

Intrinsically disordered proteins: controlled chaos or random walk

T.C. Howton,¹ Yingqian Ada Zhan,²
Yali Sun,¹ M. Shahid Mukhtar^{1,3}

¹Department of Biology, University of Alabama at Birmingham, AL; ²Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID; ³Nutrition Obesity Research Center, University of Alabama at Birmingham, AL, USA

Abstract

Traditional conventions that a protein's sequence dictates its definitive, tertiary structure, and that this fixed structure provides the protein with the ability to carry out its designated role(s) are still correct but not for all proteins. Research over the past decade discovered that several key proteins possess intrinsically disordered regions (IDRs) that are crucial to their ability to perform specific functions and are observed clustered together within important classes of proteins. In this review, we aim to demonstrate how free energy landscapes, molecular dynamics simulations, and homology modeling are helpful in understanding key conformational dynamics of intrinsically disordered proteins (IDPs). Additionally, we use a list of predicted IDPs found in *Arabidopsis* to identify chromatin organizers and transcriptional regulators as being highly enriched in IDPs. Furthermore, we focus our attention to specific proteins within these families such as HAC5, EFS, ANAC019, ANAC013, and ANAC046. Future studies are needed to experimentally identify additional IDPs and their binding mechanisms.

Introduction

Over the past two decades, the world of proteomics has undergone a significant paradigm shift. The classical approach to the study of proteins depended on the adherence to the protein structure-function model, where each protein was composed of an amino acid sequence that lead to a static structure and function. With the discovery of intrinsically disordered regions (IDRs), researchers have developed new approaches and methodologies to better understand the unstructured and dynamic world of intrinsically disordered proteins (IDPs).¹ Although all domains of life con-

tain IDPs, eukaryotic proteins tend to show a significant level of enrichment. It has been demonstrated that approximately 33% of eukaryotic proteins contain at least one long stretch of residues (30 or more) that code for an intrinsically disordered region,² and greater than 30% of eukaryotic proteins have 50 or more consecutive disordered residues.³ Proteome-wide analyses of multiple plants have shown that roughly 30% of a plant's proteome is comprised of proteins containing at least one region with 50 or more disordered residues. It has been suggested that the pronounced occurrence of IDPs in plant genomes might be advantageous for them to mount effective cellular responses under varied biotic and abiotic environmental conditions. This increased frequency of IDPs in plants could also lead to a high level of phenotypic plasticity.⁴ Intrinsically disordered proteins are found in nearly every class of proteins within the eukaryotic proteome, including transcription factors, signaling proteins, and proteins involved in chromatin remodeling.

The intrinsically disordered regions of unstructured proteins serve multiple purposes under different environmental conditions. Given that proteins generally form macromolecular complexes to execute diverse cellular functions, proteins with unstructured regions can have a wide-range of binding partners and might participate in multiple biological processes.⁵ In addition, it was also demonstrated that the speed of a protein's ability to bind to its partner(s) increased drastically in proteins with intrinsically disordered regions when compared to their structured, globular counterparts.⁶ Intrinsically disordered regions known as *linker* regions can hold two globular portions of a protein in close proximity with each other, while still allowing a large amount of flexibility in their spatial relationship with one another.⁷ This perhaps allows a protein with IDRs to bind with several binding partners simultaneously under changing cellular states.

Researchers have created several tools in order to better understand IDPs and their functions. Computational analyses of free energy landscapes (mapping all possible conformations of an entity) are vital to experimentally defining an IDP. Typically, proteins fold into their stable state at a distinct trough of the free energy landscape. However, IDPs show multiple shallow troughs due the increased number of conformational states. Free energy landscapes, along with molecular dynamics studies and homology modeling, allow predicting the conformational dynamics of IDPs. In this review, we aim to highlight computational tools that are helpful in understanding key conformational dynamics of IDPs. Distinctively, we will focus on intrinsically disordered proteins within the plant *Arabidopsis*

Correspondence: Shahid M. Mukhtar, Department of Biology, The University of Alabama at Birmingham, Campbell Hall 369, 1300 University Blvd., Birmingham, AL 35294-1170, USA.
Tel.: +1.205.934.8335 - Fax: +1.205.975.6097.
E-mail: smukhtar@uab.edu

Key words: Intrinsically disordered regions; *Arabidopsis*; chromatin organizers; transcriptional regulators.

Acknowledgments: the authors would like to acknowledge Mr. Derek Moates for his valuable comments and suggestions on this manuscript. We acknowledge funding from UAB, College of Arts and Science graduate fellowship to YS and NSF (IOS-1557796) to MSM. The authors apologize to all authors whose work was not included within this review due to space restrictions. M. S. Mukhtar was supported through funds from the Department of Biology, The University of Alabama at Birmingham.

