Annu. Rev. Phys. Chem. 1996. 47:369–95
Copyright © 1996 by Annual Reviews Inc. All rights reserved

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

INSIGHTS INTO PROTEIN FOLDING FROM NMR

H. Jane Dyson and Peter E. Wright
The Scripps Research Institute, La Jolla, California 92037

KEY WORDS: protein folding intermediates, folding kinetics, unfolded proteins, peptide folding, NMR hydrogen exchange

ABSTRACT

NMR has emerged as an important tool for studies of protein folding because of the unique structural insights it can provide into many aspects of the folding process. Applications include measurements of kinetic folding events and structural characterization of folding intermediates, partly folded states, and unfolded states. Kinetic information on a time scale of milliseconds or longer can be obtained by real-time NMR experiments and by quench-flow hydrogen-exchange pulse labeling. Although NMR cannot provide direct information on the very rapid processes occurring during the earliest stages of protein folding, studies of isolated peptide fragments provide insights into likely protein folding initiation events. Multidimensional NMR techniques are providing new information on the structure and dynamics of protein folding intermediates and both partly folded and unfolded states.

INTRODUCTION

There is currently intense interest in the mechanisms by which proteins fold into the correct three-dimensional (3D) structures required for biological activity. In the past few years, many new experimental and theoretical approaches to the protein-folding problem have been employed, leading to significant advances in our understanding of the fundamental molecular processes involved. NMR has emerged as an especially important tool for studies of protein folding because of the unique structural insights it can provide into many aspects of the folding process. Applications range from direct or indirect characterization of kinetic folding events to structural studies of equilibrium folding intermediates, partly folded states, peptide fragments, and unfolded or denatured proteins. The

0066-426X/96/1101-0369$08.00

diverse applications of NMR in protein-folding studies are illustrated in Figure 1. Future goals include the determination of solution structures of folding intermediates and the structural characterization of folding pathways. In the present article, we review these applications and the insights they provide into mechanisms of protein folding. The impact of NMR on the protein-folding field has been so dramatic and widespread that an exhaustive review is not possible here. Instead, we focus on papers that illustrate the principal applications of the NMR method, with emphasis on more recent work. Recent general reviews on the subject of protein folding include References 1–4; reviews of applications of NMR to study protein folding include References 5–7.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

DIRECT NMR STUDIES OF FOLDING AND UNFOLDING PROCESSES

The use of NMR to monitor protein folding (or unfolding) events directly, i.e. in real time, is difficult because the time scale of the NMR experiment is slow compared to the rate of folding of most proteins. However, in favorable cases, where

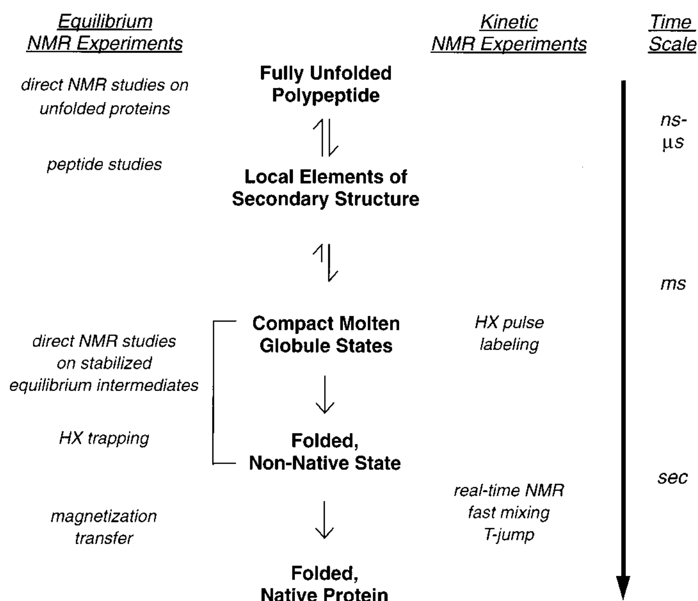

Figure 1 Schematic diagram of the folding process showing the time scale for various stages and the information that can be obtained from NMR experiments at each stage.

folding is sufficiently slow, NMR can be used to obtain valuable information. The methods that have been used vary greatly in their complexity, from simple manual mixing experiments to temperature-jump and stopped-flow NMR.

Manual Mixing Methods

Slow conformational changes can be followed directly by recording a series of NMR spectra, usually one-dimensional (1D) spectra, immediately after initiation of the folding or unfolding reaction by manual mixing to change the concentration of denaturant. This method has been used to follow slow folding events, usually associated with proline isomerization, in several proteins. For French bean apoplastocyanin, for example, folding takes place on a time scale of several hours and involves the *cis-trans* isomerization of two proline residues (8). The folding process is slow enough that it can be monitored in real time by recording 1D NMR spectra following dilution of guanidine hydrochloride denaturant by manual mixing. In this case, it was possible to measure the kinetics of the folding reaction and also to record the ^{1}H NMR spectrum of a folding intermediate in which one of the prolines remains in the incorrect *trans* configuration.

A manual mixing experiment was recently used to follow the unfolding reaction of ribonuclease A by ^{1}H NMR and revealed the presence of a previously unsuspected intermediate (9). To follow this process, it was necessary to choose conditions under which unfolding was sufficiently slow to allow recording of 1D NMR spectra. The evidence for the intermediate is indirect; its presence is inferred from the observed intensities of resolved resonances compared with the intensities expected from a consideration of the kinetics of folding obtained from circular dichroism measurements. The authors concluded that the intermediate is a "dry molten globule," in which side chains are free to rotate but water has not yet penetrated the hydrophobic core.

A transient intermediate was detected during refolding of *Escherichia coli* thioredoxin (10). This example is of interest because selective deuteration was used to simplify the aromatic region of the ^{1}H NMR spectrum; the intermediate was detected by observation of splitting of phenylalanine resonances in a spectrum recorded during refolding of the deuterated protein and could not be identified by using 1D ^{1}H NMR on unlabeled protein.

Kinetics of Folding and Unfolding at Equilibrium

Under favorable conditions, protein folding and unfolding rates can be measured by magnetization transfer between resonances arising from folded and unfolded states that interconvert slowly on the chemical shift time scale. The method, which has been reviewed elsewhere (5, 6), has been applied to several proteins, including bovine pancreatic trypsin inhibitor (BPTI) (5), lysozyme (11), and staphylococcal nuclease (12), for which interconversion between folded and

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

unfolded forms is on an appropriate time scale. When interconversion rates are sufficiently fast, exchange-broadened resonances are observed. Although most proteins fold too slowly to give rise to exchange broadening, such effects have recently been observed for a monomeric form of the phage λ repressor that folds on a submillisecond time scale (13). Rate constants for the folding and unfolding processes (3600 s^{-1} and 27 s^{-1}, respectively, at 0 M urea and 37°C) were derived by using computer simulation of the line shapes. The urea-dependence of the rate constants suggests that the transition state may be more solvent exposed than for more slowly folding proteins. Because the resonances of protons distributed throughout the molecule reflect identical denaturation behavior, Huang & Oas suggested that the folding of this protein is two-state, with no populated intermediates on the folding pathway (14). Dynamic NMR line-broadening effects have also been used to study rat intestinal fatty acid binding protein by ^{19}F NMR, where a highly populated folding intermediate was observed (15).

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

Temperature-Jump Methods

Temperature-jump methods are useful for studying kinetic intermediates in proteins that undergo reversible thermal unfolding. Temperature-jump was first used by Baldwin and coworkers (16) to follow the refolding of thermally denatured ribonuclease A; samples were rapidly cooled and then manually transferred to the NMR spectrometer, where spectra were recorded over a time scale of several minutes. The method has recently been refined for studies of staphylococcal nuclease refolding kinetics (17); following a rapid thermal quench, samples were transferred to the spectrometer, and single transient ^{1}H NMR spectra were recorded at 2.5-s intervals until folding was complete. The dead time associated with the quench and sample transfer was only 8–12 s. By monitoring time-dependent changes in histidine proton resonances, a transient intermediate with an incorrect *trans* proline peptide bond was detected. Interestingly, the rate of isomerization of this peptide bond was found to be ∼ 100-fold faster than that observed in a model peptide. Use of an NMR temperature-jump apparatus in which the temperature of the protein sample is changed rapidly within the NMR spectrometer has also been reported (18). An advantage of this method is that signals can be averaged over several heating and cooling cycles. Finally, Adler & Scheraga have described a continuous recycled flow method for studying refolding of thermally denatured proteins and used it to study a ribonuclease A folding intermediate (19). Although only 1D ^{1}H spectra were reported, extensive signal averaging is possible, and this method could in principle be used to record two-dimensional (2D) NMR spectra of refolding intermediates.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

