Assessing Protein Disorder and Induced Folding

Véronique Receveur-Bréchot,[1] Jean-Marie Bourhis,[1] Vladimir N. Uversky,[2,3] Bruno Canard,[1] and Sonia Longhi[1]*
[1]*Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités Aix-Marseille I et II, Campus de Luminy, Marseille Cedex 09, France*
[2]*Department of Biochemistry and Molecular Biology and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202*
[3]*Institute for Biological Instrumentation, Russian Academy of Sciences, Moscow Region, Russia*

ABSTRACT **Intrinsically disordered proteins (IDPs) defy the structure–function paradigm as they fulfill essential biological functions while lacking well-defined secondary and tertiary structures. Conformational and spectroscopic analyses showed that IDPs do not constitute a uniform family, and can be divided into subfamilies as a function of their residual structure content. Residual intramolecular interactions are thought to facilitate binding to a partner and then induced folding. Comprehensive information about experimental approaches to investigate structural disorder and induced folding is still scarce. We herein provide hints to readily recognize features typical of intrinsic disorder and review the principal techniques to assess structural disorder and induced folding. We describe their theoretical principles and discuss their respective advantages and limitations. Finally, we point out the necessity of using different approaches and show how information can be broadened by the use of multiples techniques. Proteins 2006;62:24–45.**

Key words: protein disorder; induced folding; structure–function paradigm

INTRODUCTION

Intrinsically disordered proteins (IDPs) are functional proteins that fulfill essential biological functions while lacking highly populated and uniform secondary and tertiary structure in physiological conditions.[1–7]

So far, regions lacking specific 3D structure have been mainly found in proteins involved in regulatory and signaling events.[2,4,8,9] The functional significance of disorder resides in an increased structural plasticity. This allows (1) to couple a high specificity with a low affinity, (2) to ensure faster association and dissociation rates, (3) to enable binding of numerous structurally distinct targets,[10] and (4) to provide the ability to overcome steric restrictions, thus leading to larger surfaces of interaction between partners.[1,3,11–15]

Although there are IDPs that carry out their function while remaining disordered all the time (e.g., entropic chains),[1] a majority undergo an unstructured-to-structured transition upon binding to their physiological partner(s), a process termed induced folding.[10,16–18] The term induced folding is often associated to a notion of gain of regular secondary structure. However, coupled folding and binding processes can also occur without such dramatic structural transitions.[19,20] In those cases, binding to the partner would nevertheless lead to a more ordered state, through selection of a conformer of the IDP out of the numerous possible conformational states adopted by the unbound form in solution. The reduced conformational entropy of the IDP would then fall in the range of structural transitions typifying induced folding. Whatever the extent and the nature of the structural transition involved, induced folding is thought to lead to an increased affinity for the partner, and to the shielding of specific regions of interaction, thus resulting in a modulation of activity.

Regions of intrinsic disorder are characterized by a distribution of Ramachandran phi and psi angles for each amino acid residue, giving rise to an ensemble of interconverting conformers. It has been shown that unfolded peptides and proteins have a tendency to adopt a poly(L-proline) II (PII) conformation, that is, an extended left-handed α-helix.[21,22] As for the functional relevance, the local concentration of residues in the PII region of the conformational space likely lowers the entropy of the unfolded protein chain, thereby facilitating folding under appropriate conditions.[21]

Nonrigid proteins show an extremely wide diversity in their structural properties. Indeed, IDPs can attain extended conformations (random coil-like) or remain globally collapsed (molten globule-like), where the latter possess regions of fluctuating secondary structure.[11] Conformational and spectroscopic analyses further showed that random coil-like proteins can be subdivided in their turn into two major groups. The first group consists of proteins with extended maximum dimensions typical of random coils with no (or little) secondary structure. The second group comprises the so-called premolten globules, which

Grant sponsor: the Commission of the European Communities, specific RTD programme "Quality of Life and Management of Living Resources;" Grant number: QLK2-CT2001-01225 ("Towards the design of new potent antiviral drugs: structure–function analysis of Paramyxoviridae RNA polymerase").

*Correspondence to: Sonia Longhi, AFMB, UMR 6098, CNRS and Universités Aix-Marseille I et II, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France. E-mail: sonia.longhi@afmb.univ-mrs.fr

Received 8 April 2005; Revised 12 July 2005;Accepted 30 July 2005

Published online 14 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20750

are more compact (but still less compact than globular or molten globule proteins) and conserve some residual secondary structure.[1,3] The residual intramolecular interactions that typify the premolten globule state may enable a more efficient start of the folding process induced by a partner.[2,18,23,24] Thus, the functional relevance of premolten globules may reside in a more pronounced propensity to undergo induced folding compared to random coil-like structures.[2,18,23,24]

Despite the rapid development of powerful computing approaches for the prediction of disorder, experimental data on the molecular mechanisms that govern the interaction of IDPs with their partners are still limited (for a data base of protein disorder, DisProt, see http://www.disprot.org).[25] Furthermore, comprehensive description of the experimental approaches for the study of protein intrinsic disorder is scarce, except for a few examples.[2,3,26,27]

Here, we review the principal experimental approaches that can be used to assess structural disorder and induced folding. We highlight the type of information that can be derived by the different approaches and point out how information can be broadened by using different methods.

Structural Overview of Partially Folded Proteins

It is useful to introduce the topic of IDPs with a brief description of different protein conformations. It has been pointed out that structurally IDPs resemble unfolded or different partially folded conformations detected for "normal" ordered globular proteins under various denaturing conditions; that is, molten globule, premolten globule, and unfolded or random coil. This presumes that nature can modulate structure of an IDP by varying its amino acid sequence in a similar way as an experimentalist can modify structure of a globular protein by changing the environmental conditions.

From the structural viewpoint, ordered globular protein can be considered as an aperiodic crystal in which the location of virtually all the atoms is fixed in space. In other words, the 3D structure of the ordered protein is characterized by slight variations in Ramachandran angles around their equilibrium positions. Such structure, being relatively rigid, experiences occasional cooperative conformational switches. In application to structural characterization this means that all the molecules of an ordered protein have nearly the same structure. Notably, some ordered proteins can yet exist in various conformations due to the occurence of flexible hinges connecting globular domains.[28]

Conversely, a sample of a disordered protein consists of a broad ensemble of molecules each having a different conformation. Therefore, averaged quantities have different meanings for ordered and disordered proteins.[27] For an ordered protein, an averaged quantity, such as the circular dichroism signal, gives information about each molecule in the sample, because nearly all the molecules are in the same structural state. For a 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.

The molten globule is the best studied partially folded conformation so far, whose structural properties have been thoroughly systematized (e.g., ref. 29). It has been established that the molten globule is compact (the average value for the increase in the molten globule hydrodynamic radius compared with that of ordered state is no more than 15%, which corresponds to a volume increase of ~50%). Molten globule is a globular conformation that is characterized by highly developed ordered secondary structure, but has no (or has only a trace of) rigid cooperatively melting tertiary structure. 2D NMR analysis revealed the presence of relatively stable folding pattern, or topology in molten globules.[30–37] Finally, molten globule is characterized by a considerable increase in the accessibility of a protein molecule to proteases[38,39] and in the affinity of a protein molecule to hydrophobic fluorescence probes (such as 8-anilinonaphthalene-1-sulfonate, ANS) due to the presence of solvent-exposed hydrophobic clusters.[40,41]

The premolten globule is characterized by the presence of some secondary structure, although much less than that of the molten globule. This conformation is nonglobular and considerably less compact than the molten globule, but it is still more compact than the random coil. It may interact with ANS, although more weakly than molten globules, suggesting that at least some hydrophobic clusters might be present in premolten globules. Importantly, premolten globules were shown to be separated from molten globules by an all-or-none transition, implying that these partially folded conformations represent discrete phases.[42] Some very flexible elements of ordered structure in premolten globules may occupy specific positions.[43] Therefore, any small variations in the protein environment; that is, changes in the thermodynamic quality of the solvent or changes induced by proton transfer, interactions with a ligand, temperature fluctuation, and so on, can trigger transition of a premolten globule to a molten globule or to an ordered conformation.

The random coil represents an ensemble of rapidly interchanging conformations, some of which are extended, and some more compact. It is possible that such stabilizing interactions can induce a more populated ensemble of chain conformations, and, if such structures exist in the unfolded state, they would probably guide the folding process and function as folding initiation sites. It has been emphasized that unfolded proteins are not strictu senso random coils, and numerous protein folding studies revealed the presence of an assured residual structure even under the most severe denaturing conditions, such as high concentrations of strong denaturants.[44–49] Thus, unfolded proteins with their complex polymeric nature exhibit nonrandom coil behavior. In fact, a total lack of intraresidue interactions would be unexpected in the unfolded state because certain (e.g., hydrophobic) side chains have high affinity for each other in the folded state.[49] In addition, some secondary structure within unfolded proteins could be expected due to the preferential distribution of phi and psi angles,[50–52] and some residual hydrophobic interactions may also be present.[49,53] All this restricts consider-

ably the conformational space of the unfolded polypeptide chain.

In the context of a living cell, although folding and self-organization of the ordered proteins is a spontaneous process, there are critical steps at which the participation of cell factors may be required. The main role of these proteinaceous factors, known as molecular chaperones or heat-shock proteins, is the maintenance of optimal conditions during the protein self-organization by removing obstacles leading to incorrect contacts, nonspecific association, and aggregation of folding intermediates.[54,55] On the other hand, despite the presence of molecular chaperones, IDPs conserve their predominant disordered state owing to their peculiar sequence properties. Nevertheless, one of the most widely studied heat-shock protein, GroEL or chaperonin-60, was shown to predominantly interact with molten globule-like proteins.[56–58] This points out the inherent capability of chaperones to recognize and bind partially ordered conformations of target proteins, without necessarily promoting their folding.

HOW CAN INTRINSIC DISORDER BE ANTICIPATED?

IDPs possess distinct sequence features that distinguish them from structured proteins. Similar to structured proteins, whose 3D structure is encoded within their primary structure,[59] the inability to fold and adopt a precise 3D structure can be anticipated by the analysis of their amino acid sequence. Thus, computational methods, recognizing these peculiar sequence features, have been developed. These methods include PONDR®,[60] DisEMBL,[61] DISOPRED,[62] GlobPlot,[63] RQA,[64] IUPred,[65] as well as the methods described by Uversky et al.,[66] Liu et al.,[67] Garbuzynskiy et al.,[68] and Coeytaux and Poupon.[69] In addition, the hydrophobic cluster analysis (HCA)[70] has been reported to be a useful complementary tool for the identification of regions of disorder,[71] and programs such as SEG[72] and CAST,[73] which detect regions of low sequence complexity, can also contribute to the identification of disordered regions.

Besides indications provided by sequence analyses, IDPs can also be recognized by a peculiar biochemical behavior. Specifically, their sequence compositional bias, combined with the lack of a titghly packed hydrophocic core, results in a different response to denaturing conditions and different hydrodynamic properties compared to structured (e.g., ordered) proteins. Thus, information about intrinsic disorder may be retrieved by the analysis of protein conformational stability and by its behavior already during the purification process.

Aberrant Electrophoretic Mobility

Electrophoresis is mostly applied either to determine the molecular mass of proteins or to elucidate the charge difference and/or form of macromolecules. It has been pointed out that due to their unique amino acid compositions, IDPs bind less sodium dodecyl sulphate (SDS) than globular proteins.[2] As a result, they possess abnormal mobility in SDS polyacrylamide gel electrophoresis experiments and their apparent molecular masses determined by this technique are often 1.2–1.8 times higher than those expected from sequence data or measured by mass spectrometry.[74–79]

High Apparent Molecular Mass

IDPs possess abnormal mobility in gel-filtration experiments. IDPs are eluted from gel filtration columns with an apparent molecular mass greater than expected from their sequence. In fact, IDPs have hydrodynamic dimensions larger than their globular counterparts due to their more extended conformation. Because size-exclusion chromatography (SEC) separates proteins by their hydrodynamic sizes rather than by their molecular masses (see also later), one should avoid one of the most common mistakes in the interpretation of SEC data for natively disordered proteins that comes from neglecting this piece of information. As a consequence, one is improperly discussing of "increased molecular mass" instead of making a conclusion about an "increased hydrodynamic volume."

Heat Resistance

The lack of ordered structure in IDPs makes many of them rather insensitive to temperature increase. As a result, such proteins can be easily separated from ordered ones due to their resistance to heat-induced aggregation. This is because of the fact that upon irreversible denaturation, hydrophobic patches, which are normally hidden in the inside of the folded structure, are exposed to the solvent, enter into uncontrolled inter- and intramolecular interactions, and finally form insoluble aggregates. The high content of hydrophilic and charged amino acid residues keeps unstructured proteins soluble even at high temperatures. This property can be exploited to purify disordered proteins.[80]

Acid Resistance

Another important peculiarity of disordered proteins associated with the lack of ordered structure is their stability toward trichloroacetic acid (TCA) and perchloric acid (PCA) treatment. It is known that in ordered proteins the protonation of negatively charged side chains (induced by the decrease in pH) leads to charge imbalances that disrupt salt bridges[46] and causes the disassociation of subunits[81] and cofactors.[82] Very often, the individual protein chains adopt a more open conformation to accommodate the charge imbalance and aggregate. However, many IDPs were shown to resist PCA- or TCA-induced precipitation.[83] Overall, these data suggest that the heat indifference and the ability to resist the PCA- or TCA-induced precipitation represent important features that can be used to purify disordered proteins.[83]

Conformational Stability

The response of a protein to denaturing conditions may be used to discriminate between collapsed (molten globule-

like) and extended intrinsic disorder (random coil-like and premolten globule-like conformations). In fact, the increase in temperature and changes in pH (as well as the increase in urea or guanidium hydrocloride concentration) will induce relatively cooperative loss of the residual ordered structure in molten globule-like disordered proteins, whereas temperature and pH may lead to formation or stabilization of residual structure in native coils and native premolten globules, respectively (see below). Furthermore, it has been shown that the cooperativity of urea- or guanidium hydrocloride-induced unfolding curves for small proteins depends strongly on whether a given protein has a rigid tertiary structure (i.e., it is ordered) or is disordered and exists as a molten globule.[84,85] To extend this type of analysis, the values of Dneff (which is the difference in the number of denaturant molecules "bound" to one protein molecule in its two states) should be determined. Then this quantity should be compared to the DneffN®U and Dneffmolten globule®U values, corresponding to the native to coil and molten globule to coil transitions in globular protein of a given molecular mass, respectively.[85]

It has been established that IDPs are characterized by "turn out" responses to changes in the environment, as their structural complexities increase at high temperature or at extreme pH. In fact, at low temperatures, IDPs 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.[3,16,42] This was observed for such natively disordered proteins as α-synuclein,[86] phosphodiesterase γ-subunit,[87] 636–771 fragment of caldesmon,[20] extracellular domain of the nerve growth factor receptor,[88] and α-casein.[89] Thus, an increase in temperature can induce the partial folding of natively disordered proteins, rather than the unfolding typical of ordered 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.

