Coupled vibrations in peptides and proteins: Structural information using 2D-IR spectroscopy
Huerta Viga, A.

Citation for published version (APA):
Predicting the folded structure of a protein from its sequence of amino acids requires a detailed understanding of the fundamental principles that drive protein folding, and obtaining this understanding is an ongoing challenge of several scientific disciplines. Here we present the basic structural features of a protein, with emphasis on secondary structure and long-range contacts, and we introduce vibrational spectroscopy, and 2D-IR spectroscopy in particular, as a powerful tool to investigate these structural features. We describe the vibrations of the backbone and how they are determined by the structure of the polypeptide chain. We also describe vibrations of some of the side chains that are ‘useful’ as structural probes, in particular those that can be involved in salt bridges.

1.1 The Protein Folding Problem

A protein is an unbranched chain of amino acids linked by peptide bonds. Amino acids are the building blocks of proteins, and they are known as ‘residues’ when they are part of one. Figure 1.1 shows part of the backbone of a protein in which two peptide bonds are highlighted and the positions of the side chains (or side groups) are only indicated. The peptide bond is planar because the C-N bond has partial double bond character.¹ The single bonds on the carbon atom that connect the two peptide groups can rotate, and the corresponding dihedral angles are generally denoted by \( \varphi \) and \( \psi \). The side groups of the amino acids are the source of the diversity of the
structures that proteins adopt during their self-assembly process. Their size, for example, restricts the possible values that the dihedral angles can adopt, giving rise to characteristic combinations of angles for different structural elements (see Chapter 6). Most proteins need to adopt a specific three-dimensional structure before they can function correctly inside the cell, and in general the sequence of amino acids of a single-domain protein uniquely defines this structure, known as native or folded conformation. A protein whose backbone conformation is completely disordered is known as a random coil, but often denatured or ‘unfolded’ proteins contain certain amount of residual structure. How proteins spontaneously self-assemble into their native conformation is a fundamental question of molecular biology, and knowing the answer would also facilitate the understanding of why folding goes ‘wrong’ some times. Many misfolded proteins accumulate in different organs of the body and are an important cause of disease.²

Estimating the time that a protein would need to fold if it had to randomly search through all the possible configurations of its backbone leads to a paradox first noted by Levinthal.³ Assuming three possible configurations for each peptide bond this time would be $\sim 10^{27}$ years, but proteins are known to fold in a milisecond timescale. However, by including a bias against locally unfavorable configurations in the random search, the time a protein would take to fold is brought to biologically-realistic time scales.⁴ The challenge is thus to determine the basic physical and chemical principles that create a bias and therefore drive the folding of a protein.⁵ A complete understanding of these principles would allow the prediction of the three-dimensional structure of a protein from its amino acid sequence, which is known as the protein folding problem and is illustrated in Figure 1.2. Many factors that determine the success of the folding process are still not well understood. Current views on protein-folding pathways describe the folding process as the sequential addition of pieces of secondary structure and tertiary contacts to an initial unstable (but highly populated) conformation.⁶ Furthermore the picture of stable folded proteins is evolving into a dynamical one, in which the folded state is rather metastable and acts as a ‘kinetic hub’.⁷

One of the main driving forces of protein folding is the hydrophobic effect. Some of the amino-acid side chains (such as alanine, valine, leucine, isoleucine, phenylalanine, tryptophan and methionine) are hydrophobic, i.e., they cannot form hydrogen bonds, which keeps them from being exposed to the solvent. This leads to the formation of hydrophobic clusters that are energetically favored. The hydrophobic effect is a key factor in the initial stages of protein folding, where hydrophobic interactions drive the formation of the core of the protein. This core then serves as a platform for the addition of other structural elements, guided by the specific sequence of amino acids. The hydrophobic core acts as a ‘kinetic hub’ where further folding proceeds, facilitating the correct formation of the native structure.

Figure 1.1: Fraction of the backbone of a protein, where two peptide bonds are highlighted in blue. The corresponding transition-dipole moments of two amide I vibrations are indicated with arrows, as well as the positions of the side chains (SC₁).