Contributions: all authors participated in developing the ideas presented in this manuscript, researching the literature, and writing parts of the text.

Conflict of interest: the authors declare no conflict of interest.

Received for publication: 10 September 2015.

Revision received: 12 September 2015.

Accepted for publication: 12 September 2015.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 International License (CC BY-NC 3.0).

©Copyright T.C. Howton et al., 2015
Licensee PAGEPress srl, Italy
International Journal of Plant Biology 2015; 6:6191
doi:10.4081/pb.2015.6191

thaliana (*Arabidopsis*). Our Gene Ontology (GO) enrichment analysis determines that proteins families belonging to chromatin remodeling and transcriptional regulation are statistically enriched in IDPs. We will discuss the potential roles of these classes of proteins in diverse biological processes with specific examples.

Free energy landscape of intrinsically disordered proteins

Due to thermal fluctuation, proteins can exhibit various conformations. The occurrence of each conformation is presented by the topography of the free energy landscape. Natively structured proteins usually fold into a native structure at the bottom of a free energy

landscape (Figure 1A). On the other hand, IDPs have no well-defined structure in the unbound state and have multiple shallow dips for weakly bound states, which makes a rugged free energy landscape (Figure 1B). Consequently, IDPs have a broad range of conformational dynamics and constantly move between different conformations.⁸⁻¹⁰ This property makes it challenging to map an IDP's structures and dynamics to its function. In the past decade, IDPs have been found prevalent among living organisms where they play essential roles in many biological processes.^{1,11-13} Although difficult, it is very important to characterize the free energy landscape of IDPs which will help us better understand how a primary sequence encodes the diverse mechanisms and links to the functions in these proteins. Furthermore, a better understanding of IDPs conformations may lead to the development of new therapeutics based on structural drug design.¹⁴⁻¹⁶

Molecular dynamic simulations and homology modeling

Intrinsically disordered proteins exhibit high levels of flexibility and span multiple different conformations. Molecular dynamics (MD) simulations provide a tool to computationally explore the conformational space and examine the dynamics of a protein over time. Although computer simulations rely on a series of cumulative approximations that can be erroneous, it has become a necessary tool in the research involving IDPs.

Combined with experimental studies, typically NMR spectroscopy, the structural ensemble of the free energy landscape of IDPs can be constructed using computer simulations. Ensemble restrained MD simulations constitute a useful and important tool for modeling IDPs.¹⁷⁻¹⁹ Constructing an ensemble based on a pre-determined structural library represents another way of completing the task. ENSEMBLE,²⁰ Select (SAS),²¹ ASTEROIDS,²² and BEGR,²³ are pieces of software falling into this category.

IDPs are considered to have high specificity and low affinity when interacting with binding, partner molecules. Mutations may harm this interaction. MD simulations can be used to study an array of mutations to predict the consequences and provide mechanistic explanations without performing experiments. Using computational alanine scanning, Massova *et al.*²⁴ suggested an approach to probe protein-protein interactions and evaluated binding free energies. In their case, they applied the method to p53 and MDM2 binding system where p53 is intrinsically disordered, and their results show excellent agreement with experi-

mental data.²⁴ Homology modeling is a useful tool to predict and study the structure of a protein from a homologue protein, where the structure of the protein in question has not been solved experimentally and the structure of its homologue protein is known. In the Arabidopsis genome, the *COR15A* gene is paired with *COR15B*.²⁵ Proteins encoded by these two genes are homologous with 70% identity in their amino acid sequences. Overexpression of the *COR15A* gene in Arabidopsis produces excess mature COR15A protein in the chloroplast stroma,²⁶ leading to enhanced freezing tolerance of chloroplasts of intact leaves and of isolated protoplasts frozen and thawed *in vitro*.^{26,27} In contrast, functional or structural information has been reported for COR15B protein. Both proteins show homology

to the Pfam LEA_4 family of LEA proteins and both were predicted to be IDPs.²⁸ Thalhammer *et al.*²⁹ showed the structural modeling of these two proteins known LEA homologues and demonstrated the interactions between MGDC and the COR15 proteins that may help keep the cell integrity through freezing stress.

Arabidopsis protein families enriched in intrinsically disordered proteins

A recent genome-wide analysis was performed to predict IDRs in Arabidopsis and compare them with the human proteome.

