NPL
National Physical Laboratory

Report

**NPL REPORT**
**DQL-AS023**

# Techniques for the measurement of the structure of proteins in solution, and their complementarity to circular dichroism

**Jascindra Ravi, Ratna Tantra, Alex Knight**

**NOT RESTRICTED**

DECEMBER 2005

dti
FUNDED BY THE DTI

National Physical Laboratory | Hampton Road | Teddington | Middlesex | United Kingdom | TW11 0LW

**Switchboard** 020 8977 3222 | **NPL Helpline** 020 8943 6880 | **Fax** 020 8943 6458 | **www.npl.co.uk**

# Techniques for the measurement of the structure of proteins in solution, and their complementarity to circular dichroism

Jascindra Ravi, Ratna Tantra, Alex Knight
Quality of Life Division

## ABSTRACT

There is a recognised requirement for techniques complementary to Circular Dichroism (CD) for the structural characterisation of biopharmaceuticals as part of the quality control process. This report reviews the available techniques for protein structural characterisation and evaluates their suitability for this application. It is concluded that Fourier Transform Infrared Spectroscopy (FTIR) is the most suitable such technique.

ISSN 1744-0602

National Physical Laboratory
Hampton Road, Teddington, Middlesex, TW11 0LW



Approved on behalf of the Managing Director, NPL
by Martyn Senè, Director, Quality of Life Division

# Contents



# Figures

# 1 Introduction

## 1.1 Context

This report has been produced as part of the project "Higher-order structure measurement to support ICH guidelines" under the Department of Trade and Industry's second "Measurements for Biotechnology" programme. Further information about this project, and others under the programme, can be obtained at: http://www.mfbprog.org.uk or http://www.npl.co.uk/biotech.

## 1.2 Background

The advent of recombinant DNA technology has resulted in an important broad class of therapeutics, biopharmaceuticals, for the treatment and prevention of life threatening and debilitating conditions [1] . About half of all new drugs approved are now biopharmaceutical compounds and include recombinant protein vaccines, hormones, enzymes, interferons, cytokines, blood clotting factors and monoclonal antibodies.

Therapeutic proteins must maintain potency, therapeutic efficacy, and stability over their shelf life, like all pharmaceuticals [1]. However, a protein drug is very susceptible to chemical and physical modifications, due to the multiple reactive sites on amino acids, and their susceptibility to structural alteration. The occurrence of subtle structural changes has major impact on their activity, stability and toxicity, and consequently compromises the efficacy and shelf life of the formulation. Proteins may adopt the incorrect structure through problems in production, the purification process, or improper storage.

A key aspect of the quality control process is the characterisation and comparison of the three-dimensional structure of batches of biopharmaceuticals. In order to reliably discriminate between 'compliant' and 'non-compliant' batches, validated physicochemical techniques are required. Currently the only techniques recognised for such structural quality control (QC) in the relevant ICH document (Q6B[1]) are circular dichroism (CD) and NMR[2] [2].

There is a general recognition that, while CD is a powerful technique, there is scope for complementary techniques for biopharmaceutical product characterisation. Indeed, at a CBER/IABS meeting in 2003[3], industry and regulators both recognised a need for alternative techniques to characterise the higher order structure of biopharmaceuticals [3].

CD is more sensitive to some structural changes than others; and the degree of sensitivity is product specific, particularly in the near-UV region. For example, the degree of changes in the near UV depends on the position of chromophores within the molecule. CD is also more sensitive to α-helix than to other structures, such as β-sheet [4]. Therefore there is scope for

---

[1] This document may be downloaded from the ICH web site at: http://www.ich.org/

[2] Abbreviations: NMR, Nuclear Magnetic Resonance; CD, Circular Dichroism; QC, Quality Control; UV, Ultraviolet; UV-Vis, Ultraviolet-visible; PCA, principal component analysis; IR, infra-red; CCD, charge-coupled device; FT, Fourier Transform; FTIR, Fourier Transform Infra-Red spectroscopy; SERS, Surface-Enhanced Raman Spectroscopy; VCD, Vibrational Circular Dichroism; ROA, Raman optical activity; ATR, Attenuated Total Reflectance

[3] "State of the Art Analytical Methods for the Characterization of Biological Products and Assessment of Comparability", Bethesda, MD, 2003. Proceedings currently in press.

additional techniques, which provide additional information, and which confirm the results of CD measurements.

In the following sections, each of the existing techniques is reviewed, with particular emphasis on suitability for biopharmaceutical characterisation and complementarity to CD. The techniques described are divided into two categories: “High resolution” techniques, which can obtain, *a priori*, atomic-level structures of proteins, are discussed first; this is followed up with a discussion of the wide variety of “low resolution” techniques that are available. These techniques do not provide such detailed information, and in some cases do not provide *a priori* structure, but may be more suitable for QC applications where comparison of structures between batches is more significant. CD is just such a low-resolution technique, and the complementarity of each technique to CD is highlighted.

# 2 High Resolution Techniques

## 2.1 X-Ray Crystallography

X-ray crystallography was the first and the most powerful technique, which was able to determine the atomic structures of protein molecules. As the name of the technique implies, the protein to be studied must be in a crystalline form with a high degree of order of the crystals, and sufficient crystal size and quality to obtain high-resolution data. Since there is no generic crystallisation protocol, studies generally proceed on a trial and error basis, or by using automated high-throughput methods [5]. The crystallisation of proteins is therefore a major hurdle in the structure determination process, more art than science.

The information about the protein's structure is obtained from the diffraction pattern produced by the X-ray beam in the presence of the crystals [5]. The X-ray source may be either a normal X-ray laboratory source or intense monochromatic radiation produced by a synchrotron. The X-ray diffraction data, which includes the position and intensity of spots for different rotations of the crystal, is now usually collected using electronic detectors.

Structural information is then obtained by reconstructing an electron density map of the molecule, using the information contained in the diffraction pattern, and by fitting the primary structure of the molecule into the envelope of the electron density map. However, the diffraction pattern alone does not contain all the necessary information, in which case the preparation of a heavy atom derivative of the protein prior to crystallisation is required.

X-ray crystallography is able to solve much larger structures than NMR (see below), however success is not guaranteed as the crystallisation of many proteins proves difficult and sometimes it is impossible for certain proteins such as glycoproteins. The steps involved are also lengthy, labour intensive and time consuming, however, crystallisation robots and automated interpretation of electron density maps are now being used with the aim of reducing the time taken [6].

