Intrinsic Disorder and Functional Proteomics

Predrag Radivojac,* Lilia M. Iakoucheva,† Christopher J. Oldfield,* Zoran Obradovic,‡ Vladimir N. Uversky,§¶ and A. Keith Dunker§

*School of Informatics, Indiana University, Bloomington, Indiana; †Laboratory of Statistical Genetics, The Rockefeller University, New York, New York; ‡Center for Information Science and Technology, Temple University, Philadelphia, Pennsylvania; §Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, Indiana; and ¶Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russia

ABSTRACT The recent advances in the prediction of intrinsically disordered proteins and the use of protein disorder prediction in the fields of molecular biology and bioinformatics are reviewed here, especially with regard to protein function. First, a close look is taken at intrinsically disordered proteins and then at the methods used for their experimental characterization. Next, the major statistical properties of disordered regions are summarized, and prediction models developed thus far are described, including their numerous applications in functional proteomics. The future of the prediction of protein disorder and the future uses of such predictions in functional proteomics comprise the last section of this article.

INTRODUCTION

Until the early 1990s, a widely, almost exclusively accepted concept of protein function was the well-known protein sequence → structure → function paradigm. According to this concept, a protein can achieve its biological function only upon folding into a unique, structured state, which represents a kinetically accessible and an energetically favorable conformation (usually the global energy minimum for the whole protein) determined by its amino acid sequence. This specific conformation has been referred to as the native state of the protein. Ample experimental evidence has been accumulated since the 1890s to support this view. Some representative supportive examples include theoretical models postulated by Pauling (1), Fischer's lock-and-key hypothesis (2), the first crystal structures of globular proteins (3,4) and of enzymes (5), and the studies that supported the refoldability proteins into their functional states (6,7), in which a protein was shown to regain its function if the necessary environmental conditions were restored after the initial perturbation. The state in which a protein loses its function, known as the denatured state, has been associated with the loss of the specific three-dimensional structure (8,9), which can lead to either monomeric conformational ensembles (both compact and noncompact) under some denaturing conditions or to insoluble aggregates under others.

Occasional counterexamples to the general view presented above have been observed over many years, but these were mostly ignored and largely overshadowed by the success of the studies of proteins with specific three-dimensional structures, or what we call ordered proteins. However, recent discoveries of intrinsically disordered proteins (IDPs) (10) (known also as natively disordered (11), natively unfolded (12), and intrinsically unstructured (13) proteins) have significantly broadened the view of the scientific community and increased the number of groups systematically studying these intriguing members of the protein world. Bioinformatics has been very helpful in transforming the disparate collection of counterexample proteins into a de facto subfield of protein science.

What is an intrinsically disordered protein?

In an ordered protein region, the Ramachandran angles and backbone atoms of each residue undergo nonisotropic small-amplitude motions relative to their local neighborhood and are characterized by the equilibrium positions defined by the time-averaged values. The atom fluctuations are caused by two factors, random thermal motion and small cooperative conformational changes of the local sequence neighborhood, and these fluctuations are known to be influenced by local residue packing (14). In contrast to ordered protein regions, ID regions are not characterized by the atom equilibrium positions and dihedral angle equilibrium values around which the residue spends most of the time. ID regions exist instead as dynamic ensembles in which atom positions and backbone Ramachandran angles vary significantly over time with no specific equilibrium values. The conformational changes of ID regions are typically noncooperative and random. Thus, the view of disorder as dynamic ensembles does not exclude the temporary presence of local secondary structure that fluctuates in absence of stabilizing forces. Associating IDPs and ID regions with structural ensembles remains a qualitative description because the degree of structural change and the number of distinct structures in the ensemble are likely to vary over a wide range for different IDPs.

Slightly $<\frac{1}{3}$ of the crystal structures in the Protein Data Bank (PDB) are completely devoid of disorder (15). Also, ID can be manifested in a variety of contexts, affecting various levels of protein structure: functional disordered segments can be as short as only a few amino acid residues, or they can

Submitted July 26, 2006, and accepted for publication November 15, 2006.

Address reprint requests to Prof. A. Keith Dunker, Tel.: 317-278-9650; E-mail: kedunker@iupui.edu; or Dr. Vladimir N. Uversky, Tel.: 317-278-9194; E-mail: vuversky@iupui.edu.

0006-3495/07/03/1439/18 $2.00

doi: 10.1529/biophysj.106.094045

occupy rather long loop regions and/or protein ends. Proteins can be partially or even wholly disordered, even large ones (16), so we define an IDP as a protein that contains at least one disordered region. However, in practice, very short disordered regions have typically been ignored, since these regions were not determined with high confidence and were not associated with particular functions. Hence, our definition of an IDP will be somewhat loose due to the experimental problems in characterizing disorder with high precision. Our current interest focuses on those regions that are sufficiently long to be readily characterized, and especially on those that have been associated with function by experiment.

Experimental characterization of disorder

The disorder in IDPs has been detected by several physicochemical methods elaborated to characterize protein self-organization. The list includes but is not limited to x-ray crystallography (17), NMR spectroscopy (11,18–21), near-ultraviolet circular dichroism (CD) (22), far-ultraviolet CD (23–26), ORD (23,26), Fourier transform infrared (26), Raman spectroscopy and Raman optical activity (27), different fluorescence techniques (28,29), numerous hydrodynamic techniques (including gel-filtration, viscometry, small angle x-ray scattering (SAXS), small angle neutron scattering (SANS), sedimentation, and dynamic and static light scattering) (28,29), rate of proteolytic degradation (30–34), aberrant mobility in SDS-gel electrophoresis (35,36), low conformational stability (28,37–40), H/D exchange (29), immunochemical methods (41,42), interaction with molecular chaperones (28), electron microscopy or atomic force microscopy (28,29), and the charge state analysis of electrospray ionization mass-spectrometry (43). (For more detailed reviews on methods used to detect intrinsic disorder, see (11,19,29,44).)

Functions of intrinsically disordered regions

Although it can be argued that IDPs occupy a continuum of structural forms, there are two major views on categorization of the form of IDPs. Dunker and Obradovic (45) proposed that functional intrinsically disordered regions may exist in two different structural forms: molten globule-like (collapsed) and random coil-like (extended) forms, whereas Uversky suggested existence of another extended form, the pre-molten globule (44), which appears to be distinct category between fully extended and molten-globular conformations and which is distinguishable by the presence of unstable secondary structure. Together with the ordered form, these ID categories form the basis of the protein trinity (45) or the protein-quartet (44) hypothesis. It follows that protein function is associated with any of the three (or four) distinct forms or with transitions between them, where conformational changes associated with function may also be brought about by alterations in environmental or cellular conditions. In short, IDPs and ID regions are typically involved in regulation, signaling and control pathways (16,46,47) and thus complement the functional repertoire of ordered regions, which in our view have evolved mainly to carry out efficient catalysis. Of course, enzymes such as kinases and phosphatases also participate in regulation, signaling, and control pathways, but for disordered proteins these activities are the direct result of their actions, whereas for enzymes these activities occur as a result of the changes brought about by the catalytic events. Indeed, it is interesting that catalytic events associated with regulation or signaling often occur in IDPs or ID regions (48) as discussed below.

Using literature searches, 90 proteins with functionally annotated IDPs and ID regions were found (48). These IDPs were shown to be involved in 28 specific functions, which were organized into four functional classes: 1), molecular recognition; 2), molecular assembly; 3), protein modification; and 4), entropic chain activities (49). The first three functions result from interactions between disordered regions and their partners. Molecular recognition is primarily represented in signaling. Protein modifications are another way of increasing the functional diversity of the proteome, in which protein modification sites can either be directly recognized by other molecules or can introduce allosteric changes that trigger a series of downstream effects. Molecular assembly is a functional class represented by proteins involved in assembly of viruses, ribosomes and the cytoskeleton. In these three functional categories, disordered regions typically undergo transitions from unfolded to folded forms. On the other hand, the functions of the fourth category, namely entropic chain activities, arise directly from the unfolded state. Common representatives of this category are linkers, spacers, bristles, springs and clocks, but it is expected that other functions depending on the unfolded state will be found as well.

The involvement of IDPs and ID regions in molecular recognition probably results from a number of capabilities enabled by this protein form (16,47) including the following: 1), decoupling of specificity and affinity due to the free energy penalty paid to fold the disordered state; 2), binding diversity in which one region folds differently to recognize differently shaped partners by different structural accommodations at the various binding interfaces; 3), binding commonality in which multiple, distinct sequences fold differently yet each recognize a common binding surface; 4), the formation of large interaction surfaces as the disordered region wraps-up or surrounds its partner; 5), faster rates of association by reducing dependence on orientation factors and by enlarging target sizes; and 6), faster rates of dissociation by unzippering mechanisms.

Computational approaches to predicting intrinsically disordered regions

In addition to laboratory experiments, a key argument about the existence and distinctiveness of ID regions came from computational analysis. Statistical comparisons of amino acid compositions and sequence complexity indicated that disordered and ordered regions are different to a significant degree. These sequence biases were then exploited to predict disordered regions with high accuracy and to estimate the commonness of IDPs and ID regions in the three kingdoms of life. Finally, in the latest wave, it has been shown that the functional repertoire, including the mechanistic properties of molecular binding show specific characteristics for disordered regions that are considerably different from the characteristics of ordered regions. We begin this section by a discussion of the public repositories of IDPs and then address the various computational approaches to ID prediction.

Database of intrinsically disordered proteins

The first public resource containing disordered protein regions was developed by Sim et al. (50). However, the ProDDO database was not curated, its contents were limited to the PDB entries only, and it did not provide information about type of disorder nor the function of disordered regions. These limitations are being overcome by DisProt, which is a database containing experimentally characterized IDPs and ID regions and their biological functions (51,52). The database contains numerous examples of IDPs characterized by several experimental techniques and includes functional information for many of the IDPs and regions. Therefore, DisProt links structure and function information for IDPs and ID regions in a systematic way. This database was developed to facilitate IDP research by collecting and organizing knowledge regarding the experimental characterization and the functional associations of IDPs. In addition to being a unique source of biological information, DisProt opens the door for bioinformatics studies. In its first public release of February 2004, DisProt contained 154 proteins (190 disordered regions), whereas in August 2006 the database contained 460 proteins (1103 disordered regions). The database can be accessed at http://www.disprot.org.

Sequence biases of disordered protein regions

Ordered and disordered regions were shown to possess distinct sequence biases. Based on the analysis of 150 IDPs and ID regions, amino acid residues were grouped into order promoting, disorder promoting and neutral (10). To illustrate this finding, Fig. 1 presents relative amino acid compositions of ID regions available in the DisProt database (51). The amino acid compositions were compared using a profiling approach (10). This figure compares the compositions of the 460 proteins currently available in the database with the compositions of the 152 proteins present in DisProt in July 2002, with the amino acids arranged in order for the larger database. Based on the new amino acid compositions of IDPs and ID regions, and using a fractional difference of 0.1 to separate the amino acid classes, the order-promoting residues are C, W, Y, I, F, V, L, H, T, and N, the disorder-promoting residues are D, M, K, R, S, Q, P, and E, and the neutral residues are A and G. Note that H, T, N, and D are borderline by the 0.1 fractional difference criterion, which is rather arbitrary, and so these residues could also be considered neutral.

Disordered regions of different length show statistical differences (53), as suggested in an earlier study (54). In addition, more rigid and less rigid regions of structured proteins also show compositional differences. Pairwise comparisons among four structural classes, namely low B-factor ordered regions, high B-factor ordered regions, short disordered regions, and long disordered regions, show each class to have a different amino acid composition from the other three, with short disordered regions and high B-factor regions having the most similar compositions. Furthermore, the compositions of these two groups were both closer to the composition of long disordered regions than to that of more rigid ordered regions (53). Particularly interesting was the analysis of charge, which showed that the short disordered and high B-factor regions were more negatively charged, whereas long disordered regions were either positively or negatively charged, but on average nearly neutral.