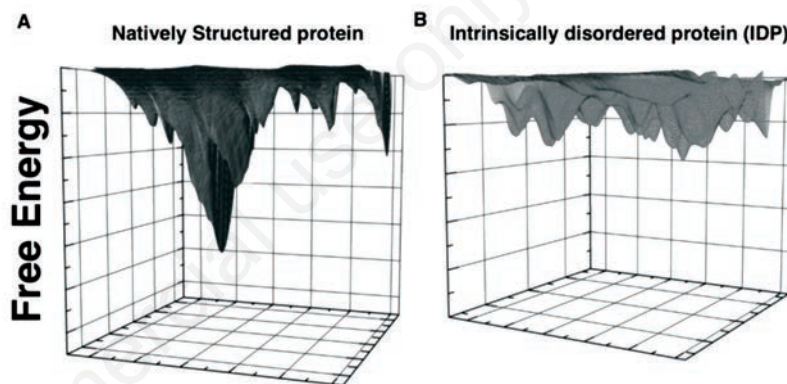


Figure 1. The free energy landscape of a natively structured protein. A natively structured protein (A) exhibits a well-defined minimum energy state corresponding to the folded conformation. In comparison, the energy landscape of an IDP (B) lacks a deep energy minimum.

Table 1. Identification of protein families enriched in intrinsically disordered regions.

	Count	P_Value	Benjamini
Annotation cluster 1 (enrichment score: 15.42)			
Chromosome organization	28	2.60E-18	1.50E-15
Chromatin modification	22	7.40E-18	2.10E-15
Chromatin organization	25	8.00E-17	2.10E-14
Chromatin regulator	15	1.40E-11	1.30E-09
Annotation cluster 2 (enrichment score: 7.94)			
DNA binding	85	1.90E-20	3.30E-18
Regulation of transcription	57	2.20E-10	3.10E-08
Nucleus	58	4.10E-09	2.00E-07
Transcription regulator activity	54	1.20E-08	6.90E-07
Transcription	40	4.20E-08	3.40E-06
Regulation of transcription, DNA-dependent	35	1.20E-07	6.90E-06
Regulation of RNA metabolic process	35	1.40E-07	7.20E-06
Transcription regulation	37	5.10E-07	1.60E-05
Transcription	37	8.30E-07	2.00E-05
Activator	17	2.70E-06	5.20E-05
Transcription factor activity	44	3.50E-06	1.00E-04
DNA binding	38	8.80E-06	1.20E-04

Intriguingly, it was discovered that specific functional classes are enriched with IDRs in *Arabidopsis*.³⁰ These functional groups include *post-translational protein modification* and *response to red or far red light*. While these broad functional classes provided insightful clues on the essential roles of these IDPs/IDRs in *Arabidopsis*, here we reevaluated these data to specifically predict protein families enriched with IDRs. To achieve this, we compiled a list of IDPs containing at least five long disordered regions (>30 residues). We subjected these IDPs to functional annotation tool, DAVID (the Database for Annotation, Visualization and Integrated Discovery)³¹ to predict the statistically enriched GO categories as well as enriched groups of different protein families (Supplementary Table S1: Functional Annotation of Predicted IDPs). DAVID employs a novel agglomeration algorithm to assemble a list of genes or associated biological terms into organized classes of related genes. Subsequently, we used the PANTHER (Protein ANalysis THrough Evolutionary Relationships)^{32,33} classification tool to organize the lists by molecular function (Figure 2). We identified the two most enriched groups contain chromatin organizers (Enrichment Score: 15.42) and transcription regulators (Enrichment Score: 7.94) (Table 1). We hypothesized that these two molecular groups play widespread roles in the downstream regulation of a vast majority of *Arabidopsis* genes under diverse cellular conditions. In the following sections of the review, we will discuss specific examples from these two groups and highlight key findings pertaining to their roles in conjunction with IDRs.

Chromatin remodeling

Even though cells within most eukaryotic organisms specialize in both structure and function, each contains the same genomic DNA of the organism. Therefore, the uniqueness of each cell is derived not by the DNA sequence but instead by the availability of portions of the DNA within that particular cell.³⁴ The accessibility of genes is regulated not only through the prevalence of transcription factors and enhancers, but also through the structure of the DNA itself. The DNA of eukaryotes is associated with histone octamers that sequester approximately 147 base pairs of DNA in a nucleosome complex. Histone octamers are eight-protein complexes consisting of two copies of four subunits (H2A, H2B, H3, and H4). The DNA that is associated with histone proteins is not accessible to the cellular machinery responsible for transcription, chromatin assembly, DNA repair, and a variety of other processes.³⁵ The tightness of this associ-