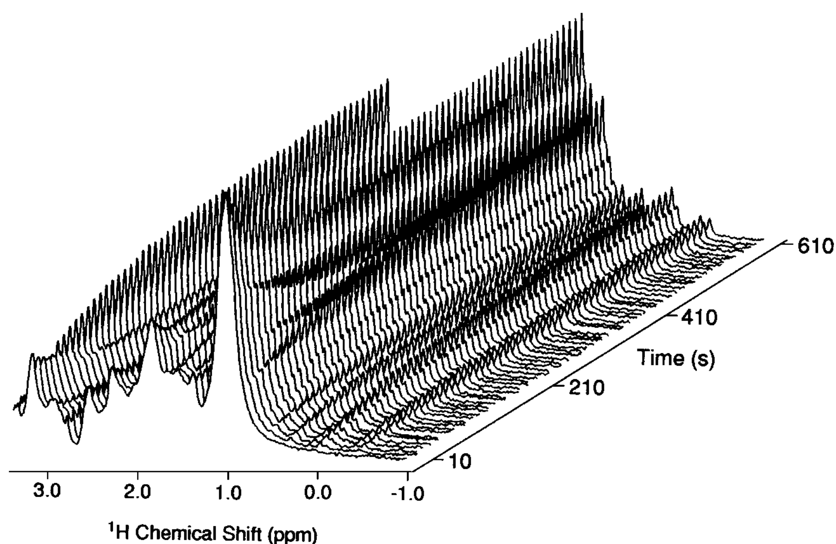

Figure 2 Stacked plot of 600-MHz 1D NMR spectra of bovine α-lactalbumin in ^2H_2O at 20°C. Spectra were recorded at time increments between 1.2 s and 10.3 min after initiation of refolding from 6 M guanidinium DCl, in 30-mM Tris DCl buffer, pH 7.2, 2 mM EDTA. For a better representation of the complete refolding kinetics, each spectrum corresponds to an average of over 8 scans (9.7 s). The region shown contains resonances from methyl and methylene protons. [Adapted with permission from Balbach et al (20).]

Stopped-Flow Methods

Few applications of NMR stopped-flow or rapid mixing devices for studies of protein folding have been reported. Frieden et al (6) described stopped-flow ^{19}F NMR experiments that monitored the unfolding of intestinal fatty acid binding protein, labeled with 6-fluorotryptophan, following mixing with urea solution. Because of the relatively long T_1 (longitudinal relaxation time) of ^{19}F, spectra could be collected at only 2-s intervals. In an important recent development, a rapid mixing NMR experiment has been utilized to follow the refolding of apo bovine α-lactalbumin (20). In this experiment, unfolded protein was rapidly injected into refolding buffer in the NMR tube and, after a dead time of 1.2 s, a continuous series of 1D ^{1}H NMR spectra were recorded (Figure 2). Spectra could thus be obtained for a transient folding intermediate that resembles the well-characterized molten globule state formed at low pH. This rapid mixing experiment could potentially be combined with heteronuclear 2D NMR experiments.

INDIRECT NMR STUDIES OF PROTEIN FOLDING PATHWAYS

NMR Hydrogen-Exchange Pulse Labeling

An important advance in the structural characterization of protein folding pathways came with the realization that a combination of quenched-flow methods, hydrogen-deuterium exchange, and NMR could give detailed site-specific information about the folding process (for reviews, see 5, 7, 21). The method was suggested by Roder & Wüthrich (22) as an extension of the ^{1}H-^{3}H exchange methods developed by Baldwin and coworkers (23), and its power was demonstrated in 1988 with the simultaneous publication of two folding studies, one on ribonuclease A (24) and one on cytochrome c (25). Since then, several proteins have been studied by the hydrogen-exchange pulse-labeling method, and representative examples are discussed below.

The method aims to label intermediates on the folding pathway in such a way that they can be studied at leisure by NMR once folding is complete. There are many advantages of working with native folded proteins, including greater stability to aggregation at NMR concentrations, better dispersion of resonances in the NMR spectra, and usually, more readily available resonance assignments. When fast folding events are to be monitored, the pulse-labeling technique requires the use of a temperature-controlled multisyringe quenched-flow apparatus, obtainable from several commercial sources. For a typical refolding experiment, the protein can be pulse-labeled either with $^{1}H_2O$ from a $^{2}H_2O$ buffer (24–26) or with $^{2}H_2O$ from a $^{1}H_2O$ buffer (27). The experiment is shown in schematic form in Figure 3. Denatured protein is rapidly diluted with refolding buffer and allowed to refold for a variable time t_f determined by the length of the delay line. The lower limit of the refolding time (dead time) is determined by the mixing characteristics of the quenched-flow apparatus and is usually 4–6 ms. The partially refolded protein is pulse-labeled for a length of time t_p (usually 30–50 ms) under conditions (high pH) in which the intrinsic peptide hydrogen-exchange time constant is much shorter than the labeling pulse. The amide protons (or deuterons) in unstructured parts of the protein are exchanged during the pulse, while those involved in hydrogen-bonded secondary structure are protected from exchange. The labeling pulse is quenched by lowering the pH to conditions in which amide proton exchange is slow, and the folding process is allowed to proceed to completion. 2D NMR spectra of the refolded protein are then recorded in $^{2}H_2O$, and the proton occupancy is measured at each refolding time for the observable amide proton resonances. A limitation of the method is that only amide protons that exchange slowly in the native protein can be used as probes.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

Further, the method does not directly identify the hydrogen bond acceptor group.

Quench-flow hydrogen-exchange studies of the folding of a number of proteins have now been reported: ribonuclease A (24), cytochrome c (25), barnase (28), ubiquitin (29), phage T4 lysozyme (30), hen lysozyme (31), interleukin (IL)-1β (32), ribonuclease T_1 (33), staphylococcal nuclease (34), apomyoglobin (27), the immunoglobulin binding domain of protein G (35), acyl CoA binding protein (36), dihydrofolate reductase (37), and λ repressor (13). Little similarity has yet been revealed in these studies between the folding mechanisms of proteins in different structural families. For ribonuclease A (disulfide bonds intact) (24, 26, 38), one early-folding intermediate is formed rapidly and contains all of the slowly exchanging amide protons in the β-sheet. These protons are only marginally protected early in the folding pathway, but they become more protected at later stages, an indication that subsequent events stabilize the formation of the sheet. There are only a few probes (i.e. only a few slowly exchanging amide protons in the native folded state) in the three helices, so information about these regions is scanty. An analysis of the results in terms of a sequential model for folding showed that the intermediate contains the majority of the β-sheet

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

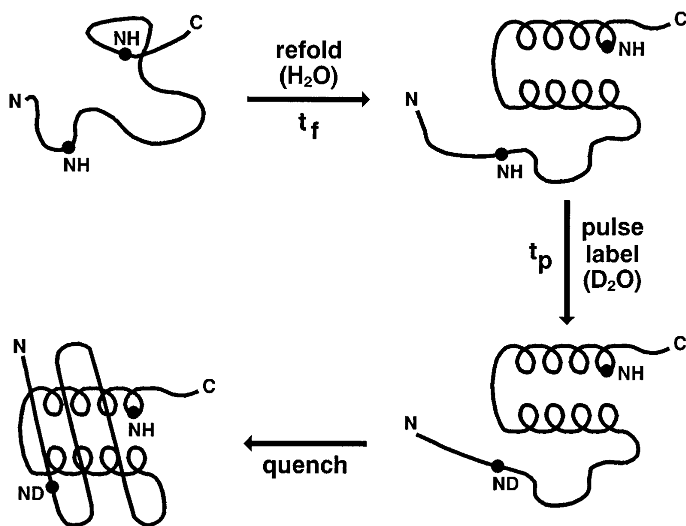

Figure 3 Schematic diagram showing the state of the protein at various stages in the folding and labeling process in a quenched-flow hydrogen-exchange experiment. [Reproduced with permission from Dyson & Wright (143).]

secondary structure of ribonuclease A but is not fully populated when first formed (26).

The structure of cytochrome c is quite different from that of ribonuclease A and consists largely of helix, with little or no β-sheet. Unlike ribonuclease A, the helices have many amide proton probes, so this molecule provides a good system for studying the folding of a helical protein (25). Several amide protons in the N- and C-terminal helices are apparently protected within the 4 ms dead time of the quench-flow apparatus. This is an indication that folding is initiated very rapidly in these regions of the protein, probably by rapid formation of helical secondary structure, followed by or perhaps concomitant with stabilization of the initial folded form by docking of the two helices. This behavior is apparently different from that of the α/β protein ribonuclease A, although it is difficult to determine whether the helices are in fact formed early in the latter protein, owing to the paucity of probes in the helices. The involvement of the covalently attached heme and its axial ligands is crucial for the initiation of the folding process in cytochrome c (39, 40). A peptide model of the cytochrome c folding intermediate showed enhanced helicity when the two component peptides were combined (41), indicating that interactions between marginally stable elements of secondary structure are important in forming tertiary subdomains (intermediates) in the folding of the intact protein.

