Unfolded Proteins and Protein Folding Studied by NMR

H. Jane Dyson* and Peter E. Wright*

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received December 10, 2003

Contents

1. Introduction

Preparation of biological macromolecules in the pure state requires that cells be disrupted, releasing and mixing the contents. Only the most stable and highly structured molecules can survive in the cellular "soup", which contains proteases and nucleases that would be tightly controlled and sequestered in a normal living cell. Thus, as long as "activity" of polypeptides had to be measured by classical biochemical methods, in test tubes, using purified and well-characterized components, the paradigm that a functional protein must be well-structured held good. However, as methods for exploring the functions of proteins (and other macromolecules) within cells and in more complex in vitro systems have arisen, it has become clear that unfolded and partly folded proteins have important roles to play in numerous cellular processes and signaling events. The extent and variety of the role of such proteins has not been determined as yet, but promises to provide a fruitful new field for thinking about the molecular mechanisms of biological processes.

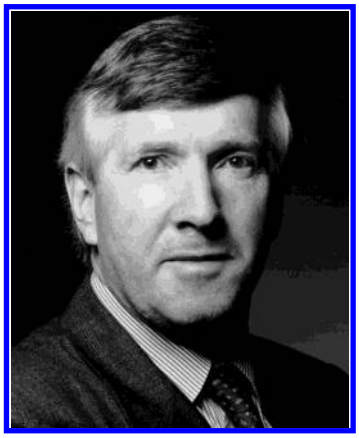
It has long been recognized that biological macromolecules are far from rigid in their structures. Motion is implicit in the normal function of such well-known molecules as serum albumin[1] and myoglobin.[2] Segmental motion of protein subunits in isolation, which become locked into specific structures in the presence of specific binding partners[3,4] is now recognized as an important component of binding specificity. Binding sites for interacting proteins are frequently more mobile than the rest of the protein.[5] However, until quite recently, the presence of functional polypeptide domains that consist of a conformational ensemble lacking an overall uniform three-dimensional structure has not been recognized.[6] Using a neural network algorithm,[7] Dunker and colleagues show that an appreciable portion of published genomes should code for proteins with unstructured regions of 50 residues or more. Certain classes of proteins, particularly those involved in transcriptional activation[8] and cell cycle regulation,[9] appear to contain domains that are intrinsically unstructured in solution, but become structured on binding to their physiological targets. Since then, the literature on "intrinsically unstructured proteins" has burgeoned, and this topic has been the subject of a number of recent reviews.[6,7,10−17] In addition, the process of protein folding itself has become of increasing interest. NMR has emerged as a particularly important tool for studies of protein folding because of the unique structural insights it can provide into the events of the protein folding process.[18]

NMR remains one of the few comprehensive sources for information on unstructured and partly structured proteins and on the protein folding process. One of the earliest NMR solution structure calculations was performed for the micelle-bound peptide hormone glucagon,[19] which is intrinsically unstructured in solution,[20] like many other small peptide hor-

* Corresponding authors: H. Jane Dyson and Peter E. Wright, Department of Molecular Biology MB2, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla California 92037. Tel.: 858-784-9721 (P.E.W.) 858-784-2223 (H.J.D.). Fax: 858-784-9822. E-mail: wright@scripps.edu; dyson@scripps.edu.

10.1021/cr030403s CCC: $48.50
Published on Web 07/10/2004

Jane Dyson received the degree of B.Sc. (hons) from the University of Sydney in 1973 and a Ph.D. from the University of Sydney in 1977, under the direction of James Beattie. She was a postdoctoral fellow at Massachusetts Institute of Technology under the direction of Paul Schimmel from 1977 to 1978, and held a Damon Runyon-Walter Winchell postdoctoral award. She was appointed as a Lecturer in Chemistry at the University of New South Wales in 1979, and joined the Scripps Research Institute in 1984, where she is presently a professor. Her research interests are in the conformation of peptides, protein folding and dynamics, and structure and functional studies of proteins using NMR and other spectroscopic techniques.



Peter Wright received his B.Sc. degree in Chemistry from the University of Auckland in 1968, M.Sc. in Chemistry from the University of Auckland in 1969, and Ph.D. from the University of Auckland in 1972. He received a New Zealand University Grants Committee Postdoctoral Fellowship to the University of Oxford, England from 1972 to 1976, working under the direction of R. J. P. Williams. Dr. Wright was appointed as Lecturer in the Department of Inorganic Chemistry at the University of Sydney in 1976, and was promoted to Senior Lecturer in 1980. He joined the faculty at The Scripps Research Institute as Professor in 1984, and was appointed in 1987 as the Chairman of the Department of Molecular Biology. He holds the Cecil and Ida Green Chair in Biomedical Research, and has acted as Editor-in-Chief of *Journal Molecular Biology* since 1990. He received an honorary M. D. degree from the Karolinska Institute, Sweden in 1995, and an honorary D.Sc. degree from the University of Sydney, Australia, in 2003. His research interests are in applications of NMR to protein folding and intrinsically unstructured proteins, solution structure and dynamics of proteins, and protein–nucleic acid interactions.

mones. Proton NMR studies of denatured proteins are difficult because of the characteristic lack of resonance dispersion. However, indirect detection heteronuclear NMR techniques overcome the problem of resonance overlap[21] and provide high-resolution information on highly denatured proteins (see following section). The earliest report of NMR observation of residual structure in a highly denatured protein appeared in 1992.[22] In most cases, unique three-dimensional structures of unstructured proteins will not be available from crystallographic studies because crystals of conformationally disordered molecules are difficult to form and if formed may not be representative of the conformational ensemble in solution. Recent attempts have been made to characterize functionally disordered systems by crystallizing them together with fusion proteins such as GST[23] or in the presence of binding partners or antibodies,[24,25] but these structures remain representative only of one particular member of the conformational ensemble of the free protein in solution. By contrast, the NMR method can give a great deal of (less specific but more accurate) information on the structural composition of members of a conformational ensemble. A few examples have been reported of solution structure calculations from NMR data of the structures present in the conformational ensembles of unfolded proteins[26] or transition state ensembles.[27]

In this review, we first survey NMR methods that are particularly applicable to the study of unfolded and partly folded proteins, and the types of information that can be obtained. The second part of the review provides some examples of protein systems in which NMR has been instrumental in the elucidation of folding pathways and mechanisms. Finally, a few examples are provided of NMR studies of functional unfolded proteins.

2. Methodologies for Studying Unfolded and Partly Folded Proteins

A recent issue of *Advances in Protein Chemistry* (Volume 62, 2002) was devoted to the study of unfolded proteins. A comprehensive review of the NMR methodology applied to unfolded and partly folded proteins is included in this volume,[15] together with other techniques such as Raman optical activity,[28] fluorescence correlation spectroscopy,[29] infrared absorption and vibrational CD,[30] and small angle scattering.[31] Other recent reviews deal with applications of NMR to study the protein folding process.[32,33]

Direct characterization of unfolded and partly folded proteins is possible on a residue-specific basis using high-field NMR spectrometers, uniformly and specifically labeled proteins, and isotope-edited and triple-resonance pulse sequences. These methodological improvements have overcome the problem of proton signal overlap that hampered early NMR studies of unfolded proteins. In many cases, resonance assignments can be made using the dispersion of the ^{13}C and ^{15}N nuclei, which are more sensitive to local amino acid sequence, rather than the protons, which are most sensitive to structural context in three dimensions.[34] Comprehensive tabulations of sequence-dependent corrections to random coil chemical shifts[35,36] should prove of particular use in detecting and quantifying residual structure in unfolded or partly folded proteins. Models for the "random coil" distributions of ϕ and ψ angles[37] and χ_1 side chain dihedral angles[38] have also been proposed, allowing estimates to be made of the presence of residual structure from the measurement of coupling constants.[38]

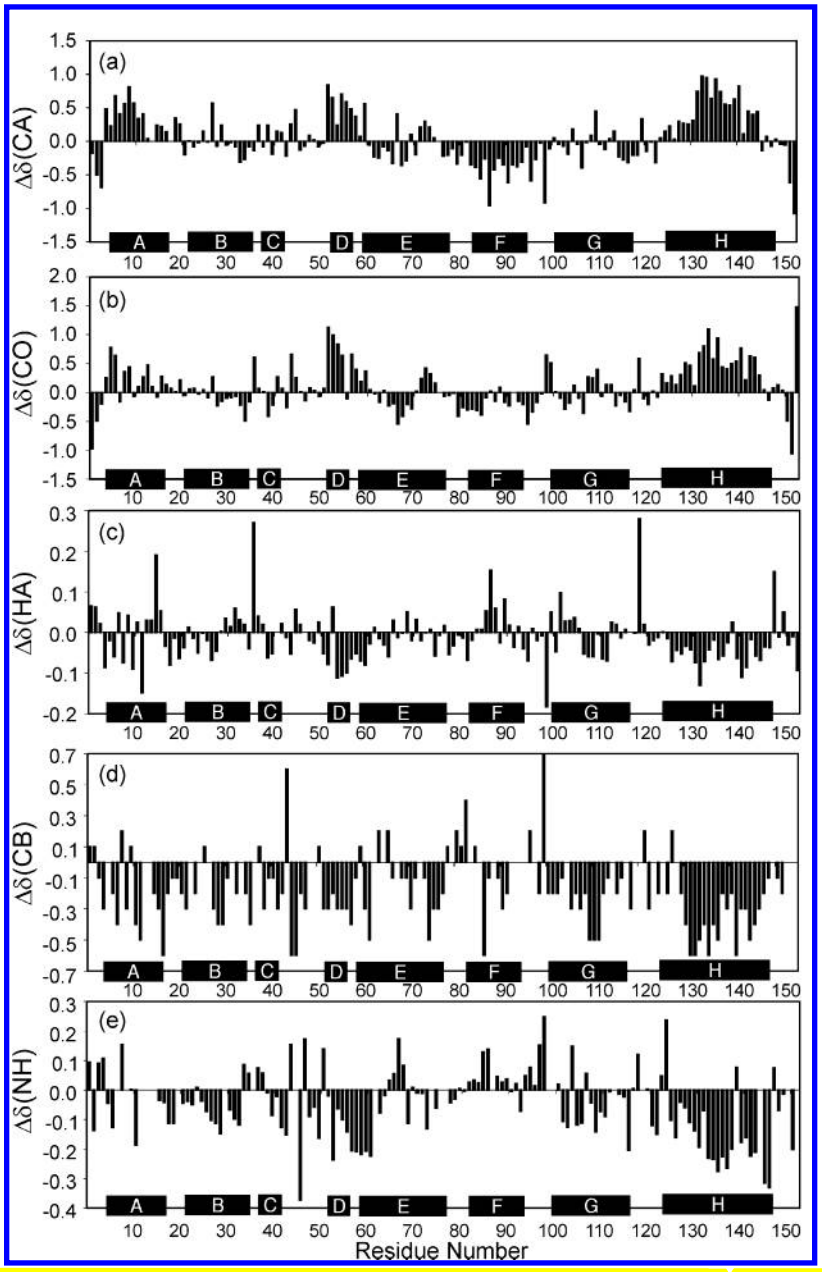


Figure 1. Secondary chemical shifts, corrected for sequence-dependent contributions,[35] of (a) $^{13}C^{\alpha}$, (b) ^{13}CO, (c) $^{1}H^{\alpha}$, (d) $^{13}C^{\beta}$, (e) $^{1}H^{N}$ resonances of acid-unfolded apomyoglobin. Regions corresponding to the helices of the native protein are marked with black bars. (Reprinted from ref 175 with permission. Copyright 2001 American Chemical Society).

