

FAST EVENTS IN PROTEIN FOLDING: The Time Evolution of Primary Processes

Robert H. Callender,¹ R. Brian Dyer,² Rudolf Gilmanshin,¹ and William H. Woodruff²

¹Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461; ²Los Alamos National Laboratory, Los Alamos, New Mexico 87545; e-mail: call@aecom.yu.edu, bdyer@lanl.gov, rgilman@aecom.yu.edu, woody@lanl.gov

KEY WORDS: temperature jump, early folding intermediates, kinetics, denaturation, infrared spectroscopy

ABSTRACT

Most experimental studies on the dynamics of protein folding have been confined to timescales of 1 ms and longer. Yet it is obvious that many phenomena that are obligatory elements of the folding process occur on much faster timescales. For example, it is also now clear that the formation of secondary and tertiary structures can occur on nanosecond and microsecond times, respectively. Although fast events are essential to, and sometimes dominate, the overall folding process, with a few exceptions their experimental study has become possible only recently with the development of appropriate techniques. This review discusses new approaches that are capable of initiating and monitoring the fast events in protein folding with temporal resolution down to picoseconds. The first important results from those techniques, which have been obtained for the folding of some globular proteins and polypeptide models, are also discussed.

INTRODUCTION

Understanding how proteins fold up into their compact three-dimensional forms is a central problem in modern structural biology. This is so because the particular structure of a protein governs its specific activity, and any biological function is supported by a corresponding protein system. It has been known for some

time that the spatial structure of a protein molecule is determined completely by the sequence of amino acids of its polypeptide chain, at least for small proteins and probably for all proteins in a general sense (1–3). Moreover, the amino acid sequence also codes the way the three-dimensional structure is reached efficiently (4). Both aspects are equally important for protein engineering purposes, because any polypeptide sequence coding a new protein must not only offer some new function but also ensure their efficient folding. Otherwise, the expressed protein might never accumulate in large quantities because slow folding processes may permit the accumulation of long-lived metastable non-functional folds. Besides, aggregation at intermediate stages or digestion by the proteinases of the host will be high for long-lived unfolded structures.

The tools of diffraction techniques and multidimensional nuclear magnetic resonance (NMR) have determined the spatial structures of many proteins, out of which perhaps about a thousand are distinctly different from one another. This achievement has facilitated the study of the molecular basis of protein function, including issues of specificity, primary stages of catalysis, and distribution of the functions within supramolecular complexes. The level of knowledge has permitted useful modifications of known proteins as well as limited ideas of protein design *ab initio*, through various semiempirical approaches (5–9). However, there is uncertainty whether such predicted structures can exist as stable folded proteins at all. One reason is that it is unknown if the final structure corresponds to the global minimum of the free energy. This is a difficult problem to assess experimentally, and there are examples where successful protein function derives from a metastable state (10). Moreover, the calculation of the energy of a polypeptide in a configuration space is a complex and very difficult physical problem because the Gibbs free energy of the folded protein is actually a small difference of large enthalpy and entropy components. At present, the components cannot be calculated with sufficient accuracy both because of the complexity of the protein molecule itself and because of its interaction with surrounding water (3, 11, 12).

It is clear that the theoretical analysis of the folding process is complex. In addition to the reasons given above, it seems necessary to deal with statistical ensembles of the unfolded state as well as ensembles of transient intermediates in the early folding stages instead of a single, stable, determined structure. There has been some progress recently in understanding the general nature of the energy landscape of folding (9, 13–18), whereby the large amount of entropy of the unfolded protein is traded for enthalpy via a folding funnel. Another approach to the folding problem is to establish empirical rules that relate known protein structures to sequence, exploiting the information contained in the protein data bank (e.g. 5, 19). Here too the situation is extremely complex because the key interactions that determine the folding pathway involve both

local as well as nonlocal interactions and the folding pathway itself is a directed process whose rules may be quite buried in the final structure.

For several decades there has been much effort toward determining the kinetics of how a protein folds from its extended unfolded conformation to its final compact form. Investigators have sought to answer the questions of what types of structures form first and on what timescales (2, 20–28). Of most importance, perhaps, are the early kinetics events (25–29). These events set up all that are to follow in the guided pathway. However, the early kinetics of the folding pathway occur generally on the submillisecond timescale, and this has been a difficult timescale to access experimentally since the typical approach, stopped-flow devices whereby two reactants are mixed and the chemical reaction launched, have been limited to a resolution of 1–10 ms.

This review discusses new experimental techniques that initiate chemical kinetics and follow the structural changes that accompany these kinetics on submillisecond timescales. Also discussed is their recent application in the study of the folding of specific proteins and model systems. Substantial technical progress has been made over the past five years or so in the study of chemical reactions, and the folding process in particular, on fast timescales. This progress has resulted in new and exciting insights into the process. This survey does not review the protein folding field, which is well discussed in several recent reviews (16, 24–28, 30). In addition, it does not review NMR relaxation spectroscopy, which also has promise of determining fast folding processes (see e.g. 31). In understanding what follows, it is well to keep in mind that the folding reaction is not a simple chemical kinetics problem with a single reactant and a single product. The unfolded state is generally thought to be vastly inhomogeneous, composed of an enormous number of different conformations of the polypeptide. In fact, it is not easy to achieve the unfolded state in the laboratory or to fully characterize it (32). Many studies of the folded/unfolded transition have involved an unfolded state that still contains substantial secondary and even tertiary structure. Also, the pathway from unfolded to folded may differ even for similar unfolded states, and there may be multiple transition states. In counterpoint, sound theoretical arguments can be made that similar unfolded structures should behave similarly and that even very inhomogeneous systems should follow relatively simple apparent kinetics (33). In either case, some events in the process must be of more importance than others, and there certainly must be a time ordering of specific events in any guided process.

TECHNICAL DEVELOPMENTS

The study of rapid chemical kinetics requires both the techniques that can initiate chemical reactions faster than milliseconds and the methods for determining

structures on such fast timescales (24, 34–38). Most of the methods for determining structure “on the fly” are optical spectroscopies: circular dichroism (CD), optical absorption, fluorescence, and infrared and Raman spectroscopies. Application of these probes to folding often presents unique problems. In folding applications, CD and fluorescence have been limited to nanosecond resolution (39–41), but the other spectroscopic methods can, in principle, follow events down to femtoseconds. In their present form, all of these probes determine structure incompletely. Nonetheless, the different approaches yield information that is generally complementary, not only mutually but with other slower techniques such as dynamic NMR or pulsed deuterium exchange (42). We first discuss the spectroscopic probes of protein structure and then discuss techniques for fast initiation of folding reactivity.

Probes

FLUORESCENCE As monitors of protein structure, both intrinsic and extrinsic fluorophores have been employed (43,44). Among the former type, the tryptophan residue (Trp) is the most convenient. It has high quantum yield and a relatively low occurrence rate in sequences. The latter characteristic often simplifies data interpretation. For example, in recent studies on apomyoglobin (from horse heart), discussed below, there are only two Trp fluorophores, and both are located at the intersection of the A, G, and H helices (AGH core), which constitute the hydrophobic core of this protein. Hence, the emission signals report specifically on this important piece of tertiary structure. In the absence of tryptophan, tyrosine fluorescence can be used successfully. Extrinsic fluorophores can be either covalently attached labels (dansyl, fluorescein, etc) or probes interacting with proteins noncovalently. Among the latter, 1-anilino-naphthalene-8-sulfonate (ANS) is very useful in protein folding studies (45). This probe can bind only to disordered hydrophobic clusters and binds neither to natively tightly packed core nor to disordered polypeptide. The fluorescence quantum yield in the bound form of this molecule is increased orders of magnitude.

Different fluorescence parameters provide different information on protein structure (44,46). The position of the fluorescence spectrum (λ_{\max}) depends on the polarity of the surroundings of the fluorophore. This parameter is important when intrinsic fluorophores are used because the extrinsic labels are usually bound to the surface groups to provide minimal disturbance of the protein structure. Screening of a tryptophan residue from solvent by the protein structure results in a blue-shifted fluorescence. Protein unfolding thus red-shifts the spectrum toward the position of the fluorescence spectrum of a solvated tryptophan residue. Therefore, the fluorescence λ_{\max} is most dependent on large-scale changes of protein structure that permit tryptophan solvation.

The fluorescence quantum yield or integrated emission intensity, in contrast, is a balance of many different quenching factors. These can be primary (e.g. resulting from a temperature change) and secondary (resulting from a protein structure transformation). The latter can be specific (e.g. resulting from the spatial proximity of a tryptophan residue to specific protein side chains within its native structure) or nonspecific (e.g. resulting from a screening from a dissolved quencher). Hence, fluorescence intensity is sensitive to *both* global and local changes in protein conformation. Therefore, when modifying the configuration of a fluorescence experiment, one can emphasize either of the two scales of the changes (47,48). If some extrinsic quencher, such as acrylamide, is added to the solvent used for the folding experiment, the collapse of a protein molecule provides screening of the intrinsic fluorophore from extrinsic quencher, and the fluorescence intensity increases with the folding. Sometimes, the quenching group (e.g. heme) is intrinsic to the protein structure, which can place the group closer to a Trp during folding, thereby quenching its fluorescence. In both cases, the change of fluorescence intensity is presumed to report the global event (collapse). In another approach, the specific quenching group that contacts a particular Trp in the native protein can be replaced with a nonquenching one by site mutation techniques. Then, studying the folding pathway by a comparison of fluorescence of mutant proteins and the wild-type protein can provide precise local structural information.

INFRARED ABSORPTION Vibrational spectra are directly related to structural features of molecules in general: the strengths of bonds, the masses bonded together, the spatial distribution of the masses, dipole moments, polarizability, and so on. A chief advantage of vibrational spectroscopy is that the motions that make up the observed spectral features are generally localized on a specific molecular moiety (local mode approximation). In addition, the frequencies of molecular vibrations are determined by the electronic nature of bonds and are sensitive to molecular structure and environmental factors that influence this structure: hydrogen bond formation, hydrophobicity, bond angle and bond length deformation, bond formation, and other factors. Numerical accuracy relating vibrational frequencies to key bond properties can be very high. For example, empirical correlations between bond vibrational frequencies have been used in some cases to determine bond lengths to within $\pm 0.005 \text{ \AA}$, bond orders to within ± 0.04 valence units, and hydrogen bonding enthalpies to within $\pm 0.5 \text{ kcal/mol}$ (49,50).