Moreover, crystallisation itself could result in a structure very different from the structure of the protein in solution. X-ray crystallography is therefore not suitable for screening or quality control applications; however, it can be used as a "one-off" technique to solve the structure of a protein. This enables this technique to provide the "gold standard" for structural information.

## 2.2 Nuclear Magnetic Resonance (NMR)

NMR spectroscopy is one of only two techniques that can provide atomic resolution structures of proteins. The major advantage of this technique is that it can be used to study the structure of proteins in solution [7]. However, the concentration of the protein solution must be high (0.5–1.0 mM) to obtain suitable spectra of sufficient quality [8]. The detailed molecular structure at atomic resolution results from a laborious examination of a number of conformationally sensitive parameters and the application of distance geometry programs [9].

Unlike X-ray crystallography, NMR is a non-destructive technique; however, it is limited to proteins of less than 40 kDa molecular weight due to the major difficulty in understanding the link between chemical shifts and structural parameters. Although recent developments in technology, experimental and computational aspects of NMR promise to increase the size of proteins that can be studied, it has not yet become possible [10, 11]. Like X-ray crystallography, it requires expensive equipment, special technical skills and moreover time-consuming data analysis. However, unlike X-ray crystallography, NMR has the potential for

rapid measurements and also provides information about the dynamics of the protein. The information obtained by rapid measurements can be partial but contains very useful information about the structure of the protein.

The most commonly used nucleus in NMR is the isotope of hydrogen, $^{1}H$, which is naturally the most abundant in macromolecules. Working with other isotopes, such as $^{13}C$ and $^{15}N$, requires the protein to be labelled with the isotopes at the time of synthesis. This would make these isotopes unsuitable for screening of batches of pharmaceuticals.

A one-dimensional NMR spectrum is the simplest and least time-consuming type of NMR measurement [12]. Protein aggregation, folding and secondary structure can be elucidated using specific chemical shift values that correspond to protons in different chemical environments. However, full secondary and tertiary structure can be solved for small proteins using a two-dimensional NMR spectrum. For larger proteins, multi dimensional NMR and isotopic labelling are required to determine the complete structure of the protein.

Although the more complex forms of NMR are not suitable for biopharmaceutical applications, one-dimensional NMR is useful for the screening of low molecular weight products. The screening process is normally accomplished by comparing spectra with reference spectra instead of interpretation.

One-dimensional NMR spectroscopy can therefore be used in screening processes of biopharmaceutical products, however, it requires extremely expensive equipment and specialized skills. As NMR is on the verge of a dramatic increase in efficiency in terms of instrumentation and automation of data analysis it will be interesting to see whether the use of NMR increases in the biotechnology field. Compared to CD, NMR can potentially provide more detailed information, but the cost of making measurements, both in terms of equipment and time, means that it is currently limited in application. Like X-ray crystallography, it is at present more suitable for obtaining reference structural information.

# 3 Low Resolution Techniques

## 3.1 Ultraviolet-Visible (UV-Vis) Absorption Spectroscopy

UV-Vis absorption spectroscopy involves the measurement of the absorption of light at a given frequency by a collection of molecules. This absorption of light causes rearrangement of electrons in a molecule, where the electron density is pushed from a starting arrangement (lower energy state) to a higher energy state resulting in electronic transitions ($\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$) within a molecule.

The broad spectral features observed in the UV-Vis region of proteins have limited the use of absorption spectroscopy for protein identification, however it is routinely used for protein concentration and purity measurements (Figure 3.1). The absorbance of proteins in the near UV region (250–300 nm), which is also described as the aromatic region, is predominantly due to electronic transitions (principally $\pi \rightarrow \pi^*$) of the aromatic amino acids tryptophan, tyrosine and phenylalanine [13]. However, disulphide bonds (cystines) in peptides also contribute to the total absorbance in this region. The usefulness of the far UV (< 250 nm) region is limited by the absorption of the amide group, peptide backbone of the protein, below 200 nm, which does not contribute any structural information.

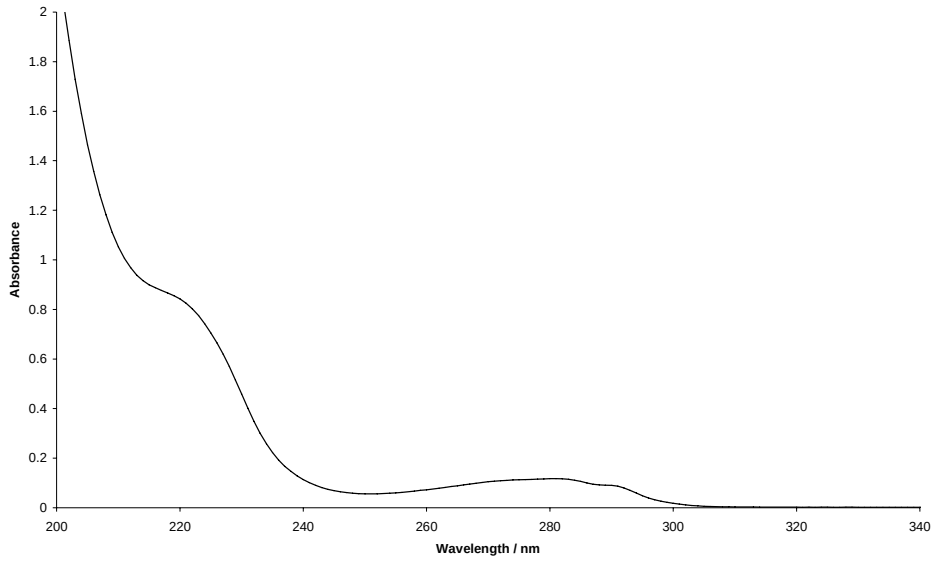



***Figure 3.1 Absorbance spectrum of lysozyme***

***Lysozyme at 0.5 mg/ml in phosphate buffer solution (pH 7, 30 mM), measured in a 1 mm path length cell. The spectrum was collected using a PerkinElmer Lambda 850 series UV-Vis spectrophotometer with scan speed ~ 260 nm/min and data pitch 1 nm. Note the broad peak, due to aromatic side-chains, at around 280 nm.***

The UV-Vis absorption can be used as a sensitive measure of subtle changes in protein structure as the molar absorptivity and peak wavelength are dependent on the polarity and chemical environment of the amino acids [14, and references therein]. The solvent polarity changes, which have greater effect on the absorption spectra of amino acids, can be used to study the solvent-exposed residues in a protein. The strong absorption of tyrosine and tryptophan enables one to determine the protein concentration by measuring the absorbance near 280 nm. One may estimate an absorption coefficient by using the known molar absorptivities of the aromatic amino acid residues at 280 nm, if the primary structure of the protein is known.

