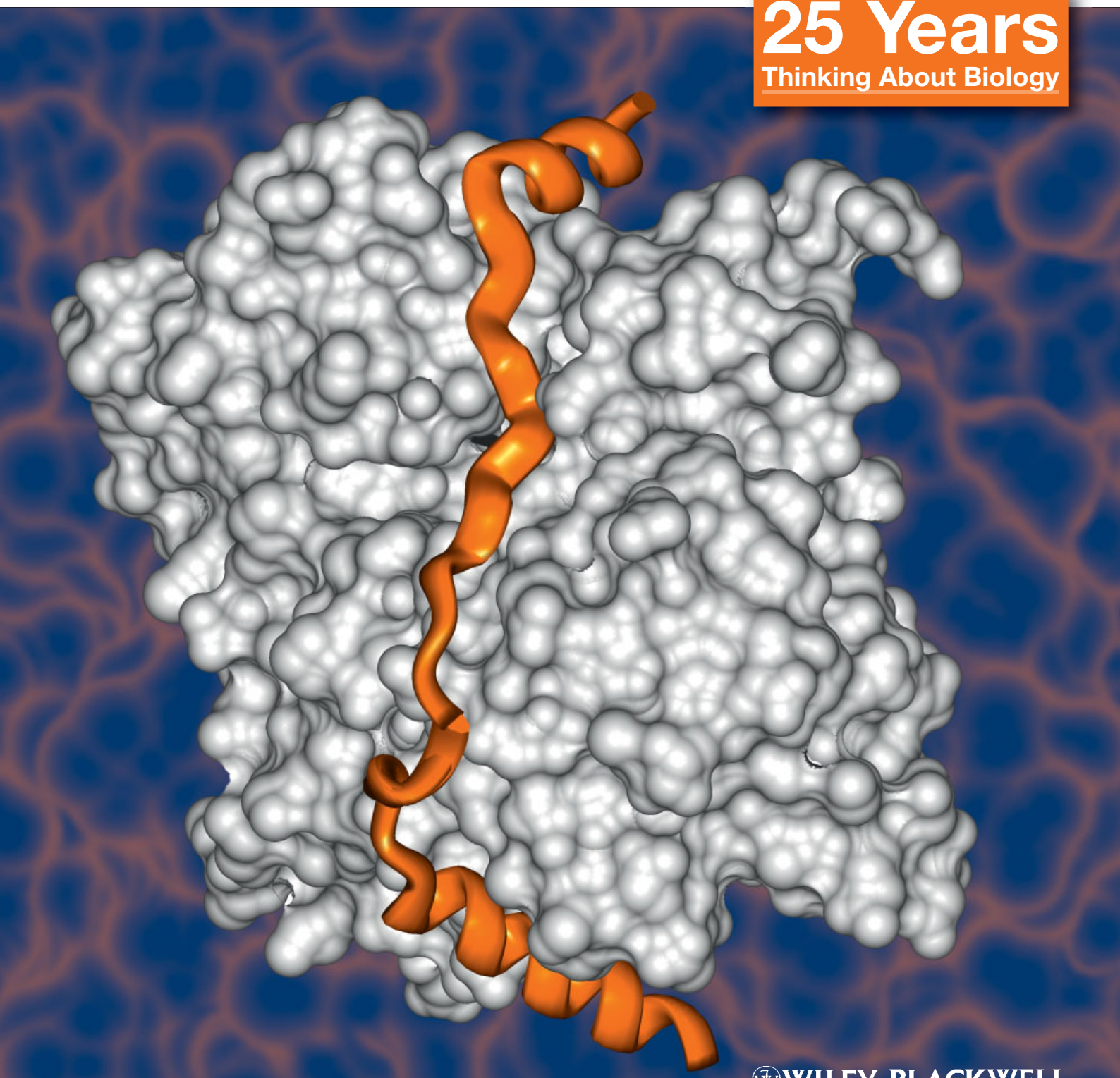
Volume 31 No. 3 March 2009 · ISSN 0265-9247

BioEssays

3/09

www.bioessays-journal.com

25 Years
Thinking About Biology



WILEY-BLACKWELL

 DOI 10.1002/bies.200800151

Close encounters of the third kind: disordered domains and the interactions of proteins

Peter Tompa,[1]* *Monika Fuxreiter,*[1] *Christopher J. Oldfield,*[2] *Istvan Simon,*[1] *A. Keith Dunker,*[2] *and Vladimir N. Uversky*[2,3,4]

[1]Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary
[2]Center for Computational Biology and Bioinformatics, School of Medicine, Indiana University, Indianapolis, IN, USA
[3]Institute for Biological Instrumentation, Russian Academy of Sciences, Moscow Region, Russia
[4]Institute for Intrinsically Disordered Protein Research, Indiana University School of Medicine, Indianapolis, IN, USA

Protein–protein interactions are thought to be mediated by domains, which are autonomous folding units of proteins. Recently, a second type of interaction has been suggested, mediated by short segments termed *linear motifs*, which are related to recognition elements of intrinsically disordered regions. Here, we propose a third kind of protein–protein recognition mechanism, mediated by disordered regions longer than 20–30 residues. Bioinformatics predictions and well-characterized examples, such as the kinase-inhibitory domain of Cdk inhibitors and the Wiskott–Aldrich syndrome protein (WASP)-homology domain 2 of actin-binding proteins, show that these disordered regions conform to the definition of domains rather than motifs, *i.e.*, they represent functional, evolutionary, and structural units. Their functions are distinct from those of short motifs and ordered domains, and establish a third kind of interaction principle. With these points, we argue that these long disordered regions should be recognized as a distinct class of biologically functional protein domains.