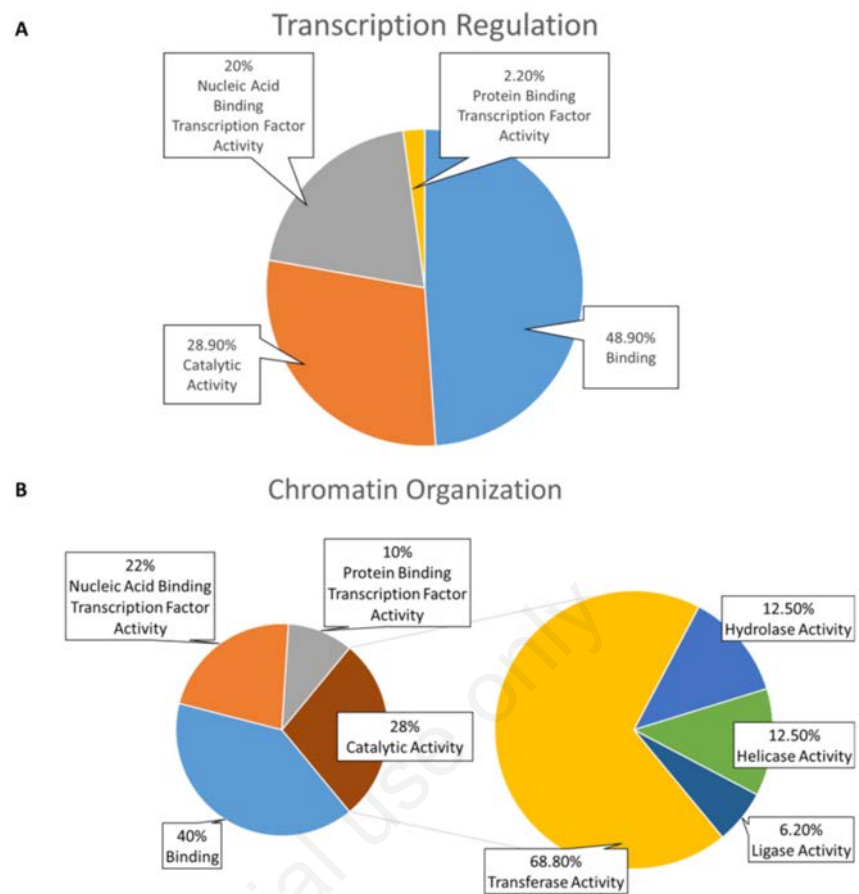


Figure 2. Classification of enrichment groups by molecular function. The two enriched groups, Transcription Regulation (A) and Chromatin Organization (B), are organized by their molecular function using the PANTHER classification tool.

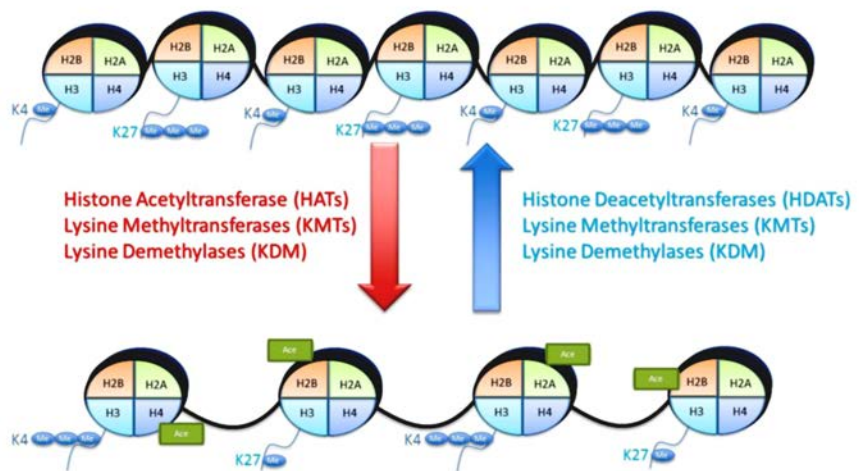


Figure 3. Chromatin organization through transferases. The transition between euchromatin and heterochromatin can be facilitated by various transferases. Acetylation of histones leads to the relaxation of chromatin structure. Histone methylation is more complex and depends on the location and extent of methylation

ation can create two different types of chromatin, heterochromatin and euchromatin. Heterochromatin occurs when the histone proteins are close together and undergo higher-ordered arrangements. This type of chromatin structure generally causes the genes associated with it to become silenced because the RNA polymerase is unable to bind to the promoter regions of the sequestered genes. Euchromatin occurs when the histone proteins disassociate with the DNA completely or spread apart, decreasing the amount of DNA directly associated with histones.

