

NPL REPORT AS 9

Higher-Order Protein Structure Measurements for Biopharmaceutical Quality Control

**Jascindra Ravi, Delphine
LePevelen, George Tranter,
Marc J. A. Bailey, Alex E.
Knight**

NOT RESTRICTED

AUGUST 2007

Higher-Order Protein Structure Measurements for Biopharmaceutical Quality Control

Jascindra Ravi, Delphine LePevelen*, George Tranter*, Alex Knight

Quality of Life Division

*Chiralabs Ltd., Begbroke Centre for Innovation and Enterprise,
Oxford University Begbroke Science Park, Sandy Lane, Yarnton,
Oxfordshire, OX5 1PF, UK

www.chiralabs.biz

ABSTRACT

This report reviews the available techniques for measuring the higher-order structure of proteins, and assesses their suitability for quality control in the biopharmaceutical industry. FTIR is found to be the technique which best complements CD, currently the dominant technique. We explore measurement issues of FTIR and make best practice recommendations. We also advise that FTIR should be used in combination with CD.

© Crown copyright 2007
Reproduced with the permission of the Controller of HMSO
and Queen's Printer for Scotland

ISSN 1754-2928

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

Extracts from this report may be reproduced provided the source is acknowledged and
the extract is not taken out of context.

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

Contents

1	Introduction	7
1.1	Context	7
1.2	Background	7
1.3	Abbreviations	8
2	Review of Techniques	9
2.1	Introduction	9
2.2	High Resolution Techniques	9
2.2.1	X-Ray Crystallography	9
2.2.2	Nuclear Magnetic Resonance (NMR)	9
2.3	Low Resolution Techniques	10
2.3.1	Ultraviolet-Visible (UV-Vis) Absorption Spectroscopy	10
2.3.2	Circular Dichroism (CD)	11
2.3.3	Fluorescence Spectroscopy	14
2.3.4	Raman Spectroscopy	15
2.3.5	Vibrational Circular Dichroism	15
2.3.6	Raman Optical Activity	16
2.3.7	Fourier Transform Infrared Spectroscopy (FTIR)	16
2.4	Conclusions	18
3	FTIR Spectroscopy	20
3.1	Introduction	20
3.2	Theory of Infrared Spectroscopy	20
3.3	Advantages of FTIR Over Dispersive Spectrometers	21
3.4	Instrumentation	21
3.4.1	Fourier-transform spectrometers	21
3.4.2	Detectors	23
3.4.3	Sampling techniques	23
3.5	Presentation of Spectra	25
3.6	Protein Secondary Structure Determination Using FTIR	26
3.7	Water interference and subtraction	27
3.8	Quantitative information from FTIR	28
3.9	Conclusions	29
4	Investigation of FTIR of proteins	30
4.1	Introduction	30

4.2	Materials and Methods	30
4.2.1	Materials	30
4.2.2	FTIR	30
4.3	Results and Discussion	30
4.3.1	Responsivity of the detector	34
4.3.2	Water subtraction	36
4.3.3	Protein adsorption onto the crystal and its kinetics with time	39
4.3.4	Cleanliness of the crystal	45
4.3.5	Evaluation of glycosylation of proteins by FTIR	46
4.3.6	Effect of salt and pH	46
4.4	Conclusions	48
5	Chemometric analysis of FTIR and CD spectra	49
5.1	Introduction	49
5.2	Collection of Data Set	49
5.2.1	Selection of panel	49
5.2.2	Materials	50
5.2.3	FTIR	50
5.2.4	CD	51
5.3	Spectra	51
5.4	Chemometric analysis	56
5.4.1	Far-UV CD	56
5.4.2	Near-UV CD	58
5.4.3	FT-IR	60
5.4.4	Comparison of techniques	63
5.4.5	Combination of CD and FT-IR Data	63
5.4.6	Conclusions	66
6	Summary	67
7	Acknowledgements	69
8	Appendix: ATR Theory	70
8.1	Theory	70
8.2	Conclusions	71
8.3	Example Calculations	72
9	References	73

Figures

Figure 2.1	Absorbance spectrum of lysozyme	11
Figure 2.2	CD Spectra of Lysozyme	13
Figure 2.3	Fluorescence spectrum of lysozyme	14
Figure 2.4	FTIR-ATR spectrum of lysozyme.....	17
Figure 2.5	Summary of techniques.....	19
Figure 3.1	Schematic illustration of FTIR system.....	22
Figure 3.2	The basic optics of a Michelson interferometer.	23
Figure 3.3	Schematic of a typical ATR accessory for liquids.	24
Figure 4.1	UV-Visible absorbance spectra of lysozyme in water.	31
Figure 4.2	FTIR-ATR spectra of lysozyme in water.....	32
Figure 4.3	FTIR-ATR spectra of lysozyme in water.....	33
Figure 4.4	Lysozyme (10 mg/ml) FTIR-ATR spectra collected on three different days.	34
Figure 4.5	Responsivity variation upon nitrogen cooling of the detector.	34
Figure 4.6	Responsivity variation with nitrogen cooling of the detector.	35
Figure 4.7	Responsivity variation with nitrogen cooling of the detector	36
Figure 4.8	Overlaid FTIR-ATR lysozyme (10mg/ml) and water spectra.	37
Figure 4.9	FTIR-ATR protein spectra after water subtraction.	38
Figure 4.10	FTIR-ATR spectra of lysozyme in water.....	40
Figure 4.11	Absorbance maxima of the amide I and II bands as a function of lysozyme concentration.	40
Figure 4.12	FTIR-ATR absorbance spectra of 6 mg/ml lysozyme in water as a function of time.....	41
Figure 4.13	Amide I absorbance of lysozyme in water (6mg/ml) with time.....	42
Figure 4.14	Absorbance spectra of 10 mg/ml lysozyme in buffer as a function of time.....	42
Figure 4.15	Amide I absorbance intensity of proteins in buffer vs. time.	43
Figure 4.16	Amide I absorbance intensity of proteins in buffer vs. time (mins) after normalising the adsorption at 1 min to zero.....	43
Figure 4.17	FTIR-ATR absorbance spectra of 1 mg/ml lysozyme in buffer as a function of time.....	44
Figure 4.18	Amide I absorbance intensity of lysozyme in water (1 mg/ml) with time (min).	44
Figure 4.19	Spectra of lysozyme after incubating for 15 mins on the ATR crystal followed by a 20 μ l buffer wash.	45
Figure 4.20	FTIR/ATR spectra of lysozyme, RNase A and RNase B.	46
Figure 4.21	Effect of salt concentration on water FTIR-ATR spectra.	47
Figure 4.22	Effect of pH on water FTIR-ATR spectrum	47
Figure 5.1	Secondary structure composition of proteins in panel	50

Figure 5.2	Far UV CD Spectra	52
Figure 5.3	Near UV CD Spectra.....	53
Figure 5.4	Background-subtracted FTIR-ATR spectra	54
Figure 5.5	FTIR-ATR spectra, normalised to Amide I peak.....	55
Figure 5.6	Amide I peaks from normalised spectra.....	56
Figure 5.7	Far-UV CD Covariance PCA.....	57
Figure 5.8	Near-UV CD Covariance PCA	59
Figure 5.9	FT-IR Covariance PCA.....	61
Figure 5.10	FT-IR 1480-1720cm ⁻¹ Covariance PCA	62
Figure 5.11	Far-UV CD & amide region FT-IR “Super-spectra” Covariance PCA	64

Tables

Table 1.1	Abbreviations used in this document	8
Table 4.1	Nominal concentration of the lysozyme solutions vs. measured concentrations.....	32
Table 4.2	The scaling factors utilized to obtain a flat baseline in the 1700–1850 cm ⁻¹ region during water subtraction on three consecutive days.....	39
Table 5.1	Protein Panel for Comparison of FTIR and CD	49
Table 5.2	Far-UV CD Regressions.....	58
Table 5.3	Near-UV CD Regressions	59
Table 5.4	FT-IR Regressions.....	61
Table 5.5	FT-IR (1480-1720cm ⁻¹) Regressions	62
Table 5.6	Far-UV CD & amide region FT-IR “Super-spectra” Regressions using the first three PCA Factors	65
Table 5.7	Far-UV CD & amide region FT-IR “Super-spectra” Regressions using the first four PCA Factors	66
Table 6.1	Issues and recommendations in FTIR analysis of protein structure.....	68

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>. Part of this report was previously issued as NPL Report DQL-AS023.

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 (Q6B¹) are circular dichroism (CD) and NMR [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 2003², 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 additional techniques, which provide additional information, and which confirm the results of CD measurements.

In this report we have set out to review the available techniques, and identify the techniques most suitable for application to biopharmaceutical quality control. We have also sought to identify the problems that may occur with these techniques, and to make recommendations as to how the best quality data may be obtained. Finally, we adopt chemometric analysis approaches to compare the structural sensitivity of the chosen technique with CD. The sections of the report are as follows:

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

² “State of the Art Analytical Methods for the Characterization of Biological Products and Assessment of Comparability”, Bethesda, MD, 2003. Proceedings currently in press.

- ❑ In Chapter 2, each of the existing techniques is reviewed, with particular emphasis on suitability for biopharmaceutical characterisation and complementarity to CD.
- ❑ In Chapter 3, a brief introduction to Fourier Transform Infrared spectroscopy is presented.
- ❑ In Chapter 4, we present the results of an investigation of the FTIR spectroscopy of proteins.
- ❑ In Chapter 5, we present a chemometric analysis and comparison of CD and FTIR data.
- ❑ The overall conclusions of the report, together with brief recommendations, are summarised in Chapter 6.

1.3 Abbreviations

Abbreviation	Definition
ATR	Attenuated Total Reflectance
CCD	Charge-coupled device
CD	Circular Dichroism
FT	Fourier Transform
FTIR	Fourier Transform Infra-Red spectroscopy
IR	Infra-red
NMR	Nuclear Magnetic Resonance
PCA	Principal component analysis
PLS	Partial Least Squares
QC	Quality Control
ROA	Raman optical activity
SERS	Surface-Enhanced Raman Spectroscopy
UV	Ultraviolet
UV-Vis	Ultraviolet-visible
VCD	Vibrational Circular Dichroism
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use

Table 1.1 **Abbreviations used in this document**

2 Review of Techniques

2.1 Introduction

In this chapter we review the available techniques for measuring protein structure. 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.

2.2 High Resolution Techniques

2.2.1 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.

2.2.2 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, ^1H , 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.

2.3 Low Resolution Techniques

2.3.1 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 (such as $\pi \rightarrow \pi^*$ 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 2.1). 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 far UV (< 250 nm) region is less informative than CD and requires precise concentration measurements to be of value [14].

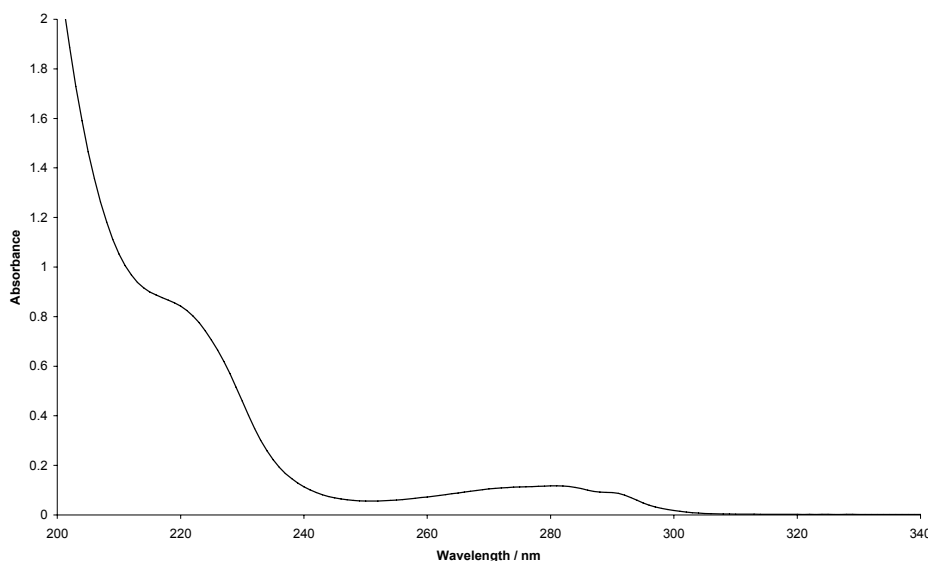


Figure 2.1 Absorbance spectrum of lysozyme

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 [15, 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 [15, 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 a suitable alternative to CD, although its utility in making accurate protein concentration measurements is undeniable. Indeed, a particularly useful feature of CD instruments is their ability to obtain absorbance and CD spectra simultaneously.

2.3.2 Circular Dichroism (CD)

CD is an absorptive phenomenon represented as the differential absorption of left and right circularly polarised light [16]. 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); protein secondary structures are also chiral; and proteins possess suitable chromophores (such as peptide bonds and aromatic side chains) which absorb in the UV-Visible region [17].

The far UV region of the CD spectrum (typically 180-260 nm, Figure 2.2(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 2.2(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 [16, 18]. 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 haem.

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, for example through exciton coupling of the repeated peptide linkage transitions. It is this effect that enables circular dichroism to indicate overall secondary structural features [14]. 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 algorithms that involve spectra of well-characterised proteins of known secondary structure [16]. 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 commonly achieved 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 2.2 was analysed using the DICHROWEB service [19, 20, and references therein]. This gave a prediction of α -helix 36%, β -sheet 19%, turns 17% and unordered structures 27%³. Compared with X-ray crystallography and NMR, the major advantages of CD come from the speed and convenience of the technique [16]. 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, [22]).

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 Settings used were algorithm CDSSTR and reference database 3. The NRMSD (goodness of fit) statistic was 0.017. The values calculated from the crystal structure are, α -helix 31%, β -sheet 6.2%, turns 23.3% 3_{10} helix 10.9%, and unordered structures 28.7% [21]. Differences between the predicted and the crystal structure values might be due to a number of factors, including the definition of the different structural classes, the quality of the spectrum and the algorithm and reference database used, or to genuine differences between the solution and crystal structures.

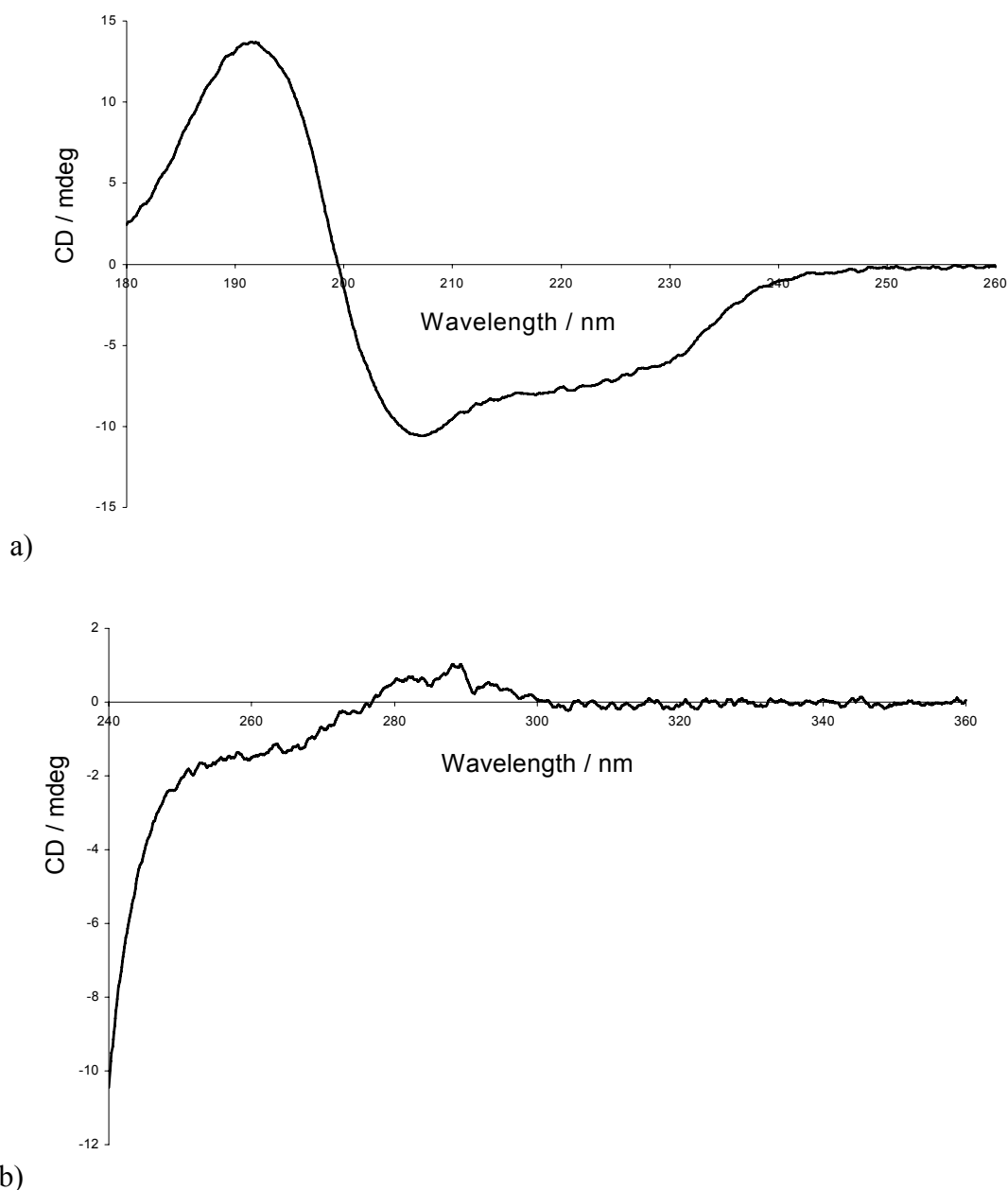


Figure 2.2 CD Spectra of Lysozyme

(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.

2.3.3 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 [23]. 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 2.3).

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 [24]. 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 [23]. 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.

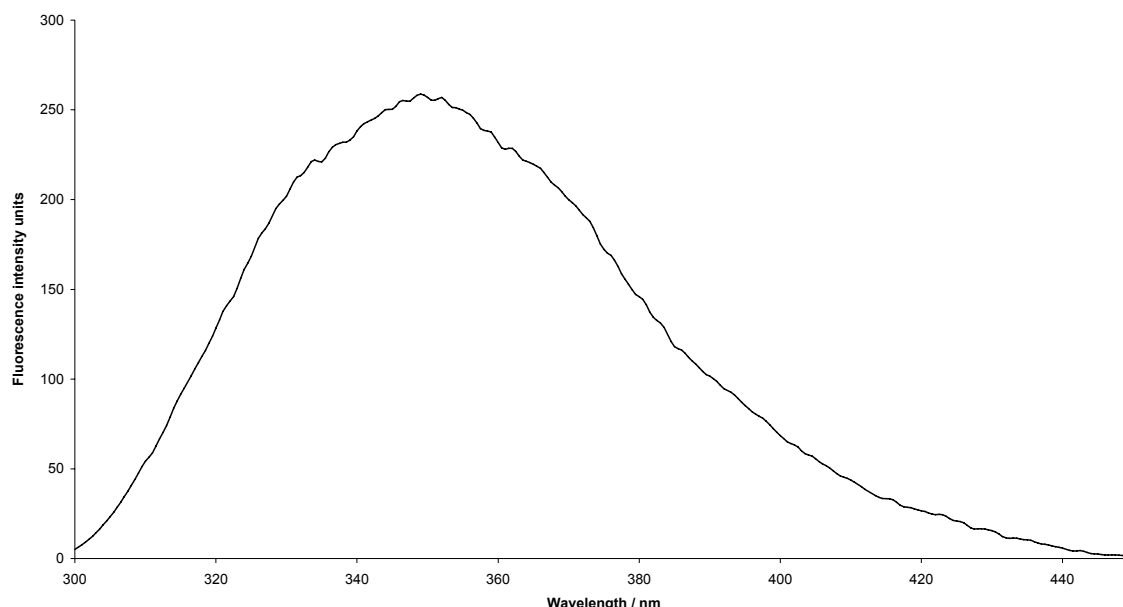


Figure 2.3 Fluorescence spectrum of lysozyme

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.

2.3.4 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 [25, 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 [26]. 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 [25, 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 [25, 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 [25, 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 gold or silver 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 further developments are required before this technique is suitable for routine quality control applications. 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 [27]).

2.3.5 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)⁴. VCD has the usual advantages of IR and CD spectroscopies; in addition it is a property of the ground electronic state of the molecule,

⁴ Circular dichroism in the UV and visible regions is sometimes called *electronic* CD, to make the distinction clear.

which offers the potential of avoiding some of the limitations of electronic CD [28, 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 [28, 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. Furthermore, as with FTIR (see below) water absorption is also a serious problem in studying biopharmaceuticals, as water has a strong absorption in the IR region. As this technique is in its initial stages of development it can be concluded that it is not suitable as a routine quality control technique complementary to CD.

2.3.6 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 [29]. Barron *et al.* first observed this effect in the liquid phase in 1973 [30]. 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, [31] 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.

2.3.7 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 absorbance (or 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 the vibrations of a 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 [32]. However, new dimensions of biological infrared spectroscopy were opened after the development of the Fourier-transform infrared (FTIR) technique (see section 3.3). 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 [33, and references therein] (Figure 2.4). 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 [34]. For example, we analysed the spectrum shown in Figure 2.1 using the Bruker OPUS software⁵, which suggests that the secondary structure content of lysozyme is 38% α -helix, and 19% β -sheet⁶.

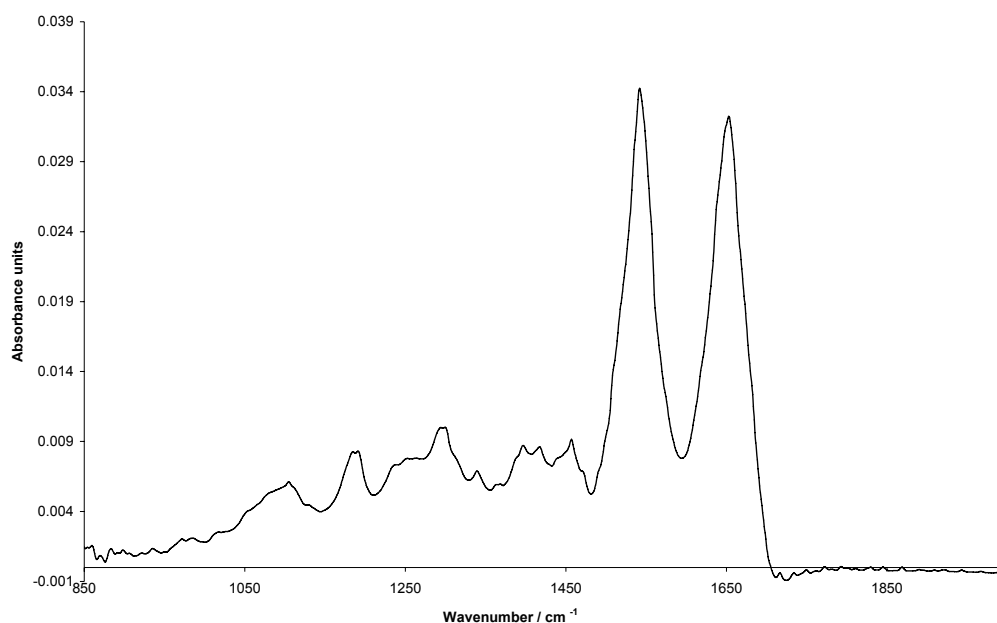


Figure 2.4 FTIR-ATR spectrum of lysozyme

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 2.3.4); 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 (see

⁵ The spectrum was analysed using the Quant Analysis 2 method of the Bruker Opus software, which is based on a partial least squares (PLS) model. The Mahalanobis distances were 0.011 (limit 0.12) for α -helix, and 0.00065 (limit 0.12) for β -sheet, indicating that the spectra are well described by the PLS model used. As noted previously (see footnote 3, p. 7), several factors may lead to discrepancies between structural percentages predicted from spectra and calculated from crystal structures.

⁶ These results are similar to those for the CD spectrum shown in Figure 2.2, see p. 12; where 36% α -helix and 19% β -sheet were predicted. However both show a discrepancy from the calculated secondary structure fractions from the crystal structure, which are α -helix 31%, β -sheet 6.2%. However, as noted previously, different methods and algorithms may classify structures in different ways; for example, 10.9% of 3_{10} helix is predicted from the crystal structure.

section 4.3.2 and Figure 4.8). However, the development of 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 [35]. 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 [36]. 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 [37]. 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 has been shown to have utility in the assessment of the β -sheet where as CD is reliable for α -helical protein structure [1, 38]. 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.

2.4 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 alternatives 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 2.5). The following chapters are dedicated to a discussion of FTIR measurements of protein structure and a comparison with CD.

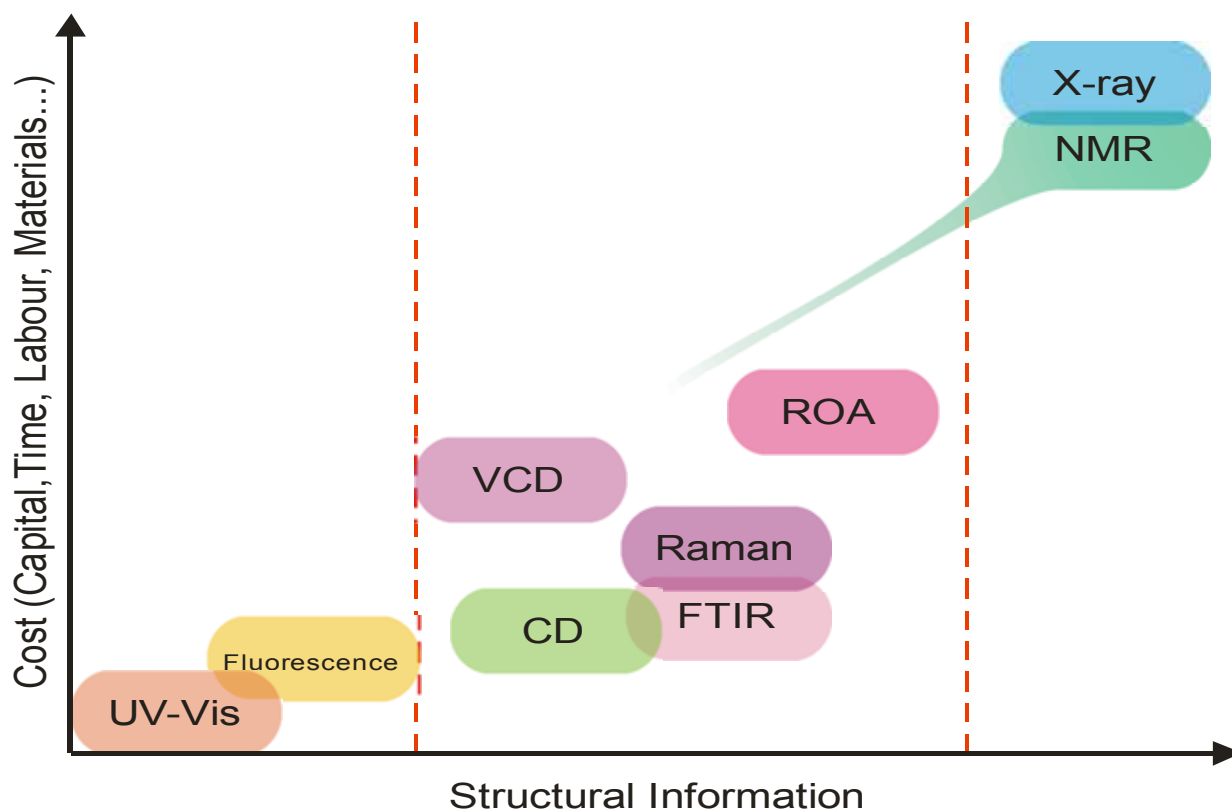


Figure 2.5 Summary of techniques

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.

3 FTIR Spectroscopy

3.1 Introduction

In the previous chapter, it was concluded that FTIR was the most suitable spectroscopic technique to complement circular dichroism in biopharmaceutical quality control. In this chapter, we present a very brief introduction to the FTIR technique, including its potentials and limitations for protein solution analysis.

The infrared (IR) region, which spans wavelengths from 800 nm to 1 mm, is divided into three sub regions: near, mid, and far IR [39]. The near IR region (NIR) is designated from the visible to 2.5 μm , mid IR from 2.5 μm to 25 μm (4000–400 cm^{-1}), and far IR (FIR) beyond 25 μm . The most attractive region is the mid IR, where most fundamental vibrations occur, making it the richest in chemical information. Although the interest in the NIR has dramatically increased over the last few years this report will concentrate only on the use of the mid-IR region.

FTIR came in to widespread use in 1980, although the first commercial instrument was introduced in 1969 and began to replace the older dispersive instruments [39]. These dispersive spectrometers initially had prisms as dispersive elements, later these were replaced with gratings. The advent of FTIR has greatly extended the capabilities of infrared spectroscopy and has been applied to many areas that are very difficult or nearly impossible to analyse by dispersive instruments [40]. The most attractive feature of FTIR spectrometers is that radiation from all wavelengths is measured simultaneously, whereas dispersive instruments measure different wavelengths successively. This enables FTIR to be a much faster and more sensitive technique.

The relationship between the infrared bands and the secondary structure of proteins was first identified by the pioneering work of Elliott and Ambrose [32]. The amide groups of proteins give rise to nine characteristic bands; however, the amide I and II bands have been the most valuable for determining the secondary structures of proteins in solution [41]. The difference in the frequency of the amide I band relates to different secondary structures of protein. However, this amide I band is severely hampered by strong water absorption. The low sensitivity of the dispersive instruments and the difficulties in accurate subtraction of the water absorption from the amide I band limited the use of infrared spectroscopy to protein solutions. The advent of the high sensitivity FTIR instrumentation and the development of methods for the spectral analysis made it feasible to measure the infrared spectrum of protein solutions with small path length cells. The problems encountered in achieving consistent path length throughout the measurements introduced the use of attenuated total reflection (ATR) sampling technique for protein solutions.

