

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Applications of Bioinformatics and Experimental Methods to Intrinsic Disorder-Based Protein-Protein Interactions

Xiaolin Sun<sup>1</sup>, William T. Jones<sup>1</sup> and Vladimir N. Uversky<sup>2,3</sup>

<sup>1</sup>*The New Zealand Institute for Plant and Food Research, Palmerston North,*

<sup>2</sup>*Department of Molecular Medicine, College of Medicine, University of South Florida,*

<sup>3</sup>*Institute for Biological Instrumentation, Russian Academy of Sciences, Moscow,*

<sup>1</sup>*New Zealand*

<sup>2</sup>*USA*

<sup>3</sup>*Russia*

## 1. Introduction

Proteins are important to the organisms and they control biochemical pathways within the cell. Each protein or group of proteins are responsible for various functions needed to maintain the living cell, including enzymatic catalysis; transporting or storing chemical compounds and energy; hormone regulation of many processes; maintaining structure of tissues and cells; antibody immune response; signal transduction by receptors and signalling proteins etc. The functions of proteins have been long related to their rigid three-dimensional (3D) structures in that a protein's biological function depended on its prior folding into a unique 3D structure. However, as it was discovered in recent years, not all biologically functional proteins fold spontaneously into globular structures. Some protein regions or entire proteins lack stable secondary and/or tertiary structures in solution yet possess crucial biological functions, and these disordered regions or proteins are key to understanding many biological processes such as transcriptional regulation, signalling and causes of diseases. In this chapter, we will focus on some basic concepts of intrinsically disordered proteins (IDPs), the recent progress of structural and functional studies of IDP, and the bioinformatics and experimental methods practically used for investigation of IDPs.

## 2. Intrinsically disordered proteins and their structural and functional studies

### 2.1 Concept of intrinsically disordered proteins (IDPs)

For over a hundred years, the structure-function relationships of proteins have been one of the central topics of protein science. In 1894, by observing the specificity of the enzymatic hydrolysis of glucosides, Fisher put forward the lock-and-key theory of protein functionality (Fischer, 1894) according to which protein functions are determined by their specific 3D structures. This lock-and-key model for protein binding was further reinforced by many experiments on the dependence of protein function on their 3D structures (Mirsky and Pauling, 1936; Wu, 1931) and was accepted widely to underlie almost all of the subsequent

work and thinking (Phillips, 1986). The flood of protein 3D structures determined by modern structural biology through X-ray diffraction and nuclear magnetic resonance (NMR) spectroscopy has also highlighted a specific 3D structure as the necessary prerequisite for protein function (Berman *et al.*, 2000). For those proteins with rigid structures, the amino acid sequence determines the protein's unique 3D structure and the sequence → structure → function paradigm have become paramount.

However, the generality of the sequence-structure-function paradigm was challenged by the observation (Karush, 1950) that the binding site of bovine serum albumin adopts a large number of configurations with similar energy levels. Upon interacting with a substrate, the best-fitting configuration was selected from the structural ensemble of bovine serum albumin, which was called configurational adaptability. Koshland further proposed this configurational adaptability as induced-fit process (Koshland, 1958), indicating that protein conformational changes are responsible for its function. At the time induced-fit was proposed, it was unclear whether the process of binding induced a new conformation or resulted in selection of the best-fit alternative from an ensemble of structures in equilibrium. The induced-fit theory was further supported by the fact that protein domains are capable of moving upon substrate binding, which were firstly observed in binding of glucose to yeast hexokinase and following numerous protein crystal structures (Bennett and Steitz, 1978; McDonald *et al.*, 1979). One example of latter comes from crystal structures of folylpolyglutamate synthetase (FPGS) with and without substrates, a crucial enzyme to retain folic acids for normal cell growth (Sheng *et al.*, 2000; Sun *et al.*, 1998). An ATP-bound FPGS is activated by binding of folate as the second substrate, triggering a large closing movement of domains of FPGS that enables the enzyme to adopt a form for binding the third substrate, L-glutamate, and effect the addition of a polyglutamate tail to the folate (Sun *et al.*, 2001). Thus, the induced-fit hypothesis not only covers structural accommodations within the binding pocket for binding a diverse set of related but structurally distinct molecules but also includes large domain movements upon ligand binding to gain deeper understanding of structure-function relationship of proteins (Koshland, 1994). The induced-fit mechanisms of protein binding are obviously involved in protein conformational changes with either small rearrangements of few interacting groups or larger domain-domain rearrangements between alternative 3D structures in multi-domain proteins.

In contrast to the views described above that function depends strictly on prior 3D structures, or on structural accommodations within a prior 3D structure, or on domain movements between different structures, it has been discovered that unfolded regions or entire proteins play crucial roles in protein functions (Bychkova *et al.*, 1996; Daughdrill *et al.*, 1997; Riek *et al.*, 1996; Uversky *et al.*, 1996; Uversky *et al.*, 1997). This indicates the diversity of protein folding and functions, and raises intriguing questions about the role of protein disorder in biological processes. Disorder in either the binding protein and/or its partner prevents the presentation of rigid 3D structures that can be bound by other rigid complementary structures or shift between distinct states of rigid structures. These ruled out the lock-and-key and induced-fit mechanisms of binding. Obviously, the predominant sequence → 3D structure → function paradigm is no longer sufficient for many unfolded functional proteins, suggesting that a more comprehensive model is needed. In the late 1990's, studies of functional unfolded proteins emerged as a new research field of protein structure-function relationships. Several research groups simultaneously and independently made the important conclusion, after systematic research, that naturally disordered

proteins, characterized by the lack of a well-defined 3D structure under physiological conditions and existing as highly dynamic ensembles of inter-converting structures, are not just rare exceptions but represent a new and very broad class of proteins with vital biological functions (Dunker *et al.*, 2001; Dunker *et al.*, 2000; Tompa, 2003; Uversky, 2002a; Uversky *et al.*, 2000; Wright and Dyson, 1999). This important conclusion was reached from different starting points using different experimental and theoretical approaches, including bioinformatics, NMR spectroscopy, multi-parametric protein folding and misfolding studies and protein structural characterization. Since then, a new protein structure-function paradigm has been established to include the novel functions of disordered proteins. The discovery and characterization of functional unfolded proteins is one of the fastest growing areas of protein science and the literature of studies on these proteins has been increasing continually and has become especially rapid during the past decade (Dunker *et al.*, 2007; Uversky, 2010). These functional unfolded proteins are known by different names, including intrinsically disordered, natively denatured, natively unfolded, intrinsically unstructured, natively disordered, inherently disordered and now are widely known as intrinsically disordered proteins (IDPs) among other names. Those proteins either forming crystals without partners or possessing ordered globular forms without partners in NMR experiments, will be termed here as “structured”, “natively folded”, or just “ordered”.

## 2.2 The protein quartet model

Historically, the protein structure-function paradigm emphasized the role of a rigid 3D structure as being a necessary prerequisite to protein function. We now know from the cumulated experimental data on intrinsic disorder that the functional protein or protein region can exist as a structural ensemble, at either secondary or tertiary level. Both unfolded regions with a little elements of secondary structure (random coils) and collapsed tertiary structures with poorly packed side chains (molten globule-like) are included in the range of intrinsic disorder. These ideas were presented as the Protein Trinity Paradigm (Dunker *et al.*, 2001; Dunker and Obradovic, 2001). The Protein Trinity Paradigm relates protein function to the three thermodynamic states of protein. In other words, the intracellular functional proteins or regions of such proteins can exist in any one of the three thermodynamic states, namely, ordered forms, molten globules, and random coils. Thus, a particular function is proposed to arise from any one of these states or a transition between any two of the states. According to this view, the native state of a protein is not just the ordered state, but any of the three states. The molten globule was initially discovered as an equilibrium structure observed in studies of protein denaturation in which the partially unfolded intermediates between the ordered state and the random coil were observed as the major species in urea, guanidine and pH titration studies. In these experiments, the protein converted from an ordered native state into a form having some liquid like characteristics with the side chains changing from rigid to non-rigid packing, while its secondary structure remains almost unchanged and the shape remains compact (Ohgushi and Wada, 1983; Ptitsyn and Uversky, 1994). The molten globule has been proposed to be responsible for biological functions. For example, molten globular state was reported to be involved in the process of translocation of proteins across membranes (Bychkova *et al.*, 1988) and the transfer of retinal from its bloodstream carrier to its cell-surface receptor (Bychkova *et al.*, 1992; Bychkova *et al.*, 1998).

By summarizing a large number of experimental results on the conformational behaviours of IDPs, one of us further revealed that the extended disordered region or entire proteins

did not possess uniform structural properties as random coils. They were split into two structurally different subclasses, intrinsic random coil-like and intrinsic pre-molten globule-like conformations (Uversky, 2002a). Proteins in pre-molten globule state are more compact than random coil, exhibiting some amount of residual secondary structure, but they are still essentially less dense than molten globule and ordered proteins. It was also noted that molten globule and pre-molten globule (as folding intermediates of globular proteins) might represent different phase states of protein, as they are separated by the first-order phase transition (Uversky, 1997; Uversky and Ptitsyn, 1994, 1996). These observations introduced the native pre-molten globule state of functional unfolded proteins, a new player on the protein functioning field. As ordered, molten globule, pre-molten globule, and random coil conformations possess clearly defined structural differences, the Protein Trinity Paradigm has been extended to the Protein Quartet Model with protein functions arising from four specific conformations (native ordered, molten globules, pre-molten globules and random coils) and transitions between any two of these states (Fig. 1). All of these four structurally defined protein native states can be characterized by using various experimental approaches and applications. For example, pre-molten globules with some residual secondary structure can be characterized by far-UV CD spectra as a typical disordered polypeptide chain with a pronounced minimum in the vicinity of 200 nm. See following sections for more information.

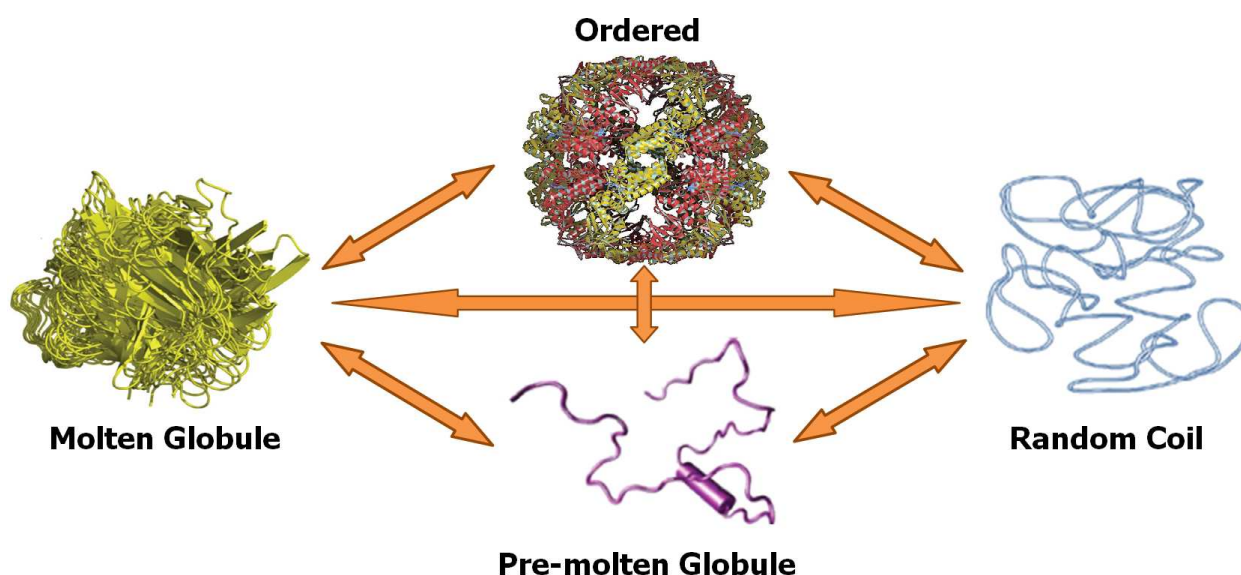


Fig. 1. The protein quartet model

### 2.3 Functional features of IDPs

Given that the intrinsic disorder represents an important structural class of proteins, and in order to meet the increasing interests in systemizing the crucial functions of IDPs, a database of disordered proteins (DisProt) has been created (Vucetic *et al.*, 2005). DisProt provides structural and functional information on proteins or regions that lack a rigid 3D structure under putatively native conditions. Verified by X-ray diffraction, NMR and CD spectra and several other biophysical techniques, each disordered protein included in the database is given the name, various aliases, accession code, amino acid sequence, location of the disordered region(s) and methods used for structural (disorder) characterization. Most



entries list the biological function(s) of each disordered region or protein, if applicable. To date, there are 643 IDPs and 1375 intrinsically disordered regions (IDRs) listed in DisProt.

