1

VIBRATIONAL SPECTROSCOPY IN MICROBIOLOGY AND MEDICAL DIAGNOSTICS

Dieter Naumann

Robert-Koch Institut, Berlin, Germany

Infrared (IR) and Raman spectroscopy are relatively old spectroscopic modalities that provide pictures of the molecular vibrations performed by molecules. Since the early experiments of Herschel, who more than 200 years ago discovered heat transporting radiation beyond the range of visible light, it took some 80 years until the first IR spectrum of an organic liquid was obtained. Since then, IR spectroscopy developed into the "workhorse" of vibrational spectroscopy in fundamental science and the industries, while Raman spectroscopy, discovered only in 1928, was initially restricted to a few laboratories in the academic area. Infrared and Raman spectroscopy, though fundamentally different in experimental design and physical background, give complementary information on molecular vibrations and should ideally be used together to attain access to the totality of all vibrational modes of a given molecule.

It has been only for the last two or three decades that both types of vibrational spectroscopy have been used systematically for the more complex building blocks of biological systems or even intact cells, tissues, and biological fluids. These scientific endeavors were facilitated by technological innovations such as the advent of Fourier transform (FT)-IR spectrometers, powerful low-cost lasers in the near-IR region, sensitive detector systems, and rapid low-cost computers, which favored new developments such as focal plane array detectors for true IR imaging systems or surface-enhanced Raman techniques based on nanostructured materials as optically active elements.

The progress achieved and the practical applications realized until now have definitely disproved the notion that IR or Raman spectroscopy are "old-fashioned technologies" useful only for pure systems and relatively small molecules. It has been convincingly proven that

Biomedical Vibrational Spectroscopy, Edited by Peter Lasch and Janina Kneipp

IR and Raman spectra of cells, tissues, or biofluids encode sufficient spectral information to distinguish between different cell types, tissue structures, and biofluids and even to detect changes in these biological materials induced by pathological processes.

1.1 VIBRATIONAL SPECTRA IN BIOMEDICINE PROVIDE FINGERPRINT-LIKE SIGNATURES OF BIOLOGICAL STRUCTURES

A rationale behind the belief that vibrational spectroscopy may be useful to diagnose diseases or pathologies in individuals is that disease processes must, generally speaking, be accompanied by changes in the chemistry/biochemistry of cells, tissues, organs, or body fluids, and vibrational spectroscopy is indeed ideally suited for sensitive detection of such changes as a diagnostic technique. It has furthermore been anticipated that these changes should be detectable also before morphological and systemic manifestation allow clinical diagnosis by conventional methods. Given the fact that sample preparation and measurement are very simple and collection times are in the range of seconds or minutes, IR and Raman spectroscopy should be ideal modalities to establish very rapid nonsubjective and cost-effective tools for early diagnosis of disease processes in individuals.

Biomedical IR and Raman spectroscopy probe biological samples in a way that the active vibrational modes of all constituents present in the mixture are observed in a single experiment, resulting in very complex spectra with broad and superimposed spectral features throughout the whole spectral range. Thus, in contrast to fluorescence spectra, obtained from a biological material labeled with some fluorescing dye, common IR and Raman spectra of intact cells, tissues, or body fluids cannot provide information on a single or even a few specific compounds present. Instead, the spectra provide spectroscopic fingerprints of the total chemical and biochemical composition of the material under study. This situation inevitably results from the fact that the complex superposition of the characteristic IR absorption or Raman signals of all constituents in biomaterials (nucleic acids, proteins, carbohydrates, lipids, and other low molecular compounds, etc.) are observed simultaneously, thereby producing spectral features that encode a vast amount of information potentially useful for biodiagnostic purposes. One peculiarity of vibrational spectroscopy is that it provides information not only on the composition of complex biological material but also on structural states of the molecules under study, since certain bands are sensitive, for example, to the secondary structure in proteins, while others report on the state of order of the membranes or the conformation of the nucleic acid structures. In this sense the total information content in vibrational spectra of biological materials is enormous. One can possibly say that there are presently no other techniques available that can provide such a huge amount of information in one single experiment. On the other hand, this fact severely limits assignments of experimentally observed bands to single discrete structures and qualifies the techniques mainly as fingerprinting methods, though the assignment of spectral bands has improved significantly in the last two centuries due to, for example, spectral resolution enhancement and "spectral feature extraction" capabilities that allow us to more efficiently visualize and resolve specific, hidden bands from the complex spectral signatures.