The folding pathways of ribonuclease A and cytochrome c appear to be rather simple, consistent with the small size of the proteins and their compact single-domain structure. Hen lysozyme shows more complex folding behavior, giving evidence of multiple pathways associated with two folding domains (31, 42); the presence of two folding pathways has been confirmed by mass spectrometry (43). NMR experiments based on competition between hydrogen exchange and the refolding process (42) showed that the rapid and slow refolding populations of lysozyme correspond to the helical and sheet regions of the molecule, respectively, indicating that the two domains fold independently. In addition, different populations of molecules fold by kinetically distinct parallel pathways, some of which may involve reorganization of incorrectly folded structure (31). Parallel-folding pathways have also been invoked to explain the results obtained for the folding of ribonuclease T_1 (33), which folds much more slowly than most other proteins studied to date by the quenched-flow hydrogen-exchange method. Two intermediates were postulated on the kinetic folding pathway of ribonuclease T_1, a native-like intermediate and a molten globule (33).

Molten globule intermediates have been suggested for many of the protein folding pathways studied so far, although the definition of what constitutes a molten globule sometimes appears to vary with the protein studied. Intermediates of this kind have been postulated for T4 lysozyme (44), ubiquitin (29),

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

barnase (28), and staphylococcal nuclease (34). For all of these $\alpha + \beta$ proteins, the majority of the stable backbone hydrogen bonds formed in the initial folding events (within a few milliseconds) belong to the β-sheets. In contrast, the α/β protein dihydrofolate reductase shows early folding of part of the β-sheet, together with part of the α-helical domain in close contact with one end of the sheet: A molten globule is not explicitly invoked for this molecule (37). Refolding of the all-β protein IL-1β also leads to rapid formation (within 25 ms) of an intermediate containing 90% of the β-sheet secondary structure but without stable hydrogen bonds (32). Formation of stable, hydrogen-bonded secondary structure begins only after ~ 1 s. Another all-β protein, plastocyanin, displays similar behavior: Hydrogen-exchange competition experiments suggest that a folding intermediate containing unstable β-sheet structure is formed, which provides only slight protection of backbone amide protons from exchange (8). Isomerization of a key proline residue, Pro 16, to the *cis* form allows correct packing of the incipient, fluxional β-sheets to form stable hydrogen-bonded secondary structure during the final slow folding step.

Although the evidence that molten globule intermediates of many types participate in the folding pathways of a number of proteins is persuasive, in none of the systems described so far was the intermediate demonstrated to participate in the actual kinetic folding pathway of the protein. Such a demonstration has recently been made for apomyoglobin. Folding of apomyoglobin (without the heme prosthetic group) has been studied extensively using both theoretical and experimental techniques. Circular dichroism (CD) spectroscopic measurements as a function of pH (45) indicated that a stable intermediate species exists at pH $\sim$ 4. The structure of this species has been characterized by amide-exchange trapping (46; see below), which showed that a region of the protein is folded in the intermediate, while the remainder of the protein is flexible or unfolded. The folded region consists of the two C-terminal helices, termed the G- and H-helices, which form a helical hairpin structure in the fully folded myoglobin and also probably in apomyoglobin (47, 48), together with the N-terminal helix, termed the A-helix. These regions are shown mapped onto the structure of native myoglobin in Figure 4. The question as to whether this intermediate is formed during the folding of apomyoglobin was addressed by Jennings & Wright (27). Using both hydrogen-exchange pulse-labeling and stopped-flow CD measurements, they showed that in less than 5 ms an intermediate was formed in which amide proton probes were fully protected in the A-, G-, and H-helical regions of the protein and a part of the B-helix. These regions are precisely the same as those previously identified as folded in the equilibrium intermediate observed at pH $\sim$ 4. Amide protons in the remainder of the B-helix, the C- and E-helices, and the CD connecting loop were protected on a slower time scale. A close structural similarity clearly exists between the

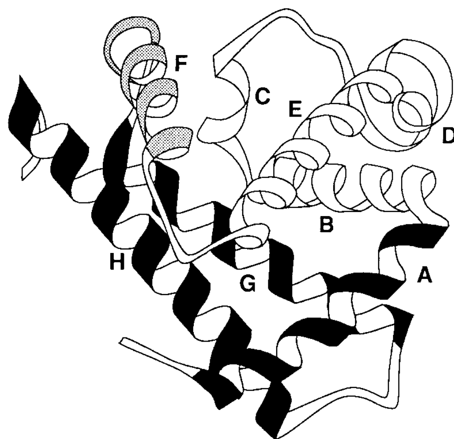

Figure 4 The location of the structured regions found in a folding intermediate of apomyoglobin mapped onto the X-ray structure (144) of the native protein (the heme group is omitted from the diagram for the sake of clarity). Helices B to E are unshaded and are proposed to be predominantly unfolded in the intermediate. The F-helix is lightly shaded to denote the absence of amide proton probes and structural information. The A-, G-, and H-helices are darkly shaded and are proposed to be folded into a native-like structure in the intermediate. [Reproduced with permission from Hughson et al (46).]

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

earliest detectable kinetic intermediate and the intermediate observed under equilibrium conditions (46); this similarity provides strong evidence that the molten globule intermediate participates in the kinetic folding pathway.

NMR STUDIES OF UNFOLDED PROTEINS

Structural Characterization of Unfolded and Denatured Proteins

NMR has been used to obtain important insights into the nature of unfolded or denatured states of proteins.[1] Because 1D NMR spectra of denatured proteins lack dispersion and resemble spectra of mixtures of free amino acids (49), the assumption has been that the denatured state comprises a random coil. However, subtle deviations from random coil spectra have been observed for many proteins, e.g. thermally denatured ribonuclease (50) or urea-denatured tryptophan synthase α subunit (51), indicating some residual structure in the

[1]Two apparently synonymous terms are used to differentiate between proteins that are chemically or thermally denatured and those that are unfolded in aqueous solution in the absence of denaturants.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

denatured state but providing little information as to its nature. It was not until the development and application of multidimensional NMR technology that detailed structural insights into denatured states were forthcoming.

The major problem to be faced in applying NMR to studies of denatured proteins is the lack of spectral dispersion, especially for ^{1}H resonances, and the consequent difficulty in making sequence-specific resonance assignments. Initial approaches utilized homonuclear (^{1}H) magnetization transfer methods to correlate resonances in the spectrum of the folded protein, which could be assigned using conventional sequential assignment methods (52), with the corresponding resonances in the denatured state. This approach has been applied successfully to obtain partial ^{1}H resonance assignments in thermally denatured BPTI (5), ribonuclease A (53), and lysozyme (54), thus providing evidence for residual hydrophobic clusters in the denatured state. Magnetization transfer has also been used to assign resonances in a urea-unfolded fragment of phage 434 repressor (55, 56). From NOE measurements in 7 M urea, specific hydrophobic interactions were identified, and the local structure of the polypeptide chain was calculated by using distance geometry methods (57). The residual structure observed in the urea-denatured protein resembled that in the corresponding region of the native folded protein, prompting the hypothesis that formation of this hydrophobic cluster could serve as an initiation site for folding of the 434 repressor.

Denatured proteins have also been studied by measuring amide proton-exchange rates. For heat-denatured BPTI (58) and ribonuclease A (59) and urea-denatured lysozyme (60), exchange rates are similar to those expected for a random coil, despite the evidence for residual structure from circular dichroism and ^{1}H NMR experiments. These observations suggest the absence of a stable hydrogen-bonded structure in the denatured states of these proteins.

A major advance in the characterization of denatured states came through use of uniform isotope labeling with ^{13}C and/or ^{15}N and the application of multidimensional heteronuclear NMR experiments. The ^{15}N chemical shift dispersion remains large in denatured proteins (56, 61), since it is influenced by both residue type and large sequence-dependent effects (62). Consequently, ^{15}N-dispersed spectra are well resolved and facilitate direct NMR characterization of denatured proteins. Heteronuclear NMR methods have now been applied to characterize unfolded states of several proteins by using 2D or 3D ^{1}H-^{15}N correlation experiments with ^{15}N-labeled protein (56, 63) or 3D triple resonance experiments with ^{13}C, ^{15}N double-labeled protein (64–67). Triple-resonance methods are applicable to larger proteins than are the simpler ^{1}H-^{15}N correlation experiments. This technology was first applied to the FK506-binding protein (107 amino acids), unfolded in both urea and guanidine hydrochloride

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

solutions, and led to complete ^{1}H, ^{13}C, and ^{15}N resonance assignments (64). Triple resonance methods have since been used successfully to obtain resonance assignments for a denatured 131-residue fragment of staphylococcal nuclease (66), for acid-unfolded barnase (65), and for the urea-denatured state of the immunoglobulin-binding domain of streptococcal protein G (67). In favorable cases, 2D homonuclear (^{1}H-^{1}H) NMR spectroscopy may be adequate for assigning resonances of unfolded states (68).