Several natively disordered proteins with extended disorder, including α-synuclein,[86] prothymosin α,[90] pig calpastatin domain I,[91] histidine rich protein II,[92] naturally occurring human peptide LL-37,[93] core histones,[94] 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 gain some 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.

Beyond the above-described peculiar biochemical behaviors, a few trivial biochemical features could be taken as hallmarks of structural disorder. Specifically, one can suspect that a protein or a protein region is intrinsically disordered if it presents one or several of these properties: it is hypervariable in sequence, it is called "proline rich" or "glycine rich" (because of its unusual composition and of its relative enrichment in disorder promoting residues), it has a very low absorbance at 280 nm (because of its depletion in aromatic residues), it has no or little predicted secondary structure, it is hypersensitive to proteolysis, and, finally, it is known to interact with multiple partners.

Atogether, these features should alert biochemists and prompt them to suspect the presence of disorder, which can then be further explored using the various experimental approaches described below.

EXPERIMENTAL METHODS FOR ASSESSING INTRINSIC DISORDER

In most cases, unique three-dimensional structures of unstructured proteins are not available from crystallographic studies because crystals of conformationally disordered molecules are difficult to obtain and in case they form may not be representative of the conformational ensemble in solution. Recently, attempts to derive high resolution structural information of IDPs have been reported, based on crystallization of such proteins as fusion with GST or in the presence of binding partners or antibodies.[95–97] However, these structures remain representative only of one particular member of the conformational ensemble of the free protein in solution. By contrast, all the methodological approaches that make use of protein samples in solution can give more representative information and are thus better suited for the study of IDPs. In the following paragraphs, we review the various available approaches that are particularly adapted to the study of unfolded (or partly folded) proteins, highlighting the type of information that can be derived.

Assessing the Global Tertiary Structure

There are several physico-chemical approaches that determine the presence or absence of a stable global tertiary structure in a protein molecule.

Differential scanning microcalorimetry (DSC)

Besides providing information on the melting temperature, DSC opens a unique possibility to directly measure the denaturation enthalpy. Another advantage of DSC is that the calorimetric and the van't Hoff (or effective) enthalpies (ΔH^{cal} and $\Delta H^{\mathrm{van't\text{-}Hoff}}$, respectively) can be determined from the analysis of a single calorimetric curve.[98] The former represents the heat required to denature a single protein molecule, whereas the latter is the heat required to melt a cooperative unit. Obviously, for the thermal denaturation of a single-domain protein, the ratio of these two quantities is equal to unity, reflecting the fact that the cooperative unit is a whole protein molecule.[98] On the other hand, the melting of multidomain or multisubunit proteins is characterized by large deviations of this ratio from the unity.[99]

Overall, the presence or absence of the heat absorption peak in the corresponding calorimetric curve represents a simple and convenient criterion indicating the presence or absence of the ordered (thus, rigid and cooperatively

melting) tertiary structure in a given protein.[29,98,100] Such an approach is especially convenient in cases where information on the presence of a specific tertiary structure in the protein molecule cannot be obtained by spectral methods (e.g., when the protein molecule does not contain aromatic amino acid residues or has a disturbed or flexible structure in their nearest spatial surrounding[56,101–103]).

DSC also allows detection of residual structure. Although the absence of any cooperative thermal transition is indicative of the absence of a rigid tertiary structure, typical of disordered proteins, a heat capacity analysis will recognize the presence of residual structure. The difference between the absolute heat capacity of the unfolded and folded state (DCp) is related to the change in solvent-exposed surface area due to the unfolding of polar and nonpolar residues. An elegant example of this method is given in ref. 104, in which the authors studied several fragments of natively disordered thioredoxin by DSC, and thereby measured their hydrophobic buried surface. Interactions could be detected between segments of the proteins.

Use of extrinsic fluorescence compounds

Compounds with extrinsic fluorescent properties, such as ANS (1-anilino-8-naphthalenesulfonic acid), can access to hydrophobic protein cavities and strongly fluoresce when bound to a nonpolar surface. Accordingly, they have been widely used to study protein conformational changes.[105–108] In particular, they can be used to probe the presence (or the appearance under denaturing conditions) of hydrophobic pockets within globular proteins. Although fully extended conformations (random coil-like) are devoid of such pockets, molten globules and premolten globules may possess hydrophobic cavities resulting from residual secondary and tertiary structure. Extrinsic fluorescent compounds are therefore useful to determine the presence of residual hydrophobic core-like structures within IDPs, thus allowing the discrimination between extended forms and molten globules.[109–113]

Limited proteolysis

Flexible protein regions are known to be easy targets for proteolysis. There has been a long-running debate whether proteolysis occurs only in unfolded or disordered regions of proteins or whether proteolysis can occur in surface-exposed, but ordered regions of proteins. There are several arguments supporting the former model. Particularly, if proteolysis is indeed happening at the surface-exposed and ordered sites, then it is unclear how a protease of known sequence specificity recognizes such a limited subset out of the many putative sites of proteolysis in a folded polypeptide chain. In fact, trypsin ought to cleave polypeptide at nearly every lysine-X and arginine-X bond (with the partial exception of proline at X), assuming that about 5–10% of the peptide bonds in a typical protein has to be susceptible to proteolytic attack. However, in native (or near-native) conditions trypsin will cut only a limited number of such bonds (or on occasion none at all) in a native protein fold. This means that the structure and dynamics of the substrate protein is playing a crucial role in limiting the proteolysis.[114,115] For several proteins it has been shown that local unfolding of at least 13 residues is needed for a set of observed cut sites to properly fit into trypsin's active site.[114,115] Furthermore, it has been established that limited proteolytic sites are typically found at flexible loop regions (as indicated by crystallographic temperature factors or B-values) that are also exposed to the solvent,[116–118] and are notably absent in regions of regular secondary structure, especially β-sheets.[114,119] They protrude from the protein surface and are expected to be found at regions where the local packing does not inhibit the local unfolding that is deemed necessary.[116] Fontana and coworkers[120] showed that apomyoglobin is digested many orders of magnitude faster than myoglobin and the sites of digestion from several different proteases all mapped to a region of intrinsic disorder. Similarly, it has been shown that while the holo-form of cytochrome *c* is fully resistant to proteolytic digestion and the apoprotein is digested to small peptides, the noncovalent complex of the apoprotein and heme exhibits an intermediate resistance to proteolysis, in agreement with the fact that the more folded structure of the complex makes the protein substrate more resistant to proteolysis.[121] Thus, limited proteolysis can be used to distinguish ordered and disordered proteins, and an increased proteolytic degradation in vitro of IDPs indirectly confirmed their increased flexibility.[1,10,39,114,122–131]

Immunochemical approaches

Immunochemical methods may also be applied toward the elucidation of protein disorder. Important to this discussion, the immunoglobulins obtained against a given protein may be specific for different levels of macromolecule: the primary structure,[132,133] the secondary structure,[132–134] or the tertiary structure. In the latter case, the antigenic determinants may reside on either the neighboring residues in the chain (loops)[132,133] or on spatially distant residues.[134] Furthermore, it has been shown that antibodies in the immune serum may possess a high affinity to the internal elements of an antigen.[134] Thus, antibodies may be successfully used to study the structural changes that a protein-immunogen undergoes upon modification of the experimental conditions.

Assessing Protein Dimension and Shape

Several methods allow the determination of the size and shape of macromolecules, and hence their degree of extension and of disorder.

Small-angle X-ray scattering (SAXS)

One of the most powerful method for assessing protein dimensions and shape is SAXS. The scattered intensity is sensitive to the size of the protein in solution but also to the conformational properties of the polypeptide chain, and the shape of the molecule can be retrieved ab initio if is well defined. The radius of gyration, R_g, is the root-mean-square distance from the center of gravity of the molecule (weighted in fact by the electron density).[135] It is inferred

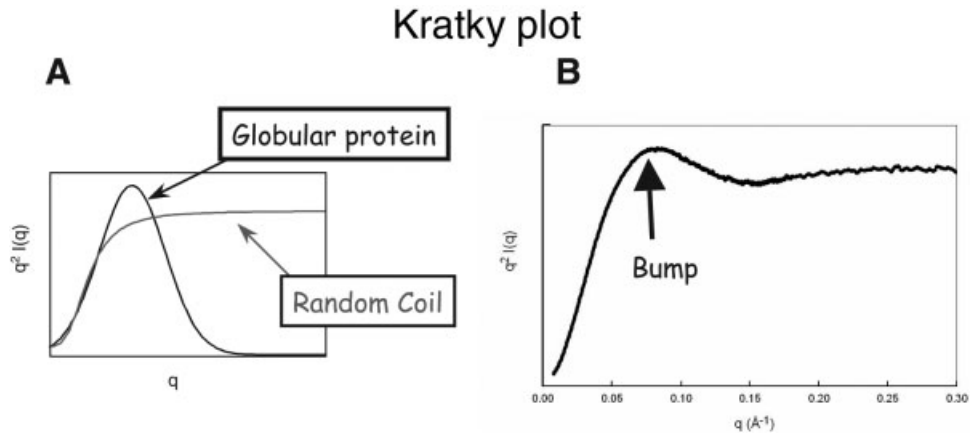

Fig. 1. Kratky plot of scattering spectra obtained from small angle X-ray scattering studies. (**A**) Reference Kratky plots for a globular and an unfolded protein. (**B**) Kratky plot of the scattered intensity of NTAIL, an IDP belonging to the premolten globule subfamily. The arrow highlights the bump at $q = 0.08$ Å^{-1}, which is indicative of the existence of some residual structure. Data for this plot are taken from ref. 78.

from the scattering intensity at low angles with the Guinier approximation:

$$I(q) = I(0)\exp(-q^2R_g^{2/3}) \quad (1)$$

where $I(q)$ is the scattering intensity at the angle q given by the modulus of the scattering vector q [$q = 4\pi\sin(q/2)$]. The forward scattering intensity $I(0)$, yielded by the Guinier approximation, is directly proportional to the molecular mass of the protein, and thus gives information on the protein oligomerization state. One can then interpret an experimental large R_g as the result either of oligomerization or of a monomeric, extended (disordered) conformation. Empirical relationships relating the R_g with the number of residues of native globular protein and of strongly denatured proteins (i.e., random coils) have been established.[136] Comparison between the experimental and expected R_g, allows discrimination between folded proteins, random coils, and premolten globules (e.g., see refs. 43, 78, 112, 113, 137, and 138).

Another important piece of information provided by SAXS is given by the distance distribution function, which is the histogram of all the interatomic distances within the molecule, and which is estimated by Fourier inversion of the scattering intensity. The observed maximum dimension of the protein can then be compared to the values found in the literature for globular and unfolded proteins.[139] Finally, the conformational properties of the polypeptide chain can be assessed with SAXS by representing the scattering pattern in a Kratky plot, that is $q2 \cdot I(q)$ versus q. Although the Kratky plot of a globular protein displays a typical bell shape, the lack of tightly packed core typical of random coils is reflected by a curve with no clear maximum and rather displaying a plateau at large angles[112,140–142] (Fig. 1). This feature readily distinguishes globular conformations (ordered and molten globule-like) from nonglobular ones (random coil- and premolten globule-like). It has been shown that IDPs with extended disorder are characterized by low (coil-like) intramolecular packing density, reflected by the absence of a maximum on their Kratky plots.[20,86,90,94,143] However, it is noteworthy that Kratky plots of premolten globules can give rise to a small bump followed by a plateau, revealing the presence of some residual structure (Fig. 1).

Dynamic light scattering (DLS)

The dimensions of IDPs can also be estimated via their hydrodynamics properties using less demanding techniques, such as DLS and gel filtration (see below), which both enable determination of the hydrodynamic radius, or Stokes radius (RS), of a particle in solution. DLS measures the translational diffusion coefficient of the protein, and its RS is inferred from the Stokes-Einstein equation, which gives the diffusion coefficient of a spherical particle with radius RS in a medium of viscosity η and at temperature T:

$$D = k_BT/6\pi\eta R_S \quad (2)$$

where k_B is Boltzmann's constant.

Gel filtration

In gel filtration (or SEC) the mobility of the particle in a matrix is measured through its elution volume from a prepacked FPLC or HPLC column, which has been calibrated with several globular proteins of known RS as standards.[144] The hydrodynamic properties of a disordered chain are significantly different from those of globular proteins, as early noticed by Kriwacki and coworkers.[125] For example, there is a well-known ~15–20% increase in the hydrodynamic radius of globular proteins corresponding to their transformation into the molten globule state, with hydrodynamic radii of premolten globules being even larger.[84,144] Noteworthy, SEC separates proteins by their hydrodynamic sizes rather than by their molecular masses. This means that for a given protein, the calibrated SEC column will provide information on RS rather than on molecular mass. However, each particular conformation of a polypeptide chain (globular, molten globules, premolten globules, and unfolded conformations) is characterized by a specific dependence of the hydrody-

namic radius on molecular mass, provided that there is no aspecific interaction between the protein and the column matrix. Indeed, as for SAXS, empirical relationships have been established between the RS and molecular masses of native globular proteins, molten globules, premolten globules, and random coils.[3,16,42,43,145] Using these equations one can compare the observed RS with the expected values and evaluate the degree of compactness of the protein. Furthermore, comparison of the R_g with the RS also allows discrimination between random coils and premolten globules. It has been shown that the ratio R_g/R_S should be $(3/5)^{1/2}$ for globular proteins, close to 1.0 for premolten globules and about 1.5 for random coils.[146]

Analytical ultracentrifugation (AU), and in particular sedimentation velocity, provides information about the hydrodynamic properties of a molecule, and thus its size and shape. Conversely, sedimentation equilibrium provides thermodynamic information about the solution molecular mass (but not the shape), association constants, stoichiometries, and solution nonideality.[147] Basically, these two complementary views of solution behavior of a macromolecule can be obtained using the same physical principle (the measurement of the concentration as a function of radial position) and the same instrument, but employing different experimental protocols.

In a sedimentation velocity experiment, a moving boundary, separating cushions of protein solutions, is formed on application of a strong centrifugal field. Analysis of the speed with which the protein molecules move toward the bottom (outermost boundary) of the cell, gives the apparent sedimentation coefficient, s^*, which can be related to the molecular weight and to the frictional coefficient, or shape of a particle. Besides the apparent sedimentation coefficient, the analysis of sedimentograms can provide estimates of the translational diffusion coefficient, D. This information is of great use, as it allows the determination of the molecular mass of dissolved macromolecules using the Svedberg equation:

$$s/D = M(1 - \rho\nu)/RT \quad (3)$$

where M is the molecular mass of the macromolecule, ν is the solute's partial specific volume, ρ is the solvent density, R is the universal gas constant, and T is the absolute temperature. Thus, information on the molecular mass, shape, and hydrodynamic dimension of the protein can be retrieved from the sedimentation velocity analysis. This method is particularly well suited for rapid estimates of the molecular weight and for determining particle size distributions in a sample.