Although various chromophores in a protein contribute to the absorbance in the near-UV and are significantly overlapped, the contributions of individual amino acids can be well resolved using derivative spectroscopy and therefore the aromatic amino acid content and the tertiary structure of the protein can be evaluated [14, and references therein]. UV-Vis spectroscopy, in conjunction with other techniques, can therefore provide insights into structure, as well as purity and concentration.

Since CD (see below) is itself a form of UV-Vis spectroscopy, and obtains more detailed information, this technique is not really complementary to CD, although its utility in making accurate protein concentration measurements is undeniable.

### 3.2 Circular Dichroism (CD)

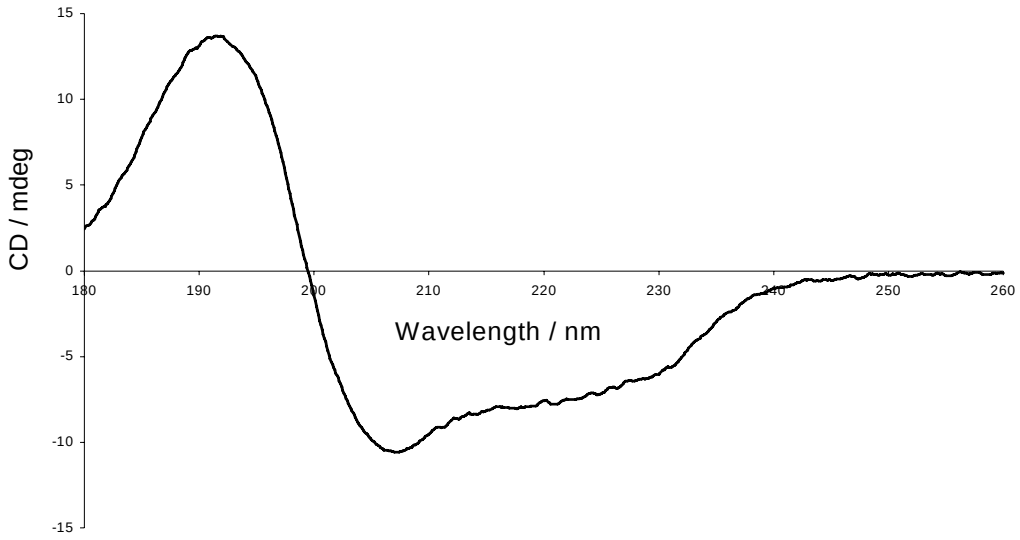
CD is an absorptive phenomenon represented as the differential absorption of left and right circularly polarised light [15]. In order to be CD active the molecule must be chiral and absorb in the spectral region of interest. Proteins or peptides are therefore CD active as most amino acids are chiral (except for glycine) and possess suitable chromophores (such as peptide bonds and aromatic side chains) to absorb in the UV-Vis region [16]. Furthermore, the secondary structural elements of the protein impose additional chirality due to the difference in hydrogen bonding.

The far UV region of the CD spectrum (typically 180-260 nm, Figure 3.2(a)), which is principally due to the absorption of peptide bonds, provides information regarding different forms of regular secondary structure, including α-helix, β-sheet and β-turn, found in proteins. The CD signal observed in the near UV region (typically 240-360 nm, Figure 3.2(b)), is principally due to the aromatic side chains of phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) and provides a detailed fingerprint of the tertiary structure [15, 17]. However, *a priori* interpretation of this region is difficult, although its sensitivity to structural changes makes it very useful for comparative purposes. Disulphide bonds also absorb in this region, and although the dihedral angle of the disulphide bond can be deduced from the CD spectrum it is difficult in the case of extensively disulphide-bonded proteins. In addition the interaction between ligands or cofactors (haem, flavin, and pyridoxal-5'-phosphate) can be deduced using changes in the CD signal in the near-UV region. The visible region is less commonly used, as most proteins do not contain chromophores that absorb in this region. The exceptions are typically prosthetic groups such as haems.

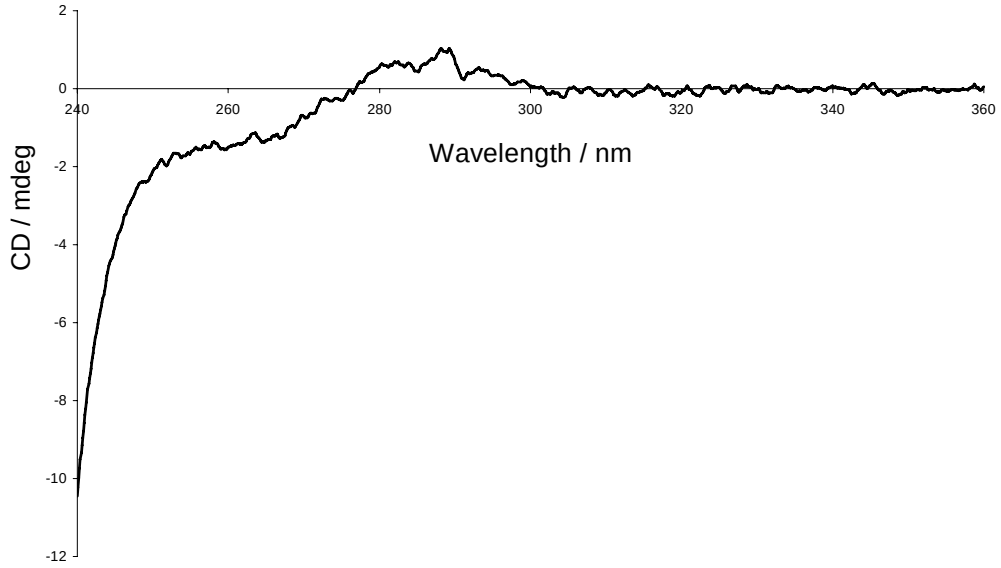
The principal absorbing group in the far UV region, the peptide bond, shows two allowed electronic transitions $\pi \rightarrow \pi^*$ at about 190 nm and $n \rightarrow \pi^*$ transitions. As a protein is a polymer of amino acids, the electronic and magnetic transitions of the chromophores in the repeating units interact and produce multiple transitions, which may cause a high degree of circular dichroism. It is this effect, and exciton coupling of $\pi \rightarrow \pi^*$ transition, that enables circular dichroism to indicate overall secondary structural features [18]. Likewise, the CD spectra of aromatic amino acids are also influenced by the rigidity of the protein, hydrogen bonding, polar groups, polarisability and exciton coupling.

