Published in final edited form as:
J Proteome Res. 2007 May ; 6(5): 1917–1932. doi:10.1021/pr060394e.

Functional Anthology of Intrinsic Disorder. III. Ligands, Postranslational Modifications and Diseases Associated with Intrinsically Disordered Proteins

Hongbo Xie[†], **Slobodan Vucetic**[†], **Lilia M. Iakoucheva**[‡], **Christopher J. Oldfield**[#], **A. Keith Dunker**[#], **Zoran Obradovic**[†], and **Vladimir N. Uversky**[*,#,§]

†*Center for Information Science and Technology, Temple University, Philadelphia, PA 19122*

‡*Laboratory of Statistical Genetics, The Rockefeller University, New York, NY 10021*

#*Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University, School of Medicine, Indianapolis, IN 46202*

§*Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia*

Abstract

Currently, the understanding of the relationships between function, amino acid sequence and protein structure continues to represent one of the major challenges of the modern protein science. As much as 50% of eukaryotic proteins are likely to contain functionally important long disordered regions. Many proteins are wholly disordered but still possess numerous biologically important functions. However, the number of experimentally confirmed disordered proteins with known biological functions is substantially smaller than their actual number in nature. Therefore, there is a crucial need for novel bioinformatics approaches that allow projection of the current knowledge from a few experimentally verified examples to much larger groups of known and potential proteins. The elaboration of a bioinformatics tool for the analysis of functional diversity of intrinsically disordered proteins and application of this data mining tool to >200,000 proteins from Swiss-Prot database, each annotated with at least one of the 875 functional keywords was described in the first paper of this series (Xie H., Vucetic S., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Obradovic Z., Uversky V.N. (2006) Functional anthology of intrinsic disorder. I. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.*). Using this tool, we have found that out of the 711 Swiss-Prot functional keywords associated with at least 20 proteins, 262 were strongly positively correlated with long intrinsically disordered regions, and 302 were strongly negatively correlated. Illustrative examples of functional disorder or order were found for the vast majority of keywords showing strongest positive or negative correlation with intrinsic disorder, respectively. Some 80 Swiss-Prot keywords associated with disorder- and order-driven biological processes and protein functions were described in the first paper (Xie H., Vucetic S., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Obradovic Z., Uversky V.N. (2006) Functional anthology of intrinsic disorder. I. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.*). The second paper of the series was devoted to the presentation of 87 Swiss-Prot keywords attributed to the cellular components, domains, technical terms, developmental processes and coding sequence diversities possessing strong positive and negative correlation with long disordered regions (Vucetic S., Xie H., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Obradovic Z., Uversky V.N. (2006)

*Correspondence should be addressed to: Vladimir N. Uversky, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, MS#4021, Indianapolis, IN 46202, USA; Phone: 317-278-9194; Fax: 317-274-4686; E-mail: vuversky@iupui.edu.

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.*). Protein structure and functionality can be modulated by various posttranslational modifications or/and as a result of binding of specific ligands. Numerous human diseases are associated with protein misfolding/misassembly/ misfunctioning. This work concludes the series of papers dedicated to the functional anthology of intrinsic disorder and describes ~80 Swiss-Prot functional keywords that are related to ligands, posttranslational modifications and diseases possessing strong positive or negative correlation with the predicted long disordered regions in proteins.

Keywords

Intrinsic disorder; protein structure; protein function; intrinsically disordered proteins; bioinformatics; disorder prediction

Introduction

The functionalities of many proteins are modulated by various types of posttranslational modifications (PTMs). Chemical modifications of a polypeptide chain after its biosynthesis extends the range of functions of the protein. Typically, PTMs are classified according to the involved mechanisms: addition of functional groups (e.g., acylation, alkylation, phosphorylation, glucosylation, etc.); attachement of other proteins and peptides (e.g., ubiquitination, SUMOylation, etc.); changing of the chemical nature of amino acids (deamidation, deimidation, oxidation, etc.); and dissection of the backbone by proteolytic cleavage. Some proteins require different types of posttranslational modifications for their function. One dramatic example of such proteins is provided by histones, which require methylation, acetylation, phosphorylation, ubiquitylation, ADP-ribosyation, and SUMOylation at different stages, with different modifications affecting histone–DNA interactions and also the histone–histone interfaces, thus providing the capacity to disrupt intra-nucleosomal interactions and to alter nucleosome stability.[1] Although the N-terminal domains of the core histones are known to contain an extraordinary number of sites that can be subjected to PTM,[2–4] over 30 histone modifications have been recently identified in the core domains too.[5–7]

We are showing here that, in addition to the known classification based on the molecular mechanisms of PTMs, the conformational state of the site where the modification would take place can also be used to group PTMs in two major classes: the first group involves modifications that are associated primarily with structured proteins and regions, whereas the second group combines modifications that are associated primarily with intrinsically disordered proteins and regions. The first group of modifications is crucial for providing moieties for catalytic functions, for modifying enzyme activities or for stabilizing protein structure. They include formylation, protein splicing, oxidation and covalent attachment of quinones and organic radicals. The second group of modifications relies on the low affinity, high specificity binding interactions between a specific enzyme and a substrate (a protein that has to be modified); i.e., they clearly represent typical signaling interactions. Among the second group of PTMs are phosphorylation, acetylation, acylation, adenylylation, ADP ribosylation, amidation, carboxylation, formylation, glycosylation, methylation, sulfation, prenylation, ubiquitination, and Ubl-conjugation (i.e., covalent attachment of ubiquitin-like proteins, including SUMO, ISG15, Nedd8, and Atg8).

Similarly, natural ligands play a number of roles in the stabilization of proteins and in the modulation of their structures. In fact, during the course of their biological function, proteins

undergo different types of structural rearrangements ranging from local to large-scale conformational changes. These changes are often triggered by protein interactions with low-molecular-weight ligands or with larger macromolecules. The interactions with natural ligands can significantly affect protein structure, a phenomenon widely used by nature. Particularly, this phenomenon represents a basis for the functioning of several cofactors. These interactions often result not only in evident local changes in the vicinity of the binding site, but also in global conformational changes. The hemoglobin-oxygen complex is a textbook example of such systems.[8–11] As a result of O_2 binding, the Fe^{2+} is shifted into the plane of the porphirine ring approximately by 0.6 Å. This slightly moves the hystidine residue, which is involved in the coordination of the iron ion. In its turn, the consequence of such local changes is an essential conformational reorganization of the whole hemoglobin molecule, accompanied by a large-scale movement (10–15° rotation of the subunits relative to each other).[12, 13]

The structure forming and stabilizing effects of such natural ligands such as electrons, oxygen, water, metal cations, hem, fatty acids and other organic compounds were recently reviewed for different proteins.[14] The range of possible structural transformations induced in a protein by ligand release is very wide, from a negligible decrease in the conformational stability to a complete protein unfolding. Furthermore, these structural alterations virtually do not depend on the nature of ligand. Comparing the structural properties and the conformational stabilities of several proteins, ligand-free forms were divided on seven major classes:[14]

Class I. The structural characteristics of apo- and holo-forms coincide; the release of ligands does not lead to the noticeable changes of the unique protein structure; only the conformational stability and the function are changed.

Class II. A detectable change in the 3D-structure takes place in the ligand-free form. However, the protein is still characterized by a pronounced tertiary structure while its secondary structure virtually does not change.

Class III. The apo-form of the protein possesses all the properties of the molten globule. In other words, tertiary structure completely disappears upon release of the ligands, the secondary structure is retained, and the protein molecule remains compact.

Class IV. The ligand-free form of the protein has specific features of the pre-molten globule; it has no rigid tertiary structure, its secondary structure is noticeably diminished, and its dimensions distinctly differ both from the random coil and also from the holo-protein dimensions.

Class V. The apo-form adopts a virtually completely unfolded polypeptide chain, similar to so-called random coil.

Class VI. The interaction with ligands induces large-scale movements of large parts (domains or subunits) of the protein molecule. These effects are frequently observed for the allosteric enzymes.

Class VII. Ligand binding destabilizes native protein conformation.

In agreement with this classification, we show here that many of the Swiss-Prot functional keywords related to the ligand binding are positively and negatively correlated with long protein regions predicted to be intrinsically disordered. In other words, using the bioinformatics tool developed in the first paper of this series[15] some natural ligands are shown to interact preferentially with ordered proteins, whereas others prefer intrinsically disordered proteins.

Finally our analysis reveals that many diseases are strongly correlated with proteins predicted to be disordered. Contrary to this, we did not find disease associated proteins to be strongly correlated with absence of disorder. Below we describe some illustrative examples obtained from a literature survey. As previously, we were able to find at least one illustrative, experimentally validated example of functional disorder or order for the vast majority of functional keywords related to PTMs, ligand binding and diseases.

Materials and methods

The first paper of this series described the assembly of the dataset for the bioinformatics analysis to find correlation between long disordered regions and Swiss-Prot keywords.[15] A bioinformatics approach was also presented that determines which functional keywords are over- or under-represented by proteins predicted to contain long (>40 consecutive amino acids) disordered regions. Using this approach, 196,326 Swiss-Prot[16] proteins longer than 40 amino acid residues have been analyzed. The redundancy of Swiss-Prot database[17] was reduced by the Markov Cluster Algorithm,[18] which was used to group Swiss-Prot proteins into 27,217 families according to sequence similarity. Each of the analyzed proteins was annotated with at least one of the 875 functional or structural Swiss-Prot keywords.

Long disordered regions in Swiss-Prot proteins were predicted using the VL3E predictor[19]. Each of the 196,326 Swiss-Prot proteins was labeled as putatively disordered if it contained a region with more than 40 consecutive amino acids predicted by VL3E to be intrinsically disordered; proteins predicted not to contain such long disordered regions were labeled as putatively ordered.

The probability, P_L, that VL3E predicts a disordered region longer than 40 consecutive amino acids in a SwissProt protein sequence of length L was estimated as the fraction of putatively disordered SwissProt proteins with lengths between 0.9L and 1.1L. TribeMCL clustering was used to reduce effects of sequence redundancy in estimation of P_L, as described previously.[15] Swiss-Prot keywords associated with disorder-(or order-) correlated functions were determined as those that contain a significantly larger (or smaller) fraction of putatively disordered proteins than what would be expected by a random selection of SwissProt sequences with the same length distribution.[15]

Results

Ligands interacting with intrinsically disordered proteins

DNA- and RNA-binding proteins—A list of keywords associated with ligands interacting with intrinsically disordered proteins is present in Table 1. One can see that the majority of ligands listed in Table 1 are related to the functionality of proteins involved in signaling, regulation and recognition. In fact, we have already pointed out that intrinsic disorder is important for functionality of different *DNA*- and *RNA-binding* proteins. The previously discussed examples include transcription factors, histones, and other proteins involved in the DNA condensation and chromosome partition.[15] Similarly, ribonucleoproteins and proteins involved in *rRNA-binding* such as ribosomal proteins were shown to possess significant intrinsic disorder.[15] Recently it has been established that serine/arginine-rich (SR) splicing factors belong to a class of intrinsically disordered proteins.[20] Several translation factors are involved in multiple interactions including *rRNA*-, *mRNA*-, and *tRNA-binding*. Analysis of the solution structure of the *Methanobacterium thermoautotrophicum* aIF2 beta, which is the archaeal homolog of eIF2 beta, a member of the initiation factor eIF2 heterotrimeric complex, implicated in the delivery of Met-tRNA(i)(Met) to the 40S ribosomal subunit, revealed that this translation factor subunit is composed of an unfolded N terminus, a mixed alpha/beta core domain and a C-terminal zinc finger.[21] Also, intrinsic disorder is crucial for many *viral*

nucleoproteins, which are responsible for the encapsulation of genomic RNA within a helical nucleocapsid. Indeed, the C-terminal domain of the nucleoprotein (N_{tail}) of measles virus, which is involved in a number important functions, is intrinsically unstructured.[22–25]

Metal-thiolate clusters—Several proteins known as metallothioneins rely on *metal-thiolate clusters* in their functions. There are at least ten known closely related metallothionein proteins expressed in humans. Metallothioneins are small proteins (< 7 kDa), that are able to coordinate a diverse range of metals. These proteins are characterized by a lack of definable secondary structure, a high cysteine content (~30%), and a degeneracy in the remaining residues (e.g. predominance of cysteine, serine, lysine and no aromatic residues).[26] The folding of metallothioneins into the functional conformation is completely determined by the coordination of the corresponding metal ions.[27] The mammalian metallothioneins make economic use of the protein synthesis machinery, as the protein chain is just long enough to wrap around the mineral cores, and as indicated above there is hardly any secondary structure.[28] Finally, the role of *zinc* binding in the folding of the intrinsically disordered zinc finger domains, which are the members of the metallothionein family, and which are primarily implicated in DNA binding, was already emphasized.[15, 29]

