Natively Disordered Proteins

[1]Gary W. Daughdrill, [2]Gary J. Pielak, [3]Vladimir N. Uversky, [4]Marc S. Cortese, and [4]A. Keith Dunker

[1]Department of Microbiology, Molecular Biology, and Biochemistry, University of Idaho, Moscow, ID 83844

[2]Department of Chemistry, Department of Biochemistry and Biophysics, and the Lineberger Comprehensive Cancer Research Center, University of North Carolina, Chapel Hill, NC 27599

[3]Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia and Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064

[4]Center for Computational Biology and Bioinformatics, University of Indiana School of Medicine, Indianapolis, IN 46202 and Molecular Kinetics, Inc., Indiana University Emerging Technology Center, 351 West 10th Street, Indianapolis, IN 45202

1. Introduction

To understand natively disordered proteins, it is first important to introduce the-structure-function paradigm, which dominates modern protein science. We then discuss the terminology used to describe natively disordered proteins and present a well-documented example of a functional disordered protein. Finally, we compare the standard structure-function paradigm with structure-function relationships for natively disordered proteins, and from this comparison, suggest an alternative for relating sequence, structure and function.

i. The Protein Structure-Function Paradigm

The structure-function paradigm is, simply stated, that the amino acid sequence of a protein determines its 3-D structure, and that the function requires the prior formation of this 3-D structure. This view has been deeply engrained in protein science long before the 3-D structure of a protein was first glimpsed almost 50 years ago. In 1893 Emil Fischer developed the “lock and key” hypothesis from his studies on different types of similar enzymes, one of which could hydrolyze α- but not β-glycosidic bonds, and another of which could hydrolyze β- but not α-glycosidic bonds[1]. By 1930 it had become clear that globular proteins lose their native biological activity (i.e., they become denatured) when solution conditions are altered by adding heat or solutes. Anson and Mirsky showed in 1925 that denatured hemoglobin could be coaxed back to its native state by changing solution conditions[2]. This reversibility is key because it means the native protein and the denatured protein can be treated as separate thermodynamic states. Most importantly, this treatment leads directly to the idea that a protein's function is determined by the definite structure of the native state. In short, the structure is known to exist because it is destroyed by denaturation. Wu stated as much in 1931, but his work was probably unknown outside China even though his papers were published in English[3]. The West had to wait until 1936, when Mirsky and Pauling published their review of protein denaturation[4].

Anfinsen and colleagues used the enzyme ribonuclease to solidify these ideas. Their work led in 1957 to the "thermodynamic hypothesis". Anfinsen wrote[5], "This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu is the one in which the Gibbs free energy of the whole system is lowest: that is, the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment." Merrifield and colleagues then performed an elegant experiment that drove home the idea; they synthesized ribonuclease in the test tube from the amino acid sequence[6]. Their experiments provided direct evidence that the amino acid sequence determines all other higher order structure, function, and stability.

As mentioned above, the observation that denaturation is reversible allows the application of equilibrium thermodynamics. It was soon realized that many small globular proteins exist in only two states, the native state or the denatured state. That is, each protein molecule is either completely in the native state or completely in the denatured state. Such two-state behavior leads directly to expressions for the equilibrium constant, K_D, and free energy, ΔG_D, of denaturation:

$$K_D = [D]/[N] \qquad (1)$$

$$\Delta G_D = -RT\ln(K_D) \qquad (2)$$

where R is the gas constant, T is the absolute temperature, and [N] and [D] represent the concentrations of the native and denatured states, respectively.

The definition of K_D, Equation 1, is straightforward, but quantifying K_D is more difficult than defining it. The difficulty arises because K_D cannot usually be quantified at the conditions of interest -- room temperature in buffered solutions near physiological pH. Specifically, most biophysical techniques can only sensitively measure K_D between values of about 10 and 0.1, but the overwhelming majority (>99.9%) of two-state globular protein molecules are in the native state at the conditions of interest. The difficulty can be overcome by extrapolation. Increasing the temperature or adding solutes such as urea or guanidinium chloride pushes K_D into the quantifiable region. Plots of $-RT\ln(K_D)$ versus temperature or solute concentration are then extrapolated back to the unperturbed condition to give ΔG_D at the conditions of interest. Many such studies indicate that small globular proteins have a stability of between about 1 and 10 kcal/mol at room temperature near neutral pH. We introduced this formal definition of stability so that later we can discuss the idea that the higher-order structure within some intrinsically disordered proteins is simply unstable.

What exactly is higher-order structure? Pauling showed that the protein chain was organized in definite local structures, helices and sheets[7], but it was unclear how these structures interact to form the native state. Because the conceptual accessibility of physical representations is greater than that of thermodynamics, the advent of X-ray crystallography strongly reinforced the sequence-structure-function paradigm. In 1960, Kendrew and Perutz used X-ray crystallography to reveal the intricate, atomic-level structures of myoglobin[8] and hemoglobin[9], effectively locking-in the sequence-to-structure paradigm. The paradigm took on the aura of revealed truth when Phillips solved the first structure of an enzyme, lysozyme, in 1965[10]. The position of the bound inhibitor revealed the structure of the active site, making it clear that the precise locations of the amino acid side chains is what facilitates catalysis.

Given all this evidence, it appeared as if the case was closed: the native state of every protein possesses a definite and stable three-dimensional structure, and this structure is required for biological function. But even early on there were worrying observations. Sometimes loops were missing from high-resolution structures, and these loops were known to be required for function[11,12]. Nuclear magnetic resonance spectroscopy also showed that some proteins with known biological functions did not possess stable, defined structure in solution[13].

Perhaps the most important difference to bear in mind when relating the sequence-structure-function paradigm to intrinsically disordered proteins is the difference between a structural state and a thermodynamic state. The native state is both a structural state and thermodynamic state, but the disordered (and denatured) state is only a thermodynamic state.

That is, all the molecules in a sample of the native state of a globular protein have nearly the same structure, and this structure is what is lost on denaturation. On the other hand, the denatured state consists of a broad ensemble of molecules - each having a different conformation. Therefore, averaged quantities have different meanings for native and disordered states. For a native globular protein, an averaged quantity, such as the CD signal (see below), gives information about each molecule in the sample, because nearly all the molecules are in the same structural state. For a denatured or disordered protein, an averaged quantity contains information about the ensemble, and this information may or may not be applicable to individual molecules in the sample.

ii. Natively Disordered Proteins

Many proteins carry out function by means of regions that lack specific 3-D structure, existing instead as ensembles of flexible, unorganized molecules. In some cases the proteins are flexible ensembles along their entire lengths, while in other cases only localized regions lack organized structure. Still other proteins contain regions of disorder without ascribed functions, but functions might be associated with these regions at a later date. Whether the lack of specific 3-D structure occurs wholly or in part, such proteins do not fit the standard paradigm that 3-D structure is a prerequisite to function.

Various terms have been used to describe proteins or their regions that fail to form specific 3-D structure including: flexible[14], mobile[15], partially folded[16], natively denatured[17], natively unfolded[18], intrinsically unstructured[19], and intrinsically disordered[20].

None of these terms or combinations is completely appropriate. Flexible, mobile, and partially folded have the longest histories and most extensive use; however, all three of these terms are used in a variety of ways including many that are not associated with proteins that exist as structural ensembles under apparently native conditions. For example, ordered regions with high B-factors are often called flexible or mobile. Partially folded is often used to describe transient intermediates involved in protein folding. With regard to the term natively, it is difficult to know whether a protein is in its native state. Even when under apparently physiological conditions, a protein might fail to acquire a specific 3-D structure due to the absence of a critical ligand or because the crowded conditions inside the cell are needed to promote folding. Because of such uncertainties, “intrinsically” is often chosen over “natively.” Unfolded and denatured are often used interchangeably, so the oxymoron “natively denatured” has a certain appeal but has not gained significant usage. Unfolded and unstructured both imply lack of backbone organization, but natively disordered proteins often have regions of secondary structure, sometimes transient and sometimes persistent. There are even examples of apparently native proteins with functional regions that resemble molten globules[21]; the molten globule contains persistent secondary structure but lacks specific tertiary structures, having instead regions of non-rigid side chain packing that leads to mobile secondary structure units[22,23]. Since disordered encompasses both extended and molten globular forms, this descriptor has some advantage, but the wide-spread use of disorder in association with human diseases complicates computer searches and gives a negative impression.

Here we use natively disordered, which has been infrequently used if at all, mainly to distinguish this work from previous manuscripts on this topic. Developing a standardized vocabulary in this field would be of great benefit. We propose collapsed-disorder for proteins and domains that exist under physiological conditions primarily as molten globules and extended-disorder for proteins and regions that exist under physiological conditions primarily as random coil.

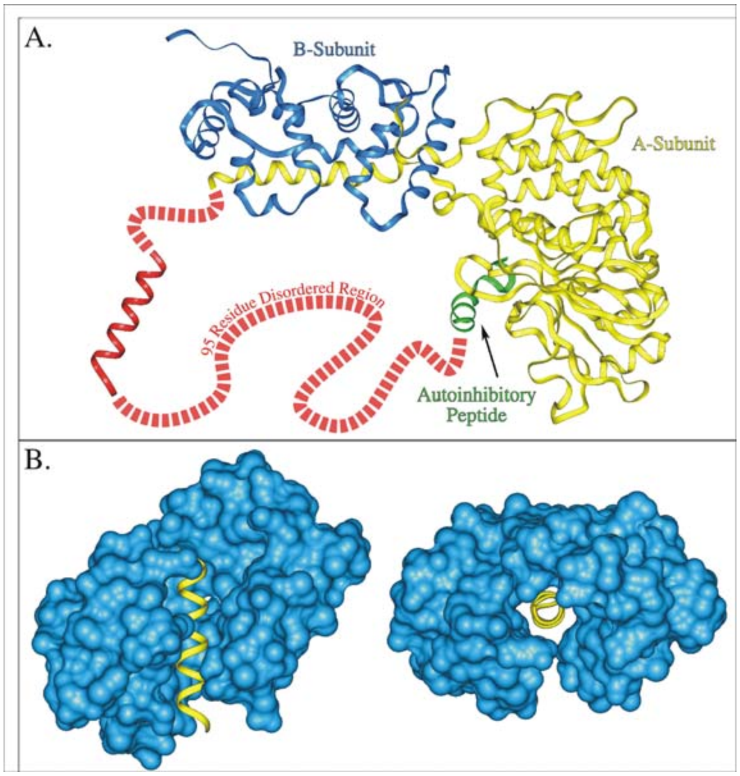


Figure 1. **(a)** 3-D structure of calcineurin, showing the A subunit (purple), The B subunit (blue), the auto-inhibitory peptide (green) and the location of a 95-residue disordered region (red). The calmodulin binding site (yellow helix) is located within the disordered region (orange). **(b)** Side and top views of calmodulin (blue) binding a target helix (yellow). Note that the calmodulin molecule surrounds the target helix when bound.

A significant body of work suggests that the unfolded state is not a true random coil, but instead possesses substantial amounts of an extended form that resembles the polyproline II helix[24,25,26] as well as other local conformations that resemble the native state. For this reason, extended-disorder may be a preferable term as compared to random coil, but the latter term continues to have widespread usage and so, for convenience, we will continue to use this term here – with the understanding that by the term random coil we do not mean the true random coil defined by the polymer chemist.

It is useful to introduce the topic of natively disordered proteins with a specific, very clear example. Calcineurin (Figure 1a) makes a persuasive case for the existence and importance of native disorder[27,28,29]. This protein contains a catalytic A subunit and a B subunit with 35% sequence identity to calmodulin. The A subunit is a serine-threonine phosphatase that becomes activated upon association with the Ca^{2+}-calmodulin complex. Thus, calcineurin, which is widespread among the eukaryotes, connects the very important signaling systems based on Ca^{2+} levels and phosphorylation/dephosphorylation. When Ca^{2+}-calmodulin binds to its target helix within calcineurin, the auto-inhibitory peptide becomes displaced from its association with the active site and by this means the phosphatase activity is turned on. Since Ca^{2+}-calmodulin wraps around its target helix (Figure 1b), this helix must lack tertiary contacts and therefore lie in a disordered region. The disordered character of the region surrounding the target helix in calcineurin has been shown by its sensitivity to trypsin digestion[27,28] and by the missing electron density of over 95 consecutive residues in the crystal structure[29]. In this example, several lines of evidence combine to support the importance of intrinsic disorder for biological function.

iii. A New Protein Structure-Function Paradigm

The standard view is that amino acid sequence codes for 3-D structure and that 3-D structure is a necessary prerequisite for protein function. In contrast, not only calcineurin discussed above, but as we have shown elsewhere[30,31], many additional proteins are natively disordered or contain natively disordered regions, and for such proteins or protein regions, lack of specific 3-D structures contributes importantly to their functions. Here we explore the distinctions between the standard and our alternative view of protein structure-function relationships.

The standard structure-function paradigm as discussed above arose initially from the study of enzymes. The original lock and key proposal[32] was based on distinctive substrate recognition by a pair of enzymes. In the approximately 110 years since this initial proposal, studies of enzyme catalysis have continuously reinforced this view.

While more than 20 proposals have been made to explain enzyme catalysis[33], it has become accepted that the most profitable way to think about the problem is in terms of transition state stabilization[34]. That is, as first suggested by Polanyi[35] and later by Pauling[36], for an enzyme to carry out catalysis, it must bind more tightly to the transition state than to the ground state. This tighter binding lowers the transition state energy and accelerates the reaction rate[34,35,36]. Transition states are derived from ground states by very slight movements of atoms, movements that are typically half the length or so of a chemical bond. Attempts to understand the mechanistic details of how an enzyme can bind more tightly to the transition state have occupied a number of protein chemists and theoreticians and are not completely clear in every aspect even today[37], with some researchers arguing that electrostatic contributions provide the dominant effects[38] and others emphasizing the importance of entropic effects[39] or other contributions[40]. Despite these uncertainties, there seems to be universal agreement that tighter binding to the transition state depends on an accurate prior positioning of the key residues in the enzyme. This prior positioning requires a well-organized protein 3-D structure. Thus, in short, the evolution of ordered structure in proteins was likely reinforced by or directly resulted from the importance of enzyme catalysis.

Not only for the calcineurin example given above, but for most of the natively disordered proteins we have studied[30,31], the function of the disorder is for signaling, regulation or control. Compared to order, disorder has several clear advantages for such functions. When disordered regions bind to signaling partners, the free energy required to bring about the disorder to order transition takes away from the interfacial, contact free energy, with the net result that a highly specific interaction can be combined with a low net free energy of association[20,41]. High specificity coupled with low affinity seems to be a useful pair of properties for a signaling interaction so that the signaling interaction is reversible. It would appear to be more difficult to evolve a highly specific yet weak interaction between two ordered structures. In addition, a disordered protein can readily bind to multiple partners by changing shape to associate with different targets[20,42,43]. Multiple interactions are now being commonly documented, and proteins having 20 or more partners are being described. In protein interaction or signaling networks, proteins with multiple partners are often called hubs. We previously suggested that the ability to interact with multiple partners may depend on regions of native disorder[44], but so far we have investigated only a limited number of examples. Whether hub proteins utilize regions of native disorder to enable their binding diversity is an important question for proteins of this class.

Given the above information, we propose a new protein structure-function paradigm. Simply put, we propose a two-pathway paradigm, with sequence to 3-D structure to function for catalysis and sequence to native disorder to function for signaling and regulation.

2. Methods used to characterize natively disordered proteins

i. NMR Spectroscopy

Intrinsically disordered proteins have dynamic structures that interconvert on a number of timescales. Nuclear magnetic resonance spectroscopy can detect this molecular motion as well as any transient secondary and tertiary structure that is present. Several reviews have focused on the use of NMR to characterize the structure and dynamics of intrinsically disordered proteins[45,46,47,48,49]. There is also a rich body of literature reviewing the use of NMR to characterize the structure and dynamics of non-native states of globular proteins[50,51,52,53,54,55]. Due to the high activation barrier for studying the structure and dynamics of intrinsically disordered proteins, much of the work reviewed here was performed on non-native states of globular proteins. Fortunately, these non-native states have many similarities to intrinsically disordered proteins. The barrier to studying intrinsically disordered proteins seems to prevail regardless of their functional relevance and the abundance of NMR techniques available to characterize their structure and dynamics. This barrier is based on an entrenched attitude of most protein chemists regarding the relationship between protein structure and function. For example, when a graduate student presents an HSQC spectrum of an intrinsically disordered protein displaying limited ^1H chemical shift dispersion and narrow resonance lineshape to their advisor, they are likely to be ridiculed for making an error during the purification of the protein. This attitude prevails despite more than a decade of overwhelming evidence that intrinsic protein disorder exists and can be essential for function. It is time to move beyond the limited view that defined 3-D structure is a requirement for protein function and to acknowledge the growing body of evidence that natively disordered proteins exist and have functions.

In this report, a systematic approach for using NMR to investigate the structure, dynamics and function of intrinsically disordered proteins will be suggested based on a subset of NMR experiments. Our objective is to develop a comprehensive picture of the structure and dynamics of intrinsically disordered proteins. This objective is most easily accomplished by unifying the different types of NMR data collected on intrinsically disordered proteins while avoiding the pitfall described in the following quotation from a classic paper on NMR and protein disorder[56]: ". . . until the amount of structural information provided by NMR methods increases by several orders of magnitude, descriptions of non-native structure will probably consist of simple lists of estimates of fractional population of secondary-structure segments and of side-chain interactions."

It is interesting that the author of this quote has provided one of the most complete descriptions of the structure and dynamics of an unfolded protein, a fragment of staphylococcal nuclease referred to as Δ131Δ[56,57,58,59,60,61,62,63,64]

The most complete description of the structure and dynamics of intrinsically disordered proteins would include the population and structure of the different members of the rapidly interconverting ensemble along with the rate of interconversion between unique structures. Some of these parameters can be estimated from NMR chemical shifts and resonance lineshape measurements[48,50,51,65,66,67]. Because liquid state NMR detects the ensemble average, chemical shifts can be treated as macroscopic equilibrium constants for secondary structure formation[50,51]. This information can be combined with measurements of hydrodynamic radii, using pulsed field gradient methods, to place limits on the conformational space that is being explored, which ultimately facilitates structure calculation of ensemble members[68,69,70]. Of course, it remains unclear how many discrete states are represented in the resonance lineshape measurements, which presents the fundamental limitation of current NMR practice and theory to provide a more detailed description of the ensemble members. Many consider this problem insoluble, given the large number of possible conformations, even for a small polypeptide. However, there is ample evidence that intrinsically disordered proteins do not explore all of these conformations. Further, several systems have been characterized where resonance lineshape measurements were deconvoluted to provide information about the number of ensemble members or to characterize the interconversion rate between conformations[71,72,73].

Chemical shifts measure the presence of transient secondary structure

Resonance assignments are the first step in characterizing the structure and dynamics of any protein using NMR. Resonance assignments specify the atomic identity of unique frequencies observed in the spectrum. These frequencies are generally normalized to some standard and reported in parts per million as the chemical shift. Chemical shift measurements for $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C(O)$ nuclei are sensitive to $\phi\Psi$ dihedral angles and deviate systematically from random coil values for helical and beta conformations[65,67,74,75]. The deviations are diagnostic for the presence of secondary structure, regardless of stability, as long as the interconversion rate is fast and the deviation from the random coil chemical shift value is greater than the spectral resolution. For stably folded proteins, the chemical shift measurement of $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C(O)$ nuclei provide a picture of secondary structure that represents a lower limit on the equilibrium constant for folding from the unfolded state[50]. The extension of this relationship to intrinsically disordered proteins makes the same two state assumption. In this case, the fraction of helix or strand present can be determined by comparing the chemical shift value observed in an intrinsically disordered protein to the value expected in a stably folded protein[50,51]. This data combined with knowledge of the Stokes radius can be used to restrict the available conformations in a structure calculation. This assumption is in agreement with a recent analysis of the effective hydrodynamic radius of protein molecules in a variety of conformational states[76]. In fact, based on the analysis of 180 proteins in different conformational states, it has been shown that the prediction of the overall protein dimension could be predicted based on the chain length; i.e., the protein molecular weight, with an accuracy of 10 %. Furthermore, it has been emphasized that the incorporation of biophysical constraints, which can be rationalized based on conventional biophysical measurements, might lead to considerable improvement of structure simulation procedures. Clearly, the size and shape of the bounding volume used for structure simulations plays a crucial role in determining the efficiency and accuracy of any algorithm. The incorporation of a size/shape constraint derived from experimental data might lead to considerable improvement of simulation procedures[76].

In a study of an intrinsically disordered negative regulator of flagellar synthesis, FlgM, chemical shift deviations from random coil values were observed for $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C(O)$ nuclei[77]. Similar deviations were not observed for $^{1}H_{\alpha}$ nuclei and this may be a general property of intrinsically disordered proteins. For FlgM, the chemical shift deviations indicated the presence of helical structure in the C-terminal half of the protein and a more extended flexible structure in the N-terminal half of the protein. Two regions with significant helical structure were identified containing residues 60-73 and 83-90. Additional NMR and genetic studies demonstrated these two helical regions were necessary for interactions with the sigma factor. The chemical shift deviations observed for FlgM exhibited a characteristic variation where the central portion of the helix had a larger helical chemical shift difference than the edges. To test whether the helical structure based on chemical shifts represented an ensemble of nonrandom conformations, the $^{13}C_{\alpha}$ chemical shifts were measured in the presence of increasing concentrations of the chemical denaturant urea. As the concentration of urea was increased, the $^{13}C_{\alpha}$ chemical shifts moved toward the random coil value in a manner characteristic of a noncooperative unfolding transition.

In a study of the basic leucine zipper transcription factor, GCN4, transient helical structure was observed in the basic region of GCN4 based on chemical shift deviations from random coil values for $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C(O)$ nuclei[78]. The temperature dependence of the three chemical shifts was also monitored. For the folded leucine zipper region, a small dependence between chemical shift and temperature was observed. Conversely, the unfolded basic region showed large change in $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C(O)$ chemical shifts when the temperature was changed[78]. The dynamic behavior of the basic region and rapidly interconverting structures are responsible for the temperature dependence of the chemical shifts[45]. In contrast, the temperature dependence of the chemical shifts in the folded leucine zipper was small and this behavior is generally observed for both folded and random coil regions.

Pulsed field gradient methods to measure translational diffusion

Another easily applied NMR method is the measurement of the translational diffusion coefficient, D, using pulsed field gradients (PFG)[68,69,70]. This approach relies on the fact that a protein molecule undergoing translational diffusion will differentially sense a gradient of signal designed to destroy any acquired magnetization. An array of gradient strength or time can be used to calculate the translational diffusion coefficient, D[69]. Knowledge of D can be used to calculate the hydrodynamic radius. The hydrodynamic radius can then be compared to the expected value based on molecular weight to determine whether the protein is compact and globular or extended and flexible. Empirical relationships were recently established between the hydrodynamic radius determined using PFG translational diffusion measurements, and the number of residues in the polypeptide chain for native folded proteins and highly denatured states[68]. This study provided evidence for significant coupling between local and global features of the conformational ensembles adopted by disordered polypeptides. As expected, the hydrodynamic radius of the polypeptide was dependent on the level of persistent secondary structure or the presence of hydrophobic clusters.

NMR relaxation and protein flexibility

NMR relaxation is the premier method to investigate protein flexibility[45,47,56,57,79]. It is the first order decay of an inductive signal back to equilibrium with the applied field. In particular; “The relaxation mechanism for the NH spin system arises from the local time-varying magnetic fields generated at the ^{1}H and ^{15}N nuclei due to global tumbling and internal mobility of the various N-H bond vectors"[80:2676].

This means relaxation of the NH spin system is sensitive to both local and global motion. Measuring the longitudinal relaxation rate, R_1, the transverse relaxation rate, R_2, and the heteronuclear Overhauser effect between the amide proton and its attached nitrogen, NH-NOE, is useful for describing the internal dynamics and molecular motions associated with proteins and typically measurable for every amide N-H pair in proteins less than 40kDa[79,81,82 83].

In particular, the NH-NOE experiment provides a fast, powerful, and easy to interpret diagnostic for the presence of intrinsic protein disorder. The sign of the NH-NOE resonance is sensitive to the rotational correlation time and is positive for N-H bond vectors with a long rotational correlation time (>1-10 ns) and negative for N-H bond vectors with a short rotational correlation time (<0.1-1 ns). The NH-NOE experiment can be performed on any uniformly ^{15}N-labeled protein sample that can be concentrated without aggregation to between 0.1 and 1.0 mM. The relatively short rotational correlation time observed for intrinsically disordered proteins results in sharp narrow lines in the NMR spectrum. Because of this property, the NH-NOE experiment can be used to detect intrinsic disorder in proteins up to 200 kDa[84].

We argue that the NH-NOE experiment is such a valuable diagnostic for the presence of intrinsic protein disorder that it should be incorporated into all screening protocols developed for structural genomics. It is generally a waste of time and resources to directly pursue crystallization of proteins that are intrinsically disordered. Crystallization meets with limited success for proteins containing intrinsically disordered regions. This is because the presence of structural interconversions for intrinsically disordered proteins prohibits the formation of an isomorphous lattice, unless there are one or a few stable low energy conformations that can be populated. During screening, proteins determined to have intrinsically disordered regions can be marked for further NMR analysis. On a technical note, a range of relaxation and saturation delays should be used when measuring the NH-NOE values in intrinsically disordered proteins[85].

Some of the most valuable contributions to understanding the structure and dynamics of intrinsically disordered proteins have come from using R_1, R_2, and NH-NOE data to model molecular motion or to solve the spectral density function. The two most successful models for relaxation data analysis are the so called ‘model-free’ approach of Lipari and Szabo[86,87] and reduced spectral density mapping introduced by Peng and Wagner[80,88].

Using the model free analysis of relaxation data to estimate internal mobility and rotational correlation time

In the interpretation of ^{15}N relaxation data it is assumed the relaxation properties are governed solely by the ^{1}H-^{15}N dipolar coupling and the chemical shift anisotropy[89]. Using these assumptions, a spherical molecule with an overall correlation time, τ_m, and an effective correlation time for fast internal motions, τ_e, will have a spectral density function of the following form[86,87]:

$$J(\omega) = \frac{2}{5}\left\{ {S^2\tau_m}\Big/{\left(1+(\omega\tau_m)^2\right)} + {\left(1+S^2\right)\tau}\Big/{\left(1+(\omega\tau)^2\right)} \right\} \qquad (3)$$

where $1/\tau = 1/\tau_m + 1/\tau_e$ and S^2 is the square of the generalized order parameter describing the amplitude of internal motions. The overall correlation time, τ_m, is determined based on an assumption of isotropic Brownian motion. For a folded protein of known structure, the diffusion tensor can be incorporated into the model to compensate for anisotropic motions[90,91,92]. It is unclear how valid the assumption of isotropic rotation is for intrinsically disordered proteins. It is assumed that anisotropic structures will populate the conformational ensemble of an intrinsically disordered protein. This is in the absence of little direct evidence on the subject. It has been suggested the effects of rapid interconversion between more isotropic structures will tend to smooth out static anisotropy and result in an isotropic average conformation[57].

Regardless of the analytical limitations, the model free analysis of several intrinsically disordered proteins has ultimately provided a useful qualitative picture of the heterogeneity in rotational diffusion observed for intrinsically disordered proteins[58,77,93,94,95]. Some general trends are observed: 1) correlation times are greater than those calculated based on polymer theory and 2) transient secondary structure tends to induce rotational correlations.