The calculation of the secondary structural content of a protein is achieved through a number of sophisticated algorithms that involve spectra of well-characterised proteins of known secondary structure [15]. However, this has limited value as the models are usually based on an over-simplification of anticipated protein structure. In addition, for the analysis of far UV CD with confidence the data should be available to low wavelengths, 175 nm, which is not possible with conventional CD instruments. However, these approaches are useful for studying changes in protein structure, or indeed where no other information is available. For example, the secondary structure content of lysozyme predicted from the spectrum shown in

Figure 3.2 was analysed using the DICHROWEB service [19, 20, and references therein]. This gave a prediction of α-helix 30.5%, β-sheet 18.3%, turns 23% and unordered structures 28.3%**.** [4]




a)




b)

***Figure 3.2 CD Spectra of Lysozyme***

***(a) Far UV CD spectrum of lysozyme (1 mg/ml) in phosphate buffer (pH 7, 30 mM) solution using 0.1 mm path length cell and 36 accumulations.***

***(b) Near UV CD spectrum of the same solution using a 1 cm path length cell and averaging over 24 accumulations.***

***Spectra were collected using Jasco J-810 CD spectrophotometer with scan speed 50 nm/min, data pitch 0.1 nm, bandwidth 1 nm and response time 1 s.***

[4] Settings used were algorithm SELCON3 and reference database 3.

Compared with X-ray crystallography and NMR, the major advantages of CD come from the speed and convenience of the technique [15]. It is an added advantage of CD that a range of cells of different path lengths can be used to study a very wide range of protein concentrations. Since it is a non-destructive technique, multiple experiments can be conducted on the same solution. Moreover, CD studies can be performed over a wide range of experimental conditions such as pH and temperature in solution. Stopped flow CD also can be exploited in the study of structural changes that occur within a few tens of milliseconds.

Because CD provides only very limited structural information, it is in fact most useful as a comparative technique; for example, in the comparison of batches of biopharmaceuticals. Here, the use of objective pattern recognition techniques such as PCA is useful. To make full use of CD in this context, it is essential to obtain comparable measurements. There are a number of pitfalls, which can be avoided by following good practice (see, for example, [21]).

To be considered complementary to CD, a technique should share its advantages of relatively rapid measurement, and should ideally provide both confirmation of the CD measurement and also additional structural information. It should also be suitable for reproducible, comparable measurement.

### 3.3 Fluorescence Spectroscopy

Fluorescence refers to the emission of radiation from an excited electronic state of a molecule at a longer wavelength than that is required for its absorption [22]. The fluorescence process is highly sensitive to perturbations caused by the chemical environment of the fluorophores. Like absorption, fluorescence also shows broad spectral features (Figure 3.3). This is due to the loss of fine vibronic structure as a result of the relatively long time between excitation and emission (known as the fluorescence lifetime).

Among biopolymers, proteins display intrinsic fluorescence due to the presence of aromatic amino acids such as tryptophan, tyrosine, and phenylalanine. The very low content of these amino acids in proteins facilitates the acquisition and analysis of spectral data. The changes in emission spectra of tryptophan, which is the dominant fluorophore, can provide tertiary structural information, as the fluorescence of tryptophan is highly sensitive to its local environment.

The main advantages of fluorescence are its exceptional sensitivity and the ability to collect dynamic information of specific sites within the protein [23]. However, the interpretation of the spectrum is not always easy in the case of multi-tryptophan proteins as the emission spectrum is the result of the sum of the emissions from each residue and they overlap at most usable wavelengths [22]. In addition, tryptophan displays complex spectral properties and is also sensitive to a variety of quenchers present in protein itself and in solution. Moreover, the local environment can change during the excited-state lifetime and result in changes in tryptophan emission.

As with many of the other low-resolution techniques, fluorescence spectroscopy is most useful in a comparative context. However, the limited information it returns in most cases severely limits its applicability to biopharmaceuticals.

More detailed information can be obtained by labelling molecules with extrinsic fluorophores. However, this is obviously a destructive process and not suitable for quality control purposes.

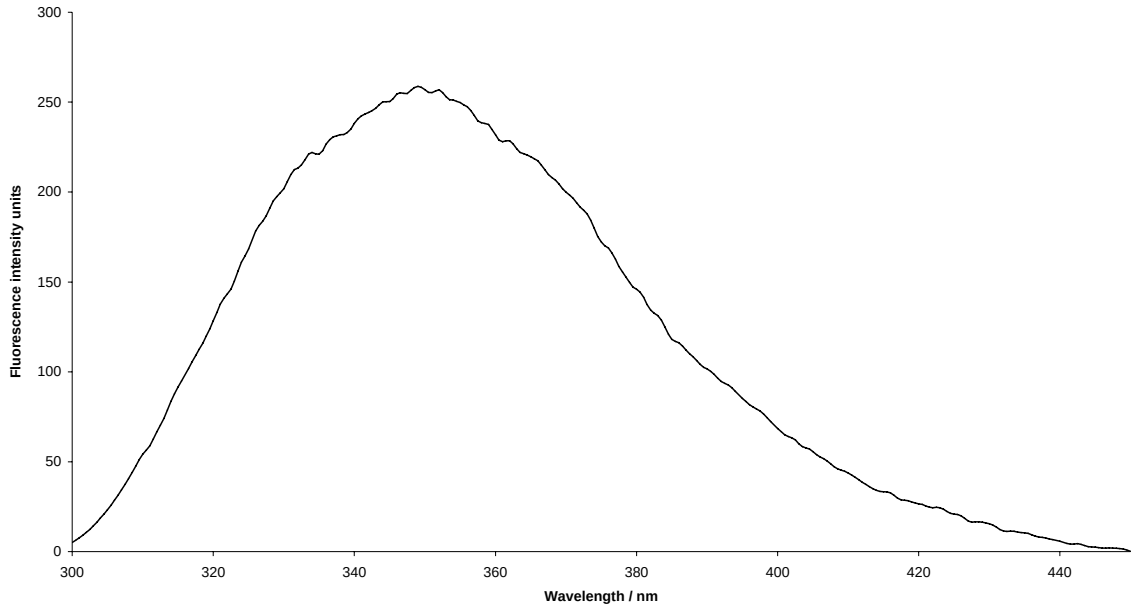



***Figure 3.3 Fluorescence spectrum of lysozyme***

***The spectrum of lysozyme at ~0.01 mg/ml in phosphate buffer (pH 7, 30 mM) was measured using a PerkinElmer LS-55 spectrofluorimeter in a 1 cm cell. The sample was excited at 280 nm and the emission spectrum was collected from 300 nm to 450 nm using a 290 nm cut off filter with emission and excitation slits at 5 nm.***

