

FTIR Analysis of Protein Structure

Warren Gallagher

A. Introduction to protein structure

The first structures of proteins at an atomic resolution were determined in the late 1950's.¹ From that time to the early 1990's approximately 300 protein structures were added to the list, primarily using X-ray crystallography. Currently over 20,000 structures are solved,² some of these using the newer technique of multidimensional nuclear magnetic resonance spectroscopy (NMR). While X-ray crystallography and NMR spectroscopy provide the greatest level of detail about a protein's structure, there are many situations where these techniques cannot be applied. Other techniques, such as FTIR, may not provide the same level of structural detail, but can be applied readily to further our understanding of how proteins function

Before describing the structural information that FTIR spectroscopy provides, a brief introduction to protein structure is given. Proteins are linear biological polymers for which the monomeric units are amino acids. (**Figure 1**). Twenty different amino acids are used to make proteins, each distinguished by the identity of the "R" group. The amino acids are linked to form a polymer by linking the amino group on one amino acid with the carboxylic acid group on another amino acid to form an amide bond (Figure 2).

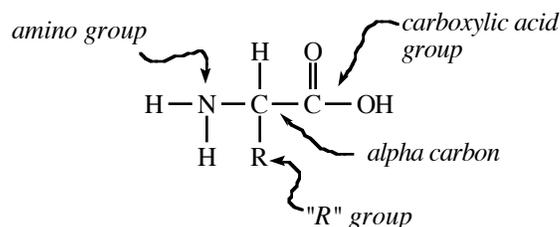


Figure 1: The structure of an amino acid; the basic building block for making proteins.

Protein chemists call the amide bond a *peptide bond*. When two amino acids are linked together by an amide bond it is called a *dipeptide*; and when many amino acids are linked together in this fashion it is called a *polypeptide*. A polypeptide consists of a *backbone* and *sidechains*. The backbone comprises the amide nitrogen, the alpha carbon and the carbonyl carbon that are contributed by each amino acid unit. The sidechains comprise the "R" groups, and dangle from the backbone like charms from a bracelet.

¹ Kendrew *et al.* (1958) *Nature* **18**, 662-666.

² Protein Data Bank, Brookhaven National Laboratory, "<http://ftp.pdb.bnl.gov>", May 16, 1997

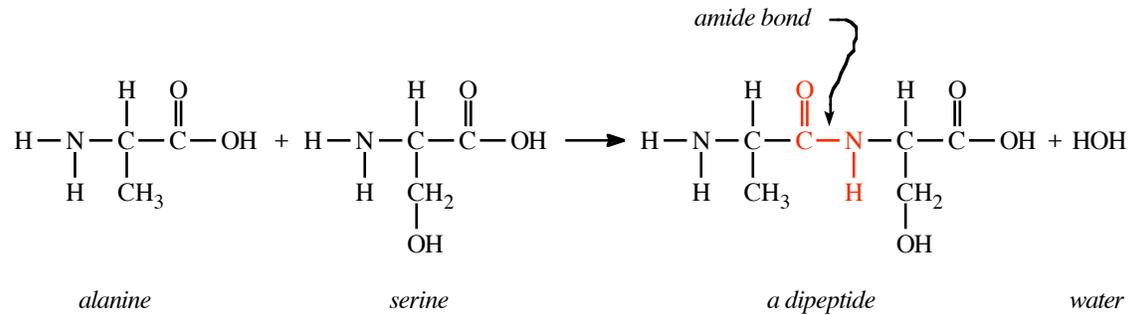


Figure 2: Linking two amino acids together by an amide bond to form a dipeptide.

Proteins are distinguished from one another by the number of amino acids units they contain and by the identity and sequential order of the amino acids. The numbers vary from 50 to hundreds of amino acid units, making the number of possible sequences astronomical. Both the number of amino acids in a protein and their sequence are genetically determined. Few polypeptides are considered proteins. For a polypeptide to be considered a protein it must be able to fold into a well-defined 3-dimensional structure. This is usually a requirement for protein function.