Changes in S^2 have been used to estimate changes in conformational entropy due to changes in ns-ps bond vector motions during protein folding and for intrinsically disordered protein folding that is coupled to binding[45,63,77,96,97]. In aqueous buffer and near neutral pH, the N-terminal SH3 domain of the Drosophila signal transduction protein, Drk, is in equilibrium between a folded, ordered structure and an unfolded, disordered ensemble. The unfolded ensemble is stabilized by 2 M guanidine hydrochloride and the folded structure is stabilized by 0.4 M sodium sulfate. Order parameters were determined for both the unfolded and folded Drk SH3 domains. The unfolded ensemble had an average S^2 value of 0.41 ± 0.10 and the folded structure had an average S^2 value of 0.84 ± 0.05. Based on the difference in S^2 between the unfolded and folded structure, the average conformational entropy change per residue was estimated to be 12 J/molK. This approach does not address additional entropy contributions from slower motional processes or the release of solvent. Interestingly, the estimate of 12 J/molK is similar to the average total conformational entropy change per residue estimated from other techniques (~14 J/molK)[96]. Another study has also suggested that changes in order parameters provide a reliable estimate of the total conformational entropy changes that occur during protein folding[63].

Using reduced spectral density mapping to assess the amplitude and frequencies of intramolecular motion

The Stokes-Einstein equation defines a correlation time for rotational diffusion of a spherical particle:

$$\tau_c = \frac{\eta V_{sph}}{Tk_B} \tag{4}$$

In Equation 4, this rotational correlation time, τ_c, depends directly on the volume of the sphere (V_{sph}) and the solution viscosity (η) and inversely on the temperature (T). The spectral density function, J(ω), describes the frequency spectrum of rotational motions of the N-H bond vector relative to the external magnetic field and is derived from the Fourier transform of the spherical harmonics describing the rotational motions[80]. For isotropic Brownian motion J(ω) is related to τ_c in a frequency dependent manner:

$$J(\omega) = \frac{2}{5}\frac{\tau_c}{1+\omega^2\tau_c^2} \tag{5}$$

Reduced spectral density mapping uses the conveniently measured ^{15}N R_1, R_2, and NH-NOE to estimate the magnitude of the spectral density function at 0, ^{1}H, and ^{15}N angular frequencies. In turn, J(0), J(ω_H), and J(ω_N) are directly related to molecular motion through Equation 3. For instance, according to Equation 5, J(ω) at 0 frequency is equal to $2/5\tau_c$. This relationship represents an upper limit on J(0), which is usually reduced by fast internal motions that may result from anisotropic rotational motions of the N-H bond vector. It is also important to note that chemical exchange that is in the microsecond to millisecond range contributes positively to J(0) but this effect can be attenuated by measuring R_2 under spin lock conditions[80].

Reduced spectral density mapping is a robust approach for analyzing intrinsic disorder because it does not depend on having a model of the molecular motions under investigation. Several groups have characterized intrinsic protein disorder using reduced spectral density mapping, providing new insights into protein dynamics[78,80,94,98,99,100,101]. In one of these studies, reduced spectral density mapping was used to help demonstrate the intrinsically disordered anti-sigma factor, FlgM, contained two disordered domains[77]. One domain, representing the N-terminal half of the protein, was characterized by fast internal motions and small J(0) values. The second domain, representing the C-terminal half of the protein, was characterized by larger J(0) values, representing correlated rotational motions induced by transient helical structure. NMR was also used to help demonstrate the C-terminal half of FlgM contained the sigma-factor binding domain and it was proposed that the transient helical structure was stabilized by binding to the sigma-factor. However, more recent NMR studies of FlgM showed that this structure appears to be stabilized in the cell due to molecular crowding, which would reduce the role of conformational entropy on the thermodynamics of the FlgM/sigma-factor interaction[102].

In another study, a temperature dependent analysis of the reduced spectral density map was performed on the GCN4 bZip DNA binding domain[78]. In the absence of DNA, GCN4 exists as a dimer formed through a coiled-coil C-terminal domain and a disordered N-terminal DNA binding domain. This disordered DNA binding domain becomes structured when DNA is added[54]. In the GCN4 study, the reduced spectral density map was evaluated at three temperatures; 290 K, 300 K, and 310 K. J(0) values increased in a manner consistent with changes in solvent viscosity induced by increasing temperature. When J(0) was normalized for changes in solvent viscosity, values for the C-terminal dimerization domain and the intrinsically disordered N-terminal DNA binding domain were identical within experimental error. This behavior suggested that the correlated rotational motion of the folded leucine zipper had the dominant influence on the solution state backbone dynamics of the basic leucine zipper of GCN4. A recently completed but not yet published study monitored the temperature dependence of the reduced spectral density map for a partially folded fragment of thioredoxin. Unlike GCN4, the thioredoxin fragment does not contain a stably folded domain. In this study, normalizing J(0) for changes in solvent viscosity induced a trend in the data, with J(0) increasing with increasing temperature. In the absence of chemical exchange, this data suggest we are monitoring an increase in the hydrodynamic volume of the fragment (G. W. Daughdrill, unpublished data).

Characterization of natively disordered proteins

In 1994 Alexandrescu and Shortle described the first complete NMR relaxation analysis of a partially folded protein under nondenaturing conditions[57,58]. It was a fragment of staphylococcal nuclease referred to as Δ131Δ. For the majority of Δ131Δ residues, their experimental data was best described by a modified "model-free" formalism that included contributions from internal motions on intermediate and fast time scales and slow overall tumbling. They observed that the generalized order parameter S^2 correlated with sequence hydrophobicity and the fractional populations of three alpha-helices in the protein. In a more recent study, residual dipolar couplings were measured for native staphylococcal nuclease and Δ131Δ.

An extensive NMR data set was used to identify an ensemble of three-dimensional structures for the N-terminal SH3 domain of the Drosophila signal transduction protein drk, with properly assigned population weights[103,104]. This was accomplished by calculating multiple unfolding trajectories of the protein using the solution structure of the folded state as a starting point. Of course this approach is limited to proteins that have a compact rigid conformation. However, the ability to integrate multiple types of data describing the structural and dynamic properties of disordered proteins is pertinent to this discussion. Population weights of the structures calculated from the unfolding trajectories were assigned by optimizing their fit to experimental data based on minimizing pseudo energy terms defined for each type of experimental constraint. This work marks the first time that NOE, J-coupling, chemical shifts, translational diffusion coefficients, and tryptophan solvent accessible surface area data were used in combination to estimate ensemble members. As seen with many other studies of the structure and dynamics of intrinsically disordered proteins, the unfolded ensemble for this domain was significantly more compact than a theoretical random coil.

ii. X-ray Crystallography

Wholly disordered proteins would not be expected to crystallize and thus could not be studied by X-ray crystallography. On the other hand, proteins with both ordered and disordered regions can form crystals, with the disordered regions occupying spaces in the crystal formed by the ordered parts of the molecule. Regions of disorder vary in location from one molecule to the next and therefore fail to scatter X-rays coherently. The lack of coherent scattering leads to missing electron density.

A given protein crystal is made up of identical repeating unit cells, where each unit cell contains one to several protein molecules depending on the symmetry of the crystal. Each atom, j, in the protein occupies a particular position, xj, yj, and zj. It is convenient to express the positions as dimensionless coordinates that are fractions of the lengths of the unit cell, yielding

Xj, Yj, and Zj as the indicators of the position of atom j. The result of an X-ray diffraction experiment is a 3-D grid of spots that are indexed by three integers, h, k, l. The positions of the spots are determined by the crystal lattice. The intensities of the spots are determined both by the symmetries in the crystal and by the positions of the atoms in the molecule. Specifically, each spot has an intensity that is the square of the magnitude of the structure factor, F(h,k,l), which is given by the following equation:

$$F_{(h,k,l)} = \sum_{j=1}^{atoms} f_{(j)} \exp\left[2\pi \bullet i\left(hX_j + kY_j + lZ_j\right)\right] \qquad (6)$$

where h, k, and l are the indices of the spots in the diffraction pattern, $f_{(j)}$ is the scattering power of atom j (dependent on the square of the number of electrons in the atom), i is the square-root of -1, and X_j, Y_j and Z_j are the coordinates of atom j given as fractions of the unit cell dimensions as mentioned above.

Each $F_{(h, k, l)}$ has both a magnitude and a phase. The magnitude of $F_{(h,k,l)}$ is determined by the square-root of the intensity of each spot in the diffraction pattern. For wavelets from 2 different spots, the phase is the shift in the peak values and is 0 degrees for two wavelets that are exactly in phase (constructive interference) and 180 degrees for two wavelets that are totally out of phase (which leads to destructive interference). The phase for two arbitrary wavelets can be found to be any value between 0 and 360 degrees. Thus, given a structure, both the magnitude and phase of each $F_{(h,k,l)}$ can be calculated by carrying out the summation in Equation 6 over all of the atoms in the structure. Taking the Fourier transform of Equation 6 gives:

$$\rho_{(x,y,z)} = \frac{1}{V} \sum_h \sum_k \sum_l F_{(h,k,l)} \exp\left[-2\pi \bullet i\left(hx + ky + lz\right)\right] \qquad (7)$$

where $\rho_{(x, y, y)}$ is the electron density of the protein at x, y, and z and V is the volume of the unit cell (and the other terms have been defined above).

Thus, to determine a structural model of the given protein, one merely has to carry out the summation of Equation 7, and then to fit the resulting electron density map with a set of atoms that correspond to the connected set of residues in the structure. To carry out this triple vector sum, it is necessary to know both the magnitude and the phase of each $F_{(h, k, l)}$. While the magnitude is obtainable simply as the square root of the intensity of the spot, the phase information is lost during data collection because the time of arrival of the peak of each wavelet relative to those of the others cannot be recorded with current technologies. The loss of this critical information is commonly called the phase problem.

Three main methods have been developed to find the missing phase information for each structure factor, $F_{(h, k, l)}$. Each of these methods has advantages and disadvantages.

The earliest successful approach for proteins was to add heavy atoms. Since scattering power depends on the number of electrons squared, even a small number of heavy atoms can perturb the diffraction values enough to allow the determination of the phase values. The mathematical details are fairly involved, but for this approach to work the positions of the heavy atoms must be determined by comparing the phase intensities with and without the heavy atoms. In addition, the addition of the heavy atoms cannot significantly perturb the structure of the protein. With regard to this second point, the protein structure with and without the heavy atom must be isomorphous. Thus, this approach is called isomorphous replacement, and at least two such heavy atom replacements must be made to determine each phase value.

A second approach is to use multiple X-ray wavelengths that traverse the absorption edge of a selected heavy atom. The change in scattering over the absorption edge becomes substantial, which enables the phases to be determined. This approach is called multiwavelength anomalous dispersion or MAD. For protein structure determination by this approach, it is common to introduce selenium in the form of selenomethionine. This is a convenient atom that has an absorption edge at an appropriate wavelength value, and often (but not always) substitution of methionine with selenomethionine does not significantly affect the structure or activity of the protein of interest.

Finally, if a structure of a closely related protein is already known, it is often possible to use the phases from the related protein with the intensities from the protein of interest to generate the initial model structure. This approach is called molecular replacement. Often the homologous structure is not similar enough so the phases are too inaccurate to give a reasonable starting structure. A second major difficulty is to determine the correct orientation of the known structure relative to that of the unknown so that the phases can be correctly associated with the intensities. Since there are a very large number of possible orientations, a complex search of the possibilities needs to be developed. Evolutionary algorithms have recently been found to be useful for this task[105].

Structure determination by X-ray crystallography typically involves not just straightforward calculation and equation solving (as indicated above) but also significant modeling and simulation. Once the phase problem has been solved by experiment, an electron density map is generated by the calculations described in Equation 7. However, there are important technical difficulties in solving the phases, so the phase values usually contain large errors and the resulting electron density map contains many mistakes. To give an example difficulty, it is unclear whether the heavy atom derivatives are truly isomorphous; even small protein movements upon heavy atom binding can lead to significant errors in the estimation of the phases. Modeling is then used to fit the amino acid sequence into the error-containing electron density map. The structural model that emerges is typically adjusted by dynamics and simulation to improve bond lengths, bond angles, and inter-atom contacts. The model is then used to calculate improved phases and the whole process reiterated until the structure no longer changes in successive cycles.

In the overall process of protein crystallography, there are no clear-cut rules for dealing with regions of low intensity or missing electron density. Some crystallographers are more aggressive than others in attempting to fit (or model) such low density or missing regions. This variability means that missing coordinates in the Protein Data Bank (PDB) have not been assigned by uniform standards, which in turn leads to variability in the assignment of disorder. All of this is compounded by the imperfections (packing defects) in real crystals, which can additionally contribute to the absence of electron density.

Crystallographers have classified regions of missing electron density as being either static or dynamic[106,107]. Static disorder is trapped into different conformations by the crystallization process, while dynamic disorder is mobile. A dynamically disordered region could potentially freeze into a single preferred position upon cooling, while static disorder would be fixed regardless of temperature. By collecting data and determining structures at a variety of temperatures, dynamic disorder sometimes becomes frozen and thus distinguished from static disorder[108].

From our point of view, more important than static or dynamic is whether the region of missing electron density has a single set of coordinates along the backbone or whether the region exists as an ensemble of structures. A missing region with one set of coordinates could be trapped in different positions by the crystal lattice (statically disordered) or could be moving as a rigid body (dynamically disorder). In either case, the wobbly domain[20] is not an ensemble of structures, e.g. is not natively disordered, but rather is an ordered region that adopts different positions due to a flexible hinge.

Given the difficulties in ascribing disorder to regions of missing electron density or to missing coordinates in the PDB, it is useful to use a second method, such as NMR or protease sensitivity, to confirm that a missing region is due to intrinsic disorder rather than some other cause. While such confirmation has been carried out for a substantial number of proteins[20], the importance of disorder has not been generally recognized so such confirmation is not routine.

Despite the uncertainties described above, missing electron density provides a useful sampling of native disorder in proteins. To estimate the frequency of such regions, a representative set of proteins was studied for residues with missing coordinates[109]. The representative set, called PDB_Select_25, was constructed by first grouping the PDB into subsets of proteins with greater than 25% sequence identity and then choosing the highest quality structure from each subset[110]. Out of 1223 chains with 239,527 residues, only 391 chains, or 32%, displayed no residues with missing backbone atoms (Table 1). Thus, 68% of the non-redundant proteins contained some residues that lacked electron density. The 832 chains with missing electron density contained 1,168 distinct regions of disorder, corresponding to ~1.4 such regions/chain. These 1,168 disordered regions contained 12,138 disordered residues, or ~10 residues/disordered region on average. While most of the disordered regions are short, a few are quite long: 68 of the disordered regions are greater than 30 residues in length. Overall, the residues with missing coordinates are about 5% of the total.

A substantial fraction of the disorder-containing proteins in PDB are fragments rather than whole proteins. Since disordered regions tend to inhibit crystallization, disordered regions are often separated from ordered domains by genetic engineering or protease digestion prior to crystallization attempts. In either case, longer regions of disorder become shortened, presumably leading to improved probability of obtaining protein crystals. Given that wholly disordered proteins do not crystallize, and given that the proteins in PDB often contain truncated disordered regions or are ordered fragments that have been separated from flanking regions of disorder, it is clear that the PDB substantially under represents the amount of protein disorder in nature.

Table 1. Disorder in PDB_Select_25*.

Parameter	Number	Percentage
Total chains	1,223	
Chains with no disorder	391	32%
Chains with disorder	832	68%
Disordered regions	1168	
Disordered regions / disordered chains	1.4	
Disordered regions > 30 residues in length	68	5.8%
Total residues	239,527	
Disordered residues	12,138	5.1%
Residues in disordered regions > 30	3,710	1.5%

*Data extracted from PDB as of 10/1/2001.

To further understand the distribution of order and disorder in crystallized proteins, we carried out a comparison between the PDB and the Swiss-Prot databases. Swiss-Prot provides an easy mechanism to extract information about various proteins[111], so this comparison allows a convenient means to study disorder across the different kingdoms. The overall results of this comparison are given in Table 2. Of 4175 exact sequence matches between proteins in PDB_Select_25 and Swiss-Prot, 2258 were from eukaryotes, 1490 were from bacteria, 170 were from archaea, and 257 were from viruses. Proteins with no disorder or only short disorder were estimated by considering proteins for which at least 95% of the primary sequence was represented as observed coordinates in the PDB structures. The number (and percent) of proteins in each set for these mostly ordered proteins were 428 (19%) for eukaryotes, 594 (40%) for bacteria, 82 (48%) for archaea, and 42 (16%) for viruses. Proteins with substantial regions of disorder were estimated by considering proteins for which the crystal structure contained less than half of the entire sequence. The number and percent of proteins in each set for these likely to contain substantial amounts of disorder were 713 (32%) for eukaryotes, 97 (6.5%) for bacteria, 4 (2.3%) for archaea, and 123 (48%) for viruses. These data suggest that both eukaryotes and viruses are mostly likely to have proteins with large regions of disorder flanking fragmentary regions of ordered, crystallizable domains.

Table 2. Comparison of disorder between the PDB_Select_25 and Swiss-Prot databases.

	Number listed in both databases	Proteins with ≥ 90% assigned coordinates		Proteins with ≤ 50% assigned coordinates	
		Number	Percent	Number	Percent
Eukaryotes	2258	428	19	713	32
Bacteria	1490	594	40	97	6.5
Archaea	170	82	48	4	2.3
Viruses	257	42	16	123	48

iii. Small Angle X-ray Diffraction and Hydrodynamic Measurements

Both small angle X-ray scattering (SAXS) and hydrodynamic methods such as gel exclusion chromatography or dynamic light scattering have been used to estimate the sizes of protein molecules in solution. Analytical ultracentrifugation is another method for determining molecular weight and the hydrodynamic and thermodynamic properties of a protein[112]. The observed size of a given protein can then be compared with the size of a globular protein of the same mass. Indeed, the size of a given globular protein can be compared under physiological conditions with its size in denaturing levels of urea or guanidine. As expected, by these methods molten globules have similar overall sizes as the globular proteins of the same mass whereas random coil forms are considerably extended and so have a significantly larger size than their globular counterparts. Natively disordered proteins and globular proteins can be compared using several features of these data.

With regard to SAXS, one approach is to plot the normalized I(S)S^2 versus S, where I is the scattering intensity at a given scattering angle, and where S is the given scattering angle. This method emphasizes changes in the signal at higher scattering angles, which in turn strongly depends on the dimensions of the scattering molecule. The resulting graph, called a Kratky plot[113,114,115], readily distinguishes random coil proteins from globular structures. That is, globular proteins give inverted parabolas, with scattering intensity increasing and then dropping sharply due to the reduced scattering intensity at large angles. On the other hand, random coil forms give monotonically increasing curves that stabilize at a plateau. Importantly, it has been shown that the natively disordered proteins with extended disorder are characterized by low (coil-like) intramolecular packing density, reflected by the absence of a maximums on their Kratky plots[116,117,118,119,120,121,122]. This statement is illustrated by Figure 2 which compares the Kratky plots of five natively disordered proteins with those of two rigid globular proteins. One can see that the Kratky plots of natively disordered proteins do not exhibit maxima. Maxima on Kratky plots are typical of the folded conformations of globular proteins such as staphylococcal nuclease. In other words, the absence of a maximum indicates the lack of a tightly packed core under physiological conditions *in vitro*[123].

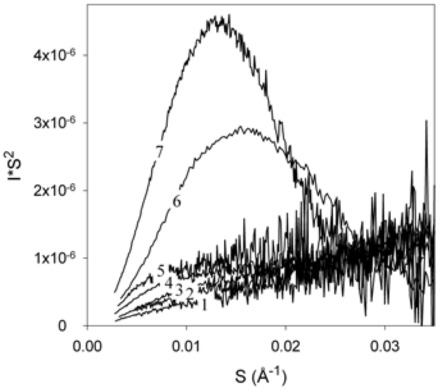


Figure 2. Kratky plots of SAXS data for natively disordered α-synuclein (1), α-synuclein (2), prothymosin α (3), caldesmon 636-771 fragment (4), and core histones (5). The Kratky plots of native globular staphylococcal nuclease (6) and hexameric insulin (7) are shown for comparison.

A second SAXS approach is to compare the radius of gyration, Rg, with that from a globular protein of the same size. The Rg value is estimated from the Guinier approximation[124]:

$$I(S) = I(0)\exp\left(-\frac{S^2 \bullet R_g^2}{3}\right) \qquad (8)$$

where I(S) is the scattering intensity at an angle S. The value of Rg can then be determined from plots of ln[I(S)] versus S^2. As tabulated by Millet, Doniach, and Plaxco[125], the radii of gyration of proteins unfolded by low pH, methanol, urea or guanidinium chloride are typically 1.5 to 2.5 times larger than the radii of gyration of the same proteins in their native globular states. Note however, that denaturation does not always result in complete conversion to extended coil formation – some structure may persist[126,127].

SAXS can also be used to estimate the maximum dimension of the protein. The distance distribution function P(r) can be estimated by Fourier inversion of the scattering intensity, I(S)[128,129], where P(r) is the probability of finding a dimension of length r. A plot of P(r) versus r yields the maximum dimension in the limit as P(r) goes to zero. The observed maximal dimension can then be compared with those observed for globular and random coil forms of various proteins[130,131].

A fourth method is to compare hydrodynamic volumes. Globular proteins differ significantly from fully or partly unstructured proteins in their hydrodynamic properties. For both native, globular proteins and for fully denatured proteins, empirical relationships have been determined between the Stoke's radius, Rs, and the number of residues in the chain[68,132,133]. These relationships in turn provide estimates of the overall hydrodynamic sizes compared to the respective controls.

Two common methods for estimating the Stoke's radius are gel exclusion chromatography and dynamic light scattering. In the former, the mobility of the protein of interest is compared to the mobilities of a collection of protein standards[132,133]. In the latter, the translational diffusion coefficient is estimated and the Stoke's radius is calculated from the Stoke's-Einstein equation.

For a given Stoke's radius, Rs, the hydrodynamic volume is given by:

$$Vh = \left(\frac{4}{3}\right)\pi(Rs)^3 \tag{9}$$

where the Stoke's radius of a given protein can be estimated by comparison to known standards using gel exclusion chromatography. Alternatively, the Stoke's radius can also be estimated from diffusion values estimated by dynamic light scattering measurements as described previously[68].

Plots of the logarithm of the hydrodynamic volume versus number of residues yield distinct straight lines for three different classes of reference proteins: 1) native, globular proteins, 2) molten globules, and 3) proteins unfolded by guanidinium chloride[123,134]. As expected, the molten globule forms are just slightly larger than their ordered, globular counterparts, while the unfolded proteins exhibit significantly larger hydrodynamic volumes.

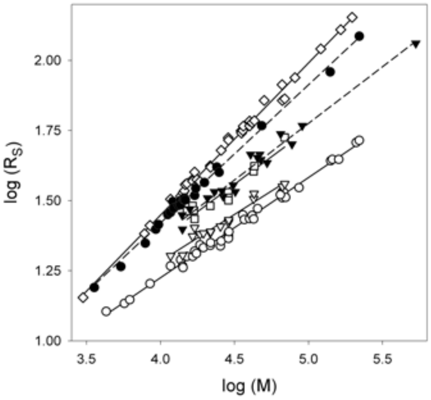


Figure 3. Dependencies of the hydrodynamic dimensions, R_S, on protein molecular mass, M, for native (open circles), molten globule (open triangles), pre-molten globule (open squares), 6M guanidinium chloride-unfolded conformational states of globular proteins (open diamonds), natively disordered proteins with coil-like (black circles) and pre-molten globule-like properties (black inverse triangles)[123].

This conclusion is illustrated by Figure 3, which compares the log(R_S) versus log(M) curves for natively disordered proteins with those of native, molten globule, pre-molten globule, and guanidinium chloride-unfolded globular proteins[123]. Additionally, Figure 3 shows that the log(R_S) versus log(M) dependencies for different conformations of globular proteins can be described by the following set of straight lines:

$$\log\left(R_S^{N}\right) = -(0.204 \pm 0.023) + (0.357 \pm 0.005) \bullet \log(M) \qquad (10)$$

$$\log\left(R_S^{MG}\right) = -(0.053 \pm 0.094) + (0.334 \pm 0.021) \bullet \log(M) \qquad (11)$$

$$\log\left(R_S^{PMG}\right) = -(0.210 \pm 0.180) + (0.392 \pm 0.041) \bullet \log(M) \qquad (12)$$

$$\log\left(R_S^{U-GdmCl}\right) = -(0.723 \pm 0.033) + (0.543 \pm 0.007) \bullet \log(M) \qquad (13)$$

where N, MG, PMG, and U-GdmCl correspond to the native, molten globule, pre-molten globule, and guanidinium chloride-unfolded globular proteins, respectively.

For the non-molten globular natively disordered proteins, their log(R_S) versus log(M) dependence may be divided in two groups, natively disordered proteins behaving as random coils in poor solvent (denoted as NU-coil in Equation 14), and essentially more compact proteins, which are similar to pre-molten globules with respect to their hydrodynamic characteristics (denoted as NU-PMG in Equation 15)[123,135]:

$$\log\left(R_S^{NU-coil}\right) = -(0.551 \pm 0.032) + (0.493 \pm 0.008) \bullet \log(M) \qquad (14)$$

$$\log\left(R_S^{NU-PMG}\right) = -(0.239 \pm 0.055) + (0.403 \pm 0.012) \bullet \log(M) \qquad (15)$$

Non-molten globular, natively disordered proteins might be expected to contain subregions of various lengths, each having differing degrees of partial folding. If this were indeed the case, plots of the logarithm of the hydrodynamic volumes versus residue numbers for a set of such natively disordered proteins would give a scatter plot having values randomly located between the lines specified by the extensively unfolded and the molten globular reference forms. Instead of this expectation, two distinct types of natively disordered proteins were observed[123]. One type closely resembles the reference proteins that were extensively unfolded by guanidinium chloride. The other type gives a log volume versus residue line between that of the fully extended protein molecules and that of the molten globules. The line for this second type is nearly superimposable[123] with a previously described denaturation intermediate called the pre-molten globule[136].

The finding of two distinct classes of extended-disorder, random coil-like and premolten globule-like, is very difficult to understand. Polymers in good solvents tend to be collapsed, but still unstructured. Uversky (2002) suggested that sequence differences could be responsible for whether a natively disordered protein behaves more like a random coil or more like a premolten globule: but a clear sequence distinction between the two classes has not been found as of yet. But even if a sequence distinction is found, it remains very difficult to understand the origin of this partition into two distinct classes. For example, if a hybrid protein were created, with one half being random coil-like and one-half being premolten globule-like, would its hydrodynamic properties lie between those of the two classes? Why haven't such chimeras been found in nature? Perhaps the simple explanation is that not enough natively disordered proteins have been characterized. But if a wide variety of natively disordered proteins has not been found in the current sampling even though such proteins do exist, what is leading to the partition observed so far? This question deserves further study.

iv. Circular Dichroism Spectropolarimetry

Circular dichroism (CD) measures the difference in the absorbance of left versus right circularly polarized light, and is therefore sensitive to the chirality of the environment[137]. There are two types of optically active chromophores in proteins, side groups of aromatic amino acid residues and peptide bonds[137,138]. CD spectra in the near ultraviolet region (250-350 nm), also known as the aromatic region, reflect the symmetry of the aromatic amino acid environment and, consequently, characterize the protein tertiary structure. Proteins with rigid tertiary structure are typically characterized by intense near-UV CD spectra, with unique fine structure, which is reflective of the unique asymmetric environment of individual aromatic residues. Thus, natively disordered proteins are easily detected since they are characterized by low intensity near-UV CD spectra with low complexity. The far-UV region of a protein's absorbance spectrum (190-240 nm) is dominated by the electronic absorbance from peptide bonds. The far-UV CD spectrum provides quantifiable information about the secondary structure of a protein because each category of secondary structure (e.g., α-helix, β-sheet) has a different effect on the chiral environment of the peptide bond. Because the time scale for electronic absorbance is so much shorter (~10^{-12} seconds) than folding or unfolding reactions (at least ~10^{-6} seconds), CD provides information about the weighted average structure of all the peptide bonds in the beam.