The major advantage of FTIR spectroscopy is that any biological material can be studied together in a wide variety of environments. This facilitates the study of proteins in different chemical environment such as crystals, aqueous solution, detergents, and lipid membranes [42]. This ability makes FTIR spectroscopy a valuable tool for the biotechnology and pharmaceutical industry where there is a need to monitor structure and stability of proteins in a variety of environments.

3.2 Theory of Infrared Spectroscopy

All the atoms in molecules are in continuous vibration when at temperatures above absolute zero [40]. The absorption of energy by stretching and bending motions (vibrations) results in the IR absorption spectrum [9]. The types of bonds (strength of bonds and mass of atoms involved in the vibration) and their modes of vibration, determine the frequencies at which the energy is absorbed. The molecule absorbs the radiation when the frequency of a specific vibration matches the frequency of the IR radiation incident on the molecule. Vibrations that produce a net change in the dipole moment may result in IR activity and those that give polarisability changes may give rise to Raman activity.

However, some vibrations can be both IR- and Raman-active. Raman spectroscopy therefore provides complementary information on molecular vibration.

Most small molecules have a unique IR spectrum as a result of the combination of the fundamental vibrations or rotations of various functional groups and the subtle interactions of these functional groups with other atoms of the molecule [40]. Generally the absorption bands observed in the mid IR region are not discrete lines as other rotational motions usually accompany the individual vibrational motion.

3.3 Advantages of FTIR Over Dispersive Spectrometers

FTIR provides fast and sensitive measurements with the same signal-to-noise ratio as the dispersive instruments (1 second vs. 10-15 minutes) [40]. This is achieved by the exploitation of the interferometer, which enables a complete spectrum to be measured during a single scan of the moving mirror, while the detector observes all frequencies simultaneously. The interferometer, which modulates all the frequencies, also eliminates problems due to stray light contribution. The added advantage of the simple design of the interferometer results in less wear and reliability.

The larger beam area of the FTIR spectrophotometer, rather than the slits used in a dispersive instrument, facilitates high throughput of light, which in turn enables the use of the energy limited ATR sampling technique. In addition, FTIR instrument does not need external wavelength calibration as it has an internal helium neon (HeNe) laser reference. This laser automatically provides a calibration with an accuracy of 0.01 cm^{-1} [40].

3.4 Instrumentation

3.4.1 Fourier-transform spectrometers

The major components of a FTIR spectrophotometer are a radiation source, an interferometer, a sample compartment, and a detector, as shown in Figure 3.1 [40]. However, the instrument is controlled by a personal computer, which handles the digital signal processing in the system. The radiation source in mid-IR spectrometers is a ceramic rod or wire heated to about 1300 K. three popular types of sources used in the mid IR are the Nernst glower (constructed of rare-earth oxides), the Globar (constructed of silicon carbide), and the Nichrome coil. Infrared light emitted from a source is directed into an interferometer [40]. The interferometer is the most important component of commercial FTIR spectrometers, as it increases the efficiency of the technique by modulating the infrared (IR) radiation. Thereafter the light passes through the sample and is then focused onto the detector, where the signal is measured as an interferogram.

The most commonly used interferometer is a Michelson interferometer. It consists of two mirrors, moving and fixed, which are perpendicular to each other, and a beam splitter (Figure 3.2) [35]. As the name implies the beam splitter divides a collimated beam of radiation into two beams of equal intensities; half the IR beam is transmitted to the fixed mirror and the remaining half is reflected to the moving mirror. The interference between these two beams is created when they recombine at the beam splitter after striking the mirrors. This interference could either be constructive or destructive depending on the variation in optical path difference caused by the moving mirror. The resulting beam will pass through the sample and is eventually focused on the detector.

The intensity of radiation reaching the detector varies in a sinusoidal manner as the mirror moves at a constant velocity [40]. The interferogram, which is the record of the interference signal, is a time domain signal, displaying intensity versus time within the mirror scan. The absorption of the sample at a particular frequency reduces the amplitude of the sinusoidal wave by an amount proportional to the amount of sample in the beam. Thus the interferogram is the continuous sum of all the interference patterns produced by each wavelength. As a result the interferogram contains information over the entire IR region to which the detector is responsive.

A mathematical operation known as Fourier transformation is applied to convert the time domain interferogram into the frequency domain spectrum, showing the intensity at different wavenumbers. Generally mid-IR spectrometers have KBr optics with a variety of sampling accessories such as attenuated total reflection (ATR), diffuse reflectance, and transmission. However, in this instrumentation section special emphasis is given to the ATR technique as it is generally exploited in protein solution characterisation studies. In most commercial spectrometers, the path difference is scanned continuously, however step-scan instruments are used in research applications where the path difference is held constant [39]. Step-scanning instruments have much wider use in addressing dynamic measurements than continuously scanning systems. Commercial spectrometers that are used for routine analysis have much of the optical path sealed and desiccated in order to reduce the problems of fluctuating levels of water and carbon dioxide.

In a dispersive spectrometer, the sample is placed between the source and the monochromator, however this arrangement is different to the FTIR where the sample is placed between the interferometer and the detector. Unlike dispersive spectrometers, which operate in a double beam mode, FTIR spectrophotometers operate in a single beam mode. As a result, FTIR does not obtain transmittance or absorbance IR spectra in real time [40]. Instead the transmittance or absorbance spectrum of the sample is calculated as the ratio of the single-beam sample spectrum against the single-beam background spectrum.

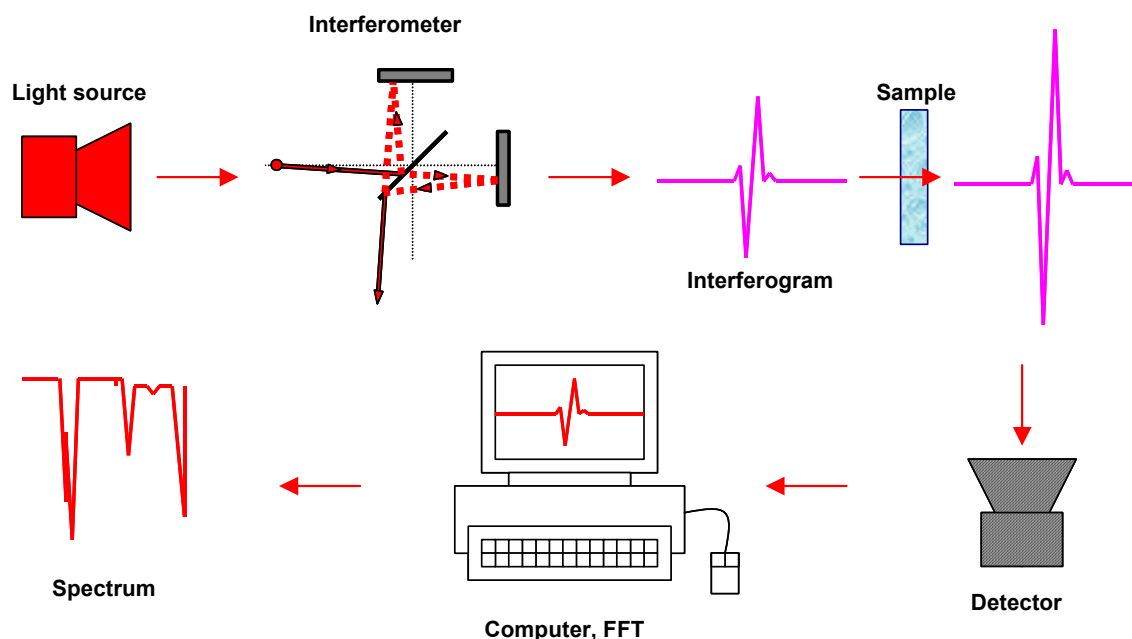


Figure 3.1 *Schematic illustration of FTIR system*

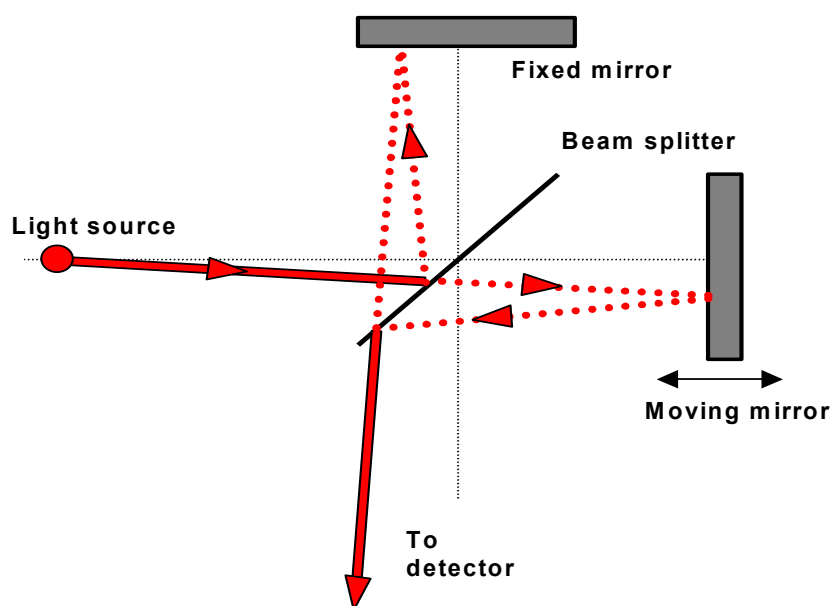


Figure 3.2 *The basic optics of a Michelson interferometer.*

FTIR spectroscopy is extensively used in pharmaceutical industry as it has the capacity to provide information at the chemical molecular level on virtually any material of interest, solid, liquid or gas, within a very short time (1-10 mins), depending on the type of instrument and the resolution required [40, 43]. FTIR is also a well-understood technique that is widely accepted and validated by a long history of use. In addition, FTIR is easy to use and results can be obtained quickly and cheaply. Therefore FTIR has the potential to be used in QC environments in the biopharmaceutical industry.

3.4.2 Detectors

Deuterated triglycine sulfate (DTGS), which is a thermal detector, is the most common detector used for routine measurements [39]. This detector produces a signal in response to the change in temperature caused by the absorption of the IR radiation, and operates at room temperature. The liquid nitrogen-cooled MCT (mercury cadmium telluride) detector, which is known as a photon (or quantum) detector, is used when higher speed or sensitivity is needed. This photon detector generates signal when electrons are excited directly by the absorption of radiation. As the name implies, the detector is cooled to liquid nitrogen temperature in order to avoid excitation of electrons by thermal motion.

DTGS detectors show linearity over a wide range of signal levels in contrast to MCT detectors, which restricts the use of MCT detectors to low signal levels, typically less than 10% transmission. The non-linearity of the MCT detector can be detected by the non-zero signal present in the region beyond the long-wavelength limit of the detector. Room temperature detectors are therefore preferred for quantitative measurements. Unlike the DTGS detector, the MCT has a long wavelength limit, $750\text{--}450\text{ cm}^{-1}$, depending on the band type of the detector, as a minimum energy is required to excite an electron [40].

3.4.3 Sampling techniques

One of the chief difficulties in obtaining FTIR spectra of native proteins is the high absorption in the region of interest by water. This means that to get sufficient transmission of light through a sample to measure a spectrum, very short path lengths – typically of the order of $6\text{ }\mu\text{m}$ – are required. This can be achieved either by using a special transmission cell with an extremely short path length, or by exploiting the phenomenon of total internal reflection, in a technique known as attenuated total

reflectance (ATR). Each technique has its particular benefits and disadvantages, which are described in the following sections.

3.4.3.1 Transmission measurements

In transmission measurements, aqueous solutions are measured as thin films between IR-transmitting windows. Calcium fluoride (CaF_2) windows are generally used for protein characterisation studies in the transmission mode, however, the choice of window material depends on solubility, chemical compatibility, and the spectral range needed [44].

Three different types of cell assembly can be used. The simplest case is a thin film sandwiched directly between the two windows. However, in this case there is no control over path length. More commonly, spacers are used between the windows to give a defined path length, either in demountable or sealed cells. Demountable cells are useful where there is a problem with cleaning the cell after use; however, the path length can vary somewhat on each re-assembly, leading to reproducibility problems. Fixed cells, in contrast, give precisely defined path lengths but are harder to clean.

Water is an extremely strong absorber that leads to signal saturation and which in turn causes non-linear detector response [45]. Therefore very short path lengths of a few μm are generally used. It is not trivial to fill such cells, as the formation of air bubbles can cause light scattering. It has also been observed that the perturbation of surface water molecules by the window material may become important and may alter the spectrum of the bulk sample as the path length is reduced. For protein work, flow cells are available with integral temperature control.

3.4.3.2 Attenuated total reflectance (ATR)

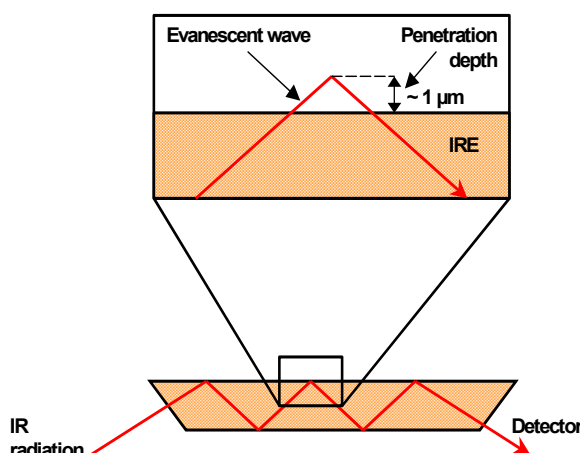


Figure 3.3 Schematic of a typical ATR accessory for liquids.

The ATR technique was developed independently by Fahrenfort and Harrick in the 1960's [35]. It provides an alternative means of obtaining a very short effective path length, by exploiting total internal reflection (For a detailed technical explanation, see the Appendix in section 8). A schematic of a typical ATR accessory is shown in Figure 3.3. Infrared radiation passes through a crystal where it undergoes multiple reflections. Liquid is introduced to the top surface, where absorption can occur in a very thin layer near the surface of the crystal.

In recent years, ATR accessories have been developed specifically for protein structure applications by introducing modifications to the design of the sample platform including temperature control and low sample volume. As water is an extremely strong IR absorber, very small path lengths, of the order of $6\ \mu\text{m}$, are required in the transmission mode to avoid signal saturation and therefore problems with reproducibility [45]. In addition, ATR is more favourable for quality control applications as it requires little or no sample preparation. Although ATR spectra can be obtained using dispersive instruments, FTIR spectrometers permit high signal-to-noise ATR spectra [40].

The signal intensity of an ATR spectrum is dependent on the penetration depth of the radiation and on the number of reflections [39]. Generally, a crystal with a high refractive index is employed to favour total internal reflection, however, the higher refractive index of the crystal reduces the penetration depth of the radiation and results in a weaker spectrum. The most commonly used crystals are ZnSe (refractive index, 2.3), diamond (refractive index, 2.3) and germanium (refractive index, 4.0). Also, the number of reflections and the penetration depth decrease with increasing angle of incidence [40]. However, the incidence angle (Θ) should be maintained at an optimum value because the band can be broadened, distorted, and shifted when it get closer to the critical angle and thus will cause serious problem in spectral interpretation [46]. Typical effective path lengths are around 6 μm (see section 8). It is important to note that the signal intensity is entirely independent of the sample thickness, provided that it is much greater than the penetration depth.

Although an IR spectrum generated using an ATR accessory is broadly similar to a spectrum produced by transmission techniques, the intensities of bands and band shapes are different [47], (see also section 8). The intensities of bands, but not the shapes, are readily corrected by software, facilitating analysis of ATR spectra and comparison to transmission spectra.

Determining the absolute path length, and therefore the absolute absorbance of a sample, is difficult because it depends on many factors (section 8). In principle the angle of incidence and number of reflections can be measured, but this is difficult to do in practice. This may not be necessary where these parameters are pre-set by the manufacturer, but for precise work it may be necessary to check that the effective path length is correct. This can be done by using a standard material of known absorbance [48].

3.5 Presentation of Spectra

Positions of IR absorption bands are generally presented as either wavenumbers ($\bar{\nu}$) or wavelength (λ) [40]. The wavenumber is the number of waves per unit length; therefore it is directly proportional to frequency (and moreover, the energy) of the IR absorption. The relationship between wavenumber and frequency can be expressed as:

$$\bar{\nu} = \frac{1}{\lambda}$$

Equation 1

Where wavelength is in m and wavenumber is in m^{-1} . However, the normal convention is to express wavenumber in cm^{-1} , whereas wavelengths in the infrared are typically expressed in μm . Therefore in this case an additional conversion factor of 10^4 must be used:

$$\bar{\nu} = \frac{10^4}{\lambda}$$

Equation 2

The IR absorption spectrum is generally presented with wavelength or wavenumber as the x-axis and absorption intensity or percent transmittance as the y-axis. Transmittance is the ratio of the intensity of light transmitted by the sample (I) to the intensity incident on the sample (I_0). Absorbance is the logarithm to the base 10 of the reciprocal of the transmittance (T).

$$A = \log_{10} \left(\frac{1}{T} \right) = -\log_{10} T = -\log_{10} \frac{I}{I_0}$$

Equation 3

The intensity differences between strong and weak bands are better represented by the transmittance spectra as transmittance ranges from 0 – 100% T whereas absorbance ranges from infinity to zero.

However, absorbance is generally used in quantitative analysis because a linear relationship exists between the sample concentration and absorbance, (known as the Beer-Lambert law). Having said that, deviations from Beer's law are common in infrared spectroscopy due to instrumental and sample effects. The instrumental effects include insufficient resolution due to the optical path difference between two beams in the FTIR, and sample effects include molecular interactions such as hydrogen bonding and aggregation due to high sample concentration.

3.6 Protein Secondary Structure Determination Using FTIR

The amide bonds of proteins show nine bands in the IR named amide A, B, I, II...VII, however, bands I and II are generally employed to study protein structure. The amide I band which is predominantly due to C=O stretch (80%) occurs around 1650 cm^{-1} ; II which is a mixture of N-H bend (60%) and C-N stretch (40%) occurs near 1550 cm^{-1} ; and III which is also a mixture of C-N stretch (40%) and N-H bend (30%) arises near 1300 cm^{-1} . The hydrogen bonding pattern between amide C=O and N-H groups, which is characteristic of the protein secondary structure, influences the exact frequency of the amide I and II absorptions [49]. It is therefore anticipated that each type of secondary structure will give rise to characteristic amide I and II absorptions and facilitate the elucidation of protein secondary structure by IR spectroscopy [50].

The amide A, $\sim 3500\text{ cm}^{-1}$, and amide B, $\sim 3100\text{ cm}^{-1}$, bands originate from a Fermi resonance between the first overtone of amide II and the N-H stretching vibration [51]. As stated, amide I and II are the most useful bands for protein structure studies by infrared spectroscopy whereas amide III and other bands are of limited use due to their complex nature. The complexity of amide III and IV arises from a mixture of several coordinate displacements and amide V, VI, and VII from the out-of-plane motions.

Amide A, which is mainly due to the N-H stretching vibration does not depend on the backbone conformation; however, the sensitivity to the strength of a hydrogen bond enables investigations of hydrogen bonding in biological macromolecules [51, 52]. The amide I band is the most intense absorption band in proteins and is mainly related to the backbone conformation. It is, however, independent of the amino acid sequence, its hydrophilic or hydrophobic properties, and charge [51]. The amide II, which is more complex than the amide I, is also sensitive to the conformation of the protein. However, the amide II band has not been as well studied as the amide I and therefore the frequency of the amide II band is poorly correlated with different protein secondary structures. However, the amide II band, which arises principally from an N-H stretching vibration, is useful for studying hydrogen-deuterium exchange of the peptide groups [53].

In the amide I region, $1600\text{--}1700\text{ cm}^{-1}$, β -sheet structure is generally attributed to be in the spectral range of $1620\text{--}1640\text{ cm}^{-1}$. However, ' β -bands' have also been observed below 1620 cm^{-1} for some proteins, due to variation in hydrogen-bonding strength as well as differences in transition dipole coupling in different β -strands [42]. Surprisingly, proteins with little or no β -sheet structure but with short extended segments that do not form β -sheet structures also give bands in this region. This region is further complicated by the possibility of β -turn absorption. The weak component in the $1670\text{--}1695\text{ cm}^{-1}$ is identified as due to anti-parallel β -sheet structure. However, the assignment of this peak is generally impossible due to the overlap of absorption from β -turn and unordered structures. The α -helical conformation gives rise to absorption in the range $1650\text{--}1658\text{ cm}^{-1}$. However, random coil structures also absorb in the same region. In order to differentiate α -helical absorption from random coil, D_2O is used, as it will shift the random coil absorption to lower frequency, 1644 cm^{-1} . It was highlighted by Haris and Severcan [42] that absorption of α -helical structures are sensitive to the environment, for example membrane proteins show α -helical absorption in the range $1656\text{--}1658\text{ cm}^{-1}$, and water soluble proteins absorb near $1650\text{--}1655\text{ cm}^{-1}$. Furthermore, the band frequency for an α -helical structure in a short, solvent exposed peptide is significantly different from the absorption of helical structure buried within a solvent inaccessible region of a highly globular protein. The

frequencies given above are not definitive and some contradictions are reported in the literature [51, 54, 55]. It is therefore obvious that assigning bands in the spectra of proteins should be carried with great care, otherwise it can lead to erroneous conclusions. Also, great care should be taken in assigning bands in the amide I region, as trifluoroacetic acid (TFA), which may be a contaminant in protein samples, gives an intense band at 1673 cm^{-1} . Natalello *et al.* claim in a recent publication that the bands in the $1200\text{--}900\text{ cm}^{-1}$ region are due to the absorption of the glucidic moieties present in the recombinant proteins [56]. This suggests that FTIR may also be useful for analysis of protein glycosylation; see also section 4.3.5.

The contribution from the amino acid residues in the $1800\text{--}1400\text{ cm}^{-1}$ region (amide I and II) was thoroughly investigated by Venyaminov and Kalnin [57]. Among the twenty amino acids, it has been established that arginine, asparagine, glutamine, aspartic and glutamic acids, lysine, tyrosine, histidine, and phenylalanine have strong absorbance in this spectral region. It is therefore essential to recognize the contribution from these amino acids before attempting to extract information from the shapes of amide I and II bands.

3.7 Water interference and subtraction

For aqueous solutions, the region around 1650 cm^{-1} , where the conformation sensitive amide I band occurs, is prone to serious errors due to the strong absorbance of water; [58] see Figure 4.8 for an example. In order to overcome this water absorbance, differential techniques with carefully matched cells were used with dispersive instruments and the solvent spectrum was subtracted from the solution spectrum using the digitised FTIR instruments [59]. However, the water interference cannot be completely eliminated because of the strong hydrogen bonding between water and protein molecules in the solution spectrum. This strong interaction often changes the frequency, width, and height of the water bands compared to water bands observed in the absence of protein and the residual features of the water spectrum that remains after subtraction in turn distorts the amide bands. Therefore, neither the older differential technique nor the computerized FTIR subtraction procedure can completely remove the residual water interference without distorting the amide bands. The ambiguity in the frequency and intensity of the amide I and II bands after correction for water contribution indicates the difficulty in conformational studies of proteins in water. The use of D_2O can eliminate problems with aqueous protein solutions because it does not absorb in this region. However, this approach is clearly not suitable for a QC environment, since the sample must be modified to make this measurement.

In order to properly correct the water absorbance in IR spectrum of aqueous solutions, it is important to select a spectral range where the water absorbance does not overlap with that of the protein and has intensity comparable with that in the region of interest. Therefore the removal of water absorption from the protein solution spectrum at 1650 cm^{-1} is generally attempted using the water association band at 2125 cm^{-1} as an internal intensity standard, by employing a scaling factor to the pure water spectrum. The scaling factor which controls the contribution of water is typically varied by an interactive procedure until a flat baseline between $1710\text{--}2000\text{ cm}^{-1}$ is judged satisfactorily [53]. The uncertainty and bias in the spectral subtraction of water from protein solutions via interactive graphic programs can cause significant differences in the spectra of the same protein and thus affects the comparability, since the determination of the baseline is done visually [60]. However, Haris and Chapman claims that spectra of proteins from different laboratories by different workers produced similar results with the above interactive method [53].

Computer-aided automatic water subtraction algorithms, are being extensively employed with the intention to remove individual bias and uncertainty and thus improves the comparability of results between users [60-62]. In addition these methods also take into account small baseline variations due to instrument drift. Most of these procedures assume that the water absorbance is same at different wavenumbers and use 2125 cm^{-1} as an internal intensity standard. However, Hübner *et al.* have indicated that there is no linear correlation between the error in the area at 1650 cm^{-1} and at 2125 cm^{-1}

as the extinction coefficient of water is different at these wavenumbers [60]. The error indicated is mainly due to phase corrections where a shift of the apodization, smoothing of the discontinuities at the beginning and at the end of the interferogram, is observed against the phase-corrected interferogram and this error seems to be additive in the single channel spectrum.

The studies on water molar absorptivities in the 1000–4000 cm^{-1} region confirms that the molar absorption at 1650 cm^{-1} is six times greater than at 2125 cm^{-1} and therefore the residual absorbance at $\sim 3645 \text{ cm}^{-1}$ can serve as a measure for precise water subtraction at 1650 cm^{-1} [63]. Therefore a second area for subtraction was considered, 4000–3650 cm^{-1} [60]. However, Arrondo *et al.* have stated that the suitable window for studying proteins in water solutions is between 3000–1000 cm^{-1} as above and below these limits water absorbance is far too strong to allow subtraction [64]. The independence of the amide I band intensity from the structural types has lead to an alternative route to analyse the complete subtraction of water by implementing a constant ratio ($V=1.29$) of the integrated intensities of the amide I and II bands along with obtaining flat baselines in the regions of 4000–3650 cm^{-1} and 2300–1800 cm^{-1} [60].

Furthermore factors such as phase error, variation in cell thickness between measurements (in transmission), dispersion effects due to reflections (in transmission), temperature fluctuations, influence of salt, and detector nonlinearity can interfere with water absorbance and/or can change spectral properties [60, 65]. This therefore indicates that complete water subtraction must be taken into account through use of appropriate controls for proper water absorbance correction.

3.8 Quantitative information from FTIR

The intrinsic shape of the amide I band is characteristic of the secondary structure of proteins. The amide I band is generally a composite, consisting of overlapping components representing α -helix, β -sheet, turns, and unordered structures [59]. Having said that, these bands cannot be resolved by conventional spectroscopic techniques because separation between these bands is less than the width of the individual bands. The pioneering work by Susi and Byler evolved the first structural determination analysis of proteins from FTIR spectra through the application of second derivative and Fourier self-deconvolution techniques [59]. However, the spectrum utilized should be of high signal-to-noise ratio because second derivative and Fourier self-deconvolution techniques enhance noise, interference fringes, impurity bands, and water vapour bands, along with the true bands of the sample. In addition, difference spectroscopy which is another powerful technique is employed in the study of small structural changes in proteins [53].

Fourier self-deconvolution requires two constants, half-width at half height of the unresolved bands (σ) and the resolution efficiency factor (K) as computer input [59]. Improper selection of σ and /or K results in serious errors and artefacts and thus leads to incorrect interpretation of the spectrum. The information about the side chains as well as the secondary structure of proteins can be obtained from the second derivative spectra; however, it does not provide quantitative information due to the complex patterns created by overlapping peaks and side lobes where as self-deconvolution offers this possibility.

Quantitative information on fraction of different secondary structural elements can be obtained by identifying the number of amide I component bands and their approximate positions from deconvoluted spectra, and then curve fitting the overall amide I band contours [53]. However, this method is not without limitations, as assignment of all bands under amide I component is impossible with complete certainty. In addition it is assumed that the molar absorptivities of the bands associated with different secondary structural elements are the same. Although this method gives reliable results, it is to some extent subjective, for example in the initial choice of input parameters.

Alternatively, pattern recognition techniques (chemometric methods), which are effective in systems with strongly overlapping bands, have proved to be effective in the quantitative analysis of protein

spectra [4]. These methods involve the use of a model comprising infrared spectra of proteins of known structure and the use of mathematical tools such as factor analysis and partial least squares analysis [42, 53]. Although this avoids deconvolution and assignment of the bands to different structures, the best results can only be obtained if amide I bands of normalized infrared spectra were used in the construction of the model.

Protein structure from IR spectra is often determined using the curve-fitting method although statistical analyses are generally accepted as the best way to analyse protein CD spectra [4]. However both methods (curve-fitting and pattern recognition approach) experience a common problem such as the influence of the absorbance of amino acid side chains in the amide I region. In addition, statistical treatment of IR data requires pre-processing such as baseline correction and normalization. The normalization procedure assumes equal absorptivity of amide bands irrespective of secondary structure variation. Therefore none of the quantitative methods currently in use can give completely reliable information as they all have their own limitations [42].

3.9 Conclusions

In this chapter we have discussed the advantages and limitations of FTIR in the context of an orthogonal technique to CD for analysing protein structure in solution. The difficulties inherent in compensating for the strong water band at $\sim 1650\text{ cm}^{-1}$ lead to uncertainty in the frequency and shape of the Amide I band, which in turn cause limitations in structure analysis of aqueous protein solutions. Although different approaches for complete subtraction of water have been stated in the literature they all fail to give unambiguous results.

Like all other low-resolution spectroscopic techniques, including CD, FTIR suffers significant limitations in the quantitative assessment of protein secondary structure. However, band narrowing methods can provide a sensitive diagnostic tool, if used cautiously, in monitoring the conformation changes of the protein backbone [38]. Assignment of bands underneath the amide I band should be carried with great care considering the surrounding environment, variation in hydrogen bonding strength, and the differences in transition dipole coupling.

Difficulty in obtaining comparable measurements between transmission and ATR spectra suggests building a universal database like those available for CD, facilitating structural analysis. However care should be taken as the quality of the reference protein database determines the efficiency of the secondary structure prediction, rather than the algorithms used [4]. Moreover, calibration of the ATR cell using a reference solution is very important for comparing spectra.

Considering all the limitations associated with this technique it can be concluded that it is possible to employ FTIR in the analysis of quality control of biopharmaceuticals if the water/buffer subtraction is carried at a high certainty. Hence the potential sources of error in the CD and FTIR spectroscopic methods are different; if at all possible they should be used in conjunction.

