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

Meuzelaar, H.

Publication date
2015

Document Version
Final published version

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Trp-cage is a synthetic 20-residue miniprotein which folds rapidly and spontaneously to a well-defined globular structure more typical of larger proteins. Due to its small size and fast folding, it is an ideal model system for experimental and theoretical investigations of protein folding mechanisms. However, Trp-cage’s exact folding mechanism is still a matter of debate. In this chapter we investigate Trp-cage’s relaxation dynamics in the amide I’ spectral region ($1530-1700 \text{ cm}^{-1}$) using time-resolved infrared spectroscopy. Residue-specific information was obtained by incorporating an isotopic label ($^{13}\text{C}=^{18}\text{O}$) into the amide carbonyl group of residue Gly11, thereby spectrally isolating an individual $3_{10}$-helical residue. The folding-unfolding equilibrium is perturbed using a nanosecond temperature-jump ($T$-jump) and the subsequent re-equilibration is probed by observing the time-dependent vibrational response in the amide I’ region. We observe bimodal relaxation kinetics with time constants of $100 \pm 10$ ns and $770 \pm 40$ ns at 322 K, suggesting that the folding involves an intermediate state, the character of which can be determined from the time- and frequency-resolved data. We find that the relaxation dynamics close to the melting temperature involve fast fluctuations in the polyproline II region, whereas the slower process can be attributed to conformational rearrangements due to the global (un)folding transition of the protein. Combined analysis of our $T$-jump data and molecular dynamics simulations indicates that the formation of a well-defined $\alpha$-helix precedes the rapid formation of the hydrophobic cage structure, implying a native-like folding intermediate, that mainly differs from the folded conformation in the orientation of the C-terminal polyproline II helix relative to the N-terminal part of the backbone. We find that

---

the main free-energy barrier is positioned between the folding intermediate and the unfolded state ensemble, and that it involves the formation of the $\alpha$-helix, the 3\textsubscript{10}-helix, and the Asp9-Arg16 salt bridge. Our results suggest that at low temperature ($T \ll T_m$) a folding path via formation of $\alpha$-helical contacts followed by hydrophobic clustering becomes more important.

### 6.1 Introduction

Understanding the mechanism by which proteins fold into their specific three-dimensional structure is a central issue in molecular biology and remains a largely unsolved problem [1, 2]. Moderately large proteins (>70 residues) are commonly believed to follow a hierarchical folding mechanism that proceeds through the occurrence of one or more intermediate states, represented by free-energy minima along the reaction coordinate [2–5]. It has been postulated that the formation of globule-like folding intermediates facilitates the fast folding process as they restrict the conformational search of the protein, thereby directing it towards a well-defined folding pathway [6]. However, since the discovery of several two-state folding proteins [7–12], of which the free-energy minima along the folding pathway are only occupied by the unfolded and folded state, the role of folding intermediates is under considerable discussion [1, 6, 13–19]. Particularly, the existence of folding intermediates involved in the folding process of small-sized, structurally simple proteins is under debate.

Probably one of the best known examples of small-sized folding problems is the 20-residue “miniprotein” Trp-cage (NLYIQ WLKDGPSSG RPPPS) designed by Neidigh et al. [20]. In spite of its short sequence and compact structure, Trp-cage folds rapidly and spontaneously into a globular structure with well-defined secondary and tertiary structure elements. It folds on the microsecond timescale [21] to a native state with an N-terminal $\alpha$-helix, followed by a short 3\textsubscript{10}-helix, a C-terminal polyproline II region, and a hydrophobic core in which the aromatic side-chains of Tyr3 and Trp6 pack against residues Gly11, Pro12, Pro18, and Pro19 (see Fig. 6.1) [20]. Enhanced fold stabilization of Trp-cage’s native state is provided through the formation of a salt bridge between residues Asp9 and Arg16.

Although Trp-cage’s folding mechanism has been extensively investigated both experimentally [20–35] and theoretically [36–48], controversies regarding the existence of intermediate states along its folding pathway remain. Circular dichroism (CD) spectroscopy, differential scanning calorimetry, and nuclear magnetic resonance (NMR) spectroscopy indicated that Trp-cage’s folding mechanism can be represented by a simple two-state model [20, 26, 49]. The absence of folding intermediates was also suggested by the results of two different temperature-jump studies that showed single-exponential relaxation kinetics monitored using tryptophan fluorescence spectroscopy [21] and amide I’ infrared spectroscopy [30].

On the other hand, fluorescence correlation spectroscopy revealed hierarchical folding through the early formation of a compact globule-like intermediate state [23]. This finding was supported by UV-resonance Raman spectroscopy
measurements that show a compact folding intermediate with a partially unfolded α-helix, in which the polyproline and/or 3\textsubscript{10}-helix is in closer proximity to Trp6, indicating reduced distances between Gly11, Pro12, and Trp6 [22]. Simulated vibrational circular dichroism and two-dimensional infrared spectroscopy indicated a similar folding pathway [48]. Additionally, temperature-jump relaxation data probed at three different frequencies in the amide I’ region revealed bimodal relaxation kinetics at temperatures below 293 K, demonstrating the existence of a low-temperature folding intermediate [31]. The authors proposed that local unfolding of the 3\textsubscript{10}-helix occurs at lower temperatures (< 293 K) than the global unfolding of the hydrophobic core structure. The existence of a low-temperature, fast-exchanging intermediate state with high similarity to the native structure and a completely formed α-helix was also indicated by recently published NMR and ECD data [35]. In contrast, other systematic investigations using NMR and CD spectroscopy suggest that the folding intermediate consists of a hydrophobically collapsed structure and that the secondary structural elements are formed in a later stage [28, 32]. Additional evidence for the early formation of tertiary contacts prior to formation of helical structure was provided by several theoretical investigations [37, 39, 40, 43]. Subsequently, it was proposed that beside a predominant pathway involving the early formation of hydrophobic contacts with less helical content, a parallel, coexisting route exists in which α-helical contacts are formed prior to hydrophobic clustering [40, 43, 50].