In all cases studied to date, coupling constants and chemical shifts indicate extensive conformational averaging in the denatured state. However, deviations from random coil chemical shifts and/or the existence of medium-range NOEs between protons on residues separated by 2–3 amino acids in the sequence, indicate the persistence of local structured regions in the denatured states of the 434 repressor (56), FK506-binding protein (64), barnase (65), the staphylococcal nuclease fragment (66), an SH3 domain (69), and reduced and unfolded BPTI (68). The residual structure involves both local clustering of hydrophobic residues and fluctuating elements of secondary structure (turns and nascent helix). Subtle differences in the location of the residual structure have been observed for different denaturants, urea or guanidine hydrochloride (64). Interestingly, for both an SH3 domain (69) and reduced BPTI (68), the unfolded forms observed in aqueous solution under nondenaturing conditions are significantly more structured than those obtained by denaturation with guanidine hydrochloride.

Polypeptide Dynamics of Unfolded States

^{15}N spin relaxation measurements have been used to characterize the backbone dynamics in several partially folded or unfolded proteins (67, 70–72). Interpretation of ^{15}N T_1 and T_2 relaxation rates and the ^{1}H-^{15}N NOE in terms of dynamical parameters is not straightforward for unfolded proteins because the usual assumption of isotropic tumbling with a single rotational correlation time is unlikely to be valid. Two approaches have been reported. For the disordered 131-residue fragment of staphylococcal nuclease (71) and the urea-denatured protein G domain (67), interpretation of the relaxation data was based on the model-free analysis by Lipari & Szabo (73, 74). In this approach, the polypeptide dynamics are described by three parameters: an effective correlation time (τ_m) for the whole molecule and a generalized order parameter (S^2) and internal correlation time (τ_e) that describe the amplitude and time scale, respectively, of the internal motion at individual amino acid residues. For the staphylococcal nuclease fragment, which was studied in aqueous solution under nondenaturing conditions, a wide range of S^2 values was observed. The order parameters provide evidence for restricted motions in local regions of the polypeptide chain and are inconsistent with a random coil conformation (71). S^2 values were also

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

found to correlate with the hydrophobicity of the sequence, suggesting a relationship between backbone flexibility and local propensities for hydrophobic collapse. It was further suggested that secondary structure may be preferentially stabilized in hydrophobic segments of the polypeptide. In contrast to the staphylococcal nuclease fragment, the measured order parameters are relatively uniform for the urea-denatured protein G, except at the N- and C-termini (67). The chemical shifts, coupling constants, and ^{1}H-^{1}H NOEs give no indications of residual structure. Taken together, the data suggest that protein G unfolded in urea forms a rather compact random coil, with relatively uniform motional properties on the subnanosecond time scale. Despite the uniformity of S^2, reductions in the magnitude of T_2 are observed that suggest restriction of motions (on a time scale longer than the effective correlation time for molecular tumbling) in local regions of the unfolded polypeptide chain that correspond to a β-hairpin and helix in the native folded protein.

A different approach has been taken for analysis of ^{15}N relaxation measurements on an unfolded SH3 domain, in aqueous solution under nondenaturing conditions (72). In this case, values of the spectral density function of 12 backbone N-H vectors were determined directly from the relaxation data; this avoids the need for simplifying assumptions about isotropic tumbling and a single overall rotational correlation time. Considerable variation was observed in the values of the spectral densities at different sites in the polypeptide chain, indicating diverse dynamic behavior in the unfolded state. Analysis of the same relaxation data by the model-free approach (73, 74) yielded wide variation in τ_m values, indicating that the overall tumbling of this unfolded protein cannot be described by a single correlation time. The spectral densities (and order parameters) suggest that the unfolded SH3 domain is relatively compact, with backbone dynamics similar to those observed for disordered regions in globular proteins.

NMR STUDIES OF PARTIALLY FOLDED STATES

For most proteins, the intermediates formed during the kinetic folding process are populated only transiently and are therefore difficult to study directly by NMR. Although some limited information has been forthcoming from real-time NMR experiments, the indirect hydrogen-exchange pulse-labeling experiment is generally the preferred method for obtaining information on kinetic intermediates. However, in some proteins, and under certain conditions, partially folded states can be stabilized at equilibrium, thereby opening the way to direct NMR analysis. Various hydrogen-exchange methods also provide valuable structural information on such equilibrium intermediates.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

Direct NMR Studies

Partly folded forms of several proteins have been studied directly using NMR. The most extensive studies to date have been on ubiquitin, α-lactalbumin, and BPTI. The partially folded A-state of ubiquitin is formed at pH ~ 2 in 60% methanol/40% water at 298 K, and three separate studies of the same state of the protein (75–77) reach broadly similar conclusions that nevertheless differ considerably in detail, an illustration of the difficulties inherent in these types of measurements. Harding et al (75) used 2D ^{1}H NMR spectroscopy to assign slowly exchanging amide proton resonances in the A-state. From their experiments, they concluded that there was no gross structural reorganization from the native state of the protein but that the A-state contained a subset of the secondary structure found in the native state, including the first two strands of the five-strand β-sheet, part of the third strand, and a partially structured α-helix packed on the hydrophobic side of the sheet. No evidence was found for the reverse turns and 3_{10} helical structure present in the native state of the protein. Pan & Briggs (76) made an extensive study of the protection factors of amide protons in the A-state of ubiquitin by using hydrogen-exchange trapping (see below). They too found a structure with native-like secondary structure, but concluded that almost all of the native secondary structure persists in the intermediate, apparently destabilized by several kilocalories per mole relative to the native state. The different results obtained in these studies were attributed to differences in the method of sample preparation (76). In a third study, which used 3D ^{1}H-^{15}N correlated NMR experiments, Stockman et al (77) were able to assign unambiguously more than 90% of the resonances of the A-state in 60% methanol, and to resolve a number of NOEs that define the limits of secondary structure. Their data indicated that, contrary to the conclusions of the two previous studies, only the first two β-strands are present in the A-state, and suggested that the C-terminus, rather than forming part of the β-sheet as in the native protein, becomes helical in the presence of methanol, which is an interesting parallel to the behavior of monellin (78).

The ^{1}H NMR spectrum of a partially folded state of α-lactalbumin (termed the A-state) formed at acid pH differs substantially from those of both the native and denatured states (79). Chemical shift dispersion in the A-state is limited, indicating that long-range tertiary structure present in the native state is absent in the intermediate (80). No evidence for native-like hydrophobic clustering is observed, but interresidue NOEs indicate that a new aromatic cluster, formed by rearrangement of the side chains of residues at positions 103, 104, and 107, is present. That this structure arises from a local interaction was confirmed by the observation of similar NOEs in a trifluoroethanol solution of a short peptide containing these residues (80). The model that emerges from these and amide

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

trapping experiments is one in which the α-lactalbumin A-state contains stable regions of secondary structure with largely disordered tertiary structure. Direct NMR studies of the α-lactalbumin A-state are hampered by line broadening arising from conformational averaging. To circumvent this problem, Dobson and coworkers have examined a partially folded form in trifluoroethanol (81), which gives rise to narrow resonances. Although this state shows some similarities to the α-lactalbumin molten globule state, a high content of secondary structure in the absence of fixed tertiary interactions, it differs in other properties such as the ability to bind 1-anilinonaphthalene-8-sulfonate. Further, the NMR spectra provide evidence for trifluoroethanol (TFE) induction of helical structure in a region that is not helical in the native folded protein. Clearly, results obtained from such TFE-stabilized states must be viewed with caution, as they might differ significantly from kinetic or equilibrium protein folding intermediates.

The most intensely studied protein, in terms of folding, is probably BPTI. Recently, a number of studies using NMR to probe the structure of the folding intermediates containing only one or two of the three disulfides of the native protein have been performed, using mutant proteins in which pairs of cysteine residues have been replaced by serine (70, 82–84), alanine (85, 86), or α-aminoisobutyric acid (87). The intermediates, which contain native disulfides, generally show many of the characteristics of the folded form of BPTI. However, Cys-Ser or Cys-Ala mutants containing only the disulfide between cysteines at positions 30 and 51 of the sequence (the 30–51 disulfide) appear from NMR studies to be largely unfolded at the N-terminus (82, 86); nevertheless this partially folded intermediate contains the major elements of native-like secondary structure stabilized by packing within the hydrophobic core of the protein. A disulfide-bridged peptide model of the 30–51 disulfide intermediate contains a similar native-like subdomain with the correct secondary structure and tertiary packing (88). The partially folded nature of the 30–51 intermediate explains its key role in the BPTI folding pathway (82, 86) and supports a model for protein folding in which the folding pathway involved sequential formation of native-like intermediates. A model for the 14–38 single-disulfide intermediate is also partially folded but retains an ordered hydrophobic core (87). NMR studies of two-disulfide intermediates, containing the native 30–51 disulfide and a second nonnative disulfide bond, have also been reported (84). These states are partially folded, and the nonnative disulfide bonds do not disrupt the native-like structure formed in the 30–51 single-disulfide intermediate.