CD provides different information about native globular versus denatured or disordered proteins. In a native globular protein, the weighted average can be applied to each molecule because nearly all the molecules are in the native state and each molecule of the native state has a similar structure. For instance, if the CD data indicate that 30% of the peptide bonds in a sample of a 100-residue native globular protein are in the α-helical conformation and 70% are in the β-sheet conformation, then 30 peptide bonds in each protein molecule are helical and the other 70 are sheet. Furthermore, it is the same 30 and 70 in each molecule. For a denatured or disordered protein, the CD spectrum provides information about the ensemble of conformations, and the information will not generally be applicable to any single molecule. For instance, changing the above example to a disordered protein, we can say that 30% of the peptide bonds in the sample are helical and 70% are sheet, but with only these data in hand nothing can be inferred about the structure of an individual molecule -- neither the helix and sheet percentage nor the location of those structures along the chain.

It has been demonstrated that a good correlation exists between the relative decrease in hydrodynamic volume and the increase in secondary structure content. This was shown for a set of 41 proteins from the literature that had been evaluated by both far-UV CD and hydrodynamic methods[139]. Study of the equilibrium unfolding of these globular proteins revealed that the Stokes radii (R_S) and secondary structure and of native and partially folded intermediates were closely correlated. Results of this analysis are presented in Figure 4 as the dependence of $(R_S^U/R_S)^3$ versus $[\theta]_{222}/[\theta]_{222}^{U}$, which represent relative compactness and relative content of ordered secondary structure, respectively. Significantly, Figure 4 illustrates that data for both classes of conformations (native globular and partially folded intermediates) can be accurately described by the common dependence (correlation coefficient r^2=0.97)[139]:

$$\left(\frac{R_S^U}{R_S}\right)^3 = (1.047 \pm 0.010) \bullet \left(\frac{[\theta]_{222}}{[\theta]_{222}^{U}}\right) - (0.31 \pm 0.12) \qquad (16)$$

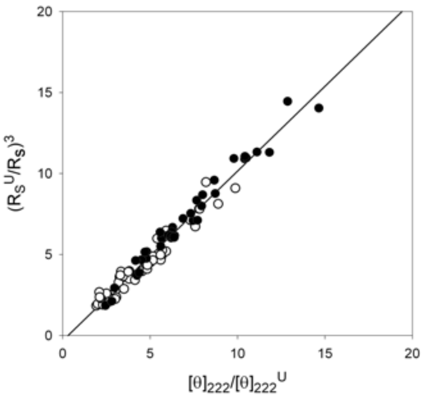


Figure 4. Correlation between the degree of compactness and the amount of ordered secondary structure for native globular proteins (filled circles) and their partially folded intermediates (open circles). The degree of compactness, $(R_S^U/R_S)^3$, was calculated for different conformational states as the decrease in hydrodynamic volume relative to the volume of the unfolded conformation while the amount of ordered secondary structure, $[\theta]_{222}/[\ \theta]_{222}^U$, was calculated from far UV CD spectra as the increase in negative ellipticity at 222 nm relative to the unfolded conformation. The data used to plot these dependencies are taken from Uversky[139].

This means that the degree of compactness and the amount of ordered secondary structure are conjugate parameters. In other words, there is no compact equilibrium intermediate lacking secondary structure or any highly ordered but non-compact species among the proteins analyzed. Therefore, hydrophobic collapse and secondary structure formation occur simultaneously rather than as two subsequent independent processes. This conclusion generalizes earlier observations made for several individual proteins including DnaK[140], apo-myoglobin[141] and staphylococcal nuclease[142].

Tiffany and Krimm noted over 35 years ago that CD spectra of reversibly denatured proteins resemble that of a left-handed polyproline II helix[143]. It was only recently, however, that Creamer revisited these data and put them on a more sound footing with studies of non-proline model peptides[144,145]. Polyproline helices made from non-proline residues have a negative CD band at 195 nm and positive band at 218 nm. These wavelengths are shorter than those of true polyproline helices, (205 nm and 228 nm, respectively) because of the different absorption maxima for primary and secondary amides. Inspection of the literature shows that many intrinsically-disordered proteins exhibit a negative feature near 195 nm with near zero ellipticity near 218 nm[17,18,146,147,148,149], and some exhibit both the negative and the positive features[150,151]. The interpretation of CD spectra in terms of polyproline II helix content remains controversial for several reasons. First, it is not yet clear what the difference is between the CD spectrum of a random coil and the CD spectrum of a polyproline II helix. Second, there is not yet a way to quantify the amount of polyproline II helix.

The conclusion is clear; many intrinsically disordered proteins resemble reversibly denatured proteins and may exist, on average, in a polyproline II helix. However, some of these proteins exhibit small amounts of other secondary structures[77]. Sometimes[102,152,153,154,155,156,157], but not always[148,158], secondary structure can be induced by molecular crowding and "structure-inducing" co-solutes.

v. Infrared Spectroscopy

Infrared (IR) spectra are the result of intramolecular movements (bond stretching, change of angles between bonds along with other complicated types of motion) of various functional groups (e.g., methyl, carbonyl, amide, etc.)[159]. The merits of spectral analysis in the IR region originate from the fact that the modes of vibration for each group are very sensitive to changes in chemical structure, molecular conformation and environment. In the case of proteins and polypeptides, two infrared bands that are connected with vibrational transitions in the peptide backbone and reflect the normal oscillations of simple atom groups are of the most interest. These bands correspond to the stretching of N-H and C=O bonds (amide I band) and the deformation of N-H bonds (amide II band). The amide I and amide II bands are characterized by the frequencies within the ranges of 1600-1700 and 1500-1600 cm^{-1}, respectively[159,160,161]. The position of the bands change due to the formation of hydrogen bonds. Thus, analysis of IR spectra allows determination of the relative content of α-helical, β and irregular structure in proteins by monitoring the intensity of the bands within the amide I and amide II regions.

The main advantage of using IR-spectroscopy to determine the secondary structure of protein is that this method is based on a simple physical phenomenon, the change of vibrational frequency of atoms upon the formation of hydrogen bonds. Thus, it is possible to calculate the parameters of normal vibrations of the main asymmetrical units forming secondary structure and to compare these calculated values with experimental data. Calculations of normal vibrational parameters for α-helices and β-sheets were done by Miyazawa[162]. Further, Chirgadze et al. have shown that there is a good correlation between the calculated and experimental data for the amide I and amide II bands[159,161]. Thus, such calculations may serve as a rational basis for the interpretation of polypeptide and protein IR spectra (Figure 5). The application of IR spectroscopy to evaluate protein secondary structure is based on the following assumptions about protein structure: 1) protein consists of a limited set of different types of secondary structure (α-helix, parallel and antiparallel β-structures, β-turns and irregular structure), 2) the IR spectra of protein is a simple sum of the spectra of these structures taken with the weights corresponding to their content in the protein, 3) the spectral characteristics of secondary structure are the same for all proteins and for all the structural elements of one type in the protein. The finding that Fourier-transform infrared spectroscopy (FTIR) exhibited a high sensitivity to the conformational state of macromolecules resulted in numerous studies where this approach was used to analyze protein molecular structure (including a number of natively disordered proteins) and to investigate the processes of protein denaturation and renaturation.

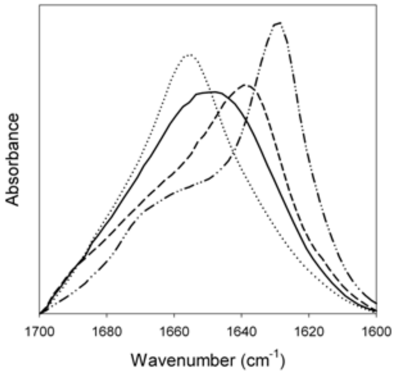


Figure 5. FTIR spectra in the amide I region measured for the natively disordered α-synuclein (solid line), α -helical human α-fetoprotein (dotted line), β-structural *Yersinia pestis* capsular protein Caf1 (dashed line) and α-synuclein amyloid fibril with cross-β-structure (dash-dot-dotted line). Spectra are normalized to have same area.

It is necessary to emphasize that FTIR also provides a means of keeping track of conformational changes in proteins. These are followed by monitoring the changes in the frequencies of the IR bands that result from deuteration (substitution of hydrogen atoms by deuterium) of the molecule[159]. Since it is usually known which band belongs to which functional group (carbonyl, oxy- or amino group), one can identify the exchangeable groups by observing the changes in band position as a result of deuteration. The rate of deuteration depends on the accessibility of a given group to the solvent that, in turn, varies according to changes in conformation. Thus, by tracking the changes in the deuterium-exchange rates for different solvent compositions and other environmental parameters, one can obtain information about the resultant conformational changes within a given protein.

vi. Fluorescence Methods

Intrinsic fluorescence of proteins

Proteins contain only three residues that have the property of intrinsic fluorescence. These chromophores form the following series: tryptophan > tyrosine > phenylalanine according to their quantum yield. The fluorescence of tryptophan is most commonly used for analysis of proteins since the quantum yield of phenylalanine fluorescence is extremely low and tyrosine fluorescence is strongly quenched in the majority of cases. Quenching of tyrosine fluorescence can be due to ionization, location near amide or carboxyl groups, or due to the energy transfer to tryptophan[163]. Application of intrinsic fluorescence to the study of protein conformational analysis relies on the fact that the parameters of tryptophan emission (intensity and wavelength of maximal fluorescence) depend essentially on environmental factors, including solvent polarity, pH, and presence or absence of quenchers[163]. For example, a completely solvated tryptophan residue (e.g., free tryptophan in water or tryptophan in an unfolded polypeptide chain) has a maximum fluorescence in the vicinity of 350 nm, whereas embedding this chromophore into the non-polar interior of a compact globular protein results in a characteristic blues shift of its fluorescence maximum (Stokes shift) by as much as 30-40 nm[163,164,165]. This means that the value of λ_{max} of tryptophan fluorescence contains some basic information about whether the given protein is compact or not under the experimental conditions. For this reason, the analysis of intrinsic protein fluorescence is frequently used for the study of protein structure and conformational change.

Dynamic quenching of fluorescence

Additional information about the accessibility of protein chromophores to solvent (and, thus on relative compactness of a protein molecule) can be obtained from the analysis of dynamic quenching of intrinsic fluorescence by small molecules. Fluorescence quenching data are frequently analyzed using the general form of the Stern-Volmer equation[166]:

$$\frac{I_0}{I} = (1 + K_{SV}[Q])e^{V[Q]} \qquad (17)$$

where I_o and I are the fluorescence intensities in the absence and presence of quencher, K_{SV} is the dynamic quenching constant, V is a static quenching constant, and $[Q]$ is the quencher concentration.

To some extent, the information obtained from dynamic quenching of intrinsic fluorescence is similar to that obtained from studies of deuterium exchange, since it reflects the accessibility of defined protein groups to the solvent. However, in distinction from the deuterium exchange, this method can be used to evaluate the amplitude and time scale of dynamic processes by using quenchers of different size, polarity and charge. In one example of this approach, it has been shown that the rate of diffusion of oxygen (which is one of the smallest and most efficient quenchers of intrinsic protein fluorescence) within a protein molecule is only 2-4 times slower than in an aqueous solution[167,168]. Furthermore, oxygen was shown to affect even those tryptophan residues that, according to X-ray structural analysis, should not be accessible to the solvent. These observations clearly demonstrated the presence of substantial structural fluctuations in proteins on the nanosecond timescale[167,168].

Acrylamide is one of the most widely used quenchers of intrinsic protein fluorescence[169,170]. Acrylamide, like oxygen, is a neutral quencher but with a much larger molecular size. This size difference results in a dramatic decrease in the rate of protein fluorescence quenching over that of oxygen[170,171]. This decrease is due to the inaccessibility of the globular protein interior to the acrylamide molecule. Thus, acrylamide actively quenches only the intrinsic fluorescence of solvent exposed residues. As applied to conformational analysis, acrylamide quenching was shown to decrease by two orders of magnitude as unstructured polypeptide chains transitioned to globular structure[170,171]. Importantly, the degree of shielding of tryptophan residues by the intramolecular environment of the molten globule state was shown to be close to that determined for the native globular proteins, whereas the accessibility of tryptophans to acrylamide in the pre-molten globule state was closer to that in the unfolded polypeptide chain[142]. The fluorescence quenching by acrylamide of the single tryptophan residue in the beta 2 subunit of tryptophan synthase was used to verify the presence of a conformational transition induced by interaction with the cofactor, pyridoxal 5'-phosphate[172].

Simultaneous application of quenchers of different size, polarity and charge (oxygen, nitrite, methylvinylketone, nitrate, acrylate, acrylamide, acetone, methylethylketone, succinimide, etc.) could be more informative since it may yield information not only about protein dynamics but also about peculiarities of the local environment of chromophores (see below). Note however, that information on the local environment of chromophores could be also retrieved from simple quenching experiments. In an example of simultaneous application of multiple quenchers, the heterogeneous fluorescence of yeast 3-phosphoglycerate kinase was resolved into two approximately equal components, one accessible and one inaccessible to the quencher succinimide[173]. The fluorescence of the inaccessible component was shown to be blue-shifted and exhibited a heterogeneous fluorescence decay which had a temperature-dependence and steady-state acrylamide quenching properties typical of a single tryptophan in a buried environment. This component was assigned to the buried tryptophan W333. The presence of succinimide greatly simplified the fluorescence, allowing the buried tryptophan to be studied with little interference from the exposed tryptophan[173].

Fluorescence polarization and anisotropy

Useful information about the mobility and aggregation state of macromolecules in solution can be obtained from analysis of fluorescence polarization or anisotropy. If excited light is polarized and passed through a protein solution, fluorescence will be depolarized or remain partially polarized. The degree of fluorescence depolarization results from the following factors that characterize the structural state of the protein molecules: 1) mobility of the chromophores, (strongly dependent on the density of their environment) and 2) energy transfer between similar chromophores[163,165,174,175,176,177,178]. Furthermore, the relaxation times of tryptophan residues determined from polarized luminescence data are a reliable indicator of the compactness of the polypetide chain. For example, it has been noted that the retention of intact disulfide bonds in the unfolded state often results in a non-essential decrease of intrinsic fluorescence polarization, whereas reduction of the disulfide bonds leads to a dramatic decrease in the luminescence polarization to values that are approximately equal for all proteins[177,178]. The relaxation times of tryptophan residues determined by fluorescence polarization for α-lactalbumin[23,179] and bovine carbonic anhydrase B[180] showed high degree of protein compactness in both the native and the molten globule states.

Fluorescence resonance energy transfer

Along with intrinsic protein fluorescence, the fluorescence of extrinsic chromophore groups is widely used in conformational studies. Extrinsic chromophores are divided into covalently attached labels and non-covalently interacting probes according to the type of interaction desired. Fluorescence labels are indispensable tools in studies of energy transfer between two chromophores. The essence of the phenomenon is that in the interaction of oscillators at small distances the electromagnetic field of the excited (donor) oscillator can induce oscillation in the non-excited (acceptor) oscillator[163,165,181]. It should be noted that the transfer of excitation energy between the donor and the acceptor originates only upon the fulfillment of several conditions: 1) the absorption (excitation) spectrum of the acceptor overlaps with the emission (luminescence) spectrum of the donor (an essential prerequisite for resonance), 2) spatial proximity of the donor and the acceptor (within a few dozen Angstroms), 3) a sufficiently high quantum yield of the donor; and 4) a favorable spatial orientation of the donor and acceptor. The biggest advantage, and hence the attractiveness, of fluorescence resonance energy transfer (FRET) is that it can be used as molecular ruler to measure distances between the donor and acceptor. According to Förster, the efficiency of energy transfer, E, from the excited donor, D, to the non-excited acceptor, A, located from the D at a distance R_{DA} is determined by the equation[181]:

$$E = \frac{1}{1+\left(\frac{R_{DA}}{R_o}\right)^6} \tag{18}$$

where R$_o$ is the characteristic donor-acceptor distance, so-called Förster distance, which has a characteristic value for any given donor-acceptor pair given by the equation[181]:

$$R_o^6 \quad = \quad \frac{9000 \ln 10}{128\pi^5 N} \frac{\langle k^2 \rangle \phi_D}{n^4} \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \tag{19}$$

where the parameter ϕ_D is the fluorescence quantum yield of the donor in the absence of the acceptor, *n* is the refractive index of the medium, *N* is the Avogadro's number, λ is the wavelength, $F_D(\lambda)$ is the fluorescence spectrum of the donor with the total area normalized to unity, and $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor. The information about these parameters were obtained directly from independent experiments. Finally, $\langle k^2 \rangle$ represents the effect of the relative orientations of the donor and acceptor transition dipoles on the energy transfer efficiency. For a particular donor-acceptor orientation this parameter is given as:

$$k = (\cos\alpha - 3\cos\beta\cos\gamma) \tag{20}$$

where α is the angle between the transition moments of the donor and the acceptor, and β and γ are the angles between the donor and acceptor transition moments and the donor-acceptor vector, respectively. Experimentally, the efficiency of direct energy transfer, *E*, is calculated as the relative loss of donor fluorescence due to the interaction with acceptor[163,181]:

$$E = 1 - \frac{\phi_{D,A}}{\phi_D} \tag{21}$$

where ϕ_D and $\phi_{D,A}$ are the fluorescence quantum yields of the donor in the absence and the presence of acceptor, respectively.

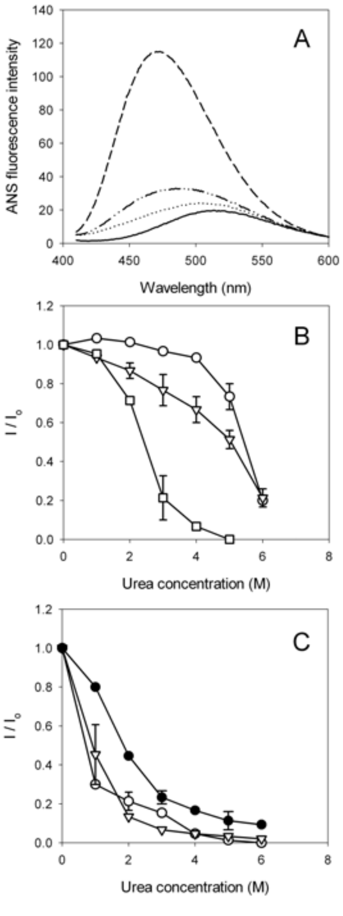
For FRET experiments, one typically uses the intrinsic chromophores (tyrosines or tryptophanes) as donors and covalently attached chromophores (such as dansyl) as acceptors. According to Equation 18, the efficiency of energy transfer is proportional to the inverse sixth power of the distance between donor and acceptor. Obviously, structural changes within a protein molecule might be accompanied by changes in this distance, giving rise to the considerable changes in energy transfer. FRET has been used to show that the urea-induced unfolding of proteins is accompanied by a considerable increase in hydrodynamic dimensions. This expansion resulted in a significant decrease in the efficiency of energy transfer from the aromatic amino acids within the protein (donor) to the covalently attached dansyl (acceptor).

An elegant approach based on the unique spectroscopic properties of nitrated tyrosine (which has maximal absorbance at ~ 350 nm, does not emit light and serves as an acceptor for tryptophan) has been recently elaborated[182,183,184,185,186,187]. For these experiments tyrosine residues are modified by reaction with tetranitromethane to convert them to a nitro-form, Tyr(NO_2). The extent of decrease of tryotophan fluorescence in the presence of Tyr(NO_2) provides a measure of average distance (R_{DA}) between these residues. This approach has been applied to study the 3-D structure of apomyolgobin in different conformational states[187]. These conformations included the native, molten globule, and unfolded conformations. The A-helix of horse myoglobin contains Trp7 and Trp14, the G-helix contains Tyr103 and the H-helix contains Tyr146. Both tyrosine residues can be converted successively into the nitro-form[186]. Comparison of tryptophan fluorescence in unmodified and modified apomyolgobin permits the evaluation of the distances from tryptophan residues to individual tyrosine residues (i.e., the distances between identified points on the protein). Employing this energy transfer method to a variety of non-native forms of horse apomyolgobin revealed that the helical complex formed by the A-, G- and H-helicies, exists in both the molten and pre-molten globule states[187].

ANS fluorescence

Hydrophobic fluorescent probes can be used to detect the hydrophobic regions of protein molecules exposed to the solvent. Hydrophobic fluorescent probes characteristically exhibit intense fluorescence upon interaction with protein and low fluorescence intensity in aqueous solutions. One such probe that is frequently used for studying the structural properties of protein molecules, is 8-anilino-1-naphthalene sulfonate (ANS)[188,189,190]. Interest in this probe reached its peak after it was shown that there was a predominant interaction of ANS with the equilibrium and kinetic folding intermediates in comparison with native and completely unfolded states of globular proteins[180,190,191,192,193]. The interaction of ANS with protein molecules in the molten globular state is accompanied by a pronounced blue shift of maximal fluorescence and a significant increase of the probe fluorescence intensity, making the latter property a useful tool for the detection of partially folded intermediates in the process of protein folding (Figure 6a). Furthermore, the interaction of ANS with protein is also accompanied by a change in the fluorescence lifetime[192,193]. Fluorescence decay of free ANS is well described by the monoexponential law. However, in the case of formation of complexes between ANS and proteins, fluorescence decay has a more complex dependence. In fact, analysis of the ANS fluorescence lifetimes of a number of proteins revealed that there are at least two types of ANS-protein complexes. The complexes of the first type are characterized by a fluorescence lifetime ranging from 1 to 5 ns. In this type, the probe molecules are bound to the surface hydrophobic clusters of the protein and remain relatively accessible to the solvent. Complexes of the second type are characterized by a fluorescence lifetime of 10 to 17 ns with the probe molecules embedding themselves inside the protein molecule and therefore poorly accessible to the solvent[192]. The interaction of ANS with both molten globules and pre-molten globules of different proteins exhibit fluorescence lifetimes characteristic of the second type with molten globules reacting more strongly than pre-molten globules[134]. Furthermore, an increased affinity for ANS was shown to be a characteristic property of several natively disordered proteins.

Figure 6. ANS interaction with proteins. **(a)** ANS fluorescence spectra measured for free dye (solid line) and in the presence of natively disordered coil-like α-synuclein (dotted line), natively disordered pre-molten globule-like caldesmon 636-771 fragment (dash-dot-dotted line) and molten globule state of α-lactalbumin (dashed line). Note native molten globular domain of clusterin has an ANS spectrum comparable with that of the molten globular state of α-lactalbumin. **(b)** Dependence of the ANS fluorescence intensity on urea concentration measured for a series of rigid globular proteins capable of ANS binding: bovine serum albumin (circles); apomyoglobin, pH 7.5 (squares); and hexakinase (inverse triangles). **(c)** Dependence of the ANS fluorescence intensity on urea concentration measured for a series of molten globular proteins: apomyoglobin, pH 2.0 (open circles); β-lactalbumin, pH 2.0 (open inverse triangles) and clusterin, a protein with a native molten globular domain (black circles).



It is known that in addition to partially folded conformations, some native proteins also possess significant affinity to ANS[188,189,190]. However, ordered proteins unfold cooperatively, whereas the unfolding of molten globular forms are typically much less cooperative[194]. It has been found that urea titration of ANS fluorescence could be used to distinguish between the binding of ANS to the hydrophobic pocket of an ordered protein and the binding of ANS to a molten globule. This experimental approach has been successfully applied in studies of clusterin. The results presented in Figure 6b indicate that this protein likely contains a molten globule-like domain in its native state[21]. This conclusion follows from the comparison of denaturation profiles of clusterin with those for rigid and molten globular forms of a globular protein (Figure 6c).

An alternative method is based on comparing Stern-Volmer quenching curves for a polar quencher, acrylamide, with the quenching curves for a nonpolar quencher, tricholorenthanol (TCE). The essence of this method is based on the following reasoning. If the hydrophobic groups surrounding a chromophore are rigidly packed, then both acrylamide and TCE are excluded from the potential contact with chromophore and therefore exhibit little quenching. On the other hand, if the hydrophobic groups surrounding the chromophore are loosely packed and dynamic, then the hydrophilic quencher, acrylamide, is still excluded and continues to show little quenching. However, the hydrophobic quencher, TCE, partitions into the hydrophobic region surrounding the chromophore, leading to quenching that is much stronger than if the chromophores were completely exposed on the protein surface. These concepts were proven during the characterization of three different forms of fd phage. The fluorescence emission maxima and intensities for the tryptophans in all three forms are nearly identical, suggesting that all of the environments had very similar overall polarities. For the fd filament it was shown that there was a very little difference in the quenching by TCE or acrylamide, suggesting that the indole rings were in tightly packed environments. On the other hand, for the two contracted forms, quenching by TCE was much stronger than the quenching by acrylamide. Furthermore, the quenching of the contracted forms by TCE was shown to be even stronger than the quenching of a free indole ring in water. This result was interpreted as indicating that the residues surrounding the tryptophans in the contracted forms of fd phage were likely to be highly dynamic (similar to the inside of an SDS micelle), thus leading to accumulation of TCE and therefore increased quenching[195]. We believe that this approach has useful features that provided additional insights into the molten globular state.

vii. Conformational Stability

Intrinsic disorder may be detected by the analysis of protein conformational stability. For example, the presence or absence of a cooperative transition on the calorimetric melting curve for a given protein is a simple and convenient criterion indicating the presence or absence of a rigid tertiary structure[196,197,198]. Furthermore, the response of a protein to denaturing conditions may be used to discriminate between collapsed (molten globule-like) and extended intrinsic disorder (coil-like and pre-molten globule-like conformations). In fact, the increase in temperature and changes in pH (as well as the increase in urea or guanidinium chloride concentration) will induce relatively cooperative loss of the residual ordered structure in molten globule-like disordered proteins, whereas temperature and pH will bring formation of residual structure in native coils and native pre-molten globules (see below). Furthermore, it has been shown that the steepness of urea- or guanidinium chloride-induced unfolding curves depends strongly on whether a given protein has a rigid tertiary structure (i.e., it is native) or is already denatured and exists as a molten globule[194,199]. To perform this type of analysis, the values of $\Delta\nu^{eff}$ (which is the difference in the number of denaturant molecules "bound" to one protein molecule in each of the two states) should be determined. Then this quantity should be compared to the $\Delta\nu^{\mathrm{eff}}_{\mathrm{N}\rightarrow\mathrm{U}}$ and $\Delta\nu^{\mathrm{eff}}_{\mathrm{MG}\rightarrow\mathrm{U}}$ values corresponding to the native to coil and molten globule to coil transitions in globular protein of a given molecular mass, respectively[194].

Effect of temperature on proteins with extended disorder

It has been pointed out that at low temperatures, natively disordered proteins with extended type of disorder show far-UV CD spectra typical of an unfolded polypeptide chain. However, as the temperature increases, the spectrum changes, consistent with temperature-induced formation of secondary structure[123,135]. In fact, such behavior was observed for such natively disordered proteins as α-synuclein[119], phosphodiesterase γ-subunit[200], caldesmon 636-771 fragment[121], extracellular domain of the nerve growth factor receptor[201] and α_s-casein[202]. Thus, an increase in temperature can induce the partial folding of natively disordered proteins, rather than the unfolding typical of globular proteins. The effects of elevated temperatures may be attributed to increased strength of the hydrophobic interaction at higher temperatures, leading to a stronger hydrophobic driving force for folding.

