Intrinsic Disorder and Protein Function

A. Keith Dunker,[‡,]* Celeste J. Brown,[‡] J. David Lawson,[‡]
Lilia M. Iakoucheva,[‡] and Zoran Obradović[§]

[‡]School of Molecular Biosciences
Washington State University, Pullman, WA 99164
[§]Center for Information Science and Technology
Temple University, Philadelphia, PA 19122

*Corresponding Author:
A. Keith Dunker
School of Molecular Biosciences
Washington State University
Pullman, WA 99164-4660
Telephone: (509) 335-5322
Facsimile: (509) 335-9688
E-mail: dunker@disorder.chem.wsu.edu

†This research was funded by Molecular Kinetics, Inc., by NSF-CSE-IIS-9711532 and NSF-IIS-0196237 to Z.O. and A.K.D and by NIH-1R01-LM06916 to A.K.D. and Z.O.

Key words: native protein, random coil, molten globule, intrinsic disorder, protein structure-function, genomics, proteomics, and bioinformatics

Abbreviations: circular dichroism (CD), limited proteolysis (LP), nuclear magnetic resonance (NMR), nuclear Overhauser effect (NOE), predictor of natural disordered regions (PONDR), Protein Data Bank (PDB), Disordered Protein Database (DisProt)

Textual Footnotes:
[1]This predictor can be accessed at http://www.ponder.com.
(footnote for page 4)

The dominant view of protein structure-function is that an amino acid sequence specifies a (mostly) fixed three-dimensional (3-D) structure that is a prerequisite to protein function. In contrast to the dominant view, many proteins display functions *requiring* the disordered state. Our purpose here is to provide a catalogue of disorder-function relationships. The very important molecular details in each example can be obtained from the references provided or from several excellent reviews and commentaries (*1-9*).

For ordered protein, the ensemble members all have the same time-averaged canonical set of Ramachandran angles along their backbones. For intrinsically disordered protein, the ensemble members have different (and typically dynamic) Ramachandran angles. Such disorder has been characterized by a variety of methods including x-ray crystallography, NMR spectroscopy, CD spectroscopy, and protease sensitivity to name several. Each of these methods has advantages and limitations that are discussed in more detail elsewhere (*10*). Although a few disordered proteins and regions have been characterized by several methods as noted below, it would be useful to have more examples with multiple methods of characterization.

In attempts to discover generalities from the known disorder examples, we recently used bioinformatics coupled with data mining (*11-15*). The results suggested that thousands of natively disordered proteins exist, representing a very substantial fraction of the proteins in the commonly used sequence databases (*13, 16*). From these and related database predictions and from a set of functionally important disordered proteins, Wright and Dyson (*17*) called for a re-assessment of the view that 3-D structure is always a prerequisite to protein function.

In this article, we discuss the following topics: 1. how common is intrinsic disorder?; 2. intrinsic disorder in vivo; 3. functional annotations for 90 proteins having physically characterized regions of disorder; 4. disordered regions without known function 5. a structure-function proposal called "the protein trinity"; 6. the functional repertoires of ordered and disordered protein, and 7. the need for a Disordered Protein Database (DisProt) to complement the Protein Data Bank (PDB).

How Common is Intrinsic Disorder? A series of predictors of natural disordered regions (PONDRs) have been developed using amino acid sequence as inputs and giving intrinsic order or disorder tendencies as outputs (*11, 14, 15, 18, 19*). The various PONDRs are distinguished by different training sets, by different data representations for their inputs, and by different machine learning models for their development.

For PONDR VL-XT[1], currently the best characterized of the PONDRs, only 6% of more than 900 non-homologous proteins spanning PDB gave false positive predictions of disorder ≥ 40 consecutive amino acids in length. Even this 6% may be an over-estimate of the false positive error rate, however, because many of these predicted disordered regions are involved in ligand binding or in crystal contacts (Lawson and Dunker, unpublished data). For several of these, the binding regions are intrinsically disordered when the parent proteins are in solution without their ligands. Furthermore, only ~0.4% of more than 220,000 ordered residues from these 900 non-homologous proteins were falsely predicted to be in long disordered regions (*17*). In contrast ~11% of a dataset containing more than 17,000 disordered residues from over 140 proteins were falsely predicted to be in correspondingly long ordered regions. Because the estimated false negative error rate is much greater than the estimated false positive rate, PONDR VL-XT probably underpredicts the occurrence of long disordered regions in nature.

PONDR VL-XT has been applied to the proteomes from more than 30 organisms. The estimates of disorder in the various genomes were as follows: bacteria = 9 to 37%, archaea = 9 to 37%, eukaryota = 35 to 51% (*20*). A possible reason for the increased disorder in the eukaryota as compared to the prokaryota is the greater need for protein-mediated signaling, regulation and control in the former (*21*). Other explanations, such as the possibility that the increased disorder in eukaryotes is an artifact arising from training set biases, also need to be explored.

Intrinsic Disorder in vivo. Although Karush's report on the binding plasticity of serum albumin pointed out the importance of structural ensembles for function as early as 1950 (*22*), and although many additional examples have been featured in literature reviews for more than a decade (*1-3, 9*), intrinsic disorder has been mostly ignored as evidenced by the lack of even one carefully discussed example in any of the popular biochemistry text books. Perhaps one reason for this omission was a belief that disorder was an artifact because protease digestion would eliminate such proteins *in vivo*. Kim and Frankel, however, argued that sequestering proteases by compartmentalization would enable protein disorder to exist *in vivo* (*3*). Tight regulation of intracellular protease activity can also be understood in terms of protecting protease sensitive sites as discussed previously (*10, 17*).

Additional mechanisms help disordered proteins avoid proteolysis *in vivo*. Some disordered regions are inaccessible to proteases due to steric factors; other disordered regions lack protease sensitive residues; and still other disordered regions may exist only transiently as they jump from one partner to another. For example, at low calcium concentrations *in vitro*, calmodulin becomes sensitive to protease digestion (*23*). As calcium concentration drops *in vivo*, calmodulin leaves its enzymatic partners and binds to proteins with IQ motifs (*24*). Similarly, some chaperones may function as bodyguards for the protection of intrinsically disordered regions in addition to their proposed function of promoting protein folding (*10*).

The protection of disorder by molecular association predicts that removal of a partner by genetic manipulation would lead to increased protease sensitivity of the remaining, unprotected protein. Indeed, in genetic dissection of protein signaling pathways, deletion of a known partner often leads to the disappearance of the remaining protein, although the cause of the disappearance has typically not been determined nor the possibility of increased protease sensitivity considered (Kendal Broadie and Jessica Tyler, personal communications). Here we suggest that every protein-protein interaction identified by the yeast-two-hybrid or other assays should be tested for the involvement of disorder by determining whether deletion of one partner leads to increased *in vivo* protease sensitivity (i.e. increased ubiquitination) of the other.

Some proteins that are highly disordered in the laboratory have short lifetimes in the cell for functional reasons. These examples, including proteins that participate in critical cellular control mechanisms, provide a further argument for the existence of disorder *in vivo* (*17*).

Molecular crowding favors more compact forms over more extended ones and has been shown to markedly shift the equilibrium toward the folded state (*25*). Two observations, however, suggest that molecular crowding inside the cell probably cannot be used to rule out the existence of intrinsic disorder *in vivo*. First, steric factors would prevent many complexes from folding before binding, so for such structures, binding and folding must be concomitant. Examples of concomitant binding and folding include calmodulin binding to its target helix (*26*) and TFIIIA binding to its target DNA (*27*). In both cases the protein wraps around its partner. Second, molecular crowding should be able to induce folding into a particular 3-D structure only

if the protein were to have an appropriately shaped energy landscape. Proteins with functions that depend on lack of folding would be expected to have evolved energy landscapes incommensurate with folding into a specific structure. Consistent with this idea, even though molecular crowding can induce a reduction in hydrodynamic radius (*28*), such crowding does not induce regular secondary structure for three intrinsically disordered proteins: c-Fos (*29*), p27kip1 (*30*), and α-synuclein (*28*).

For ligand-binding proteins, the energy landscapes change in the presence of their partners, leading to disorder-to-order transitions upon binding. A common theme is a coil-to-helix transition concomitant with binding to another molecule (*10*). Molecular crowding could certainly shift a disordered region's equilibrium from coil to helix. However, in the absence of its binding partner, a helix induced by excluded volume effects would probably not assume a unique tertiary structure, but rather would transiently sample many orientations relative to the main body of the protein.