In addition to the first-order statistics, more recent studies also addressed higher-order patterns. Lise and Jones (55) investigated sequence patterns that are statistically overrepresented in disordered regions. They examined the patterns in amino acid sequence space and also analyzed the space of various physicochemical properties. Their analysis confirmed

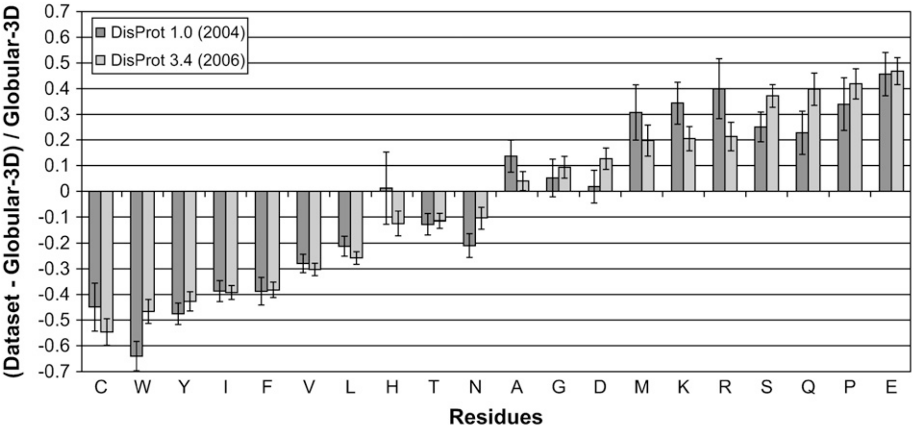

FIGURE 1 Amino-acid composition, relative to the set of globular proteins Globular-3D, of intrinsically disordered regions 10 residues or longer from the DisProt database. Dark gray indicates DisProt 1.0 (152 proteins), whereas light gray indicates DisProt 3.4 (460 proteins). Amino acid compositions were calculated per disordered regions and then averaged. The arrangement of the amino acids is by peak height for the DisProt 3.4 release. Confidence intervals were estimated using per-protein bootstrapping with 10,000 iterations.

that disordered sequences characterized to date were enriched in proline and contained both positively and negatively charged patterns.

Prediction of disordered protein regions

The first predictor of intrinsically disordered regions was constructed in 1997 by Romero et al. (54), based only on 67 disordered regions (1,340 residues) and a number of ordered regions (16,543 residues) manually extracted from PDB (56). Based on these data, a two-layer feed-forward neural-network was constructed that achieved a surprising accuracy of ~70%. This work was significant because it for the first time indicated that the lack of fixed protein three-dimensional structure is predictable from the amino acid sequence alone. In addition, it not only provided the first clues into the compositional differences between ordered and disordered protein regions, but it also indicated that disordered regions of different lengths (short, medium and long) are compositionally different from each other. The predictive model was later extended into the VLXT predictor (57), a combination of an interior disordered region predictor (VL1) and a separate predictor trained only at protein termini, XT (58). The VLXT predictor was later named the Predictor Of Natural Disordered Regions VLXT (PONDR VLXT).

Interestingly, the existence of a significant difference in the compositional complexity between the globular and nonglobular regions of protein sequences was recognized more than a decade ago (59), several years before the first order/disorder predictor. The sequences corresponding to the crystal structures in PDB were shown to differ only slightly from randomly shuffled sequences in the distribution of statistical properties such as local compositional complexity. On the other hand, ~¼ of the residues in the SWISS-PROT database was shown to occur in segments of nonrandomly low complexity (59,60). Several classes of proteins with known, experimentally defined nonglobular regions have been analyzed, including coiled-coils, elastins, histones, nonhistone proteins, mucins, proteoglycan core proteins and proteins containing long single solvent-exposed α-helices. Based on the results of these analyses it was concluded that globular and nonglobular regions of these sequences can be effectively discriminated using the difference in their compositional complexity (60). All this led to the development of a computational method, the SEG algorithm, which aimed to divide sequences into contrasting segments of low- and high-complexity (60–63).

Subsequent studies indicate that sequence regions with low complexity nearly always correspond to nonfolding segments, or to proteins and regions that form fibrous or extended structures (57), whereas IDPs or ID regions do not always possess low sequence complexity (57,64). Overall, both SEG analysis for complexity and order-disorder prediction are useful and complementary in the analysis of protein sequences. These two approaches have been recently combined into a single plot, which provides an important new method for characterizing IDPs and ID regions (65).

In 2000, Uversky et al. (26) noticed that proteins disordered over their entire lengths can be separated from ordered proteins by considering their average net charge and hydropathy. A separation line in the charge-hydropathy phase space was determined, indicating that a protein is more likely to be entirely disordered than ordered if $H > (R + 1.151)/2.785$, where H is its mean hydropathy (66) and R is its mean absolute net charge over the entire sequence (R was calculated as the absolute value of the difference between the number of lysines and arginines and the number of aspartic and glutamic acids, normalized by the sequence length). In its original form, the charge-hydropathy plot (CH-plot) did not have the sensitivity to predict disordered regions on a per residue basis, but recently charge-hydropathy analysis has been modified and extended to identify local ID regions using a sliding window approach (67).

Several of the predictors developed in the early 2000s used different definitions of disordered regions. For example, there are three versions on the DisEMBL server (68), trained on three proposed types of disorder: 1), loops/coil, i.e., structured regions missing regular secondary structure of helix and strand; 2), hot-loops, i.e., structured regions other than helix or strand, but having high C_α B-factors; and 3), remark465, i.e., regions with missing electron density from PDB. The predictor of NORS regions by Liu et al. (69,70) used a similar definition to that of loops/coil type to predict regions devoid of secondary structure. Indeed, NORS stands for NOn-Regular secondary Structure. Throughout this review, all regions that have fixed three-dimensional structure are considered to be ordered regions, regardless of their B-factor values or secondary structure assignments.

In time, more sophisticated methods based on various statistical and machine learning techniques have emerged (71,72). It is worth mentioning that in addition to the method by Uversky et al. (26), some other approaches also exploited the ideas of reduced sets of amino acids (73) or physicochemical properties, e.g., hydropathy scale only (74) or expected number of contacts per residue (75), to predict disordered regions without significant loss of accuracy. The development of different ID predictors was dramatically stimulated by including disorder prediction as a separate category in the CASP experiments (76,77). As a result, more than 20 different ID predictors have been developed, with many of them being recently reviewed (78). The list of these predictors includes but is not limited to: several PONDR models (15,53,79–81); DISOPRED models (82–84); GlobPlot (85); DisEMBL (68); NORS (69; 70); IUPred (86; 87); FoldIndex (67); RONN (88); PreLink (89); DISpro (90); SPRITZ (91), Wiggle (92), etc.

The predictors developed so far have been based on a spectrum of computational approaches relying on amino acid compositions, derived properties (such as secondary structure prediction) or simple physicochemical properties (such as charge) of the local sequence neighborhood. Almost all of the above-mentioned predictors are available as web servers. Links to these servers, when available, can be found in DisProt (51,52). The relevant information regarding these models is summarized in Table 1. We selected only those models that were scientifically novel and/or published and that are readily accessible. Various other predictors may exist at other private or commercial web sites.

Protein-protein interactions: surface area, interface area, and binding-induced folding

The structures of protein complexes formed by binding-induced folding differ from structures of complexes formed by the association of structured monomers. Disorder in the unbound state leads to bound-state structures with larger normalized monomer surface areas and with larger normalized interface areas compared to the same features for complexes assembled from structured monomers. Indeed, if the normalized monomer surface area is plotted against the normalizing interface area, a simple straight line separates complexes arising from structured proteins from those arising from the binding-induced folding of intrinsically disordered proteins (93).

Besides the fact that the monomer surface area versus interface area plot clearly distinguishes between the two classes of proteins, the disordered proteins, with variable extended shapes and with variable interface areas, are observed to distribute sparsely over the plot. On the other hand, ordered proteins, being globular, compact, and rather similar to each other, occupy a more localized region on the plot. The authors emphasized that

TABLE 1 Summary of the web servers offering prediction of intrinsically disordered proteins

Server name	URL	Approach	References
VLXT (PONDR)	http://www.pondr.com	Feed-forward neural network with separate N-/C-terminus predictor. Based on amino-acid compositions and physicochemical properties.	(54,57,58)
FoldIndex	http://bip.weizmann.ac.il/fldbin/findex	Charge/hydrophobicity score based on a sliding window.	(26,67)
NORSp	http://rostlab.org/services/NORSp/	Rule-based using a set of several neural-networks. Amino acid compositions and sequence profiles used as features.	(69,70)
VL2/VL3	http://www.ist.temple.edu/disprot/predictor.php	Ordinary least-squares linear regression (VL2) and bagged feed-forward neural-network (VL3).	(15,72,79)
	http://www.pondr.com	All models use amino-acid compositions and sequence complexity. VL3 series uses sequence profiles.	
DISOPRED	http://bioinf.cs.ucl.ac.uk/disopred/	Feed-forward neural network (DISOPRED) and linear support vector machine (DISOPRED2) based on sequence profiles.	(82–84)
GlobPlot	http://globplot.embl.de/	Autoregressive model based on amino-acid propensities for disorder/globularity.	(85)
DisEMBL	http://dis.embl.de/	Ensemble of feed-forward neural networks.	(68)
IUPred	http://iupred.enzim.hu/index.html	Linear model based on the estimated energy of pairwise interactions in a window around a residue.	(86,87)
PreLink	http://genomics.eu.org/spip/PreLink	Rule-based. Ratio of multinomial probabilities (for linker and structured regions) combined with the distance to the nearest hydrophobic cluster.	(89)
RONN	http://www.strubi.ox.ac.uk/RONN	Feed-forward neural network in the space of distances to a set of prototype sequences of known fold state.	(88)
DISpro	http://www.igb.uci.edu/servers/psss.html	Recursive neural network based on sequence profiles, predicted secondary structure and relative solvent accessibility.	(90)
VSL	http://www.ist.temple.edu/disprot/predictorVSL2.php	Logistic regression (VSL1) and linear support vector machine (VSL2) based on sequence composition, physicochemical properties and profiles. Combination of short and long disorder predictors.	(80,81)
DRIP-PRED	http://www.sbc.su.se/~maccallr/disorder/	Kohonen's self-organizing maps based on sequence profiles.	—
SPRITZ	http://protein.cribi.unipd.it/spritz/	Nonlinear support vector machine based on multipally aligned sequences. Separate predictors for short and long disorder regions.	(91)

this approach, being structure-based, can be extended to proteins with homology-modeled structures. Finally, they pointed out that their finding can be utilized for the de novo design of stable monomeric proteins and peptides (93).

Recently, as shown in Fig. 2 (94), the monomer area versus interface area plot has been used to test for the presence of binding-induced disorder-to-order transitions in a set of polypeptides having molecular recognition features (MoRFs). These are short, intrinsically disordered peptides that undergo disorder-to-order transitions upon partner recognition (94,95). As Fig. 2 shows, almost all of the MoRFs in the dataset collected from PDB were on the intrinsic disorder side of the boundary that was developed in the original study using a completely different set of proteins (93). These results suggest that these peptides responsible for recognition are likely to be disordered in isolation, which was further supported by high disorder predictions in regions flanking the MoRFs of these polypeptides (94).

Prediction of disorder in computer-aided functional proteomics

In this section we review various applications of the predictors of IDPs and ID regions. We distinguish three major situations in which ID predictors were used: 1), to improve estimation of commonness of disorder and its functional repertoire; 2), to facilitate or improve prediction of other protein features such as protein post-translational modification sites or other types of binding regions; and 3), as a tool to gain insight into structural and dynamic properties of the proteins of interest, both in individual and high-throughput experiments.

