Article No. jmbi.1999.3110 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) **293**, 321–331

JMB

Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm

Peter E. Wright* and H. Jane Dyson*

Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA 92037, USA

**Corresponding authors*

A major challenge in the post-genome era will be determination of the functions of the encoded protein sequences. Since it is generally assumed that the function of a protein is closely linked to its three-dimensional structure, prediction or experimental determination of the library of protein structures is a matter of high priority. However, a large proportion of gene sequences appear to code not for folded, globular proteins, but for long stretches of amino acids that are likely to be either unfolded in solution or adopt non-globular structures of unknown conformation. Characterization of the conformational propensities and function of the non-globular protein sequences represents a major challenge. The high proportion of these sequences in the genomes of all organisms studied to date argues for important, as yet unknown functions, since there could be no other reason for their persistence throughout evolution. Clearly the assumption that a folded three-dimensional structure is necessary for function needs to be re-examined. Although the functions of many proteins are directly related to their three-dimensional structures, numerous proteins that lack intrinsic globular structure under physiological conditions have now been recognized. Such proteins are frequently involved in some of the most important regulatory functions in the cell, and the lack of intrinsic structure in many cases is relieved when the protein binds to its target molecule. The intrinsic lack of structure can confer functional advantages on a protein, including the ability to bind to several different targets. It also allows precise control over the thermodynamics of the binding process and provides a simple mechanism for inducibility by phosphorylation or through interaction with other components of the cellular machinery. Numerous examples of domains that are unstructured in solution but which become structured upon binding to the target have been noted in the areas of cell cycle control and both transcriptional and translational regulation, and unstructured domains are present in proteins that are targeted for rapid destruction. Since such proteins participate in critical cellular control mechanisms, it appears likely that their rapid turnover, aided by their unstructured nature in the unbound state, provides a level of control that allows rapid and accurate responses of the cell to changing environmental conditions.

© 1999 Academic Press

Keywords: unstructured proteins; protein structure; protein folding; transcriptional regulation; translation initiation

Structure and function: a necessary partnership?

When the first issue of *Journal of Molecular Biology* was published in 1959, only one three-dimensional protein structure, that of myoglobin, had been solved, in the laboratory of the founding editor of this Journal (Kendrew *et al*., 1958). In the intervening 40 years, the field of structural biology has grown explosively and in 1999 more than 9000

Abbreviations used: CBP, CREB binding protein; HAT, histone acetyltransferase; DBD, DNA-binding domain; KID, kinase-inducible activation domain; pKID, phosphorylated KID; TBP, TATA box-binding protein; sTAR, steroid acute regulatory protein; LEF-1 lymphoid enhancer binding factor; HMG, high mobility group; RXR, retinoid X receptor.

E-mail address of the corresponding authors: wright@scripps.edu; dyson@scripps.edu

0022-2836/99/420321–11 $30.00/0 © 1999 Academic Press

sets of atomic coordinates of proteins have been deposited in the Protein Data Bank. In parallel with this dramatic growth in structural knowledge, there has been an increasing conviction that the biological function of proteins is intrinsically encoded in their detailed 3D structures. In short, the central dogma of structural biology is that a folded protein structure is necessary for biological function. While there can be no doubt that protein structure and function are intimately linked, there is a growing awareness that not all biologically functional proteins encoded by the various genomes fold spontaneously into stable globular structures. Here, we will review some of the evidence that establishes unfolded or partly folded states of proteins as being key to understanding certain biological processes.

Disorder in protein structures can be either local or global. Locally disordered regions are common and have been observed in numerous X-ray and NMR structures of proteins; in some cases, such disordered regions have been linked to biological function. Until recently, well-characterized examples of global disorder have been rare, although it has long been known that small polypeptide hormones, such as monomeric glucagon, are flexible random coils in aqueous solution (Bösch *et al*., 1978) which presumably fold upon binding to their receptors. However, it is now recognized that many protein domains or even full-length proteins are intrinsically unstructured. Indeed, examination of amino acid sequence data shows that many proteins contain extended segments of low sequence complexity and compositional bias that are unlikely to adopt globular folds. These observations raise intriguing questions about the role of protein disorder in biological processes.

Database predictions

A survey performed in 1994 showed that half of the protein sequences in the Swiss Protein Database at that time contained segments of low complexity that correspond to non-globular regions (Wootton, 1994). Such regions are not necessarily disordered, as they include repetitive structural elements like coiled coils. Disordered regions of proteins are characterized by low sequence complexity, with amino acid compositional bias and high predicted flexibility (Dunker *et al*., 1998; Garner *et al*., 1998). A neural network algorithm has been developed to screen proteins for long disordered regions, using primary sequence information alone (Romero *et al*., 1998). Using this algorithm, more than 15,000 proteins in the Swiss Protein Database were predicted to contain long disordered segments 40 or more residues in length. Thus, local or even global disorder appears to be quite prevalent in the proteins encoded by the various genomes. Indeed, the "average" genomic protein is significantly different in length, composition, and secondary structure from those of known 3D structure in the Protein Data Bank (Gerstein, 1998). Determining the structure, or lack thereof, of the non-globular protein sequences in the genomes represents a major challenge for structural biology.

As an example, let us consider the transcriptional coactivator, CBP (CREB binding protein) and its paralog p300. CBP and p300 are central mediators of transcription, integrating numerous signaling pathways in the cell with the selective induction of gene expression (Shikama *et al*., 1997; Janknecht & Hunter, 1996). These proteins function by binding the activation domains of many transcription factors and recruiting RNA polymerase through direct interactions with components of the basal transcriptional complex. They also appear to exert control over chromatin structure by way of histone acetyltransferase (HAT) activity. CBP is a modular protein of length 2441 amino acid residues. Analysis of the CBP amino acid sequence reveals only seven potentially globular domains, with more than half of the sequence corresponding to regions of low complexity and high predicted flexibility. Three of the putative globular domains are implicated in protein-protein interactions (the Cys/His-rich C/H1 and C/H3 domains, and the KIX domain). In addition, there are HAT and bromodomains associated with chromatin remodeling (Ogryzko *et al*., 1996; Jeanmougin *et al*., 1997) and putative PHD and ZZ zinc finger domains of unknown function (Aasland *et al*., 1995; Ponting *et al*., 1996). Sequence analysis predicts that long segments of the protein at the N and C termini and between many of the identifiable domains are non-globular and highly flexible, and hence are likely to be intrinsically disordered. These segments include the nuclear receptor interaction domain (residues 1-101), and a long glutamine-rich C-terminal "transactivation" domain (residues 1846-2441) that interacts with p160^{SRC-1} (Janknecht & Hunter, 1996). The unstructured receptor interaction domain includes the LXXLL motif that has been directly implicated in nuclear receptor binding (Torchia *et al*., 1997; Heery *et al*., 1997), and which has been shown, for other coactivators, to form an α-helix upon complex formation (Nolte *et al*., 1998; Darimont *et al*., 1998; Shiau *et al*., 1998). The entire glutamine-rich region between residues 1846-2441, which contains another LXXLL motif in the region that binds p160^{SRC-1} and other coactivators, is likely to be disordered.

