Structural Investigations of Nematode Anhydrobiotic and Rab11-FIP Proteins by Far-UV Circular Dichroism and Mid-IR Fourier Transform Infra-Red Spectroscopy

Patrick Lall, B. Sc.

Thesis submitted to the National University of Ireland in fulfilment of the requirements for the Degree of Doctor of Philosophy

NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Department of Chemistry
N.U.I. Maynooth
January 2009

Head of Department: Professor John P. Lowry
Supervisor: Dr. John G. McCaffrey

In memory of my mother

Mary Angela Lall

Acknowledgements

I would like to thank my supervisor Dr. John McCaffrey for his assistance, advice and patience during the course of this PhD thesis.

I would like to thank Prof. Ann Burnell for her assistant supervision and extended patience throughout our collaboration.

A special thanks must go to both Dr. Jackie Boyde and Bridget Culleton for their never-ending production of protein over the past number of years.

Thanks are also owed to Dr. Mary McCaffrey and Dr. Arek Oleksy for supplying the material for the Rab11-FIP3 work and Dr. Amir Kahn for supplying the Rab11-FIP2 protein samples.

A large debt of gratitude is owed to Mr. Noel Williams for his continuous assistance with all manner of computer and equipment related difficulties.

I would like also to thank my lab colleagues Dr. Maryanne Ryan, Dr. Martin Collier, Owen Byrne and Ciaran Murray for their help and company over the years.

I wish to thank all the staff of the Chemistry Department, especially Dr. Ken Maddock, for their help during my time at Maynooth.

I would like to thank all the Department postgraduates of whom Gillian, Catherine, Claire, Linda, Declan, Rob and Denis are but a few.

Lastly, thanks goes to Billy and Derek for their friendship over the years.

Table of contents

Part A: Rab11-FIP3

Part B: Rab11-FIP2

Abstract

Analysis of the secondary structures of the standard proteins myoglobin, hemoglobin, lysozyme and α-chymotrypsin by CD spectroscopy resulted in excellent agreement of estimated fractional composition with that observed by x-ray crystallography. CD spectroscopy was found to be very accurate for estimating α-helix and random coil content, but considerably less so for estimating β-sheet. Conversely, FTIR spectroscopic analysis of the same proteins proved to be more accurate for β-sheet estimation than either α-helix or random coil. Overall, CD spectroscopy was found to be superior to FTIR spectroscopy for the quantitative analysis of protein secondary structure, however, the two techniques are highly complimentary for protein structural studies. Both CD and FTIR methods were found to be useful for protein stability studies, where CD is suited to analysing α-helix stability through CD_{222} protein melts, and FTIR is capable of investigating protein aggregation phenomenon (together with 2D correlation spectroscopy) and β-sheet stability.

The anhydrobiotic AavLEA1 nematode protein was discovered to be a natively unfolded protein with an extended tertiary conformation. From FTIR melt experiments this protein was shown to resist temperature-induced aggregation and to act synergistically with the trehalose disaccharide in retarding the aggregation of cytochrome c. AavLEA1 was also shown to be capable of producing protein fibrils by adopting a completely helical structure in the presence of high concentrations of TFE. Additionally, titration with Ca^{2+} resulted in an increase of ordered secondary structure and demonstrated AavLEA1's ability to sequester cations.

A second nematode anhydrobiotic protein (P. sup DJ-1) was found to be structurally very similar to the human DJ-1 homolog and exist as an α/β-mixed β-sandwich protein. Stability studies revealed that P. sup DJ-1 showed considerable temperature stability, both in terms of its helix domains and its resistance to aggregation. The structure of P. sup DJ-1 was found to be unaffected by high concentrations of H_2O_2, as was its temperature stability, leading to the conclusion that this protein functions as a molecular chaperone to relieve oxidative and/or heat stress, similar to current opinions on the functionality of the human homolog.[1]

The Rab11-FIP3 homo-dimer predicted coiled coil protein fragment was confirmed by CD spectroscopy and TFE titration. For the Rab11-FIP2 homo-dimer, mutation of the valine residue 456 and the leucine residue 457 to glycine residues was found to destabilise the mutant relative to the wild-type. As such, these residues are concluded to be of central importance in the formation of the coiled-coil cap that protects the hydrophobic core from the aqueous environment.

References:

1 Shendelman S, Jonason A, Martinat C, Leete T, Abeliovich A. DJ-1 is a redox-dependent molecular chaperone that inhibits alpha-synuclein aggregate formation. Plos Biology 2(11):1764-1773, 2004.

Chapter I

Introduction

As a compliment to the discovery of new anhydrobiotic genes in nematodes it was desirable to develop a spectroscopic suite through which structural analysis of the proteins encoded by these genes could be performed. Because there was a need to examine the solution behaviour of protein structure, the spectroscopic techniques of circular dichroism (CD) and Fourier-transform infrared (FTIR) analyses were chosen as the most powerful methods available. In general, CD and FTIR spectroscopic protein structural analysis can be divided into those methods designed to study the secondary structures of proteins and those capable of investigating their tertiary structures. The spectrometer to be used for each type of analysis must be optimised for that particular spectroscopic technique and, therefore, it was decided to initially focus on designing a system capable of investigating the secondary structures of proteins. Therefore, far-UV CD and mid-IR FTIR spectrometers were selected as the primary tools for analysing protein secondary structure content.

The study of protein structure by either CD or FTIR spectroscopy is a more global and less specific approach than the more involved and exhaustive methods of nuclear magnetic resonance (NMR) or x-ray crystallography spectroscopy. While both the latter techniques are capable of furnishing the precise molecular coordinates of a protein molecule (and, hence, a detailed description of both its secondary and tertiary structure), they nonetheless possess limitations that are important to the investigation of protein molecules. For example, neither technique is capable of studying a protein’s secondary structure as a dynamic system that responds to changes in its physical environment, but instead treat the protein as a static rigid structure. This is acceptable when only the absolute structure of a protein is required, but disqualifies both NMR and x-ray crystallography techniques from being able to examine proteins for structural variation. Because both CD and FTIR techniques can be rapidly performed on proteins in solution, they are ideal for studying their structural behaviour over a range of varying physiological conditions - a property that is of great interest to the biological characterisation of proteins in general.

The stringency of each spectroscopic technique in determining the conditions that the protein sample must fulfil, coupled with the difficulty for the biologist to achieve these conditions without compromising the protein's structure, represents a considerable barrier to structural investigations. For x-ray crystallography the biggest challenge is to crystallise the protein and in proteins that contain significant random coil content this is very often impossible. Analysis of proteins by NMR requires very large amounts of protein in high concentration. In many cases it is difficult for the protein biologist to generate the amount of protein required and in such instances, analysis by NMR cannot be achieved. Where protein synthesis is difficult CD is of particular use, since the sensitivity of the CD technique is such that it requires only approximately 0.04 mg of protein to obtain its CD spectrum. CD analysis is somewhat restricted, however, in that it requires the protein sample to be free from contaminating detergents and other artefacts and is selective in terms of what buffer solutions may be used. Where such restrictions prevent analysis of a particular protein by CD, it is likely that the same protein sample may be analysed by FTIR – although FTIR analysis requires considerably more protein. In practical terms, therefore, protein CD and FTIR are highly complimentary techniques.

Biofluorescence is a phenomenon that can be used to measure distances between chromophores via the fluorescent resonance energy transfer technique (FRET) on a scale that is applicable to protein molecules. In the case of the AavLEA1 protein, time-resolved biofluorescence was used to further investigate its solution-phase protein structure, using a pulsed laser light source. Similar to CD and FTIR, FRET is a dynamic technique and it is most effective when used to investigate the binding of proteins to other molecules, insofar as energy transfer might occur upon protein-ligand association, but will be absent where no interaction is present. FRET may also be used to investigate the association of protein domains with one another as a function of the folding or unfolding of a protein or with the occurrence of a structural transition.

Preparation of protein samples capable of being analysed by CD, FTIR or FRET within this work, involved considerable effort in the development of protein expression, purification and site-directed mutation (SDM) systems. Protein expression requires the identification of a specific protein gene from within the organism's genome and the engineering of this gene into a vector plasmid. The plasmid is then transformed into an *E. coli* cell and the cell's mechanisms are used to express the protein from the introduced gene. Purification generally relies on a modification of the protein in question at the genetic level, such that it contains an additional amino-acid sequence tag. An example of this is the introduction of a His-tag region, which is composed of repeated histidine residues, at one end of the resulting fusion protein. The His-tag can then be used to purify the protein by affinity chromatography. Investigation of the structural consequences of altering single residues within a protein's primary sequence may be achieved using SDM. Considering that some genetic diseases are the result of a single point mutation within a gene, it is important to compare the structural profile of the native and mutated protein in such cases.

Upon installation of the CD, FTIR and FRET instrumentation the immediate task of validating each technique was necessary to provide confidence in the accuracy of each system. For the CD and FTIR systems a comprehensive analysis of a number of standard commercial proteins was performed. This included a quantitative analysis of each protein's secondary structure by both CD and FTIR methods and the results of these tests could be compared with the known crystal structure of each protein. Protein FTIR melt experiments were also performed for each standard protein in order to examine the protein unfolding process and two dimensional correlation spectroscopy was used to investigate the sequence of this process. The FRET experiment was validated using fluorescently-labelled BSA, whereby FRET was detected by both steady-state and time-resolved luminescence spectroscopy.

An investigation of the secondary structure of the anhydrobiotic AavLEA1 nematode protein from the nematode *Aphelenchus avenae*, which is up-regulated in response to desiccation, was performed using CD and FTIR spectroscopy. A subsequent examination of the possible functioning of AavLEA1 was carried out by monitoring its structural variation in response to a variety of physiological changes and the results of these tests were considered in the context of current models describing the possible roles of similar anhydrobiotic proteins. To this end the tertiary structure of AavLEA1 was also examined using the FRET technique by introducing a single cysteine residue at a specific location via SDM and labelling this cysteine with one of two different fluorophores. The occurrence of energy transfer between the lone tryptophan residue within AavLEA1 and the site-specifically incorporated fluorophore was used to determine their distance apart and from this the degree of extendedness or compactness of the protein backbone was revealed. The role that AavLEA1 might play in stabilising the secondary structures of other proteins was investigated by FTIR melt experiments and the possible synergistic functioning with the disaccharide trehalose was considered in this regard.

A DJ-1 homolog was discovered as a second anhydrobiotic protein that became up-regulated in response to desiccation of the *panagolaimus superbus* nematode. Deleterious mutations within an homologous human DJ-1 have been associated with the genetic condition of early-onset Parkinson’s disease. One of the primary functions of the DJ-1 family of proteins is to combat oxidative stress within cells by acting as a molecular chaperone to preserve the structures of complex protein molecules. The x-ray crystal structure of the human DJ-1 protein has been solved and, in an effort to understand possible structural and functional similarity of the novel nematode DJ-1, its secondary structure was examined by both CD and FTIR methods.

The Rab11-FIP proteins are a group of Rab11 interacting proteins that complex with the Rab11 protein, via a Rab binding domain possessed by all Rab11-FIP’s, to create a protein complex that functions in the trafficking of vesicles to different regions of the cell. The Rab11-FIP2 and Rab11-FIP3 proteins are understood to first self-associate and form homo-dimers before they interact with Rab11 proteins to generate the functional complex. X-ray crystallography has shown that the self-association of the Rab11-FIP2 protein is by way of a coiled-coil interaction. As a confirmation to the nature of Rab11-FIP3 homo-dimer formation, the structure of the predicted coiled-coil homo-dimer protein fragment was examined using CD spectroscopy. For the Rab11-FIP2 protein, for which the coiled-coil interaction has been observed, the importance of the coiled-coil “cap” was examined by performing SDM on residues in this region and then analysing the structures and stabilities of the resulting mutants by both CD and FTIR spectroscopy.

Chapter II

Far-UV Circular Dichroism and Fluorescent Resonance Energy Transfer Spectroscopy

II.1 The Use of Polarised Light in Spectroscopy

II.1.1 The Polarisation of Light

Plane polarised light differs from normal non-polarised light in that the electric vectors of all photons, which describe both the phase and the polarity of the light, are confined to a single *xy*-plane, whereas for non-polarised light a photon's electric vector may have a random direction. For circularly polarised light the electric vector rotates uniformly about the direction of propagation of the light by 2π during each cycle and, when observed opposite to the direction of light propagation, rotation is either clockwise or anticlockwise for right or left circularly polarised light, respectively. Therefore, the electric vector of circularly polarised light traces out either a right- or left-handed helix, depending on whether it is right or left circularly polarised. Figure II.1 illustrates a vector representation of the polarisation character of both plane polarised and circularly polarised light. Circular dichroism (CD) spectroscopy involves stimulating molecules to undergo electronic transitions from lower to higher energy levels using circularly polarised light.

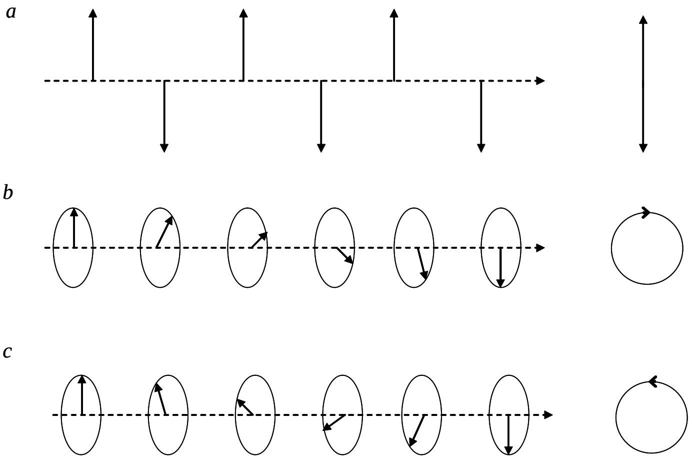


Figure II.1. Illustration of the polarisation of light. *a*, *b* and *c* illustrate plane polarised, right circularly polarised and left circularly polarised light, respectively. The solid arrows show the instantaneous spatial direction of the electric vector, while the dashed arrows depict the propagation direction of the light. The variation in the polarity of the electric vector is shown to the right of each illustration.

II.1.2 Optical Retardation and the Generation of Circularly Polarised Light

Figure II.2 shows that plane polarised light can be decomposed into left and right circularly polarised components that are in-phase and of equal magnitude. When incident at 45^{o} to the optic axis of an optically active crystal plane polarised light can be resolved into two orthogonal components: the extraordinary wave and the ordinary wave. The extraordinary wave lies in the same plane as the original plane polarised light, the so-called optic axis, whereas the ordinary wave lies in a perpendicular plane. A retardation plate affects plane polarised light by selectively retarding the ordinary wave and the degree to which the ordinary wave is retarded depends upon the thickness of the retardation plate. Passing plane polarised light through a λ/2 retardation plate causes a complete offset in the phases of the extraordinary and ordinary waves, with the result that the emerging plane polarised beam has its optic axis rotated by an angle of 90^{o}. If, however, plane polarised light is passed through a λ/4 retardation plate the extraordinary and ordinary waves are offset so that the zero of one wave corresponds with the peak of the other. The result is that plane polarised light passing through a λ/4 retardation plate exits as circularly polarised light.

In arriving at a classical expression for left and right circularly polarised light, one starts by considering the extraordinary and ordinary waves as *x*- and *y*-polarised electric vectors, $\mathbf{E}_{\mathrm{x}}$ and $\mathbf{E}_{\mathrm{y}}$, which propagate along the z direction and are described by the equations

$$\mathbf{E}_x = \mathbf{i}E_0 \cos\left[2\pi\nu\left(t - \frac{nz}{c}\right)\right] \qquad \text{(II.1)}$$

and

$$\mathbf{E}_y = \mathbf{j}E_0 \cos\left[2\pi\nu\left(t - \frac{nz}{c}\right)\right] \qquad \text{(II.2)}$$

where **i** and **j** are unit vectors along the *x* and *y* axes, respectively, ν is the radiation frequency, *t* is time, *n* is the refractive index of the medium and *c* is the velocity of light in vacuo. To simulate the condition for the relative phases of the extraordinary and ordinary waves for circularly polarised light a phase difference of π/2 is introduced to the *y*-polarised electric vector, leading to

$$\mathbf{E}_{\mathrm{y}} = \pm\mathbf{j}E_0 \sin\left[2\pi\nu\left(t - \frac{nz}{c}\right)\right] \qquad \text{(II.3)}$$

The vector addition of Equations II.1 and II.3 gives the required expressions for left and right circularly polarised light, $\mathbf{E}^{+}$ and $\mathbf{E}^{-}$, respectively, where

$$\mathbf{E}^{+} = E_0\left\{\mathbf{i}\cos\left[2\pi\nu\left(t-\frac{nz}{c}\right)\right]+\mathbf{j}\sin\left[2\pi\nu\left(t-\frac{nz}{c}\right)\right]\right\} \quad \text{(II.4)}$$

and

$$\mathbf{E}^{-} = E_0\left\{\mathbf{i}\cos\left[2\pi\nu\left(t-\frac{nz}{c}\right)\right]-\mathbf{j}\sin\left[2\pi\nu\left(t-\frac{nz}{c}\right)\right]\right\} \quad \text{(II.5)}$$

II.1.3 Optical Rotatory Dispersion

Plane polarised light exiting an optically active transparent medium has an optic axis that is non-parallel to the original plane polarised light. This ability of an optically active medium to rotate the plane of plane polarised light depends on it having an unequal refractive index for left and right circularly polarised light, i.e. $n_L \neq n_R$, where n refers to the refractive index. This phenomenon, measured over a range of wavelengths, is referred to as optical rotatory dispersion (ORD) and ORD is commonly used to study the optical activities of materials by characterising them in terms of the angle through which they rotate the optic axis of plane polarised light. The optical activity of a material is quantified by its specific rotation, [α], as

$$[\alpha] = \frac{\alpha}{ld\rho} = \frac{\alpha}{lb'} \quad \text{(II.6)}$$

where α is the observed rotation in degrees, l is the pathlength in decimetres, d is the material density in grams per cubic centimetre (gcm^{-3}), ρ is the fraction per weight of the optically active substance and b' is the concentration in gcm^{-3}. Fresnel's equation for α in radians per unit pathlength, as measured in the same units as λ, is given by

$$\alpha = \frac{(n_L - n_R)\pi}{\lambda} \quad \text{(II.7)}$$

The occurrence of a concentration parameter in Equation II.6 allows an expression for the molar rotation, [M], to be defined as

$$[M] = M_r[\alpha]\times 10^{-2} = \frac{M_r\alpha}{(100lb')} = \frac{100\alpha}{lb} \quad \text{(II.8)}$$

where M_r is the molecular weight of the optically active material, l is the pathlength in cm and b is the concentration in mol.L^{-1}. The molar rotation is then expressed in units of degree square centimetres per decimole (deg cm^2 dmol^{-1}). Figure II.2*a* illustrates the resolution of the plane polarised resultant electric vector into its left and right circularly polarised components. Figure II.2*b* shows the rotation of the plane of resultant plane polarised light by an angle α when circularly polarised light is passed through a transparent dextrorotatory medium.

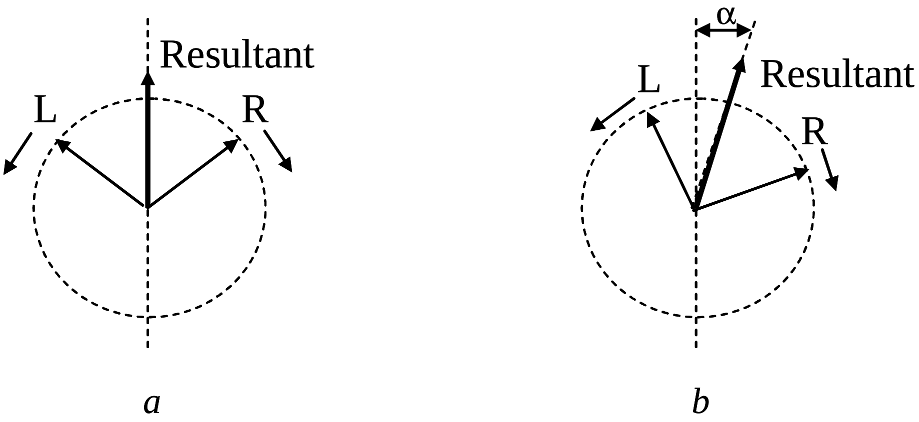


Figure II.2. Illustration of ORD for a dextrorotatory medium. *a* Before entering the dextrorotatory medium the electric vectors of both right and left circularly polarised light combine to produce a plane polarised resultant parallel to the optical axis, as defined by the dotted vertical line. *b* After passing through the dextrorotatory medium both right and left circularly polarised light combine to produce a plane polarised resultant, which is now oriented at an angle α from the optical axis.

II.1.4 Circular Dichroism and Elliptically Polarised Light

Since CD stimulates the upward transitions of electrons from lower to higher energy levels, an absorption process must be involved and, accordingly, circular dichroism is defined as the absorption differential of left versus right circularly polarised light, where dichroism refers to the differential absorption of light polarised in two directions as a function of frequency.

An important correlation exists between an optically active material's refractive index and its absorption extinction coefficient for circularly polarised light, such that performing a Kronig-Kramers transformation on the ORD spectrum of an optically active material returns its absorption spectrum and vice versa. The absorption indices k_L and k_R, for left and right circularly polarised light, respectively, are related to the refractive indices n_L and n_R such that, at wavelengths where an optically active material is transparent, if $n_L > n_R$ then $k_L > k_R$ at wavelengths where the material absorbs. Therefore, optically active materials that are non-transparent in certain wavelength regions necessarily exhibit CD spectra. Optical rotation over such non-transparent regions is known as anomalous ORD.

Figure II.3 illustrates the combined effects of anomalous ORD and CD on circularly polarised light as it exits a dextrorotatory absorbing material. Due to the greater absorption of left over right circularly polarised light, the resultant electric vector is no longer composed of equal amounts of left and right circularly polarised components and this inequality gives rise to a resultant that is no longer plane polarised, but instead is elliptically polarised. The angle α between the major axis of the ellipse and the plane of the incident polarised light is the effect of anomalous ORD, whereas the angle ψ is the material's ellipticity, which is defined as the tangent of the ratio of the major and minor axes of the ellipse described by the resultant of the exiting elliptically polarised light, as follows

$$\tan\psi = \frac{(E_R - E_L)}{(E_R + E_L)} \qquad \text{(II.9)}$$

where E_L and E_R are the magnitudes of the left and right circularly polarised components of the elliptically polarised light. The major and minor ellipse axes correspond to the sum and difference in the amplitudes of the two circularly polarised components on emerging from the material, respectively. For the purpose of clarity, the magnitude of the elliptically polarising effect of CD has been greatly exaggerated with respect to the anomalous ORD effect in Figure II.3.

The amplitude of the electric vector over a unit pathlength decreases as the light passes through the absorbing material according to

$$E = E_0 \exp\left(\frac{-2\pi k}{\lambda}\right) \qquad \text{(II.10)}$$

Since both (k_L – k_R) and ψ are always very small, approximating $\tan\psi$ as ψ and considering only the first two terms in the exponential of Equation II.10 is reasonable, with the result that from Equations II.9 and II.10 the following analogous expression arises for the ellipticity, in terms of the absorption indices, k_L and k_R, as to that of Fresnel's equation II.7 for optical rotation, α, which is expressed in terms of the refractive indices, n_L and n_R;

$$\psi = \frac{(k_L - k_R)\pi}{\lambda} \tag{II.11}$$

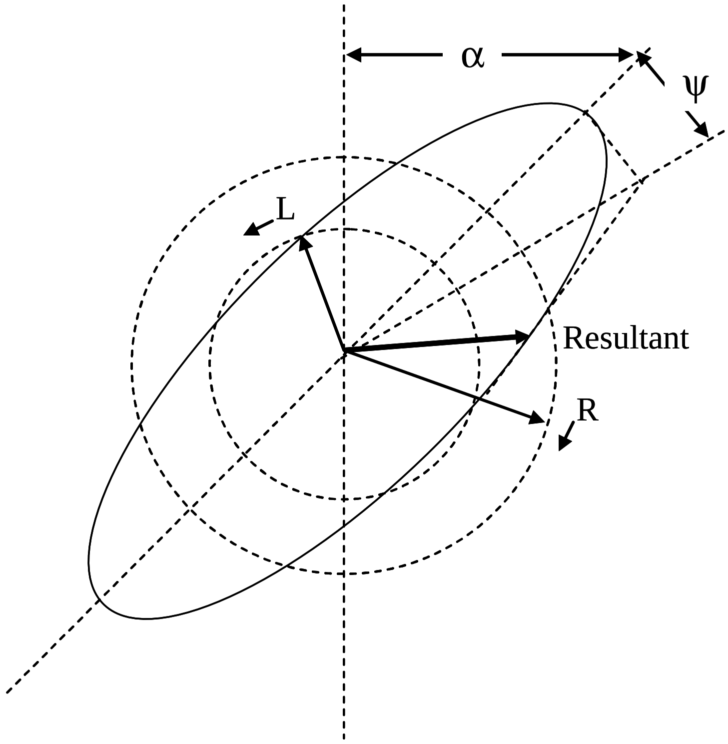


Figure II.3. Illustration of the combined effects of ORD and CD on the polarisation of plane polarised light after passing through a dextrorotatory optically active absorbing medium. The vertical dotted line indicates the optical axis of resultant plane polarised light before entering the medium and prior to ORD. α is the angle by which the optical axis is rotated when exiting the medium, resulting from ORD. The two dotted circles represent the elliptically polarised component left and right circularly polarised electric vectors exiting the medium over a complete cycle. The ellipse represents the electric vector resultant of the left and right circularly polarised light exiting the medium over a complete cycle. The magnitudes of all electric vectors are proportional to the lengths of the arrows.

Also, because the intensity of light, *I*, is proportional to the square of the amplitude of the electric vector, the following equation can be written;

$$A = \log_{10}\left(\frac{I_0}{I}\right) = \log_{10}\left[\exp\left(\frac{4\pi k}{\lambda}\right)\right] \tag{II.12}$$

Combining Equations II.11 and II.12 gives

$$\Delta A \approx 3.0 \times 10^{-2}\,\psi \tag{II.13}$$

where ψ is in degrees. As for the case of optical activity, a material's CD is quantified in terms of its molar ellipticity, [θ], as defined by the equation

$$[\theta] = \frac{M_r\psi}{(100lb')} = \frac{100\psi}{lb} \tag{II.14}$$

where [θ] has the same units as the molar rotation, [M], of deg cm^2 dmol^{-1}. The CD extinction coefficient, $\Delta\varepsilon$, is defined as the difference in the left and right circularly polarised extinction coefficients and is given by

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \frac{\Delta A}{lb} \tag{II.15}$$

Therefore, the relationship between molar ellipticity and $\Delta\varepsilon$ can be expressed as

$$\left[\theta\right] = 2298\Delta\varepsilon \tag{II.16}$$

II.1.5 The Relationship Between Absorption, CD and ORD Spectra

The relationship between the absorption, CD and ORD spectra of a dextrorotatory molecule is shown in Figure II.4. Because the CD spectrum represents the differential between the absorption of left versus right circularly polarised light, CD bands of an optically active molecule will, therefore, be less intense than the corresponding absorption bands. To a reasonably good approximation the anomalous ORD spectrum results from taking the second derivative of the CD spectrum with respect to frequency. More precisely, the ORD and CD spectra are related by a Kronig-Kramers transformation, but by using the relation between dispersion and absorption, a simple semi-empirical conversion between ORD and CD can be arrived at and is given as

$$[A] = [M]_{\max} - [M]_{\min} = 4028 \cdot \Delta\varepsilon_{\max} \tag{II.17}$$

and

$$\Delta\nu = 0.925(\nu_{\max} - \nu_{\min}) \tag{II.18}$$

where $\Delta\nu$ is the CD bandwidth at half-maximum intensity and $\nu_{\max}$ and $\nu_{\min}$ are the maximum and minimum of the anomalous ORD.

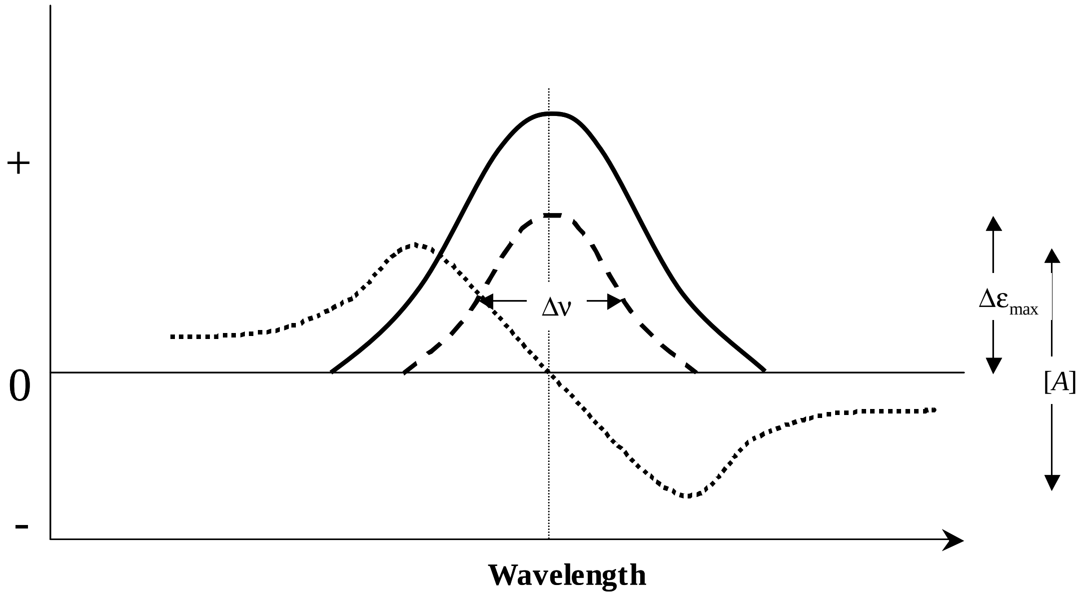


Figure II.4. Absorption, CD and ORD spectra for a single electronic transition of a dextrorotatory optically active molecule. The solid, dashed and dotted lines represent the absorption, CD and ORD simulated spectra, respectively. $\Delta\varepsilon_{max}$ is the extinction coefficient differential characterising the CD spectrum and Δν is the CD bandwidth at half maximum intensity. In moving from higher to lower wavelength the ORD spectrum describes a minimum followed by a maximum and, as such, displays a negative Cotton effect.

II.1.6 The Stimulation of Chiral Electronic Transitions by Circularly Polarised Light

An optically active electronic transition involves a helical displacement of charge, which can be understood in terms of the superposition of a linear charge displacement component and a circular charge displacement component. This description of a chiral transition makes sense in that photons may only interact with electrons to stimulate transitions from lower to higher electronic energy levels if the path of the electron in travelling from the lower to the higher energy level coincides with the path of the photon, which is the case for both left and right circularly polarised light, since the electric vectors of both left and right circularly polarised light in fact describe a helix in three dimensional space when the propagation of light is taken into account. Since a circular motion of charge gives rise to a magnetic moment, a chiral electronic transition has both an electronic and a magnetic transition dipole moment, corresponding to the linear and circular charge displacement, respectively, with both described by electric vectors that run either parallel or antiparallel to each other, according to whether the electron traverses a right or left handed helical path, respectively.

The first selection rule for an optically active electronic transition is that there must exist non-zero electronic and magnetic transition dipole moments. The electric and magnetic dipole intensities, $\boldsymbol{D}_e$ and $\boldsymbol{D}_m$, respectively, determine the strength of absorption of circularly polarised light, which are reflected in the probability of the electronic transition and are given by the square of the electric and magnetic dipole transition moments:

$$\mathbf{D}_e = \langle \Psi_i | \hat{\mu}_e | \Psi_j \rangle^2 \qquad \text{(II.19)}$$

and

$$\mathbf{D}_m = \langle \Psi_i | \hat{\mu}_m | \Psi_j \rangle^2 \qquad \text{(II.20)}$$

where $\hat{\mu}_e$ and $\hat{\mu}_m$ are the electric and magnetic dipole moment operators and Ψ_i and Ψ_j are the ground and excited state wavefunctions, respectively. $\boldsymbol{D}_e$ is generally much larger than $\boldsymbol{D}_m$ and, therefore, is mostly responsible for the absorption intensity.

II.1.7 The Rotational Strength

The absorption intensity and, hence, the area of the CD band for an optically active electronic transition of a chiral material is quantified by means of a parameter known as the rotational strength, *R*, in a similar manner to which the electric dipole strength, or oscillator strength, of a non-chiral transition determines its absorption band area. The rotational strength has the units erg·cm^3 and is experimentally defined as

$$R = \frac{hc}{32\pi^3 N_A} \int \left(\frac{\Delta\varepsilon}{\lambda} \right) d\lambda = 2.295 \times 10^{-39} \int \left(\frac{\Delta\varepsilon}{\lambda} \right) d\lambda \qquad \text{(II.21)}$$

and theoretically defined as the imaginary part of the scalar product of the electric and magnetic dipole transition moments:

$$R = \mathrm{Im}\left\{ \langle \Psi_i | \hat{\mu}_e | \Psi_j \rangle \cdot \langle \Psi_i | \hat{\mu}_m | \Psi_j \rangle \right\} \qquad \text{(II.22)}$$

where Im indicates that the imaginary part is to be taken, since $\hat{\mu}_m$ is purely imaginary. Alternatively, Equation II.21 can be written as

$$R = \mu_e \mu_m \cos\phi \qquad \text{(II.23)}$$

where μ_e and μ_m are the magnitudes of the electric and magnetic transition dipole moments and ϕ is the angle between them.

II.2 The Circular Dichroism of Peptides and Proteins

Protein CD consists of both vibrational and electronic CD. Biologists have selected electronic CD, referred to as ECD or, more commonly just CD, over vibrational CD (VCD) as the method most suited to the study of proteins and their structures and, therefore, ECD spectra comprise the majority of the literature relating to the chiral properties of proteins. Protein CD is divided into three distinct wavelength regions, according to what aspect of the protein is under investigation: far-UV CD involves spectroscopic investigations at wavelengths below 250 nm, near-UV CD is between 250 and 300 nm and near-UV-visible CD is between 300 and 700 nm. Near-UV CD is less commonly employed for protein investigations, although it does find application when information about a protein's tertiary structure is sought.

The chromophores for far- and near-UV CD protein spectroscopy are the amide groups of the peptide backbone and the aromatic residue side-chains, respectively. The peptide group has approximately a tenfold larger extinction coefficient than an aromatic amino-acid side-chain in the far-UV wavelength region and, therefore, is the dominant chromophore giving rise to far-UV CD spectra. Near-UV-visible protein CD can be used for secondary structural investigations of proteins containing chromophoric prosthetic groups. For example, iron-sulfur proteins, heme proteins and flavo proteins all display CD bands above 300 nm, where the chromophores are the prosthetic groups.

II.2.1 Protein Far-UV CD

The most widely used application of electronic CD spectroscopy to proteins is far-UV CD and this is used to characterise a protein in terms of its secondary structure. Before interpreting a protein's far-UV CD spectrum, the spectral intensity must be normalised with respect to protein concentration and this is achieved by expressing the spectral intensity in terms of molar ellipticity [θ], in accordance with Equation II.14, as

$$[\theta] = \frac{\theta}{(10 \times c \times l)} \tag{II.24}$$

where θ is the spectral intensity in millidegrees, c is the concentration of the protein sample in moles/L and l is the optical pathlength in cm. The units of $[\theta]$ are, therefore, degrees cm^2 per decimole. Sometimes, the additional step of expressing the CD of a protein in terms of its mean residue molar ellipticity $[\theta]_{\mathrm{MRW}}$ is performed, whereby

$$[\theta]_{MRW} = \frac{\theta}{(10 \times c_r \times l)} \tag{II.25}$$

where c_r is the mean residue molar concentration and corresponds to the molar concentration of the protein multiplied by the number of peptide bonds in the protein.

Although the peptide group in isolation exhibits a plane of symmetry and is, therefore, achiral and does not exhibit intrinsic CD, its location within the highly unsymmetrical peptide backbone acts to confer optical activity, making it an extrinsic chromophore. The aromatic amino-acid side-chain chromophores for near-UV CD spectroscopy are considered as intrinsic chromophores. The classification of protein CD into either intrinsic or extrinsic CD is done on the basis of whether the chromophore is, or is not, part of the protein, and not on whether that chromophore exhibits either an intrinsic or extrinsic CD signal. Therefore, near- and far-UV CD are examples of intrinsic CD, with near-UV-visible CD representing extrinsic CD.

II.2.2 Electronic Transitions of the Peptide Backbone

Figure II.5*c* shows the far-UV peptide backbone amide chromophore along with the magnetic dipole and electric dipole transition moments. The condensed phase amide far-UV CD is characterised by one $n\pi^*$ and two $\pi\pi^*$ electronic transitions, as shown in Figure II.6. The $n\pi^*$ transition gives rise to a CD band centred around 220 nm and the $\pi_0\pi^*$ and $\pi_+\pi^*$ transitions give CD bands at around 190 and 140 nm, respectively.

Owing to the lack of symmetry of the amide chromophore within a peptide backbone, the $n\pi^*$ transition is not strictly electronically forbidden, as is the case for aldehydes and ketones, and exists as a very weak transition. The energy of the peptide backbone amide $n\pi^*$ transition is sensitive to hydrogen bond formation, to the extent that in apolar solvents the $n\pi^*$ band is red-shifted to 230 nm, while in polar environments it is blue-shifted to 210 nm. The $n\pi^*$ transition involves a circular displacement of charge and, therefore, it is magnetically allowed. The magnetic dipole transition moment is oriented along the carbonyl bond with a magnitude of approximately one bohr magneton. The two $\pi\pi^*$ transitions are oriented approximately parallel ($\pi_0\pi^*$, NV$_1$) and perpendicular ($\pi_+\pi^*$, NV$_2$) to the C-N bond direction. Since both $\pi\pi^*$ transitions involve a linear displacement of charge, they are both electrically allowed and magnetically forbidden, but when the amide group is embedded within a peptide backbone the $\pi\pi^*$ transitions give rise to CD bands at 190 and 140 nm for the NV$_1$ and NV$_2$ transitions, respectively.

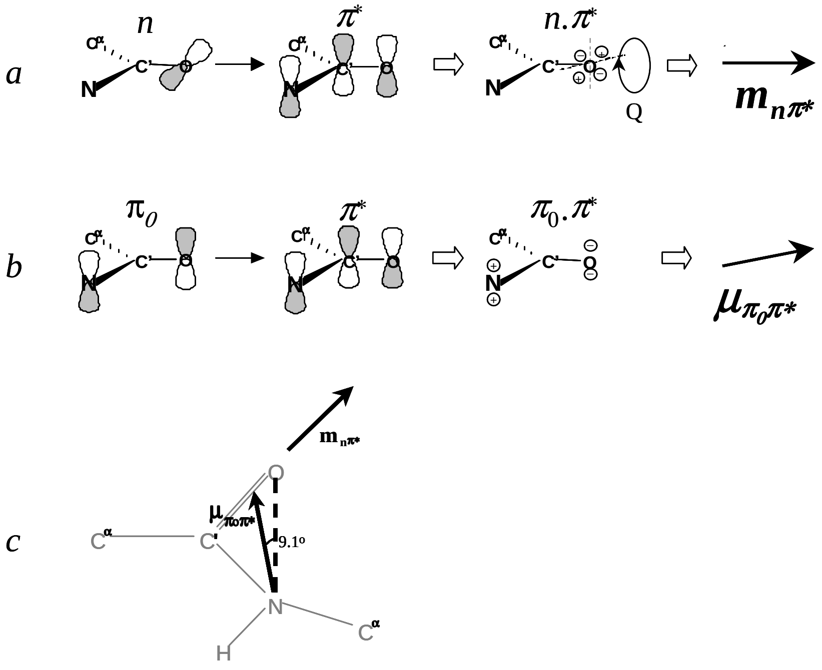


Figure II.5. The $n\pi^*$ and $\pi_0\pi^*$ electronic transitions stimulated by the absorption of far-UV circularly polarised light. *a* The amide $n\pi^*$ transition involves the electronic transition of the lone pair electrons on the carbonyl oxygen into the π^* carbonyl antibonding orbital. This transition involves a circular displacement of electric charge, corresponding to an electric quadrupole moment, Q, and, therefore, gives rise to a magnetic transition dipole moment, $m_{n\pi^*}$. As such, the amide $n\pi^*$ transition is electrically forbidden and magnetically allowed. *b* The $\pi_0\pi^*$ transition involves a linear displacement of charge from the nitrogen atom towards the centre of the carbonyl bond, therefore, giving rise to an electronic transition dipole moment, $\mu_{\pi 0\pi^*}$. The $\pi_0\pi^*$ transition is, therefore, electrically allowed and magnetically forbidden. *c* The amide chromophore responsible for the far-UV CD signal of peptides. $m_{n\pi^*}$ extends directly from and in the same direction as the carbonyl bond while $\mu_{\pi 0\pi^*}$ is offset from the N-O axis by 9.1^o and points towards the centre of the carbonyl bond. Although neither the $n\pi^*$ nor the $\pi_0\pi^*$ transitions are intrinsically chiral, since for the $n\pi^*$ transition an electric transition dipole moment is lacking and the $\pi_0\pi^*$ transition is without a magnetic dipole transition moment, the lack of symmetry of the peptide backbone generates a chiral environment and so induces chirality to both transitions.

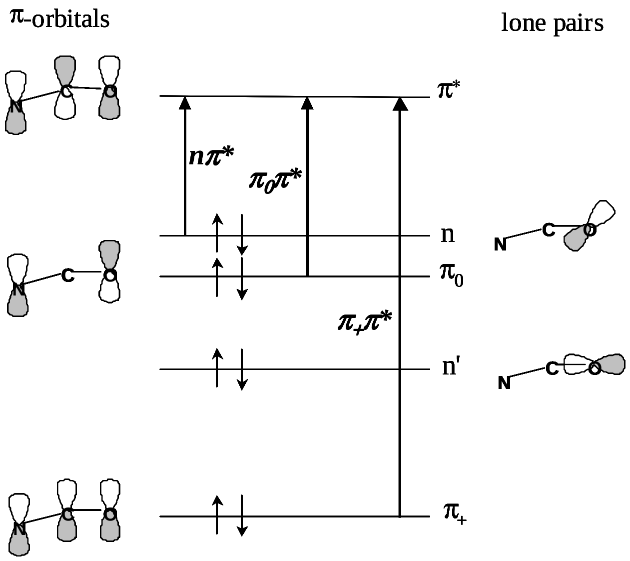


Figure II.6. Energy level diagram illustrating the optically active peptide backbone $n\pi^*$, $\pi_0\pi^*$ (NV$_1$) and $\pi_+\pi^*$ (NV$_2$) electronic transitions. The lower energy $n\pi^*$ transition occurs at an approximate wavelength of 220 nm in condensed phase and the higher energy $\pi_0\pi^*$ transition occurs at approximately 190 nm, with the $\pi_+\pi^*$ transition falling at 140 nm.

II.2.3 Exciton Coupled-Oscillator Interactions and Protein Secondary Structure

The peptide backbone amide $\pi\pi^*$ and $n\pi^*$ transitions also interact with each other in any of three types of mechanisms: Coupled oscillator exciton interactions can occur between degenerate or nearly degenerate $\pi\pi^*$ transitions on separate peptide groups, in a mechanism referred to as *μ-μ coupling*; mixing between the $n\pi^*$ and $\pi\pi^*$ transitions on the same peptide group give rise to the so called *one-electron effect*; and where the same mixing is between separate peptide groups there occurs the effect known as *μ-m coupling*. Each of these three mechanisms has a geometrical dependence and it is this dependence that links the far-UV CD spectrum of a protein to its secondary-structure, since the energies of the $n\pi^*$ and $\pi\pi^*$ transitions are perturbed by the extent to which they interact. Moffitt was able to explain the CD signal of an α-helical polypeptide by taking account of the exciton coupled-oscillator effect, or *μ-μ coupling,* in a pioneering work that laid the foundations for the interpretation of protein CD spectra.[3]

Protein CD spectra are interpreted, on some level at least, by means of CD spectra generated by model peptide fragments, which are representative for each of the characteristic structural motifs that combine to form the individual and varied secondary-structures of all proteins. These characteristic secondary structures are represented as α-helix, β-sheet, β-turn, poly(pro)*II* and unordered conformation. The model CD spectrum of each is shown in Figure II.7. Importantly, at 222 nm the CD spectrum of a protein is dominated by the α-helical secondary structure contribution and this phenomenon is widely used in stability studies of a protein's α-helical domains .

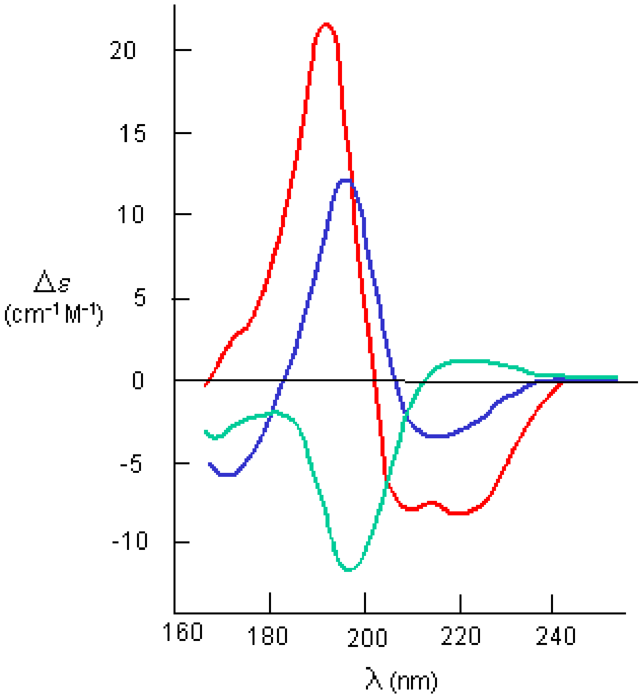


Figure II.7. CD spectra of model polypeptides adopting the α-helix, β-sheet and random-coil secondary structures. The overlay shows the relative intensities of a residues contribution to the CD spectrum at each wavelength when in an α-helix (red), β-sheet (blue) and random-coil (green) conformation. (Redrawn from [2])

II.2.4 The CD Spectrum of an α-Helix

The α-helix CD spectrum is characterised by negative minima located at 222 and 208 nm and a positive maximum centred at 190 nm. As shown in Figure II.8, the minimum at 222 nm results from the $n\pi^*$ transition. The other negative band at 208 nm and the positive band at 190 nm result from the exciton splitting of the $\pi_0\pi^*$ transition by the *μ-μ coupling* mechanism, whereby the negative long wavelength band is polarised along the helix axis and the positive short wavelength band is polarised perpendicular to it. The $\pi_+\pi^*$ transition for the α-helix has been observed as a positive band at 140 nm, while the shoulder feature observed at 175 nm is at present unassigned. The negative band at 222 nm resulting from the $n\pi^*$ transition derives most of its intensity from static field mixing with the $\pi_0\pi^*$ transition.

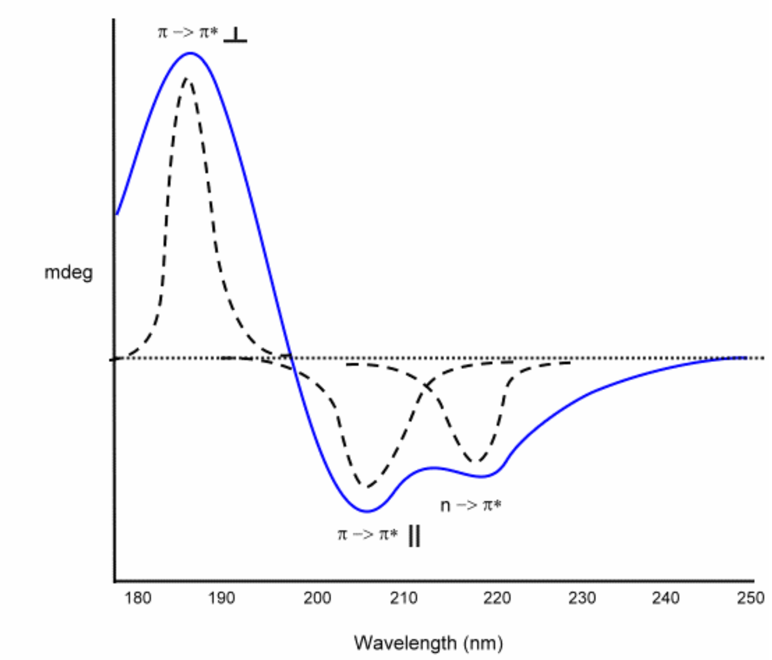


Figure II.8. The α-helical CD signal. The α-helix CD signal can be constructed from three Gaussian bands centred at 222, 208 and 190 nm. The $n\pi^*$ transition gives rise to the negative band at 222 nm, while the $\pi_0\pi^*$ transition is split into two transitions polarised perpendicular and parallel to the helix axis, giving rise to the positive band located at 190 nm and the negative band at 208 nm, respectively. [1]

The dependence of the α-helix CD spectrum on the length of the helix has been examined by Applequist using the dipole interaction model and it was observed that when the helix length was varied from 4 to 12 residues there occurred a corresponding 25-30% increase in CD intensity.[1] This chain-length dependence of the α-helix CD spectrum is extremely important when a quantitative analyses of a protein’s secondary structure is sought from its CD spectrum and the empirical equation

$$V_r = \frac{V_\infty (r-k)}{r} \qquad \text{(II.26)}$$

was proposed by Chen et al., where V_r is the CD amplitude at a specific wavelength for a helix of r residues, V_∞ is the CD of an infinite helix at the same wavelength, and k is an empirical parameter.[4] The parameter k is the number of residues considered missing due to end effects at either end of the helix and k values of between 3.7 and 4.6 have been suggested from the analysis the CD magnitudes at 222 nm of experimental data.[5] This corresponds to the expected two-residue-at-either end distortion of the α-helix observed by Sreerama et al. when they analysed a set of protein CD spectra in terms of estimating secondary structure fractions.[6]

II.2.5 The CD Spectrum of a β-Sheet

The β-sheet CD spectrum displays a negative band around 215 nm, a positive band at approximately 198 nm and a negative band near 175 nm (Figure II.9). The negative band at around 215 nm has been assigned to the $n\pi^*$ transition. The amplitude of this band, however, comes predominantly from *μ-m coupling* between the $n\pi^*$ and $\pi_0\pi^*$ transitions. The bands at 198 and 175 nm are the components of the $\pi_0\pi^*$ exciton split transition, via the *μ-μ coupling* mechanism. Differences in the CD spectrum of a parallel versus anti-parallel β-sheet conformation are predicted from exciton coupling theory, with the bands offset by approximately 5 to 15 nm. The absolute difference between the CD spectra of the two conformations is understood to be inversely proportional to the sheet widths of the compared conformations.

The difficulty with the interpretation of a protein’s CD spectrum in terms of its β-sheet content arises from the ease with which β-sheets become distorted within proteins, giving rise to somewhat variable CD signatures. Also, the efficiency of exciton coupling within β-sheets is greatly diminished compared to the α-helix, with the result that in proteins with a mixture of α-helical and β-sheet (αβ-proteins) content the major component to the protein’s CD spectrum are the α-helix domains. This further complicates the accurate interpretation of a protein’s CD spectrum for β-sheet content in αβ-proteins. That β-sheet proteins have limited solubilities in aqueous solution also hinders their interpretation by CD spectroscopy.

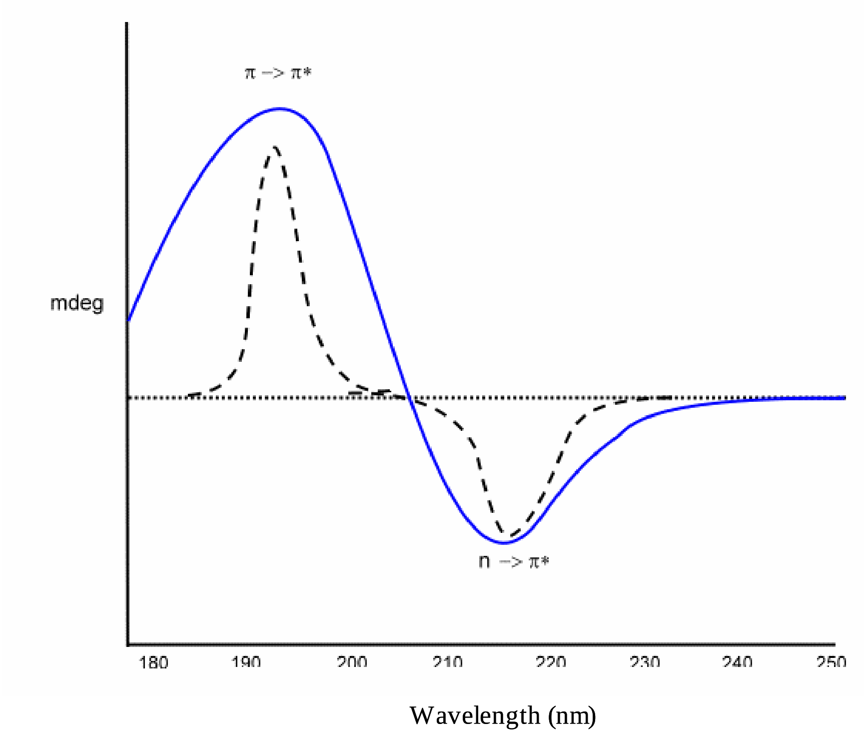


Figure II.9. The β-sheet CD signal. The β-sheet CD signal can be constructed from two Gaussian bands centred at 215 and 196 nm. The $n\pi^*$ transition gives rise to the negative band at 215 nm, while the $\pi_0\pi^*$ transition is exciton split into two bands; a positive band at around 198 nm and a negative band at approximately 175 nm. The amplitude of the negative band at 215 nm derives mainly from the mixing of the $n\pi^*$ and $\pi_0\pi^*$ transitions via the $\mu\text{–}m$ mechanism.[1]

The variations in the degree of twisting of the β-sheet geometry has been investigated by Manning *et al.* with the conclusion that increased twisting leads to increased amplitudes of the CD bands.[7] Whereas weakly twisted β-sheet conformations have relatively unchanged CD bands, strongly twisted β-sheets display a significantly more intense $\pi_0\pi^*$ band at around 198 nm. Also, the CD dependence on the β-sheet length has been examined by Woody [8] and by Madison and Schellman [9], and they found that with increased sheet length there occurs an increase in the magnitudes of both the $n\pi^*$ and $\pi_0\pi^*$ bands.

II.2.6 The CD Spectrum of a β-Turn

The CD spectra of β-turns is also somewhat variable, as expected, since there are in total eight types of β-turn conformation. A β-turn is understood as a non-repetitive structure spanning a total of three residues, which reverses the polypeptide chain direction. Of the different β-turn conformations the most common are the type I, II and II turns and theoretical calculations suggest that type I and II turns have CD spectra similar to an undistorted β-sheet, but with red-shifted band maxima.[8]

II.2.7 The CD Spectrum of Poly(Pro) Structures

Poly(Pro) structures are of either cis or trans conformations, forming either a right handed helix with 3.3 residues per turn or a left handed helix with 3 residues per turn, respectively. The cis conformation is refered to as poly(pro)I and is favoured in nonpolar solvents, whereas the trans poly(pro)II conformation is favoured in polar solvents. The CD spectrum of the poly(pro)I helix is characterised by a weak negative band at around 232 nm and a positive band at lower wavelengths, with the crossover point around 205 nm. In contrast, the poly(pro)II conformation CD spectrum exhibits a weak positive band near 226 nm and an intense negative band near 206 nm. As such, the poly(pro)II CD signature is very similar to that of a random coil CD spectrum (Figure II.10). From this, it has been proposed that unordered polypeptides contain significant amounts of poly(pro)II-like conformation, irrespective of the presence of proline residues within the polypeptide primary sequence.[10]

II.2.8 The CD Spectrum of a Random Coil

Random coil or unordered secondary-structures are peptide regions that display non-repeat peptide-peptide orientations. As such, randomly coiled regions tend to exist in an extended rather than globular tertiary conformation. Figure II.10 shows the typical CD spectrum of a randomly coiled protein. A very weak positive band is seen at around 215 nm and an intense negative band appears near 197 nm. As expected, the 215 nm band corresponds to the the $n\pi^*$ transition and the negative band around 197 nm results from the $\pi_0\pi^*$ transition. The weakness of the the $n\pi^*$ transition is a consequence of the lack of mixing between the the $n\pi^*$ and $\pi_0\pi^*$ transitions.

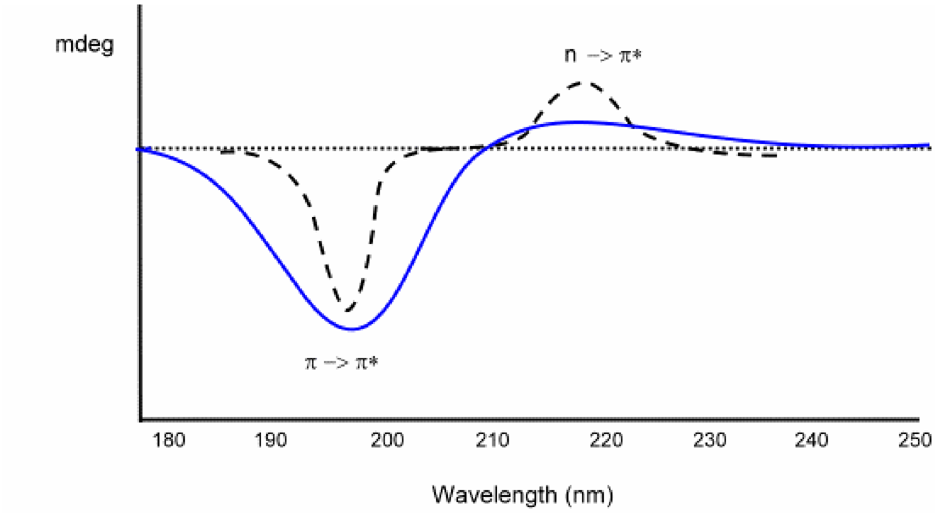


Figure II.10. The CD spectrum of a random coil. The random coil CD spectrum can be reproduced by a weak positive Gaussian component at around 215 nm and a strong negative component at around 197 nm. The positive and negative components correspond to the $n\pi^*$ and $\pi_0\pi^*$ transitions, respectively.[1]

II.2.9 The Calculation of Protein CD Spectra

The calculation of a protein’s CD spectrum is possible using theoretical methods, which can be divided into two approaches; the classical approach, which is based upon the interaction of electromagnetic energy with matter and the subsequent changes in bond polarizability; and the quantum mechanical method, which focuses on the characteristics of the upward electronic transitions. The most commonly employed method for the calculation of protein CD spectra is the matrix method, which is a quantum mechanical method, since it depends on the quantum mechanical description of the chromophores undergoing transitions.[11] The matrix method is in line with both the coupled-oscillator and exciton models in that it begins with the construction of a matrix with a secular determinant based on the energies of and interactions between the transitions of different chromophores. For example, the secular matrix for a two-chromophore system having one transition on each chromophore is

$$\begin{pmatrix} E_1 & V_{12} \\ V_{12} & E_2 \end{pmatrix}$$

where E_1 and E_2 are the energies of the transitions on chromophores 1 and 2, respectively, and V_{12} is the energy of interaction between the two transition dipole moments. Two excited states, corresponding to linear combinations of the individual transitions, result when the matrix is diagonalized. As such, the matrix method employs the ‘divide and conquer’ approach, where the molecule is divided into independent chromophores with the potential for mutual interaction. When the calculation of protein CD spectra is considered it makes sense to construct a matrix consisting of the same amount of independent chromophores as equals three times the amount of residues within the protein. This is to allow for the treatment of the $n\pi^*$ and two $\pi\pi^*$ transitions as separate chromophores. For a protein containing N residues, considering the first $n\pi^*$ and the first two $\pi\pi^*$ transitions leads to the following interaction matrix, **H**,

$$\begin{pmatrix}
E_{11} & V_{11,12} & V_{11,13} & V_{11,21} & V_{11,22} & V_{11,23} & \bullet & \bullet & \bullet & V_{11,N1} & V_{11,N2} & V_{11,N3} \\
\bullet & E_{12} & V_{12,13} & V_{12,21} & V_{12,22} & V_{12,23} & & & & V_{12,N1} & V_{12,N2} & V_{12,N3} \\
\bullet & & E_{13} & V_{13,21} & V_{13,22} & V_{13,23} & & & & V_{13,N1} & V_{13,N2} & V_{13,N3} \\
\bullet & & & E_{21} & V_{21,22} & V_{21,23} & & & & V_{21,N1} & V_{21,N2} & V_{21,N3} \\
\bullet & & & & E_{22} & V_{22,23} & & & & V_{22,N1} & V_{22,N2} & V_{22,N3} \\
\bullet & & & & & E_{23} & & & & V_{23,N1} & V_{23,N2} & V_{23,N3} \\
\bullet & & & & & & \bullet & & & \bullet & \bullet & \bullet \\
\bullet & & & & & & & \bullet & & \bullet & \bullet & \bullet \\
\bullet & & & & & & & & \bullet & \bullet & \bullet & \bullet \\
\bullet & & & & & & & & & E_{N1} & V_{N1,N2} & V_{N1,N3} \\
\bullet & & & & & & & & & & E_{N2} & V_{N2,N3} \\
\bullet & & & & & & & & & & & E_{N3}
\end{pmatrix}$$

where the energies of the localized transitions form the diagonal elements and the interactions between the localised transitions form the off-diagonal elements. E_{ij} represents the energy of the transition j on the chromophore I and $V_{ij,kl}$ represents the interaction between the transition j on chromophore i and the transition l on chromophore k. As a rule, the second entry to the subscript number represents either the $n\pi^*$, $\pi_0\pi^*$ and $\pi_+\pi^*$ transitions as 1, 2 or 3, respectively, for both diagonal and off-diagonal entries. The appropriate charge distribution is used to represent a localised transition. Therefore, since the $\pi\pi^*$ transitions involve linear displacements of charge, they are represented by dipolar charge distributions and the $n\pi^*$ transition is represented by a set of quadrupolar charges centred on the relevant atoms. The $V_{ij,kl}$ terms are calculated using the monopole-monopole approximation in considering the Coulomb interactions between localised transitions as follows:

$$V_{ij,kl} = \sum_m \sum_n \frac{q_{ijm} q_{knl}}{r_{ijm,knl}} \qquad \text{(II.27)}$$

where the indices *m* and *n* correspond to the point charges for the transitions *j* and *l*, respectively, and the distance between the point charges is given by *r*. For the interaction matrix, **H**, the terms with $i \neq k$, and *j* and $l = 2$ or 3 denote interactions which are electrically allowed and, therefore, these interactions are by the μ–μ mechanism. Where $i \neq k$, $j = 1$, and $l = 2$ or 3, the interaction is between an $n\pi^*$ and $\pi\pi^*$ transition on different chromophores and, therefore, occurs by the μ–m mechanism. When $i = k$, the interaction mechanism is the *one-electron-effect*, since both transitions lie on the same chromophore and undergo mixing of their excited states in the static field of the molecule. If $i \neq k$, $j = 1$ and $l = 1$, the interaction is between quadrupolar charge distributions of two $n\pi^*$ transitions on two separate chromophores.

Group-states result from the diagonalization of **H** and these represent the eigenvalues and eigenvectors that describe the excited states of the composite molecule, as follows

$$\mathbf{H} = \mathbf{U}\,\mathbf{H_d}\,\mathbf{U^T} \qquad \text{(II.28)}$$

The diagonal entries of the matrix $\mathbf{H_d}$ give the eigenvalues, which correspond to the energies of the transitions. The elements of the unitary matrix **U** ($\mathbf{U}\,\mathbf{U^T} = \mathbf{I}$, where $\mathbf{U^T}$ is the transpose of **U** and **I** is the identity matrix), describe the mixing of the localized transitions. The all important rotational strength can then be calculated by determining the transition moments for each group-state of the composite system, and this is done by combining the eigenvectors with the properties of the localized transitions, $\boldsymbol{\mu}_k^0$ and $\mathbf{m}_k^0$, as follows:

$$\boldsymbol{\mu}_i = \sum\nolimits_k \mathbf{U}_{ki}\boldsymbol{\mu}_k^0$$

and (II.30)

$$\mathbf{m}_i = \sum\nolimits_k \mathbf{U}_{ki}\mathbf{m}_k^0$$

Figure II.11 shows the effectiveness of the matrix method in reproducing the CD spectra of a number of well-studied proteins. The calculation of CD spectra from the crystal structures of proteins can be seen from Figure II.11 to enjoy a reasonable level of accuracy, particularly in terms of the bandshapes predicted.

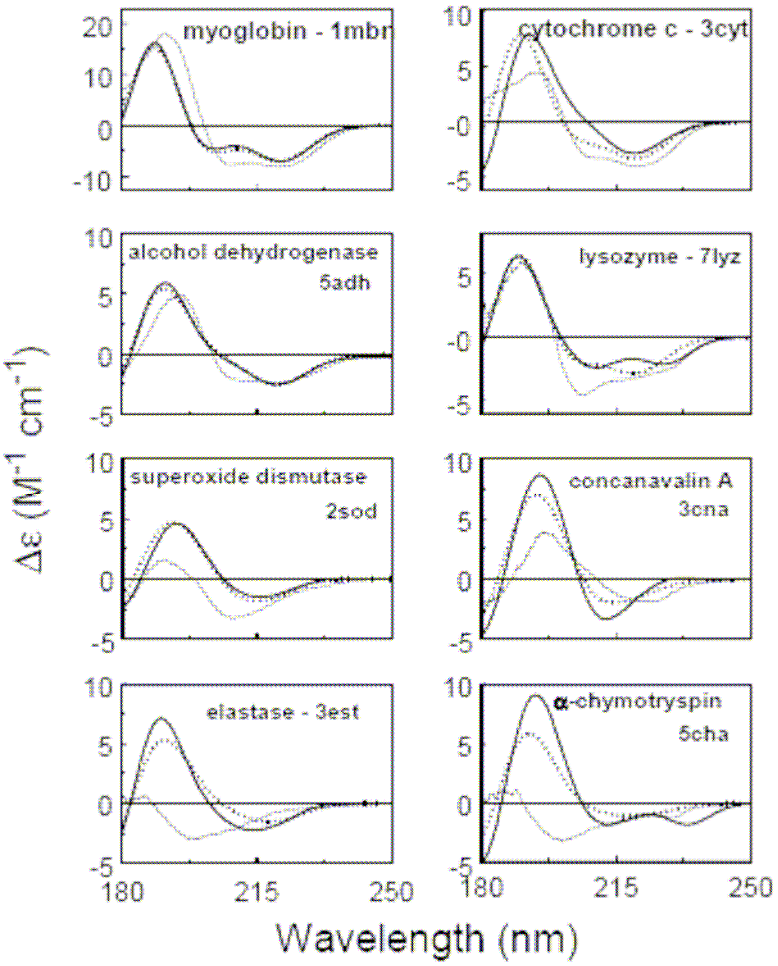


Figure II.11. Comparison of the calculated and experimental CD for selected proteins. The thin line represents the experimental CD spectra, while the dotted and bold lines give the side-chain excluded and side-chain included theoretical CD spectra, respectively, as calculated by the matrix method and using the Protein Data Bank (PDB) crystal structures for each protein. The PDB code is listed for each protein. (Reproduced from [2])

II.3 Fluorescent Resonance Energy Transfer

II.3.1 Energy Transfer

The theory behind fluorescence resonance energy transfer (FRET) was primarily developed by Förster [12], with Stryer and Haugland extending these principles to the lab to show that FRET could be used as an effective and reliable "molecular ruler" for measuring distances on the order of between 10-70 Å [13]. FRET is a non-radiative dipole-dipole coupled oscillator interactive process, whereby an applied electric (radiation) field generates an induced dipole moment in the donor molecule by raising it to an excited state and this induced dipole moment then induces a dipole moment of equal energy in the acceptor molecule. It is at this resonance point that energy transfer between the donor and acceptor dipole moments becomes possible. The FRET process can be explained both classically, in terms of oscillating dipoles, or quantum mechanically, in terms of donor and acceptor ground and excited states.

Where labelling with either donor or acceptor proceeds to 100 %, the efficiency of energy transfer, *E*, from the donor to the acceptor oscillator is given by

$$E = 1 - \frac{I_{DA}}{I_D} = 1 - \frac{\tau_{DA}}{\tau_D} \qquad \text{(II.31)}$$

where I_{DA} and τ_{DA} are the intensity and lifetime of the donor in the presence of the acceptor, respectively, and I_D and τ_D are the donor intensity and lifetime in the absence of the acceptor. In cases where the acceptor is also fluorescent, it is possible to measure *E* from the increase in fluorescence of the acceptor using the equation

$$E = \left(\frac{I_{AD}}{I_A} - 1 \right) \left(\frac{\varepsilon_A}{\varepsilon_D} \right) \qquad \text{(II.32)}$$

where I_{AD} is the emission of the acceptor in the presence of the donor, I_A is the fluorescence of the acceptor only and ε_A and ε_D are the molar extinction coefficients of the acceptor and donor at the wavelength of excitation.

II.3.2 The Förster Distance

The Forster distance, R_0, is the distance at which half of the energy absorbed by the donor is transferred to the acceptor via FRET. The value of R_0 is heavily dependent on the overlap between the donor emission (f_D) and the acceptor absorption (ε_A) (Figure II.12). The normalised spectral overlap (*J*) at a given wavelength (λ) is given by the equation

$$J = \frac{\int \varepsilon_A(\lambda) f_D(\lambda) \lambda^4 d\lambda}{\int f_D(\lambda) d\lambda} \qquad (\mathrm{M^{-1}.cm^{-1}.nm^4}) \qquad \text{(II.33)}$$

The value of R_0 is then described by the equation

$$R_0 = \left(8.79 \times 10^{-5} J q_D n^{-4} \kappa^2 \right)^{\frac{1}{6}} \qquad (\text{Å}) \qquad \text{(II.34)}$$

where q_D is the quantum yield of the donor emission in the absence of acceptor, n is the refractive index and κ^2 is a geometric factor related to the relative angle of the two transition dipoles. By calculating the value of E from the spectral analysis of the labelled and unlabelled systems the distance between the donor and acceptor molecules (R) can be calculated as

$$R = R_0\left(\frac{1}{E}-1\right)^{\frac{1}{6}} \qquad (\text{Å}) \qquad (\text{II.35})$$

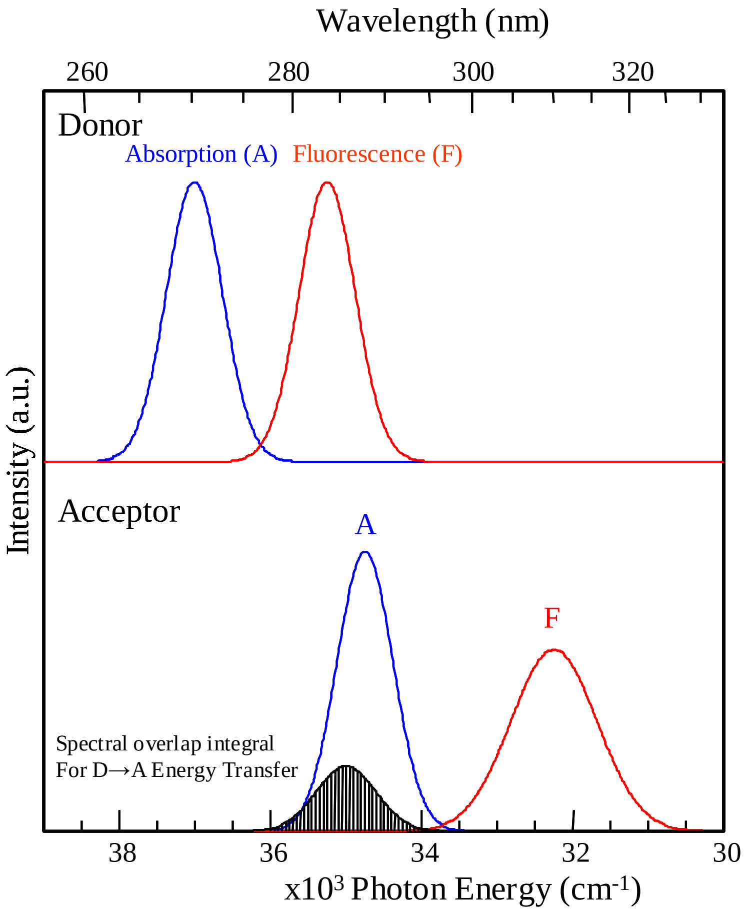


Figure II.12. Donor-acceptor spectral overlap. The overlap of the donor fluorescence and acceptor absorption bands, as illustrated by the black spectral overlap integral band, is a necessary condition for FRET.

II.3.3 The Rate of Energy Transfer

The rate of energy transfer (k_T) is given by the equation

$$k_T = \left(\frac{R_0}{R}\right)^6 \Big/ \tau_D \qquad \text{(II.36)}$$

From this equation it can be seen that k_T has an R^{-6} dependence, such that as the separation between the donor and acceptor becomes greater than the R_0 value the rate of energy transfer drops off sharply. The quantum mechanical definition of FRET can be formulated in describing the energy of interaction between the donor and acceptor by the Hamiltonian

$$H = \frac{(\mu_D \cdot \mu_A)}{R^3} + \frac{(\mu_D \cdot \mathbf{R})(\mu_A \cdot \mathbf{R})}{R^5} \qquad \text{(II.37)}$$

where **R** is the vector separating the centres of the donor and acceptor and μ_D and μ_A are the transition dipole moments of the donor and acceptor, respectively. According to Fermi's rule, the rate of energy transfer is proportional to the square of the Hamiltonian matrix element between final and initial states, such that

$$k_T \propto \left[\left\langle D^* \middle| \left\langle A \middle| \left[\frac{(\mu_D \cdot \mu_A)}{R^3} - \frac{3(\mu_D \cdot \mathbf{R})(\mu_A \cdot \mathbf{R})}{R^5}\right] \middle| D \right\rangle \middle| A^* \right\rangle\right]^2 \qquad \text{(II.38)}$$

where the initial state is the product of the excited state of the donor (<D*|) and the ground state of the acceptor (<A|) and the final state is the product of the ground state of the donor (<D|) and the excited state of the acceptor (<A*|). In this quantum mechanical description the assumption is made that the individual wavefunctions for the ground and excited states of the donor and acceptor are not perturbed by the dipole-dipole coupling that occurs during FRET. This is valid since such coupling would be expected to be very weak. By separating out the quantities that depend on the donor wavefunctions, the acceptor wavefunctions and those that depend on the relative orientation of the donor and acceptor, we arrive at the equation

$$k_T \propto R^{-6} \left\langle D^* \middle| \mu_D \middle| D \right\rangle^2 \left\langle A \middle| \mu_A \middle| A^* \right\rangle^2 \left[\left\langle D^* \middle| \left\langle A \middle| (\cos\theta_{DA} - 3\cos\theta_D \cos\theta_A) \middle| D \right\rangle \middle| A^* \right\rangle\right]^2 \quad \text{(II.39)}$$

where θ_{DA} is the angle between the donor and acceptor transition dipole moments and θ_D and θ_A are the angles between the donor and acceptor transition dipole moments and the **R** vector joining the two molecules, respectively. Here the rate of energy transfer can be seen to depend on the square of the transition dipole moments of the donor and acceptor, which can be related to the acceptor absorption and donor emission properties. This equation can be rewritten as

$$k_T \propto R^{-6} \left\langle D^* \middle| \mu_D \middle| D \right\rangle^2 \left\langle A \middle| \mu_A \middle| A^* \right\rangle^2 \kappa^2 \tag{II.40}$$

where the orientation factor κ is $\left\langle D^* \middle| \left\langle A \middle| \left(\cos\theta_{DA} - 3\cos\theta_D \cos\theta_A \right) \middle| D \right\rangle \middle| A^* \right\rangle$. Where the transition dipole moments are not fixed in space, but move rapidly, with respect to the donor lifetime, the value of κ must take account of this. As such, the value of κ^2 represents an average orientation value and, by taking the motion of the transition dipoles as random, a value of κ^2 as 2/3 is generally accepted. This approximation is thought to introduce an uncertainty in distance determination by FRET measurements of up to 11%.

References

1 Applequist J. Full Polarizability Treatment of the Pi-Pi-Star Absorption and Circular Dichroic Spectra of Alpha-Helical Polypeptides. Journal of Chemical Physics 71(11):4332-4338, 1979.

2 Sreerama N, Woody RW. Computation and analysis of protein circular dichroism spectra. Numerical Computer Methods, Pt D 383:318-351, 2004.

3 Moffitt W. Optical Rotatory Dispersion of Helical Polymers. Journal of Chemical Physics 25(3):467-478, 1956.

4 Chen YH, Yang JT, Chau KH. Determination of Helix and Beta-Form of Proteins in Aqueous-Solution by Circular-Dichroism. Biochemistry 13(16):3350-3359, 1974.

5 Gans PJ, Lyu PC, Manning MC, Woody RW, Kallenbach NR. The Helix-Coil Transition in Heterogeneous Peptides with Specific Side-Chain Interactions - Theory and Comparison with Cd Spectral Data. Biopolymers 31(13):1605-1614, 1991.

6 Sreerama N, Venyaminov SY, Woody RW. Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. Protein Science 8(2):370-380, 1999.

7 Manning MC, Illangasekare M, Woody RW. Circular-Dichroism Studies of Distorted Alpha-Helices, Twisted Beta-Sheets, and Beta-Turns. Biophysical Chemistry 31(1-2):77-86, 1988.

8 Woody RW. The Circular-Dichroism of Oriented Beta Sheets - Theoretical Predictions. Tetrahedron-Asymmetry 4(3):529-544, 1993.

9 Madison V, Schellma.J. Optical-Activity of Polypeptides and Proteins. Biopolymers 11(5):1041-&, 1972.

10 Tiffany ML, Krimm S. New Chain Conformations of Poly(Glutamic Acid) and Polylysine. Biopolymers 6(9):1379-&, 1968.

11 Bode KA, Applequist J. Globular protein ultraviolet circular dichroic spectra. Calculation from crystal structures via the dipole interaction model. Journal of the American Chemical Society 120(42):10938-10946, 1998.

12 Forster T. Experimentelle Und Theoretische Untersuchung Des Zwischenmolekularen Ubergangs Von Elektronenanregungsenergie. Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences 4(5):321-327, 1949.

13 Stryer L, Haugland RP. Energy Transfer - a Spectroscopic Ruler. Proceedings of the National Academy of Sciences of the United States of America 58(2):719-&., 1967.

Chapter III

Protein Mid-IR Fourier Transform Spectroscopy

III.1 The Absorption of Infra-Red Radiation

The energies of the discrete vibrational energy levels, v, of the potential energy curve for an harmonic oscillator (Figure III.1), where the energy spacing between successive energy levels is the same and equal to *hv*, are given by

$$E_v = (v + \tfrac{1}{2})h\cdot v \quad v = 0, 1, 2, 3\ldots. \qquad \text{(III.1)}$$

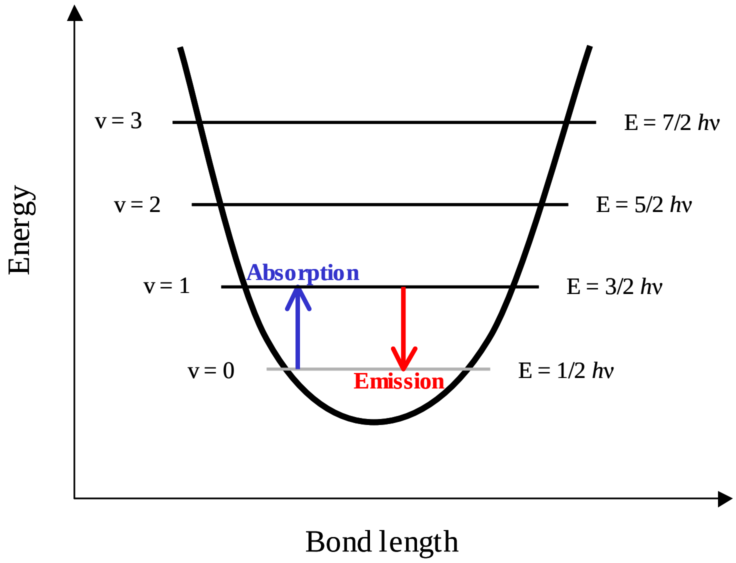


Figure III.1. Potential energy curve for an harmonic oscillator. The absorption and emission vibrational transitions are shown in blue and red, respectively

Vibrational transitions occur only between consecutive energy levels in the harmonic oscillator model and at room temperature most vibrational transitions occur between the ground vibrational state (v = 0) and the first excited vibrational state (v = 1). The vibration of a molecule involves either a change in the length or geometry of its internal bonds and this is illustrated in Figure III.2 by the stretching and bending vibrational modes of the water molecule.

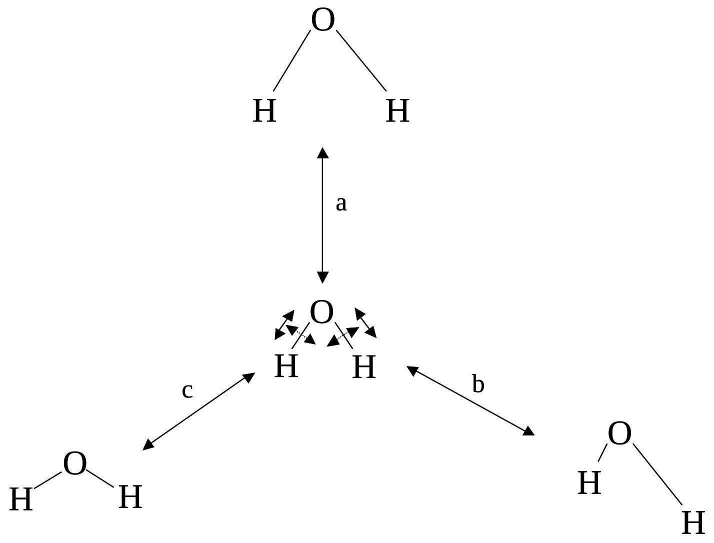


Figure III.2. The normal modes of vibration of H_2O. The symmetric and asymmetric stretching modes (3400 cm^{-1}) are labelled a and b, respectively, and the bending mode (1600 cm^{-1}) is labelled c.

IR absorption involves interaction of electromagnetic waves with molecular vibrations. The absorption of IR light by a vibrating molecule can be understood by defining the molecule in terms of the classical picture of two vibrating point charges +q and –q, which are connected by a spring. When the frequencies of the electromagnetic wave and the vibrating point charges are exactly equal, the electric field can interact so as to amplify the movement of the partial charges, as a photon of light of frequency equal to the frequency of the vibrating point charges becomes absorbed into the system. Implicit in the above description is that the absorption of IR radiation by a molecule depends upon there existing a dipole moment to the bond involved in the absorption process. Also, the larger the dipole moment – corresponding to the size of the point charges at each end of the spring - the stronger will be the interaction between the electric field and the system. As such, the polarity of a bond is directly related to its IR extinction coefficient. In the simplest case of a two-atomic oscillator, the absolute frequency of absorption of the vibrating bond depends both on the bond strength and the mass of the atoms on either end of the bond. In the above analogy the bond strength corresponds to the tightness of the spring and this is referred to as the force constant, k. The frequency, υ, of a two-atomic oscillator is given by

$$\upsilon = \frac{\sqrt{\frac{k}{\mu_r}}}{2\pi} \qquad \text{(III.2)}$$

where μ_r is the reduced mass, defined by the equation $\mu_r = m_1 .\ m_2 /(m_1 + m_2)$, where m is the mass of either atom 1 or 2 at each end of the bond.

As mentioned, the value of k is one of the factors governing the absorption frequency. Since we are discussing the vibrations of covalent bonds, and a covalent bond is simply a sharing of electrons, the bond strength and, hence, the value of k, is proportional to the electron density within the bond. Any internal or external molecular interaction adding or subtracting electron density to or from the vibrating bond in question will have the effect of respectively increasing or decreasing its vibrational frequency. As can be seen from Equation III.2, an alternative mechanism to alter the frequency of a vibrating bond is to change the mass of one or both of the bonded atoms. Increasing the mass of either atom has the effect of reducing the vibrational frequency of the bond. The dependence of the bond vibrational frequency on the value of m_1 or m_2 is often exploited by using IR spectroscopy to interpret effects of isotopically labelled proteins. A more common use of the mass-frequency dependence in protein IR spectroscopy, however, is the use of D_2O as a replacement solvent for H_2O. This procedure will be discussed in depth in Chapter VII, but at this point it is enough to remark that the use of heavy water has the effect of shifting the solvent band so that it no longer directly overlaps the protein amide-I IR signal, making protein IR spectral acquisition considerably less problematic.

According to the harmonic oscillator rule, the quantum mechanical description of an IR absorption process is that a photon of frequency υ and energy $h\upsilon$ (h is Planck’s constant) is absorbed by a molecule, promoting it from a lower lying vibrational energy level, E, to a higher energy level, $E + h\upsilon$. The aforementioned condition that the frequency of the absorbed photon must match the vibrational frequency is ensured, since the spacing of successive vibrational energy levels is $h\upsilon$. For the harmonic oscillator model, all vibrational transitions are subject to the selection rule $\Delta v = \pm 1$, where v are the vibrational energy levels.

III.2 The Transition Dipole Moment

A transition dipole moment (TDM) is defined as the change in the bond dipole in making the transition from one energy level to another. Fermi's golden rule states that the square of this TDM can be taken as the transition probability between the respective vibrational energy levels. The TDM is expressed by the following equation:

$$TDM = \left\langle \psi_0 \phi_m \middle| \mu \middle| \psi_0 \phi_n \right\rangle \qquad \text{(III.3)}$$

where the transition occurs between energy levels *n* and *m* of the electronic ground state, ψ_0. Equation III.3 makes use of the Born-Oppenheimer approximation in that it separates the nuclear wavefunctions ϕ_n and ϕ_m from the electronic wavefunction ψ_0 and so allows the TDM to be considered as a simple function of the spatial coordinates of each of the nuclei of the vibrating atoms. The interaction of the electric field with the dipole moment as a function of time, t, is taken into account by using the interaction potential operator

$$\mathbf{U} = \boldsymbol{\mu}(\mathrm{t})\mathrm{E}(\mathrm{t}) \qquad \text{(III.4)}$$

where E(t) represents the electric field and $\boldsymbol{\mu}$(t) represents the dipole moment operator. Since the majority of oscillators that absorb mid-IR radiation at room temperature are not thermally excited, a diatomic oscillator TDM can be expressed as

$$TDM = \left\langle \frac{\partial \mu}{\partial R}(R_0) \right\rangle \left(\frac{h}{8\pi^2 m_r \upsilon} \right)^{0.5} \qquad \text{(III.5)}$$

where *h* is Planck's constant, m_r is the reduced mass of the diatomic oscillator and υ is the frequency of oscillation. The selection rule that IR absorption only occurs when there is a change in the molecular dipole moment arises out of the left hand term, which denotes the transition's electronic component. The expectation value for this term is the change of dipole moment at equilibrium position, R_0, and with a larger expectation value comes a stronger IR absorption. This can be seen by again using the simple classical picture of vibrating point charges $+q$ and $-q$ with a connecting spring. The dipole moment between the point charges separated by the length of the spring, *R*, is $\mu = qR$ and the change of this dipole moment is $\partial\mu/\partial R = q$. And, since larger charges give stronger IR absorption, a larger change in the dipole moment also must give stronger absorption. As bond polarity is simply an alternative expression for dipole moment, bonds with higher polarities will have more intense IR absorptions. For example, a C-H bond is significantly less polar than a C-O bond, with their respective electronegativity differences of 0.3 and 1. Consistent with this is that strong IR bands are observed for C-O bonds, whereas only weak bands result from C-H bonds.

III.3 Protein Absorption of Infra-Red Radiation

For an averaged sized protein there exists many thousand normal modes of vibration, where there is 3N-6 normal modes for a molecule containing N atoms. For example, even for the small AavLEA1 protein the value of N is 1325 and there exist 3969 normal modes. At first glance it would seem that the vibrational spectrum of a protein would prove far too complicated to be able to extract any meaningful secondary structural information. However, it is the C═O and N—H bonds on the peptide backbone that comprise the majority of the more polar bonds within a peptide and, as such, they make the largest contribution to its IR spectrum. Furthermore, the normal vibration used to infer protein secondary structure from its IR spectrum is dominated by only one internal coordinate - the C═O stretching vibration. Despite this a protein’s vibrational spectrum is very complex, with many of the vibrational bands overlapping one another. It is possible, however, to select different spectral regions to give answers to specific questions. For example, the region of interest for information about a protein’s secondary structure is the amide-I band, located between 1600 and 1700 cm^{-1}.

The six highest frequency normal modes of vibration of the amide group are, in descending order: the amide-A and -B modes, centred at ~3300 and 3170 cm^{-1}, respectively; the amide-I mode, centred at ~1650 cm^{-1}; the Amide-II mode, centred at ~1550 cm^{-1}; the amide-III mode, located between 1400-1200 cm^{-1}; and the skeletal stretch, giving a band between 1200-880 cm^{-1}. A model compound used to understand these different modes of vibration is N-methylacetamide (NMA) (Figure III.3).[4] Although the contribution of internal coordinates to the six previously mentioned normal modes of vibration would be expected to differ for NMA compared with an amide group incorporated into a protein backbone, a general correlation between the two should persist. Each of the six normal modes of NMA are, therefore, discussed below and are expected to be roughly equivalent to an amide group within a protein.

Figure III.4 shows the contribution of each of the above vibrational modes to the IR absorption spectrum of aqueous hemoglobin.

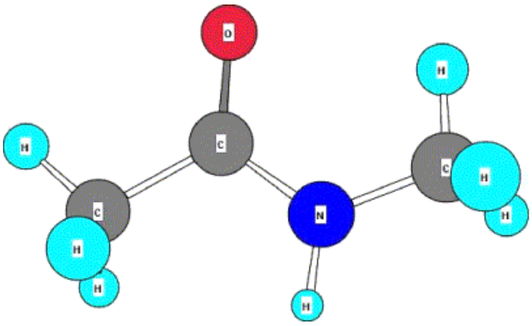


Figure III.3. The molecular structure for NMA. Carbon, nitrogen, oxygen and hydrogen are shown as grey, blue, red and light blue, respectively.

III.3.1 The Amide-A and –B Vibrational Modes

Both the amide-A and -B modes result from the NH stretching vibration and are localised entirely on the NH group. Of the two the amide-A band between 3310 and 3270 cm^{-1} is the most intense and is usually part of a Fermi resonance doublet, with the amide-B band occurring between 3100 and 3030 cm^{-1}. The NH stretching vibration can be resonant with an overtone of the amide-II vibration.

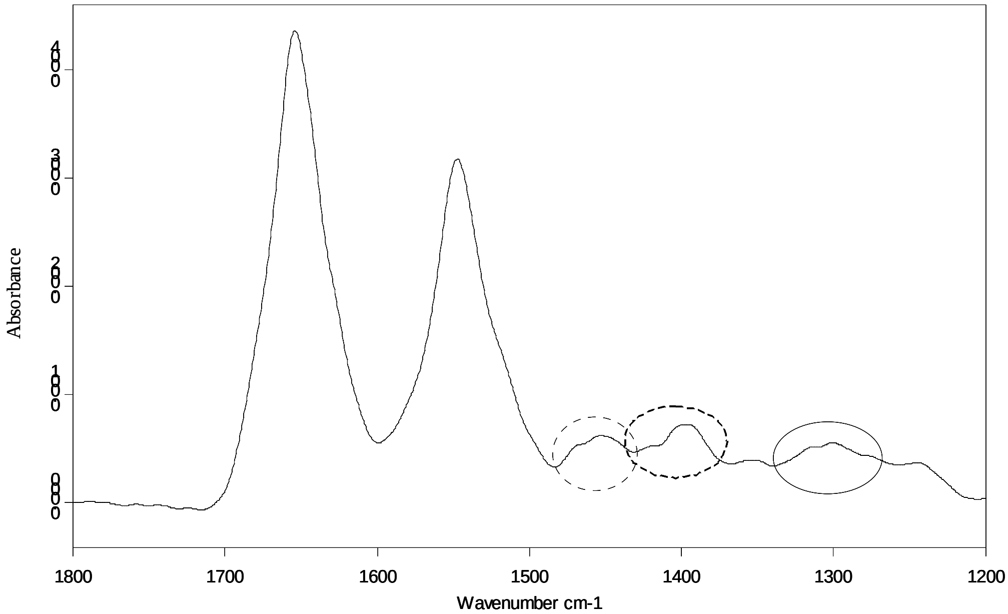


Figure III.4. Protein amide absorption bands. The hemoglobin FTIR absorption spectrum shows the amide-I band centred at approximately 1650 cm^{-1}, resulting from the peptide backbone CO symmetric stretch. The amide-II band results from a combination of the protein backbone NH bend and the CO asymmetric stretch and is centred at approximately 1540 cm^{-1}. The solid circle marks the location of the amide-IIII band, which results from the CN symmetric stretch and NH bending vibrations. The band enclosed by the dotted circle results from amino-acid side-chain groups and corresponds to the asymmetric stretch of NH_3^+ and the bend of CH_2. The dashed circle shows the location of the symmetric stretching of COO^- of side-chain groups.

III.3.2 The Amide-I Vibrational Mode

The NMA amide-I vibration is centred at ~1650 cm^{-1} and arises predominantly from the C═O stretching vibration. Importantly, the CCN deformation, the out-of-phase CN stretching vibration and the NH in-plane bend do, however, make minor contributions to the amide-I band and the exact extent to which each of these three internal coordinates contribute to the amide-I normal mode in a protein is determined by the protein's secondary structure.[5] This dependency, along with the fact that the amide-I vibration is largely unaffected by the nature of the amino-acid side-chain, is what makes the amide-I band the band of choice for inferring protein secondary structure from its IR absorption spectrum.

III.3.3 The Amide-II Vibrational Mode

The NMA amide-II mode is centred at ~1550 cm^{-1} and arises mainly from the out-of-phase combination of the NH in-plane bend and the CN stretching vibrations. The CO in-plane bend, the CC stretch and the NC stretch also make minor contributions to the amide-II mode. Although the amide-II mode is also not affected by amino-acid side-chains, the relationship between a protein's secondary structure and its amide-II vibrational mode is considerably more complex than for the amide-I mode. As such, the amide-II mode is generally not considered when interpreting protein secondary structure using IR spectroscopy. The amide-II band is either weak or absent in a protein's Raman spectrum.

Figure III.5 shows an FTIR absorption spectral overlay of Myoglobin in both ^1H_2O and ^2H_2O. The comparison reveals that the amide-II band, centred at ~1550 cm^{-1} in ^1H_2O, becomes significantly red-shifted to ~1455 cm^{-1} in ^2H_2O. The reason for this is that N-deuteration converts the amide-II mode to a predominantly CN stretching vibration, with this altered mode referred to as the amide-II' mode. Since the N^2H bending vibration is substantially red-shifted relative to the N^1H bend, it is no longer capable of coupling with the CN stretching vibration and instead mixes with other modes in the 1070-900 cm^{-1} region. The fundamental difference between the amide-II and -II' modes means they will be affected differently by the environment of the amide group. In other words, they will exhibit non-cooperative dependencies on protein secondary structure. This effect is not witnessed for the corresponding amide-I' mode because of the lack of any large contribution from H atoms to this mode and from Figure III.5 the locations of the myoglobin amide-I and amide-I' bands can be seen to be similar. Given that it is the amide-I band that is used for protein secondary structure IR spectroscopic analysis, protein IR spectroscopy is commonly performed using ^2H_2O solvent, which shifts the OH bending vibration such that it no longer overlaps the protein amide-I band.

III.3.4 The Amide-III Vibrational Mode

The NMA amide-III mode is located between 1400 to 1200 cm^{-1} and arises predominantly from a combination of the in-phase combination of the NH bending and the CN stretching vibrations, with minor contributions from the CO in-plane bending and the CC stretching vibrations. Because the NH bending vibration contributes to several modes of vibration of the polypeptide backbone in the 1400 to 1200 cm^{-1} region, protein amide-III IR bands are very complex. Furthermore, unlike the amide-I and –II modes, a protein's amide-III band is very much affected by its particular amino-acid side-chains. Among different proteins the contributions of backbone and side-chains to the amide-III mode may vary considerably and this makes the amide-III band unsuitable for protein secondary structure analysis. Upon ^1H_2O to ^2H_2O solvent exchange the amide-III mode behaves in a complex and less predictable manner when compared to the amide-I and –II modes.

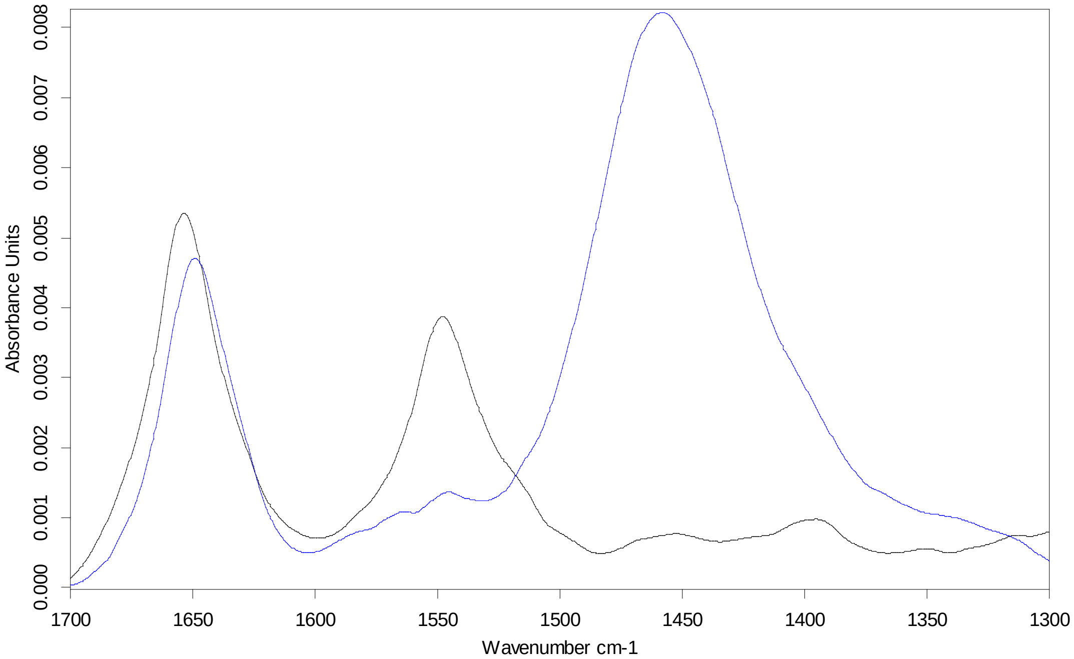


Figure III.5. Comparison of the protein amide band locations in ^1H_2O and ^2H_2O solvent. The myoglobin FTIR absorption spectrum is shown when using either ^1H_2O (black) or ^2H_2O (blue) as a solvent.

III.3.5 The Skeletal Stretch Vibrational Mode

The skeletal stretching vibrations for NMA and polypeptides are considerably different. NMA consists of two weakly absorbing and well-defined skeletal stretching vibrations, a predominantly NC_α stretching mode at ~1096 cm^{-1} and a mixed mode at ~881 cm^{-1}. Polypeptides lack any distinct NC mode and the skeletal stretching vibrations are distributed over a number of modes that, depending on side-chain interactions, absorb between 1180 to 920 cm^{-1}. In both NMA and polypeptides the skeletal stretching vibrations produce strong Raman bands between 960 and 880 cm^{-1}.

III.4 The Origin and Character of the Amide-I Vibrational Mode

Figure III.6 shows the amide-I absorption bands for the characteristic protein secondary structures α-helix, β-sheet and random-coil. The splitting of the β-sheet amide-I band by about 50 cm^{-1} into a major component centred at ~1630 cm^{-1} and a minor component centred at ~1680 cm^{-1} posed a considerable challenge to theoreticians attempting to account for the dependence of the amide-I vibrational mode on protein secondary structure. Eventually, in 1972 Abe and Krimm proposed the transition dipole coupling (TDC) mechanism, which was able to explain the unusual amide-I bands of β-sheet proteins.[6] Consideration of through-bond coupling and hydrogen bonding effects also helped investigators understand the dependence of a protein's amide-I bandshape on its secondary structure.

Since the C_α atom is not displaced any considerable distance during the amide-I CO stretch, the through-bond interaction of the vibration of one amide group to that of its neighbours is not expected to have much of an affect on a protein's overall amide-I band. This is not the case for hydrogen bonding, the effects of which have been investigated for NMA by both *ab initio* calculations and time-resolved matrix IR spectroscopy.[7; 8] In the case of a hydrogen bond to the C═O group the amide-I frequency is red-shifted by 20-25 cm^{-1}. Where the NH group is hydrogen bonded a smaller red-shift of 10-20 cm^{-1} occurs.