Estimation of commonness of disorder and its functional repertoire

The first application of the predictors appeared as soon as the first model was trained. Romero et al. (96,97) estimated the commonness of protein disorder in the Swiss-Prot database (98) with the finding that 25% of proteins in Swiss-Prot had predicted ID regions longer than 40 consecutive residues and that at least 11% of residues in Swiss-Prot were likely to be disordered. Given the existence of a few dozen experimentally characterized disordered regions at the time, this work had significant influence on the recognition of the importance of studying disordered proteins. If indeed 25% of all proteins contained long disordered regions, the natural question to ask was, what biological functions are carried out by these IDPs?

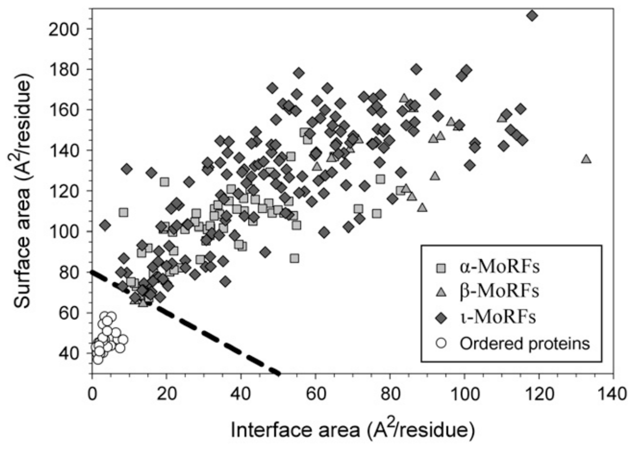

FIGURE 2 Surface and interface area normalized by the number of residues in each chain for MoRF and the ordered complexes datasets. Modified from Mohan et al. (94).

Vucetic et al. (72) developed a supervised clustering algorithm in an attempt to discover possible types or ''flavors'' of disorder and applied these flavor-specific predictors to 28 available genomes from the three kingdoms of life. First, this work revealed that there indeed were distinct types of disorder (three flavors were found) and even more interestingly that various types of disorder could be responsible for different protein functions. In addition, even though archaea and bacteria seemed to have similar relative frequency of disordered proteins, the distribution of the flavor of their disorder was largely different. Confirming the initial analysis by Garner et al. (99) and Dunker et al. (10), it has been shown that disordered proteins were involved in protein-nucleic acid and protein-protein binding and that different flavors were associated with different types of molecular functions (72).

Ward et al. (83) have refined and systematized such an analysis and concluded that the fraction of proteins containing disordered regions of 30 residues or longer (predicted using DISOPRED) were 2% in archaea, 4% in bacteria, and 33% in eukarya. In addition, a complete analysis of the yeast proteome with respect to the three Gene Ontology (GO) categories was performed (100). In terms of molecular function, transcription, kinase, nucleic acid and protein binding activity were the most distinctive signatures of disordered proteins. The most overrepresented GO terms characteristic for the biological process category were transposition, development, morphogenesis, protein phosphorylation, regulation, transcription, and signal transduction. Finally, with respect to cellular component, it appeared that nuclear proteins were significantly enriched in disorder, whereas terms membrane, cytosol, mitochondrion and cytoplasm were distinctively overrepresented in ordered proteins (100).

Recently, a novel data-mining tool that identifies ID-correlated functional keywords in the Swiss-Prot database has been elaborated (101–103). An application of this method to a set of over 200,000 Swiss-Prot proteins revealed that out of 711 functional keywords associated with at least 20 proteins, 262 keywords were found to be strongly positively correlated with predictions of long, intrinsically disordered regions, whereas 302 keywords were strongly negatively correlated with such regions. A significant fraction of these predictions were verified by comparing the inferred correlations to information found in the literature. That is, at least one illustrative example of functional disorder or functional order was found for a large majority of the keywords showing the strongest positive or negative correlation with predicted intrinsic disorder, respectively (101–103).

In the next few years, with further improvement of the existing computational approaches and the development of novel bioinformatics tools, we anticipate that prediction of disorder-dependent functions will be made for the proteomes of all the model organisms and for proteins from all major databases. This initial work will be followed by laboratory experiments to verify or disprove these prediction-based annotations. Using prediction to guide experiments will become especially important for accelerating the characterization of IDPs and ID regions (19).

Prediction of functional sites and sites of post-translational modifications

Molecular recognition features

Various predictors of intrinsic disorder have been used to facilitate prediction of functional properties of proteins. The first use of a disorder predictor to find protein-binding sites was performed by Garner et al. (104) who noticed that sharp dips in disorder prediction could indicate short loosely structured binding regions that undergo disorder-to-order transitions upon binding to a partner. Interestingly, these dips in disorder prediction were originally noticed for the 4E binding protein (4EBP1, see Fig. 3) (104), which had been shown to be completely disordered by NMR (105). However, a short stretch of 4EBP1 undergoes a disorder-to-order transition upon binding to eukaryotic translation initiation factor 4E (106). A different example of the same process is shown in Fig. 4, which represents the disorder-to-order transition in a disordered region of Bad (ribbon) induced by its binding to Bcl-XL (globular). The commonness of such interactions is supported by Fig. 2 and the associated work leading to this figure (94).

Additional work has further validated the use of these distinctive downward spikes in VLXT curves to locate functional binding regions. The follow-up study by Oldfield et al. led to the development of a predictor of short helical regions, termed Molecular Recognition Elements (MoREs) (95) or Molecular Recognition Features (MoRFs) (94). A large decrease in conformational entropy that accompanies disorder-to-order transition uncouples specificity from binding strength. This phenomenon has the effect of making highly specific interactions easily reversible, which is beneficial for cells, especially in the inducible responses typically involved in signaling and regulation. A recent computational study of such binding illustrated that the disordered partner contains a ‘‘conformational preference’’ for the structure it will take upon binding, and that these so-called ‘‘preformed elements’’ tend to be helices (107). This research validates previous findings for individual protein-protein interactions, such as p27$^{\mathrm{Kip1}}$ (108,109) and p53 (110), both of which have disordered regions containing significant helical content and with the likely result that these transient α-helices become stabilized upon binding to their partners. Several MoRFs or downward spikes have been first noticed by prediction and later confirmed by experiment to be involved in protein-protein interactions (111–113).

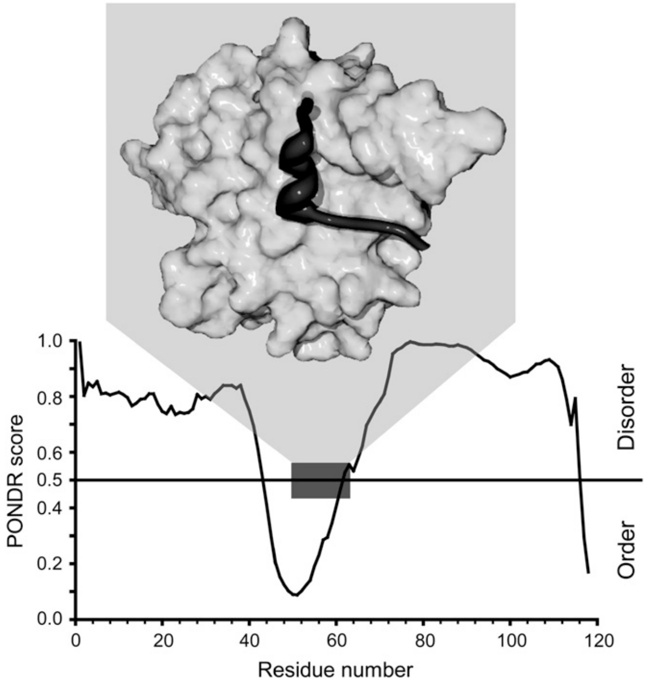

FIGURE 3 Example of a binding region and its positions relative to the regions of predicted order (PONDR VLXT score) and α-MoRF. Eukaryotic initiation factor (*yellow*) and the binding region of 4EBP1 (*dark red*) are shown above the PONDR VLXT plot for 4EBP1, where the binding region and the predicted α-MoRF region are shown as dark red and blue bars, respectively. Modified from Oldfield et al. (95).

Recently, by searching PDB, 1,261 MoRFs were found that were clustered into 372 families by sequence similarity (94). Based on the structure adopted upon binding, at least three basic types of MoRFs were found: α-MoRFs, β-MoRFs, and ι-MoRFs, which form α-helices, β-strands, and irregular secondary structure when bound, respectively (94). Furthermore, the details of the MoRF-partner interactions were compared with other types of protein-protein interactions and several very significant differences were found (114). One of the most striking differences is that MoRF-partner interfaces have a much higher fraction of hydrophobic side chains as compared to interfaces between structured domains. This result is remarkable and interesting because, in the unbound state, MoRF sequences are significantly depleted in hydrophobic groups compared to the sequences of globular proteins (94). Thus, overall a very high percentage of the hydrophobic groups in MoRFs become involved in the binding interfaces with protein partners. These higher numbers of hydrophobic groups and their specific sequence patterns within predicted or experimentally identified regions of intrinsic disorder

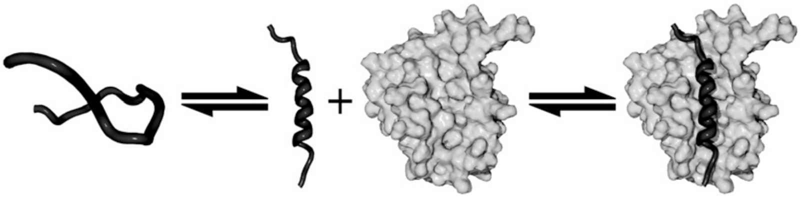

FIGURE 4 Illustration of disorder-to-order transition upon binding. This example shows the binding of a disordered region of Bad (*ribbon*) binding to Bcl-XL (*globular*). Modified from Oldfield et al. (95).

should provide the basis for the development of predictors of MoRFs from sequence. When combined with experiment, these future predictors will be especially helpful in identifying the subregions within longer ID regions that are involved in binding to partners.

Calmodulin binding targets

Calmodulin (CaM), a ubiquitous Ca^{2+} sensor (115), is a highly conserved intracellular protein, which is heavily involved in numerous regulatory processes (116–118). CaM is known to be recruited by at least 180 different proteins and enzymes (119), by which these target proteins express Ca^{2+} sensitivity in their biological functions (120,121). Based on the analysis of the solved structures of CaM associated with several of its binding targets, the distinctive binding mechanism of CaM, and the significant trypsin sensitivity of the binding targets, it has been concluded that the process of association likely involves coupled binding and folding for both CaM and its binding targets (122). To further validate this hypothesis, a set of 287 MoRFs that were known to be CaM binding targets (CaMBTs) has been recently collected (122). Based on this dataset, a predictor of CaMBTs was developed in which the prediction of disorder was used as an input feature to the system. Feature selection has isolated disorder as one of the dominant characteristics of CaMBTs, in addition to the high helical propensity, aromaticity and positive charge (122). Per residue accuracy of this predictor reached 81%, which, in combination with a high recall/precision balance at the binding region level, suggests high predictability of CaM-binding partners. Application of this predictor to yeast and human proteomes revealed that CaMBTs are highly abundant in various activators and repressors, nuclear proteins, DNA- and RNA-binding proteins, helicases, ribosomal proteins, coiled coils, homeobox proteins, protein involved in transcription regulation, development and ATP binding, variants produced by alternative splicing, and proteins with activities regulated by phosphorylation (122).