4 Investigation of FTIR of proteins

4.1 Introduction

Having established the potential suitability of FTIR as a technique for biopharmaceutical QC, we set out to perform a limited validation study of the technique to identify the issues and problems that arise when attempting to make reproducible measurements of protein structure. The results of this study, and recommendations arising from it, are presented in this chapter.

4.2 Materials and Methods

4.2.1 Materials

Proteins cytochrome *c* (equine heart), lysozyme (chicken egg white), myoglobin (equine heart), IgG (bovine), ribonuclease B (RNase B, bovine pancreas), and ribonuclease A (RNase A, type XII-A, bovine pancreas) were used in this study. All proteins except RNase A (Fluka) were purchased from Sigma-Aldrich and were used without further purification.

4.2.2 FTIR

For these studies, proteins were dissolved in water. Initial analysis on FTIR was conducted using protein solutions with nominal concentrations ranging from 1 to 10 mg/ml. The actual concentrations of protein solutions were calculated using the 280 nm absorption band using a high performance UV-Vis spectrophotometer (Lambda 850, PerkinElmer).

All FTIR spectra were collected using a Tensor-27 series FTIR spectrophotometer from Bruker Optics with a BioATR II unit as the sampling platform with a photovoltaic MCT detector and a Bruker Optics workstation, which was equipped with OPUS software. Aqueous samples of very low volume (15 μ l) were placed in a circular sampling area of radius 2 mm with a path length of 6 μ m. This multi-reflection ATR accessory is based on a dual crystal technology, which has an upper silicon crystal and a hemispherical zinc-selenide (ZnSe) lower crystal that does not come into contact with the sample. The temperature of the sample was maintained at 20°C by means of flow connectors to a circulating water bath. This accessory was purged continuously throughout the experiment with dry nitrogen via telescopic inserts that seals the optical path inside the spectrometer sample compartment. All FTIR spectra were collected with resolution 4 cm^{-1} , scanner velocity 20 kHz, 256 scans, phase resolution 32 and zero filling factor 4.

4.3 Results and Discussion

Initial studies on FTIR were conducted using hen egg white lysozyme at concentrations ranging from 1 to 10 mg/ml in order to determine the possible protein concentration range for spectral measurements and to assess the quality of the spectra obtained with our instrument. Lysozyme was chosen for its ready availability, stability, and well-characterised spectral and structural properties. Concentrations of these solutions were measured by UV absorbance (Figure 4.1), assuming the extinction coefficient of lysozyme at 280 nm is 2.39 $\text{mg}\cdot\text{ml}^{-1}\cdot\text{cm}^{-1}$ [66] (Table 4.1).

In FTIR, a spectrum is generated as the ratio of the single-beam sample spectrum against the single-beam background spectrum and therefore the protein spectrum was obtained by subtracting the water spectrum from the protein solution spectrum where we assumed both the spectra were obtained under identical conditions. Two different subtraction methods, without and with scaling factors, were utilized to analyse the difference between them as the selection of a scaling factor is subjective and may introduce an artefact into the water corrected spectrum. Initially, the FTIR spectra of lysozyme (Figure 4.2) of different concentrations were obtained and thereafter an appropriate water blank spectrum was digitally subtracted from the protein spectra without applying any scaling factors (1:1

subtraction). All spectra were normalised to obtain the minimum absorbance unit, 0, at 1850 cm^{-1} using the offset normalisation available within the OPUS software. This differential method should give a zero baseline throughout the $1700\text{--}1850$ region, however, most of these spectra show baseline distortion where the absorbance is decreasing with decreasing frequencies until the amide I band is reached (Figure 4.2). This unexpected behaviour in the baselines might be due to a small reduction in the effective path length, due to an increase in the refractive index of the solution with increasing protein concentration [67]. This in turn results in a slight reduction in the water absorption in the protein sample compared to the water blank and gives rise to a negative band. We therefore followed the approach of applying a scaling factor to the blank spectrum to minimize the difference between the protein spectrum and the blank spectrum in the $1700\text{--}1850\text{ cm}^{-1}$ region and the removal of the water band near 2130 cm^{-1} where there is no absorbance from protein. Subsequently this procedure resulted in corrected lysozyme spectra with flat baselines (Figure 4.3) through the spectral subtraction function of the OPUS software [61].

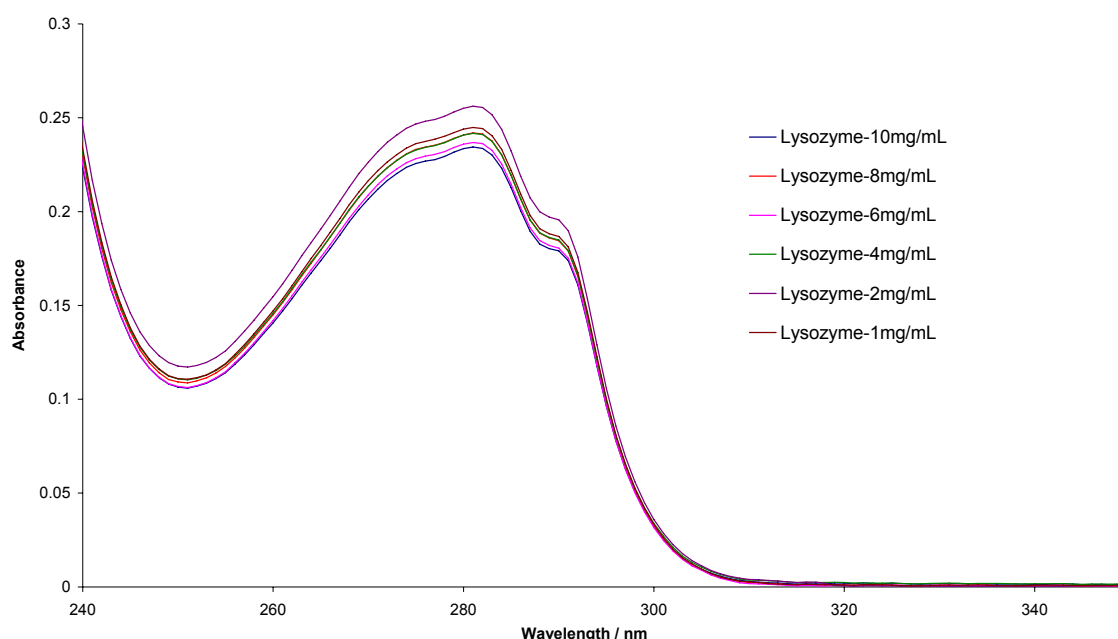


Figure 4.1 UV-Visible absorbance spectra of lysozyme in water.

Lysozyme (nominal concentrations of 1, 2, 4, 6, 8, and 10 mg/ml) was prepared in water and then diluted to 1 mg/ml. Absorbance spectra were collected in a 1 mm path length rectangular cell using a PerkinElmer Lambda 850 series UV-Vis spectrophotometer with scan speed 260 nm/min and data pitch 1 nm.

Nominal concentration (mg/ml)	Measured concentration (mg/ml)
1	1.02
2	2.14
4	4.03
6	5.92
8	8.06
10	9.77

Table 4.1 *Nominal concentration of the lysozyme solutions vs. measured concentrations*
True concentrations of the solutions were calculated using Beer's law from the absorbance at ~ 280 nm and assuming the extinction coefficient of lysozyme to be $2.39 \text{ mg}^{-1} \cdot \text{ml} \cdot \text{cm}^{-1}$. See Figure 4.1.

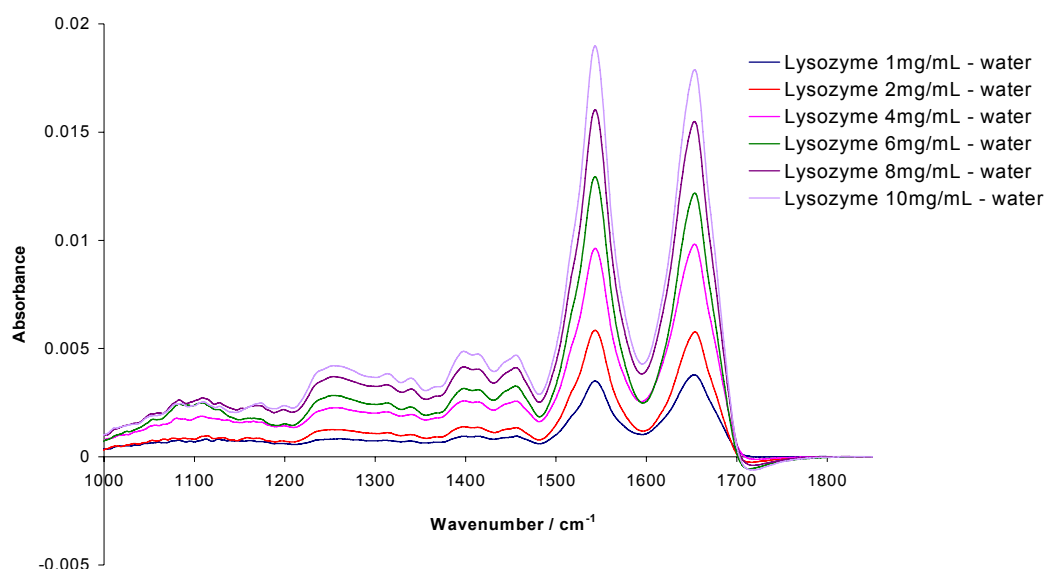


Figure 4.2 *FTIR-ATR spectra of lysozyme in water*

Lysozyme (1, 2, 4, 6, 8, and 10 mg/ml) was prepared in water and FTIR spectra collected on a Tensor-27 series FTIR spectrophotometer from Bruker Optics. Water absorption was corrected by subtracting the appropriate water blank spectrum and normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available with OPUS software.

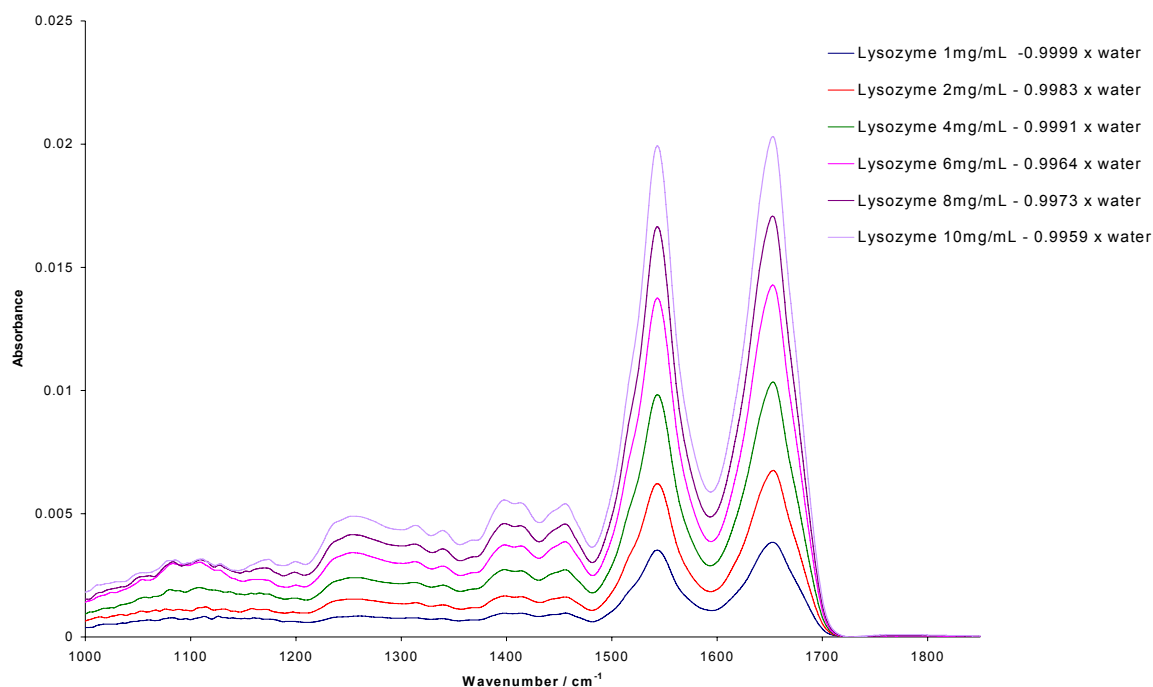


Figure 4.3 FTIR-ATR spectra of lysozyme in water.

Lysozyme was prepared as in Figure 4.2. Water absorption was achieved by subtracting the appropriate water blank spectrum utilizing a weight factor obtaining a flat baseline in the region 1700–1850 cm⁻¹ and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm⁻¹. Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available with OPUS software.

It is possible to acquire spectra from a low concentration protein solution (e.g. 1 mg/ml) with ATR/FTIR, however, the fine structure information in the fingerprint region is only clear with higher concentration protein sample, such as 10 mg/ml. The crucial step in obtaining a protein spectrum is the correct subtraction of water from the protein spectrum. Although the contribution of water can be digitally subtracted, the signal shows strong absorption near 1645 cm⁻¹ which overlaps the amide I region. The blank spectrum therefore must be obtained under exactly the same conditions (temperature, number of scans, resolution etc) as that used for the sample spectra. It is very important for samples prepared in water to obtain sample and blank spectrum at a constant temperature to avoid band shifts [53]. In this study all spectra were obtained at 20°C by means of flow connectors to a circulating water bath. Furthermore the atmospheric water absorption was controlled by purging the ATR accessory continuously with dry nitrogen via telescopic inserts that seals the optical path inside the spectrometer sample compartment. Also the ATR geometry is factory defined and remains perfectly constant. The ‘optical path’ is therefore dependant on the crystal and solution refractive indices, which is constant for a given solution. We would therefore expect the measurements to be highly reproducible. However, in reality we observed reproducibility issues with FTIR spectra of proteins (Figure 4.4).

The lysozyme spectra collected on three different consecutive days, (Figure 4.4), shows variation in the intensity of the absorbance but the shape of the spectra appears to be the same. This suggests that FTIR can still be used as a qualitative tool in the quality control process of the production of biopharmaceuticals. The factors that might be influencing the intensity discrepancy were investigated and given as follows: detector sensitivity, improper water subtraction, adsorption of protein onto the crystal and its kinetics with time, and the cleanliness of the crystal.

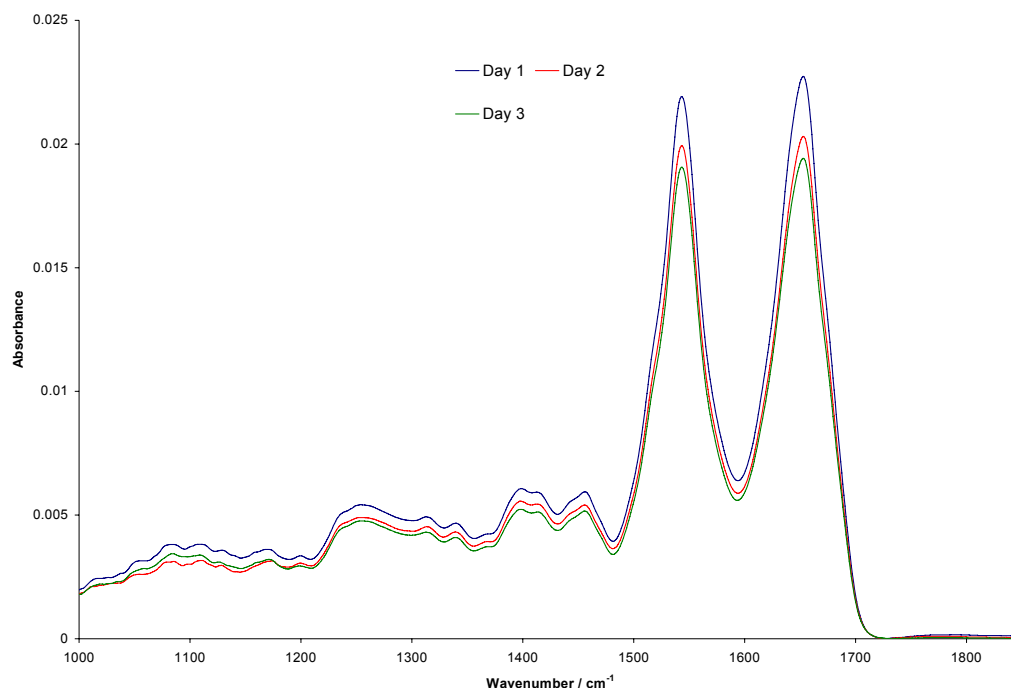


Figure 4.4 *Lysozyme (10 mg/ml) FTIR-ATR spectra collected on three different days. Corrected for atmospheric water absorption and contribution of water by introducing scaling factors.*

4.3.1 Responsivity of the detector

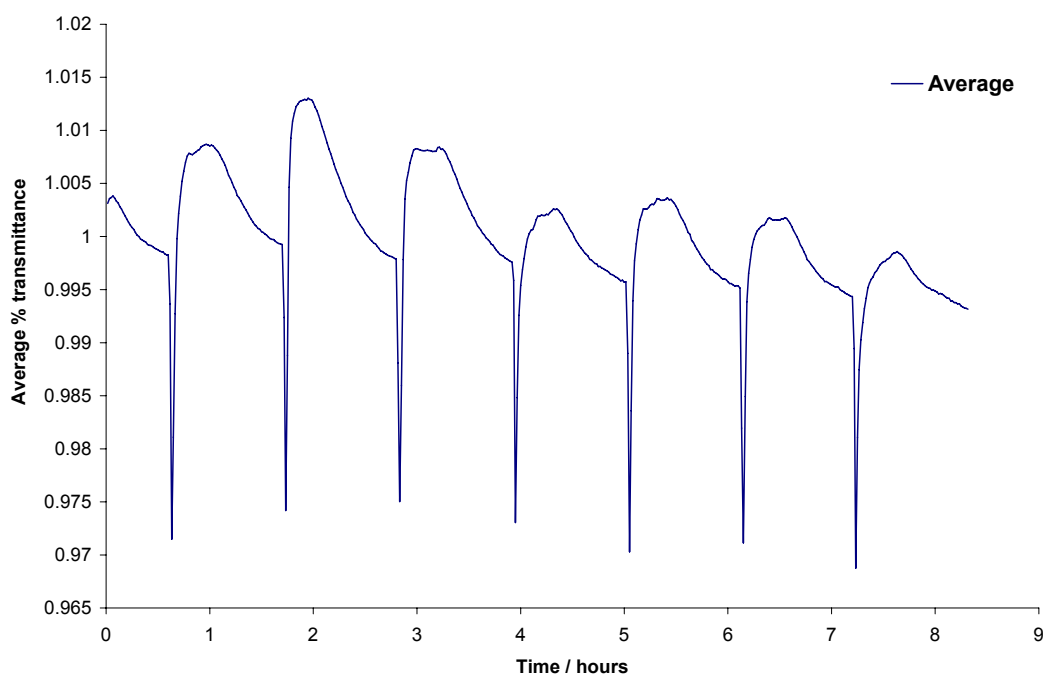


Figure 4.5 *Responsivity variation upon nitrogen cooling of the detector. Background spectra were collected every minute and averaged over the entire wavenumber region. During the experiment, liquid nitrogen was added every hour. The sharp drops in the intensity indicate the addition of liquid nitrogen.*

We found that on our instrument the responsivity of the detector varied with time, in a fashion that appeared to be related to the cooling of the detector.

Background spectra were collected every minute after the liquid nitrogen was added to the detector and averaged over the whole region to produce the spectrum shown in Figure 4.5. This clearly shows that the signal varies with time. This is also true for the amide I region (Figure 4.6). The detector stability was measured over an hour without adding liquid nitrogen (Figure 4.7) and it suggests that the detector responsivity does seem to decrease with time, which may cause uncertainty in measurements.

Our findings about the responsivity of our detector are consistent with those of Theocharous [68]. He suggests that the reason for the drop in spectral responsivity is the formation of thin films of ice on the surface of the infrared detector, which then absorbs wavelengths coinciding with a number of strong ice absorption bands and causes responsivity reduction in those wavelengths over time. The icy films are formed as a result of moisture seeping past the O-ring seals of the Dewar, where the detector is mounted, and depositing on cold windows and on the detector when cooled by liquid nitrogen. Moisture in the form of ice has strong dielectric properties and interacts with the multi-layer dielectric antireflection coatings of the photodetector and causes their transmission to change and thus also reflects on the drift in responsivity with time [68]. Therefore Theocharous recommends evacuating and baking the Dewar at approximately 60 °C to help eliminate the effect temporarily. However, the effect seems to reappear and grow in magnitude. Although the responsivity drift of the cryogenically cooled infrared detectors cannot be eliminated, it can be lowered when using detectors mounted in Dewars with metal O-rings and cemented windows, preventing moisture leakage. Moreover the coolant hold time of a Dewar deteriorates with time and this also can have an effect on the drift of the detector. In this situation it is suggested that the Dewar should be evacuated to bring back the hold time. However, this depends on the age of the detector.

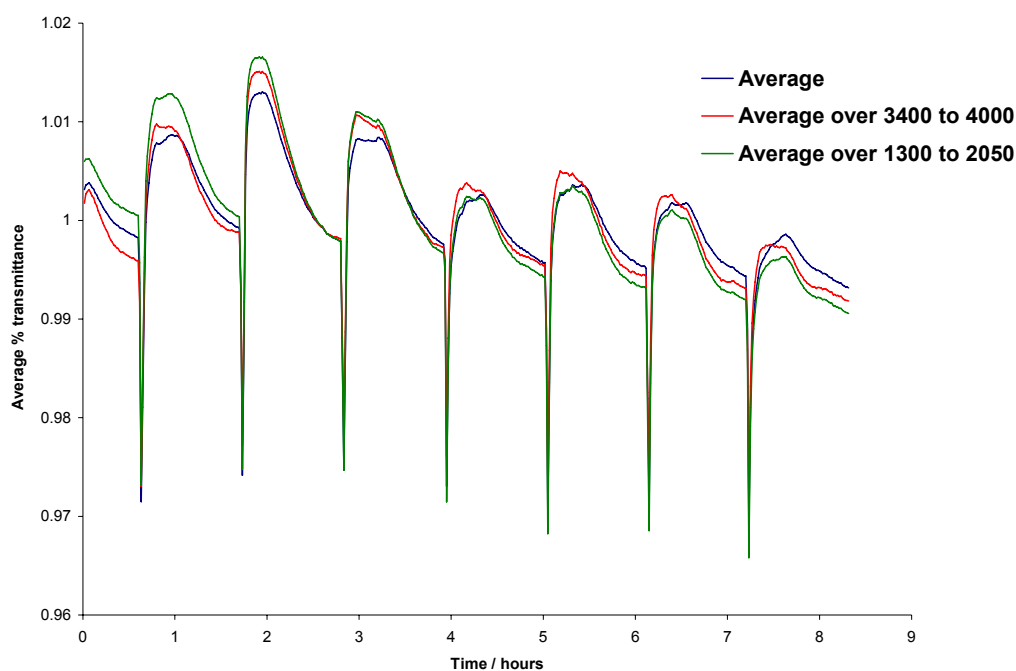


Figure 4.6 *Responsivity variation with nitrogen cooling of the detector.*

Background spectra were collected every minute and averaged over the entire wavenumber region while the detector was cooled approximately every hour by adding liquid nitrogen (blue), averaged over 3400–4000 cm^{-1} (red), and over 1300–2050 cm^{-1} (green). The sharp drop in the intensity indicates the addition of liquid nitrogen.

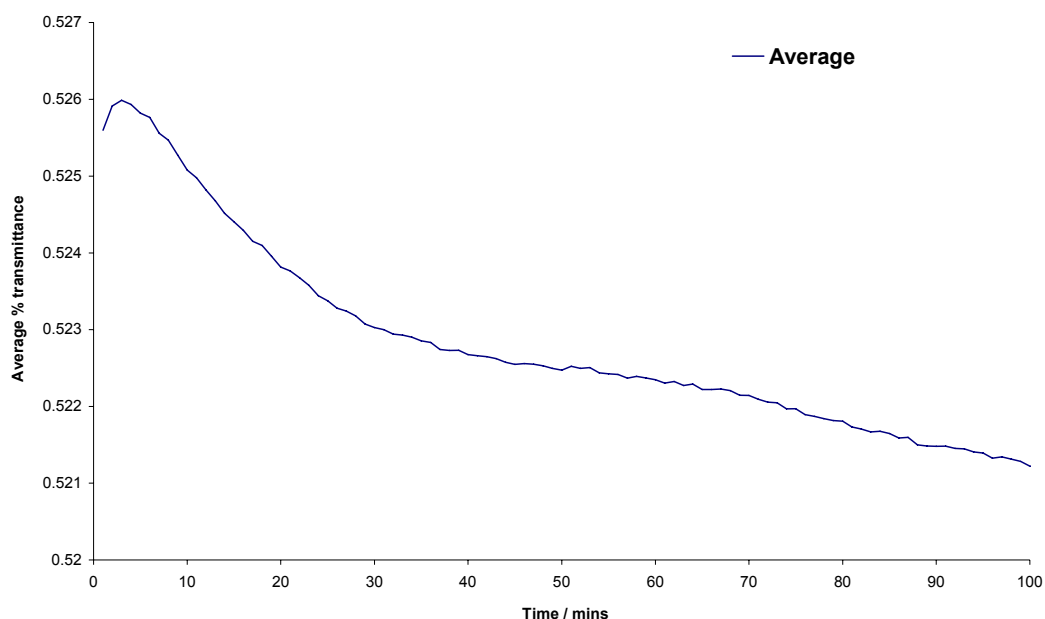


Figure 4.7 **Responsivity variation with nitrogen cooling of the detector**

Background spectra were collected every minute and averaged over the entire wavenumber region to see the variation of average apparent % transmittance with time.

4.3.2 Water subtraction

Water absorbs strongly in the $1550\text{--}1650\text{ cm}^{-1}$ region and obscures the secondary structural information contained in the amide I and II absorption bands of the peptide bonds (centred at ~ 1650 and 1550 cm^{-1} respectively) [67]. The overlaid water and protein spectra (Figure 4.8) show only a very slight difference, as the protein signal is extremely small compared to the water signal due to the very small concentration of peptide bonds compared to the water concentration [67]. This therefore suggests that the water absorption should be subtracted from the protein spectrum with great care. If the signal is incorrectly subtracted by over or under estimation, it can affect the intensity and the shape of the amide I band. As discussed earlier, water subtraction is not a straightforward procedure; therefore we followed the approach of applying a scaling factor to the blank spectrum to minimize the difference between the protein spectrum and the blank spectrum in the $1700\text{--}1850\text{ cm}^{-1}$ region (see section 3.7 above).

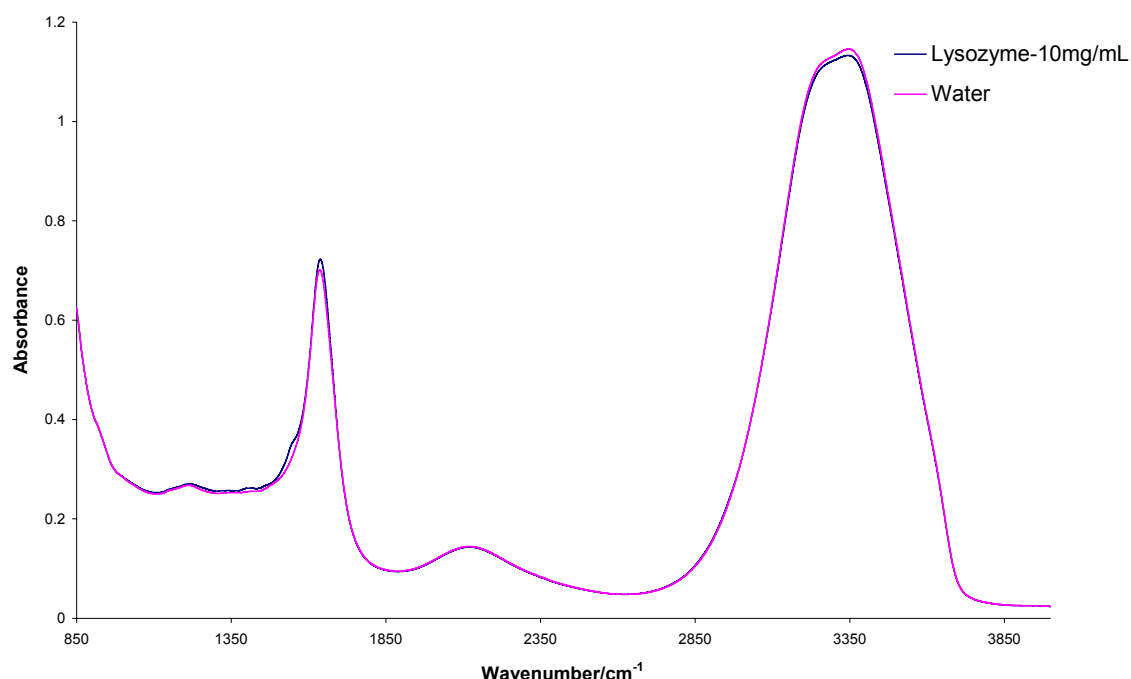


Figure 4.8 Overlaid FTIR-ATR lysozyme (10mg/ml) and water spectra. Both spectra corrected for atmospheric water absorption using the tool available in the OPUS software.

Although Harris and Chapman have highlighted that spectra of proteins obtained by this method in different laboratories by multiple users produce comparable results, we find that this approach can lead to reproducibility problems, as the flat baseline beyond the amide I region is achieved subjectively rather than objectively [53]. A small change in the scaling factor can cause significant variation in the amide I intensity. In Figure 4.9, we have shown how the amide I intensity changes with the variation in the scaling factor when it is varied between 1 and 0.986. The change in the intensity of the amide I band, which is 0.0098, is quite significant when compared to the intensity of the amide I band. The repeat measurements of a range of concentrations of lysozyme on three consecutive days show random variations of the scaling factors (Table 4.2). This therefore implies that the water or buffer subtraction from the protein solution should be carried objectively rather than subjectively in order to avoid reproducibility problems.