Keywords: disordered domain; disorder in pfam; intrinsically disordered; intrinsically unstructured; pfam domain; unstructured domain

Introduction

Proteins are macromolecules that carry out most of the basic functions in the cell. A single protein in itself can contain substantial functional complexity, with distinct elements of its function usually ascribed to autonomous regions called domains.[1–3] Approaching protein function from the level of domains represents one of the most powerful paradigms in the molecular description of the cell. Basically, domains can exhibit catalytic, regulatory and recognition functions, and their combinatorial varieties supply an inexhaustible repertoire of building blocks for complex signaling/regulatory systems.[4,5] Domains can be grouped into families by sequence similarities that suggest a common evolutionary origin, and establishing such relationships has become an important first line of analysis in any attempt to uncover the function of a novel protein, or even to characterize an entire genome.[6]

There are three operative definitions for protein domains. The first definition is an autonomous structural unit of a protein,[7] as first suggested for the nucleotide-binding element of dehydrogenases, the Rossman fold.[8] This structural approach has given rise to the concept of a *protein fold*, which emphasizes the ability of a domain to acquire a well-defined tertiary structure on its own.[9] The second definition of a domain is a protein segment that can be recognized in distinct genetic contexts by virtue of sequence similarity. Such a segment is often called a *module*.[10] The third definition is that a domain is an interchangeable element of a protein with *functional autonomy*.[3] These three definitions highlight structural, evolutionary and functional aspects of protein modularity, but these different aspects do not necessarily all exist concurrently. Domains might not serve as individual functional units, for example, if an active site forms at the interface of two subunits, as in the case of pancreatic α-amylase.[11]

Given the distinctions among the three definitions, domains can be identified in multiple ways. The Pfam database[12] is based on hidden Markov models and multiple sequence alignments, emphasizing the evolutionary conservation of protein domains. The SMART approach for identifying domains focuses on genetic mobility.[13] Other domain databases employ structural definitions, classifying domains as autonomous folding units of proteins. For example, the CATH database contains a hierarchical

Abbreviations: ACTR, activator for thyroid hormone and retinoid receptors; BC, beta-catenin; CBD, catenin-binding domain; Cdk, cyclin-dependent kinase; CDP, conserved disorder prediction; ELM, eukaryotic linear motif; GBD, GTPase-binding domain; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; IFSU, intrinsically folded structural unit; KID, kinase-inhibitory domain (Cdk inhibitors); KID, kinase-inducible domain (CREB); MoRF, molecular recognition feature; MT, microtubule; MTB, MT-binding domain; NCBD, nuclear-receptor coactivator-binding domain; PCS, primary contact site; PSE, preformed structural element; Tcf3, T-cell factor3; WASP, Wiskott–Aldrich syndrome protein; WH2, WASP-homology domain 2.

***Correspondence to:** P. Tompa, Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, 1518 Budapest, Hungary.
E-mail: tompa@enzim.hu

 BioEssays 31:328–335,

classification of protein domain structures studied at four different levels,[14] and the SCOP database is based on an evolutionary classification that builds on conserved structural features.[15] In these last two cases, the domain concept is biased toward a structural definition, in which the domain is considered to be an autonomous folding unit of the protein.

Intrinsic disorder of proteins

The relationships among the different definitions of the term ''domain'' need to be examined in light of the newly emerging concept of protein disorder, because proteins–protein regions lacking a well-defined three dimensional structure constitute a large fraction of eukaryotic proteomes.[16–18] These intrinsically disordered proteins (IDPs) or regions (IDRs) exist as structural ensembles, and thus defy the classical concept that function of a protein derives from a well-defined 3D structure. IDPs/IDRs carry out important functions in key biological processes such as signal transduction and the regulation of transcription,[19] and their prevalence increases with increase in complexity of the organism. Bioinformatics analysis shows that that there is a high frequency of disorder in eukaryotic proteins, with about 50% of human proteins predicted to have at least one long (>30 consecutive residues) disordered region.[20] IDPs/IDRs function by a variety of mechanisms that can classified into two major categories. IDPs/IDRs either function by molecular recognition *via* binding to a partner, or they function as entropic chains by exploiting the physical benefits of the disordered state, as will be detailed below.

Upon binding to a macromolecular partner, the disordered regions undergo induced folding,[21] which is thought to confer functional advantages such as uncoupling specificity from binding strength, increasing the speed of interaction and providing adaptability, which results in functional promiscuity.[22,23] However, disordered proteins may remain disordered in the bound state.[24] Often, the recognition function of an IDP is mediated by a short segment of the protein, termed a eukaryotic linear motif (ELM),[25] molecular recognition feature, MoRF,[26] or preformed structural element (PSE).[27] All these recognition motifs have similar lengths (3–20 for ELMs, 4–15 for PSEs), but may function *via* different mechanisms.[26,28] PSEs and MoRFs benefit from transient secondary structural elements, while ELMs are exposed sites with unique physicochemical features. The function of entropic chains stems directly from the disordered state of proteins. They can provide conformational freedom by serving as linkers between subunits of multi-domain proteins.