Among the rapidly increasing number of publications on IDPs, bioinformatics studies have predicted that about 25-30% of eukaryotic proteins are mostly disordered (Oldfield *et al.*, 2005a); more than 50% eukaryotic proteins have long regions of disorder (Dunker *et al.*, 2000; Oldfield *et al.*, 2005a); and more than 70% of signalling proteins and the vast majority of cancer associated proteins have long disordered regions (Iakoucheva *et al.*, 2002); 82-94% of transcription factors from three transcription factor datasets possess extended disordered regions (Liu *et al.*, 2006). IDPs are now widely accepted as ubiquitously existing in all kingdoms of life (Dunker *et al.*, 2000; Ward *et al.*, 2004). Since IDPs and IDRs have amazing conformational variability with a variety of functions, the terms “unfoldome” and “unfoldomics” have been recently introduced (Cortese *et al.*, 2005; Dunker *et al.*, 2007; Midic *et al.*, 2009). Unfoldome is attributed to a large set of functional IDPs and disordered regions within the proteome while Unfoldomics deals with both the identification of the set of proteins or regions in the unfoldome of a given organism and their functions, structures, interactions and evolution (Uversky *et al.*, 2009).

Literature search and comprehensive survey on functions of IDPs characterized by using different experiments suggested that IDPs or IDRs fall into broad functional classes including: (i) entropic chains activities stemming directly from disorder; (ii) molecular recognition via binding to other proteins or to nucleic acids; (iii) scavengers which store and/or neutralize small ligands; (iv) molecular assembly which assemble, stabilize and regulate large multi-protein complexes; (v) various protein modifications (acetylation, hydroxylation, ubiquitination, methylation, phosphorylation) and proteolysis etc. (Dunker *et al.*, 2002a; Dunker *et al.*, 2002b; Dyson and Wright, 2005b; Tompa, 2002; Uversky, 2010). Some illustrative biological functions of IDPs have also been collected in numerous literatures including cell division regulation, transcriptional and translational regulation, molecular chaperoning and cell signalling etc. (Dunker *et al.*, 2005; Dunker *et al.*, 2001; Radivojac *et al.*, 2007; Tompa, 2005; Wright and Dyson, 1999). Obviously, structural flexibility and plasticity originating from the lack of a rigid 3D structure probably represents a major functional advantage for IDPs, reflecting from the fact that IDPs or IDRs can interact with a broad range of binding partners including protein, membranes, nucleic acids, and small molecules (Oldfield *et al.*, 2008; Tompa and Csermely, 2004).

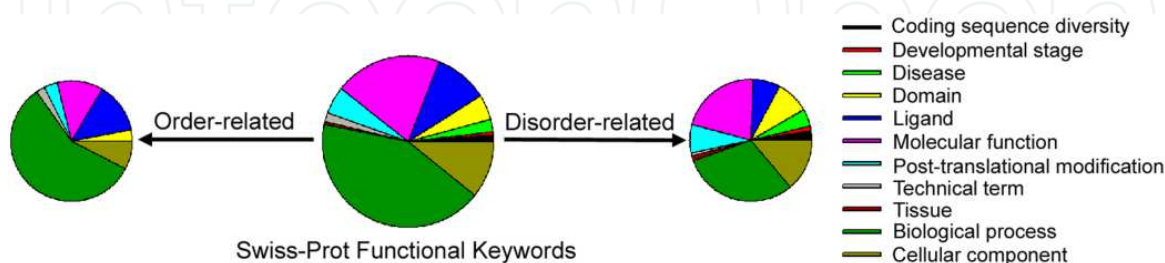


Fig. 2. Functional anthology of IDPs. Modified from Uversky *et al.* (2009) *Bmc Genomics*, 10 (Suppl 1): S7.

Specifically, a majority of the IDPs or IDRs characterized are involved in regulation, cell-signalling, and control pathways via interactions with multiple partners using high specificity/low affinity strategy, and such disordered regions often become folded upon

binding to their partners, confirming that prior 3D structure is not required for molecular recognition (Dunker *et al.*, 2005; Dyson and Wright, 2002; Haynes *et al.*, 2006; Uversky *et al.*, 2005). The crucial role of IDPs in cell-signalling is further confirmed by the finding that eukaryotic proteomes which have developed extensive interaction networks possess higher frequency of IDPs than bacteria and archaea (Dunker *et al.*, 2000; Ward *et al.*, 2004). The close correlation between IDPs and cell-signalling implies that IDPs would play a critical role in protein interaction networks (see following section for details).

Although functional studies of the IDPs characterized experimentally have made significant progress in revealing the functional diversity of IDPs, they are bound to provide only an incomplete view due to limited number of known IDPs. By taking advantage of bioinformatics methods, statistical approaches utilizing a novel data mining tool have been used for a comprehensive study of functional roles of IDPs or IDRs from Swiss-Prot database containing over 200,000 proteins. In the mean time, at least one illustrative and experimentally validated example of functional disorder or order for the vast majority of functional keywords were found to support the bioinformatics analyses. These studies represent a functional anthology of IDPs or IDRs and provide researchers with a novel theoretical tool that could be used to strengthen the understanding of functional diversity of IDPs and protein structure-function relationships (Vucetic *et al.*, 2007; Xie *et al.*, 2007a; Xie *et al.*, 2007b). In these studies, it was shown that many protein functions are associated with long disordered regions; the 262 of 710 Swiss-Prot functional keywords were found to be strongly positively correlated with long IDRs; whereas 302 were strongly negatively correlated with such regions. Those Swiss-Prot functional keywords used in the analyses are associated with various biological processes, cellular components, domains, technical terms, developmental processes, coding sequence diversities, ligands, molecular function, post-translational modifications, tissue and diseases. When all of the functional keywords were classified into eleven functional categories, disorder-associated keywords were found for all eleven categories while order-associated keywords were found only for seven of the eleven categories (Vucetic *et al.*, 2007; Xie *et al.*, 2007b) (Fig. 2). Among them, coding sequence diversities, developmental processes, diseases and tissue functional keywords were exclusively strongly correlated with IDPs or IDRs. Therefore, the functional diversity provided by disordered regions complements functions of proteins with ordered structures, implying that IDPs and IDRs are characterized by a wide functional repertoire.

## 2.4 Protein-protein interactions involving IDPs

One of the most important functions of IDPs and IDRs is molecular recognition in regulatory and cell-signalling processes via protein-protein interactions (Dunker and Obradovic, 2001; Wright and Dyson, 1999). Protein - protein interactions are organized into complex networks which are central to many processes in regard to the physiology and function of cells. Studies of the protein interaction map proposed that most proteins interact with just a few partners and a small number of proteins interact with many partners. Such a small number of proteins, called hubs, represent a few highly connected nodes in the protein-protein interaction networks (Rual *et al.*, 2005; Stelzl *et al.*, 2005). Hubs can interact with multiple partners to connect various biological molecules in the network either simultaneously (party hubs) or at different times and locations (date hubs). It has been suggested that date hubs organize the proteome, connecting biological processes to each other, whereas party hubs act inside functional modules, forming scaffolds for various

molecular machines or coordinated processes (Han *et al.*, 2004). With their ability to interact with multiple partners, hubs play a central role in various cellular biological processes by defining the properties of the protein interaction network (Barabasi and Oltvai, 2004). The high level of hub connectivity should be reflected in protein structures that render hub proteins the ability to carry out highly specific interactions with multiple, structurally diverse partners. Statistical investigation on protein interaction databases and bioinformatics studies have revealed that intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes and disordered domains confer hubs with the ability to interact with multiple structurally diverse partners in interaction networks (Dunker *et al.*, 2005; Haynes *et al.*, 2006; Patil and Nakamura, 2006). IDRs provide hubs the required binding promiscuity to interact with a large number of small molecules, proteins or nucleic acids. Alternatively, structured hubs bind to disordered regions in their many interaction partners (Oldfield *et al.*, 2008).

The abilities to interact with multiple partners (binding promiscuity) and to carry out binding-induced folding to accommodate diverse binding sites of different partners (binding plasticity) make IDPs and IDRs central in signalling and functional regulation of the cells (Uversky *et al.*, 2005). The p53 protein, regulating more than 150 genes and binding to over 100 partners (Zhao *et al.*, 2000), represents a typical example showing that intrinsic disorder is critical for function through binding promiscuity and binding plasticity (Oldfield *et al.*, 2008). Such protein interactions involving p53 known as one-to-many binding mode are illustrated in Fig. 3, in which the interactions with ten partners are mediated by protein regions experimentally confirmed as IDRs (Uversky *et al.*, 2009). Fig. 3 also indicates that protein binding sites can be predicted, using disorder predictor PONDR®VL-XT, to correspond to some short rigid regions (downward spikes) within predicted long regions of disorder. Actually, short rigid segments within long disordered regions are subject to bioinformatics screening to be the potential protein binding sites known as Molecular Recognition Features (MoRFs).

In addition to promote binding diversity by interacting with numerous partners, molecular recognition involving IDPs provides other important functional advantages over globular proteins with 3D structure for signalling and regulation: disordered regions can bind their partners with high specificity and low affinity. To permit specific recognition, disordered regions usually undergo binding-induced folding during protein interactions (Dyson and Wright, 2002, 2005b), involving a disorder-to-order transition in which IDPs or IDRs adopt a highly structured conformation upon binding to their biological partners. This binding-induced folding can occur for the whole IDPs, or large or short IDRs. It is known that a large decrease in conformational entropy due to folding of disordered regions in the disorder-to-order transition can uncouple specificity from binding strength (Dunker *et al.*, 2001; Schulz, 1979). With such a high specificity/low affinity, the regulatory interaction between an IDP and its partner is both highly specific and easily dispersed – activating and terminating a signal are equally important (Dunker *et al.*, 2002a). An IDP has been suggested to contain a “conformational preference” for the structure it will take upon binding (Fuxreiter *et al.*, 2004). This preferred conformation could be  $\alpha$ -helix ( $\alpha$ -MoRFs) (Cheng *et al.*, 2007),  $\beta$ -strand ( $\beta$ -MoRFs) or an irregular structure ( $\iota$ -MoRFs) (Mohan *et al.*, 2006; Vacic *et al.*, 2007a). Fig. 3 shows these different conformational preferences in that a single intrinsic disordered region of p53 (residues 374 - 388 in the C-terminal regulatory domain) forms all three major secondary structure types in the bound state:  $\alpha$ -helix when associating with S100 $\beta$ , a  $\beta$ -



sheet with sirtuin and different irregular structures with CBP and cyclin A. The set of residues involved in these interactions exhibit a high extent of overlap along the sequence. However, p53 utilizes different residues for the interactions with four different binding partners, suggesting that the same intrinsic disordered region sequence is induced by the different partners in entirely different ways (Oldfield *et al.*, 2008).