Effect of pH on proteins with extended disorder

Several natively disordered proteins with extended disorder, including α-synuclein[119], prothymosin α[116], pig calpastatin domain I[203], histidine rich protein II[204], naturally occurring human peptide LL-37[205] and several other proteins show intriguing dependence of their structural parameters on pH. In fact, these proteins possess low structural complexity at neutral pH, but were shown to have significant residual structure under conditions of extreme pH. These observations show that a decrease (or increase) in pH induces partial folding of natively disordered proteins due to the minimization of their large net charge present at neutral pH, thereby decreasing charge/charge intramolecular repulsion and permitting hydrophobic-driven collapse to partially folded conformations.

viii. Mass Spectrometry-Based High Resolution Hydrogen-Deuterium Exchange

In order to obtain high-resolution structural information using X-ray crystallography, it is necessary to produce crystals. This can be impossible for proteins with substantial amounts of intrinsic disorder. High resolution structural information can also be provided by NMR, but this technique is limited by protein size and the need for high concentrations of protein. Mass spectrometry-based high resolution hydrogen-deuterium exchange (MSHDX) shows promise in becoming a third source of high resolution structural information.

Monitoring the exchange rate between main chain amides and solvent hydrogens as a method to study the structure of proteins has seen increased usage over the past 40 years[206]. Hydrogen deuterium exchange (HDX) rates are dependent on thermodynamics and dynamic behavior and thus yield information regarding the structural stability of the protein under study. The study of HDX as applied to proteins was initiated by Kaj Linderstrøm-Lang in the 1950s as a method to investigate Pauling's newly-postulated α-helix and β-sheet secondary structures[207]. The two-step model of HDX is described as[208]:

$$C^H \underset{k_{cl}}{\overset{k_{op}}{\Leftrightarrow}} O^H \underset{k_{ch}}{\overset{k_{ch}}{\Leftrightarrow}} O^D \underset{k_{op}}{\overset{k_{cl}}{\Leftrightarrow}} C^D \qquad (22)$$

where H and D denote protonated and deuterated forms, respectively of the "closed" or C (nonexchanging) and "open" or O (exchanging) conformations, k_{op} is the rate constant for the conformational change that exposes the hydrogen, k_{ch} is the rate constant for the chemical exchange reaction, and k_{cl} is the rate constant for the return to the closed conformation.

The fundamental equation for the observed exchange rate formulated by Linderstrøm-Lang still applies:

$$k_{ex} = \left(\frac{k_{op} \cdot k_{ch}}{k_{cl} + k_{op}} \right) \tag{23}$$

The EX2 HDX mechanism, where $k_{cl} > k_{ch}$, describes most proteins at or below neutral pH. In this case, k_{ex} can be found using the following equation:

$$k_{ex} = \left(\frac{k_{op} \cdot k_{ch}}{k_{cl}} \right) = K_{op} \cdot k_{ch} \tag{24}$$

where K_{op} is the unfolding equilibrium constant which is also equal to the ratio of the observed rate (k_{ex}) to that of a random coil amide (k_{ch}).

HDX rates for random coil amides (k_{ch}), which are dependent on pH, primary sequence, and temperature, can be obtained from published studies of peptide behavior[209]. The ratio of the observed rate of exchange (k_{ex}) to that of a random coil (k_{ch}) relates to the free energy of unfolding by the following equation:

$$\Delta G_{op} = -RT \ln\left(\frac{k_{ex}}{k_{ch}} \right) = -RT \ln\left(K_{op} \right) \tag{25}$$

Equation 25 shows that there is a direct correlation between the exchange rate and the conformation stability; i.e., the faster the exchange rate, the less stable the folded state. Being initially developed to study protein folding on a global scale, this formalism could also be used to analyze the intrinsic dynamic behavior of proteins and, what is the most attractive, to quantify the stability of localized regions within protein molecules.

Hydrogen bonding, accessibility to the protein surface, and flexibility of the immediate and adjacent regions affect the HDX rate. The combined contribution of these structural and dynamic factors are termed the “protection factor”, which has been observed to vary as much as

10^8 fold[210]. Based on these observations, it has been suggested that each amide hydrogen in the polypeptide can be viewed as a sensor of the thermodynamic stability of localized regions within the protein structure[211]. Potentially, application of HDX could yield data to near single amide resolution.

In the 1960s, HDX rates were determined using tritium incorporation, size-exclusion chromatography and liquid scintillation counting[212]. Structural resolution remained a limitation of this procedure. Improvements in resolution were facilitated by separating fragments using HPLC and analysis of tritium or deuterium incorporation using mass spectrometric methods[212,213,214]. Recent improvement in automation, consisting of solid phase proteolysis, automated liquid handling and sample preparation, online ESI MS and specialized data reduction software have recently improved throughput and sequence coverage of MSHDX[211]. As high as six amide resolution have been achieved[211]. The use of multiple proteases to generate more overlaps in the peptide map can be used to increase the resolution even further.

Crystallization success rates can be improved by detecting and excluding sequences coding for unstructured regions when designing recombinant expression constructs. As part of a structural genomics effort, 24 proteins from *Thermotoga maritima* were analyzed using MSHDX in order to detect unstructured regions[215]. Prior to HDX analysis, parameters affecting fragmentation were optimized in order to obtain to maximize the number of fragments available for analysis. These conditions included: denaturant concentration, protease type and duration of incubation. As a prerequisite for MSHDX analysis, the ability to generate complete fragment maps using SYQUEST (Thermo Finnigan, San Jose, CA) and DXMS data reduction software was tested for each protein. Satisfactory fragment maps were generated for 21 of the 24 proteins. Next, labeling conditions were optimized for discrimination of fast-exchanging amide protons.

In this procedure, the proteins were subjected to proteolysis prior to exchange and HDX analysis. An exchange duration of 10 s was determined to be sufficient to allow differentiation between deuterated freely solvated and nondeuterated inaccessible amides. MSHDX analysis detected regions of rapidly changing amides that were greater than 10% of total protein length in five of the proteins. From these five, MSHDX analysis led to the construction of deletion mutants for two recalcitrant proteins. Proteins produced from these two deletion mutants were subsequently crystallized, and the structures were solved[215].

ix. Protease Sensitivity

Numerous enzymes catalyze the hydrolysis of peptide bonds, including trypsin, pepsin, carboxy peptidase, and so on. Those that cleave at specific sites within protein chains have long been used to probe protein structure. In the 1920s Wu and others showed that native (ordered) proteins are typically much more resistant to protease digestion than are denatured (disordered) forms[216]. These early indications were substantiated by further experiments over the next ~20 years[217]. Increasing the stability of a protein, for example by ligand or substrate binding, in turn often leads to reduction in the digestion rates[218].

In essentially all biochemistry textbooks, the structure of proteins is described as a hierarchy. The amino acid sequence is the primary structure; the local folding of the backbone into helices, sheets, turns and so on is the secondary structure; while the overall arrangement and interactions of the secondary structural elements is the tertiary structure. Many details of this hierarchy were determined from the first protein 3-D structure[16] but the original proposal for the primary-secondary-tertiary hierarchy was based on differential protease digestion rates over the three types of structure[217].

While acknowledging that flexible protein regions are easy targets for proteolysis, some researchers argued that surface-exposure of susceptible amino acids in appropriate conformations could be important sites for accelerated digestion rates[219]. Indeed, proteins that inhibit proteases, such as soybean trypsin inhibitor[220], have local regions of 3-D structure that fit the active sites and bind irreversibly to their target proteases. Despite the examples provided by the inhibitors, susceptible surface digestion sites in ordered proteins are not very common[221,222]. Despite the paucity of examples, the importance of surface exposure rather than flexibility or disorder seems to have become widely accepted as the main feature determining protease sensitivity.

To consider one well-studied protease, trypsin, most of the lysine and arginine target residues are located on protein surfaces, but as mentioned above very few of these are actually sites of digestion when located in regions of organized structure. Starting with a few protein examples having specific surface accessible digestion sites, attempts were made to use molecular simulation methods to dock the backbones of susceptible residues into the active site of trypsin. These studies suggested the need to unfold more than 10 residues in order to fit each target residue's backbone into trypsin's active site[223]. Indeed, recent studies show that backbone hydrogen bonds are protected from water by being surrounded or wrapped with hydrophobic side chains from nearby residues, with essentially every ordered residue being well wrapped[224]. Backbone hydrogen bond wrapping and binding the backbone deep within trypsin's active site are mutually exclusive, and so the requirement for local unfolding (unwrapping) is easily understood.

Direct support for the importance of local unfolding has been provided by comparison of the digestion rates of myoglobin and apomyoglobin[225]. Myoglobin is digested very slowly by several different proteases. Removal of the heme leads to local unfolding of the F helix. This disordered loop becomes the site of very rapid digestion by several different proteases including trypsin. The trypsin digestion rate following the order to disorder transition of the F helix is at least five orders of magnitude faster.

Many fully ordered proteins, such as the myoglobin example given above, are digested by proteases without observable intermediates of significant size. That is, if digestion progress were followed using PAGE, myoglobin would be seen to disappear and small fragments would appear but mid-sized intermediates would not be observed. In contrast, digestion of the F helix in apomyoglobin leads to two relatively stable mid-sized fragments. The likely cause of the lack of observable mid-sized intermediates for fully ordered proteins is that digestion proceeds at multiple sites within the ordered regions of the protein. Access to these sites is mediated by transient unfolding events within ordered regions that allow initial digestion events. These initial cuts perturb the local structure and inhibit refolding. The sum of these changes leads to the exposure of multiple protease sites, which are rapidly cleaved in a random order. Additional transient unfolding events elsewhere in the structure lead to the same outcome. Thus the regions of the protein where structure has been disturbed become rapidly digested into a multitude of different-sized intermediate fragments without the accumulation of an observable quantity of any particular mid-sized intermediate[225].

The molten globular form has been probed by protease digestion[226]. Protease digestion of molten globules leads rapidly to multiple mid-sized fragments; upon further digestion, the intermediates convert to smaller-sized fragments consistent with nearly complete digestion. The transient stability of several mid-sized intermediates suggests that molten globules have multiple regions of high flexibility that are easy to digest interspersed with more stable structured regions that are resistant to digestion.

At this time we have not found published, systematic studies of fully unfolded, natively disordered proteins. One would expect such proteins to undergo digestion in a random fashion with all of the target residues equally accessible to digestion, with only sequence-dependent effects to modulate the digestion rates at the various sites. If this general picture were true, then proteolytic digestion would lead to rapid conversion of the primary protein into small fragments, without significant amounts of transient, mid-sized intermediates. Although not yet published, the highly disordered phosphatase inhibitor (Calyculin A) apparently undergoes digestion to small fragments without any observable intermediates (S. Shenolikar, personal communication)[227] just as suggested herein. The digestion patterns of extensively unfolded proteins likely resemble those of fully folded proteins: both evidently lack the accumulation of mid-sized intermediates. However, the two types of proteins can be distinguished by their vastly different digestion rates: extensively unfolded proteins would be expected to be digested at least 10^5 times faster than fully folded forms.

x. Prediction from Sequence

Amino acid sequence codes for protein 3-D structure[5]. Since native disorder can be viewed as a type of structure, we wanted to test whether the amino acid compositions of disordered regions differed from the compositions of ordered regions, and, if so, whether the types of enrichments and depletions gave insight into the disorder. Thus, we studied amino acid compositions of collections of both ordered and natively disordered protein regions longer than 30 amino acids. We chose such long regions to lessen the chance that nonlocal interactions would complicate our analysis.

Natively disordered sequences characterized by X-ray diffraction, NMR spectroscopy, and CD spectroscopy all contain similar amino acid compositions and these are very different from the compositions of the ordered parts of proteins in PDB. For example, compared to the ordered proteins, natively disordered proteins contain (on average) 50% less W, 40% less F, 50% less isoleucine, 40% less Y, and 25% or so less C, V, and L, but 30% more K, 40% more E, 40% more P, and 25% more S, Q and R[31]. Thus, natively disordered proteins and regions are substantially enriched in typically hydrophilic residues found on protein surfaces and significantly depleted of hydrophobic, especially aromatic, residues found in protein interiors. These data help to explain why natively disordered proteins fail to form persistent, well organized 3-D structure.

Because ordered and disordered proteins contain significantly different amino acid compositions, it is possible to construct order/disorder predictors that use amino acid sequence data as inputs. By now, several researchers have published predictors of order and disorder[13,30,109,227,228,229,230,231,232,233]. The first predictor, developed by R. J. P. Williams, separated just two natively disordered proteins from a small set of ordered proteins by noting the abnormally low charge/hydrophobic ratio for the two natively disordered proteins. This paper is significant as being the first indication that natively disordered proteins have amino acid compositions that differ substantially from those of proteins with 3-D structure. The predictor by Uversky, et al. [227], also utilizes the relative abundance of charged and hydrophobic groups, but on a much larger set of proteins. Most of the other predictors[30,109,229,230,231,232,233] utilized neural networks.

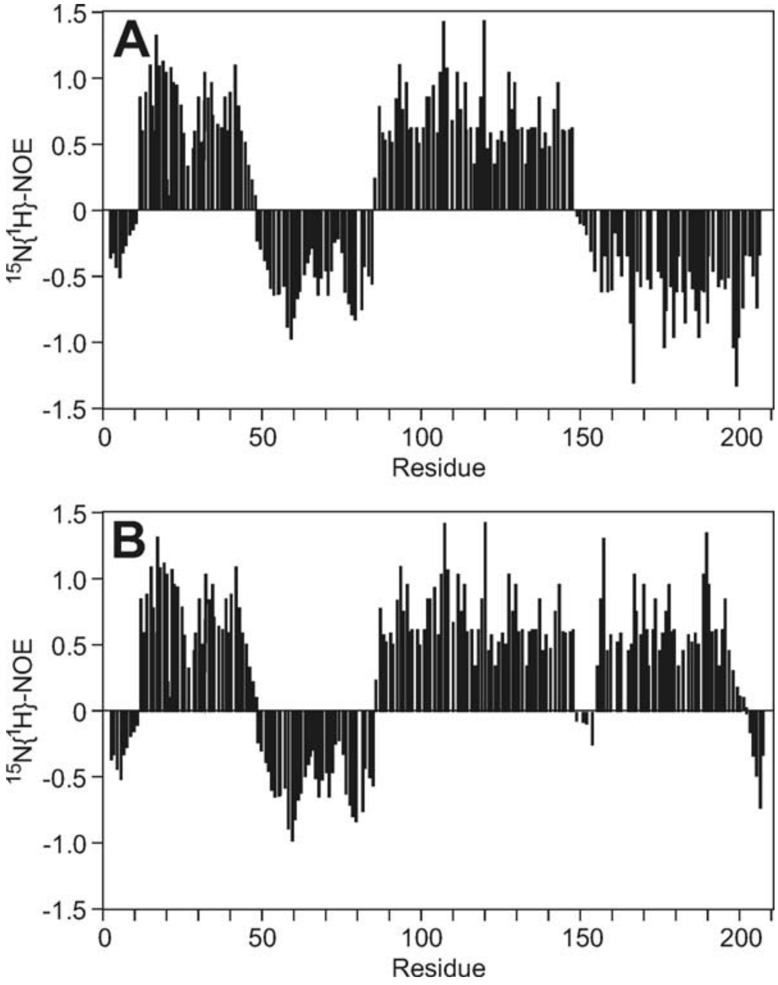
Predicting order and disorder was included in the most recent cycle of the Critical Assessment of Protein Structure Prediction (CASP)[234]. While the summary publication contained papers from two sets of researchers[109,233], several additional groups participated. All of the groups achieved similar levels of accuracy[234]. Furthermore, prediction accuracies were on the order of 90% or so for regions of order and 75-80% for regions of intrinsic disorder. These values are significantly above what could be expected by chance. The predictability of native disorder from sequence further supports the conjecture that natively disordered proteins and regions lack specific 3-D structures as a result of their amino acid sequences. That is, amino acid sequence codes for both order and disorder.

xi. Advantage of Multiple Methods

Given the limitations of the various physical methods, it is useful if natively disordered proteins are characterized by multiple methods. Each approach gives a slightly different view, with better understanding arising from the synthesis of the different perspectives.

A significant difficulty with long disordered regions identified by X-ray diffraction is that the missing coordinates could be indicating either native disorder or a wobbly domain that moves as a folded, ordered, rigid body. Thus, it is especially useful if X-ray-indicated disorder is confirmed by additional methods. The X-ray-indicated regions of disorder in calcineurin[29,235], clotting factor Xa[236,237,238], histone H5 (C. Crane-Robinson, personal communication)[239,240,241], and topoisomerase II (J.M. Berger, personal communication)[242,243,244] were all confirmed by limited proteolysis. The X-ray-indicated regions of disorder in Bcl-X_L[245,246,247,248], the gene 3 protein (g3p) of filamentous phage[249,250,251], and the negative factor of HIV1 (Nef)[252,253,254,255] were all confirmed by NMR spectroscopy. Given the importance of natively disordered regions, it would be helpful if disordered regions indicated by X-ray were routinely subjected to further study by an alternative method.

Figure 7. Idealized steady-state heteronuclear $^{15}N\{^{1}H\}$-NOE for the backbone amides of the hypothetical protein discussed in text. The domain profile of this protein based on the X-ray crystal structure is: short disordered N-terminal region / ordered region / disordered region / ordered region / long disordered C-terminal region. **(a)** $^{15}N\{^{1}H\}$-NOE that is agreement with X-ray determined structure: all regions of missing electron density in the X-ray crystal structure have negative values. **(b)** $^{15}N\{^{1}H\}$-NOE indicating that the C-terminal domain is a wobbly domain (a structured domain attached via a short flexible linker) and not totally disordered as determined by X-ray crystallography.



To illustrate the advantages of combining X-ray and NMR analysis, a hypothetical example is given here. Suppose an X-ray structure reveals that a protein has a short N-terminal disordered region, a central region composed of an ordered region flanked by two disordered segments and a long C-terminal disordered region. Figure 7 shows the expected $^{15}N\{^{1}H\}$-NOE spectra for two different possibilities. Figure 7a shows the expected result if all three regions of missing electron density are truly disordered with very little secondary structure: in this case all three regions of missing electron density give negative values, indicating unfolded, peptide-like motions for these segments. In Figure 7b, the C terminal region shows a short region of disorder followed by a large region of order; this would be consistent with an ordered, but wobbly domain that would lead to missing electron density in the X-ray structure.

The nucleoprotein N of the measles virus contains more than 100 disordered residues at its C-teriminus, a region called Ntail. The disordered character of Ntail was first identified by prediction and was confirmed by Ntail's hypersensitivity to protease digestion and by its NMR and CD spectra[256]. Predictions on the N proteins of related Paramyxoviridae indicate that the Ntail regions, which have hypervariable sequences, are all predicted to be natively disordered[257]. Cloning, expressing and isolation of the Ntail region have enabled multiple studies of this region. The native disorder of this region by now has been confirmed by CD spectroscopy, SAXS, and dynamic light scattering (for the determination of Rs)[257]. The Ntail is apparently of the premolten globule type. Furthermore, although natively disordered and separated from the remainder of the molecule, Ntail retains its biological function of binding to a protein partner[258].

3. Do natively disordered proteins exist inside cells?

i. Evolution of Ordered and Disordered Proteins are Fundamentally Different

The evolution of natively disordered proteins

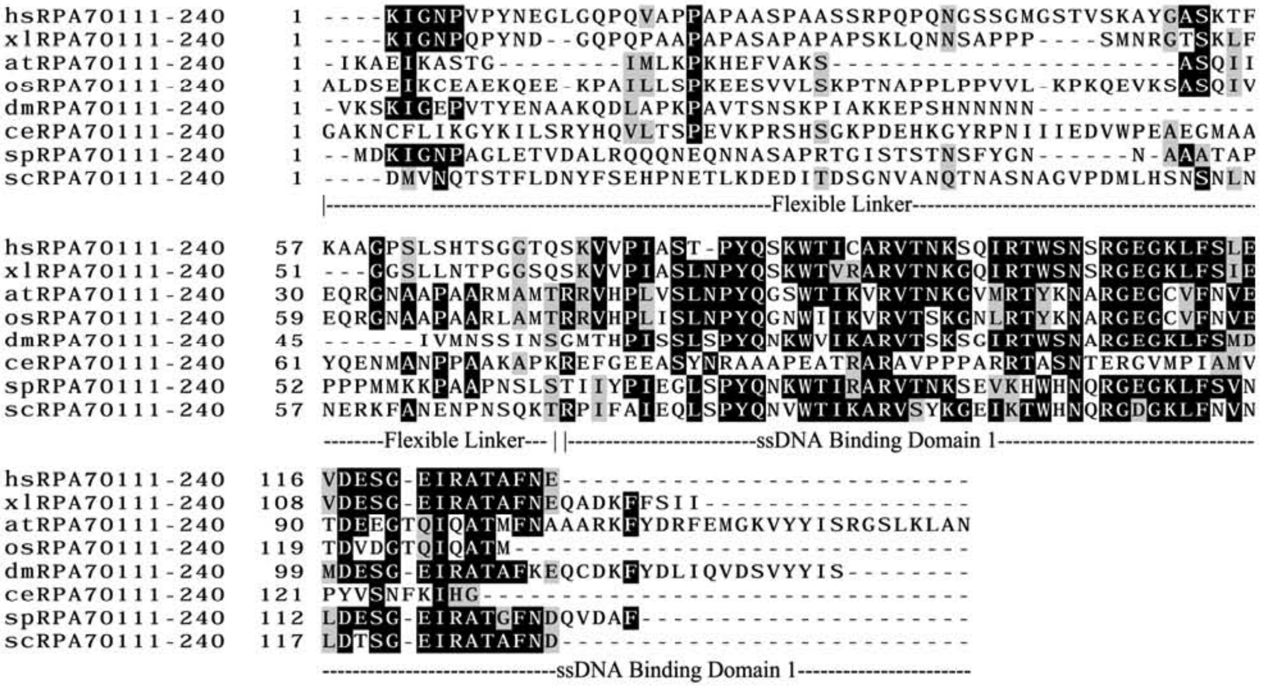
Low resolution structural models of novel proteins can be generated based on sequence similarity and evolutionary relationships to proteins with known structures[259,260,261,262,263,264,265]. This process assumes that the proteins being compared adopt compact rigid structures. Natively disordered proteins have important biological functions and analysis of genome sequence data has revealed proteins with intrinsically disordered segments longer than fifty amino acids are common in nature[20,30,266,267,268,269,270,271]. The lack of information characterizing the partially collapsed flexible structures of proteins limits our ability to predict their existence based on sequence data. It also limits our understanding of how the sequences of such regions specify function, the presence or absence of residual structure, and degree of flexibility.

The evolution of globular protein structures depends on the maintenance of a nonpolar interior and a polar exterior to promote collapse and folding through the hydrophobic effect[261]. This is achieved by the placement and distribution of nonpolar amino acids so that those making favorable nonpolar contacts in the folded structure will be far apart in the linear sequence (so called long range interactions). Therefore, the evolution of globular protein structure is partly dependent on selection for this property. In addition to the hydrophobic effect, desolvation of backbone hydrogen bonds appears to be of similar importance[224].

The evolution of intrinsically disordered protein structure seems to depend on selection for other properties. Natively disordered proteins with extended disorder do not form globular structures and therefore do not have a requirement to maintain long range interactions such as those required by globular proteins. This creates a potential for these sequences to accumulate more variation and generate more sequence divergence than globular proteins (see below). For globular proteins, the selection for structural motifs is apparent by sequence and structural comparison of evolutionarily-related protein structures. In an attempt to develop evolutionary relationships for intrinsically disordered proteins, the Daughdrill lab is currently investigating the structure, dynamics, and function of a conserved flexible linker from the 70 kDa subunit of replication protein A (RPA70). The flexible linker for human RPA70 is ~70 residues long[269,272]. For the handful of sequenced RPA70 homologues, the similarity among linker sequences varies significantly, going from 43% sequence identity between *Homo sapiens* and *Xenopus laevis* to no significant similarity between *Homo sapiens* and *Saccharomyces cerevisiae*. It is unclear what selective processes have resulted in the observed sequence variations among RPA70 linkers. It is also unclear how the observed sequence variation affects the structure and function of the linkers. If natural selection works to preserve flexible structures then one would expect that the linkers from different species have evolved to the same level of flexibility although adopting different sequences. By testing this hypothesis, we will begin understanding the rules governing the evolution of natively disordered protein sequences.