Actin-binding proteins—Thymosin β4 is a small 5-kDa *actin-binding* protein with a diverse range of activities, including its function as an actin monomer sequestering protein, an antiinflammatory agent, and an inhibitor of bone marrow stem cell proliferation. Thymosin β4 is a typical natively unfolded protein with extremely low level of ordered structure in solution.[30] It is believed that the flexible structure of thymosin β4 may facilitate the recognition of a variety of molecular targets, thus explaining numerous functions attributed to this interesting protein. Furthermore, it has been hypothesized that thymosin β4 has a unique integrative function that links the actin cytoskeleton to important immune and cell growth signaling cascades.[31] Troponin, another *actin-binding* protein, is a part of the native tropomyosin, a complex located on actin filaments and involved in the contraction of striated muscle. Troponin consists of three components, each performing specific functions: troponin C binds Ca^{2+}, troponin I inhibits the ATPase activity of actomyosin, and troponin T provides for the binding of troponin to tropomyosin.[32] Intrinsic disorder is abundant in troponins. In fact, troponin C possesses a relatively flexible linker providing high relative mobility between the two globular domains.[33] Troponin I, a polar protein with a high excess of positively charged residues (calculated pI is ~9.9), is highly extended in its functional state.[34] Finally, troponin T is also a highly polar protein but its charged residues are unevenly distributed within the amino acid sequence: the N-terminal part including residues 1–59 is enriched in negatively charged residues, whereas the C-terminal part is enriched in positively charged residues.[32] Similarly to troponin I, troponoin T has an extended, rod-like structure even being involved in the tertiary complex with troponins C and I.[34]

Calmodulin-binding—Table 1 shows the Swiss-Prot keyword *calmodulin-binding* to be strongly correlated with predicted disorder. Calmodulin (CaM) binding is involved in various important signaling pathways.[35–37] Numerous structural studies revealed that CaM target proteins interact with CaM in a common manner.[38–40] In fact, the interactions between CaM and its binding targets involve disorder-to-order transitions for the CaM molecule: a flexible linker, which becomes structured upon complex formation, enables the two globular domains to wrap around the CaM binding target.[38–40] The helix-helix interactions within the two globular domains of CaM are not completely rigid, so the helix-helix packing interfaces in these domains vary in a manner that depends on the detailed interactions with the different CaM binding targets.[37, 41] Finally, the CaM target-binding surface is rich in methionines, which adopt different configurations when CaM associates with CaM binding target having different sequences. The end result of this structural plasticity is that CaM binds to a very large

number of different sequences with high affinity. Furthermore, recent studies revealed that intrinsic disorder is a crucial feature of the CaM-binding targets and can be used to improve the accuracy of the calmodulin binding target predictions.[35]

Heparin-binding proteins—Many proteins with widely diverse structures and functions are capable of binding heparin.[42] Analysis of *heparin-binding* sites (HBSs) revealed that they are often disordered. For example, there are two HBSs in the anticoagulant annexin V: HBS-1 is formed by two of the calcium-binding loops, and HBS-2 includes the N terminus and nearby loop and helix regions.[43] The conformational behavior and fibrillation of a Parkinson's disease related natively unfolded protein α-synuclein[44] and Alzheimer's disease-related natively denatured protein tau[45, 46] are both strongly modulated by heparin binding.

Growth factor binding—Insulin-like *growth factor binding* proteins (IGFBPs) are carriers and regulators of the insulin-like growth factors. Analysis of the solution structure of the C-terminal domain of IGFBP-6 (residues 161–240) revealed that it has substantial flexible regions (e.g., three hightly disordered loops), which include 33 of 79 residues.[47]

cGMP and cGMP-binding—Cyclic GMP (cGMP) is a second messenger that regulates various metabolic processes at the cellular level and that mediates the action of certain hormones. cGMP is synthesized from GTP by guanylate cyclase (GC) and is degraded to 5'-GMP by cyclic nucleotide phosphodiesterases (PDEs 1-6). *cGMP-binding* PDE from retinal rods plays a key role in the process of visual signal transduction. The PDE holoenzyme is a functionally inactive heterotetramer of $\alpha\beta\gamma_2$ composition. The catalytically active dimer of two large homologous α- and β-subunits is reversibly inhibited by the small γ-subunit, PDEγ.[48] It has been established that PDEγ has little or no ordered secondary structure under physiological conditions *in vitro*, being characterized by hydrodynamic dimensions typical for the completely unfolded polypeptide of corresponding molecular mass. This suggests that PDEγ belongs to the family of natively unfolded proteins.[49]

Sialic acids—The members of this family of acidic amino carbohydrates are derived from a nine-carbon monosaccharide and are components of *mucoproteins* and *glycoproteins*. Naturally occurring sialic acids possess large structural diversity, which explains their involvement in a variety of biologically important processes. The analysis of solution structure of ovine submaxillary mucin (OSM) by NMR spectroscopy revealed that the mucin is characterized by high internal segmental flexibility and it exists in solution as a random coil.[50] Similarly, bovine submaxillary mucin (BSM) was shown to be significantly disordered, as its far-UV CD spectrum is typical of a highly unfolded polypeptide chain.[51]

cAMP and cAMP-binding—Cyclic AMP (*cAMP*) is another second messenger used for intracellular signal transduction. cAMP is produced from ATP by adenylate cyclase and is decomposed into AMP by the enzyme phosphodiesterase. cAMP is known to act through cAMP-dependent protein kinase A (also known as cAPK or PKA), and also able to directly regulate ion channels and the Rap GTPase guanine exchange factors Epac 1 and 2. The cAMP-protein kinase A (PKA) pathway is an important intracellular signal transduction cascade activated by various stimuli. Activation/inhibition of this pathway is known to affect the *transcriptional regulation* of various genes through distinct responsive sites. In vertebrates, the best-characterized nuclear targets of PKA are the cAMP response element-binding (CREB) proteins. One of the CREB proteins is the activating transcription factor 2 (ATF-2).[52] Solution structure analysis of the N-terminal transactivation domain of ATF-2 revealed that the in the absence of zinc this domain is completely disordered.[53] However, addition of zinc induced partial folding of the ATF-2 transactivation domain.[53] Furthermore, the N-subdomain (residues 19–54) is well-folded, whereas the C-subdomain (residues 55–106) is highly flexible

and disordered.[53] Other illustrative examples of *cAMP-binding* proteins are the Epac1 and 2 (exchange protein directly activated by cAMP) protein, which are the guanine nucleotide-exchange factors (GEFs) that activate Rap GTPase upon binding to cAMP.[54] Rap1 is a small GTPase which serves as a hub protein involved in numerous aspects of cell adhesion, including integrin-mediated cell adhesion and cadherin-mediated cell junction formation.[55] The roles of intrinsic disorder in the functions of hub proteins have been recently recognized and supported by the additional bioinformatics analysis.[56, 57]

IgG-binding protein—Some strains of the anaerobic bacteria *Peptostreptococcus magnus* express a multidomain *IgG-binding* protein L at their surface. Protein L interacts with variable IgG light chain domain through the five homologous repeats (B1-B5) located in the N-terminal part of the protein. The 3-D solution structure of the 76 amino acid residue long B1 domain was analyzed using NMR spectroscopy. The domain was shown to contain a 15 amino acid residue long disordered N-terminus followed by a folded portion consisting of an α-helix packed against a four-stranded β-sheet.[58]

IgE-binding protein—Galectins are members of the lectin family of carbohydrate-binding proteins. So far 14 mammalian galectins have been identified, all containing a conserved carbohydrate-recognition-binding domain (CRD) of about 130 amino acids. The members of galectin family have been classified into three subtypes: the prototype group (galectins-1, -2, -5, -7, -10, -11, -13, and -14), which contains one CRD; the chimera group (galectin-3), that contains one CRD and a long N-terminal proline- and glycine-rich domain; and the tandem repeat group (galectins-4, -6, -8, -9, and -12), the members of which have two CRDs separated by an non-conserved linker sequence of up to 70 amino acids.[59] The vertebrate galectins were found in the cytoplasm and the nucleus, on the cell surface, and in the extracellular space. They are present in numerous cell and tissue types, and possess various functions.[60] Galectin-3 is a 29-to 35-kDa beta-galactoside binding protein containing a C-terminal CRD and a flexible N-terminal domain which is composed of 110–130 amino acids and contains multiple homologues repeats (7–14) with a consensus sequence Pro–Gly–Ala–Tyr–Pro–Gly, followed by three additional amino acids.[60] This protein is involved in mRNA splicing activity, control of the cell cycle, regulation of cell adhesion, the modulation of allergic reactions, and the binding of the advanced glycosylation end products (AGE). These multiple functions are exerted intra- and extracellularly due to the ability of galectin-3 to bind the β-galactoside residues of cell surface and matrix glycoproteins via the CRD domain and to also interact with intracellular proteins via peptide–peptide associations mediated by its N-terminus domain.[61] Among numerous functions of the extracellular galectin-3 is its *IgE binding* and interaction with the IgE receptor, which provide the molecular means to accomplish the modulation of inflammation.[62] Importantly, NMR analysis of hamster galectin-3 (residues 1–245) revealed that its N-terminus (residues 1–125) is mostly unfolded in aqueous media.[63]

tRNA-binding—Ribosomal protein L16 is an essential component of the bacterial ribosome organizing the architecture of the aminoacyl *tRNA-binding* site in the ribosome 50 S subunit. Although in solution L16 forms an α+β sandwich structure combined with two additional β-sheets located at the loop regions connecting the two layers, its terminal regions, Met1-Asp25 and Asp138-Gln141, and the internal loop region Thr75-Glu91 do not possess any ordered structure.[64]

Ligands interacting with ordered proteins

***Iron* homeostasis, *heme* and *iron-sulfur* or *4Fe-4S* clusters—**Table 2 lists keywords related to ligands interacting with ordered proteins. Strikingly, most of the ligands and cofactors listed in the Table 2 are related to enzyme function or to the oxygen/electron transport. These and other cofactors and ligands are recruited by protein molecules to extend the chemistry

available within the active sites. For example, besides the well-known *heme*-containing proteins (cytochromes, myoglobin and hemoglobin), transferrins, lactoferrin, ceruloplasmin and ferritins, numerous other proteins are involved in *iron* homeostasis.[65] The vast majority of these iron-binding proteins are known to possess well-organized rigid structure. Similarly, *iron-sulfur* or *4Fe-4S* cluster containing proteins are almost always highly organized, as these clusters are perhaps the most abundant and the most diversely employed cofactor.[66] The simplest iron-sulfur center is comprised of a single iron atom liganded within a polypeptide by four cysteine residues, whereas more common iron-sulfur clusters have two, three or four iron atoms coordinated to polypeptide residues and bridged by inorganic sulfide. Iron-sulfur clusters serve most prominently in redox centers involved in the electron transfer reactions, as well as in several dehydratases, biotin synthase and lipoate synthase.[66]

Flavoproteins—More than 25 *flavoproteins* (mostly enzymes, including different oxidases, oxidocyclases, dehydrogenases, methylhydroxylases and methylenehydroxylases) have been identified, in which the flavin adenine dinucleotide (*FAD*) or the flavin mononucleotide (*FMN*) is employed as a covalently bound cofactor.[67] Furthermore, many enzymes (e.g., different dehydrogenases, reductases, oxidases, etc.) use nicotinamide adenine dinucleotide (*NAD*) and nicotinamide adenine dinucleotide phosphate (*NADP*) as their cofactors.

Pyridoxal phosphate—Like the iron-sulfur clusters, the *pyridoxal phosphate* cofactor is very widely used. Indeed, the *pyridoxal phosphate* (*PLP*)-dependent enzymes catalyze a wider variety of reactions than those containing any other cofactor. These PLP-dependent enzymes include different amino-acid decarboxylases and aminotransferases, serine hydroxymethyltransferase, tryptophan synthase, glycogen phosphorylase etc.[68]

Magnesium—Enzymes belonging to the enolase superfamily are characterized by the highly conserved active site carboxylate residues that bind an essential *magnesium* ion and mediate proton transfer reactions from the carbon acid substrate to the resulting enolate ion intermediate.[69, 70] Structures were determined for members of the superfamily that catalyze eight different reactions, including enolase, mandelate racemase, muconate lactonizing enzymes I and II, D-glucarate dehydratase, D-galactonate dehydratase, *o*-succinylbenzoate synthases, L-Ala-D/L-Glu epimerases, and 3-methylaspartate ammonia lyase.[69, 70]

Plastoquinone—Finally, *plastoquinone* is frequently found in proteinaceous complexes associated with the electron transport chain in the light-dependent reactions of photosynthesis. For example, cytochrome b_6f complex from the thermophilic cyanobacterium *Mastigocladus laminosus* and the green alga *Chlamydomonas reinhardtii* is a highly organized machine, which also includes such prosthetic groups as hemes (*c*-type, *b*-type, and a new *x*-type heme), [2Fe-2S] cluster, chlorophyll *a*, β-carotene and plastoquinone.[71]

Thiamine pyrophosphate (TPP)—TPP is the derivative of vitamin B1, which is a cofactor of different enzymes performing catalysis in pathways of energy production. High resolution crystal structures have been determined for several TPP-dependent enzymes, including 2-oxoisovalerate dehydrogenase,[72] branched-chain α-ketoacid dehydrogenase,[73] bacterial[74] and human[75] pyruvate dehydrogenases, transketolase,[76] pyruvate decarboxylase,[77] benzoylformate decarboxylase,[78] acetohydroxyacid synthase,[79] pyruvate oxidase,[80] and pyruvate:ferredoxin oxidoreductase.[81]

