Biochemistry

A.12

Volume 41, Number 21 May 28, 2002

Current Topics

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, Washington 99164 and Center for Information Science and Technology, Temple University, Philadelphia, Pennsylvania 19122

Received December 14, 2001; Revised Manuscript Received April 2, 2002

The dominant view of protein structure−function is that an amino acid sequence specifies a three-dimensional (3-D) structure that is a prerequisite for protein function. In contrast, many proteins display functions *requiring* intrinsic disorder. Our purpose here is to catalog and analyze disorder−function relationships. Molecular details can be obtained from the references provided or from several excellent reviews and commentaries (*1*−*4*).

For ordered proteins, most ensemble members have the same time-averaged canonical set of Ramachandran angles. For intrinsically disordered protein, the ensemble members have quite different (and typically dynamic) Ramachandran angles. Such disorder has been characterized by X-ray crystallography, NMR spectroscopy, CD[1] spectroscopy, and protease sensitivity. Each method has advantages and limitations (*5*).

To search for generalities from known disorder examples, we used bioinformatics coupled with data mining (*6*−*10*). The results suggest that thousands of natively disordered proteins exist, representing a substantial fraction of the commonly used sequence databases (*8*, *11*). From these and related database predictions and from a set of functionally important disordered proteins, Wright and Dyson called for a reassessment of the view that 3-D structure is prerequisite to protein function (*12*).

How Common Is Intrinsic Disorder? A series of *p*redictors *o*f *n*atural *d*isordered *r*egions (PONDRs) have been developed that take amino acid sequence inputs and give disorder tendency outputs (*6*, *9*, *10*, *13*, *14*). The various PONDRs are distinguished by training sets, data representations for their inputs, and machine learning models for their development.

For PONDR VL-XT,[2] currently the best characterized of the PONDRs, only 6% of more than 900 nonhomologous proteins spanning PDB give false positive predictions of disorder ≥ 40 consecutive amino acids in length. Even this may overestimate the false positive error rate. Many of the regions of apparent prediction errors are involved in ligand binding or in crystal contacts (Lawson and Dunker, unpublished data) and are actually intrinsically disordered as predicted when the parent proteins are in solution without their ligands. Furthermore, only ∼0.4% of more than 177 000 regions of 40 from these 900 nonhomologous proteins were falsely predicted to be disordered (*5*). 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. Disorder was estimated by the %

[†] 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.

* 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.

[‡] Washington State University.

[§] Temple University.

[1] 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)

[2] This predictor can be accessed at http://www.pondr.com.

10.1021/bi012159+ CCC: $22.00
Published on Web 05/01/2002

of proteins in each genome with ≥40 consecutive disorder predictions, giving the following results: bacteria = 6−33%, archaea = 9−37%, eukaryota = 35−51% (*15*). 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 (*16*). The possibility that the increased disorder in eukaryotes is an artifact arising from training or prediction biases also needs 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 (*17*), and additional examples are featured in literature reviews (*1*−*3*), intrinsic disorder has been mostly ignored. Perhaps one reason was a belief that disorder is an artifact because protease digestion would eliminate such proteins in vivo. Kim and Frankel, however, argued that sequestering proteases by compartmentalization would protect disordered proteins (*3*). Tight regulation of intracellular protease activity can also be understood as a mechanism to protect protease-sensitive sites (*5*, *12*).

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 hop from one partner to another. For example, at low calcium concentrations in vitro, calmodulin becomes sensitive to protease digestion (*18*). As calcium concentration drops in vivo, calmodulin leaves its enzymatic partners and binds to proteins with IQ motifs (*19*). Similarly, some chaperones may function as bodyguards for the protection of intrinsically disordered regions in addition to their proposed function of promoting protein folding (*5*).

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 partner often leads to the disappearance of the remaining protein (K. Broadie, J. Tyler, and L. Hall, personal communications), perhaps from increased protease sensitivity. Protein−protein interactions 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 (*12*).

Molecular crowding favors more compact forms over more extended ones and has been shown to markedly shift the equilibrium toward the folded state (*20*). 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 (*21*) and TFIIIA binding to its target DNA (*22*). In both cases, the protein wraps around its partner. Second, molecular crowding should induce folding into a particular 3-D structure only if the protein has 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 (*23*), such crowding does not induce regular secondary structure for three intrinsically disordered proteins: c-Fos (*24*), p27kip1 (*25*), and α-synuclein (*23*).

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 (*5*). 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.

We 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 (*26*). 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. More than 150 proteins, 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).

The proteins described in Table 1 contain numerous disordered regions involved in molecular recognition, indicated in the function column of Table 1 by the letters (a−j) and are further characterized in Table 2. These molecular interactions involve binding to other proteins (a), such as to kinases, transcription factors, and translation inhibitors, or 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) (*202*). 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).