Adaptive evolution and protein flexibility

The ability to align multiple sequences and reconstruct phylogenies based on sequence data is essential to understanding molecular evolution. The development of reliable algorithms that can align multiple sequences and reconstruct phylogenies is encumbered by the presence of highly divergent segments within otherwise obviously related sequences[259,260,273,274]. This problem appears to be especially acute for totally disordered proteins and disordered protein domains. We hypothesize that intrinsically disordered regions will have a higher rate of evolution than compact rigid structures because their rate of evolution is not constrained by by the requirement to maintain long-range interactions. Genetic distance measurements of the RPA70 homologues lend support to this hypothesis and suggest that the linker has evolved at a rate 1.5 to 5 times faster than the rest of RPA70[275]. This study tested the evolutionary rate heterogeneity between intrinsically disordered regions and ordered regions of proteins by estimating the pairwise genetic distances among the ordered and the disordered regions of 26 protein families that have at least one member with region of disorder of 30 or more consecutive residues that has been characterized by X-ray crystallography or NMR. For five of the protein families, there were no significant differences in pairwise genetic distances between ordered and disordered sequences. Disordered regions evolved more rapidly than ordered regions for 19 of the 26 families. The known functions for some of these disordered regions were diverse, including flexible linkers and binding sites for protein, DNA, or RNA. The functions of other disordered regions were unknown. For the two remaining families, the disordered regions evolved significantly more slowly than the ordered regions. The functions of these more slowly evolving disordered regions included sites for DNA binding. According to the authors, much more work is needed to understand the underlying causes of the variability in the evolutionary rates of intrinsically ordered and disordered proteins. Figure 8 illustrates the point, showing the contrast between the protein sequence alignment for the flexible linker and a fragment of the ssDNA binding domain from the eight sequenced RPA70 homologues. Because of the known functional consequences of linker deletion in human RPA70[272,276,277,278,279] we are interested in how the level of functional selection for flexible regions is related to their evolutionary rates. As is observed for folded proteins, most likely this relationship depends on the function under selection[259,260,261,273]. However, because the substitution rate for flexible regions can be uncoupled from the maintenance of long-range interactions, the exact dependence between the rate of evolution and functional selection should differ significantly from that observed for folded proteins. Experiments planned to test the functional consequences of swapping divergent flexible linkers between species will begin to address this subject.



```
hsRPA70111-240     1  ----KIGNPVPYNEGLGQPQVAPPAPAASPAASSRPQPQNGSSGMGSTVSKAYGASKTF
xlRPA70111-240     1  ----KIGNPQPYND--GQPQPAAPAPASAPAPAPSKLQNNSAPPP----SMNRGTSKLF
atRPA70111-240     1  -IKAEIKASTG--------IMLKPKHEFVAKS---------------------ASQII
osRPA70111-240     1  ALDSEIKCEAEKQEE-KPAILLSPKEESVVLSKPTNAPPLPPVVL-KPKQEVKSASQIV
dmRPA70111-240     1  -VKSKIGEPVTYENAAKQDLAPKPAVTSNSKPIAKKEPSHNNNNN-------------
ceRPA70111-240     1  GAKNCFLIKGYKILSRYHQVLTSPEVKPRSHSGKPDEHKGYRPNIIIEDVWPEAEGMAA
spRPA70111-240     1  --MDKIGNPAGLETVDALRQQQNEQNNASAPRTGISTSTNSFYGN------N-AAATAP
scRPA70111-240     1  ----DMVNQTSTFLDNYFSEHPNETLKDEDITDSGNVANQTNASNAGVPDMLHSNSNLN
                      |-----------------------------Flexible Linker--------------------------

hsRPA70111-240    57  KAAGPSLSHTSGGTQSKVVPIAST-PYQSKWTICARVTNKSQIRTWSNSRGEGKLFSLE
xlRPA70111-240    51  ---GGSLLNTPGGSQSKVVPIASLNPYQSKWTVRARVTNKGQIRTWSNSRGEGKLFSIE
atRPA70111-240    30  EQRGNAAPAARMAMTRRVHPLVSLNPYQGSWTIKVRVTNKGVMRTYKNARGEGCVFNVE
osRPA70111-240    59  EQRGNAAPAARLAMTRRVHPLISLNPYQGNWIIKVRVTSKGNLRTYKNARGEGCVFNVE
dmRPA70111-240    45  ------IVMNSSINSGMTHPISSLSPYQNKWVIKARVTSKSCIRTWSNARGEGKLFSMD
ceRPA70111-240    61  YQENMANPPAAKAPKREFGEEASYNRAAAPEATRARAVPPPARRTASNTERGVMPIAMV
spRPA70111-240    52  PPPMMKKPAAPNSLSTIIYPIEGLSPYQNKWTIRARVTNKSEVKHWHNQRGEGKLFSVN
scRPA70111-240    57  NERKFANENPNSQKTRPIFAIEQLSPYQNVWTIKARVSYKGEIKTWHNQRGDGKLFNVN
                      --------Flexible Linker--- | |-----------------------ssDNA Binding Domain 1--------------------------------

hsRPA70111-240   116  VDESG-EIRATAFNE-----------------------
xlRPA70111-240   108  VDESG-EIRATAFNEQADKFFSII------------------
atRPA70111-240    90  TDEEGTQIQATMFNAAARKFYDRFEMGKVYYISRGSLKLAN
osRPA70111-240   119  TDVDGTQIQATM-----------------------------
dmRPA70111-240    99  MDESG-EIRATAFKEQCDKFYDLIQVDSVYYIS--------
ceRPA70111-240   121  PYVSNFKIHG-------------------------------
spRPA70111-240   112  LDESG-EIRATGFNDQVDAF---------------------
scRPA70111-240   117  LDTSG-EIRATAFND-------------------------
                      -----------------------------ssDNA Binding Domain 1---------------------
```

Figure 8. Protein sequence alignment of the RPA70 flexible linker and a fragment of the first ssDNA binding domain from *Homo sapiens* (hs), *Xenopus laevis* (xl), *Arabidopsis thaliana* (at), *Oriza sativa* (os), *Drosophila melanogaster* (dm), *Caenorhabditis elegans* (ce), *Saccharomyces pombe* (sp), and *Saccharomyces cerevisiae* (sc). Dark shading indicates identity and light shading indicates conservative substitutions. The alignment was performed using Clustal 1.8 for residues 111-240 of all eight sequences. The alignment shows the stark contrast in the level of sequence similarities between the ssDNA binding domain and the flexible linker regions.

Phylogeny reconstruction and protein structure

We are interested in developing reliable algorithms that can align multiple sequences and reconstruct phylogenies for highly divergent sequences. We hypothesize that flexible linkers and other disordered protein domains will often be highly divergent. Recent efforts at phylogeny reconstruction using models that incorporate the effects of secondary structure and solvent accessibility on amino acid substitution rate and type have yielded significantly improved maximum likelihood scores[280,281,282]. These models depend on an empirically determined three dimensional structure for at least one member in a homologous family and incorporate four classes of secondary structure (helix, sheet, turn, and coil) as well as two solvent accessibility classes (buried and exposed). In particular, the inclusion of a flexible secondary structure class (coil) is necessary to model the evolution of natively disordered protein domains. The coil category was based on relatively short regions (≤ 5 residues) whose ends are constrained by the compact globular part of the protein. For proteins with longer flexible regions, the current model must be adapted to incorporate structural classes that are based on the presence of residual secondary structure and the degree of flexibility estimated from NMR relaxation measurements.

Another important feature of protein evolution is the pattern of amino acid substitutions observed over time. For ordered proteins, the change of an amino acid into one of a similar chemical type is commonly observed whereas the change to a chemically dissimilar one is rare. For example, isoleucine to leucine, aspartate to glutamate, and arginine to lysine are all commonly observed in related ordered proteins while tryptophan to glycine, lysine to aspartate, and leucine to serine are all very uncommon changes, especially for buried residues.

Patterns of amino acid substitutions are readily observed in substitution matrices used to assign values to various possible sequence alignments. These scoring matrices are constructed by first assembling a set of aligned sequences, typically with low rates of change in pairs of sequences so that there is confidence in the alignments. Given such a set of high confidence alignments, the frequencies of the various amino acid substitutions are then calculated and used to build a scoring matrix. Important and commonly used scoring matrices are the PAM[283] and the Blossum[284] series.

To improve alignments of disordered regions, a scoring matrix for disordered regions was developed[285]. First, homologous groups of disordered proteins were aligned by standard protocols using the Blossum62 scoring matrix. From this set of aligned sequences, a new scoring matrix was calculated then realignment was carried out with this new scoring matrix. As compared to the first alignment, the new alignment was improved as estimated by a reduction in the sizes and number of gaps in the pairwise alignments. The new set of alignments was then used to develop a new scoring matrix, and a new set of alignments was generated. These steps were repeated in an iterative manner until little or no change was observed in successive sets of alignments and in successive scoring matrices.

The resulting scoring matrix for disordered regions showed significant differences from the Blossum 62 or PAM 250 scoring matrices[286]. Glycine/tryptophan, serine/glutamate, alanine/lysine substitutions, for example, were much more common in aligned regions of native disorder as compared to aligned regions of ordered proteins. Since natively disordered regions lack specific structure and the accompanying specific amino acid interactions, substitutions of disparate amino acids are not so strongly inhibited by the need to conserve structure. Thus the commonness of such disparate substitutions in natively disordered regions is readily understood.

The commonly observed higher rates of amino acid change[287] and the distinctive pattern of amino acid substitutions[286] both strongly support the notion that native disorder exists inside the cell.

ii. Direct Measurement by NMR

The inside of cells are extremely crowded, and proteins themselves do most of the crowding since they occupy 40% of a cell's volume and achieve concentrations of greater than 500 g/L[288,289,290]. Despite the crowded nature of the cell's interior, almost all proteins are studied outside cells and in dilute solutions--i.e., total solute concentrations of less than 1 g/L. It is also clear that macromolecular and small-molecule crowding can increase protein stability[291,292,293].

Could some disordered proteins be artifacts of the way proteins are studied? Are some just unstable proteins that unfold in dilute solution? A recent study by Dedmon et al.[102] on a protein called FlgM shows that the answer to the question is yes and no. Some disordered proteins probably have structure in cells, others do not.

Bacteria use rotating flagella to move through liquids[294,295]. The protein FlgM is part of the system controlling flagellar synthesis. It binds the transcription factor σ^{28}, arresting transcription of the genes encoding the late flagellar proteins. Transcription can resume when FlgM leaves the cell, most probably via extrusion through the partially assembled flagellum.

Free FlgM is mostly disordered in dilute solution, but NMR studies in dilute solution indicate that the C-terminal half of FlgM becomes structured on binding σ^{28} as shown by the disappearance of crosspeaks from residues in the C-terminal half of FlgM in the FlgM-σ^{28} complex[77,267]. One signature of protein structure can be the absence of crosspeaks in ^{1}H-^{15}N HSQC NMR spectra because of conformational exchange[296,297,298]. The disappearance of cross peaks results from chemical exchange between a disordered and more ordered form. Specifically, cross peaks broaden until they are undetectable when the rate of chemical exchange between states is about the same as the difference in the resonance frequencies of the nuclei undergoing exchange. The bipartite behavior of FlgM (i.e., disappearance of crosspeaks from the C-terminal half with retention of crosspeaks from the N-terminal half) provides a valuable built-in control for studying the response of FlgM to different solution conditions.

How do these observations about the ability of FlgM to gain structure on binding its partner relate to what might happen in cells? Until recently, all protein NMR was performed *in vitro* on purified protein samples in dilute solution. Two years ago, Dötsch and colleagues showed the feasibility of obtaining the spectra of ^{15}N-labeled proteins inside living *Escherichia* *coli* cells[299,300,301]. Overexpression is key to the success of in-cell NMR. The protein of interest must contain a large proportion of the ^{15}N in the sample so that the spectrum of the overexpressed protein can be observed on top of signals arising from other ^{15}N-enriched proteins in the cell, which contribute to a uniform background.

^{15}N-enriched FlgM was found to give excellent in-cell NMR data[102]. About half the crosspeaks disappear in cells. Most importantly, the crosspeaks that disappear are the same ones that disappear on σ^{28} binding in simple buffered solution, and the crosspeaks that persist are the same ones that persist on σ^{28} binding. These data suggest that FlgM gains structure all by itself under the crowded conditions found in the cell, but it is important to rule out alternative explanations. There is a homolog of *S. typhimurium* σ^{28} in *E. coli* but there is not enough of the homolog present in *E. coli* (i.e., FlgM is overexpressed, the σ^{28} homolog is not) for σ^{28}-FlgM binding to explain the results. Furthermore, the same behavior is observed *in vitro* -- in the complete absence of σ^{28} -- when intracellular crowding was mimicked by using 450 g/L glucose, 400 g/L bovine serum albumin, or 450 g/L ovalbumin. The lack of cross peaks from the C-terminal half of the protein is not caused by degradation because the FlgM can be isolated intact at the end of the in-cell experiment. The gain of structure in cells does not seem to be an artifact of FlgM overexpression because the total protein concentration is independent of FlgM expression. Two observations show that the presence or absence of cross peaks is not simply a matter of viscosity. First, crosspeaks from the N-terminal half of FlgM are present under all conditions tested even though the relative viscosities of the solutions differ dramatically. Second, the absence of crosspeaks does not correlate with increased viscosity. Taken together, these data strongly suggest that even in the absence of its binding partner, the C-terminal portion of FlgM gains structure in cells and in crowded *in vitro* conditions.

Does this observation about the C-terminal part of FlgM mean all intrinsically disordered proteins will gain structure under crowded conditions? No. First, the N-terminal part of FlgM remains disordered under crowded conditions both inside and outside the cell. Second, the same observation has been made for another protein under crowded conditions *in vitro*[302]. Third, as discussed below, several functions of disordered proteins require the absence of stable structure. Perhaps the N-terminal half of FlgM only gains structure upon binding some yet unknown molecule, or maybe it needs to remain disordered to ensure its exit from the cell. It is also important to note, however, that crowding can induce compaction even when it does not introduce structure[158].

In summary, some - but certainly not all - so-called natively disordered proteins will gain structure in cells. In terms of the equilibrium thermodynamics of protein stability (Equations 1 and 2), these proteins are best considered as simply very unstable. This instability may be essential for function, for example allowing rapid degradation or facilitating exit from the cell. And, finally, it is important to consider this discussion in terms of Anfinsen's thermodynamic hypothesis. Specifically the last four words of his statement[5]: "the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, *in a given environment*."

4. Functional Repertoire

i. Molecular Recognition

The coupling of folding and binding

Molecular recognition is an essential requirement for life. Protein/protein, protein/nucleic acid, and protein ligand interactions initiate and regulate most cellular processes.

Many natively disordered proteins can fold upon binding to other proteins or DNA. The loss of conformational entropy that occurs during folding can influence the kinetics and thermodynamics of binding[78,267,303,304,305]. The most appealing model for the coupling of folding and binding proposes an initial encounter complex that forms nonspecifically while the protein is still unfolded[303]. The release of solvent will provide a favorable entropic contribution to the binding and be dependent on the amount of hydrophobic burial that occurs during the formation of the nonspecific encounter complex. This encounter complex will undergo a sequential selection to achieve a specific complex. Sequential selection is achieved by consecutive structural interconversions that increase the surface complimentarity. A subsequent increase in surface complimentarity leads to a more stable complex. Of course, this model does not address the very real possibility that folding to a single structure does not occur and instead the bound structure is dynamic and the overall stability is governed by a collection of competing interactions.

Structural plasticity for the purpose of functional plasticity

There is increasing evidence that multiple binding modes can be accommodated in protein/ligand, protein/protein, and protein/DNA interactions[306,307,308,309,310,311]. In some cases, different segments of a single polypeptide can be used for recognition of different substrates. In other cases a single protein surface can accommodate the coupled binding and folding of multiple polypeptide sequences into different structures. For instance, the phosphotyrosine-binding domain of the cell fate determinant Numb can recognize peptides that differ in both primary and secondary structure by engaging various amounts of the binding surface[308].

The importance of structural plasticity in molecular recognition is made more concrete by considering a specific example: the activation of calcineurin (CaN) by calmodulin (CaM). The calcium-dependent binding of CaM to CaN brings about exposure of CaN's serine-threonine phosphatase active site by displacement of the autoinhibitory peptide and thereby turns on the phosphatase activity (Figure 1). This interaction forms a bridge between phosphorylation-dephosphorylation-based signaling and calcium-based signaling. The CaN-CaM interaction plays important roles in a wide variety of eukaryotic cells. For example, dephosphorylation of Nfat by Ca^{2+}-CaM-activated CaN leads to killer T-cell proliferation and foreign tissue rejection[312]: blockage of CaN activation by complexation with FK506-FK binding protein leads to suppression of the rejection[313].

The intrinsic disorder (plasticity) of CaM binding domains enables them to bind to a wide variety of target sequences[314,315]. The four EF-hands of CaM undergo disorder-to-order transitions upon Ca^{2+} binding[316]. The two domains of CaM are connected by a helix in the crystal, but NMR shows the central region to be melted in solution, thus providing a flexible hinge that enables CaM to wrap around its target helix[317]. On the CaN side, intrinsic disorder flanks the CaM target and thus provides space for CaM to wrap around its target helix[235]. Indeed, before the CaN disordered region was revealed as missing electron density in its X-ray crystal structure, trypsin digestion had already indicated disorder in CaN's CaM binding region[318,319]. Similar trypsin digestion analyses show that many CaM binding sites are within disordered regions.

Systems where disorder increases upon binding

Two studies have shown that disorder can increase upon binding to hydrophobic ligands and proteins[320,321]. Increased protein backbone conformational entropy was observed for the mouse major urinary protein (MUP-I) upon binding the hydrophobic mouse pheromone 2-*sec*-butyl-4,5-dihydrothiazole[320]. ^{15}N relaxation measurements of free and pheromone bound MUP-I were fitted using the Lipari-Szabo model free approach[86,87]. Order parameter differences were observed between free and bound MUP-I that were consistent with an increase in the conformational entropy. Out of 162 MUP-I residues, 68 showed significant reductions in S^2 upon pheromone binding. The changes were distributed throughout the β-barrel structure of MUP-I with a particular prevalence in a helical turn proposed to form a ligand entry gate at one end of the binding cavity. The authors called into question assumptions that the protein backbone becomes more restricted upon ligand binding.

An increase in side chain entropy facilitated effector binding for the signal transduction protein, Cdc42Hs[321]. Cdc42Hs is a member of the Ras superfamily of GTP-binding proteins that displays a wide range of side chain flexibility. Methyl axis order parameters ranged from 0.3 ± 0.1 (highly disordered) in regions near the effector binding site to 0.9 ± 0.1 in some helices. Upon effector binding, the majority of methyl groups showed a significant reduction in their order parameters, indicating increased entropy. Many of the methyl groups that showed increased disorder were not part of the effector binding interface. The authors propose that increased methyl dynamics balance entropy losses as the largely unstructured effector peptide folds into an ordered structure upon binding.

Actually, there is an entire class of proteins, for which ligand binding may be accompanied by destabilization of the native state[306,307]. For example, the introduction of a Ca^{2+}-binding amino-acid sequence did not affect the structure or stability of the T4 lysozyme in the absence of calcium. However, in the presence of this cation the stability of the mutant protein was detectably less than that of wild-type T4 lysozyme. This instability suggests that the binding of Ca^{2+} might be accompanied by considerable conformational changes in the modified loop that lead to destabilization of the protein[322]. Similar effects have been described for the calcium binding N-domain fragments of Paramecium calmodulin[323], rat calmodulin[324] and isolated domains of troponin C[325].

Similarly, the tertiary structure of calreticulin, a 46.8 kDa chaperone involved in the conformational maturation of glycoproteins in the lumen and endoplasmic reticulum, was distorted by Zn^{2+} binding, that resulted in a concomitant decrease in the conformational stability of this protein[326]. The zinc-induced structural perturbations and destabilization described above are characteristic of several other calcium binding proteins, including α-lactalbumin, parvalbumin and recoverin[307].

Of special note are some biomedical implications of the observed destabilization upon binding. For example, it has been shown that β2-microglobulin is able to bind Cu^{2+}. The binding is accompanied by a significant destabilization of the protein, suggesting that the ion has a higher affinity for the unfolded form[327]. β2-microglobulin is a 12 kDa polypeptide that is necessary for the cell surface expression of the class I major histocompatibility complex (MHC). Turnover of MHC results in release of soluble β2-microglobulin followed by its catabolism in the kidney. In patients suffering from kidney disease treated by dialysis, β2-microglobulin forms amyloid deposits principally in the joints, resulting in a variety of arthropathies. Importantly, the zinc-induced destabilization of the β2-microglobulin native structure was implicated as the driving force of this amyloidosis[327].

ii. Assembly/Disassembly

Protein crystallographers are often frustrated in attempts to ply their trade because disordered N- and C- termini prevent crystallization. The idea that such sequences have been retained by nature to frustrate crystallographers, although enticing, is probably invalid since the termini have been around longer than crystallographers. Instead, it has been shown that disordered termini are often conserved to facilitate assembly and disassembly of complex objects, like viruses.

The main idea of disorder-assisted assembly is that sequences from several different proteins are required for the disordered termini to fold into a defined structure that stabilizes the assembly. In some instances, only one player is disordered, in others disordered chains must interact to gain structure. Disorder-assisted assembly accomplishes two beneficial goals. First, it ensures assembly only when all the players are in their correct positions. Second, it prevents aggregation of individual components. For example, the N-terminal 20 residues of porcine muscle lactate dehydrogenase forms a disordered tail in the monomer that is essential for tetramerization[328]. Namba has reviewed several other of these systems, including examples from the assembly of tobacco mosaic virus, bacterial flagella, and icosahedral viruses, and DNA/RNA complexes[329].

iii. Highly-Entropic Chains

These are the elite of protein disorder—proteins, whole or in part, that function only when disordered. Terms such as "entropic bristles", "entropic springs", and "entropic clocks" have been used to describe these systems, but these can be misleading because all matter, except perfect crystals at 0 Kelvin, has entropy.

Hoh set down the main concept, in his contribution about entropic bristle domains[330]. Disordered regions functioning as entropic bristles within a binding site will block binding until the bristle is modified. The modification causes the disordered region to move to one side of the binding site. Members of this class can be recognized by the effects of deletion. That is, removal of the bristle should lead to permanent activation[330]. The C-terminal region of p53 has been shown to function as an entropic bristle domain of this type[330,331]. Bristles can also act as springs when two sets of disordered proteins are brought into contact with each other. The interactions of neurofilaments may be controlled in this way[332,333,334].

An extended unfolded region is important for the timed inactivation of some voltage-gated potassium channels[335]. The extended disordered region functions as one component of an entropic clock. Charge migrations within the tetrameric pore proteins are associated with the majority of state changes of voltage-gated K$^+$ ion channels[336]. However, the timing of the inactivation step is determined by the time it takes for a mobile domain to find and block the channel. The movement of the mobile domain is restricted by a tether composed of ~ 60 disordered residues (Figure 9). The timing of channel inactivation is a function of the length of the disordered tether[337]. Since ion channels serve to modulate the excitability of nerve cells, their malfunction can have substantial impact on human health.

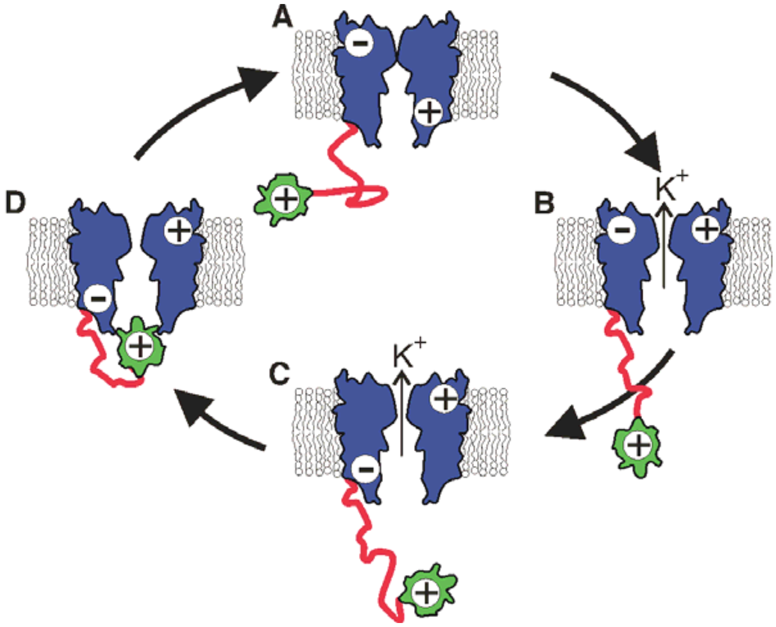


Figure 9. Example of an entropic clock. Simplified model of a Shaker-type voltage-gated K$^+$ ion channel (blue) with ‘ball and chain’ timing mechanism. The ‘ball and chain’ is comprised of an inactivation, or ball, domain (green) that is tethered to the pore assembly by a disordered chain (red) of ~ 60 residues. For simplicity, only four of the proposed ten states are shown [336]. The cytoplasmic side of the assembly is oriented downward. **(a)** Closed state prior to membrane depolarization. Note that conformational changes in the pore have sealed the channel and a positive charge on the cytoplasmic side of the pore assembly excludes binding of the ball domain. **(b)** Open state following membrane depolarization. **(c)** After depolarization, the cytoplasmic side of the pore opening assumes a negative charge that facilitates interaction with the positively charged ball domain. **(d)** Inactivation of the channel occurs when the ball domain occludes the pore. The transition from (c) to (d) does not involve charge migration and can be modeled as a random walk of the ball domain towards the pore opening. (Portions of figure based on Antz et al.[396]).

One further example of entropic disordered region function is length adjustment within the muscle protein[338]. Please note, however, that for each of the three examples described above, there are counter examples from related systems or proteins where either the region is absent or is replaced by a globular domain.

iv. Protein Modification

As discussed above, protease digestion occurs preferentially in unfolded regions of proteins. The need to protect backbone hydrogen bonds in folded structure[224] and the need for extensive contacts with the backbone residues to bring about hydrolysis[222] are mutually exclusive. This can explain the observations that there is a very strong preference for protease sensitive regions to be located in disordered regions[339,340]. It is much less clear whether protein modification involving side chains would occur preferentially in ordered or disordered regions.

In a study of the functions associated with more than 100 long disordered regions, many were found to contain sites of protein modification[30,31]. These modifications included phosphorylation, acetylation, fatty acylation, methylation, glycosylation, ubiquitination, and ADP-ribosylation. These observations suggest the possibility that protein modifications commonly occur in regions of disorder.

Phosphorylation by the kinases and dephosphorylation by the phosphatases provides an extremely important signaling system for eukaryotic cells, with an estimate that up to one-third of eukaryotic proteins are phosphorylated[341]. As mentioned above, many sites of protein phosphorylation were found to be in regions structurally characterized as natively disordered[30,31]. Thus, further study of the relationship between phosphorylation and disorder seemed appropriate.

Several lines of evidence support the view that protein phosphorylation in eukaryotic cells occurs primarily in regions of disorder. These include: 1) despite the very high interest in phosphorylation, very few structures in PDB exist for both the unphosphorylated and phosphorylated forms of the same protein[342,343] (a possible explanation is that the prevalence of disorder in proteins that become phosphorylated tends to inhibit their crystallization[343,344]); 2) nine structures of eukaryotic kinase substrates in their unphosphorylated forms show that the residues of the phosphorylation site have extended, irregular conformation that are consistent with disordered structure[343]; 3) the structures of substrate or inhibitor polypeptides indicate that the residues corresponding to the sites of phosphorylation are within segments that lack secondary structure and that are held in place not only by side chain burial but also by backbone hydrogen bonds to the surrounding kinase side chains[345,346,347,348,349,350] (just as for protease digestion, this is a strong indicator that the substrate must be locally unfolded before binding to its enzyme partner); 4) in a database of more than 1,500 well characterized sites of phosphorylation and a larger number of sites that are not phosphorylated, the residues flanking the sites of phosphorylation are substantially and systematically enriched in the same amino acids that promote protein disorder and are depleted in the amino acids that promote protein order[343]; and 5) the sequence complexity distribution of the residues flanking phosphorylation sites matches almost exactly the complexity distribution obtained for a collection of experimentally characterized regions of disorder while the complexity distribution of the residues flanking nonphosphorylation sites matches almost exactly the complexity distribution obtained for a collection of ordered proteins. In addition there is a high correspondence between the prediction of disorder and the occurrence of phosphorylation and, conversely, the prediction of order and the lack of phosphorylation (unpublished observations) - this is expected from the amino acid compositions of the residues flanking of phosphorylation and nonphosphorylation sites. A new predictor of phosphorylation exhibited small, but significant improvement if predictions of order and disorder were added[343]. These data and observations support the suggestion that sites of protein phosphorylation occur preferentially in regions of native disorder.

Data support the suggestion that protease digestion and phosphorylation both occur preferentially within regions of disorder. Also, several other types of protein modification, such as acetylation, fatty acid acylation, ubiquitination, and methylation, have also been observed to occur in regions of intrinsic disorder[30,31]. From these findings, it is tempting to suggest that sites of protein modification in eukaryotic cells universally exhibit a preference for natively disordered regions.

What might be the basis of a preference for locating sites of modification within regions of native disorder? For all of the examples discussed above, the modifying enzyme has to bind to and modify similar sites in a wide variety of proteins. If all the regions flanking these sites are disordered before binding to the modifying enzyme, it is easy to understand how a single enzyme could bind to and modify a wide variety of protein targets. If instead, all these regions bind as ordered structures, then there is the complicating feature that the proteins being regulated by the modification must all adopt the same local structure at the site of modification. This imposes significant constraints on the site of modification. The structural simplification that arises from locating the sites of modification within regions of disorder is herein proposed to be an important principle. Elsewhere we point out that a particular advantage of disorder for regulatory and signaling regions is that changes, such as protein modification, lead to large-scale disorder to order structural transitions: such large-scale structural changes are not subtle and so could be an advantage for signaling and regulation as compared to the much smaller changes that would be expected from the decoration of an ordered protein structure.