Nucleotide-binding* and *ATP-binding*—Nucleotide-binding* in general, and *ATP-binding* in particular, are crucial events for the function of numerous well-folded and highly ordered nucleotide-binding enzymes. For example, different ATP- and NAD(+)-dependent DNA ligases, ATP-dependent RNA ligases and GTP-dependent mRNA capping enzymes

catalyze nucleotidyl transfer to polynucleotide 5' ends via covalent enzyme-(lysyl-N)-NMP intermediates. These enzymes share a core tertiary structure, which is composed minimally of a nucleotidyltransferase domain and an OB-fold domain.[82]

Molybdenum—While only a minor constituent of the earth's crust, *molybdenum* is easily available to biological systems due to the water solubility of its high-valent oxides. Molybdenium is an integral component of the multinuclear M center of nitrogenases[83] and is also found in the mononuclear active sites of a diverse group of enzymes that catalyze transfer of an oxygen atom either to or from a physiological acceptor or donor molecule, respectively.[84] Crystal structures of several molybdenium-containing enzymes, e.g., aldehyde oxidoreductase, xanthine oxidase, xanthine dehydrogenase, aldehyde oxidase, sulfite oxidase, nitrate reductase, DMSO reductase, etc., have been determined.[84, 85]

Nickel—*Nickel*-iron hydrogenases ([NiFe] hydrogenases, or H_2ases) are the best studied members of the hydrogenase (hydrogen acceptor oxido-reductase) family of enzymes that metabolize the most simple of chemical compounds, molecular hydrogen. The X-ray structure of a dimeric [NiFe] hydrogenase revealed that the large subunit contains the bimetallic [Ni-Fe] active site, with biologically uncommon CO and CN ligands bound to the iron, whereas the small subunit contains three iron-sulfur clusters. During catalysis, the *nickel* atom is most likely responsible for a base-assisted heterolytic cleavage of the hydrogen molecule, whereas the iron atom could be redox active.[86]

Manganese—Catalases, also known as hydroperoxidases, are one of the most studied classes of enzymes. These proteins catalyze the degradation of two molecules of hydrogen peroxide to water and oxygen. Although the most widespread in nature and the most extensively characterized class of catalases is the monofunctional heme-containing enzymes, several hydroperoxidases are the non-heme, or *manganese*-containing enzymes (also known as dimanganese catalases). High resolution crystal structures of two *manganese*-containing catalases revealed that the catalytic center of these homohexameric enzymes is a dimanganese group.[87]

Sodium—The major ion of the extracellular fluid is *sodium*. Transport of *sodium* ions across different membranes is crucial for life. In fact, production of electrical impulses in living organisms or tissues requires synchronized opening of transmembrane Na^+ channels possessing a *sodium* selectivity-filter, a high-throughput ion-conductance pathway, and a voltage-dependent gating function.[88] For example, voltage-gated sodium channels (VGSCs), which are highly ordered transmembrane proteins, are important for the generation and propagation of rapid electrical signals in electrically excitable tissues such as muscle, heart, and nerve.[89] The epithelial sodium channels (ENaCs), transmembrane proteins, which are composed of three partly homologous subunits, α, β and γ, inserted into the membrane with a proposed stoichiometry of 2α:1β:1γ, are crucial for the control of sodium fluxes in epithelial cells.[90] Mutations in genes encoding voltage-gated sodium channels have been correlated with numerous inherited human disorders affecting skeletal muscle contraction, heart rhythm, and nervous system function.[89] Similarly, the role of ENaC in the overall control of sodium balance, blood volume and thereby of blood pressure is demonstrated by genetic disorders of sodium-channel activity (such as Liddle's syndrome), a rare inheritable form of hypertension associated with gain-of-function mutations in the β and γ subunits of the ENaC.[91]

Cobalt—The functional activities of several crucial enzymes depend on *cobalt*. For example, the catalytic activity of adenosylcobalamin-dependent isomerases proceeds through the formation of free radical intermediates generated by homolysis of the cobalt-carbon bond of the coenzyme.[92] Members of the one of the three classes of ribonucleotide reductases (RNR),

the enzymes responsible for the conversion of the four standard ribonucleotides, to their 2'-deoxyribonucleotide counterparts, thus providing the precursors needed for both synthesis and repair of DNA, are *cobalt*-containing enzymes. These proteins utilize a cobaltous cofactor, adenosylcobalamin, a vitamin B_{12} derivative, that interacts directly with an active site cysteine to form the reactive cysteine radical needed for ribonucleotide reduction.[93]

Posttranslational modification and intrinsically disordered proteins

Posttranslational modification keywords that are strongly correlated with intrinsic disorder are listed in Table 3. Many sites of posttranslational modifications of the kinds given in Table 3 have been experimentally associated with regions of intrinsic disorder.[94]

Phosphorylation—Protein *phosphorylation* is known to represent an important regulatory mechanism in eukaryotic cells. At least one-third of all eukaryotic proteins are estimated to undergo reversible phosphorylation.[95] Phosphorylation modulates the activity of numerous proteins involved in signal transduction, and regulates the binding affinity of transcription factors to their coactivators and DNA, thereby altering gene expression, cell growth and differentiation.[96] Recently it has been reported that amino acid compositions, sequence complexity, hydrophobicity, charge and other sequence attributes of regions adjacent to phosphorylation sites are very similar to those of intrinsically disordered protein regions.[97] These observations were employed to develop a new web-based tool for the prediction of protein phosphorylation sites, DISPHOS (DISorder-enhanced PHOSphorylation predictor, http://www.ist.temple.edu/DISPHOS).[97] Using this predictor in association with kinase substrate preference-based predictors such as ScanSite[98] appears to be a useful combination (work in progress).

Cytoplasmic domains of several immune receptors, which are members of the family of multichain immune recognition receptors (MIRRs) (e.g., T-cell receptors (TCRs), B-cell receptors (BCRs), and the high-affinity IgE receptor) represent an illustrative example of the functional importance of intrinsic disorder in protein phosphorylation.[99–101] MIRRS were shown to be intrinsically disordered.[102, 103] They contain signaling subunits with the immunoreceptor tyrosine-based activation motif (ITAM). Phosphorylation of Tyr residues in these ITAMs takes place upon antigen binding and represents an early and obligatory event in the signaling cascade.[99–101]

Protein phosphorylation sites are the substrates for specific enzymes, kinases.[94] Typically, binding of these substrates to their cognate enzymes is characterized by low affinity, and yet phosphorylation by each kinase is a highly specific process.[104] Combination of high specificity with low affinity, being ideal for signaling, can be achieved via the coupled binding and folding.[105] The low net affinity arises because the positive free energy associated with the disorder-to-order transition must be deducted from the magnitude of the negative free energy arising from the interactions within the contact interface. The usefulness of protein disorder for such high specificity/low affinity signaling interactions was pointed out almost 25 years ago.[105]

Similar to phosphorylation, many other types of posttranslational modification including *acetylation*, *acylation*, *adenylylation*, *ADP ribosylation*, *amidation*, *carboxylation*, *formylation*, *glycosylation*, *methylation*, *sulfation*, and *ubiquitination* are controlled by specific enzymes.[106] Each of these examples is discussed below.

Lipidation—is a covalent attachement of fatty acids to the protein moiety. This type of posttranslational modification includes myristoylation, S-prenylation, and S-palmitoylation. *Myristylation* is the N-terminal attachment of a *myristoyl* lipid anchor to a glycine residue. This

posttranslational modification regulates protein–membrane and protein–protein interactions. The modification is catalyzed by the enzyme N-myristoyltransferase.[107]

GPI-anchor—Several membrane proteins contain a covalently linked lipid, glycosylphosphatidylinositol anchor (*GPI-anchor*), which tethers the proteins to the extracellular face of eukaryotic plasma membranes. GPI-anchored proteins are involved in a number of functions ranging from enzymatic catalysis to adhesion.[108] Both biosynthesis of GPI precursors and posttranslational protein modification with GPI occur in the endoplasmic reticulum. Upon GPI modification, the carboxyl-terminal signal peptide is split off from the protein and the resulting new carboxyl-terminal is then combined with the amino group of ethanolamine residue in the GPI precursors. The whole process of cleavage and GPI attachment is catalyzed by the GPI-transamidase complex.[109]

Proteoglycans—These represent a special class of *glycoproteins* whose protein cores are heavily glycosylated. Proteoglycans consist of a core protein with one or more covalently attached glycosaminoglycan chain(s) (e.g., heparin and heparan sulfate). These glycosaminoglycan chains are long, linear carbohydrate polymers that are negatively charged under physiological conditions. Glycosylation of the proteoglycan occurs in the Golgi apparatus in multiple enzymatic steps, where one sugar unit is added at each step.[110] In a more general form, a *glycoprotein* composed of a polypeptide and a carbohydrate (or oligosaccharide, which could be glucose, glucosamine, galactose, galactosamine, mannose, fucose and sialic acid). There are two types of glycosylation: N-glycosylation, where protein is modified at asparagines; and O-glycosylation, originating from the glycosylation at hydroxylysine, hydroxyproline, serine or threonine. It has been established that the glycosylation is a site-specific enzymatic process that involve several specific enzymes. Glycosylated proteins are often disordered, which may help to explain the observation that although it is estimated that >50% of all eukaryotic proteins are glycosylated,[111, 112] only ~5% of all PDB entries have attached glycan chains.[113]

***Gamma-carboxyglutamic acid* (Gla)—**This amino acid contains a dicarboxylic acid side chain, which occurs in a number of calcium-binding proteins. Gla has been discovered in blood coagulation proteins (prothrombin, Factor X, Factor IX, and Factor VII), plasma proteins of unknown function (Protein C, Protein S, and Protein Z), and proteins from calcified tissue (osteocalcin and bone-Gla protein).[114] Gla is synthesized by the post-translational modification of glutamic acid residues. This reaction is catalyzed by a specific enzyme, the vitamin K-dependent carboxylase.[115] Osteocalcin possesses a highly flexible structure which may be essential for the function of the protein.[116]

***Pyrrolidone carboxylic acid* (PCA, pGlu)—**This acid is formed either during the later stages of protein biosynthesis at the terminal phases of translation or as a post-translational event, just prior to cellular secretion of protein with amino-terminal pGlu.[117] The pGlu moiety results from the cyclization of an amino terminal glutamyl or glutaminyl residue by a specific enzyme, glutaminyl cyclase.[118] Many proteins and bioactive peptides exhibit an amino terminal pGlu residue, which subsequently minimizes their susceptibility to degradation by aminopeptidases.[119]

The binding of the above-mentioned enzymes to corresponding substrates represents high specificity/low affinity signaling interactions. The molecular mechanism of kinase interaction with the modified protein might serve as aprototype for these enzyme-driven posttranslational modifications that are specific for signaling and that involve the high specificity/low affinity interactions characteristic for their function. In fact, the validity of this assumption has been recently supported for protein *methylation*[120] and *ubiquitination* (Radivojak and Iakoucheva, work in progress). In addition to ubiqutin, a number of distinct ubiquitin-like proteins (Ubls)

such as SUMO, ISG15, Nedd8, and Atg8, function as protein modifiers. Post-translational covalent attachment of Ubls, *Ubl-conjugation*, is critical for many cellular processes, including transcription, DNA repair, signal transduction, autophagy, and cell-cycle control.[121] Ubl-conjugation cascades are initiated by activating enzymes, which also coordinate the ubls with their downstream pathways.[122]

Posttranslational modification keywords associated with ordered proteins

Table 4 lists keywords related to posttranslational modifications that occur in structured or ordered proteins.