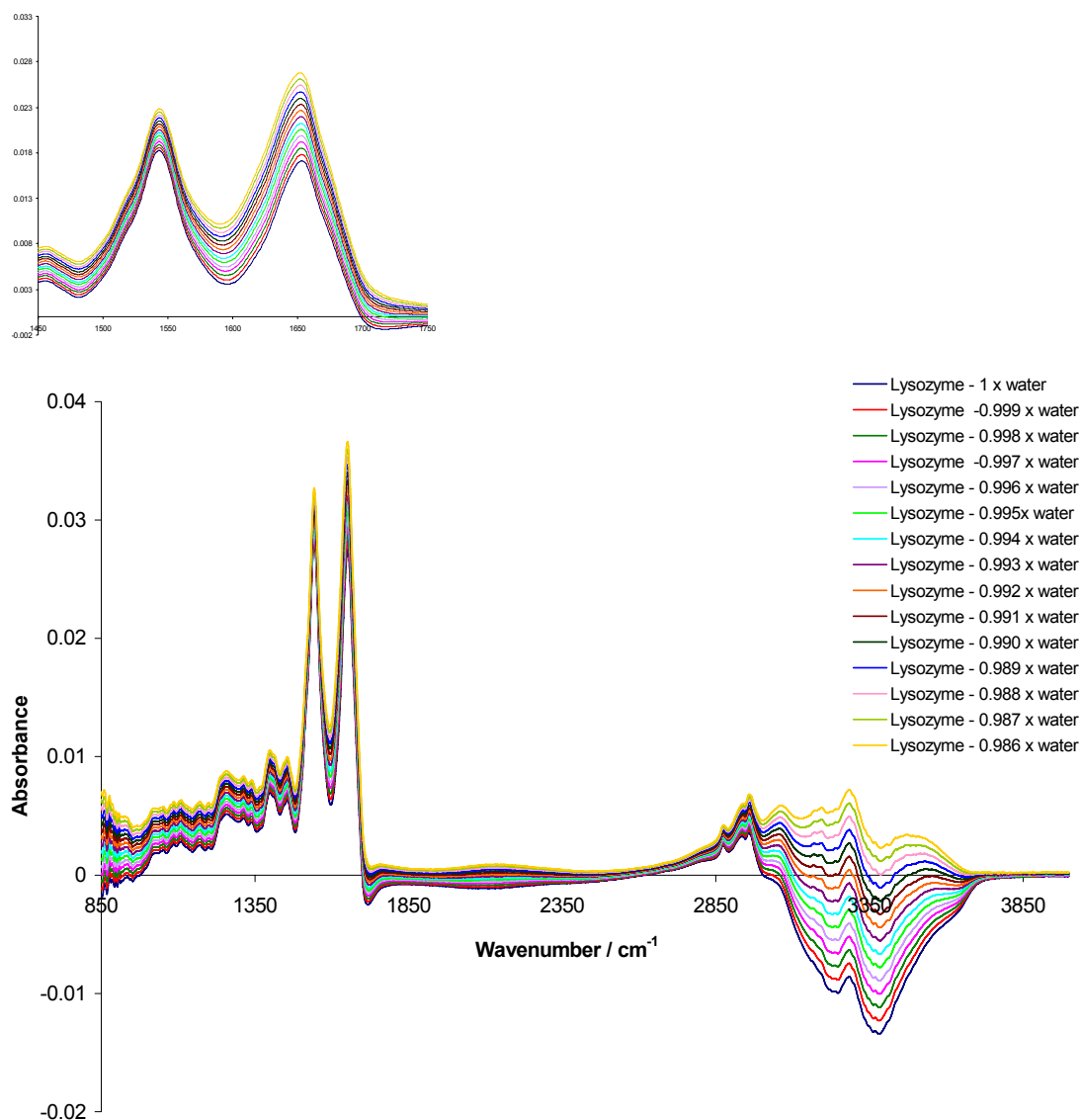


Figure 4.9 *FTIR-ATR protein spectra after water subtraction.*

The water spectrum was subtracted using a range of scaling factors from 1 to 0.986. Insertion is the enlarged amide I and II bands.

In order to reduce the variability in the amide I intensity it is good practice to collect the protein buffer or solvent spectra under exactly the same conditions (temperature, number of scans, resolution etc.) as that used for the sample spectra [53]. Furthermore, the sample compartment needs continuous nitrogen purging in order to minimise the water vapour absorption. In our case the telescopic inserts attached to the ATR accessory seals the light path from getting exposed to the environment and also the light path was continuously purged with dry nitrogen during the measurements. Moreover the variation in the blank and sample spectra due to water vapour absorption was kept minimal by collecting these spectra within a very short interval from the background spectrum. However, the variation between spectra in Figure 4.4 is quite significant.

Concentration of Lysozyme (mg/ml)	Day 1	Day 2	Day 3
1.02	0.9965	0.9999	0.9972
2.14	0.9956	0.9983	0.9995
4.03	0.9991	0.9976	0.9991
5.92	0.9964	0.9971	0.9978
8.06	0.997	0.9973	0.9972
9.77	0.9953	0.9959	0.9956

Table 4.2 *The scaling factors utilized to obtain a flat baseline in the 1700–1850 cm⁻¹ region during water subtraction on three consecutive days.*

4.3.3 Protein adsorption onto the crystal and its kinetics with time

The visual inspection of lysozyme spectra in Figure 4.10 reveals that the intensities of the amide I and II bands are not proportional to the protein concentration. This was also confirmed by plotting the observed absorbance intensity of the amide I and II bands as a function of the protein concentration (Figure 4.11). The linear regression lines for the amide I and II bands do not pass through the origin. This might be due to protein adsorption on the silicon crystal of the ATR accessory, as protein adsorption is a common phenomenon on glass surfaces [67]. This suggests that the evanescent wave would have to cross the adsorbed protein before penetrating the solution, which may have different conformation compared to the protein in solution. This phenomenon can also cause repeatability problems depending on the kinetics of adsorption of the protein. The adsorption behaviour of protein on ATR crystal was investigated by studying the FTIR spectra of lysozyme with time (Figure 4.12).

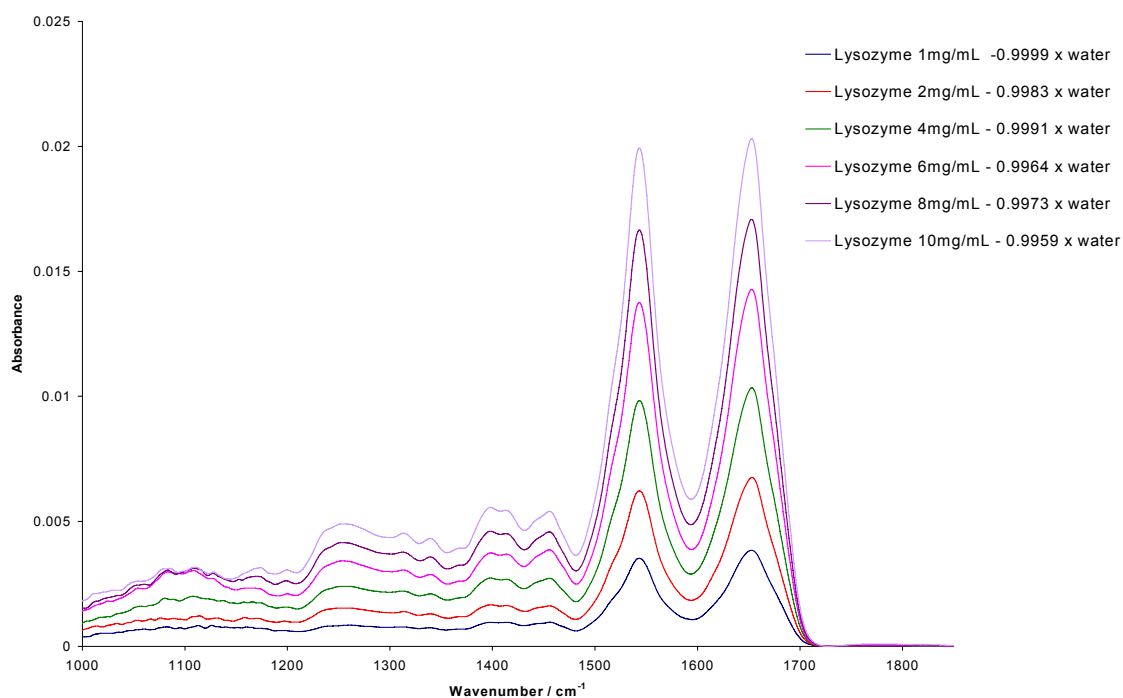


Figure 4.10 FTIR-ATR spectra of lysozyme in water.

Water absorption was corrected by subtracting the appropriate water blank spectrum utilizing a scaling factor which yielded a flat baseline in the region 1700–1850 cm^{-1} and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available with OPUS software.

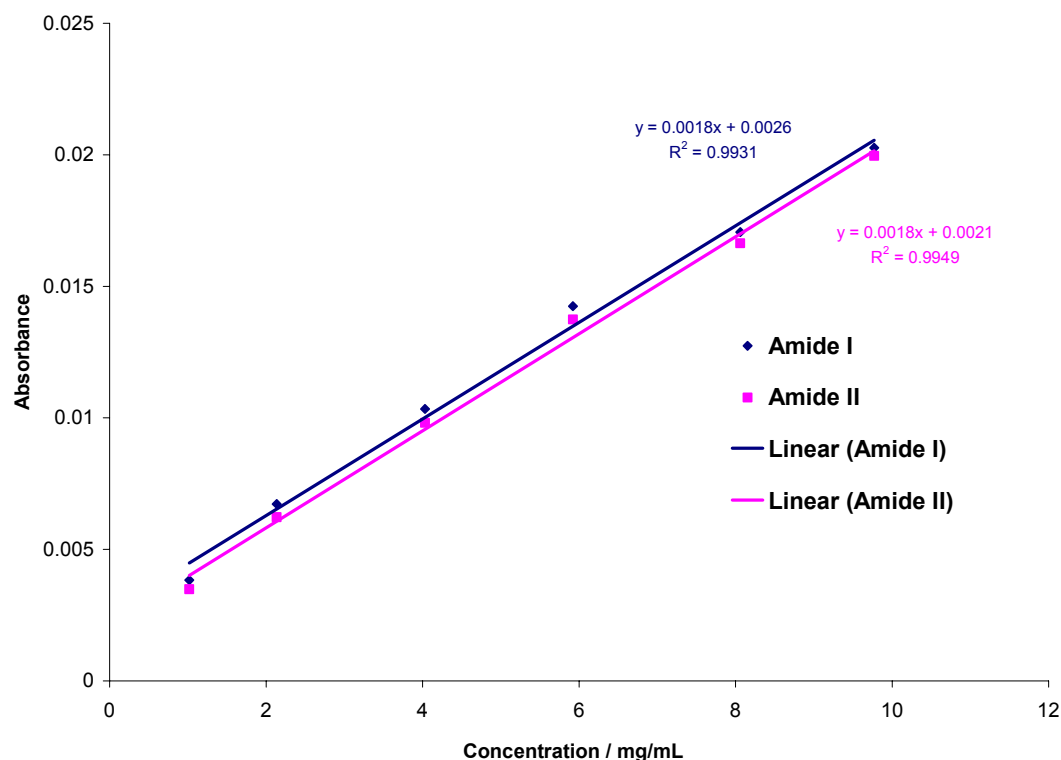


Figure 4.11 Absorbance maxima of the amide I and II bands as a function of lysozyme concentration.

The absorption spectra of lysozyme in water (Figure 4.12) shows increase in absorbance intensity of amide I and II bands with time. This is also confirmed by plotting the amide I absorbance intensity

versus time (Figure 4.13). This suggests that the adsorption phenomenon of protein is dependent on time. Similar experiments were carried out with 10 mg/ml of lysozyme in phosphate buffer (pH 7.2, 30 mM) and similar results were observed (Figure 4.14). The adsorption behaviour of different proteins utilised in this study were also investigated. (Figure 4.15, Figure 4.16).

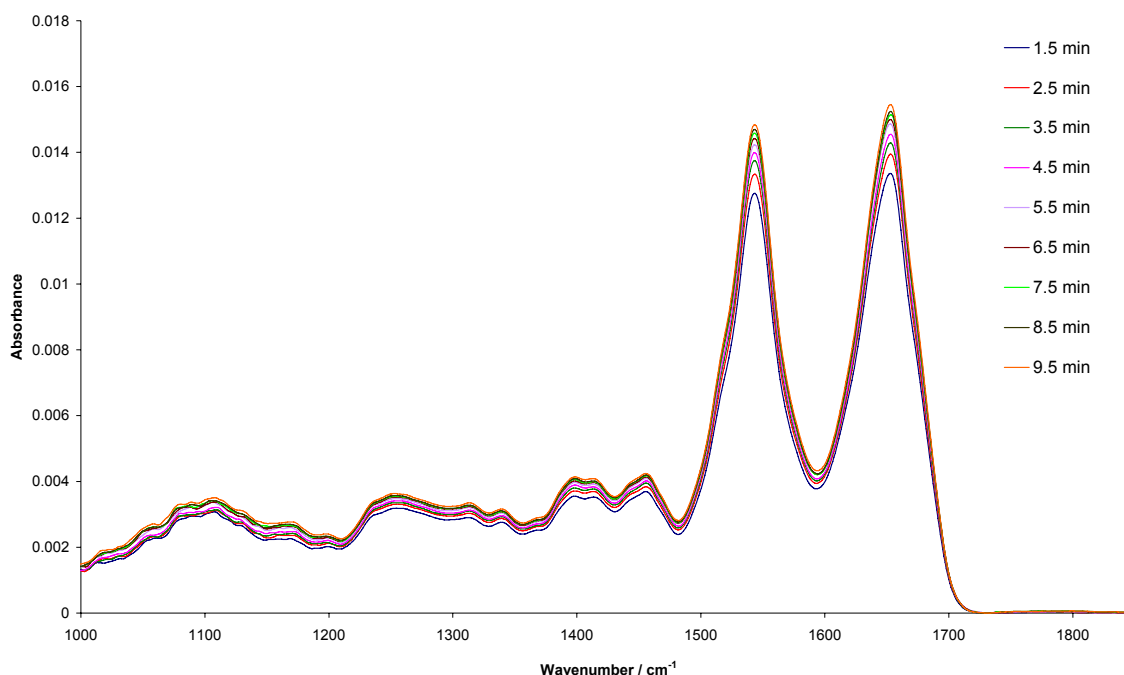


Figure 4.12 FTIR-ATR absorbance spectra of 6 mg/ml lysozyme in water as a function of time
 Water absorption was corrected by subtracting the appropriate water blank spectrum utilizing a scaling factor obtaining a flat baseline in the region 1700–1850 cm^{-1} and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available with OPUS software.

The amount of protein adsorption onto the silicon ATR crystal varies with different proteins and thus might depend on the overall hydrophobicity, charge and size of the protein. This therefore indicates that proteins do not adsorb to the same extent, however, all show adsorption increments with time and this effect will make the correction more complex.

When protein adsorption occurs, there are two components of the overall signal. One is concentration independent, corresponding to the adsorbed material and the second is proportional to the protein concentration, and corresponds to the protein in solution [67]. Goldberg and Chaffotte [67] eliminated the adsorbed protein contribution by subtracting the absorption spectrum at low concentration from that obtained at high concentration, assuming the adsorption for the two concentrations are the same and the protein is saturating after the same time period. However, this is not what we observed, as the adsorption of a highly concentrated protein increases with time and does not reach saturation in a reasonable time for analysis (as seen in Figure 4.13 and Figure 4.16). This observation is also true for low concentrations such as ~ 1 mg/ml lysozyme in buffer (Figure 4.17). In addition, the extent of adsorption for the 1mg/ml concentration is not the same as for the 10 mg/ml lysozyme (Figure 4.14 and Figure 4.17). This therefore implies that the correction is not simple as given in reference [67]; this may be due to the different surface involved – diamond in their case, whereas in our experiments, silicon was used.

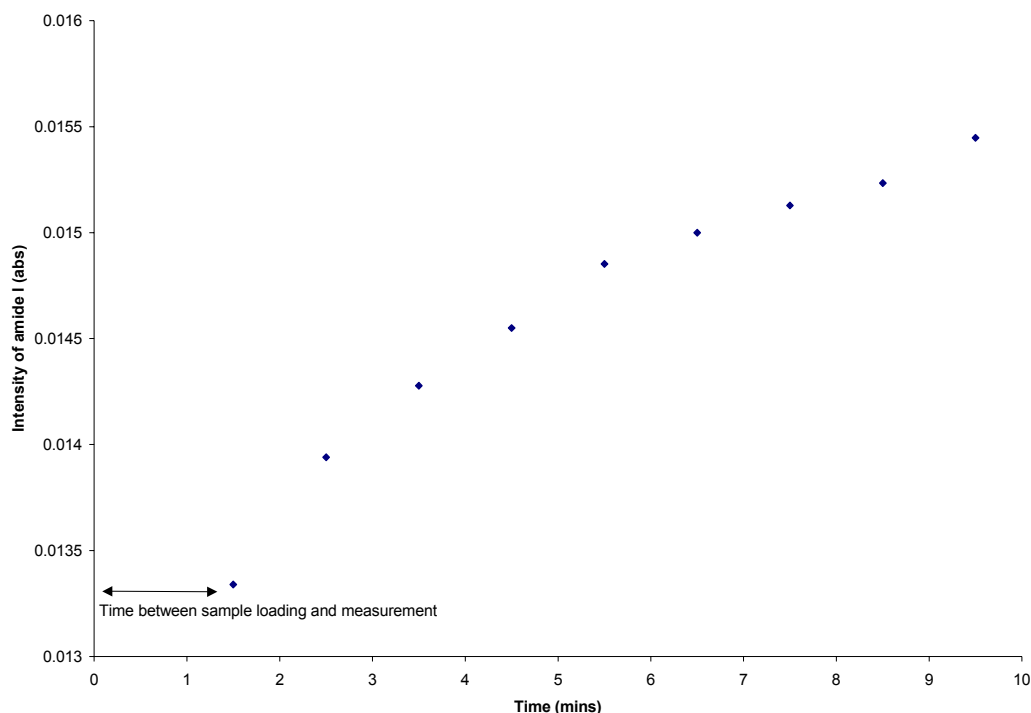


Figure 4.13 Amide I absorbance of lysozyme in water (6mg/ml) with time.

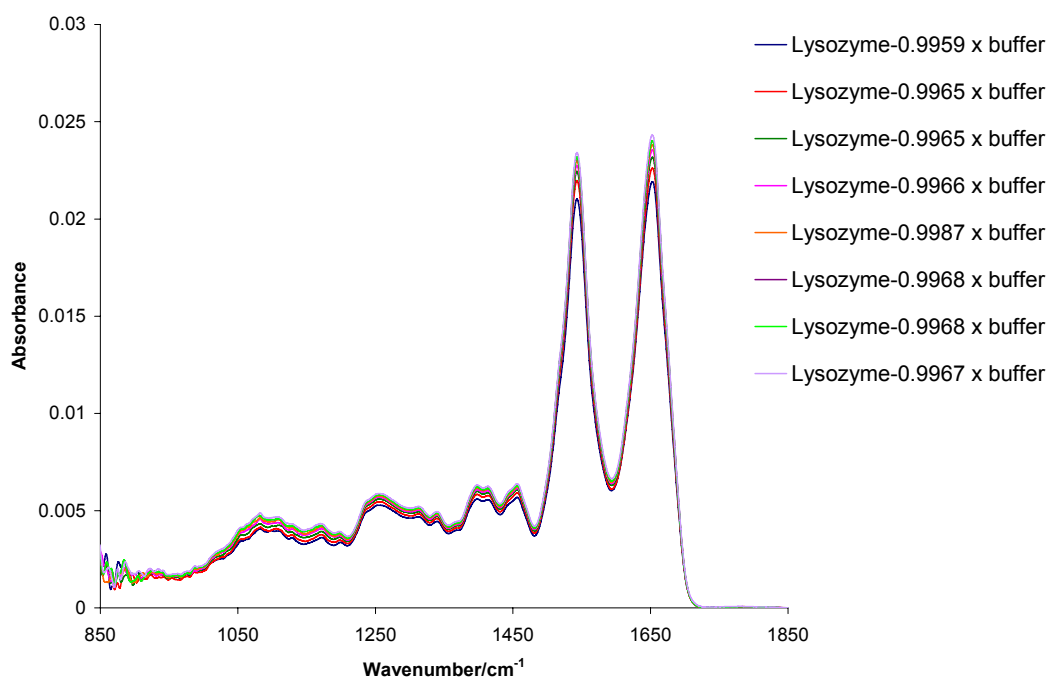


Figure 4.14 Absorbance spectra of 10 mg/ml lysozyme in buffer as a function of time.

Water absorption was corrected by subtracting the appropriate buffer blank spectrum utilizing a scaling factor to obtain a flat baseline in the region 1700–1850 cm^{-1} and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available with OPUS software.

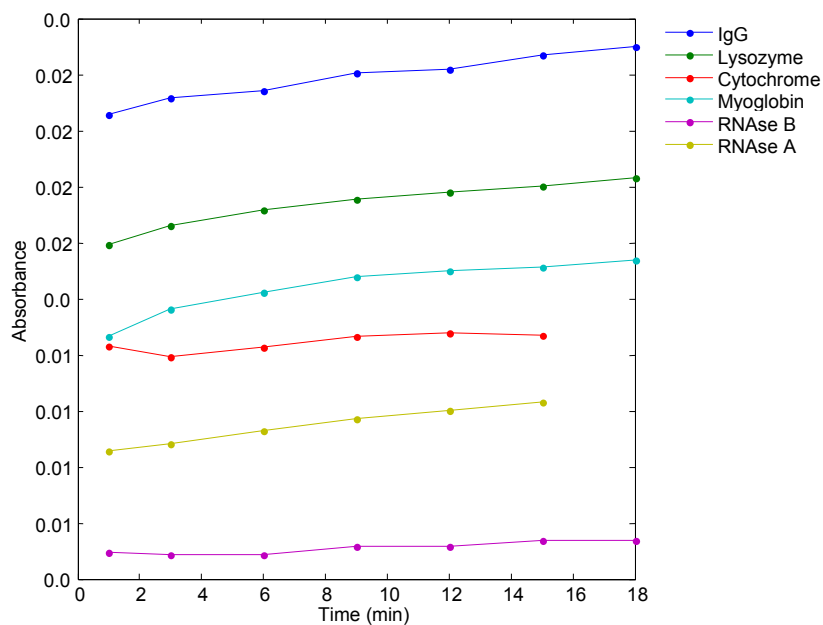


Figure 4.15 Amide I absorbance intensity of proteins in buffer vs. time.

Absorbance intensity of 10 mg/ml IgG (blue solid line), lysozyme (green solid line), cytochrome c (red solid line), myoglobin (turquoise solid line), RNase B (purple solid line), and RNase A (gold solid line) in buffer as a function of time.

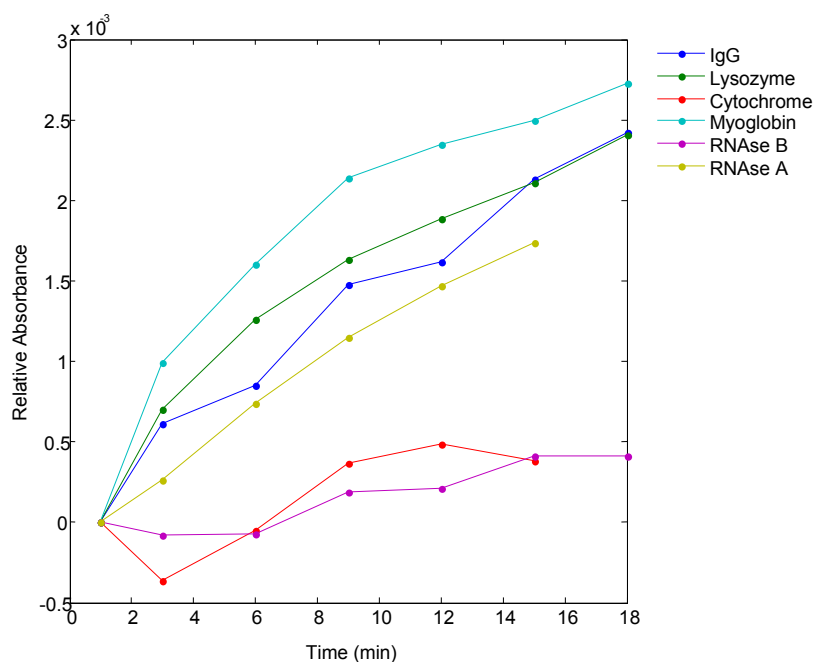


Figure 4.16 Amide I absorbance intensity of proteins in buffer vs. time (mins) after normalising the adsorption at 1 min to zero.

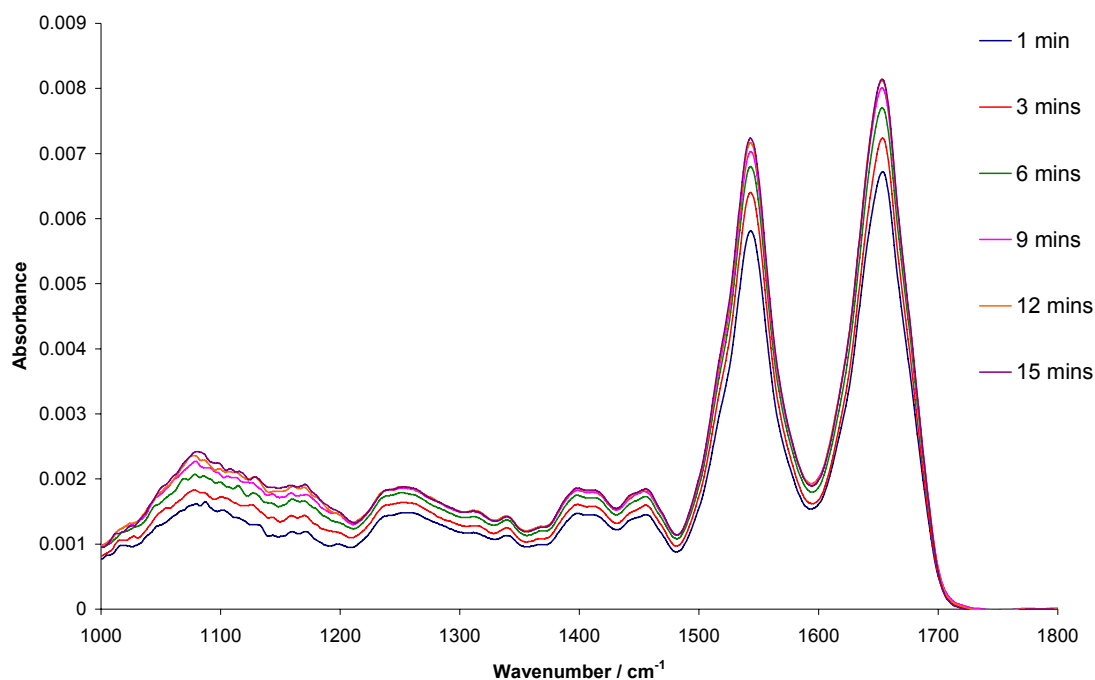


Figure 4.17 FTIR-ATR absorbance spectra of 1 mg/ml lysozyme in buffer as a function of time. Water absorption was corrected by subtracting the appropriate buffer blank spectrum utilizing a scaling factor to obtain a flat baseline in the region 1700–1850 cm^{-1} and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available with OPUS software.

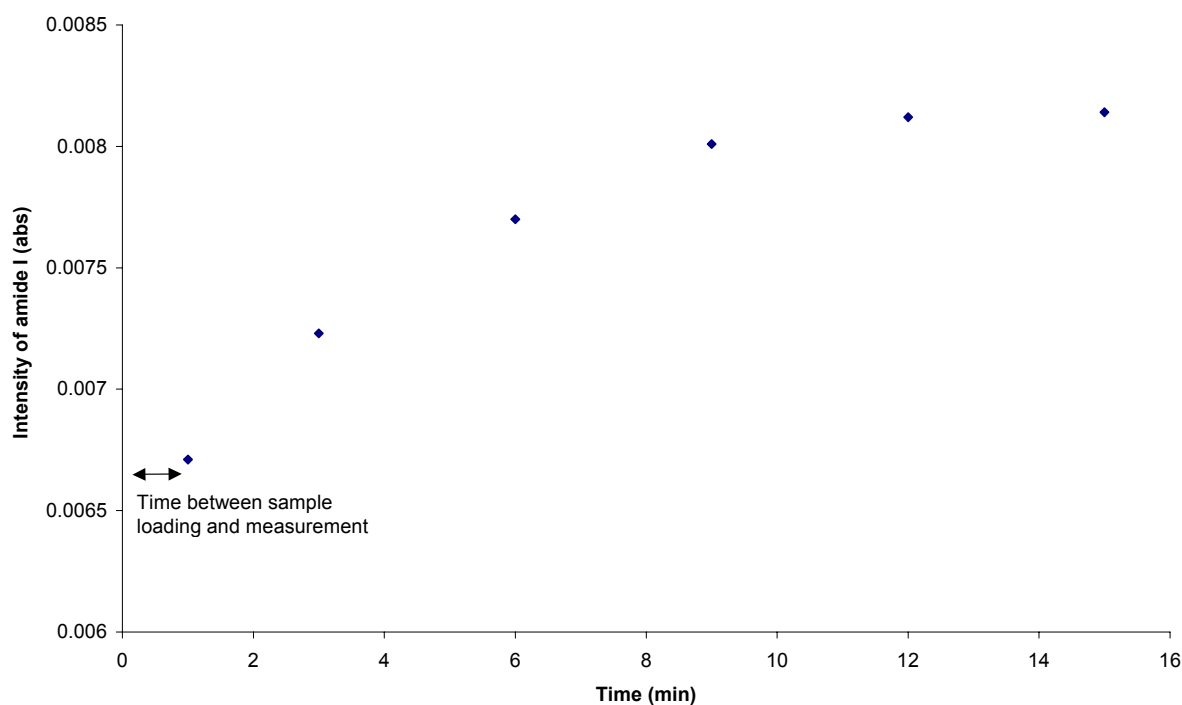


Figure 4.18 Amide I absorbance intensity of lysozyme in water (1 mg/ml) with time (min).