Sites of post-translational modifications

Recently, various studies showed the importance of intrinsic disorder prediction for the prediction of protein post-translational modification sites. Iakoucheva et al. (123) used prediction of intrinsic disorder to predict phosphorylation sites, whereas Daily et al. (124) used a similar approach to identify protein methylation sites. Our experiments also reveal that protein ubiquitination sites are located within disordered regions and that prediction of disorder was found useful for this important modification (P. Radivojac and L. Iakoucheva, unpublished data).

In all three of the above-mentioned applications, prediction of disorder was used simply as an input feature to the system and was shown to be useful, increasing the accuracy by 2-3 percentage points. However, disorder prediction can also be used in other ways. For example, Radivojac et al. (125) used a predictor of intrinsically disordered regions to cluster protein residues into two groups (disordered and ordered) and then used different thresholds on the raw scores to assign phosphorylated residues. This approach eliminated many false positives that were otherwise found in ordered protein regions. In addition, Beltrao and Serrano (126) showed that SH3 binding domains prefer binding targets that are located within intrinsically disordered regions and showed that an analysis of conservation of linear peptide sequences in combination with prediction of intrinsic disorder can be used to screen for protein-protein interactions.

How does disorder prediction in the above-described problems improve the prediction accuracy? In other words, why would generalized disorder prediction improve accuracy for models specifically trained on their own, problem-specific datasets? We believe that the main reason for this phenomenon results from the small dataset sizes for each of these problems coupled with the ‘‘prior knowledge’’ that disorder is related to each of these functions. For example, in predicting protein phosphorylation sites, only 136 tyrosine and 141 threonine sites had been retained for the predictor construction after redundancy removal (123). On the other hand, predictors of disorder were trained on more than 20,000 nonredundant residues (15). If indeed intrinsic disorder is related to protein phosphorylation, then disorder propensity could be expected to significantly reduce the number of false positive predictions. In this way the datasets used for prediction of disorder are indirectly contributing to the increased accuracy of prediction of other phenomena. In the early stages when only a small number of experimentally verified positive sites or binding regions is available, predictors of disordered regions can be expected to play an important role for those processes for which prior knowledge indicates that disorder is important.

We anticipate that an important future direction will be to combine sequence motif-based prediction, which is commonly used to identify potential binding sites or potential sites of protein modification (127), with disorder-based prediction to improve annotations of the proteomes of various model organisms. If a binding sequence motif or a sequence-motif-based identification of a posttranslational modification site is experimentally characterized to reside in intrinsically disordered regions, then disorder predictions can be used to help focus efforts on experiments that are more likely to be productive. Although in our view prediction of disorder will become increasingly useful for functional proteomics (19), in the end, laboratory experiments will always be essential for unambiguously identifying the sites or regions of interest.

Prediction of disorder as a tool in determining protein structural and dynamic properties

ID, protein crystallization, and structural genomics projects

Due to rapid DNA sequencing, the number of translated protein sequences is growing substantially faster than the number of determined three-dimensional structures. That is, whereas the number of translated protein sequences has surpassed the 4,000,000 mark, the number of protein structures in PDB is nearing the much lower 40,000 number, corresponding to only ~1% of currently determined protein sequences. The discrepancy between these two figures can be partly attributed to the time-intensive and difficult process of producing a protein crystal and then the subsequent labor-intensive process of interpreting the resulting diffraction pattern. Furthermore, a number of bottlenecks have been identified in structural genomic high throughput pipelines (128). A major challenge results from the finding that ~70% of selected targets are predicted to be unsuitable for structural determination using current methods (129). Application of methods that account for protein disorder can greatly reduce these bottlenecks. Close examination of sequences that failed to crystallize may reveal intrinsically disordered regions interspersed with regions of order. Thus, accounting for protein disorder can improve target selection and prioritization. In fact, implicit ID predictions have been used by structural genomics centers to prioritize target selection. For example, proteins with low complexity, coil-coil proteins or very long proteins are typically assigned low priority in structure determination (130). However, IDPs and ID regions are not necessarily low-complexity nor do all multi-domain proteins contain a disordered region. Oldfield et al. (131) explicitly utilized predictions of protein disorder to pre-screen 71 proteins in the pipeline from *Arabidopsis thaliana*. The authors showed clear benefits of using disorder predictions in the analysis as compared to simple sequence complexity analysis. This result is especially important in light of the fact that an emphasis in structural determination is given to the discovery of new folds. Alternative analyses of disordered protein regions, for example by identifying regions of low sequence conservation, have been used by crystallographers for many years to change expression constructs in attempts to avoid difficult-to-crystallize protein regions.

Researchers can utilize disorder prediction at the level of individual proteins as well. Recently it has been shown that crystallization trials for full-length NEIL1, a human homolog of *E. coli* DNA glycosylase endonuclease VIII, failed to yield any crystals. This inability to grow crystals was corroborated by the fact that the protein was polydisperse regardless of the temperature or buffer conditions used, based on dynamic light scattering (DLS) experiments (132). To resolve this problem, the VLXT predictor was used to indicate possible disordered region(s) in NEIL1 that might have hindered crystallization. The analysis showed that this protein likely had a disordered C-terminal region (106 residues). A set C-terminal deletion constructs were cloned and checked for expression. A NEIL1 construct missing the C-terminal 100 amino acids (NEIL1C_100) was successfully crystallized, whereas deletions of >100 residues did not yield any protein expression (132). This study clearly illustrates the usefulness of serious consideration of ID for successful crystallization of proteins and protein fragments. With the set of tools to be developed in the near future, researchers will be able to identify those proteins or portions of proteins which are more likely to be soluble (133) and which are more likely to crystallize (134), with higher accuracy.

As a further illustration of the use of disorder prediction, based on previous reports that many viral proteins have a modular organization containing hydrophobic and disordered regions that are often not compatible with the crystallization process (135,136), the ''viral enzyme module localization'' (VaZyMolO) tool was recently developed which serves to define and classify viral protein modularity (137). Among different attributes used by VaZyMolO to produce modules suitable for crystallization, protein regions that may contain hydrophobic (peptide signal, hydrophobic domain and transmembrane) or natively disordered patterns were precisely defined. In the absence of three-dimensional data, a systematic bioinformatics analysis was performed to define globular and disordered regions. Disordered regions were identified by combining the results from the analysis of the mean hydrophobicity/mean charge ratio (26), as well as from VLXT (57) and DisEMBL (68) predictions.

Besides the crucial role of the prediction of intrinsic disorder in finding new targets for structural analysis, various disorder predictors have proved their usefulness for gaining insight into structural and dynamic properties of different proteins and protein families and for better understanding protein function. This is truly an exploding field with several studies describing new usage of intrinsic disorder published each week. A few illustrative examples are outlined below.

IDPs in DNA repair and cancer

One of the first applications of the disorder predictors for structural characterization of proteins is exemplified by the analysis of the *Xeroderma pigmentosum* group A (XPA) DNA repair protein using the VLXT predictor, limited proteolysis and mass-spectrometry (138). The disorder predictions indicated that XPA carries extended disordered regions on its N- and C-termini with an ordered central core. These predictions agreed well with the partial proteolysis results; the trypsin cleavage sites were observed in XPA termini but not within its internal region despite the presence of 14 possible cut sites in this region. Furthermore, the NMR structure of the internal core confirmed the prediction of order for this segment. Thus, disorder analysis helped provide a better insight into structural properties of this important DNA repair protein. In agreement with this example, it has been established that ID is also very common in cancer-associated proteins. Of cancer-associated proteins, 79% contain predicted regions of disorder of 30 residues or longer (46). In contrast, only 13% of a set of proteins with well-defined ordered structures contained such long regions of predicted disorder. In experimental studies, the presence of disorder has been directly observed in several cancer-associated proteins, including p53 (110), $p57^{kip2}$ (139), Bcl-X_L and Bcl-2 (140), c-Fos (141), and most recently, a thyroid cancer associated protein, TC-1 (142).

IDPs and human papillomaviruses

A recent comparison of the proteomes of the oncogenic and benign types of human papillomaviruses (HPV) provided additional evidence of a correlation between ID and cancer (143). In humans, there are more than 100 different types of HPVs. Some of them are the causative agents of benign papillomas/warts, whereas other HPVs are cofactors in the development of carcinomas of the genital tract, the head and neck, and the epidermis. Specific types of HPV play causal role in cervical cancer, a major cause of women's death worldwide, with ~200,000 women dying of this disease each year (144–146). With respect to their association with cancer, HPVs are grouped into two classes, known as low- (e.g., HPV-6 and HPV-11) and high-risk (e.g., HPV-16 and HPV-18) types (144,147).

The papillomaviruses (PV) are small nonenveloped icosahedral viruses found in many animals as well as in man. These viruses have a circular double stranded DNA genome of ~8 kb that encode eight to nine proteins, including six nonstructural proteins [E1, E2, E4, E5, E6 and E7 (the latter two are known to function as oncoproteins in the high-risk HPVs)] and two structural proteins (L1, and L2) (145,146,148). Similar to other DNA viruses, these viruses are dependent upon the cellular machinery to replicate their nucleic acid and complete a productive life cycle. HPVs achieve the proper cellular environment by inducing cells to enter S phase (146,148).

To understand whether ID plays a role in the oncogenic potential of different HPVs and thus to differentiate the cancer-related and benign HPVs, a detailed bioinformatics analysis of proteomes of high-risk and low-risk HPVs was performed with the major focus on the E6 and E7 oncoproteins (143). This analysis indicates that high-risk HPVs are characterized by a significantly increased amount of predicted intrinsic disorder in transforming proteins E6 and E7 (143). The results of ID prediction in E7 oncoprotein are consistent with the solution structure recently determined for this protein from the high-risk HPV-45 (149), as both the NMR analysis and the predicted disorder distribution showed that the N-terminal fragment of E7 (residues 1-54) is completely disordered.

IDPs in cardiovascular disease

The high abundance of ID in proteins associated with cardiovascular disease (CVD), which has been recognized as the No. 1 killer in the United States, has been recently established using the bioinformatics analysis of a dataset of 487 CVD-related proteins extracted from the Swiss-Prot using keyword searches (150). This analysis suggests that CVD-related proteins are depleted in major order-promoting residues (W, F, Y, I, and V) and are enriched in several disorder-promoting residues (R, Q, S, P, and E). The application of several ID predictors (including VLXT, CH-plot, CDF analysis, and α-MoRF indicator) revealed that CVD-related proteins are highly enriched in intrinsic disorder, with many proteins being predicted to be wholly disordered (150). This high level of ID could be important for the functions of CVD-related protein and for the control and regulation of processes associated with CVD. In agreement with this hypothesis, 198 α-MoRFs were predicted in 101 proteins from CVD dataset. A comparison of disorder predictions with the experimental structural and functional data for a subset of the CVD-associated proteins indicated good agreement between predictions and observations (150).

ID in PEST proteins

PEST sequences, which have been indicated to be protein degradation targeting signals, are enriched in proline (P), glutamic acid (E), serine (S), and threonine (T). PEST sequences were first observed in rapidly degraded, eukaryotic intracellular proteins (151) and are believed to confer rapid instability to many proteins (151,152). Various experimental approaches including deletion, transfer, and mutation of PEST sequences have shown the role and importance of PEST regions for the stability of proteins (153,154). There are the two major protein degradation pathways that are implicated in PEST-mediated proteolysis, the ubiquitin-proteasome degradation and the calpain cleavage (155,156).

