plane array detectors and aim of the experiment, respectively. Principally, point mode mapping describes a raster scan approach where use of an infrared aperture allows spectral acquisition at localised sample points in stepwise fashion. Conversely, wide-field FTIR microspectroscopy describes exposure of a deposited sample to an infrared light source of increased divergence, where multi-element detectors allow simultaneous acquisition of considerable quantities of spectra representative of larger sample areas. Critically, selection of either spectroscopic approach has practical implications for experimentation and influences analytical performance, and should therefore be carefully considered prior to implementation for biofluid studies.

Point mode mapping measures spectra with a high SNR with spatial resolution defined by the diameter of the infrared aperture with respect to the diffraction limit. However, the nature of point mode mapping approaches results in considerable analytical time for analysis of entire biofluid samples, without reduction in the scan time per pixel that would consequently produce lower SNR spectra. In contrast, wide-field FTIR microspectroscopy is a rapid technique that collects spectra with good SNR where spectral resolution is limited by the diffraction limit. Therefore, wide-field FTIR microspectroscopy is strongly advocated for high volume biofluid applications where timely

diagnosis is imperative in clinical laboratories. Alternatively, point mode FTIR microspectroscopy facilitates detailed acquisition of spectroscopic information confined to a particular sample location and therefore may be particularly useful in research activities for identification of biomolecular signatures within heterogeneous biofluid films.

Consideration should also be given to instrumentation when selecting the appropriate FTIR microspectroscopy platform for biofluid studies. The light source is fundamental to spectral acquisition and may feature global, synchrotron or QCL sources. Currently, global sources are commonly used for FTIR microspectroscopy where emission of black body radiation at a specified temperature provides a stable and economical source with high SNR and peak energy density in the mid-IR region [93]. However, synchrotron radiation provides significant advantages for FTIR microspectroscopy when the spatial resolution approaches the diffraction limit, where the increased source brightness relative to global light beams improves SNR of acquired spectra [94]. Furthermore, QCLs have recently demonstrated significant potential for FTIR microspectroscopy and operate on the principle of inter-sub-band transitions where the coherent light source emits photons of narrow line widths with increased power densities compared to global sources [53]. Practically, QCLs offer the prospect of reduced sample scan time given the ability to produce discrete frequencies of mid-infrared light that can be engineered to correspond to pertinent diagnostic frequencies of biofluid samples [95]. The discrete frequency approach further circumvents the need for a Michelson interferometer, resulting in low cost, compact instrumentation. Nevertheless, global sources represent a mature and economical infrared technology where component simplicity currently offers an attractive and accessible tool for integration in to spectrometers within clinical laboratories.

The choice of infrared detector further influences spectroscopic experimentation and may feature either a thermal or photonic component. Thermal detectors utilise pyroelectric materials, commonly deuterated triglycine sulphate, which convert temperature fluctuations to readable electrical signals and offer primary advantages of low cost and room temperature operation at the expense of reduced response time and sensitivity. Photonic detectors utilise semiconductor materials with narrow band gaps, commonly mercury cadmium telluride (MCT), which produce electronic excitations in response to incident infrared photons, and provide superior spectral SNR although require cooling with liquid nitrogen. Overall, there is a trade-off between practicality, cost and analytical performance when selecting appropriate instrumentation for FTIR microspectroscopy platforms and respective parameters should be carefully considered prior to conducting biofluid studies.

2.3 | Spectral acquisition

It is imperative that a background spectrum is obtained prior to the deposition of biological samples to account for atmospheric conditions, particularly changes in carbon dioxide and water vapour, which may otherwise negatively influence spectroscopic data analysis. For standard FTIR approaches, one background spectrum is performed prior to repeat and replicate measurements of a particular biofluid sample to reduce the impact of fluctuating laboratory environments. For FTIR microspectroscopy, one background spectra is typically acquired for every 5-10 consecutive sample spectra, hence, multiple background measurements must be collected at specified sample points for experiments given the longer experimental time scales.

Spectral quality is of paramount importance for spectroscopic disease diagnostics and is characterised by SNR of infrared spectra, which is dependent on analytical time and scan resolution, as summarised by the FTIR trading rules. The first rule states spectral quality increases at the expense of analytical time, where the resultant SNR is proportional to the square root of accumulated scans [93]. The second rule states that spectral quality decreases as the scan resolution increases, attributed to the increase in spectral noise with the acquisition of additional information, where the SNR is proportional to scan resolution [93].

2.4 | Data processing

There are numerous pre-processing options to explore when it comes to large spectroscopic datasets, to improve multivariate and classification algorithms by reducing computational burden. Raw spectral data can be difficult to interpret as it often contains unwanted noise and artefacts, therefore pre-processing is usually required [96–98]. Some common pre-processing steps include spectral cut, binning, smoothing, normalisation, and baseline corrections such as extended multiplicative signal correction (EMSC) [4, 96, 97].

Selection of the spectral region of interest is often the first step in pre-processing. Many studies will use the full IR spectrum for their data analysis; however, it is common practice to cut the spectra to a desired region to enable faster analysis with less data points. When interrogating biofluids, the spectra are often cut to the fingerprint region where biomolecules are known to vibrate ($1800\text{--}900\text{ cm}^{-1}$). Many studies also include the high wavenumber region ($\sim 3700\text{--}2700\text{ cm}^{-1}$) which relates to proteinaceous and lipidic vibrations.

Binning involves reducing the number of data points by averaging adjacent data points, to lower the dimensionality of the dataset. The number of data points to be averaged is represented by a bin factor [96, 97]. For example, a bin factor of 1 is considered as no binning, but a bin factor of 4 results in each 4 data points being averaged and is substituted by the mean value. Binning is useful in both reducing the computational burden of a dataset and improving the SNR by reducing the impact of small fluctuations between adjacent data points. It is particularly useful with datasets that contain a large range of wavenumbers, therefore thousands of data points; however, decimation of data points can increase the SNR by increasing the data spacing within a spectrum, ultimately losing spectral resolution. Therefore, when choosing a bin factor it is important to keep these potential issues in mind.

Smoothing removes high frequency noise while preserving low frequency components in order to reduce the appearance of noise within the dataset. Occasionally, spectral features or unresolved peaks can be mistaken for high frequency noise, therefore each smoothing technique comes with a substantial risk [96, 99]. Three common smoothing techniques include Savitzky-Golay (SG) filtering, wavelet denoising and local polynomial fitting with Gaussian weighting. The SG filtering technique is commonly used within FTIR spectroscopy and is a method based upon a local least-squares polynomial approximation. The least-squares smoothing maintains peak morphology while minimising high frequency noise and involves fitting a polynomial within a moving window of a fixed degree [96, 97]. Wavelet denoising is particularly useful within FTIR spectroscopy for spectra that have high SNRs and is known to improve visual spectral quality. This technique estimates the smoothed result by looking at one data point and a collection of data points following, resulting in a noise-free signal as the output [96, 100]. The third smoothing technique for FTIR spectroscopy involves estimating and fitting Gaussian curves to the spectrum at varying bandwidths, which can be useful for general denoising [97].

Normalisation is a common pre-processing technique in FTIR spectroscopy that further allows the removal of any spectral artefacts, which includes min-max scaling (between 0 and 1), vector normalisation in biological samples [96, 97]. Min-max scaling sets the minimum and maximum absorbances to 0 and 1, respectively, which results in all areas of the spectrum being scaled in relation to each other. For vector normalisation, each wavenumber variable is initially averaged and then subtracted from the original spectrum to equal zero. Following this, each wavenumber is squared, then divided by the square root of the total sum of squared wavenumber variables, thus normalising the spectral dataset to a

magnitude of one [101]. Additionally, peak normalisation can be used where the intensity corresponding to a particular absorbance band is used as a reference, such as the Amide II or Amide A bands, or most commonly the Amide I band. Normalisation to the Amide I band ensures all spectra is scaled according to the maximum intensity of the peak within the Amide I wavenumber region ($1700\text{-}1600\text{ cm}^{-1}$) [96, 97, 102]. The choice of normalisation technique is often in conjunction with the baseline corrections that are applied and are discussed below.

Baseline corrections for pre-processing also have numerous options including Savitzky-Golay derivative filters (or differentiation—first and second), rubberband baseline correction, polynomial baseline correction and EMSC [97]. Derivative filters are very commonly used within FTIR spectroscopy as they are a straightforward mathematical transformation which reduces baseline differences and improves spectral resolution. The shape of the spectra will be altered by differentiation; however, it has the ability to resolve any overlapping bands [10]. Rubberband baseline corrections are useful when the background noise within a spectrum is non-linear and adjusts the baseline within specific areas by fitting a convex polygonal to the troughs of the spectrum [99]. Polynomial baseline correction is more widely used within Raman spectroscopy as baselines are often less consistent than with FTIR spectroscopy; however, can still be utilised for this type of analysis [96]. EMSC scales each data point according to a reference spectrum, while irrelevant polynomial trends are subtracted. The addition of prior knowledge about spectral patterns from the reference spectrum allows for an improvement in data quality by correcting unwanted additives and multiplicative effects [103, 104]. There is no universal method approach to which is the best baseline correction, or any pre-processing technique, and choices are typically based on visual problems within the individual spectral dataset [10].