The involvement of intrinsic disorder in molecular recognition may at first seem counterintuitive. Upon further reflection, however, the involvement of intrinsic disorder in molecular recognition is useful for enabling (i) high specificity coupled with low affinity (*203*) because the free energy arising from the contacts of protein with ligand is reduced by the free energy needed to fold the intrinsic disorder; (ii)

Table 1: Intrinsically Disordered Protein and Their Functions

	protein (GenPept ID)	detection method[a]	location	chain length	protein length	class[b]	function[c]	refs
1a	acidic ribosomal stalk protein P1(133069)	NMR, CD LP, other	1−106	106	106	D, MG	a, cR, h, mP, l	(*25*, *27*−*29*)
1b	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[a]	b, n, t	(*30*−*32*)
3	antibacterial protein LL-37 (1706745)	CD	1−37	37	37	D−O	a, g	(*33*)
4	antitermination protein N of bacteriophage λ (132276)	NMR, CD	73−107	107	107	D−O	a, cM	(*34*, *35*)
5	apocytochrome *c* (117995)	CD	1−104	104	104	D−O	i	(*36*)
6	apurinic/apyrimidinic endonuclease (299037)	X-ray, LP	1−40	279	317	U	x	(*37*−*39*)
7	B cell-specific transcription co-activator (1150493)	CD, LP	1−65	65	256	D−O	a, b	(*40*)
			66−256	256	256	D−O[e]	a	
8	Bcl-x_L antiapoptotic protein (2392082)	X-ray, NMR	28−80	196	233	D	mP, l	(*41*−*43*)
9	Bcl-2 antiapoptotic protein (13786963)	other	24−93	196	239	U[e]	h, mP[e], l[e]	(*42*, *44*−*47*)
10	calcineurin (1352673)	X-ray, LP	374−468	521	521	O−D, D−O[e]	a, k[e]	(*48*, *49*)
			487−521			U	x	
11	α_s-casein (1070620)	CD	1−190	190	190	D	v	(*50*)
12	catecholamine sulfotransferase (1711609)	X-ray	216−261	295	295	D−O[e]	h	(*51*, *52*)
13	CFTR (14753227)	CD, LP	708−831	124	1480	D	k, mP	(*53*)
14	chorionic gonadotropin (116184)	X-ray	B:112−145	145	145	D	mG, w	(*54*, *55*)
15	clusterin (461756)	LP	66−97	447	447	D, MG	v	(*56*)
			386-445				v	
16	clotting factor Xa (119761)	X-ray, LP	L:1−46	146	488	U	x	(*57*−*59*)
17	cAMP-dependent protein kinase inhibitor (417194)	CD, X-ray	1−76	77	76	D−O	a	(*60*, *61*)
18a	cyclin-dependent kinase inhibitor p21 (729143)	NMR, CD, LP	1−164	164	164	U, D−O	a	(*62*)
18b	cyclin-dependent kinase inhibitor p57 (11440665)	CD, other	1−316	316	316	D	a	(*63*)
18c	cyclin-dependent kinase inhibitor p27 (1168871)	CD, X-ray, other	23−106	84	198	D−O	a	(*24*, *64*)
19	cytochrome bc1 complex (1351360)	X-ray	subunit I: 1−43	78	78	U	x	(*65*)
			subunit E: 76−196	196	196			
20	DNA-binding protein RAP1 (730473)	X-ray, LP	482−512	246	827	U	x	(*66*)
21	elongation factor G (1827912)	X-ray	400−480	691	691	U	cR[e], k, n[e]	(*67*−*69*)
22	embryonic abundant protein from carrot (119316)	NMR	1−92	92	92	U	h[e]	(*70*)
23	estrogenic 17-β hydroxysteroid dehydrogenase (2392375)	X-ray	285−327	327	327	U	f[e], w	(*71*)
24	E7 protein from HPV16 (6469700)	CD	1−93	93	93	D−O	a, j	(*72*)
25	fibronectin binding protein (120457)	NMR	745−873	130	1018	D−O	a	(*73*−*75*)
26	flagellin (96744)	NMR, X-ray[d], CD, other	1−65	various	494	D, D−O	a, g, s	(*76*−*80*)
			451−494					
27	flagellum specific σ factor (120306)	NMR	1−97	97	97	D, D−O	a, s	(*81*, *82*)
28	4E-binding protein 1 (4758258)	NMR, CD, X-ray	1−118	118	118	D, D−O	a, mP	(*83*−*85*)
29	glial cell-derived neurotrophic factor (729568)	X-ray	1−40	135	135	D−O[e]	a[e]	(*86*−*88*)
30	glucocorticoid receptor (4758482)	CD, other	77−262	186	835	D−O	a	(*89*)
31	glycine *N*-methyltransferase (121328)	X-ray	1−40	292	292	U, D−O	h	(*90*, *91*)
32	glycyl-tRNA synthetase (2829475)	X-ray	96−158	442	505	U	x	(*92*, *93*)
33	growth hormone receptor (121180)	X-ray	1−31	238	620	D−O	a, w	(*94*, *95*)
34	g3p (fd phage minor coat protein) (5822481)	X-ray, NMR	218−256	225	406	D	n	(*96*−*98*)
35	G protein $G_{i\alpha1}$ (121020)	X-ray	1−33	353	353	D, D−O	a, f, mF	(*99*, *100*)
36	heat shock transcription factor (123686)	NMR, CD	1−193	282	677	D	a	(*101*)
37	heparin-binding EGF-like growth factor (544477)	X-ray	73−106	79	189	U	mG	(*102*, *103*)