In a series of pioneering studies, Jonas and coworkers (89–91) demonstrated that high-pressure, high-resolution NMR techniques hold considerable promise

for stabilization and characterization of partly folded states of proteins. The Arc repressor dimer was observed to denature with increasing pressure via a predissociated state to a molten globule monomer (89, 90). Two-dimensional NOESY spectra showed that the molten globule monomer retains substantial secondary structure. In another application, pressure unfolding of ribonuclease A was investigated, and a state with the characteristics of a molten globule intermediate was identified (91).

Amide-Exchange Trapping

In the amide-exchange trapping experiment, an intermediate state of a protein is formed in H_2O solution, and following dilution into ^2H_2O, amide proton exchange is allowed to proceed. The solution conditions are then altered to allow the protein to refold to its native state, and NMR spectra are recorded to identify amide protons trapped in stable secondary structure in the intermediate state. By varying the exchange time, amide-exchange rates and protection factors can be measured for the intermediate. As for hydrogen-exchange pulse labeling, the method only provides information on sites where amide proton exchange is slow in the native protein. Also, the hydrogen bond acceptor in the intermediate is not identified, although it can often be inferred by comparing patterns of protection in the intermediate with those in the native protein (92). The amide-exchange trapping method was first used to investigate the low-pH molten globule of apomyoglobin (46). Amide protons were found to be protected in the A-, G-, and H-helix regions (Figure 4) but were not protected in other regions of the molecule. This led to the conclusion that the acid intermediate of apomyoglobin contains stable, native-like helical secondary structure within a compact subdomain, while the remainder of the protein is apparently unfolded. As mentioned above, the low-pH intermediate appears to correspond closely to the first intermediate detected on the kinetic folding pathway of apomyoglobin (27). More recently, an intermediate in which a fourth helix, the B-helix, participates in the molten globule has been identified by amide-exchange trapping in the presence of trichloroacetate (93). This intermediate appears to correspond to the second kinetic intermediate found on the folding pathway (27).

An extensive study of cytochrome c using hydrogen-exchange trapping has been made (94–96). It appears that native-like secondary structure is formed in the cytochrome c–folding intermediates (termed submolecular folding units) that are formed at low pH in high salt. On the basis of a comparison of the hydrogen-exchange behavior of the helices in the molten globule state of cytochrome c with those expected for the isolated helical segments, it has been concluded that loose (nonnative) interactions between helices are present in the molten globule (96).

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

The A-state of α-lactalbumin has been characterized by amide-exchange trapping (97). Residual secondary structure present in the A-state (79) is thought to be native-like, and to consist of two helices stabilized by packing via hydrophobic interactions in a compact intermediate state. The highly protected nature of this hydrophobic region is shown by the high degree of protection of an indole NH from a tryptophan residue in the A-state (97).

The amide-exchange trapping method has been applied to proteins denatured by alcohols, monellin in ethanol or trifluoroethanol (78) and lysozyme in trifluoroethanol (98). The characteristic feature of these alcohol-denatured states appears to be a higher content of helical secondary structure than found in the native state. In the case of monellin, the experiment showed that a local region had apparently changed from β-sheet secondary structure in the native protein to α-helix in the alcohol solution (78). In contrast, the protected amides in alcohol-denatured lysozyme were present in regions that were helical in the native folded state, and little protection was found for the sheet and loop regions of the native protein (98).

Detection of Intermediates by Using Hydrogen Exchange in Denaturant

Variation of amide-exchange rates as a function of denaturant concentration has been proposed as a method for identification of folding intermediates (99, 100). In the case of ribonuclease A, it was found that in addition to the major unfolding events at higher guanidine hydrochloride concentrations, partial unfolding occurs at low concentrations of denaturant, giving rise to a second mechanism of exchange that is relatively independent of guanidine hydrochloride concentration (99). A similar method was applied to cytochrome c (100); the presence of partially folded intermediates was inferred from plots of thermodynamic parameters derived from hydrogen-exchange rates as a function of denaturant concentration. The authors conclude that these partially unfolded forms may define the major pathway for cytochrome c folding.

NMR OF PEPTIDE FRAGMENTS OF PROTEINS

The power of NMR methods in the study of protein folding is emphasized by the quantity and variety of research reported to date and by the continuing development of new techniques, as briefly described above. NMR is able to describe folding intermediates and processes in a site-specific manner that cannot be rivaled by any other technique, but is presently limited to a time scale of milliseconds or longer. For most proteins, some steps in the folding process may be completed on a shorter time scale; in apomyoglobin, for example, the first intermediate is already fully formed within the dead time (~ 5 ms) of

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

stopped-flow or quenched-flow experiments (27). Information on the earliest events in the folding process is thus not accessible to NMR studies of the whole protein. Wright et al (101) have proposed an approach to characterize the probable initiation events in protein folding in which the protein is dissected into peptide fragments and NMR methods are used to identify elements of secondary structure or hydrophobic clustering in aqueous solution. The majority of linear peptides are unstructured in aqueous solution, that is, they are present as a conformational ensemble containing conformers with widely different structures. For some peptides, a conformational preference for a folded form can be observed. However, it should be stressed firmly that, in the absence of covalent conformational constraints, all linear peptides must be considered to consist of conformational ensembles in solution and the NMR data interpreted accordingly.

The S-peptide of ribonuclease A was the first peptide for which a conformational preference, for helix, was observed by CD spectroscopy in aqueous solution (102). One-dimensional NMR was later used to demonstrate the formation and dissolution of secondary structure in the S-peptide (103) and in three-residue peptides from other regions of ribonuclease A (104). Turn-like conformations were inferred from the latter data; their existence is consistent with the hypothesis that the formation of β-turns is important in the initiation of folding of ribonuclease A. A conformational preference for a well-defined turn was found by using 2D NMR methods in a nine-residue peptide derived from influenza virus hemagglutinin (105). Since then, many studies of peptide fragments of proteins have been reported, a selection of which is summarized below. Several reviews have recently appeared on this subject, including a summary of the techniques used for identifying structured states in the conformational ensemble (106). The connection between the NMR observation of structured conformers in peptide fragments of proteins and the initiation of protein folding was initially suggested in 1988 (101); since that time accumulated evidence has established that peptides can be used as models for the initiation of protein folding.

Conformational Preferences Observed in Peptides: A Model for Folding Initiation?

Reverse turns have been identified by NMR methods in a number of peptide fragments and, more recently, in several unfolded proteins. Turn-like structures may play an important role in the earliest steps involved in initiation of protein folding (34, 101, 107) and have recently been shown by mutagenesis to be crucial to the correct folding of plastocyanin (108). Although much is known of the sequence preferences of turn conformations in folded proteins

(109), the factors that stabilize turns in peptides and unfolded proteins are less well understood. An extensive study of the sequence preferences of turns was made in several series of five-residue peptides derived from the peptide YPGDV (110). A correlation has been observed between the turn propensity of the peptides in aqueous solution and the empirical sequence preferences obtained from protein structures (110), an important link between the behavior of peptides in solution and structure in proteins. This correlation indicates that interactions specified by the local amino acid sequence, rather than medium-to-long-range interactions in the folded protein, are often sufficient to determine the reverse turn conformation. Such sequences are excellent candidates for protein folding initiation sites. However, local elements of structure observed in peptides or unfolded proteins need not necessarily be present in the final folded protein, even if they play a role in initiation of protein folding. Local structures formed during the earliest initiation events are necessarily unstable and, despite their importance in directing chain folding, could easily be eliminated from the final folded structure by subsequent folding events that make other conformations energetically more favorable.

Type VI turn conformations have been observed at high population in peptides containing *cis*-proline (110). The turn conformation is highly favored when there are aromatic residues on either side of the *cis*-proline and when Asp is present at position 5, following the turn (111). A 3D structure of the folded form of one of the peptides has been calculated from NMR data (112) and reveals a specific stacking interaction between the two aromatic rings and the proline ring (Figure 5). The stability and high population of the structure is attested to by the results of an NOE-ROE study of the hydration of the peptide (113), which shows that the proline-ring protons with the furthest upfield-shifted resonances are partially buried from solvent, in accord with the 3D structures, which show these proline protons closely stacked against the adjacent aromatic rings. The turn is stabilized primarily by hydrophobic packing of the proline and aromatic side chains, with some additional contribution from electrostatic interactions between the positively charged N-terminus and the aspartate side chain carboxylate group.

Specific turn conformations in proteins and other polypeptides have been modeled by studying peptides in solution. In particular, Bansal & Gierasch (114) found using NOE experiments that the internalization sequence of the LDL receptor adopts a turn conformation in solution (114); this finding is important both for understanding surface recognition and export and for molecular design. A result relevant to the problem of protein folding in vivo has emerged from recent studies of the conformational preferences of peptides from the P22 phage tailspike protein. A region of the sequence that is altered in temperature-

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

sensitive folding mutants (115) contains a conformational preference for a turn in a peptide corresponding to the wild-type sequence, but not in the corresponding peptide with the mutant sequence (116). These results provide support for the role of turn formation in the initiation of folding of proteins.