Each of these pre-processing techniques discussed here can be used in conjunction with each other and the order they are applied to the raw data can be vital towards the output of pre-processed data, as shown by Butler *et al.* [96] Generally they are applied as a *trial-and-error approach* depending on the individual dataset, analysis goal and the computing power available [10]. Schematic overview as displayed by Butler *et al.* is shown in Figure 1.

2.5 | Exploratory analysis

After the raw spectral data has been pre-processed it is common for principal component analysis (PCA) to be

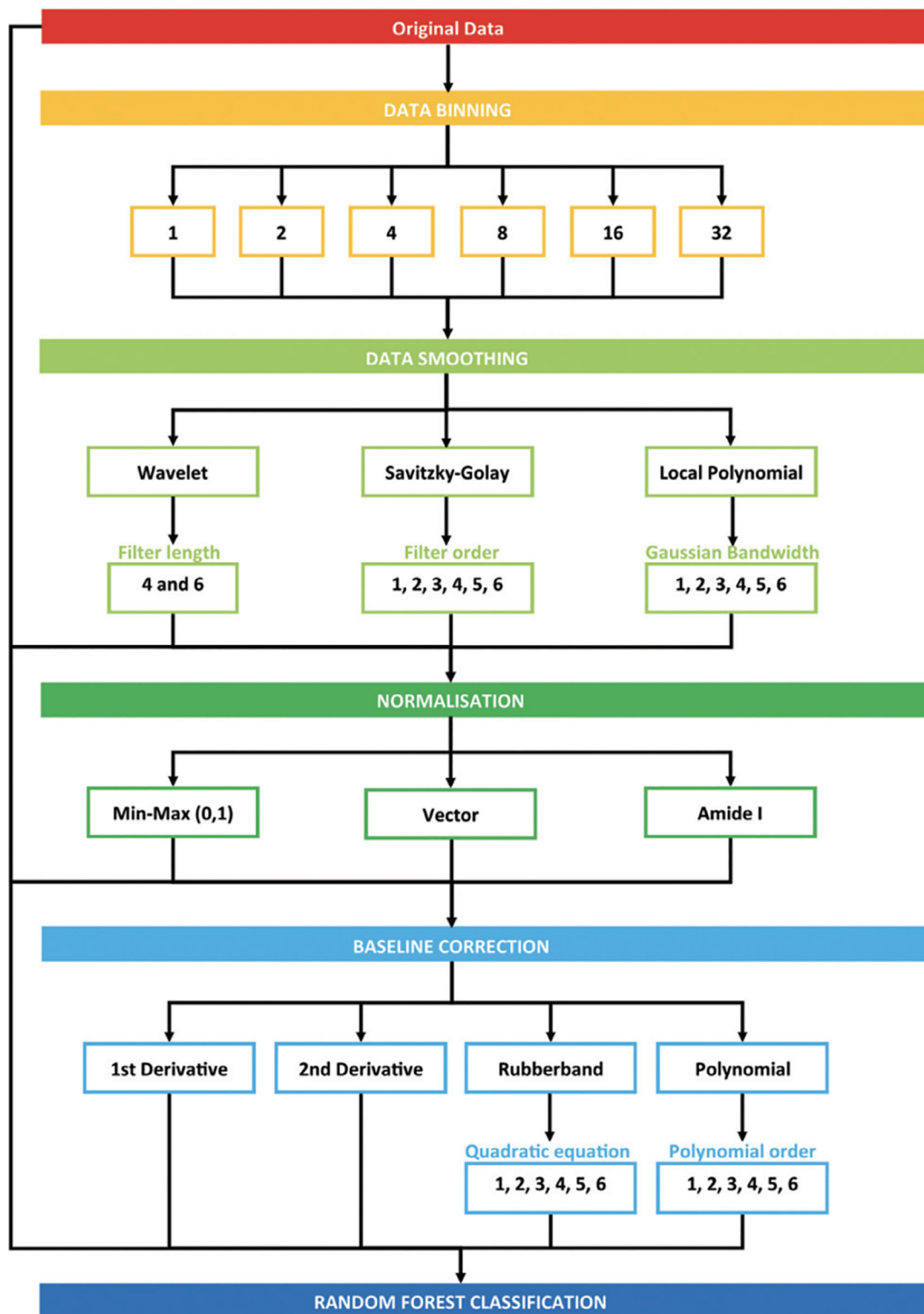


FIGURE 1 Schematic overview of pre-processing steps including binning, smoothing, normalisation and baseline corrections; to be applied before exploratory or classification analysis (Random forest as an example) [96]

used as an exploratory technique to determine variance between classes. It involves an orthogonal linear transformation which reduces the dimensionality of the data set

and establishes covariance between variables within the spectra [96, 102]. The covariance matrix is converted into scores and loadings to allow each covariance to be

displayed as principal components. The first principal component represents the greatest covariance within the dataset and is typically the greatest interest for FTIR spectroscopic data. The PCA scores plots are used to visualise the covariance between certain principal components and the corresponding PCA loadings identify the wavenumber regions that contain the covariance between spectral classes [102].

Another technique commonly employed in spectroscopic research is hierarchical clustering analysis (HCA), which is used to explore the similarity between observations and/or clusters [105, 106]. HCA repeatedly performs the following two steps: (a) identify the two observations that are closest together, and (b) merge the two most similar clusters. This iterative process continues until all the clusters are merged together, and the result can be visualised using heat maps or dendrograms, which are tree diagrams describing the hierarchical relationship between objects [107].

Exploratory techniques are classified as unsupervised methods of analysis as they only have input variables and will not be influenced by corresponding output variables, training and test sets; however, they can be useful in preliminary analyses to gain an understanding of the wavenumber regions (and therefore biological components) that are responsible for the covariance within the dataset.

2.6 | Classification analysis

For disease diagnostics there are a variety of machine learning algorithms for classifications. For example, neural networks [108], random forest (RF), linear discriminant analysis (LDA), partial least squares-discriminant analysis (PLS-DA), soft independent modelling by class analogy (SIMCA) [109] and support vector machine (SVM), are all supervised techniques that create a classification function from training data [35, 102, 110, 111]. Prior to classification, datasets are generally split into two parts: a training and a test set. The training set is used to identify disease biosignatures, and the test set is used for the prediction. Moreover, cross-validation is often used to achieve a significant approximation of how the classification would perform in a real-world environment. In this process, training and test sets are sampled multiple times (ie, *k-fold*), and the test set is examined against the training set in every fold to evaluate the true performance of the model and minimise the bias in sample selection [112]. RF, PLS-DA, SIMCA and SVM are four classification techniques that have shown excellent performance and reliability for disease diagnostics [35, 102, 110, 111, 113]. Each of them can be optimised by tuning their

parameters according to the cross-validation performance. To build a robust diagnostic model, the classification should be repeated a number of times (ie, *resampling*) in order to minimise the error in the results. Bootstrapping analysis can be utilised to determine an acceptable number of iterations required to obtain reliable results.

RF is a widely used machine learning algorithm and is a robust, accurate technique for spectral diagnostics. This method averages the predictions of several independent base models for an output prediction that is a binary classification. It involves using the Classification and Regression Trees (CART) algorithm to build an ensemble of decision trees where the classification prediction is a result of majority vote of all the decision trees within the forest [97, 114]. Within this technique there are three main training parameters; *ntree*, *mtry* and *nodesize*. The number of trees is represented by *ntree*, the number of descriptors available at each split is represented as *mtry* and the depth of the trees is referred to as *nodesize* [97, 115]. Palmer *et al.* investigated the effect of RF classifications with a variety of training parameters (*ntree* from 1 to 5000, *mtry* from 1 to 126 and *nodesize* from 1 to 50) and ultimately observed little variation within the ranges; *ntree* from 250 upward, *mtry* between 40 and 126, and *nodesize* between 5 and 10 [115]. When choosing a value for *ntree*, it is suggested not to use less than 250, as results did deteriorate; however, higher values (above 500) did not see an improvement in classification only an added increase on computational burden. Using a smaller value for *mtry* (<40), the predictive quality decreases as not enough descriptors are available at each split, therefore values of 40 and above are recommended. While a *nodesize* that is too big increases the size and range of the trees, and ultimately there is a decrease in predictive accuracy [115].

PLS-DA combines PLS regression and LDA. It reduces the dimensionality of the data that in turn reveals hidden patterns and can be used to extract important information out [116]. Classes are separated into two distinct regions by a straight line, where data points are projected perpendicular to the line, known as a discriminator. Discriminant scores are the distances from the discriminator and provides new variables called PLS components [116]. The first component (PLS1) will account for the greatest variation in the dataset, the second (PLS2) will account for the second greatest variation, and so on. Like PCA, PLS has scores and loadings plots that outline the general inconsistencies and which wavenumber regions have the highest disparity [102, 116].