Infrared absorption spectroscopy of the amide-I mode of peptides and proteins, largely the stretching motion of the backbone peptide C=O group with some contribution from the C-N stretch, is a structurally useful probe. This polar mode, which lies at $1610\text{--}1680 \text{ cm}^{-1}$, is an established indicator of secondary

structure changes because of its sensitivity to hydrogen bonding, dipole-dipole interactions, and geometry of the peptide backbone (51, 52). For IR absorption studies of proteins in solution, D₂O is often used instead of H₂O to shift the strong water band, which normally overlaps with the amide-I band, down to 1200 cm⁻¹. Deuteration of the amide group shifts the amide-I vibrations by a small amount and changes slightly the nature of the mode. Therefore, a distinction is made between protonated and deuterated amide-I modes. The standard nomenclature is amide-I', where the prime indicates the deuterated amide group. Previous correlations between band position and secondary structure (53, 54) have shown that bands centered at approximately 1638 and 1654 cm⁻¹ correspond to ₃10- 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 to the disordered parts of polypeptide backbones. Recently it has been demonstrated that intensity near 1625–1635 cm⁻¹ can result from α -helical structure exposed to solvent [the so-called solvated helix (55–58 and references therein)]. Hence, different portions of the amide-I' band report on different parts of the structure of a folded peptide or protein. In principle, great structural specificity can be obtained from isotope editing of the vibrational spectrum. For example, the amide-I' bands shift about 40 cm⁻¹ for ¹³C-labeled C=O. Site-specific ¹³C labeling of the polypeptide backbone should yield very detailed pictures of the dynamics of peptides and of proteins as protein-labeling methods are developed.

Surprisingly, the marker band position for an unfolded protein is not yet clear. The position of the amide I component as a result of irregular structure within the native protein (51) as well as the spectral maximum of denatured proteins (59) is generally found to be around 1645–1649 cm⁻¹. Spectra of the long disordered polypeptides exhibit two components: a strong one at 1646–1650 cm⁻¹ and a weak one at 1669–1675 cm⁻¹ (60). However, small peptides that are known to have a random structure and studies that have followed the melting of proteins suggest that the unfolded protein may be characterized by a broad marker band centered near 1665 cm⁻¹ (57, 59, 61). This may be because it is very difficult to unfold a protein completely, even under extremes of denaturing conditions. In addition, we note that the vibrational frequencies of irregular structures inside proteins are almost certainly different from solvent-exposed nonrigid structures of unfolded proteins.

As an example of an application, the thermal denaturation curves of native apomyoglobin are presented in Figure 1 (61, 62). Cooperative melting is seen on both the temperature dependencies monitored by fluorescence intensity at 335 nm, I₃₃₅, (T_m = 56 ± 3°C) and by IR absorbance at 1650 cm⁻¹ (T_m = 59 ± 3°C). The beginning of cold denaturation is also visible on the I₃₃₅ dependence below 10°C. Neither of these two transitions appears on the λ_{\max} denaturation curve because the Trp residues in the AGH core are not becoming

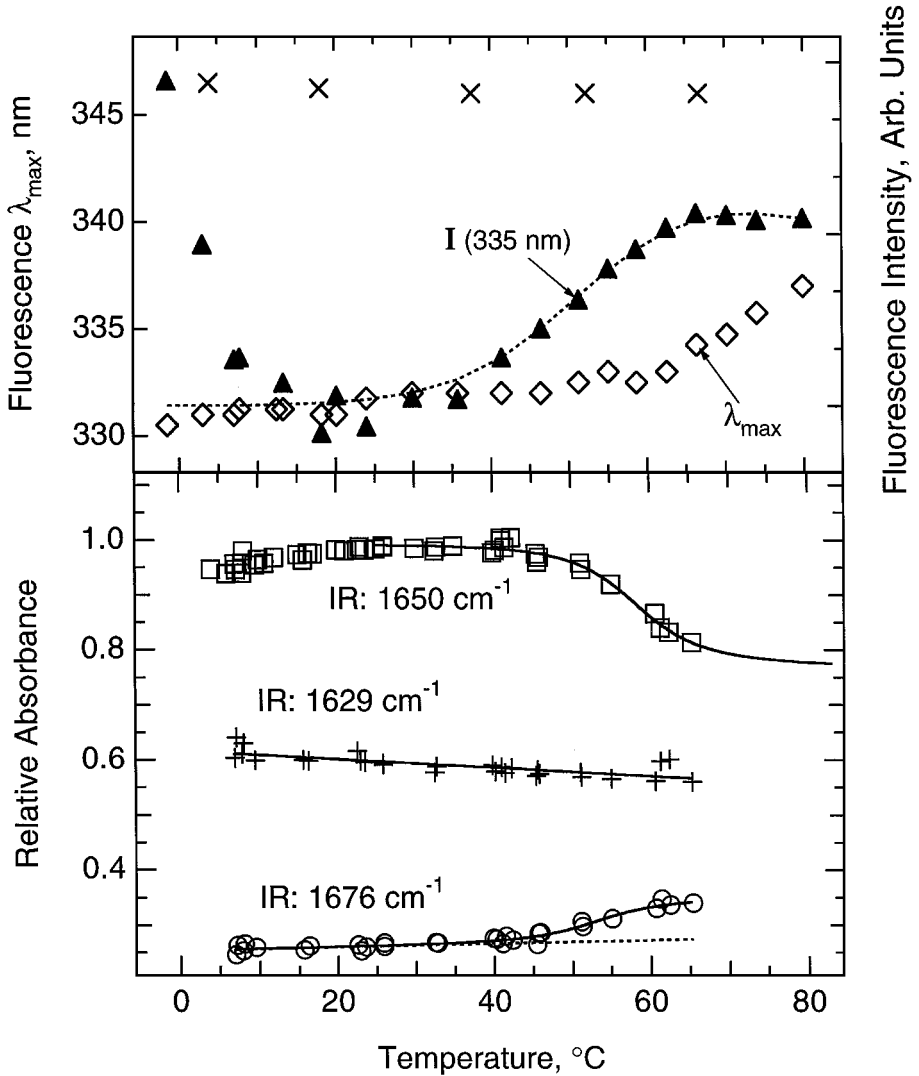


Figure 1 Thermal denaturation curves of native apomyoglobin, monitored by maximum position (λ_{\max}) of its intrinsic fluorescence (\diamond), fluorescence intensity at 335 nm (I_{335}) (\blacktriangle), and IR absorption at 1650 (\square , native helix), 1629 ($+$, solvated helix), and 1676 (\circ , disordered coil) cm^{-1} (61, 62). The λ_{\max} of L-tryptophan in D_2O (\times) is shown for comparison.

significantly exposed to solvent as temperature is increased for these conditions. Contrary to the IR absorbance at 1650 cm^{-1} , where the main change is due to the melting of nativelike helices (see below), the temperature dependence of the IR absorbance at 1629 cm^{-1} is determined mainly by solvated helices and shows only linear, monotonic change. The appearance of random structure monitored at 1675 cm^{-1} has inputs from both linear and cooperative melting because both types of helices melt into disordered structure.

RESONANCE RAMAN Raman spectroscopy, as a vibrational technique, enjoys advantages similar to infrared in terms of structural specificity. Time-resolved Raman measurements have been made on subpicosecond and longer timescales (63 and references therein). From the first such studies, Raman investigations of protein dynamics have generally utilized resonance Raman scattering (64), wherein the intensity of Raman scattering is enhanced by resonance with an electronic chromophore, because normal Raman scattering is insufficiently intense. The chromophore can be a prosthetic group or intrinsic to the polypeptide (the amide groups or the aromatic side chains). Resonance enhancement of the polypeptide chromophores reports directly on the protein structure.

CIRCULAR DICHROISM Circular dichroism (CD) has been one of the major techniques used to study static structures, or slow kinetics, in protein folding (65 and references therein). Kligler and coworkers (66) have devised elegant time-resolved techniques that allow application of CD to nanosecond and longer timescale phenomena, and Xie & Simon (67) have observed time-resolved CD on picosecond timescales.

Initiation of the Folding Reaction

Several methods have recently been developed to initiate folding reactivity faster than the 1–10 ms dead time of conventional stopped flow. These include laser-induced temperature jump (34,36), laser photolysis (68), laser-initiated electron transfer (69), and submillisecond rapid mixing techniques (38). Earlier studies have used ultrasound and Joule-heating temperature jump as reaction initiation schemes for folding (70).

LASER-INDUCED TEMPERATURE-JUMP Proteins generally have a temperature range over which the functional structure is stable. Temperatures either above or below this range will unfold the structure. Thus by judicious choice of temperature and solution conditions, one can cause the equilibrium between folded and unfolded forms to be poised at any point, and an increase in temperature can displace the equilibrium in the direction of folding or unfolding. This approach is applicable to virtually all proteins and peptides, and it does not necessarily require the introduction of extraneous reagents into the folding system.

The use of fast temperature-jump methods (T-jump) to displace chemical equilibria and measure relaxation rates was pioneered by Eigen & De Maeyer (70), whereby rapid capacitance discharge induced Joule heating in conducting solutions. This approach has generally been limited to the microsecond and longer timescale. 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 kinetics of inorganic reactions by Sutin & coworkers (71). Vibrational relaxation takes place on the picosecond timescale both in water and in proteins, and experimental conditions can be devised wherein the distances between the molecules that absorb the energy responsible for the T-jump are small. Therefore, complete thermalization of solvent and solute can occur in 10–20 ps (72–74). The underlying equations governing this relaxation process have been developed (36).

