REVIEWS OF MODERN PHYSICS VOLUME 31, NUMBER 1 JANUARY, 1959

13
Application of Infrared Spectroscopy to Biological Problems

G. B. B. M. SUTHERLAND

National Physical Laboratory, Teddington, Middlesex, England

INTRODUCTION

PROGRESS in biology at the molecular level is becoming increasingly dependent on physical and physicochemical techniques such as x-ray analysis, radioactive tracers, chromatography, light scattering, and spectroscopy. The purpose of this paper is to give a broad review of the actual and potential contributions of infrared spectroscopy to biology. Since several excellent articles and papers have already been written giving detailed accounts of the practical application of infrared methods to selected problems in biology, detail is avoided as much as possible in this presentation. In the writer's opinion, infrared methods have not been sufficiently exploited in the biological field and, since this may well be because many biologists find it hard to assess their value, it is hoped that a rather general discussion of the advantages and disadvantages of infrared analysis will be of value to biologists.

The infrared region of the spectrum is conventionally defined as extending from the limit of the visible (about 7500 A or 0.75 μ) to the present lower wavelength limit of the microwave region (about 1000 μ), but only a relatively small range of these wavelengths is of much interest to biologists, *viz.*, 2.5 to 15 μ. There are two reasons for this: the first is that within these limits are found the vast majority of spectra which throw any light on molecular structure; the second is that this region can be conveniently scanned and recorded in about ten minutes by a commercially available instrument with a rocksalt prism as the dispersing element. The exclusion of infrared wavelengths beyond 15 μ does not seem so strange when expressed in frequency* rather than wavelength units (i.e., 670 to 10 cm^{-1}), since the majority of the fundamental vibration frequencies of molecules which can be readily assigned to highly localized motions of one or both of the atoms in a chemical bond lies between 4000 cm^{-1} (2.5 μ) and 600 cm^{-1}. The limits of 670 and 4000 cm^{-1} given in the foregoing are set by the fact that beyond 670 cm^{-1} absorption by the rocksalt prism rapidly becomes complete, while above 4000 cm^{-1} the dispersion is rather low. However, it is becoming easy to replace prisms by gratings, and so these limits need no longer be set by the experimental technique. The first reason given is the dominant one, *viz.*, most of the interesting absorption bands which can be interpreted with certainty lie between 4000 and 600 cm^{-1}.

* The conventional frequency unit in infrared spectroscopy is the wave number (cm^{-1}), which expresses the number of waves in 1 cm.

The general interpretation of the infrared spectra of molecules is very straightforward; even the detailed interpretation is not difficult, provided the objective is, for instance, the identification of a compound in a mixture, or the establishment of the presence of a characteristic chemical grouping in a molecule. But, to determine the configuration of a biological polymer such as a protein is a difficult and complex problem which requires the same degree of expertise as that needed in the x-ray analysis of protein structure.

GENERAL INTERPRETATION OF INFRARED SPECTRA OF MOLECULES

The average polyatomic molecule of biological interest has an absorption spectrum in the region between 600 and 4000 cm^{-1} consisting of 20 to 30 discrete absorption bands of which about 5 to 10 will usually be much more intense than the remainder. This is because certain of the modes of vibration of such a molecule cause a periodic change in the dipole moment of the molecule. Thus, radiation of the appropriate frequency incident on the molecule can be absorbed and excite the corresponding mode of vibration. Some bands are more intense than others because the change in dipole moment is greater for some of the modes of vibration. Theoretically, a polyatomic molecule (of n atoms) has an infinite number of modes of vibration, all of which can be built up from the $3n$-6 fundamental modes of vibration. However, the change in dipole moment associated with overtone and combination frequencies is usually an order of magnitude less than that associated with the fundamental modes, and in the region of the spectrum under discussion the majority of the observed bands can be assigned to fundamental modes.