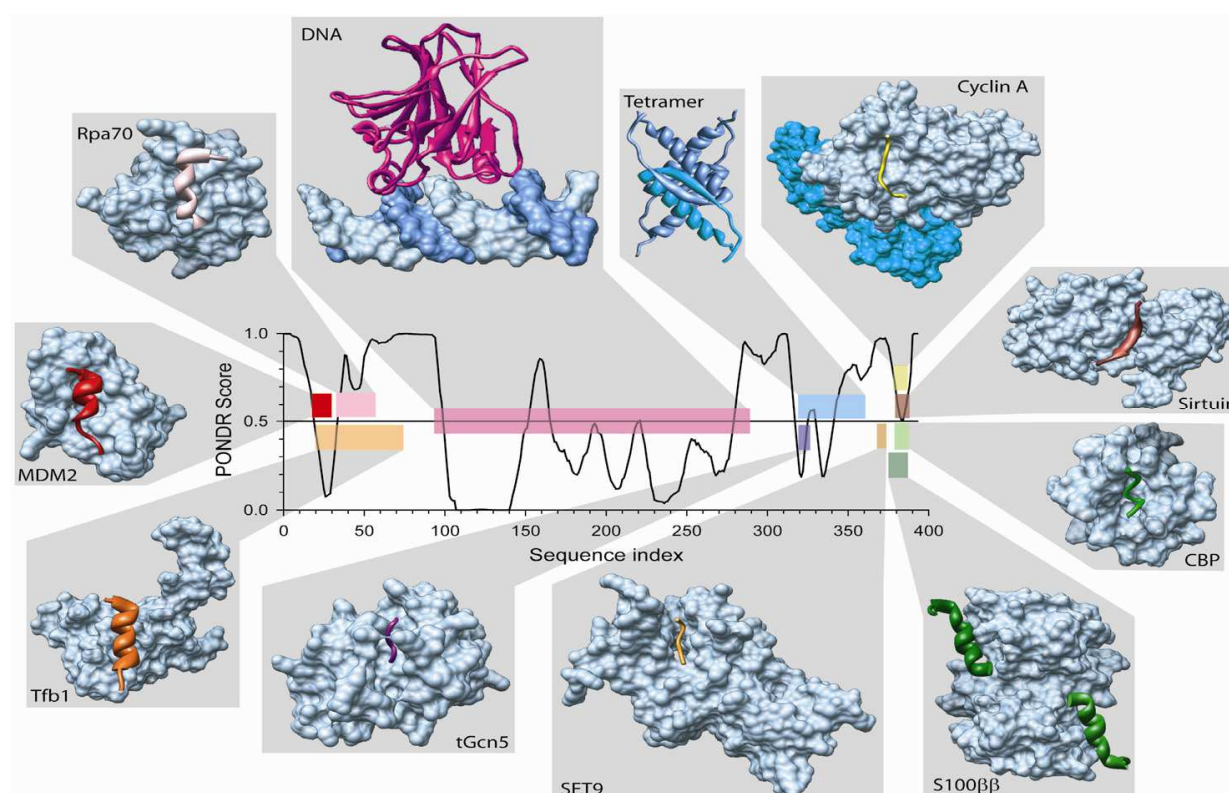


Fig. 3. The protein interactions involving p53. PONDR scores of intrinsic disorder was predicted by PONDR®VL-XT predictor. Residues with scores above 0.5 (threshold) are disordered and those below 0.5 are ordered. The N-domain (residues 1-100) and C-domain (residues 290-390) were predicted as disordered regions while the central DNA binding domain was ordered. The ten binding sites in both N- and C-domains of p53 are at or near the downward spikes in the plot of disorder scores. The complex structures containing various p53 binding regions are displayed around the predicted disorder pattern. In complexes, the structures of p53 segments bound to their partners are shown in different colours. And the same colours are used for the bars in the plot of disorder scores to indicate the positions of the segments in the sequence of p53. The Protein Data Bank IDs and partner names for complex structures are as follows: (1tsr DNA), (1q2d tGcn5), (3sak p53 (tet dom)), (1xqh set9), (1h26 cyclinA), (1ma3 sirtuin), (1jsp CBP bromo domain), (1dt7 s100ββ), (2gs0 Tfb1), (1ycr MDM2), and (2b3g rpa70). Reproduced from Uversky *et al.* (2009) *Bmc Genomics*, 10 (Suppl 1): S7.

### 3. Bioinformatics methods for predicting structures of IDPs

Bioinformatics has contributed greatly to the studies of IDPs. Driven by a rapidly increasing number of experimentally verified IDPs in late 1990's, the bioinformatics research of IDPs have promoted the correlation between protein sequence analyses and characterization of

intrinsic disorder, and made it possible to investigate the intrinsic disorder nature of proteins from large databases such as interactomes, genomes and Swiss-Prot database. Bioinformatics analyses of IDPs have provided a conceptual framework for experimental studies of molecular function and protein-protein interactions (Sun *et al.*, 2011). Below we will focus on some basic sequence analysis tools for prediction of IDPs or IDRs.

### 3.1 Amino acids compositional profile of IDPs

The propensity for disorder is encoded in the peculiarities of protein amino acid sequences. By comparing the compositions of the disordered protein datasets with each other and with ordered protein datasets, it was found that IDPs are generally enriched in polar and charged residues and are depleted of hydrophobic residues except for proline (Dunker *et al.*, 2001; Uversky *et al.*, 2000). The relative fractional differences in composition for each amino acid residue between the studied set and a set of ordered proteins are calculated as  $(C_x - C_{\text{order}})/C_{\text{order}}$ , where  $C_x$  is the percentage of a given amino acid in the studied set, and  $C_{\text{order}}$  is the corresponding percentage in a set of ordered proteins. The compositional profile can be visualized by plotting this relative fractional difference in composition against each of twenty amino acids of protein (Dunker *et al.*, 2001; Vacic *et al.*, 2007b). Thus, in the studied set, negative peaks correspond to the amino acids which are depleted in comparison to the set of ordered proteins, and positive peaks indicate the amino acids which are enriched. Most IDPs are substantially depleted in amino acids W, C, F, I, Y, V, L, H, T and N (order-promoting residues) and enriched in amino acids K, E, P, S, Q, R, D and M (disorder-promoting residues). Amino acids A and G are neutral in regards to order and disorder (Radivojac *et al.*, 2007). For the order-promoting residues, the hydrophobic (I, L, and V) and aromatic amino acid residues (W, Y and F) normally form the hydrophobic core of an ordered globular protein. The order-promoting cysteine is known to have a significant contribution to the protein conformational stability via disulfide bond formation or being involved in coordination of different prosthetic groups. On the other hand, disorder-promoting residues (R, Q, S, E, D and K) are polar and charged, i.e. their abundance defines a large net charge of an IDP at physiological pH. Although disorder-promoting proline is hydrophobic, it is well known as a structure terminator (Romero *et al.*, 2001). These compositional biases of IDPs are characterized as low overall hydrophobicity and high net charge (Uversky *et al.*, 2000), and widely used as one of the criteria for IDPs prediction.

The DELLA proteins (DELLAs), a plant-specific protein family, function as repressors of gibberellin (GA)-responsive plant growth and are the key regulatory targets in the GA signalling pathway. The N-domains of DELLAs have been experimentally verified to be intrinsically disordered, and play an important role in molecular recognition (Sun *et al.*, 2008; Sun *et al.*, 2010). Similar to disordered proteins from the DisProt database (Sickmeier *et al.*, 2007), the N-domains of DELLAs showed an overall lack of order-promoting residues and enrichment in disorder-promoting residues, in particular S, M and D, a characteristic of IDPs (Fig. 4). A special feature of the N-domains of DELLAs was a depletion of K, Q and R residues, indicating that DELLAs are a special group of IDPs lacking these three disorder-promoting residues. This analysis can also be performed using a web Composition Profiler tool (<http://www.cprofiler.org/>) which automates composition profiling with graphical output (Vacic *et al.*, 2007b). There are four different background datasets available, including both intrinsic disordered and ordered datasets, as a comparison to the query sequences.

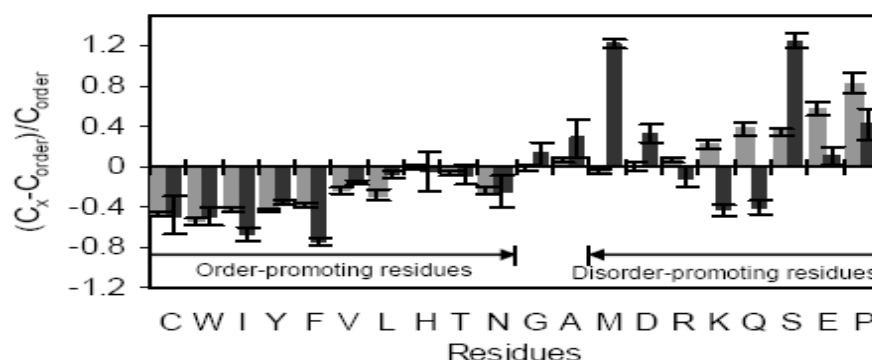


Fig. 4. The compositional profile of the N-domains of eight DELLA proteins (black bars) and disordered proteins of Dis-Pro database (grey bars) in comparison to the ordered globular proteins from the protein data bank. The eight DELLA proteins are from *Arabidopsis* (AtGAI, AtRGA, AtRGL1, AtRGL2 and AtRGL3), rice (SLR1), barley (SLN1) and wheat (RHT1).

### 3.2 Low sequence complexity of IDPs

The proteins with long disordered regions exhibit a close relationship with low sequence complexity. An investigation of Swiss-Prot database for both low sequence complexity and long disorder regions showed that nearly all the identified low-complexity segments are also predicted as disordered (Dunker *et al.*, 2001). In addition, IDPs all exhibit lower sequence complexity compared to, but partly overlapping with, the distribution of the sequence complexity for ordered proteins (Romero *et al.*, 2001). It was further revealed that IDPs or IDRs and low complexity sequences have similar compositional bias – more disorder-promoting residues (R, K, E, P, and S) and less order-promoting residues (C, W, Y, I, and V). Therefore the low sequence complexity is frequently accompanied by disorder, though not exclusively since low sequence complexity can sometimes occur in structurally ordered proteins (Romero *et al.*, 2001). Overall, simultaneous use of sequence complexity analysis and other disorder predictions will provide a better view of protein disorder.

As an example, plant-specific GRAS proteins, including DELLA subfamily, play critical roles in plant development and various signalling processes (Sun *et al.*, 2011; Sun *et al.*, 2010). One common feature of GRAS proteins is that all of the N-domains contain homopolymeric stretches of certain amino acid residues such as S, T, P, Q, G, D or A. Most of these amino acids are disorder-promoting residues and observed in low sequence complexity segments. By using an iterative algorithm for the complexity analysis of sequence (CAST) (Promponas *et al.*, 2000), the segments with low complexity in all GRAS proteins are mostly located within the N-domains which have previously been proven both experimentally and theoretically to be intrinsically disordered (Sun *et al.*, 2011; Sun *et al.*, 2010).

### 3.3 Charge-hydrophobicity (CH) and cumulative distribution function (CDF) plots

IDPs, as shown in the compositional profile, are characterized to have low overall hydrophobicity and high net charge. The combination of low mean hydrophobicity and high net charge may represent a prerequisite under physiological conditions for lack of folding in some kinds of IDPs. Statistical analysis of both intrinsically ordered and disordered protein datasets resulted in a plot of the net charge of a protein against its mean

hydropathy (CH-plot), showing that ordered and disordered proteins tend to occupy two different areas within the charge-hydrophobicity phase space, separated by a linear boundary line (Uversky *et al.*, 2000):  $\langle R \rangle = 2.785 \langle H \rangle - 1.151$ , where the mean net charge  $\langle R \rangle$  of the protein is calculated as the absolute value of the difference between the numbers of positively charged and negatively charged residues divided by the total number of amino acids, the mean hydrophobicity  $\langle H \rangle$  is defined as the sum of the normalized hydrophobicity (Kyte and Doolittle approximation with a window size of 5 and normalization on the scale from 0 to 1) of all residues divided by the total number of residues minus 4. Figure 5A represents the original charge-hydrophobicity phase space; an IDP with a given mean net charge will most likely locate above the green boundary line. Further statistics with a wider range of IDPs has shown that the mean net charge and mean hydrophobicity of IDPs can be scattered over the charge-hydrophobicity phase space, and sometimes cross into the area of ordered proteins (Oldfield *et al.*, 2005a). Therefore, an added boundary margin allowed the accuracy of the estimation to reach to 95%. As an example of CH-plot analysis, four of the N-domains of eight DELLAs fit into the disordered area but all of them are located within a boundary margin (Fig. 5B) (Sun *et al.*, 2010).