---

²Some proteins perform function while disordered, and they are known as intrinsically disordered (or unstructured) proteins.
bonds with water molecules around them. As a result, the hydrogen bond network of water rearranges and molecules that constitute the solvation shell effectively slow down their re-orientation dynamics.\textsuperscript{8,9} This effect reduces the entropy of the system, and therefore the hydrophobic side chains of proteins tend to pack against each other, reducing in this manner the hydrophobic surface of the protein that is exposed to the solvent and minimizing the disruption of the hydrogen bond network of water. As a result of the hydrophobic effect, many proteins have regions from which water molecules are completely excluded, known as a hydrophobic core, which stabilize the native state of the protein. Solvent-exposed surfaces on the other hand are often constituted of polar or hydrophilic residues.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{folding_process.png}
\caption{Illustration of the folding process of a protein in which an intermediate structure can be linked to two possible end folds. One labeled as the native state of the protein, and the second as a fibril structure. Random coils are represented in black, $\alpha$-helices in red, and $\beta$-sheets in green.}
\end{figure}

1.2 Vibrational Spectroscopy

Addressing the protein-folding problem requires techniques that are sensitive to the conformation of proteins, and vibrational spectroscopy is well suited for this purpose. The backbone of proteins and some amino acid side chains have vibrational modes whose frequencies and (or) intensities depend on the protein conformation. An important advantage compared to NMR\textsuperscript{†} is that the intrinsic timescale of vibrational spectroscopy is such that different conformers can be detected separately (as long as exchange between them tanks longer than the dephasing rate of the vibrations). Despite the frequency specificity of FTIR spectroscopy, conformational disorder and solvation effects often result in broad, featureless bands. 2D-IR spectroscopy provides additional sensitivity and information about the vibrational couplings (see Chapter 2). Furthermore, it allows the study of picosecond-solvation dynamics of

\textsuperscript{†}Nuclear Magnetic Resonance
1. Introduction

the vibrational modes, and folding experiments that probe structural changes from nanoseconds to milliseconds can be easily implemented. An other strength of vibrational spectroscopy is that by using isotopic labels, the vibrational frequencies of selected vibrations can be shifted to lower frequencies in order to eliminate spectral crowding or to obtain site-specific structural information (see Chapter 7 for a less conventional usage of an isotopic label).

1.3 Secondary Structure

Backbone vibrations give rise to the amide I band, which is very sensitive to the secondary structure of proteins and is therefore very commonly used for secondary-structure analyses. The amide I band arises mainly from the C=O stretching vibration of the backbone, with minor contributions from the out-of-phase CN stretching vibration, the CCN deformation and the NH in-plane bend.\(^9\) The most abundant structural element in proteins is the \(\alpha\)-helix. It is a periodic structure with local-sequence character that is stabilized by intra-helical hydrogen bonds between amino acids that are four units apart in the sequence. These hydrogen bonds are formed by an NH group of residue \(i\) and a CO group of a residue \(i + 4\), resulting in a helical backbone conformation with 3.6 residues per turn. Some amino acids have a natural helix propensity, and particularly alanine and leucine are encountered in \(\alpha\)-helical structures.\(^{11}\) \(\alpha\)-helices tend to bundle together or pack against each other in order to bury hydrophobic side chains on their surfaces.\(^{12}\) Bundles of \(\alpha\)-helices known as coiled-coils are often encountered in membrane proteins and act as ion channels.\(^{13}\) In Chapter 5 we investigate one of such coiled-coils. The individual backbone amide I vibrations of an \(\alpha\)-helix are coupled, resulting in a collective vibration of the helix, or more precisely, in vibrational modes that are delocalized along the helix. The vibrational delocalization gives rise to two characteristic vibrations, whose transition dipole moment is oriented parallel and perpendicular to the helix. The frequency of these vibrations is illustrated in Figure 1.3. In the linear absorption spectrum, these two modes overlap and cannot be separated, but with 2D-IR spectroscopy it is possible to detect the coupling between them (see Chapter 7).