The sedimentation equilibrium is characterized by the steady-state concentration gradient that develops as the molecule's tendency to sediment in the centrifugal field is counterbalanced by its tendency to diffuse against the concentration gradient so established. In this case a smooth gradient is seen instead of formation of distinct boundaries. Hydrodynamics affect the time needed to reach equilibrium, and thermodynamics affect the equilibrium concentration gradient. Different approaches have been used to derive the equations describing sedimentation equilibrium.[148–151] They can yield valuable thermodynamic and stoichiometric information on the interaction between molecules forming complex structures: the monomer molecular mass, the complex molecular mas, the stoichiometry of heterogeneous components, the strength of interactions between components (interaction constants), and the thermodynamic non-ideality of the solution (virial coefficients).

Viscometry

Viscosity is the degree to which a fluid resists flow under an applied force. It describes the internal friction of a moving fluid and thus depends on the diffusion of the dissolved macromolecules. Macromolecules increase the viscosity of a solution already at very low concentrations. The increase of the viscosity is not only dependent on the molecular mass but also on the shape of the macromolecules. In fact, the value of the intrinsic viscosity, $[\eta]$, provides a further measure of the conformation (and hydration) of the macromolecule in solution. The dependence of $[\eta]$ of a spherical molecule on its Stokes radius, R_S, was established by Einstein and further developed by Tanford:[152]

$$[\eta] = (2.5N_A/M)(4/3\pi R_S^3) \quad (4)$$

where N_A is the Avogadro's number and M is the molecular mass. For polymer molecules the relationship between $[\eta]$ and M is expressed by the Mark-Houwink-equation:

$$[\eta] = KM^a \quad (5)$$

where K and a are constants, depending on the solvent. For proteins unfolded in 6 M guanidinium hydrochloride these constants are well known from numerous publications and are $K = 0.716$ and $a = 0.66$ ($[\eta]$ expressed in g/cm^3). Thus, the viscosity of unfolded proteins depends on their molecular mass. By contrast, native globular proteins are characterized by a relatively constant viscosity ($[\eta] \sim 3$ g/cm^3) independent of their molecular mass. Determination of the viscosity of a protein can thus inform on its folded/unfolded state. In particular, it was shown that molten globules can be almost just as compact as the native conformations of the respective proteins.[153–156]

Pulsed field gradient NMR experiments also gives access to hydrodynamic parameters of proteins in solution, although this technique is less frequently used than those above-mentioned.[145] A gradient is generated across the sample, thus allowing measurement of translational diffusion coefficients. Hence, the hydrodynamic radius and the molecular mass of the protein can be estimated. Discrepancies between the experimentally measured radius and the theoretical radius expected for a globular protein, are indicative of an extended conformation, consistent with the disordered state.

Estimation of the Secondary Structure Content

Estimation of the secondary structure content allows molten globules to be distinguished from random coils and premolten globules.

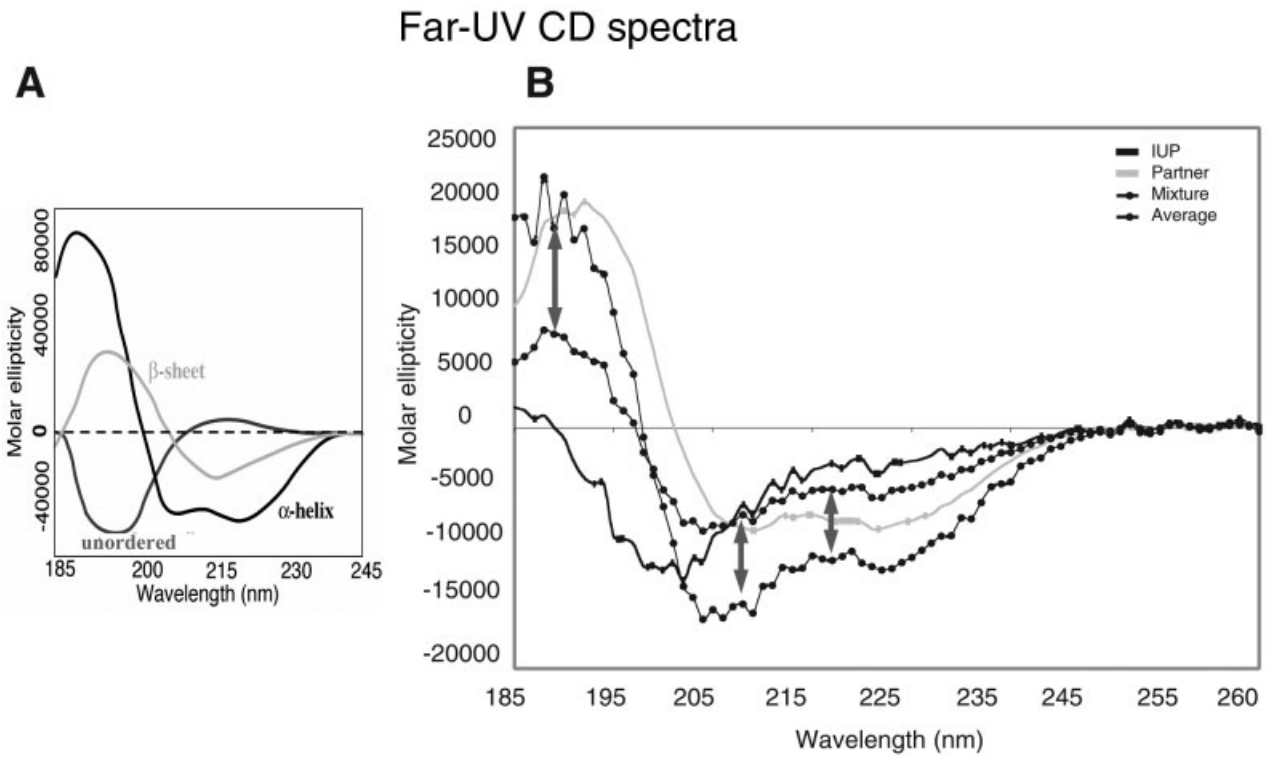

Fig. 2. (**A**) Far-UV CD reference spectra of an unordered protein, and of proteins with a prevalent α-helical or β content. (**B**) Induced folding of the intrinsically unstructured C-terminal domain (aa 401–525) of measles virus nucleoprotein (N_{TAIL}) in the presence of its partner (X domain of measles virus phosphoprotein, aa 459–507) assessed by far-UV CD spectroscopy. CD spectra of the IDP, of the folded partner, and of a 1:2 stoechiometric mixture. The theoretical average curve, derived by averaging the two individual spectra and thus corresponding to the expected curve in case no structural transition takes place, is also shown. The arrows highlight deviations from the average curve that are indicative of a random coil to α-helix transition. Data are taken from ref. 227.

Far-UV circular dichroism (CD) is a spectroscopic technique that allows the estimation of the secondary structure content of a protein in solution, and requires protein amounts in the microgram range. It relies on the interaction of circularly polarized light with optically active (chiral) compounds, as it is the case for proteins because of the peptide bonds. There are two types of optically active chromophores in proteins, side groups of aromatic amino acid residues, and peptide bonds.[157,158] CD spectra in the far-UV region (peptide part of the spectrum, 180–250 nm) reflect the symmetry of the peptide bond environment and indicate the content of secondary structure in the protein molecule. In the far-UV region, the different forms of regular secondary structure found in proteins give rise to characteristic CD spectra. The CD spectrum of an IDP is typified by a largely negative ellipticity near 200 nm, a negligible ellipticity at 222 nm, and by an ellipticity close to zero at 185 nm [Fig. 2(A)]. The CD spectra of many unordered polypeptides show a remarkable resemblance to that of poly(L-proline) II conformation (PII), as judged by their more pronounced minima at 200 nm compared to random coils (-5000 deg cm^2 dmol^{-1}) [see Fig. 2(A)]. Accordingly, it has been suggested that many unordered peptides have significant amounts (up to 50%) of PII helix, in the form of short helical regions.[21] Other unordered polypetides also shows a strong negative band below 200 nm, but the long wavelength region (i.e., 220–230 nm) shows a negative shoulder rather than a positive band. This type of spectrum reflects a higher extent of disorder compared to that of PII-like spectra. As reported by Uversky,[3] far-UV CD also enables discrimination between random coils and premolten globules, based on the ratio of the ellipticity values at 200 and 222 nm.

Fourier transform infrared spectroscopy (FTIR) also provides quantitative information on the secondary structures motifs of proteins through decomposition of the Amide I absorption band (1700–1600 cm^{-1} region) into its components via curve-fitting procedures.[159] This band originates from the C═O stretching vibration of the peptide group, the frequency of which is sensitive to the protein conformation. In contrast to CD, FTIR enables monitoring protein aggregation and discrimination of helices with different extents of flexibility.[160] However it requires larger protein amounts (microgram range) than CD. Notably, FTIR is one of the few techniques that allow study of dry proteins (protein films).

Raman optical activity (ROA)

Another method sensitive to the secondary structure of proteins is ROA, which measures vibrational optical activity by means of a small difference in the intensity of Raman scattering from chiral molecules in right and left circularly polarized incident laser light. The ROA spectra of a wide range of biomolecules in aqueous solution can

now be measured routinely. Because of its sensitivity to the chiral elements of biomolecular structure, ROA provides new information about structure and dynamics in solution complementary to that supplied by conventional spectroscopic techniques.[161] Also, ROA is well suited to measure the extent of PII conformation,[162] a conformation that could be dominant in unfolded peptides and proteins.[21]

Deep-UV resonance Raman spectroscopy (DUVRR)

Raman spectroscopy and particularly resonance Raman spectroscopy is widely used for structural analysis of biological systems.[163,164] Until quite recently, the application of this technique was limited to characterization of heme proteins via the excitation on heme vibrational modes.[165] Recent achievements in the DUVRR development opened the possibility to retrieve direct quantitative information about the protein secondary structure,[166,167] thus rendering this technique promising for studying IDPs. This new method measures deep-UV excitation enhancement of Raman scattering from the amide chromophore.[168,169]

Nuclear magnetic resonance (NMR)

The secondary structure content of proteins can also be assessed by NMR spectroscopy. NMR measures the magnetic properties of some atom nuclei (including ^{1}H, ^{15}N, ^{13}C), which possess a spin, that is, they can precess around a magnetic field. The absorption of energy by nuclear spins results in an excited state. NMR experiments follow the return to equilibrium (resonance phenomenon). A nuclear spin creates its own magnetic field. The resonance frequency reflects the local magnetic field, and therefore, the local chemical environment, and hence, the secondary structure. Thus, several parameters can be derived through various experiments, including ^{1}H, ^{15}N, and ^{13}C chemical shift values, ^{1}H–^{1}H and ^{1}H–^{15}N NOE (Nuclear Overhauser Effect) values, and hydrogen/deuterium exchange rates. The different NMR methods, which provide different structural insights into IDPs, will be briefly described below highlighting their specific applications.

NMR has the unique property of providing atomic resolution, and thus residue-specific, information of macromolecules in solution. For IDPs, the inherent flexibility of the polypeptide chain and the rapid interconversion between multiple conformations results in poor chemical shift dispersion of most resonances, especially of protons. Amino acid residues within IDPs share a similar chemical environment, being all prevalently exposed to the solvent. Consequently, they share similar NMR frequencies, thus resulting in a strong overlap of the ^{1}H resonances. Therefore, while bidimensional NMR spectra of folded proteins display a broad distribution of NMR frequencies, NMR spectra of IDPs are typified by a very low spread of the resonance frequencies of amide protons.[170,171] This feature renders the characterization of disordered states or proteins particularly challenging. Since the earliest report of NMR observation of residual structure in a highly denatured protein appeared in 1992,[172] the advances in multidimensional NMR technology and the development of methods for uniform labeling of proteins with ^{15}N and ^{13}C have enabled studies of unfolded states of full-length proteins. HSQC (Heteronuclear Single Quantum Correlation) spectra of uniformly ^{15}N,^{13}C-labeled proteins can, in fact, provide residue and atom specific structural information: indeed, in 3D triple resonance experiments, the chemical shifts of the ^{15}N and ^{13}CO resonances are well dispersed even in fully unfolded states, thus allowing unambiguous resonance assignments. Among the first reports on the use of multidimensional NMR technology in studying IDPs, we mention the work from the Kay and Shortle group.[173–181] The application of NMR to characterize IDPs has been the topic of a few excellent recent reviews (see refs. 27, 170, 171, and 182). Briefly, when exchange between folded and unfolded states occurs on an appropriate time scale, at least backbone resonance assignments can often be made. These assignments make use of homonuclear ^{1}H magnetization transfer methods to correlate resonances in the spectrum of the folded protein with the corresponding resonances in the denatured state. An assignment strategy that utilizes the superior chemical shift dispersion of the ^{13}CO resonance in unfolded proteins is especially valuable, given the excellent resolution and long T_2 relaxation time of these resonances. A set of high-resolution constant-time triple resonance experiments that transfer magnetization sequentially along the amino acid sequence using carbonyl ^{13}C homonuclear isotropic mixing, have been developed specifically for assignment of unfolded proteins. These experiments have the advantage that the slow ^{13}CO relaxation rates enable correlations to be established across proline residues.

One of the advantages of working with unfolded proteins is that the intrinsic flexibility of the polypeptide generally causes the resonances to be much narrower than they would be in a folded protein of comparable molecular weight, that is, T_2 is longer than for the natively folded protein. This characteristic gives special advantages in the implementation of pulse sequences characterized by a large number of delays. For folded proteins of molecular mass greater than 15–20 kDa, this experiment is often rather insensitive because of decay of magnetization during the pulse sequence. For unfolded proteins, however, because T_2 is significantly longer, this type of experiment is extremely sensitive and can therefore be successfully used. Unfortunately, this is not the case for partly folded proteins or molten globules that give rise to extremely broad resonances that preclude atomic resolution NMR analysis.