Experimental methods for characterizing unstructured states

Spectroscopic methods are a powerful tool for the detection of unfolded and partially folded states of proteins. Recent advances in NMR technology, including the advent of ultra-high field instruments and the application of heteronuclear multidimensional experiments, have allowed resonances of unfolded and partly folded proteins to be fully assigned (Logan *et al*., 1994; Zhang *et al*., 1994; Eliezer *et al*., 1998), opening the way to

detailed structural and dynamic characterization of these states. Circular dichroism and fluorescence spectra can also provide valuable insights into the secondary structure of non-globular proteins and can be used to monitor folding transitions associated with binding. On the basis of spectroscopic experiments, many groups have recently reported proteins that appear to be unfolded or only partly folded under normal conditions in the cell. These include the cyclin-dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$ (Kriwacki *et al*., 1996a), a fibronectin-binding protein (Penkett *et al*., 1997), the steroidogenic acute regulatory protein (Bose *et al*., 1999), and the unliganded state of the $p21^{H\text{-}ras}$ protein (Zhang & Matthews, 1998), to name but a few. Many unstructured or partially structured proteins are extremely sensitive to protease degradation, and differential protease cleavage patterns in the presence and absence of the target provide a powerful method for identifying the binding region (Kriwacki *et al*., 1996b).

Folding transitions in DNA recognition

Eukaryotic transcription factors contain a minimum of two domains which mediate sequence-specific DNA recognition and transcriptional activation. The structures of numerous DNA-binding domains (DBDs) are known. Most DBDs adopt folded structures in the absence of DNA, although folding of locally disordered segments does accompany DNA binding in a significant number of cases (Spolar & Record, 1994). A particularly well-characterized example is the basic DNA-binding region of the leucine zipper protein GCN4; the basic region is largely unstructured in the absence of DNA, but folds into a stable helical structure upon binding to its cognate AP-1 site (Weiss *et al*., 1990). NMR experiments show that the basic region transiently forms nascent helical structures in the unbound state that significantly reduce the entropic cost of DNA binding (Bracken *et al*., 1999). Nevertheless, there is a large loss of conformational entropy upon DNA binding, associated with the folding of the basic region helix. This requirement for DNA-induced folding appears to enhance the specificity of DNA binding; the helical content of the basic region is greater when bound to a specific *versus* non-specific DNA site (O'Neil *et al*., 1990). Thus, GCN4 provides an example of local, DNA-induced folding required for sequence-specific nucleic acid recognition. A further example is the DNA-binding domain of NFATC1, which undergoes substantial folding transitions, required for both DNA recognition and for the establishment of cooperative protein-protein contacts upon complexation with its DNA target (Zhou *et al*., 1998).

The role played by DNA-induced folding in cooperative binding processes is illustrated by the nuclear hormone receptors. The nuclear receptor DBDs bind with high cooperativity and high affinity as either homo- or heterodimers to response elements consisting of two direct repeats of the sequence AGGTCA (Perlmann *et al*., 1993; Mader *et al*., 1993). The DBDs do not dimerize spontaneously in solution and bind only weakly and without cooperativity to an isolated DNA half-site; dimerization is mediated by DNA binding. NMR studies of the free retinoid X receptor (RXR) DBD show that the dimer interface is disordered in the absence of DNA (Holmbeck *et al*., 1998a; van Tilborg *et al*., 1999). However, upon binding to a DNA half-site, the structure is induced and stabilized in the region of the protein that forms the dimerization interface (Holmbeck *et al*., 1998b); the induced structure is very similar to that seen in the X-ray structure of the RXR-TR heterodimer assembled on DNA (Rastinejad *et al*., 1995). Thus, the molecular basis for cooperativity is the local induced folding that accompanies DNA binding and creates an appropriate molecular surface to facilitate dimer formation with a second DBD recruited to a neighboring half-site.

An even more striking example of the coupling between protein folding and DNA binding is exhibited by the high mobility group (HMG) domain of the lymphoid enhancer-binding factor LEF-1. LEF-1 regulates the T cell receptor-α gene enhancer by inducing a major bend in the DNA (Love *et al*., 1995). The HMG domain is poorly folded in the absence of its cognate DNA and several major structural elements are stabilized only upon binding to and bending the DNA target (Love, 1999). Thus, DNA recognition and gene regulation by LEF-1 involves mutual folding transitions of both protein and DNA.

Unstructured domains in transcriptional activation

In contrast to the DBDs, which for the most part exhibit rather localized folding transitions on binding to DNA, it appears that many transcriptional activation domains are both non-globular and intrinsically unstructured. There is growing evidence that many activation domains fold only upon binding to their protein targets. Thus, early NMR studies of the isolated acidic activation domains of the Vmw65 (VP16) protein of herpes simplex virus (Donaldson & Capone, 1992; O'Hare & Williams, 1992), NF-κb (Schmitz *et al*., 1994), and the glucocorticoid receptor (Dahlman-Wright *et al*., 1995) indicated a lack of structure. Similarly, the isolated kinase-inducible activation domain of the cAMP regulated transcription factor CREB and the acidic activation domains of p53 and VP16 are intrinsically disordered; however, all fold into amphipathic helical structures upon binding to their target proteins CBP, MDM2, and $TAF_{II}31$, respectively (Radhakrishnan *et al*., 1997, 1998; Kussie *et al*., 1996; Uesugi *et al*., 1997). In a particularly interesting example, the N-terminal activation domain of the yeast heat shock transcription factor has been shown to be unstructured even in the presence of the sequentially adjacent DNA-binding domain (Cho *et al*., 1996). Thus, intrinsic structural

disorder appears to be a rather common property of transcriptional activation domains.