In recent studies, the pump laser wavelength has been tuned either to the absorption band of dye molecules (36) in solution with the protein sample or to weak overtone bands of water or D₂O (34, 35, 40, 75, 76). In the former case, the separation between the dye molecules is relatively large and the thermal equilibration time has been estimated to be around 70 ps; thermal relaxation is much faster in the latter case because water molecules much closer together are excited. After excitation and thermalization, the heat contained in the laser interaction volume gradually escapes to nearby cold surfaces by thermal diffusion; this determines the long-timescale limit of the laser-induced T-jump. Depending on the volume of the heated solution, decay times for the T-jump ranging from a few to tens of milliseconds have been reported (both faster and longer times are possible).

One well-known problem with fast T-jumps has to do with the thermal expansion of water. The sudden change in temperature induces a shock wave that propagates through the medium with a characteristic time of 1–10 μ s. Although this effect tends to be small near 4°C (H₂O) and 12°C (D₂O), where the thermal coefficient of expansion is close to zero, it is not negligible even for symmetric temperature jumps about these temperatures.

As an example of the laser-induced T-jump setups, Figure 2 shows the apparatus we used (35). Other variants that have been developed successfully include producing near-IR light of sufficient power to pump weak water bands by Raman shifting the fundamental Nd:YAG laser emission (77–79). Our apparatus probes structure using IR absorbance in the amide-I spectral region. The change of IR absorbance of an alpha-helical 21-peptide sample after a T-jump in a typical experiment is presented in Figure 3.

It is important to note that the jump from one temperature to another changes the equilibrium position of the ensemble of possible states faster than do the chemical processes that maintain the equilibrium. This is the fundamental

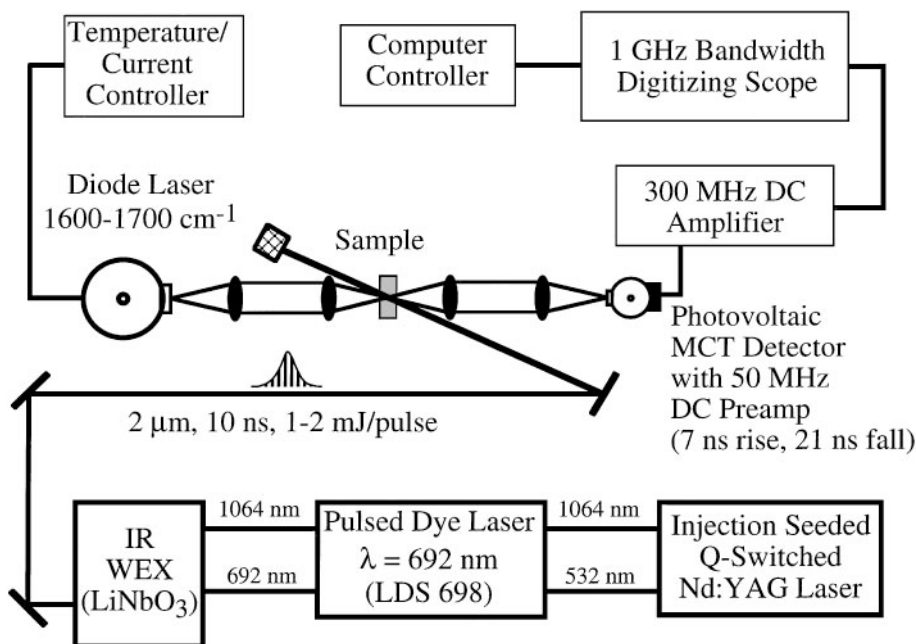


Figure 2 The scheme of our temperature-jump apparatus with the transient IR spectrometer (35). A Nd:YAG and dye lasers with frequency differencing module produce the pump radiation at $2\ \mu\text{m}$, which is the source for the temperature jump. The induced transient changes in the transmission of the probe IR beam through the sample are detected by a HgCdTe detector, digitized, and signal averaged. The combined total instrument response time is currently limited by the 21-ns fall time of detector/preamplifier system. A split IR cell is employed, one compartment containing the sample solution and the other containing reference. The cell is translated so that each compartment is probed under identical conditions. The size of the temperature jump is determined by the temperature dependence of the D_2O absorbance band.

feature of relaxation methods (70). Subsequent to the T-jump, the system relaxes to the new equilibrium point. In general, the kinetics that are observed involve both 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 (80).

FAST MIXING The major advantages of fast mixing, an important approach, are that it is completely general, that the folding occurs under near-native conditions, and that extreme concentrations of denaturant can be used to fully unfold the protein. Disadvantages are that physical limitations of mixing place a lower limit on the time resolution of the technique ($1\text{--}10\ \mu\text{s}$), that the large

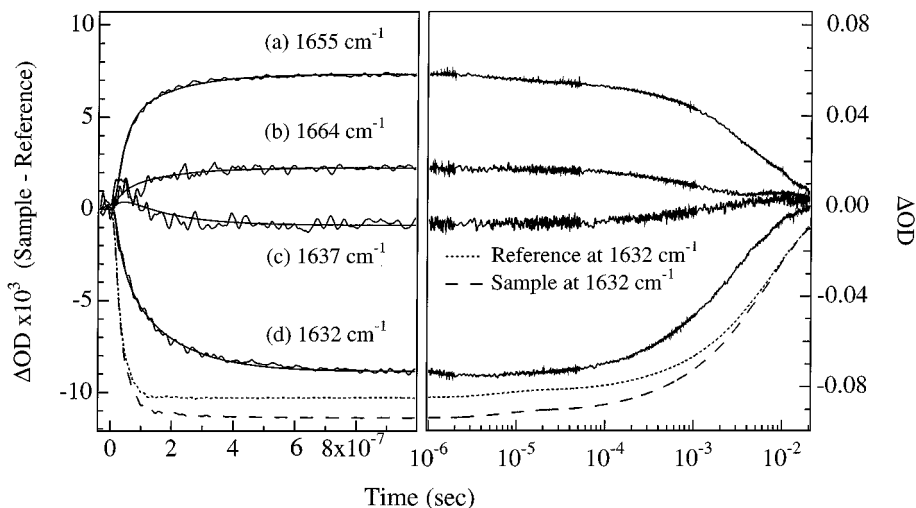


Figure 3 Change in optical density (ΔOD) at 1632 cm^{-1} of D_2O (reference, dotted line) and the 2 mM F_5 -peptide solution in D_2O , $pH^* 4.1$ (sample, dashed line), right vertical axis, and change in optical density (ΔOD) of sample minus reference at different wavenumbers, left vertical axis, as a function of time in response to a temperature jump from 9.3 to 27.4°C. The solid curves represent fits of the data to two-phase exponential kinetics. The first phase is instrument limited (faster than 10 ns), and the time constant of the second phase is 160 ± 60 ns. Reprinted with permission from (35). Copyright © 1996 American Chemical Society.

shear forces present in the mixing volume may affect the folding process of the protein, and that the turbulent nature of the flow can affect the optical quality needed for some probes.

Conventional stopped-flow mixers have a dead time of about 1 ms. The dead time of mixing techniques can be improved at least two orders of magnitude with a continuous-flow approach (81). Recently, Rousseau & colleagues adapted an early mixer design to a fast-mixing, continuous-flow technique for folding (38,48). The sample is probed downstream of mixing. Because the volume of this continuous-flow system is kept small to minimize the sample solution volume required and to improve time resolution, laser-based probes that can be tightly focused are used: for example, optical absorption, resonance Raman, and fluorescence. The best-demonstrated dead time to date is ca. 100 μs , an improvement of 1–2 orders of magnitude over routine commercial stopped-flow instruments.

PHOTOCHEMICAL TRIGGERING In a limited number of systems, folding or unfolding can be initiated by the absorption of light. The timescale accessible

to this approach is generally limited by the chemistry that light absorption induces, because the timescale of the absorption event itself is limited only by the duration of the laser pulse. The first successful method of triggering a folding reaction, that of cytochrome *c*, relied on the fact that carbon monoxide binds preferentially to the heme of the denatured state (68). Therefore, appropriate concentrations of guanidine hydrochloride were found that poised the CO form of the protein in a partially unfolded state but allowed refolding to proceed once CO was photodissociated from the heme. The photolysis event of heme-CO is a very fast process (<1 ps); hence, rapid folding could be initiated by absorption of visible light by the heme group, limited in timescale only by the duration of the laser pulse. Aside from the fact that it is applicable to very few systems, this method has two primary disadvantages: (a) The range of concentrations of denaturant required to poise the folding equilibrium at the required point for photolytic initiation to be effective is relatively narrow, and (b) the recombination of CO with the heme poses an intractable interference to observation of the slower (ca. 1 ms and longer) folding processes.

A second method of optical triggering, also employed on cytochrome *c*, involves photochemical oxidation or reduction of the heme group (69). The folded structure of cytochrome *c* is considerably less stable when the heme is oxidized compared to the reduced form of the protein. Therefore, triggering the folding reaction can be induced by the transfer of an electron to the heme iron of oxidized cytochrome *c*, which is poised in the unfolded state by suitable denaturing conditions. Photoreduction can be brought about by optically exciting a suitable compound present in solution, such as *tris*-(2,2'-bipyridine)ruthenium(II), which serves as an electron donor in the excited state and reduces cytochrome *c* at rates up to diffusion control. Like the photolysis of heme-bound CO, the folding times are determined under denaturing conditions, although a much wider range of concentrations of denaturant is possible. This approach is potentially applicable to a very wide range of timescales. Timescales shorter than diffusion control can be attained by covalently bonding the photoredox agent to an external protein residue, and arbitrarily long times can be accessed by reductive quenching of the oxidized photoproduct to prevent thermal back-electron transfer. This approach is applicable in principle to redox proteins wherein the oxidized and reduced forms have different stabilities with respect to folding; thus the approach enjoys wide but not general applicability.

A novel photochemical trigger has been used by DeGrado, Hochstrasser & coworkers (82, 83) to study formation of α -helix within a short model peptide. The peptide is constrained to a disordered conformation via a disulfide linkage that covalently bonds its ends together. Photodissociation of the link by a laser pulse eliminates the structural constraint on the peptide and initiates α -helix formation.