## 3.4 Raman Spectroscopy

Raman spectroscopy is based on the measurement of the wavelength and intensity of the inelastically scattered light from molecules. In Raman scattering, the incident and scattered photons differ by the energies of molecular vibrations and therefore it occurs at wavelength shifts corresponding to that small energy difference. Although Raman spectroscopy provides information about the vibrational characteristics of molecules it is a complementary technique to IR spectroscopy that occurs via a different mechanism [24, and references therein]. The Raman effect is quite challenging to observe as it only contains a small fraction of the incident photons (about 1 in $10^7$) at wavelengths very close to the incident radiation [25]. The advent of accessible and considerably less expensive laser sources and sensitive detectors have made Raman a viable protein structure determination technique.

Raman spectroscopy can provide information regarding the ratios of various amide conformations, but not the conformation of a specific peptide bond. It is possible to extract dihedral angles and their related conformers from Raman spectrum. Not only can conformational information about amino acids such as tyrosine and tryptophan be determined from Raman, it also provides a sensitive indication of changes in protein structure.

The major advantage of Raman spectroscopy, compared to infrared spectroscopy, is that water does not give strong signals; therefore it can readily obtain information from aqueous solutions [24, and references therein]. However, it cannot provide such detailed information about peptides and proteins as NMR. The main reason for this is that, like water, proteins and peptides show low polarisability, and therefore exhibit weak Raman spectra. However, the amide bands, as the main component of the peptide chain, show comparable intensity. As a result the protein concentration needs to be one to two orders of magnitude more concentrated than samples used with CD.

A problem frequently encountered in protein Raman spectroscopy is that proteins exhibit background fluorescence, which can mask the Raman scattering [24, and references therein]. This may arise from tryptophan residues in the proteins or from fluorescent impurities. It may be possible to remove the impurities by further purification of the proteins. Fluorescence may also be avoided by the selection of longer wavelength lasers that do not excite the fluorescence. For example, a red or IR laser coupled with Fourier transform (FT-Raman) or CCD detection may be suitable. Also, care must be taken to keep proteins from adsorbing onto glass walls of the container during heating by the incident beam.

Two refinements of Raman spectroscopy have been developed, and applied to biological molecules [24, and references therein]. Resonance Raman spectroscopy uses an excitation wavelength that coincides with the absorption of the molecule, and thereby increases the amount of Raman scattering. In Surface-Enhanced Raman spectroscopy (SERS), the Raman effect is enhanced, and fluorescence quenched, in proximity to metal surfaces such as colloids.

Although a potential candidate for a complementary technique to CD, the low sensitivity and problems of background fluorescence are significant problems, and suggest that this technique is still in its infancy, and not suitable for routine quality control. Surface-enhanced Raman solves both of these problems, but the use of metal colloids introduces additional complications and means that the technique is no longer non-destructive (see also [26]).

### 3.5 Vibrational Circular Dichroism

The differential absorption of left and right handed circularly polarised light in the infrared region of the spectrum is referred to as vibrational circular dichroism (VCD)[5]. VCD has the usual advantages of IR and CD spectroscopies; in addition it is a property of the ground electronic state of the molecule, which offers the potential of avoiding some of the limitations of electronic CD [27, and references therein]. Thereby electronic CD and VCD are considered to have a complementary relationship that could enhance the structural information gleaned from either.

The main disadvantage of this technique is the very low intensity of the signals measured, i.e. the very low signal to noise ratio [27, and references therein]. These signals are a few orders of magnitude lower than in electronic CD and an order of magnitude lower than in conventional IR spectroscopy. A very sensitive instrument is therefore required for VCD measurement. However, the development in technology has resulted in instrumentation that makes the measurement of VCD routine over much of the IR region. Having said that, there are only a few commercial systems available, and they cannot be considered equivalent to commercial conventional CD instruments, as most VCD instruments are constructed “in house” by assembling various commercial components together.

Since VCD is the combination of IR and CD it is capable of determining structural characteristics of a wider range of molecules, such as glycoproteins. However, the real challenge of VCD is the reliable and efficient extraction of structural information of proteins, as detailed theoretical analysis is still developing. Although VCD reveals new insights into protein structure, there are many aspects such as measurement conditions and sampling methods, which need to be investigated. As this technique is in its initial stages of

[5] Circular dichroism in the UV and visible regions is sometimes called *electronic* CD, to make the distinction clear.

development it can be concluded that it is not suitable as a routine quality control technique complementary to CD.

## 3.6 Raman Optical Activity

Raman optical activity (ROA) is an analogous technique to VCD, as Raman scattering and infrared absorbance spectroscopy are complementary in terms of the selection rules for activity. ROA refers to the differential Raman scattering of right and left-circularly polarised light in the presence of chiral molecules [28]. Barron *et al.* first observed this effect in the liquid phase in 1973 [29]. Since then, few practical applications have been found for ROA due to the lack of sensitivity. However, the advent of instruments based on back scattering and charge-coupled device (CCD) detection have introduced a wider range of applications for ROA studies including biological macromolecules in aqueous solution.

Like many of these techniques, ROA spectra are difficult to interpret *a priori*, although they do contain detailed structural information. The technique is also significantly less sensitive than Raman spectroscopy, and acquisition times of several hours are not unusual.

Although ROA has potential to analyse the extended secondary structure and dynamics of biopolymers, including glycoproteins, in solution, there is only one commercial system available. This, and the long acquisition times, means that this technique is not yet suitable for quality control testing of biopharmaceuticals, although it is to be hoped that technological improvements will realise the potential of this technique.