2.1. Information from Chemical Shifts

Once resonance assignments have been made for an unfolded or partly folded protein, a number of NMR measurements can be made to further characterize the conformational ensemble. The primary observable in NMR studies of unfolded and partly folded proteins remains the chemical shift. Variations in chemical shift from "random coil" values established by various means[39−43] give important insights into the structures populated in the conformational ensemble in incompletely folded proteins. An example of this is shown in Figure 1, which shows the secondary chemical shifts, corrected for sequence-dependence,[35] for several nuclei in the NMR spectra of apomyoglobin unfolded at pH 2.3. All of the data are consistent in showing the presence of a residual propensity for helical structure in the A and H helices of the fully folded protein, as well as a small area of non-native helical propensity in the D/E helix linker.

2.2. Dynamic Information from Relaxation Data

Estimates of backbone and side chain dynamics using NMR relaxation provide a unique insight into protein motions. A number of techniques have been used to characterize the dynamics of unfolded proteins, and the results have been used to make inferences about the local and global motion of unfolded molecules. The most commonly used NMR techniques for studying polypeptide chain dynamics involve the measurement of T_1, T_2, and heteronuclear NOE for backbone resonances. Most studies use the convenience of the backbone amide proton in ^{15}N-labeled proteins.

Recent innovations include R_2 relaxation dispersion,[44] NMR relaxation coupled with MD simula-

tion,[45] dipole−dipole cross-correlated spin relaxation,[46] and off-resonance $R_{1\rho}$.[47] All of these methods are in general similar to those used for folded proteins, producing only a different range of values. Side chain dynamics can provide a particularly useful insight into the development of hydrophobic interactions in unfolded states: these have been measured using ^{2}H and ^{13}C relaxation.[48−50]

For data analysis, the use of the model-free formalism[51,52] is not usually valid for highly unfolded proteins because the assumption of a single overall correlation time and the temporal deconvolution of internal motions and molecular tumbling is invalid. Model-free calculations have been used to analyze relaxation measurements on unfolded states,[53,54] but most studies use the method of reduced spectral density mapping.[55,56] Recent variations on the model-free method take into account the distribution of correlation times.[57,58]

2.3. Distance Information: NOEs and Spin Labels

The three-dimensional structures of folded proteins are elucidated in solution primarily with the aid of long-range distance information available from the nuclear Overhauser effect. Regions of the protein that are distant in the primary sequence but close together in space in the folded protein structure give rise to NOEs that can be utilized to determine the global fold. The NOE depends on the inverse sixth power of the internuclear distance, and thus, the complete set of distance restraints consists of a list of proton contacts within about 5 Å. Detection and assignment of long-range NOEs in unfolded proteins is extremely difficult, although elegant pulse sequences have been devised for this purpose.[59,60] Medium-range NOEs indicative of helical or turn-like structure have been observed in peptides and in unfolded proteins, but long-range interactions indicative of the presence of transient tertiary structure at equilibrium[61] have not in general been well-authenticated.[62] It is likely that, for most systems so far studied, either the population of the transiently structured forms is too low, or the ensemble containing them is too heterogeneous, for the NOE to be observable. Other NMR evidence for transient long-range interactions, for example, from relaxation data, is strong in several cases (see later section).

Long-range distance information to complement and extend these observations has recently been obtained from the use of covalently attached nitroxide spin labels. Paramagnetic nitroxide spin labels cause broadening of nuclear spins within a radius of about 15 Å and were originally used to determine interatomic distances in folded proteins.[63,64] Extension of the method for use in unfolded and partly folded proteins was pioneered by Shortle, studying a fragment of staphylococcal nuclease (see later section)[65,66] and more recently used on protein L,[67] acyl CoA-binding protein,[68] and apomyoglobin.[69]

The method involves engineering of a single Cys residue into the protein by site-directed mutagenesis (the same method can be used to substitute nondisulfide Cys residues already present in the sequence). The thiol group of the Cys residue reacts with a spin-label reagent, for example, PROXYL (1-oxyl-2,2,5,5-tetramethyl-3-pyrrolidinyl) activated as an iodoacetamide or methanethiosulfonate. Spin label sites must be chosen with care not to disrupt or influence structure formation. Distance information is obtained by assessing the broadening effect of the spin label: a spectrum, for example, a ^{1}H-^{15}N HSQC spectrum is recorded for the paramagnetic (oxidized spin label) sample, then a second spectrum is obtained with the spin label reduced to the diamagnetic state. Differences in the line width, relaxation rates, or intensity in these two spectra give an estimate of the distance of the spin label site from any given amide. An example of such an experiment is shown in Figure 2. Results may be analyzed either qualitatively, utilizing a plot that compares the intensities of cross-peaks in the spin labeled and reduced spectrum or by a more quantitative analysis that gives actual distance ranges.[66] Figure 3 shows the type of information that can be obtained from spin label experiments on unfolded and partly folded proteins. Unfolded apomyoglobin shows evidence of transient long-range interactions within the C-terminal 50 residues and between the N- and C-terminal regions (Figure 3B), while other parts of the molecule show no evidence of such interactions (Figure 3A).

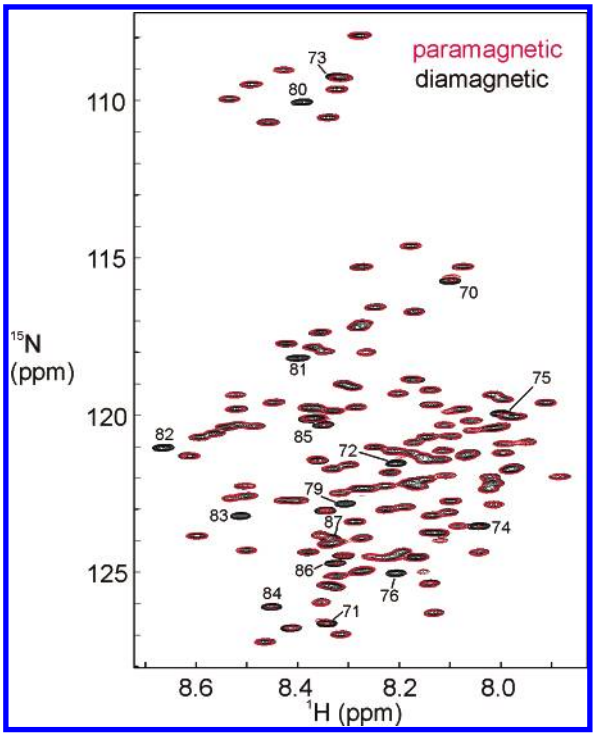


Figure 2. 750 MHz ^{15}N-HSQC spectra at pH 2.3, 25 °C, of the K77C* derivative of apomyoglobin spin-labeled by reaction of the spin label (1-oxyl-2,2,5,5-tetramethyl-Δ^{3}-pyrrolin-3-yl)methyl methanethiosulfonate with the cysteine thiol. The spectrum of the paramagnetic form (red) is superimposed on that of the diamagnetic state obtained by reduction of the spin label with ascorbic acid (black). Labeled cross-peaks are for resonances that are broadened by the presence of the spin label. (Adapted from ref 69 with permission. Copyright 2002 Elsevier).

2.4. Structural Information: Dipolar Couplings

A powerful new method for the characterization of overall structure in biological macromolecules utilizes

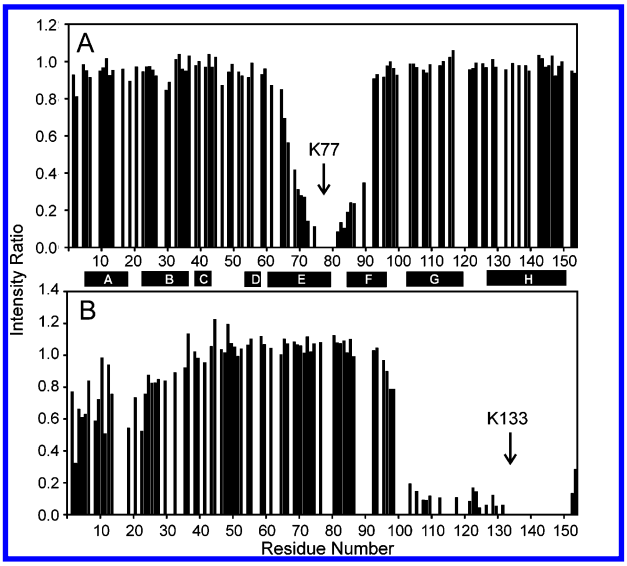


Figure 3. Paramagnetic enhancement to nuclear spin relaxation for unfolded apomyoglobin at pH 2.3. The histograms show the experimental intensity ratios ($I = I_{para}/I_{dia}$) for each residue with an adequately resolved cross-peak in the ^{1}H-^{15}N HSQC spectrum of (A) K77C* and (B) K133C*. (Adapted from reference 69 with permission. Copyright 2002 Elsevier).

residual dipolar couplings in partially aligned media.[70,71] These methods and the relationship between residual dipolar couplings and macromolecular structure have been recently extensively reviewed.[72−74] Partial alignment may be accomplished in a number of ways: by direct induction in the magnetic field,[75] by the use of dilute solutions of lipid bicelles,[71] using filamentous bacteriophages,[76] or by incorporation of the sample in stressed polyacrylamide gels.[77,78] The majority of the literature utilizing residual dipolar couplings for structure determination by NMR has focused on their application to the refinement of protein structure, particularly in cases in which the overall topology of the molecule is not well determined due to a paucity of restraints that would fix long-range order. For example, residual dipolar couplings can be used to determine the relative orientation of independently folded protein domains[79,80] or the bending of nucleic acid structures.[81]

Application of dipolar couplings to unfolded and partly folded states could, in principle, give valuable information on the overall backbone topology, without relying on NOEs or spin labels.[74] Care is required, however, to ensure that the media used to obtain partial alignment do not interact with the unfolded or partly folded proteins of interest. When a protein is incompletely folded, there may be extensive areas of exposed hydrophobic groups, which cause binding to bicellar media and consequent loss of signal. The most successful medium for the study of unfolded and partly folded proteins has been the polyacrylamide gel.[82,83]

2.5. Amide Proton Hydrogen Exchange

One of the earliest methods of NMR characterization of unfolded states was the measurement of amide proton hydrogen exchange, and the comparison of exchange behavior in native and non-native states.[84−86] Information on the integrity of structural domains under weak and strong denaturing conditions, as well as an indication of polypeptide motional changes, have been demonstrated by equilibrium hydrogen exchange measurements on a number of systems.[87−92] Quench-flow hydrogen exchange experiments analyzed by NMR[93,94] and mass spectrometry[95] remain the basis for most kinetic studies of protein folding.

2.6. NMR under Pressure

The ensemble of conformational states sampled by a protein in solution under a given set of conditions can be perturbed by increasing the pressure. This occurs because the ensemble contains conformers with different effective volumes, in rapid equilibrium with each other. An increase in pressure favors states of the system with smaller volume, and thus shifts the conformational equilibrium in the direction of lower-volume conformers. In a number of cases, this perturbation of the conformational equilibrium can result in the population of higher energy states under normal conditions of pH, temperature, and denaturant concentration. These higher energy states can approximate folding intermediates and unfolded states of the protein. Variable-pressure NMR studies have been published on a number of proteins, including lysozyme,[96] α-lactalbumin,[97] myoglobin,[98] and ubiquitin.[99] Detailed reviews of variable-pressure NMR have recently been published.[100,101]

2.7. Real-Time NMR

Since NMR is intrinsically a slow technique, in which it may take minutes to hours for measurements to be made, the study of protein folding in real time has been limited to those systems in which the folding process is exceptionally slow. Even within this class, which includes proteins where disulfide exchange and proline isomerization slow the folding rate dramatically, some systems are not amenable to study by this method, due to problems with exchange-mediated line broadening and resonance overlap. An important innovation was made by the combined use of rapid mixing techniques and NMR to monitor different aspects of the folding of α-lactalbumin.[102−105] Other systems studied by real-time NMR include triple helical collagen peptide models,[106,107] apoplastocyanin,[108] and barstar.[109] Such studies confirm the highly cooperative nature of the protein folding process, with probes at all sites in the molecule displaying identical folding kinetics.