Table 1 (Continued)

	protein (GenPept ID)	detection method[a]	location	chain length	protein length	class[b]	function[c]	refs
38a	high mobility group − 14 (4826758)	NMR, CD, other	1−100	100	100	D, D−O	a, b, mP, mA	(*104*−*107*)
38b	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	(*108*−*110*)
40a	high mobility group − T (123382)	NMR, CD	1−204	204	204	D−O	b	(*111*)
40b	high mobility group − H6 (462245)		1−69	69	69			
41	histone H3 (211857)	X-ray	1−40	136	136	D	mP, mA, mM	(*112*, *113*)
42	histone H5 (70678)	X-ray, LP	101−185	185	185	D−O[e]	b[e]	(*114*−*117*)
43	hypoxanthine phosphoribosyltransferase (6435814)	X-ray	190−221	221	221	U	x	(*118*)
44	lymphoid enhancer factor-1 (6537322)	NMR	296−380	86	397	D−O	b, u	(*119*)
45	myelin basic protein (126796)	CD	1−169	169	169	D−O	f	(*120*)
46	negative factor, HIV1 (128023)	X-ray, NMR	2−73	various	206	D, D−O	a, f, mP, mF, l	(*121*−*124*)
			149−178			U	a	
47	neural zinc finger factor 1 (1511632)	NMR	1−99	99	1187	D−O	b, j	(*125*)
48	neurofilament H (128127)	other	409−1087	1087	1087	D	mP, mG, p	(*126*, *127*)
49	neuromodulin (548347)	NMR, CD	1−239	239	239	D, D−O	a, f, mP, mR, mF	(*128*)
50	ornithine decarboxylase (7404357)	X-ray	1−36	425	425	U	x	(*129*)
51	osteonectin (129284)	CD	23−68	285	285	D−O	j	(*130*)
52	phenylalanyl-tRNA synthetase (135112)	X-ray	α1−84	350	350	D−O	cT	(*131*, *132*)
53	phosphatidylinositol phosphate kinase (3745771)	X-ray	1−33	416	416	U	x	(*133*)
			307−341					
54	phospholipase c−δ1 (130228)	X-ray, LP	135−205	622	756	U	k[e]	(*134*, *135*)
			446−484			U	w, x	
55	prion (200527)	NMR	23−120	219	241	D−O	j	(*136*)
56	protamines (123705)	CD	1−57	57	57	D−O	b, j	(*137*)
57	prothymosin α (135836)	CD	1−109	109	109	U	a	(*138*)
58	Myc proto-oncogene protein (127619)	CD	1−143	143	439	D−O	a	(*139*)
59	*Pvu*II DNA methyltransferase (6729995)	X-ray	179−216	336	336	D−O[e]	b[e], l[e]	(*140*)
60	replication protein A (1350579)	NMR	109−168	168	616	D	n	(*141*)
61	retinoid X receptor α (4506755)	NMR	130−212	93	462	D−O	b, j	(*142*)
62	30S ribosomal proteins	X-ray, CD	various	various	various	D, D−O	a, cR, r	(*143*)
63	50S ribosomal proteins	X-ray, CD	various	various	various	D, D−O	a, cR, r	(*143*, *144*)
64	signal transduction inhibitor RGS4 (1710149)	X-ray	1−59	205	205	U	a[e]	(*145*−*147*)
			176−205			U	x	
65	southern bean mosaic virus capsid (116795)	X-ray	1−38	260	260	D−O[e]	cG	(*148*, *149*)
66	serine aspartamine repeat protein D (3550594)	CD	569−1123	555	1315	D−O	j, n[e]	(*150*)
67	Sindbis virus capsid (1942972)	X-ray	1−105	264	264	D−O[e]	cG	(*151*, *152*)
68	small heat-shock protein HSP16.5 (2495337)	X-ray	1−33	147	147	U	a[e]	(*153*, *154*)
69	SNAP-25 (134583)	CD, LP, other, X-ray	1−83	various	206	D−O	a	(*155*−*159*)
			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	(*156*−*160*)
71	α-synuclein (586067)	CD	1−140	140	140	D−O	a, g, j	(*161*, *162*)
72	thyroid transcription factor 1 (136462)	LP	1−156	156	372	D−O	a	(*163*)
73a	titin, skeletal (1017427)	other, CD	PEVK domain (2174 aa's)	n.a,	∼33,000	O−D	o	(*164*−*167*)
73b	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[e]	cG	(*168*)
75	topoisomerase I (135989)	CD	1−174	various	765	U	a, l, w	(*169*, *170*)
			632−680			D	b[e], n	
76	topoisomerase II (1633273)	X-ray, LP	1077−1106	various	1429	U	k, mP	(*170*−*173*)
			1203−1429			U	a, k, mP, w	
77	transcription factor ADR1 (113450)	NMR	75−159	various	1323	D−O	b, j	(*174*−*177*)
78	transcription factor c-Fos (4063509)	CD, NMR	216−380	85	380	D−O	a	(*178*)
79	transcription factor c-Jun (135298)	NMR, CD X-ray, other	61−98	37	331	D−O	mP	(*179*)
			257−314	58		D−O	a, b	(*180*−*183*)
80	transcription factor CREB (117435)	CD	1−265	341	341	D−O[e]	a, mP	(*184*)
81	transcription factor $dTAF_{II}230$ (1705691)	NMR	11−77	67	2068	D−O	a	(*185*)

Table 1 (Continued)

	protein (GenPept ID)	detection method[a]	location	chain length	protein length	class[b]	function[c]	refs
82	transcription factor eIF-4G (1170510)	NMR	393−490	98	1395	D−O	a	(*186*)
83	transcription factor GCN4 (121066)	NMR	225−250 251−281	58	281	D−O D−O	b a	(*187*)
84	transcription factor MAX (126776)	CD, other	2−110	109	160	D−O	a, b	(*188*)
85	transcription factor NF-kB p65 (417924)	NMR	428−551	123	551	D−O	a	(*189*)
86	transcription factor p53 (129369)	CD	1−73	73	393	D−O	a	(*190*, *191*)
87	transcription factor VP16 (2827761)	NMR, CD X-ray, LP, other	350−394 403−490	various	490	D−O D−O	a, b a	(*192*−*197*)
88	ubiquinol oxidase (118072; 118071)	X-ray	subunit I: 1−51 subunit I: 553−663	663	663	U	x	(*198*, *199*)
			subunit II: 284−315	315	315	U	x	
89	voltage-gated potassium channel (4557685)	NMR	1−62	495	495	D	k, q	(*200*)
90	xeroderma pigmentosum group A (139817)	LP	1−84 181−265	265	265	D	a[e]	(*201*)