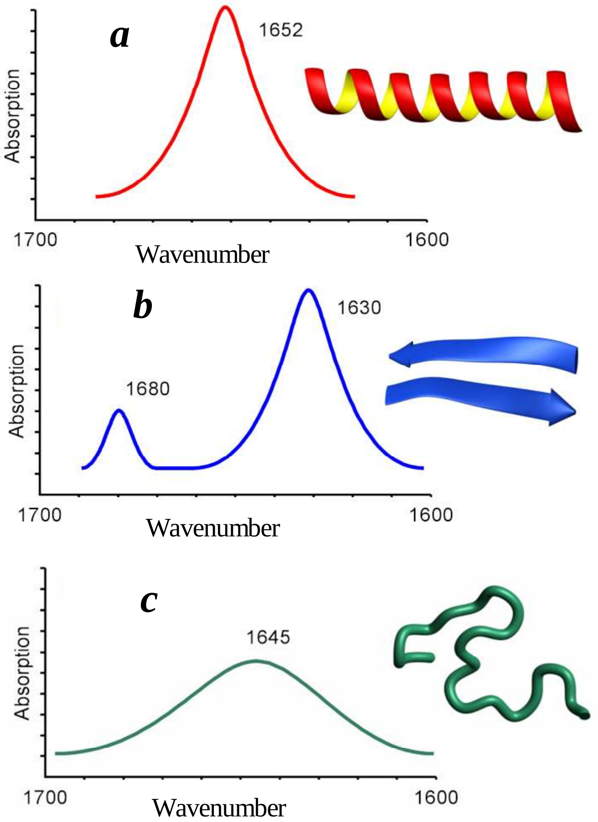


Figure III.6. Theoretical IR amide-I absorption bands of model protein secondary-structures. (Modified from Bruker) ***a*** The calculated amide-I band of an undistorted α-helix has a distinct Lorenztian bandshape with a peak maximum at 1652 cm^{-1}. ***b*** The amide-I band of an intramolecular antiparallel β-sheet structure showing its splitting into a major and minor component centred at 1630 and 1680 cm^{-1}, respectively. Both bands are predominantly Lorenztian in shape. ***c*** The characteristically broad Guassian bandshape of a random-coil secondary structure with a maximum at 1645 cm^{-1}.

Krimm and Bandekar studied the two model β-sheet polypeptides poly-β-L-Ala and poly-β-L-Glu and observed amide-I peak maxima of 1632 and 1624 cm^{-1}, respectively.[2] They attributed the red-shifted amide-I peak maxima of poly-β-L-Glu to its ability to form stronger hydrogen bonds. For each of the characteristic secondary structures, the relative frequency order of the amide-I peak maxima is consistent with the strengths of hydrogen bonding present and, in decreasing order, they are as follows: intermolecular extended chains (1610 - 1628 cm^{-1}), intramolecular antiparallel β-sheets (1630 - 1640 cm^{-1}), α-helices (1648 - 1658 cm^{-1}), 3_{10}-helices (1660 - 1666 cm^{-1}) and non-hydrogen bonded amide groups in DMSO (1660 - 1665 cm^{-1}).[9] Important to the use of IR spectroscopy in studying the folding and unfolding of protein tertiary structures, Parrish and Blout discovered that the vibrational frequency of solvated helices is approximately 20 cm^{-1} lower than that of non-solvated helices.[10] The additional hydrogen bonds to the solvated helices by surrounding water molecules are understood to be responsible for this effect.

III.5 Transition Dipole Coupling

It is widely accepted by theoreticians that the transition dipole coupling (TDC) mechanism lies at the heart of a protein's amide-I bandshape dependence on its secondary structure.[3] This mechanism is a resonance interaction involving the oscillating dipoles of adjacent amide groups. TDC is strongest when the coupled oscillators vibrate with the same frequency and the overall effect of the interaction depends on both the distance between the oscillating dipoles and their relative orientation. The two effects of TDC are exciton transfer and exciton splitting. Exciton transfer is a process whereby energy absorbed by one oscillator is transferred to a second nearby oscillator and, therefore, the excited state is understood as being delocalised over the two coupled oscillators. For an α-helix the delocalisation has a typical time constant of 0.5 ps and is over a distance of 8 Å.[11] Exciton splitting shifts the amide-I frequency as a function of the distance, orientation and relative phases of the coupled oscillators. As shown in Figure III.7, for the case of two coupled oscillators, the exciton splitting effect gives rise to a splitting of the excited state energy level into an in-phase and an out-of-phase energy level. The energy separation, or splitting, between these resulting energy levels is dramatically demonstrated in the case of β-sheet structures, where there is a splitting of 70 cm^{-1}. A similar effect is seen in far-UV circular dichroism protein spectroscopy, where $\pi-\pi^*$ electronic transitions couple in a correspondingly relative-orientation-dependent manner, yielding secondary structure dependent circular dichroism between 160-250 nm. Implicit in the description above is that TDC is a through-space rather than a through-bond mechanism and that it is mediated by the Coulomb interactions between moving partial charges.

The existence of a permanent dipole moment is not a prerequisite for the occurrence of TDC and its formalism starts by giving an account of a non-stationary excited state dipole-dipole interaction, as follows: Given an inter-dipole distance R, where R > 3Å, the interaction between two oscillating dipoles, $\mathbf{U}_{AB}$, can be expanded in a multipole series, with the leading term being the dipole-dipole interaction. In SI units $\mathbf{U}_{AB}$ is given by

$$\mathbf{U}_{AB} = (4\pi\varepsilon_r\varepsilon_0)^{-1}[(\boldsymbol{\mu}_A\boldsymbol{\mu}_B)/\mathbf{R}^3 - 3(\boldsymbol{\mu}_A\mathbf{R})\,(\boldsymbol{\mu}_B\,\mathbf{R})/\mathbf{R}^5] \qquad \text{(III.6)}$$

where $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$ are the dipole moment operators describing the varying oscillations of oscillators A and B, respectively, $\mathbf{U}_{AB}$ is the operator of the interaction potential of A with B and **R** is the operator describing the varying distance between A and B.

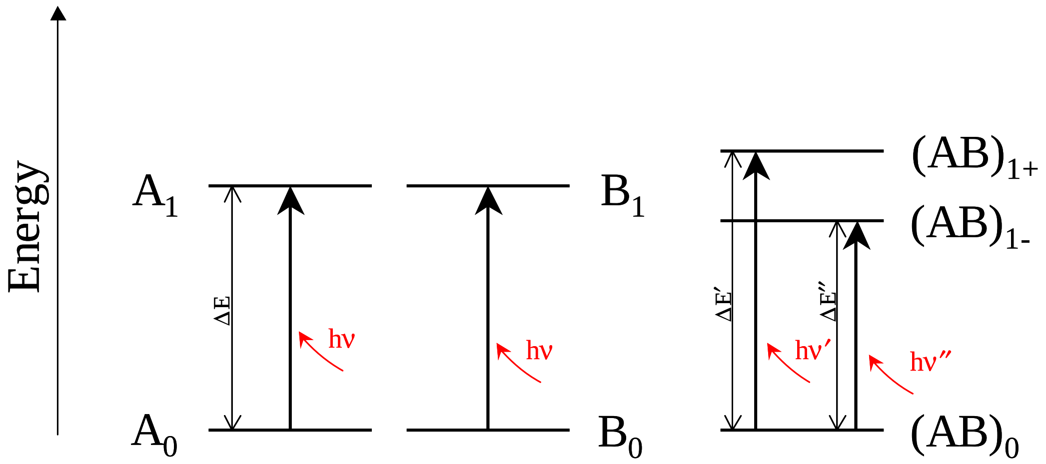


Figure III.7. Energy level diagram of the IR stimulated coupled amide-I transition. The identical amide-I vibrational transition of each unperturbed oscillator from the ground state (A_0 and B_0) to the first excited state (A_1 and B_1) occur with absorption of an IR photon of energy E = hν = ΔE. TDC leads to a splitting of the first excited state energy levels of the coupled oscillators into two new energy levels $(AB)_{1+}$ and $(AB)_{1-}$. When a photon of energy hν′ = ΔE′ is absorbed by a coupled oscillator the transition $(AB)_0 \rightarrow (AB)_{1+}$ occurs. Absorbing a photon of energy hν″ = ΔE″ stimulates the $(AB)_0 \rightarrow (AB)_{1-}$ transition. The result of TDC is that the IR absorption spectrum shifts from showing a single absorption band at the frequency ν to displaying two new absorption bands at frequencies ν′ and ν″.

Although using the dipole-dipole interaction returns values slightly larger than observed for adjacent oscillators, it is nonetheless used in explaining TDC.[12] The energy transfer between oscillators can be written as follows

$$V_{AB} = \langle A_0B_1|\mathbf{U}_{AB}|A_1B_0\rangle \qquad \text{(III.7)}$$

where the subscripts 0 and 1 refer to the ground and first excited vibrational states, respectively, and V_{AB} corresponds to the wavefunction of energy transfer between the oscillators A and B. As such, A_0B_1 refers to oscillator A being in the ground state and oscillator B being in the first excited state, and the reverse is true for A_1B_0. Since V_{AB} is a wavefunction, according to Fermi's golden rule the probability of an energy transfer transition between oscillators A and B is given by $|\, V_{AB} \,|^2$. Because the dipole moment operator of each oscillator acts only on that oscillator, the scalar product of Equation III.7 can be separated into $<A_1|\boldsymbol{\mu}_A|A_0>$ and $<B_1|\boldsymbol{\mu}_B|B_0>$. Also, because both A and B refer to the same amide-I mode on adjacent amide groups both of these scalar products can be replaced by $<1|\boldsymbol{\mu}|0>$, which is non-zero for any IR-active vibrational transition. In evaluating the scalar products, the operator **R** is replaced by a constant equal to the distance between the geometric centres. Substituting Equation III.6 into Equation III.7 gives the following:

$$V_{AB} = (4\pi\varepsilon_r\varepsilon_0)^{-1}|<1|\boldsymbol{\mu}|0>|^2 X_{AB} \tag{III.8}$$

where X_{AB} represents the energy transfer dependence on the relative orientation of the coupled oscillators. Krimm and Abe have determined that

$$X_{AB} = (\cos\alpha - 3\cos\beta\cos\gamma)/R^3 \tag{III.9}$$

where α is the angle between the two TDM's, β is the angle between the line joining the centres of the TDM's with the TDM of oscillator A, γ is the angle between the same line with the TDM of oscillator B and R is the distance between the centres of the TDM's.[3] Therefore, from Equation III.8, the probability of an energy transfer transition between two coupled oscillators depends upon their distance apart, their relative orientations and the TDM of the isolated oscillator, and the stronger the IR absorbance of the particular normal vibrational mode in question (in this case the amide-I mode) the more probable this transition will be. Whenever the value of X_{AB} is non-zero, the energy transfer transition $|A_1B_0> \rightarrow |A_0B_1>$ will have some probability and, consequently, the eigenstates $|A_1B_0>$ and $|A_0B_1>$ will be non-stationary. It is also apparent from Equation III.8 that, since different characteristic secondary structures contain different coupled oscillator relative orientations, each structural type will have its own unique extinction coefficient and, hence, the overall protein IR amide-I absorption band will be a weighted sum of all the individual secondary structure amide-I bands, as resulting from the presence of those secondary structures within the protein.

It is helpful at this point to separate the scalar product into its electronic and nuclear contributions to the TDM of the isolated oscillator, as follows:

$$|<1|\boldsymbol{\mu}|0>| = |<\partial\boldsymbol{\mu}/\partial\mathbf{q}><\phi_1|\mathbf{Q}|\phi_0>| = |<\partial\boldsymbol{\mu}/\partial\mathbf{q}>|(h/8\pi^2\nu)^{0.5} \tag{III.10}$$

The electronic contribution ($<\partial\boldsymbol{\mu}/\partial\mathbf{q}>$) represents the change of the dipole moment brought about by the change of the normal coordinate q at equilibrium position. The nuclear contribution is given by $<\phi_1|\mathbf{Q}|\phi_0>$, with ϕ_0 and ϕ_1 representing the ground and first excited state wavefunctions, respectively, and the frequency of the isolated oscillator is given by ν. Substituting Equation III.10 into Equation III.8 gives the final expression for V_{AB} as

$$V_{AB} = (4\pi\varepsilon_r\varepsilon_0)^{-1}(h/8\pi^2\upsilon)|<\partial\boldsymbol{\mu}/\partial\mathbf{q}>|^2 X_{AB} \qquad \text{(III.11)}$$

The next step in formalising TDC is to describe the excited states of the coupled oscillators. The occurrence of non-stationary excited states presents an obstacle when attempting to describe the energy levels of the individual oscillators. As such, we shall confine ourselves to the less detailed description of the excited states of the coupled oscillators combined. Similar to the approach taken when accounting for a molecular orbital in terms of the contributing atomic orbitals, the combined excited state of the coupled oscillators can be expressed as a linear combination of the excited states of the separate oscillators in the absence of energy transfer, $|A_1B_0>$ and $|A_0B_1>$, giving two new states $|1+>$ and $|1->$. These can then be expressed as

$$\left|1+\right\rangle = \frac{1}{\sqrt{2}}\left(\left|A_0B_1\right\rangle + \left|A_1B_0\right\rangle\right) \qquad \text{and}$$

$$\left|1-\right\rangle = \frac{1}{\sqrt{2}}\left(\left|A_0B_1\right\rangle - \left|A_1B_0\right\rangle\right) \qquad \text{(III.12)}$$

Evaluating the scalar product $<1-|\mathbf{U}_{AB}|1+>$ returns a value of zero and, therefore, transitions from $|1+> \rightarrow |1->$ and vice versa do not occur. As such, these newly generated states are stationary and can be used to calculate the energy eigenvalues of the excited states. Disregarding the scenario of there being any permanent dipole moments affecting the vibrational mode in question, the energy eigenvalues are then given as

$$E_{|1+>} = <1+|\mathbf{H}_A + \mathbf{H}_B + \mathbf{U}_{AB}|1+> = E_1 + E_0 + V_{AB} \quad \text{and}$$

$$E_{|1->} = <1-|\mathbf{H}_A + \mathbf{H}_B + \mathbf{U}_{AB}|1-> = E_1 + E_0 - V_{AB} \qquad \text{(III.13)}$$

where $E_0 = <0|\mathbf{H}|0>$ and $E_1 = <1|\mathbf{H}|1>$ are the energies of the ground and first excited states of the oscillators in the absence of energy transfer. Since the energy of the ground states of the coupled oscillators is unaffected by the coupling interaction and since TDC produces the two new excited state energy levels $E_1 + E_0 \pm V_{\mathrm{AB}}$, the energy difference between the ground and excited states is

$$\Delta E = \Delta E_{\mathrm{noIA}} \pm V_{\mathrm{AB}} \qquad \text{(III.14)}$$

where ΔE_{noIA} is the energy difference between the ground and excited states of the uncoupled oscillator. Given the relation $E = c^2\tilde{\nu}$, the absorption band of the isolated oscillator is correspondingly split into two bands, defined by the equation

$$\tilde{\nu} = \tilde{\nu}_{\mathrm{noIA}} \pm V_{\mathrm{AB}}/hc \qquad \text{(III.15)}$$

Substituting Equation III.11 into Equation III.15 gives the band splitting affect of TDC as

$$\tilde{\nu} = \tilde{\nu}_{\mathrm{noIA}} \pm (4\pi\varepsilon_r\varepsilon_0)^{-1}(8\pi^2\nu c)^{-1}|\langle\partial\boldsymbol{\mu}/\partial\mathbf{q}\rangle|^2 X_{\mathrm{AB}} \qquad \text{(III.16)}$$

Equation III.16 yields three important consequences: Firstly, TDC results in a splitting of the isolated oscillator absorption band into two. Secondly, since the displacement of the dipole moment with vibration ($\partial\boldsymbol{\mu}/\partial\mathbf{q}$) is proportional to the IR absorption intensity, the greater the IR absorption of the isolated oscillator the larger will be the band splitting effect. Thirdly, the magnitude of band splitting as a result of TDC is heavily dependent on the relative oscillator orientation, X_{AB}.

Assuming that each operator $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$ acts only on the states of oscillators A and B, respectively, means that terms like $\langle A_1B_0|\boldsymbol{\mu}_A|A_0B_0\rangle$ are equal to $\langle A_1|\boldsymbol{\mu}_A|A_0\rangle$. Since we are considering the case for oscillators possessing no permanent dipole moment, i.e. the terms $\langle A_0|\boldsymbol{\mu}_A|A_0\rangle$ and $\langle B_0|\boldsymbol{\mu}_B|B_0\rangle$ equal zero, the following expression for the polarization of a coupled oscillator transition results:

$$|\langle 1\pm|\boldsymbol{\mu}_A + \boldsymbol{\mu}_B|0\rangle|^2 = \tfrac{1}{2}\,|\langle B_1|\boldsymbol{\mu}_B|B_0\rangle \pm \langle A_1|\boldsymbol{\mu}_A|A_0\rangle|^2 \qquad \text{(III.17)}$$

And because both oscillators A and B are the same, this gives

$$\tfrac{1}{2}\,|\langle B_1|\boldsymbol{\mu}_B|B_0\rangle \pm \langle A_1|\boldsymbol{\mu}_A|A_0\rangle|^2 = |\langle 1|\boldsymbol{\mu}|0\rangle|^2\,(1 \pm \cos\boldsymbol{\Theta}) \qquad \text{(III.18)}$$

where $\boldsymbol{\Theta}$ is the angle between the two TDMs of A and B. From Equation III.18 it can be seen that the polarization of a coupled oscillator transition is not identical to that of the isolated oscillators and that the polarization of the coupled oscillator system can be obtained by adding or subtracting the TDMs of the isolated oscillators when the oscillators oscillate in-phase or out-of-phase, respectively. This relationship will be useful in understanding the particular IR absorbencies of coupled oscillators for various relative orientations.

The localisation of the amide-I and –II modes on the peptide group within a protein means that these modes can be considered as arising from separate molecules within a crystal. IR and Raman active modes within crystals have the constraint that the vibrational modes common to each unit cell must be in-phase, where a unit cell is taken as the two coupled adjacent oscillators, which translates as two neighbouring residues within a protein. This selection rule for the optically allowed vibrational transitions places considerable restraints on the allowed phase differences between the individual oscillators *within* each unit cell, i.e. the allowed phase differences (δ) between each of the coupled oscillators. For β-sheets this phase difference is denoted as δ', taking account of the hydrogen bonding to groups in adjacent β-sheet chains. When the TDMs of a unit cell's amide groups sum to zero the optically allowed transition becomes IR inactive, as sometimes turns out to be the case.

III.6 Amide-I TDC for the Infinite Parallel β-Sheet

Figure III.9 shows a representation of the unit cell of a single chain of an infinite parallel β-sheet, which contains two adjacent peptide groups. As mentioned above, all other chains within the β-sheet must move in-phase for optically active vibrations, i.e. $\delta = 0$. Also, since adjacent unit cells of a chain must move in-phase, they are restricted to a phase difference of 0 or 2π and, because there are two amide groups per unit cell, the phase difference between amide groups can, therefore, be either 0 or π. Depending on the symmetry of the coupled oscillators, they are referred to as either A or B. As such, A(0,0) and B(π,0) refer to the in-phase and out-of-phase combinations. As can be seen from Figure III.8*a*, the in-phase combination has identical individual oscillator contributions to the overall TDM, where the TDM of each individual oscillator is represented by a vector with a centre close to the C=O bond and pointing away from the C=O bond towards the C-N bond by 20^o.[13] These contributions cancel perpendicular to the chain and add parallel to it, resulting in a small overall TDM polarized parallel to the chain. This small overall TDM is responsible for the minor β-sheet IR amide-I band centred at ~1685 cm^{-1}. Figure III.8*b* shows the out-of-phase combination and, in this case, the contribution made by the π group must be multiplied by a phase factor of –1, with the result that the contributions cancel parallel to the chain and add perpendicular to it. The highly polarized overall TDM produced is oriented perpendicular to the chain. It is this transition which gives rise to the main β-sheet IR amide-I band centred at ~1635 cm^{-1}.

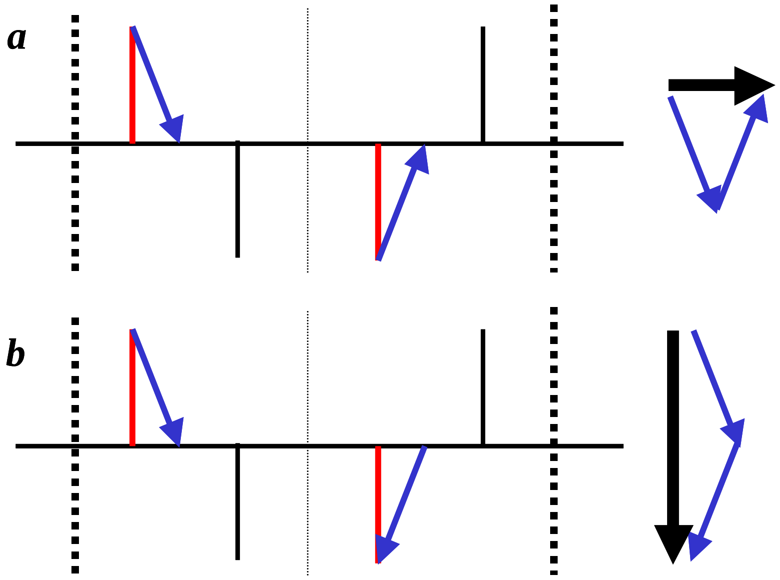
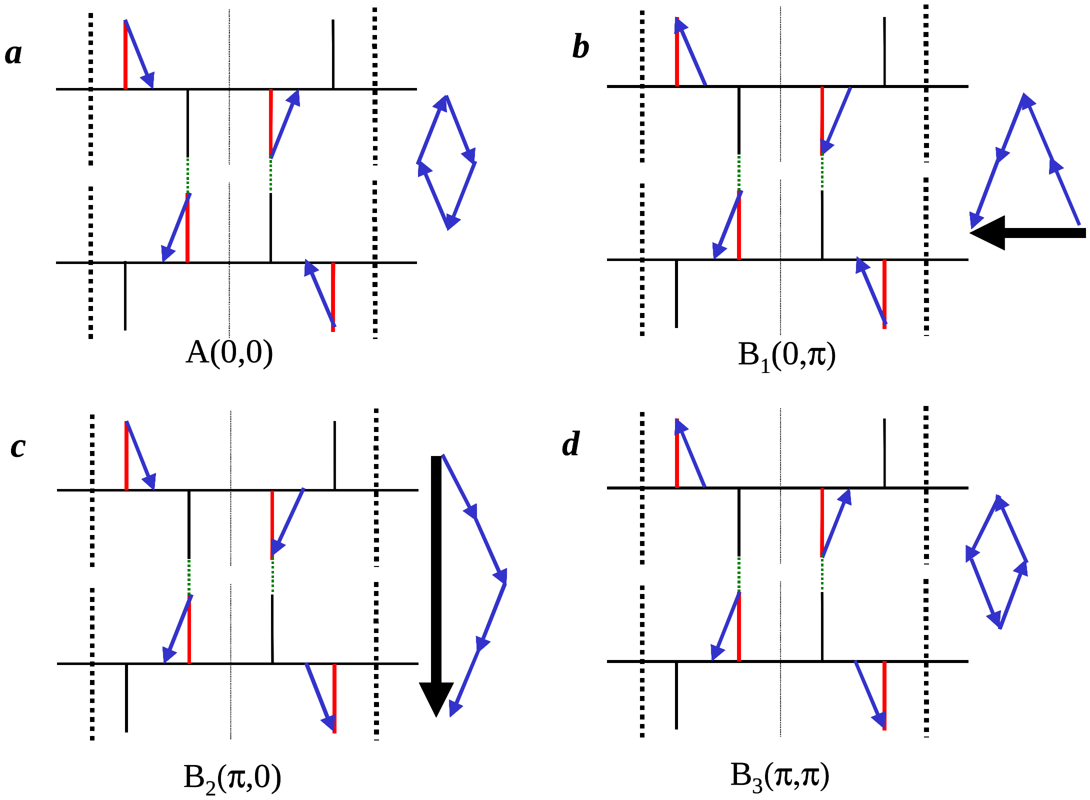


Figure III.8. Scheme of the parallel β-sheet unit cell showing the amide-I vibrational modes A(0,0) and B(π,0). (Modified from[1]) The two peptide groups contained in each unit cell are drawn as red vertical lines, while the individual TDM contributions of each oscillator are drawn as blue arrows. The direction of the arrows indicates the phase of the TDM, with a δ = 0 shown as an arrow pointing towards the chain and a δ = π as an arrow pointing away. The overall TDM is shown as a bold arrow on the right of each schematic, with the direction indicating the direction of polarisation and the length indicating its magnitude. The NH bonds are also shown as black vertical lines joining the horizontal chain. The bold and faint dashed lines indicate that the chain is either rising out of or pointing in towards the plane of the paper, respectively. ***a*** A phase difference of δ = 0 between the two coupled oscillators results in the A(0,0) vibrational mode and the minor component of the amide-I band. ***b*** A phase difference of δ = π produces the B(π,0) mode and the major component of the amide-I band.

III.7 Amide-I TDC for the Infinite Anti-parallel β-Sheet

Figure III.9 shows the unit cell of an antiparallel β-sheet containing four peptide groups and two chains. The requirement that the two unit cells must have either a phase difference of 0 or 2π places restrictions on the allowed phase combinations for an IR active transition. In summing the contributions for each of the four TDMs an overall TDM of either 0 or 2π must result. There is only one possible way to achieve an overall TDM of zero and that is for all the TDM contributions to be themselves zero. This scenario is depicted in Figure III.9*a* and gives rise to the transition A(0,0).

Because individual TDMs may have a phase of either 0 or π, there are three ways to achieve an overall TDM of 2π. Each of these possibilities are shown as Figures III.9*b*, *c* and *d* and the corresponding transitions are referred to as $B_1(0,\pi)$, $B_2(\pi,0)$ and $B_3(\pi,\pi)$, respectively. Each of the four vibrational modes are Raman active.[5]



A(0,0)

$B_1(0,\pi)$

$B_2(\pi,0)$

$B_3(\pi,\pi)$

Figure III.9. Schematic representation of the antiparallel β-sheet unit cell showing the amide-I vibrational modes A(0,0), $B_1(0,\pi)$, $B_2(\pi,0)$ and $B_3(\pi,\pi)$. (Modified from [1; 3]. Figure details are as given for Figure III.8. The hydrogen bonds linking the two chains of the unit cell are shown as dotted green lines.

For the A(0,0) vibrational mode Figure III.9*a* shows that the individual TDMs cancel each other and so the A(0,0) mode is IR inactive. Figure III.9*b* shows that for the $B_1(0,\pi)$ vibrational mode adjacent chains vibrate with a phase difference of π. Both chains have TDMs that cancel perpendicular to the chain and add parallel to it. As such, the overall TDM is small and polarized parallel to the chain. The $B_1(0,\pi)$ vibrational mode gives rise to a weak IR amide-I band centred ~1695 cm^{-1}. Figure III.9*c* shows that the overall TDM of the $B_2(\pi,0)$ vibrational mode consists of individual TDM contributions where there is a phase difference of π between adjacent groups on each chain. For each chain the TDMs add perpendicular to the chain and cancel parallel to it. Therefore, the overall TDM is very large and is highly polarised perpendicular to the β-sheet chains, giving rise to the major β-sheet IR amide-I band centred ~1630 cm^{-1}. From Figure III.9*d* it would appear that the individual TDMs for the $B_3(\pi,\pi)$ vibrational mode, having a similar phase difference of π between adjacent groups on the same chain as for the $B_2(\pi,0)$ vibrational mode, would cancel to give an overall TDM of zero, as in the case of the A(0,0) vibrational mode. While it is true that they do cancel both parallel and perpendicular to the chain, perpendicular to the plane of the paper they add and, as such, the overall TDM is very small and polarized perpendicular to the plane of the paper. The $B_3(\pi,\pi)$ transition produces only a very weak IR absorption band and is rarely detected in a protein's IR spectrum.

III.8 Amide-I TDC for the Infinite α-Helix

Figure III.10 shows a scheme of a unit cell of a hypothetical α-helix having four residues per helix turn. In reality α-helices have five helix turns for every 18 residues and, therefore, the unit cell for an infinite α-helix consists of 18 amino-acid residues. There are two vibrational modes of α-helices arising from there being two phase difference values between adjacent peptide groups, $\delta_1 = 0$ and $\delta_2 = 2\pi/3.6$, with the latter value corresponding to the angle between the lines connecting two adjacent peptide groups to the helix axis. Multiplying δ_2 by 18 gives a value of 10π and this satisfies the condition that all unit cells must be in-phase for an IR active vibration.

Figure III.10*a* illustrates the individual TDMs contributing to the overall TDM of the A species vibrational mode of the four-residue-per-turn model, with a phase difference of zero between adjacent groups. The TDMs cancel perpendicular to the helix axis and add parallel to it, resulting in an overall TDM polarized parallel to the helix axis. Figure III.10*b* shows the E_1 species vibrational mode, where the phase difference between adjacent residues is 2π/4. Therefore, coupling between adjacent residues is forbidden and, in complying with the condition that the phase difference between coupling vibrations must be either 0 or π, coupling is instead to the second next neighbour. As such, energy transfer occurs in such a way as to produce the following excited state transitions: $|A_0C_1> \leftrightarrow |A_1C_0>$ and $|B_0D_1> \leftrightarrow |B_1D_0>$. The result is that two degenerate vibrations are possible, with the overall TDMs of both being polarized perpendicular to the helix axis and to each other. The final vibrational mode for the infinite α-helix is the E_2 species. For this mode $\delta_2 = 4\pi/3.6$ and, therefore, adjacent groups have a phase difference of π and can couple to one another. The individual TDMs contribute in such a way as to cancel both perpendicular and parallel to the helix axis and so the E_2 vibrational mode is IR inactive. Both the E_1 and E_2 species are Raman active.

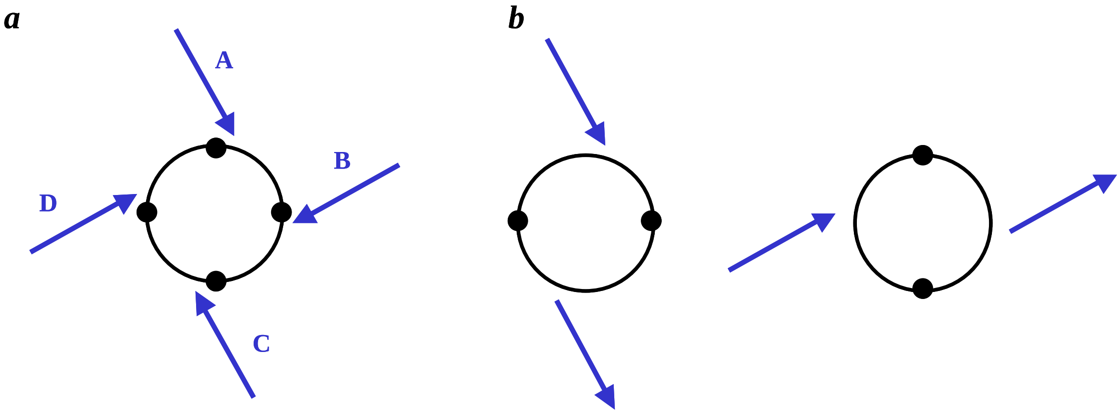


Figure III.10. Schematic representation of the vibrational modes of a hypothetical four-peptides-per-turn helix. (Modified from [2]) The figure details are as given in Figure III.8 with the unfilled circles representing one complete turn about the helix major axis, which is defined by the circle's origin, and the filled circles representing the C=O oscillator. ***a*** δ_1=0. Coupling occurs between adjacent oscillators when the phase difference is zero, giving rise to the A species vibrational mode whose overall TDM is oriented parallel to the helix major axis. ***b*** δ_2=2π/4. Coupling is to the next nearest neighbour when the phase difference between adjacent oscillators is 2π/4, resulting in two degenerate vibrational modes with overall TDMs perpendicular to each other and to the helix major axis, known as the E_2 species vibrational mode.

III.9 Fourier Transform Infrared Spectroscopy

III.9.1 The Michelson Interferometer

Over 95 % of all mid-infrared spectrometers in use today are Fourier transform infrared (FTIR) spectrometers. In the same way that a monochromator is central to a dispersive spectrometer, the heart of an FTIR spectrometer is the Michelson interferometer, a schematic of which is shown in Figure III.11. Radiation emitted from the IR source (commonly a glow bar) is reflected by a mirror onto a semi-permeable beamsplitter. Half of the IR radiation passes through this beamsplitter to contact a moving mirror and half is reflected onto a fixed mirror. The radiation is then reflected from both mirrors and recombines at the beamsplitter. From there, half of the light is directed back to the source, and is lost, and half is sent through the sample and then focused onto the detector using another mirror.

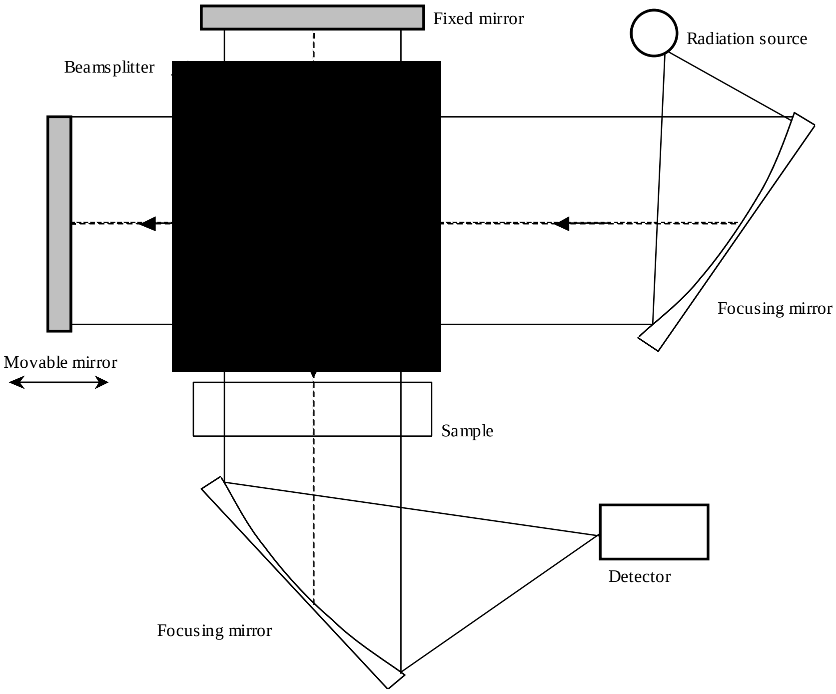


Figure III.11. Schematic representation of an FTIR spectrometer with a Michelson interferometer. The arrows show the direction of light propagation.

III.9.2 The Interferogram

FTIR spectroscopy is an interference technique. An interference effect occurs when photons of radiation interact with one another, where photons can either constructively or destructively interfere. The degree to which each photon is in-phase determines the degree to which it constructively interferes and the reverse is true for destructively interfering out-of-phase photons. Because of the wave-like motion of light, the phase of a photon at the beamsplitter after returning from either the fixed or moving mirror depends upon the absolute distance it has travelled in making the journey from the beamsplitter to the mirror and back again. Therefore, since the movement of the moving mirror introduces a variable pathlength for the radiation travelling from the beamsplitter to the moving mirror and back again, there will also occur a correspondent variation in radiation phase at the beamsplitter. And because the pathlength between the beamsplitter and the fixed mirror is constant, radiation arriving at the beamsplitter from the fixed mirror will also be of a constant phase. When radiation from the moving mirror recombines with radiation from the fixed mirror at the beamsplitter, the degree to which constructive and destructive interference occurs depends upon the relative phases of the photons arriving from each mirror and, therefore, on the position of the moving mirror. The precise constructive and destructive photon interactions for each incremental position of the moving mirror along its complete path determine what frequencies of IR light survive, and to what relative intensities, to be focused onto the sample. The sum total of all the radiation received by the detector when the moving mirror has made a complete movement along its path is referred to as the interferogram. In this manner the exciting radiation is encoded into an interferogram and it is the function of the Fourier transform mathematical procedure to decode this interferogram into a spectrum.

A mid-IR interferogram is shown in Figure III.12, where the moving mirror has moved through a full displacement, as defined by the distance extremities of the moving mirror from the beamsplitter. The x-coordinate corresponds to the pathlength from the beamsplitter to the moving mirror and the y-coordinate is the overall photon intensity produced by interference at each position of the moving mirror. The peak maximum located at the centre of the interferogram is referred to as the centre-burst and it corresponds to the position of the moving mirror at which the pathlength between moving mirror and beamsplitter is precisely equal to the pathlength between the beamsplitter and the fixed mirror. As such, all photons from the moving mirror will be in-phase with photons from the fixed mirror when they recombine at the beamsplitter and so all photons will interact constructively, giving rise to a maximum of photon intensity. The so-called wings of the interferogram correspond to positions of the moving mirror that give rise to partial destructive interference at the beamsplitter and, therefore, a decreased photon intensity signal. In terms of FTIR spectroscopy, a good interferogram has an intense centre-burst, with a small number of wings and a flat baseline, and spectra produced from the Fourier transforms of such interferograms are deemed reliable.

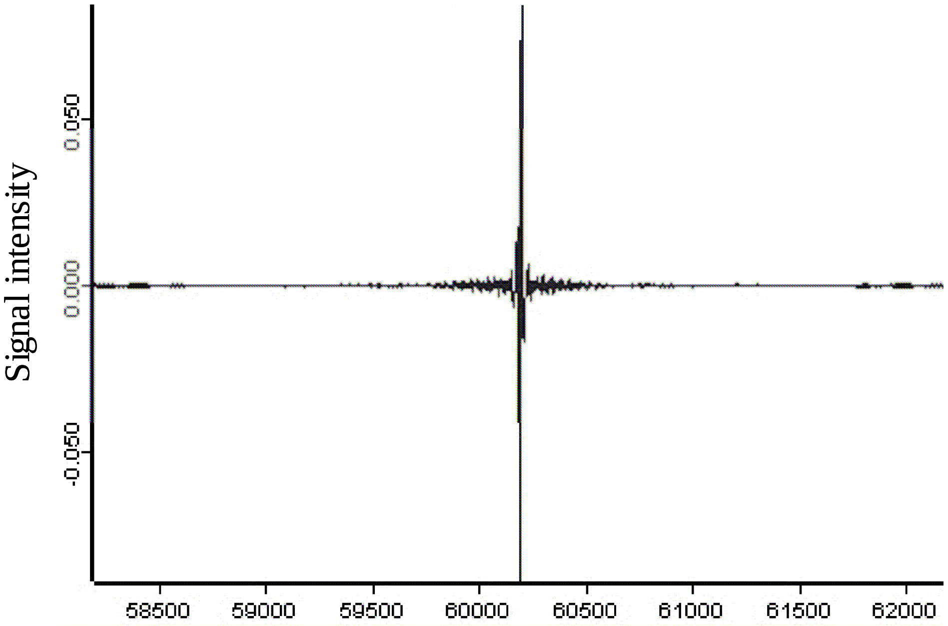


Figure III.12. A Mid-IR interferogram from a Tensor 27 FTIR

Modern FTIR spectrometers produce digitised interferograms and the digitised Fourier transform (DFT) is given by

$$S(k\cdot\Delta\tilde{\nu}) = \Sigma I(n\cdot\Delta x)\exp(i2\pi nk/N) \tag{III.19}$$

where $n\cdot\Delta x$ and $k\cdot\Delta\tilde{\nu}$ represent the discrete interferogram and spectral points, respectively, and N is the number of discrete points within the interferogram. Unlike the regular Fourier transform, the DFT is not a continuous function; since it is only calculated for a limited number of values, n, and this leads importantly to the picket-fence effect and leakage.

I.9.3 The Picket-Fence Effect and Zero-Filling

When the interferogram contains frequency components that do not exactly coincide with the data point positions of the spectrum, $k\cdot\tilde{\nu}$, the result is the picket-fence effect. It is so named because the spectrum produced is absent of those frequencies lying between the data points $k\cdot\tilde{\nu}$, in analogy to a spectrum as viewed from behind a picket-fence, in that certain regions are shielded from view. To correct for this effect, which in the worst case can lead to a signal reduction of around 36 %, a procedure known as zero-filling is performed. Zero-filling reduces the picket-fence effect by adding zeros to the end of the interferogram, with the result that the DFT decoding returns an interpolated spectrum that has a greater number of points per wavenumber. A greater amount of zeroes added to the end of the interferogram causes a correspondingly greater reduction of the picket-fence effect. Since zero-filling interpolates by using the instrument line-shape function (ILS), it should not be regarded as simply a smoothing process and it is considered to be more appropriate than either polynomial or spline interpolation methods, which act directly on the spectrum, as opposed to the interferogram.

I.9.4 Apodization and the Leakage Effect

Apodization is a process that smoothly brings the interferogram down to zero at the edges of the sampled region. However, apodization is performed at the expense of spectral resolution, since it has the effect of widening spectral lines. Because the moving mirror travels a finite pathlength, the experimental interferogram is available only over a limited length. To arrive at this finite interferogram mathematically requires that the infinite interferogram function be multiplied by a rectangular function that is unity over the experimental interferogram region and zero everywhere else – referred to as a boxcar truncation apodization function. An apodization function is an artificial weighting of the interferogram, such that when the Fourier transform of the weighted interferogram (apodization function) is taken the ILS is returned. In order to achieve the best possible spectral resolution it is desirable that a boxcar apodization function be used, since this allows for the entire interferogram to be used in reproducing the spectrum – at all points a weighting factor of one. However, as can be seen from Figure III.13, using the boxcar truncation apodization function gives rise to a maximum leakage effect.

Leakage is a process that leads to oscillations around the base of the ILS band, giving rise to the sidelobes shown in Figure III.13. Side lobes result from measuring the interferogram only up to a certain point and these are minimized by increasingly reducing the weighting of the interferogram when approaching its edges. Figure III.13 also shows other apodization functions available for determining the ILS function that are better able to reduce the leakage effect, but in all cases there exists a trade off between minimizing the leakage effect and obtaining maximum spectral resolution, and achieving one is always at the expense of sacrificing the other. It is a matter of choice and the particular experiment system in question as to which is the best apodization function to be used, with more drastic apodization functions being required when quantitative evaluation of closely spaced spectral lines is required and minimum apodization when a qualitative analysis is wanted. Because protein FTIR does not require very high spectral resolution, the Blackman apodization function was chosen as the most suitable, since this function involves relatively heavy apodization, thus minimizing the leakage effect. The apodization and ILS formulae for each of the apodization functions are shown in Table III.1.

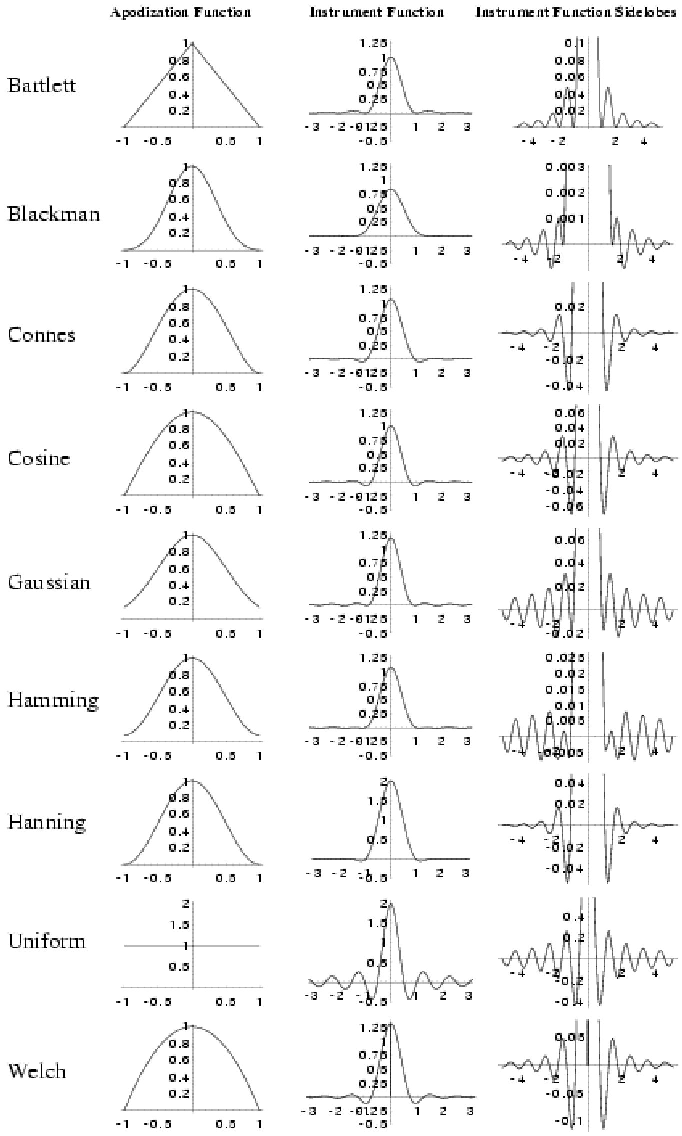


Figure III.13. Apodization and instrument function plots. (Wolfram Math World) The side lobes of each ILS are shown on a magnified scale on the right of the figure. Uniform corresponds to the Boxcar apodization.

Apodization	**Function**	**ILS function**	**ILS FWHM**
Bartlett	$1-\frac{\lvert x\rvert}{a}$	$a\ \mathrm{sinc}^2(\pi k a)$	1.77179
Blackman	$B_A(x)$	$B_I(k)$	2.29880
Connes	$\left[1-\frac{x^2}{a^2}\right]^2$	$8a\sqrt{2\pi}\ \frac{J_{S/2}[2\pi k a]}{[2\pi k a]^{S/2}}$	1.90416
Cosine	$\cos\left(\frac{\pi x}{2a}\right)$	$\frac{4a\cos[2\pi a k]}{\pi\left(1-16a^2k^2\right)}$	1.63941
Gaussian	$e^{\frac{-x^2}{[2\sigma^2]}}$	$2\int_0^a \cos(2\pi k x)e^{-\frac{x^2}{(2\sigma^2)}}dx$	
Hamming	$\mathrm{Hm}_A(x)$	$\mathrm{Hm}_I(k)$	1.81522
Hanning	$\mathrm{Hn}_A(x)$	$\mathrm{Hn}_I(k)$	2.00000
Boxcar	1	$2\ a\ \mathrm{sinc}\left(2\pi k a\right)$	1.20671
Welch	$1-\frac{x^2}{a^2}$	$W_I(k)$	1.59044

Table III.1. FTIR apodization functions (Modified from Wolfram Math World)

III.9.5 Phase-Correction

Owing to particular beamsplitter characteristics and, also, to electronic effects introduced when processing the interferogram before it becomes digitised, interferograms routinely display asymmetry about the centreburst brought about by a frequency-dependent phase delay effect. This effect causes a slight shifting of the interference patterns of different wavelengths with respect to one another and introduces sinusoidal components into the cosinusoidal output from the interferometer. It is corrected for by using a “phase-correction” algorithm, the most commonly used is the “Mertz phase correction”.

III.9.6 FTIR Spectral Resolution

The resolution of an IR dispersive spectrum depends upon the monochromaticity of the sampling radiation and, therefore, on the focal length slit width of the monochromator. Narrowing the slit width allows a smaller range of frequencies of light to pass through to the sample compartment, thereby increasing the monochromaticity of the sampling radiation. The nominal resolution of an FTIR spectrum, by contrast, depends upon the pathlength of the moving mirror in acquiring that spectrum, according to the equation $R = 1/L$, where L is the moving mirror pathlength. Therefore, in order to improve the spectral resolution by a factor of two one needs to double the distance over which the moving mirror must move. Because for any FTIR spectrometer there is a maximum moving mirror pathlength, there is a correspondent maximum resolution obtainable for a given instrument.

III.9.7 Advantages of FTIR Spectroscopy

The main advantages of using FTIR over dispersive IR are the increased signal-to-noise ratio of FTIR spectra, their superior accuracy and their shorter acquisition times. Also, because the sampling position of FTIR spectrometers is typically located behind the modulating interferometer, the scattering effects for FTIR spectra are considerably less than for dispersive IR spectra. Increased signal-to-noise ratio is a consequence of using a greater proportion of the radiation source output in collecting the IR spectrum. This technique is referred to as multiplexing and it gives the so-called FELLGETT’S advantage.

With dispersive IR spectroscopy a monochromator is used to select a single wavelength of the entire source output and this monochromatic light is then passed through the sample. Therefore, only a small percentage of the produced light goes towards generating the spectral band. However, the spectral noise is independent of what percentage of the produced light is used during spectral acquisition and, therefore, in the case of dispersive IR spectroscopy, the signal-to-noise ratio tends to suffer as a consequence of discarding so much light within the monochromator. The multiplexing technique of FTIR uses a far higher percentage of the light produced by the source in generating the interferogram and it is the interferogram that is used to investigate the light absorbing properties of the sample, being subsequently decoded by the Fourier transform procedure to reproduce the IR spectrum. Therefore, the signal-to-noise ratio of an FTIR spectrum is significantly enhanced compared to that of an IR spectrum.

Another advantage of using FTIR rather than dispersion IR spectroscopy comes from the fact that the circular apertures used in FTIR spectrometers have larger areas than the slits used in grating monochromators and, therefore, higher radiation throughput is achieved with FTIR. This is referred to as the JACQUINOT advantage and it acts in concert with the FELLGETT advantage in improving the signal-to-noise ratio of FTIR spectra, which can be up to ten times less noisy than corresponding dispersion IR spectra.

The superior accuracy of the wavenumber scale of FTIR spectrometers arises from its dependence on the interferogram of a built-in HeNe laser. The distance between zero-crossings of the HeNe laser interferogram determines the sampling interval of the FTIR interferogram, and the point spacing in the resulting spectrum is inversely proportional to this. Assuming that the FTIR spectrometer is correctly zeroed it should have a wavenumber scale precise to a few hundredths of a wavenumber, and this intrinsic precision of FTIR spectroscopy is referred to as the CONNES' advantage. Generally, if the position of the HeNe laser is centred as 15,798.002 cm^{-1} the FTIR spectrometer is regarded as being accurate to ±0.1 cm^{-1}.[14]

Because during the acquisition of a dispersive IR spectrum the sample has to be separately exposed to radiation of each wavelength, the acquisition time of a dispersive IR spectrum is greatly increased compared to that of an FTIR spectrum. This increased acquisition time becomes problematic where there are issues of sample stability and, also, makes the performance of time-resolved IR spectroscopy implausible. With FTIR, spectra can be acquired in approximately one second for a single scan spectrum. Also, with expensive and complex instrumentation, a step-scan experiment can be performed for the acquisition of time-resolved FTIR spectra with nanosecond time-resolution, and this is currently being developed at NUI Maynooth. Time-resolved FTIR has been used by many researchers in studying the folding and unfolding mechanisms of proteins at nanosecond time-resolution. For example, along with laser excitation, time-resolved FTIR has been used in studying the conformational changes of the light-driven proton pump of bacterior rhodopsin [15], the redox-driven proton pump cytochrome-c-oxidase [16] and the photosynthetic reaction centre [17].

III.10 Attenuated Total Reflectance Spectroscopy

Attenuated total reflectance (ATR) spectroscopy is a technique used to ensure reliable pathlength reproducibility where very short optical pathlengths must be used. It makes use of the particular refractive indices of various materials and samples in generating evanescent waves that penetrate only very small, but highly reproducible distances into samples placed in contact with ATR crystals.

III.10.1 Snell's Law

The direction of a beam of light is altered upon passing from one medium to another in a manner that depends upon the angle the light travels relative to the boundary between the two media. This angle is referred to as the incidence angle (Θ) and is related to the refraction angle (Φ), the angle by which the path of light is altered, by Snell's Law, which holds that

$$n_1 . \sin\Theta = n_2 . \sin\Phi \qquad \text{(III.20)}$$

where n_1 and n_2 are the refractive indices of medium 1 and 2, respectively. Figure III.14 illustrates this relationship when the media consist of an ATR crystal in contact with a sample solution and from this it can be seen that beam c has an incidence angle at which the beam is neither reflected back into the ATR crystal nor propagates through the sample solution. This angle is referred to as the critical angle (Θ_G) and total reflection of an incident beam occurs at the boundary between two media at any angle that exceeds this critical angle. At the critical angle the angle of refraction is exactly $90^o$and, therefore, using Snell's Law

$$\sin\Theta_G = \frac{n_2}{n_1} \qquad \text{(III.21)}$$

Where ZnSe is used as the ATR crystal material, Θ_G equals 38^o when samples having an n value of 1.5 are used. The reflection process is illustrated by beam d and it is by this mechanism that radiation may be passed through an ATR crystal and then be redirected and collected by a detector. Typically the beam is returned to the boundary between medium 1 and 2 numerous times by the same reflection principle by placing another material at the opposite side of the ATR crystal that has the same n value as that of the sample solution.

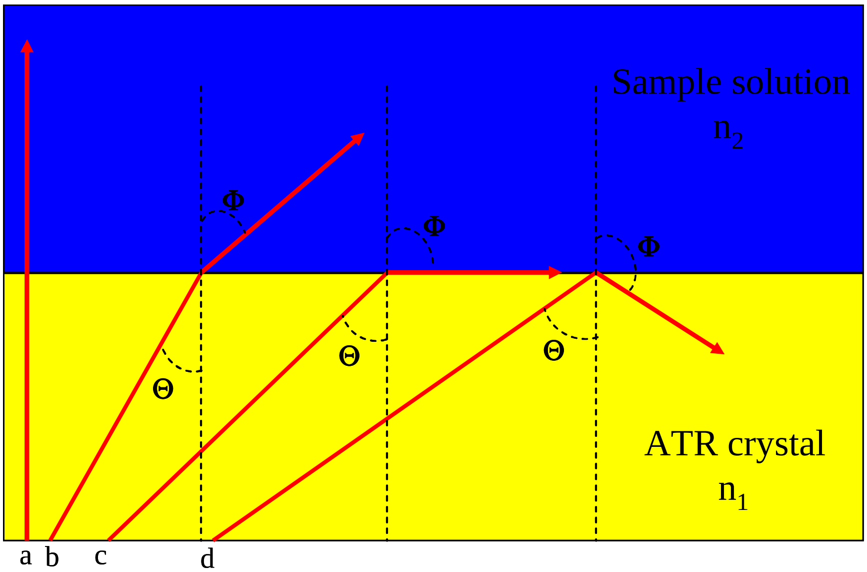


Figure III.14. Refraction and reflection of light within an ATR crystal. n_1 and n_2 are the refractive indices of the ATR crystal and the sample solution, respectively. Θ and Φ are the angles of incidence and reflection, respectively.

III.10.2 The Harrick Approximation and Penetration Depth

At the boundary point between the two media where the beam is reflected an evanescent wave is generated that travels a very small distance into the sample solution, before travelling the same distance back to the boundary and re-entering the ATR crystal at the same angle with which it exited. This distance of penetration (d_p) can be calculated from the Harrick approximation as

$$d_p = \frac{\lambda}{2\pi n_p \left(\sin^2\theta - n_{sp}{}^2\right)^{\frac{1}{2}}} \qquad \text{(III.22)}$$

where λ is the wavelength, n_p is the crystal refractive index, θ is the incidence angle and n_{sp} is the refraction index ratio between the sample and crystal. d_p is defined as the distance between the sample surface and the position where the intensity of the penetrating evanescent wave dies off to $(1/e)^2$ or 13.5 %, or its amplitude has decayed to 1/e. The Harrick approximation predicts that ATR crystals having higher refractive indices will generate evanescent waves with shallower penetration depths. For ZnSe crystals the refractive index at 1000 cm^{-1} is 2.4 and the penetration depths of beams with incidence angles of 45° and 60° are 1.66 and 1.04 µm, respectively.

It is also evident from the Harrick approximation that the value of d_p has a linear dependence on the wavelength of the incident light. The spectral effect of this wavelength dependence of d_p can be seen from the comparison of the transmission and ATR spectrum of the same sample (Figure III.15). At shorter wavelengths (higher frequency) the penetration depth is smaller and so the absorbance of the bands from 2800-3000 cm^{-1} is reduced, relative to the transmission spectrum, due to a decrease in the sample pathlength. The opposite is true at longer wavelengths (lower frequency) and the bands from 1400-1500 cm^{-1} have higher absorbance for the ATR spectrum than for the transmission spectrum. The overall effect on the ATR absorption spectrum is to alter the relative intensities of bands in a wavelength dependent manner, increasing the band intensities at lower frequency while decreasing the intensities of bands at higher frequencies. It is possible to correct for this effect by creating an algorithm that takes account of the frequency dependence of the penetration depth of the evanescent wave and which post-processes the ATR spectrum to return the transmission spectrum.

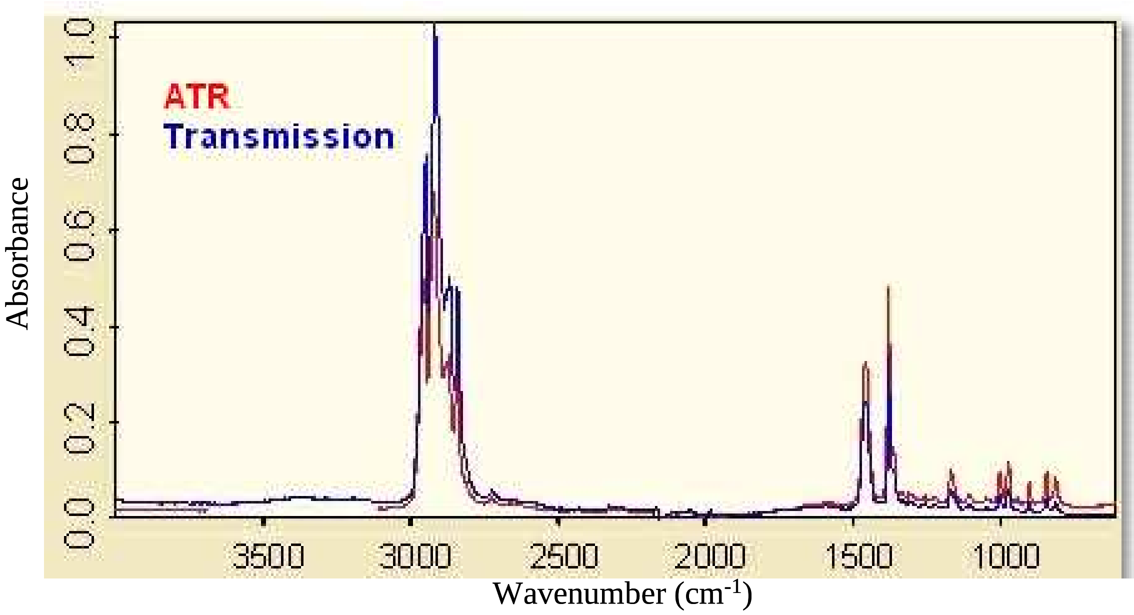


Figure III.15. The wavelength dependence of penetration depth for ATR FTIR. The ATR and transmission absorption spectra are shown as red and blue, respectively (Bruker OPUS software).

III.11 Perturbation Based Two-Dimensional Spectroscopy

Perturbation two-dimensional (2D) spectroscopy relies upon the application of a perturbing influence to a system that is being monitored spectroscopically in order to stimulate spectral variation along the perturbation coordinate and the generation of dynamic spectra, which are then used to generate a 2D spectrum, as described below. Examples of such perturbing influences are temperature, pH, pressure, electricity, magnetism and ligand titration, and their effects are detected by changes induced in the spectrum of the system. This spectral variation is referred to as a dynamic spectrum and it may be defined as a direct function of the imposed physical effect over any range of the total applied perturbation or it might be created to reflect the temporal spectral profile of the system at each incremental point of the applied perturbation. The dynamic spectrum is usually defined for each spectral wavelength of the system's spectrum and the interval between successive dynamic spectra determines the resolution of the 2D plot.

III.11.1 The Dynamic Spectrum

The dynamic spectrum $\tilde{y}(\upsilon,t)$of a system that has experienced some perturbation by an external variable t is formally defined as

$$\tilde{y}(\upsilon,t) = y(\upsilon,t) - \tilde{y}(\upsilon) \quad \text{for } T_{\min} \le t \le T_{\max},$$

$$0 \qquad \text{otherwise} \qquad \text{(III.23)}$$

where $\tilde{y}(\upsilon)$ is the reference spectrum of the system and T_{min} and T_{max} define the range of variation of the external variable. Where $\tilde{y}(\upsilon)$ is set equal to zero the dynamic spectrum corresponds to the observed variation of spectral intensity at a given wavelength, or spectral variable. $\tilde{y}(\upsilon)$ may also be defined as the system spectrum before the application of the external perturbation (the system ground state). Alternatively, an averaged spectrum may be used for $\tilde{y}(\upsilon)$, given by

$$\tilde{y}(\upsilon) = \frac{1}{T_{\max} - T_{\min}} \int_{T_{\min}}^{T_{\max}} y(\upsilon,t)dt \qquad \text{(III.24)}$$

III.11.2 The 2D Correlation Spectrum

2D correlation spectroscopy can be understood as a quantitative comparison of the patterns of spectral intensity variation along the external variable t observed at two different spectral variables, υ_1 and υ_2, over some finite observation interval between T_{min} and T_{max}. The 2D correlation spectrum is formally defined as

$$\mathrm{X}(\upsilon_1,\upsilon_2) = \left\langle \tilde{y}(\upsilon_1,t) \cdot \tilde{y}(\upsilon_2,t') \right\rangle \qquad \text{(III.25)}$$

The intensity of the 2D correlation spectrum depends upon the similarity or dissimilarity of the dynamic spectra generated at υ_1 and υ_2. The $\langle \; \rangle$ symbol denotes a cross-correlation function that compares the dependence of two different quantities on *t*.

Treating the 2D correlation spectrum as a complex number function gives

$$\mathrm{X}\left(\upsilon_1,\upsilon_2\right)=\Phi\left(\upsilon_1,\upsilon_2\right)+i\Psi\left(\upsilon_1,\upsilon_2\right) \tag{III.26}$$

where $\Phi\left(\upsilon_1,\upsilon_2\right)$ is the real component that is orthogonal to the imaginary component $i\Psi\left(\upsilon_1,\upsilon_2\right)$ and these are known respectively as the synchronous and asynchronous 2D correlation intensities. The overall similarity between the two dynamic spectra at υ_1 and υ_2 is reflected in the synchronous 2D correlation intensity $\Phi\left(\upsilon_1,\upsilon_2\right)$, while the overall dissimilarity between these dynamic spectra determines the asynchronous 2D correlation intensity $i\Psi\left(\upsilon_1,\upsilon_2\right)$.

The formal definition of the synchronous and asynchronous correlation intensities is given by

$$\Phi\left(\upsilon_1,\upsilon_2\right)+i\Psi\left(\upsilon_1,\upsilon_2\right)=\frac{1}{\pi\left(T_{\max}-T_{\min}\right)}\int_0^{\infty}\tilde{Y}_1\left(\omega\right)\cdot\tilde{Y}_2\left(\omega\right)d\omega \tag{III.27}$$

where $\tilde{Y}_x$ is the forward Fourier transform of the dynamic spectrum $\tilde{y}\left(\upsilon_x,t\right)$, as given by

$$\begin{aligned}\tilde{Y}_x\left(\omega\right)&=\int_{-\infty}^{\infty}\tilde{y}\left(\upsilon_x,t\right)e^{-i\omega t}\\&=\tilde{Y}_x^{\mathrm{Re}}\left(\omega\right)+\tilde{Y}_x^{\mathrm{Im}}\left(\omega\right)\end{aligned} \tag{III.28}$$

where $\tilde{Y}_x^{\mathrm{Re}}\left(\omega\right)$ and $\tilde{Y}_x^{\mathrm{Im}}\left(\omega\right)$ are the real and imaginary components of the Fourier transform, respectively. $\tilde{Y}_x^{\mathrm{Re}}\left(\omega\right)$ corresponds to an even function of ω, while $\tilde{Y}_x^{\mathrm{Im}}\left(\omega\right)$ is an odd function.

III.11.3 Properties of the Synchronous 2D Correlation Spectrum

Figure III.16 shows a schematic representation of a sample synchronous 2D contour map. The diagonal line is defined by coordinates where $\upsilon_1 = \upsilon_2$. Correlation peaks are shown as concentric circles and they can be either positive (clear) or negative (shaded) in sign. Correlation peaks occurring along the diagonal line are referred to as autopeaks, since they correspond mathematically to the autocorrelation function of spectral intensity variation observed during the interval between T_{min} and T_{max}. Taking a slice through the diagonal of a synchronous 2D map returns the autopower spectrum. The magnitude of autopeaks is always positive and it represents the overall extent of spectral intensity variation observed at the specific spectral variable υ during the observation interval between T_{min} and T_{max}. Strong autopeaks will be present for the spectral regions that undergo considerable spectral variation between T_{min} and T_{max}; while for regions where there occurs only minor intensity variations autopeaks will be weak or non-existent.

Peaks located off the diagonal line are referred to as cross peaks and, in a synchronous 2D map, they represent simultaneous or coincidental changes of spectral intensity observed at two different spectral variables υ_1 and υ_2. The coincidence of such spectral variation suggests a possible coupling between the events giving rise to the intensity changes. In order to illustrate the coherent variation of spectral intensities at these spectral variables, a correlation square can be constructed, whereby the pair of cross peaks located at opposite side of the diagonal line are connected to the autopeaks located at each of the spectral variables. In Figure III.16 correlation squares show the correlation of A with B and C with D.

Unlike autopeaks, cross peaks may be of either positive or negative sign. When the spectral intensities at both υ_1 and υ_2 (corresponding to the coordinates of the cross peak on the 2D correlation map) either increase or decrease together as functions of the external variable t the sign of the cross peak is positive. If the intensity variation at υ_1 occurs in the opposite direction to that of υ_2, however, the cross peak sign is negative. Therefore, in Figure III.16 B and D spectral regions either increase or decrease together while moving from T_{min} to T_{max}. Whereas the spectral regions labelled A and C experience spectral variation such that A is increasing in intensity while C is decreasing, or vice versa.

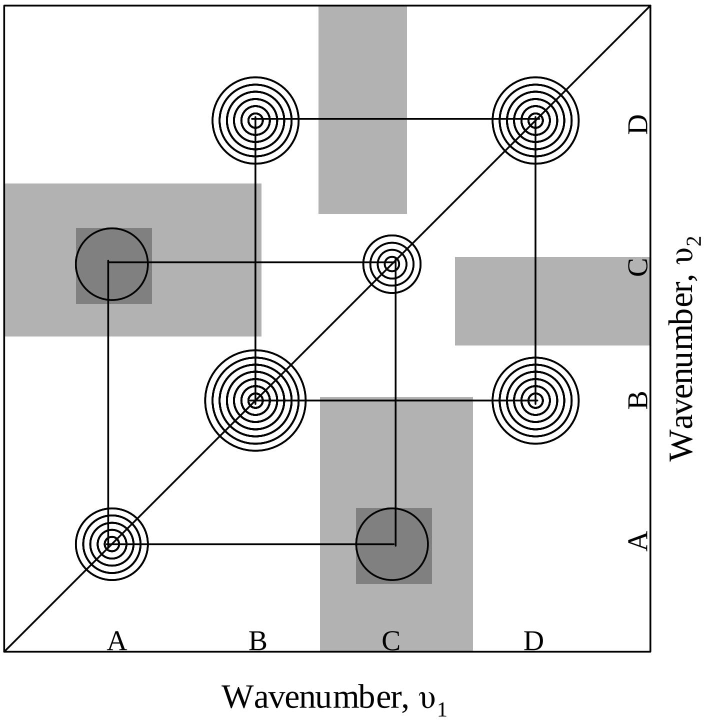


Figure III.16. Schematic representation of a synchronous 2D correlation spectrum. Where the map is positive the area is left clear and where it is negative the map is shaded. Therefore, shaded peaks are peaks of negative sign, whereas positive peaks are clear.

III.11.4 Properties of the Asynchronous 2D Correlation Spectrum

Figure III.17 shows a schematic representation of an asynchronous 2D correlation spectrum. Whereas the synchronous 2D correlation spectrum represents coincidental changes in intensity variations, the asynchronous 2D correlation spectrum represents sequential or successive changes measured separately at υ_1 and υ_2. Unlike a synchronous spectrum, an asynchronous spectrum is anti-symmetric with respect to the diagonal line and contains no autopeaks, consisting exclusively of cross peaks. In the same manner as above, asynchronous correlation squares can be constructed and from Figure III.17 it can be seen that asynchronous correlation is observed for band pairs A and B, A and D, B and C, and C and D, and that four asynchronous correlation squares can be drawn.

An asynchronous cross peak will only result if the intensity variations at υ_1 and υ_2 occur out of phase with each other (i.e. at a lower or higher temperature range, if temperature is the external variable and the dynamic spectrum is a direct function of the temperature perturbation). As such, asynchronous cross peaks are very effective in resolving overlapped bands that have different origins. For example, in protein FTIR there occurs considerable overlap in amide-I signals arising from the symmetric CO stretching vibration of the peptide backbone of CO groups residing in different protein secondary structure environments. Altering the physical environment of the protein may result in the loss of some secondary structural types and the formation of others. By generating an asynchronous 2D correlation spectrum it is possible to resolve the amide-I bands of the separate secondary structural types if the loss and gain of the different structural types occur sequentially. The resolving potential of asynchronous 2D correlation spectroscopy is very high and the resolution is determined by the degree to which spectral variations occur sequentially.

Similar to the synchronous 2D correlation spectrum, the sign of asynchronous cross peaks can be either positive or negative and this distinction is very useful because it provides information on the sequential order of events stimulated by performing the perturbation over a given range. If the asynchronous cross peak is positive it can be taken that the intensity change at υ_1 occurs predominantly before that at υ_2. On the other hand, if the cross peak is negative the reverse is true. A caveat to these rules is that the synchronous correlation spectrum must be greater or equal to zero at the same coordinate as the asynchronous cross peak. Where this is not the case and $\Phi\left(\upsilon_1,\upsilon_2\right) < 0$ then the rules are reversed and a positive (negative) asynchronous cross peak should be taken as evidence that spectral intensity variation at υ_2 occurs predominantly before (after) that at υ_1. Therefore, in Figure III.17 the intensity changes at bands A and C occur after those at B and D. It should be noted, however, that the above set of sequential order rules can only be applied if the patterns of spectral intensity variations during the perturbation range selected for composition of the 2D correlation spectrum are generally monotonic. In other words, spectral intensity variation should be unidirectional at each spectral variable of interest.

The description of 2D correlation spectroscopy given above in some ways assumes ideal spectral behaviour along the perturbation coordinate. Where spectral features such as band position shifts or line shape changes occur conventional rules of 2D spectroscopy do not apply. It is possible to modify the analysis in order to take account of such anomalous behaviours by analysing correlation peaks in clusters instead of individual peaks and a description of this procedure is provided by Noda.[18]

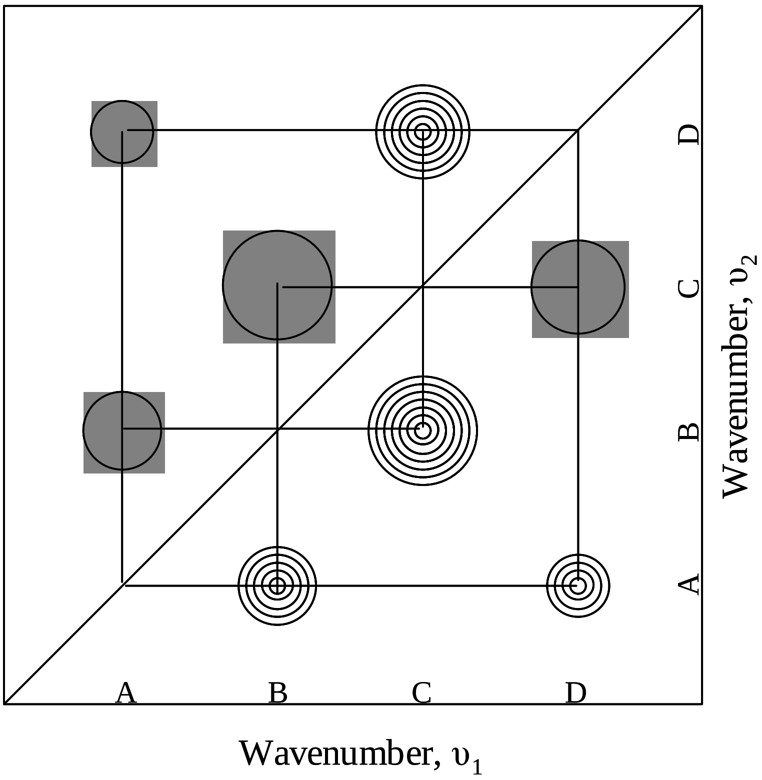


Figure III.17. Schematic representation of an asynchronous 2D correlation spectrum. Positive cross peaks are shown as clear and negative cross peaks are shaded.

References

1 Miyazawa T. Perturbation Treatment of the Characteristic Vibrations of Polypeptide Chains in Various Configurations. Journal of Chemical Physics 32(6):1647-1652, 1960.

2 Barth A, Zscherp C. Substrate binding and enzyme function investigated by infrared spectroscopy. Febs Letters 477(3):151-156, 2000.

3 Chirgadze YN, Nevskaya NA. Infrared-Spectra and Resonance Interaction of Amide-One Vibration of Anti-Parallel-Chain Pleated Sheet. Biopolymers 15(4):607-625, 1976.

4 Barth A, Zscherp C. What vibrations tell us about proteins. Quarterly Reviews of Biophysics 35(4):369-430, 2002.

5 Krimm S, Bandekar J. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. Advances in Protein Chemistry 38:181-364, 1986.

6 Abe Y, Krimm S. Normal Vibrations of Crystalline Polyglycine-I. Biopolymers 11(9):1817-&,1972.

7 TorII H, Tatsumi T, Tasumi M. Effects of hydration on the structure, vibrational wavenumbers, vibrational force field and resonance Raman intensities of N-methylacetamide. Journal of Raman Spectroscopy 29(6):537-546, 1998.

8 TorII H, Tatsumi T, Kanazawa T, Tasumi M. Effects of intermolecular hydrogen-bonding interactions on the amide I mode of N-methylacetamide: Matrix-isolation infrared studies and ab initio molecular orbital calculations. Journal of Physical Chemistry B 102(1):309-314, 1998.

9 Jackson M, Mantsch HH. Beware of Proteins in Dmso. Biochimica Et Biophysica Acta 1078(2):231-235, 1991.

10 Parrish JR, Blout ER. Conformation of Poly-L-Alanine in Hexafluoroisopropanol. Biopolymers 11(5):1001-&, 1972.

11 Hamm P, Lim MH, Hochstrasser RM. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. Journal of Physical Chemistry B 102(31):6123-6138, 1998.

12 Lee SH, Krimm S. General treatment of vibrations of helical molecules and application to transition dipole coupling in amide I and amide II modes of alpha-helical poly(L-alanine). Chemical Physics 230(2-3):277-295, 1998.

13 TorII H, Tasumi M. Model-Calculations on the Amide-I Infrared Bands of Globular-Proteins. Journal of Chemical Physics 96(5):3379-3387, 1992.

14 Bertie JE. Specification of components, methods and parameters in Fourier transform spectroscopy by Michelson and related interferometers (Technical Report). Pure and Applied Chemistry 70(10):2039-2045, 1998.

15 Deisenhofer J, Michel H. The Photosynthetic Reaction Center from the Purple Bacterium Rhodopseudomonas-Viridis. Science 245(4925):1463-1473, 1989.

16 Wikstrom M. Proton-coupled electron transfer in cytochrome c oxidase. Abstracts of Papers of the American Chemical Society 216:U95-U95, 1998.

17 Mantele, W. Infrared Vibrational Spectroscopy of Reaction Centres. The Photosynthetic Reaction Center Vol.II, eds J. Deisenhofer and J. Norris, Academic Press, New York, 23-282, 1993.

18 Noda, I. and Ozaki, Y. TWO-DIMENSIONAL CORRELATION SPECTROSCOPY, Applications in Vibrational and Optical Spectroscopy. Wiley, Chapter 4, 2004.

Chapter IV

Experimental Instrumentation

IV.1 Luminescence Instrumentation

IV.1.1 Steady-State Luminescence

Figure IV.1 shows a schematic of the instrumentation used for the acquisition of steady-state luminescence for the fluorescence resonance energy transfer (FRET) experiments. The set-up consists of an L6310 D_2 ultra-violet light source (Hamamatsu), a 300i SpectraPro dispersal monochromator (Acton Research Corporation) for excitation, an SC-447 Sample Chamber (Acton Research Corporation), a single photon counting (SPC) R928-P photon multiplier tube (PMT) (Hamamatsu) detector for excitation, a 500i SpectraPro dispersal monochromator (Acton Research Corporation) for emission, an IP28 PMT (Hamamatsu) detector for emission and an NCL Spectral Measuring System (Acton Research Corporation) for controlling both the excitation and emission monochromators.

The 500i and 300i SpectraPro monochromators have focal lengths of 500 and 300 mm, respectively, and are capable of operating in the wavelength range of 180 nm to the far infrared region. They are capable of accommodating up to three different gratings, which are mounted on a rotatable turret. In terms of (grooves/mm)/Blazed (nm), the three gratings fitted were 1200/300, 600/600 and 150/300, for the 500i SpectraPro, and 1200/500, 150/500 and 150/300, for the 300i SpectraPro. A larger number of grooves provides increased spectral resolution, but at the expense of the spectral acquisition wavelength range. The wavelength at which each grating is blazed correlates its optimum wavelength for generating monochromatic light. Using the 1200/300 and 1200/500 gratings, the 500i SpectraPro and 300i SpectraPro have an accuracy of ±0.2 nm. The NCL Spectral Measuring System controls both monochromators via a computer using the SpectraSense software (Acton Research Corporation).

The IP28 emission PMT has a wavelength range between 185-650 nm and is optimised at 340 nm. It has an anode sensitivity of 200 A/Im and a cathode sensitivity of 40 μA/Im. The R928-P SPC excitation PMT has a wavelength range between 185-900 nm and is optimised at 400 nm. It has an anode sensitivity of 2000 A/Im and a cathode sensitivity of 200 μA/Im.

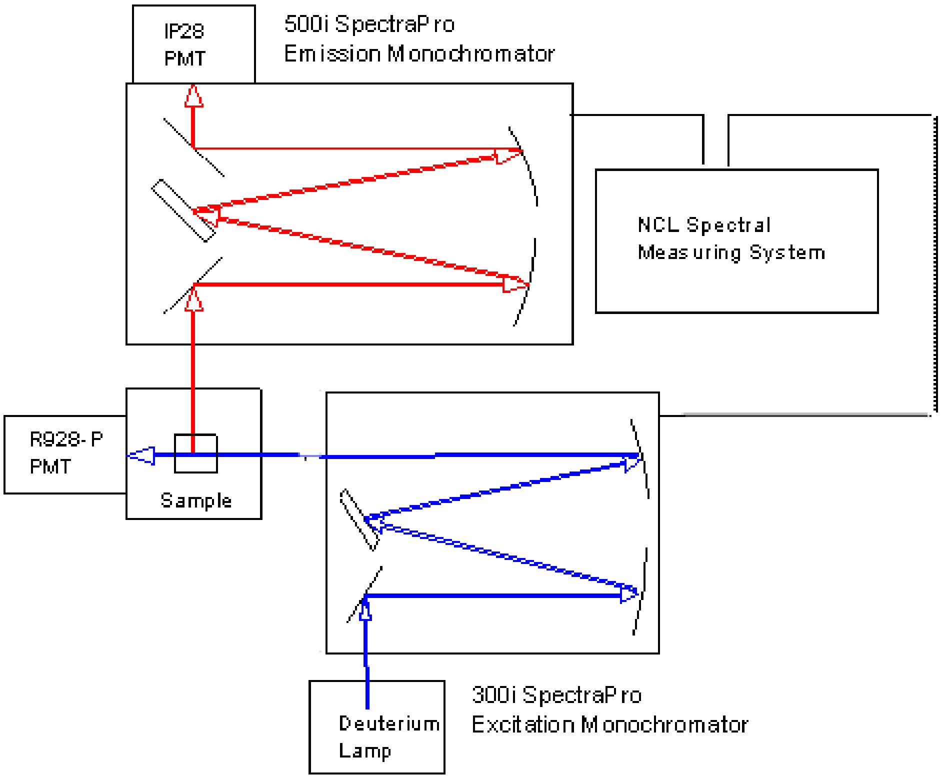


Figure IV.1. Instrument schematic of steady-state luminescence for the detection of FRET. Excitation radiation is shown in blue and emission radiation is coloured red.

IV.1.2 Time-Resolved Emission

Figure IV.2 shows a schematic of the instrumentation used to record both steady-state and time-resolved emission during the FRET analyses. Excitation was achieved by taking the second harmonic (532 nm) of a YG-980E pulsed solid state neodymium: yttrium aluminium garnet (Nd:YAG) laser (Quantel) and using it to pump a TDL-90 dye laser (Quantel) containing a Rhodamine 590 dye (Exciton) to generate an output of 565 nm. This beam was then frequency-doubled by passing it through a pair of DCC3 doubling crystals (Quantel) to generate the excitation radiation of 282.5 nm. A 300i SpectraPro monochromator (Acton Research Corporation) (described above) was used to collect emission radiation and detection was by means of an iStar DH-720-25F intensified charged couple device (iCCD) camera (Andor Technology).

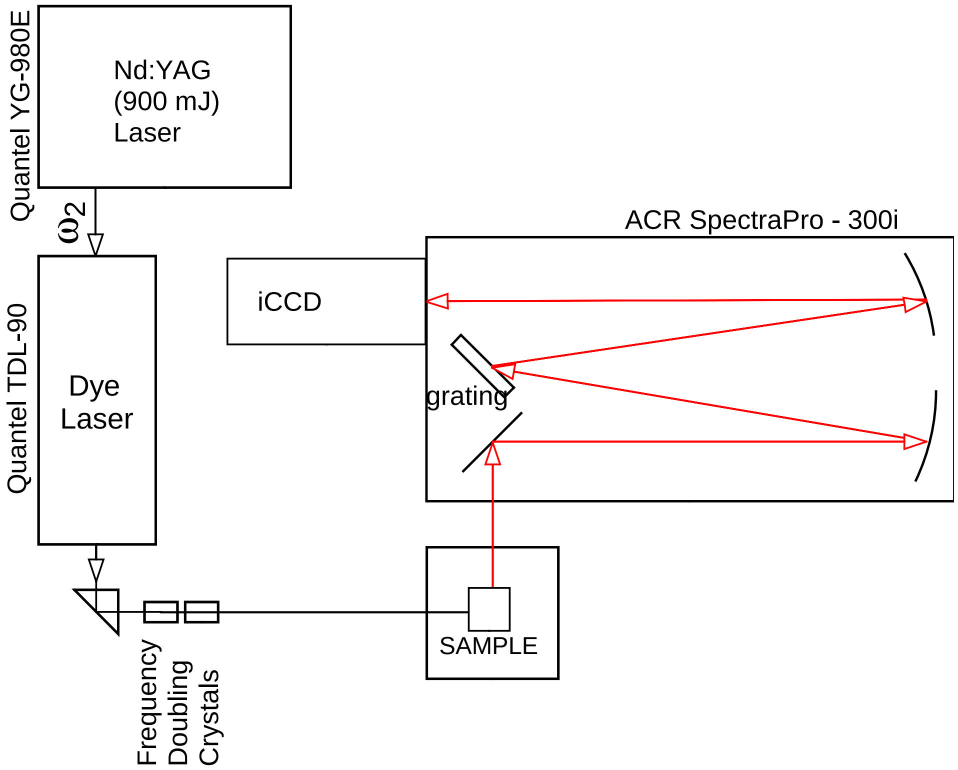


Figure IV.2. Schematic representation of the instrumentation used to acquire steady-state and time-resolved emission spectra for the detection of FRET. The emission radiation is shown in red.

Figure IV.3 shows a schematic representation of the iStar DH-720-25F iCCD image intensifier. The 2D CCD Array, composed of a 2D grid of 256 x 1024 pixels, allows for the simultaneous collection of a wavelength range of between 180-850 nm by a process known as vertical binning, whereby the charge collected by each column of pixels is summed to generate the overall spectral intensity at the corresponding wavelength. The three major components to the image intensifier are the photocathode, the microchannel plate (MCP) and the phosphor screen. When a photon of a particular energy strikes the photocathode at a characteristic location it stimulates the emission of an electron from the opposite side of the photocathode. This electron then travels in a straight line across a distance of approximately 0.2 mm down a voltage gradient of -200 volts to strike the MCP. The MCP consists of a thin disk of about 1 mm thickness that is composed of numerous honeycombed glass channels of approximately 10 μm diameter. Within each channel there exists a high potential of up to 1000 volts, so that an entering electron produces a cascading effect that results in the generation of up to 10,000 secondary electrons in a process known as gain. These secondary electrons then exit the MCP and travel in a straight line across another gap of 0.5 mm and down a potential of several thousand volts to strike the phosphor screen, where they are converted into an image by a fibre optic window. The voltage gradient between the photocathode and the MCP can be regulated between on and off. At the on-position the -200 volt gradient mentioned above exists and at the off-position there exists an opposite gradient of +50 volts. When in the off-position electrons generated by photons striking the photocathode are prevented from reaching the MCP by the positive voltage gradient, so that by turning the voltage to either on or off acts like a shutter mechanism. This has the advantage that toggling between the two voltages, or ‘gating’, can be performed on a ns time-scale with a high degree of accuracy, which provides a mechanism whereby time-resolved detection of radiation can be achieved. The DH-720-25F is specified to a minimum optical gate-width of 4.6 ns and this defines the limit to which time-resolved detection can be performed.

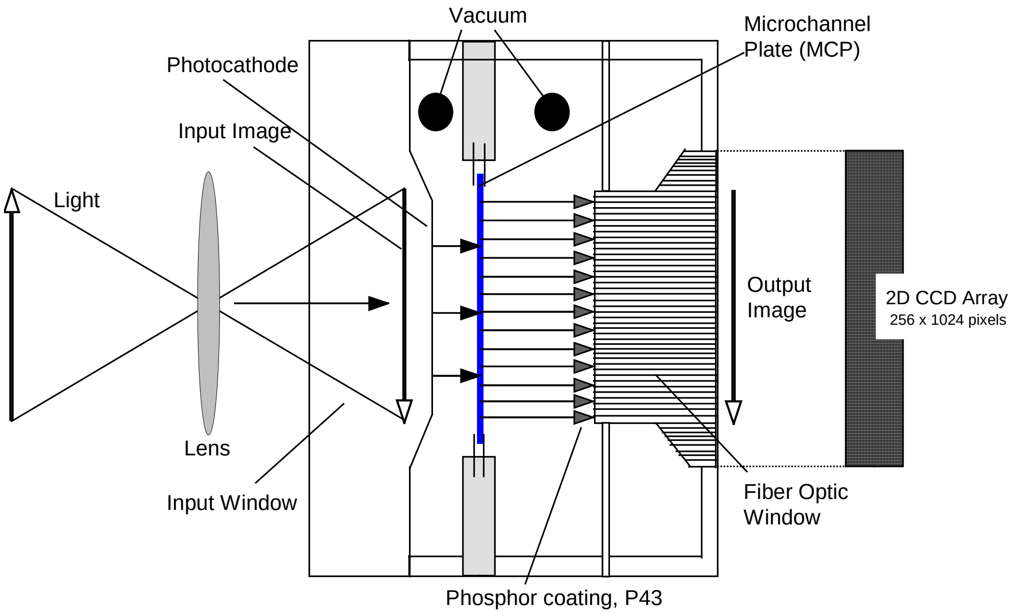


Figure IV.3. Schematic representation of the iStar DH-720-25F iCCD image intensifier.

IV.2 Circular Dichroism Spectroscopy Instrumentation

CD spectra were acquired using a Model 400 Circular Dichroism Spectrometer (Aviv Biomedical Inc.). The Model 400 is functionally divided into the following subsystems: lamp and lamp power supply; monochromator; polarizer compartment; sample compartment; and detector compartment. A schematic of the Model 400 Circular Dichroism Spectrometer is shown in Figure IV.4.

The light source for the Model 400 is a Xenon Short Arc XBO® 150/4 Bulb (OSRAM). The front panel of the power supply contains several indicators as to the instruments status, including an Hour Meter and Lamp Ready, Instrument Power On, Lamp Temperature On and N2 Flow OK lights. The lamp is automatically shut down after 30 seconds if the N2 gas flow (see below) drops below a certain value, during which an alarm sounds.

The white light produced by the xenon bulb is directed into the monochromator, where it is dispersed using two fused silica prisms in series in a “w” double monochromator design, thereby producing superior wavelength resolution and containing less stray light than would result from using a single monochromator. The prisms have been optimised to produce dispersion of light in the UV region and the wavelength range of the Model 400 is between 165 and 1200 nm. Rotation of the two prisms is linked to a wavelength cam, which converts the non-linear dispersion of the fused silica prisms into a linear motion of a computer-controlled stepped external gear drive mechanism. The motor step size determines wavelength specificity, with each step corresponding to a 0.05 nm change in wavelength. The motor slew speed is approximately 60 nm per minute. Backlash in the drive mechanism is eliminated by the monochromator drive mechanism automatically overshooting by 10 nm when moving from short to long wavelengths and this ensures wavelength accuracy during scanning. The monochromator can only scan from longer to shorter wavelengths.

The use of a linear wavelength scale makes it possible to calibrate the instrument for wavelength accuracy by precisely calibrating it at a single wavelength. This single wavelength corresponds to the position of an optical beam switch linked to the motion of the wavelength cam. All other wavelengths are then taken by counting the number of motor steps the wavelength cam is removed from this position. The entrance and exit slits of the monochromator are not fixed in position, but rather they move synchronously and are guided by the motion of an external drive system, with the exit slit width (mm) determining the fraction of the light allowed through to the sample. A stepping motor also controls the motion of the slits and the maximum slit width is 3.6 mm.

Non-polarised light is converted to linear-polarised light by an $MgFI_2$ polarizer, located within the polarizer compartment. Of the ordinary and extra-ordinary beams produced by the polarizer only the ordinary beam is used for the CD measurement, while the extra-ordinary beam is filtered out by a filter located beyond the photoelastic modulator (PEM). Circularly polarised light is produced from this linear polarised light by a PEM, which is driven by a 50 kHz oscillator, producing a strain-induced birefringence oriented 45 degrees to the axis of the linear polarised light. The oscillation amplitude is adjusted such that the PEM alternately produces +90 and –90 degrees retardation along one birefringent axis of the oscillation peaks. This leads to the alternate production of left and right circularly polarised light. Within the polarizer compartment an achromatic lens focuses light from the monochromator exit slit to a suitable point within the sample compartment and the lens is so designed that the location of this focal point is independent of wavelength. Because of this wavelength independent focal point a variety of micro cells can be used within the sample compartment without the need for special accessories.

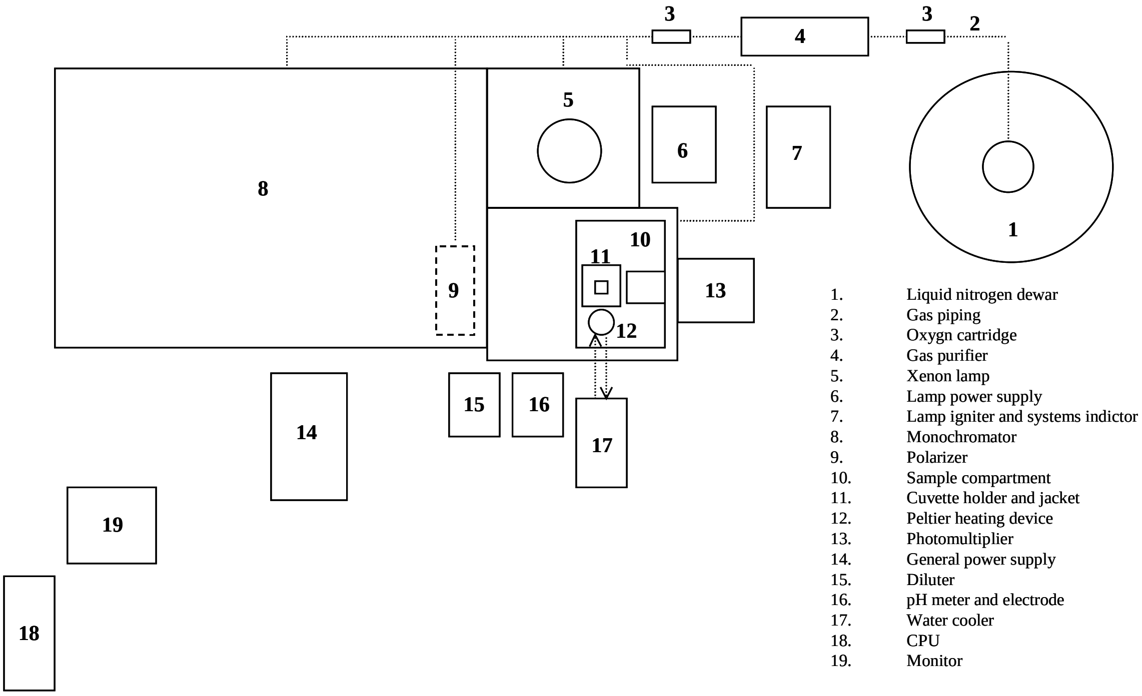
The sample compartment is of sufficient volume to facilitate the use of a number of accessories. It is possible to monitor fluorescent CD by attaching a PMT at an angle of 90 degrees to the excitation beam. The sample holder is equipped with a Peltier heating system, a magnetic stirrer and ports for a stand-alone thermometer (to monitor the temperature precisely at the quartz cuvette surface) and for the insertion of a microelectrode. Provision is also made for the supply and removal of titration solutions by directing thin tubes into the sample compartment.

The detector compartment houses a high speed, high sensitivity end-on photomultiplier tube (PMT) (Hamamatsu). The output current is converted to a voltage by a preamplifier within the PMT housing and is then sent to the electronics for processing. The PMT is operated in such a way that the output current is kept constant and, since the light energy received by the PMT varies as a function of wavelength and sample absorption, a dynode voltage is applied to the PMT to account for this variation and produce a constant current output. Mounted close to the PMT housing is a high voltage generator containing a DC to DC converter, which produces a PMT dynode voltage of 280 volts for each volt of input supplied by a D/A converter. The dynode voltage is transmitted to the PMT via a voltage driver chain at the PMT socket.

The Model 400 is kept under a continuous N2 purge while operational and this is supplied by the boil-off from liquid nitrogen contained within a PV-120 liquid nitrogen dewar (Wessington Cryogenics). The boil-off is routed from the dewar through an Oxygen Indicator Cartridge (Aviv Biomedical), is purified in an model RGP-R1-3000B Rechargeable Gas Purifier (Aviv Biomedical), then passes through another Oxygen Indicator Cartridge before supplying a separate purge to the lamp housing, monochromator, polarizer and sample compartment of the Model 400.

A MICROLAB 500 series diluter (Hamilton) equipped with two 500 μl syringes (one for solution delivery and one for removal) and fine bore no.240010 tubing (Hamilton) is interfaced with the Instrument Control Version 3.05 CD software (Aviv Biomedical Inc.) to allow for automated titration of solutions into the sample. The magnetic stir device in the cuvette holder can control a miniature stir bar placed within a 10 mm cuvette and this device is also integrated with the Instrument Control software, so that the stir speed can be varied and the experimenter can select to stir between or during spectral acquisition. Because the stepping motors that control the injection pumps have in total 2000 steps available, the minimum volume of titrant that can be injected into the sample is the volume of the syringe divided by 2000, which corresponds to 0.25 μl for 500 μl syringes. Also interfaced with the Instrument Control software is an Orion 720A+ pH meter (Thermo Electron Corporation) equipped with an MI-710 Micro-combination pH microelectrode (MICROELECTRODES INC.), allowing the pH of the sample to be monitored during titration experiments.

The sample is housed within rectangular quartz SUPRASIL sample cuvettes of nominal pathlengths 10, 5, 1, 0.1 (strain-free) and 0.01 (strain-free) mm (HELLMA UK LTD.) to allow for samples of different protein concentrations and within a range of buffer solutions to be scanned. Because of the uniform dimension of the Model 400 sample holder different pathlength cuvettes must be matched with their appropriate adaptors. Since both the 0.1 and 0.01 mm cells are demountables they fit into a common adaptor. The 10 mm cuvettes can be sealed with either a stopper or a lid. Both are necessary, since titration experiments require that the open-ended lid-designed cuvette be used to facilitate the delivery tubes, the microelectrode and the magnetic bar, while temperature ramp experiments require the cuvette to be sealed with an air-tight stopper throughout in order to prevent solvent evaporation and consequent changes in sample concentration.



1. Liquid nitrogen dewar
2. Gas piping
3. Oxygn cartridge
4. Gas purifier
5. Xenon lamp
6. Lamp power supply
7. Lamp igniter and systems indictor
8. Monochromator
9. Polarizer
10. Sample compartment
11. Cuvette holder and jacket
12. Peltier heating device
13. Photomultiplier
14. General power supply
15. Diluter
16. pH meter and electrode
17. Water cooler
18. CPU
19. Monitor

Figure IV.4. Schematic representation of the Model 400 Circular Dichroism Spectrometer. The dotted lines indicate piping through which the N_2 purge is supplied.

IV.3 Fourier Transform Infra-Red Instrumentation

IV.3.1 The Tensor 27

Figure IV.5 shows a schematic representation of the instrumentation used to acquire FTIR spectra. The major component is the Tensor 27 FTIR spectrometer (Bruker), which is equipped with a mid-IR glow-bar light source and a liquid nitrogen cooled photovoltaic mercury cadmium telluride (MCT) mid-IR detector. The Tensor 27 spectrometer was customised to have an operational wavenumber range of between 4000 to 1000 cm^{-1} and has a maximum resolution of about 1 cm^{-1}. The Michelson interferometer and MCT detector compartments of the Tensor 27 are both air-tight sealed and in both there is a column containing silicon beads that acts as a drying agent to remove water vapour from these areas. The sample compartment is supplied with a dry-air purge from a PG28L air drier (PEAK SCIENTIFIC). Protein FTIR spectral acquisition was achieved using either the AquaSpec or BioATR-IV sample accessories (Bruker), with both customised to fit the Tensor 27 and interface electronically to the spectrometer. Both accessories are designed to be able to be purged with the dry air supply to the Tensor 27 sample compartment and to be temperature regulated with a HAAKE DC 30 temperature control module and HAAKE K20 bath vessel (Thermo Electron Corporation), the functioning of which are automated with the OPUS software used to control the Tensor 27.

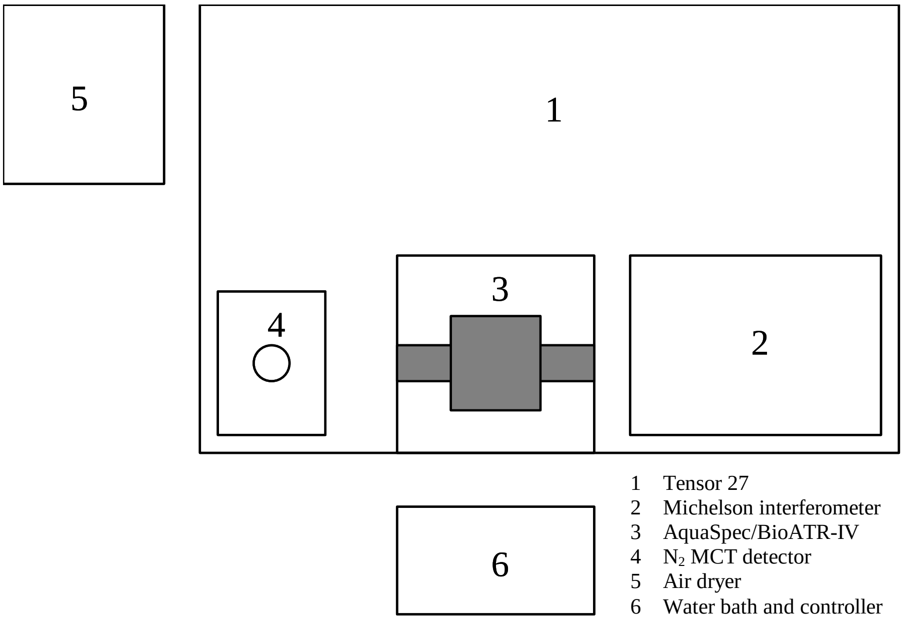


Figure IV.5. A schematic representation of the instrumentation used for protein FTIR spectroscopy.