A related technique involves the use of time-resolved photo-CIDNP (chemically induced dynamic nuclear polarization) NMR, in which laser pulses are used to excite a dye present in the protein solution as it sits in the NMR probe. Depending on their solvent exposure, tyrosine, histidine, and tryptophan side chains in the protein are excited to form short-lived radical pairs, which can be detected in 1D[110] or 2D NMR experiments.[111] This method has been used to explore the configuration of several proteins during the folding process, including α-lactalbumin[112] and lysozyme[111] and single-tryptophan mutants of HPr.[113]

Detailed reviews on real-time methods have been published over the past few years.[114−116]

2.8. Diffusion-Based Methods for Determination of Hydrodynamic Radius

The effective hydrodynamic radius of a macromolecule, defined as the radius of a sphere with the same diffusion coefficient,[117] can be estimated by measuring the diffusion coefficient by gradient NMR methods. In the case of unfolded proteins, which consist of a rapidly interconverting conformational ensemble, the diffusion coefficient is a population-weighted average. In general, the results obtained by NMR methods agree extremely well with those obtained by small-angle X-ray scattering, as seen for lysozyme[117−119] and the drk SH3 domain.[120] Pulsed field gradient diffusion methods have been used to elucidate the composition of unfolded states,[121] and have been utilized in questions of association state.[122−124]

3. Toward an Understanding of the Protein Folding Process

The processes of DNA replication, gene transcription, and mRNA translation, all necessary preliminaries to the formation of a polypeptide chain, are extremely complex, and involve layers of control that are only now being elucidated. However, the process of producing a polypeptide remains the same, no matter what the primary sequence of the protein. The details of the final step in the process, the folding of the protein, depend strongly on the actual composition and primary sequence—although the physical principles that govern the folding process are the same for all proteins, folding pathways may differ significantly for different proteins. Some proteins require extra help, such as prosequences and chaperones, to fold to the correct conformation.[125,126] Many different protein systems are under study at present, as a number of groups endeavor in the long term to discern the common features in the various systems and to determine general principles. NMR has an important role to play in such studies, but there is a fundamental disconnect between the process being studied and the basic attributes of the technique itself. This disconnect is due to the different time scales of the folding process, which is complete in milliseconds for many proteins, and of the NMR experiment (the fastest 2D spectra can only be accumulated in minutes). NMR experiments to study aspects of the folding process must therefore be creatively tailored to give valid information. Valuable information on the kinetics of protein folding can be obtained from quench-flow hydrogen exchange methods, detected by NMR[93,94] or using mass spectrometry.[95] Magnetization transfer methods have been successfully used to study fast-folding proteins.[127−130] Recently, equilibrium approaches, in which stable unfolded or partly folded states can be studied in solution over a relatively long period, have been extremely fruitful for a number of protein systems, to be discussed in the following sections.

The importance of the composition of the denatured state in the study of protein folding processes was first recognized by Tanford.[131,132] The following section reviews approaches to the characterization of the unfolded states of proteins, in the context of the information obtained on particular systems. Often, a number of methods, both NMR and non-NMR, are combined to give a more complete description. In the following paragraphs, we attempt to distill the extensive research that has been published on four major protein systems, the SH3 domain of drk, staphylococcal nuclease, lysozyme, and apomyoglobin. NMR studies have been published on many other unfolded proteins: examples include FK506-binding protein,[133] barnase,[134−136] Protein G,[137] and acyl coenzyme A-binding protein.[68]

3.1. Folded and Unfolded States in Equilibrium: An SH3 Domain of drk

NMR studies of the N-terminal src-homology (SH3) domain of the *Drosophila* signal adapter protein drk have provided novel insights into the behavior of unfolded proteins in solution. Because this domain is present in solution under normal conditions (pH 6.0−7.5, 50 mM sodium phosphate, 23−37 °C) as an equilibrium mixture in slow exchange on the NMR time scale between a fully folded form and one which is largely unfolded,[138−141] many aspects of the structure and dynamics of the unfolded protein, as well as the nature of the transition to the folded form, can be examined conveniently. Following assignment[140] and structural characterization[139] of both folded and unfolded forms of the protein, backbone dynamics[138,142] of the unfolded form could be extensively studied. NMR experiments were designed to determine the rates of exchange between the two forms, under conditions in which at least one of the exchanging forms has rapid exchange of amide protons with solvent.[141] To discriminate the NMR characteristics of the two forms, the equilibrium between folded and unfolded forms was perturbed toward stabilization of the folded form by the addition of a proline-containing peptide, and toward the stabilization of the unfolded form by the addition of a denaturant.[143] These studies showed the presence of residual structure, mainly turn-like structures, in both unfolded states (in the presence and absence of chemical denaturant), but there were differences in detailed structural features that prompted a caution regarding extrapolation of results obtained from studies of chemically denatured proteins to native folding conditions. Most recently, this system has been used as a vehicle for the elucidation of nonnative residual structure in the unfolded form[144] and for the evaluation of the influence of electrostatic interactions on pK_a and protein stability.[145] The N-terminal SH3 domain of the drk protein has proved to be an excellent model system for the NMR study of unfolded states of proteins under nondenaturing conditions.

3.2. Staphylococcal Nuclease: Fragments and Mutations Define an Unfolded State

To examine in detail a denatured state under nondenaturing conditions, a destabilized mutant of

staphylococcal nuclease[146,147] was prepared for NMR study. Later work involved a large fragment, termed Δ131Δ,[148] consisting of residues 1−3 fused to residues 13−140, forming a 131-residue protein in which the native state is destabilized relative to the wild-type protein. Under nondenaturing conditions, the fragment is largely unfolded, according to CD spectra[148] and water-amide proton exchange rates,[91] but residual helical structure is observed in the NMR spectra of the fragment, in rapid equilibrium with extended forms,[148] and the backbone dynamics[53] are inconsistent with a totally random coil polypeptide. The authors reported that the regions of elevated S^2 values, corresponding to regions of reduced backbone mobility, were better correlated with highly hydrophobic portions of the polypeptide, rather than with regions of high intrinsic helical propensity. Nevertheless, the propensity for helix formation was observed in hydrophobic regions that contained helical structure in the native folded state.[53]

Like the drk SH3 domain system, the staphylococcal nuclease system is relatively well-behaved in solution, but problems inherent in the study of unfolded proteins have necessitated creative innovations. The general dearth of long-range structural information was addressed in an innovative manner by Gillespie and Shortle,[65,66] who used covalently attached spin labels to elucidate long-range contacts in Δ131Δ. Ensembles of structures compatible with these restraints were generated, providing insights into the likely conformations present in the ensemble of the unfolded protein.[66] Most significantly, the authors concluded that the denatured nuclease fragment exhibits a nativelike topology in the unfolded state, despite the apparent absence of tertiary structure stabilization by fixed hydrophobic packing, hydrogen bonds, or salt bridges. More recently, Shortle and Ackerman, using dipolar coupling data obtained from partial alignment of the unfolded fragment in strained polyacrylamide gels,[83] reported the observation of a persistence of nativelike topology in the denatured state in 8 M urea.[82] This result was surprising, given the common assumption that high concentrations of denaturants would generate states that correspond to "random coil" for all proteins. These authors suggested that the "nativelike topology" persists even when there is extensive mutagenesis of the Δ131Δ fragment,[149] although the backbone dynamics of these variants did not appear to show any overall conformational restriction consistent with a nativelike topology.[150] Ohnishi and Shortle report the observation of similar dipolar coupling results for short peptides.[151] Is this phenomenon limited to the staphylococcal nuclease system, or will this apparent persistence of nativelike topology under all circumstances be commonly observed? A recent analysis of the origins of residual dipolar couplings in unfolded proteins[152] and analysis of the data obtained for unfolded apomyoglobin[153] (see following section) indicate that these observations may be more related to intrinsic properties of unfolded chains than to propensities for nativelike structure in highly unfolded states.

3.3. Apomyoglobin: Equilibrium and Kinetic Dissection of a Folding Pathway

As well as being amenable to kinetic folding studies, the apomyoglobin system provides a number of excellent equilibrium models for states along the folding pathway, and has been a major focus for several laboratories. A large number of techniques other than NMR have been brought to bear on the apomyoglobin system in recent years,[154−166] giving detailed insights into its folding pathway and partially folded states. In this section, we concentrate on NMR studies of apomyoglobin, which has proved to be one of the best-behaved protein systems for folding studies in solution.

Quench-flow pulse labeling[93,94] has proved a particularly valuable technique for kinetic studies of apomyoglobin folding. An initial study utilizing ^{1}H 2D experiments[167] provided definitive evidence for the similarity of the kinetic intermediate observed by quench-flow pulse labeling and the well-known pH 4 equilibrium intermediate previously identified by CD spectroscopy[168] and amide proton exchange measurements.[85] This result formed the basis for a series of NMR studies aimed at dissecting the folding pathway of apomyoglobin. The quench-flow kinetic studies have been repeated and extended using ^{15}N-labeled apomyoglobin,[169,170] and an extensive set of mutant proteins has been examined,[169,171,172] as well as a variant protein, leghemoglobin, a plant oxygen carrier.[173] All of these proteins, even the evolutionarily distant leghemoglobin, fold using a mechanism that involves a burst phase intermediate. However, the detailed pathways of folding vary widely from protein to protein. For example, the H64F mutant apomyoglobin[169] is more stable than the wild-type protein, and folds more rapidly. These effects are likely due to the added hydrophobicity of the phenylalanine side chain, which stabilizes the E helix of the protein in the kinetic intermediate and the final folded state. The histidine residue at position 64 (the distal histidine) is important in the heme- and oxygen-binding function of myoglobin. Thus, in this case, the presence of the hydrophilic histidine side chain exacts a price in protein stability and folding rate for its importance in function.