Once backbone resonance assignments have been made, various NMR parameters can be used to characterize residual structure in unfolded and partly folded states. However, it should be pointed out that these parameters are a population-weighted average over all structures in the conformational ensemble. Conformational preferences are identified by comparison of experimental NMR parameters to those expected for a random coil state. Thus, deviations of the chemical shifts from random coil values, especially for $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{1}H^{\alpha}$, can be used to calculate

the relative population of dihedral angles in the α or β regions (for reviews, see refs. 170, 171, and 182). Notably, secondary $^{13}C^{\alpha}$ chemical shift values ($\delta\Delta^{13}C^{\alpha}$) have been successfully used to identify residues exhibiting residual α-helical structure (e.g., see refs. 24 and 183). Similarly, abnormally low amide proton temperature coefficients, relative to random coil values, and low hydrogen exchange rates (see next section) are hallmarks of residual structure within IDPs (for reviews, see refs. 170, 171, and 182). Medium-range NOEs indicative of helical or turn-like structure have been observed in peptides and unfolded proteins. However, long-range NOEs, indicative of transient tertiary structure, are very difficult to observe in IDPs. This is likely due to the fact that either the population of the transiently structured forms is too low, or the ensemble containing them is too heterogeneous. For IDPs, long range distance information can rather be derived by the use of covalently attached paramagnetic nitroxide spin labels (see next section). Distance information is obtained by assessing the broadening effect of the unique spin label between the spectra of the reduced (diamagnetic) and oxidized (paramagnetic) samples, as well illustrated in the case of apomyoglobin.[184] Alternatively, by inserting spin labels at multiple sites, a sufficient number of long-range distance constraints can be obtained to allow determination of the global topology.[185,186] Estimates of backbone and side-chain dynamics can be derived from measurements of relaxation data (i.e., T_1 and T_2 and heteronuclear NOE for backbone resonance in uniformly ^{15}N-labeled proteins), where ^{1}H–^{15}N heteronuclear NOE values, which indicate backbone amide dynamics on the nanosecond time scale, provide indications about residual secondary structure (see ref. 24). Recent advances include R2 relaxation dispersion and NMR relaxation coupled with MD simulation (for a review, see ref. 182). Finally, we would like to mention a few recently developed NMR methods that might be very well suited for studying IDPs, namely dipolar couplings in partially aligned media, NMR under pressure, and real-time NMR (reviewed in ref. 182).

Assessing Local Tertiary Structure

Assessing local tertiary structure can provide crucial information on the location of residual structures that could represent possible starting sites for induced folding.

Circular dischroism spectra in near ultraviolet region (near-UV CD). CD spectra in the near-UV (250–350 nm), also called the aromatic region, reflects the symmetry of the environment of aromatic amino acid residues and, consequently, characterize the tertiary structure of an ordered protein. Each aromatic residue tends to have its own wavelength profile. Signals in the 250–270-nm region are attributable to phenylalanine residues, while signals in the 270–290 and 280–300-nm regions are attributable to tyrosine and tryptophan residues, respectively. Disulfide bonds give rise to broad weak signals throughout the near-UV spectrum.[187] The near-UV CD spectrum therefore provides a detailed fingerprint of the tertiary structure around these specific residues. 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. On the other hand, the lack of ordered structure in a protein containing aromatic residues may be easily detected by near-UV CD. In fact, diminished near-UV CD spectrum with low intensity and low complexity was shown to be a characteristic feature of many partially folded proteins. However, in some cases IDPs can possess rather complex near-UV CD spectra, suggesting that their aromatic amino acid residues are in relatively asymmetric environment; that is, these proteins contains some residual tertiary structure in these regions (e.g., natively unfolded caldesmon; ref. 20).

Fluorescence spectroscopy is a widely used technique to study protein structural transitions. Proteins contain only three intrinsic fluorescent chromophores, which, according to their quantum yield, form the following series: tryptophan > tyrosine > phenylalanine. The most commonly used in practice is the fluorescence of tryptophan, because the quantum yield of phenylalanine fluorescence is extremely low and tyrosine fluorescence is strongly quenched in the majority of cases due to its ionization or location near the amide or carboxyl groups, or due to the energy transfer to tryptophan.[188] Tryptophan has an absorption maximum close to 280 nm and an emission maximum that is highly dependent on the polarity of its environment. In particular, a tryptophan in a non polar environment (i.e., buried inside a protein), has an emission maximum close to 320 nm, while in a polar environment (i.e., solvent exposed) it has an emission maximum close to 350 nm, so-called Stokes shift.[188,189] This means that the value of λ_{max}, that is, the wavelength of maximum emission, can give some basic information as to whether the given protein is compact under the given conditions. Moreover, the intensity of tryptophan fluorescence is dependent on interactions with neighboring groups, which can lead to sharing and transfer of the excitation energy.[190] This phenomenon results in a consequent reduction in the quantum yield and diminution of fluorescence intensity, a phenomenon known as quenching of the fluorescence. Therefore, the fluorescence spectrum of an IDP provides useful information on the environment of the fluorophore, and hence on the presence of residual ordered protein structure in its vicinity.[191]

Dynamic quenching of fluorescence

The intensity of tryptophan and tyrosine fluorescence can be modulated by different compounds, as for instance iodide, cesium ions, or succinimide.[192] These compounds allow the chromophore environment to be probed. The presence of these molecules in physical contact with an excited chromophore can lead to sharing the excitation energy, with a consequent reduction of the fluorescence intensity, a phenomenon named fluorescence quenching. There are two different types of quencher: polar and nonpolar. Thus, they can be used to differentiate between buried and solvent-accessible tryptophans.[109] Furthermore, some neutral quenchers, like acrylamide, have the ability to interact with both exposed and buried trypto-

phans. Analysis of dynamic quenching of intrinsic fluorescence by small molecules can provide additional information on the accessibility of protein chromophores to solvent (and, thus on relative compactness of a protein molecule). Fluorescence quenching data are frequently analyzed using the general form of the Stern-Volmer equation:[192]

$$\frac{I_0}{I} = (1 + K_{SV}[Q])e^{V[Q]} \quad (6)$$

where I_0 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. Overall, the information retrived by the experiments on dynamic quenching of intrinsic fluorescence is to some extend close to that from studies on deuterium exchange, because it reflects the accessibility of defined protein groups to molecules of quencher contained in the solvent. However, in distinction from the deuterium exchange, this method evaluates the amplitude and time scale of dynamic processes using the quencher of different sizes, polarity and charge.

Fluorescence resonance energy transfer (FRET)

Fluorescence labels are indispensable tools in studies on energy transfer between two chromophores. The essence of the phenomenon is that, at interaction of oscillators at a small distance, the electromagnetic field of the exited oscillator can induce oscillation with the same frequency in the nonexcited oscillator.[188,193] The transfer of excitation energy between the donor and the acceptor originates only with the fulfillment of several conditions: (1) the absorption (excitation) spectrum of the acceptor overlaps with the emission (luminescence) spectrum of the donor. This is an important prerequisite for resonance; (2) a sufficient spatial proximity of the donor and the acceptor is necessary; they must be at distance not exceeding a few dozens of Angstroms; (3) a sufficiently high quantum yield of the donor is also necessary; (4) spatial orientation of donor and acceptor also plays an important role for the effective energy transfer. The biggest advantage and attractiveness of FRET are that this method could be used as molecular ruler to measure distances between the donor and acceptor. In fact, according to Förster, the efficiency of energy transfer, E, from the excited donor, D, to the nonexcited acceptor, A, located from the D at a distance R_{DA} is determined by an equation:[193]

$$E = \frac{1}{1 + \left(\frac{R_{DA}}{R_o}\right)^6} \quad (7)$$

where R_o is the characteristic donor–acceptor distance, so-called Förster distance. Usually, in FRET experiments, one uses intrinsic chromophores (tyrosines or tryptophanes as donors and covalently attached chromophores emitting light in visible region) as acceptors. As it follows from Equation (7), 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 the changes of this distance, giving rise to the considerable changes of this parameter.

An elegant approach based on the unique spectroscopic properties of nitrated tyrosine (which has maximal absorbance in the vicinity of 350 nm, does not emit light and renders as an acceptor for Trp electronic energy) has been recently elaborated.[43,194–197] For these experiments Tyr residues have to be modified by reaction with tetranitromethane to convert them to a nitro form, Tyr(NO_2). The extent of decrease of Trp fluorescence in the presence of Tyr(NO_2) provides a measure of average distance R_{DA} between these residues. This decrease reflects variations in the distance between these residues as resulting from conformational changes.

Fluorescence polarization and anisotropy

Important information about mobility and aggregation state of macromolecules in solution can be retrieved from the analysis of fluorescence polarization or anisotropy. If excited light is polarized, fluorescence will be only partially polarized or completely depolarized. The degree of fluorescence depolarization results from the following factors characterizing the structural state of the protein molecule: (1) mobility of the chromophore, which strongly depends on the density of its environment, and (2) energy transfer between similar chromophores.[40,188,198–203] Furthermore, relaxation times of tryptophan residues determined from polarized luminescence data are a reliable test for compactness of the polypetide chain. For example, the retention of intact disulfide bonds in unfolded proteins containing disulfide bonds was shown to result in a nonessential decrease of the intrinsic fluorescence polarization, whereas the reduction of the SS-bridges leads to a dramatic decrease in fluorescence polarization.[40,201] Relaxation times of tryptophan residues determined by fluorescence polarization for α-lactalbumin[153,154] and bovine carbonic anhydrase B[156] showed high degree of protein compactness both in the ordered and the molten globule states.

Electron paramagnetic resonance (EPR)

Another powerful and sensitive technique to probe protein structure is provided by EPR spectroscopy. This method has the advantage of requiring rather low amounts of proteins (typically nanomoles). This technique is based on the covalent modification of a cysteine side chain to yield a nitroxide side chain that possesses an unpaired electron. Because the mobility of a spin label covalently attached to a protein is influenced by its environment, analysis of its EPR spectra provides insights into the local protein structure,[204] and into the equilibria between ordered and disordered conformations within poorly structured proteins.[205]

Electrospray-ionization mass spectrometry (ESI-MS)

A valuable tool to monitor protein conformational transitions has also proven to be ESI-MS. In particular, it

enables the detection of poorly populated states, thus allowing analysis of dynamic equilibria involving different conformational states.[206,207] This technique generates multiple protonated protein ions in the gas phase, where the number of actual protein charges progressively increases as the protein unfolds.[208] Time-resolved ESI-MS allows the detailed study of folding kinetics and mechanism, and can also be used in conjunction with proton–deuterium exchange methods (see below).[209,210]

The hydrogen-exchange technique gives information on the accessibility of hydrogens. The accessibility of amide hydrogens is largely determined by the presence of backbone hydrogen bonds. Thus, this technique is sensitive both to local tertiary interactions and to the formation of secondary structure elements. Therefore, a good indication of the extent of protein flexibility can be derived through measurement of the rate of exchange of amide hydrogens in D_2O. Amide protons located in α-helices, β-sheets, and in the hydrophobic core of a protein (where access to solvent is limited) exchange slowly, in contrast to hydrogens located in "floppy loops" on the protein surface, which exchange much faster.[211] Therefore, fast exchange rates reflect a high extent of accessibility, which in turn, is indicative of the lack of an ordered structure. NMR can give information about the rates of exchange of both amide and side-chain hydrogens, provided that these have been assigned in the spectrum.[212] This is possible because ^2H nuclei do not give rise to a signal, in contrast to ^1H. Mass spectrometry (MS) is an alternate and convenient approach to monitor proton–deuterium exchange, where folding processes of proteins can be followed by measuring the gradually increasing number of protected hydrogens.[213] FTIR spectroscopy can also be used to monitor hydrogen–deuterium exchange by following variations of the vibrational frequency of the carbonyl C═O group of the peptide bonds involved in hydrogen bonds.

All these techniques highlight the presence of residual structure, thus enabling not only premolten globules to be discriminated from random coils, but also to localize these residual structures, whenever evidences pointing out the presence of partially buried aromatic residues, of low-exchanging protons, and of partially protonated forms are provided.

ASSESSING INDUCED FOLDING

Assessing Folding Propensities

Before addressing induced folding in the presence of a partner, one can assess the folding propensity of an IDP. Secondary structure stabilizers, such as trifluoroethanol (TFE), are widely used as a probe to identify protein regions with propensity to fold.[77,78,214] Trimethylamine N-oxide (TMAO) can also be used to this endeavor.[215] TMAO and other osmolytes may fold unstructured proteins due to the osmophobic effect, a solvophobic thermodynamic force, arising from the unfavorable interaction between the osmolyte and the peptide backbone.[216–219] Although both solutes, TMAO and TFE, act on the peptide backbone, the molecular mechanisms underlying their effects are different. It has long been known that TFE increases the propensity of amino acids to form an α-helix, presumably by strengthening peptide hydrogen bonds in TFE/H_2O mixtures and through favorable interactions of hydrophobic amino acid side chains with TFE.[220,221] Peptide hydrogen bonds in helices are believed to be stabilized indirectly by weakening the hydrogen bonding of water molecules to the peptide backbone in the coil form.[221] As a result of weakening the hydrophobic interactions within the protein interior, TFE might promote helical structure in most peptides and proteins, even though this helical structure is nonnative.[222–224] In contrast, TMAO increases the driving forces for protein folding due to its solvophobic effect on the backbone, forcing thermodynamically unstable proteins to fold without altering the rules for folding to a native-like conformation.[217] Furthermore, in opposition to TFE solution, the propensities of hydrophobic groups to interact with solvent are essentially the same in water as they are in TMAO solution.[56,225] Thus, due to the weakening of hydrophobic interactions, the dominant effect of TFE on proteins is protein denaturation accompanied by the preferential formation of α-helices as a result of the strengthening of peptide hydrogen bonds. Contrarily to that, TMAO promotes folding of unfolded proteins by providing an additional force for folding that has no preference for any particular secondary structure.[219] Based on the molecular origin of TMAO-driven protein folding, if biologically relevant structure can be induced in any intrinsically unstructured protein without its target molecule, it is more likely to be induced by solutes (such as TMAO) that have been selected by nature for their ability to fold and stabilize proteins than by alcohols.[218]

It should be pointed out, however, that even if TFE is known to stabilize α-helices more than β-strands, some proteins, as for instance the GCN4 acidic activation domain,[226] form little or no α-helix in TFE concentrations as high as 30% and fold as β-sheets at higher TFE concentrations. This observation points the ability of TFE to show β propensities. Nevertheless, the general reliability of information derived from CD or FTIR studies that make use of TFE is still a matter of debate. Indeed, no systematic study attempting to validate or to refute this approach has been undertaken so far. Definitive conclusions await studies focused on structural comparisons between TFE-induced structures and structures arising from induced folding.

Assessing Folding Induced by a Partner

When the folding partner is known, induced folding can be detected using any of the different spectroscopic methods described above. In the case of most spectroscopic techniques, such as far-UV CD, FTIR, or SAXS, deviations of the experimentally measured signal from the theoretical average spectrum indicate a structural transition[20,78,227] [Fig. 2(B)], whereas with fluorescence spectroscopy or near-UV CD, gain of structure of the IDP can be monitored through progressive burying of aromatic residues and changes in the fluorescence intensity[20,228] (Fig. 3). The major difficulty lies in extracting the signal of the folded IDP from that arising from the partner and from the uncomplexed form of the IDP that can remain unbound.

