What is the interrelation of protein structure and physical behaviour and how is it influenced by environmental conditions?

Daniel J. Belton and Aline F. Miller*

School of Chemical Engineering and Analytical Science, The University of Manchester, PO Box 88, Manchester, UK M60 1QD

*To whom correspondence should be addressed. E-mail: aline.miller@manchester.ac.uk

Abstract

Proteins are essential for life and as such have been deemed useful as therapeutic agents. However, the complexities that make proteins important for biological processes can also make their behaviour complicated. Such behaviour can make production, processing and formulation of protein pharmaceuticals problematic. A therapeutic protein test system has been examined using infrared spectroscopy and differential scanning calorimetry to increase understanding of its behaviour. It has been show that the protein thermally denatures at 47 °C and that this is accompanied by changes to the structure.

Introduction

Proteins are long unbranched chain molecules that can adopt complicated three-dimensional structures. Their structure is crucial for their biological activity and functionality. However, this structure can change under different environmental conditions; the structure can breakdown (denaturation) or be subtly altered to regulate biological activity (allosteric effect) or higher order structures can be formed (e.g. dimers and aggregates). Understanding this behaviour is required for precise control over protein structure and will allow improved processing of biotech products. Also, any understanding gained in vitro, motivated by the biotech impetus, would have the additional spin off of a better understanding of in vivo processes, with corresponding implications for health.
A great deal of work has been done on determining the structure of proteins. Separate work has involved examining physical behaviour that results from changes in structure. But how can we re-examine these in order to increase our understanding and answer questions such as:
How and why do environmental conditions affect subtleties of structure?
How and why do environmental conditions and any resulting structural changes affect behaviour?

The area of protein structure and the importance of structure for function is well covered in the literature (Alberts et al., 2002, Branden and Tooze, 1991). Proteins are made up of amino acids joined together by peptide bonds. They commonly contain more than one hundred amino acids and are essential elements of living systems. The complicated three-dimensional structures that proteins adopt allow them to have varying functional properties. Some exhibit mechanical properties such as elasticity or tensile strength; for example:

- Keratins: form elastic fibres that confer elasticity to the skin
- Collagen: forms tough fibrils that give connective tissue tensile strength

Alternatively, proteins might bind to specific molecules for catalysis, molecular recognition or transportation, examples include:

- Enzymes: bind to other molecules to catalyse reactions
- Calmodulin: binds to calcium in order to alter activity of other proteins

Proteins can also build up in to larger assemblies; for example:

- Viruses: identical proteins pack together to form a spherical shell
- Amyloid: large well ordered assembly of proteins resulting in fibres

The sequence of amino acids in the long molecular protein chain is referred to as the primary structure. Sections of the protein chain can coil, fold or loop. The nature of these localised arrangements is a direct result of the amino acid sequence and is referred to as secondary structure. If the backbone coils

this is referred to as helix, whereas if it forms pleated sheet this is referred to as β-sheet. The localised arrangements or secondary structure elements are well documented in the literature. There are two common means of defining protein secondary structure, using either backbone torsion angles (Ramachandran and Sasisekharan, 1968) or hydrogen bonding pattern (Kabsch and Sander, 1983).
The spatial arrangement of all the atoms of a protein subunit is referred to as tertiary structure (IUPAC, 1970). A subunit refers to a distinct chain component within a protein that may be separated from other chain components, whether identical or different, without breaking main chain covalent bonds. The inter-subunit contacts and interactions of a protein are referred to as quaternary structure (IUPAC, 1970). A protein molecule that does not have any subunits that can be separated without the braking of covalent bonds possesses no quaternary structure.

The classical and simplistic view of protein behaviour is shown in Figure 1; folded protein unfolds and then aggregates. This over simplified picture tells us nothing about the nature of the denaturation, aggregation or protein structures and how these are interrelated. There are examples of protein behaviour in the literature that allow a more universal approach to be taken. This is proposed and neatly summarised in Figure 2; links between behaviour are demonstrated and it also allows trends to be identified. This is the first time such a representation has been proposed, to the author's knowledge. Whilst the generalisations may not apply to all systems it is the initiation of a more global model and approach.