[a] Method of detection: X-ray: X-ray crystallography; NMR: nuclear magnetic resonance spectroscopy; CD: circular dichroism; LP: limited proteolysis; other: other techniques. [b] 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. [c] 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. [d] Structure solved only after removal of disordered region. [e] Hypothesized.

one molecule to bind to differently shaped partners by structural accommodations at the binding interfaces (*4*, *17*, *62*, *204*); (iii) different disordered sequences to fold (perhaps differently) so as to bind to a common binding site (*201*); (iv) the creation of very large interaction surfaces as the disordered protein wraps-up (*22*) or surrounds its partner (*21*); (v) faster rates of association by reducing dependence on orientation factors and by enlarging target sizes (*205*); and (vi) faster rates of dissociation by unzippering mechanisms (*5*).

Chemical modification of side chains requires close association between the target protein and the modifying enzyme. If the side chain being modified is within a structured region, steric factors would typically prevent or slow the association. On the other hand, a side chain within a disordered region facilitates substrate binding because the disordered region can fold onto the modifying enzyme. Several types of chemical modification 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., pseudosubstrates). 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 apparently carry out function without becoming ordered, indicated as (n−q) in Tables 1 and 2, e.g., flexible linkers/spacers (n) between domains. Flexible linkers allow two domains to move relative to each other, and some also act as spacers that regulate the distance between adjacent domains (*141*, *150*). The functional, native state of flexible linkers/spacers is likely to be a random coil, or the polypeptide approximation of the random coil (*206*). Alternatively, the linkers/spacers can have local or transient secondary structure (*207*), but in either case, such regions can carry out function without 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 (*201*, *208*, *209*). Indeed, when digestion occurs in an ordered region of a protein, local unfolding not just surface exposure is necessary for efficient in vitro proteolysis (*210*). The likely explanation is similar to that given above for chemical modification. That is, the polypeptide segment being cleaved must form a specific structure with the associated protease, a process that is substantially enhanced by intrinsic disorder. Given 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

Table 2: Number of Disordered Regions Exhibiting Each Function

code[a]	function	no. 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	16

[a] Code in function column of Table 1.

from *Escherichia coli* show that 10 large- or small-subunit proteins are substantially disordered (*143*). 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 (*211*). 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 (*212*). 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 (*56*), while casein forms micellular aggregates that solubilize hydrophobically aggregated proteins (*50*). 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 (*80*). 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 (*81*).

Another evolutionary niche occupied by disordered regions involves *Staphylococcus aureus* fibronectin binding protein. This protein is completely disordered in solution (*73*). The regions that bind fibronectin, however, undergo disorder-to-order transitions upon forming this high-affinity interaction (*74*). Interestingly, the flexibility of this protein was proposed to account for the lack of an immune response during infection by *S. aureus* (*213*) because highly flexible polypeptides are evidently not immunogenic (M. 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 (*5*).

Disordered Regions Without Known Function. Given the wide range of functions already discovered or proposed for intrinsically disordered proteins, from molecular recognition (*203*) to protection against desiccation (*70*) to detergent action (*50*, *56*), functional annotation may turn out to be especially vexing for proteins with intrinsically disordered regions of sequence. Tables 1 and 2 include six disordered regions claimed to lack essential function (w) as well as 10 proteins with 16 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 six 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 unexpected functions not related to their initially known ones (*214*). 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. As demonstrated here, intrinsic disorder is common and important for protein function. To organize these observations, we propose a structure−function trinity, according to which, *native proteins* can adopt any of three forms: fully folded, partially disordered, or fully disordered (*16*). 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. These factors are under dynamic control by the cell, so native proteins can change back and forth among the different forms.

Function can arise from any one of the three forms. Examples involving fully folded proteins need no further discussion. Functions involving completely disordered proteins include flexible linkers or spacers, which allow the connected domains to move relative to each other to allow simultaneous binding to two or more components with separations that vary over space and time (*96*, *141*, *159*, *170*); entropic bristles, which use excluded volume effects to keep

molecules apart (*126*, *127*); entropic springs, which provide a restoring force resulting from randomization of bond torsion angles that become restricted upon stretching (*164*−*167*); entropic clocks, which provide a timing mechanism (*5*) arising from random searches such as that observed for the ball-and-chain model for closure of voltage-gated ion channels (*200*). Functions involving partially disordered regions include the binding of hydrophobic groups by dynamic detergent-like proteins (*56*) and the multiple binding of ions by static polymorphic ensembles (*215*). Disorder-to-order transitions upon binding, whether starting from partially disordered or from completely disordered forms, are an important use of disorder for function (*4*, *12*).

Functional Repertoires of Ordered and Disordered Proteins. The functions of disordered protein fall into four broad categories: molecular recognition, molecular assembly, protein modification, and entropic chains, each of which can be compared with similar 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; such kinetic delays should be sensitive to low amounts of urea or guanidine. Testing for the frequency of 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 (*202*). For assembly based on ordered subunits, the pairwise interactions would be unlikely to be precise enough to bring about closure when the last subunit associates 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 a loose overall assembly, which could then undergo disorder-to-order transitions to tighten up overall complex.