When a protein folds to form a well defined 3-dimensional structure it exhibits three levels of structure: *primary*, *secondary* and *tertiary*. The genetically determined sequence of amino acids is the *primary structure*. The primary structure is often modeled as beads on a string, where each bead represents one amino acid unit. The *tertiary structure* might be modeled as a tightly-packed snowball, where each atom in the protein has a well defined location. Protein folding might be likened to scrunching up the string of beads (the primary structure) into a tightly-packed ball (the tertiary structure) (**Figure 3a**). A problem that proteins have in forming a tertiary structure is that many of the amino acids find themselves on the inside of the tightly-packed ball and no longer in direct contact with water. However, amides like to be exposed to water, because they can form “*hydrogen bonds*” with water (**Figure 3b**). This problem is solved by providing alternative ways for the buried amides to hydrogen bond. Primary among these is to have the amide bonds hydrogen bond to each other (**Figure 4**).

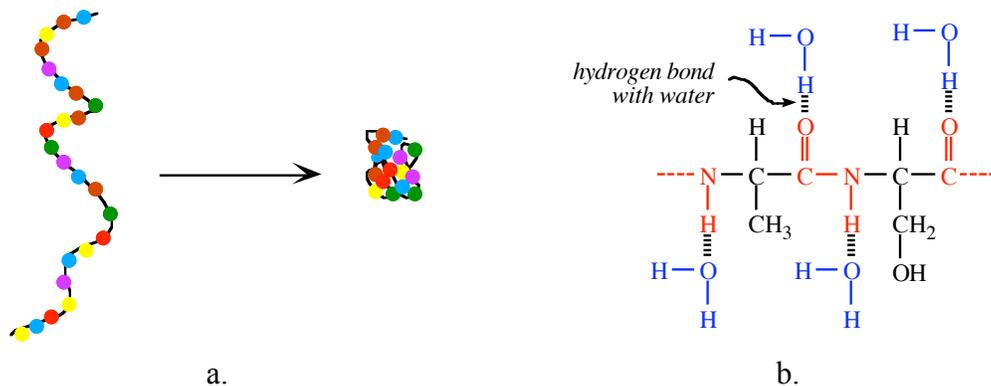


Figure 3: a. The crudest of protein folding models: scrunching up a string of beads. b. The hydrogen bonds that form between the amide bond and water.

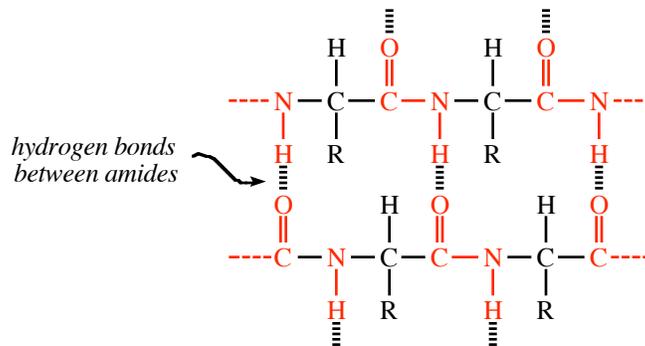


Figure 4: The hydrogen bonds that form between amide bonds buried inside a folded protein.

This leads to an intermediate level of protein structure called *secondary structure*. The types of secondary structure includes the α helices and β sheets, which allow the amides to hydrogen bond very efficiently with one another. Both are periodic structures. In an α -helix the polypeptide backbone is coiled in a right-handed helix where the hydrogen bonding occurs between successive turns of the helix. **Figure 5(left)** shows a segment of α helix from the small protein *bovine pancreatic trypsin inhibitor* (BPTI). In β sheets, the strands of polypeptide are stretched out and lie either parallel or antiparallel to one another. The hydrogen bonds form between the strands. **Figure 5(right)** illustrates this with a piece of antiparallel β -sheet from BPTI. The other elements of secondary structure include β turns and *unordered* structure. β turns are sharp turns that connect the adjacent strands in an antiparallel β sheet. Unordered structure is generally a catch-all for regions that do not fall into one of the other categories. These are often loops which form near the surface of proteins and join the other elements of secondary structure.