In datasets where there are many variables, PLS is particularly useful as it replaces the original variables

with fewer latent variables (LVs), which is a tunable parameter. Over-fitting can occur when too many LVs are selected and the output information includes data as well as noise, while too little LVs provide insufficient output data information and represent an under-fitted model. Selecting the number of LVs can be a crucial step in building the model as both over- and under-fitting is undesirable. The most common method for prediction ability is to employ a cross-validation in order to tune the PLS model [117].

SIMCA is a well-established supervised classification model for spectroscopic data that employs PCA to evaluate spectra within a data training set. SIMCA performs independent PCA analysis on spectra of respective sample classes in the training set such that sample groups have a distinct PC space, based only on significant principal components, for classification of unknown samples. In a binary classification model, constructed PC spaces are utilised to assign samples in a test data set to one of four outcomes, in particular, samples may be assigned to one of respective classes, or sample belongs to both or neither class, from which performance statistics are subsequently calculated [118].

SVM provides an optimal dimension, known as the hyperplane, for the separation of the data. Each support vector is a co-ordinate of an individual observation and the hyperplane is used to categorise samples. Tuning parameters for SVM classifications can have a significant effect on the output. The parameter *cost*, for example, is the trade-off between the ability to classify data and the smooth boundaries [97, 102].

Each classification model described here will provide an output of sensitivity, specificity, kappa and balanced accuracy [102]. The sensitivity and specificity refer to the ability of predicting true positives and true negatives, respectively, within the dataset. The kappa value gives a measurement of the agreement between observers and helps to understand the reliability of the model. While the balanced accuracy illustrates the overall performance of the model by averaging the accuracy of either class [102]. Examples from the literature highlight the importance and applicability of each of these models for classifications within biofluid studies. This includes examples such as a study by Smith *et al.* who utilised RF to identify spectral features within serum for distinction between cancer and non-cancer patients [110]; a study by Dickens *et al.* who relied on PLS-DA to differentiate between disease stages in multiple sclerosis (MS) patients [119]; and a study by Zhang *et al.* who developed an SVM based algorithm for the classification and prediction of breast cancer in peripheral blood [120]. Table 2 highlights the main advantages and disadvantages for each model described here.

TABLE 2 Advantages and disadvantages of RF, PLS-DA, SIMCA and SVM algorithms (Partially adapted from Gromski *et al.* [121] and Granato *et al.* [122])

Model type	Advantages	Disadvantages
RF	<ul style="list-style-type: none"> • Can handle noisy and missing data • Robust to outliers • Estimates what variables are important in classifications 	<ul style="list-style-type: none"> • Sensitive to the number of variables selected at each node
PLS-DA	<ul style="list-style-type: none"> • Reduces dimensionality • Can handle noisy data • Can draw confidence intervals around groups • Can predict either continuous or categorical variables 	<ul style="list-style-type: none"> • Often results in over-fitting • Model validation is essential
SIMCA	<ul style="list-style-type: none"> • Good for high-within class variability • Can evaluate model performance based on classification outcomes • Can recognise outliers 	<ul style="list-style-type: none"> • Neglects between-class information that could aid in classification • Metrics used in classifications are pre-determined • Results dependant on number of chosen PCs
SVM	<ul style="list-style-type: none"> • Less likely to over-fit • No local minima • Robust to outliers • Useful in both linear and non-linear data 	<ul style="list-style-type: none"> • Computationally expensive with large datasets • Binary classifier

3 | MATERIALS

Materials necessary for optimum sample preparation and analysis will be dependent upon the study aim and design. The authors guide the readers the Experimental Design section for further information and provide recommendations below for key materials that may be useful.

3.1 | Reagents

- Biofluid sample, such as; blood, blood derivatives, saliva, sputum, urine, bile, sweat, breast milk, tissue fluid, cerebrospinal fluid, pleural fluid.

CAUTION: All human biofluid samples must be obtained with ethical approval, usually granted by local research ethics committee, following review of study design and patient participation consent forms.

- Fixative reagents, such as; formalin, ethanol, or methanol, where sample fixation may be required.

CAUTION: Fixatives are often irritants, and may have additional associated hazards, so careful risk assessment (RA) using material safety data sheets (MSDSs), and control of substances hazardous to health (COSHH) is essential.

- Physiological water, or equivalent dilutive agent, for dilution studies where concentrated biofluid sample may be unsuitable for analysis, such as; transmission investigations with limited pathlengths, or to prevent the impact of the Vroman effect as thoroughly discussed by Cameron *et al.* [44]
- Reagents for containing spillages, such as; ethanol (70%) or disinfectant agents like Virkon (Antec, DuPont), that can be applied in the circumstance of sample escape, and/or general cleaning of surfaces and substrates.

3.2 | Equipment

3.2.1 | Sample handling

- Storage facilities, including refrigerator (4°C), freezer (−20°C), or ultra-low freezers (−80°C)

CAUTION: Human tissue samples will degrade at room temperature over time, which will be spectrally apparent. A consistent low temperature is preferred for stable sample storage. For samples containing cellular matter, ultra-low temperature storage is critical.

- Sample storage containers such as; Eppendorf tubes and cryogenic vials, dependent upon chosen storage method. Sample acquisition containers may differ, such as blood collection tubes, which may act as temporary storage of sample.
- Sample substrates for sample deposition and containment, such as; BaF₂ and CaF₂ slides, or silicon multi-well plates for transmission measurements; low-E slides (Kevley Technologies, UK) for transfection measurements; or IREs made of germanium, diamond, silicon carbide, or silicon.
- Sample handling accessories such as; microtitre pipette and pipette tips.

3.2.2 | Spectrometer

- Commercial FTIR spectrometer, including but not limited to those produced by the following original equipment

manufacturers; Agilent Technologies, Bruker Optics, JASCO, Perkin Elmer, Thermo Fisher and Shimadzu.

- FTIR accessory units for analysis modes, such as; ATR, specular and diffuse reflectance, accessories, transmission cells and novel optical systems. These are provided by spectrometer manufacturers, as well as specific accessory providers including, but not limited to, Specac Ltd and Pike Technologies.

3.2.3 | Software

- Data acquisition software, often provided by instrument manufacturer.
- Data analysis software, available either with limited functionality within instrument manufacturer acquisition software, or by using external statistical programming packages, such as Matlab, Python, R, and custom analysis packages. Many spectral analysis functions are available as open source packages within the aforementioned programming languages; however, some are provided commercially such as the Cytospec image analysis software.

4 | PROCEDURE

The following protocols represent examples of commonly employed methods. However, it is important to note that the parameters may vary depending on the instrumentation, project aim and the needs of the analyst. Optimisation studies should be carried out to determine the best methodology for the desired application.

4.1 | Sample preparation

Biofluids should be stored in a −80°C freezer in cryovials or Eppendorf tubes after collection from pathology laboratories or biobanks. The period of time concerning long-term storage of biofluid samples does not appear to influence acquired spectral data as established by previous work exploring pre-analytical factors with human serum samples [123]. Whole biofluid samples are most often used, but some studies may involve dilution [86], filtration or centrifugation (eg, to examine the low molecular weight fraction of serum) [68, 87]. Similarly, whole blood samples are commonly subjected to centrifugation processes to extract serum or plasma constituents prior to clinical testing. Differences in centrifugation speeds and time intervals between protocols have not been found to influence spectral data, although it is important to highlight that minor spectral differences have been observed

between different serum collection tube types [123]. Volumes of required sample are dependent on mode of spectral collection. ATR analysis only requires minute volumes of samples, whereas greater volumes are often needed for transmission measurements due to the longer pathlength [10]. For quantification purposes sample concentrations are required to lie within the linear range of the Beer Lambert Law, given that the Law deviates from linearity at higher concentrations due to alterations in the absorption characteristics and refractive index of the solution [10]. Biofluids may be spotted directly onto the IRE (eg, diamond, silicon, germanium etc.) for ATR measurements, which are generally left to dry on the ATR crystal to negate the spectral interference of water and ensure sufficient sample-IRE contact [44]. Likewise, biofluid films can be developed for transmission measurements by dehydrating samples onto IR transparent substrates, such as CaF_2 and BaF_2 . Optimal sample drying procedures should be determined through preliminary analysis [9].

Examples of a typical *step-by-step sample preparation procedures* include:

1. Remove biofluid sample from -80°C storage and allow to thaw at room temperature;
2. (a) Deposit blood serum directly onto ATR substrate with an appropriate volume to fully cover the entire surface area of the ATR substrate and leave to dry for 10 minutes prior to ATR mode spectral analysis [78];
(b) (i) Dilute each serum sample with distilled water to obtain a 2-fold dilution;
(ii) Deposit 5 mL aliquots of diluted sample onto wells of a 96-well silicon plate, and dry for 30 minutes at room temperature prior to transmission mode spectral analysis [124].