The most detailed insights to date into the structural basis for activation domain function come from studies of the kinase-inducible activation (KID) domain of CREB. This activation domain binds with high affinity to the KIX domain (residues 586-672) of CBP, but only when it is induced by phosphorylation at Ser133 of KID (to form pKID). The pKID domain is disordered in the free state but folds into a pair of orthogonal helices (αA and αB) upon binding to the KIX domain of CBP (Radhakrishnan *et al.*, 1997) (Figure 1). Detailed NMR analysis of the intrinsic conformational propensities of the free KID domain show that the helical propensity in the αB region is very small (<10 %) in both the non-phosphorylated and phosphorylated (pKID) states (Radhakrishnan *et al.*, 1998), whereas the αA region spontaneously populates helical populations to a significant extent (~50 %). Binding of pKID to the KIX domain is entropically disfavored ($\Delta S = -6$ cal mol^{-1} K^{-1}), consistent with the coil to helix folding transition that must accompany complex formation (Parker *et al.*, 1999). Formation of the pKID complex is driven by a large enthalpy change ($\Delta H = -10.6$ kcal mol^{-1}), presumably associated with the favorable hydrogen bonding interactions made by the phosphoryl group attached to Ser133. This group makes hydrogen bonding and electrostatic interactions with tyrosine and lysine side-chains of the KIX domain as well as forming an N-cap for the αB helix (Radhakrishnan *et al.*, 1997).

The KIX domain of CBP interacts with numerous transcriptional activation domains in addition to the pKID domain of CREB. The activation domain of the proto-oncogene c-Myb binds constitutively, i.e. without requiring activation by phosphorylation, to the same site on the KIX domain as pKID. Unlike pKID, the c-Myb activation domain spontaneously adopts helical structure, even as an isolated peptide (Parker *et al.*, 1999). Binding to KIX is now entropically favored ($\Delta S = +7.5$ cal mol^{-1} K^{-1}), but the enthalpy change that accompanies binding ($\Delta H = -4.1$ kcal mol^{-1}) is smaller than for pKID, presumably because the hydrogen bonding interactions made by the phosphoryl moiety of pKID are absent for c-Myb (Parker *et al.*, 1999).

These comparative studies of the KID and c-Myb domains (Parker *et al.*, 1999) reveal the underlying molecular mechanisms of inducible and constitutive activation (Figure 2). Inducibility derives from the fact that the KID domain is intrinsically unstructured and must undergo a coil → helix induced folding transition upon binding to the KIX domain of CBP. The substantial entropic cost of folding reduces the binding affinity of the unphosphorylated KID domain, which has a dissociation constant of the order of millimolar (I. Radhakrishnan, G. Perez-Alvarado, P.E.W., unpublished results). In the activated state, however, the phosphoryl group can make numerous hydrogen bonding interactions with the target protein, leading to a large negative binding enthalpy, which compensates for the entropy loss and renders binding thermodynamically favorable. In contrast, the c-Myb

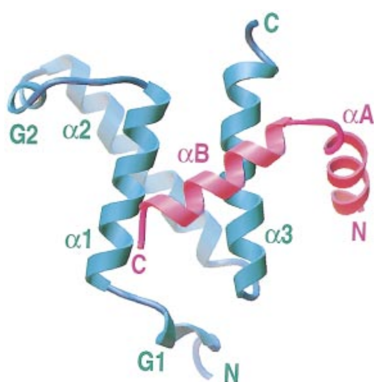

Figure 1. Ribbon diagram of the NMR structure of the complex formed between the KIX domain of the coactivator CBP and the kinase inducible activation domain (pKID) of the transcription factor CREB. The backbone of the KIX domain is shown in cyan, and that of the pKID domain is in pink. The pKID domain is intrinsically unstructured and folds into a pair of orthogonal helices upon binding to KIX. Reproduced with permission from Radhakrishnan *et al.* (1997).

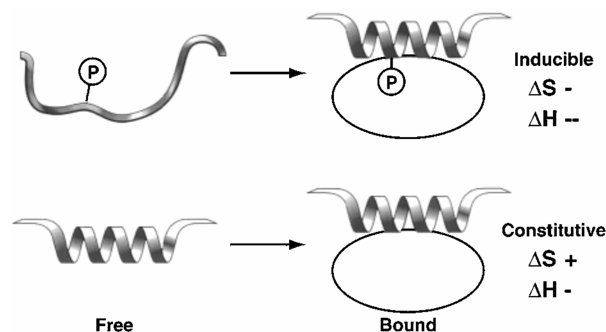

Figure 2. Schematic diagram showing how differences in intrinsic structural propensities of a protein domain can determine whether binding is constitutive or inducible (Parker *et al.*, 1999). Inducible binding is typified by the kinase inducible activation domain of CREB, which binds to the KIX domain of the coactivator CBP only in the phosphorylated (pKID) form (see the text). The free pKID domain is intrinsically unstructured, but folds on binding to its target. The entropic penalty associated with the folding transition is counterbalanced by the large enthalpy of binding, attributable in part to the complementary intermolecular hydrogen bonds formed by the phosphoserine group of pKID. In the unphosphorylated state, binding of KID is very weak since the smaller enthalpy of binding cannot compensate for the entropic cost of the folding transition. Thus, inducible binding is a consequence of the fine thermodynamic balance that arises from the coupling of folding and binding events. By contrast, the transactivation domain of the c-Myb oncoprotein is folded into a helical structure in its free state and can bind constitutively to its target protein since both the ΔH and ΔS of association are favorable (Parker *et al.*, 1999).

activation domain is intrinsically helical and binds constitutively to KIX. Thus, inducible or constitutive binding behavior reflects directly the intrinsic secondary structural propensities of the individual protein domains.