NMR studies of equilibrium states of apomyoglobin that model stages in the kinetic folding pathway have yielded a number of insights into structural and dynamic changes that occur as the protein folds. Extensive structural and dynamic characterization has been carried out on five states of apomyoglobin, modeling various stages of polypeptide chain compaction. The C^α secondary chemical shifts ($\delta\Delta = \delta_{\text{observed}} - \delta_{\text{random coil}}$), calculated using sequence-corrected random coil shift values,[35] shown in Figure 4, clearly show an increase in the helical structure of apomyoglobin from solutions in 8 M urea at pH 2.3,[174] pH 2.3,[175] pH 4.1,[176] and pH 6.[177] Further, the location of the helical structure corresponds, in the main, to the locations of ordered helix in the fully folded myoglobin structure. It is clear from Figure 4 that there is very little helical structure in the urea-denatured state. The acid-denatured state contains some propensity for helix, mainly in the regions of

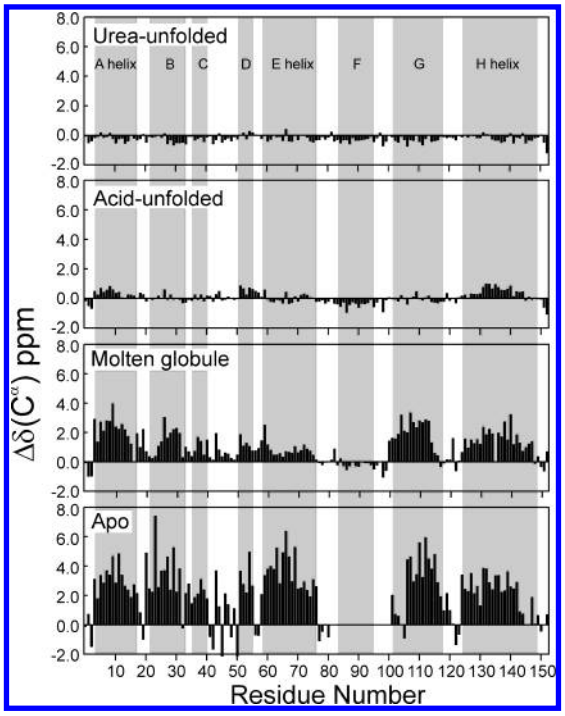


Figure 4. Plot of the secondary $^{13}C^{\alpha}$ chemical shift ($\delta_{observed} - \delta_{random\ coil}$) as a function of residue number for apomyoglobin under solution conditions shown. The urea-unfolded state[174] is in 8 M urea, pH 2.3, 20 °C. The acid-unfolded state[175] is at pH 2.3, 25 °C. The molten globule state[176] is in acetate buffer (10% ethanol) at pH 4.1, 50 °C. The apoprotein (without the heme prosthetic group)[178] is in 10 mM acetate buffer, pH 6.1, 35 °C.

the sequence corresponding to the A and H helices, and in a region that encompasses the end of the D helix and beginning of the E helix. The so-called "molten globule state" at pH 4.1 has helical structure in all of the major helical regions of the folded state, with the exception of the F helix. No signals are observed for the F helix in the folded apomyoglobin at pH 6.1, most likely due to a conformational equilibrium on a time-scale comparable to the chemical shift time scale, resulting in broadening of the resonances beyond detection. It is thought that this conformational equilibrium may be between folded helical forms similar to that in the heme-containing holoprotein and unfolded states: such an equilibrium would be conducive to proper formation of the functional holoprotein by allowing access of the bulky heme prosthetic group to the interior of the molecule.[178]

The completely unfolded protein is modeled by the state in 8 M urea at pH 2.3.[174] Perhaps surprisingly, variations were observed in the relaxation parameters for the backbone of urea-denatured apomyoglobin. These variations were strongly correlated with intrinsic properties of the local amino acid sequence, as illustrated in Figure 5. Clusters of glycine and alanine residues show a greater mobility than the average, while persistence of local hydrophobic interactions causes slight restriction of backbone motions in other regions of the polypeptide. Since these interactions depend only on the most basic attributes of the amino acid sequence, the surface area buried upon folding, they are candidates for the fundamental events that initiate protein folding.[174]

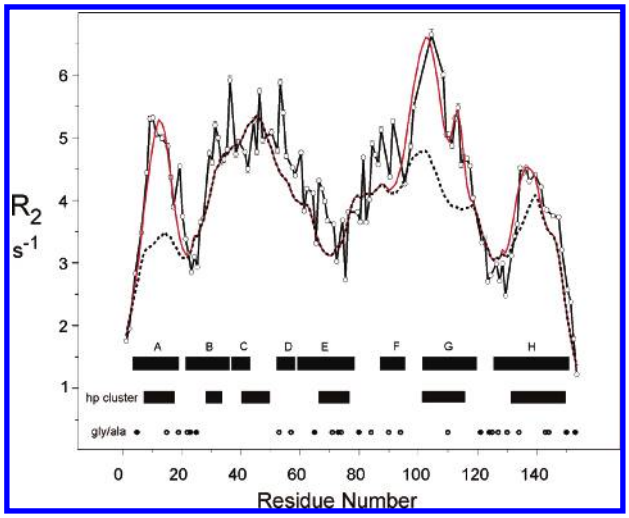


Figure 5. Relaxation rate R_2 for apomyoglobin unfolded in 8 M urea, pH 2.3, 20 °C (open circles joined by thin lines), plotted as a function of residue number. The dashed line shows the R_2 values calculated according to a simple model incorporating only side chain radius of gyration and with persistence length $\lambda_j = 7$ except for Gly and Ala ($\lambda_j = 1$). The red line shows the result of the calculation incorporating both the radius of gyration and four clusters centered at residues 12, 103, 113, and 135. Calculated values have been scaled for comparison with the experimental data. Black bars represent the positions of the helices A−H in the fully folded holoprotein. The positions of hydrophobic clusters (black bars) and of glycine residues (filled circles) and alanine residues (open circles) in the sequence of sperm whale apomyoglobin are also indicated. (Reproduced from ref 174 with permission. Copyright 2002 American Chemical Society).

Propensities for secondary structure formation provide the next level of folding initiation. This is demonstrated by the behavior of apomyoglobin at acid pH (2.3) in the absence of denaturant.[175] A small but significant propensity for helix formation is observed in three parts of the sequence (Figure 4). In addition, motional restriction is observed in these regions, and localized regions of increased R_2 relaxation rate in the A and G helices raised the intriguing possibility that these regions of the protein were in transient contact in the conformational ensemble at pH 2.3.[175] This hypothesis was borne out by the results of experiments in which spin labels were covalently attached to various sites in the molecule.[69] With the spin label in the region corresponding to the H helix in the folded protein, clear evidence of broadening was observed in the region corresponding to the G helix (Figure 3B). By contrast, no broadening was observed in the E helix region, which had shown no increase of R_2 relaxation rate.[175] This observation was corroborated by the results of a similar experiment with the spin label in the E helix: only local broadening was observed at pH 2.3 in this protein (Figure 3A). These results demonstrate not only that there are areas of nativelike secondary structure occurring in this highly unfolded state of apomyoglobin, but that transient nativelike tertiary interactions occur within the conformational ensemble.

The use of strained polyacrylamide gels[77,78] has successfully allowed residual dipolar coupling measurements to be made for unfolded apomyoglobin at pH 2.3 and in 8 M urea at pH 2.3.[153] Significant

residual dipolar couplings were observed for the protein, even in 8 M urea, as was previously seen for staphylococcal nuclease.[82] However, rather than an interpretation invoking persistence of a nativelike topology in the presence of a high concentration of denaturant,[82] the results for unfolded apomyoglobin have been interpreted[153] on the basis of a model that incorporates the notion of "jointed statistical segments"[179,180] that occur in unfolded proteins, but not in folded proteins. This analysis resembles the theoretical treatment of Louhivuori et al.[152] for denatured proteins in liquid crystal media, and is consistent with recent results on ACBP.[181] For folded proteins, the magnitude and direction of the dipolar couplings depend on the orientation of each bond vector to the alignment tensor of the whole molecule. For unfolded proteins, the residual dipolar couplings are a consequence of local conformational restriction of the backbone. The chain has a characteristic persistence length, the length of chain over which successive residues tend to adopt similar (ϕ,ψ) dihedral angles.[180] For several unfolded proteins, the persistence length has been shown by relaxation data to be about seven residues.[174,182] Each of these "statistical segments" is described by its own alignment tensor; these tensors move independently of each other. This interpretation relies only on the polymer-like nature of the unfolded polypeptide chain, and thus appears more feasible than a hypothesis that would invoke specific nativelike topology in high concentrations of chemical denaturant, in which all other measurements conspicuously point to a highly flexible chain and the absence of persistent structure.

3.4. And Amyloid Too? Versatile Lysozyme

Just as lysozyme was one of the earliest systems for the development of NMR techniques for studying folded proteins in solution, it was also the subject of one of the earliest NMR studies of unfolded proteins,[21] and work with lysozyme was instrumental in the development of a number of techniques for the study of unfolded proteins.[110,183] It continues to be utilized extensively today, as much for its inherently interesting properties as for its stability and amenability to experimentation under a wide variety of solution conditions.

Amide exchange in lysozyme was measured to probe the unfolded/folded transition,[90,184] and quench-flow studies showed the presence of folding domains corresponding to the structural domains of the protein.[185−187] Lysozyme was the model protein used in the exploratory NMR studies of the differing effects of denaturants, solvents, calcium ions, heat, and pH.[188−194] In particular, hen lysozyme in trifluoethanol (TFE) and equine lysozyme at pH 2 were found to form partially folded states, which were extensively characterized by NMR. Amide hydrogen exchange measurements of the equine lysozyme at pH 2 showed that the helices of lysozyme differ greatly in their protection, defining a compact cluster in the core of the protein that was identified with the molten globule folding intermediate.[195−197] Resonance assignments for the partly folded state in TFE showed that it contains extensive secondary structure, with different stabilities in different parts of the polypeptide;[198] this information was further amplified by extensive studies of backbone dynamics on this partly unfolded state of lysozyme.[54] Interestingly, the presence of TFE appears to accelerate the folding of hen lysozyme, an indication that the slowest steps of the folding process are associated primarily with the reorganization of hydrophobic side chains in the core of the molten globule, rather than with the formation of hydrogen bonded secondary structure.[199] The folding of lysozyme is complicated by disulfide bridge formation; the reduced protein refolds much more slowly than when the disulfide bridges are left intact, but folding remains highly cooperative.[200]

Both oxidized and reduced forms of lysozyme have been studied by a number of different NMR techniques in highly unfolded states. Complete ^{1}H and ^{15}N resonance assignments for lysozyme in 8 M urea at low pH were utilized to provide a basis for comparison of NOE, coupling constant, and relaxation time data obtained both by NMR and by molecular dynamics simulations.[201] The experimental and theoretical results were generally in good agreement, and indicated that only local interactions were occurring in lysozyme under these conditions. The presence or absence of the disulfide bonds appeared to make little difference to the conformational ensemble in this high concentration of denaturant. The relaxation data were interpreted to indicate the presence of weak hydrophobic clustering in the vicinity of tryptophan side chains. A recent reexamination of the relaxation data, together with mutagenesis data, has led these authors to postulate the presence of long-range interactions in this state of lysozyme.[182]

Lysozyme continues to participate in important advances in the protein-folding field. It provided the means for an exhaustive recent examination of the distributions of side chain rotamers,[202] providing a baseline set of values of great utility in the characterization of protein random coils and more structured states. Lysozyme provided the test case for the use of CIDNP-pulse labeling in probing partially folded states along the folding pathway.[203] Finally, consistent with its continuing role at the forefront of research into protein folding, variants of lysozyme have recently been found to undergo fibrillogenesis, in a reaction analogous to the formation of fibrils in amyloid diseases such as Alzheimer's disease and prion diseases.[204−207]

3.5. Molten Globules − A Challenge for NMR

The molten globule state was first described by Ptitsyn and co-workers on the basis of spectroscopic data of α-lactalbumin.[208] The molten globule was conceived as a state containing nativelike secondary structure but loosened or "molten" tertiary structure. This concept has proved to be a very useful one in the description of folding pathways and intermediates, but it appears that the definition has been stretched in many directions, to encompass a number of different states as different authors study different molecular systems. The study of molten globule states by NMR is exceptionally difficult. This is because molten globule states are fluctional, mostly

on time scales that correspond with intermediate exchange broadening in NMR. Unfolded states also populate many conformational states, but the exchange between them is fast, giving rise to (usually) a single set of resonances corresponding to an ensemble average. Slow exchange between multiple states gives rise to multiple sets of resonances, which can be deconvoluted and separated if necessary (as in the case, for example, of the drk SH3 domain, see section 3.1). However, when states are in intermediate exchange, the resonances are broadened, sometimes beyond detection. Even if the resonances can be observed, their line width is such that resonance overlap becomes very much worse than it would otherwise be, while the intensity of the signal is correspondingly decreased. These problems were seen in early NMR studies of the molten globule state of α-lactalbumin.[84] A number of clever techniques were developed to obtain information on the α-lactalbumin molten globule, using amide trapping experiments and subsequent NMR experiments on the native state of the protein,[84,209] peptide studies coupled with 2D ^{1}H NMR experiments,[210] and R_2 measurements.[211] Denaturant-induced unfolding of the α-lactalbumin molten globule was followed by 2D ^{1}H-^{15}N HSQC spectra, providing residue-specific information on a highly stable core that is present in the human protein under strongly denaturing conditions, but not in the homologous bovine protein.[212] Extremes of pH,[213] temperature,[214] and pressure,[97] as well as site-directed mutagenesis[215] have been used to access excited states approximating the α-lactalbumin molten globule, and have enabled excellent NMR data to be obtained for this difficult system.