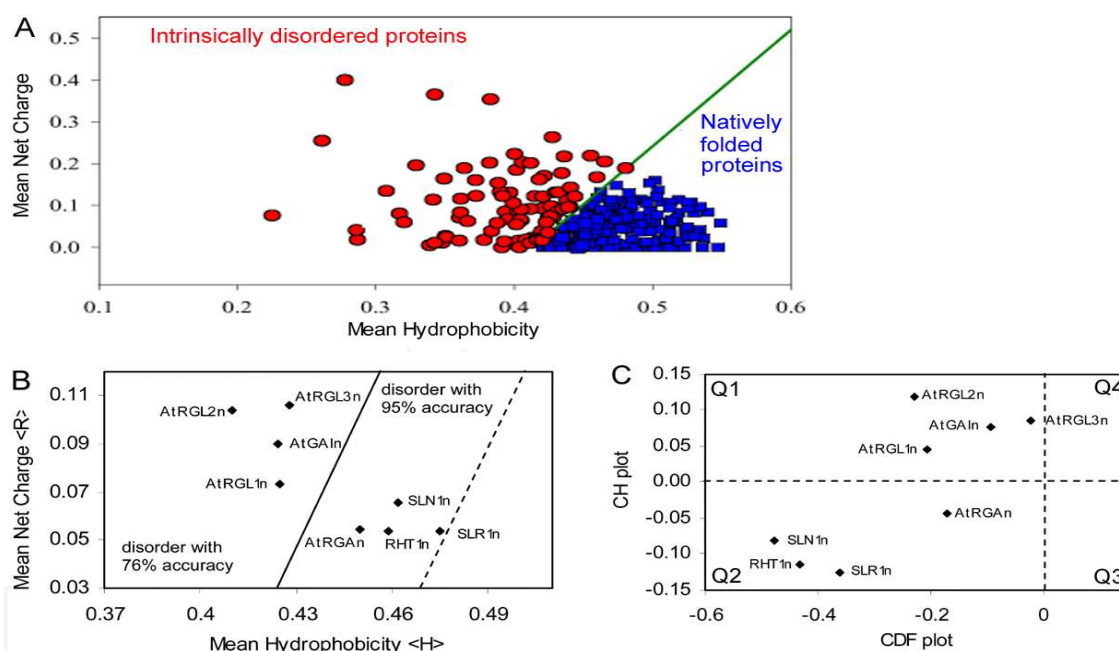


Fig. 5. (A) Mean net charge versus mean hydrophobicity plot for the set of 275 folded (blue squares) and 91 natively unfolded proteins (red circles). (B) Mean net charge versus mean hydrophobicity plot of the N-domains of eight DELLAs. A boundary margin of +0.045 (dotted) extends the disorder estimation accuracy to 95%. (C) Combined CDF-CH plot of the N-domains of eight DELLAs. Modified from Sun *et al.* (2010) *J. Biol. Chem.*, 285, 11557-11571.

A combination of low hydrophobicity and high net charge as a prerequisite for IDPs can be explained from a physical viewpoint in that high net charge leads to charge-charge repulsion, and low hydrophobicity indicates less driving force for protein compaction. However, the CH-plot is a linear disorder classifier that takes into account only two parameters of the particular sequence - charge and hydrophobicity - and is predisposed to discriminate proteins with substantial amounts of extended disorder (random coils and pre-molten globules) from proteins with globular conformations (molten globule-like and rigid



well-structured proteins) (Oldfield *et al.*, 2005a; Uversky *et al.*, 2000). Another binary disorder classifier, cumulative distribution function (CDF) analysis, discriminates all disordered conformations, including molten globules, from rigid well-folded proteins (Oldfield *et al.*, 2005a; Xue *et al.*, 2009). Therefore, simultaneous CDF-CH plot analysis gives a more accurate prediction for a wider range of sequences. CDF is a cumulated histogram of disordered residues at various disordered scores that are obtained from the disorder predictor PONDR-VSL2 (Peng *et al.*, 2006) (see next section for details). The cumulated histogram for structured proteins increases faster in the range of smaller disordered scores and then flattens at larger disordered scores. The cumulated histogram for disordered proteins increases slightly in the range of lower disordered scores but significantly at higher disordered scores. So, there is also a boundary line identified in the CDF plot. The distances to the boundary lines in both CH and CDF plots for a specific protein are further used to build up the CH-CDF plot. In the resulting CH-CDF plot, coordinates of each spot are calculated as a distance of the corresponding protein in the CH-plot from the boundary (as Y-coordinate in the CH-CDF plot) and an average distance of the respective cumulative distribution function (CDF) curve from the CDF boundary (as X-coordinate in the CH-CDF plot). Positive and negative Y values in the CH-CDF plot correspond to proteins predicted within CH-plot analysis to be intrinsically disordered or ordered, respectively. In contrast, positive and negative X values are attributed to proteins predicted within CDF analysis to be ordered or intrinsically disordered, respectively. Thus, the resultant quadrants of CDF-CH phase space correspond to the following expectations: Q1, proteins predicted to be disordered by both methods; Q2, proteins predicted to be disordered by CDFs but compact by CH-plots (i.e., putative molten globules); Q3, ordered proteins; Q4, proteins predicted to be disordered by CH-plots, but ordered by CDFs. All of the N-domains of eight DELLAs are, located in Q1 and Q2 quadrants (Fig. 5C), intrinsically disordered with different levels of compactness, which is consistent with physical and biological evidence (Sun *et al.*, 2010).

### 3.4 Prediction of intrinsic disorder and potential binding sites

Disorder prediction has been one of important approaches in IDPs and IDRs research. It is a powerful tool for study of IDPs, especially considering time and cost, compared to the experimental methods. Furthermore, it can be easily used to investigate large datasets such as proteome, interactome etc. So far, more than fifty predictors of disorder have been developed to evaluate intrinsic disorder of a given sequence on a per-residue basis (He *et al.*, 2009). These predictors utilize biased amino acid compositions of IDRs, various datasets derived from experiments and different computing techniques, and many of them are accessible on public servers. A partial list of various predictors can be found on DisProt website (<http://www.disprot.org/predictors.php>). In this chapter, we only focus on a series of Predictors Of Natural Disordered Regions (PONDRs).

A basic strategy for developing these predictors includes constructing training datasets of both ordered and disordered segments, selecting sequence attributes of order and disorder and applying a neural networks model in training. Predictor PONDR®VL-XT applies three different neural networks, one trained on Variously characterized Long disordered regions for internal region of the sequence and two trained on X-ray characterized Terminal disordered regions for N- and C-terminal regions ( $\geq 5$  amino acids) (Romero *et al.*, 2001). The PONDR®VL-XT has outputs from the first to the last residue in a sequence, and furthermore it provides the basis for CDF plot and potential binding site (MoRFs) predictions. PONDR®

VSL2 combines neural networks for both short ( $\leq 30$  residues) and long disordered regions with each neural network trained by the dataset of that specific length. This predictor gives relatively higher accuracy of prediction in the PONDR series (Peng *et al.*, 2006). PONDR®VL3 applies ten neural networks and selects the final prediction by simple majority voting. The input features of these predictors are various sequence profiles. This predictor has higher accuracy in predicting longer disordered regions. All of these PONDR predictors have relatively high accuracy ( $> 80\%$ ), and the accuracy has been further improved by a meta-predictor PONDR-FIT that combines PONDR-VLXT, PONDR-VSL2, PONDR-VL3 and other three different predictors and is so far one of the predictors with higher accuracy of disorder prediction (Xue *et al.*, 2010).

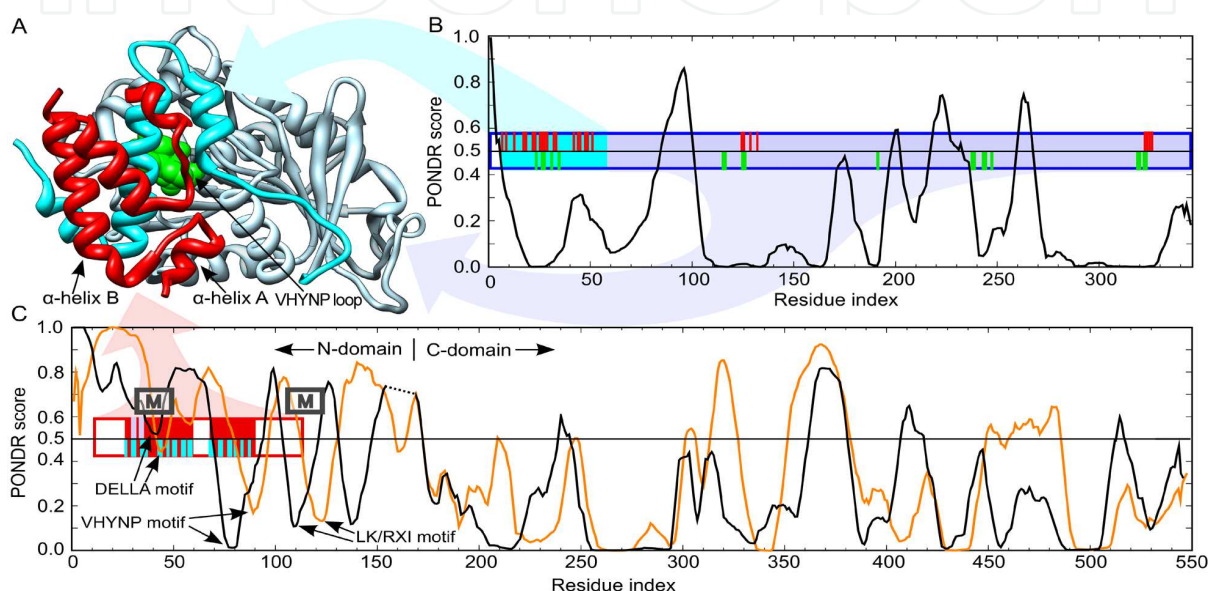


Fig. 6. PONDR disorder predictions for AtGID1a, AtGAI and AtRGL2. **(A)** X-ray crystal structure of AtGAI-AtGID1a/GA<sub>3</sub> complex (PDB 2ZSH) with ribbon representation of AtGAI (red), the N-terminal binding pocket of AtGID1a (cyan), GA receptor domain of AtGID1a (light blue) and GA<sub>3</sub> (green van der Waals surface) **(B)** Disorder prediction for AtGID1a **(C)** Disorder predictions for AtGAI (black line, its C-domain is shifted right to align with that of AtRGL2) and AtRGL2 (orange line). Disorder predictions were made with PONDR®VL-XT with a threshold 0.5 ( $\geq 0.5$  for disorder and  $< 0.5$  for order). Boxes indicate the fragments of AtGAI and AtGID1a crystallized in the complex, with filled positions indicating region of defined density in the crystal structure. Ticks indicate residues that interact with the portion of the complex with the corresponding colours. Approximate positions of  $\alpha$ -MoRFs predicted for most DELLA proteins are indicated by boxes labelled 'M'. Reproduced from Sun *et al.* (2010) *J. Biol. Chem.*, 285, 11557-11571.

As an example, PONDR®VL-XT was used to predict the disorders for two DELLAs (AtGAI and AtRGL2) together with the GA-receptor (AtGID1a) (Fig. 6). The PONDR scores of AtGID1a indicated a folded structure (Fig. 6B). The PONDR scores of AtGAI and AtRGL2, similar to each other, indicated that the C-domains of both DELLAs are dominated by ordered structures with most residues having PONDR score  $< 0.5$ , the threshold for order/disorder (Fig. 6C). In contrast, the N-domains of both DELLAs, AtGAI and AtRGL2n, are clearly intrinsically disordered except for some short rigid segments corresponding to the DELLA, VHYNP and LK/RXI motifs (Fig. 6C, indicated by arrows).

These disorder predictions of the N-domains AtGAI<sub>n</sub> and AtRGL2<sub>n</sub> are consistent with the results from compositional profile (Fig. 4) and CH-CDF plot analysis (Fig. 5).

As discussed in Section 2.4 (protein-protein interactions involving IDPs), such short ordered segments within long disordered regions detected as downward spikes in disorder prediction using PONDR®VL-XT (Fig. 6C) are termed **Molecular Recognition Features** (MoRFs). They are potential binding sites for protein interactions, and responsible for molecular recognition via a disorder-to-order transition upon binding to their interacting partners. By utilizing this specific sequence pattern of IDPs, a unique bioinformatics tool dedicated to the identification of MoRFs or potential protein-protein interaction sites in IDPs has been developed.  $\alpha$ -MoRFs-I and its updated form  $\alpha$ -MoRFs-II, the identifiers of  $\alpha$ -helix forming MoRFs, are focused on short binding regions within long regions of disorder that are likely to form helical structure upon binding (Cheng *et al.*, 2007; Oldfield *et al.*, 2005 b). The  $\alpha$ -MoRFs predictor defines a heuristic for binding-associated downward spikes and removes false positive predictions. It assigns relatively short segments (20 residues) that gain functionality through a disorder-to-order transition induced upon binding to a partner. The identifiers of  $\beta$ -sheet or irregular structure forming MoRFs are under development.