The extent of protein adsorption on to the crystal was investigated by removing the sample from the ATR well by aspiration after incubating for 15 mins. 20 μL of buffer was added, and the spectrum was recorded as soon as possible to minimize protein desorption (Figure 4.19). The spectra obtained shows the characteristics of a protein spectrum with the two bands at the amide I and II positions. The same

procedure was repeated three times, the data suggested that considerable amount of protein was strongly adsorbed to the crystal. In order to visualize the post-adsorption changes in conformation, it is vital to obtain an adsorption corrected spectrum for comparison. This therefore implies that a protein ATR/FTIR spectrum not only contains solution spectrum but also contains the spectrum of adsorbed protein of non-native conformation, as partial denaturation upon adsorption is a common phenomenon.

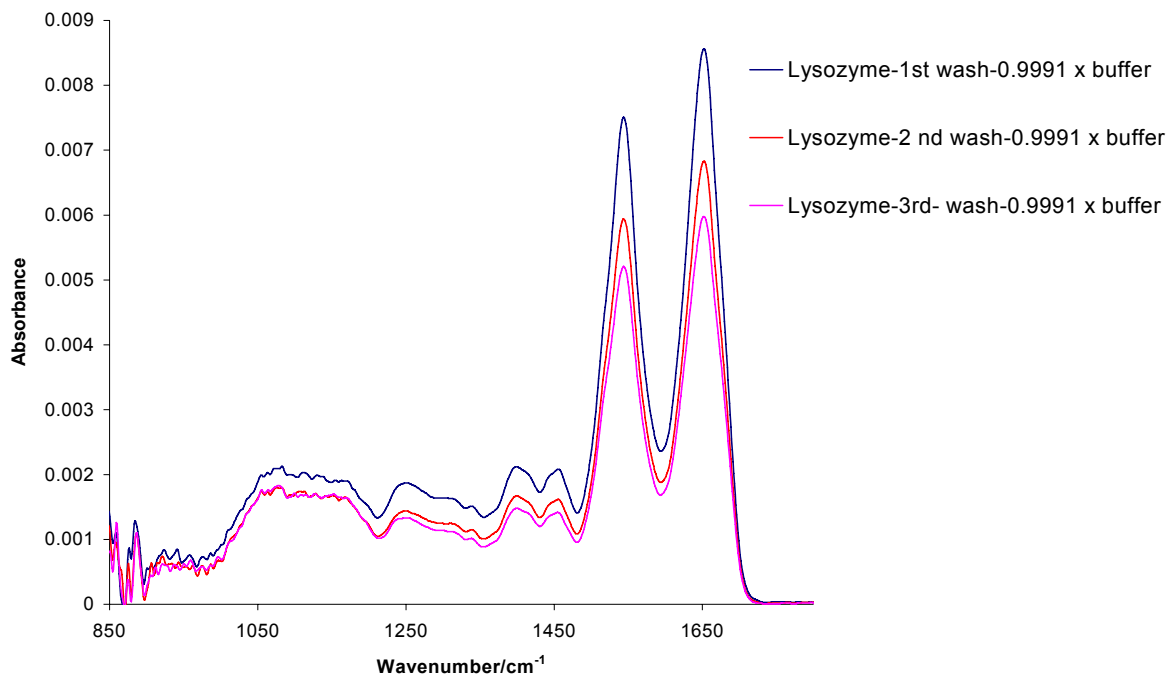


Figure 4.19 Spectra of lysozyme after incubating for 15 mins on the ATR crystal followed by a 20 μ l buffer wash.

Water absorption was corrected by subtracting the appropriate buffer blank spectrum utilizing a scaling factor to obtain a flat baseline in the region 1700–1850 cm^{-1} and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available within the OPUS software.

4.3.4 Cleanliness of the crystal

A considerable amount of protein is adsorbed onto the ATR crystal during each analysis. It is therefore important to make sure that the ATR crystal is protein free before another sample is analysed. Just washing with water does not remove the proteins completely therefore a slightly stronger detergent, 2% Hellmanex, has proved successful in removing adsorbed protein. Before placing the next sample it has to be verified that the crystal is free of proteins by collecting an air spectrum.

4.3.5 Evaluation of glycosylation of proteins by FTIR

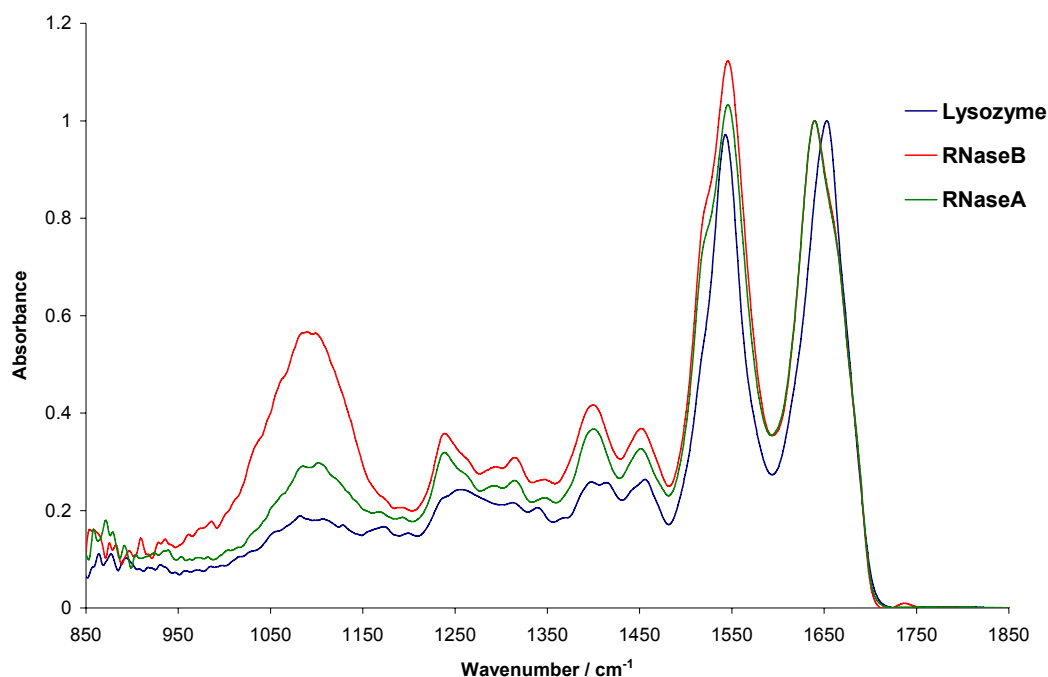


Figure 4.20 FTIR/ATR spectra of lysozyme, RNase A and RNase B.

All proteins were at 10 mg/ml in phosphate buffer. Water absorption was corrected by subtracting the appropriate buffer blank spectrum utilizing a scaling factor to obtain a flat baseline in the region 1700–1850 cm^{-1} and off-set normalized to obtain a minimum absorbance, 0, at 1850 cm^{-1} . Atmospheric water vapour correction for sample and blank spectra were carried prior to subtraction using the facility available within the OPUS software. Spectra were normalised to the amide I bands to give a clear view of intensity differences across the fingerprint region.

FTIR is potentially a useful tool to study the glycosylation of proteins. It has been shown that the bands located within the “fingerprint” region, from 1000–1200 cm^{-1} , are due to the composite modes of vibration of the sugar rings [69]. In addition, the bands of mono- and disaccharides in this region are specific to each sugar and to the environment of the sugar molecules [69]. Figure 4.20 shows the spectra of lysozyme and RNase A (non glycosylated), and RNase B, (glycosylated) and highlights the differences between the proteins in the fingerprint region other than the characteristic amide I and II bands. The broad band in the finger print region of the RNase B may be due to the glycosyl group on the protein. It is interesting to note that although RNase A is marketed as a deglycosylated protein, according to the IR spectrum, it may be somewhat glycosylated compared to the lysozyme spectrum. The small broad band in the finger print region of the lysozyme spectrum is due to the water absorption as water absorbs in this region and it is used as the background absorption for quantifying glycosylation [70]. The high sensitivity of FTIR to carbohydrates in aqueous solution suggests that various degrees of glycosylation of proteins can be studied using FTIR bands [56]. This suggests FTIR may be useful to analyse some aspects of glycosylation [56].

4.3.6 Effect of salt and pH

Salt has an effect on the shape of the water ATR/FTIR spectra. When the salt concentration was increased to 200 mM there is a clear difference in the water spectrum. This might be due to the interaction between water molecules and the sodium and chloride ions. This interaction will affect the stretching and bending modes of the water molecule, which then affects the water spectrum. This has also been observed with changes in pH of the buffer (Figure 4.21).

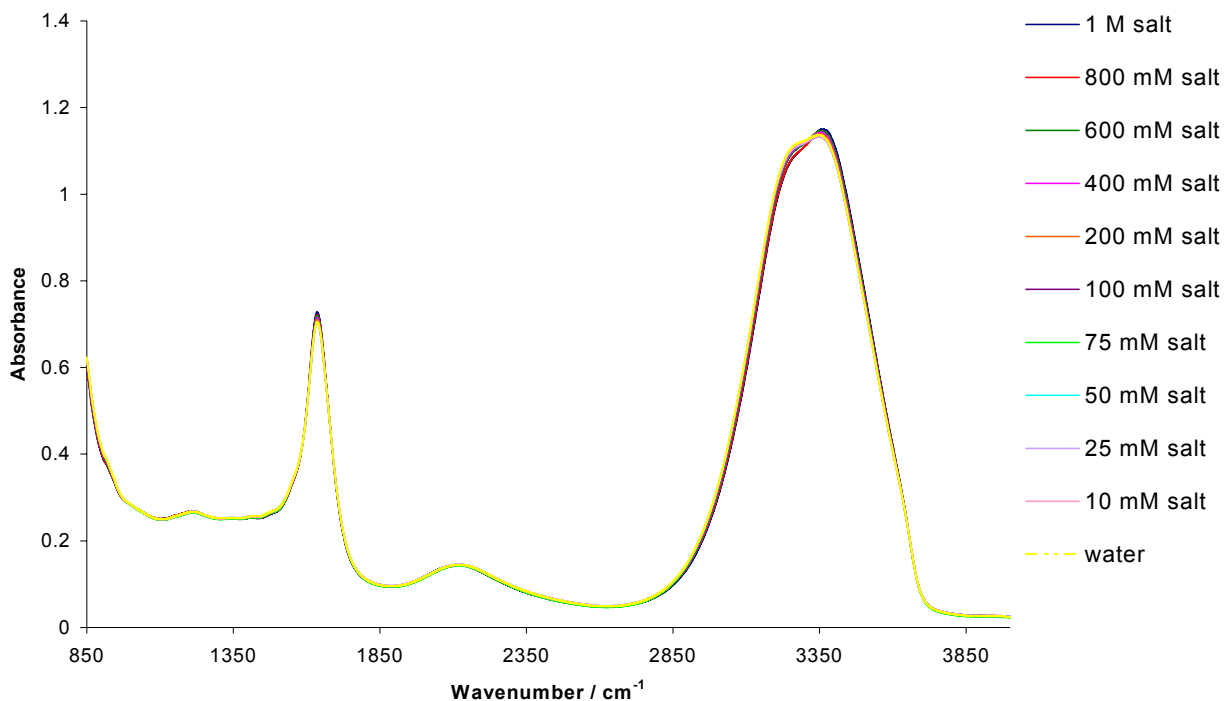


Figure 4.21 Effect of salt concentration on water FTIR-ATR spectra.
Comparison of water spectra in salt concentrations from 10 mM to 1 M.

The same argument is applicable for the pH as it is the negative log of $[H^+]$, however, there is hardly any change observed in the water spectra within the pH range studied (Figure 4.22). Moreover, we might see a change in the shape of the water spectrum if the pH of the water was adjusted using buffer salts rather than using acid or base. Again this difference is more likely to be due to the interaction between the salt ions and water molecules.

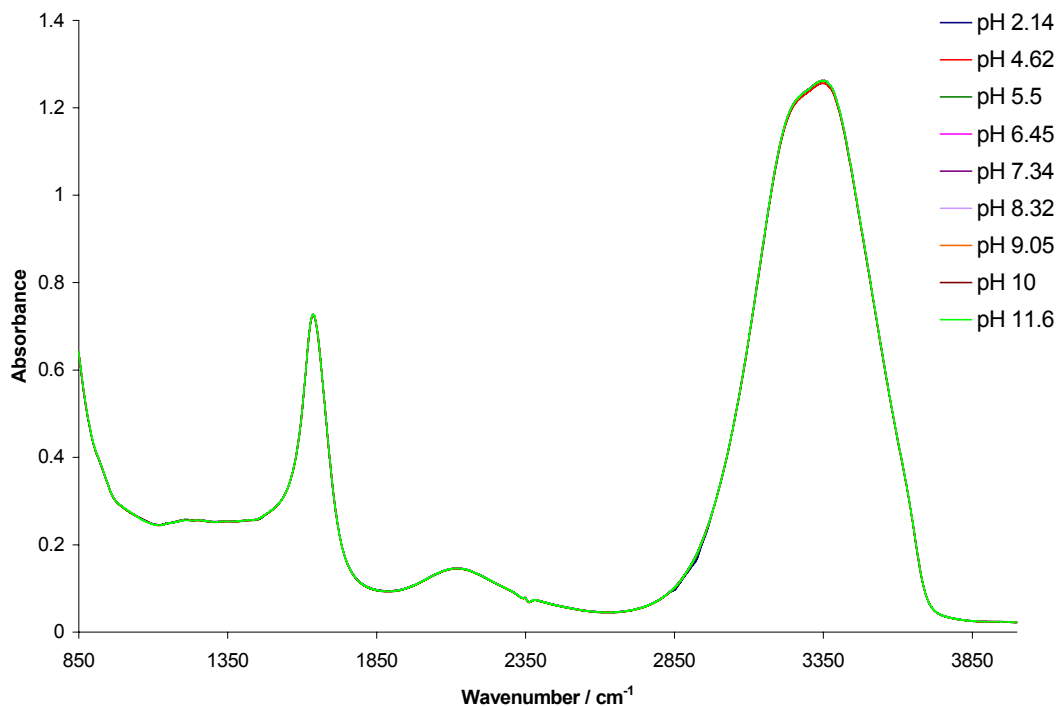


Figure 4.22 Effect of pH on water FTIR-ATR spectrum

4.4 Conclusions

In this chapter we have explored some of the major issues in using FTIR to analyse the secondary structures of proteins in their native aqueous environment. As a result we have come up with a set of recommendations, and these are presented in Table 6.1. Briefly, the key issues are:

- Variations in detector responsivity
- Problems with water subtraction
- Protein adsorption onto the ATR crystal
- Cleaning of the ATR crystal after use

ATR is generally used in protein solution studies; however, there is a growing interest in using transmission cells with very small path lengths. This is because ATR sampling technology requires further processing before analysis of the spectrum and also it has some reproducibility problems. Provided these limitations are taken into account, however, FTIR-ATR is a suitable orthogonal spectroscopic technique to CD for obtaining information about the secondary structure content of proteins in their native aqueous environment.

5 Chemometric analysis of FTIR and CD spectra

5.1 Introduction

In this section, we compare CD and FTIR as techniques for measuring protein structure. We also look at what complementarity is offered where both techniques are used in combination. To address this question in a rigorous way, we applied a chemometric approach to the analysis of a dataset obtained from CD and FTIR measurements of a panel of model proteins that exhibited a variety of structural compositions. Previous studies of this type [4] have not specifically addressed the question of measurement reproducibility.

5.2 Collection of Data Set

5.2.1 Selection of panel

A panel of six proteins was chosen on the basis of their structural diversity, stability and commercial availability. The proteins used are listed in Table 5.1; their structural compositions are presented graphically in Figure 5.1. For each spectral type, six repeat spectra were collected.

Protein	Concentration (mg/ml)	Structural Class	Secondary structure (%)				
			α	β	Turn	3_{10}	Rem.
Cytochrome	13.8	α	41.4	0.0	17.3	0.0	41.3
IgG ⁷	9.4	β	2.2	50.0	13.5	5.5	28.8
Lysozyme	10.3	$\alpha+\beta$	31.0	6.2	23.3	10.9	28.7
Myoglobin	17.8	α	73.9	0.0	13.1	0.0	13.1
RNAse A	8.7	$\alpha+\beta$	17.7	33.1	14.5	3.2	31.5
RNAse B ⁸	9.8						

Table 5.1 Protein Panel for Comparison of FTIR and CD

Structural percentages are from [21], and indicate assignments based on X-ray crystallography data. Classes are α -helix, β -strand, turns, 3_{10} helix, and the remainder (minor structural classes or unassigned). The concentration is the measured concentration of the solution used for FTIR (nominally 10 mg/ml).

⁷ IgG used was bovine; structural percentages are for human.

⁸ Structural percentages are for RNAse A. RNAse B is the glycosylated form of the protein.

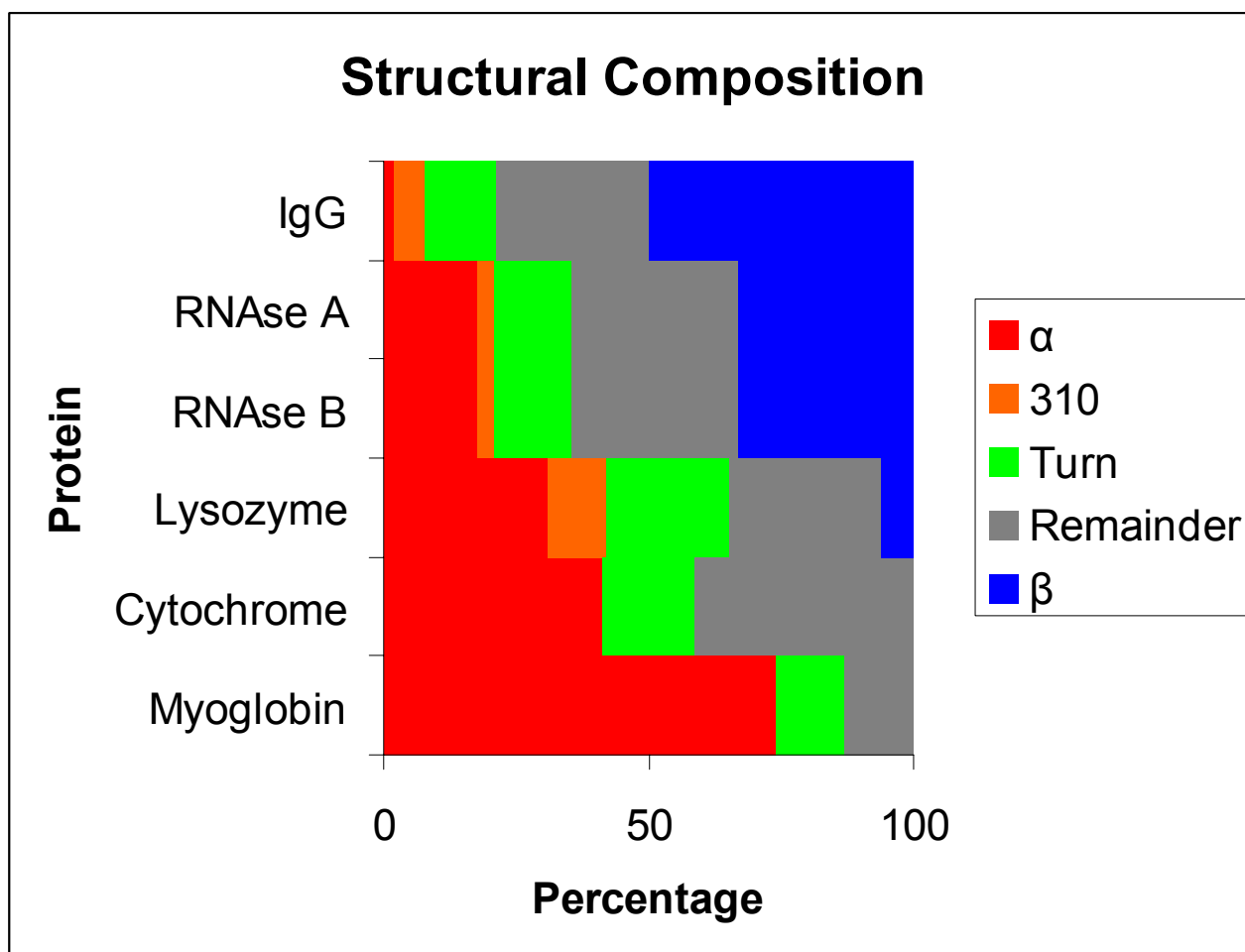


Figure 5.1 Secondary structure composition of proteins in panel

5.2.2 Materials

Proteins cytochrome *c* (equine heart), lysozyme (chicken egg white), myoglobin (equine heart), IgG (bovine), ribonuclease B (RNase B, bovine pancreas), and ribonuclease A (RNase A, type XII-A, bovine pancreas) were used in this study. All proteins except RNase A (Fluka) were purchased from Sigma-Aldrich and were used without further purification. For spectroscopy, proteins were dissolved in phosphate buffer (pH 7.2, 30 mM) at a nominal concentration of 10 mg/ml for FTIR and 1 mg/ml for CD. The actual concentrations of protein solutions were calculated using the 280 nm absorption band using a high performance UV-Vis spectrophotometer (PerkinElmer Lambda 850). Extinction coefficients were obtained from manufacturer's documentation or from the literature [66].

5.2.3 FTIR

All FTIR spectra were collected using a Tensor 27 series FTIR spectrophotometer from Bruker Optics with a BioATR II sampling platform, photovoltaic MCT detector, and a Bruker Optics workstation equipped with OPUS software. Aqueous samples of very low volume (20 μ l) were placed in a circular sampling area of radius 2 mm with a path length of 6 μ m. This multi-reflection ATR accessory is based on a dual crystal technology, which has an upper silicon crystal and a hemispherical zinc-selenide (ZnSe) lower crystal that does not come into contact with the sample. The temperature of the sample was maintained at 20°C by means of flow connectors to a circulating water bath. This accessory was purged continuously throughout the experiment with dry nitrogen via telescopic inserts that seal the optical path inside the spectrometer sample compartment. All FTIR spectra were collected

with resolution 4 cm^{-1} , scanner velocity 20 kHz, 256 scans, phase resolution 32 and zero filling factor 4.

5.2.4 CD

The CD spectra of proteins were collected with ten-fold dilutions of the protein solutions (to $\sim 1\text{ mg/ml}$) that were used to collect FTIR spectra. All CD spectra were obtained on a Jasco J-810 spectrophotometer using the following parameters: standard sensitivity, data pitch 0.1 nm, scanning speed 50 nm/min, band width 1 nm, response time 1 s, and 6 accumulations. The far UV spectra were collected using a 0.01 cm nominal path length cylindrical quartz cell and the near UV using a 0.1 cm nominal path length rectangular quartz cell.

5.3 Spectra

The spectra included in the dataset are presented in the following figures. The far UV CD spectra are presented in Figure 5.2. It is immediately apparent that the variability of the repeat spectra of each protein is confined to a relatively narrow band. The different shapes of the spectra correspond to what would be expected from the known secondary structures of the proteins. For example, the strongly α -helical proteins (myoglobin, cytochrome and lysozyme) display the characteristic α -helix CD spectrum with two negative maxima. The near UV spectra are presented in Figure 5.3, and again show reasonably good superimposition of the repeat spectra, especially when the smaller amplitude of the signals is taken into account. Note the differences between the RNase A and B spectra are of intensity rather than shape, in both near and far UV. This reflects the fact that the RNase B contains a significant glycan component, which contributes to the molecular weight, but not to the CD spectrum.

The FTIR spectra of the proteins are presented in Figure 5.4. Note that there is much more variability between the repeats than is seen for the CD spectra. The spectra shown are not corrected for concentration. The peak wavelengths and peak shapes are crucial for FTIR whereas for CD the peak intensity is also of equal importance. For ease of visual comparison between the spectra, they are presented in normalised form in Figure 5.5. This removes some of the variation between repeat spectra, and also makes it easier to compare the spectra of the different proteins. A close-up of the Amide I peaks is presented in Figure 5.6; in this view, the different peak positions are readily apparent. As would be expected, the strongly α -helical proteins (myoglobin, cytochrome and lysozyme) show peaks between 1650 cm^{-1} and 1658 cm^{-1} , whereas the predominantly β proteins IgG and RNase show peaks at approximately 1630 cm^{-1} to 1640 cm^{-1} . It is also apparent that there is a shoulder in the lysozyme spectrum, which probably corresponds to its small (6.2%) β component. Shoulders visible in some of the peaks in the 1670 cm^{-1} to 1695 cm^{-1} region possibly correspond to antiparallel β -sheet.

Some features of the reproducibility and structural information obtained by these spectroscopic techniques are therefore apparent from a visual inspection of the spectra. However, for a more rigorous and quantitative analysis, we turned to chemometric approaches, as described in the following section.

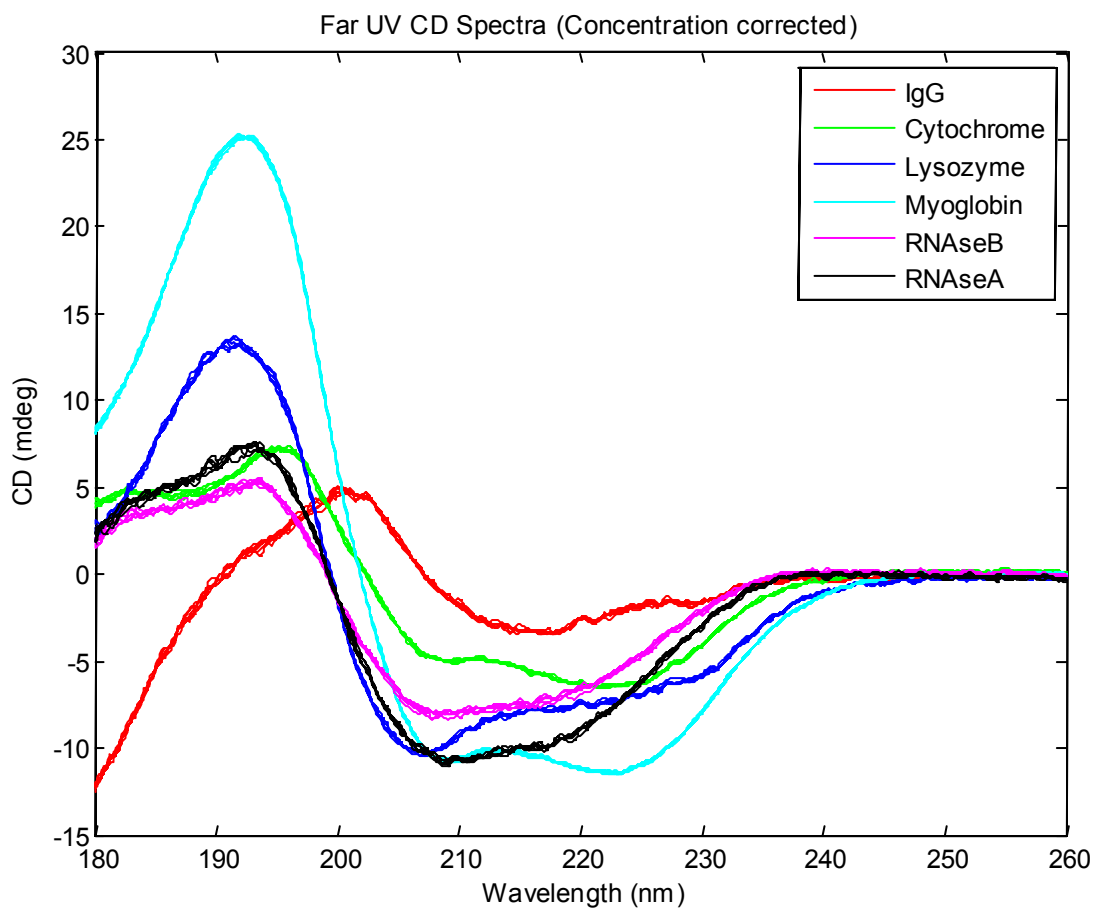


Figure 5.2 Far UV CD Spectra

Spectra are presented in measurement units (mdeg), corrected to nominal concentration of 10 mg/ml using absorbance data (see Table 5.1). Note the close superimposition of repeat spectra.

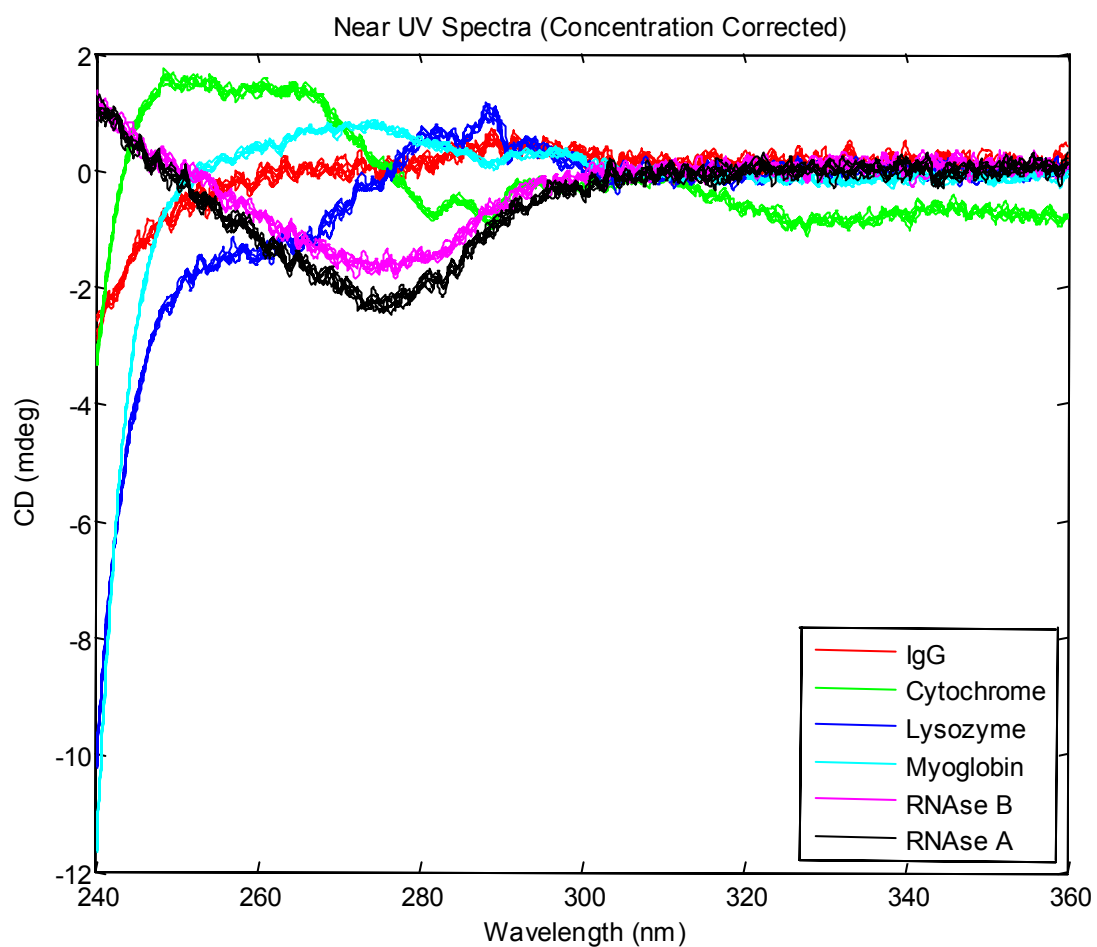


Figure 5.3 Near UV CD Spectra

Spectra are presented in measurement units (mdeg), corrected to nominal concentration of 10 mg/ml using absorbance data (see Table 5.1). Note again the close superimposition of spectra.