Many peptide fragments that adopt helical structure in aqueous solution have now been identified. Indeed, local elements of helix can often be more readily identified by NMR through observation of medium range NOEs than by circular dichroism (106, 117). Further, the site-specific information obtained from the NMR experiment allows the location and, to a certain extent, the local population of helix to be determined (117): In most cases it appears that the helix is most populated in the center of the peptide and frayed towards the ends. Even in the most highly helical peptides, however, unfolded forms or alternately folded conformations are present in solution (118, 119).

An interesting structure detected by NMR in linear peptides is the so-called nascent helix (120), which is characterized by $d_{NN}(i,i+1)$ NOEs, intermediate to small $^3J_{HN\alpha}$ values, intermediate to low amide proton temperature coefficients, and the presence of a number of $d_{\alpha N}(i,i+2)$ NOEs, which are normally

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

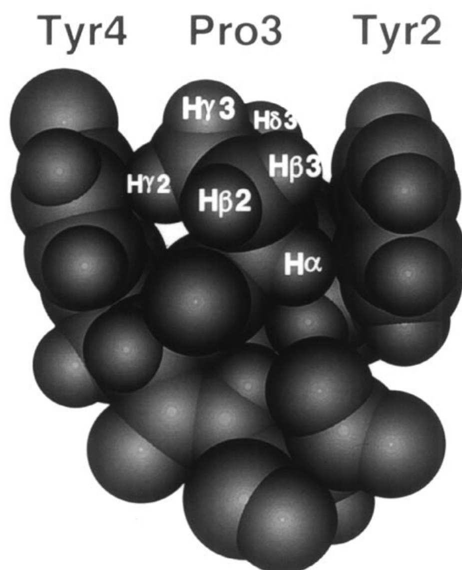

Figure 5 Structure of the Type VI turn conformation in a five-residue peptide SYPYD (112) showing the hydrophobic interaction between the aromatic and proline rings that gives rise to the unique properties and stabilization of this conformation in aqueous solution. [Reproduced with permission from Yao et al (113).]

associated with turns (106). The nascent helical peptide can readily be induced to form ordered helix by the addition of solvents such as trifluoroethanol (TFE). These results are interpreted in terms of a molecular model in which folding (and unfolding) of helices proceeds by way of turn-like intermediates. Indeed, such structures are observed in molecular dynamics simulations of the unfolding of helical peptides [reviewed by Brooks & Case (121)].

Recently, β-hairpin structures have been observed by NMR in peptide fragments of ubiquitin and the B domain of protein G (122–125). Interestingly, the hairpin structure is stabilized in solution by the addition of alcohols, either TFE or methanol (122, 123). A comparison was made of the secondary structure preferences of two peptide fragments of different lengths from the N-terminal sequence of ubiquitin (122). Although the peptides were unstructured in water solution, NMR studies in methanol solution show that they adopt a native-like β-hairpin structure, as is also found in the A-state of ubiquitin in aqueous methanol. Thus, although the β-hairpin is stabilized by long-range interactions in native ubiquitin, its structure is encoded entirely by the N-terminal amino acid sequence.

Local structures stabilized by hydrophobic interactions have been observed in a number of peptides derived from protein sequences. The diagnostic observations include upfield-shifted resonances and medium-range NOE connectivities involving side-chain protons. Such observations were used to infer the presence of a hydrophobic cluster in peptides derived from plastocyanin (126); a similar structure has been observed in a peptide from BPTI and in unfolded BPTI itself (127). An unusual interaction between the amide proton of a glycine residue and an aromatic ring preceding it by two residues in the amino acid sequence has been observed in several peptides (128, 129). The glycine amide proton is highly sequestered from solvent as a result of its interaction with the aromatic ring. Formation of local elements of structure, either local side-chain clusters or elements of secondary structure, stabilized by hydrophobic interactions probably plays a major role in the initiation of protein folding. The importance of specific hydrophobic interactions dictated by the sequence in stabilizing elements of secondary structure in peptides and unfolded proteins should not be underestimated. Indeed, hydrophobic interactions have been directly implicated in the stabilization of a Type VI turn (111, 112), and Alexandrescu & Shortle (71) have recently shown that hydrophobicity is a better measure than helix propensity of the likelihood that a particular sequence in a protein will contain residual structure in the unfolded state.

Proteins in Pieces and Pieces of Proteins

In order to delineate protein folding initiation sites for different protein structural motifs, several proteins have been examined by dissecting their sequences

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

into a series of peptide fragments and investigating the fragments by using NMR in aqueous solution. The protein sequences studied include myohemerythrin (120, 128), plastocyanin (126), myoglobin (117, 130–133), ubiquitin (122), BPTI (129, 134), barnase (135, 136), chymotrypsin inhibitor 2 (137), protein G B1 domain (138), and the helical regions of the α/β proteins CheY, flavodoxin, and P21-Ras (139). A weak correlation has been observed in these studies between conformational preferences for helical or nascent-helical structures in the peptides and the presence of helix in the folded protein. In particular, the peptides derived from the four-helix bundle protein myohemerythrin (128) showed a marked preference for secondary structure formation in solution, including turn, helix, and nascent helix conformational preferences. By contrast, all-β protein plastocyanin showed very little preference for such folded conformations (126). These results appear to indicate that there is a prepartitioning of conformational space sampled by the polypeptide backbone that is related to the secondary structure in the folded state (126, 128).

Several peptide fragments of the myoglobin sequence in the region of the G- and H-helices have been studied by both NMR and CD spectroscopy (117, 132, 133) in an attempt to determine which segments of the early intermediate formed on the folding pathway of apomyoglobin fold spontaneously and might therefore constitute folding initiation sites. The H-helix peptide forms a significant population of ordered helical conformations in water solution, whereas the G-helix peptide is not helical in water but becomes strongly helical in TFE solution (117). The peptide corresponding to the G-H turn region contains a significant population of a reverse turn, which persists in fragments of varying lengths (132, 133) and under a variety of solution conditions. Thus, both the H-helix and the G-H turn sequences exhibit a strong propensity for spontaneous secondary structure formation and probably play a key role in the initiation of folding of apomyoglobin.

An in vitro peptide model for the folding of a nascent polypeptide chain in vivo has been studied in the barley chymotrypsin inhibitor 2 system (140). A series of peptides of increasing length from the N-terminus were structurally characterized by CD and NMR spectroscopy to determine folding propensities. Consistent with the hypothesis that simple proteins fold as they are generated from the ribosome during translation, progressively greater populations of folded forms were detected in the longer peptides.

Peptide Models of Protein Folding Intermediates

Apart from the peptide fragments described above, a number of ingenious peptide mimics of protein structures and folding intermediates have been designed, including the PαPβ covalently linked heterodimer from BPTI (88) and a noncovalent peptide complex designed as a model of an early folding intermediate

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

of cytochrome c (41). These models represent a considerable achievement in design, since many factors can influence whether usable information can be derived from such systems, including solubility and the possibility that the desired association between the peptide components does not occur.

Effect on the Peptide Conformational Ensemble of Receptor-Protein Binding

Peptides interact with larger molecules in many systems, and it appears that a binding interaction with a receptor protein can cause a peptide to fold into a defined conformation. Binding to the receptor has traditionally been thought to mediate the induced-fit conformation of peptide hormones, which are in general largely unstructured in solution unless conformationally restricted. For complexes in which the dissociation rate of the peptide is favorable, the transferred NOE can be used to infer the structure of the bound peptide. In systems in which the dissociation rate is too slow for the transferred NOE to be observed, other NMR experiments are available for determination of the conformation of the bound peptide ligand. For example, isotope-edited NMR methods have been used to investigate the conformation of a peptide, part of the myohemerythrin C-helix, bound to an antibody fragment (141). The peptide, labeled site-specifically with ^{15}N and ^{13}C, was bound to the Fab fragment ($M_r \sim 50,000$) of an antibody that had been raised against the C-helix peptide. The peptide conformation appears to be helical when bound to the antibody, although it forms only a nascent helix when free in solution. It is encouraging that NMR experiments can be used to gain important information on folding of peptides in systems as large as Fab complexes. In another approach, exemplified by the structure determination of a calmodulin-target peptide complex (142), isotope-edited multidimensional NMR methods have been applied to a system consisting of ^{13}C,^{15}N-labeled protein and unlabeled peptide. High-resolution structures were obtained for both the protein and the bound peptide, which was found to fold into a helix upon binding to calmodulin.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

FUTURE PROSPECTS

Major advances in understanding the molecular mechanisms by which proteins fold into their correct tertiary structures have come from applications of NMR technology to the protein folding problem. The advent of quenched-flow hydrogen-exchange pulse labeling has contributed significantly to our understanding of millisecond time scale folding events, and future applications to additional structural classes may ultimately result in a generalized model of the slower steps in the protein folding pathway. With its ability to identify the structures populated in the conformational ensemble of peptide fragments

of proteins in aqueous solution under conditions that favor folding, NMR has provided new insights into the probable events occurring at the earliest stages of the initiation of protein folding. Applications to additional proteins and to more complex peptide models of early folding species are to be expected. Relatively new multidimensional NMR experiments now allow complete resonance assignments to be made for unfolded proteins under a variety of denaturing conditions. Such experiments hold exceptional promise for identification of residual structure in denatured proteins and will allow detailed structural characterization of the initial states in the folding process. Finally, heteronuclear multidimensional NMR studies of stabilized molten globule intermediates are now possible for some proteins and promise to provide important new insights into the 3D structure and dynamics of these states. There can be little doubt that NMR will continue to play a vital role in unraveling the protein folding puzzle.