Quinone

Several classes of enzymes that contain post-translationally modified amino acid residues have been recently discovered.[123, 124] One of the best characterized examples of such enzymes is the copper-containing amine oxidases (CuAOs), which contain a covalently bound cofactor, 2,4,5-trihydroxyphenylalanine *quinone* (TPQ), which is derived from the modification of an endogenous tyrosine residue[125] and a single copper ion located in the active site. CuAOs have been found in different organisms ranging from bacteria to mammals. Crystal structures solved for several CuAOs from different sources revealed dimeric tightly folded complexes.[126]

Organic radicals

The activities of several crucial enzymes depend on *organic radicals* covalently attached to the protein moiety. Three illustrative examples include the following: 1) ribonucleotide reductase which contains a stable organic free radical located on a tyrosine residue in the small subunit of the enzyme;[127] 2) S-adenosylmethionine radical enzymes (e.g., coproporphyrinogen-III oxidase and biotin synthase) in which S-adenosylmethionine (SAM) serves as a precursor to organic radicals, generated by one-electron reduction of SAM and subsequent fission to form 5'-deoxyadenosyl radical and methionine;[128] and 3) galactose oxidase in which a cofactor is derived from active-site amino acid residues via the autocatalytic formation of a thioether bond between Cys-228 and Tyr-272.[129]

Covalent protein-RNA linkage

The linkage between RNA and protein is found in some viral proteins that are attached to the end of a replicating viral RNA and that are necessary for RNA replication. 52% similarity between the distribution of hydrophobic and hydrophilic residues in 188 residues of the genome-linked viral protein (VPg) cistron located in the central part of potato virus Y genome and the fragment of cytoplasmic malate dehydrogenase of known crystal structure, has been used to propose a 3-D structural model of VPg.[130]

PQQ

Several quinoproteins are involved in the long-range interprotein electron transfer. Quinoproteins that possess *pyrroloquinoline quinone* (*PQQ*), tryptophan tryptophylquinone (TTQ), and cysteine tryptophylquinone (CTQ) are dehydrogenases.[131] PQQ is tightly but non-covalently bound to the enzyme, whereas TTQ and CTQ are derived from amino acid residues of the polypeptide chain. Illustrative examples of a PQQ-dependent enzyme are methanol dehydrogenase (MEDH), which catalyzes the oxidation of methanol to formaldehyde,[132] and soluble quinoprotein alcohol dehydrogenase, which is a monomeric enzyme with one PQQ and one *c*-type heme cofactor.[131]

Formylation

Another posttranslational modification of enzymes is provided by *formylation*. For example, according to the RESID database of protein modifications,[133] N-formyl-L-methionine is present in several cytochrome oxidases and phosphatidylserine decarboxylase. Crystal structures of several cytochrome c oxidases, containing N- formyl-L-methionine, have been solved.[134] N-formylglycine is found in N_{α}-formyl melittin and N-formyl-L-lysine is a part of some peptides from bee venom.[133, 135]

Zymogen

Many important proteolytic enzymes are synthesized as *zymogens*, i.e., inactive precursors or proenzymes decorated with a prosequence-inhibitor that has to be cleaved to make the enzyme functionally active. This is done to ensure precise regulation of the proteolytic enzyme activity, which is an essential requirement for cells and tissues because proteolysis at the wrong time or place may be lethal.[136] The analysis of crystal structures of several zymogens revealed that the inhibition is achieved via a specific mode of propeptide interaction with the proenzyme whereby the prosequence covers the active site cleft in a non-productive orientation.[136] An absolutely different mechanism of inhibition has been described for the malarial aspartic proteinases (plasmepsins), which are produced from inactive zymogens, proplasmepsins, having unusually long N-terminal prosegments of more than 120 amino acids.[137] Comparison of the crystal structures of plasmepsin and proplasmepsin from *Plasmodium vivax* revealed that a dramatic refolding of the mature N-terminus and a large (18°) reorientation of the N-domain between *P. vivax* proplasmepsin and plasmepsin produce a severe distortion of the active site region of the zymogen relative to that of the mature enzyme.[137]

Autocatalytic cleavage

This is an important step in activation of different enzymes, including proteolytic enzymes synthesized as *zymogens* and in the production of inteins via *protein splicing* (see above). Furthermore, autocatalytic cleavage is known to produce individual proteins from polyproteins synthesized by picornaviruses.[138] Finally, γ-glutamyltranspeptidase (GGT), a heterodimic enzyme that catalyzes the hydrolysis of γ-glutamyl bonds in γ-glutamyl compounds such as glutathione and/or the transfer of the γ-glutamyl group to other amino acids and peptidesis, is generated from the precursor protein through the posttranslational autocatalytic cleavage.[139] The crystal structure of GGT from *E. coli* has been recently determined at 1.95 Å resolution.[140]

Protein splicing

This is an intriguing posttranslational modification, where an intervening protein sequence (intein, internal proteins) is self-catalytically excised out from a protein precursor and the two flanking sequences (N- and C-exteins) are ligated to produce two mature enzymes.[141, 142] This phenomenon was first described in 1990 for the TFP1 gene product of *Saccharomyces cerevisiae*,[143] and several years later for the recA protein of *Mycobacterium tuberculosis*,[144] an archaeal DNA polymerase,[145] and for the production of Vma1p (a catalytic 70-kDa subunit of the vacuolar H^+-ATPase) and VDE (a 50-kDa DNA endonuclease or VMA1 intein) from the nascent 120-kDa translational product of the *VMA1* gene in *Saccharomyces cerevisiae*.[146] The crystal structure of VDE has been determined at 2.1 Å resolution.[146] Protein splicing is not unique for the VMA1 precursor. Most inteins consist of two domains: one is involved in autocatalytic splicing, and the other is an endonuclease that is important in the spread of inteins.[147] The mechanism of this process has been elucidated[148, 149] based on the well-established chemistry for the non-enzymatic cleavage of asparagine residues in proteins resulting in the formation of succinimide. Recently, a database of protein splicing,

InBase, has been compiled[150] and is currently maintained by the New England BioLabs (http://www.neb.com/neb/inteins.html).

Oxidation

Proteins can undergo different types of *oxidation*. These include carbonylation (an irreversible process that targets different amino acids including lysine, arginine, proline and threonine[151, 152]), nitration of tyrosine[152] and oxidation of methionine to methionine sulfoxide.[153] Methionine is easily oxidized by H_2O_2, hypochlorite, chloramines, and peroxynitrite; all these oxidants are produced in biological systems.[153] However, this modification can be repaired by methionine sulfoxide reductase, which catalyzes the thioredoxin-dependent reduction of methionine sulfoxide back to methionine.[154] It has been pointed out that the reversible oxidation of methionine has the potential to modulate intracellular signaling under conditions involving oxidative stress in a manner analogous to other regulatory posttranslational modifications such as those involving disulfide bond formation or phosphorylation.[153]

Protein cysteines can undergo various forms of oxidation, some of them reversible (disulphide formation, glutathionylation, cysteinylation, S-nitrosylation and formation of sulphenic and sulphinic acids).[155] Intraprotein disulphide bonds are viewed, in classical textbooks, as part of the well ordered tertiary structure of the protein and their formation is an important step in protein folding. Similarly, interprotein disulphide bonds are important in the quaternary structure of proteins and in the formation of homo- or hetero-multimers.

Hypusine

Some proteins are posttranslationally modified on lysine residues to form hypusine (N-epsilon-(4-aminobutyl)lysine). Hypusine (a molecule comprises moieties of HYdroxyPUtrescine and lySINE) was first isolated from bovine brain in 1971.[156] eIF5A is the only protein in eukaryotes and archaebacteria known to contain hypusine. Hypusine is formed in eIF5A by a novel post-translational spermidine-dependent modification reaction that involves two enzymatic steps. In the first step, deoxyhypusine synthase catalyzes the cleavage of the polyamine spermidine and transfer of its 4-aminobutyl moiety to the epsilon-amino group of one specific lysine residue of the eIF5A precursor to form a deoxyhypusine intermediate. In the second step, deoxyhypusine hydroxylase converts the deoxyhypusine-containing intermediate to the hypusine-containing mature eIF5A.[157] The structure and mechanisms of deoxyhypusine synthase and deoxyhypusine hydroxylase have been extensively characterized.[158] Deoxyhypusine hydroxylase is a HEAT-repeat protein with a symmetrical superhelical structure consisting of 8 helical hairpins (HEAT motifs). It is a novel metalloenzyme containing tightly bound iron at the active sites.[158]

Intrinsic disorder in disease associated proteins

In the Swiss-Prot disease category, 10 disease-related keywords have strong correlation with intrinsically disorder proteins and no disease-associated keywords were found to be related to intrinsic order (see Table 5).

Oncoproteins

Important to this study, disorder is very common in cancer-associated proteins (or *oncoproteins*).[159] In fact, in previous studies we found that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted regions of disorder of 30 residues or longer.[159] In contrast, in a control set of proteins with well-defined ordered structures, which was extracted from protein data bank (PDB), the content of such long disordered regions was much smaller, as only 13% of these proteins contained long regions of predicted disorder. In experimental

studies, the presence of disorder has been directly observed in many cancer-associated proteins, a few examples of which include p53,[160] $p57^{kip2}$,[161] $Bcl\text{-}X_L$ and Bcl-2,[162] c-Fos,[163] proto-oncogene securin,[164] a breast cancer associated protein BRCA1.[165] We have recently established that the E6 and E7 *oncoproteins* from the high-risk types of human papillomaviruses (HPVs) possess more predicted intrinsic disorder than proteins from the low-risk HPVs.[166]

Malaria

Biophysical analysis of the merozoite surface protein 3 (MSP3) of *Plasmodium falciparum* (the parasitic agent that causes most cases of fatal *malaria*) revealed that MSP3 polypeptides contain a large amount of α-helix and random coil secondary structure and form highly elongated dimers and tetramers.[167] The MSP3 dimer was assumed to be formed via a parallel *coiled-coil* interaction between the *leucine zipper*-like regions of two monomers. Importantly, MSP3 contains numerous features associated with intrinsic disorder: its central domain includes three blocks of imperfect Ala heptad *repeats*, a second central domain is a very long Glu-rich region (242–294 fragment), whereas the C-terminal domain contains a *leucine zipper* motif.[167] Apical membrane antigen 1 (AMA1) of the malarial parasite *Plasmodium falciparum* is a merozoite antigen that is considered a strong candidate for inclusion in a malaria vaccine. The solution structure of AMA1 domain III, a 14kDa protein, has been determined using NMR spectroscopy.[168] It was shown that the structure has a well-defined disulfide-stabilized core region separated by a disordered loop, and both the N and C-terminal regions of the molecule are unstructured.[168] In another study the solution structure of a 16kDa construct corresponding to the putative second domain of AMA1 have been reported.[169] Interestingly, while CD and hydrodynamic data were consistent with a folded structure for domain II, its NMR spectra were characterized by broad lines and significant peak overlap, more typical of a molten globule.[169]

Trypanosomiasis

Several surface antigens of *trypanosome* are *glycoproteins*. For example, the glutamic acid and alanine-rich procyclin from *Trypanosoma congolense* is a *glycoprotein* containing a galactosylated *GPI-anchor* and carrying two large mannose- and galactose-containing oligosaccharides linked to threonine residues via phosphodiester linkages.[170] Furthermore, in *Trypanosoma brucei*, the procyclic and epimastigote stage glycoproteins belong to the family of procyclic acidic repetitive proteins (procyclins) containing numerous tandem *repeats* of EP or GPEET.[171–173] The procyclins provide a highly acidic coat that is proposed to help the parasite survive in the harsh environment of the tsetse fly gut.[170] The isolated GPEET procyclin from the *T. brucei* is highly susceptible to proteolytic treatment,[172] suggesting one possible need for the protective coat.

HIV and *AIDS*

The human immunodeficiency virus (*HIV*) is the causative agent for acquired immunodeficiency syndrome (*AIDS*). The HIV genome encodes a total of three structural proteins, two envelope proteins, three enzymes, and six accessory proteins.[174] One of the HIV structural proteins, nucleocapsid (NC) protein, is involved in interactions with *nucleic acids*. The solution behavior of NC may be best considered as a rapid equilibrium between conformations with weakly interacting and non-interacting knuckle (*zinc finger*-like) domains, and this inherent conformational flexibility plays a crucial role in the adaptive binding of NC to different nucleic acid targets.[175] A basic domain (Arg35 to Arg50) of the accessory protein Rev was shown to be largely unstructured in solution, but forms an α-helix upon binding to the SLIIB RNA stem.[176, 177] NMR structural analysis of another accessory protein, Tat (trans-activator) protein from HIV-1, revealed that it possesses two highly flexible domains

corresponding to a cysteine-rich and a basic sequence region.[178] Furthermore, recent multinuclear NMR analysis of uniformly ^{15}N and $^{15}N/^{13}C$-labelled $Tat_{1–72}$ region under reduced conditions (pH 4.1) revealed that that it exists in a random coil conformation.[179] The absence of a fixed conformation and the fast dynamics are consistent with the ability of Tat protein to interact with a wide variety of proteins and nucleic acid and support the concept of a natively-unfolded protein.[179] The envelope *glycoprotein* of HIV forms trimers on the virion surface, with each monomer consisting of two subunits, gp120 and gp41.[180] During the initial step of HIV infection, the gp41/gp120 complex associates with the CD4 receptor located on the surface of a human cell. The mature gp120 consists of 5 conserved regions (C1–C5) and five variable regions (V1–V5). The C1 and C5 domains of gp120 are involved in direct interaction with gp41 but are largely missing from the available X-ray structure,[181] suggesting that these regions can be disordered. In agreement with this disordered region hypothesis, the HIV gp120 C5 domain (residues 489–511 of HIV-1 strain HXB2), which corresponds to the carboxy terminal region of gp120, has been recently reported to be unstructured in aqueous solutions.[182]

Deafness

Connexins are integral membrane proteins that oligomerize to form intercellular channels called gap junctions between adjacent cells, and these channels promote intercellular communication. Connexin proteins are involved in pathological conditions in humans, mainly in hearing loss, neurodegenerative disorders and skin diseases. There are over 100 mutations in genes encoding connexins that are associated with *deafness*. Most prominent is the remarkable involvement of connexin 26 in hearing loss. Mutations in the gene GJB2, encoding connexin 26, are responsible for around 50% of genetic cases of severe to profound non-syndromic hearing loss in some parts of the world.[183] Conformational analysis of two intracellular domains of connexin43 (Cx43), cytoplasmic loop 95–144 and C-terminal domain (amino acids 254–382), revealed that they are mostly disordered and possess short transient α-helices that are connected by long, highly flexible loops of random coil.[184–186] Numerous binding partners to the carboxyl-terminal domain of Cx43 have been identified; these include tubulin, v-Src, c-Src, ZO-1, casein kinase 1 (CK1), mitogen-activated protein kinase (MAPK), cGMP-dependent protein kinase, cAMP-dependent protein kinase, and protein kinase C.[187] Interestingly, other than the possible binding of CK1 to Ser-325 within one of the α-helices, all these proteins interact with Cx43 in disordered regions outside the two α-helical domains.[186] This firther emphasizes that the disordered structure of the carboxy-terminal domain of Cx43 is advantageous for signaling between different binding partners.