As an example,  $\alpha$ -MoRFs have been predicted at or near the DELLA motif and the LK/RXI motif in the N-domains of DELLAs (Fig. 6C), suggesting that the DELLA motif is a binding site that undergoes disorder-to- $\alpha$ -helix transition upon binding to GID1 receptors. The formation of this  $\alpha$ -helix has been confirmed and shown as  $\alpha$ -helix A and  $\alpha$ -helix B in the crystal structure of the AtGID1a-GA<sub>3</sub>-AtGAI<sub>n</sub> ternary complex (Fig. 6A) (Murase *et al.*, 2008). Although the VHYNP motif appears as a large downward spike in the PONDR score pattern (Fig. 6C), it was not identified in our  $\alpha$ -MoRFs prediction for any of the DELLAs. It is clearly shown in Fig. 6A that the VHYNP motif binds to AtGID1a, forming an irregular VHYNPSD loop. Therefore, it is  $\tau$ -MoRFs rather than  $\alpha$ -MoRFs involved in the binding-induced folding. The binding-induced folding of the DELLA and VHYNP motifs have also been supported by other biochemical evidence (Sun *et al.*, 2010). The LK/RXI motif was also identified in our  $\alpha$ -MoRFs prediction. It is possibly a potential binding site of DELLAs for an unknown component in DELLA signalling pathway as this motif was not involved in interactions with GID1 receptors.

## 4. Experimental methods for investigation of IDPs

Having specific amino acid sequences, IDPs possess a number of distinctive structural characteristics that can be utilized for their identification. This includes but not limited to, sensitivity to proteolysis, disorder characteristics of CD and NMR spectroscopy, small-angle X-ray scattering, hydrodynamic measurement, dynamic light scattering, deuterium-hydrogen exchange mass spectrometry, Raman and infrared spectroscopy, monoclonal antibody based immunoassays and so on (Uversky, 2002a). While every method emphasizes different structural features of IDPs, simultaneous application of several techniques mentioned above will provide unambiguous evidence for the presence of varying degrees of unfoldedness in a given protein. Here, we will focus on some of frequently used techniques.

### 4.1 Hydrodynamic dimensions of IDPs

Hydrodynamic dimension is the most definite characteristic of the conformational state of a protein. According to Uversky (Uversky, 1993, 2002b), a protein molecule can exist in

natively folded (NF), molten globule (MG), pre-molten globule (PMG), native coil (Coil) and completely unfolded states ( $U_{\text{urea}}$ ) and is characterized by the following relationships of its Stokes radius ( $R_s$ ) and theoretic molecular weight ( $MW^{\text{Theo}}$ ): natively folded,  $\log R_s^{\text{NF}} = 0.369 \times \log (MW^{\text{Theo}}) - 0.254$ ; molten globule,  $\log R_s^{\text{MG}} = 0.334 \times \log (MW^{\text{Theo}}) - 0.053$ ; pre-molten globule,  $\log R_s^{\text{PMG}} = 0.403 \times \log (MW^{\text{Theo}}) - 0.239$ ; native coil,  $\log R_s^{\text{Coil}} = 0.493 \times \log (MW^{\text{Theo}}) - 0.551$ ; complete unfolded random coil in urea,  $\log R_s^{\text{Urea}} = 0.521 \times \log (MW^{\text{Theo}}) - 0.649$ . Therefore, experimentally determined apparent molecular weight ( $MW^{\text{App}}$ ) and Stokes radius ( $R_s^{\text{D}}$ ) can be used to identify and classify IDPs into one of three subclasses: molten globule (MG)-like, pre-molten globule (PMG)-like or random-coil like, depending on their hydrodynamic characteristics.

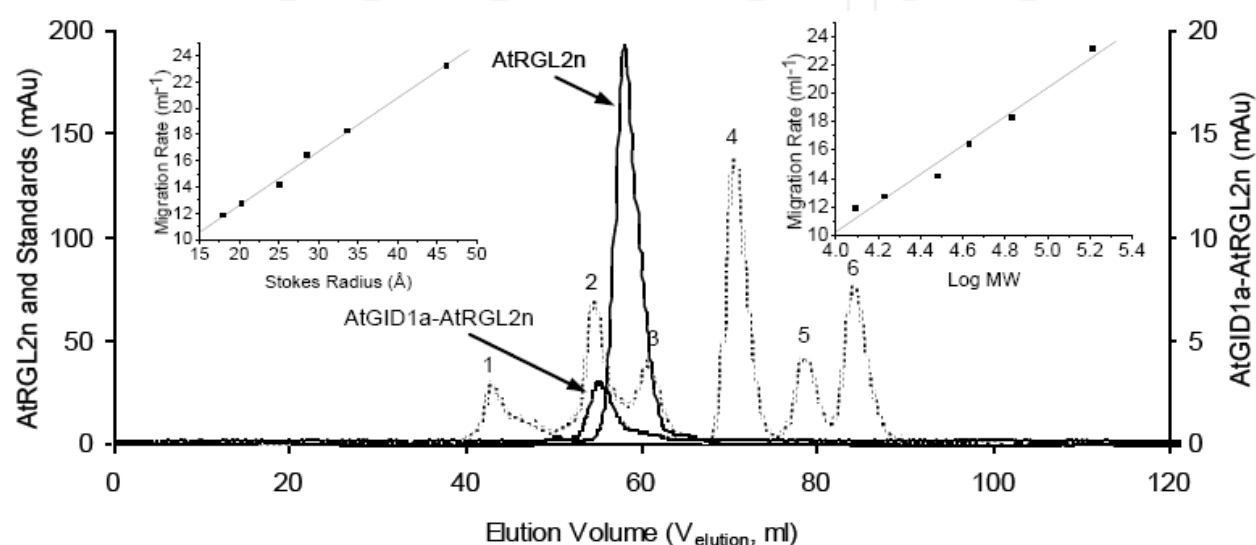


Fig. 7. Size-exclusion chromatography of AtRGL2n and AtGID1a-AtRGL2n complex. Six natively folded globular proteins (1. IgG, 158 kDa; 2. BSA, 67 kDa; 3. Ovalbumin, 43 kDa; 4. Carbonic Anhydrase, 30 kDa; 5. Myoglobin, 17 kDa; 6. Cytochrome c, 12.3 kDa corresponding to numbered peaks of hatched line) were used as standards to calibrate the Superdex 75-16/60 column. The resultant migration rate ( $1000/V_{\text{elution}}$ ) plotted against Stokes radius (left inset) and molecular weight (right inset) were used to determine Stokes radius ( $R_s^{\text{D}}$ ) and apparent molecular weight ( $MW^{\text{App}}$ ) of DELLAs and the complex. As examples, the peaks of AtRGL2n and AtGID1a-AtRGL2n complex are shown in solid and dotted lines, respectively. Adopted from Sun et al. (2010) *J. Biol. Chem.*, 285, 11557-11571.

As an example, we investigated hydrodynamic dimensions of the N-domains of DELLAs using size-exclusion chromatography (Fig. 7) in which six natively folded globular proteins were used as standards to calibrate the Superdex 75-16/60 column (see insets). AtRGL2n and its ternary complex AtGID1a-GA<sub>3</sub>-AtRGL2n were then applied on to the gel filtration column using the same buffer and flow rate. The determined Stokes radius ( $R_s^{\text{D}}$ ) and apparent molecular weight ( $MW^{\text{App}}$ ) of AtRGL2n and the ternary complex were calculated using migration rate ( $1000/V_{\text{elution}}$ ). The  $MW^{\text{App}}$  of free AtRGL2n is approximately two to three times larger than its theoretical molecular weight ( $MW^{\text{Theo}}$ ), indicating that it has an extended conformation with low compactness. The value of  $R_s^{\text{D}}$  reveals that free AtRGL2n belongs to the PMG-like IDPs. In contrast, the  $MW^{\text{App}}$  and  $R_s^{\text{D}}$  values show that the AtGID1a-GA<sub>3</sub>-AtRGL2n complex becomes natively folded, supporting the hypothesis that



intrinsic disordered AtRGL2n must undergo a binding-induced folding during the complexing of AtGID1a/GA<sub>3</sub> and AtRGL2n (Sun *et al.*, 2010).

4.2 CD spectra of IDPs

The far-UV circular dichroism (CD) spectra have been widely used for identification of IDPs due to its quick operation and easy accessibility. Contrary to  $\alpha$ -helix and  $\beta$ -sheet, random coil displays specific shape of the far-UV CD spectrum, with a large negative ellipticity in the vicinity of 200 nm, low ellipticity at 190 nm and an ellipticity close to zero in the vicinity of 222 nm (Johnson, 1988; Kelly and Price, 1997). This is a very useful graphical criterion for identification of IDPs. Double wavelength statistics showed the far-UV CD spectra characteristics of IDPs: some random coil-like IDPs having averaged ellipticities at 200 nm (-18900 deg cm<sup>2</sup> · dmol<sup>-1</sup>) and at 222 nm (-1700 deg cm<sup>2</sup> · dmol<sup>-1</sup>) while some pre-molten globule-like IDPs having averaged ellipticities at 200 nm (-10700 deg cm<sup>2</sup> · dmol<sup>-1</sup>) and at 222 nm (-3900 deg cm<sup>2</sup> · dmol<sup>-1</sup>) (Uversky, 2002a).

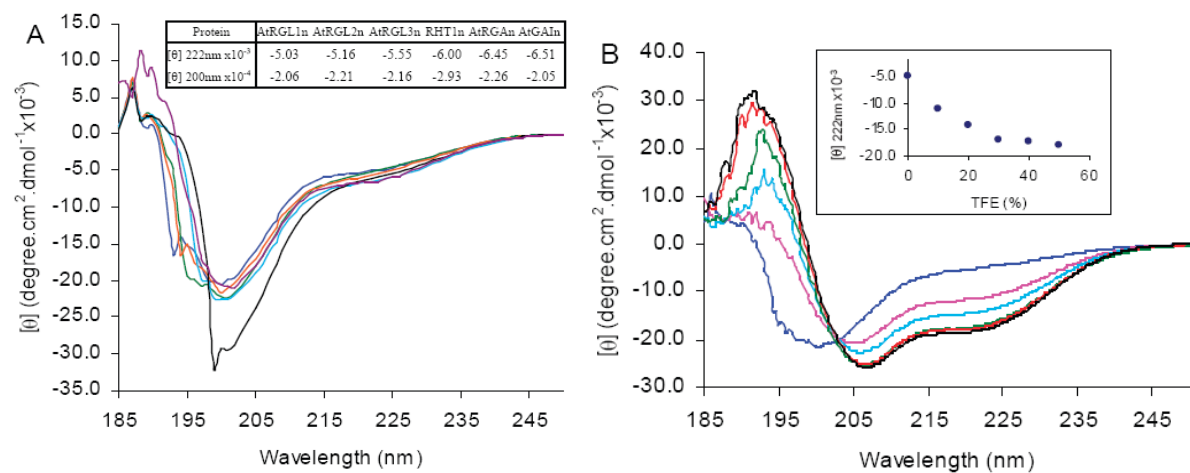


Fig. 8. Far-UV CD spectra of the N-domains of DELLAs (A) CD of AtRGL1n (blue), AtRGL2n (green), AtRGL3n (orange), AtRGAn (cyan), RHT1n (black) and AtGAI n (purple). The ellipticity [θ] at both 200 nm and 222 nm are shown in inset (B) 2,2,2-trifluoroethanol (TFE)-induced folding of AtRGL2n with TFE percentage: 0% (blue), 10% (pink), 20% (cyan), 30% (green), 40% (red) and 50% (black). The ellipticity [θ] at 222 nm versus TFE is shown in inset. Adopted from Sun *et al.* (2010) *J. Biol. Chem.*, 285, 11557-11571.

As an example of using CD spectra to characterize IDPs, the far-UV CD spectra of the N-domains of six DELLAs (Fig. 8A) display a large negative ellipticity at 200 nm and low ellipticity at 190 nm, characteristic of proteins in a largely disordered conformation. The ellipticity values at 222 nm are close to that of pre-molten globule-like IDPs, on the other hand, the ellipticity values at 200 nm are close to that of random coil-like IDPs (Fig. 8A, inset table). Alternatively, the solvent 2,2,2-Trifluoroethanol (TFE) mimics the hydrophobic environment experienced by protein - protein interactions and has therefore been widely used as a probe for the propensity of IDPs to undergo an induced folding upon target binding (Dyson and Wright, 2002). To test the potential binding-induced folding of the N-domains of DELLAs, far-UV CD spectra of AtRGL2n were recorded in the presence of increasing concentrations of TFE (Fig. 8B). AtRGL2n showed an increased  $\alpha$ -helicity upon

the addition of TFE, as indicated by the characteristic peak at 192 nm and double minima at 208 and 222 nm. Most of the disorder-to- $\alpha$ -helix transition takes place in presence of 30% TFE, at which point the  $\alpha$ -helical content reaches 51.5% as estimated from the ellipticity at 222 nm. The TFE results alone reveal a potential of AtRGL2n to form  $\alpha$ -helices upon binding to its interacting partners (Sun *et al.*, 2010).