4.2 | Spectral Acquisition

The acquisition of IR spectra can be achieved through standard laboratory benchtop FTIR instruments or by using more complex synchrotron facilities. For example, ATR-FTIR spectroscopy (A) and transmission FTIR microspectroscopy (B). In general, there are several collection parameters that must be selected prior to acquisition, such as the number of background/sample scans and the spectral resolution. Collection settings may be altered to suit the needs of the user. Some examples of spectral acquisition include:

4.2.1 | (A) FTIR spectroscopy

1. Open instrument software (eg, Perkin Elmer Spectrum acquisition software);
 2. Insert the appropriate sample accessory into the FTIR spectrometer depending on intended measurement technique, either ATR, transmission or transflection FTIR spectroscopy;
 3. If utilising an ATR accessory, ensure the fixed IRE crystal is thoroughly cleaned with ethanol and then distilled water and air dried prior to the acquisition of background measurements. Similarly, if using disposable IRE's [35], ensure the substrate is clean prior to spectroscopic analysis. Go to step 6.
 4. If utilising a transmission FTIR accessory, ensure IR transparent optical windows (eg, CaF_2 , BaF_2) are cleaned with isopropanol using a cotton-tipped applicator to prevent abrasive scratches. Go to step 6.
 5. If utilising a transflection FTIR accessory, ensure the reflective substrate (eg, MirrIR Low-e microscope slides, silicon) are cleaned successively with ethanol and distilled water and air dried prior to collection of background measurements.
 6. Apply instrumental settings (eg, wavenumber range of $4000\text{--}450\text{ cm}^{-1}$, 4 cm^{-1} resolution, data spacing of 1 cm^{-1} and 16 co-added scans for background and sample);
 7. Input appropriate naming convention for sample scans and select path location for files to be saved;
 8. Ensure instrument is calibrated and ready for acquisition;
- NOTE: follow accessory calibration procedure if signal is low.*
9. Acquire background measurement, with no sample present on the substrate. Ensure a background measurement is recorded prior to respective sample depositions to account for possible changes in atmospheric conditions over time.
 10. For ATR measurements, deposit biofluid sample onto the IRE and ensure the sample fully covers the exposed IRE surface area. The minimum sample thickness should be 3-4 times the depth of penetration of the evanescent wave to prevent spectral artefacts arising from interactions between evanescent waves and substrate surfaces. Go to step 13.
 11. For transmission FTIR measurements, deposit biofluid sample directly onto IR transparent optical windows for dried spectroscopic analysis and allow spotted sample to fully air dry before placing optical windows in sample holder. For wet biofluid analysis, introduce the biofluid sample into a liquid cell where the sample pathlength is determined by the defined spacer thickness. Go to step 13.
 12. For transflection FTIR measurements, deposit biofluid sample directly onto reflective substrate mounted to transflection FTIR accessory and allow

sample to fully air dry. Preliminary investigations may be required to determine optimal sample volume for specific biofluids.

13. Acquire sample spectrum, either when the sample is in liquid state or fully dried depending on experimental design. *NOTE:* Transfection FTIR spectra may promote significant Mie scattering depending on dried sample morphology and may further give rise to standing wave artefacts, hence, transfection FTIR measurements should be considered with caution.
14. Repeat steps 7-13 for biological and technical replicates, as required.

4.2.2 | (B) FTIR microspectroscopy

1. Switch on microscope and instrument;
2. Fill the detector with liquid nitrogen (N_2) to ensure cooling, if using an MCT detector;

NOTE: allow detector to stabilise for ~ 20 minutes before continuing and top up the detector with N_2 as required (eg, every 7 hours).

3. Open instrument software (eg, Bruker Opus acquisition software);
4. Apply instrumental settings (eg, wavenumber range of $4000-600\text{ cm}^{-1}$, 8 cm^{-1} resolution, 256 background scans and 128 co-added sample scans);
5. Place the sample onto the microscope stage and focus the microscope as explained in the microscope instructions manual;
6. Move to a sample-free area and check signal quality by adjusting the stage position to bring the substrate surface into focus;
7. Choose the aperture size (eg, $10 \times 10\ \mu\text{m}$);

NOTE: use the smallest possible aperture size in order to acquire spectra with high SNR.

8. Select a clean sample-free area of the substrate and acquire background spectrum;
9. Move the joystick to move the sample slide around the microscope stage and identify points of interest.

NOTE: background measurements should be taken at regular intervals to account for atmospheric changes.

10. Acquire sample measurement (eg, point spectra or image map);

NOTE: ensure measurement does not exceed time frame for liquid N_2 cooling.

11. Save spectra and/or image map until data processing.

4.3 | Data pre-processing

When approaching the pre-processing of a set of data, it must be kept in mind that the outcome of each step depends on the specific features of the dataset, but also on the data acquisition settings. It is always advised to perform spectral correction to eliminate unwanted minor spectral interferences when analysing a considerable set of data, such as baseline variation and spectral noise. However, the selection and order of pre-processing parameters depends entirely on the dataset chosen for analysis.

- Depending on the size of the dataset and the chosen steps, pre-processing can require from less than an hour to several hours. It is often a *trial-and-error approach*, which can be time consuming. By using a grid search, it is possible to examine several combinations of pre-processing to individuate the optimal method.

1. Selection of the spectral region of interest.
2. Baseline correction. SG differentiation (first or second), rubberband, polynomial or EMSC.
3. Normalisation of the spectra. Min-max (0-1), vector or normalisation to Amide I (or II) band.
4. Smoothing of the spectra. SG filtering, wavelet denoising or local polynomial fitting with Gaussian weighting.
5. Binning of the spectra.

Further theoretical details about the abovementioned pre-processing techniques can be examined in the Data Processing section of the Experimental Design.

4.4 | Data analysis

The goal of data analysis for diagnostic purposes is represented by great outcomes in discriminating between diseased and healthy patients. Exploratory analysis and classification are both valid methods to achieve further knowledge on the set of data. Exploratory analysis techniques are useful to obtain information on discriminating features in the dataset through interpretation of plots; while, when looking at building a strong classification model for future clinical applications, it is important to find the approach that guarantees the best and most reliable performance.

- Depending on the size of the dataset and the chosen steps, data processing can require from several hours to multiple days.

- A. Exploratory analysis. PCA (or similar) for correlation or pattern analysis.
- B. Classification analysis. PLS-DA, SVM or RF.

A typical *step-by-step classification process* carried out with a classification software or through programming, would be the following:

1. Split the dataset in training and test set, according to the size of the set of data;

2. Choose your supervised classification method (eg, RF, PLS-DA, SVM etc.);
3. Perform bootstrapping and assess the correct number of resampling iterations required;
4. Perform some preliminary classifications with a k-fold cross-validation procedure (eg, 5-fold) and determine optimal model tuning parameters;
5. Perform a classification with a k-fold cross-validation procedure, the correct number of resampling and the appropriate tuning.