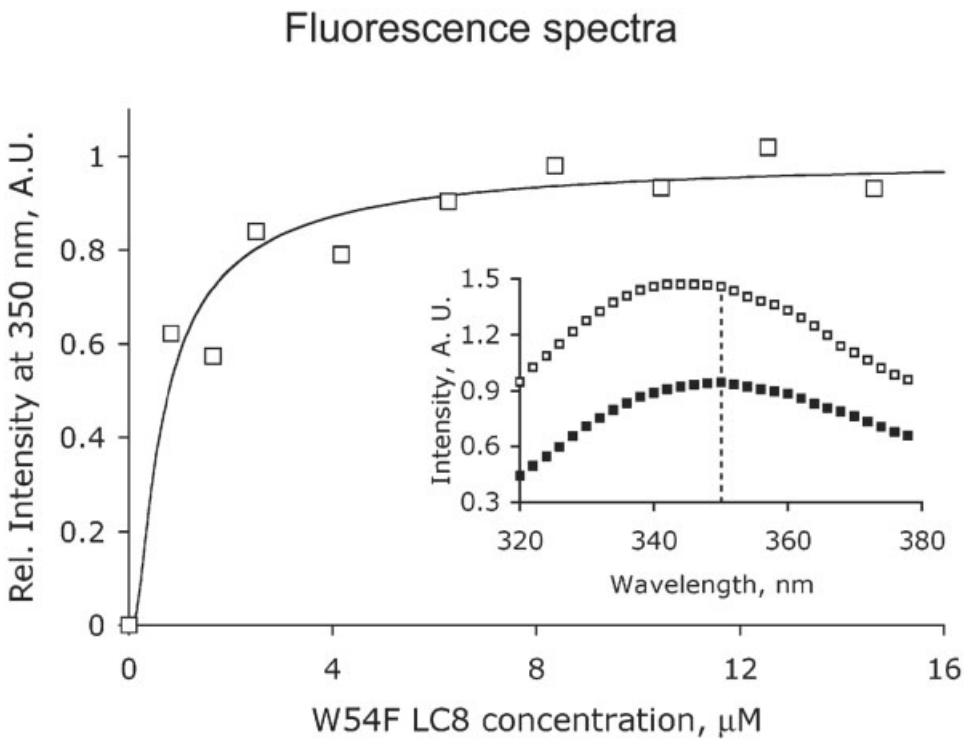

Fig. 3. Titration of the intrinsically disordered N-terminal domain of dynein intermediate chain ($IC^{114-260}$) with an increasing concentration of dynein light chain (LC8) bearing the W54F substitution, monitored by relative fluorescence intensity at 350 nm. Fluorescence intensity was normalized to 0 in the absence of LC8 and to 1 in the presence of a threefold molar excess of this latter. The inset shows fluorescence emission spectra of free $IC^{114-260}$ (full squares) and a 2:1 mixture of LC8/ $IC^{114-260}$ (open squares). The increase in intensity and the shift in λ_{max} from 350 to 343 nm indicate an increase in the compactness of $IC^{114-260}$. Reproduced from ref. 228 with permission of Elisar Barbar and the American Chemical Society.

An elegant method is the singular value decomposition, which allows estimation of the contribution of each component to the observed signal.[139]

Heteronuclear NMR

Although much more demanding in terms of protein amounts (micromolar range) compared to CD, FTIR, EPR, ESI-MS, and fluorescence spectroscopy, heteronuclear NMR has, however, the advantage to provide a more comprehensive estimation of induced folding (including the number of residues involved in the structural transition). It is noteworthy that although such information can be obtained without making full sequence assignment or structure calculation, NMR is the only technique that has the potentiality of providing atom specific information of IDPs in solution. Thus, heteronuclear NMR studies have been used to monitor folding on binding events as early as in 1996,[10] because they have been extensively used to study induced folding (see ref. 182 and references therein cited). Briefly, the IDP is isotopically labeled with ^{15}N, and an HSQC spectrum is recorded in the absence of the partner. In this type of experiments, only correlations between the amide nitrogens and their geminal proton are monitored, and shift of these N—H peaks after addition of the partner indicates the occurrence of structural transitions (Fig. 4). It is possible to estimate whether α or β transitions are concerned, and to quantify the number of residues involved. If the protein undergoes considerable structural transition, then structure determination can be envisaged (for a list of such protein complexes see refs. 3, 8, 13, 18, and 66).

ESI-MS

This technique provides an alternate convenient approach to monitor induced folding,[229,230] where the gain of structure is reflected by a decrease in the number of actual IDP charges (Fig. 5). This approach allows the characterization of the species distributions in heterogeneous samples, thus leading to the detection of possible folding intermediates (Fig. 5).

Furthermore, it has to be noted that quench-flow hydrogen exchange methods, detected by both NMR and MS, remain the basis for most kinetics studies of protein folding, and are thus very well suited for the study of induced folding events.

Isothermal titration calorimetry (ITC)

Another method that establishes the thermodynamic parameters of the interaction between two partners is ITC. ITC measures the heat produced by the binding from which the binding constant, the enthalpy of the reaction

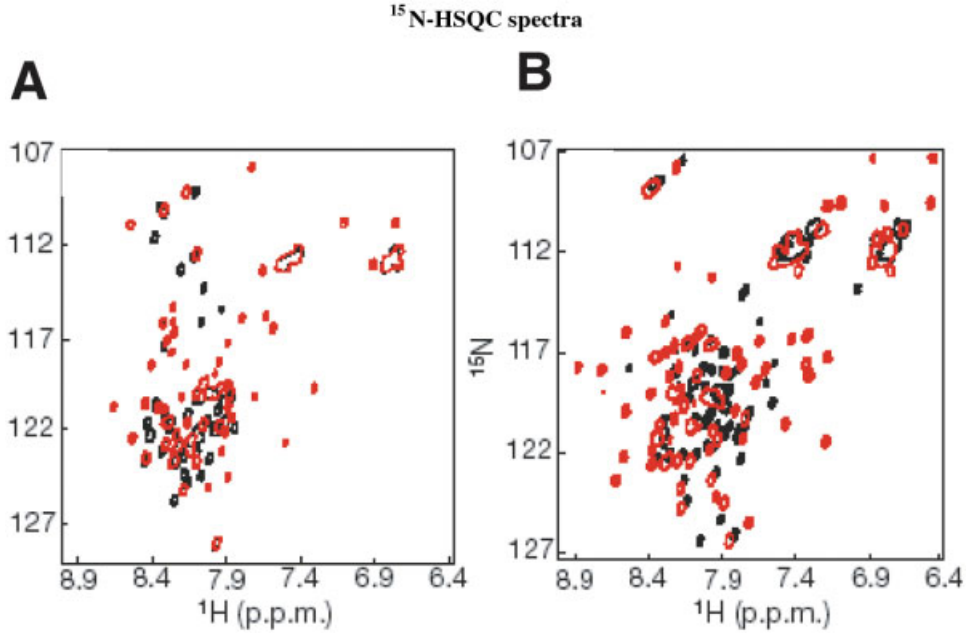

Fig. 4. ^{15}N-HSQC spectra of (**A**) the human ^{15}N-ACTR (activator for thyroid hormone and retinoid receptors) interaction domain (residues 1018–1088) free (black) and in the presence of unlabeled CBP (cAMP-responsive binding protein) (interaction domain, residues 2059–2117) (red). (**B**) ^{15}N-CBP (aa 2059–2152) free (black) and in the presence of unlabeled ACTR (1018–1088) (red). Both ^{1}H-^{15}N spectra show gain of significant amide proton dispersion upon formation of the complex, indicative of the synergistic folding of these two intrinsically unstructured protein domains. Reproduced from ref. 242 with permission of Peter E. Wright and Nature Publishing Group (http://www.nature.com). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

and the heat capacity change can be inferred. This last parameter reflects the polar/nonpolar nature of the surface involved in the binding and can bring valuable information about the binding mechanisms of the IDP to its partner.[231,232] However, in the case of induced folding events, interpretation of the results is generally biased by the fact that the measured ΔH accounts not only for binding but also for folding. The heat capacity change and the variation in entropy also allow estimation of the number of residues involved in folding, as well illustrated in the case of p27 upon binding to the cyclin–CDK complexes.[24] Furthermore, thermodynamic studies were shown to be extremely powerful for investigating protein folding coupled to DNA binding.[233] In particular, determination of the entropy change ($\Delta S°_{ass}$) allows discrimination between "rigid body" associations and binding events involving coupled folding transitions.[233] Furthermore, analysis of the heat capacity change ($\Delta C°_{ass}$) during site-specific DNA recognition by proteins lead the authors to propose that the nature of the structural transition is dictated by the DNA sequence.[233]

Surface plasmon resonance (SPR)

Further, it has recently been reported that SPR can be successfully used to monitor processes of protein folding and unfolding, as the SPR signal variations reflect the conformational changes of an immobilized protein.[234] Therefore, this technique should provide a useful tool to assess intrinsic structural disorder and induced folding in the presence of a ligand partner, provided that one can set up experimental conditions allowing to neglect the contribution of the ligand.

EPR spectroscopy

Processes of protein folding and unfolding can also be successfully monitored by EPR spectroscopy, as variations in the EPR lineshape reflect conformational changes of the spin labeled-derivatized protein.[235] Therefore, this technique should, in principle, provide a useful tool to assess intrinsic structural disorder and induced folding in the presence of a ligand partner, provided that the nitroxide side chain does not impair formation of the complex.

Furthermore, induced folding, or more generally any binding event resulting in a selection of a given conformer, can also be detected by any method able to reveal changes in (1) protein conformational stability, (2) mode of IDP interaction with antibodies or chaperones, and (3) in susceptibility to proteolytic attack, and (4) in hydrodynamic properties and shape.

IDENTIFICATION OF REGIONS UNDERGOING INDUCES FOLDING

Identification of regions undergoing induced folding is the ultimate goal of the structural study of an IDP. Such regions can be identified using both computational and biochemical approaches. The group of Dunker currently develops a program for the identification of Molecular Recognition Elements (MoREs), which are regions within

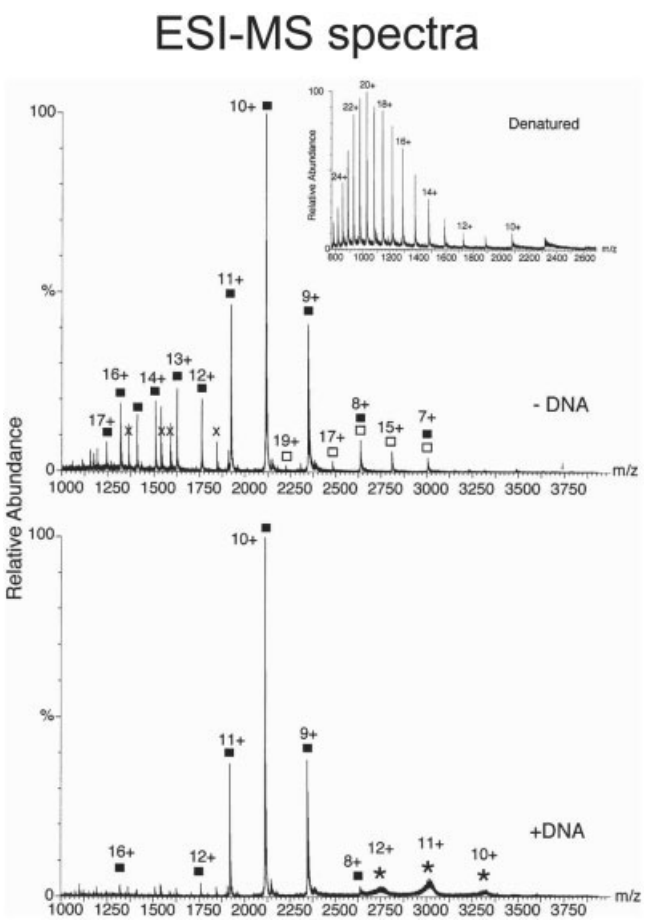

Fig. 5. Electrospray mass spectra of C170 (the catalytic domain of bacteriophage I integrase) in the absence (top) and in the presence (bottom) of the DNA partner. The ions corresponding to the indicated protonated charge states of the free protein are indicated with filled squares, the dimeric form of C170 with open squares, and the C170–DNA complex with asterisks. Ions indicated with an "x" are from a minor contaminant. The spectrum of the free protein (top) exhibits a bimodal distribution of charge states, one corresponding to the folded protein (centered at the 10+ state), and a broader distribution at a higher charge (centered at 15+). These signals arise from ionization of unfolded or partially folded protein. The signals corresponding to the unfolded protein are absent in the spectra recorded in the presence of DNA (bottom). The spectrum of acid-denatured C170 is shown in the inset, displaying the typical broad distribution of highly charged species. Reproduced from ref. 230 with permission of Mark P. Foster and Cold Spring Harbor Laboratory Press (http://www.cshlpress.com).

an IDP with a certain propensity to bind to a partner and thereby to undergo induced folding.[236] This approach has been successfully used to identify a region undergoing a random coil to α-helix transition upon binding to its physiological partner,[79] and laid the basis for the crystallographic structure determination of a chimeric complex mimicking the folded region of the IDP and its partner.[232]

The group of Gibson has recently introduced the concept of Eukaryotic Linear Motifs (ELMs). ELMs are short linear peptide motifs essentially acting independently of protein tertiary structure and are used for various cellular functions, ranging from cell compartment targeting, post-translational modifications, and protein–protein interactions.[237] The ELM server (http://elm.eu.org) comprises >80 motif patterns, and allows access to basic annotation of candidate ELMs.

Fuxreiter and coworkers have recently reported that regions of consistently predicted secondary structure, and in particular helices, correspond to folded regions in the bound form, while regions with weak conformational preferences represent linkers that remain unfolded even after interaction with the partner.[18] Moreover, the authors propose that these regions correspond to preformed structural elements in the unbound state, and that they serve as initial contact points, thus facilitating binding to the partner.

Such predictions can be experimentally assessed through mutational studies, that is, by introducing either point mutations or deletions in an IDP.

Also, regions with a propensity to induced folding can be identified by limited proteolysis studies. As already discussed, IDPs are more prone to proteolysis because they are solvent accessible and devoid of a buried core. The addition of a binding partner leads to a protein form that is less solvent exposed and thus less sensitive to protease degradation. This protection arises from shielding of specific protein regions, regardeless of the nature of the structural transition that the IDP undergoes upon binding to its target. Indeed, limited proteolytic digestion, coupled to MALDI-TOF mass spectrometry and heteronuclear protein labeling, allowed the identification of the regions of the cyclin-dependent kinase inhibitor p27 involved in the interaction with its various physiological partners (see refs. 125 and 238–240). Moreover, comparison of the proteolytic patterns obtained with different partners provided structural insights into the way p27 interacts with its various targets, and provided a molecular basis for its binding specificity[240] (Fig. 6).