Tertiary structure arises when various elements of secondary structure pack tightly together to form the well-defined 3dimensional structure. The tertiary structure of a protein is defined by the spacial coordinates of all its atoms, held together by favorable interactions between the sidechains. These interactions are weak, which is why proteins are easily “denatured” by heat and exposure to certain chemicals. The sidechains buried inside a folded protein are packed tightly together and must interact favorably in order to remain folded. Few

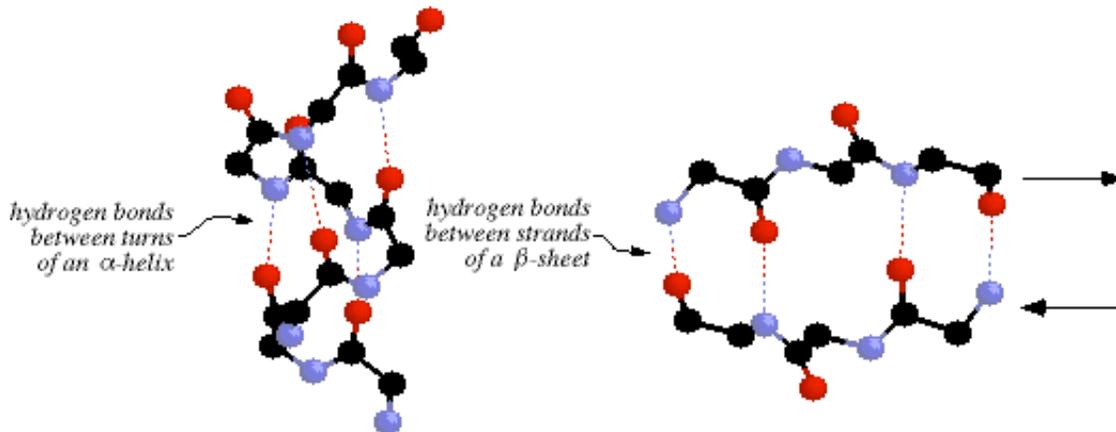


Figure 5: Segments of an α -helix (left), and a β -sheet (right) taken from BPTI. The hydrogens and sidechains have been omitted for clarity. Carbons are black, oxygens are red and nitrogens are blue. The hydrogen bonds between the amide oxygens and nitrogens are shown as dotted lines.

of the possible polypeptide sequences meet these stringent requirements for folding into the well-defined 3-dimensional structures. **Figure 6** shows both a spacefilling and a ribbon model for BPTI. In the space-filling model the atoms are represented by spheres scaled to their true relative sizes, illustrating the tightly packed folded protein. The ribbon model shows the polypeptide backbone tracing its way through the folded structure. It highlights the various types of secondary structure.

B. Obtaining structural information about proteins using FTIR spectroscopy

FTIR spectroscopy provides information about the secondary structure content of proteins, unlike X-ray crystallography and NMR spectroscopy which provide information about the tertiary structure. FTIR spectroscopy works by shining infrared radiation on a sample and seeing which wavelengths of radiation in the infrared region of the spectrum are absorbed by the sample. Each compound has a characteristic set of absorption bands in its infrared spectrum. Characteristic bands found in the infrared spectra of proteins and polypeptides include the Amide I and Amide II. These arise from the amide bonds that link the amino acids. The absorption associated with the Amide I band leads to stretching vibrations of the C=O bond of the amide, absorption associated with the Amide II band leads primarily to bending vibrations of the N—H bond (**Figure 7**). Because both the C=O and the N—H bonds are involved in the hydrogen bonding that takes place between the different elements of secondary structure, the locations of both the Amide I and Amide II bands are sensitive to the secondary structure content of a protein. Studies with proteins of known structure have been used to correlate systematically the shape of the Amide I band to secondary structure content.^{3,4} The Amide II band, though sensitive

³ Byler, D.M. and Susi, H. (1986) *Biopolymers* **25**, 469-487.