N (native) ↔ U (unfolded/denatured) → A (aggregated)

Figure 1 Simplistic view of protein behaviour

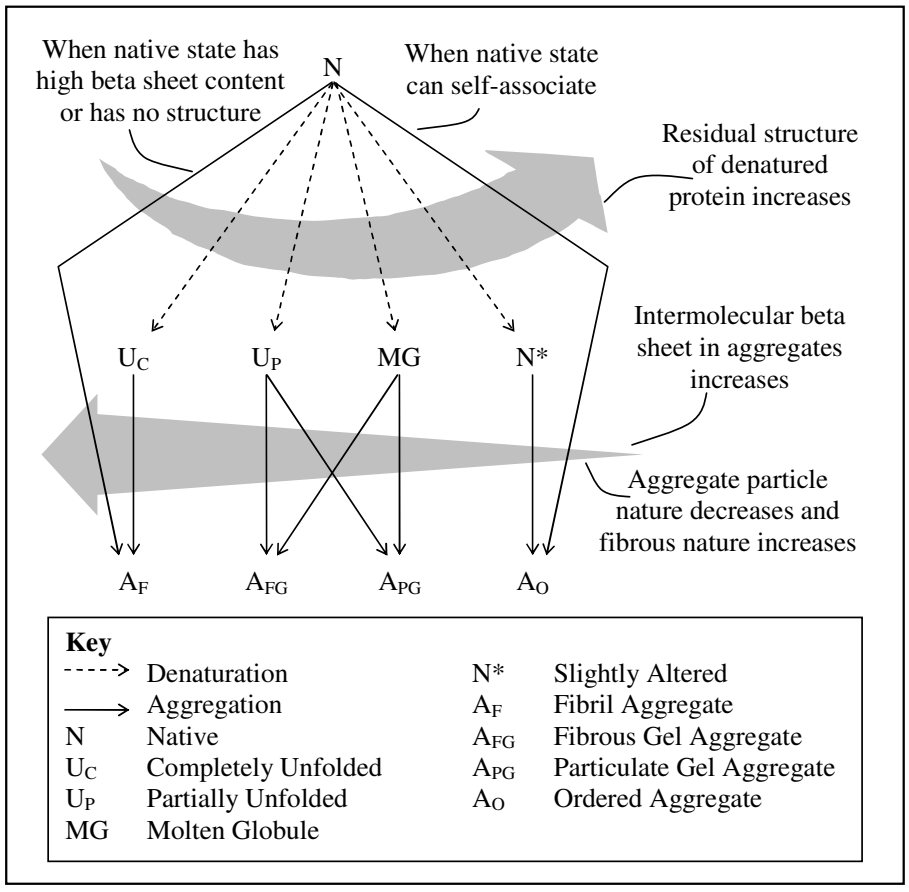

Figure 2 The interrelation and diversity of protein behaviour

The behaviour of proteins is made complicated by their complex three dimensional structures. A comprehensive review of protein behaviour exists (Wang, 1999).

The transition from a random protein chain to a protein with structure is referred to as folding. Proteins must be properly folded to exhibit the correct biological activity; this is their native state (N). Folding can be a complicated process so when an organism synthesizes a protein it requires the assistance of molecular chaperones to ensure the correct conformation.

The alteration of the native structure, without change to the primary structure but with a loss of biological activity, is termed denaturation. Protein can be denatured by chemical, thermal or mechanical action. Chemical denaturants include urea and guanidine hydrochloride, which disrupt hydrophobic interactions and hydrogen bonds by preferentially binding to the protein to reduce the free energy and chemical potential of the denaturant. Protein structure can also be altered by extremes of pH, which cause increased repulsion between like charges. A protein will denature when heated, this occurs at a temperature specific to the primary structure. Shaking a protein solution can create foam films since proteins are able to stabilise the air water interface; a requirement of foam formation. In this location the protein structure can then be perturbed by hydrophobic and hydrophilic regions being pulled away from each other (Clarkson et al., 1999). The extent to which the structure is perturbed by denaturation can vary. A protein might loose its entire ordered structure and become completely unfolded, U_C (Uversky, 2002). Alternatively, a protein may loose its tertiary structure but exhibit residual secondary structure and so become partially unfolded, U_P (Uversky, 2002). If most of the native structure is retained but residue side chains become mobile this is referred to the molten globule state, MG (Pande and Rokhsar, 1998). Finally, the structure might only be altered slightly, N*, as with the allosteric effect. The allosteric effect occurs when a protein becomes ligand bound and under goes a conformation change in order to induce or inhibit biological activity (Allaby, 1998).

Aggregation is the grouping and holding together of protein monomers by intermolecular interactions. There are various interactions between residues from adjoining monomers (Jones and Thorton, 1995), these include the formation of salt bridges between oppositely charged residues, the formation of sulphur bridges between cysteine residues and the formation of hydrogen bonds between residues. The latter is the most important and is referred to as intermolecular beta sheet. The aggregation process is driven by the thermodynamically unfavourable exposure of hydrophobic residues to water. By drawing upon various examples it becomes clear that the nature of aggregate structure depends upon the state of the protein monomer prior to aggregation.