A much more complex induced folding process has been demonstrated for the N-terminal domain of *Drosophila* $TAF_{II}230$, which interacts with the concave saddle-shaped surface of the TATA box-binding protein (TBP) and inhibits DNA binding (Liu *et al*., 1998). Both NMR and CD spectra indicate that the free $TAF_{II}230$ domain is unfolded. On binding to TBP, however, it folds into a globular domain comprised of three α-helices and a small β-sheet, presenting an extensive surface that matches the hydrophobicity and charge of the TBP and mimics the properties of the exposed minor groove of the TBP-bound TATA element (Liu *et al*., 1998).

Binding-induced folding transitions have also been identified recently for the Bob-1 transcriptional coactivator (Chang *et al*., 1999). Bob-1 regulates B cell-specific expression of immunoglobulin genes by forming a ternary complex with the POU domain of Oct-1 bound to the octamer DNA element. In isolation, full-length Bob-1 has little intrinsic structure. In the ternary complex, Bob-1 makes specific contacts with both the POU domain and the DNA major groove that induce a folding transition in a 65 residue recognition domain. However, the C-terminal region, which is implicated in transcriptional activation by Bob-1, remains poorly structured even in the ternary complex.

A further, well-characterized example of the involvement of unstructured proteins in transcriptional regulation is found in the bacterial sigma factor inhibitor FlgM (Daughdrill *et al*., 1997, 1998). FlgM functions as a negative regulator of σ^{28}, a transcription factor that controls the expression of bacterial flagella and chemotaxis genes. The N-terminal half of the free FlgM closely resembles a random coil under physiological conditions, while the C-terminal region appears to form a partially collapsed state that transiently samples helical secondary structural elements (Daughdrill *et al*., 1998). Upon binding to σ^{28}, the C-terminal domain becomes more rigid and more highly structured; the N-terminal region remains unstructured and flexible, even in the complex (Daughdrill *et al*., 1997). It has been proposed that the lack of stable folded structure in FlgM, in both its free and complexed states, may facilitate its export from the cytoplasm to the exterior of the cell.

Induced folding in protein-RNA recognition

Induced folding of both protein and RNA is emerging as a common theme in protein-RNA recognition. Thus, the RNA recognition regions of the HIV-1 Rev protein and the BIV Tat protein are disordered in solution but adopt stable α-helical and β-hairpin conformations on binding to the RRE and TAR recognition elements, respectively (Tan & Frankel, 1994; Battiste *et al*., 1996; Puglisi *et al*., 1995; Ye *et al*., 1995). NMR studies of the transcriptional antitermination protein N of bacteriophage λ show that the full-length protein is disordered in solution (Mogridge *et al*., 1998). Interaction with the cognate *boxB* RNA induces helical structure in the amino-terminal RNA-binding domain, but the activating regions remain disordered and are presumed to fold in the presence of their target proteins. It has been suggested that the intrinsically disordered state of protein N might be advantageous *in vivo*; protease degradation could regulate the concentration and thereby inhibit the *boxB*-independent antitermination activity that can occur at high concentrations of N (Mogridge *et al*., 1998; Van Gilst *et al*., 1997).

Certain ribosomal proteins are known to undergo folding transitions upon RNA binding and many may depend upon an RNA scaffold for stabilization of a defined structure (Yonath & Franceschi, 1997). In the ribosomal protein L11-C76, for example, a disordered and highly flexible 15-residue loop becomes structured upon binding to RNA (Markus *et al*., 1997). An even more extreme example is the yeast ribosomal protein YP2β, which displays all of the characteristics of a molten globule state in the absence of RNA (Zurdo *et al*., 1997). Plasticity may play a key role in the assembly of complex ribonucleoprotein structures such as the ribosome, with mutually induced protein and RNA folding events directing the assembly pathway and ensuring addition of the multiple components in the correct order.

Unfolded proteins in translational processes

The initiation of mRNA translation in eukaryotes is a highly regulated process involving assembly of a multisubunit protein-RNA complex that recruits the ribosome to the initiation codon. As in transcriptional activation, coupled folding and binding events have been identified amongst proteins involved in initiation of translation. Recruitment of the small ribosomal subunit to the 5′ end of the mRNA is facilitated by eIF4G, which acts as a molecular bridge between eIF4E (the m^7GpppN cap binding protein) and the RNA helicase eIF4A. Interactions with eIF4E are mediated by a 98 residue domain of eIF4G that, according to both NMR and proteolysis experiments, is unfolded in the absence of eIF4E (Hershey *et al*., 1999). The entire domain undergoes a transition to a folded state upon binding to its eIF4E target. In mammalian cells, the association of eIF4G with eIF4E is regulated by three eIF4E-binding proteins (termed 4E-BPs), which compete with eIF4G for binding. NMR studies of 4E-BP1 show that it is intrinsically unstructured and, unlike the eIF4G binding domain, undergoes only local folding on binding to eIF4E (Fletcher *et al*., 1998). Both eIF4G and the 4E-BPs utilize a common Y-X-X-X-L-ϕ recognition motif (where ϕ = Leu, Met, or Phe) (Mader *et al*., 1995; Altmann *et al*., 1997). A recent structural study has shown that 17 residue peptides from

eIF4G and 4E-BP1 containing this motif are unstructured yet bind with high affinity to eIF4E (Marcotrigiano *et al*., 1999). Both peptides interact with the same surface of eIF4E, undergoing a folding transition on binding to form an L-shaped structure comprised of an extended chain region and a short α-helix (Figure 3). Thus, in both the full-length proteins and in peptide fragments, the eIF4E recognition motif undergoes a disorder-order transition upon binding to its target, somewhat reminiscent of the pKID-KIX interaction.

Cell cycle regulation

Cell cycle regulation is another area in which unstructured states of proteins are functionally implicated. In particular, it is thought that rapid turnover of some of these proteins, presumably facilitated by their unstructured state in the absence of the appropriate ligands, is a means of ensuring the sensitivity of the cell cycle to external conditions. One example of such a protein is the cyclin dependent kinase inhibitor $p21^{Waf1/Cip1/Sdi1}$. A recombinant construct of this protein is soluble and stable in solution in the absence of proteases, but shows no evidence by NMR or CD of the presence of secondary or tertiary structure. In itself this is a rather remarkable observation, since most globular proteins are susceptible to aggregation when unfolded in aqueous solution. Addition of the cyclin dependent kinase Cdk2 results in folding of p21 (Kriwacki *et al*., 1996a; Pavletich, 1999), a circumstance which leads to the hypothesis that the flexibility and disorder of p21 may be a vital component of its function, since it is a central player in the cell cycle, binding to a number of different components (Johnson & Walker, 1999).