In this chapter, we report a systematic investigation of Trp-cage’s (TC9b) folding mechanism using time-resolved infrared (IR) spectroscopy following a laser-induced temperature-jump (T-jump). Time- and structural resolution is obtained

![Figure 6.1: Stick (a) and ribbon (b) representation of Trp-cage’s (TC9b) folded structure (PdB entry: 1L2Y) showing the α-helix (blue), the 3\textsubscript{10}-helix (green), the polyproline II region (red), the hydrophobic side-chains, and the salt-bridging side-chains of Asp9 and Arg16. An isotopic label (\textsuperscript{13}C=\textsuperscript{18}O) was incorporated into the amide carbonyl group of residue Gly11 (yellow). Pictures rendered with Chimera.[51]
in the amide I' spectral region (1530–1700 cm⁻¹) which is very sensitive to conformational changes [52]. Four key frequencies providing structural specificity are studied in particular, viz., the carbonyl stretch of ¹³C=¹⁸O labeled Gly11 (1554 cm⁻¹), the side-chain carboxylic group of residues Asp9 (1575 cm⁻¹), the imide groups of the polyproline (II) helix (1624 cm⁻¹), and the high frequency amide I' mode (1664 cm⁻¹) related mainly to the amount of α-helical content. Our T-jump IR-probe study at a final T-jump temperature close to Trp-cage’s melting temperature reveals relaxation kinetics that can be well described by a bi-exponential function, and involves fast fluctuations in polyproline II helical structure. The observed relaxation time constants derived from a global bi-exponential fit are 100 ± 10 ns and 770 ± 40 ns at 322 K. Accordingly, our findings do not support that Trp-cage’s folding pathway can be represented by a simple two-state transition. In contrast, SVD and global fitting analysis provide evidence that Trp-cage follows a hierarchical folding mechanism and proceeds through the formation of an effective intermediate state, the nature of which can be characterized from a combined analysis of the time- and frequency-resolved T-jump data, and molecular dynamics (MD) simulations.

6.2 Unfolding in thermal equilibrium

The equilibrium temperature-induced unfolding transition of Trp-cage was characterized using CD and FTIR spectroscopy. Temperature-dependent CD spectra were measured between 257.5–190 nm in a temperature range of 273–368 K (see Fig. 6.9). SVD and global fitting analysis of the molar ellipticity as a function of temperature resulted in a thermal melting point (T_m) of 323.1 ± 1.7 K, in quantitative agreement with the transition midpoint of 324.0 K reported previously for TC9b [30].

The amide I’ region of the FTIR spectra of unlabeled and isotopically labeled Trp-cage are shown in Fig. 6.2a. The major peak centered at 1645 cm⁻¹ arises from the amide I’ vibrations of the backbone residues involved in an α-helical conformation (this peak contains a small contribution from the 3₁₀-helix) [52, 53], whereas the shoulder appearing at 1624 cm⁻¹ is due to contributions of the backbone amide residues that adopt a polyproline II helix [54–56]. The minor component at the low-frequency edge of the amide I’ peak that occurs at 1584 cm⁻¹, is due to side-chain contributions of the carboxylic groups (COO⁻) of residue Asp9 and the C-terminus [52, 57]. In addition, in the spectrum of the isotope-labeled Trp-cage, the carbonyl stretching frequency of the ¹³C=¹⁸O label incorporated at residue Gly11 is shifted to a lower energy relative to the main amide I’ peak and appears at 1554 cm⁻¹, thereby spectrally isolating an individual 3₁₀-helical residue involved in the hydrophobic packing of the Trp-cage.

To study the equilibrium spectral changes produced by thermal unfolding, FTIR spectra were collected between 274 K and 363 K with incremental steps of 5 K. FTIR difference spectra were generated by subtracting the spectrum recorded at 274 K from those collected at higher temperatures (Fig. 6.2b,c). By
Figure 6.2: Equilibrium FTIR absorption spectra of unlabeled (violet) and isotopically labeled (green) Trp-cage in the Amide I' region measured at 298 K (a). The peak corresponding to the carbonyl stretch of the $^{13}\text{C}=^{18}\text{O}$ label for residue Gly11 occurs at 1554 cm$^{-1}$. Temperature-dependent FTIR difference spectra were measured from 274 K and 363 K of unlabeled (b) and isotopically labeled (c) Trp-cage. FTIR difference spectra were generated by subtracting the FTIR spectrum collected at 274 K from those measured at higher temperatures. All spectra were corrected for D$_2$O background. The inset shows the solvent-corrected IR absorption of Trp-cage as a function of temperature, monitored at 1664 cm$^{-1}$ and fitted to a two-state model (eq. 2.2) [58]. pH* = uncorrected pH meter reading in D$_2$O.
comparing with the temperature-dependent FTIR response of an aqueous solution of N-methylacetamide (NMA), which contains a single trans-peptide group (Fig. 6.9), we conclude that the temperature-induced IR spectral changes of Trp-cage are to a large extent due the intrinsic temperature dependence of the amide I’ mode. We find that the FTIR response of NMA shows a linear temperature dependence, so we can assign the sigmoidal transition observed in Trp-cage’s IR response (inset Fig. 6.2b) to the temperature-induced conformational changes upon unfolding in thermal equilibrium. Comparison between the thermal melting curves obtained from the UV-CD and FTIR measurements (Fig. 6.9), indicates that Trp-cage’s FTIR thermal unfolding curve is a combination of a linear function (intrinsic temperature-dependence of the amide I’ mode) and a sigmoid function (conformational changes). The negative spectral features observed in Trp-cage’s FTIR difference spectra (Fig. 6.2b,c) upon increasing temperature (1600–1650 cm$^{-1}$) are indicative of the loss of secondary and tertiary structural contacts, whereas the positive features (1650–1700 cm$^{-1}$) are attributed to an increase of the unfolded population due to melting of the helical structure [52, 53]. Specifically, the broad negative feature centered at 1624 cm$^{-1}$ follows the loss of polyproline II structure and breaking of hydrogen bonds to the proline imide group [54, 56]. The shoulder appearing at the high-frequency edge of this negative feature observed at 1642 cm$^{-1}$, involves contributions of both polyproline II and α-helical unfolding due to spectral overlap of the polyproline II and α-helical subcomponents [52, 53]. The region of increasing intensity centered at 1664 cm$^{-1}$ is indicative of the decrease in α- and 3_10-helical content in response to unfolding [53]. Moreover, the negative feature found at 1575 cm$^{-1}$ arises from the change in absorbance of the side-chain carboxylic groups of residue Asp9 and the C-terminus [52, 57]. Accordingly, this spectral signature provides unique information about the unfolding of the hydrophobic cage structure due to the salt bridge breaking [31, 59]. The absorbance change of the $^{13}$C=$^{18}$O amide stretching mode shows a negative feature centered at 1554 cm$^{-1}$, indicating the loss of secondary structure involving residue Gly11 (Fig. 6.2c), and uniquely reflecting the thermal melting of an individual 3_10-helical residue that is involved in the formation of the hydrophobic cage structure of Trp-cage. It should be noted that unlike the main amide I’ peak, no positive features upon thermal unfolding arise from the $^{13}$C=$^{18}$O label in the FTIR difference spectrum of isotopically labeled Trp-cage (Fig. 6.2c). Possibly, the positive features that presumably arise from the temperature-induced blue-shift of the $^{13}$C=$^{18}$O amide stretching mode of residue Gly11 are canceled due to spectral overlap with the carboxylate groups (Fig. 6.2b,c).