IV.3.2 The AquaSpec Transmission Cell

The AquaSpec transmission cell sample accessory is shown in Figure IV.6. It is mounted on a QuickLock^{TM} base plate that locks into place with a fitting in the Tensor 27 sample compartment and ensures optimised and reproducible transmission of radiation through the cell. An approximate airtight seal is achieved for the beam path by telescopic tubes that bridge the distance between the KBr windows of the Tensor 27 sample compartment and CaF_2 windows of the AquaSpec sealed cell. Water vapour is removed from within these tubes by the presence of tubing that redirects the dry-air purge of the Tensor 27 sample compartment. An aperture fitting is located within the telescopic tube that carries radiation to the cell, which reduces the occurrence of spurious reflections within the tube. Figure IV.6*c* shows an expanded schematic of the AquaSpec sealed cell, revealing the presence of a thermostating plate, which is supplied with water at a controlled temperature from the K20 bath vessel through an inlet and outlet aperture. A surrounding Teflon jacket, visible as the white region of Figure IV.6*b*, insulates the sealed cell. The cell consists of two CaF_2 windows that are sealed by a special resin, which is capable of small expansion and retraction movements. At the top of the cell is located an injection aperture and at the bottom there is another aperture for the cell to drain. The cell is filled using a 50 μl HPLC syringe with a dull point, open-end needle of size 22 gauge x 2 inches. The sample is first passed through a filter with a 2-micron frit, located above the inlet port, before passing into the cell. The cell has a very small pathlength of 6 μm, which enables aqueous protein FTIR spectral acquisition, and in order to inject aqueous samples into a cavity of this size a considerable injection pressure must be used. The cell withstands this pressure by temporarily expanding upon injection and then contracting again once the sample has been injected and movement of the resin that creates the seal facilitates this.

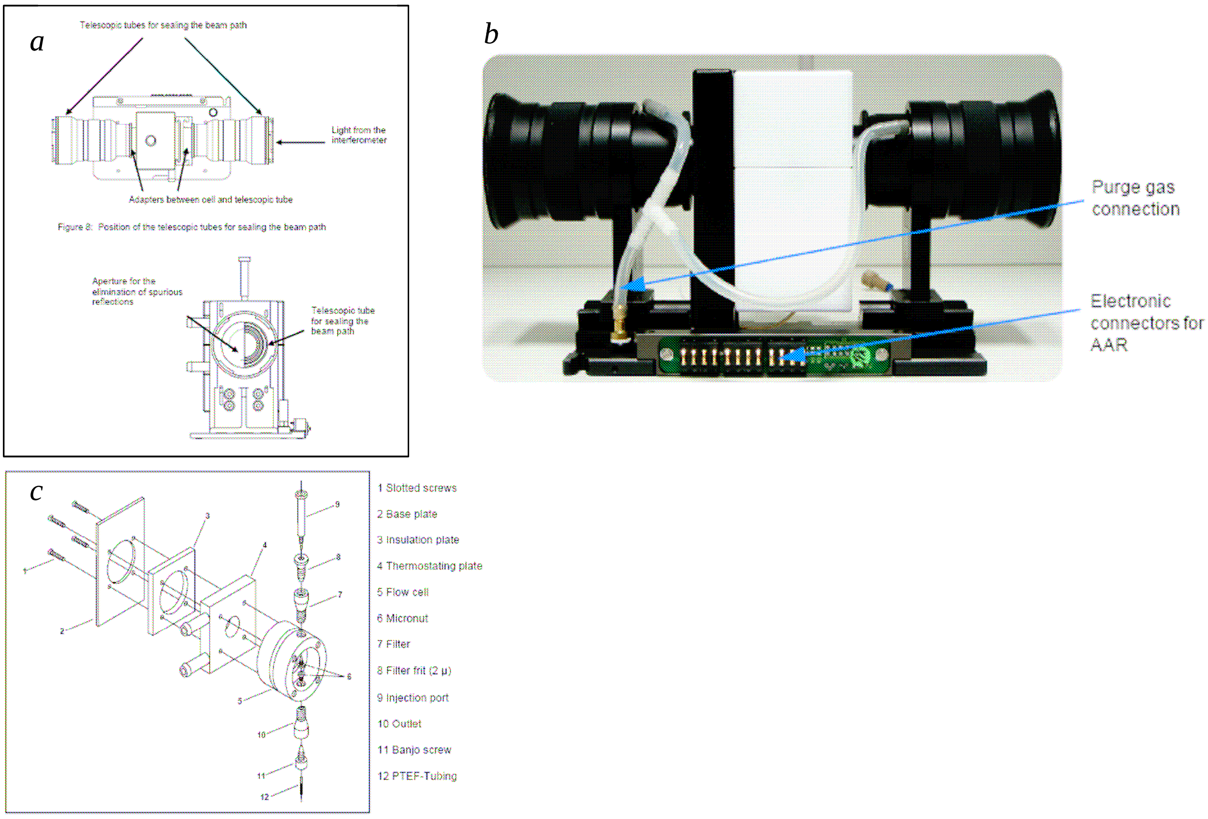


Figure IV.6. The AquaSpec transmission cell. *a* shows an end-on and top view of a schematic of the AquaSpec accessory, photographed in *b*. *c* shows an expanded schematic of the temperature regulated sealed cell assembly.

IV.3.3 The BioATR-IV Accessory

The BioATR-IV sample accessory is depicted in Figure IV.7. Similar to the AquaSpec accessory, the BioATR-IV is mounted on a QuickLock^{TM} base plate and has telescopic tubes that seal the beam path and these are supplied with a dry air purge from the Tensor 27 sample compartment purge. The sample is placed on a silicon wafer located in a well at the centre of the Main Plate of the BioATR-IV and securing a Top Plate directly on top of the Main Plate encloses the sample area. Figure IV.7*b* shows both the Main and Top Plates and the sample well is also labelled. Both the Main and Top Plates have cavities through which water is passed by connecting tubes from the K20 bath vessel to connectors on both plates. The Main Plate receives water directly from the K20 bath vessel and then transmits this water to the Top Plate via a tubular inter-connect linking the two plates and from the Top Plate the water exits and returns to the K20 bath vessel. The BioATR-IV is thermo-regulated by this circulation of water and a Teflon jacket fits over the Top Plate and helps to insulate the accessory. The Top Plate has a removable piece at its centre on the underside and by replacing the standard fitting with a dialysis insert it is possible to perform titration experiments using the BioATR-IV. Protruding from the Top Plate is a stem with a small hole bored through its centre, through which an M15(G)300 K thermocouple (Farnell) is placed in order to monitor the temperature directly at the sample area.

The mid-IR beam is deflected by an adjustable mirror into the ATR-crystal assembly, as shown in Figure IV.7*a*. The beam then undergoes multiple ‘bounces’ within the ATR-crystal before being directed onto another adjustable mirror, located on the opposite side of the crystal assembly, from where it is directed to the MCT detector. The adjustable mirrors are shown in Figure IV.7*d*, along with the screws used to position the mirrors so as to optimise the throughput of radiation through the BioATR-IV.

A schematic of the ATR-crystal assembly is shown in Figure IV.7*c* and it consists of a ZnSe ATR-crystal joined to a silicon wafer by a pressure exerted from a securing screw. The pressure applied is finely adjusted so as to ensure maximal contact between the ZnSe crystal and the silicon wafer, without cracking the silicon wafer. The throughput of the BioATR-IV is heavily dependent upon optimal contact between the ZnSe crystal and the silicon wafer. Figure IV.8 shows an overlay of a transmission spectrum of an empty sample compartment and a sample compartment with the BioATR-IV in place, within which no sample has been injected. From this it can be seen that the throughput of the BioATR-IV at 2000 cm^{-1} is approximately 3.5 %.

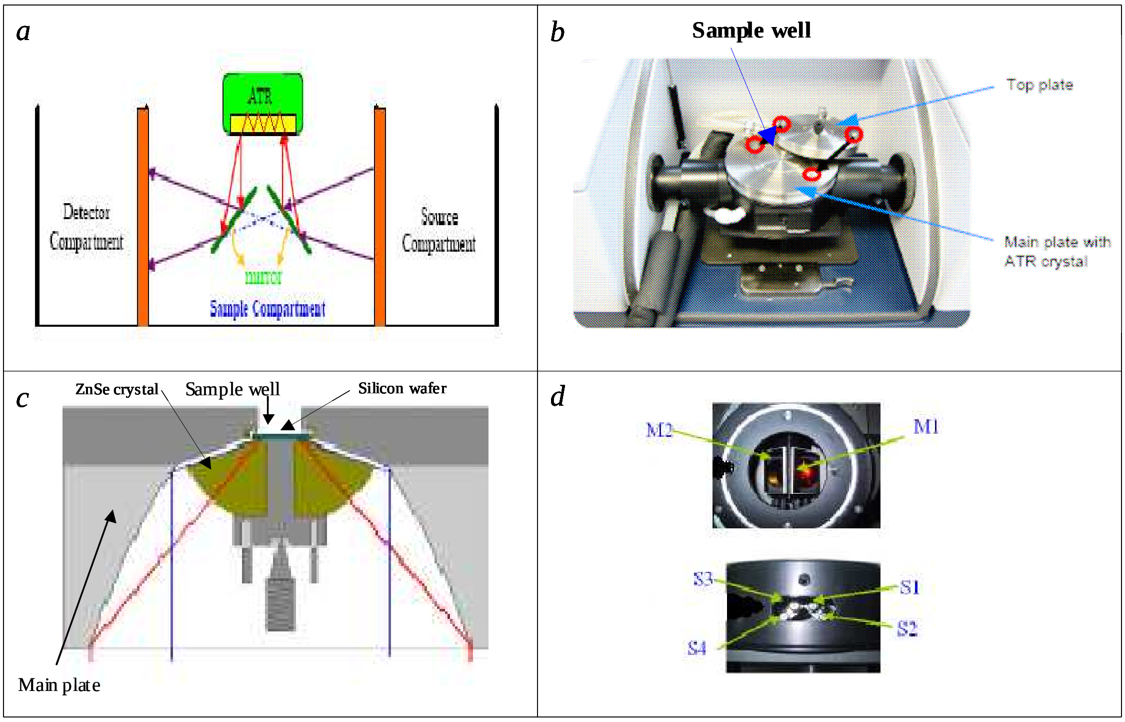


Figure IV.7. The BioATR-IV accessory. A schematic representation of the BioATR-IV is shown in *a* and is photographed when positioned in the Tensor 27 in *b*. A schematic representation of the ATR ZnSe crystal assembly is shown in *c*, while *d* gives a side and top view photograph of the directing mirrors with the adjusting screws, as represented in *a*.

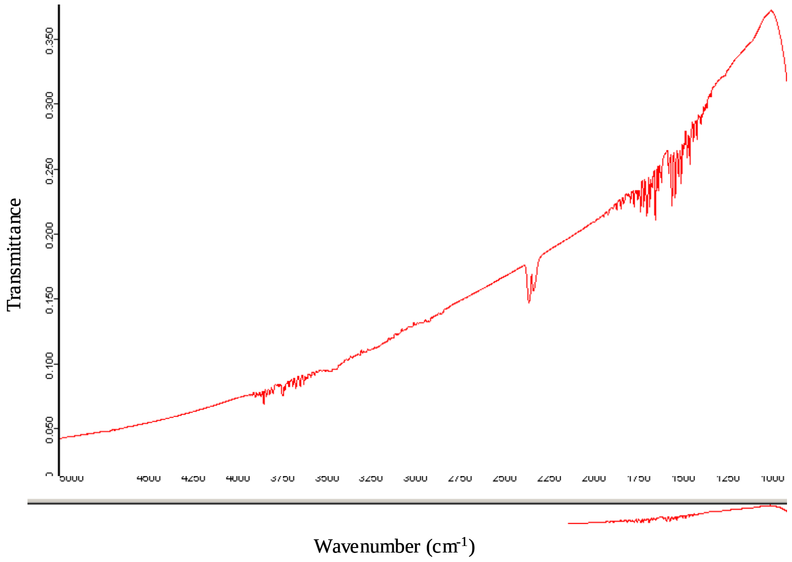


Figure IV.8. BioATR-IV throughput. The transmission spectrum of the empty sample compartment and the sample-free BioATR-IV is shown as red and blue, respectively. At 2000 cm^{-1} the BioATR-IV has a throughput of 3.5 % of that of the empty sample compartment.

IV.3.4 The Demountable Liquid Cell for Transmission FTIR

The DLC-M13 demountable liquid cell (Harrick), along with two 13 mm CaF_2 windows (Korth Crystal) and a 100 μm Mylar spacer (Harrick) is shown photographed in Figure IV.9. Also shown in Figure IV.9 is a sample holder and QuickLock^{TM} base plate for reproducible and optimised positioning within the Tensor 27 spectrometer. The DLC-M13 can be filled while completely assembled using a syringe that attaches to either the top or bottom inlets of the cell. However, this is only possible when cell pathlengths above 200 μm are used. Spectral acquisition under purged conditions is possible by closing the lid of the Tensor 27 sample compartment and purging the entire sample compartment with the DLC-M13 in place.

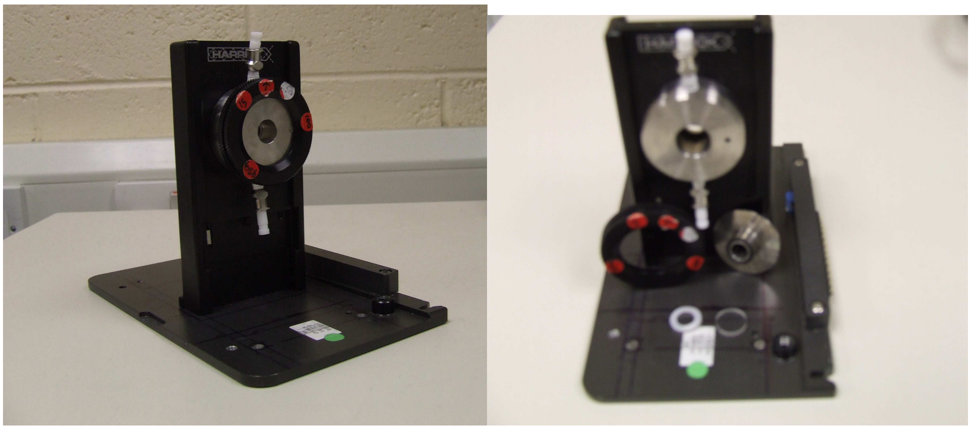


Figure IV.9. The DLC-M13 demountable liquid cell. The assembled and disassembled FTIR transmission cell is shown as left and right, respectively, along with the QuickLock^{TM} base plate and transmission cell holder.

Chapter V

Quantitative Analysis of Protein Secondary Structure by CD Spectroscopy

V.1 Methods for Deconvoluting Protein CD Spectra

Numerous approaches have been taken in the design of programs used for the quantitative analysis of protein secondary structure from its CD spectrum. In order of their appearance they are as follows: multilinear regression; singular value decomposition; ridge regression; convex constraint analysis; the self-consistent method; and neural networks. The ridge regression method is used in creating the CONTIN program, while variable selection is at the heart of the VARSLC program (CDSSTR) and the self-consistent method underlies the SELCON program. All of these methods make the assumption that the CD spectrum of a protein can be represented as a linear combination of the spectra of the secondary structural elements, as follows:

$$\theta_\lambda = \sum F_i S_{\lambda i} + \text{noise} \qquad \text{(V.1)}$$

where θ_λ is the CD of the protein as a function of wavelength, λ, F_i is the fraction of each i^{th} secondary structure and $S_{\lambda i}$ is the ellipticity at each wavelength of each i^{th} secondary structural element. The noise term includes the contribution made to the CD spectrum by any aromatic chromophores present.

V.1.1 Multilinear Regression

Using the multilinear regression method, the first attempts at deconvoluting the CD spectra of proteins made use of the CD spectra of model polypeptides of pure single classed secondary structure and fitted them to the protein CD spectrum by a simple least squares method.[1; 6] When a sufficient number of proteins had been structurally characterised using X-ray crystallography, the CD spectra of these same proteins were used to create a protein data bank and from this basis curves were extracted using multilinear regression. The basis spectra were generated so as to represent the CD spectrum of an α-helix, β-sheet (both parallel and antiparallel), β-turn and random coil as they would appear when these structures were present within a protein.[9-13] These basis spectra were then substituted for the model polypeptide CD spectra in analysing novel proteins by CD spectroscopy.

The multilinear regression approach has been used to generate programs designed to perform quantitative secondary structural analysis based on protein CD spectra and these programs can be categorised based upon whether the least-squares fit is constrained or non-constrained. If a constrained fit is used then the sum of all F_i values must equal 1, the logic behind this being that the fractional sum of all secondary structures that combine to produce the protein's CD spectrum cannot possibly be greater than unity. Constrained least-squares analysis was used in creating the multilinear regression LINCOMB and G&F programs. By using a non-constrained least-squares analysis the fit is not restricted such that the sum of the F_i values must equal unity. Instead, the coefficients are allowed to be normalised after the fitting procedure. Non-constrained analysis has the advantage that it allows for more flexibility in the fitting procedure and so can take account of the possibility of the basis spectra being less than perfect for fitting the protein CD being analysed. The MLR program was developed by the non-constrained multilinear regression method.

The G&F and the LINCOMB programs are superior to the MLR program for the prediction of β-sheet and β-turns and the use of the model polypeptide standard curves gives more accurate predictions than the use of basis curves extracted from protein databases.[1; 9; 12] However, the prediction of both β-sheet and β-turn content using the multilinear regression method is generally quite poor. A number of factors combine to produce this lack of performance. Firstly, the contributions made by aromatic groups to the protein CD are ignored when generating the basis spectra from the protein reference database, and, as such, model polypeptide standard spectra when used with multilinear regression enhances the estimation by this method of β-sheet and β-turn content. Secondly, because the basis or model polypeptide spectra are not well distinguished over several wavelength regions, the multilinear regression method of analysing the protein CD spectrum at each wavelength independently is open to a significant degree of error. The G&F, LINCOMB and MLR programs perform equally well when predicting α-helix content. The main benefit for using the multilinear regression approach for quantitative secondary structural analysis is that the programs can work with CD spectra acquired over a quite narrow wavelength range (240-200 nm) with only a slight drop in performance. Since MLR uses a non-constrained analysis, it is the only program that can be used for quantitative secondary structural analysis in the absence of accurate knowledge of protein concentration.

V.1.2 Singular Value Decomposition

To address the problem of the similarity of basis spectra over certain wavelength regions, the singular value decomposition (SVD) approach was developed, whereby basis spectra were no longer generated by considering the CD spectra of the reference proteins at each wavelength independently, but rather that orthogonal basis curves were extracted from the reference base as eigenvectors in a method of multicomponent analysis.[14] Here, the component is the CD spectrum at a single wavelength and, therefore, the spectral magnitude at each wavelength is no longer considered independently of its magnitude at other wavelengths, with the result that a combination of co-dependent spectral magnitudes at several wavelengths is much better suited to being able to characterise the CD signature of a structural component. SVD also constrains the fractional weights of each conformation to equal 1 during the fitting procedure. The SVD approach has an improved performance over the multilinear regression method when predicting the α-helix content of a protein from CD data, but it is less able to estimate β-turn content. Both approaches perform equally well in attempting to estimate β-sheet content. SVD suffers in that to have enough information for conformational analysis basis spectra must have a sufficient number of nodes (maxima and minima) and as a result the CD data must be collected to at least 184 nm.

V.1.3 Convex Constraint Analysis

Convex constraint analysis (CCA) is an approach to spectral deconvolution similar to the SVD method, in that it uses multicomponent criteria for generating the basis spectra from the protein CD spectra compiling the reference database.[15-17] Similar to SVD, the sum of the fractional components is constrained to equal 1 for CCA. CCA differs from SVD in the manner by which the basis curves are extracted. Instead of relying on spectral nodes to define each basis spectrum, a volume minimization procedure is performed, introducing a constraint that results in a finite number of component curves being extracted from the reference base. CCA also differs from

SVD in that no X-ray crystallographic data are used in generating the basis spectra. Instead the basis curves are assigned to specific secondary structures by correlation of the fractional weights of each secondary structure in the protein reference set with the fractional weight of each extracted basis spectrum. CCA gives a very good estimate of α-helix, which is largely independent of the wavelength range of the CD spectrum under investigation. However, estimates of β-sheets and β-turns are poor. It is supposed that the weakness of CCA in this regard is attributable to secondary structures within a protein not being truly independent of one another, as is assumed by the CCA method.

V.1.4 The CONTIN Program

To test the theory that the secondary structure of an unknown protein can be best estimated by interpreting its CD spectrum with respect to other proteins that have similar CD spectra and are of known structure, methods were developed that could select from the protein reference data base only those proteins that have CD spectra similar to the test spectrum and, from these, compose a new reference set with which to analyse the test spectrum. Such selection methods include ridge regression, variable selection and neural networks. Selection is seen as important in that proteins may display unusual CD spectra, which have been distorted by features such as a high presence of aromatic amino acids or the occurrence of disulfide bridges.[18] Even the adoption of a rare tertiary conformation by a protein can greatly distort its CD spectrum. Therefore, the desire to have as comprehensive a protein reference set should be balanced by including a selection method in the generation of the basis spectra.

CONTIN uses ridge regression analysis to fit the test protein CD spectrum directly from the CD spectra of a large protein reference set, without the generation of basis spectra.[19] In doing this it keeps the contribution of each spectrum small unless it contributes to a good agreement between the theoretical best-fit curve and the test spectrum. Therefore, spectra most similar to the test spectrum are given greater weight than those that are less similar. The main improvement in using CONTIN to estimate secondary structural content is that it gives a much better estimate of β-turn content than previous methods based on multilinear regression, or either SVD or CCA. Also, CONTIN doesn’t suffer from limiting the test spectrum to wavelengths longer than 200 nm. The most recent version of CONTIN uses a locally linearised method instead of the ridge regression method in selecting the weighting of proteins within the reference set and is called CONTIN/LL.

V.1.5 The VARSLC and CDSSTR Programs

Combining variable selection with the SVD method led to the creation of the VARSLC program.[20] Here proteins are systematically eliminated from the reference set, creating numerous smaller reference sets. These reduced sets are used in evaluating the conformation of the test protein and the results obtained using each are examined. Selection criteria are then applied to all the results and only those yielding a good fit are chosen. These results are then averaged to give the final result. The VARSLC program gives excellent evaluation of protein secondary structure from CD spectra, but suffers from requiring the test spectrum to be collected to at least 184 nm. In the VARSLC program there are a number of parameters that can be varied by the user for each analysis, such as the amount of calculations to be made, the constraint on the total sum of the fractional composition, the number of basis proteins to be used and the maximum root mean square deviation acceptable for a solution. Solutions that do not lie within the boundaries set by the user are ignored by the program, regardless of whether they happen to be good solutions. The CDSSTR program is essentially the same as the VARSLC program, but is designed so that all of the above-mentioned boundaries are given maximum freedom. The result is that calculations done using the CDSSTR program generally take longer than the VARSLC program, but the solutions are arrived at in a more comprehensive manner and so improve the probability of achieving a successful estimation of the test protein's secondary structure.

V.1.6 The SELCON Program

The self-consistent method continues on from the variable selection method, improving its speed and accuracy, and it is the basis on which the SELCON program has been develop.[21-23] The SELCON program arranges the proteins of the reference set in order of increasing root-mean-square difference from the test spectrum and then the most dissimilar proteins to the test data are systematically eliminated so that a reduced protein reference set is produced. This serves to increase the speed of finding the best solutions. Also, included in the reduced reference set is the test protein spectrum and an initial guess is made at its fractional secondary structural composition. This serves to improve the estimation of the structure of the test protein, since the reduced reference set is made biased towards the test protein structure. The test spectrum is then deconvoluted by the SVD method and the protein’s secondary structure is determined. The solution serves to replace the initial guess of the test protein’s fractional composition and the procedure is repeated. This is continued until self-consistency is reached and at that point the final structure is estimated.

The latest version of the SELCON program is SELCON3 and it includes in the analysis the constraint that the sum of the estimated secondary structures is equal to one and that each estimated secondary structure fraction be greater than –0.5. A final constraint is introduced in the SELCON3 program that only solutions conforming to the helix limit theorem be included in the final solutions, whereby the solutions must fall within the range of helix content suggested by the Hennessey and Johnson method.[14]

The SELCON program returns very good predictions of α-helix, β-sheet and β-turn structures, even when the test data is only collected between 240 and 200 nm. The SELCON program has also been augmented so as to be able to determine the P2 conformation of globular proteins.[23] The program, however, performs poorly when attempting to estimate fractional contents of proteins with very high β-sheet contents. The origin of this lack of performance lies within the proteins that represent the reference set from which the SELCON program currently operates and, specifically, can be traced to the magnitude of the ellipticity of a pure “infinite” β-sheet being much higher than that of a short β-sheet, the kind of which are represented in the protein reference set.

V.1.7 The K2D and CDNN Programs

The third selection method used for estimating protein secondary structure is the neural net method and this has been used in generating the K2D and CDNN programs.[24; 25] In general, a neural network consists of three different types of units, called input, hidden and output units. Input units receive information from the external into the network. Hidden units pass information between different layers within the network and output units pass information from the network to the external. The units are organised into layers and “neurons” connect each unit to other units within the same layer and/or to other units of a different layer. These connections are numerically weighted in what is referred to as the “training” or “learning” phase, in which the neural net is exposed to the protein reference set, and it is this phase that represents the selection step in the analysis.

The CDNN neural network of Bohm *et al.* consists of 83 input units, corresponding to the CD data at 83 wavelengths between 260 and 178 nm, a single hidden layer of 45 units and an output layer of 5 units, representing the α-helix, parallel and antiparallel β-sheet, β-turn and remainder conformations.[26] The K2D program differs from CDNN in that it relies on a procedure called proteinotopic mapping to create a database of weights and then uses a recall program to determine the α-helix and β-sheet structures of a test protein. Also, the K2D neural net utilizes data only between 240 and 200 nm as input, but it is unable to estimate β-turn content. Of all the programs used for estimating the fractional composition of protein secondary structure K2D is best able to estimate β-sheet content when only a limited wavelength range of data has been acquired.

V.1.8 The Root Mean Square Deviation and Correlation Coefficient

The general approach taken to assess the performance of a program in correctly predicting the fractional composition of a protein’s secondary structure is to analyse the ability of the program in estimating the fractional composition of each protein of known secondary structure contained within the protein reference set, after that protein has been removed from the reference set. The two categories of performance indices are the root mean square deviation (RMSD) (δ) and the correlation coefficient (r) between the crystal structure and the CD predicted values. The performance indices are given separately for each secondary structure estimate (e.g. δ_α and r_α for the α-helix prediction) and then again for all estimated secondary structures combined, representing an overall performance index. The RMSD and correlation coefficients are calculated by the following equations

$$\delta = \sqrt{\frac{\sum_i \left(f_i^{CD} - f_i^{X} \right)^2}{N}} \qquad \text{(V.2)}$$

and

$$r = \frac{N\sum_{i}\left(f_i^{CD}\times f_i^{X}\right)-\sum_{ij}\left(f_i^{CD}\times f_j^{X}\right)}{\sqrt{\left[N\sum_{i}\left(f_i^{CD}\right)^2-\left(\sum_{i} f_i^{CD}\right)^2\right]\times\left[N\sum_{i}\left(f_i^{X}\right)^2-\left(\sum_{i} f_i^{X}\right)^2\right]}} \qquad \text{(V.3)}$$

where f_i^{CD} and f_i^{X} are CD and X-ray estimates of secondary structure types of N reference proteins, respectively. An exact fit of the sample CD spectrum will return a δ value of 0 and an r value of 1. The closeness of each to these ideal values indicates the degree of confidence associated with the estimation of fractional composition of the sample protein. Confidence in the estimation of the fractional composition of a protein is further enhanced if similar results are obtained when analysing the data by different methods, and, therefore, analyses should be performed using a variety of methods. This, however, assumes that all methods be used with a common protein reference set and finding the best reference set is generally a matter of trial and error. It should be noted, however, that reference sets that consist of a larger number of proteins and represent a larger coverage of characteristic secondary structures, along with CD data for each protein extending well into the far-UV, as far as 178 nm, generally give superior estimates of fractional composition.

V.1.9 The Protein Reference Sets

The CDPro software package consists of the CDSSTR, SELCON3, CONTIN/LL and CLUSTER programs, and is available online to non-commercial users at http://lamar.colostate.edu/~sreeram/CDPro/main.html. There are 10 different protein reference sets to choose from when analysing the CD spectrum of a sample protein. These are listed as SP29, SP22X, SP37, SP43, SP37A, SDP42, SDP48, CLSTR, SMP50 and SMP56. The reference sets labelled with the initials SP consist only of soluble native proteins and the number indicates the amount of proteins that make up the reference set. Similarly, reference sets labelled SDP and SMP consist of soluble proteins along with either denatured or membrane proteins, respectively. The proteins are characterised according to their fractional compositions with respect to six secondary structures, which have been assigned by the DSSP method for interpreting X-ray crystallographic spectra and these represent regular α-helix (α_R), distorted α-helix (α_D), regular β-strand (β_R), distorted β-strand (β_D), β-turns (T) and unordered (U) conformations.[27] Exceptions are when the reference set is designed to be able to predict polyproline or 3_{10}-helix conformations. Table V.1 lists the fractional composition of each of the 56 proteins used by the CDPro software.

V.1.10 The Relative Performance of CONTIN(/LL), SELCON3 and CDSSTR

Sreerama and Woody have tested the performances of three of the most widely used methods for interpreting protein fractional composition from CD spectroscopy, namely, CONTIN(/LL), SELCON3 and CDSSTR, using a variety of reference sets (Table V.1) and conclude that the performances of all three methods are comparable.[3] The performance for each method is gauged by estimating the fractional composition of each protein in the reference set, with that protein being excluded from the reference set during the analysis, and by then comparing these estimates with the X-ray crystal structure and calculating the δ and r values, using equations V.2 and V.3. As shown in Table V.2 they found that CDSSTR performed the best when a reference set consisting of a smaller number of proteins, but containing data over a larger wavelength range was used. When the number of reference proteins was increased at the expense of the wavelength range available for the analysis it was found that CONTIN/LL performed best. The quality of performance for estimating the content of individual secondary structures was found to be mixed between the different methods. The authors, therefore, recommend that all three methods should be used conjointly when analysing data by any of the three methods to improve the reliability of the final result.[3]

Protein	PDB code	Protein class	Wavelength range (nm)	Reference set	α_R	α_D	β_R	β_D	T	U
Myoglobin	4mbn	αα	260-178	1,2,3,4,5,6,7,8,9	0.582	0.222	0.000	0.000	0.052	0.144
Hemoglobin	2mhb	αα	260-178	1,2,3,4,5,6,7,8,9	0.537	0.223	0.000	0.000	0.105	0.136
Hemerythrin	2hmz	αα	260-178	1,2,3,4,5,6,7,8,9	0.478	0.197	0.000	0.000	0.111	0.215
T4 lysozyme	2lzm	αα	260-178	1,2,3,4,5,6,7,8,9	0.421	0.244	0.049	0.037	0.116	0.134
Triose phosphate isomerase	3tim	αβ	260-178	1,2,3,4,5,6,7,8,9	0.236	0.210	0.090	0.064	0.124	0.276
Lactate dehydrogenase	6ldh	αβ	260-178	1,2,3,4,5,6,7,8,9	0.277	0.161	0.088	0.073	0.155	0.246
Lysozyme	1lys	αβ	260-178	1,2,3,4,5,6,7,8,9	0.202	0.217	0.016	0.047	0.298	0.221
Thermolysin	8tln	αβ	260-178	1,2,3,4,5,6,7,8,9	0.282	0.133	0.070	0.095	0.215	0.206
Cytochrome c	5cyt	αα	260-178	1,2,3,4,5,6,7,8,9	0.214	0.194	0.000	0.000	0.233	0.359
Phosphoglycerate kinase	3pgk	αβ	260-178	2,3,4,5,6,7,8,9	0.210	0.135	0.043	0.067	0.231	0.313
EcoRI endonuclease	1eri	αβ	260-178	2,3,4,5,6,7,8,9	0.192	0.127	0.098	0.080	0.210	0.293
Flavodoxin	1fx1	αβ	260-178	1,2,3,4,5,6,7,8,9	0.209	0.108	0.108	0.108	0.264	0.203
Substilisin BPN′	1SBT	αβ	260-178	2,3,4,5,6,7,8,9	0.171	0.131	0.098	0.080	0.225	0.295
Glyceraldehyde 3-P dehydrogenase	3gpd	αβ	260-178	2,3,4,5,6,7,8,9	0.172	0.102	0.115	0.093	0.217	0.301
Papain	9pap	αβ	260-178	1,2,3,4,5,6,7,8,9	0.137	0.123	0.094	0.075	0.175	0.396
Substilisin Novo	2SBT	αβ	260-178	2,3,4,5,6,7,8,9	0.113	0.102	0.065	0.073	0.295	0.353
Ribonuclease A	3rn3	αβ	260-178	1,2,3,4,5,6,7,8,9	0.113	0.097	0.218	0.113	0.218	0.242
Pepsinogen	2psg	αβ	260-178	1,2,3,4,5,6,7,8,9	0.051	0.154	0.235	0.151	0.165	0.243
β-lactoglobulin	1beb	αβ	260-178	1,2,3,4,5,6,7,8,9	0.056	0.111	0.287	0.123	0.216	0.207
α-chymotrypsin	5cha	ββ	260-178	1,2,3,4,5,6,7,8,9	0.069	0.045	0.208	0.106	0.200	0.371
Azurin	5azu	αβ	260-178	1,2,3,4,5,6,7,8,9	0.047	0.062	0.141	0.109	0.312	0.328
Elastase	3est	ββ	260-178	1,2,3,4,5,6,7,8,9	0.021	0.087	0.225	0.117	0.208	0.342
γ-crystallin	4gcr	ββ	260-178	2,3,4,5,6,7,8,9	0.006	0.086	0.299	0.161	0.109	0.339
Prealbumin	2pab	ββ	260-178	1,2,3,4,5,6,7,8,9	0.031	0.031	0.307	0.142	0.165	0.323
Concanavalin A	2ctv	ββ	260-178	1,2,3,4,5,6,7,8,9	0.000	0.038	0.329	0.135	0.236	0.262
Bence-Jones protein	1rei	ββ	260-178	1,2,3,4,5,6,7,8,9	0.000	0.028	0.294	0.196	0.229	0.252
Tumor necrosis factor	1tnf	ββ	260-178	2,3,4,5,6,7,8,9	0.000	0.019	0.293	0.140	0.219	0.329
Superoxide dismutase	2sod	ββ	260-178	1,2,3,4,5,6,7,8,9	0.000	0.018	0.248	0.119	0.298	0.316
α-bungarotoxin	2abx	ββ	260-178	2,3,4,5,6,7,8,9	0.000	0.000	0.014	0.095	0.284	0.608
Trypsin		αβ	260-178	1	0.088	0.031	0.193	0.136	0.140	0.412
Colicin A	1col	αα	240-185	3,4,5,7,8,9	0.529	0.225	0.000	0.000	0.044	0.202
Rat intestinal fatty acid binding protein	1ifc	ββ	240-185	3,4,5,7,8,9	0.053	0.061	0.432	0.152	0.152	0.152
Green fluorescent protein	1ema	ββ	240-185	3,4,5,7,8,9	0.004	0.064	0.347	0.093	0.191	0.301
Staphylococcal nuclease	2sns	αβ	240-190	4,6,7,8,9	0.094	0.101	0.081	0.107	0.289	0.328
Insulin	4ins	αα	240-190	4,6,7,8,9	0.294	0.235	0.020	0.040	0.050	0.361
Parvalbumin	5cpv	αα	240-190	4,6,7,8,9	0.278	0.287	0.000	0.037	0.194	0.204
Carboxypeptidase A	5cpa	αβ	240-190	4,6,7,8,9	0.254	0.127	0.111	0.052	0.212	0.244
Bovine pancreatic trypsin inhibitor	5pti	αβ	240-190	4,6,7,8,9	0.069	0.138	0.172	0.069	0.190	0.362
Adenylate kinase	3adk	αβ	240-190	4,6,7,8,9	0.340	0.206	0.077	0.052	0.012	0.313
α-chymotrypsinogen	2cga	ββ	260-178	3,4,5,6,7,8,9	0.053	0.082	0.210	0.110	0.210	0.335
Alcohol dehydrogenase	8adh	αβ	260-178	3,4,5,6,7,8,9	0.139	0.115	0.139	0.096	0.214	0.297
Carbonic anhydrase	1ca2	αβ	260-178	3,4,5,6,7,8,9	0.058	0.104	0.170	0.116	0.240	0.312
Glutathione reductase	3grs	αβ	260-178	3,4,5,6,7,8,9	0.188	0.142	0.140	0.096	0.172	0.262
Rhodanese	1rhd	αβ	260-178	3,4,5,6,7,8,9	0.150	0.147	0.041	0.068	0.235	0.359
Reaction center*	1prc	αα	245-185	8,9	0.291	0.186	0.024	0.042	0.194	0.263
Photosystem I*	1jb0	αα	245-185	8,9	0.363	0.193	0.025	0.029	0.167	0.222
Reaction center*	1qov	αα	245-185	8,9	0.341	0.185	0.035	0.035	0.138	0.263
Antenna complex*	1nkz	αα	245-185	8,9	0.569	0.161	0.000	0.000	0.086	0.183
Ubiquinol-cytochrome c reductase*	1bgy	αα	245-185	8,9	0.355	0.163	0.056	0.034	0.165	0.228
Cytochrome oxidase*	1occ	αα	245-185	8,9	0.434	0.146	0.031	0.022	0.141	0.226
Rhodopsin*	1f88	αα	245-185	8,9	0.482	0.153	0.012	0.025	0.160	0.166
Bacteriorhodopsin*	1qhi	αα	245-185	8,9	0.605	0.154	0.035	0.017	0.109	0.079
Ca^{2+}ATPase*	1eu1	αβ	245-185	8,9	0.286	0.154	0.087	0.058	0.203	0.211
Porin*	2omf	ββ	245-185	8,9	0.010	0.035	0.462	0.118	0.223	0.153
Porin*	2por	ββ	245-185	8,9	0.027	0.040	0.462	0.106	0.193	0.172
Maltoporin*	1af6	ββ	245-185	8,9	0.000	0.028	0.482	0.114	0.159	0.216
Phosphoporin*	1pho	ββ	245-185	8,9	0.000	0.020	0.433	0.115	0.236	0.194

Table V.1. Reference proteins included in the CDPro software suite. The protein reference sets are as follows: 1=SP22X; 2=SP29; 3=SP37; 4=SP43; 5=SP37A; 6=SDP42; 7=SDP48; 8=SMP50; 9=SMP56. Reference sets 3 and 5 contain the same set of proteins, but differ in the assignment of secondary structures. For reference set 5 the secondary structures are redefined as α, β, PII, T and U, for the purpose of being able to estimate PII content of sample proteins. Similarly, the secondary structures of reference set 2 are defined as α, 3_{10}, β, PII, T and U for the purpose of estimating 3_{10} helical content. The secondary structures of all other reference sets are defined as given in the table. Membrane proteins are labelled as *.

V.1.11 The Wavelength Dependence of Protein CD

Comparing the cross referencing of SP29:260-178 with itself and with SP29:240-190 gives a clear indication of the effect of extending the data acquisition to lower wavelengths. It can be seen from Table V.2 that acquisition to 178 nm significantly improves the correlation coefficients of all fractions, except that of the regular helix, when compared with the same correlation coefficients for data collected to 190 nm. The RMSD values show a similar trend when the two sets of performance indices are compared. Both trends are reflected in the overall RMSD and correlation coefficients for the two wavelength ranges. Cross correlating the reference sets SP29:240-190 and SP29:240-200 with both SP43:240-190 and SDP48:240-190 gives similar results, further emphasising the importance of collecting data to as short a wavelength as possible before attempting an estimation of the secondary structure fractional composition of a protein from its CD spectrum. These results were found to be independent of the particular method used for the analysis.

V.1.12 The Dependence of Reference Set Size

Assessing the performance of each method while varying the number of proteins in the reference set can be done by comparing the performance indices returned by the self correlation of each of the five reference sets used, and in particular those of SP29, SP37 and SP43. For both SELCON3 and CONTIN/LL an improvement in performance was seen when increasing the number of reference proteins from 29 to 37. For the same methods, increasing the number of proteins to 43 showed no improvement over the case when 37 proteins were used. CDSSTR performed best when only 29 proteins were used and showed no difference between using 37 or 43 proteins.

Greenfield has gone on to perform a similar analysis on a selection of proteins chosen from the SP29 reference set (1αα, 2αβ and 1ββ class2) with a more comprehensive selection of analysis methods.[8] Also, the estimates of fractional compositions were compared when data was collected to different shorter wavelength limits. Table V.3 shows the accuracy that can be expected when using CD spectroscopy to estimate protein secondary structure contents. In general, the LINCOMB and MLR methods did a poor job of estimating the fractional compositions of the four proteins used, regardless of wavelength lower limit. This could be due to the fact that the polypeptide reference set of Brahms and Brahms was used during the analysis.[1] CONTIN showed equal or better performance when analysing data collected to 200 nm compared to data collected to 190 nm for all proteins except α-chymotrypsin. SELCON gave approximately equally good estimates for all four proteins regardless of whether data was collected to 200 or 178 nm. The same was true of the VARSLC program, which overall did not perform quite as well as the SELCON or CONTIN programs. CONTIN, SELCON and VARSLC all performed well when estimating the α-helical content of each of the proteins, but performed less well when trying to estimate β-sheet content of all but the αα protein of myoglobin. In complete contrast, the neural net K2D program performed very well when estimating the β-sheet contents of all proteins except for myoglobin. The CDNN neural net program also performed excellently well in estimating the fractional compositions of each of the four proteins and the pure α-helix polypeptide. CDNN performed slightly better when interpreting data collected to 200 nm compared to 180 nm. It did a poor job in analysing the pure β-sheet polypeptide, underestimating the β-sheet content and overestimating the α-helix content.

Ref. set	**Method**	**Cross ref. set: wavelength range**	$\delta_{\alpha R}$	$r_{\alpha R}$	$\delta_{\alpha D}$	$r_{\alpha D}$	$\delta_{\beta R}$	$r_{\beta R}$	$\delta_{\beta D}$	$r_{\beta D}$	δ_T	r_T	δ_U	r_U	δ	r
SP29	SELCON3	SP29: 260-178	0.054	0.946	0.052	0.717	0.087	0.646	0.034	0.742	0.062	0.482	0.101	0.300	0.073	0.795
		SP29: 240-190	0.052	0.949	0.053	0.689	0.102	0.547	0.036	0.709	0.075	0.302	0.118	0.268	0.078	0.773
	CDSSTR	SP29: 260-178	0.050	0.955	0.053	0.805	0.079	0.706	0.029	0.810	0.060	0.536	0.099	0.478	0.066	0.836
		SP29: 240-190	0.059	0.938	0.052	0.785	0.083	0.655	0.030	0.790	0.074	0.337	0.097	0.491	0.070	0.817
	CONTIN	SP29: 260-178	0.046	0.960	0.050	0.727	0.099	0.489	0.031	0.783	0.060	0.476	0.100	0.397	0.070	0.812
	CONTIN/LL	SP29: 260-178	0.050	0.952	0.056	0.695	0.099	0.533	0.034	0.734	0.065	0.448	0.103	0.350	0.072	0.802
		SP29: 240-190	0.058	0.936	0.055	0.679	0.102	0.486	0.035	0.719	0.074	0.323	0.103	0.317	0.075	0.784
SP37	SELCON3	SP37:240-185	0.050	0.952	0.043	0.767	0.084	0.705	0.037	0.664	0.056	0.570	0.108	0.154	0.068	0.819
		SP29: 240-190	0.047	0.960	0.050	0.715	0.094	0.638	0.036	0.704	0.063	0.538	0.116	0.142	0.073	0.795
	CDSSTR	SP37:240-185	0.055	0.946	0.044	0.830	0.096	0.600	0.028	0.811	0.065	0.448	0.101	0.323	0.070	0.809
		SP29: 240-190	0.059	0.939	0.047	0.811	0.087	0.648	0.030	0.801	0.066	0.452	0.098	0.413	0.069	0.819
	CONTIN	SP37:240-185	0.056	0.940	0.042	0.773	0.101	0.529	0.030	0.787	0.064	0.362	0.087	0.380	0.068	0.814
	CONTIN/LL	SP37:240-185	0.052	0.948	0.047	0.745	0.098	0.577	0.031	0.763	0.066	0.418	0.094	0.279	0.069	0.811
		SP29: 240-190	0.054	0.944	0.052	0.706	0.093	0.624	0.033	0.753	0.066	0.447	0.095	0.328	0.069	0.813
SP43	SELCON3	SP43:240-190	0.053	0.941	0.044	0.776	0.086	0.663	0.031	0.743	0.073	0.367	0.098	0.216	0.068	0.811
		SP29: 240-190	0.051	0.953	0.048	0.747	0.086	0.659	0.034	0.746	0.073	0.382	0.110	0.181	0.072	0.802
		SP29: 240-200	0.056	0.938	0.044	0.809	0.094	0.550	0.037	0.682	0.074	0.273	0.105	0.256	0.073	0.799
	CDSSTR	SP43:240-190	0.065	0.918	0.045	0.771	0.092	0.611	0.028	0.807	0.068	0.463	0.088	0.369	0.068	0.810
		SP29: 240-190	0.064	0.929	0.042	0.792	0.081	0.704	0.028	0.843	0.067	0.462	0.089	0.444	0.065	0.833
		SP29: 240-200	0.063	0.933	0.043	0.789	0.088	0.609	0.030	0.804	0.076	0.357	0.088	0.452	0.068	0.820
	CONTIN	SP43:240-190	0.059	0.927	0.046	0.753	0.088	0.631	0.029	0.782	0.078	0.213	0.082	0.397	0.067	0.815
	CONTIN/LL	SP43:240-190	0.057	0.930	0.043	0.793	0.087	0.649	0.029	0.774	0.077	0.333	0.089	0.253	0.068	0.814
		SP29: 240-190	0.053	0.942	0.048	0.756	0.084	0.674	0.031	0.781	0.076	0.373	0.096	0.262	0.069	0.817
		SP29: 240-200	0.065	0.920	0.047	0.771	0.084	0.658	0.030	0.793	0.081	0.239	0.097	0.311	0.071	0.805
SDP42	SELCON3	SDP42:240-185	0.047	0.956	0.043	0.794	0.082	0.672	0.037	0.690	0.064	0.650	0.140	0.769	0.077	0.873
		SP29: 240-190	0.048	0.957	0.050	0.724	0.091	0.641	0.037	0.694	0.066	0.527	0.125	0.316	0.076	0.798
	CDSSTR	SDP42:240-185	0.052	0.950	0.042	0.847	0.093	0.620	0.029	0.819	0.069	0.585	0.140	0.774	0.080	0.864
		SP29: 240-190	0.059	0.940	0.047	0.813	0.088	0.616	0.029	0.811	0.073	0.343	0.116	0.495	0.074	0.805
	CONTIN	SDP42:240-185	0.054	0.940	0.047	0.755	0.095	0.618	0.032	0.769	0.091	0.187	0.157	0.712	0.090	0.712
	CONTIN/LL	SDP42:240-185	0.049	0.950	0.045	0.781	0.088	0.671	0.030	0.799	0.071	0.575	0.120	0.836	0.074	0.885
		SP29: 240-190	0.055	0.943	0.052	0.711	0.088	0.650	0.031	0.788	0.069	0.401	0.097	0.485	0.069	0.823
SDP48	SELCON3	SDP48:240-190	0.052	0.942	0.044	0.806	0.082	0.694	0.034	0.719	0.076	0.505	0.129	0.775	0.076	0.866
		SP29: 240-190	0.053	0.949	0.049	0.746	0.081	0.678	0.034	0.748	0.072	0.431	0.119	0.394	0.073	0.810
		SP29: 240-200	0.065	0.920	0.045	0.813	0.099	0.505	0.040	0.638	0.076	0.255	0.132	0.281	0.082	0.768
	CDSSTR	SDP48:240-190	0.060	0.930	0.047	0.822	0.087	0.640	0.031	0.770	0.078	0.455	0.135	0.766	0.080	0.852
		SP29: 240-190	0.062	0.933	0.051	0.819	0.082	0.664	0.028	0.821	0.074	0.341	0.116	0.500	0.074	0.807
		SP29: 240-200	0.056	0.948	0.049	0.825	0.084	0.644	0.033	0.754	0.079	0.322	0.121	0.465	0.076	0.803
	CONTIN	SDP48:240-190	0.055	0.934	0.049	0.750	0.091	0.594	0.035	0.685	0.092	0.099	0.154	0.672	0.089	0.814
	CONTIN/LL	SDP48:240-190	0.053	0.941	0.041	0.840	0.081	0.697	0.031	0.765	0.076	0.512	0.114	0.833	0.072	0.884
		SP29: 240-190	0.053	0.946	0.047	0.785	0.078	0.710	0.031	0.780	0.071	0.441	0.093	0.527	0.065	0.843
		SP29: 240-200	0.066	0.919	0.048	0.769	0.090	0.618	0.033	0.753	0.085	0.139	0.126	0.314	0.080	0.771

Table V.2. The performance of SELCON3, CDSSTR and CONTIN(/LL) programs for analysing protein CD spectra when different reference sets are used. The table rows are coloured clear, grey and yellow where the reference set is cross correlated with itself, the reference set SP29 with data between 240-190 nm and the reference set SP29 with data between 240-200 nm, respectively. The latter two performance indices are included in order to compare the performance effect with wavelength range and with varying the number of proteins in the reference set. (Modified from [3])

Method		**X-ray**	**LINCOMB***		**MLR***		**CONTIN**		**SELCON**		**VARSLC**		**K2D**	**CDNN**	
λ_{min} (nm)			***200***	***178***	***200***	***178***	***200***	***190***	***200***	***178***	***200***	***178***	***200***	***200***	***180***
Myoglobin	*α-helix*	**78**	96	93	89	97	67	89	73	79	76	74	74	83	84
	β-sheet	**0**	0	0	0	0	0	0	-3	0	0	0	8	3	2
	Turn	**10**	4	5	8	3	16	0	13	20	4	18	ND	9	9
Lactate dehydrogenase	*α-helix*	**37**	46	40	63	42	46	40	41	39	40	39	55	42	44
	β-sheet	**14**	21	29	15	33	7	39	12	27	15	28	11	13	10
	Turn	**25**	15	11	13	9	26	1	22	27	17	13	ND	15	14
Chymotrypsin	*α-helix*	**10**	15	21	33	28	11	9	7	15	24	16	12	19	20
	β-sheet	**38**	25	14	6	0	16	32	16	26	0	11	33	29	50
	Turn	**26**	10	16	5	9	44	32	13	16	42	39	ND	21	22
Bence Jones protein	*α-helix*	**3**	0	0	0	0	6	0	3	9	14	0	3	13	16
	β-sheet	**50**	43	40	68	40	42	81	47	34	7	69	50	39	22
	Turn	**24**	25	29	17	28	25	10	23	40	28	17	ND	22	22
λ_{min} (nm)			*200*	*190*	*200*	*190*	*200*	*190*	*200*	*178*	*200*	*178*	*200*	*200*	*180*
Poly(lys-leu)n In 0.5 M NaF, pH 7	*α-helix*	**0**	0	9	0	2	12	9	31	24	31	30	5	34	39
	β-sheet	**100**	89	65	89	99	73	91	24	39	51	55	89	18	13
	Turn	**0**	11	26	7	0	0	0	26	26	1	3	ND	16	13
λ_{min} (nm)			*200*	*190*	*200*	*190*	*200*	*190*	*200*	*190*	*200*	*190*	*200*	*190*	*190*
Poly-L-lysine 0.01 % helical form	*α-helix*	**100**	97	100	-	-	100	100	99	89	100	96	100	92	94
	β-sheet	**0**	3	0	-	-	0	0	0	1	6	18	0	1	1
	Turn	**0**	0	0	-	-	0	0	1	6	13	15	ND	7	7

Table V.3. Comparison of the performances of various methods in estimating the fractional composition of 4 proteins and 2 polypeptides. *The reference set used for the LINCOMB and MLR methods when analysing the 4 proteins was that of Brahms and Brahms.[1] The polypeptide reference set of Reed and Reed was used for estimating the composition of poly(lys-leu)n and poly-L-lysine.[4] A common reference set from the CDPro software was used for all other methods on the 4 proteins. (Reproduced from[8])

V.1.13 Analysis of Membrane Proteins by CD

The ability of soluble protein reference sets to analyse the fractional composition of membrane proteins has been a topic of some recent debate. Wallace *et al.* examined the performance of the CDPro software in analysing the fractional compositions of eight membrane proteins and concluded that the soluble protein reference sets were inadequate for estimating the fractional composition of membrane proteins from their CD spectra.[28] They suggested that the CD spectra of membrane proteins were in some way fundamentally different from those of soluble proteins and that in order to be able to solve for their fractional composition a separate membrane protein reference set would be required. Sreerama and Woody subsequently created such a reference set and incorporated it within the CDPro software.[2] Their analysis of membrane proteins, however, suggested no such fundamental difference between membrane protein CD spectra and those of soluble proteins. The inclusion of 13 membrane proteins to the soluble protein reference sets of CDPro enhanced the performances of CONTIN, SELCON and CDSSTR for both soluble and membrane proteins alike, especially in the estimation of β-sheet content. The authors suggest that the enhanced performance in analysing membrane proteins resulted from there being a greater number of proteins in the new reference sets, rather than being due to some uniqueness of membrane protein CD spectra, since, contrary to the findings of Wallace *et al.*, the soluble protein reference sets performed as well when analysing both membrane and soluble proteins. Table V.4 compares the performance of the soluble reference set with the combined soluble and membrane reference set in estimating the fractional compositions of both soluble and membrane proteins.

Method	Reference set	δ_α	r_α	δ_β	r_β	δ_T	r_T	δ_U	r_U	δ	r
SELCON3	SP37: 240-185	0.08	0.94	0.11	0.71	0.06	0.57	0.11	0.15	0.09	0.80
	SP43: 240-190	0.08	0.93	0.11	0.71	0.03	0.37	0.10	0.22	0.09	0.80
	SMP50: 240-185	0.09	0.92	0.11	0.72	0.06	0.50	0.11	0.18	0.09	0.79
	SMP56: 240-190	0.07	0.93	0.10	0.77	0.07	0.44	0.10	0.22	0.09	0.82
	SP37: 240-185	0.10	0.96	0.13	0.93	0.04	0.70	0.07	0.25	0.09	0.92
	SP43: 240-190	0.09	0.97	0.13	0.97	0.04	0.64	0.08	0.26	0.09	0.92
	SMP50: 240-185	0.07	0.97	0.08	0.96	0.03	0.68	0.06	0.25	0.06	0.95
	SMP56: 240-190	0.09	0.94	0.08	0.97	0.04	0.57	0.06	0.29	0.07	0.95
CONTIN/LL	SP37: 240-185	0.08	0.93	0.12	0.65	0.07	0.42	0.09	0.28	0.09	0.81
	SP43: 240-190	0.08	0.93	0.11	0.71	0.08	0.33	0.09	0.25	0.09	0.80
	SMP50: 240-185	0.08	0.94	0.11	0.75	0.05	0.58	0.09	0.28	0.09	0.83
	SMP56: 240-190	0.07	0.94	0.10	0.80	0.07	0.40	0.09	0.27	0.08	0.84
	SP37: 240-185	0.10	0.94	0.15	0.81	0.07	0.47	0.10	-0.06	0.11	0.85
	SP43: 240-190	0.10	0.93	0.10	0.96	0.05	0.52	0.08	0.25	0.09	0.91
	SMP50: 240-185	0.09	0.95	0.07	0.96	0.04	0.62	0.08	-0.01	0.07	0.94
	SMP56: 240-190	0.13	0.88	0.07	0.96	0.05	0.47	0.08	0.12	0.09	0.91
CDSSTR	SP37: 240-185	0.08	0.94	0.12	0.69	0.07	0.45	0.10	0.32	0.09	0.81
	SP43: 240-190	0.09	0.92	0.11	0.71	0.07	0.46	0.09	0.37	0.09	0.80
	SMP50: 240-185	0.09	0.94	0.11	0.75	0.06	0.59	0.09	0.44	0.09	0.84
	SMP56: 240-190	0.09	0.94	0.10	0.76	0.07	0.50	0.09	0.43	0.09	0.84
	SP37: 240-185	0.09	0.96	0.14	0.96	0.04	0.76	0.08	0.09	0.09	0.90
	SP43: 240-190	0.08	0.96	0.10	0.97	0.04	0.64	0.08	0.18	0.08	0.93
	SMP50: 240-185	0.08	0.96	0.08	0.99	0.05	0.75	0.08	0.42	0.08	0.93
	SMP56: 240-190	0.09	0.94	0.08	0.99	0.04	0.64	0.07	0.22	0.07	0.93

Table V.4. Performance indices of soluble and membrane proteins by the SELCON, CONTIN/LL and CDSSTR methods, with and without the inclusion of membrane proteins in the reference set. The greyed data corresponds to the performance of the 13 membrane proteins.(Reproduced from [2])

V.1.14 The Effect of Tertiary Structure on Protein Far-UV CD

It is well understood that protein far-UV CD spectra are not exclusively defined by their secondary structural fractional compositions, but are influenced, sometimes heavily, by the tertiary structure the protein adopts. As such, any lack of representation in the protein reference set of each of the different classes of tertiary structure could significantly reduce a method's performance when estimating fractional compositions. Figure V.1 shows the CD spectra of six protein pairs, wherein the proteins have similar fractional compositions, but differ in their tertiary structures. Another source of error in estimating fractional compositions from CD spectra is thought to lie with the accuracy of the X-ray crystal structures assigned to each protein within the reference set. Oberg *et al.* has developed a rationally selected basis protein reference set consisting of 50 proteins (RaSP50) in order to account for both of these factors.[5] The RaSP50 reference set was constructed in order to represent the greatest coverage of tertiary structure fold space, the greatest coverage of secondary structure fractional composition space and to include only proteins whose crystal structure had been solved to a high degree of accuracy. The condition that all of the included proteins be commercially available was also applied.

On comparing the performance of the RaSP50 basis set with that of the basis sets used in the CDPro software for the SELCON3 method, the correlation coefficients for each structure do not vary considerably, suggesting that neither insufficient tertiary structure space coverage nor inaccurate X-ray crystallographic data for the CDPro basis sets significantly degrade the estimation of secondary structure fractional composition. In light of the spectral differences shown in Figure V.1 this is somewhat surprising. Oberg *et al.* tested the performance of a reduced RaSP16 basis set in estimating the structure of a protein occupying a region of tertiary structure space outside of that covered by the RaSP16 basis set and concluded that in regions where tertiary space coverage was poor for a basis set, the basis set was inadequate for accurate fractional composition analysis by CD spectroscopy. This result suggests that a study of the tertiary structure space coverage of the basis sets within CDPro should be performed and that the software should only be used for analysing proteins whose tertiary structures occupy space covered by the basis sets.

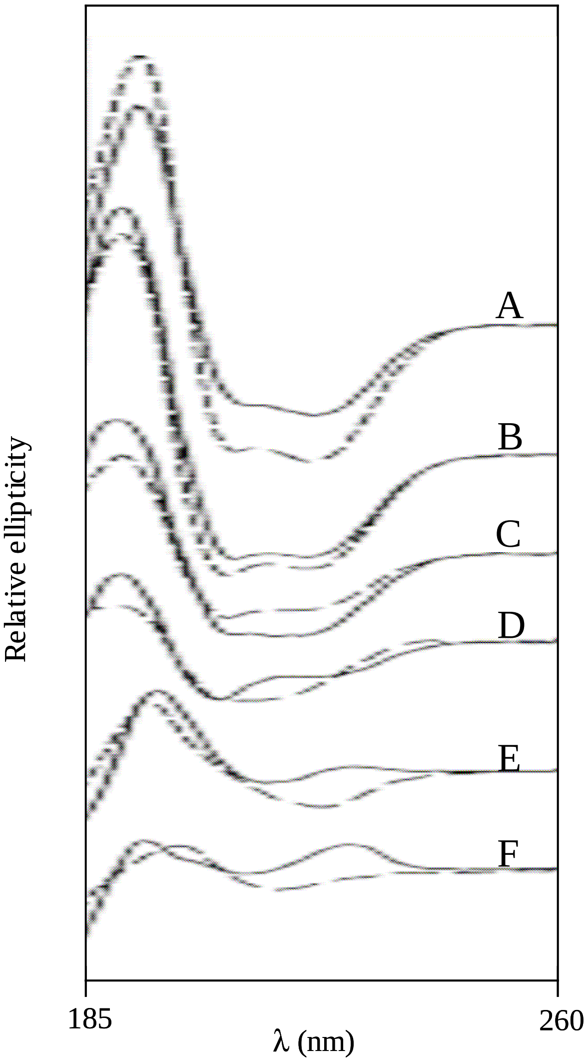


Figure V.1. The effect of tertiary structure on protein far-UV CD spectra. The relative intensities of the spectra are proportional to their mean residue ellipticities. The protein pairs are as follows, where the protein CD represented by the solid line is listed first and the protein CD plotted as a dotted line is listed second and the protein PDB codes are also listed: A Ferritin (FTN) and Hemoglobin (HBN); B Citrate synthase (CSA) and Troponin (TRO); C Lipoxygenase (LOX) and Phosphoglycerate kinase (PGK); D Ubiquitin (UBQ) and Ribonuclease A (RNA); E Erabutoxin (BTE) and Concanavalin A (CAN); F Avidin (AVI) and γ immunoglobulin (IGG). (Modified from [5])

V.2 Characterisation of Cuvettes for CD Use

Considering the optical range over which CD spectra are acquired (260-178 nm) it is essential to choose a sample cuvette (cell) composed of material which is transparent over this range. SUPRASIL quartz is the most commonly used material used to make cells for CD spectroscopy because of its wide wavelength range of transparency, cost effectiveness, tolerance of exposure to a range of solvents and relative durability. When performing protein CD for the purpose of quantitative analyses of secondary structure all spectra must be in units of molar ellipticity. As such, it is essential that both the concentration of the sample solution be known to as precise a value as possible and that the optical pathlength be precisely known. When purchasing cells of pathlengths smaller than 1 mm, however, it is common that there can be as much as a 50 % difference between the actual and nominal pathlengths. Therefore, it is essential that all cells to be used in quantitative analyses of protein secondary structure by CD spectroscopy have their pathlengths experimentally verified. Cell pathlength determination can be by either the K_2CrO_4 method or the interference fringe method.

The K_2CrO_4 method is used for determining the pathlengths of longer pathlength cells and relies on the known extinction coefficient of K_2CrO_4 at 372 nm of 4830 Mol^{-1} cm^{-1}. By using Beers Law of $A = \varepsilon cl$, where A is the absorption at a specified wavelength, ε is the extinction coefficient at the same wavelength, c is the sample molar concentration and l is the cell pathlength in cm, the cell pathlength can be determined from the UV/visible absorption spectrum of a standard solution of K_2CrO_4. The following experiment was performed in order to determine the pathlengths of cells of nominal pathlength greater than 0.01 mm.

V.2.1 Method

Five stock solutions of 0.02 M K_2CrO_4 were made up by accurately weighing out 1.9419 g of K_2CrO_4 and dissolving it in 500 ml of 0.01 M KOH solvent made up with deionised water. Serial dilutions of x10 and x100 were then made of these stock solutions, producing five K_2CrO_4 solutions of 0.002 and 0.0002 M, respectively. Additionally, five solutions of 0.2 M K_2CrO_4 were made up by dissolving 1.9419 g of K_2CrO_4 in 50 ml of 0.01 M KOH solvent. Using Hellmanex®II cleaning solution (HELLMA UK INC.) and 105 Lens Cleaning Tissue (Whatman®) all cuvettes were thoroughly cleaned and dried, whilst wearing gloves, until they showed up as spotless under lighting.

To determine the pathlengths of the nominal 1, 5 and 10 mm cells the 0.002, 0.0002 and 0.0002 M K_2CrO_4 solutions were used, respectively, and the absorption spectrum of each recorded on a Cary 50 Conc (Varian) absorption spectrometer (Table V.5). For pathlength determination of the nominal 0.1 and 0.01 mm demountable cuvettes, the 0.02 and 0.2 M K_2CrO_4 solutions were used, respectively. A 1 ml volume of the 0.02 M K_2CrO_4 solution was degassed for approximately 10 minutes and a 40 μl drop was placed onto the bevelled half of the 0.1 mm demountable cuvette. The cuvette was assembled by placing the blank half on top of the bevelled half with the solution drop in between, being careful to ensure that the bevelled volume filled completely, that the cuvette did not overflow and that no gas bubbles were present within the cell. The nominal 0.01 mm cuvette was filled with a drop of volume 6 μl in a similar manner. In an attempt to prevent leakage from the 0.01 mm demountable the added precaution of sealing the edges of the assembled cell with a thin strip of Parafilm was taken. Beers Law was used together with the absorbance at 372 nm and the extinction coefficient of K_2CrO_4 at 372 nm of 4830 $cm^{-1}M^{-1}$ to calculate the value of *l* according to each of the five solutions for each cell. The largest and smallest calculated values of *l* were ignored and the average of the three remaining intermediate values was taken to be the true value of *l*.

V.2.2 Results

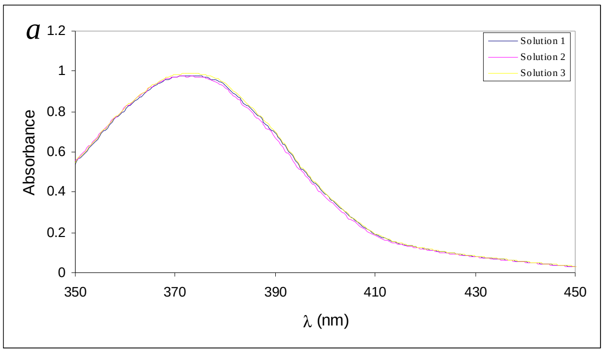
Table V.5 summarises the results of the determined pathlengths of each of the cuvettes to be used in CD spectroscopy according to the K_2CrO_4 method. With the exception of the 0.01 mm nominal cuvette, the determined pathlengths for all cuvettes were within 2 % of their nominal values. From Figures V.2, V.3, V.4 and V.5 it can be seen that the deviation from the calculated average of the three most similar absorption spectra was very small, so that in most cases the absorption spectra of the different K_2CrO_4 solutions almost exactly overlap.

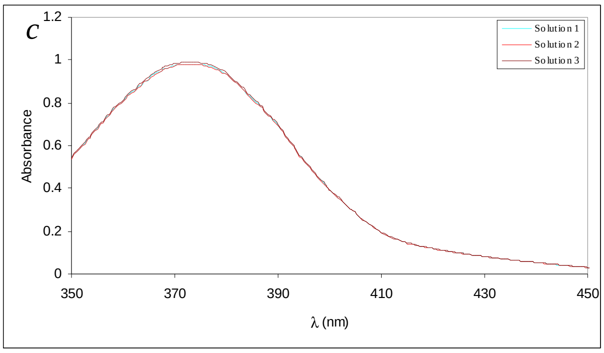
The variation of the pathlength of the demountable nominal 0.1 and 0.01 mm cuvettes due to either evaporation or leakage of solution from the cell can be seen from Figures V.4*b* and V.5, respectively. From the inset in Figure V.4*b* it seems that there occurs an initial loss of solution from the cell and a decrease of the cell pathlength from an initial maximum of 0.1033 mm (corresponding to an absorbance of 0.998) to 0.1019 mm, after a time of approximately 20 minutes. This latter pathlength is maintained for a period of at least 80 minutes. In contrast, the nominal 0.01 mm cuvette showed a continual and non-uniform decrease in cell pathlength due to solution loss, which can be seen from the decreasing intensity absorption spectra with time shown in Figure V.5. The insets in Figure V.5 show that in some cases the determined pathlength actually increases between time points. Also, the initial absorption spectra for different fills of the nominal 0.01 mm cell with a common K_2CrO_4 solution showed significant variation (data not shown). Figure V.5 shows that, although the pathlength cannot be assumed constant for the nominal 0.01 mm cell, it does decrease to a final pathlength of approximately 0.013 mm (corresponding to an absorbance of 1.28) after a time of 2 hours. Although the application of Parafilm to the edges of the cell did not prevent leakage from the cell (Figures V.5*c* and *d*) it has been seen that it does improve the retention of solution within the cell over very long time periods (greater than 3 hours)(data not shown) and so perhaps retards the evaporation of solution from the cell.

For very short pathlength cells the interference fringe method can be used to determine the pathlength and Figure V.6 shows the interference spectrum of the nominal 0.01 mm cuvette. Using the formula

$$\left(\frac{n\left(W_1 \cdot W_2\right) \Big/ 2\left(W_2 - W_1\right)}{1000} \right) \qquad \text{(V.4)}$$

where W_1 and W_2 are the wavelengths of fringe peaks at least ten peaks apart, n is the number of fringe peaks separating W_1 and W_2. When the wavelengths are given in nm equation V.4 provides the pathlength in μm. From Figure V.6 the pathlength of the nominal 0.01 mm cell is calculated as 15.983 μm. However, since this is the pathlength of the cell when empty, it might be more accurate to use the cell pathlength determined by the K_2CrO_4 method for CD spectroscopy purposes.





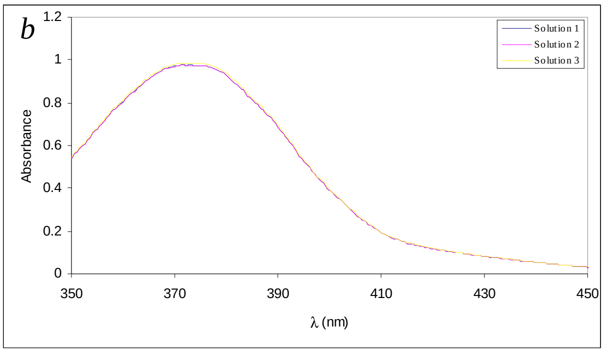
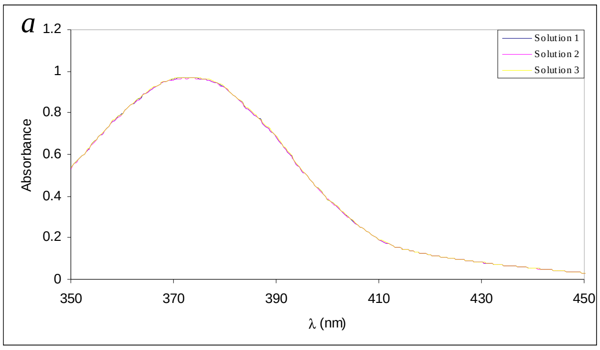
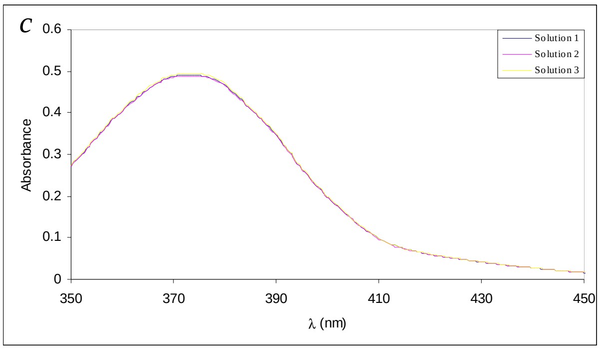


Figure V.2. Pathlength determination of the nominal 1 cm pathlength cuvettes by the K_2CrO_4 method. The absorption spectral overlays of three separate 0.0002 M K_2CrO_4 solutions for the nominal 1 cm pathlength cuvettes, numbered 1, 2 and 3, are shown as *a*, *b* and *c*, respectively.





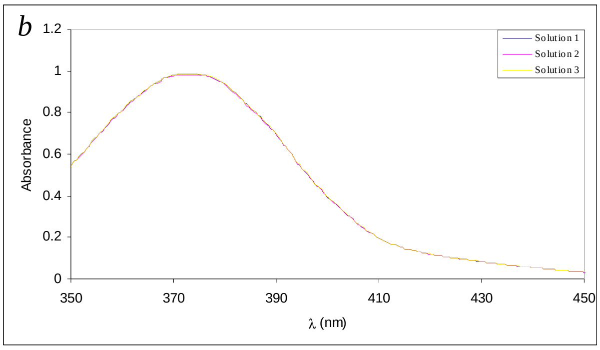


Figure V.3. Pathlength determination for nominal 1 mm and 5 mm pathlength cuvettes by the K_2CrO_4 method. *a* and *b* are the absorption spectral overlays of three different 0.002 M K_2CrO_4 solutions using the nominal 1 mm cuvette numbers 1 and 2, respectively. *c* is the absorption spectral overlay of three different 0.0002 M K_2CrO_4 solutions using the 5 mm cuvette.

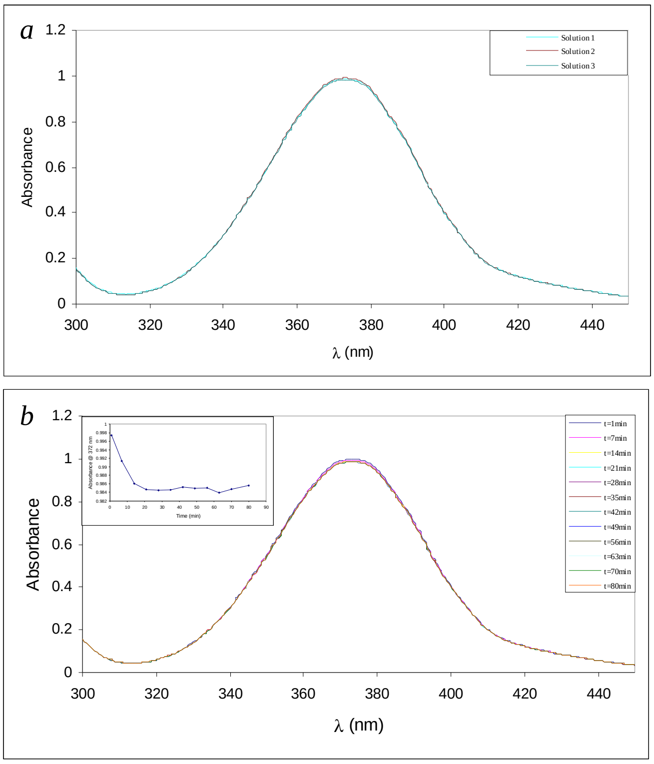


Figure V.4. Pathlength determination of the nominal 0.1 mm pathlength cell by the K_2CrO_4 method. *a* shows the absorption spectral ovelay of three different 0.02 M K_2CrO_4 solutions and *b* shows the variation with time of the filled 0.1 mm cell absorption spectra over a range of 80 minutes. Inset to *b* is this variation as measured at the peak maximum of 372 nm.

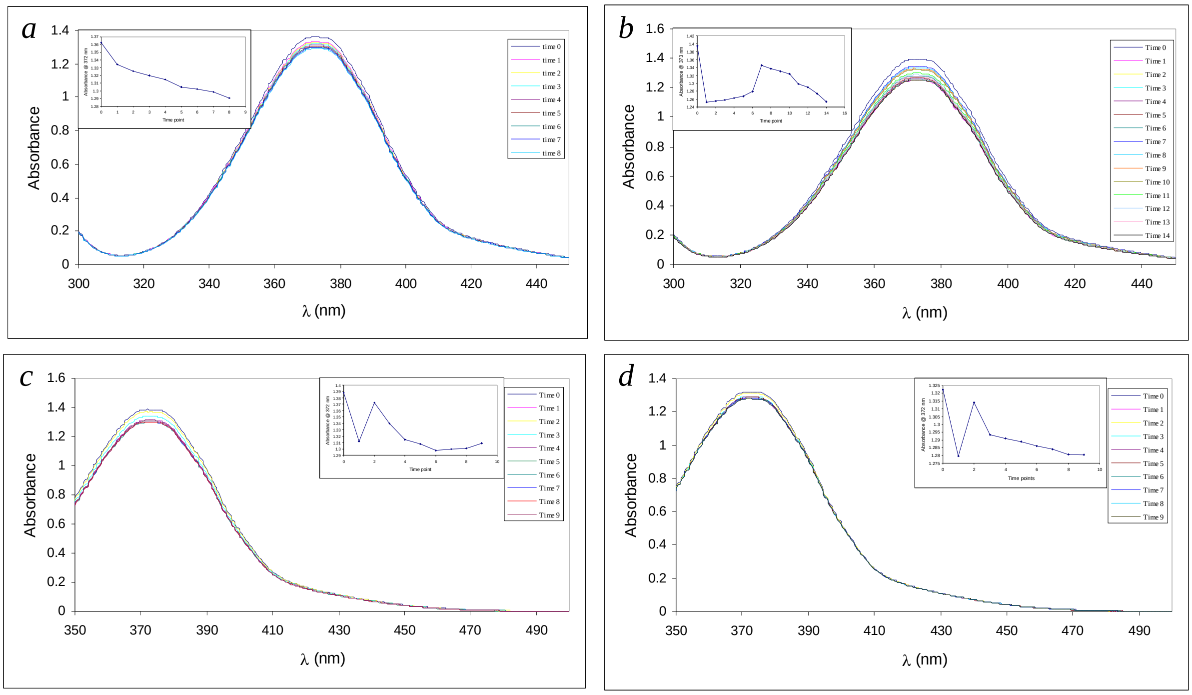


Figure V.5. Pathlength determination of the 0.01 mm nominal pathlength cell by the K_2CrO_4 method. Overlay of consecutive absorption spectra of a single 0.2 M solution of K_2CrO_4 acquired over a time of approximately 2 hours, with equal time intervals between each scan. Inset are the variations in absorbance at the peak maximum at 372 nm over the time interval. *a* and *b* were acquired without the use of Parafilm, while *c* and *d* were recorded after the cell had been sealed at the edges with Parafilm.

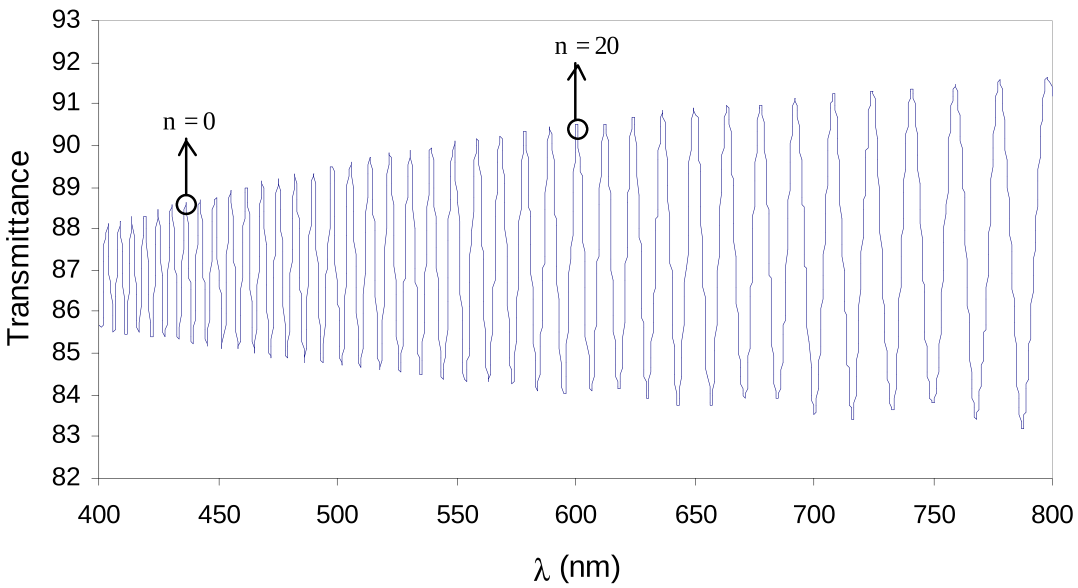


Figure V.6. Pathlength determination of the nominal 0.01 mm pathlength cell by the interference fringe method. The fringe labelled n = 0 is centred at 436.5 nm, while the n = 20 fringe is centred at 600.5 nm.

Cuvette	K_2CrO_4 sol. conc. (M)	$Abs_{372\,nm}$	Pathlength (cm)
1 cm No.1	0.0002	0.9797	1.014
1 cm No.2	0.0002	0.9793	1.0138
1 cm No.3	0.0002	0.9827	1.0173
5 mm	0.0002	0.4905	0.5078
1 mm No.1	0.002	0.9676	0.10017
1 mm No.2	0.002	0.9843	0.10189
0.1 mm	0.02	0.9845	0.01019
0.01 mm	0.2	1.39 to 1.26	0.001439 to 0.001304

Table V.5. Pathlength determination of SUPRASIL quartz cuvettes used for CD spectroscopy.

A further requirement for cells to be used for CD spectroscopy is that they should be strain free and thereby exhibit no intrinsic CD. As such, cells giving rise to deviations from the instrument air baseline of more than 5 millidegrees should be discarded for CD use, since these cells do not provide equal enough transmission between left and right circularly polarised light through the cell. SUPRASIL quartz cells certified as strain free by manufacturers are only certified over the recommended wavelength range for the use of this material, from 200 to 2500 nm. Because the effect of strain becomes more pronounced when working at lower wavelengths, it is required that all cells be tested for strain over the entire wavelength range used during CD spectroscopy. Furthermore, cells should not exhibit significantly different CD signals upon reversal of orientation. Figures V.7 and V.8 show the CD spectra of both the 0.1 and 0.01 mm cuvettes, respectively, in each of the four orientations possible for a rectangular demountable cell. The different orientations of both the 0.1 and 0.01 mm cells show approximately similar CD spectra. While the deviations in ellipticity over the wavelength range for protein CD is less than ideal for both cells, the requirement that these deviations be less than 5 millidegrees is satisfied. The degree of dependence of the cell CD spectrum on its orientation requires that when obtaining protein CD spectra the sample and reference spectra should be acquired with the cuvette in the same orientation.

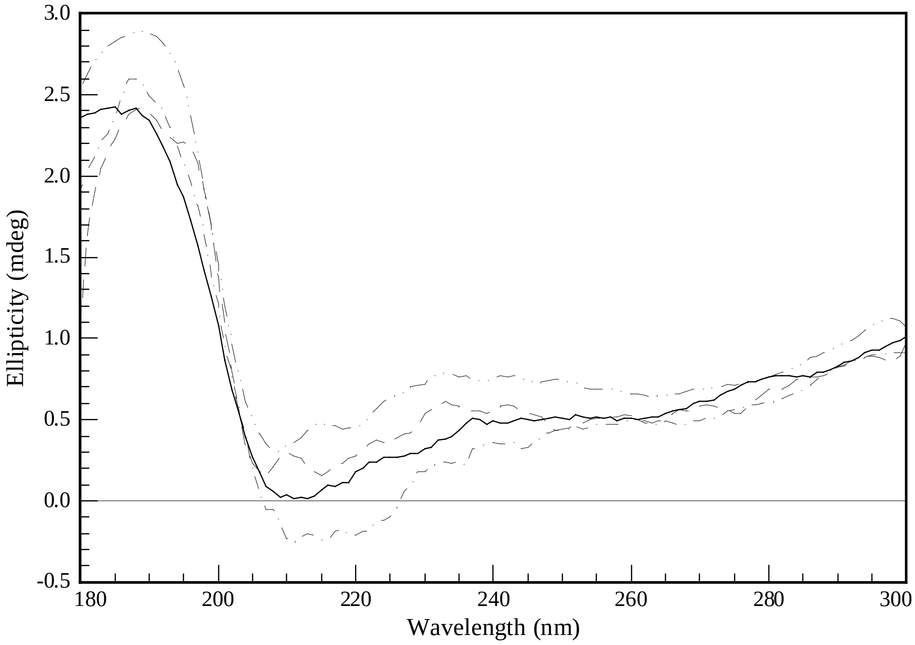


Figure V.7. CD of 0.1 mm cuvette in each of its four possible orientations.

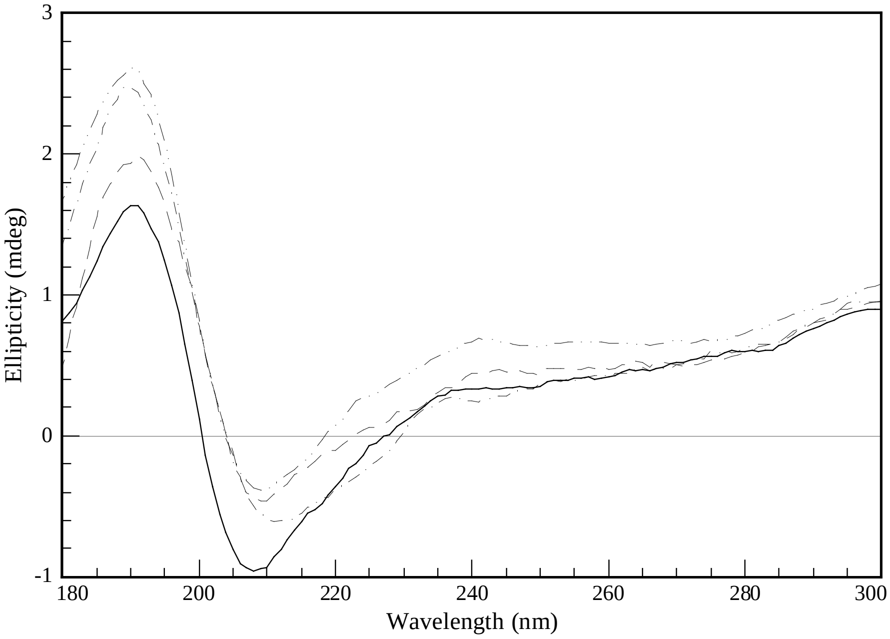


Figure V.8. CD of 0.01 mm cuvette in each of its four possible orientations.

V.3 Calibration of the Model 400 CD Spectrometer

The calibration of a CD spectrometer involves the standardisation of the spectrometer in terms of both its wavelength accuracy and its magnitude sensitivity or ellipticity. Instrument calibration is vital when performing CD spectroscopy both to maintain internally consistent spectra for the same instrument over a given time span and to insure that spectra obtained by one instrument are comparable to those obtained using other CD spectrometers. The latter consideration is of special importance when performing quantitative analysis of protein secondary structure because of the reliance of this procedure on a database of protein CD spectra that have been acquired on different CD spectrometers. The level of importance of proper calibration of CD spectrometers to the study of protein secondary structure by CD has been emphasised by the carrying out of a UK-wide study designed to compare the absolute calibrations of 27 CD spectrometers (NPL REPORT DQL-AS 009). A subsequent report based on the findings of this comparative study was designed in order to establish a standard method for the calibration of CD spectrometers (NPL REPORT DQL-AS 010). Both of these reports make use of the data contained within a third report, which was carried out by the European Chirality Services (E. C. S.) (TECHNICAL REPORT N^{O} 43) in order to characterise standards to be used for CD. Table V.6 summarises the findings of this report.

Compound	**Molar ellipticity** (mdeg)	**Molecular weight** (g)
(1S)-(+)-10-camphorsulfonic acid (CSA)	+7820	232.30
(1R)-(-)-10-camphorsulfonic acid (CSA)	-7820	232.30
(1S)-(+)-10-camphorsulfonic acid, ammonium salt (ACS)	+7910	249.33
(1R)-(-)-10-camphorsulfonic acid, ammonium salt (ACS)	-7910	249.33
(R)-(-)-Pantolactone	-16160	130.15
(S)-(+)-Pantolactone	+16160	130.15
Epiandrosterone	+10880	290.45
$2(+)_D$-[Coen$_3$]Cl$_3$•NaCl•6H$_2$O	+6431	857.78
$2(-)_D$-[Coen$_3$]Cl$_3$•NaCl•6H$_2$O	-6431	857.78

Table V.6. Standards for CD magnitude calibration.

The general standard used for the magnitude calibration of CD spectrometers to be used for the far-UV analysis of protein molecules is (1S)-(+)-10-camphorsulfonic acid, ammonium salt (ACS). ACS is preferred over (1S)-(+)-10-camphorsulfonic acid (CSA) due to the extreme hygroscopic nature of CSA, while ACS does not display any significant level of hygroscopicity. When calibrating CD spectrometers it is important to use calibrants that display optical activity within the wavelength range over which the spectrometer is intended to be used because calibration can decrease as one moves away from the wavelength of calibration. ACS is ideal for protein CD calibration because it displays a band at 290.5 nm and another at 192 nm and, therefore, it enables the calibration of the spectrometer over much of the wavelength range to be used for protein structural investigation. Ideally, calibration should be performed at as many distinct wavelengths as possible and this has been shown to be important in the standardisation and comparability of spectra obtained on multiple instruments.[29] As such, (R)-(-)-pantolactone is also used as a calibrant in addition to ACS, since it produces a CD band centred at 219 nm that dissects the two bands produced by ACS.

The linearity of response of the CD instrument at each of the separate wavelengths used for the magnitude calibration as a function of the concentration of the calibrant must also be established if the instrument is going to be assumed to perform independently of sample concentration. This is also a valuable test for the overall stability of the instrument.

The calibration of the CD spectrometer for wavelength accuracy is done by obtaining the CD spectrum of a solution of holmium oxide. Holmium oxide produces a series of sharp absorption bands between 680 and 230 nm and the precise wavelengths of each of these have been published by The National Institute of Standards and Technology (NIST). In addition to this the observed band maxima of the magnitude calibrants can be compared to the expected band maxima.

V.3.1 Method

A 0.029 M stock solution of 1S-ACS (Katayama Chemical) was made up by accurately weighing out 144.7 mg of 1S-ACS and dissolving it in 20 ml of deionised H_2O. Samples of x2, x4 and x8 dilutions were then made from this stock solution. A stock solution of approximately 7 mg/ml of 1S-CSA was made by dissolving 700 mg of CSA in 100 ml of deionised H_2O. The concentration of this stock solution was then adjusted such that it had an absorbance value close to 1. Samples of x2, x4 and x8 dilutions of the CSA stock solution were then made using deionised H_2O. A 0.0115 M (R)-(-)-pantolactone (Aldrich) stock solution was made up by accurately weighing out 149.6 mg of pantolactone and dissolving it in 100 ml of deionised H_2O. Samples of x2, x4, x8 and x16 dilutions were then made from this stock solution. The CD of all samples were recorded consecutively and on the same day as the samples were made and all samples were kept in a dark box until being scanned.

CD spectra were recorded using the 1 mm No.2 SUPRASIL cuvette (HELLMA UK). Each of both the 1S-ACS and 1S-CSA samples were scanned over the wavelength ranges 188 to 195 nm and 288 to 293 nm. Each of the pantolactone samples was scanned over the wavelength range 216 to 222 nm. Scans were acquired at a temperature of 23 ^{o}C, a bandwidth of 1 nm, an averaging time of 1 second and a scan stepwidth of 0.2 nm. Spectra were calculated as an average of between 5 and 20 scans and in all cases they were background corrected using a deionised water background acquired over the same wavelength range, in the same cell and orientation and using the same spectral parameters. All spectra were obtained under conditions whereby the dynode voltage was below the 600 volt limit.

The calculation of the expected ellipticities of each solution was performed using the equation

$$\theta_{\text{expected}} = \Delta\varepsilon \times 33,000 \times c \times l \qquad \text{(V.5)}$$

where θ_{expected} is the expected ellipticity at the wavelength of interest, $\Delta\varepsilon$ is the delta epsilon differential extinction coefficient value of the substance at the same wavelength, c is the molar concentration of the sample solution and l is the sample pathlength in cm. The $\Delta\varepsilon$ values for 1S-CSA at 290.5 nm, ACS at 291 nm and pantolactone at 219 nm are 2.37, 2.40 and –4.9 mdeg M^{-1} cm^{-1}, respectively. The $\Delta\varepsilon$ value for both CSA and ACS at 192 nm is –4.72 mdeg M^{-1} cm^{-1}. For the CSA solutions the concentrations were calculated from the absorption spectra using the extinction coefficient of 34.5 M^{-1} cm^{-1} at 285 nm and the Beer-Lambert equation

$$\text{Abs} = \varepsilon \times c \times l \qquad \text{(V.6)}$$

where ε is the extinction coefficient, c is the sample molar concentration and l is the sample pathlength in cm. Since the No.2 1 mm nominal pathlength cuvette was used to record all CD spectra and the No.1 10 mm nominal pathlength cell was used to record the CSA absorption spectra. Therefore, the value of l should be taken as 0.10189 cm when calculating the θ_{expected} values and 1.0140 cm when calculating the concentrations of the CSA solutions.

V.3.2 Results

When determining the calibration performance of a CD spectrometer it is important to identify the instrument’s calibration both in terms of its wavelength and absorbance accuracy. Both of these checks should be carried out over the wavelength region for which sample spectra are to be recorded. For protein CD work this region corresponds to between 260 and 178 nm. Furthermore, the absorbance accuracy determination of the instrument should be performed at more than one point within this wavelength range in order to rule out the possibility of wavelength dependent absorption aberrations. When performing the absorption accuracy test the instrument absorption should be examined as a function of calibrant concentration at each of the wavelengths investigated. When operating correctly there should exist a linear relationship between calibrant concentration and spectral absorbance. Due to the alteration of the instrument dynode voltage during spectral measurement it is important to carefully monitor this value over all wavelengths and ensure that it does not exceed the instrument specification of approximately 600 volts, beyond which the detector experiences saturation.

Figures V.9, V.10 and V.11 give the CD spectral overlays for the serial dilutions of the 1S-ACS, 1S-CSA and pantolactone CD calibrants, respectively. For each overlay the corresponding dynode voltage overlay is also presented. The signal-to-noise ratios of all spectra can be seen to be good. Since both ACS and CSA exhibit well-resolved CD bands of opposite ellipticity and centred at two wavelengths located at either end of the wavelength range relevant to protein CD spectroscopy, these standards are of particular used when calibrating instruments intended for this purpose. The CD spectral overlays are presented for both bands in Figures V.9 and V.10, however, due to both ACS and CSA having significantly greater extinction coefficient values for the lower wavelength band compared to the higher wavelength band, only the two most dilute solutions resulted in spectra having dynode voltage values below the 600 volt saturation point for the low-wavelength band. The use of pantolactone in conjunction with either CSA or ACS is recommended when calibrating the instrument for protein CD work, since it exhibits a band that is somewhat intermediate in wavelength between the two bands of these calibrants. As such, the recommended three-point calibration of Miles *et al.* is used here for the Model 400 absorbance calibration.[29]

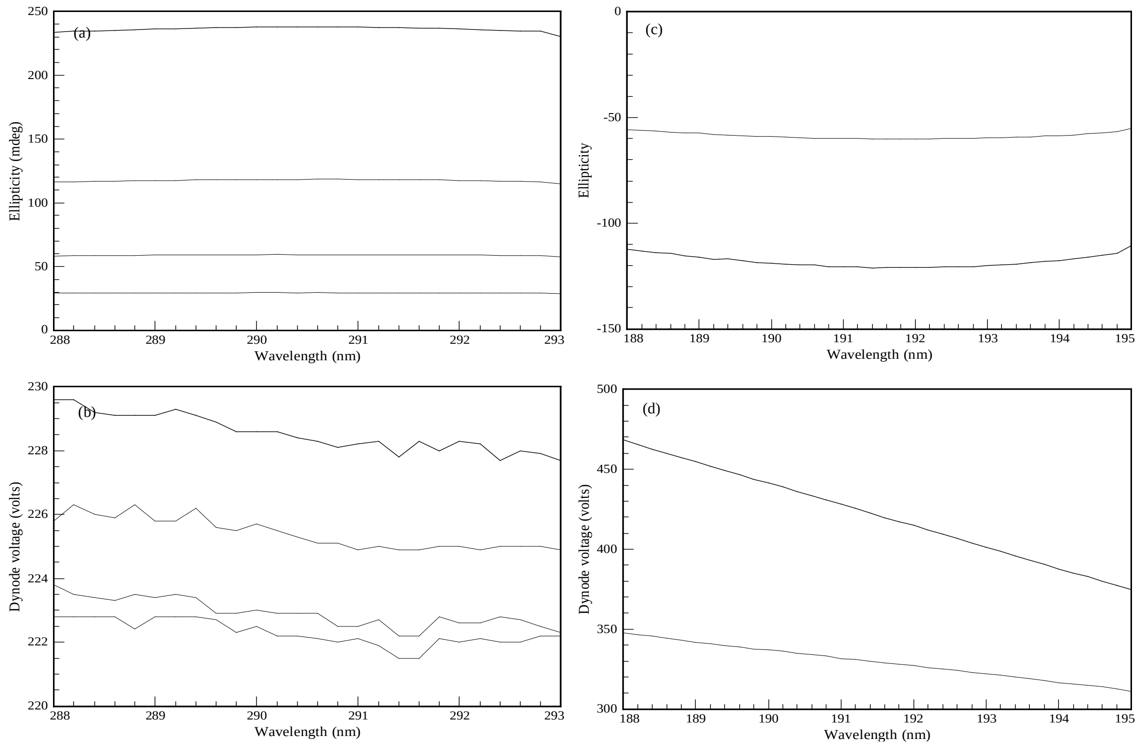


Figure V.9. Calibration of the Model 400 using ACS. (a) In descending order, the CD spectral overlay of ACS solutions of concentrations 0.029, 0.0145, 0.00725 and 0.003625 M showing the positive 290 nm band maximum. (b) In descending order, the CD dynode voltage spectral overlay of the respective solutions mentioned in (a). (c) In ascending order, the CD spectral overlay of the 0.00725 and 0.003625 M ACS solutions showing the negative 192.5 nm band minimum. (d) In descending order, the CD dynode voltage spectral overlay of the 0.00725 and 0.003625 M ACS solutions, as shown in (c).

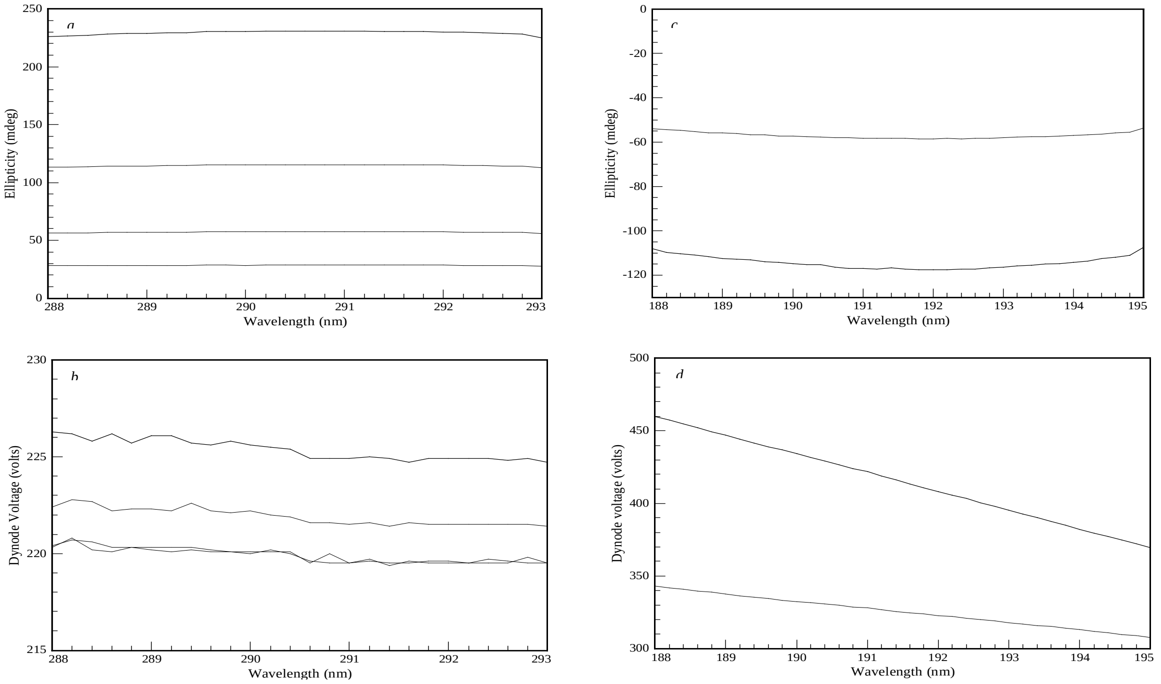


Figure V.10. Calibration of the Model 400 using CSA. *a* In descending order, the CD spectral overlay of a CSA stock solution along with x2, x4 and x8 dilutions showing the positive 290 nm band maximum. *b* In descending order, the CD dynode voltage spectral overlay of the respective solutions mentioned in *a*. *c* In ascending order, the CD spectral overlay of the x4 and x8 CSA stock solution dilutions showing the negative 192.5 nm band minimum. *d* In descending order, the CD dynode voltage spectral overlay of the x4 and x8 CSA stock solution dilutions, as shown in *c*.