IDPs/IDRs can span several hundred residues in length.[18] In many cases their distinct parts are associated with different functions, and are therefore referred to as domains in the literature.[16] The trans-activator domains of transcription factors are the classical example. These domains are responsible for communication with other transcriptional regulatory proteins and were noted long ago to be disordered.[29]

Pfam domains are often disordered

The foregoing points raise the possibility that disordered functional regions of proteins may conform to one or more of the definitions of domains. Corroborating this idea, all of the almost 3,000 domains in the InterPro database were predicted to have conserved, disordered segments of at least 20–30 amino acid residues.[30,31] Here, we further analyzed the level of structural disorder in Pfam domains (version 22.0, July 2007). In this analysis, domains were taken from SwissProt proteins in Pfam-A seed alignments. To ensure that the Pfam-A domains had known functions, we selected only domains that had GO annotations (*via* the InterPro database) or that had at least one literature citation. These restricted selections resulted in 71,974 examples of domains classified into 6,857 domain types. We found that 12.14% of the domains have >50% predicted disorder and 4.15% are fully disordered (95–100% predicted disorder, Supplementary Table S1) (Fig. 1). This implies that a noteworthy fraction of disordered proteins function as domains and suggests that the structural view of domains needs to be extended to encompass states that cannot be characterized by a well-defined fold. Since it is estimated that >50% of eukaryotic proteins have disordered segments of 30 or more residues, it is also possible that many disordered

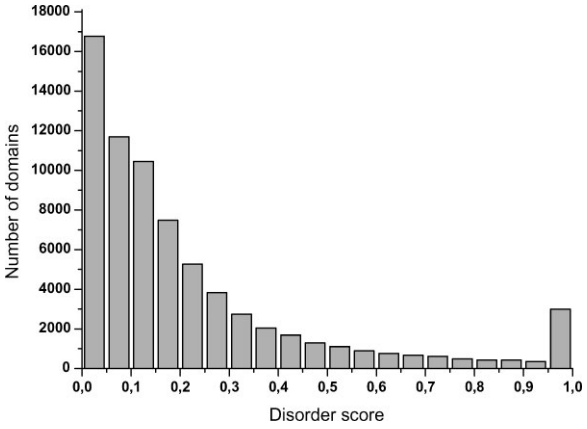


Figure 1. Distribution of Pfam domains with given fraction of disorder. Pfam domains were analyzed for the presence of predicted disorder (prediction was carried out using the PONDR® VSL2 algorithm). Domains were taken from SwissProt proteins in Pfam seed alignments, which gave 71,974 domain examples consisting of 6,857 domain types. The small peak at fully disordered domains (95–100% predicted disorder) results from the uneven length distribution of domains, with somewhat shorter domains dominating at high percentage of disorder (*cf.* text).

domains are not represented in Pfam. This may be due to evolutionary variability in the sequences of the disordered regions.[32]

Similarly to InterPro,[31] conserved protein segments that are consistently predicted to be disordered were identified in 40% of Pfam domains. Such regions, previously described as regions of conserved disorder prediction (CDP), were found in proteins from all taxonomic kingdoms, including viruses. The number of proteins containing long disordered regions was an order of magnitude higher in eukaryotes than in archaea and bacteria.[31] Functionally, CDP regions were found to be associated with DNA/RNA binding, protein binding, signaling/regulation, and complex formation. They were also frequently found in ribosomal proteins.[33] Hence, the abundance and biological relevance of CPDs in Pfam domains underscores the inherent functionality of disordered regions or even whole domains in the Pfam collection of sequences.

Examples of disordered domains

From the work described above, many Pfam domains are predicted to be disordered, and there are several examples in the literature characterizing such regions in detail with respect to both structural disorder and function. The classification of these regions as domains derives from both their length and noted functional autonomy. A collection of these domains are described here in some detail (*cf.* Fig. 2 and Table 1), with the primary focus on examples for which there is solid evidence of structural disorder (reference to disorder in Table 1), functional autonomy and evolutionary divergence. The most common function associated with these domains is molecular recognition, during which they undergo a disorder-to-order transition, but in some cases they may remain disordered and function as entropic chains.[16,34]

Several different proteins contain homologous disordered[35] binding regions called catenin-binding domains (CBD, Fig. 2B). The partner of the CBD, beta-catenin (BC) is a regulatory protein that functions in the WNT signaling pathway. In the cell, BC is found in complex with two distinct partners, the disordered intracellular domains of either E-cadherin or APC/axin.[36] Activation of the WNT pathway results in the release of BC from the latter complex and its translocation to the nucleus, where it acts as a transcription cofactor for the CBD-containing T-cell factors LEF-1, Tcf3, and Tcf4.

The small, disordered[37] Wiskott–Aldrich syndrome protein (WASP)-homology domain 2 (WH2, Fig. 2C), which is about 40 amino acids in length, appears in several actin-binding proteins such as thymosin-β4, ciboulot, and WASP.[38] This domain occurs in different sequence contexts, but is almost always embedded in a Pro-rich region. Conserved features of the domain (Fig. 3) suggest that in all of its homologs it functions in actin binding, but the outcome of this binding is context-dependent. A single WH2 domain (*e.g.*, in thymosin-β4) inhibits actin polymerization and sequesters G-actin, whereas several tandem WH2 domains (*e.g.*, in ciboulot) promote actin polymerization, most likely due to binding several actin monomers in close proximity.[39]

The modular nature of the fully disordered[40,41] microtubule (MT)-associated proteins tau, MAP2, and MAP1c has been