## 3.7 Fourier Transform Infrared Spectroscopy (FTIR)

Infrared spectroscopy is the measurement of the incident infrared radiation, 0.78–1000 μm, that is absorbed at a particular energy when it interacts with electric dipoles within a molecule. The infrared spectrum is basically a plot of transmittance of a sample as a function of wave number. In contrast to Raman, a molecule is IR active when a vibration causes a net change in the dipole moment of the molecule.

Molecular vibrations determine the characteristic frequency at which the absorption takes place, depending upon the masses of atoms of the molecule, their spatial geometry, and strengths of the connecting bonds. Therefore an infrared spectrum is characteristic of a vibration of a chemical bond in that molecule.

Infrared spectroscopy has developed into a major tool for protein structure analysis since Elliott and Ambrose discovered the correlation between protein secondary structure and the position of the infrared bands [30]. However, new dimensions of biological infrared spectroscopy were opened after the development of the Fourier-transform infrared (FTIR) technique. This has in turn led to high accuracy and reproducibility for IR measurements, dramatic increase in signal-to-noise ratio, and the ability to perform measurements with low-transmittance samples such as aqueous solutions of proteins [31, and references therein] (Figure 3.4). The combination of FTIR instruments and advanced mathematical data analysis methods have enabled the separation of overlapping bands by Fourier self-deconvolution procedures and thereby enabled the quantitation of different conformational structures present in a protein. The different secondary structural content of protein can be revealed not only utilizing the amide I band but also amide II and III bands [32]. For example, we analysed the spectrum shown in Figure 3.1 using the Bruker OPUS software[6], which suggests that the secondary structure content of lysozyme is 37% α-helix, and 19% β-sheet.

[6] The spectrum was analysed using the Quant Analysis 2 method of the Bruker Opus software, which is based on a partial least squares model.

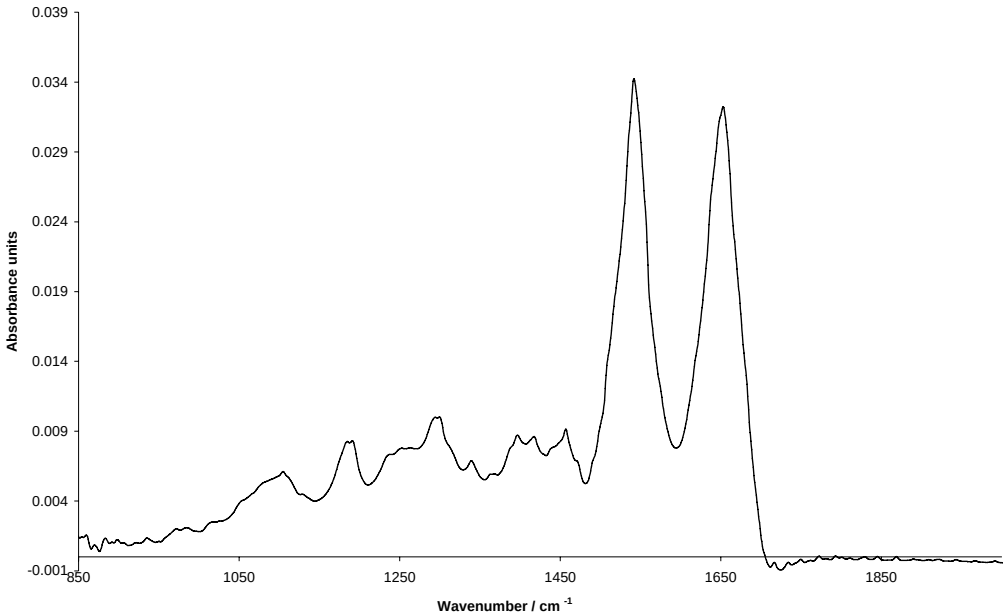



***Figure 3.4 FTIR spectrum of lysozyme***

***Lysozyme was at (20 mg/ml) in phosphate buffer solution (pH 7, 30 mM). The spectrum was collected using a Tensor 27 series FTIR instrument from Bruker Optics with a resolution of 4 $cm^{-1}$, 128 scans, 20 kHz scanner velocity and an aperture setting of 6 mm.***

As mentioned previously, FTIR and Raman spectroscopies are considered complementary to each other (section 3.4); however, FTIR spectroscopy offers several advantages over Raman. FTIR spectroscopy is less sensitive to fluorescence artefacts than Raman and also causes less damage to the sample. The major limitation, which has restricted the use of this technique in protein analysis in the past, was the absorption of water, which strongly overlaps with the Amide I band of the protein. However, the development of a new sampling technology, attenuated total reflectance (ATR), has made a substantial contribution to overcoming this problem.

The ATR sampling method is based on a total internal reflection principle. The incident light enters a high refractive index crystal, and undergoes multiple total internal reflections at the interface with a sample of lower refractive index [33]. The beam then emerges at the output end of the crystal, and subsequently enters the normal beam path of the spectrometer.

In each reflection at the sample interface, an evanescent wave is generated in the sample, and absorption by the sample can occur. The evanescent wave decays exponentially away from the interface, and there is no path length as such [34]. However, an effective path length can be derived which is equivalent to the path length in transmission mode. This depends on the refractive indices of the sample and crystal, the angle of incidence, the number of internal reflections and the wavelength of the incident light [35]. This effective path length is very small, of the order of 5 μm, and so the interference from water absorption is highly reduced, enabling the use of aqueous solutions and non-transparent samples. However, the wavelength dependence of this effective path length, and other factors, mean that spectra must be corrected before they can be compared to spectra collected in transmission mode.

FTIR is highly reliable in the assessment of the β-sheet where as CD is reliable for α-helical protein structure [1, 36]. Since the potential sources of error in the CD and FTIR techniques are different they may be considered to be complementary techniques. It is therefore recommended to use these two techniques in conjunction, as the samples used for CD can be recovered and analysed by FTIR. Moreover, the use of two independent techniques (FTIR

and CD) increases the confidence in the results obtained by spectral deconvolution or other analysis techniques. However, the discrepancy in the quantitative analysis of structural components of lysozyme using CD and FTIR strongly highlights the need for the validation of the FTIR technique against the well-established CD technique. Modern FTIR instrumentation is reasonably economic to buy, and good-quality spectra may be acquired in a few minutes. The software for spectral analysis and deconvolution is also relatively well developed. The above considerations strongly suggest that FTIR is a suitable complementary technique to CD for quality control of biopharmaceuticals.