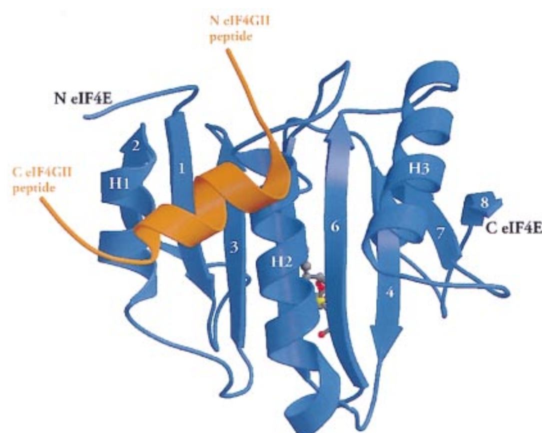

Figure 3. Structure of the ternary complex of the mRNA cap-binding protein eIF4E (blue), 7-methyl-GDP (ball-and-stick representation), and the recognition motif from eIF4GII (orange) (Marcotrigiano *et al*., 1999). The eIF4GII peptide is intrinsically unstructured but folds into an L-shaped structure comprised of a helix and an N-terminal extended chain upon binding to the convex surface of eIF4E. Figure provided by Dr S. K. Burley and reproduced with permission from Marcotrigiano *et al*. (1999).

Membrane fusion and membrane transport

Vesicle fusion in all eukaryotes is mediated by the SNARE proteins, which interact and assemble into a complex that brings the membrane surfaces into close proximity to facilitate fusion. An X-ray structure has been determined for the core synaptic SNARE complex formed by syntaxin-1A, synaptobrevin-II, and SNAP-25B; the ternary complex consists of a twisted four-helix bundle ~120 Å in length (Sutton *et al*., 1998). NMR and CD studies of the yeast homologues of syntaxin, synaptobrevin, and SNAP-25 (Sso1, Snc1, and Sec9, respectively) show that the latter two are intrinsically unstructured in isolation, while the syntaxin homolog Sso1 is only partly folded (Fiebig *et al*., 1999). Thus, assembly of the SNARE complex involves binding-induced structure formation. Such coupling of folding and binding may help to direct the assembly of the ternary complex, beginning distant from the membrane surfaces and proceeding towards them, thereby bringing the membranes into close proximity and facilitating membrane fusion (Fiebig *et al*., 1999).

Compact but incompletely folded states of proteins that contain much of the native secondary structure but lack fixed tertiary interactions are termed molten globules (Ohgushi & Wada, 1983; Dolgikh *et al*., 1981). Although these forms are most familiar in the context of protein folding, where they are thought to represent intermediates in the folding process, it has become increasingly clear that molten globules play an important role in many cellular processes. This was first suggested in 1988 by the group of Ptitsyn (Bychkova *et al*., 1988). Since then, molten globules have been shown to participate in the recognition of proteins by chaperones (Martin *et al*., 1991; van der Vies *et al*., 1992), to facilitate interaction and penetration of proteins into membranes (van der Goot *et al*., 1991; Bychkova *et al*., 1996) and to facilitate release of protein ligands (Bychkova *et al*., 1992).

A recent report suggests that unfolded states play a functional role in a protein associated with transport through membranes. The steroid acute regulatory protein (sTAR), which is involved in transport of cholesterol from the outer to the inner mitochondrial membrane, has been shown to form a molten globule in its active state (Bose *et al*., 1999). It is postulated that the molten globule state of sTAR facilitates mitochondrial entry and allows the protein to act as a switch to control cholesterol entry into the mitochondrion. Proximity to the membrane appears to be the determining factor in the formation of the molten-globule structure of the protein. Membrane transport of proteins provides an excellent example of the mechanisms by which the cell contrives to achieve desired ends by manipulating the structure of the protein. It has been suggested (Bychkova *et al*., 1996) that in the case of membrane transport,

the negative electrostatic potential and locally lowered pH in the vicinity of the membrane surface causes a local region where proteins are denatured to molten globule states, which can pass more readily through the membrane than folded, native proteins. Indeed, proteins that must be translocated to specific organelles may remain in a largely unfolded state in the cytoplasm (Randall & Hardy, 1989; Hartl & Neupert, 1990). An example of this is the photosynthetic protein plastocyanin, which folds extremely slowly *in vitro* (Koide *et al*., 1993), and which appears to remain unfolded as a high-molecular mass precursor protein before transport into the chloroplast (Merchant & Bogorad, 1986; Li *et al*., 1990).

Unfolded domains and amyloid formation

Amyloid diseases such as Alzheimer's disease and the prion diseases (BSE or ''mad cow'' disease, scrapie, Creutzfeld-Jacob disease) are characterized by the deposition of protein plaques that arise from misfolding of proteins normally found in the brain and other tissues. The prion protein, for example, is thought to undergo a conformational transition from a normal cellular form containing a preponderance of helical secondary structure to a plaque-forming conformer containing a greater proportion of β-sheet (Pan *et al*., 1993). Structure determination of fragments of the prion protein (Riek *et al*., 1996; James *et al*., 1997) revealed that ~ 100 residues at the C terminus are folded into a largely helical domain. Somewhat surprisingly, NMR studies of the full-length protein (~200 residues) indicated that the N-terminal half of the protein is completely unfolded under normal solution conditions (Donne *et al*., 1997). Partial folding of a local region containing four octapeptide repeats occurs in the presence of Cu(II) ions (Viles *et al*., 1999) and may give a clue as to the overall physiological function of the prion protein, which is at present unknown. If this protein functions as a copper storage or transport protein, the extreme flexibility of the N terminus is probably of functional significance, as the membrane-anchored protein picks up copper ions from the extracellular fluid.