P, E, S, and T are among the disorder-promoting amino acids (Fig. 1), thus sequences rich in these amino acids would be expected to be intrinsically disordered. This was validated in a recent study (157), which showed that PEST motifs are associated disordered regions more often than with globular proteins. Furthermore, analysis of representative PDB entries revealed very few structures containing PEST sequences, with the vast majority of the PEST-containing regions of PDB entries being characterized by the lack of ordered secondary structure. Other important findings based on a proteome-wide analysis included the following observations:

1), PEST proteins are prevalent in eukaryotic proteomes; 2), they comprise a large fraction of the unfolded proteome in completely sequenced eukaryotes; and 3), the PEST-containing proteins show an over- and an underrepresentation in functions related to regulation and metabolism, respectively (157). More recently, the disorder of the PEST motif of the suppressor of cytokine signaling SOCS3 has been confirmed experimentally by NMR (158).

ID in nuclear localization signals

A nuclear localization signal (NLS) is a short amino-acid sequence that mediates transport of nuclear proteins into the nucleus of the cell. The classical NLS was first discovered in the simian virus 40 (SV40) T-antigen and consisted of a string of seven basic amino-acid residues (PKKKRKV) (159). The discovery of the bipartite NLSs soon followed. The bipartite NLSs comprise two strings of basic amino acid residues separated by a short intervening sequence (reviewed in (160)). These classical NLSs bind the adaptor protein Kapα, which forms a heterodimer with Kapβ1, which in turn mediates nuclear import (161).

In addition to the previous examples, many of the proteins imported into the nucleus do not utilize such an adaptor but rather bind directly to a Kapβ. These proteins contain a more complex and diverse set of NLS sequences. In humans ten distinct import Kapβs carry a diverse set of macromolecular substrates into the nucleus, and each Kapβ appears to bind distinct sets of substrates (162). The very large sequence diversity among various substrates together with a limited number of substrates that have been identified for most import Kapβs has so far prevented identification of NLSs for most Kapβs.

A recent study for import by one of the karyopherins, Kapβ2, led to three rules for this protein's NLS recognition: 1. NLSs are structurally disordered in free substrates; 2. they have overall basic character; and 3. they contain a set of consensus sequences (163). Application of these three rules was used to first computationally identify and then to biochemically confirm NLSs in seven known Kapβ2 substrates (163). Furthermore, 81 new candidate import substrates for Kapβ2 were predicted, and five of them were confirmed to bind Kapβ2 through the predicted NLS. This example demonstrates how disorder predictions aided in understanding the mechanism of substrate recognition by Kapβ2 and supports our thesis that the combination of disorder prediction and biophysical experiments to confirm the disorder provides a new avenue for the understanding of regulation, signaling and control.

IDPs in apicomplexan parasite proteomes

Malaria, being present in areas where ~40% of the world's population lives and causing up to 2.7 million deaths each year, remains a major and growing threat to the public health (164). Malaria is caused by infection with the apicomplexan parasite *Plasmodium falciparum*, the sequencing of which has been completed recently (165). The abundance of IDPs in *P. falciparum* and several apicomplexan parasites, together with the variation in the IDP content associated with four stages of the life cycle of *P. falciparum* were analyzed using the DisEMBL predictor (166). The apicomplexan species are extremely enriched in proteins containing long disordered regions. Furthermore, the disorder contents in mammalian *Plasmodium* species were higher than in most other apicomplexan parasites. Finally, the proteome of the *P. falciparum* sporozoite was shown to be distinct from the other life cycle stages in having an even higher content of disordered proteins (166).

ID in voltage-activated potassium channels

Voltage-activated potassium channels (known also as K_v channels) are modular proteins composed of several domains including a ball-and-chain inactivation domain, a tetramerization (T1) domain, membrane-spanning voltage-sensor and pore domains, and an intracellular C-terminal segment. K_v are allosteric pore-forming proteins that undergo conformational transitions between closed and open states thus underlying many fundamental biological processes (167–169). The crucial role of ID and high conformational flexibility in the functioning of a ball-and-chain inactivation domain was recognized long ago (16). Specifically, a ''ball'' on the end of a flexible (disordered) polypeptide ''chain'' was suggested to plug the open channel, thereby converting the channel from the open to the inactive state (170–175). Furthermore, the length and flexibility of a disordered polypeptide ''chain'' were shown to be responsible for the control of the rate of channel inactivation (174).

In addition to this well-established role of ID in the inactivation/activation cycle of the K_v channels, the C-terminal segments of K_v channels have been suggested recently to be disordered as indicated by CH-plots and the FoldIndex predictor. The ID at the C-terminus is suggested to enable K^+ channel binding to scaffold proteins by means of an intermolecular, fishing rod-like mechanism (176).

ID and histones

The core (H2A, H2B, H3, H4) and linker (H1 family) histones are the major protein components of chromatin fibers (177,178). The nucleosome core particle represents the elemental subunit in the hierarchy of DNA packaging in chromatin. The eukaryotic core nucleosome contains eight histone proteins, two dimers of H2A–H2B that serve as molecular caps for the central $(H3–H4)_2$ tetramer. The sequence of a given type of histone is highly conserved from yeast to mammals, but there is minimal sequence identity, at the level of 4–6%, between the histone proteins (179). Linker histones comprise a family of nucleosome-binding proteins that stabilize condensed chromatin and regulate genome function (177,180). The linker histones of most eukaryotes have a very simple domain organization, consisting of a central winged helix fold, a short N-terminal extension, and a long basic C-terminal domain, which is ~100 residues in length, enriched in K, A, and P, and unstructured in aqueous solution (181). Simple bioinformatics analysis using CH-plots and FoldIndex predictor revealed that bovine core histones H2A, H2B, H3, and H4 are also significantly enriched in intrinsic disorder. This prediction was corroborated by subsequent experimental analysis showing that the bovine core histones are natively unfolded proteins in solutions with low ionic strength due to their high net positive charge at pH 7.5 (182). The N-terminal ''tail'' domains (NTDs) of the core histones and the C-terminal tail domain (CTD) of linker histones are intrinsically disordered, and this property likely facilitates their binding to many different macromolecular partners in chromatin (183).

ID and hub proteins

The crucial role of intrinsic disorder for the function of several individual hub proteins (i.e., proteins with a high degree of connectivity) with known disordered regions was recently reviewed (16,47,184). Furthermore, recent systematic computational analysis of proteins with various numbers of interacting partners from four eukaryotic organisms (*C. elegans, S. cerevisiae, D. melanogaster*, and *H. sapiens*) revealed that for all four studied organisms, hub proteins, defined as those that interact with ≥10 partners, were significantly more disordered than end proteins, defined as those that interact with just one partner (185). A study by Ekman et al. reports a similar finding in which the difference between hubs and nonhubs is created predominantly by the date hubs (as opposed to the party hubs), thus suggesting the importance of ID in transient binding (186). Two other recent studies indicate that ID is an important property for enabling hub proteins to interact with many partners (187,188). These various results provide strong support for the hypothesis that ID represents a distinctive and common characteristic of hub proteins, likely serving as an important determinant of protein interactivity.

ID in serine/arginine-rich splicing factors

In a recent study (189) a disorder predictor was used to estimate the disorder content of proteins involved in RNA splicing. Serine/arginine-rich (SR) splicing factors are essential for both constitutive and alternative splicing of pre-mRNAs. These proteins have modular organization, consisting of RNA recognition motifs (RRMs), located on their N-terminus, and an arginine-serine-rich (RS) domain, located on the C-terminus. Both domains have a broad binding specificity, e.g., they are involved in numerous protein-protein and protein-RNA interactions. The previous structural knowledge about SR proteins has been limited to only RRM domains. The application of the disorder predictor showed that the members of this protein family belong to a class of intrinsically disordered proteins. The amino acid composition and sequence complexity of SR proteins are very similar to those of disordered protein regions. Furthermore, the RS domains and the Gly-rich regions of these splicing factors are predicted to be completely disordered, whereas RRM domains are predicted to be ordered in agreement with previous structural studies. The disorder of RS domains may play an important role in several functions of SR proteins such as binding to multiple partners (proteins and RNA), in mediating interactions of spliceosome components during the assembly process, and in facilitating post-translational modifications that are abundant in the RS domains.

Intrinsic disorder of 14-3-3 proteins partners

The application of various disorder predictors with the aim of gaining biologically important insights is reflected in yet another recent study (190). The authors discovered that the distinctive feature of seemingly unrelated binding partners of the 14-3-3 proteins is high disorder content. Based on the results from three different disorder predictors (VL3H, VLXT, and DISOPRED2), >90% of 14-3-3 binding partners were indicated to contain disordered regions. Since almost all 14-3-3 proteins bind to a specific phosphoserine/phosphothreonine-containing peptide motif within their targets, the analysis also demonstrated that the binding sites of 14-3-3 proteins were located inside disordered regions. Also, the structures of two peptides bound to 14-3-3 exhibit extended backbones with their backbone hydrogen bonds largely formed by interactions with the side chains of 14-3-3 but with slightly different hydrogen bonding patterns for the two different peptides (191). These structures are entirely consistent with the peptides being unfolded before binding to 14-3-3. Thus, the mode of interaction between 14-3-3 proteins and their targets is proposed to involve disorder-to-order transition upon binding (190).

ID and transcription factors

Transcription factors (TFs) regulate the activation of transcription via the recognition of specific DNA sequences coupled with the recruitment and assembly of the transcription machinery. This implies that both protein-DNA and protein-protein recognition play key roles in TF function. Available experimental data points to a central role of ID in the function of TFs (192). For example, it has been reported that protein-protein and protein-DNA interactions are typically accompanied by a local folding of TF molecules (192). Furthermore, the high degree of backbone mobility of the *lac* repressor was shown to facilitate its association with nonspecific DNA, whereas the binding to specific DNA was accompanied by a considerable decrease in the backbone mobility (193). In addition to these instances, several other well-characterized examples of the individual ID proteins involved in transcriptional regulation have been described in the literature (184,194). The overwhelming prevalence of ID in TFs was been recently established using a set of ID predictors (195). This analysis revealed that >90% of transcription factors might possess extended regions of ID. Furthermore, the analysis of ID distribution in different TFs and their domains revealed that the eukaryotic TFs are essentially more enriched in ID and α-MoRFs that prokaryotic TFs. Interestingly, the AT-hooks and basic regions of TF DNA-binding domains where predicted to be highly disordered, whereas the degree of disorder in transactivation regions was even higher (195).

The abundance of ID in TF has been further confirmed by the detailed comparison of the human transcriptional regulation factors (including activators, repressors, and enhancer-binding factors) with their prokaryotic counterparts (196). These comparison revealed that human and prokaryotic TFs are different in at least two respects: the average TF sequence in human is more than twice as long as that in prokaryotes, whereas the fraction of sequence aligned to domains of known structure in human TFs (31%) is <½ of that in bacterial TFs (72%). Furthermore, it has been established that ID regions occupy a high fraction of sequence in the eukaryotic TFs, but not in prokaryotes (196). This suggests that the efficiency of the well-developed gene transcription machinery of eukaryotes relies to a significant degree on the TF flexibility.

Similar analyses have been applied to numerous other proteins. For example, the disorder predictions aided in structural and/or functional characterization of the retinal tetraspanin (197), nicotinic acetylcholine receptor (198), DBE (199), proapoptotic BH domain-containing family of proteins (200), transcriptional corepressor CtBP (201), colicin E9 (202), troponin I (203), secA (204), Notch signaling pathway proteins (205) and many others.

CONCLUSIONS

In the last 10–15 years, the field of intrinsically disordered proteins has transitioned from its infancy into an important and dynamic field of protein science. As summarized in previous sections, this field has grown rapidly in part due to a potent synergy between experimental and computational techniques. Although the importance of intrinsically disordered proteins is established and will continue to grow, especially in the fields of evolution and drug design, it is yet to reach the textbook level and ultimate recognition. Indeed, current biochemistry textbooks ignore disordered proteins (206), and in our view, this omission has serious consequences, leading to a significant retardation in the understanding of protein structure/function relationships.