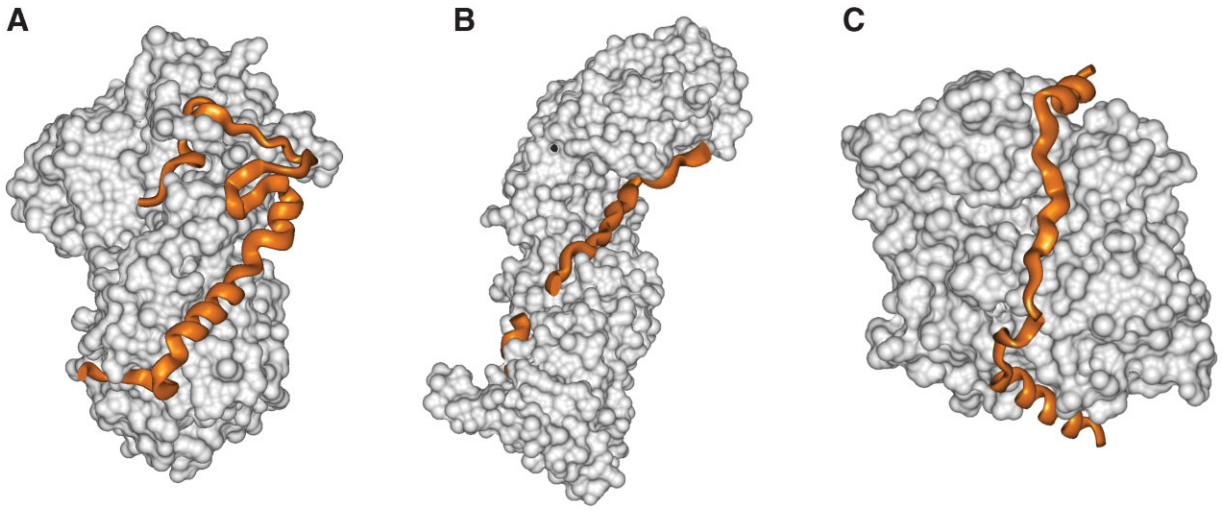


Figure 2. Structures of select examples of disordered domains. Structures of disordered regions of proteins in complex with their partners are shown. **(A)** The kinase-inhibitory domain (KID) of Cdk inhibitor p27Kip1 bound to the CycA/Cdk2 complex (pdb 1jsu). **(B)** The CBD of T-cell factor3 (Tcf3) in complex with BC (pdb 1g3j). **(C)** The WH2 domain of thymosin-β4 in complex with G-actin (a composite of two structures, pdb 1t44 and 1sqk). In each case there is conclusive experimental evidence for the disorder and functional autonomy of these regions, termed "domains" in the literature.

Table 1. Selected examples of disordered domains

Disordered domain	Protein	Partner	Function	Kingdom coverage[a]	Ref./Interpro
WH2	Thymosin-β4, ciboulot, WASP, WIP, verprolin, spire	Actin	m.r.	444/0/32	[38], IPR003124
CKI	p21, p27, p57	cyclin/Cdk	m.r.	39(118)/0/0	[53], PF02234
MT-binding domain (MTB)	Tau, MAP2, MAP1c	MT	m.r.	96/0/0	[42,43], IPR001084
Projection domain	Tau, MAP2, MAP1c	N/A	Entropic spacer	16/0/0	[47], IPR013588
Sidearm domain	Neurofilaments	N/A	Entropic spacer	14/1/16	[64,65], IPR010790
CBD	E-cadherin, LEF-1, Tcf3, Tcf4	BC	m.r.	299/0/0	[35], IPR000233
Sma-binding domain (SBD)	Smad anchor for receptor activation (SARA)	Smad2	m.r.	11/0/0	[66], IPR017165
GBD	WASP	GTPase cdc42, VCA domain/WASP	m.r.	23/0/0	[62], IPR015116
Kinase-inducible domain (KID)	CREB	CREB-binding protein (CBP)	m.r.	120/0/0	[53], IPR003102
Regulatory domain (R)	CFTR	Nucleotide-binding domain of CFTR	m.r.	106/0/0	[67], IPR009147
Prion domain	Yeast prion Sup35p, Ure2p	Autoaggregation	m.r.	587/0/1	[46], IPR000817
FG-repeat domain of nucleoporins	FG-NUPs, *e.g.*, Nup153	Karyopherins	Entropic filter, m.r.	13/0/0	[68], IPR013913
NCBD	CBP	ACTR domain of p160	m.r.	67/0/0	[60], IPR009110

m.r., molecular recognition; N/A, non-applicable.
[a]Number of domains in Eukaryotes/Archaea/Bacteria.

inferred from limited proteolysis, which separates these proteins into two large segments:[42,43] the projection domain and the MT-binding domain (MTB). MAPs are thought to associate with MTs through their homologous MTBs, which are composed of three to four copies of a 32-amino acid repeat. Binding of MTB domains to MTs promotes their assembly.

An illustrative example of a functional CDP region within a larger disordered domain is found in prion proteins (Fig. 4). The CDP region found in the prion family extends, on average, from residues 36 to 122.[33] The disordered state of this region was confirmed by NMR in a number of prion proteins, including those from humans and hamsters.[44] Prion proteins are capable of undergoing a self-sustaining conformational change, from a soluble cellular form to an insoluble aggregate, in an autocatalytic manner, and can be transmitted to different organisms.[45] Full

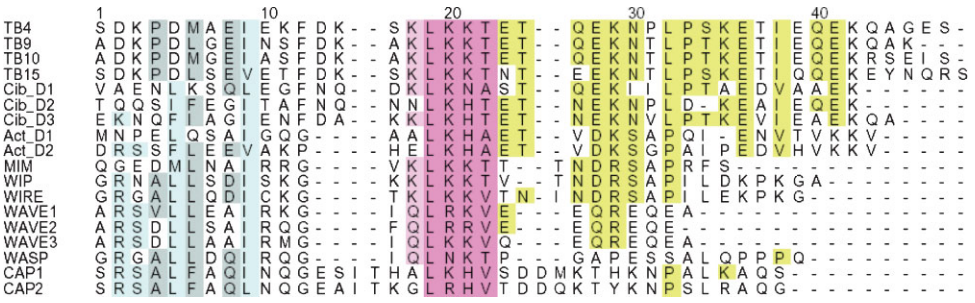


Figure 3. Multiple sequence alignment of the WH2 domain. Sequence alignment of the WH2 domains generated by the Clustalw program. Residues are colored as follows (based on ref.[55]): the critical conserved binding element LKKT is magenta, residues homologs to thymosine-β4 are gray, conserved residues in the WH2 family are blue and conserved residues in thymosine-β4 are yellow. Conservation outside the short recognition motif illustrates the difference between a linear motif and a disordered domain.