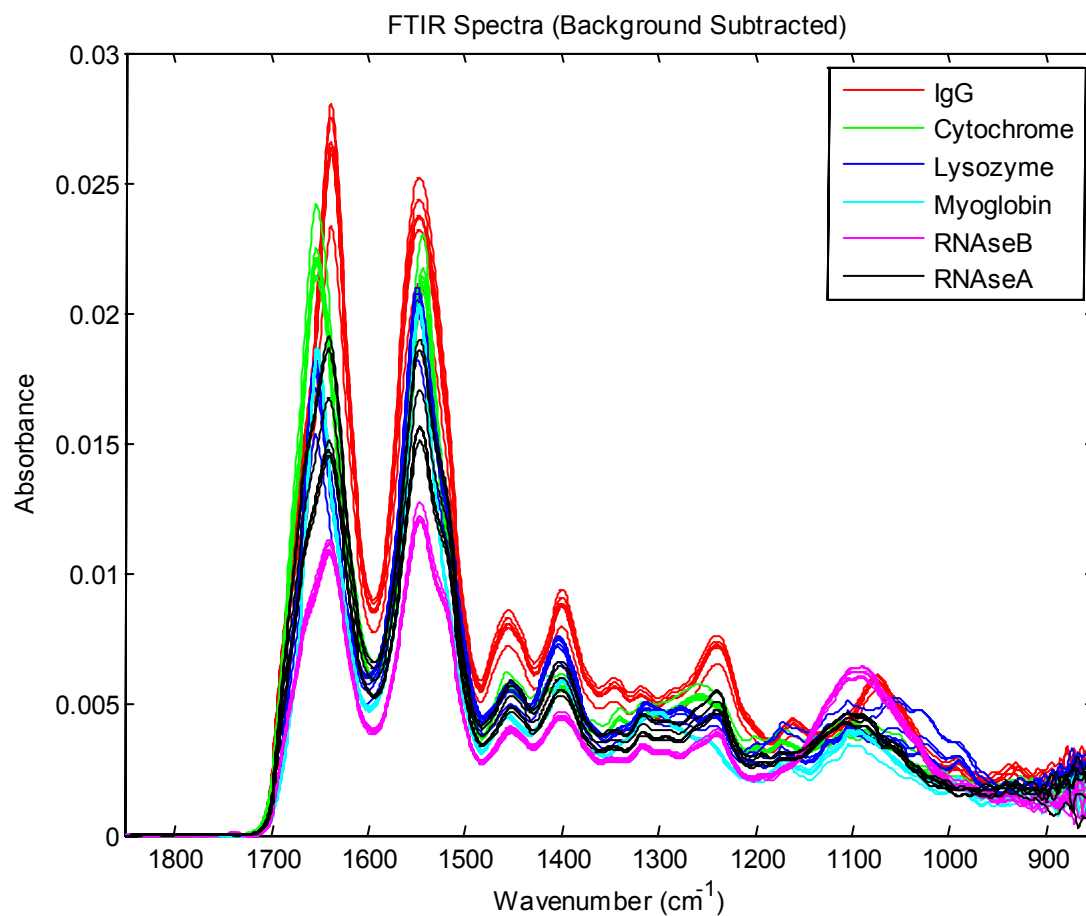


Figure 5.4 *Background-subtracted FTIR-ATR spectra*
 Note the significant variability between repeat spectra.

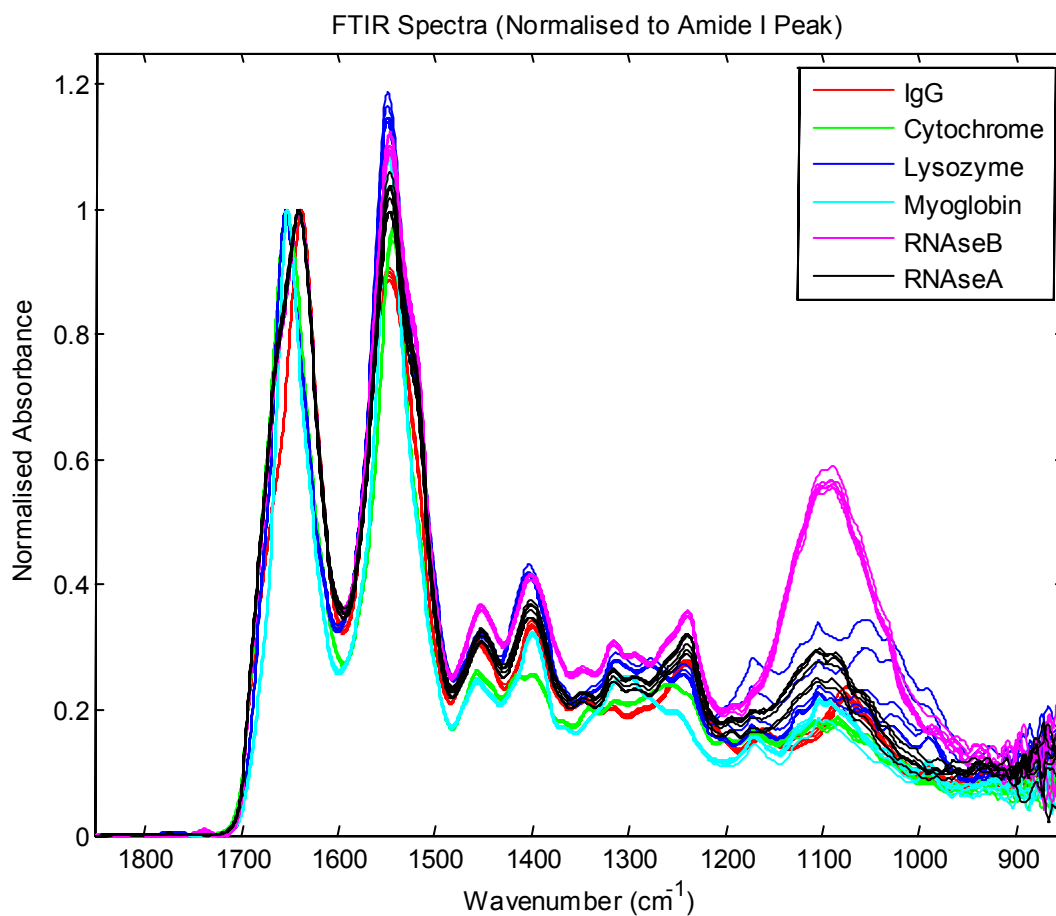


Figure 5.5 *FTIR-ATR spectra, normalised to Amide I peak*

Normalisation to the Amide I peak removes some of the variability and facilitates comparison of spectra between different proteins. Note, in most cases, the closer superposition of the spectra for each protein. Notice also the additional peak for RNase B, which may correspond to the glycans.

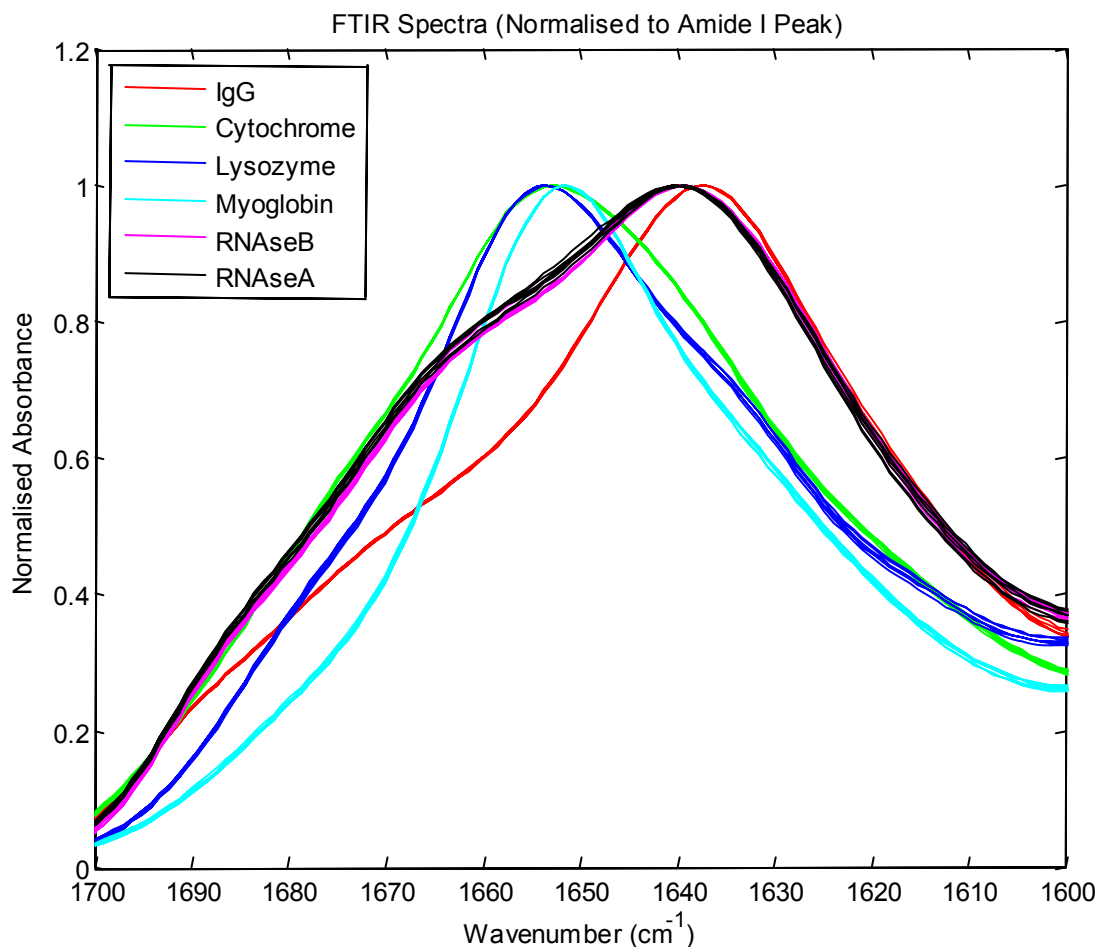


Figure 5.6 Amide I peaks from normalised spectra

This detail view shows clearly the variations within the Amide I peaks of the different proteins.

5.4 Chemometric analysis

5.4.1 Far-UV CD

The Far-UV CD spectra, corrected for concentration (by A_{280} , see Figure 5.2) are very consistent for each protein, and show reasonable discrimination between the different structures, as can be seen from the Covariance Principal Components Analysis (PCA) plot for the first two Factors (Principal Components) presented in Figure 5.7. The first three Factors are sufficient to account for 98.7% of the variance in the spectra, and thus there are, at maximum, three features of the proteins that can be delineated independently from the spectra.

Potential multiple regressions between the first three Factors and the structural features of the proteins are summarised in Table 5.2. The secondary structure features and chromophoric residues of the proteins have been categorized using a scheme developed by Chiralabs optimised for the characterisation of proteins using spectro-structural analysis techniques. It can be seen that, from the limited data set of just 6 proteins, alpha helix, isolated beta strand, BL-type turns and LG-type turns are quantifiable. Moreover, beta sheets and other forms of helices and turns have reasonable regressions. However, great care should be taken, given that we have only 6 proteins (that are not independent) and are regressing with three parameters plus an intercept. This approach should be just used as a guide for comparing the techniques within the limitations of the data set.

It should also be remembered that there are strong correlations and anti-correlations (negative correlations) between the various structural features of the protein, thus some of the apparent ability to quantify structural features is actually a result of this inter-relationship between the structural features. For example, in the data set, phenylalanine is strongly related to the amount of AA and aA turns.

Nonetheless, it is apparent that Far-UV CD is giving consistent spectra that can be interpreted in terms of structural features.

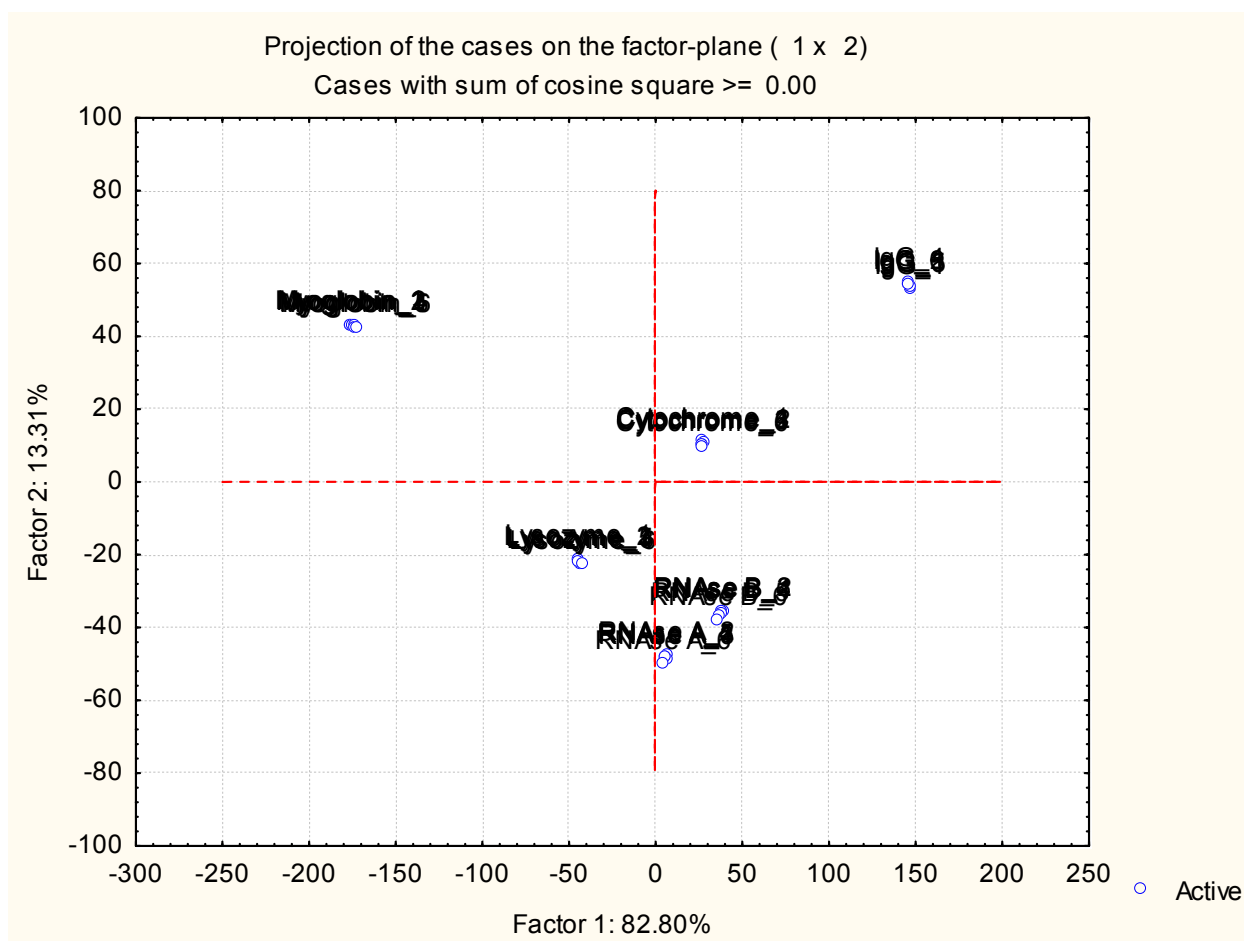


Figure 5.7 Far-UV CD Covariance PCA

Structural Feature	Multiparameter Regression Quality						Regression Equation Coefficients			
	Multiple R	Multiple R ²	Adjusted R ²	F(3,32)	p	Standard Error of Estimate	Intercept	B ₁	B ₂	B ₃
Basic Secondary Structure										
Helix: total	0.9892	0.9784	0.9764	484.22	0.0000	0.0400	0.3327	-0.0025	0.0013	0.0036
Strand: total	0.9424	0.8881	0.8776	84.68	0.0000	0.0662	0.2895	0.0018	-0.0004	-0.0013
Turn: total	0.5604	0.3140	0.2497	4.88	0.0066	0.0668	0.2058	0.0002	-0.0006	-0.0016
Other: total	0.8995	0.8091	0.7913	45.22	0.0000	0.0286	0.1785	0.0005	-0.0002	-0.0017
Secondary Structure Details										
helix: alpha terminal	0.9663	0.9338	0.9276	150.37	0.0000	0.0244	0.1455	-0.0009	-0.0001	0.0016
helix: alpha internal	0.9849	0.9700	0.9672	345.44	0.0000	0.0301	0.1305	-0.0016	0.0015	0.0011
helix: 3-10	0.8739	0.7637	0.7415	34.46	0.0000	0.0191	0.0502	-0.0001	-0.0002	0.0018
helix: pi	0.8842	0.7817	0.7613	38.20	0.0000	0.0072	0.0065	0.0000	0.0000	-0.0007
helix: left-handed P2										
strand: isolated	0.9834	0.9671	0.9640	313.45	0.0000	0.0089	0.0727	0.0004	0.0000	-0.0014
strand: parallel	0.8861	0.7852	0.7651	38.99	0.0000	0.0176	0.0188	0.0001	0.0002	-0.0017
strand: anti-parallel	0.8693	0.7556	0.7327	32.98	0.0000	0.0796	0.1980	0.0013	-0.0006	0.0018
turn: AA	0.9832	0.9667	0.9635	309.31	0.0000	0.0052	0.0847	0.0000	-0.0006	0.0006
turn: Aa	0.8671	0.7518	0.7286	32.31	0.0000	0.0047	0.0080	0.0000	-0.0002	-0.0002
turn: aA	0.9415	0.8863	0.8757	83.17	0.0000	0.0040	0.0085	0.0000	0.0001	-0.0005
turn: AB	0.7423	0.5510	0.5089	13.09	0.0000	0.0055	0.0153	0.0000	-0.0001	-0.0001
turn: PG	0.6278	0.3941	0.3373	6.94	0.0010	0.0143	0.0152	0.0001	0.0001	-0.0005
turn: PL	0.6109	0.3732	0.3144	6.35	0.0017	0.0216	0.0128	0.0000	-0.0001	0.0009
turn: BL	0.9917	0.9835	0.9819	634.85	0.0000	0.0005	0.0017	0.0000	0.0001	0.0001
turn: LL										
turn: LG	0.9581	0.9179	0.9102	119.27	0.0000	0.0025	0.0065	-0.0001	0.0001	0.0002
turn: other	0.8636	0.7457	0.7219	31.28	0.0000	0.0251	0.0532	0.0002	0.0000	-0.0021
uncharacterised	0.8995	0.8091	0.7913	45.22	0.0000	0.0286	0.1785	0.0005	-0.0002	-0.0017
Other Contributions										
disulphide pr pos.	0.7158	0.5124	0.4667	11.21	0.0000	0.0035	0.0073	0.0000	-0.0001	0.0001
disulphide pr neg.	0.9569	0.9156	0.9077	115.78	0.0000	0.0032	0.0117	0.0000	-0.0002	0.0002
tryptophan pr	0.5959	0.3552	0.2947	5.87	0.0026	0.0134	0.0138	0.0000	0.0000	0.0005
tyrosine pr	0.8545	0.7301	0.7048	28.86	0.0000	0.0079	0.0293	0.0001	-0.0003	0.0002
phenylalanine pr	0.9870	0.9742	0.9718	403.57	0.0000	0.0017	0.0332	0.0000	0.0002	-0.0004
porphyrin pr	0.9461	0.8952	0.8853	91.07	0.0000	0.0008	0.0011	0.0000	0.0000	0.0000
Copper pr	0.8842	0.7817	0.7613	38.20	0.0000	0.0014	0.0013	0.0000	0.0000	-0.0001
Iron pr	0.9461	0.8952	0.8853	91.07	0.0000	0.0008	0.0011	0.0000	0.0000	0.0000

Table 5.2 Far-UV CD Regressions
Blue bands indicate good regressions.

5.4.2 Near-UV CD

The Near-UV CD spectra, corrected for concentration (A_{280} , see Figure 5.3) are likewise very consistent for each protein and show reasonable discrimination between the different structures, as can be seen from the Covariance Principal Components Analysis (PCA) plot for the first two Factors presented in Figure 5.8. The first three Factors are sufficient to account for 98.5% of the variance in the spectra and thus there are, at maximum, three features of the proteins that can be delineated independently from the spectra.

Potential multiple regressions between the first three Factors and the structural features of the proteins are summarised in Table 5.3. It can be seen that, from the limited data set of just 6 proteins, the various chromophoric residues are quantifiable, particularly tyrosine and phenylalanine. Again, care must be taken in interpreting the results. This spectral region, which does not cover any significant signals from the peptidic backbone and is dominated by the residue chromophores, but is apparently capable to quantifying some of the secondary structural elements. This is predominantly just a result of the fortuitous correlation between the chromophoric residues and secondary structure in this limited data set.

Nonetheless, it is apparent that Near-UV CD is giving consistent spectra that can be interpreted in terms of protein features.

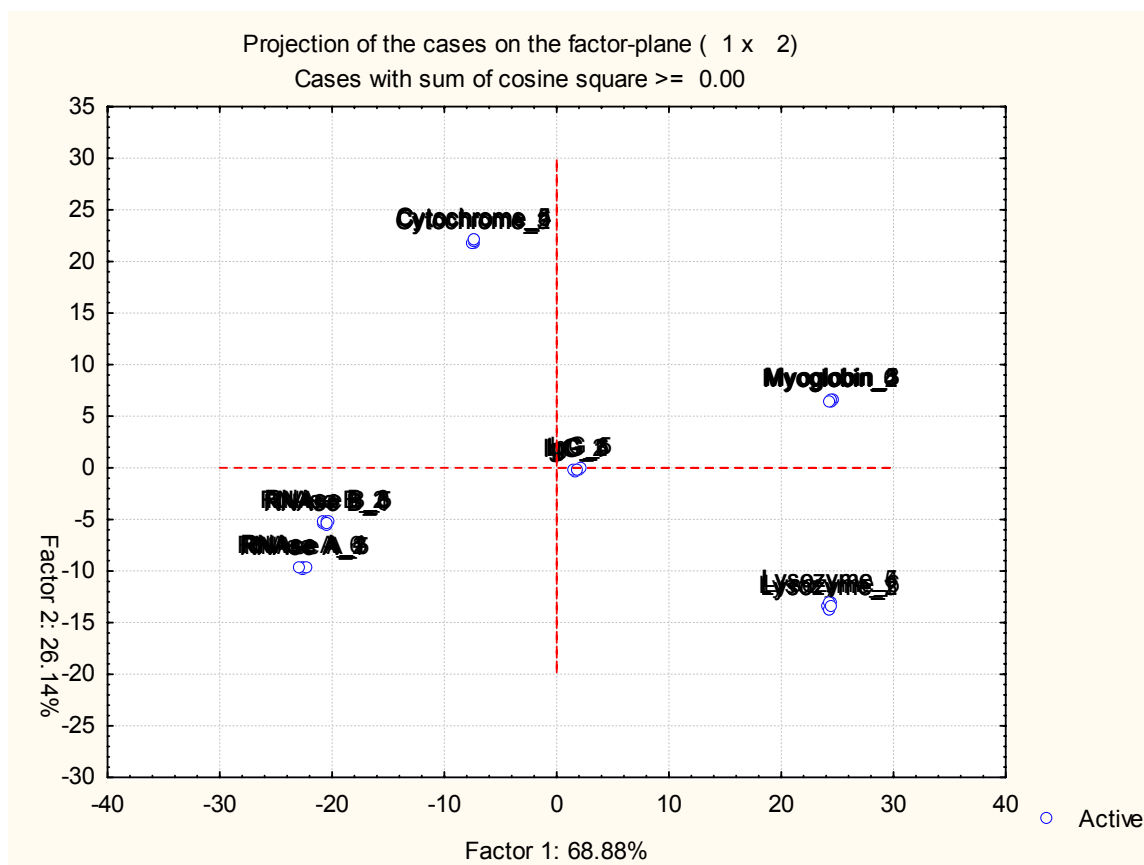


Figure 5.8 Near-UV CD Covariance PCA

Structural Feature	Multiparameter Regression Quality						Regression Equation Coefficients			
	Multiple R	Multiple R ²	Adjusted R ²	F(3,32)	p	Standard Error of Estimate	Intercept	B ₁	B ₂	B ₃
Basic Secondary Structure										
Helix: total	0.8378	0.7019	0.6740	25.12	0.0000	0.1486	0.3327	0.0094	-0.0018	0.0270
Strand: total	0.9762	0.9530	0.9486	216.19	0.0000	0.0429	0.2895	-0.0076	0.0001	-0.0259
Turn: total	0.2788	0.0777	-0.0087	0.90	0.4524	0.0775	0.2058	-0.0004	0.0012	0.0033
Other: total	0.6562	0.4306	0.3772	8.07	0.0004	0.0494	0.1785	-0.0016	0.0018	-0.0036
Secondary Structure Details										
helix: alpha terminal	0.7950	0.6320	0.5976	18.32	0.0000	0.0575	0.1455	0.0025	-0.0020	0.0109
helix: alpha internal	0.7889	0.6223	0.5869	17.58	0.0000	0.1069	0.1305	0.0058	0.0015	0.0149
helix: 3-10	0.9694	0.9398	0.9342	166.53	0.0000	0.0096	0.0502	0.0012	-0.0023	0.0006
helix: pi	0.8760	0.7674	0.7456	35.20	0.0000	0.0074	0.0065	-0.0001	0.0010	0.0006
helix: left-handed P2										
strand: isolated	0.8978	0.8061	0.7879	44.35	0.0000	0.0216	0.0727	-0.0017	0.0016	-0.0045
strand: parallel	0.8707	0.7581	0.7354	33.42	0.0000	0.0187	0.0188	-0.0003	0.0026	0.0004
strand: anti-parallel	0.9869	0.9739	0.9715	397.99	0.0000	0.0260	0.1980	-0.0056	-0.0041	-0.0218
turn: AA	0.9394	0.8825	0.8715	80.09	0.0000	0.0097	0.0847	-0.0004	-0.0019	0.0025
turn: Aa	0.9814	0.9632	0.9598	279.47	0.0000	0.0018	0.0080	-0.0004	-0.0002	0.0007
turn: aA	0.9815	0.9633	0.9599	280.01	0.0000	0.0023	0.0085	0.0000	0.0009	0.0002
turn: AB	0.3257	0.1061	0.0223	1.27	0.3027	0.0077	0.0153	-0.0001	-0.0001	0.0001
turn: PG	0.6221	0.3871	0.3296	6.74	0.0012	0.0144	0.0152	0.0001	0.0009	-0.0006
turn: PL	0.7856	0.6172	0.5813	17.20	0.0000	0.0169	0.0128	0.0008	-0.0011	0.0006
turn: BL	0.9506	0.9036	0.8946	100.02	0.0000	0.0012	0.0017	0.0000	0.0000	-0.0008
turn: LL										
turn: LG	0.9710	0.9428	0.9374	175.84	0.0000	0.0021	0.0065	0.0004	-0.0001	0.0004
turn: other	0.7517	0.5651	0.5243	13.86	0.0000	0.0328	0.0532	-0.0008	0.0027	0.0003
uncharacterised	0.6562	0.4306	0.3772	8.07	0.0004	0.0494	0.1785	-0.0016	0.0018	-0.0036
Other Contributions										
disulphide pr pos.	0.5313	0.2823	0.2150	4.20	0.0130	0.0042	0.0073	0.0000	-0.0002	0.0002
disulphide pr neg.	0.9843	0.9689	0.9660	332.69	0.0000	0.0019	0.0117	-0.0003	-0.0007	0.0004
tryptophan pr	0.8553	0.7315	0.7063	29.06	0.0000	0.0087	0.0138	0.0007	-0.0004	-0.0001
tyrosine pr	0.9886	0.9774	0.9753	461.04	0.0000	0.0023	0.0293	-0.0006	-0.0007	0.0000
phenylalanine pr	0.9684	0.9379	0.9320	161.01	0.0000	0.0027	0.0332	0.0002	0.0008	0.0000
porphyrin pr	0.6728	0.4527	0.4013	8.82	0.0002	0.0019	0.0011	0.0001	0.0001	0.0001
Copper pr	0.8760	0.7674	0.7456	35.20	0.0000	0.0015	0.0013	0.0000	0.0002	0.0001
Iron pr	0.6728	0.4527	0.4013	8.82	0.0002	0.0019	0.0011	0.0001	0.0001	0.0001

Table 5.3 Near-UV CD Regressions
Blue bands indicate good regressions.