If the native protein structure is able to aggregate without structural change then an ordered aggregate will be formed as with insulin dimers and hexamers (Brangex et al., 1997) or virus protein capsids (Branden and Tooze, 1991), the aggregate size being limited by the regular nature of the structure. If the native protein structure is slightly altered by the allosteric effect again ordered aggregates can form as with the protein TLR4 (Saitoh et al., 2004).

If the native protein structure is denatured to form a molten globule then upon aggregation a gel is formed and in this case the aggregate size is not limited by the nature of structure. The type of the gel formed depends upon the pH of the system (Gosal and Ross-Murphy, 2000, Clark et al., 2001), if the pH is close to the isoelectric point (pI) of the protein then a particulate gel will form or if the pH is well below the isoelectric point then a fibrilliar network gel will form. The isoeletric point of a protein is the pH at which the proteins net charge is zero and so repulsion between monomers is minimised. If the native protein structure becomes partially unfolded then a fibrilliar network gel is formed, with the initial aggregation being linear to form β-sheet fibrils followed by random physical cross-linking of the fibres (Clark et al., 2001).

If the native protein structure is either denatured to a completely unfolded state or if it possesses high beta sheet content or if it has no structure then upon aggregation a fibril is formed and again the aggregate size is not limited. Fibril formation without denaturation is exemplified by certain peptides that have no structure in the native state (Zanuy et al., 2003) and also by prion proteins that have high beta sheet content (Prusiner, 1998).

The work in this study focused on the protein ‘protective antigen’ (PA). This protein is being produced as the main constituent in a new and improved Anthrax vaccine to replace the existing vaccine (AVA), which is subject to flaws (Young and Collier, 2002) and safety concerns (Pittman et al., 2001). The US government is to spend $700 million on stockpiling the new vaccine in contracts with two companies, one of which is Avecia (Gillis, 2004).

The PA is produced using genetically modified bacterial which were created by a DNA recombinant technique. Normally, PA is produced and released by *Bacillus anthracis* the bacterium that causes anthrax (Stubbs, 2002). PA is part of the anthrax toxin system. Its role is to bind to a cell surface, after which it self-associates to form a surface bound complex that can transport lethal factor (LF) and oedema factor (EF) into the cell. LF and EF cause cell damage by disrupting essential cell signalling which subsequently leads to death of the host. The rational behind a PA vaccine is to enhance the immune response against the PA and so render the anthrax toxin ineffective (Jendrek et al., 2003).
The structure of PA has been determined by X-ray diffraction (Petosa et al., 1997). Various features of PA structure can be found on the Protein Data Bank (PDB) (Berman et al., 2000), which is available at http://www.pdb.org. The PDB identity code for PA is 1ACC. The images in Figure 3 and Figure 4 and the information in Table 1 has been acquired or derived from the PDB. The image of PA in Figure 3 shows the arrangement of the non-hydrogen atoms, whilst its quality and usefulness are limited, it does give an idea of the shape and complexity of this molecule. The image of PA in Figure 4 is clearer and highlights α-helix and β-sheet elements of the secondary structure; α-helix represented by cylinders and β-sheet by arrows.

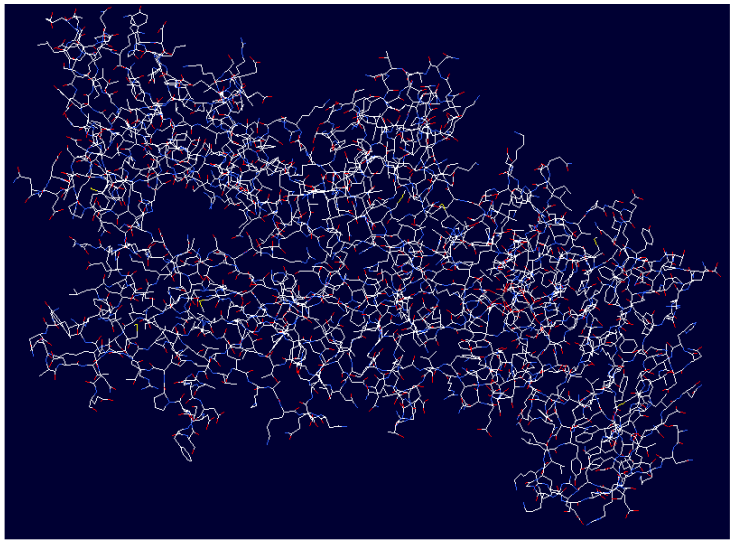

Figure 3 Stick model of PA from PDB; carbon in white, oxygen in red, nitrogen in blue and sulphur in yellow