Finally, we have recently compared the evolutionary rates of ordered and disordered regions that exist within the same protein. Faster rates of evolution were observed for several disordered regions (*31*). Functionally crucial side chain interactions within the ordered cores are thought to be responsible for slow rates of evolutionary change for ordered proteins. The faster rates of evolution imply that disordered regions have a lack of crucial side chain interactions, and thus provide additional support for the existence of disorder *in vivo*.

Functions of Intrinsic Disorder. Our searches have identified more than 140 proteins that, under apparently native conditions, contain disordered regions of 30 consecutive residues or longer. Careful literature searches were carried out for a portion of these proteins (Table 1). Twenty-eight separate functions were identified for 98 out of 115 disordered regions (Table 2). Cataloguing the functions of the disordered regions in these proteins provides insight regarding relationships between intrinsic disorder and protein function.

The proteins described in Table 1 contain numerous disordered regions involved in molecular recognition. These are indicated in the function column of Table 1 by the letters (a-j) and are characterized further in Table 2. Some of these molecular interactions involve binding to other proteins (a), such as to kinases, transcription factors, and translation inhibitors, while others involve binding to nucleic acid polymers including DNA (b), rRNA (cR), tRNA (cT), mRNA (cM) and genomic RNA (cG). Some of the DNA binding regions are also involved in DNA unwinding (t) and bending (u). Membrane-associating peptides are often disordered in solution and acquire helical secondary structure upon binding to the membrane (f). Protein-protein interactions leading to polymer formation often involve intrinsically disordered regions (g) as has been recently reviewed elsewhere (*207*). Several receptors undergo disorder-to-order transitions upon binding their ligands (h). Likewise, enzymes can have mobile regions that become structured upon binding their substrate (h) or cofactor (i). Other proteins undergo disorder-to-order transitions upon binding hemes (i) or metal ions (j).

Given our conditioning to think in terms of the lock-and-key model, the involvement of intrinsic disorder in molecular recognition at first seems counterintuitive. Upon further reflection, however, the use of intrinsic disorder for molecular recognition exhibits several important features, such as 1. enabling high specificity coupled with low affinity (*208*) because the free energy arising from the contacts of protein with ligand is reduced by the free energy needed to fold the intrinsic disorder; 2. enabling one molecule to bind to differently shaped

partners by structural accommodations at the binding interfaces (*22, 67, 209*); 3. enabling different disordered sequences to fold (perhaps differently) in order to bind to a common binding site (*206*); 4. enabling the creation of very large interaction surfaces as the disordered protein wraps-up (*27*) or surrounds its partner (*26*); 5. enabling faster rates of association by reducing dependence on orientation factors and by enlarging target sizes (*210*); and 6. enabling faster rates of dissociation by unzippering mechanisms (*10*).

Chemical modification of side chains requires close association between the protein being modified and the enzyme doing the modification. If the side chain being modified were within a structured region, steric factors would typically prevent or slow down the association between the enzyme and target. On the other hand, locating the side chain undergoing modification within a disordered region facilitates substrate binding because the disordered region can fold onto the surface of the modifying enzyme. Thus, several types of chemical modification have been found to occur in intrinsically disordered regions. These are indicated as (mA - mR) in Tables 1 and 2 and include acetylation (mA), fatty acid acylation (mF), glycosylation (mG), methylation (mM), phosphorylation (mP), and ADP-ribosylation (mR). An open question is whether chemical modification *universally requires* regions of intrinsic disorder just prior to association with the modifying enzymes.

Disordered regions involved in regulating the activity of their parent protein are designated as autoregulatory (k) in Tables 1 and 2. Many of these regions are modified as part of the regulatory process (*e.g.* phosphorylated or acetylated). As noted above, disorder facilitates such modifications. Other regions regulate activity through differential binding (*e.g.* psuedosubstrates). Again, disordered regions are useful for such functions because of their lower binding affinity (the off rates are thus compatible with quick adaptation to signal changes) and binding plasticity (by definition, the region must bind multiple partners).

Some disordered regions carry out function without the necessity of assuming an ordered state, indicated as (n-q) in Tables 1 and 2. Flexible linkers/spacers (n) between domains (*146*) comprise a major group of this type. Flexible linkers allow two domains to move relative to each other, with such movement being essential for function. Some of these linkers also act as spacers that regulate the distance between adjacent domains (*155*). The functional, native state of flexible linkers/spacers is likely to be a random coil, or the polypeptide approximation of the random coil (*211*). Alternatively, the linkers/spacers can have local or transient secondary structure (*212*), but in either case, such regions can carry out function without necessarily undergoing a disorder-to-order transition. A similar lack of a requirement for an ordered state characterizes proteins that function as entropic springs (o), entropic bristles (p) and entropic clocks (q).

Careful *in vitro* studies demonstrate that disordered regions undergo protease digestion orders of magnitude faster than do ordered regions (*206, 213, 214*). Indeed, when digestion occurs in an ordered region of a protein, local unfolding, not just surface exposure, is necessary for efficient *in vitro* proteolysis (*215*). The likely explanation is similar to that given above for chemical modification. That is, the polypeptide segment being cleaved must form a specific structure on the protease surface; disordered regions can almost always fold correctly onto the protease surface while ordered regions typically don't (and maybe can't) have the right shape. Given these extensive studies on the importance of disorder for *in vitro* protease digestion, it is not surprising to find examples of intrinsic disorder that are associated with *in vivo* proteolysis, indicated as (l) in Tables 1 and 2.

Numerous proteins comprising the ribosome are disordered or have disordered regions when removed from the ribosome. For example, CD studies of individual ribosomal proteins from *Escherichia coli* show that 10 large- or small-subunit proteins are substantially disordered (*148*). X-ray crystallography of the large ribosomal subunit from *Haloarcula marismortui* confirms that many of these proteins occur in an extended form within the ribosome (*216*). The structure of the *H. marismortui* ribosome also indicates that several proteins have both globular domains and extended regions. The authors suggest that the proteins of the ribosome act as mortar (indicated as (r) in Tables 1 and 2), filling the gaps and cracks between loops of the rRNA. For these proteins, binding does not induce a disorder-to-order transition in the typical sense of formation of globular structure, but rather it involves the capture of one member from the ensemble of extended structures. Aside from being a structural mortar, ribosomal proteins may help to alleviate rRNA misfolding by selecting the native RNA structure from the ensemble of nonnative forms (*217*). Extended, disordered protein can make multiple contacts that bridge multiple folded RNA domains, thereby assuaging RNA misfolding better than globular proteins.

Two functions of disordered regions that cannot be readily grouped with other functions are protein detergent (v) and self-transport through membrane channels (s). Both clusterin and casein have detergent-like properties. Clusterin's promiscuous binding to hydrophobic moieties depends on a molten globular domain (*218*), while casein forms micellular aggregates that solubilize hydrophobically aggregated proteins (*55*). The disordered regions of flagellin and the flagellum specific σ binding factor, FlgM, both exist in part to allow the movement of the proteins through the narrow channel in the core of the flagellum. Disorder is necessary for flagellin to move through the tube to the tip of the growing filament where polymerization occurs (*85*). Likewise, FlgM's disorder facilitates its export through the flagellum, thus controlling FlgM's intracellular concentration, which in turn regulates the transcription of more than 50 genes associated with flagellar assembly (*86*).

Another evolutionary niche occupied by disordered regions has come to light in the study of *Staphylococcus aureus* fibronectin binding protein. This protein is completely disordered in solution (*78*). The regions involved in binding fibronectin, however, undergo disorder-to-order transitions upon forming this high-affinity interaction (*79*). Interestingly, the flexibility of this protein was proposed to account for the lack of an immune response during infection by *S. aureus* (*219*) because highly flexible polypeptides are evidently not immunogenic (Höök, personal communication). Indeed, highly flexible attachment domains have been reported for a number of pathogens, suggesting that such use of disorder may be a general strategy (*10*).

Disordered Regions Without Known Function. Given the wide range of functions already discovered or proposed for intrinsically disordered proteins, from molecular recognition (*208*) to protection against desiccation (*75*) to detergent action (*55, 218*), functional annotation may turn out to be especially vexing for proteins with intrinsically disordered regions of sequence. Tables 1 and 2 include 6 disordered regions claimed to lack essential function (w) as well as 11 proteins with 17 disordered regions with unknown functions (x). With regard to the latter group, 9 of the 10 proteins have enzymatic functions; the other protein is the signal transduction inhibitor called RGS4. The disordered regions in the enzymes are not at their respective catalytic sites, suggesting the possibility of regulatory functions. With regard to the former 6 disordered regions claimed to lack essential function, the regions in question were cleaved, deleted, or otherwise removed from the protein without affecting its activity. An experiment ruling out one particular

function, however, does not rule out all possible functions. In this regard, moonlighting proteins have functions completely unrelated to their originally discovered functions (*220*). Given the potential for moonlighting, proving that a disordered region has no function at all would be extremely difficult. These and other intrinsic disorder orphans may provide useful starting points for the discovery of novel disorder-associated functions.