EMERGING TECHNIQUES In addition to T-jump, mixing of solutions and changing pH can be used in almost all systems to initiate folding or unfolding. As noted above, the shortest mixing dead times achieved in folding experiments so far have been ca. 100 μ s. However, dead times nearly two orders of magnitude shorter have been achieved by the pulsed accelerated flow technique (PAF), the fastest mixing technique developed to date (84, 85).

The photoexcitation of acidic or basic dyes often leads to very large changes in the pKa in the excited state of the dye; transient pH changes of 1–3 pH units in ca. 100 ps, for the purpose of initiating fast protonation-sensitive reactivity of species in solution other than the dyes, has been demonstrated (86–88). In the same vein, the stabilities of some proteins are affected by metal ion concentration. For many metal ions, caged metal systems have been identified whereby fast photolytic release of metal ions is possible (e.g. 89).

Ultraviolet excitation in resonance Raman spectroscopy allows resonance enhancement of protein vibrations, including those of amino acid side chains and the amide vibrations of the polypeptide backbone (90, 91). It has the structural specificity of vibrational spectroscopy but, unlike infrared, is not subject to interferences from water. In addition, the resonance effect allows one to choose, through selection of the excitation frequency, which specific ultraviolet chromophore (e.g. tyrosine, tryptophan, or amide) will be observed.

SYSTEMS

Peptide Models

Proteins are composed of runs of secondary structure. Clearly, our understanding of the overall folding process in proteins will be aided by study of the dynamics of secondary structure 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 and sheet structures. There was a period of substantial experimental and theoretical work in the mid-1960s to early 1970s concerning the thermodynamics and dynamics of helix formation, primarily in homopolymers. The kinetic studies employed dielectric loss (92), ultrasonic absorption (93–96), resistive temperature jump (97–99), and electric-field jump (100–102). These studies have been reviewed by Gruenewald et al (94). 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 (which included fluorescence and ultrasound or UV light absorption) lacked structural specificity and because the helix formation rates were inferred indirectly from the data in most cases. In addition, the systems that were studied involved long homopolypeptides (e.g. poly-L-glutamic acid) in

solution. These are probably not good models for the short runs of secondary structure usually found in proteins.

Recently, this problem was revisited on a few short peptides through the use of laser-induced T-jump techniques. The S-peptide of ribonuclease A, obtained by cleaving off the first 21 amino acid residues, forms a helical structure in solution (about 30% helix and 70% unfolded at temperatures near 5°C) (103). Through the use of the amide-I' band to monitor helix formation, a relaxation time of faster than 300 ns has been observed for the S-peptide (104).

A study has been reported of the C-terminal β -hairpin of protein G B1 (residues 41–56) (105). By monitoring tryptophan fluorescence and Förster energy transfer in a dansylated derivative, relaxation rates of 4 μ s at 15°C and 0.9 μ s at 55°C were observed. Thus the folding relaxation times are factors of 10–100 slower for this β -structure than any that have been reported on helical model systems.

The peptide system that has been studied the most thoroughly is the synthetic F_S peptide, a small 21-residue peptide that has the sequence X-(A)₅-(AAARA)₃-A-NH₂. Alanine has strong helix propensity, and about 70–90% of this peptide is found in a helical conformation in solution at room temperature (106, 107). In one study, wherein the N-terminal residue was succinyl (X = Suc), the amide-I' bands were used as an indication of helix content (35). In a second study, X was a reporter fluorophore (X = 4-(methylamino)benzoic acid, MABA), and the fluorescence quenching of MABA was probed (79). Hence, these two studies are somewhat complementary in that the IR probes the average amount of helix found along the chain (all amide I' backbone C=O stretches are probed equally) while the MABA fluorophore is sensitive to changes at the end of the peptide (this assumes, of course, that the changes in the chemical nature of the X group do not affect, in a serious way, helix stability or kinetics). In the IR study, the relaxation kinetics were well described by a biexponential model with components of ca. \sim 10 ns (small amplitude, possibly resulting from fraying of the helix ends) and 160 ns (dominant amplitude, resulting from helix melting) for a T-jump from 9.3 to 27.4°C (Figure 4) (35). No changes in helicity were observed by the amide I criterion between 50 ps and 10 ns. The relaxation time is weakly temperature dependent (107a). In the fluorescence study, a single relaxation time was found near 20 ns, about eight times faster at the same temperature than that found in the IR study, which was also weakly temperature dependent (Figure 4).

Thompson et al (79) published a statistical model to account for the experimental kinetics based on a picture of helix formation (103) that assumes that the activation barrier for the helix formation is largely entropic. Two elementary processes, nucleation and propagation, are supposed; a distinction is made between the probability of forming the first loop of helix (nucleation) and the

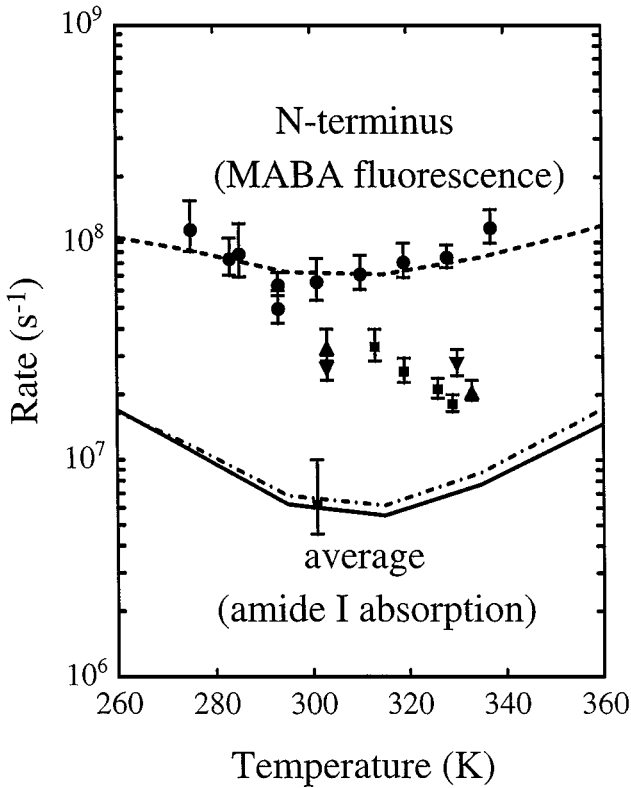


Figure 4 Temperature dependence of the relaxation rates of the Fs-peptide measured by fluorescence intensity of the 4-(methylamino)benzoic acid (MABA)-labeled peptide (●) (79) and IR absorption (■) (35). The curves are the fast (*dashed line*) and slow (*dot-dashed line*) calculated relaxation rates for the decay of N-terminal content and the calculated relaxation rate for the average helix content (*continuous line*) (79). Our data for the fast relaxation phase of the apomyoglobin forms are included: N-form, (▲); I-form, (▼); E-form, (■) (61, 104, 107b). Adapted from Reference 79. The drawing is kindly provided by J Hofrichter, National Institutes of Health.

formation of subsequent loops (propagation). Nucleation is viewed as less probable than helix propagation. Thompson et al additionally assumed that each peptide contains either no helical residues or only a single continuous region of helix. This reduced the number of possible conformations enormously because two or more segments of helix separated by nonhelical regions are excluded. Calculation from this model of the relaxation kinetics induced by a temperature jump yielded two distinct processes with characteristic rates: (a) a faster process involving rapid folding/unfolding equilibrium of the helix ends and

(b) a slower process involving an equilibrium that must cross the nucleation free energy barrier between helix-containing and non-helix-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 (Figure 4). The model, however, does not account for the essentially single exponential character that was found in the fluorescence study.

Hochstrasser & coworkers (83) recently used amide I infrared absorbance to study the conformational dynamics of peptides having the sequence Ac-X-(AAAAL)₃-X-CONH₂, where Ac is acetyl and X is a synthetic amino acid having an aryl thiol sidechain (82). Consistent with the results on the F_S-peptide, no significant formation of α -helix occurs on times as long as 2 ns, the longest times accessed in this experiment.

Proteins

APOMYOGLOBIN The folding of apomyoglobin (apoMb, myoglobin without its heme group) has received intense scrutiny, in part because the protein is relatively simple. It serves as an archetype for folding of proteins that are small, single-domain, and globular. The holoprotein contains 153 residues organized into eight strands of mostly α -helical segments, labeled A–H (108). Near neutral pH, apoMb adopts a structure that is natively like according to the available NMR, CD, and calorimetric evidence (109–112). Thus native apoMb 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. The major differences between the apo- and holoprotein appear to occur in the C, D, and F helices, which are considerably more disordered in the apoprotein.

Under varying conditions of low pH and specific salt concentrations, apoMb forms various destabilized structures that are variably extended, depending on conditions. Three different folded structures can be defined (62, 109). The structure of ApoMb-N (Native form), formed near neutral pH, is believed to resemble closely the holoprotein. It is the most stable and the most compact of the three structures. ApoMb-I (Intermediate form), formed at lower pH and specific salt conditions, has a looser structure. NMR techniques coupled with hydrogen-deuterium exchange measurements, small-angle scattering, and more recent mutant and FT-IR studies show that the I form contains a tight core consisting of the intersection of the AGH helices and loose solvated helical regions

(104, 113). A species formed at still lower pH and low salt, the E form (extended form), contains essentially only the AGH core and is otherwise very extended (62). Thus, another attractive feature of apoMb in folding studies is the ability to access different equilibrium forms of the protein having intermediate stability between the native and unfolded states. Furthermore, recent evidence suggests that these equilibrium intermediate states are structurally close to kinetic intermediates on the folding pathway (native, molten globule, and unfolded) (109). A striking structural similarity between the I form and an early folding intermediate has been demonstrated (114, 115).

Experimental studies by two groups have addressed the fast folding dynamics of apoMb. Both groups used laser-induced T-jump to displace the folded/unfolded equilibrium. However, one used fluorescence and the other used IR absorbance as probes, the details of the T-jump regime employed were different, and solution conditions differed so that different apoMb structures were being observed. It is important to realize that both of these studies involved displacement of the folded/unfolded equilibrium by T-jump and thus unavoidably were relaxation kinetics studies. Accordingly, the observed relaxation dynamics contain contributions from both the folding and unfolding rates, regardless of whether the process observed in the relaxation involves the formation of a more folded or less folded ensemble. The extraction of rate constants or lifetimes specifically describing folding or unfolding depends on knowledge of the equilibria involved and adoption of a kinetics model within which the relaxation data can be treated.