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 (Iakoucheva et al., submitted), whereas recognition by ordered proteins may be primarily used for catalysis (Lawson et al., manuscript in preparation). 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 overrepresented in PDB and which dominate our thinking about protein structure and function, have evolved to be completely folded to carry out catalysis efficiently, not for the molecular recognition aspects of their functions.

Need for a Disordered Protein Database. Given the numerous examples discussed herein, an annotated database of intrinsically disordered protein is crucial to current structural genomics and proteomics efforts. Tables 1 and 2 provide a start, but a major expansion and a transformation from a flat file into an appropriate data structure are needed. During the expansion phase, we invite researchers to deposit their disorder examples into 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.

ACKNOWLEDGMENT

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.

REFERENCES

1. Sigler, P. B. (1988) *Nature 333*, 210−212.
2. Churchill, M. E., and Travers, A. A. (1991) *Trends Biochem. Sci. 16*, 92−7.
3. Frankel, A. D., and Kim, P. S. (1991) *Cell 65*, 717−719.
4. Dyson, H. J., and Wright, P. E. (2002) *Curr. Opin. Struct. Biol. 12*, 54−60.
5. 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.
6. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., and Dunker, A. K. (1997) *Proc. IEEE Intl. Conf. Neural Networks 1*, 90−95.
7. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) *Gen. Inform. 10*, 30−40.
8. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) *Proteins: Struct., Funct., Gen. 42*, 38−48.
9. Romero, P., Obradovic, Z., and Dunker, A. K. (2000) *Artif. Intelligence Rev. 14*, 447−484.
10. Vucetic, S., Radivojac, P., Obradovic, Z., Brown, C. J., and Dunker, A. K. (2001) *Intl. Joint INNS−IEEE Conf. Neural Networks 4*, 2718−2723.
11. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Guilliot, S., Garner, E., and Dunker, A. K. (1998) *PSB 1998 3*, 437−448.
12. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
13. Romero, P., Obradovic, Z., and Dunker, A. K. (1997) *Gen. Inform. 8*, 110−124.
14. Li, X., Obradovic, Z., Brown, C. J., Garner, E. C., and Dunker, A. K. (2000) *Gen. Inform. 11*, 172−184.
15. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) *Gen. Inform. 11*, 161−171.
16. Dunker, A. K., and Obradovic, Z. (2001) *Nat. Biotech. 19*, 805−806.
17. Karush, F. (1950) *J. Am. Chem. Soc. 72*, 2705−2713.
18. Yazawa, M., Matsuzawa, F., and Yagi, K. (1990) *J. Biochem. (Tokyo) 107*, 287−291.
19. Chakravarthy, B., Morley, P., and Whitfield, J. (1999) *Trends. Neurosci. 22*, 12−16.
20. Minton, A. P. (2000) *Curr. Opin. Struct. Biol. 10*, 34−39.
21. Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) *Science 257*, 1251−1255.
22. Choo, Y., and Schwabe, J. W. (1998) *Nat. Struct. Biol. 5*, 253−255.

