Untangling Fibrillogenesis: Investigations of the Mechanisms of Amyloid Formation

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of life sciences

2014
Benjamin Gardner
Table of Contents

List of Figures ................................................................................................................. 4
List of Tables......................................................................................................................... 13
List of Equations .................................................................................................................. 14
List of Abbreviations ......................................................................................................... 16
Abstract ................................................................................................................................. 17
Declarations ............................................................................................................................. 18
Copyright Statement ........................................................................................................... 18
Acknowledgements ............................................................................................................. 19

Chapter 1: General Introduction ......................................................................................... 20

Chapter 2: Literature Review ............................................................................................... 22

2.1. Amino Acids ................................................................................................................. 22
2.2. Peptide Structure ......................................................................................................... 24
   2.2.1. Alpha helices ...................................................................................................... 25
   2.2.2. Beta sheet (β-sheet) ......................................................................................... 26
   2.2.3. Turns ................................................................................................................ 28
   2.2.4. PolyProline Type II (PPII) ............................................................................... 29
   2.2.5. Disordered ....................................................................................................... 29
2.3. Fibril Formation and Disease ....................................................................................... 30
   2.3.1. Fibrillogenesis ................................................................................................. 31
   2.3.2. The Mature Fibril ......................................................................................... 34
   2.3.3. The Toxic Species in Disease ........................................................................ 35
   2.3.4. Therapeutic intervention ............................................................................... 36
2.4. Theory of Spectroscopy ............................................................................................... 38
   2.4.1. Vibrational Spectroscopy ............................................................................... 38
   2.4.2. Raman Spectroscopy ..................................................................................... 40
   2.4.3. Raman Optical Activity (ROA) ..................................................................... 41
   2.4.4. UV Raman ..................................................................................................... 44
   2.4.5. Tip Enhanced Raman Spectroscopy (TERS) .................................................. 45
2.5. Summary ....................................................................................................................... 47
2.6. References .................................................................................................................... 48

Chapter 3: A Raman and Raman Optical Activity (ROA) study of amino acids
as a function of pH ............................................................................................................... 55

3.0. Declarations ................................................................................................................. 56
3.1. Abstract ......................................................................................................................... 57
Chapter 4: A Matlab toolbox to aid the analysis of Raman and Raman Optical Activity (ROA) Spectra

4.1. Abstract .................................................................................................................. 88
4.2. Introduction ............................................................................................................ 88
4.3. Two dimensional correlation theory .................................................................... 90
4.4. Toolbox Design and Demonstration .................................................................... 92
4.5. Conclusions ........................................................................................................... 94
4.6. References ............................................................................................................ 95
4.7. Appendix ............................................................................................................... 102

Chapter 5: Monitoring in vitro insulin fibrillogenesis with Raman spectroscopy

5.1. Abstract ................................................................................................................. 113
5.2. Introduction .......................................................................................................... 114
5.3. Materials and Methods ....................................................................................... 115
5.4. Results and Discussion ....................................................................................... 116
5.5. Conclusions .......................................................................................................... 120
5.6. References ........................................................................................................... 121
5.7. Supplementary information .................................................................................. 136
Chapter 6: Investigations of the independent fibrillation of the A and B chains of bovine insulin

6.0. Declarations ................................................................. 142
6.1. Abstract ........................................................................ 143
6.2. Introduction .................................................................... 144
6.3. Materials and Methods .................................................. 145
6.3.1. Sample preparation ..................................................... 145
6.3.2. Conventional Raman Spectroscopy 532 nm excitation. .................................................. 146
6.3.3. Raman optical activity (ROA). ...................................... 146
6.3.4. DUVRR. .................................................................... 146
6.3.5. Atomic force microscopy (AFM). ................................. 147
6.3.6. TERS ........................................................................ 147
6.3.7. Data processing .......................................................... 148
6.3.8. Principal component analysis (PCA) and 2D-Correlation ............................................ 148
6.4. Results and Discussion .................................................. 149
6.4.1. A Raman pH study of A and B chains of Insulin .............. 149
6.4.2. Raman and ROA measurements of intact insulin and its A- and B-chains in the non aqueous environment DMSO-d₆ ............................................................ 151
6.4.3. In situ Raman measurements of A and B chain fibril formation ..................................... 153
6.4.4. DUVRR of A and B chain insulin .................................. 155
6.4.5. AFM measurement of A and B chain insulin fibrils .......................................................... 156
6.4.6 TERS of A and B-chain insulin fibrils ............................. 157
6.5. Conclusions ................................................................... 159
6.5. References .................................................................... 161
6.6. Supplementary Information ............................................. 179

Chapter 7: Investigating the thermal stability of Lysozyme with Raman spectroscopy

7.0. Declarations ................................................................... 189
7.1. Abstract ........................................................................ 190
7.2. Introduction .................................................................... 191
7.3. Materials and Methods .................................................. 192
7.4. Results and Discussion .................................................. 193
7.5. Conclusions ................................................................... 196
7.6. References .................................................................... 197
7.7. Supplementary information ............................................. 207

Chapter 8: Conclusions and further work ................................................. 221

8.1. Conclusions .................................................................... 221
8.2. Future work

8.3. References

Word count: 69,827.

List of Figures

Figure 1. 1: Diagram of L and D forms of an amino acid in general form.

Figure 2. 1: Generalised amino acid structure (A), peptide bond formation (B).

Figure 2. 2: A) Ramachandran plot with cartoon representation of protein structures that occupy indicated regions. B) General amino acid structure, with red arrow describing Φ angle around N-Cα bond, while the blue arrow represents the Ψ angle around the Cα-C bond.

Figure 2. 3: Model helical structures (310, α and π, from left to right) modelled from 10 alanine residues. A) as viewed down the structure from the C terminus while B) shows the side view.

Figure 2. 4: Models of β-sheet structure, the top diagram show three β-strands organised parallel to one another to form parallel β-sheet, while the bottom diagram shows strand 2 anti-parallel to that of strands 1 and 3, thus anti-parallel β-sheet.

Figure 2. 5: Two models representing different turns which occur in protein structure.

Figure 2. 6: Funnel shaped free energy landscape of protein folding (blue) and protein misfolding (red). The blue scheme illustrates protein folding with partially folded intermediates (1 and 2), and the energetically stable folded protein (3). The red scheme illustrates protein aggregation, where amorphous aggregation (1), fibril intermediate oligomer (2) and full fibril (3) are shown respectively. The aggregation landscape is concentration dependent, thus separating it from energy states sampled during free protein folding.

Figure 2. 7: Schematic diagram highlighting key pathways of fibrillogenesis.

Figure 2. 8: Model of fibril structure with a general schematic (A) showing the propagation of β-sheet with respect to fibril axis. Different stages of fibril growth are presented (B–D): filaments, protofibrils and mature fibrils respectively.

Figure 2. 9: Energy diagram of Infrared absorption, Raman and Rayleigh scattering processes and resonance. Electronic states are labelled E and vibrational states are labelled V.

Figure 2. 10: Simple schematic representation of tip enhanced Raman scattering setup in back scattering mode (A). (B) shows a magnified schematic representing the size relationship between a polypeptide chain and the TERS tip, typical sampling distances indicated.

Figure 3. 1: Raman spectra of the amino acids Gly, Ala Val and Leu measured in their fully protonated state (black) and zwitterions (red) and the corresponding ROA spectra for Ala Val and Leu for the same experimental conditions.

Figure 3. 2: Raman spectra of amino acids Ile, Pro, Phe, Trp and Tyr measured in their fully protonated state (black) and zwitterions (red) and the corresponding ROA spectra for Ile, Pro, Phe, Trp and Tyr under the same experimental conditions. Due to the low solubility of Tyr, it was only measured at one pH.
Figure 3.3: Raman spectra of amino acids Arg, Lys, Asn and Gln measured in their fully protonated state (black) and zwitterions (red) (A,C,E and G) and the corresponding ROA spectra for Arg, Lys, Asn and Gln under the same experimental conditions (B, D, F and H).

Figure 3.4: Raman spectra of amino acids Ser, Thr, Cys and Met measured in their fully protonated state (black) and zwitterions (red) (A,C,E and G) and the corresponding ROA spectra for Ser, Thr, Cys and Met measured under the same experimental conditions (B, D, F and H).

Figure 3.5: Raman spectra of amino acids Asp, Glu and His measured in their fully protonated state (black) and zwitterions (red) or green in the case of His (A,C and E) and the corresponding ROA spectra for Asp, Glu and His under the same experimental conditions (B, D, and F).

Figure 3.6: Calculated (Calc) Raman spectra of Val in both its zwitterionic and fully protonated forms (A and C) and the corresponding experimental (Exp) spectra (B and D).

Figure 3.7: Experimental (Exp) and calculated (Calc) ROA spectra of Val, in both its protonated forms (A) and its zwitterionic forms (B).

Figure 3.8: 2D potential energy surface plot of Phi Psi rotation of Val in its zwitterionic form (A) and protonated (B) forms. The colour scale represents energy (kcal/mol).

Figure 3.9: Molecular representation of Val in its zwitterionic form (I) and fully protonated state (II), highlighting the changes in intramolecular hydrogen bonding.

Figure 3.10: Pictorial representation of vibrational modes of Val in its zwitterionic form, for bands between 1300 and 1500 cm\(^{-1}\). For a particular vibrational mode the vibrational intensity is proportional to the sphere volume, and the direction of the atomic displacement is perpendicular to the junction between the two hemispheres (blue or yellow).

Figure 3.11: Pictorial representation of vibrational modes of Val in its protonated form, for bands between 1300 and 1500 cm\(^{-1}\). For a particular vibrational mode the vibrational intensity is proportional to the sphere volume, and the direction of the atomic displacement is perpendicular to the junction between the two hemispheres (blue or yellow).

Figure S3.1: Suggested structures of His under different protonation states and tautomers\(^{[33]}\), \(pK_a\) values taken from\(^{[35]}\).

Figure 4.1: Screen shots of the main GUI (A) and two other sub GUIs which can be ran independently to the main program for Raman baselines (B) and processing ROA spectra (C).

Figure 4.2: A demonstration of the 2D correlation module of the toolbox. A matrix containing 30 spectra (A), each consisting of three Gaussian peaks: peak 1 increasing simultaneously as peak 2 decreases, while peak 3 is invariant for 10 arbitrary perturbation steps before decreasing. The 2D correlation examples include a variance plot (B), the synchronous and asynchronous plots (C and D) a 2D-moving windows plot (E) and 2D perturbation correlation moving windows plot (F). Red indicates positive and large intensity, while blue negative or small intensity change.

Figure 4.3: Four sets of bisignate data (A–D), each of which contains 10 sets of spectra which are undergoing different changes of intensity along arbitrary perturbations. Peaks one and two are progressively increasing in intensity (A), conversely the two peaks are both decreasing in intensity (B), while peak one decreases as peak two increases (C), and (D) where peak one decreases in
each spectrum, while peak 2 is invariant for the first 5 arbitrary perturbation steps before decreasing in intensity.................................................................................................................99
Figure 4. 4: Plots (A–D) display the central intensity of band one (blue) and band two (green) with respect to the arbitrary perturbation range for the four sets of generated data. ........................................100
Figure 4. 5: PCMW plots of the four spectral sets (A–D).................................................................................................................................101
Figure S4. 1: The layout of main GUI workspace. .................................................................................................................................102
Figure S4. 2: SPC spectra import option box.................................................................................................................................103
Figure S4. 3: The layout of the main data processing options provided in the toolbox. .................................................................104
Figure S4. 4: The sub GUI which launches when selecting baseline routine for either Raman or ROA spectra.................................................................105
Figure S4. 5: The layout of the independent Raman baseline sub GUI. ..............................................................................105
Figure S4. 6: The layout of the independent ROA sub GUI.......................................................................................................................106
Figure S4. 7: The data averaging sub GUI.................................................................................................................................107
Figure S4. 8: Smoothing Sub GUI. .................................................................................................................................107
Figure S4. 9: The sub GUI for Interpolation of data.................................................................................................................................108
Figure S4. 10: Window size selection GUI. .................................................................................................................................109
Figure 5. 1: The Raman (A) and ROA (B) spectra of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) at pH 2 (blue) and pH 3 (red) measured at room temperature ..............................................124
Figure 5. 2: FWHH analysis of the amide I region of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) as a function of time (log\(_{10}\) s) at pH 3 (red) and pH 2 (blue) while undergoing heating at 80°C. .................................................................124
Figure 5. 3: Raman spectra of four phases (A–D) of bovine insulin’s (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) fibril formation being monitored \textit{in situ} at pH 3 at 80°C. The first 60 s of heating are contained in (A and B), both containing 15 spectra, respectively. An invariant phase accounting for ~850 s (C) containing 450 spectra, while D contains ~200 spectra accounting for ~400 s ..........125
Figure 5. 4: Raman spectra of four phases (A–D) of bovine insulin’s (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) fibril formation being monitored \textit{in situ} at pH 2 at 80°C. The first 80 s of heating are contained in (A and B), with A containing 15 spectra and B containing 25 spectra, respectively. An invariant phase accounting for ~440 s (C) containing 220 spectra, while D contains ~200 spectra accounting for ~400 s ........................................................................................................126
Figure 5. 5: PCMW analysis of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 3) under heating at 80°C, where the first ~60 s are accounted for in A, while B monitors ~800–1300 seconds. Colour scale identifies direction and intensity of change. ..................................................127
Figure 5. 6: PCMW analysis of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 2) under heating at 80°C, where the first ~60 s are accounted for in A, while B monitors ~400–800 seconds. Colour scale identifies direction and intensity of change. .................................................................128
Figure 5. 7: PC loadings of three phases of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) fibril formation at pH 3 (A–C). ........................................................................................................129
Figure 5. 8: PC loadings of three phases of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 2) fibril formation (A–C). ........................................................................................................130
Figure 5. 9: Figure 5. 9: FWHH of the amide I region of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) as a function of time ($\log_{10}$ s). Sample initially under heating at 80°C prior to quenching to 20°C. Four phases (A–D) are indicated. Phase A accounts for ~0–80 s, B ~80–360 s, C ~470–600 s and finally D, 600 s to 14 hours.

Figure 5. 10: Real time Raman spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) monitoring spectral response as a function of temperature and time: A, the first 40 spectra while undergoing heating at 80°C, accounting for ~0–80 s; B, 140 Raman spectra ~80–360 s, while sample still undergoing heating at 80°C; C, 65 Raman spectra ~470–600 s following quenching of temperature to 20°C; D, 26 Raman spectra (1 every 1000) taken from 600 s – 14 hours.

Figure 5. 11: PCMW of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) under heating (40 spectra) at 80°C (A) and refolding (26 spectra) (B).

Figure 5. 12: PCA scores plots of four sets of 10 Raman spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) under unfolding and refolding conditions, taken at different time points: initial spectra comprise of the first 10 spectra under heating at 80°C; intermediate 1, 10 spectra from the lag phase while being heated at 80°C; intermediate 2, 10 spectra following temperature quenching to 20°C; final, the last 10 spectra of the data set. Loadings plot of PC1 (B).

Figure 5. 13: FWHH of the amide I region of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 2) as a function of time. Sample initially under heating at 80°C, prior to quenching to 20°C. Temperature remained at 20°C for the rest of the experiment.

Figure S5. 1: Cartoon representation of Insulin (pdb:2A3G), with the A chain in green is shown at the front (A), while the B chain is shown at the front (B).

Figure S5. 2: ROA spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) prior to heating (blue), and after initial heating and refolding (red).

Figure S5. 3: Real time Raman spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 2) monitoring spectral response as a function of temperature and time: A, the first 35 spectra while undergoing heating at 80°C, accounting for ~0–70 s; B, 90 Raman spectra ~70–250 s, while sample still undergoing heating at 80°C; C, 89 Raman spectra ~250–430 s following quenching of temperature to 20°C; D, 26 Raman spectra (1 every 1000) taken from 430 s – 14 hour.

Figure S5. 4: PCMW of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 2) under heating (35 spectra) at 80°C (A) and refolding at 20°C (26 spectra) (B).

Figure S5. 5: PCA scores plots of four sets of 10 Raman spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) under unfolding and refolding conditions, taken at different time points: initial spectra comprise of the first 10 spectra under heating at 80°C; intermediate 1, 10 spectra from the lag phase while being heated at 80°C; intermediate 2, 10 spectra following temperature quenching to 20°C; final, the last 10 spectra of the data set. The loadings plot of PC1 (B).

Figure 6. 1: Bulk solution Raman spectra measured at 532 nm of the A chain of insulin (10 mg ml$^{-1}$ in H$_2$O) from pH 1.05 to 4.89.

Figure 6. 2: PCA scores plot along PC1 and PC3 (A) comprising five Raman measurements at each pH for the A-chain of insulin (10 mg ml$^{-1}$ in H$_2$O). Panel B shows PCA loadings plot along PC 1 from the PCA scores plot (A).
Figure 6. 3: Perturbation correlation moving widows plot of the Raman spectra of A-chain insulin (10 mg ml\(^{-1}\) in H\(_2\)O) as a function of increasing pH. The direction and intensity of spectral changes is indicated by the colour bar.

Figure 6. 4: Bulk solution Raman spectra measured at 532 nm of the B-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O) from pH 1.06 to 3.91.

Figure 6. 5: PCA scores plot along PC1 and PC2 (A) comprising five Raman measurements at each pH for the B-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O). Panel B shows The PCA loadings plot along PC 1 from the PCA scores plot (panel A).

Figure 6. 6: Perturbation correlation moving widows plot of the Raman spectra of B-chain insulin (10 mg ml\(^{-1}\) in H\(_2\)O) as a function of increasing pH of. Colour scale indicates the direction and intensity of spectral changes.

Figure 6. 7: Raman spectra (532 nm excitation) of intact insulin, A-chain insulin and B-chain insulin all at 20 mg ml\(^{-1}\) dissolved in DMSO-d6 (A) and the corresponding ROA spectra (B).

Figure 6. 8: The in situ Raman measurements (A) (532 nm excitation) of the A-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O HCl adjusted to pH 2.00) while being heated at 30 °C. The corresponding FWHH of the amide I region as a function of time is shown in panel B. The first and last spectra are shown in panel C.

Figure 6. 9: The in situ Raman measurements (A) (532 nm excitation) of the B-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O HCl adjusted to pH 3.00) while being heated at 60 °C. The corresponding FWHH of the amide I region as a function of time is shown in panel B. The first and last spectra are shown in panel C.

Figure 6. 10: The in situ Raman measurements (A) (532 nm excitation) of the B-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O HCl adjusted to pH 2.00) while being heated at 37.5 °C. The corresponding FWHH of the amide I region as a function of time (B), with transitions indicated (I–IV). The 25 spectra corresponding to the period between I and II as indicated in FWHH (B) are shown in panel C. Panel D shows a comparison of the spectrum from the start of phase (III) and the final spectrum at point IV.

Figure 6. 11: In situ Raman measurements (A) (532 nm) of the B-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O HCl adjusted to pH 1.00) while being heated at 60 °C. The corresponding FWHH of the amide I region as a function of time is shown in panel B. Smaller sub sets of the first 140 spectra are presented (C–E), where the f first 10 spectra are shown (C), the 60 following spectra (D) and the final 70 spectra (E).

Figure 6. 12: Deep UV Raman spectra (199.7 nm excitation) of the A (red) and B (blue) chain monomers of insulin (10 mg ml\(^{-1}\) in H\(_2\)O pH adjusted to pH 2 with HCl).

Figure 6. 13: Deep UV Raman spectra (199.7 nm excitation) of the initial A-chain insulin structure (10 mg ml\(^{-1}\) in H\(_2\)O at pH 2.00) (blue) and its fibril after 24 hours incubation at 30°C (red).

Figure 6. 14: The comparison of the DUVRR spectra of B-chain insulin (10 mg ml\(^{-1}\) in H\(_2\)O at pH 2) initial structure (blue) and its fibril (red) (panel A) after incubation at 60°C for 24 hours. The spectral features of the fibril core of the B-chain fibril (panel B) are revealed through H/D exchange with D\(_2\)O and aromatic contribution subtraction.
Figure 6. 15: AFM images of B-chain insulin fibrils formed through incubation at 65°C for 24 hours at two different pHs, ~pH 3 (panels A–D) and ~pH 1 (panels E–H). Samples diluted 100x and 200x respectively from an initial concentration of 10 mg ml⁻¹. Colour scale indicates height in nm, whereas white scale bar depicts distance.

Figure 6. 16: AFM images (panels A–F) of A-chain insulin fibrils formed through incubation at 30°C for 24 hours at pH 2. Samples were diluted 100x from an initial concentration of 10 mg ml⁻¹. Colour scale indicates height in nm, whereas white scale bar depicts distance.

Figure 6. 17: AFM images (A–C) of B-chain insulin fibrils through incubation at 60°C for 24 hours at pH 2. Samples were diluted 100x from an initial concentration of 10 mg ml⁻¹. Colour scale indicates height in nm, whereas white scale bar depicts distance.

Figure 6. 18: Two sets of TERS spectra (A and B) measured along a fibril formed from the A-chain of insulin (C). Measurements were taken at 0.5 nm steps from each other, with panel A corresponding to the region 9.5 nm from the start point to 19.5 nm and panel B, the region 21 to 39 nm from the start point.

Figure 6. 19: Two sets of TERS spectra (A and B) measured along a fibril formed from the B-chain of insulin (C). Measurements were taken at 2 nm steps from each other, with A corresponding to the first 20 measurements, 0 nm to 19.5 nm and panel B, from 21 to 39 nm. The data for panel A are displayed in a rotated 3D plot for better visualisation of the bands.

Figure 6. 20: Schematic model of the pH influence on the fibrillation of the B-Chain of insulin. B-chain of insulin (PDB: 1AI0).

Figure S6. 1: A the cartoon representation of bovine insulin, dimmer shown (PDB: 1AI0). The A-chain is shown in red and the B-chain is shown in green. B, shows the insulin monomer (PDB: 4INS), with the amino acids of the hydrophobic core displayed in red for the A-chain and blue for the B-chain.

Figure S6. 2: Bulk solution Raman spectra (532 nm excitation) of intact insulin (50 mg ml⁻¹ in H₂O) measured under a series as a function of pH, from pH 1.60 to 3.00.

Figure S6. 3: PCA scores plot along PC1 and PC2, with the associated variance explained by PC1 and PC2 (A) of the three samples A chain insulin, B-chain insulin and intact insulin measured at five separate time intervals in DSMO-d6. The PCA loadings plot (B) along PCA 1 from the PCA scores plot (A).

Figure S6. 4: Deconvolution of the amide I region of A-chain Insulin (A), B-chain Insulin (B) and intact insulin (C) in DMSO-d6. Data: blue line; cumulative fit: red line; components of secondary structure: green circles.

Figure S6. 5: Width (A) and height (B) profiles of 20 different fibrils of B-chain insulin formed at pH 3. Each fibril was measured at 7–14 independent points, of which the average is displayed with the corresponding standard deviation.

Figure S6. 6: Width (A) and height (B) profiles of 25 different spherulites of B-chain insulin formed at pH 1. Each spherulite was measured at 3–5 independent points, of which the average is displayed with the corresponding standard deviation.
Figure S6. 7: Width (A) and height (B) profiles of 6 different fibrils of B-chain insulin formed at pH 1. Each fibril was measured at 8–15 independent points, of which the average is displayed with the corresponding standard deviation.

Figure S6. 8: Width (A) and height (B) profiles of 11 different fibrils of B-chain insulin formed at pH 2. Each fibril was measured at 8–15 independent points, of which the average is displayed with the corresponding standard deviation. These fibrils correspond to those for which TERS was recorded.

Figure S6. 9: Width (A) and height (B) profiles of 11 different fibrils of A-chain insulin formed at pH 2. Each fibril was measured at 8–15 independent points, of which the average is displayed with the corresponding standard deviation. These fibrils correspond to those for which TERS was recorded.

Figure S6. 10: The periodicity (A) in height of an A-chain fibril (B) formed at pH 2 along its axial length.

Figure S6. 11: Schematic diagram of the sequence of the A and B chains of bovine insulin. Amino acids unique to one chain are displayed red, while the SRE element of each sequence is indicated by a dashed line.

Figure S6. 12: TERS spectra measured along a fibril formed from the B-chain of insulin. Measurements were taken at 1 nm steps from each other.

Figure 7. 1: The average Raman spectra for HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2) at 20°C as a function of concentration (0, 25, 50, 100, 200, 300, 400, 500 mM) of NaCl (A) and CaCl$_2$ (B), respectively.

Figure 7. 2: PCA scores plot of the non heated control measurements of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 maintained at 20°C) in NaCl (A) and CaCl$_2$ (B), respectively, showing PC1 plotted against PC2. The concentrations of NaCl and CaCl$_2$ given in the keys.

Figure 7. 3: Raman spectra of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 with 0mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue, green and red respectively. The loadings plot of PC1 is shown in panel D.

Figure 7. 4: The amide I region of the initial Raman spectrum (black) and the final (red) for HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2) during heating at 60°C (0 mM NaCl).

Figure 7. 5: The average FWHH analysis of the amide I region of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2) for a series of NaCl concentrations (0, 25, 50, 100, 200, 300, 400 and 500 mM) observed as a function of time while under heating at 60°C for ~15 minutes (A). The first 160 seconds are observed in B, where horizontal error bars representing the standard errors are shown. Sigmoidal fits of the FWHH data of HEWL as a function of time (s) is shown (C) for 0–400 mM NaCl.

Figure 7. 6: PCA Scores plot of PC1 and PC3 for sub sets (10 spectra chosen at 3 time points) of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2) undergoing heating at 60°C in 300 mM NaCl (A). Loadings plot of PC1 (B) and PC3 (C).

Figure 7. 7: The Raman spectra of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 and 400 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA
scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure 7. 8: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 500 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC3 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure 7. 9: The average (of triplicate measurements) FWHH of the amide I region of HEWL (50 mg ml⁻¹ in H₂O at pH 2) being heated at 60°C over a range of CaCl₂ concentrations.

Figure S7. 1: Cartoon representation of HEWL (PDB: LYZ1) visualised in PyMOL (a). Sidechains are displayed in blue for Trp (b), Tyr (c) and Phe(d).

Figure S7. 2: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 25 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st and 2nd runs are blue red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 3: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 50 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 4: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 100 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 5: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 200 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 6: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 300 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 7: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 0 mM CaCl₂) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 8: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 25 mM CaCl₂) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure S7. 9: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 50 mM CaCl₂) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA
scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D. ..........................215

Figure S7. 10: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 100mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D. ..........................216

Figure S7. 11: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 200mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D. ..........................217

Figure S7. 12: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 300mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D. ..........................218

Figure S7. 13: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 400mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D. ..........................219

Figure S7. 14: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 500mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 2nd and 3rd runs are shown as blue and green respectively. The loadings plot of PC1 is shown in panel D. ..........................220
List of Tables

Table 2.1: Structures of the twenty common amino acids. *Glycine lacks a chiral centre. **Proline forms acyclic structure through side chain linkage to the α-amine. Element key, black C, red O, white H, blue N and yellow S. .................................................................23
Table 2.2: Typically observed dihedral angles for β-sheet[11]. .................................................................27
Table 2.3: Diseases involving protein misfolding .........................................................................................30
Table 3.1: A summary of Raman bands and positions for 20 Amino acids in a low pH environment and in their zwitterionic form. Band intensity described in lowercase: weak (wk), medium (m), strong (s) and very strong (vs) while band widths are described in upper case Sharp (Sh), medium (M) and broad (Br). If a band appears as a shoulder it is designated Sho. ..............................................83
Table S5.1: Raman band assignments for insulin[15, 16]. ...........................................................................136
Table S6.1: Raman band assignments for Insulin and its independent chains[56]. ............................179
Table S6.2: TERS band assignments[35, 36]. .........................................................................................180
Table S6.3: Peaks identified (position and percentage areas) through the deconvolution of the amide I region of the A- and B-chain of insulin as well as intact insulin. .................................183
List of Equations

2.1 (a-c) General equations of light................................................................. 38
2.2. Electric dipole.......................................................................................... 40
2.3. Polarizability tensor.................................................................................. 40
2.4. Isotropic invariant...................................................................................... 40
2.5. Anisotropic invariant.................................................................................. 40
2.6. Circular intensity difference...................................................................... 41
2.7. Raman optical activity electric dipole....................................................... 41
2.8. Magnetic dipole.......................................................................................... 41
2.9. Electric quadrupole moment...................................................................... 41
2.10. Electric field vector.................................................................................. 41
2.11. Real oscillating electric dipole................................................................. 42
2.12 Real oscillating magnetic dipole............................................................... 42
2.13. Real oscillating quadrupole dipole........................................................... 42
2.14. Electric dipole-electric dipole tensor....................................................... 42
2.15. Electric dipole-magnetic dipole tensor..................................................... 42
2.16. Electric dipole-electric quadrupole tensor.............................................. 42
2.17. Electric isotropic invariant....................................................................... 42
2.18. Magnetic isotropic invariant.................................................................... 42
2.19. Electric anisotropic invariant.................................................................. 43
2.20. Magnetic anisotropic invariant................................................................ 43
2.21. Quadrupole anisotropic invariant........................................................... 43
2.22. Circular intensity difference for 0° scattering geometry.......................... 43
2.23. Circular intensity difference for 180° scattering geometry....................... 43
2.24. Circular intensity difference for Δ, 90° scattering geometry.................... 43
2.25. Circular intensity difference for Δ, 90° scattering geometry.................... 43
2.26. Reduced circular intensity difference for 0° scattering geometry................ 43
2.27. Reduced circular intensity difference for 180° scattering geometry.......... 43
2.28. Reduced circular intensity difference for Δ, 90° scattering geometry........ 43
2.29. Reduced circular intensity difference for Δ, 90° scattering geometry........ 44
2.30. Optical resolution equation..................................................................... 47
4.1. Dynamic spectrum..................................................................................... 90
4.2. Formal 2D correlation definition............................................................. 90
4.3. The forward Fourier transform of spectral intensity variations observed for the spectral variable ν₁................................................................. 90
4.4. The synchronous spectrum calculated using the Hilbert transform.......... 90
4.5. The asynchronous spectrum calculated using the Hilbert-Noda transform 90
4.6. Hilbert-Noda transform........................................................................... 91
4.7. Calculating the reference spectrum in moving windows.......................... 91
4.8. The dynamic spectrum corresponding to the jth window........................ 91
4.9. The auto correlation intensity of the jth window ................................................................. 91
4.10. The average perturbation of the jth window ...................................................................... 91
4.11. The dynamic perturbation of the jth window .................................................................... 91
4.12. The synchronous perturbation correlation moving window spectra ................................. 92
4.13. The asynchronous perturbation correlation moving window spectra ............................... 92
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D-Cos</td>
<td>Two dimensional correlation</td>
</tr>
<tr>
<td>Aβ</td>
<td>Amyloid-β peptide</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic force microscopy</td>
</tr>
<tr>
<td>APP</td>
<td>Amyloid precursor protein</td>
</tr>
<tr>
<td>B3LYP</td>
<td>Becke 3-parameter, Lee Yang and Parr</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CID</td>
<td>Circular intensity difference</td>
</tr>
<tr>
<td>CT</td>
<td>Charge transfer</td>
</tr>
<tr>
<td>DMSO-d₆</td>
<td>Dimethyl sulfoxide –d₆</td>
</tr>
<tr>
<td>DUVRR</td>
<td>Deep UV resonance Raman</td>
</tr>
<tr>
<td>EM</td>
<td>Electromagnetic radiation</td>
</tr>
<tr>
<td>EPR</td>
<td>Electron paramagnetic resonance</td>
</tr>
<tr>
<td>FWHH</td>
<td>Full width half height</td>
</tr>
<tr>
<td>IDP</td>
<td>Intrinsically disordered protein</td>
</tr>
<tr>
<td>GC</td>
<td>Glutaminyl cyclase</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>LNOs</td>
<td>Low n oligomers</td>
</tr>
<tr>
<td>MW</td>
<td>Moving windows</td>
</tr>
<tr>
<td>NCC</td>
<td>Nucleated conformational conversion</td>
</tr>
<tr>
<td>NICE</td>
<td>National Institute for Health and Clinical Excellence</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PCMW</td>
<td>Perturbation correlation moving windows</td>
</tr>
<tr>
<td>PPII</td>
<td>Polyproline type II helix</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>ROA</td>
<td>Raman optical activity</td>
</tr>
<tr>
<td>ROS</td>
<td>Reductive oxidative stress</td>
</tr>
<tr>
<td>RR</td>
<td>Resonance Raman</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small angle X-ray scattering</td>
</tr>
<tr>
<td>SERS</td>
<td>Surface enhanced Raman scattering</td>
</tr>
<tr>
<td>SNOM</td>
<td>Scanning near field optical microscopy</td>
</tr>
<tr>
<td>SPM</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>SRE</td>
<td>Self recognition element</td>
</tr>
<tr>
<td>ssNMR</td>
<td>Solid state nuclear magnetic resonance</td>
</tr>
<tr>
<td>ThT</td>
<td>Thiflavin T</td>
</tr>
<tr>
<td>TERS</td>
<td>Tip enhanced Raman spectroscopy</td>
</tr>
<tr>
<td>VCD</td>
<td>Vibrational circular dichroism</td>
</tr>
</tbody>
</table>
Abstract

A number of diseases share a common mechanism, which is the misfolding of proteins leading to the formation of insoluble fibrils through a process termed fibrillogenesis. Diseases where fibrillogenesis is central to the pathology include CJD, Huntington’s disease and Alzheimer’s disease, all three being irreversible fatal neurodegenerative disorders. Therefore, an improved understanding of the mechanistic process involved is crucial if there are to be developments in the treatment for these diseases. However, the conventional approaches routinely used in the structural analysis of biomolecules such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopies are restricted in their application to dynamic systems; such as for unfolded proteins, or in studying rapid changes in structure, both of these situations are usually observed in protein misfolding and fibrillogenesis. Vibrational spectroscopic techniques including Raman spectroscopy, Raman optical activity (ROA), tip enhanced Raman spectroscopy (TERS) and deep UV resonance Raman spectroscopy (DUVRR) have been shown to be particularly well suited for the study of biomolecules. This has been found to be especially so for monitoring structural transitions and protein dynamics. Therefore, these techniques are being increasingly used in the study of fibrillogenesis.

The work in this thesis focuses on the use of vibrational spectroscopies to investigate the mechanisms of fibrillogenesis for bovine insulin and lysozyme. Through monitoring the fibrillogenesis of bovine insulin in real time with Raman spectroscopy, it was possible to readily identify distinct phases in the transitions from the native monomeric state to β-sheet rich fibrils. This was possible through combining different analytical methods including: full width half height analysis (FWHH), 2D correlation approaches and principal component analysis (PCA). It was also shown that the pH of the sample, while not appearing to affect the initial structure, had a significant effect on the rate of fibril formation. Moreover, temperature quenching of the sample when in the β-sheet intermediate was shown to disrupt fibril formation and allow disassociation and recovery of the initial native state.

The structure of the independent A and B chains of insulin was shown to be sensitive to pH, which also had a significant effect on fibrillogenesis mechanism. This was especially so for the B chain of insulin where at pH 3 full fibrils were formed, while at the extreme of pH 1 a mixture of spherulites and fragmented fibrils formed. Moreover, using TERS it was shown that there were significant differences in the surface chemistry of the A and B chain fibrils with respect to secondary structure composition.

Raman spectroscopy was used to show that different Hofmeister ions including Na$^+$ and Ca$^{2+}$ affect the initial unfolding of lysozyme differently, indicating different mechanisms of interaction. Further to this, ion concentration had a dramatic effect on the stability of the partially unfolded intermediate, whereby increases in salt concentration rapidly increased the rate of β-sheet formation.

The effect of the protonation state on the structure of the 20 common amino acids was also investigated with Raman and ROA spectroscopy. This study extended to the use of density functional theory (DFT) calculations on Val, which confirmed that changes in pH have an extensive influence on the structure and dynamics of amino acids. The changes were shown to beyond the simple addition of a proton, with variations in side chain environment and hydrogen bonding also being observed.

In order to facilitate spectral treatment and post-acquisition analysis of the large number of Raman and ROA spectra collected in this project, a graphical user interface (GUI) toolbox was developed for use within Matlab. This toolbox contained several common data treatment pre-requisites which are crucial for reliable post-acquisition analysis.

Overall, this work has shown the suitability of vibrational spectroscopies in the conformational analysis of amino acids and the different protein species formed during fibrillogenesis. Specifically, it was shown that through the combination of techniques and chemometric approaches, insightful knowledge can be obtained with regards to mechanisms of fibrillogenesis, be that through changes in pH, ion concentration or thermal induced change.
Declarations

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other University or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DoculInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgements

Firstly, I would like to thank my supervisor Ewan Blanch, for not only giving me the opportunity to undertake a PhD with him, but for his office door always being open for a discussion about work or otherwise. I would like to say a big thank you to all the members of the Blanch research group past and present for their continued support, advice and friendship over the past 4 years, with a particular acknowledgement to Christian Johannessen whose support has gone above and beyond.

I would like to thank Steve Marsden for his help in using equipment. Also, a large thank you to Lorna Ashton, who advised and trained me in the intricacies of chemometrics, but whom also, gave valued feedback and advice. Additionally, I would like to thank Volker and Tanja Deckert and Igor Lednev, and all the members of their two respective labs for allowing me to visit their labs and being incredibly supportive and welcoming.

I would like to thank all my friends and family for their continued support. In particular I need to acknowledge Tom Waterson, Claire McKenzie and Stephen Boothroyd for continued advice support and friendship over the course of the PhD. Finally, a huge thank you to Hannah, who has continually supported me and was my source of motivation and encouragement throughout the write up.
Chapter 1: General Introduction

In structural biology, understanding of function has often linked to the 3D determination of the structure of the protein of interest, and the tool of choice for this has been X-ray diffraction. While this has been very successful over the years, it has its limitations especially in the field of dynamic proteins and proteins that will not crystallize. Unfortunately, both these criteria are met in the continually growing area of protein misfolding and fibrillogenesis. Fibrillogenesis of proteins is an active area of research due to its implication in a number of diseases including the neurodegenerative Alzheimer's disease. Therefore, to be able to investigate fibrillogenesis, emerging techniques and complementary combinations of these techniques are necessary to improve understanding of this complex process.

Optical spectroscopy has been proven as a suitable technique in studying a range of biological samples and is able to supply the crucial structural information that is sought-after. In addition to optical spectroscopy being applicable to a large number of samples types, certain applications are sensitive to the chiral environment of samples, which is important in the analysis of bio molecules. Throughout nature, it has been observed that biomolecules often adhere to certain absolute chemical configurations e.g. chirality. Chirality or handedness is a term to describe when two chemically identical compounds are non super imposable mirror images of each other. Chirality is often observed in biology due to an asymmetric central carbon, as shown in Figure 1.1. which illustrates how amino acids, the building blocks of proteins, are chiral.

![Diagram of L and D forms of an amino acid in general form.](image)

This study aims to better understand fibrillogenesis through the use of spectroscopic studies in a bottom up fashion. Through this work, a range of topics are covered ranging from the importance of data analysis, investigations of amino acids and the fibrillogenesis of proteins.
The second chapter of this thesis is a detailed review of literature that relates to this work, and it can be broken down into three main sections: I, an introduction to peptide and protein science which includes a discussion of amino acids, the formation of peptides and the varied structures they adopt; II, the process of fibrillation, and its role in disease specifically focussing on Alzheimer’s disease and the development of therapeutics; III, a review of the theory behind spectroscopic techniques with a focus on Raman spectroscopy and its derivative techniques.

In chapter 3, a combined computational and Raman and Raman optical activity investigation of the common amino acids as a function of pH is investigated. There were two main aims of this work: firstly, to produce a reference database of amino acids in different pH environments to ease spectral interpretation and secondly, to understand the changes that are observed in spectra with respect to structural changes occurring with the amino acids under differing protonation states.