The nature of information obtained in biomedical vibrational spectroscopy is represented best by the notion of "spectral fingerprints." Thus, the analysis of these spectral signatures by evaluating peak intensities, frequencies, or half-widths of a few bands that can by some means be resolved will fail in most cases. Moreover, taking into account that thousands of spectra have to be analyzed at a given time, the availability of intelligent data

evaluation concepts is a virtual necessity that should ideally include efficient data pretreatment algorithms such as quality testing, normalization, filtering, and adequate multivariate statistical techniques to achieve data reduction and finally the classification of patterns. With such methods, even hundreds of thousands of spectra – as is the case in spectroscopic imaging – can be analyzed.

Vibrational spectra of cells, tissues, and biofluids are obviously the expression of the sum of cellular chemistry/biochemistry and structure. Therefore they provide an "OMICS"-like view of the total chemical/biochemical status of the samples and give a snapshot on cell division, differentiation, growth and metabolism. In this view, vibrational spectroscopic techniques provide information on phenotypes and mirror transcriptional and translational up- and down-regulation processes and post-translational modifications. In a strict sense, vibrational spectroscopies as applied to biofluids, cells, or tissues are, however, not typical metabolomic techniques. Their advantage is possibly that in some way the totality of all chemical/biochemical changes including those in the pool of nucleic acids, proteins, or low molecular metabolic compounds are reflected in the spectra, constituting a technique that cannot easily be assigned to one of the known "OMICS" disciplines in life science such as genomics, transcriptomics, proteomics, or metabolomics. But, as do the common "OMICS" methods, they deal with complex systems in their entirety and with the simultaneous analysis of many biological individuals or objects rather than a single property of a single gene or metabolic product. In many cases the situation might be similar to global metabolic fingerprinting, but one has to bear in mind that the basis of changes observed does not necessarily have to be purely metabolic. This definition qualifies vibrational spectroscopies as explorative and rapid analysis techniques par excellence, which can be used to diagnose disease or dysfunction via spectral biomarkers that change as indicators of the presence of a particular disease or in response to drug intervention, environmental stress, or genetic modification. When nothing or little is known about an observed phenomenon, vibrational spectroscopy may provide a first hint for further, possibly more specific investigations. This is particularly the case when changing systems, whether it is a cell suspension of synchronized cells or cells treated with some specific drug are measured time-dependently. Such experiments can, however, be done with vibrational spectroscopic techniques in a few minutes compared to serial measurements using, for example, fluorescence labels, testing many genes or separating and analyzing proteins or metabolites from complex mixtures. Therefore, the fundamental fingerprinting nature of vibrational spectra of complex biological samples is a big advantage. It is, however, a disadvantage at the same time, since comprehensive understanding of these spectra is desirable but not achievable in most cases.

1.2 DIFFERENT TECHNICAL OPTIONS TO OBTAIN THE SPECTRAL INFORMATION

The most important step forward in biomedical vibrational spectroscopy within the last two decades is certainly the coupling of spectrometers to light microscopes to obtain spectral information from single cells or to achieve spatial resolution in tissue analysis in a way that is familiar to biologists or pathologists. Since then the technological progress has been enormous and high-quality IR and Raman microscopes are available on the market, which can be used to image tissues and single cells and even analyze subcellular compartments. Raman imaging systems that do not rely on spectral point-by-point mapping are not yet on the market, thus precluding Raman imaging under clinical constrains. Notwithstanding,

tissue or subcellular imaging by various different Raman microspectroscopic modalities provides a wealth of biological information not available by other techniques. Today, focal plane array detectors for mid-IR imaging allow rapid segmentation of histological structures without any tissue staining and to image larger cells. Using focal plane array systems, pioneering applications have been published on IR imaging of various soft and hard tissues and a vast number of pathologies. Infrared synchrotron radiation sources coupled with IR microscopes allowed the analysis of single living cells growing in culture with unprecedented high signal-to-noise ratio and reproducibility, opening up the possibility to perform strict difference spectroscopic investigations on viable cells – for example, after treatment with drugs or other chemicals.