In our laboratories, we have used IR absorbance to study the dynamics of the transition from more folded to less folded ensembles in successively more denatured (N, I, and E) apoMb states. The folding pathway can thus be “peeled away,” and processes on both fast and slower timescales 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, but, as noted above, at least the I form has been identified with an intermediate on the folding pathway under conditions where the N form is stable.

From a technical point of view, apoMb has been a fortuitous choice for the study of protein folding. The horse heart protein that has been studied contains just two Trp residues, and these are located in the interior of the protein, on the A helix at the points where the A helix intersects the G-H hairpin to form the AGH core (108). Thus in studies that have monitored Trp fluorescence, the structure of the monitor is strategically located, and the state of folding of the AGH core can be specifically monitored. Also, because the protein consists essentially only of helix or disordered structures, the amide-I' band of apoMb consists largely of marker bands that involve only buried native helix, solvated helix, or unfolded conformations (61). Hence, the amide-I' IR absorption band

contains a rich amount of information that is readily interpreted. Finally, the IR results of static and kinetic properties, which can arise from any part of the protein (in the absence of isotope labeling), can be compared to Trp fluorescence studies that report only on events that affect the AGH core. In this way many of the IR results can be associated with specific substructures of the protein, and the complementary IR and fluorescence spectroscopies lead to a very detailed picture (104). Figure 5 shows representative kinetic relaxation data of the native form of apomyoglobin using IR absorbance as a probe.

With the assumption that the structures of the I and E forms mimic structures along the folding pathway (as indicated above), Figure 6 shows the folding pathway in schematic form. We have found that apoMb consists of substructures or subdomains that have various thermodynamic stabilities and kinetic labilities (61, 104). The different substructures can be recognized by the secondary structure elements that have different IR spectra: nativelike helix; solvated helix wherein the secondary structure is intact but not as rigid, with the amide carbonyls exposed to the surrounding solvent; and unfolded structure (as indicated by the marker amide-I' IR bands). The native helices constitute a tightly packed substructure that shows cooperative melting transitions, while the solvated helical substructures show nearly linear melting curves. In response to a temperature jump, the helix folding/unfolding relaxation of the solvated helical substructure exhibits relaxation times typical of isolated helices such as the model peptide described above (ca. 30–150 ns; see the comparison of relaxation times of the solvated helical moiety of different apoMb forms with those of the model peptide, Figure 4). 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 tens of nanoseconds timescale (61).

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 (116, 117). Moreover, the native tertiary fold involving the intersection of the A, G, and H helices forms rapidly, but subsequent to the formation of the helices themselves. No relaxations involving the AGH core have been observed in our experiments on the I or N states, but the formation time of the AGH tertiary structure has been seen to be ca. 90 μ s in the E state at 35°C (both infrared and fluorescence measurements are required to measure the timescale of this relaxation and to show that the AGH core is the structure involved; 107b).

The AGH intersection at the time of its initial formation may involve only about 20 residues, but it is nevertheless a stable structure with a large negative enthalpy of formation, as shown by its cooperative melting (62). The fast formation of a nativelike structure that acts as an energy trap has several advantages in the folding of the protein. The speed may ensure that other non-native but

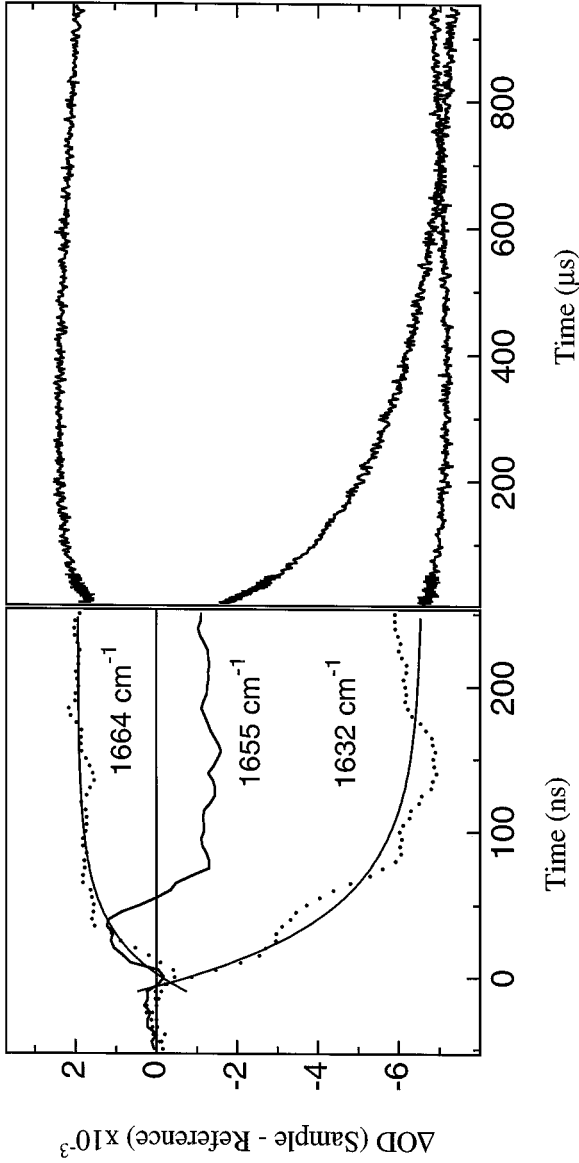


Figure 5 Kinetic IR response for a T-jump from 45 to 60°C at 1664 cm^{-1} (maximal positive effect, disordered structure input presumably), 1655 cm^{-1} (near the maximal bleaching effect, native helix input presumably), and 1632 cm^{-1} (solvated helix input presumably). The region from 0.2 to 8 μs is obscured by a photoacoustic artifact and is not shown. Two clear kinetic phases are present, one on the nanosecond timescale and the other on the microsecond timescale. Single-exponential fits to the data, shown by *solid lines*, yielded 48 ns and 120 μs for the two lifetimes. All three signals begin to show recovery on the millisecond timescale with a ≈ 7 -ms time constant resulting from a cooling after the T-jump. Reprinted with permission from Reference 61. Copyright © 1997 National Academy of Sciences, USA.

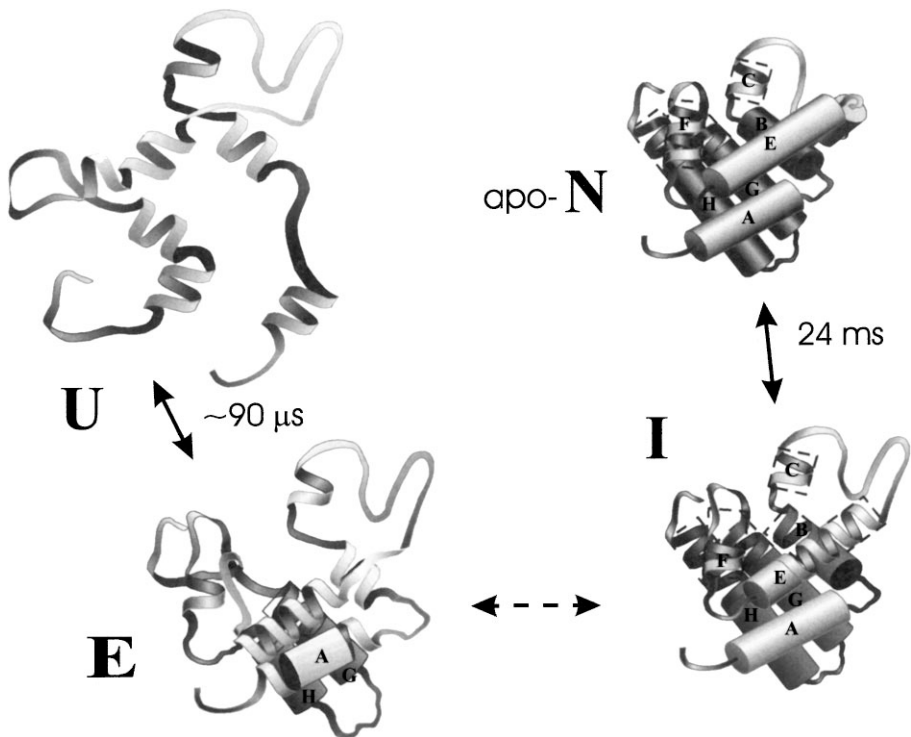


Figure 6 A cartoon of the folding of apomyoglobin starting from an unfolded (U) state. There are good experimental reasons to assume that the I and E states of apoMb mimic structures found along the folding pathway (see text), and this forms the basis of the schematic of the folding pathway. The folding time of $90 \mu\text{s}$ 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 relaxation signals are also observed (see Figure 4), which arise from the melting of the solvated helical portion of the E state. The folding times under native chemical conditions may differ. The folding time of 25 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 relaxation measurements of the N and I forms of the protein also show melting of their solvated helical portions on the nanosecond timescale (Figure 4). The folding kinetics from the E to I state have not yet been determined, but they are likely to be fast.

metastable folds do not form. In the specific case of apoMb, tying the two ends together reduces enormously the amount of conformational space that needs to be searched in further 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 because the helices form on a timescale two to three orders of magnitude faster than the tertiary contacts (35, 61, 79).

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-F loop, which has the effect of compacting the protein further. Also, the tertiary contacts in the A helix and the G-H hairpin are likely to 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. These two events, formation of the core and then its growth, could be termed nucleation followed by growth (27). 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 for the association of the AGH core and is temperature sensitive, 280 μ s at 60°C (61) and ca. 1 s at 5°C (114), indicating a substantial formation energy barrier.