How can these fundamental modes be vizualized? Although, strictly speaking, each fundamental mode involves some motion of every atom in the molecule and would, therefore, seem to be hard to predict and describe, it turns out that an appreciable fraction of the fundamental modes of any molecule can be assigned to easily vizualized vibrations of individual chemical bonds, or of small structural units in the molecule. This is especially true of the fundamental modes which are associated with the motions of hydrogen atoms. For instance, any molecule containing an NH group will be found to have a fundamental vibration near 3300 cm^{-1} owing to the stretching and contraction of the NH bond and two other fundamentals between 1600 and 600 cm^{-1} in which the motion of the hydrogen atoms is almost perpendicular to the NH bond. More generally, in the

region between 3700 and 2500 cm^{-1} are found fundamental modes owing to such "hydrogenic-stretching frequencies" in the following order:

OH 3650–3150 cm^{-1},
NH 3500–3150 cm^{-1},
CH 3150–2850 cm^{-1},
SH 2650–2550 cm^{-1},
PH 2450–2300 cm^{-1}.

No other fundamental modes appear in this region of the spectrum, and so identification of any of these bonds in a compound is generally easy. There are occasional limitations, e.g., difficulty of distinguishing OH from NH, and the weakness of the SH fundamental; these are not discussed here.

The corresponding "hydrogenic deformation frequencies" for OH, NH, and CH occur between 1650 and 600 cm^{-1}, but cannot be characterized so simply. Moreover, many other fundamental modes occur within this range whose assignment and identification present special problems. However, several groups (such as CH_2 and CH_3 and the peptide link) can readily be identified by characteristic fundamentals of this type.

Before leaving the hydrogenic frequencies, mention should be made of the ease with which hydrogen bonding can be detected through its effect on these frequencies. The stretching frequencies are lowered by a few hundred cm^{-1}, and the corresponding absorption bands become quite broad and diffuse. The deformation frequencies are increased to a much smaller degree and the broadening is not always so marked.

In the region between the hydrogenic-stretching and hydrogenic-deformation frequencies (*viz.*, 2300 to 1650 cm^{-1}) occur two other separable classes of fundamentals: firstly, those owing to the stretching of triple bonds such as C≡N and C≡C occurring between 2250 and 2100 cm^{-1}, and, secondly, those largely localized in the stretching vibrations of double bonds such as C=O, C=C, C=N. The latter start about 1800 cm^{-1} and overlap a little the high-frequency end of the hydrogenic-deformation frequencies at 1650 cm^{-1}. This means that it is occasionally difficult to decide whether a band occurring between 1650 and 1500 cm^{-1} is the result of the stretching of a double bond or of the deformation of a hydrogen atom. If the hydrogen atom can be substituted by deuterium, this uncertainty is often resolved, as a hydrogenic frequency will be reduced by almost a factor of $\sqrt{2}$, whereas the double-bond frequency will be unaffected. It should be added that occasionally a hydrogenic-deformation frequency interacts with a double-bond frequency, and the two fundamentals resulting can only be described as a superposition of these two motions.

A very large number of the chemical bonds in any polyatomic molecule are of the type C–C, C–O, C–N. Such bonds do not give rise to localized fundamental vibrations. The reason is that the masses of these atoms are not very different, nor are the restoring forces between them. This means that their characteristic stretching frequencies (which can be observed in small molecules such as ethane or methyl alcohol, where they occur in isolation) are all about the same magnitude (approximately 900 to 1100 cm^{-1}). Consequently, in a polyatomic molecule, strong coupling occurs and the resulting fundamental modes involve simultaneous motions of all of the bonds of this type in the molecule. Such fundamentals are usually referred to as "skeletal modes" and have a range between about 800 and 1250 cm^{-1}. The pattern of these skeletal frequencies is often the most characteristic physical property of a molecule. Two hydrocarbons or two steroids, which may be very hard to differentiate by chemical means or by other physical properties (e.g., refractive index or melting point), can usually be recognized instantly by comparing their infrared spectra between 700 and 1300 cm^{-1}.

APPLICATION TO BIOLOGICAL PROBLEMS

There are three principal ways in which infrared spectroscopy can be of help in molecular biology, *viz.*:

(a) as an analytical tool, (b) as a means of establishing the structural formula of a biologically important compound, and (c) as a means of determining the spatial configuration of biological polymers.