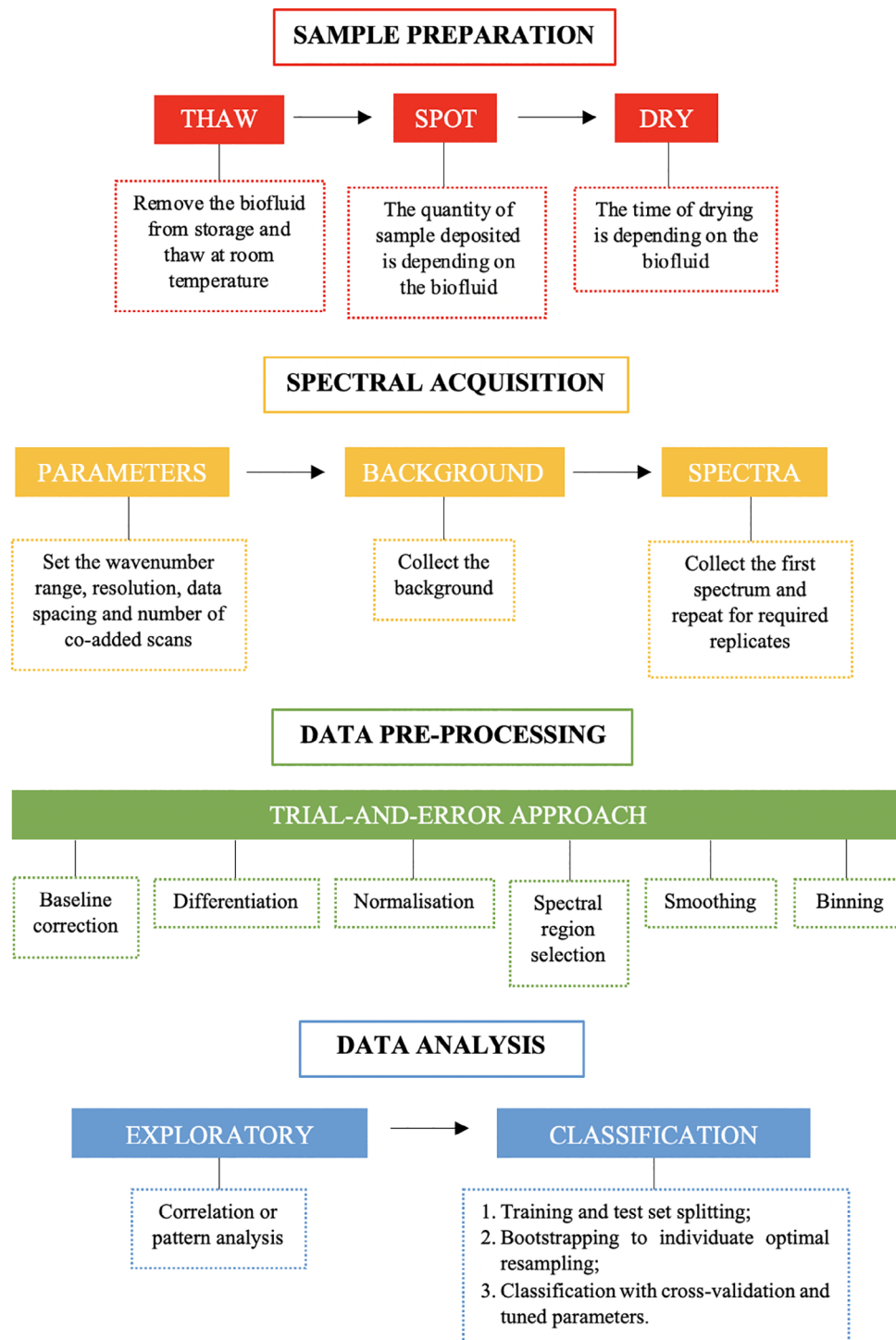


FIGURE 2 Stepwise process of each stage required in biofluid analysis, from sample preparation to data analysis

Further theoretical details about the abovementioned data analysis techniques can be examined in the Data Processing section of the Experimental Design.

Figure 2 summarises all the general steps that need to be performed from sample preparation to data analysis.

5 | FUTURE APPLICATIONS

Infrared spectroscopy shows significant promise in the medical field aiding in the improvement of patient diagnosis, prognosis and medical observation while enabling clinical management of disease, directing type and duration of treatment. The application of spectroscopic techniques for “liquid biopsies” is a rapidly progressing field in diagnostics. Traditional diagnostic methods typically require invasive tests such as surgical biopsies or medical imaging techniques which are expensive, require specialised technicians and consist of subjective histopathological examinations. Furthermore, patients who are deemed at risk of an underlying malignancy are subjected to significant waiting times to receive medical imaging tests for diagnosing disease such as cancer [125]. Cancer screening tests have proven to be effective for early diagnosis of certain cancers allowing for improved survival rates for these patients [126–129]. However, there is a distinctive gap in cancer diagnostics, with routine screening tests currently only available for a small proportion of cancers. Numerous cancers, particularly those involving internal body organs, do not employ routine screening programmes where late stage diagnosis is common and associated with advanced disease and poorer patient survival. Therefore, infrared spectroscopy has significant potential to provide rapid and low-cost screening tools for several cancers, which could either be integrated with established diagnostic pathways, or adopted within primary care facilities to identify patients at an early disease stage who require further clinical investigations. To date, numerous proof-of-concept studies have demonstrated the potential of infrared spectroscopy as a powerful analytical clinical tool for the diagnosis of cancer using blood samples. ATR-FTIR spectroscopy has been utilised to diagnose ovarian [130], brain [78], melanoma [47], and breast cancers [38] to name a few. Furthermore, FTIR spectroscopy has potential to monitor and predict the therapeutic response in cancer treatment aiding towards the delivery of precision medicine to promote successful treatment for the patient [131].

Infrared spectroscopy has demonstrated to be capable of the accurate identification of bacteria with significant diagnostic capabilities in the field of microbiology [132–134]. The limitations of the current

methods of pathogen identification have been well documented [135, 136]. Gram-negative bacilli clones are responsible for the most frequent healthcare associated infection outbreaks [137]. Recently, Martak *et al.* demonstrated that FTIR spectroscopy was able to identify and type bacterial clones within short time frames to allow infection controls to be quickly implemented [138]. Similarly, FTIR spectroscopy has allowed various *Pseudomonas*, *Escherichia* and *Bacillus* strains to be identified at both strain and species level [139]. Furthermore, infrared spectroscopy has been utilised for guiding treatment for infections well as in the determination of antibiotic resistance [140, 141]. Current diagnostic procedures for infections and sepsis are insufficient and fail to diagnose patients at early stages causing delays in the diagnosis and initiation of treatment. There is great potential for translation of infrared spectroscopy into clinical environments to allow for the rapid identification of infections in both the community and hospital care settings.

Viruses are typically identified by the detection of antibodies and antigens using serological assays and molecular polymerase chain reaction based assays which are time consuming, expensive, require bulky equipment and specialised training [142]. Numerous studies have demonstrated infrared spectroscopy to overcome these limitations of conventional virus diagnosis by providing label-free, non-destructive analytical tool which requires little sample preparation [143]. FTIR spectroscopy has been evaluated for the detection of hepatitis B and C in serum samples [144] and in the determination of human immunodeficiency virus (HIV) in plasma samples [145]. ATR-FTIR spectroscopy has been applied as a point-of-care test for identifying malaria parasites in whole blood samples in malaria-endemic countries [146]. Furthermore, with the increasing risk of pandemic viruses, like Covid-19, spectroscopic methods can also be used to gain an understanding into the evolution of new and existing viruses by understanding how they mutate and spread to allow for their effective control and eradication [147].

FTIR spectroscopy techniques can detect the early stages of disease including before clinical symptoms have presented by analysing samples at the molecular level. This has been particularly well demonstrated in the case of degenerative neurological diseases including Alzheimer's disease. ATR-FTIR analysis is shown to be able to distinguish between different types of dementia and neurodegenerative diseases using blood samples [148]. Mordechai *et al.* found that FTIR analysis of plasma and white blood cell samples whole blood samples can be used as an early diagnosis tool for

Alzheimer's disease, improving on the current subjective tests used in the diagnosis [149]. FTIR spectroscopy has also been investigated as a tool for the identification of traumatic axonal injury [150], Parkinson's disease [151] and in differentiating relapsing-remitting MS from clinically isolated syndrome (CIS) as well as identifying those CIS who will progress to relapsing-remitting MS. [23]

Infrared spectroscopy has clearly demonstrated significant potential as a rapid, economical and clinically effective diagnostic tool suitable for several diagnostic pathways within our current healthcare system. Nevertheless, infrared spectroscopy remains in the preliminary stage on the roadmap to clinical translation, and several challenges must be addressed before clinical adoption of infrared spectroscopic technologies within healthcare can be realised. Firstly, future research must demonstrate the clinical utility of infrared spectroscopy in large, multi-centre studies on both retrospective and prospective data with significantly greater sample populations to validate the potential of proposed technologies. Secondly, comprehensive health economic studies must be performed for proposed clinical scenarios to demonstrate the health economic benefits associated with introduction of infrared spectroscopy within clinical environments. Lastly, there is a dire need for implementation of standardised methodologies for different infrared spectroscopic techniques to ensure adopted technologies are simple to use and robust for clinical laboratories. To this extent, we hope this tutorial has provided valuable insights and discussions into optimal methodologies for infrared spectroscopy of biofluids, which we envisage will disrupt the current diagnostic pathway and transform the way in which healthcare is delivered in the clinical environment.

ACKNOWLEDGMENTS

The authors would like to acknowledge EPSRC (EP/L505080/1, EP/L015595/1) and CRUK (A28345) for funding.

CONFLICT OF INTEREST

There are no conflicts to declare.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no new data was created or analysed within this tutorial.

