

Making Sense of Intrinsically Disordered Proteins

H. Jane Dyson^{1,*}

¹Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California

Proteins form the molecular scaffolding of life and are essential to catalyzing the chemical reactions that sustain living systems. These characteristics have led us to think that proteins function only when folded into the right structure. The central dogma of molecular biology states that genetic information encoded in the DNA sequence is transcribed into messenger RNA and then translated into a sequence of amino acids, which folds into a protein. The mechanisms that govern how a linear sequence of amino acids folds into the correct three-dimensional structure are still not well understood. Biophysical techniques have been indispensable to unraveling how protein structures fold, and many of the major factors that determine how the amino-acid sequence codes for the folded protein structure are beginning to be understood.

The genomic era that began at the end of the 20th century gave scientists access to complete genome sequences. Scientists observed that some of the predicted protein sequences derived from genomes were not expected to fold into normal globular protein structures (1). At the same time, experimental studies began to uncover examples of important protein molecules and domains that were incompletely structured or completely disordered in solution yet remained perfectly functional (2,3) (Fig. 1). In the following years, an explosion of experimental data and genome annotation studies mapped the extent of this intrinsic disorder phenomenon and explored the possible biological reasons for its widespread occurrence. Answers to the question of why a particular domain would need to be unstructured are as varied as the systems where such domains are found.

One of the hallmarks of intrinsically disordered proteins (IDPs) is a marked bias in the amino-acid composition, including a relatively low proportion of hydrophobic and aromatic residues, and a relatively high proportion of charged and polar residues (Fig. 2). The high frequency of small hydrophilic amino acids renders these sequences as unlikely candidates for membrane or scaffolding proteins. Yet many of the proteins identified in surveys, as well as in concurrent NMR experiments, showed that these proteins were involved in important cellular processes such as control of the cell cycle, transcriptional activation, and signaling (4,5), and they frequently interacted with or functioned as central hubs in protein interaction networks (6). The

amounts of various IDPs in the cell are tightly regulated to ensure fidelity in signaling. Altered abundance of IDPs is associated with disease (7).

Disordered sequences can also be found in proteins that contain ordered, structured domains, and these disordered sequences are termed intrinsically disordered regions (IDRs). Some IDRs function as linkers between interaction domains (Fig. 2), and in some cases, their properties as polymers contribute to their function (8). Many IDRs contain sequence elements that interact with partners and frequently fold upon binding. For example, the intrinsically disordered interaction domain of the transcription factor STAT2 folds upon binding to its partner, the TAZ1 domain of CREB-binding protein (CBP) (Fig. 3) (9). Backbone flexibility of an IDR in its free state enables it to bind to multiple targets, which increases its potential repertoire of responses, as exemplified in the binding of the hypoxia-inducible factor HIF-1 α . The transactivation domain of HIF-1 α binds to its partner TAZ1 as a helix (10), whereas the same HIF-1 α sequence binds to the hydroxylating enzyme FIH as a β -strand (Fig. 4) (11).

Disorder makes IDR sequences accessible to posttranslational modification and IDRs are rich in modification sites. IDRs facilitate efficient protein-protein interactions using only a small number of residues. A folded protein would need to be much larger to provide an interaction surface area equivalent to that seen with IDRs, as illustrated in Fig. 4. This efficiency is important in signaling, as it translates into the ability to bind with high specificity but only modest affinity, enabling dissociation of the IDR after signaling is complete. Signaling can be turned off by competition between IDRs for a particular physiological partner, mediated by slightly different binding sites (Fig. 5). The reaction of cells to hypoxia (low oxygen) is a good example of this phenomenon (Fig. 6). Under normal conditions, the HIF protein is synthesized in the cell, but is degraded upon hydroxylation of two prolines. Interaction with the transcriptional coactivator CBP is further interdicted by the hydroxylation of an asparagine in the C-terminal activation domain (CTAD). Under hypoxic conditions, the hydroxylation reactions no longer occur, so the protein is stable to degradation and the CTAD can interact with the TAZ1 domain of CBP, leading to transcription of hypoxia-response genes such as VEGF, which promotes growth of blood vessels (12). Such a response is dangerous if not constrained, however, and the signal must be turned off before adverse physiological effects occur. One of the genes

Submitted July 30, 2015, and accepted for publication October 29, 2015.

