Acc. Chem. Res. **1998,** *31,* 709−716

Infrared Studies of Fast Events in Protein Folding

R. BRIAN DYER,* FENG GAI, AND WILLIAM H. WOODRUFF

Chemical Science and Technology Division, CST-4, Mail Stop J586, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

RUDOLF GILMANSHIN AND ROBERT H. CALLENDER

Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

Received December 17, 1997

I. Introduction

The specific function of a protein is determined by its structure and the ability of the structure to evolve with time. The functional three-dimensional structure is generally determined by the amino acid sequence,[1,2] and understanding how the sequence folds into the functional structure is a central problem in modern structural biology.[3−8] This problem may be stated as two questions. First, how is the three-dimensional structure encoded in the one-dimensional amino acid sequence? At present, the folded structure cannot be predicted ab initio with sufficient reliability, both because of the complexity of protein molecule itself and because of its interaction with surrounding water.[9−11] Second, what are the dynamics of the folding process? In addition to determining the three-dimensional structure of a protein, the amino acid sequence also codes the mechanistic paths that folding actually follows through the vast conformational space available to the polypeptide, so that the three-dimensional structure is reached in a reasonable time. Understanding the critical early events in folding is key to answering both of these questions.

Intense experimental effort has been devoted to determining the kinetics of protein folding, revealing that the early kinetics events on the folding pathway are of central importance.[4,6−8,12] These fast events set up all that are to follow in the guided pathway through conformational space. However, the earliest events occur on a submillisecond time scale, which has been difficult to access experimentally. Recent advances in reaction initiation methods and time-resolved, structure specific probes made in our laboratories[13−16] and in others[17−21] have made it possible for the first time to resolve some of the earliest, most fundamental steps along the folding pathway.

This account discusses the application of temperature-jump (T-jump) reaction initiation methods and time-resolved infrared (TRIR) spectroscopy techniques developed in our laboratories to the earliest events in protein folding. Of particular interest is the critical early step of secondary structure formation. We have used these methods to study the formation of helices in model peptides and in apoMb, a globular, helical protein.

II. Initiating and Probing Fast Folding Events

Initiation Problem. The primary impediment to studying the earliest events in protein folding has been the 1−10 ms dead time of conventional stopped-flow methods for initiating folding reactions. Our approach to the initiating folding/unfolding chemistry is the laser-induced T-jump technique, the most general and simplest of the fast initiation approaches. Proteins usually have a temperature range over which the functional structure is stable. Temperatures either above or below this range denature the structure. Thus, by judicious choice of temperature and solution conditions, one can cause the equilibrium between the native and denatured forms to be poised at any point, and an increase in temperature then perturbs the equilibrium toward a more folded or more unfolded position. This approach is applicable to virtually all proteins and peptides and does not necessarily require the introduction of extraneous reagents into the folding system.

The essential characteristic of the temperature jump is that it perturbs the equilibrium position of the ensemble

---

R. Brian Dyer received the Ph.D. degree in inorganic chemistry from Duke University in 1985. After spending 2 years in the Inorganic and Structural Chemistry Group (INC-4) at Los Alamos National Laboratory as a postdoc, he moved to the chemical and laser sciences division as a staff member. He is currently on staff in the biosciences and biotechnology group (CST-4). He was recently awarded the Los Alamos Fellows Prize for his work on molecular dynamics and protein folding.

Feng Gai received the Ph.D. degree in 1994 from Iowa State University. He spent 2 years at Harvard University as a postdoctoral associate before moving to Los Alamos National Laboratory as a LANL postdoctoral fellow.

William H. Woodruff received the M.S. (1969) and Ph.D. (1972) degrees from Purdue University and was awarded an NIH Postdoctoral Fellowship for study at Princeton University. He then served on the faculties of Syracuse University (1974−1977) and the University of Texas at Austin before moving to Los Alamos National Laboratory in 1984. He is currently a Laboratory Fellow in the Bioscience and Biotechnology Group (CST-4). In 1997 Woody was awarded the Bomem−Michelson Award by the Coblentz Society for advances in spectroscopy and the Elisabeth Roberts Cole Award by the Biophysical Society for advances in biophysics.

Rudolf Gilmanshin received the Ph.D. degree in biophysics from the Institute of Biophysics, Russian Academy of Science (RAN), in 1987. His postdoctoral training was at the Institutes of Biological Research Center, RAN, where he later was appointed a Group Leader in 1990. He worked as Research Associate at the University of Virginia (1992−1993) and at the City College of City University (1993−1997). Now he is Instructor of Biochemistry at the Albert Einstein College of Medicine, New York.

Robert H. Callender received the Ph.D. degree in applied physics from Harvard University in 1969. After a year of postdoctoral studies at the University of Paris, he joined the physics faculty at the City College of City University in 1970 and was appointed as a Distinguished Professor of Biophysics in 1989. He joined the faculty at the Albert Einstein College of Medicine, Bronx, New York, in 1996, where he is Professor of Biochemistry.