5. Importance of Disorder for Protein Folding

Interest in the unfolded protein state has increased markedly in recent years. A major motivation has been to better understand the structural transitions that occur as a protein acquires 3-D structure, both from the point of view of the mechanism of folding and from the point of view of the energetics. A major effort has been to connect structural models of the unfolded polypeptide chain with experimental data supporting the given model. The overall sizes of guanidinium chloride-denatured proteins fit the values expected for random coils when excluded volume effects were taken into account[351]. For a true random coil, the Φ and Ψ angles of a given dipeptide are independent of the angles of the dipeptides before and after. This is often called the Flory isolated pair hypothesis[352]. Both experiments and calculations call the isolated pair hypothesis into question[353]. For example, using primarily repulsive terms, Pappu et al.[354], computed the effects of steric clash (or excluded volume) on blocked polyalanines of various lengths. Contrary to Flory's isolated hypothesis, they found that excluded volume effects were sufficient to lead to preferred backbone conformations, with the polyproline II helix being among the most preferred. Representing unfolded proteins as polymer chain models, which have the advantage of simplifications that result from ignoring most atomic details, are proving useful for modeling experimental data pertaining to protein stability, folding and interactions[355].

While there has been substantial emphasis on the study of unfolded proteins as precursors to 3-D structures as briefly described above, to our knowledge there has been no systematic discussion or studies of how natively disordered regions affect protein 3-D structure and folding kinetics. Here we will consider several possibilities, each of which are simple hypotheses that have not yet been tested by experiment. These hypotheses form the basis for experiments into the effects of native disorder on protein structure and folding.

First a well known example with ample experimental support is provided. Trypsinogen folds into a stable 3-D structure, but compared to trypsin, the folding is incomplete and the protein is inactive. Trypsinogen's folding does involve the formation of the crucial catalytic triad, and yet trypsinogen remains inactive evidently because the binding pocket for the substrate lysine or arginine fails to form completely; e.g. remains natively disordered[12,356]. Thus, trypsinogen, the precursor to trypsin, remains inactive and so does not harm the interior of the cell. The folding into active trypsin is inhibited by a short region of native disorder at the amino terminus; this region is completely missing in the electron density map of trypsinogen[357]. Once trypsinogen has been exported from the cell, this disordered region is cleaved off by trypsin. The amino terminus changes from a highly charged moiety into a hydrophobic terminus of isoluecine followed by valine (IV). In the absence of the natively disordered, charged extension, this IV terminus becomes capable of binding into a particular site elsewhere in the structure. This binding of the IV moieties brings about a disorder to order transition of trypsin's binding pocket, now enabling the protein to bind arginine or lysine and thereby converting inactive trypsinogen into active trypsin[358]. Even in the absence of the cleavage, high levels of an IV dipeptide can stimulate the protease activity of uncleaved trypsinogen by binding into the site used by the IV terminus[12]. We speculate that proteolytic removal of a natively disordered region may be a common mechanism for regulating protein folding and function, but we have not yet searched systematically for other examples.

While trypsinogen provides an example of a protein for which a region of native disorder inhibits protein folding, we speculate that there are proteins for which regions of native disorder promote protein folding. Suppose, for example, a protein contained a short, highly charged region of native disorder and a folded domain with an opposite charge of significant magnitude. From various analyses suggesting that a high net charge can lead to a natively unfolded protein[123,227,359], the region of oppositely charged disorder might be essential for overall charge balance and, if so, would be required for protein folding. A simple experiment is proposed here: 1) identify example proteins in the PDB with short, highly charged tails and oppositely charged, folded domains; 2) determine which domains are likely to not fold without their oppositely charged tail by means of the net charge versus hydropathy plot[227]; 3) remove the charged tail by proteolysis or genetic engineering; 4) compare the folding rate of the shorter protein with that of the full-length protein. The prediction is that, if located in the unfolded region of the net charge hydropathy plot, the shorter protein would fail to fold or would require higher ionic strength to fold as compared to the full-length protein.

Various theoretical analyses on protein folding have suggested relationships among the size, stability and topology of a protein fold and the rate and mechanism by which the fold is achieved. The characterization of the folding of a large number of simple, single domain proteins enabled detailed studies to test the various models and assertions regarding the mechanism of protein folding. The simple proteins in this set are characterized by having single domains, by the lack of a prosthetic groups, disulfide bonds, and cis proline residues, and by two-state refolding kinetics. Despite the simplicity of this set, more than a million-fold variation in refolding rates was observed. A remarkable finding was that the statistically significant correlations among the folding rate, the transition state placement and the relative contact order are observed, where the transition state placement is an estimate of the fraction of the burial of the hydrophobic surface area in the transition state as compared to the native state, and where the relative contact order is the length-normalized sequence separation between contacting residues in the native state[360]. A number of alternative empirical measures of topology were later shown to correlate about equally well with folding rates as does the relative contact order, including the number of sequence-distant contacts per residue[361], the fraction of contacts that are distant in the sequence[362], and the total contact distance[363]. The quality of these alternative measures supports the conjecture that contact order predicts rates, not because it is directly related to the mechanism of folding, but rather because it is related to an alternative physical parameter of importance[364].

Attempts to reconcile the observed relationship between contact order or other measures of topology and folding rates has led to the topomer search model[365], which was based substantially on similar prior models[366,367]. Simply put, the topomer search model stipulates that the rate limiting step in protein folding is the search for an unfolded conformation with the correct overall topology. The unfolded form with the correct topology then rapidly folds into the native state[364].

Assuming for the sake of discussion the basic correctness of the topomer search model, the expected effects of native disorder can be readily described. First let us consider natively disordered regions at the amino or carboxy terminus that do not affect the overall protein stability and that don't stabilize misfolded intermediates. Such natively disordered regions would be expected to have very small effects on the folding rates. On the other hand, natively disordered internal loops would be expected to slow the rate of folding in a length-dependent manner. Indeed, a systematic study of the folding rates of simple, two-state proteins with natively disordered loops of varying lengths could provide a useful test of the topomer search model.

The relationship between the log of the folding rates of simple two-state proteins and the contact order value exhibits an r^2 value of about 0.8, suggesting that this measure captures in excess of 3/4 of the variance in the logarithms of the reported folding rates[364]. The topomer search model, which evidently captures the dependence of protein folding on the contact order, ignores variation in foldability along the sequence. On the other hand, predictors of order and disorder capture local sequence tendencies for order or disorder. Such a local tendency for disorder could be over-ridden by non-local interactions in the final native structure, and so a local tendency for disorder could be important for a folded protein even if the native structure does not exhibit actual disorder. We wonder whether there is any relationship between protein folding rates and variations in order/disorder tendencies along an amino acid sequence. That is, we wonder whether the 1/4 variability in the logarithms of the folding rate that is not captured by relative contact order could be related to differences in the amounts or in the organization of disorder tendencies along the amino acid sequences. Local regions with high tendencies for disorder could, for example, lead to very non-uniform polymer chain models by local alterations in the stiffness or persistence length; locating such anomalies at topologically critical sites could greatly speed-up or slow down protein folding rates. These ideas could be tested both computationally and experimentally.

6. Protocols

In this section we include several protocols for NMR spectroscopy, X-ray crystallography and circular dichroism spectropolarimetry. We focused on these methods because they are traditionally the most used to characterize native disorder in proteins. Protocols for small angle X-ray diffraction, hydrodynamic measurement, fluorescence methods, conformational stability, mass spectrometry-based high resolution hydrogen-deuterium exchange, protease sensitivity, and prediction from sequence are available from the references cited in the above methods section.

i. NMR Spectroscopy

General requirements

The following methods assume the experimenter has access to a high field digital nuclear magnetic resonance spectrometer and a soluble, homogeneous protein sample at a concentration of ~1 mM. These methods also assume that the incorporation of a NMR-active isotope such as ^{15}N and/or ^{13}C is possible.

Measuring transient secondary structure in secondary chemical shifts

Resonance assignments are the first step in the analysis of protein structure using NMR. A convenient outcome of these measurements is the determination of protein secondary structure. The C_α, C_β, and C' chemical shifts are the most sensitive to phi and psi angles. To measure these chemical shifts, sensitivity enhanced HNCACB and HNCO experiments can be performed on uniformly ^{15}N and ^{13}C labeled protein samples[368,369,370]. An appropriate digital resolution for the HNCACB and HNCO experiments is 9.8 Hz/pt in ^{1}H with 512 complex points, 47.1 Hz/pt in ^{13}C with 128 complex points, and, 34.4 Hz/pt in ^{15}N with 32 complex points. After transformation of the data, it is essential to apply the appropriate referencing before comparing the protein chemical shifts to random coil standards[67].

Measuring the translational diffusion coefficient using pulsed field gradient diffusion experiments

Pulsed field gradient diffusion measurements can be made using a variation of the water-sLED sequence developed by Altieri et al. 1995[69]. Collect data at an appropriate field strength, temperature, sweep width, gradient pulse width, and delay time. The amplitude of the gradient pulses should be varied from ~1-31 gauss/cm in increments of 2 gauss/cm. However, this range will vary depending on the protein and typically must be determined empirically. It is also possible to perform the experiment by varying the length of the gradient pulses and holding the magnitude constant[371]. The NMR data must be processed so that resonance intensity measurements can be made. An entire region of the spectrum can be integrated or individual resolved resonances can be measured. Resonance assignments are not necessary, the experiment can be performed on an unlabeled sample, and it is probably best to integrate the resonances of nonlabile nuclei. Resonance intensity measurements should be normalized, averaged and fit to the function relating the normalized resonance intensity, A, to the translational diffusion coefficient, D:

$$A = \exp(-Dg^2\delta^2\gamma^2(\Delta - \delta/3)) \qquad (26)$$

where g and δ are respectively the magnitude and duration of the gradient pulses, Δ is the time between gradient pulses, and γ is the gyromagnetic ratio of the observed nucleus[372]. Nonlinear least-squares regression of a decaying exponential function onto the data can be used to extract D. It is useful to combine these measurements with another technique like sedimentation to obtain information about the shape of the protein.

Relaxation experiments

For ^{15}N relaxation experiments the spin-lattice relaxation rates (R_1), spin-spin relaxation rates (R_2), and ^1H-^{15}N NOE's can be measured by inverse-detected two-dimensional NMR methods[81]. Spin-lattice relaxation rates are typically determined by collecting 8-12 two dimensional spectra using relaxation delays from 10-1500 milliseconds. Spin-spin relaxation rates can determined by collecting 8-12 two dimensional spectra using spin-echo delays of 10-300 milliseconds. Peak heights from each series of relaxation experiments are then fitted to a single decaying exponential. To measure the ^{1}H-^{15}N NOE's, one spectrum is acquired with a 3 second mixing time for the NOE to buildup and another spectrum is acquired with a 3 second recycle delay for a reference. It is often necessary to optimize this delay for intrinsically disordered proteins and longer delays will attenuate larger than predicted negative NHNOE values for highly dynamic regions of the protein[85]. For all experiments, water suppression can be achieved by using pulsed field gradients. Uncertainties in measured peak heights are usually estimated from baseline noise level and with good tuning and shimming are typically less than 1% of the peak heights from the first R_1 and R_2 delay points.

Relaxation data analysis using reduced spectral density mapping

Relaxation data were analyzed using the reduced spectral density mapping approach[80,88,373,374]. The ^{15}N chemical shift anisotropy and the dipolar coupling between the amide ^{15}N nucleus and the attached proton have the greatest influence on the ^{15}N nuclear relaxation[89]:

$$R_1 = \left(\frac{d^2}{4}\right)\left[3J(\omega_N) + 6J(\omega_H + \omega_N) + J(\omega_H - \omega_N)\right] + c^2 J(\omega_N) \qquad (27)$$

$$R_2 = \left(\frac{d^2}{8}\right)\left[4J(0) + 3J(\omega_N) + 3J(\omega_H + \omega_N) + 6J(\omega_H) + J(\omega_H - \omega_N)\right] + \left(\frac{c^2}{6}\right)\left[J(0) + 6J(\omega_N)\right] + R_{ex} \qquad (28)$$

$$NOE = 1 + \left(\frac{d^2}{4R_1}\right)\left(\frac{\gamma_H}{\gamma_N}\right)\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \qquad (29)$$

where $d = \left(\frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2}\right)\left\langle r_{NH}^{-3}\right\rangle$ and $c = \frac{\omega_N \Delta\sigma}{\sqrt{3}}$, μ_0 is the permeability of free space, *h* is Planck's constant, γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N respectively, r_{NH} is the amide bond length

(1.02 Å), $\Delta\sigma$ is the chemical shift anisotropy (-160 ppm) and R_{ex} is the chemical exchange contribution to R_2. J(ω) is the power spectral density function defining the reorientation of the N-H bond vector by stochastic (global) and intramolecular motions. Reduced spectral density mapping uses an average value of $J(\omega_H)$ for the linear combinations of $J(\omega_H+\omega_N)$, $J(\omega_H)$, and $J(\omega_H-\omega_N)$ leading to values of J(0), J(ω$_N$), and $J(\omega_H)$ that are given by:

$$\sigma_{NH} = R_1(NOE-1)\frac{\gamma_N}{\gamma_H} \tag{30}$$

$$J(\omega_H) = \frac{4\sigma_{NH}}{5d^2} \tag{31}$$

$$J(\omega_N) = \frac{[4R_1 - 5\sigma_{NH}]}{[3d^2 + 4c^2]} \tag{32}$$

$$J(0) = \frac{[6R_2 - 3R_1 - 2.72\sigma_{NH}]}{[3d^2 + 4c^2]} \tag{33}$$

This approach estimates the magnitude of the spectral density function at the given frequencies, making no assumptions about the form of the spectral density function or about the molecular behavior giving rise to the relaxation.

In-cell NMR

Although a potentially powerful technique, there are few published studies involving in-cell NMR[102,375 299,300,301]. Therefore, the suggestions here will have to be quite general. Most of the following suggestions are distilled from publications by Volker Dötsch's group, and a few are from the Pielak laboratory.

Most importantly, protein expression is, at present, more of an art than a science; the suitability of each protein expression system (i.e., inducer concentration, induction time, cell density at induction) must be determined empirically. Although yeast and insect cells have been tried, the technique currently works best for proteins expressed in *E. coli*. Over expression of the protein to be studied in ^{15}N- or ^{13}C- enriched media is essential. Over expression is operationally defined as the ability to isolate greater than 10 mg of pure protein from 1 L of saturated culture, using the same media and conditions that will be used for the in-cell NMR experiment. It is best to check expression using unenriched media first. The HSQC experiment works well for in-cell NMR. It is important to know what the "background" spectrum looks like when the experiment fails. Background spectra contain crosspeaks from metabolites that become enriched in ^{15}N. That is, collect a spectrum without protein overexpression (i.e., untransformed cells). It can often take as long as 15 h to obtain a spectrum with a conventional probe. Using a cryoprobe can dramatically decrease this time. When the experiment is over, it is important to perform dilution and plating experiments to show the cells are still alive and the protein of interest has remained inside the cells. Gentle centrifugation of the NMR sample followed by SDS-PAGE analysis of both the supernatant and the pellet is the best way to show that the protein was overexpressed and did not leave the cells.

ii. X-ray Crystallography

Determination of protein structure by X-ray crystallography has become almost routine, once crystals are obtained. Many excellent books and papers provide the details of protein structure determination by X-ray crystallography. Two books that emphasize practical experimental approaches are by Stout and Jensen[376] and by McRee[377] In this protocol section, we will concentrate on how to obtain crystals in the face of native disorder.

The presence of large regions of disorder can block attempts to crystallize proteins. Failure to obtain crystals is the single greatest experimental problem in X-ray crystallography.

Here, we present calsequestrin as a case study illustrating the successful crystallization of a natively disordered protein.

Calsequestrin is a calcium storage protein located within the sarcoplasmic reticulum that binds 40-50 calcium ions with ~ 1mM affinity. Calsequestrin is a highly acidic protein with many of the acidic residues located in clusters. The physiochemical properties of purified calsequestrin as revealed by tryptophan fluorescence[378,379,380,381], circular dichroism[378,380,381,382,383], Raman spectroscopy[380,384], NMR[380] and proteolytic digestion[381,384,385] indicated that calsequestrin was mostly unfolded at low ionic strength. As ionic strength was increased, folding into a compact structure was observed. This structure can be induced by calcium as well as other ions such as Na^+, Zn^{2+}, Sr^{2+}, Tb^{2+}, K^+, and H^+. The high Ca^{2+} binding capacity of calsequestrin was believed to require the formation of aggregates, thus transitioning from a soluble disordered to a solid crystalline form[381,386,387]. However, crystallization attempts in the presence of Ca^{2+} resulted in needle-like crystals. These liner polymeric forms were also observed in vivo suggesting that this was the physiologically relevant form of the complex[388]. These narrow crystals were unsuitable for structure determination but did demonstrate that crystallization of calsequestrin was possible. The fact that calsequestrin adopted structure in the presence of mono- and di- valent cations other than calcium and that these forms were not observed to aggregate and precipitate as the needle-like crystals did suggested that alternate crystal forms could be possible.

The following hypothesis was formed: the growth of needle-like crystals of calsequestrin is a two step process[389]. The first step was the induction of structure by high ionic strength. The second was the calcium-specific cross-bridging of individual calsequestrin molecules by means of the unneutralized charges remaining after initial non-specific binding of ions by the monomers. This cross-bridging could account for growth in a single direction, thereby producing the needles. To test this hypothesis, crystallization in the absence of Ca^{2+} was attempted[389,390].

Calsequestrin was purified from the skeletal muscle of New Zealand white rabbits as previously described[381]. Approximately 500 initial crystallization experiments were conducted using the hanging drop vapor diffusion method[391] with an incomplete factorial approach[392] to cover a wide range of conditions. Each condition was tested in a volume of 1µl containing 5-10 µg of calsequestrin. Conditions with high monovalent cation concentrations were emphasized along with a range of 16 different precipitation reagents. Good-sized non-needle crystals were obtained when 2-methyl-2,4 pentane diol (MPD) was used as the precipitating reagent. The best crystals were grown from a solution containing 10% (v/v) MPD, 0.1 M sodium citrate, 0.05 M sodium cacodylate, pH 6.5 and 5 mg/ml of calsequestrin. The nominal intitial Na^{+} concentration was 0.35 M. The rectangular crystals formed within one week and grew to 0.2 x 0.2 x 0.8 mm by the second week. Structure determination details can be found elsewhere[389,390].

Surprisingly, the resulting structure of calsequestrin exhibited three identical domains, each with a thioredoxin protein fold. The three domains interact to form a disk-like shape with an approximate radius of 32 Å and a thickness of 35 Å. No clues to this three domain structure were obtained from sequence analysis[393,394]. Rather, analysis pointed towards a globular N-terminal domain and a C-terminal disordered region[381]. Even in hindsight, no significant similarities among the three similar domains could be deduced from the sequence.

A common approach to disordered regions in proteins is to remove the coding regions from the recombinant expression constructs so that these regions do not prevent the protein from crystallizing[344]. In the case of calsequestrin, the fact that structure could be induced in the disordered protein by increasing cation concentration lead to attempts at crystallization under non-intuitive conditions. The end result was the elucidation of the an interesting and important structure.

iii. Circular Dichroism Spectropolarimetry

There exists an excellent book on protein circular dichroism that contains several sections on the collection and interpretation of spectra[137]. As discussed above, the interpretation of spectra from disordered proteins remains controversial. However, the two most important experimental concepts are well agreed upon.

First, collect data as far into the ultraviolet as possible. The far UV region, from 260 nm to 190 nm contains a great deal of information about secondary structure. With a powerful UV source, which often means a new lamp, and a well-behaved sample, data can be obtained down to wavelengths below 190 nm. Second, only the electronic transitions of the protein being studied should contribute to the absorbance. Any extrinsic absorbance degrades the instrument's ability to detect the small differences between left and right circularly polarized light absorbed by the protein under study.

The two most common sources of unwanted absorbance are light scattering and buffer/co-solute absorbance. The protein must not aggregate to such an extent that the sample scatters light. It is important to pass every sample through a 0.2-micron filter prior to data acquisition. A common mistake is to use a buffer that absorbs in the ultraviolet region. Histidine is a common buffer because it is used to elute his-tagged proteins from Ni^{2+}-affinity columns, but at the concentrations used for elution, the histidine contributes to an excessive amount of ultraviolet absorbance. Tris is also a poor choice. Phosphate and acetate are more useful buffers, at least in terms of absorbance. Absorbance can also be a problem when collecting spectra at high co-solute (*e.g.*, sugars, urea, etc.) concentrations.

The best way to proceed is to collect a spectrum of filtered water or buffer and then compare this spectrum to that of the solution. After choosing appropriate solution conditions, there is no better protocol for preparing the sample than thorough dialysis followed by filtration. The reference solution is then made from a filtered sample of the solution from outside the dialysis bag.

CD analysis is especially sensitive to error due to misestimation of protein concentration. It is important that extreme care be taken during concentration determination. Additionally, we prefer the method of Gill and von Hippel[395] rather than any of the popular colorimetric assays. This method is based on the calculation of a molar extinction coefficient based on the amino acid content of the protein under study and is typically accurate to ± 5%.

Acknowledgements

AKD and MSC are supported by NIH grant R01 LM 007688 (Indiana University School of Medicine). AKD is additionally supported by NIH grants R43 CA099053, R43 CA 097629, R43 CA 097568 and R43 GM 066412 (Molecular Kinetics, Inc.). GJP thanks the NSF for support and Trevor Creamer for helpful discussions. GWD is supported by NIH grant P20 RR 16448 from the COBRE Program of the National Center for Research Resources.

References

1. Fisher, E. (1894). Einfluss der configuration auf die wirkung den enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2985-2993.

2. Anson, M. L. & Mirsky, A. E. (1925). On some general properties of proteins. *J. Gen. Physiol.* **9**, 169-179.

3. Edsall, J. T. (1995). Hsien Wu and the first theory of protein denaturation (1931). *Adv. Protein Chem.* **46**, 1-5.

4. Mirsky, A. E. & Pauling, L. (1936). On the structure of native, denatured and coagulated proteins. *Proc. Natl. Acad. Sci. U. S. A.* **22**, 439-447.

5. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

6. Gutte, B. & Merrifield, R. B. (1969). The total synthesis of an enzyme with ribonuclease A activity. *J. Am. Chem. Soc.* **91**, 501-2.

7. Pauling, L., Corey, R. B. & Branson, R. H. (1951). The structure of proteins: two hydrogen-bonded configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37**, 205-210.

8. Kendrew, J. C., Dickerson, R. E. & Strandberg, B. E. (1960). Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **206**, 757-763.

9. Perutz, M. F., Rossmann, M. P., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature* **185**, 416-422.

10. Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Ångstrøm resolution. *Nature* **206**, 757-761.

11. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). Protein disk of tobacco mosiac virus at 2.8Å resolution showing the interactions within and between subunits. *Nature* **276**, 362-368.

12. Bode, W., Schwager, P. & Huber, R. (1978). The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine

trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9Å resolution. *J. Mol. Biol.* **118**, 99-112.

13. Williams, R. J. (1978). The conformational mobility of proteins and its functional significance. *Biochem. Soc. Trans.* **6**, 1123-1126.

14. Pullen, R. A., Jenkins, J. A., Tickle, I. J., Wood, S. P. & Blundell, T. L. (1975). The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen. *Mol. Cell. Biochem.* **8**, 5-20.

15. Cary, P. D., Moss, T. & Bradbury, E. M. (1978). High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur. J. Biochem.* **89**, 475-82.

16. Linderstrøm-Lang, K. U. & Schellman, J. A. (1959). Protein structure and enzyme activity. In *The Enzymes* (Boyer, P. D., Lardy, H. & Myrback, K., eds.), Vol. 1, pp. 443-510. Academic Press, New York.

17. Schweers, O., Schönbrunn-Hanebeck, E., Marx, A. & Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for ß-structure. *J. Biol. Chem.* **269**, 24290-24297.

18. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T., Jr. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709-13715.

19. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321-331.

20. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. & Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26-59.

21. Bailey, R. W., Dunker, A. K., Brown, C. J., Garner, E. C. & Griswold, M. D. (2001). Clusterin: a binding protein with a molten globule-like region. *Biochemistry* **40**, 11828-11840.

22. Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87-103.

23. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. & Ptitsyn, O. B. (1981). Alpha-lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett.* **136**, 311-315.

24. Tiffany, M. L. & Krimm, S. (1968). New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **6**, 1379-82.

25. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002). Is polyproline II a major backbone conformation in unfolded proteins? *Adv Protein Chem* **62**, 163-240.

26. Creamer, T. P. & Campbell, M. N. (2002). Determinants of the polyproline II helix from modeling studies. *Adv. Protein Chem.* **62**, 263-82.

27. Klee, C. B., Crouch, T. H. & Krinks, M. H. (1979). Calcineurin: a calcium- and calmodulin-binding protein of the nervous system. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 6270-6273.

28. Klee, C. B., Draetta, G. F. & Hubbard, M. J. (1988). Calcineurin. *Adv. Enzymol. Relat. Areas Mol. Biol.* **61**, 149-200.

29. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., Gastinel, L. N., Habuka, N., Chen, X., Maldanado, F., Barker, J. E., Bacquet, R. & Villafranca, J. E. (1995). Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* **378**, 641-644.

30. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradovic, Z. (2002). Intrinsic disorder and protein function. *Biochemistry* **41**, 6573-82.

31. Dunker, A. K., Brown, C. J. & Obradovic, Z. (2002). Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* **62**, 25-49.

32. Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Ber. Dt. Chem. Ges.* **27**, 2985-2993.

33. Page, M. I. (1987). Enzyme Mechanisms. In *Enzyme Mechanisms* (Page, M. I. & Williams, A., eds.), pp. 1-13. Royal Society of Chemistry, London.

34. Kraut, J. (1988). How do enzymes work? *Science* **242**, 533-40.

35. Polanyhi, M. (1921). Absorption catalysis. *Zeitshrifte fur Electrochemie* **27**, 142-150.

36. Pauling, L. (1946). Molecular architecture and biological reactions. *Chemical Engineering News* **24**, 1375-1377.

37. Borman, S. (2004). Much ado about enzyme mechanisms. *Chemical and Engineering News* **82**, 35-39.

38. Warshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* **273**, 27035-27038.

39. Bruice, T. C. & Benkovic, S. J. (2000). Chemical basis for enzyme catalysis. *Biochemistry* **39**, 6267-74.

40. Hur, S. & Bruice, T. C. (2003). Just a near attack conformer for catalysis (chorismate to prephenate rearrangements in water, antibody, enzymes, and their mutants). *J Am Chem Soc* **125**, 10540-2.

41. Schulz, G. E. (1979). Nucleotide Binding Proteins. In *Molecular Mechanism of Biological Recognition* (Balaban, M., ed.), pp. 79-94. Elsevier/North-Holland Biomedical Press, New York.

42. Karush, F. (1950). Heterogeneity of the binding sites of bovine serum albumin. *J. Am. Chem. Soc.* **72**, 2705-2713.

43. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. (1996). Structural studies of p21$^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11504-11509.

44. Dunker, A. K. & Obradovic, Z. (2001). The protein trinity - linking function and disorder. *Nat. Biotechnol.* **19**, 805-806.

45. Bracken, C. (2001). NMR spin relaxation methods for characterization of disorder and folding in proteins. *J. Mol. Graph. Model.* **19**, 3-12.

46. Dyson, H. J. & Wright, P. E. (1998). Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Struct. Biol.* **5 Suppl**, 499-503.

47. Dyson, H. J. & Wright, P. E. (2001). Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* **339**, 258-270.

48. Dyson, H. J. & Wright, P. E. (2002). Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv. Protein Chem.* **62**, 311-40.

49. Barbar, E. (1999). NMR characterization of partially folded and unfolded conformational ensembles of proteins. *Biopolymers* **51**, 191-207.

50. Yao, J., Chung, J., Eliezer, D., Wright, P. E. & Dyson, H. J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* **40**, 3561-71.