(a) Analysis

Infrared spectroscopy is now a well-established and widely used analytical tool in any up-to-date biochemical laboratory. The principal advantage of infrared over visible or ultraviolet spectroscopy is that the spectrum is much richer, since it is the result of vibrations in every part of the molecule. Spectra in the visible and ultraviolet arise from the excitation of an electron in one part of the molecule, usually in a double bond such as CO group. In analytical work, one therefore has to be certain that this group is not present in any of the other compounds in the mixture. On the other hand, infrared methods will usually reveal every one of the compounds in a mixture. The only serious disadvantage of infrared analysis is that the use of water as a solvent is generally impossible because of its intense absorption over a good deal of the working range. This can frequently be overcome either by the use of other solvents, or (in the case of a solid) by making a pressed disk of the compound ground very finely in an excess of KBr. One other point should be mentioned, since it is frequently very important in much biological work, *viz.*, the minimum quantity which is necessary in order to make an identification. It is impossible to give a precise figure here, because the intrinsic intensity of absorption in the infrared varies over such a wide range between compounds, and reference must be made to the literature on this subject. As a general guide, one may say that, by using specially designed cells, it is possible to get a spectrum from about a milligram of material, but that, by using a reflecting

microscope attachment, this limit can be reduced to below one microgram.

(b) Structural Formulas

One of the earliest applications of infrared spectroscopy to the structure of a biologically important molecule occurred in the case of penicillin. Here, the chemists were unable to agree on which of the three possible formulas was the correct one. Although x-ray analysis gave the first and most unequivocal proof that the β-lactam formula was the correct one, infrared methods gave an independent proof and demonstrated that no structural changes took place in going from the solid state (required for x-ray work) into solution, in which penicillin is known to be a rather labile molecule. This work depended on differences in the type and in the environment of the double bonds in the three possible structures. By investigating the spectra of model compounds containing double bonds of various types in appropriate environments, a decision could finally be made in favor of the lactam structure.

The biological molecules which have been most extensively investigated by this method are the steroids. In such compounds, it is now possible to answer quite a variety of important structural questions concerning the position of CO, OH, CH_2, and C=C groups. On the other hand, it is still not possible to recognize the class of steroids by any common feature running through all of their spectra, although certain closely related groups of steroids do show common features. We find the opposite extreme in the proteins which, as a class, are easily recognized and differentiated from nucleic acids and lipids, but show remarkably little variation among one another. The reason is that the spectrum of any protein is dominated by several intense bands arising from the identical peptide links. Only in the few cases where two or three of the constituent amino acids are in great excess is it possible to identify individual amino-acid residues (e.g., glycine and alanine in silk).

The general line of attack is to compare the spectrum of the compound whose structure is unknown with the spectra of compounds of known structure containing groups of atoms identical with and in a similar environment to those suspected of occurring in the unknown. Extensive collections of infrared spectra have now been compiled by various laboratories and organizations, and the comparisons can be made relatively rapidly by various mechanized sorting devices. At first, it is usually advisable to call in an experienced spectroscopist for final confirmation, but any laboratory which takes up infrared analysis in a particular chemical field is not long in developing its own expert. The logic behind the whole process is not unlike that behind the chemical attack on the same problem. The chemist identifies certain groups in an unknown structure by their reactive properties, and he degrades the molecules into simpler ones which he can identify from their well-known chemical or physical properties. The spectroscopist identifies individual chemical bonds in an unknown structure from their known spectroscopic earmarks and tries to supplement this by identifying groups or constituent units in a structure (e.g., peptide link, benzene ring) in a similar way. The most efficient approach is, of course, to combine the two methods.

(c) Determination of Spatial Configuration of Biological Polymers

In order to understand many key biological phenomena, such as the mode of action of genes, protein synthesis, or the production of antibodies, at the molecular level, it will be necessary to establish in detail the molecular configuration of proteins and nucleic acids. This must be done by physical methods, of which the most successful so far has been x-ray analysis. However, in spite of the intensive efforts of several groups of very able workers in various countries over the past twenty years, the spatial configuration of any globular protein is still unknown. While it is true that one now seems to be very close to this goal in the case of myoglobin, the extension of the methods used for that protein to other proteins presents formidable problems. At present, infrared analysis seems to be the second most powerful physical method. Although it has not yielded such precise results, it has often given independent confirmation of the main features of a structure. More important is the fact that it can often give a lead at any early stage, either in laying down certain conditions, which the correct model structure must fulfill, or in ruling out certain models which might appear to fit a preliminary analysis of the x-ray data. Just as the x-ray method requires the polymers to be arranged in an orderly pattern (preferably in a single crystal), so the infrared method can be more precise the closer the arrangement of the polymers approaches that found in a single crystal. This is because the infrared method is based on observing the dichroism associated with key absorption bands when the spectrum is obtained using polarized radiation.