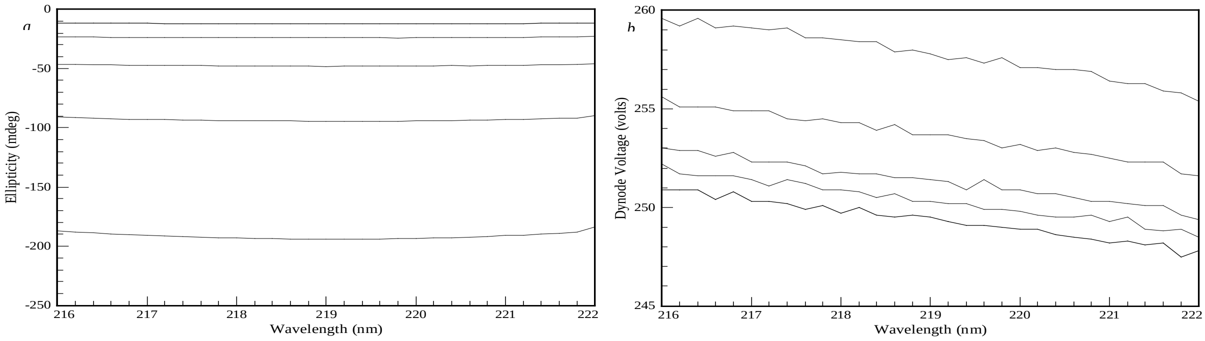


Figure V.11. Calibration of the CD spectrometer using the calibrant pantolactone. *a* In ascending order, the CD spectral overlay of a pantolactone solution of concentrations 0.0115, 0.00575, 0.002875, 0.0014375 and 0.00071875 M. *b* In descending order, the dynode voltage spectral overlay of the respective solutions mentioned in *a*.

The use of both ACS and CSA to calibrate the Model 400 at identical wavelengths was intended to make the calibration process somewhat independent of the concentration determining steps of each of the serial dilution solutions. The concentration of ACS can be determined gravimetrically, however, this method is vulnerable to the purity of the compound preparation and the calibration of the scales used. In contrast, the determination of the concentration of the CSA solutions must be done by absorption measurements, owing to the hygrophobicity of CSA rendering gravimetric determination non-feasible. This method is viable because of the accuracy with which the extinction coefficient of CSA has been determined, however, it is dependent upon the absolute calibration of the absorption spectrometer used in recording the solution absorbance values. Figure V.12 shows the absorption overlay of each of the serial dilutions of CSA used in the calibration. Using a combination of both ACS and CSA calibration was intended to highlight any possible errors in either the purity of the ACS or the calibration of the absorption spectrometer. Figure V.13 shows the linearity-of-response of the absorption spectrometer with concentration. The R^2 value of 1 combined with a y-axis intercept of –0.0002 suggests that the absorption spectrometer displays excellent linearity of absorbance with concentration change.

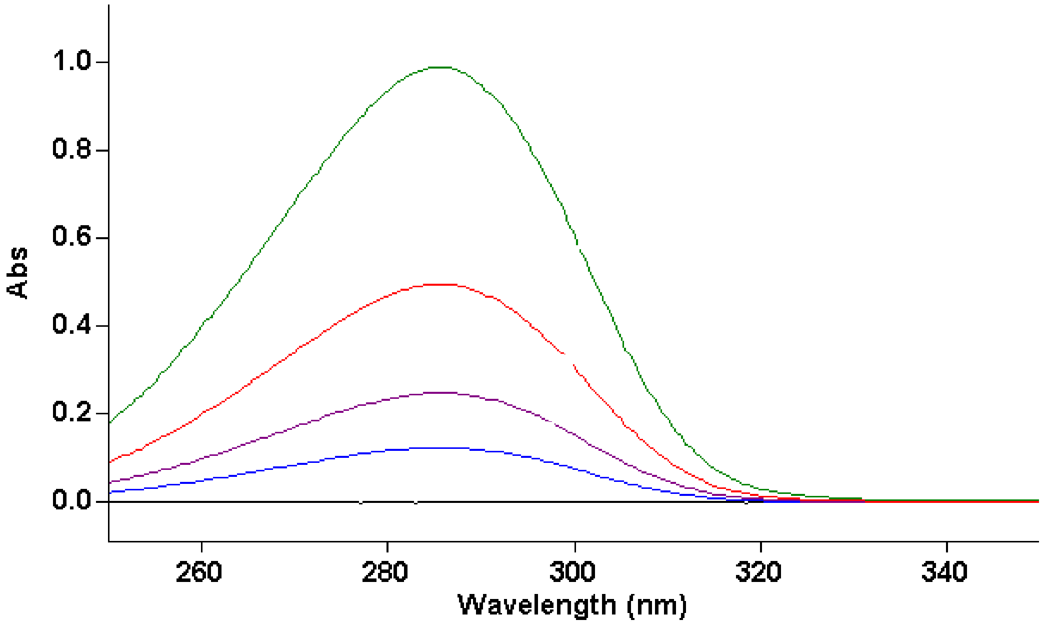


Figure V.12. The absorption spectral overlay of the CSA calibrant solutions. The stock solution and x2, x4 and x8 dilutions spectra appear in descending order, respectively, within the overlay. The water baseline absorption spectrum is also shown as a relatively flat line at zero absorbance.

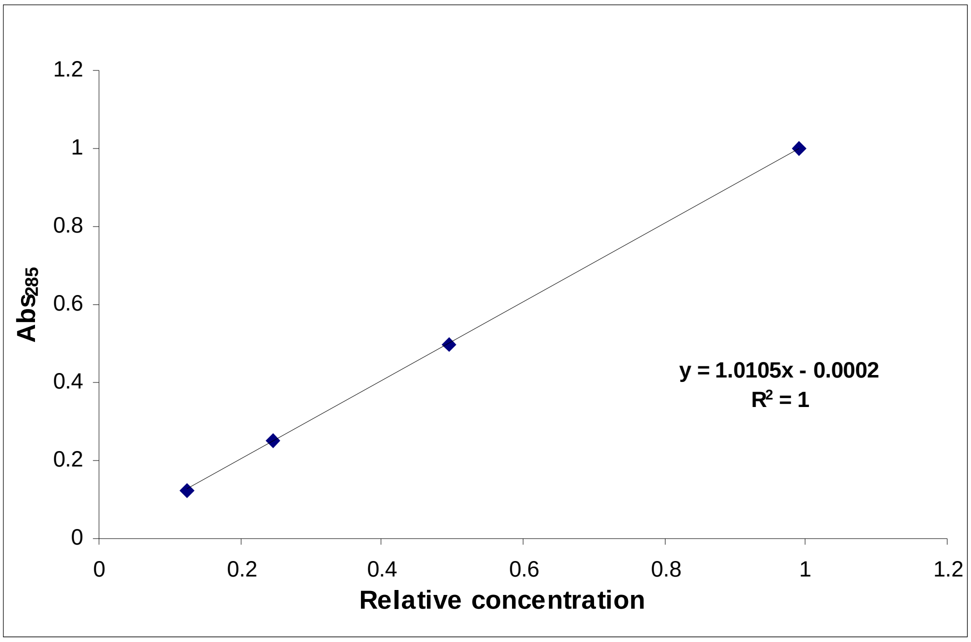


Figure V.13. Linearity of response of absorption spectrometer with 1S-CSA concentration. The equation of the straight-line fit of the four absorption values of the four CSA solutions at 285 nm is shown along with the error of this fit as the R^2 value.

Tables V.7, V.8 and V.9 show the accuracy to which the Model 400 is calibrated in terms of its absorbance by comparing the expected and the experimental ellipticity for the relevant bands of the ACS, CSA and pantolactone calibrant CD spectra, respectively. At 290 nm, for both the ACS and CSA calibrants, and at 219 nm, for the pantolactone calibrant, the percent error decreased considerably when only a small concentration of the calibrant was used. Because the accuracy of the PMT detector decreases at higher ellipticity values, the calibration accuracy of a CD instrument should be assessed at magnitudes comparable with the range over which samples are to be measured. The most dilute calibrant solutions for each of the solutions is most representative of this ellipticity region and so it is from these solutions that the Model 400 absolute calibration accuracy should be determined. As such, errors of approximately 0.8 and 1.5 % were observed at 290.5 nm for the ACS and CSA calibrants, respectively.

The ratio of the band magnitudes at 192 and 290 nm for either of the ACS and CSA calibrants is an important performance indicator of CD spectrometers. Because of solvent absorption and reduced short-wavelength lamp intensity effects, the ACS and CSA CD short-wavelength spectral bands tend to contain significant amounts of noise. Although the number of scans to be averaged was considerably increased the averaged spectra still retained some noise contribution. Due to such problems the literature is undecided as to the exact expected magnitude of the ACS/CSA short-wavelength band. Instead it is suggested that a ratio of the long- and short-wavelength bands should be greater than 2 for an accurately calibrated CD spectrometer. Table V.7 shows that this ratio for the ACS calibrant is consistently greater than 2. Table V.9 shows that the absolute magnitude calibration error at 219 nm of the Model 400 is approximately 0.9 %.

1S-ACS solution	**Relative conc.**	**Expected CD $\theta_{290.6}$ (mdeg)**	**Actual CD $\theta_{290.6}$ (mdeg)**	**$\theta_{290.6}$ Actual/Expected**	**Actual CD -$\theta_{191.5}$ (mdeg)**	**Actual $\theta_{191.5}$/ $\theta_{290.6}$**
1	1	237.721	233.648	1.0174	-	-
2	0.5	118.293	116.924	1.0117	-	-
3	0.25	59.324	58.462	1.0147	120.748	2.0354
4	0.125	29.444	28.231	1.0073	60.287	2.0475

Table V.7. The Model 400 calibration at 290.5 nm using the 1S-ACS CD calibrant.

1S-CSA solution	**Relative conc.**	**Abs$_{286}$**	**Expected CD $\theta_{290.6}$ (mdeg)**	**Actual CD $\theta_{290.6}$ (mdeg)**	**Actual/Expected**
1	1	0.9899	223.163	230.848	1.0344
2	0.5	0.4952	111.638	115.571	1.0352
3	0.25	0.2462	55.503	57.588	1.0376
4	0.125	0.1250	28.180	28.595	1.0147

Table V.8. The Model 400 spectrometer calibration at 290.5 nm using the 1S-CSA CD calibrant.

Pantolactone solution	**Relative conc.**	**Expected CD -θ_{219} (mdeg)**	**Actual CD -θ_{219} (mdeg)**	**-θ_{219} Actual/Expected**
1	1	189.448	192.208	1.0146
2	0.5	94.724	94.480	0.9974
3	0.25	47.362	48.219	1.0181
4	0.125	23.681	24.161	1.0203
5	0.0625	11.841	11.943	1.0087

Table V.9. The Model 400 spectrometer calibration at 290.5 nm using the pantolactone CD calibrant.

Figure V.14 reveals the relationship between the response of the Model 400 with CSA concentration at the 290.5 nm band, as measured by the Abs$_{285}$ value of each solution. At this wavelength the Model 400 was found to show excellent linearity of response by virtue of the R^2 value being equal to 1. The equation of the straight-line fit reveals that it passes almost precisely through the origin and this is as expected for an ideal magnitude-concentration instrument response.

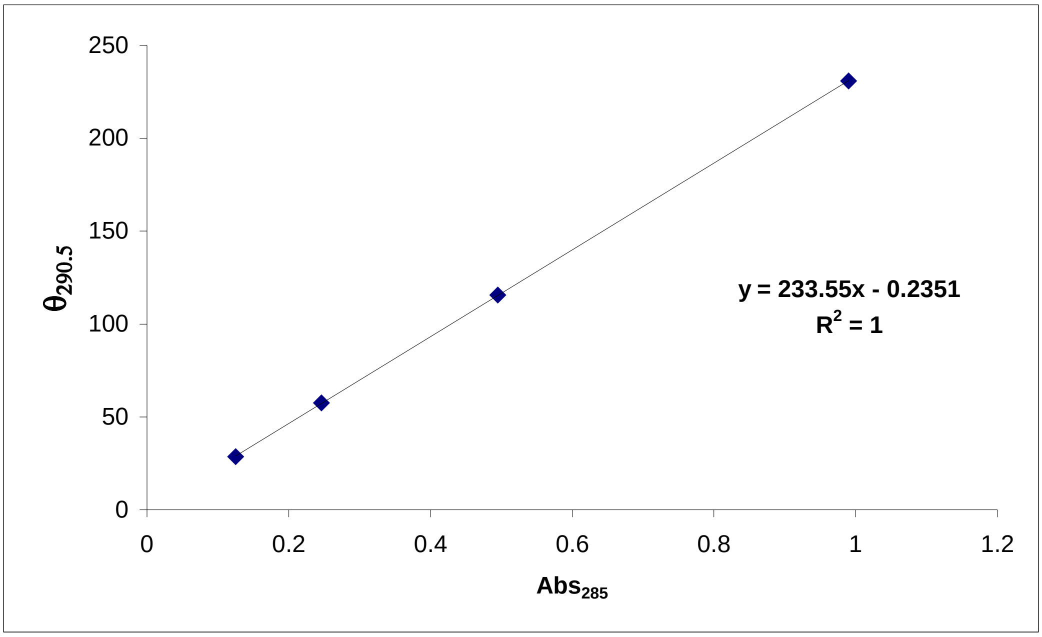


Figure V.14. Linearity of response of the Model 400 CD spectrometer at 290.5 nm with the CARY 50 CONC absorption spectrometer using 1S-CSA serial dilutions. The equation of the straight-line fit of the four data points is shown along with the R^2 fit error value.

In order to assess the linearity of response of the Model 400 ellipticity magnitude at 290.5 nm independent from the absorption spectrometer calibration, the ACS calibrant was substituted for CSA and the same plot was recreated (Figure V.15). Again, the Model 400 demonstrated excellent linearity of response with calibrant concentration, as can be seen from the $R^2 = 1$ value. The expected interception of the straight-line fit with the origin is again revealed by the line equation.

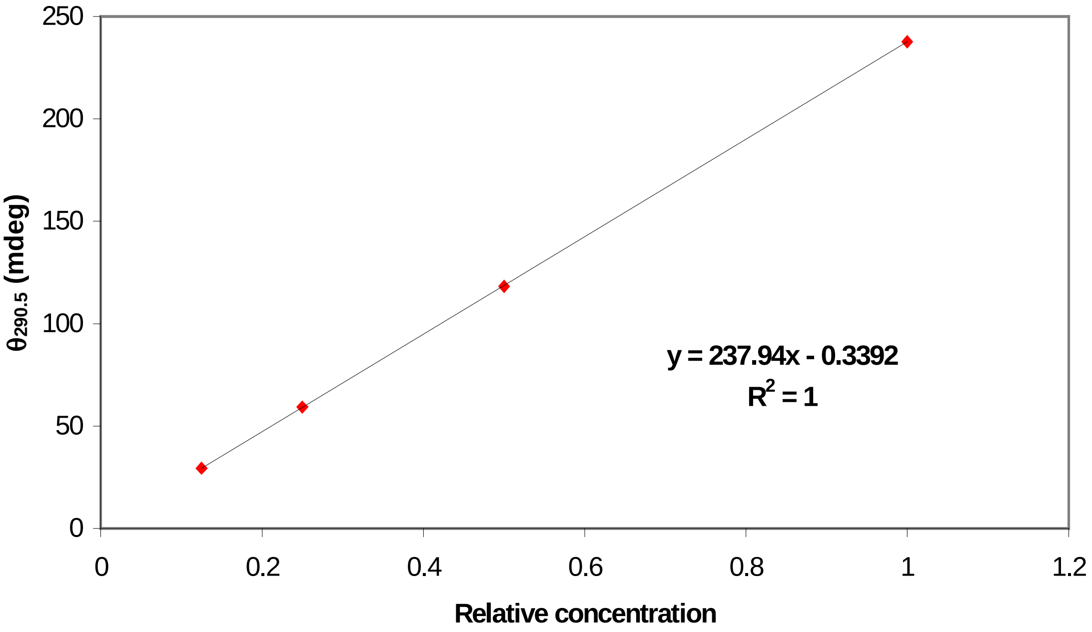


Figure V.15. Linearity of response of the Model 400 at 290.5 nm using the 1S-ACS CD calibrant. The equation of the straight-line fit to the four data points is shown along with the R^2 fit error value.

Examining the same concentration response at the wavelength of 219 nm for the Model 400 using the pantolactone calibrant resulted in the plot given as Figure V.16. As with the response at 290.5 nm, an almost exact response with calibrant concentration variation was observed, whereby an R^2 value of 1 was achieved for the straight-line fit, which passed through the origin, as expected.

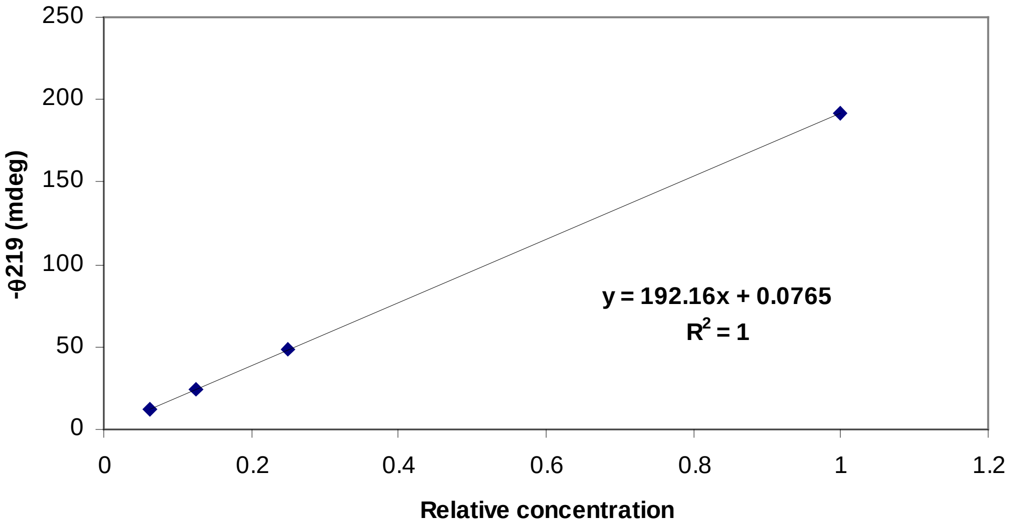


Figure V.16. Linearity of response of the Model 400 at 219 nm using the pantolactone CD calibrant. The equation of the straight-line fit to the four data points is shown along with the R^2 fit error value.

The wavelength calibration accuracy of the Model 400 is represented by Figure V.17 and Table V.10. The instrument calibration is accurate to approximately 0.5 nm at around 250 nm, but deteriorates considerably at longer wavelengths. Since protein CD spectroscopy involves far-UV radiation beyond 250 nm the Model 400 was optically optimised for use in this region. An accuracy of 0.5 nm is taken as within the accepted error for carrying out protein CD.

Band	**Literature λ_{max} (nm)**	**Model 400 λ_{max} (nm)**	**Model 400 wavelength error**
1	241.12	240.6	0.52
2	249.89	249.4	0.49
3	278.13	277.2	0.93
4	287.22	286.4	0.82
I		292.8	
5	333.48	332.2	1.28
6	345.38	343.6	1.78
7	361.25	359.4	1.85
8	385.61	383.6	2.01
9	416.25	414.0	2.25
10 (a)	450.96	448.2	2.76
11	467.82	463.2	4.62
Q		471.8	
12	485.23	481.4	3.83
13	536.56	532.6	3.96
14	640.50	635.0	5.50
U	652.67	649.0	3.67

Table V.10. Quantification of the wavelength calibration accuracy of the Model 400. The peak assignments are as given in Figure IV.10.

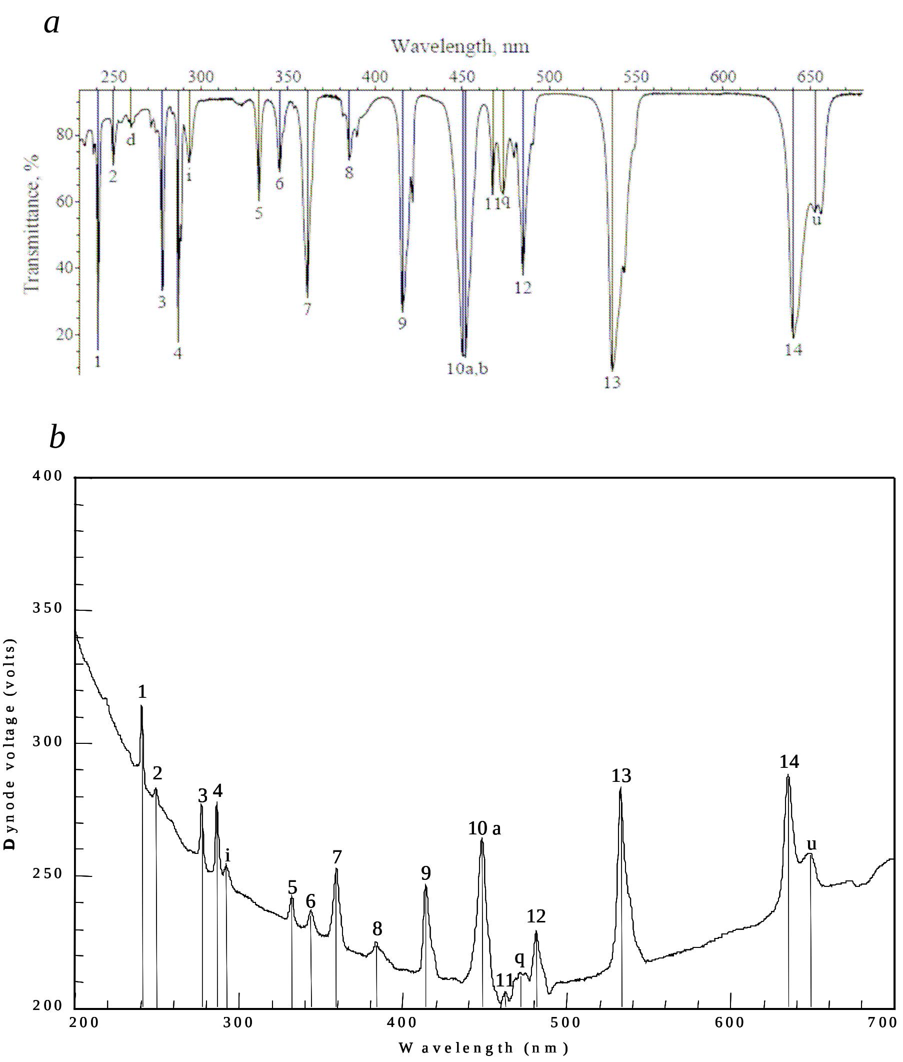


Figure IV.17. Wavelength calibration of Model 400 CD spectrometer. *a* The transmission spectrum of holmium oxide was taken from [7] and represents the accepted band positions of each of the labelled bands. *b* The CD dynode voltage of the holmium oxide standard as acquired on the Model 400. The precise wavelengths of each of the labelled bands are compared in Table IV.10.

V.3.3 Discussion

The absolute magnitude calibration accuracy of the Model 400 over the wavelength range applicable to protein CD and at magnitudes comparable to those of protein CD sample spectra is approximately to within a 0.8 % error (corresponding to the errors of the most dilute calibrant solutions). The higher error value of 1.5 % returned when using the CSA calibrant is attributed to a lack of accuracy of the absolute calibration of the absorption spectrometer used to determine the concentrations of the CSA samples. This conclusion is based on the error agreement measured by both the ACS and pantolactone calibrants. Although an absolute magnitude calibration error of 0.8

% is considerable it does lie within the acceptable limits for the performance of protein CD spectroscopy. The wavelength calibration error of the Model 400 was found to be approximately 0.5 nm over the wavelength range applicable to protein CD and this corresponds to a well-calibrated spectrometer. The response of the Model 400 ellipticity with sample concentration was found to be very accurate at both 219 and 290.5 nm. Calibration measurements performed subsequent to the above analysis confirm that the Model 400 displays a high level of stability over time (data not shown).

V.4 Quantitative Secondary Structure Analysis of Standard Proteins

With the aim of establishing a reliable and validated procedure for performing quantitative analysis of secondary structure by CD spectroscopy for novel proteins whose crystal structures have not yet been solved, it was first necessary to demonstrate that the experimental set-up was capable of accurately predicting the secondary structural fractional compositions of proteins with known structures. To this end a CD investigation into the structures of myoglobin, hemoglobin, lysozyme and α-chymotrypsin was performed. This involved accurately recording the CD spectrum of each protein to as low a wavelength as possible while obtaining a precise measure of the concentration of the protein within the scanned sample so that its CD spectrum could be analysed by the various programs discussed above.

V.4.1 Method

Samples of myoglobin (sperm whale), hemoglobin (horse), lysozyme (hen) and α-chymotrypsin (bovine) (Sigma) were prepared by dissolving lyophilised protein in 10 mM PBS buffer. The protein concentration of each sample was adjusted to a minimum value such that when placed in a 0.1 mm cuvette and scanned the resulting CD spectrum showed an adequate signal-to-noise ratio. Each sample was then sent for concentration determination by quantitative amino-acid (QAA) analysis (Cambridge University).

The CD spectrum of each protein was acquired using a 0.1 mm pathlength quartz cuvette. Spectra were acquired over a wavelength range of 250-178 nm, with an averaging time of 1 s, a step increment of 0.2 nm, a bandwidth of 1 nm and at a temperature of 23 ^{o}C. The step increment for the hemoglobin spectral acquisition was 1 nm. All spectra represented are an average of between 3 and 6 scans and were background corrected by subtracting the CD spectrum of the PBS buffer contained in the same 0.1 mm cell. Samples were centrifuged prior to scanning to remove any suspended particles and thoroughly degassed to prevent air bubble formation within the demountable cell. Spectra were converted to units of $\Delta\varepsilon$ from the machine units of θ by using the following equation:

$$\Delta\varepsilon = \theta \times \frac{\mathrm{MRW}}{10 \times lc} \qquad \text{(V.7)}$$

where MRW is the mean residue weight of the protein in g/mol, *l* is the optical pathlength in cm and *c* is the protein concentration in mg/ml. Quantitative analysis of the secondary structure of each of the proteins was performed using the Dichroweb server located at http://www.cryst.bbk.ac.uk/cdweb. The SELCON, CONTIN/LL, CDSSTR, VARSLC and K2d programs were used to analyse the fractional compositions of each protein over a range of protein reference sets (neither the VARSLC or K2d programs require a protein reference set). A spectral scaling factor value was adjusted such that the NRMSD value for each analysis reached a minimum.

V.4.2 Results

Figure V.18 shows the concentration determination results for each of the standard proteins by the QAA analysis method. The absolute concentrations of each sample were computed by fitting the area of each peak and in general the fitting was to better than 5 % accuracy for each amino acid. The concentrations of myoglobin, hemoglobin, lysozyme and α-chymotrypsin were calculated to be 1.69, 1.42, 1.90 and 1.54 mg/ml, respectively.

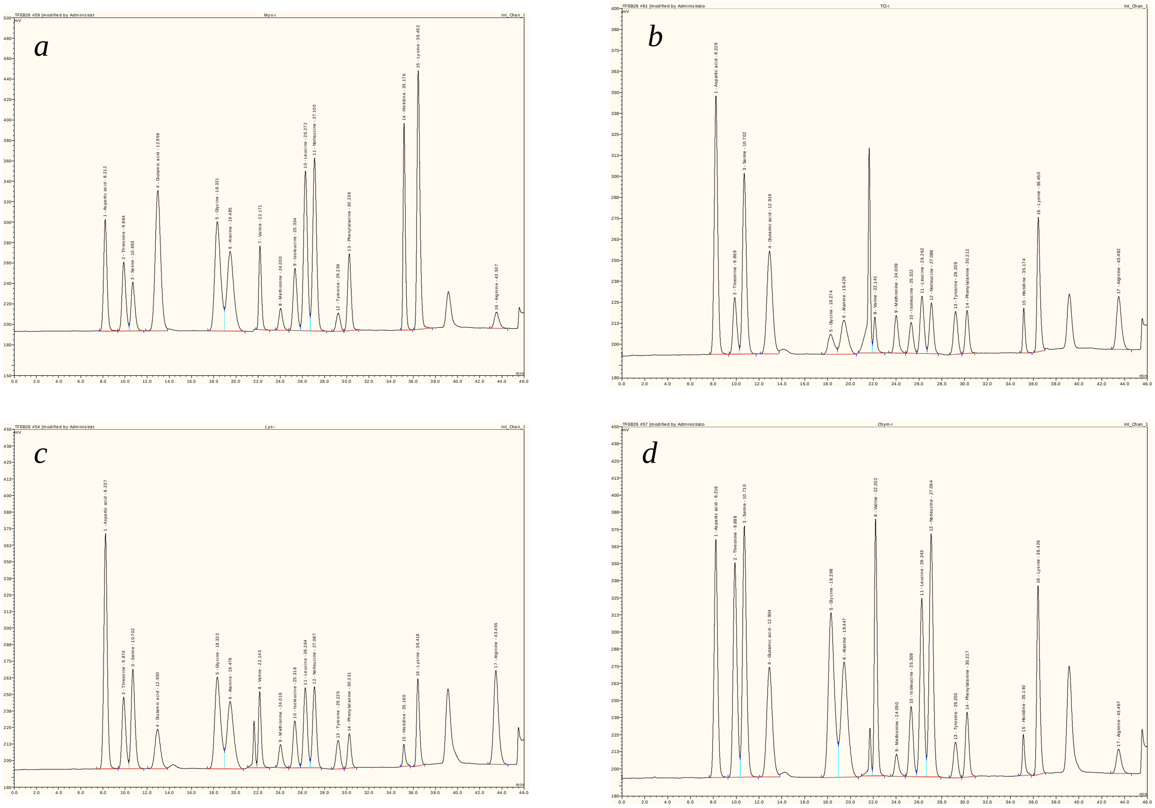


Figure V.18. The QAA analysis results for the four standard proteins. The amino acid analysis of myoglobin, hemoglobin, lysozyme and α-chymotrypsin are shown as *a*, *b*, *c* and *d*, respectively.

Figure V.19 shows the unsmoothed CD spectra of the four standard proteins under investigation. The CD error values are shown as overlaid. The signal-to-noise ratio for each spectrum can be seen to be very good. The CD error values are for the most part negligible, but they do increase slightly towards the shorter wavelength regions of each spectrum. From the presence of a band having intense minima at 222 and 208 nm of comparable magnitude and an intense maximum between 190-195 nm in both the myoglobin and hemoglobin spectra (Figures V.19*a* and *b*) it is clear that these proteins are predominantly α-helical in structure. The absence of a minimum at 222 nm in the spectrum of α-chymotrypsin reveals a lack of any large presence of α-helix within its structure (Figure V.19*d*). Its CD spectrum is somewhat irregular and of weak intensity and both of these are characteristic of CD spectra of proteins with high β-sheet contents. The CD spectrum of lysozyme displays moderate intensity at 222 and 190 nm, which is consistent with it having some significant α-helix contribution to its structure (Figure V.19*c*).

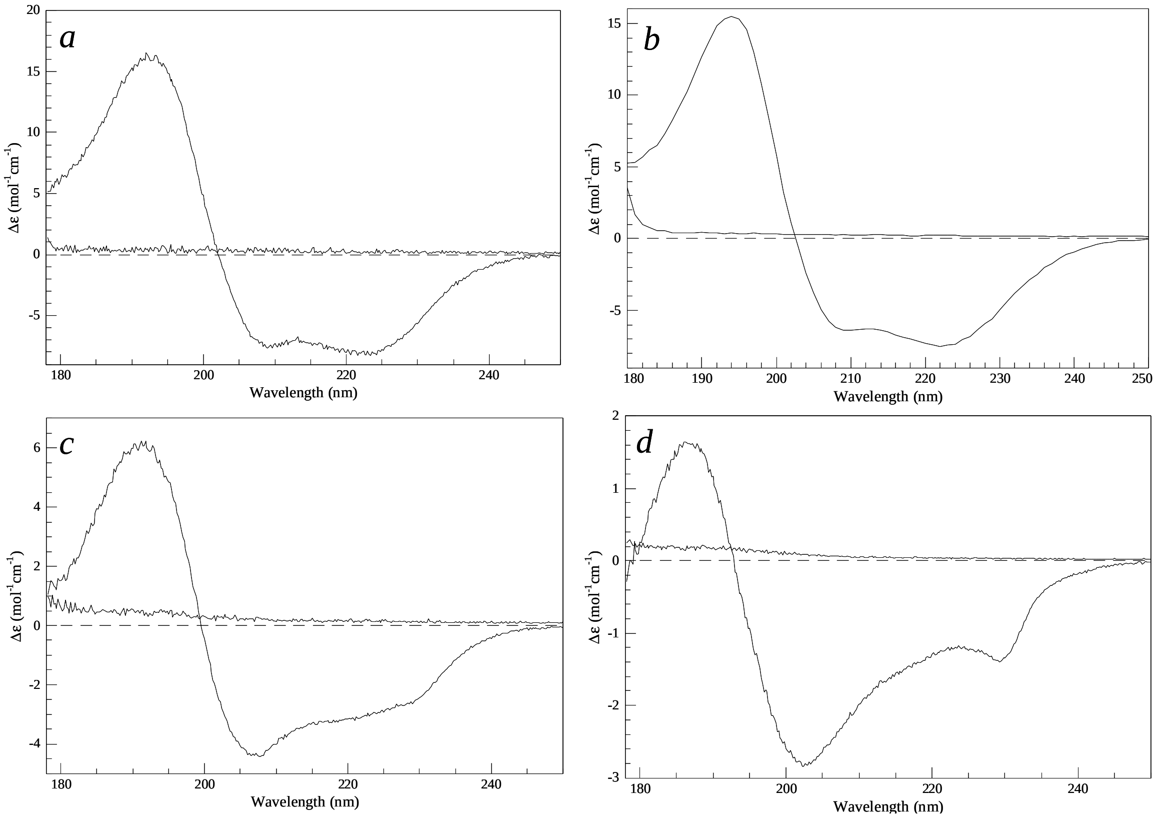


Figure V.19. CD spectra of standard proteins. The CD spectra of myoglobin, hemoglobin, lysozyme and α-chymotrypsin are given in units of Δε as *a*, *b*, *c* and *d*, respectively. The CD error bars for each wavelength are overlaid within each protein spectrum.

Table V.11 shows the fractional compositions of each of the four standard proteins as revealed by x-ray crystallography. Tables V.12, V.13, V.14, V.15 and V.16 give the quantitative analysis of secondary structure of the standard proteins using the SELCON, CONTIN/LL, CDSSTR, VARSLC and K2d programs, respectively. The separate analyses for each protein between each reference set (for the SELCON, CONTIN/LL and CDSSTR programs) and across all programs were very consistent, which suggests that the fitting procedure had been carried out successfully in each case. The NRMSD values were within the required limit of 0.25 given for an acceptable fit in the data presented below. The CDSSTR program was unable to perform analyses with reference sets containing data down to 178 nm (ref. Sets 1, 2 and 5). For Table V.17 an average of the fit results over all reference sets for the SELCON, CONTIN/LL and CDSSTR programs and the fit results of the VARSLC and K2d programs were used to calculate the accuracy of each analysis for each protein based on its crystal structure.

Protein	**α-helix**	**β-sheet**	**β-turn**	**P_{II}**	**Other**
Myoglobin	0.804	0.000	0.000	0.052	0.144
Hemoglobin	0.760	0.000	0.000	0.105	0.136
Lysozyme	0.419	0.016	0.047	0.298	0.221
α-chymotrypsin	0.114	0.208	0.106	0.200	0.371

Table V.11. The crystal structure fractional compositions of the four standard proteins from x-ray crystallography.

Protein	**Ref. set**	**H1**	**H2**	**S1**	**S2**	**T**	**U**	**Total**	**NMRSD**
Myoglobin	1	0.539	0.238	-.017	-.021	0.065	0.180	0.983	0.067
	3	0.507	0.225	0.017	0.013	0.075	0.159	0.995	0.102
	4	0.546	0.226	0.025	0.010	0.022	0.168	0.998	0.090
	6	0.507	0.225	0.017	0.013	0.075	0.159	0.995	0.102
	7	0.546	0.226	0.025	0.010	0.022	0.168	0.998	0.091
Hemoglobin	1	0.550	0.247	-.020	-.021	0.069	0.149	0.974	0.038
	3	0.535	0.256	0.001	-.009	0.076	0.152	1.011	0.030
	4	0.517	0.239	0.008	0.003	0.096	0.141	1.004	0.034
	6	0.537	0.255	0.005	-.008	0.070	0.148	1.006	0.027
	7	0.517	0.239	0.008	0.003	0.096	0.141	1.004	0.034
Lysozyme	1	0.200	0.156	0.078	0.066	0.209	0.275	0.984	0.189
	3	0.208	0.184	0.049	0.061	0.230	0.260	0.992	0.105
	4	0.213	0.186	0.053	0.062	0.212	0.270	0.996	0.212
	6	0.207	0.181	0.047	0.060	0.228	0.267	0.990	0.113
	7	0.201	0.189	0.045	0.059	0.222	0.267	0.982	0.182
α chymotrypsin	1	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-
	4	0.068	0.051	0.184	0.103	0.203	0.348	0.957	0.082
	6	0.039	0.039	0.168	0.089	0.181	0.452	0.968	0.161
	7	0.032	0.048	0.175	0.095	0.197	0.409	0.956	0.126
		αH	**3_{10}H**	**S**	**P_{II}**	**T**	**U**	**Total**	**NMRSD**
Myoglobin	2	0.721	0.112	0.000	0.018	0.054	0.109	1.014	0.048
	5	0.781	-	-.001	0.049	0.072	0.100	1.001	0.029
Hemoglobin	2	0.720	0.092	-.011	0.015	0.052	0.126	0.994	0.036
	5	0.687	-	0.055	0.040	0.112	0.113	1.007	0.089
Lysozyme	2	0.300	0.079	0.116	0.046	0.131	0.321	0.994	0.202
	5	0.359	-	0.144	0.227	0.083	0.186	0.999	0.226
α-chymotrypsin	2	0.092	0.033	0.180	0.122	0.125	0.437	0.989	0.158
	5	0.158	-	0.221	0.159	0.211	0.153	0.904	0.118

Table V.12. Analysis of the secondary structure fractional composition of standard proteins using the SELCON program. The various secondary structures are defined as α-helix (αH), α-helix type 1 (H1), α-helix type 2 (H2), 3_{10}-helix (3_{10}-H), β-sheet (S), anti-parallel β-sheet (S1), parallel β-sheet (S2), poly-proline 2 conformation (P_{II}), β-turn (T) and unordered (U). The reference sets are numbered as given in the preceding text. Data is omitted for fits where the NMRSD value fell outside the 0.25 limit. The 3_{10}-helix structure is not defined for reference set 5.

The prediction of α-helix content had accuracy to greater than 5 % error for all proteins over all five programs and often this accuracy was better than 3 % (Table V.17). The error in β-sheet estimation for both of the highly α-helical proteins myoglobin and hemoglobin was comparable with that of the α-helix with all programs. However, for both lysozyme and α-chymotrypsin the error in β-sheet content estimation was considerably higher than it was for the α-helix contents of these proteins. In general, the β-sheet estimation error was found to vary considerably depending on the program used in the analysis. The error in β-turn estimation was relatively consistent across each of the various programs (except in the case of myoglobin) and found to be approximately between 5-15 % accurate. The accuracy of P_{II} conformation was found to vary from one protein to another. The estimation of myoglobin P_{II} content was to within 2 %, whereas for lysozyme it was around 20 %. For hemoglobin and α-chymotrypsin this accuracy was between 5-7 %. The accuracy for the estimation of random coil content was almost as good as that for α-helix and generally it was within 5 % error for all proteins.

Protein	**Ref. set**	**H1**	**H2**	**S1**	**S2**	**T**	**U**	**Total**	**NMRSD**
Myoglobin	3	0.56	0.28	0.03	0.01	0.04	0.08	1	0.007
	4	0.55	0.22	0.02	0.01	0.06	0.13	0.99	0.007
	6	0.57	0.29	0.02	0.01	0.05	0.06	1	0.006
	7	0.56	0.28	0.02	0.01	0.06	0.08	1.01	0.005
Hemoglobin	3	0.55	0.27	0.02	0.00	0.07	0.09	1	0.009
	4	0.53	0.25	0.02	0.01	0.08	0.11	1	0.009
	6	0.55	0.26	0.01	0.00	0.08	0.09	0.99	0.009
	7	0.53	0.30	0.01	0.01	0.10	0.05	1	0.010
Lysozyme	3	0.25	0.18	0.07	0.06	0.20	0.25	1.01	0.011
	4	0.24	0.16	0.08	0.06	0.19	0.26	0.99	0.012
	6	0.25	0.17	0.07	0.06	0.18	0.27	1	0.011
	7	0.25	0.18	0.07	0.05	0.18	0.26	0.99	0.010
α-chymotrypsin	3	0.04	0.05	0.23	0.11	0.22	0.35	1	0.026
	4	0.03	0.07	0.19	0.11	0.24	0.34	0.98	0.027
	6	0.03	0.04	0.13	0.07	0.14	0.58	0.99	0.035
	7	0.03	0.04	0.17	0.09	0.18	0.47	0.98	0.028

Table V.13. Analysis of the secondary structure fractional composition of standard proteins using the CONTIN/LL program. The different secondary structures are as listed for Table V.12.

Protein	Ref. set	H1	H2	S1	S2	T	U	Total	NMRSD
Myoglobin	1	0.531	0.257	0.000	0.000	0.067	0.145	1	0.035
	3	0.515	0.248	0.000	0.001	0.078	0.158	1	0.017
	4	0.522	0.256	0.000	0.000	0.065	0.157	1	0.017
	6	0.522	0.260	0.000	0.006	0.106	0.106	1	0.017
	7	0.524	0.270	0.000	0.007	0.100	0.100	1.001	0.017
Hemoglobin	1	0.569	0.278	0.000	0.000	0.053	0.100	1	0.042
	3	0.499	0.253	0.000	0.000	0.114	0.134	1	0.021
	4	0.478	0.263	0.000	0.008	0.100	0.152	1.001	0.021
	6	0.495	0.245	0.000	0.000	0.096	0.164	1	0.021
	7	0.478	0.270	0.000	0.012	0.117	0.123	1	0.021
Lysozyme	1	0.246	0.163	0.081	0.054	0.199	0.258	1.001	0.068
	3	0.236	0.162	0.088	0.060	0.202	0.252	1	0.022
	4	0.234	0.173	0.088	0.058	0.187	0.260	1	0.026
	6	0.232	0.144	0.065	0.046	0.160	0.353	1	0.022
	7	0.240	0.153	0.055	0.038	0.132	0.383	1.001	0.026
α-chymotrypsin	1	0.077	0.105	0.175	0.085	0.217	0.341	1	0.048
	3	0.067	0.098	0.181	0.094	0.217	0.344	1.001	0.052
	4	0.051	0.095	0.179	0.102	0.216	0.358	1.001	0.043
	6	0.058	0.067	0.142	0.072	0.149	0.513	1.001	0.045
	7	0.061	0.069	0.136	0.075	0.144	0.515	1	0.046
		αH	**3_{10}H**	**S**	**P_{II}**	**T**	**U**	**Total**	**NMRSD**
Myoglobin	2	0.681	0.102	0.000	0.010	0.065	0.141	0.999	0.026
	5	0.757	-	0.001	0.047	0.090	0.099	0.999	0.026
Hemoglobin	2	0.691	0.095	0.000	0.016	0.065	0.133	1	0.037
	5	0.748	-	0.000	0.051	0.103	0.098	1	0.044
Lysozyme	2	0.296	0.091	0.093	0.069	0.131	0.319	0.999	0.095
	5	0.363	-	0.173	0.099	0.229	0.137	1.001	0.050
α-chymotrypsin	2	0.108	0.043	0.168	0.115	0.136	0.429	0.999	0.057
	5	0.166	-	0.222	0.186	0.229	0.195	0.998	0.073

Table V.14. Analysis of the secondary structure fractional composition of standard proteins using the CDSSTR program. The different secondary structures are as listed for Table V.12.

Protein	αH	S1	S2	T	Other	Total
Myoglobin	0.76	-0.03	0.01	0.11	0.14	1.01
Hemoglobin	0.76	-0.06	0.00	0.09	0.12	0.96
Lysozyme	-	-	-	-	-	-
α-chymotrypsin	0.11	0.30	0.01	0.20	0.40	1.02

Table V.15. Analysis of the secondary structure fractional composition of standard proteins using the VARSLC program. The different secondary structures are as listed for Table V.12. The VARSLC program failed to adequately analyse the lysozyme CD spectrum.

Protein	αH	S	U	NMRSD
Myoglobin	0.81	0.00	0.19	0.057
Hemoglobin	0.79	0.00	0.21	0.056
Lysozyme	0.37	0.16	0.47	0.106
α-chymotrypsin	0.09	0.35	0.56	0.115

Table V.16. Analysis of the secondary structure fractional composition of standard proteins using the K2d program. The different secondary structures are as listed for Table V.12.

Protein	**Program**	**α-helix error**	**β-sheet error**	**β-turn error**	**P_{II} error**	**Other error**
Myoglobin	SELCON	-0.049	0.013	0.055	-0.019	0.005
	CONTIN/LL	-0.041	0.002	0.082	0.005	-0.015
	CDSSTR	0.024	0.033	0.053	-	-0.057
	VARSLC	-0.044	-0.02	0.11	-	-0.004
	K2d	0.006	0.00	-	-	-0.006
Hemoglobin	SELCON	-0.003	0.002	0.082	-0.070	0.003
	CONTIN/LL	-0.008	0.012	0.093	-0.072	-0.007
	CDSSTR	0.050	0.020	0.083	-	-0.051
	VARSLC	0.00	-0.06	0.09	-	-0.016
	K2d	0.03	0.00	-	-	-0.031
Lysozyme	SELCON	-0.050	0.104	0.141	-0.162	0.043
	CONTIN/LL	-0.042	0.112	0.130	-0.214	0.059
	CDSSTR	0.001	0.114	0.140	-	0.039
	VARSLC	-	-	-	-	-
	K2d	-0.049	0.144	-	-	0.049
α-chymotrypsin	SELCON	-0.004	0.048	0.081	-0.060	-0.017
	CONTIN/LL	0.032	0.025	0.081	-0.050	0.014
	CDSSTR	-0.032	0.067	0.089	-	0.064
	VARSLC	-0.004	0.102	0.094	-	0.029
	K2d	-0.024	0.142	-	-	-0.011

Table V.17. The prediction error of the fractional composition of the standard proteins from their CD spectra. For SELCON, CONTIN/LL and CDSSTR the errors were calculated based on the average structure predicted over all protein reference sets. The errors are given as + or – based on whether the programs over or under predicted the amount of each secondary structure, respectively.

V.4.3 Discussion

The accuracy of the quantitative secondary structural analysis for each of the proteins studied serves to validate the experimental set-up and procedure as capable of reliably analysing novel proteins in terms of their secondary structure fractional compositions. It should be noted that the superior accuracy of estimates of secondary structure content in the above experiment, relative to the correlation coefficients for each program, most likely arises from the fact that the reference sets used in the analyses contain each of the proteins studied above. CD spectroscopy has been seen to be an excellent technique for analysing protein secondary structure, particularly in terms of its α-helix and random-coil content. It is a little less suited for analysing both β-sheet and β-turn content and this arises from the weakness in intensity of these bands in the CD spectrum and the lack of uniformity of these structures within proteins, giving rise to non-uniform CD bands. During the fitting procedure the test spectra were scaled in order to return the lowest NMRSD value. The accuracy of the estimates at this lowest NMRSD qualifies this parameter as a good gauge of the success in estimating a protein’s secondary structure composition and this is consistent with the findings of

Miles *et al.* in their investigation into the correlation of the NMRSD value and the accuracy of the estimated protein secondary structure content.[30]

References

1 Brahms S, Brahms J. Determination of Protein Secondary Structure in Solution by Vacuum Ultraviolet Circular-Dichroism. Journal of Molecular Biology 138(2):149-178, 1980.

2 Sreerama N, Woody RW. On the analysis of membrane protein circular dichroism spectra. Protein Science 13(1):100-112, 2004.

3 Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Analytical Biochemistry 287(2):252-260, 2000.

4 Reed J, Reed TA. A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism. Analytical Biochemistry 254(1):36-40, 1997.

5 Oberg KA, Ruysschaert JM, Goormaghtigh E. Rationally selected basis proteins: A new approach to selecting proteins for spectroscopic secondary structure analysis. Protein Science 12(9):2015-2031, 2003.

6 Greenfie.N, Fasman GD. Computed Circular Dichroism Spectra for Evaluation of Protein Conformation. Biochemistry 8(10):4108-&, 1969.

7 Travis JC, Acosta JC, Andor G, Bastie J, Blattner P, Chunnilall CJ, Crosson SC, Duewer DL, Early EA, Hengstberger F and others. Intrinsic wavelength standard absorption bands in holmium oxide solution for UV/visible molecular absorption spectrophotometry. Journal of Physical and Chemical Reference Data 34(1):41-56, 2005.

8 Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols 1(6):2876-2890, 2006.

9 Yang JT, Wu CSC, Martinez HM. Calculation of Protein Conformation from Circular-Dichroism. Methods in Enzymology 130:208-269, 1986.

10 Wetlaufer DB, Saxena VP, Ahmed AK, Schaffer SW, Pick PW, Oh KJ. Protein Thiol-Disulfide Interchange and Interfacing with Biological Redox Systems. Abstracts of Papers of the American Chemical Society 172(SEP3):78-78, 1976.

11 Chen YH, Yang JT. New Approach to Calculation of Secondary Structures of Globular Proteins by Optical Rotatory Dispersion and Circular Dichroism. Biochemical and Biophysical Research Communications 44(6):1285-&,1971.

12 Chang CT, Wu CSC, Yang JT. Circular Dichroic Analysis of Protein Conformation - Inclusion of Beta-Turns. Analytical Biochemistry 91(1):13-31, 1978.

13 Bolotina IA, Lugauskas VY. Determination of the Secondary Structure of Proteins from the Circular-Dichroism Spectra .4. Consideration of the Contribution of Aromatic Amino-Acid-Residues to the Circular-Dichroism Spectra of Proteins in the Peptide Region. Molecular Biology 19(5):1154-1166, 1985.

14 Hennessey JP, Johnson WC. Information-Content in the Circular-Dichroism of Proteins. Biochemistry 20(5):1085-1094, 1981.

15 Perczel A, Hollosi M, Tusnady G, Fasman GD. Convex Constraint Analysis - a Natural Deconvolution of Circular-Dichroism Curves of Proteins. Protein Engineering 4(6):669-679, 1991.

16 Perczel A, Park K, Fasman GD. Deconvolution of the Circular-Dichroism Spectra of Proteins - the Circular-Dichroism Spectra of the Antiparallel Beta-Sheet in Proteins. Proteins-Structure Function and Genetics 13(1):57-69, 1992.

17 Perczel A, Park K, Fasman GD. Analysis of the Circular-Dichroism Spectrum of Proteins Using the Convex Constraint Algorithm - a Practical Guide. Analytical Biochemistry 203(1):83-93, 1992.

18 Venyaminov SY, Baikalov IA, Wu CSC, Yang JT. Some Problems of Cd Analyses of Protein Conformation. Analytical Biochemistry 198(2):250-255, 1991.

19 Provencher SW, Glockner J. Estimation of Globular Protein Secondary Structure from Circular-Dichroism. Biochemistry 20(1):33-37, 1981.

20 Manavalan P, Johnson WC. Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular-Dichroism Spectra. Analytical Biochemistry 167(1):76-85, 1987.

21 Sreerama N, Woody RW. A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular-Dichroism. Analytical Biochemistry 209(1):32-44, 1993.

22 Sreerama N, Woody RW. Protein Secondary Structure from Circular-Dichroism Spectroscopy - Combining Variable Selection Principle and Cluster-Analysis with Neural-Network, Ridge-Regression and Self-Consistent Methods. Journal of Molecular Biology 242(4):497-507, 1994.

23 Sreerama N, Woody RW. Poly(Pro)Ii Helices in Globular-Proteins - Identification and Circular Dichroic Analysis. Biochemistry 33(33):10022-10025, 1994.

24 Andrade MA, Chacon P, Merelo JJ, Moran F. Evaluation of Secondary Structure of Proteins from Uv Circular-Dichroism Spectra Using an Unsupervised Learning Neural-Network. Protein Engineering 6(4):383-390, 1993.

25 Merelo JJ, Andrade MA, Prieto A, Moran F. Proteinotopic Feature Maps. Neurocomputing 6(4):443-454, 1994.

26 Bohm G, Muhr R, Jaenicke R. Quantitative-Analysis of Protein Far Uv Circular-Dichroism Spectra by Neural Networks. Protein Engineering 5(3):191-195, 1992.

27 Kabsch W, Sander C. Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers 22(12):2577-2637, 1983.

28 Wallace BA, Lees JG, Orry AJW, Lobley A, Janes RW. Analyses of circular dichroism spectra of membrane proteins. Protein Science 12(4):875-884, 2003.

29 Miles AJ, Wien F, Lees JG, Rodger A, Janes RW, Wallace BA. Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers. Spectroscopy-an International Journal 17(4):653-661, 2003.

30 Miles AJ, Whitmore L, Wallace BA. Spectral magnitude effects on the analyses of secondary structure from circular dichroism spectroscopic data. Protein Science 14(2):368-374, 2005.

Chapter VI

Secondary Structure Analysis of Standard Proteins by FTIR Spectroscopy

VI.1 Introduction

The performance of protein FTIR spectroscopy is by no means straightforward and much practise and experimental adjustment is needed before a reliable procedure is obtained which is capable of routinely analysing the precious and expensive protein samples prepared by biological researchers. In satisfying this requirement it is essential to optimise all experimental procedures using standard proteins of known secondary structure, since such proteins have previously been investigated using FTIR spectroscopy by other researchers and so they provide a standard by which one can develop and validate ones own experimental setup and technique. This approach also makes economic sense, since standard proteins are commercially available and can be relatively inexpensive. The following work was performed for these reasons and, also, to take account of the various considerations, advantages and disadvantages involved with using alternative applications of FTIR spectroscopy in investigating protein secondary structure. The main disadvantage of the FTIR method is the large amounts of protein required.

The four well-ordered proteins selected for FTIR analysis are those highlighted in the CD study, namely myoglobin, hemoglobin, lysozyme and α-chymotrypsin. These proteins were selected so as to provide a broad coverage of secondary structure protein types to be analysed. Myoglobin and hemoglobin both contain very high levels of α-helix conformation and are absent of β-sheet content. Myoglobin has an important difference from hemoglobin, however, in that myoglobin is monomeric while hemoglobin is a tetramer consisting of four subunits. Lysozyme is also a predominantly α-helical protein that has considerably less α-helix content than either hemoglobin or myoglobin, but it does contain a small amount of β-sheet structure. In contrast, α-chymotrypsin is a predominantly β-sheet protein and has only a minor amount of α-helix content.

VI.2 Method

Samples of myoglobin, hemoglobin, lysozyme and α-chymotrypsin (Sigma) were made up to concentrations of approximately 15 mg/ml by weighing out 1.5 mg of each protein and dissolving each in 100 μl of 10 mM PBS made up with double processed tissue culture water (Sigma). These samples were used directly for ATR-FTIR analysis, while samples to be analysed by transmission FTIR were made by taking 50 μl of each ATR sample and diluting it with 150 μl of 10 mM PBS. Stock solutions of 5 mg/ml concentration were made up for each protein using a 10 mM PBS buffer prepared in D_2O solvent. Prior to spectral analysis all samples were centrifuged at 9000 RPM for 2 minutes, were thoroughly degassed and were kept at 4 ^{o}C.

Transmittance FTIR spectra were acquired using a temperature controlled AquaSpec™ accessory (Bruker) on a Bruker TENSOR 27 FTIR spectrometer. Spectral parameters were set as follows: a temperature of 23 ^{o}C, an aperture of 1.5 mm, a resolution of 4 cm^{-1} and as an average of 100 scans. Spectra were acquired in absorbance mode and, as such, were background corrected by taking the transmission spectrum of the solution in which the samples were made up. Before the acquisition of transmittance spectra the sealed cell was repeatedly cleaned with Hellmanex ® II detergent (Hellma UK) alternated with continuous rinsing using tissue culture water. To help prevent the occurrence of gas bubbles forming within the cell it was filled at the optimal filling rate of 5-10 μl per second using a 50 μl glass syringe (Hamilton). As a check for the presence of any interfering gas bubbles within the filled cell prior to spectral acquisition a transmission spectrum of the cell over the wavenumber range 3200-3600 cm^{-1} was taken. Spectral acquisition was only performed if over this range the sample was fully absorbing - a condition which can only occur in the absence of any trapped air bubbles in the cell. The AquaSpec accessory was continually purged with dry air from a gas purifier both before and during spectral acquisition.

ATR-FTIR spectra were acquired using a temperature controlled BioATR II™ accessory (Bruker) on a TENSOR 27 FTIR spectrometer with the following spectral parameters: a temperature of 23 ^{o}C, an aperture of 3 mm, a resolution of 4 cm^{-1} and as an average of 100 scans. The ATR-FTIR spectra were acquired in absorbance mode by background correcting each spectrum using the transmission spectrum of the solution used to make the sample. Between each measurement the silicon ATR wafer was cleaned 3 times with a 3 M guanidine/HCl solution, alternated with thorough rinsing with deionised H_2O. Prior to spectral acquisition the silicon wafer was checked for cleanness by recording the transmission spectrum over the wavenumber range 2700-3000 cm^{-1} of the empty sample compartment to ensure that it was free from any CH-stretching vibrations. The BioATR II accessory was continually purged with dry air both before and during spectral acquisition.

For both ATR- and transmittance FTIR absorption spectra the appropriateness of the background transmission spectrum to that of the sample was ensured by applying the condition that the absorption spectrum in the wavenumber range 1715-1750 cm^{-1} correspond approximately to a flat line at zero absorbance. Spectral post-processing consisted of zeroing all absorption spectra between 1750 and 1800 cm^{-1}. To remove any interfering water vapour lines an atmospheric correction procedure was applied to each protein absorption spectrum. The line narrowing procedure of taking the spectral 2^{ND} derivative was applied to all spectra in the amide-I region between 1600 and 1700 cm^{-1}. In some cases the alternative line narrowing procedure of taking the Fourier self-deconvolution was also performed using a bandwidth of 52 nm and a noise reduction factor of 0.25. When comparing the transmittance and ATR-FTIR spectra of each protein as spectral overlays, both spectra were normalised such that the area beneath the amide-I absorption band between 1600 and 1700 cm^{-1} was set to 1.

Standard protein temperature melt experiments were performed using the Bio-ATR II accessory. Temperature melts were performed by incrementally ramping the temperature from low to high temperatures while taking scans at pre-specified temperatures. A corresponding temperature ramp of the sample reference solution was performed in order to take account of the temperature effect on the transmission spectrum of water.

The myoglobin melt experiment was measured at a protein concentration of 20 mg/ml dissolved in tissue culture water. The start temperature was 20 ^{o}C and the final temperature was 90 ^{o}C. The temperature was stepped at 5 ^{o}C with an equilibration time of 2 minutes before beginning spectral acquisition.

The hemoglobin melt experiment was measured at a protein concentration of 10 mg/ml dissolved in tissue culture water. The start temperature was 10 ^{o}C and the final temperature was 90 ^{o}C. Two temperature ramps were used to cover this range with the first ramping at 2 ^{o}C increments between 10 and 60 ^{o}C and the second ramping at 5 ^{o}C between 60 and 90 ^{o}C. A temperature equilibration time of 2 minutes was used across both ramps before the beginning of spectral acquisition.

The α-chymotrypsin melt experiment was performed at a protein concentration of 15 mg/ml. The start temperature was 20 ^{o}C and the end temperature was 90 ^{o}C. Three temperature ramps were used within this region: from 20 to 35 ^{o}C in increments of 5 ^{o}C and with a temperature equilibration time of 2 minutes: from 35 to 65 ^{o}C in increments of 1 ^{o}C and with an equilibration time of 1 minute: and from 65 to 90 ^{o}C in increments of 5 ^{o}C and with an equilibration time of 2 minutes.

The lysozyme melt experiment was performed at a protein concentration of 10 mg/ml. The temperature was varied from 18 to 51 ^{o}C and stepped at 3 ^{o}C intervals. A temperature equilibration time of 2 minutes was used once the sample had reached the scan temperature.

Spectral post-processing for each of the standard protein melt experiments consisted of atmospheric correction for water vapour and CO_2 bands, along with offset correction at a wavenumber between 1700 and 1750 cm^{-1}, area normalisation of the amide-I band and subsequent 2ND derivative analysis within the amide-I region of each absorption spectrum.

The $T_{1/2}$ value of each protein was measured by constructing a model designed to quantify the relative degree of aggregation at each temperature. This was done by taking the 2ND derivative spectrum of the aggregation band between 1615 and 1630 cm^{-1} of the spectrum at the start and end temperatures of the melt experiment and giving these a numeric value of 0 and 100 % aggregation, respectively. Each individual model created for each protein was then used to analyse the aggregation content of the protein at temperatures intermediate to these extremes in terms of the percent aggregation at a given temperature. The ability of each model to analyse the test spectra is reflected in a quantity known as the Mahanalobis distance.[1] A Mahanalobis limiting distance of 0.5 defines whether or not the model is capable of evaluating each of the test spectra. Only spectra having Mahanalobis distances within the 0.5 limiting value are presented below for analysis.

VI.3 Results

VI.3.1 Transmittance FTIR of Standard Proteins

Table VI.1 gives the secondary structure fractional composition of four standard proteins, as determined by x-ray crystallography, and Figure VI.1 gives cartoon illustrations of their secondary structures in terms of α-helix, β-sheet and random coil contents. The secondary structures of these proteins have been interpreted according to α-helix, β-sheet, β-turn, P_{II}-helix and other, with the 'other' class being approximately equivalent to that of randomly coiled structure. The β-sheet class represents both parallel and anti-parallel β-sheet structures. From the x-ray data both hemoglobin and myoglobin are understood to be highly α-helical and contain no β-sheet or β-turn regions. Lysozyme is also a predominantly α-helical protein, but it has considerably less α-helix content than either hemoglobin or myoglobin, having approximately half the α-helix content of myoglobin. It also has a minor amount of β-sheet and β-turn structure. In terms of its secondary structure, lysozyme's most distinct difference from hemoglobin and myoglobin is that it has a considerable P_{II}-helix contribution to its secondary structure. α-chymotrypsin, on the other hand, is a predominantly β-sheet protein, although it does have some amount of α-helix. It is, therefore, referred to as a mixed αβ-protein, while hemoglobin, myoglobin and lysozyme are known as α-proteins.

Protein	**α-helix**	**β-sheet**	**β-turn**	**P_{II}-helix**	**other**
Myoglobin	80.4	0.0	0.0	5.2	14.4
Hemoglobin	76.0	0.0	0.0	10.5	13.6
Lysozyme	41.9	1.6	4.7	29.8	22.1
α-chymotrypsin	11.4	20.8	10.6	20.0	37.1

Table VI.1. The x-ray crystal structures of standard proteins. The relative amounts of each characteristic secondary structure within each protein are given as a percentage.

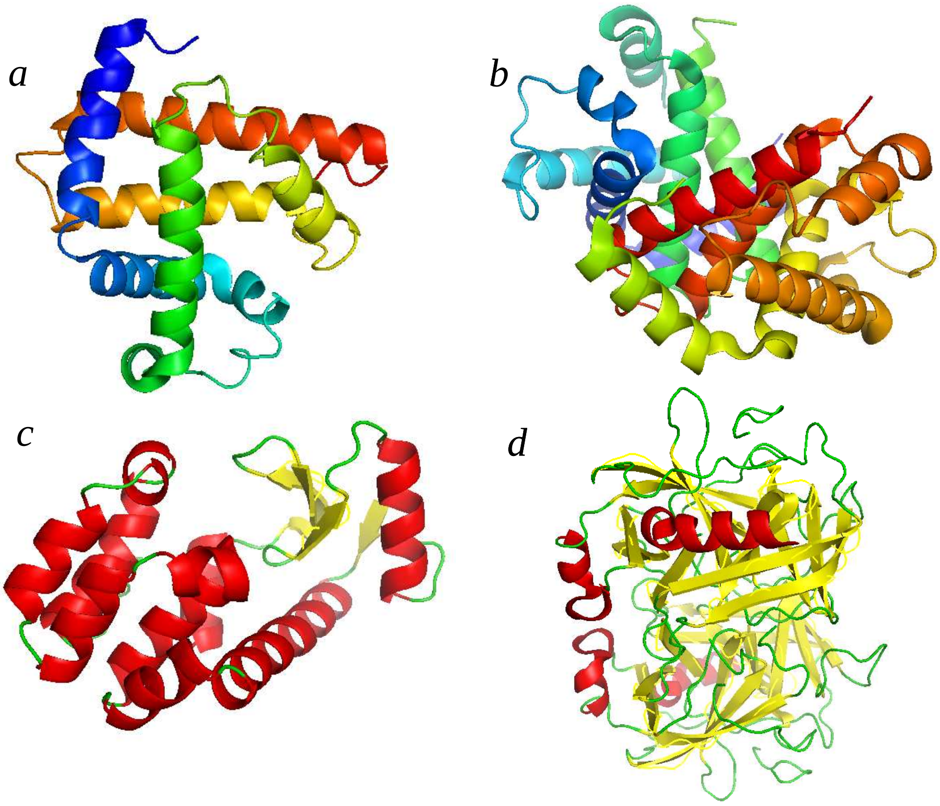


Figure VI.1. Crystal structures of standard proteins. The crystal structures of myoglobin, hemoglobin, lysozyme and α-chymotrypsin are represented as *a*, *b*, *c* and *d*, respectively. α-helix protein regions are shown as ribbons and β-sheet regions are shown as yellow arrows, with the orientation of each β-sheet given by the direction of the arrow. (Generated using the PyMol Molecular Viewer program (http://pymol.org/))

Figure VI.2 shows the absorption spectrum of myoglobin, as acquired using transmission FTIR. The amide-I band maximum occurs at 1654 cm^{-1}, which corresponds to the amide-I of the α-helix conformation. The distinctly lorenztian bandshape - especially to the higher wavenumber side of the band maximum – also agrees well with the characteristic α-helix amide-I. The approximate flat line shape of the spectrum past 1700 cm^{-1} indicates the accuracy with which the water band has been referenced out of the absorption spectrum. The Fourier deconvolution of the myoglobin amide-I band (Figure VI.2*b*) has the effect of resolving the composite protein amide-I band into its constituent characteristic secondary structure components. The Fourier-deconvoluted amide-I band has its band maximum at 1656 cm^{-1}, as expected. The minor random coil content of myoglobin predicted from its x-ray crystal structure is observable as the band shoulder to the lower wavenumber side of this band maximum and centred at 1645 cm^{-1}, at the expected location of the random coil amide-I band maximum. The shoulder occurring with a band centre of

1638 cm^{-1}, and also the small band at 1682 cm^{-1}, is evidence for the presence of some β-sheet content to the myoglobin sample. The band centred at 1627 cm^{-1} reveals that a small level of aggregation was present within the sample, with the minor aggregation band also visible at 1695 cm^{-1}. A small band at 1672 cm^{-1} suggests a minor presence of β-turn structure. The 2ND derivative of the transmission myoglobin amide-I band reveals the same spectral features as that observed using Fourier deconvolution (Figure VI.2*c*). The main difference between the two, however, is that 2ND derivative analysis does not clearly resolve the α-helix and random coil bands, although the presence of both spectral features is clear from the shape of the main band in the 2ND derivative spectrum centred at 1655 cm^{-1}. A residual aggregate band is again represented by a band centred at 1627 cm^{-1}, along with the expected minor band at 1692 cm^{-1}. The presence of a small amount of β-sheet in the myoglobin sample is represented as a small shoulder to the aggregation band located at 1638 cm^{-1}. The β-turn band at 1672 cm^{-1} is also evident within the 2ND derivative spectrum.

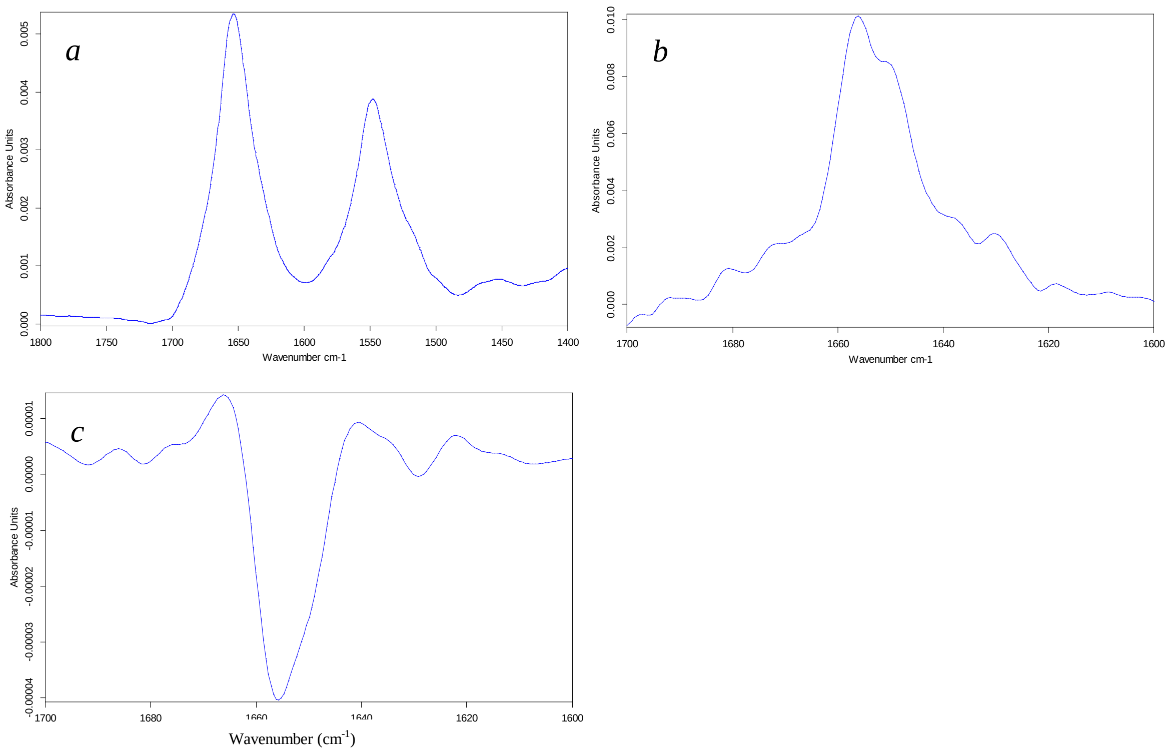


Figure VI.2. Myoglobin secondary structure analysis from transmission FTIR spectroscopy. *a* The transmission FTIR spectrum of myoglobin. *b* The Fourier-deconvoluted myoglobin transmission FTIR spectrum. *c* The 2ND derivative spectrum of the unconvoluted transmission FTIR spectrum.

Figure VI.3*a* shows the amide-I and –II bands of the hemoglobin FTIR absorption spectrum acquired using transmission FTIR. The band maximum of the amide-I band is at 1656 cm^{-1} and is consistent with that of a predominantly α-helical protein. The bandwidth of the hemoglobin amide-I band is somewhat broader than that of myoglobin, reflecting the smaller α-helical content of the former. A flat baseline beyond 1700 cm^{-1} confirms accurate subtraction of the background water band. Figure VI.3*b* shows the hemoglobin Fourier-deconvoluted amide-I band. The amide-I band can be seen to be resolved into a number of component bands, most prominent amongst them being the α-helix band centred at 1656 cm^{-1}. A shoulder band centred at 1642 cm^{-1} is attributable to the random coil secondary structure. The extent of protein aggregation within the sample is indicated by the band centred at 1625 cm^{-1} and the band shoulder at approximately 1692 cm^{-1}. The occurrence of bands at around 1666 and 1678 cm^{-1} indicates a certain amount of β-turn present within the structure and bands occurring at 1618 and 1612 cm^{-1} are likely the result of side-chain absorption overlapping the amide-I band. All of the above bands identified by the Fourier-deconvolution procedure were also revealed by using 2$^{\mathrm{ND}}$ derivative analysis of the amide-I band (Figure VI.3*c*). The relative intensities of the deconvoluted bands, however, were found to be different when using 2$^{\mathrm{ND}}$ derivative analysis versus that of Fourier deconvolution, the most notable difference between the two being the higher relative intensity of the α-helix amide-I contribution to those of other secondary structures, for instance that of the random coil band centred at 1641 cm^{-1}, within the 2$^{\mathrm{ND}}$ derivative spectrum.

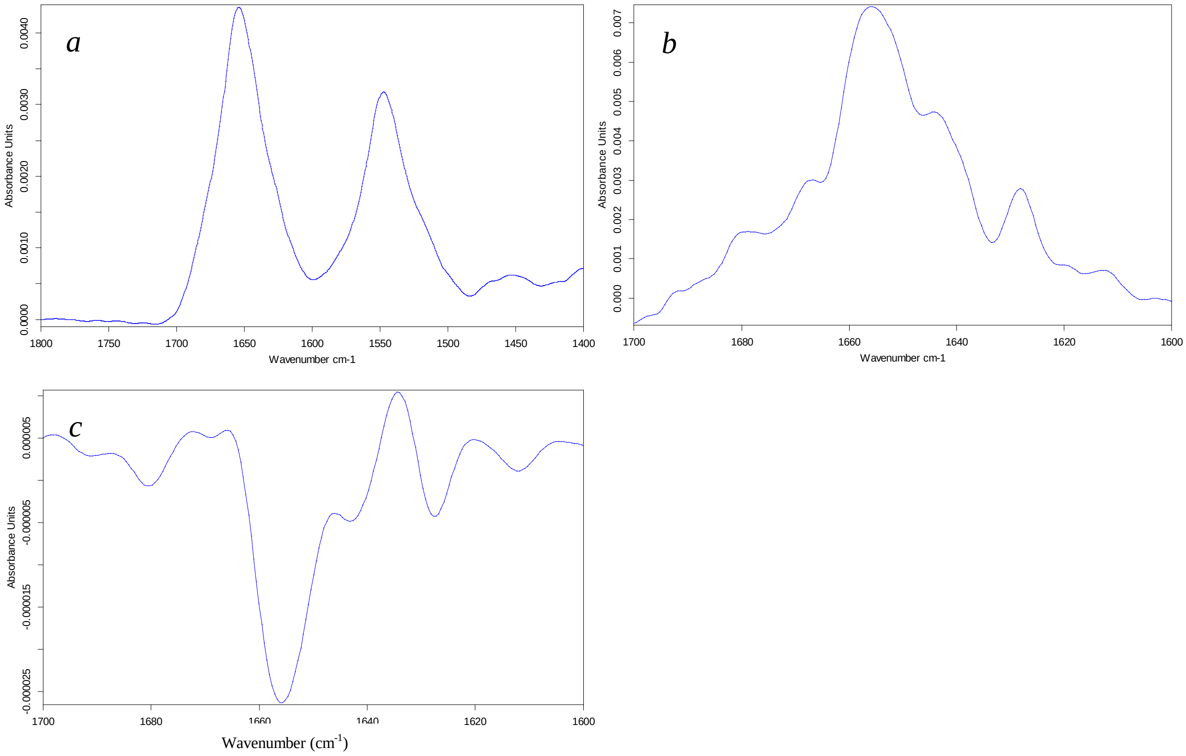


Figure VI.3. Hemoglobin secondary structure analysis from transmission FTIR spectroscopy. *a* The unconvoluted transmission FTIR spectrum of hemoglobin. *b* The Fourier-deconvoluted hemoglobin transmission FTIR spectrum. *c* The 2ND derivative spectrum of the unconvoluted transmission FTIR spectrum.

The FTIR absorption spectrum of lysozyme acquired using transmission FTIR in the amide-I and –II band region is shown in Figure VI.4*a*. The band maximum of the amide-I band was found at 1652 cm^{-1}, as expected for a predominantly α-helical protein. Figure VI.4*b* shows the lysozyme amide-I band Fourier-deconvolution. The prominent deconvoluted band centred at 1652 cm^{-1} corresponds to that of the α-helix amide-I band. To shorter wavenumbers there appears a broad shoulder, which is most likely a result of both random coil and β-sheet structure. A minor band shoulder at around 1627 cm^{-1} can be safely assigned to the residual presence of protein aggregates, while shoulders at 1669 and 1677 cm^{-1} are probably a result of 3_{10}-helix and β-turn structures, respectively. The shoulder occurring around 1688 cm^{-1} is predominantly from β-sheet structure, with some contribution form the residual aggregates present. The shoulder around 1616 cm^{-1} can be understood as amino-acid side-chain absorption overlapping the amide-I band. The same bands as mentioned above are identified using 2ND derivative analysis (Figure VI.4*c*) as compared to Fourier-deconvolution of the lysozyme amide-I band. However, the 2ND derivative spectrum achieves superior resolution of the α-helix band from its shoulder at lower wavenumbers, in contrast to the case of myoglobin (Figures VI.2*b* and *c*). The effect of this is to shift the α-helix band maximum to the expected position of 1656 cm^{-1}, while revealing the band maximum of the shoulder to be approximately 1642 cm^{-1}. This allows the confident assignment of this shoulder to the random coil structure. The β-turn bands occurring between 1670 and 1680 cm^{-1} are also present within the 2ND derivative spectrum.

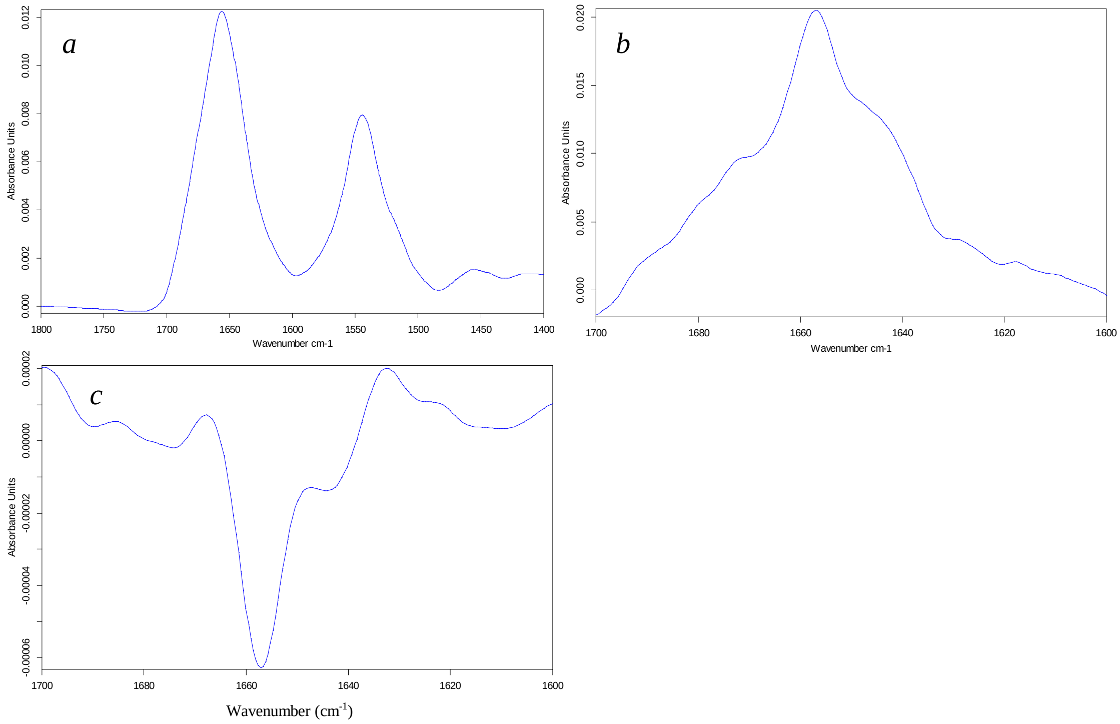


Figure VI.4. Lysozyme secondary structure analysis from transmission FTIR spectroscopy. *a* The unconvoluted transmission FTIR spectrum of lysozyme. *b* The Fourier-deconvoluted lysozyme transmission FTIR spectrum. *c* The 2ND derivative spectrum of the unconvoluted transmission FTIR spectrum.

Figure VI.5*a* shows the amide-I and –II bands of the α-chymotrypsin FTIR absorption spectrum acquired using transmission FTIR. The irregular shape of the amide-I band suggests an even mixture of conformations contributing to the secondary structure of α-chymotrypsin. Figure VI.5*b* shows the Fourier-deconvoluted α-chymotrypsin amide-I band, revealing significant contributions from a number of component bands summing to form the composite protein amide-I band. The most intense band component was found to be centred at 1632 cm^{-1} and can be assigned to the β-sheet conformation. The band intensity between 1680 and 1690 cm^{-1} is sufficient to allow for the presence of the required minor β-sheet band in order to make this assignment. Of almost equal intensity to the β-sheet band is a component centred at 1643 cm^{-1}. This can be safely assigned to random coil content within the α-chymotrypsin secondary structure. Such a significant level of random coil conformation gives an explanation for the relatively broad amide-I bandshape, since the random coil amide-I band is very much gaussian in shape. The occurrence of a component band at 1654 cm^{-1} reveals the presence of a small but significant level of α-helix present in α-chymotrypsin. The band centred at 1666 cm^{-1}, along with its shoulder at 1672 cm^{-1}, are clear indications of β-turn content to the α-chymotrypsin secondary structure. The absence of any component band between 1620 and 1625 cm^{-1} is evidence for their being no appreciable degree of aggregation within the protein sample. Side-chain absorption overlapping the amide-I band can be seen by the presence of bands at 1610 and 1618 cm^{-1}. The 2ND derivative of the α-chymotrypsin amide-I band reveals similar bands to those identified using Fourier-deconvolution (Figure VI.5*c*). The main difference in using 2ND derivative analysis over Fourier-deconvolution is the relative intensity of the β-sheet component with respect to other conformation component bands, particularly that of random coil. As such, the intensities of the major and minor β-sheet bands located at 1636 and 1683 cm^{-1}, respectively, dominate the α-chymotrypsin amide-I 2ND derivative spectrum. The random coil, α-helix and β-turn components are readily identifiable within the 2ND derivative spectrum at 1643, 1656 and 1672 cm^{-1}, respectively.

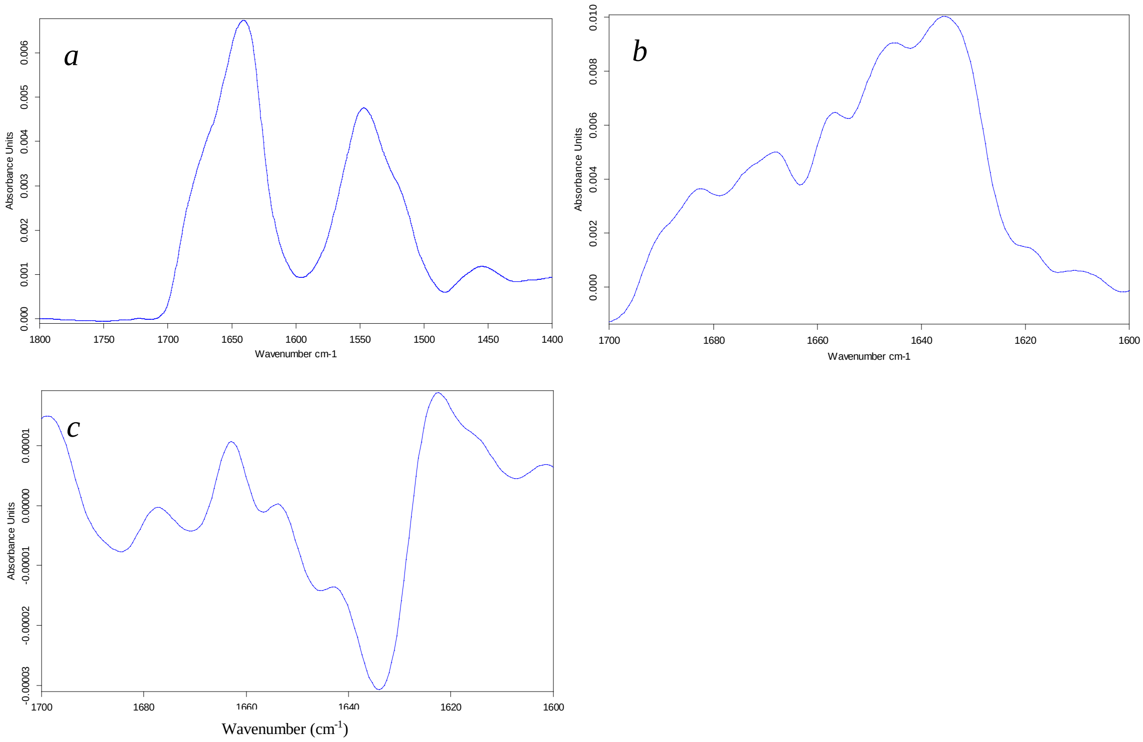


Figure VI.5. α-chymotrypsin secondary structure analysis from transmission FTIR spectroscopy. *a* The unconvoluted transmission FTIR spectrum of α chymotrypsin. *b* The Fourier-deconvoluted α chymotrypsin transmission FTIR spectrum. *c* The 2^{ND} derivative spectrum of the unconvoluted transmission FTIR spectrum.

VI.3.2 ATR-FTIR of Standard Proteins

Figure VI.6*a* shows the amide-I and -II bands of the myoglobin FTIR absorption spectrum acquired using ATR-FTIR spectroscopy. The accuracy with which the water background has been referenced out when generating the absorption spectrum is revealed by the closeness of the spectral shape to a flat line at zero absorption beyond 1700 cm^{-1}. A very small negative deviation to this ideal seen in Figure VI.6*a*, described as a negative water signal, is an undesirable effect and arises from their being a higher "water concentration" within the reference compared to that present in the sample solution. (This is a result of the necessity of using high protein concentrations when working with ATR-FTIR spectroscopy. The effect of having different water concentrations for the reference and sample solutions on the amide-I bandshape is considerable and so any difference should be kept to an absolute minimum.) The Fourier deconvolution of the myoglobin ATR amide-I band is shown in Figure VI.6*b*. The principal component band to the myoglobin amide-I band can be seen centred at 1652 cm^{-1}, corresponding to the α-helix conformation. To the lower wavenumber side of this band there occurs a shoulder at approximately 1641 cm^{-1}, which is attributable to the random coil conformation. A band located at 1625 cm^{-1} reveals the presence of aggregates within the protein sample. Side-chain absorption is evident from the bands located at 1618 and 1611 cm^{-1}. A band at 1678 cm^{-1} suggests the presence of β-turn content within the myoglobin secondary structure. The 2ND derivative of the myoglobin ATR amide-I band accentuates the α-helical component over that of the random coil, with the latter being only recognisable as a slight shoulder to the lower wavenumber side of the α-helix band centred at 1654 cm^{-1} (Figure VI.6*c*). β-turn content is revealed by the occurrence of a band at 1678 cm^{-1}, while the presence of aggregation within the protein sample can be seen from the small band at 1624 cm^{-1}. Side-chain absorption is again evident by bands at 1618 and 1610 cm^{-1}.

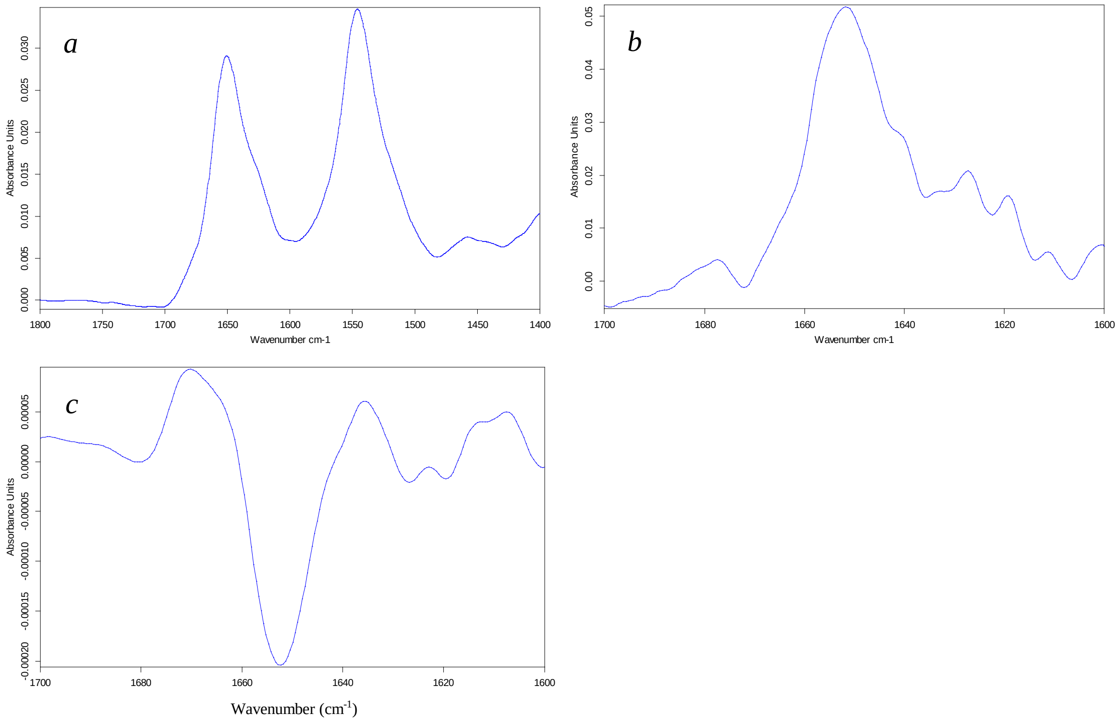


Figure VI.6. Myoglobin secondary structure analysis from ATR-FTIR spectroscopy. *a* The unconvoluted ATR-FTIR spectrum of myoglobin. *b* The Fourier-deconvoluted myoglobin ATR-FTIR spectrum. *c* The 2ND derivative spectrum of the unconvoluted ATR-FTIR spectrum.

The hemoglobin ATR-FTIR absorption spectrum showing the amide-I and –II bands is shown in Figure VI.7*a*. The very slight presence of a negative water band is apparent at wavenumbers beyond 1700 cm^{-1}. The lorenztian character to the amide-I band, along with its band maximum of 1654 cm^{-1}, suggests that the hemoglobin protein sample is predominantly α-helical in structure. Figure VI.7*b* shows the Fourier-deconvoluted amide-I band. The component band at 1654 cm^{-1} is as expected for a largely α-helical protein. The band at 1642 cm^{-1} of almost equal intensity is evidence for a considerable amount of random coil to the protein structure, while the higher wavenumber shoulder at 1672 cm^{-1} suggests β-turn content. Bands at 1634 and 1685 cm^{-1} are evidence that β-sheet structure is present within the protein secondary structure. Aggregation within the protein sample can be seen from the band at 1625 cm^{-1}. Side-chain absorption results in the bands appearing at 1610 and 1618 cm^{-1}. Figure VI.7*c* shows the hemoglobin ATR-FTIR amide-I 2^{ND} derivative spectrum. Compared to the Fourier deconvolution analysis, the relative intensity of the α-helix component at 1656 cm^{-1} of the amide-I band is significantly increased, with respect to the other structural components, particularly when compared to that of the random coil component at 1642 cm^{-1} (Figure VI.7*c*) . The aggregation band is again present at 1625 cm^{-1} as are the β-sheet and β-turn bands at 1636 and 1672 cm^{-1}, respectively. The side-chain absorption can be seen at 1610 cm^{-1}.

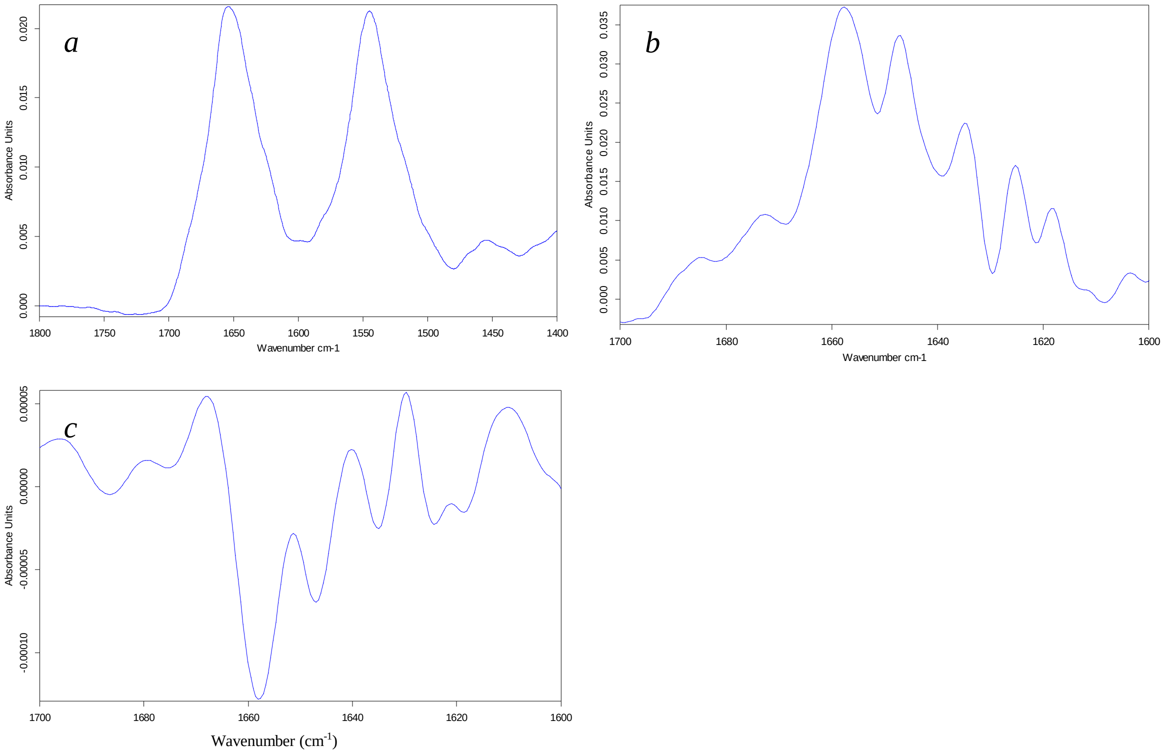


Figure VI.7. Hemoglobin secondary structure analysis from ATR-FTIR spectroscopy. *a* The unconvoluted ATR-FTIR spectrum of hemoglobin. *b* The Fourier-deconvoluted hemoglobin ATR-FTIR spectrum. *c* The 2^{ND} derivative spectrum of the unconvoluted ATR-FTIR spectrum.

The lysozyme ATR-FTIR absorption amide-I and –II bands are shown in Figure VI.8*a*. A very small positive water band can be seen from the spectrum at wavenumbers longer than 1700 cm^{-1}. The result of this was most likely a less than completely clean ATR surface for the acquisition of the reference transmission spectrum. The consequence of this would be to very marginally reduce the amount of water sampled by the IR beam in collecting the reference. The effect is, however, small enough to be of little significance to the amide-I bandshape. The amide-I band maximum of 1653 cm^{-1} suggests that there is a large amount of α-helix present within the lysozyme secondary structure. However, the reasonably broad amide-I bandshape indicates the presence of other component structures to the overall protein structure. Figure VI.8*b* shows the ATR-FTIR absorption Fourier-deconvoluted amide-I band, revealing that the main component band is, as expected, that of the α-helix at 1652 cm^{-1}, while the only slightly less intense band at 1643 cm^{-1} is that arising from random coil content. The pronounced band at 1632 cm^{-1} indicates a substantial level of β-sheet to the protein conformation and this is corroborated by the presence of an expected accompanying minor band at 1685 cm^{-1}. The band at 1666 cm^{-1} is evidence for the presence of β-turn content, while side-chain absorption can be clearly seen from bands located at 1618 and 1610 cm^{-1}. There does not appear to be significant levels of aggregates within the protein sample from the absence of a band of any real intensity around 1625 cm^{-1}. The lysozyme ATR-FTIR amide-I band 2ND derivative spectrum is shown as Figure VI.8*c*. This differs from that of the Fourier-deconvoluted spectrum in that it accentuates the intensity of the α-helix component relative to that of other conformations. All of the bands present within the Fourier-deconvoluted spectrum are, however, again clearly evident in the 2ND derivative spectrum.

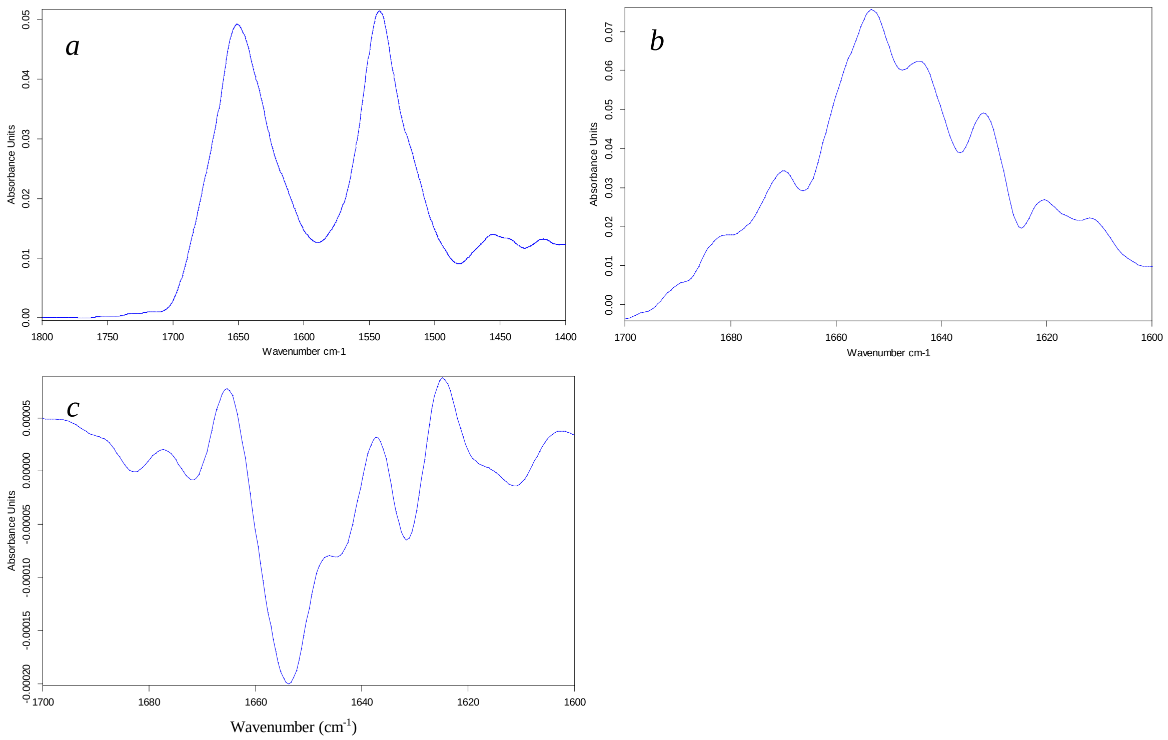


Figure VI.8. Lysozyme secondary structure analysis from ATR-FTIR spectroscopy. *a* The unconvoluted ATR-FTIR spectrum of lysozyme. *b* The Fourier-deconvoluted lysozyme ATR-FTIR spectrum. *c* The 2^{ND} derivative spectrum of the unconvoluted ATR-FTIR spectrum.

Figure VI.9*a* shows the amide-I and –II bands of the ATR-FTIR absorption spectrum of α-chymotrypsin. A small negative water band occurs to higher wavenumbers than 1700 cm^{-1}. The amide-I band maximum at around 1640 cm^{-1} suggests a large contribution of β-sheet to the protein secondary structure. Figure VI.9*b* shows the Fourier deconvolution of the α-chymotrypsin ATR-FTIR amide-I band. The band of highest intensity is that centred at 1632 cm^{-1} and corresponds to the major β-sheet amide-I component, with the minor β-sheet component present at 1682 cm^{-1}. The higher wavenumber shoulder to this band is also likely a result of β-sheet structure. A lower wavenumber shoulder to the highest intensity band, centred at 1625 cm^{-1} indicates a degree of aggregation within the protein sample and the minor aggregation band located at 1692 cm^{-1} corroborates this. The broad band between 1655 and 1665 cm^{-1} can be attributed to P_{II}-helix conformation. Bands at 1645 and 1672 cm^{-1} most likely result of random coil and β-turn content, respectively. A side-chain absorption band can be seen at 1612 and 1602 cm^{-1}. The 2^{ND} derivative ATR-FTIR amide-I spectrum is shown in Figure VI.9*c*. The intense β-sheet component band at 1636 cm^{-1} and its minor counterpart at 1682 cm^{-1} are again evident and a slight shoulder to the lower wavenumber side of this band indicates the presence of protein aggregates. The random coil, P_{II}-helix, and β-turn components are present at 1646, 1658 and 1672 cm^{-1}, respectively The side-chain absorption bands at 1612 and 1602 cm^{-1} can also be seen in the 2^{ND} derivative spectrum.