Gruebele & coworkers have used Trp fluorescence to measure the relaxation kinetics from a cold-denatured state to higher temperatures, which under the conditions used induces folding of the protein to the N state (40, 47, 118). In this case, the process being probed is a transition from a less folded to a more folded ensemble. Studies are now under way to characterize the structure of the cold-denatured state (117), but the fluorescence data are consistent with partial unfolding of the AGH core, particularly a change in the proximity of Trp14 and Met131, at -8°C. Upon raising the temperature from -8°C to 10°C, substantial refolding occurs. The dynamics of this refolding process were followed by fluorescence, and a single relaxation was observed with a 5 μ s lifetime (47). This result was assigned to the closing of the loop between the GH hairpin and the A helix. Subsequently, improved data allowed the observation of a second, faster relaxation at ca. 250 ns (40), which, in accord with our earlier results (35, 75, 76), was assigned to the formation of α -helix.

The results on apoMb from our laboratories and those of Gruebele & coworkers suggest two timescales for the formation of the AGH core, depending on conditions: 90 μ sec in the E state at 35°C and 5 μ sec in the N state at 10°C. Considering that the E-state results concern an AGH structure much less stable (denaturing conditions) than that of the N state at 10°C (native conditions), a formation rate more than a factor of 10 faster for the AGH core in the latter case, despite the lower temperature, may constitute satisfactory agreement. These times are fast. They are close to or shorter than that predicted by a simple

polymer-diffusion model (41, 119), 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 (more than 100 amino acid residues). This is particularly so given that a number of tertiary contacts are also established in forming the natively like AGH core. However, the short timescale of this tertiary folding event may not be unreasonable if it is realized that the protein has had ample time to form helical segments in the intervening B-F structure as well as the GH peptide before the AGH core is formed and that these helices can guide the formation of the AGH tertiary contacts in ways that are directed and faster than diffusion of random structures.

CYTOCHROME C The results from the fast folding processes of this protein have been reviewed recently, and the reader is referred to these reviews for more detail than is given here (24, 26). Three groups have studied the fast folding reactions of cytochrome c, and each has adopted different approaches to reaction initiation: rapid mixing, photodissociation of CO, and photoinitiated electron transfer. Different spectroscopic probes have also been used: Trp fluorescence, heme resonance Raman scattering, and heme electronic absorption.

Quenching by Förster energy transfer to the heme of the fluorescence of the only tryptophan in cytochrome c has been taken as a measure of the compactness of the protein (48). Rapid mixing experiments using Trp fluorescence quenching as a probe have shown that there is a burst phase collapse of the unfolded protein in less than 100 μ sec followed by slower processes (48).

One of the slower processes appears to be ligand exchange wherein the heme first coordinates non-native histidine side chains and then exchanges these for the native histidine ligand on the submillisecond timescale (38, 120) in a process that shows nonexponential behavior. The initial formation of non-native tertiary contacts that are necessary for the coordination of "incorrect" histidine imidazoles to the heme must be unmade to form the native fold. The use of imidazole in solution as an extrinsic heme ligand blocks the non-native heme-histidine coordination and prevents the concomitant formation of the non-native folds. Under these conditions, the slower kinetics become exponential and the unresolved $<100 \mu$ s feature is followed by a single 600- μ s process. The two processes have been interpreted as a partial collapse of the protein to a more compact but still denatured state in an energy-barrierless process (fast phase) followed by reorganization to the folded conformation over a small energy barrier (about 4 kcal/mol or less). This sequence of events is shown schematically in Figure 7.

Studies of the initiation of ferrocycytochrome c folding by nanosecond-timescale photodissociation of CO have shown that the rate of heme binding to Met80, located about 50 residues from the heme, occurs with a lifetime of 40 μ s

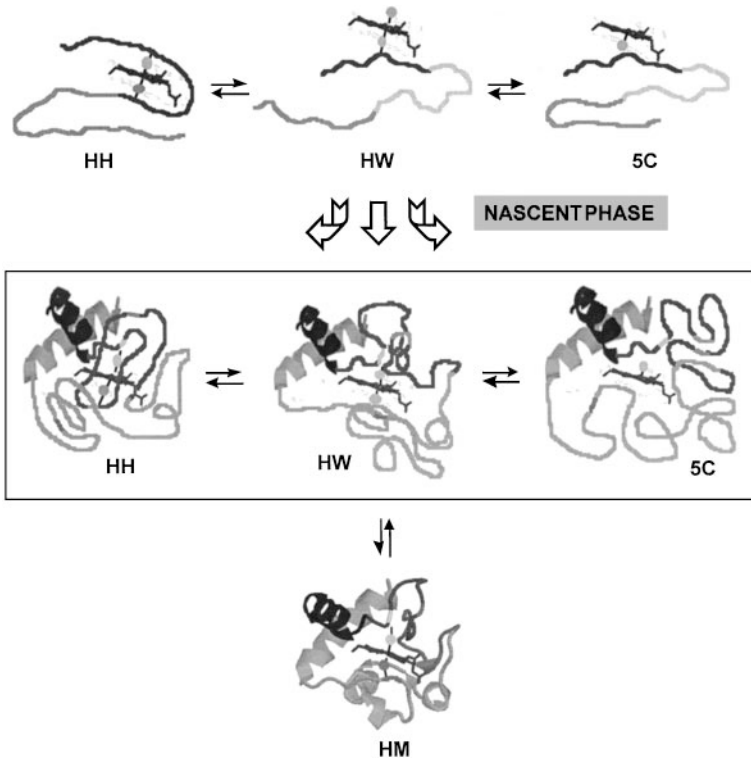


Figure 7 Schematic diagram representing the folding of cytochrome c. The unfolded protein can contain a mixture of histidine-histidine (HH), histidine-water (HW), and five-coordinate (5C) heme ligation states, depending on pH conditions. When the protein is placed under folding conditions, a redistribution of these states, termed the nascent phase, occurs within the 100- μ s dead time of the rapid mixer. During this phase the protein collapses to a semicompact structure so that a thermodynamically controlled ligand exchange phase can take place in the subsequent phase, leading to the native histidine-methionine (HM) structure. The two helices at the C and N terminal were assumed to be present at the early stage of folding based on prior reports. Although the water ligand for 5C can be on either side of the heme plane, for simplicity, only one configuration is illustrated. The drawing is kindly provided by D Rousseau, Albert Einstein College of Medicine.

(24,26, and references therein). Kinetics experiments in which the binding of extrinsic methionine with a small polypeptide containing the heme group was studied were performed and yielded nearly identical results. These results were taken as evidence that the kinetics of methionine rebinding are controlled by polymer diffusion and that the 40- μ s rebinding lifetime is as expected for diffusion-limited first contact of the ends of a polypeptide 50 amino acid residues in length (41, 119). This time is related to that for formation of a 6-10 residue

loop, a typical size for loops in proteins, and suggests that the smaller loop could be formed in about 1 μs (29). These results and the interpretation pose the question of whether diffusion-controlled loop formation is the fastest time for the formation of compact structures in proteins (viz. a “speed limit for protein folding”; 29). An affirmative answer to this question asserts that passive diffusion is the rate-limiting step in forming compact protein structures and that faster processes such as helix formation do not play an active role. This may be correct for the special case of this specific folding process (methionine rebinding to the heme in cytochrome c), wherein the loop between Cys17, which is covalently bound to the heme, and Met80 contains very little helix. However, the experimental results on the formation of the AGH core in apoMb cited in the previous section and the arguments presented therein suggest that this particular speed limit is not, in general, strictly enforced.

The folding of ferrocycytochrome c has also been observed following reaction initiation by photochemically generated reductants (69). The folding reaction initiated by the ruthenium chemistry allowed an observation window from $<1 \mu\text{s}$ to ca. 1 ms, while that based on cobalt chemistry allowed times $>1 \text{ ms}$ to be observed. On the faster timescale, the authors observed a change in heme absorbance with a lifetime of 40 μs , which, as in the CO-photodissociation study described above, was ascribed to a change in heme ligation. In this study the authors suggested that hydrophobic collapse accompanied the 40- μs ligation event. On the longer timescales, a single-exponential folding reaction was observed that was assigned to the final folding reaction of the protein. The rate constant for this reaction was dependent on denaturant concentration and was extrapolated to zero denaturant concentration, where the rate was estimated to be 2000 s^{-1} . The free energies for folding in the two oxidation states of the protein, the redox free energies in the folded and unfolded states, and the dependence of the folding rate on conditions allowed inferred comparisons of the ferrocycytochrome c folding kinetics to those determined by others for ferricytochrome c.

The conclusions of Pascher et al that the 40- μs rate represents hydrophobic collapse has been challenged by Eaton & coworkers (121), who assert that such collapse should be accompanied by almost total quenching of Trp fluorescence. In their studies (initiated by photodissociation of CO from the reduced, unfolded protein), no such quenching occurs at 40 μs , nor on any other timescale between 10 ns and 10 ms. This argument involves the implicit assumption that the folding process is similar for initiation by photodissociation of ferroheme-CO as it is for initiation by photoreduction of ferriheme. Winkler & Gray have responded (122) by pointing out that there is reasonable evidence that the structures of the unfolded structures of ferrous and ferric cytochrome c are indeed different and that there are other significant differences between the two experiments.

BARSTAR The fast folding kinetics from the cold-denatured state of barstar, a relatively small 89-residue 10-kDa protein, have been investigated by Fersht & colleagues through resistive (microsecond) T-jump techniques (123, 124). The native structure of this protein has four α -helices and three strands of parallel β -sheets (125). The cold-denatured state of this protein has been characterized by NMR spectroscopy and contains nativelike structure, which is, however kinetically labile, in two of the four helices (126, 127). The folding kinetics are characterized by several discrete stages: a fast folding intermediate formed in a few hundreds of microseconds where some nativelike contacts have formed; a slower millisecond intermediate that is quite nativelike; and a very slow process arising from a proline isomerization.