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

ACKNOWLEDGMENTS

The authors' research on protein folding is supported by grants GM38794 (HJD, PEW) and DK34909 (PEW) from the National Institutes of Health.

Any *Annual Review* chapter, as well as any article cited in an *Annual Review* chapter, may be purchased from the Annual Reviews Preprints and Reprints service.
1-800-347-8007; 415-259-5017; email: arpr@class.org
Visit the *Annual Reviews home page* at
http://www.annurev.org.

Literature Cited

1. Matthews CR. 1993. *Annu. Rev. Biochem.* 62:653–83
2. Creighton TE. 1992. *Protein Folding.* New York: Freeman
3. Fink AL. 1995. *Annu. Rev. Biophys. Biomol. Struct.* 24:495–522
4. Freire E. 1995. *Annu. Rev. Biophys. Biomol. Struct.* 24:141–65
5. Roder H. 1989. *Meth. Enzymol.* 176:446–73
6. Frieden C, Hoeltzli SD, Ropson IJ. 1993. *Protein Sci.* 2:2007–14
7. Englander SW, Mayne L. 1992. *Annu. Rev. Biophys. Biomol. Struct.* 21:243–65
8. Koide S, Dyson HJ, Wright PE. 1993. *Biochemistry* 32:12299–310
9. Kiefhaber T, Labhardt AM, Baldwin RL. 1995. *Nature* 375:513–16
10. Wishart DS, Sykes BD, Richards FM. 1993. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1164:36–46
11. Dobson CM, Evans PA. 1984. *Biochemistry* 23:4267–70
12. Evans PA, Kautz RA, Fox RO, Dobson CM. 1989. *Biochemistry* 28:362–70
13. Huang GS, Oas TG. 1995. *Proc. Natl. Acad. Sci. USA* 92:6878–82
14. Huang GS, Oas TG. 1995. *Biochemistry* 34:3884–92
15. Ropson IJ, Frieden C. 1992. *Proc. Natl. Acad. Sci. USA* 89:7222–26
16. Blum AD, Smallcombe SH, Baldwin RL. 1978. *J. Mol. Biol.* 118:305–16
17. Kautz RA, Fox RO. 1993. *Protein Sci.* 2:851–58
18. Akasaka K, Naito A, Nakatani H. 1991. *J. Biomol. NMR* 1:65–70
19. Adler M, Scheraga HA. 1988. *Biochemistry* 27:2471–80
20. Balbach J, Forge V, van Nuland NAJ, Winder SL, Hore PJ, et al. 1995. *Nature Struct. Biol.* 2:865–70

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

21. Baldwin RL. 1993. *Curr. Opin. Struct. Biol.* 3:84–91
22. Roder H, Wüthrich K. 1986. *Proteins* 1:34–42
23. Kim PS, Baldwin RL. 1980. *Biochemistry* 19:6124–29
24. Udgaonkar JB, Baldwin RL. 1988. *Nature* 335:694–99
25. Roder H, Elöve GA, Englander SW. 1988. *Nature* 335:700–4
26. Udgaonkar JB, Baldwin RL. 1990. *Proc. Natl. Acad. Sci. USA* 87:8197–201
27. Jennings PA, Wright PE. 1993. *Science* 262:892–96
28. Bycroft M, Matouschek A, Kellis JT, Serrano L, Fersht AR. 1990. *Nature* 346:488–90
29. Briggs MS, Roder H. 1992. *Proc. Natl. Acad. Sci. USA* 89:2017–21
30. Lu J, Dahlquist FW. 1992. *Biochemistry* 31:4749–56
31. Radford SE, Dobson CM, Evans PA. 1992. *Nature* 358:302–7
32. Varley P, Gronenborn AM, Christensen H, Wingfield PT, Pain RH, et al. 1993. *Science* 260:1110–13
33. Mullins LS, Pace CN, Raushel FM. 1993. *Biochemistry* 32:6152–56
34. Jacobs MD, Fox RO. 1994. *Proc. Natl. Acad. Sci. USA* 91:449–53
35. Kuszewski J, Clore GM, Gronenborn AM. 1994. *Protein Sci.* 3:1945–52
36. Kragelund BB, Robinson CV, Knudsen J, Dobson CM, Poulsen FM. 1995. *Biochemistry* 34:7217–24
37. Jones BE, Matthews CR. 1995. *Protein Sci.* 4:167–77
38. Udgaonkar JB, Baldwin RL. 1995. *Biochemistry* 34:4088–96
39. Elöve GA, Bhuyan AK, Roder H. 1994. *Biochemistry* 33:6925–35
40. Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. *Struct. Biol.* 1(3):149–56
41. Wu LC, Laub PB, Elöve GA, Carey J, Roder H. 1993. *Biochemistry* 32:10271–76
42. Miranker A, Radford SE, Karplus M, Dobson CM. 1991. *Nature* 349:633–36
43. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM. 1993. *Science* 262:896–900
44. Brems DN. 1988. *Biochemistry* 27:4541–46
45. Griko YV, Privalov PL, Venyaminov SY, Kutyshenko VP. 1988. *J. Mol. Biol.* 202:127–38
46. Hughson FM, Wright PE, Baldwin RL. 1990. *Science* 249:1544–48
47. Cocco MJ, Kao YH, Phillips AT, Lecomte JTJ. 1992. *Biochemistry* 31:6481–91
48. Cocco MJ, Lecomte JTJ. 1990. *Biochemistry* 29:11067–72
49. McDonald CA, Phillips DC. 1969. *J. Am. Chem. Soc.* 91:1513
50. Matthews CR, Westmoreland DG. 1975. *Biochemistry* 14:4532–38
51. Saab-Rincon G, Froebe CL, Matthews CR. 1993. *Biochemistry* 32:13981–90
52. Billeter M, Braun W, Wüthrich K. 1982. *J. Mol. Biol.* 155:321–46
53. Akasaka K, Naito A, Imanari M. 1991. *J. Am. Chem. Soc.* 113:4688–89
54. Evans PA, Topping KD, Woolfson DN, Dobson CM. 1991. *Proteins* 9:248–66
55. Neri D, Wider G, Wüthrich K. 1992. *FEBS Lett.* 303:129–35
56. Neri D, Wider G, Wüthrich K. 1992. *Proc. Natl. Acad. Sci. USA* 89:4397–401
57. Neri D, Billeter M, Wider G, Wüthrich K. 1992. *Science* 257:1559–63
58. Roder H, Wagner G, Wüthrich K. 1985. *Biochemistry* 24:7407–11
59. Robertson AD, Baldwin RL. 1991. *Biochemistry* 30:9907–14
60. Buck M, Radford SE, Dobson CM. 1994. *J. Mol. Biol.* 237:247–54
61. Egan DA, Logan TM, Liang H, Matayoshi E, Fesik SW, et al. 1993. *Biochemistry* 32:1920–27
62. Braun D, Wider G, Wüthrich K. 1994. *J. Am. Chem. Soc.* 116:8466–69
63. Nicholls A, Sharp KA, Honig B. 1991. *Proteins* 11:281–96
64. Logan TM, Thériault Y, Fesik SW. 1994. *J. Mol. Biol.* 236:637–48
65. Arcus VL, Vuilleumier S, Freund SMV, Bycroft M, Fersht AR. 1994. *Proc. Natl. Acad. Sci. USA* 91:9412–16
66. Alexandrescu AT, Abeygunawardana C, Shortle D. 1994. *Biochemistry* 33:1063–72
67. Frank MK, Clore GM, Gronenborn AM. 1995. *Protein Sci.* 4:2605–15
68. Pan H, Barbar E, Barany G, Woodward C. 1995. *Biochemistry* 34:13974–81
69. Zhang O, Forman-Kay JD. 1995. *Biochemistry* 34:6784–94
70. van Mierlo CPM, Darby NJ, Keeler J, Neuhaus D, Creighton TE. 1993. *J. Mol. Biol.* 229:1125–46
71. Alexandrescu AT, Shortle D. 1994. *J. Mol. Biol.* 242:527–46
72. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. 1995. *Biochemistry* 34:868–78
73. Lipari G, Szabo A. 1982. *J. Am. Chem. Soc.* 104:4559–70
74. Lipari G, Szabo A. 1982. *J. Am. Chem. Soc.* 104:4546–59
75. Harding MM, Williams DH, Woolfson DN. 1991. *Biochemistry* 30:3120–28

