

# BREAKING THE PROTEIN RULES

*If dogma dictates that proteins need a structure to function, then why do so many of them live in a state of disorder?*

BY TANGUY CHOUARD

**K**eith Dunker's life is a mess. His desk is so swamped with books, old chocolate bars, half-reviewed manuscripts, pens, diet coke bottles and — somewhere — a stray sock, that he ends up printing papers again rather than wading in to find the original. "I'm so disorganized," he crows, "some people have called me Dr Disorder." But he remembers with great precision the moment that disorder invaded his scientific life. It was 15 November 1995, at 12:40 p.m., halfway through a seminar by crystallographer Chuck Kissinger, at Washington State University in Pullman, where Dunker was then a biochemist. Dunker was staring at a slide showing the atomic structure of calcineurin, an enzyme targeted by immunosuppressive drugs. What caught his attention wasn't the intricate structure but something missing from it: a dotted line representing a string of amino acids with a position too variable to be determined by X-ray crystallography, as the rest of the protein had been. And Kissinger was insisting that this loop had to remain flexible for calcineurin to serve its crucial function in the human immune system.

"It hit me like a brick," says Dunker: this wayward piece of protein flouted a century of dogma. A central tenet in molecular biology is that the function of a protein depends critically on its fixed three-dimensional structure; for example, enzymes bind to specific substrates because their shapes match perfectly, as immortalized in the 'lock-and-key' model proposed by chemist Emil Fischer as early as 1894. But this part of calcineurin seemed to disobey these rules, by providing function without structure. Now Dunker was wondering how many other proteins were ignoring the rules too.

To find out, he and his colleagues wrote a bioinformatics program that predicted which protein segments are 'intrinsically disordered' — meaning that they do not fold spontaneously into a unique three-dimensional shape. Today, this and other similar programs predict that about 40% of all human proteins contain at least one intrinsically disordered segment of 30 amino acids or more, and that some 25% are likely to be

disordered from beginning to end<sup>1</sup>. This part of the protein universe had largely been ignored because disordered protein segments impede crystal formation — a prerequisite for X-ray diffraction, the predominant way structures are deduced — and structural biologists clip them out whenever they can.

Today, though, "the recognition of disorder has grown dramatically", Peter Wright, a protein biophysicist at the Scripps Research Institute in La Jolla, California, told the American Association for the Advancement of Science meeting in Washington DC last month. A large part of that recognition has come from studies using nuclear magnetic resonance (NMR) spectroscopy, which allows researchers to determine the structures of small proteins even as they twist and turn in solution. Such work has shown that disorder can actually be essential to function by helping a signalling protein to recognize and react to a protein partner, or by allowing a regulatory protein to interact with multiple targets. Still, says Wright, "that hasn't got through to the textbooks".

Many structural biologists see no need for revision. "My mantra has been: function requires structure," declares Tom Steitz, a crystallographer at Yale University in New Haven, Connecticut. "Some flexibility can be required, it may be an essential part of the assembly process, but it's not interesting until the proteins get to do their job." Critics argue that the computer programs predicting high levels of disorder are fundamentally flawed because they identify proteins that are well-known to become perfectly ordered — and to crystallize — when bound to their proper molecular partners. They say that unfolded protein chains cannot persist for long in living cells, and some want the concept of intrinsically disordered proteins to be ditched altogether. That seems unlikely. Data are fast accumulating from all fronts — biophysics, bioinformatics and cell biology — in support of widespread disorder, and disorder aficionados are calling for a complete reassessment of the structure–function paradigm. "Biology uses disorder to bring about its various functions," Wright says.

Since the late 1950s, newly made proteins have been assumed to fold up immediately and spontaneously into a unique three-dimensional shape — their most energetically stable conformation and the only functional

**➔ NATURE.COM**  
Read about protein folding that can cause disease:  
[go.nature.com/s6w5mm](http://go.nature.com/s6w5mm)

one<sup>2</sup>. The few proteins known to remain unfolded “were pointed to as oddities”, Wright says. But that started to change in 1999, when Wright and fellow NMR spectroscopist Jane Dyson, also at Scripps, wrote a review<sup>3</sup> pointing to the growing collection of proteins that seemed to function despite their disordered state. It has been “the big-dog paper in the field”, says Dunker, now at Indiana University School of Medicine in Indianapolis.

One burning question, then and now, is how a protein can function if it has no fixed shape. “We all accept flexibility,” says structural biologist Joël Janin at the CNRS Laboratory of Enzymology and Structural Biochemistry at Gif-sur-Yvette, France. “The question is: how can you get recognition with flexibility?” The whole concept of disorder seems incompatible with the lock-and-key model. You might as well try to open the door with cooked spaghetti.

In 2007, postdoc Kenji Sugase in Wright’s lab found an answer: the spaghetti uses the lock to mould itself into the shape of the key, rather than forming the key beforehand. Sugase focused on CREB, a gene-regulatory protein involved in many processes including learning and memory. Once bound to DNA, CREB also needs to recognize and bind a protein partner called CBP before it can switch on a gene. But the part

of CREB involved with CBP enters the game in a disordered state. How could a thing like this possibly work?

To find out, Sugase developed the equivalent of a super-fast NMR camera so that he could capture frequent snapshots of CREB’s wriggling chain, at atomic resolution, as it tested out points of contact within itself and with CBP. What he saw was that several bonds had to form cooperatively within CREB and with CBP for the whole complex to snap into shape<sup>4</sup>. That’s exactly how the average ‘globular’ protein folds: its internal segments need to establish long-range chemical bonds with one another, to pull the whole thing suddenly into shape<sup>2</sup>. CREB forms such interactions externally, by bonding to CBP — and if this bonding is just a little weaker, then the key cannot form and there is no binding with the lock at all. Wright and his colleagues think that disorder is therefore advantageous because it allows CREB to partner with CBP more exclusively than a rigid protein would. And Wright thinks that this type of process allows many signalling proteins to engage in fast-changing yet selective interactions.

## DISORDERLY CONDUCT

The structure–function mantra took an even bigger hit recently from a protein with parts that never seem to fold at all. The Sic1 signalling protein is a key regulator of the cell cycle that puts the brakes on DNA replication until the cell is ready to divide. In 2001, a team led by Mike Tyers, a yeast cell-cycle expert at the University of Toronto, Canada, began unpicking the mechanism of the switch. The group found that when phosphate groups are added to six sites on Sic1 it can then hook up with a second protein, Cdc4, which pushes Sic1 into the cell’s protein-disposal pathways<sup>5</sup>. Once Sic1 is degraded, DNA replication can forge ahead. But unless that degradation occurs at precisely the right time, DNA replication goes haywire and the cell may eventually die. The cell achieves that precision by ensuring that it takes exactly six phosphates to flip the switch, not four or five. But there’s a rub: Cdc4 has only one high-affinity binding pocket for a phosphate group. How can Cdc4 ‘count’ up to six with effectively only one finger to count on?

After all attempts to crystallize Sic1 failed, Tyers’s team called NMR spectroscopist Julie Forman-Kay, also at the University of Toronto. In 2008, Tanja Mittag, a postdoc in Forman-Kay’s lab, showed that Sic1 was disordered<sup>6</sup> — not only in its free state but, astoundingly, also when bound to Cdc4. The complex seemed to be a mixture of different conformations shifting around in constant, dynamic equilibrium. And the most stunning part was that each of the six phosphate groups on Sic1 could be found to occupy the single Cdc4 pocket, one after the other, as in a constant dance around the fire (see ‘Orders of disorder’).

The researchers then developed a computer model and fed it with every scrap of experimental data about the proteins’ structures that they could gather<sup>7</sup>. They concluded that, even though Sic1 is disordered when it is bound, it maintains a rather compact structure, which keeps all the phosphate groups sufficiently close together to form an average electrostatic field that glues Sic1 to Cdc4. Only when six phosphates are present is the glue strong enough for Cdc4 to hold Sic1 close and force-feed it into the cell’s disposal machinery. And one reason that Sic1 has to be disordered during all this, the team proposes, is to enable the rather rigid disposal machinery to reach all parts of Sic1 and carpet it with the chemical tags that mark the protein for destruction. It takes a nimble protein indeed to make all these connections at once.

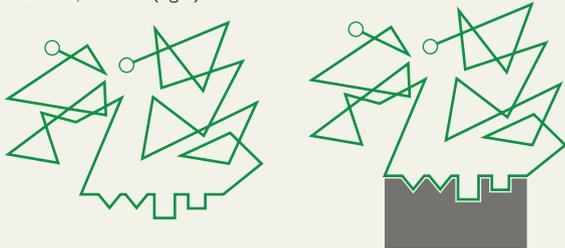
“The result is interesting,” says structural biologist Stephen Harrison of Harvard Medical School in Boston, “because interaction motifs are often found more or less repeated along unstructured segments, and the work shows how such multiplicity can function.” And Forman-Kay thinks that adopting ‘multi-structural’ states could allow other proteins to constantly probe and sense signals from many partners at once. This is particularly important for ‘hub’ proteins, which are central to vast networks of rapidly changing molecular interactions. “There is a complexity people haven’t talked about,” she says. “These hub proteins need to very rapidly sample the complex cellular environment.”

## ORDERS OF DISORDER

A central tenet in molecular biology states that the function of a protein depends on its fully folded three-dimensional structure. In the new view, protein segments can function when transiently or durably disordered.

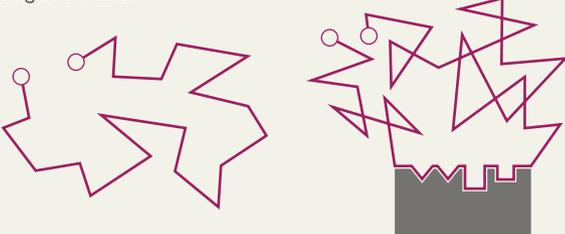
### LOCK AND KEY

In the conventional view, an enzyme folds up immediately into a unique and stable 3D shape, the key (left). Its shape perfectly matches and allows it to bind its substrate, the lock (right).



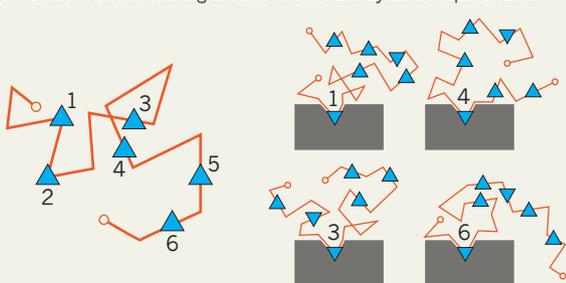
### FOLD AS YOU BIND

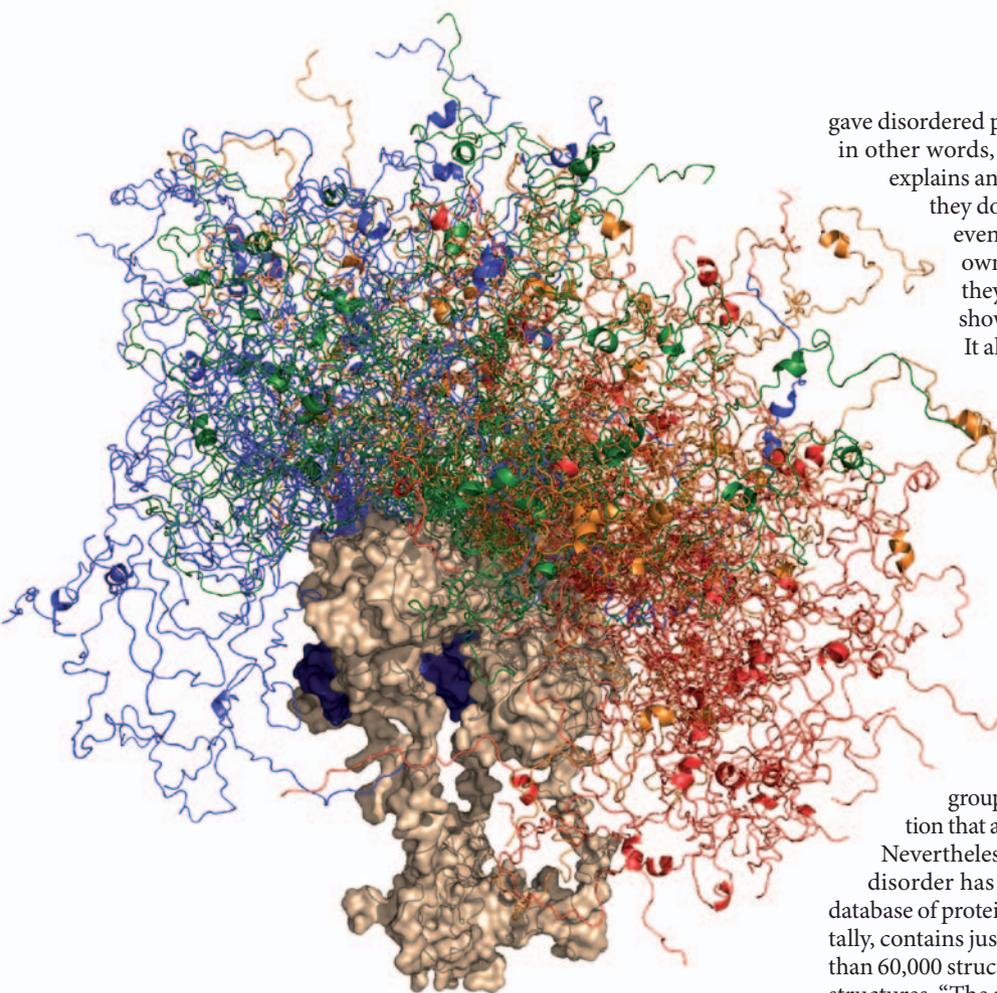
A disordered part of the gene-regulatory protein CREB (left) uses the lock to mould itself into the shape of the key when the two meet (right), rather than folding beforehand.



### SHAPE SHIFTING

The signalling protein Sic1 remains disordered in its bound state, and each of six phosphate groups occupies the binding site in turn. The protein is a mix of different conformations shifting around in constant dynamic equilibrium.





Tumour-suppressor protein p53 contains an ordered, globular domain (brown); its disordered segments (colours) help it interact with hundreds of partners.

One extreme example can be seen in the tumour suppressor p53, an extraordinarily well connected hub in multiple signalling networks, and the protein most frequently implicated in human cancer. Part of the explanation for p53's promiscuity seems to lie in its versatile structure, which features every possible conformation from order to disorder. The core domain is globular and binds to DNA and just a few other proteins; its two flanking wings are mostly disordered and can bind to hundreds of signalling partners; and a segment within one wing shows a 'chameleon' status which can flip between four different ordered states, depending on which partner it binds to<sup>8</sup>. Alan Fersht, a biophysicist and NMR expert at the University of Cambridge, UK, says that he is "absolutely sure" that long parts of p53 remain largely disordered in the cell: "I don't think there's any doubt about that whatsoever."

### PREDICTION PROBLEMS

Yet many researchers question how widespread disordered proteins can be. That is mainly because biochemists, over more than 100 years of preparing tissue extracts, have struggled to prevent proteins from unfolding and tangling into insoluble clumps or getting digested by enzymes called proteases. "It's hard to believe that disordered proteins could produce anything else than a mess," says Janin.

Researchers who study these processes, however, say that such fears are largely irrelevant. In human cells, for example, nonspecific proteases are locked away in compartments called lysosomes, which should allow disordered proteins to survive everywhere else, explains Ulrich Hartl, an expert on protein quality control at the Max Planck Institute in Martinsried, Germany. Disordered proteins should also be protected from aggregation because, unlike globular proteins, they contain few hydrophobic amino acids, which tend to stick together — and are instead rich in 'polar' amino acids that are happy swimming in water.

Hartl thinks that natural selection against aggregation probably

gave disordered proteins this particular amino-acid composition — in other words, it is not a signature for disorder per se. And this explains an apparent inconsistency of disorder predictors: that they do not pick internal segments from globular proteins, even though these too are incapable of folding on their own when sliced out of a protein<sup>2</sup>. The reason is that they are rich in hydrophobic amino acids and so do not show the signature sequence that the predictors detect.

It also explains why the predictors select some proteins that are in fact ordered, another point of controversy surrounding these programs: because proteins can lack hydrophobic amino acids for many different reasons. "This makes perfectly good sense," agrees Dunker.

Overall, the numerous programs claim an 80% success rate at predicting whether any individual amino acid in a protein will be surrounded by order or disorder<sup>1</sup>, as compared to the 50% success rate expected by chance — and crystallographers rely on them to sideline proteins expected to resist crystallization. "Disorder predictors are a massive oversimplification, but they are very useful," says Adam Godzik, at the Burnham Institute in La Jolla, California, and head of the bioinformatics group for the Protein Structure Initiative, a large collaboration that aims to solve large numbers of protein structures.

Nevertheless, debate about the prevalence and importance of disorder has probably slowed progress in the field. DisProt, a database of proteins whose disorder has been established experimentally, contains just over 500 proteins, a number dwarfed by the more than 60,000 structures in the Protein Data Bank<sup>1</sup>, the database for 3D structures. "The main reason why DisProt is so small," says Dunker, "is because it has taken so many damn years to get it funded." But in the past few years, major consortia aimed at exploring intrinsically disordered proteins have been set up in several countries. Interest in disordered proteins as drug targets is also on the rise because so many of them, like p53, are crucially implicated in disease.

Little by little, a fundamentally new picture of the relationships between protein sequence, structure and function is emerging: a continuum running from the most rigid 'lock-and-key' enzymes and molecular machines at one extreme through to durably unstructured spaghetti such as Sic1 at the other, and spanning all degrees of structural ambiguity in between. Figuring out how all these disordered proteins really work is a long way off, if Sic1 is anything to go by: determining its mode of action involved several, often arcane, biophysical techniques, new computer tools and statistical physics theory — plus at least ten years of work by six labs. Multi-structural biology isn't going to be simple.

Still, Dunker, Wright and other doctors of disorder are optimistic. As is Martin Blackledge, an NMR spectroscopist at the Institute of Structural Biology in Grenoble, France, who compares the excitement now to that surrounding the first crystal protein structures in the 1950s. "Every new case is fascinating at the moment," he says. Blackledge looks forward to a day when it might be possible to predict where on the structural continuum a protein segment falls, from its amino-acid sequence — to crack the full code of disorder. "This is exactly what I'm aiming for," he says, "this is my dream."

Perhaps the rules of disorder are needed, before disorder can rule. ■

**Tanguy Chouard** is an editor for *Nature* in London.

1. Uversky, V. N. & Dunker, A. K. *Biochim. Biophys. Acta* **1804**, 1231–1264 (2010).
2. Anfinsen, C. B. *Science* **181**, 223–230 (1973).
3. Wright, P. E. & Dyson, H. J. *J. Mol. Biol.* **293**, 321–331 (1999).
4. Sugase, K., Dyson, H. J. & Wright P. E. *Nature* **447**, 1021–1025 (2007).
5. Nash, P. et al. *Nature* **414**, 514–521 (2001).
6. Mittag, T. et al. *Proc. Natl Acad. Sci. USA* **105**, 17772–17777 (2008).
7. Mittag, T. et al. *Structure* **18**, 494–506 (2010).
8. Oldfield, C. J. et al. *BMC Genomics* **9** (Suppl. 1), S1 (2008).