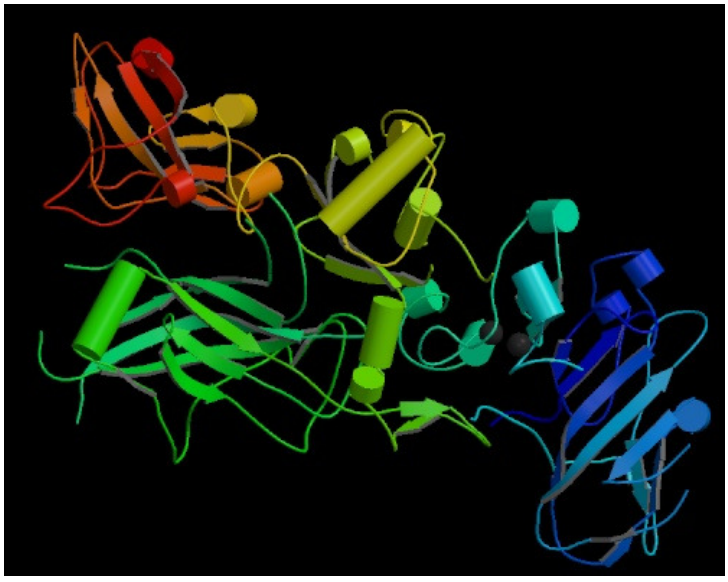

Figure 4 Backbone and secondary structure arrangement of PA from PDB; α-helix represented by cylinders and β-sheet by arrows

The percentages of secondary structural elements in PA are given in Table 1; structures assigned using STRIDE (Heinig and Frishman, 2004), an algorithm for determining secondary structure from the X-ray structure based upon H-bonding pattern and backbone torsion angles.

Table 1 Secondary structure of PA

2° Structure Name	Percentage /%
α-helix	11.3
Isolated beta bridge	3.0
Extended beta strand	33.8
3_{10}-helix	3.3
π-helix	0.00
Turn	28.1
Random coil	20.5

PA is a large protein with a molecular weight of 82667 Daltons. The pH at which PA has zero net charge is theoretically 5.58; calculated from first principles assuming that all applicable species contribute to the charge.

The PA protein is prone to aggregation which is a problem for processing and functionality. Therefore, the long term goal of the work is to examine the effect of environmental factors and additives that influence aggregation. The aim is to guide improvements in processing, formulation and storage of PA in order to reduce the degradation. The preliminary work which will be discussed here examines:

- Thermal behaviour: using micro differential scanning calorimetry
- Secondary structure changes: using infrared spectroscopy

Methods and Materials

The instruments used to examine PA structure and behaviour were a micro differential scanning calorimeter (DSC) and an infrared spectroscope (IR).
A Setaram Micro DSC III was used to monitor thermal events. The instrument is capable of a temperature range of −20 to 120 °C, scanning rates between 0.001 to 1.2 °C min^{-1}, a maximum sample volume of 850 µl and a resolution of 0.03 µW. Samples of 0.5 ml were used and de-ionised water was used as the reference. A scan rate of 0.5 °C min^{-1} was used.
The IR spectroscope used was a Nicolet Avitar 360 FTIR spectroscope, capable of scanning from 400 to 4000 cm^{-1} at a resolution of 1 to 32 cm^{-1}. The instrument was used in transmission mode with calcium fluoride windows and a 6 µm Mylar spacer used to create a thin film of sample through which the IR radiation passes. A resolution of 4 cm^{-1} and a range of 1100 to 4000 cm^{-1} were used.
The protein samples were provided by Avecia and contained phosphate buffered saline pH adjusted to 7.4, they also contained Tween-20, a non-ionic surfactant, at less than 0.1 %v/v. The samples were stored at −18 °C and used as supplied. The concentration of PA in the sample used for micro DSC was 1.9 g l^{-1}, allowing the subtle thermal events to be monitored; a higher concentration might saturate the heat flow capability of the instrument. The concentration of PA in the sample used for IR was 18 g l^{-1} to allow the protein contributions to be observed in the spectrum; the technique is limited to a concentration of at least 10 g l^{-1} when using H_2O.
The protein concentration was determined using ultraviolet (UV) absorbance at 280 nm in 8 M urea. An absorbance coefficient of 75845 M^{-1} cm^{-1} was determined from the primary structure of PA and the UV absorbance of model compounds in 8 M urea reported elsewhere (Pace et al., 1995).

Results

The 1.9 g l^{-1} PA sample was subjected to two heating and cooling cycles and results are given in Figure 5. An exothermic event was observed during the heating on the first cycle at around 47 °C, however, the same was not observed on the second cycle thus suggesting it is an irreversible thermal

denaturation, an event for which similar results are reported elsewhere (Kudou et al., 2003). Initially, the sample appeared clear but once heat treated the sample appeared white suggesting the formation of a gel. The enthalpy change of the denaturation was 6 joules per gram of protein. The temperature and enthalpy of the transition during the heating on the first cycle were reproducible. The endothermic event at around 65 °C present during the cooling on both cycles is attributed to the behaviour of the gel and, as yet, has not been investigated further.