10.1021/ar970343a CCC: $15.00
Published on Web 07/30/1998

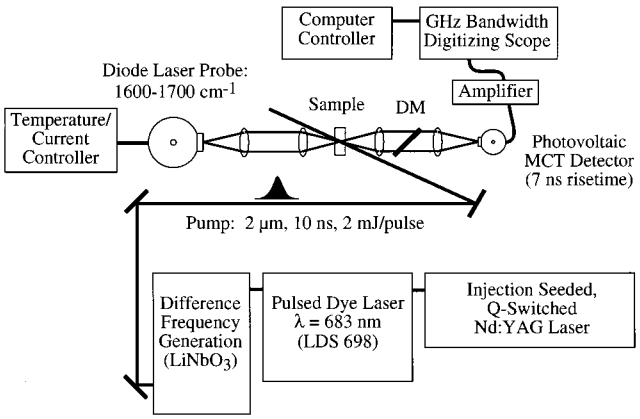

FIGURE 1. Nanosecond laser temperature-jump, time-resolved infrared apparatus.

of possible states faster than the chemical processes that comprise the equilibrium. This is the fundamental characteristic of relaxation methods.[22] After the T-jump, the system "relaxes" to the new equilibrium point. In general, the kinetics that are observed involve both the forward and back reactions. For simple two-state kinetics, the observed relaxation rate is the sum of the forward and back reaction rates, and each can be determined separately if the equilibrium constant is known.[23]

The use of fast T-jump methods to displace chemical equilibria and measure relaxation rates was pioneered by Eigen and De Mayer,[22] using rapid capacitance discharge to produce Joule heating in conducting solutions. This approach is limited to the microsecond time scale. More recent work has focused on rapid absorbance of large amounts of laser energy by the sample solution, generally using near-infrared solvent absorbances as first developed for studying conformational kinetics of inorganic complexes.[24] In experiments of this type, the T-jump and subsequent complete thermalization of solvent and solute can occur in 10−20 ps.[25]

Our approach for generating a rapid, laser-induced T-jump is shown schematically in Figure 1.[13−15] An injection-seeded Nd:YAG laser pumps a pulsed dye laser (694 nm) which is differenced with the Nd:YAG fundamental (1064 nm) in a nonlinear $LiNbO_3$ crystal to produce the T-jump pump pulse at 2 μm (10 ns full width at half-maximum (fwhm) pulse width). The wavelength of the T-jump pulse is chosen to transmit 80% of the light through a 100 μm path length cell, which ensures a nearly uniform temperature profile in the laser interaction volume. The laser energy is absorbed by the D_2O, and the temperature of the interaction volume reaches its maximum value in approximately 20 ns. The size of the T-jump is calibrated using the change of D_2O absorption with temperature, which acts as an internal thermometer in the range of 1600−1700 cm^{-1}. Temperature jumps of 20 °C are routinely obtained with this apparatus using 2 mJ of laser energy at 2 μm. The heat contained in the interaction volume gradually escapes to nearby cold surfaces, the rate governed by thermal diffusion. The diffusion of heat out of the interaction volume takes several milliseconds in our cells. Thus, the apparatus generates a temperature jump within 20 ns that remains nearly constant up to approximately 1 ms.

The short time resolution of this instrument is limited by both the laser pulse width and the detector rise time. We have used a separate apparatus to extend the earliest time accessible with a T-jump to 50 ps. The picosecond T-jump apparatus is based on a regeneratively amplified, mode-locked Nd:YAG laser, which is shifted to 2 μm (50 ps fwhm) by the stimulated Raman effect in H_2 gas. Time resolution is achieved by optical delay between the pump pulse and a probe pulse, which circumvents the problem of ultrafast (ps) IR detection. A long, multipass delay line allows us to scan out to 20 ns, overlapping the earliest time accessible to our nanosecond instrument. Other groups have used the Raman shift approach in different media to generate 1.4−1.5 μm pulses for pumping the analogous near-IR band in H_2O.[24,26,27]

Probing Fast Folding Dynamics with Structural Specificity. Structural characterization of early folding events requires spectroscopic tools with sufficient time resolution and structural sensitivity; for this reason, we emphasize vibrational spectroscopy. The advantages of time-resolved vibrational measurements as structure-specific probes of protein dynamics are well-established. The structural specificity derives from the connection of molecular vibrations to specific structures which determine the frequency, intensity, and line widths of the absorptions. Protein folding can involve changes in the polypeptide backbone conformation and H-bonding, in the orientation and packing of side chains, in the solvation of these structures, and so forth, all of which will be reflected in changes in the IR spectrum. Finally, the IR spectral changes track molecular dynamics down to ultrashort time scales because vibrational transitions respond to changes in structure or environment on time scales as short as the 10−100 fs periods of molecular vibrations.