Other technical developments such as fiber-optic probes have dramatically increased the possibilities to use Raman spectroscopy as a diagnostic biomedical tool. Fiber-optic applications useful for *in vivo* applications have made greatest progress in Raman spectroscopy, since the production of Raman compatible fiber probes can be based on materials already developed for fiber-based telecommunications or fiber-based chemical sensors. Compared to this situation, optical halide fibers necessary for mid-IR spectroscopy are only available for a few laboratories apart from the detrimental fact that IR radiation has too small penetration depths and problems with strong water absorptions to be useful for *in vivo* experiments.

SERS is a very sensitive Raman modality that can detect and characterize extremely small amounts of nucleic acids, proteins, or virus particles and can also characterize biomolecular events in subcellular compartments. The attractiveness of SERS relies on detection limits close to immunoassay sensitivities with femtomolar detection of, for example, prostate-specific antigen. Tip-enhanced Raman spectroscopy (TERS), another SERS modality, combines SERS spectroscopy with scanning probe technologies and provides lateral resolutions of around 20 nm and thus provides the possibility to study the surface chemistry and structure or composition of cell membranes and cell walls.

Many scientists have realized that IR spectroscopy has great potentials as a fingerprinting technique, useful for the very rapid diagnosis of disease or dysfunction in humans and animals with high-throughput screening capabilities. At present, however, IR and Raman spectroscopy seem to be best developed in microbiology and clinical chemistry, and first dedicated systems for use under practical conditions are already on the market; also, the development of vibrational spectroscopy based diagnostics for *in vivo* glucose screening is near to practical translation. It has also been recognized that vibrational spectroscopies are simple and economical techniques to screen for changes in cells or body fluids in response to drug-based intervention, environmental stress, or genetic modifications in organisms. The results obtained with bone, cartilage, and dental tissues are impressive, and the possibility of practical applications developed for clinical or other medical settings seems to be obvious. The FT-IR imaging data obtained on colon, prostate, or brain cancer are also significant and could be good candidates for translation to routine applications using benchtop IR imaging system as the technical platform.

1.3 THE NEED FOR AND BENEFIT FROM DATA EVALUATION

The necessity to use multivariate pattern recognition methodologies when dealing with spectral data of complex biomedical materials has been realized by the spectroscopic community more than 20 years ago. Among the first who recognized this problem were scientists working with IR spectroscopic data of intact microorganisms. While

univariate statistical analysis considered only a single property of a given selection of microbial species (e.g., a single intensity or frequency value at a given wavenumber or peak), multivariate statistical methods allowed the evaluation of several, if not all, properties of the spectra at the same time. Only in this way the interrelations between the sample properties and the spectra could be figured out. This learning process has been facilitated at that early time by the need to handle thousands of measurements on hundreds of different microbial species and strains, to evaluate these data systematically for spectral similarity, and to exchange data between different laboratories.

Out of the large number of pattern recognition techniques that are presently used for, or have been adapted to, vibrational spectroscopic data, *factor analysis techniques* like principal component analysis (PCA) and hierarchical clustering analysis (HCA) or *classification methodologies* such as artificial neural nets (ANN), support vector machines (SVM), and linear discriminant analysis (LDA) have experienced broad acceptance. Factor analysis is frequently used to achieve data reduction and the classification of patterns in large data sets, and hierarchical clustering (a so-called unsupervised or data-driven classification method) also attempts to find intrinsic similarity structures within the data sets without the need for any *a priori* class assignment, while ANN analysis as a supervised or concept--driven classification technique needs the class assignment of each individual object from the beginning. Partitioning of the whole data set into a so-called training and internal validation data subset is needed to train the system for optimal performance. It took some years by the spectroscopic community to learn that only independent data sets from ideally blinded samples should be used to objectively test the performance and robustness of the classifier and to evaluate the accuracy of the established models.

Meanwhile, nearly the whole arsenal of multivariate bioinformatic techniques is used, and multivariate statistical analysis of spectroscopic data constitutes an own discipline within the scientific area of biomedical spectroscopy. As for any other scientific discipline, these methods not only can be used to evaluate given data sets, but also allow completely new problem solutions to be addressed. New applications arose, for example, when it was realized that determining the covariance between different large data matrices obtained from the same sample populations with fundamentally different techniques is not only a challenge *per se*, but also provides insight into the interlink between biological structures. One of these new applications recently published was the use of genetic algorithms in combination with partial least-square regression (PLSR) analysis to correlate genes selected from gene expression profiles obtained by microarray technologies to metabolic markers from spectral data sets measured from the same samples by IR spectroscopy. The analysis of covariance patterns in these very complex mixed data sets helped to rapidly recognize and visualize the interrelationships and trends in a developing and changing biological system that is not easily achieved by any other means.