A common characteristic of these disordered regions is that functions are often carried out by a few localized residues within the disordered regions. Independent of the biophysical, structure-based work described herein, there has been a substantial body of work in which functional motifs are determined sequence analysis and molecular biology experiments without biophysical structural characterization. Indeed servers exist for using sequence comparisons to find such functional motifs (127,208). The discovered function-associated motifs are often short sequences, which are called eukaryotic linear motifs (ELMs) by one research group (127), and these function-associated motifs resemble in many ways the functional regions found to reside within long ID regions. The PEST (157) and NLS (163) examples discussed above suggest a possible correlation between ID and the functional motifs found by sequence analysis. Indeed, we anticipate that functional ELMs and other functional motifs will usually map to regions of disorder, whereas the same sequence motifs that are found to be nonfunctional in some proteins will likely map to regions of structure. Clearly, examining functional motifs with disorder prediction followed by systematic biophysical studies to determine the order-disorder status of the various functional motifs should be carried out.

When looking into the future, some questions regarding the computational techniques become legitimate. What is the future of the prediction of intrinsic disorder? Has disorder prediction reached its maximum accuracy or can prediction accuracy still be improved? Our internal experiments indicate that sequence-based prediction of intrinsically disordered regions is indeed nearing its upper limit of ~85–90% (A. Mohan and P. Radivojac, unpublished data). To reach this limit, however, high quality data and possibly even novel computational methods will be required. For example, exploiting other types of data such as text (for use in text mining), interaction data, expression patterns, or functional annotation could certainly lead to even higher accuracy of prediction. In addition, methods based on first-principles may start gaining importance as computational power grows in the future. For example, methods such as SnapDRAGON (207), used in prediction of protein domains, are expected to play an important role. Indeed, models for predicting both structural and dynamic properties of proteins together with predictions of interactions with partners are ambitious goals currently being actively pursued. Prediction of disorder is likely to be an important piece for achieving these goals.

The Indiana Genomics Initiative, funded in part by the Lilly Endowment, and National Institutes of Health grant No. 1 R01 LM007688-0A1 provided support for P.R., V.N.U, Z.O, and A.K.D. This work received additional support from the Programs of the Russian Academy of Sciences for ''Molecular and Cellular Biology'' and ''Fundamental Science for Medicine'', especially for V.N.U. L.M.I. was supported by National Science Foundation grant No. MCB0444818.

REFERENCES

1. Pauling, L., R. B. Corey, and R. H. Branson. 1951. The structure of proteins: two hydrogen-bonded configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA*. 37:205–210.
2. Fischer, E. 1894. Einfluss der configuration auf die wirkung der enzyme. *Ber. Dt. Chem. Ges*. 27:2985–2993.
3. Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. 181:662–666.
4. Kendrew, J. C., R. E. Dickerson, and B. E. Strandberg. 1960. Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution. *Nature*. 206:757–763.
5. Blake, C. C., D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma. 1965. Structure of hen egg-white

lysozyme. A three-dimensional Fourier synthesis at 2 Ångström resolution. *Nature*. 206:757–761.

6. Anson, M. L., and A. E. Mirsky. 1925. On some general properties of proteins. *J. Gen. Physiol.* 9:169–179.
7. Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science*. 181:223–230.
8. Wu, H. 1931. Studies on denaturation of proteins. XIII. A theory of denaturation. *Chin. J. Physiol.* 1:219–234.
9. Mirsky, A. E., and L. Pauling. 1936. On the structure of native, denatured and coagulated proteins. *Proc. Natl. Acad. Sci. USA*. 22: 439–447.
10. Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* 19:26–59.
11. Daughdrill, G. W., G. J. Pielak, V. N. Uversky, M. S. Cortese, and A. K. Dunker. 2005. Natively disordered protein. *In* Protein Folding Handbook. J. Buchner and T. Kiefhaber, editors. Wiley-VCH:Verlag, Weinheim, Germany. 271–353.
12. Weinreb, P. H., W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury, Jr. 1996. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*. 35:13709–13715.
13. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321–331.
14. Halle, B. 2002. Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. USA*. 99:1274–1279.
15. Obradovic, Z., K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, and A. K. Dunker. 2003. Predicting intrinsic disorder from amino acid sequence. *Proteins*. 53:566–572.
16. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2005. Showing your ID: intrinsic disorder as an ID for recognition, regulation, and cell signaling. *J. Mol. Recogn.* 18:343–384.
17. Ringe, D., and G. A. Petsko. 1986. Study of protein dynamics by x-ray diffraction. *Methods Enzymol.* 131:389–433.
18. Dyson, H. J., and P. E. Wright. 2002. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv. Protein Chem.* 62:311–340.
19. Bracken, C., L. M. Iakoucheva, P. R. Romero, and A. K. Dunker. 2004. Combining prediction, computation and experiment for the characterization of protein disorder. *Curr. Opin. Struct. Biol.* 14:570–576.
20. Dyson, H. J., and P. E. Wright. 2004. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* 104:3607–3622.
21. Dyson, H. J., and P. E. Wright. 2005. Elucidation of the protein folding landscape by NMR. *Methods Enzymol.* 394:299–321.
22. Fasman, G. D. 1996. Circular Dichroism and the Conformational Analysis of Biomolecules. Plenum Press, New York.
23. Adler, A. J., N. J. Greenfield, and G. D. Fasman. 1973. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol.* 27:675–735.
24. Provencher, S. W., and J. Glockner. 1981. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*. 20:33–37.
25. Woody, R. W. 1995. Circular dichroism. *Methods Enzymol.* 246:34–71.
26. Uversky, V. N., J. R. Gillespie, and A. L. Fink. 2000. Why are ''natively unfolded'' proteins unstructured under physiologic conditions? *Proteins*. 41:415–427.
27. Smyth, E., C. D. Syme, E. W. Blanch, L. Hecht, M. Vasak, and L. D. Barron. 2001. Solution structure of native proteins with irregular folds from Raman optical activity. *Biopolymers*. 58:138–151.
28. Uversky, V. N. 1999. A multiparametric approach to studies of self-organization of globular proteins. *Biochemistry (Mosc.)*. 64:250–266.
29. Receveur-Brechot, V., J. M. Bourhis, V. N. Uversky, B. Canard, and S. Longhi. 2006. Assessing protein disorder and induced folding. *Proteins*. 62:24–45.
30. Markus, G. 1965. Protein substrate conformation and proteolysis. *Proc. Natl. Acad. Sci. USA*. 54:253–258.
31. Mikhalyi, E. 1978. Application of Proteolytic Enzymes to Protein Structure Studies. CRC Press, Boca Raton, FL.
32. Hubbard, S. J., F. Eisenmenger, and J. M. Thornton. 1994. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.* 3:757–768.
33. Fontana, A., P. P. de Laureto, V. de Filippis, E. Scaramella, and M. Zambonin. 1997. Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* 2:R17–R26.
34. Fontana, A., P. P. de Laureto, B. Spolaore, E. Frare, P. Picotti, and M. Zambonin. 2004. Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* 51:299–321.
35. Iakoucheva, L. M., A. L. Kimzey, C. D. Masselon, R. D. Smith, A. K. Dunker, and E. J. Ackerman. 2001. Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Sci.* 10:1353–1362.
36. Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27:527–533.
37. Privalov, P. L. 1979. Stability of proteins: small globular proteins. *Adv. Protein Chem.* 33:167–241.
38. Ptitsyn, O. 1995. Molten globule and protein folding. *Adv. Protein Chem.* 47:83–229.
39. Ptitsyn, O. B., and V. N. Uversky. 1994. The molten globule is a third thermodynamical state of protein molecules. *FEBS Lett.* 341:15–18.
40. Uversky, V. N., and O. B. Ptitsyn. 1996. All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. *Fold. Des.* 1:117–122.
41. Westhof, E., D. Altschuh, D. Moras, A. C. Bloomer, A. Mondragon, A. Klug, and M. H. Van Regenmortel. 1984. Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature*. 311:123–126.
42. Berzofsky, J. A. 1985. Intrinsic and extrinsic factors in protein antigenic structure. *Science*. 229:932–940.
43. Kaltashov, I. A., and A. Mohimen. 2005. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal. Chem.* 77:5370–5379.
44. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.
45. Dunker, A. K., and Z. Obradovic. 2001. The protein trinity—linking function and disorder. *Nat. Biotechnol.* 19:805–806.
46. Iakoucheva, L. M., C. J. Brown, J. D. Lawson, Z. Obradovic, and A. K. Dunker. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323:573–584.
47. Dunker, A. K., M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272:5129–5148.
48. Dunker, A. K., C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic. 2002. Intrinsic disorder and protein function. *Biochemistry*. 41:6573–6582.
49. Dunker, A. K., C. J. Brown, and Z. Obradovic. 2002. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* 62:25–49.
50. Sim, K. L., T. Uchida, and S. Miyano. 2001. ProDDO: a database of disordered proteins from the Protein Data Bank (PDB). *Bioinformatics*. 17:379–380.
51. Vucetic, S., Z. Obradovic, V. Vacic, P. Radivojac, K. Peng, L. M. Iakoucheva, M. S. Cortese, J. D. Lawson, C. J. Brown, J. G. Sikes, C. D. Newton, and A. K. Dunker. 2005. DisProt: a database of protein disorder. *Bioinformatics*. 21:137–140.
52. Sickmeier, M., J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, and A. K. Dunker. 2006. DisProt: the database of disordered proteins. *Nucleic Acids Res.* In press.

53. Radivojac, P., Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, and A. K. Dunker. 2004. Protein flexibility and intrinsic disorder. *Protein Sci.* 13:71–80.

54. Romero, P., Z. Obradovic, C. R. Kissinger, J. E. Villafranca, and A. K. Dunker. 1997. Identifying disordered regions in proteins from amino acid sequences. *IEEE Int. Conf. Neural Netw.* 1:90–95.

55. Lise, S., and D. T. Jones. 2005. Sequence patterns associated with disordered regions in proteins. *Proteins.* 58:144–150.

56. Berman, H., T. N. Bhat, P. Bourne, Z. Feng, G. Gilliand, and H. Weissig. 2000. The Protein DataBank and the challenge of structural genomics. *Nat. Struct. Biol.* 7:957–959 (Structural Genomics supplement).

57. Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. 2001. Sequence complexity of disordered protein. *Proteins.* 42:38–48.

58. Li, X., P. Romero, M. Rani, A. K. Dunker, and Z. Obradovic. 1999. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.* 10:30–40.

59. Wootton, J. C. 1993. Statistic of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* 17:149–163.

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

61. Wootton, J. C. 1994. Sequences with ''unusual'' amino acid compositions. *Curr. Opin. Struct. Biol.* 4:413–421.

62. Wootton, J. C. 1997. Evaluating the effectiveness of sequence analysis algorithms using measures of relevant information. *Comput. Chem.* 21:191–202.

63. Wootton, J. C., and S. Federhen. 1996. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* 266:554–571.

64. Romero, P., Z. Obradovic, and A. K. Dunker. 1999. Folding minimal sequences: the lower bound for sequence complexity of globular proteins. *FEBS Lett.* 462:363–367.

65. Weathers, E. A., M. E. Paulaitis, T. B. Woolf, and J. H. Hoh. 2007. Insights into protein structure and function from disorder-complexity space. *Proteins.* 66:16–28.

66. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

67. Prilusky, J., C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman, and J. L. Sussman. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics.* 21:3435–3438.

68. Linding, R., L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell. 2003. Protein disorder prediction: implications for structural proteomics. *Structure.* 11:1453–1459.

69. Liu, J., H. Tan, and B. Rost. 2002. Loopy proteins appear conserved in evolution. *J. Mol. Biol.* 322:53–64.

70. Liu, J., and B. Rost. 2003. NORSp: predictions of long regions without regular secondary structure. *Nucleic Acids Res.* 31:3833–3835.