The molten globule state of apomyoglobin has proved to be a much more straightforward subject for NMR studies.[176,177] While it is not easy to work with, and requires relatively high temperatures (50 °C) for optimal NMR spectra and the presence of a small amount of cosolvent (ethanol, 10%) to protect against aggregation during acquisition of multidimensional NMR spectra,[216] double- and triple-labeled samples of the apomyoglobin molten globule state at pH 4.2 give excellent triple-resonance spectra, from which almost all of the resonances can be assigned. Chemical shift deviations from random coil, together with relaxation measurements, delineated a picture of a classic molten globule state: most of the helices in the molecule populate helical secondary structure significantly, but fixed tertiary structure does not form. Compared to the fully folded native state, the helices range in population from about 70% for each of the A, G, and H helices to about 30% for the D and E helices. All of the helices are significantly frayed at the ends. Mobility of the polypeptide backbone is greatest in the areas that contained the least secondary structure, and ^{15}N spin relaxation data show evidence for packing of the A, G, and H helices, and part of the B helix to form a compact hydrophobic core. The picture that emerges is of a molten globule state that varies in stability throughout the polypeptide. The remaining helices populate a significant amount of helical structure, but are more loosely associated with the better-structured core.

Extensive NMR studies have also been performed for the well-behaved A-state of ubiquitin, which is formed at low pH in mixed methanol−water. A comparison of the slowly exchanging amide protons in native and A-states[217] showed that the overall topology of the A-state remains similar to that of the native state, while certain regions of the protein, particularly in the central β-sheet, retain the same amide hydrogen bonding interactions as the native form. According to heteronuclear NMR measurements, only the first two strands of the β-sheet, together with a few turns of helix, remain fully structured in the A-state,[218] constituting an N-terminal folded subdomain. The C-terminal half of the molecule undergoes a methanol-induced transition to a poorly structured, highly dynamic state.[219] More recently, the contributions of non-native structure to the A-state has been evaluated by site-directed mutagenesis and NMR.[220] Interestingly, the mutation of a residue in the N-terminal region of the protein from a sequence favoring a Type I turn to that favoring a Type II turn showed a preference for Type II turn in isolated peptides (representing the fully unfolded state of ubiquitin), but when the mutation was incorporated into the full-length protein, the structural preference in the A-state was for the Type I turn characteristic of the native protein, indicating that the structural specificity and stability of the A-state are modulated to a great extent by nonlocal, long-range interactions.

Other molten globule states studied by NMR include a CheY mutant,[221] cystatins,[222] and carbonic anhydrase.[223]

4. NMR Studies of Intrinsically Unstructured, Functional Proteins

NMR has been instrumental in identifying and characterizing unfolded and partly folded protein domains that are functional. This field is rapidly expanding. We provide three examples to show the range of contributions that the NMR method has made in this area. Intrinsically unstructured, functional proteins show different degrees of disorder in their native states. The first example to be discussed is a case in which the individual domains of a DNA-binding protein are connected by flexible linkers and tumble independently in the absence of DNA, but adopt a rigid, ordered structure in complex with DNA, providing a mechanism for high-affinity, sequence-specific binding. The second example is a case in which a protein is unfolded in isolation but folds when bound to its (folded) interaction partner. This is a very common pattern for the interactions of unstructured proteins, including small peptide hormones. The third example shows the case of two unstructured proteins that are mutually folded when they interact.

4.1. Sequence Specific DNA Binding by Zinc Finger Proteins: The "Snap-Lock"

The DNA-binding interactions of zinc finger proteins provide an excellent example of the interplay of structure and flexibility in the promotion of

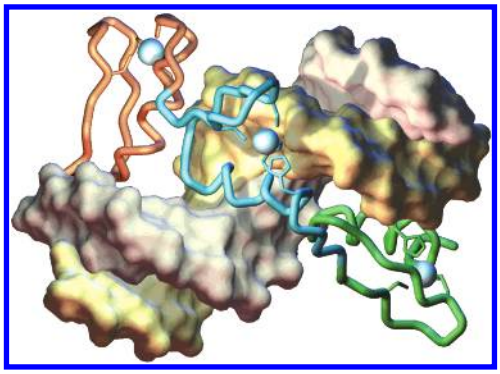


Figure 6. The three amino-terminal zinc fingers of transcription factor IIIA bound to DNA, as determined by NMR spectroscopy.[230] DNA strands are shown in white and yellow. Zinc finger 1 is colored orange, finger 2 is cyan, and finger 3 is green. (Adapted from ref 230 with permission. Copyright 1997 Elsevier).

sequence-specific binding. Zinc fingers are small zinc-binding motifs first identified as a repeating homologous sequence in the transcription factor TFIIIA, the factor specific for the internal control region of the 5S RNA gene.[224] In the absence of zinc, the finger is unstructured;[225] the structure of the individual zinc finger is stabilized by coordination of a zinc ion to four ligands (generally a combination of Cys and His side chains) in a tetrahedral arrangement.[226,227] Sequence-specific DNA recognition is achieved by the concerted interaction of several fingers, connected by short but highly conserved linker sequences.[228] Structures of tandem zinc finger proteins bound to DNA, for example, zif268[229] and TFIIIA,[230] show a characteristic wrapping of the fingers around the DNA, making base-specific contacts in the major groove (Figure 6). It is clear from this mode of binding that the structure of the free zinc finger protein must differ in important ways in solution and in the DNA complex, since the DNA forms a central, integral part of the structure of the complex. The polypeptide chain dynamics of a construct containing three zinc fingers showed that the individual fingers behave largely as "beads on a string", with minimal interaction between them, but with long-range motional restriction mandated by the anisotropy of the molecule.[231] The role of the conserved linker sequences was recently explored by comparing the NMR spectra and dynamics of zinc finger proteins free and in complex with DNA.[3] A comparison of the superposition of the linker sequences of 12 structures of a construct containing the first three zinc fingers of TFIIIA, free in solution (X. Liao and P. E. W., unpublished data) and complexed to the cognate DNA sequence[230] is shown in Figure 7. It appears that the linker provides in the complex but not in the free protein a stabilizing capping motif, complete with a side chain−backbone hydrogen bond, for the helix of the N-terminal zinc finger. Heteronuclear ^{1}H-^{15}N NOE measurements (Figure 8) show that the linker sequences have

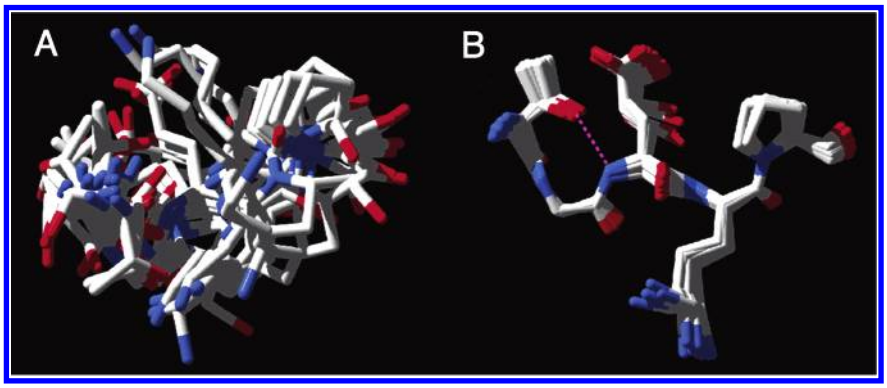


Figure 7. Comparison of the $^{13}C^{\alpha}$ chemical shifts of the protein in the DNA complex and free in solution ($\Delta\delta^{13}C^{\alpha}_{bound-free}$) for the α-helical regions of the four fingers of the DNA-binding isoform of the Wilms' tumor zinc finger transcription factor. A schematic representation of the corresponding hydrogen bonding pattern is shown below the bar plot for each finger. For each finger, the hydrogen bonds between residues that are both in regular helix are shown underneath the amino acid sequence, while hydrogen bonds that include one or two residues outside the regular helix (i.e., the capping hydrogen bonds) are shown above the sequence. The hydrogen bonds that exist in both the free and the DNA-bound forms of the proteins are represented with solid lines. The DNA-induced C-capping hydrogen bonds for each finger are represented by dashed lines. (Adapted from ref 3 with permission. Copyright 2000 Elsevier).

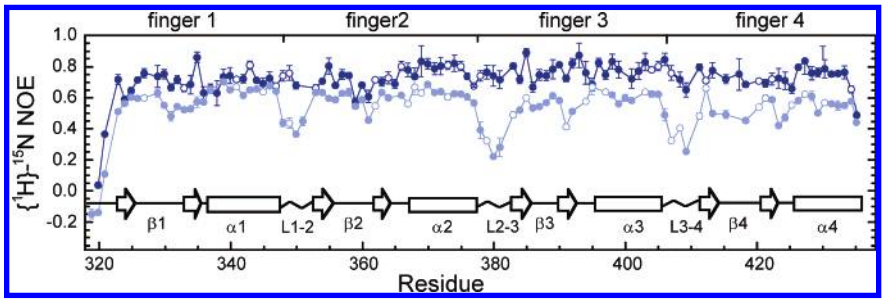


Figure 8. Backbone ^{1}H-^{15}N NOE values as a function of residue number for the four fingers of the DNA-binding isoform of the Wilms' tumor zinc finger transcription factor, free (light symbols) and DNA bound (dark symbols). The β-sheet (open arrows), α-helix (open rectangles), linker regions (zigzag lines) in the zinc finger domain of WT1 are indicated schematically.

considerable flexibility in the free form of the protein, but are rigidified to the same extent as the folded zinc finger domains in the DNA complex. Thus, one might envision a mechanism for the binding of tandem zinc fingers to DNA consisting initially of nonspecific interaction involving primarily the sugar and phosphate moieties. The protein, with individual zinc fingers structured but segmentally disordered, is able to search along the DNA for the correct base sequence because the nonspecific association is weak. Once the correct base sequence is encountered, base-specific contacts are initiated in the major groove, followed by the formation of the helix capping interaction in the linker sequence. This concept of the formation of a specific structure in a flexible protein in response to binding to the correct base sequence has been termed the "snap-lock" mechanism.[3]

4.2. Folding of Unstructured Proteins upon Binding

A number of systems have been shown to contain components that are unstructured in isolation, but which fold to defined structures upon binding to physiological targets. This area has recently been reviewed.[14] The protein CBP (CREB-binding protein) and its homologue p300 are large proteins that contain both structured domains and long intrinsically unstructured regions. Structural and dynamic information for these domains and their complexes with physiological ligands have been obtained almost exclusively by NMR. NMR studies show that a number of the domains are relatively well-folded in isolation in the absence of binding partners.[8,232,233]

The TAZ domains of CBP/p300 are zinc-containing domains with a distinctive fold. Each of the three zinc binding motifs consists of two helices joined by a connecting loop, with one histidine and three cysteine ligands that are arranged sequentially to form an HCCC-type zinc-binding motif.[232] TAZ1 and TAZ2 share significant sequence homology, but bind a different subset of transcription factors.[234] The structures of the two domains are quite similar, but differ significantly in the third zinc site.[233] A number of TAZ domain ligands appear to be unstructured in the free state, for example, the minimal binding domain of the hypoxia-inducible transcription factor HIF-1α is unstructured in solution, but binds with nanomolar affinity to TAZ1.[233,235] This is indicated by the dispersion of the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled HIF-1α free and in complex with TAZ1 (Figure 9a). The structure of the complex (Figure 9b) illustrates a likely reason for the necessity for HIF-1α to be unstructured in the free state: the ligand polypeptide is wound around the TAZ1 core to give maximal binding surface area. It is clear that a stable tertiary fold in free HIF-1α would not be conducive to complex formation.