6.3 Temperature-jump dynamics

6.3.1 $T$-jump dynamics at the melting temperature

We probed the relaxation dynamics of Trp-cage following a nanosecond $T$-jump in the amide I’ spectral region (1530–1700 cm$^{-1}$). Fig. 6.3 shows a solvent-corrected three-dimensional representation of the time- and frequency-dependent relaxation
kinetics in response to a $T$-jump of 5 K. The final $T$-jump temperature of 322 K is close to the melting temperature of the unfolding transition ($T_m = 323.1 \pm 1.7$ K). We find that our data exhibit relaxation kinetics that can be well described by a bi-exponential function with a fast and a slow phase.

To validate that the fast phase is related to the (un)folding of Trp-cage, reference transient absorption data were obtained by measuring the $T$-jump response of an aqueous solution of NMA. The transient absorption changes of NMA following a $T$-jump (Fig. 6.10), show an instantaneous response that rises concomitantly with the time-resolution of the $T$-jump pulse and does not exhibit a kinetic phase. The instantaneous response arises from the temperature-induced shift of amide I’ band [60]. Accordingly, we conclude that the fast relaxation rate is due to conformational changes of Trp-cage as it progresses through re-equilibration. The presence of two well-separated relaxation phases (Fig. 6.3) suggest that Trp-cage’s folding pathway cannot be represented by a simple two-state transition. For a strict two-state folder, the observed relaxation constant ($k_R$) is simply the sum of the folding ($k_F$) and the unfolding ($k_U$) rate constants, and single-exponential relaxation kinetics is detected (see section 1.3.2) [7, 61]. However, when the folding involves an intermediate state, the $T$-jump disturbs multiple steps along the folding pathway and multiple-exponential relaxation kinetics will be observed [4, 62–64], and the observed relaxation time constants depend in a complicated manner on the underlying rate constants of the $U \rightleftharpoons I \rightleftharpoons F$ interconversions (see section 1.3.2) [64]. From the multiple-exponential relaxation kinetics, we conclude that at a temperature close to the transition midpoint, Trp-cage’s relaxation dynamics involves at least one intermediate state, which is consistent with earlier findings [34].

SVD analysis of our kinetic data were performed to confirm the presence of two relaxation phases and to model Trp-cage’s folding mechanism in an objective manner. As can be seen in Fig. 6.11 a, four principal components were required to describe the $T$-jump relaxation data. The first target vector represents a relatively slow phase with the largest signal amplitude, and reports on the absorption changes due to the global (un)folding transition. The second and third target vectors both represent a fast phase. One of these should account for an instantaneous response arising from the intrinsic temperature-dependence of the amide I’ mode. However, the other fast component represents the absorption changes induced by the formation of a folding intermediate as Trp-cage progresses through re-equilibration. The fourth target vector shows a fast phase as well, but considering the small signal amplitude relative to the amplitude of the first component ($\approx 0.5\%$), this target vector is probably not related to the (un)folding kinetics and possibly due to experimental noise introduced by the $T$-jump, e.g., cavitation effects [65], and/or electronic noise introduced by the Q-switch of the Nd:YAG laser. The fifth and higher vectors of the SVD all represent uncorrelated noise contributions, and the corresponding amplitudes were negligible relative to the amplitude of the first component ($< 0.5\%$). Consistent with these findings, we conclude that Trp-cage’s folding transition involves the formation of an effective intermediate state and model it according to three-state analysis ($U \rightleftharpoons I \rightleftharpoons F$) using a bi-exponential function [64].
Global fitting of the time- and frequency-resolved data as presented in Fig. 6.3 shows two well-separated phases that can be described by a bi-exponential function. To this purpose, the relaxation kinetics probed at the different frequencies in the 1530–1700 cm\(^{-1}\) region were fitted simultaneously to a bi-exponential function with global (frequency-independent) time constants (\(\tau_1\) and \(\tau_2\)). The observed \(T\)-jump relaxation time constants derived from this global fit are \(\tau_1 = 100 \pm 10\) ns and \(\tau_2 = 770 \pm 40\) ns, respectively. The fast relaxation time constant likely represents fast fluctuations due to a rapid pre-equilibrium in the folded or unfolded free-energy basin, whereas the slower process reflects crossing the rate-limiting free-energy barrier of the folding process and represents the global (un)folding of the Trp-cage miniprotein.

### 6.3.2 Transient absorption changes

Fig. 6.4 depicts solvent-corrected transient IR spectra of unlabeled (a) and isotopically labeled (b) Trp-cage at several time delays after the \(T\)-jump pulse. At a short time delay of 20 ns, we observe that Trp-cage’s spectrum is dominated by a negative peak centered at 1624 cm\(^{-1}\), indicating the loss of polyproline II structure. The frequency-dependent absorption changes become larger as a func-
6.3 Temperature-jump dynamics

Figure 6.4: Solvent-corrected transient absorption changes of unlabeled (a) and isotopic labeled (b) Trp-cage at several time delays after the T-jump pulse (T-jump from 317 to 322 K). The curves are a guide to the eye. The different structure-sensitive vibrational modes are the carboxylate group (1575 cm$^{-1}$), the proline II helix (1624 cm$^{-1}$), the $\alpha$- and $3_{10}$-helices (1664 cm$^{-1}$) (a), and the carbonyl stretch of residue Gly11 (1554 cm$^{-1}$) (b).

tion of time as the system re-equilibrates and reaches its maximum value when the new equilibrium conditions are achieved (time delays > 3000 ns). At these longer time delays, the observed transient spectrum is essentially identical to the equilibrium FTIR difference spectrum (Fig. 6.2). The different structure-sensitive vibrational modes as considered in subsection 6.2 (Fig. 6.2), are clearly present in the transient spectra observed after longer time delays (Fig. 6.4. a,b), and arise from the structural changes involving residue Gly11 (1554 cm$^{-1}$), the side-chain carboxylic group (COO$^-$) of residue Asp9 (1575 cm$^{-1}$), the polyproline II helix (1624 cm$^{-1}$), and the $\alpha$- and $3_{10}$-helices (1664 cm$^{-1}$).