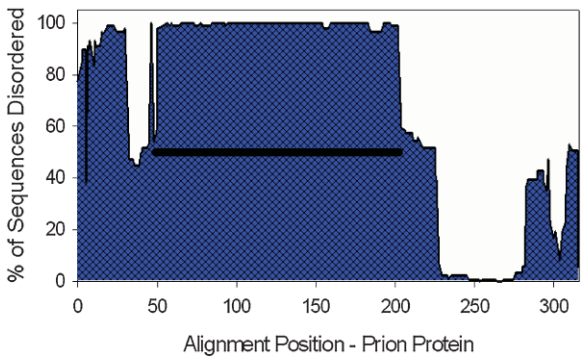


Figure 4. Conservation of disorder in the disordered domains of over 100 prion sequences. Plot shows residue-based disorder prediction for a prion family containing a long CDP region (shown as a bold black line). The prediction was performed with the PONDR-VL-XT predictor, using a set of aligned sequences including gaps. The nature (disordered/ordered) of the gaps was decided by a coin-flip procedure. Due to the large sequence variability of the disordered N-terminal region many gaps were introduced, considerably increasing the length of this segment. Therefore, the length of the aligned sequence is longer than the longest sequence analyzed and the position of the CDP region on the graph does not correspond exactly to the position of the CDP region in a single (*e.g*, human, *cf.* text) protein. Modified from ref.[37].

functional autonomy of the prion CDP is demonstrated by the fact that these disordered prion domains are transportable, *i.e.*, they can confer prion characteristics on other proteins,[46] and they can be detached from the associated globular domain without impairing the structure and/or function of either domain.

Disordered domains may also function as entropic chains. One example is the projection domain of MAPs, mentioned above, which protrudes from the surface of MTs and exerts long-range force against extension or compression, thus ensuring proper spacing in the cytoskeleton.[47] The FG repeat domain of nuclear pore subunits (Nups), which form the nuclear pore complex (NPC), is another example of entropic chains. The NPC is a huge assembly of approximately 50 MDa that selectively transports molecular cargo across the nuclear envelope.[48] In yeast, the NPC is made up of about 30 different Nups, 13 of which contain long phenylalanine-glycine repeats (hence termed FG Nups), which are intrinsically disordered both *in vitro* and *in vivo*.[49] These disordered appendages act in NPC gating by entropic exclusion.[50]

Further, examples of disordered regions that have been called domains and carry autonomous functions are collected in Table 1. In the most extensive repository of disordered proteins currently available, the DisProt database,[18] 69 out of the 484 entries have been designated as 'domains' by the researchers who study them, which underscores the likely prevalence of this structural phenomenon.

Disordered domains differ from short recognition motifs

The disordered Pfam domains are somewhat shorter than their ordered counterparts (the average length of domains with 0–25% predicted disorder is 176.1 residues, while those with 75–100% predicted disorder average only 98.7 residues.), but they are definitely longer than the short recognition motifs (ELMs, MoRFs, and PSEs) described above. We propose that disordered domains which are above a threshold length of about 20–30 residues and which are involved in partner recognition/binding are principally different from motifs. Motifs are effective evolutionary switches that can be turned on and off more-or-less at random by point mutations.[25,51] Their occurrence within different contexts probably represents an example of evolutionary convergence. Domains, in contrast, are functional units that spread in the genome by inheritance, and their presence indicates divergence.[1]

A whole range of functional observations underscore this distinction. First, there is a clear functional difference between the short motifs and the short conserved segments within disordered domains. As epitomized by the binding of Pro-rich segments to SH3 domains,[52] such short motifs appear as a few conserved residues embedded within a variable stretch of amino acids, which can confer a recognition function within the context of many independent molecular settings.[25] Conserved segments within disordered domains, on the other hand, only carry function as a part of the entire domain. For example, a conserved Leu^{32}-Phe-Gly motif in the KID domain of Cdk inhibitors p21, p27, and p57 (Fig. 2A) initiates binding at the cyclin subunit, but it can only function as an inhibitor together with the remote element Tyr^{88}, which blocks the active site of the associated kinase.[53] Phosphorylation of Tyr^{88} can relieve the inhibition of the kinase, thereby promoting unimolecular phosphorylation at Tyr-187 in the C-terminus, which in turn serves as a signal for degradation of p27.[54] This illustrates that the intrinsic disorder of the p27 domain allows independent regulation of different segments of the Cdk inhibitor proteins, which can alter protein function. In the WH2 domain of thymosin-β4 the critical, conserved binding element is the segment Leu^{17}-Lys-Lys-Thr-Glu-Thr (*cf.* Fig. 3). G-actin sequestration, however, also requires binding and capping of G-actin by two distinct helical regions of thymosin-β4, between residues 5–16 (helix 1) and 30–39 (helix 2), which prevent G-actin from joining either the pointed or barbed ends of actin filaments.[55] Residues Glu^{24}-Glu^{28} constitute a critical "hot spot" for catenin binding by the CBD of Tcf4. Mutation of this regions results in a significant reduction of binding enthalpy.[56] Mutations of Leu^{41}, Val^{44}, and Leu^{48} within a remote amphipathic helix show that this region is also critically involved in the binding of BC.