ORCID

Ashton G. Theakstone  <https://orcid.org/0000-0001-8754-5094>

Alexandra Sala  <https://orcid.org/0000-0001-6417-9706>

TWITTER

Matthew J. Baker  @Chemistrybaker

REFERENCES

- [1] Z. Kaszowska, K. Malek, M. Pańczyk, A. Mikołajska, *Vibrational Spectroscopy* **2013**, 65, 1.
- [2] M. Sponheimer, C. M. Ryder, H. Fewlass, E. K. Smith, W. J. Pestle, S. Talamo, *Sci. Rep.* **2019**, 9, 1.
- [3] Z. Bacsik, J. Mink, G. Keresztury, *Applied Spectroscopy Reviews* **2005**, 40, 327.
- [4] K. Gajjar, J. Trevisan, G. Owens, P. J. Keating, N. J. Wood, H. F. Stringfellow, P. L. Martin-Hirsch, F. L. Martin, *Analyst* **2013**, 138, 3917.
- [5] J. Ollesch, S. L. Drees, H. M. Heise, T. Behrens, T. Brüning, K. Gerwert, *Analyst* **2013**, 138, 4092.
- [6] M.-M. Blum, H. John, *Drug Test. Anal.* **2012**, 4, 298.
- [7] A. Rohman, A. Windarsih, E. Lukitaningsih, M. Rafi, K. Betania, N. A. Fadzillah, *Biomed. Spectrosc. Imaging* **2020**, 8, 55.
- [8] B. C. Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC Press, Boca Raton, FL **2011**.
- [9] M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszewski, G. Thiéfin, G. D. Sockalingum, *Chem. Soc. Rev.* **2016**, 45, 1803.
- [10] M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner, F. L. Martin, *Nat. Protoc.* **2014**, 9, 1771.
- [11] M. Jackson, H. H. Mantsch, *Crit. Rev. Biochem. Mol. Biol.* **1995**, 30, 95.
- [12] H. Fabian, P. Lasch, D. Naumann, *J. Biomed. Opt.* **2005**, 10, 031103.
- [13] J. N. Adkins, S. M. Varnum, K. J. Auberry, R. J. Moore, N. H. Angell, R. D. Smith, D. L. Springer, J. G. Pounds, *Mol. Cell. Proteomics* **2002**, 1, 947.
- [14] T. D. Veenstra, *Mol. Cell. Proteomics* **2005**, 4, 409.
- [15] E. F. Petricoin, C. Belluco, R. P. Araujo, L. A. Liotta, *Nat. Rev. Cancer* **2006**, 6, 961.
- [16] L. Lovergne, P. Bouzy, V. Untereiner, R. Garnotel, M. J. Baker, G. Thiéfin, G. D. Sockalingum, *Faraday Discuss.* **2015**, 187, 521.
- [17] I. C. C. Ferreira, E. M. G. Aguiar, A. T. F. Silva, L. L. D. Santos, L. Cardoso-Sousa, T. G. Araújo, D. W. Santos, L. R. Goulart, R. Sabino-Silva, Y. C. P. Maia, *J. Oncol.* **2020**, 2020, 1.
- [18] P. D. Lewis, K. E. Lewis, R. Ghosal, S. Bayliss, A. J. Lloyd, J. Wills, R. Godfrey, P. Kloer, L. A. Mur, *BMC Cancer* **2010**, 10, 640.
- [19] I. Maitra, C. L. M. Morais, K. M. G. Lima, K. Ashton, R. S. Date, F. L. Martin, *Analyst* **2019**, 144, 7447.
- [20] K. V. Oliver, A. Vilasi, A. Maréchal, S. H. Moochhala, R. J. Unwin, P. R. Rich, *Sci. Rep.* **2016**, 6, 34737.
- [21] K.-Z. Liu, T. C. Dembinski, H. H. Mantsch, *Am. J. Obstet. Gynecol.* **1998**, 178, 234.
- [22] V. Untereiner, G. Dhruvananda Sockalingum, R. Garnotel, C. Gobinet, F. Ramaholimihaso, F. Ehrhard, M.-D. Diebold, G. Thiéfin, *J. Biophotonics* **2014**, 7, 241.

- [23] D. Yonar, L. Ocek, B. I. Tiftikcioglu, Y. Zorlu, F. Severcan, *Sci. Rep.* **2018**, *8*, 1025.
- [24] S. Abbas, N. Simsek Ozek, S. Emri, D. Koksall, M. Severcan, F. Severcan, *J. Biomed. Opt.* **2018**, *23*, 105003.
- [25] H. M. Heise, R. Marbach, T. Koschinsky, F. A. Gries, *Appl. Spectrosc.* **1994**, *48*, 85.
- [26] J. Wang, M. Sowa, H. H. Mantsch, A. Bittner, H. M. Heise, *TrAC Trends Anal. Chem.* **1996**, *15*, 286.
- [27] D. Rohleder, G. Kocherscheidt, K. Gerber, W. Kiefer, W. Köhler, J. Möcks, W. Petrich, *J. Biomed. Opt.* **2005**, *10*, 031108.
- [28] R. A. Shaw, S. Kotowich, M. Leroux, H. H. Mantsch, *Ann. Clin. Biochem.* **1998**, *35*, 624.
- [29] E. Diessel, S. Willmann, P. Kamphaus, R. Kurte, U. Damm, H. M. Heise, *Appl. Spectrosc.* **2004**, *58*, 442.
- [30] E. Diessel, P. Kamphaus, K. Grothe, R. Kurte, U. Damm, H. M. Heise, *Appl. Spectrosc.* **2005**, *59*, 442.
- [31] V. R. Kondepati, U. Damm, H. M. Heise, *Appl. Spectrosc.* **2006**, *60*, 920.
- [32] A. Staib, B. Dolenko, D. J. Fink, J. Früh, A. E. Nikulin, M. Otto, M. S. Pessin-Minsley, O. Quarder, R. Somorjai, U. Thienel, G. Werner, W. Petrich, *Clin. Chim. Acta* **2001**, *308*, 79.
- [33] W. Petrich, K. B. Lewandrowski, J. B. Muhlestein, M. E. Hammond, J. L. Januzzi, E. L. Lewandrowski, R. R. Pearson, B. Dolenko, J. Früh, M. Haass, M. M. Hirschl, W. Köhler, R. Mischler, J. Möcks, J. Ordóñez-Llanos, O. Quarder, R. Somorjai, A. Staib, C. Sylvén, G. Werner, R. Zerback, *Analyst* **2009**, *134*, 1092.
- [34] J. R. Hands, G. Clemens, R. Stables, K. Ashton, A. Brodbelt, C. Davis, T. P. Dawson, M. D. Jenkinson, R. W. Lea, C. Walker, M. J. Baker, *J. Neurooncol* **2016**, *127*, 463.
- [35] H. J. Butler, P. M. Brennan, J. M. Cameron, D. Finlayson, M. G. Hegarty, M. D. Jenkinson, D. S. Palmer, B. R. Smith, M. J. Baker, *Nat. Commun.* **2019**, *10*, 1.
- [36] J. Backhaus, R. Mueller, N. Formanski, N. Szlama, H.-G. Meerpohl, M. Eidt, P. Bugert, *Vib. Spectrosc.* **2010**, *52*, 173.
- [37] F. Elmi, A. F. Movaghar, M. M. Elmi, H. Alinezhad, N. Nikbakhsh, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2017**, *187*, 87.
- [38] V. E. Sitnikova, M. A. Kotkova, T. N. Nosenko, T. N. Kotkova, D. M. Martynova, M. V. Uspenskaya, *Talanta* **2020**, *214*, 120857.
- [39] J. Ollesch, M. Heinze, H. M. Heise, T. Behrens, T. Brüning, K. Gerwert, *J. Biophotonics* **2014**, *7*, 210.
- [40] A. C. O. Neves, P. P. Silva, C. L. M. Morais, C. G. Miranda, J. C. O. Crispim, K. M. G. Lima, *RSC Adv.* **2016**, *6*, 99648.
- [41] D. K. R. Medipally, D. Cullen, V. Untereiner, G. D. Sockalingum, A. Maguire, T. N. Q. Nguyen, J. Bryant, E. Noone, S. Bradshaw, M. Finn, M. Dunne, A. M. Shannon, J. Armstrong, A. D. Meade, F. M. Lyng, *Ther. Adv. Med. Oncol.* **2020**, *12*, 175883592091849.
- [42] C. Petitbois, A.-M. Melin, A. Perromat, G. Cazorla, G. Délérís, *J. Lab. Clin. Med.* **2000**, *135*, 210.
- [43] C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker, N. Stone, *Analyst* **2009**, *134*, 1029.
- [44] J. M. Cameron, H. J. Butler, D. S. Palmer, M. J. Baker, *J. Biophotonics* **2018**, *11*, e201700299.
- [45] E. Socha, M. Koba, P. Kośliński, *Amino Acids* **2019**, *51*, 367.
- [46] K. Spalding, F. Bonnier, C. Bruno, H. Blasco, R. Board, I. Benz-De Bretagne, H. J. Byrne, H. J. Butler, I. Chourpa, P. Radhakrishnan, M. J. Baker, *Vibrational Spectroscopy* **2018**, *99*, 50.
- [47] R. Minnes, M. Nissinmann, Y. Maizels, G. Gerlitz, A. Katzir, Y. Raichlin, *Sci. Rep.* **2017**, *7*, 4381.
- [48] S. Gok, O. Z. Aydin, Y. S. Sural, F. Zorlu, U. Bayol, F. Severcan, *J. Biophotonics* **2016**, *9*, 967.
- [49] D. C. Caixeta, E. M. G. Aguiar, L. Cardoso-Sousa, L. M. D. Coelho, S. W. Oliveira, F. S. Espindola, L. Raniero, K. T. B. Crosara, M. J. Baker, W. L. Siqueira, R. Sabino-Silva, *PLoS One* **2020**, *15*, e0223461.
- [50] J. Sulé-Suso, D. Skingsley, G. D. Sockalingum, A. Kohler, G. Kegelaer, M. Manfait, A. J. El Haj, *Vibrational Spectroscopy* **2005**, *38*, 179.
- [51] A. Roth, F. Dornuf, O. Klein, D. Schneditz, H. Hafner-Gießauf, W. Mäntele, *Anal. Bioanal. Chem.* **2012**, *403*, 391.
- [52] M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, M. Miljković, C. Krafft, J. Popp, *J. Biophotonics* **2013**, *6*, 855.
- [53] A. Schwaighofer, M. Brandstetter, B. Lendl, *Chem. Soc. Rev.* **2017**, *46*, 5903.
- [54] A. Schwaighofer, M. Montemurro, S. Freitag, C. Kristament, M. J. Culzoni, B. Lendl, *Anal. Chem.* **2018**, *90*, 7072.
- [55] C. K. Akhgar, G. Ramer, M. Žbik, A. Trajnerowicz, J. Pawluczyk, A. Schwaighofer, B. Lendl, *Anal. Chem.* **2020**, *92*, 9901.
- [56] K. Isensee, N. Müller, A. Pucci, W. Petrich, *Analyst* **2018**, *143*, 6025.
- [57] A. Nijssen, S. Koljenović, T. C. B. Schut, P. J. Caspers, G. J. Puppels, *J. Biophotonics* **2009**, *2*, 29.
- [58] L. B. Leal, M. S. Nogueira, R. A. Canevari, L. F. C. S. Carvalho, *Photodiagnosis Photodyn. Ther.* **2018**, *24*, 237.
- [59] M. De, S. Rana, H. Akpınar, O. R. Miranda, R. R. Arvizo, U. H. F. Bunz, V. M. Rotello, *Nat. Chem.* **2009**, *1*, 461.
- [60] M. M. Lubran, *Ann. Clin. Lab. Sci.* **1978**, *8*, 106.
- [61] K. Zheng, L. Wu, Z. He, B. Yang, Y. Yang, *Measurement* **2017**, *112*, 16.
- [62] P. Atkins, J. d. Paula, *Physical Chemistry*, Oxford University Press, Oxford **2006**.
- [63] A. Gupta, S. L. Stockham, *Vet. Clin. Pathol.* **2014**, *43*, 422.
- [64] G-Biosciences *Biuret Protein Assay—Handbook*, **2018**. https://www.gbiosciences.com/image/pdfs/protocol/402B_Biuret_Protein_Assay.pdf
- [65] C. K. Dixit, S. K. Vashist, B. D. MacCraith, R. O'Kennedy, *Nat. Protoc.* **2011**, *6*, 439.
- [66] B. D. Hames, *Gel Electrophoresis of Proteins—A Practical Approach*, New York, Oxford University Press, **1998**.
- [67] T. Wehr, W. Strong, S. Adachi, A. Paulus, *Electrophoresis, Bio-Rad*, **2008**. https://www.bio-rad.com/webroot/web/pdf/lsl/literature/Bulletin_5719A.pdf
- [68] F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L. F. C. S. Carvalho, D. Bertrand, M. J. Baker, H. J. Byrne, I. Chourpa, *Analyst* **2017**, *142*, 1285.
- [69] C. Hughes, G. Clemens, B. Bird, T. Dawson, K. M. Ashton, M. D. Jenkinson, A. Brodbelt, M. Weida, E. Fotheringham, M. Barre, J. Rowlette, M. J. Baker, *Sci. Rep.* **2016**, *6*, 20173.
- [70] H. J. Butler, J. M. Cameron, C. A. Jenkins, G. Hithell, S. Hume, N. T. Hunt, M. J. Baker, *Clin. Spectrosc.* **2019**, *1*, 100003.