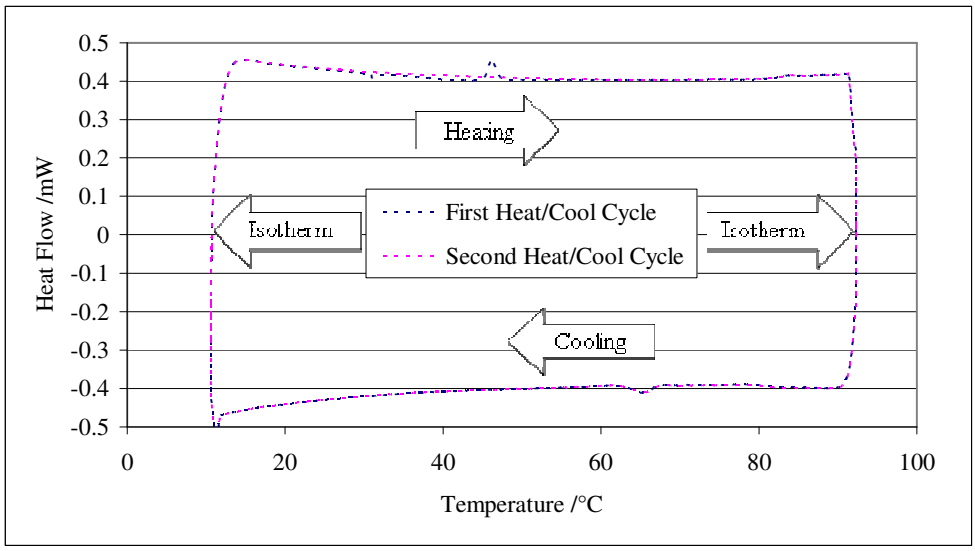

Figure 5 DSC trace of 1.9 g l^{-1} PA sample being heated and cooled at 0.5 °C min^{-1}

The 18 g l^{-1} PA sample was heated and cooled once in the IR. A spectrum was taken at various temperatures; the spectrum for 15 °C is shown in Figure 6. The heating rate cannot be quantified since the heating was stepwise; each spectrum taken once the new temperature had stabilised.

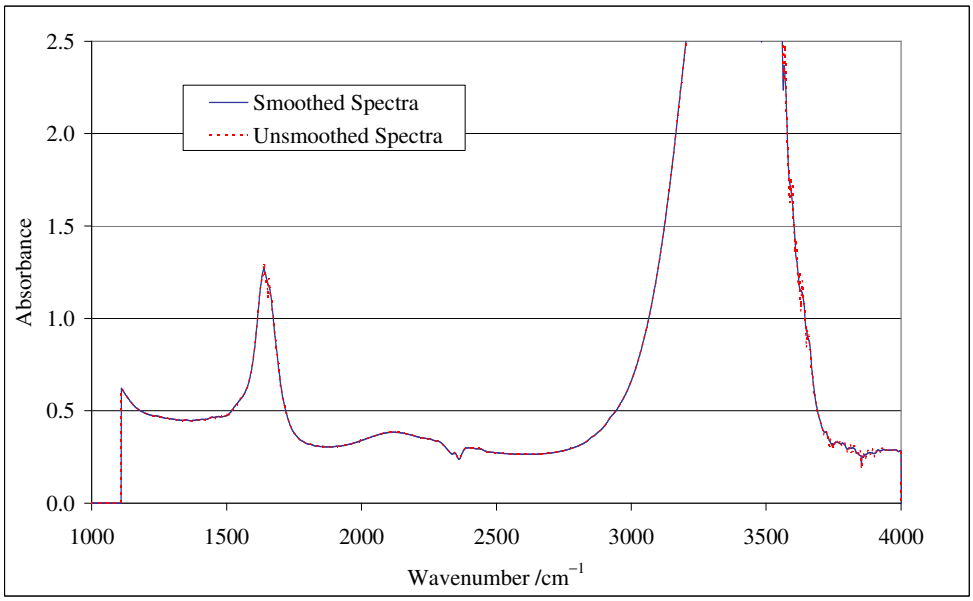

Figure 6 Full IR spectrum of 18 g l^{-1} PA sample at 15 °C, unsmoothed and smoothed

The spectrum in Figure 6 is typical of water, which strongly absorbs IR radiation, and only a small contribution from the protein. This presents a challenge for examining protein in solution particularly since the important protein information is located around 1600 to 1700 cm^{-1} were there is a strong water absorption peak. The water contribution can be subtracted using a spectrum of sample without the protein present. This method is problematic since the water spectrum is affected by pH and

temperature, which must be exactly matched between samples in order to accurately retrieve the protein contributions. In order to avoid these difficulties, the water contribution was subtracted using a new method devised by the author, which uses a gauss curve to model the contribution in the region important for protein structure, see Figure 7.