Chapter 4 focuses on the importance of standardisation of data processing in spectroscopy with high throughput data analysis, in which the advantages of a standard toolkit for the processing of Raman and Raman optical activity spectra are detailed.

In Chapter 5, the fibrillogenesis of human and bovine insulin are studied with Raman spectroscopy, with the aim of understanding how changes in fibrillation environment and sequence of the amino acids affects fibrillogenesis.

In Chapter 6, the independent peptide chains of bovine insulin (A and B) are investigated with combined spectroscopic techniques including: Raman, Deep UV resonance Raman and Tip enhanced Raman spectroscopy. There are two aims of this: firstly, to demonstrate the usefulness in the combination of Raman techniques and, secondly, to further investigate the fibrillation of insulin by the removal of its disulphide bonds.

Chapter 7, the early unfolding events of lysozyme are monitored with Raman spectroscopy, with the aim of understanding how salt concentration, and its placement in the Hofmeister series alters this process.

Chapter 8, the general conclusions of this investigation and a discussion of recommended future work.
Chapter 2: Literature Review

This review has three focus points, firstly an introduction to protein science, building from amino acids to full protein structure. Secondly, an in depth review of fibrillogenesis covering current understanding of protein misfolding and its implications in disease. Finally, this chapter presents a discussion of Raman theory and its application in a range of techniques relevant to this project.

2.1. Amino Acids

Amino acids, which are the building blocks of proteins, consist of an amine (NH$_2$) and a carboxylate (COOH) group attached to a CαHR, where R denotes the side chain which is specific to the amino acid (Figure 2.1:A). Excluding glycine, amino acids contain a chiral centre at the Cα and in biology exist as the L-stereoisomers. There are hundreds of amino acids which occur in vivo\textsuperscript{[1]}; However, only 20 constitute the "common" amino acids or proteiongenic.

![Generalised amino acid structure (A), peptide bond formation (B)](image)

The side chain of the amino acid plays a crucial role in their chemistry and allows its classification into types, including:

- Acidic: Asp, Glu.
- Basic: His, Lys, Arg.
- Polar: Asn, Cys, Gln Gly, Ser, Thr, Tyr.
- Non-polar: Ala, Ile, Leu, Met, Phe, Pro, Trp, Val.

This represents only one such way to group amino acids, others include groupings based upon their tendency to interact with water e.g. hydrophobic and hydrophilic, or chemical groupings e.g. aromatic amino acids, sulphur containing etc. The structures of the 20 common amino acids including the side chains are given in table 2.1.
Table 2. 1: Structures of the twenty common amino acids. *Glycine lacks a chiral centre. **Proline forms acyclic structure through side chain linkage to the α-amine. Element key, black C, red O, white H, blue N and yellow S.
The side chains of amino acids play a varied role in structure and function, including: driving self assembly through stacking interactions\cite{2}, forming hydrogen bonds, formation of disulphide bonds, and influencing structure through conformational freedom and/ or rigidity e.g. Glycine and Proline respectively. Moreover, a number of side chains undergo post translational modification, by the addition of biological moieties, including the addition of: acetyl groups, glycol groups, phosphate groups. The range and function of post translational modifications are numerous, and beyond the scope of this review, however they are well documented\cite{3, 4}.

Amino acids can join together through a condensation reaction between amine and carboxylate groups, forming a peptide bond (Figure 2.1: B). A short chain of amino acids is generally termed a peptide, whilst longer polypeptide chains (50+ amino acids) are called proteins. However, ambiguity remains between the distinction of a large peptide or, conversely, a small protein.

### 2.2. Peptide Structure

The primary structure of a peptide/protein is the amino acid sequence, while the secondary structure is the result of folding of the chain into a structure which is usually the energetically most favourable conformation; exceptions to this rule will be discussed in more detail below (Section 2.3.). The final structure of a protein or peptide is usually determined by its primary structure (Anfinsens thermodynamic hypothesis)\cite{5}, and folding is driven by factors such as the hydrophobic effect.

The structures adopted by peptides can be demonstrated by considering the Phi (Φ) N-C rotational bond angle and the Psi (Ψ) C-C rotational bond angle (Figure 2.2B) where the angles can range from -180° to +180°. These angles can be plotted to produce a Ramachandran plot\cite{6} (Figure 2.2A), showing common types of structures including: α-helix, β-sheet and PPII, and the regions of the Ramachandran plot they occupy. As is evident, not all areas of the Ramachandran plot are or can be occupied, as to do so would be energetically unfavourable e.g. due to steric hindrance and clashes between side chains.
Figure 2.2: A) Ramachandran plot with cartoon representation of protein structures that occupy indicated regions. B) General amino acid structure, with red arrow describing Φ angle around N-Cα Bond, while the blue arrow represents the Ψ angle around the Cα-C bond.

2.2.1. Alpha helices

The α-helix represents the most common type of secondary structure observed in proteins with ≈30% of all residues found in this type of structure. The canonical α-helix (Figure 2.3A & B) defined by its dihedral angles are Φ -57° and Ψ -47°, which forms a right handed helix, in which each residue accounts for a 100° turn and a translation of 1.5 Å (one complete turn is 3.6 amino acids). The α-helix is stabilised by the formation of internal hydrogen bonds between a backbone amide group and back a bone carbonyl group 4 residues earlier, forming an n+4 pattern. The canonical helix described is the most abundant form of helix observed, owing to its stable bonding pattern where the hydrogen bonds formed are linear. Other helical forms include the 3_10 and π helix (Figure 2.3A & B), which exhibit a different hydrogen bonding pattern, n+3 and n+5 respectively. In the case of the 3_10 helix, it causes narrowing of the helix, and now suboptimal hydrogen bonding when compared to canonical α-helix. The 3_10 helix accounts for ≈3% of observed helices, and is usually observed at the C and N terminus of a canonical helix, and has been proposed to be a helical
The π helix is a rare form of helix, due to it being energetically unfavourable, with its angles of $\Phi -57^\circ$ and $\Psi -69.7^\circ$ transitioning the allowed region. However, given its rarity it is most likely to be associated with function in proteins.

![Model helical structures (3_10, α and π, from left to right) modelled from 10 alanine residues. A) as viewed down the structure from the C terminus while B) shows the side view.](image)

2.2.2. Beta sheet (β-sheet)

The β-sheet is another example of a secondary structure found in peptides and proteins, albeit less commonly than the α-helix. In a β-sheet conformation the polypeptide chain is in an extended conformation, and hydrogen bonding occurs between backbones of two adjacent chains or, a folded chain. β-sheet is observed in two forms, parallel and anti-parallel (Figure 2.4.), which occupy angles $\Phi -119^\circ$ and $\Psi 113^\circ$ and $\Phi -139^\circ$ $\Psi +135^\circ$ respectively on the Ramachandran plot. Of the two forms of β-sheet anti-parallel is the most commonly observed, owing to the more ideal
hydrogen bond angle i.e. linear. While idealised examples of β-sheet are shown (Figure 2.4.), more often β-sheet is observed to have a twist, more so in the case of parallel β-sheet, due to the polypeptide chain not being fully extended[11]. Due to the irregularities of structure observed in β-sheet it occupies a large region on the Ramachandran plot as shown in Table 2.2.

**Parallel β-Sheet**

![Parallel β-Sheet Diagram]

**Anti-Parallel β-Sheet**

![Anti-Parallel β-Sheet Diagram]

**Figure 2. 4: Models of β-sheet structure, the top diagram show three β- strands organised parallel to one another to form parallel β-sheet, while the bottom diagram shows strand 2 anti-parallel to that of strands 1 and 3, thus anti-parallel β-sheet.**

**Table 2. 2: Typically observed dihedral angles for β-sheet [11].**

<table>
<thead>
<tr>
<th>β-Sheet</th>
<th>$\Phi^\circ$</th>
<th>$\Psi^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel (Ideal)</td>
<td>-119</td>
<td>+113</td>
</tr>
<tr>
<td>Parallel (non twisted)</td>
<td>-132 to -100</td>
<td>+132 to +103</td>
</tr>
<tr>
<td>Parallel (heavily twisted)</td>
<td>-94 to -65</td>
<td>+158 to +143</td>
</tr>
<tr>
<td>Anti-parallel (Ideal)</td>
<td>-139</td>
<td>+135</td>
</tr>
<tr>
<td>Anti-parallel (non twisted)</td>
<td>-160 to -133</td>
<td>+161 to +118</td>
</tr>
<tr>
<td>Anti-Parallel (heavily twisted)</td>
<td>-108 to -84</td>
<td>+176 to +140</td>
</tr>
</tbody>
</table>
2.2.3. Turns

Turns are an essential part of secondary structure, as they allow for changes in direction of the polypeptide chain, and importantly allowing for adoption of a compact structure\(^{[12]}\). They differ from above described secondary structure elements as they do not maintain a constant dihedral angle through the structure, conversely angles differ markedly between residues. Turns are defined by the number of residues which separate the first and last residue of the turn, classifications of turns include: \(\sigma\), \(\gamma\), \(\beta\), \(\alpha\) & \(\pi\), whose separations range from \(i \rightarrow i \pm 1\alpha\) 5 respectfully. Furthermore, there are further classifications within each main class of turn. Two common types of turn include the \(\beta\)-turn and \(\alpha\)-turn (Figure 2.5.), while shown in the \(\beta\)-turn, all residues taking part i.e. first to last, the \(\alpha\)-turn only shows the connecting residue of the turn. As shown in the example of a \(\beta\)-turn, due to the close proximities, it allows for the formation of hydrogen bonds, providing a stabilising effect, although this is not a requirement for a turn.

![Type I \(\beta\)-Turn](image)

**Figure 2.5:** Two models representing different turns which occur in protein structure.

Another import type of turn is a \(\beta\)-hairpin, which reverses the direction of the peptide backbone, allowing interaction between two secondary structure elements e.g. allowing hydrogen bonding in anti parallel \(\beta\)-sheets. The hairpin is defined by the number of residues in the turn which are not involved in the hydrogen bonding of the sheet. This type of turn can contain previously described turns to achieve this rearrangement of the peptide chain.
2.2.4. PolyProline Type II (PPII)

The polyproline type II (PPII) helix is observed as a left handed helix ($\Phi$ -75° and $\Psi$ 150°) with three residues per complete turn. Due to PPII's extended conformation this inhibits the formation of organized interchain hydrogen bonding that is observed in previously described structures. Moreover, due to this extended conformation it makes the backbone N-H and the C=O groups accessible to water for hydrogen bonding, and hydrogen bonding to side chains. The PPII helix is not restricted to its namesake, and has been observed to be adopted by all amino acid types, with the PPII helix frequently occurring in small segments of proteins, especially extended loops\[13\]. Due to lacking organized hydrogen bonding and inherent flexibility, PPII is indistinguishable from unorganized protein structure using classic methods such as NMR; however, spectroscopic techniques have been successful in recognizing this structure and the frequency with which it is present in proteins\[14\]. Due to the properties of the PPII helix, that is its extended structure, it can readily interconvert between other structural types e.g. $\alpha$-helix and $\beta$-sheet. It is for this reason it is often associated with the fibrillogenesis of proteins\[14\].

2.2.5. Disordered

While a disordered sequence of amino acids is not strictly a type of secondary structure element, but classified by the lack of a statically defined structure, it is important for later discussion points. When considering the Ramachandran plot the previously described protein structures occupy well defined regions with the backbone dihedral angles occurring around an equilibrium conformation, with inter-conversion between regions energetically unfavourable. However, unstructured sequences are not restricted to a strictly defined region of the Ramachandran plot, and are best thought about in terms of the ensemble populations of the plot which can dynamically interchange\[15, 16\]. Disordered structures can range from a sequence of amino acids to an entire protein, which are usually designated intrinsically disordered proteins (IDPs)\[16-18\]. The importance of unstructured regions of proteins and their general presence and occurrence was unappreciated for a long time; this in part is due to the static view of a protein structure that is observed when determined by using techniques such as X-ray crystallography\[16\]. However, the prevalence of unstructured regions, estimates range of up to 30% in eukaryotes\[19\], and their importance in biological function is now widely accepted. The biological role of unstructured regions of a protein to whole proteins is extensive, examples ranging from cell signalling to protein scaffolds\[15\]. The sequence of amino acids is essential in disordered structure, and there are clear predicates separating a disordered structure to that of an ordered sequence. Two commonly observed features of unstructured sequence is a high unbalanced net charge, which is often negative, and a normalized low hydrophobicity i.e. low average mean of hydrophobic residues (e.g. Leu, Ile, Val)\[16, 19\]. The rationalisation of these observations is clear, a low average hydrophobicity will decrease the drive for the structure to compact, while a high unbalanced net charge will lead to charge-charge repulsion. As with organized structure e.g. $\alpha$-helix, $\beta$-sheet etc, the disordered structures of IDPs have been categorized into distinguishing types such as collapsed and extended. Defining
distinguishing types of disorder has been taken further by other groups in which algorithms have been developed to distinguish types of disorder termed flavours arbitrarily designated V, C and S\cite{20}. In this system disorder types were not only separated based upon amino acid sequence, but disorder location and biological function\cite{20}.

2.3. Fibril Formation and Disease

The seminal work of Anfinsen showed that the primary sequence of a protein dictates its 3D structure\cite{5}. The majority of proteins fold into a specific 3D structure; exceptions to this rule of protein folding include previously discussed IDPs, which lack a defined tertiary structure yet still have functional roles\cite{21}. The tertiary structure of a protein is essential to its function, and disruption to the way proteins fold and or a deviation to their final structure can be detrimental, and can result in a disease phenotype. Diseases which result from the misfolding of proteins are numerous, classic examples being cystic fibrosis\cite{22} and type II diabetes\cite{23}. Protein misfolding is a broad term covering a variety of occurrences, such as the inability of a protein to fold into its thermodynamically stable native state, a specific subset involves protein self aggregation to form fibrils termed “amyloid fibrils” through a process termed fibrillogenesis. A number of proteins (approximately 20) have been shown to undergo fibrillogenesis under physiological conditions\cite{24}. Furthermore it has been shown that under specific in vitro conditions, numerous proteins can form fibrils, showing an inherent ability for polypeptides to form fibrils\cite{25}. While proteins usually comprise varied structures, fibrils have a shared homology consisting of a core cross β structure, which is comprised of continuous β sheets which run perpendicular to the axis of the fibril\cite{26}. Due to fibrils and the general fibrillogenesis pathway’s involvement in multiple diseases (Table 2.3.) it makes their study an active area of research\cite{27}.

Table 2.3: Diseases involving protein misfolding

<table>
<thead>
<tr>
<th>Disease</th>
<th>Protein Featured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s disease</td>
<td>β-Amyloid\cite{28]</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>α-Synuclein\cite{29]</td>
</tr>
<tr>
<td>Transmissible Spongiform Encephalopathy (CJD)</td>
<td>Prion\cite{30]</td>
</tr>
<tr>
<td>Type II Diabetes Mellitus</td>
<td>Islet Amyloid Polypeptide (IAPP)\cite{31}</td>
</tr>
<tr>
<td>Huntington’s Disease</td>
<td>Huntingtonin\cite{32]</td>
</tr>
<tr>
<td>Cardiac Arrhythmias</td>
<td>Atrial Natriuretic Peptide\cite{33]</td>
</tr>
<tr>
<td>Haemodialysis-Associated Amyloidosis</td>
<td>β2-Microglobulin\cite{34, 35}</td>
</tr>
</tbody>
</table>

Just focusing on the best known of this set of diseases, Alzheimer’s disease (AD), the clinical burden is staggering. AD is a neurodegenerative disorder and a common cause of dementia in the elderly. Dementia is estimated to affect 700,000 people in the UK, with annual associated costs of £23 billion (a continually rising figure)\cite{36}. Just considering AD the necessity to improve understanding of fibrillogenesis is clear, with the following sections exploring fibrillogenesis in greater detail with an emphasis on its role in diseases such as AD.
2.3.1. Fibrillogenesis

There is considerable interest vested in understanding how functional proteins can misfold and self associate to form fibrils as this process is often associated with several disease pathologies. By being able to understand the properties of proteins that lead to an increased propensity to misfold and why will allow the rational design of drug to intervene and stabilise the native structure.

A conceptually useful way to describe the folding of proteins is through the folding funnel hypothesis (Figure 2.6.). In this scenario the fully folded protein occupies a deep free energy well, surrounded by steep walls. The folding profile can be rough including various non-native local minima which might contain partially folded intermediates, or, include saddle points which the folding species must pass through to reach the global minima i.e. the native protein structure[37]. Under normal conditions intramolecular interactions are kinetically favoured, thus a protein folds into its native state[38]; however, due to the dynamic and transient nature of protein structure, non native structures can be populated which might be prone to aggregation. While proteins usually fold in such a way as to bury sequences which are prone to self associate, transient aggregation prone species might have these exposed. With aggregation sequences exposed, proteins are then able to self associate with other likewise proteins; therefore allowing the intermolecular association of monomers in a concentration dependent fashion (figure 2.6 red scheme).

![Figure 2.6: Funnel shaped free energy landscape of protein folding (blue) and protein misfolding (red). The blue scheme illustrates protein folding with partially folded intermediates (1 and 2), and the energetically stable folded protein (3). The red scheme illustrates protein aggregation, where amorphous aggregation (1), fibril intermediate oligomer (2) and full fibril (3) are shown respectively. The aggregation landscape is concentration dependent, thus separating it from energy states sampled during free protein folding.](image-url)
A structurally diverse range of proteins have been shown to misfold and form fibrils. Investigations to identify amyloidogenic sequences and residues in proteins have been done. An important determinant of amyloidogenic propensity is the hydrophobicity. Moreover, aromatic residues are observed to appear in amyloidogenic sequences, with stacking of aromatic residues including Phe playing a crucial role. While aromatic residues, especially when in groups act to increase aggregation probability, other residues have the opposite effect. Specifically, Gly and Pro are known to be β-sheet destabilizers. Moreover, it is assumed their conservation in protein structures such as a β-sandwich is to inhibit aggregation. Additionally, the overall net charge of the protein is important, with a high charge acting to inhibit fibrillation propensity.

The tendency for a protein to form fibrils has been shown to originate from a small sequence of the overall protein termed the self recognition element (SRE). The SRE constitutes the core of the fibril that is formed through fibrillogenesis. These sequences are generally quite short, with three residues (VYK) responsible for the self association in Tau protein. Moreover, a number of these short sequences have been identified in other proteins including: β-amyloid (KLVFFAE), medin (NFGSVQ), calcitonin (DFNKF), islet amyloid polypeptide (FGAIL & NFGAIL), insulin (VEALYL & LYQLEN) and α-synuclein (VAQKT). While one model might not be sufficient to explain fibrillogenesis for all proteins, certain models have been shown to be applicable to a number of proteins e.g. the NCC model. The NCC model incorporates aspects of other models and has been used to explain experimental results of the assembly of the prion protein Sup35 and insulin fibrillogenesis. The NCC model suggests that oligomers act to alter the normal folding pathway, allowing polypeptides to: (I) rearrange into nuclei, (II) undergo conformational conversion upon association with a structured nuclei, (III) alter structure to a species not part of a more rigid structure, which can rapidly assemble upon nuclei addition. While these models are useful in giving a qualitative overview of fibrillogenesis, they fall significantly short in being useful in any quantitative fashion. Specifically their shortcomings are due to the polymorphic nature of fibrillation, with emerging evidence indicating it occurs through multidimensional energy landscapes. As eluded to, the process of fibrillogenesis is not a simple one step transition from a native protein to that of a fibril, but proceeds through a number of transient polymorphic intermediate structures that reside both on and off the fibril pathway. A full structural and mechanistic understanding of this process is lacking, however key intermediaries of the process have been identified.

The first key step in fibrillogenesis is exposing the aggregation prone segments of the protein through transient local unfolding, or through the native proteins destabilisation. There is extensive...
evidence linking destabilising conditions, and mutations to increased amyloidogenic potential. Moreover folding intermediates have also been recognised as a bridge between the folding and aggregation pathways\cite{53}. The initiation of fibrillogenesis is going to differ between proteins as their initial structural content will differ; a purely α-helix containing protein will have to unfold, acting as a barrier to fibrillation. Conversely a protein high in β-sheet content acts to increase aggregation propensity\cite{39}, while an IDP will have the SRE exposed for interaction. Through this alteration of a proteins native structure, in its now aggregation “active state” it can go onto associate with other monomers\cite{54}. While “active monomer” can enter the fibril pathway, there are other competing pathways including amorphous aggregation and globulomer formation (Figure 2.7.). The early amyloidogenic pathway is complex, and the current consensus is that it's occupied by a number of transient species which may be on or off the pathway. A key identified intermediate structure is termed the oligomer, which describes small soluble species on the fibrillogenesis pathway. Due to the transient heterogeneous nature of these species they are very difficult to study structurally, ergo any definitive information remains obscure\cite{38}. Moreover, the oligomer is the species linked to the toxicity in a number of amyloid diseases, which is discussed in more detail below (Section 2.3.3.). A key aspect of fibrillogenesis is the formation of a “nucleus” from which fibril propagation occurs; from a kinetic point of view the time precluding this is the lag phase of fibrillogenesis, however what constitutes this nucleus is often quite vague. Larger insoluble intermediaries of the fibrillogenesis pathway include protofibrils, which are short and thin fibrilar species; however, a consensus whether these form from oligomer polymerisation or monomer addition is still lacking. Finally, it's postulated that the protofibrils laterally associate, with possible minor structural rearrangement to form the full fibrils\cite{37}.

Figure 2. 7: Schematic diagram highlighting key pathways of fibrillogenesis.
While this initial phase is very complex a general overview can be summarised by the following, as fibrillogenesis progresses the β-sheet content of the species increases as does their overall stability. These species still remain elusive, however progress is being made in characterising these structures [55-57].

2.3.2. The Mature Fibril

While the fully mature fibril might not represent the most toxic aspect of these amyloidogenic disease pathologies, it has received considerable scientific interest. It is now believed that under the correct in vitro conditions most, possibly all, proteins can form fibrils [39, 58]. Furthermore, regardless of the proteins sequence, size or initial structural content the mature fibrils share a similar morphology [39]. The classic definition of a fibril is a cross-β structure, where the peptide chain is in a β-sheet conformation which is parallel to the longitudinal fibril axis, while the β-stands constituting the sheets run perpendicular to the fibril axis (Figure 2.8: A) [39]. Additionally fibrils can be defined by certain chemical and physical properties i.e. binding of certain dyes (Congo red, ThT etc) or their characteristic X-ray fibre diffraction pattern. The X-ray pattern consists of an X-ray reflection of ≈4.8 Å resolution along the fibril direction, and a second X-ray reflection ≈8-11 Å resolution that’s perpendicular to the fibril’s direction [26, 59].

![Figure 2.8: Model of fibril structure with a general schematic (A) showing the propagation of β-sheet with respect to fibril axis. Different stages of fibril growth are presented (B-D): filaments, protofibrils and mature fibrils respectively.](image)

As discussed fibrils result from organized assembly of protein monomers to form the higher ordered structure. Through this process of self association there are a number of morphologically unique prerequisite stages to the final insoluble fibril including filaments and protofibrils (Figure 2.8: B – D). Moreover, the mature fibrils can form even larger deposits known as plaques and inclusion
bodies, the hallmark of various amyloidogenic diseases\textsuperscript{[61]}. While fibrils are associated with pathology of a number of diseases, they are used in nature in a functional way by a diverse range of organisms (bacteria, fungi and insects) with varied physiological functions\textsuperscript{[62, 63]}

While a characteristic structure is adopted by fibrils, morphological differences arise between fibrils of different proteins with respect to the cross $\beta$-structure and strand orientation. Furthermore, the absolute position of amyloid core with respect to the protein is influenced by the position of the SRE in the overall peptide sequence. This has lead to a classification system of how the SRE or “steric zipper” are orientated with respect to each other, in which there are eight overall classes\textsuperscript{[44]}. In this class system the sequences are ordered dependent upon three key criteria: I) the orientation of the strands in the sheet i.e. the type of $\beta$-sheet e.g. parallel or anti-parallel, II) whether the sequences are packed face to face, or face to back and III) the order of sheets with respect to each other e.g. parallel is up-up with anti-parallel being up-down.

A number of fibril forming protein sequences have been resolved, and been shown to pack differently, while insulins two SRE’s (LYQLEN and VEALYL) packing as class VII, while the two sequences of $\beta$-amyloid (GGVVIA and MVGGVV) packed as class IV and VIII respectively. Interestingly, their observation of multiple SRE sequences observed in fibrils of the same sequence can be used to rationalise fibril morphology, which is often observed in fibrils of the same polypeptide sequence. Another important aspect of the fibril structure is the supramolecular chirality, a term used to describe the orientation of the fibril twist e.g. does the fibril contain a left handed or right twist \textsuperscript{[64, 65]}. The supramolecular chirality of fibrils has shown to be detectable by techniques that are sensitive to chirality of molecules, such as circular dichroism (CD) and vibrational circular dichroism (VCD)\textsuperscript{[66]}. Varied studies have shown that the overall twist of the fibril can be influenced by the \textit{in vitro} conditions, such as pH and agitation\textsuperscript{[67, 68]}. Moreover, it has been shown that an interconversion between fibril twists can occur when the conditions are changed\textsuperscript{[69, 70]}. This last observation is significant, as fibrils represent the most energetically stable structure and are seen as being extraordinarily stable; thus, their ability to interconvert between types presents a new paradigm for fibrils. However, while these are interesting \textit{in vitro} observations how they relate to fibrils that form \textit{in vivo} where conditions such as pH, ion concentration and temperature are tightly regulated and whether these observations can be exploited for a therapeutic benefit remains to be seen.

\textbf{2.3.3. The Toxic Species in Disease}

To be able to understand successfully disease pathogenesis and develop therapeutic agents, the toxic aspect of the disease needs to be identified. Our understanding of what constitutes the toxic species in amyloidogenic diseases such as AD has changed from an initial belief that the senile plaques were responsible for AD pathogenesis; however, in the last 15 – 20 years there has been a substantial shift in this opinion, with soluble intermediates of fibrillogenesis being identified as the more toxic aspect of the AD pathogenesis\textsuperscript{[71]}. In AD the protein that is heavily implicated in the disease state is the peptide fragment $\beta$-amyloid. While the soluble intermediates of fibrillogenesis
have been identified as being the toxic species, there is a complex and confusing body of work describing their precise role in disease pathogenesis. There are a plethora of low n-oligomers (LNOs) of β-amyloid (Aβ) that have been identified, including: dimers and trimers, dodecamers, Aβ derived diffusible ligands (ADDLs) and annular assemblies. All of these aforementioned assemblies have been attributed to neuronal disruption through different possible mechanisms. A number of these LNOs have been shown to inhibit in a reversible fashion hippocampal long term potentiation (LTP) which is responsible for learning and memory. It is believed that these assemblies interact with and interrupt neuronal cellular receptors directly through binding, and indirectly through altering their intracellular recycling rates, although how Aβ alters this intracellular property is not fully understood. Furthermore, there is evidence that certain assemblies might embed into the cell wall forming a pore, which could lead to widespread intracellular disruption and changes including: perturbed ionic homeostasis of e.g. calcium, reactive oxygen species production and altered cellular signalling and eventual cytotoxicity. While the discussed assemblies involved either Aβ1–40 or Aβ1–42, cytotoxicity is not restricted to these two forms of Aβ. New observations of Aβ’s toxicity have correlated an increased toxicity and propensity to self associate through certain types of post translational modification. One investigation looked at the effects of how modification of the ε amino group of lysine residues (protein N homocysteinylated), of which Aβ contains one in the SRE element, with a highly reactive cyclic thioester (homocysteine thiolactone) altered fibrillogenesis. Neutralisation of the positive amino group and introduction of a thiol group was seen to stabilise the oligomer intermediates and therefore increase neurotoxicity. Other studies have looked at toxicity of amino truncated Aβ3(pE)-42, with the now N terminal glutamate undergoing a cyclic condensation reaction via glutaminyl cyclase (GC) to form pyroglutamate. This modification has observed to be more neurotoxic, and oligomerise more readily. Furthermore, a mechanism describing Aβ3(pE)-42 neurotoxicity has been developed, with Aβ3(pE)-42 acting as a template driving misfolding of the monomeric Aβ1–42, ergo eluding that Aβ3(pE)-42 might act as the primary step in AD pathogenesis. From this work it is clear a more thorough investigation the role PTM’s play in AD and other related diseases might be a fruitful endeavour in better understanding their pathogenesis. While the role Aβ plays in AD is pivotal, it is not the only cause for the neurological disruption, other aspects include: aberrant phosphorylation of Tau protein, inflammatory response, and reactive oxygen species (ROS) to name a few, which are well covered in various reviews.

2.3.4. Therapeutic intervention

Despite the extensive work of the scientific community in 30 years of active research, a treatment to delay and prevent the onset of Alzheimer’s disease and similar diseases is currently lacking. This is reflected in patient treatment. Looking at the latest National Institute for Health and Clinical Excellence (NICE) guidelines for Alzheimer’s disease the only treatments available act to alleviate symptoms by acting on receptors in the brain: Donepezil, Galantamine and Rivastigmine act on acetylcholine receptors and Memantine on glutamate receptors. However, these treatments
do nothing with regards to disease progression, highlighting how research has failed to translate to patient treatment. To successfully translate research to therapeutic treatment there are a number of areas which require discussion, which rationalize the current failings and discuss future progress.

While some of the causative genetic factors of Alzheimer’s disease are well understood i.e. mutations to presenilins\(^{[61]}\) or the amyloid precursor protein (APP)\(^{[89]}\) and inheritance of certain ε4 apolipoprotein alleles\(^{[90]}\) other genetic aspects are not so well understood. Understanding what upstream biochemical process affect the Aβ steady state, leading to Alzheimer’s, still remains unknown\(^{[27]}\), while there are most likely to be multi-factorial changes in biological activity through ageing it is paramount to understand and isolate them for any chance of preventative medicines to be developed. Furthermore, understanding the role of long term lifestyle choices and Alzheimers susceptibility has been lagging\(^{[27]}\), but this is starting to come under more focus i.e. the role that long term intellectual stimulation\(^{[91]}\), aerobic physical activity\(^{[92]}\) and how levels of social engagement relates to Alzheimer’s susceptibility\(^{[93]}\). While, all these aspects of life style might alter the susceptibility of developing Alzheimer’s over a lifetime, short term changes are unlikely to be of benefit to a symptomatic patient\(^{[27]}\). Arguably the most important area that needs improvement in Alzheimer’s disease is its detection in the clinical setting, specifically being able to identify at risk groups, through routine genetic screening and biochemical assessments to start early preventative treatments. It has been reported that biochemical imbalances are detectable in inherited Alzheimer’s up to 20 years prior to symptomatic presentation\(^{[94]}\), and it has been suggested the same is true for the more common case of sporadic AD\(^{[95]}\), thus giving a large clinical window for intervention. It is critical to be able to detect patients in the pre-symptomatic phase of the disease because, firstly, irreversible neurological changes occur prior to detectable cognitive decline, and, secondly, targeting patients that are symptomatic has shown to be too late for therapeutic intervention to be a success, which is reflected in recent clinical drug trials\(^{[27]}\).

Due to the complex nature of Alzheimer’s disease there are a number of rational drug targets, the foremost targeting Aβ. A number of drugs have targeted Aβ in different ways from monoclonal antibodies increasing the removal of Aβ, to drugs acting to inhibit the enzymes (β and γ secretase) that produce Aβ. Further Aβ targets include targeting and inhibiting its self association i.e. preventing the formation of oligomers; there is an ever increasing list of potential compounds that can achieve this including: polythenols such as Curcumin\(^{[96]}\), rationally designed N-methylated peptides\(^{[97]}\) and collagen type IV\(^{[98]}\) to name a few, for a more extensive list see Amijee et al\(^{[97]}\). Finally, disruption and removal of fibril and plaque deposits have been targeted. However, for any real progress to be made in the treatment of Alzheimers disease it is going to take more years of perseverance within the field to identify the truly toxic species (Section 2.3.3), the reasons for changes in steady states of Aβ and improvements in early clinical diagnosis for any meaningful progress to be made.
2.4. Theory of Spectroscopy

The purpose of this thesis is the application of spectroscopy to studying biological problems, especially protein misfolding and fibrillogenesis. Therefore, the following section details the main experimental approaches used during the thesis, and the theories that govern them.

2.4.1. Vibrational Spectroscopy

Optical spectroscopy is the study of how electromagnetic radiation (EM) interacts with matter; when photons pass through matter they can be absorbed, scattered or not interact\(^ {99}\). When photons do interact with matter this interaction will be informative of the molecule being probed, providing information relating to the molecular structure, which in disciplines such as structural biology can be very important by elucidating the 3D structure of a protein. Vibrational spectroscopy is a specific subsection of optical spectroscopy in which the vibrations of a molecule are probed with EM radiation, typically visible and infrared radiation.

While vibrational spectroscopy is still not the most conventional method to study protein structure, compared to X-ray crystallography\(^ {100-102}\) nuclear magnetic resonance (NMR) or circular dichroism (CD), it provides distinct advantages especially in the area of studying IDPs and protein dynamics\(^ {101}\). While X-ray and NMR are hindered by aspects such as a protein’s dynamic structure and size, these are not obstructions to vibrational techniques.

Two of the most common types of vibrational spectroscopy are infrared (IR) absorbance and Raman spectroscopy. In the case of IR spectroscopy, if the energy of a photon matches that of the energy gap between a ground state and a vibrationally excited state, the photon can be absorbed. Therefore, in essence IR investigates the annihilation of photons when passed through a sample. Conversely a photon can be scattered, when viewed as a propagating oscillating dipole, forming a short lived virtual state from which the photon is quickly re-radiated; in this process the incident photon is annihilated and a new photon is radiated\(^ {99}\). Most scattered photons have the same energy as the incident source therefore the same angular frequency, this type of scattering is termed elastic scattering or Rayleigh scattering. However, a small number of photons (typically 1 in \(10^6 - 10^8\)\(^ {99}\) will interact with the molecules being irradiated, causing polarization of the electron cloud and inducing nuclear motion in which an energy transfer from the incident photon to the molecule occurs, or molecule to the re-radiated photon can occur. In this case the scattered photons have a different energy than that of the incident photon (inelastic scattering) a process termed Raman scattering after C.V. Raman who first demonstrated this phenomenon in 1928\(^ {103}\).

Some useful equations relating radiation to its wavelength (\(\lambda\)), velocity (\(c\)) frequency (\(\nu\)) energy (\(E\)) and wavenumber (\(\alpha\)) are shown (Equations 2.1a – c). From these it is clear to see how energy relates to electromagnetic radiation, and changes in energy affect frequency of scattered photons.

\[
\lambda = \frac{c}{\nu} \tag{2.1a}
\]
Figure 2.9 presents an energy diagram representing absorption, Raman and Rayleigh scattering and Resonance scattering processes. When light interacts with a molecule in its ground vibrational state the scattered photon is described as being Stokes scattered, while if the molecule is in a vibrationally excited state the scattered light is termed anti-Stokes. At room temperature most scattering will be Stokes scattered as most molecules will be in their ground state. However, the ratio between Stokes and anti-Stokes is temperature dependent and can be determined using the Boltzman equation.

\[ \alpha = \frac{\nu}{c} = \frac{1}{\lambda} \]  

\[ \Delta E = E_2 - E_1 = \hbar \nu \]

There are a number of ways in which a molecule can vibrate e.g. stretching (symmetric and asymmetric), rocking, wagging, twisting, scissoring and bending. A molecule’s movement can be described in terms of translation, rotation and vibration. There are three degrees of freedom used in describing a molecule’s translation and usually three for describing rotation, except when describing a linear molecule then there are only two rotational degrees of freedom. Therefore all other degrees of freedom are vibrational, e.g. the number of vibrations of a non linear molecule of N atoms is 3N-6, while for a linear it is 3N-5. Now considering a small peptide consisting of 10
Alanine residues would consist of over 100 atoms, this would give rise to 300+ vibrations. While not all are going to be IR and/or Raman active, or, have a measurable intensity, it is from this large number of vibrations that IR and Raman spectra provide a very information rich probe of a molecule’s structure.

2.4.2. Raman Spectroscopy

As discussed above Raman scattering occurs when light interacts with a molecule, causing distortion to the electron cloud and inducing nuclear motion. The probability to induce a distortion to the electron cloud and create a dipole moment (μ) i.e. a separation of charge within the molecule, is described by the molecules polarizability (α) and the electric field vector (E).

\[ \mu = \alpha E \]  \hspace{1cm} (2.2)

When a photon distorts the electron cloud it does so in all directions, thus when describing the dipole moment all spatial directions need to be considered, e.g. it must be expanded into a tensor using Cartesian notation (x, y, and z). This can be denoted as shown for Raman scattering (Equation 2.3) \[^99\], in the notation $\alpha_{xx}$ the first subscript represents the direction of polarizability of the molecule, while the second refers to the polarization of the incident light.

\[
\begin{pmatrix}
\mu_x \\
\mu_y \\
\mu_z \\
\end{pmatrix} =
\begin{pmatrix}
\alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\
\alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\
\alpha_{zx} & \alpha_{zy} & \alpha_{zz} \\
\end{pmatrix}
\begin{pmatrix}
E_x \\
E_y \\
E_z \\
\end{pmatrix} \]  \hspace{1cm} (2.3)

For samples which are measured in a liquid or gas phase, there is no ordering of the samples axes with regards to the polarization direction of light, for such samples it is possible to express terms as averages of the polarizability tensor which are invariant to rotation, i.e. isotropic ($\bar{\alpha}$) and anisotropic scattering ($\gamma^2$) (2.4 – 2.5) \[^99, 104\]. Where in this description, isotropic scattering is measured with an analyser parallel to the plane of the incident radiation; whereas, the anisotropic scattering the analyser is perpendicular to the incident radiations plane.

\[ \bar{\alpha} = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \]  \hspace{1cm} (2.4)

\[ \gamma^2 = \frac{1}{2} \left[ (\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{yy} - \alpha_{zz})^2 + (\alpha_{zz} - \alpha_{xx})^2 + 6(\alpha_{xy}^2 + \alpha_{xz}^2 + \alpha_{yz}^2) \right] \]  \hspace{1cm} (2.5)
2.4.3. Raman Optical Activity (ROA)

Raman optical activity (ROA) was originally defined by the small observable difference ($\Delta$-values $\sim 10^{-3}$ at best)\cite{105} between the scattered intensity of right and left circularly polarized incident light when interacting with a chiral molecule\cite{105-109}, now termed incident circular polarization (ICP)\cite{105,110}.