⁴ Surewicz, W.K. and Mantsch, HH. (1988) *Biochim. Biophys. Acta* **952**, 115-130.

to secondary structure content, is not as good a predictor for quantitating the secondary structure of proteins. One difficulty with analyzing the Amide I band for secondary structure is that the shifts in

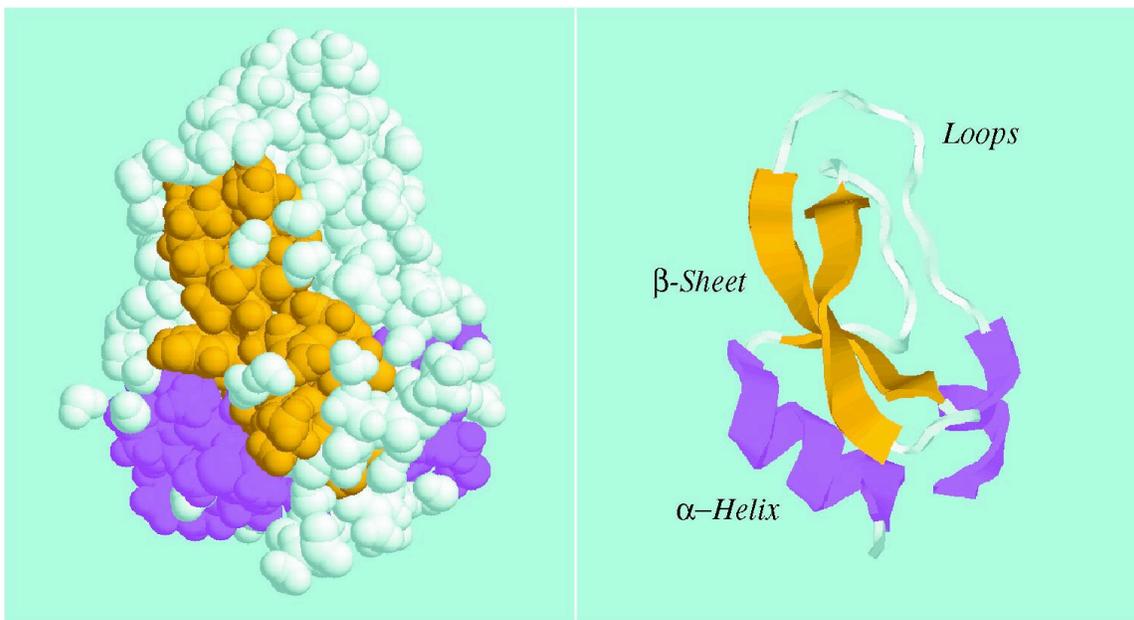


Figure 6: Bovine pancreatic trypsin inhibitor. Left, spacefilling model; right, ribbon model. Colors are used to label α -helix (purple), β -sheet (yellow), and loops (white).

the Amide I band are small compared to the intrinsic width of the band. Instead of a series of nicely resolved peaks for each type of secondary structure, one broad lumpy peak is observed (**Figure 8a**). Several numerical methods are used to increase the apparent resolution of the Amide I band so that estimates can be made of the secondary structure content. **Figure 8b** shows results we obtained for the Amide I band of BPTI using the Fourier self-deconvolution method developed by Kauppenin *et al.*⁵

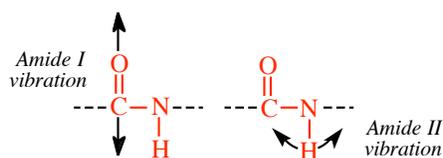


Figure 7: The vibrations responsible for the Amide I and Amide II bands in the infrared spectra of proteins and polypeptides. The Amide I band is due to carbonyl stretching vibrations while the Amide II is due primarily to NH bending vibrations.

C. Monitoring hydrogen isotope exchange by FTIR spectroscopy

When a protein is dissolved in water the hydrogens attached to the amide nitrogen can readily exchange for hydrogens attached to water molecules. For unfolded

⁵ Kauppinen, J.K.; Moffatt, D.J.; Mantsch, H.H. and Cameron, D.G. (1981) *Appl. Spectr.* **35**, 271-276.

polypeptides this exchange can occur thousands of times a second. When a polypeptide folds to form the well-defined 3dimensional structure of a protein, many of the amide hydrogens become buried in the interior of the protein and no longer come into direct contact with the solvent water. This causes a marked decrease in the rate of exchange for these hydrogens. The most deeply buried amide hydrogens can take weeks, even months, to exchange. The fact that they do exchange eventually is taken as evidence that folded proteins are somewhat flexible, *i.e.*, their structures fluctuate with time to allow buried regions eventual access to the solvent. Hydrogen isotope exchange is becoming a powerful tool for studying the protein folding process.⁶