The structure of chromatin can be dynamic, and it is regulated by a group of proteins and complexes known as chromatin remodelers. These chromatin remodelers can modify the histone octamers in a variety of ways including acetylation, ubiquitination, phosphorylation, methylation, and sumoylation. The histone modifications either relax or tighten the structure of the localized DNA thereby permitting or restricting the transcription of nearby genes. Histone acetylation causes the relaxation of chromatin, thereby permitting an increase in the localized amount of transcription.³⁶ The acetylation of the N-terminal tails of histones removes the overall positive charge. When the histone's tail loses its positive charge, it loses its attraction to the negatively charged phosphate backbone of DNA as well as its attraction to other histones. Histone acetylation generally involves the transfer of an acetyl group from a molecule (such as acetyl coenzyme A) to a lysine amino acid residue on the tail of a histone octamer. Two opposing enzymes, histone acetyltransferases (HAT) and histone deacetyltransferases (HDAT), regulate the histone acetylation state (Figure 3).

One HAT with a predicted IDR is Histone Acetyltransferase of the CBP Family (HAC5) (Supplementary Table S1: Functional Annotation of Predicted IDPs). The HAT protein family of Arabidopsis consists of five CREB-binding proteins including HAC5.³⁷ The HAC family of proteins is implemented in a variety of cellular processes including cell differentiation, growth and homeostasis.³⁸ HAC5, in conjunction with HAC1, has been shown to be significant in the ethylene-signaling pathway. Interestingly, *hac1hac5* double mutants are hypersensitive in their ethylene signaling response in both dark and light conditions.³⁹ HAC5 has also implemented in a diverse group of developmental functions including flowering.^{40,41}

Histone methylation is not as straightforward as histone acetylation. Histone methyltransferases (HMT) can methylate lysine (K) and arginine (R) amino acid residues on both the H3 and H4 subunits of the histone complex. Additionally, histone methylation can lead to either transcriptional activation or repression, depending on the location of the

methylation and its surroundings. Nonetheless, histone methylation is an important key in epigenetic regulation of transcription. H3K4me3 is a key indicator of active transcription that is conserved in all eukaryotes.^{42,43} Early flowering in short days (EFS) is a well-studied SET-domain containing histone-lysine N-methyltransferase with predicted intrinsically disordered regions. EFS regulates flowering by actively promoting the flowering inhibitor *Flowering locus C* (FLC) via trimethylation of the localized histones at the H3K4 location.⁴⁴

Transcription factors

Transcription factors are modular proteins that contain one or more DNA-binding domains, recognize and bind to specific DNA sequences, therefore regulating the rate of transcription of genetic information from DNA to RNA. Generally, a prototypical transcription factor contains a DNA-binding domain (DBS), signal-sensing domain (SSD), and a transactivation domain (TAD). And the two processes involved in transcriptional regulations: protein-protein interaction and protein-nucleotide interaction have been reported to be accompanied often by a local folding in a protein molecule,⁴⁵ which may adjust the transcription factor flexibility correspondingly.

Seventy-eight different families of transcription factors contain different numbers of IDRs, according to our analysis. The first 10 enriched transcription factor families are MYB, AP2-EREBP, bHLH, MADS, C2H2, NAC, HB, WRKY, bZIP and C3H.

Four features of ID regions are very important for molecular recognition, including the combination of high specificity and low affinity in their interactions with functional partners, intrinsic plasticity, formation of large interaction surfaces and rapid turnover and reduced lifetime.⁴⁶ Besides, analysis of the distribution of disorder on transcription factors reflected the requirement of disorder regions for their flexibility and efficiency. It has been found in eukaryotes that the AT-hooks and basic regions of transcription factor DNA-binding domains are highly disordered, the degree of disorder in transcription factor activation regions is much higher than that in DNA-binding domains, and the level of -MoRFs (molecular recognition feature) prediction is much higher in transcription factors. MoRF is a specific structural element functioning in the recognition of protein or nucleic acid partners.