IR absorption spectroscopy of the amide-I mode of peptides and proteins, largely the stretch motion of the backbone peptide C═O group with some contribution from the C−N stretch, has been shown to be strongly correlated to both secondary and tertiary structural features. We have found amide I exceptionally valuable in determining structural changes in recent fast folding studies by our group. This polar mode, which lies at 1610−1680 cm^{-1}, is an established indicator of secondary structural changes because of its sensitivity to hydrogen bonding, dipole−dipole interactions, and geometry of the peptide backbone.[28,29] For IR absorption studies of proteins in solution, D_2O is often used instead of H_2O in order to shift the strong water band, that normally overlaps with the amide I band, down to 1200 cm^{-1}. The mode for deuterated amide groups is labeled amide I′ by convention.

Previous correlations between the amide I′ band position and the secondary structure in proteins[30,31] have shown that bands centered at approximately 1638 and 1654 cm^{-1} correspond to 3_{10}- and α-helices, respectively,

1629 and 1675 cm^{-1} bands to the low- and high-frequency components of β-sheet structure, 1668, 1675, and 1686 cm^{-1} peaks to turns, and a component at 1645 cm^{-1} results from the disordered parts of polypeptide backbones. Recently it has been demonstrated that intensity near 1630−1640 cm^{-1} can result from α-helical structure exposed to solvent (the so-called "solvated" helix[32−34]). Hence, different portions of the amide I′ band report on different parts of the structure of a folded peptide or protein.

III. α-Helix Formation in Peptide Models

Characteristic Rate of Helix Formation. The α-helix is the most common secondary structural motif found in proteins. Thus, our understanding of the overall folding process in proteins will be aided by study of the dynamics of helix formation in peptide fragments or model peptides. Very little is known about this problem, either experimentally or theoretically, despite the relative structural simplicity of helices.

The thermodynamics and kinetics of the helix coil transition in homopolymers were studied intensely in the mid-1960s to early 1970s, both experimentally and theoretically (reviewed by Gruenewald).[35] The fastest relaxation times observed ranged from 50 ns to microseconds, depending on the system that was studied and the method. However, these experiments were difficult to interpret because the probes that were used (e.g. ultrasound) lacked structural specificity and because the helix formation rates were inferred indirectly from the data. In addition, the very long homopolymers (e.g. poly-(L-glutamic acid)) studied are poor models for the short runs of helical structure usually found in proteins.

Recently, we revisited this problem using our T-jump approach on short helical peptides. The first system we studied was the "S-peptide" of ribonuclease A, obtained by cleaving off the first 21 amino acid residues, which forms a helical structure in solution (about 30% helix and 70% unfolded at temperatures near 5 °C.[36] Using the amide I′ band to monitor helix content, we observed a relaxation time of faster than 300 ns for the S-peptide in response to a 20 °C T-jump.[14] The earliest version of our T-jump apparatus was unable to resolve the actual relaxation rate.

Following these initial results, we built the picosecond and nanosecond T-jump setups described above and characterized the folding of the synthetic "F_s peptide", a small 21-residue peptide that has the sequence X-$(A)_5$-$(AAARA)_3$-ANH_2 (X = succinyl). Alanine has a strong helix propensity, and about 70−90% of this peptide is helical in room temperature solution.[37] Guided by the S-peptide result, we first examined the relaxation kinetics using the picosecond apparatus (Figure 2A). We found no response of the amide backbone (detectable in the amide I′ absorbance) to the T-jump from 50 ps out to about 10 ns. As we soon discovered with our faster ns apparatus, we had bracketed the actual response (∼160 ns, blue trace in Figure 2A) in these two experiments. The relaxation kinetics are well-described by a biexponential model with components of 10 ns (small amplitude) and 160 ns (dominant amplitude) for a T-jump from 8 to 27 °C.[15] More recently, a single, nearly temperature independent relaxation time of about 20 ns was observed for the F_s peptide using a laser-induced T-jump and fluorescence spectroscopy to probe the folding dynamics.[27] The latter study used F_s peptide labeled at the N-terminus with the fluorescent "reporter" 4-(methylamino)benzoic acid (MABA). The fluorescence of the MABA probe appears to be sensitive to only the last turn of helix, in contrast with the IR sensitivity to the helical content of the entire peptide.

These results are best understood in terms of statistical models for helix formation developed by several groups.[27,36] These models assume that the activation barrier for helix formation is largely entropic. Two elementary processes, nucleation and propagation, are supposed, a distinction being made between the probability of forming the first turn of helix (nucleation, a relatively less probable event) and the formation of subsequent turns (propagation). In this view, the dominant rate observed in the T-jump measurements would be the propagation rate, since the measurements are made under conditions where most of the peptide molecules are always partially folded. Calculation from this model of the relaxation kinetics induced by a T-jump yielded two distinct processes with characteristic rates; a faster process involving the rapid folding/unfolding equilibrium of the helix ends and a slower process involving an equilibrium which must cross the nucleation free energy barrier between helix-containing and nonhelix-containing structures. The relaxation of the equilibrium determining the fraction of N-terminal residues that are helical was predicted to contain approximately equal contributions from both processes, while that determining the average helix content was found to be dominated by the slower process. The model thus successfully predicts the differences in rates between the IR and fluorescence time constants and also, if the barrier to helix formation is taken as purely entropic, the weak temperature dependence of the relaxation rates. The approximations of this model and a thorough investigation of the dynamics of helix formation can be explored by isotopic labeling of individual turns in the middle and ends of the peptide, to determine the rates of formation. These experiments are underway in our laboratories.