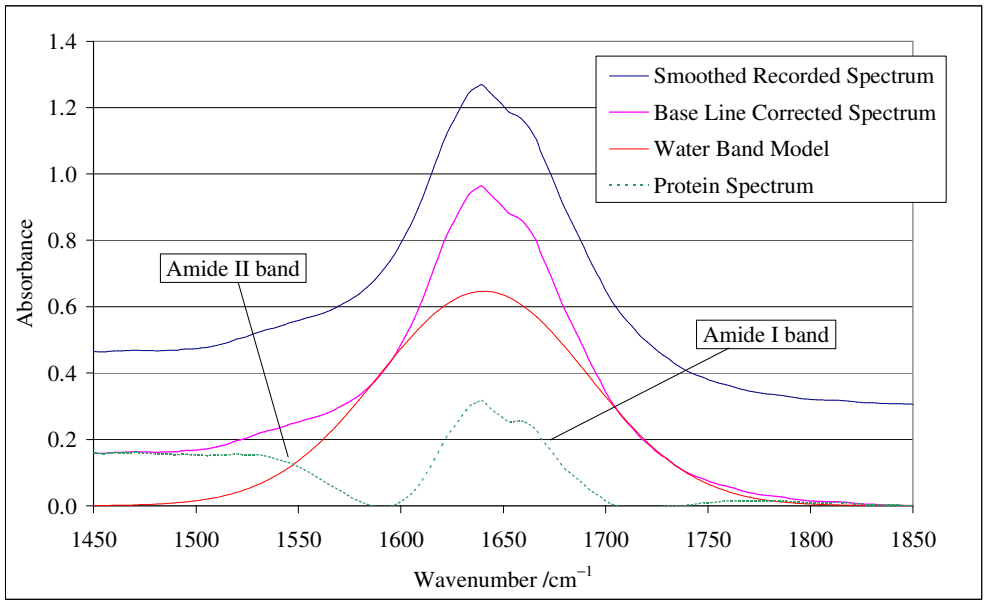

Figure 7 IR spectrum of 18 g l^{-1} PA sample at 15 °C with baseline and water band subtraction

The gauss curve used to model the water band was fitted by adjusting its width, height and position so that it matched the recorded data in a region, 1700 to 1750 cm^{-1}, where the protein does not absorb but where the water does. The subtraction revealed amide I band of the protein information which can be used to quantify secondary structure (Goormaghtigh et al., 1994, Barth and Zscherp, 2002). The amide bands arise from vibrations of bonds in the polypeptide backbone and coupling of these vibrations make them sensitive to the configuration of the backbone. Gauss curves were fitted to the amide I band, 1600 to 1700 cm^{-1}, to determine secondary structure, see Figure 8; the position of a component curve relates the secondary structure type as shown by theory and experiment and the percentage area of the component curve directly relates to the quantity of the particular structure. The curves were fitted by locating their positions with fine structure enhancement (FSE) (Barth, 2000), the height and width of the curves was iteratively adjusted until the sum of all the curves matched the recorded spectrum.

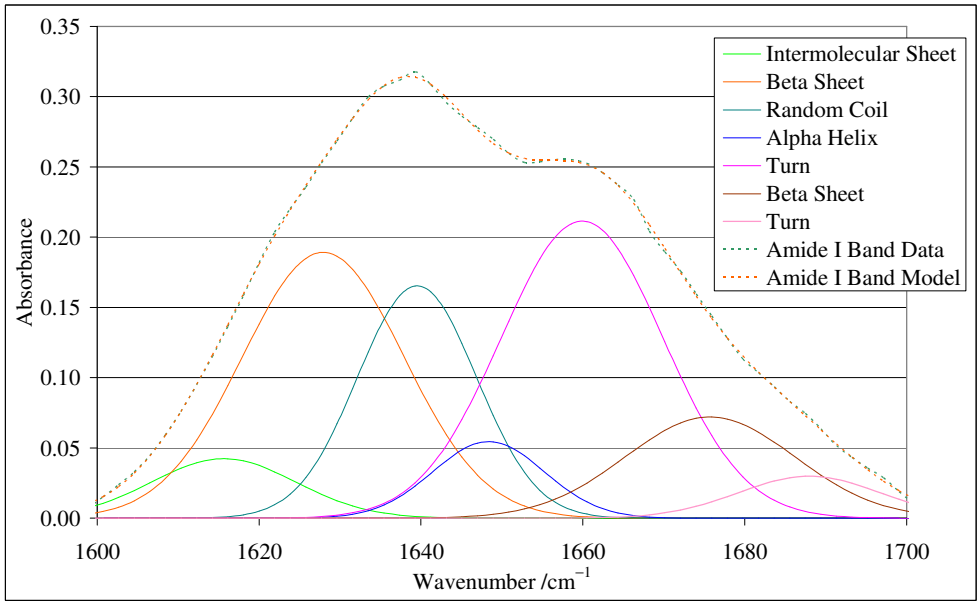

Figure 8 Amide I band data and model with component curves for 18 g l^{-1} PA sample at 15 °C