### 4.3 NMR spectra of IDPs

As discussed above, CD spectroscopy is a very useful technique for identification of IDPs. However, this technique aims to detect overall tendency of conformational ensemble of intrinsic disorder and the resultant spectroscopic characteristics reflect local structural propensities averaged over the whole protein molecule. It does not provide conformational information about local residual structures such as those retained in pre-molten globules. In past decade, NMR spectroscopy has rapidly developed into a key technique for studying conformational ensemble and dynamics of proteins in unfolded and partially folded states (Dyson and Wright, 2005a). NMR spectroscopy investigates both local and long-range conformational behaviour at atomic resolution on timescales varying over many orders of magnitude, it has been used together with small-angle X-ray scattering (SAXS) to characterize the conformational ensemble of IDPs (Wells *et al.*, 2008). The residual dipolar couplings (RDCs) of NMR spectra has become a powerful tool to describe quantitatively the level of local structure and transient long-range order in IDPs, see review (Jensen *et al.*, 2009) for details.

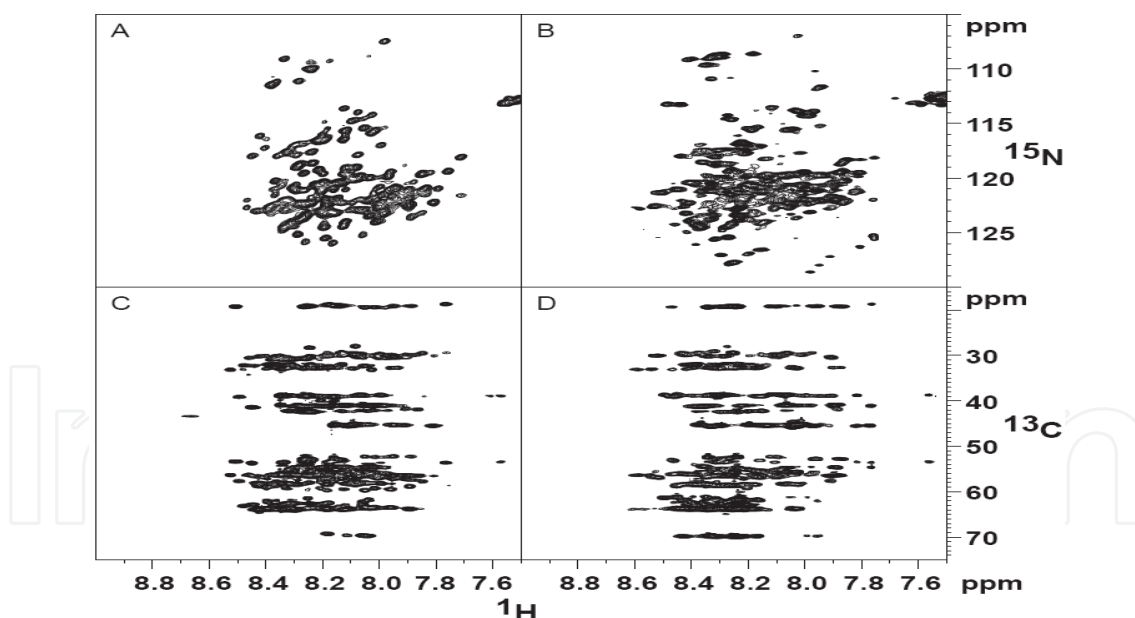


Fig. 9. NMR spectra showing chemical shift dispersion of AtRGL2n and AtRGAn. (A) and (B) are  $^1\text{H}$ - $^{15}\text{N}$  HSQC of AtRGL2n and AtRGAn, respectively. (C) and (D) show the  $^1\text{H}$ - $^{13}\text{C}$  plane of a CBCA(CO)NH experiment for AtRGL2n and AtRGAn, respectively. Reproduced from Sun *et al.* (2010) *J. Biol. Chem.*, 285, 11557-11571.

Here, we just show examples of characteristic NMR spectra of IDPs. The two-dimensional  $^{15}\text{N}$ ,  $^1\text{H}$ -HSQC spectra and  $^1\text{H}$ - $^{13}\text{C}$  planes from a CBCA(CO)NH experiment have been collected for two N-domains of DELLAs, AtRGL2n and AtRGAn (Fig. 9). The narrow ranges

of chemical shifts in the HSQC spectra for AtRGL2n and AtRGAn are characteristic of unstructured proteins (Fig. 9A and 9B, respectively). The CBCA(CO)NH planes (Fig. 9C and 9D) correlate the amide proton chemical shift with the C $\alpha$  and C $\beta$  shifts of the previous residue, in which horizontal rows of peaks at chemical shifts typical of random coil can be seen. Neither of these rows displays a significant spread in  $^{13}\text{C}$  shift values, suggesting a nearly uniform and therefore disordered environment. While the narrow range of  $^{13}\text{C}$  shifts seen here is consistent with disorder, the chemical dispersion of AtRGAn (Fig. 9B) does appear to be slightly higher than that of AtRGL2n (Fig. 9A), indicating that AtRGAn may have relatively more local residual structures (Sun *et al.*, 2010).

4.4 Deuterium / hydrogen exchange mass spectra

Deuterium/hydrogen exchange mass spectrometry (DHXMS) provides insights into protein structure and dynamics on a per amino acid basis. Compared to NMR spectra, the DHXMS technique requires much less time for sample preparation and can be routinely applied to larger proteins. The backbone amide hydrogens of proteins reversibly interchange with deuterium in D $_2$ O solvent and the exchange rate of each amide hydrogen in a protein directly and precisely reports solvent accessibility to it, revealing the protein’s conformational states on the scale of individual amino acids. Amide hydrogens that are involved in intramolecular H-bonds via secondary structures and/or permanently buried inside the protein are protected from the exchanging. Conversely, exchanging occurs readily at sites that are solvent-exposed and involved no hydrogen bond of secondary structures. Therefore, the exchange rates determined by mass spectrometry following the deuterium/hydrogen exchanging allow direct localization of structured or unfolded regions of the protein. DHXMS has been used in structural genomics to identify the disordered regions within large number of crystallographic targets (Pantazatos *et al.*, 2004), in exploring protein conformational dynamics and the structural aspect of solution-phase proteins (Konermann *et al.*, 2008).

As an example, Fig. 10 shows the hydrogen/deuterium exchange of AtRGL1n determined by MS. The MBP moiety of a well folded structure exhibited inaccessibility to solvent for most of the protein except the two terminal regions. In contrast, AtRGL1n showed instantly high exchange rate for the whole polypeptide chain, implying that free AtRGL1n without interacting partners is totally unfolded (Sheerin *et al.*, 2011).

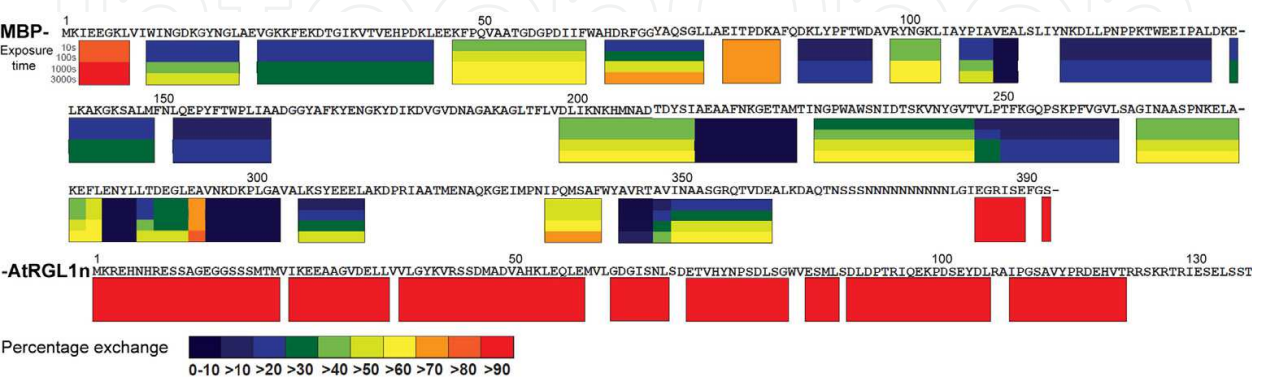


Fig. 10. The solvent accessibility of a maltose-binding protein (MBP) fused AtRGL1n. The exchange rates of backbone amide hydrogens were determined by MS following exposure to D $_2$ O for 10-3000 s intervals. Adopted from Sheerin et al. (2011) *Biochem. J.*, 435(3): 629-639.

#### 4.5 Monoclonal antibody immunoassays for intrinsic-disorder based protein interactions

Monoclonal antibodies (mAb) have long been recognized and used as powerful molecular probes to monitor protein folding and conformational changes (Goldberg, 1991). The conformational specific mAb that recognizes conformation of epitopes (small groups of sequential or distal amino acids) allows a direct insight into the conformation of antigenic proteins at individual amino acids level. This method was utilized in conformational characterization of paired helical filaments (PHF) of tau protein, an IDP in its monomeric form and polymerizes through binding-induced folding into insoluble PHF, causing Alzheimer's disease (AD) and related tauopathies. A special mAb MN423 raised against the PHF core of tau protein recognizes three spatially close amino acid segments which reside on a nearly 90 amino acid-long polypeptide chain including the C-terminus. The disclosure of the spatial proximity of these segments represents constraints for intra-molecular folding of the PHF core, leading to propose the model for folding of tau polypeptide chain in the PHF core (Skrabana *et al.*, 2006). The conformational specific mAb has often been used to monitor conformational changes of targeted proteins. Bax is a pro-apoptotic member of the B-cell lymphoma-2 (Bcl-2) family which are either IDP or contain IRDs that are critical to their function (Rautureau *et al.*, 2010). Bax undergoes a conformational change triggered by a TNF-related apoptosis-inducing ligand (TRAIL), leading to effective induction of mitochondrial apoptosis (Sundararajan *et al.*, 2001). This critical conformational change can be efficiently detected by using a mAb that recognizes only the three-dimensional epitope of conformationally changed Bax and was immobilized on a chip for surface plasmon resonance (SPR) detection (Kim *et al.*, 2005).

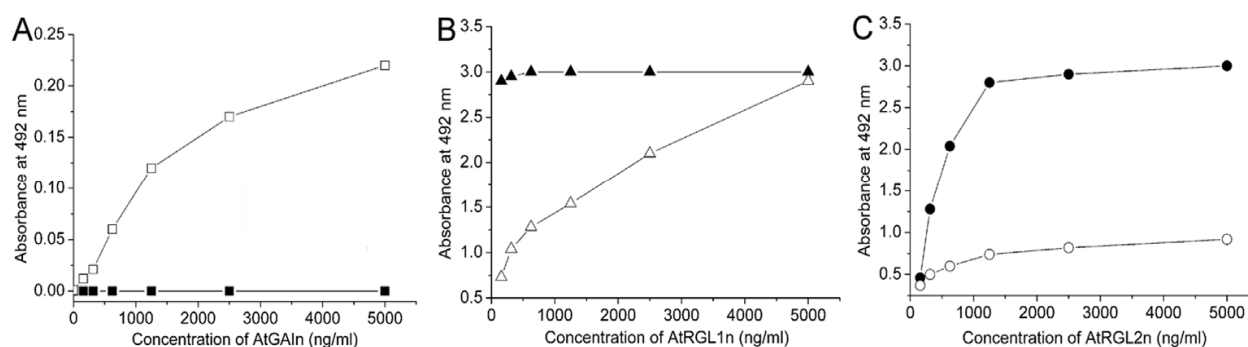


Fig. 11. Interaction of AtGID1 receptors extracted from *Arabidopsis quadruple-DELLA* mutant (*QUAD*) with the recombinant N-domains of DELLAs using mAb AD7 in sandwich immunoassays. All assays were carried out for AtGAI (□), AtRGL1 (△), AtRGL2 (○) mixed with *QUAD* extract and AtGAI (■), AtRGL1 (▲), AtRGL2 (●) mixed with extraction buffer only. Adopted from Sun *et al.* (2010) *J. Biol. Chem.*, 285, 11557-11571.