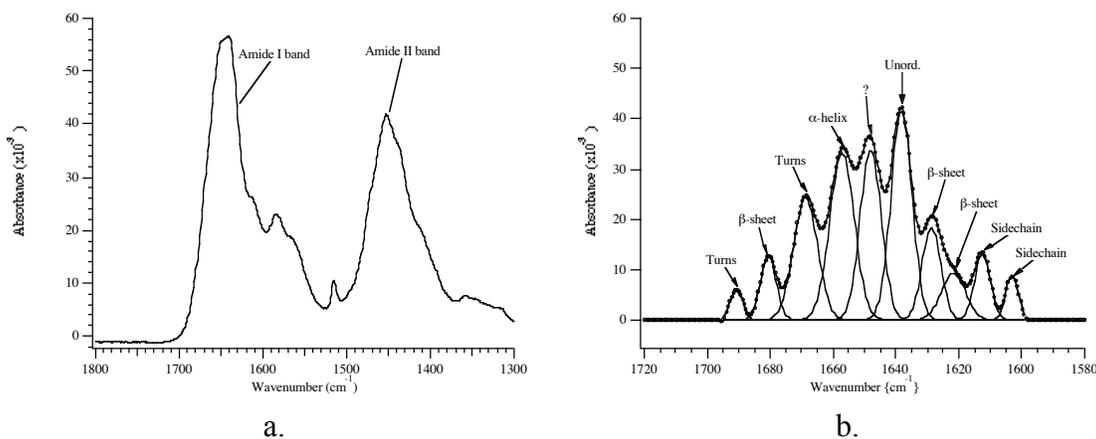


Figure 8: a. FTIR spectrum of bovine pancreatic trypsin inhibitor showing the Amide I and Amide II bands. b. Results of a Fourier self-deconvolution analysis on the Amide I band for bovine pancreatic trypsin inhibitor. The assignments of the various peaks to the different elements of secondary structure are shown.

To observe the rate of hydrogen exchange a different isotope of hydrogen is used for the solvent water than that initially present on the protein. For our experiments we use heavy water, which contains the deuterium isotope of hydrogen (^2H) instead of the more abundant ^1H isotope of hydrogen. The ^2H isotope has nearly identical chemical properties to that of the ^1H isotope, but twice the mass. Like adding mass to the end of a pendulum, the exchange of the ^1H isotope for the ^2H isotope affects the vibrational frequencies of the amide bond and therefore affect the Amide I and Amide II bands of the infrared spectrum of a protein. **Figure 9** shows the results of a hydrogen isotope exchange experiment on BPTI, monitored using FTIR..

We are carrying out a systematic study to see if we can resolve hydrogen isotope exchange for the different types of secondary structure using FTIR spectroscopy. We chose to do our initial study on BPTI for several reasons: it is a very stable protein that is easy to work with; it contains representative examples of each of the major elements of

⁶ Kunihiro, K., Kim, P. and Baldwin, R.L. (1984) *Biopolymers* **22**, 59-67.

secondary structure (see **Figure 6(right)**); and extensive hydrogen isotope exchange experiments were carried out on this protein using NMR spectroscopy, so that exchange rates are available for nearly every amide hydrogen in the protein.⁷ Results of an experiment carried out at *pH* 6.04 and 25°C are shown in **Figure 10a**. To see which elements of secondary structure were affected the most during this experiment, **Figure 11** plots the relative intensity of each peak as a function of time along with assignments for each. A logarithmic scale is used for time scale because the exchange is an exponentially decaying process. From this experiment we made peak assignments for both the hydrogenated and deuterated forms of each of the major elements of secondary structure. As shown in **Figure 11**, the assigned peaks mirror each other's behavior during the exchange process. The elements exhibiting the greatest change in this experiment are the unordered and α -helical elements. The turns and β -sheet elements show the least change. This is expected for BPTI. The turns are located at the surface of the protein and most likely completed their exchange before the first time point was taken. Conversely, the β -sheet in BPTI is buried in the core of the protein, so most of the β -sheet hydrogens should have exchanged very little by the end of this experiment. Under the conditions of *pH* and temperature used in this experiment these hydrogens should take years to exchange. However, if the temperature is raised to near the boiling point, exchange from all sites is complete within minutes. Many