The Protein Trinity. Intrinsic disorder is common (*13, 16, 20*), and this paper supports its importance for function as discussed previously (*10*). To organize these observations, we are proposing a structure-function relationship called the protein trinity (Figure 1). According to this proposal, *native proteins* can adopt any one of 3 states: fully folded, collapsed, or extended. The type of native structure for a given protein is determined by its amino acid sequence, the presence or absence of chemical modification, the presence or absence of ligands, and the conditions of the surrounding medium. Since the various factors affecting the type of native structure are under dynamic control inside the cell, the native form of the protein can change.

With regard to function, it can arise from any one of the three states. Examples involving fully folded proteins need no further discussion. Functions involving the extended-disordered state include flexible linkers or spacers, which allow the connected domains to move relative to each other in order to allow simultaneous binding to two or more components with separations that vary over space and time (*101, 146, 164, 175*), entropic bristles, which use excluded volume effects to keep molecules apart (*131, 132*), entropic springs, which provide a restoring force resulting from randomization of bond torsion angles that become restricted upon stretching (*169-172*), entropic clocks, which provide a timing mechanism (*10*) arising from random searches such as that observed for the ball-and-chain model for closure of voltage-gated ion channels (*205*). Functions involving collapsed disorder include the binding of hydrophobic groups by dynamic detergent-like proteins (*218*) and the multiple binding of ions by static polymorphic ensembles (*221*).

Disorder-to-order transitions upon binding, whether starting from collapsed-disorder or from extended-disorder, are an important use of disorder for function. Such interactions and transitions have been extensively reviewed elsewhere (*8, 17*).

Functional Repertoires of Ordered and Disordered Proteins: The functions listed above for disordered protein can be organized into four broad categories: molecular recognition, molecular assembly, protein modification, and entropic chains. These will be discussed in reverse and compared with the functions of ordered proteins.

Entropic chains carry out functions that depend directly on the disordered state, and so such functions are simply outside the capabilities of fully folded structures. The particular entropic chain functions identified so far – linkers/spacers, bristles, springs, and clocks – are unlikely to represent the complete set for this protein form. Future efforts in this area may lead to the identification of new functions that depend on disordered chains.

Sites of protein modification, whether by chemical additions or protease cleavage, evidently occur with very strong preference for regions of disorder. Perhaps the few modification sites identified as ordered beforehand actually undergo order-to-disorder transitions just prior to the modification. Such events would be revealed experimentally as kinetic delays associated with the local unfolding events. Furthermore, these kinetic delays should be sensitive to low amounts of urea or guanidine. Testing for the universality of locating protein modification sites in regions of disorder would be important.

The use of partially folded subunits for molecular assembly appears to have significant advantages compared to the use of ordered subunits. For an assembly based on ordered subunits, the pair-wise interactions would unlikely be precise enough to bring about closure when the last subunit of the assembly has to associate with the first, not to mention the steric problems of inserting the last subunit. On the other hand, partially folded subunits would have the flexibility to form an overall assembly, which would then be followed by disorder-to-order transitions to give a tight overall complex. For more details, a recent review of the involvement of intrinsic disorder in molecular assembly should be consulted (*207*).

Of the four broad functional categories defined above, molecular recognition appears to be a common function of both ordered and disordered proteins. Molecular recognition by disordered proteins may be primarily used for signaling whereas recognition by ordered proteins may be primarily used for catalysis (Dunker and Lawson, manuscript in preparation). Simply put, disordered regions can bind to multiple targets with low affinity – ideal properties for signal transduction. However, binding by a flexible region cannot lead to efficient catalysis because much of the binding energy is used for folding and so would be unavailable for inducing the transition state. In this view, enzymes, which are over-represented in PDB and which dominate our thinking about protein structure and function, have evolved to be completely folded in order to carry out catalysis efficiently, not for the molecular recognition aspects of their functions.

The Need for a Disordered Protein Database: As foretold many years ago (*222*), one of the most important strategies for identifying the function of a new protein is by sequence matching. For this reason, and given the specific examples discussed herein, having an annotated database of intrinsically disordered protein is crucial to current structural genomics and proteomics efforts.

Tables 1 and 2 provide a start for a database, but a major expansion and a transformation from a flat file into an appropriate architecture are obviously both needed. With regard to the expansion phase, we hope that researchers will submit their disorder examples via our recently established website: http://DisProt.wsu.edu. We especially encourage submission by researchers working on intrinsically disordered proteins, but will also accept submissions from anyone who happens to know about a specific example. Once an expanded set of functionally annotated, intrinsically disordered proteins is in place, DisProt will be converted into a searchable database.

Acknowledgments

Ya-Yue Van is thanked for suggesting the name “the protein trinity,” Clare Woodward for encouraging the writing of this review, and the anonymous reviewer for many helpful comments. This work was supported by grants from NSF (CSE-IIS-971152) and NIH (1R01 LM06916).

References

1. Bennett, W. S., and Huber, R. (1984) *Crit. Rev. Biochem. 15*, 291-384.
2. Sigler, P. B. (1988) *Nature 333*, 210-212.
3. Frankel, A. D., and Kim, P. S. (1991) *Cell 65*, 717-719.
4. Bychkova, V., and Ptitsyn, O. (1993) *Chemtracts - Biochem. Mol. Biol. 4*, 133-163.
5. Triezenberg, S. J. (1995) *Curr. Opin. Genet. Dev. 5*, 190-196.
6. Plaxco, K. W., and Gross, M. (1997) *Nature 386*, 657, 659.
7. Uversky, V. N. (2002) *Eur J Biochem 269*, 2-12.
8. Dyson, H. J., and Wright, P. E. (2002) *Curr. Opin. Struct. Biol. 12*, 54-60.
9. Churchill, M. E., and Travers, A. A. (1991) *Trends Biochem. Sci. 16*, 92-7.
10. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) *J. Mol. Graph. Model. 19*, 26-59.
11. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., and Dunker, A. K. (1997) *Proc. IEEE Intl. Conf. Neural Networks 1*, 90-95.
12. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) *Genome Informatics 10*, 30-40.
13. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) *Proteins: Struc., Funct., Gen. 42*, 38-48.
14. Romero, P., Obradovic, Z., and Dunker, A. K. (2000) *Artificial Intelligence Rev. 14*, 447-484.
15. Vucetic, S., Radivojac, P., Obradovic, Z., Brown, C. J., and Dunker, A. K. (2001) *Intl. Joint INNS-IEEE Conf. Neural Networks 4*, 2718-2723.
16. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Guilliot, S., Garner, E., and Dunker, A. K. (1998) *Pacific Symp. Biocomputing 1998 3*, 437-448.
17. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321-331.
18. Romero, P., Obradovic, Z., and Dunker, A. K. (1997) *Genome Informatics 8*, 110-124.
19. Li, X., Obradovic, Z., Brown, C. J., Garner, E. C., and Dunker, A. K. (2000) *Genome Informatics 11*, 172-184.
20. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) *Genome Informatics 11*, 161-171.
21. Dunker, A. K., and Obradovic, Z. (2001) *Nature Biotech. 19*, 805-806.
22. Karush, F. (1950) *J. Am. Chem. Soc. 72*, 2705-2713.
23. Yazawa, M., Matsuzawa, F., and Yagi, K. (1990) *J. Biochem. (Tokyo) 107*, 287-291.
24. Chakravarthy, B., Morley, P., and Whitfield, J. (1999) *Trends Neurosci. 22*, 12-16.
25. Minton, A. P. (2000) *Curr. Opin. Struct. Biol. 10*, 34-39.
26. Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) *Science 257*, 1251-1255.
27. Choo, Y., and Schwabe, J. W. (1998) *Nat. Struct. Biol. 5*, 253-255.
28. Morar, A. S., Wang, X., and Pielak, G. J. (2001) *Biochemistry 40*, 281-285.
29. Flaugh, S. L., and Lumb, K. J. (2001) *Biomacromolecules 2*, 538-540.
30. Zurdo, J., Sanz, J. M., Gonzalez, C., Rico, M., and Ballesta, J. P. (1997) *Biochemistry 36*, 9625-9635.

31. Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T., Oldfield, C. J., Williams, C. J., and Dunker, A. K. (2002) *J. Mol. Evol. in press*.
32. Zurdo, J., Gonzalez, C., Sanz, J. M., Rico, M., Remacha, M., and Ballesta, J. P. (2000) *Biochemistry 39*, 8935-8943.
33. Nusspaumer, G., Remacha, M., and Ballesta, J. P. (2000) *EMBO J. 19*, 6075-6084.
34. Ballesta, J. P., personal communication.
35. Tucker, P. A., Tsernoglou, D., Tucker, A. D., Coenjaerts, F. E., Leenders, H., and van der Vliet, P. C. (1994) *EMBO J. 13*, 2994-3002.
36. Dekker, J., Kanellopoulos, P. N., van Oosterhout, J. A., Stier, G., Tucker, P. A., and van der Vliet, P. C. (1998) *J. Mol. Biol. 277*, 825-838.
37. van Breukelen, B., Kanellopoulos, P. N., Tucker, P. A., and van der Vliet, P. C. (2000) *J. Biol. Chem. 275*, 40897-40903.
38. Johansson, J., Gudmundsson, G. H., Rottenberg, M. E., Berndt, K. D., and Agerberth, B. (1998) *J. Biol. Chem. 273*, 3718-3724.
39. Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E., and Greenblatt, J. (1998) *Mol. Cell 1*, 265-275.
40. Legault, P., Li, L., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998) *Cell 93*, 289-299.
41. Stellwagen, E., Rysavy, R., and Babul, G. (1972) *J. Biol. Chem. 247*, 8074-8077.
42. Strauss, P. R., and Holt, C. M. (1998) *J. Biol. Chem. 273*, 14435-14441.
43. Gorman, M. A., Morera, S., Rothwell, D. G., de La Fortelle, E., Mol, C. D., Tainer, J. A., Hickson, I. D., and Freemont, P. S. (1997) *EMBO J. 16*, 6548-6558.
44. Mol, C. D., Izumi, T., Mitra, S., and Tainer, J. A. (2000) *Nature 403*, 451-456.
45. Chang, J. F., Phillips, K., Lundback, T., Gstaiger, M., Ladbury, J. E., and Luisi, B. (1999) *J. Mol. Biol. 288*, 941-952.
46. Yamamoto, K., Ichijo, H., and Korsmeyer, S. J. (1999) *Mol. Cell. Biol. 19*, 8469-8478.
47. Cheng, E. H., Kirsch, D. G., Clem, R. J., Ravi, R., Kastan, M. B., Bedi, A., Ueno, K., and Hardwick, J. M. (1997) *Science 278*, 1966-1968.
48. Chang, B. S., Minn, A. J., Muchmore, S. W., Fesik, S. W., and Thompson, C. B. (1997) *EMBO J. 16*, 968-977.
49. Clem, R. J., Cheng, E. H., Karp, C. L., Kirsch, D. G., Ueno, K., Takahashi, A., Kastan, M. B., Griffin, D. E., Earnshaw, W. C., Veliuona, M. A., and Hardwick, J. M. (1998) *Proc. Natl. Acad. Sci. USA 95*, 554-559.
50. Rodi, D. J., Janes, R. W., Sanganee, H. J., Holton, R. A., Wallace, B. A., and Makowski, L. (1999) *J. Mol. Biol. 285*, 197-203.
51. Rodi, D. J., and Makowski, L. (1999) *Pacific Symp. Biocomputing 4*, 532-541.
52. Fang, G., Chang, B. S., Kim, C. N., Perkins, C., Thompson, C. B., and Bhalla, K. N. (1998) *Cancer Res. 58*, 3202-3208.
53. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., Gastinel, L. N., Habuka, N., Chen, X., Maldanado, F., Barker, J. E., Bacquet, R., and Villafranca, J. E. (1995) *Nature 378*, 641-644.
54. Yang, S. A., and Klee, C. B. (2000) *Biochemistry 39*, 16147-16154.
55. Bhattacharyya, J., and Das, K. P. (1999) *J. Biol. Chem. 274*, 15505-15509.

56. Bidwell, L. M., McManus, M. E., Gaedigk, A., Kakuta, Y., Negishi, M., Pedersen, L., and Martin, J. L. (1999) *J. Mol. Biol. 293*, 521-530.
57. Dajani, R., Cleasby, A., Neu, M., Wonacott, A. J., Jhoti, H., Hood, A. M., Modi, S., Hersey, A., Taskinen, J., Cooke, R. M., Manchee, G. R., and Coughtrie, M. W. (1999) *J. Biol. Chem. 274*, 37862-37868.
58. Ostedgaard, L. S., Baldursson, O., Vermeer, D. W., Welsh, M. J., and Robertson, A. D. (2000) *Proc. Natl. Acad. Sci. USA 97*, 5657-5662.
59. Lapthorn, A. J., Harris, D. C., Littlejohn, A., Lustbader, J. W., Canfield, R. E., Machin, K. J., Morgan, F. J., and Isaacs, N. W. (1994) *Nature 369*, 455-461.
60. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994) *Structure 2*, 545-558.
61. Bailey, R. W., Yang, J., Dunker, A. K., and Griswold, M. D. (2000) *Biophysical J. 78*, 152A.
62. Brandstetter, H., Bauer, M., Huber, R., Lollar, P., and Bode, W. (1995) *Proc. Natl. Acad. Sci. USA 92*, 9796-9800.
63. Brandstetter, H., Kuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K., and Engh, R. A. (1996) *J. Biol. Chem. 271*, 29988-29992.
64. Padmanabhan, K., Padmanabhan, K. P., Tulinsky, A., Park, C. H., Bode, W., Huber, R., Blankenship, D. T., Cardin, A. D., and Kisiel, W. (1993) *J. Mol. Biol. 232*, 947-966.
65. Thomas, J., Van Patten, S. M., Howard, P., Day, K. H., Mitchell, R. D., Sosnick, T., Trewhella, J., Walsh, D. A., and Maurer, R. A. (1991) *J. Biol. Chem. 266*, 10906-10911.
66. Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Xuong, N. H., Taylor, S. S., and Sowadski, J. M. (1991) *Science 253*, 414-420.
67. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) *Proc. Natl. Acad. Sci. USA 93*, 11504-11509.
68. Adkins, J. N., and Lumb, K. J. (2002) *Proteins 46*, 1-7.
69. Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J., and Pavletich, N. P. (1996) *Nature 382*, 325-331.
70. Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B., Link, T. A., Ramaswamy, S., and Jap, B. K. (1998) *Science 281*, 64-71.
71. Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell 85*, 125-136.
72. AEvarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., al-Karadaghi, S., Svensson, L. A., and Liljas, A. (1994) *EMBO J. 13*, 3669-3677.
73. Laurberg, M., Kristensen, O., Martemyanov, K., Gudkov, A. T., Nagaev, I., Hughes, D., and Liljas, A. (2000) *J. Mol. Biol. 303*, 593-603.
74. Martemyanov, K. A., and Gudkov, A. T. (2000) *J. Biol. Chem. 275*, 35820-35824.
75. Eom, J., Baker, W., Kintanar, A., and Wurtele, E. (1996) *Plant Sci. 115*, 17-24.
76. Ghosh, D., Pletnev, V. Z., Zhu, D. W., Wawrzak, Z., Duax, W. L., Pangborn, W., Labrie, F., and Lin, S. X. (1995) *Structure 3*, 503-513.
77. Pahel, G., Aulabaugh, A., Short, S. A., Barnes, J. A., Painter, G. R., Ray, P., and Phelps, W. C. (1993) *J. Biol. Chem. 268*, 26018-26025.
78. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A., Dobson, C. M., and Smith, L. J. (1997) *J. Mol. Biol. 274*, 152-159.