Moreover, this effect is not limited to isotropic chiral molecules, but to all molecules under the influence of a magnetic field, and certain orientated materials\cite{105}. The mechanisms describing this effect were developed in the late sixties by Atkins and Barron\cite{111}, and proven experimentally in the early seventies with the small chiral molecule 1-phenylethylamine\cite{108}. In the development of ROA theory a useful experimental observable (Equation 2.6) was defined as the circular intensity difference (CID)\cite{109} which relates the ratio of the difference between the intensity of right ($I^R$) and left ($I^L$) circularly polarized light divided by the sum of both components; in this representation, Raman scattering is equivalent to the denominator ($I^R + I^L$) while ROA intensity is ($I^R - I^L$).

\[
\Delta = \frac{(I^R - I^L)}{(I^R + I^L)}
\]  
(2.6)

In the theory describing Raman scattering usually only the electric dipole term needs to be included, however, to adequately explain the experimentally observable ROA not only does the electric dipole moment ($\mu_\alpha$) (Equation 2.7) need to be considered, but additionally the magnetic dipole moment ($m_\alpha$) (Equation 2.8) and the electric quadrupole moment ($\Theta_{\alpha\beta}$) (Equation 2.9) which are semi classically defined below\cite{105}:

\[
\mu_\alpha = \sum_i e_i r_{i\alpha},
\]  
(2.7)

\[
m_\alpha = \sum_i \frac{e_i}{2m_i} \varepsilon_{\alpha\beta\gamma} r_{i\beta} l_{i\gamma},
\]  
(2.8)

\[
\Theta_{\alpha\beta} = \frac{1}{2} \sum_i e_i (3r_{i\alpha}r_{i\beta} - r_{i\gamma}^2 \delta_{\alpha\beta}).
\]  
(2.9)

in which the particle $i$ at $r_i$ has charge $e_i$, mass $m_i$ and linear momentum $P_i$. Moreover $\varepsilon_{\alpha\beta\gamma}$ represents the third rank unit antisymmetric tensor or Levi-Civita tensor and $\delta_{\alpha\beta}$ is a substitution tensor or Kronecker delta. The electric field vector (Equation 2.10) of a plane-wave light beam having angular frequency $\omega = 2\pi c/\lambda$ travelling in the direction of a unit vector $n$ with velocity $c$:

\[
E = E^{(0)} \exp[-i\omega (t - n \cdot r/c)].
\]  
(2.10)
To simplify ROA theory, it is described in the far from resonance (FFR) approximation. In doing so any contributions from electronic excited states are ignored. The real components as opposed to the imaginary components induced in the molecule, i.e. the oscillating electric dipole, magnetic dipole and electric quadrupole moments by the real part of the electric field vector, with an associated magnetic vector $\mathbf{B} = \mathbf{n} \times \mathbf{E}/c$ and corresponding electric field gradient $\nabla \alpha E^\beta$ in the FFR limit are the following\cite{105,112}:

\begin{equation}
\mu_\alpha = \alpha_{\alpha\beta} E^\beta + \frac{1}{\omega} G'_{\alpha\beta} E^\beta + \frac{1}{3} A_{\alpha\beta\gamma} \nabla \alpha E^\beta + \ldots ,
\end{equation}

\begin{equation}
m_\alpha = -\frac{1}{\omega} G'_{\alpha\beta} E^\beta + \ldots ,
\end{equation}

\begin{equation}
\theta_{\alpha\beta} = A_{\gamma\alpha\beta} E^\gamma + \ldots .
\end{equation}

Utilisation of time-dependent perturbation theory allows development of a quantum-mechanical expression to describe the dynamic molecular property tensors as shown\cite{105,112,113}:

\begin{equation}
\alpha_{\alpha\beta} = \frac{2}{\hbar} \sum_{j=\pi} \frac{\omega_{jn}}{\omega_j^2 - \omega_n^2} \text{Re}(\langle n|\mu_\alpha|j\rangle|\mu_\beta|n\rangle),
\end{equation}

\begin{equation}
G'_{\alpha\beta} = \frac{2}{\hbar} \sum_{j=\pi} \frac{\omega}{\omega_j^2 - \omega_n^2} \text{Im}(\langle n|\mu_\alpha|j\rangle|\mu_\beta|n\rangle).
\end{equation}

\begin{equation}
A_{\alpha\beta\gamma} = \frac{2}{\hbar} \sum_{j=\pi} \frac{\omega_{jn}}{\omega_j^2 - \omega_n^2} \text{Re}(\langle n|\mu_\alpha|j\rangle|\theta_{\beta\gamma}|n\rangle).
\end{equation}

In the above expressions $n$ and $j$ denotes the initial state and the virtual intermediate state of the molecules. Where the virtual state is an intermediate state, which does not correspond to a real energy level of the molecule\cite{99}. The angular frequency separation is given by $\omega_{jn} = \omega_j - \omega_n$ and $\hbar$ is Plank’s constant. Furthermore, $\alpha_{\alpha\beta}$ denotes the electric dipole-electric dipole tensor, while $G'_{\alpha\beta}$ is the electric dipole-magnetic dipole tensor, and finally $A_{\alpha\beta\gamma}$ the electric dipole-electric quadrupole tensor.

Through orientation averaging of the tensor products (Equations 2.14 – 2.16), products invariant to rotation can be described as\cite{105,106}:

Isotropic:

\begin{equation}
\alpha = \frac{1}{3} \alpha_{\alpha\alpha} = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz}),
\end{equation}

\begin{equation}
G' = \frac{1}{3} G'_{\alpha\alpha} = \frac{1}{3} (G'_{xx} + G'_{yy} + G'_{zz}).
\end{equation}
Anisotropic:

\[
\beta(\alpha)^2 = \frac{1}{2} \left( 3\alpha_{\alpha\beta}\alpha_{\alpha\beta} - \alpha_{\alpha\beta}\beta_{\alpha\beta} \right) \tag{2.19}
\]

\[
\beta(G')^2 = \frac{1}{2} \left( 3\alpha_{\alpha\beta}G'_{\alpha\beta} - \alpha_{\alpha\beta}G'_{\alpha\beta} \right) \tag{2.20}
\]

\[
\beta(A)^2 = \frac{1}{2} \omega\alpha_{\alpha\beta}\varepsilon_{\alpha\beta\gamma}\gamma_{\gamma\delta} \tag{2.21}
\]

One important consideration in ROA is the scattering geometries, as the scattering geometries are not equivalent. As shown in Equations (2.22–2.25) the ICP CID ratio can be expressed in terms of the ROA tensors for four geometries for an isotropic chiral sample, in regard to the geometry of where the detector is placed with respect to the direction of the incident light e.g. at 0° the detector is directly behind the sample with respect to the incident light, while conversely at 180° scattering also termed backscattering, where the photons are scattered in the direction of the incident radiation.

\[
\Delta(0°) = \frac{4 [45\alpha G' + \beta(G')^2 - \beta(A)^2]}{c [45\alpha^2 + 7\beta(\alpha)^2]} \tag{2.22}
\]

\[
\Delta(180°) = \frac{24 \left[ \beta(G')^2 + \frac{1}{2} \beta(A)^2 \right]}{c [45\alpha^2 + 7\beta(\alpha)^2]} \tag{2.23}
\]

\[
\Delta_\alpha(90°) = \frac{2[45\alpha G' + 7\beta(G')^2 + \beta(A)^2]}{c [45\alpha^2 + 7\beta(\alpha)^2]} \tag{2.24}
\]

\[
\Delta_\alpha(90°) = \frac{2 \left[ \beta(G')^2 - \left( \frac{1}{2} \right) \beta(A)^2 \right]}{c\beta(\alpha)^2} \tag{2.25}
\]

These equations can be further simplified upon the following assumption, that if the sample molecule consists of only idealized axially symmetrical bonds for which \( \beta(G')^2 = \beta(A)^2 \) and \( \alpha G' = 0 \), the resulting ROA generated is exclusively anisotropic and the equations reduce to the following\[^{105, 107}\]:

\[
\Delta(0°) = 0 \tag{2.26}
\]

\[
\Delta(180°) = \frac{32\beta(G')^2}{c [45\alpha^2 + 7\beta(\alpha)^2]} \tag{2.27}
\]

\[
\Delta_\alpha(90°) = \frac{16\beta(G')^2}{c [45\alpha^2 + 7\beta(\alpha)^2]} \tag{2.28}
\]
From this it is clear that the optimal experimental setup for ROA is that the scattered light is detected at 180° to maximize intensity, as it is double that of scattering at 90°.

As discussed above (Section 2.4.1), in vibrational spectroscopy a large number of vibrations occur in a molecule, which in Raman can cause spectra to be complicated especially when investigating biological molecules. In Raman investigations of proteins, the spectra are dominated by contributions from the side chains of the amino acids, which can act to obscure other contribution such as those arising from the backbone conformation. However, in ROA the opposite is true due to the ROA signal arising from the most chiral and rigid parts of the molecule the resulting signal is rich in information from the backbone, and less influenced by sidechains. However, ROA is not blind to sidechains and can be very sensitive to their absolute conformation such as for tryptophan[114] or the chiral side chain environment of a molecule[115]. Due to the time scale of Raman scattering (≈3.3 × 10^-14 s when considering a band shift at 1000cm^-1) being much smaller than that of protein conformational dynamics e.g. side chain rotation (10^-12–10^-9 s), helix formation (10^-9–10^-6 s), loop motion (10^-12–10^-3 s) and protein folding (10^-6–1 s)[116], ROA provides a snapshot of all the chiral conformations present, making it a sensitive tool in the study of protein dynamics[106, 117].

2.4.4. UV Raman

Usually in respect to Raman scattering the light interacting with a molecule has a frequency which is far from that of any electronic transition, resulting in the molecule being in an excited vibrational state within the ground electronic state. In the theory described above it was assumed that the system was far from resonance, however electronic transitions can contribute to scattering properties, and in doing so can lead to strong signal enhancement, known as the resonance Raman (RR) effect. While in normal Raman scattering there is no interaction with a real excited electronic state, therefore the polarizabilities are considered to be in the electronic ground state (Figure 2.9.). This contrasts with RR scattering, in which the incident light is resonant with allowed electronic transitions. Therefore, the RR bands initiate at electronic ground state vibrational frequencies but their intensities include information on the exited electronic state[118]. The vibration coupled to this electronic transition is greatly enhanced compared to that of non-resonance scattering in the order of 10^6-10^8[102]. Several mechanisms have been shown to contribute to the RR effect, and have been grouped into terms e.g. A (Frank-Codon overlap), and B (Herzber-Teller vibronic coupling) etc[99, 118]. The different terms have different dependencies to be satisfied, for example the A term has two general requirements: the transition dipole must be non zero i.e. the resonant electronic transition must be electric-dipole allowed. For the first criteria to be satisfied the excitation light’s wavelength must be close to that of a charge transfer (CT) or \( \pi-\pi^* \) type

\[
\Delta \omega(90°) = \frac{8\beta(G')^2}{6c\beta(\alpha)^2} \tag{2.29}
\]
absorbance band. Secondly the vibrational overlap integral (F-C factor) between the ground and excited electronic states must be greater than zero. The A term is generally observed to be the dominant cause of large intensities observed in the RR effect \cite{99, 118}. However, the mentioned terms are not the only contributions for the RR effect, and other theories have been developed e.g. time dependent theory \cite{119-121}. For a more in depth discussion of resonance Raman, and the theory governing this phenomenon refer to a number of excellent articles\cite{118, 122, 123}. As discussed, in order for resonance to occur the wavelength of incident light must correspond to a particular electronic transition, for biological molecules such as proteins this usually falls in the UV region. When considering the selectivity of resonance, using a tuneable laser would allow for specific probing of chromophores present in the molecule. In protein investigations this is achievable using available lasers\cite{124}, a range of 190 nm to 240 nm is achievable, which allows specific probing of side chains such as Phe and Tyr at 240 nm whilst lower wavelengths in the deep UV specifically probe \( \pi \rightarrow \pi^* \) of the amide peptide bonds with particular enhancement of C-N bonds, and less so with carbonyl stretch. Thus, by using a tuneable UV laser it is possible to selectively probe the molecule instead of just getting an average spectra consisting of all vibrational modes.

2.4.5. Tip Enhanced Raman Spectroscopy (TERS)

While resonance provides one way to enhance the Raman signal it is not the only approach. An additional way is surface enhanced Raman scattering (SERS). SERS was developed in the 1970s through observation of an increased Raman scattering of pyridine when bound to a roughened silver electrode\cite{125}. From this pioneering work and effort from other many other groups, there has been a rapid expansion of the use of surface enhancement in Raman, and it has been applied in a number of ways in biology from detection of bacterial DNA\cite{126} to imaging in tissue\cite{127}. A full review of SERS and its applications is beyond the scope of this review, however, recent reviews can be found here \cite{128-131}. The SERS enhancement has been attributed to two mechanisms: an electromagnetic effect and a chemical effect; which effect is the most important, and the full origins of both effects are, however, still under debate\cite{132}. While a full understanding of the SERS effect is ongoing, it is clear SERS provides a great enhancement to the Raman signal ranging from \( 10^2 \) - \( 10^6 \) with reported enhancements of up to \( 10^{13} \)\cite{133}. Such enhancements are of great benefit given the fact that the intrinsic Raman effect is weak, allowing limitations such as concentration dependence to be mitigated, with studies of single molecule detections being reported\cite{133, 134}. While SERS can provide a great enhancement of signals, it is by no means a golden bullet for spectroscopy, as it is plagued with issues such as reproducibility, thus any meaningful quantitative analysis can be very challenging\cite{131, 135}.

Due to the aforementioned issues, the technique of Tip Enhanced Raman Spectroscopy (TERS) has been developed. Commonly in SERS (in bulk solution) the analyte is introduced to metal nanoparticles (typically gold or silver) through interaction with a colloidal suspension in order to achieve enhancement of the signal; while in TERS a single metal nanoparticle is attached to an atomic force microscopy (AFM) tip which is then introduced to the sample (Figure 2.10: A), a
methodology first developed by Wessel\textsuperscript{[136]}. By only introducing one nanoparticle there is only a single uniform source of enhancement, whilst in bulk SERS the enhancement is non uniform which causes the reproducibility issues. Thus TERS can provide a quantitative approach to SERS. As with SERS, the TERS effect is caused by two mechanisms, chemical and electromagnetic. The electromagnetic contribution is the larger of the two enhancement mechanisms ($10^5-10^{10}$)\textsuperscript{[135]}, which is due to excitation of localized surface plasmon resonances (SPRs) i.e. collective oscillations of the surface valence electrons of the metal, when irradiated by the laser. The chemical contribution in comparison is much smaller ($10^1-10^2$ enhancement)\textsuperscript{[135]}, and arises from charge transfer interactions between the metal and absorbed molecule. In TERS by coupling vibrational spectroscopy and scanning near field optical microscopy (SNOM), chemical and topographical information about the sample is gathered in unison, with resolution dictated by the tip apex with resolutions of 15 nm and lower being reported\textsuperscript{[135, 137]}.

The great potential of TERS has been demonstrated with studies ranging from monitoring of catalytic reactions on the nanoscale\textsuperscript{[138]} to the sequencing of ribonucleic acid (RNA) molecules\textsuperscript{[139]}. Figure 2.10:B demonstrates a typical TERS experiment in which a TERS tip probes a polypeptide chain, it is expected that only a few amino acids within the chain are exposed to the enhancement field at any one time\textsuperscript{[140]}.  

Figure 2.10: Simple schematic representation of tip enhanced Raman scattering setup in back scattering mode (A). (B) shows a magnified schematic representing the size relationship between a polypeptide chain and the TERS tip, typical sampling distances indicated.
One important technical aspect to appreciate is that the resolution achieved in TERS is in fact far lower than the best resolution that should be achievable (≈200 nm) according to the diffraction limit of visible light; with typical resolutions reported in the range of 15 nm and smaller\textsuperscript{[135]}. The Rayleigh criterion (Equation 2.30) relates the distance, \(d\), that is resolvable to the wavelength of light, \(\lambda\), and the numerical aperture of the lens, \(NA\).

\[
d = 0.61 \frac{\lambda}{NA}
\]

(2.30)

The rules of diffraction normally apply to light which travels through an aperture; there are exceptions, specifically when light travels through an aperture which has a diameter less than the wavelength of the light source. Under these circumstances diffraction theory no longer applies and electromagnetic theory has to be applied. Moreover, when the distance between aperture and sample is very close (range of nm), the area is defined by the aperture and is the so-called near field region. Near field standing, or evanescent, waves exhibit an intensity that exponentially decays as a function of distance from the boundary where they are created. Allowing propagation of evanescent waves to the far field therefore allows breaking of the diffraction limit. In the backscattering TERS setup shown (Figure 2.10: A) the TERS tip not only acts as the source of SERS enhancement for the signal, but also as the aperture-less antenna allowing for the propagation of near field waves at the surface of the sample to the far field allowing for the high resolutions achieved\textsuperscript{[141]}.

2.5. Summary

This review has covered the basics of protein science and conveyed why it is so critical to understand the misfolding of proteins and fibrillogenesis due to its implication in a number of diseases, particularly Alzheimer’s disease. Moreover there was a discussion of spectroscopic techniques that will be central to investigating this phenomenon.

While there have been significant steps forward in the area of fibrillogenesis, there still remain a number of questions that need to be answered, especially concerning the elusive oligomers.
2.6. References


135. Deckert-Gaudig, T. and V. Deckert, Tip-enhanced Raman scattering (TERS) and high-resolution bio nano-analysis-a comparison. Physical Chemistry Chemical Physics, 2010. 12(38).


Chapter 3: A Raman and Raman Optical Activity (ROA) study of amino acids as a function of pH
3.0. Declarations

As the first author of this planned publication, I carried out all of the experimental and computational work, as well as spectral analysis. Christian Johannessen provided training in computational calculation of vibrational spectra.
A Raman and Raman Optical Activity (ROA) study of amino acids as a function of pH

Benjamin Gardner, a Christian Johannessenb and Ewan W. Blancha

a Faculty of Life Sciences, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.
E-mail: E.Blanch@manchester.ac.uk; Fax: +44 (0)161 236 0409; Tel: +44 (0)161 306 5819
b Department of Chemistry, University of Antwerp, 171 Groenenborgerlaan, Antwerp, Belgium, B2020.

3.1. Abstract

Due to the complexity of interpreting Raman spectra of biological molecules such as proteins, reference spectra of the constituting amino acids are invaluable for their detailed analysis. However, the usefulness of these reference spectra can be limited, especially when they consist of measurements not undertaken in an aqueous environment. Furthermore, understanding how these spectra change in different environments is particularly important when considering experimental parameters such as pH. In order to begin to address these issues, we present a Raman and Raman optical activity (ROA) investigation of the 20 common amino acids in solution in both their zwitterionic and fully cationic states. This work showed that protonation led to extensive, and dramatic, changes in the Raman and ROA spectra as a function of pH. This indicated that the structures of amino acids are extensively changed upon protonation, in a manner that goes beyond the simple addition or subtraction of a proton; with, systemic changes occurring, including changes in the side chains and hydrogen bonding. To validate this work further, quantum mechanical (QM) calculations were performed on Val at the density functional theory (DFT) level. The calculations on Val confirmed that changes in vibrational modes extend past those directly implicated in protonation, and disruptions to hydrogen bonding and side chain modes were also evident. These studies enhance the interpretation of spectral changes as a function of pH in Raman and ROA spectra of amino acids and proteins.

3.2. Introduction

Raman spectroscopy provides a fast, non destructive probe of conformational dynamics and is sparing of the amount of sample required, making it a powerful tool for the analysis of biomolecules such as amino acids, sugars and proteins[1-3]. Additionally, through the continual development of better cameras, baseline subtraction routines, methods to suppress fluorescence etc, Raman has overcome many of the classical limiting factors associated with it[4]. The derivative technique Raman optical activity (ROA), which measures the small difference between left (I^L) and right (I^R) circularly polarised Raman scattered light of chiral molecules, allows the stereochemistry of a molecule to be probed[5]. This is especially important with biomolecules, as chirality is observed in
various forms including: optical isomers, axial chirality and helical chirality\cite{8}. Over recent years, Raman and ROA have been increasingly applied to complex biological problems including the study of structural disorder in proteins\cite{9-11}.

The simultaneous measurement of Raman and ROA spectra of samples such as polypeptides provides data that are complementary. Raman spectra are rich in structural information regarding the side chains, while ROA spectra provide detailed information relating to the polypeptide backbone. However, the result of this can be complex to analyse fully\cite{10}. The most limiting aspect of spectroscopy is often the simplest, the correct assignment of bands observed in spectra, which can often be very complex when considering large biomolecules.

Biomolecules, especially amino acids, have a number of chemical groups which can act as proton donors or acceptors, including the N terminal amine and C terminal carboxylate moieties. In addition to these, a number of the sidechains in amino acids are also free to donate and accept protons including: Arg, Asp, Cys, Glu, His, Lys and Tyr. Furthermore, common post-translational modifications such as phosphorylated side chains have been shown to be particularly sensitive to the pH of the environment they are in\cite{12-14}. These chemical behaviours are often essential for the function of proteins and enzymes, especially amino acids such as His. His, is commonly observed in catalytic centres of enzymes due to its variable protonation states around a small pH range\cite{15-17}. The effect of multiple protonation states of amino acids has consequences for the Raman and ROA spectra observed, when the selection rules for vibrational spectra are taken into account. For example (Gly) can exist in three ionisation states: \(\text{NH}_3^+\text{CH}_2\text{COOH}\), \(\text{NH}_2\text{CH}_2\text{COOH}\) and \(\text{NH}_2\text{CH}_2\text{COO}^-\); which would give rise to 27, 24 or 21 vibrational modes, respectively. While this is a general simplification of the problem, as not all vibrations might be Raman active, or have a measurable intensity, this does illustrate the potential effect on the spectra. Moreover, while the number of vibrations will be altered through variable protonation states, this can also affect the chemical behaviour of amino acids, such as the ability to hydrogen bond, which can lead to large variations in the vibrational spectra. In this paper, the effects of protonation on the Raman and ROA spectra of the common amino acids are investigated in solution. To further understand the changes in the Raman and ROA spectra upon protonation, Val was chosen as a case example to perform quantum mechanical (QM) calculations of the Raman and ROA spectra. These calculations were facilitated by density functional theory (DFT)\cite{18} and polarisable continuum models (PCM) of the solvent, in this case water. By combining computational approaches with experimental results, a thorough investigation of protonation effect on the Raman and ROA spectra of the naturally occurring amino acids is presented here.

3.3. Materials and Methods

All L-amino acids and the imino acid Pro were purchased from Sigma Aldrich and used without further purification. Solutions were prepared over a range of concentrations, from 7 to 150 mg ml\(^{-1}\) dependent on the solubility of the specific amino acid. Samples were prepared in MilliQ double
distilled (dd) H$_2$O and were left in their zwitterionic forms, or had their pH adjusted using either HCl or NaOH until the desired pH was obtained. The pH of each sample was measured using a digital pH meter (HANNAH model), to a precision of ± 0.1 pH units. Samples were then loaded into quartz micro-fluorescence cells and all Raman ($I_R + I_L$) and ROA ($I_R - I_L$) spectra were measured using a scattered circular polarization ChiralRaman spectrometer (BioTools Inc., Jupiter, FL, USA) $^{[19]}$. Experimental conditions were as follows: laser excitation 532 nm, spectral resolution 7 cm$^{-1}$, laser power at the sample ~600mW, acquisition times varied from 3–48 hours depending on the sample concentration. Raman and ROA spectra were processed using Matlab 2010 and the documented in-house toolbox$^{[20]}$. All Raman spectra were baseline corrected using the previously described asymmetric least squares smoothing approach$^{[21]}$. This is to remove differences in the baseline attributed to instrumentation imperfection between measurements$^{[10]}$. Furthermore, the Raman spectra were intensity normalised, through numerical scaling for comparative purposes, and all spectra have had the Raman spectrum of water subtracted. The ROA spectra were baseline corrected through the subtraction of an 80 point median filter of the spectra. Any residual noise in the spectra was removed using a 15 point Savitzky-Golay filter.

In order to make comparison of the Raman spectra more quantitative, we present an analysis of the spectral features in terms of their relative intensities, which were ranked to distinguish between bands, central peak position and bandwidths (FWHH). This was accomplished by combining visual inspection of the spectra, 2$^{\text{nd}}$ derivatives and using the Matlab bioinformatics toolbox$^{[22]}$. Gaussian 09$^{[18]}$ was used for all quantum mechanical (QM) calculations described here. Geometry optimisation of Val in both zwitterionic and cationic forms was performed using the DFT functional B3LYP, with the 6311G** basis set. Raman and ROA calculations were then performed on the optimised structures, and the double diffuse function was included in the basis set (B3LYP/6311++G**). Gaussian’s implicit CPCM was used for water solvent model, where water is described as a stationary body which is defined by its dielectric constant$^{[23, 24]}$. A frequency scaling factor of 0.98 was used to present the calculated Raman and ROA spectra. Finally, the rotational energy barrier was calculated for NH$_3^+$ and CO$_2^-$ for the Val zwitterion, and NH$_3^+$ and COOH for Val cation. This was accomplished through a two dimensional scan around the Phi/Psi angles of the NH$_3^-$-Cα bond and C$_\alpha$-COOH/COO$^-$ bond, respectively. The NH$_3^+$ group of Val was rotated in 10° steps through 120° while the COO$^-$ group of Val was rotated in 10° steps through 180° and COOH through 360°.

### 3.4. Results and Discussion

The Raman and ROA spectra for the common amino acids measured in both their zwitterionic form and a fully protonated cationic state are presented in Figures 3.1–3.5; with the exception of Gly, where no ROA was measured due to it being achiral, and Tyr where no Raman or ROA spectra were recorded for its zwitterionic form due to its low solubility at its isoelectric point at the higher pH. Table 3.1 details the observed Raman bands in which their central positions, relative intensities and widths are listed. All subsequent descriptions of spectral observations and changes
are described in terms of protonation, i.e. going from the zwitterion to the fully cationic form of the amino acid, unless otherwise specified.

### 3.4.1. Aliphatic amino acids

For the simplest of the aliphatic amino acids, Gly and Ala, there are considerable differences induced in the Raman spectra upon protonation; most noticeably a decrease in intensity, and down shift in wavenumber of around 25 cm\(^{-1}\) of the \(\nu(\text{CNC})\) symmetric stretch of the amide group occurs. This particular vibration is initially positioned at \(-900\) and \(850\) cm\(^{-1}\) for Gly and Ala, respectively. Other noticeable changes include reduction in the intensity of bands assigned to \(\sigma(\text{CH}_2)\) and \(\sigma(\text{CH}_3)\) deformations and changes (decreases in intensity and positional shifts) in the bands associated with \(\text{CH}_2\) and \(\text{NH}_3\) twisting. Finally, as expected in the low pH environment as the COO group becomes protonated there is a large increase in intensity of the band from C=O stretching found in the region between 1730–1750 cm\(^{-1}\). The ROA spectrum for Ala (Figure 3.1C) shows similar trends to the Raman spectrum upon protonation, specifically a down shift of spectral positions upon protonation, and a reduction of the dominant spectral features between 1300 and 1450 cm\(^{-1}\) where prominent bands at 1305 (+ve), 1356 (+ve) and 1418 (-ve) cm\(^{-1}\) are diminished.

The differences observed between the Raman spectra of Val (Figure 3.1D) in its two forms are similar for those observed for Gly and Ala. Specifically, there is a down shift of bands upon protonation and a general decrease in the intensity of the previously described bands between 1300–1450 cm\(^{-1}\). In contrast, the two bands at 1450 and 1475 cm\(^{-1}\) show little difference in intensity; which indicates that the corresponding vibrational modes of the side chain group, including tertiary \(\sigma(\text{CH})\) deformations are invariant to pH. The changes observed in the ROA spectra of Val (Figure 3.1E) are more subtle with few noticeable differences, except for the loss (upon protonation) of a very intense positive band of unknown origin observed at 1360 cm\(^{-1}\) in the Val zwitterion.

The Raman spectra for the longer chained aliphatic amino acids Leu and Ile (Figure 3.1F–G and Figure 3.2A–B) share many similarities upon protonation with those measured for Val in either the shift in position or loss of bands. The main observation of interest in the fully protonated ROA spectrum of Leu is the downshift and decreased intensity of the band at 1340 (+ve) cm\(^{-1}\) by 10 cm\(^{-1}\), while the ROA spectrum of Ile shows a loss of the bands at 1360 (+ve) and 1420 (+ve) cm\(^{-1}\) respectively. It is noteworthy that of the aliphatic amino acids, only the zwitterion of Val exhibits an unusually intense peak relative to the others in its ROA signature.

### 3.4.2. Cyclic imino acid

The chemistry of Pro differs from that of the other amino acids and is in fact an imino acid, due to the amine group being bound to the R group alkyl chain, therefore making it a secondary amine. When considering its chemical differences, the observed trends in its spectra upon protonation are not dissimilar when compared to those of the other amino acids. In the Raman spectrum of
protonated Pro (Figure 3.2C), there is a splitting of the peaks at 863 cm\(^{-1}\) into two smaller peaks 835 and 866 cm\(^{-1}\), as well as a decrease in the intensity of peaks at 1339 and 1414 cm\(^{-1}\). While a down shift of a number of lower wave number Raman bands was observed in the aliphatic amino acids upon protonation, this is not observed in Pro. The main changes in the ROA spectrum upon protonation (Figure 3.2D) of Pro are the reduction in intensity of bands at positions: 963 (-ve), 996 (-ve), 1042 (+ve) and 1336 (+ve) cm\(^{-1}\). The pH induced structural transitions of Pro have previously been investigated\(^{[26]}\), where it was shown that cationic and zwitterionic forms share similar torsion angles in the pyrrolidine ring. These structural similarities were reflected in the positive band positioned at 940 cm\(^{-1}\), which can be used to monitor changes to the pyrrolidine ring. This band appeared invariant in our measurements, in agreement with the earlier study.

### 3.4.3. Aromatic amino acids

For the amino acids containing an aromatic side chain (Phe, Trp and Tyr), the Raman spectra are dominated by intense bands originating from the ring structure (Figures 3.2E, G and I) which appears invariant with respect to the pHs examined here. In Phe, differences in the Raman spectra (Figure 3.2E) are mainly observed in two regions: 700–900 and 1250–1450 cm\(^{-1}\). In the aforementioned regions, observed changes upon protonation include down shift of bands in the lower wavenumber region, and a decrease of band intensities at positions: 1273, 1341 and 1414 cm\(^{-1}\). The ROA spectrum of protonated Phe (Figure 3.2F) shows a down shift of 10–15 cm\(^{-1}\) of bands between 700 to 850 cm\(^{-1}\) as well as an increase in the intensity of bands at 820 (-ve) and 920 (+ve) cm\(^{-1}\); apart from these differences, there are no other changes which are greater than the signal to noise ratios observed for the two spectra.

The Raman spectra of Trp (Figure 3.2G) are dominated by a number of strong bands associated with the sidechain, and their contribution to the spectra masks weaker Raman bands. However, noticeable changes upon protonation include a down shift of the band at 714 cm\(^{-1}\) by ~7 cm\(^{-1}\) as well as reduction in the intensity of bands at 810 and 866 cm\(^{-1}\). Additionally, a reversal of relative intensities between two bands at 1349 and 1364 cm\(^{-1}\), where the intensity of the latter band decreases and that of the former band increases with protonation. This change is attributed to a Fermi resonance between the in plane N\(_1\)-C\(_8\) stretch and other out of plane vibrations\(^{[17]}\). This marker bands acts as a probe of the hydrophobicity of the local environment where the higher the hydrophobicity the higher the ratio between I\(_{1360}\)/I\(_{1340}\). Therefore, the observed change in the ratio of I\(_{1360}\)/I\(_{1340}\) indicates a decreasing hydrophobicity in the local environment upon decreasing pH.

A further interesting observation is the invariance of the position and intensity of the indole ring vibration at 1550 cm\(^{-1}\), which is termed the W3 mode. The main contribution to this band is the C=C stretch; however, the position of the band is affected by the torsional angles around the C\(_\alpha\)-C\(_\beta\) (\(x_1\)) and C\(_\beta\)-C\(_3\) (\(x_2\)) linkages\(^{[17, 26, 27]}\). Previous work has shown how sensitive the position of the W3 mode is to the described tortional angles\(^{[17, 26, 27]}\); therefore, these results show upon protonation there is no apparent variance in these torsional angles of the sidechain group.
The ROA spectra of Trp (Figure 3.2H) in the two protonation states share a similar profile. However, there are increases in the absolute intensity of peaks: 1085 (+ve), 1140 (-ve), 1250 (-ve) and 1310 (+ve) cm\(^{-1}\). Furthermore, there is an additional positive band observed at 1015 cm\(^{-1}\). These observations show that the chiral centre is sensitive to changes which are apparently hidden by the intense invariant Raman bands of the side chain. Although, there is again no variance observed in the ROA feature at +1550 cm\(^{-1}\) attributed to the previously described W3 mode\(^{[17]}\). In agreement with the Raman spectra this implies there is no significant change in the \(x^{2,1}\) torsional angle for the indole group upon protonation.

Due to the solubility of Tyr being too low (0.5 mg m\(^{-1}\)) for Raman and ROA measurements in its zwitterionic form, only the low pH spectra were measurable (Figure 3.2I,J). Therefore, only a more limited comparison can be made. When comparing the Raman spectrum of Tyr (Figure 3.2I) in solution at low pH, with previous measurements on Tyr as a solid sample\(^{[1, 28, 29]}\) there are changes in the relative intensities of the bands. The most prominent band in solid state measurements of Tyr is at ~830 cm\(^{-1}\) (ring breathing mode) while in solution this is not the case and the band at 1620 cm\(^{-1}\) (in phase ring stretching) is most intense. Comparing the ROA spectrum of Tyr with that of the closely related Phe at low pH similarities observed include bands at 910 (+ve), 1220 (+ve), 1270 (-ve), 1380 (-ve) and 1440 (-ve) cm\(^{-1}\). However, the spectrum of Tyr appears more complex, with a number of additional peaks between 600–1800 cm\(^{-1}\) as well as a number of peaks having the opposite sign to that observed for Phe, including three negative peaks at 855, 1009 and 1060 cm\(^{-1}\).

### 3.4.4. Basic amino acids

For both the basic amino acids (Lys and Arg) the isoelectric point for both is above the \(pK_a\) of the amino group, due to the contribution from the basic side chain. Therefore, the measurements close to the isoelectric point for both Lys and Arg results in the amino group being mostly deprotonated. Therefore, the comparisons for both is the full anion and cation.

The Raman spectrum of Arg (Figure 3.3A) shows a number of changes once fully protonated, including the appearance of a band at 740 cm\(^{-1}\) and the reduced intensity of many bands between 850 and 1000 cm\(^{-1}\). There is a decrease in the ratio \(I_{1089}/I_{1062}\) between the two bands at 1089 (\(C_6N_7H_2\) asym bend) and 1062 cm\(^{-1}\) (\(N_7\text{-C}_6\text{-N}_7\text{symmetric stretch}\))\(^{[30]}\). Other intensity decreases are observed for the bands at 1290, 1320 and 1267 cm\(^{-1}\), which are attributed to a number of vibrations including COO\(^{-}\) symmetric stretch. The loss of COO\(^{-}\) symmetric stretch is as expected upon protonation, which is shown with the loss of the band at 1420 cm\(^{-1}\).

The ROA spectra of Arg (Figure 3.3B) show extensive changes between the two measured pHs. Upon protonation there is a down shift of 30 cm\(^{-1}\) of the feature 910 (-ve), 970 (+ve), 960 (-ve) and 996 (+ve) cm\(^{-1}\), as well as the gain of a negative band at 1026 cm\(^{-1}\) and an increase of intensity of the feature 1060 (-ve) and 1110 (+ve) cm\(^{-1}\). There is an apparent reversal in the chiral signal at a couplet of 1330 (+ve), 1350 (-ve) and a loss of the negative band at 1374 cm\(^{-1}\). The change of sign for the described couplet indicates a significant change in the chiral centre of Arg upon protonation.

There are a number of changes in the Raman spectrum of Lys (Figure 3.3C) upon protonation, including shifts in peak position and changes in intensity from 600 to 1000 cm\(^{-1}\). For the band at
1016 cm\(^{-1}\), which has been assigned to either C\(\beta\)-wag and C\(\gamma\)-scissor \[^{30}\] there is a large increase in the intensity. There are also large increases in the intensity of bands at 1150 and 1189 cm\(^{-1}\) which are attributed to (NH\(_3\) asymmetric rock and Ca-Ha) and N\(_2\)H3 asymmetric rocking with C\(\varepsilon\)-rocking motions, respectively\[^{30}\]. The COO- symmetric stretch related bands observed at 1360 and 1416 cm\(^{-1}\) disappear upon protonation and a small decrease in the band attributed C\(\beta\) bend at 1450 cm\(^{-1}\) \[^{30}\] is also observed. The differences in the ROA signals of lysine upon protonation (Figure 3.3D) are predominantly small shifts in band positions between 600–1000 cm\(^{-1}\) and the relative intensities of the bands in this spectral window. Further differences upon protonation include: the appearance of a negative band at 1075 cm\(^{-1}\), the loss of a sharp negative band positioned at 1120 cm\(^{-1}\), the gain of two negative features at 1287 and 1323 cm\(^{-1}\) and the loss of a sharp negative band at 1455 cm\(^{-1}\).

**3.4.5. Carboxamide containing amino acids**

The Raman spectrum of Asn (Figure 3.3E) shows a number of changes upon protonation including: a down shift of 20 cm\(^{-1}\) of the band at 768 cm\(^{-1}\), the two bands at 805 and 831 cm\(^{-1}\) are replaced by a very intense band at 812 cm\(^{-1}\), the disappearance of the bands at 871, 1331 and 1361 cm\(^{-1}\) as well as a decrease in the peaks between 1400 and 1500 cm\(^{-1}\) and finally the gain of the band arising from C=O stretching at 1743 cm\(^{-1}\). There are corresponding changes in the ROA spectrum of Asn upon protonation (Figure 3.3F) between 700 and 900 cm\(^{-1}\) which consist of a down shift of \(\approx 30\) cm\(^{-1}\) from the bands of the zwitterions at 772 (-ve), 803 (+ve), 866 (-ve), and 890 (+ve) cm\(^{-1}\). Further changes in the spectrum are observed between 1250 and 1500 cm\(^{-1}\), including a slight shifts in band positions and a greatly reduced intensity (2 ½ times) of the negative band at 1405 cm\(^{-1}\).

In comparison to the case of Asn, the Raman spectra of Gln (Figure 3.3: G), which contains an additional CH\(_2\) group in the side chain, shows less change upon protonation. However, a number of shifts in band position and intensity between 700 and 850 cm\(^{-1}\) and the loss of the band at 1350 cm\(^{-1}\) are observed upon protonation. As with Asn, changes are also observed in the band profile between 1400 and 1500 cm\(^{-1}\), specifically changes in the relative intensities of the bands that constitute it, and a shift in the entire profile in its position by 5–10 cm\(^{-1}\). Interestingly, the ROA spectrum of Gln (Figure 3.3H) shows few changes upon protonation, with only small shifts observed in peak widths.

**3.4.6. Hydroxyl containing amino acids**

The Raman spectrum of fully protonated Ser (Figure 3.4A), compared with that of its zwitterion, shows a 15–20 cm\(^{-1}\) down shift of the bands between 700 and 900 cm\(^{-1}\). In addition to these positional differences, there are relative intensity differences in the bands, including a large increase in the band at 815 cm\(^{-1}\). There is a switch in the ratio of intensities of two bands assigned to OH deformations at 1046 and 1059 cm\(^{-1}\), respectively, with the latter decreasing and the former increasing upon protonation. Further changes upon protonation include the loss of COO-
symmetric stretching bands between 1350 and 1450 cm\(^{-1}\). The main differences in the ROA spectra of Ser (Figure 3.4B) upon protonation includes a small reduction in the intensity of bands at 920 (+ve), 980 (-ve), 1050 (-ve), 1250 (-ve) and 1285 (-ve) cm\(^{-1}\). However, there are larger relative reductions in intensity of the bands at 1310 (+ve), 1290 (-ve) and 1415 (-ve) cm\(^{-1}\) with the band at 1351 (+ve) cm\(^{-1}\) showing a very large reduction in intensity. However, one band appears in the protonated form of Ser that is absent in the zwitterion which is at 1480 (+ve) cm\(^{-1}\). Previous studies have highlighted the importance that side chain hydrogen bonding has in the ROA spectra of serine\(^{[31]}\). In particular, this work showed the relationship of the ROA signal arising from the ν(O-H) and ν(C-O) vibrations and hydrogen bonding. Therefore, the decrease in the negative peak at ~1050 cm\(^{-1}\) upon protonation might be attributed to a decrease or disruption of short strong hydrogen bonding of the sidechain.