23. Morar, A. S., Wang, X., and Pielak, G. J. (2001) *Biochemistry 40*, 281−285.
24. Flaugh, S. L., and Lumb, K. J. (2001) *Biomacromolecules 2*, 538−540.
25. Zurdo, J., Sanz, J. M., Gonzalez, C., Rico, M., and Ballesta, J. P. (1997) *Biochemistry 36*, 9625−9635.
26. 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. 55*, 102−107.
27. Zurdo, J., Gonzalez, C., Sanz, J. M., Rico, M., Remacha, M., and Ballesta, J. P. (2000) *Biochemistry 39*, 8935−8943.
28. Nusspaumer, G., Remacha, M., and Ballesta, J. P. (2000) *EMBO J. 19*, 6075−6084.
29. Ballesta, J. P. (personal communication).
30. 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.
31. 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.
32. van Breukelen, B., Kanellopoulos, P. N., Tucker, P. A., and van der Vliet, P. C. (2000) *J. Biol. Chem. 275*, 40897−40903.
33. Johansson, J., Gudmundsson, G. H., Rottenberg, M. E., Berndt, K. D., and Agerberth, B. (1998) *J. Biol. Chem. 273*, 3718−3724.
34. Mogridge, J., Legault, P., Li, J., Van Oene, M. D., Kay, L. E., and Greenblatt, J. (1998) *Mol. Cell 1*, 265−275.
35. Legault, P., Li, L., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998) *Cell 93*, 289−299.
36. Stellwagen, E., Rysavy, R., and Babul, G. (1972) *J. Biol. Chem. 247*, 8074−8077.
37. Strauss, P. R., and Holt, C. M. (1998) *J. Biol. Chem. 273*, 14435−14441.
38. 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.
39. Mol, C. D., Izumi, T., Mitra, S., and Tainer, J. A. (2000) *Nature 403*, 451−456.
40. Chang, J. F., Phillips, K., Lundback, T., Gstaiger, M., Ladbury, J. E., and Luisi, B. (1999) *J. Mol. Biol. 288*, 941−952.
41. Yamamoto, K., Ichijo, H., and Korsmeyer, S. J. (1999) *Mol. Cell. Biol. 19*, 8469−8478.
42. 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.
43. Chang, B. S., Minn, A. J., Muchmore, S. W., Fesik, S. W., and Thompson, C. B. (1997) *EMBO J. 16*, 968−977.
44. 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. U.S.A. 95*, 554−559.
45. 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.
46. Rodi, D. J., and Makowski, L. (1999) *Pacific Symp. Biocomputing 4*, 532−541.
47. Fang, G., Chang, B. S., Kim, C. N., Perkins, C., Thompson, C. B., and Bhalla, K. N. (1998) *Cancer Res. 58*, 3202−3208.
48. 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.
49. Yang, S. A., and Klee, C. B. (2000) *Biochemistry 39*, 16147−16154.
50. Bhattacharyya, J., and Das, K. P. (1999) *J. Biol. Chem. 274*, 15505−15509.
51. 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.
52. 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.
53. Ostedgaard, L. S., Baldursson, O., Vermeer, D. W., Welsh, M. J., and Robertson, A. D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5657−5662.
54. 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.
55. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994) *Structure 2*, 545−558.
56. Bailey, R. W., Dunker, A. K., Brown, C. J., Garner, E. C., and Griswold, M. D. (2001) *Biochemistry 40*, 11828−11840.
57. Brandstetter, H., Bauer, M., Huber, R., Lollar, P., and Bode, W. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9796−9800.
58. 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.
59. 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.
60. 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.
61. 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.
62. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504−11509.
63. Adkins, J. N., and Lumb, K. J. (2002) *Proteins 46*, 1−7.
64. Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J., and Pavletich, N. P. (1996) *Nature 382*, 325−331.
65. 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.
66. Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell 85*, 125−136.
67. 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.
68. Laurberg, M., Kristensen, O., Martemyanov, K., Gudkov, A. T., Nagaev, I., Hughes, D., and Liljas, A. (2000) *J. Mol. Biol. 303*, 593−603.
69. Martemyanov, K. A., and Gudkov, A. T. (2000) *J. Biol. Chem. 275*, 35820−35824.
70. Eom, J., Baker, W., Kintanar, A., and Wurtele, E. (1996) *Plant Sci. 115*, 17−24.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.
76. Vonderviszt, F., Kanto, S., Aizawa, S., and Namba, K. (1989) *J. Mol. Biol. 209*, 127−133.
77. Aizawa, S. I., Vonderviszt, F., Ishima, R., and Akasaka, K. (1990) *J. Mol. Biol. 211*, 673−677.
78. Mimori-Kiyosue, Y., Yamashita, I., Fujiyoshi, Y., Yamaguchi, S., and Namba, K. (1998) *J. Mol. Biol. 284*, 521−530.
79. Mimori-Kiyosue, Y., Vonderviszt, F., and Namba, K. (1997) *J. Mol. Biol. 270(2)*, 222−237.
80. Samatey, F. A., Imada, K., Nagashima, S., Vonderviszt, F., Kumasaka, T., Yamamoto, M., and Namba, K. (2001) *Nature 410*, 331−337.
81. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T., and Dahlquist, F. W. (1997) *Nat. Struct. Biol. 4*, 285−291.