79. Penkett, C. J., Redfield, C., Jones, J. A., Dodd, I., Hubbard, J., Smith, R. A., Smith, L. J., and Dobson, C. M. (1998) *Biochemistry 37*, 17054-17067.
80. Penkett, C. J., Dobson, C. M., Smith, L. J., Bright, J. R., Pickford, A. R., Campbell, I. D., and Potts, J. R. (2000) *Biochemistry 39*, 2887-2893.
81. Vonderviszt, F., Kanto, S., Aizawa, S., and Namba, K. (1989) *J. Mol. Biol. 209*, 127-133.
82. Aizawa, S. I., Vonderviszt, F., Ishima, R., and Akasaka, K. (1990) *J. Mol. Biol. 211*, 673-677.
83. Mimori-Kiyosue, Y., Yamashita, I., Fujiyoshi, Y., Yamaguchi, S., and Namba, K. (1998) *J. Mol. Biol. 284*, 521-530.
84. Mimori-Kiyosue, Y., Vonderviszt, F., and Namba, K. (1997) *J. Mol. Biol. 270(2)*, 222-237.
85. Samatey, F. A., Imada, K., Nagashima, S., Vonderviszt, F., Kumasaka, T., Yamamoto, M., and Namba, K. (2001) *Nature 410*, 331-337.
86. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T., and Dahlquist, F. W. (1997) *Nat. Struct. Biol. 4*, 285-291.
87. Daughdrill, G. W., Hanely, L. J., and Dahlquist, F. W. (1998) *Biochemistry 37*, 1076-1082.
88. Fletcher, C. M., McGuire, A. M., Gingras, A. C., Li, H., Matsuo, H., Sonenberg, N., and Wagner, G. (1998) *Biochemistry 37*, 9-15.
89. Fletcher, C. M., and Wagner, G. (1998) *Protein Sci. 7*, 1639-1642.
90. Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R., and Sonenberg, N. (1999) *Genes Dev. 13*, 1422-1437.
91. Eigenbrot, C., and Gerber, N. (1997) *Nat. Struct. Biol. 4*, 435-438.
92. Eketjall, S., Fainzilber, M., Murray-Rust, J., and Ibanez, C. F. (1999) *EMBO J. 18*, 5901-5910.
93. Ibanez, C. F., personal communication.
94. Baskakov, I. V., Kumar, R., Srinivasan, G., Ji, Y. S., Bolen, D. W., and Thompson, E. B. (1999) *J. Biol. Chem. 274*, 10693-10696.
95. Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., and Takusagawa, F. (2000) *J. Mol. Biol. 298*, 149-162.
96. Takusagawa, F., personal communication.
97. Logan, D. T., Mazauric, M. H., Kern, D., and Moras, D. (1995) *EMBO J. 14*, 4156-4167.
98. Arnez, J. G., Dock-Bregeon, A. C., and Moras, D. (1999) *J. Mol. Biol. 286*, 1449-1459.
99. Clackson, T., Ultsch, M. H., Wells, J. A., and de Vos, A. M. (1998) *J. Mol. Biol. 277*, 1111-1128.
100. Clackson, T., and Wells, J. A. (1995) *Science 267*, 383-386.
101. Holliger, P., Riechmann, L., and Williams, R. L. (1999) *J. Mol. Biol. 288*, 649-657.
102. Holliger, P., and Riechmann, L. (1997) *Structure 5*, 265-275.
103. Nilsson, N., Malmborg, A. C., and Borrebaeck, C. A. (2000) *J. Virol. 74*, 4229-4235.
104. Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G., and Sprang, S. R. (1994) *Science 265*, 1405-1412.
105. Coleman, D. E., and Sprang, S. R. (1999) *J. Biol. Chem. 274*, 16669-16672.
106. Cho, H. S., Liu, C. W., Damberger, F. F., Pelton, J. G., Nelson, H. C., and Wemmer, D. E. (1996) *Protein Sci. 5*, 262-269.
107. Louie, G. V., Yang, W., Bowman, M. E., and Choe, S. (1997) *Mol. Cell 1*, 67-78.

108. Higashiyama, S., Lau, K., Besner, G. E., Abraham, J. A., and Klagsbrun, M. (1992) *J Biol Chem 267*, 6205-6212.
109. Cary, P. D., King, D. S., Crane-Robinson, C., Bradbury, E. M., Rabbani, A., Goodwin, G. H., and Johns, E. W. (1980) *Eur. J. Biochem. 112*, 577-580.
110. Abercrombie, B. D., Kneale, G. G., Crane-Robinson, C., Bradbury, E. M., Goodwin, G. H., Walker, J. M., and Johns, E. W. (1978) *Eur. J. Biochem. 84*, 173-177.
111. Louie, D. F., Gloor, K. K., Galasinski, S. C., Resing, K. A., and Ahn, N. G. (2000) *Protein Sci 9*, 170-179.
112. Bergel, M., Herrera, J. E., Thatcher, B. J., Prymakowska-Bosak, M., Vassilev, A., Nakatani, Y., Martin, B., and Bustin, M. (2000) *J. Biol. Chem. 275*, 11514-11520.
113. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans, J. N., Reeves, R., Gronenborn, A. M., and Clore, G. M. (1997) *Nat. Struct. Biol. 4*, 657-665.
114. Pierantoni, G. M., Fedele, M., Pentimalli, F., Benvenuto, G., Pero, R., Viglietto, G., Santoro, M., Chiariotti, L., and Fusco, A. (2001) *Oncogene 20*, 6132-6141.
115. Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G., and Thanos, D. (2001) *Science 293*, 1133-1136.
116. Cary, P. D., Crane-Robinson, C., Bradbury, E. M., and Dixon, G. H. (1981) *Eur. J. Biochem. 119*, 545-551.
117. Lambert, S. J., Nicholson, J. M., Chantalat, L., Reid, A. J., Donovan, M. J., and Baldwin, J. P. (1999) *Acta. Crystallogr. D. Biol. Crystallogr. 55*, 1048-1051.
118. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001) *Science 292*, 110-113.
119. Aviles, F. J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C., and Bradbury, E. M. (1978) *Eur. J. Biochem. 88*, 363-371.
120. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) *Nature 362*, 219-223.
121. Graziano, V., Gerchman, S. E., Wonacott, A. J., Sweet, R. M., Wells, J. R., White, S. W., and Ramakrishnan, V. (1990) *J. Mol. Biol. 212*, 253-257.
122. Crane-Robinson, C., personal communication.
123. Focia, P. J., Craig, S. P., 3rd, Nieves-Alicea, R., Fletterick, R. J., and Eakin, A. E. (1998) *Biochemistry 37*, 15066-15075.
124. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) *Nature 376*, 791-795.
125. Polverini, E., Fasano, A., Zito, F., Riccio, P., and Cavatorta, P. (1999) *Eur. Biophys. J. 28*, 351-355.
126. Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996) *Cell 85*, 931-942.
127. Arold, S., Franken, P., Strub, M. P., Hoh, F., Benichou, S., Benarous, R., and Dumas, C. (1997) *Structure 5*, 1361-1372.
128. Geyer, M., Munte, C. E., Schorr, J., Kellner, R., and Kalbitzer, H. R. (1999) *J. Mol. Biol. 289*, 123-138.
129. Arold, S. T., and Baur, A. S. (2001) *Trends Biochem. Sci. 26*, 356-363.
130. Berkovits, H. J., and Berg, J. M. (1999) *Biochemistry 38*, 16826-168230.
131. Brown, H. G., and Hoh, J. H. (1997) *Biochemistry 36*, 15035-15040.
132. Hoh, J. H. (1998) *Proteins 32*, 223-228.

133. Zhang, M., Vogel, H. J., and Zwiers, H. (1994) *Biochem. Cell Biol. 72*, 109-116.
134. Grishin, N. V., Osterman, A. L., Brooks, H. B., Phillips, M. A., and Goldsmith, E. J. (1999) *Biochemistry 38*, 15174-15184.
135. Engel, J., Taylor, W., Paulsson, M., Sage, H., and Hogan, B. (1987) *Biochemistry 26*, 6958-6965.
136. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G. (1995) *Nat. Struct. Biol. 2*, 537-547.
137. Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997) *Structure 5*, 59-68.
138. Rao, V. D., Misra, S., Boronenkov, I. V., Anderson, R. A., and Hurley, J. H. (1998) *Cell 94*, 829-839.
139. Grobler, J. A., Essen, L. O., Williams, R. L., and Hurley, J. H. (1996) *Nat. Struct. Biol. 3*, 788-795.
140. Essen, L. O., Perisic, O., Cheung, R., Katan, M., and Williams, R. L. (1996) *Nature 380*, 595-602.
141. Lopez Garcia, F., Zhan, R., Riek, R., and Wulthrich, K. (2000) *Proc Natl. Acad. Sci. USA 97*, 8334-8339.
142. Gatewood, J. M., Schroth, G. P., Schmid, C. W., and Bradbury, E. M. (1990) *J. Biol. Chem. 265*, 20667-20672.
143. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Muller-Frohne, M., Zirwer, D., Czarnecki, J., and Damaschun, G. (1995) *Biochemistry 34*, 13211-13218.
144. McEwan, I. J., Dahlman-Wright, K., Ford, J., and Wright, A. P. (1996) *Biochemistry 35*, 9584-93.
145. Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997) *Nucleic Acids Res. 25*, 2702-2715.
146. Jacobs, D. M., Lipton, A. S., Isern, N. G., Daughdrill, G. W., Lowry, D. F., Gomes, X., and Wold, M. S. (1999) *J. Biomol. NMR 14*, 321-331.
147. Holmbeck, S. M., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J., and Wright, P. E. (1998) *J. Mol. Biol. 281*, 271-284.
148. Venyaminov, S., Gudkov, A., Gogia, Z., and Tumanova, L. (1981) pp 128, Biological Research Center, Institute of Protein Research, Pushchino: AS USSR.
149. Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B., and Steitz, T. A. (1999) *Nature 400*, 841-847.
150. Tesmer, J. J., Berman, D. M., Gilman, A. G., and Sprang, S. R. (1997) *Cell 89*, 251-261.
151. Moy, F. J., Chanda, P. K., Cockett, M. I., Edris, W., Jones, P. G., Mason, K., Semus, S., and Powers, R. (2000) *Biochemistry 39*, 7063-7073.
152. Chatterjee, T. K., and Fisher, R. A. (2000) *J. Biol. Chem. 275*, 24013-24021.
153. Silva, A. M., and Rossmann, M. G. (1985) *ACTA Cryst. B41*, 147-157.
154. Rossmann, M. G., Chandrasekaran, R., Abad-Zapatero, C., Erickson, J. W., and Arnott, S. (1983) *J. Mol. Biol. 166*, 73-80.
155. Josefsson, E., O'Connell, D., Foster, T. J., Durussel, I., and Cox, J. A. (1998) *J. Biol. Chem. 273*, 31145-31152.
156. Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G., and Wengler, G. (1991) *Nature 354*, 37-43.