4.3. Mutual Synergistic Folding of Two Unstructured Domains

The response to nuclear hormones occurs via nuclear hormone receptors that regulate gene expression, and is mediated by recruitment of coactivators such as p160, which recruits the general coactivator CBP. The nuclear receptor coactivator binding domain of CBP is located at the C-terminus and is intrinsically unstructured in isolation;[236] its spectroscopic properties and heat capacity are characteristic of a molten globule state. The interaction region of its binding partner, activator for thyroid hormone and retinoid receptors (ACTR), is also unstructured in solution.[236] When the two domains interact, the NMR spectra (Figure 10a) clearly show that both domains are well-folded. Solution structure determination of the complex[236] revealed an intertwined helical bundle (Figure 10b) with an extensive leucine-rich hydrophobic core. Within the complex, the three helices of the CBP domain form a small

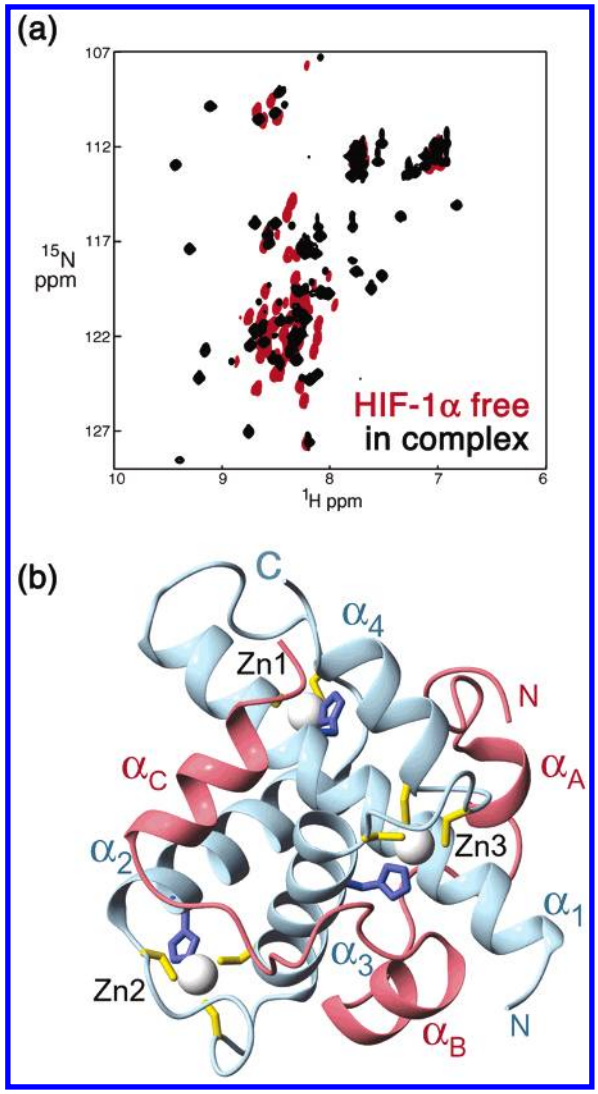


Figure 9. (a) 600 MHz ^{1}H-^{15}N HSQC spectrum of HIF-1α (776−826) free (red) and bound to unlabeled TAZ1 (black). The spectrum of the free HIF-1α domain shows the limited ^{1}H resonance dispersion characteristic of unstructured proteins. (Adapted from reference 233 with permission. Copyright 2002. National Academy of Sciences of the United States of America). (b) Ribbon representation of the structure of a single structure from the family of NMR solution structures of the TAZ1-HIF-1α complex. TAZ1 is shown in blue, with zinc ions as silver balls. Zinc ligands are shown in blue (His) and yellow (Cys). HIF-1α is shown in red. (Adapted from ref 233 with permission. Copyright 2002. National Academy of Sciences of the United States of America).

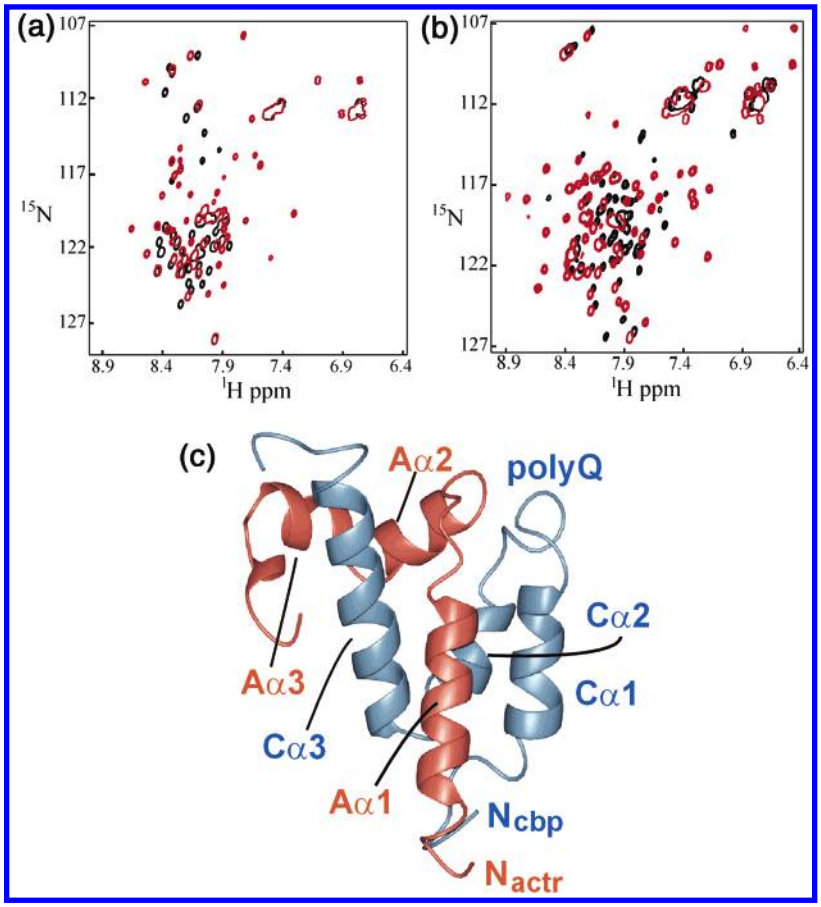


Figure 10. ^{15}N-HSQC spectra of (a) ^{15}N-ACTR:1018-1088 free (black) and in the presence of excess unlabeled CBP:2059−2117 (red). (b) ^{15}N-CBP:2059-2152 free (black) and in the presence of excess unlabeled ACTR:1018-1088 (red). (Reproduced with permission from *Nature* (http://www.nature.com), ref 236. Copyright 2002 Nature Publishing Group). (c) Ribbon representation of the solution structure of the ACTR/CBP complex. ACTR is pink and CBP blue. (Reproduced from *Nature* (http://www.nature.com), ref 236. Copyright 2002 Nature Publishing Group).

intramolecular hydrophobic core, but there are no long-range intramolecular contacts in the ACTR domain. These observations demonstrate that the stability of the complex is dependent largely, if not completely, on the intermolecular interactions that are potentiated when the two domains fold mutually and synergistically.

5. Conclusion

This review has attempted to summarize some of the many contributions of NMR studies to our knowledge of unfolded and partly folded proteins. Structural and dynamic characterization of intrinsically unstructured, functional proteins both free and in complex with their binding partners is an important area that we anticipate will receive a great deal of attention in the future, as the products of unknown genes from published genomes are expressed and characterized. As well, the process of folding of proteins that are well structured in their functional state remains a hurdle in our understanding of cellular processes. Indeed, the correct folding of proteins is a most important process, which appears to go awry as the organism ages. Misfolding of proteins too is a harbinger of disease. We need to know as much as we can about these processes. It is clear that, because of the unique information it provides in solution under physiologically relevant conditions, NMR will continue to play a pivotal role in the elucidation of the structure and function of these important systems.

6. Acknowledgment

We are grateful to all members of the Wright/Dyson lab, past and present, for their contributions to this research, which was supported by a number of grants from the National Institutes of Health, primarily GM57374 (H.J.D.) and DK34909, GM36643, and CA96865 (P.E.W.).

7. References

(1) Karush, F. *J. Am. Chem. Soc.* **1950**, *72*, 2705.
(2) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355.
(3) Laity, J. H.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2000**, *295*, 719.
(4) Young, M. A.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. *Cell* **2001**, *105*, 115.
(5) Tainer, J. A.; Getzoff, E. D.; Alexander, H.; Houghten, R. A.; Olson, A. J.; Lerner, R. A.; Hendrickson, W. A. *Nature* **1984**, *312*, 127.
(6) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321.
(7) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K. *Pac. Symp. Biocomput.* **1998**, *3*, 437.
(8) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741.
(9) Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11504.