71. Vucetic, S., P. Radivojac, Z. Obradovic, C. J. Brown, and A. K. Dunker. 2001. Methods for improving protein disorder prediction. *In* International Joint INNS-IEEE Conference on Neural Networks. Washington, DC. 2718–2723.

72. Vucetic, S., C. J. Brown, A. K. Dunker, and Z. Obradovic. 2003. Flavors of protein disorder. *Proteins.* 52:573–584.

73. Weathers, E. A., M. E. Paulaitis, T. B. Woolf, and J. H. Hoh. 2004. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett.* 576:348–352.

74. Stoffer, D. A., and L. G. Volkert. 2005. A neural network for predicting protein disorder using amino acid hydropathy values. *In* IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. San Diego, CA. 482–490.

75. Garbuzynskiy, S. O., M. Y. Lobanov, and O. V. Galzitskaya. 2004. To be folded or to be unfolded? *Protein Sci.* 13:2871–2877.

76. Melamud, E., and J. Moult. 2003. Evaluation of disorder predictions in CASP5. *Proteins.* 53(Suppl 6):561–565.

77. Jin, Y., and R. L. Dunbrack, Jr. 2005. Assessment of disorder predictions in CASP6. *Proteins.* 61(Suppl 7):167–175.

78. Ferron, F., S. Longhi, B. Canard, and D. Karlin. 2006. A practical overview of protein disorder prediction methods. *Proteins.* 65:1–14.

79. Peng, K., S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, and Z. Obradovic. 2005. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol.* 3:35–60.

80. Obradovic, Z., K. Peng, S. Vucetic, P. Radivojac, and A. K. Dunker. 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins.* 61(Suppl 7):176–182.

81. Peng, K., P. Radivojac, S. Vucetic, A. K. Dunker, and Z. Obradovic. 2006. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics.* 7:208.

82. Jones, D. T., and J. J. Ward. 2003. Prediction of disordered regions in proteins from position specific score matrices. *Proteins.* 53:573–578.

83. Ward, J. J., J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337:635–645.

84. Ward, J. J., L. J. McGuffin, K. Bryson, B. F. Buxton, and D. T. Jones. 2004. The DISOPRED server for the prediction of protein disorder. *Bioinformatics.* 20:2138–2139.

85. Linding, R., R. B. Russell, V. Neduva, and T. J. Gibson. 2003. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31:3701–3708.

86. Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon. 2005. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* 347:827–839.

87. Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 21:3433–3434.

88. Yang, Z. R., R. Thomson, P. McNeil, and R. M. Esnouf. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics.* 21:3369–3376.

89. Coeytaux, K., and A. Poupon. 2005. Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics.* 21:1891–1900.

90. Cheng, J., M. J. Sweredoski, and P. Baldi. 2005. Accurate prediction of protein disordered regions by mining protein structure data. *Data Mining Knowl. Disc.* 11:213–222.

91. Vullo, A., O. Bortolami, G. Pollastri, and S. C. E. Tossato. 2006. Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. *Nucleic Acids Res.* 34(Web server issue):W164–168.

92. Gu, J., M. Gribskov, and P. E. Bourne. 2006. Wiggle-predicting functionally flexible regions from primary sequence. *PLoS Comput. Biol.* 2:e90.

93. Gunasekaran, K., C. J. Tsai, and R. Nussinov. 2004. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J. Mol. Biol.* 341:1327–1341.

94. Mohan, A., C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, and V. N. Uversky. 2006. Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.* 362:1043–1059.

95. Oldfield, C. J., Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, and A. K. Dunker. 2005. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry.* 44:12454–12470.

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

97. Romero, P., Z. Obradovic, and A. K. Dunker. 2000. Intelligent data analysis for protein disorder prediction. *Artif. Intel. Rev.* 14:447–484.

98. Bairoch, A., R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. S. Yeh. 2005. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 33:D154–D159.

99. Garner, E., P. Cannon, P. Romero, Z. Obradovic, and A. K. Dunker. 1999. Predicting disordered regions in protein from amino acid sequence: common themes despite differing structural characterization. *Genome Inform. Ser. Workshop Genome Inform.* 9:201–213.

100. Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25:25–29.

101. Xie, H., S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic, and V.N. Uversky. 2007. Functional anthology of intrinsic disorder. I. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* In press.

102. Xie, H., S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic, and V.N. Uversky. 2007. Functional anthology of intrinsic disorder. II. Cellular components, domains, technical terms, developmental processes and coding sequence diversities correlated with long disordered regions. *J. Proteome Res.* In press.

103. Xie, H., S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic, and V.N. Uversky. 2007. Functional anthology of intrinsic disorder. III. Ligands, posttranslational modifications and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* In press.

104. Garner, E., P. Romero, A. K. Dunker, C. Brown, and Z. Obradovic. 1999. Predicting binding regions within disordered proteins. *Genome Inform. Ser. Workshop Genome Inform.* 10:41–50.

105. Fletcher, C. M., and G. Wagner. 1998. The interaction of eIF4E with 4E–BP1 is an induced fit to a completely disordered protein. *Protein Sci.* 7:1639–1642.

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

107. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338:1015–1026.

108. Bienkiewicz, E. A., J. N. Adkins, and K. J. Lumb. 2002. Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). *Biochemistry.* 41:752–759.

109. Lacy, E. R., I. Filippov, W. S. Lewis, S. Otieno, L. Xiao, S. Weiss, L. Hengst, and R. W. Kriwacki. 2004. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol.* 11:358–364.

110. Lee, H., K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi, and K. H. Han. 2000. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem.* 275:29426–29432.

111. Longhi, S., V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard. 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.* 278:18638–18648.

112. Callaghan, A. J., J. P. Aurikko, L. L. Ilag, J. Gunter Grossmann, V. Chandran, K. Kuhnel, L. Poljak, A. J. Carpousis, C. V. Robinson, M. F. Symmons, and B. F. Luisi. 2004. Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *J. Mol. Biol.* 340:965–979.

113. Bourhis, J. M., K. Johansson, V. Receveur-Brechot, C. J. Oldfield, K. A. Dunker, B. Canard, and S. Longhi. 2004. The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res.* 99:157–167.

114. Vacic, V., C.J. Oldfield, A. Mohan, P. Radivojac, M.S. Cortese, V.N. Uversky, and A.K. Dunker. 2007. Characterization of molecular recognition features, MoRFs, and MoRF-binding proteins. *J. Mol. Biol.* In press.

115. Chin, D., and A. R. Means. 2000. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* 10:322–328.

116. Van Eldik, L. J., and D. M. Watterson, editors. 1998. Calmodulin and Signal Transduction. Academic Press, San Diego, CA.

117. Stull, J. T. 2001. Ca^{2+}-dependent cell signaling through calmodulin-activated protein phosphatase and protein kinases. Minireview series. *J. Biol. Chem.* 276:2311–2312.

118. Vetter, S. W., and E. Leclerc. 2003. Novel aspects of calmodulin target recognition and activation. *Eur. J. Biochem.* 270:404–414.

119. Yap, K., J. Kim, K. Truong, M. Sherman, T. Yuan, and M. Ikura. 2000. Calmodulin target database. *J. Struct. Funct. Genomics.* 1:8–14.

120. James, P., T. Vorherr, and E. Carafoli. 1995. Calmodulin-binding domains: just two faced or multi-faceted? *Trends Biochem. Sci.* 20:38–42.

121. Yuan, T., H. J. Vogel, C. Sutherland, and M. P. Walsh. 1998. Characterization of the Ca^{2+}-dependent and -independent interactions between calmodulin and its binding domain of inducible nitric oxide synthase. *FEBS Lett.* 431:210–214.

122. Radivojac, P., S. Vucetic, T. R. O'Connor, V. N. Uversky, Z. Obradovic, and A. K. Dunker. 2006. Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. *Proteins.* 63:398–410.

123. Iakoucheva, L. M., P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, and A. K. Dunker. 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 32:1037–1049.

124. Daily, K. M., P. Radivojac, and A. K. Dunker. 2005. Intrinsic disorder and protein modifications: building an SVM predictor for methylation. *In* IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2005. San Diego, CA, 475–481.

125. Radivojac, P., N. V. Chawla, A. K. Dunker, and Z. Obradovic. 2004. Classification and knowledge discovery in protein databases. *J. Biomed. Inform.* 37:224–239.

126. Beltrao, P., and L. Serrano. 2005. Comparative genomics and disorder prediction identify biologically relevant SH3 protein interactions. *PLoS Comput. Biol.* 1:e26.

127. Puntervoll, P., R. Linding, C. Gemund, S. Chabanis-Davidson, M. Mattingsdal, S. Cameron, D. M. Martin, G. Ausiello, B. Brannetti, A. Costantini, F. Ferre, V. Maselli, A. Via, G. Cesareni, F. Diella, G. Superti-Furga, L. Wyrwicz, C. Ramu, C. McGuigan, R. Gudavalli, I. Letunic, P. Bork, L. Rychlewski, B. Kuster, M. Helmer-Citterich, W. N. Hunter, R. Aasland, and T. J. Gibson. 2003. ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* 31:3625–3630.

128. Goh, C. S., N. Lan, S. M. Douglas, B. Wu, N. Echols, A. Smith, D. Milburn, G. T. Montelione, H. Zhao, and M. Gerstein. 2004. Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. *J. Mol. Biol.* 336:115–130.

129. Adams, M., A. Joachimiak, R. Kim, G. T. Montelione, and J. Norvell. 2004. Meeting review: 2003 NIH protein structure initiative workshop in protein production and crystallization for structural and functional genomics. *J. Struct. Funct. Genomics.* 5:1–2.

130. Chandonia, J. M., S. H. Kim, and S. E. Brenner. 2006. Target selection and deselection at the Berkeley Structural Genomics Center. *Proteins.* 62:356–370.

131. Oldfield, C. J., E. L. Ulrich, Y. Cheng, A. K. Dunker, and J. L. Markley. 2005. Addressing the intrinsic disorder bottleneck in structural proteomics. *Proteins.* 59:444–453.

132. Bandaru, V., W. Cooper, S. S. Wallace, and S. Doublie. 2004. Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that recognizes oxidative DNA damage. *Acta Crystallogr. D Biol. Crystallogr.* 60:1142–1144.

133. Idicula-Thomas, S., A. J. Kulkarni, B. D. Kulkarni, V. K. Jayaraman, and P. V. Balaji. 2006. A support vector machine-based method for

predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in *Escherichia coli*. *Bioinformatics*. 22:278–284.

134. Smialowski, P., T. Schmidt, J. Cox, A. Kirschner, and D. Frishman. 2006. Will my protein crystallize? A sequence-based predictor. *Proteins*. 62:343–355.

135. Ferron, F., S. Longhi, B. Henrissat, and B. Canard. 2002. Viral RNA-polymerases–a predicted 2′-O-ribose methyltransferase domain shared by all *Mononegavirales*. *Trends Biochem. Sci.* 27:222–224.

136. Karlin, D., F. Ferron, B. Canard, and S. Longhi. 2003. Structural disorder and modular organization in *Paramyxovirinae* N and P. *J. Gen. Virol.* 84:3239–3252.

137. Ferron, F., C. Rancurel, S. Longhi, C. Cambillau, B. Henrissat, and B. Canard. 2005. VaZyMolO: a tool to define and classify modularity in viral proteins. *J. Gen. Virol.* 86:743–749.

138. Iakoucheva, L. M., A. L. Kimzey, C. D. Masselon, J. E. Bruce, E. C. Garner, C. J. Brown, A. K. Dunker, R. D. Smith, and E. J. Ackerman. 2001. Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.* 10:560–571.