The globally fitted relaxation dynamics monitored at these different key frequencies are shown in Fig. 6.5. Interestingly, we find that the relative amplitudes of the kinetic responses differ for the relaxation probed at different structure-sensitive modes. The transient absorption changes monitored at 1554 cm$^{-1}$ (Gly11), 1575 cm$^{-1}$ (COO$^-$), and 1664 cm$^{-1}$ ($\alpha$- and $3_{10}$-helix) are dominated by a single slow kinetics phase, whereas significant amplitude of the fast phase is observed in the transients measured at 1624 cm$^{-1}$ (Pro-helix) and 1642 cm$^{-1}$ (Pro- and $\alpha$-helix).

To quantify the frequency-dependence of the relative contribution of the two kinetic phases, decay-associated spectra (Fig. 6.6) were constructed by plotting the relative amplitudes derived from the global bi-exponential fit as a function of frequency. As can be seen in Fig. 6.6a, the decay-associated spectra confirm that the relative amplitude of the slow phase is predominating at 1575 cm$^{-1}$ (COO$^-$) and 1664 cm$^{-1}$ ($\alpha$- and $3_{10}$-helix), whereas significant contribution of the fast
phase is observed at frequencies where the polyproline II helix absorption change occurs (1624 and 1642 cm⁻¹). Thus, the fast kinetic phase likely follows rapid changes in the conformational state of the polyproline II helix [54–56] and/or its environmental conditions [56, 66, 67] (proline cis-trans isomerization can be excluded, as it typically occurs on a time scale of seconds [68]). We find that the decay-associated spectrum of isotope-labeled Trp-cage (Fig. 6.6b) shows no contribution of the fast phase at the frequency where the $^{13}$C=O amide stretching mode of residue Gly11 appears (1554 cm⁻¹). Consequently, our results obtained at a temperature close to the melting temperature indicate that re-equilibration occurs through fast fluctuations in the content and/or environmental conditions of the polyproline II structure, whereas the conformational rearrangement associated with formation of the $\alpha$-helix, the salt bridge, and the secondary structure elements involving residue Gly11 ($\beta$-helix) are represented by a slow kinetic phase. The fast fluctuations observed in the polyproline II region might have been hidden in previously reported $T$-jump studies, possibly due to insufficient structural resolution. Selective monitoring of the $T$-jump response of specific backbone residues might locally appear as two-state behavior. Qiu et al. [21] reported single-exponential relaxation kinetics monitored by probing the $T$-jump-induced changes in tryptophan fluorescence, thereby selectively probing the relaxation dynamics involving residue Trp6. On the other hand, Culik et al. [31] measured
6.3 Temperature-jump dynamics

Figure 6.6: Decay-associated spectra of unlabeled (a) and isotope-labeled (b) Trp-cage, obtained by plotting the pre-exponential factors of the fast (blue) and slow (cyan) phase derived from the global bi-exponential fit as a function of frequency. The curves are a guide to the eye.

The $T$-jump relaxation kinetics in the amide I’ region at three different frequencies, viz. 1580 cm$^{-1}$ (COO$^-$), 1612 cm$^{-1}$ (13C-labeled Ala2 and Ala4 residues), and 1664 cm$^{-1}$ ($\alpha$- and 310-helix), but not at the frequency where the polyproline II helix participates (1624 cm$^{-1}$). Single-exponential relaxation kinetics were reported at all three frequencies for temperatures above 293 K, indicative of a two-state folding process. At final $T$-jump temperatures below 293 K, bi-exponential relaxation kinetics was observed at 1664 cm$^{-1}$, which provided evidence for the occurrence of a low-temperature folding intermediate characterized by a partially molten 310-helical structure.

6.3.3 $T$-jump dynamics at low temperature

Additional $T$-jump experiments were performed at lower temperatures. At low temperature ($T \ll T_m$), the contribution of the folding process to the $T$-jump relaxation kinetics is dominating (see section 1.3.2). Because the $T$-jump-induced population change is relatively small at low temperature ($T \ll T_m$), correspondingly small transient signals due to Trp-cage’s conformational redistribution process are observed, and the major contribution to the transient absorbance arises from the intrinsic temperature-dependence of the amide I’ mode. As a consequence, even our maximum achievable signal-to-noise ratio at low initial $T$-jump temperatures is not sufficient to derive unambiguous conclusions from the data obtained at
1575 cm$^{-1}$ (salt bridge) and 1554 cm$^{-1}$ (Gly11). In Fig. 6.7a we compare Trp-cage's relaxation dynamics at different final $T$-jump temperatures monitored at the frequency where the polyproline helix participates. SVD analysis again reveals the presence of two kinetic phases (Fig. 6.11 b), and global fitting to a bi-exponential function results in observed $T$-jump relaxation time-constants of $\tau_1 = 90 \pm 25 \text{ ns}$ and $\tau_2 = 2.5 \pm 0.7 \text{ µs}$ at 285 K, $\tau_1 = 150 \pm 25 \text{ ns}$ and $\tau_2 = 2.2 \pm 0.2 \text{ µs}$ at 300 K, and $\tau_1 = 180 \pm 25 \text{ ns}$ and $\tau_2 = 1.3 \pm 0.1 \text{ µs}$ at 313 K (the uncertainties in these time-constants obtained from the global fit were estimated using a fit to the time-dependency at a single frequency. Since we use an array detector, the experimental errors in the data points measured at different frequencies but at the same time-delay are correlated. Hence, the uncertainty obtained from the global fit would be an underestimate). The temperature dependence of the observed relaxation constants is not straightforward to interpret, as these constants depend in a complicated manner on the underlying rate constants of the folded $\rightarrow$ intermediate and intermediate $\rightarrow$ unfolded re-equilibration processes (see section 1.3.2) [64].