Obesity, cardiovascular disease, and *diabetes mellitus*

Several human diseases including *cardiovascular disease, diabetes mellitus*, hyperlipidemia and hypertension and so-called metabolic syndrome are *obesity*-related disorders that are associated with the visceral fat accumulation.[188] Our recent bioinformatics analysis of a set of 487 *cardiovascular disease* (CVD)-related proteins revealed that these proteins are enriched in intrinsic disorder.[189] The percentage of proteins with long disordered regions was shown to be 61(±5)% for CVD-associated proteins, which is less than the value described earlier for human cancer-associated and signaling proteins (79±5% and 66±6%, respectively),[159] but which is significantly larger than that in eukaryotic proteins from SWISS-PROT (47±4%) and in non-homologous protein segments with well-defined 3-D structure (13±4%). Furthermore, 120 out of 487 proteins in the CVD dataset (~25%) were predicted to be wholly disordered.[189] This high level of intrinsic disorder could be important for function of CVD-related protein and for the control and regulation of processes associated with cardiovascular disease. In agreement with this hypothesis, [198] α-helical molecular recognition fragments α-MoRFs[190] were predicted in 101 proteins from CVD dataset.[189]

Albinism

is a complex genetic disease, which is a result of the melanin pigment deficiency in the skin, hair, and eye [oculocutaneous albinism (OCA)], or primarily in the eye [ocular albinism (OA)]. Mutations in six genes have been reported to be responsible for different types of oculocutaneous and ocular albinism, including the tyrosinase gene (*TYR*), the *OCA2* gene, the gene (*TYRP1*), the Hermansky-Pudlak syndrome (*HPS*) gene, the Chediak-Higashi syndrome (*CHS*) gene, and the X-linked ocular albinism gene.[191] Tyrosinase is a *glycosylated* transmembrane *copper-containing* enzyme that is responsible for conversion of tyrosine to dopaquinone, which is the rate-limiting step in the melanin pathway.[192] The enzyme contains 529 amino acids, including an 18-amino acid *signal peptide*, two putative copper-binding sites, and a hydrophobic transmembrane region at the C-terminal end.[193] Tyrosinases are activated in vivo by limited *proteolytic cleavage*.[194] All these proteins are predicted to have long regions of intrinsic disorder.

Prion

The central event in the pathogenesis of *prion* diseases is a major conformational change of the prion protein (PrP) from an alpha-helical (PrP^{C}) to a beta-sheet-rich isoform (PrP^{Sc}). The mature PrP^{C} species consists of an N-terminal region of about 100 amino acids, which is unstructured in the isolated molecule in solution, and a folded C-terminal domain, also approximately 100 amino acids in length. It has been emphasized that the most striking feature of the full-length PrP is the random-coil nature of chemical shifts for its residues 30–124.[195] The C-terminal domain is folded into a largely α-helical conformation (three α-helices and a short antiparallel β-sheet) and stabilized by a single disulfide bond linking helices 2 and 3.[196] Although unstructured in the isolated molecule, the N-terminal region contains tight binding sites for Cu^{2+} ions and therefore may acquire structure following copper binding.[197, 198]

Disease-associated mutations

Functional protein represents a tightly balanced machine, structure and performance of which can be easily distorted by point mutations. For example, p53 was shown to play crucial role in the development of most cancers by condemning damaged cells to death or quarantining them for repair.[199] p53 activity relies on its intact native conformation, which can be lost following mutation of a single nucleotide. In fact, over 10,000 somatic tumorigenic mutations in p53 gene were found,[200] with 95% of these lie in the core DNA-binding domain.[199]

Furthermore, even single changes of amino acid in protein sequences can change the rates at which they aggregate by an order of magnitude or more,[201, 202] thus dramatically accelerating development of protein depositions and related diseases. In fact, the changes in aggregation rates cause by such mutations were shown to correlate with changes in simple properties that result from such substitutions, such as charge, secondary structure propensities and hydrophobicity.[202] Mutations modulate the aggregation propensities of both, well-folded and intrinsically disordered proteins. Numerous neurodegenerative diseases originate from misfolding and neurotoxic aggregation of specific proteins. It is established that Aβ, α-synuclein and prion protein, the major players involved in the pathogenesis of such famous diseases as Alzheimer's and Parkinson's diseases and prion diseases, respectively, are either completely disordered (Aβ, α-synuclein) or contain long disordered regions (prion protein).[203] For example, detailed structural analysis revealed that α-synuclein, a conservative presynaptic protein, the aggregation and fibrillation of which is assumed to be involved into the pathogenesis of Parkinson's disease and several other neurodegenerative disorders, synucleinopathies, is characterized by the lack of rigid well-defined structure under the physiological conditions *in vitro*. However, this protein is characterized by a remarkable conformational plasticity as it adopts a series of different conformations depending on the

environment, being able to either stay substantially unfolded, or adopt an amyloidogenic partially folded conformation, or fold into α-helical or β-structural species, both monomeric and oligomeric. Its aggregated forms possess astonishing morphological diversity, ranging from oligomers (spheres or doughnuts) to amorphous aggregates or amyloid-like fibrils. This unusual conformational behavior and exceptional structural plasticity of α-synuclein have led to the protein-chameleon concept, according to which a polypeptide chain can gain different structures depending on its environment.[204] There are three point mutations in α-synuclein, A30P,[205] E49K,[206] and A53T,[207] which are associated with the early onset of the Parkinson's disease and were shown to accelerate the α-synuclein aggregation (but not necessarily fibrillation) *in vitro*.[208]

Conclusions

We have established that many natural ligands, PTMs and diseases are associated with proteins predicted to possess long disordered regions. Strong positive or negative correlations of different PTMs with intrinsically disordered regions is of special interest. Our data are consistent with the existence of two major protein groups subjected to PTMs. The first group involves modifications that are associated primarily with structured proteins and regions. These include the following modifications: formylation, protein splicing, oxidation and covalent attachment of quinones and organic radicals. These modifications are important for providing moieties for catalytic functions, for modifying enzyme activities or for stabilizing protein structure.

The second group involves modifications that are associated primarily with intrinsically unstructured or disordered proteins and regions. These include the following modifications: phosphorylation, acetylation, acylation, adenylylation, ADP ribosylation, amidation, carboxylation, formylation, glycosylation, methylation, sulfation, prenylation, ubiquitination, Ubl-conjugation (i.e., covalent attachment of ubiquitin-like proteins, including SUMO, ISG15, Nedd8, and Atg8). These modifications involve low affinity, high specificity binding interactions between a specific enzyme and a substrate (the protein that is modified). Combination of high specificity with low affinity, being ideal for signaling, can be achieved via coupled binding and folding. Importantly, posttranslational modifications associated with intrinsically disordered proteins and regions are especially important for signaling and regulation. For example, protein phosphorylation is known to represent a crucial regulatory mechanism in eukaryotic cells. Another illustrative example is Ubl-conjugation, which is critical for many cellular processes, including transcription, DNA repair, signal transduction, autophagy, and cell-cycle control. Ubl-conjugation cascades are initiated by activating enzymes, which also coordinate the ubls with their downstream pathways. In fact, conjugation of ubiquitin-like proteins (the Ubl conjugation pathway) to components of the transcriptional machinery is an important regulatory mechanism allowing switching between different activity states. While ubiquitination of transcription factors is associated with transcriptional activation, their SUMOylation is most often connected with transcriptional repression.

Overall, comparison of the posttranslational modifications in structured and intrinsically disordered regions revealed that the latter are more diverse and more often tend to be reversible. Thus, the irreversible posttranslational modifications are mostly used to increase stability and allow catalytic functions of the ordered proteins, whereas the reversible modifications are more frequently used for the signaling activities of intrinsically disordered proteins.

Acknowledgment

The authors express their deepest gratitude to Celeste Brown and Predrag Radivojak for numerous valuable discussions. This work was supported by grants from the National Institutes of Health LM007688-0A1 (A. K. D., and Z. O.) and GM071714-01A2 (A.K.D and V.N.U.), and by the Indiana Genomics Initiative (INGEN) (A. K. D.). INGEN

is supported in part by Lilly Endowment Inc. The Programs of the Russian Academy of Sciences for the "Molecular and cellular biology" and "Fundamental science for medicine" provided partial support to V. N. U., and L. M. I. was supported by the NSF grant MCB 0444818.

References

1. Mersfelder EL, Parthun MR. The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. Nucleic Acids Res 2006;34:2653–2662. [PubMed: 16714444]
2. Grant PA. A tale of histone modifications. Genome Biol 2001;2REVIEWS0003
3. Goll MG, Bestor TH. Histone modification and replacement in chromatin activation. Genes Dev 2002;16:1739–1742. [PubMed: 12130533]
4. Turner BM. Cellular memory and the histone code. Cell 2002;111:285–291. [PubMed: 12419240]
5. Cocklin RR, Wang M. Identification of methylation and acetylation sites on mouse histone H3 using matrix-assisted laser desorption/ionization time-of-flight and nanoelectrospray ionization tandem mass spectrometry. J Protein Chem 2003;22:327–334. [PubMed: 13678296]
6. Zhang K, Tang H, Huang L, Blankenship JW, Jones PR, Xiang F, Yau PM, Burlingame AL. Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. Anal Biochem 2002;306:259–269. [PubMed: 12123664]
7. Zhang L, Eugeni EE, Parthun MR, Freitas MA. Identification of novel histone post-translational modifications by peptide mass fingerprinting. Chromosoma 2003;112:77–86. [PubMed: 12937907]
8. Stryer, L. Biochemistry. Vol. 1. Moscow: Mir; 1985.
9. Metzler, DE. Biochemistry. The Chemical Reactions of Living Cell. Vol. 1. Moscow: Mir; 1980.
10. Fersht, AR. Enzyme Structure and Mechanism. Moscow: Mir; 1980.
11. Ovchinnikov, YA. Bioorganic Chemistry. Moscow: Prosveshchenie; 1987.
12. Hoard, JL. Stereochemistry of porphyrins. In: Chance, B.; Estabrook, RW.; Yonetani, T., editors. Hemes and Homoproteins. New York: Academic Press; 1966. p. 9-24.
13. Perutz MF. Stereochemistry of cooperative effects in haemoglobin. Nature 1970;228:726–739. [PubMed: 5528785]
14. Uversky, VN. A rigidifying union: The role of ligands in protein structure and stability. In: Pandalai, SG., editor. Recent Research Developments in Biophysics & Biochemistry. Kerala, India: Transworld Research Network; 2003. p. 711-745.
15. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional anthology of intrinsic disorder. I. Biological processes and functions of proteins with long disordered regions. J Proteome Res. 2006
16. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31:365–370. [PubMed: 12520024]
17. O'Donovan C, Martin MJ, Glemet E, Codani JJ, Apweiler R. Removing redundancy in SWISS-PROT and TrEMBL. Bioinformatics 1999;15:258–259. [PubMed: 10222414]
18. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 2002;30:1575–1584. [PubMed: 11917018]
19. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 2005;3:35–60. [PubMed: 15751111]
20. Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. Nucleic Acids Res 2006;34:305–312. [PubMed: 16407336]
21. Gutierrez P, Osborne MJ, Siddiqui N, Trempe JF, Arrowsmith C, Gehring K. Structure of the archaeal translation initiation factor aIF2 beta from Methanobacterium thermoautotrophicum: implications for translation initiation. Protein Sci 2004;13:659–667. [PubMed: 14978306]
22. Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B. 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 2003;278:18638–18648. [PubMed: 12621042]

23. Bourhis JM, Johansson K, Receveur-Brechot V, Oldfield CJ, Dunker KA, Canard B, Longhi S. 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 2004;99:157–167. [PubMed: 14749181]
24. Bourhis JM, Receveur-Brechot V, Oglesbee M, Zhang X, Buccellato M, Darbon H, Canard B, Finet S, Longhi S. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. Protein Sci 2005;14:1975–1992. [PubMed: 16046624]
25. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. J Gen Virol 2003;84:3239–3252. [PubMed: 14645906]
26. Vasak M, Hasler DW. Metallothioneins: new functional and structural insights. Curr Opin Chem Biol 2000;4:177–183. [PubMed: 10742189]
27. Ejnik J, Robinson J, Zhu J, Forsterling H, Shaw CF, Petering DH. Folding pathway of apo-metallothionein induced by Zn2+, Cd2+ and Co2+ J Inorg Biochem 2002;88:144–152. [PubMed: 11803035]
28. Blindauer CA, Sadler PJ. How to hide zinc in a small protein. Acc Chem Res 2005;38:62–69. [PubMed: 15654738]
29. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional anthology of intrinsic disorder. II. Cellular components, domains, technical terms, developmental processes and coding sequence diversity associated with long disordered regions. J Proteome Res. 2006
30. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41:415–427. [PubMed: 11025552]
31. Bubb MR. Thymosin beta 4 interactions. Vitam Horm 2003;66:297–316. [PubMed: 12852258]
32. Filatov VL, Katrukha AG, Bulargina TV, Gusev NB. Troponin: structure, properties, and mechanism of functioning. Biochemistry (Mosc) 1999;64:969–985. [PubMed: 10521712]
33. Slupsky CM, Sykes BD. NMR solution structure of calcium-saturated skeletal muscle troponin C. Biochemistry 1995;34:15953–15964. [PubMed: 8519752]
34. Heller WT, Abusamhadneh E, Finley N, Rosevear PR, Trewhella J. The solution structure of a cardiac troponin C-troponin I-troponin T complex shows a somewhat compact troponin C interacting with an extended troponin I-troponin T component. Biochemistry 2002;41:15654–15663. [PubMed: 12501194]
35. Radivojac P, Vucetic S, O'Connor TR, Uversky VN, Obradovic Z, Dunker AK. Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. Proteins 2006;63:398–410. [PubMed: 16493654]
36. Stull JT. Ca2+-dependent cell signaling through calmodulin-activated protein phosphatase and protein kinases minireview series. J Biol Chem 2001;276:2311–2312. [PubMed: 11096124]
37. Vetter SW, Leclerc E. Novel aspects of calmodulin target recognition and activation. Eur J Biochem 2003;270:404–414. [PubMed: 12542690]
38. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry 1992;31:5269–5278. [PubMed: 1606151]
39. Ikura M, Barbato G, Klee CB, Bax A. Solution structure of calmodulin and its complex with a myosin light chain kinase fragment. Cell Calcium 1992;13:391–400. [PubMed: 1505004]
40. Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, Bax A. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. Science 1992;256:632–638. [PubMed: 1585175]
41. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput 1998:473–484. [PubMed: 9697205]
42. Mulloy B, Linhardt RJ. Order out of complexity--protein structures that interact with heparin. Curr Opin Struct Biol 2001;11:623–628. [PubMed: 11785765]