157. Choi, H. K., Lee, S., Zhang, Y. P., McKinney, B. R., Wengler, G., Rossmann, M. G., and Kuhn, R. J. (1996) *J. Mol. Biol. 262*, 151-167.
158. Kim, K. K., Kim, R., and Kim, S. H. (1998) *Nature 394*, 595-599.
159. Kim, R., Kim, K. K., Yokota, H., and Kim, S. H. (1998) *Proc. Natl. Acad. Sci. USA 95*, 9129-9133.
160. Fasshauer, D., Bruns, D., Shen, B., Jahn, R., and Brunger, A. T. (1997) *J. Biol. Chem. 272*, 4582-4590.
161. Fasshauer, D., Otto, H., Eliason, W. K., Jahn, R., and Brunger, A. T. (1997) *J. Biol. Chem. 272*, 28036-28041.
162. Fasshauer, D., Eliason, W. K., Brunger, A. T., and Jahn, R. (1998) *Biochemistry 37*, 10354-10362.
163. Canaves, J. M., and Montal, M. (1998) *J. Biol. Chem. 273*, 34214-32221.
164. Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) *Nature 395*, 347-353.
165. Hazzard, J., Sudhof, T. C., and Rizo, J. (1999) *J. Biomol. NMR 14*, 203-207.
166. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T., Jr. (1996) *Biochemistry 35*, 13709-13715.
167. Nielsen, M. S., Vorum, H., Lindersson, E., and Jensen, P. H. (2001) *J. Biol. Chem. 18*, 18.
168. Tell, G., Perrone, L., Fabbro, D., Pellizzari, L., Pucillo, C., De Felice, M., Acquaviva, R., Formisano, S., and Damante, G. (1998) *Biochem. J. 329*, 395-403.
169. Trombitas, K., Greaser, M., Labeit, S., Jin, J. P., Kellermayer, M., Helmes, M., and Granzier, H. (1998) *J. Cell Biol. 140*, 853-859.
170. Labeit, S., and Kolmerer, B. (1995) *Science 270*, 293-296.
171. Kellermayer, M. S. Z., Smith, S. B., Granzier, H. L., and Bustamante, C. (1997) *Science 276*, 1112-1116.
172. Helmes, M., Trombitas, K., Centner, T., Kellermayer, M., Labeit, S., Linke, W. A., and Granzier, H. (1999) *Circ. Res. 84*, 1339-1352.
173. Hopper, P., Harrison, S. C., and Sauer, R. T. (1984) *J. Mol. Biol. 177*, 701-713.
174. Stewart, L., Ireton, G. C., Parker, L. H., Madden, K. R., and Champoux, J. J. (1996) *J. Biol. Chem. 271*, 7593-7601.
175. Shaiu, W. L., Hu, T., and Hsieh, T. S. (1999) *Pacific Symp. Biocomputing 4*, 578-589.
176. Berger, J. M., Gamblin, S. J., Harrison, S. C., and Wang, J. C. (1996) *Nature 379*, 225-232.
177. Caron, P. R., Watt, P., and Wang, J. C. (1994) *Mol. Cell Biol. 14*, 3197-3207.
178. Berger, J. M., personal communication.
179. Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., and Klevit, R. E. (1988) *Science 241*, 1489-1492.
180. Klevit, R. E., Herriott, J. R., and Horvath, S. J. (1990) *Proteins 7*, 215-226.
181. Hyre, D. E., and Klevit, R. E. (1998) *J. Mol. Biol. 279*, 929-943.
182. Bowers, P. M., Schaufler, L. E., and Klevit, R. E. (1999) *Nat. Struct. Biol. 6*, 478-485.
183. Campbell, K. M., Terrell, A. R., Laybourn, P. J., and Lumb, K. J. (2000) *Biochemistry 39*, 2708-2713.
184. John, M., Briand, J. P., and Schnarr, M. (1996) *Pept Res 9*, 71-78.
185. Krebs, D., Dahmani, B., el Antri, S., Monnot, M., Convert, O., Mauffret, O., Troalen, F., and Fermandjian, S. (1995) *Eur. J. Biochem. 231*, 370-380.
186. Glover, J. N., and Harrison, S. C. (1995) *Nature 373*, 257-61.

187. Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S., and King, G. F. (1996) *J. Biol. Chem. 271*, 13663-13667.
188. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature 392*, 42-48.
189. Richards, J. P., Bachinger, H. P., Goodman, R. H., and Brennan, R. G. (1996) *J. Biol. Chem. 271*, 13716-13723.
190. Liu, D., Ishima, R., Tong, K. I., Bagby, S., Kokubo, T., Muhandiram, D. R., Kay, L. E., Nakatani, Y., and Ikura, M. (1998) *Cell 94*, 573-583.
191. Hershey, P. E., McWhirter, S. M., Gross, J. D., Wagner, G., Alber, T., and Sachs, A. B. (1999) *J. Biol.Chem. 274*, 21297-21304.
192. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) *Nature 347*, 575-578.
193. Horiuchi, M., Kurihara, Y., Katahira, M., Maeda, T., Saito, T., and Uesugi, S. (1997) *J. Biochem. (Tokyo) 122*, 711-716.
194. Schmitz, M. L., dos Santos Silva, M. A., Altmann, H., Czisch, M., Holak, T. A., and Baeuerle, P. A. (1994) *J. Biol. Chem. 269*, 25613-25620.
195. Chang, J., Kim, D. H., Lee, S. W., Choi, K. Y., and Sung, Y. C. (1995) *J. Biol. Chem. 270*, 25014-25019.
196. Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., and Pavletich, N. P. (1996) *Science 274*, 948-953.
197. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996) *J. Biol. Chem. 271*, 4827-4837.
198. Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J., and Verdine, G. L. (1997) *Science 277*, 1310-1313.
199. Liu, Y., Gong, W., Huang, C. C., Herr, W., and Cheng, X. (1999) *Genes Dev. 13*, 1692-1703.
200. Hayes, S., and O'Hare, P. (1993) *J. Virol. 67*, 852-862.
201. O'Hare, P., and Williams, G. (1992) *Biochemistry 31*, 4150-4156.
202. Donaldson, L., and Capone, J. P. (1992) *J. Biol. Chem. 267*, 1411-1414.
203. Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E., and Saraste, M. (1995) *Proc. Natl. Acad. Sci. USA 92*, 11955-11959.
204. Abramson, J., Riistama, S., Larsson, G., Jasaitis, A., Svensson-Ek, M., Laakkonen, L., Puustinen, A., Iwata, S., and Wikstrom, M. (2000) *Nat. Struct. Biol. 7*, 910-917.
205. Wissmann, R., Baukrowitz, T., Kalbacher, H., Kalbitzer, H. R., Ruppersberg, J. P., Pongs, O., Antz, C., and Fakler, B. (1999) *J. Biol. Chem. 274*, 35521-35525.
206. Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Bruce, J. E., Garner, E. C., Brown, C. J., Dunker, A. K., Smith, R. D., and Ackerman, E. J. (2001) *Prot. Sci. 10*, 560-571.
207. Namba, K. (2001) *Gen. Cells 6*, 1-12.
208. Schulz, G. E. (1979) in *Molecular Mechanism of Biological Recognition* (Balaban, M., Ed.) pp 79-94, Elsevier/North-Holland Biomedical Press, New York.
209. Meador, W. E., Means, A. R., and Quiocho, F. A. (1993) *Science 262*, 1718-1721.
210. Pontius, B. W. (1993) *Trends Biochem. Sci. 18*, 181-186.
211. Pappu, R. V., Srinivasan, R., and Rose, G. D. (2000) *Proc. Natl. Acad. Sci. USA 97*, 12565-12570.