*Correspondence: dyson@scripps.edu

© 2016 by the Biophysical Society
0006-3495/16/03/1013/4



<http://dx.doi.org/10.1016/j.bpj.2016.01.030>

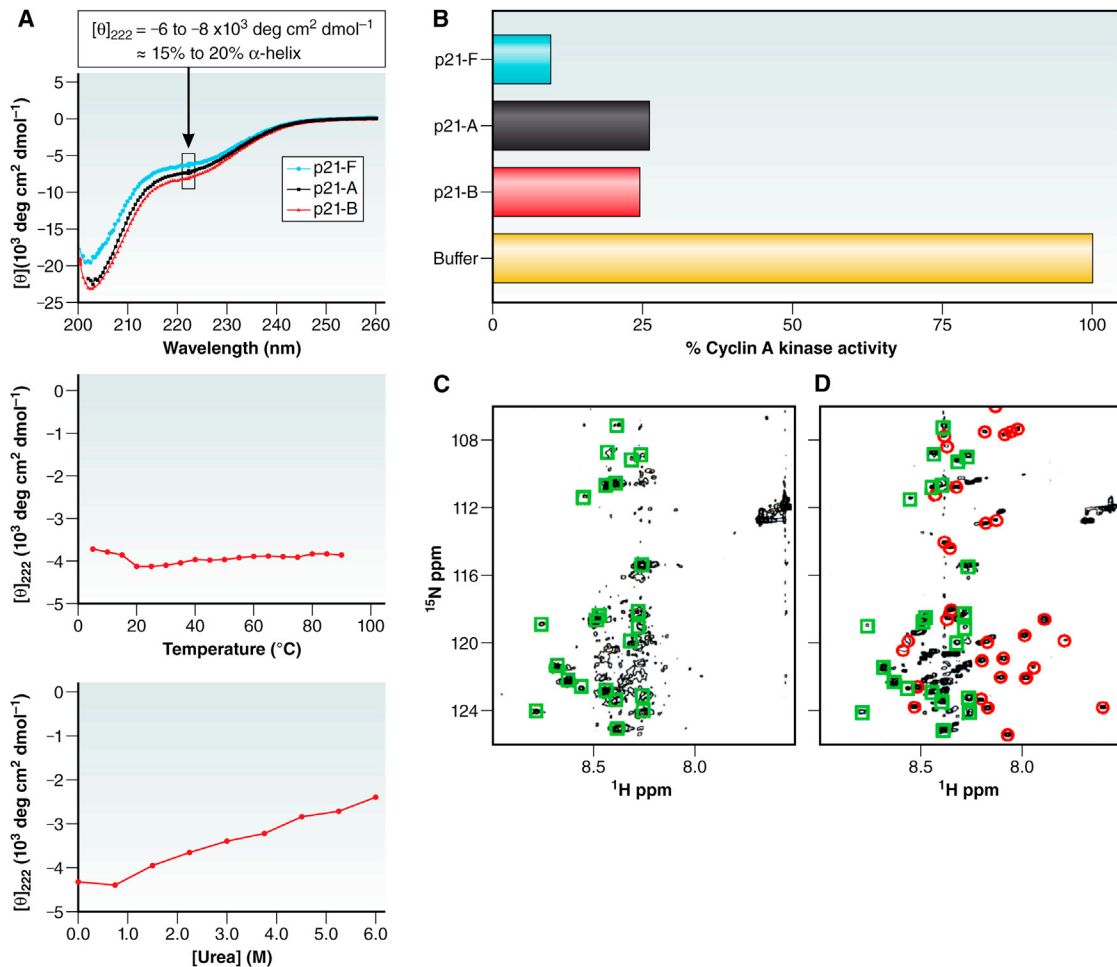


FIGURE 1 (A) The cyclin-dependent kinase inhibitor p21 does not form recognizable structure in solution, indicated by the far-UV circular dichroism spectrum of constructs of various lengths (*top panel*), by the absence of temperature-induced (*middle panel*), or urea-induced (*bottom panel*) unfolding transitions. (B) All of the p21 constructs are active in the inhibition of cyclin-A kinase activity. (C and D) ^1H - ^{15}N HSQC NMR spectra of free p21 (C) and p21 bound to cyclin-dependent kinase-2. (D) Squares indicate cross peaks present in the same place in the two spectra, and circles denote new cross peaks at positions that indicate that folding has occurred (adapted from Kriwacki et al. (2) with permission, © 1996 National Academy of Sciences, USA). To see this figure in color, go online.

transcribed in response to the hypoxia signal encodes an inhibitor, CITED2, which removes the HIF CTAD from the CBP TAZ1 domain, turning off the hypoxia response.