An oversimplified example will make this clearer. In the infrared spectrum of deoxyribonucleic acid is an absorption band which can be assigned to vibrations localized in the planes of the purine and pyrimidine bases. In the Watson-Crick model of DNA, these bases are nearly perpendicular to the axis of the double helix. If a highly oriented specimen of DNA is prepared in which the axes of the helices are roughly parallel to the direction of orientation, then this absorption band is very intense when the incident infrared radiation is polarized with the electric vector perpendicular to the direction of orientation, and very weak when the direction of polarization is rotated through a right angle (Fig. 1). This gives a general confirmation of one aspect of the Watson-Crick model. If it were possible to produce a single crystal of DNA of suitable dimensions in which all of the bases were parallel to each other, then the dichroism would become perfect (i.e., no absorption for

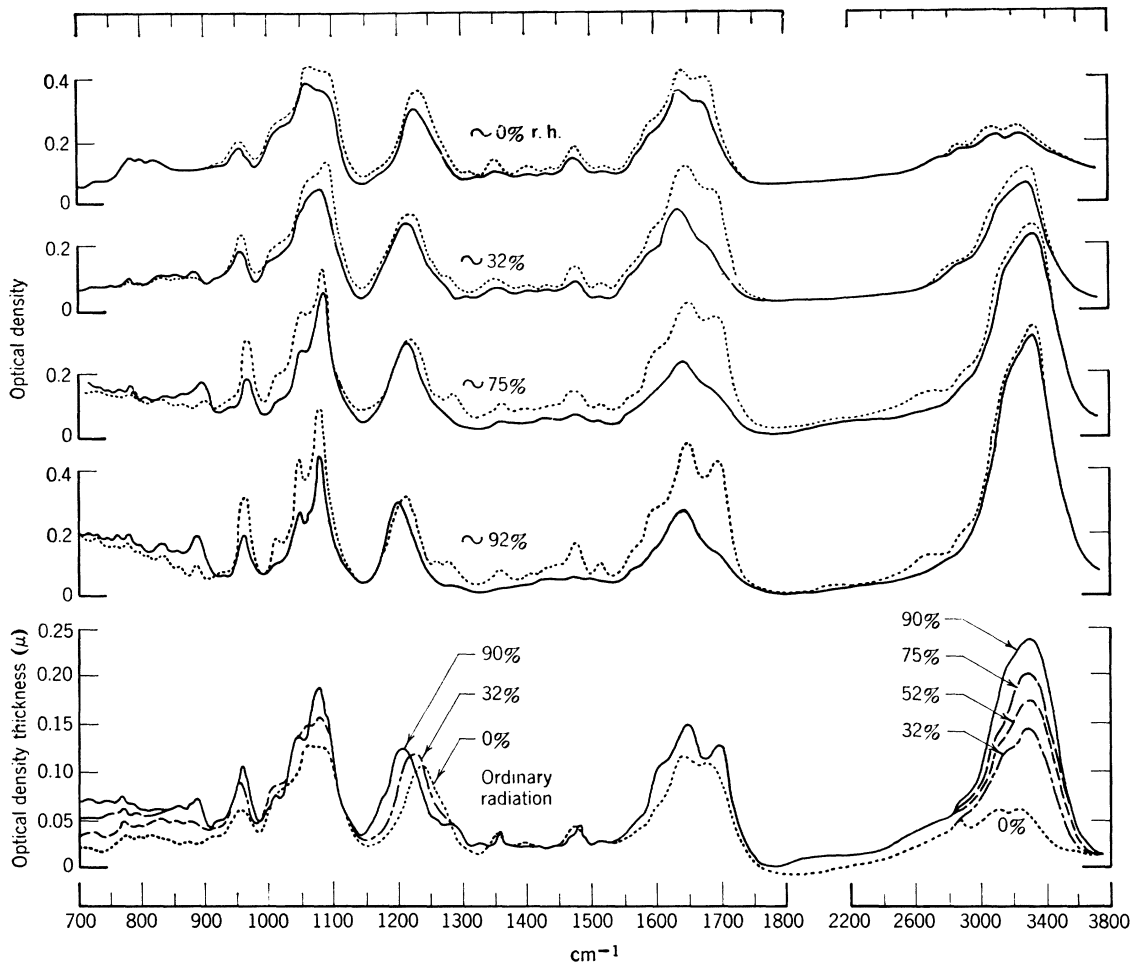

FIG. 1. The infrared absorption spectrum of NaDNA at various relative humidities. ——— indicates electric vector parallel to the orientation direction; ..... indicates perpendicular to the same direction. The bands referred to in the text are those between 1600 and 1700 cm^{-1} [from G. B. B. M. Sutherland and M. Tsuboi, Proc. Roy. Soc. (London) **A239**, 446 (1957)].

one polarization direction) and a precise confirmation of this aspect of the model could be given.

From the foregoing, it might appear that the infrared methods should, therefore, be applied to single crystals of biological polymers. Unfortunately, there are severe practical difficulties about such an approach. The first is that the "suitable dimensions" referred to in the foregoing hardly ever can be realized, *viz.*, a flat crystal about 10 μ thick and 2×5 mm in area. A second difficulty is that many biological polymers crystallize only in association with a high proportion of water. This limits observation to the bands which are not obscured by the spectrum of H_2O or D_2O. The third is that the molecular configuration is unknown and assumptions have, therefore, to be made about the orientations of recurring chemical groups or bonds in the polymer. The number of reasonable assumptions from which a choice has to be made can be quite large, and infrared analysis may not prove to be a sensitive enough discriminant between the various possibilities. In cases where the number of ways of fitting a long polymer into the unit cell is severely restricted, this method should give a precise answer, and it may be that the first experimental difficulty referred to can be overcome by investigating with a microspectrometer the reflection spectrum, instead of the absorption spectrum. Another line of attack is to measure the dichroism of the overtone and combination bands above 4000 cm^{-1}, since these are much weaker than the fundamentals and can be conveniently studied in a reasonably thick crystal. It should be added that there are still some theoretical problems to be overcome in the interpretation of such spectra.