- [71] M. J. Baker, C. S. Hughes, K. A. Hollywood, *Biophotonics: Vibrational Spectroscopic Diagnostics*, San Rafael, CA, Morgan & Claypool Publishers, **2016**.
- [72] M. Milosevic, *Appl. Spectrosc.* **2013**, *67*, 126.
- [73] M. Milosevic, *Appl. Spectrosc. Rev.* **2004**, *39*, 365.
- [74] H. Schumacher, U. Künzelmann, B. Vasilev, K.-J. Eichhorn, J. W. Bartha, *Appl. Spectrosc.* **2010**, *64*, 1022.
- [75] E. Karabudak, R. Kas, W. Ogieglo, D. Rafieian, S. Schlautmann, R. G. H. Lammertink, H. J. G. E. Gardeniers, G. Mul, *Anal. Chem.* **2013**, *85*, 33.
- [76] M. Koc, E. Karabudak, *Applied Spectroscopy Reviews* **2017**, *53*, 420.
- [77] J. Haas, A. Muller, L. Sykora, B. Mizaikoff, *Analyst* **2019**, *144*, 3398.
- [78] J. R. Hands, K. M. Dorling, P. Abel, K. M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M. D. Jenkinson, R. W. Lea, C. Walker, M. J. Baker, *J. Biophotonics* **2014**, *7*, 189.
- [79] D. Perez-Guaita, Z. Richardson, P. Heraud, B. Wood, *Anal. Chem.* **2020**, *92*, 2409.
- [80] J. Dubois, R. A. Shaw, *Anal. Chem.* **2004**, *76*, 360A.
- [81] Y. Li, Q. Yang, M. Li, Y. Song, *Sci. Rep.* **2016**, *6*, 24628.
- [82] W. Bou Zeid, J. Vicente, D. Brutin, *Colloids Surf. A Physicochem. Eng. Asp.* **2013**, *432*, 139.
- [83] J. Grdadolnik, *Acta Chim. Slov.* **2002**, *49*, 631.
- [84] A. Sala, K. E. Spalding, K. M. Ashton, R. Board, H. J. Butler, T. P. Dawson, D. A. Harris, C. S. Hughes, C. A. Jenkins, M. D. Jenkinson, D. S. Palmer, B. R. Smith, C. A. Thornton, M. J. Baker, *J. Biophotonics* **2020**, *13*, 1.
- [85] M. R. Alcaráz, A. Schwaighofer, H. Goicoechea, B. Lendl, *Anal. Bioanal. Chem.* **2016**, *408*, 3933.
- [86] L. Lovregne, G. Clemens, V. Untereiner, R. A. Lukaszewski, G. D. Sockalingum, M. J. Baker, *Anal. Methods* **2015**, *7*, 7140.
- [87] C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N. W. Clarke, P. Gardner, *J. Biophotonics* **2014**, *7*, 180.
- [88] R. Bhargava, *Anal. Bioanal. Chem.* **2007**, *389*, 1155.
- [89] B. Bird, M. Miljković, S. Remiszewski, A. Akalin, M. Kon, M. Diem, *Lab. Invest.* **2012**, *92*, 1358.
- [90] C. Hughes, J. Iqbal-Wahid, M. Brown, J. H. Shanks, A. Eustace, H. Denley, P. J. Hoskin, C. West, N. W. Clarke, P. Gardner, *J. Biophotonics* **2013**, *6*, 73.
- [91] I. I. Patel, D. A. Shearer, S. W. Fogarty, N. J. Fullwood, L. Quaroni, F. L. Martin, J. Weisz, *Cancer Biol. Ther.* **2014**, *15*, 225.
- [92] G. Clemens, B. J. Bird, M. J. Weida, J. Rowlette, M. J. Baker, *Spectrosc. Eur.* **2015**, *26*, 14.
- [93] P. R. Griffiths, J. A. de Haseth, *Fourier Transform Infrared Spectroscopy*, Hoboken, New Jersey, John Wiley & Sons, **2007**.
- [94] W. D. Duncan, G. P. Williams, *Appl. Optics* **1983**, *22*, 2914.
- [95] M. R. Kole, R. K. Reddy, M. V. Schulmerich, M. K. Gelber, R. Bhargava, *Anal. Chem.* **2012**, *84*, 10366.
- [96] H. J. Butler, B. R. Smith, R. Fritzsche, P. Radhakrishnan, D. S. Palmer, M. J. Baker, *Analyst* **2018**, *143*, 6121.
- [97] B. R. Smith, M. J. Baker, D. S. Palmer, *Chemom. Intel. Lab. Syst.* **2018**, *172*, 33.
- [98] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, F. L. Martin, *Nat. Protoc.* **2016**, *11*, 664.
- [99] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott, F. L. Martin, *Analyst* **2012**, *137*, 3202.
- [100] B. K. Alsberg, A. M. Woodward, M. K. Winson, J. Rowland, D. B. Kell, *Analyst* **1997**, *122*, 645.
- [101] P. Lasch, *Chemom. Intel. Lab. Syst.* **2012**, *117*, 100.
- [102] J. M. Cameron, H. J. Butler, B. R. Smith, M. G. Hegarty, M. D. Jenkinson, K. Syed, P. M. Brennan, K. Ashton, T. Dawson, D. S. Palmer, M. J. Baker, *Analyst* **2019**, *144*, 6736.
- [103] J. Skogholt, K. H. Liland, U. G. Indahl, *J. Raman Spectrosc.* **2019**, *50*, 407.
- [104] N. K. Afseth, A. Kohler, *Chemom. Intel. Lab. Syst.* **2012**, *117*, 92.
- [105] P. Lasch, W. Haensch, D. Naumann, M. Diem, *Biochim. Biophys. Acta* **2004**, *1688*, 176.
- [106] D. Naumann, in *Infrared Spectroscopy of Cells, Tissues, and Biofluids* (Ed: G. C. K. Roberts), Springer Berlin Heidelberg, Berlin, Heidelberg **2013**, p. 1057.
- [107] Z. Zhang, F. Murtagh, S. V. Van Poucke, S. Lin, P. Lan, *Ann. Transl. Med.* **2017**, *5*, 9.
- [108] N. Pizzi, L. P. Choo, J. Mansfield, M. Jackson, W. C. Halliday, H. H. Mantsch, R. L. Somorjai, *Artif. Intell. Med.* **1995**, *7*, 67.
- [109] M. Khanmohammadi, A. B. Garmarudi, K. Ghasemi, H. K. Jaliseh, A. Kaviani, *Med. Oncol.* **2009**, *26*, 292.
- [110] B. R. Smith, K. M. Ashton, A. Brodbelt, T. Dawson, M. D. Jenkinson, N. T. Hunt, D. S. Palmer, M. J. Baker, *Analyst* **2016**, *141*, 3668.
- [111] A. Sala, D. J. Anderson, P. M. Brennan, H. J. Butler, J. M. Cameron, M. D. Jenkinson, C. Rinaldi, A. G. Theakstone, M. J. Baker, *Cancer Lett.* **2020**, *477*, 122.
- [112] G. C. Cawley, N. L. C. Talbot, *J. Mach. Learn. Res.* **2010**, *11*, 2079.
- [113] Q.-B. Li, L.-M. Yang, X.-F. Ling, J.-S. Wang, X.-S. Zhou, J.-S. Shi, J.-G. Wu, *Guang Pu Xue Yu Guang Pu Fen Xi* **2004**, *24*, 414.
- [114] J. Ali, R. Khan, N. Ahmad, I. Maqsood, *International Journal of Computer Science Issues* **2012**, *9*, 272.
- [115] D. S. Palmer, N. M. O'Boyle, R. C. Glen, J. B. O. Mitchell, *J. Chem. Inf. Model.* **2007**, *47*, 150.
- [116] M. Otto, *Pattern Recognition and Classification*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim **2016**, p. 135.
- [117] B.-C. Deng, Y.-H. Yun, Y.-Z. Liang, D.-S. Cao, Q.-S. Xu, L.-Z. Yi, X. Huang, *Anal. Chim. Acta* **2015**, *880*, 32.
- [118] C. Márquez, M. I. López, I. Ruisánchez, M. P. Callao, *Talanta* **2016**, *161*, 80.
- [119] A. M. Dickens, J. R. Larkin, J. L. Griffin, A. Cavey, L. Matthews, M. R. Turner, G. K. Wilcock, B. G. Davis, T. D. W. Claridge, J. Palace, D. C. Anthony, N. R. Sibson, *Neurology* **2014**, *83*, 1492.
- [120] F. Zhang, H. L. Kaufman, Y. Deng, R. Drabier, *BMC Med. Genomics* **2013**, *6*, S4.
- [121] P. S. Gromski, H. Muhamadali, D. I. Ellis, Y. Xu, E. Correa, M. L. Turner, R. Goodacre, *Anal. Chim. Acta* **2015**, *879*, 10.
- [122] D. Granato, P. Putnik, D. B. Kovačević, J. S. Santos, V. Calado, R. S. Rocha, A. G. D. Cruz, B. Jarvis, O. Y. Rodionova, A. Pomerantsev, *Compr. Rev. Food Sci. Food Saf.* **2018**, *17*, 663.
- [123] J. M. Cameron, H. J. Butler, D. J. Anderson, L. Christie, L. Confield, K. E. Spalding, D. Finlayson, S. Murray, Z. Panni,