In the Raman spectra of Thr (Figure 3.4C), as with Ser, there is an extensive down shift of bands between 700 and 900 cm\(^{-1}\) upon protonation, as well as changes in intensity. These changes include an increase in the intensity of the band down shifted by 30 cm\(^{-1}\) from 780 cm\(^{-1}\) in the zwitterion and a change in the ratio of the bands at 850 and 875 cm\(^{-1}\), with the latter decreasing and the former increasing upon protonation. Other changes include the loss of COO- bands (1350 and 1410 cm\(^{-1}\)) and the appearance of the band assigned to C=O stretch (1737 cm\(^{-1}\)). The ROA spectra of Thr (Figure 3.4D) shows more differences upon protonation compared to that of Ser. However, it is important to remember that Thr also contains an additional chiral centre at the C\(_\beta\) atom. Therefore, Thr contains two probes of the sterochemical structure of the amino acid compared to Ser’s one. While there is an invariant quartet (1000 (+ve), 1042 (-ve), 1104 (+ve) and 1134 (-ve) cm\(^{-1}\) observed, other regions of the spectrum shows multiple differences. The observed changes include: a down shift of 30 cm\(^{-1}\) of a negative band at 780 cm\(^{-1}\), the loss of intensity of a triplet at 870 (-ve), 890 (+ve) 930 (+ve) cm\(^{-1}\), the appearance of a negative band at 1180 cm\(^{-1}\), extensive changes in spectral profile at 1124 (-ve), 1361 (+ve) 1418 (-ve), and 1455 (+ve), extensive reduction in intensity of the bands at 1361 (+ve) and 1455 (-ve) cm\(^{-1}\), and finally a possible new positive band may also emerge at 1305 cm\(^{-1}\). However, this feature at 1305 cm\(^{-1}\) is approaching the noise level so might not be reliable.

3.4.7. Sulphur containing amino acids

Upon protonation, the Raman spectrum of Cys (Figure 3.4E) shows a slight intensity increase and shift in the C-S stretch band at 688 cm\(^{-1}\), while between 750 and 1250 cm\(^{-1}\) the spectral details are similar, excluding small reductions in the intensity of the peaks in this region. However, the intense peaks observed for the zwitterion at 1313, 1354, 1404 and 1434 cm\(^{-1}\) show a large reduction in their intensity upon protonation. Finally, the C=O stretch is observed at a slightly higher position (1742 cm\(^{-1}\)) than observed for the other amino acids. There are few observable changes in the ROA spectrum of Cys (Figure 3.4F) upon protonation, with most of the features in the two spectra appearing to be the same, except for minor differences in absolute intensity. The main change observed upon protonation is a reduction in the intensity of the features at 1353 (+ve), 1405 (-ve) and 1433 (+ve) cm\(^{-1}\).
The Raman spectra of Met (Figure 3.4G) shows that the bands between 650 and 720 cm\(^{-1}\), relating to the sulphur atom\(^{1,28}\), vary with pH in their relative intensities. The C-S symmetric stretch mode at 656 cm\(^{-1}\) appears to show an increase. However, this observation also corresponds to other signatures including O-C=O deformation and COO\(^{-}\) scissoring\(^{28}\). Furthermore, the ratio between the bands at \(~700\) and \(~720\) cm\(^{-1}\) which correspond to C-S asymmetric stretch and C-S-C asymmetric stretch, respectively, changes upon protonation. Both spectra contain several overlapping peaks between 800 and 1000 cm\(^{-1}\), with a number of bands varying in intensity and position upon protonation, making assignment challenging. Other clear changes include the loss of the COO\(^{-}\)- stretching modes at 1350 and 1414 cm\(^{-1}\). The differences between the ROA spectra (Figure 3.4H) of Met upon protonation are quite complex, with a number of differences in position and intensity of features. There is a down shift of the spectral profile from \(-680\) (-ve) with the peaks at 730 (+ve), 752 (-ve) and 792 (+ve) shifted by 10–20 cm\(^{-1}\). In addition, the complex spectral response between 1300 and 1400 cm\(^{-1}\) shows variance upon protonation. The intensity of the peaks in the aforementioned shows both increases and decreases in the intensity at different points. Additionally, the ROA spectrum of the zwitterion shows two positive peaks at 1331 and 1364 cm\(^{-1}\), whereas only one is observed at 1340 (+ve) cm\(^{-1}\) for the fully protonated Met.

3.4.8. Acidic amino acids

Due to the lower solubility of the two acidic amino acids (Asp and Glu) at a higher pH, both have a poor signal to noise ratio in the ROA spectra compared to that reported here for the other amino acids. Moreover, Asp and Glu were measured at a lower concentrations in zwitterionic form compared to that of the fully protonated form due to the extreme difference in solubility. Therefore, assignments and comparisons might not be as reliable as found with the other amino acids. Moreover, at the measured pH there is still partial protonation of the OH group (15\%) for Glu.

The Raman spectrum of Asp (Figure 3.5A) shows a number of changes upon protonation, with a new band appearing at 750 cm\(^{-1}\), there is also a large intensity increase of the band at 830 cm\(^{-1}\). Furthermore, there is the loss of a band at 1560 cm\(^{-1}\), as well as a narrowing and intensity decrease of the band at 1416 cm\(^{-1}\), and the emergence of the C=O stretching band at 1740 cm\(^{-1}\). The main differences observed in the ROA spectra of Asp (Figure 3.5B) upon protonation include a down shift of 20 cm\(^{-1}\) of the couplet at 775 (-ve) / 815 (+ve) cm\(^{-1}\), an increase in the intensity of the negative band at 1230 cm\(^{-1}\) and a large decrease in the intensity of the band observed at 1400 cm\(^{-1}\). Unfortunately, due to the signal to noise problems, it is not possible to assign other differences between the two ROA spectra of Asp with confidence.

The Raman spectra of Glu (Figure 3.5C) show similarities to those of Asp in the changes upon protonation, with an additional band appearing at 746 cm\(^{-1}\). However, there are only smaller changes in intensity of other bands at 863 and 922 cm\(^{-1}\), which are associated with COOH deformation and CCN stretching respectively. Furthermore, there is a loss of the band at 1356 cm\(^{-1}\), and a change in position and shape of the bands between 1400 and 1500 cm\(^{-1}\). Finally, there is an increase in the C=O stretching mode at 1737 cm\(^{-1}\). The two ROA spectra of Glu (Figure 3.5D) show a similar profile; however, after intensity scaling of the spectra there are noticeable
differences between the region of 1250 and 1450 cm\(^{-1}\) upon protonation. These changes include an increased relative intensity of this region, as well as changes in absolute position with a downshift of by ~10 cm\(^{-1}\) of the band originating at 1360 cm\(^{-1}\) in the zwitterion. This is notable, as with all of the other amino acids, a decrease in the relative intensity of this region of the spectra was observed. Therefore, either Glu behaves differently from the other amino acids, or, this observation is due to the previously described signal to noise problem.

3.4.9. Imidazole containing amino acid

While His is classified as a basic amino acid, the imidazole ring plays a pivotal role in its chemical properties so that we classify it separately. Previous studies have investigated His\(^{15,17,32}\), including ones where QM calculations were performed on His in its zwitterionic state, and these assist in the discussion of the spectral changes observed. Due to the multiple protonation states that occur from H\(_4\)His\(^2+\) to His\(^2-\), the structures of which from previous strides are shown (Supplementary information Figure S3.1)\(^{33}\), the Raman and ROA spectra of His were measured for its zwitterion H\(_2\)His as well as its H\(_2\)His\(^+\) and H\(_2\)His\(^2+\) states. At each of the specified pHs, >95% of the His molecules populate the described protonation state, according to previously reported fractional distributions of His as a function of pH\(^{15}\). Interpretation of His spectra is complicated by two factors. Firstly, in the zwitterionic form, two tautomers exist where either the N\(_\tau\) or N\(_\pi\) of the imidazole ring is protonated. Secondly, six geometrical forms exist due to rotation around the two bonds C\(_\alpha\)-C\(_\beta\) and C\(_\beta\)-C\(_\gamma\)\(^{15}\) i.e. rotation around the first carbon of the side chain with respect to C\(_\alpha\) and rotation around the bond between the first C\(_\beta\) and second carbon atom of the His sidechain C\(_\gamma\).

There are extensive changes in the Raman and ROA spectra between H\(_2\)His and H\(_3\)His\(^+\) in part due to now both nitrogens (N\(_\tau\) and N\(_\pi\)) of the imidazole ring being protonated\(^{17}\). Therefore, where there were two independent chemical species present, there is now one. The spectral fingerprint of the imidazole ring in the zwitterionic form is spread over a number of bands, while protonation causes a redistribution of this profile into fewer bands at 1200, 1275 and 1500 cm\(^{-1}\) (Figure 3.5E). Other observations upon protonation from H\(_2\)His to H\(_3\)His\(^+\) include: the loss of a band at 670 cm\(^{-1}\), a downshift of 10 cm\(^{-1}\) of the band at 720 cm\(^{-1}\) and two smaller bands at 923 and 946 cm\(^{-1}\) being replaced with a more intense band at 930 cm\(^{-1}\). As alluded to, the region between 1100–1600 cm\(^{-1}\), which contains the fingerprint of the imidazole ring, sees many changes in the profile upon protonation. This includes the loss of the intensity, or complete loss of the bands at (1161, 1200, 1288, 1327, 1261 and 1575 cm\(^{-1}\)). While new bands, or a marked increase in the intensity, as observed in the following positions 1200, 1275, 1500 and 1636 cm\(^{-1}\). Differences between the ROA spectra of H\(_2\)His and H\(_3\)His\(^+\) (Figure 3.5F) are extensive; while the ROA spectrum of H\(_2\)His is dominated by the main bands at 1284 (-ve), 1350 (+ve), 1412 (-ve) and 1447 (+ve) cm\(^{-1}\), this profile is markedly different upon protonation, decreased intensity and changes in band positions to the following: 1244 (-ve), 1290 (-ve), 1351 (+ve), 1409 (-ve) and +1441 (+ve) cm\(^{-1}\). Other changes observed in the ROA spectrum of H\(_3\)His\(^+\) when compared to H\(_2\)His, include the loss of the positive band at 1075 cm\(^{-1}\) and the appearance of a positive band at 990 cm\(^{-1}\). While there were extensive
changes in the Raman and ROA spectra between H$_2$His and H$_3$His$^+$, there are fewer changes between H$_3$His$^+$ and H$_3$His$^{2+}$. In the Raman spectrum upon protonation to H$_3$His$^{2+}$ (Figure 3.5E), a band is gained at 780 cm$^{-1}$ and there is a small relative increase in the intensity of bands at 1272 and 1500 cm$^{-1}$. Furthermore, there is a loss of the bands at 1348 and 1415 cm$^{-1}$ and the gain of C=O stretching at 1741 cm$^{-1}$. The changes in the ROA spectra (Figure 3.5F) between H$_3$His$^+$ and H$_3$His$^{2+}$ show as a decrease in the intensity of bands, especially at a triplet at 1351 (+ve), 1410 (-ve) and 1441 (+ve) cm$^{-1}$ and the additional loss of the negative signed feature at 1660 cm$^{-1}$.

3.4.10. QM Calculations of Valine

In order to confirm and explain the variance observed in the spectral features of the amino acids upon protonation, Val was chosen as a target for QM as it was representative of the other amino acids in the changes observed. Changes which not only include band positional shifts and loss and addition of spectral features, but it also showed the previously discussed unusual and unexplained intense positive feature in the ROA. Furthermore, the side chain cannot be protonated; therefore, it would help to confirm or disprove whether the observations are through simple addition or subtraction of a proton, or are due to more widespread changes in the molecule.

The computed Raman spectra (Figure 3.6A and C) generated for Val show good agreement with their corresponding experimental spectra (Figure 3.6B and D). Specifically, upon protonation, there are a number of changes throughout the spectra observed, including shifts in position, a number of changes through 1300 to 1500 cm$^{-1}$ and the appearance of C=O stretching mode, although this is slightly shifted to the higher position of 1780 cm$^{-1}$ in the calculation.

In comparing the experimental and computational ROA spectra for protonated Val (Figure 3.7A), they show a good agreement, in both the sign and position of bands. However, the intensity does not always appear to be in close agreement with the experimental spectrum. A number of the bands between 600–1200 cm$^{-1}$ in the computational spectra show agreement in intensity and sign with that of the experimental spectrum. A noticeable difference between the two spectra includes the peaks observed at 1267 (-ve), 1308 (+ve), 1334 (-ve) and 1351 (+ve) cm$^{-1}$ in the experimental spectrum, where the calculated spectrum underestimates the positive bands at 1308 and 1364 cm$^{-1}$ and overestimates the positive contribution at 1460 cm$^{-1}$. At the higher end of the spectrum, an intense negative feature is predicted at 1620 cm$^{-1}$, which is absent in the experimental spectrum, while the weak negative feature in the experimental spectrum at 1655 cm$^{-1}$ is observed in the calculated spectrum albeit with greater intensity. Finally, a weak positive feature at 1730 cm$^{-1}$ in the experimental spectrum is observed at a slightly shifted position of 1745 cm$^{-1}$ in that of the calculated spectrum.

The calculated ROA spectrum of zwitterionic Val (Figure 3.7B) also shows a good level of agreement to its corresponding experimental spectrum. In the low wavenumber region a band observed in the experimental spectrum at ~660 cm$^{-1}$ appears to be down shifted by 40 cm$^{-1}$ in the calculated spectrum. The triplet at 900 (-ve), 954 (+ve), 989 (-ve) cm$^{-1}$ also appears in the calculated spectrum, while the ROA spectral feature at 1076 (-ve), 1124 (+ve), 1160 (-ve) and 1200 (+ve) cm$^{-1}$ is also down shifted in the calculated spectrum by ≈40 cm$^{-1}$. The very intense
couplet at 1340 (-ve) and 1360 (+ve) cm\(^{-1}\) in the experimental spectrum is present in the calculated spectrum, at the same wavenumber position; however, while in the experimental spectrum there was a positive bias in intensity of this couplet, in the calculated spectrum the reverse is true, with the negative component being more intense. Finally, the negative band at 1470 cm\(^{-1}\) shows a close match and the final couplet of 1580 (-ve), 1604 (+ve) appears slightly upshifted in the calculated spectrum by 15 cm\(^{-1}\). Both calculated spectra of Val in the two protonation states explored show a good resemblance to the corresponding experimental spectra; especially when considering that only single conformations of Val were used, and no explicit water modelling was included.

Upon comparing the 2D potential energy scans presented in Figure 3.8 for both the zwitterionic and protonated forms of Val, free rotation of COO\(^-\) / COOH around C\(_\alpha\) is restricted in both cases, with a barrier above 5 kcal mol\(^{-1}\). Conversely, the NH\(_3\) group would be able to rotate around its bond to C\(_\alpha\). Furthermore, as expected, differences are observed between the two energy landscapes, with a more restricted profile being found upon protonation to COOH.

For the two optimised structures of Val (Figure 3.9), the most obvious difference observed is the disruption of an intra hydrogen bond between the COO\(^-\) and NH\(_3\)\(^+\) groups that is no longer predicted when Val is fully protonated.

Extensive changes in the Raman spectra upon protonation for all of the amino acids were observed between the region of 1300–1500 cm\(^{-1}\). The calculations offer insight into the vibrational modes responsible, therefore, this region is now focused upon in more detail. Figures 3.10 and 3.11 present a pictorial representation of the contributions of the vibrational modes, in which the atom involved and intensity of the vibration is proportional to the size of the sphere. The two figures highlight a number of differences in the vibrational modes between zwitterionic Val and protonated, such as the mode at 1326 cm\(^{-1}\) in Figure 3.10 involving N-H and COO and a large N-H mode at 1400 cm\(^{-1}\). Furthermore, changes in vibrational modes upon protonation are not restricted to the N and C termini of Val. Changes are observed in the side chain, such as changes in intensity of C-H modes of Val corresponding to the Raman band observed at 1355 ± 2 cm\(^{-1}\). These figures elegantly illustrate that protonation acts to disrupt a number of vibrational modes throughout the molecule, and not simply those directly involved at the site of protonation.

3.5. Conclusions

Previous work which details the Raman signature of the common amino acids has only presented a partial listing, and measurements were not carried out in solution\(^{[1]}\). Whereas, other work where the amino acids were in solution provided Raman spectra of such poor signal to noise the extent of the discussion was limited to discussing accompanying measurements not in solution\(^{[28]}\). Furthermore, to date ROA studies of amino acids have usually only focussed on one amino acid\(^{[15, 25, 34]}\). Whereas, this study presents a detailed analysis of Raman and ROA spectra of the common amino acids in solution at both their isoelectric point and in their fully cationic forms. This provides detailed band assignments for all the common amino acids, for both Raman and ROA. Moreover, it has been shown that significant differences are observed between the amino acids in the Raman and ROA spectra, but also that extensive changes occur in the spectra upon
protonation. These changes observed in the Raman and ROA spectra of the common amino acids upon protonation consistently involve more than just the protonation site; extensive changes through the entire molecule are observed. These characteristic differences include changes in the side chain vibrations, even when they themselves are not proton donors or acceptors, and changes in the hydrogen bonding. The QM calculations of Val further demonstrated this, with possible intra-hydrogen bonding patterns becoming disrupted upon protonation. Moreover, further trends became evident, with more complex amino acids in some circumstances showing less variance upon protonation than their simpler counterparts. This indicates that as the distance between the end of the side chain increases with respect to Cα, the chiral sensitivity to the micro environment diminishes. Finally establishing reference spectra of amino acids in solution is critical, not only for band assignments; but, also to better differentiate changes observed between higher order structure observed in polypeptides and proteins, from those of the side chain under different protonation states.
3.6 References

19. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, Gaussian 09, Revision A02; Gaussian, Inc. Wallingford CT, 2009.


23. Matlab and Bioinformatics toolbox, R2010a; The MathWorks, Inc.: Natick, Massachusetts, United States., 2010.


Figure 3.1: Raman spectra of the amino acids Gly, Ala Val and Leu measured in their fully protonated state (black) and zwitterions (red) (A, B, D and F) and the corresponding ROA spectra for Ala Val and Leu for the same experimental conditions (C, E and G).
Figure 3.2: Raman spectra of amino acids Ile, Pro, Phe, Trp and Tyr measured in their fully protonated state (black) and zwitterions (red) (A,C,E, G and I) and the corresponding ROA spectra for Ile, Pro, Phe, Trp and Tyr under the same experimental conditions (B, D, F, H and J). Due to the low solubility of Tyr, it was only measured at one pH.
Figure 3.3: Raman spectra of amino acids Arg, Lys, Asn and Gln measured in their fully protonated state (black) and zwitterions (red) (A,C,E and G) and the corresponding ROA spectra for Arg, Lys, Asn and Gln under the same experimental conditions (B, D, F and H).
Figure 3.4: Raman spectra of amino acids Ser, Thr, Cys and Met measured in their fully protonated state (black) and zwitterions (red) (A,C,E and G) and the corresponding ROA spectra for Ser, Thr, Cys and Met measured under the same experimental conditions (B, D, F and H).
Figure 3.5: Raman spectra of amino acids Asp, Glu and His measured in their fully protonated state (black) and zwitterions (red) or green in the case of His (A,C and E) and the corresponding ROA spectra for Asp, Glu and His under the same experimental conditions (B, D, and F).
Figure 3. 6: Calculated (Calc) Raman spectra of Val in both its zwitterionic and fully protonated forms (A and C) and the corresponding experimental (Exp) spectra (B and D)
Figure 3. 7: Experimental (Exp) and calculated (Calc) ROA spectra of Val, in both its protonated forms (A) and its zwitterionic forms (B).
Figure 3.8: 2D potential energy surface plot of Phi Psi rotation of Val in its zwitterionic form (A) and protonated (B) forms. The colour scale represents energy (kcal/mol).
Figure 3.9: Molecular representation of Val in its zwitterionic form (I) and fully protonated state (II), highlighting the changes in intramolecular hydrogen bonding.
Figure 3. 10: Pictorial representation of vibrational modes of Val in its zwitterionic form, for bands between 1300 and 1500 cm\(^{-1}\). For a particular vibrational mode the vibrational intensity is proportional to the sphere volume, and the direction of the atomic displacement is perpendicular to the junction between the two hemispheres (blue or yellow).
Figure 3.11: Pictorial representation of vibrational modes of Val in its protonated form, for bands between 1300 and 1500 cm\(^{-1}\). For a particular vibrational mode the vibrational intensity is proportional to the sphere volume, and the direction of the atomic displacement is perpendicular to the junction between the two hemispheres (blue or yellow).
Table 3. 1: A summary of Raman bands and positions for 20 Amino acids in a low pH environment and in their zwitterionic form. Band intensity described in lowercase: weak (wk), medium(m), strong (s) and very strong (vs) while band widths are described in uppercase: Sharp (Sh), medium (M) and broad (Br). If a band appears as a shoulder it is designated Sho.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Reference bands cm⁻¹ (Low pH)</th>
<th>Reference bands cm⁻¹ (Zwitterion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Alanine (Ala)</td>
<td>533 mM, 620 sBr, 751 sM, 825 vsM, 842 Sho, 924 sSh, 1021 sM, 1123 sM, 1152 Sho, 1223 wkM, 1264 mM, 1308 wSh, 1333 mSh, 1361 mM, 1418 MS, 1465 sM, 1531 wkM, 1609 wkBr, 1732 sBr, 1754 Sho</td>
<td>534 mM, 647 wkBr, 780 mSh, 851 vsSh, 924 mM, 1007 sM, 1117 sM, 1114 Sho, 1220 wSh, 1305 sSh, 1356 vsSh, 1381 mSh, 1418 vsSh, 1465 vsSh, 1381 mSh, 1418 vsSh, 1465 vsSh, 1523 SM, 1602 mBr</td>
</tr>
<tr>
<td>L-Arginine (Arg)</td>
<td>531 sM, 609 MB, 650 Sho, 740 mM, 775 Sho, 817 MB, 832 Sho, 857 mM, 885 SM, 908 wkM, 938 sM, 982 mM, 1018 mSh, 1062 vsM, 1097 VIm, 1179 sM, 1221 wSh, 1273 wkM, 1301 Sho, 1326 vsM, 1351 Sho, 1374 sM, 1394 Sho, 1419 Sho, 1449 vsM, 1473 wSh, 1590 mM, 1633 LB, 1664 mM, 1737 vsM.</td>
<td>531 mBr, 555 Sho, 609 mM, 670 wkM, 760 wkM, 783 Sho, 837 Sho, 866 mM, 905 mM, 938 sM, 1029 wkM, 1059 Sho, 1089 VIB, 1142 Sho, 1182 sM, 1218 mSh, 1290 vsM, 1321 vsM, 1367 vsM, 1387 Sho, 1416 vsM, 1449 vsM, 1480 Sho, 1567 mM, 1593 MB, 1673 sBr.</td>
</tr>
<tr>
<td>L-Asparagine (Asn)</td>
<td>524 wkM, 602 mM, 678 wkM, 745 mSh, 788 Sho, 811 vsM, 885 mM, 924 sM, 998 mM, 1080 Sho, 1128 sBr, 1168 Sho, 1233 sM, 1274 mM, 1310 wkM, 1340 mM, 1373 Sho, 1426 Sho, 1437 vsBr, 1530 wkM, 1623 vsBr, 1688 sM, 1742 VIIm.</td>
<td>527 wkM, 611 wkM, 696 wkBr, 771 mSh, 805 mM, 831 vsM, 870 sM, 887 mM, 932 sM, 993 mM, 1079 sM, 1130 sBr, 1157 Sho, 1233 mM, 1269 wkM, 1330 vsM, 1360 vsM, 1408 vsM, 1415 vsM, 1530 wSh, 1618 mBr, 1685 sM.</td>
</tr>
<tr>
<td>L-Aspartic Acid (Asp)</td>
<td>575 wSh, 617 mM, 681 wkM, 754 sSh, 794 mSh, 828 vsM, 865 wSh, 896 mSh, 932 vsM, 1001 sM, 1085 sM, 1122 nM, 1162 mBr, 1233 sM, 1261 mBr, 1333 wSh, 1370 mM, 1415 vsM, 1528 wkM, 1618 mBr, 1738 vsBr.</td>
<td>530 mM, 575 wSh, 623 mBr, 693 wkM, 783 mSh, 842 vsM, 868 wSh, 904 wSh, 937 vsM, 995 mM, 1079 sM, 1144 mM, 1151 mM, 1230 sM, 1256 mM, 1307 wSh, 1358 vsM, 1415 vsM, 1601 mM, 1610 sSh, 1616 sM, 1729 mSh.</td>
</tr>
<tr>
<td>L-Cysteine (Cys)</td>
<td>530 mBr, 617 wkM, 684 vsM, 774 vsBr, 808 Sho, 862 vsM, 876 Sho, 940 sM, 1001 mM, 1075 mM, 1120 Sho, 1152 mBr, 1220 sM, 1272 wkM, 1313 Sho, 1333 mSh, 1353 sM, 1373 Sho, 1406 Sho, 1431 vsM, 1526 wkM, 1620 wkBr, 1741 LM.</td>
<td>533 mBr, 626 mM, 687 vsM, 780 sM, 820 mM, 879 sM, 940 sM, 998 mM, 1069 mM, 1115 Sho, 1149 mBr, 1218 mBr, 1262 Sho, 1274 wkM, 1313 sM, 1353 vsSh, 1403 vsM, 1433 vsSh, 1515 wkBr, 1604 wkBr.</td>
</tr>
<tr>
<td>L-Glutamic Acid (Glu)</td>
<td>503 mM, 539 mM, 590 wkM, 643 mBr, 675 Sh 754 sM, 803 sM, 842 Sho, 859 vsM, 918 vsM, 979 mSh, 1006 mM, 1052 mSh, 1072 Sho, 1090 sM, 1162 mM, 1201 wSh, 1256 wSh, 1310 mM, 1350 sBr, 1425 vsM, 1447 vsM, 1532 wkM, 1611 wSh, 1733 sBr.</td>
<td>539 wSh, 611 wkM, 649 mBr, 780 mBr, 820 mM, 862 vsM, 921 vsM, 1001 mM, 1044 mM, 1085 sM, 1157 mM, 1199 wSh, 1294 mM, 1330 sM, 1355 vsM, 1418 vsM, 1452 sM, 1620 sBr.</td>
</tr>
<tr>
<td>L-Glutamine (Gln)</td>
<td>537 wkM, 638 wkBr, 746 mM, 792 mM, 846 vsM, 880 wSh, 908 vsM, 947 Sho 974 Sho 1007 mM, 1045 mSh, 1088 mM, 1123 sBr, 1163 Sho, 1207 mSh, 1303 Sho 1257 Sho 1341 sBr, 1431 vsBr, 1451 Sho, 1533 wSh, 1619 vsBr, 1679 sM, 1739 sM.</td>
<td>537 mM, 655 wkBr, 772 sM, 812 wSh, 857 sM, 910 vsM, 944 Sho 999 mM, 1040 mSh, 1083 mM, 1126 sBr, 1155 sM, 1205 mM, 1293 mSh, 1331 Sho, 1353 mSh, 1369 Sho, 1414 Sho, 1426 vsBr, 1448 vsM, 1536 wSh, 1612 sBr, 1677 mSh.</td>
</tr>
<tr>
<td>Amino Acid</td>
<td>mM</td>
<td>sM</td>
</tr>
<tr>
<td>------------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>L-Glycine (Gly)</td>
<td>506</td>
<td>575</td>
</tr>
<tr>
<td>L-Histidine (His)</td>
<td>525</td>
<td>635</td>
</tr>
<tr>
<td>L-Isoleucine (Ile)</td>
<td>536</td>
<td>606</td>
</tr>
<tr>
<td>L-Leucine (Leu)</td>
<td>542</td>
<td>605</td>
</tr>
<tr>
<td>L-Lysine (Lys)</td>
<td>504</td>
<td>549</td>
</tr>
<tr>
<td>L-Methionine (Met)</td>
<td>539</td>
<td>655</td>
</tr>
<tr>
<td>L-Phenylalanine (Phe)</td>
<td>623</td>
<td>734</td>
</tr>
<tr>
<td>L-Proline (Pro)</td>
<td>534</td>
<td>576</td>
</tr>
<tr>
<td>L-Serine (Ser)</td>
<td>614</td>
<td>739</td>
</tr>
</tbody>
</table>
### L-Threonine (Thr)

<table>
<thead>
<tr>
<th>pH</th>
<th>mM</th>
<th>wkM</th>
<th>vsM</th>
<th>wSh</th>
<th>mSh</th>
<th>sBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1358</td>
<td>1415</td>
<td>1474</td>
<td>1535</td>
<td>1625</td>
<td>1355</td>
<td>1413</td>
<td>1472</td>
<td>1528</td>
<td>1609</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### L-Tryptophan (Trp)

<table>
<thead>
<tr>
<th>pH</th>
<th>mM</th>
<th>wkM</th>
<th>vsM</th>
<th>wSh</th>
<th>mSh</th>
<th>sBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
</tr>
</thead>
<tbody>
<tr>
<td>546</td>
<td>581</td>
<td>708</td>
<td>762</td>
<td>788</td>
<td>542</td>
<td>560</td>
<td>625</td>
<td>713</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### L-Tyrosine (Tyr)

<table>
<thead>
<tr>
<th>pH</th>
<th>mM</th>
<th>wkM</th>
<th>vsM</th>
<th>wSh</th>
<th>mSh</th>
<th>sBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
</tr>
</thead>
<tbody>
<tr>
<td>530</td>
<td>560</td>
<td>646</td>
<td>716</td>
<td>734</td>
<td>545</td>
<td>565</td>
<td>627</td>
<td>710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### L-Valine (Val)

<table>
<thead>
<tr>
<th>pH</th>
<th>mM</th>
<th>wkM</th>
<th>vsM</th>
<th>wSh</th>
<th>mSh</th>
<th>sBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
<th>mBr</th>
<th>vsM</th>
<th>wkM</th>
</tr>
</thead>
<tbody>
<tr>
<td>545</td>
<td>634</td>
<td>734</td>
<td>754</td>
<td>820</td>
<td>545</td>
<td>593</td>
<td>649</td>
<td>831</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.7. Supplementary information

**Figure S3. 1** Suggested structures of His under different protonation states and tautomers[33], pKa values taken from[35]
Chapter 4: A Matlab toolbox to aid the analysis of Raman and Raman Optical Activity (ROA) Spectra
4.0. Declarations

As the first author of this planned publication, I carried out all the design, programming and implementation of the work, as well as analysis of the test spectra. Lorna Ashton and Ewan Blanch provided feedback of the manuscript.
A Matlab toolbox to aid the analysis of Raman and Raman Optical Activity (ROA) Spectra for 2D correlation

Ben Gardner, Lorna Ashton and Ewan W. Blanch

Faculty of Life Sciences, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.
E-mail: E.Blanch@manchester.ac.uk; Fax: +44 (0)161 236 0409; Tel: +44 (0)161 306 5819

School of Chemistry, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, UK, M1 7DN.

4.1. Abstract

Presented here is an easy to use graphical user interface (GUI) toolbox, designed for the purpose of pre-processing of Raman and Raman optical activity (ROA) spectra. This toolbox allows users to perform a number of spectral data treatment prerequisites for post acquisition analysis, such as generalised 2D correlation which is also a function of the toolbox. Whilst there are a number of available packages for the analysis of Raman spectra, these do not always provide a good range of data-processing techniques and often require specialist knowledge to use effectively. Furthermore, no user friendly toolboxes are available for 2D correlation analysis of ROA data, which urgently needed addressing. An advantage of designing a toolbox for Matlab is that it allows users to easily modify it to their specific needs, and to continually add and expand the functionality of the toolbox; it can also be ran as an executable program therefore providing access for those without Matlab.

4.2. Introduction

With the continual improvement in instrument development, providing high quality Raman and Raman optical activity (ROA) spectra; furthered by improvements in baseline approaches and increased computational power, previous limitations of the field are increasingly mitigated. These advancements easily allow large scale handling and processing of spectra (100,000+) for use in common post acquisition analysis methods such as multivariate analysis[1] and 2D correlation[2, 3]. Generalised two dimensional correlation (2D-Cos) is a powerful analytical technique developed by Noda[4], where cross-correlation is applied to spectra. This involves the cross-correlation of spectra as a function of two independent wavenumbers, thereby spreading the data over an additional dimension. The generalised method of 2D-Cos that Noda introduced is applicable to probing dynamic changes in spectra by any external perturbation, thus providing a powerful analysis tool for an extensive set of experiments[2, 5-9]. 2D-Cos has proven to be a very popular method of analysis and has been used in a number of studies[2, 5-10]. Furthermore, it has led to a number of variant analytical methods including: moving windows (MW), perturbation correlation moving
windows (PCM), double 2D-Cos and moving window principal component analysis to name a few. While there is an ever increasing demand for 2D-Cos and its variants, 2D-Shige is the only freely available software for performing this type of analysis, with a commercially available alternative being SpectraCarr 2DCOS (Thermo scientific). Due to the lack of other software alternatives (such as other freely available toolboxes) the result of this limitation is that most users calculate 2D-Cos on in house developed software, usually in the Matlab environment. While this provides a user with additional flexibility, not all users have the necessary computer skills to do this. Furthermore, in house developed software needs to be thoroughly examined with test data to ensure that common mistakes are avoided including the correct results are given under all circumstances.

The availability of sophisticated analysis methods such as 2D-Cos is especially important in the realm of biological studies, as it allows for the development of meaningful complex experiments and an interpretation going past a vague qualitative account of the observed changes in spectra. However, in order for advanced post acquisition analysis methods to generate reliable results there is always the prerequisite of data treatment, which comes in many forms including: baseline correction procedure, solvent subtraction, smoothing, data normalisation or scaling, and interpolation. All of these steps are essential as they address and correct for unavoidable problems such as varying backgrounds, noise, variances in intensity between data sets, and redistribution of data so that it is evenly spaced. However, to date there is no consensus on a “correct” approach, and likely will never be due to the inherent subjectiveness and differing requirements a user has in their data pre-treatment. The result of which is a minefield of approaches for spectral pre-treatment, which can be daunting for the new and casual spectroscopist. In addition there is a steep learning curve for those used to a graphical user interface (GUI) and not experienced in using a command line environment provided by popular software such as Matlab, compounded by a time consuming search to identify a method that suits their needs. It is for the reasons discussed above that an easy to use GUI interface for Matlab has been developed and is demonstrated here. This toolbox not only provides means to carry out commonly utilised approaches for processing of both Raman and Raman optical activity (ROA) spectra but, additionally can be used for 2D analysis all in one easy to use package. When new chemometric approaches are implemented, they are usually only applied to a spectral set with only positive values, and it is from these general rules of analysis are developed. However, previously it was shown with the case of generalised 2D-correlataion when applied to bisignate data these general rules often need to be re-evaluated. Due to this it becomes critically important to apply new methods of analysis such as perturbation correlation moving windows (PCMW) to model data sets of bisignate data. This is essential to understand how the developed rules might need to be amended for application to bisignate data such as that obtained through ROA or vibrational circular dichroism (VCD). Therefore, potential pitfalls in the application of certain 2D correlation methods, specifically (PCM), have been investigated when applied to bisignate data.
4.3. Two dimensional correlation theory

The methodologies used in the toolbox are discussed below (Generalised 2D-Cos, MW and PCMW). For a spectral set which is perturbed by an external change (e.g. time, pH concentration, temperature etc) in which the intensity of the set is expressed as \( y(v, t) \) where the spectral variable is \( v \) and \( t \) the external perturbation. Due to practicality only a finite range of \( t \) can be measured and a dynamic spectrum \( \bar{y}(v, t) \) is defined as:

\[
\bar{y}(v, t) = y(v, t) - \bar{y}(v) \text{ for } T_{\text{min}} \leq t \leq T_{\text{max}},
\]  

(4.1)

where \( \bar{y}(v) \) is the reference spectra for the system, and the limits of the perturbation variable lay between \( T_{\text{min}} \) and \( T_{\text{max}} \). A commonly chosen reference spectrum is the average of the spectral set, however a number of options are available including choosing the first or last spectrum, or none at all \[^{17} \). Generalised 2D-Cos is defined as:

\[
\Phi(v1, v2) + i\Psi(v1, v2) = \frac{1}{\Pi(T_{\text{max}} - T_{\text{min}})} \int_0^{\infty} \bar{Y}_1(\omega), \bar{Y}_2(\omega) d\omega,
\]

(4.2)

where the synchronous spectrum, or spectrum of changes is represented by \( \Phi(v1, v2) \) while the asynchronous spectrum or spectrum of differences \( \Psi(v1, v2) \) relating to the perturbation induced spectral variations at two points \( (v1, v2) \) between the limits of \( T_{\text{min}} \) and \( T_{\text{max}} \). As shown in equation (4.2), the expression \( (\bar{Y}_1(\omega) \) refers to a Fourier transformation of the spectral intensity variations of \( \bar{y}(v, t) \) observed at \( v1 \) and can be expressed as:

\[
\bar{Y}_1(\omega) = \int_{-\infty}^{\infty} \bar{y}(v_1, t)e^{-i\omega t} dt
\]

\[
= \bar{Y}_1^{\text{Re}}(\omega) + i\bar{Y}_1^{\text{Im}}(\omega).
\]

(4.3)

The expression for \( \bar{Y}_2(\omega) \) is described in a similar manner, but an inverse Fourier transformation is utilised instead. Due to the time consuming and cumbersome nature of this approach, the discrete Hilbert transform was adapted for use in generalised 2D correlation \[^{18, 19} \], with the synchronous and asynchronous correlations being calculated as follows:

\[
\Phi(v1, v2) = \frac{1}{m - 1} \sum_{j=1}^{m} \bar{y}_j(v1).\bar{y}_j(v2),
\]

(4.4)

where \( m \) is the number of spectra that are equally spaced along the perturbation variable \( t \), and

\[
\Psi(v1, v2) = \frac{1}{m - 1} \sum_{j=1}^{m} \bar{y}_j(v1) \sum_{k=1}^{m} N_{jk}.\bar{y}_j(v2),
\]

(4.5)

where \( N_{jk} \) are the elements of the Hilbert-Noda matrix, a modified Hilbert matrix where \( j \) and \( k \) correspond to row and column number, respectively.
Moving windows is a development of 2D-Cos where the data are divided into smaller data sets, allowing the location of important transitions to be determined; therefore, this method directly links the 2D-Cos to the perturbation\[^4\]. In this process the data is subdivided into “windows”, the size of which are dependent on the overall spectral sets size e.g. if a window size of three is chosen the 1\(^{st}\) window would contain spectra one to three, the 2\(^{nd}\) would contain spectra two to four and so on, until all spectra have been included. For each sub-matrix generalised 2D-Cos is utilised to calculate the autocorrelation intensities, which are also known as the slice spectrum or power spectrum, equivalent to the variance induced by the perturbation. In this incremented sliding window, which ranges between \( j = 1 + m \) to \( N - m \) equations (4.7–9) are calculated for each window\[^4\] where: \( j \) is index number of the window, \( J \) the spectrum in the window, \( P \) the perturbation variable, \( \Omega \) the 2D-Cos autocorrelation intensities and \( 2m + 1 \) is the window size.

\[
\begin{align*}
N_{jk} &= \begin{cases} 
0, & \text{for } j = k \\
\frac{1}{\Pi(k-j)}, & \text{Otherwise.}
\end{cases}
\end{align*}
\]  

\( (4.6) \)

\[
\begin{align*}
\bar{y}_j(v) &= \frac{1}{2m+1} \sum_{j=j-m}^{j+m} y_j(v, p_j) \\
\bar{y}_j(v, p_j) &= y_j(v, p_j) - \bar{y}_j(v) \\
\Omega_{aj}(v, p_j) &= \frac{1}{2m} \sum_{j=-m}^{j+m} \bar{y}_j^2(v, p_j)
\end{align*}
\]  

\( (4.7), (4.8), (4.9) \)

While now spectral variance is directly linked to the perturbation, only the intensity of the variance is displayed, and not its directional property i.e. if a peak is increasing or decreasing. Due to this the method of perturbation correlation moving windows (PCMW) was developed\[^4\]. PCMW is based upon the correlation between changes observed in spectral intensity and the corresponding changes in the perturbation\[^4\]. In PCMW the synchronous plot correlates spectral intensity change with perturbation direction, i.e. positive changes correlate to increments and negative changes with decreases in spectral intensity; whereas, the asynchronous plot provides detail on the rate of change. In PCMW the reference (or average spectrum) and dynamic perturbation in the \( j^{th} \) window can be expressed as shown in the following in equations (4.10–11).

\[
\begin{align*}
\bar{P}_j(v) &= \frac{1}{2m+1} \sum_{j=j-m}^{j+m} P_j \\
\bar{P}_j &= P_j - \bar{P}_j
\end{align*}
\]  

\( (4.10), (4.11) \)

From this the synchronous \( \Pi \Phi \) (Equation 4.12) and \( \Pi \Psi \) asynchronous (Equation 4.13) are the following:
\[ \Pi \Phi, j(v, p_j) = \frac{1}{2m} \sum_{j=j-m}^{j+m} \bar{y}(v, p_j) \cdot \bar{p}_j, \]  

\[ \Pi \Psi(v1, v2) = \frac{1}{2m} \sum_{j=j-m}^{j+m} \bar{y}(v, p_j) \cdot \sum_{k=k-m}^{k+m} N_{jk} \cdot \bar{p}_k, \]

It is these values which are used (\( \Pi \Phi \) and \( \Pi \Psi \)) instead of the perturbation change correlated against the variance of spectral intensity at \( v_c \) as is the case in moving windows\(^{[14]}\).