43. Capila I, Hernaiz MJ, Mo YD, Mealy TR, Campos B, Dedman JR, Linhardt RJ, Seaton BA. Annexin V--heparin oligosaccharide complex suggests heparan sulfate--mediated assembly on cell surfaces. Structure 2001;9:57–64. [PubMed: 11342135]
44. Cohlberg JA, Li J, Uversky VN, Fink AL. Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from alpha-synuclein in vitro. Biochemistry 2002;41:1502–1511. [PubMed: 11814343]
45. Friedhoff P, Schneider A, Mandelkow EM, Mandelkow E. Rapid assembly of Alzheimer-like paired helical filaments from microtubule-associated protei tau monitored by fluorescence in solution. Biochemistry 1998;37:10223–10230. [PubMed: 9665729]
46. Paudel HK, Li W. Heparin-induced conformational change in microtubule-associated protein Tau as detected by chemical cross-linking and phosphopeptide mapping. J Biol Chem 1999;274:8029–8038. [PubMed: 10075702]
47. Headey SJ, Keizer DW, Yao S, Brasier G, Kantharidis P, Bach LA, Norton RS. C-terminal domain of insulin-like growth factor (IGF) binding protein-6: structure and interaction with IGF-II. Mol Endocrinol 2004;18:2740–2750. [PubMed: 15308688]
48. Ovchinnikov YA, Lipkin VM, Kumarev VP, Gubanov VV, Khramtsov NV, Akhmedov NB, Zagranichny VE, Muradov KG. Cyclic GMP phosphodiesterase from cattle retina. Amino acid sequence of the gamma-subunit and nucleotide sequence of the corresponding cDNA. FEBS Lett 1986;204:288–292. [PubMed: 3015681]
49. Uversky VN, Permyakov SE, Zagranichny VE, Rodionov IL, Fink AL, Cherskaya AM, Wasserman LA, Permyakov EA. Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein. J Proteome Res 2002;1:149–159. [PubMed: 12643535]
50. Gerken TA. The solution structure of mucous glycoproteins: proton NMR studies of native and modified ovine submaxillary mucin. Arch Biochem Biophys 1986;247:239–253. [PubMed: 3013090]
51. Shanmugam G, Polavarapu PL. Structures of intact glycoproteins from vibrational circular dichroism. Proteins 2006;63:768–776. [PubMed: 16498615]
52. Hai TW, Liu F, Coukos WJ, Green MR. Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev 1989;3:2083–2090. [PubMed: 2516827]
53. Nagadoi A, Nakazawa K, Uda H, Okuno K, Maekawa T, Ishii S, Nishimura Y. Solution structure of the transactivation domain of ATF-2 comprising a zinc finger-like subdomain and a flexible subdomain. J Mol Biol 1999;287:593–607. [PubMed: 10092462]
54. McPhee I, Gibson LC, Kewney J, Darroch C, Stevens PA, Spinks D, Cooreman A, MacKenzie SJ. Cyclic nucleotide signalling: a molecular approach to drug discovery for Alzheimer's disease. Biochem Soc Trans 2005;33:1330–1332. [PubMed: 16246111]
55. Bos JL. Linking Rap to cell adhesion. Curr Opin Cell Biol 2005;17:123–128. [PubMed: 15780587]
56. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. Febs J 2005;272:5129–5148. [PubMed: 16218947]
57. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic Disorder is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. PLoS Comput Biol. 2006in press
58. Wikstrom M, Drakenberg T, Forsen S, Sjobring U, Bjorck L. Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. Biochemistry 1994;33:14011–14017. [PubMed: 7947810]
59. Hirabayashi J, Kasai K. The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. Glycobiology 1993;3:297–304. [PubMed: 8400545]
60. Dumic J, Dabelic S, Flogel M. Galectin-3: an open-ended story. Biochim Biophys Acta 2006;1760:616–635. [PubMed: 16478649]
61. Iacobini C, Amadio L, Oddi G, Ricci C, Barsotti P, Missori S, Sorcini M, Di Mario U, Pricci F, Pugliese G. Role of galectin-3 in diabetic nephropathy. J Am Soc Nephrol 2003;14:S264–S270. [PubMed: 12874444]

62. Liu FT, Frigeri LG, Gritzmacher CA, Hsu DK, Robertson MW, Zuberi RI. Expression and function of an IgE-binding animal lectin (epsilon BP) in mast cells. Immunopharmacology 1993;26:187–195. [PubMed: 8288440]
63. Birdsall B, Feeney J, Burdett ID, Bawumia S, Barboni EA, Hughes RC. NMR solution studies of hamster galectin-3 and electron microscopic visualization of surface-adsorbed complexes: evidence for interactions between the N- and C-terminal domains. Biochemistry 2001;40:4859–4866. [PubMed: 11294654]
64. Nishimura M, Yoshida T, Shirouzu M, Terada T, Kuramitsu S, Yokoyama S, Ohkubo T, Kobayashi Y. Solution structure of ribosomal protein L16 from Thermus thermophilus HB8. J Mol Biol 2004;344:1369–1383. [PubMed: 15561149]
65. Sargent PJ, Farnaud S, Evans RW. Structure/function overview of proteins involved in iron storage and transport. Curr Med Chem 2005;12:2683–2693. [PubMed: 16305465]
66. Imlay JA. Iron-sulphur clusters and the problem with oxygen. Mol Microbiol 2006;59:1073–1082. [PubMed: 16430685]
67. Mewies M, McIntire WS, Scrutton NS. Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: the current state of affairs. Protein Sci 1998;7:7–20. [PubMed: 9514256]
68. John RA. Pyridoxal phosphate-dependent enzymes. Biochim Biophys Acta 1995;1248:81–96. [PubMed: 7748903]
69. Neidhart DJ, Kenyon GL, Gerlt JA, Petsko GA. Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. Nature 1990;347:692–694. [PubMed: 2215699]
70. Gerlt JA, Babbitt PC, Rayment I. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. Arch Biochem Biophys 2005;433:59–70. [PubMed: 15581566]
71. Cramer WA, Zhang H, Yan J, Kurisu G, Smith JL. Evolution of photosynthesis: time-independent structure of the cytochrome b6f complex. Biochemistry 2004;43:5921–5929. [PubMed: 15147175]
72. AEvarsson A, Seger K, Turley S, Sokatch JR, Hol WG. Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. Nat Struct Biol 1999;6:785–792. [PubMed: 10426958]
73. AEvarsson A, Chuang JL, Wynn RM, Turley S, Chuang DT, Hol WG. Crystal structure of human branched-chain alpha-ketoacid dehydrogenase and the molecular basis of multienzyme complex deficiency in maple syrup urine disease. Structure 2000;8:277–291. [PubMed: 10745006]
74. Arjunan P, Nemeria N, Brunskill A, Chandrasekhar K, Sax M, Yan Y, Jordan F, Guest JR, Furey W. Structure of the pyruvate dehydrogenase multienzyme complex E1 component from Escherichia coli at 1.85 A resolution. Biochemistry 2002;41:5213–5221. [PubMed: 11955070]
75. Ciszak EM, Korotchkina LG, Dominiak PM, Sidhu S, Patel MS. Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase. J Biol Chem 2003;278:21240–21246. [PubMed: 12651851]
76. Fiedler E, Thorell S, Sandalova T, Golbik R, Konig S, Schneider G. Snapshot of a key intermediate in enzymatic thiamin catalysis: crystal structure of the alpha-carbanion of (alpha,betadihydroxyethyl)-thiamin diphosphate in the active site of transketolase from Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 2002;99:591–595. [PubMed: 11773632]
77. Dyda F, Furey W, Swaminathan S, Sax M, Farrenkopf B, Jordan F. Catalytic centers in the thiamin diphosphate dependent enzyme pyruvate decarboxylase at 2.4-A resolution. Biochemistry 1993;32:6165–6170. [PubMed: 8512926]
78. Hasson MS, Muscate A, McLeish MJ, Polovnikova LS, Gerlt JA, Kenyon GL, Petsko GA, Ringe D. The crystal structure of benzoylformate decarboxylase at 1.6 A resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes. Biochemistry 1998;37:9918–9930. [PubMed: 9665697]
79. Pang SS, Duggleby RG, Guddat LW. Crystal structure of yeast acetohydroxyacid synthase: a target for herbicidal inhibitors. J Mol Biol 2002;317:249–262. [PubMed: 11902841]
80. Muller YA, Schumacher G, Rudolph R, Schulz GE. The refined structures of a stabilized mutant and of wild-type pyruvate oxidase from Lactobacillus plantarum. J Mol Biol 1994;237:315–335. [PubMed: 8145244]

81. Kern D, Kern G, Neef H, Tittmann K, Killenberg-Jabs M, Wikner C, Schneider G, Hubner G. How thiamine diphosphate is activated in enzymes. Science 1997;275:67–70. [PubMed: 8974393]
82. Shuman S, Lima CD. The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. Curr Opin Struct Biol 2004;14:757–764. [PubMed: 15582400]
83. Howard JB, Rees DC. Structural Basis of Biological Nitrogen Fixation. Chem Rev 1996;96:2965–2982. [PubMed: 11848848]
84. Hille R. The Mononuclear Molybdenum Enzymes. Chem Rev 1996;96:2757–2816. [PubMed: 11848841]
85. Hille R. Molybdenum and tungsten in biology. Trends Biochem Sci 2002;27:360–367. [PubMed: 12114025]
86. Fontecilla-Camps JC, Frey M, Garcin E, Hatchikian C, Montet Y, Piras C, Vernede X, Volbeda A. Hydrogenase: a hydrogen-metabolizing enzyme. What do the crystal structures tell us about its mode of action? Biochimie 1997;79:661–666. [PubMed: 9479448]
87. Chelikani P, Fita I, Loewen PC. Diversity of structures and properties among catalases. Cell Mol Life Sci 2004;61:192–208. [PubMed: 14745498]
88. Rogerson FM, Brennan FE, Fuller PJ. Dissecting mineralocorticoid receptor structure and function. J Steroid Biochem Mol Biol 2003;85:389–396. [PubMed: 12943727]
89. George AL Jr. Inherited disorders of voltage-gated sodium channels. J Clin Invest 2005;115:1990–1999. [PubMed: 16075039]
90. Gormley K, Dong Y, Sagnella GA. Regulation of the epithelial sodium channel by accessory proteins. Biochem J 2003;371:1–14. [PubMed: 12460120]
91. Schild L, Lu Y, Gautschi I, Schneeberger E, Lifton RP, Rossier BC. Identification of a PY motif in the epithelial Na channel subunits as a target sequence for mutations causing channel activation found in Liddle syndrome. Embo J 1996;15:2381–2387. [PubMed: 8665845]
92. Marsh EN, Drennan CL. Adenosylcobalamin-dependent isomerases: new insights into structure and mechanism. Curr Opin Chem Biol 2001;5:499–505. [PubMed: 11578922]
93. Kolberg M, Strand KR, Graff P, Andersson KK. Structure, function, and mechanism of ribonucleotide reductases. Biochim Biophys Acta 2004;1699:1–34. [PubMed: 15158709]
94. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41:6573–6582. [PubMed: 12022860]
95. Marks, F. Protein Phosphorylation. New York, Basel, Cambridge, Tokyo: VCH Weinheim; 1996.
96. Zor T, Mayr BM, Dyson HJ, Montminy MR, Wright PE. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. J Biol Chem 2002;277:42241–42248. [PubMed: 12196545]
97. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 2004;32:1037–1049. [PubMed: 14960716]
98. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 2003;31:3635–3641. [PubMed: 12824383]
99. Sigalov AB. Multichain immune recognition receptor signaling: different players, same game? Trends Immunol 2004;25:583–589. [PubMed: 15489186]
100. Sigalov A. Multi-chain immune recognition receptors: spatial organization and signal transduction. Semin Immunol 2005;17:51–64. [PubMed: 15582488]
101. Sigalov AB. Immune cell signaling: a novel mechanistic model reveals new therapeutic targets. Trends Pharmacol Sci 2006;27:518–524. [PubMed: 16908074]
102. Sigalov A, Aivazian D, Stern L. Homooligomerization of the cytoplasmic domain of the T cell receptor zeta chain and of other proteins containing the immunoreceptor tyrosine-based activation motif. Biochemistry 2004;43:2049–2061. [PubMed: 14967045]
103. Sigalov AB, Aivazian DA, Uversky VN, Stern LJ. Lipid-Binding Activity of Intrinsically Unstructured Cytoplasmic Domains of Multichain Immune Recognition Receptor Signaling Subunits. Biochemistry 2006;45:15731–15739. [PubMed: 17176095]

104. Gould C, Wong CF. Designing specific protein kinase inhibitors: insights from computer simulations and comparative sequence/structure analysis. Pharmacol Ther 2002;93:169–178. [PubMed: 12191609]

105. Schulz, GE. Nucleotide binding proteins. In: Balaban, M., editor. Molecular mechanism of biological recognition. New York: Elsevier/North-Holland Biomedical Press; 1979. p. 79-94.