51. Eliezer, D., Yao, J., Dyson, H. J. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5**, 148-155.

52. Arcus, V. L., Vuilleumier, S., Freund, S. M., Bycroft, M. & Fersht, A. R. (1994). Toward solving the folding pathway of barnase: the complete backbone 13C, 15N, and 1H NMR assignments of its pH-denatured state. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9412-6.

53. Arcus, V. L., Vuilleumier, S., Freund, S. M., Bycroft, M. & Fersht, A. R. (1995). A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J. Mol. Biol.* **254**, 305-21.

54. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575-578.

55. Dobson, C. & Hore, P. (1998). Kinetic studies of protein folding using NMR spectroscopy. *Nat. Struct. Biol.* **5**, 504-507.

56. Shortle, D. R. (1996). Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* **6**, 24-30.

57. Alexandrescu, A. T. & Shortle, D. (1994). Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J. Mol. Biol.* **242**, 527-46.

58. Alexandrescu, A. T., Abeygunawardana, C. & Shortle, D. (1994). Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. *Biochemistry* **33**, 1063-1072.

59. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**, 170-84.

60. Gillespie, J. R. & Shortle, D. (1997a). Characterization of long-range structure in the denatured state of *staphylococcal* nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158-169.

61. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**, 487-9.

62. Sinclair, J. F. & Shortle, D. (1999). Analysis of long-range interactions in a model denatured state of staphylococcal nuclease based on correlated changes in backbone dynamics. *Protein Sci.* **8**, 991-1000.

63. Wrabl, J., Shortle, D. & Woolf, T. (2000). Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: molecular dynamics simulations of native and denatured staphylococcal nuclease. *Proteins* **38**, 123-133.

64. Wrabl, J. O. & Shortle, D. (1996). Perturbations of the denatured state ensemble: modeling their effects on protein stability and folding kinetics. *Protein Sci.* **5**, 2343-52.

65. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-51.

66. Wishart, D. S. & Sykes, B. D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4**, 171-80.

67. Wishart, D. S. & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363-92.

68. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424-31.

69. Altieri, A. S., Hinton, D. P. & Byrd, R. A. (1995). Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* **117**, 7566-67.

70. Pan, H., Barany, G. & Woodward, C. (1997). Reduced BPTI is collapsed. A pulsed field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor. *Protein. Sci.* **6**, 1985-92.

71. Davis, D. G., Perlman, M. E. & London, R. E. (1994). Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1 rho and T2 (CPMG) methods. *J. Magn. Reson. B* **104**, 266-75.

72. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1994). A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4**, 727-34.

73. Feeney, J., Batchelor, J. G., Albrand, J. P. & Roberts, G. C. K. (1979). Effects of intermediate exchange processes on the estimation of equilibrium-constants by NMR. *J. Magn. Reson.* **33**, 519-529.

74. Spera, S. & Bax, A. (1991). Empirical correlation between protein backbone conformation and C-alpha and C-beta C-13 nuclear-magnetic-resonance chemical-shifts. *J. Am. Chem. Soc.* **113**, 5490-5492.

75. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311-33.

76. Tcherkasskaya, O., Davidson, E. A. & Uversky, V. N. (2003). Biophysical constraints for protein structure prediction. *J. Proteome Res.* **2**, 37-42.

77. Daughdrill, G. W., Hanely, L. J. & Dahlquist, F. W. (1998). The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry* **37**, 1076-1082.

78. Bracken, C., Carr, P. A., Cavanagh, J. & Palmer, A. G., 3rd. (1999). Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.* **285**, 2133-2146.

79. Palmer, A. G., 3rd. (1993). Dynamic properties of proteins from NMR spectroscopy. *Curr. Opin. Biotechnol.* **4**, 385-91.

80. Lefevre, J. F., Dayie, K. T., Peng, J. W. & Wagner, G. (1996). Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry* **35**, 2674-86.

81. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-9.

82. Skelton, N. J., Palmer, A. G., Akke, M., Kordel, J., Rance, M. & Chazin, W. J. (1993). Practical aspects of 2-dimensional proton-detected N-15 spin relaxation measurements. *J. Magn. Reson. B* **102**, 253-264.

83. Kordel, J., Skelton, N. J., Akke, M., Palmer, A. G. & Chazin, W. J. (1992). Backbone dynamics of calcium-loaded calbindin D9k studied by 2-dimensional proton-detected N15 NMR spectroscopy. *Biochemistry* **31**, 4856-4866.

84. McEvoy, M. M., de la Cruz, A. F. & Dahlquist, F. W. (1997). Large modular proteins by NMR. *Nat. Struct. Biol.* **4**, 9.

85. Renner, C., Schleicher, M., Moroder, L. & Holak, T. A. (2002). Practical aspects of the 2D N-15-{H-1}-NOE experiment. *J. Biomol. NMR* **23**, 23-33.

86. Lipari, G. & Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J. Am. Chem. Soc.* **104**, 4546-4559.

87. Lipari, G. & Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **104**, 4559-4570.

88. Peng, J. W. & Wagner, G. (1992). Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. *Biochemistry* **31**, 8571-86.

89. Abragam, A. (1961). *The principles of nuclear magnetism*. The International series of monographs on physics, Clarendon Press, Oxford.

90. Bruschweiler, R., Liao, X. & Wright, P. E. (1995). Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886-9.

91. Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566.

92. Lee, L. K., Rance, M., Chazin, W. J. & Palmer, A. G. (1997). Rotational diffusion anisotropy of proteins from simultaneous analysis of N-15 and C-13(alpha) nuclear spin relaxation. *J. Biomol. NMR* **9**, 287-298.

93. Buck, M., Schwalbe, H. & Dobson, C. M. (1996). Main-chain dynamics of a partially folded protein: 15N NMR relaxation measurements of hen egg white lysozyme denatured in trifluoroethanol. *J. Mol. Biol.* **257**, 669-83.

94. Buevich, A. V., Shinde, U. P., Inouye, M. & Baum, J. (2001). Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J. Biomol. NMR* **20**, 233-249.

95. Farrow, N. A., Zhang, O., Muhandiram, R., Formankay, J. D. & Kay, L. E. (1995). A comparative-study of the backbone dynamics of the folded and unfolded forms of an SH3 domain. *J. Cell. Biochem.*, 44-44.

96. Yang, D. & Kay, L. E. (1996). Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **263**, 369-82.

97. Yang, D., Mok, Y. K., Forman-Kay, J. D., Farrow, N. A. & Kay, L. E. (1997). Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. *J. Mol. Biol.* **272**, 790-804.

98. Viles, J. H., Donne, D., Kroon, G., Prusiner, S. B., Cohen, F. E., Dyson, H. J. & Wright, P. E. (2001). Local structural plasticity of the prion protein. Analysis of NMR relaxation dynamics. *Biochemistry* **40**, 2743-53.

99. Landry, S. J., Steede, N. K. & Maskos, K. (1997). Temperature dependence of backbone dynamics in loops of human mitochondrial heat shock protein 10. *Biochemistry* **36**, 10975-86.

100. Bhattacharya, S., Falzone, C. J. & Lecomte, J. T. (1999). Backbone dynamics of apocytochrome b5 in its native, partially folded state. *Biochemistry* **38**, 2577-89.

101. Campbell, A. P., Spyracopoulos, L., Irvin, R. T. & Sykes, B. D. (2000). Backbone dynamics of a bacterially expressed peptide from the receptor binding domain of *Pseudomonas aeruginosa* pilin strain PAK from heteronuclear 1H-15N NMR spectroscopy. *J. Biomol. NMR* **17**, 239-55.

102. Dedmon, M. M., Patel, C. N., Young, G. B. & Pielak, G. J. (2002). FlgM gains structure in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12681-12684.

103. Choy, W. Y., Mulder, F. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S., Forman-Kay, J. D. & Kay, L. E. (2002). Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J. Mol. Biol.* **316**, 101-12.

104. Choy, W. Y. & Forman-Kay, J. D. (2001). Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308**, 1011-1032.

105. Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). Molecular replacement by evolutionary search. *Acta. Crystallogr. D Biol. Crystallogr.* **57**, 1474-9.

106. Huber, R. (1987). Flexibility and rigidity, requirements for the function of proteins and protein pigment complexes. Eleventh Keilin memorial lecture. *Biochem. Soc. Trans.* **15**, 1009-20.

107. Huber, R. & Bennett, W. S., Jr. (1983). Functional significance of flexibility in proteins. *Biopolymers* **22**, 261-279.

108. Douzou, P. & Petsko, G. A. (1984). Proteins at work: "stop-action" pictures at subzero temperatures. *Adv. Protein Chem.* **36**, 245-361.

109. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J. & Dunker, A. K. (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins* **53**, 566-572.

110. Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522-524.

111. Bairoch, A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45-48.

112. Schachman, H. K. (1959). *Ultracentrifugation in Biochemistry*, Academic Press, New York.

113. Glatter, O. & Kratky, O. (1982). *Small Angle X-ray Scattering*, Academic Press, London, England.

114. Doniach, S., Bascle, J., Garel, T. & Orland, H. (1995). Partially folded states of proteins: characterization by X-ray scattering. *J. Mol. Biol.* **254**, 960-7.

115. Kataoka, M. & Goto, Y. (1996). X-ray solution scattering studies of protein folding. *Fold. Des.* **1**, R107-14.

116. Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasiliev, A. M., Chernovskaya, T. V., Vasilenko, R. N., Kozlovskaya, G. D., Dolgikh, D. A., Fink, A. L., Doniach, S. & Abramov, V. M. (1999). Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* **38**, 15009-16.

117. Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasilenko, R. N., Vasiliev, A. M., Rodionov, I. L., Kozlovskaya, G. D., Dolgikh, D. A., Fink, A. L., Doniach, S., Permyakov, E. A. & Abramov, V. M. (2000). Zn(2+)-mediated structure formation and compaction of the "natively unfolded" human prothymosin alpha. *Biochem. Biophys. Res. Commun.* **267**, 663-668.

118. Li, J., Uversky, V. N. & Fink, A. L. (2001). Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human alpha-synuclein. *Biochemistry* **40**, 11604-13.

119. Uversky, V. N., Li, J. & Fink, A. L. (2001). Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **276**, 10737-10744.

120. Uversky, V. N., Li, J., Souillac, P., Jakes, R., Goedert, M. & Fink, A. L. (2002). Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma- synucleins. *J. Biol. Chem.* **25**, 25.

121. Permyakov, S. E., Millett, I. S., Doniach, S., Permyakov, E. A. & Uversky, V. N. (2003). Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins* **53**, 855-62.

122. Munishkina, L. A., Fink, A. L. & Uversky, V. N. (In press). Formation of amyloid fibrils from the core histones *in vitro*. *J. Mol. Biol.*

123. Uversky, V. N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2-12.

124. Guinier, A. & Fournet, G. (1955). *Small-angle scattering of X-rays*, John Wiley & Sons.

125. Millett, I. S., Doniach, S. & Plaxco, K. W. (2002). Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Protein Chem.* **62**, 241-62.

126. Jaenicke, R. & Seckler, R. (1997). Protein misassembly in vitro. *Adv. Protein Chem.* **50**, 1-59.

127. Rose, G. D. (2002). *Unfolded Proteins*. Advances in Protein Chemistry (F.M., R., S., E. D. & J., K., Eds.), 62, Academic Press, New York.

128. Severgun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495-503.

129. Bergmann, A., Orthaber, D., Scherf, G. & Glatter, O. (2000). Improvement of SAXS measurements on Kratky slit systems by Göbel mirrors and imaging-plate detectors. *J. Appl. Cryst.* **33**, 869-875.

130. Panick, G., Malessa, R., Winter, R., Rapp, G., Frye, K. J. & Royer, C. A. (1998). Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle X-ray scattering and Fourier-transform infrared spectroscopy. *J. Mol. Biol.* **275**, 389-402.

131. Pérez, J., Vachette, P., Russo, D., Desmadril, M. & Durand, D. (2001). Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. *J. Mol. Biol.* **308**, 721-43.

132. Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* **32**, 13288-13298.

133. Uversky, V. N. (1994). Gel-permeation chromatography as a unique instrument for quantitative and qualitative analysis of protein denaturation and unfolding. *Int. J. Bio-Chromatography* **1**, 103-114.

134. Uversky, V. N. (2003). Protein folding revisited. A polypeptide chain at the folding - misfolding - no-folding crossroads: Which way to go? *Cell. Mol. Life Sci.*

135. Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739-756.

136. Uversky, V. N. & Ptitsyn, O. B. (1996). Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **255**, 215-28.

137. Fasman, G. D. (1996). *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenem Press, New York.

138. Adler, A. J., Greenfield, N. J. & Fasman, G. D. (1973). Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol.* **27**, 675-735.

139. Uversky, V. N. & Fink, A. L. (2002). The chicken-egg scenario of protein folding revisited. *FEBS Lett.* **515**, 79-83.

140. Palleros, D. R., Reid, K. L., McCarty, J. S., Walker, G. C. & Fink, A. L. (1992). DnaK, hsp73, and their molten globules. Two different ways heat shock proteins respond to heat. *J. Biol. Chem.* **267**, 5279-85.

141. Fink, A. L., Oberg, K. A. & Seshadri, S. (1998). Discrete intermediates versus molten globule models for protein folding: characterization of partially folded intermediates of apomyoglobin. *Fold. Des.* **3**, 19-25.

142. Uversky, V. N., Karnoup, A. S., Segel, D. J., Seshadri, S., Doniach, S. & Fink, A. L. (1998). Anion-induced folding of *Staphylococcal* nuclease: characterization of multiple equilibrium partially folded intermediates. *J. Mol. Biol.* **278**, 879-894.

143. Krimm, S. & Tiffany, M. L. (1974). The circular dichroism spectrum and structure of unordered polypeptides and proteins. *Israeli J. of Chem.* **12**, 189-200.

144. Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F. & Creamer, T. P. (2001). Host-guest study of left-handed polyproline II helix formation. *Biochemistry* **40**, 14376-83.

145. Rucker, A. L. & Creamer, T. P. (2002). Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci.* **11**, 980-5.

146. House-Pompeo, K., Xu, Y., Joh, D., Speziale, P. & Hook, M. (1996). Conformational changes in the binding MSCRAMMs are induced by ligand binding. *J. Biol. Chem.* **271**, 1379-1384.

147. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Müller-Frohne, M., Zirwer, D., Czarnecki, J. & Damaschun, G. (1995). Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry* **34**, 13211-13218.

148. Flaugh, S. L. & Lumb, K. J. (2001). Effects of macromolecular crowding on the intrinscially disordered proteins c-Fos and p27(Kip 1). *Biomacromolecules* **2**, 538-540.

149. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. (2003). Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2450-2455.

150. Bothner, B., Aubin, Y. & Kriwacki, R. W. (2003). Peptides derived from two dynamically disordered proteins self-assemble into amyloid-like fibrils. *J. Am. Chem. Soc.* **125**, 3200-1.

151. Mackay, J. P., Matthews, J. M., Winefield, R. D., Mackay, L. G., Haverkamp, R. G. & Templeton, M. D. (2001). The hydrophobin EAS is largely unstructured in solution and functions by forming amyloid-like structures. *Structure (Camb)* **9**, 83-91.

152. Baskakov, I. & Bolen, D. W. (1998). Forcing thermodynamically unfolded proteins to fold. *J. Biol. Chem.* **273**, 4831-4834.

153. Qu, Y., Bolen, C. L. & Bolen, D. W. (1998). Osmolyte-driven contraction of a random coil protein. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9268-9273.

154. Baskakov, I. V., Kumar, R., Srinivasan, G., Ji, Y. S., Bolen, D. W. & Thompson, E. B. (1999). Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. *J. Biol. Chem.* **274**, 10693-10696.

155. Qu, Y. & Bolen, D. W. (2002). Efficacy of macromolecular crowding in forcing proteins to fold. *Biophys. Chem.* **101-102**, 155-65.

156. Davis-Searles, P. R., Morar, A. S., Saunders, A. J., Erie, D. A. & Pielak, G. J. (1998). Sugar-induced molten-globule model. *Biochemistry* **37**, 17048-17053.

157. Saunders, A. J., Davis-Searles, P. R., Allen, D. L., Pielak, G. J. & Erie, D. A. (2000). Osmolyte-induced changes in protein conformational equilibria. *Biopolymers* **53**, 293-307.

158. Morar, A. S., Olteanu, A., Young, G. B. & Pielak, G. J. (2001). Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome c. *Protein Sci.* **10**, 2195-2199.

159. Chirgadze, Y. N., Shestopalov, B. V. & Venyaminov, S. Y. (1973). Intensities and other spectral parameters of infrared amide bands of polypeptides in the beta- and random forms. *Biopolymers* **12**, 1337-51.

160. Susi, H., Timasheff, S. N. & Stevens, L. (1967). Infrared spectra and protein conformations in aqueous solutions. I. The amide I band in H2O and D2O solutions. *J. Biol. Chem.* **242**, 5460-6.

161. Chirgadze, Y. N. & Brazhnikov, E. V. (1974). Intensities and other spectral parameters of infrared amide bands of polypeptides in the alpha-helical form. *Biopolymers* **13**, 1701-12.

162. Miyazawa, T. & Blout, E. R. (1961). The Infrared Spectra of Polypeptides in Various Conformations: Amide I and II Bands. *J. Am. Chem. Soc.* **83**, 712-719.

163. Lackowicz, J. (1999). *Principals of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publishers, New York.

164. Stryer, L. (1968). Fluorescence spectroscopy of proteins. *Science* **162**, 526-33.

165. Permyakov, E. A. (1993). *Luminescence spectroscopy of proteins*, CRC PRess, London.

166. Eftink, M. R. & Ghiron, C. A. (1981). Fluorescence quenching studies with proteins. *Anal. Biochem.* **114**, 199-227.

167. Lakowicz, J. R. & Weber, G. (1973). Quenching of fluorescence by oxygen. A probe for structural fluctuations in macromolecules. *Biochemistry* **12**, 4161-70.

168. Lakowicz, J. R. & Weber, G. (1973). Quenching of protein fluorescence by oxygen. Detection of structural fluctuations in proteins on the nanosecond time scale. *Biochemistry* **12**, 4171-9.

169. Eftink, M. R. & Ghiron, C. A. (1976). Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. *Biochemistry* **15**, 672-80.

170. Eftink, M. R. & Ghiron, C. A. (1975). Dynamics of a protein matrix revealed by fluorescence quenching. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 3290-4.

171. Eftink, M. R. & Ghiron, C. A. (1977). Exposure of tryptophanyl residues and protein dynamics. *Biochemistry* **16**, 5546-51.

172. Chaffotte, A. F. & Goldberg, M. E. (1984). Fluorescence-quenching studies on a conformational transition within a domain of the beta 2 subunit of *Escherichia coli* tryptophan synthase. *Eur. J. Biochem.* **139**, 47-50.

173. Varley, P. G., Dryden, D. T. & Pain, R. H. (1991). Resolution of the fluorescence of the buried tryptophan in yeast 3-phosphoglycerate kinase using succinimide. *Biochim. Biophys. Acta.* **1077**, 19-24.

174. Weber, G. (1952). Polarization of the fluorescence of macromolecules. II. Fluorescent conjugates of ovalbumin and bovine serum albumin. *Biochem. J.* **51**, 155-67.

175. Weber, G. (1952). Polarization of the fluorescence of macromolecules. I. Theory and experimental method. *Biochem. J.* **51**, 145-55.

176. Weber, G. (1953). Rotational Brownian motion and polarization of the fluorescence of solutions. *Adv. Protein Chem.* **8**, 415-59.

177. Weber, G. (1960). Fluorescence-polarization spectrum and electronic-energy transfer in proteins. *Biochem. J.* **75**, 345-52.

178. Semisotnov, G. V., Zikherman, K. K., Kasatkin, S. B., Ptitsyn, O. B. & V., A. E. (1981). Polarized luminescence and mobility of tryptophan residues in polypeptide chains. *Biopolymers* **20**, 2287-2309.

179. Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Gilmanshin, R. I., Lebedev Yu, O., Semisotnov, G. V., Tiktopulo, E. I., Ptitsyn, O. B. & et al. (1985). Compact state of a protein molecule with pronounced small-scale mobility: bovine alpha-lactalbumin. *Eur. Biophys. J.* **13**, 109-21.

180. Rodionova, N. A., Semisotnov, G. V., Kutyshenko, V. P., Uverskii, V. N. & Bolotina, I. A. (1989). [Staged equilibrium of carbonic anhydrase unfolding in strong denaturants]. *Mol Biol (Mosk)* **23**, 683-92.

181. Förster, T. (1948). Intermolecular energy migration and fluorescence. *Ann. Physik.* **2**, 55-75.

182. Lundblad, R. L. (1991). *Chemical reagents for protein modification*, CRC Press, Boca Raton.

183. Rischel, C. & Poulsen, F. M. (1995). Modification of a specific tyrosine enables tracing of the end-to-end distance during apomyoglobin folding. *FEBS Lett.* **374**, 105-9.

184. Uversky, V. N. & Fink, A. L. (1999). Do protein molecules have a native-like topology in the pre-molten globule state? *Biochemistry (Mosc)* **64**, 552-5.

185. Tcherkasskaya, O. & Ptitsyn, O. B. (1999). Molten globule versus variety of intermediates: influence of anions on pH-denatured apomyoglobin. *FEBS Lett.* **455**, 325-31.

186. Tcherkasskaya, O. & Ptitsyn, O. B. (1999). Direct energy transfer to study the 3D structure of non-native proteins: AGH complex in molten globule state of apomyoglobin. *Protein Eng.* **12**, 485-90.

187. Tcherkasskaya, O. & Uversky, V. N. (2001). Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins* **44**, 244-254.

188. Stryer, L. (1965). The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol.* **13**, 482-495.

189. Turner, D. C. & Brand, L. (1968). Quantitative estimation of protein binding site polarity. Fluorescence of N-arylaminonaphthalenesulfonates. *Biochemistry* **7**, 3381-90.

190. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-128.

191. Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J. & Ptitsyn, O. B. (1987). Sequential mechanism of refolding of carbonic anhydrase B. *FEBS Lett.* **224**, 9-13.

192. Uversky, V. N., Winter, S. & Lober, G. (1996). Use of fluorescence decay times of 8-ANS-protein complexes to study the conformational transitions in proteins which unfold through the molten globule state. *Biophys. Chem.* **60**, 79-88.

193. Uversky, V. N., Winter, S. & Lober, G. (1998). Self-association of 8-anilino-1-naphthalene-sulfonate molecules: spectroscopic characterization and application to the investigation of protein folding. *Biochim. Biophys. Acta.* **1388**, 133-142.

194. Uversky, V. N. & Ptitsyn, O. B. (1996). All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. *Fold. & Des.* **1**, 117-122.

195. Roberts, L. M. & Dunker, A. K. (1993). Structural changes accompanying chloroform-induced contraction of the filamentous phage fd. *Biochemistry* **32**, 10479-10488.

196. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Adv. Protein. Chem.* **33**, 167-241.

197. Ptitsyn, O. (1995). Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83-229.

198. Uversky, V. N. (1999). A multiparametric approach to studies of self-organization of globular proteins. *Biochemistry (Mosc)* **64**, 250-266.

199. Ptitsyn, O. B. & Uversky, V. N. (1994). The molten globule is a third thermodynamical state of protein molecules. *FEBS Lett.* **341**, 15-18.

200. Uversky, V. N., Permyakov, S. E., Zagranichny, V. E., Rodionov, I. L., Fink, A. L., Cherskaya, A. M., Wasserman, L. A. & Permyakov, E. A. (2002). Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein. *J. Proteome Res.* **1**, 149-159.

201. Timm, D. E., Vissavajjhala, P., Ross, A. H. & Neet, K. E. (1992). Spectroscopic and chemical studies of the interaction between nerve growth factor (NGF) and the extracellular domain of the low affinity NGF receptor. *Protein Sci.* **1**, 1023-31.

202. Kim, T. D., Ryu, H. J., Cho, H. I., Yang, C. H. & Kim, J. (2000). Thermal behavior of proteins: heat-resistant proteins and their heat-induced secondary structural changes. *Biochemistry* **39**, 14839-46.

203. Konno, T., Tanaka, N., Kataoka, M., Takano, E. & Maki, M. (1997). A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. *Biochim. Biophys. Acta.* **1342**, 73-82.

204. Lynn, A., Chandra, S., Malhotra, P. & Chauhan, V. S. (1999). Heme binding and polymerization by *Plasmodium falciparum* histidine rich protein II: influence of pH on activity and conformation. *FEBS Lett.* **459**, 267-71.

205. Johansson, J., Gudmundsson, G. H., Rottenberg, M. E., Berndt, K. D. & Agerberth, B. (1998). Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. *J. Biol. Chem.* **273**, 3718-3724.

206. Englander, S. W. & Krishna, M. M. (2001). Hydrogen exchange. *Nat. Struct. Biol.* **8**, 741-2.

207. Linderstrøm-Lang, K. (1955). Deuterium exchange between peptides and water. *Chem. Soc. (London) Spec. Publ.* **2**, 1-20.

208. Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287-386.

209. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86.

210. Hamuro, Y., Coales, S. J., Southern, M. R., Nemeth-Cawley, J. F., Stranz, D. D. & Griffin, P. R. (2003). Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry. *J Biomol Tech* **14**, 171-82.

211. Hamuro, Y., Coales, S. J., Southern, M. R., Nemeth-Cawley, J. F., Stranz, D. D. & Griffin, P. R. (2003). Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry. *J. Biomol. Tech.* **14**, 171-82.

212. Englander, J. J., Rogero, J. R. & Englander, S. W. (1985). Protein hydrogen exchange studied by the fragment separation method. *Anal. Biochem.* **147**, 234-44.

213. Rosa, J. J. & Richards, F. M. (1979). An experimental procedure for increasing the structural resolution of chemical hydrogen-exchange measurements on proteins: application to ribonuclease S peptide. *J. Mol. Biol.* **133**, 399-416.

214. Zhang, Z. & Smith, D. L. (1996). Thermal-induced unfolding domains in aldolase identified by amide hydrogen exchange and mass spectrometry. *Protein Sci* **5**, 1282-9.

215. Pantazatos, D., Kim, J. S., Klock, H. E., Stevens, R. C., Wilson, I. A., Lesley, S. A. & Woods, V. L., Jr. (2004). Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 751-6.

216. Wu, H. (1931). Studies on denaturation of proteins XIII A theory of denaturation. *Chin. J. Physiol.* **1**, 219-234.

217. Linderstrøm-Lang, K. (1952). Structure and enzymatic break-down of proteins. *Lane Medical Lectures* **6**, 117-126.

218. Markus, G. (1965). Protein substrate conformation and proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **54**, 253-258.

219. Wright, H. T. (1977). Secondary and conformational specificities of trypsin and chymotrypsin. *Eur. J. Biochem.* **73**, 567-578.

220. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-A resolution. *Biochemistry* **13**, 4212-4228.

221. Hubbard, S. J., Campbell, S. F. & Thornton, J. M. (1991). Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.* **220**, 507-530.

222. Hubbard, S. J., Beynon, R. J. & Thornton, J. M. (1998). Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures. *Protein Eng.* **11**, 349-359.

223. Hubbard, S. J., Eisenmenger, F. & Thornton, J. M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.* **3**, 757-768.

224. Fernandez, A. & Scheraga, H. A. (2003). Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 113-8.

225. Fontana, A., de Laureto, P. P., de Filippis, V., Scaramella, E. & Zambonin, M. (1997). Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* **2**, R17-R26.

226. Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A. & Scaramella, E. (1997). Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol.* **266**, 223-230.

227. Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415-27.

228. Williams, R. J. (1978). Energy states of proteins, enzymes and membranes. *Proc. R. Soc. Lond. B Biol. Sci.* **200**, 353-389.

229. Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701-8.

230. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure (Camb)* **11**, 1453-1459.

231. Romero, P., Obradovic, Z. & Dunker, A. K. (1997). Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform. Ser. Workshop Genome Inform.* **8**, 110-124.

232. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins* **42**, 38-48.

233. Jones, D. T. & Ward, J. (2003). Prediction of disordered regions in proteins from position specific score matrices. *Proteins* **53**, 573-578.

234. Melamud, E. & Moult, J. (2003). Evaluation of disorder predictions in CASP5. *Proteins* **53 Suppl 6**, 561-5.

235. Yang, S. A. & Klee, C. B. (2000). Low affinity Ca2+-binding sites of calcineurin B mediate conformational changes in calcineurin A. *Biochemistry* **39**, 16147-16154.

236. Brandstetter, H., Bauer, M., Huber, R., Lollar, P. & Bode, W. (1995). X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and hemophilia B. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 9796-9800.

237. Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. & Engh, R. A. (1996). X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J. Biol. Chem.* **271**, 29988-29992.

238. Padmanabhan, K., Padmanabhan, K. P., Tulinsky, A., Park, C. H., Bode, W., Huber, R., Blankenship, D. T., Cardin, A. D. & Kisiel, W. (1993). Structure of human des(1-45) factor Xa at 2.2 A resolution. *J. Mol. Biol.* **232**, 947-966.

239. Aviles, F. J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C. & Bradbury, E. M. (1978). The conformation of histone H5. Isolation and characterisation of the globular segment. *Eur. J. Biochem.* **88**, 363-371.

240. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219-223.

241. Graziano, V., Gerchman, S. E., Wonacott, A. J., Sweet, R. M., Wells, J. R., White, S. W. & Ramakrishnan, V. (1990). Crystallization of the globular domain of histone H5. *J. Mol. Biol.* **212**, 253-257.

242. Shaiu, W. L., Hu, T. & Hsieh, T. S. (1999). The hydrophilic, protease-sensitive terminal domains of eucaryotic DNA topoisomerases have essential intracellular functions. *Pac. Symp. Biocomput.* **4**, 578-589.

243. Berger, J. M., Gamblin, S. J., Harrison, S. C. & Wang, J. C. (1996). Structure and mechanism of DNA topoisomerase II. *Nature* **379**, 225-232.

244. Caron, P. R., Watt, P. & Wang, J. C. (1994). The C-terminal domain of Saccharomyces cerevisiae DNA topoisomerase II. *Mol. Cell. Biol.* **14**, 3197-3207.

245. Muchmore, S. W., Sattler, M., Liang, H., Meadows, R. P., Harlan, J. E., Yoon, H. S., Nettesheim, D., Chang, B. S., Thompson, C. B., Wong, S. L., Ng, S. L. & Fesik, S. W. (1996). X-ray and NMR structure of human Bcl-x_L, an inhibitor of programmed cell death. *Nature* **381**, 335-341.

246. Yamamoto, K., Ichijo, H. & Korsmeyer, S. J. (1999). BCL-2 is phosphorylated and inactivated by an ASK1/Jun N-terminal protein kinase pathway normally activated at G(2)/M. *Mol. Cell. Biol.* **19**, 8469-8478.

247. Cheng, E. H., Kirsch, D. G., Clem, R. J., Ravi, R., Kastan, M. B., Bedi, A., Ueno, K. & Hardwick, J. M. (1997). Conversion of Bcl-2 to a Bax-like death effector by caspases. *Science* **278**, 1966-1968.

248. Chang, B. S., Minn, A. J., Muchmore, S. W., Fesik, S. W. & Thompson, C. B. (1997). Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2. *EMBO J.* **16**, 968-977.

249. Holliger, P., Riechmann, L. & Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 A: evidence for conformational lability. *J. Mol. Biol.* **288**, 649-657.

250. Holliger, P. & Riechmann, L. (1997). A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure* **5**, 265-275.

251. Nilsson, N., Malmborg, A. C. & Borrebaeck, C. A. (2000). The phage infection process: a functional role for the distal linker region of bacteriophage protein 3. *J. Virol.* **74**, 4229-4235.

252. Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T. & Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell* **85**, 931-942.

253. Arold, S., Franken, P., Strub, M. P., Hoh, F., Benichou, S., Benarous, R. & Dumas, C. (1997). The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. *Structure* **5**, 1361-1372.

254. Geyer, M., Munte, C. E., Schorr, J., Kellner, R. & Kalbitzer, H. R. (1999). Structure of the anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. *J. Mol. Biol.* **289**, 123-138.

255. Arold, S. T. & Baur, A. S. (2001). Dynamic Nef and Nef dynamics: how structure could explain the complex activities of this small HIV protein. *Trends Biochem. Sci.* **26**, 356-363.

256. Karlin, D., Longhi, S., Receveur, V. & Canard, B. (2002). The N-terminal domain of the phosphoprotein of *morbilliviruses* belongs to the natively unfolded class of proteins. *Virology* **296**, 251-262.

257. Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **84**, 3239-52.

258. Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S. & Canard, B. (2003). The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.* **278**, 18638.

259. Lesk, A. M., Levitt, M. & Chothia, C. (1986). Alignment of the amino acid sequences of distantly related proteins using variable gap penalties. *Protein Eng.* **1**, 77-8.

260. Lesk, A. M. & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225-70.

261. Chothia, C. & Lesk, A. M. (1987). The evolution of protein structures. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 399-405.

262. Harrison, R. W., Chatterjee, D. & Weber, I. T. (1995). Analysis of six protein structures predicted by comparative modeling techniques. *Proteins* **23**, 463-71.

263. Skolnick, J. & Fetrow, J. S. (2000). From genes to protein structure and function: novel applications of computational approaches in the genomic era. *Trends Biotechnol.* **18**, 34-9.

264. Skolnick, J., Fetrow, J. S. & Kolinski, A. (2000). Structural genomics and its importance for gene function analysis. *Nat. Biotechnol.* **18**, 283-287.

265. Thornton, J. M., Orengo, C. A., Todd, A. E. & Pearl, F. M. (1999). Protein folds, functions and evolution. *J. Mol. Biol.* **293**, 333-42.

266. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. & Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **11**, 161-171.

267. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T. & Dahlquist, F. W. (1997). The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nat. Struct. Biol.* **4**, 285-291.

268. Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (1997). Structure of the recombinant full-length hamster prion protein PrP(29- 231): the N terminus is highly flexible. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13452-13457.

269. Jacobs, D. M., Lipton, A. S., Isern, N. G., Daughdrill, G. W., Lowry, D. F., Gomes, X. & Wold, M. S. (1999). Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker. *J. Biomol. NMR* **14**, 321-331.

270. Lee, H., Mok, K. H., Muhandiram, R., Park, K. H., Suk, J. E., Kim, D. H., Chang, J., Sung, Y. C., Choi, K. Y. & Han, K. H. (2000). Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem.* **275**, 29426-29432.

271. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527-33.

272. Daughdrill, G. W., Ackerman, J., Isern, N. G., Botuyan, M. V., Arrowsmith, C., Wold, M. S. & Lowry, D. F. (2001). The weak interdomain coupling observed in the 70 kDa subunit of human replication protein A is unaffected by ssDNA binding. *Nucleic Acids Res.* **29**, 3270-3276.

273. Chothia, C. & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823-826.

274. Phillips, A., Janies, D. & Wheeler, W. (2000). Multiple sequence alignment in phylogenetic analysis. *Mol. Phylogenet. Evol.* **16**, 317-30.

275. Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T., Oldfield, C. J., Williams, C. J. & Dunker, A. K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* **55**, 104-110.

276. Shen, J. C., Lao, Y., Kamath-Loeb, A., Wold, M. S. & Loeb, L. A. (2003). The N-terminal domain of the large subunit of human replication protein A binds to Werner syndrome protein and stimulates helicase activity. *Mech. Ageing. Dev.* **124**, 921-30.

277. Longhese, M. P., Plevani, P. & Lucchini, G. (1994). Replication factor A is required in vivo for DNA replication, repair, and recombination. *Mol. Cell. Biol.* **14**, 7884-90.

278. Umezu, K., Sugawara, N., Chen, C., Haber, J. E. & Kolodner, R. D. (1998). Genetic analysis of yeast RPA1 reveals its multiple functions in DNA metabolism. *Genetics* **148**, 989-1005.

279. Wold, M. S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61-92.

280. Goldman, N., Thorne, J. L. & Jones, D. T. (1998). Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics* **149**, 445-58.

281. Lio, P., Goldman, N., Thorne, J. L. & Jones, D. T. (1998). PASSML: combining evolutionary inference and protein secondary structure prediction. *Bioinformatics* **14**, 726-33.

282. Thorne, J. L., Goldman, N. & Jones, D. T. (1996). Combining protein evolution and secondary structure. *Mol. Biol. Evol.* **13**, 666-73.

283. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978). A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure* **5**, 345-352.

284. Henikoff, S. & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10915-9.

285. Radivojac, P., Obradovic, Z., Brown, C. J. & Dunker, A. K. (2002). Improving sequence alignments for intrinsically disordered proteins. *Pac. Symp. Biocomput.* **7**, 589-600.

286. Radivojac, P., Obradovic, Z., Brown, C. J. & Dunker, A. K. (2002). *Pac. Symp. Biocomput.*

287. Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J. & Dunker, A. K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* **55**, 104-10.

288. Luby-Phelps, K. (2000). Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *Int. Rev. Cytol.* **192**, 189-221.

289. Zimmerman, S. B. & Minton, A. P. (1993). Macromolecular crowding: biochemical, biophysical and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 27-65.

290. Fulton, A. B. (1982). How crowded is the cytoplasm? *Cell* **30**, 345-347.

291. Davis-Searles, P. R., Saunders, A. J., Erie, D. A., Winzor, D. J. & Pielak, G. J. (2001). Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 271-306.

292. Minton, A. P. (2001). The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276**, 10577-80.

293. Sasahara, K., McPhie, P. & Minton, A. P. (2003). Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J. Mol. Biol.* **326**, 1227-37.

294. Elston, T. C. & Oster, G. (1997). Protein turbines. I: The bacterial flagellar motor. *Biophys. J.* **73**, 703-21.

295. Hughes, K. T., Gillen, K. L., Semon, M. J. & Karlinsey, J. E. (1993). Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator. *Science* **262**, 1277-80.

296. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol.* **4**, 630-634.

297. Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S. & Dobson, C. M. (1999). Alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat. Struct. Biol.* **6**, 948-52.

298. McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. (2002). Structural properties of an amyloid precursor of beta(2)-microglobulin. *Nat. Struct. Biol.* **9**, 326-331.

299. Shimba, N., Serber, Z., Ledwidge, R., Miller, S. M., Craik, C. S. & Dötsch, V. (2003). Quantitative identification of the protonation state of histidines *in vitro* and *in vivo*. *Biochemistry* **42**, 9227-34.

300. Serber, Z. & Dötsch, V. (2001). In-cell NMR spectroscopy. *Biochemistry* **40**, 14317-23.

301. Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. (2001). Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **123**, 8895-901.

302. Honnappa, S., Cutting, B., Jahnke, W., Seelig, J. & Steinmetz, M. O. (2003). Thermodynamics of the Op18/stathmin-tubulin interaction. *J. Biol. Chem.* **278**, 38926-34.

303. Demchenko, A. P. (2001). Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.* **14**, 42-61.

304. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54-60.

305. Spolar, R. S. & Record, M. T., Jr. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-84.

306. Uversky, V. N. & Narizhneva, N. V. (1998). Effect of natural ligands on the structural properties and conformational stability of proteins. *Biochemistry (Mosc)* **63**, 420-33.

307. Uversky, V. N. (2003). A rigidifying union: The role of ligands in protein structure and stability. In *Recent Research Developments in Biophysics & Biochemistry* (Pandalai, S. G., ed.), Vol. 3, pp. 711-745. Transworld Research Network, Kerala, India.

308. Zwahlen, C., Li, S. C., Kay, L. E., Pawson, T. & Forman-Kay, J. D. (2000). Multiple modes of peptide recognition by the PTB domain of the cell fate determinant Numb. *EMBO J.* **19**, 1505-15.

309. Li, S. C., Zwahlen, C., Vincent, S. J., McGlade, C. J., Kay, L. E., Pawson, T. & Forman-Kay, J. D. (1998). Structure of a Numb PTB domain-peptide complex suggests a basis for diverse binding specificity. *Nat. Struct. Biol.* **5**, 1075-83.

310. Jen-Jacobson, L., Engler, L. E. & Jacobson, L. A. (2000). Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure Fold. Des.* **8**, 1015-23.

311. Wester, M. R., Johnson, E. F., Marques-Soares, C., Dansette, P. M., Mansuy, D. & Stout, C. D. (2003). Structure of a substrate complex of mammalian cytochrome P450 2C5 at

2.3 A resolution: evidence for multiple substrate binding modes. *Biochemistry* **42**, 6370-9.

312. Furuke, K., Shiraishi, M., Mostowski, H. S. & Bloom, E. T. (1999). Fas ligand induction in human NK cells is regulated by redox through a calcineurin-nuclear factors of activated T cell-dependent pathway. *J. Immunol.* **162**, 1988-93.

313. Liu, J., Farmer, J. D., Jr., Lane, W. S., Friedman, J., Weissman, I. & Schreiber, S. L. (1991). Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP- FK506 complexes. *Cell* **66**, 807-815.

314. O'Day, D. H. (2003). CaMBOT: profiling and characterizing calmodulin-binding proteins. *Cell Signal.* **15**, 347-54.

315. Zhang, L. & Lu, Y. T. (2003). Calmodulin-binding protein kinases in plants. *Trends Plant Sci.* **8**, 123-7.

316. Urbauer, J. L., Short, J. H., Dow, L. K. & Wand, A. J. (1995). Structural analysis of a novel interaction by calmodulin: high-affinity binding of a peptide in the absence of calcium. *Biochemistry* **34**, 8099-8109.

317. Sandak, B., Wolfson, H. J. & Nussinov, R. (1998). Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. *Proteins* **32**, 159-74.

318. Yang, S. A. & Klee, C. (2002). Study of calcineurin structure by limited proteolysis. *Methods Mol. Biol.* **172**, 317-34.

319. Manalan, A. S. & Klee, C. B. (1983). Activation of calcineurin by limited proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4291-4295.

320. Zidek, L., Novotny, M. V. & Stone, M. J. (1999). Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.* **6**, 1118-21.

321. Loh, A. P., Pawley, N., Nicholson, L. K. & Oswald, R. E. (2001). An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics in Cdc42Hs. *Biochemistry* **40**, 4590-600.

322. Leontiev, V. V., Uversky, V. N., Permyakov, E. A. & Murzin, A. G. (1993). Introduction of Ca(2+)-binding amino-acid sequence into the T4 lysozyme. *Biochim. Biophys. Acta.* **1162**, 84-8.

323. VanScyoc, W. S. & Shea, M. A. (2001). Phenylalanine fluorescence studies of calcium binding to N-domain fragments of Paramecium calmodulin mutants show increased calcium affinity correlates with increased disorder. *Protein Sci.* **10**, 1758-68.

324. Sorensen, B. R. & Shea, M. A. (1998). Interactions between domains of apo calmodulin alter calcium binding and stability. *Biochemistry* **37**, 4244-53.

325. Fredricksen, R. S. & Swenson, C. A. (1996). Relationship between stability and function for isolated domains of troponin C. *Biochemistry* **35**, 14012-26.

326. Li, Z., Stafford, W. F. & Bouvier, M. (2001). The metal ion binding properties of calreticulin modulate its conformational flexibility and thermal stability. *Biochemistry* **40**, 11193-201.

327. Eakin, C. M., Knight, J. D., Morgan, C. J., Gelfand, M. A. & Miranker, A. D. (2002). Formation of a copper specific binding site in non-native states of beta-2-microglobulin. *Biochemistry* **41**, 10646-10656.

328. Opitz, U., Rudolph, R., Jaenicke, R., Ericsson, L. & Neurath, H. (1987). Proteolytic dimers of porcine muscle lactate dehydrogenase: characterization, folding, and reconstitution of the truncated and nicked polypeptide chain. *Biochemistry* **26**, 1399-406.

329. Namba, K. (2001). Roles of partly unfolded conformations in macromolecular self-assembly. *Genes Cells* **6**, 1-12.

330. Hoh, J. H. (1998). Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins* **32**, 223-228.

331. Hupp, T. R., Meek, D. W., Midgley, C. A. & Lane, D. P. (1992). Regulation of the specific DNA binding function of p53. *Cell* **71**, 875-86.

332. Brown, H. G. & Hoh, J. H. (1997). Entropic exclusion by neurofilament sidearms: a mechanism for maintaining interfilament spacing. *Biochemistry* **36**, 15035-15040.

333. Kumar, S., Yin, X., Trapp, B. D., Hoh, J. H. & Paulaitis, M. E. (2002). Relating interactions between neurofilaments to the structure of axonal neurofilament distributions through polymer brush models. *Biophys. J.* **82**, 2360-2372.

334. Kumar, S., Yin, X., Trapp, B. D., Paulaitis, M. E. & Hoh, J. H. (2002). Role of long-range repulsive forces in organizing axonal neurofilament distributions: evidence from mice deficient in myelin-associated glycoprotein. *J. Neurosci. Res.* **68**, 681-690.

335. Wissmann, R., Baukrowitz, T., Kalbacher, H., Kalbitzer, H. R., Ruppersberg, J. P., Pongs, O., Antz, C. & Fakler, B. (1999). NMR structure and functional characteristics of the hydrophilic N terminus of the potassium channel beta-subunit Kvbeta1.1. *J. Biol. Chem.* **274**, 35521-35525.

336. Armstrong, C. M. & Bezanilla, F. (1977). Inactivation of the sodium channel. II. Gating current experiments. *J. Gen. Physiol.* **70**, 567-590.

337. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. (1990). Biophysical and molecular mechanisms of *Shaker* potassium channel inactivation. *Science* **250**, 533-538.

338. Helmes, M., Trombitas, K., Centner, T., Kellermayer, M., Labeit, S., Linke, W. A. & Granzier, H. (1999). Mechanically driven contour-length adjustment in rat cardiac titin's unique N2B sequence. *Circ. Res.* **84**, 1339-1352.

339. Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M. & Zambonin, M. (1986). Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **25**, 1847-1851.

340. Fontana, A., de Laureto, P. P., de Filippis, V., Scaramella, L. & Zambonin, M. (1999). Limited proteolysis in the study of protein conformation. In *Proteolytic Enzymes: Tool and Targets*, pp. 257 - 284, Springer Verlag.

341. Marks, F. (1996). *Protein Phosphorylation*, VCH Weinheim, New York, Basel, Cambridge, Tokyo.

342. Johnson, L. N. & Lewis, R. J. (2001). Structural basis for control by phosphorylation. *Chem. Rev.* **101**, 2209-2242.

343. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037-49.

344. Kwong, P. D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J. S., Hellmig, B. D., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1999). Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem.* **274**, 4115-23.

345. Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H. & Huber, R. (1993). Phosphotransferase and substrate binding mechanism of the cAMP- dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 A structure of the complex with Mn2+ adenylyl imidodiphosphate and inhibitor peptide PKI(5-24). *EMBO J.* **12**, 849-59.

346. Narayana, N., Cox, S., Shaltiel, S., Taylor, S. S. & Xuong, N. (1997). Crystal structure of a polyhistidine-tagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5-24) and adenosine. *Biochemistry* **36**, 4438-4448.

347. Lowe, E. D., Noble, M. E., Skamnaki, V. T., Oikonomakos, N. G., Owen, D. J. & Johnson, L. N. (1997). The crystal structure of a phosphorylase kinase peptide substrate complex: kinase substrate recognition. *Embo. J.* **16**, 6646-6658.

348. ter Haar, E., Coll, J. T., Austen, D. A., Hsiao, H. M., Swenson, L. & Jain, J. (2001). Structure of GSK3beta reveals a primed phosphorylation mechanism. *Nat. Struct. Biol.* **8**, 593-596.

349. Hubbard, S. R. (1997). Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *Embo. J.* **16**, 5572-5581.

350. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.

351. Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* **23**, 121-282.

352. Flory, P., J. (1969). *Statistical Mechanics of Chain Molecules*, John wiley.

353. Baldwin, R. L. (2002). A new perspective on unfolded proteins. *Adv Protein Chem* **62**, 361-7.

354. Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000). The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12565-12570.

355. Zhou, H. X. (2004). Polymer models of protein stability, folding, and interactions. *Biochemistry* **43**, 2141-54.

356. Huber, R. (1979). Conformational flexibility in protein molecules. *Nature* **280**, 538-539.

357. Kossiakoff, A. A., Chambers, J. L., Kay, L. M. & Stroud, R. M. (1977). Structure of bovine trypsinogen at 1.9 A resolution. *Biochemistry* **16**, 654-664.

358. Bennett, W. S. & Huber, R. (1984). Structural and functional aspects of domain motions in proteins. *Crit. Rev. Biochem.* **15**, 291-384.

359. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity and disordered protein. *Proteins: Structure, Function, Genetics* **42**, 38-49.

360. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985-94.

361. Gromiha, M. M. & Selvaraj, S. (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J Mol Biol* **310**, 27-32.

362. Mirny, L. & Shakhnovich, E. (2001). Protein folding theory: from lattice to all-atom models. *Annu Rev Biophys Biomol Struct* **30**, 361-96.

363. Zhou, H. & Zhou, Y. (2002). Folding rate prediction using total contact distance. *Biophys J* **82**, 458-63.

364. Makarov, D. E. & Plaxco, K. W. (2003). The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Sci* **12**, 17-26.

365. Makarov, D. E., Keller, C. A., Plaxco, K. W. & Metiu, H. (2002). How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc Natl Acad Sci U S A* **99**, 3535-9.

366. Debe, D. A. & Goddard, W. A., 3rd. (1999). First principles prediction of protein folding rates. *J Mol Biol* **294**, 619-25.

367. Debe, D. A., Carlson, M. J. & Goddard, W. A., 3rd. (1999). The topomer-sampling model of protein folding. *Proc Natl Acad Sci U S A* **96**, 2596-601.

368. Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663.

369. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. of Magn. Reson. B* **3**, 203-216.

370. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. of Magn. Reson. B* **2**, 201-205.

371. Dingley, A. J., Mackay, J. P., Chapman, B. E., Morris, M. B., Kuchel, P. W., Hambly, B. D. & King, G. F. (1995). Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *J. Biomol. NMR* **6**, 321-8.

372. Stejskal, E. O. & Tanner, J. E. (1965). Spin diffusion measurements spin echoes in the presence of a time dependent field gradient. *J. Chem. Phys.* **42**, 288-292.

373. Peng, J. W. & Wagner, G. (1995). Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry* **34**, 16733-52.

374. Farrow, N. A., Zhang, O. W., Szabo, A., Torchia, D. A. & Kay, L. E. (1995). Spectral density-function mapping using N-15 relaxation data exclusively. *J. Biomol. NMR* **6**, 153-162.

375. Planson, A. G., Guijarro, J. I., Goldberg, M. E. & Chaffotte, A. F. (2003). Assistance of maltose binding protein to the in vivo folding of the disulfide-rich C-terminal fragment from Plasmodium falciparum merozoite surface protein 1 expressed in *Escherichia coli. Biochemistry* **42**, 13202-11.

376. Stout, G. H. & Jensen, L. H. (1989). *X-Ray Structure Determination: A Practical Guide*, Wiley-Interscience, New York.

377. McRee, D. E. (1997). *Practical Protein Crystallography*. 2nd edit, Academic Press, New York.

378. Ikemoto, N., Nagy, B., Bhatnagar, G. M. & Gergely, J. (1974). Studies on a metal-binding protein of the sarcoplasmic reticulum. *J. Biol. Chem.* **249**, 2357-65.

379. Ostvald, T. V., MacLennon, D. H. & Dorrington, K. J. (1974). Effects of Cation Binding on the Conformation of Calsequestrin and the High Affinity Calcium-binding Protein of Sarcoplasmic Reticulum. *J. Biol. Chem.* **249**, 567-5871.

380. Aaron, B. M., Oikawa, K., Reithmeier, R. A. & Sykes, B. D. (1984). Characterization of skeletal muscle calsequestrin by 1H NMR spectroscopy. *J. Biol. Chem.* **259**, 11876-11881.

381. He, Z., Dunker, A. K., Wesson, C. R. & Trumble, W. R. (1993). Ca(2+)-induced folding and aggregation of skeletal muscle sarcoplasmic reticulum calsequestrin. The involvement of the trifluoperazine-binding site. *J. Biol. Chem.* **268**, 24635-24641.

382. Ikemoto, N., Bhatnagar, G. M., Nagy, B. & Gergely, J. (1972). Interaction of divalent cations with the 55,000-dalton protein component of the sarcoplasmic reticulum. Studies of fluorescence and circular dichroism. *J. Biol. Chem.* **247**, 7835-7.

383. Cozens, B. & Reithmeier, R. A. (1984). Size and shape of rabbit skeletal muscle calsequestrin. *J. Biol. Chem.* **259**, 6248-6252.

384. Ohnishi, M. & Reithmeier, R. A. (1987). Fragmentation of rabbit skeletal muscle calsequestrin: spectral and ion binding properties of the carboxyl-terminal region. *Biochemistry* **26**, 7458-65.

385. Mitchell, R. D., Simmerman, H. K. & Jones, L. R. (1988). Ca2+ binding effects on protein conformation and protein interactions of canine cardiac calsequestrin. *J Biol. Chem.* **263**, 1376-81.

386. Williams, R. W. & Beeler, T. J. (1986). Secondary structure of calsequestrin in solutions and in crystals as determined by Raman spectroscopy. *J. Biol. Chem.* **261**, 12408-12413.

387. Maurer, A., Tanaka, M., Ozawa, T. & Fleischer, S. (1985). Purification and crystallization of the calcium binding protein of sarcoplasmic reticulum from skeletal muscle. *Proc. Natl. Acad. Sci. U. S. A. U.S.A.* **82**, 4036-40.

388. Franzini-Armstrong, C., Kenney, L. J. & Varriano-Marston, E. (1987). The structure of calsequestrin in triads of vertebrate skeletal muscle: a deep-etch study. *J. Cell Biol.* **105**, 49-56.

389. Kang, C. H., Trumble, W. R. & Dunker, A. K. (2001). Crystallization and structure-function of calsequestrin. In *Calcium Binding Protein Protocols: volume 1 reviews and case studies* (Vogel, H. J., ed.), Vol. 172, pp. 281-294. Humana Press, Totowa, New Jersey.

390. Wang, S., Trumble, W. R., Liao, H., Wesson, C. R., Dunker, A. K. & Kang, C. H. (1998). Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. *Nat. Struct. Biol.* **5**, 476-483.

391. McPherson, A. (1990). Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189**, 1-23.

392. Jancarik, J. & Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409-411.

393. Fliegel, L., Ohnishi, M., Carpenter, M. R., Khanna, V. K., Reithmeier, R. A. F. & MacLennan, D. H. (1987). Amino acid sequence of rabbit fast-twitch skeletal muscle calsequestrin deduced from cDNA and peptide sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 1167-1171.

394. Scott, B. T., Simmerman, H. K., Collins, J. H., Nadal-Ginard, B. & Jones, L. R. (1988). Complete amino acid sequence of canine cardiac calsequestrin deduced by cDNA cloning. *J Biol. Chem.* **263**, 8958-64.

395. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.

396. Antz, C., Geyer, M., Fakler, B., Schott, M. K., Guy, H. R., Frank, R., Ruppersberg, J. P. & Kalbitzer, H. R. (1997). NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. *Nature* **385**, 272-275.