The distinction between motifs and domains is also evident upon comparison of the mode and affinity of binding. Short

motifs often exhibit low binding affinities, typically in the range of 1–10 μM,[25] whereas disordered domains may demonstrate much stronger binding, often in the nanomolar range, as in the following interactions: p27-KID–CycA-Cdk2 (3.5 nM),[53] Tcf4-CBD–BC (8.4 nM)[57], and WASP-GBD–Cdc42 (63 nM).[58] In analyzing these differences, it appears that short motifs are often involved in transient interactions during post-translational modification, whereas binding of domains almost always results in stable complexes. This point is illustrated by the example of the disordered KID domain of CREB binding to the ordered KID-binding (KIX) domain of CBP (*cf*. Table 1). This interaction is promoted by PKA-dependent phosphorylation of KID at Ser133,[59] which mediates cAMP-dependent activation of the transcription factor by interaction with its coactivator. KID acts as a domain, because the whole conserved region is required for stable KIX binding and function. The PKA phosphorylation site, on the other hand, relies on a few residues only, engages in a transient interaction with the active site of PKA, and should be considered to be an ELM. In theory, this ELM could be converted by a few point mutations from a PKA to PKC-regulated site, while leaving the functional readout, as assessed by the interaction of KID and KIX domains, intact.

The principal distinction between ELMs and domains is that disordered domains, but not ELMs, are able to bind not only an ordered partner, but also to each other in a process of mutual induced folding. For example, the action of nuclear hormone receptors is mediated through recruitment and interaction of p160 receptor coactivator and the general transcriptional coactivator CBP/p300, mediated by the nuclear-receptor coactivator-binding domain (NCBD) of CBP and the activator for thyroid hormone and retinoid receptors (ACTR) domain of p160.[60] Both domains are disordered prior to binding (Table 1), and form a complex (Fig. 5A) in a process termed mutual synergistic folding.[60]

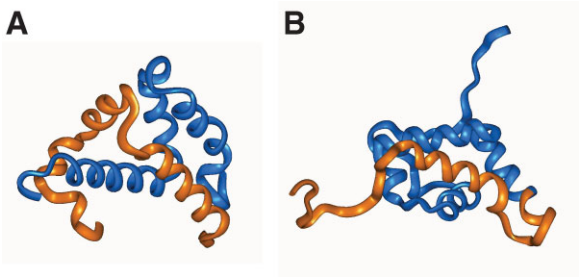


Figure 5. Mutual synergistic folding or cofolding of two disordered regions/domains. Disordered domains may not only bind an ordered partner, but may also be engaged in interaction with another disordered domain. This mutual synergistic folding (also termed cofolding) is exemplified by **(A)** the structure of the NCBD of CBP (blue) in complex with the ACTR domain of p160 (orange, pdb 1kbh) and **(B)** the structure of region C of WASP (orange) in complex with the GBD (blue) of WASP (pdb 1ej5).

Another example is the WASP, which mediates the effect of Cdc42 on cytoskeleton assembly.[61] WASP has a disordered GTPase-binding domain (GBD, Table 1), which can bind either the disordered carboxy-terminal VCA region of the same molecule (Fig. 5B) or Cdc42, in a mutually exclusive fashion.[62,63]

Conclusion

Disordered domains represent a third kind of protein–protein recognition

The foregoing examples and discussions constitute a compelling argument for the classification of certain disordered regions of proteins as domains. These regions conform to all three domain definitions, which often do not hold true simultaneously even in the case of ''classical'' domains. Disordered domains, by definition, serve as functional elements, and they are also often found in different genetic contexts. Although it may seem perplexing, they are also structurally autonomous, because they can be taken out of context while preserving their structural state and function.

Given all of the above, we propose that disordered domains are a new type of protein structure. To clarify whether a given protein or region is a member of this new structural classification, we define disordered domains as lengthy regions that display the following four characteristics: (1) they are structurally and functionally independent of the remainder of the protein molecule (or they constitute the entire protein); (2) they can be recognized by homology due to evolutionary conservation of sequence; (3) they are structurally characterized and/or predicted as disordered; and (4) they possess at least one specific biological function. Actually, regions with several of these characteristics (and sometimes with all four) have been termed domains in the literature for a long time, without reference to their actual structural status. Our key point is that with the recent advent of the concept of protein disorder, the disordered domain ought to be recognized as a distinct element of protein functionality.

Acknowledgments: This research was supported by grants OTKA K60694 and NK71582 from the Hungarian Scientific Research Fund, ETT 245/2006 from the Hungarian Ministry of Health, International Senior Research Fellowship ISRF 067595 from the Wellcome Trust (to P. T.), OTKA K72569 (to I. S.), a Bolyai Fellowship (M. F.) and MRTN-CT-2005-019566 of the European FP6 (to M.F.), and R01 LM007688-01A1 (to A. K. D., C. J. O., and V. N. U.) and R01 GM071714-01A2 (to A. K. D., C. J. O., and V. N. U.) from the National Institutes of Health. The authors gratefully acknowledge the support of the IUPUI Signature Centers Initiative.

