New light on protein folding: Unraveling folding and unfolding mechanisms using time-resolved and two-dimensional vibrational spectroscopy

Meuzelaar, H.

Citation for published version (APA):
Proteins are biological macromolecules that play an essential role in virtually all chemical processes that occur in nature. A living cell contains thousands of different kinds of proteins, that fulfill a large variety of functions, including catalyzing metabolic reactions, controlling and carrying out cell-division, regulating gene expression, participating in immune protection, responding to stimuli, providing structural-mechanical support, conveying messages from one cell to another, and controlling the transport and storage of a wide variety of target molecules. In spite of their functional diversity, proteins are a relatively homogeneous class of polymeric biomolecules: they differ from one another primarily in length and sequence of amino-acid residues. In order to fulfill its specific biological function, a protein’s unique sequence of typically several hundreds of amino acids needs to fold into a well-defined three-dimensional structure, a process that usually occurs spontaneously in cells. However, with an astronomically large number of ways in which the floppy chains of amino acids can potentially fold, the question arises as to how a protein folds on biologically relevant time scales (typically on the micro- to millisecond time scale) from its one-dimensional sequence of amino acids into its highly organized functional shape. This is one of the most challenging questions currently investigated in several scientific disciplines and is known as the protein folding problem. Gaining knowledge about the mechanisms that control protein folding becomes increasingly relevant, as incorrect or incomplete folding of proteins is believed to be involved in several important human diseases and neurological disorders, such as particular types of cancer, Alzheimer’s disease, Parkinson’s disease, and mad cow disease.
Introduction

Figure 1.1: Amino-acid residues (indicated by the dotted line) are linked together in a protein sequence by peptide bonds (highlighted in yellow) [1]. The central carbon atoms (Cα) and the residue side-chain groups (R) of an amino acid and its two neighboring residues are displayed schematically. The backbone torsion angles (φ, ψ) and the side-chain torsion angle (χ) are indicated by the circular arrows. The straight arrows indicate the direction of the transition-dipole moments of two amide I vibrations.

1.1 The structure and conformation of proteins

The structural building blocks of protein molecules are amino acids. Each amino acid consists of a central carbon atom (Cα) that is covalently bonded to a carboxylic acid group, an amino group, a hydrogen atom, and a specific side-chain group. Natural proteins are built of different sequences of the same set of twenty amino-acid residues that each have a unique side-chain group varying in size, charge, hydrogen-bonding capacity, hydrophobic character, and chemical reactivity, giving each type of amino acid its distinctive chemical and physical properties [2]. Adjacent amino-acid residues are linked together by peptide bonds that are formed between the α-carboxyl of one residue and the α-amino group of the next (Fig. 1.1). The specific sequence of amino-acid residues is referred to as the primary structure (Fig. 1.2a). Depending on the number of residues and their relative positions in the polypeptide chain, a protein folds into its specific functional conformation. The conformational flexibility of the protein’s backbone, i.e., the repetitive chain of carbon, nitrogen, and oxygen atoms along the polypeptide chain (see Fig. 1.1), depends on the degrees of freedom of each individual amino-acid residue. Due to delocalization of the lone pair of electrons of the α-nitrogen onto the α-carbonyl oxygen, the C–N bond of the peptide group has a partial double-bond character, resulting in a planar geometry of this bond [3, 4]. As a result, internal rotation about the peptide bond is restricted, and the conformational space of the polypeptide chain is primarily characterized by the degrees of freedom of the backbone torsion angles (φ, ψ) of the two σ-bonds of each single amino-acid residue in the protein sequence (see Fig. 1.1). The freedom of rotation about the two φ and ψ bonds of each amino acid allows even a relatively short polypeptide chain to fold into an astronomical number of possible conformations, which increases exponentially with chain length. The unfolded state or ‘random coil’ conformation of a
protein consists of a wide distribution of structures, and can be defined as the
ensemble of conformers in which internal rotations about \( \phi \) and \( \psi \) can occur with
the same freedom compared to a molecule of low molecular weight containing an
isolated peptide bond [5]. Local torsion \((\phi, \psi)\) angle fluctuations can take place on
time scales as fast as \( \sim 10^{-11} \) s, implying that the conformations of an unfolded
protein sequence rapidly interconvert with each other [5]. Within the \((\phi, \psi)\) con-
figuration space, two dominant, broad free energy minima exist, corresponding to
the two most common secondary-structure elements found in native protein struc-
tures, viz. \( \alpha \)-helices and \( \beta \)-sheets (Fig. 1.2b). These secondary structures adopted
by local regions of the polypeptide backbone arise from the repetition of specific
torsion angles along its sequence, resulting in a specific intramolecular hydrogen-
bonding structure [5]. For example, amino-acid residues involved in an \( \alpha \)-helical
backbone conformation typically adopt torsion \((\phi, \psi)\) angles near \((-62^\circ, -41^\circ)\),
permitting the NH group of each backbone amide nitrogen to donate a hydrogen
bond to the C=O group of the amino acid positioned four residues earlier [4].
On the other hand, the energetically preferred backbone torsion \((\phi, \psi)\) angles of
residues in a \( \beta \)-strand are around \((-135^\circ, 135^\circ)\). Individual \( \beta \)-strands are energet-
ically not stable due to the lack of intramolecular interactions, but when incorpor-
ated into a \( \beta \)-sheet a hydrogen-bonding pattern involving the peptide groups of
adjacent strands forms a stable conformation [4]. The secondary-structure motifs
adopted by a single polypeptide backbone pack together to form the overall three-
dimensional conformation of the folded protein (tertiary structure), which allows
the protein to fulfill its biological function(s) (Fig. 1.2c). The tertiary structure is
well-defined for most globular proteins, and the structural transitions within the
folded state resulting from molecular motions occur typically on a scale of less than
2 Å [6]. The characteristic three-dimensional shape of proteins that contain more
than one polypeptide chain is determined by the spatial arrangement of different
tertiary subunits, forming the quaternary structure of the protein (Fig. 1.2d).

\[
\begin{array}{c}
\text{Val} \\
\text{Lys} \\
\text{Phe} \\
\text{Gly} \\
\text{Arg} \\
\text{Leu} \\
\text{Ala} \\
\text{Glu} \\
\text{Met} \\
\text{Cys} \\
\text{Ala} \\
\text{Ala} \\
\text{Arg} \\
\text{His} \\
\text{Gly} \\
\text{Leu} \\
\text{Asp} \\
\text{Asn} \\
\text{Tyr}
\end{array}
\]

\textbf{Figure 1.2:} Schematic overview of the four levels of protein structure [7]. (a) Primary
structure: the unique sequence of different amino-acid residues. (b) Secondary structure:
the conformation of local regions of the polypeptide backbone. (c) Tertiary structure:
the overall folding of the protein. (d) Quaternary structure: the three-
dimensional shape of a complex of two or more proteins.
1.2 Understanding how proteins fold

The three-dimensional, biologically active conformation of a protein is specified by its sequence of amino-acid residues. Because the number of configurations accessible to the disordered polypeptide chain grows exponentially with chain length, the theoretical number of unfolded conformations available to a protein is astronomically large and the general question arises as to how and why proteins fold. In the early 1960’s, the classical experiments of Anfinsen demonstrated that denatured (or unfolded) proteins refold spontaneously into their well-defined globular structure as soon as the optimal environmental conditions are restored [8]. Anfinsen concluded that the overall three-dimensional structure of a protein must be encoded in its sequence. He also concluded that under physiological conditions, the folded conformation of a protein corresponds to a unique, stable, and kinetically accessible minimum of the Gibbs free energy [8]. The question of how a disordered sequence of amino acids determines a protein’s three-dimensional conformation and reaches its corresponding minimum free energy is far from simple to answer and became known as the protein folding problem. Unraveling the folding problem requires a detailed understanding about the fundamental forces that control the folding and unfolding processes of proteins and peptides.

1.2.1 Thermodynamics of protein folding

The stability of the folded state of proteins and peptides depends on a complicated interplay of a large number of intrinsically weak, noncovalent interactions, including hydrophobic effects, Van der Waals forces, hydrogen bonding, and electrostatic interactions [4, 8]. Although each of these individual interaction energies is relatively small, because of their large number and their high degree of cooperativity, the sum of their individual strengths add up to a significantly large total interaction energy of the folded state conformation [3, 4]. However, since the energetically favorable interactions formed during the folding process is counteracted by a loss of conformational entropy of the polypeptide backbone as the protein folds, the overall thermodynamic stability of proteins (the free energy difference between the folded and unfolded states) is typically very small, ranging from about $-20$ to $-60$ kJ mol$^{-1}$ under physiological conditions [3]. The equilibrium Gibbs free energy of folding ($\Delta G_{F-U}^0 = G_F - G_U$) is given by:

$$\Delta G_{F-U}^0 = \Delta H_{F-U} - T \Delta S_{F-U} = -RT \ln K_{eq}$$

(1.1)

where $\Delta H_{F-U} = H_F - H_U$ is the enthalpy change and $\Delta S_{F-U} = S_F - S_U$ is the entropy change upon folding of the protein. $K_{eq}$ denotes the folding equilibrium constant, i.e., the ratio between the folded and unfolded state populations in equilibrium, $R$ is the gas constant, and $T$ the absolute temperature.

The major force opposing protein folding is the loss of configurational freedom of the disordered protein backbone when it folds into a well-defined, conformationally restricted three-dimensional structure. This loss of entropy must be balanced
by a negative change in enthalpy for the protein to fold \((\Delta G_F^{\text{F-U}} < 0)\) [3]. As a result, spontaneous folding processes are driven by large negative values of \(\Delta H_F^{\text{F-U}}\). At increasing temperature, the magnitude of the entropic term \(T \Delta S_F^{\text{F-U}}\) increases, resulting in an increasing value of \(\Delta G_F^{\text{F-U}}\). The transition midpoint or melting temperature \((T_m)\) defines the temperature at which the equilibrium free energy of folding equals zero \((\Delta G_F^{\text{F-U}} = 0)\), with equally populated folded and unfolded states. At temperatures above \(T_m\), the entropic term \(-T \Delta S_F^{\text{F-U}}\) becomes greater than \(\Delta H_F^{\text{F-U}}\), resulting in positive values of the free energy of folding \((\Delta G_F^{\text{F-U}} > 0)\) and increasing the population of the unfolded state. Thus the thermally induced unfolding of a protein is mainly an entropy-driven process [3, 9]. Note that we assume that the enthalpic \(\Delta H_F^{\text{F-U}}\) and entropic \(\Delta S_F^{\text{F-U}}\) contributions to the free energy of folding \(\Delta G_F^{\text{F-U}}\) are temperature-independent, which is expected for the small protein systems or peptides as studied in this thesis (see section 2.1.1).

### 1.2.2 The folding problem: Levinthal’s paradox

The number of conformations that an unfolded polypeptide chain can adopt primarily depends on the freedom of rotation about the two \(\phi\) and \(\psi\) bonds of each single amino-acid residue in its sequence (see Fig. 1.1). While this number can only be roughly estimated due to a certain restriction of the accessible \((\phi, \psi)\) configuration space arising from geometric constrains between spatially overlapping residue side-chain groups, the unfolded polypeptide chain of even a relatively short protein has a tremendous number of potential conformations [3, 4]. In 1968, Levinthal pointed out that if each of the \(\phi\) and \(\psi\) bond angles can be in one of three possible stable conformations, a relatively small protein consisting of 100 amino-acid residues could adopt approximately \(10^{100}\) possible conformations [10, 11]. The time needed to randomly explore this astronomically large number of available protein configurations would be on the order of many millions of years \((\sim 10^{10}\) years), even if we assume every single sampling step would be as short as \(10^{-12}\) s [6, 11]. This implies that an unbiased random search through the entire conformational space available to an unfolded polypeptide chain does not permit a protein to fold within the lifetime of the universe [10, 11]. This reasoning is in strong contrast with the observation that the folding process of many naturally occurring proteins proceeds typically on a time scale of milliseconds to seconds, and several small proteins even fold on the microsecond time scale [6]. These apparently contradicting facts following from Levinthal’s thought experiment became known as the **Levinthal paradox** or the **protein folding paradox** (see Fig. 1.3). Levinthal argued that the protein folding paradox can be resolved if there exists a conformational bias in the protein’s sequence towards the folded structure, providing the folding process with a high degree of directionality [10]. As a consequence, there must exist specific folding pathways that direct the protein molecule from an unfolded conformation to its folded structure [6, 10]. The rapid formation of local substructures and the occurrence of one or a few obligatory intermediate structures along the folding pathway may guide subsequent events, allowing the folding process to proceed far more efficiently than sequential sampling of all accessible conformations [6].
Figure 1.3: Illustration of the protein folding paradox and its solution [6]. For a protein consisting of 100 amino acids, approximately $10^{100}$ conformations in the maximally unfolded state exists. An unbiased random search through all of these accessible configurations would take the protein many millions of years to fold. In reality, the folding occurs typically on a µ—ms time scale. The question arises as to the underlying mechanism that allows protein folding processes to proceed far more efficiently than a random search.

1.2.3 Folding pathways and folding funnels

In the early 1990’s, the concept of protein folding pathways evolved to the more general idea that the folding free-energy landscape of a protein is shaped like a funnel [12–19]. Instead of envisioning the folding process as a well-defined sequence of events in which each individual protein molecule follows essentially the same pathway on its way from the unfolded to the folded structure, the folding funnel hypothesis puts more emphasis on the coexistence of conformational sub-ensembles and multiple folding trajectories. From the free-energy landscape perspective, the process of protein folding is analogous of going downhill in the funnel-like energy landscape through a multiplicity of folding routes, until reaching the global free-energy minimum at the narrowest bottom of the funnel, corresponding to the native conformation. Thus, individual polypeptide chains can fold by funneling down the free-energy landscape following different folding trajectories, to eventually find the same, final folded state at the global energy minimum [13, 16]. Fig. 1.4 shows different types of free-energy landscapes that represent different folding scenarios [13]. In these free-energy landscapes, the vertical axis corresponds to the free energy, while the two horizontal axes schematically represent the conformational degrees of freedom of the polypeptide chain. The top of the funnel is wide, representing the high configurational entropy of the unfolded state ensemble, while it increasingly narrows as the system funnels down and approaches
the bottom, reflecting the reduction of the accessible configurational space of near-native states and the native structure. The Levinthal ‘golf-course’ model of the free-energy landscape (see Fig. 1.4a) emphasizes that protein folding cannot occur on biologically relevant time scales if the folding process would involve randomly searching all accessible conformations \[13\]. When performing a random search through the featureless ‘golf-course’ energy landscape, the probability to find the folded state at the position of the ‘golf-hole’ is extremely small. The folding-funnel view readily explains how reducing the dimensionality of the accessible configurational space could solve the search problem, as steep downhill slopes guide the protein relatively quickly from the unfolded to its folded state. Fig. 1.4b shows a smooth, moated free-energy landscape that illustrates how multiple folding routes can coexist: a fast-folding two-state process that proceeds directly from the unfolded state ensemble to the native state, in parallel with a slow-folding trajectory that occurs through the formation of an intermediate state, because the protein molecule gets trapped in a local energy minimum on its way to the bottom of the funnel \[13\]. For larger systems, folding funnels may not be as simple as in Fig. 1.4b. A more realistic free-energy landscape can be found in Fig. 1.4c, showing a rugged energy landscape involving different non-native local energy minima, free-energy barriers, and a variety of possible folding routes \[13\].

### 1.2.4 Transition-state theory

The question arises as to what determines the rate of folding of a protein when it funnels down the energy landscape. Transition-state theory connects the rate constant \(k\) to the Gibbs free energy of activation \(\Delta G^\ddagger\) and assumes that the initial (unfolded, U) and final (folded, F) states are separated by a free-energy barrier \[5, 6\]. The transition state corresponds to the state of the highest free en-

---

**Figure 1.4:** Representation of different types of free-energy landscapes (adapted from Ref. \[13\]). (a) The Levinthal ‘golf-course potential’ representing a random search through a featureless energy landscape. (b) A moated energy landscape with two parallel folding pathways: a fast-folding two-state transition, and a slow-folding process involving a local energy minimum. (c) A rugged energy landscape involving different kinetic traps, free-energy barriers, and some narrow folding routes.
energy along the reaction coordinate (denoted with $\ddagger$), and the height of the barriers $\Delta G^{\ddagger}_{U \rightarrow F}$ and $\Delta G^{\ddagger}_{F \rightarrow U}$ determines the magnitude of the rate constants $k_F$ and $k_U$, respectively (see Fig. 1.5a) [5, 6]. The temperature dependence of the folding or unfolding rate constant $k$ is given by the Eyring equation [3]:

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right) = \frac{k_B T}{h} \exp\left(-\frac{\Delta H^{\ddagger}}{RT} + \frac{\Delta S^{\ddagger}}{R}\right)$$

(1.2)

where $h$ is Planck’s constant, $\Delta G^{\ddagger}$ the Gibbs free energy of activation, $\Delta H^{\ddagger}$ the transition enthalpy, and $\Delta S^{\ddagger}$ the transition entropy. Note that for protein systems with a nonzero change in heat capacity ($\Delta C_p$), the $\Delta C_p$ terms needs to be added to the Eyring equation (see Eq. 17.1 and 17.2 in Ref. [3]). In addition, Eyring theory can only be applied to protein folding if the thermodynamic equilibrium can be represented by two experimentally distinguishable states, viz., the unfolded and folded states, implying that the folding process occurs through crossing of a single kinetic activation barrier (see Fig. 1.5a). However, if the folding pathway leads through one or more non-native free energy minima, and the folding thus proceeds through the formation of one or more intermediate states (see Fig. 1.5b), the protein exhibits non-Arrhenius kinetics.

### 1.3 Initiating and probing fast folding events

Experimental investigation of protein folding dynamics requires experimental methods that combine conformational specificity with a time resolution that is sufficient to probe the rapid evolution from one backbone conformation to another, on time scales ranging from nanoseconds to milliseconds. Vibrational spectroscopy is a well-suited technique for the structural characterization of proteins, in particular due to its sensitivity to the conformation of the polypeptide backbone in the amide I spectral region. In addition, vibrational spectroscopy can be performed.
1.3 Initiating and probing fast folding events

with picosecond time resolution, which can be exploited for the experimental observation of rapidly evolving protein conformations during (un)folding experiments. In this thesis, folding and unfolding processes are initiated using a laser-induced temperature-jump (T-jump). In the following subsections, we discuss how time-resolved IR spectroscopy in the amide I spectral region can be utilized to study the conformational dynamics during the (un)folding of proteins and peptides in response to a nanosecond T-jump.

1.3.1 Amide I vibrations as a spectral signature of protein structure

The amide I absorbance of proteins and peptides (1600–1700 cm⁻¹), arising from the vibrational excitation of the amide I modes of the polypeptide backbone, provides distinct spectroscopic signatures of secondary structure [20]. Amide I vibrations involve primarily backbone C=O stretch vibrations of the peptide-linkage groups (highlighted in yellow in Fig. 1.1), with minor contributions from the out-of-phase CN stretching vibration, the CCN deformation, and the NH in-plane bend [21]. The sensitivity of amide I vibrations to secondary structure is due to transition-dipole coupling between individual amide I oscillators of neighboring peptide units, which can result in delocalization of vibrational states over large regions of the protein backbone [20, 21]. The interaction between coupled vibrational states gives rise to characteristic amide I normal modes which are determined by the relative orientation of, and the distance between, the coupled transition dipoles [21]. Structural changes of the protein backbone alter the relative position and orientation of the coupled amide I transition-dipole moments, thereby affecting the energy levels of the coupled system and leading to a shift of the amide I frequency. In addition to conformation-dependent vibrational couplings between adjacent peptide groups, the local-mode frequency of the carbonyl C=O stretching vibrations of the amide units is also correlated with the strength of hydrogen bonding [21]. Generally, when peptide backbone groups are involved in specific hydrogen-bonding patterns such as an α-helical structure, a proton-donor (backbone NH groups) hydrogen bonds to the backbone C=O group, thereby withdrawing electron density from the C=O bond [22]. This effect weakens the C=O force constant and shifts the position of the amide I absorbance maximum to lower frequencies [21, 22]. Consequently, the frequency and width of the amide I band of a protein directly reflects its secondary structure.

Fig. 1.6 summarizes the well-established empirical relationships between the vibrational frequency of the amide I mode and protein conformation [23]. The unstructured regions of a protein backbone in a random-coil conformation give rise to randomly oriented amide I transition-dipole moments, and appear in the IR-absorbance spectrum as a single broad amide I peak centered at approximately 1645 cm⁻¹ [21]. When involved in an α-helix, the transition dipoles of the coupled amide oscillators are directed in a specific manner along the helical axis, and the ensuing vector summation give rise to resultant transition-dipole moments that are oriented either parallel or perpendicular relative to the helical axis. As a consequence, the peptide groups in an α-helical backbone conformation give rise
Amide I spectroscopy of proteins: the well-established correlation between the position of the amide I absorbance maximum and different secondary structural motifs [1, 23].

Figure 1.6: Amide I spectroscopy of proteins: the well-established correlation between the position of the amide I absorbance maximum and different secondary structural motifs [1, 23].

to two characteristic vibrational modes, viz., $A$-($\|\|$) and $E$-($\perp\|$) symmetry modes, centered around 1638 and 1650 cm$^{-1}$, respectively [21]. The frequency splitting of the $\alpha$-helix modes cannot be resolved using linear IR spectroscopy, however, the coupling between the two modes can be measured using polarization-controlled 2D-IR experiments [24] (see section 2.1.2 and chapter 7). The coupling between backbone amide I vibrations in $\beta$-sheets also results in the appearance of two characteristic IR spectral features centered at approximately 1680 and 1630 cm$^{-1}$, which are assigned to the parallel ($\nu_\parallel$) and perpendicular ($\nu_\perp$) alignment of the transition-dipole moments relative to the direction of the $\beta$-strands [21]. The low-frequency $\nu_\perp$ peak red-shifts and gains intensity with increasing size of the $\beta$-sheet [25]. Another local probe of protein structure arises from backbone peptide units in a polyproline helix [26, 27]. Because proline’s backbone nitrogen forms an imide unit when involved in a peptide linkage group, and is bonded to three carbon atoms instead of two carbon atoms and a hydrogen, the amide I vibration of proline has a low amide I frequency centered around 1620 cm$^{-1}$ [27]. This red-shift largely decouples the proline imide vibration from the remaining amide I vibrations of the polypeptide backbone [27]. As a consequence of this weak coupling, spectral changes in IR absorption at frequencies where proline participates mainly indicate changes in the degree of hydrogen bonding to the proline C=O imide groups, for instance due to changes in solvent exposure [27] (see chapter 6).

Despite the high sensitivity of amide I vibrations to the secondary structure of proteins and peptides, the structural interpretation of linear amide I spectra of globular proteins is often not straightforward. The presence of a large number of delocalized modes often leads to broad, featureless amide I absorption lineshapes (see Fig. 1.6). As a result, linear amide I spectra commonly exhibit limited variation for different proteins, and due to the spectral overlap of the various subcomponents difficulties may arise as to the assignment of different secondary structural motifs. Besides the amide I band of the peptide backbone, site-specific resolution can be achieved by probing the IR response of several amino acid side-chain groups, which can provide information about the specific tertiary packing interactions that stabilize the folded protein conformation. For example, the C–C aromatic ring stretching vibration of tyrosine gives rise to a relatively sharp absorption peak
centered at \( \sim 1515 \text{ cm}^{-1} \), which significantly blue-shifts upon solvation \([28, 29]\). Because tyrosine’s aromatic side chain is often involved in the formation of the hydrophobic core of proteins, this spectral feature can be indicative for the loss of globular structure caused by water penetration into the hydrophobic interior of the protein and the unfolding of its hydrophobic core structure \([28]\). Another spectral signature arising from side-chain functional groups are the symmetric and antisymmetric CN-stretch vibrations of the guanidinium \((\text{CN}_3\text{H}_5^+)\) group of arginine \((\text{Arg})\) appearing at 1585 and 1607 cm\(^{-1}\), respectively \([21]\). The guanidinium group of \(\text{Arg}^+\) can interact with the side-chain carboxylic \((\text{COO}^-)\) groups of glutamate \((\text{Glu}^-)\) or aspartate \((\text{Asp}^-)\) to form a salt bridge, i.e., a tertiary interaction between oppositely charged residue side-chain groups. The contribution of the CO-stretching modes of the carboxylate groups of \(\text{Glu}^-\) and \(\text{Asp}^-\) arise at \(\sim 1565\) and \(\sim 1585\) cm\(^{-1}\), respectively \([21]\). Experimental evidence has indicated that the infrared response of both the side-chain guanidinium and carboxylate groups slightly changes upon salt-bridge formation due to a coupling between the low- and high-frequency CN stretch modes of \(\text{Arg}^+\) and the COO\(^-\) mode of \(\text{Glu}^-\) or \(\text{Asp}^-\) when involved in a salt bridge \([30]\). Therefore, the spectral changes observed at frequencies where \(\text{Arg}^+\) and \(\text{Glu}^-/\text{Asp}^-\) participate can provide unique information about tertiary structural contacts formed between salt-bridging side chains (see chapters 6 and 8). In addition, because the contributions of the COO\(^-\) mode of \(\text{Glu}^-\) and \(\text{Asp}^-\) at 1565 and 1586 cm\(^{-1}\) shift upon protonation to 1705 and 1713 cm\(^{-1}\), respectively, the IR spectrum can serve as a probe of the \(\text{Glu}\) and \(\text{Asp}\) protonation state \([21]\) (see chapters 3 and 4). Increased structural specificity can also be obtained by isotopic substitution of a single or multiple backbone \(^{12}\text{C}^{16}\text{O}\) groups with \(^{13}\text{C}^{18}\text{O}\) groups (see chapter 6). Due to their larger reduced mass, the amide I vibrations of the isotopically labeled groups have a

![Figure 1.7](image-url)

**Figure 1.7:** Equilibrium IR spectral changes in the amide I’ region produced by thermal unfolding of the Trp-cage miniprotein. (a) Schematic representation of the folded structure of Trp-cage showing the different secondary structure elements. The thermal difference IR spectrum (b, yellow) reflects the conformational changes upon thermal unfolding.
lower frequency relative to the nonisotopically edited protein backbone fragments, thereby spectrally isolating individual residues involved in a particular secondary structure motif of the protein. Finally, additional structural resolution in the amide I spectral region can also be achieved by spreading the amide I absorption spectra over a second frequency dimension using two-dimensional infrared (2D-IR) spectroscopy [23], which give rise to characteristic cross peaks resulting from the vibrational coupling between different modes (see section 2.1.2 and chapters 5, 7 and 8).

1.3.2 Temperature-jump infrared-probe spectroscopy

As discussed in the previous subsection, the amide I response of proteins and peptides, primarily arising from C=O stretch vibrations of the backbone amide groups, is a sensitive probe of secondary structure and can be directly related to the conformation of the protein backbone [20, 21]. Thus, when a protein undergoes structural changes, for example during thermal denaturation, the amide I spectrum changes. The sensitivity of the amide I mode to secondary structural changes is illustrated in Fig. 1.7, in which we show the temperature-dependent IR absorbance in the amide I spectral region of the Trp-cage miniprotein. The low-temperature spectrum (Fig. 1.7b, cyan) exhibits a major amide I peak, and shows various IR spectral signatures of different secondary and tertiary structure elements of Trp-cage’s folded structure (see Fig. 1.7a). When the folded=unfolded equilibrium is shifted in the direction of the unfolded state by increasing the temperature, the main amide I absorption maximum shifts to higher frequencies and decreases in intensity while broadening, maintaining approximately the same integrated intensity (Fig. 1.7b, magenta). The underlying spectral changes produced by thermal unfolding are better visible in the equilibrium IR difference spectrum (Fig. 1.7b, yellow), which is generated by the subtraction of the IR spectrum at low temperature from the spectrum at higher temperature. The resulting temperature-difference spectrum exhibits both negative (1600–1650 cm⁻¹) and positive (1650–1700 cm⁻¹) contributions reflecting the conformational changes of the miniprotein upon thermal unfolding. The negative spectral feature indicates the loss of secondary and tertiary structural contacts, whereas the region of increasing intensity is indicative of an increase of the unfolded population due to melting of the helical subcomponents of the miniprotein [21, 31]. In addition, the minor negative feature appearing at 1575 cm⁻¹ arises from the change in absorbance of the side-chain carboxylic groups of residue Asp [21], and provides unique information about the unfolding of Trp-cage’s hydrophobic cage structure due to breaking of the salt bridge formed between Asp and Arg (see Fig. 1.7a). Hence, the equilibrium temperature-difference IR spectrum of proteins and peptides can be seen as a spectral fingerprint of the conformational changes produced by thermal unfolding.

The structural and conformational specificity of temperature-dependent IR spectroscopy can be used to study protein folding events not only in thermal equilibrium, but also dynamically after initiation of the (un)folding process us-
1.3 Initiating and probing fast folding events

Figure 1.8: Principle of T-jump experiments for a two-state folder under conditions that favor folding ($\Delta G_{F-U}^0 < 0$). (a) The temperature dependence of the equilibrium free energy ($\Delta G^\circ$). $\ddagger$ indicates the transition state. (b) A nanosecond T-jump ($\Delta T$) perturbs the free-energy surface of the protein and initiates a relaxation process to a new equilibrium position. The population re-distribution of the unfolded (U) and folded (F) states follows exponential relaxation kinetics [3].
to interpret experimental $T$-jump findings and to derive the underlying folding and unfolding rate constants.

**Kinetics of two-state folding**

The simplest type of folding transition is that of a two-state process. The two-state model applies if the thermodynamic equilibrium of the protein can be effectively characterized by two states, viz., the unfolded and folded states, that can be regarded as two broad free-energy minima in the conformational free-energy landscape, which are separated by a single, relatively high kinetic activation barrier (see Fig. 1.8a). As a consequence, the folding kinetics of a simple two-state folder proceeds directly from the unfolded ($U$) to the folded ($F$) state without the occurrence of an intermediate state [5, 6]:

\[
U \xrightarrow{k_F/k_U} F
\]

(1.3)

where $k_F$ and $k_U$ correspond to the rate constants for folding and unfolding, respectively. Because no intermediate species accumulate during a two-state folding transition, the change of the unfolded and folded state populations with time after perturbation from equilibrium at $t = 0$ can be described by the following rate equations [6]:

\[
\frac{d[U]}{dt} = k_U[F] - k_F[U]
\]

(1.4)

\[
\frac{d[F]}{dt} = k_F[U] - k_U[F]
\]

where $[U]$ and $[F]$ are the time-dependent concentrations of the $U$ and $F$ conformations, respectively. The general solution of these rate equations is a single-exponential function [6]:

\[
[F](t) = C_1 + C_2 \exp(-(k_F + k_U)t)
\]

\[
[U](t) = [UF] - [F](t)
\]

\[
C_1 = [F]_0 - [UF]k_U/(k_F + k_U)
\]

\[
C_2 = [UF]k_F/(k_F + k_U)
\]

\[
k_R = k_F + k_U.
\]

(1.5)

where $[UF]$ is the (time-independent) total peptide/protein concentration $[U]+[F]$, and $[F]_0$ the initial concentration of $F$ at $t = 0$. Fig. 1.9d presents the time-dependence of the unfolded and folded state populations after perturbation from equilibrium by a $T$-jump under conditions that favor unfolding ($\Delta G^o_{F-U} > 0$). Note that the change in the populations of $U$ and $F$ after a $T$-jump perturbation shows a single-exponential time dependence with a relaxation rate constant $k_R$ [5, 6]. After the new equilibrium conditions have been established at $t \gg k_R^{-1}$,
1.3 Initiating and probing fast folding events

Figure 1.9: Schematic free energy profiles of a two-state folder (a) under conditions that favor folding \((T \ll T_m)\), (b) at the melting temperature \((T = T_m)\), and (c) under conditions that favor unfolding \((T \gg T_m)\). (d) In a reversible two-state transition \((U \rightleftharpoons F)\), the population of the unfolded \((U)\) and folded \((F)\) states changes according to single-exponential kinetics after perturbation of the folding equilibrium at \(t = 0\) [6]. The conditions chosen for this example favor unfolding \((k_F = 1, k_U = 3, k_R = k_F + k_U = 4, [F_0]/[UF] = 0.85)\).

The concentrations of the folded and unfolded populations reach a constant value that depends on the relative size of \(k_F\) and \(k_U\). In order to determine \(k_F\) and \(k_U\) it is necessary to also determine the equilibrium constant \((K_{eq})\). For a two-state folding process \(K_{eq}\) is related to \(k_F\) and \(k_U\) by [5]:

\[
K_{eq} = \frac{[F]_{eq}}{[U]_{eq}} = \frac{k_F}{k_U}
\]

where \([F]_{eq}\) and \([U]_{eq}\) indicate the equilibrium concentrations of the F and U conformations, respectively. \(K_{eq}\) can be determined from temperature-dependent equilibrium experiments, whereas \(k_R\) can be determined from time-resolved experiments after a nanosecond \(T\)-jump perturbation. This implies that when the folding transition is a two-state process, a combined analysis of the equilibrium and kinetic folding data provides a system of two equations with two unknowns, which allows the determination of the folding \(k_F\) and unfolding \(k_U\) rates. This will be used in chapters 3 and 4.

**Sequential three-state transitions**

The folding mechanism of moderately large globular proteins (>70 residues) often involves the occurrence of one or more transiently populated intermediate states, represented by local free-energy minima along the reaction coordinate [32, 33]. The formation of folding intermediates in such hierarchical folding mechanism may facilitate a faster folding process as they restrict the conformational search of the protein, thereby directing it towards a well-defined folding pathway (see section 1.2.3) [34]. In case of a reversible sequential three-state transition, the
folding proceeds through the formation of a single on-pathway intermediate (I) state, represented by a local free-energy minimum in the conformational free-energy landscape that is located on the reaction coordinate between the unfolded (U) and folded (F) state (see Fig. 1.5b). The kinetic scheme for a reversible sequential three-state process is given by [6]:

\[
U \xrightarrow{k_{U\rightarrow I}} I \xrightarrow{k_{I\rightarrow F}} F
\]

(1.7)

where \(k_{U\rightarrow I}, k_{I\rightarrow F}, k_{I\rightarrow U}, \) and \(k_{F\rightarrow I}\) are the rate constants of each of the individual transitions. The population changes of the unfolded, intermediate, and folded states after perturbation from equilibrium at \(t = 0\) is in this case given by the following rate equations [6]:

\[
\begin{align*}
\frac{d[U]}{dt} &= k_{I\rightarrow U}[I] - k_{U\rightarrow I}[U] \\
\frac{d[I]}{dt} &= k_{U\rightarrow I}[U] - k_{I\rightarrow U}[I] + k_{F\rightarrow I}[F] - k_{I\rightarrow F}[I] \\
\frac{d[F]}{dt} &= k_{I\rightarrow F}[I] - k_{F\rightarrow I}[F].
\end{align*}
\]

(1.8)

The general solution of this set of coupled differential rate equations can be found by linear algebra methods, and is given by two bi-exponential functions [6]:

\[
\begin{align*}
[F](t) &= C_1 + C_2 \exp(-k_{R_1}t) + C_3 \exp(-k_{R_2}t) \\
[U](t) &= C_4 + C_5 \exp(-k_{R_1}t) + C_6 \exp(-k_{R_2}t)
\end{align*}
\]

(1.9)

The constants \(C_n\) are determined by the conservation relationship \([UIF] = [U] + [I] + [F]\), and by the initial concentrations \([U](0), [I](0), \) and \([F](0)\) [6]. Hence, the time-dependent change of the populations of U, I, and F after perturbation from equilibrium in a reversible sequential three-state transition follow a bi-exponential function (see Fig. 1.10 for an illustration). The corresponding relaxation rates \(k_{R_1}\) and \(k_{R_2}\) depend in a complicated manner on the underlying rate constants of the \(U \rightleftharpoons I \rightleftharpoons F\) interconversions [6]:

\[
k_{R_{1,2}} = 0.5(k_{U\rightarrow I} + k_{I\rightarrow U} + k_{I\rightarrow F} + k_{F\rightarrow I}) \pm \left((k_{U\rightarrow I} + k_{I\rightarrow U} + k_{I\rightarrow F} + k_{F\rightarrow I})^2 - 4\sqrt{(k_{U\rightarrow I}k_{I\rightarrow F} + k_{U\rightarrow I}k_{F\rightarrow I} + k_{I\rightarrow U}k_{F\rightarrow I})}\right)
\]

(1.10)

A difficulty arises as to the determination of the individual rate constants associated with crossing of each separate energy barrier, as this would require knowledge about the relative positions of the free-energy levels for the U, I, and F states. As a consequence, and in contrast to the two-state case in the previous subsection, it is generally not possible to determine the microscopic folding rate constants from the observed bi-exponential \(T\)-jump decays. This situation will be encountered in chapters 6 and 8. As can be seen in Fig. 1.10a, the intermediate state accumulates kinetically during the folding transition under conditions that favor folding (\(\Delta G_{F\rightarrow U}^0 < 0\), because it becomes rapidly populated by fast \(U \rightleftharpoons I\) interconversions
Figure 1.10: The population changes of the unfolded (U), intermediate (I), and folded (F) state in a reversible sequential three-state transition \((U \rightleftharpoons I \rightleftharpoons F)\) follows a bi-exponential time-course after disruption from equilibrium at \(t = 0\) [6]. The conditions chosen for these examples favor (a) folding, and (b) unfolding. The parameters in (a) are \(k_{U \rightarrow I} = 4, k_{I \rightarrow U} = 2, k_{I \rightarrow F} = 2, k_{F \rightarrow I} = 1, [F_0]/[UIF] = 0, [U_0]/[UIF] = 1\), and in (b) \(k_{U \rightarrow I} = 4, k_{I \rightarrow U} = 2, k_{I \rightarrow F} = 4, k_{F \rightarrow I} = 1, k_{F \rightarrow U} = 2, [F_0]/[UIF] = 1, [U_0]/[UIF] = 0\). The insets show the corresponding free energy profiles [6].

and then gets partially trapped by the relatively high kinetic barrier that separates I and F [6]. On the contrary, under conditions that favor unfolding \((\Delta G_{F-U}^c > 0)\), the folding intermediate is not transiently accumulated, and its population level never exceeds the equilibrium population level (see Fig. 1.10b) [6].

1.4 Outline of this thesis

This thesis describes experimental investigations of protein folding mechanisms. In the previous sections, we have introduced the basic structural features of proteins, the protein folding problem, and how protein folding processes can be followed in real time using time-resolved vibrational spectroscopy. A detailed description of the experimental techniques used in the acquisition of equilibrium and non-equilibrium protein folding data can be found in chapter 2. In chapter 3, we study the influence of \(\text{Glu}^-/\text{Arg}^+\) salt bridges on the folding and unfolding dynamics of different \(\alpha\)-helical model systems. We extend this investigation in chapter 4 by studying the kinetic effect of salt bridges involving different type of charge pairs, viz. \(\text{Glu}^-/\text{Arg}^+, \text{Asp}^-/\text{Arg}^+,\) or \(\text{Glu}^-/\text{Lys}^+\). This chapter also includes complementary molecular dynamics simulations that have been carried out to obtain detailed insights into the correlation between the salt bridge population and the \(\alpha\)-helical content of the investigated peptides. In chapter 5, we study the chemical-denaturation mechanism of guanidinium, and particularly its potential effect on salt-bridge interactions. A detailed investigation of the folding and unfolding mechanism of the Trp-cage miniprotein using temperature-jump infrared-probe spectroscopy and molecular dynamics simulations is presented in chapter 6.
In chapter 7, we investigate the temperature-jump-induced (un)folding dynamics of an α-helical peptide and show how transient infrared dispersed pump-probe spectroscopy can be applied to observe time-dependent 2D-IR cross peaks. In the last chapter, we use this method to investigate the folding and unfolding pathways of the zinc finger ββα-motif FSD-1.
References