Peptide Model of Helix−Helix Interactions. The next level of complexity in proteins involves tertiary interactions, such as the packing of helices together in a globular protein. We have studied helix−helix interactions in a peptide fragment of the GCN4 leucine zipper protein. The GCN4-p1 peptide (NH_2−RMKQLEDKVEELLSKNYHLENEVARLKKLVGER−COOH) forms a well-characterized, dimeric coiled-coil (a leucine zipper). The structure of the leucine zipper protein has been solved by X-ray crystallography[38] (Figure 2A), and the folding has been studied in solution by CD and NMR spectroscopies.[39,40]

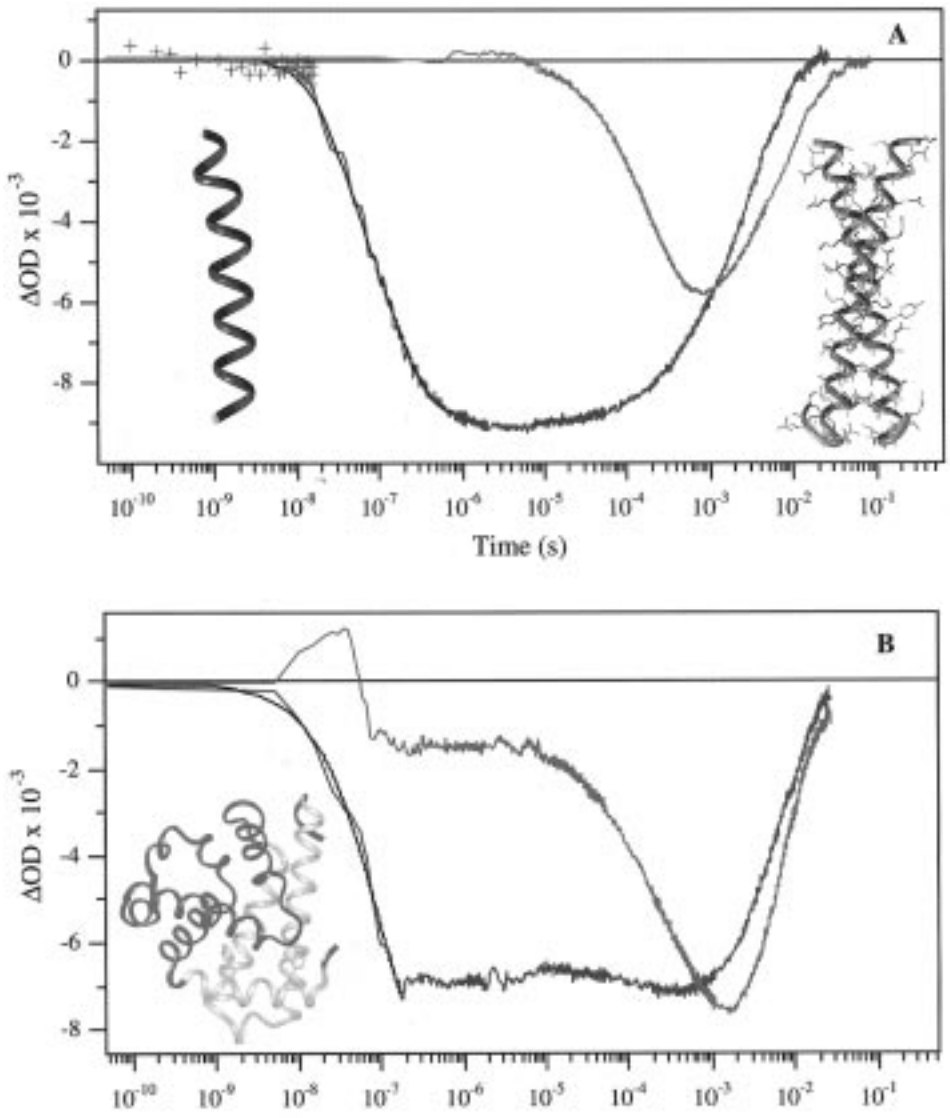

FIGURE 2. (A) Time-resolved IR relaxation kinetics following a laser-induced T-jump for two peptide model systems. Blue symbols (+) and blue trace: picosecond and nanosecond responses, respectively, of the F_s peptide to a T-jump from 8 to 27 °C, monitored at the peak of the helix absorbance (1632 cm^{-1}). The blue structure is a ribbon trace of the backbone of a hypothetical, fully α-helical conformation of the peptide. Green trace: the response of the GCN4-p1 coiled-coil to a T-jump from 51 to 60 °C monitored at the peak of the helix absorbance at 1648 cm^{-1}. The structure is from the crystal structure of the GCN4 dimer,[38] representing the peptide fragment used in this study. (B) Time-resolved IR of the native state of apoMb (4.2 mg/mL, 0.01 M NaCl, pH* = 5.3) for a T-jump from 45 to 60 °C. Blue trace: IR response monitored at the peak of the solvated helix absorbance (1632 cm^{-1}). Green trace: IR response monitored near the peak of the native helix absorbance (1655 cm^{-1}). The structure is from the crystal structure of the holoprotein.