The second important secondary-structure element of proteins are \(\beta\)-sheets, which are less compact than helices. They are held together by hydrogen bonds between adjacent strands, which can be located in non-adjacent segments of the protein. \(\beta\)-

![Figure 1.3: Infrared-absorption bands which arise from delocalized amide I modes of the backbone. The basic secondary-structure elements of proteins, \(\alpha\)-helices, \(\beta\)-sheets and random coils, give rise to bands at the indicated frequencies.](image-url)
sheets are usually further stabilized by hydrophobicity, as they allow packing of hydrophobic side chains against each other above and below the sheets.\textsuperscript{12} Depending on the relative orientation of the strands, β-sheets can be parallel, and antiparallel (the later are always formed by non-adjacent segments of the protein). β-sheets also give rise to two coupled vibrational modes whose transition-dipole moments are oriented parallel and perpendicular to the direction of the sheets,\textsuperscript{10} and the separation between these modes and their relative intensity is determined by the size of the β-sheets.\textsuperscript{14} Secondary structure elements often group together to form specific geometric arrangements or super-secondary structure elements, which are known as structural ‘motifs’. In this thesis we study two of such motifs, a coiled coil (Chapter 5) and a zinc finger (Chapter 8).

1.4 Tertiary Structure

The tertiary structure of a protein is determined by how helices and β-sheets are spatially arranged with respect to each other. Vibrations of the side chains can provide tertiary-structure probes because their vibrational frequencies and 2D-IR response vary upon changes in local environment. Figure 1.4 shows an example of infrared absorption in a range that includes the amide I’ and amide II’ bands. H\textsubscript{2}O has a strong absorption band at the same frequency as the amide I mode, which limits the thickness of the samples that can be used. To overcome this problem D\textsubscript{2}O is commonly used as a solvent, and the labile hydrogens will exchange by deuterium, leading to shifts in the amide I and amide II bands, the deuterated versions of which are labeled with a prime. The amide II band arises from backbone vibrations, mainly an out-of-phase combination of the NH in-plane bend and the CN stretching vibration.\textsuperscript{10} This band is insensitive to the secondary structure of proteins, but it is very sensitive to the hydrogen-deuterium exchange, resulting in a red shift of about 100 cm\textsuperscript{−1}. Therefore, the amide II’ is a sensitive probe of solvent exposure and of the presence of hydrophobic cores,\textsuperscript{15} see Chapter 9 for an example. Figure 1.4 shows the frequency at which several amino-acid side chains absorb.\textsuperscript{16} The COOD mode at ~1710 cm\textsuperscript{−1} arises from carboxylic acid and aspartic acid, and it is an indicator of the pH of the sample, because this band appears at pH values below the pK\textsubscript{a} of these amino acids. The side chain of proline is bound to the nitrogen of its backbone, forming a tertiary amide unit, and as a result its vibrational mode (which is centered at ~1610 cm\textsuperscript{−1}) has a lower frequency than amide I modes. Proline often appears in γ- and β-turns of proteins, and because of its localized character, it is often used as a structural probe of proteins.\textsuperscript{17,18} The guanidinium side chain of arginine has two modes that absorb at ~1605 cm\textsuperscript{−1} and ~1585 cm\textsuperscript{−1}. The low-frequency arginine mode overlaps with the absorption of the COO\textsuperscript{−} moiety of aspartate, which is also at ~1585 cm\textsuperscript{−1}. The COO\textsuperscript{−} group of glutamate has an absorption at slightly lower frequencies, normally ~1565 cm\textsuperscript{−1}. The arginine and carboxylate side chains can interact and form what is known as a salt bridge, which is a long-range or tertiary interaction. Characterizing the infrared and 2D-IR response of the vibrational modes of the two moieties that
1. Introduction

Figure 1.4: FTIR spectrum of the zinc-finger mutant FSD1 in D2O (solvent subtracted). The amide I' and amide II' bands are due to vibrations of the backbone. Other side-chain vibrations that are visible are also indicated. Note that FSD1 does not contain proline, but the position of the IR band is nevertheless indicated.

form a salt bridge is the topic of Chapters 4 and 5. Arginine can also form a salt bridge with aspartate, but the vibrational frequencies overlap too much and a spectroscopic study is not possible without isotope labels. The mode at \(1517\ \text{cm}^{-1}\) arises from a ring vibration of the side chain of tyrosine, and it is sensitive to solvation. Tyrosine is often found inside the hydrophobic core of proteins, and this ring vibration can therefore serve as a probe of loss of globular structure and fast dynamics of water penetration into the hydrophobic core during time-resolved experiments.\(^{19}\) The mode at \(1400\ \text{cm}^{-1}\) arises from carboxylate and aspartate. There are many other vibrations of amino acid side chain,\(^{16}\) in particular ring vibrations of aromatic side chains, but they normally have a smaller absorption than the ones mentioned above, and therefore might not be very useful in determining local structure of proteins.