5.4.3 FT-IR

The FT-IR spectra, corrected for concentration (A_{280} , see Figure 5.4) are much poorer in their consistency for a given protein, although they do show a reasonable discrimination between the different structures, as can be seen from the Covariance Principal Components Analysis (PCA) plot for the first two Factors presented in Figure 5.9. The first three Factors are sufficient to account for 98.9% of the variance in the spectra and thus there are, at maximum, three features of the proteins that can be delineated independently from the spectra.

A major source of inconsistency would seem to be a variation in either concentration or detection sensitivity.

Potential multiple regressions between the first three Factors and the structural features of the proteins are summarised in Table 5.4. It can be seen that the ability of the FT-IR spectra to quantify protein features is much poorer in comparison with CD in either wavelength region, although it has success with anti-parallel beta strand.

The lack of quantitation and distinguishing power of the FT-IR is, in part, due to the inconsistencies of the spectra. There are two approaches to improving these results:

- Counter the inconsistencies in spectral magnitude
- Explore using smaller wavenumber ranges, for example just the amide peaks.

For the purposes of this study, the corresponding PCA and subsequent multiple regressions were attempted, but using just the wavenumber region $1480\text{-}1720\text{cm}^{-1}$. This region covers the two main amide bands of protein peptidic backbones and is therefore most relevant to the assessment of secondary structure rather than other structural or compositional features.

The Covariance Principal Components Analysis (PCA) plot for the first two Factors is presented in Figure 5.10. Again, the spectra display considerable discrimination between the structures, albeit still with a substantial level of inconsistency in the spectra for any given protein. The first three Factors are sufficient to account for 99.4% of the variance in the spectra and thus there are, at maximum, three features of the proteins that can be delineated independently from the spectra using this restricted wavenumber region.

Potential multiple regressions between the first three Factors and the structural features of the proteins are summarised in Table 5.5. It can be seen that the ability of the FT-IR spectra to quantify protein features is much improved by the restriction to the $1480\text{-}1720\text{cm}^{-1}$ wavenumber range. In particular, the method has reasonable correlations with various helices, beta strand and several types of turns, although, again, care should be taken in over-interpreting, due to the fortuitous correlations between the structures in the dataset.

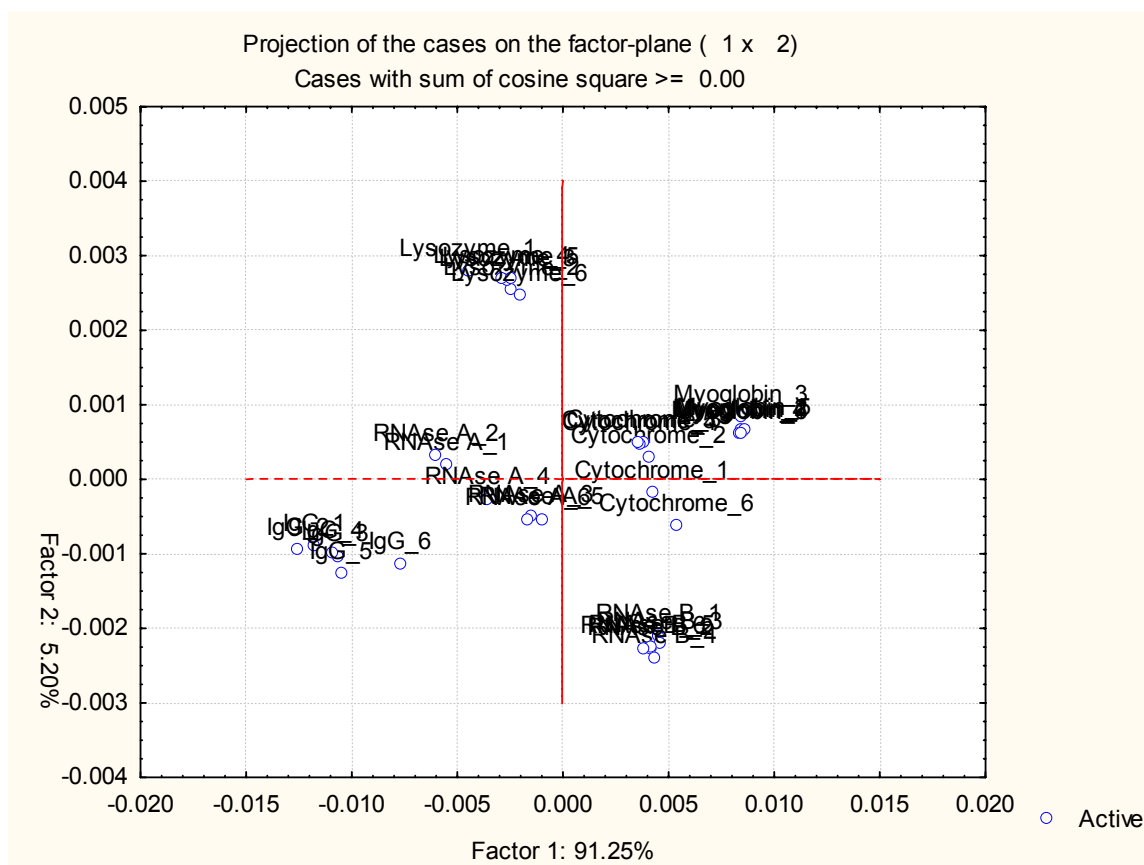


Figure 5.9 FT-IR Covariance PCA

Structural Feature	Multiparameter Regression Quality						Regression Equation Coefficients			
	Multiple R	Multiple R ²	Adjusted R ²	F(3,32)	p	Standard Error of Estimate	Intercept	B ₁	B ₂	B ₃
Basic Secondary Structure										
Helix: total	0.7581	0.5747	0.5348	14.41	0.0000	0.1775	0.3327	22.7822	80.0043	38.7381
Strand: total	0.9336	0.8717	0.8596	72.45	0.0000	0.0709	0.2895	-17.0376	-88.8017	-2.4093
Turn: total	0.5162	0.2665	0.1977	3.88	0.0180	0.0691	0.2058	-1.3441	17.6588	-26.1393
Other: total	0.4287	0.1838	0.1073	2.40	0.0859	0.0591	0.1785	-3.5434	-8.0162	-5.6166
Secondary Structure Details										
helix: alpha terminal	0.6791	0.4612	0.4107	9.13	0.0002	0.0696	0.1455	7.8736	22.0245	-2.9677
helix: alpha internal	0.8018	0.6428	0.6093	19.20	0.0000	0.1040	0.1305	15.7031	41.8893	53.4300
helix: 3-10	0.8155	0.6651	0.6337	21.18	0.0000	0.0227	0.0502	-1.4447	15.6365	-15.2776
helix: pi	0.3861	0.1491	0.0693	1.87	0.1548	0.0142	0.0065	0.6502	0.4540	3.5533
helix: left-handed P2										
strand: isolated	0.6561	0.4305	0.3771	8.06	0.0004	0.0370	0.0727	-2.7164	-15.9843	4.5925
strand: parallel	0.3763	0.1416	0.0611	1.76	0.1748	0.0352	0.0188	0.9811	0.0331	11.4601
strand: anti-parallel	0.9856	0.9713	0.9686	361.44	0.0000	0.0272	0.1980	-15.3023	-72.8506	-18.4619
turn: AA	0.9459	0.8948	0.8850	90.73	0.0000	0.0092	0.0847	-0.3796	2.1388	-24.0193
turn: Aa	0.8158	0.6656	0.6342	21.23	0.0000	0.0055	0.0080	0.2169	-3.4182	-4.7104
turn: aA	0.7131	0.5085	0.4624	11.04	0.0000	0.0083	0.0085	0.6571	0.4761	6.4420
turn: AB	0.6560	0.4304	0.3770	8.06	0.0004	0.0062	0.0153	-0.4247	0.7751	-3.9255
turn: PG	0.4008	0.1607	0.0820	2.04	0.1277	0.0168	0.0152	-0.3960	3.3466	3.7938
turn: PL	0.9228	0.8516	0.8377	61.20	0.0000	0.0105	0.0128	-1.1136	12.7229	-11.0903
turn: BL	0.9341	0.8726	0.8606	73.04	0.0000	0.0014	0.0017	-0.4351	-0.7385	1.6875
turn: LL										
turn: LG	0.7749	0.6004	0.5630	16.03	0.0000	0.0056	0.0065	0.3689	3.7773	1.4496
turn: other	0.1072	0.0115	-0.0812	0.12	0.9453	0.0494	0.0532	0.1619	-1.4214	4.2332
uncharacterised	0.4287	0.1838	0.1073	2.40	0.0859	0.0591	0.1785	-3.5434	-8.0162	-5.6166
Other Contributions										
disulphide pr pos.	0.9238	0.8534	0.8397	62.11	0.0000	0.0019	0.0073	-0.2111	1.2681	-3.4501
disulphide pr neg.	0.9044	0.8179	0.8009	47.92	0.0000	0.0047	0.0117	-0.3391	-1.7601	-8.2863
tryptophan pr	0.8685	0.7543	0.7313	32.75	0.0000	0.0083	0.0138	-0.6162	8.4078	-2.4379
tyrosine pr	0.8886	0.7896	0.7699	40.04	0.0000	0.0069	0.0293	-0.5379	-5.2730	-8.7982
phenylalanine pr	0.9361	0.8763	0.8647	75.58	0.0000	0.0038	0.0332	0.6755	0.3764	8.0291
porphyrin pr	0.8120	0.6593	0.6274	20.64	0.0000	0.0015	0.0011	0.2224	0.3161	1.2182
Copper pr	0.3861	0.1491	0.0693	1.87	0.1548	0.0028	0.0013	0.1300	0.0908	0.7107
Iron pr	0.8120	0.6593	0.6274	20.64	0.0000	0.0015	0.0011	0.2224	0.3161	1.2182

Table 5.4 FT-IR Regressions
Blue bands indicate good regressions.

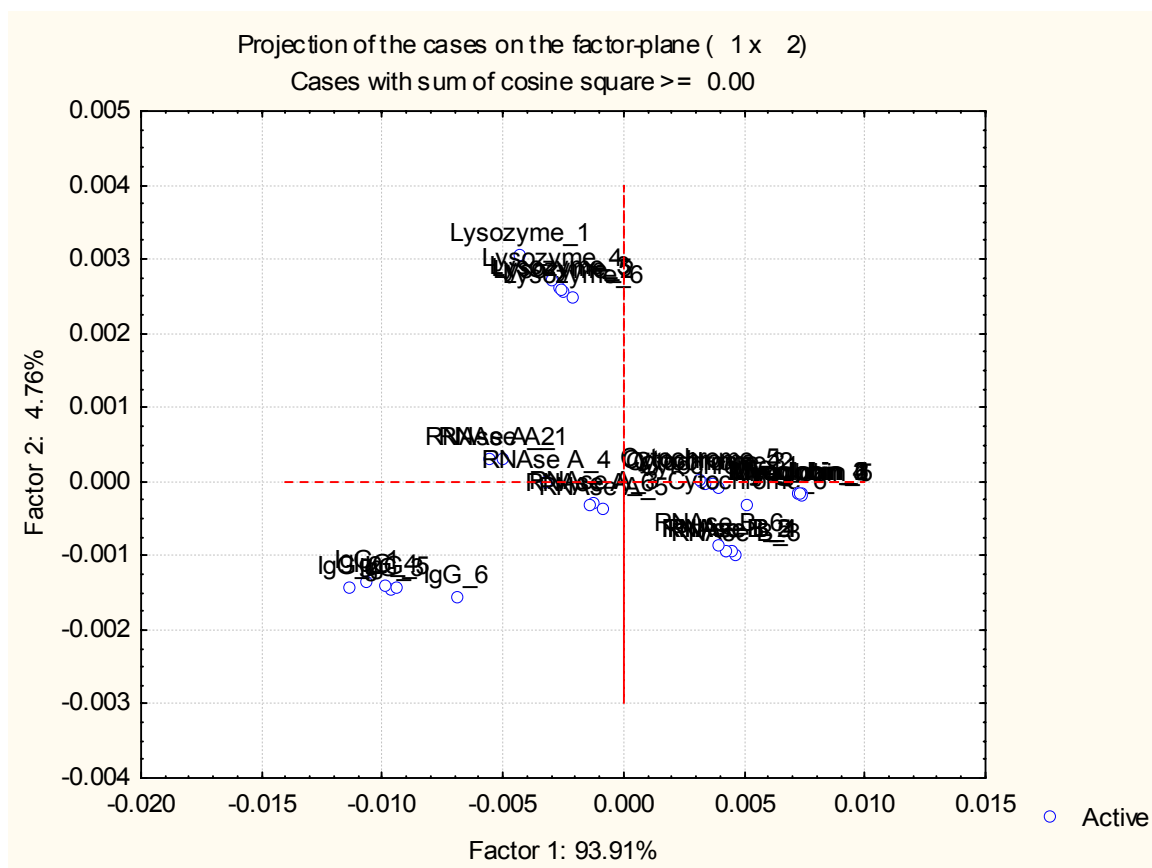


Figure 5.10 FT-IR 1480-1720cm⁻¹ Covariance PCA

Structural Feature	Multiparameter Regression Quality						Regression Equation Coefficients			
	Multiple R	Multiple R ²	Adjusted R ²	F(3,32)	p	Standard Error of Estimate	Intercept	B ₁	B ₂	B ₃
Basic Secondary Structure										
Helix: total	0.8110	0.6578	0.6257	20.50	0.0000	0.1592	0.3327	23.7927	68.7565	252.0933
Strand: total	0.8855	0.7841	0.7639	38.75	0.0000	0.0919	0.2895	-17.6933	-93.3989	-85.9691
Turn: total	0.8677	0.7529	0.7297	32.50	0.0000	0.0401	0.2060	-1.4570	29.7110	-104.3190
Other: total	0.8556	0.7321	0.7070	29.15	0.0000	0.0339	0.1785	-3.6895	-5.1725	-94.4727
Secondary Structure Details										
helix: alpha terminal	0.8926	0.7967	0.7777	41.81	0.0000	0.0427	0.1455	8.3197	23.0728	110.6822
helix: alpha internal	0.7300	0.5328	0.4890	12.17	0.0000	0.1189	0.1305	16.4919	24.6584	127.6372
helix: 3-10	0.9755	0.9517	0.9471	210.08	0.0000	0.0086	0.0502	-1.7458	21.2956	40.8168
helix: pi	0.9846	0.9695	0.9666	339.02	0.0000	0.0027	0.0065	0.7269	-0.2704	-27.0429
helix: left-handed P2										
strand: isolated	0.9097	0.8275	0.8113	51.17	0.0000	0.0204	0.0727	-2.7675	-17.7704	-61.8237
strand: parallel	0.9761	0.9527	0.9483	214.86	0.0000	0.0083	0.0188	1.1101	-2.7848	-67.6028
strand: anti-parallel	0.8883	0.7891	0.7694	39.92	0.0000	0.0739	0.1980	-16.0359	-72.8437	43.4574
turn: AA	0.6630	0.4396	0.3871	8.37	0.0003	0.0213	0.0847	-0.3746	10.2865	22.5575
turn: Aa	0.3499	0.1225	0.0402	1.49	0.2363	0.0089	0.0080	0.2930	-2.0040	0.5076
turn: aA	0.9160	0.8390	0.8239	55.59	0.0000	0.0047	0.0085	0.7135	-1.4076	-18.1024
turn: AB	0.7415	0.5498	0.5076	13.03	0.0000	0.0055	0.0153	-0.4522	2.3170	-8.0435
turn: PG	0.8948	0.8006	0.7819	42.83	0.0000	0.0082	0.0152	-0.4543	2.7270	-29.3653
turn: PL	0.9429	0.8890	0.8786	85.46	0.0000	0.0091	0.0128	-1.3305	17.2792	8.4825
turn: BL	0.8915	0.7947	0.7755	41.30	0.0000	0.0018	0.0017	-0.4776	-1.4035	0.2341
turn: LL										
turn: LG	0.7299	0.5327	0.4889	12.16	0.0000	0.0060	0.0065	0.3522	3.4297	6.9400
turn: other	0.9461	0.8952	0.8853	91.07	0.0000	0.0161	0.0532	0.2732	-1.5136	-87.5299
uncharacterised	0.8556	0.7321	0.7070	29.15	0.0000	0.0339	0.1785	-3.6895	-5.1725	-94.4727
Other Contributions										
disulphide pr pos.	0.7843	0.6151	0.5790	17.05	0.0000	0.0031	0.0073	-0.2319	2.5688	-1.0090
disulphide pr neg.	0.5067	0.2567	0.1870	3.68	0.0219	0.0094	0.0117	-0.3331	0.8438	9.3088
tryptophan pr	0.8638	0.7461	0.7223	31.35	0.0000	0.0084	0.0138	-0.7638	9.7458	3.7251
tyrosine pr	0.4957	0.2457	0.1750	3.47	0.0272	0.0131	0.0293	-0.5133	-2.7605	10.4932
phenylalanine pr	0.7194	0.5176	0.4723	11.44	0.0000	0.0074	0.0332	0.7202	-2.2011	-10.2380
porphyrin pr	0.6581	0.4331	0.3800	8.15	0.0004	0.0019	0.0011	0.2350	-0.1051	1.6147
Copper pr	0.9846	0.9695	0.9666	339.02	0.0000	0.0005	0.0013	0.1454	-0.0541	-5.4086
Iron pr	0.6581	0.4331	0.3800	8.15	0.0004	0.0019	0.0011	0.2350	-0.1051	1.6147

Table 5.5 FT-IR (1480-1720cm⁻¹) Regressions
Blue bands indicate good regressions

5.4.4 Comparison of techniques

The Near-UV CD, while not relevant to the determination of protein secondary structure, provides a valuable insight into the nature of chromophoric components. Likewise, taken individually, Far-UV CD and FT-IR (restricted to the amide region) provide a degree of discrimination between secondary structural classes and can be used to characterise protein folding. However, the reproducibility problems of FT-IR spectra are limiting for this technique.

However, it is notable that the successful secondary structure correlations derived from the amide region FT-IR spectra are complementary to those achieved with Far-UV CD. The combination of Far-UV CD and amide region FT-IR together is therefore a natural extension to enhance the determination of protein secondary structure. This may be achieved either by

- i) interpreting the two complementary sets of data separately and then combining the interpretations to provide a overall characterisation, or
- ii) combining the CD and FT-IR spectra together to give a “super-spectrum”, which can be subjected to various pattern recognition and multidimensional analyses

The latter requires careful consideration of the relative weighting of the IR and Far-UV region signals, as they will typically be on significantly different magnitude scales. Nonetheless, it provides an opportunity to improve protein characterisation (assuming the practicalities of obtaining reproducible FT-IR spectra of aqueously dissolved proteins can be surmounted).

5.4.5 Combination of CD and FT-IR Data

On appending the Far-UV CD spectrum (re-digitised to 1 nm data intervals to give a data size comparable with the FT-IR) with the FT-IR spectrum (restricted to the amide 1720-1480 cm^{-1} region and scaled by a factor of 10^4 to give comparable magnitudes to that of the CD), one obtains a “super-spectrum” for each sample. The set of super-spectra for all the samples can be subjected to a Covariance Principal Components Analysis, analogous to that done for the CD or FT-IR spectra alone.

Such an analysis shows reasonable discrimination between the proteins, as can be seen from the PCA plot for the first two Factors presented in Figure 5.11. As previously with the FT-IR alone, there is still considerable dispersion for each protein, deriving from the inconsistency of the FT-IR spectra. Although this can be further corrected by, say, normalising against peak height, this would artefactually bias the analysis and is not done here with the limited dataset available. As might be expected, myoglobin has a greater distinction from the other proteins than with FT-IR alone, due to the strong relationship between Far-UV CD and alpha helix, myoglobin having a substantial quantity of this particular secondary structure.

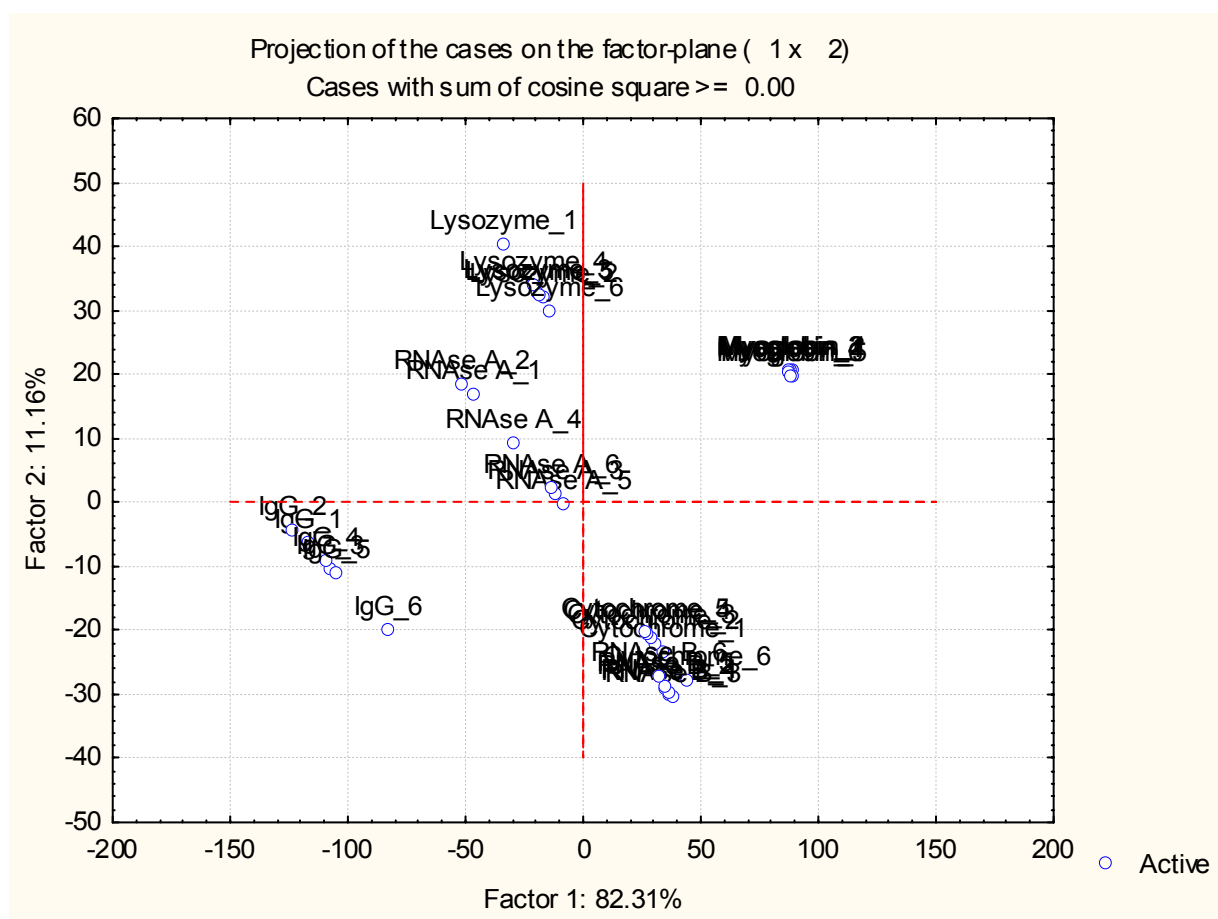


Figure 5.11 Far-UV CD & amide region FT-IR “Super-spectra” Covariance PCA

From the PCA, the first three, four and five Factors cumulatively account for 97.8%, 99.4% and 99.9% of the variance respectively, indicating that there may be marginally more information content than that seen with either Far-UV CD or FT-IR alone, where three Factors covered 98.7% and 99.4% of the spectral variance respectively.

Potential multiple regressions between the first three Factors and the structural features of the proteins are summarised in Table 5.6. It can be seen that, from the limited data set of just 6 proteins, an improvement over Far-UV CD alone is evident regarding correlation with total strand, but at the expense of relationships with total helix and “other” (i.e. uncharacterised structures). Moreover, inspecting the relationships between the detailed structural elements and the Factors, it can be seen that the strong relationships with Far-UV CD alone have been predominantly lost excepting for internal alpha helix and BL-type turns. Moreover, none of the previous relationships with the amide region FT-IR have been preserved. There has therefore been little overall improvement in ability to structurally characterise the proteins over and above that achievable with Far-UV CD alone.

Structural Feature	Multiparameter Regression Quality						Regression Equation Coefficients			
	Multiple R	Multiple R ²	Adjusted R ²	F(3,32)	p	Standard Error of Estimate	Intercept	B ₁	B ₂	B ₃
Basic Secondary Structure										
Helix: total	0.9702	0.9413	0.9358	170.92	0.0000	0.0660	0.3327	0.0026	0.0075	-0.0053
Strand: total	0.9677	0.9365	0.9306	157.39	0.0000	0.0498	0.2895	-0.0019	-0.0058	0.0008
Turn: total	0.6145	0.3777	0.3193	6.47	0.0015	0.0637	0.2058	-0.0002	-0.0004	0.0031
Other: total	0.8006	0.6410	0.6073	19.05	0.0000	0.0392	0.1785	-0.0004	-0.0016	0.0014
Secondary Structure Details										
helix: alpha terminal	0.9131	0.8338	0.8182	53.52	0.0000	0.0387	0.1455	0.0009	0.0025	-0.0008
helix: alpha internal	0.9866	0.9734	0.9709	390.75	0.0000	0.0284	0.1305	0.0018	0.0040	-0.0052
helix: 3-10	0.8284	0.6863	0.6569	23.34	0.0000	0.0220	0.0502	-0.0001	0.0012	0.0007
helix: pi	0.4952	0.2452	0.1745	3.47	0.0275	0.0134	0.0065	0.0001	-0.0003	0.0000
helix: left-handed P2										
strand: isolated	0.9069	0.8225	0.8059	49.44	0.0000	0.0207	0.0727	-0.0003	-0.0016	0.0003
strand: parallel	0.4934	0.2434	0.1725	3.43	0.0285	0.0331	0.0188	0.0001	-0.0007	-0.0001
strand: anti-parallel	0.8677	0.7529	0.7297	32.49	0.0000	0.0800	0.1980	-0.0017	-0.0035	0.0007
turn: AA	0.9103	0.8287	0.8126	51.60	0.0000	0.0118	0.0847	0.0000	0.0003	0.0016
turn: Aa	0.8022	0.6435	0.6101	19.25	0.0000	0.0057	0.0080	0.0000	-0.0002	0.0004
turn: aA	0.5965	0.3558	0.2954	5.89	0.0025	0.0095	0.0085	0.0001	-0.0002	-0.0003
turn: AB	0.8365	0.6997	0.6716	24.85	0.0000	0.0045	0.0153	0.0000	-0.0001	0.0004
turn: PG	0.2843	0.0808	-0.0054	0.94	0.4339	0.0176	0.0152	-0.0001	-0.0002	0.0000
turn: PL	0.7867	0.6188	0.5831	17.32	0.0000	0.0168	0.0128	-0.0001	0.0007	0.0007
turn: BL	0.9839	0.9681	0.9651	323.19	0.0000	0.0007	0.0017	0.0000	0.0000	-0.0002
turn: LL										
turn: LG	0.9367	0.8774	0.8659	76.30	0.0000	0.0031	0.0065	0.0000	0.0003	-0.0002
turn: other	0.5702	0.3251	0.2618	5.14	0.0052	0.0408	0.0532	0.0000	-0.0011	0.0006
uncharacterised	0.8006	0.6410	0.6073	19.05	0.0000	0.0392	0.1785	-0.0004	-0.0016	0.0014
Other Contributions										
disulphide pr pos.	0.9063	0.8213	0.8046	49.04	0.0000	0.0021	0.0073	0.0000	0.0000	0.0003
disulphide pr neg.	0.7918	0.6270	0.5920	17.93	0.0000	0.0067	0.0117	0.0000	0.0000	0.0006
tryptophan pr	0.7246	0.5251	0.4805	11.79	0.0000	0.0115	0.0138	-0.0001	0.0005	0.0001
tyrosine pr	0.7068	0.4995	0.4526	10.65	0.0001	0.0107	0.0293	-0.0001	-0.0002	0.0006
phenylalanine pr	0.8331	0.6940	0.6654	24.20	0.0000	0.0059	0.0332	0.0001	-0.0001	-0.0005
porphyrin pr	0.9790	0.9585	0.9546	246.44	0.0000	0.0005	0.0011	0.0000	0.0000	-0.0001
Copper pr	0.4952	0.2452	0.1745	3.47	0.0275	0.0027	0.0013	0.0000	-0.0001	0.0000
Iron pr	0.9790	0.9585	0.9546	246.44	0.0000	0.0005	0.0011	0.0000	0.0000	-0.0001

Table 5.6 Far-UV CD & amide region FT-IR “Super-spectra” Regressions using the first three PCA Factors

Blue bands indicate good regressions. Good regressions seen with Far-UV CD alone or amide region FT-IR alone are highlighted in green and pink respectively.

However, as noted above, the variance coverage of the Factors for the super-spectra suggest that consideration of more than three Factors may be more appropriate in this case, although great care should be taken in interpreting any such results, due to the limited number of proteins in the dataset, which are not fully independent. The results of potential multiple regressions between the first four Factors and the structural features of the proteins are summarised in Table 5.7.

It can be immediately seen that this leads to a dramatic improvement in many of the relationships, with total helix, strand and turns (and hence the remaining “other”) being well related to the super-spectra Factors. In addition, many of the detailed structural elements are better related to the super-spectra.