We have characterized the structure and dynamics of the GCN4-p1 peptide using infrared spectroscopy. We observe the unique infrared signature of a coiled-coil that has also been reported for other systems.[41] At low temperature, the amide I spectrum of the folded coiled coil contains a major component at ~1650 cm^{-1}, which is typical of α-helix desolvated by helix−helix interactions in a globular protein (in contrast with the monomeric helical peptide models, which show only a 1630 cm^{-1} solvated helix band). Raising the temperature causes the 1650 cm^{-1} component to diminish with the concomitant appearance of a peak at ~1670 cm^{-1}, which we interpret as the unfolding of the coiled-coil to produce a disordered structure.

TRIR measurements of the relaxation dynamics of GCN4-p1 (green trace in Figure 2A) are strikingly different from the relaxation dynamics of the monomeric, helical F_s peptide. The observed relaxation kinetics probed at 1648 cm^{-1} are 3 orders of magnitude slower than the F_s peptide kinetics. Folding and unfolding rates extracted from the relaxation rate using a two-state analysis also differ from their F_s peptide counterparts by over 3 orders of magnitude. It is highly unlikely that the intrinsic helix formation and melting rates in these two systems differ by this extent. We attribute the differences to the formation of tertiary structure in the coiled-coil. The rate measured in the monomeric peptide models is the intrinsic rate of the helix−coil transition, whereas the rate