Because of all of these difficulties, most of the infrared investigations to date have been made on thin sections of fibrous proteins such as porcupine quill, silk, etc. In such cases, the orientation is far from perfect and obviously the degree of disorientation must be estimated quantitatively before any reliable deductions can be made from the partial dichroism. A method of doing this has been recently worked out and applied to a number of proteins, *viz.*, porcupine quill, elephant hair, horse hair, horn, feather quill, silkworm gut, and colla-

gen. For each of these proteins, seven different structures were tested, and it was possible to reject all but one or two of the model structures in most cases.

It must be emphasized that this method of attack depends on certain assumptions which may not always be correct. These assumptions are as follows:

(1) The direction of change of electric moment (associated with a particular absorption band) bears the same relation to the repeat unit of the polymer as that determined from infrared studies on single crystals of small molecules which contain the repeat unit; (2) the spatial distribution of the polymer chains has a certain degree of symmetry; (3) the specimen studied consists of one structural species.

The first assumption is necessary, because it is found from work on single crystals of small molecules containing a peptide link that the change of dipole moment in a bond stretching fundamental vibration (e.g., NH or CO) is not exactly along the bond in question, but may make an angle of as much as 20° with it. Thus, the directions of these bonds (and consequently the molecular configuration) cannot be determined unless the corresponding angles are known for a peptide link in a protein. Up to the present, single crystals of only two different molecules have been investigated. The angles in these two crystals did not differ by more than five degrees. However, more work on single crystals of a few other molecules is required in order to estimate accurately the degree of uncertainty arising from this assumption. The second assumption can be shown to be a reasonable one if the sample is prepared in a certain way. At first sight, the third assumption seems reasonable and is made by the x-ray analysts. The reason for doubting it is that the structure of many of the infrared bands of proteins is complex. Whether this arises from the presence of two different configurations, or from differences in frequency between a paracrystalline and less-ordered arrangements of the polymer molecules, cannot easily be determined. In simple synthetic polymers, such as polyethylene, differences in frequency are found between crystalline and amorphous forms of the polymer. Structure can also occur in the absorption band of a single crystal because of interaction between neighboring molecules.