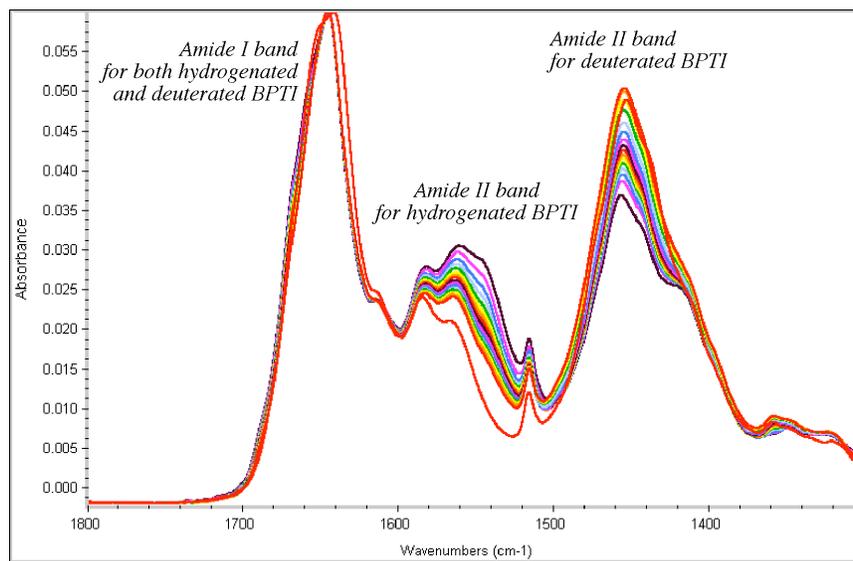


Figure 9: Time course for hydrogen/deuterium from BPTI, from 3 minutes to 374 minutes after initiating the exchange. The Amide II band at 1450 cm^{-1} is for deuterated BPTI and is increasing with time. The

⁷ Hilton, B.D., and Woodward, C.K. (1978) *Biochemistry* **17**, 3325-3332; Hilton, B.D., and Woodward, C.K. (1979) *Biochemistry* **18**, 5834-5841; Wüthrich, K. and Wagner, G. (1979) *J. Mol. Biol.* **130**, 1-18; Richarz, R., Sehr, P., Wagner, G. and Wüthrich, K. (1979) *J. Mol. Biol.* **130**, 19-30; Wüthrich, K. and Wagner, G. (1979) *J. Mol. Biol.* **130**, 31-37; Tüchsen, E. and Woodward, C. (1985) *J. Mol. Biol.* **185**, 405-419; Tüchsen, E. and Woodward, C. (1985) *J. Mol. Biol.* **185**, 421-430; Roder, H., Wagner, G. and Wüthrich, K. (1985) *Biochemistry* **24**, 7396-7407; Tüchsen, E. and Woodward, C. (1987) *J. Mol. Biol.* **193**, 793-802; Gallagher, W.H., Tao, F., and Woodward, C. (1992) *Biochemistry* **31**, 4673-4680

Amide II band at 1550 cm^{-1} is for hydrogenated BPTI and is decreasing with time. The bottom red spectrum at 1550 cm^{-1} was obtained after complete exchange; the residual absorbency in this region is due to groups on the protein other than the amide group.

changes occur between the last time point taken in the experiment carried out at pH 6.04 and 25°C and an infinite time when the exchange will be complete (see **Figure 10b**). During this time hydrogen isotope exchange is expected to be primarily from the β -sheet region of the BPTI.

We propose to extend these studies to look at the effects that hydrogen isotope exchange from the β -sheet of BPTI has on the Amide I band. This will be done by carrying out the experiments at higher pH values. At pH values above 5, hydrogen isotope exchange is base catalyzed. Exchange rates are expected to increase 10fold for every 1 unit increase in the pH . This will allow us to bring exchange from the β -sheet into the time window accessible with our procedures.