106. Han KK, Martinage A. Post-translational chemical modifications of proteins--III. Current developments in analytical procedures of identification and quantitation of post-translational chemically modified amino acid(s) and its derivatives. Int J Biochem 1993;25:957–970. [PubMed: 8365549]

107. Maurer-Stroh S, Eisenhaber F. Myristoylation of viral and bacterial proteins. Trends Microbiol 2004;12:178–185. [PubMed: 15051068]

108. Sharom FJ, Lehto MT. Glycosylphosphatidylinositol-anchored proteins: structure, function, and cleavage by phosphatidylinositol-specific phospholipase C. Biochem Cell Biol 2002;80:535–549. [PubMed: 12440695]

109. Ikezawa H. Glycosylphosphatidylinositol (GPI)-anchored proteins. Biol Pharm Bull 2002;25:409–417. [PubMed: 11995915]

110. Grobe K, Ledin J, Ringvall M, Holmborn K, Forsberg E, Esko JD, Kjellen L. Heparan sulfate and development: differential roles of the N-acetylglucosamine N-deacetylase/N-sulfotransferase isozymes. Biochim Biophys Acta 2002;1573:209–215. [PubMed: 12417402]

111. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1999;1473:4–8. [PubMed: 10580125]

112. Ben-Dor S, Esterman N, Rubin E, Sharon N. Biases and complex patterns in the residues flanking protein N-glycosylation sites. Glycobiology 2004;14:95–101. [PubMed: 14514714]

113. Lutteke T, Frank M, von der Lieth CW. Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. Carbohydr Res 2004;339:1015–1020. [PubMed: 15010309]

114. Burnier JP, Borowski M, Furie BC, Furie B. Gamma-carboxyglutamic acid. Mol Cell Biochem 1981;39:191–207. [PubMed: 6458761]

115. Furie B, Bouchard BA, Furie BC. Vitamin K-dependent biosynthesis of gamma-carboxyglutamic acid. Blood 1999;93:1798–1808. [PubMed: 10068650]

116. Atkinson RA, Evans JS, Hauschka PV, Levine BA, Meats R, Triffitt JT, Virdi AS, Williams RJ. Conformational studies of osteocalcin in solution. Eur J Biochem 1995;232:515–521. [PubMed: 7556201]

117. Abraham GN, Podell DN. Pyroglutamic acid. Non-metabolic formation, function in proteins and peptides, and characteristics of the enzymes effecting its removal. Mol Cell Biochem 1981;38 Spec No:181–190. [PubMed: 6117006]

118. Fischer WH, Spiess J. Identification of a mammalian glutaminyl cyclase converting glutaminyl into pyroglutamyl peptides. Proc Natl Acad Sci U S A 1987;84:3628–3632. [PubMed: 3473473]

119. Cummins PM, O'Connor B. Pyroglutamyl peptidase: an overview of the three known enzymatic forms. Biochim Biophys Acta 1998;1429:1–17. [PubMed: 9920379]

120. Daily, KM.; Radivojac, P.; Dunker, AK. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. San Diego, California, U.S.A.: CIBCB; 2005. p. 475-481.

121. Kerscher O, Felberbaum R, Hochstrasser M. Modification of Proteins by Ubiquitin and Ubiquitin-Like Proteins. Annu Rev Cell Dev Biol. 2006

122. Huang DT, Walden H, Duda D, Schulman BA. Ubiquitin-like protein activation. Oncogene 2004;23:1958–1971. [PubMed: 15021884]

123. Okeley NM, van der Donk WA. Novel cofactors via post-translational modifications of enzyme active sites. Chem Biol 2000;7:R159–R171. [PubMed: 10903941]

124. Mure M. Tyrosine-derived quinone cofactors. Acc Chem Res 2004;37:131–139. [PubMed: 14967060]

125. Janes SM, Mu D, Wemmer D, Smith AJ, Kaur S, Maltby D, Burlingame AL, Klinman JP. A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of bovine serum amine oxidase. Science 1990;248:981–987. [PubMed: 2111581]

126. Brazeau BJ, Johnson BJ, Wilmot CM. Copper-containing amine oxidases. Biogenesis and catalysis; a structural perspective. Arch Biochem Biophys 2004;428:22–31. [PubMed: 15234266]

127. Eklund H, Eriksson M, Uhlin U, Nordlund P, Logan D. Ribonucleotide reductase--structural studies of a radical enzyme. Biol Chem 1997;378:821–825. [PubMed: 9377477]

128. Marsh EN, Patwardhan A, Huhta MS. S-adenosylmethionine radical enzymes. Bioorg Chem 2004;32:326–340. [PubMed: 15381399]

129. Firbank S, Rogers M, Guerrero RH, Dooley DM, Halcrow MA, Phillips SE, Knowles PF, McPherson MJ. Cofactor processing in galactose oxidase. Biochem Soc Symp 2004:15–25. [PubMed: 15777009]

130. Plochocka D, Welnicki M, Zielenkiewicz P, Ostoja-Zagorski W. Three-dimensional model of the potyviral genome-linked protein. Proc Natl Acad Sci U S A 1996;93:12150–12154. [PubMed: 8901548]

131. Davidson VL. Electron transfer in quinoproteins. Arch Biochem Biophys 2004;428:32–40. [PubMed: 15234267]

132. Davidson VL. Pyrroloquinoline quinone (PQQ) from methanol dehydrogenase and tryptophan tryptophylquinone (TTQ) from methylamine dehydrogenase. Adv Protein Chem 2001;58:95–140. [PubMed: 11665494]

133. Garavelli JS. The RESID Database of Protein Modifications as a resource and annotation tool. Proteomics 2004;4:1527–1533. [PubMed: 15174122]

134. Michel H, Behr J, Harrenga A, Kannt A. Cytochrome c oxidase: structure and spectroscopy. Annu Rev Biophys Biomol Struct 1998;27:329–356. [PubMed: 9646871]

135. Farriol-Mathis N, Garavelli JS, Boeckmann B, Duvaud S, Gasteiger E, Gateau A, Veuthey AL, Bairoch A. Annotation of post-translational modifications in the Swiss-Prot knowledge base. Proteomics 2004;4:1537–1550. [PubMed: 15174124]

136. Wiederanders B. Structure-function relationships in class CA1 cysteine peptidase propeptides. Acta Biochim Pol 2003;50:691–713. [PubMed: 14515150]

137. Bernstein NK, Cherney MM, Yowell CA, Dame JB, James MN. Structural insights into the activation of P. vivax plasmepsin. J Mol Biol 2003;329:505–524. [PubMed: 12767832]

138. Seipelt J, Guarne A, Bergmann E, James M, Sommergruber W, Fita I, Skern T. The structures of picornaviral proteinases. Virus Res 1999;62:159–168. [PubMed: 10507325]

139. Tate SS, Meister A. gamma-Glutamyl transpeptidase: catalytic, structural and functional aspects. Mol Cell Biochem 1981;39:357–368. [PubMed: 6118826]

140. Okada T, Suzuki H, Wada K, Kumagai H, Fukuyama K. Crystal structures of gamma-glutamyltranspeptidase from Escherichia coli, a key enzyme in glutathione metabolism, and its reaction intermediate. Proc Natl Acad Sci U S A 2006;103:6471–6476. [PubMed: 16618936]

141. Anraku Y, Mizutani R, Satow Y. Protein splicing: its discovery and structural insight into novel chemical mechanisms. IUBMB Life 2005;57:563–574. [PubMed: 16118114]

142. Perler FB. Protein splicing mechanisms and applications. IUBMB Life 2005;57:469–476. [PubMed: 16081367]

143. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science 1990;250:651–657. [PubMed: 2146742]

144. Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedgwick SG. Protein splicing in the maturation of M. tuberculosis recA protein: a mechanism for tolerating a novel class of intervening sequence. Cell 1992;71:201–210. [PubMed: 1423588]

145. Hodges RA, Perler FB, Noren CJ, Jack WE. Protein splicing removes intervening sequences in an archaea DNA polymerase. Nucleic Acids Res 1992;20:6153–6157. [PubMed: 1475179]

146. Kawasaki M, Makino S, Matsuzawa H, Satow Y, Ohya Y, Anraku Y. Folding-dependent in vitro protein splicing of the Saccharomyces cerevisiae VMA1 protozyme. Biochem Biophys Res Commun 1996;222:827–832. [PubMed: 8651930]

147. Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E. Inteins: structure, function, and evolution. Annu Rev Microbiol 2002;56:263–287. [PubMed: 12142479]

148. Clarke ND. A proposed mechanism for the self-splicing of proteins. Proc Natl Acad Sci U S A 1994;91:11084–11088. [PubMed: 7972014]

149. Xu MQ, Perler FB. The mechanism of protein splicing and its modulation by mutation. Embo J 1996;15:5146–5153. [PubMed: 8895558]

150. Perler FB. InBase: the Intein Database. Nucleic Acids Res 2002;30:383–384. [PubMed: 11752343]

151. Dalle-Donne I, Giustarini D, Colombo R, Rossi R, Milzani A. Protein carbonylation in human diseases. Trends Mol Med 2003;9:169–176. [PubMed: 12727143]

152. Stadtman ER. Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. Annu Rev Biochem 1993;62:797–821. [PubMed: 8352601]

153. Vogt W. Oxidation of methionyl residues in proteins: tools, targets, and reversal. Free Radic Biol Med 1995;18:93–105. [PubMed: 7896176]

154. Moskovitz J, Weissbach H, Brot N. Cloning the expression of a mammalian gene involved in the reduction of methionine sulfoxide residues in proteins. Proc Natl Acad Sci U S A 1996;93:2095–2099. [PubMed: 8700890]

155. Ghezzi P. Oxidoreduction of protein thiols in redox regulation. Biochem Soc Trans 2005;33:1378–1381. [PubMed: 16246123]

156. Shiba T, Mizote H, Kaneko T, Nakajima T, Kakimoto Y. Hypusine, a new amino acid occurring in bovine brain. Isolation and structural determination. Biochim Biophys Acta 1971;244:523–531. [PubMed: 4334286]

157. Park MH, Cooper HL, Folk JE. The biosynthesis of protein-bound hypusine (N epsilon - (4-amino-2-hydroxybutyl)lysine). Lysine as the amino acid precursor and the intermediate role of deoxyhypusine (N epsilon -(4-aminobutyl)lysine). J Biol Chem 1982;257:7217–7222. [PubMed: 6806267]

158. Park MH. The post-translational synthesis of a polyamine-derived amino acid, hypusine, in the eukaryotic translation initiation factor 5A (eIF5A). J Biochem (Tokyo) 2006;139:161–169. [PubMed: 16452303]

159. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323:573–584. [PubMed: 12381310]

160. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem 2000;275:29426–29432. [PubMed: 10884388]

161. Adkins JN, Lumb KJ. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. Proteins 2002;46:1–7. [PubMed: 11746698]

162. Chang BS, Minn AJ, Muchmore SW, Fesik SW, Thompson CB. Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2. Embo J 1997;16:968–977. [PubMed: 9118958]

163. Campbell KM, Terrell AR, Laybourn PJ, Lumb KJ. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. Biochemistry 2000;39:2708–2713. [PubMed: 10704222]

164. Sanchez-Puig N, Veprintsev DB, Fersht AR. Human full-length Securin is a natively unfolded protein. Protein Sci 2005;14:1410–1418. [PubMed: 15929994]

165. Mark WY, Liao JC, Lu Y, Ayed A, Laister R, Szymczyna B, Chakrabartty A, Arrowsmith CH. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? J Mol Biol 2005;345:275–287. [PubMed: 15571721]

166. Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein intrinsic disorder and human papillomaviruses: Increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res. 2006In press

167. Burgess BR, Schuck P, Garboczi DN. Dissection of merozoite surface protein 3, a representative of a family of Plasmodium falciparum surface proteins, reveals an oligomeric and highly elongated molecule. J Biol Chem 2005;280:37236–37245. [PubMed: 16135515]