Both the slow and fast relaxation rates increase with temperature, probably because all underlying rate constants increase with temperature. The faster $\tau_1$-value at the lowest temperature, which deviates from this trend, might be a result of the additional free-energy minimum at these temperatures discussed below (section 6.4). The relative amplitude of the fast kinetic phase compared to that of the slow phase becomes significantly smaller at lower temperature.

The corresponding decay-associated spectra at different final $T$-jump temperatures are shown in Fig. 6.7b and c. The slow kinetic phase contains contributions of both the polyproline II structure and the $\alpha$-helix at all different temperatures (Fig. 6.7b). The transient absorbance due to the polyproline II helix shifts to a slightly lower frequency at decreasing temperature (in the spectra of both the fast and slow phase), due to the increased number of hydrogen bonds with the proline imide group [56]. The improved spectral resolution due to the red-shift of the polyproline II transient absorbance compared to the transient absorbance arising from the $\alpha$- and 3$_{10}$-helices results in two well-resolved peaks in the decay-associated spectra of the slow phase at 285 K. The decay-associated spectra of the fast kinetic phase at different temperatures differ significantly (Fig. 6.7c). At low temperature (285 K) these spectra contain similar contributions of the polyproline II (1618 cm$^{-1}$) and $\alpha$- and 3$_{10}$-helical (1664 cm$^{-1}$) frequency ranges. This indicates that the rapid fluctuations at low temperature (285 K) involve both the polyproline II structure and the remaining, mostly $\alpha$-helical structure. With increasing temperature, the relative amplitude of the fast kinetic phase observed at 1624 cm$^{-1}$ (Pro-helix) increases significantly more compared to its relative amplitude observed at 1664 cm$^{-1}$ ($\alpha$- and 3$_{10}$-helix). Eventually, the decay-associated spectrum of the fast phase at 322 K is dominated by contributions involving the polyproline II helix. These results suggest that the fast fluctuations observed during re-equilibration at low temperature involve rapid structural rearrangements of both the polyproline II structure and the $\alpha$- and 3$_{10}$-helices. As the temperature increases, the relative contribution of the rapid fluctuations involving the polyproline II region becomes dominant.
6.4 Comparison with molecular dynamics simulations

In order to complement and further interpret our $T$-jump relaxation results, we performed replica exchange molecular dynamics (REMD) simulations (see Ref. [47] and section 6.6.2). A major objective for revisiting these simulations was to elucidate the structural nature of the slow and fast phase, and whether the fast polyproline helix phase is related to a transition close to the native state, or rather to one inside the unfolded state. While Ref. [47] focused on ambient condition, we present here the REMD results for 281 K and 320 K, to compare with the $T$-jump experiments at low and high temperature. The trajectories from the REMD simulations were analyzed by calculating the free-energy landscapes (see section 6.6.2). Fig. 6.8a-d shows the free-energy landscapes at 281 K and 320 K as a function of $\text{rmsd}_{\text{Pro}}$, $\text{rmsd}_{\text{Alx}}$, and $\text{rmsd}_{\text{Co}}$. The darker blue areas on the plots represent free energy minima, while the lighter blue and green areas are higher in energy. Less
area is sampled in Fig. 6.8a,c than Fig. 6.8b,d since the lower temperature results in fewer thermal fluctuations. The free-energy landscapes in Figs. 6.8a and b are a function of rmsd_{ohx}, and rmsd_{Ca}. Small values of these parameters indicate a folded structure and large values, the unfolded structure, as indicated by the labels in the plot. Fig. 6.8e and f depict structures which are representative of the minima as determined by clustering analysis. For each minimum this structure is shown twice, viewed from different angles. As there are many possible unfolded structures, only one conformation is shown for each temperature for comparison with the folded structure.

Fig. 6.8d shows the free-energy landscape of Trp-cage in explicit solvent close to its transition temperature (T = 320 K) as a function of order parameters rmsd_{Pro} and rmsd_{ohx}. The free-energy minimum located at small values of rmsd_{Pro} and rmsd_{ohx} represents the native structure, whereas the unfolded state ensemble in which the content of well-defined secondary structure elements is minimal, is found at high values of rmsd_{Pro} > 0.25nm and rmsd_{ohx} > 0.2nm. A local free-energy minimum representing an intermediate state is located at rmsd_{ohx} ≈ 0.02 nm and at rmsd_{Pro} ≈ 0.20 nm, with the main free-energy barrier positioned between the folding intermediate and the unfolded state ensemble. Rapid fluctuations (indicated by the red arrow) can occur between the minima in the folded free-energy basin (approximately at rmsd_{ohx} < 0.08 nm and at rmsd_{Pro} < 0.40 nm), while the main free-energy barrier between the intermediate and the unfolded state ensemble, correspond to a relatively slow component of the folding process. This indicates that a possible folding pathway proceeds through a native-like intermediate with well-defined α-helical structural elements, but mainly differs from the folded structure in the polyproline II region. These observations agree well with our T-jump relaxation data, that show fast fluctuations at frequencies where the polyproline II helix participates (1624 and 1642 cm\(^{-1}\)), whereas the relaxation dynamics observed at frequencies where the α- and 3_{10}-helices appear (1664 cm\(^{-1}\)) can be well-described by a single, slow kinetic phase. We cannot determine with certainty that the intermediate state observed in the T-jump experiments is an obligatory folding intermediate. However, the simulations indicate that a majority of the folding/unfolding paths do go through the intermediate from the native to the unfolded state. While there is the possibility that paths go directly from the native state towards the unfolded state, via the so-called loop state [40, 47], the simulations suggest that the majority of the paths follow the route via the experimentally observed native-like intermediate.