The quantities of secondary structures identified for the spectrum at 15 °C are 5.4 % α-helix and 37.8 % β-sheet and 34.0 % turns; all structure assignments with percentages are given in Table 2. These IR results are comparable to the quantities determined by the X-ray data if it is assumed that the intermolecular β-sheet has been formed by random coil and if it is assumed that the 3_{10}-helix is contributing to the component curve for the turns.

Table 2 Secondary structure of 18 g l^{-1} PA sample from IR spectrum at 15°C

2° Structure Name	Percentage /%
Intermolecular β-Sheet	5.4
β-Sheet	37.8
Random Coil	17.4
α-Helix	5.4
Turn	34.0

The quantity of secondary structure has also been determined using the spectra taken at higher temperatures during heating; these are represented graphically in Figure 9. The results for cooling are not shown.

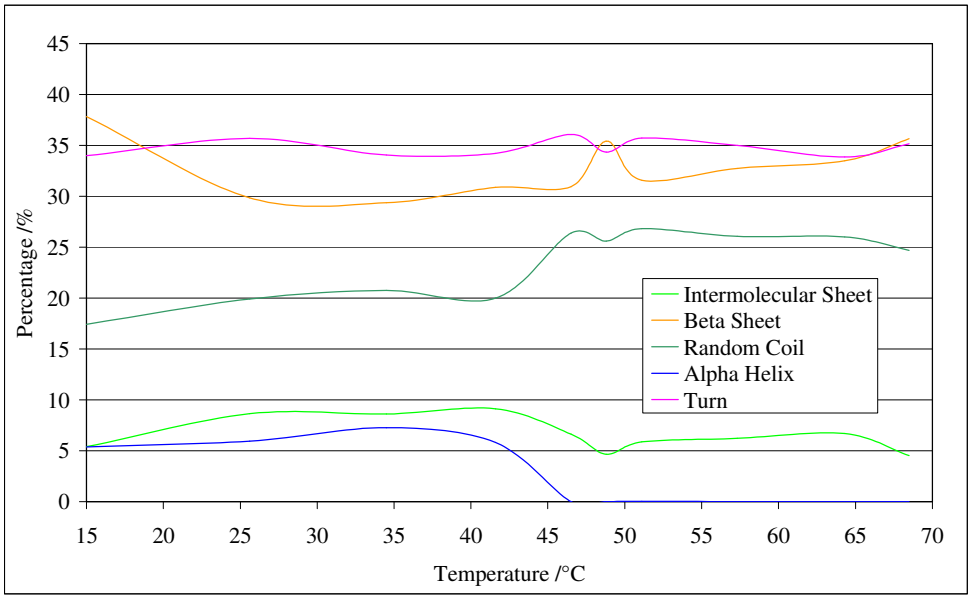

Figure 9 Percentages of secondary structure of 18 g l^{-1} PA sample as a function of temperature

It can be seen in Figure 9 that as the temperature increases there are alterations to secondary structure, most notably at around 48 °C where there is multiple shifts in the structure quantities and also notably for α-helix which increases in quantity at around 30 °C and is then completely lost around 46 °C.

Discussion

The change in appearance of the PA sample subsequent to heating indicates the formation of a gel, which is a large structure built up by aggregation. The infrared results show that aggregation was present even prior to heating as indicated by the presence of intermolecular β-sheet. Aggregation is a degradation event that can severely reduce product quality; reducing biological activity and altering immunogenicity (Wang, 1999). Investigating the prevention of aggregation is therefore an important

requirement for the formulation of protein pharmaceuticals, a task to which IR is able to lend itself given that it is able to detect aggregation without interfering with the sample.
The IR event at around 48 °C coincides with thermal denaturation observed with DSC. The secondary structure was not completely lost; there was however a substantial increase in random coil; a change that results from the break down of ordered secondary structures. Given that the correct structure is important for biological activity and that denaturation is the disruption of structure resulting in a loss of activity it would appear that the changes in secondary structure quantity around 48 °C are consistent with a denaturation event. It is important to observe structural changes that occur with degradation events in order to improve protein pharmaceutical stability. IR has demonstrated a capability that could be extended and used to investigate conditions for improving product stability.
The gain and loss of α-helix content observed with increasing temperature coincides with the gain and loss of biological activity with increasing temperature as demonstrated elsewhere (Radha et al., 1996), the same work also reports that PA shows an increase α-helix content coinciding with the gain of biological activity. This suggests that the presence of the α-helix could be important for biological activity and therefore, might be important for the correct immunogenicity of PA in a vaccine. Understanding structural features required for activity is important for formulating effective protein pharmaceuticals, another application for which IR can be used.