Additionally, limited proteolysis experiments in the presence of TFE can also provide hints about regions of an IDP with a propensity to fold. These experiments are particularly indicated in those cases where the partner of the IDP is not available. Thermolysin, a TFE resistant enzyme, has a broad substrate specificity, and cleaves protein regions as a function of their exposure. Thus it is particularly adapted for this type of studies. N-terminal sequencing of thermolysin-resistant fragments obtained upon digestion of the intrinsically disordered N-terminal domain of the measles virus phosphoprotein (P) in the presence of TFE, allowed the identification of a protein region likely corresponding to an α-helix arising from induced folding.[77] Bioinformatic analysis of P confrmed the strong helical propensity of this region.[71]

Finally, heteronuclear NMR can lead to the identification of the residues involved in the unstructured-to-structured transition provided that backbone resonance assignment has been completed. Similarly, the information on the location of sites undergoing induced folding can be extracted from analysis of changes in MS-assisted analysis of hydrogen–deuterium exchange profiles.

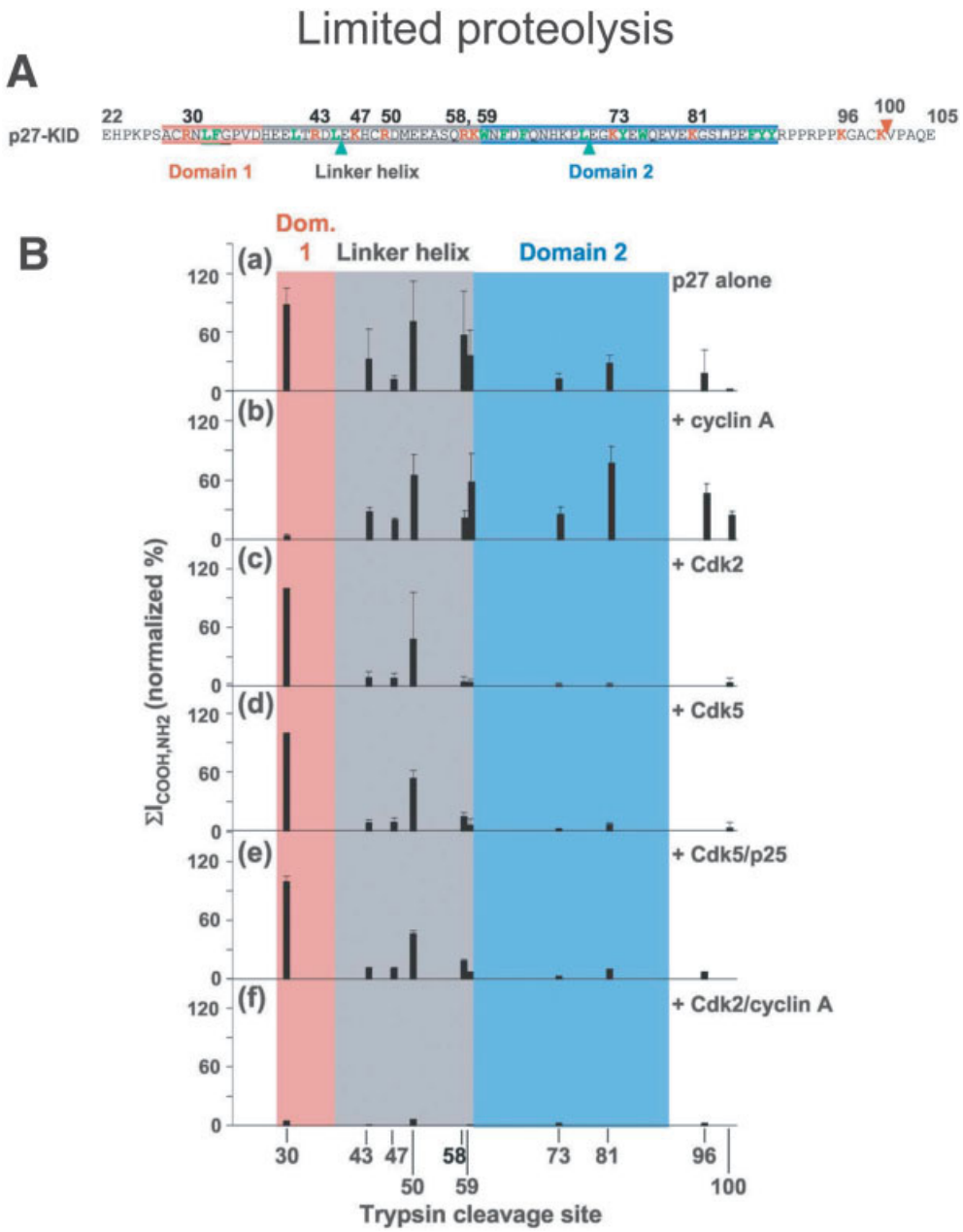

Fig. 6. Limited proteolysis of cyclin-dependent kinase inhibitor p27 (p27-KID) in the presence and absence of its various induced folding partners. (**A**) Domain structure of p27-KID. The residues colored in red are predicted trypsin cleavage sites, and those in green are predicted chymotrypsin cleavage sites. The LFG motif within domain 1 that binds cyclin A is underlined. Residue numbers for the N and C termini, and trypsin cleavage sites refer to the sequence of human p27. The red and green arrows identify sites not cleaved experimentally by trypsin and chymotrypsin, respectively. Domain 1 and domain 2 represent the segments of p27 that interact with cyclin A and Cdk2, respectively 24, 125, and 238. (**B**) Trypsin cleavage patterns based on analysis by MALDI-TOF mass spectrometry for (a) free p27-KID, and (b) p27-KID in the presence of cyclin A, (c) Cdk2, (d) Cdk5, (e) Cdk5/p25 and (f) Cdk2/cyclin A. Values on the Y-axis correspond to MALDI-TOF mass spectral peak intensities expressed as the sum of ion intensities for all fragments with N- and C-termini, corresponding to cleavage at a particular site; these values were normalized to 100% with respect to the largest peak for a particular reaction. Data for all possible trypsin cleavage sites are included, as indicated on the X-axis. The location of domain 1, the linker helix and domain 2 relative to the cleavage sites are indicated by the red, gray, and blue shading, respectively. Reproduced from ref. 240 with permission of Richard W. Kriwacki and Science Direct (http://www.sciencedirect.com).

CONCLUSIONS

The panel of different biophysical techniques herein described, even though it cannot be exhaustive, provides multiple tools to assess the lack of ordered, regular structure. These techniques yield complementary information and allow to distinguish between different "flavors" of protein disorder.[241]

The lack of ordered secondary and tertiary structure in a recombinant protein could of course also arises from a purification artefact. This point has therefore to be addressed, and assignment of a protein to the family of IDPs can only be done when its sequence and physicochemical properties both converge to show that the protein is intrinsically disordered. In this regard, the computational approaches mentioned above represent valuable tools.

Once the possibility of an artefact has been ruled out, a comprehensive understanding of the interaction of the IDP of interest with its partner(s) can be addressed. This requires a global approach to circumvent the apparent lack of information due to the inherent deficit in regular structure of the IDP. The first step consists in determining whether the isolated protein possesses knots of residual structure likely to gain further structure upon binding to a partner (and if yes, where). The presence of residual structure typifies molten globules and premolten globules, by contrast to random coils. A comparison of the global tertiary structure with the secondary structure content of the IDP, investigated using the appropriate methods, should provide this information. The identification of the residues potentially involved in binding requires a relevant combination of different experimental and computational techniques, among those described above, together with shrewdness and intuition from the investigator. Mutagenesis studies and methods assessing local tertiary structure, adequately chosen, enable the identification of protein regions involved in coupled binding and folding processes (e.g., see refs. 79 and 104).

The second step in dissecting the mode of action of the IDP addresses the molecular mechanisms that govern the interaction with its partner(s). Identifying the putative partners can be tricky, especially in the context of structural genomics projects, where IDPs can be detected among OrFans. When an in-depth knowledge of the system is available, the identification of the folding partner(s) is in principle easier. Then, to fully describe the mechanisms of the interaction, one has to gather as many information as possible, so as to give an accurate picture of the binding event. In this view, the nature and thermodynamics of the interaction can be monitored for example by ITC, the strength of the interaction can be estimated by SPR, the residues involved can be identified by NMR and/or mutagenesis, and the gain of secondary structure can be assessed by CD or FTIR. Again, the investigator must elaborate a strategy according to the properties of the IDP and its partner, to the methods available, and to his/her "creativity." The grail is, of course, to be able to crystallize the complex, to describe the interaction at the atomic scale. The field of experimentally characterized IDPs is continuously expanding, and any novel original approach will enrich the repertoire of available means to depict the structure of such refractory proteins, thus shedding light on their original mode of action.

ACKNOWLEDGMENTS

This study has been carried out with financial support from the Commission of the European Communities. It does not necessarily reflect its views, and in no way anticipates the Commission's future policy in this area. We thank Rita Grandori, Hervé Darbon, Silvia Maria Doglia, and André Fournel for useful discussions on their respective field of expertise. We are also grateful to Michel Desmadril for useful suggestions on the manuscript.

REFERENCES

1. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model 2001;19:26–59.
2. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002;27:527.
3. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci 2002;11:739–756.
4. Tompa P. The functional benefits of disorder. J Mol Structure (Theochem) 2003;666–667:361–371.
5. Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15:35–41.
6. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197–208.
7. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognition 2005;18:343–384.
8. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41: 6573–6582.
9. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584.
10. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci USA 1996;93:11504–11509.
11. Dunker AK, Obradovic Z. The protein trinity-linking function and disorder. Nat Biotechnol 2001;19:805–806.
12. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998;3:473–484.
13. Wright PE, Dyson HJ. Intrinsically unstructured proteins: reassessing the protein structure–function paradigm. J Mol Biol 1999;293:321–331.
14. Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 2003;28:81–85.
15. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005;579:3346–3354.
16. Uversky VN. What does it mean to be natively unfolded? Eur J Biochem 2002;269:2–12.
17. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 2002;12:54–60.
18. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 2004;338:1015–1026.
19. Bourhis JM, Receveur-Bréchot V, Oglesbee M, Zhang X, Buccellato M, Darbon H, Canard B, Finet S, Longhi S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Sci 2005;14:1975–1992.
20. Permyakov SE, Millett IS, Doniach S, Permyakov EA, Uversky

VN. Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. Proteins 2003;53:855–862.
21. Shi Z, Woody RW, Kallenbach NR. Is polyproline II a major backbone conformation in unfolded proteins? Adv Protein Chem 2002;62:163–240.
22. Sreerama N, Woody RW. Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. Biochemistry 1994; 33:10022–10025.
23. Bienkiewicz EA, Adkins JN, Lumb KJ. Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). Biochemistry 2002;41:752–759.
24. Lacy ER, Filippov I, Lewis WS, Otieno S, Xiao L, Weiss S, Hengst L, Kriwacki RW. p27 binds cyclin–CDK complexes through a sequential mechanism involving binding-induced protein folding. Nat Struct Mol Biol 2004;11:358–364.
25. Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, Cortese MS, Lawson JD, Brown CJ, Sikes JG, Newton CD, Dunker AK. DisProt: a database of protein disorder. Bioinformatics 2005;21:137–140.
26. Bracken C, Iakoucheva LM, Romero PR, Dunker AK. Combining prediction, computation and experiment for the characterization of protein disorder. Curr Opin Struct Biol 2004;14:570–576.
27. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK. Natively disordered proteins. In: Buchner J, Kiefhaber T, editors. Handbook of protein folding. Volume 3. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co. KGaA; 2005. p 271–353.
28. Graille M, Zhou CZ, Receveur-Brechot V, Collinet B, Declerck N, van Tilbeurgh H. Activation of the LicT transcriptional antiterminator involves a domain swing/lock mechanism provoking massive structural changes. J Biol Chem 2005;280:14780–14789.
29. Ptitsyn OB. Molten globule and protein folding. Adv Protein Chem 1995;47:83–229.
30. Baum J, Dobson CM, Evans PA, Hanley C. Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. Biochemistry 1989;28:7–13.
31. Bushnell GW, Louie GV, Brayer GD. High-resolution three-dimensional structure of horse heart cytochrome *c*. J Mol Biol 1990;214:585–595.
32. Jeng MF, Englander SW, Elove GA, Wand AJ, Roder H. Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR. Biochemistry 1990;29:10433–10437.
33. Chyan CL, Wormald C, Dobson CM, Evans PA, Baum J. Structure and stability of the molten globule state of guinea-pig alpha-lactalbumin: a hydrogen exchange study. Biochemistry 1993;32:5681–5691.
34. Wu LC, Laub PB, Elove GA, Carey J, Roder H. A noncovalent peptide complex as a model for an early folding intermediate of cytochrome *c*. Biochemistry 1993;32:10271–10276.
35. Eliezer D, Yao J, Dyson HJ, Wright PE. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. Nat Struct Biol 1998;5:148–155.
36. Bose HS, Whittal RM, Baldwin MA, Miller WL. The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. Proc Natl Acad Sci USA 1999;96:7250–7255.
37. Bracken C. NMR spin relaxation methods for characterization of disorder and folding in proteins. J Mol Graph Model 2001;19:3–12.
38. Merrill AR, Cohen FS, Cramer WA. On the nature of the structural change of the colicin E1 channel peptide necessary for its translocation-competent state. Biochemistry 1990;29:5829–5836.
39. Fontana A, Polverino de Laureto P, De Philipps V. Molecular aspects of proteolysis of globular proteins. In: van den Tweel W, Harder A, Buitelear M, editors. Protein stability and stabilization. Amsterdam: Elsevier Science; 1993. p 101–110.
40. Semisotnov GV, Zikherman KK, Kasatkin SB, Ptitsyn OB, V. AE. Polarized luminescence and mobility of tryptophan residues in polypeptide chains. Biopolymers 1981;20:2287–2309.
41. Uversky VN, Winter S, Lober G. 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 1996;60:79–88.
42. Uversky VN. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? Cell Mol Life Sci 2003;60:1852–1871.
43. Tcherkasskaya O, Uversky VN. Denatured collapsed states in protein folding: example of apomyoglobin. Proteins 2001;44:244–254.
44. Shortle DR. Structural analysis of non-native states of proteins by NMR methods. Curr Opin Struct Biol 1996;6:24–30.
45. Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J 1996;10:27–34.
46. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.
47. Baldwin RL, Zimm BH. Are denatured proteins ever random coils? Proc Natl Acad Sci USA 2000;97:12391–12392.
48. Plaxco KW, Gross M. Unfolded, yes, but random? Never! Nat Struct Biol 2001;8:659–660.
49. Hammarstrom P, Carlsson U. Is the unfolded state the Rosetta Stone of the protein folding problem? Biochem Biophys Res Commun 2000;276:393–398.
50. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem 1968;23:283–438.
51. Smith LJ, Fiebig KM, Schwalbe H, Dobson CM. The concept of a random coil. Residual structure in peptides and denatured proteins. Fold Des 1996;1:R95–R106.
52. Shortle D. The expanded denatured state: an ensemble of conformations trapped in a locally encoded topological space. Adv Protein Chem 2002;62:1–23.
53. Shimizu S, Chan HS. Origins of protein denatured state compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force. Proteins 2002;49:560–566.
54. Pelham HR. Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell 1986;46:959–961.
55. Ellis RJ, Hemmingsen SM. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. Trends Biochem Sci 1989;14:339–342.
56. Uversky VN, Kutyshenko VP, Protasova N, Rogov VV, Vassilenko KS, Gudkov AT. Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. Protein Sci 1996;5:1844–1851.
57. Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU. Chaperonin-mediated protein folding at the surface of groEL through a "molten globule"-like intermediate. Nature 1991;352: 36–42.
58. Hayer-Hartl MK, Ewbank JJ, Creighton TE, Hartl FU. Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. EMBO J 1994;13:3192–3202.
59. Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–230.
60. Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. Genome Inform Ser Workshop Genome Inform 1999;10:30–40.
61. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Structure (Camb) 2003;11:1453–1459.
62. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. Bioinformatics 2004;20:2138–2139.
63. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Res 2003;31:3701–3708.
64. Giuliani A, Benigni R, Zbilut JP, Webber CL Jr, Sirabella P, Colosimo A. Nonlinear signal analysis methods in the elucidation of protein sequence–structure relationships. Chem Rev 2002;102: 1471–1492.
65. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21:3433–3636.
66. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427.
67. Liu J, Tan H, Rost B. Loopy proteins appear conserved in evolution. J Mol Biol 2002;322:53–64.