82. Daughdrill, G. W., Hanely, L. J., and Dahlquist, F. W. (1998) *Biochemistry 37*, 1076−1082.
83. Fletcher, C. M., McGuire, A. M., Gingras, A. C., Li, H., Matsuo, H., Sonenberg, N., and Wagner, G. (1998) *Biochemistry 37*, 9−15.
84. Fletcher, C. M., and Wagner, G. (1998) *Protein Sci. 7*, 1639−1642.
85. 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.
86. Eigenbrot, C., and Gerber, N. (1997) *Nat. Struct. Biol. 4*, 435−438.
87. Eketjall, S., Fainzilber, M., Murray-Rust, J., and Ibanez, C. F. (1999) *EMBO J. 18*, 5901−5910.
88. Ibanez, C. F. (personal communication).
89. Baskakov, I. V., Kumar, R., Srinivasan, G., Ji, Y. S., Bolen, D. W., and Thompson, E. B. (1999) *J. Biol. Chem. 274*, 10693−10696.
90. Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., and Takusagawa, F. (2000) *J. Mol. Biol. 298*, 149−162.
91. Takusagawa, F. (personal communication).
92. Logan, D. T., Mazauric, M. H., Kern, D., and Moras, D. (1995) *EMBO J. 14*, 4156−4167.
93. Arnez, J. G., Dock-Bregeon, A. C., and Moras, D. (1999) *J. Mol. Biol. 286*, 1449−1459.
94. Clackson, T., Ultsch, M. H., Wells, J. A., and de Vos, A. M. (1998) *J. Mol. Biol. 277*, 1111−1128.
95. Clackson, T., and Wells, J. A. (1995) *Science 267*, 383−386.
96. Holliger, P., Riechmann, L., and Williams, R. L. (1999) *J. Mol. Biol. 288*, 649−657.
97. Holliger, P., and Riechmann, L. (1997) *Structure 5*, 265−275.
98. Nilsson, N., Malmborg, A. C., and Borrebaeck, C. A. (2000) *J. Virol. 74*, 4229−4235.
99. Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G., and Sprang, S. R. (1994) *Science 265*, 1405−1412.
100. Coleman, D. E., and Sprang, S. R. (1999) *J. Biol. Chem. 274*, 16669−16672.
101. 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.
102. Louie, G. V., Yang, W., Bowman, M. E., and Choe, S. (1997) *Mol. Cell 1*, 67−78.
103. Higashiyama, S., Lau, K., Besner, G. E., Abraham, J. A., and Klagsbrun, M. (1992) *J. Biol. Chem. 267*, 6205−6212.
104. 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.
105. 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.
106. Louie, D. F., Gloor, K. K., Galasinski, S. C., Resing, K. A., and Ahn, N. G. (2000) *Protein Sci. 9*, 170−179.
107. 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.
108. 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.
109. 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.
110. Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G., and Thanos, D. (2001) *Science 293*, 1133−1136.
111. Cary, P. D., Crane-Robinson, C., Bradbury, E. M., and Dixon, G. H. (1981) *Eur. J. Biochem. 119*, 545−551.
112. 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.
113. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001) *Science 292*, 110−113.
114. Aviles, F. J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C., and Bradbury, E. M. (1978) *Eur. J. Biochem. 88*, 363−371.
115. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) *Nature 362*, 219−223.
116. 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.
117. Crane-Robinson, C. (personal communication).
118. Focia, P. J., Craig, S. P., 3rd, Nieves-Alicea, R., Fletterick, R. J., and Eakin, A. E. (1998) *Biochemistry 37*, 15066−15075.
119. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) *Nature 376*, 791−795.
120. Polverini, E., Fasano, A., Zito, F., Riccio, P., and Cavatorta, P. (1999) *Eur. Biophys. J. 28*, 351−355.
121. Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996) *Cell 85*, 931−942.
122. Arold, S., Franken, P., Strub, M. P., Hoh, F., Benichou, S., Benarous, R., and Dumas, C. (1997) *Structure 5*, 1361−1372.
123. Geyer, M., Munte, C. E., Schorr, J., Kellner, R., and Kalbitzer, H. R. (1999) *J. Mol. Biol. 289*, 123−138.
124. Arold, S. T., and Baur, A. S. (2001) *Trends Biochem. Sci. 26*, 356−363.
125. Berkovits, H. J., and Berg, J. M. (1999) *Biochemistry 38*, 16826−168230.
126. Brown, H. G., and Hoh, J. H. (1997) *Biochemistry 36*, 15035−15040.
127. Hoh, J. H. (1998) *Proteins 32*, 223−228.
128. Zhang, M., Vogel, H. J., and Zwiers, H. (1994) *Biochem. Cell Biol. 72*, 109−116.
129. Grishin, N. V., Osterman, A. L., Brooks, H. B., Phillips, M. A., and Goldsmith, E. J. (1999) *Biochemistry 38*, 15174−15184.
130. Engel, J., Taylor, W., Paulsson, M., Sage, H., and Hogan, B. (1987) *Biochemistry 26*, 6958−6965.
131. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G. (1995) *Nat. Struct. Biol. 2*, 537−547.
132. Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997) *Structure 5*, 59−68.
133. Rao, V. D., Misra, S., Boronenkov, I. V., Anderson, R. A., and Hurley, J. H. (1998) *Cell 94*, 829−839.
134. Grobler, J. A., Essen, L. O., Williams, R. L., and Hurley, J. H. (1996) *Nat. Struct. Biol. 3*, 788−795.
135. Essen, L. O., Perisic, O., Cheung, R., Katan, M., and Williams, R. L. (1996) *Nature 380*, 595−602.
136. Lopez Garcia, F., Zhan, R., Riek, R., and Wulthrich, K. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 8334−8339.
137. Gatewood, J. M., Schroth, G. P., Schmid, C. W., and Bradbury, E. M. (1990) *J. Biol. Chem. 265*, 20667−20672.
138. 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.
139. McEwan, I. J., Dahlman-Wright, K., Ford, J., and Wright, A. P. (1996) *Biochemistry 35*, 9584−93.
140. Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997) *Nucleic Acids Res. 25*, 2702−2715.
141. 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.
142. 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.
143. Venyaminov, S., Gudkov, A., Gogia, Z., and Tumanova, L. (1981) Biological Research Center, Institute of Protein Research, Pushchino: AS USSR.
144. Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B., and Steitz, T. A. (1999) *Nature 400*, 841−847.
145. Tesmer, J. J., Berman, D. M., Gilman, A. G., and Sprang, S. R. (1997) *Cell 89*, 251−261.
146. 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.
147. Chatterjee, T. K., and Fisher, R. A. (2000) *J. Biol. Chem. 275*, 24013−24021.
148. Silva, A. M., and Rossmann, M. G. (1985) *Acta Crystallogr. B41*, 147−157.