Conclusions

The work has demonstrated that IR and DSC can be used to examine the degradation of protein products. The methods demonstrated and conditions surveyed have gone some way to meeting the stated aims of improving processing, formulation and storage. The work has exemplified the effect of temperature on the product and opened up opportunities for other environmental factors to be examined in relation to protein degradation and this will form the basis of future work.

Acknowledgements

The authors wish to thank EPSRC and Avecia for funding, Allan Watkinson for supplying PA samples and Geoff Parr for guidance.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) *Molecular biology of the cell,* Garland Science, New York.

Allaby, M. (1998) *A Dictionary of Plant Sciences,* Oxford University Press.

Barth, A. (2000) *Spectrochimica Acta Part A,* **56,** 1223-1232.

Barth, A. and Zscherp, C. (2002) *Quarterly Reviews of Biophysics,* **35,** 369–430.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) *Nucleic Acids Research,* **28,** 235-242.

Branden, C. and Tooze, J. (1991) *Introduction to protein structure,* Garland Publishing, New York.

Brangex, J., Andersen, L., Laursen, E. D., Meyn, G. and Rasmussen, E. (1997) *Journal of Pharmaceutical Sciences,* **86,** 517 - 525.

Clark, A. H., Kavanaghb, G. M. and Ross-Murphy, S. B. (2001) *Food Hydrocolloids,* **15,** 383-400.

Clarkson, J. R., Cui, Z. F. and Darton, R. C. (1999) *Journal of Colloid and Interface Science,* **215,** 323–332.

Gillis, J. (2004) In *Washington Post*, pp. A01.

Goormaghtigh, E., Cabiaux, V. and Ruysschaert, J. (1994) *Subcellular Biochemistry,* **23,** 329-362.

Gosal, W. S. and Ross-Murphy, S. B. (2000) *Current Opinion in Colloid and Interface Science,* **5,** 188-194.

Heinig, M. and Frishman, D. (2004) *Nucleic Acids Research,* **32,** W500-W502.

IUPAC (1970) *Biochemistry,* **9,** 3471-3479.

Jendrek, S., Little, S. F., Hemc, S. and Giardina, G. M. (2003) *Vaccine,* **21,** 3011–3018.

Jones, S. and Thorton, J. M. (1995) *Progress in Biophysics and Molecular Biology,* **63,** 31-65.

Kabsch, W. and Sander, C. (1983) *Biopolymers,* **22,** 2577-2637.

Kudou, M., Shiraki, K., Fujiwara, S., Imanaka, T. and Takagi, M. (2003) *European Journal of Biochemistry,* **270,** 4547-4554.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) *Protein Science,* **4,** 2411-2423.

Pande, V. S. and Rokhsar, D. S. (1998) *Proc. Natl. Acad. Sci. USA,* **95,** 1490–1494.

Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. and Liddington, R. C. (1997) *Nature,* **385,** 833-838.

Pittman, P. R., Gibbs, P. H., Cannon, T. L. and Friedlander, A. M. (2001) *Vaccine,* **20,** 972-978.

Prusiner, S. B. (1998) *Proc. Natl. Acad. Sci. USA,* **95,** 13363–13383.

Radha, C., Salotra, P., Bhat, R. and Bhatnagar, R. (1996) *Journal of Biotechnology,* **50,** 235-242.

Ramachandran, G. N. and Sasisekharan, V. (1968) *Advances in Protein Chemistry,* **23,** 283-438.

Saitoh, S., Akashi, S., Yamada, T., Tanimura, N., Kobayashi, M., Konno, K., Matsumoto, F., Fukase, K., Kusumoto, S., Nagai, Y., Kusumoto, Y., Kosugi, A. and Miyake, K. (2004) *International Immunology,* **16,** 961-969.

Stubbs, M. T. (2002) *Trends in Pharmacological Sciences,* **23,** 539-541.

Uversky, V. N. (2002) *Protein Science,* **11,** 739–756.

Wang, W. (1999) *International Journal of Pharmaceutics,* **185,** 129-188.

Young, J. A. T. and Collier, R. J. (2002) *Scientific American,* **286,** 48-50, 54-59.

Zanuy, D., Ma, B. and Nussinov, R. (2003) *Biophysical Journal,* **84,** 1884–1894.