Long hypocotyl 5 (HY5) is a basic leucine zipper (bZIP) transcription factor family, involved in light-regulated transcriptional activation of G-box-containing promoters. It acts downstream of the light receptor network and directly affects transcription of light-induced genes, specifically involved in the blue light

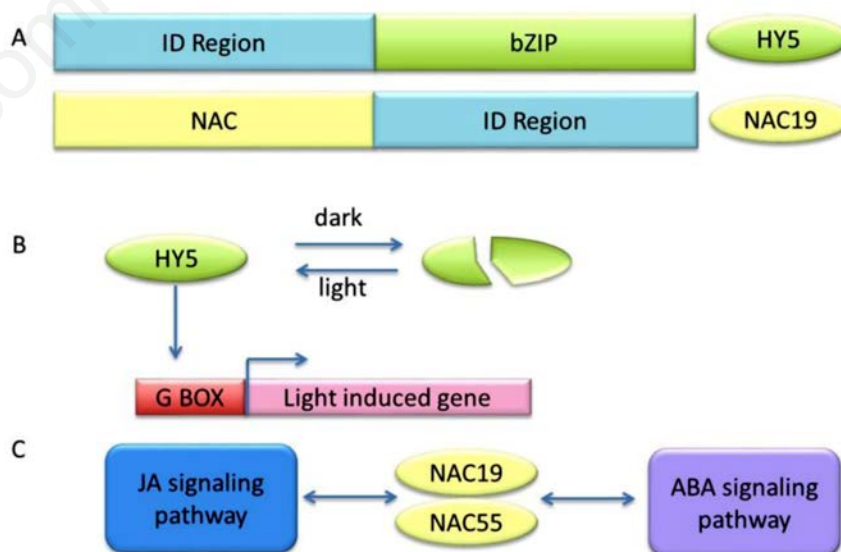


Figure 4. Model of the function of NAC and bZIP transcription factors. (A) Schematic organization of HY5 and NAC19 transcription factors. The ID regions are in blue, bZIP region is in green, NAC region is in yellow. (B) Model of the function of HY5. In darkness, the degradation of HY5 inhibits the activation of light-induced genes. After illumination, the accumulation of HY5 activates G-Box containing promoters and directly affects transcription of light-induced genes. (C) Role of NAC19 and NAC55 in the crosstalk between ABA and JA signaling pathway.

pathway, suggesting that it may participate in the transmission of cryptochromes (CRY1 and CRY2) signals. In darkness, the degradation of HY5 inhibits the activation of light-induced genes. It has been reported that HY5 is negatively regulated by COP1, a light-inactivable repressor of photomorphogenic development interacting directly and specifically with HY5.⁴⁷ Interestingly, the plant hormone cytokinin also induces similar phenotypes as the cryptochrome flacin-type photoreceptors. *hy5* mutants show a reduced induction of anthocyanin accumulation in blue light by cytokinins. It has been shown that cytokinins can increase the levels of HY5 protein accumulation, hinting that cytokinin could play a role in stabilizing HY5 protein, and that the regulation of HY5 stability could act at intersection of cytokinin signaling pathway and cryptochrome pathways.⁴⁸ Abscisic acid (ABA), another phytohormone, regulates seed germination and seedling development as light. It was found that HY5 binds to the promoter of the transcription factor ABI5, which is significantly enhanced by ABA, while overexpression of ABI5 led to increased light response.⁴⁹

NAC (NAM, ATAF, CUC) transcription factors share an N-terminal NAC domain and regulate stress perception and developmental programs. The crystal structure of the ANAC019 NAC domain consists of twisted β -sheet packing against a α -helix on both sides.^{50,51} The NAC transcription regulatory domains (TRDs) contain group-specific sequence motifs and have a high degree of intrinsic disorder.⁵² Both full-length and truncated ANAC019 are able to induce the expression of stress-responsive marker genes [COR47 (cold-responsive 47), RD29b (responsive-to-desiccation 29b) and ERD11 (early-responsive-to-dehydration11)]. Replacing the NAC domain of ANAC019 with the analogous regions from other NAC transcription factors still keeps the ability to regulate ABA signaling, while replacing the ANAC019 TRD with other TRDs loses the ABA signaling regulation ability.⁵² Further it has been shown that ANAC019 interacts with the RING-finger H2-type E3 ubiquitin-protein ligase RHA2a.⁵³ ANAC019 may play a dual role in regulating ABA and jasmonate response with the other RHA2a-interacting protein ANAC055. These two signaling pathways are involved in the activation of defense responses to both biotic and abiotic stresses, and ANAC019 and ANAC055 could serve as players linking the crosstalk between these two signaling pathways (Figure 4).⁵³ Additionally, two other ANAC proteins, ANAC046 and ANAC013, have recently been experimentally identified as IDPs. It has been shown that both of these aforementioned ANAC transcription factors are capable of interacting with the small hub protein known as Radical-induced Cell Death 1 (RCD1), at least partially due to their intrinsi-

cally disordered region. Both proteins are involved in plant senescence. Interestingly, the interaction between the two TFs and RCD1 does not appear to be dependent on a disorder-to-order transition.⁵⁴ Future research could focus on determining if ANAC019 interacts with its binding partners in a similar way.