149. Rossmann, M. G., Chandrasekaran, R., Abad-Zapatero, C., Erickson, J. W., and Arnott, S. (1983) *J. Mol. Biol. 166*, 73−80.
150. Josefsson, E., O'Connell, D., Foster, T. J., Durussel, I., and Cox, J. A. (1998) *J. Biol. Chem. 273*, 31145−31152.
151. Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G., and Wengler, G. (1991) *Nature 354*, 37−43.
152. 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.
153. Kim, K. K., Kim, R., and Kim, S. H. (1998) *Nature 394*, 595−599.
154. Kim, R., Kim, K. K., Yokota, H., and Kim, S. H. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9129−9133.
155. Fasshauer, D., Bruns, D., Shen, B., Jahn, R., and Brunger, A. T. (1997) *J. Biol. Chem. 272*, 4582−4590.
156. Fasshauer, D., Otto, H., Eliason, W. K., Jahn, R., and Brunger, A. T. (1997) *J. Biol. Chem. 272*, 28036−28041.
157. Fasshauer, D., Eliason, W. K., Brunger, A. T., and Jahn, R. (1998) *Biochemistry 37*, 10354−10362.
158. Canaves, J. M., and Montal, M. (1998) *J. Biol. Chem. 273*, 34214−32221.
159. Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) *Nature 395*, 347−353.
160. Hazzard, J., Sudhof, T. C., and Rizo, J. (1999) *J. Biomol. NMR 14*, 203−207.
161. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T., Jr. (1996) *Biochemistry 35*, 13709−13715.
162. Nielsen, M. S., Vorum, H., Lindersson, E., and Jensen, P. H. (2001) *J. Biol. Chem. 18*, 18.
163. 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.
164. Trombitas, K., Greaser, M., Labeit, S., Jin, J. P., Kellermayer, M., Helmes, M., and Granzier, H. (1998) *J. Cell Biol. 140*, 853−859.
165. Labeit, S., and Kolmerer, B. (1995) *Science 270*, 293−296.
166. Kellermayer, M. S. Z., Smith, S. B., Granzier, H. L., and Bustamante, C. (1997) *Science 276*, 1112−1116.
167. Helmes, M., Trombitas, K., Centner, T., Kellermayer, M., Labeit, S., Linke, W. A., and Granzier, H. (1999) *Circ. Res. 84*, 1339−1352.
168. Hopper, P., Harrison, S. C., and Sauer, R. T. (1984) *J. Mol. Biol. 177*, 701−713.
169. Stewart, L., Ireton, G. C., Parker, L. H., Madden, K. R., and Champoux, J. J. (1996) *J. Biol. Chem. 271*, 7593−7601.
170. Shaiu, W. L., Hu, T., and Hsieh, T. S. (1999) *Pacific Symp. Biocomput. 4*, 578−589.
171. Berger, J. M., Gamblin, S. J., Harrison, S. C., and Wang, J. C. (1996) *Nature 379*, 225−232.
172. Caron, P. R., Watt, P., and Wang, J. C. (1994) *Mol. Cell Biol. 14*, 3197−3207.
173. Berger, J. M. (personal communication).
174. Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., and Klevit, R. E. (1988) *Science 241*, 1489−1492.
175. Klevit, R. E., Herriott, J. R., and Horvath, S. J. (1990) *Proteins 7*, 215−226.
176. Hyre, D. E., and Klevit, R. E. (1998) *J. Mol. Biol. 279*, 929−943.
177. Bowers, P. M., Schaufler, L. E., and Klevit, R. E. (1999) *Nat. Struct. Biol. 6*, 478−485.
178. Campbell, K. M., Terrell, A. R., Laybourn, P. J., and Lumb, K. J. (2000) *Biochemistry 39*, 2708−2713.
179. John, M., Briand, J. P., and Schnarr, M. (1996) *Pept. Res. 9*, 71−78.
180. 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.
181. Glover, J. N., and Harrison, S. C. (1995) *Nature 373*, 257−61.
182. Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S., and King, G. F. (1996) *J. Biol. Chem. 271*, 13663−13667.
183. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature 392*, 42−48.
184. Richards, J. P., Bachinger, H. P., Goodman, R. H., and Brennan, R. G. (1996) *J. Biol. Chem. 271*, 13716−13723.
185. 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.
186. Hershey, P. E., McWhirter, S. M., Gross, J. D., Wagner, G., Alber, T., and Sachs, A. B. (1999) *J. Biol. Chem. 274*, 21297−21304.
187. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) *Nature 347*, 575−578.
188. Horiuchi, M., Kurihara, Y., Katahira, M., Maeda, T., Saito, T., and Uesugi, S. (1997) *J. Biochem. (Tokyo) 122*, 711−716.
189. 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.
190. Chang, J., Kim, D. H., Lee, S. W., Choi, K. Y., and Sung, Y. C. (1995) *J. Biol. Chem. 270*, 25014−25019.
191. Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., and Pavletich, N. P. (1996) *Science 274*, 948−953.
192. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996) *J. Biol. Chem. 271*, 4827−4837.
193. Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J., and Verdine, G. L. (1997) *Science 277*, 1310−1313.
194. Liu, Y., Gong, W., Huang, C. C., Herr, W., and Cheng, X. (1999) *Genes Dev. 13*, 1692−1703.
195. Hayes, S., and O'Hare, P. (1993) *J. Virol. 67*, 852−862.
196. O'Hare, P., and Williams, G. (1992) *Biochemistry 31*, 4150−4156.
197. Donaldson, L., and Capone, J. P. (1992) *J. Biol. Chem. 267*, 1411−1414.
198. Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E., and Saraste, M. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 11955−11959.
199. 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.
200. 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.
201. 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) *Protein Sci. 10*, 560−571.
202. Namba, K. (2001) *Genes Cells 6*, 1−12.
203. Schulz, G. E. (1979) In *Molecular Mechanism of Biological Recognition* (Balaban, M., Ed.) pp 79−94, Elsevier/North-Holland Biomedical Press, New York.
204. Meador, W. E., Means, A. R., and Quiocho, F. A. (1993) *Science 262*, 1718−1721.
205. Pontius, B. W. (1993) *Trends Biochem. Sci. 18*, 181−186.
206. Pappu, R. V., Srinivasan, R., and Rose, G. D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 12565−12570.
207. Gutierrez-Cruz, G., Van Heerden, A. H., and Wang, K. (2001) *J. Biol. Chem. 276*, 7442−7449.
208. Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) *J. Mol. Biol. 266*, 223−230.
209. Fontana, A., de Laureto, P. P., de Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Des. 2*, R17−R26.
210. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) *Protein Sci. 3*, 757−768.
211. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905−920.
212. Treiber, D. K., and Williamson, J. R. (2001) *Curr. Opin. Struct. Biol. 11*, 309−314.
213. House-Pompeo, K., Xu, Y., Joh, D., Speziale, P., and Hook, M. (1996) *J. Biol. Chem. 271*, 1379−1384.
214. Jeffery, C. J. (1999) *Trends Biochem. Sci. 24*, 8−11.
215. 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.

BI012159+