References

1. **Ponting, C. P., Schultz, J., Copley, R. R., Andrade, M. A. and Bork, P.,** Evolution of domain families. *Adv Protein Chem* 2000. **54**: 185–244.
2. **Copley, R. R., Goodstadt, L. and Ponting, C.,** Eukaryotic domain evolution inferred from genome comparisons. *Curr Opin Genet Dev* 2003. **13**: 623–628.
3. **Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. and Teichmann, S. A.,** Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 2004. **14**: 208–216.
4. **Pawson, T. and Nash, P.,** Assembly of cell regulatory systems through protein interaction domains. *Science* 2003. **300**: 445–452.
5. **Seet, B. T., Dikic, I., Zhou, M. M. and Pawson, T.,** Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 2006. **7**: 473–483.
6. **Xie, L. and Bourne, P. E.,** Functional coverage of the human genome by existing structures, structural genomics targets, and homology models. *PLoS Comput Biol* 2005. **1**: e31.
7. **Wetlaufer, D. B.,** Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA* 1973. **70**: 697–701.
8. **Rao, S. T. and Rossmann, M. G.,** Comparison of super-secondary structures in proteins. *J Mol Biol* 1973. **76**: 241–256.
9. **Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A. and Clarke, J.,** The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* 2007. **8**: 319–330.
10. **Patthy, L.,** Exon shuffling and other ways of module exchange. *Matrix Biol* 1996. **15**: 301–310.
11. **Larson, S. B., Greenwood, A., Cascio, D., Day, J. and McPherson, A.,** Refined molecular structure of pig pancreatic alpha-amylase at 2.1 A resolution. *J Mol Biol* 1994. **235**: 1560–1584.
12. **Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L.,** *et al.* The Pfam protein families database. *Nucleic Acids Res* 2002. **30**: 276–280.
13. **Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., Bork P.,** SMAR 5: domains in the context of genomes and networks. *Nucleic Acids Res* 2006. **34**: D257–D260.
14. **Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., Thornton J. M.,** CATH–a hierarchic classification of protein domain structures. *Structure* 1997. **5**: 1093–1108.
15. **Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C., Murzin, A. G.,** SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* 2004. **32**: D226–D229.
16. **Tompa, P.,** Intrinsically unstructured proteins. *Trends Biochem Sci* 2002. **27**: 527–533.
17. **Uversky, V. N., Oldfield, C. J. and Dunker, A. K.,** Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 2005. **18**: 343–384.
18. **Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S.,** *et al.* DisProt: the Database of disordered proteins. *Nucleic Acids Res* 2007. **35**: D786–D793.
19. **Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K.,** *et al.* Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 2007. **6**: 1882–1898.
20. **Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. and Brown, C. J.,** Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 2000. **11**: 161–171.
21. **Dyson, H. J. and Wright, P. E.,** Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 2002. **12**: 54–60.
22. **Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. and Wright, P. E.,** Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* 1996. **93**: 11504–11509.
23. **Tompa, P., Szasz, C. and Buday, L.,** Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 2005. **30**: 484–489.
24. **Tompa, P. and Fuxreiter, M.,** Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem Sci* 2008. **33**: 2–8.
25. **Neduva, V. and Russell, R. B.,** Linear motifs: evolutionary interaction switches. *FEBS Lett* 2005. **579**: 3342–3345.
26. **Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S.,** *et al.* Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* 2007. **6**: 2351–2366.
27. **Fuxreiter, M., Simon, I., Friedrich, P. and Tompa, P.,** Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 2004. **338**: 1015–1026.
28. **Fuxreiter, M., Tompa, P. and Simon, I.,** Structural disorder imparts plasticity on linear motifs. *Bioinformatics* 2007. **23**: 950–956.
29. **Sigler, P. B.,** Transcriptional activation. Acid blobs and negative noodles. *Nature* 1988. **333**: 210–212.
30. **Dosztanyi, Z., Sandor, M., Tompa, P. and Simon, I.,** Prediction of protein disorder at the domain level. *Curr Protein Pept Sci* 2007. **8**: 161–171.
31. **Chen, J. W., Romero, P., Uversky, V. N. and Dunker, A. K.,** Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J Proteome Res* 2006. **5**: 879–887.
32. **Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W.,** *et al.* Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* 2002. **55**: 104–110.
33. **Chen, J. W., Romero, P., Uversky, V. N. and Dunker, A. K.,** Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J Proteome Res* 2006. **5**: 888–898.
34. **Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. and Obradovic, Z.,** Intrinsic disorder and protein function. *Biochemistry* 2002. **41**: 6573–6582.
35. **Huber, A. H., Stewart, D. B., Laurents, D. V., Nelson, W. J. and Weis, W. I.,** The cadherin cytoplasmic domain is unstructured in the absence of beta-catenin. A possible mechanism for regulating cadherin turnover. *J Biol Chem* 2001. **276**: 12301–12309.
36. **Shapiro, L.,** Beta-catenin and its multiple partners: promiscuity explained. *Nat Struct Biol* 2001. **8**: 484–487.
37. **Domanski, M., Hertzog, M., Coutant, J., Gutsche-Perelroizen, I., Bontems, F.,** *et al.* Coupling of folding and binding of thymosin beta4 upon interaction with monomeric actin monitored by nuclear magnetic resonance. *J Biol Chem* 2004. **279**: 23637–23645.
38. **Paunola, E., Mattila, P. K. and Lappalainen, P.,** WH2 domain: a small, versatile adapter for actin monomers. *FEBS Lett* 2002. **513**: 92–97.
39. **Chereau, D., Kerff, F., Graceffa, P., Grabarek, Z., Langsetmo, K., Dominguez, R.,** Actin-bound structures of Wiskott–Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc Natl Acad Sci USA* 2005. **102**: 16644–16649.
40. **Schweers, O., Schonbrunn-Hanebeck, E., Marx, A. and Mandelkow, E.,** Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J Biol Chem* 1994. **269**: 24290–24297.
41. **Hernandez, M. A., Avila, J. and Andreu, J. M.,** Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. *Eur J Biochem* 1986. **154**: 41–48.
42. **Steiner, B., Mandelkow, E. M., Biernat, J., Gustke, N., Meyer, H. E.,** *et al.* Phosphorylation of microtubule-associated protein tau: identification of the site for Ca2(+)-calmodulin dependent kinase and relationship with tau phosphorylation in Alzheimer tangles. *EMBO J* 1990. **9**: 3539–3544.
43. **Wille, H., Mandelkow, E. M. and Mandelkow, E.,** The juvenile microtubule-associated protein MAP2c is a rod-like molecule that forms antiparallel dimers. *J Biol Chem* 1992. **267**: 10737–10742.
44. **Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L.,** *et al.* Structure of the recombinant full-length hamster prion protein PrP(29–231): the N terminus is highly flexible. *Proc Natl Acad Sci USA* 1997. **94**: 13452–13457.
45. **Chien, P., Weissman, J. S. and DePace, A. H.,** Emerging Principles of Conformation-based prion inheritance. *Annu Rev Biochem* 2004. **73**: 617–656.
46. **Wickner, R. B., Taylor, K. L., Edskes, H. K. and Maddelein, M. L.,** Prions: portable prion domains. *Curr Biol* 2000. **10**: R335–R337.
47. **Mukhopadhyay, R. and Hoh, J. H.,** AFM force measurements on microtubule-associated proteins: the projection domain exerts a long-range repulsive force. *FEBS Lett* 2001. **505**: 374–378.
48. **Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J.,** *et al.* The molecular architecture of the nuclear pore complex. *Nature* 2007. **450**: 695–701.