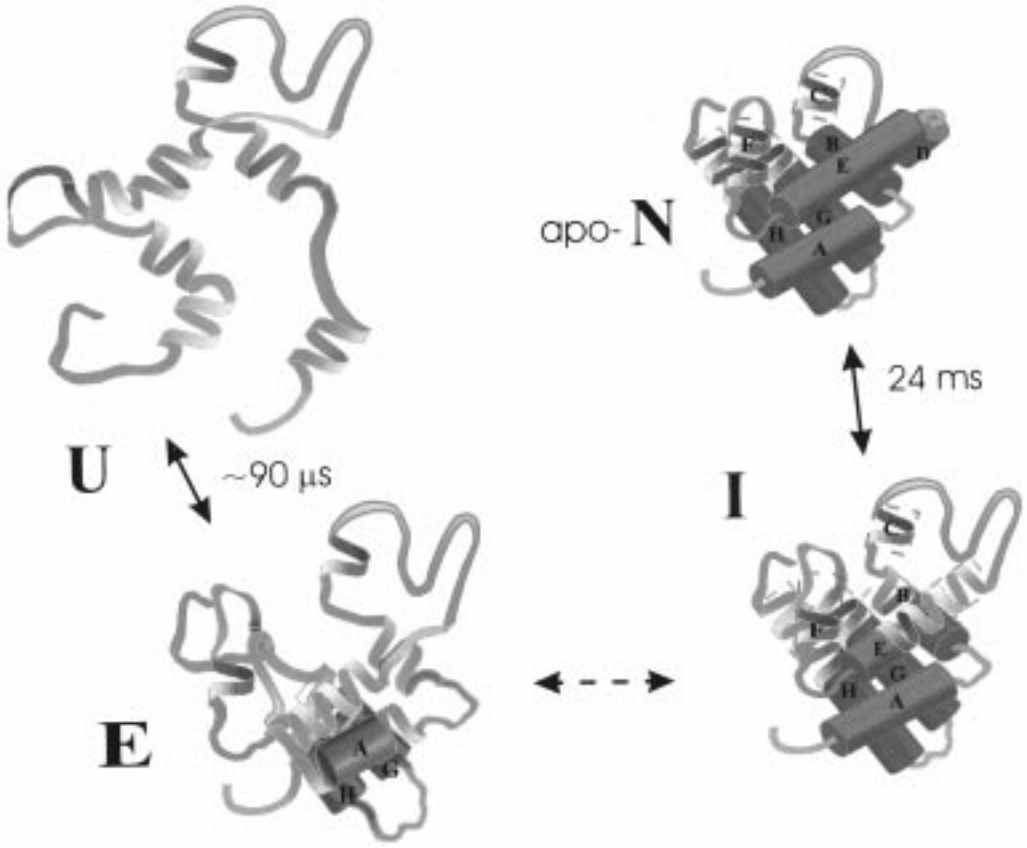

FIGURE 3. Schematic representation of the folding of apomyoglobin starting from an unfolded (U) state. The folding time of 90 μs for the U to E transition is that measured at 35 °C from T-jump relaxation measurements of the E state (see text), which involves the formation of a tightly packed AGH core. In addition, nanosecond transients are also observed, which arise from the relaxation of the solvated helical portion of the E state. The folding times under native chemical conditions are likely faster due to increased driving force. The folding time of 24 ms for the I to N transition, which involves formation of nativelike contacts of the B and E helices (and possibly others), has been extrapolated to 35 °C from the measured rates at 60 and 5 °C (see text) using an Arrhenius (log) scale. Like the E state, T-jump measurements of the N and I forms of the protein also show relaxation dynamics of their solvated helical portions on the nanosecond time scale.

measured for the GCN4-p1 peptide is clearly limited by the much slower packing of two helices together to form the leucine zipper and the coiled-coil.

IV. Apomyoglobin

The folding of apomyoglobin has received intense scrutiny. The protein is relatively simple, and it serves as an archetype for folding of proteins which are small, single-domain, and globular. The holoprotein contains 153 residues organized into eight strands of mostly α-helical segments, labeled A−H (Figure 3).[42] Near neutral pH, apoMb adopts a structure that is nativelike according to the available NMR, CD, and calorimetric evidence.[43−46] Thus "native" apoMb (ApoMb-N) has a compact hydrophobic core consisting of the very stable A, G, and H helices, significant tertiary structure involving the B and E helices, and roughly the same secondary structure content and tertiary fold as the holoprotein.

Another attractive feature of apoMb is the accessibility of different equilibrium forms of the protein having intermediate stability between the native and unfolded states.[43,47] ApoMb-I (intermediate form), formed at low pH and specific salt conditions, has a less compact structure than apoMb-N. NMR techniques coupled with hydrogen−deuteron exchange measurements and more recent mutant and FTIR studies show that the I form contains a tight core consisting of the intersection of the AGH helices and loose "solvated" helical region(s).[48,49] A species formed at still lower pH and low salt, apoMb-E (extended form), contains essentially only the AGH core and is otherwise very extended.[47] Furthermore, recent evidence suggests that these equilibrium intermediate states are structurally close to transient intermediates on the folding pathway.[43] A striking structural similarity between the I form and an early folding intermediate has been demonstrated.[50,51]

We have obtained the thermal denaturation curves of apoMb-N using both fluorescence and infrared spectroscopies.[16,47] A cooperative transition is observed both in the fluorescence intensity at 335 nm, I_{335}, ($T_m = 56 \pm 3$ °C) and by IR absorbance at 1650 cm^{-1} ($T_m = 59 \pm 3$ °C). The beginning of the cold denaturation transition also appears in the I_{335} dependence below 10 °C. Neither of these two transitions shows up on the fluorescence λ_{max} denaturation curve, indicating that the AGH core remains intact through this transition. A second component of the amide I′ band at 1629 cm^{-1} decreases monotonically with

increasing temperature and is assigned to the unfolding of solvated helices.

The environments and the conformations of the various substructures lead to different amide I′ frequencies that allow these substructures to be distinguished uniquely by the infrared spectra, a major advantage of this probe; native helix absorbs near 1650 cm^{-1}, solvated helix near 1630 cm^{-1}, and the unfolded structures above 1660 cm^{-1}. Using these infrared absorbances, the relaxation kinetics of the substructures (as well as the equilibrium melting behavior) can be ascertained independently of one another. The relaxation kinetics of ApoMb-N in response to a T-jump are shown in Figure 2B. The solvated helical substructure (blue trace, Figure 2B) exhibits a relaxation time (ca. 50 ns) typical of isolated helices such as the model peptide described above. This is the case even though these helices may be in contact with each other, held together by loose hydrophobic forces. Thus, the α-helical secondary structures can be formed early on the folding pathway of apoMb, on the time scale of tens of nanoseconds.[16] In contrast, the relaxation time observed for the native helices of ApoMb-N (monitored at 1650 cm^{-1}) is some 3 orders of magnitude slower (Figure 2B, green trace), analogous to the relaxation rate of the coiled-coil structure of GCN4-p1 (Figure 2A, green trace). This slower relaxation is not limited by the intrinsic helix formation rate but rather reflects the slower rates of the melting and formation of helix−helix interactions.

We have used the approach outlined above to study the dynamics of the transition from more folded to less folded ensembles in successively more denatured (N, I, and E) apoMb states. In this way, the folding pathway can be "peeled away" and processes on both fast and slower time scales determined. Of course, the chemical conditions that give rise to the I and E forms are not those under which the N state is formed. Nevertheless, as noted above, equilibrium intermediate states have been identified with transient intermediates on the folding pathway under conditions where the N form is stable. Figure 3 shows the folding pathway in schematic form, assuming that the I and E states are transient intermediates under native conditions.