4.4. Toolbox Design and Demonstration

The toolbox was implemented in Matlab 2010, primarily using GUIDE Matlab's inbuilt function for designing a GUI. A number of functions incorporated into the toolbox are independently available through GNU license, and their use is cited as such in the user guide (4.7. Appendix); or are modified and adapted from previously described work\(^{[4, 13, 14, 20-23]}\). The toolbox can be executed from within Matlab, or can be sent as an executable program. When the toolbox is distributed as an executable the user is required to install Matlab Compiler Runtime (MCR), which is royalties free, thus providing access for those without Matlab. However, some functions will differ, and usage is more restricted without a full Matlab environment available. The layout of the main GUI (Figure 4.1A) is laid out so similar functions are grouped together into modules aiming to be as intuitive for the user as possible, with a streamlined user input for simplicity. Some of the more important functions, i.e. baseline subtraction procedures, allow for full user input and can be run independently to the main GUI (Figure 4.1A and B). The toolbox is designed to be flexible, allowing a number of file types to be directly imported (.PLT and .SPC), as well as automatically uploading data already in the Matlab workspace. Additionally, data can be directly exported to an excel spreadsheet or saved as a Matlab variable. A range of plotting tools are provided, allowing a user to quickly view the data from within the GUI, and plotting results from 2D analysis. Graphs can be saved in a number of formats as well as Matlabs .fig, allowing the user to edit them to their liking.

A range of functions are provided in the data processing module, which allows for all common types of spectral processing which are a prerequisite for post-acquisition analysis, including: Baseline procedures, which for Raman spectra utilises the baseline method using asymmetric least squares developed by Eilers et al.\(^{[20]}\), while a baseline for ROA data is created using a heavy median smoothing filter; smoothing, the smoothing option uses the well established Savitzky-Golay smoothing filter; normalisation and scaling, the two normalisation routines provided are standard normal variate (SNV) and a method for standardising the intensity of a single band, data can also be scaled [0, 1] to remove artificially induced negative values; interpolation, this routine fits a surface and interpolates the required points from this. These routines have been selected over others available on three main criteria, their effectiveness at the problem they address, the computational cost and ease of implementing them in a user friendly fashion for the toolbox. A full description of the methods for these procedures is provided in the user guide. The 2D correlation
module allows for routine 2D approaches to be calculated including: Variance plot, Synchronous and asynchronous plot, MW and PCMW.

A demonstration of the 2D functionality is shown (Figure 4.2) where 30 spectra were generated, each containing three Gaussian peaks (Figure 4.2A) with varying intensities along an arbitrary perturbation axis. In this data set peaks one and two have a mirrored simultaneous response with peak one growing while peak two decreases; while peak three is invariant for the first 10 arbitrary perturbation steps before decreasing in intensity. The variance plot is shown (Figure 4.2B) and is also known as the auto-intensities or the diagonal of the synchronous plot where \( \nu_1 = \nu_2 \). As expected the observed intensity variance for peaks one and two are identical, while peak three is roughly a third of the intensity of the first two peaks. The synchronous plot (Figure 4.2C) shows three auto peaks along the diagonal and three cross peaks, while the asynchronous plot (Figure 4.2D) shows two peaks one negative and one positive. In accordance to the rules established by Noda for the interpretation of the signs of synchronous and asynchronous plots\(^4\) this shows peaks at two and three are changing in the opposite direction to that of peak one, and peak two and three are changing in the same direction. While the asynchronous plot indicates that peaks one and two are changing before that of peak three, and changing simultaneously with respect to one another. Furthermore, both the moving windows (Figure 4.2E) and perturbation correlation moving windows plot (Figure 4.2F) correctly display the results; with the former showing the changes relating to the perturbation and the latter correlating both the nature of the change and duration.

While the rules of interpretation established by Noda hold true for the analysis of 2D-Cos for data of only one sign e.g. positive, they required expanding when applied to data containing features of two signs i.e. bisignate data e.g. the positive and negative signals that are observed in ROA. A modified set of rules for the interpretation of 2D-Cos performed on bisignate data have been established through the work of Aston et al\(^16\). Specifically, in the interpretation of the synchronous plot (the asynchronous plot is not affected) if there is a cross peak where the two bands have opposite signs, the rules need to be reversed e.g. a negative peak now indicates a change in the same direction with regard to absolute intensity, while a positive cross peak indicates they are changing in the opposite direction.

In PCMW both the synchronous and asynchronous data are plotted in the same plane, therefore it is important to understand the effect bisignate spectra might have on this form of analysis. To thoroughly test the response of PCMW with regards to bisignate data, four sets of bisignate Lorentzian peaks were produced (Figure 4.3A–D). Each data set contains 10 spectra exhibiting changes in intensity along an arbitrary perturbation series. These spectral sets include a number of examples of peak response, such as: peaks one and two both increasing in absolute intensity e.g. peak one becomes more positive while peak two becomes more negative, peak one decreases as peak two decreases in absolute intensity, peak one decreases in intensity while peak two increases and, finally, peak one decreases while peak two is initially invariant before decreasing.
The central intensities of the peaks are plotted against the arbitrary perturbation step (Figure 4.4), to better display the relationship between perturbation series and absolute intensity. Looking at the first PCMW plots (Figure 4.5A) in this scenario both peaks are increasing in intensity, and for the +ve peak, this corresponds to a positive plot, however the -ve peak which is increasing in intensity i.e. becoming more negative shows a negative response on the PCMW plot. The reverse of this scenario is shown in Figure 4.5B, where both peaks are decreasing in intensity. Peak one which has a +ve sign is shown as decreasing while the -ve peak which is decreasing towards zero, is shown as increasing in the PCMW plot. In the third scenario (Figure 4.5C), peak one, which is +ve, is decreasing at the same rate that peak two is increasingly becoming negative, and they show the same result, with a very intense decrease observed. Finally, in Figure 4.5D where peak one, with a +ve sign, is observed to decrease in intensity before peak two with a -ve intensity that is decreasing, is shown as decreasing for peak one and increasing for peak two.

From these results it is clear that caution is required in the interpretation of bisignate data from PCMW, especially with respect to perspective of the change. If a peak is changing in the direction of $-\infty$, which could be from either a +ve peak losing intensity or a -ve peak increasing in intensity (so becoming more negative) it shows as a negative change in PCMW. While the reciprocal is true for peaks where intensity is changing towards $+\infty$, in this scenario the change is seen as an increase. Thus, when interpreting bisignate data using PCMW care has to be taken to relate it to the initial spectra in describing the observed changes. Another option though is to attribute changes purely in a directional sense regarding $+\infty$ or $-\infty$, therefore negating possible confusion with +ve and -ve peak variance with respect to the zero intensity line.

4.5. Conclusions

The Matlab GUI toolbox presented here provides spectroscopists with a range of methods to process spectra, for the use of post-acquisition analysis like 2D-Cos. Due to its open design in Matlab it can easily be modified by users for their more specific requirements. Additionally, through continual development and feedback more functionality can be added including a broader range of pre-processing techniques, expanding the 2D-Cos capabilities and providing a module for PCA analysis. Furthermore, the demonstration of the 2D-Cos functionality highlights that caution has to be taken in the interpretation of bisignate data not only with general 2D-cos, but its derivatives such as PCMW.
4.6. References


22. Smirnova, D.S., Doctor of Philosophy, Applications of two dimensional correlation analysis of explosives detection and polymer crystallization, 2009, California Institute of Technology, California
23. Pazderska, T. and V. Kopecký, 2D Correlation spectroscopy and its application in vibrational spectroscopy using Matlab, Charles University, Prague
Figure 4.1: Screen shots of the main GUI (A) and two other sub GUIs which can be ran independently to the main program for Raman baselines (B) and processing ROA spectra (C).
Figure 4. 2: A demonstration of the 2D correlation module of the toolbox. A matrix containing 30 spectra (A), each consisting of three Gaussian peaks: peak 1 increasing simultaneously as peak 2 decreases, while peak 3 is invariant for 10 arbitrary perturbation steps before decreasing. The 2D correlation examples include a variance plot (B), the synchronous and asynchronous plots (C and D) a 2D-moving windows plot (E) and 2D perturbation correlation moving windows plot (F). Red indicates positive and large intensity, while blue negative or small intensity change.
Figure 4.3: Four sets of bisignate data (A-D), each of which contains 10 sets of spectra which are undergoing different changes of intensity along arbitrary perturbations. Peaks one and two are progressively increasing in intensity (A), conversely the two peaks are both decreasing in intensity (B), while peak one decreases as peak two increases (C), and (D) where peak one decreases in each spectrum, while peak 2 is invariant for the first 5 arbitrary perturbation steps before decreasing in intensity.
Figure 4. 4: Plots (A-D) display the central intensity of band one (blue) and band two (green) with respect to the arbitrary perturbation range for the four sets of generated data.
Figure 4.5: PCMW plots of the four spectral sets (A-D).
4.7. Appendix

Document 1 - Toolbox User Guide

**Installation**

Unzip the Raman toolbox files into a folder of choice and add the folder + sub folder to your Matlab path. To do this go (file- setpath-add with subfolders), select the folder you extracted the toolbox to, then save (you might need to launch Matlab as administrator).

**Useful Matlab tips and using the toolbox**

- Transpose Matrix A e.g. rows to columns and vice versa $B = A'$;
- Combine two Matrices $A$ and $B$. $C = [A, B]$;
- Subtract 1 matrix from another e.g. a water spectra from a matrix containing multiple spectra. $C = \text{bsxfun}(@\text{minus},a,b)$; where $C$ is the output, $a$ is the matrix containing your spectra, and $b$ is the spectra you want to subtract e.g. water.

Data should be orientated in columns, with the X-axis containing one column of data, while spectra should consist of columns of data.

Data is listed alphabetically in the main listbox, and is also selected in this order, so when multiple variables are selected ensure they are ordered as specified by the function.

**Using the Toolbox**

To launch the toolbox (Figure S4.1) type "workspace_v2" in the Matlab command window. Upon launching any variables currently in the main Matlab workspace will be uploaded into the toolbox. In discussing the functions the corresponding line number in the code is given next to each function.

![Figure S4.1: The layout of main GUI workspace.](image)
Work Space

Upload Workspace (137) - While the toolbox automatically uploads variables present in the workspace when launching, if any data has been manually added into the main Matlab workspace post launching, or any variables you have personally edited (such as transposing a variable, or renaming it) this uploads them into the toolbox.

Import SPCs (144) - SPC files can be uploaded via the "import SPC files" button. This function makes use of the GSTools toolbox[1], and the function GSImportspec. Copyright (c) 2004-2009, Kris De Gussem All rights reserved. This is used under the terms under the GNU license, please see GSTools folder for further information.

The import of SPC files is adapted for importing SPCs generated from a biotools Chiral Raman instrument tm, therefore correctly importing from other sources might require altering the code (line 144 in the main file) specifically lines addressing the imported structure array.

Figure S4. 2: SPC spectra import option box.

The user must choose whether you are importing from a folder containing just Raman spectra or ROA files or both (Figure S4.2). After making this choice and pressing OK navigate to a folder containing SPC files, and select the first file, and all files in the containing folder will be uploaded. The imported variables include a structure array off all the imported data, as well as the variables Raman, ROA and X_axis. If there is an error accessing the imported structure array the user will receive a notification and have to manually extract their data from it.

Import PLT (894) - PLT files can also be imported using the import plt button, again this is adapted for use with the aforementioned instrument, so might need adjustment for other users

Save Workspace (205) - Saves all data in the toolbox in the Matlab file.

Export to Excel (840) - Exports all data into a spreadsheet (.xlsx) with each variable in the workspace becoming its own sheet, with each sheet named according to the variable name in the toolbox.

Delete Variable (180) - Deletes selected variable from the toolbox and the Matlab workspace.

Plotting

Plot Figure (226) - The plot figure button requires selection of two variables, e.g. your data and its corresponding X-axis. (Plotting requires the data to appear in the workspace in the following order e.g. Raman spectra, X-Axis. If your x-axis precedes the data you're trying to plot it will be plotted incorrectly).
Plot Mesh (271)- This requires the user to select 3 items, the appropriate data e.g. PCMW, the X-axis and the z axis (the z-axis should be a single column, and of equal size to the number of spectra).

Save Figure (241)- This will save the current figure displayed in the figure window of the toolbox. Output options include Matlabs .fig allowing a user to customise the graphs to personal preference.

Plot Contour (245)- Plot contour launches a window allowing for interactively altering the contour level. This function can be launched by selecting two variables e.g. Synchronous spectra and the corresponding X-axis. Or, by selecting three variables e.g. moving windows, X-axis and Z-data (In this order). Again for plotting purposes the z-data should be in column format and be the size of total spectral number. When the window launches select a contour level (default is 10) and then plot the figure, the figures generated can also be saved.

Data Processing

![Data Processing](image)

Figure S4. 3: The layout of the main data processing options provided in the toolbox.

Strike Removal (297) - Strike removal removes anomalous spikes from spectra, it requires a minimum of 3 spectra to be present. Utilises a moving median filter along each wavenumber point as a function of the number of spectra E.g. if spectra are columns in a matrix of 1024x10 it would move in a window along each row. This approach is should be avoided for dynamic measurements where a lot of change is expected between each measurement.

Baseline (294) – This function is for the baseline correction of Raman spectra using a previously described method developed by Eilers et al\textsuperscript{[2]}.

Upon selecting the data you will be presented with an option for either automated or manual baseline correction. If you choose Automated the Raman spectra will be baseline corrected using the following parameters:

\[\text{lambda} = 1000;\]
\[p=0.01;\]
\[\text{Iterations} = 10;\]

Producing a new variable called Raman_Baselines in the workspace, Data is automatically numbered to prevent data being overwritten therefore if Raman_Baselines already exists the output would be Raman_Baselines1 and so on.
Figure S4. 4: The sub GUI which launches when selecting baseline routine for either Raman or ROA spectra.

If Manual is chosen a new GUI will open (Figure S4.5), this sub GUI can also be ran independently to the Raman Toolbox by executing Raman_Baseline in the Matlab command window. This allows the user to view the baseline and change the main variables which affect its fit. Only select the data you want to baseline correct when adjusting the baseline the average of the spectral set is displayed (to prevent a lag if the user has a large spectral set). When the user finds the baseline acceptable by pressing “Baseline Data” will baseline correct the entire matrix. This will produce a workspace variable called Raman_Baselined, and a structural array contacting the initial data, the baselined data, the baselines and the users chosen parameters. Data is sequentially numbered therefore if Raman_Baselined already exists the output would be Raman_Baselined1 and so on.

Figure S4. 5: The layout of the independent Raman baseline sub GUI.

Process ROA (344) – This is for the processing of ROA spectra, again the user is presented with two options automated and Manual (Figure S4.4). The automated option applies an automated processing procedure for ROA data. The automated function is designed deal with a series of raw ROA measurements, and carries out a number of functions to leave 1 singular ROA spectrum. These include using a median filters to eliminate any anomalous “spikes” in the data, summing the
spectra, generating and subtracting a baseline by using an 80 point median filter, and then smoothing the data with 15 point sgolay smoothing function.

If Manual is chosen a new SUB GUI will open (Figure S4.6). As with the Raman sub GUI this can be ran independently to the main toolbox by typing ROA_Baseline_sub_GUI into the Matlab command window. Upon launching, all variables from the main Matlab workspace are loaded into this GUI. Additionally, SPCs can be imported using the import ROA function (set for a folder containing both Raman and ROA .spc files). This sub GUI allows the user to inspect their data after each stage of processing by plotting it (select spectra and X-Axis).

1) The strike removal works the same as in the main GUI, only select the data to be processed.

2) The user is given the option to either average or sum their data, again only select the data to be process.

3) The final module concerns the baseline and smoothing the ROA data.

Firstly select the ROA spectra and X-axis, toggle which process you want to carry out e.g. baseline or smooth, and type in the editable box a value. This value is the number of points the smoothing window is carried out over. For a baseline, a larger value is suggested such as 80. Whereas smoothing data a smaller value such as 11 (needs to be odd) and a default polynomial order of 2 is used. Update will plot either the ROA data and the corresponding baseline, or the smoothed data. When the user is happy with the output press save. The output for baseline is ROA_Baseline (which is just the baseline produced) and ROA_Baselined which is the ROA data with the generated baseline subtracted from it. The output from smoothing is ROA_smoothed.

![Figure S4.6: The layout of the independent ROA sub GUI.](image-url)
Sum Spectra (Average_sub_GUI.M) – This launches a sub GUI (Figure S4.7), which gives you four options in total:

1) An option to average all data, leaving a singular spectrum.
2) An option to sum all data, leaving a singular spectrum.
3) An option to average the data into groups e.g. if a matrix consisted of 15 spectra and 3 was chosen, the output would consist of 5 spectra, where the first is the average of spectra 1-3, the 2nd 4-6 and so on. *Note the user has to be aware how many spectra are in their matrix, and that it’s divisible into a whole number by their grouping choice.
4) An option to sum the data into groups following the procedure described above.

Figure S4.7: The data averaging sub GUI.

Smooth Spectra (431) - This applies a Savitzky-Golay filtering to the variable selected. The user is promoted to select the window size which has to be odd (larger the window the higher the level of smoothing). Secondly a polynomial order is required (Figure S4.8)

Type “help sgolay” in Matlab command window for more information.

Figure S4.8: Smoothing Sub GUI.

Normalise Spectra (451) - This launches a window providing the user with three normalisation options:

1) SNV normalisation, which uses the standard deviation to normalise the spectra.
2) Band normalisation, if chosen the window will close and display the data in the main toolbox graph area, you then need to select the point through which you want to normalise.
3) Thirdly normalise $[0, 1]$ which scales data between 0 and 1 (useful for removing negative values which can arise through data processing).

**Interpolate Data (495)** – An important step in processing data especially for 2D correlation is making it evenly spaced, for this you need 3 variables: your spectral data set, X-axis i.e. Raman shift range and Z matrix which contains your perturbation variables in a row. E.g. if your perturbation was pH the Z matrix would contain a row of the pHs i.e. 1.0,1.3,1.7,1.9,2.5,3.4 etc. Once executed you are asked how many new data sets are required (Figure S4.9). If you just want to evenly space your data without introducing new sets chose 0, otherwise enter a value for how many more data sets you require.

The output is three variables X_New, Z_New and Interpolated_Y

![Interpolation Sub GUI](image)

**Figure S4. 9: The sub GUI for Interpolation of data.**

**Derivatives** - To calculate either first or second derivative select the data, and press the corresponding button. Derivatives are calculated using a 2 point central difference$^{[3, 4]}$.

**Resize Spectral Window** –

**Automated (594)** - To resize you need to select the variable to resize and the corresponding X-Axis. Automated removes generally unused regions of spectra resulting in spectra between 600-1800 cm$^{-1}$. The values used correspond will need changing for other users.

**Interactive resize** - Select both Spectra and X-Axis, this plots the data in the graph window of the Main GUI. Following this select the start of the region of interest, and click again at the end. This crops the data outside of the selected range, giving you newly resized data + X-Axis.

**Two Dimensional Correlation**

The Matlab code for the 2D functions are adapted from a number of sources including original publications$^{[5-9]}$

**Dynamic Spectra** - There are two options here, "Average" and "non", you are required to choose one (usually average) before calculating: Splice Spectrum, Synchronous Spectrum and Asynchronous Spectrum.
Slice Spectrum - Select the applicable variable and it calculates the auto peaks of the data, which identifies regions of intensity change. The result is a spectrum which corresponds to the diagonal of the synchronous plot where $v_1=v_2$.

Synchronous & Asynchronous Spectrum – Select the appropriate variable and press the corresponding button to calculate the synchronous or asynchronous matrix.

Moving Windows - After selecting the data and pressing the button, a sub window will open requiring you to choose a window size. After choosing a value the moving windows matrix will be calculated.

![Windows size sub GUI](image)

**Figure S4. 10:** Window size selection GUI.

PCMW- Requires selecting the data i.e.(Raman) and your perturbation range (Z) which should consist as a single row, in which the number of columns equals the number of columns in your data range. With correct data selected the same sub window from moving windows will launch, asking you to choose your window size (Figure S4.10). After choosing a value the PCMW matrix will be calculated.

**Additional Notes**

The program is currently not bug free, two main ones include zooming using mesh plot. Alternatively save the mesh plot, then alter it as the saved Matlab figure. Additionally if the workspace has all variables deleted from it can cause problems, requires closing and reopening the workspace.

**References**

8. Pazderska, T. and V. Kopecký, 2D Correlation spectroscopy and its application in vibrational spectroscopy using Matlab, Charles University, Prague.

Chapter 5: Monitoring *in vitro* insulin fibrillogenesis with Raman spectroscopy
5.0. Declarations

As the first author of this planned publication I performed all experimental work, and carried out all spectral analysis.
Monitoring *in vitro* insulin fibrillogenesis with Raman spectroscopy

Benjamin Gardner\(^a\) and Ewan W. Blanch\(^a\)

\( ^a \) Faculty of Life Sciences, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.

E-mail: E.Blanch@manchester.ac.uk; Fax: +44 (0)161 236 0409; Tel: +44 (0)161 306 5819

5.1. Abstract

Understanding the mechanisms of protein misfolding and fibrillogenesis is crucial due to these processes being implicated in a number of pathological conditions, including Alzheimer’s disease (AD) and type II diabetes. Raman spectroscopy has been shown to be sensitive to the structure of proteins and monitoring changes in protein conformation; therefore, it is particularly suited in monitoring protein misfolding and fibrillogenesis. In this investigation, the real time *in vitro* fibrillogenesis of insulin was monitored as a function of pH. Moreover, the reversibility of this process was also explored. This work showed the benefit of monitoring the fibrillogenesis of proteins in real time, especially when analysed with chemometric approaches including 2D correlation and principal component analysis. It was possible to easily identify rapid transitions in protein structure, and monitor differences in the kinetics as a function of pH. This is significant, as it expands the mechanistic understanding that would otherwise be missed when only a limited set of spectra are analysed.
5.2. Introduction

Diseases linked to the misfolding of proteins, or “protein conformational” diseases, are numerous and provide a serious burden to healthcare systems worldwide. These diseases include a number of neurological degenerative diseases, such as: Parkinson’s, Huntington’s, CJD and Alzheimer’s disease\textsuperscript{[1-3]}. Furthermore, these diseases are not restricted to neurological conditions, with type II diabetes a well known example of a non-neuropathic disease\textsuperscript{[1]}. While the structures of the proteins implicated in these diseases are diverse, they undergo a similar process which is linked to their pathology. Through the process of fibrillogenesis, proteins lose their native structure, and aggregate into elongated insoluble protein fibrils. Moreover, it is believed that in vitro all proteins can form fibrils under the right conditions; this is regardless of the initial structure and function of the particular protein or peptide in question\textsuperscript{[2, 4]}. Fibrils have been defined through their biophysical characteristics, in particular their cross-β X-ray diffraction pattern\textsuperscript{[5]}. Through this it has been shown that the structure of the fibril consists predominantly of intermolecular β-sheet, which is ordered so that the stands are orientated perpendicularly to the fibril axis. Fibrils have received much attention\textsuperscript{[5-9]}, and X-ray diffraction has been used to study micro crystals of short fibrillogenetic peptide sequences termed self recognition elements (SRE)\textsuperscript{[10]}. These investigations showed that amyloidogenic sequences can pack in a number of orientations, which might account for certain polymorphisms that are witnessed between fibrils\textsuperscript{[10]}. While the mentioned neurological diseases are defined by the accumulation of these insoluble β-sheet containing fibrils, they themselves are not directly responsible for the disease pathology. A number of studies have found that it is the soluble intermediate species formed during fibrillogenesis that are linked to their pathology\textsuperscript{[3, 11-14]}. Therefore, improving the knowledge of the mechanism of fibrillogenesis is essential for the development of therapeutic agents to treat and prevent these diseases.

Vibrational spectroscopic techniques such as Raman spectroscopy, Raman optical activity (ROA), deep UV Raman spectroscopy and vibrational circular dichroism (VCD) have been shown to be particularly suited to the study of protein structure and dynamics. These techniques are being increasingly applied in the study of fibrillogenesis, as they can monitor all aspects of the process, whereas, the classic techniques of structural biology i.e. NMR and X-ray crystallography are more restricted. Moreover, when vibrational spectroscopy is coupled with powerful chemometric approaches such as principal component analysis (PCA) and 2D-correlation (2D-Cos) it allows the elucidation of important details, such as, transition points of structural unfolding, or identification of small differences in structures.

Insulin, a small peptide hormone consisting of two peptide chains linked by two disulfide bonds has received considerable attention due to its fibril forming propensity. Insulin has been shown to be particularly sensitive to its in vitro conditions, with fibrils being formed under a range of conditions including temperature, pressure, agitation and shear flow\textsuperscript{[15-20]}. Moreover, the end products of fibrillogenesis have also been shown to be particularly sensitive to the in vitro conditions. When the conditions are altered changes are induced in the supramolecular structure of fibrils i.e. reorganisation can occur between left and right twist of the fibrils\textsuperscript{[21, 22]}. In the case of
insulin it was shown that below pH 2.1 one twist of fibril is observed and above pH 2.4 the opposite twist is observed. An additional example of the sensitivity of insulin has been shown when the in vitro conditions significantly alter the end product, to that of spherulites.

Work that investigates the fibrillation of proteins often does so in snap shots, where the sample is re-measured at specific time points. This is due to a number of reasons: such as the duration of fibrillation, the time a measurement takes on a specific piece of equipment, or how invasive the measurement is to the sample. While any detail to explain fibrillation mechanism is important, it is clear this approach is not sufficient as key mechanistic detail could easily be missed.

Raman spectroscopy is particularly suited to monitor processes such as protein fibrillation in real time. The measurements are non-destructive, a high quality spectrum can be gather in less than a second and measurements can be carried out indefinitely.

For the above reasons Raman spectroscopy has been used in real time to monitor the time dependent reversible unfolding and fibrillation of bovine insulin in vitro. This work demonstrates the suitability of Raman spectroscopy for monitoring fibrillation in real time. Additionally, fibrillation has monitored above and below pH values that have been previously reported to significantly influence fibrillation as discussed above. Finally, chemometric analysis including PCA and 2D-Cos have been used in combination to maximise the information of the mechanistic process of fibrillogenesis. It is possible to probe the sequence of changes in the structural unfolding of insulin and subsequent aggregation / association, as well as identifying further key transitions in fibrillogenesis.

5.3. Materials and Methods

Bovine insulin was purchased from Sigma Aldrich (UK), and used without further purification. Samples were prepared in a 0.1 M glycine buffer which was pH adjusted using HCl to either pH 2 or 3. Samples were then prepared to a final protein concentration of 50 mg ml\(^{-1}\) at either pH 2 or 3. Using a high concentration of insulin allows a high quality Raman signal to be gathered in around 1 second. It has been previously shown that lowering the pH increases the concentration of insulin monomers. Moreover, the fibrillation of insulin is dependent on insulin being in the monomer form. Samples were filtered with a 0.22 µm Durapore membrane, prior to loading into a quartz micro-fluorescence cell. All Raman and ROA spectra were recorded on a previously described ChiralRaman instrument (Biotools Inc, Juipier, FL, USA). Raman and ROA spectra were gathered for insulin samples under the following conditions: laser excitation 532 nm, spectral resolution 7 cm\(^{-1}\), laser power at the sample ~600 mW and acquisition time ~6–36 hours.

In situ Raman measurements monitoring fibril formation of insulin were carried out at 80°C, with an acquisition time of ~1.2 s for ~1½ hours (~2600 individual spectra). Previous work has shown insulin fibrils to form at various temperatures, with fibrillation occurring quicker at higher temperatures. Therefore, 80°C was chosen as for the fibrillation experiments.

The temperature of the samples was controlled using a custom cell holder, which was coupled to a Julabo F12 heating/cooling unit. This allowed precise control of the temperature of the sample to within 0.1°C, which was verified using a Jenway digital temperature meter connected to a
thermocouple. During the experiments monitoring temperature induced unfolding / refolding, the samples were heated at 80°C for 4 and 8 minutes, for bovine insulin samples at pH 2 and 3, respectively. Following the heating period the temperature was quenched from 80°C at a rate of ~20°C / min to a final temperature of 20°C, which was maintained for the remainder of the experiment.

The Raman and ROA spectra measured were processed in Matlab 2010, using a previously described in-house toolbox [28] (detailed in Chapter 4) and the bioinformatics toolbox available in Matlab [29]. All Raman spectra were baseline corrected using the asymmetric least squares smoothing method developed by Eilers et al [30], to remove background intensity variations between measurements. Moreover, aberrant spikes attributed to cosmic rays were removed from the spectra and the Raman profile of the glycine buffer was subtracted. Raman spectra monitoring in situ changes were further smoothed using the Matlab function mslowess [31], which uses a locally weighted regression approach to smooth the data. The individual cycles of ROA data had aberrant spikes attributed to cosmic rays removed prior to being averaged. The spectra were then baseline corrected using a median filter, before being smoothed using a Savitzky-Golay smoothing filter to further improve the quality of the data.

Chemometric approaches including principal component analysis (PCA) and 2D correlation (2D-Cos) provide a powerful means for sample analysis and are widely applied in spectroscopy [16, 32-39]. PCA allows the reduction in the dimensionality of a data set while preserving the variance. In PCA a data set is reduced into a series of principal components (PCs), where each PC accounts for a decreasing amount of the variance of the original data set [16, 36]. When displaying PC scores, each data point relates to an original spectrum. Plotting the PC scores allows differences or similarities between spectra to be easily observed on a 2D axis, through either their clustering or separation on the plot. An additional way to display the data from PCA is through the use of loadings plots. In a loadings plot the variance of a specific PC, i.e. PC1 is related to the original variables i.e. wavenumber range. This allows identification of spectral feature attributed to variance in the data set. 2D-Cos approaches including perturbation correlation moving windows (PCMW) are suited to monitoring changes in spectra as a function of a perturbation e.g. time, temperature, pH, etc. PCMW was calculated in Matlab using a previously described in house toolbox [28]. PCMW was applied to subsets of Raman spectra (~30 spectra) monitoring different phases of insulin heating. The processed Raman spectra were mean centred prior to the calculation of PCA scores and loadings plots in Matlab. PC loadings were used to analyse phases of insulin heating to explain the variance of the data set. While, scores plots were used to compare insulin at specified time points during heating and temperature quenching experiments.

5.4. Results and Discussion

The Raman and ROA spectra of Bovine insulin are shown (Figure 5.1), at both pH 3 and 2. The spectra of both the Raman and ROA are very similar between the two pHs. However, minor differences are observed in the Raman spectra, which is likely due to a result of buffer subtraction.
A number of band assignments are known for the Raman spectra of insulin, which are detailed in supplementary information (Table S5.1). The amide I region of a protein Raman spectrum (1620–1700 cm\(^{-1}\)) contains a number of secondary structure marker bands, including those of, α-helix, β-sheet, disorder and turn-type structure. Moreover, the structures that contribute to the amide I region have relative differences in their bandwidths and scattering intensities\(^{[40]}\). Therefore, monitoring changes in the full width at half height (FWHH) of the amide I region can be informative of changes in secondary structure of the sample.

The FWHH of the amide I region of insulin was monitored as a function of time at pH 2 and 3 (Figure 5.2), while the samples were heated at 80°C. A similar trend is observed at both pHs, where a brief increase in the FWHH is observed prior to a rapid decrease to ~25 cm\(^{-1}\); the FWHH then appears to show no change for ~10 minutes and ~16 minutes for pH 2 and 3, respectively, after which the FWHH decreases further to ~23 cm\(^{-1}\). The Raman spectra are shown in four panels (A–D) Figure 5.3 and 5.4 for pH 3 and 2, respectively. To facilitate the analysis of the spectral changes as a function of time PCMW and PCA analysis have been used in combination.

The PCMW contour plot (Figure 5.5A) for insulin at pH 3 monitoring the first 60 s of heating shows a rapid decrease (~20 s) in the intensity of a number of marker bands. Following, these decreases at ~45 s three bands positioned at ~1000, ~1200 and ~1670 cm\(^{-1}\), show an intense increase in their intensity. The PCMW plot monitoring insulin at pH 2 (Figure 5.6A) show a very similar response with respect to the position of the intensity changes and the time the kinetics of the process change.

Monitoring the 2\(^{nd}\) phase of change with PCMW (Figure 5.5B), as indicated by the FWHH analysis of the amide I region, shows that the most intense changes are centred around ~1150 s. During this time period the most intense change is observed in the amide I region(Figure 5.5B). A number of other bands are also possibly changing, however, the noise in the spectra prevents determining whether this is the case through this plot. The changes observed in the PCMW plot for pH 2 (Figure 5.6B) are slightly clearer. Simultaneous changes are observed at ~625 s, nearly half the time before the corresponding transition occurred at pH 3. An intense increase is observed in the amide I region, while decreases are observed in the Raman bands at ~ 600, ~800 and1000 cm\(^{-1}\).

Due to the difficulty of determining band positions from contour plots the loadings intensities from PC1 were used. As under initial heating the PCMW analysis showed changes occurring in two distinct phases this was investigated further in terms of these two periods using loadings plots. These help to clearly identify the bands that are changing and their significance, and prevent the analysis being dominated by the intense increases of certain bands. As the 2\(^{nd}\) phase of change appeared to show a simultaneous change in the bands it was not divided into smaller spectral sub-sets.

The loadings intensities correlate the variance in a principal component (PC), to the initial variables i.e. wavenumber range. A strong correlation between a variable and a PC, indicates that the variable is responsible for the variance in the PC. PC1accounts for most the variance observed (87.2%), for the initial heating phase of insulin at pH 3, and the loadings plot of PC1 is displayed in
Figure 5. 7A. In the amide I region of Figure 5. 7A changes are observed at two bands centred at 1658 and 1681 cm\(^{-1}\), which are assigned to α-helix and intramolecular β-sheet, respectively\cite{15, 16}. Further secondary structure sensitive bands are observed at 1348 and 1285 cm\(^{-1}\) for α-helix, at 1260 cm\(^{-1}\) indicating disordered conformations and at 1230 cm\(^{-1}\) for intermolecular β-sheet. Additionally, changes in the band arising from the v(C-C) stretching vibration at ~890 and from the v(C-N) at ~1115 cm\(^{-1}\) show that changes in the conformations exhibited by the protein backbone are occurring. Smaller changes in the intensity of side chain bands including those for Phe and Tyr can be observed at 1620, ~1205, ~1180 and ~1007 cm\(^{-1}\). A nearly identical loadings plot is observed for the same phase for insulin at pH 2 (Figure 5. 8A).

The loadings plot of PC1 (99.7 %) for the 2nd phase of insulin fibrillogenesis (Figure 5. 7B), is dominated by three bands at 1675, 1228 and 1007 cm\(^{-1}\). The first two bands are assigned to intermolecular β-sheet and the latter to Phe. Further bands of interest are positioned at ~830 and ~850 cm\(^{-1}\), which are assigned to a Fermi doublet of Tyr\cite{15, 41, 42}. The relative intensities of the two bands \(I_{850} / I_{830}\) is indicative of the hydrogen bonding state of the phenoxyl group e.g. a ratio of 0.3 is indicative of hydrogen bond donor, while a value of 1.3 indicates phenoxyl groups that can both donate and accept hydrogen bonds. Referring back to the original Raman spectra, the ratio of \(I_{850} / I_{830}\) is observed to change from ~1.2 to ~0.9. This indicates a significant change in the environment of the four Tyr residues, specifically a decrease in their average exposure to solvent. It is noteworthy that the average change in the Tyr environment is coincident with intermolecular β-sheet formation, and a change in the Tyr Fermi doublet is not observed during the loss of α-helix. As is shown in the supplementary information (Figure S5. 1), three of the four Tyr residues are located within α-helical structure. Therefore, it is interesting that the overall environment does not change prior to the formation of intermolecular β-sheet. Again, similar trends are observed in the loadings intensities for PC1 for insulin at pH 2 (Figure 5. 8B).

The loadings plot for PC1 (90.9 %) for the final phase of change (as indicated by the FWHH and PCMW analysis), is dominated by a band in the amide I region at 1679 cm\(^{-1}\). Further changes in the protein backbone are evident from the band at ~1020 cm\(^{-1}\), and a further β-sheet band ~1050 cm\(^{-1}\). Changes in the side chains of Phe residues are observed from the band at 996 cm\(^{-1}\), while there is a further change Tyr Fermi doublet at ~1050 cm\(^{-1}\), with an increase in intensity of the band at ~856 cm\(^{-1}\). This results in an increase is of the ratio of \(I_{850} / I_{830}\) from 0.9 to 1.03. This indicates that there is an increase in the solvent exposure for the average Tyr residues during this phase. Furthermore, there is a slight downshift of ~2 cm\(^{-1}\) of the band previously found at 834 cm\(^{-1}\). There are slight differences between the loadings of PC1 at pH3 of the same phase when compared to pH 2 (Figure 5. 8C). This difference is most noticeable for the two bands at ~1020 and 1050 cm\(^{-1}\), where at pH 2 the intensity of the band at 1020 cm\(^{-1}\) is far greater than the band at ~1050 cm\(^{-1}\), while at pH 3 the two bands share a similar relative intensity. At both pHs there is change indicated ~790 cm\(^{-1}\) in the loadings intensity, however, at pH 3 there appears to be a change in the baseline with no observable band, while at pH 2 there is the emergence of a band at this position. This is likely highlighting a mechanistic difference between the two samples as a function of pH.
Previous studies that have monitored the formation of insulin fibrils with spectroscopic techniques including Raman and Infrared spectroscopy (IR), have done so in a more limited way; where either, only a comparison of the initial and final spectrum of the sample is made, or, a more limited sampling as a function of time is made\textsuperscript{15, 18, 19, 43}. However, as has been shown here, where rapid transitions occur, mechanistic detail will be missed when only a limited series of spectra are collected and analyzed.