Furthermore, the conformational specific mAb has advantage over other techniques in monitoring conformational changes of antigenic proteins involved in a real time reaction system with multiple interacting components. This method has been successfully applied to gain an insight into the folding of DELLAs upon binding to the AtGID1, a GA receptor extracted from *Arabidopsis* tissue and binding to both the DELLA and VHYNP motifs of DELLAs (Sun *et al.*, 2010). Both biophysical data and biological functions suggested that the *Arabidopsis* DELLA family may be further divided into two subgroups: AtGAI, AtRGA



(RGA-group) and AtRGL1, AtRGL2 and AtRGL3 (RGL-group). This classification was reinforced by different conformations of unbound *Arabidopsis* DELLAs with regard to the VHYNP motifs. The mAb AD7, conformational specific to the VHYNP motif, does not recognize unbound AtGAIn and AtRGAn but AtRGL1n, AtRGL2n and AtRGL3n (Fig. 11), implying the different conformations of VHYNP motif between these two subgroups in their unbound form. Furthermore, the ELISA assays of the recombinant N-domains (AtGAIn, AtRGL1n and AtRGL2n), AtGID1/GA<sub>3</sub> and mAb AD7 showed that the AtGID1 only partially blocks mAb AD7 from binding to both AtRGL1n and AtRGL2n (Fig. 11B, 11C). In contrast, binding of AtGID1/GA<sub>3</sub> to AtGAIn renders mAb AD7 binding to AtGID1/GA<sub>3</sub>-AtGAIn complex (Fig. 11A). This indicates that the VHYNPSD loop in the unbound AtGAIn undergoes conformational changes induced by binding of AtGID1/GA<sub>3</sub>, resulting in at least the partial conformational epitope recognized by mAb AD7.

## 5. Conclusion

As one of the fastest growing areas of protein science in the past decade, IDPs or IDRs lack secondary and/or tertiary structures yet possess crucial cellular functions under physiological conditions. Protein interactions involving IDPs or IDRs can trigger binding-induced folding of IDPs or IDRs via a disorder-to-order transition. Such important characteristics enable IDPs or IDRs to play critical roles in molecular recognition. The intrinsic disorder based protein interactions will be a key factor in elucidating mechanisms of biological processes and regulations such as disease, various signal transductions, plant growth and development, transcriptional regulation and so on.

## 6. References

- Barabasi, A L and Oltvai, Z N (2004) Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5, 101-U115. doi: 10.1038/nrg1272.
- Bennett, W S and Steitz, T A (1978) Glucose-induced conformational change in yeast hexokinase. *Proc. Natl. Acad. Sci. U. S. A.*, 75, 4848-4852.
- Berman, H M, Westbrook, J, Feng, Z, Gilliland, G, Bhat, T N, Weissig, H, Shindyalov, I N and Bourne, P E (2000) The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242.
- Bychkova, V E, Berni, R, Rossi, G L, Kutysenko, V P and Ptitsyn, O B (1992) Retinol-binding protein is in the molten globule state at low pH. *Biochemistry*, 31, 7566-7571.
- Bychkova, V E, Dujsekina, A E, Fantuzzi, A, Ptitsyn, O B and Rossi, G L (1998) Release of retinol and denaturation of its plasma carrier, retinol-binding protein. *Folding & Design*, 3, 285-291.
- Bychkova, V E, Dujsekina, A E, Klenin, S I, Tiktopulo, E I, Uversky, V N and Ptitsyn, O B (1996) Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface. *Biochemistry*, 35, 6058-6063.
- Bychkova, V E, Pain, R H and Ptitsyn, O B (1988) The molten globule state is involved in the translocation of proteins across membranes. *Febs Letters*, 238, 231-234.
- Cheng, Y G, Oldfield, C J, Meng, J W, Romero, P, Uversky, V N and Dunker, A K (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry*, 46, 13468-13477. doi: 10.1021/bi7012273.

- Cortese, M S, Baird, J P, Uversky, V N and Dunker, A K (2005) Uncovering the unfoldome: Enriching cell extracts for unstructured proteins by acid treatment. *Journal of Proteome Research*, 4, 1610-1618. doi: 10.1021/pr050119c.
- Daughdrill, G W, Chadsey, M S, Karlinsey, J E, Hughes, K T and Dahlquist, F W (1997) The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma(28). *Nature Structural Biology*, 4, 285-291.
- Dunker, A K, Brown, C J, Lawson, J D, Iakoucheva, L M and Obradovic, Z (2002a) Intrinsic disorder and protein function. *Biochemistry*, 41, 6573-6582. doi: 10.1021/bi012159+.
- Dunker, A K, Brown, C J and Obradovic, Z (2002b) Identification and functions of usefully disordered proteins. *Unfolded Proteins*, 62, 25-49.
- Dunker, A K, Cortese, M S, Romero, P, Iakoucheva, L M and Uversky, V N (2005) Flexible nets - The roles of intrinsic disorder in protein interaction networks. *Febs Journal*, 272, 5129-5148. doi: 10.1111/j.1742-4658.2005.04948.x.
- Dunker, A K, Lawson, J D, Brown, C J, Williams, R M, Romero, P, Oh, J S, Oldfield, C J, Campen, A M, Ratliff, C R, Hipps, K W, Ausio, J, Nissen, M S, Reeves, R, Kang, C H, Kissinger, C R, Bailey, R W, Griswold, M D, Chiu, M, Garner, E C and Obradovic, Z (2001) Intrinsically disordered protein. *Journal of Molecular Graphics & Modelling*, 19, 26-59.
- Dunker, A K and Obradovic, Z (2001) The protein trinity - linking function and disorder. *Nature Biotechnology*, 19, 805-806.
- Dunker, A K, Obradovic, Z, Romero, P, Garner, E C and Brown, C J (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform*, 11, 161-171.
- Dunker, A K, Oldfield, C J, Meng, J W, Romero, P, Yang, J Y, Chen, J W, Vacic, V, Obradovic, Z and Uversky, V N (2007) The unfoldomics decade: an update on intrinsically disordered proteins. *Bmc Genomics*, 9. doi: 10.1186/1471-2164-9-s2-s1.
- Dyson, H J and Wright, P E (2002) Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology*, 12, 54-60.
- Dyson, H J and Wright, P E (2005a) Elucidation of the protein folding landscape by NMR. In *Nuclear Magnetic Resonance of Biological Macromolecules, Part C Vol. 394*. San Diego: Elsevier Academic Press Inc, pp. 299-+.
- Dyson, H J and Wright, P E (2005b) Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6, 197-208. doi: 10.1038/nrm1589.
- Fischer, E (1894) Einfluss der configuration auf die wirkung der enzyme. *Ber. Dt. Chem. Ges.*, 27, 2985-2993.
- Fuxreiter, M, Simon, I, Friedrich, P and Tompa, P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of Molecular Biology*, 338, 1015-1026. doi: 10.1016/j.jmb.2004.03.017.
- Goldberg, M E (1991) Investigating protein conformation, dynamics and folding with monoclonal-antibodies. *Trends in Biochemical Sciences*, 16, 358-362.
- Han, J D J, Bertin, N, Hao, T, Goldberg, D S, Berriz, G F, Zhang, L V, Dupuy, D, Walhout, A J M, Cusick, M E, Roth, F P and Vidal, M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430, 88-93. doi: 10.1038/nature02555.
- Haynes, C, Oldfield, C J, Ji, F, Klitgord, N, Cusick, M E, Radivojac, P, Uversky, V N, Vidal, M and Iakoucheva, L M (2006) Intrinsic disorder is a common feature of hub

- proteins from four eukaryotic interactomes. *Plos Computational Biology*, 2, 890-901. doi: 10.1371/journal.pcbi.0020100.
- He, B, Wang, K J, Liu, Y L, Xue, B, Uversky, V N and Dunker, A K (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Research*, 19, 929-949. doi: 10.1038/cr.2009.87.
- Iakoucheva, L M, Brown, C J, Lawson, J D, Obradovic, Z and Dunker, A K (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*, 323, 573-584. doi: 10.1016/s0022-2836(02)00969-5.
- Jensen, M R, Markwick, P R L, Meier, S, Griesinger, C, Zweckstetter, M, Grzesiek, S, Bernado, P and Blackledge, M (2009) Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure*, 17, 1169-1185. doi: 10.1016/j.str.2009.08.001.
- Johnson, W C (1988) Secondary structure of proteins through circular-dichroism spectroscopy. *Annual Review of Biophysics and Biophysical Chemistry*, 17, 145-166.
- Karush, F (1950) Heterogeneity of the binding sites of bovine serum albumin. *J. Am. Chem. Soc.*, 72, 2705-2713.
- Kelly, S M and Price, N C (1997) The application of circular dichroism to studies of protein folding and unfolding. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, 1338, 161-185.
- Kim, M, Jung, S O, Park, K, Jeong, E J, Joung, H A, Kim, T H, Seol, D W and Chung, B H (2005) Detection of Bax protein conformational change using a surface plasmon resonance imaging-based antibody chip. *Biochemical and Biophysical Research Communications*, 338, 1834-1838. doi: 10.1016/j.bbrc.2005.10.155.
- Konermann, L, Tong, X and Pan, Y (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *Journal of Mass Spectrometry*, 43, 1021-1036. doi: 10.1002/jms.1435.
- Koshland, D E (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.*, 44, 98-104.
- Koshland, D E (1994) The key-lock theory and the induced fit theory. *Angew. Chem.-Int. Edit. Engl.*, 33, 2375-2378.
- Liu, J G, Perumal, N B, Oldfield, C J, Su, E W, Uversky, V N and Dunker, A K (2006) Intrinsic disorder in transcription factors. *Biochemistry*, 45, 6873-6888. doi: 10.1021/bi0602718.
- McDonald, R C, Steitz, T A and Engelman, D M (1979) Yeast hexokinase in solution exhibits a large conformational change upon binding glucose or glucose-6-phosphate. *Biochemistry*, 18, 338-342.
- Midic, U, Oldfield, C J, Dunker, A K, Obradovic, Z and Uversky, V N (2009) Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *Bmc Genomics*, 10. doi: 10.1186/1471-2164-10-s1-s12.
- Mirsky, A E and Pauling, L (1936) On the structure of native, denatured and coagulated proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 13, 739-766.
- Mohan, A, Oldfield, C J, Radivojac, P, Vacic, V, Cortese, M S, Dunker, A K and Uversky, V N (2006) Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology*, 362, 1043-1059. doi: 10.1016/j.jmb.2006.07.087.
- Murase, K, Hirano, Y, Sun, T P and Hakoshima, T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature*, 456, 459-464.

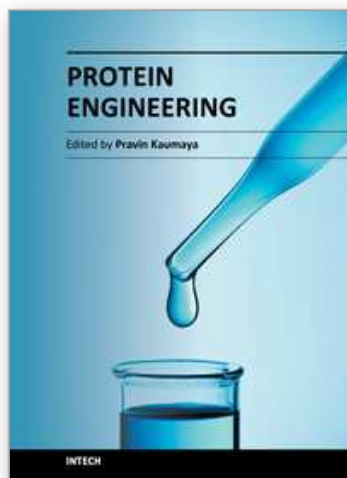
- Ohgushi, M and Wada, A (1983) Molten-globule state - a compact form of globular-proteins with mobile side-chains. *Febs Letters*, 164, 21-24.
- Oldfield, C J, Cheng, Y, Cortese, M S, Brown, C J, Uversky, V N and Dunker, A K (2005a) Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, 44, 1989-2000. doi: 10.1021/bi047993o.
- Oldfield, C J, Cheng, Y G, Cortese, M S, Romero, P, Uversky, V N and Dunker, A K (2005b) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, 44, 12454-12470. doi: 10.1021/bi050736e.
- Oldfield, C J, Meng, J, Yang, J Y, Yang, M Q, Uversky, V N and Dunker, A K (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *Bmc Genomics*, 9 Suppl 1, S1.
- Pantazatos, D, Kim, J S, Klock, H E, Stevens, R C, Wilson, I A, Lesley, S A and Woods, V L (2004) Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc. Natl. Acad. Sci. U. S. A.*, 101, 751-756.
- Patil, A and Nakamura, H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *Febs Letters*, 580, 2041-2045. doi: 10.1016/j.febslet.2006.03.003.
- Peng, K, Radivojac, P, Vucetic, S, Dunker, A K and Obradovic, Z (2006) Length-dependent prediction of protein intrinsic disorder. *Bmc Bioinformatics*, 7. doi: 10.1186/1471-2105/7/208.
- Phillips, D C (1986) Development of concepts of protein-structure. *Perspectives in Biology and Medicine*, 29, S124-S130.
- Promponas, V J, Enright, A J, Tsoka, S, Kreil, D P, Leroy, C, Hamodrakas, S, Sander, C and Ouzounis, C A (2000) CAST: an iterative algorithm for the complexity analysis of sequence tracts. *Bioinformatics*, 16, 915-922.
- Ptitsyn, O B and Uversky, V N (1994) The molten globule is a 3rd thermodynamical state of protein molecules. *Febs Letters*, 341, 15-18.
- Radivojac, P, Iakoucheva, L M, Oldfield, C J, Obradovic, Z, Uversky, V N and Dunker, A K (2007) Intrinsic disorder and functional proteomics. *Biophysical Journal*, 92, 1439-1456. doi: 10.1529/biophysj.106.094045.
- Rautureau, G J P, Day, C L and Hinds, M G (2010) Intrinsically Disordered Proteins in Bcl-2 Regulated Apoptosis. *International Journal of Molecular Sciences*, 11, 1808-1824.
- Riek, R, Hornemann, S, Wider, G, Billeter, M, Glockshuber, R and Wuthrich, K (1996) NMR structure of the mouse prion protein domain PrP(121-231). *Nature*, 382, 180-182.
- Romero, P, Obradovic, Z, Li, X H, Garner, E C, Brown, C J and Dunker, A K (2001) Sequence complexity of disordered protein. *Proteins-Structure Function and Genetics*, 42, 38-48.
- Rual, J F, Venkatesan, K, Hao, T, Hirozane-Kishikawa, T, Dricot, A, Li, N, Berriz, G F, Gibbons, F D, Dreze, M, Ayivi-Guedehoussou, N, Klitgord, N, Simon, C, Boxem, M, Milstein, S, Rosenberg, J, Goldberg, D S, Zhang, L V, Wong, S L, Franklin, G, Li, S M, Albala, J S, Lim, J H, Fraughton, C, Llamasas, E, Cevik, S, Bex, C, Lamesch, P, Sikorski, R S, Vandenhaute, J, Zoghbi, H Y, Smolyar, A, Bosak, S, Sequerra, R, Doucette-Stamm, L, Cusick, M E, Hill, D E, Roth, F P and Vidal, M (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437, 1173-1178.