It appeared at one time that infrared analysis might be able to distinguish between various configurations of a protein molecule (specifically between the α-helix and the fully extended form of the polypeptide chain), even when the polypeptide chains were not arranged in any ordered pattern. This method depended on a change in the frequency of the CO absorption band in synthetic polypeptides which could be prepared in either of these two forms by precipitation from the appropriate solvent. However, there are so many possible reasons for a small change in the frequency of an absorption band that the application of any such rule to proteins (even on an empirical basis) leads to serious difficulties and inconsistencies. Until the reasons for these changes are well understood from the study of simpler model compounds, it seems very unwise to use them as a guide to protein structure.

The general situation with the infrared analysis of protein structure is similar to that which existed at an earlier stage in the x-ray analysis of the same problem. Before the interatomic distances in the peptide link were established by careful work on small molecules, it was impossible to establish the correctness of any protein structure by x-ray methods. One could only say that such and such a structure was generally consistent with the diffraction pattern. Now that these distances are known and that single crystals of proteins can be prepared containing a heavy atom, it is becoming feasible to locate the individual atoms. Similarly, before the infrared method can become precise, more work is required on single crystals of simple compounds containing the peptide link and on the experimental problems of obtaining satisfactory spectra from single crystals of proteins.

BIBLIOGRAPHY

The following bibliography is not meant to be comprehensive. It gives references to most of the general articles on the subject within which references can be found to the original papers. A few references have been given to very recent papers which are not dealt with in the general articles.

M. Beer, G. B. B. M. Sutherland, K. N. Tanner, and D. L. Wood, "Infrared spectra and structure of proteins," Proc. Roy. Soc. (London) **A249**, 147 (1959).

L. J. Bellamy, *The Infrared Spectra of Complex Molecules* (Methuen and Company, Ltd., London, 1954).

E. R. Blout, M. Parrish, Jr., G. R. Bird, and M. J. Abbate, "Infrared microspectroscopy," J. Opt. Soc. Am. **42**, 966 (1952).

C. Clark, "Infrared spectrophotometry" in *Physical Techniques in Biological Research*, G. Oster and A. W. Pollister, editors (Academic Press, Inc., New York, 1955), p. 205. (This is the best general review of the subject.)

K. Dobriner, E. R. Katzenellenbogen, and R. N. Jones, *Infrared Absorption Spectra of Steroids* (Interscience Publishers, Inc., New York, 1953).

A. Elliott, "The structure of polypeptides and proteins" in *Proceedings of the Third International Congress of Biochemistry, Brussels, 1955* (Academic Press, Inc., New York, 1956), p. 106.

R. D. B. Fraser, "Infrared spectra of biologically important molecules" in *Progress in Biophysics*, J. A. V. Butler and J. T. Randall, editors (Pergamon Press, London, 1953), Vol. III, p. 47.

Biological Applications of Infrared Spectroscopy, Ann. New York Acad. Sci., **69**, 1–254 (1957).

H. M. Randall, R. G. Fowler, N. Fuson, and J. R. Dangl, *Infrared Determination of Organic Structures* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1949).

G. B. B. M. Sutherland, "Infrared analysis of the structure of proteins and nucleic acids" in *Convegno Antonio Baselli le Macromolecole Viventi*, Ist. Lombardo sci. lettere **89**, 67 (1955).

G. B. B. M. Sutherland and M. Tsuboi, "Infrared spectrum and molecular configuration of sodium deoxyribonucleate," Proc. Roy. Soc. (London) **A239**, 446 (1957).

D. L. Wood, "Infrared spectrophotometry" in *Methods in Enzymology* (Academic Press, Inc., New York, 1957), Vol. IV, p. 104.