Intrinsically disordered proteins

Peter Tompa and Kyou-Hoon Han

Researchers are discovering an ever-increasing number of proteins that perform key cellular tasks without having the fixed, three-dimensional structure once thought mandatory for a protein to do its job.

Peter Tompa is a professor at the Institute of Enzymology of the Hungarian Academy of Sciences in Budapest and at the Flanders Institute for Biotechnology (VIB) department of structural biology in Brussels, Belgium. **Kyou-Hoon Han** is a professor at the University of Science and Technology and a senior investigator at the Korea Research Institute of Bioscience and Biotechnology, both in Daejeon, South Korea.

Proteins are the protagonists of life. They catalyze reactions, transmit information, provide scaffolding both inside and outside cells, and transport material throughout the body; they even respond to environmental threats. Despite their bewildering multiplicity of functions, proteins all share the same simple design: Every protein is a linear chain, each element, or residue, of which is one of the 20 amino acids. The variety of proteins comes from the many possible permutations of those amino acids and the resulting ability of the protein chains to fold into three-dimensional structures of seemingly limitless diversity.

The natural disorder of things

Once upon a time, it was dogma that proteins function only if they acquire their natural, or native, structure. Indeed, much of the community's understanding of protein function rests on our ability to deduce those structures by such methods as x-ray crystallography and nuclear magnetic resonance (NMR). The immense success and explanatory power of the structure–function paradigm is witnessed by the more than 80 000 protein structures in the Protein Data Bank and by countless works that have elucidated the function of enzymes, receptors, and so-called structural proteins. Nonetheless, we now understand that for many important proteins, or at least for regions of many proteins, the native, functional state is unstructured. Those intrinsically disordered proteins (IDPs) defy the structure–function model.

Rather than folding into a single, stable, 3D structure, IDPs exist as an ensemble of rapidly interconverting conformations that resemble the denatured states of ordered proteins—the states of proteins exposed, for example, to acids or other harmful chemicals. Structural disorder occurs in some of the most important and amply studied proteins involved in fatal disease. Those include α-synuclein and prion protein, both of which may be implicated in neurodegenerative disorders, and the tumor-suppressing anticancer protein p53.

But proteins connected to diseases are only the tip of the iceberg: Structural disorder is rather common in the higher eukaryotes, organisms whose cells have nuclei. In humans, 10% of all proteins are fully disordered, and an additional 40% have at least one disordered region more than 30 residues long. Structural disorder is most prevalent in proteins having regulatory roles. Such tasks include the determination of the cell's response to an external stimulus; transcription, which is the process of creating an RNA strand complementary to a given DNA strand; assistance in the folding and unfolding of macromolecular structures in the cell; and translation, the conversion of genetic information into proteins. On the other hand, structural disorder is rarely observed in enzymes, receptors, and structural proteins that, like the hemoglobin protein shown in panel a of the figure, require the precise spatial positioning of residues involved in ligand binding and catalysis.

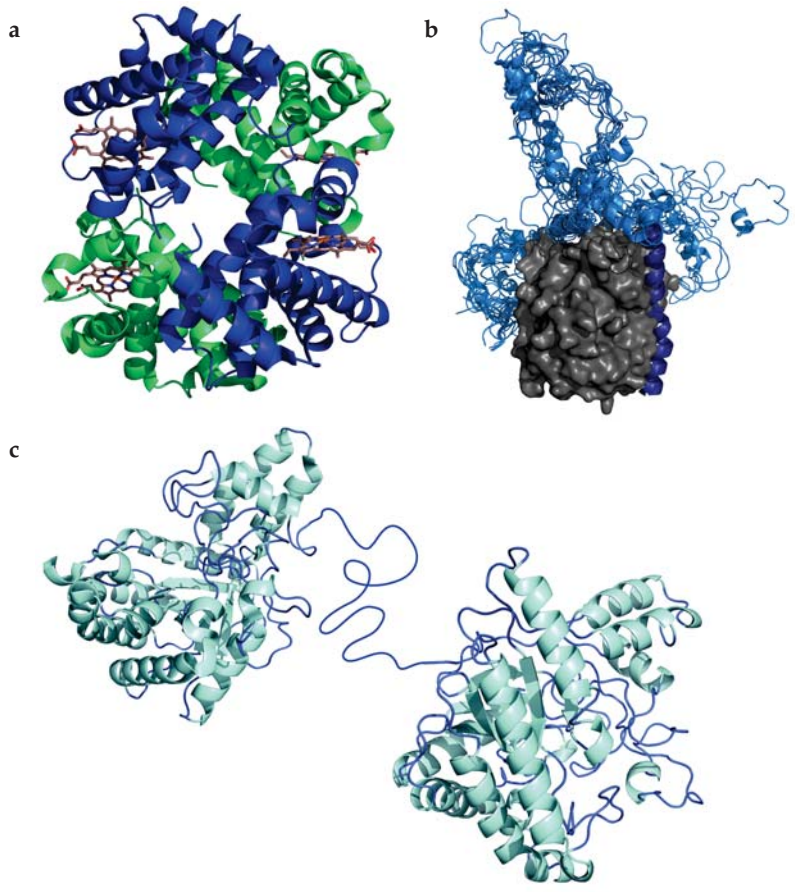
A long experimental road

For about a century, protein studies tended to address the objects that were most prevalent and easy to assay: enzymes. As recognized by Emil Fischer in 1894, enzyme specificity comes from a lock-and-key mechanism for substrate binding, which requires that the enzyme be well structured. Fischer's view was later substantiated by observations that when an enzyme is denatured by environmental conditions, it suffers a loss of function; moreover, the activity of the denatured enzyme returns only when its structure is restored. All in all, the spectacular success of structural biology and the functional insight it generated appeared to make an airtight case for the importance of structure in protein action.