# 4 Conclusions

The “high resolution” structural techniques discussed above are the “gold standard” for obtaining reference structural information for protein molecules such as biopharmaceuticals. However, they are both time-consuming procedures requiring extensive, technically demanding, data collection and analysis. With NMR there is the possibility that, with further technical developments, rapid one-dimensional NMR spectroscopy may become a useful quality control tool. However, for the present the cost and time required for measurements preclude its widespread application.

Among the “low resolution” techniques, unpolarised UV-Vis absorbance and fluorescence spectroscopies return very limited structural information, and as such are not suitable complements to CD. However, both may play a niche role; for example, UV-Vis is useful for accurate protein concentration measurements. Raman spectroscopy, Raman optical activity and Vibrational CD all return more detailed information, but suffer from limited sensitivity, and in the case of the latter two, limited availability of instrumentation and problems with data analysis.

FTIR spectroscopy returns detailed structural information and the instrumentation and software for protein structural analysis are relatively well developed. The cost of the measurements – both in terms of the instrumentation and the time required to collect data – is reasonably economic, and comparable with CD. The information obtained is in some respects complementary, and in others confirmatory, of CD measurements. Although some issues with obtaining comparable measurements need to be addressed, we conclude that FTIR is at present the most suitable measurement technique complementary to CD (Figure 4.1). Further work under this project will address the outstanding issues pertaining to this technique.

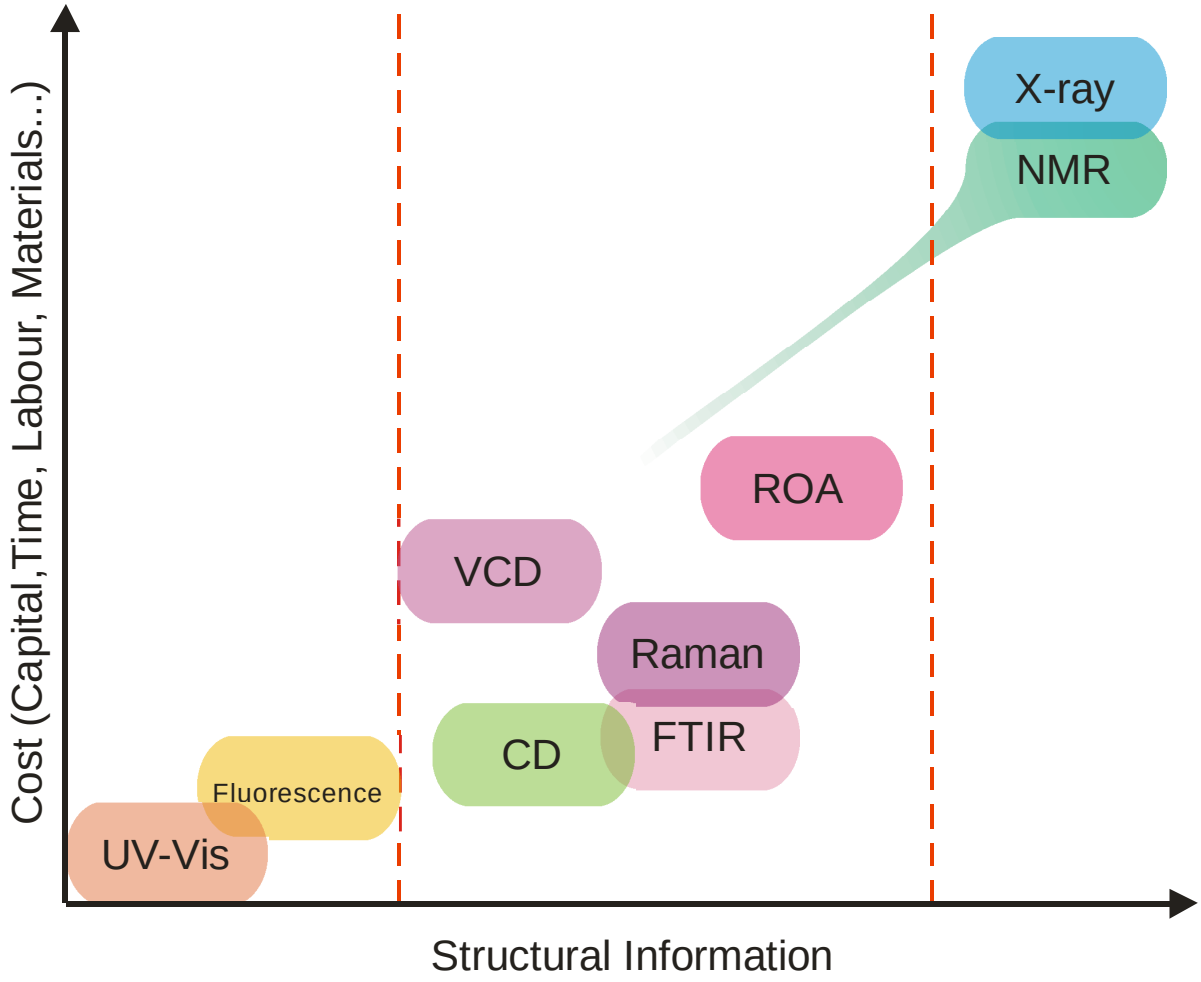



***Figure 4.1 Summary of techniques***

***A qualitative map illustrating the suitability of protein structural determination techniques for biopharmaceutical quality control. Each technique has a different cost (in terms of capital, labour, etc.) and can return a given amount of structural information. At top right, X-ray crystallography and NMR return intricate detail but are relatively expensive. In the centre section are techniques that return a lesser degree of structural detail, typically including overall secondary structure and some information about amino-acid side chains. Finally, in the left hand panel, UV absorbance spectroscopy and native fluorescence return minimal structural detail. The various forms of vibrational spectroscopy bridge the gap between CD and NMR, but of these FTIR is the cheapest and best-developed technique.***