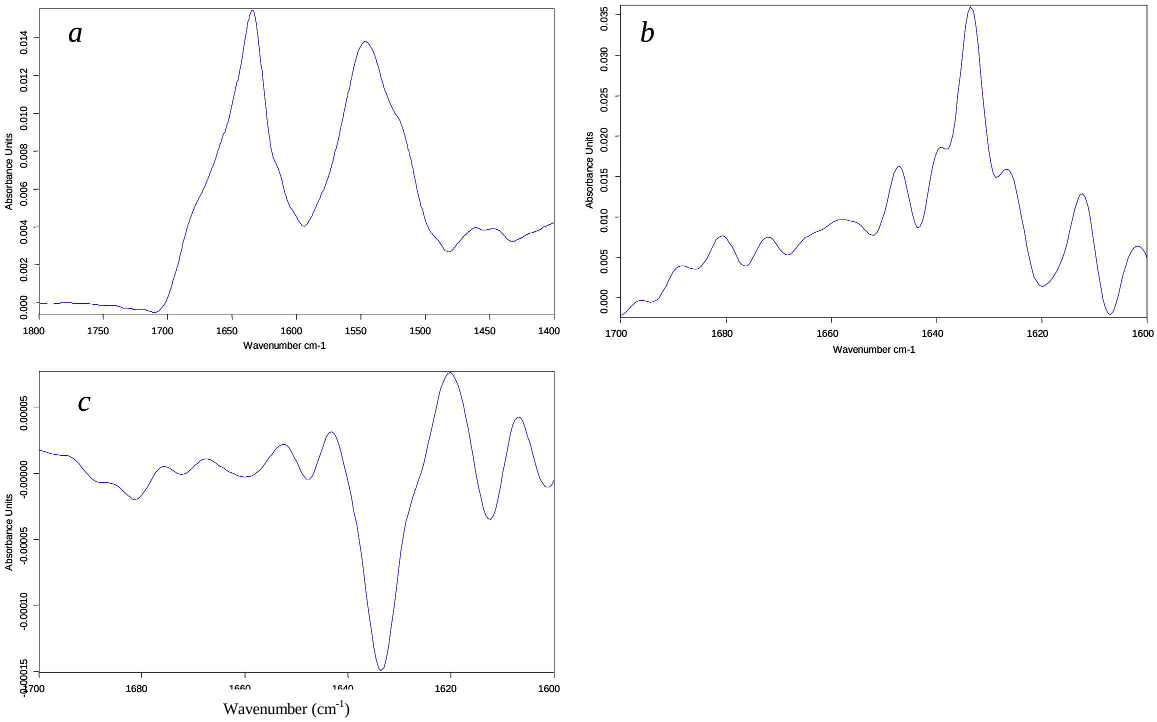


Figure VI.9. α chymotrypsin secondary structure analysis from ATR-FTIR spectroscopy. *a* The unconvoluted ATR-FTIR spectrum of α chymotrypsin. *b* The Fourier-deconvoluted α chymotrypsin ATR-FTIR spectrum. *c* The 2^{ND} derivative spectrum of the unconvoluted ATR-FTIR spectrum.

VI.3.3 Quantitative Analysis of Secondary Structure from Protein FTIR Spectra

The ability of the protein reference set to account for the structures of the test proteins, along with the suitability of the mathematical model used to relate the structures present within the reference set to those of the test spectra is given by the Mahanalobis distance parameter. The closer this value is to zero the better is the secondary structure fitting procedure and the greater is the degree of confidence to be had in the estimation. A failed result is returned when the Mahanalobis distance exceeds the acceptable limit of 0.12. The component value density is a measure of the number of calibration spectra within the protein reference set that have values close to that of the test spectrum. For example, a component value density of 2 for the prediction of a protein’s β-sheet content indicates that 2 proteins within the reference set have similar β-sheet content to that of the protein being analysed. A higher value of the component density value indicates a greater ability of the reference set to interpret the spectrum of the test protein.

Table VI.2 presents the results for the estimation of the α-helix and β-sheet content of the four standard proteins from their transmission FTIR spectra. By comparing the estimated structural contents of each protein with its x-ray crystal structure it can be seen that β-sheet estimation is accurate to within 10 %. The accuracy for the estimation of α-helix content is roughly the same as that of β-sheet except when attempting to estimate the helical content of highly α-helical proteins, such as myoglobin and hemoglobin. Transmission FTIR tends to substantially underestimate the α-helical fractional compositions of such proteins with high α-helix contents.

Protein	**Structure**	**x-ray**	**Prediction (%)**	**Error**	**Mahanalobis distance**	**Component value density**
α chymotrypsin	α helix	10	9.19	-0.81	0.019	0.88
	β sheet	34	33.944	-0.056	0.024	0.56
Hemoglobin	α helix	75	54.01	-20.99	0.057	0.12
	β sheet	0	9.66	9.66	0.02	0.59
Lysozyme	α helix	36	43.36	7.36	0.022	1.63
	β sheet	9	2.45	-6.55	0.052	0.71
Myoglobin	α helix	80.4	62.99	-17.41	0.099	0.15
	β sheet	0	5.78	5.78	0.035	0.33

Table VI.2. The prediction of protein secondary structure by transmission FTIR spectroscopy. A Mahanalobis distance limit of 0.12 defines whether predictions are to be taken as valid.

VI.3.4 Comparison of the ATR and Transmission Techniques for Protein FTIR

Figure VI.10 overlays the amide-I and –II bands of each of the four standard proteins when either the transmission or ATR techniques of acquiring the FTIR absorption spectrum of each of the proteins were used. The most noticeable difference between the transmission and ATR spectra is the relative intensities of the amide-I and amide-II bands. Transmission FTIR represents the true relative intensities of these bands and it can be seen from Figure VI.10 that the amide-I band is always of higher intensity than that of the amide-II band. An effect of using the ATR accessory is that radiation of lower frequency penetrates to a greater extent beyond the crystal surface and into the contacting sample solution than does radiation of higher frequency. This produces an effect whereby the sample pathlength is not uniform, but rather is a function of the irradiating frequency. As such, bands occurring at lower frequencies experience an intensity increase relative to those located at higher frequencies. This effect can be seen in the altered relative intensities of the amide-I and –II bands for the ATR-FTIR spectra of the standard proteins shown in Figure VI.10. The amide-I bandshape is consequently altered in the ATR absorption spectra for each of the four proteins studied due to the overlap between the amide-I and –II bands, in that the amide-I band intensity is increased at lower frequencies, in the region of this overlap, relative to its intensity at higher frequency, where there exists no overlap between the amide-I and -II bands. This effect can be clearly seen for each of the overlays in Figure VI.10.

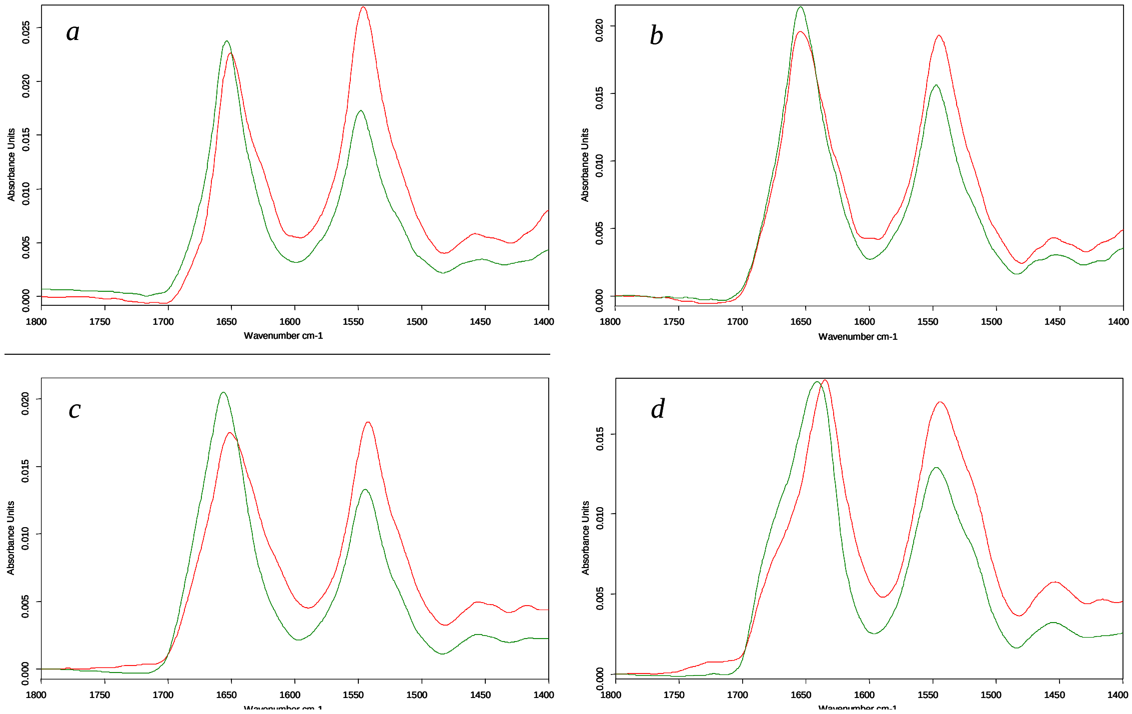


Figure VI.10. A comparison of protein transmission and ATR FTIR spectra. The amide-I, -II and –III bands are compared for both transmission (green) and ATR (red) FTIR techniques for the proteins myoglobin, hemoglobin, lysozyme and α–chymotrypsin as given by *a*, *b*, *c* and *d*, respectively. (Normalisation of the amide-I band area was carried out between 1700 and 1600 cm^{-1} for each spectrum.)

Figure VI.11 makes a similar comparison, but instead the 2$^{\text{ND}}$ derivative spectra are compared in order to examine possible protein structural variation associated with using ATR-FTIR spectroscopy. From the decrease in intensity of the bands at approximately 1655 cm^{-1} the total amount of α-helix structure present within each of the proteins can be seen to decrease upon adsorption of each protein to the ATR crystal surface, relative to that seen using transmittance FTIR. The emergence of prominent bands at 1632 cm^{-1} for both hemoglobin and lysozyme is evidence that both proteins experience pronounced increases in their β-sheet content upon adsorption to the ATR surface, with the expected minor β-sheet bands at 1682 cm^{-1} also present in both spectra. Similarly, the ATR 2ND derivative spectrum of α-chymotrypsin shows an increased level of β-sheet content relative to that of the transmittance equivalent, as seen by the increase in the intensity of the band at 1636 cm^{-1}. For the highly α-helical proteins of myoglobin and hemoglobin, it appears that both proteins become substantially more randomly coiled when adsorbed onto the ATR crystal surface. The red-shifting of the main band in the myoglobin 2ND derivative spectrum from 1656 cm^{-1} in transmittance mode to 1652 cm^{-1} in ATR mode, along with a band broadening effect in the ATR spectrum, can be understood as an increase in the random coil content of myoglobin. For hemoglobin the effect is even more striking, with a marked increase in the random coil band, along with a blue-shift to a more characteristic location of 1643 cm^{-1} for this band. No significant increase in the extent of aggregate formation was observed upon protein adsorption to the crystal surface, as can be seen from the 2ND derivative overlays in the frequency range of between 1625 and 1620 cm^{-1}.

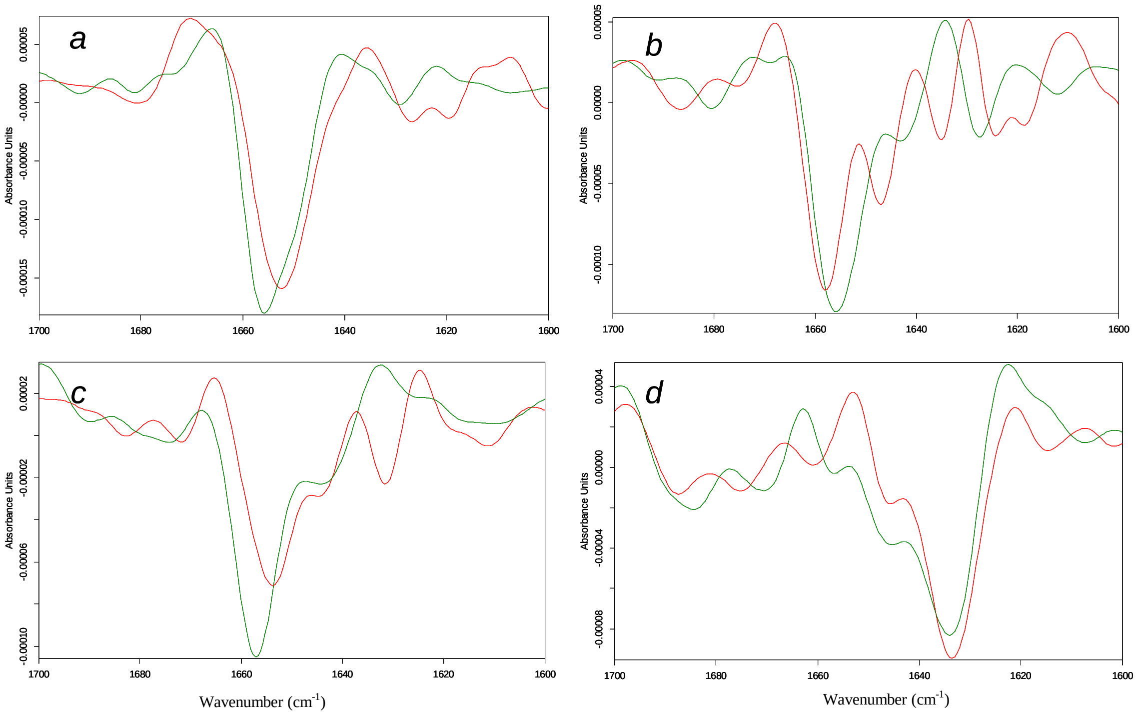


Figure VI.11. Comparison of transmittance and ATR 2ND derivative amide-I FTIR absorption spectra of standard proteins. The amide-I band absorption FTIR 2ND derivative spectral overlay of myoglobin, hemoglobin, lysozyme and α-chymotrypsin when using either transmittance (green) or ATR (red) spectroscopic techniques are shown as *a*, *b*, *c* and *d*, respectively.

VI.3.5 Protein FTIR in D_2O Solvent

Figure VI.12 shows the amide-I bands of the four standard proteins myoglobin, hemoglobin, lysozyme and α-chymotrypsin when dissolved in D_2O solvent. The peak maxima of between 1650 and 1655 cm^{-1} and the overall amide-I bandshapes for both myoglobin and lysozyme are as expected for predominantly α-helical proteins (Figure VI.12*a* and *c*). The amide-I band maximum of 1638 cm^{-1} for the α-chymotrypsin spectrum is as expected for a predominantly β-sheet protein (Figure VI.12*d*). The amide-I band maximum of hemoglobin in D_2O solvent is found to be located at approximately 1645 cm^{-1}, demonstrating the sizable shift in vibrational frequency that can occur upon isotopic substitution of peptide hydrogen atoms with deuterium (Figure VI.12*b*).[2]

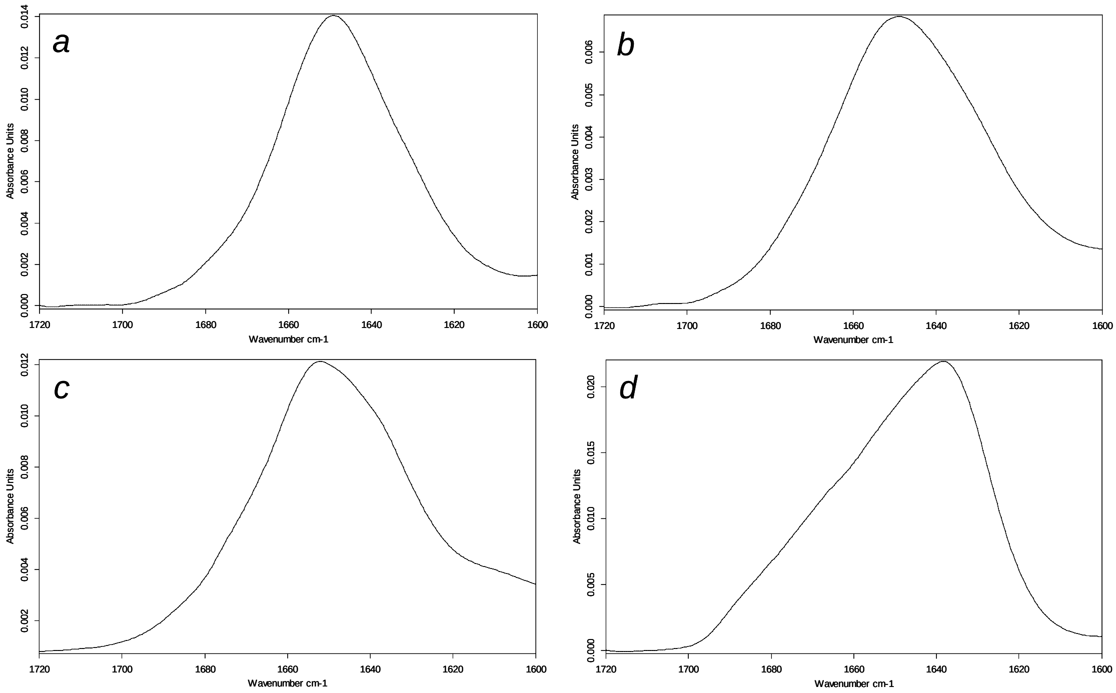


Figure VI.12. The amide-I bands of standard proteins in D_2O solution. The amide-I bands of myoglobin, hemoglobin, lysozyme and α-chymotrypsin are shown as *a*, *b*, *c* and *d*, respectively. The FTIR absorption spectra were acquired using transmittance FTIR.

Figure VI.13 compares the 2^{ND} derivative spectra of the amide-I bands of each of the four standard proteins when dissolved in D_2O versus H_2O solvent. A general feature of the D_2O protein 2^{ND} derivative bands with respect to their H_2O counterparts is that superior band resolution is achieved when analysing proteins in D_2O solvent, as evident from Figure VI.13*a*, *c* and *d*. The expected red-shift in the amide-I band maximum for predominantly α-helical proteins when dissolved in D_2O versus H_2O is apparent from the 2^{ND} derivative spectral overlays of myoglobin, hemoglobin and lysozyme (Figure VI.13*a*, *b* and *c*, respectively). From Figure VI.13*a* it can be seen that the secondary structure of myoglobin is unaltered when moving from an aqueous to D_2O solvent. Figure VI.13*c* reveals that lysozyme retains all of its native band components when moving from H_2O to D_2O, however, a considerable increase in band intensity at 1636 cm^{-1} is observed in D_2O solvent. Also, a dramatic increase at the same wavenumber is seen in moving from H_2O to D_2O solvent for the predominantly β-sheet protein α-chymotrypsin (Figure VI.13*d*). Hemoglobin displays a considerable decrease in the intensity of the helical band when dissolved in D_2O compared to H_2O solvent (Figure VI.13*b*). Residual aggregation bands are again present in the D_2O spectra of both myoglobin and hemoglobin.

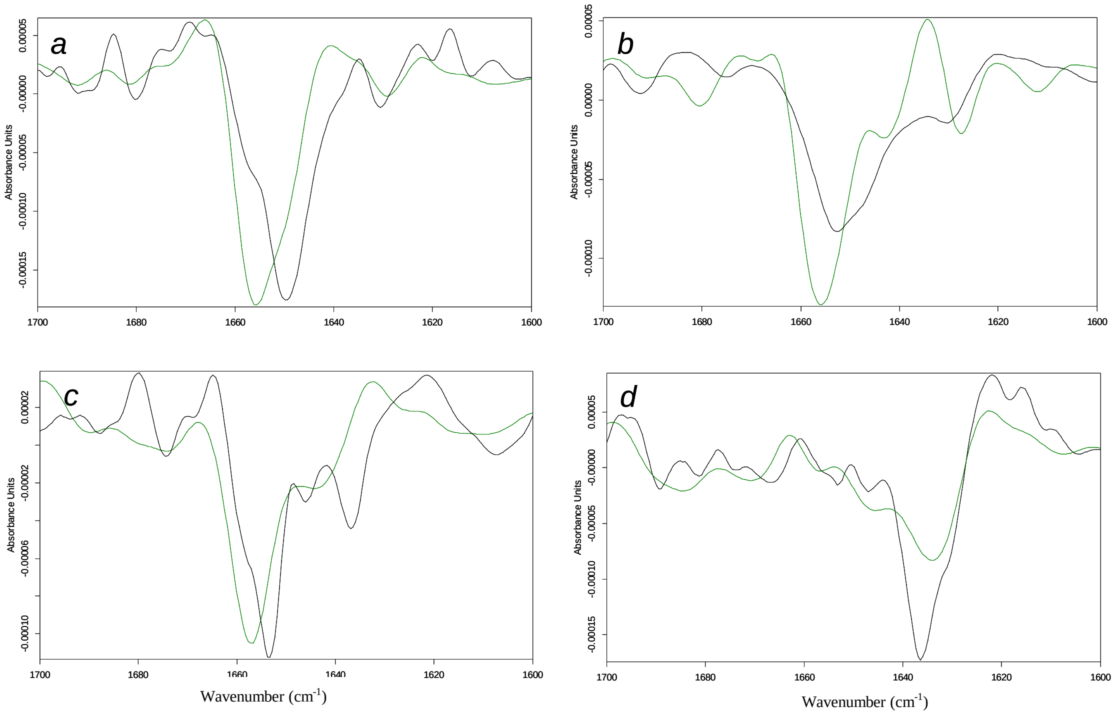


Figure VI.13. The amide-I band 2^{ND} derivative spectra of standard proteins in H_2O versus D_2O solution. The amide-I band 2^{ND} derivative spectral overlays of H_2O (green) and D_2O (black) solutions of myoglobin, hemoglobin, lysozyme and α-chymotrypsin are shown as *a*, *b*, *c* and *d*, respectively. All FTIR absorption spectra were acquired using transmittance FTIR.

VI.3.6 Protein Melts and Two-Dimensional FTIR Spectroscopy

Figure VI.14*a* shows the absorption spectral overlay of the myoglobin amide-I band variation as a function of temperature. From Figure 14*a* and *b* myoglobin can be seen to demonstrate high stability of secondary structure content up to a temperature of 45 ^{o}C, illustrated by the blue coloured spectra. Beyond 45 ^{o}C there occurs a marked degree of degradation of α-helicity, as can be seen by the decrease in intensity of the amide-I band at the α-helix band maximum of 1655 cm^{-1}, shown by the green coloured spectra. The decrease in α-helix content at higher temperatures accompanies an increase in amide-I band intensity in the aggregation band regions of between 1620 to 1628 cm^{-1} and 1690 to 1695 cm^{-1}. There is also a considerable increase in band intensity in the amide-I region corresponding to 3_{10}-helix and β-turn conformation between 1660 and 1682 cm^{-1}.

The variation of the 2ND derivative myoglobin amide-I spectra gives more precise insight into the structural changes brought about by increasing temperature (Figure VI.14*b*). The decrease in the band centred at 1655 cm^{-1} corresponds to the loss of α-helix structure with increased temperature. The major and minor aggregation bands at 1622 and 1696 cm^{-1}, respectively, have clear temperature dependence. The emergence of a band centred at 1666 cm^{-1} with increasing temperature is evidence that myoglobin acquires 3_{10}-helix content at elevated temperatures. The myoglobin 2ND derivative spectra used for generating its aggregation profile (Figure VI.14*d*) is shown in Figure VI.14*c* and the calculated $T_{1/2}$ value of 68 ^{o}C is consistent with it being a very temperature-stable protein.

Figure VI.15 shows the two-dimensional (2D) correlation analysis of the myoglobin amide-I band for the melt experiment shown in Figure VI.14. Figure VI.15*a* shows the synchronous plot revealing an intense positive auto peak at the α-helix band position of approximately 1650 cm^{-1}. The asynchronous plot is shown as Figure VI.15*b* and features a number of cross peaks. Cross peaks occurring at the coordinates (1655, 1632), (1655, 1663) and (1655, 1678) correlate a decrease in the spectral intensity at 1655 cm^{-1} with increases in spectral intensities at 1632, 1663 and 1678 cm^{-1}, respectively. This correlates the loss of α-helix content to the gain of β-sheet, 3_{10}-helix and β-turn content, respectively, for myoglobin when subjected to increasing temperature variation. The lack of a significant cross peak correlating the aggregation band at 1623 cm^{-1} with any other secondary structure in the asynchronous map suggests that aggregation formation in myoglobin occurs in-phase with other structural changes. It can, therefore be assumed that loss of α-helix structure occurs simultaneous to the formation of protein aggregates.

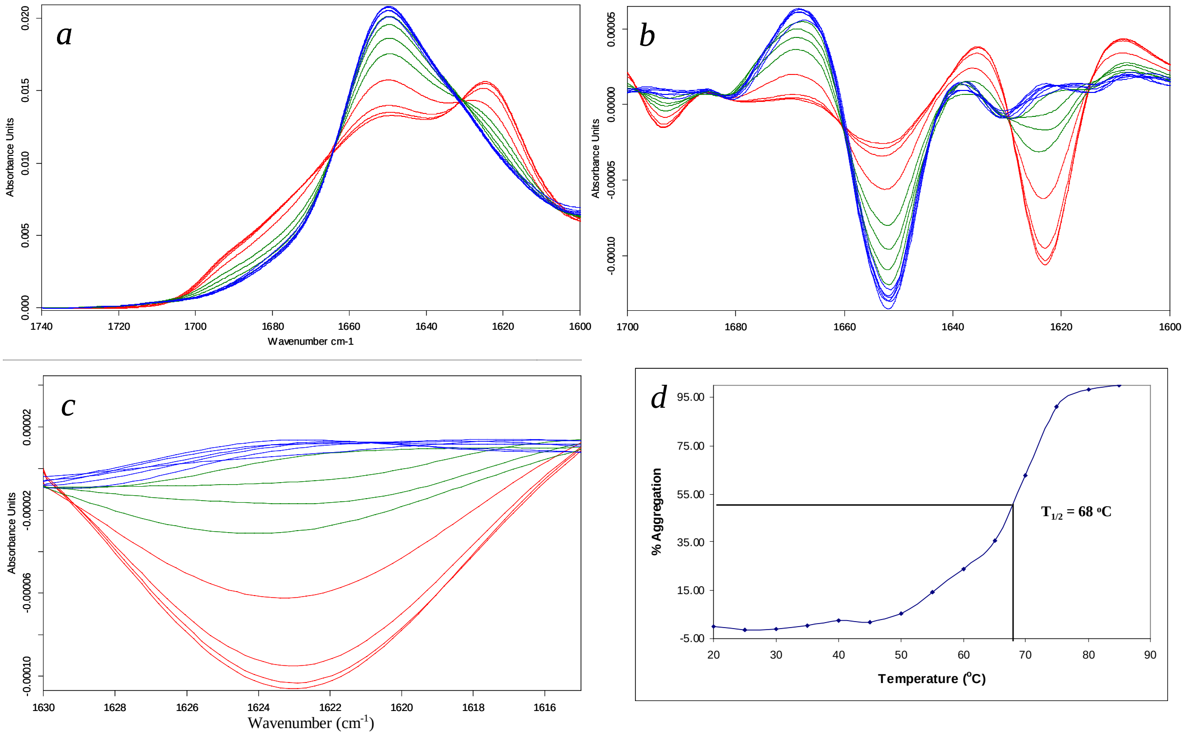


Figure VI.14. Myoglobin ATR-FTIR melt experiment. *a* Myoglobin amide-I absorption band spectral overlay with varying temperature. The spectra coloured blue, green and red correspond to the temperature ranges 20-45, 50-65 and 70-85 ^{o}C, respectively. *b* Myoglobin amide-I 2^{ND} derivative spectral overlay. *c* The 2^{ND} derivative spectra over the wavenumber range used in generating the model for determining the aggregation profile of myoglobin. *d* The myoglobin aggregation profile showing the $T_{1/2}$ value as 68 ^{o}C.

The negative sign of the cross peaks at (1655, 1632) and (1655, 1663) reveal that both β-sheet and 3_{10}-helix secondary structures, respectively, are formed before the loss of α-helix content and, therefore, that these newly formed structures are from residues not originally involved in forming the native helix domains. The positive sign of the cross peak at (1655, 1678) suggests that β-turn structures are formed after the loss of α-helix content and, therefore, most likely involve those residues which originally had formed the α-helix protein domains.

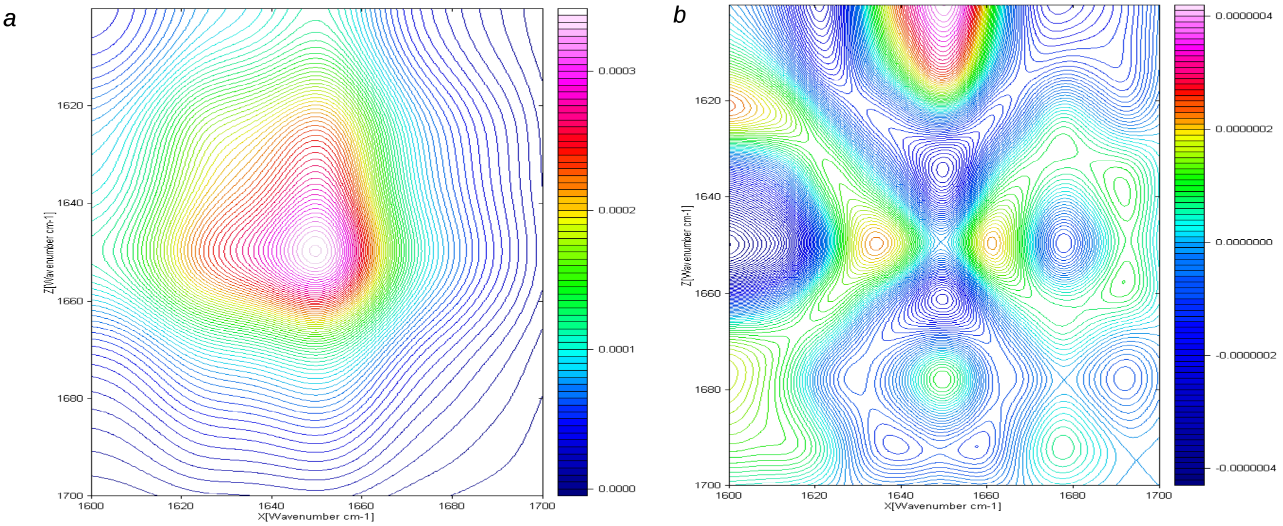


Figure VI.15. Two-dimensional correlation analysis of myoglobin from ATR-FTIR spectroscopy. The myoglobin synchronous and asynchronous 2D correlation spectra are shown as *a* and *b*, respectively.

Figure VI.16*a* and *b* show the amide-I absorption and 2^{ND} derivative overlays, respectively, of the hemoglobin melt experiment. Spectra coloured blue, green and red represent the temperature ranges 10-36, 38-52 and 52-90 °C, respectively. The blue coloured spectra represent the temperature range over which hemoglobin is predominantly stable with respect to its secondary structure. Some variation in the spectral intensity at 1655 cm^{-1} corresponding to the α-helix amide-I signal is evident, but overall the protein appears to retain its native secondary structure. The green spectra represent the temperature range over which hemoglobin begins to unfold and, subsequently, aggregate. Red spectra represent the complete unfolding and intense aggregation of hemoglobin. Band decreases at 1655 and 1638 cm^{-1}, representing loss of α-helix and β-sheet content, respectively, and increases at 1622 and 1693 (aggregation), 1666 (3_{10}-helix) and 1678 cm^{-1} (β-turn) with increasing temperature can clearly be seen in Figure VI.16*b*. The $T_{1/2}$ value for hemoglobin was calculated as 54 °C as shown in Figure VI.16*d*, defining it as a reasonably temperature-stable protein.

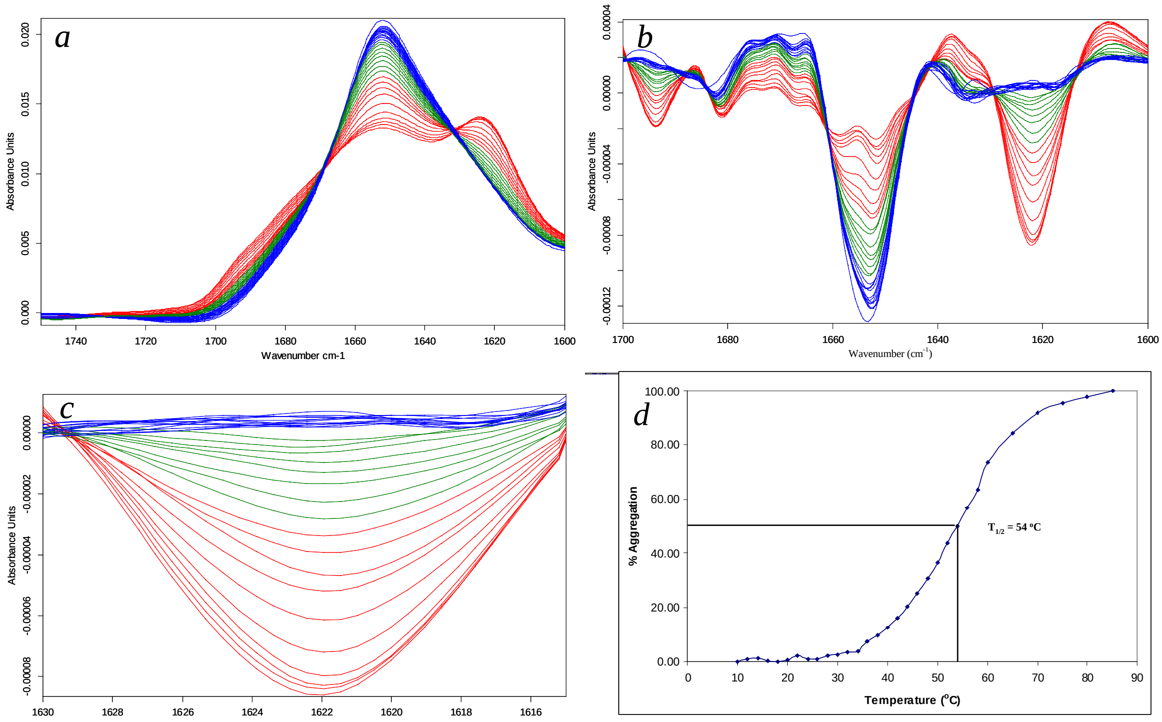


Figure VI.16. Hemoglobin ATR-FTIR melt experiment. *a* The hemoglobin amide-I absorption spectral overlay. The spectra coloured blue, green and red were acquired over the temperature ranges 20-34, 35-39 and 40-85 ^{o}C, respectively. *b* The 2ND derivative spectral overlay. *c* The 2ND derivative spectral overlay for the wavenumber range used in generating the model for determining the $T_{1/2}$ value of hemoglobin. *d* Hemoglobin aggregation profile showing the $T_{1/2}$ value as 54^{o}C .

Figure VI.17 shows the 2D correlation analysis of the hemoglobin melt experiment given in Figure VI.16. The synchronous and asynchronous plots are shown as Figure VI.17*a* and *b*, respectively. Similar to myoglobin, the synchronous plot contains one major auto peak at the α-helix band location of 1655 cm^{-1}. The asynchronous plot shows cross peaks correlating the intensity variation at 1655 cm^{-1} with that at both 1620 and 1678 cm^{-1}. The negative cross peaks at coordinates (1655, 1620) and (1655, 1678) suggest that the formation of aggregate and β-turn structures, respectively, occur predominantly before the loss of α-helix content. This is the reverse of that observed for myoglobin, where it was found that aggregation and β-turn formation occurred either simultaneous to or after the loss of α-helix secondary structure.

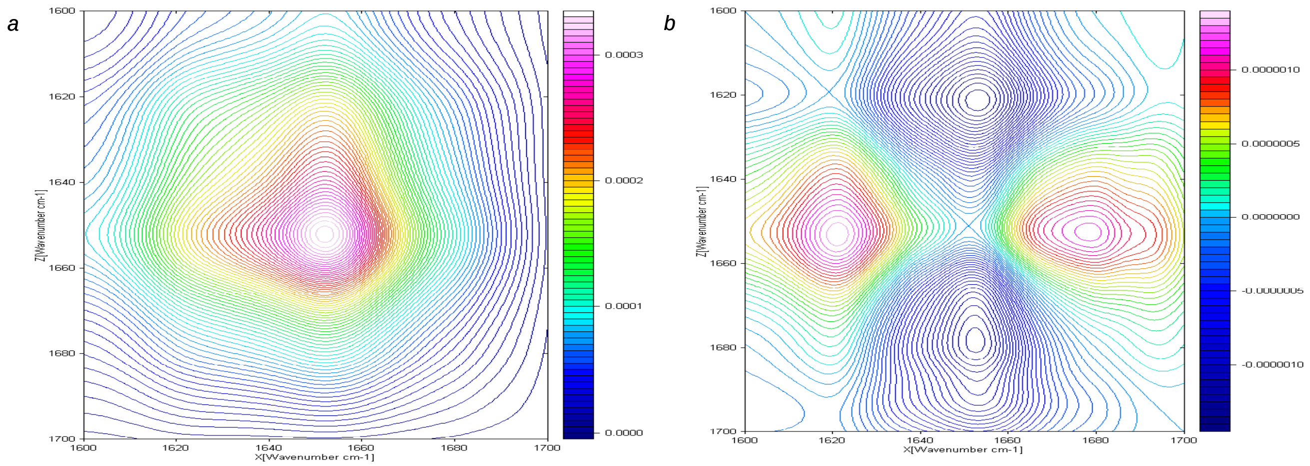


Figure VI.17. Two-dimensional correlation analysis of hemoglobin ATR-FTIR spectroscopy. The hemoglobin synchronous and asynchronous 2D correlation spectra are shown as *a* and *b*, respectively.

Figure VI.18 shows the melt experiment of α-chymotrypsin. Over the temperature range 20-34 ^{o}C the β-sheet content of α-chymotrypsin becomes somewhat reduced, as can be seen for the blue spectra from the decreasing intensity of the amide-I band at approximately 1635 cm^{-1}. No other structural variation is observed over this temperature range. Spectra recorded over the temperature range 35-39 ^{o}C are coloured green and reveal an increased loss of β-sheet content over this range. The red spectra acquired from 40-85 ^{o}C reveal a further loss of β-sheet content along with the accumulation of protein aggregates, as revealed by the emergence of a band at 1626 cm^{-1}. The less than complete loss of β-sheet content is reflected in the maximum of the aggregation band being slightly blue-shifted from its expected location of 1622 cm^{-1}, as can be seen from Figure VI.18*b*. The transition from native intra-molecular β-sheet conformation to inter-molecular β-sheet aggregate formation with increasing temperature can also be seen from the decrease in the minor inter-molecular β-sheet band at 1680 cm^{-1} and the growth of the minor intra-molecular β-sheet band at 1693 cm^{-1} (Figure VI.18*b*). An increase in helix and/or random-coil content seems evident from the increase in the 2ND derivative spectral intensity between 1645 and 1665 cm^{-1} with increased temperature. The α-chymotrypsin T$_{1/2}$ value of 55 ^{o}C is characteristic of that of a reasonably temperature-stable protein.

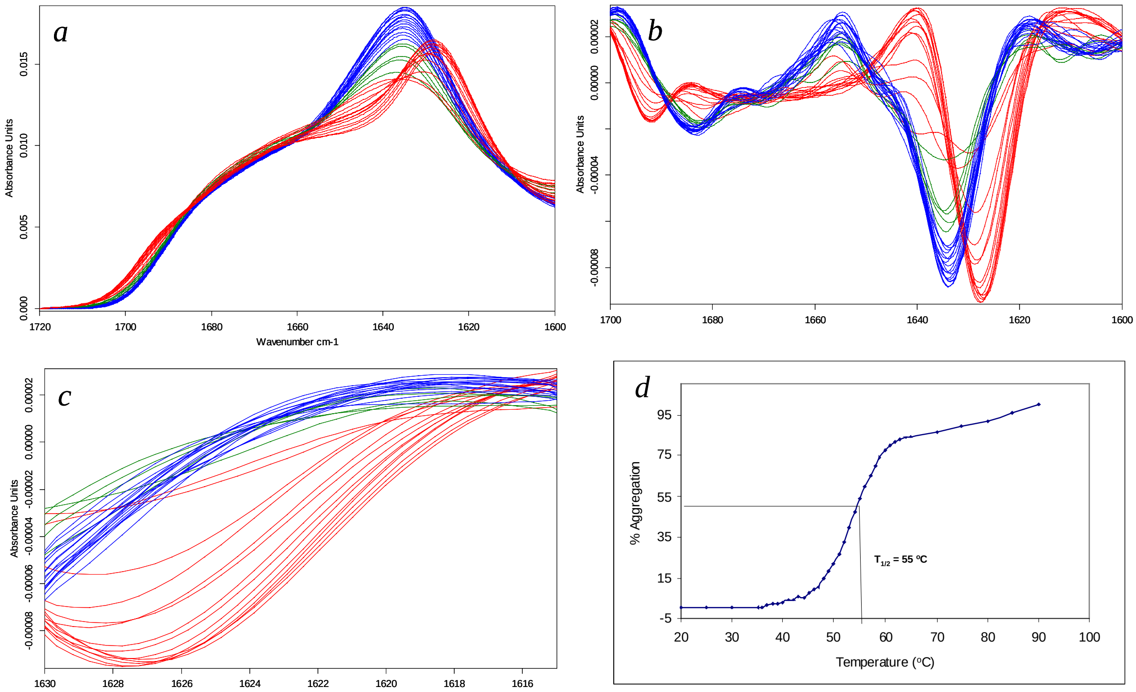


Figure VI.18. α-chymotrypsin ATR-FTIR melt experiment. *a* The amide-I absorption spectral overlay of α-chymotrypsin. Blue, green and red spectra represent the temperature ranges 20-34, 35-39 and 40-90 °C, respectively. *b* The 2ND derivative spectral overlay of the absorption spectra shown in *a*. *c* The 2ND derivative absorption spectral overlay over the wavenumber region used to generate the model for calculating the α-chymotrypsin $T_{1/2}$ value. *d* The aggregation profile of α-chymotrypsin.

The 2D correlation analysis of α-chymotrypsin is shown in Figure VI.19. A single auto peak at 1638 cm^{-1} can be seen in the synchronous plot and corresponds to the intensity change brought about by the loss of native β-sheet structure with increasing temperature (Figure VI.19*a*). The most prominent feature of the asynchronous plot is the cross peak at (1638, 1626), correlating the loss of β-sheet structure with the gain of protein aggregates (Figure VI.19*b*). The positive sign of this cross peak indicates that the β-sheet structure is lost before the protein aggregates appear, as is apparent from Figure VI.18. Smaller cross peaks at (1640, 1664) and (1640, 1678) indicate a correlation between the loss of β-sheet content and the increase in 3_{10}-helix and β-turn content, respectively. The negative sign of these cross peaks suggest that both of these structures are formed before the loss of the native β-sheet content and so that neither of these structures are formed from residues involved in forming native β-sheets. Two other cross peaks located at identical wavenumbers to the former two, but correlated to the aggregation band at 1626 cm^{-1} appear as cross peaks at (1626, 1664) and (1626, 1678). These cross peaks relate the decrease of spectral intensity at 1664 and 1678 cm^{-1} to the growing of the aggregation band at

1626 cm^{-1}. The positive sign of these cross peaks suggest that both 3_{10}-helix and β-turn conformations form before the completion of protein aggregation.

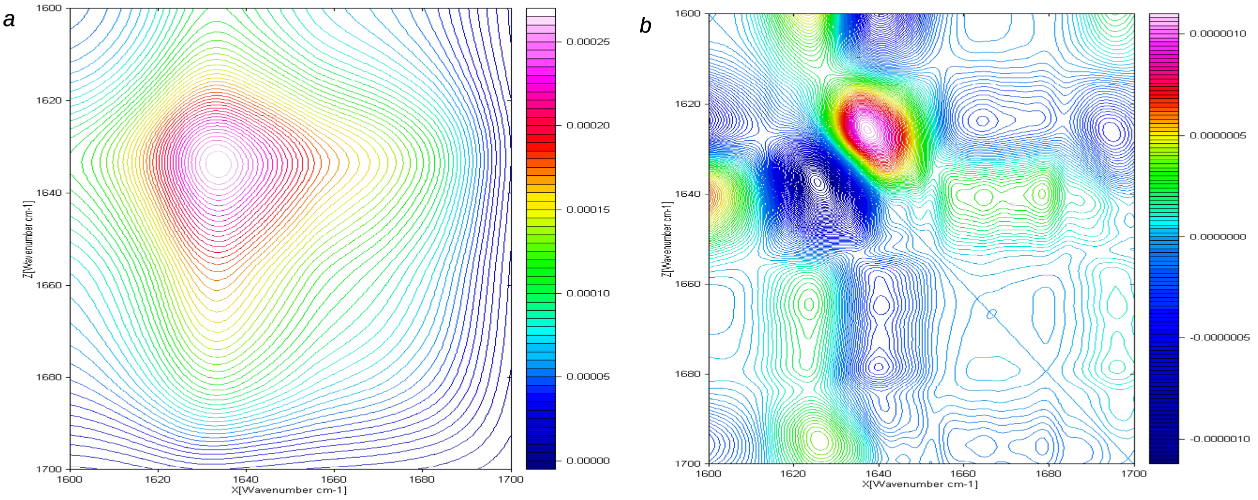


Figure VI.19. Two-dimensional correlation analysis of α-chymotrypsin ATR-FTIR spectroscopy. The α-chymotrypsin synchronous and asynchronous 2D correlation spectra are shown as *a* and *b*, respectively.

Despite repeated attempts to monitor the aggregation process of lysozyme the temperature dependence of lysozyme's secondary structure was only achieved up to a temperature of 51 °C. From Figure VI.20*b* it appears that considerable variation of the α-helix content of lysozyme occurs over this temperature from the magnitude of the band at 1655 cm^{-1}. Similarly, the non-uniform behaviour of a small band between 1640-1635 cm^{-1} indicates a variation of β-sheet content. A slight increase in band intensity can be seen at approximately 1680 cm^{-1}, indicating a small increase in the β-turn content of lysozyme over this temperature range. The apparent onset of aggregation can be seen in Figure VI.20*b* from a band emerging at approximately 1620 cm^{-1} in the red coloured spectrum acquired at 51°C. This is accompanied by a further reduction in α-helix content. The failure to successfully acquire well-behaved spectra beyond 51 °C for lysozyme is most likely related to anomalous behaviour of the protein at the ATR-crystal surface at higher temperatures during spectral acquisition. The non-uniform behaviour towards the loss or formation of particular secondary structures makes lysozyme a troublesome protein for the performance of a 2D correlation analysis over this temperature range.

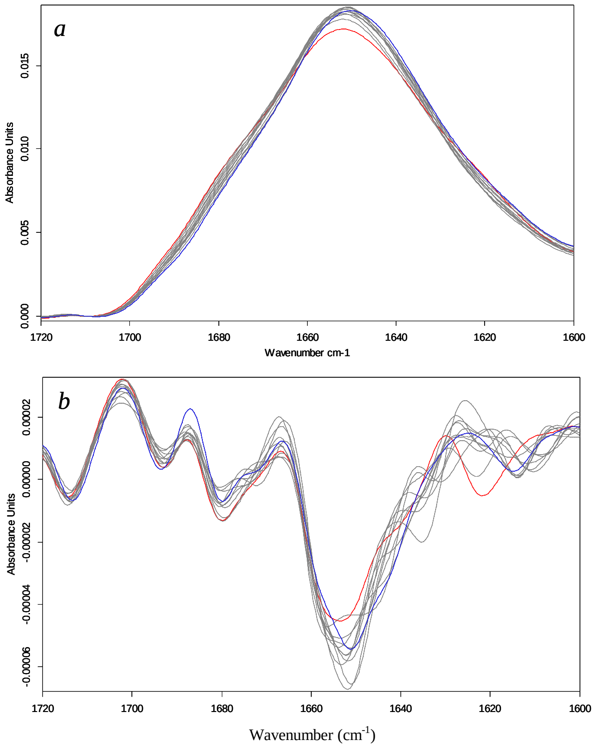


Figure VI.20. The ATR-FTIR melt experiment of lysozyme. *a* and *b* show the amide-I absorption band and 2ND derivative spectral overlays, respectively. The blue and red spectra were acquired at 18 and 51 ^{o}C, respectively. Spectra taken at intermediate temperatures are coloured grey.

VI.4 Discussion

VI.4.1 Quantitative Analysis of Secondary Structure from Protein FTIR

One of the most challenging problems when working with FTIR spectroscopy for studying protein secondary structure is achieving the high protein concentrations necessary for this technique while at the same time preventing protein aggregation. FTIR protein spectroscopy routinely demands protein concentrations of between 5 and 20 mg/ml. Since proteins do not remotely approach these concentrations *in vitro* this exposes them to higher risks of interacting with each other and forming aggregates. Despite careful steps to reduce the risks of aggregates being present within the sample, a small amount of aggregation was unavoidable for some of the proteins studied. Particularly sensitive to this effect were the highly α-helical proteins of myoglobin and hemoglobin (Figures VI.2 and VI.3). This likely contributed to a decrease in performance when trying to estimate the α-helix and β-sheet contents of these proteins. Where aggregation was found not to occur to large extents, as in the case of lysozyme and α-chymotrypsin, the estimation of both α-helix and β-sheet was found to be very good (Table VI.2).

Although no exact attempt at solving for the fractional composition of random coil within the proteins studied was made, the intensity of bands centred between 1640 and 1645 cm^{-1} correlates well with the extent of random coil expected for each of the proteins studied (Figure VI.11). Of the proteins studied, α-chymotrypsin has the largest band at 1643 cm^{-1}, followed by lysozyme, hemoglobin and myoglobin. The relative intensities of these bands to other bands present in the amide-I 2^{ND} derivative spectrum are in good agreement with the expected fractional compositions from x-ray crystallographic data (Table VI.1), which define the random coil content of myoglobin, hemoglobin, lysozyme and α-chymotrypsin as 14.4, 13.6, 22.1 and 37.1 %, respectively, of the total protein structure. Also, the presence of the expected β-turn fractional content within lysozyme and α-chymotrypsin of 4.7 and 10.6 %, respectively, agrees well with the intensities of bands present at 1672 cm^{-1} in both transmission FTIR 2^{ND} derivative spectra (Figures VI.4 and VI.5).

VI.4.2 Fourier-deconvolution versus 2^{ND} derivative analysis

The composite nature of a protein's amide-I band, representing all of the secondary structures present within the protein, is clear from the FTIR spectra presented above. From the unprocessed amide-I band it is difficult to distinguish the different secondary structures present within a particular protein. To this end the band-narrowing techniques of Fourier-deconvolution and 2^{ND} derivative analysis were employed and their individual and relative effectiveness can be seen above. Both techniques were seen to be highly effective in deconvoluting the individual secondary structure components from the composite amide-I absorption band, in that the locations of all deconvoluted bands correspond well to the expected locations of the characteristic secondary structural motifs. In examining the relative performance of Fourier-deconvolution and 2^{ND} derivative procedures in deconvoluting the overlapping bands of the composite protein amide-I band it was found that both techniques performed approximately equally well. The most difficult bands to accurately resolve are those arising from the α-helix and random coil conformations. This results from the closeness of the peak maxima of the amide-I bands of each (1660-1650 cm^{-1} for α-helix and 1645 cm^{-1} for random coil), but also from the broad bandshape of the random coil amide-I. Fourier-deconvolution showed itself to be slightly superior to 2^{ND} derivative analysis for resolving the α-helix and random coil band components, particularly in the case of myoglobin (Figure VI.2*b* and *c*). The reason for this is the apparent bias 2^{ND} derivative analysis has for the α-helix component, which tends to dominate other close lying components, and the source of this bias is most likely attributable to the distinctly lorentzian bandshape of the α-helix amide-I.

VI.4.3 H_2O versus D_2O solvent for Protein FTIR

Performing FTIR analysis of aqueous protein samples is very challenging due to the overlapping water band with the protein amide-I band. Since there are many more water molecules than protein molecules, the OH bending vibration mode from the water within the sample absorbs far more strongly than the protein amide vibrational modes. The extreme temperature dependence of the energy of the water OH bending vibration adds a further complication to performing aqueous FTIR spectroscopy of protein molecules. In order to successfully reference out the water background from the protein absorption spectrum the preservation of a constant pathlength between reference and sample acquisitions is essential, along with accurate thermostating of the cell. The need to disassemble and reassemble demountable transmission cells between collecting the reference and sample transmission spectra introduces a pathlength uncertainty, making demountable cells unsuitable for aqueous protein FTIR spectroscopy. This pathlength uncertainty is accentuated by the need to use very small pathlengths, of the order of 5 to 8 μm, when working with aqueous protein samples in order that the sample doesn’t become fully absorbing due to the water content of the sample. Demountable cells are also unsuitable for aqueous protein FTIR because the necessary precise regulation of temperature between sample and reference is difficult to achieve. The transmittance FTIR spectra presented above were, therefore, acquired with a non-demountable sealed cell.

The manufacture of sealed cells with very small pathlengths is a challenging process and, therefore, such cells are quite costly. The main difficulty lies in the injection of the aqueous sample into the very thin cavity of the cell without having to apply extreme pressures that would be liable to break the cell windows. The AquaSpec cell addresses this problem by sealing the two windows with an expandable material that allows the cell to be filled by expanding the distance between the windows and then, once the cell is filled, the material contracts to its original dimension, thus reducing the pathlength to the required distance of approximately 6 μm.

The reliable relationship between a protein's amide-I band and its secondary structure has been used extensively for analysing proteins over the past twenty years. However, because of the relatively recent development of small pathlength temperature-regulated FTIR cells and their associated cost, most of the literature concerning protein FTIR is for proteins in D_2O solvent. Given the widely acknowledged sensitivity of protein secondary structure to solvent effects, the question of the effect on a protein's secondary structure of exchanging the native H_2O solvent with non-native D_2O is an important one. Comparing the amide-I bands of the four standard proteins used above in H_2O versus D_2O sheds light on this effect.

Figure VI.13 shows the effects of exchanging H_2O with D_2O solvent on the secondary structures of the four standard proteins studied herein. The well ordered secondary structure regions of myoglobin, lysozyme and α-chymotrypsin all appear stable with respect to solvent exchange. However, for the lysozyme and α-chymotrypsin amide-I spectra in D_2O there appears a considerable increase in band intensity at 1636 cm^{-1}. This is consistent with observations made with the nicotinic acetylcholine receptor upon substitution of H_2O with D_2O and the authors attribute this effect as predominantly arising from the sizable frequency shift of the random coil amide-I band upon H-D isotopic substitution, which occurs rapidly for the solvent exposed residues comprising the random coil protein content.[2] The authors also site the possibility that the increase in band intensity at this position may result from β-sheet regions having increased spectral absorption when dissolved in D_2O compared to H_2O. The considerable reduction in intensity of the helix band for hemoglobin in the H_2O versus the D_2O spectrum is difficult to explain, but is likely related to the broadening of the band width that may result from the significant red-shift of the random coil amide-I band (Figure VI.13*b*). Again, this effect has been seen for the nicotinic acetylcholine receptor.[2] In conclusion, therefore, D_2O may be used with some confidence for studying proteins by FTIR. It should be noted, however, that the large shift of the random coil amide-I band from being located close to the helix amide-I in H_2O to directly overlapping the β-sheet amide-I in D_2O should be taken into account when studying proteins that contain significant levels of random coil secondary structure. The considerable red-shift of the random coil amide-I upon H-D isotopic substitution can be used as an effective means of quantifying the degree of protein random coil content, especially since the random coil amide-I band is located very close to the helix amide-I in aqueous solution and it can be difficult to resolve the two bands. Once this shift is taken into account, the spectra presented in Figure VI.13 indicate that protein secondary structure is largely unaffected by exchanging H_2O with D_2O solvent.

The extensive overlap between the water and amide-I band introduces an unwanted artefact into the FTIR absorption spectrum, which results from a difference between the concentration of the water in the reference and sample solutions. This concentration difference is as a result of the high concentration of protein molecules within the sample solution, which acts to decrease the concentration of water. Therefore, the intensity of the water transmission band will be greater for the reference than for the sample and if this difference is large enough it can result in a deformed amide-I absorption band and improper determination of protein secondary structure. In order to minimise this effect it is desirable to have the protein concentration within the sample solution at close to the minimum level from which a reproducible absorption spectrum may be achieved. In order to protect against inappropriate referencing when generating the absorption spectrum a condition that the protein absorption spectrum have a flat line appearance between 1700 and 1750 cm^{-1} is standard. Inspecting the absorption spectra presented above, it can be seen that the flat line appearance of the spectra beyond 1700 cm^{-1} is well maintained, especially for absorption spectra acquired in transmittance mode, for which the protein concentration is less than that needed for the ATR mode. The occurrence of very small negative water bands can be seen in some of the ATR data presented above and is a result of having to use higher protein concentrations when working with ATR spectroscopy.

VI.4.4 Removal of Water Vapour from Protein FTIR Spectra

A common mistake made when performing protein FTIR spectroscopy is to neglect to remove the water vapour lines from the protein amide-I band before analysing it for its secondary structure components. Unless considerable effort is taken to remove every trace of water vapour from the beam path, it is inevitable that some water gas-phase lines will be present in the protein spectrum. Any water vapour bands present in the protein spectrum are extremely detrimental to the secondary structural analysis of the protein because it happens that certain water vapour bands are located at wavenumbers matching those of all secondary structural elements. To illustrate the possible error introduced by water vapour bands to the protein amide-I band, Figure VI.20 overlays the same protein spectrum of lysozyme both before (red) and after (blue) the performance of the water vapour removing post-processing step, carried out for each of the protein spectra presented above. Despite continuous purging of the sample compartment with dry air it can be seen that some water vapour bands are present in the original spectrum. Within the amide-I band region water vapour signals can be seen at approximately 1615, 1620, 1627, 1636, 1643, 1653, 1675 and 1682 cm^{-1}. Since an approximate coincidence to the band positions of all of the secondary structural elements exists, a common mistake is to assign the water vapour bands to secondary structure components. It is obvious that this would lead to serious errors in the prediction of the secondary structural fractional composition of the protein being investigated.

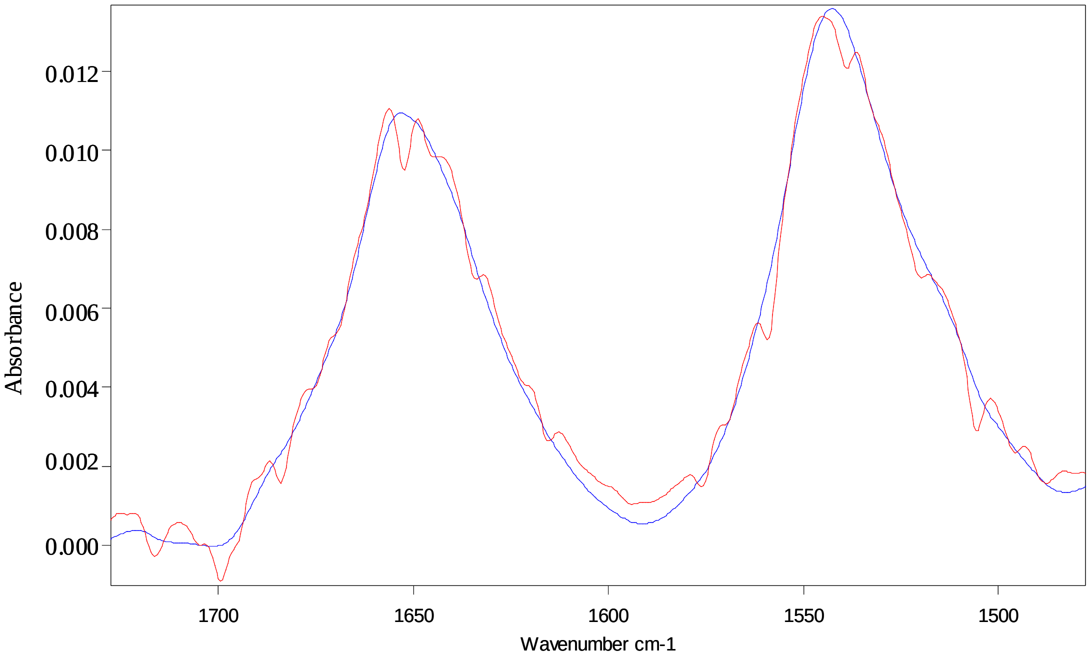


Figure VI.20. The effect of water vapour bands to the protein amide FTIR spectrum. The ATR-FTIR absorption spectrum of lysozyme in the region of the amide-I and –II bands both before (red) and after (blue) the performance of the water vapour removing post-processing procedure.

VI.4.5 Transmission versus ATR for Protein FTIR Spectroscopy

Comparing the absorption spectra of the standard proteins acquired with transmission versus ATR-FTIR spectroscopy reveals a clear shift in the relative intensities of the amide-I and –II bands. Since only the amide-I band is used for protein structural analysis this does not seem to be of immediate importance. However, spectral overlap between the two bands results in this change in amide-I and –II relative intensity leading to a change in the bandshape of the amide-I band. Also, the ATR effect of making the absolute optical pathlength dependent on the radiation frequency continues to apply over the wavenumber range defining the amide-I band and, therefore, not only alters the relative intensities of the neighbouring amide-I and –II bands, but also changes the amide-I bandshape itself. This might be reflected by a comparison of the amide-I bandshapes of the standard proteins involved in this investigation when acquired by either ATR or transmission FTIR spectroscopy. However, in practise this comparison is difficult to achieve, since other effects occurring at the ATR surface are likely to make a like-with-like comparison of ATR and transmission spectra non-feasible when trying to measure this effect.

The viability of using ATR-FTIR spectroscopy as an effective replacement for transmittance FTIR spectroscopy is examined above. As previously mentioned, the problem of having to use increased protein concentrations when working with ATR compared to transmittance FTIR spectroscopy is potentially troublesome due to the possibility of increased aggregation at higher protein concentrations. The extent of this problem is revealed in Figure VI.11, which overlays the 2^{ND} derivative deconvoluted amide-I of both transmission and ATR FTIR spectra of each of the standard proteins. From this it can be seen that aggregation levels do not significantly increase for any of the proteins studied when prepared at the higher concentrations needed for ATR-FTIR, relative to the samples used for transmission FTIR. It would, therefore, seem that the increased protein concentration requirements for ATR over transmission FTIR spectroscopy are not of the order to rule it out as an alternative technique and that protein aggregation levels are comparable when using either technique.

Another factor to consider when working with ATR-FTIR spectroscopy is the possible effect adsorption of the protein to the ATR surface might have on its secondary structure. The ratio of adsorbed protein molecules to protein molecules in solution that are sampled by the irradiating light beam as it passes through the BioATR-II accessory is dependent on the penetration depth of the evanescent wave (between 1 and 2 μm for the BioATR-II accessory) beyond the ATR crystal surface into the contacting sample solution. This ratio is of crucial importance to the overall effect on the amide-I bandshape that will be caused by any structural perturbation that occurs upon protein adsorption to the ATR-crystal surface. In order to gain insight into the approximate scale of protein adsorption the AavLEA1 protein's adsorption profile with increasing temperature was examined (Figure VI.21). The AavLEA1 protein was chosen for this test due to its lack of structural perturbation and consequent protein precipitation with temperature variation. As such, the change in the amide-I band intensity with temperature increase should be exclusively related to the adsorption process onto the ATR crystal surface. A comparison of this intensity variation with the initial amide-I band intensity gives an idea as to the level of contribution made to the amide-I band intensity by protein molecules adsorbed onto the crystal surface. From Figure VI.21 it can be seen that the amide-I band intensity approximately doubles due to the temperature dependent adsorption of AavLEA1. This is clear evidence that the penetration depth of the IR beam into the sample solution when using the BioATR II is such that most of the protein molecules sampled are either adsorbed onto or lie extremely close to the ATR crystal surface. Therefore, any protein structural perturbation introduced by the process of surface adsorption is likely to be highly represented in the ATR spectrum and represents a serious consideration when using ATR spectroscopy for studying protein secondary structure.

The effect of protein structural variation upon surface adsorption was investigated by Sethuraman and Belfort, by examining the structural perturbation and aggregation effects on globular proteins such as lysozyme when adsorbed onto monolayer surfaces.[3] They observed a conformational shift from α-helix to β-turn and/or random coil and an overall increase in β-sheet content within the secondary structures of proteins adsorbed onto monolayer surfaces at room temperature. Comparing the transmittance and ATR amide-I 2^{ND} derivative absorption spectra for each of the standard proteins examined above yields information about any such changes that result from adsorption to the ATR crystal surface (Figure VI.11). It can be seen that the ATR-FTIR protein spectra display decreased α-helix and increased β-sheet band intensities with respect to their transmission counterparts. Therefore, it would seem that a marked drop in the stability of α-helices and an increase in the stability of β-sheets occurs as a result of protein adsorption to the ATR crystal surface for each of the proteins studied. The random coil content of highly α-helical proteins is also increased upon surface adsorption. In general, adsorption of a protein to an ATR surface leads to non-native secondary structure conformation for that protein and, as such, ATR-FTIR spectroscopy should be used with caution for estimating the secondary structure contents of proteins.

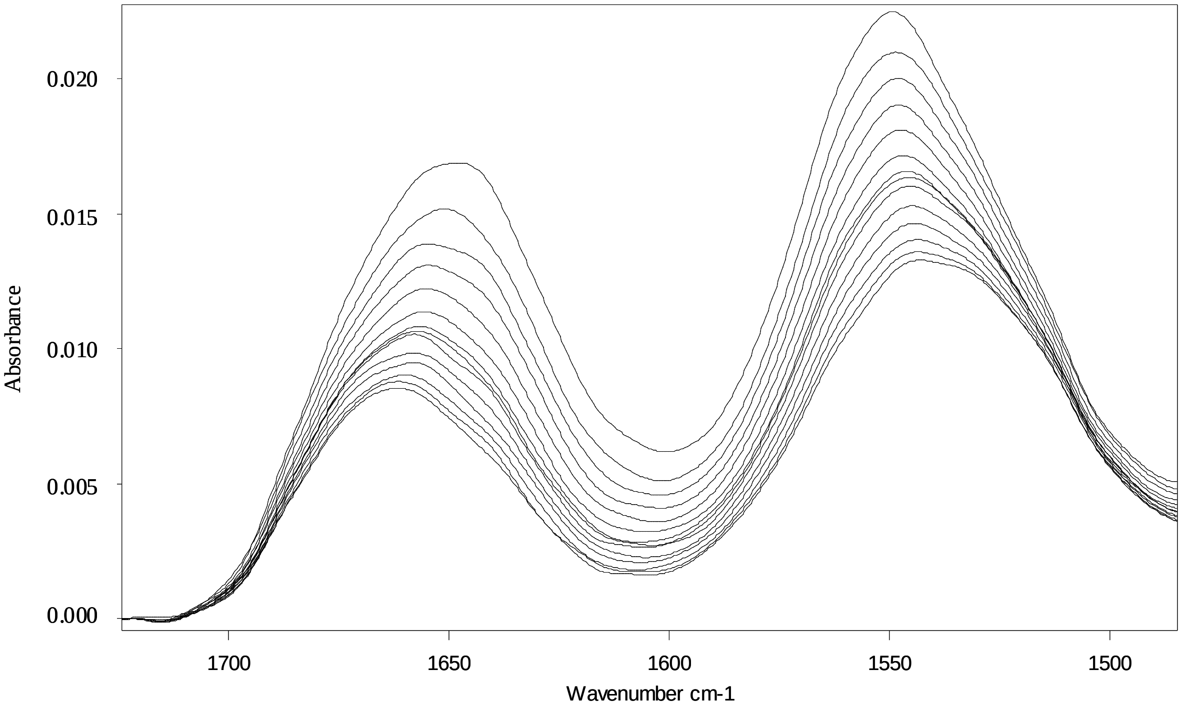


Figure VI.21. Protein adsorption to the ATR crystal surface with increased temperature. The temperature was ramped from 20-90 °C using the AavLEA1 protein, which shows little structural variation with temperature.

VI.4.6 Protein Aggregation Analysis from ATR-FTIR

Studying protein aggregation processes from temperature melt experiments is impractical using transmission FTIR with a sealed cell because the formation of aggregates within the cell cause it to become blocked, which is a serious problem for a sealed cell that cannot be disassembled for cleaning and which relies on an easy flow of liquid for effective gas-bubble-free sample delivery. Another problem with using sealed cells for melt experiments is that the material creating the cell seal often experiences dimensional changes (either expanding or contracting) with temperature variation. Such variations can be difficult to reproduce from one temperature ramp to another and this creates a problem when attempting to subtract the water background in generating the protein absorption spectra over a series of discrete temperatures. Therefore, the ATR-FTIR approach is most often taken when performing protein melt experiments because the ATR cell can easily be cleaned of aggregated sample and the technique offers high temperature stability of pathlength between the collection of the background and sample temperature spectral series. This latter advantage can be seen in the melt experiments presented above by the absence of either positive or negative water signals for each of the temperature series absorption spectra of the four standard proteins studied.

A note of caution should be made, however, when using ATR-FTIR for studying protein melt experiments in that proteins routinely vary in their adsorption to the ATR crystal surface with temperature variation and, as seen from Figure VI.21, this can have a profound effect on the intensity of the amide-I band. This intensity variation with temperature increase can be exacerbated by the occurrence of protein precipitation at elevated temperatures, which has the effect of increasing the concentration of protein close to the ATR crystal surface if a horizontal crystal assembly is used. Therefore, for ATR protein melt experiments a band-intensity-normalising procedure must be performed in order to correct for these effects. However, such normalising procedures are prone to introducing relative errors into the spectral data series since they rely on the approximation that the extinction coefficients of all band components within the composite amide-I band are equal. Combining this approximation with the potential for structural variation of proteins when adsorbed onto the ATR surface and the variation of this adsorption process with temperature detracts from the sensitivity of the ATR technique for analysing protein melt experiments. As such, ATR-FTIR should only really be used for investigating relatively large structural shifts that occur during melt experiments, such as protein aggregation.

ATR-FTIR spectroscopy was successful in studying the aggregation behaviours of three of the four standard proteins studied above. Both α-chymotrypsin and hemoglobin were found to display approximately equal temperature stability and they returned $T_{1/2}$ values of 55 and 54 ^{o}C, respectively. Myoglobin was discovered to be considerably more temperature stable and its $T_{1/2}$ value was measured as 68 ^{o}C.

The acquisition of high quality data for the temperature melt experiments for myoglobin, hemoglobin and α-chymotrypsin allowed for the use of 2D correlation analysis to try to understand the sequence of events involved in the unfolding and aggregation processes of each of the proteins. The question of whether proteins become aggregated in their premolten globule form - whereby they retain most of their secondary structure and adopt a more flexible tertiary conformation - or whether the aggregation process is primarily stimulated by the loss or denaturation of protein secondary structure is of considerable interest to biologists studying protein aggregation behaviours. Since 2D correlation analysis is capable of determining whether secondary structure is lost before or after the onset of aggregation it should provide a valuable tool for investigating protein aggregation phenomena.

For the myoglobin 2D asynchronous plot there occurs no well defined cross peak correlating the loss of α-helix structure with the onset of aggregation (Figure VI.15*b*). Since from Figure VI.14 it is clear that aggregation proceeds at the expense of α-helix band intensity the two processes must, therefore, occur simultaneously. The positive sign of the cross peak in the α-chymotrypsin asynchronous 2D plot correlating the loss of native β-sheet content with the accumulation of protein aggregates means that the β-sheet content is lost before protein aggregates are formed (Figure VI.19*b*). In contrast, the negative sign of the cross peak correlating the spectral intensity change at the α-helix wavenumber with that at the aggregate band wavenumber in the hemoglobin asynchronous 2D plot is evidence that the native α-helix secondary structure is lost predominantly after the formation of protein aggregates (Figure VI.17*b*). From the 2D spectral data it seems that in proteins where there exist high levels of tertiary structure, such as hemoglobin, disruption of this tertiary structure can lead to the aggregation of the premolten globule protein form of the protein. Whereas in proteins containing low levels of tertiary structure a denaturation of secondary structural conformation is necessary to initiate the aggregation process.

VI.5 Protein Concentration Determination by FTIR Spectroscopy

VI.5.1 Introduction

The accurate determination of protein concentration within a sample of soluble protein is of crucial importance when using CD spectroscopy to determine secondary structure content. The most popular method used in biology labs for protein concentration determination is the Bradford technique, but this is unsatisfactory because of the relatively high degree of error involved.[4] The determination of protein concentration by UV absorption spectroscopy, whereby a protein absorption spectrum is produced from the absorption properties of its aromatic residues phenylalanine, tyrosine and tryptophan, is less than ideal, since the absorption extinction coefficient of each of the aromatic residues depends upon its environment within the protein and so a protein's overall extinction coefficient can only be calculated with approximate accuracy from the protein's primary sequence. It is considered that the most accurate methods for determining protein concentrations are by quantitative amino acid (QAA) analysis, which usually returns values that are accurate to within 5-10 %, or the determination of peptide backbone concentration by the measurement of biuret [5] or by that of total nitrogen [6].[7] The performance of each of these procedures, however, requires specialized equipment and this presents a practical barrier for accurate protein concentration determination.

The use of FTIR spectroscopy for protein concentration determination may be feasible, since the intensity of a protein's amide-I band may not be dependent on its secondary or tertiary structures, but rather solely upon the protein's concentration and size – i.e. the total amount of peptide backbone CO groups. As such, it may be possible to generate a concentration calibration curve in terms of protein mg/ml that would be applicable to a protein of any size or structure. From this curve the molar concentration of a protein may be calculated by taking account of the protein's molecular weight. Of course, the absolute concentrations of the protein samples used to generate such a curve would have to be first established by one of the above methods, but the need to continually employ such methods to achieve accurate protein concentration determination could be averted by creating a standard calibration curve. The following work was carried out to investigate the possibility of creating an accurate and reliable protein concentration calibration curve by FTIR spectroscopy.

VI.5.2 Method

Concentrated stock solutions of α-chymotrypsin, lysozyme and myoglobin were made up using 10 mM PBS buffer in D_2O solvent and aliquots of these were sent for concentration determination by QAA analysis. Serial dilutions were made from each protein stock solution using 10 mM PBS buffer in D_2O solvent and the concentration of each dilution was calculated from the initial concentration of the stock solution and the dilution factor used. FTIR spectra were acquired using two 13 mm CaF_2 windows separated by a 100 μm Mylar spacer (Harrick) to create a demountable cell and placed in a DCL-M13 Demountable Liquid Cell (Harrick) sample holder. Spectra were acquired at room temperature, at a resolution of 4 cm^{-1}, using an aperture of 1 mm, over a wavenumber range of between 4000-1000 cm^{-1} and as an average of 50 scans. Absorption spectra were generated by subtracting the transmission spectrum of 10 mM PBS buffer in D_2O solvent from each sample transmission spectrum. The y-coordinate of each sample within the concentration calibration curve plots were calculated by integrating each spectrum between 1600-1700 cm^{-1} to obtain its amide-I band area and this was then plotted along with the sample's concentration in mg/ml.

VI.5.3 Results

Figure VI.22 shows the FTIR amide-I spectral overlays for each of the serial dilutions of each protein. The differences in secondary structure between each of the three proteins used to generate the concentration calibration curve are revealed by comparing their amide-I bands (Figure VI.22*a*). The highly α-helical myoglobin amide-I band (red) is distinctly Lorentzian and has a ν_{max} of approximately 1650 cm^{-1}. The moderately α-helical lysozyme amide-I (green) has a decreased Lorentzian bandshape relative to myoglobin and has an α-helical ν_{max} of approximately 1655 cm^{-1}. Also present in the lysozyme amide-I band is a shoulder at about 1640 cm^{-1}, which reflects the considerable proportion of random coil within its structure. The amide-I band of α-chymotrypsin (blue) displays a ν_{max} of around 1638 cm^{-1}, which is characteristic for β-sheet secondary structure. The minor β-sheet band is just about visible as a slight shoulder around 1685 cm^{-1}. Small discrepancies can be seen in the protein FTIR spectra of Figure VI.22*b*, *c* and *d* in the departure from a straight-line appearance of the spectra at zero intensity at wavenumbers beyond 1750 cm^{-1}. This is not thought to have a large effect on the protein amide-I band.

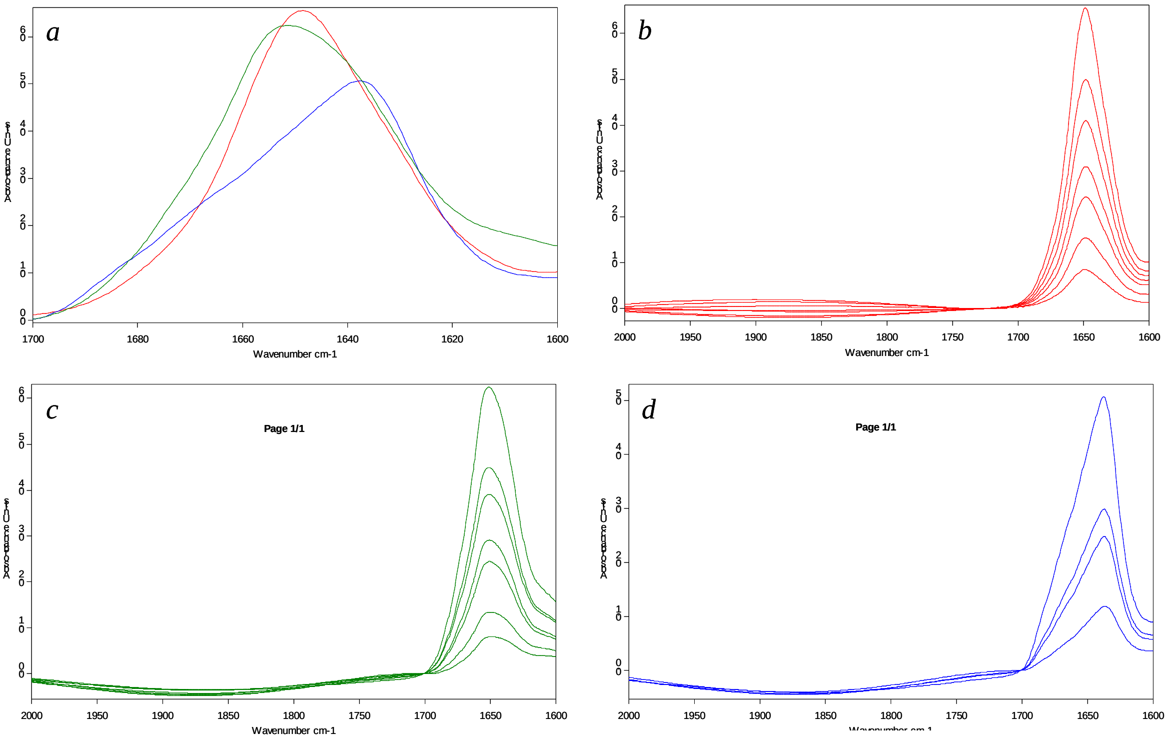


Figure VI.22. The protein amide-I FTIR spectra used to generate the concentration calibration curve. An overlay of the amide-I bands of myoglobin (red), lysozyme (green) and α-chymotrypsin (blue) recorded from the stock solutions of each is shown in *a*. The serial dilution amide-I overlays of myoglobin, lysozyme and α-chymotrypsin are shown as *b*, *c* and *d*, respectively.

Figure VI.23 shows the concentration calibration curve generated by the spectra of Figure VI.22. A straight-line fit of the data points for each of the three proteins was carried out and the equations of these lines are shown in Figure VI.23. The R^2 values for each line are all better than 0.996 and, therefore, there is a very reliable relationship between each protein's amide-I band area and the sample's protein concentration. From each equation it can be seen that the y-intercept was very close to zero for each of the fits, as should be the case. Inspection of Figure VI.23 reveals a close similarity between all three fits and this is evidence that the amide-I band area is largely independent of protein secondary and tertiary structure.

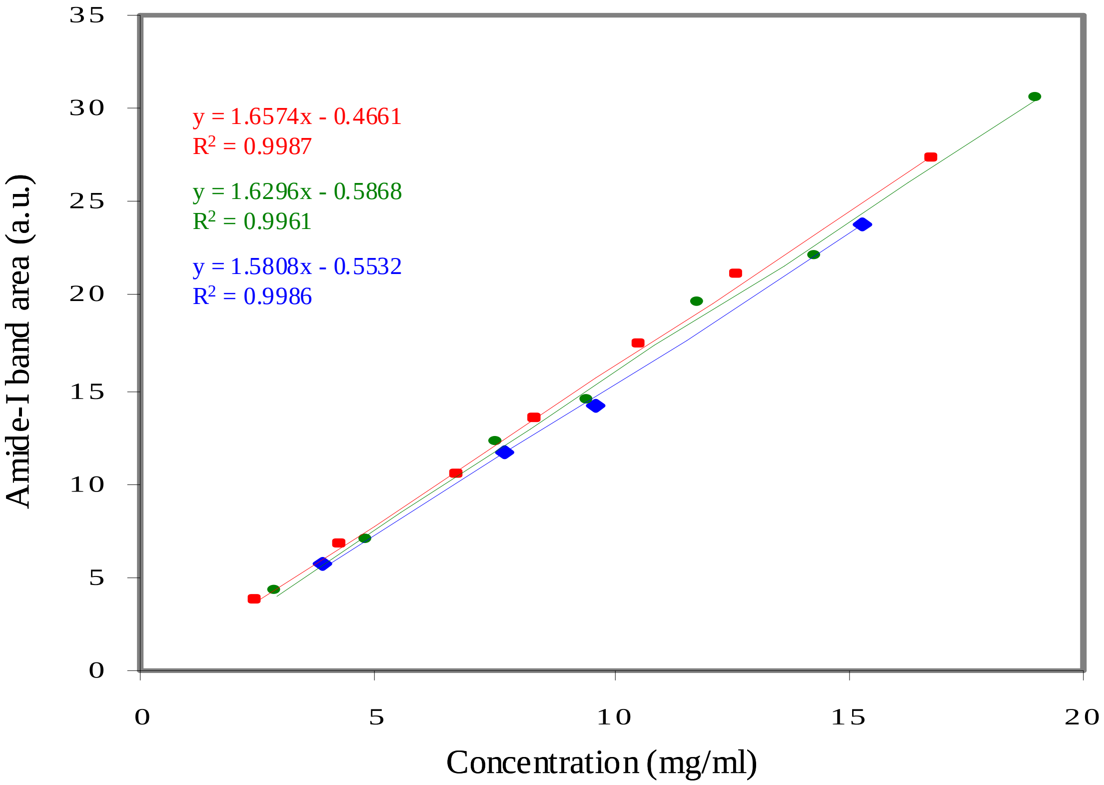


Figure VI.23. Protein amide-I concentration calibration curve. The concentration calibration curve for myoglobin, lysozyme and α-chymotrypsin is shown as the red, green and blue straight-line fits to their data points, respectively. The equations for each are shown in the same colour.

Figure VI.24 shows the straight-line fit of all the data points shown in Figure VI.23, where the y-intercept has been forced through the origin, and represents the final concentration calibration curve. The R^2 value of 0.9949 once again illustrates the independence of amide-I band area upon protein structure. By using the equation

$$c = A/1.5841 \qquad \text{(VI.1)}$$

where c is the protein concentration in mg/ml and A is the amide-I band area, it is possible to accurately calculate the concentration of any protein solution from its FTIR absorption spectrum, provided the spectrum is acquired under the same conditions as those used to generate Figure VI.24.

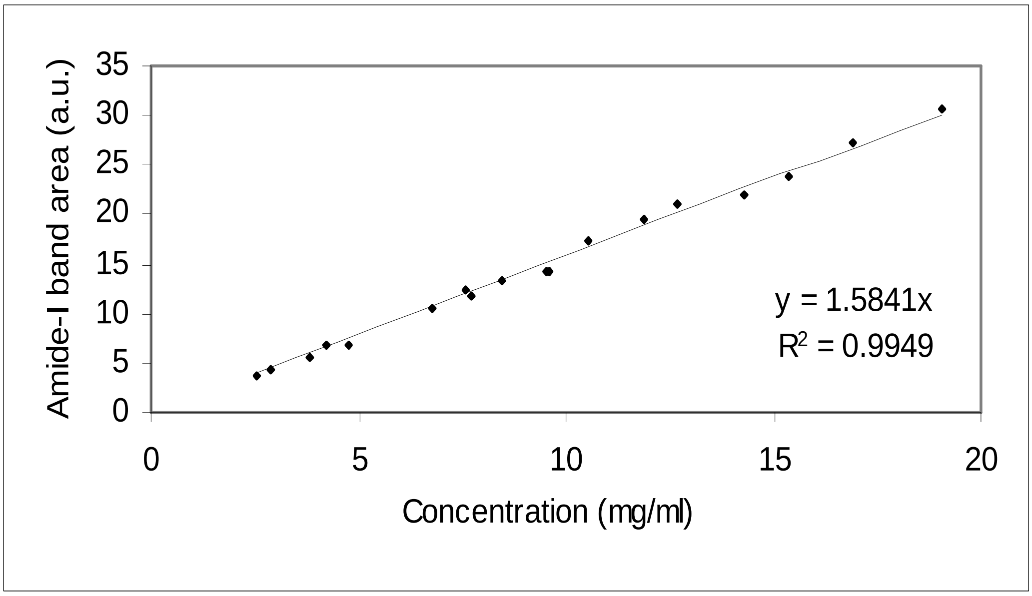


Figure VI.24. Averaged concentration calibration curve. The data points for the myoglobin, lysozyme and α-chymotrypsin concentration calibration curves were used to generate an average concentration calibration curve straight-line fit, where the y-intercept was set to zero.

VI.5.4 Discussion

When using demountable cells for protein FTIR there is a real concern for being able to reproduce the same pathlength for the sample and reference transmission spectrum. This concern is compounded by the weakness of the amide-I absorption band. Where ^1H_2O is used as a solvent, the solvent absorption band directly overlaps that of the protein amide-I and very small pathlengths must be used to prevent spectral saturation in the amide-I region. Achieving adequate pathlength reproducibility for such small pathlengths, where the pathlength is of accentuated consequence due to the strength of ^1H_2O absorption in the amide-I region, is not possible using a demountable cell and so a ^1H_2O solvent could not be used for this study. The vibrational absorption band of D_2O becomes shifted to lower frequencies relative to that of the ^1H_2O band, with the result that it no longer directly overlaps the amide-I band absorption. The effect is to increase the maximum sample pathlength that can be used before spectral saturation in the amide-I 1700-1600 cm^{-1} wavenumber range occurs. In moving to higher pathlength demountable cells, the fractional pathlength error becomes ever smaller and, therefore, the pathlength reproducibility is improved. Also, because D_2O absorption intensity in the amide-I region is considerably less than that of ^1H_2O, there is less stringency to the pathlength reproducibility requirement and accurate protein spectra can be achieved in cases where there is only a slight discrepancy between the pathlength of the sample and the reference.

The results show that their exists no significant dependence of amide-I band area on protein structure. Therefore, accurate protein concentration determination can be achieved using FTIR spectroscopy, once an accurate concentration calibration curve has been established for the particular experimental system. The data above was generated using a 100 μm pathlength cell and well-behaved protein FTIR spectra could only be obtained above a concentration of 2.5 mg/ml. Although by protein solution standards this corresponds to a high sample concentration, another calibration curve could be generated using a longer pathlength, which would allow for the acquisition of well-behaved protein FTIR to lower sample concentrations.

From the biological standpoint, the problem of only being able to determine protein concentrations in D_2O by FTIR might appear very limiting, since it is the aqueous properties of a protein that are of most immediate biological interest. However, it is possible to combine FTIR and UV spectroscopy to accurately determine the UV extinction coefficient at a particular wavelength of a given protein in D_2O solution and, since most protein's secondary structures only change slightly when in D_2O versus ^1H_2O solution, the UV extinction coefficient should be very similar in both cases. Therefore, the protein concentration of subsequent samples prepared in ^1H_2O could be accurately determined from their UV spectra using the D_2O sample extinction coefficient.

While the concentration calibration curve calculated above contains an inherent error from the QAA analysis that was used to determine the concentrations of the stock solutions of the three proteins used, the practice of using FTIR to calculate protein concentrations need not be limited in this way. For example, it is possible to purchase proteins in lyophilised form that are almost 100 % pure and, from these, stock solutions of known concentration could be created and used to generate a concentration calibration curve. Therefore, in theory FTIR should be able to offer an improvement over QAA analysis for protein concentration determination. In the interests of accuracy it may be desirable to obtain FTIR spectra over a wider range of proteins for a more comprehensively averaged concentration calibration curve.

References

1 Scheding S, Dissanayake G, Nebot EM, Durrant-Whyte H. An experiment in autonomous navigation of an underground mining vehicle. Ieee Transactions on Robotics and Automation 15(1):85-95, 1999.

2 Baenziger JE, Methot N. Fourier-Transform Infrared and Hydrogen-Deuterium Exchange Reveal an Exchange-Resistant Core of Alpha-Helical Peptide Hydrogens in the Nicotinic Acetylcholine-Receptor. Journal of Biological Chemistry 270(49):29129-29137, 1995.

3 Sethuraman A, Belfort G. Protein structural perturbation and aggregation on homogeneous surfaces. Biophysical Journal 88(2):1322-1333, 2005.

4 Bradford MM. Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding. Analytical Biochemistry 72(1-2):248-254, 1976.

5 Goa J. A Micro Biuret Method for Protein Determination - Determination of Total Protein in Cerebrospinal Fluid. Scandinavian Journal of Clinical & Laboratory Investigation 5(3):218-222, 1953.

6 Lang CA. Simple Microdetermination of Kjeldahl Nitrogen in Biological Materials. Analytical Chemistry 30(10):1692-1694, 1958.

7 Greenfield NJ. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. Analytical Biochemistry 235(1):1-10, 1996.

Chapter VII

Spectroscopic Investigations into the Structure of AavLEA1

VII.1 Introduction

One of the most basic tenants of biology is that water is essential for all living organisms. It has long been known, however, that certain organisms are capable of surviving the absence of water by entering into a dormant phase for an indefinite period of time and, upon contact with water, they can reactivate and continue along their lifecycles. Such an ability of suspended animation is known as anhydrobiosis and organisms that are capable of this are said to be anhydrobiotic. Anhydrobiotic organisms have been discovered in all of the biological kingdoms and a familiar example is that of a plant seed, which requires the presence of water for germination.[2-4] Less familiar and more surprising anhydrobiotic organisms are to be found within the animal kingdom in organisms as complex as nematodes, one such example being *Aphelenchus avenae*. Over the past fifteen years there has been much research into anhydrobiosis and in particular into nematode anhydrobiosis.

At the sub-cellular level there are many problems associated with anhydrobiosis. Within the cell all of the organelles and other components are suspended in a water matrix. The entire structural integrity of the cell is reliant on the positive pressure exerted by its water content on the semi-permeable outer cell membrane. At the molecular level still more complications are associated with anhydrobiosis. Because water is a polar solvent molecules that are soluble in water must also be somewhat hydrophilic. If the water is removed from within the cell the likelihood of dissolved molecules interacting with each other electrostatically would be considerable, since these molecules would be unstable with respect to charge. Also, the chances of them coming into contact with each other would be increased by decreasing the cell water content.

The molecular structure of many macromolecules depends to a large extent on the solvent shell surrounding the molecule and this is especially the case when the molecule contains charged groups. Where molecular structure is essential to the cellular function of the molecule, for example, as occurs with most proteins, this represents a serious obstacle for anhydrobiosis. This is particularly the case, since electrostatic interactions tend to be strong in nature and, therefore, tend to be irreversible. In this situation rehydration of a desiccated cell would not be sufficient to restore the molecular structures of molecules that have been distorted by the removal of their solvent shells and subsequent bonding of their various charged regions.

It is known that cells in preparation for entering into a desiccated state greatly increase their concentrations of disaccharides.[3; 5] Plant cells accumulate sucrose, while animal cells generally increase their levels of trehalose. It has been theorised that the function of high concentrations of these disaccharides is to form a glassy matrix that might replace the lost water matrix and, thereby, retain the structural integrity of the cell while distancing molecules from one another within the cell.[6; 7] If the primary condition for anhydrobiosis is the preservation of the cell's protein molecules upon dehydration and rehydration, since neither trehalose nor sucrose alone appear to directly stabilise protein molecules against the effects of desiccation, it would seem that any possible protein stabilising effect must be indirect, such as the formation of a glassy matrix. That both trehalose and sucrose readily form stable glasses upon dehydration when present at concentrations found in predesiccation cells of anhydrobiotic organisms highlights the possible importance of the production of such glasses to anhydrobiosis. The glassy matrix mechanism is now widely accepted as being the most likely mechanism through which anhydrobiosis becomes possible.

Important in the anhydrobiotic process is the upregulation of certain genes in anhydrobiotic organisms in situations of water-stress. Since glass formation can occur irrespective of cellular protein content, where sufficiently high disaccharide levels exist, these anhydrobiotic proteins would seem to be providing some function other than glass formation to the desiccating cell. Common features of many anhydrobiotic proteins are that they contain large amounts of charged residues and that they have no distinct secondary structure. They also generally have high pI values and so exist as charged molecules in aqueous solution.[8]

An effect of desiccation is to remove from intracellular molecules their water solvent shells and, as discussed above, this exposes such molecules to damaging effects. If some way could be found to re-establish molecular solvent shells it would likely be of tremendous benefit towards facilitating anhydrobiosis. Possible candidates for replacing macromolecular solvent shells are the anhydrobiotic proteins. These proteins display a large degree of plasticity, since they are without rigid secondary structure, and they also exhibit the necessary polarity to be able to electrostatically interact with other soluble polar molecules when forming the solvent shell. If anhydrobiotic proteins were to provide this function they would fit in nicely with the glassy matrix model of the anhydrobiotic mechanism, in that the two functions of water, as providing a cellular matrix maintaining the cell's structure and a solvent shell that maintains the structure of intracellular molecules, can be replaced by anhydrobiotic proteins surrounding important, previously dissolved macromolecules, where both are contained within a sugar glass matrix.

In studying the nematode *Aphelenchus avenae* it was discovered that the anhydrobiotic protein AavLEA1 was produced by the upregulation of the *Aav-lea-I* gene upon desiccation.[9] The 18 kDa 143 residue AavLEA1 protein has been classified as belonging to the group 3 of a class of proteins known as the late embryogenesis abundant (LEA) proteins and its primary sequence is given in Figure VII.1. These proteins are relatively small, having low sequence complexity and containing the repeat consensus 11-mer TAE/QAAKE/QKAXE in plants, or more broadly ΦΦE/QXΦKE/QKΦXE/D/Q (where Φ and X represent a hydrophobic and random residue, respectively), along with their being over-expressed in the late embryonic phase of maturing seeds or in other desiccating environments. In contrast, group 2 LEA proteins have repeats of either or both of DEYGNP and EEKK, with most containing a poly-serine stutter.

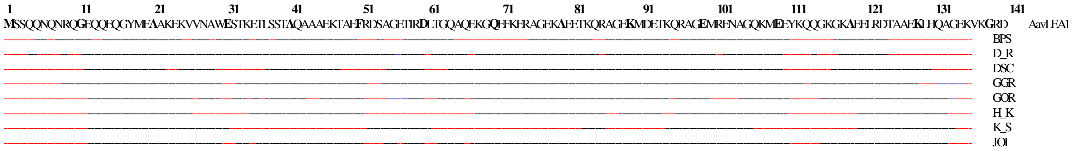


Figure VII.1. AavLEA1 secondary structure predictions. Beneath the AavLEA1 primary sequence are given the different secondary structure predictions according to various predicting programs, which are labelled with initials to the left. α-helix, random coil and β-sheet are represented as black, red and blue dashes, respectively. The secondary structure prediction listed JOI is a "winner takes all" sum of the 7 predictions above it (Modified from [1])

In general, Group 3 LEA proteins have yet to be well characterised in terms of their secondary structures. Figure VII.1 shows the results of seven different programs used to predict the secondary structure of AavLEA1 from its primary sequence. In contrast to the general understanding of LEA proteins being predominantly randomly coiled, the bottom "winner takes all" prediction of AavLEA1 predicts it to be predominantly α-helical in nature. Across all seven programs a large α-helical content is predicted for AavLEA1. According to this evidence, AavLEA1 would seem an unlikely candidate to replace the solvent shells of dissolved proteins, since it would lack the necessary plasticity required to carry out this function.

Dure has shown by using a computer modelling program that group 3 LEA proteins are predicted to adopt a right-handed dimeric coiled-coil structure that have a periodicity defined by the 11-mer repeat motif that characterises this.[10] Such coiled-coils have been observed experimentally in a surface layer protein from *Staphylothermus marinus*.[11] The MultiCoil program was used to predict the structure of a Group 3 LEA-like wheat protein and returned a trimeric coiled-coil structure.[12] In the case of AavLEA1, the MultiCoil program predicts a 40 % probability of coiled-coils, although in this case the structure is predicted to be the more usual left-handed orientation.[1] From this evidence it is possible that in the absence of water AavLEA1 might form a network of reinforcing fibrils to the sugar glass matrix, thus giving it an added stability.

A D-7 LEA protein from pollen has been shown by Wolkers *et al.* to stabilize sugar glasses in vitro by means of FTIR spectroscopy.[13] By monitoring the wavelength maximum of the OH stretching band at around 3360 cm^{-1} it is possible determine the transition temperature (T_g) at which the glass goes from a solid phase to a liquid phase. The T_g temperature can then be used as an indicator as to the density of a sugar glass. It was observed that a glass produced by drying solutions of sucrose, 0.5 mg D-7 LEA/ mg sucrose, and 0.8 mg D-7 LEA/ mg sucrose had T_g temperatures of 60, 69 and 79 ^{o}C, respectively. Therefore, the conclusion to be made is that D-7 LEA protein produces denser and more stable sucrose glasses than sucrose alone and the stabilising effect of D-7 LEA increases with increased relative concentration to sucrose. This data suggests that LEA proteins may indeed form reinforcing coiled-coil fibril structures. In addition to this, gel electrophoresis experiments reveal that AavLEA1 readily produces both dimer and trimer bands, in addition to the more intense monomer band, and this is consistent with the formation of both dimeric and trimeric coiled-coil fibrils.[1]

In contrast to both the structural results of the prediction programs and any evidence suggesting that AavLEA1 might form coiled-coil structures in the presence of water are the results of hydrodynamic experiments performed on AavLEA1.[1] A single boundary was observed across a range of sample protein concentrations, suggesting that there is only one species present – that of the monomer. A sedimentation coefficient of 1.20 for AavLEA1 indicates unusually high drag forces for an 18 kDa molecule. The frictional ratio of 2.28 was calculated for AavLEA1, which suggests either that AavLEA1 exists as an unstructured, highly hydrated swollen protein or that it has a highly extended structure and this is evidence that AavLEA1 is more likely to function in solvent shell replacement rather than as a reinforcement to sugar glass matrices.

In view of the uncertainty concerning the actual structure of AavLEA1, a comprehensive investigation into the structure of the protein was carried out using CD, FTIR and luminescence spectroscopic techniques.

VII.2 Method

VII.2.1 Protein Expression and Purification

The *Aav-lea-1* sequence was amplified by polymerase chain reaction (PCR) using oligonucleotides containing engineered NdeI and BamHI restriction sites. The PCR product was cloned into a pCR2.1-TOPO vector (Invitrogen), cut out with the appropriate restriction enzymes and ligated into the pET15b vector (Novagen). The plasmid was transformed into *E. coli* BL21(DE3) cells. His-tagged recombinant AavLEA1 protein was produced in BL21(DE3) cells in the following manner. A single bacterial colony was used to inoculate 3 ml of Luria Bertani (LB) medium containing 50 μg/ml carbenicillin (Sigma) and grown at 37 ^{o}C and 250 rpm until the OD_{600} reached 0.5. This culture was then used to inoculate 100 ml of LB plus antibiotic (50 μg/ml carbenicillin). At an OD_{600} of 0.6 the culture was split and 50 ml was induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) (Sigma) for 4 h at 37 ^{o}C. Cells were harvested by centrifugation at 10,000 g for 10 min at 4 ^{o}C. Bacterial pellets were lysed with Bugbuster extraction reagent (Novagen) and 1KU/ml rLyzozyme (Novagen). The cell suspension was incubated at room temperature for 20 min and centrifuged at 16,000 g for 20 min at 4 ^{o}C. The supernatant was filtered through a 0.45 µM filter (Sarstedt AG & Co.) before being mixed with Ni-NTA bind slurry (Novagen) with gentle shaking at 4 °C for 60 min. The mixture was poured into a His·Bind column (Novagen) and washed with Ni-NTA wash buffer (50 mM sodium phosphate buffer, 300 mM NaCl, 20 mM imidazole). The AavLEA1 His-tagged recombinant protein was removed from the column by adding an elution buffer and collected as a separate fraction. The purified fraction was analysed by SDS-PAGE gel electrophoresis with Coomassie blue staining. A buffer exchange was performed by passing the purified fraction through a PD-10 size-exclusion column (Sigma) and eluting with 10 mM PBS buffer solution.

Site directed mutagenesis (SDM) of the pET15b plasmid containing the *Aav-lea-1* gene was carried out using an SDM kit (QuikChange ™ kit, Stratagene) to produce two mutated strains: L37C and L132C *AavLEA1*. The plasmids containing these genes were transformed into *E. coli* BL21(DE3) cells and protein expression and purification were carried out in the same manner as that of the wild-type AavLEA1.

VII.2.2 Protein Fluorescent Labelling

The mutated AavLEA1 proteins were labelled by adding molar equivalents of the IAEDANS or CPM fluorophores and allowing the protein solution to sit at 4 °C overnight. In the case of CPM, the CPM had to be first dissolved in 100 µl of dimethyl sulfoxide (Fluka) before addition to the protein solution. Unbound fluorophore was removed using size-exclusion chromatography (Sephadex G-25 spin columns (Sigma)) and repeat separations were performed when necessary.

VII.2.3 CD spectroscopy

All CD spectra were acquired on an Aviv Model 215 CD spectrometer. Spectra were acquired in quartz cuvettes of pathlengths of either 0.1, 1 or 10 mm and with spectral parameters of 1 nm bandwidth, at least 1 S averaging time, a stepwidth of between 0.2 and 1 nm, as an average of at least 3 scans and at a temperature 23 °C. Each spectrum was background corrected by subtracting the spectrum of the buffer solution as acquired in the same cuvette, held in the same orientation and scanned using identical spectral parameters. For pH, and titration experiments spectra were acquired while the sample was continually stirred at a quarter the instruments maximum stir rate, by means of a stir-bar placed within the cuvette. The concentration of the AavLEA1 sample solution was determined using quantitative amino acid analysis (QAAA) (Cambridge, Peter Sharratt) and the analysis was checked for reproducibility. The CD spectrum of AavLEA1 was then plotted in terms of molar ellipticity.

The AavLEA1 dependence on pH was determined by adding volumes of no greater than 10 μl of either standard solutions of HCl or NaOH of varying concentrations between measurements to the sample solution that was contained within a 10 mm quartz cuvette by using a 10 μl adjustable pipette. The pH was recorded after each addition using a microelectrode coupled to a pH meter that was inserted into the cuvette and out of the beam path during each measurement. Each pH dependence experiment was performed twice as a check for reproducibility.

The AavLEA1 temperature dependence experiment was performed using a 1 mm pathlength cuvette. A temperature equilibration time of 1 minute was set once the desired temperature had been reached and temperatures of 20, 40, 55, 75 and 95 ^{o}C were investigated. The ellipticity at 222 nm was taken and used to measure the relative helicity of AavLEA1 at each temperature.

The effect of cation concentration on the CD spectrum of AavLEA1 was measured by titrating a sample of AavLEA1 within a 10 mm cuvette with a standard solution of a 100 mM $AlCl_3$. The titration was performed by adding the titrant in 3 μl additions and a total of 8 such additions were made over the course of the experiment. An equilibration time of 1 minute was allowed under continuous stirring between each addition before spectral acquisition. The relative helicity of AavLEA1 over the course of the experiment was determined by plotting the ellipticity as a function of $AlCl_3$ concentration. The experiment was run twice to check for reproducibility.

The effect of trifluoroethanol (TFE) (Sigma) to the secondary structure of AavLEA1 was examined by recording the spectra of AavLEA1 solutions of uniform approximate concentration of 2.75 mg/ml in volume per volume TFE concentrations of 15, 30, 37 and 45 %. The CD spectrum of each solution was recorded using a cuvette of pathlength 0.1 mm and then calculated in terms of its molar ellipticity. The AavLEA1 helix formation with TFE concentration was determined by plotting the ellipticity at 222 nm against TFE concentration.

VII.2.4 FTIR spectroscopy

FTIR spectra were acquired on a Tensor 27 FTIR spectrometer equipped with a mid-IR nitrogen-cooled MCT detector. AavLEA1 spectra were acquired as ATR-FTIR transmission spectra, using the temperature regulated BioATR-II accessory and then converted to absorption spectra by subtracting the transmission spectra of the buffer solution, as acquired under identical spectral parameters as that of the sample. Both reference and sample spectra were acquired in single-sided mode, as an average of 100 scans, over a frequency range of 4000-1000 cm^{-1}, with a Norton-Beer apodization function, a zero-filling factor of 8, at a resolution of 4 cm^{-1}, using an aperture size of 6 mm and held at a constant temperature. The N_2 MCT detector was allowed to cool for at least 45 minutes before scanning and the sample compartment and BioATR-II accessory were continually purged with dry air. All solutions were thoroughly degassed prior to being placed within the BioATR-II accessory. Protein kinetic measurements were performed at a temperature of 54 ^{o}C with a time interval of 5 minutes between scans and over a duration of 8 h. Reverse protein melt experiments were performed over a temperature range of 20-90-20 ^{o}C, at 5 ^{o}C increments and with an equilibration time of 2 minutes for each temperature increment.

Post processing of spectral data included atmospheric correction for any water vapour vibrational bands appearing in the absorption spectrum. Also, a computer algorithm included in the OPUS software was used to correct the absorption spectra for an effect arising from a small difference in water concentration between the sample and buffer solutions. Spectra were offset corrected at a wavenumber between 1700 and 1750 cm^{-1} and then normalised according to the amide-I band area. The second derivative of absorption spectra was calculated as a band narrowing technique and a mild band smoothing procedure accompanied this procedure. Between measurements the BioATR-II silicon wafer surface was thoroughly cleaned using a 3 M guanidine HCl solution alternated with deionised water rinsing for a total of three cycles. The cleanness of the silicon surface was inspected prior to scanning and the absence of any residual adsorbed protein from previous scans was confirmed by acquiring the spectrum of the empty BioATR-II sample region after cleaning and making sure that no CH stretching bands were present between 3000 and 2800 cm^{-1}.

VII.2.5 Luminescence Spectroscopy

Absorption spectra were acquired using a Spectronic Array 3000 (Milton Roy) absorption spectrometer and a 1 cm quartz cuvette. Steady-state and time-resolved luminescence spectra were acquired using a pulsed laser source (a TDL-90 dye laser (equipped with a rhodamine 590 dye (Exciton)) pumped by a YG-980E Nd:YAG solid-state laser (Quantel)), a 0.3 metre SpectraPro®-300i monochromator (Acton Research) equipped with diffraction gratings, having 150 grooves and blazed at 300 and 500 nm, and an intensified-gated charge coupled device (iCCD) (Andor). The excitation wavelength of 282.5 nm was achieved by frequency-doubling of the output of Rodamine 590 emission of 565 nm. Steady-state emission spectra were acquired using a gate-pulse delay of 74 ns, a gate-pulse width of 30 ns, an exposure time of 0.017 s and as an average of 50 scans. The same spectral parameters were used for collecting time-resolved emission spectra, except that the gate width was reduced to 3 ns. Excitation spectra were acquired using a tungsten/deuterium light source (Acton Research) and by passing the exciting radiation through a SpectraPro® 0.3 metre monochromator (primary) controlled by an NCL spectral measurement system (Acton Research). Radiation was channelled into a sampling compartment (Acton Research) containing a 1 cm quartz cell in which the sample was placed. A SpectraPro® 0.5 metre monochromator (secondary) was positioned such that it collected radiation emitted by the sample at right angles from the excitation beam, which had passed through a focusing lens. This monochromator was fixed at the emission radiation wavelength of interest and radiation was detected using a single photon counting R928-P photon multiplier tube (Hamamatsu). Slit-widths of 2 mm were set for both the primary and secondary monochromators. A stepwidth of 0.1 nm and an integration time of 500 ms were used during acquisition of excitation spectra.

In designing a FRET experiment considerable attention must be paid to selecting appropriate donor and acceptor fluorophores. A number of spectroscopic requirements that must be satisfied for FRET to be possible. The most immediate of these is that there should exist extensive overlap between the emission band of the donor molecule and the absorption band of the acceptor. It is desirable that both the donor and acceptor molecules exhibit a large stokes-shift such that the spectral features of each are clearly resolved. Ideally, there should occur minimal overlap between the donor and acceptor absorption bands, as this allows for the selective excitation of the donor molecules and, consequently, the detection of FRET is much simplified as simply any emission intensity of the acceptor molecule resulting from excitation at a wavelength of donor absorption. Where overlap does occur between the donor and acceptor absorption bands it is necessary to use either excitation or time-resolved spectroscopy to determine whether FRET has occurred. For the range of fluorophores available for use in biological systems this is most commonly the case and the donor-acceptor pairs used in this experiment do have some absorption band overlap. An additional consideration for FRET systems is that both fluorophores should have high extinction coefficients and quantum yields if the occurrence of FRET is to be detectable. Lastly, the Forster distance parameter defining the distance over which FRET is possible for certain donor-acceptor combinations should be on a scale suited for the experiment in question.

The usual way to site-specifically label a protein with a fluorescent tag is to make use of the thiol group of cysteine residues by attaching a thiol-specific acceptor tag, as represented in Figure VII.2 for the CPM and IAEDANS fluorophores. However, the AavLEA1 protein does not contain any cysteine residues and so it was necessary to perform SDM to prime the protein for labelling with the fluorophores given in Figure VII.2. Although this involves extra work, the end product of a protein with a single tryptophan donor molecule and a single CPM or IAEDANS acceptor molecule is very useful for FRET measurements. In addition the locations of both the donor and acceptor molecules on the protein would be precisely known and the distance between these could be varied by introducing the cysteine mutation at different points on the AavLEA1 protein.

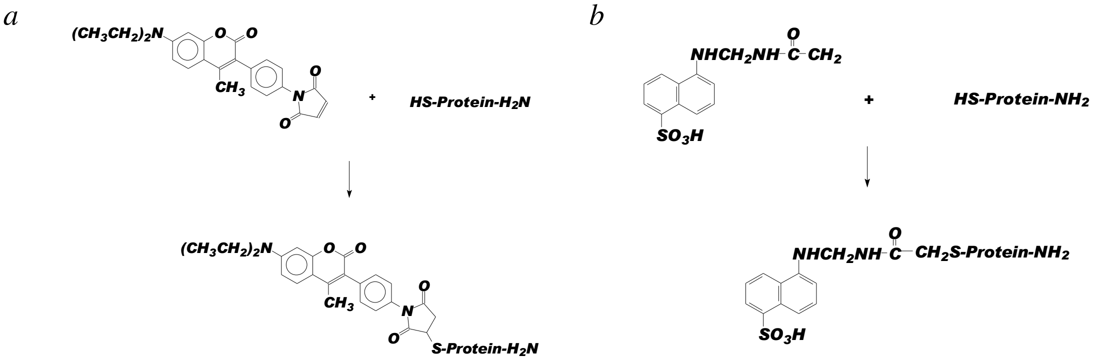


Figure VII.2. Protein site-specific fluorescent labelling. The reaction schemes for labelling of a protein with CPM and IAEDANS thiol-specific fluorophores is shown in *a* and *b*, respectively.

VII.3 Results

VII.3.1 Protein purity

Figure VII.3 shows an SDS-PAGE gel of a batch synthesis of wild-type AavLEA1. The purified AavLEA1 fraction was run in lane 3, in which there can be seen an intense band at approximately 18 kDa, according to the ladder in lane 1. This is as expected for the AavLEA1 protein and is taken as confirmation of a successful protein synthesis. Upon close inspection very faint additional bands can be seen in lane 3 at higher molecular weights of Figure VII.3, suggesting the possibility of very minor contamination of the AavLEA1 sample, but it also possible that these bands are the result of overflow of material from the loading of lane 2. In any event, by protein purification standards, the AavLEA1 sample analysed in Figure VII.3 would be regarded as highly purified. The structural investigation of AavLEA1 required that many batches of protein be produced and Figure VII.3 shows the SDS-PAGE analysis of only a single batch. However, for each batch of either wild-type or mutant AavLEA1 protein a separate gel was run and the protein was only passed fit for structural analysis if the gel was of equal quality to that shown in Figure VII.3.

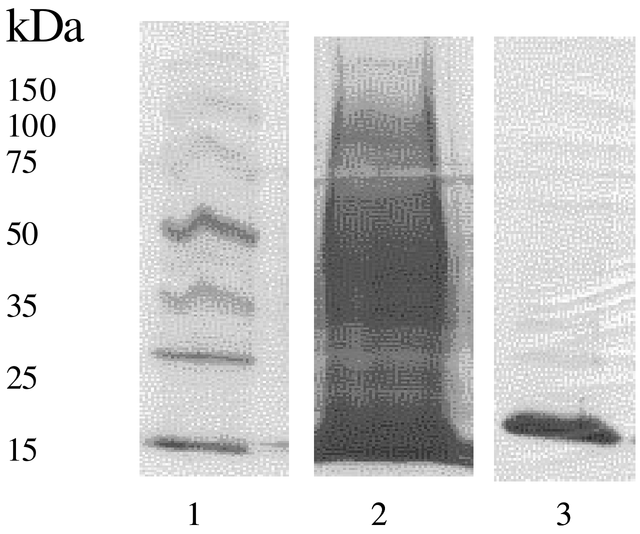


Figure VII.3 SDS-PAGE gel of wild-type AavLEA1. The protein ladder, unpurified cell lysate and purified AavLEA1 fraction were run in lanes 1, 2 and 3, respectively. The sizes of the bands of the protein ladder are listed on the left.

VII.3.2 The Structure of AavLEA1 from CD Spectroscopy

The concentration of the AavLEA1 sample solution that was used to generate the spectrum presented in Figure VII.4 was determined by QAAA to be 67.22 nanomoles/ml, as shown in Table VII.1. The similarity of the results of experiment 1 and 2 to one another suggests that the concentration determination was accurate. The good agreement between the numbers of amino acids found during QAAA and those predicted by the primary sequence of AavLEA1 indicate that the sample was mostly free from biological impurities.

The CD spectrum of AavLEA1 presented in Figure VII.4 is characteristic of a random coil secondary structure protein, with an intense negative band centred at approximately 200 nm and no pronounced negative bands between 210 and 230 nm or positive bands at wavelengths shorter than 195 nm, both of which are an indication for the presence of α-helical or β-sheet ordered secondary structure. The magnitude of the smoothing residuals to lower wavelengths than 195 nm decreases the confidence that can be placed in the accuracy of the CD spectrum in this range, although it is evident from the spectrum that there is an absence of any bands of significant intensity in this wavelength range. The presence of some small spectral intensity at 222 nm suggests that there exists a very minor degree of α-helix content within AavLEA1.

Amino acid	**Found Amino Acids**					
	Expected	**Experiment 1**	**Experiment 2**	**Average**	**Std Dev**	**Closeness of fit**
Cysteine	0	0.00	0.00	0.00	0.00	-
Asparigine	9	8.51	8.49	8.50	0.01	Within 5-10 %
Threonine	7	8.02	8.13	8.07	0.08	>10 % out
Serine	6	8.99	8.97	8.98	0.01	>10 % out
Glutamic acid	42	38.99	39.08	39.04	0.06	Within 5-10 %
Glycine	13	14.26	14.16	14.21	0.07	Within 5-10 %
Alanine	19	17.55	17.41	17.48	0.10	Within 5-10 %
Valine	3	3.43	3.45	3.44	0.01	>10 % out
Methionine	4	4.29	4.26	4.28	0.02	Within 5-10 %
Isoleucine	1	1.07	1.10	1.08	0.02	Within 5-10 %
Leucine	4	4.75	4.78	4.76	0.02	>10 % out
Tyrosine	2	1.85	1.89	1.87	0.03	Within 5-10 %
Phenylalanine	2	1.93	1.93	1.93	0.00	Better than 5 %
Histidine	excluded	-	-	-	-	-
Lysine	17	15.06	15.06	15.06	0.01	>10 % out
Arginine	9	9.30	9.30	9.30	0.00	Better than 5 %
Proline	0	0.00	0.00	0.00	0.00	-
Tryptophan	excluded	-	-	-		
Total sample		**24.20 μg**	**24.06 μg**	**24.13 μg**	**0.10**	
Concentration		67.22 nmoles/ml (1209.92 μg/ml)	66.83 nmoles/ml (1203.02 μg/ml)	**67.03 nmoles/ml (1206.47 μg/ml)**	0.27 (4.88)	

Table VII.1. Concentration determination of AavLEA1 by QAA analysis.

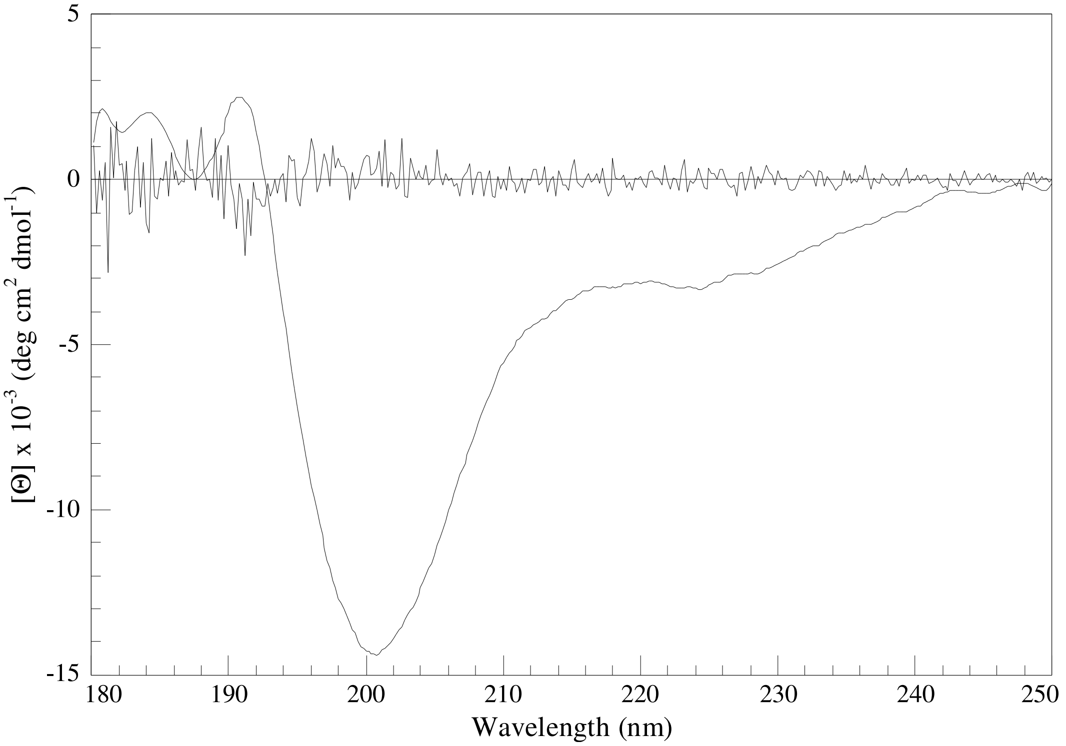


Figure VII.4. The CD spectrum of AavLEA1. The smoothing residuals are overlaid.

VII.3.3 Temperature Dependence of AavLEA1 Secondary Structure from CD

Figure VII.5 shows the AavLEA1 CD spectral overlay when recorded over a range of temperatures. With increasing temperature there occurs an increase in the negative magnitude of the AavLEA1 CD spectrum at 222 nm and a correlated decrease in negative magnitude at the band minimum located at approximately 200 nm. Taken together, this behaviour suggests that with increasing temperature an overall increase in the α-helix content of AavLEA1 occurs accompanied by a decrease in the random coil content.

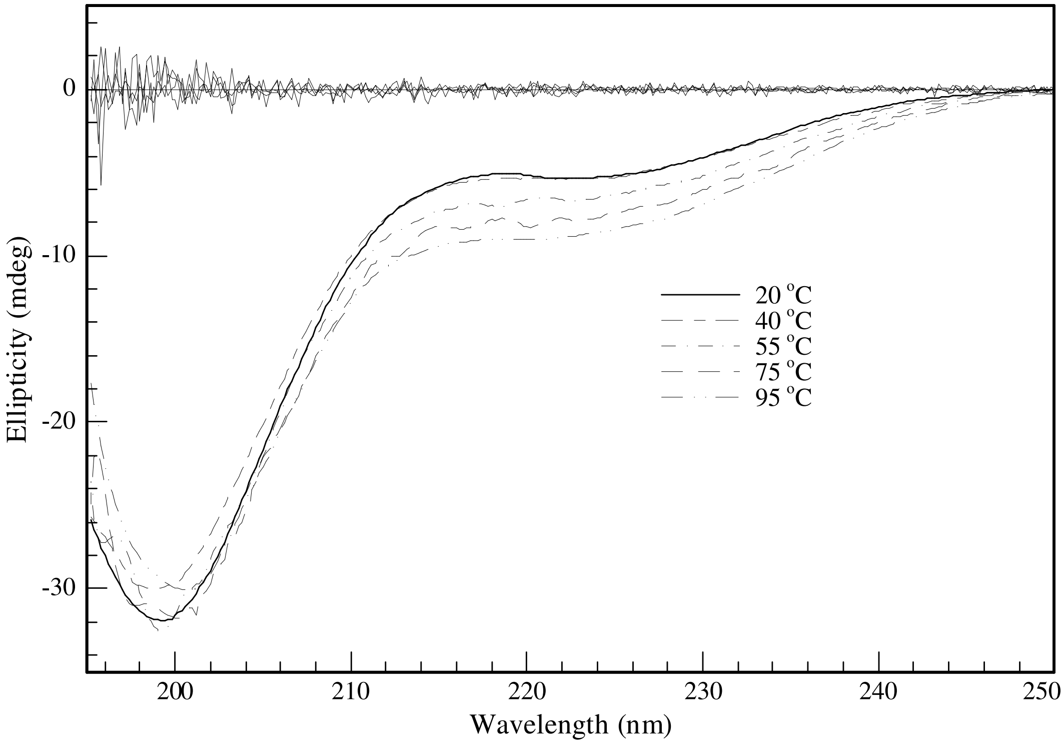


Figure VII.5. AavLEA1 CD spectral overlay with temperature increase. Smoothing residuals are shown as overlaid.

VII.3.4 AavLEA1 Structural Variation in the Presence of Cations

The ability of AavLEA1 to potentially sequester potentially deleterious cations and the effects of any such process on its secondary structure can be seen by observing the AavLEA1 CD in response to cation concentration (Figure VII.6). The AavLEA1 CD is significantly altered upon the introduction of Al^{3+} cations to the sample solution and this change can be seen to depend on the cation concentration. An increase in the negative amplitude of the AavLEA1 CD at 222 nm suggests that AavLEA1 experiences an increase in α-helical content as a function of increased cation concentration. AavLEA1 CD spectra could not be accurately acquired to wavelengths lower than 205 nm when in the presence of $AlCl_3$ due to the strong absorbance of the chloride ions at these wavelengths. Therefore, the expected corresponding decrease in the negative CD amplitude at approximately 200 nm, as random coil protein regions become α-helical in structure, could not be confirmed.

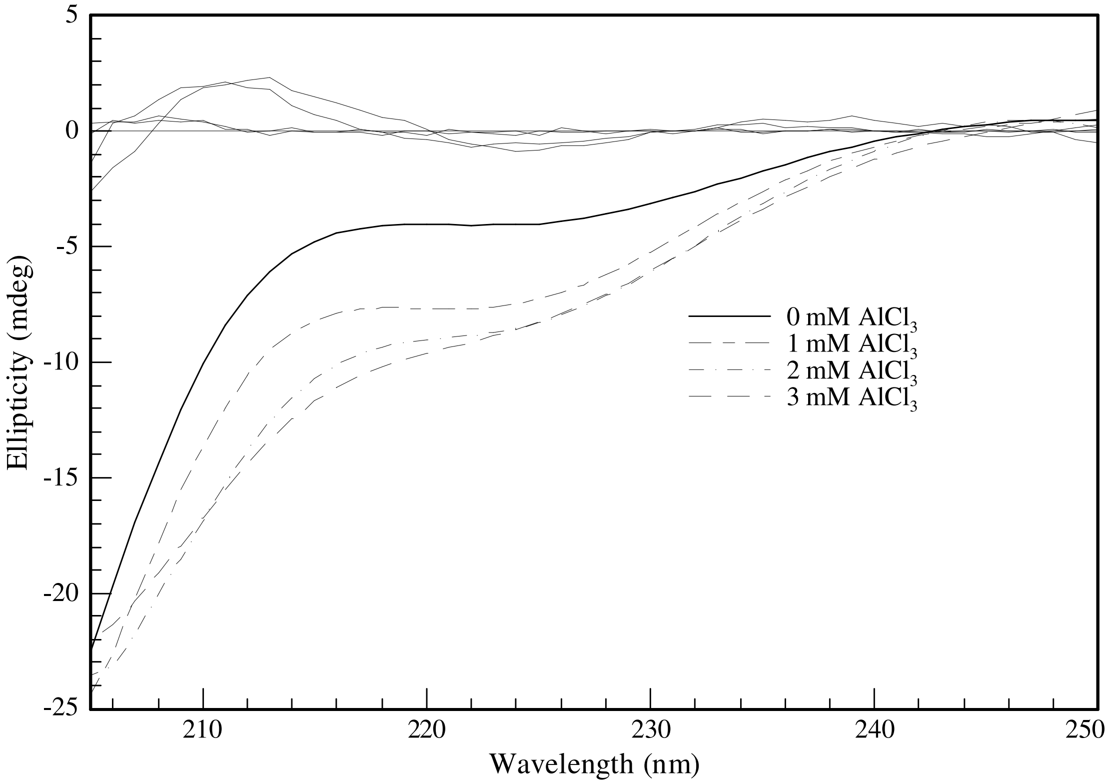


Figure VII.6. AavLEA1 CD with varying cation concentration. Smoothing residuals are shown as overlaid.

VII.3.5 The Potential for AavLEA1 to Adopt Ordered Secondary Structure

The potential for AavLEA1 to adopt significant levels of ordered secondary structure, which would be a likely prerequisite for it to interact with ligands within the cell, and, thereby, carry out its function in desiccation tolerance by virtue of this newly acquired structure was tested by observing the response of the AavLEA1 CD as a function of the helix inducing reagent TFE (Figure VII.7). The increase in the negative AavLEA1 CD amplitude at 208 and 222 nm and the emergence of a positive band centred at 192 nm when the concentration of TFE is increased indicate that AavLEA1 adopts significant α-helix content with the addition of TFE. The ratios of the magnitudes of the bands at 222 and 192 nm indicate that AavLEA1 becomes almost completely α-helical at TFE concentrations of 45 % v/v. This is corroborated by the complete disappearance of the characteristic random coil band centred at approximately 200 nm.

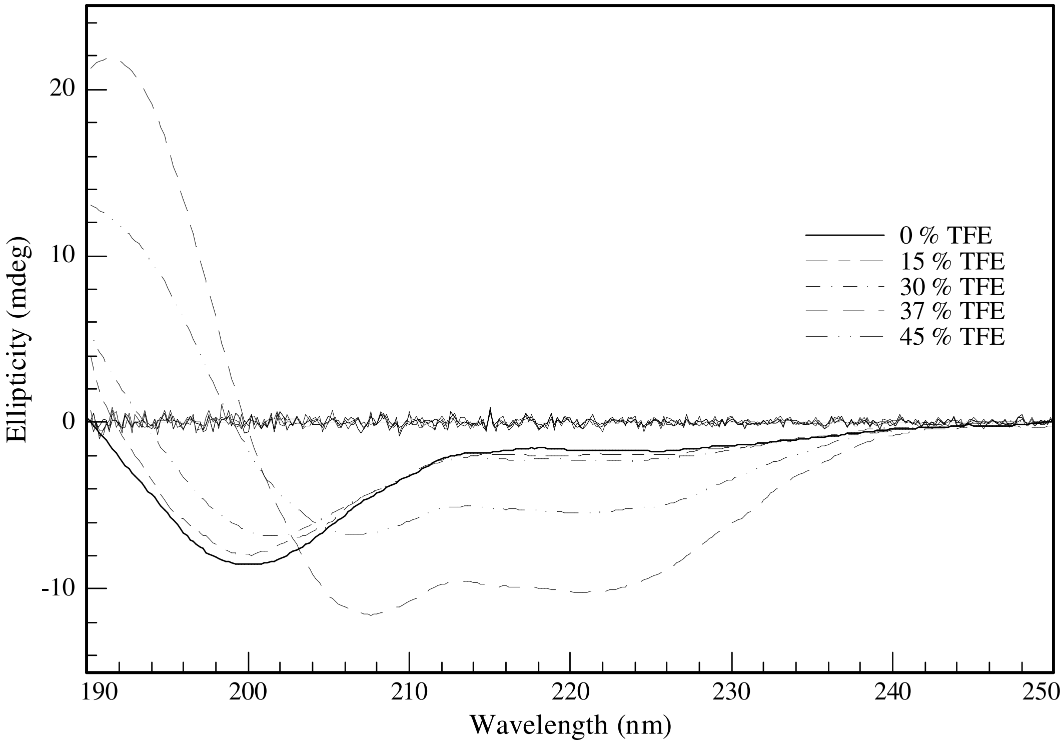


Figure VII.7. The effect of TFE on the secondary structure of AavLEA1. Overlaid are the smoothing residuals for each of the smoothed CD spectra.

VII.3.6 AavLEA1 Secondary Structure Variation with pH

Figure VII.8 summaries the AavLEA1 helix formation with environmental change, as discussed above. In addition, the effect of pH upon the helix content of AavLEA1 is presented. Under acidic conditions AavLEA1 can be seen to increasingly become more α-helical as the pH is reduced, while exposure to basic conditions appears to have no effect on the α-helical content.

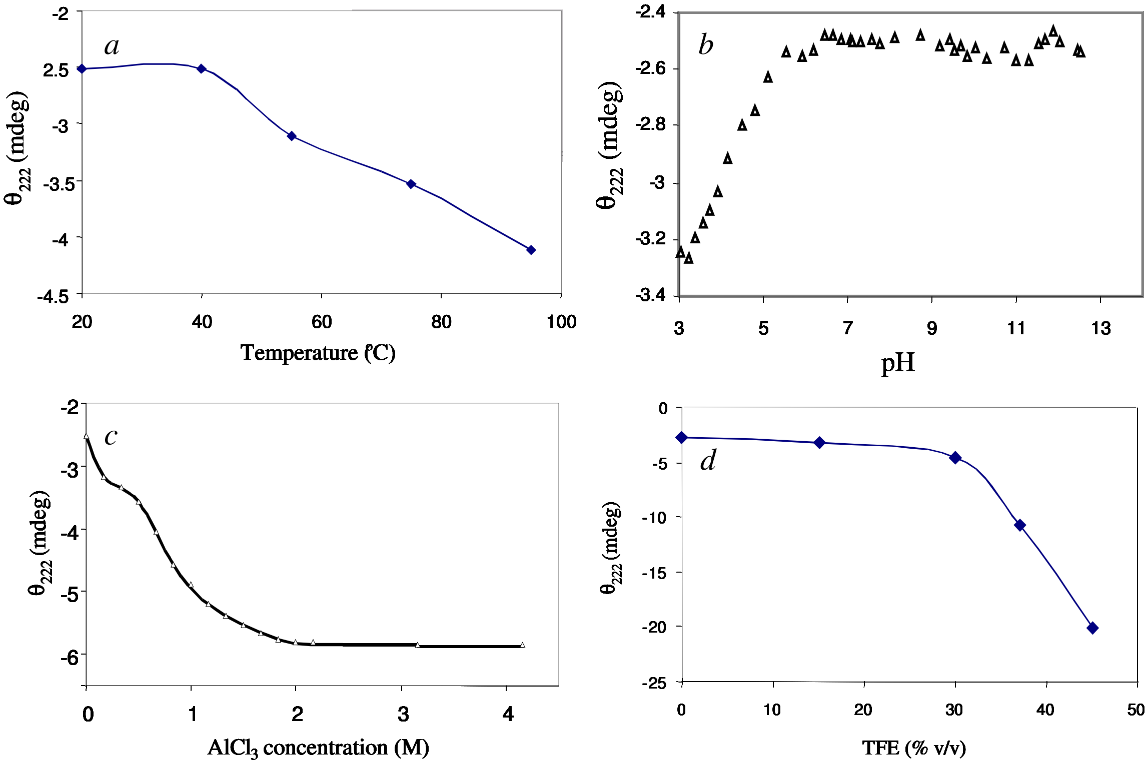


Figure VII.8. The change in helicity of AavLEA1 as a function of environmental change, as measured by CD$_{222}$. The change in the AavLEA1 CD$_{222}$ as a function of temperature, pH, cation concentration and TFE is presented as *a*, *b*, *c* and *d*, respectively.

VII.3.7 Temperature Dependence of AavLEA1 Secondary Structure from FTIR

Figure VII.9 shows the analysis of AavLEA1 by ATR-FTIR spectroscopy. The broad Gaussian-shaped amide-I band centred at approximately 1645 cm^{-1} is evidence that AavLEA1 is a predominantly an unstructured protein (Figure VII.9*a* and *b*). The effect of trehalose on the secondary structure of AavLEA1 is investigated in Figure VII.9*b* and the protein appears to be unaltered in the presence of trehalose. The temperature dependence of AavLEA1 secondary structure is shown in Figure VII.9*c* and *d*, where the temperature was consecutively increased from 20 to 90 ^{o}C and decreased from 90 to 20 ^{o}C, respectively. The amide-I band is expected to experience some degree of red-shift with increasing temperature and blue-shift with decreasing temperature. Notwithstanding this, the degree to which the amide-I band can be seen to shift in Figure VII.9*c* and *d* is significantly greater than what would be expected from the temperature effect. It can also be seen from Figure VII.9*c* and *d* that the amide-I bandwidth decreases with increasing temperature and this is reflected in the amide-I band having a more intense band maximum with increased temperature, since the amide-I band for all temperatures was area-normalised. Both the frequency shift and decrease in bandwidth of the amide-I band in moving to higher temperatures can be explained by AavLEA1 adopting a larger α-helical content with increased temperature, since the band maximum of the helix amide-I is of higher frequency than the random coil amide-I and the bandwidth of the random coil amide-I is larger than that of the helix. From Figure VII.9*d*, it can be seen that this structural shift is fully reversible.

Inspection of Figure VII.9*c* and *d* reveals that AavLEA1 is unusual in that it does not experience any aggregation effect when heated to a temperature of 90 °C. This is confirmed by an absence of an aggregate amide-I band at approximately 1620 cm^{-1} in the higher temperature spectra and the general reversibility of the variation in the amide-I band with temperature. This is convenient for investigating the possible protein anti-aggregation properties of AavLEA1 by ATR-FTIR, since any contribution to the amide-I band by aggregated proteins must arise from the protein whose aggregate formation characteristics are being investigated, and not from AavLEA1.

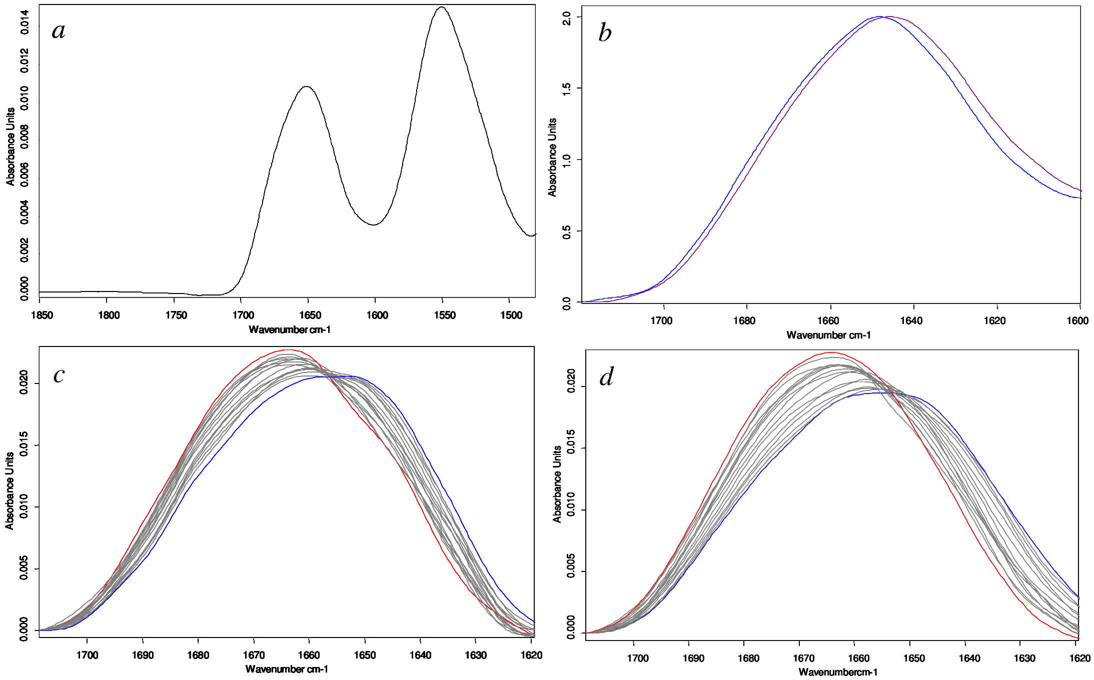


Figure VII.9. ATR FTIR spectral analyses of AavLEA1. *a* AavLEA1 amide-I and -II bands, centred at approximately 1650 and 1540 cm^{-1}, respectively. *b* Overlay of area-normalised, offset-corrected (1720 cm^{-1}) AavLEA1 amide-I band, both with (black) and without (blue) 0.375 M trehalose. *b* and *d* Offset-corrected (1720 cm^{-1}) AavLEA1 amide-I band overlay as recorded consecutively over temperature ramps of 20-90 and 90-20 °C, respectively, and with a ΔT of 5 °C. Bands recorded at 20 and 90 °C are shown in blue and red, respectively. Spectra presented above were acquired using a protein concentration of approximately 3 mg/ml.

VII.3.8 AavLEA1 Antagonism of Heat-Induced Aggregation of Cytochrome c

The possible stabilising effect that AavLEA1 might have on other proteins was investigated using a host of standard proteins by heating them both in the presence and absence of AavLEA1 and comparing the aggregation profiles in both cases. AavLEA1 was shown to confer no added stability to myoglobin, hemoglobin, lysozyme and α-chymotrypsin when subjected to FTIR melt experiments (data not shown), either alone or together with trehalose. The presence of trehalose alone also had no effect on the stabilities of these proteins. This is somewhat in contrast to the findings of Goyal *et al.*, who found that AavLEA1 and trehalose acted synergistically to prevent the heat-induced aggregation of citrate synthase.[14] However, in agreement with this study were the results of the aggregation behaviour of cytochrome c when heated to a temperature of 54 ^{o}C (Figures VII.10 and VII.11). Figure VII.10 shows the amide-I absorption spectra of cytochrome c when in the presence and absence of both trehalose and AavLEA1, alone and together. The spectra are coloured according to the time at which they were acquired, with the chronological sequence being blue, pink, tan, green, orange and black.

Figure VII.11 shows the second derivative analysis of the spectra shown in Figure VII.10. From a band at 1620 cm^{-1} in Figure VII.11*a* it can be seen that cytochrome c suffers a moderate and gradual degree of aggregation at 54 ^{o}C. This same band evolution can be seen in Figure VII.11*b* and, in contrast to the findings of Goyal *et al.* this demonstrates the inability of trehalose alone to retard protein aggregation at increased temperatures. Figure VII.11*c* shows a possible reduction in the intensity of the emerging aggregation band and is evidence that AavLEA1 alone may have some small stabilising influence on other proteins with respect to temperature-induced aggregation. When both AavLEA1 and trehalose are added to the cytochrome c sample, however, the emergence of an aggregation band at 1620 cm^{-1} is almost completely suppressed at 54 ^{o}C (Figure VII.11*d*) and this is in agreement with the findings of Goyal *et al.* and their study of the aggregation profile of citrate synthase.

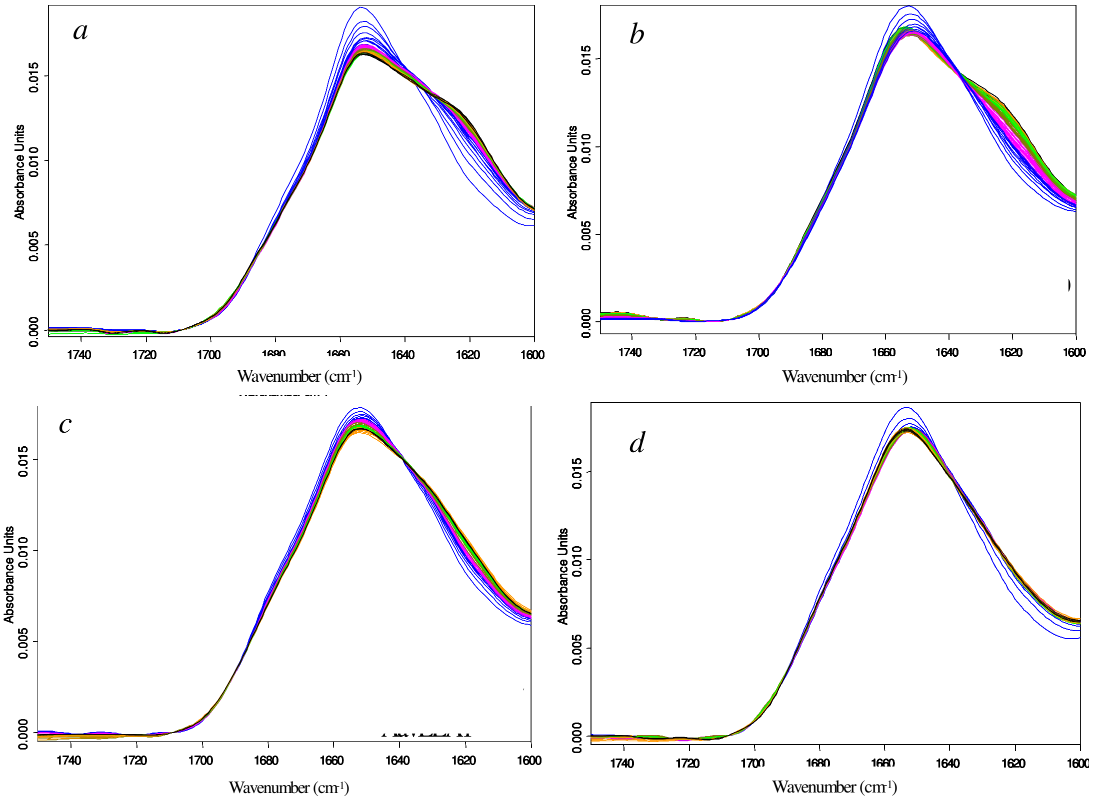


Figure VII.10. The aggregation process of cytochrome c at 54 °C measured by ATR-FTIR. Spectra were recorded at intervals of 5 minutes over a total time of 8 hours. Cytochrome c in 10 mM PBS buffer aggregation profile was recorded alone, in the presence of 500 mM trehalose, with an equimolar AavLEA1 sample and with both an equimolar AavLEA1 solution and in the presence of 500 mM trehalose, the results of which are shown in spectral overlay as *a*, *b*, *c* and *d*, respectively. The concentration of cytochrome c was kept constant at approximately 10 mg/ml over the four different samples. The times at which spectra were recorded during each of the kinetic experiments are colour coded, with the start to finish sequence being blue, pink, tan, green, orange and black.

The helix stabilising properties of AavLEA1 and trehalose for the case of cytochrome c can be understood by examining the intensity variations of the α-helix band at approximately 1655 cm^{-1} as a function of time at 54 °C. Figure VII.11*a* shows a gradual loss of helix content with time when cytochrome c is held at 54 °C. In contrast, Figure VII.11*b* shows an initial loss of helicity followed by a subsequent gain in the helix content of cytochrome c. This is perhaps evidence for a role of trehalose in providing a helix-stabilising effect for proteins at increased temperature. The effect of introducing AavLEA1 on its own to the cytochrome c sample was also to suppress the level of helix degradation (Figure VII.11*c*). Where both AavLEA1 and trehalose were added (Figure VII.11*d*) the helix content varied in a manner similar to the case were trehalose alone was added, in that there was an initial loss of helix content followed by a subsequent gain.

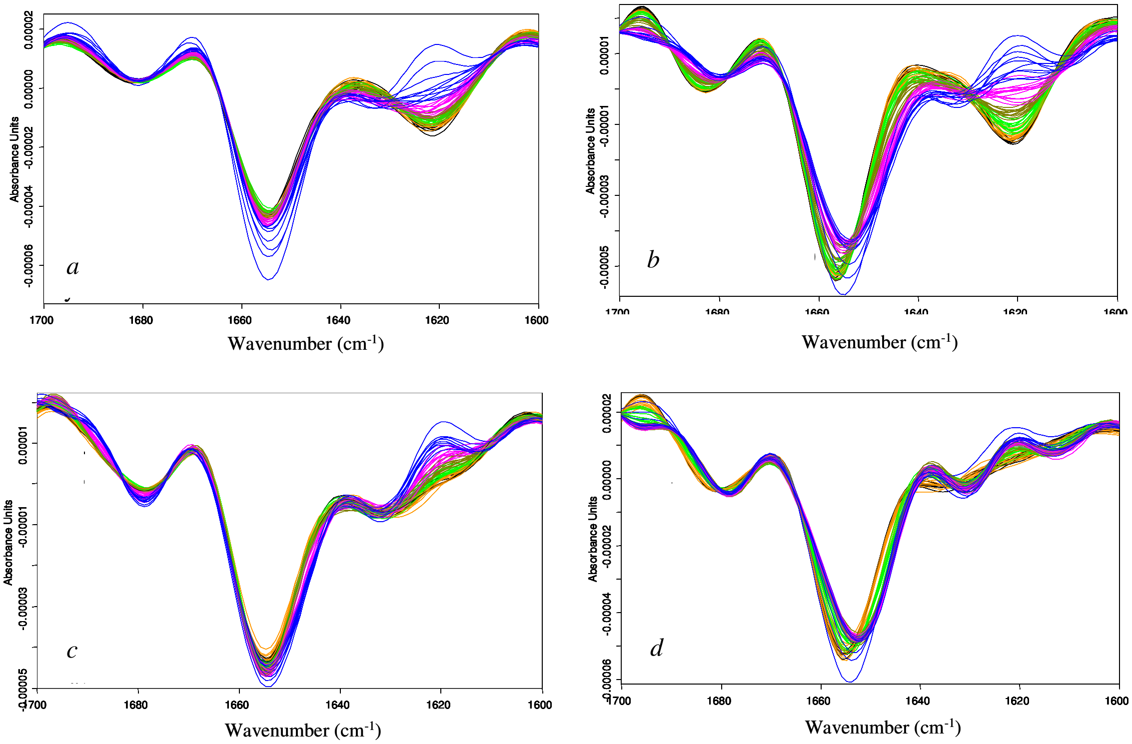


Figure VII.11. Cytochrome c aggregation analysis by 2nd derivative ATR-FTIR spectroscopy. The 2nd derivatives of the respective spectra presented in Figure VII.9 are shown as *a*, *b*, *c* and *d*. The time-colour correlation is as given in Figure VII.9.

VII.3.9 FRET for Fluorescently Labelled BSA by Steady-State Luminescence Spectroscopy

Before proceeding with the SDM of AavLEA1 it was first desirable to check whether FRET could be demonstrated for the commercial and inexpensive BSA protein using the IAEDANS and CPM fluorophores and the spectroscopic instrumentation to be used in the investigation of AavLEA1. BSA was a good candidate for this validation procedure because it is abundant in both tryptophan and cysteine residues and, as such, should exhibit FRET when fluorescently labelled. Figure VII.12 reveals the extent to which BSA's luminescent properties are determined by the tryptophan residues it contains. The suitability of the IAEDANS and CPM fluorophores for energy transfer from tryptophan is revealed by the steady-state luminescence profiles of BSA, IAEDANS and CPM, and these are given in Figures VII.12, VII.13 and VII.14, respectively. More explicitly, the overlay of the emission bands of both BSA and AavLEA1 with the absorption bands of IAEDANS and CPM show the degree of donor-emission/acceptor-absorption band overlap (Figure VII.13). From Figures VII.13 and VII.14 it can be seen that both IAEDANS and CPM both have a minimum of absorption at around 280 nm. Therefore, exciting a labelled protein with radiation around 280 nm should preferentially, but not specifically, excite the tryptophan donor molecules. Crucially, although a small number of acceptor molecules will experience direct excitation when irradiating with 280 nm light, most will remain in their ground-states and be available to receive energy from the excited tryptophan molecules via FRET.

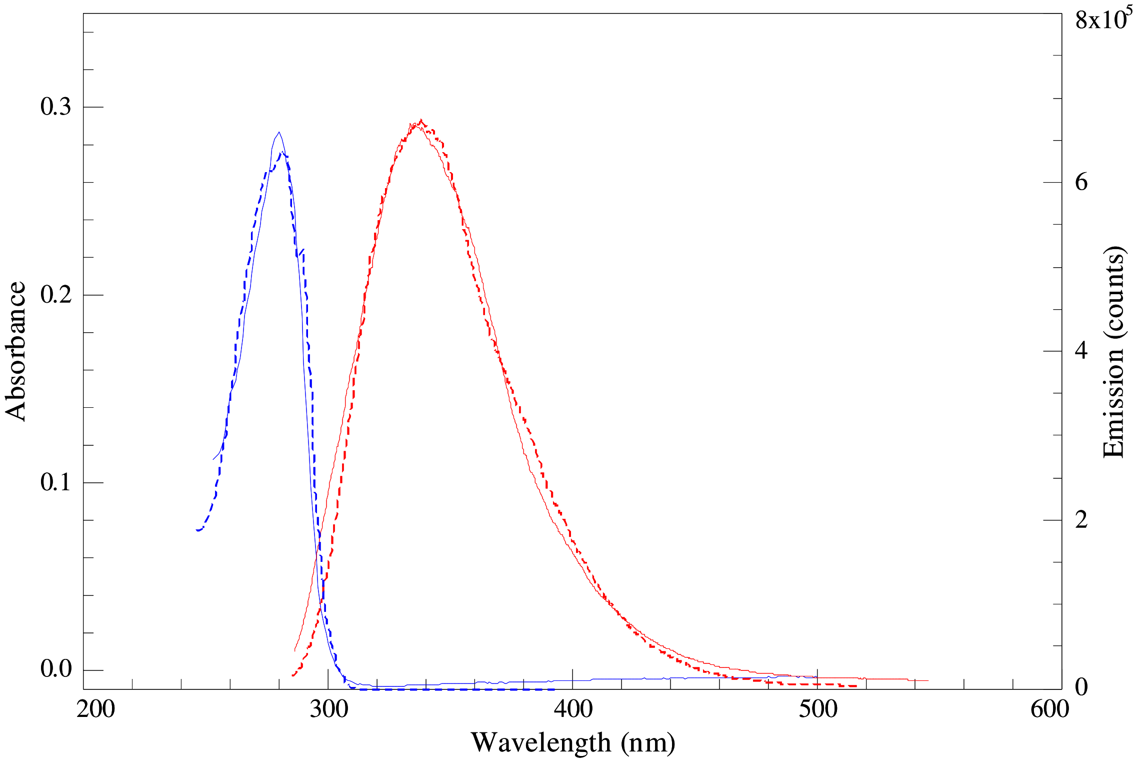


Figure VII.12. Steady-state luminescence of tryptophan and BSA. The absorption spectrum of tryptophan (dashed) and BSA (solid) is shown in blue and the emission spectrum of tryptophan (dashed) and BSA (solid) with excitation at 282.5 nm is in red.

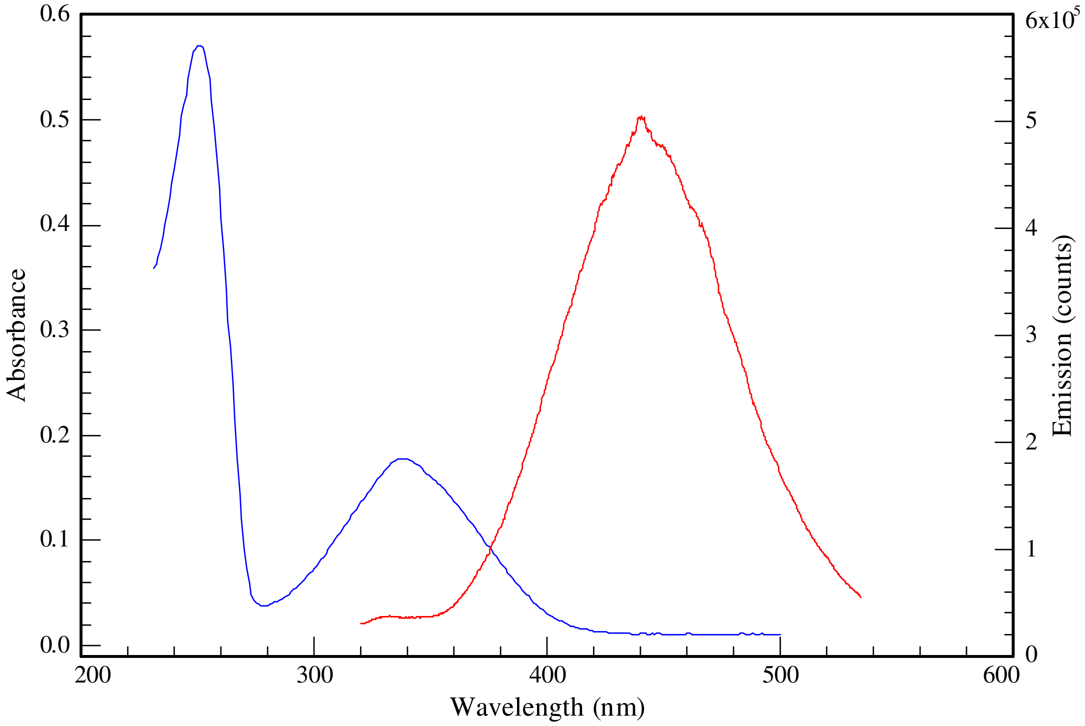


Figure VII.13. Steady-state luminescence of IAEDANS. The absorption spectrum is shown in blue and the emission spectrum with excitation at 282.5 nm is in red.

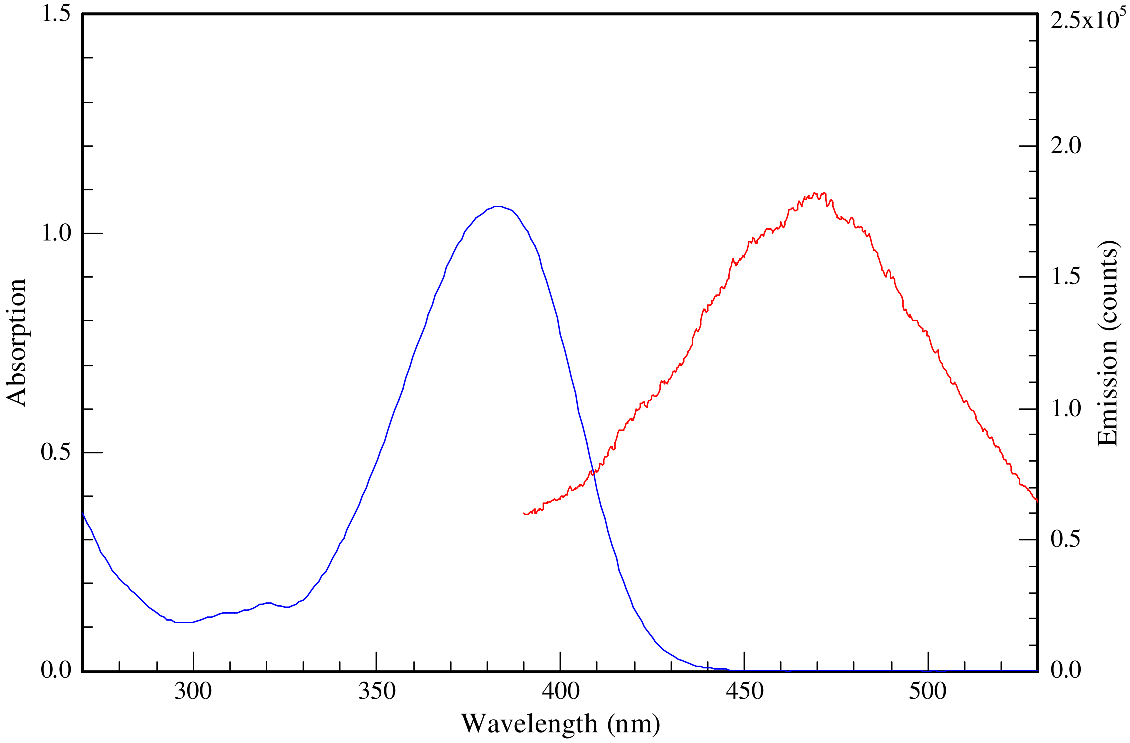


Figure VII.14. Steady-state luminescence of CPM. The absorption spectrum is shown in blue and the emission spectrum with excitation at 282.5 nm is in red.

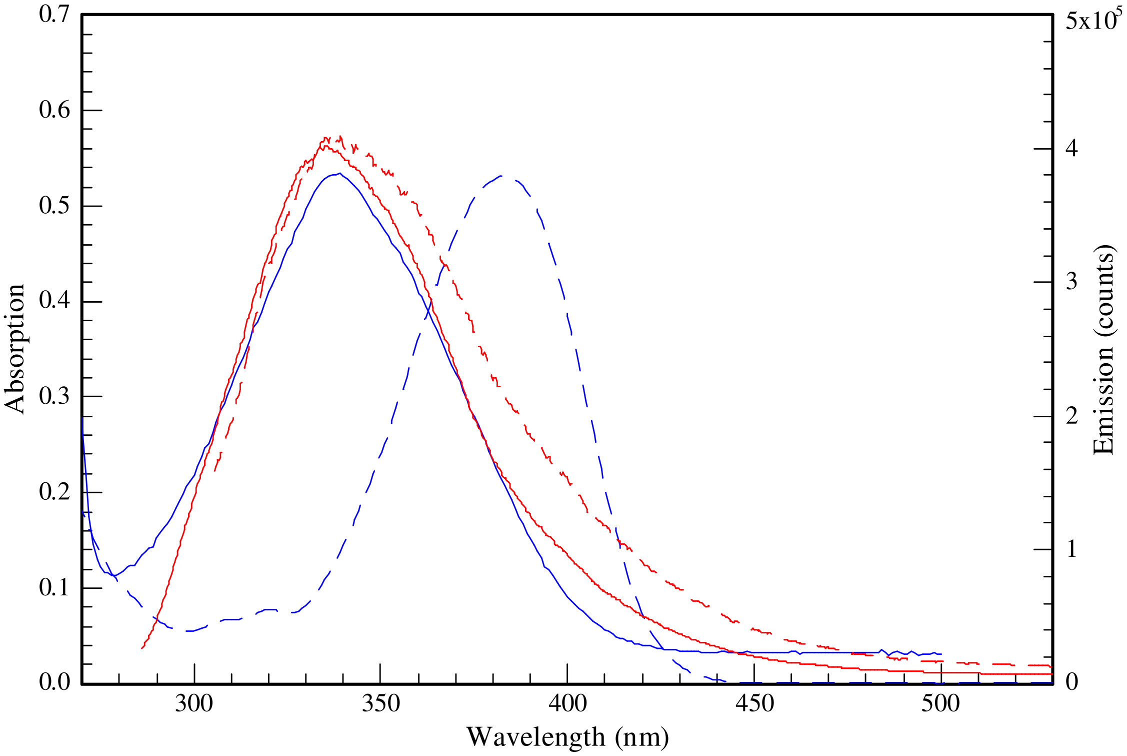


Figure VII.15. Donor-acceptor absorption/emission overlap. The blue absorption spectra of IAEDANS (solid line) and CPM (dashed line) overlap the red emission spectra of both BSA (solid line) and LEA (dashed line) when excited at 282.5 nm.

Figure VII.16 shows the emission spectrum of IAEDANS-labelled BSA when excited at 282.5 nm. Two distinct bands are clearly present, corresponding to the tryptophan emission at around 350 nm and the IAEDANS emission at around 480 nm. This assignment can be made based upon the emission bands shown in Figures VII.12 and VII.13 when exciting at the same wavelength. Overlaid in Figure VII.16 are the excitation spectra acquired while monitoring emission at both the tryptophan (blue) and IAEDANS (red) band maxima. The blue excitation spectrum shows bands located around 295 and 275 nm. Tryptophan displays bands at both these wavelengths and inspection of Figure VII.12 allows this band to be safely assigned to the tryptophan absorption. The red excitation spectrum shows three distinct bands. Both the low and high wavelength bands are present in the IAEDANS absorption spectrum (Figure VII.13) and are, therefore, assigned to IAEDANS absorption. The in-between band located at around 295 nm is not present in the IAEDANS absorption spectrum and corresponds almost exactly in both location and bandshape to that assigned to the tryptophan absorption in the blue excitation spectrum. This is evidence that at least some of the emission intensity at the band maximum of the IAEDANS acceptor resulted from excitation of the tryptophan donor by the process of FRET.

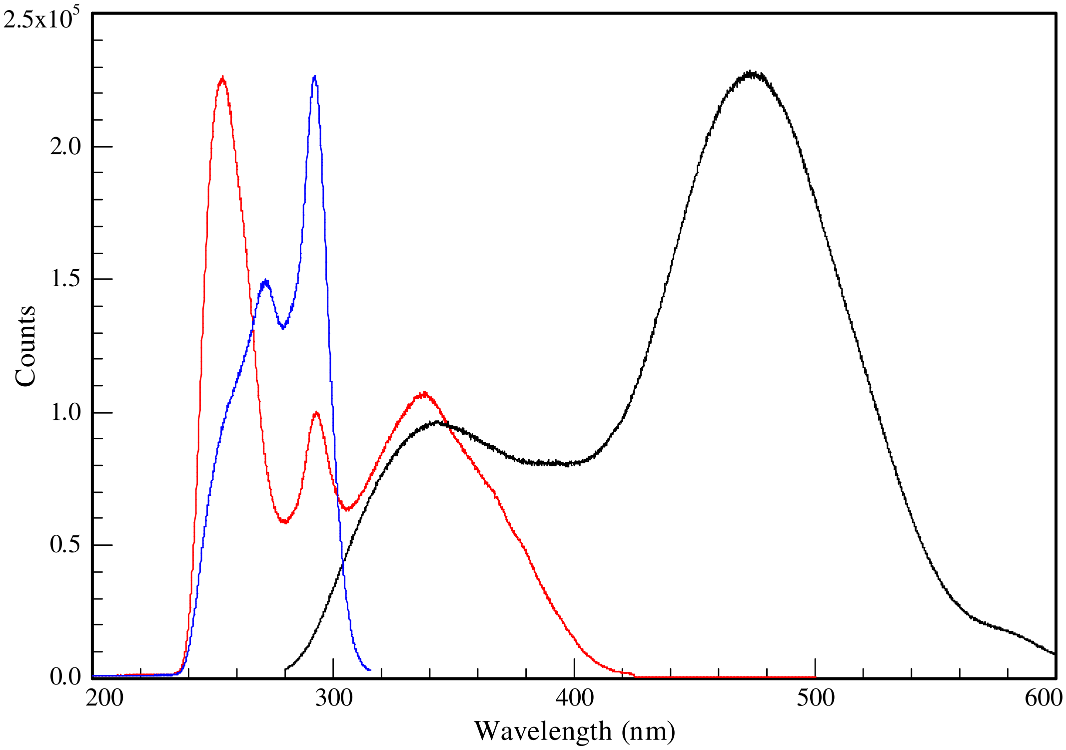


Figure VII.16. Steady-state luminescence of IAEDANS-labelled BSA. The black spectrum is the emission of the labelled BSA when excited at 282.5 nm. The red and blue spectra are the excitation spectra recorded while monitoring emission at 473 and 340 nm, respectively.

Figure VII.17 shows the steady-state luminescence profile of CPM-labelled BSA. The emission spectrum (red) contains two distinct bands centred at approximately 340 and 460 nm. The band at 340 nm is assigned as tryptophan emission and by inspection of Figure VII.14 the 460 nm band can be seen to arise from CPM emission. The relative intensities of the two emission bands are consistent with the higher quantum yield of the CPM acceptor over that of the tryptophan donor. The familiar tryptophan absorption band dominates the excitation spectrum recorded while monitoring emission at 342 nm (black). This is as expected and confirms the assignment of the 340 nm emission band to tryptophan fluorescence. The excitation spectrum generated through the monitoring of emission at 461 nm reveals the presence of at least three distinct bands. The broadest and high wavelength band centred at around 380 nm is identifiable from Figure VII.14 as representing CPM absorption. The low wavelength band centred at around 260 nm is also present in the

CPM absorption spectrum, however, due to its intensity it leads to saturation of the spectrometer detector and for this reason it does not appear in Figure VII.14. The middle band at around 295 nm does not appear in Figure VII.14 and so cannot reflect CPM absorption. It is, however, present in the tryptophan absorption spectrum (Figure VII.12) and, therefore, the occurrence of this band is evidence for the excitation of the CPM acceptor via energy transfer from the tryptophan donor via FRET.

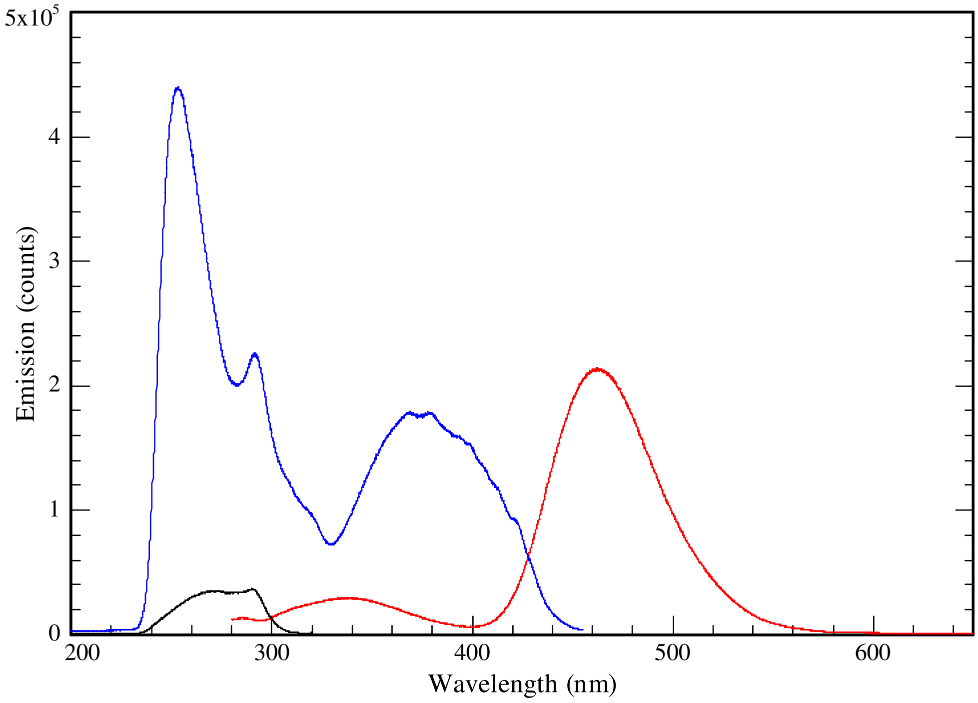


Figure VII.17. Steady-state luminescence of CPM-labelled BSA. The emission spectrum with excitation at 282.5 nm is shown in red. The excitation spectra while monitoring emission at 342 and 461 nm are shown as black and blue, respectively.

VII.3.10 FRET for Fluorescently Labelled BSA by Time-Resolved Luminescence Spectroscopy

An alternative method for detecting the presence of FRET is by performing time-resolved emission spectroscopy to monitor the decay time of the donor excited state. The occurrence of FRET has the effect of creating an alternative relaxation pathway for the excited donor molecules and, therefore, the average decay time for the many excited state donor molecules should become decreased in the presence of FRET. Figures VII.18 and VII.19 show the time-resolved emission spectrum of BSA and IAEDANS-labelled BSA when excited at 282.5 nm. The laser exciting radiation is represented in both spectra as a sharp band at this wavelength. An emission band centred at approximately 340 nm is also present in both spectra and arises from the decay of excited state tryptophan molecules. An additional emission band centred at approximately 470 nm is present in Figure VII.19 and this is created by the radiative decay of excited IAEDANS molecules. Figure VII.20 compares temporal slices taken at 340 nm from both Figures VII.18 and VII.19, which correspond to the BSA tryptophan decay profiles in both the non-labelled (red triangles) and IAEDANS-labelled (black squares) conditions, respectively. From Figure VII.20, the greater slope of the IAEDANS-labelled tryptophan decay profile beyond the emission maximum relative to the case of the non-labelled tryptophan decay is clear. The only explanation for this decrease in the decay time of the BSA tryptophan residues within the IAEDANS-labelled BSA is that energy transfer from excited tryptophan residues to ground state IAEDANS molecules has provided an additional relaxation pathway, resulting in a decrease in lifetime of the tryptophan excited state.

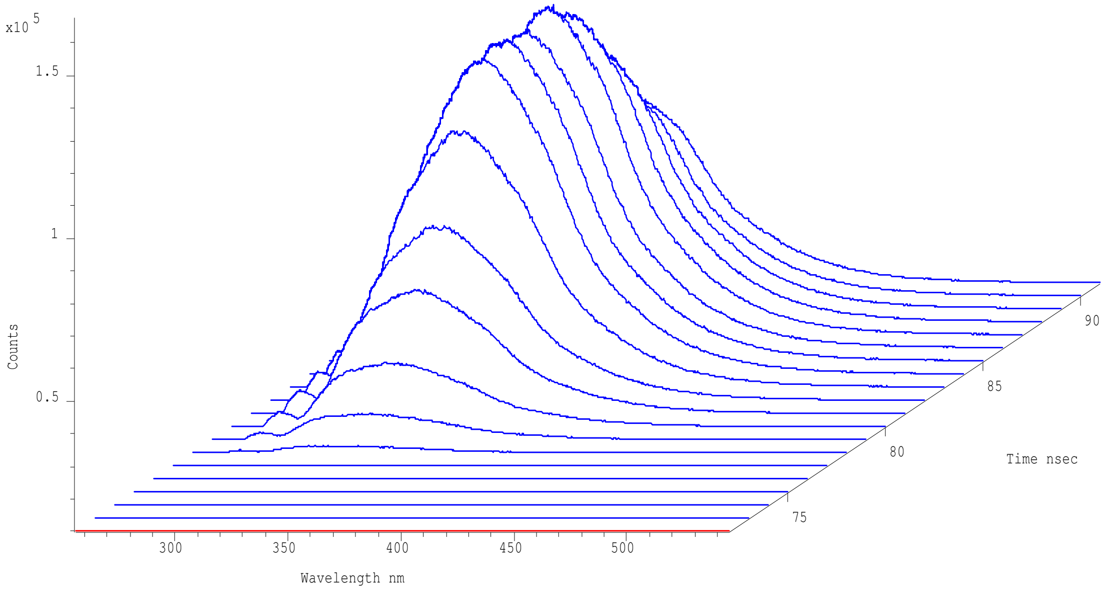


Figure VII.18. Time-resolved luminescence of BSA. Emission was recorded after excitation at 282.5 nm.

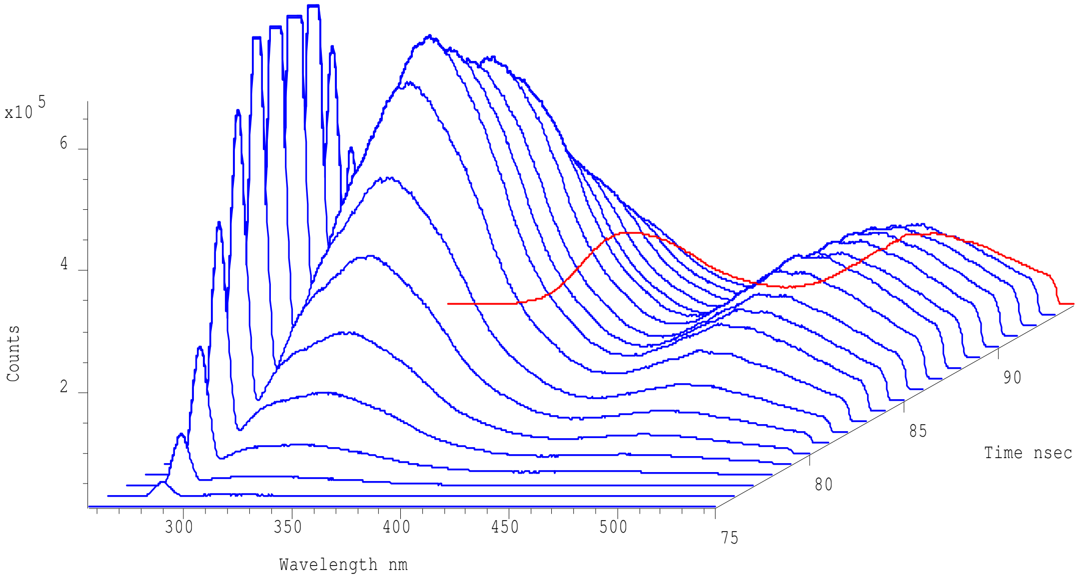


Figure VII.19. Time-resolved luminescence of IAEDANS-labelled BSA. Emission was recorded after excitation at 282.5 nm. Note the prompt emission of the tryptophan at 340 nm and the delayed emission of IAEDANS at 480 nm.

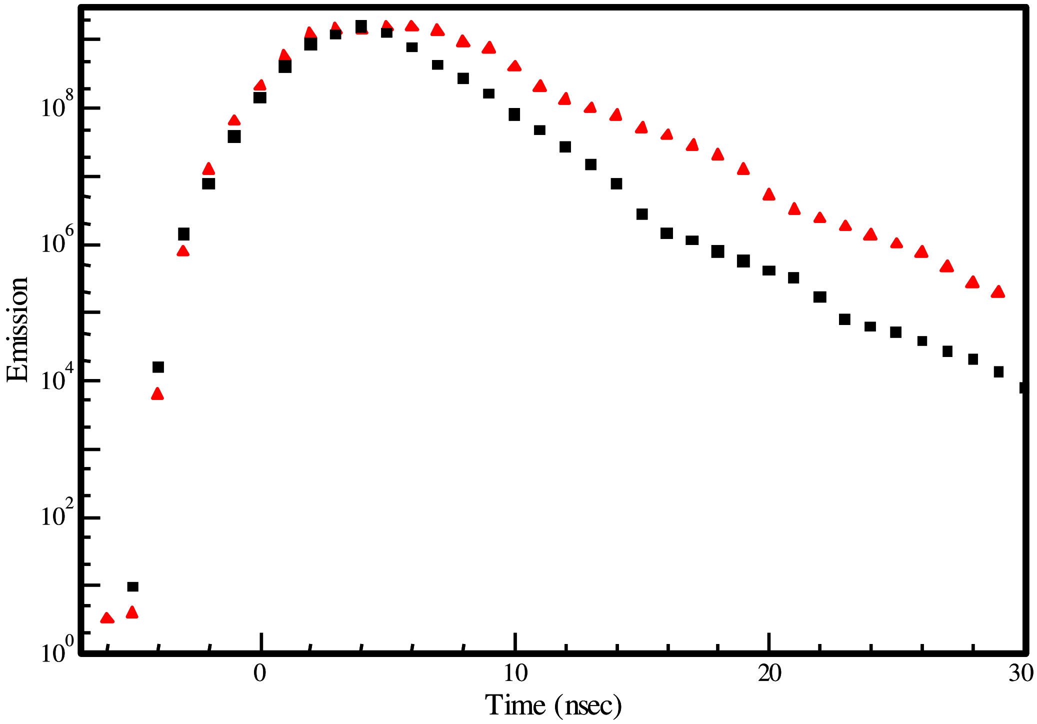


Figure VII.20. Overlay of the emission intensity temporal profiles at 340 nm for the BSA and IAEDANS-labelled BSA. The rise and decay of the tryptophan emission of BSA is shown on a semi-log plot for unlabelled (red) and labelled (black) BSA after excitation at 282.5 nm.

Figure VII.21 shows the time-resolved emission spectrum of CPM-labelled BSA with excitation at 282.5 nm. The tryptophan emission band is again present, but with low intensity at 340 nm, along with the exciting laser radiation at 282.5 nm. The more intense band centred at approximately 460 nm is from radiative relaxation of excited stated CPM molecules. Figure VII.22 compares a temporal slice taken through the tryptophan emission band at 340 nm of Figure VII.21 with that of Figure VII.18. In Figure VII.22 the blue stars represent the decay time profile of tryptophan residues within non-labelled BSA and the black stars show the same tryptophan decay time profile within CPM-labelled BSA. As in the case of the IAEDANS-labelled BSA, the increased slope of the tryptophan decay profile beyond the emission maximum relative to that of the non-labelled tryptophan decay is again clear. Similarly, the explanation for this is the occurrence of FRET between the tryptophan donor and CPM acceptor molecules.

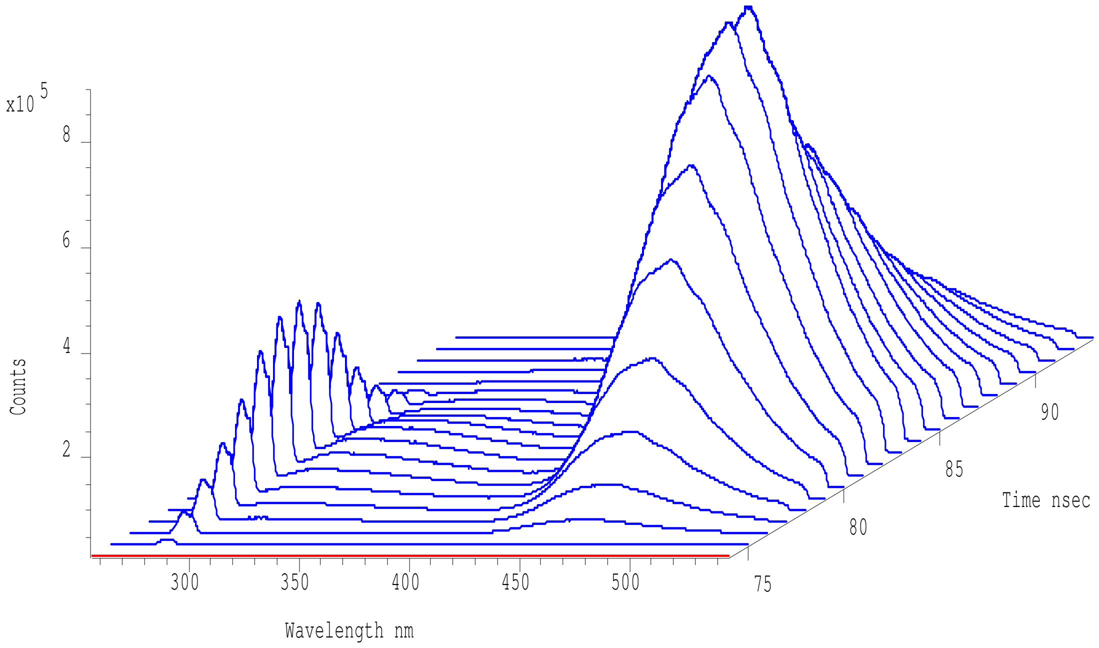


Figure VII.21. Time-resolved emission of CPM-labelled BSA. Emission was acquired with excitation at 282.5 nm. The tryptophan emission at 340 nm is almost entirely quenched in this spectrum.

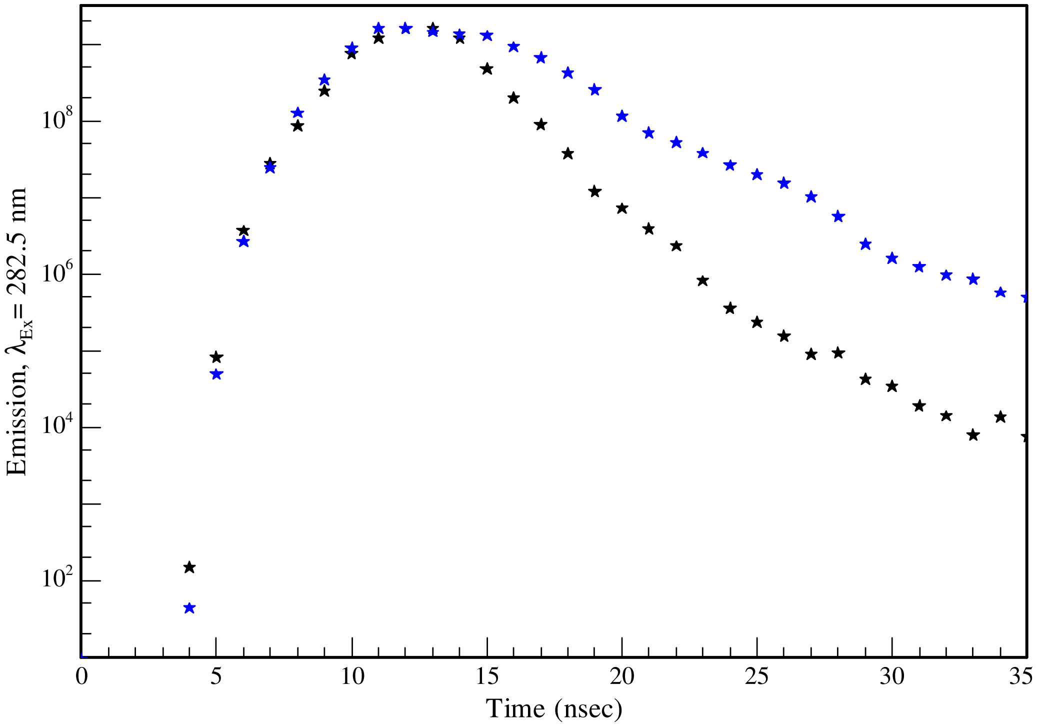


Figure VII.22. Time profile of tryptophan in native and CPM-labelled BSA. The rise and decay in emission intensity at 342 nm, following excitation at 282.5 nm, of both native and CPM-labelled BSA is shown as blue and black stars, respectively.

VII.3.11 Steady-State Luminescence and Time-Resolved Emission of AavLEA1

Having established that our technique for engineering the conditions for FRET arising from tryptophan excitation and that our methods for detecting this event were both valid for the BSA protein, the procedure was then transferred to the AavLEA1 mutants L37C, S38C and L132C. While expression of the L132C mutant was very successful, both the L37C and S38C mutants expressed at considerably lower levels than that of the wild-type protein. Although this did not prevent investigation of both the L37C and S38C mutants for the occurrence of FRET, it did restrict the examination to either the IAEDANS or CPM fluorophores, and not both. Also, it was decided that of the two techniques of steady-state and time-resolved spectroscopy for detecting FRET, the latter was superior for use in the AavLEA1 system, since it allowed for the determination of distance between the donor and acceptor molecules where both are singly present, as in the AavLEA1 case. Therefore, only the investigation of FRET within mutants of AavLEA1 using time-resolved emission spectroscopy is presented below.

Figure VII.23 shows the absorption spectrum and emission produced with excitation at 282.5 nm of the wild-type AavLEA1. The presence of tryptophan within the small AavLEA1 protein is evidenced by the absorption at around 290 nm and emission at 340 nm bands. Figure VII.24 shows the time-resolved emission spectrum of wild-type AavLEA1 when excited at 282.5 nm. The growth and decay of the emission band of tryptophan is again present at 340 nm, along with the sharp laser band at 282.5 nm. Taking a temporal slice through both the laser and tryptophan emission bands of Figure VII.24 allowed for the generation of Figure VII.25. Although Figure VII.24 shows the laser band as saturated, this is only an effect from showing the spectrum on a scale matching the tryptophan emission band and in reality care was taken to ensure that the laser emission was always below saturation. Both the laser (red triangles) and tryptophan (black stars) emission temporal slices are overlaid in Figure VII.25, along with a single exponential fit of the tryptophan emission (solid line) that returns the lifetime of the tryptophan excited state as 2.20 ns. The single exponential fit is achieved by deconvoluting the effect of the time profile of the laser from that of the tryptophan emission. The accuracy of this fit to the experimental data is revealed by how closely the two overlap and this determines the confidence to be had in the estimated tryptophan excited state lifetime. Unfortunately, the occurrence of a “jitter” in the iCCD detector had the effect of introducing small errors into some of the experimental data and, where a linear decay in time of emission intensity beyond the emission maximum is expected for a single exponential excited state decay, this is not always reflected in the data. Such deviation in linearity is attributed to this “jitter” effect and not to some multi exponential decay pattern of the tryptophan excited state.

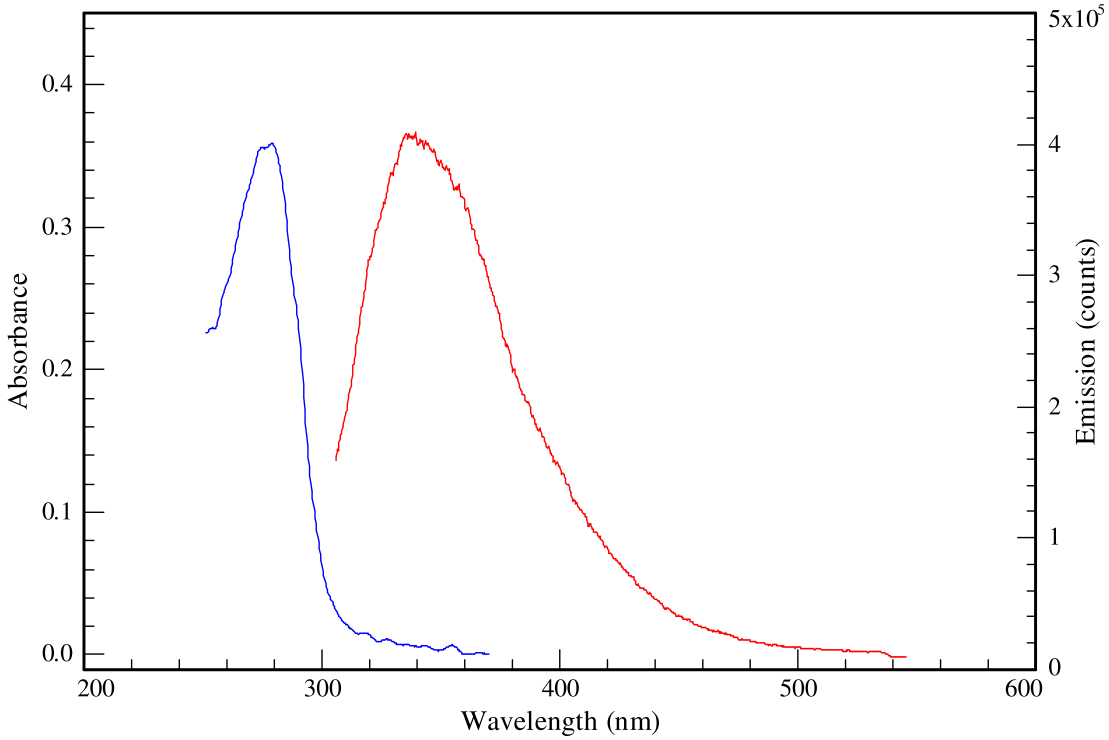


Figure VII.23. AavLEA1 steady-state luminescence. The blue spectrum is the AavLEA1 absorption spectrum and the red is the emission spectrum when excited at 282.5 nm.

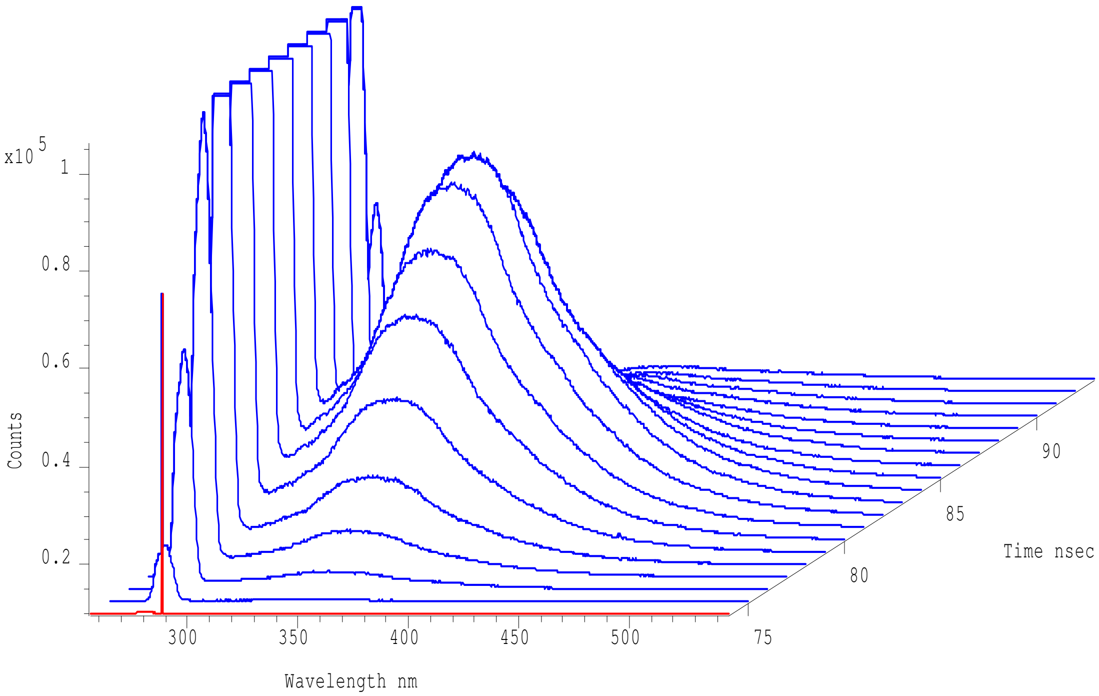


Figure VII.24. Time-resolved emission of AavLEA1. Emission was recorded after excitation at 282.5 nm.

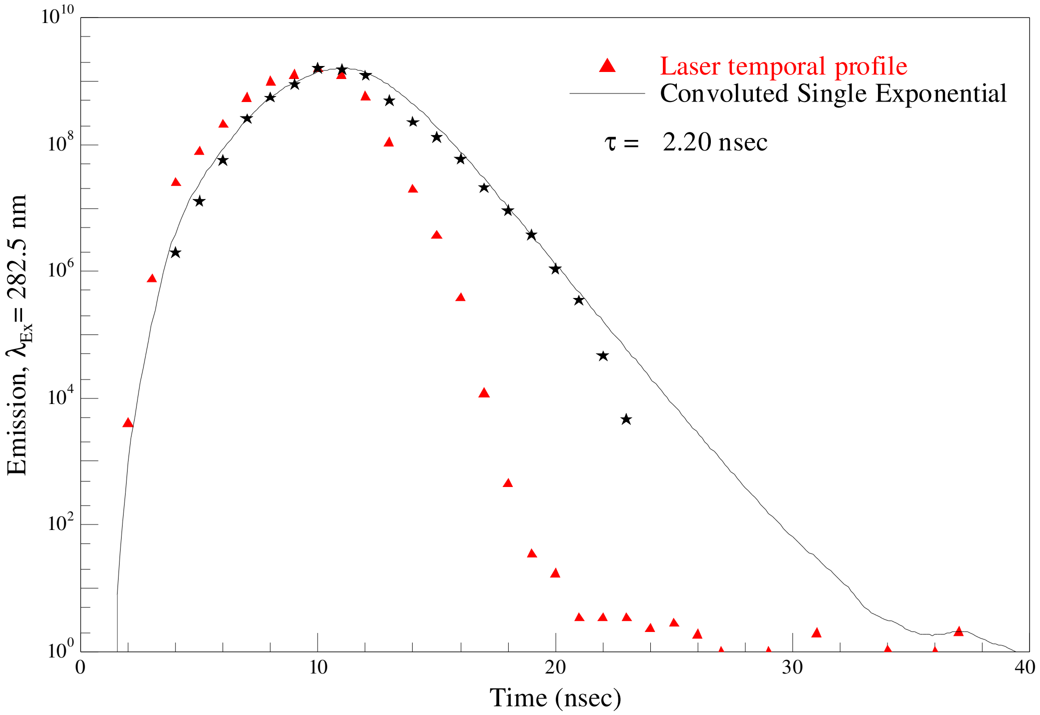


Figure VII.25. AavLEA1 emission time profile at 340 nm. The laser-convoluted single exponential fit of the temporal profile of the AavLEA1 emission is shown as the solid line.

VII.3.12 Steady-State and Time-Resolved Emission of IAEDANS-Labelled AavLEA1 Mutants

The steady-state emission spectral overlay of IAEDANS-labelled L37C (blue) and L132C (black) AavLEA1 with excitation at 282.5 nm is shown in Figure VII.26. Both the tryptophan and IAEDANS emission bands are present in both spectra at 340 and 460 nm, respectively. The relative intensities of these two bands are markedly different between the two spectra. For the L37C spectrum the relative intensity of the tryptophan and IAEDANS band is almost equal, whereas the L132C spectrum shows a considerably more intense tryptophan emission band than that of IAEDANS. This is consistent with FRET occurring for the IAEDANS-labelled L37C protein, but not for that of the L132C, since FRET would lead to a non-radiative decay of a fraction of tryptophan excited state molecules and, therefore, reduce the intensity of the tryptophan emission band relative to that of IAEDANS. At the same time energy transfer from tryptophan to IAEDANS would generate a larger number of excited state IAEDANS molecules and so increase the IAEDANS emission band intensity relative to that of tryptophan.

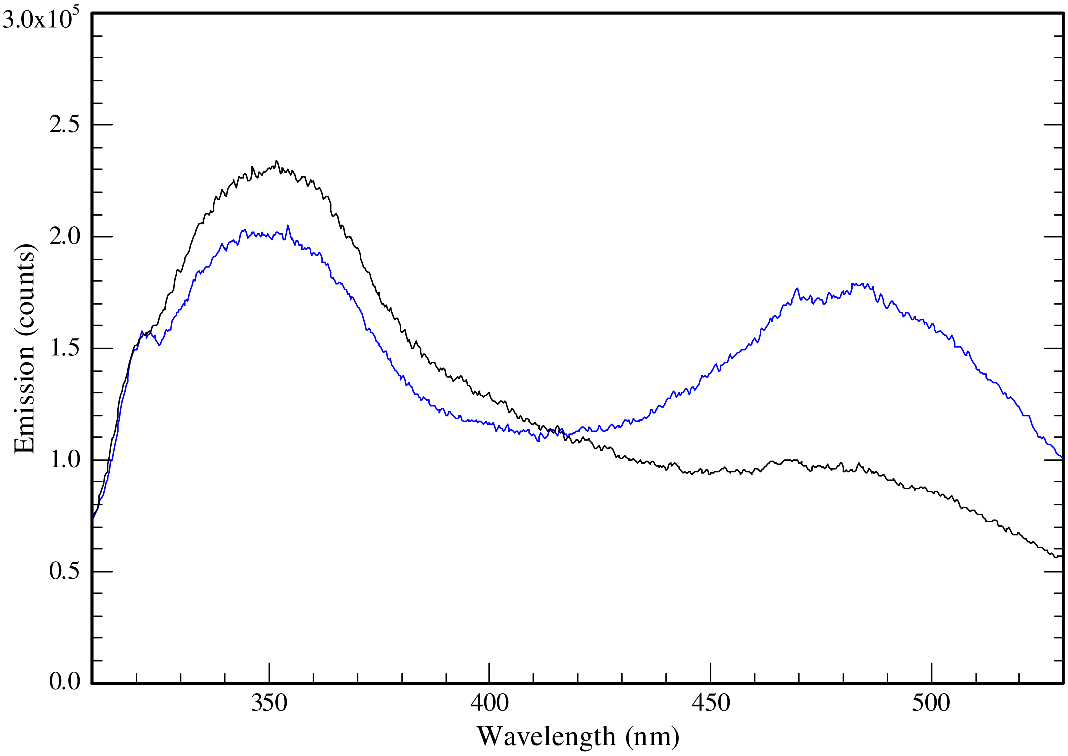


Figure VII.26. Steady-state emission of IAEDANS-labelled AavLEA1. The blue and black spectra show emission of the IAEDANS-labelled L37C and L132C AavLEA1 mutants, respectively, with excitation at 282.5 nm.

Figure VII.27 shows the time-resolved emission of L132C AavLEA1 with excitation at 282.5 nm. The laser band saturation at 282.5 nm is, once again, only apparent, resulting from showing the spectrum on a scale appropriate to the tryptophan (340 nm) and IAEDANS (460 nm) bands. Taking a temporal slice through Figure VII.27 at 282.5 and 340 nm gives the time profile of emission from the laser and tryptophan residues, respectively. Figure VII.28 shows both emission profiles, along with the single exponential fit of the tryptophan decay. The lifetime of the tryptophan excited state in the IAEDANS-labelled L132C AavLEA1 was calculated to be 2.30 ns, a lifetime roughly equal to tryptophan in the wild-type AavLEA1. From this there does not seem to occur any FRET for the IAEDANS-labelled L132C AavLEA1 protein.

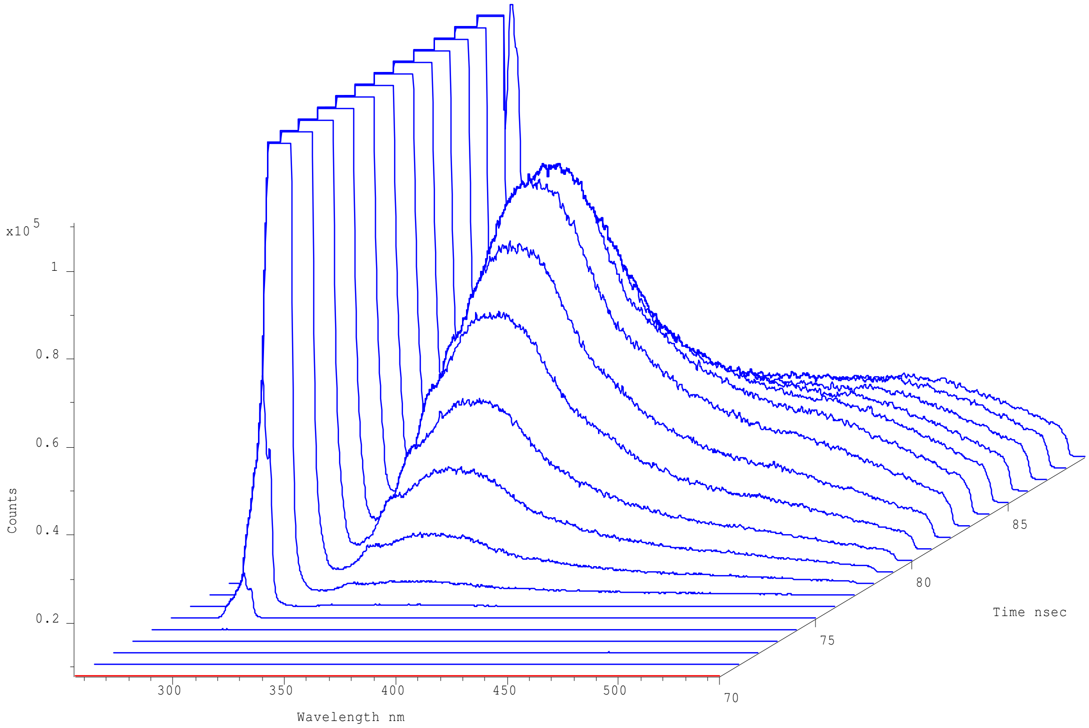


Figure VII.27. Time-resolved luminescence of IAEDANS-labelled L132C AavLEA1. Emission was recorded after excitation at 282.5 nm.

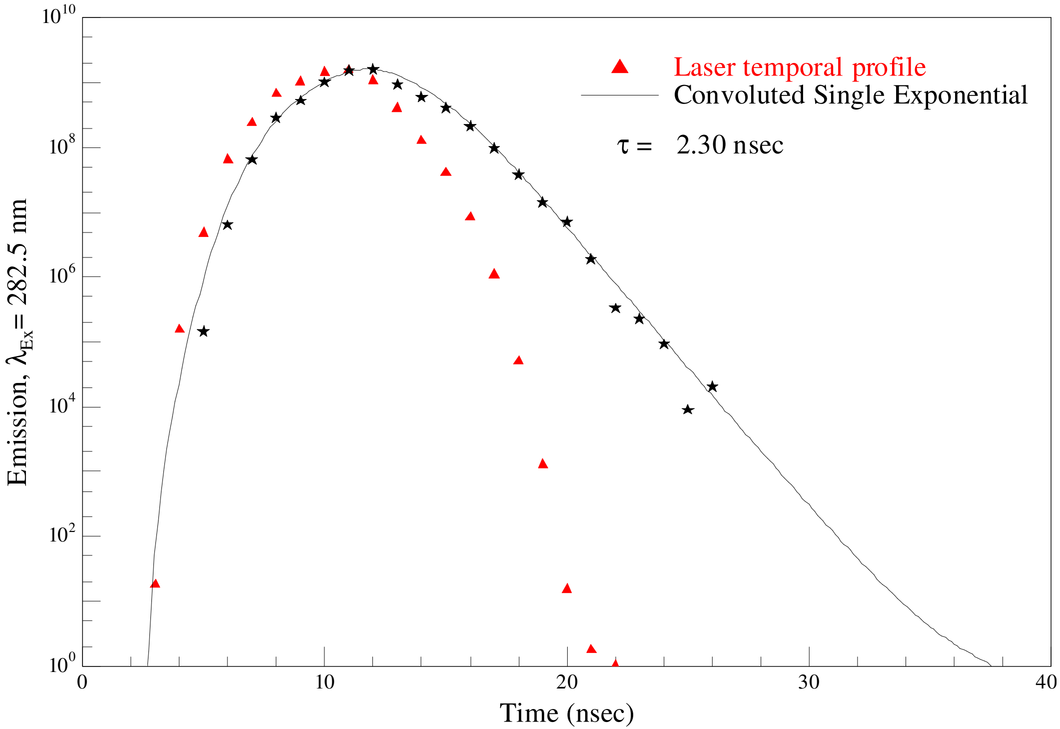


Figure VII.28. IAEDANS-labelled L132C AavLEA1 emission time profile at 340 nm.