(10) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573.
(11) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(12) Uversky, V. N. *Eur. J. Biochem.* **2002**, *269*, 2.
(13) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739.
(14) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54.
(15) Dyson, H. J.; Wright, P. E. *Adv. Protein Chem.* **2002**, *62*, 311.
(16) Bright, J. N.; Woolf, T. B.; Hoh, J. H. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 131.
(17) Barbar, E. *Biopolymers* **1999**, *51*, 191.
(18) Dyson, H. J.; Wright, P. E. *Annu. Rev. Phys. Chem.* **1996**, *47*, 369.
(19) Braun, W.; Bösch, C.; Brown, L. R.; Go, N.; Wüthrich, K. *Biochim. Biophys. Acta* **1981**, *667*, 377.
(20) Bösch, C.; Bundi, A.; Oppliger, M.; Wüthrich, K. *Eur. J. Biochem.* **1978**, *91*, 209.
(21) Dobson, C. M.; Evans, P. A.; Williamson, K. L. *FEBS Lett.* **1984**, *168*, 331.
(22) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559.
(23) Zhan, Y.; Song, X.; Zhou, G. W. *Gene* **2001**, *281*, 1.
(24) Buck, M. *Structure* **2003**, *11*, 735.
(25) Loris, R.; Marianovsky, I.; Lah, J.; Laeremans, T.; Engelberg-Kulka, H.; Glaser, G.; Muyldermans, S.; Wyns, L. *J. Biol. Chem.* **2003**, *278*, 28252.
(26) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011.
(27) Paci, E.; Vendruscolo, M.; Dobson, C. M.; Karplus, M. *J. Mol. Biol.* **2002**, *324*, 151.
(28) Barron, L. D.; Blanch, E. W.; Hecht, L. *Adv. Protein Chem.* **2002**, *62*, 51.
(29) Frieden, C.; Chattopadhyay, K.; Elson, E. L. *Adv. Protein Chem.* **2002**, *62*, 91.
(30) Keiderling, T. A.; Xu, Q. *Adv. Protein Chem.* **2002**, *62*, 111.
(31) Millett, I. S.; Doniach, S.; Plaxco, K. W. *Adv. Protein Chem.* **2002**, *62*, 262.
(32) Juneja, J.; Udgaonkar, J. B. *Curr. Sci.* **2003**, *84*, 157.
(33) Brockwell, D. J.; Smith, D. A.; Radford, S. E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 16.
(34) Yao, J.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1997**, *419*, 285.
(35) Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J. Am. Chem. Soc.* **2001**, *123*, 2970.
(36) Wang, Y.; Jardetzky, O. *J Am. Chem. Soc.* **2002**, *124*, 14075.
(37) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661.
(38) West, N. J.; Smith, L. J. *J. Mol. Biol.* **1998**, *280*, 867.
(39) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285.
(40) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311.
(41) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363.
(42) Merutka, G.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **1995**, *5*, 14.
(43) Braun, D.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1994**, *116*, 8466.
(44) Tollinger, M.; Skrynnikov, N. R.; Mulder, F. A.; Forman-Kay, J. D.; Kay, L. E. *J Am. Chem. Soc.* **2001**, *123*, 11341.
(45) Prompers, J. J.; Bruschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.
(46) Yang, D. W.; Mok, Y. K.; Muhandiram, D. R.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 3555.
(47) Meekhof, A. E.; Freund, S. M. V. *J. Mol. Biol.* **1999**, *286*, 579.
(48) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439.
(49) Skrynnikov, N. R.; Millet, O.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6449.
(50) Yang, D. W.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. *J. Mol. Biol.* **1998**, *276*, 939.
(51) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559.
(52) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(53) Alexandrescu, A. T.; Shortle, D. *J. Mol. Biol.* **1994**, *242*, 527.
(54) Buck, M.; Schwalbe, H.; Dobson, C. M. *J. Mol. Biol.* **1996**, *257*, 669.
(55) Peng, J. W.; Wagner, G. *J. Magn. Reson.* **1992**, *98*, 308.
(56) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. *J. Biomol. NMR* **1995**, *6*, 153.
(57) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **1999**, *121*, 8671.
(58) Buevich, A. V.; Shinde, U. P.; Inouye, M.; Baum, J. *J. Biomol. NMR* **2001**, *20*, 233.
(59) Zhang, O.; Kay, L. E.; Shortle, D.; Forman-Kay, J. D. *J. Mol. Biol.* **1997**, *272*, 9.
(60) Zhang, O.; Forman-Kay, J. D.; Shortle, D.; Kay, L. E. *J. Biomol. NMR* **1997**, *9*, 181.
(61) Mok, Y. K.; Kay, C. M.; Kay, L. E.; Forman-Kay, J. *J. Mol. Biol.* **1999**, *289*, 619.
(62) Crowhurst, K. A.; Choy, W. Y.; Mok, Y. K.; Forman-Kay, J. D. *J. Mol. Biol.* **2003**, *329*, 185.
(63) Schmidt, P. G.; Kuntz, I. D. *Biochemistry* **1984**, *23*, 4261.
(64) Kosen, P. A.; Scheek, R. M.; Naderi, H.; Basus, V. J.; Manogaran, S.; Schmidt, P. G.; Oppenheimer, N. J.; Kuntz, I. D. *Biochemistry* **1986**, *25*, 2356.
(65) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158.
(66) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170.
(67) Yi, Q.; Scalley-Kim, M. L.; Alm, E. J.; Baker, D. *J. Mol. Biol.* **2000**, *299*, 1341.
(68) Teilum, K.; Kragelund, B. B.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *324*, 349.
(69) Lietzow, M. A.; Jamin, M.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2002**, *322*, 655.
(70) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279.
(71) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111.
(72) Tolman, J. R. *Curr. Opin. Struct. Biol.* **2001**, *11*, 532.
(73) MacDonald, D.; Lu, P. *Curr. Opin. Struct. Biol.* **2002**, *12*, 343.
(74) Bax, A. *Protein Sci.* **2003**, *12*, 1.
(75) Prestegard, J. H. *Nat. Struct. Biol.* **1998**, *5*, 517.
(76) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571.
(77) Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303.
(78) Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340.
(79) Tsui, V.; Zhu, L.; Huang, T. H.; Wright, P. E.; Case, D. A. *J. Biomol. NMR* **2000**, *16*, 9.
(80) Al Hashimi, H. M.; Valafar, H.; Terrell, M.; Zartler, E. R.; Eidsness, M. K.; Prestegard, J. H. *J. Magn Reson.* **2000**, *143*, 402.
(81) Wu, Z.; Delaglio, F.; Tjandra, N.; Zhurkin, V. B.; Bax, A. *J. Biomol. NMR* **2003**, *26*, 297.
(82) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487.
(83) Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 3089.
(84) Baum, J.; Dobson, C. M.; Evans, P. A.; Hanley, C. *Biochemistry* **1989**, *28*, 7.
(85) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. *Science* **1990**, *249*, 1544.
(86) Jeng, M.-F.; Englander, S. W.; Elöve, G. A.; Wand, A. J.; Roder, H. *Biochemistry* **1990**, *29*, 10433.
(87) Robertson, A. D.; Baldwin, R. L. *Biochemistry* **1991**, *30*, 9907.
(88) Kim, K.-S.; Fuchs, J. A.; Woodward, C. K. *Biochemistry* **1993**, *32*, 9600.
(89) Kim, K.-S.; Woodward, C. *Biochemistry* **1993**, *32*, 9609.
(90) Buck, M.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1994**, *237*, 247.
(91) Mori, S.; van Zijl, P. C. M.; Shortle, D. *Proteins* **1997**, *28*, 325.
(92) Chung, E. W.; Nettleton, E. J.; Morgan, C. J.; Gross, M.; Miranker, A.; Radford, S. E.; Dobson, C. M.; Robinson, C. V. *Protein Sci.* **1997**, *6*, 1316.
(93) Udgaonkar, J. B.; Baldwin, R. L. *Nature* **1988**, *335*, 694.
(94) Roder, H.; Elöve, G. A.; Englander, S. W. *Nature* **1988**, *335*, 700.
(95) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896.
(96) Refaee, M.; Tezuka, T.; Akasaka, K.; Williamson, M. P. *J. Mol. Biol.* **2003**, *327*, 857.
(97) Lassalle, M. W.; Li, H.; Yamada, H.; Akasaka, K.; Redfield, C. *Protein Sci.* **2003**, *12*, 66.
(98) Kitahara, R.; Yamada, H.; Akasaka, K.; Wright, P. E. *J. Mol. Biol.* **2002**, *320*, 311.
(99) Kitahara, R.; Akasaka, K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3167.
(100) Akasaka, K. *Pure Appl. Chem.* **2003**, *75*, 927.
(101) Akasaka, K. *Biochemistry* **2003**, *42*, 10875.
(102) Balbach, J.; Forge, V.; van Nuland, N. A. J.; Winder, S. L.; Hore, P. J.; Dobson, C. M. *Nat. Struct. Biol.* **1995**, *2*, 865.
(103) Balbach, J.; Forge, V.; Lau, W. S.; van Nuland, N. A. J.; Brew, K.; Dobson, C. M. *Science* **1996**, *274*, 1161.
(104) Balbach, J.; Forge, V.; Lau, W. S.; Jones, J. A.; van Nuland, N. A. J.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7182.
(105) Balbach, J. *J. Am. Chem. Soc.* **2000**, *122*, 5887.
(106) Buevich, A. V.; Dai, Q. H.; Liu, X.; Brodsky, B.; Baum, J. *Biochemistry* **2000**, *39*, 4299.
(107) Baum, J.; Brodsky, B. *Folding Des.* **1997**, *2*, R53−R60.
(108) Mizuguchi, M.; Kroon, G. J.; Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **2003**, *328*, 1161.
(109) Killick, T. R.; Freund, S. M. V.; Fersht, A. R. *Protein Sci.* **1999**, *8*, 1286.
(110) Broadhurst, R. W.; Dobson, C. M.; Hore, P. J.; Radford, S. E.; Rees, M. L. *Biochemistry* **1991**, *30*, 405.
(111) Lyon, C. E.; Jones, J. A.; Redfield, C.; Dobson, C. M.; Hore, P. J. *J. Am. Chem. Soc.* **1999**, *121*, 6505.
(112) Wirmer, J.; Kuhn, J.; Schwalbe, H. *Angew. Chem., Int. Ed.* **2001**, *40*, 4248.
(113) Canet, D.; Lyon, C. E.; Scheek, R. M.; Robillard, G. T.; Dobson, C. M.; Hore, P. J.; van Nuland, N. A. J. *J. Mol. Biol.* **2003**, *330*, 397.
(114) van Nuland, N. A. J.; Forge, V.; Balbach, J.; Dobson, C. M. *Acc. Chem. Res.* **1998**, *31*, 773.
(115) Dobson, C. M.; Hore, P. J. *Nat. Struct. Biol.* **1998**, *5*, 504.