# 5 References


1. Nguyen, L. T., Wiencek, J. M., and Kirsch, L. E., Characterisation methods for the physical stability of biopharmaceuticals. *PDA Journal of Pharmaceutical Science and Technology* **2003,** 57, (6), 429-445.
2. *Specifications: Test Procedures And Acceptance Criteria for Biotechnological/Biological Products (Q6B)*; International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use: 1999.
3. *State of the art analytical methods for the characterisation of biological products and assessment of comparability*. S. Karger A.G.: Basel, 2003; Developments in Biologicals Vol. 122.
4. Oberg, K. A.; Ruysschaert, J.-M.; Goormaghtigh, E., The optimization of protein secondary structure determination with infrared and circular dichroism spectra. *European Journal of Biochemistry* **2004,** 271, (14), 2937-2948.
5. Abola, E., Kuhn, P., Earnest, T., and Stevens, R. C., Automation of X-ray crystallography. *Nature Structural Biology* **2000,** Structural genomics supplement, 973-977.
6. Goodwill, K. E., Tennant, M. G., and Stevens, R. C., High-throughput x-ray crystallography for structure-based drug design. *Drug Discovery Today* **2001,** 6, (15 (Suppl.)), 113-118.
7. Gronenborn, A. M., and Clore, G. M., Nuclear magnetic resonance studies of small and medium sized proteins in solution. In *Spectroscopic methods for determining protein structure in solution*, Havel, H. A., Ed. John Wiley and Sons: 1996; pp 190-246.
8. Price, N. C., Conformational issues in the characterization of proteins. *Biotechnology and Applied Biochemistry* **2000,** 31, 29-40.
9. Pelton, J. T., and McLean, L. R., Spectroscopic methods for analysis of protein secondary structure. *Analytical Biochemistry* **2000,** 277, (2), 167-176.
10. Wider, G., and Wuthrich, K., NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Current Opinion in Structural Biology* **1999,** 9, (5), 594-601.
11. Clore, G. M., and Gronenborn, A. M., Determining the structures of large proteins and protein complexes by NMR. *Trends in Biotechnology* **1998,** 16, (1), 22-34.
12. Grzesiek, S., *Characterisation of biotechnology pharmaceutical products*. International association of biological standardisation: 1998; Structural characterisation of proteins by NMR Vol. 96, p 37-41.
13. Rodger, A., and Norden, B., *Circular dichroism and linear dichroism*. Oxford University Press: 1997.
14. Havel, H. A., Derivative near-ultraviolet absorption techniques for investigating protein structure. In *Spectroscopic methods for determining protein structure in solution*, Havel, H. A., Ed. John Wiley & Sons: 1996; pp 62-68.
15. Kelly, S. M.; Price, N. C., The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science* **2000,** 1, (4), 349-384.
16. Drake, A. F., *Protein-ligand interactions: Structure and spectroscopy-A practical approach*. Oxford University Press: 2001; Circular dichroism Vol. 2, p 123-167.

17. Greenfield, N. J., *Encyclopedia of spectroscopy and spectrometry*. Academic press: 2000; Biomacromolecular applications of circular dichroism and ORD Vol. 1, p 117-130.

18. Mulkerrin, M. G., Protein structure analysis using circular dichroism. In *Spectroscopic Methods for determining protein structure in solution*, Havel, H. A., Ed. John Wiley & Sons: 1996; pp 5-27.

19. Lobley, A.; Whitmore, L.; Wallace, B. A., DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **2002,** 18, (1), 211-2.

20. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **2004,** 32, (Web Server issue), W668-73.

21. Jones, C.; Schiffmann, D.; Knight, A.; Windsor, S. *Val-CiD Best Practice Guide: CD spectroscopy for the quality control of biopharmaceuticals*; DQL-AS 008; National Physical Laboratory: Teddington, October, 2004.

22. Lakowicz, J. R., *Principles of fluorescence spectroscopy*. Second ed.; Kluwer academic/Plenum publishers: 1999.

23. Ladokhin, A. S., Fluorescence Spectroscopy in Peptide and Protein Analysis. *Encylopedia of Analytical Chemistry* **2000**.

24. Tensmeyer, L. G., and Kauffman, E. W., Protein structure as revealed by nonresonance Raman spectroscopy. In *Spectroscopic methods for determining protein structure in solution*, Havel, H. A., Ed. John Wiley & sons: 1996; pp 69-95.

25. Raman, C. V., and Krishnan, K. S., A new type of secondary radiation. *Nature* **1928,** 121, (3048), 501.

26. Tantra, R.; McCabe, A.; Bailey, M.; Knight, A.; Smith, E. *Comparable Raman Spectroscopy Across Different Instruments and Excitation Wavelengths*; DQL-AS 012; National Physical Laboratory: Teddington, November, 2004.

27. Keiderling, T. A., Protein structural studies using vibrational circular dichroism spectroscopy. In *Spectroscopic methods for determining protein structure in solution*, Havel, H. A., Ed. John Wiley & Sons: 1996; pp 163-189.

28. Barron, L. D., Hecht, L., Bell, A. F., and Wilson, G., Recent developments in Raman optical activity of biopolymers. *Applied spectroscopy* **1996,** 50, (5), 619-629.

29. Barron, L. D.; Bogaard, M. P.; Buckingham, A. D., Raman scattering of circularly polarized light by optically active molecules. *Journal of the American Chemical Society* **1973,** 95, (2), 603-605.

30. Elliott, A., and Ambrose, E. J., Structure of synthetic polypeptides. *Nature* **1950,** 165, 921-2.

31. Surewicz, W. K., and Mantsch, H. H., Infrared absorption methods for examining protein structure. In *Spectroscopic methods for determining protein structure in solution*, Havel, H. A., Ed. John Wiley & sons: 1996; pp 135-162.

32. van de Weert, M., Haris, P. I., Hennink, W. E., Crommelin, D. J., Fourier transform infrared spectrometric analysis of protein conformation: effect of sampling method and stress factors. *Anal Biochem* **2001,** 297, (2), 160-9.

33. Parker, F. S., *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry*. Plenum press, New York: 1983.

34. Mirabella, F. M., Jr; Harrick, N. J., Correction of Internal Reflection Spectra for the Wavelength Dependence. In *Internal Reflection Spectroscopy: Review and Supplement*, Harrick Scientific Corporation: Ossining, NY, 1985; pp 84-91.

35. Belton, P. S., New methods for monitoring changes in proteins. *Food Reviews International* **1993,** 9, (4), 551-573.

36. Surewicz, W. K., Mantsch, H. H., Chapman, D., Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment. *Biochemistry* **1993,** 32, (2), 389-94.