Another example of an unfolded amyloidogenic protein is the NACP protein implicated in Alzheimer's disease. The NACP protein is a precursor of the non-Aβ component (NAC) of the amyloid plaque and the 14 kDa precursor protein is found by a series of biophysical experiments to be unfolded under normal solution conditions (Weinreb *et al*., 1996). As for the prion protein, this characteristic may be an important potentiator of the protein-protein interactions that lead to plaque formation and disease.

How can unfolded proteins survive in the cell?

As the above examples should serve to demonstrate, there is now overwhelming evidence that numerous eukaryotic proteins (and some prokaryotic proteins) are intrinsically unstructured or unfolded in their functional states. A critical question that immediately arises is how unfolded proteins can survive in the cell, successfully avoiding the protein degradation machinery at least for long enough to perform their cellular functions. The most likely explanation is that proteolytic degradation is tightly regulated in eukaryotic cells, being carried out by the proteasomes in response to ubiquitination or other such signals (Hochstrasser, 1996). Clearly, unfolded proteins can be functional under such circumstances, although one may suspect that the lifetime of such molecules might well be shorter than for other cellular components that are well-folded at all times. This reduced lifetime may also constitute a component of the regulation of these proteins, a possibility made more plausible by the vital roles many of them play in cell cycle regulation and in transcriptional and translational processes. Conceivably, the sensitivity and short response time implicit in many of these processes could be mediated by a system where the regulatory molecules are targeted for rapid turnover.

Rapid turnover has recently been recognized as a common means of regulation in the cell. In many cases this turnover is mediated by a PEST sequence, rich in Pro, Glu, Ser and Thr, which targets the protein for degradation (Rechsteiner & Rogers, 1996). PEST sequences are very often present as unstructured regions of the protein, and there appears to be a statistical preference for the location of these sequences in the C-terminal regions of proteins involved in cell cycle regulation. A preference has also been observed for highly charged (and therefore likely unstructured) peptide sequences such as PEST, KEKE and AQP motifs in antigenic peptides presented on MHC I molecules (Realini *et al*., 1994), which is perhaps not unexpected if these sequences predispose the protein for proteolysis.

Biological advantages of induced folding in molecular recognition and signaling

Certain biological functions, such as enzyme catalysis, immunological recognition, or molecular discrimination by receptors, absolutely demand exquisite control of three-dimensional structure. In contrast, functions such as signaling can be achieved by linear sequences, simple sequence patterns, or by isolated secondary structural motifs. Intrinsically unstructured proteins, which are induced to fold by interactions with other molecules, offer several important advantages in systems involved in cellular signaling and regulation. Unstructured proteins are inherently flexible and both their local and global structures can easily be shaped by their environment. Such intrinsic plasticity could allow a single protein to recognize a large number of biological targets without sacrificing specificity. The ability of $p21^{Waf1/Cip1/Sdi1}$ to inhibit multiple cyclin-Cdk complexes, for example, appears to stem not from the complemen-

tarity of preformed protein surfaces but from p21's ability to adopt multiple conformations that mediate different binding events (Kriwacki *et al*., 1996a). Indeed, the requirement for a folding transition upon binding of a disordered protein domain to its target can contribute significantly to the specificity of molecular recognition, as discussed in detail by Spolar & Record (1994). The thermodynamic consequence of coupled folding and binding is that free energy must be expended to bring about an induced folding transition; this comes at the cost of a reduction in the overall free energy of complex formation, so that high affinity binding can only occur when the complementarity of the target is maximal. This fine thermodynamic balance means that binding can be regulated by cellular signals or through interactions with other components of the cellular machinery. For example, binding of the CREB activation domain to its CBP target is induced by phosphorylation, which dramatically increases affinity by maximizing complementarity. In a similar way, the affinity with which an unstructured protein binds to its target can also be regulated by folding transitions induced by interactions with other proteins or nucleic acids within a multimeric complex. In this way, the pathway of assembly of multiprotein complexes and higher-order nucleoprotein complexes can be directed, with mutually induced folding events ensuring incorporation of the individual components in the correct order. Finally, the biological actions of disordered proteins are open to regulation at another level, through control of their concentration by the proteolytic machinery within the cell.

Acknowledgments

We thank current and former members of our research groups for invaluable contributions at many stages of this work. This work was supported by grants GM36643 and DK34909 (P.E.W.) and GM57374 (H.J.D.) from the National Institutes of Health and by the Skaggs Institute for Chemical Biology.

References

Aasland, R., Gibson, T. J. & Stewart, A. F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20**, 56-59.

Altmann, M., Schmitz, N., Berset, C. & Trachsel, H. (1997). A novel inhibitor of cap-dependent translation initiation in yeast: p20 competes with eIF4G for binding to eIF4E. *EMBO J.* **16**, 1114-1121.

Battiste, J. L., Mao, H. Y., Rao, N. S., Tan, R. Y., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). α-Helix-RNA major groove recognition in an HIV-1 Rev peptide RRE RNA complex. *Science*, **273**, 1547-1551.

Bösch, C., Bundi, A., Oppliger, M. & Wüthrich, K. (1978). ^{1}H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur. J. Biochem.* **91**, 209-214.

Bose, H. S., Whittal, R. M., Baldwin, M. A. & Miller, W. L. (1999). The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. *Proc. Natl Acad. Sci. USA*, **96**, 7250-7255.

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

Bychkova, V. E., Pain, R. H. & Ptitsyn, O. B. (1988). The 'molten globule' state is involved in the translocation of proteins across membranes? *FEBS Letters*, **238**, 231-234.

Bychkova, V. E., Berni, R., Rossi, G. L., Kutyshenko, V. P. & Ptitsyn, O. B. (1992). Retinol-binding protein is in the molten globule state at low pH. *Biochemistry*, **31**, 7566-7571.

Bychkova, V. E., Dujsekina, A. E., Klenin, S. I., Tiktopulo, E. I., Uversky, V. N. & Ptitsyn, O. B. (1996). Molten globule-like state of cytochrome *c* under conditions simulating those near the membrane surface. *Biochemistry*, **35**, 6058-6063.