1.4 PERSPECTIVES OF BIOMEDICAL VIBRATIONAL SPECTROSCOPY

Despite all the fascinating potential and technological developments and the vast amount of exciting research papers in the literature, progress toward factual translation of vibrational spectroscopic techniques to practical applications is less evident. Moreover, the present situation of a multiplicity of different vibrational spectroscopic modalities, which are viewed by the nonspecialists as competing technologies, is possibly confusing.

The use of IR and Raman spectroscopy for microbial characterization and identification is presently the best developed and most frequent application of biomedical vibrational

spectroscopy. It is especially remarkable that both spectroscopies are applied in microbiological laboratories not only for research purposes but also for routine analysis, for example, in the food industry for microbiological quality control to guide adequate production measures. This situation has greatly been promoted by dedicated high-throughput IR and Raman instrumentation available now on the market. New avenues of microbiological applications can be expected from the use of IR or Raman microscopes, whether it will be for (a) the microspectroscopic analysis of microcolonies to speed up identification of microorganisms and analyze mixed populations of cells or (b) the identification of single cells directly from environmental samples. The combination of IR focal plane array detectors and microarray printing technologies may contribute to make microbiological IR analysis an extremely rapid, cost-effective and unprecedented high-throughput technology for microbiological analyses. This technology may not only help to scale down the number of cells needed for analysis, to investigate mixed cultures, and to perform population analyses, but also help to detect light-microscopic and spectroscopic features simultaneously, with the prospect of a fully automated IR microscopic system combining detection, enumeration, and identification of microorganisms in one single instrument. One particular aspect of vibrational spectroscopy in microbiology which constitutes its attractiveness is the possibility to achieve subspecies differentiation and the ability to analyze all kind of cells that can be grown in culture. No other technique is currently available that can trace microbiological contaminations in food microbiology or perform epidemiological investigation in clinical settings similarly quickly and easily. It is interesting to note that this potential is currently evaluated in several laboratories and that dedicated instrumentations are being designed for microbial subspecies differentiations in collaboration with industrial partners.

It is the author's personal belief that best perspectives for practical applications will arise in those fields where the various vibrational spectroscopic modalities are used as "coupled" techniques – for example, in the form of spectroscopy and microscopy, microspectroscopy and nanoparticles, spectroscopy and optical fibers, or spectroscopy and optical tweezers. In the case of microbiology, to give an example, this will not only allow us to scale down the number of cells needed for analysis to a few or even only a single cell to perform, for example, population analyses in complex habitats, but also allow to detect light microscopic and spectroscopic features of cells simultaneously, which is impossible for other techniques presently in use.

Immense future applications in cell biology, virology, and microbiology may arise from the use of Raman spectroscopy with optical tweezers. Raman tweezers is a relatively new technology that couples Raman spectroscopy with optical tweezers that are already routinely used for the noninvasive manipulation of biological particles to achieve previously impossible sample control. This combination represents a new category of application and may become a modality for flow cytometry to identify cells on the basis of intrinsic molecular properties instead of the particles' size, shape, or fluorescence.

Noninvasive methods to image single live cells are resonance Raman scattering (RRS) and coherent anti-Stokes Raman scattering (CARS) microscopy, which provide intrinsic molecular-vibration-based contrast with a sensitivity that is orders of magnitude higher than conventional Raman microscopy. CARS technology has recently been used to track lipid metabolism in live cells and may become a significant tool in environmental and medical microbiology.

SERS will most probably gain greatest attention reaching far beyond the relatively small community of vibrational spectroscopists, since it may provide biological information that is not available by any other means. SERS used with biocompatible gold nanoparticles incorporated as sensors by cells holds great promise to sensitively and specifically test

molecules in selected subcellular compartments in femtoliter-scaled volumes. This Raman spectroscopic modality could greatly benefit from the fact that defined SERS-active nanoparticles are routinely available and already used along with fluorescence techniques or electron microscopy in cell biology.