Figure VII.29 shows the time-resolved emission of IAEDANS-labelled L37C AavLEA1 with excitation at 282.5 nm. Once more, the laser saturation at 282.5 nm is due to scaling. The second-order laser emission band is visible at 565 nm. Both the tryptophan (340 nm) and IAEDANS (460 nm) bands are clearly present. At first glance the relative intensity of these bands seems inconsistent with the steady-state emission spectrum shown in Figure VII.26. This can be explained by noting that the IAEDANS excited state lifetime is considerably longer than that of tryptophan and, therefore, the IAEDANS emission is spread out over a broader time range than that of tryptophan. Taking a temporal slice through the laser and tryptophan bands allows the fitting of the tryptophan decay and the calculation of its excited-state lifetime in IAEDANS-labelled L37C AavLEA1 as 1.80 ns (Figure VII.30). This lifetime value is considerably less than the 2.20 ns of tryptophan in wild-type AavLEA1 and, therefore, represents evidence of energy transfer from tryptophan to IAEDANS in the labelled L37C mutant.

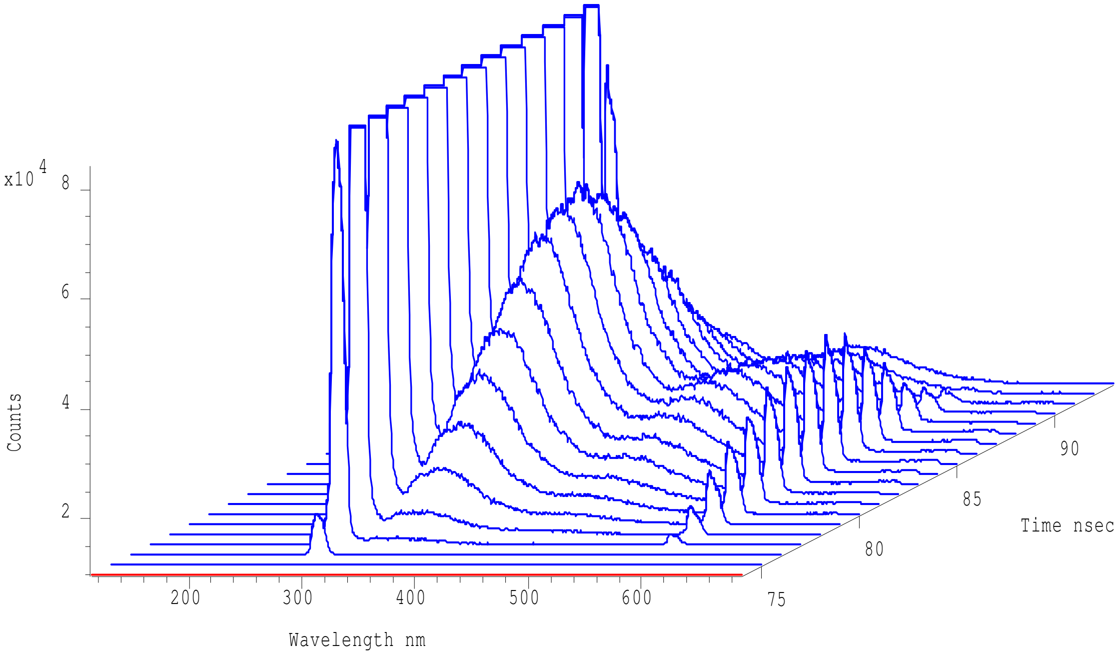


Figure VII.29. Time-resolved luminescence of IAEDANS-labelled L37C AavLEA1. Emission was recorded after excitation at 282.5 nm.

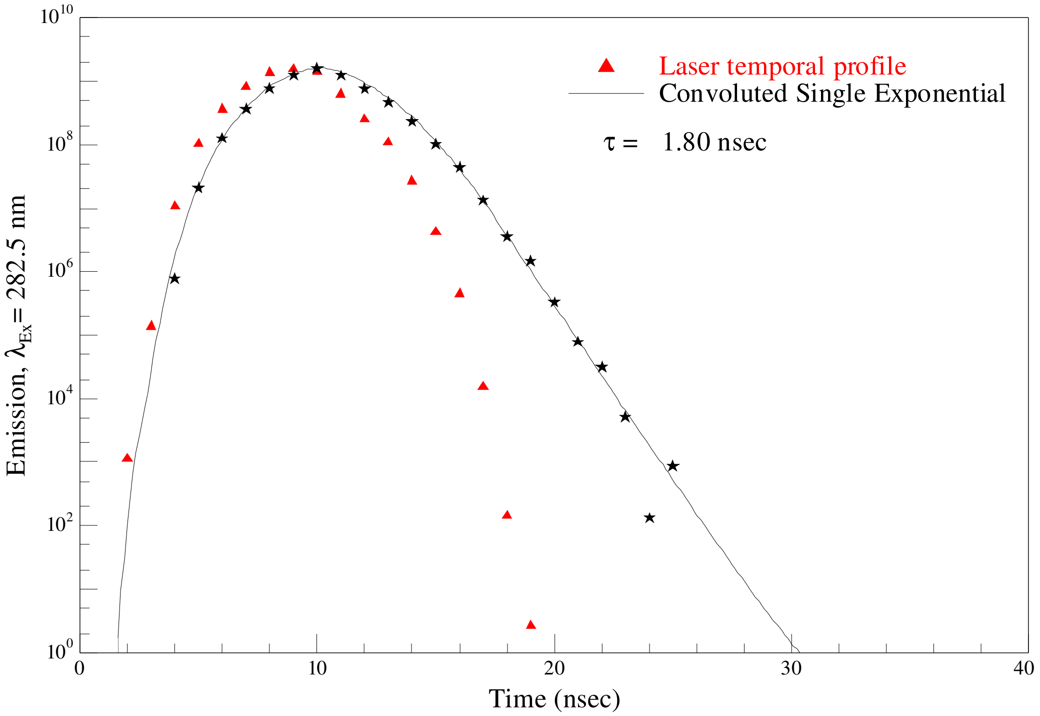


Figure VII.30. IAEDANS-labelled L37C AavLEA1 emission time profile at 340 nm.

VII.3.13 Time-Resolved Emission of CPM-Labelled AavLEA1 Mutants

The time-resolved emission of CPM-labelled L132C AavLEA1 is shown in Figure VII.31. An anomalous band is observed at around 300 nm and, because of both the sharp bandshape and the similarity of the time profile of this band to that of the laser band, this was seen to represent laser emission. The tryptophan and CPM emission bands are visible at 340 and 460 nm, respectively. Taking a temporal slice through the laser and tryptophan band maximum allowed the calculation of the tryptophan excited state lifetime to be calculated as 2.20 ns (Figure VII.32). This is identical to that of wild AavLEA1 and, therefore, suggests that no FRET occurs between tryptophan and CPM in the L132C mutant.

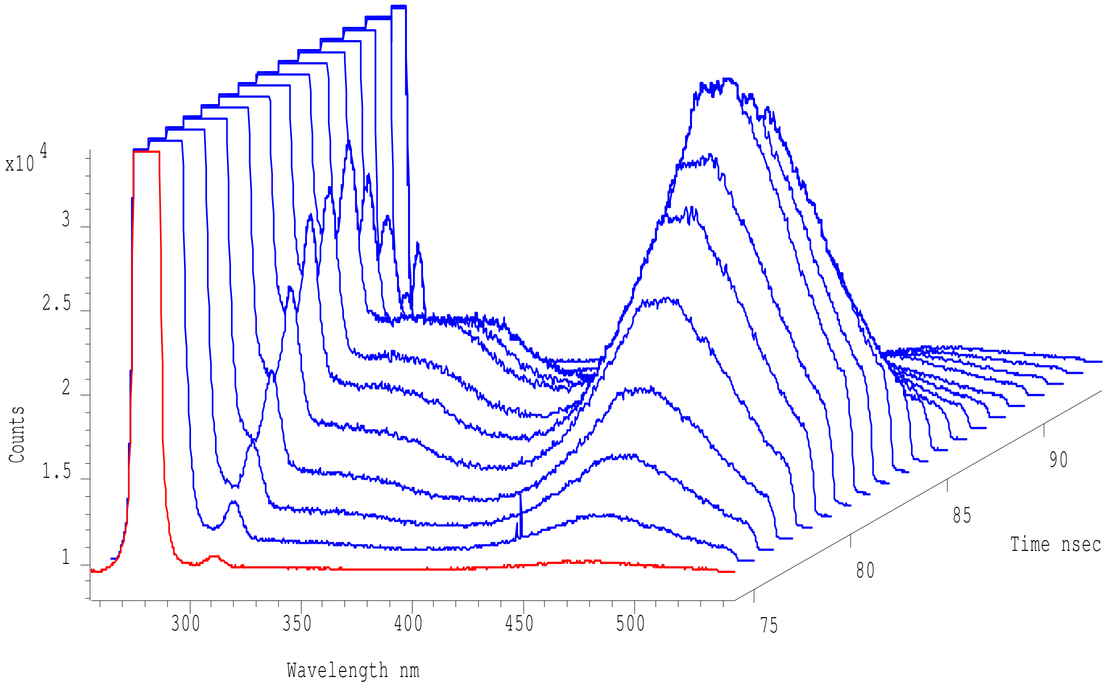


Figure VII.31. Time-resolved luminescence of CPM-labelled L132C AavLEA1. Emission was recorded after excitation at 282.5 nm

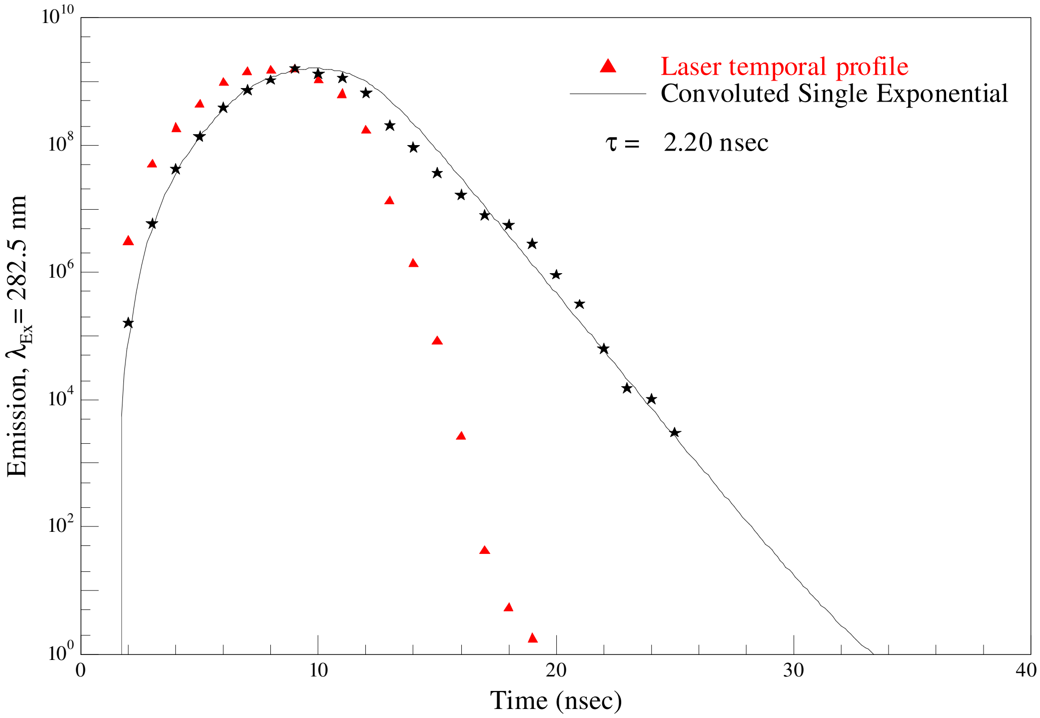


Figure VII.32. CPM-labelled L132C AavLEA1 emission decay at 340 nm.

The same laser, tryptophan and CPM emission bands can be seen in Figure VII.33 for the time-resolved emission of CPM-labelled S38C AavLEA1. Calculating the tryptophan excited state lifetime by the same method given above (Figure VII.34) gives a value of 2.25 ns. Again, this value is very close to that of the wild-type AavLEA1 and, therefore, it would appear that there occurs no energy transfer between tryptophan and CPM in the S38C mutant.

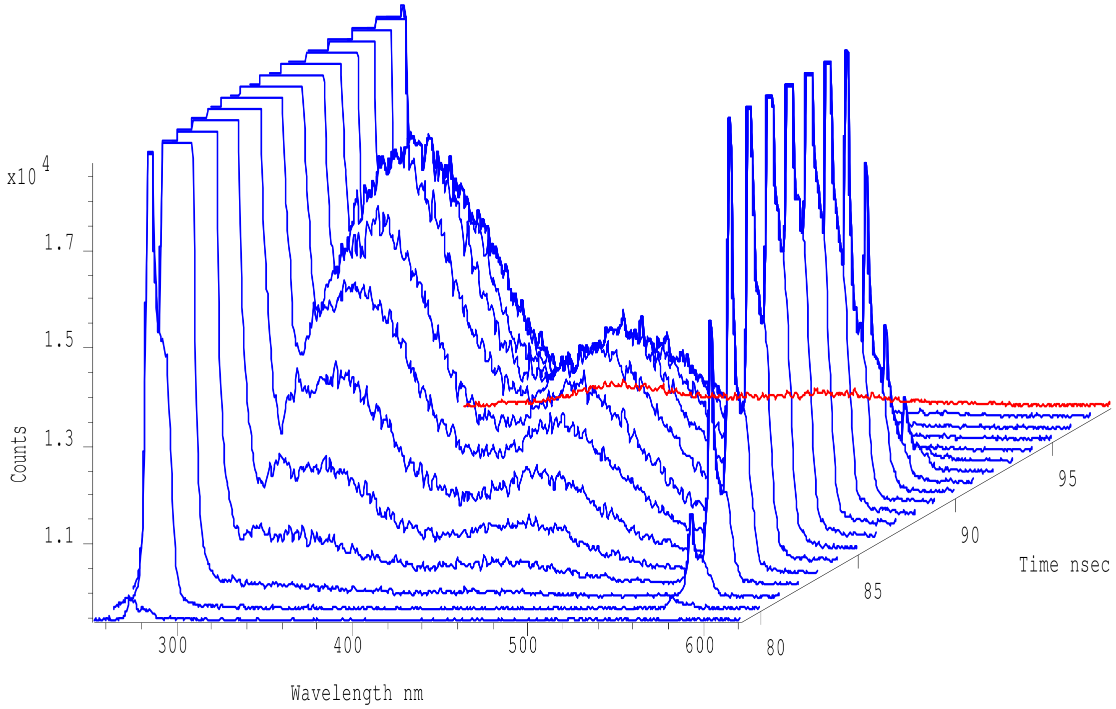


Figure VII.33. Time-resolved luminescence of CPM-labelled S38C AavLEA1. Emission was recorded after excitation at 282.5 nm

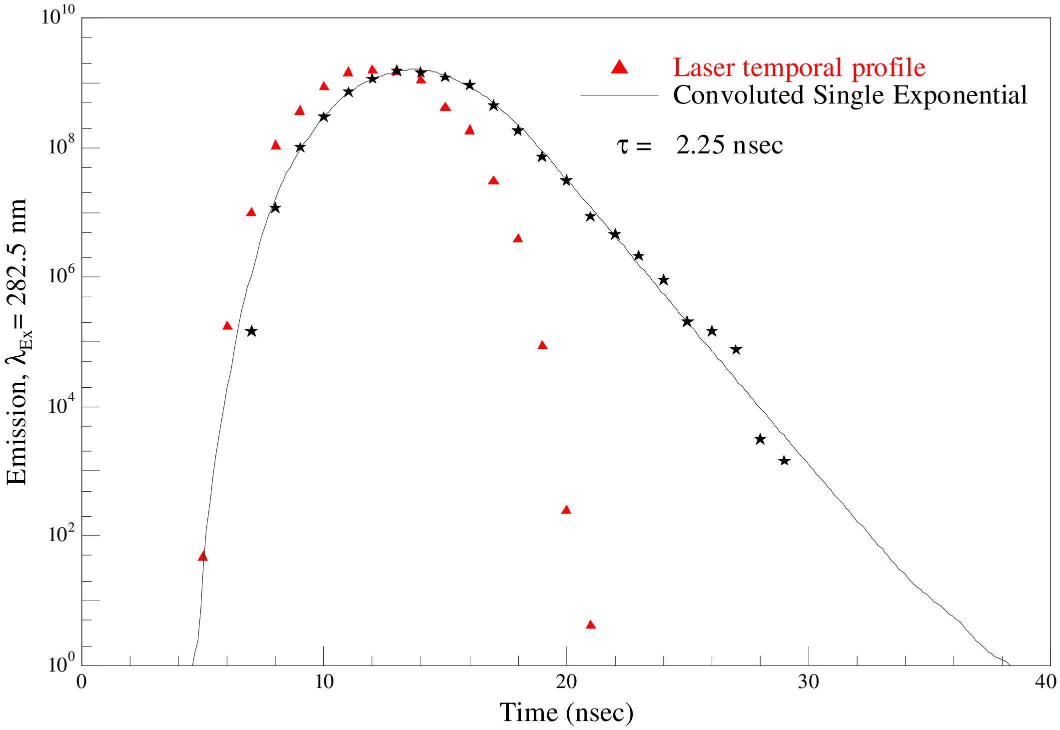


Figure VII.34. CPM-labelled S38C AavLEA1 emission decay at 340 nm.

VII.3.14 AavLEA1 Inter-Residue Distance Determination by FRET Measurements

From the FRET observed for the IAEDANS-labelled L37C it is possible to calculate the distance, *r*, between the tryptophan donor and the IAEDANS acceptor. This distance is calculated using the equations

$$r = R_0\left(\frac{1}{E}-1\right)^{\frac{1}{6}} \qquad \text{(VII.1)}$$

and

$$E = 1-\frac{\tau_{DA}}{\tau_D} \qquad \text{(VII.2)}$$

where R_0 is the Forster distance, τ_{DA} is the lifetime of the donor in the presence of the acceptor, τ_D is the lifetime of the donor in the absence of the acceptor and *E* is the efficiency of energy transfer. For the tryptophan-IAEDANS donor-acceptor pair R_0 is 22 Å. τ_D is the lifetime of tryptophan in the unlabelled AavLEA1 (2.20 ns) and τ_{DA} is the tryptophan lifetime in the IAEDANS-labelled L37C AavLEA1 (1.80 ns). Therefore, *E* has a value of 0.182 and *r* is 28.26 Å. There is a separation of 7 residues between the tryptophan residue at position 30 in the AavLEA1 protein and the mutated cysteine residue of L37C. Therefore, there is an inter-residue average distance of approximately 4 Å within this particular segment of the L37C AavLEA1 mutant. Considering the average inter-residue distances of 1.5 Å for a right-handed α-helix, 3.4 Å for an anti-parallel β-sheet and 3.2 Å for a parallel β-sheet, it would seem that in the region investigated above AavLEA1 adopts a fully-extended random-coil secondary structure.

The lack of any detected FRET for the CPM-labelled S38C mutant is difficult to explain considering the clear presence of FRET in IAEDANS-labelled L37C. This is especially true given that the Forster distance for the tryptophan-CPM donor-acceptor pair is 29 Å, while that of tryptophan-IAEDANS is only 22 Å. It is possible that the CPM fluorophore failed to attach to the mutated protein. This is somewhat plausible given the hydrophobic nature of CPM and its very limited water solubility. However, the labelling of BSA with CPM did not seem to be affected by this.

For the L132C mutant, FRET could only be expected to occur if the protein adopts a condensed tertiary structure, such that distant residues come to be located close to each other in space. This is true because FRET is a through-space rather than a through-bond phenomenon. Even if AavLEA1 were to adopt the most compact secondary structure, that of an α-helix, the 132 mutated cysteine residue is far too distant from the tryptophan donor for FRET to occur in the absence of a compact tertiary structure. The absence of FRET in either the IAEDANS- or CPM-labelled L132C AavLEA1 mutant is evidence that AavLEA1 does not adopt any compact tertiary structure and this is consistent with the data from the L37C mutant, which suggests that AavLEA1 exists as a fully-extended natively unfolded protein.

VII.4 Discussion

In a polar environment the main driving forces that determine whether a protein will exist in a folded or unfolded state are the protein’s mean hydrophobicity and its mean net charge. Hydrophobic regions will tend to become folded and charged regions will tend towards existing in an unfolded state. Uversky has made use of this understanding by constructing a charge-hydrophobicity phase space diagram with the protein's mean net charge on the y-axis and its mean hydrophobicity on the x-axis.[15] Upon populating his charge-hydrophobicity phase space diagram with a range of well studied proteins of known secondary structure that consisted of similar numbers of both structured and unfolded proteins, Uversky observed that a boundary line could be drawn that clearly separated folded from unfolded proteins. This line can be expressed by the equation

$$\langle H \rangle_b = \frac{\langle R \rangle + 1.151}{2.785} \qquad \text{(VII.3)}$$

where $<H>_b$ and $<R>$ represent a protein's mean hydrophobicity and mean net charge, respectively. Based upon which side of the boundary line a protein falls, one can accurately predict whether that protein will be either folded or unfolded. AavLEA1, with its low mean hydrophobicity value and high net charge, lies well within the region populated by unfolded proteins.

Both CD and FTIR spectral analyses confirm the secondary structure of AavLEA1 as predominantly random coil. Also, the response of AavLEA1 structure to changes in physiological conditions, such as temperature, pH and cation concentration is entirely consistent with that of other natively unfolded proteins, in that AavLEA1 displays a 'turn-out' effect at elevated temperatures, acidic pH and increased cation concentrations, whereby it experiences an increase in its ordered secondary structure content.[16] This is in direct contrast to what is observed for structured proteins under these conditions. Given that AavLEA1 exists as a natively unfolded protein it is assumed that this confers substantial plasticity to the protein molecule and it is likely that this plasticity is in some way related to AavLEA1's biological function in anhydrobiosis.

Although AavLEA1 is natively unfolded in aqueous solution it is still possible that it might carry out its anhydrobiotic function as a structured protein, since under desiccating conditions the protein would lose its water content and this may have a profound effect on the AavLEA1 secondary structure. The potential for AavLEA1 to adopt a well-ordered structure was examined using TFE and it was seen that under these conditions AavLEA1 shifted its secondary structure from being predominantly random coil to being almost completely α-helical. Furthermore, Goyal *et al.* has shown that when dried AavLEA1 becomes considerably helical in structure.[1]

Considering the structural prediction that AavLEA1 has a 40 % chance of adopting a coiled-coil super secondary structure, it is tempting to imagine that AavLEA1 carries out its anhydrobiotic function by forming coiled-coil type polymeric fibrils in the absence of water that reinforce the trehalose glassy matrix that forms upon desiccation.

From the results of the FRET analysis it appears that AavLEA1 adopts an extended tertiary structure rather than existing as a compact protein and this is consistent with the notion that protein plasticity is involved in its functioning and fits well with the model of AavLEA1 functioning as a protein solvent shell replacement by moulding itself around other well-structured proteins. AavLEA1 has demonstrated antagonistic properties towards heat-induced aggregation of both cytochrome c and citrate synthase and it is plausible that it does this by masking regions of these proteins that become exposed upon heating, which are likely to interact with other proteins and so form protein aggregates. That AavLEA1 contributes no temperature-induced aggregation antagonism to four other standard proteins studied suggests that the main role of AavLEA1 is not that of a heat-shock protein. Goyal *et al.* have also observed that AavLEA1 offers cryoprotection against inactivation and aggregation to both citrate synthase and lactate dehydrogenase and that it acts synergistically with trehalose to protect lactate dehydrogenase activity upon drying.[14] From this it seems possible that AavLEA1 functions in a less specific and more general way, in that it may facilitate a host of beneficial processes that help the cell during anhydrobiosis and at the root of this general functioning is the protein's molecular plasticity.

References

1 Goyal K, Tisi L, Basran A, Browne J, Burnell A, Zurdo J, Tunnacliffe A. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. Journal of Biological Chemistry 278(15):12977-12984, 2003.

2 Crowe JH, Hoekstra FA, Crowe LM. Anhydrobiosis. Annual Review of Physiology 54:579-599, 1992.

3 Oliver MJ, Tuba Z, Mishler BD. The evolution of vegetative desiccation tolerance in land plants. Plant Ecology 151(1):85-100, 2000.

4 Clegg JS. Cryptobiosis - a peculiar state of biological organization. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 128(4):613-624, 2001.

5 Perry RN. Desiccation survival of parasitic nematodes. Parasitology 119:S19-S30, 1999.

6 Crowe JH, Carpenter JF, Crowe LM. The role of vitrification in anhydrobiosis. Annual Review of Physiology 60:73-103, 1998.

7 Bolen DW, Baskakov IV. The osmophobic effect: Natural selection of a thermodynamic force in protein folding. Journal of Molecular Biology 310(5):955-963, 2001.

8 Wise MJ. LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. Bmc Bioinformatics 4:-, 2003.

9 Browne JA, Dolan KM, Tyson T, Goyal K, Tunnacliffe A, Burnell AM. Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode Aphelenchus avenae. Eukaryotic Cell 3(4):966-975, 2004.

10 Dure L. A Repeating 11-Mer Amino-Acid Motif and Plant Desiccation. Plant Journal 3(3):363-369, 1993.

11 Stetefeld J, Jenny M, Schulthess T, Landwehr R, Engel J, Kammerer RA. Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. Nature Structural Biology 7(9):772-776, 2000.

12 Wolf E, Kim PS, Berger B. MultiCoil: A program for predicting two- and three-stranded coiled coils. Protein Science 6(6):1179-1189, 1997.

13 Wolkers WF, McCready S, Brandt WF, Lindsey GG, Hoekstra FA. Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro.

Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1544(1-2):196-206, 2001.

14 Goyal K, Walton LJ, Tunnacliffe A. LEA proteins prevent protein aggregation due to water stress. Biochemical Journal 388:151-157., 2005.

15 Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins-Structure Function and Genetics 41(3):415-427, 2000.

16 Uversky VN. What does it mean to be natively unfolded? European Journal of Biochemistry 269(1):2-12, 2002.

Chapter VIII

Spectroscopic Investigation of the Rab11-FIP2 and Rab11-FIP3 Protein Dimers

VIII.1 Background

The process by which cells selectively internalise material from their environments is known as endocytosis. Cells do this by incorporating specially constructed protein molecules into their outer plasma membranes, which are composed of a hydrophobic portion that is embedded in the non-polar lipid membrane and a hydrophilic region that extends into the cell's aqueous environment. Such proteins are generally referred to as receptor proteins and the hydrophilic region of each receptor protein is specific to bind certain ligand molecules that are present in the cell's environment and are required to be transported to its interior. Upon binding of ligands to the protein receptors endocytosis occurs by the formation of a clathrin-coated vesicle that is produced by the internal folding of the plasma membrane. After the dissociation of the ligand-receptor complex there is a need to return the receptor molecules to the plasma membrane and this is facilitated by one of two pathways: the fast recycling pathway, which involves direct transport of molecules from peripheral sorting or early endosomes to the plasma membrane, and the slow recycling pathway, whereby molecules are returned to the plasma membrane indirectly through a tubulovesicular endosomal-recycling compartment (ERC) that is located close to the microtubule organising centre of the cell. Effective regulation of trafficking of materials into and within the cell depends largely on the complex recycling of receptor molecules within the slow recycling pathway.

VIII.1.1 Rab GTPase's

Recent cytological investigations have highlighted the importance of the Rab GTPase family of proteins (Rabs) in regulating vesicular trafficking (trafficking of vesicles and their cargos within the cell) through their role in membrane trafficking (trafficking of proteins/lipids to cellular membranes).[6-16] For example; Rab1 and Rab2 appear to play a role in the transport of proteins from the endoplasmic reticulum to the Golgi [7; 8]; Rab6 functions in intra-Golgi transport [9]; and Rabs 4, 5, 7, 9, 11 and 15 help regulate the trafficking and sorting of endocytosed material between lysosomes, endosomes and the plasma membrane.[10-12] Of the Rab GTPases Rab11 has been amongst the most extensively researched and is recognised as regulating a host of membrane transport pathways (e.g., phagocytosis [13], apical targeting in epithelial cells [14], protein transport from endosomes to the trans-Golgi network [16] and insulin-dependent glucose transporter 4 transport to plasma membranes). A well-known mechanism of vesicular trafficking is the myosin motor docking to transferrin receptor-containing transport vesicles. Yeast two-hybrid experiments have revealed that Rab11 interacts with the myosin Vb globular tail domain and, also, in the absence of wild-type Rab11, myosin Vb fails to localise to transport vesicles [17], highlighting the importance of Rab11 in membrane trafficking.

VIII.1.2 The Family of Rab11-Interacting Proteins

Numerous so-called effector proteins have been shown to selectively bind to specific Rabs and it is supposed that Rabs are functional only when complexed to these effector proteins. The supposed mechanism of Rab-effector protein complex formation is the GDP(inactive)-/GTP(active)-bound cycling of the Rab GTPase, whereby the GTP-bound Rab alters the conformation of the Rab switch 2 domain relative to the GDP-bound Rab, in a Rab specific manner, and as a result permits binding of the effector protein to the Rab switch 1 domain.[4] A number of effector proteins that bind the Rab11 Rab GTPase have recently been discovered and, based on the similarities of their predicted tertiary structures, have been grouped into what is known as the family of Rab11-interacting proteins (FIPs). Within mammals, FIPs genes show a broad distribution across chromosomes and comprise an evolutionarily conserved protein family.[3] It has been shown that FIP-Rab11 complexes are selectively localised to various membranes within the cell (e.g., recycling endosomes) [18], whereas in the absence of complex formation both proteins remain in the cytosol. It has also been shown that different FIPs compete with each other for binding to Rab11 [19] and, as such, may control Rab11 functioning within the cell - depending on the relative amounts of each Rab11-FIP, Rab11 could be preferentially assigned to one or other of the numerous trafficking pathways in which it functions. It is from these understandings that FIPs, along with their interactions with Rab11, have recently become the subject of considerable cytological and biophysical research.

So far, six FIPs have been identified: Rab11-FIP1 (FIP1) [20], Rab11-FIP2 (FIP2) [20], Rab11-FIP3 (FIP3/eferin) [20], Rip11 [21], Rab coupling protein (RCP) [22] and Rab11-FIP4 [3]. FIPs are characterised by the presence of a C-terminal (CT) Rab11-interacting site, known as the Rab binding domain (RBD) (Figure VIII.1C), which partially overlaps a predicted coiled-coil domain (Figure VIII.1A). FIPs are further classified according to the presence of other structural domains located at the N-terminus. Class I FIPs (Rip11, FIP2 and RCP) contain an N-terminal (NT) C2-phospholipid binding domain (Figure VIII.2A, Figure VIII.1B), class II FIPs (FIP3 and FIP4) contain an NT EF-hand calcium binding motif (Figure VIII.2B) and an ezrin/radixin/moesin (ERM) domain (Figure VIII.2C, Figure VIII.1D), and class III FIPs (FIP1) have no NT domains with homology to other domains (Figure VIII.1A). Despite the above similarities, the degree of homology between the various FIPs is low (Figure VIII.1E) and, therefore, they are considered as a family of diverse proteins with certain common functional motifs.

The Rab11 subfamily of proteins are known to be enriched on the ERC of the slow recycling pathway and are thought to be instrumental in the recycling of receptor proteins. The class II effector protein of Rab11a, Rab11-FIP3, has been seen to localise to the pericentriolar ERC during the interphase stage of cell division and, when complexed to Rab11a, is understood to be involved in the delivery, targeting and/or fusion of the ERC with the cleavage furrow/midbody during cell division.[23-25]

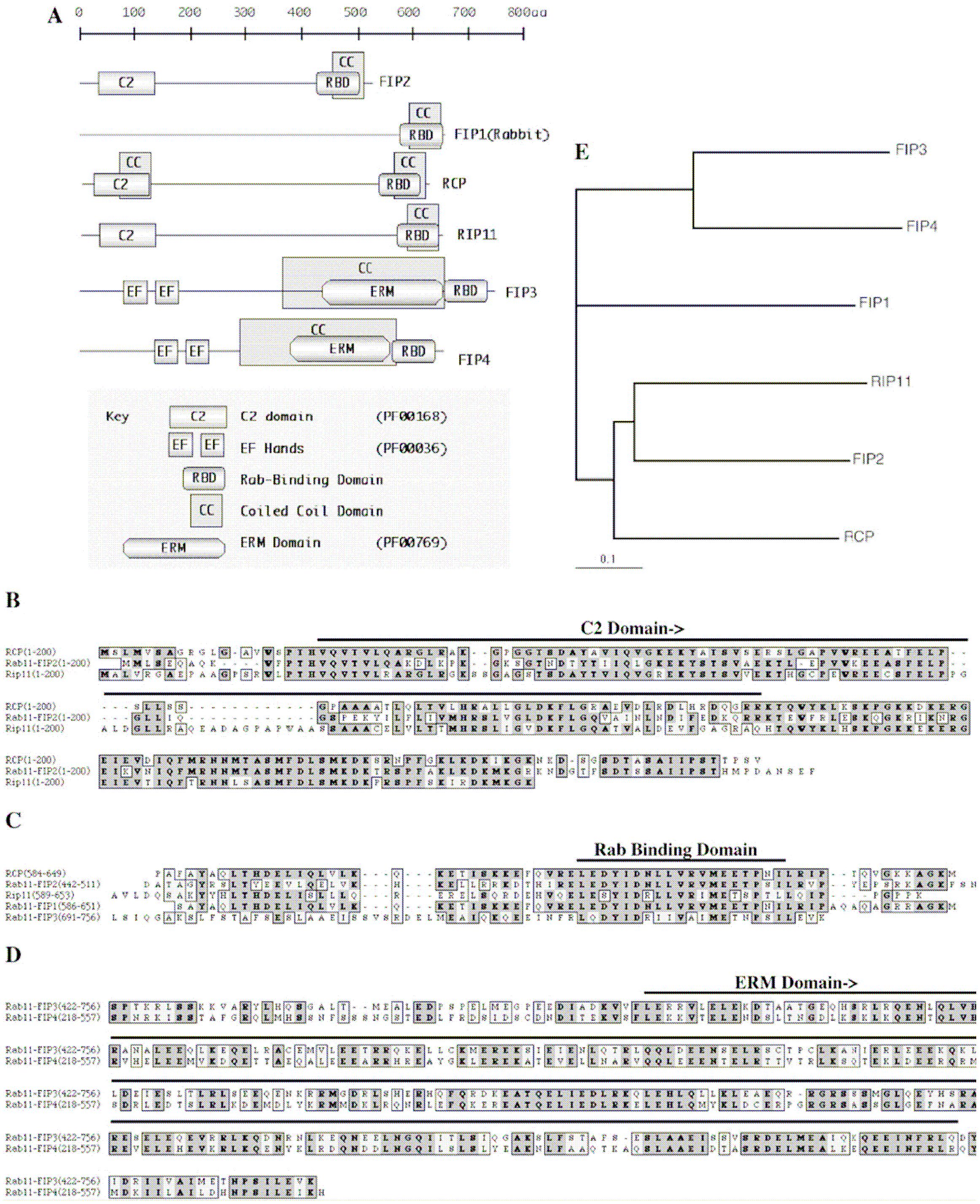


Figure VIII.1. The Rab11-FIP protein family. A Location representation of the various C2, EF, ERM and RBD functional domains within the human Rab11-FIP proteins, including the Pfam reference numbers for each domain. **B** ClustalW alignment of the amino terminal 200 amino acids of RCP, Rab11-FIP2, and Rip11; **C** ClustalW alignment of the Rab Binding Domains of RCP, Rab11 FIP2, Rip11, Rab11-FIP1, Rab11-FIP3, and Rab11-FIP4; **D** Alignment of the ERM domains of Rab11-FIP3 and Rab11-FIP4; **E** Phylogenetic tree. The scale bar (0.1) represents the distance along the tree corresponding to a change in 10% of the amino acids. (Reproduced from [3])

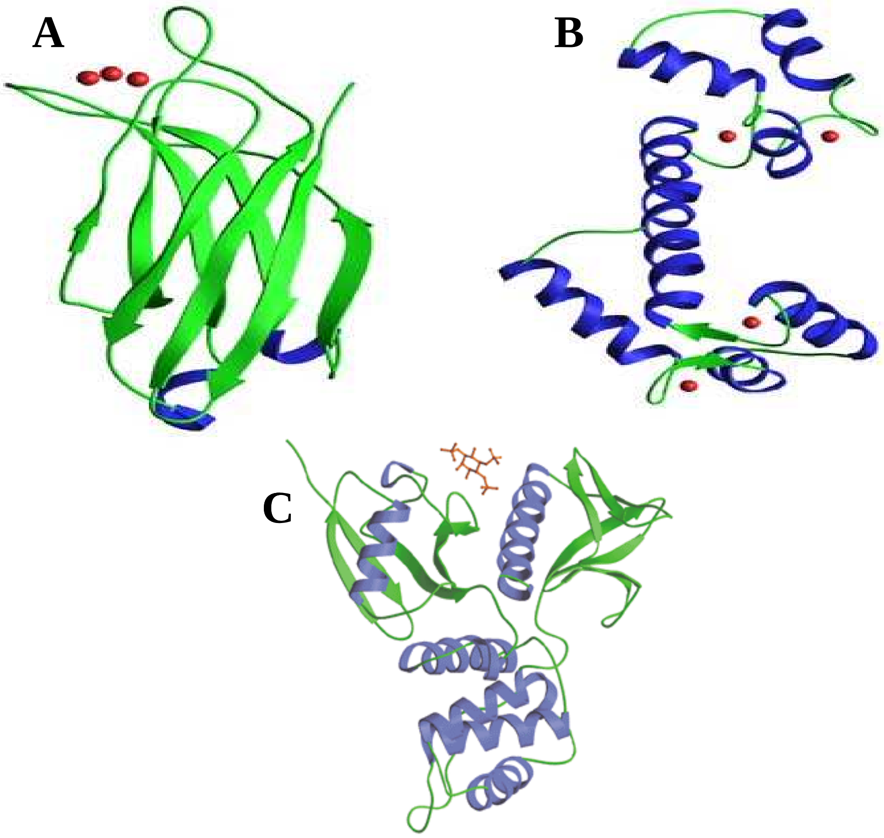


Figure VIII.2. The conserved domains of the Rab11-FIP protein family. α-helices, β-sheets, loop regions and calcium ions/ligands are shown as blue spirals, green ribbons, green strings and in red colour, respectively. **A** The C2 domain is composed of two four-stranded β-sheets, creating three loops at the top of the domain and four at the bottom. Upper loops 1 and 3 are involved in the binding of 3 calcium ions required for phospholipid binding. **B** Each EF-hand domain consists of two perpendicular 10 to 12 residue α-helices with a 12-residue loop region between, forming a single calcium-binding site (helix-loop-helix). Calcium ions interact with residues contained within the loop region. **C** The ERM domain is approximately 150 amino acids in length and is found in a number of cytoskeletal-associated proteins that are found at the interface between the plasma membrane and the cytoskeleton. *(C) 2006 Nash Lab - University of Chicago*

VIII.1.3 The Coiled-Coil Protein Interaction within Rab11-FIPs

In the simplest case coiled-coils involve two right-handed α-helical segments aligning either parallel or anti-parallel to one another and spiralling around each other in a left-handed manner to form a two-stranded superhelix (Figure VIII.3*a*). Each helix region involved in a coiled-coil can be interpreted in terms of both the hydrophobicity and charge of its residues, revealing heptad repeat sequences along the length of coiled-coil forming helices. Such regions are defined by [**a**bc**de**fg]$_n$, with **a** and **d** representing hydrophobic residues and **e** and **f** denoting charged residues (Figure VIII.3*b*). The coiled-coil is stabilised by both hydrophobic and ion-pair interactions (Figure VIII.3*c*) and has at its centre a hydrophobic core (Figure VIII.3*d*).

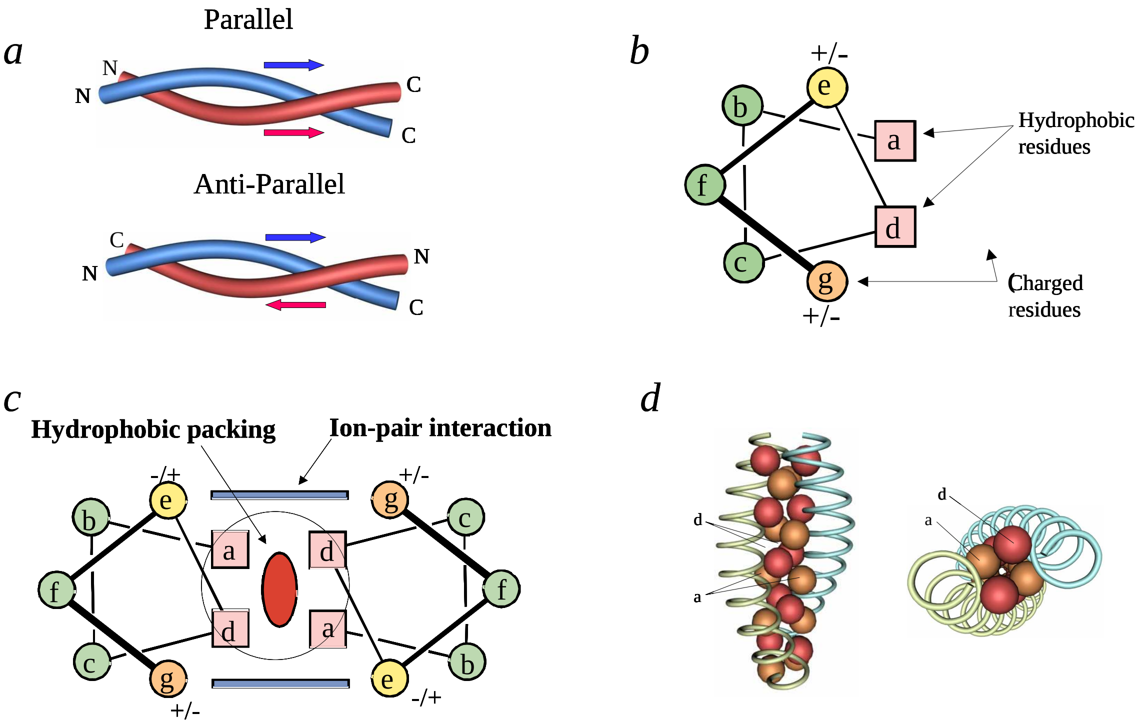


Figure VIII.3. Illustration of a dimeric coiled-coil. *a* Parallel and anti-parallel left-handed spiralling coiled-coil with right-handed helices. *b* Helical wheel representation of a heptad repeat sequence. *c* Coiled-coil stabilising ion-pair interaction and hydrophobic packing. *d* Hydrophobic core within the coiled-coil. (Modified from [1])

According to current understandings the coiled-coil is amongst the most common protein-protein interaction motif. It is particularly common amongst structural proteins and proteins involved in the tethering of vesicles to target organelles – a process that allows subsequent fusion of the vesicle membrane with that of the organelle. As such, FIPs are suspected to function either as scaffolding proteins – facilitating the formation of a membranous protein complex – or as docking proteins. Using immunolabelling, the function of the coiled-coil containing Rab11-FIP3 has been investigated.[5] When complexed to Rab11a, Rab11-FIP3 is seen to localise to membranes of recycling endosomes. The conserved prediction of CT coiled-coils and their overlap with the RBD suggests an importance of coiled-coil domains for the complexing of FIPs to Rab11. For example, Rab11-FIP2 has been shown to form dimers *in vitro* [2], with monomers held together by an intermolecular coiled-coil. Complexing of this homodimer with Rab11 at the RBD has been shown by x-ray crystallography to result in a heterotetramer with dyad symmetry, arranged as a Rab11-(FIP2)$_2$-Rab11 complex [4]. The intermolecular FIP2 coiled-coil domain is, therefore, important in determining both the composition of the complex and its symmetry. Figure VIII.4 gives a cartoon illustration of how the coiled-coil region of Rab11-FIP3 might be instrumental in determining how it interacts with Rab11, based on what is known about complex formation involving the Rab11-FIP2 dimer. Figures VIII.5 and VIII.1A show the locations of the predicted coiled-coil domains for each FIP. Both FIP3 and FIP4 are exceptional in that they appear to contain extensive coiled-coil domains.

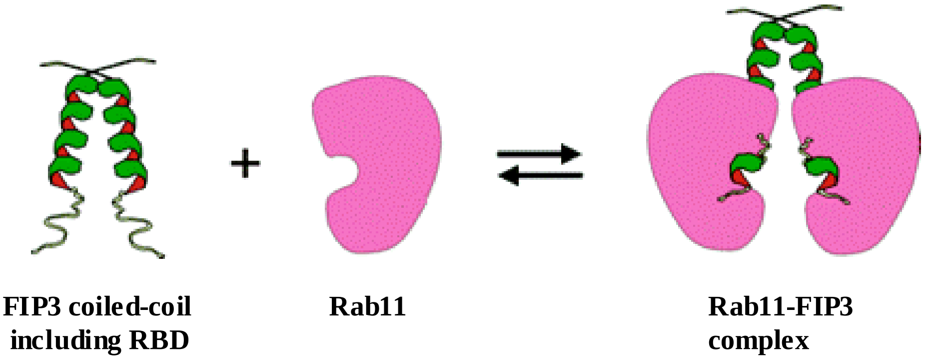


Figure VIII.4. Illustration of the possible symmetry determining influence of the coiled-coil region of the Rab11-FIP3 dimer for the FIP3-Rab11 tetrameric complex. (Modified from [2])

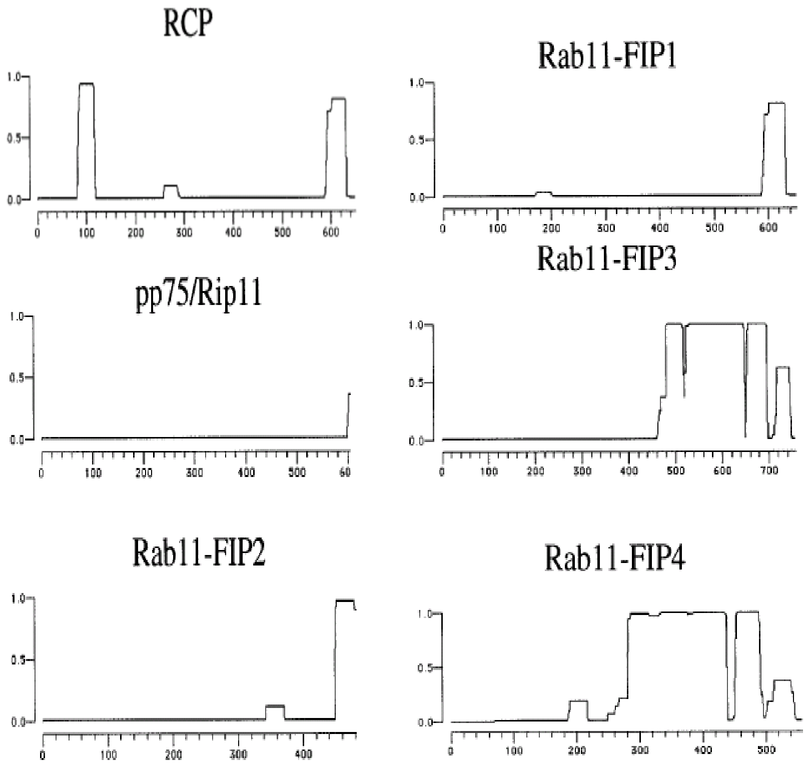


Figure VIII.5. Prediction of coiled-coils within the Rab11-FIP protein family. The residue number is plotted on the x-axis and the probability of forming a coiled-coil (0–1) is on the y-axis. (Copied from [3])

Although C-terminal coiled-coils are unanimously predicted for FIPs, the veracity of these predictions have, for most FIPs, yet to be experimentally determined. The prediction of both the size and location of the Rab11-FIP3 coiled-coil is corroborated by the low probability of disorder for the exact region of the coiled-coil prediction (Figure VIII.6*a*), which must be α-helical if the prediction is correct, and the relative probabilities of the coiled-coil existing as either a dimer or trimer is shown in Figure VIII.6*b*. From this it is evident that Rab11-FIP3 has a much greater probability of existing as a dimer molecule.

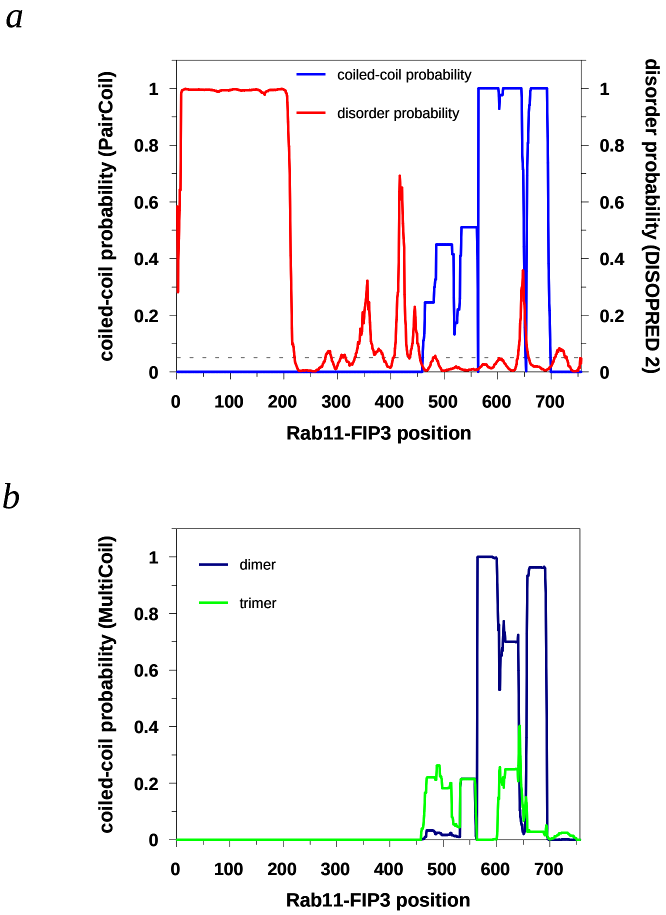


Figure VIII.6. The coiled-coil prediction profile for Rab11-FIP3. *a* Overlay of the probability of coiled-coil domains and the probability of disorder, from PairCoil and DISOPRED 2 algorithms, respectively. *b* Overlay of the probabilities of both dimer and trimer coiled-coil domains, from the Multicoil algorithm. (Copied from [5])

Part A: Rab11-FIP3

VIII.A.1 Introduction

The following far-UV CD work was designed to investigate the existence of a CT coiled-coil for Rab11-FIP3 for the region of the protein showing a high probability for coiled-coil domains. In addition, the structural effect of mutating a specific residue located within the RBD region – a mutation that causes dramatic changes within examined cells - was investigated by comparing the CD spectrum of the wild-type and mutant proteins. The secondary structure fractional composition for both was also estimated.

VIII.A.2 Method

VIII.A.2.1 Protein Expression and Purification

Purification of hexahistidine-fused Rab11-FIP3 wild-type and Rab11-FIP3 I738E was performed by transforming XL1 cells with either pTrcHisC/Rab11-FIP3 or pTrcHisC/Rab11-FIP3 I738E. Transformants were grown to an optical density at 600 nm of approximately (OD_{600}) 0.6 and then induced using 0.1 mM isopropyl-b-D-thiogalactopyranoside (IPTG) (Melford) for 1 hour at 30 ^{o}C. Purification of the predicted coiled-coil structured recombinant fragment Rab11-FIP3$_{(463-692)}$ containing a six-histidine tag was done by first transforming the pTrcHisC/Rab11-FIP3$_{(463-692)}$ plasmid into BL21 (DE3) *e. coli* cells. After growing a single transformant to an OD_{600} of approximately 0.6 the culture was induced with 0.3 mM IPTG for a period of 12 hours at 20 ^{o}C. All proteins were affinity purified using columns containing Ni^{2+}-NTA agarose beads (Qiagen). The purity of each protein preparation was analysed by SDS-PAGE gel electrophoresis with coomassie blue staining.

VIII.A.2.2 CD spectroscopy

All spectra were recorded on a Model 400 Spectrophotometer equipped with a Peltier temperature controller (Aviv Biomedical Inc.). Spectra were obtained using either a 10, 1 or 0.1 mm quartz Suprasil cuvette (Hellma® UK) and at a temperature of 23 ^{o}C, a bandwidth of 1 nm, a stepwidth of 0.2 nm, a response time of 1 s and as an average of at least 5 scans. Protein melt experiments were recorded at 222 nm in a 10 mm cell with a bandwidth of 1.5 nm, a response time of at least 100 s, an equilibration time of

30 s, a temperature step of 2 °C, a heating rate of 2 °C /min and under continuous stirring. The protein structural response to trifluoroethanol (TFE) addition was investigated by the addition of 0, 5, 10, 15, 20, 30, 40, 50 and 60 % volume per volume TFE to protein samples while keeping the concentration of protein constant across all samples. CD spectra were expressed in terms of mean residue molar ellipticity using the equation $[Q]_{MRW} = q/(10.c_r.l)$, where c_r and l are the mean residue molar concentration and the sample pathlength (cm), respectively. Sample protein concentrations were determined by quantitative amino acid analysis (Protein and Nucleic Acid Chemistry Facility, Cambridge).

VIII.A.3 Results

VIII.A.3.1 Protein Expression

An SDS-PAGE gel for a single batch of Rab11-FIP3$_{453\text{-}692}$ protein is shown in Figure VIII.9 and from the protein ladder the sample can be seen to be of the correct size based on the protein primary sequence. An absence of any other bands within the sample lane confirms that the protein sample is free from any peptide contamination.

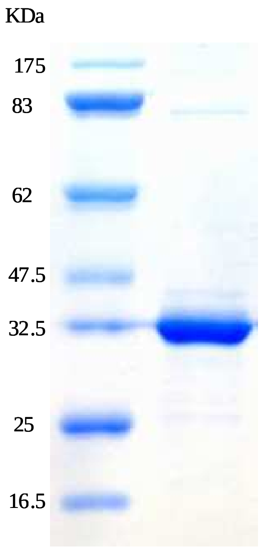


Figure VIII.9. SDS-PAGE gel of the purified Rab11-FIP3$_{463\text{-}692}$ fragment. The protein ladder is shown in the left lane and the purified protein sample is in the right lane.

VIII.A.3.2 Mutant and Wild-Type Rab11-FIP3 CD Spectra

Figures VIII.10 and VIII.11 show the smoothed CD spectra of the wild-type Rab11-FIP3 and Rab11-FIP3 I738E mutant, respectively. The overlaid smoothing residuals reveal both spectra to contain good signal-to-noise, however, this is somewhat diminished at the lower wavelength end of each spectrum. The concentration determination of both the wild-type and mutant samples as 1.27 and 1.25 mg/ml, respectively, enabled the CD spectra to be plotted in terms of the mean residue molar ellipticity. A comparison of the wild-type and mutant CD reveals a close similarity between the two spectra, suggesting that the I738E mutation did not dramatically alter the secondary structure of the Rab11-FIP3 protein. A discrepancy between the bands located at 208 nm for the wild-type versus the mutant CD suggests that there is some structural consequence to carrying out this mutation.

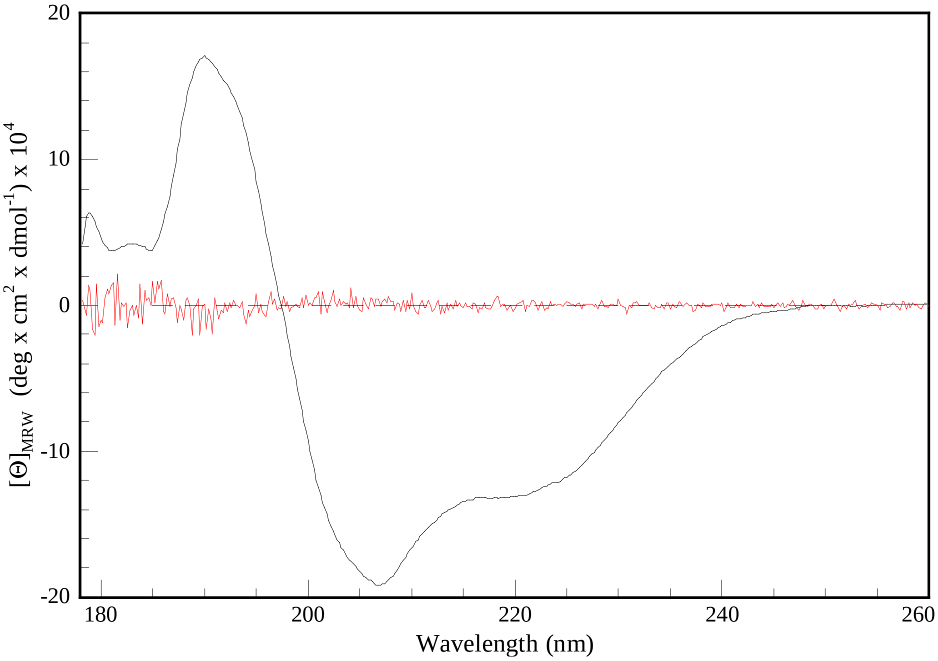


Figure VIII.10. CD spectrum of wild-type Rab11-FIP3. The smoothed CD spectrum is overlaid with the smoothing residuals (red).

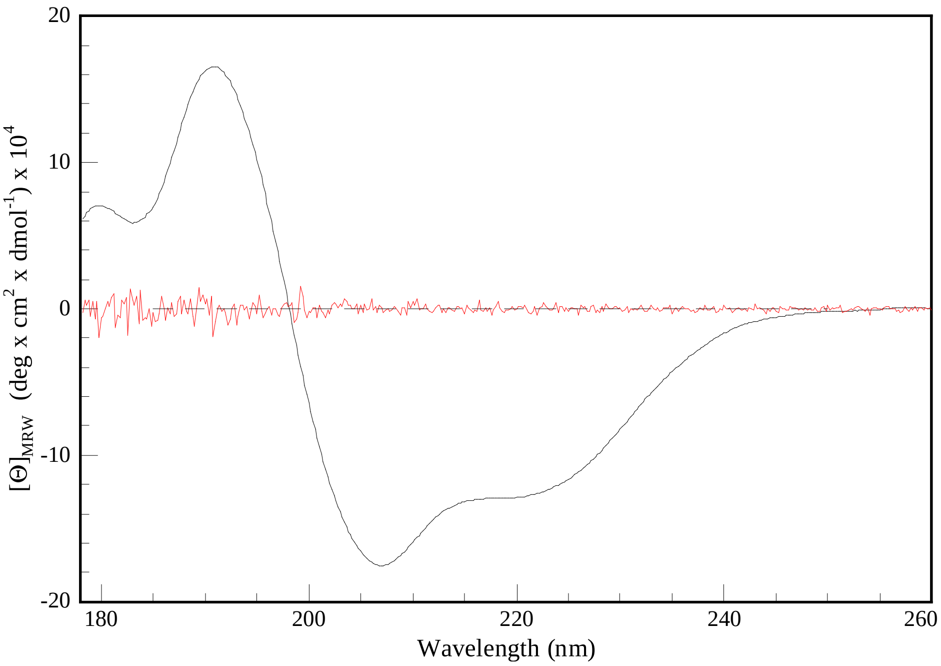


Figure VIII.11. CD spectrum of the Rab11-FIP3 I738E mutant. The smoothed CD spectrum is overlaid with the smoothing residuals (red).

VIII.A.3.3 Quantitative Analysis of Secondary Structure of Rab11-FIP3

Analysis of the fractional compositions of both the mutant and wild-type full-length Rab11-FIP3 proteins with most of the programs contained within the DichroWeb software resulted in NMRSD values outside of the limits for an acceptable deconvolution fit. The coiled-coil nature of these proteins along with the presence of C2-phospholipid, EF-hand and ERM domains are somewhat particular to FIP proteins and are not well represented within the protein reference sets used by the DichroWeb software. It is likely, therefore, that the impact of these domains on the Rab11-FIP3 CD spectra was responsible for the failure of the fractional composition analysis. The CDSSTR and K2D programs did return NMRSD values below the cut-off limit and the results are shown in Table VIII.1. For each program the estimated fractional composition values are consistent between the two proteins, as expected from the similarity of both protein's CD spectra. The α-helix content of both proteins is likely to be closer to the 29 % value predicted by the K2D program, since the CDSSTR program tends to over estimate α-helical content. The uncertainty in the estimation of β-sheet content is a feature of protein CD and all that can be concluded is that both proteins contain small but significant levels of β-sheet structure. The β-turn estimation of both proteins from the CDSSTR program is comparable to that of β-sheet. The K2D program does not estimate β-turn content, but instead includes all non-helical and non-sheet conformations as unordered structure. Reasonable agreement between the unordered contents of both proteins for the K2D and CDSSTR programs is seen when this is taken into account.

Protein	**Program**	**α-helix**	**β-sheet**	**β-turn**	**Unordered**	**NMRSD**
WT Rab11-FIP3	CDSSTR (ref. set 3)	49	15	14	22	0.015
	K2D	29	28	-	44	0.179
Rab11-FIP3 I738E	CDSSTR (ref. set 3)	46	13	17	24	0.019
	K2D	29	19	-	52	0.162

Table VIII.1. Fractional composition analysis of wild-type Rab11-FIP3 and Rab11-FIP3 I738E.

VIII.A.3.4 Detection of Rab11-FIP3 Coiled-Coil Structure by TFE Analysis

The protein concentration of the predicted coiled-coil domain of the Rab11-FIP3$_{463\text{-}692}$ fragment used for the TFE analysis was determined by QAA analysis to be 0.11 mg/ml. The purity of the protein sample is indicated by a single intense band being present on the SDS-PAGE gel at the expected location relative to the protein ladder (Figure VIII.9). Figure VIII.12 shows the CD spectral overlay of the effect of TFE on the Rab11-FIP3$_{463\text{-}692}$ fragment. TFE acts to stabilise peptides towards helix formation by increasing the strength of the H-bonds formed within the helix conformation.[26] Since TFE induces α-helix conformation for protein regions that are of other conformations, the lack of spectral intensity increase at 222 nm up to 20 % v/v TFE indicates that the Rab11-FIP3$_{463\text{-}692}$ fragment is already completely α-helical, as expected. It is known that the spectral signature of an α-helix that has been distorted by its participation within a coiled-coil conformation displays a slightly different CD spectrum than that seen with an undistorted non-interacting typical α-helix. This difference is characterised by the value of the band ratio at 222 and 208 nm. For a non-interacting α-helix the 222:208 ratio is approximately 0.83, but for α-helices with coiled-coil super secondary structure this value changes to about 1.03. The observed 222:208 ratio of Rab11-FIP3$_{463\text{-}692}$ in the absence of TFE is 0.968. This is very close to the coiled-coil expected value of 1.03, indicating that the fragment is composed almost entirely of helices with coiled-coil super secondary structure. Although TFE stabilises secondary structure, it conversely destabilises tertiary structure and as such, titration with TFE might be expected to disrupt any coiled-coil content within the

Rab11-FIP3$_{463\text{-}692}$ fragment and this disruption should be evidenced by a change in the 222:208 ratio from the value of the coiled-coil to that of the non-interacting α-helix. Figure VIII.13 shows the plot of the 222:208 ratio with TFE concentration and from this it is clear that at TFE concentrations beyond 20 % v/v there occurs a dramatic shift of the 222:208 ratio to that expected for non-interacting α-helices. At 50 % v/v TFE it appears from the 222:208 ratio value of 0.88 that almost all coiled-coil structures have unravelled to form non-interacting α-helix structures. As expected, higher TFE concentrations had no additional effect on the 222:208 ratio.

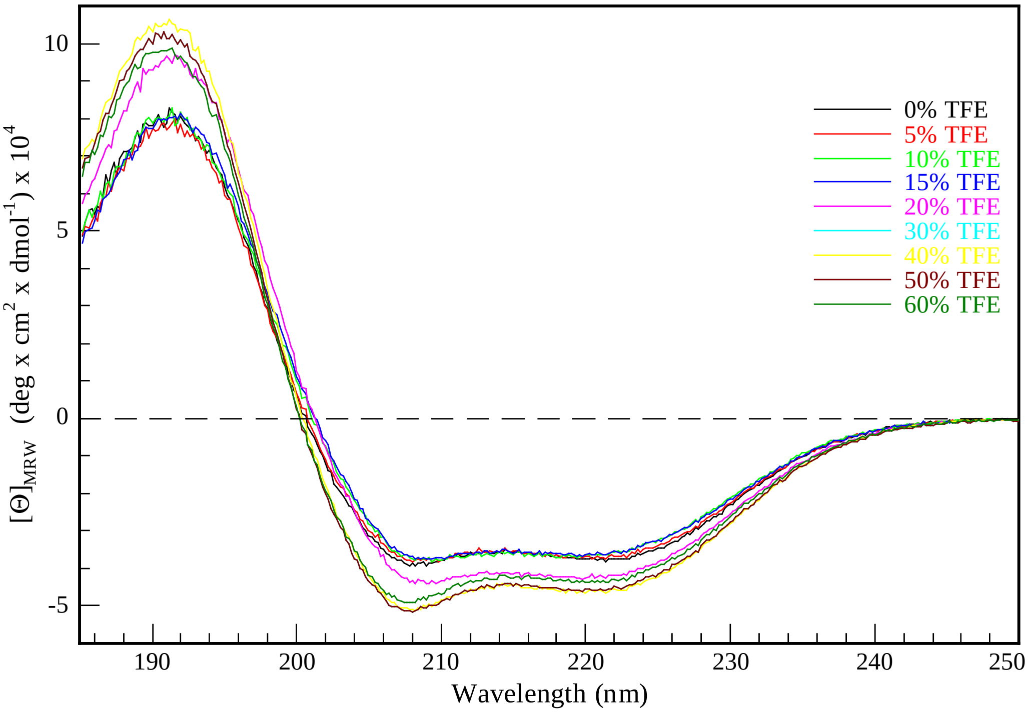


Figure VIII.12. CD spectra of Rab11-FIP3$_{463\text{-}692}$ fragment as a function of TFE concentration.

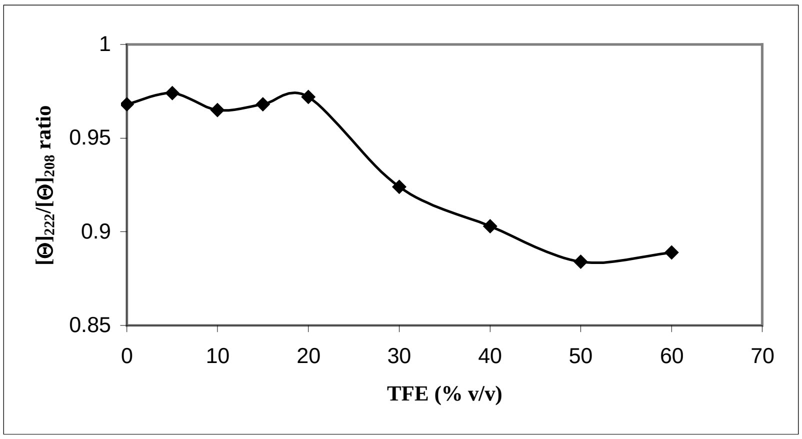


Figure VIII.13. Coiled coil conformation of the Rab11-FIP3$_{463\text{-}692}$ as a function of TFE concentration. The ratio of the CD spectral intensity at 222 and 208 nm is indicative for the presence of coiled coil conformation at lower TFE concentrations.

VIII.A.3.5 The Effect of Protein Concentration on the Coiled-Coil Structure of Rab11-FIP3

The formation of coiled-coil polymers from the monomeric peptide chain raises the question of the effect of protein concentration on protein polymerisation. Although there exists evidence that FIPs interact to form dimeric coiled-coils, it is at least feasible that higher order polymerisation is involved in generating the functional form of the Rab11-FIP3 complex. In order to investigate this, the effect of protein concentration on the structure of Rab11-FIP3 was examined. Initially the CD_{222} value dependence on protein concentration was measured in order to establish the stability of the α-helix secondary structure of the Rab11-FIP3$_{463-692}$ fragment. Figure VIII.14 shows that the Rab11-FIP3$_{463-692}$ helical secondary structure is stable over approximately a 60-fold concentration range. The ability of the Rab11-FIP3$_{463-692}$ fragment to form higher order polymers might be revealed from increasing the protein concentration while observing the $[\Theta]_{222:}[\Theta]_{208}$ ratio, which is dependent on the tertiary structure of the protein. From Figure VIII.15 it would seem that this ratio is somewhat stable over the concentration range investigated and that the Rab11-FIP3$_{463-692}$ fragment is, therefore, resistant to forming higher order polymers.

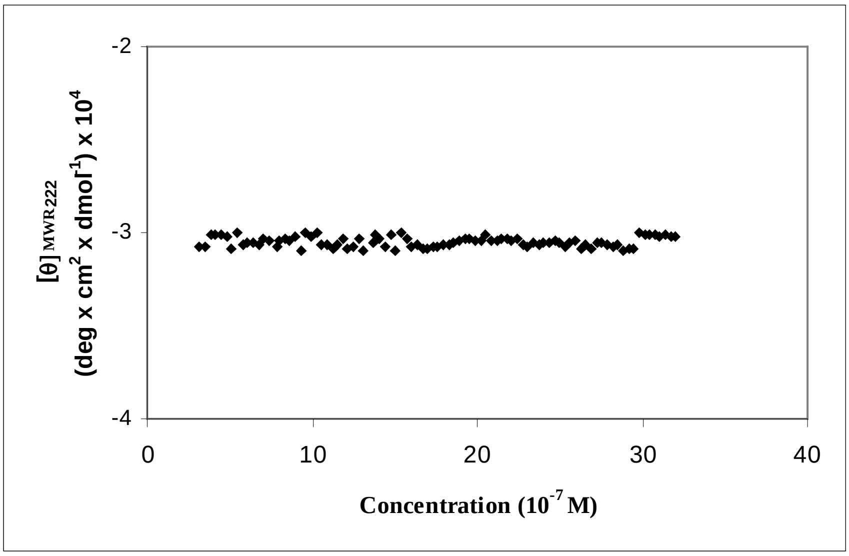


Figure VIII.14. Rab11-FIP3$_{463-692}$ secondary structure dependence on concentration. The dependence of Rab11-FIP3$_{463-692}$ structure on concentration can be assessed by monitoring the CD_{222}, since it is a completely α-helical protein.

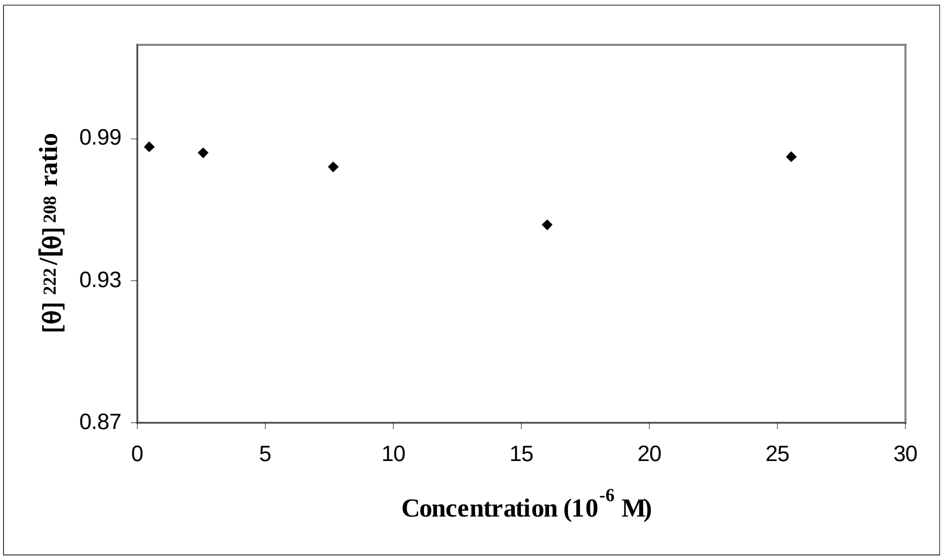


Figure VIII.15. Coiled-coil structural dependence on protein concentration.

VIII.A.4 Discussion

The mutation of the amino acid number 738 from the wild-type hydrophobic isoleucine to the hydrophilic glutamic acid falls within the RBD and has been previously shown to prevent complex formation of Rab11a with Rab11-FIP3.[25; 27] Since Rab11-FIP3 only functions when complexed to Rab11a the effect on cellular activity due to this mutation might be expected to be dramatic. Indeed it has been shown that A431 cells containing this mutation behaved as though they had a complete absence of Rab11-FIP3 protein, in that the morphology of the ERC is profoundly altered, as evidenced by the exclusion of the ERC-marker proteins RCP, Rab11a and Rab11-FIP4 from the pericentrosomal region of the cell, revealed by confocal immunofluorescence microscopy.[5] The fact that the CD spectrum of the Rab11-FIP3 I738E mutant is only marginally different from that of the wild-type protein demonstrates the high structural similarity between the two proteins. Therefore, it can be concluded that the binding of Rab11a to Rab11-FIP3 exhibits very high structural specificity. This is somewhat expected considering the highly conserved sequences of the RBD domains between the known Rab11 effector proteins (Figure VIII.1C). In fact the isoleucine 738 hydrophobic residue forms part of a non-polar pocket in which the isoleucine 44 residue of switch 1 of the Rab11a fits upon FIP3-Rab11 complex formation.[28] Substitution with the polar glutamic acid at this position should, therefore, constitute a severe destabilisation towards complex formation, since this would bring about a strong repulsive interaction with the non-polar Ileu44 of Rab11a. The coiled-coil nature of Rab11-FIP3 as confirmed by the CD analysis of the Rab11-FIP3$_{463\text{-}692}$ fragment is consistent with structural predictions from the PairCoil and MultiCoil algorithms and is in good agreement with the crystal structure of the related Rab11-FIP2 protein.[4] The estimated α-helix content of 29 % by the K2D program for both the wild-type and mutant Rab11-FIP3 proteins is in excellent agreement with the 28.50 and 26.13 % estimations, respectively, from FTIR spectroscopy.[5]

Part B: Rab11-FIP2

VIII.B.1 Introduction

The class 1 FIP Rab11-FIP2 also binds Rab11a and is a 512 residue protein that contains a myosin Vb binding region, an RBD at the C terminus and a C2 domain at the N terminus. The FIP2-Rab11 complex containing the Rab11a (1-173) and Rab11-FIP2 (410-512) fragments has been recently characterised by x-ray crystallography and its structure is shown in Figure VIII.16. The FIP2-Rab11 complex is composed of a homodimer of Rab11-FIP2 with each Rab11-FIP2 monomer bound to a single Rab11a such that the overall structure of the complex is represented as a heterotetramer with dyad symmetry. Binding of each Rab11-FIP2 monomer to a Rab11a protein occurs at the RBD region of the Rab11-FIP2 and the switch 1 region of the Rab11a. The Rab11a switch 1 region is made accessible to the Rab11-FIP2 RBD by a structural shift of the Rab11a upon binding of a phosphate group (converting it from the GDP to the GTP state) such that the blocking Rab11a switch 2 region is relocated. No binding between either of the two Rab11a proteins occurs in the complex and the FIP2-Rab11 heterotetramer is held together by the coiled-coil interaction of the Rab11-FIP2 homodimer. This coiled-coil domain is, therefore, thought to be crucial to the cellular function of the FIP2-Rab11 complex.

As illustrated above, coiled-coil domains are stabilised by the shielding of the hydrophobic residue side-chains from the aqueous environment to form a hydrophobic core running along the axis of the coiled-coil. This hydrophobic core is further shielded from the environment by the formation of a cap at the top of the coiled-coil. Figure VIII.17 shows the N-terminal segment of the Rab11-FIP2 homodimer and from this it can be seen that the Tyr453 phenyl ring forms a lid over the coiled-coil. The hydroxyl group of Tyr453 is hydrogen bonded to the oxygen of the carbonyl group of Leu451 and the coiled-coil lid is completed by participation of Val456, which lies one helix turn down from Leu451. The residues Thr452-Val456 are considered to all be involved in the coiled-coil cap formation and they are highly conserved in FIPs.

Three distinct FIP2-Rab11 complexes are examined in the following work. The complex termed ”A1” consists of two Rab11a$_{1\text{-}173}$ wild-type fragments and the Rab11-FIP2$_{410\text{-}512}$ wild-type fragment homodimer. The complex designated as “Classic” corresponds to two Rab11a$_{1\text{-}173}$ Q70L mutant fragments bound to the Rab11-FIP2$_{410\text{-}512}$ wild-type fragment homodimer. The “M1” complex is composed of two Rab11a$_{1\text{-}173}$ Q70L mutant fragments bound to the Rab11-FIP2$_{410\text{-}512}$ V456G L457G twice-mutated fragment homodimer. In addition two Rab11-FIP2 homodimers were investigated. These were the full-length versions of the Rab11-FIP2 dimer fragments present in the Classic and M1 complexes and, therefore, they are referred to as either the Classic or M1 homodimer. Here, the Classic homodimer is just the Rab11-FIP2 wild-type version.

The purpose of performing the Rab11a$_{1\text{-}173}$ Q70L mutation is to “activate” the Rab11a fragment towards complex formation with the Rab11-FIP2 homodimer, such that the Rab11a fragment adopts a conformation similar to the GTP-Rab11-FIP2 version, with the switch 2 domain delocalised from the switch 1 domain, allowing the switch 1 domain to interact with the Rab11-FIP2 homodimer. It is understood, therefore, that the A1 complex sample will not consist of a heterotetrameric complex at all, but rather will be made up from a mixture of the Rab11-FIP2$_{410\text{-}512}$ wild-type fragment homodimer and non-interacting Rab11a$_{1\text{-}173}$ wild-type fragments, since the Rab11a fragment is present in its inactive GDP form and complex formation with Rab11-FIP2 will be prevented by the inhibiting effect of switch 2 on switch 1.

The reason for replacing the valine and leucine residues of the hydrophobic cap of Rab11-FIP2 with glycine residues is to restrict the steric hindrance effects, created by the side-chains of valine and leucine, to water molecules from coming into contact with the hydrophobic core of the coiled-coil. Glycine is an effective replacement because it does not have any side-chain, other than that of a hydrogen atom, and, therefore, allows maximum exposure of the hydrophobic core to water molecules with mutation of the above residues.

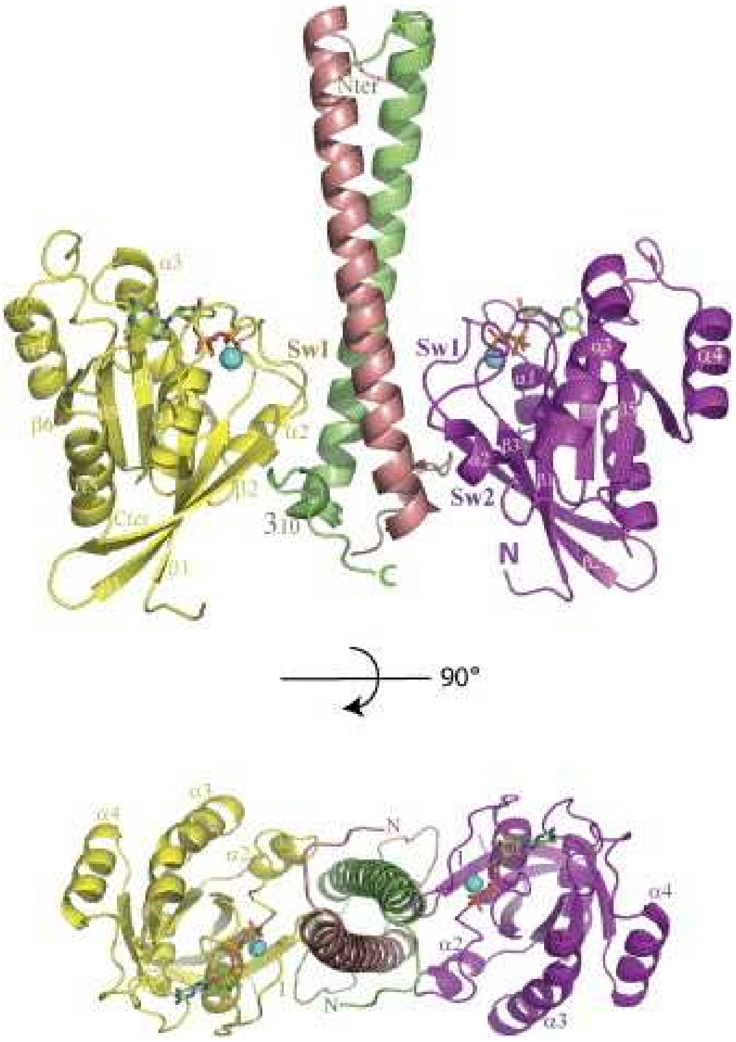


Figure VIII.16. Ribbon Model of the FIP2-Rab11 Complex. (Top) Rab11a molecules are yellow and magenta, while FIP2 is coloured dark pink and green. Switch 1 and switch 2 are indicated, GTP is represented as a stick model, and the conserved Mg^{2+} ion is drawn as a sphere. The short 3_{10} helix at the C terminus of FIP2 (green) is also labelled. (Bottom) View of FIP2-Rab11 rotated 90° in order to show the 2-fold b axis in the crystal. (Copied from [4])

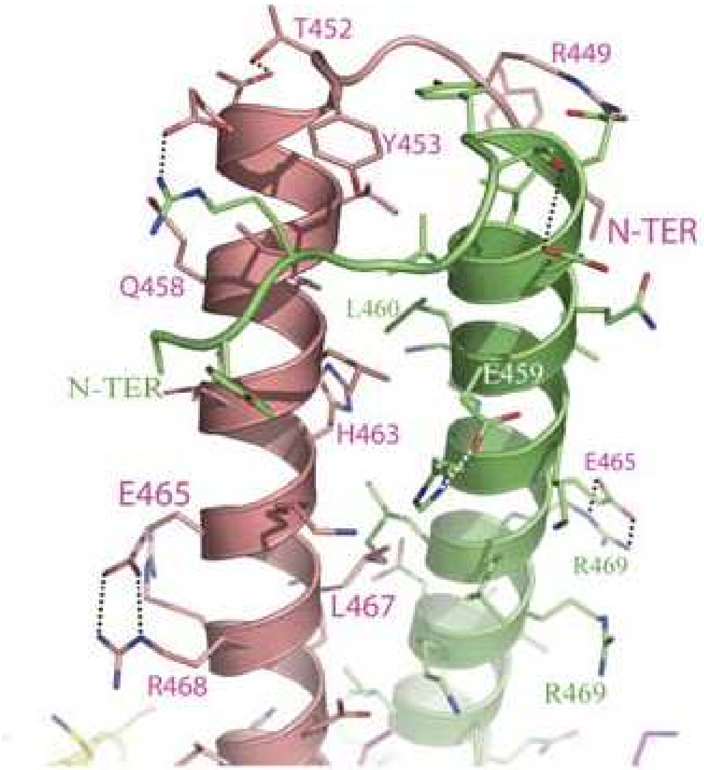


Figure VIII.17. The N-terminal coiled-coil cap for the Rab11-FIP2 homodimer. The tyrosine residue 453 can be seen to form a lid over the hydrophobic core with its aromatic ring side-chain. The orientation of this side-chain, however, is understood to depend on interactions with other cap-forming residue side-chains. (Copied from [4])

In order to examine the stabilising effect of the Rab11-FIP2 cap on its coiled-coil domain, mutants of the complex shown in Figure VIII.16 were created and the structures and stabilities of these mutants were then analysed by both CD and FTIR spectroscopy.

VIII.B.2 Method

VIII.B.2.1 Protein expression and purification

For each individual complex mutant the relevant Rab11-FIP2 and Rab11 plasmids were co-transformed into BL21(DE3) cells and the proteins were co-expressed. For expression of the Rab11-FIP2 homodimer mutants only the relevant Rab11-FIP2 plasmid was transformed into the BL21(DE3) cells. Protein expression was performed using SeMet Media (Molecular Dimensions) with addition of 100 μg/ml ampicillin and 30 μg/ml kanamycin at a temperature of 37 °C. After cell growth had proceeded to an OD_{600} of 0.6 the culture was induced by addition of 0.5 mM IPTG. Cells were incubated at 37 °C for a further 3 hours, harvested by centrifugation and stored at –20

^{o}C. Frozen pellets were re-solublised in MBP extraction buffer (20 mM Tris-HCl, 200 mM NaCl, 5 mM $MgCl_2$ and 10 mM β-mercaptoethanol (pH 7.8)) and then sonicated twice for 1 minute at room temperature. Lysates were centrifuged at 20,000 x g to remove cell debris and the resultant supernatant was applied to an amylose resin (New England Biolabs). Following repeat washing with MBP extraction buffer, the proteins were eluted using the same buffer but with the addition of 10 mM maltose. The eluted protein was then dialysed against 10 mM Tris buffer (pH 8.0), 25 mM NaCl in the presence of rTEV protease (20 μg/ml fusion protein) overnight. The cleaved protein was loaded onto an ion-exchange column (MonoQ GL 5/50, GE Healthcare) and a 10-500 mM NaCl salt gradient was applied over a 20-fold excess column volume. The FIP2-Rab11 complex protein fractions were collected, pooled and re-purified using a Superdex 200 16/60 column (GE Healthcare) equilibrated in column buffer (10 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT (pH 8.0)).

VIII.B.2.2 Site-Directed Mutagenesis

Site-directed mutagenesis was performed using the QuikChange kit (Stratagene) using a pEGFP-C1 Rab11-FIP2 template and incorporation of the mutations was confirmed by sequencing. In total three varieties of the FIP2-Rab11 complex were produced; Rab11$_{1\text{-}173}$ WT complexed to Rab11-FIP2$_{410\text{-}512}$ WT; Rab11$_{1\text{-}173}$ Q70L complexed to Rab11-FIP2$_{410\text{-}512}$ WT; and Rab11$_{1\text{-}173}$ Q70L complexed to Rab11-FIP2$_{410\text{-}512}$ V456G L457G. Two varieties of full-length Rab11-FIP2 homodimer were expressed; Rab11-FIP WT and Rab11-FIP2 V456G L457G.

VIII.B.2.3 CD spectroscopy

All CD spectra were acquired on a Model 215 CD spectrometer. The Classic and M1 CD spectra were acquired using a 0.01 mm pathlength quartz demountable rectangular cell at a temperature of 25 ^{o}C, a bandwidth of 1 nm, an averaging time of 1 sec, a scan increment of 0.2 nm and as an average of 12 scans. Both spectra were background corrected with the CD spectrum of the buffer flow-through from the protein purification step, as acquired under identical spectral parameters as those of the sample. Protein concentration determination was by Bradford assay. Protein CD melt experiments were performed by recording the CD of each protein at 222 nm while varying the temperature from 20 to 90 ^{o}C in 2 ^{o}C increments using a 10 mm pathlength quartz cell. A bandwidth of 1 nm and an averaging time of 20 s were used for each melt, with an equilibration time set at 2 minutes between each temperature change step. The protein concentration for each of the three melts was adjusted so that the CD_{222} at the start temperature of 20 ^{o}C was the same in each case. $T_{1/2}$ values were calculated for each melt by observing the temperature at which the CD_{222} had decreased by 50 % of the total CD_{222} decrease in going from 20 to 90 ^{o}C. For all CD measurements the dynode voltage was monitored and kept to within the instrument specification of 500 volts.

VIII.B.2.4 FTIR spectroscopy

All FTIR spectra were collected by the ATR technique using the BioATR II accessory, on a Tensor 27 FTIR spectrometer and using the OPUS software. Spectra were acquired in double-sided forward-reverse absorption mode by subtracting the transmittance spectrum of the protein sample buffer solution from the sample transmittance. Spectral parameters were set at a Blackman-Harris Fourier transform, an aperture of 6 mm, a resolution of 4 cm^{-1}, a wavelength range of between 1000-4000 cm^{-1}, a temperature of 20 ^{o}C and as an average of 700 scans. The BioATR II was purged with dry air before and during spectral acquisition. Spectra were processed by zeroing at 1750 cm^{-1} and correcting for atmospheric gas contribution using an existing software algorithm. Spectral second derivatives were calculated within the OPUS software as a band-narrowing technique.

Protein melt experiments were conducted by ramping the temperature of the BioATR II accessory using a water bath in increments over the temperature range of the melt. A temperature equilibration time of 2 minutes was set for each temperature step. Protein concentrations were adjusted such that they were approximately equal for the different melt experiments. The Classic FIP2-Rab11 complex melt was performed over a temperature range of 20-90 ^{o}C in 5 ^{o}C increments. The M1 complex melt was over a temperature range of 20-85 ^{o}C with an increment of 5 ^{o}C between 20-50 and 65-85 ^{o}C and 1 ^{o}C between 50-65 ^{o}C. The Classic and M1 homodimer melts were between 20-85 ^{o}C with a 5 ^{o}C increment between 20-30 and 60-85 ^{o}C and a 1 ^{o}C increment between 30-60 ^{o}C. To account for the variable protein adsorption to the ATR crystal surface with temperature increase, the amide-I band was area-normalised at each temperature increment before the calculation of second derivative spectra. The helix stability profiles from FTIR data were created using a Quant algorithm individually generated for each melt experiment. This made use of the helix band areas of the second derivative FTIR spectra in order to establish the helix content present at each temperature as a function of the total change in helix content over the entire melt temperature range.

VIII.B.3 Results

VIII.B.3.1 Rab11-FIP2 Complex CD Spectra

Figure VIII.18 shows the CD spectral overlay of the Classic and M1 FIP2-Rab11 complexes. For both spectra the signal to noise ratio is good and the CD errors are low in both cases, although they do increase slightly in moving from higher to lower wavelengths. Unfortunately accurate protein concentration determination was not performed and both spectra were created using values obtained from the Bradford technique. This creates the danger of introducing an error into the absolute magnitude values of both protein CD spectra. However, the ATR-FTIR analyses of both samples reveal them to be of almost identical protein concentration (Figure VIII.19) and so an accurate comparison of the two CD spectra can be made.

The CD spectra of the Classic and M1 complexes are very similar and from each spectrum it can be seen that both complexes are relatively rich in α-helix content, as indicated by the magnitude of the negative CD bands at 222 and 208 nm. From the relative magnitudes of these negative bands to the positive CD band(s) below 200 nm, both complexes appear to contain a minor, but considerable, degree of random coil content. The broadness and band-shape of these positive CD bands suggest that they are a composite of more than one band. Since both α-helices and β-sheets display positive CD bands in this region, which have different band maxima, it would seem that each complex has some β-sheet content. There does occur a quite sizeable difference between the two CD spectra at the low wavelength end of each spectrum and this may reflect a real difference of structure between the Classic and M1 complexes. However, it should be noted that the CD error values are largest in this region.

As discussed above for the case of Rab11-FIP3, the presence of a coiled-coil super-secondary structure within a protein has the effect of altering the magnitude ratios of the 222 and 208 CD bands resulting from the α-helical secondary structure, since a coiled-coil domain involves slightly distorted α-helix segments. Both complexes have a 222:208 band ratio of 1.12, a value that is closer to the coiled-coil expected value of 1.03 than to the non-interacting helix value of 0.83. However, the presence of other structures within each complex will have an effect on this 222:208 ratio and, as such, this value cannot be taken to be purely reflective of the extent of coiled-coil interaction within the complex. In spite of this, the similarity of the 222:208 ratios of the Classic and M1 complexes indicate a lack of any significant difference of coiled-coil content.

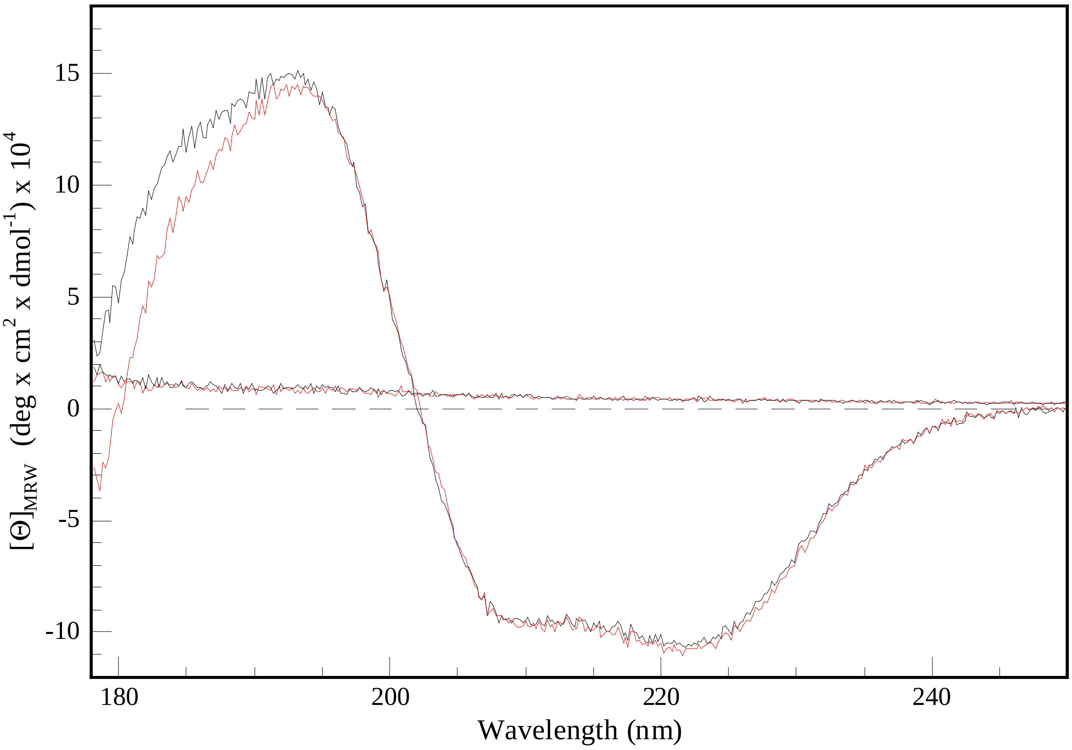


Figure VIII.18. CD spectrum of Classic and M1 FIP2-Rab11 complexes. The Classic CD and CD error are plotted in red and the M1 CD and CD error are in black.

VIII.B.3.2 Rab11-FIP2 Complex FTIR Spectra

The ATR-FTIR absorption spectral overlay in the amide-I band region of both the Classic and M1 complexes is shown in Figure VIII.19. The similarity of both amide-I bands is consistent with the results of the CD analysis. The approximately equal intensities of both bands reveals how similar the protein concentration is for both the Classic and M1 protein samples. From the bandshape of each amide-I spectra both complexes can be seen to contain significant levels of both α-helix and β-sheet secondary structures, as revealed by the band intensities at around 1655 and 1632 cm^{-1}, respectively.

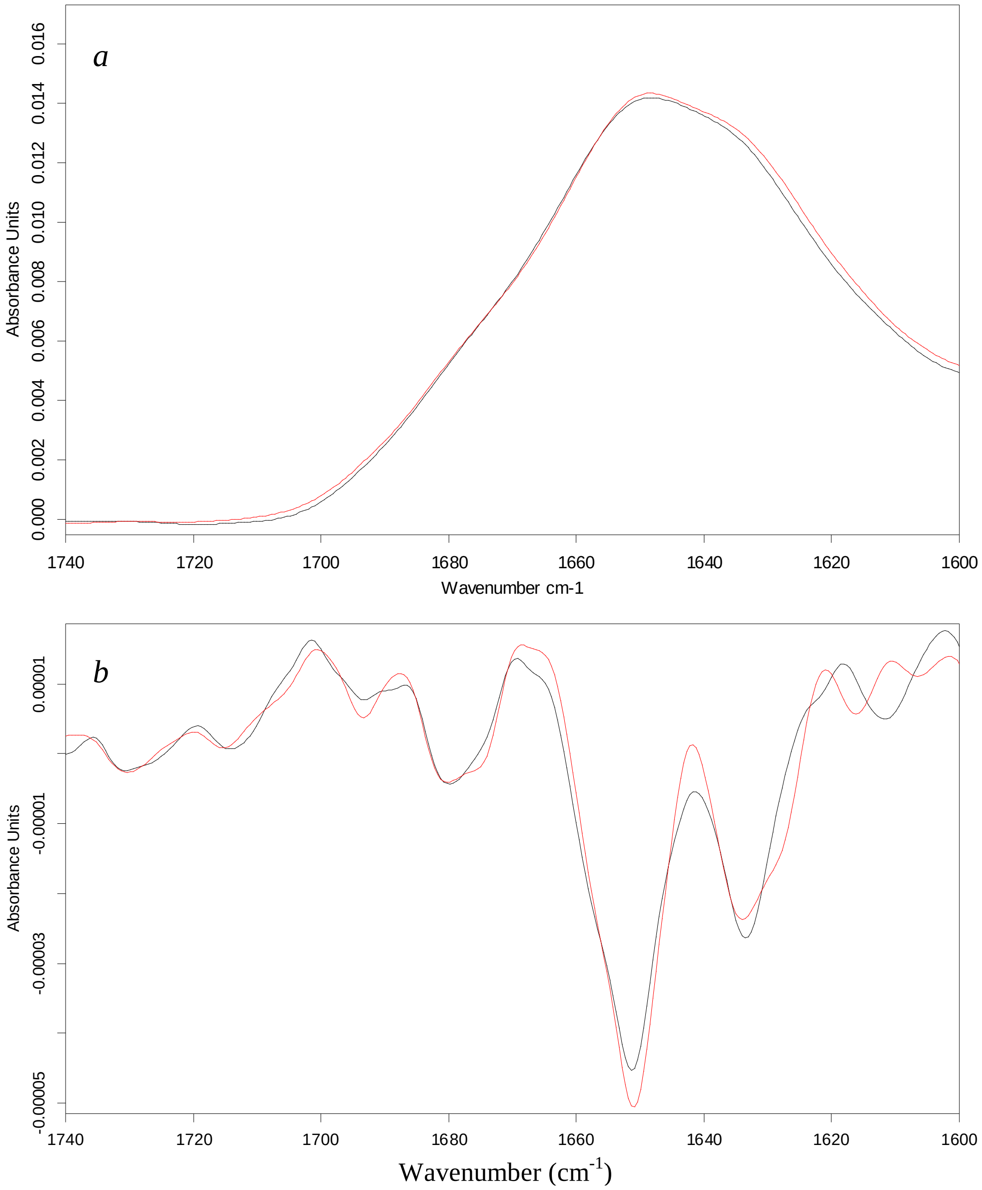


Figure VIII.19. The amide-I bands of the Classic and M1 Rab11-FIP2 complexes. Shown are the absorption (*a*) and 2nd derivative amide-I spectral overlays (*b*) of the Classic (red) and M1 (black) complexes.

The amide-I 2nd derivative spectrum reveals a major band at 1652 cm^{-1}, corresponding to the α-helical amide-I signal. The next most intense band is that located at 1632 cm^{-1} produced by the presence of intra-molecular β-sheet structure. The minor β-sheet band is also likely present within the band occurring around 1680 cm^{-1}. The majority of the intensity of this band, however, probably results from the occurrence of β-turn regions within both complexes. The lack of protein inter-molecular β-sheet aggregates within both samples is revealed by a general absence of band intensity between 1620-1625 cm^{-1}. The most notable difference between the two spectra is the seemingly higher α-helical content of the Classic complex compared to the M1 complex, as seen from the relative intensities of each of the bands centred at 1652 cm^{-1}. This difference can also be detected, although less noticeably, in the CD data (Figure VIII.18).

VIII.B.3.3 Rab11-FIP2 Homodimer FTIR Spectra

The structural perturbation created by mutating two of the coiled-coil cap-forming residues of the Rab11-FIP2 protein was also investigated by recording the ATR-FTIR spectra of both the Classic (Figure VIII.20) and M1 (Figure VIII.21) full-length Rab11-FIP2 homodimers. As expected, comparison of the respective homodimers with their complex equivalents shows, from the increased relative intensity of the helix amide-I component, that the homodimers contain considerably higher fractions of α-helix over other structures than do the complexes. Performing the line-narrowing second derivative procedure for both spectra allows comparison of the fractional contents for both homodimers (Figure VIII.22). Similar to the case of the complexes, there exists a close similarity between the second derivative spectra for both proteins. In contrast to that seen for the complexes, the M1 homodimer seems to contain a slightly higher proportion of α-helix than the Classic. This is also against expectation, since cap mutation would be expected to destabilise the α-helix segments involved in forming the coiled-coil for the M1 mutant.