As discussed above, insulin has a propensity to form fibrils under a number of conditions. However, removing the driving force of fibrillation, i.e. elevated temperature, can allow the native state to reform. The amide I FWHH response as a function of time, under initial heating at 80°C followed by temperature quenching to 20°C for bovine insulin at pH 3 is shown in Figure 5. 9. The FWHH initially shows a similar profile as to what was seen previously (Figure 5. 2), where there is an initial increase in the FWHH followed by a drop to ~25 cm\(^{-1}\) upon heating. However, when the temperature is quenched to 20°C there is an initial rise in the amide I FWHH which is followed by a sigmoidal response in the FWHH returning back to the initial FWHH of ~40 cm\(^{-1}\) over the course of the experiment. The corresponding Raman spectra (Figure 5. 11) are separated into four panels (A–D), representing the four key phases of the experiment: A, the initial unfolding and formation of β-sheet intermediate; B, the lag phase where no changes in the Raman spectra are apparent; C, the spectra where temperature quenching from 80°C to 20°C occurs; D, the spectra monitoring the refolding. Of these four panels most change is observed in panels A and D, respectively; where the changes observed in the spectra look similar, albeit the reverse. Upon temperature quenching, panel C, changes are observed in a few regions of the spectra. Specifically change is observed in the amide I and III regions, as well as a shift in band position occurring at ~1410 cm\(^{-1}\).

The PCMW analysis of the Raman spectra from the initial phase (Figure 5. 12 A), while the sample is undergoing heating, shows that unfolding precedes the initial β-sheet formation. This is in agreement with the data observed previously when insulin is undergoing heating. However, the refolding of insulin to the native state (Figure 5. 12 B), occurs simultaneously with the loss of β-sheet. Moreover, the rates of heat induced unfolding vs refolding are significantly different. Where the formation of β-sheet intermediate occurs in ~80 s, while for the entire sample to refold to the “native state” takes ~14 hours. It is likely that dissociation of β-sheet intermediates is the slowest step and refolding of the disassociated monomer occurs quickly. Therefore, monitoring the “refolding” is likely to involve monitoring the change in the heterogeneous population of protein structures i.e. soluble oligomers and other aggregates.

PCA was used to analyse the four key phases of this process (with 10 replicates of each): initial spectra, lag phase (intermediate 1), spectra following temperature quenching (intermediate 2) and the final spectra i.e. refolded native insulin. The scores plots of PC 1 plotted against PC2 (Figure 5. 13A) showed a clear separation between the two intermediate sets of spectra and the initial and final spectra. Moreover, there is an overlapping of the initial and final data points. The loadings plot of PC1 (Figure 5. 13B) shows that most of the variance is accounted for in bands assigned for β-sheet, α-helix and side chain bands including Phe and Tyr, as seen previously. Finally, to confirm the stability and structure of the refolded insulin sample the ROA was re-measured (Figure S5. 2).
Due to the increased rate of kinetics of aggregation at pH 2, the sample was heated for a shorter period before the temperature was quenched to 20°C. Initially the FWHH of the amide I region (Figure 5.14) shows similarities with that of the pH 3 sample. Specifically, a rapid drop in FWHH occurs through heating, however, upon temperature quenching a further drop in FWHH is observed prior to the gradual increase observed previously for the pH 3 sample. Moreover, while the pH 3 sample was stable after refolding, the sample at pH2 was not. As can be seen from the FWHH of the amide I region, after returning to near the FWHH of the start point, the FWHH then drops down to ~23 cm$^{-1}$. It is significant that there is no observed lag phase in the amide I region decreasing directly to 23 cm$^{-1}$, indicating that this has likely been abolished due to seeding through β-sheet intermediates formed through the initial heating that did not disassociate and refold. The analysis of the Raman spectra of the initial unfolding and refolding was similar to what was observed at pH 3. Please refer to Supplementary Information Figure S5.2–4 for further details.

5.5. Conclusions

By studying both the fibrillogenesis of insulin, and probing the stability of this process a better understanding of the fibrilogenetic pathway can be developed. For example, these results clearly establish the unfolding of insulin precedes β-sheet formation. It was shown at both pH 2 and 3 that bovine insulin rapidly unfolds losing the predominant α-helical structure when heated, which is rapidly followed by formation of a β-sheet structure containing intermediate product. However, both unfolding and initial β-sheet formation appear to be fully reversible. This separates insulin from other model proteins used in fibrillogenesis studies, such as lysozyme, where initial unfolding is irreversible[44]. Moreover, fibrillogenesis of insulin clearly shows a more complex relationship than a simple two state transition. At both pHs a lag period is observed following initial β-sheet formation, which is significantly shorter at the lower pH. This lag phase is then followed by further changes in β-sheet structure and the environment of the side chains including Tyr and Phe.

The importance of aromatic side chains in the self assembly of proteins and fibrillation has been shown previously[9, 45–47]. While aromatic residues are not a necessity for fibrillation, they do have a strong influence on the kinetics of the process. Moreover, it has been shown that for insulin during fibrillogenesis aromatic residues such as Tyr and Phe are buried as has been observed for the protofibrils[48] and mature fibrils[49, 50]. Therefore, the second phase of changes in the aromatic residues could represent irreversible association of smaller soluble species into the insulin protofibril structure.

While most the spectral changes observed were very similar for Insulin at both pHs there were observable differences in the Raman spectra. Specifically, there was a band at 790 cm$^{-1}$ that is observed to appear when fibrils are formed at pH 2, that is absent at pH 3. Through the work of Chapter 3, this appears not to be attributed to a side chain as no bands were observed at this position. Moreover, the Raman and ROA measurements of insulin at both the pHs investigated prior to heating showed no presence of a band at 790 cm$^{-1}$. Therefore, this indicates a mechanistic difference in the process when fibrils are formed at different pHs. This is interesting, as previous work has commented on differences in the kinetics as a simple function of insulin and the
concentration of the monomer.\textsuperscript{[19, 26]} However, observing discrete differences in the Raman spectra as a function of pH during fibrillogenesis likely indicates a more complex relationship. This work builds on other studies investigating the behaviour of insulin fibrils as a function of pH.\textsuperscript{[22]} Where previous studies have shown a difference in the Raman spectra as a function of pH during fibrillogenesis likely indicates a more complex relationship. This work builds on other studies investigating the behaviour of insulin fibrils as a function of pH.\textsuperscript{[22]} Where previous studies have shown a difference in the chiral signal observed for fibrils formed at different pHs, this work shows that there are observable differences in the Raman signal during fibrillation.

This work clearly shows the advantage in combining different methods of analysis, i.e. FWHH, PCMW and PCA; as a detailed analysis of the fibrillogenesis of insulin fibril formation monitored in real time is readily feasible. Through the FWHH analysis it is possible to easily identify transition points as a function of time, where PCMW and the loadings intensities were then used to explain the variance that is occurring and the sequence of events as a function of time. This type of analysis is crucial if the mechanistic understanding how chemical moieties rationally designed therapeutic agents perturb fibrillogenesis.

5.6. References


29. Matlab and Bioinformatics toolbox,R2010a;The MathWorks, Inc.:Natick, Massachusetts, United States.,2010


Figure 5. 1: The Raman (A) and ROA (B) spectra of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) at pH 2 (blue) and pH 3 (red) measured at room temperature.

Figure 5. 2: FWHH analysis of the amide I region of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) as a function of time (log\(_{10}\) s) at pH 3 (red) and pH 2 (blue) while undergoing heating at 80°C.
Figure 5.3: Raman spectra of four phases (A–D) of bovine insulin’s (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) fibril formation being monitored in situ at pH 3 at 80°C. The first 60 s of heating are contained in (A and B), both containing 15 spectra, respectively. An invariant phase accounting for ~850 s (C) containing 450 spectra, while D contains ~200 spectra accounting for ~400 s.
Figure 5.4: Raman spectra of four phases (A–D) of bovine insulin’s (50 mg ml⁻¹ in 0.1 M glycine buffer) fibril formation being monitored in situ at pH 2 at 80°C. The first 80 s of heating are contained in (A and B), with A containing 15 spectra and B containing 25 spectra, respectively. An invariant phase accounting for ~440 s (C) containing 220 spectra, while D contains ~200 spectra accounting for ~400 s.
Figure 5.5: PCMW analysis of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 3) under heating at 80°C. where the first ~60 s are accounted for in A, while B monitors ~800 – 1300 seconds. Colour scale identifies direction and intensity of change.
Figure 5. 6: PCMW analysis of bovine insulin (50 mg ml⁻¹ in 0.1 M glycine buffer at pH 2) under heating at 80°C. where the first ~60 s are accounted for in A, while B monitors ~400 – 800 seconds. Colour scale identifies direction and intensity of change.
Figure 5.7: PC loadings of three phases of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer) fibril formation at pH 3 (A–C) under heating at 80°C.
Figure 5.8: PC loadings of three phases of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 2) fibril formation (A–C).
Figure 5.9: Figure 5.10: FWHH of the amide I region of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) as a function of time (log$_{10}$ s). Sample initially under heating at 80°C prior to quenching to 20°C. Four phases (A–D) are indicated. Phase A accounts for ~0-80 s, B ~80–360 s, C ~470–600 s and finally D, 600 s to 14 hours.
Figure 5.11: Real time Raman spectra of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 3) monitoring spectral response as a function of temperature and time: A, the first 40 spectra while undergoing heating at 80°C, accounting for ~0–80 s; B, 140 Raman spectra ~80–360 s, while sample still undergoing heating at 80°C; C, 65 Raman spectra ~470–600 s following quenching of temperature to 20°C; D, 26 Raman spectra (1 every 1000) taken from 600 s – 14 hours.
Figure 5. 12: PCMW of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 3) under heating (40 spectra) at 80°C (A) and refolding (26 spectra) (B).
Figure 5.13: PCA scores plots of four sets of 10 Raman spectra of bovine insulin (50 mg ml⁻¹ in 0.1 M glycine buffer at pH 3) under unfolding and refolding conditions, taken at different time points: initial spectra comprise of the first 10 spectra under heating at 80°C; intermediate 1, 10 spectra from the lag phase while being heated at 80°C; intermediate 2, 10 spectra following temperature quenching to 20°C; final, the last 10 spectra of the data set. Loadings plot of PC1 (B).
Figure 5. 14: FWHH of the amide I region of bovine insulin (50 mg ml\textsuperscript{-1} in 0.1 M glycine buffer at pH 2) as a function of time. Sample initially under heating at 80°C, prior to quenching to 20°C. Temperature remained at 20°C for the rest of the experiment.
5.7. Supplementary information

Table S5. 1: Raman band assignments for insulin$^{[15,16]}$.  

<table>
<thead>
<tr>
<th>Structure</th>
<th>Wavenumber (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-sheet (intramolecular)</td>
<td>1682</td>
</tr>
<tr>
<td>β-sheet (intermolecular)</td>
<td>1677</td>
</tr>
<tr>
<td>α-helix</td>
<td>1658</td>
</tr>
<tr>
<td>Disorder</td>
<td>1648 &amp; 1678</td>
</tr>
<tr>
<td>Tyr</td>
<td>1620</td>
</tr>
<tr>
<td>Phe &amp; Tyr</td>
<td>1605</td>
</tr>
<tr>
<td>Phe</td>
<td>1585</td>
</tr>
<tr>
<td>CH$_2$ deformation</td>
<td>1450</td>
</tr>
<tr>
<td>Symm CO$_2^-$ stretching</td>
<td>1418</td>
</tr>
<tr>
<td>α-helix</td>
<td>1340</td>
</tr>
<tr>
<td>CH deformation</td>
<td>1314</td>
</tr>
<tr>
<td>α-helix</td>
<td>1270,1284</td>
</tr>
<tr>
<td>Disorder</td>
<td>1257</td>
</tr>
<tr>
<td>β-sheet (intramolecular)</td>
<td>1236,1250</td>
</tr>
<tr>
<td>β-sheet (intermolecular)</td>
<td>1230</td>
</tr>
<tr>
<td>Phe</td>
<td>1207</td>
</tr>
<tr>
<td>Tyr</td>
<td>1182</td>
</tr>
<tr>
<td>Tyr</td>
<td>1173</td>
</tr>
<tr>
<td>ν(C-N)</td>
<td>1110</td>
</tr>
<tr>
<td>β-sheet</td>
<td>1052</td>
</tr>
<tr>
<td>Phe</td>
<td>1000</td>
</tr>
<tr>
<td>disordered</td>
<td>955</td>
</tr>
<tr>
<td>α-helix</td>
<td>930</td>
</tr>
<tr>
<td>ν(C-C)</td>
<td>895</td>
</tr>
<tr>
<td>Tyr Fermi Doublet</td>
<td>830 &amp; 855</td>
</tr>
<tr>
<td>ν(C-S) of disulphide link</td>
<td>664</td>
</tr>
<tr>
<td>Tyr</td>
<td>643</td>
</tr>
<tr>
<td>Phe</td>
<td>620</td>
</tr>
<tr>
<td>Skeletal bending</td>
<td>560</td>
</tr>
<tr>
<td>ν(S-S)</td>
<td>515</td>
</tr>
</tbody>
</table>
Figure S5. 1: Cartoon representation of Insulin (pdb:2A3G), with the A chain in green is shown at the front (A), while the B-chain is shown at the front (B).

Figure S5. 2: ROA spectra of bovine insulin (50 mg ml\(^{-1}\) in 0.1 M glycine buffer at pH 3) prior to heating (blue), and after initial heating and refolding. (red).
Figure S5. 3: Real time Raman spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 2) monitoring spectral response as a function of temperature and time: A, the first 35 spectra while undergoing heating at 80°C, accounting for ~0–70 s; B, 90 Raman spectra ~70–250 s, while sample still undergoing heating at 80°C; C, 89 Raman spectra ~250–430 s following quenching of temperature to 20°C; D, 26 Raman spectra (1 every 1000) taken from 430 s – ~14 hour
Figure S5. 4:PCM of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 2) under heating (35 spectra) at 80°C (A) and refolding at 20°C (26 spectra) (B).
Figure S5. 5: PCA scores plots of four sets of 10 Raman spectra of bovine insulin (50 mg ml$^{-1}$ in 0.1 M glycine buffer at pH 3) under unfolding and refolding conditions, taken at different time points: initial spectra comprise of the first 10 spectra under heating at 80°C; intermediate 1, 10 spectra from the lag phase while being heated at 80°C; intermediate 2, 10 spectra following temperature quenching to 20°C; final, the last 10 spectra of the data set. The loadings plot of PC1 (B).
Chapter 6: Investigations of the independent fibrillation of the A and B chains of bovine insulin
6.0. Declarations

As the first author of this publication I performed all Raman and ROA experiments measured at 532 nm, as well as all the sample preparation for all the work described. Additionally the AFM measurements of samples at pH 2 were performed by me, however, AFM measurements of B-chain fibrils at pH 3 and 1 were carried out by D. Kurouski. Deep UV resonance Raman measurements were performed by V. Sikirzhytski and K. Serada, while TERS experiments were performed by T. Deckert. I performed the analysis and processing of all the spectra.
Investigations of the independent fbrillation of the A and B chains of bovine insulin

Benjamin Gardner, Vitali Sikirzytks, Katerina Serada, Dmitry Kurouski, Tanja Deckert, Igor Lednev, and Ewan W. Blanch

a Faculty of Life Sciences, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.
b Institute for Physical Chemistry, University of Jena, Helmholtzweg 4, 07743 Jena, Germany.
c Department of Chemistry, University at Albany, SUNY, 1400 Washington Ave., Albany, NY 12222, USA.
E-mail: E.Blanch@manchester.ac.uk; Tel: +44 (0)161 306 5819; Fax: +44 (0)161 236 0409

6.1. Abstract

The formation of protein fibrils through the process of fibrillogenesis is known to be the pathological cause of a number of neurodegenerative diseases, including Alzheimer’s disease. Therefore, understanding fibrillogenesis at the molecular level is important, for the development of therapies to prevent or cure these diseases. We demonstrated through combining Raman and Raman optical activity (ROA), deep UV Raman and Tip enhanced Raman spectroscopies it was possible to probe the fibrillogenesis of the independent A- and B-chains of insulin in detail. Key findings from this study included identification that the A- and B-chains of insulins secondary structure are highly sensitive to pH, which influenced the pathway of fibrillogenesis from fibrils to spherulites in the B-chain. Moreover, the two chains are structurally dissimilar, which is reflected in their respective behaviours during fibrillogenesis and the morphology of the fibrils formed. Raman and ROA were used to investigate the structure of the independent chains in the non-aqueous environment DMSO-d6. Further differences in structure of the two were observed, in what is termed the stabilised intermediate, which might rationalise differences observed in the fibrils. Deep UV Raman was used to probe the initial A- and B-chain momomers, as well as the fibril core, where differences were observed between the two with respect to secondary structure. For the first time we showed TERS analysis of the surface chemistry of the independent A- and B-chain fibrils. Through this it was observed that the surface chemistry of the fibrils of the A- and B-chains were different; not only in the preferences of amino acids at the surface, but in the secondary structure composition. Overall, this work demonstrated the value and sensitivity of Raman spectroscopies in the understanding of protein fibrillogenesis.
6.2. Introduction

The pathology of over 40 diseases have a shared commonality, which is the misfolding and aggregation of natively soluble proteins into elongated insoluble amyloid fibrils\(^1\). These fibrillogenic diseases includes Alzheimer’s, Parkinson’s, Huntingdon’s and type II diabetes\(^2-5\). While the process of fibril formation of specific proteins \textit{in vivo} has been linked to the pathological cause of a number of diseases, it is believed that \textit{in vitro} fibril formation, or fibrillogenesis, is possible for all proteins under the right conditions\(^6-9\); regardless of the initial size, sequence or structure of the protein\(^1\). Different biophysical properties of Amyloid fibrils has lead to them being defined in various ways. These properties include their ability to bind certain dyes, such as Congo red and thioflavin T (ThT) and a distinctive cross-β X-ray diffraction pattern\(^10\). This cross-β pattern has revealed that the constituent polypeptides are arranged predominantly into β-strands which are oriented perpendicularly to the fibril axis, while the β-sheet runs parallel to the fibril axis. While fibril structures share this overall similarity, they exhibit polymorphism in their architecture. Polymorphism has been reported between fibrils formed from different proteins\(^1\), for fibrils of the same protein formed under different conditions\(^11\), and polymorphism can also be an inherent phenomenon of the polypeptide sequence\(^12, 13\). Fibril polymorphs can exist in a number of forms, such as differences in the supramolecular structure i.e. a fibril can posses either a left or right handed twist\(^14\). Other polymorphisms can include the type of β-sheet constituting the fibril i.e. parallel vs anti parallel\(^15\), or differences in the height, width and the periodicity of the fibril\(^16-18\).

Another important way in which fibrils can differ is in the composition of the fibril core, specifically the sequence recognition element (SRE) that constitutes it\(^12\). The SRE is the minimum sequence of amino acids from the amyloidogenic protein or polypeptide that is required for self association into fibrils\(^1, 12\). This sequence has been shown to be as small as three amino acids in the case of tau (VYK). Other examples include (KLVFFAE) for β-amyloid, (VAQKTV) for α-synuclein and (SNQNNF) for the human prion protein\(^1\).

An additional interesting case is that of insulin, which is a small predominantly α-helical peptide hormone comprised of two polypeptide chains termed the A- and B-chain. The two polypeptide chains of insulin are linked by two disulfide bonds A\(_7\)cys–B\(_7\)cys and A\(_20\)cys–B\(_19\)cys. In addition to the two disulfide bonds between the two chains, the A chain contains an internal disulfide bond between A\(_6\)cys–A\(_11\)cys. Under physiological conditions and zinc coordination insulin forms a hexamer through three dimers\(^19\). The dimerisation of insulin is controlled through the C-termini of the B-chain, as is shown in Supplementary Information Figure S6. 1. In insulin one SRE is found in the A chain (LYQLEN) while a second is present in the B chain (VEALYL). The work of Sawaya et al.\(^12\) explored how a number of SREs pack when investigating these short sequences in micro-crystals. From this work they produced a classification system where eight classes of packing were established. However, a number of the SREs contain amino acids which have a side chain in which the protonation state will change as a function of pH. An example of this is the pair of insulin SREs, where both sequences contain Glu; the protonation state of Glu might act to influence or alter the packing of the SRE. The influence that charged side chains can have on secondary structure in
model studies is well established, with poly Glu and poly Lys\textsuperscript{20-23} being thoroughly examined. Therefore, pH influences on SRE conformation may be significant and important for fibrillogenesis.

While understanding the nuances of fibril structure is important, the mature fibrils themselves are not thought to be the cause of the pathology of the disease they are implicated in. Increasing evidence shows the toxic species in these diseases are the soluble intermediates oligomers of fibrillogenesis \cite{4, 24-26}.

However, due to the heterogeneous, and short lived transient nature of these intermediaries, and large insoluble nature of fully formed fibrils it is a challenging phenomenon to study\textsuperscript{27}. The more “classic” methods for protein structural determinations are finally elucidating structural information, with approaches including: solid state nuclear magnetic resonance (ssNMR), electron paramagnetic resonance (EPR), small angle X-ray scattering (SAXS) and X-ray crystallography\textsuperscript{27}. However, methods based on vibrational spectroscopy have shown their sensitivity and suitability for studying all aspects of fibrillogenesis. Such methods include Raman spectroscopy\textsuperscript{28, 29}, Raman optical activity\textsuperscript{30, 31}, deep UV Raman spectroscopy\textsuperscript{13, 32, 33}, vibrational circular dichroism (VCD)\textsuperscript{14, 34} and tip enhanced Raman spectroscopy (TERS)\textsuperscript{35, 36}. This range of techniques provides complementary biophysical information including the chiral properties of the moieties (VCD and ROA), structural details of the fibril core deep UV resonance Raman (DUVRR), and spatially resolved surface chemical information (TERS). While a lot of work has investigated the fibrillation of insulin when the disulfide bonds remain intact, far less work has investigated the fibrillation of the independent A- and B-chains\textsuperscript{37-39}. The aim of this work is to provide further mechanistic detail of the fibrillation of the independent A- and B-chains of insulin. The two chains of insulin have amino acids specific to a particular chain. Specifically, the B-chain of insulin contains amino acids not present in the A-chain including three incidences of Phe. While the two chains remain connected through disulphide bonds this is irrelevant, however when investigating the two chains independently differences in specific amino acids are important. This is especially so with regards to aromatic amino acids such as Phe, which have been implicated in the self association process of fibrillation\textsuperscript{11}. Differences in the constituent could influence both the kinetics of fibril formation and morphology of the subsequent fibrils. In this study we used a combination of Raman spectroscopies to characterise the independent fibrillation of the A- and B-chains of insulin. The individual merit of each these techniques is demonstrated, but through combining these methods a more complete picture of the mysterious fibrillation process can be obtained.

6.3. Materials and Methods

6.3.1. Sample preparation

The sample preparation for the Raman investigation of the independent A- and B-chains of insulin (Sigma Aldrich, UK) as a function of pH is the following. The lyophilised samples were dissolved into double distilled H\textsubscript{2}O (Milli-Q) at a concentration of 10 mg ml\textsuperscript{-1}, from which the pHs of the respective solutions were measured with a digital pH meter with an accuracy of ± 0.10 pH units. The pH of the solution was then incrementally decreased using concentrated HCl from pH 4.90 to
pH 1.00 for the A-chain and pH 3.90 to pH 1.00 for the B-chain. At each adjusted pH Raman spectra (532 nm) were recorded for 15 minutes.

Samples of insulin bovine insulin, A-chain insulin and B-chain insulin (Sigma Aldrich, UK) were prepared in dimethyl sulphoxide-d$_6$ (DMSO-d$_6$, Sigma Aldrich). Samples were prepared at a concentration of 20 mg ml$^{-1}$, gentle shaking, and in the case of the B-chain slight warming (30°C) was required to facilitate solubilisation. Raman and ROA spectra were measured for each sample. A higher concentration of sample was required due to the strong Raman bands of the solvent (DMSO-d$_6$). Moreover, the sample was more readily soluble in DMSO-d$_6$ than water.

Fibrils formed from the A- and B-chains were prepared by dissolving the lyophilised material into double distilled (dd) H$_2$O and the pH was adjusted with HCl to a final concentration of 10 mg ml$^{-1}$ and pH 2.0. The samples were then heated at 30°C for the case of A-chain and 60°C for B-chain for up to 72 hours and 48 hours, respectively. Fibrils were also produced for B-chain at the same concentration but at pH ~3.0 and ~1.0. It has been previously shown the two independent chains of insulin form fibrils under different heating conditions$^{[37]}$. Therefore, this was the basis of choosing different fibrillation temperatures for the two samples.

6.3.2. Conventional Raman Spectroscopy 532 nm excitation.

All bulk solution Raman measurements at 532 nm were carried out on a ChiralRaman spectrometer (Biotools, Inc., Jupiter, FL, USA)$^{[40]}$ operating in back scattering mode under the following conditions: laser excitation 532 nm, laser power at the sample 0.6 W, spectral resolution ~7 cm$^{-1}$. Temperature controlled measurements were performed using a custom designed cell holder which was thermostatically controlled using a Julabo F12 heating / cooling unit. This allowed control of sample temperature to 0.1°C which was verified using a Jenway digital temperature meter. Samples of the A and B chains of insulin were monitored in situ under fibril forming conditions where Raman spectra were collected over one minute intervals (comprising 32 scans) for the duration of the fibrillation process monitored.

6.3.3. Raman optical activity (ROA).

The Raman measurements ($I_R + I_L$) and ROA ($I_R - I_L$) of insulin samples in DMSO-d$_6$ were measured on the previously described scattered circular polarised (SCP) Biotools ChiralRaman instrument$^{[40]}$. Samples were prepared as described and loaded into quartz micro-fluorescence cells. Measurements were carried out at ambient temperature under the following conditions: laser power at the sample 0.4 W, excitation 532 nm, spectral resolution ~7 cm$^{-1}$ acquisition times varied from ~24–48 h.

6.3.4. DUVRR.

DUVRR spectra were acquired for the monomeric A- and B-chains of insulin (10 mg ml$^{-1}$ at pH 2 in H$_2$O) as well as the fibrillar species. Measurements were acquired on a home built Raman instrument with deep UV excitation capabilities, which has been previously described$^{[41]}$. Samples
were loaded into a suprasil NMR tube (~50 μl) which rotates during the measurement, while the following laser conditions were used: excitation wavelength 199.7 nm and power of 1mW. Hydrogen-deuterium exchange (HX) can be a useful tool in the characterisation of protein structure, especially when combined with vibrational spectroscopy\cite{33, 42, 43}. In proteins the carbon bound hydrogen does not exchange, whereas NH, O, N and S bound protons can readily exchange\cite{42}. The hydrophobic core of a protein or strongly hydrogen bonded secondary structure i.e. cross-β structure of a fibril the HX rate is greatly diminished. Owing to this, the signature of the cross-β core can be identified through it remaining unchanged in comparison to the water accessible features of overall structure.

Hydrogen-deuterium exchange experiments samples were centrifuged at 16,000 rcf for 30 minutes. Following this, the fibril containing precipitate was resuspended into D₂O prior to DUVRR measurements.

### 6.3.5. Atomic force microscopy (AFM).

AFM measurements of the fibrils from A and B chains of insulin at pH 2.0 were measured on a Nanowizard instrument (JPK, Germany). Samples were diluted from 10 mg ml⁻¹ by 1:100 and deposited onto glass slides and washed with 1 ml of ddH₂O dried under argon. While the AFM measurements of B-chain fibrils formed at pH 3.0 and 1.0 were diluted 1:200 and 1:100 respectively and deposited onto freshly cleaved mica. The solution was allowed to settle for 2–3 minutes before the excess solution was washed away, and the sample was dried under a stream of nitrogen. Measurements were carried out on a MFP-3D™ Bio Asylum Research microscope (Asylum Research, CA, USA) using Olympus 160TS tips.

### 6.3.6. TERS.

The TERS experiments involving the A and B chain insulin fibrils were performed on a glass slide which had 1:100 diluted fibril solutions deposited on them. The glass slide was then washed with 1 ml⁻¹ of ddH₂O and dried under argon before further use. TERS experiments were performed as previously described\cite{44} on the following instrumental setup: AFM (JPK, Germany) with an inverted Raman microscope (Horiba, France). TERS spectra were measured at 530 nm with each spectrum the result of a 10s accumulation. TERS spectra were measured in a number of grids across the fibril surface in which the step size varied from 0.5 nm to 2 nm, depending on the fibril and dimensions of the grid in question. To confirm the tip was not contaminated during the measurements a reference spectrum of the glass next to the fibril was taken after each measurement.
6.3.7. Data processing

Raman spectra was processed using Matlab 2010 and a documented in-house toolbox\cite{45}(Chapter 4), while Matlab's bioinformatics toolbox was used for calculating FWHH. Conventional Raman spectra were baseline corrected using the method developed by Eilers et al\cite{46} and normalised to an invariant band, in addition the Raman spectra of water was subtracted. For the Raman spectra recorded in DMSO-$d_6$, the solvent was subtracted in addition to baseline procedures and intensity normalisation for comparative purposes. The ROA spectra had aberrant spikes removed and were baseline corrected using a median filter, before smoothing with a second level Savitzky-Golay filter. Spectral deconvolution was carried out in OriginPro 8.5.0 using its inbuilt peak fitting and deconvolution tool. Second derivatives of the data were used to pick peak locations where Gaussian peaks were iteratively fitted until convergence criteria were met. Deep UV Raman spectra were processed using GRAMS /AI (7.01) software (Thermo Fischer Scientific). Analysis and presentation of the AFM data from both instruments was performed using Gwyddion software\cite{47}. Finally, the raw spectra for TERS data are presented, excluding the use of an in-house aberrant spike removal routine. Cosmic rays were identified through the first derivative spectra intersecting a chosen threshold and being replaced by the median value of the adjacent data points.

6.3.8. Principal component analysis (PCA) and 2D-Correlation

Principal component analysis (PCA) is a multivariate technique in which the dimensionality of a data set is reduced into a series of principal components (PCs). Each PC accounts for a decreasing proportion of the variance of the original data set\cite{28, 48}. A scores plot displays each spectrum as a single point on a 2D Axis, allowing similarities and differences between spectra to be observed. In a loadings plot the variance of a specific PC is observed in terms of the original variables that constitute it, i.e. the wavenumber values. Spectra were processed as detailed above (6.3.7), with the addition of being mean-centered prior to PCA. PCA was applied to the Raman spectra A-chain and B-chain insulin that was perturbed as a function of pH. The data from PCA analysis is displayed in the form of scores and loading plots. Moreover, PCA was applied to the samples of insulin in DMSO, with the data displayed in the same way. Generalised 2D-correlation is a cross correlation approach\cite{49} where the cross-correlated data is visualised over two independent wavenumber, thereby spreading the data over an additional dimension. This methodology has been further adapted to include moving windows (MW)\cite{50}, and perturbation correlation moving windows (PCMW)\cite{51}. In PCMW the correlation between changes in spectral intensity and the perturbation are visualised. This can help identify complex changes in spectra over a perturbation range. PCMW was applied to the Raman spectra of A- and B-chain insulin that was monitored as a function of pH. The Raman Spectra of A- and B-chains of Insulin were interpolated so that pH intervals were evenly spaced, and smoothed to remove noise.
6.4. Results and Discussion

6.4.1. A Raman pH study of A and B chains of Insulin

To understand better the structural properties of the independent chains of insulin, their response to changes in pH was monitored in bulk solution using conventional Raman spectroscopy. To elucidate the most significant changes, and their sequence of change, principal component analysis (PCA) and perturbation correlation moving windows (PCMW) were used in combination.

A table of the Raman band assignments for insulin and its independent chains are contained in the Supplementary Information (Table S6.1). For the A-chain of insulin, it is clear that there is a complex relationship between pH and the measured Raman spectra (Figure 6.1), as a number of bands change in shape, intensity and position. The most intense band in the spectra occurs at 1048 cm$^{-1}$. This band is assigned to sulphonate groups present on each Cys residue of the A-chain (four in total)$^{[52, 53]}$ acting as a capping molecule on the sulphurs preventing disulphide bond formation$^{[54]}$. The PCA analysis (Figure 6.2A) of the A-chain of insulin uses the first and third coefficients (PCs 1 and 3), as this showed a better separation with respect to pH, while the loadings of PC1, which accounts for most of the variance is shown Figure 6.2B. The PCA plot shows three main groupings of the data with respect to pH, which are observed for pH 4.89–4.03, pH 3.08–1.61 and pH 1.34–1.04. The loading plot for PC1 (Figure 6.2B) shows a number of features that relate to structurally induced changes of the A-chain of insulin as a function of pH. Bands which can be assigned to Tyr show variances in a number of regions (644, 835, 857, 1170 and 1620 cm$^{-1}$). Changes in the intensity of the bands at ~830 and ~850 cm$^{-1}$ are informative of changes in the hydrogen bonding of the phenoxyl group of Tyr$^{[55, 56]}$, which acts as a probe for changes in environment of the Tyr residues. In the Raman spectra the ratio of $I_{850}/I_{830}$ is observed to change from 1.06 at the lowest pH to 0.94 at pH 4.89. This indicates that as the pH increases the Tyr residues become buried, and less solvent exposed. Key regions of the spectrum are informative of the secondary structure including the C-C stretch (900–1000 cm$^{-1}$), C-N stretch (1050–1150 cm$^{-1}$) and the amide I and III regions (1620–1700 and 1230–1300 cm$^{-1}$, respectively). From the loadings plot there are changes in all of these regions. Assignable changes include those of $\alpha$-helix markers at 930 and 1340 cm$^{-1}$, and bands at 950 cm$^{-1}$ and 1260 cm$^{-1}$ which have been observed in previous studies of protein structural disorder$^{[21, 56]}$. Correlating these changes with the PCMW plot (Figure 6.3) it is apparent that $\alpha$-helix content increases with increasing pH. There are also changes for a number of other peaks.

The most intense changes occur below pH 2, however changes are observed throughout the entire perturbation range. This indicates that the structural changes occurring are doing so in more than one transition. Moreover, this correlates to the observations in the PCA clustering. There are bands which lose intensity upon increasing pH including a band at 1712 cm$^{-1}$, which arises from C=O stretching$^{[21]}$, which is sensitive to the protonation state. Three further bands at 789, 1468 and 1607 cm$^{-1}$ show a decrease in intensity with increase pH, however definitive assignments for these bands are not yet available. An interesting observation in the Raman spectra of A-chain of insulin
when measured at a higher pH, is a Raman band at 1250 cm\(^{-1}\). This band is indicative of poly-
proline type II helix (PPII)\(^{[21]}\), indicating that at higher pH both \(\alpha\)-helix and PPII are predominant in
its structure, which is not observed in fully intact insulin\(^{[57]}\). This difference could be the result of
losing the internal disulphide bond usually observed between Cys6 and Cys11, which might act to
stabilise the helix between the first and eighth residues \(^{[57]}\).

Raman spectra of the B-chain of insulin were measured at thirteen pHs ranging from pH 1.06 to
pH 3.91 in increments \(\sim\)0.2 pH units. From the Raman spectra (Figure 6. 4), it is evident that
changes in pH lead to changes in the Raman spectra of the B-chain of insulin with a number of
bands across the spectral wind showing differences in both intensity and position. The scores plot
of PC1 vs. PC2 (Figure 6. 5A) shows a tight clustering of the repeat measurements at the same
pH, yet separates each individual pH into their own clusters. The only two pH measurements that
nearly overlap are observed at pH 2.50 and pH 2.80.

A PCA loadings plot relates the importance of a variable i.e. wavenumber to the variance of a
principal component. For PC1 (Figure 6. 5B), which accounts for most of the variance, is rich in
information with regard to the changes observed. A number of these changes relate to bands
attributed to the sidechain of the two Phe residues including the bands at 630, 1010, 1586 and
1605 cm\(^{-1}\)\(^{[56]}\). Further identifiable changes in side chain environment include those for the two Tyr
residues, with changes in bands at 830, 850, 1170 and 1620 cm\(^{-1}\). The changes observed for the
Tys are far smaller than those observed for the Phes, however, as previously described the
changes in the intensity of the bands at 830 and 850 cm\(^{-1}\) are informative about changes in the
hydrogen bonding of Tyr\(^{[55, 56]}\). The ratio of these intensities \(I_{850}/I_{830}\) is 1.12 at pH 1.06, while it
decreases to 0.74 at pH 3.91. This indicates that decreasing the pH significantly increases the
exposure to solvent of the Tyr residues in B-chain insulin, further indicating changes in the
secondary structure of B-chain insulin.

Further assignable changes observed in the loadings plot include those of C-C stretching (890, 924 and 960 cm\(^{-1}\)) and C-N stretching modes (1115, 1130 and 1150 cm\(^{-1}\)), respectively. These
particular marker bands are sensitive to secondary structure, which further indicates that as pH
changes the structure of B-chain insulin also changes\(^{[56]}\). Additional changes in secondary structure
are observed at 1680 cm\(^{-1}\), indicating possible differences in the propensity of intramolecular \(\beta\)-
sheet / \(\beta\)-strand phi/psi conformations. Finally, there are two significant changes in conjunction with
the pH at \(\sim 780\) and \(\sim 1460\) cm\(^{-1}\); while an assignment could not be made for the former, the latter
might arise from changes in methyl deformations. If this assignment is correct it would indicate a
significant reordering of the aliphatic residues of the B-chain.

The PCMW plot shown in Figure 6. 6 shows that the most intense spectral changes are
occurring between pH 1.3 to pH 2.5. However, further changes are observed around pH 3.5,
indicating that these transitions are not continuous and occur over two stages. As pH increases the
intensity of the Phe and \(\beta\)-sheet marker increase, while other bands show a decrease in intensity.
The data from the PCMW correlates well with that of the loadings plot (Figure 6. 5B), with respect
to which vibrational modes are undergoing the greatest change as a function of pH.
In intact insulin the B-chain is responsible for dimerisation, especially the C-terminal region, including Tyr26\[^{56}\]. The dimerisation of native insulin is shown (Supplementary Information Figure S6. 1)

Therefore, an interesting postulation would be that the B-chain of insulin can still dimerise, which is observed at the higher pH, through the increase in β-strand. However, further studies would be require to confirm this.

It is evident that the structure of both the A- and B-chains of insulin are sensitive to pH, especially the A-chain. This exceptional sensitivity to pH is not observed in the structure of fully intact insulin (supplementary information Figure S6. 2). This suggests that the disulphide bonds play a stabilising role to external perturbations such as pH to the structure, especially so for the A chain. Moreover, when the two chains are no longer disulphide linked there is a significant change in the structure adopted by the two chains compared to that of intact insulin (data not shown). It is likely that residues such as Tyr16 of the β-chain, which usually resides in the core of the protein, become solvent exposed in the single chain\[^{56}\].

**6.4.2. Raman and ROA measurements of intact insulin and its A- and B-chains in the non aqueous environment DMSO-d\(_6\)**

The solubility of the A- and B-chain of insulin in water was found to be too low for a ROA measurement. Moreover, the rate of aggregation for the A- and B-chains was shorter than the length of time required for a ROA measurement (especially with the samples having a low concentration), meaning an ROA measurements of the monomers were not possible in water. However, previous work involving insulin in the non-aqueous environment of DMSO-d\(_6\) has shown that intact insulin (where the A and B chains are still linked by disulphide bonds) and insulin fibrils, consist of the same structure that is rich in PPII type helix as determined with vibrational spectroscopies\[^{34}\]. Moreover, the structure or population of structures are predicted through molecular dynamic (MD) studies to be rich in PPII type helix\[^{34}\], and is suggested to represent the stabilised intermediate structure of fibril formation. Due to the strong interaction between DMSO-d\(_6\) and the protein of interest, the conformation is prevented from immediate aggregation\[^{34}\] and remains stable. Therefore, studies in DMSO-d\(_6\) provide an opportunity for studying stabilised conformations of aggregating proteins.