It is known that the "hydrophobic core" of apoMb, comprising the A helix and the G−H hairpin, is mostly α-helical and is the most stable part of the protein.[52] The formation time of the AGH core measured by IR is about 90 μs in the E state at 35 °C[53] and perhaps as fast as 5 μs in the N state at 10 °C.[54] The latter measurement was made using fluorescence and T-jump from a cold denatured state. It is difficult to compare these two rates directly due to the different chemical conditions and temperatures. The starting and ending states have substantially different structures in the two experiments. The folded E-state of apoMb is highly destabilized, containing essentially only the AGH core in contrast to the nativelike structure near neutral pH, while the heat denaturation and cold denaturation transitions are inherently different, yielding different denatured states (for example the cold denatured N-state retains substantial secondary structure, whereas the thermal denatured E-state does not). Nonetheless, it is clear that the formation time of the AGH core is very fast, on the order of what is expected from a simple polymer-diffusion model,[53,55,56] considering only the time it would take for the two ends of the molecule to come together by diffusion given the size of the loop intervening between the A and G helices. However, the short time scale of this tertiary folding event is not unreasonable if it is realized that the protein has had ample time to form helical segments in the intervening B−F structure (which are likely not set structures but rather interconverting helical runs of varying length and position) before the AGH core is formed, which shortens the effective loop length. The fast formation of intervening secondary structure may facilitate the formation of the AGH tertiary contacts in ways that are directed and faster than random diffusion control.

The AGH intersection at the point of its initial formation may only involve about 20 residues, but it is nevertheless a very stable structure with a large negative enthalpy of formation, as shown by its cooperative, sigmoidal melting curve. The fast formation of a such a stable nativelike structure would seem to have several purposes in the folding strategy of the protein. The speed may ensure that other nonnative but metastable folds do not form. Tying the two ends together reduces enormously the amount of conformational space that may be searched in subsequent folding. The two ends of the protein (at least part of the A and G−H structures) must be preformed into helices before the correct tertiary contacts can form; this presents no problem in timing since the helices form on a time scale 2−3 orders of magnitude faster than the tertiary contacts.[15,16]

Following the formation of the AGH core, the next step in the process is the rapid growth of the existing solvated helical portions of the B−E loop, which has the effect of compacting the protein further. Also, it is likely that the tertiary contacts in the A helix and the G−H hairpin grow out from the AGH intersection in relatively short times following the formation of the intersection. These processes form a structure similar to the I form, as shown in Figure 3. These two events, formation of the core and then its growth, could be termed "nucleation" followed by growth.[7] Finally, the B and E helices, and possibly parts of others, come together to form a second native fold. The timing of this event is much slower than that for the association of the AGH core and quite temperature sensitive, 280 μs at 60 °C[16] and ca. 1 s at 5 °C,[50] indicating a substantial formation energy barrier. This model of a sequence of events U → E → I → N that leads to progressively more compact and helical structures is consistent with a recent NMR study which gives an exquisitely detailed picture of the structure and dynamics of the U, I, and N states.[57]

The energy surface for folding of apomyoglobin appears to be such that, very early, the protein condenses into a compact topology with a very small cooperative unit (the AGH core), found in the totally folded protein, which remains during the rest of the folding process and which

perhaps serves as a template. It seems likely that the presence of secondary structure is required before the nucleation of this tertiary structure can take place, and the required helical structure has, indeed, already formed. The next stage is further growth of the ordered structure surrounding the small AGH core, in both the number of residues in helices and an extension of tertiary contacts spreading out from the small core volume. Perhaps surprisingly for a small protein like apomyoglobin, a second kinetics phase whereby other residues condense into a native fold occurs much later, on relatively slow time scales. Hence, 2-folding units describe the folding of this protein. From this degree of detail, it is fairly easy to see how the early structures, even if local and fluctuating in nature, result in the required constriction in conformational space that accelerates and directs the folding process.

The authors are grateful for the support of this work by the NIH (R.B.D.) and the NSF (R.H.C.).