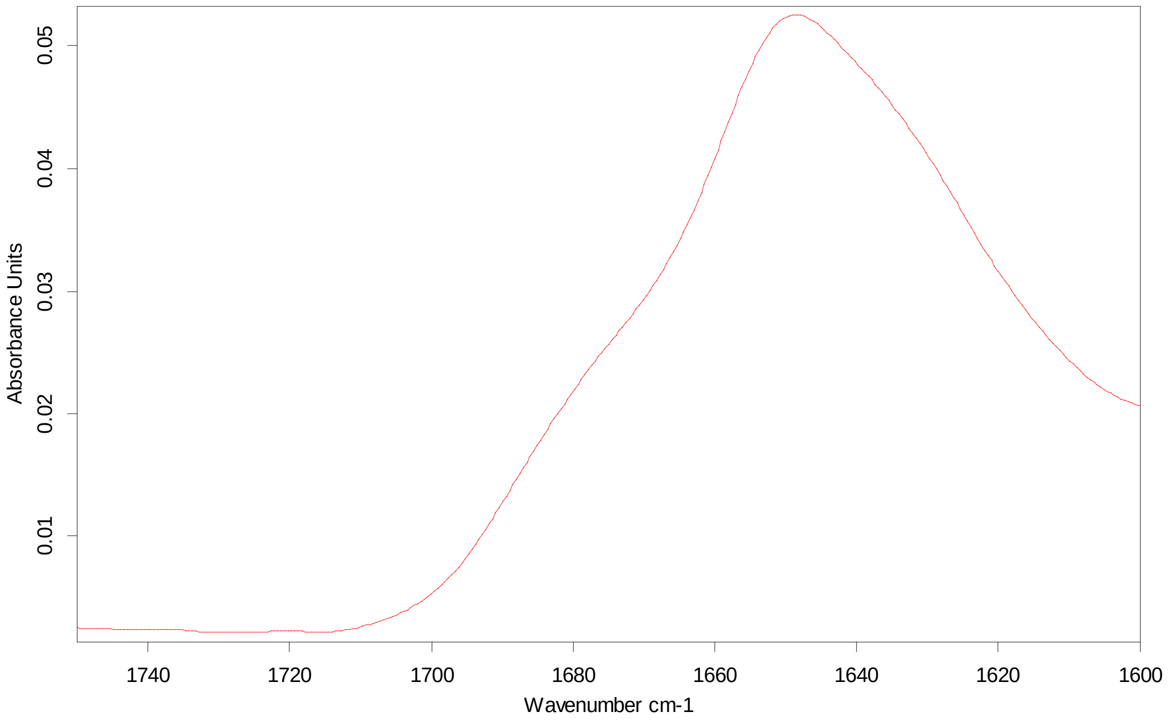


Figure VIII.20. Classic Rab11-FIP2 homodimer amide-I FTIR absorption spectrum.

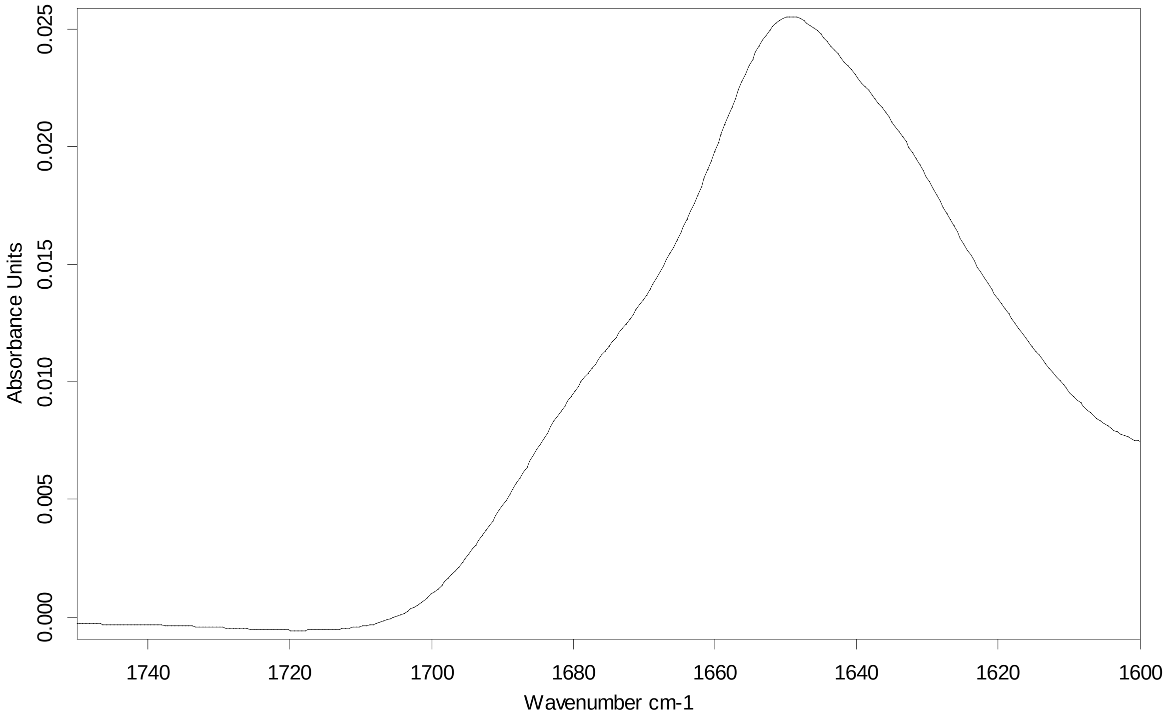


Figure VIII.21. M1 Rab11-FIP2 homodimer amide-I FTIR absorption

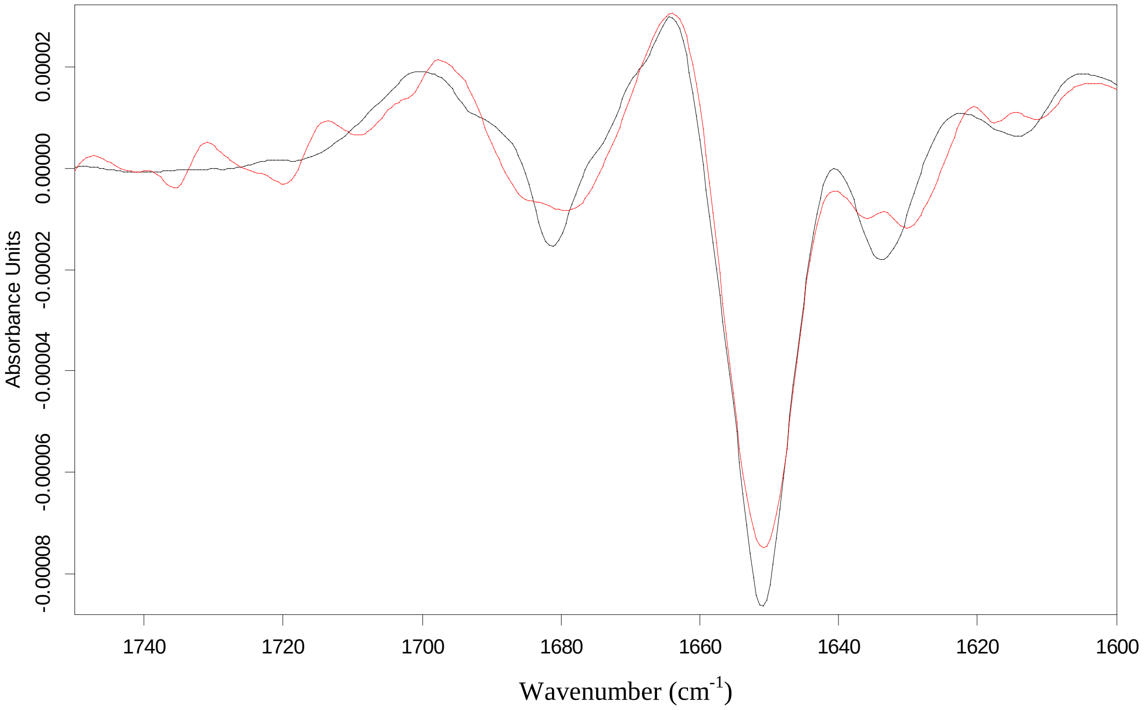


Figure VIII.22. Classic and M1 homodimer second derivative area-normalised amide-I FTIR spectral overlay. The Classic and M1 spectra are plotted in red and black, respectively.

VIII.B.3.4 Rab11-FIP2 Complex CD Melts

It often occurs that the effect of performing a mutation of one or two residues is not sufficient to substantially disrupt the secondary or tertiary structure of an entire domain, but instead the biggest influence of such mutations is to alter the stability of the domain. In order to establish the relative stabilities of each of the complexes a CD melt experiment was conducted, where the CD signal at 222 nm was recorded as a function of temperature. It was decided to monitor the change in ellipticity at 222 nm because this is a characteristic wavelength for the α-helix CD and the mutations performed on the Rab11-FIP2 monomers were such as to disrupt the helical regions of each protein. Inspection of Figure VIII.23 reveals a large shift in the stability of the M1 complex relative to the Classic. For the case where the coiled-coil cap has been disrupted by mutations it would be expected that the resulting complex would display a less stable CD melt curve, since the stabilising effect of the coiled-coil hydrophobic core is likely to be lessened in such an instance. The protein CD melts are, therefore, consistent with the identification of the mutated residues as being highly important for the formation of a hydrophobic-core-protecting cap. The intermediate stability of the A1 mutant is evidence that the formation of the FIP2-Rab11 complex confers greater stability to the Rab11-FIP2 homodimer and/or Rab11a monomers than if these proteins existed as non-interacting proteins in solution.

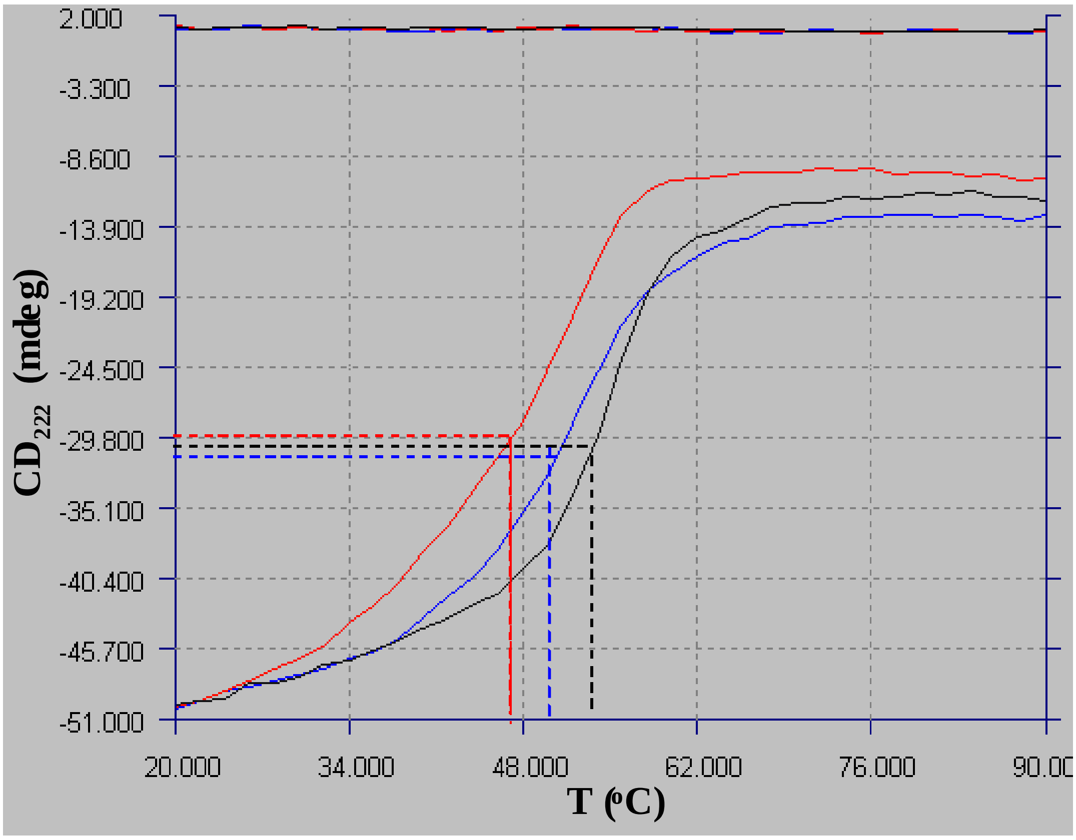


Figure VIII.23. CD temperature melt experiments of the three FIP2-Rab11 complexes. The ellipticity was recorded at 222 nm for each temperature for the three complexes. The Classic, M1 and A1 temperature melts are plotted as black, red and blue, respectively. $T_{1/2}$ values for the Classic, M1 and A1 complexes were calculated as 53.5, 47 and 51 °C, respectively. The CD error values are overlaid along the zero-line with the same colour scheme as given above.

VIII.B.3.5 Rab11-FIP2 Complex FTIR Melts

The relative stabilities of the Classic and M1 complexes were also investigated by ATR-FTIR protein melt experiments. Figure VIII.24 shows the amide-I bands recorded during both melt experiments. A greater number of spectra are present within the M1 overlay since the temperature interval between scans was smaller for the intermediate temperatures of the M1 melt. The variation in band intensity seen in Figure VIII.24*a* and *b* is due to adsorption of the protein to the ATR-crystal surface. Figure VIII.24*c* and *d* correct for this by normalising the amide-I band area. The variation in the spectral series at frequencies higher than 1700 cm^{-1} in Figure VIII.24*d* result from the occurrence of a slight negative water signal due to improper referencing of the aqueous background. Although this would introduce some error into the amide-I bandshape, inspection of Figure VIII.24*d* shows that the amide-I bandshape remains interpretable. From the blue coloured spectra in both Figure VIII.24*c* and *d* it can be seen that both the Classic and M1 complexes experience a clear shift in structure when the temperature is varied from 20 to 30 °C. This is consistent with the Rab11a proteins within each complex having an inherent flexibility in the switch 1 and 2 regions.

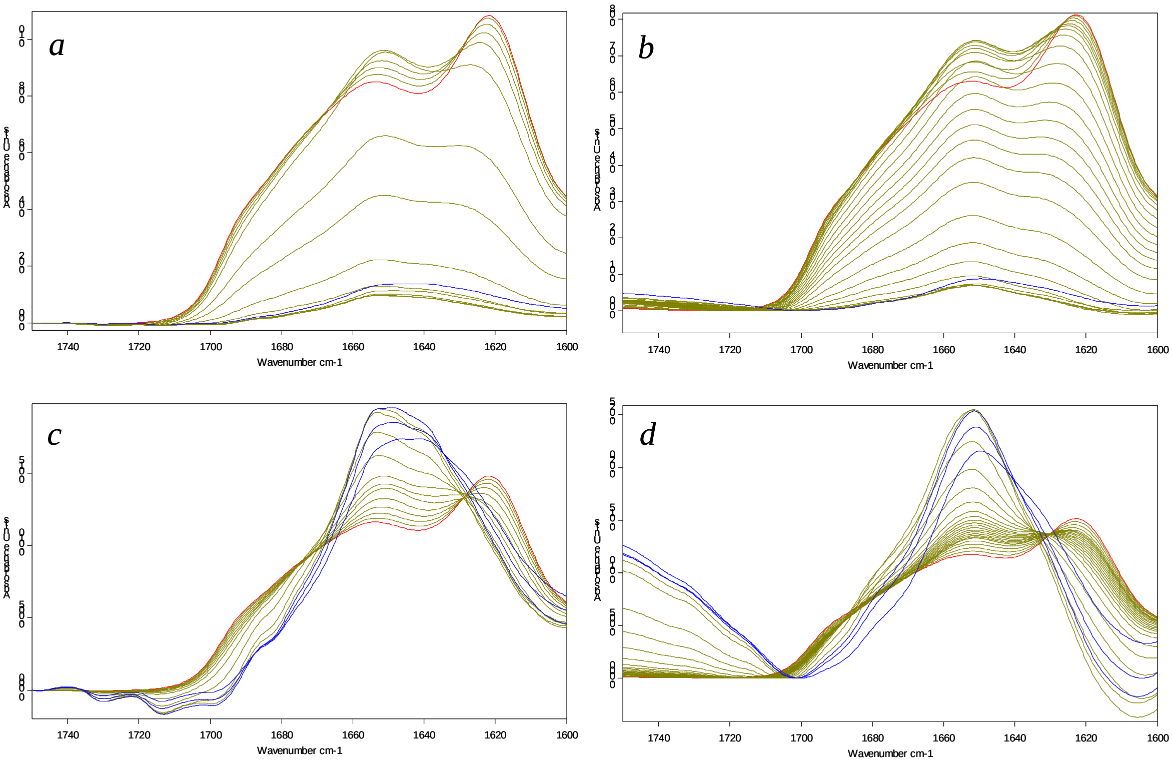


Figure VIII.24. Classic and M1 FIP2-Rab11 complex amide-I absorption FTIR melts. The blue and red spectra correspond to the low and high temperature extremes, respectively, of the *a* Classic 20-90 °C and *b* M1 20-85 °C melts, with the kaki spectra taken at intermediate temperatures. *c* and *d* are the same melt experiments as *a* and *b* where the amide-I band has been area-normalised and the first three temperature step spectra are coloured blue.

Figure VIII.25 shows the second derivative spectra of the complex melt experiments shown in Figure VIII.24. In both cases a decrease in the bands corresponding to α-helix and β-sheet content, located at 1652 and 1632 cm^{-1}, respectively, occur with increased temperature. The emergence of a band at approximately 1620 cm^{-1} confirms the formation of intermolecular β-sheet aggregates at elevated temperatures during both melt experiments.

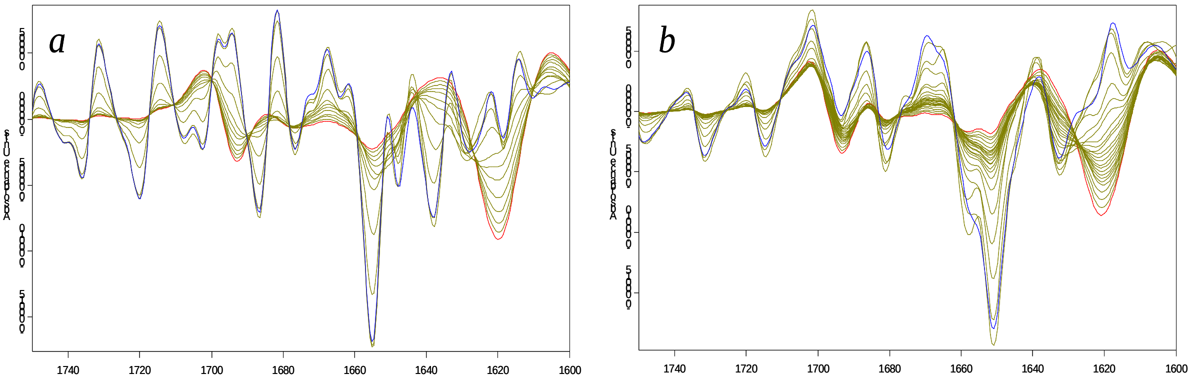


Figure VIII.25. Classic and M1 FIP2-Rab11 complex amide-I area-normalised FTIR second derivative spectral overlay. The spectra corresponding to the first two temperature steps have been excluded from both the *a* Classic and *b* M1 overlays.

Using a precalibrated Quant model, the stability of the helix contents of both complexes can be plotted as a percentage of the total loss of α-helicity experienced during the melt experiment. Figure VIII.26 compares the stability of the Classic and M1 α-helix content and from this it can be seen that both complexes appear to contain helices of roughly equal stabilities.

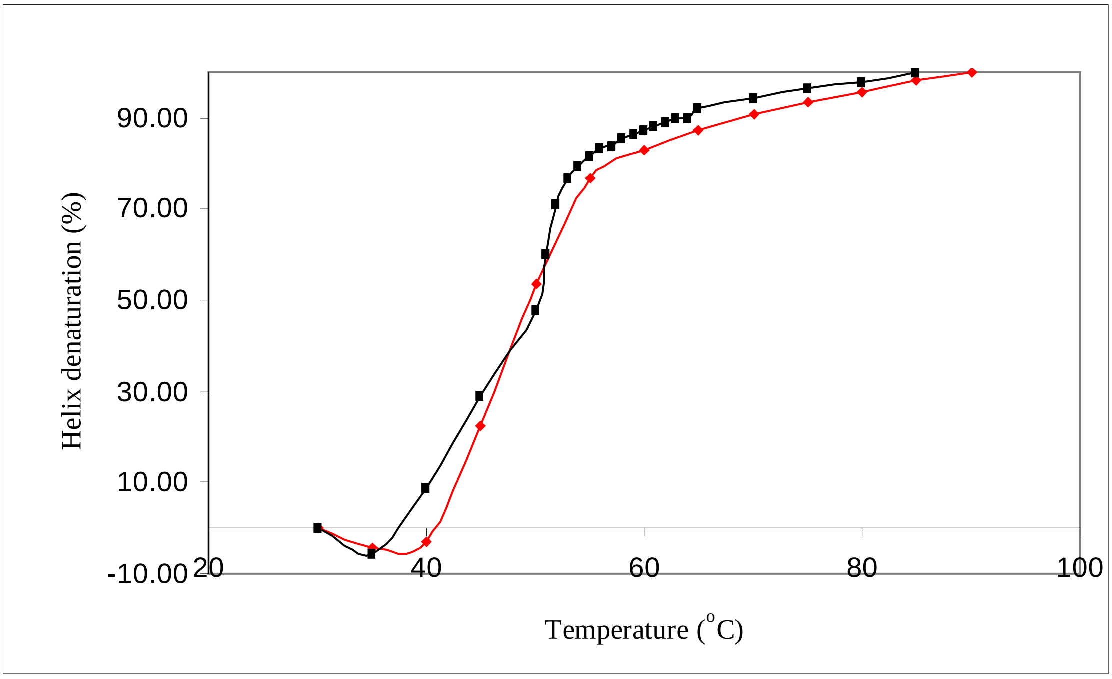


Figure VIII.26. Classic and M1 FIP2-Rab11 complex helix stability. The Classic and M1 curves are plotted in red and black, respectively

VIII.B.3.6 Rab11-FIP2 Homodimer FTIR Melts

ATR-FTIR melt experiments were also performed for the Classic and M1 Rab11-FIP2 homodimers and the spectral overlays of both are shown in Figure VIII.27. Similar to the complex melts, both the Classic and M1 homodimers show a degradation of α-helix and β-sheet content and the formation of protein aggregates when heated, as revealed by the decreased intensity of bands at 1652 and 1632 cm^{-1} and the increased intensity of a band at 1620 cm^{-1}, respectively, in the second derivative spectral overlays (Figure VIII.27*c* and *d*). Figure VIII.28 compares the helix stabilities of the homodimers and from this it appears that the Classic helix content has a slightly greater stability than that of the M1 homodimer at temperatures beyond 36 ^{o}C.

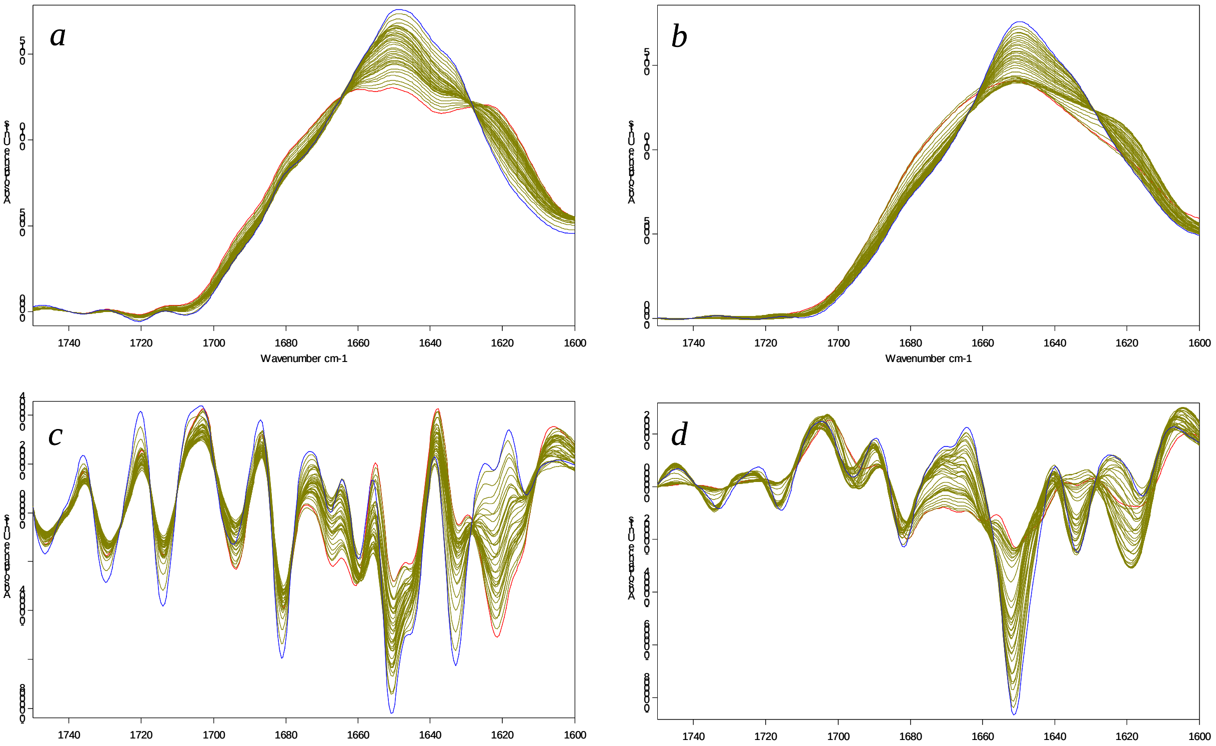


Figure VIII.27. Classic and M1 Rab11-FIP2 homodimer ATR-FTIR melts. The area-normalised amide-I absorption spectral overlays for the *a* Classic and *b* M1 melts are presented along with the *c* Classic and *d* M1 second derivative spectra for the temperature range 20-85 ^{o}C.

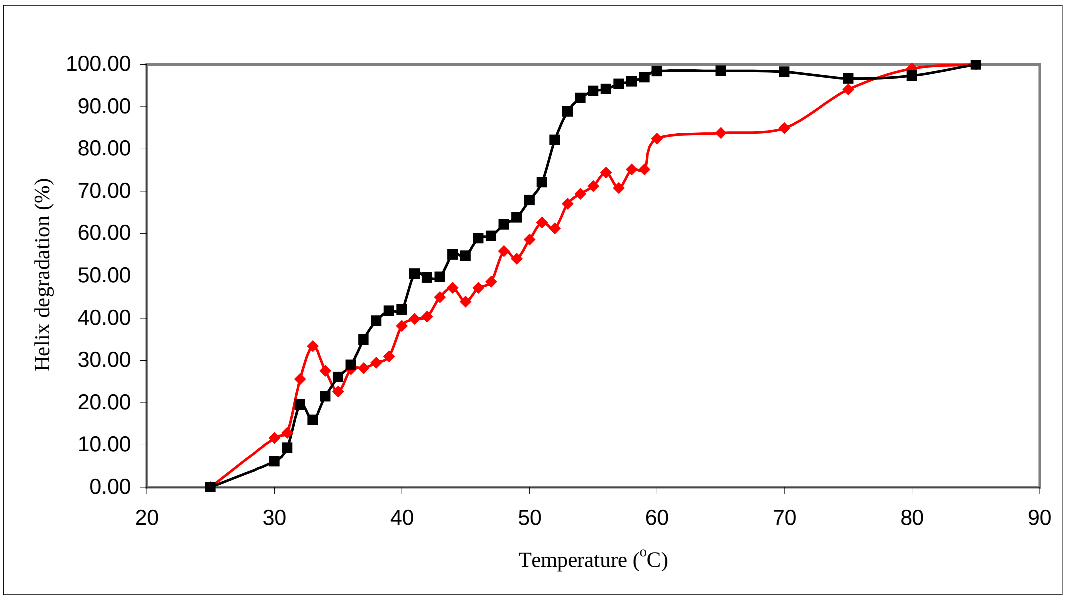


Figure VIII.28. Classic and M1 Rab11-FIP2 homodimer helix stability. The Classic and M1 curves are plotted in red and black, respectively

VIII.B.4 Discussion

The CD and FTIR analyses of the FIP2-Rab11 complexes are consistent both with each other and with the x-ray crystal structure illustrated in Figure VIII.16, in that the spectra contain bands at the expected locations and intensities to account for the secondary structures present in the crystallised complex. Although small differences between the Classic and M1 complexes were present in both the CD and FTIR spectra, giving the possibility that the Classic complex contains slightly more α-helix content than that of the M1 complex, the magnitude of these differences are too small to lend confidence to such a conclusion and, therefore, both complexes are considered to have similar secondary structures. Detection of the presence of the coiled-coil domain within each complex was not possible by CD spectroscopy because of interference from the CD signals of residues within each complex not involved in the coiled-coil domain. In spite of this, a comparison of the Classic and M1 complex CD spectra revealed that the 222:208 ratio was equal in both cases and, therefore, it can be concluded that mutation of the coiled-coil cap residues to produce the M1 complex did not prevent the formation of the coiled-coil domain and, consequently, the heterotetrameric complex. This suggests that either the structural effects on the coiled-coil cap produced by carrying out the mutations to generate the M1 complex were of only minor importance in terms of exposing the hydrophobic core to the surrounding water molecules or that the hydrophobic core is not one of the main stabilising forces behind the formation of the coiled-coil domain. The former conclusion would seem the more likely and, therefore, it appears that the coiled-coil cap might be able to tolerate considerable structural variation and yet still provided a barrier to keep out water molecules. The FTIR spectra of the Classic and M1 homodimers were also extremely similar, confirming that the cap mutations do not prevent formation of the M1 homodimer or significantly distort its structure.

An examination of the stability of the coiled-coil region in both the Classic and M1 complexes by both CD and FTIR produced conflicting results. Both the CD and FTIR demonstrate approximate agreement for the temperature range over which the helix content for both complexes becomes destabilised. The CD data suggests that the Classic complex contains considerably more stable helix content than that of the M1 complex. However, when examined by FTIR both complexes appear to have helix contents of approximately equal stabilities. This result is somewhat unexplainable, although it is possible that the complexes show varying stabilities when in solution, as is the case for the CD measurement, or when adsorbed onto the ATR-crystal surface. Adding further confusion to matters is that the FTIR analysis of the helix stabilities of the Classic versus M1 homodimers reveals that the Classic homodimer helix content is somewhat more stable than that of the M1 homodimer. In conclusion, it would appear that if the CD data is to believed the coiled-coil cap plays an important role in stabilising the coiled-coil domain of the FIP2-Rab11 complex, but according to the FTIR data this stabilising effect is only marginal.

References

1 Cohen C, Parry DAD. Alpha-Helical Coiled Coils and Bundles - How to Design an Alpha-Helical Protein. Proteins-Structure Function and Genetics 7(1):1-15, 1990.

2 Wei J, Fain S, Harrison C, Feig LA, Baleja JD. Molecular dissection of Rab11 binding from coiled-coil formation in the Rab11-FIP2 C-terminal domain. Biochemistry 45(22):6826-6834, 2006.

3 Wallace DM, Lindsay AJ, Hendrick AG, McCaffrey MW. The novel Rab11-FIP/Rip/RCP family of proteins displays extensive homo- and hetero-interacting abilities. Biochemical and Biophysical Research Communications 292(4):909-915, 2002.

4 Jagoe WN, Lindsay AJ, Read RJ, Mccoy AJ, McCaffrey MW, Khan AR. Crystal structure of Rab11 in complex with Rab11 family interacting protein 2. Structure 14(8):1273-1283, 2006.

5 Horgan CP, Oleksy A, Zhdanov AV, Lall PY, White IJ, Khan AR, Futter CE, McCaffrey JG, McCaffrey MW. Rab11-FIP3 is critical for the structural integrity of the endosomal recycling compartment. Traffic 8(4):414-430, 2007.

6 Gonzalez L, Scheller RH. Regulation of membrane trafficking: Structural insights from a Rab/effector complex. Cell 96(6):755-758, 1999.

7 Tisdale EJ, Bourne JR, Khosravifar R, Der CJ, Balch WE. GTP-Binding Mutants of Rab1 and Rab2 Are Potent Inhibitors of Vesicular Transport from the Endoplasmic-Reticulum to the Golgi-Complex. Journal of Cell Biology 119(4):749-761, 1992.

8 Plutner H, Cox AD, Pind S, Khosravifar R, Bourne JR, Schwaninger R, Der CJ, Balch WE. Rab1b Regulates Vesicular Transport between the Endoplasmic-Reticulum and Successive Golgi Compartments. Journal of Cell Biology 115(1):31-43, 1991.

9 Martinez O, Schmidt A, Salamero J, Hoflack B, Roa M, Goud B. The Small Gtp-Binding Protein Rab6 Functions in Intra-Golgi Transport. Journal of Cell Biology 127(6):1575-1588, 1994.

10 Vandersluijs P, Hull M, Webster P, Male P, Goud B, Mellman I. The Small Gtp-Binding Protein Rab4 Controls an Early Sorting Event on the Endocytic Pathway. Cell 70(5):729-740, 1992.

11 Ullrich O, Reinsch S, Urbe S, Zerial M, Parton RG. Rab11 regulates recycling through the pericentriolar recycling endosome. Journal of Cell Biology 135(4):913-924, 1996.

12 Chavrier P, Goud B. The role of ARF and Rab GTPases in membrane transport. Current Opinion in Cell Biology 11(4):466-475, 1999.

13 Cox D, Lee DJ, Dale BM, Calafat J, Greenberg S. A Rab11-containing rapidly recycling compartment in macrophages that promotes phagocytosis. Proceedings of the National Academy of Sciences of the United States of America 97(2):680-685, 2000.

14 Wang XY, Kumar R, Navarre J, Casanova JE, Goldenring JR. Regulation of vesicle trafficking in Madin-Darby canine kidney cells by Rab11a and Rab25. Journal of Biological Chemistry 275(37):29138-29146, 2000.

15 Kessler A, Tomas E, Immler D, Meyer HE, Zorzano A, Eckel J. Rab11 is associated with GLUT4-containing vesicles and redistributes in response to insulin. Diabetologia 43(12):1518-1527, 2000.

16 Wilcke M, Johannes L, Galli T, Mayau V, Goud B, Salamero J. Rab11 regulates the compartmentalization of early endosomes required for efficient transport from early endosomes to the trans-golgi network. Journal of Cell Biology 151(6):1207-1220, 2000.

17 Lapierre LA, Kumar R, Hales CM, Navarre J, Bhartur SG, Burnette JO, Provance DW, Mercer JA, Bahler M, Goldenring JR. Myosin Vb is associated with plasma membrane recycling systems. Molecular Biology of the Cell 12(6):1843-1857, 2001.

18 Meyers JM, Prekeris R. Formation of mutually exclusive Rab11 complexes with members of the family of Rab11-interacting proteins regulates Rab11 endocytic targeting and function. Journal of Biological Chemistry 277(50):49003-49010, 2002.

19 Lindsay AJ, McCaffrey MW. Characterisation of the Rab binding properties of Rab coupling protein (RCP) by site-directed mutagenesis. Febs Letters 571(1-3):86-92, 2004.

20 Hales CM, Griner R, Hobdy-Henderson KC, Dorn MC, Hardy D, Kumar R, Navarre J, Chan EKL, Lapierre LA, Goldenring JR. Identification and characterization of a family of Rab11-interacting proteins. Journal of Biological Chemistry 276(42):39067-39075, 2001.

21 Prekeris R, Klumperman J, Scheller RH. A Rab11/Rip11 protein complex regulates apical membrane trafficking via recycling endosomes. Molecular Cell 6(6):1437-1448, 2000.

22 Lindsay AJ, Hendrick AG, Cantalupo G, Senic-Matuglia F, Goud B, Bucci C, McCaffrey MW. Rab coupling protein (RCP), a novel Rab4 and Rab11 effector protein. Journal of Biological Chemistry 277(14):12190-12199, 2002.

23 Horgan CP, Walsh M, Zurawski TH, McCaffrey MW. Rab11-FIP3 localises to a Rab11-positive pericentrosomal compartment during interphase and to the cleavage furrow during cytokinesis. Biochemical and Biophysical Research Communications 319(1):83-94, 2004.

24 Horgan CP, Zurawski TH, McCaffrey MW. Purification and functional properties of Rab11-FIP3. Gtpases Regulating Membrane Targeting and Fusion 403:499-+, 2005.

25 Wilson GM, Fielding AB, Simon GC, Yu XZ, Andrews PD, Hames RS, Frey AM, Peden AA, Gould GW, Prekeris R. The FIP3-Rab11 protein complex regulates recycling endosome targeting to the cleavage furrow during late cytokinesis. Molecular Biology of the Cell 16(2):849-860, 2005.

26 Luo PZ, Baldwin RL. Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. Biochemistry 36(27):8413-8421, 1997.

27 Fielding AB, Schonteich E, Matheson J, Wilson G, Yu XZ, Hickson GRX, Srivastava S, Baldwin SA, Prekeris R, Gould GW. Rab11-FIP3 and FIP4 interact with Arf6 and the Exocyst to control membrane traffic in cytokinesis. Embo Journal 24(19):3389-3399, 2005.

28 Eathiraj S, Mishra A, Prekeris R, Lambright DG. Structural basis for Rab11-mediated recruitment of FIP3 to recycling endosomes. Journal of Molecular Biology 364(2):121-135, 2006.

Chapter IX

An Investigation into the Structure of a Novel Anhydrobiotic Nematode DJ-1 Protein

IX.1 Introduction

Studies of anhydrobiosis within nematodes led to the discovery of an anhydrobiotic protein within the *Panagolaimus superbus* species that belongs to the DJ-1/ThiJ/Pfpl protein superfamily. The human form of DJ-1 has been considerably researched since it was found that mutations within the DJ-1 gene were responsible for the recessively inherited Parkinson’s disease. The structure of human DJ-1 has been solved by x-ray crystallography and is shown in Figure IX.1.[2] Other DJ-1 orthologs are to be found in a variety of eukaryotic species and these are reported as being responsible for a variety of functions. For example, DJ-1 proteins have been associated with cellular transformation [3], transcriptional effects [4], control of mRNA stability [5], chaperone activity [6] and response to oxidative stress [7; 8]. DJ-1 proteins show sequence homology to numerous proteins, which are related by their possession of a ThiJ domain ($\alpha/\beta/\alpha$ sandwich motif) and amongst these are included protein chaperones [9], catalases [10], proteases [11; 12]and ThiJ kinases [13; 14]. Similar to the glutamine amidotransferase protein superfamily (GAT), members of the DJ-1/ThiJ/PfpI superfamily have an overall α/β sandwich structure, as can be seen for human DJ-1 from Figure IX.1.[10; 15] Bandyopadhyay and Cookson performed a detailed analysis of several hundred sequences of the DJ-1/ThiJ/PfpI superfamily members and found the bacterial ThiJ genes to be the nearest homologous primary sequences.[15] From this they supposed that DJ-1 might have evolved from thiamine synthesis genes that have been dispensed with in eukaryotes.

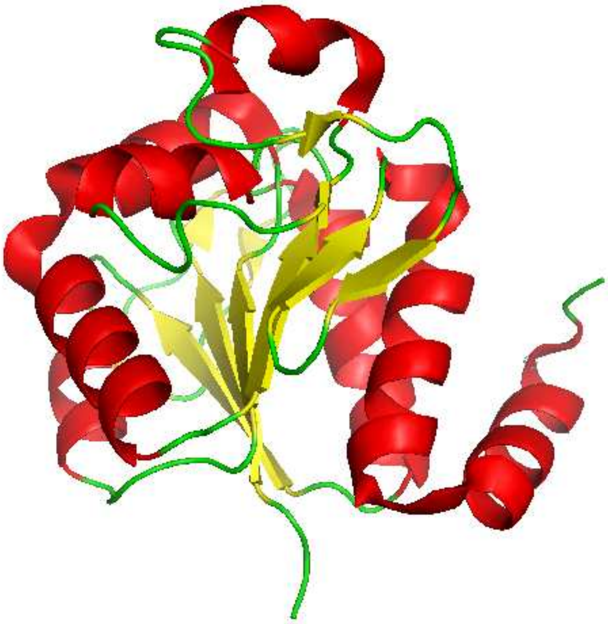


Figure IX.1. Ribbon representation of the secondary structure of human DJ-1. The α-helix and β-sheet regions are coloured red and yellow, respectively. (PyMol)

The neurodegenerative Parkinson's disease (PD) is characterised by the loss of dopaminergic neurons in the substantia nigra and by the presence of intracellular inclusions called Lewy bodies.[16] Lewy bodies are found to be particularly enriched for α-synuclein fibrils. α-synuclein is a natively unfolded protein that exists in high concentrations in dopaminergic neurons and autosomal mutations within α-synuclein lead to a rare form of primary Parkinsonism.[17] Human DJ-1 has been found to perturb α-synuclein aggregate formation in a redox-dependent manner by functioning as a molecular chaperone.[18] It is, therefore, thought that DJ-1 exerts its effect on Parkinson's disease by preventing the formation of Lewy bodies, which are the result of α-synuclein aggregation - a process that is exacerbated under oxidative stress conditions. Such oxidising conditions are present within dopaminergic neurons, due to the highly reactive nature of dopamine.[19-21]

Human DJ-1 is thought to respond to oxidative stress by becoming oxidised at its cysteine residues, leading to the formation of cysteine sulfinic or sulfonic acids, and this is particularly true for the highly conserved solvent-exposed cysteine-106 residue.[22; 23] The modification of these residues is understood to have consequences relative to the functioning of the protein. Studies performed on cell cultures and animal models suggest that DJ-1 function may be modulated by oxidation of cysteine residues and that this alteration potentially leads to the activation of chaperone activity.[18; 24; 25] Shendelman *et al.* performed mutation knockdown investigations into the importance of Cys-106 and Cys-53 for the chaperone functionality of human DJ-1.[18] It was discovered that Cys-53 was essential for DJ-1 chaperone activity in preventing α-synuclein aggregation, whereas Cys-106 was not. Zhou *et al.* went on to demonstrate that oxidation of Cys-106 in human DJ-1 was essential for its chaperone activity towards α-synuclein and that where such oxidation did not occur that native DJ-1 had no effect upon α-synuclein fibrillation.[25] Also, Zhou and Freed found that DJ-1 up-regulated glutathione synthesis during oxidative stress and inhibited human A53T α-synuclein toxicity through increased expression of heat shock protein 70, leading the authors to conclude that DJ-1 has multiple specific mechanisms for protecting dopamine neurons from cell death.[26] In addition, human DJ-1 has recently been discovered to enhance the cellular oxidative stress response by regulating the activity of the antioxidant transcription factor Nrf2 [27] and it may also exert additional influence over cell fate through its involvement in the PTEN/Akt signalling pathway.[28; 29]

Having discovered a novel DJ-1 protein within *panagolaimus superbus* (P. sup DJ-1) , it was interesting to examine this protein’s structure and to compare it to the structure of human DJ-1. It was also valuable to determine whether this protein was capable of adopting a similar function to that of the human variety and, as such, its thermo-stability and structural response to oxidation was examined.

IX.2 Method

IX.2.1 Protein synthesis and purification

A pET-30-DJ-1 plasmid was generated by ligating the entire cDNA coding region of DJ-1 from *p. sup.* into a pET-30 Ek/LIC vector. The plasmid was transformed into BL21 (DE3) pLYS *E. coli* cells, a single colony of these cells was used to inoculate an LB broth containing kanamycin (Sigma) and the culture was grown at 37 ^{o}C until the OD_{600} had reached 0.6. This culture was used to inoculate 100 ml of LB broth containing kanamycin and grown at 37 ^{o}C to an OD_{600} of 0.8. Induction was by addition of 1 mM IPTG (Sigma) with incubation at 37 ^{o}C for 4 hrs. Cells were harvested by centrifugation at 4 ^{o}C and stored at –20 ^{o}C until needed. After defrosting, cells were lysed using Bugbuster reagent and Benzonase Nuclease (Merck Biosciences). Protein purification was achieved initially by using a vector encoded His-Tag and a Ni-NTA resin column (Merck Biosciences). Before a second purification step was applied to the eluted protein, using an AKTA purifier equipped with HiTrap Q HP anion exchange column, the eluted protein was dialyzed against 75 mM Tris 10 mM NaCl pH 8.0 (bind buffer). After loading, the ion exchange column was washed with five column volumes of bind buffer and eluted with a shallow gradient of 0-50 % of 75 mM Tris 1 M NaCl pH 8.0 in 15 column volumes. Purified protein eluted as a single peak at 35 % gradient. Protein purity was analysed by SDS-PAGE gel electrophoresis (12 %) using a Precision Plus All Blue marker (Biorad) with Coomassie blue staining. The presence of recombinant P. sup DJ-1 was confirmed by both Western Blot - using anti-His tag antibody (Merck Biosciences) and 3,3’-Diaminobenzidine (DAB) (Sigma) - and MALDI-TOF mass spectrometry procedures (data not shown). The sample was prepared for CD analysis by dialyzing against 10 mM PBS, 100 mM NaF pH 7.4. Samples were concentrated to a protein concentration of approximately 2 mg/ml using a 10K MW cut-off Amicon Ultra filter (Millipore), as measured by A_{280} and an extinction coefficient estimated from the protein primary sequence by the Protparam program (http://www.expasy.ch/tools/protparam.html).

IX.2.2 CD Spectroscopy

Spectra were acquired on a Model 400 CD Spectrometer (Aviv Biomedical), using a rectangular demountable quartz cell of 0.01 cm pathlength (Hellma UK), as an average of 10 scans, in the wavelength range 260-178 nm (250-190 nm where DTT was present), at a temperature of 23 ^{o}C, with a bandwidth of 1 nm, an acquisition time of 1 s and at intervals of 0.2 nm. Spectra were background corrected by subtracting the CD spectrum of the protein sample buffer solution, which was acquired under identical conditions to that of the sample spectrum. Spectra were acquired with DJ-1 in its native state, in the presence of 600 mM H_2O_2 in its oxidised state and in the presence of 5 mM DTT in its reduced state. CD_{222} melt experiments were performed for DJ-1 in both its native and oxidised state using a 1 cm quartz cell (Hellma UK) equipped with a magnetic stirrer, at a wavelength of 222 nm, with a spectral acquisition time of 1 s, over the temperature range of 10-88 ^{o}C, with a temperature step of 3 ^{o}C and with a 2 minute equilibration time at each temperature prior to acquisition. The $T_{1/2}$ value was determined as the temperature at which the CD_{222} intensity had decreased by half of its total intensity shift over the temperature range of the experiment. All samples were thoroughly degassed before spectral acquisition. Sample protein concentration determination was by the QAAA method (Cambridge University) and was taken as an average of 2 analyses. Quantitative secondary structure analysis was performed using the SELCON3, CONTIN/LL, CDSSTR, VARSLC and K2D programs and the DichroWeb server.

IX.2.3 FTIR Spectroscopy

FTIR spectra were acquired using a Tensor27 FTIR spectrometer equipped with a sensitive N_2 cooled MCT detector (Bruker). Transmission spectra were acquired using an AquaSpec transmission cell accessory (Bruker), as an average of 150 scans, at a resolution of 4 cm^{-1}, with an aperture of 0.5 cm, over a wavenumber range of 4000-1000 cm^{-1}, with a Norton-Beer Fourier transform apodization function and at a temperature of 20 ^{o}C. Absorption spectra were generated by subtracting the transmission spectrum of the sample buffer solution (obtained using the same acquisition parameters) from that of the sample.

FTIR melt experiments were performed using a BioATR-II ATR accessory (Bruker), as an average of 500 scans, at a resolution of 4 cm^{-1}, with an aperture of 4 cm, over a wavenumber range of 4000-1000 cm^{-1}, with a Blackman-Harris Fourier transform apodization function, over a temperature range of 13-88 ^{o}C and with a 2 minute equilibration time. Absorption spectra were generated by subtracting the spectrum of the buffer from that of the sample at each temperature. Samples were thoroughly degassed before injection into either the AquaSpec or BioATR-II cells.

All spectra were post-processed within the OPUS software (Bruker) using an atmospheric correction algorithm to remove any water vapour bands from the protein spectrum. Spectra were zeroed between 1800-1750 cm^{-1}, area-normalised between 1700-1600 cm^{-1} (for the protein melt spectra only) and the spectral 2ND derivatives were then calculated. The transmission mode spectrum was used to determine the α-helix and β-sheet content of the sample protein using Quant macrofiles that employ a protein spectral database within the OPUS software. The protein aggregation profile was determined by generating a Quant macrofile that analysed the area of a band between 1626-1620 cm^{-1} in the 2ND derivative spectra. $T_{1/2}$ values were obtained by observing the temperature at which the intensity of this band had increased to half of its maximal value.

IX.3 Results

IX.3.1 P. sup. DJ-1 Protein Sample Purity

Figure IX.2*a* shows the SDS-PAGE gel analysis of a P. sup DJ-1 protein sample. The presence of an intense band at approximately 25 kDa in lanes 3-8 is evidence of the isolation of a highly concentrated P. sup DJ-1 sample. The lack of any other significant bands within these lanes suggests that the protein was present in high purity. Figure IX.2*b* reveals the results of the ion-exchange purification procedure. The presence of a single peak in the UV_{280} and UV_{264} time resolved absorption; along with the narrow peak width suggests that the protein sample is of excellent purity.

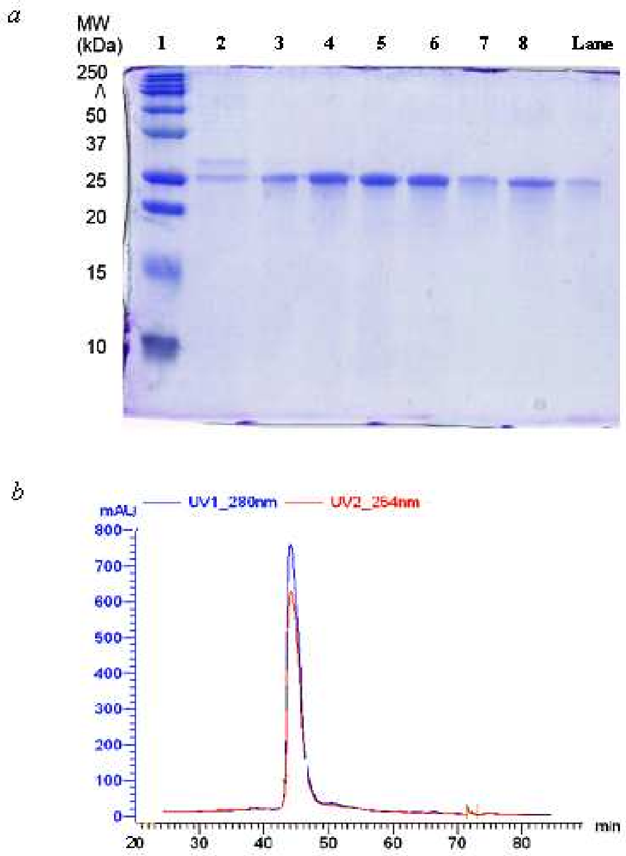
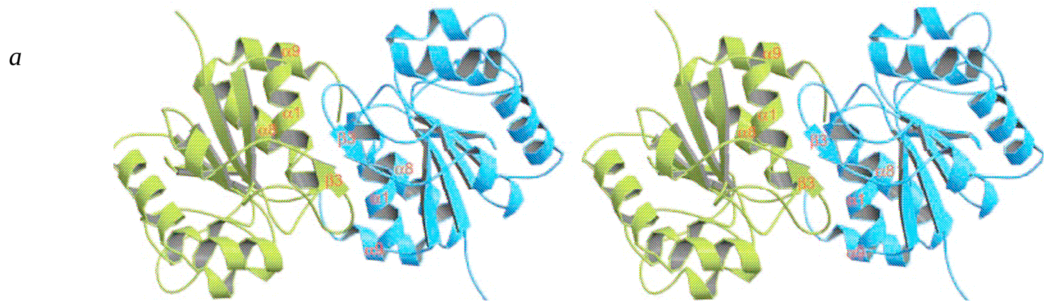


Figure IX.2. SDS-PAGE gel and ion exchange purification of P. sup DJ-1. ***a*** For the SDS-PAGE gel the protein ladder is shown in lane 1 and the purified protein is in lanes 3-8. *b* The sample elution fraction from the ion exchange column shows both a single UV_{280} and UV_{264} peak that occurred after 45 minutes flow time.

IX.3.2 Structural Prediction of P. Sup. DJ-1

X-ray crystallography studies of human DJ-1 reveal that the protein consists of a dimer composed of two identical monomeric peptide units.[30] Residues Met-17, Ile-21, His-126, Pro-127, Pro-158 and Phe-162 are thought to be involved in dimer formation and these residues are completely conserved within human, rat, mouse, *Xenopus*, *Drosophila* and nematoda DJ-1 proteins.[31] Of these, residues Met-17 and Phe-162 are considered to be essential for dimer fomation because they represent the core of the monomeric hydrophobic interaction. Figure IX.3 compares the primary sequences of human DJ-1 and the nematode *Caenorhabditis elegans* (*C. elegans*) and *Panagolaimus superbus* DJ-1, where the residues that are involved in forming the DJ-1 dimer are highlighted in red. From this it seems plausible that P. sup DJ-1 exists as a dimer, since all but two of the residues listed above are present within the *P. superbus* primary sequence. However, this conclusion is made somewhat doubtful due to the absence of residue Met-17 within P. sup DJ-1 and, in addition, the hydrophobic methionine is replaced by a hydrophilic threonine residue. A similar replacement of the hydrophobic His-126 with the hydrophilic tyrosine residue for P. sup DJ-1 raises further doubts about the dimeric nature of P. sup DJ-1.



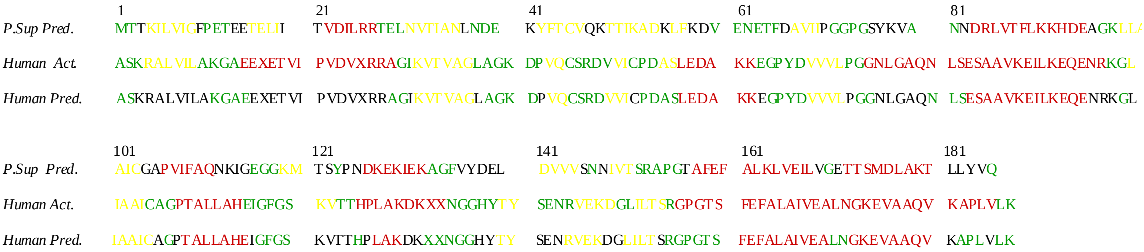
b

	1	21	41	61	81
P.Sup	MTTKILVIGFPETEETELII	TVDILRRTELNVTIANLNDE	KYFT**C**VQKTTIKADKLFKDV	ENETFDAVIIPGGPGSYKVA	NNDRLVTFLKKHDEAGKLLA
Human	ASKRALVILAKGAEEXETVI	PVDVXRRAGIKVTVAGLAGK	DPVQ**C**SRDVVI**C**PDASLEDA	KKEGPYDVVVLPGGNLGAQN	LSESAAVKEILKEQENRKGL
C. elegans	MAQKSALIILAAEGAEEMEV	IITGDVLARGEIRVVYAGLD	GAEPVK**C**ARGAHIVPDVKLE	DVETEKFDIVILPGGQPGSN	TLAESLLVRDVLKSQVESGG

	101	121	141	161	181
P.Sup	AI**C**GAPVIFAQNKIGEGGKM	TSYPNDKEKIEKAGFVYDEL	DVVVSNNIVTSRAPGTAFEF	ALKLVEILVGETTSMDLAKT	LLYVQ
Human	IAAI**C**AGPTALLAHEIGFGS	KVTTHPLAKDKXXNGGHYTY	SENRVEKDGLILTSRGPGTS	FEFALAIVEALNGKEVAAQV	KAPLVLK
C. Elegans	LIGAI**C**AAPIALLSHGVKAE	LVTSHPSVKEKLEKGGYKYS	EDRVVVSGKIITSRGPGTAF	EFALKIVELLEGKDKATSLI	APMLLKL

Figure IX.3. DJ-1 dimer formation. *a* A stereo pair of a ribbon diagram of the human DJ-1 dimer. The ribbons are coloured blue and green for monomer A and B, respectively, and the four structural elements involved in the dimer formation are labelled in red. The β3 domain is formed upon dimerization. *b* Comparison of the primary sequences of human , P. sup and C. elegans DJ-1. The residues important for dimer formation are coloured red and the cysteine residues are shown as bold and underlined. Where the dimer forming residues are not conserved within P. sup DJ-1 the replacing residue is shown in green.

Figure IX.4 shows the predicted secondary structures of P. sup. and human DJ-1 from their primary structures, along with the crystal structure of human DJ-1 from x-ray crystallography.[2] Comparing the predicted and x-ray structure of human DJ-1 reveals the accuracy of the PROF program in predicting secondary structure and it is clear that there exists a good correlation between the predicted and experimentally determined structure for human DJ-1. Comparing the predicted structure of P. sup DJ-1 with the crystal structure of human DJ-1 reveals a high level of structural homology between the two. It is likely, therefore, that P. sup DJ-1 adopts a very similar secondary structure to human DJ-1.



	1	21	41	61	81
P.Sup Pred.	MTTKILVIGFPETEETELII	TVDILRRTELNVTIANLNDE	KYFTCVQKTTIKADKLFKDV	ENETFDAVIIPGGPGSYKVA	NNDRLVTFLKKHDEAGKLL
Human Act.	ASKRALVILAKGAEEXETVI	PVDVXRRAGIKVTVAGLAGK	DPVQCSRDVVICPDASLEDA	KKEGPYDVVVLPGGNLGAQN	LSESAAVKEILKEQENRKGL
Human Pred.	ASKRALVILAKGAEEXETVI	PVDVXRRAGIKVTVAGLAGK	DPVQCSRDVVICPDASLEDA	KKEGPYDVVVLPGGNLGAQN	LSESAAVKEILKEQENRKGL

	101	121	141	161	181
P.Sup Pred.	AICGAPVIFAQNKIGEGGKM	TSYPNDKEKIEKAGFVYDEL	DVVVSNNIVTSRAPGTAFEF	ALKLVEILVGETTSMDLAKT	LLYVQ
Human Act.	IAAICAGPTALLAHEIGFGS	KVTTHPLAKDKXXNGGHYTY	SENRVEKDGLILTSRGPGTS	FEFALAIVEALNGKEVAAQV	KAPLVLK
Human Pred.	IAAICAGPTALLAHEIGFGS	KVTTHPLAKDKXXNGGHYTY	SENRVEKDGLILTSRGPGTS	FEFALAIVEALNGKEVAAQV	KAPLVLK

Figure IX.4. Comparison of the primary sequence of P. Sup and human DJ-1. The predicted secondary structure of P. Sup DJ-1 is shown (PROF)[1], along with the x-ray crystal and predicted secondary structures of human DJ-1 [2], with β-sheet and α-helix residues coloured yellow and red, respectively, and loop region residues coloured green. For the predicted secondary structures, residues coloured black correspond to protein regions where the PROF program could not assign secondary structure.

IX.3.3 CD Analysis of P. sup DJ-1 Secondary Structure

The CD spectrum of P. sup DJ-1 is characteristic of a well-structured α/β mixed protein, whereby the typical helix band with minima at 222 and 208 nm is present, along with a disproportionately intense positive band at wavelengths shorter than 200 nm - indicating the presence of significant sheet content (Figure IX.5). The CD error values in Figure IX.5 are very small, even at the short wavelength region of the spectrum. Table IX.1 shows the results of analysing the native DJ-1 spectrum shown in Figure IX.5 using the programs contained within the DichroWeb software. Since the CD spectrum was acquired to 178 nm, each of the eight protein databases could be used in conjunction with the SELCON3 and CONTIN/LL programs. The software would not allow analysis below 185 nm when using the CDSSTR program, however, and this prevented the use of reference sets 1, 2 and 5 with this program.

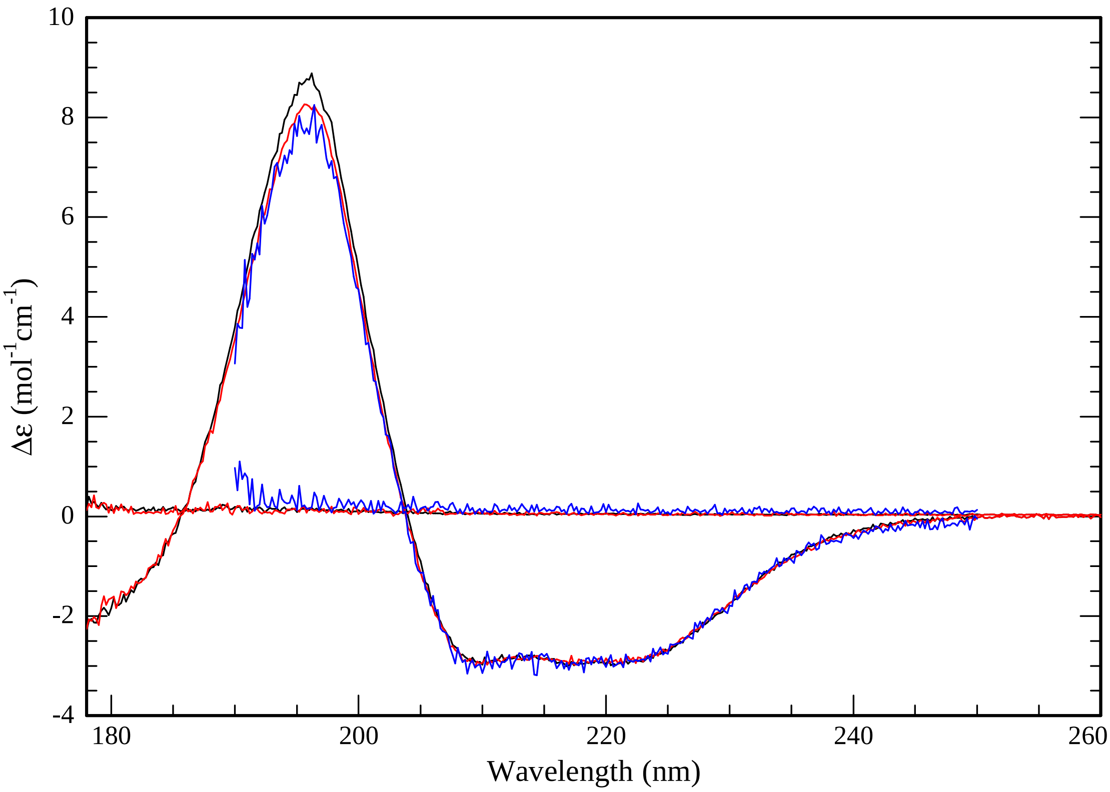


Figure IX.5. CD spectrum of P. Sup. DJ-1. The black, red and blue spectrum was recorded under native, oxidising and reducing conditions, respectively. The CD error values are shown in their corresponding colours as overlaid.

For each analysis there is listed the NMRSD value, which approximates the success achieved when analysing the protein's structure by the particular method, with lower values indicating a superior analysis. With the exception of the SELCON3 program, NMRSD values were always within the limits specified for the program in order for the result to be valid. However, the NMRSD values for each analysis using the SELCON3 program were very close to the 0.25 cut-off limit and SELCON3 usually produces the largest NMRSD values when compared to the other programs. The close comparison between results using the SELCON3 program and those using the CONTIN/LL program, for which the NMRSD values were well within the required limit, suggests that SELCON3 was capable of analysing the P. sup DJ-1 CD spectrum.

Program	**Ref. Set**	α_1	α_2	β_1	β_2	P_{II}	**Turn**	**Other**	**NMRSD**
K2D	-	0.38	-	0.15	-	-	-	0.47	0.113
SELCON3	*1*	0.194	0.127	0.110	0.091	-	0.249	0.221	0.360
	2	0.276	0.070 (3_{10}H)	0.202	-	0.040	0.124	0.297	0.335
	3	0.203	0.136	0.192	0.083	-	0.202	0.200	0.277
	4	0.243	0.170	0.115	0.066	-	0.145	0.265	0.277
	5	0.353	-	0.207	-	0.068	0.228	0.124	0.314
	6	0.204	0.137	0.196	0.086	-	0.209	0.183	0.259
	7	0.246	0.173	0.119	0.063	-	0.132	0.262	0.258
	SP175	0.215	0.156	0.139	0.077	-	0.090	0.321	0.346
CONTIN/LL	*1*	0.220	0.114	0.142	0.088	-	0.235	0.200	0.109
	2	0.262	0.062 (3_{10}H)	0.227	-	0.032	0.117	0.300	0.111
	3	0.223	0.128	0.158	0.074	-	0.215	0.201	0.095
	4	0.244	0.167	0.120	0.065	-	0.174	0.230	0.056
	5	0.344	-	0.236	-	0.044	0.224	0.152	0.160
	6	0.228	0.138	0.171	0.078	-	0.231	0.155	0.095
	7	0.241	0.177	0.137	0.075	-	0.202	0.168	0.056
	SP175	0.238	0.157	0.133	0.077	-	0.097	0.299	0.044
VARSLC	-	0.41	-	0.08	0.12	-	0.14	0.20	0.294
CDSSTR	*3*	0.26	0.17	0.10	0.09	-	0.23	0.15	0.007
	4	0.29	0.17	0.11	0.06	-	0.15	0.21	0.008
	6	0.27	0.17	0.18	0.09	-	0.19	0.12	0.007
	7	0.27	0.20	0.12	0.08	-	0.14	0.20	0.008
	SP175	0.27	0.19	0.10	0.07	-	0.07	0.29	0.012

Table IX.1. Quantitative secondary structure analysis of P. sup DJ-1 from its CD spectrum.

From Table IX.1 it can be seen that results were consistent over the range of reference sets used for each program. There was good correlation also between the average results from each of the programs. As expected, the CDSSTR program estimated the highest total helix content compared to the other programs and also achieved the lowest NMRSD values. Given that CDSSTR generally overestimates the helix content of proteins, P. sup DJ-1 is likely to contain between 38-42 % α-helix, as suggested by the estimates of all other programs. The total β-sheet estimate was consistently between 18-24 % over all analyses. Therefore, the proportion of non-helix/non-sheet structure in P. sup DJ-1 is estimated as between 34-44 %.

Table IX.2 shows the values estimated for both the number of α-helices and β-sheets and their average lengths within P. sup DJ-1. Since the protein contains 185 residues, analysis of its CD spectrum estimates it to contain approximately 7 α-helices of an average length of 10 residues and 6.5 β-sheets of an average length of 5.5 residues. Comparison of the estimated structure of P. sup DJ-1 from its CD spectrum with the x-ray crystal structure of human DJ-1 (Figures IX.1 and IX.3) shows excellent agreement for both the number of helices and their average length. Although the estimated number of β-sheets for P. sup DJ-1 is considerably less than that observed in human DJ-1, this is not surprising considering that many of the human DJ-1 β-sheets are comprised of only 2 or 3 residues. If these are excluded, there is excellent agreement between P. sup and human DJ-1 in terms of both the number of β-sheets and their average length. Taken together, this is strong evidence for both proteins having very similar secondary structures.

Program	**Ref. Set**	**H per 100 res.**	**Av. H length (Res)**	**β per 100 res.**	**Av. β length (Res)**
SELCON3	*1*	3.173	10.099	4.528	4.429
	3	3.396	9.963	4.165	6.615
	4	4.245	9.722	3.276	5.500
	6	3.417	9.973	4.287	6.568
	7	4.315	9.709	3.173	5.738
	SP175	3.908	9.502	3.827	5.627
CONTIN/LL	*1*	3.100	10.312	4.661	4.996
	3	3.800	9.974	3.850	7.143
	4	4.183	9.996	3.359	5.672
	6	3.800	9.974	3.850	7.143
	7	4.308	9.844	3.398	5.810
	SP175	4.029	9.607	3.882	5.696
CDSSTR	*3*	4.350	10.039	4.306	4.246
	4	4.202	10.923	3.185	5.586
	6	4.230	10.293	4.278	6.170
	7	5.009	9.306	3.772	5.064
	SP175	4.831	9.598	3.289	5.057

Table IX.2. Analysis of the helical and sheet properties of P. sup DJ-1 from its CD spectrum. The average helix (H) and sheet (β) lengths are given in residue numbers (Res).

IX.3.4 FTIR Analysis of P. sup DJ-1 Secondary Structure

The amide-I band of P. sup DJ-1 is shown in Figure IX.6 and 2^{ND} derivative analysis of the absorption spectrum reveals the presence of two intense bands centred at 1653 and 1628 cm^{-1}, corresponding to the α-helix and β-sheet amide-I bands, respectively. The minor β-sheet band can also be seen in the 2^{ND} derivative spectrum as a band at around 1687 cm^{-1}. A band located at approximately 1674 cm^{-1} most likely results from β-turn protein content. Inspection of Figure IX.1 reveals the presence of these structures within human DJ-1. A quantitative analysis of secondary structure estimates the α-helix and β-sheet content as 37.831 and 29.167 %, respectively.

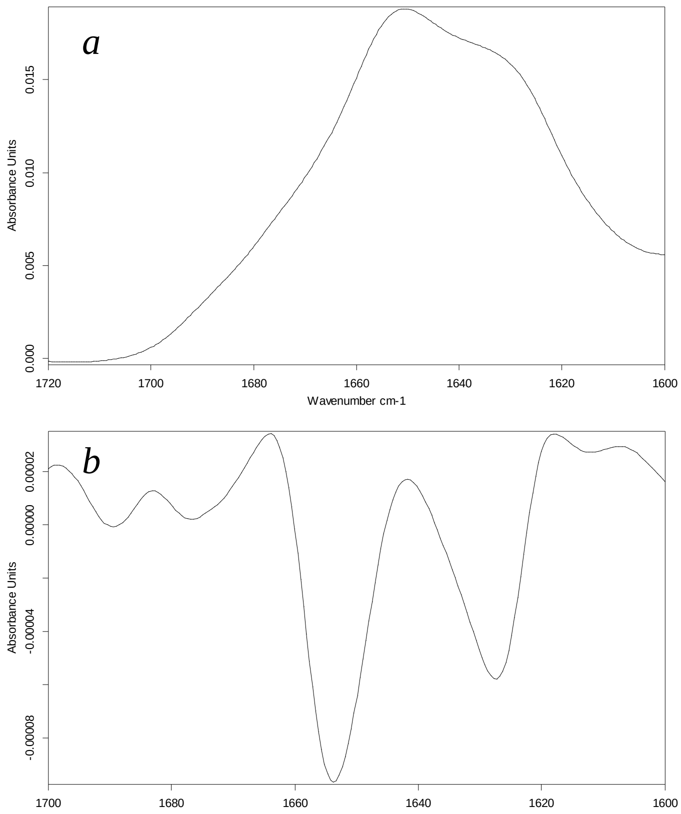


Figure IX.6. The amide-I band of P. sup DJ-1. The amide-I absorption spectrum is shown in *a* and the 2ND derivative is given in *b*.

IX.3.5 P. sup DJ-1 Thermo-stability

Given the close relation between DJ-1 proteins and the heat-shock Hsp31 *E. coli* chaperone protein, the thermo-stability of the helix domains of P. sup DJ-1 is significant if a similar functionality is to be ascribed to it. Figure IX.7 shows the $CD_{222\ nm}$ melt of P. sup DJ-1 and from this a $T_{1/2}$ value of 76 ^{o}C was determined. This is in almost exact agreement with the $T_{1/2}$ value of 75 ^{o}C of human DJ-1.[18] Also, the P. sup DJ-1 aggregation profile was investigated by way of an FTIR melt analysis (Figure IX.8). Figure IX.9 tracks the increasing intensity of the aggregation band, centred at approximately 1622 cm^{-1} in Figure IX.8, and from this a $T_{1/2}$ value of 70 ^{o}C was determined. Therefore, P. sup DJ-1 exhibits very high thermo-stability with regard to both its helix domains and its aggregation profile and this is consistent with what is observed for human DJ-1 and is as expected for a heat shock protein.[18]

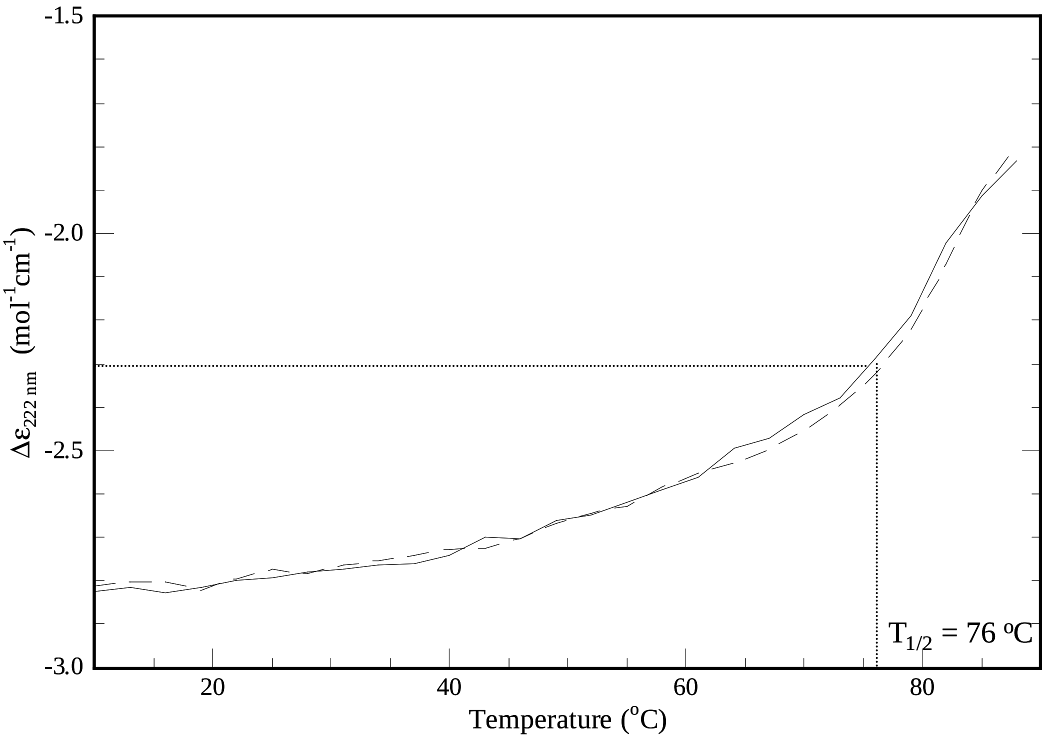


Figure IX.7. P. sup DJ-1 helix stability from CD. The helix stability of P. sup DJ-1 alone (solid) and in the presence of 600 mM H_2O_2 (dashed) is revealed by monitoring the $\Delta\varepsilon_{222\ \mathrm{nm}}$ as a function of temperature, with a $T_{1/2}$ value of 76 ^{o}C.

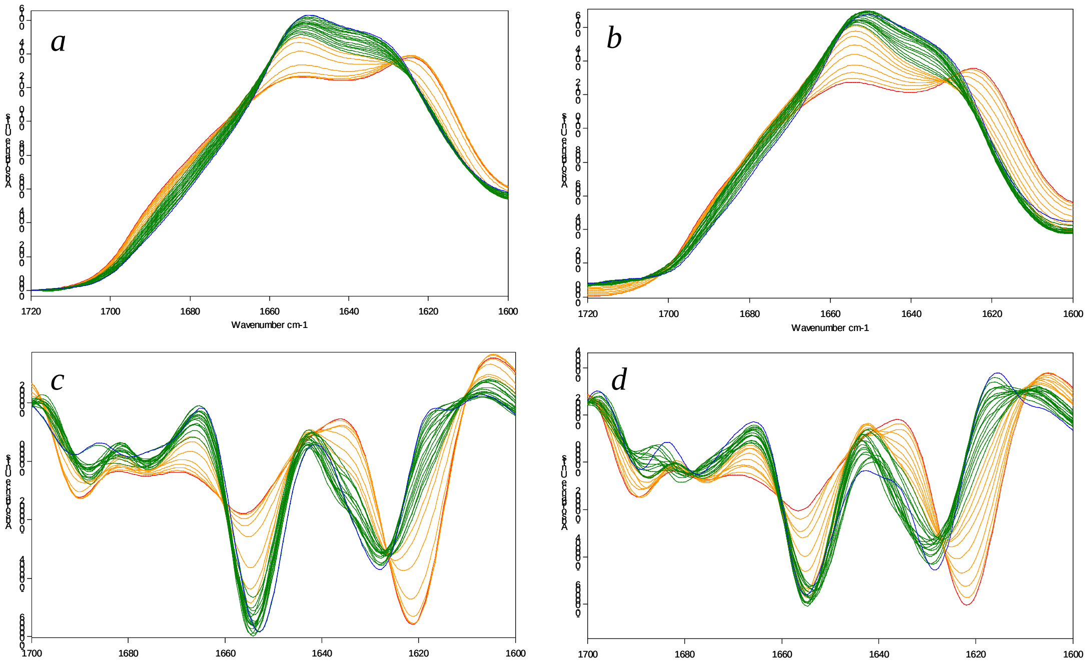


Figure IX.8. FTIR melts of P. sup DJ-1. The amide-I absorption melt overlays for P. sup DJ-1 alone and in the presence of 600 mM H_2O_2 are shown as *a* and *b*, respectively, while *c* and *d* are the corresponding 2ND derivative overlays. Spectra shown as blue, green, orange and red were taken at 13, 16-61, 64-85 and 88 ^{o}C, respectively.

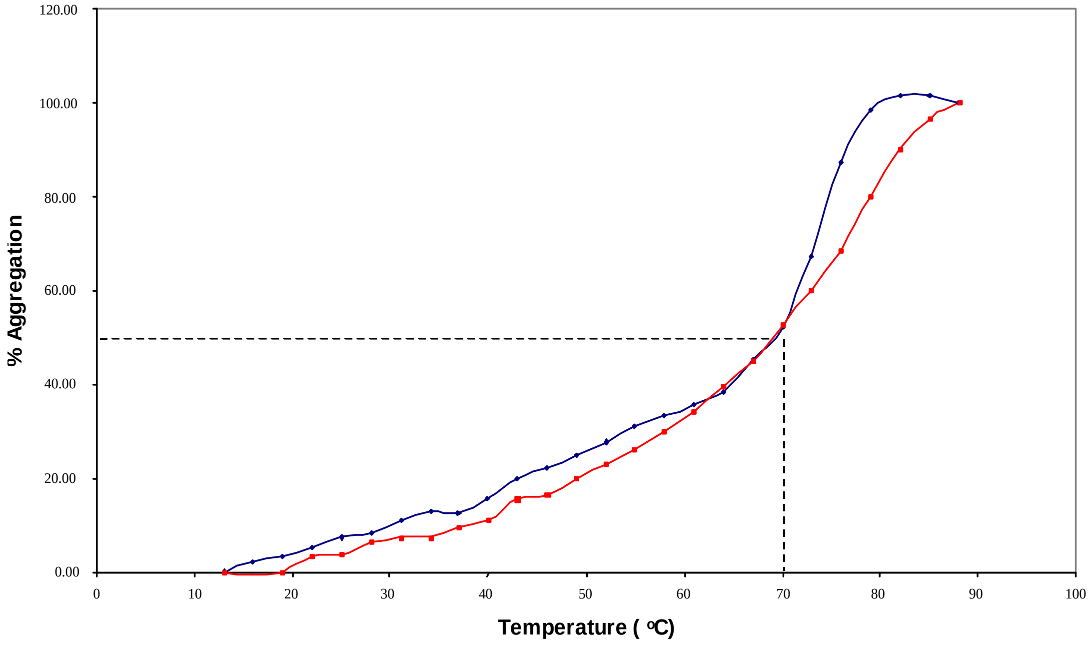


Figure IX.9. Aggregation profile of P. sup DJ-1. The aggregation profile of P. sup DJ-1 is shown both in the absence (blue) and presence (600 mM) (red) of H_2O_2, with both having a $T_{1/2}$ value of 70 °C.

IX.3.6 The effect of Oxidation and Reduction to P. Sup DJ-1

From Figure IX.5 it can be seen that subjecting P. sup DJ-1 to high concentrations of H_2O_2 (600 mM) or DTT (5 mM) had little or no effect on its secondary structure. Also, Figures IX.7, IX.8 and IX.9 show that P. sup DJ-1 does not become destabilised under oxidising conditions, either in respect to its helix domains or its ability to withstand aggregation beyond the limits of most other proteins. This data supports the theory that P. sup DJ-1 functions as a chaperone protein that protects other proteins from denaturing under oxidative stress.

IX.4 Discussion

Table IX.3 shows the x-ray and predicted quantitative secondary structure composition of human DJ-1, along with the predicted, CD and FTIR estimates of the fractional composition of P. sup DJ-1. From this the PROF program significantly underestimated the helix content, but gave very good correlation of sheet content for human DJ-1. Good agreement is seen when comparing the FTIR and CD estimation of helix content for P. sup DJ-1. In this case the amount of helix is likely to be closer to the minimum of the 36-44 % range, since the CDSSTR program is responsible for defining the range's upper limit and this program tends to over estimate protein helix content. The 37.83 % helix content predicted by FTIR supports this interpretation. Using CD to determine the extent of sheet structure within a protein is less than ideal, due to the irregularity and low intensity of sheet CD signal, and this is reflected in the relatively large 15-25 % range of estimated sheet content from CD data. FTIR is considered a superior method for evaluating the total amount of β-sheet content than CD and, therefore, the value of 29.17 % is taken as the best estimate for sheet content. According to the data given in Table IX.3, therefore, P. sup DJ-1 contains less helix and more sheet content than the human form. Notwithstanding this, the two proteins are predicted to be structurally homologous (Figure IX.4). Similar to that of the human form, the significant under-prediction of helix content is again seen for P. sup DJ-1, whereas the prediction for sheet content is in good agreement with that suggested from FTIR data.

Protein	**% Helix**	**% Sheet**	**% Other**
H. DJ-1 (predicted)	25.13	22.99	51.88
H. DJ-1 (x-ray)	48.66	20.86	30.48
P. sup DJ-1 (predicted)	28.65	25.95	45.50
P. sup DJ-1 (CD)	36-44	15-25	31-49
P. sup DJ-1 (FTIR)	37.83	29.17	33.00

Table IX.3. Quantitative secondary structures of human and P. sup DJ-1.

The up-regulation of chaperone heat-shock or oxidative stress proteins in response to cell desiccation is consistent with expectation, since the loss of water represents a considerable stress from which the cell's native protein molecules must be protected. It is, therefore, not surprising that a DJ-1 homolog should have been discovered as an anhydrobiotic protein within P. sup**.** From the FTIR and CD examination of P. sup DJ-1, it is clear that both the α-helix and β-sheet contents are roughly similar to that of human DJ-1 and in addition there is good agreement between the two homologs in both the estimated number of α-helices and β-sheets, along with their respective average lengths. Taken together, this data lends strong support to the predicted P. sup DJ-1 structure, which can be seen to be in good agreement with the human form, in terms of the locations of helices and sheets within both proteins.

Given the structural similarity between the human and P. sup DJ-1 homologs, it seems reasonable to speculate that they may possess similar functionalities. Previous examinations of human DJ-1 suggest that it is capable of carrying out a variety of specific functions. Among the most important of its functions is its ability to aid other proteins to retain their structures during conditions of oxidative stress and to perturb their heat-induced denaturation. If P. sup DJ-1 can be assigned such functionality it should show tolerance to both oxidative and heat stress. According to both the CD and FTIR melt experiments, P. sup DJ-1 demonstrates high thermo-stability and this is unaffected by the introduction of oxidative stress conditions. Furthermore, the structure of P. sup DJ-1 was seen to be unchanged when subjected to either oxidative or reductive stress. In conclusion, therefore, from the data presented here it seems plausible at least that the novel P. sup DJ-1 protein functions during anhydrobiosis by acting as a molecular chaperone to relieve oxidative and/or thermal stress from other native proteins within *Panagolaimus superbus*.

The significance of the absence of Cys-53 from both the *P. superbus* and *C. elegans* DJ-1 proteins could be related to an absence of redox control for their functioning as chaperones during oxidative stress. Studies being carried out at present indicate that P. sup DJ-1 does indeed inhibit the heat induced aggregation of α-synuclein, insulin and citrate synthase and that this chaperone activity is not effected by the protein's redox state (in press). This is in contrast to findings made with human DJ-1, where chaperone activity was only demonstrated when the protein was in its oxidised state. Furthermore, human DJ-1 chaperone activity was abrogated upon mutation of Cys-53. It was also discovered by CD spectroscopy that human DJ-1 secondary structure was considerably affected by exposure to 50 mM H_2O_2.[25]This is in contrast to P. sup DJ-1, which demonstrated structural stability up to 600 mM H_2O_2 concentration. These data indicate that human DJ-1 possess a sensitivity towards its redox environment where non exists for P. sup DJ-1 and it is possible that Cys-53 may have some involvement in human DJ-1 being a redox-dependent molecular chaperone.

References

1 Rost B. PHD: Predicting one-dimensional protein structure by profile-based neural networks. Computer Methods for Macromolecular Sequence Analysis 266:525-539, 1996.

2 Tao X, Tong L. Crystal structure of human DJ-1, a protein associated with early onset Parkinson's disease. Journal of Biological Chemistry 278(33):31372-31379, 2003.

3 Nagakubo D, Taira T, Kitaura H, Ikeda M, Tamai K, IguchiAriga SMM, Ariga H. DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras. Biochemical and Biophysical Research Communications 231(2):509-513, 1997.

4 Takahashi K, Taira T, Niki T, Seino C, Iguchi-Ariga SMM, Ariga H. DJ-1 positively regulates the androgen receptor by impairing the binding of PIASx alpha to the receptor. Journal of Biological Chemistry 276(40):37556-37563, 2001.

5 Hod Y, Pentyala SN, Whyard TC, El-Maghrabi MR. Identification and characterization of a novel protein that regulates RNA-protein interaction. Journal of Cellular Biochemistry 72(3):435-444, 1999.

6 Lee SJ, Kim SJ, Kim IK, Ko J, Jeong CS, Kim GH, Park C, Kang SO, Suh PG, Lee HS and others. Crystal structures of human DJ-1 and Escherichia coli Hsp31, which share an evolutionarily conserved domain. Journal of Biological Chemistry 278(45):44552-44559, 2003.

7 Mitsumoto A, Nakagawa Y, Takeuchi A, Okawa K, Iwamatsu A, Takanezawa Y. 2001. Oxidized forms of peroxiredoxins and DJ-1 on two-dimensional gels increased in response to sublethal levels of paraquat. Free Radical Research 35(3):301-310.

8 Mitsumoto A, Nakagawa Y. DJ-1 is an indicator for endogenous reactive oxygen species elicited by endotoxin. Free Radical Research 35(6):885-893, 2001.

9 Quigley PM, Korotkov K, Baneyx F, Hol WGJ. The 1.6-angstrom crystal structure of the class of chaperones represented by Escherichia coli Hsp31 reveals a putative catalytic triad. Proceedings of the National Academy of Sciences of the United States of America 100(6):3137-3142, 2003.

10 Horvath MM, Grishin NV. The C-terminal domain of HPII catalase is a member of the type I glutamine amidotransferase superfamily. Proteins-Structure Function and Genetics 42(2):230-236, 2001.

11 Halio SB, Blumentals II, Short SA, Merrill BM, Kelly RM. Sequence, expression in Escherichia coli, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon Pyrococcus furiosus. Journal of Bacteriology 178(9):2605-2612, 1996.

12 Du XL, Choi IG, Kim R, Wang WR, Jancarik J, Yokota H, Kim SH. Crystal structure of an intracellular protease from Pyrococcus horikoshii at 2-angstrom resolution. Proceedings of the National Academy of Sciences of the United States of America 97(26):14079-14084, 2000.

13 Mizote T, Tsuda M, Nakazawa T, Nakayama H. The thiJ locus and its relation to phosphorylation of hydroxymethylpyrimidine in Escherichia coli. Microbiology-Uk 142:2969-2974, 1996.

14 Mizote T, Tsuda M, Smith DDS, Nakayama H, Nakazawa T. Cloning and characterization of the thiD/J gene of Escherichia coli encoding a thiamin-synthesizing bifunctional enzyme, hydroxymethylpyrimidine kinase phosphomethylpyrimidine kinase. Microbiology-Sgm 145:495-501, 1999.

15 Bandyopadhyay S, Cookson MR. Evolutionary and functional relationships within the DJ1 superfamily. Bmc Evolutionary Biology 4:-, 2004.

16 Lang AE, Lozano AM. Parkinson's disease - First of two parts. New England Journal of Medicine 339(15):1044-1053, 1998.

17 Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R and others. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276(5321):2045-2047, 1997.

18 Shendelman S, Jonason A, Martinat C, Leete T, Abeliovich A. DJ-1 is a redox-dependent molecular chaperone that inhibits alpha-synuclein aggregate formation. Plos Biology 2(11):1764-1773, 2004.

19 Stokes AH, Hastings TG, Vrana KE. Cytotoxic and genotoxic potential of dopamine. Journal of Neuroscience Research 55(6):659-665, 1999.

20 Floor E, Wetzel MG. Increased protein oxidation in human substantia nigra pars compacta in comparison with basal ganglia and prefrontal cortex measured with an improved dinitrophenylhydrazine assay. Journal of Neurochemistry 70(1):268-275, 1998.

21 Yoritaka A, Hattori N, Uchida K, Tanaka M, Stadtman ER, Mizuno Y. Immunohistochemical detection of 4-hydroxynonenal protein adducts in Parkinson disease. Proceedings of the National Academy of Sciences of the United States of America 93(7):2696-2701, 1996.

22 Kinumi T, Kimata J, Taira T, Ariga H, Niki E. Cysteine-106 of DJ-1 is the most sensitive cysteine residue to hydrogen peroxide-mediated oxidation in vivo in human umbilical vein endothelial cells. Biochemical and Biophysical Research Communications 317(3):722-728, 2004.

23 Wilson MA, Collins JL, Hod Y, Ringe D, Petsko GA. The 1.1-angstrom resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease. Proceedings of the National Academy of Sciences of the United States of America 100(16):9256-9261, 2003.

24 Canet-Aviles RM, Wilson MA, Miller DW, Ahmad R, McLendon C, Bandyopadhyay S, Baptista MJ, Ringe D, Petsko GA, Cookson MR. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. Proceedings of the National Academy of Sciences of the United States of America 101(24):9103-9108, 2004.

25 Zhou WB, Zhu M, Wilson MA, Petsko GA, Fink AL. The oxidation state of DJ-1 regulates its chaperone activity toward alpha-synuclein. Journal of Molecular Biology 356(4):1036-1048, 2006.

26 Zhou WB, Freed CR. DJ-1 up-regulates glutathione synthesis during oxidative stress and inhibits A53T alpha-synuclein toxicity. Journal of Biological Chemistry 280(52):43150-43158, 2005.

27 Clements CM, McNally RS, Conti BJ, Mak TW, Ting JPY. DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2. Proceedings of the National Academy of Sciences of the United States of America 103(41):15091-15096, 2006.

28 Kim RH, Peters M, Jang YJ, Shi W, Pintilie M, Fletcher GC, DeLuca C, Liepa J, Zhou L, Snow B and others. DJ-1, a novel regulator of the tumor suppressor PTEN. Cancer Cell 7(3):263-273, 2005.

29 Tang BS, Xiong H, Sun P, Zhang YH, Wang DL, Hu ZM, Zhu ZH, Ma H, Pan Q, Xia JH and others. Association of PINK1 and DJ-1 confers digenic inheritance of early-onset Parkinson's disease. Human Molecular Genetics 15(11):1816-1825, 2006.

30 Honbou K, Suzuki NN, Horiuchi M, Taira T, Niki T, Ariga H, Inagaki F. Crystallization and preliminary crystallographic analysis of DJ-1, a protein associated with male fertility and parkinsonism. Acta Crystallographica Section D-Biological Crystallography 59:1502-1503, 2003.

31 Honbou K, Suzuki NN, Horiuchi M, Niki T, Taira T, Ariga H, Inagaki F. The crystal structure of DJ-1, a protein related to male fertility and bParkinson's disease. Journal of Biological Chemistry 278(33):31380-31384, 2003.

Chapter X

Conclusion

From this work it is clear that both CD and FTIR spectroscopy are appropriate for the investigation of protein secondary structure, since both techniques return structurally characteristic and reproducible protein spectra. Together with reference sets containing proteins of known structure, both CD and FTIR can be used to accurately determine secondary structure fractional composition. However, due to the intense overlapping OH bending vibration for the water solvent molecules with that of the protein amide-I band, it was found that in general FTIR was a little less reliable in terms of achieving precisely reproducible spectra and, consequently, secondary structure composition estimates. The superiority of FTIR over CD for revealing protein β-sheet content, along with the relative weakness of CD in this regard, makes FTIR a highly complimentary technique to that of CD for protein analysis. The procedure for jointly using the CD and FTIR techniques should be, therefore, to use CD to assess the protein helix content and FTIR to determine its sheet composition.

A difficulty with using FTIR to examine the helix content of proteins in aqueous solution is encountered when the protein also contains significant amounts of random coil content. This problem arises because of the closeness of the helix (1648-1655 cm^{-1}) and random coil (1640-1645 cm^{-1}) amide-I bands, along with the broad bandshape of the latter, which makes it hard to resolve the two. It is possible to perform a separate analysis in D_2O solvent, in which the random coil amide-I band experiences a disproportionately large red-shift, such that it no longer overlaps the helix amide-I band, but instead overlaps that of the β-sheet. This effect was observed for both lysozyme and α-chymotrypsin. In practice, however, FTIR should be used with caution for determining protein helix content, as evidenced by the large errors observed for the myoglobin and hemoglobin FTIR structural estimates.

Somewhat contrary to current opinion, it was found that CD spectral acquisition to wavelengths shorter than 185 nm did not have a pronounced effect on the accuracy of estimation of protein structure. Since it is difficult to achieve accurate spectral measurements at these short wavelengths unless vacuum spectrometers are used, it is perhaps a safer practise to restrict conventional CD analyses to between 260-185 nm. Analysis of fractional compositions of standard and novel proteins using each of the programs contained within the DichroWeb server was found to be quite consistent. For each program there occurred high consistency across all of the protein reference sets available, regardless of the amount of proteins contained within the reference set. These findings are as expected for a reliable structural analysis and such a result lends confidence to the accuracy of the estimated structural fractional composition.

In order to determine a protein's secondary structure from its CD spectrum it is essential that the concentration of the protein within the sample be accurately measured. Because of the dependence of a protein's UV extinction coefficient on its secondary structure, sample concentrations of novel proteins of unknown structure cannot be definitively obtained by UV spectroscopy. Protein concentration is, therefore, most accurately estimated using methods such as the quantitative amino acid analysis technique, prior to CD spectral analysis. By examining the area of the amide-I bands of proteins with distinctly different secondary structures, it was found that the amide-I extinction coefficient was not significantly affected by protein secondary structure. As such, it was possible to generate a concentration calibration curve using FTIR spectroscopy that could be used to accurately determine the concentration of samples of novel proteins.

In terms of performing protein melt experiments, both CD and FTIR spectroscopy should be employed. Because FTIR melt experiments in general require the use of ATR FTIR, this technique is not suitable for accurately measuring small structural changes with temperature. The reason for this is that the process of becoming adsorbed onto the ATR crystal surface distorts the protein's secondary structure in a manner that decreases its helix and increases its sheet content. Since protein adsorption is a function of temperature, the difficulty of identifying spectral changes due to adsorption and those due to temperature effects makes ATR FTIR only suitable for monitoring large changes in protein structure, such as protein aggregation. This, however, is a valuable measurement, since aggregation is one of the principal means by which proteins suffer loss of function and, therefore, stability against aggregation is of crucial importance. While CD cannot measure aggregation phenomenon, due to light scattering effects, it is capable of faithfully monitoring small structural shifts that occur with temperature variation. This is of considerable benefit when examining the possible destabilising influence that a particular mutation may have to a protein's structure. The temperature stability of helix domains should also be principally investigated by CD_{222} melts.

2D correlation spectroscopy is a useful technique for investigating the sequence of structural changes for proteins that occur when some external perturbation is applied to the sample. The most obvious perturbation is that of ramping the temperature from low to high values in order to investigate protein unfolding. The level of detail that can be achieved using this procedure with ATR FTIR is limited, however, due to protein-surface adsorption. Use of this technique over large temperature ranges does not lend itself to CD analysis because of the light scattering caused by protein aggregates. With FTIR, 2D correlation spectroscopy can be used to examine a protein's aggregation process, where it is of biological importance whether aggregation occurs predominantly before or after the loss of its secondary structure. For both myoglobin and α-chymotrypsin, aggregation was found to be preceded by the loss of protein secondary structure, whereas for hemoglobin aggregation occurs at the premolten globule stage, before secondary structural degradation.

While both CD and FTIR only give information on a protein's overall structural composition, as opposed to a detailed description of the geometry of the molecule, these techniques can be used effectively to support structural predictions. Both CD and FTIR were used in this way to study the novel Panagolaimus superbus (P. sup) DJ-1 protein. Since many novel proteins correspond to homologs of other proteins of known structure, this information can be used along with CD and FTIR to create a detailed estimate of the protein's general structure. Additionally, by generating protein fragments it should be possible to use either technique to accurately determine local secondary structure. This was demonstrated for the case of the Rab11-FIP3 protein, where a coiled-coil domain was predicted to exist in the region of residues 463-692. Analysis of the Rab11-$FIP3_{463\text{-}692}$ protein fragment by CD spectroscopy confirmed the presence of a coiled-coil domain. By taking suitably small protein fragments the distribution of structure along the protein backbone can be built up using CD and FTIR spectroscopy. Although this involves considerable effort on the part of the biologist and would be impractical for most studies, it does exist as a possibility for the system in question.

Both CD and FTIR can also be useful in cases where the protein structure has been previously solved, for example by x-ray studies. This was demonstrated with the coiled-coil cap region of Rab11-FIP2, whereby x-ray crystallography identified crucial residues within Rab11-FIP2 that most likely act to stabilise the hydrophobic core of the coiled-coil by shielding it from the external aqueous environment. By generating mutated fragments of the Rab11-FIP2 coiled-coil domain, it was possible to assess each mutant in terms of both its structure and its stability using CD and FTIR. Although mutation of the valine$_{456}$ and leucine$_{457}$ residues to glycines had little effect on the overall secondary structure of the coiled-coil (helix content), CD$_{222}$ melt experiments revealed that performing these mutations had a marked effect on its stability, such that the mutated protein fragment helix content was considerably less stable than that of the wild type. This acts as strong evidence that the valine$_{456}$ and/or leucine$_{457}$ residue(s) plays a vital role in enclosing the coiled-coil hydrophobic core in Rab11-FIP2.

Analysis of protein tertiary structure is generally the preserve of x-ray crystallography or nmr spectroscopy, although some information is possible using near-UV CD to examine amino acid side-chain absorption. As an aid to these techniques, luminescent spectroscopy can be employed to answer specific biologically significant questions. For example, by the site-specific labelling of AavLEA1 mutants it was possible to determine that it exists in a highly extended configuration, rather than as a compacted protein, through use of the FRET technique. Since FRET is a through-space phenomenon, it can be used to examine the spatial relation of important residues within a protein. Alternatively, FRET can be used to detect protein substrate interaction by separately labelling either the protein or substrate with donor or acceptor fluorescent tags.

Structural investigations of two proposed anhydrobiotic proteins supported them as being identified as such. AavLEA1 from the *aphelencus avenae* nematode was found to be a natively unfolded protein that exists in an extended tertiary conformation. It was found that this protein acted synergistically with the trehalose disaccharide to stabilise cytochrome A against temperature induced aggregation and, therefore, may have some heat-shock functionality. In addition, this protein was found to change structure by becoming more helical in response to cation concentration increase, supporting the possibility that it may offer cation-sequestering protection during anhydrobiosis. The likelihood of AavLEA1 forming coiled-coil fibrils in the absence of water that can act as reinforcement to a sugar glass matrix was examined by suspending it in TFE solution. AavLEA1 was discovered to become completely helical in TFE, suggesting that in the absence of a polar environment - as would be the case in a sugar glass - AavLEA1 is perfectly suited towards forming coiled-coil fibrils.

P. sup DJ-1 from the nematode *Panagrolamus superbus* was discovered to have high structural homology to the human DJ-1 form and, as such, exists as an α/β-sandwich protein. P. sup DJ-1 was discovered to have high thermostability, both in terms of its helical domains and its ability to withstand aggregation and also demonstrated extreme tolerance to redox stress, being structurally unaltered when placed at H_2O_2 concentrations of up to 600 mM and DTT concentrations of 5 mM. Furthermore, protein thermostability was shown not to be effected by oxidative conditions. Since human DJ-1 is known to act as a molecular chaperone to proteins such as α-synuclein in a manner that stabilises it against temperature and oxidative stress induced aggregation, it is, therefore, likely that P. sup DJ-1 adopts a similar function. Recent studies on the effect of P. sup DJ-1 to α-synuclein, citrate synthase and insulin heat-induced aggregation processes seem to confirm this conclusion. Also, from these studies it appears that P. sup DJ-1 chaperone activity towards these proteins is not under redox control, in contrast to that of human DJ-1. Because of the extreme nature of the anhydrobiotic process, molecular chaperone activity could be essential to cell survival and DJ-1 proteins may very well be instrumental in this regard.