The A and B chains of insulin dissolved in DMSO-d\(_6\) appeared to be stable with no noticeable differences in their Raman spectra as a function of time. Figure 6. 7A shows the comparison of the Raman spectra of the A and B chains of insulin and of fully intact insulin when dissolved in DMSO-d\(_6\). While the three spectra share similarities in the Raman profile, there are however noticeable differences in a number of regions of the spectra; which is reflected in the complete separation of the three samples using PCA, shown in Supplementary Information (Figure S6. 3). All three spectra show a similar profile between 1150 to 1300 cm\(^{-1}\) with peaks at 1180, 1210, 1236 and 1272 cm\(^{-1}\). Examining the region between 1300–1350 cm\(^{-1}\) for the B-chain of insulin and for intact insulin, an identical profile is observed. In comparison the A-chain appears at a slightly lower intensity and the two peaks at 1320 and 1343 cm\(^{-1}\) are both downshifted by 5–10 cm\(^{-1}\). However, the main
differences observed between the two sets of spectra corresponding to the A and B chain occurs between 1600 and 1700 cm\(^{-1}\). This difference is also reflected in the loadings plot of PC1 (see Supplementary Information Figure S6. 3). The main contribution to Raman bands between 1580 and 1620 cm\(^{-1}\) are attributed to the aromatic side chains (Tyr and Phe). Both the A- and B-chains contain the same number of Tyr residues (two), therefore it is interesting to observe a significant difference in the intensity of the band at 1620 cm\(^{-1}\). This indicates that the Tyr residues of A- and B-chain insulin are in significantly different environments, further demonstrating the differences in the secondary structures of the two chains.

While the side chain environment can be used to infer structural differences between protein species, the amide I region provides the clearest confirmation of this. On first appearance all three samples contained a similar Raman profile of the amide I region, with the profile intensity centred at 1688 cm\(^{-1}\). As has been previously reported, this profile does not correspond to either fibril or native insulin amide I band assignments\(^{[34]}\) with respect to either position or shape.

Deconvolution of the amide I region for the three samples (see Supplementary Information Figure S6. 4) identifies three possible bands contributing in the amide I region, centred around ~1662, ~1680 and ~1690 cm\(^{-1}\) respectively. While the percentage fit for the band at 1662 cm\(^{-1}\) (which is usually the dominant α-helix structural marker band in fully intact insulin\(^{[56]}\)) is similar for all three, there are bigger differences for the other contributing bands. It is interesting to note that in the intact insulin, the percentage attributed to the band at 1680 vs 1690 cm\(^{-1}\) is similar (45% vs 39%), while for the A and B chain insulin these two values are not similar. In the A-chain, the contribution is 28% at 1680 cm\(^{-1}\) and 50% at 1690 cm\(^{-1}\), while the B-chain shows the reverse, 54% at 1680 cm\(^{-1}\) and 28% 1690 cm\(^{-1}\). These differences may represent gross differences in the structural content of PPII type helix, β-turn and β-strand between the two chains of insulin when in the non aqueous environment DMSO-d\(_6\). However, difficulties in assigning these bands from the literature and the reliability of spectrum deconvolution means that caution is should be exercised in interpreting this observation.

Comparison of the three measured ROA spectra for insulin and the independent A and B chains (Figure 6. 7B) in DMSO-d\(_6\), reveals that they all share similar features. All three spectra contain a +ve/-ve couplet at 1315 /1350 cm\(^{-1}\) and the presence of a purely positive band at ~1690 cm\(^{-1}\). The couplet, this has been identified as a marker for PPII helical structure\(^{[30, 34]}\). In the ROA spectrum of the intact insulin sample this couplet appears slightly broader, with a slight shoulder at 1290 cm\(^{-1}\), while a shoulder is evident for the A-chain that is downshifted 10 cm\(^{-1}\) from this position. Due to the poor signal to noise ratio in the ROA spectrum of B-chain insulin, a reliable comparison with the other spectra is not possible for this region. The positive band in the amide I region of the ROA spectra is very similar in full insulin and the B-chain, in both intensity and shape, while for the A-chain profile it appears to be less intense and broader. An additional difference between the ROA spectra can be observed at 1445 cm\(^{-1}\), for which intact insulin and the B-chain show a significant negative peak, while it is diminished in the A-chain insulin spectrum. It is clear from the Raman and ROA spectra that the A- and B-chains of insulin in this non aqueous stabilised
“prefibrillar” structure adopt different conformations from each other and, of the two, the B-chain insulin appears to be more similar to that of intact insulin.

6.4.3. In situ Raman measurements of A and B chain fibril formation

Previous work has shown that insulin fibrils supramolecular chirality to be sensitive to the initial pH\(^{[58]}\). Specifically it was shown that initiation of fibrils above ~pH2 to be significantly different to those formed below that pH. Therefore, the in situ formation of fibril formation in water was monitored for the A-chain at pH 2 (sample limitations), and for the B-chain at three pHs (3.0, 2.0 and 1.0). By choosing three pHs one above and one below this transition point around pH 2 should show if this type of phenomenon is observed for the B-chain of insulin.

The literature regarding the independent fibrillation of the A and B chains of insulin is mixed\(^{[37,54,59]}\), with different conditions being recommended. In this case the B-chain was observed to be more liable to form fibrils, with the A-chain being less consistent in its fibril forming propensity.

The Raman spectra gathered during the fibrillation of the A-chain of insulin are shown in Figure 6. 8A. To aid in the analysis of the changes the full width half height (FWHH) profile of the amide I region is shown (Figure 6. 8B). FWHH of the amide I region is used because this region is a good marker of secondary structure in proteins, with different types of structure having differing band profiles. Therefore, changes in the FWHH of the region can be reflective of changes in the samples secondary structure.

From Figure 6. 8: B it appears there was an initial lag, with no evident changes in the Raman spectra, followed by a consistent decrease in the FWHH. Therefore, only the first and last spectra are used in the discussion of the observed changes (Figure 6. 8: C). There are a number of bands which appear to change through the fibrillation process, however one appears invariant. This band at 1048 cm\(^{-1}\) is assigned to the sulphonic acid group capping the Cys residues. The two Tyr marker bands, the importance of which has previously been discussed, at ~835 and ~855 cm\(^{-1}\) show variance through the fibrillation process. Specifically, the ratio of \(I_{855}/I_{835}\) changes from the initial value of 1.03 to 1.1. This indicates a small increase in solvent exposure during fibrillation. Other changes in the spectra include the decrease in intensity of the band at 896 cm\(^{-1}\). This has been assigned to \(\nu\)(C-C) modes\(^{[56]}\), further changes are seen in the shift from a broad feature positioned ~966 cm\(^{-1}\) to a peak centred at 950 cm\(^{-1}\) containing a shoulder at 960 cm\(^{-1}\). A band at ~1010 cm\(^{-1}\) also emerges as fibrillation progresses, which also represents changes in \(\nu\)(C-C) modes. Changes in \(\nu\)(C-N) modes are observed at 1130 cm\(^{-1}\), with an increase in the intensity of the band as fibrillation progresses. Finally, there are large increases in the marker bands for \(\beta\)-sheet at 1230 and 1675 cm\(^{-1}\). By the end of the monitoring process these two bands dominate the amide III and I regions, respectively. These secondary structure changes are characteristic of those that have been previously observed during the fibrillation of intact insulin\(^{[28,56]}\). However, slight differences are observed with respect to overall changes observed in the marker bands for the Tyr residues\(^{[56]}\).

The in situ fibrillation of the B-chain of insulin was monitored at three separate pHs (3.00, 2.00 and1.00). At pH 3 the Raman spectra show a number of changes as a function of time during fibrillation (Figure 6. 9A). From the FWHH analysis of the amide I region (Figure 6. 9B), the
changes appear to occur in one transition, with no apparent lag in fibrillation. Therefore, only the first and last spectra (Figure 6. 9C) will be utilised in the following discussion of the overall spectral changes of fibrillation. Changes are observed in the Tyr marker bands at 835 and 855 cm\(^{-1}\), where the \( I_{855} / I_{835} \) ratio changes from 0.91 to 1.35. This indicates an increase in solvent exposure during the fibrillation process. As fibrillation progresses there is a shift in the band originally observed at 1043 cm\(^{-1}\) by 10 cm\(^{-1}\), additionally the relative intensity also increases with respect to the band at 1006 cm\(^{-1}\) originating from Phe. The two main \( \beta \)-sheet marker bands at 1230 and 1675 cm\(^{-1}\), show a large increase in intensity with time. The amide III region becoming dominated by the band at 1230 cm\(^{-1}\) and other spectral features being reduced; while the amide I feature starts off as a small broad band centred at the same position 1675 cm\(^{-1}\) but becomes narrower and eventually dominates the spectrum.

A number of spectral changes are observed accompanying the fibrillation of B-chain insulin at pH 2 (Figure 6. 10A), with a lag phase being indicated in the FWHH analysis of the amide I region (Figure 6. 10B). A lag period is expected due to the lower temperature of the experiment with respect to fibrillation at both pH 3 and 1. The FWHH analysis indicated two initial changes in the amide I region, however closer indication of this region (between time points I and II in Figure 6. 10C) shows quite a minor change in intensity of this region, with no obvious changes in the other regions of the spectra. Therefore, a comparison between representative spectra at the start of the indicated phase III (Figure 6.10: B) and the final spectrum are discussed (Figure 6. 10D).

At pH 2 slight only small changes in the initial Raman profile of the B-chain are observable, compared to pH3. These changes include the emergence of a small peak at 786 cm\(^{-1}\), the splitting of the peaks in the region between 1030 and 1050 cm\(^{-1}\) and emergence of a band at 1468 cm\(^{-1}\). Upon fibrillation these bands are either lost, in the case of the band at 786 and 1468 cm\(^{-1}\), or undergo change, such as for, the peak initially at 1035 cm\(^{-1}\) which increases in intensity and downshifts slightly in position by 3 cm\(^{-1}\) relative to its neighbouring band at 1046 cm\(^{-1}\) which appears invariant. As has been seen previously there is a change in the relative intensities of the Tyr marker bands at 835 and 855 cm\(^{-1}\). As fibrillation progresses the ratio of \( I_{855} / I_{835} \) changes from 0.76 to 1.01. This is indicating an increase in solvent exposure during the fibrillation process. Finally, as fibrillation progresses \( \beta \)-sheet marker bands at 1230 and 1675 cm\(^{-1}\) in the amide III and I regions, respectively emerge.

As can be seen from the Raman spectra of B-chain of insulin at pH 1 there is a complex sequence of changes upon fibril formation (Figure 6. 11A). The FWHH analysis on the amide I region (Figure 6. 11B) helps to focus on time periods of interest i.e. transitions in the spectra. This analysis shows that upon heating there are immediate structural changes, which is followed by a lag period. Looking at the first 10 spectra (Figure 6. 11C) these changes include the loss of intensity of the \( \alpha \)-helix marker band at 1344 cm\(^{-1}\), and relative increase in \( \beta \)-sheet marker bands at 1230 cm\(^{-1}\) and 1673 cm\(^{-1}\). As shown (Figure 6. 11D) following the initial changes in the spectra of B-chain insulin at pH1, there is a lag phase where no changes in the spectra are observed. Following this lag, as shown Figure 6. 11E, there are a number of further changes that occur in the spectra. These include further increases in the intensity of \( \beta \)-sheet marker bands at 1230 cm\(^{-1}\) and
1677 cm$^{-1}$. There is a slight shift in the position of this amide I $\beta$-sheet marker band, when compared with the B-chain fibrillation at higher pHs. However, this shift is within the resolution limits of the instrument, therefore if this represents a significant difference between the types of $\beta$-sheet formed at these different pHs cannot be confirmed.

The two Tyr marker bands at 835 and 854 cm$^{-1}$ can be observed to change in their relative intensities, with the $I_{855}/I_{835}$ ratio changing from 1.38 to 0.98 during this phase. The latter ratio remains the same until the end of the experiment. The observed change show that the Tyr sidechains are becoming buried and less solvent exposed through the fibrillation process. This is interesting, as it is the reverse as to what was observed at pH 3 of B-chain insulin, indicating differences in the fibrillation process at the two pHs. Between 900–1000 cm$^{-1}$ a positional shift (~3 cm$^{-1}$) of the band at 930 cm$^{-1}$ occurs and a small reduction in intensity. Furthermore, a band centred at 960 cm$^{-1}$ assigned to disordered structure appears. The bands assigned to Phe also show changes, with the relative intensity of the band at 1006 cm$^{-1}$ increasing while the band at ~1040 cm$^{-1}$ decreases.

Interestingly, two bands that are observed to appear as the pH was lowered, in the B-chain of insulin, at 786 cm$^{-1}$ and 1463 cm$^{-1}$ both decrease in intensity during fibrillation of the sample. This suggests that these two bands correspond to structural elements within the B-chain, rather than monitoring the protonation event of a side chain or the N and C termini of the molecule. Additionally, it might imply that there are structural elements in the B-chain whose loss is a prerequisite for the fibrillation of insulin to occur.

6.5.4. DUVRR of A and B chain insulin

The structures of the A and B chains of insulin were monitored with DUVRR (199.7 nm) at pH 2.00 before and after fibrillation. However, due to the low yield of fibrils formed for the A-chain, it was not possible to carry out H/D exchange experiments. Therefore, this experiment was only carried out for the B-chain of insulin. Additionally, the A-chain of insulin appeared to be sensitive to photo degradation, especially with respect to the aromatic bands in both monomer and fibril form; however, this behaviour was not observed for the B-chain (data not shown).

Comparing the DUVRR spectra of the A and B chain of insulin (Figure 6. 12) differences are apparent between the two spectra. As expected, the Phe band (which the A-chain lacks) is present in the B-chain at ~1000 cm$^{-1}$, and further differences in the aromatic bands(A-chain Tyr, B-chain Phe and Tyr) are observed at 1620 cm$^{-1}$; where this band is more intense for the A-chain spectrum relative to the band at 1610 cm$^{-1}$, while for the B-chain these two bands are of a similar intensity.

Differences are also apparent between the secondary structures of the A- and B-chain. In the Amide III regions of these spectra, the central intensity for both chains is ~1260 cm$^{-1}$, indicating significant disorder, however both also contain a shoulder at ~1240 indicative of PPI$^\parallel[60]$. In the amide I region the band profile for the A-chain appears broader and is centred at 1668 cm$^{-1}$; while the B-chains amide I profile appears to be sharper and is centred at 1674 cm$^{-1}$. This shows agreement with the measurements of the Raman spectra at 532 nm. Further supporting the
analysis that there are significant conformational differences between the two chains, with β-sheet appearing more prominent in the B-chain of insulin.

The low yield of A chain fibrils is reflected in the increased noise in the corresponding fibril spectrum, and more discreet differences observed between the initial DUVRR spectrum of A-chain and the fibril spectrum (Figure 6. 13). There is a decrease in the intensity of the bands at 1610 and 1620 cm⁻¹ from aromatic residues upon fibrillation. Further changes are observed in the amide I and III regions of the spectrum. Whereas in the amide III region there is a loss of the shoulder at ~1240 cm⁻¹ upon fibril formation, the amide I region shows an increase in intensity of the β-sheet marker band at 1674 cm⁻¹.

The changes observed for the fibrillation of the B-chain at pH 2.00, observed with DUVRR (Figure 6. 14A) show similar changes to those for the A-chain. There are changes in the intensity of the aromatic residues, showing a marked decrease at ~1600 cm⁻¹; furthermore, there are changes in the secondary structure, now showing classic hallmarks of a fibrillated protein i.e. intense β-sheet band at 1675 cm⁻¹. Characterisation of the B-chain fibril core was achieved through hydrogen/deuterium exchange, which showed that the fibrillated B-chain structure has elements that are accessible to exchange, causing a loss of the amide III region due to the loss in coupling between N-H and C-H vibrations. There is also a down shift for the bands in the amide II region (amide II')[33]. Therefore, the amide III bands that remain are through N-H groups which are present in the fibril core and therefore protected from exchanging. The DUVRR spectrum of the fibril core of the B-chain (Figure 6. 14B) shows that the core shares a classic β-sheet core, as has been seen in previous fibril studies using this approach[33].

6.4.5. AFM measurement of A and B chain insulin fibrils

AFM enables topographic information of a sample to be determined, providing the dimensions of a structure on a nanometre scale. This approach is routinely used to characterise the morphology of protein fibrils[11, 37]. Averages of the width and height were calculated from the average of several measurements per fibril and over several fibrils as detailed in the supplementary information (Figure S6.5–9). The height of fibrils was measured relative to the background, while measurements across the fibrils were taken for the width.

AFM measurements were made on the A-chain fibrils that were produced at pH 2, and on the B-chain fibrils produced at pH 3, 2 and 1. As has been demonstrated, pH has an effect on the Raman spectra of the independent chains; therefore, a sound hypothesis would be this might be reflected in the structures of the fibrils when produced under different conditions, as is seen in fully intact insulin[14, 61, 62].

As the fibrils formed at pH 2 were analysed on a different AFM instrument, and were the fibrils for which TERS was performed on, they are discussed separately. Examining the fibrils formed from B-chain insulin at pH 3 (Figure 6. 15A–D), consistent fibrils are produced, with an average width of 48.6 nm ± 3.8 nm and a height of 2.5 nm ± 0.6 nm (Supplementary Information Figure S6. 5). In comparison, the AFM images of the B-chain fibrils produced at pH 1 (Figure 6. 15E–F), are dominated by a heterogeneous ensemble of structures. These structures vary between an ellipsoid
deposits 94 nm ±20 nm in height and 558 nm ±68 nm width (Supplementary Information Figure S6. 6), and fragmented fibrils rarely longer than a few μm in length with widths 48 nm ± 10 nm and heights of 2 nm ± 0.2 nm (Supplementary Information Figure S6. 7). The formation of ellipsoid amylloid deposits (spherulites) in insulin fibrillation has been shown before[^61, 62]. However, this phenomenon has not been previously reported for the independent B-chain of insulin[^37, 54]. While previous work has investigated the conditions that affect the distribution of spherulite and fibril assembly (concentration, pH and salt concentration), for the B-chain of insulin pH clearly plays a critical role in the control of this process.

The AFM data for the A-chain of insulin (Figure 6. 16A–F) formed at pH 2 show an ensemble of fibril type morphologies, which likely includes a mixture of both protofibrils and mature fibrils. This mixture of fibrils shows an extraordinary variance, in both their height (4–12 nm) and width (40–200 nm) (Supplementary Information Figure S6. 9), which is quite different to that seen for the B-chain fibrils. The range of morphologies for the A-chain of insulin has been discussed previously[^37], in which Devlin et al. also observed a diverse set of fibrils at pH 2. Furthermore, in accordance with their observations a periodicity is seen in some of the A-chain fibrils, which can be clearly observed in Figure 6. 16D. In this example the fibril shows a periodicity of ~90 nm (supplementary information Figure S6. 10). Compared to the A-chain fibrils, those of the B-chain at pH 2 are more homogeneous. Furthermore, they share similarities with those formed at pH 3 by the B-chain and those reported for full chain insulin[^37] with regards to their dimensions. However, variance in morphology is observed, and it is quite likely for this particular sample both proto-fibrils, and more mature fibrils are present. This conclusion is based upon the width profiles of the fibrils observed.

### 6.4.6 TERS of A and B-chain insulin fibrils

TERS has been shown to be particularly suited to the study of fibrils, with previous studies investigating fully intact insulin[^35, 36]. These studies were able to probe the secondary structure along the fibril surface, and the prevalence of different amino acids at the surface. As is shown (Supplementary Information Figure S6. 11), there are a number of differences in the sequence of amino acids between the A and B chains. TERS measurements on the fibrils formed by the independent chains can provide spatially resolved chemical information about the fibril surface i.e. the location of specific amino acids and secondary structure present.

A table of TERS band assignments is given in the Supplementary information (Table S6. 2). Protein structure assignments of the amide I region were assigned on the following criteria, a band observed between 1635–1659 cm\(^{-1}\) was assigned to α-helix/disordered type structure; whereas, β-sheet was assigned between the region of 1660–1679 cm\(^{-1}\)[^35, 36]. If bands were observed in both regions that spectrum is described as having a mixed amide I contribution[^35, 36]. The percentage contribution is calculated from the total number of spectra showing a particular band assignment, divisible by the total number of spectra showing an active amide I region.

The TERS spectra which were gathered over two sections of the A-chain fibril surface are shown in Figure 6. 18A and B. These consecutive spectral sets (A and B) correspond to 10 and 18 nm stretches along the fibril axis, respectively (Figure 6. 18C), where each spectrum is measured.
0.5 nm apart. While this is a small sampling of the surface area of a whole fibril, as fibrils extend for several micro-meters at a time; as the polypeptide chains are orientated perpendicular to the sampling direction, it is representative of a number of peptide chains. While there are a number of bands which vary between spectra, there is a consistent amide I profile. In both, the amide I region is dominated by marker bands for α-helix/disorder. From the measurable spectra along this particular fibril (~80), 43% showed α-helix/disordered structure, while 14% showed β-sheet. This is informative in two ways. Firstly, it shows the consistency of α-helix along the fibril axis. Secondly, the suppression of the amide I region in SERS is a well documented phenomenon, which has been recently been explained in terms of bulky side chains, e.g. Phe and Tyr preventing the tip getting close enough to enhance the amide bond. Therefore, a consistent amide I response, and no observable Tyr marker bands suggest Tyr is not found in these α-helical/disordered segments on the surface.

For the A-chain the two Tyr residues are present closely together (Tyr14 and Tyr19) in the sequence, with Tyr 14 being in the SRE, while Tyr19 flanks the sequence. Therefore, it appears that the amino acids constituting the SRE are buried in the fibril core, especially at pH 2, while this example of surface α-helix might consist of residues (1–8) which are helical in native insulin. It was previously reported for intact insulin fibrils, that Tyr showed a far higher propensity to be found within β-sheet compared to α-helix. TERS spectra were gathered from another 10 fibrils (479 spectra in total), from which the breakdown of secondary structure from the amide I region is 8% β-sheet, 43% α-helix and 49% a mixture of both α-helix and β-sheet.

The two sets of TERS spectra displayed for the B-chain of insulin (Figure 6. 19A–B), correspond to two regions of the same fibril. The first region A, corresponds to the first 20 measurements, where each spectrum is measured 2 nm apart, incorporating an area of ~80 nm². The second region (B) comprises 14 spectra, covering a slightly smaller surface area of the fibril. From these snapshots, it is clear the B-chain fibril exhibits a complex surface chemistry.

In the first part of the mapped regions are (A) identifiable side chains include His, Phe, and Cys, and amide I marker bands for secondary structure. The secondary structure bands predominantly showed α-helix/disordered structure, while in the second area (B), the TERS spectra again dominated by side chain bands, particularly those of Tyr and Pro. However, the amide I region is absent in this series of spectra.

An interesting observation for B-chain insulin fibrils was a series of spectra showing both α-helix in the Amide I region, and bands corresponding to Tyr and Phe (Supplementary Information Figure S6. 12). As has been shown previously, Tyr is predominantly observed to be present with β-sheet, which makes sense given three of the Tyr residues are located either in, or close to the proposed SRE. Therefore, the corresponding observation of Tyr observed with α-helix/disordered structure may originate from Tyr26 in the B-chain. Marker bands for Phe are also observed, and with both Phe and Tyr residues located close together in the sequence of the B-chain further supporting this hypothesis. It is also interesting to see that Tyr and Phe side chains appear to be localised together on the surface of the fibril, as this was not observed in the investigation of the fully intact insulin fibrils. This indicates that while the morphology of B-chain and full insulin fibrils
are similar, the surface chemistries are different. However, caution has to be taken in analysing these particular observations at this level of detail, given they are based upon a small number of spectra.

Overall, for the measured TERS spectra (491 in total), which were gathered from 6 different B-chain fibrils, the secondary structure from the amide I region is as follows: α-helix/disorder 66%, mixed (α-helix/disorder and β-sheet) 32% and β-sheet 2%. The presence of side chains when normalised by their occurrence in the sequence appears as might be expected, except for Phe; this particular side chain appeared less frequently than expected. Phe was observed in 3% of spectra, while accounting for 10% of the amino acids of the B-chain. This is interesting as it is the opposite of what was observed in the studies for intact insulin[36].

6.5. Conclusions

The work presented here provides a detailed investigation into the structural properties and fibrillogenesis of the independent A- and B-chains of insulin, using a multi disciplinary spectroscopic approach.

It was shown that the independent chains of insulin are highly sensitive to pH, undergoing structural changes not observed in intact insulin. For the A-chain of insulin there are two ionisable amino acids (Glu) in addition to the C-terminus liable to a change in protonation state under the pH range investigated. While for the B-chain of insulin contains two negatively charged residues (Asp + Glu) in addition to the C-terminus. These protonation changes might contribute to the structural changes observed. However, other aspects of the separation of the chains would be more likely the cause for such sensitivity to pH. One result of the chains being separated is exposure of the hydrophobic core, which in intact insulin is buried between the A- and B-chains of insulin (Figure S6. 1B). It is likely through this combination why such sensitivity is observed in both chains to pH, which is not observed in the intact insulin. Moreover, there is a loss of several salt bridges, hydrogen bonds and van der Waals interactions which will all influence the stability of the intact insulin monomer which is lost upon removal of the disulfide bonds. Furthermore, pH was shown to play a critical role in fibrillogenesis in the case of the B-chain. It was observed that the pH induced structural changes in the B-chain, were reflected in a disruption of the fibrillogenesis pathway i.e. from the classic formation of fibrils, to that of spherulites as is summarised in Figure 6. 20. At pH 2 and above regular fibrils were formed, which were similar in appearance to that of intact insulin[37, 63]. However, when the pH was lowered = pH 1 small fragmented fibrils and “spherulite” like particles formed. The B-chain of insulin contains five aromatic amino acids (3 × Phe and 2 × Tyr). The importance of aromatic residues in fibril formation has been previously demonstrated[1, 64, 65]. Moreover, there was a large change in the bands attributed to aromatic residues as a function of pH for the B-chain. Therefore, it is quite possible as was observed with Marek et al[65] for the islet amyloid peptide that the aromatic residues influence the morphology of the fibrils formed for the B-chain of insulin when pH is altered. The ROA spectra measured in DMSO-d_6 of the independent chains of insulin highlighted differences between A- and B-chain of insulin in this stabilised state. These differences in the ROA spectra might explain some of the morphological differences between the A- and B-chain fibrils when prepared at the same pH. Finally, TERS showed that
differences in the morphology of the fibrils are reflected in their surface chemistry. Specifically, there were noticeable differences between the A-chain and B-chain fibrils in terms of secondary structure content, and how side chains are distributed on the surface.

In summary, a combination of spectroscopic techniques have been used to probe the independent fibrillation of the A- and B-chains of insulin. This work demonstrates the benefits of a multi-disciplinary approach to studying protein fibrillogenesis, as it has shown detail from all stages of the fibrillogenesis pathway can be gathered. While this work is preliminary with respect to monitoring a model system, it shows the benefit of this kind of approach to studying the complex problem of fibrillogenesis.
6.5. References


Figure 6.1: Bulk solution Raman spectra measured at 532 nm of the A-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O) from pH 1.05 to 4.89.

Figure 6.2: PCA scores plot along PC1 and PC3 (A) comprising five Raman measurements at each pH for the A-chain of insulin (10 mg ml\(^{-1}\) in H\(_2\)O). Panel B shows PCA loadings plot along PC 1 from the PCA scores plot (A).
Figure 6.3: Perturbation correlation moving widows plot of the Raman spectra of A-chain insulin (10 mg ml⁻¹ in H₂O) as a function of increasing pH. The direction and intensity of spectral changes is indicated by the colour bar.

Figure 6.4: Bulk solution Raman spectra measured at 532 nm of the B-chain of insulin (10 mg ml⁻¹ in H₂O) from pH 1.06 to 3.91.
Figure 6. 5: PCA scores plot along PC1 and PC2 (A) comprising five Raman measurements at each pH for the B-chain of insulin (10 mg ml$^{-1}$ in H$_2$O). Panel B shows The PCA loadings plot along PC 1 from the PCA scores plot (panel A).

Figure 6. 6: Perturbation correlation moving widows plot of the Raman spectra of B-chain insulin (10 mg ml$^{-1}$ in H$_2$O) as a function of increasing pH of. Colour scale indicates the direction and intensity of spectral changes.
Figure 6.7: Raman spectra (532 nm excitation) of intact insulin, A-chain insulin and B-chain insulin all at 20 mg ml\(^{-1}\) dissolved in DMSO-d6 (A) and the corresponding ROA spectra (B).
Figure 6.8: The in situ Raman measurements (A) (532 nm excitation) of the A-chain of insulin (10 mg ml$^{-1}$ in H$_2$O HCl adjusted to pH 2.00) while being heated at 30 °C. The corresponding FWHH of the amide I region as a function of time is shown in panel B. The first and last spectra are shown in panel C.
Figure 6. 9: The in situ Raman measurements (A) (532 nm excitation) of the B-chain of insulin (10 mg ml$^{-1}$ in H$_2$O HCl adjusted to pH 3.00) while being heated at 60 °C. The corresponding FWHH of the amide I region as a function of time is shown in panel B. The first and last spectra are shown in panel C.
Figure 6. 10: The in situ Raman measurements (A) (532 nm excitation) of the B-chain of insulin (10 mg ml⁻¹ in H₂O HCl adjusted to pH 2.00) while being heated at 37.5 °C. The corresponding FWHH of the amide I region as a function of time (B), with transitions indicated (I–IV). The 25 spectra corresponding to the period between I and II as indicated in FWHH (B) are shown in panel C. Panel D shows a comparison of the spectrum from the start of phase (III) and the final spectrum at point IV.
Figure 6. 11: In situ Raman measurements (A) (532 nm) of the B-chain of insulin (10 mg ml$^{-1}$ in H$_2$O HCl adjusted to pH 1.00) while being heated at 60 °C. The corresponding FWHH of the amide I region as a function of time is shown in panel B. Smaller sub sets of the first 140 spectra are presented (C–E), where the first 10 spectra are shown (C), the 60 following spectra (D) and the final 70 spectra (E).
Figure 6. 12: Deep UV Raman spectra (199.7 nm excitation) of the A (red) and B (blue) chain monomers of insulin (10 mg ml\textsuperscript{-1} in H\textsubscript{2}O pH adjusted to pH 2 with HCl).

Figure 6. 13: Deep UV Raman spectra (199.7 nm excitation) of the initial A-chain insulin structure (10 mg ml\textsuperscript{-1} in H\textsubscript{2}O at pH 2.00) (blue) and its fibril after 24 hours incubation at 30°C (red).
Figure 6.14: The comparison of the DUVRR spectra of B-chain insulin (10 mg ml$^{-1}$ in H$_2$O at pH 2) initial structure (blue) and its fibril (red) (panel A) after incubation at 60°C for 24 hours. The spectral features of the fibril core of the B-chain fibril (panel B) are revealed through H/D exchange with D$_2$O and aromatic contribution subtraction.
Figure 6.15: AFM images of B-chain insulin fibrils formed through incubation at 65°C for 24 hours at two different pHs, ~pH 3 (panels A-D) and ~pH 1 (panels E–H). Samples diluted 100x and 200x respectively from an initial concentration of 10 mg ml⁻¹. Colour scale indicates height in nm, whereas white scale bar depicts distance.
Figure 6. 16: AFM images (panels A–F) of A-chain insulin fibrils formed through incubation at 30°C for 24 hours at pH 2. Samples were diluted 100x from an initial concentration of 10 mg ml⁻¹. Colour scale indicates height in nm, whereas white scale bar depicts distance.

Figure 6. 17: AFM images (A–C) of B-chain insulin fibrils through incubation at 60°C for 24 hours at pH 2. Samples were diluted 100x from an initial concentration of 10 mg ml⁻¹. Colour scale indicates height in nm, whereas white scale bar depicts distance.
Figure 6.18: Two sets of TERS spectra (A and B) measured along a fibril formed from the A-chain of insulin (C). Measurements were taken at 0.5 nm steps from each other, with panel A corresponding to the region 9.5 nm from the start point to 19.5 nm and panel B, the region 21 to 39 nm from the start point.
Figure 6.19: Two sets of TERS spectra (A and B) measured along a fibril formed from the B-chain of insulin (C). Measurements were taken at 2 nm steps from each other, with A corresponding to the first 20 measurements, 0 nm to 19.5 nm and panel B, from 21 to 39 nm. The data for panel A are displayed in a rotated 3D plot for better visualisation of the bands.
Figure 6. 20: Schematic model of the pH influence on the fibrillation of the B-Chain of insulin. B-chain of insulin (PDB: 1AI0)
6.6. Supplementary Information

Table S6. 1: Raman band assignments for Insulin and its independent chains\cite{56}.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Wavenumber (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-sheet (intramolecular)</td>
<td>1682</td>
</tr>
<tr>
<td>β-sheet (intermolecular)</td>
<td>1677</td>
</tr>
<tr>
<td>α-helix</td>
<td>1658</td>
</tr>
<tr>
<td>Disorder</td>
<td>1648 &amp; 1678</td>
</tr>
<tr>
<td>Tyr</td>
<td>1620</td>
</tr>
<tr>
<td>Phe &amp; Tyr</td>
<td>1605</td>
</tr>
<tr>
<td>Phe</td>
<td>1585</td>
</tr>
<tr>
<td>CH$_2$ deformation</td>
<td>1450</td>
</tr>
<tr>
<td>Symm CO$_2^-$ stretching</td>
<td>1418</td>
</tr>
<tr>
<td>α-helix</td>
<td>1340</td>
</tr>
<tr>
<td>CH deformation</td>
<td>1314</td>
</tr>
<tr>
<td>α-helix</td>
<td>1270,1284</td>
</tr>
<tr>
<td>Disorder</td>
<td>1257</td>
</tr>
<tr>
<td>β-sheet (intramolecular)</td>
<td>1236,1250</td>
</tr>
<tr>
<td>β-sheet (intermolecular)</td>
<td>1230</td>
</tr>
<tr>
<td>Tyr</td>
<td>1207</td>
</tr>
<tr>
<td>Phe</td>
<td>1182</td>
</tr>
<tr>
<td>Tyr</td>
<td>1173</td>
</tr>
<tr>
<td>v(C-N)</td>
<td>1110</td>
</tr>
<tr>
<td>β-sheet</td>
<td>1052</td>
</tr>
<tr>
<td>Phe</td>
<td>1000</td>
</tr>
<tr>
<td>disordered</td>
<td>955</td>
</tr>
<tr>
<td>α-helix</td>
<td>930</td>
</tr>
<tr>
<td>V(C-C)</td>
<td>895</td>
</tr>
<tr>
<td>Tyr Fermi Doublet</td>
<td>830 &amp; 855</td>
</tr>
<tr>
<td>v(C-S) of disulphide link</td>
<td>664</td>
</tr>
<tr>
<td>Tyr</td>
<td>643</td>
</tr>
<tr>
<td>Phe</td>
<td>620</td>
</tr>
<tr>
<td>Skeletal bending</td>
<td>560</td>
</tr>
<tr>
<td>v(S-S)</td>
<td>515</td>
</tr>
</tbody>
</table>
Table S6. 2: TERS band assignments\textsuperscript{[35, 36]}

<table>
<thead>
<tr>
<th>Structure</th>
<th>Wavenumber (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-sheet</td>
<td>1660-1680</td>
</tr>
<tr>
<td>α-helix /disorder</td>
<td>1620-1660</td>
</tr>
<tr>
<td>CH(_2), CH, CH(_3)</td>
<td>1475, 1457, 1372-</td>
</tr>
<tr>
<td></td>
<td>1377, 1355</td>
</tr>
<tr>
<td>COOH/COO(^-)</td>
<td>1400, 1687</td>
</tr>
<tr>
<td>NH(^{3+})/NH(^{3+})</td>
<td>1144</td>
</tr>
<tr>
<td>His</td>
<td>1187, 1335, 1495</td>
</tr>
<tr>
<td>Phe</td>
<td>1000</td>
</tr>
<tr>
<td>Pro</td>
<td>899-902</td>
</tr>
<tr>
<td>Tyr</td>
<td>856-860, 826-828</td>
</tr>
<tr>
<td>Cys</td>
<td>650-690, 750-790, 800</td>
</tr>
</tbody>
</table>
Figure S6. 1: A the cartoon representation of bovine insulin, dimmer shown (PDB: 1AI0). The A-chain is shown in red and the B-chain is shown in green. B, shows the insulin monomer (PDB: 4INS), with the amino acids of the hydrophobic core displayed in red for the A-chain and blue for the B-chain.

Figure S6. 2: Bulk solution Raman spectra (532 nm excitation) of intact insulin (50 mg ml⁻¹ in H₂O) measured under a series as a function of pH, from pH 1.60 to 3.00.
Figure S6. 3: PCA scores plot along PC1 and PC2, with the associated variance explained by PC1 and PC2 (A) of the three samples A chain insulin, B-chain insulin and intact insulin measured at five separate time intervals in DSMO-d6. The PCA loadings plot (B) along PCA 1 from the PCA scores plot (A).
Figure S6. 4: Deconvolution of the amide I region of A-chain Insulin (A), B-chain Insulin (B) and intact insulin (C) in DMSO-d6. Data: blue line; cumulative fit: red line; components of secondary structure: green circles.

Table S6. 3: Peaks identified (position and percentage areas) through the deconvolution of the amide I region of the A- and B-chain of insulin as well as intact insulin.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Peak 1</th>
<th>Peak 2</th>
<th>Peak 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Position cm(^{-1})</td>
<td>Area %</td>
<td>Position cm(^{-1})</td>
</tr>
<tr>
<td>A-Chain</td>
<td>1667</td>
<td>20</td>
<td>1680</td>
</tr>
<tr>
<td>B-Chain</td>
<td>1662</td>
<td>17</td>
<td>1682</td>
</tr>
<tr>
<td>Full insulin</td>
<td>1662</td>
<td>15</td>
<td>1680</td>
</tr>
</tbody>
</table>
Figure S6. 5: Width (A) and height (B) profiles of 20 different fibrils of B-chain insulin formed at pH 3. Each fibril was measured at 7-14 independent points, of which the average is displayed with the corresponding standard deviation.

Figure S6. 6: Width (A) and height (B) profiles of 25 different spherulites of B-chain insulin formed at pH 1. Each spherulite was measured at 3-5 independent points, of which the average is displayed with the corresponding standard deviation.
Figure S6. 7: Width (A) and height (B) profiles of 6 different fibrils of B-chain insulin formed at pH 1. Each fibril was measured at 8-15 independent points, of which the average is displayed with the corresponding standard deviation.

Figure S6. 8: Width (A) and height (B) profiles of 11 different fibrils of B-chain insulin formed at pH 2. Each fibril was measured at 8-15 independent points, of which the average is displayed with the corresponding standard deviation. These fibrils correspond to those for which TERS was recorded.
Figure S6. 9: Width (A) and height (B) profiles of 11 different fibrils of A-chain insulin formed at pH 2. Each fibril was measured at 8–15 independent points, of which the average is displayed with the corresponding standard deviation. These fibrils correspond to those for which TERS was recorded.

Figure S6. 10: The periodicity (A) in height of an A-chain fibril (B) formed at pH 2 along its axial length.
Figure S6. 11: Schematic diagram of the sequence of the A and B chains of bovine insulin. Amino acids unique to one chain are displayed red, while the SRE element of each sequence is indicated by a dashed line.