Several anti-cancer treatments include inhibition of the hypoxia response, which prevents vascularization of tumors, thus restricting their growth (12).



FIGURE 2 Illustration of the sequence bias found in disordered sequences. The amino-acid sequence of a portion of the multidomain transcriptional coactivator CBP is classified by amino-acid type: green, small hydrophilic amino acids (G, A, S, T, N, Q, P); yellow, hydrophobic amino acids (V, L, I, M, F, Y, W); red, acidic amino acids (D, E); blue, basic amino acids (K, R, H); pink, cysteine (C). The sequences of two folded domains, TAZ1 (blue) and KIX (yellow), show much greater sequence diversity than the disordered flanking and linker domains, which are predominantly green, indicating a heavy bias toward small hydrophilic amino acids (adapted from Dyson and Wright (4)).

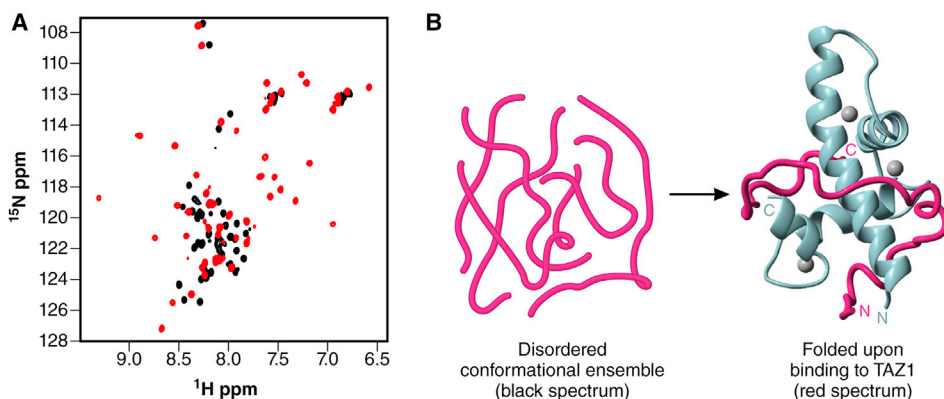


FIGURE 3 Coupled folding and binding of the transcription factor STAT2 on the TAZ1 domain of CBP. (A) Disorder in the free STAT2 is shown in the small resonance dispersion in the ^1H dimension of the black ^1H - ^{15}N HSQC spectrum. The structured nature of the bound STAT2 is shown by the increased ^1H dispersion of the gray spectrum. (B) Schematic diagram illustrating the conversion of the disordered conformational ensemble of free STAT2 into a structured form on the TAZ1 (adapted from Wojciak et al. (9)). To see this figure in color, go online.

IDPs offer novel advantages as therapeutic targets. Their central role in key cellular signaling pathways (5), their frequent association with disease (7), and the reversible nature of their intermolecular interactions, by which they bind with high specificity but modest affinity, makes them extremely attractive targets for small molecule drugs or stapled peptide mimetics (13,14). Indeed, many viruses hijack the host cell by using their own viral IDPs, e.g., the adenovirus E1A or papillomavirus E7 oncoproteins (15,16), to compete with cellular IDPs for binding to key regulatory proteins (17). IDPs commonly bind to concave grooves in the surface of their target proteins; the interactions are predominantly hydrophobic and the fit is more intimate than to their globular protein counterparts. Finally, it may prove possible to design drugs targeted against the IDP itself, rather than its globular target.