The nature of the transition state between the cold-denatured state and the early folded state was probed with a procedure called Φ analysis. In Φ analysis, the changes in free energy of folding between the wild-type protein and selected mutants are compared to the changes in the folding kinetics induced by the same mutations. This analysis yields the changes in the transition state free energy and shows that the first helix, and the second to a lesser extent, have become substantially consolidated in the transition state of the fast folding intermediate. On the other hand, peptide fragments of the first two helices show little propensity toward helical structures under conditions favoring helix formation. It was thus suggested that the early folding stages involve a folding nucleus that consists of a sequence of neighboring residues whose stability as a (helical) unit is, however, maintained by long-range interactions within the protein (the nucleation-condensation mechanism; 124). The formation of the folding nucleus and its stabilization by long-range intraprotein interactions occur simultaneously. The mechanism of this early stage appears to be very similar to that proposed for the smaller chymotrypsin inhibitor 2 (27, 128). This small protein contains only a single module of structure, and its kinetics are simple two state. This is in contrast to studies on barnase (27), where substantial helical structure is present at the very earliest stage and where the protein folds in a more stepwise and multistate fashion.

RIBONUCLEASE A Hochstrasser & coworkers (36) have combined laser T-jump on the picosecond timescale with ultrafast infrared probe to observe the events in the first few nanoseconds after the temperature jump is applied to ribonuclease A. The T-jump had a risetime of 70 ps, and temperature changes in the range of 3–6°C were achieved. At pH 5.7 and 59°C, an amide I absorbance change within the T-jump rise time was observed and tentatively ascribed to solvation effects. Following this initial change, no amide I changes were observed before 1 ns had elapsed. Similar induction periods for response of polypeptide structure to picosecond perturbations were described in the foregoing section

on peptide dynamics, for two different classes of peptide (the F_S peptide and the $(AAAAL)_3$ systems) and two different reaction initiation schemes (vide supra); this may turn out to be a common phenomenon. In ribonuclease A, infrared changes attributed to melting of β -structure are observed between 1 and 5.5 ns. The transient IR difference signal resembles the static difference spectrum observed upon unfolding of the protein but has only ca. 15% of the amplitude of the (native unfolded) difference.

CYTOCHROME b_{562} Gray & coworkers (129) have used photoinduced electron transfer to initiate rapid folding of a four-helix bundle heme protein, cytochrome b_{562} . Unlike cytochrome c, the heme of cytochrome b_{562} is not covalently bound to the polypeptide other than via the axial heme ligands, Met7 and His102. As observed by heme electronic absorption, the folding process obeyed single-exponential kinetics with a rate constant of 800 s^{-1} at 2.5 M guanidine hydrochloride. This is almost a factor of 100 faster than ferrocyanochrome c and is comparable to the nonheme four-helix-bundle protein acyl-coenzyme A binding protein.

PHYTOCHROME The photoinduced transformation of switch-off phytochrome A (Pr) to the switch-on forms is triggered by red light (e.g. 660 nm) and is accompanied by the folding of N-terminal residues 20–50 into an α -helical structure. Time-resolved CD studies (130) revealed a decrease in α -helical content of the protein with a lifetime of 310 μs , suggesting that the rate of helix unfolding must be limited by the dissociation of tertiary contacts because the unfolding of unconstrained α -helix should be 100 to 1000 times faster (see above).

SUMMARY

The development and use of new techniques to study the fast folding kinetics of proteins is now just 4 or 5 years old. In that time, a number of important discoveries have been made. For the first time, experiment is now ahead of theory in describing primary folding processes. Still, it is clear that such studies are just beginning, that the techniques need further development particularly in regard to specific structural sensitivity (certainly not an impossible demand experimentally), that the systems that have been studied need more work, and that a wider range of systems need to be studied for there to be general rules formulated on how proteins fold. It has been found that helical assemblies in proteins as well as model systems form on the 10- to 100-ns timescale while β -sheet formation of a model required times of about 1 ms. Formation of tertiary structure from a fully extended unfolded protein has been demonstrated to occur within 50–100 μs and perhaps 10 times faster. Most

interesting, it has been found that a small portion and small number of residues can form a cooperative folding unit, one whose enthalpy is high enough to show a sigmoidal melting transition, in about 100 μ s (apomyoglobin). In apomyoglobin, the protein folds in parts; that is, secondary structure precedes tertiary structure formation, which proceeds to further secondary and tertiary structure formation.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Science Foundation, for grant MCB-9727439 to RHC, and the support of the National Institutes of Health, for grant GM53640 to RBD, grant GM35183 to RHC, and grant GM45807 to WHW. We thank Martin Gruebele, James Hofrichter, and Denis Rousseau for the reprints of their papers prior to publication and for copies of their figures.

Visit the *Annual Reviews* home page at
<http://www.AnnualReviews.org>.

Literature Cited

1. Anfinsen CB. 1973. *Science* 181:223–30
2. Kim PS, Baldwin RL. 1990. *Annu. Rev. Biochem.* 59:631–60
3. Dill KA. 1990. *Biochemistry* 29:7133–51
4. Levinthal C. 1968. *J. Chem. Phys.* 65:44–45
5. Montal M. 1995. *Annu. Rev. Biophys. Biomol. Struct.* 24:31–57
6. Rosenfeld R, Vajda S, DeLisi C. 1995. *Annu. Rev. Biophys. Biomol. Struct.* 24:677–700
7. Cordes MHJ, Davidson AR, Sauer RT. 1996. *Curr. Opin. Struct. Biol.* 6:3–10
8. Betz SF, Liebman PA, DeGrado WF. 1997. *Biochemistry* 36:2450–58
9. Shakhnovich EI. 1997. *Curr. Opin. Struct. Biol.* 7:29–40
10. Baker D, Agard DA. 1994. *Biochemistry* 33:7505–9
11. Murphy KP. 1995. In *Protein Stability and Folding: Theory and Practice*, ed. BA Shirley, 40:1–34. Totowa, NJ: Humana
12. Friesner RA, Gunn JR. 1996. *Annu. Rev. Biophys. Biomol. Struct.* 25:315–42
13. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. *Proteins: Struct. Funct. Genet.* 21:167–95
14. Karplus M, Sali A. 1995. *Curr. Opin. Struct. Biol.* 5:58–73
15. Shortle D, Wang Y, Gillespie JR, Wrabl JO. 1996. *Protein Sci.* 5:991–1000
16. Dill KA, Chan HS. 1997. *Nat. Struct. Biol.* 4:10–19
17. Guo Z, Brooks CL III, Boczek EM. 1997. *Proc. Natl. Acad. Sci. USA* 94:10161–66
18. Onuchic JN, Luthey-Schulten Z, Wolynes PG. 1997. *Annu. Rev. Phys. Chem.* 48:545–600
19. Chothia C, Hubbard T, Brenner S, Barns H, Murzin A. 1997. *Annu. Rev. Biophys. Biomol. Struct.* 26:597–627
20. Evans PA, Radford SE. 1994. *Curr. Opin. Struct. Biol.* 4:100–6
21. Ptitsyn OB. 1995. *Curr. Opin. Struct. Biol.* 5:74–78
22. Creighton TE, Darby NJ, Kemmink J. 1996. *FASEB J.* 10:110–18
23. Dyson HJ, Wright PE. 1996. *Annu. Rev. Phys. Chem.* 47:369–95
24. Eaton WA, Thompson PA, Chan CK, Hagen SJ, Hofrichter J. 1996. *Structure* 4:1133–39
25. Miranker AD, Dobson CM. 1996. *Curr. Opin. Struct. Biol.* 6:31–42
26. Eaton WA, Munoz V, Thompson PA, Chan CK, Hofrichter J. 1997. *Curr. Opin. Struct. Biol.* 7:10–14
27. Fersht AR. 1997. *Curr. Opin. Struct. Biol.* 7:3–9