Chang, J. F., Phillips, K., Lundbäck, T., Gstaiger, M., Ladbury, J. E. & Luisi, B. (1999). Oct-1 POU and octamer DNA co-operate to recognise the Bob-1 transcription co-activator *via* induced folding. *J. Mol. Biol.* **288**, 941-952.

Cho, H. S., Liu, C. W., Damberger, F. F., Pelton, J. G., Nelson, H. C. M. & Wemmer, D. E. (1996). Yeast heat shock transcription factor N-terminal activation domains are unstructured as probed by heteronuclear NMR spectroscopy. *Protein Sci.* **5**, 262-269.

Dahlman-Wright, K., Baumann, H., McEwan, I. J., Almlöf, T., Wright, A. P. H., Gustafsson, J.-Å. & Härd, T. (1995). Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. *Proc. Natl Acad. Sci. USA*, **92**, 1699-1703.

Darimont, B. D., Wagner, R. L., Apriletti, J. W., Stallcup, M. R., Kushner, P. J., Baxter, J. D., Fletterick, R. J. & Yamamoto, K. R. (1998). Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev.* **12**, 3343-3356.

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

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

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

Donaldson, L. & Capone, J. P. (1992). Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. *J. Biol. Chem.* **267**, 1411-1414.

Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (1997). Structure of the recombinant full-length hamster prion protein PrP(29-

231): the N terminus is highly flexible. *Proc. Natl Acad. Sci. USA,* **94**, 13452-13457.

Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C. & Villafranca, J. E. (1998). Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.* **3**, 473-484.

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

Fiebig, K. M., Rice, L. M., Pollock, E. & Brunger, A. T. (1999). Folding intermediates of SNARE complex assembly. *Nature Struct. Biol.* **6**, 117-123.

Fletcher, C. M., McGuire, A. M., Gingras, A. C., Li, H. J., Matsuo, H., Sonenberg, N. & Wagner, G. (1998). 4E binding proteins inhibit the translation factor eIF4E without folded structure. *Biochemistry,* **37**, 9-15.

Garner, E., Cannon, P., Romero, P., Obradovic, Z. & Dunker, A. K. (1998). Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. *Genome Inform.* **9**, 201-214.

Gerstein, M. (1998). How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold. Design,* **3**, 497-512.

Hartl, F.-U. & Neupert, W. (1990). Protein sorting to mitochondria: evolutionary conservations of folding and assembly. *Science,* **247**, 930-938.

Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature,* **387**, 733-736.

Hershey, P. E., McWhirter, S. M., Gross, J. D., Wagner, G., Alber, T. & Sachs, A. B. (1999). The cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1. *J. Biol. Chem.* **274**, 21297-21304.

Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405-439.

Holmbeck, S. M. A., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J. & Wright, P. E. (1998a). High resolution solution structure of the retinoid X receptor DNA binding domain. *J. Mol. Biol.* **281**, 271-284.

Holmbeck, S. M. A., Dyson, H. J. & Wright, P. E. (1998b). DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J. Mol. Biol.* **284**, 533-539.

James, T. L., Liu, H., Ulyanov, N. B., Farr-Jones, S., Zhang, H., Donne, D. G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S. B. & Cohen, F. E. (1997). Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc. Natl Acad. Sci. USA,* **94**, 10086-10091.

Janknecht, R. & Hunter, T. (1996). Transcriptional control: versatile molecular glue. *Curr. Biol.* **6**, 951-954.

Jeanmougin, F., Wurtz, J.-M., Le, Douarin B., Chambon, P. & Losson, R. (1997). The bromodomain revisited. *Trends. Biochem. Sci.* **22**, 151-153.

Johnson, D. G. & Walker, C. L. (1999). Cyclins and cell cycle checkpoints. *Annu. Rev. Pharmacol. Toxicol.* **39**, 295-312.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. W. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature,* **181**, 662-666.

Koide, S., Dyson, H. J. & Wright, P. E. (1993). Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry,* **32**, 12299-12310.

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

Kriwacki, R. W., Wu, J., Siuzdak, G. & Wright, P. E. (1996b). Probing protein/protein interactions with mass spectrometry and isotopic labeling: analysis of the p21/Cdk2 complex. *J. Am. Chem. Soc.* **118**, 5320-5321.

Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J. & Pavletich, N. P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science,* **274**, 948-953.

Li, H.-M., Theg, S. M., Bauerle, C. M. & Keegstra, K. (1990). Metal-ion-center assembly of ferredoxin and plastocyanin in isolated chloroplasts. *Proc. Natl Acad. Sci. USA,* **87**, 6748-6752.

Liu, D., Ishima, R., Tong, K. I., Bagby, S., Kokubo, T., Muhandiram, D. R., Kay, L. E., Nakatani, Y. & Ikura, M. (1998). Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell,* **94**, 573-583.

Logan, T. M., Thériault, Y. & Fesik, S. W. (1994). Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.* **236**, 637-648.

Love, J. J. (1999). Biophysical characterization of the HMG-1 box domain of the lymphoid enhancer binding factor-1. PhD thesis, University of California, San Diego.

Love, J. J., Li, X. A., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature,* **376**, 791-795.

Mader, S., Chen, J.-Y., Chen, Z., White, J., Chambon, P. & Gronemeyer, H. (1993). The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificities of the DNA binding domains. *EMBO J.* **12**, 5029-5041.

Mader, S., Lee, H., Pause, A. & Sonenberg, N. (1995). The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol. Cell Biol.* **15**, 4990-4997.

Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1999). Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell,* **3**, 707-716.

Markus, M. A., Hinck, A. P., Huang, S. R., Draper, D. E. & Torchia, D. A. (1997). High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nature Struct. Biol.* **4**, 70-77.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L. & Hartl, F.-U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. *Nature,* **352**, 36-42.

Merchant, S. & Bogorad, L. (1986). Rapid degradation of apoplastocyanin in Cu(II)-deficient cells of *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **261**, 15850-15853.

Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E. & Greenblatt, J. (1998). Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol. Cell,* **1**, 265-275.

Nolte, R. T., Wisely, G. B., Westin, S., Cobb, J. E., Lambert, M. H., Kurokawa, R., Rosenfeld, M. G., Willson, T. M., Glass, C. K. & Milburn, M. V. (1998). Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature,* **395**, 137-143.

Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell,* **87**, 953-959.

O'Hare, P. & Williams, G. (1992). Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using ^{1}H NMR. *Biochemistry,* **31**, 4150-4156.

Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side chains. *FEBS Letters,* **164**, 21-24.

O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990). Design of DNA-binding peptides based on the leucine zipper motif. *Science,* **249**, 774-777.

Pan, K.-M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R. J., Cohen, F. E. & Prusiner, S. B. (1993). Conversion of α-helices into β-sheets features in the formation of the scrapie prion proteins. *Proc. Natl Acad. Sci. USA,* **90**, 10962-10966.

Parker, D., Rivera, M., Zor, T., Henrion-Caude, A., Radhakrishnan, I., Kumar, A., Shapiro, L. H., Wright, P. E., Montminy, M. & Brindle, P. K. (1999). Role of secondary structure in discrimination between constitutive and inducible activators. *Mol. Cell Biol.* **19**, 5601-5607.

Pavletich, N. P. (1999). Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J. Mol. Biol.* **287**, 821-828.

Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M. & Smith, L. J. (1997). NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *J. Mol. Biol.* **274**, 152-159.

Perlmann, T., Rangarajan, P. N., Umesono, K. & Evans, R. M. (1993). Determinants for selective RAR and TR recognition of direct repeat HREs. *Genes Dev.* **7**, 1411-1422.

Ponting, C. P., Blake, D. J., Davies, K. E., Kendrick-Jones, J. & Winder, S. J. (1996). ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins. *Trends Biochem. Sci.* **21**, 11-13.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science,* **270**, 1200-1203.

Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R. & Wright, P. E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell,* **91**, 741-752.

Radhakrishnan, I., Pérez-Alvarado, G. C., Dyson, H. J. & Wright, P. E. (1998). Conformational preferences in the Ser^{133}-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Letters,* **430**, 317-322.

Randall, L. L. & Hardy, S. J. (1989). Unity in function in the absence of consensus in sequence: role of leader peptides in export. *Science,* **243**, 1156-1159.

Rastinejad, F., Perlmann, T., Evans, R. M. & Sigler, P. B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature,* **375**, 203-211.

Realini, C., Rogers, S. W. & Rechsteiner, M. (1994). KEKE motifs: proposed roles in protein-protein association and presentation of peptides by MHC class I receptors. *FEBS Letters,* **348**, 109-113.

Rechsteiner, M. & Rogers, S. W. (1996). PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267-271.

Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. & Wüthrich, K. (1996). NMR structure of the mouse prion protein domain PrP(121-231). *Nature,* **382**, 180-182.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S. & Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* **3**, 437-448.

Schmitz, M. L., dos Santos, Silva M. A., Altmann, H., Czisch, M., Holak, T. A. & Baeuerle, P. A. (1994). Structural and functional analysis of the NF-kappa B p65 C terminus. An acidic and modular transactivation domain with the potential to adopt an alpha-helical conformation. *J. Biol. Chem.* **269**, 25613-25620.

Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A. & Greene, G. L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell,* **95**, 927-937.

Shikama, N., Lyon, J. & La Thangue, N. B. (1997). The p300/CBP family: integrating signals with transcription factors and chromatin. *Trends Cell Biol.* **7**, 230-236.

Spolar, R. S. & Record, M. T. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science,* **263**, 777-784.

Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature,* **395**, 347-353.

Tan, R. & Frankel, A. D. (1994). Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry,* **33**, 14579-14585.

Torchia, J., Rose, D. W., Inostroza, J., Kamei, Y., Westin, S., Glass, C. K. & Rosenfeld, M. G. (1997). The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature,* **387**, 677-684.

Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J. & Verdine, G. L. (1997). Induced α helix in the VP16 activation domain upon binding to a human TAF. *Science,* **277**, 1310-1313.

van der Goot, F. G., González-Mañas, J. M., Lakey, J. H. & Pattus, F. (1991). A 'molten-globule' membrane-insertion intermediate of the pore-forming domain of colicin A. *Nature,* **354**, 408-410.

van der Vies, S. M., Viitanen, P. V., Gatenby, A. A., Lorimer, G. H. & Jaenicke, R. (1992). Conformational states of ribulosebisphosphate carboxylase and

their interaction with chaperonin 60. *Biochemistry*, **31**, 3635-3644.

Van Gilst, M. R., Rees, W. A., Das, A. & von Hippel, P. H. (1997). Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry*, **36**, 1514-1524.

van Tilborg, P. J. A., Mulder, F. A. A., De Backer, M. M. E., Nair, M., Van Heerde, E. C., Folkers, G., van der Saag, P. T., Karimi-Nejad, Y., Boelens, R. & Kaptein, R. (1999). Millisecond to microsecond time scale dynamics of the retinoid X and retinoic acid receptor DNA-binding domains and dimeric complex formation. *Biochemistry*, **38**, 1951-1956.

Viles, J. H., Cohen, F. E., Prusiner, S. B., Goodin, D. B., Wright, P. E. & Dyson, H. J. (1999). Copper binding to the prion protein: structural implications of four identical cooperative binding sites. *Proc. Natl Acad. Sci. USA*, **96**, 2042-2047.

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

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

Wootton, J. C. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269-285.

Ye, S., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* **2**, 827-840.

Yonath, A. & Franceschi, F. (1997). New RNA recognition features revealed in ancient ribosomal proteins. *Nature Struct. Biol.* **4**, 3-5.

Zhang, J. & Matthews, C. R. (1998). Ligand binding is the principal determinant of stability for the $p21^{H\text{-}ras}$ protein. *Biochemistry*, **37**, 14881-14890.

Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994). Backbone ^{1}H and ^{15}N 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*, **4**, 845-858.

Zhou, P., Sun, L. J., Dötsch, V., Wagner, G. & Verdine, G. L. (1998). Solution structure of the core NFATC1/DNA complex. *Cell*, **92**, 687-696.

Zurdo, J., Sanz, J. M., Gonzalez, C., Rico, M. & Ballesta, J. P. (1997). The exchangeable yeast ribosomal acidic protein YP2beta shows characteristics of a partly folded state under physiological conditions. *Biochemistry*, **36**, 9625-9635.