Disordered regions of proteins provide a uniquely versatile and useful toolbox for reactions in the cell. Interestingly, the majority of IDRs so far characterized are from eukaryotic systems, in which they are intimately involved with the signaling and physiological control required for multicellular organisms. IDRs are found in prokaryotes, but

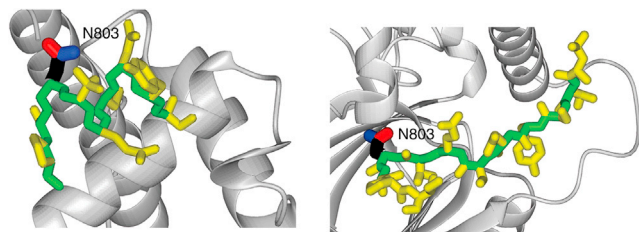


FIGURE 4 Structural differences between the transactivation domain of HIF-1 α bound to the TAZ1 domain of CBP (10) (left panel), where the sequence containing the regulatory asparagines appears as an α -helix, and the same sequence bound to the hydroxylating enzyme FIH (11) (right panel), where it appears as a β -strand. To see this figure in color, go online.

they tend to be associated with unusual functions in particular bacteria, for example the toxin-antitoxin systems of phage-infected *Escherichia coli* (18).

Protein molecules are rarely, if ever, completely rigid. Dynamic motions of backbone and side chains, independent of the tumbling of the whole molecule, can be estimated by various spectroscopic means, and are frequently associated with the function of enzymes. Disordered regions and fully disordered proteins can be thought of as a continuation of this characteristic, by which functional disorder, an extreme form of local protein dynamics, is functional through the particular advantages bestowed by the disordered state. The continuum between rigidity and complete disorder provides an expanded proteome, allowing proteins to perform multiple tasks through interactions with different partners or under different conditions. Disorder occupies an important biological niche that promises

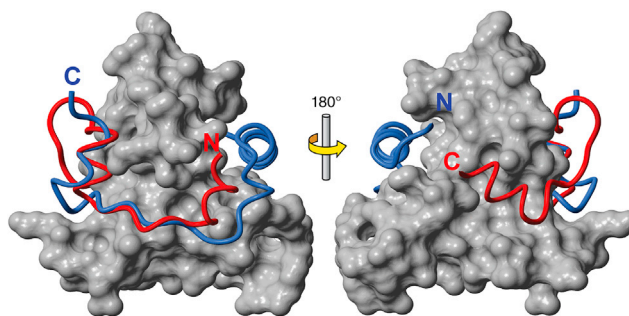


FIGURE 5 Comparison of the structures of ligands bound to the CBP TAZ1 domain. The surface of TAZ1 (almost identical in both complexes) is shown in gray, with the backbone of HIF-1 α -C-terminal activation domain (10) in red (labeled N in the left image and C in the right image) and the CITED2-*trans*-activation domain (19) in blue (labeled C in the left image and N in the right image). (Left and right panels) 180° rotation around the vertical axis in the plane of the page (adapted from Wojciak et al. (9)). To see this figure in color, go online.