212. Gutierrez-Cruz, G., Van Heerden, A. H., and Wang, K. (2001) *J. Biol. Chem. 276*, 7442-7449.
213. Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) *J. Mol. Biol. 266*, 223-230.
214. Fontana, A., de Laureto, P. P., de Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Design 2*, R17-R26.
215. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) *Protein Sci. 3*, 757-768.
216. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905-920.
217. Treiber, D. K., and Williamson, J. R. (2001) *Curr Opin Struct Biol 11*, 309-314.
218. Bailey, R. W., Dunker, A. K., Brown, C. J., Garner, E. C., and Griswold, M. D. (2001) *Biochemistry 40*, 11828-11840.
219. House-Pompeo, K., Xu, Y., Joh, D., Speziale, P., and Hook, M. (1996) *J. Biol. Chem. 271*, 1379-1384.
220. Jeffery, C. J. (1999) *Trends Biochem Sci 24*, 8-11.
221. Kang, C. H., Trumble, W. R., and Dunker, A. K. (2001) in *Methods in Molecular Biology* (Vogel, H. J., Ed.) pp 281-294, Humana Press, Totowa, New Jersey.
222. Doolittle, R. F. (1986) *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences*, University Science Books.

Table 1. Intrinsically disordered protein and their functions.

	Protein	Detection Method	Location	Chain Length	Protein Length	Class	Function	References
1	Acidic ribosomal stalk protein P1(133069)	NMR, CD LP, other	1-106	106	106	D, MG	a, cR, h, mP, l	(*30, 32-34*)
	Acidic ribosomal stalk protein P2 (133071)		1-110	110	110			
2	Adenovirus ssDNA binding protein (1633461)	X-ray, other	297-331	356	529	D-O†	b, n, t	(*35-37*)
3	Antibacterial protein LL-37 (1706745)	CD	1-37	37	37	D-O	a, g	(*38*)
4	Antitermination protein N of bacteriophage λ (132276)	NMR, CD	73-107	107	107	D-O	a, cM	(*39, 40*)
5	Apocytochrome c (117995)	CD	1-104	104	104	D-O	i	(*41*)
6	Apurinic/apyrimidinic endonuclease (299037)	X-ray, LP	1-40	279	317	U	x	(*42-44*)
7	B cell-specific transcription co-activator (1150493)	CD, LP	1-65	65	256	D-O	a, b	(*45*)
			66-256	256	256	D-O†	a	
8	Bcl-x_L antiapoptotic protein (2392082)	X-ray, NMR	28-80	196	233	D	mP, l	(*46-48*)
9	Bcl-2 antiapoptotic protein (13786963)	Other	24-93	196	239	U†	h, mP†, l†	(*47, 49-52*)
10	Calcineurin (1352673)	X-ray, LP	374-468	521	521	O-D, D-O†	a, k†	(*53, 54*)
			487 - 521			U	x	
11	α_s-Casein (1070620)	CD	1-190	190	190	D	v	(*55*)
12	Catecholamine sulfotransferase (1711609)	X-ray	216-261	295	295	D-O†	h	(*56, 57*)
13	CFTR (14753227)	CD, LP	708-831	124	1480	D	k, mP	(*58*)
14	Chorionic gonadotropin (116184)	X-ray	B:112-145	145	145	D	mG, w	(*59, 60*)
15	Clusterin (461756)	LP	66-97	447	447	D, MG	v	(*61*)
			386-445				v	
16	Clotting factor Xa (119761)	X-ray, LP	L:1-46	146	488	U	x	(*62-64*)
17	cAMP-dependent protein kinase inhibitor (417194)	CD, X-ray	1-76	77	76	D-O	a	(*65, 66*)
18	Cyclin-dependent kinase inhibitor p21 (729143)	NMR, CD, LP	1-164	164	164	U, D-O	a	(*67*)
	Cyclin-dependent kinase inhibitor p57 (11440665)	CD, other	1-316	316	316	D	a	(*68*)

	Cyclin-dependent kinase inhibitor p27 (1168871)	CD, X-ray, other	23-106	84	198	D-O	a	(*29, 69*)
19	Cytochrome bc1 complex (1351360)	X-ray	Subunit I: 1-43	78	78	U	x	(*70*)
			Subunit E: 76-196	196	196			
20	DNA-binding protein RAP1 (730473)	X-ray, LP	482-512	246	827	U	x	(*71*)
21	Elongation factor G (1827912)	X-ray	400-480	691	691	U	cR†, k, n†	(*72-74*)
22	Embryonic abundant protein from carrot (119316)	NMR	1-92	92	92	U	h†	(*75*)
23	Estrogenic 17-β hydroxysteroid dehydrogenase (2392375)	X-ray	285-327	327	327	U	f†, w	(*76*)
24	E7 protein from HPV16 (6469700)	CD	1-93	93	93	D-O	a, j	(*77*)
25	Fibronectin binding protein (120457)	NMR	745-873	130	1018	D-O	a	(*78-80*)
26	Flagellin (96744)	NMR, X-ray*, CD, other	1-65	various	494	D, D-O	a, g, s	(*81-85*)
			451-494					
27	Flagellum specific σ factor (120306)	NMR	1-97	97	97	D, D-O	a, s	(*86, 87*)
28	4E-binding protein 1 (4758258)	NMR, CD X-ray,	1-118	14	118	D, D-O	a, mP	(*88-90*)
29	Glial cell-derived neurotrophic factor (729568)	X-ray	1-40	135	135	D-O†	a†	(*91-93*)
30	Glucocorticoid receptor (4758482)	CD, other	77-262	186	835	D-O	a	(*94*)
31	Glycine N-methyltransferase (121328)	X-ray	1-40	292	292	U, D-O	h	(*95, 96*)
32	Glycyl-tRNA synthetase (2829475)	X-ray	96-158	442	505	U	x	(*97, 98*)
33	Growth hormone receptor (121180)	X-ray	1-31	238	620	D-O	a, w	(*99, 100*)
34	g3p (fd phage minor coat protein) (5822481)	X-ray, NMR	218-256	225	406	D	n	(*101-103*)
35	G protein $G_{i\alpha 1}$ (121020)	X-ray	1-33	353	353	D, D-O	a, f, mF	(*104, 105*)
36	Heat shock transcription	NMR, CD	1-193	282	677	D	a	(*106*)

	factor (123686)							
37	Heparin-binding EGF-like growth factor (544477)	X-ray	73-106	79	189	U	mG	(*107, 108*)
38	High mobility group - 14 (4826758)	NMR, CD, other	1-100	100	100	D, D-O	a, b, mP, mA	(*109-112*)
	High mobility group - 17 (5031749)		1-89	89	89			
39	High mobility group - I(Y) (123377)	NMR	1-107	various	107	D, D-O	a, b, mP, mA	(*113-115*)
40	High mobility group - T (123382)	NMR, CD	1-204	204	204	D-O	b	(*116*)
	High mobility group - H6 (462245)		1-69	69	69			
41	Histone H3 (211857)	X-ray	1-40	136	136	D	mP, mA, mM	(*117, 118*)
42	Histone H5 (70678)	X-ray, LP	101-185	185	185	D-O†	b†	(*119-122*)
43	Hypoxanthine phosphoribosyltransferase (6435814)	X-ray	190-221	221	221	U	x	(*123*)
44	Lymphoid enhancer factor-1 (6537322)	NMR	296-380	86	397	D-O	b, u	(*124*)
45	Myelin basic protein (126796)	CD	1-169	169	169	D-O	f	(*125*)
46	Negative factor, HIV1 (128023)	X-ray, NMR	2-73	various	206	D, D-O	a, f, mP, mF, l	(*126-129*)
			149-178			U	a	
47	Neural zinc finger factor 1 (1511632)	NMR	1-99	99	1187	D-O	b, j	(*130*) 2
48	Neurofilament H (128127)	Other	409-1087	1087	1087	D	mP, mG, p	(*131, 132*)
49	Neuromodulin (548347)	NMR, CD	1-239	239	239	D, D-O	a, f, mP, mR, mF	(*133*)
50	Ornithine decarboxylase (7404357)	X-ray	1-36	425	425	U	x	(*134*)
51	Osteonectin (129284)	CD	23-68	285	285	D-O	j	(*135*)
52	Phenylalanyl-tRNA synthetase (135112)	X-ray	α1-84	350	350	D-O	cT	(*136, 137*)
53	Phosphatidylinositol phosphate kinase (3745771)	X-ray	1-33	416	416	U	x	(*138*)
			307-341					
54	Phospholipase c-δ1 (130228)	X-ray, LP	135-205	622	756	U	k†	(*139, 140*)
			446-484			U	w, x	