- C. Rinaldi, A. Sala, A. G. Theakstone, M. J. Baker, *Vib. Spectrosc.* **2020**, *109*, 103092.
- [124] X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon, G. D. Sockalingum, *Transl. Res.* **2013**, *162*, 279.
- [125] NHS England, *Diagnostic Imaging Dataset Release*, England, NHS, **2017**.
- [126] M. Quinn, P. Babb, J. Jones, E. Allen, *BMJ* **1999**, *318*, 904.
- [127] S. M. Collin, R. M. Martin, C. Metcalfe, D. Gunnell, P. C. Albertsen, D. Neal, F. Hamdy, P. Stephens, J. A. Lane, R. Moore, J. Donovan, *Lancet Oncol.* **2008**, *9*, 445.
- [128] R. G. Blanks, *BMJ* **2000**, *321*, 665.
- [129] R. J. C. Steele, *BMJ* **2004**, *329*, 133.
- [130] G. Theophilou, K. M. G. Lima, P. L. Martin-Hirsch, H. F. Stringfellow, F. L. Martin, *Analyst* **2016**, *141*, 585.
- [131] S. Kalmodia, S. Parameswaran, W. Yang, C. J. Barrow, S. Krishnakumar, *Sci. Rep.* **2015**, *5*, 16649.
- [132] D. Naumann, D. Helm, H. Labischinski, *Nature* **1991**, *351*, 81.
- [133] D. Naumann, S. Keller, D. Helm, C. Schultz, B. Schrader, *J. Mol. Struct.* **1995**, *347*, 399.
- [134] P. Lasch, D. Naumann, *Infrared Spectroscopy in Microbiology*, Chichester, John Wiley and Sons, **2015**, p. 1.
- [135] L. Bissonnette, M. G. Bergeron, *Clin. Microbiol. Infect.* **2010**, *16*, 1044.
- [136] J. Vila, M. D. Gomez, M. Salavert, J. Borsch, *Enferm. Infecc. Microbiol. Clin.* **2017**, *35*, 41.
- [137] A. Y. Peleg, D. C. Hooper, *N. Engl. J. Med.* **2010**, *362*, 1804.
- [138] D. Martak, B. Valot, M. Sauget, P. Cholley, M. Thouverez, X. Bertrand, D. Hocquet, *Front. Microbiol.* **2019**, *10*, 1.
- [139] J. Lee, M. S. Ahn, Y.-L. Lee, E. Y. Jie, S.-G. Kim, S. W. Kim, *J. Appl. Microbiol.* **2019**, *126*, 864.
- [140] U. Sharaha, E. Rodriguez-Diaz, K. Riesenber, I. J. Bigio, M. Huleihel, A. Salman, *Anal. Chem.* **2017**, *89*, 8782.
- [141] A. Salman, U. Sharaha, E. Rodriguez-Diaz, E. Shufan, K. Riesenber, I. J. Bigio, M. Huleihel, *Analyst* **2017**, *142*, 2136.
- [142] P. Rajapaksha, A. Elbourne, S. Gangadoo, R. Brown, D. Cozzolino, J. Chapman, *Analyst* **2019**, *144*, 396.
- [143] M. C. D. Santos, C. L. M. Morias, Y. M. Nascimento, *Trends Anal. Chem.* **2017**, *97*, 244.
- [144] S. Roy, D. Perez-Guaita, S. Bowden, P. Heraud, B. R. Wood, *Clinical Spectroscopy.* **2020**, *1*, 1.
- [145] C. Pizarro, I. Esteban-Diez, I. Arenzana-Ramila, J. M. Gonzalez-Saiz, *J. Biophotonics* **2018**, *11*, 1.
- [146] S. Roy, D. Perez-Guaita, D. W. Andrew, J. S. Richards, D. McNaughton, P. Heraud, B. R. Wood, *Anal. Chem.* **2017**, *89*, 5238.
- [147] A. Sakudo, Y. Suganuma, T. Kobayashi, T. Onodera, K. Ikuta, *Biochem. Biophys. Res. Commun.* **2006**, *341*, 279.
- [148] M. Paraskevaidi, C. L. M. Morais, K. M. G. Lima, J. S. Snowden, J. A. Saxon, A. M. T. Richardson, M. Jones, D. M. A. Mann, D. Allsop, P. L. Martin-Hirsch, F. L. Martin, *Proc. Natl. Acad. Sci.* **2017**, *114*, E7929.
- [149] S. Mordechai, E. Shufan, B. S. Porat Katz, A. Salman, *Analyst* **2017**, *142*, 1276.
- [150] J. Zhang, P. Huang, Z. Wang, H. Dong, *Biosci. Rep.* **2017**, *37*, BSR20170720.
- [151] X. Wang, Q. Wu, C. Li, Y. Zhou, F. Xu, L. Zong, S. Ge, *Infrared Phys. Technol.* **2020**, *106*, 103279.

How to cite this article: Theakstone AG, Rinaldi C, Butler HJ, et al. Fourier-transform infrared spectroscopy of biofluids: A practical approach. *Translational Biophotonics*. 2021;3:e202000025. <https://doi.org/10.1002/tbio.202000025>