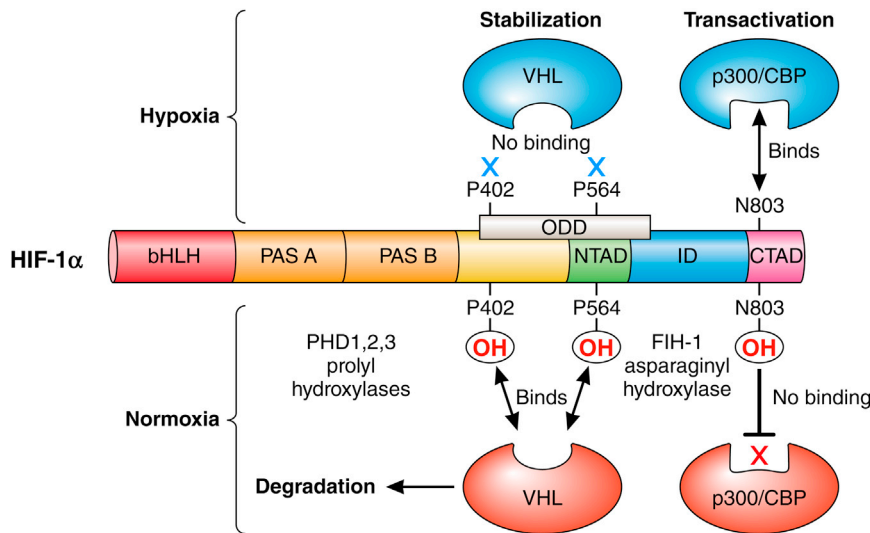


FIGURE 6 Schematic diagram showing the regulation of the HIF-1 α transcription factor under normal oxygenation conditions (*bottom*), where proline hydroxylation in the central oxygen-dependent degradation domain recruits the von Hippel-Lindau factor, leading to degradation, and asparagine hydroxylation in the C-terminal activation domain lowers the affinity for transcriptional activators. In hypoxic conditions (*top*), neither the prolines nor the asparagines are hydroxylated, with the result that HIF-1 α is stabilized and binds to CBP/p300 to promote transcription of hypoxia-response genes. (Reproduced from Dyson (20) with permission.) To see this figure in color, go online.

to yield important new insights into how biological systems operate.

ACKNOWLEDGMENTS

This work was supported by grant Nos. GM071862 and GM113251 from the National Institutes of Health.

REFERENCES

- Romero, P., Z. Obradovic, ..., A. K. Dunker. 1997. Identifying disordered regions in proteins from amino acid sequences. *Proc. IEEE*. 1:90–95.
- Kriwacki, R. W., L. Hengst, ..., P. E. Wright. 1996. Structural studies of p21^{Waf1/Cip1/Sd1} in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. USA*. 93:11504–11509.
- Daughdrill, G. W., M. S. Chadsey, ..., F. W. Dahlquist. 1997. The C-terminal half of the anti- σ factor, FlgM, becomes structured when bound to its target, σ 28. *Nat. Struct. Biol.* 4:285–291.
- Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.
- Wright, P. E., and H. J. Dyson. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16:18–29.
- Dunker, A. K., M. S. Cortese, ..., V. N. Uversky. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272:5129–5148.
- Babu, M. M., R. van der Lee, ..., J. Gsponer. 2011. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* 21:432–440.
- Tompa, P. 2010. Structure and Function of Intrinsically Disordered Proteins. Chapman & Hall, Boca Raton, FL.
- Wojciak, J. M., M. A. Martinez-Yamout, ..., P. E. Wright. 2009. Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *EMBO J.* 28:948–958.
- Dames, S. A., M. Martinez-Yamout, ..., P. E. Wright. 2002. Structural basis for Hif-1 α /CBP recognition in the cellular hypoxic response. *Proc. Natl. Acad. Sci. USA*. 99:5271–5276.
- Elkins, J. M., K. S. Hewitson, ..., C. J. Schofield. 2003. Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 α . *J. Biol. Chem.* 278:1802–1806.
- Krock, B. L., N. Skuli, and M. C. Simon. 2011. Hypoxia-induced angiogenesis: good and evil. *Genes Cancer*. 2:1117–1133.
- Vassilev, L. T., B. T. Vu, ..., E. A. Liu. 2004. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*. 303:844–848.
- Chang, Y. S., B. Graves, ..., T. K. Sawyer. 2013. Stapled α -helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. USA*. 110:E3445–E3454.
- Ferreon, J. C., M. A. Martinez-Yamout, ..., P. E. Wright. 2009. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl. Acad. Sci. USA*. 106:13260–13265.
- Jansma, A. L., M. A. Martinez-Yamout, ..., P. E. Wright. 2014. The high-risk HPV16 E7 oncoprotein mediates interaction between the transcriptional coactivator CBP and the retinoblastoma protein pRb. *J. Mol. Biol.* 426:4030–4048.
- Davey, N. E., G. Travé, and T. J. Gibson. 2011. How viruses hijack cell regulation. *Trends Biochem. Sci.* 36:159–169.
- Garcia-Pino, A., S. Balasubramanian, ..., R. Loris. 2010. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell*. 142:101–111.
- De Guzman, R. N., M. A. Martinez-Yamout, ..., P. E. Wright. 2004. Interaction of the TAZ1 domain of the CREB-binding protein with the activation domain of CITED2: regulation by competition between intrinsically unstructured ligands for non-identical binding sites. *J. Biol. Chem.* 279:3042–3049.
- Dyson, H. J. 2011. Expanding the proteome: disordered and alternatively folded proteins. *Q. Rev. Biophys.* 44:467–518.