55	Prion (200527)	NMR	23-120	219	241	D-O	j	(*141*)
56	Protamines (123705)	CD	1-57	57	57	D-O	b, j	(*142*)
57	Prothymosin α (135836)	CD	1-109	109	109	U	a	(*143*)
58	Myc proto-oncogene protein (127619)	CD	1-143	143	439	D-O	a	(*144*)
59	*Pvu*II DNA methyltransferase (6729995)	X-ray	179-216	336	336	D-O†	b†, l†	(*145*)
60	Replication protein A (1350579)	NMR	109-168	168	616	D	n	(*146*)
61	Retinoid X receptor α (4506755)	NMR	130-212	93	462	D-O	b, j	(*147*)
62	30S ribosomal proteins	X-ray, CD	various	various	various	D, D-O	a, cR, r	(*148, 149*)
63	50S ribosomal proteins	X-ray, CD	various	various	various	D, D-O	a, cR, r	(*148, 149*)
64	Signal transduction inhibitor RGS4 (1710149)	X-ray	1-59	205	205	U	a†	(*150-152*)
			176-205			U	x	
65	Southern bean mosaic virus capsid (116795)	X-ray	1-38	260	260	D-O†	cG	(*153, 154*)
66	Serine aspartamine repeat protein D (3550594)	CD	569-1123	555	1315	D-O	j, n†	(*155*)
67	Sindbis virus capsid (1942972)	X-ray	1-105	264	264	D-O†	cG	(*156, 157*)
68	Small heat-shock protein HSP16.5 (2495337)	X-ray	1-33	147	147	U	a†	(*158, 159*)
69	SNAP-25 (134583)	CD, LP, other, X-ray	1-83	various	206	D-O	a	(*160-164*)
			84-130			D	f, mF, n	
			131-206			D-O	a	
70	Synaptobrevin (135093)	NMR, CD X-ray, LP, other	1-96	various	116	D-O	a	(*161-165*)
71	α-synuclein (586067)	CD	1-140	140	140	D-O	a, g, j	(*166, 167*)
72	Thyroid transcription factor 1 (136462)	LP	1-156	156	372	D-O	a	(*168*)
73	Titin, skeletal (1017427)	Other, CD	PEVK domain (2174 aa's)	n.a,	~33,000	O-D	o	(*169-172*)
	Titin, cardiac (2136280)	Other	N2B domain (572 aa's)	n.a.	~27,000	O-D	o	
74	Tomato bushy stunt virus coat protein (116805)	X-ray	1-100	387	387	D-O†	cG	(*173*)

75	Topoisomerase I (135989)	CD	1-174	various	765	U	a, l, w	(*174, 175*)
76	Topoisomerase II (1633273)	X-ray, LP	632-680	various	1429	D	b†, n	(*175-178*)
			1077-1106			U	k, mP	
			1203-1429			U	a, k, mP, w	
77	Transcription factor ADR1 (113450)	NMR	75-159	various	1323	D-O	b, j	(*179-182*)
78	Transcription factor c-Fos (4063509)	CD,NMR	216-380	85	380	D-O	a	(*183*)
79	Transcription factor c-Jun (135298)	NMR, CD X-ray, other	61-98	37	331	D-O	x	(*184*)
			257-314			D-O	a, b	(*185-188*)
80	Transcription factor CREB (117435)	CD	1-265	341	341	D-O†	a, mP	(*189*)
81	Transcription factor $dTAF_{II}230$ (1705691)	NMR	11-77	67	2068	D-O	a	(*190*)
82	Transcription factor eIF-4G (1170510)	NMR	393-490	98	1395	D-O	a	(*191*)
83	Transcription factor GCN4 (121066)	NMR	225-250	58	281	D-O	b	(*192*)
			251-281			D-O	a	
84	Transcription factor MAX (126776)	CD, other	2-110	109	160	D-O	a, b	(*193*)
85	Transcription factor NF-κB p65 (417924)	NMR	428-551	123	551	D-O	a	(*194*)
86	Transcription factor p53 (129369)	CD	1-73	73	393	D-O	a	(*195, 196*)
87	Transcription factor VP16 (2827761)	NMR, CD X-ray, LP, other	350-394	various	490	D-O	a, b	(*197-202*)
			403-490			D-O	a	
88	Ubiquinol oxidase (118072; 118071)	X-ray	Subunit I: 1-51	663	663	U	x	(*203, 204*)
			Subunit I: 553-663					
			Subunit II: 284-315	315	315	U	x	
89	Voltage-gated potassium channel (4557685)	NMR	1-62	495	495	D	k, q	(*205*)
90	Xeroderma pigmentosum Group A (139817)	LP	1-84	265	265	D	a†	(*206*)
			181-265					

[1]Method of detection: X-ray: X-ray crystallography; NMR: Nuclear magnetic resonance spectroscopy; CD: Circular dichroism; LP: Limited proteolysis; Other: Other techniques

[2]Class: (D) Function arises from the disordered (extended) state; (D-O) Function arises via a disorder to order transition; (O-D) Function arises via an order to disorder transition; (MG) Function arises from the molten globule (collapsed) state; (U) Known to exist in disordered state, relationship to function unknown

[3]Function: (a) Protein-protein binding; (b) Protein-DNA binding; (cR) Protein-rRNA binding; (cT) Protein-tRNA binding ; (cM) Protein-mRNA binding ; (cG) Protein-genomic RNA binding; (f) Protein-lipid interaction; (g) Polymerization; (h) Autoregulatory; (i) Substrate/ligand binding; (j) Cofactor/heme binding; (k) Metal binding; (l) Regulation of proteolysis *in vivo*; (mP) Phosphorylation; (mA) Acetylation; (mG) Glycosylation; (mM) Methylation; (mF) Fatty acylation (myristolation and palmitoylation); (mR) ADP-ribosylation; (n) Flexible linkers/spacers; (o) Entropic spring; (p) Entropic bristle; (q) Entropic clock; (r) Structural mortar; (s) Transport thru channel; (t) DNA unwinding; (u) DNA bending; (v) Protein detergent; (w) Disordered region not essential for protein function; (x) Unknown

[4]Additional notes: *Structure solved only after removal of disordered region; †Hypothesized

Table 2. Number of disordered regions exhibiting each function.

Code[1]	Function	# of examples
a	Protein-protein binding	54
b	Protein-DNA binding	19
cR	Protein-rRNA binding	5
cT	Protein-tRNA binding	1
cM	Protein-mRNA binding	1
cG	Protein-genomic RNA binding	3
f	Protein-lipid interaction	6
g	Polymerization	4
h	Substrate/ligand binding	6
i	Cofactor/heme binding	1
j	Metal binding	9
k	Autoregulatory	7
l	Regulation of proteolysis *in vivo*	7
mA	Acetylation	4
mF	Fatty acylation (myristolation and palmitoylation)	4
mG	Glycosylation	3
mM	Methylation	1
mP	Phosphorylation	16
mR	ADP-ribosylation	1
n	Flexible linkers/spacers	7
o	Entropic spring	2
p	Entropic bristle	1
q	Entropic clock	1
r	Structural mortar	>10
s	Self-transport through channel	3
t	DNA unwinding	1
u	DNA bending	1
v	Protein detergent	3
w	Disordered region is not essential for protein function	6
x	Unknown	17

[1]Code in Function column of Table 1

Figure Legend

Figure 1. The Protein Trinity: in this proposal, structure-function relationships of native proteins can depend on any of the three forms - folded (having an ordered 3-D structure), collapsed (molten globule-like, but perhaps having regions exhibiting protection from hydrogen exchange), or extended (random coil-like, but perhaps having regions of secondary structure) - and can involve transitions between these forms.