28. Roder H, Colon W. 1997. *Curr. Opin. Struct. Biol.* 7:15–28
29. McCammon JA. 1996. *Proc. Natl. Acad. Sci. USA* 93:11426–27
30. Bai Y, Englander SW. 1996. *Proteins: Struct. Funct. Genet.* 24:145–51
31. Huang GS, Oas TG. 1995. *Proc. Natl. Acad. Sci. USA* 92:6878–82
32. Smith LJ, Fiebig KM, Schwalbe H, Dobson CM. 1996. *Fold. Des.* 1:R95–106
33. Zwanzig R. 1997. *Proc. Natl. Acad. Sci. USA* 94:148–50
34. Callender RH, Gilmanshin R, Dyer RB, Woodruff WH. 1994. *Phys. World.* August:41–45
35. Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, et al. 1996. *Biochemistry* 35:691–97
36. Phillips CM, Mizutani Y, Hochstrasser RM. 1995. *Proc. Natl. Acad. Sci. USA* 92:7292–96
37. Pascher T, Chesick JP, Winkler JR, Gray HB. 1996. *Science* 271:1558–60
38. Takahashi S, Yeh SR, Das TK, Chan CK, Gottfried DS, et al. 1997. *Nat. Struct. Biol.* 4:44–50
39. Chen EF, Goldbeck RA, Kliger DS. 1997. *Annu. Rev. Biophys. Biomol. Struct.* 26:327–55
40. Ballew RM, Sabelko J, Gruebele M. 1996. *Proc. Natl. Acad. Sci. USA* 93:5759–64
41. Hagen SJ, Hofrichter J, Szabo A, Eaton WA. 1996. *Proc. Natl. Acad. Sci. USA* 93:11615–17
42. Engelhard M, Evans PA. 1996. *Fold. Des.* 1:R31–37
43. Beechem JM, Brand L. 1985. *Annu. Rev. Biochem.* 54:43–71
44. Eftink MR. 1994. *Biophys. J.* 66:482–501
45. Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, et al. 1991. *Biopolymers* 31:119–28
46. Lakowicz JR. 1983. *Principles of Fluorescence Spectroscopy*. New York: Plenum. 496 pp.
47. Ballew RM, Sabelko J, Gruebele M. 1996. *Nat. Struct. Biol.* 3:923–26
48. Chan CK, Hu Y, Takahashi S, Rousseau DL, Eaton WA, Hofrichter J. 1997. *Proc. Natl. Acad. Sci. USA* 94:1779–84
49. Spiro TG, ed. 1987. *Raman Spectra and the Conformations of Biological Molecules*, Vol. 1. New York: Wiley. 352 pp.
50. Callender R, Deng H. 1994. *Annu. Rev. Biophys. Biomol. Struct.* 23:215–45
51. Susi H, Byler DM. 1986. *Methods Enzymol.* 130:290–311
52. Arrondo JLR, Muga A, Castresana J, Goni FM. 1993. *Prog. Biophys. Mol. Biol.* 59:23–56
53. Prestrelski SJ, Byler DM, Thompson MP. 1991. *Int. J. Pept. Protein Res.* 37:508–12
54. Prestrelski SJ, Byler DM, Thompson MP. 1991. *Biochemistry* 30:8797–804
55. Haris PI, Chapman D. 1995. *Biopolymers* 37:251–63
56. Martinez G, Millhauser G. 1995. *J. Struct. Biol.* 114:23–27
57. Gilmanshin R, Van Beek J, Callender R. 1996. *J. Phys. Chem.* 100:16754–60
58. Barron LD, Hecht L, Wilson G. 1997. *Biochemistry* 36:13143–47
59. Reinstadler D, Fabian H, Backmann J, Naumann D. 1996. *Biochemistry* 35:15822–30
60. Chirgadze YN, Shestopalov BV, Venyaminov SY. 1973. *Biopolymers* 12:1337–51
61. Gilmanshin R, Williams S, Callender RH, Woodruff WH, Dyer RB. 1997. *Proc. Natl. Acad. Sci. USA* 94:3709–13
62. Gilmanshin R, Dyer RB, Callender RH. 1997. *Protein Sci.* 6:2134–42
63. Lau A, ed. 1994. *Time-Resolved Vibrational Spectroscopy*, Vol. 74. Berlin: Springer-Verlag. 268 pp.
64. Woodruff WH, Farquharson S. 1978. *Science* 201:831–33
65. Fasman G, ed. 1996. *Circular Dichroism and the Conformational Analysis of Biomolecules*. New York: Plenum. 738 pp.
66. Wen YX, Chen EF, Lewis JW, Kliger DS. 1996. *Rev. Sci. Instrum.* 67:3010–16
67. Xie XL, Simon JD. 1990. *J. Am. Chem. Soc.* 112:7802–3
68. Jones CM, Henry ER, Hu Y, Chan CK, Luck SD, et al. 1993. *Proc. Natl. Acad. Sci. USA* 90:11860–64
69. Pascher T, Chesick JP, Winkler JR, Gray HB. 1996. *Science* 271:1558–60
70. Eigen M, De Maeyer LD. 1963. In *Technique of Organic Chemistry*, ed. SL Friess, ES Lewis, A Weissberger, 8:895–1054. New York: Interscience
71. Beitz JV, Flynn GW, Turner DH, Sutin N. 1970. *J. Am. Chem. Soc.* 92:4130–33
72. Anfinsen PA, Han C, Hochstrasser RM. 1989. *Proc. Natl. Acad. Sci. USA* 86:8347–91
73. Genberg L, Richard L, McLendon G, Miller RJD. 1991. *Science* 251:1051–54
74. Richard L, Genberg L, Deak J, Chiu H-L, Miller RJD. 1992. *Biochemistry* 31:10703–15

75. Woodruff WH, Dyer RB, Callender RH, Paige K, Causgrove T. 1994. *Biophys. J.* 66:A397
76. Williams S, Causgrove TP, Gilmanshin R, Dyer RB, Woodruff WH, et al. 1995. In *Spectrosc. Biol. Mol., Eur. Conf., 6th*, ed. JC Merlin, S Turrell, JP Huvenne, pp. 105–6. Dordrecht, Netherlands: Kluwer
77. Turner DH, Flynn GW, Sutin N, Beitz JV. 1972. *J. Am. Chem. Soc.* 94:1554–59
78. Ballew RM, Sabelko J, Reiner C, Gruebele M. 1996. *Rev. Sci. Instrum.* 67:3694–99
79. Thompson PA, Eaton WA, Hofrichter J. 1997. *Biochemistry* 36:9200–10
80. Cantor CR, Schimmel PR. 1980. *Bio-physical Chemistry*. San Francisco: Freeman. 1371 pp.
81. Regenfuss P, Clegg RM, Fulwyler MJ, Barrantes FJ, Jovin TM. 1985. *Rev. Sci. Instrum.* 56:283–90
82. Lu HSM, Volk M, Kholodenko Y, Gooding E, Hochstrasser RM, et al. 1997. *J. Am. Chem. Soc.* 119:7173–80
83. Volk M, Kholodenko Y, Lu HSM, Gooding E, DeGrado WF, et al. 1997. *J. Phys. Chem.* 101:8607–16
84. Owens GD, Taylor RW, Ridley TV, Margerum DW. 1980. *Anal. Chem.* 52:130–38
85. Nemeth MT, Fogelman KD, Ridley TV, Margerum DW. 1987. *Anal. Chem.* 59:283–91
86. Clark JH, Shapiro SL, Campillo AJ, Winn KR. 1979. *J. Am. Chem. Soc.* 101:746–48
87. Gutman M, Huppert D, Pines E. 1981. *J. Am. Chem. Soc.* 103:3709–13
88. Pines E, Huppert D, Gutman M, Nachliel N, Fishman M. 1986. *J. Phys. Chem.* 90:6366–70
89. Ellis-Davies GCR, Kaplan JH, Barsotti RJ. 1996. *Biophys. J.* 70:1006–16
90. Jordan T, Eads JC, Spiro TG. 1995. *Protein Sci.* 4:716–28
91. Holtz JSW, Bormett RW, Chi ZH, Cho NJ, Chen XG, et al. 1996. *Appl. Spectrosc.* 50:1459–68
92. Schwarz G, Seelig J. 1968. *Biopolymers* 6:1263–77
93. Barksdale AD, Stuehr JE. 1972. *J. Am. Chem. Soc.* 94:3334–38
94. Gruenewald B, Nicola CU, Lustig A, Schwarz G, Klump H. 1979. *Biophys. Chem.* 9:137–47
95. Hammes GG, Roberts PB. 1969. *J. Am. Chem. Soc.* 91:1812–16
96. Schwarz G. 1965. *J. Mol. Biol.* 11:64–77
97. Bosterling B, Engel J. 1979. *Biophys. Chem.* 9:201–9
98. Hamori E, Scheraga HA. 1967. *J. Phys. Chem.* 71:4147–50
99. Lumry R, Legare R, Miller WG. 1964. *Biopolymers* 2:489–500
100. Cummings AL, Eyring EM. 1975. *Biopolymers* 14:2107–14
101. Sano T, Yasunaga T. 1980. *Biophys. Chem.* 11:377–86
102. Tsuji Y, Yasunaga T, Sano T, Ushio H. 1976. *J. Am. Chem. Soc.* 98:813–18
103. Scholtz JM, Baldwin RL. 1992. *Annu. Rev. Biophys. Biomol. Struct.* 21:95–118
104. Gilmanshin R, Williams S, Callender RH, Woodruff WH, Dyer RB. 1997. *Biochemistry* 36:15006–12
105. Munoz V, Thompson PA, Hofrichter J, Eaton WA. 1997. *Nature* 390:196–99
106. Lockhart DJ, Kim PS. 1992. *Science* 257:947–51
107. Lockhart DJ, Kim PS. 1993. *Science* 260:198–202
- 107a. Dyer RB, Williams S. Unpublished results
- 107b. Gilmanshin R, Callender RH, Dyer RB. 1998. *Nat. Struct. Biol.* In press
108. Evans SV, Brayer GD. 1988. *J. Biol. Chem.* 263:4263–68
109. Barrick D, Baldwin RL. 1993. *Protein Sci.* 2:869–76
110. Cocco MJ, Lecomte J TJ. 1994. *Protein Sci.* 3:267–81
111. Johnson RS, Walsh KA. 1994. *Protein Sci.* 3:2411–18
112. Lin L, Pinker RJ, Forde K, Rose GD, Kallenbach NR. 1994. *Nat. Struct. Biol.* 1:447–52
113. Kay MS, Baldwin RL. 1996. *Nat. Struct. Biol.* 3:439–45
114. Jennings PA, Wright PE. 1993. *Science* 262:892–96
115. Eliezer D, Jennings PA, Wright PE, Doniach S, Hodgson KO, et al. 1995. *Science* 270:487–88
116. Shin HC, Merutka G, Waltho JP, Tennant LL, Dyson HJ, et al. 1993. *Biochemistry* 32:6356–64
117. Sabelko J, Ervin J, Gruebele M. 1998. *J. Phys. Chem.* Submitted
118. Gruebele M, Sabelko J, Ballew R, Ervin J. 1998. *Acc. Chem. Res.* In press
119. Pastor RW, Zwanzig R, Szabo A. 1996. *J. Chem. Phys.* 105:3878–82
120. Yeh SR, Takahashi S, Fan B, Rousseau DL. 1997. *Nat. Struct. Biol.* 4:51–56
121. Chan CK, Hofrichter J, Eaton WA. 1996. *Science* 274:628–29
122. Winkler JR, Gray HB. 1996. *Science* 274:629
123. Nolting B, Golbik R, Fersht AR. 1995. *Proc. Natl. Acad. Sci. USA* 92:10668–72

124. Nolting B, Golbik R, Neira JL, Soler-Gonzalez AS, Schreiber G, Fersht AR. 1997. *Proc. Natl. Acad. Sci. USA* 94: 826–30
125. Lubienski MJ, Bycroft M, Freund SMV, Fersht AR. 1994. *Biochemistry* 33:8866–77
126. Freund SM, Wong KB, Fersht AR. 1996. *Proc. Natl. Acad. Sci. USA* 93:10600–3
127. Wong KB, Freund SM, Fersht AR. 1996. *J. Mol. Biol.* 259:805–18
128. Fersht AR. 1995. *Proc. Natl. Acad. Sci. USA* 92:10869–73
129. Wittung-Stafshede P, Gray HB, Winkler JR. 1997. *J. Am. Chem. Soc.* 119:9562–63
130. Chen EF, Lapko VN, Song PS, Kliger DS. 1997. *Biochemistry* 36:4903–8