Conclusions

The discovery of intrinsically disordered proteins created a novel area of proteomics. Specifically, the increased knowledge of protein folding dynamics may lead to a better understanding of plants' phenotypic plasticity. Because plants are fixed in soil and unable to move, they create complex mechanisms for coping with biotic and abiotic environmental stresses. The concept of flexible proteins, which are able to change their interaction profile based on cellular conditions, leads to a new way of thinking about plant plasticity. We have highlighted tools that are useful for characterizing IDPs' conformational dynamics. Additionally, we have demonstrated that Arabidopsis contains two highly enriched groups of IDPs; both of which are key components in the transcriptional regulation landscape. Further studies need to focus on experimentally verifying proteins as IDPs, and describing the advantages and disadvantages of intrinsically disordered regions over their more structured counterparts. Furthermore, IDPs have been recent targets for therapeutic strategies for mammalian diseases such as Parkinson's disease.⁵⁵ Similarly, IDPs found in plants could be targets for decreasing crops' susceptibility to disease or increasing overall yield. Additional research could focus on identifying key IDP targets for both therapeutics and plant resistance.

References

- 1 Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci* 2002;27:527-3.
- 2 Ward JJ, Sodhi JS, McGuffin LJ, et al. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004;337:635-45.
- 3 Dunker AK, Lawson JD, Brown CJ, et al. Intrinsically disordered protein. *J Mol Graph Model* 2001;19:26-59.
- 4 Marin M, Ott T. Intrinsic disorder in plant proteins and phytopathogenic bacterial effectors. *Chem Rev* 2014;114:6912-32.
- 5 Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*

1999;293:321-31.

- 6 Huang Y, Liu Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the fly-casting mechanism. *J Mol Biol* 2009;393:1143-59.
- 7 Tantos A, Han KH, Tompa P. Intrinsic disorder in cell signaling and gene transcription. *Mol Cell Endocrinol* 2012;348:457-65.
- 8 Dunker AK, Oldfield CJ, Meng J, et al. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 2008;9:S1.
- 9 Higo J, Umezawa K. Free-energy landscape of intrinsically disordered proteins investigated by all-atom multicanonical molecular dynamics. *Adv Exp Med Biol* 2014;805:331-51.
- 10 Jensen MR, Zweckstetter M, Huang JR, Blackledge M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev* 2014;114:6632-60.
- 11 Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 2002;11:739-56.
- 12 Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005;6:197-208.
- 13 Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta* 2010;1804:1231-64.
- 14 Metallo SJ. Intrinsically disordered proteins are potential drug targets. *Curr Opin Chem Biol* 2010;14:481-8.
- 15 Uversky VN. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept. *Expert Rev Proteomics* 2010;7:543-64.
- 16 Wang J, Cao Z, Zhao L, Li S. Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). *Int J Mol Sci* 2011;12:3205-19.
- 17 Ganguly D, Chen J. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J Mol Biol* 2009;390:467-77.
- 18 Allison JR, Varnai P, Dobson CM, Vendruscolo M. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J Am Chem Soc* 2009;131:18314-26.
- 19 Lindorff-Larsen K, Best RB, Depristo MA, et al. Simultaneous determination of protein structure and dynamics. *Nature* 2005;433:128-32.
- 20 Krzeminski M, Marsh JA, Neale C, et al. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 2013;29:398-9.
- 21 Chen Y, Campbell SL, Dokholyan NV. Deciphering protein dynamics from NMR