- Schulz, G E (1979) Nucleotide Binding Proteins. In *Molecular Mechanism of Biological Recognition* (Balaban, M., ed. New York: Elsevier/North-Holland, pp. 79-94.
- Sheerin, D J, Buchanan, J, Kirk, C, Harvey, D, Sun, X, Spagnuolo, J, Li, S, Liu, T, Woods, V A, Foster, T, Jones, W T and Rakonjac, J (2011) Inter- and intra-molecular interactions of *Arabidopsis thaliana* DELLA protein RGL1. *Biochemical Journal*, 435, 629-639. doi: 10.1042/bj20101941.
- Sheng, Y, Sun, X L, Shen, Y, Bogнар, A L, Baker, E N and Smith, C A (2000) Structural and functional similarities in the ADP-forming amide bond ligase superfamily: Implications for a substrate-induced conformational change in folylpolyglutamate synthetase. *Journal of Molecular Biology*, 302, 427-440. doi: 10.1006/jmbi.2000.3987.
- Sickmeier, M, Hamilton, J A, LeGall, T, Vacic, V, Cortese, M S, Tantos, A, Szabo, B, Tompa, P, Chen, J, Uversky, V N, Obradovic, Z and Dunker, A K (2007) DisProt: the database of disordered proteins. *Nucleic Acids Research*, 35, D786-D793. doi: 10.1093/nar/gkl893.
- Skrabana, R, Sevcik, J and Novak, M (2006) Intrinsically disordered proteins in the neurodegenerative processes: Formation of tau protein paired helical filaments and their analysis. *Cell. Mol. Neurobiol.*, 26, 1085-1097. doi: 10.1007/s10571-006-9083-3.
- Stelzl, U, Worm, U, Lalowski, M, Haenig, C, Brembeck, F H, Goehler, H, Stroedicke, M, Zenkner, M, Schoenherr, A, Koeppen, S, Timm, J, Mintzlaff, S, Abraham, C, Bock, N, Kietzmann, S, Goedde, A, Toksoz, E, Droege, A, Krobitsch, S, Korn, B, Birchmeier, W, Lehrach, H and Wanker, E E (2005) A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, 122, 957-968. doi: 10.1016/j.cell.2005.08.029.
- Sun, X, Xue, B, Jones, W T, Rikkerink, E, Dunker, A K and Uversky, V N (2011) A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol. Biol.*, 77, 205-223. doi: 10.1007/s11103-011-9803-z.
- Sun, X L, Bogнар, A L, Baker, E N and Smith, C A (1998) Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase. *Proc. Natl. Acad. Sci. U. S. A.*, 95, 6647-6652.
- Sun, X L, Cross, J A, Bogнар, A L, Baker, E N and Smith, C A (2001) Folate-binding triggers the activation of folylpolyglutamate synthetase. *Journal of Molecular Biology*, 310, 1067-1078. doi: 10.1006/jmbi.2001.4815.
- Sun, X L, Frearson, N, Kirk, C, Jones, W T, Harvey, D, Rakonjac, J, Foster, T and Al-Samarrai, T (2008) An *E. coli* expression system optimized for DELLA proteins. *Protein Expression and Purification*, 58, 168-174. doi: 10.1016/j.pep.2007.09.003.
- Sun, X L, Jones, W T, Harvey, D, Edwards, P J B, Pascal, S M, Kirk, C, Considine, T, Sheerin, D J, Rakonjac, J, Oldfield, C J, Xue, B, Dunker, A K and Uversky, V N (2010) N-terminal Domains of DELLA Proteins Are Intrinsically Unstructured in the Absence of Interaction with GID1/Gibberellic Acid Receptors. *J. Biol. Chem.*, 285, 11557-11571. doi: 10.1074/jbc.M109.027011.
- Sundararajan, R, Cuconati, A, Nelson, D and White, E (2001) Tumor necrosis factor-alpha induces Bax-Bak interaction and apoptosis, which is inhibited by adenovirus E1B 19K. *J. Biol. Chem.*, 276, 45120-45127.

- Tompa, P (2002) Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27, 527-533.
- Tompa, P (2003) Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, 25, 847-855. doi: 10.1002/bies.10324.
- Tompa, P (2005) The interplay between structure and function in intrinsically unstructured proteins. *Febs Letters*, 579, 3346-3354. doi: 10.1016/j.febslet.2005.03.072.
- Tompa, P and Csermely, P (2004) The role of structural disorder in the function of RNA and protein chaperones. *Faseb Journal*, 18, 1169-1175. doi: 10.1096/fj.04-1584rev.
- Uversky, V N (1993) Use of fast protein size-exclusion liquid-chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, 32, 13288-13298.
- Uversky, V N (1997) Diversity of equilibrium compact forms of denatured globular proteins. *Protein and Peptide Letters*, 4, 355-367.
- Uversky, V N (2002a) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.*, 11, 739-756. doi: 10.1110/ps.4210102.
- Uversky, V N (2002b) What does it mean to be natively unfolded? *European Journal of Biochemistry*, 269, 2-12.
- Uversky, V N (2010) The Mysterious Unfoldome: Structureless, Underappreciated, Yet Vital Part of Any Given Proteome. *Journal of Biomedicine and Biotechnology*, 2010, 568068. doi: 10.1155/2010/568068.
- Uversky, V N, Gillespie, J R and Fink, A L (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics*, 41, 415-427.
- Uversky, V N, Kutysenko, V P, Protasova, N Y, Rogov, V V, Vassilenko, K S and Gudkov, A T (1996) Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. *Protein Sci.*, 5, 1844-1851.
- Uversky, V N, Narizhneva, N V, Ivanova, T V, Kirkitadze, M D and Tomashevski, A Y (1997) Ligand-free form of human alpha-fetoprotein: Evidence for the molten globule state. *Febs Letters*, 410, 280-284.
- Uversky, V N, Oldfield, C J and Dunker, A K (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition*, 18, 343-384. doi: 10.1002/jmr.747.
- Uversky, V N, Oldfield, C J, Midic, U, Xie, H B, Xue, B, Vucetic, S, Iakoucheva, L M, Obradovic, Z and Dunker, A K (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *Bmc Genomics*, 10. doi: 10.1186/1471-2164-10-s1-s7.
- Uversky, V N and Ptitsyn, O B (1994) Partly folded state, a new equilibrium state of protein molecules - 4-state guanidinium chloride-induced unfolding of beta-lactamase at low-temperature. *Biochemistry*, 33, 2782-2791.
- Uversky, V N and Ptitsyn, O B (1996) Further evidence on the equilibrium "pre-molten globule state": Four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *Journal of Molecular Biology*, 255, 215-228.
- Vacic, V, Oldfield, C J, Mohan, A, Radivojac, P, Cortese, M S, Uversky, V N and Dunker, A K (2007a) Characterization of molecular recognition features, MoRFs, and their binding partners. *Journal of Proteome Research*, 6, 2351-2366. doi: 10.1021/pr0701411.

- Vacic, V, Uversky, V N, Dunker, A K and Lonardi, S (2007b) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *Bmc Bioinformatics*, 8. doi: 10.1186/1471-2105-8-211.
- Vucetic, S, Obradovic, Z, Vacic, V, Radivojac, P, Peng, K, Iakoucheva, L M, Cortese, M S, Lawson, J D, Brown, C J, Sikes, J G, Newton, C D and Dunker, A K (2005) DisProt: a database of protein disorder. *Bioinformatics*, 21, 137-140. doi: 10.1093/bioinformatics/bth476.
- Vucetic, S, Xie, H B, Iakoucheva, L M, Oldfield, C J, Dunker, A K, Obradovic, Z and Uversky, V N (2007) Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *Journal of Proteome Research*, 6, 1899-1916. doi: 10.1021/pr060393m.
- Ward, J J, Sodhi, J S, McGuffin, L J, Buxton, B F and Jones, D T (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, 337, 635-645. doi: 10.1016/j.jmb.2004.02.002.
- Wells, M, Tidow, H, Rutherford, T J, Markwick, P, Jensen, M R, Mylonas, E, Svergun, D I, Blackledge, M and Fersht, A R (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.*, 105, 5762-5767. doi: 10.1073/pnas.0801353105.
- Wright, P E and Dyson, H J (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293, 321-331.
- Wu, H (1931) Studies on denaturation of proteins XIII. A theory of denaturation. *Chinese J. Physiol.*, 1, 219-234.
- Xie, H B, Vucetic, S, Iakoucheva, L M, Oldfield, C J, Dunker, A K, Obradovic, Z and Uversky, V N (2007a) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *Journal of Proteome Research*, 6, 1917-1932. doi: 10.1021/pr060394e.
- Xie, H B, Vucetic, S, Iakoucheva, L M, Oldfield, C J, Dunker, A K, Uversky, V N and Obradovic, Z (2007b) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *Journal of Proteome Research*, 6, 1882-1898. doi: 10.1021/pr060392u.
- Xue, B, Dunbrack, R L, Williams, R W, Dunker, A K and Uversky, V N (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta*, 1804. doi: :10.1016/j.bbapap.2010.01.011.
- Xue, B, Oldfield, C J, Dunker, A K and Uversky, V N (2009) CDF it all: Consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *Febs Letters*, 583, 1469-1474. doi: 10.1016/j.febslet.2009.03.070.
- Zhao, R B, Gish, K, Murphy, M, Yin, Y X, Notterman, D, Hoffman, W H, Tom, E, Mack, D H and Levine, A J (2000) Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes & Development*, 14, 981-993.



## **Protein Engineering**

Edited by Prof. Pravin Kaumaya

ISBN 978-953-51-0037-9

Hard cover, 344 pages

**Publisher** InTech

**Published online** 24, February, 2012

**Published in print edition** February, 2012

A broad range of topics are covered by providing a solid foundation in protein engineering and supplies readers with knowledge essential to the design and production of proteins. This volume presents in-depth discussions of various methods for protein engineering featuring contributions from leading experts from different countries. A broad series of articles covering significant aspects of methods and applications in the design of novel proteins with different functions are presented. These include the use of non-natural amino acids, bioinformatics, molecular evolution, protein folding and structure-functional insight to develop useful proteins with enhanced properties.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Xiaolin Sun, William T. Jones and Vladimir N. Uversky (2012). Applications of Bioinformatics and Experimental Methods to Intrinsic Disorder-Based Protein-Protein Interactions, Protein Engineering, Prof. Pravin Kaumaya (Ed.), ISBN: 978-953-51-0037-9, InTech, Available from: [http://www.intechopen.com/books/protein-engineering/applications-of-bioinformatics-and-experimental-methods-to-intrinsic-disorder-based-protein-](http://www.intechopen.com/books/protein-engineering/applications-of-bioinformatics-and-experimental-methods-to-intrinsic-disorder-based-protein-protein-)

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen