

the orderly chaos of proteins

To do their magic in the cell, proteins must fold into rigid shapes—or so standard wisdom says. But a more tangled story is beginning to emerge

By A. Keith Dunker and Richard W. Kriwacki

IN BRIEF

According to conventional wisdom, proteins must fold into rigid shapes to perform such tasks as binding to specific target molecules. But recent work suggests that one third of the types that exist in humans are partially or completely unstructured.

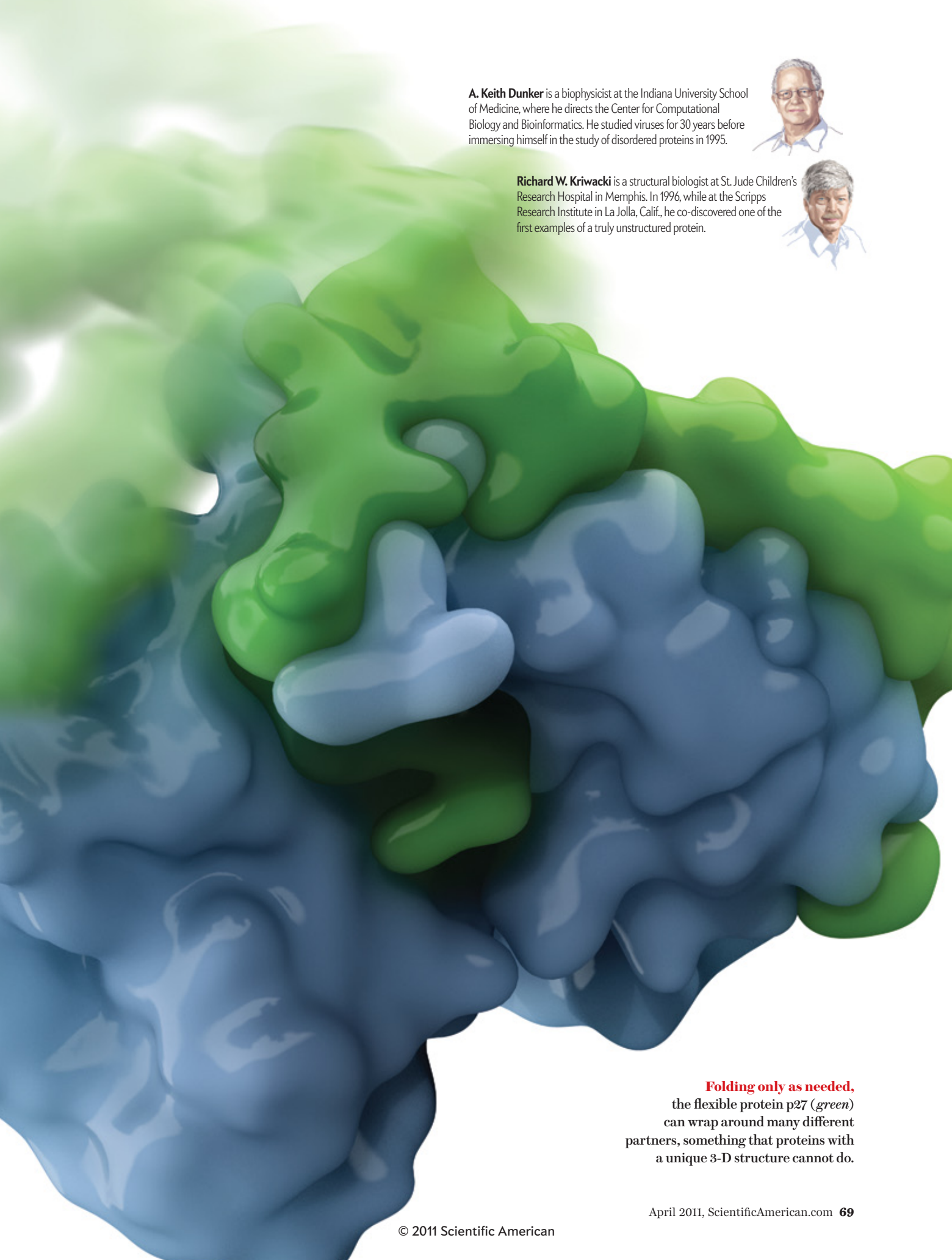
Although lack of folding was long considered a pathology, it need not hamper functionality—and it is in fact often crucial to a protein's workings.

Unstructured proteins may have played important roles during evolution, and a better understanding of their true nature may also lead to the design of novel drugs.

PROTEINS ARE THE STUFF OF life. They are the eyes, arms and legs of living cells. Even DNA, the most iconic of all molecules in biology, is important first and foremost because it contains the genes that specify the makeup of proteins. And the cells in our body differ from one another—serving as neurons, white blood cells, smell sensors, and so on—largely because they activate different sets of genes and thus produce different mixtures of proteins.

Given these molecules' importance, one would think biologists would have long figured out the basic picture of what they look

like and how they work. Yet for decades scientists embraced a picture that was incomplete. They understood, quite properly, that proteins consist of amino acids linked together like beads on a string. But they were convinced that for a protein to function correctly, its amino acid chain first had to fold into a precise, rigid configuration. Now, however, it is becoming clear that a host of proteins carry out their biological tasks without ever completely folding; others fold only as needed. In fact, perhaps as many as one third of all human proteins are “intrinsically disordered,” having at least some unfolded, or disordered, parts.



A. Keith Dunker is a biophysicist at the Indiana University School of Medicine, where he directs the Center for Computational Biology and Bioinformatics. He studied viruses for 30 years before immersing himself in the study of disordered proteins in 1995.



Richard W. Kriwacki is a structural biologist at St. Jude Children's Research Hospital in Memphis. In 1996, while at the Scripps Research Institute in La Jolla, Calif., he co-discovered one of the first examples of a truly unstructured protein.



Folding only as needed,
the flexible protein p27 (*green*)
can wrap around many different
partners, something that proteins with
a unique 3-D structure cannot do.

To be sure, biologists have known for a while that enzymes such as the polymerases that copy DNA or transcribe it into RNA are complicated nanomachines consisting of many moving parts, with hinges that allow different segments of a protein to pivot around one another. But those proteins are often pictured as combinations of rigid parts, like the sections of a folding chair. Intrinsically disordered proteins look more like partially cooked spaghetti constantly jiggling in a pot of boiling water.

Fifteen years ago this assertion would have seemed downright heretical. Today scientists are realizing that such amorphous and flexible features probably helped life on earth get started and that their flexibility continues to play critical roles in cells, for instance, during cell division and gene activation. And this new understanding offers more than startling new insights into the basic biology of cells. Equally exciting, it hints at new ways for treating disease, including cancer.

PERFECT MATCHES

THE NOTION that a rigid, three-dimensional structure determines a protein's function first emerged in 1894. Emil Fischer, a chemist at the University of Berlin, proposed that enzymes—the catalysts of biochemical reactions—interact with other molecules by binding to specific shapes on their outer surface; at the same time, enzymes would completely ignore any molecules whose surface features are only slightly different. In other words, an enzyme and its binding partner fit together like a key and a lock.

At the time Fischer formulated his model, the nature of proteins was unknown. Over the next 60 or so years biologists learned that proteins were chains of amino acids and concluded that they had to fold into a precise shape to work properly. In 1931 Chinese biochemist Hsien Wu lent strong support to that view, showing that protein denaturation, or loss of natural 3-D structure, led to a complete loss of function. Since then, starting with the 3-D structure of sperm whale myoglobin in 1958, researchers have determined the architecture of more than 50,000 types of protein, usually by first coaxing their rigid structure into forming crystals and then scattering x-rays off those crystals.

Not all was static in this structured, lock-and-key protein world, though. As far back as the early 1900s, scientists knew that many antibodies can bind to multiple targets, or antigens—an observation that did not fit neatly with the lock-and-key model. In the 1940s the great chemist Linus Pauling speculated that certain antibodies can fold up in any of several ways, with the folding of each configuration guided by the fit between antibody and antigen.

From about the 1940s on, various other observations indicated that not all proteins abided by the dogma that function follows from a rigid, 3-D structure. But those that did not were usually regarded as isolated, freak exceptions to the rule. One of us (Dunker) was among the first researchers to collect such examples and to note that perhaps the dogma itself needed revision. In 1953, for instance, scientists noticed that the milk protein casein is largely unstructured; this pliability probably facilitates its digestion by infant mammals. In the early 1970s a protein called fibrinogen was found to contain a region of significant size having no fixed structure; this region, along with similar but smaller ones discovered later, plays a key role in blood clotting. Later in the 1970s, the protein that forms the outer casing, or capsid, of the tobacco mosaic virus offered another striking example. When the capsid is empty, the protein has large, unstructured regions hanging loose inside the capsid's cavity; that looseness enables

newly minted RNA, made during viral reproduction in an infected cell, to pack inside. But as the RNA gets in, the protein binds to it and sets into a rigid shape.

Meanwhile experimenters who could not induce certain proteins to fold in their test tubes assumed they were doing something wrong: surely the amino acid chains would find a “correct” folded shape in the environment of the cell. For example, when researchers placed solutions containing isolated proteins into vials and scanned them with a nuclear magnetic resonance (NMR) spectrometer—a workhorse of protein studies—they would sometimes get blurry data, which they interpreted as indicating that the proteins had failed to fold.

But those data had a richer story to tell. NMR spectroscopy involves the application of powerful radio-frequency pulses to induce the atomic nuclei of particular elements, such as hydrogen, to spin in sync. Slight frequency shifts in the nuclei's response correlate tightly to the atoms' positions inside amino acids and to the positions of those amino acids with respect to one another. Thus, from these frequency shifts investigators can often piece together the structure of a rigid protein. But if the amino acids move a lot—as would be the case in an unfolded protein—the frequency shifts become blurry.

In 1996 one of us (Kriwacki, then at the Scripps Research Institute) was performing NMR spectroscopy on a protein called p21, involved in controlling cell division, when he noticed something shocking. According to his NMR data, p21 was almost entirely disordered. The amino acids freely rotated about the chemical bonds that held them together, never staying in one conformation for more than a fraction of a second. And yet—and this was the shocking part—p21 was still able to perform its critical regulatory function. It was the first convincing demonstration that lack of structure does not make a protein useless.

NMR spectroscopy remains the primary technique to determine whether a protein is folded or disordered, and together with other technologies it has now confirmed that many proteins are intrinsically disordered. These molecules constantly morph under the action of Brownian motion and their own thermal jitters, and yet they are perfectly functional.

PROTEIN SPAGHETTI

THIS NEW, BROADER VIEW is well illustrated by the protein p27, which is known to exist in most vertebrates. Like p21, p27 is one of the crucial proteins that regulate cell division so that cells do not multiply uncontrollably. NMR shows that p27 is highly flexible, with sections that rapidly fold and unfold into short-lived corkscrew- or sheet-shaped structures. Most cancer cells in humans have reduced amounts of p27, and the greater the loss, the poorer the prognosis for a patient's survival.

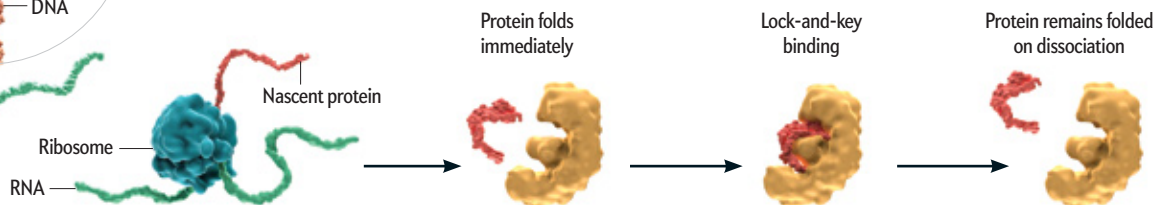
The p27 molecule acts as a brake on cell division by binding to and inhibiting the activities of at least six different types of kinase enzymes. Kinases are the master regulators of DNA replication and cell division. They attach phosphate (PO_4) to other proteins (“phosphorylate” them), a move that sets off a cascade of events. In carrying out its task, the stringlike, dynamic p27 molecule wraps around a kinase—which has a mostly rigid structure—and covers a significant portion of its surface, including its chemically reactive, or “active,” sites [*see illustration on preceding pages*]. This blockage prevents phosphorylation and so arrests cell division. Thanks to its flexibility, then, p27 can mold itself around, and inhibit, different types of enzymes. Proteins with

Order vs. Disorder

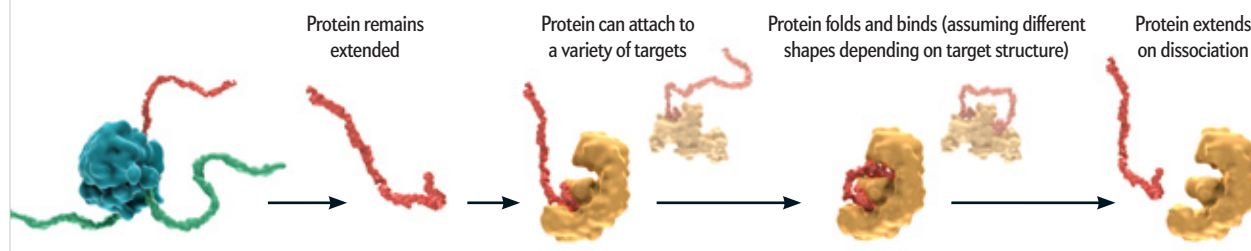
The molecular machinery of cells transcribes information encoded in DNA sequences—the genes—into RNA and translates the RNA into the long chains of amino acids that compose proteins. Biology textbooks say a protein must then fold into a unique

shape (*top row*) to do its job properly, for instance, binding to a specific molecule like a key fits into a lock. Many proteins, however, stay at least partly unfolded. This flexibility enables them to bind to various molecules (*bottom row*) or to perform other tasks [see box on next page].

Classic View of Protein Activity



Newly Uncovered Process



such an ability are described as promiscuous or moonlighting.

The p27 protein, being almost completely unstructured, falls near the disordered end of a scale that ranges from complete disorder (totally unstructured) to complete order (totally rigidly folded). The kinases themselves fall near the opposite end of this scale. Many other proteins lie somewhere in between, having both structured and unstructured regions. Calcineurin, which is involved in immune responses (and is the target of antirejection drugs), is the reverse of a kinase: it removes phosphates from particular proteins that have been phosphorylated. It has a structured region that is the enzyme's active site and operates in the classic lock-and-key-manner to remove phosphates from other proteins. But it also has an unstructured region that binds to and inactivates the enzyme's own active site when phosphate removal is not needed. Thus, calcineurin is like two proteins in one: the structured region performs catalysis, and the unstructured region regulates this catalytic function.

The examples we have discussed so far are proteins that fold—either on themselves or around other proteins—when they perform their function. But disorder is often part of a protein's working gear. In one known example, the length of an unstructured region acts as a timing device, controlling how fast two binding sites come together: if the unstructured region is longer, the two binding sites spend more time searching for each other than when the unstructured region is shorter. In another instance, being unstructured enables a particular protein to thread through a narrow opening and cross the cell membrane. And unstructured proteins occur in the axons of nerve cells, where they form brushlike structures that prevent the axons from collapsing.

Unexpectedly, some proteins remain unstructured even after

binding. At the Hospital for Sick Children in Toronto, Tanja Mittag (now in the faculty of Kriwacki's department) recently discovered an inhibiting protein in yeast, called Sic1, that stays attached to its partner through several small segments that continuously hop on and off a single binding site, while the rest of Sic1 remains disordered.

Disorder also exists in the proteins of simpler organisms and even of viruses. Some viruses known as phages, which specialize in infecting bacteria, attach to a host's membrane via proteins that connect to the main body of the phage through flexible linkers. The attachment protein, which is smaller and faster-moving than the entire phage, can then rapidly reorient to optimize its alignment during docking.

WIDESPREAD PROMISCUITY

TO DATE, ROUGHLY 600 partially or totally unstructured proteins have been directly identified and their functions understood by researchers at laboratories around the world. But we suspect many more exist. After all, scientists have so far learned the structure of just a small fraction of the estimated 100,000 or so proteins that exist in the human body alone. Also, new "bioinformatics" studies by Dunker and his collaborators point in that direction.

The bioinformatics approach builds on earlier theoretical studies of individual proteins, which suggested that after a cell synthesizes a chain of amino acids to make a protein, the chain folds in a way that depends on its composition. In particular, the amino acids that are bulky and hydrophobic—meaning they "dislike" the water molecules that naturally surround proteins—tend to end up in the interior. In contrast, the ones that end up on the surface of a given folded protein are generally small and hydro-

The Living Cell's Flexible Workers

In their roles as enzymes, structural components, molecular machines, and so on, proteins drive virtually everything a cell does. Here a view inside a human cell shows three important examples of proteins in which the lack of a rigid, predetermined structure is crucial to the proteins' functions.

Monorail Transporter

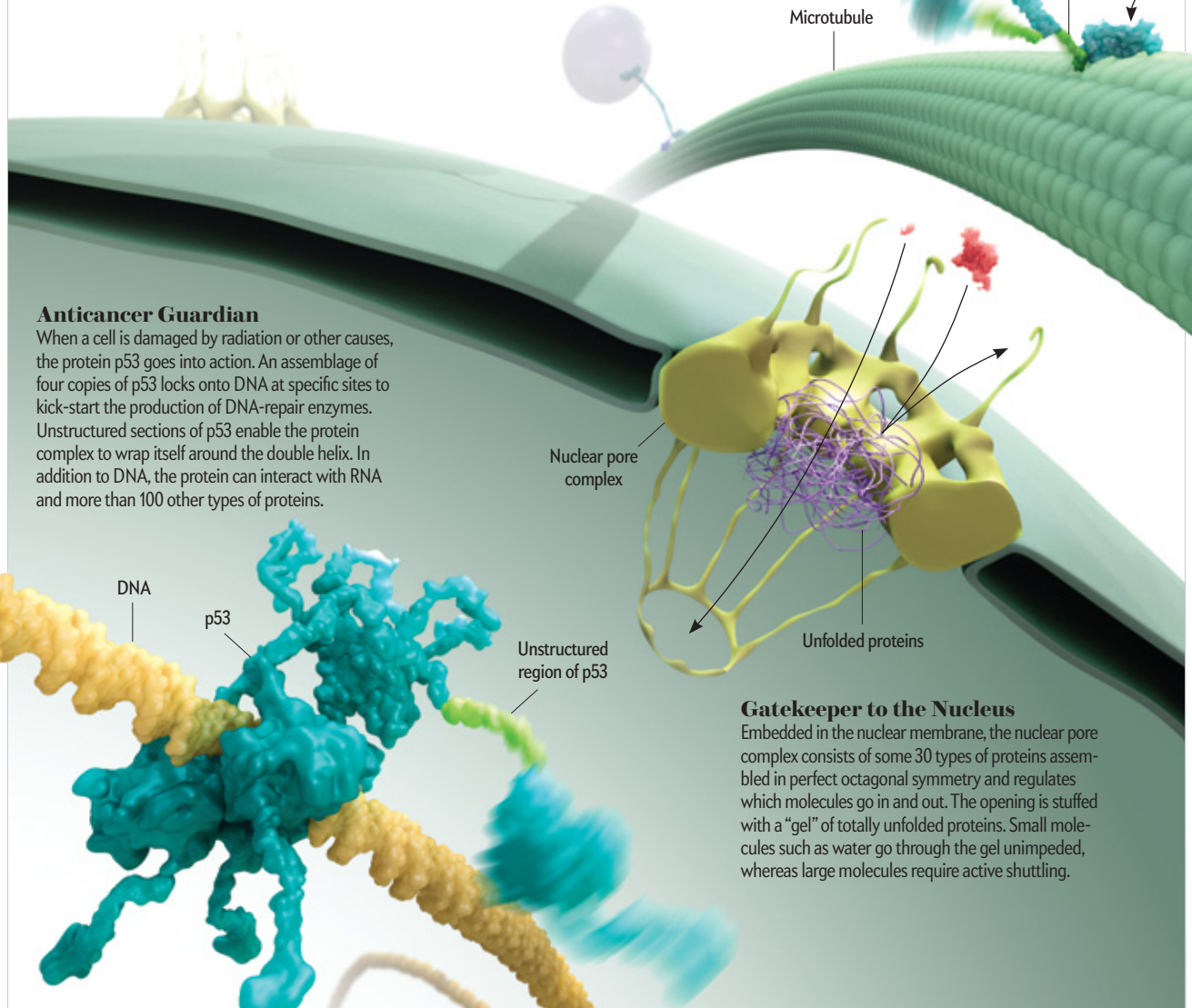
A molecular machine made of two copies of the protein kinesin "walks" along microtubules to drag a vesicle or other cargo from one part of a cell to another. Each step is powered by an ATP molecule, which reacts with the front "foot" and forces an unstructured "linker" section of the protein to fold onto that foot. The linker, meanwhile, also pulls the back foot, forcing it to twist forward, where it reattaches to the microtubule.

Anticancer Guardian

When a cell is damaged by radiation or other causes, the protein p53 goes into action. An assemblage of four copies of p53 locks onto DNA at specific sites to kick-start the production of DNA-repair enzymes. Unstructured sections of p53 enable the protein complex to wrap itself around the double helix. In addition to DNA, the protein can interact with RNA and more than 100 other types of proteins.

Gatekeeper to the Nucleus

Embedded in the nuclear membrane, the nuclear pore complex consists of some 30 types of proteins assembled in perfect octagonal symmetry and regulates which molecules go in and out. The opening is stuffed with a "gel" of totally unfolded proteins. Small molecules such as water go through the gel unimpeded, whereas large molecules require active shuttling.



philic—they tend to stick to the surrounding water molecules.

Dunker's idea was to compare the amino acid sequences of proteins known to be intrinsically disordered with those of proteins known to have rigidly folded shapes. What his team found in 1997, using computer algorithms, was that intrinsically disordered proteins tend to be richer in hydrophilic amino acids when compared with rigid proteins. Thus, the balance of hydrophilic and hydrophobic amino acids could predict whether a given protein would fold only partially or not at all.

To explore the biological implications of its earlier findings, in 2000 Dunker's team made a comparison across the kingdoms of life. The researchers examined the genomes of various organisms with algorithms that looked for stretches of DNA coding for long chains of hydrophilic amino acids. The corresponding proteins would be top candidates to be at least partially unstructured. In the simplest organisms, bacteria and archaea, fewer proteins were predicted to be intrinsically disordered. But in eukaryotes—the more complex organisms such as yeast, fruit flies

SOURCE: "THE TUMOR SUPPRESSOR p53: FROM STRUCTURES TO DRUG DISCOVERY," BY ANDREAS C. JOEBGER AND ALAN R. FISHT, IN *COLD SPRING HARBOR PERSPECTIVES IN BIOLOGY*, VOL. 3, NO. 2, FEBRUARY 2011

and humans, which have nucleated cells—unstructured proteins seem to be much more prevalent.

These results were extended in 2004 by a team led by David T. Jones of University College London, who used similar comparisons that included human data. Strikingly, the investigators found that as many as 35 percent of all human proteins may have long unstructured regions. Thus, about one third of our proteins may have large regions for which the lock-and-key concept is simply irrelevant.

The reasons for this discrepancy are unclear, but a possible explanation is that proteins with lock-and-key structural features are optimized for functions such as enzymatic activity, whereas intrinsically disordered proteins are best at signaling and regulation. Simple bacteria have everything in one container; complex organisms have multiple intracellular containers such as the nucleus, the Golgi, the mitochondria, and so on and thus need more signaling among their various parts and require more extensive regulation. Multicellular organisms also require signaling schemes to coordinate actions among various cells and tissues. In the example of p27 discussed earlier, thanks to its flexibility the protein can carry chemical messages along a cell's signaling pathways: the messages are encoded in its conformation, in its chemical modifications such as phosphorylation, and in the partners it binds to (and thus inhibits or regulates).

EVOLUTION'S BEST-KEPT SECRET

THE DEARTH of intrinsically disordered proteins in bacteria might seem to imply that these proteins arose only late in evolution. Several lines of investigation, however, suggest that they arose early. For one thing, numerous important bacterial signaling systems do use unstructured rather than structured proteins. Furthermore, in evolutionarily ancient molecular machines that are made of RNA and proteins assembled together, nearly all the proteins are partially or entirely unstructured when not bound to their RNA partners. These ancient hybrid complexes include the spliceosome (a molecular machine that edits, or splices, RNA as a step toward producing proteins) and the ribosome (the complex that strings amino acids together into proteins).

Research into the origin of life also hints at the antiquity of unstructured proteins. A leading hypothesis is that the first organisms were based on RNA. The RNA acted both as a catalytic molecule and as a repository of genetic information—the roles that in modern cells are played by proteins and DNA, respectively. One significant problem with this “RNA world” theory is that RNA folds very inefficiently into its catalytically active form and often gets stuck in inactive conformations. In today's cells proteins called RNA chaperones help the RNA fold correctly. Other proteins stabilize a given RNA in its active conformation, raising the possibility that the advent of such proteins solved the sticking problem of RNA folding. Both the chaperone and the stabilizer proteins lack stable structure before binding to RNA.

Yet more support for the early evolution of unstructured proteins comes from analyzing the origin of the genetic code. The genetic code is the set of rules cells use to translate the information stored in nucleic acids (RNA or DNA) into an amino acid sequence. Researchers believe that certain amino acids were encoded early in the evolution of life, whereas others came later. The bulky, hydrophobic amino acids that drive a protein to fold likely came late, so proteins made from just the early amino ac-

ids would very likely remain unfolded if left alone. If these ideas on the evolution of the genetic code are correct, then the first proteins on earth folded poorly or not at all. The amino acids that arose later evidently enabled proteins to form structure, providing the basis for the formation of lock-and-key enzyme active sites and enabling proteins, over millions of years, to replace RNA as the catalytic powerhouse in all living cells.

DOUBLE-EDGED SWORD

GIVEN HOW CENTRAL proteins are to biology, it should be no surprise that many of them are involved in disease. The new paradigm of intrinsic disorder in proteins will thus profoundly affect how we understand and treat human illnesses.

For starters, in some cases a protein's lack of structure may be harmful: if a cell produces them in excess, certain unstructured proteins are prone to jumble up and form plaques. In the brain, such plaques are major suspects in several devastating neurodegenerative diseases, including Alzheimer's, Parkinson's and Huntington's. More generally, it seems unstructured proteins need to be kept scrupulously in check to avoid trouble: a large-scale study of yeast, mice and humans led by M. Madan Babu of the Medical Research Council's Laboratory of Molecular Biology in Cambridge, England, showed in 2008 that cells regulate disordered proteins more tightly compared with folded proteins.

The realization that intrinsically disordered proteins can be involved in certain diseases is also leading to new ideas for potential treatments. Protein-protein interactions underlie virtually every biological process and thus have long been attractive targets for drug discovery, but with little success so far, compared with the approach that targets interactions of enzymes with smaller molecules. Proteins that interact with unstructured proteins often offer their partners anchoring nooks, which researchers might exploit to insert new drugs. In particular, molecules that block an interaction between an important gene for suppressing cancer and one of its regulatory partners have shown success at fighting cancer in lab animals and are now undergoing clinical trials in humans. Kriwacki, with his colleagues, is developing a similar line of attack to treat retinoblastoma, a cancer of the eye that especially affects children. Early tests in animals have given promising results. Other labs are working on similar projects.

Scientists interested in understanding how proteins work are beginning to dispel past biases represented by the lock-and-key model of protein function. They are recognizing that some biological functions are best performed by rigid proteins and others by highly dynamic ones. The dawning of a new era in protein structure and function has the potential to transform our understanding of life—and perhaps to save a life. ■

VIDEOS OF
FLEXIBLE PROTEINS
[ScientificAmerican.com/
apr2011/flexible-proteins](http://ScientificAmerican.com/apr2011/flexible-proteins)

MORE TO EXPLORE

Structural Studies of p21^{Waf1/Cip1/5dfl} in the Free and Cdk2-Bound State: Conformational Disorder Mediates Binding Diversity. Richard W. Kriwacki et al. in *Proceedings of the National Academy of Sciences USA*, Vol. 93, No. 21, pages 11504–11509; October 15, 1996.

Identification and Functions of Usefully Disordered Proteins. A. Keith Dunker et al. in *Advances in Protein Chemistry*, Vol. 62, pages 25–49; 2002.

Intrinsically Unstructured Proteins and Their Functions. H. Jane Dyson and Peter E. Wright in *Nature Reviews Molecular Cell Biology*, Vol. 6, pages 197–208; March 2005.

Structural Disorder Throws New Light on Moonlighting. Peter Tompa et al. in *Trends in Biochemical Sciences*, Vol. 30, No. 9, pages 484–489; September 1, 2005.

Database of Protein Disorder. Known unstructured proteins are searchable online at www.disprot.org. Protein Data Bank (www.rcsb.org/pdb) is a database of structured proteins.