The development of technologies for subwavelength spectroscopy of cells and tissues is presently a major point of interest, and different approaches are being evaluated by several groups. The coupling of atomic force microscopy (AFM) with SERS, the so-called tip-enhanced Raman spectroscopy (TERS), seems to be very promising. The possibility to obtain compositional and structural information at a nanoscale level is the most attractive aspect of this new methodology and could provoke as much attention as AFM did some 20 years ago. Also, the coupling of IR lasers with AFM technology, which can probe in a photothermal deflection near-field experiment the local transient deformation induced by an IR pulsed laser tuned to different absorbing wavelengths, may be developed into a microscopic technique that yields chemical contrast at lateral resolutions not accessible by any IR far-field optical technique.

The use of Raman fiber-optic probes may open new avenues for routine *in vivo* use in clinical settings, since the high specificity of Raman spectroscopy can be combined with the possibility of immediate visualization. For practical applications, such fibers will most reasonably be used in multimodal fashion with other optical techniques such as light scattering, optical coherence tomography, or fluorescence spectroscopy, since wide-field Raman imaging still needs to be developed. Further technological progress will be necessary, because fiber-optic technologies are not routinely compatible with existing endoscopic technologies and because of fundamental physical limitations. Though no technical advances are in sight that could allow retrieval of spectra from several centimeters below the tissue surface, very efficient *in vivo* skin analyses based on confocal Raman spectroscopy are already on the market and in practical use.

Bench-top instrumentation for routine IR imaging of diseased tissue sections is available. The vast amount of applications so far published clearly prove that segmentation of histological structures is possible without any staining, and the identification of cancerous lesions within tissues may be achieved in an objective way using extensive reference data bases. Possibly, the *x*th publication of data showing that vibrational spectroscopic imaging can identify pathologies in tissues is not only lacking novelty hereafter, but even counterproductive. To push biomedical vibrational spectroscopy forward, multicenter clinical trials focusing on selected clinical indications are needed to attract the attention of the clinicians and to establish sensitivity and specificity parameters under practical constraints. At present, however, the following questions remain: Who could conduct such trials? Which relevant cancer types or clinical samples (fresh patient biopsies or archive material) should be used? Which technological platforms should be used?

The use of vibrational spectroscopy together with accepted genomic or metabolomic methods such as DNA/RNA microarrays or mass spectroscopies can be of profit when data sets obtained by fundamentally different experimental techniques from the same selection of samples are combined to analyze the covariance patterns in these complex data blocks. The combined analysis, for example, of gene expression and biomolecular response data to external stress factors in microorganisms would help to close the gap between different disciplines, since they can inherently only be done in cooperation between groups that are able to professionally deal with complex technologies. The results of such joint efforts would immediately be recognized by a much broader range of scientists and potential users of the new vibrational spectroscopic techniques.

Obviously, no killer application has been found yet that could pave the way for further steps forward and that cannot be done with any other type of technology. Although vibrational spectroscopy may be superior to competing methods in some cases, no major application could be found to date that can be done in no other way or which is so much superior to replace present technologies in practical use. The scientific community in biomedical vibrational spectroscopy is perhaps at a turning point where practical applications must arise. It will probably not be easy to invest such a high amount of enthusiasm, money, and time for another 10 or 15 years. Indeed, it will instead become more difficult to attract funding for this scientific field, unless significant progress will be made in the transfer of basic science to important practical applications accepted by biologists or clinicians. The gap between enthusiasm and optimism on the one side and the necessity to significantly contribute to the present practical needs of the medical or biological community on the other side must be closed. It should also be clear that series of nice publications will not be enough to close this gap. What must be paramount are joint efforts that combine experience, manpower, and budgets of several groups to bring selected applications to practical applications and patents to the industries.

A similarly important point is the necessity to define standards to exchange data and compare reproducibility levels between the groups and to establish criteria, for example, how sensitivity and specificity values are determined for objective evaluation of spectral data. Without the definition of standards, protocols, and quality control measures, the value of large amounts of data will be rapidly lost after completion of the primary research and increase the probability of reinventing the wheel. This will be critical for the successful development and maturation of an emerging technology like vibrational spectroscopy.