(116) Mok, K. H.; Nagashima, T.; Day, I. J.; Jones, J. A.; Jones, C. J.; Dobson, C. M.; Hore, P. J. *J. Am. Chem. Soc.* **2003**, *125*, 12484.
(117) Jones, J. A.; Wilkins, D. K.; Smith, L. J.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 199.
(118) Chen, L. L.; Hodgson, K. O.; Doniach, S. *J. Mol. Biol.* **1996**, *261*, 658.
(119) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424.
(120) Choy, W. Y.; Mulder, F. A. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101.
(121) Pan, H.; Barany, G.; Woodward, C. *Protein Sci.* **1997**, *6*, 1985.
(122) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 7566.
(123) Ilyina, E.; Roongta, V.; Pan, H.; Woodward, C.; Mayo, K. H. *Biochemistry* **1997**, *36*, 3383.
(124) Mayo, K. H.; Ilyina, E. *Protein Sci.* **1998**, *7*, 358.
(125) Eder, J.; Fersht, A. R. *Mol. Microbiol.* **1995**, *16*, 609.
(126) Grantcharova, V.; Alm, E. J.; Baker, D.; Horwich, A. L. *Curr. Opin. Struct. Biol.* **2001**, *11*, 70.
(127) Huang, G. S.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6878.
(128) Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Fullbright, P. W.; Oas, T. G. *J. Mol. Biol.* **1996**, *263*, 311.
(129) Vugmeyster, L.; Kroenke, C. D.; Picart, F.; Palmer, A. G.; Raleigh, D. P. *J. Am. Chem. Soc.* **2000**, *122*, 5387.
(130) Spector, S.; Raleigh, D. P. *J. Mol. Biol.* **1999**, *293*, 763.
(131) Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 122.
(132) Tanford, C. *Adv. Protein Chem.* **1970**, *24*, 1.
(133) Logan, T. M.; Thériault, Y.; Fesik, S. W. *J. Mol. Biol.* **1994**, *236*, 637.
(134) Arcus, V. L.; Vuilleumier, S.; Freund, S. M. V.; Bycroft, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9412.
(135) Arcus, V. L.; Vuilleumier, S.; Freund, S. M. V.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 305.
(136) Wong, K. B.; Clarke, J.; Bond, C. J.; Neira, J. L.; Freund, S. M. V.; Fersht, A. R.; Daggett, V. *J. Mol. Biol.* **2000**, *296*, 1257.
(137) Sari, N.; Alexander, P.; Bryan, P. N.; Orban, J. *Biochemistry* **2000**, *39*, 965.
(138) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1995**, *34*, 868.
(139) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1995**, *34*, 6784.
(140) Zhang, O.; Kay, L. E.; Olivier, J. P.; Forman-Kay, J. D. *J. Biomol. NMR* **1994**, *4*, 845.
(141) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *J. Biomol. NMR* **1994**, *4*, 727.
(142) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1997**, *36*, 2390.
(143) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1997**, *36*, 3959.
(144) Crowhurst, K. A.; Tollinger, M.; Forman-Kay, J. D. *J. Mol. Biol.* **2002**, *322*, 163.
(145) Tollinger, M.; Crowhurst, K. A.; Kay, L. E.; Forman-Kay, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4545.
(146) Shortle, D.; Meeker, A. K. *Biochemistry* **1989**, *28*, 936.
(147) Shortle, D.; Abeygunawardana, C. *Structure* **1993**, *1*, 121.
(148) Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, *33*, 1063.
(149) Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 13791.
(150) Ohnishi, S.; Shortle, D. *Protein Sci.* **2003**, *12*, 1530.
(151) Ohnishi, S.; Shortle, D. *Proteins* **2003**, *50*, 546.
(152) Louhivuori, M.; Paakkonen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647.
(153) Borges, R. M.; Goto, N. K.; Kroon, G.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, in press.
(154) Jamin, M.; Geierstanger, B.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6127.
(155) Jamin, M.; Antalik, M.; Loh, S. N.; Bolen, D. W.; Baldwin, R. L. *Protein Sci.* **2000**, *9*, 1340.
(156) Jamin, M.; Yeh, S. R.; Rousseau, D. L.; Baldwin, R. L. *J. Mol. Biol.* **1999**, *292*, 731.
(157) Jamin, M.; Baldwin, R. L. *J. Mol. Biol.* **1998**, *276*, 491.
(158) Geierstanger, B.; Jamin, M.; Volkman, B. F.; Baldwin, R. L. *Biochemistry* **1998**, *37*, 4254.
(159) Jamin, M.; Baldwin, R. L. *Nat. Struct. Biol.* **1996**, *3*, 613.
(160) Abbruzzetti, S.; Crema, E.; Masino, L.; Vecli, A.; Viappiani, C.; Small, J. R.; Libertini, L. J.; Small, E. W. *Biophys. J.* **2000**, *78*, 405.
(161) Bondos, S. E.; Sligar, S.; Jonas, J. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2000**, *1480*, 353.
(162) Chance, M. R. *Biochem. Biophys. Res. Commun.* **2001**, *287*, 614.
(163) Gulotta, M.; Gilmanshin, R.; Buscher, T. C.; Callender, R. H.; Dyer, R. B. *Biochemistry* **2001**, *40*, 5137.
(164) Gilmanshin, R.; Gulotta, M.; Dyer, R. B.; Callender, R. H. *Biochemistry* **2001**, *40*, 5127.
(165) Tcherkasskaya, O.; Uversky, V. N. *Proteins* **2001**, *44*, 244.
(166) Kamei, T.; Oobatake, M.; Suzuki, M. *Biophys. J.* **2002**, *82*, 418.
(167) Jennings, P. A.; Wright, P. E. *Science* **1993**, *262*, 892.
(168) Griko, Y. V.; Privalov, P. L.; Venyaminov, S. Y.; Kutyshenko, V. P. *J. Mol. Biol.* **1988**, *202*, 127.
(169) Garcia, C.; Nishimura, C.; Cavagnero, S.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2000**, *39*, 11227.
(170) Nishimura, C.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2002**, *322*, 483.
(171) Cavagnero, S.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **1999**, *285*, 269.
(172) Nishimura, C.; Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **2003**, *334*, 293.
(173) Nishimura, C.; Prytulla, S.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **2000**, *7*, 679.
(174) Schwarzinger, S.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2002**, *41*, 12681.
(175) Yao, J.; Chung, J.; Eliezer, D.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2001**, *40*, 3561.
(176) Eliezer, D.; Chung, J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2000**, *39*, 2894.
(177) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **1998**, *5*, 148.
(178) Eliezer, D.; Wright, P. E. *J. Mol. Biol.* **1996**, *263*, 531.
(179) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W. H. Freeman & Company: New York, 1980.
(180) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley: New York, 1969.
(181) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, in press.
(182) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.
(183) Evans, P. A.; Topping, K. D.; Woolfson, D. N.; Dobson, C. M. *Proteins* **1991**, *9*, 248.
(184) Radford, S. E.; Buck, M.; Topping, K. D.; Dobson, C. M.; Evans, P. A. *Proteins* **1992**, *14*, 237.
(185) Miranker, A.; Radford, S. E.; Karplus, M.; Dobson, C. M. *Nature* **1991**, *349*, 633.
(186) Bieri, O.; Kiefhaber, T. *J. Mol. Biol.* **2001**, *310*, 919.
(187) Bai, P.; Peng, Z. *J. Mol. Biol.* **2001**, *314*, 321.
(188) Lumb, K. J.; Dobson, C. M. *J. Mol. Biol.* **1992**, *227*, 9.
(189) Lumb, K. J.; Cheetham, J. C.; Dobson, C. M. *J. Mol. Biol.* **1994**, *235*, 1072.
(190) Buck, M.; Radford, S. E.; Dobson, C. M. *Biochemistry* **1993**, *32*, 669.
(191) Van Dael, H.; Haezebrouck, P.; Morozova, L.; Arico-Muendel, C.; Dobson, C. M. *Biochemistry* **1993**, *32*, 11886.
(192) Kotik, M.; Radford, S. E.; Dobson, C. M. *Biochemistry* **1995**, *34*, 1714.
(193) Haezebrouck, P.; Joniau, M.; Van Dael, H.; Hooke, S. D.; Woodruff, N. D.; Dobson, C. M. *J. Mol. Biol.* **1995**, *246*, 382.
(194) Bhattacharjya, S.; Balaram, P. *Proteins* **1997**, *29*, 492.
(195) Morozova, L. A.; Haynie, D. T.; Arico-Muendel, C.; Van Dael, H.; Dobson, C. M. *Nat. Struct. Biol.* **1995**, *2*, 871.
(196) Morozova-Roche, L. A.; Arico-Muendel, C. C.; Haynie, D. T.; Emelyanenko, V. I.; Van Dael, H.; Dobson, C. M. *J. Mol. Biol.* **1997**, *268*, 903.
(197) Morozova-Roche, L. A.; Jones, J. A.; Noppe, W.; Dobson, C. M. *J. Mol. Biol.* **1999**, *289*, 1055.
(198) Buck, M.; Schwalbe, H.; Dobson, C. M. *Biochemistry* **1995**, *34*, 13219.
(199) Lu, H.; Buck, M.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1997**, *265*, 112.
(200) Roux, P.; Delepierre, M.; Goldberg, M. E.; Chaffotte, A. F. *J. Biol. Chem.* **1997**, *272*, 24843.
(201) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977.
(202) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288*, 705.
(203) Lyon, C. E.; Suh, E. S.; Dobson, C. M.; Hore, P. J. *J. Am. Chem. Soc.* **2002**, *124*, 13018.
(204) Booth, D. R.; Sunde, M.; Bellotti, V.; Robinson, C. V.; Hutchinson, W. L.; Fraser, P. E.; Hawkins, P. N.; Dobson, C. M.; Radford, S. E.; Blake, C. C. F.; Pepys, M. B. *Nature* **1997**, *385*, 787.
(205) Canet, D.; Sunde, M.; Last, A. M.; Miranker, A.; Spencer, A.; Robinson, C. V.; Dobson, C. M. *Biochemistry* **1999**, *38*, 6419.
(206) Krebs, M. R. H.; Wilkins, D. K.; Chung, E. W.; Pitkeathly, M. C.; Chamberlain, A. K.; Zurdo, J.; Robinson, C. V.; Dobson, C. M. *J. Mol. Biol.* **2000**, *300*, 541.
(207) Canet, D.; Last, A. M.; Tito, P.; Sunde, M.; Spencer, A.; Archer, D. B.; Redfield, C.; Robinson, C. V.; Dobson, C. M. *Nat. Struct. Biol.* **2002**, *9*, 308.
(208) Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, V. E.; Bychkova, G. V.; Semisotnov, G. V.; Venyaminov, S. Y.; Ptitsyn, O. B. *FEBS Lett.* **1981**, *136*, 311.
(209) Chyan, C.-L.; Wormald, C.; Dobson, C. M.; Evans, P. A.; Baum, J. *Biochemistry* **1993**, *32*, 5681.
(210) Alexandrescu, A. T.; Evans, P. A.; Pitkeathly, M.; Baum, J.; Dobson, C. M. *Biochemistry* **1993**, *32*, 1707.
(211) Kim, S.; Bracken, C.; Baum, J. *J. Mol. Biol.* **1999**, *294*, 551.
(212) Wijesinha-Bettoni, R.; Dobson, C. M.; Redfield, C. *J. Mol. Biol.* **2001**, *312*, 261.

(213) Horng, J. C.; Demarest, S. J.; Raleigh, D. P. *Proteins* **2003**, *52*, 193.
(214) Ramboarina, S.; Redfield, C. *J. Mol. Biol.* **2003**, *330*, 1177.
(215) Demarest, S. J.; Horng, J. C.; Raleigh, D. P. *Proteins* **2001**, *42*, 237.
(216) Eliezer, D.; Jennings, P. A.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1997**, *417*, 92.
(217) Harding, M. M.; Williams, D. H.; Woolfson, D. N. *Biochemistry* **1991**, *30*, 3120.
(218) Stockman, B. J.; Euvrard, A.; Scahill, T. A. *J. Biomol. NMR* **1993**, *3*, 285.
(219) Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *Biochemistry* **1997**, *36*, 13043.
(220) Platt, G. W.; Simpson, S. A.; Layfield, R.; Searle, M. S. *Biochemistry* **2003**, *42*, 13762.
(221) Garcia, P.; Serrano, L.; Rico, M.; Bruix, M. *Structure* **2002**, *10*, 1173.
(222) Staniforth, R. A.; Giannini, S.; Higgins, L. D.; Conroy, M. J.; Hounslow, A. M.; Jerala, R.; Craven, C. J.; Waltho, J. P. *EMBO J.* **2001**, *20*, 4774.
(223) Kutyshenko, V. P.; Cortijo, M. *Protein Sci.* **2000**, *9*, 1540.
(224) Miller, J.; McLachlan, A. D.; Klug, A. *EMBO J.* **1985**, *4*, 1609.
(225) Lee, M. S.; Gottesfeld, J. M.; Wright, P. E. *FEBS Lett.* **1991**, *279*, 289.
(226) Parraga, G.; Horvath, S. J.; Eisen, A.; Taylor, W. E.; Hood, L.; Young, E. T.; Klevit, R. E. *Science* **1988**, *241*, 1489.
(227) Lee, M. S.; Gippert, G.; Soman, K. Y.; Case, D. A.; Wright, P. E. *Science* **1989**, *245*, 635.
(228) Laity, J. H.; Lee, B. M.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39.
(229) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.
(230) Wuttke, D. S.; Foster, M. P.; Case, D. A.; Gottesfeld, J. M.; Wright, P. E. *J. Mol. Biol.* **1997**, *273*, 183.
(231) Brüschweiler, R.; Liao, X.; Wright, P. E. *Science* **1995**, *268*, 886.
(232) De Guzman, R. N.; Liu, H. Y.; Martinez-Yamout, M.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2000**, *303*, 243.
(233) Dames, S. A.; Martinez-Yamout, M.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5271.
(234) Goodman, R. H.; Smolik, S. *Genes Devel.* **2000**, *14*, 1553.
(235) Freedman, S. J.; Sun, Z. Y.; Poy, F.; Kung, A. L.; Livingston, D. M.; Wagner, G.; Eck, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5367.
(236) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. *Nature* **2002**, *415*, 549.

CR030403S