68. Garbuzynskiy SO, Lobanov MY, Galzitskaya OV. To be folded or to be unfolded? Protein Sci 2004;13:2871–2877.
69. Coeytaux K, Poupon A. Prediction of unfolded segments in a protein sequence based on amino acid composition. Bioinformatics 2005;21:1891–1900.
70. Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci 1997;53:621–645.
71. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 2003;84(Pt 12):3239–3252.
72. Wootton JC, Federhen S. Analysis of compositionally biased regions in sequence databases. Methods Enzymol 1996;266:554–571.
73. Promponas VJ, Enright AJ, Tsoka S, Kreil DP, Leroy C, Hamodrakas S, Sander C, Ouzounis CA. CAST: an iterative algorithm for the complexity analysis of sequence tracts. Complexity analysis of sequence tracts. Bioinformatics 2000;16:915–922.
74. Hu CC, Ghabrial SA. The conserved, hydrophilic and arginine-rich N-terminal domain of cucumovirus coat proteins contributes to their anomalous electrophoretic mobilities in sodium dodecylsulfate-polyacrylamide gels. J Virol Methods 1995;55:367–379.
75. Armstrong DJ, Roman A. The anomalous electrophoretic behavior of the human papillomavirus type 16 E7 protein is due to the high content of acidic amino acid residues. Biochem Biophys Res Commun 1993;192:1380–1387.
76. Iakoucheva L, Kimzey A, Masselon C, Smith R, Dunker A, Ackerman E. Aberrant mobility phenomena of the DNA repair protein XPA. Protein Sci 2001;10:1353–1362.
77. Karlin D, Longhi S, Receveur V, Canard B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. Virology 2002;296:251–262.
78. Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem 2003;278:18638–18648.
79. Bourhis J, Johansson K, Receveur-Bréchot V, Oldfield CJ, Dunker AK, Canard B, Longhi S. The C-terminal domain of measles virus nucleoprotein belongs to the classof intrinsically disordered proteins that fold upon binding to their pohysiological partner. Virus Research 2004;99:157–167.
80. Kalthoff C. A novel strategy for the purification of recombinantly expressed unstructured protein domains. J Chromatogr B Analyt Technol Biomed Life Sci 2003;786:247–254.
81. Neumann S, Matthey U, Kaim G, Dimroth P. Purification and properties of the F1F0 ATPase of Ilyobacter tartaricus, a sodium ion pump. J Bacteriol 1998;180:3312–3316.
82. Haines DC, Sevrioukova IF, Peterson JA. The FMN-binding domain of cytochrome P450BM-3: resolution, reconstitution, and flavin analogue substitution. Biochemistry 2000;39:9419–9429.
83. Cortese MS, Baird JP, Uversky VN, Dunker AK. Uncovering the unfoldome: Enriching extracts for unstructured proteins by acid treatment. J Proteome Res 2005;4:1610–1618.
84. Ptitsyn OB, Uversky VN. The molten globule is a third thermodynamical state of protein molecules. FEBS Lett 1994;341:15–18.
85. Uversky VN, Ptitsyn OB. All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. Fold Des 1996;1:117–122.
86. Uversky VN, Li J, Fink AL. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. J Biol Chem 2001;276:10737–10744.
87. Uversky VN, Permyakov SE, Zagranichny VE, Rodionov IL, Fink AL, Cherskaya AM, Wasserman LA, Permyakov EA. Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein. J Proteome Res 2002;1:149–159.
88. Timm DE, Vissavajjhala P, Ross AH, Neet KE. Spectroscopic and chemical studies of the interaction between nerve growth factor (NGF) and the extracellular domain of the low affinity NGF receptor. Protein Sci 1992;1:1023–1031.
89. Kim TD, Ryu HJ, Cho HI, Yang CH, Kim J. Thermal behavior of proteins: heat-resistant proteins and their heat-induced secondary structural changes. Biochemistry 2000;39:14839–14846.
90. Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM. Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. Biochemistry 1999;38:15009–15016.
91. Konno T, Tanaka N, Kataoka M, Takano E, Maki M. A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. Biochim Biophys Acta 1997;1342:73–82.
92. Lynn A, Chandra S, Malhotra P, Chauhan VS. Heme binding and polymerization by Plasmodium falciparum histidine rich protein II: influence of pH on activity and conformation. FEBS Lett 1999;459:267–271.
93. Johansson J, Gudmundsson GH, Rottenberg ME, Berndt KD, Agerberth B. Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. J Biol Chem 1998;273:3718–3724.
94. Munishkina LA, Fink AL, Uversky VN. Conformational prerequisites for formation of amyloid fibrils from histones. J Mol Biol 2004;342:1305–1324.
95. Zhan Y, Song X, Zhou GW. Structural analysis of regulatory protein domains using GST-fusion proteins. Gene 2001;281:1–9.
96. Buck M. Crystallography: embracing conformational flexibility in proteins. Structure (Camb) 2003;11:735–736.
97. Loris R, Marianovsky I, Lah J, Laeremans T, Engelberg-Kulka H, Glaser G, Muyldermans S, Wyns L. Crystal structure of the intrinsically flexible addiction antidote MazE. J Biol Chem 2003;278:28252–28257.
98. Privalov PL. Stability of proteins: small globular proteins. Adv Protein Chem 1979;33:167–241.
99. Privalov PL. Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem 1982;35:1–104.
100. Uversky VN. A multiparametric approach to studies of self-organization of globular proteins. Biochemistry (Mosc) 1999;64:250–266.
101. Uversky VN, Kirkitadze MD, Narizhneva NV, Potekhin SA, Tomashevski A. Structural properties of alpha-fetoprotein from human cord serum: the protein molecule at low pH possesses all the properties of the molten globule. FEBS Lett 1995;364:165–167.
102. Abdullaev ZK, Latypov RF, Badretdinov AY, Dolgikh DA, Finkelstein AV, Uversky VN, Kirpichnikov MP. S6 permutein shows that the unusual target topology is not responsible for the absence of rigid tertiary structure in de novo protein albebetin. FEBS Lett 1997;414:243–246.
103. Aphasizheva IY, Dolgikh DA, Abdullaev ZK, Uversky VN, Kirpichnikov MP, Ptitsyn OB. Can grafting of an octapeptide improve the structure of a de novo protein? FEBS Lett 1998;425:101–104.
104. Mendoza C, Figueirido F, Tasayco ML. DSC studies of a family of natively disordered fragments from *Escherichia coli* thioredoxin: surface burial in intrinsic coils. Biochemistry 2003;42:3349–3358.
105. Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers 1991;31:119–128.
106. Naeem A, Khan KA, Khan RH. Characterization of a partially folded intermediate of papain induced by fluorinated alcohols at low pH. Arch Biochem Biophys 2004;432:79–87.
107. Christensen H. The contribution of the molten globule model, in Mechanism of protein folding. In: RH Pain, editor. Mechanisms of protein folding. New York: Oxford University Press; 1994. p 55–79.
108. Agashe VR, Shastry MC, Udgaonkar JB. Initial hydrophobic collapse in the folding of barstar. Nature 1995;377:754–757.
109. Roberts LM, Dunker AK. Structural changes accompanying chloroform-induced contraction of the filamentous phage fd. Biochemistry 1993;32:10479–10488.
110. Uversky VN, Ptitsyn OB. "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. Biochemistry 1994;33:2782–2791.
111. Uversky VN, Ptitsyn OB. 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 1996;255:215–228.
112. Uversky VN, Karnoup AS, Segel DJ, Seshadri S, Doniach S, Fink AL. Anion-induced folding of Staphylococcal nuclease: character-

ization of multiple equilibrium partially folded intermediates. J Mol Biol 1998;278:879–894.
113. Uversky VN, Karnoup AS, Khurana R, Segel DJ, Doniach S, Fink AL. Association of partially-folded intermediates of staphylococcal nuclease induces structure and stability. Protein Sci 1999;8:161–173.
114. Hubbard SJ, Eisenmenger F, Thornton JM. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. Protein Sci 1994;3:757–768.
115. Hubbard SJ, Beynon RJ, Thornton JM. Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures. Protein Eng 1998;11:349–359.
116. Hubbard SJ, Campbell SF, Thornton JM. Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. J Mol Biol 1991;220:507–530.
117. Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zambonin M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. Biochemistry 1986;25:1847–1851.
118. Novotny J, Bruccoleri RE. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. FEBS Lett 1987;211:185–189.
119. Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. Fold Des 1997;2:R17–R26.
120. Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. J Mol Biol 1997;266:223–230.
121. Spolaore B, Bermejo R, Zambonin M, Fontana A. Protein interactions leading to conformational changes monitored by limited proteolysis: apo form and fragments of horse cytochrome *c*. Biochemistry 2001;40:9460–9468.
122. Markus G. Protein substrate conformation and proteolysis. Proc Natl Acad Sci USA 1965;54:253–258.
123. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci USA 1996;93:11504–11509.
124. Lisse T, Bartels D, Kalbitzer HR, Jaenicke R. The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Craterostigma plantagineum* displays no defined three-dimensional structure in its native state. Biol Chem 1996;377:555–561.
125. Kriwacki RW, Wu J, Tennant L, Wright PE, Siuzdak G. Probing protein structure using biochemical and biophysical methods. Proteolysis, matrix-assisted laser desorption/ionization mass spectrometry, high-performance liquid chromatography and size-exclusion chromatography of p21Waf1/Cip1/Sdi1. J Chromatogr A 1997;777:23–30.
126. Horiuchi M, Kurihara Y, Katahira M, Maeda T, Saito T, Uesugi S. Dimerization and DNA binding facilitate alpha-helix formation of Max in solution. J Biochem (Tokyo) 1997;122:711–716.
127. Hubbard SJ. The structural aspects of limited proteolysis of native proteins. Biochim Biophys Acta 1998;1382:191–206.
128. Hershey PE, McWhirter SM, Gross JD, Wagner G, Alber T, Sachs AB. The Cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1. J Biol Chem 1999;274:21297–21304.
129. Ratnaswamy G, Koepf E, Bekele H, Yin H, Kelly JW. The amyloidogenicity of gelsolin is controlled by proteolysis and pH. Chem Biol 1999;6:293–304.
130. Iakoucheva LM, Kimzey AL, Masselon CD, Bruce JE, Garner EC, Brown CJ, Dunker AK, Smith RD, Ackerman EJ. Identification of intrinsic order and disorder in the DNA repair protein XPA. Protein Sci 2001;10:560–571.
131. Bouvier M, Stafford WF. Probing the three-dimensional structure of human calreticulin. Biochemistry 2000;39:14950–14959.
132. Amit AG, Mariuzza RA, Phillips SE, Poljak RJ. Three-dimensional structure of an antigen-antibody complex at 6 A resolution. Nature 1985;313:156–158.
133. Wilson IA, Haft DH, Getzoff ED, Tainer JA, Lerner RA, Brenner S. Identical short peptide sequences in unrelated proteins can have different conformations: a testing ground for theories of immune recognition. Proc Natl Acad Sci USA 1985;82:5255–5259.
134. Fujio H, Takagaki Y, Ha YM, Doi EM, Soebandrio A, Sakato N. Native and non-native conformation-specific antibodies directed to the loop region of hen egg-white lysozyme. J Biochem (Tokyo) 1985;98:949–962.
135. Glatter O, Kratzy O. Small angle X-ray scattering. London: Academic Press, London; 1982.
136. Millett IS, Doniach S, Plaxco KW. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. Adv Protein Chem 2002;62:241–262.
137. Uversky VN, Segel DJ, Doniach S, Fink AL. Association-induced folding of globular proteins. Proc Natl Acad Sci USA 1998;95:5480–5483.
138. Moncoq K, Broutin I, Craescu CT, Vachette P, Ducruix A, Durand D. SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer? Biophys J 2004;87:4056–4064.
139. Perez J, Vachette P, Russo D, Desmadril M, Durand D. Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. J Mol Biol 2001;308:721–743.
140. Kataoka M, Hagihara Y, Mihara K, Goto Y. Molten globule of cytochrome *c* studied by small angle X-ray scattering. J Mol Biol 1993;229:591–596.
141. Kataoka M, Goto Y. X-ray solution scattering studies of protein folding. Fold Des 1996;1:R107–114.
142. Eliezer D, Chiba K, Tsuruta H, Doniach S, Hodgson KO, Kihara H. Evidence of an associative intermediate on the myoglobin refolding pathway. Biophys J 1993;65:912–917.
143. Uversky VN, Li J, Souillac P, Jakes R, Goedert M, Fink AL. Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins. J Biol Chem 2002;25:25.
144. Uversky VN. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. Biochemistry 1993;32:13288–13298.
145. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry 1999;38:16424–16431.
146. Gast K, Damaschun H, Misselwitz R, Muller-Frohne M, Zirwer D, Damaschun G. Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. Eur Biophys J 1994;23:297–305.
147. Laue TM, Stafford WF, 3rd. Modern applications of analytical ultracentrifugation. Annu Rev Biophys Biomol Struct 1999;28:75–100.
148. Casassa EF, Eisenberg H. Thermodynamic analysis of multicomponent solutions. Adv Protein Chem 1964;19:287–395.
149. Fujita H. Foundations of ultracentrifugal analysis. New York: Wiley; 1975.
150. Goldberg RJ. Sedimentation in the ultracentrifuge. J Phys Chem 1953;57:194–202.
151. Williams JW, Van Holde KE, Baldwin RL, Fujita H. The theory of sedimentation analysis. Chem Rev 1958;58:715–806.
152. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. Adv Protein Chem 1970;24:1–95.
153. Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, Venyaminov S, Ptitsyn OB. Alpha-Lactalbumin: compact state with fluctuating tertiary structure? FEBS Lett 1981;136:311–315.
154. Dolgikh DA, Abaturov LV, Bolotina IA, Brazhnikov EV, Bychkova VE, Gilmanshin RI, Lebedev Yu O, Semisotnov GV, Tiktopulo EI, Ptitsyn OB, et al. Compact state of a protein molecule with pronounced small-scale mobility: bovine alpha-lactalbumin. Eur Biophys J 1985;13:109–121.
155. Wong KP, Hamlin LM. Acid denaturation of bovine carbonic anhydrase B. Biochemistry 1974;13:2678–2683.
156. Rodionova NA, Semisotnov GV, Kutyshenko VP, Uverskii VN, Bolotina IA. Staged equilibrium of carbonic anhydrase unfolding in strong denaturants. Mol Biol (Mosk) 1989;23:683–692.
157. Adler AJ, Greenfield NJ, Fasman GD. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. Methods Enzymol 1973;27:675–735.
158. Fasman GD. Circular dichroism and conformational analysis of biomolecules. New York: Plenum Press; 1996.
159. Haris P, Severcan F. FTIR spectroscopic characterization of protein structure in acqueous and non-acqueous media. J Mol Catal B-Enzym 1999;7:207–221.
160. Natalello A, Ami D, Brocca S, Lotti M, Doglia S. Secondary