49. **Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. and Rexach, M.,** Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc Natl Acad Sci USA* 2003. **100**: 2450–2455.
50. **Patel, S. S., Belmont, B. J., Sante, J. M. and Rexach, M. F.,** Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* 2007. **129**: 83–96.
51. **Davey, N. E., Shields, D. C. and Edwards, R. J.,** SLiMDisc: short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res* 2006. **34**: 3546–3554.
52. **Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W.,** *et al.* Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* 1994. **76**: 933–945.
53. **Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L.,** *et al.* p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat Struct Mol Biol* 2004. **11**: 358–364.
54. **Galea, C. A., Nourse, A., Wang, Y., Sivakolundu, S. G., Heller, W. T.,** *et al.* Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J Mol Biol* 2008. **376**: 827–838.
55. **Irobi, E., Aguda, A. H., Larsson, M., Guerin, C., Yin, H. L.,** *et al.* Structural basis of actin sequestration by thymosin-beta4: implications for WH2 proteins. *EMBO J* 2004. **23**: 3599–3608.
56. **Fasolini, M., Wu, X., Flocco, M., Trosset, J. Y., Oppermann, U.,** *et al.* Hot spots in Tcf4 for the interaction with beta-catenin. *J Biol Chem* 2003. **278**: 21092–21098.
57. **Knapp, S., Zamai, M., Volpi, D., Nardese, V., Avanzi, N.,** *et al.* Thermodynamics of the high-affinity interaction of TCF4 with beta-catenin. *J Mol Biol* 2001. **306**: 1179–1189.
58. **Rudolph, M. G., Bayer, P., Abo, A., Kuhlmann, J., Vetter, I. R.,** *et al.* The Cdc42/Rac interactive binding region motif of the Wiskott–Aldrich syndrome protein (WASP) is necessary but not sufficient for tight binding to Cdc42 and structure formation. *J Biol Chem* 1998. **273**: 18067–18076.
59. **Parker, D., Ferreri, K., Nakajima, T., LaMorte, V. J., Evans, R.,** *et al.* Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism. *Mol Cell Biol* 1996. **16**: 694–703.
60. **Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W.,** *et al.* Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 2002. **415**: 549–553.
61. **Caron, E.,** Regulation of Wiskott–Aldrich syndrome protein and related molecules. *Curr Opin Cell Biol* 2002. **14**: 82–87.
62. **Abdul-Manan, N., Aghazadeh, B., Liu, G. A., Majumdar, A., Ouerfelli, O.,** *et al.* Structure of Cdc42 in complex with the GTPase-binding domain of the 'Wiskott-Aldrich syndrome' protein. *Nature* 1999. **399**: 379–383.
63. **Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A. and Rosen, M. K.,** Autoinhibition and activation mechanisms of the Wiskott–Aldrich syndrome protein. *Nature* 2000. **404**: 151–158.
64. **Brown, H. G. and Hoh, J. H.,** Entropic exclusion by neurofilament sidearms: a mechanism for maintaining interfilament spacing. *Biochemistry* 1997. **36**: 15035–15040.
65. **Kumar, S., Yin, X., Trapp, B. D., Hoh, J. H. and Paulaitis, M. E.,** Relating interactions between neurofilaments to the structure of axonal neurofilament distributions through polymer brush models. *Biophys J* 2002. **82**: 2360–2372.
66. **Wu, G., Chen, Y. G., Ozdamar, B., Gyuricza, C. A., Chong, P. A.,** *et al.* Structural basis of Smad2 recognition by the Smad anchor for receptor activation. *Science* 2000. **287**: 92–97.
67. **Baker, J. M., Hudson, R. P., Kanelis, V., Choy, W. Y., Thibodeau, P. H.,** *et al.* CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat Struct Mol Biol* 2007. **14**: 738–745.
68. **Lim, R. Y., Huang, N. P., Koser, J., Deng, J., Lau, K. H.,** *et al.* Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc Natl Acad Sci USA* 2006. **103**: 9512–9517.