References

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223−30.
(2) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631−60.
(3) Bai, Y.; Englander, S. W. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 145−51.
(4) Miranker, A. D.; Dobson, C. M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 31−42.
(5) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−9.
(6) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10−4.
(7) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.
(8) Roder, H.; Colon, W. *Curr. Opin. Struct. Biol.* **1997**, *7*, 15−28.
(9) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−51.
(10) Murphy, K. P. In *Protein Stability and Folding: Theory and Practice*; Shirley, B. A., Ed.; Humana Inc.: Totowa, NJ, 1995; Vol. 40, pp 1−34.
(11) Friesner, R. A.; Gunn, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 315−42.
(12) McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11426−7.
(13) Callender, R.; Gilmanshin, R.; Dyer, R. B.; Woodruff, W. *Phys. World* **1994** (Aug), 41−5.
(14) Woodruff, W. H.; Dyer, R. B.; Callender, R. H.; Paige, K.; Causgrove, T. *Biophys. J.* **1994**, *66*, A397.
(15) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691−7.
(16) Gilmanshin, R.; Williams, S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3709−13.
(17) Phillips, C. M.; Mizutani, Y.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7292−6.
(18) Jones, C. M.; Henry, E. R.; Hu, Y.; Chan, C. K.; Luck, S. D.; Bhuyan, A.; Roder, H.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11860−4.
(19) Pasher, T.; Chesick, J. P.; Winkler, J. R.; Gray, H. B. *Science* **1996**, *271*, 1558−60.
(20) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5759−64.
(21) Chen, E.; Lapko, V. N.; Song, P. S.; Kliger, D. S. *Biochemistry* **1997**, *36*, 4903−8.
(22) Eigen, M.; De Maeyer, L. D. In *Technique Organic Chemistry*; Friess, S. L., Lewis, E. S. Weissberger, A., Eds.; Interscience: New York, 1963; Vol. 8, pp 895−1054.
(23) Schmid, F. X. In *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman and Co.: New York, 1992; pp 197−241.
(24) Beitz, J. V.; Flynn, G. W.; Turner, D. H.; Sutin, N. *J. Am. Chem. Soc.* **1970**, *92*, 4130−33.
(25) Richard, L.; Genberg, L.; Deak, J.; Chiu, H.-L.; Miller, R. J. D. *Biochemistry* **1992**, *31*, 10703−15.
(26) Ballew, R. M.; Sabelko, J.; Reiner, C.; Gruebele, M. *Rev. Sci. Instrum.* **1996**, *67*, 3694−9.
(27) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200−10.
(28) Susi, H.; Byler, D. M. *Methods Enzymol.* **1986**, *130*, 290−311.
(29) Arrondo, J. L. R.; Muga, A.; Castresana, J.; Goni, F. M. *Prog. Biophys. Mol. Biol.* **1993**, *59*, 23−56.
(30) Prestrelski, S. J.; Byler, D. M.; Thompson, M. P. *Int. J. Pept. Protein Res.* **1991**, *37*, 508−12.
(31) Prestrelski, S. J.; Byler, D. M.; Thompson, M. P. *Biochemistry* **1991**, *30*, 8797−804.
(32) Haris, P. I.; Chapman, D. *Biopolymers* **1995**, *37*, 251−63.
(33) Martinez, G.; Millhauser, G. *J. Struct. Biol.* **1995**, *114*, 23−7.
(34) Gilmanshin, R.; Van Beek, J.; Callender, R. *J. Phys. Chem.* **1996**, *100*, 16754−60.
(35) Gruenewald, B.; Nicola, C. U.; Lustig, A.; Schwarz, G.; Klump, H. *Biophys. Chem.* **1979**, *9*, 137−47.
(36) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95−118.
(37) Lockhart, D. J.; Kim, P. S. *Science* **1993**, *260*, 198−202.
(38) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.
(39) Harbury, P.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401−6.
(40) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538.
(41) Heimburg, T.; Schuenemann, J.; Weer, K.; Geisler, N. *Biochemistry* **1996**, *35*, 1375−82.
(42) Evans, S. V.; Brayer, G. D. *J. Biol. Chem.* **1988**, *263*, 4263−8.
(43) Barrick, D.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 3790−6.
(44) Cocco, M. J.; Lecomte, J. T. J. *Protein Sci.* **1994**, *3*, 267−81.
(45) Johnson, R. S.; Walsh, K. A. *Protein Sci.* **1994**, *3*, 2411−8.
(46) Lin, L.; Pinker, R. J.; Forde, K.; Rose, G. D.; Kallenbach, N. R. *Nat. Struct. Biol.* **1994**, *1*, 447−52.
(47) Gilmanshin, R.; Dyer, R. B.; Callender, R. H. *Protein Sci.* **1997**, *6*, 2134−42.
(48) Kay, M. S.; Baldwin, R. L. *Nat. Struct. Biol.* **1996**, *3*, 439−45.
(49) Gilmanshin, R.; Williams, S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1997**, *36*, 15006−12.
(50) Jennings, P. A.; Wright, P. E. *Science* **1993**, *262*, 892−6.
(51) Eliezer, D.; Jennings, P. A.; Wright, P. E.; Doniach, S.; Hodgson, K. O.; Tsuruta, H. *Science* **1995**, *270*, 487−8.
(52) Shin, H. C.; Merutka, G.; Waltho, J. P.; Tennant, L. L.; Dyson, H. J.; Wright, P. E. *Biochemistry* **1993**, *32*, 6356−64.

(53) Gilmanshin, R.; Callender, R. H.; Dyer, R. B. *Nat. Struct. Biol.* **1998**, *5*, 363−5.
(54) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Nat. Struct. Biol.* **1996**, *3*, 923−6.
(55) Pastor, R. W.; Zwanzig, R.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 3878−82.
(56) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615−7.
(57) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **1998**, *5*, 148−155.

AR970343A