Structural Feature	Multiparameter Regression Quality						Regression Equation Coefficients				
	Multiple R	Multiple R ²	Adjusted R ²	F(3,32)	p	Standard Error of Estimate	Intercept	B ₁	B ₂	B ₃	B ₄
Basic Secondary Structure											
Helix: total	0.9816	0.9635	0.9588	204.53	0.0000	0.0528	0.3327	0.0026	0.0075	-0.0053	-0.0044
Strand: total	0.9903	0.9808	0.9783	395.75	0.0000	0.0279	0.2895	-0.0019	-0.0058	0.0008	-0.0045
Turn: total	0.9805	0.9614	0.9564	192.91	0.0000	0.0161	0.2058	-0.0002	-0.0004	0.0031	0.0066
Other: total	0.9513	0.9050	0.8927	73.80	0.0000	0.0205	0.1785	-0.0004	-0.0016	0.0014	0.0036
Secondary Structure Details											
helix: alpha terminal	0.9739	0.9484	0.9418	142.56	0.0000	0.0219	0.1455	0.0009	0.0025	-0.0008	-0.0035
helix: alpha internal	0.9948	0.9896	0.9882	736.03	0.0000	0.0180	0.1305	0.0018	0.0040	-0.0052	-0.0024
helix: 3-10	0.8321	0.6923	0.6526	17.44	0.0000	0.0221	0.0502	-0.0001	0.0012	0.0007	0.0003
helix: pi	0.8465	0.7165	0.6799	19.59	0.0000	0.0083	0.0065	0.0001	-0.0003	0.0000	0.0011
helix: left-handed P2											
strand: isolated	0.9339	0.8722	0.8557	52.89	0.0000	0.0178	0.0727	-0.0003	-0.0016	0.0003	0.0012
strand: parallel	0.8564	0.7334	0.6991	21.32	0.0000	0.0199	0.0188	0.0001	-0.0007	-0.0001	0.0029
strand: anti-parallel	0.9970	0.9950	0.9940	1479.12	0.0000	0.0120	0.1980	-0.0017	-0.0035	0.0007	-0.0085
turn: AA	0.9606	0.9227	0.9127	92.50	0.0000	0.0080	0.0847	0.0000	0.0003	0.0016	-0.0009
turn: Aa	0.9732	0.9472	0.9404	139.00	0.0000	0.0022	0.0080	0.0000	-0.0002	0.0004	-0.0006
turn: aA	0.8644	0.7472	0.7146	22.90	0.0000	0.0060	0.0085	0.0001	-0.0002	-0.0003	0.0008
turn: AB	0.9944	0.9888	0.9873	682.20	0.0000	0.0009	0.0153	0.0000	-0.0001	0.0004	0.0005
turn: PG	0.9592	0.9200	0.9096	89.08	0.0000	0.0053	0.0152	-0.0001	-0.0002	0.0000	0.0018
turn: PL	0.9081	0.8246	0.8019	36.43	0.0000	0.0116	0.0128	-0.0001	0.0007	0.0007	0.0013
turn: BL	0.9839	0.9681	0.9639	234.83	0.0000	0.0007	0.0017	0.0000	0.0000	-0.0002	0.0000
turn: LL											0.0000
turn: LG	0.9476	0.8979	0.8847	68.14	0.0000	0.0029	0.0065	0.0000	0.0003	-0.0002	0.0001
turn: other	0.8807	0.7757	0.7467	26.80	0.0000	0.0239	0.0532	0.0000	-0.0011	0.0006	0.0036
uncharacterised	0.9513	0.9050	0.8927	73.80	0.0000	0.0205	0.1785	-0.0004	-0.0016	0.0014	0.0036
Other Contributions											
disulphide pr pos.	0.9814	0.9631	0.9584	202.50	0.0000	0.0010	0.0073	0.0000	0.0000	0.0003	0.0002
disulphide pr neg.	0.9758	0.9522	0.9461	154.48	0.0000	0.0024	0.0117	0.0000	0.0000	0.0006	-0.0007
tryptophan pr	0.9140	0.8354	0.8141	39.33	0.0000	0.0069	0.0138	-0.0001	0.0005	0.0001	0.0010
tyrosine pr	0.9940	0.9881	0.9866	644.64	0.0000	0.0017	0.0293	-0.0001	-0.0002	0.0006	-0.0011
phenylalanine pr	0.9258	0.8571	0.8386	46.47	0.0000	0.0041	0.0332	0.0001	-0.0001	-0.0005	0.0005
porphyrin pr	0.9990	0.9980	0.9980	3556.27	0.0000	0.0000	0.0011	0.0000	0.0000	-0.0001	-0.0001
Copper pr	0.8465	0.7165	0.6799	19.59	0.0000	0.0017	0.0013	0.0000	-0.0001	0.0000	0.0002
Iron pr	0.9990	0.9980	0.9980	3556.27	0.0000	0.0000	0.0011	0.0000	0.0000	-0.0001	-0.0001

Table 5.7 Far-UV CD & amide region FT-IR “Super-spectra” Regressions using the first four PCA Factors

Blue bands indicate good regressions. Good regressions seen with Far-UV CD alone or amide region FT-IR alone are highlighted in green and pink respectively.

It is therefore suggestive that the super-spectra approach is a valuable avenue for enhancing the characterisation of protein secondary structure, with the proviso that this needs to be confirmed with a substantially larger data set of proteins.

5.4.6 Conclusions

Experience with the limited protein data set of this study suggests that super-spectra created by concatenating Far-UV CD and amide region FT-IR spectra, with appropriate scaling to give corresponding magnitudes, and re-digitisation to give similar numbers of data points for each spectral type, can be fruitfully used to derive relationships with the main secondary structures (helix, strand, turns and “other”). However, it is important that this be regarded as a preliminary result that warrants further confirmation with a larger data set of proteins with a wide variety of secondary structures, as the regressions described above are tantamount to fitting a five-parameter equation to six data points.

6 Summary

There is an established requirement for techniques to complement circular dichroism as tools in the quality control of biopharmaceuticals. Therefore, we reviewed the available alternative techniques for obtaining information about protein structure.

Alternative techniques were divided into “high resolution” techniques, which can provide atomic resolution structures of proteins, and “low resolution” techniques, which provide less detailed structural information. The “high resolution” techniques are more suited to deriving reference structural data, rather than quality control applications, although cost reductions in NMR may make it a viable approach in the future. Several “low resolution” techniques were discussed. Fluorescence spectroscopy and UV-Vis spectroscopy do not typically return sufficiently detailed information, although UV-Vis is an important technique for measuring protein concentrations. Raman spectroscopy is not particularly sensitive, and suffers from fluorescence background problems. Surface-enhanced Raman is not non-destructive and insufficiently reproducible. Two techniques that can provide information about chirality, Vibrational CD and Raman Optical Activity, have great power but current instruments are too slow to make this a practical QC technique. FTIR spectroscopy, however, is a well-established technique that offers excellent complementarity to CD, and is relatively rapid and cheap to perform.

The chief difficulty in applying FTIR as a tool in biopharmaceutical quality control is the absorption due to water. In a quality control context, the proteins will inevitably be dissolved in aqueous solutions⁹. This means that short path lengths and high protein concentrations are required. The former is achieved by use of the ATR sampling method, or cells with very short path lengths. The latter is not usually a concern since biopharmaceuticals are usually at very high concentrations. Even when using very short path lengths, however, the water absorption dominates the spectrum, and great care is required when subtracting it. Various bands in the FTIR spectrum of proteins contain structural information, but the best understood is the Amide I band, which is therefore the one that is usually used. This consists of a number of overlapping peaks corresponding to different secondary structures, and therefore the method of data analysis used is critical to obtaining accurate and reproducible results.

With these concerns in mind, we set out to evaluate the use of FTIR to measure protein structure. We chose to investigate the ATR sampling method as most suitable for routine use. We identified a number of measurement issues that potential users should be aware of, and are able to recommend measures to avoid or minimise these difficulties (Table 6.1).

We also applied chemometric methods to a limited dataset of CD and FTIR spectra, to compare the utility of the two methods and to identify any complementarity between them. We found that Near UV CD and Far UV CD both gave very clear discrimination between the different proteins. The Far UV spectra showed excellent correlations with alpha helix, isolated beta strand, BL-type turns and LG-type turns. The near UV region provides information on the amino acid side chains, rather than the peptide backbone. The FTIR spectra did provide good discrimination between the proteins, although they showed much greater variability than the CD spectra. They also showed good correlations with various helices, beta strand and several types of turns, and therefore return some information which is apparently complementary to that obtained by CD. To explore this, we investigated the utility of combining the Far UV CD spectra with the Amide I FTIR spectra. We found that this appeared to give better structural information than either technique alone. However, this needs to be confirmed with a larger data set of proteins with a wide range of secondary structures.

⁹ Biopharmaceutical products are also likely to contain excipients, which may also interfere with FTIR analysis. This issue was not examined in the present study.

Finally, it is clear that FTIR is a technique with great potential to supplement CD in biopharmaceutical quality control and related applications. However, a number of problems with the reproducibility of the technique mean that it should only be used in combination with CD and with considerable care. However, this combination of the two techniques can yield more detailed information about protein structure.

Issue	Recommendations
1. The responsivity of the detector was found to vary rapidly with time, linked to the replenishment of the liquid nitrogen coolant. This may be caused by a number of factors, including the formation of ice.	• Background, buffer, and sample spectra should be collected as close together as possible • Detector dewars should regularly be checked for problems and evacuated or baked, as appropriate and in accordance with manufacturer's recommendations
2. Correct subtraction of the water/buffer spectrum is difficult, as several factors can cause changes in the water spectrum between the buffer solution and the protein sample. Incorrect subtraction can lead to distortions in spectral shape and therefore interpretation of structure.	• Objective methods should be used when scaling the water spectrum before subtraction • The buffer and protein spectra should be collected using identical conditions
3. With ATR sampling, adsorption of protein onto the crystal surface can bias results. This will be influenced by the protein, buffer and crystal material. It is a particular concern where adsorption causes a structural change in the protein.	• Users should perform a control to check whether this is a concern with their particular samples • If possible, the conditions should be chosen so as to minimise protein adsorption • Alternatively, a transmission cell may be used
4. Protein may remain bound to the crystal after the experiment	• Careful cleaning of the crystal is critical. Test spectra may be run to check for protein contamination.

Table 6.1 *Issues and recommendations in FTIR analysis of protein structure*

7 Acknowledgements

We would like to thank the following for their support for this project:

- Warwick Analytical Services
- M-Scan
- Lonza
- Bruker Optics
- Jasco (UK)
- EBI
- NIBSC
- GlaxoSmithKline

Finally, we would like to thank the DTI's National Measurement System for funding this project.

8 Appendix: ATR Theory

This appendix reviews some of the important theoretical aspects of the ATR sampling method, and is based upon a review by Harrick [36]. We conclude with some example calculations, which relate to the experimental work carried out under this study.

8.1 Theory

ATR relies upon the phenomenon of total internal reflection, where light impinging on an interface between two media from the denser medium is completely reflected when its angle of incidence Θ (measured from the normal) is greater than the critical angle, Θ_c , which is given by:

$$\Theta_c = \sin^{-1} n_{21}$$

Equation 4

Where $n_{21}=n_2/n_1$ and n_1 and n_2 are the refractive indices of the first (denser) and second (rarer) media respectively. (In the ATR context, the first medium is the ATR crystal and the second, the sample solution). Although the radiation is completely reflected, an oscillating electric field, known as the evanescent wave, penetrates a short distance into the second medium. In fact the field decays exponentially, and its amplitude E at a depth Z is given by:

$$E = E_0 e^{-\gamma Z}$$

Equation 5

Where E_0 is the amplitude at the interface, and γ is a decay coefficient. The reciprocal of γ is referred to as the penetration depth, d_p , and is the depth at which the amplitude falls to $1/e$ times the value at the interface:

$$E = E_0 e^{-1}$$

Equation 6

and therefore:

$$\gamma Z = 1$$

Equation 7

And so:

$$d_p = Z = \frac{1}{\gamma}$$

Equation 8

However, the penetration depth is not the same as the effective path length of the ATR device, in part because the field decays exponentially away from the surface. The effective path length is defined as the thickness of medium that would give the same absorption in a transmission cell. The two are related by the following equation:

Equation 9

It is immediately apparent that the effective path length is dependent on the refractive indices of both media, the angle of incidence, and the electric field amplitude at the interface, as well as the

penetration depth. However, the electric field amplitude itself is dependent on the polarisation as well as refractive index and angle of incidence. Furthermore, the penetration depth is given by:

$$d_p = \frac{\lambda_1}{2\pi\sqrt{\sin^2 \Theta - n_{21}^2}}$$

Equation 10

Where λ_1 is the wavelength of the light in the first medium (where $\lambda_1 = \lambda/n_1$, λ being the wavelength in a vacuum). Again, it is apparent that the penetration depth is dependent on angle of incidence and refractive index, as well as wavelength.

Putting all this together, for perpendicular polarisation (perpendicular to the plane of incidence):

$$d_{e\perp} = \frac{n_{21}\lambda_1 \cos \Theta}{\pi(1 - n_{21}^2)\sqrt{(\sin^2 \Theta - n_{21}^2)}}$$

Equation 11

And for parallel polarisation:

$$d_{e\parallel} = \frac{n_{21}\lambda_1(2 \cdot \sin^2 \Theta - n_{21}^2)\cos \Theta}{\pi(1 - n_{21}^2)[(1 + n_{21}^2)\sin^2 \Theta - n_{21}^2]\sqrt{(\sin^2 \Theta - n_{21}^2)}}$$

Equation 12

And for unpolarised radiation,

$$d_{eu} = \frac{d_{e\parallel} + d_{e\perp}}{2}$$

Equation 13

In addition, in most ATR designs, there are multiple reflections at the surface of the crystal (Figure 3.3). The effective path length equations above give the path length *per reflection*, therefore to get the total effective path length for a device, we must multiply by the number of reflections:

$$d_t = Nd_e$$

Equation 14

Where N is the number of reflections, d_e is the effective path length per reflection, and d_t is the total effective path length. It is worth noting that the equations presented above make a number of simplifying assumptions – for example, that the absorption in the second medium is negligible and that the refractive indices do not change with wavelength. However, they are a sufficiently good approximation to illustrate the main points.

8.2 Conclusions

- The effective path length, and thereby the absorbance, are linearly related to the wavelength of the incident light (Equation 10, Equation 11). Therefore ATR spectra are not directly comparable to transmission spectra, where the path length is constant for different wavelengths. The heights of the peaks require a simple linear intensity correction, which is typically available in instrument manufacturer's software. However, there are also changes to the peak shape, which are more difficult to correct for.
- The effective path length is also influenced by the refractive indices of the media (Equation 10, Equation 11). For example, if the refractive index of the sample, n_2 , increases, n_{21} also increases

and so does d_e . This becomes especially important where, for example, a buffer spectrum is to be subtracted from a protein spectrum, as concentrated protein solutions typically have a higher refractive index than the buffer they are made in.

- Although the angle of incidence, θ , number of reflections, N , and the refractive index of the first medium also have a strong influence on d_e or d_t , in a commercial instrument, these are typically fixed and do not affect the reproducibility of data.
- Polarisation of the incident light also has a strong influence on d_e , but as unpolarised light is typically used for protein solution studies, this is not a significant factor in this context. In some applications, however, polarised light is used and this should be borne in mind.
- Because of the exponential decay of the evanescent wave (Equation 5), protein nearer to the surface, and particularly deposited on the surface, will be disproportionately represented in measured spectra. This may cause reproducibility problems. Furthermore, if the protein adsorbed on the crystal undergoes a structural change, this will bias the results of structural analysis. Additionally, since the amplitudes of the different components of the electric field differ, even for unpolarised light, artefacts could be caused if the proteins become oriented on the surface.

8.3 Example Calculations

For the instrument used in this study, we can calculate our effective path length as an illustration. The crystal in the ATR is coated with silicon, refractive index 3.45. The optics are set up to give 8 reflections with a 45° ($\pi/4$ rad) angle of incidence. Assuming a solution refractive index of 1.333 and a wavenumber of 1000 cm^{-1} , we calculate an effective path length of $6.0041\text{ }\mu\text{m}$.

How strong is the influence of refractive index? In a typical case of a 10 mg/ml protein solution, we can calculate the expected refractive index. The accepted literature value for dn/dc , the refractive index increment, is $0.186\text{ cm}^3/\text{g}$ for proteins [71]. Therefore we would expect the refractive index of a 10 mg/ml solution in the previous buffer to be approximately 1.3349 (about a 0.14% increase). This would yield, under the same conditions, a total effective path length of $6.0190\text{ }\mu\text{m}$, or a 0.25% increase. This seems tiny, until one remembers that the difference between the Amide I band and the water background in a protein spectrum is about 6% (see Figure 4.8 and Figure 4.9). It is clear, then, that there can be a significant difference in effective path length between a protein solution and the corresponding buffer.

9 References

1. Nguyen, L.T., J.M. Wiencek, and L.E. Kirsch, *Characterization methods for the physical stability of biopharmaceuticals*. PDA J Pharm Sci Technol, 2003. **57**(6): p. 429-45.
2. Ascoli, G.A., P. Pergami, K.X. Luu, D.L. Alkon, E. Bramanti, C. Bertucci, L. Di Bari, and P. Salvadori, *Use of CD and FT-IR to determine the secondary structure of purified proteins in the low-microgram range*. Enantiomer, 1998. **3**(4-5): p. 371-81.
3. *Specifications: Test Procedures And Acceptance Criteria for Biotechnological/Biological Products (Q6B)*. 1999, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. p. 1-16.
4. Oberg, K.A., J.M. Ruyschaert, and E. Goormaghtigh, *The optimization of protein secondary structure determination with infrared and circular dichroism spectra*. Eur J Biochem, 2004. **271**(14): p. 2937-48.
5. Abola, E., P. Kuhn, T. Earnest, and R.C. Stevens, *Automation of X-ray crystallography*. Nat Struct Biol, 2000. **7 Suppl**: p. 973-7.
6. Goodwill, K.E., Tennant, M. G., Stevens, R. C., *High-throughput x-ray crystallography for structure-based drug design*. DDT, 2001. **6**(15): p. 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*, H.A. Havel, Editor. 1996, John Wiley and Sons. p. 190-246.
8. Price, N.C., *Conformational issues in the characterization of proteins*. Biotechnol Appl Biochem, 2000. **31 (Pt 1)**: p. 29-40.
9. Pelton, J.T. and L.R. McLean, *Spectroscopic methods for analysis of protein secondary structure*. Anal Biochem, 2000. **277**(2): p. 167-76.
10. Wider, G., and Wuthrich, K., *NMR spectroscopy of large molecules and multimolecular assemblies in solution*. Current Opinion in Structural Biology, 1999. **9**(5): p. 594-601.
11. Clore, G.M. and A.M. Gronenborn, *Determining the structures of large proteins and protein complexes by NMR*. Trends Biotechnol, 1998. **16**(1): p. 22-34.
12. Grzesiek, S., *Characterisation of biotechnology pharmaceutical products*. Structural characterisation of proteins by NMR, ed. F. Brown, Lubiniecki, A., and Murano, G. Vol. 96. 1998: International association of biological standardisation.
13. Rodger, A., and Norden, B., *Circular dichroism and linear dichroism*. 1997: Oxford University Press.
14. Mulkerrin, M.G., *Protein structure analysis using circular dichroism*, in *Spectroscopic Methods for determining protein structure in solution*, H.A. Havel, Editor. 1996, John Wiley & Sons. p. 5-27.
15. Havel, H.A., *Derivative near-ultraviolet absorption techniques for investigating protein structure*, in *Spectroscopic methods for determining protein structure in solution*, H.A. Havel, Editor. 1996, John Wiley & Sons. p. 62-68.
16. Kelly, S.M. and N.C. Price, *The use of circular dichroism in the investigation of protein structure and function*. Curr Protein Pept Sci, 2000. **1**(4): p. 349-84.

17. Drake, A.F., *Protein-ligand interactions: Structure and spectroscopy-A practical approach* Circular dichroism, ed. S.E. Harding, and Chowdhry, B. Z. Vol. 2. 2001: Oxford University Press.
18. Greenfield, N.J., *Biomacromolecular Applications of Circular Dichroism and ORD*, in *Encyclopedia of Spectroscopy and Spectrometry*, J.C. Lindon, G.E. Tranter, and J.L. Holmes, Editors. 2000, Academic Press. p. 117-130.
19. Lobley, A., L. Whitmore, and B.A. Wallace, *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra*. *Bioinformatics*, 2002. **18**(1): p. 211-2.
20. Whitmore, L. and B.A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data*. *Nucleic Acids Res*, 2004. **32**(Web Server issue): p. W668-73.
21. Oberg, K.A., J.-M. Ruyschaert, and E. Goormaghtigh, *Rationally selected basis proteins: A new approach to selecting proteins for spectroscopic secondary structure analysis*. *Protein Sci*, 2003. **12**(9): p. 2015-2031.
22. Jones, C., D. Schiffmann, A. Knight, and S. Windsor, *Val-CiD Best Practice Guide: CD spectroscopy for the quality control of biopharmaceuticals*. 2004, National Physical Laboratory.
23. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. 2nd ed. 1999, New York: Kluwer Academic/Plenum Publishers.
24. Ladokhin, A.S., *Fluorescence Spectroscopy in Peptide and Protein Analysis*. *Encyclopedia of Analytical Chemistry*, 2000.
25. Tensmeyer, L.G., and Kauffman, E. W., *Protein structure as revealed by nonresonance Raman spectroscopy*, in *Spectroscopic methods for determining protein structure in solution*, H.A. Havel, Editor. 1996, John Wiley & sons. p. 69-95.
26. Raman, C.V., and Krishnan, K. S., *A new type of secondary radiation*. *Nature*, 1928. **121**(3048): p. 501.
27. Tantra, R., A. McCabe, M. Bailey, A. Knight, and E. Smith, *Comparable Raman Spectroscopy Across Different Instruments and Excitation Wavelengths*. 2004, National Physical Laboratory: Teddington. p. 29.
28. Keiderling, T.A., *Protein structural studies using vibrational circular dichroism spectroscopy*, in *Spectroscopic methods for determining protein structure in solution*, H.A. Havel, Editor. 1996, John Wiley & Sons. p. 163-189.
29. Barron, L.D., Hecht, L., Bell, A. F., and Wilson, G., *Recent developments in Raman optical activity of biopolymers*. *Applied spectroscopy*, 1996. **50**(5): p. 619-629.
30. Barron, L.D., M.P. Bogaard, and A.D. Buckingham, *Raman scattering of circularly polarized light by optically active molecules*. *Journal of the American Chemical Society*, 1973. **95**(2): p. 603-605.
31. Zhu, F.J., Isaacs, N. W., Tranter, G. E., Barron, L. D., *Raman optical activity of proteins, carbohydrates and glycoproteins*. *Chirality*, 2006. **18**: p. 103-115.
32. Elliott, A. and E.J. Ambrose, *Structure of synthetic polypeptides*. *Nature*, 1950. **165**(4206): p. 921-2.

33. Surewicz, W.K., and Mantsch, H. H. , *Infrared absorption methods for examining protein structure*, in *Spectroscopic methods for determining protein structure in solution*, H.A. Havel, Editor. 1996, John Wiley & sons. p. 135-162.
34. van de Weert, M., P.I. Haris, W.E. Hennink, and D.J. Crommelin, *Fourier transform infrared spectrometric analysis of protein conformation: effect of sampling method and stress factors*. Anal Biochem, 2001. **297**(2): p. 160-9.
35. Parker, F.S., *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry*. 1983: Plenum press, New York.
36. Mirabella, F.M., Jr and N.J. Harrick, *Correction of Internal Reflection Spectra for the Wavelength Dependence*, in *Internal Reflection Spectroscopy: Review and Supplement*. 1985, Harrick Scientific Corporation: Ossining, NY. p. 84-91.
37. Belton, P.S., *New methods for monitoring changes in proteins*. Food Reviews International, 1993. **9**(4): p. 551-573.
38. Surewicz, W.K., H.H. Mantsch, and D. Chapman, *Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment*. Biochemistry, 1993. **32**(2): p. 389-94.
39. Spragg, R.A., *IR Spectrometers*, in *Encyclopedia of Spectroscopy and Spectrometry*, G. Tranter, J. Holmes, and J. Lindon, Editors. 1999, Academic Press. p. 1048-1057.
40. Sherman Hsu, C.P., *Infrared Spectroscopy*, in *Handbook of instrument techniques for analytical chemistry*, F. Settle, Editor. 2000, Pearson education. p. 247-283.
41. Dong, A., Huang, p., and Caughey, W. S., *Protein secondary structures in water from second-derivative amide I infrared spectra* Biochemistry, 1990. **29**: p. 3303-3308.
42. Haris, P.I. and F. Severcan, *FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media*. Journal of Molecular Catalysis B-Enzymatic, 1999. **7**(1-4): p. 207-221.
43. Werner Hug, G.H., *A novel high-throughput Raman spectrometer for polarization difference measurements*. Journal of Raman Spectroscopy, 1999. **30**(9): p. 841-852.
44. Spragg, R.A., *IR spectroscopy and sample preparation methods*, in *Encyclopedia of Spectroscopy and Spectrometry*, G. Tranter, Holmes, J., and Lindon, J, Editor. 1999, Academic Press. p. 1058-1066.
45. Grdadolnik, J., *ATR-FTIR spectroscopy:its advantages and limitations*. Acta Chim. Slov 2002. **49**: p. 631-642.
46. Belali, R., Vigoureux, J, M., and Morvan, J., *Dispersion effects on infrared spectra in attenuated total reflection*. J. Opt. Soc. Am. B, 1995. **12**: p. 2377-2381.
47. Grdadolnik, J., *ATR-FTIR spectroscopy:its advantages and limitations*. Acta Chim. Slov 8 2002. **49**: p. 631-642.
48. Sperline, R.P., Muralidharan, S., and Freiser, H., *New quantitative technique for ATR spectrometry; Calibration of the "CIRCLE" ATR device in the infrared*. Applied spectroscopy, 1986. **40**: p. 1019-1022.
49. Jackson, M. and H.H. Mantsch, *The use and misuse of FTIR spectroscopy in the determination of protein structure*. Crit Rev Biochem Mol Biol, 1995. **30**(2): p. 95-120.

50. Elliott, A., and Ambrose, E. J., *Structure of synthetic polypeptides*. Nature, 1950. **165**: p. 921-2.
51. Jabs, A. *Determination of secondary structure in proteins by Fourier transform infrared spectroscopy*. 2005 [cited 2005 03rd of August].
52. Susi, H., *Infrared spectroscopy-conformation*. Methods in enzymology, 1972. **26**: p. 455-472.
53. Haris, P.I., and Chapman, D. , *Analysis of polypeptide and protein structures using Fourier transform infrared spectroscopy*. Methods in molecular biology: microscopy, optical spectroscopy, and macroscopic techniques, ed. C. Jones, Mulloy, B., and Thomas, A.H. Vol. 22. 1994.
54. Arrondo, J.L.R., and Goni, F.M., *Structure and dynamics of membrane proteins as studied by infrared spectroscopy*. Progress in biophysics & molecular biology, 1999. **72**: p. 367-405.
55. Nevskaya, N.A., and Chirgadze, Yu, N., *Infrared spectra and resonance interactions of amide I and II vibrations of alpha helix*. Biopolymers, 1976. **15**: p. 637-648.
56. Natalello, A., Ami, D., Brocca, S., Lotti, M., and Doglia, S. M., *Secondary structure, conformational stability and glycosylation of a recombinant Candida rugosa lipase studied by FTIR spectroscopy*, in *Biochemical journal immediate publication* 2004. p. 1-20.
57. vanyaminov, S.Y., and Kalnin, N. N. , *Quantitative IR spectrophotometry of peptide compounds in water solutions. I. spectral parameters of amino acid residue absorption bands*. Biopolymers, 1990. **30**: p. 1243-1257.
58. Byler, D.M. and H. Susi, *Examination of the secondary structure of proteins by deconvolved FTIR spectra*. Biopolymers, 1986. **25**(3): p. 469-87.
59. Susi, H., and Byler, D. M., *Resolution-enhanced fourier transform infrared spectroscopy of enzymes*. Methods in enzymology, 1986. **130**: p. 290-311.
60. Rahmelow, K., and Hubner, W., *Infrared spectroscopy in aqueous solution: Difficulties and accuracy of water subtraction*. Applied spectroscopy, 1997. **51**(2): p. 160-170.
61. Dousseau, D., Therrien, M., and Pezolet, M., *On the spectral subtraction of water from the FTIR spectra of aqueous solutions of proteins*. Applied spectroscopy, 1989. **43**(3): p. 538-542.
62. Powell, J.R., Wasacz, F. M., and Jakobsen, R.J., *An algorithm for the reproducible spectral subtraction of water from the FTIR spectra of proteins in dilute solutions and adsorbed monolayers*. Applied spectroscopy, 1986. **40**(3): p. 339-344.
63. Venyaminov, S.Y. and J.T. Yang, *Determination of Protein Secondary Structure*, in *Circular Dichroism and the Conformational Analysis of Biomolecules*, G.D. Fasman, Editor. 1996, Plenum Press: New York. p. 69-107.
64. Arrondo, J.L., A. Muga, J. Castresana, and F.M. Goni, *Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy*. Prog Biophys Mol Biol, 1993. **59**(1): p. 23-56.
65. Venyaminov, S.Y. and F.G. Prendergast, *Water (H₂O and D₂O) Molar Absorptivity in the 1000-4000 cm⁻¹ Range and Quantitative Infrared Spectroscopy of Aqueous Solutions*. Analytical Biochemistry, 1997. **248**(2): p. 234-245.
66. Noble. J.E., K.A.E., Reason. A.J., Di Matola. A., and Bailey. M.J.A., *A comparison of protein quatitation assays for biopharmaceutical applications*. Molecular biotechnology 2007. **in press**.

67. Goldberg, M.E. and A.F. Chaffotte, *Undistorted structural analysis of soluble proteins by attenuated total reflectance infrared spectroscopy*. Protein Sci, 2005. **14**(11): p. 2781-2792.
68. Theocharous, E., *Drift in cryogenically cooled photodetectors: causes and cures*. Photonics, 2006. **40**(4): p. 70-72.
69. Khajehpour, M., J.L. Dashnau, and J.M. Vanderkooi, *Infrared spectroscopy used to evaluate glycosylation of proteins*. Anal Biochem, 2006. **348**(1): p. 40-8.
70. Khajehpour, M., J.L. Dashnau, and J.M. Vanderkooi, *Infrared spectroscopy used to evaluate glycosylation of proteins*. Anal Biochem, 2006. **348**(1): p. 40-8Also the various degrees of glycosylation of proteins can be studied using FTIR bands. This suggests that FTIR can be used to analyse the posttranslational
71. J. A. De Feijter, J.B.F.A.V., *Ellipsometry as a tool to study the adsorption behavior of synthetic and biopolymers at the air-water interface*. Biopolymers, 1978. **17**(7): p. 1759-1772.