In retrospect, though, it becomes clear that experimental work from the late 20th century conclusively demonstrated structural disorder in proteins and protein domains. As early as the 1950s, Raman spectroscopy showed that milk caseins are unfolded. Many proteins could not be crystallized, but those failures did not carry much weight because many proteins known to be structured also could not be crystallized. Small-angle x-ray scattering (SAXS) of proteins indicated an unexpectedly large size (the technical term is hydrodynamic dimension)—a consequence, we now know, of unfolded regions. And a number of proteins displayed notably strange behavior: Unlike structured, "globular," proteins, they were heat or acid resistant.

Even collectively, though, such observations failed to tip the community consensus to acknowledging disorder in protein structure. That feat was accomplished by NMR studies, which demonstrated physiologically relevant protein disorder under native conditions and provided structural and dynamical information at the residue level. For example, NMR studies showed that although IDPs are largely disordered, they have significant local structure in the form of transient structural motifs. Those are functional regions and may be considered as the active sites of IDPs. Other NMR work revealed the consequences of the extremely high concentration of macromolecules in the intracellular milieu; the ensuing crowding can shift the conformational equilibria of IDPs toward more compact, folded states.

Nuclear magnetic resonance studies have also contributed to the most exciting recent advance in the IDP field: the description of structural ensembles. In practice, biophysicists take NMR observations of proteins in action, sometimes



Form—and disorder—follows function. The traditional structure–function paradigm asserts that a protein has to acquire a unique structure to ensure the proper spatial localization of residues important in catalysis or binding. **(a)** Highly structured hemoglobin, the protein responsible for oxygen transport in the blood, attests to the structure–function idea. **(b)** On the contrary, many proteins lack a well-defined structure; they are intrinsically disordered. Inhibitor 2 (blue), the regulator of protein phosphatase 1 (gray), is significantly disordered even when it is bound to its partner. Indeed, the inhibitor does not act in a single conformation but rather as an ensemble of conformations. This image shows an ensemble of 10 different conformations selected on the basis of nuclear magnetic resonance data. The disordered state and independence of short binding motifs enable complex regulation of the enzyme. **(c)** Bacterial cellulase has two structured domains connected by a disordered linker that enables the domains to move relatively independently.

supplemented by SAXS experiments, and fit the data with as many as 100 structures selected from a random, computer-generated set. Panel b of the figure shows an example in which the binding of the protein inhibitor 2 is displayed as an ensemble of 10 different configurations.

The peculiar structural state of IDPs resembles not only denatured globular proteins but also chemical polymers. There is a crucial difference, of course; only the IDPs have biological function. Still, the IDP field has learned much from classical studies of polymer chemistry, and indeed, a good deal of the field's concepts and terminology came from earlier studies of polymers, via studies of protein denaturation and folding. For example, the structural elements of IDPs are often termed residual structure in witness to that legacy.

Disorder enables independence

Some IDPs derive their functionality directly from structural disorder and are described as having entropic chain functionality; panel c shows an example from the enzyme bacterial cellulase. Chains such as those seen in the figure often appear as linkers in multidomain proteins and enable the IDP to flexibly search for binding partners. That flexibility influences the kinetics, thermodynamics, and specificity of the action of the protein. In the case of bacterial cellulase, the entropic chain functionality allows the enzyme to cleave its macroscopic cellulose substrate many times without having to release.

Other IDPs operate through a process called molecular recognition, in which the active sites of the IDP weakly bind to a target molecule. The protein's disorder increases the interaction speed and allows the IDP to adapt to distinct partners. Inhibitor 2, as shown in panel b, is an example of such an IDP. Its interaction with the enzyme protein phosphatase 1 results in a complex regulation as, in response to diverse cellular signals, the enzyme–inhibitor complex transits between inhibited, de-inhibited, and activated states. The complex regulation is possible only because of the structural disorder in inhibitor 2, which enables various segments of the protein to independently bind or release.

For almost 100 years, the structure–function paradigm served biologists well, and indeed, it is still the best paradigm for understanding enzymes. But nature is subtle, and also acts through IDPs. The varied, important, and fascinating functions of those proteins fully justifies the increasing study they are receiving.

Additional resources

▶ H. Lee et al., "Local structural elements in the mostly unstructured transcriptional activation domain of human p53," *J. Biol. Chem.* **275**, 29426 (2000).

▶ V. N. Uversky, "Natively unfolded proteins: A point where biology waits for physics," *Protein Sci.* **11**, 739 (2002).

▶ P. Tompa, M. Fuxreiter, "Fuzzy complexes: Polymorphism and structural disorder in protein–protein interactions," *Trends Biochem. Sci.* **33**, 2 (2008).

▶ P. Tompa, *Structure and Function of Intrinsically Disordered Proteins*, CRC Press, Boca Raton, FL (2010).

▶ S.-H. Lee et al., "Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins," *Curr. Protein Pept. Sci.* **13**, 34 (2012).

▶ J. A. Marsh et al., "Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators," *Structure* **18**, 1094 (2010).