139. Adkins, J. N., and K. J. Lumb. 2002. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57(Kip2). *Proteins*. 46:1–7.

140. Chang, B. S., A. J. Minn, S. W. Muchmore, S. W. Fesik, and C. B. Thompson. 1997. Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2. *EMBO J.* 16:968–977.

141. Campbell, K. M., A. R. Terrell, P. J. Laybourn, and K. J. Lumb. 2000. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. *Biochemistry*. 39:2708–2713.

142. Sunde, M., K. C. McGrath, L. Young, J. M. Matthews, E. L. Chua, J. P. Mackay, and A. K. Death. 2004. TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer. *Cancer Res.* 64:2766–2773.

143. Uversky, V. N., A. Roman, C. J. Oldfield, and A. K. Dunker. 2006. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J. Proteome Res.* 5:1829–1842.

144. zur Hausen, H. 2002. Papillomaviruses and cancer: from basic studies to clinical application. *Nat. Rev. Cancer*. 2:342–350.

145. Longworth, M. S., and L. A. Laimins. 2004. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol. Mol. Biol. Rev.* 68:362–372.

146. Munger, K., A. Baldwin, K. M. Edwards, H. Hayakawa, C. L. Nguyen, M. Owens, M. Grace, and K. Huh. 2004. Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol.* 78:11451–11460.

147. Munoz, N., F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders, and C. J. Meijer. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* 348:518–527.

148. Tommasino, M., and L. Crawford. 1995. Human papillomavirus E6 and E7: proteins that deregulate the cell cycle. *Bioessays*. 17:509–518.

149. Ohlenschlager, O., T. Seiboth, H. Zengerling, L. Briese, A. Marchanka, R. Ramachandran, M. Baum, M. Korbas, W. Meyer-Klaucke, M. Durst, and M. Gorlach. 2006. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene*. 25:5953–5959.

150. Cheng, Y., T. LeGall, C. J. Oldfield, A. K. Dunker, and V. N. Uversky. 2006. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry*. 45:10448–10460.

151. Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*. 234:364–368.

152. Rechsteiner, M., and S. W. Rogers. 1996. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* 21:267–271.

153. Yaglom, J., M. H. Linskens, S. Sadis, D. M. Rubin, B. Futcher, and D. Finley. 1995. p34Cdc28-mediated control of Cln3 cyclin degradation. *Mol. Cell. Biol.* 15:731–741.

154. Berset, C., P. Griac, R. Tempel, J. La Rue, C. Wittenberg, and S. Lanker. 2002. Transferable domain in the G(1) cyclin Cln2 sufficient to switch degradation of Sic1 from the E3 ubiquitin ligase SCF(Cdc4) to SCF(Grr1). *Mol. Cell. Biol.* 22:4463–4476.

155. Bordone, L., and C. Campbell. 2002. DNA ligase III is degraded by calpain during cell death induced by DNA-damaging agents. *J. Biol. Chem.* 277:26673–26680.

156. Gregory, M. A., and S. R. Hann. 2000. c-Myc proteolysis by the ubiquitin-proteasome pathway: stabilization of c-Myc in Burkitt's lymphoma cells. *Mol. Cell. Biol.* 20:2423–2435.

157. Singh, G. P., M. Ganapathi, K. S. Sandhu, and D. Dash. 2006. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins*. 62:309–315.

158. Babon, J. J., E. J. McManus, S. Yao, D. P. DeSouza, L. A. Mielke, N. S. Sprigg, T. A. Willson, D. J. Hilton, N. A. Nicola, M. Baca, S. E. Nicholson, and R. S. Norton. 2006. The structure of SOCS3 reveals the basis of the extended SH2 domain function and identifies an unstructured insertion that regulates stability. *Mol. Cell.* 22:205–216.

159. Kalderon, D., B. L. Roberts, W. D. Richardson, and A. E. Smith. 1984. A short amino acid sequence able to specify nuclear location. *Cell*. 39:499–509.

160. Dingwall, C., and R. A. Laskey. 1991. Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.* 16:478–481.

161. Conti, E., and E. Izaurralde. 2001. Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.* 13:310–319.

162. Mosammaparast, N., and L. F. Pemberton. 2004. Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends Cell Biol.* 14:547–556.

163. Lee, B. J., A. E. Cansizoglu, K. E. Suel, T. H. Louis, Z. Zhang, and Y. M. Chook. 2006. Rules for nuclear localization sequence recognition by karyopherin $\beta 2$. *Cell*. 126:543–558.

164. Breman, J. G. 2001. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* 64:1–11.

165. Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 419:498–511.

166. Feng, Z. P., X. Zhang, P. Han, N. Arora, R. F. Anders, and R. S. Norton. 2006. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. *Mol. Biochem. Parasitol.* 150:256–267.

167. Bezanilla, F. 2000. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* 80:555–592.

168. Sigworth, F. J. 1994. Voltage gating of ion channels. *Q. Rev. Biophys.* 27:1–40.

169. Yellen, G. 1998. The moving parts of voltage-gated ion channels. *Q. Rev. Biophys.* 31:239–295.

170. Armstrong, C. M., and F. Bezanilla. 1977. Inactivation of the sodium channel. II. Gating current experiments. *J. Gen. Physiol.* 70:567–590.

171. Bezanilla, F., and C. M. Armstrong. 1977. Inactivation of the sodium channel. I. Sodium current experiments. *J. Gen. Physiol.* 70:549–566.

172. Antz, C., M. Geyer, B. Fakler, M. K. Schott, H. R. Guy, R. Frank, J. P. Ruppersberg, and H. R. Kalbitzer. 1997. NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. *Nature*. 385:272–275.

173. Bentrop, D., M. Beyermann, R. Wissmann, and B. Fakler. 2001. NMR structure of the ''ball-and-chain'' domain of KCNMB2, the $\beta 2$-subunit of large conductance Ca^{2+}- and voltage-activated potassium channels. *J. Biol. Chem.* 276:42116–42121.

174. Hoshi, T., W. N. Zagotta, and R. W. Aldrich. 1990. Biophysical and molecular mechanisms of *Shaker* potassium channel inactivation. *Science*. 250:533–538.

175. Zagotta, W. N., T. Hoshi, and R. W. Aldrich. 1990. Restoration of inactivation in mutants of *Shaker* potassium channels by a peptide derived from ShB. *Science*. 250:568–571.

176. Magidovich, E., S. J. Fleishman, and O. Yifrach. 2006. Intrinsically disordered C-terminal segments of voltage-activated potassium channels: a possible fishing rod-like mechanism for channel binding to scaffold proteins. *Bioinformatics*. 22:1546–1550.

177. Wolffe, A. 1998. Chromatin: Structure and Function. Academic Press, New York.

178. Hansen, J. C. 2002. Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu. Rev. Biophys. Biomol. Struct.* 31:361–392.

179. Arents, G., and E. N. Moudrianakis. 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. USA*. 92:11170–11174.

180. Bustin, M., F. Catez, and J. H. Lim. 2005. The dynamics of histone H1 function in chromatin. *Mol. Cell*. 17:617–620.

181. van Holde, K. 1988. Chromatin. Springer-Verlag, New York.

182. Munishkina, L. A., A. L. Fink, and V. N. Uversky. 2004. Conformational prerequisites for formation of amyloid fibrils from histones. *J. Mol. Biol.* 342:1305–1324.

183. Hansen, J. C., X. Lu, E. D. Ross, and R. W. Woody. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.* 281:1853–1856.

184. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.

185. Haynes, C., C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, and L. M. Iakoucheva. 2006. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.* 2:e100.

186. Ekman, D., S. Light, A. K. Bjorklund, and A. Elofsson. 2006. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol.* 7:R45.

187. Patil, A., and H. Nakamura. 2006. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.* 580:2041–2045.

188. Dosztányi, Z., J. Chen, A. K. Dunker, I. Simon, and P. Tompa. 2006. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* 5:2985–2995.

189. Haynes, C., and L. M. Iakoucheva. 2006. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.* 34:305–312.

190. Bustos, D. M., and A. A. Iglesias. 2006. Intrinsic disorder is a key characteristic in partners that bind 14–3-3 proteins. *Proteins*. 63:35–42.

191. Rittinger, K., J. Budman, J. Xu, S. Volinia, L. C. Cantley, S. J. Smerdon, S. J. Gamblin, and M. B. Yaffe. 1999. Structural analysis of 14–3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14–3-3 in ligand binding. *Mol. Cell*. 4:153–166.

192. Spolar, R. S., and M. T. Record II. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science*. 263:777–784.

193. Kalodimos, C. G., N. Biris, A. M. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, and R. Kaptein. 2004. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*. 305:386–389.

194. Dyson, H. J., and P. E. Wright. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12:54–60.

195. Liu, J., N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker. 2006. Intrinsic disorder in transcription factors. *Biochemistry*. 45:6873–6888.

196. Minezaki, Y., K. Homma, A. R. Kinjo, and K. Nishikawa. 2006. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359:1137–1149.

197. Ritter, L. M., T. Arakawa, and A. F. Goldberg. 2005. Predicted and measured disorder in peripherin/RDS, a retinal tetraspanin. *Protein Pept. Lett.* 12:677–686.

198. Kukhtina, V., D. Kottwitz, H. Strauss, B. Heise, N. Chebotareva, V. Tsetlin, and F. Hucho. 2006. Intracellular domain of nicotinic acetylcholine receptor: the importance of being unfolded. *J. Neurochem.* 97:S63–S67.

199. Yiu, C. P., R. L. Beavil, and H. Y. Chan. 2006. Biophysical characterization reveals structural disorder in the nucleolar protein, Dribble. *Biochem. Biophys. Res. Commun.* 343:311–318.

200. Hinds, M. G., C. Smits, R. Fredericks-Short, J. M. Risk, M. Bailey, D. C. Huang, and C. L. Day. 2007. Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets. *Cell Death Differ.* 14:128–136.

201. Nardini, M., D. Svergun, P. V. Konarev, S. Spano, M. Fasano, C. Bracco, A. Pesce, A. Donadini, C. Cericola, F. Secundo, A. Luini, D. Corda, and M. Bolognesi. 2006. The C-terminal domain of the transcriptional corepressor CtBP is intrinsically unstructured. *Protein Sci.* 15:1042–1050.

202. Loftus, S. R., D. Walker, M. J. Mate, D. A. Bonsor, R. James, G. R. Moore, and C. Kleanthous. 2006. Competitive recruitment of the periplasmic translocation portal TolB by a natively disordered domain of colicin E9. *Proc. Natl. Acad. Sci. USA*. 103:12353–12358.

203. Hoffman, R. M., T. M. Blumenschein, and B. D. Sykes. 2006. An interplay between protein disorder and structure confers the Ca^{2+} regulation of striated muscle. *J. Mol. Biol.* 361:625–633.

204. Keramisanou, D., N. Biris, I. Gelis, G. Sianidis, S. Karamanou, A. Economou, and C. G. Kalodimos. 2006. Disorder-order folding transitions underlie catalysis in the helicase motor of SecA. *Nat. Struct. Mol. Biol.* 13:594–602.

205. Roy, S., S. Schnell, and P. Radivojac. 2006. Unraveling the nature of the segmentation clock: intrinsic disorder of clock proteins and their interaction map. *Comput. Biol. Chem.* 30:241–248.

206. Dyson, H. J., and P. E. Wright. 2006. According to current textbooks, a well-defined three-dimensional structure is a prerequisite for the function of a protein. Is this correct? *IUBMB Life*. 58:107–109.

207. George, R. A., and J. Heringa. 2002. SnapDRAGON: a method to delineate protein structural domains from sequence data. *J. Mol. Biol.* 316:839–851.

208. Obenauer, J. C., L. C. Cantley, and M. B. Yaffe. 2003. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31:3635–3641.