- data using explicit structure sampling and selection. *Biophys J* 2007;93:2300-6.
- 22 Nodet G, Salmon L, Ozenne V, et al. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc* 2009;131:17908-18.
- 23 Daughdrill GW, Kashtanov S, Stancik A, et al. Understanding the structural ensembles of a highly extended disordered protein. *Mol Biosyst* 2012;8:308-19.
- 24 Massova I, Kollman P. Computational alanine scanning to probe protein-protein interactions: a binding free energies. *J Am Chem Soc* 1999;121:10.
- 25 Wilhelm KS, Thomashow MF. *Arabidopsis thaliana* cor15b, an apparent homologue of cor15a, is strongly responsive to cold and ABA, but not drought. *Plant Mol Biol* 1993;23:1073-7.
- 26 Artus NN, Uemura M, Steponkus PL, et al. Constitutive expression of the cold-regulated *Arabidopsis thaliana* COR15a gene affects both chloroplast and protoplast freezing tolerance. *Proc Natl Acad Sci USA* 1996;93:13404-9.
- 27 Steponkus PL, Uemura M, Joseph RA, et al. Mode of action of the COR15a gene on the freezing tolerance of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 1998;95:14570-5.
- 28 Hundertmark M, Hinch DK. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 2008;9:118.
- 29 Thalhammer A, Hundertmark M, Popova AV, et al. Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state. *Biochim Biophys Acta* 2010;1798:1812-20.
- 30 Pietrosevoli N, Garcia-Martin JA, Solano R, Pazos F. Genome-wide analysis of protein disorder in *Arabidopsis thaliana*: implications for plant environmental adaptation. *PLoS One* 2013;8:e55524.
- 31 Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1-13.
- 32 Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 2013;41:D377-86.
- 33 Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 2013;8:1551-66.
- 34 Zhou Y, Kim J, Yuan X, Braun T. Epigenetic modifications of stem cells: a paradigm for the control of cardiac progenitor cells. *Circ Res* 2011;109:1067-81.
- 35 Marino-Ramirez L, Kann MG, Shoemaker BA, Landsman D. Histone structure and nucleosome stability. *Expert Rev Proteomics* 2005;2:719-29.
- 36 Gorisch SM, Wachsmuth M, Toth KF, et al. Histone acetylation increases chromatin accessibility. *J Cell Sci* 2005;118:5825-34.
- 37 Pandey R, Müller A, Napoli CA, et al. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res* 2002;30:5036-55.
- 38 Bordoli L, Netsch M, Luthi U, et al. Plant orthologs of p300/CBP: conservation of a core domain in metazoan p300/CBP acetyltransferase-related proteins. *Nucleic Acids Res* 2001;29:589-97.
- 39 Li C, Xu J, Li J, et al. Involvement of *Arabidopsis* histone acetyltransferase HAC family genes in the ethylene signaling pathway. *Plant Cell Physiol* 2014;55:426-35.
- 40 Li C, Xu J, Li J, et al. Involvement of *Arabidopsis* HAC family genes in pleiotropic developmental processes. *Plant Signal Behav* 2014;9:e28173.
- 41 Han SK, Song JD, Noh YS, Noh B. Role of plant CBP/p300-like genes in the regulation of flowering time. *Plant J* 2007;49:103-14.
- 42 Berr A, Xu L, Gao J, et al. Set domain group25 encodes a histone methyltransferase and is involved in flowering locus C activation and repression of flowering. *Plant Physiol* 2009;151:1476-85.
- 43 Santos-Rosa H, Schneider R, Bannister AJ, et al. Active genes are tri-methylated at K4 of histone H3. *Nature* 2002;419:407-11.
- 44 Shafiq S, Berr A, Shen WH. Combinatorial functions of diverse histone methylations in *Arabidopsis thaliana* flowering time regulation. *New Phytol* 2014;201:312-22.
- 45 Spolar RS, Record MT, Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 1994;263:777-84.
- 46 Liu J, Perumal NB, Oldfield CJ, et al. Intrinsic disorder in transcription factors. *Biochemistry* 2006;45:6873-88.
- 47 Ang LH, Chattopadhyay S, Wei N, et al. Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of *Arabidopsis* development. *Mol Cell* 1998;1:213-22.
- 48 Vandenbussche F, Habricot Y, Condiff AS, et al. HY5 is a point of convergence between cryptochrome and cytokinin signalling pathways in *Arabidopsis thaliana*. *Plant J* 2007;49:428-41.
- 49 Chen H, Zhang J, Neff MM, et al. Integration of light and abscisic acid signaling during seed germination and early seedling development. *Proc Natl Acad Sci USA* 2008;105:4495-500.
- 50 Ernst HA, Olsen AN, Larsen S, Lo Leggio L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Rep* 2004;5:297-303.
- 51 Greve K, La Cour T, Jensen MK, et al. Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization. *Biochem J* 2003;371:97-108.
- 52 Jensen MK, Kjaersgaard T, Nielsen MM, et al. The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem J* 2010;426:183-96.
- 53 Jiang H, Li H, Bu Q, Li C. The RHA2a-interacting proteins ANAC019 and ANAC055 may play a dual role in regulating ABA response and jasmonate response. *Plant Signal Behav* 2009;4:464-6.
- 54 O'Shea C, Kryger M, Stender EG, et al. Protein intrinsic disorder in *Arabidopsis* NAC transcription factors: transcriptional activation by ANAC013 and ANAC046 and their interactions with RCD1. *Biochem J* 2015;465:281-94.
- 55 Toth G, Gardai SJ, Zago W, et al. Targeting the intrinsically disordered structural ensemble of alpha-synuclein by small molecules as a potential therapeutic strategy for Parkinson's disease. *PLoS One* 2014;9:e87133.