\(k\)-Medoid clustering analysis shows that the intermediate state at rmsd_{ohx} = 0.02 nm and rmsd_{Pro} = 0.20 nm found in Fig. 6.8d is a native-like structure, in which the C-terminal polyproline II helix is not completely packed against the N-terminal α-helix, thereby partially disrupting the native hydrophobic core cluster (Fig. 6.8f). The different orientations of the Pro rings are clearly visible when comparing both views of the folded and intermediate conformations. This is supported by other free-energy landscapes calculated for Trp-cage (Fig. 6.12), which show that the polyproline II helix stays folded (small values of rmsd_{Pro fit}) regardless of the overall structure of the protein. Therefore, the fast kinetic phase
6.4 Comparison with molecular dynamics simulations

Figure 6.8: (a)-(d) Calculated free-energy landscapes as a function of rmsd$_{C\alpha}$ and rmsd$_{\alpha\beta}$ and (c,d) rmsd$_{\text{Pro}}$ (entire backbone fit to NMR structure before rmsd calculation) and rmsd$_{\alpha\beta}$ for Trp-cage folding at $T = 281$ K (a,c), and $T = 320$ K (b,d). Contours are separated by $1k_B T$. Color scale: blue (lowest free-energy) to green (highest free-energy). The area in white was not sampled. (e,f) Structures of representative folded and intermediates from clustering for the indicated minima in (a,c). The two rows depict each structure from a different angle. One of many possible unfolded structures is presented for comparison. The residues containing the $\alpha$-helix, $3_{10}$-helix, and polyproline II helix are colored blue, green, and red, respectively. The side chains of Trp6, Pro17, Pro18, and Pro19 are represented as sticks.

observed in our $T$-jump relaxation data at frequencies where polyproline II helix absorption change occurs, likely involves rapid environmental rearrangements of the (conformationally intact) polyproline II helix. These fluctuations of the intact
polyproline II helical structure will increase its hydration, a process which is known to result in an IR intensity loss of the proline peak [56], in agreement with the observed decay-associated spectrum (Fig. 6.7c).

In Fig. 6.8c and 6.8d, we compare Trp-cage’s free-energy landscapes calculated at 320 K and at 281 K, respectively. We note that it is often difficult to extract reliable free-energy barriers from projection onto arbitrary order parameters [47]. Nevertheless it is clear that there is a qualitative difference between the free-energy landscapes at the two temperatures. The relative free-energy barrier between the folded and unfolded state ensemble is $2k_B T$ higher in energy at 281 K compared to 320 K. Both energy landscapes exhibit a local free-energy minimum positioned at rmds$_{ohx}$ 0.02 nm and at rmds$_{Pro}$ 0.20 nm, representing a native-like intermediate conformation. The relative free-energy difference between this folding intermediate and the native state at low temperature (281 K) is similar to that close to the melting temperature (320 K). The accessible conformational space in the folded energy basin at low temperature ranges from approximately rmds$_{ohx}$ 0 to 0.10 nm and rmds$_{Pro}$ 0 to 0.30 nm. An additional, relatively shallow, local free-energy minimum positioned at rmds$_{ohx}$ 0.09 and at rmds$_{Pro}$ 0.20 is found at low temperature (Fig. 6.8a and c), in which both the helical structure elements and the orientation of the polyproline II helix are less well-defined. The corresponding conformation (structure Pro12d, see Fig. 6.8e) lacks well-defined $3_{10}$-helical structure elements and shows that apart from the polyproline II helix (Pro17, 18, and 19) residue Pro12 is also detached from the hydrophobic core [40, 47]. Indeed, our $T$-jump relaxation data suggest that re-equilibration at low temperature ($T \ll T_m$) involves fast fluctuations in not only the polyproline II region, but also in the remainder of the backbone ($\alpha$- and/or $3_{10}$-helical region, see Fig. 6.7c). This is in agreement with the $T$-jump results reported by Culik et al. [31], who found evidence for the occurrence of a low-temperature folding intermediate lacking a fully formed $3_{10}$-helix.

Taken together, a combined analysis of our $T$-jump and REMD results provide evidence that Trp-cage’s folding pathway proceeds through the formation of a native-like intermediate conformation. The $T$-jump relaxation dynamics monitored at different structure-sensitive modes close to the transition temperature can be well-described by a bi-exponential function, and involves fast fluctuations of the polyproline II region (1624 cm$^{-1}$) (Fig. 6.5). The relaxations kinetics observed at frequencies where the $\alpha$-helix (1664 cm$^{-1}$), the $3_{10}$-helix (Gly11, 1554 cm$^{-1}$), and the Asp9-Arg16 salt bridge (COO$^-$, 1575 cm$^{-1}$) participate, can be described by a single slow phase (Fig. 6.6). The REMD results (Fig. 6.8) indicate the presence of native-like intermediate, with a fully formed $\alpha$-helix, but that differs mainly in the geometrical orientation of the (intact, but detached) polyproline II helix with respect to the rest of the peptide. The main free-energy barrier is positioned between the folding intermediate and unfolded state ensemble. Possibly, the folding intermediate contains a well-defined $\alpha$-helix, but lacks a completely formed hydrophobic cage structure due to the larger average distance of the polyproline II helix relative to the aromatic residues Tyr3 and Trp6. These findings are consistent with the simulation results published recently by Lai et al. [48], who
reported that the α-helical structure is formed prior to the formation of tertiary contacts (the coupling between N-terminal α-helix and the C-terminus). On the other hand, our $T$-jump results show that at low temperature ($T \ll T_m$) the rapid fluctuations involve both the polyproline II and the α- and $3_{10}$-helical frequency region (Fig. 6.7c). The REMD results suggest that this is due to an additional local free-energy minimum (Fig. 6.8c), in which the α-helix and the orientation of the polyproline II helix are less well-defined, and in which residue Pro12 is detached from the hydrophobic cage structure thereby disrupting the $3_{10}$-helix (see structure Pro12d, Fig. 6.8e). Indeed, the formation of such a low-temperature folding intermediate has been reported previously [31, 69].

### 6.5 Conclusions

Our $T$-jump IR-probe study shows bimodal relaxation dynamics, providing kinetic evidence that Trp-cage's folding process cannot be represented by a simple two-state transition and that the folding proceeds through the formation of an intermediate state. The nature of the folding intermediate was characterized based on the time- and frequency-resolved $T$-jump relaxation data combined with the results of REMD simulations. We find that the re-equilibration dynamics at temperatures close to the transition midpoint involve rapid fluctuations in the polyproline II content, whereas the conformational rearrangement involving the α-helix, the formation of the salt bridge between Asp9 and Arg16, and the secondary structure elements involving residue Gly11 ($3_{10}$-helix), are represented by a single slow kinetic phase. Our REMD data suggest that formation of α-helical structure precedes the formation of the tertiary structure: the correct folding of the polyproline II helix to complete the hydrophobic core. The folding intermediate contains a well-defined α-helix, but lacks a completely formed hydrophobic cage structure due to a larger average distance of the polyproline II helix relative to the aromatic residues Tyr3 and Trp6. However, our $T$-jump results show that at low temperature ($T \ll T_m$), rapid fluctuations occur due to rearrangements not only of the polyproline II region, but also of the other parts of the miniprotein. The REMD results indicate that this is due to an additional local free-energy minimum, in which the α-helix and the hydrophobic cage structure are less well-defined. This suggest the existence of an additional, low-temperature folding intermediate, in agreement with previous observations [31, 69].

### 6.6 Additional methods and data

#### 6.6.1 Materials and methods

**Peptide synthesis and purification**

A mutant of the Trp-cage miniprotein (TC9b) consisting of the sequence NAYAQ WLKDG GPSSG RPPPS is used in all experiments [30]. The Trp-cage minipro-
tein was purchased from GL Biochem (Shanghai). Peptide purity (≥95%) was assessed by reversed-phase HPLC and the identity of the purified product was confirmed by mass spectrometry. An isotopic label (13C=18O) was incorporated into the amide carbonyl group of the peptide backbone of residue Gly11. For introduction of the 13C=18O isotopic label, 18O/16O exchange was carried out starting from commercially acquired 13C(1)-labeled glycine (99% 13C, Sigma-Aldrich) according to a previously reported method [70] (progress of the 18O/16O exchange was monitored via 13C-NMR of the C(1)-signal). After Fmoc-protection, the thus obtained 13C=18O isotopic labeled glycine was used to prepare the labeled Trp-cage NAYAQ WLKD G*PSSG RPPPS using standard Fmoc solid phase peptide chemistry on a Protein Technologies Prelude automated synthesizer. After purification by preparative reversed-phase HPLC, the peptide was lyophilized against a 35% DCl/D2O solution to remove residual trifluoroacetic acid (TFA) and to achieve H/D-exchange. Trp-cage stock solutions of 15–20 mM were prepared by directly dissolving dry peptide in D2O. The pH* of the peptide solutions was adjusted to neutral pH level (pH*= 6.9) by addition of NaOD solution (pH* = uncorrected pH meter reading in D2O).

**Equilibrium temperature-dependent measurements**

UV circular dichroism (CD) spectra were measured using an Olis DSM-1000 CD Spectrophotometer equipped with a Julabo (CF31) temperature controller. The peptide concentration used for the CD measurements was 40 µM in a 20 mM phosphate buffer (pH = 6.8) and all measurements were performed using 2-mm quartz cuvettes. The temperature-dependency of the ellipticity was monitored between 257.5–190 nm with incremental steps of 0.75 nm and averaging over 5 s of data acquisition time (0.15 nm/min scanning speed). Singular value decomposition (SVD) and global fitting of the temperature-dependent dataset to an apparent two-state model was performed using the Olis GlobalWorks software package.

Temperature-dependent equilibrium Fourier-transform infrared (FTIR) spectra were collected using a Bruker Vertex 70 spectrometer. The peptides were kept at 15–20 mM concentration between 2 mm thick CaF2 windows separated by a 50 µm Teflon spacer. Temperature control was achieved by a circulating water bath. Typically, 75 scans were signal averaged at a resolution of 2 cm⁻¹ to generate one spectrum.

**Time-resolved T-jump IR experiments**

The frequency-dependent transient absorption changes induced in the sample in response to a nanosecond T-jump are detected at various time delays after the T-jump by mid-infrared (mid-IR) probe pulses, using the optical setup described in section 2.2.1. Because the frequency-dispersed detection allows measurements in a spectral window of approximately 100 cm⁻¹, two consecutive measurements with different mid-IR center frequencies were required to detect the transient absorption changes in the amide I’ spectral region between 1530 and 1700 cm⁻¹ and
to construct the transient spectra as shown in Fig. 6.3 and 6.4. From comparison of data sets with overlapping frequency ranges, we find that no scaling of the transient signals measured in different spectral windows is necessary.

### 6.6.2 Molecular dynamics simulations

Although Trp-cage is a very small protein, the accessible timescale of standard MD simulations (microseconds) is too short to allow for many transitions between the folded and unfolded states at room temperature. Replica exchange molecular dynamics (REMD) [71] overcomes this problem by allowing transitions at higher temperatures. In REMD, a number of parallel MD simulations (replicas) are run, each with a different temperature. After a short time, exchanges are attempted between the different replicas, and exchanges are accepted according to $P_{\text{acc}} = \min[1, e^{\Delta \beta E}]$. $E$ is the potential energy and $\beta = 1/k_BT$, where $k_B$ is the Boltzmann constant and $T$ is the temperature. REMD simulations of Trp-cage TC5b (NLYIQ WLKDG GPSSG RPPPS) were performed over the temperature range 270.0 to 556.0 K using 64 replicas [47]. Exchanges were attempted every 2 ps and 50,000 cycles were performed. Two independent simulations were started from a folded and unfolded Trp-cage structure for an aggregate simulation time of 200 ns for each temperature. In the simulation box, Trp-cage TC5b was solvated with 2796 TIP3P waters and one Cl-atom to neutralize the charge. The MD simulations were performed with the GROMACS MD engine [72] using the AMBER99SB force field [73] and a perl wrapper script was used to perform the exchanges. See Ref. [47] for further details.

The simulation trajectories of Ref. [47] were analyzed further by calculating several order parameters, $\lambda$, which were then used to calculate two-dimensional free-energy landscapes from probability histograms, $F(\lambda_1, \lambda_2) = k_BT \ln P(\lambda_1, \lambda_2)$ In order to improve statistics, data from the rejected exchange attempts was included in $P(\lambda_1, \lambda_2)$ through virtual move parallel tempering [74]. The order parameters used for this work were chosen to characterize the formation of the $\alpha$-helix and the polyproline II helix. They include the root mean square deviation of the $\mathrm{C}\alpha$ atoms (rmsd$_{\mathrm{C}\alpha}$) with respect to the NMR structure (PDB entry 1L2Y) and the $\alpha$-helical residues (rmsd$_{\alpha\text{hx}}$) with respect to an ideal $\alpha$-helix. In addition to these order parameters, which were also used in Ref. [47], we computed the rmsd of the polyproline II helix in two ways. First, the rmsd$_{\mathrm{Pro}}$ is calculated by fitting the entire protein backbone to the equilibrated NMR structure, so a large value indicates that the polyproline II helix is far from its position in the equilibrated structure and not that the polyproline II helix itself is unfolded. In order to quantify how much the helix is unfolded, the rmsd (rmsd$_{\mathrm{Pro fit}}$) is also calculated by first fitting only the three Pro residues of the polyproline II helix to the equilibrated, folded Trp-cage structure. Low values of rmsd$_{\mathrm{Pro fit}}$ indicate the helix is formed, while higher values indicate that it is unfolded. To find representative structures of the minima in the free energy landscape, k-medoid clustering was performed on the trajectories at 281 K and 320 K using the same procedure as in Ref. [47].
6.6.3 Additional data

Unfolding in thermal equilibrium using UV-CD and FTIR

Fig. 6.9 shows the thermal melting curves of Trp-cage (TC9b) detected at 222 nm using UV-CD (blue), and at 1664 cm$^{-1}$ using FTIR spectroscopy (red). Both UV-CD and FTIR thermal unfolding curves show a sigmoidal transition reflecting the temperature-induced structural changes of Trp-cage in thermal equilibrium. The temperature-dependency of Trp-cage’s FTIR response is dominated by the intrinsic temperature-induced blue-shift of the amide I’ band, as indicated by the temperature-dependent FTIR response of N-methylacetamide (NMA, green). The FTIR thermal unfolding curve is a combination of a linear function (intrinsic temperature-dependence of the amide I’ mode) and a sigmoid function (conformational changes). SVD and global fitting analysis of the CD data revealed a thermal melting point $T_m = 323.1 \pm 1.7$ K, which is in quantitative agreement with the transition midpoint of 324.0 K reported previously for Trp-cage TC9b [30].

Figure 6.9: Thermal melting curves of Trp-cage (TC9b) detected at 222 nm using UV-CD (blue), and at 1664 cm$^{-1}$ using FTIR spectroscopy (red). The green data points corresponds to the temperature-dependent FTIR response of N-methylacetamide (NMA, green; for better comparison of the IR spectral responses, the NMA concentration was set equal to the total concentration of amide groups in the Trp-cage sample) The solid curves represent least-square fits to a two-state unfolding transition (eq. 2.2) [58] (Trp-cage, CD and FTIR), and a linear fit (NMA, FTIR).

\[ T \text{-jump control measurements} \]

To verify that the origin of the kinetic phase arises from the folding/unfolding re-equilibration of Trp-cage, we measure the transient absorption changes of an aqueous solution of N-methylacetamide (NMA, which contains a single transpeptide group) in response to a nanosecond $T$-jump (Fig. 6.10). As can be seen in Fig. 6.10 a, the transient absorption changes of NMA shows an instantaneous $T$-jump response. However, the $T$-jump relaxation of NMA does not exhibit a kinetic phase, and we conclude that the observed kinetic phase as observed for the peptides arises from the conformational redistribution process of Trp-cage during re-equilibration.
Figure 6.10: (a) Transient absorption changes (uncorrected for solvent contributions) monitored at 1624 cm\(^{-1}\) for N-methylacetamide (NMA, green), D\(_2\)O (cyan), and Trp-cage (red) in response to temperature-jump (\(\Delta T = 5\) K). (b) Solvent-corrected transient spectra of NMA at several time delays after the T-jump pulse (\(\Delta T = 5\) K). The curves are a guide to the eye. The inset compares the solvent-corrected T-jump response of NMA (green) and Trp-cage (red) monitored at 1624 cm\(^{-1}\).

SVD analysis of the T-jump data

Figure 6.11: The weighted time-dependence of the first four target vectors resulting from singular-value decomposition of (a) the dataset as shown in Fig. 6.3 (\(T\)-jump from 317 to 322 K), and (b) the dataset as shown in Fig. 6.7 (\(T\)-jump from 277 to 285 K).
Calculated free-energy landscapes indicative of a folding intermediate containing an intact, but detached polyproline II helix

Fig 6.12 compares the calculated free-energy landscape as a function of $\text{rmsd}_{\text{Pro fit}}$ (rmsd of the Pro-helix with respect to its native local conformation) and $\text{rmsd}_{\text{ahx}}$ for Trp-cage folding at $T = 320$ K (top left), and $T = 281$ K (bottom left). The calculated free energy landscape as a function of $\text{rmsd}_{\text{Pro fit}}$ (entire backbone fit to NMR structure before rmsd calculation) and $\text{rmsd}_{\text{Pro fit}}$ (rmsd of the Pro-helix residues with respect to its native local conformation) at $T = 320$ K (top), and $T = 281$ K (bottom) are shown in Fig 6.12 (right). Comparison of the free-energy landscapes as a function of the different order parameters (left versus right), shows that the polyproline II helix stays folded (small values of $\text{rmsd}_{\text{Pro fit}}$) regardless of the overall structure of Trp-cage, indicating that the native-like folding intermediate differs mainly in the geometrical orientation of the (intact, but detached) polyproline II helix with respect to the rest of the peptide.

**Figure 6.12:** Left: calculated free-energy landscape as a function of $\text{rmsd}_{\text{Pro fit}}$ (rmsd of the Pro-helix with respect to its native local conformation) and $\text{rmsd}_{\text{ahx}}$ for Trp-cage folding at $T = 320$ K (top), and $T = 281$ K (bottom). Right: calculated free energy landscape as a function of $\text{rmsd}_{\text{Pro fit}}$ (entire backbone fit to NMR structure before rmsd calculation) and $\text{rmsd}_{\text{Pro fit}}$ (rmsd of the Pro-helix residues with respect to its native local conformation) at $T = 320$ K (top), and $T = 281$ K (bottom). Contours are separated by $1k_BT$. Color scale: blue (lowest free energy) to green (highest free energy). The area in white was not sampled.
References