168. Nair M, Hinds MG, Coley AM, Hodder AN, Foley M, Anders RF, Norton RS. Structure of domain III of the blood-stage malaria vaccine candidate, Plasmodium falciparum apical membrane antigen 1 (AMA1). J Mol Biol 2002;322:741–753. [PubMed: 12270711]

169. Feng ZP, Keizer DW, Stevenson RA, Yao S, Babon JJ, Murphy VJ, Anders RF, Norton RS. Structure and inter-domain interactions of domain II from the blood-stage malarial protein, apical membrane antigen 1. J Mol Biol 2005;350:641–656. [PubMed: 15964019]

170. Thomson LM, Lamont DJ, Mehlert A, Barry JD, Ferguson MA. Partial structure of glutamic acid and alanine-rich protein, a major surface glycoprotein of the insect stages of Trypanosoma congolense. J Biol Chem 2002;277:48899–48904. [PubMed: 12368279]

171. Treumann A, Zitzmann N, Hulsmeier A, Prescott AR, Almond A, Sheehan J, Ferguson MA. Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of Trypanosoma brucei. J Mol Biol 1997;269:529–547. [PubMed: 9217258]

172. Butikofer P, Ruepp S, Boschung M, Roditi I. 'GPEET' procyclin is the major surface protein of procyclic culture forms of Trypanosoma brucei brucei strain 427. Biochem J 1997;326(Pt 2):415–423. [PubMed: 9291113]

173. Acosta-Serrano A, Cole RN, Mehlert A, Lee MG, Ferguson MA, Englund PT. The procyclin repertoire of Trypanosoma brucei. Identification and structural characterization of the Glu-Pro-rich polypeptides. J Biol Chem 1999;274:29763–29771. [PubMed: 10514452]

174. Turner BG, Summers MF. Structural biology of HIV. J Mol Biol 1999;285:1–32. [PubMed: 9878383]

175. Lee BM, De Guzman RN, Turner BG, Tjandra N, Summers MF. Dynamical behavior of the HIV-1 nucleocapsid protein. J Mol Biol 1998;279:633–649. [PubMed: 9641983]

176. Tan R, Frankel AD. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. Biochemistry 1994;33:14579–14585. [PubMed: 7981219]

177. Battiste JL, Tan R, Frankel AD, Williamson JR. Assignment and modeling of the Rev Response Element RNA bound to a Rev peptide using 13C-heteronuclear NMR. J Biomol NMR 1995;6:375–389. [PubMed: 8563466]

178. Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rosch P. Structural studies of HIV-1 Tat protein. J Mol Biol 1995;247:529–535. [PubMed: 7723010]

179. Shojania S, O'Neil JD. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1–72) by NMR spectroscopy. J Biol Chem 2006;281:8347–8356. [PubMed: 16423825]

180. Zaitseva M, Peden K, Golding H. HIV coreceptors: role of structure, posttranslational modifications, and internalization in viral-cell fusion and as targets for entry inhibitors. Biochim Biophys Acta 2003;1614:51–61. [PubMed: 12873765]

181. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 1998;393:648–659. [PubMed: 9641677]

182. Guilhaudis L, Jacobs A, Caffrey M. Solution structure of the HIV gp120 C5 domain. Eur J Biochem 2002;269:4860–4867. [PubMed: 12354117]

183. Sabag AD, Dagan O, Avraham KB. Connexins in hearing loss: a comprehensive overview. J Basic Clin Physiol Pharmacol 2005;16:101–116. [PubMed: 16285463]

184. Duffy HS, Sorgen PL, Girvin ME, O'Donnell P, Coombs W, Taffet SM, Delmar M, Spray DC. pH-dependent intramolecular binding and structure involving Cx43 cytoplasmic domains. J Biol Chem 2002;277:36706–36714. [PubMed: 12151412]

185. Sosinsky GE, Nicholson BJ. Structural organization of gap junction channels. Biochim Biophys Acta 2005;1711:99–125. [PubMed: 15925321]

186. Sorgen PL, Duffy HS, Sahoo P, Coombs W, Delmar M, Spray DC. Structural changes in the carboxyl terminus of the gap junction protein connexin43 indicates signaling between binding domains for c-Src and zonula occludens-1. J Biol Chem 2004;279:54695–54701. [PubMed: 15492000]

187. Giepmans BN. Gap junctions and connexin-interacting proteins. Cardiovasc Res 2004;62:233–245. [PubMed: 15094344]

188. Matsuzawa Y. The metabolic syndrome and adipocytokines. FEBS Lett 2006;580:2917–2921. [PubMed: 16674947]

189. Cheng Y, Le Gall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in proteins associated with cardiovascular disease. Biochemistry. 2006Submitted

190. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44:12454–12470. [PubMed: 16156658]

191. Oetting WS, King RA. Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism. Hum Mutat 1999;13:99–115. [PubMed: 10094567]

192. Cooksey CJ, Garratt PJ, Land EJ, Pavel S, Ramsden CA, Riley PA, Smit NP. Evidence of the indirect formation of the catecholic intermediate substrate responsible for the autoactivation kinetics of tyrosinase. J Biol Chem 1997;272:26226–26235. [PubMed: 9334191]

193. Hearing VJ, Jimenez M. Mammalian tyrosinase--the critical regulatory control point in melanocyte pigmentation. Int J Biochem 1987;19:1141–1147. [PubMed: 3125075]

194. van Gelder CW, Flurkey WH, Wichers HJ. Sequence and structural features of plant and fungal tyrosinases. Phytochemistry 1997;45:1309–1323. [PubMed: 9237394]

195. Donne DG, Viles JH, Groth D, Mehlhorn I, James TL, Cohen FE, Prusiner SB, Wright PE, Dyson HJ. Structure of the recombinant full-length hamster prion protein PrP(29#x02013;231): the N terminus is highly flexible. Proc Natl Acad Sci U S A 1997;94:13452–13457. [PubMed: 9391046]

196. Riek R, Hornemann S, Wider G, Billeter M, Glockshuber R, Wuthrich K. NMR structure of the mouse prion protein domain PrP(121–321). Nature 1996;382:180–182. [PubMed: 8700211]

197. Aronoff-Spencer E, Burns CS, Avdievich NI, Gerfen GJ, Peisach J, Antholine WE, Ball HL, Cohen FE, Prusiner SB, Millhauser GL. Identification of the Cu2+ binding sites in the N-terminal domain of the prion protein by EPR and CD spectroscopy. Biochemistry 2000;39:13760–13771. [PubMed: 11076515]

198. Burns CS, Aronoff-Spencer E, Dunham CM, Lario P, Avdievich NI, Antholine WE, Olmstead MM, Vrielink A; Gerfen GJ, Peisach J, Scott WG, Millhauser GL. Molecular features of the copper binding sites in the octarepeat domain of the prion protein. Biochemistry 2002;41:3991–4001. [PubMed: 11900542]

199. Bullock AN, Fersht AR. Rescuing the function of mutant p53. Nat Rev Cancer 2001;1:68–76. [PubMed: 11900253]

200. Hainaut P, Hollstein M. p53 and human cancer: the first ten thousand mutations. Adv Cancer Res 2000;77:81–137. [PubMed: 10549356]

201. Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci 1999;24:329–332. [PubMed: 10470028]

202. Dobson CM. Principles of protein folding, misfolding and aggregation. Semin Cell Dev Biol 2004;15:3–16. [PubMed: 15036202]

203. Uversky VN, Fink AL. Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim Biophys Acta 2004;1698:131–153. [PubMed: 15134647]

204. Uversky VN. A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. J Biomol Struct Dyn 2003;21:211–234. [PubMed: 12956606]

205. Kruger R, Kuhn W, Muller T, Woitalla D, Graeber M, Kosel S, Przuntek H, Epplen JT, Schols L, Riess O. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 1998;18:106–108. [PubMed: 9462735]

206. Zarranz JJ, Alegre J, Gomez-Esteban JC, Lezcano E, Ros R, Ampuero I, Vidal L, Hoenicka J, Rodriguez O, Atares B, Llorens V, Gomez Tortosa E, del Ser T, Munoz DG, de Yebenes JG. The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol 2004;55:164–173. [PubMed: 14755719]

207. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997;276:2045–2047. [PubMed: 9197268]

208. Li J, Uversky VN, Fink AL. Effect of familial Parkinson's disease point mutations A30P and A53T on the structural properties, aggregation, and fibrillation of human alpha-synuclein. Biochemistry 2001;40:11604–11613. [PubMed: 11560511]

Table 1

All (17) ligand keywords strongly correlated with predicted disorder

Keywords	Number of proteins	Number of families	Average sequence length	Z-score	P-value
DNA-binding	13224	1751	417.25	18.22	1
RNA-binding	9781	552	261.95	17.7	1
Zinc	10661	1135	521.74	13.49	1
rRNA-binding	5852	109	153.48	11.46	1
Metal-thiolate cluster	228	11	64.67	6.99	1
Actin-binding	817	98	785.55	6.41	1
Calmodulin-binding	503	58	986.49	6.35	1
Viral nucleoprotein	437	39	573.29	6.14	1
Heparin-binding	257	51	441.11	5.53	1
Growth factor binding	53	7	562.76	3.33	1
cGMP-binding	52	4	736.37	3.07	1
cGMP	107	7	661.02	2.96	1
Sialic acid	43	17	479.95	2.17	0.97
cAMP-binding	104	6	573.42	2.06	0.97
cAMP	180	16	556.52	2.03	0.99
IgG-binding protein	30	6	349.87	1.9	0.99
tRNA-binding	1046	22	313.94	1.78	0.96

Table 2

Top 20 of ligand keywords strongly correlated with predicted order

Keywords	Number of proteins	Number of families	Average sequence length	Z-score	P-value
Iron	9417	641	345.12	−22.78	0
NADP	3878	231	403.78	−21.72	0
Flavoprotein	2926	228	461.23	−19.39	0
NAD	5864	333	365.4	−16.48	0
FAD	2128	184	496.26	−16.47	0
Iron-sulfur	2874	227	394.78	−13.99	0
Pyridoxal phosphate	2362	81	432.38	−13.93	0
Magnesium	6186	407	441.77	−12.92	0
Plastoquinone	260	11	372.09	−12.59	0
4Fe-4S	2097	165	416.69	−12.48	0
Heme	4617	212	322.96	−10.59	0
FMN	791	74	389.7	−9.85	0
Thiamine pyrophosphate	396	16	594.21	−9.84	0
ATP-binding	19639	687	568.44	−8.95	0
Nucleotide-binding	23736	780	545.67	−8.62	0
Molybdenum	232	32	626.79	−8.51	0
Nickel	446	72	320.5	−7.34	0
Manganese	1913	194	416.88	−7.19	0
Sodium	611	81	569.15	−6.69	0
Cobalt	560	82	363.38	−6.51	0

Table 3

All (17) posttranslational modifications keywords strongly correlated with predicted disorder

Keywords	Number of proteins	Number of families	Average sequence length	Z-score	P-value
Phosphorylation	10895	1663	592.18	27.12	1
Cleavage on pair of basic residues	867	192	194.23	14.07	1
Amidation	836	457	81.19	10.83	1
Ubl conjugation	806	154	466	8.65	1
Myristate	681	71	559.39	6.97	1
Glycoprotein	16202	1993	540.4	6.84	1
Sulfation	245	72	483.69	6.42	1
Proteoglycan	189	27	748.35	6.37	1
Lipoprotein	4337	633	393.84	5.99	1
Prenylation	723	47	265.25	5.37	1
Heparan sulfate	48	10	673	4.7	1
Gamma-carboxyglutamic acid	106	25	254.63	3.86	1
Covalent protein-DNA linkage	26	8	474.74	3.37	1
Methylation	1417	101	328.37	2.85	1
GPI-anchor	590	146	460.29	2.73	1
ADP-ribosylation	150	11	437.05	1.98	0.98
Pyrrolidone carboxylic acid	795	272	223.71	1.74	0.96

Table 4

All (11) posttranslational modifications keywords strongly correlated with predicted order

Keywords	Number of proteins	Number of families	Average sequence length	Z-score	P-value
Quinone	449	17	380.29	−16.17	0
Organic radical	54	3	484.74	−8.76	0
Covalent protein-RNA linkage	108	4	2365.56	−5.63	0
TPQ	22	3	645.09	−5.35	0
PQQ	26	7	663.31	−4.21	0
Formylation	57	30	168.74	−3.66	0
Zymogen	1680	128	436.22	−3.22	0
Autocatalytic cleavage	325	51	502.94	−2.42	0
Protein splicing	84	30	1027.94	−2.08	0.03
Oxidation	32	14	295.63	−1.98	0.04
Hypusine	58	1	147.19	−1.94	0.02

Table 5

All (11) disease keywords strongly correlated with predicted disorder

Keywords	Number of proteins	Number of families	Average sequence length	Z-score	p-value
Proto-oncogene	405	135	567.12	8.86	1
Malaria	110	35	615.59	5.46	1
AIDS	405	17	371.9	4.95	1
Trypanosomiasis	26	19	431.85	4.06	1
Deafness	94	65	900.21	3.75	1
Cardiomyopathy	31	28	941.26	2.75	1
Albinism	25	14	638.28	2.71	1
Diabetes mellitus	36	28	566.14	2.33	0.99
Obesity	48	21	376.83	2.22	0.99
Prion	74	5	253.41	1.72	0.97
Disease mutation	1269	763	722.24	1.59	0.96