structure, conformational stability and glycosylation of a recombinant *Candida rugosa* lipase studied by Fourier transform infrared spectroscopy. Biochem J 2004;385:511–517.
161. Barron LD, Hecht L, Blanch EW, Bell AF. Solution structure and dynamics of biomolecules from Raman optical activity. Prog Biophys Mol Biol 2000;73:1–49.
162. Smyth E, Syme CD, Blanch EW, Hecht L, Vasak M, Barron LD. Solution structure of native proteins with irregular folds from Raman optical activity. Biopolymers 2001;58:138–151.
163. Carey PR. Molecular biology: Biochemical applications of Raman and resonance Raman spectroscopies. New York: Academic Press; 1982.
164. Asher SA. Handbook of vibrational spectroscopy. New York: Wiley; 2001.
165. Palaniappan V, Bocian DF. Acid-induced transformations of myoglobin. Characterization of a new equilibrium heme-pocket intermediate. Biochemistry 1994;33:14264–14274.
166. Xu M, Ermolenkov VV, He W, Uversky VN, Fredriksen L, Lednev IK. Lysozyme fibrillation: Deep UV Raman spectroscopic characterization of protein structural transformation. Biopolymers 2005;79:58–61.
167. Lednev IK, Ermolenkov VV, He W, Xu M. Deep-UV Raman spectrometer tunable between 193 and 205 nm for structural characterization of proteins. Anal Bioanal Chem 2005;381:431–437.
168. Copeland RA, Spiro TG. Secondary structure determination in proteins from deep (192–223-nm) ultraviolet Raman spectroscopy. Biochemistry 1987;26:2134–2139.
169. Chi Z, Chen XG, Holtz JS, Asher SA. UV resonance Raman-selective amide vibrational enhancement: quantitative methodology for determining protein secondary structure. Biochemistry 1998;37:2854–2864.
170. Dyson HJ, Wright PE. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. Methods Enzymol 2001;339:258–270.
171. Dyson HJ, Wright PE. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem 2002;62:311–340.
172. Neri D, Billeter M, Wider G, Wuthrich K. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. Science 1992;257:1559–1563.
173. Zhang O, Kay LE, Olivier JP, Forman-Kay JD. Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR 1994;4:845–858.
174. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J Biomol NMR 1994;4:727–734.
175. Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. Spectral density function mapping using 15N relaxation data exclusively. J Biomol NMR 1995;6:153–162.
176. Zhang O, Forman-Kay JD. Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. Biochemistry 1995;34:6784–6794.
177. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. Biochemistry 1995;34:868–878.
178. Zhang O, Kay LE, Shortle D, Forman-Kay JD. Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Delta131Delta, using NMR methods with improved resolution. J Mol Biol 1997;272:9–20.
179. Zhang O, Forman-Kay JD. NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. Biochemistry 1997;36:3959–3970.
180. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. Biochemistry 1997;36:2390–2402.
181. Zhang O, Forman-Kay JD, Shortle D, Kay LE. Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. J Biomol NMR 1997;9:181–200.
182. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chem Rev 2004;104:3607–3622.
183. Green TB, Ganesh O, Perry K, Smith L, Phylip LH, Logan TM, Hagen SJ, Dunn BM, Edison AS. IA3, an aspartic proteinase inhibitor from *Saccharomyces cerevisiae*, is intrinsically unstructured in solution. Biochemistry 2004;43:4071–4081.
184. Lietzow MA, Jamin M, Jane Dyson HJ, Wright PE. Mapping long-range contacts in a highly unfolded protein. J Mol Biol 2002;322:655–662.
185. Gillespie JR, Shortle D. 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 1997;268:170–184.
186. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. J Mol Biol 1997;268:158–169.
187. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci 2000;1:349–384.
188. Lackowicz J. Principles of fluorescence spectroscopy. New York: Kluwer Academic/Plenum Publishers; 1999.
189. Stryer L. Fluorescence spectroscopy of proteins. Science 1968;162:526–533.
190. Chen Y, Barkley MD. Toward understanding tryptophan fluorescence in proteins. Biochemistry 1998;37:9976–9982.
191. Lakowitz J. Principles of fluorescence spectroscopy. New York: Plenum Press; 1983.
192. Eftink MR, Ghiron CA. Fluorescence quenching studies with proteins. Anal Biochem 1981;114:199–227.
193. Förster T. Intermolecular energy migration and fluorescence. Ann Physik 1948;2:55–75.
194. Rischel C, Poulsen FM. Modification of a specific tyrosine enables tracing of the end-to-end distance during apomyoglobin folding. FEBS Lett 1995;374:105–109.
195. Uversky VN, Fink AL. Do protein molecules have a native-like topology in the pre-molten globule state? Biochemistry (Mosc) 1999;64:552–555.
196. Tcherkasskaya O, Ptitsyn OB. Molten globule versus variety of intermediates: influence of anions on pH-denatured apomyoglobin. FEBS Lett 1999;455:325–331.
197. Tcherkasskaya O, Ptitsyn OB. Direct energy transfer to study the 3D structure of non-native proteins: AGH complex in molten globule state of apomyoglobin. Protein Eng 1999;12:485–490.
198. Weber G. Polarization of the fluorescence of macromolecules. I. Theory and experimental method. Biochem J 1952;51:145–155.
199. Weber G. Polarization of the fluorescence of macromolecules. II. Fluorescent conjugates of ovalbumin and bovine serum albumin. Biochem J 1952;51:155–167.
200. Weber G. Rotational Brownian motion and polarization of the fluorescence of solutions. Adv Protein Chem 1953;8:415–459.
201. Weber G. Fluorescence-polarization spectrum and electronic-energy transfer in proteins. Biochem J 1960;75:345–352.
202. Weber G. Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds. Biochem J 1960;75:335–345.
203. Permyakov EA. Luminescence spectroscopy of proteins. London: CRC Press; 1993.
204. Hubbell WL, Gross A, Langen R, Lietzow MA. Recent advances in site-directed spin labeling of proteins. Curr Opin Struct Biol 1998;8:649–656.
205. Cai K, Langen R, Hubbell WL, Khorana HG. Structure and function in rhodopsin: topology of the C-terminal polypeptide chain in relation to the cytoplasmic loops. Proc Natl Acad Sci USA 1997;94:14267–14272.
206. Grandori R. Detecting equilibrium cytochrome *c* folding intermediates by electrospray ionisation mass spectrometry: two partially folded forms populate the molten-globule state. Protein Sci 2002;11:453–458.
207. Simmons DA, Konermann L. Characterization of transient protein folding intermediates during myoglobin reconstitution by time-resolved electrospray mass spectrometry with on-line isotopic pulse labeling. Biochemistry 2002;41:1906–1914.
208. Grandori R. Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry. J Mass Spectrom 2003;38:11–15.
209. Simmons DA, Dunn SD, Konermann L. Conformational dynamics of partially denatured myoglobin studied by time-resolved electrospray mass spectrometry with online hydrogen-deuterium exchange. Biochemistry 2003;42:5896–5905.
210. Wilson DJ, Konermann L. Mechanistic studies on enzymatic

reactions by electrospray ionization MS using a capillary mixer with adjustable reaction chamber volume for time-resolved measurements. Anal Chem 2004;76:2537–2543.
211. Engen JR, Smith DL. Investigating protein structure and dynamics by hydrogen exchange MS. Anal Chem 2001;73:256A–265A.
212. Nishimura C, Dyson HJ, Wright PE. Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. Proc Natl Acad Sci USA 2005;102:4765–4770.
213. Konermann L, Simmons DA. Protein-folding kinetics and mechanisms studied by pulse-labeling and mass spectrometry. Mass Spectrom Rev 2003;22:1–26.
214. Hua QX, Jia WH, Bullock BP, Habener JF, Weiss MA. Transcriptional activator-coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. Biochemistry 1998;37:5858–5866.
215. Watt K, Jess TJ, Kelly SM, Price NC, McEwan IJ. Induced alpha-helix structure in the aryl hydrocarbon receptor transactivation domain modulates protein–protein interactions. Biochemistry 2005;44:734–743.
216. Baskakov I, Wang A, Bolen DW. Trimethylamine-N-oxide counteracts urea effects on rabbit muscle lactate dehydrogenase function: a test of the counteraction hypothesis. Biophys J 1998;74:2666–2673.
217. Baskakov I, Bolen DW. Forcing thermodynamically unfolded proteins to fold. J Biol Chem 1998;273:4831–4834.
218. Baskakov IV, Kumar R, Srinivasan G, Ji YS, Bolen DW, Thompson EB. Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. J Biol Chem 1999;274:10693–10696.
219. Bolen DW, Baskakov IV. The osmophobic effect: natural selection of a thermodynamic force in protein folding. J Mol Biol 2001;310:955–963.
220. Luo P, Baldwin RL. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. Biochemistry 1997;36:8413–8421.
221. Cammers-Goodwin A, Allen TJ, Oslick SL, McClure KF, Lee JH, Kemp DS. Mechanism of stabilization of helical conformations of polypeptides by water containing trifluoroethanol. J Am Chem Soc 1996;118:3082–3090.
222. Buck M, Schwalbe H, Dobson CM. Characterization of conformational preferences in a partly folded protein by heteronuclear NMR spectroscopy: assignment and secondary structure analysis of hen egg-white lysozyme in trifluoroethanol. Biochemistry 1995;34:13219–13232.
223. Fan P, Bracken C, Baum J. Structural characterization of monellin in the alcohol-denatured state by NMR: evidence for beta-sheet to alpha-helix conversion. Biochemistry 1993;32:1573–1582.
224. Luo Y, Baldwin RL. Trifluoroethanol stabilizes the pH 4 folding intermediate of sperm whale apomyoglobin. J Mol Biol 1998;279:49–57.
225. Wang A, Bolen DW. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. Biochemistry 1997;36:9101–9108.
226. Van Hoy M, Leuther KK, Kodadek T, Johnston SA. The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. Cell 1993;72:587–594.
227. Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. J Biol Chem 2003;278:44567–44573.
228. Nyarko A, Hare M, Hays TS, Barbar E. The intermediate chain of cytoplasmic dynein is partially disordered and gains structure upon binding to light-chain LC8. Biochemistry 2004;43:15595–15603.
229. Veenstra TD, Johnson KL, Tomlinson AJ, Craig TA, Kumar R, Naylor S. Zinc-induced conformational changes in the DNA-binding domain of the vitamin D receptor determined by electrospray ionization mass spectrometry. J Am Soc Mass Spectrom 1998;9:8–14.
230. Kamadurai HB, Subramaniam S, Jones RB, Green-Church KB, Foster MP. Protein folding coupled to DNA binding in the catalytic domain of bacteriophage lambda integrase detected by mass spectrometry. Protein Sci 2003;12:620–626.
231. Honnappa S, Cutting B, Jahnke W, Seelig J, Steinmetz MO. Thermodynamics of the Op18/stathmin-tubulin interaction. J Biol Chem 2003;278:38926–38934.
232. Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. Structural basis for the attachment of a paramyxoviral polymerase to its template. Proc Natl Acad Sci USA 2004;101:8301–8306.
233. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. Science 1994;263:777–784.
234. Zako T, Harada K, Mannen T, Yamaguchi S, Kitayama A, Ueda H, Nagamune T. Monitoring of the refolding process for immobilized firefly luciferase with a biosensor based on surface plasmon resonance. J Biochem (Tokyo) 2001;129:1–4.
235. Persson M, Hammarstrom P, Lindgren M, Jonsson BH, Svensson M, Carlsson U. EPR mapping of interactions between spin-labeled variants of human carbonic anhydrase II and GroEL: evidence for increased flexibility of the hydrophobic core by the interaction. Biochemistry 1999;38:432–441.
236. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44:12454–12470.
237. Puntervoll P, Linding R, Gemund C, Chabanis-Davidson S, Mattingsdal M, Cameron S, Martin DM, Ausiello G, Brannetti B, Costantini A, Ferre F, Maselli V, Via A, Cesareni G, Diella F, Superti-Furga G, Wyrwicz L, Ramu C, McGuigan C, Gudavalli R, Letunic I, Bork P, Rychlewski L, Kuster B, Helmer-Citterich M, Hunter WN, Aasland R, Gibson TJ. ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. Nucleic Acids Res 2003;31:3625–3630.
238. Kriwacki RW, Wu J, Siuzdak G, Wright PE. Probing protein/protein interactions with mass spectrometry and isotopic labeling: analysis of the p21/Cdk2 complex. J Am Chem Soc 1996;118:5320–5321.
239. Kriwacki RW, Siuzdak G. Probing protein–protein interactions with mass spectrometry. Methods Mol Biol 2000;146:223–238.
240. Lacy ER, Wang Y, Post J, Nourse A, Webb W, Mapelli M, Musacchio A, Siuzdak G, Kriwacki RW. Molecular basis for the specificity of p27 toward cyclin-dependent kinases that regulate cell division. J Mol Biol 2005;349:764–773.
241. Vucetic S, Brown C, Dunker K, Obradovic Z. Flavors of protein disorder. Proteins 2003;52:573–584.
242. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, Wright PE. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature 2002;415:549–553.