Figure S6. 12: TERS spectra measured along a fibril formed from the B-chain of insulin. Measurements were taken at 1 nm steps from each other.
Chapter 7: Investigating the thermal stability of Lysozyme with Raman spectroscopy
7.0. Declarations

As the first author of this planned publication I performed all the experimental work, and analysis of the spectra.
Investigating the thermal stability of lysozyme with Raman spectroscopy

Benjamin Gardner\textsuperscript{a}, and Ewan W. Blanch\textsuperscript{a}

\textsuperscript{a} Faculty of Life Sciences, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK.
E-mail: E.Blanch@manchester.ac.uk; Fax: +44 (0)161 236 0409; Tel: +44 (0)161 306 5819

7.1. Abstract

The complex process of hen egg white lysozyme (HEWL) aggregation under different concentrations of Hofmeister series ions (NaCl and CaCl\textsubscript{2}) has been investigated using Raman spectroscopy. To facilitate analysis of the data, full width half height (FWHH) of the amide I region and principal component analysis (PCA) were used. In this way we were able to show the sensitivity of HEWL to different concentrations of Hofmeister ions, which were found to directly affect the unfolding rate of HEWL in the case of NaCl. Furthermore, the concentration of Hofmeister ions alters the stability of the unfolded HEWL, which was observed in the rate at which intermolecular $\beta$-sheet observed to form. This work not only highlights the effect different Hofmeister ions have on aggregation; but also the importance of their concentration dependence when comparing kinetics of aggregation experiments.
7.2. Introduction

Understanding the conditions which affect protein aggregation is essential for a number of reasons. Most notably, fibrillation and aggregation are implicated in a number of neurodegenerative diseases including Parkinson’s, Huntington’s and Alzheimer’s disease [1-3]. These phenomena also presents a problem in the biotechnology industry, where protein aggregation can lead to “product fouling” [4-6]. Protein aggregation has been defined in a number of ways, however a recent review by Morris et al. differentiated aggregation into three types [7] where each type was dependent on its function and setting i.e. biological vs. industrial. Furthermore, five mechanistic groups explaining aggregation were put forward, wherein each class contained several sub classes [5, 7]. The aggregation of proteins is an incredibly complex process, with the products of aggregation ranging from unorganised protein aggregates, to highly organised protein fibrils. The formation of fibrils has been studied extensively due to its involvement in the mentioned diseases. Fibril formation has been shown to occur in vitro regardless of the native structure and sequence of the protein, and has been suggested to be a general property of proteins under the right conditions [2]. The structures of fibrils formed by different proteins are similar and defined by their classic cross-β structure [8]. This cross-β structure consists primarily of β-sheet in which the β-sheet runs in parallel to the propagation of the fibril axis, while the constituent β-strands are aligned perpendicularly to the fibril axis. However, fibrils only represent one end product structure of protein aggregation. Many other end product structures have been observed, some well defined such as spherulites [9, 10] formed by insulin, or globulomer [11] as formed by the Aβ peptide implicated in Alzheimer’s disease. Not all end products of protein aggregation are organised, as is the case of amorphous aggregation. In amorphous aggregation unordered aggregates form through intermolecular interactions, which in the native state were previously utilised in intramolecular interactions [12]. Therefore, understanding how the native states of proteins are destabilised is essential for understanding the process of aggregation. Investigating the conditions that destabilise and perturb protein structure has been an active area of research; conditions that have been investigated include: pH, temperature, mechanical stress, chaotropes, salts etc. [4, 13-18] The Hofmeister series classifies salts upon their ability to salt in and out proteins [19]. Moreover, these salts have been shown to influence both secondary and the tertiary structure of proteins, with anions having been demonstrated to have more of an effect than that of cations. Partial listings of the anions including F⁻, SO₄²⁻, acetate (Ac), Cl⁻ and cations K⁺, Na⁺, Mg²⁺, Mn²⁺. Understanding the role of the Hofmeister series of ions plays in protein structure has been one particular area of focus [20-22]. However, there still remain many questions as to their effects e.g. their role in disruption of bulk water and direct ion interactions [20-22].

Due to the heterogeneous and transient nature of protein aggregation, classic methods of protein structural determination, i.e. NMR and X-ray crystallography, are limited in their ability to monitor this problem [23]. A protein is required to be crystallised for X-Ray crystallography, or, be sufficiently small and soluble for NMR, none of these conditions are met in protein aggregation. However, these limitations are not observed for vibrational spectroscopy based methods, including:
Raman and Raman optical activity (ROA)\cite{4,24,25}, deep UV resonance Raman spectroscopy (DUVRR)\cite{23} vibrational circular dichroism (VCD)\cite{26} and tip enhanced Raman spectroscopy (TERS)\cite{27}. These approaches have been shown to be suitable for monitoring protein aggregation, especially when coupled with chemometric approaches including 2D correlation (2D-Cos)\cite{14,28} and principal component analysis (PCA)\cite{4,24}. This investigation uses Raman spectroscopy to monitor the influence of select Hofmeister salts influence on the initial structural unfolding of lysozyme while undergoing heating. The Hofmeister salts chosen for this investigation are Na\textsuperscript{+} and Ca\textsuperscript{2+}. These choices are separated in the Hofmeister list in their ranking of affect. Therefore, differences might be expected to be observed in their influence on the proteins stability when unfolding. Moreover, investigations of fibrillation often use differing conditions i.e. salt free or in buffer such as PBS. Therefore, establishing the significance of concentration of Hofmeister salts on unfolding is important.

7.3. Materials and Methods

Lysozyme has been studied extensively as a model protein for amyloid research\cite{24,29-33}. This work has shown the stability to be lowered with decreasing pH. An optimum pH for decreasing the stability and promoting fibril formation has been shown to be at pH 2 and raising the temperature to 60°C. Therefore, pH 2 has been selected to investigate the influence on Na\textsuperscript{+} Ca\textsuperscript{2+} on stability when their respective concentrations are changed and then the sample is heated. Concentration was selected on the basis of previous work, in addition to allowing Raman spectra with a high signal to noise to be gathered in real time. The sample of hen egg white lysozyme (HEWL) was purchased from Sigma Aldrich (UK), and used without further purification. A stock solution was prepared and pH adjusted (using HCl); this was then used to produce a series of samples containing different concentrations of NaCl and CaCl\textsubscript{2}. Samples contained a final HEWL concentration of 50 mg ml\textsuperscript{-1} at pH 2.00; which included either 0, 25, 50, 100, 200, 300, 400 or 500 mM NaCl or CaCl\textsubscript{2}. Samples were filtered using a 0.22 μm Durapore membrane filter to remove fluorescent impurities. Samples were then loaded into quartz micro-fluorescence cells and Raman spectra were recorded on a ChiralRaman instrument (Biotools Inc., Jupiter, Fl, USA)\cite{34} under the following conditions: laser excitation 532 nm, spectral resolution 7 cm\textsuperscript{-1}, laser power at sample ~600 mW, acquisition time 1.2 s. Temperature control of the samples was maintained by a thermostatically controlled custom cell holder using a Julabo F12 heating/cooling unit. This allowed control of the temperature to within 0.1°C. The temperature at the cell was verified using a Jenway digital temperature meter connected to a thermocouple. Control samples were measured at 20°C, while samples that were measured under heating were at 60°C and measured for 10 minutes (~350 spectra) in triplicate for each salt concentration.

The pretreatment of raw spectra is well established as a prerequisite for detailed chemometric analysis\cite{35-37}. Data was processed in Matlab 2010, using a previously described in-house toolbox\cite{36} and Matlabs bioinformatics toolbox\cite{38}. Data processing included baseline correction, using the previously described method developed by Ellers et al\cite{39}. Baseline correction is an
essential step as it helps to eliminate background variations in measurements, therefore primarily leaving only the changes due to experimental conditions. Other stages in the pretreatment of data included: the removal of aberrant spikes attributed to "cosmic rays", normalisation and subtraction of the Raman signal originating from water, smoothing to remove noise from the data, and mean centered for principal component analysis (PCA). PCA is a powerful analysis technique which is commonly used to discriminate between samples, or monitor changes in samples under a perturbation \[^4, 24, 36, 40\]. This technique allows for the reduction in the dimensionality of a data set, while the variance is maintained. The data goes through a linear orthogonal transformation, which is then ordered into principal components (PCs); where the first PC contains the highest variability of the data set and subsequent PCs account for a decreasing amount of variation. Therefore, this approach allows the variance in all spectral changes of a data set to be monitored, instead of focus being restricted to a small number of bands.

7.4. Results and Discussion

For the non-heated “native state” control samples maintained at 20°C not there were no observable visible differences between the series of HEWL samples in NaCl, or obvious differences in the Raman spectra (Figure 7.1A). While the HEWL samples between 0 to 400 mM CaCl\(_2\) remained clear, the sample in 500mM CaCl\(_2\) became cloudy, likely indicating precipitation of the sample. Furthermore, there was a noticeable difference in the Raman intensity of the sample at 500 mM CaCl\(_2\) prior to normalisation. Additionally, the PCA analysis shows a general clustering of the NaCl HEWL samples regardless of the concentration of salt (Figure 7.2A). However, for HEWL in CaCl\(_2\) the samples between 0 and 400 mM CaCl\(_2\) group together and are clearly separated from that of the 500 mM CaCl\(_2\) sample (Figure 7.2B).

The Raman spectra of HEWL (0mM NaCl) undergoing heating (60°C) (Figure 7.3A) show a series of changes. These changes include those in the intensity and position of several marker bands, which are easier to observe in Figure 7.3B, where only the first and last spectra are shown. PCA was used to monitor changes in HEWL under heating. The scores plots of PC1 (81.8%) and PC2 (8.1 %) (Figures 7.3C) how an initial clustering to the left hand side of the plot. This is followed by a transition to the right hand side, where the majority of the measurements are seen to cluster. This indicates that the changes that occur in HEWL upon heating do so rapidly. Furthermore, all three repeats show a similar response, indicating reproducibility of the process. The loadings plot (Figure 7.3D) contains information regarding the variance observed due to heating. The most intense change is seen in the amide I region, through two bands at 1660 and 1689 cm\(^{-1}\). Initially the amide I region is centred at \(-1660 \text{ cm}^{-1}\), through heating there is a change in the position and width of this feature. The structure of HEWL (Figure S7.1) consists of 39% α-helix and 10% β-sheet organised into two domains, α and β. The α-domain predominantly consists of α-helix, with four α-helices and a 3\(_{10}\)-helix. While the β-domain consists of β-sheet, loops and an additional 3\(_{10}\)-helix. Changes in the amide I region will be a result of the loss of the constituting α-helix and relative gain of other features; such the band at 1689 cm\(^{-1}\), which is usually assigned to turn type structure. HEWL contains 6 Trp residues, which are predominantly contained in the interior of the α-domain.
of the protein core (Figure S7.1B). A number of bands in the loadings plot can be attributed to Trp, in particular the band at ~1550 cm\(^{-1}\). The band at ~1550 cm\(^{-1}\) is due to the indole ring vibration termed W3\(^{[41]}\), however, the position is dependent on the torsion angle with Ca. Through heating there is a ~3 cm\(^{-1}\) downshift of W3, indicating a significant perturbation of the \(\alpha\)-domain core resulting in a change in the average environment of Trp residues due to heating. HEWL contains three Tyr and Phe residues, as shown in Figure S7.1 (C &D). Where the Trp residues were predominantly buried in the protein core, Tyr and Phe are located more on the protein surface. Further side chain changes can be observed at ~1007 cm\(^{-1}\) for Phe and 834 and 854 cm\(^{-1}\) for Tyr; where for Phe, a small gain in absolute intensity is observed after heating, while for Tyr there is an increase in the intensity of the band at ~854 cm\(^{-1}\). The relative intensity of two Tyr marker bands at ~830 and ~850 cm\(^{-1}\) have been shown to be an important indicator of the hydrogen bonding pattern\(^{[42]}\). The relative increase in the band at ~854 cm\(^{-1}\) compared to the band at 830 cm\(^{-1}\) indicates an increase in the solvent exposure of the Tyr residues of HEWL through heating. Additionally, variations in secondary structure bands include those for \(\alpha\)-helix at ~930 cm\(^{-1}\), disorder at ~960 cm\(^{-1}\) and \(\beta\)-sheet 1230 cm\(^{-1}\); where decreases in \(\alpha\)-helix and increase in disorder and \(\beta\)-sheet content is observed.

All of these observations are in agreement with previous work that has shown HEWL unfolds into a partially unfolded state under heating. This partially unfolded state has been shown to contain less \(\alpha\)-helix and tertiary structure\(^{[43]}\). As previously discussed the amide I region of a protein spectrum contains information regarding the secondary structure, with \(\alpha\)-helix, \(\beta\)-sheet, turns and disordered structure all contributing to the profile\(^{[44-46]}\).

As is shown (Figure 7.4) there is a significant difference in the amide I region after heating. Therefore, monitoring changes in the full width half height (FWHH) of the amide I region can be insightful into gross changes in secondary structure; if these changes are monitored as a function of time, this will be reflective of the rate of change. The average FWHH as a function of time (Figure 7.5A) is shown for HEWL under heating for all NaCl concentrations. This shows, in agreement with the PCA analysis, that the change in the amide I region occurs rapidly upon heating at 60°C. However, it also highlights the differences in the response between samples. For sample concentrations between 0 and 200 mM the initial change in FWHH, from ~36 cm\(^{-1}\) to 46 cm\(^{-1}\), is followed with a plateau, where no further changes in the FWHH are evident. At 300mM there is a gradual decrease in the FWHH after the initial increase to 46 cm\(^{-1}\); while at 400 mM following the initial increase in FWHH, it rapidly decreases at ~100 s while at 500 mM this decrease in FWHH occurs at ~80 s. Focusing on the first 160 s upon heating it is apparent that the concentration of NaCl directly affects initial unfolding rate. A trend is observed where the gradient of the transition increases with increasing NaCl concentration, which is shown more clearly in the fitted data (Figure 7.5C). As the samples between 0 and 200 mM NaCl show a similar response, only differing in the rate of change, only the samples containing concentrations above 200 mM NaCl will be discussed in more detail. For spectral data and analysis of the HEWL samples in 25 to 200 mM NaCl please see supplementary information (Figures S7.2–5).
For HEWL in 300 mM NaCl there is a gradual decrease in the FWHH from 46 cm$^{-1}$ to about 42 cm$^{-1}$ after the initial unfolding event. From the initial analysis of HEWL in 300 mM NaCl (Figure S7.6A–D), there was only a slight difference observed in the loadings of PC1 (Figure S7.6D). This was a slight increase in the intensity of the $\beta$-sheet band at 1230 cm$^{-1}$. However, it was possible through using PCA analysis carried out smaller subsets of data to determine the cause of this decrease in FWHH of the amide I region. When comparing the initial 10 spectra and 10 spectra following unfolding the loadings plot (Figure 7.6C) is similar to that previously observed. However, when introducing the final 10 spectra the three sets of data can be separated when plotting PC1 against PC3. In the loadings of PC3, there is an intense band centred at 1673 cm$^{-1}$, which is not observed in any of the loadings when only the initial and intermediate spectra are compared (data not shown). The band at 1673 cm$^{-1}$ is a marker band for $\beta$-sheet, which has a narrower FWHH compared to other structural markers such as $\beta$-turns$^{[44]}$.

As discussed the FWHH analysis of HEWL at 400 and 500 mM NaCl (Figure 7.5A) indicated that there was a significant difference in the development of the amide I region, compared to the other HEWL NaCl samples. This difference was not only in the kinetics of the change, but in the development of the FWHH. Both samples showed an initial increase in FWHH (44 and 46 cm$^{-1}$, respectively), followed by a decrease. These changes are reflected in differences in the spectra, and the PCA analysis (Figures 7.7–8 A–D) compared with those of the previous samples. There are similarities in the Raman spectra between HEWL in 400 and 500 mM of NaCl (Figures 7.7–8 A–B), which include the spectra becoming dominated by $\beta$-sheet bands. This $\beta$-sheet is observed at $\sim$1240 cm$^{-1}$ in the amide III region and 1675 cm$^{-1}$ in the amide I region (Figures 7.7–8B). Additionally, comparing the scores plots of the experiments at the two concentrations (Figure 7.7–8C), shows a similar trend. It is noted however, for 400 mM NaCl PCs 1 and 2 are used, while PCs 1 and 3 are used for 500 mM; as there is a significant difference between the two sets of spectral changes accounting for the variance. These differences include two peaks at 798 and 830 cm$^{-1}$ developing in the 500 mM NaCl sample, which is not observed in the 400 mM NaCl sample. These peaks can be seen to increase in intensity after initial unfolding and coincide with the increase in $\beta$-sheet; as shown (Figure 7.8B) these peaks are not present in the initial or intermediate spectra. Furthermore, these two peaks dominate the loadings plot of PC1 for HEWL in 500 mM NaCl (Figure 7.8D), which is not observed for PC1 of HEWL being heated in 400 mM NaCl (Figure 7.7D). Additional changes are observed in the NH stretch vibrations at $\sim$1078, $\sim$1113 and $\sim$1130 cm$^{-1}$, where the peak at 1078 cm$^{-1}$ becomes more intense compared to the latter two. This indicates significant changes occurring in the protein backbone structure rapidly after initially unfolding when in the presences of a higher NaCl concentration.

As shown in the FWHH analysis of the amide I region for HEWL in CaCl$_2$ (Figure 7.9), these results share similarities with those of HEWL being heated with different NaCl concentrations. For samples measured at 0–100 mM CaCl$_2$ there is a similar response, to that of NaCl, where within 200 s a transition occurs in the amide I width from $\sim$36 cm$^{-1}$ to $\sim$45 cm$^{-1}$. Interestingly, there appears to be no obvious correlation between CaCl$_2$ concentration and the shape of this transition observed, as was seen for the NaCl containing HEWL samples. However, at 200 mM CaCl$_2$ a
decrease in the FWHH is apparent after the initial increase. Increasing the CaCl$_2$ concentrations further (300–500 mM) sees a dramatic change in the FWHH response of the amide I region. Whereas for 300 and 400 mM CaCl$_2$ FWHH starts to decrease at ~90 s, this is seen for 500 mM CaCl$_2$ at ~60 s. Interestingly, for HEWL in 400 and 500 mM CaCl$_2$ changes in the width of the amide I profile coincide with the emerging band at ~798 cm$^{-1}$. This observation differs to what was observed for HEWL in NaCl at the same concentration. In that sample the broad amide I region was observed prior to the emergence of the band at ~798 cm$^{-1}$. 

While there are differences in the kinetics of changes observed, there were no differences in the observed spectra compared to the HEWL NaCl samples i.e. for either samples that showed a stable FWHH after unfolding, or, samples that showed a decreasing FWHH, the spectral response was equivalent to the counterpart HEWL NaCl samples. Therefore, specific details of the spectra are not discussed in any more detail (please see Supplementary Information S7–14 for spectra and PCA analysis).

7.5. Conclusions

Through the Raman spectra, FWHH analysis and PCA analysis the unfolding of HEWL in different concentrations of NaCl and CaCl$_2$ has been monitored. It has been shown that both NaCl and CaCl$_2$ act to destabilise the native structure and promote an accelerated unfolding of HEWL under denaturing conditions, in this case high temperature (60°C). Furthermore, as the concentration of NaCl and CaCl$_2$ were increased (>200 and >100 mM respectively) this promoted and accelerated the formation of β-sheet aggregates. A significant finding includes the correlation between rate of unfolding and NaCl concentration which was not observed for CaCl$_2$. The implication of this is there are differences between how the sodium and calcium interact with HEWL. While specific interactions have not been reported for either cations (Na$^+$ or Ca$^{2+}$) HEWL, Cl$^-$ has been shown to interact with Tyr23[1]. Theoretical work has shown differences in the interaction energy between cations, and aromatic residues[2]. Specifically, it was shown that the interaction energy of Ca$^{2+}$ > Na$^+$ for these types of interactions. This might be one explanation of some of the differences observed, as a number of changes corresponded to that of aromatic residues. Therefore, a difference in how the cations interact with the aromatic residues might influence overall aggregation. However, to fully understand these results further work would be required, that is beyond the scope of this initial investigation. With the ever increasing performance of computer molecular dynamics (MD) is being applied to more complex problems. MD has been previously applied to looking at Lysozyme where it was used to investigate pH, concentration point mutations and salt concentration[3]. Therefore, this would be a prime technique that might be used in furthering the results observed here[4].

This work also indicates that unfolding occurs prior to β-sheet formation even under extreme conditions. Furthermore, these results also show that the effects of even low salt concentrations need to be considered when performing investigations of the kinetics of unfolding. This is especially so if comparisons with other work in the field are to be made.
7.6. References


38. Matlab and Bioinformatics toolbox,R2010a;The MathWorks, Inc.;Natick, Massachusetts, United States.,2010


Figure 7.1: The average Raman spectra for HEWL (50 mg ml⁻¹ in H₂O at pH 2) at 20°C as a function of concentration (0, 25, 50, 100, 200, 300, 400, 500 mM) of NaCl (A) and CaCl₂ (B), respectively.
Figure 7.2: PCA scores plot of the non heated control measurements of HEWL (50 mg ml\(^{-1}\) in \(H_2O\) at pH 2 maintained at 20°C) in NaCl (A) and CaCl\(_2\) (B), respectively, showing PC1 plotted against PC2. The concentrations of NaCl and CaCl\(_2\) given in the keys.
Figure 7. 3: Raman spectra of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 with 0mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1$^{st}$, 2nd and 3rd runs are blue, green and red respectively. The loadings plot of PC1 is shown in panel D.

Figure 7. 4: The amide I region of the initial Raman spectrum (black) and the final (red) for HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2) during heating at 60°C (0 mM NaCl).
Figure 7.5: The average FWHH analysis of the amide I region of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2) for a series of NaCl concentrations (0, 25, 50, 100, 200, 300, 400 and 500 mM) observed as a function of time while under heating at 60°C for ~15 minutes (A). The first 160 seconds are observed in B, where horizontal error bars representing the standard errors are shown. Sigmoidal fits of the FWHH data of HEWL as a function of time (s) is shown (C) for 0–400 mM NaCl.
Figure 7. 6: PCA Scores plot of PC1 and PC3 for sub sets (10 spectra chosen at 3 time points) of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2) undergoing heating at 60°C in 300 mM NaCl (A). Loadings plot of PC1 (B) and PC3 (C).
Figure 7.7: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 400 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\), 2\(^{nd}\) and 3\(^{rd}\) runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure 7.8: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 500 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC3 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.

Figure 7.9: The average (of triplicate measurements) FWHH of the amide I region of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2) being heated at 60°C over a range of CaCl\(_2\) concentrations.
Figure S7. 1: Cartoon representation of HEWL (PDB: LYZ1) visualised in PyMOL (a). Sidechains are displayed in blue for Trp (b), Tyr (c) and Phe(d).
Figure S7. 2: The Raman spectra of HEWL (50 mg mL\(^{-1}\) in H\(_2\)O at pH 2 and 25 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\) and 2nd runs are blue red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 3: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 50 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\), 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 4: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 100 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 5: The Raman spectra of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 and 200 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 6: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 300 mM NaCl) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\), 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 7: The Raman spectra of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 and 0mM CaCl$_2$) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7.8: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 25 mM CaCl₂) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 9: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 50mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\), 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 10: The Raman spectra of HEWL (50 mg ml⁻¹ in H₂O at pH 2 and 100mM CaCl₂) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1st, 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 11: The Raman spectra of HEWL (50 mg ml$^{-1}$ in H$_2$O at pH 2 and 200mM CaCl$_2$) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1$^{st}$, 2$^{nd}$ and 3$^{rd}$ runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 12: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 300mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\), 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 13: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 400mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 1\(^{st}\), 2nd and 3rd runs are blue green and red, respectively. The loadings plot of PC1 is shown in panel D.
Figure S7. 14: The Raman spectra of HEWL (50 mg ml\(^{-1}\) in H\(_2\)O at pH 2 and 500mM CaCl\(_2\)) undergoing heating at 60°C (A), where the first (blue) and last spectra (red) are shown (B). PCA scores plot (C) displaying PC1 and PC2 where the 2nd and 3rd runs are shown as blue and green respectively. The loadings plot of PC1 is shown in panel D.
Chapter 8: Conclusions and further work

8.1. Conclusions

Multidisciplinary spectroscopic studies of biology provide an excellent way to investigate challenging problems, including protein fibrillogenesis. Through combining spectroscopic techniques i.e. Raman and ROA, TERS and DUVRR, each provide complementary information i.e. real time monitoring, chiral information, information regarding the protein backbone and the surface chemistry, respectively, which when used together can be incredibly insightful. Moreover, when the spectral information obtained from techniques is analysed with chemometric approaches, including 2D-Cos approaches and principal component analysis (PCA) a more detailed understanding can be developed. This is the central premise of the thesis presented here: to improve the understanding of the structural characteristics and aggregation mechanisms of fibrillogenesis, through combining a range of complementary Raman spectroscopies and analytical methods.

Chapter 3 reported the effect that protonation has on the common amino acids, which was monitored using Raman and ROA spectroscopies. This work worked showed that protonation has a significant effect on the spectra, and therefore the conformations of the amino acids. Changes upon protonation were shown to be more complex than can be explained by simply the addition or subtraction of a proton. It was shown that the protonation state not only influences hydrogen bonding, but also the side chain environment. The influence of protonation on the side chain conformation was regardless of their ability to accept or donate a proton i.e. Val showed changes in its side chain conformation even though the side chain is not directly involved in protonation. This was especially interesting, as it highlighted not only the sensitivity of the overall amino acid structure to pH, but the importance in accounting for the pH. The principal aim of this work was to expand the spectral assignments on account of protonation, which is essential in experiments not carried out at a neutral pH, which previous studies have focused on\(^1\text{–}^4\). Moreover, this is the first time all amino acids have been measured for both Raman and ROA in solution in one study. This work is especially useful in the assignment of changes in protein structure where pH effect is investigated (Chapters 5 and 6), as it allows distinction between changes in secondary structure and those of the side chains. This is crucial with Raman spectroscopy, as the side chains contribute significantly to the overall spectra of proteins. Therefore, it is important to characterise spectral responses from individual amino acids as a function of protonation from changes induced in the secondary structure. This is in order to be able to disentangle what changes in a spectrum can be reliably attributed to side chains, from those of the secondary structure of a protein undergoing a complex structural change such as fibrillogenesis.

Chapter 4 addressed the need for a combined toolbox for easily processing large data sets of Raman and ROA spectra and their post-acquisition analysis, including 2D correlation. This was due to the lack of free toolboxes, and the limitations of the available software i.e. not being suitable for
processing large data sets, in either processing spectra or the post-acquisition analysis. Therefore, a toolbox was developed for Matlab, which incorporated a number of common data processing tools, as well as a module to carry out 2D-Cos. This development was a necessity for the work detailed in chapters 5–7, to enable the efficient analysis of the large number of Raman spectra generated from real-time measurements (some in excess of 100,000); as opposed to a limited spectral set based on measurements at a few time points. Furthermore, it was shown in the demonstration of PCMW, that caution is required when it is applied to bisignate data. It was shown that a peak changing towards $+\infty$ is always displayed as a positive change; conversely, a peak changing in the direction of $-\infty$ is always negative. Therefore, this can cause confusion when considering the negative half of the spectrum; as a negative peak increasing towards $-\infty$ will show as a negative change, while a negative peak decreasing i.e. towards the zero line will be displayed as a positive change. Therefore, when interpreting this type of analysis, it is required to either refer back to the original spectra, or discuss the changes with respect to direction towards positive or negative infinity, to prevent errors in the analysis of bisignate data.

Chapter 5 built on the characterisation of amino acid spectra as a function of pH conducted in Chapter 3, by using Raman spectroscopy for monitoring the in vitro fibril formation of bovine insulin at pH 2 and 3 in real-time; this was possible through use of the Matlab toolbox described in Chapter 4. Moreover, the effect of temperature quenching during the initial stages of fibril formation was also investigated at both selected pHs. These pHs were chosen building on work in the literature showing these pH changes influence insulin fibril morphology\cite{5}. The usefulness of monitoring changes in the FWHH of selected spectral regions was demonstrated. This approach was used to highlight transitions in spectral response as a function of time. This work showed that at both selected pHs the changes observed in the Raman spectra during fibrillogenesis were predominantly the same, where rapid unfolding preceded $\beta$-sheet formation. Following initial changes in the spectra a lag phase was observed, prior to a 2$^{nd}$ transition period in the spectra, where further changes in $\beta$-sheet and Phe and Tyr side chain environments occurred. While most of the spectral changes observed were very similar for Insulin at both pHs, there were some notable differences. Specifically, there was a band at 790 cm$^{-1}$ that is observed to appear when fibrils are formed at pH 2, that is absent at pH 3. Through the work described in Chapter 3, this appears not to be attributed to a side chain as no bands from individual amino acids were observed at this position. Therefore, this band is either a shifted band from a side chain band, or, this is a marker band for secondary structure. Secondary structure marker bands in this region are not as well documented compared to other spectral regions, e.g. the amide I region. As has been previously reported for insulin fibrils that were produced under different pH conditions, the sign of the chiral signal measured by VCD has been observed to change\cite{6}. This has been postulated to be due to differences in the orientation of the fibrils and the packing of protofibrils\cite{5}. The observation of a specific band at 790 cm$^{-1}$ at pH 2 but not at pH 3 certainly indicates specific mechanistic differences in the fibril assembly process when formed at different pHs. This spectral observation has not been previously discussed in regards to insulin fibrillogenesis, this therefore represents a significant novel finding. Differences in the kinetics were also observed, the lag phase was
noticeably shorter at pH 2. Differences in the kinetics of the fibril formation of insulin as a function
of pH are not entirely novel\cite{7}, and have been discussed previously in terms of the monomer
concentration; when at a higher pH insulin forms a dimer which firstly needs to disassociate prior to
fibrillation\cite{7, 8}. However, as was shown, minimal differences were observed between insulin's initial
structure at pH 2 and 3; yet, differences were observed in the fibrillogenesis mechanism.
Therefore, it is likely there is a more complex relationship between insulin fibrillation and pH than
previously appreciated. Furthermore, when insulin was heated for a short time period prior to
temperature quenching samples at pH 3 showed full recovery of the native state of insulin, while
this was not observed at pH 2 even when heated for a shorter time period. This work shows the
advantage of monitoring the fibrillogenesis of proteins in real-time with Raman spectroscopy, by
providing a more detailed understanding of the mechanistic process, compared to a simple
comparison of the initial and final spectra\cite{9, 10}. This is especially so when monitoring rapid changes
in the structure, as is demonstrated in this work.

Chapter 6 built on the work established in Chapter 5, which demonstrated the benefit and
feasibility of real-time monitoring of fibrillogenesis; by combining this approach with additional
techniques in a multidisciplinary investigation of fibrillogenesis. In this investigation, the
independent A and B-chains of insulin were studied using Raman, ROA, TERS and DUVRR; these
techniques provided different complementary information, rapid monitoring of structural changes,
chiral information, detailed surface chemistry and selective probing of secondary structure,
respectively. This work showed that the A and B-chains of insulin are structurally dissimilar, and
that the conformations of both are sensitive to pH. It is interesting that for both of these chains a
band positioned at 790 cm$^{-1}$ was observed upon lowering the pH, which was not observed for intact
insulin as a function of pH. This is the same position as was found for a band observed in Chapter
5 in the latter stage of fibrillogenesis at pH 2. This is noteworthy, as it shows that pH leads to the
same conformational change in both A and B-chains as fibrillogenesis does for intact insulin. It was
observed that the initial pH greatly influenced fibrillogenesis in the case of the B-chain. It was seen
that at pH 1 the B-chain formed spherulites, while at pH 2 and 3 fibrils were observed. Therefore,
pH in the case of the B-chain was seen to directly alter the pathway of fibrillogenesis. This is
important, as pH has been shown to influence the fibril morphology\cite{5, 6, 11} of intact insulin, but the
effect of protonation is apparently more significant for the isolated B-chain of insulin. Previous work
investigating spherulite formation of insulin showed protein concentration and salt to be critical to
spherulite formation\cite{12-14}. However, with the B-chain, just low pH was sufficient to drive
fibrillogenesis primarily to this. The ROA investigation of the A and B-chains dissolved in DMSO-d$_6$
showed further differences between the behaviour of the two independent chains, in what has been
described as a stabilised intermediate state of fibrillogenesis\cite{15}. Monitoring proteins in DMSO-d$_6$
induced conformational states provides an interesting avenue for fibrillogenesis studies\cite{15}. As
DMSO studies provide access to a structure or population of structures that would normally be
transient in nature. Further studies on a number of other amyloidogenic proteins, as well as non-
amyloidogenic proteins could be important in establishing if these structures adopted in the non-
aqueous DMSO environment are observed across all proteins, or if they are conserved to
amyloidogenic proteins. In accordance with previous work, differences were observed in the morphology of the A and B-chain fibrils\textsuperscript{[16-18]} as observed with AFM. While the B-chain fibrils were generally quite consistent in their dimensions, the A-chain fibrils however were more heterogeneous and showed a range of structures. Differences in the morphology of the fibrils translated to differences in the fibril structure at the molecular level. As was shown with the DUVRR measurements, differences were observed between the fibril cores of the A and B-chain fibrils. Furthermore, TERS showed that the A and B-chain fibrils exhibit differences in the amount of secondary structure present at the surface. Specifically, it was seen that the B-chain consisted of more α-helix and disordered at the surface in comparison of that of the A-chain fibril.

As the suitability of Raman spectroscopy had been demonstrated in fibril studies of insulin detailed in chapters 5 and 6, a different fibrillogeneic protein (lysozyme) was investigated in Chapter 7. As pH influence on unfolding and fibrillation is well documented for lysozyme, Chapter 7 investigated the influence of Hofmeister ions on the unfolding and aggregation of hen egg white lysozyme (HEWL) with Raman spectroscopy. The Hofmeister series is a series of ions which act to influence the solubility and structure of proteins through different interactions. Raman spectroscopy was used in real time to monitor the unfolding and aggregation of HEWL as a function of NaCl and CaCl\textsubscript{2} concentration. It was shown that NaCl concentration directly affected the kinetics of the initial partial unfolding of HEWL. However, this direct relationship was not apparent with changes in the concentration of CaCl\textsubscript{2}, as the CaCl\textsubscript{2} concentration did not affect the kinetics. Previously Ca\textsuperscript{2+} has been shown to have a destabilising effect on the structure of HEWL\textsuperscript{[19]}. This work showed changes in the concentration of both NaCl and CaCl\textsubscript{2} influenced the stability of the HEWL unfolded intermediate. As both NaCl and CaCl\textsubscript{2} concentrations were increased, the rapid aggregation of intermolecular β-sheet containing species was observed. This work highlights differences between the way NaCl and CaCl\textsubscript{2} interact with HEWL, leading to differences in the unfolding. Hofmeister salts have been shown to have a complex interaction with the fibrillogenesis pathway. In insulin the presence of small amounts of NaCl has been shown to be critical for the formation of spherulites\textsuperscript{[12,20]}. Moreover, with bovine apo-α-lactalbumin removal of NaCl was seen as critical for the spontaneous refolding of the fibrils from one polymorph to another\textsuperscript{[21]}. There are a number of suggested methods for how Hofmeister salts could be interacting with proteins to modulate stability. These include electrostatic interactions with side chains, and competing with water hydration during unfolding\textsuperscript{[19]}. It might be impossible to predict on general terms how a protein will respond to Hofmeister salts, as is seen with insulin and apo-α-lactalbumin above. This is because the presence of Hofmeister salts causes a very different response in these examples. Taking the case of electrostatic interactions with charged side chains, these could be either stabilising or destabilising for the peptide backbone on a site specific basis. Moreover, the result of this interaction would also be specific to the anion or cation involved\textsuperscript{[19]}. Overall, this work shows the importance of accounting for the concentration of Hofmeister ions in studies of aggregation kinetics. The observations reported here also add to the current and sizeable body of work investigating the Hofmeister series and its influence on protein structure and fibrillogenesis.
The combined body of work reported in this thesis has demonstrated the suitability of complementary Raman spectroscopies for the study of fibrillogenesis. In the case of conventional Raman spectroscopy, protein structural changes were monitored in real-time; this was facilitated by the toolbox that was developed, allowing efficient analysis of large scale spectral sets. The other complementary techniques used in this work included: ROA, which provided chiral information and acted as a sensitive probe of secondary structure; TERS, which provided a detailed chemical analysis of the surface of fibrils (e.g. the secondary structure found on the surface of a fibril and the side chains found on the fibril surface); and finally, DUVRR which provided a method to probe the fibril core of a sample. The additional mechanistic information these techniques can obtain was detailed in Chapters 5–7. Through being able to monitor processes i.e. fibrillogenesis in real time, this has most importantly led to an improved understanding of the mechanistic changes observed in two model systems of fibrillogenesis (Insulin and HEWL) when variables such as pH or salt concentration are changed. This has allowed us to develop a more detailed model of the conformational changes driving fibrillogenesis, especially for insulin as discussed in Chapter 6, which has further expanded our understanding of this complex process. This work is important because it provides a foundation for further studies to be carried out monitoring fibrillogenesis with these approaches. Specifically, the ability to monitor and identify transient aggregates of fibrillogenesis mechanisms is arguably one of the most significant outcomes of this research; as such early stage intermediates are now key targets for therapeutic intervention.

8.2. Future work

Providing enabling software that allows increased access to vibrational spectroscopies is vital, to further increase its use in the scientific community. Therefore, further and continual development of the toolbox described in Chapter 4 will be important. While the toolbox provides the necessary tools for spectral analysis it can be expanded in a number of ways. Firstly, expanding the data processing routines to include additional approaches to baseline, smooth and normalise data would be appropriate. Furthermore, as discussed in the chapter itself, an essential way to take it forwards would be introducing PCA; therefore, expanding the chemometric capabilities with a complementary clustering method. This would be particularly useful for researchers that work with large data sets, where quickly discriminating between different classes of samples i.e. where medical diagnostics is the objective. Further additions might include an expansion of the plotting functions, and include straightforward approaches to label peaks and tabulate the information; this would be especially useful for the analysis of TERS data. Through these introductions it would greatly expand the versatility of the toolbox, and hopefully improve its usability and suitability for the wider spectroscopy community.

The work contained within this thesis has demonstrated the usefulness of vibrational spectroscopy, and chemometric approaches, in the studying of the complex structural changes underlying fibrillogenesis. However, the work contained in this thesis focussed on understanding model systems. Therefore, the most obvious way to take this work forwards would be to focus on
more clinically relevant samples i.e. the peptide fragments implicated in Alzheimer’s disease (AD), Aβ40 and Aβ42. As was discussed in the introduction, not only are these peptide sequences implicated in AD, but also their post translational modified variants. By combining different spectroscopic approaches i.e. conventional Raman, ROA, TERS and DUVR as was demonstrated in chapter 6, it could be possible to investigate the role post translational modification plays in fibrillogenesis. As these techniques could be used to answer specific questions e.g. how the fibril core is perturbed and how the surface chemistry is altered through different modifications? Furthermore, these multidisciplinary approaches would be suited to investigating pharmaceutical interruptions of fibrillation. TERS in particular could be suited to this specific purpose; being able to not only investigate the morphology of the resulting off pathway products of fibrillogenesis, but identifying where these pharmaceuticals bind. This information could be crucial in the development of effective treatments for Alzheimer’s disease, which as discussed are currently lacking.

One obvious omission from this thesis is the absence of ROA data for fibrillar species, which was due to linear birefringence induced artefacts that plagued any attempts to measure fibril ROA spectra. Similar birefringent artefacts have been previously reported to be 1000 times stronger than optical activity signals. Artefacts caused unreliability in the spectral features, which manifests in the absolute intensity and sign of most peaks in a spectrum being un-reproducible between measurements. With more time two possible solutions might have been explored to overcome the problems associated with birefringence. One potential solution is the development of a cell holder that rotates the sample perpendicularly to the laser beam. The rationale for this is that this axial rotation would cause an averaging of the linear birefringence, therefore allowing a non biased spectrum of the sample to be acquired. This approach has been successfully applied in VCD, where it was shown to aid the removal of birefringence; therefore, it would likely help in this setting too. A second more technical method would to be use an ROA instrument that is optimised to detect light that is scattered in the Δz (90°) geometry. The reason for this is that in measuring depolarized ROA, the contribution is limited to the anisotropic contribution, which is less susceptible to artefacts. Therefore, it might provide a more reliable way to measure ROA for these challenging samples.

8.3. References