76. Pan Y, Briggs MS. 1992. *Biochemistry* 31:11405–12
77. Stockman BJ, Euvrard A, Scahill TA. 1993. *J. Biomol. NMR* 3:285–96
78. Fan P, Bracken C, Baum J. 1993. *Biochemistry* 32:1573–82
79. Baum J, Dobson CM, Evans PA, Hanley C. 1989. *Biochemistry* 28:7–13
80. Alexandrescu AT, Evans PA, Pitkeathly M, Baum J, Dobson CM. 1993. *Biochemistry* 32:1707–18
81. Alexandrescu AT, Ng Y-L, Dobson CM. 1994. *J. Mol. Biol.* 235:587–99
82. van Mierlo CPM, Darby NJ, Neuhaus D. 1991. *J. Mol. Biol.* 222:353–71
83. van Mierlo CPM, Darby NJ, Neuhaus D, Creighton TE. 1991. *J. Mol. Biol.* 222:373–90
84. van Mierlo CPM, Kemmink J, Neuhaus D, Darby NJ, Creighton TE. 1994. *J. Mol. Biol.* 235:1044–61
85. Staley JP, Kim PS. 1992. *Proc. Natl. Acad. Sci. USA* 89:1519–23
86. Staley JP, Kim PS. 1994. *Protein Sci.* 3:1822–32
87. Ferrer M, Barany G, Woodward C. 1995. *Nature Struct. Biol.* 2:211–17
88. Oas TG, Kim PS. 1988. *Nature* 336:42–48
89. Peng X, Jonas J, Silva JL. 1993. *Proc. Natl. Acad. Sci. USA* 90:1776–80
90. Peng X, Jonas J, Silva JL. 1994. *Biochemistry* 33:8323–29
91. Zhang J, Peng X, Jonas A, Jonas J. 1995. *Biochemistry* 34:8631–41
92. Baldwin RL, Roder H. 1991. *Curr. Biol.* 1:218–20
93. Loh SN, Kay MS, Baldwin RL. 1995. *Proc. Natl. Acad. Sci. USA* 92:5446–50
94. Jeng M-F, Englander SW. 1991. *J. Mol. Biol.* 221:1045–61
95. Jeng M-F, Englander SW, Elöve GA, Wand AJ, Roder H. 1990. *Biochemistry* 29:10433–37
96. Kuroda Y, Endo S, Nagayama K, Wada A. 1995. *J. Mol. Biol.* 247:682–88
97. Chyan C-L, Wormald C, Dobson CM, Evans PA, Baum J. 1993. *Biochemistry* 32:5681–91
98. Buck M, Radford SE, Dobson CM. 1993. *Biochemistry* 32:669–78
99. Mayo SL, Baldwin RL. 1993. *Science* 262:873–76
100. Bai Y, Sosnick TR, Mayne L, Englander SW. 1995. *Science* 269:192–97
101. Wright PE, Dyson HJ, Lerner RA. 1988. *Biochemistry* 27:7167–75
102. Brown JE, Klee WA. 1971. *Biochemistry* 10:470–76
103. Kim PS, Baldwin RL. 1984. *Nature* 307:329–34
104. Montelione GT, Arnold E, Meinwald YC, Stimson ER, Denton JB, et al. 1984. *J. Am. Chem. Soc.* 106:7946–58
105. Dyson HJ, Cross KJ, Houghten RA, Wilson IA, Wright PE, et al. 1985. *Nature* 318:480–83
106. Dyson HJ, Wright PE. 1991. *Annu. Rev. Biophys. Biophys. Chem.* 20:519–38
107. Skolnick J, Kolinski A. 1989. *Annu. Rev. Phys. Chem.* 40:207–35
108. Ybe JA, Hecht MH. 1996. *Protein Sci.* 5:814–24
109. Chou PY, Fasman GD. 1977. *J. Mol. Biol.* 115:135–75
110. Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE. 1988. *J. Mol. Biol.* 201:161–200
111. Yao J, Feher VA, Espejo BF, Reymond MT, Wright PE, et al. 1994. *J. Mol. Biol.* 243:736–53
112. Yao J, Dyson HJ, Wright PE. 1994. *J. Mol. Biol.* 243:754–66
113. Yao J, Brüschweiler R, Dyson HJ, Wright PE. 1995. *J. Am. Chem. Soc.* 116:12051–52
114. Bansal A, Gierasch LM. 1991. *Cell* 67:1195–201
115. Yu MH, King J. 1988. *J. Biol. Chem.* 263:1424–31
116. Stroup AN, Gierasch LM. 1990. *Biochemistry* 29:9765–71
117. Waltho JP, Feher VA, Merutka G, Dyson HJ, Wright PE. 1993. *Biochemistry* 32:6337–47
118. Osterhout JJ Jr, Baldwin RL, York EJ, Stewart JM, Dyson HJ, et al. 1989. *Biochemistry* 28:7059–64
119. Merutka G, Morikis D, Brüschweiler R, Wright PE. 1993. *Biochemistry* 32:13089–97
120. Dyson HJ, Rance M, Houghten RA, Wright PE, Lerner RA. 1988. *J. Mol. Biol.* 201:201–17
121. Brooks CL III, Case DA. 1993. *Chem. Rev.* 93:2487–502
122. Cox JPL, Evans PA, Packman LC, Williams DH, Woolfson DN. 1993. *J. Mol. Biol.* 234:483–92
123. Blanco FJ, Jiménez MA, Pineda A, Rico M, Santoro J, et al. 1994. *Biochemistry* 33:6004–14
124. Blanco FJ, Jiménez MA, Herranz J, Rico M, Santoro J, et al. 1993. *J. Am. Chem. Soc.* 115:5887–88
125. Blanco FJ, Rivas G, Serrano L. 1994. *Nat. Struct. Biol.* 1:584–90
126. Dyson HJ, Sayre JR, Merutka G, Shin H-C, Lerner RA, et al. 1992. *J. Mol. Biol.*

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.

226:819–35
127. Lumb KJ, Kim PS. 1994. *J. Mol. Biol.* 236:412–20
128. Dyson HJ, Merutka G, Waltho JP, Lerner RA, Wright PE. 1992. *J. Mol. Biol.* 226:795–817
129. Kemmink J, van Mierlo CPM, Scheek RM, Creighton TE. 1993. *J. Mol. Biol.* 230:312–22
130. Waltho JP, Feher VA, Wright PE. 1990. In *Current Research in Protein Chemistry,* ed. JJ Villafranca, pp. 283–93. New York: Academic
131. Waltho JP, Feher VA, Lerner RA, Wright PE. 1989. *FEBS Lett.* 250:400–4
132. Shin H-C, Merutka G, Waltho JP, Wright PE, Dyson HJ. 1993. *Biochemistry* 32:6348–55
133. Shin H-C, Merutka G, Waltho JP, Tennant LL, Dyson HJ, et al. 1993. *Biochemistry* 32:6356–64
134. Kemmink J, Creighton TE. 1993. *J. Mol. Biol.* 234:861–78
135. Sancho J, Neira JL, Fersht AR. 1992. *J. Mol. Biol.* 224:749–58
136. Ikura T, Go N, Kohda D, Inagaki F, Yanagawa H, et al. 1993. *Proteins* 16:341–56
137. De Prat Gay G, Fersht AR. 1994. *Biochemistry* 33:7957–63
138. Blanco FJ, Serrano L. 1995. *Eur. J. Biochem.* 230:634–49
139. Muñoz V, Serrano L, Jiménez MA, Rico M. 1995. *J. Mol. Biol.* 247:648–69
140. De Prat Gay G, Ruiz-Sanz J, Neira JL, Itzhaki LS, Fersht AR. 1995. *Proc. Natl. Acad. Sci. USA* 92:3683–86
141. Tsang P, Rance M, Fieser TM, Ostresh JM, Houghten RA, et al. 1992. *Biochemistry* 31:3862–71
142. Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, et al. 1992. *Science* 256:32–38
143. Dyson HJ, Wright PE. 1996. In *Encyclopedia of NMR*, ed. DM Grant, RK Harris. pp. 3811–20. Chichester: Wiley
144. Kuriyan J, Wilz S, Karplus M, Petsko GA. 1986. *J. Mol. Biol.* 192:133–54

Annu. Rev. Phys. Chem. 1996.47:369-395. Downloaded from www.annualreviews.org
by Kanazawa University Medical Library Branch on 05/28/14. For personal use only.