



STACEY RAVEN PHOTOGRAPHY

# Anarchy in the proteome

Just 15 years ago, the idea that proteins might be functional without a well-ordered 3D structure was heretical. But as Michael Gross discovers, a little flexibility can go a long way

The golden or classical era of molecular biology, ranging from the 1950s when the structure of DNA was solved, through to the early 1980s, when molecular biology practically replaced traditional biochemistry, had a very simple world view, epitomised in the so-called central dogma. According to this dogma, biological information flows only in one direction. The gene sequence determines the messenger RNA sequence, which determines the amino acid sequence of the protein, which defines the 3D structure of the protein, which defines its function.

In the last three decades, this world view has cracked in a number of places. The discovery that retroviruses such as HIV make DNA based on RNA templates was in blatant violation of the dogma. The discovery that RNA can itself have catalytic function didn't fit very well into that view either. And now the tail end of the dogmatic chain of command, namely 'structure determines function', must also be taken with a pinch of salt.

#### Necessary flexibility

The trouble started when NMR began to establish itself as an alternative way to solve protein structures. Crystallographers had developed a tradition of ignoring all parts of the molecule that were too poorly ordered to show up in their electron density map as irrelevant. For them it was as if disorder simply didn't exist, by definition. NMR studies of protein structures, by contrast, led to a more dynamic view of the polypeptide, obtained in solution. Not only could one see the less-than-perfectly-ordered parts of a protein chain, but one could also quantify the degree of flexibility.

Since the early 1990s, more and more NMR studies have reported that large parts of some proteins appeared to have no firm structural framework, even under conditions where the protein was known to be functional. As this violated both the central dogma of molecular biology and the world view of structural biology shaped by x-ray crystallography, people tried to make the problem go away. Maybe the conditions used in those NMR experiments weren't quite physiological enough. Maybe under the right set of conditions the protein would adopt a well-defined solid structure like the ones seen in crystal structures.

A very clear case proving these objections wrong emerged in the signalling protein FlgM found in bacteria that have flagella, such as *Salmonella typhimurium*, which causes food poisoning. The bacterial flagellum is a hollow tube, and while it is under construction it remains open to the outside world. During assembly the bacterium constantly exports the signalling protein FlgM through this hollow channel. When building work on the new flagellum has finished, the end is sealed. FlgM can no longer be exported, and will accumulate in the cell, where it suppresses the genes responsible for making the building blocks of flagella.

Intriguingly, FlgM can only be exported in its unfolded state. The fully folded protein is simply too large to fit through the channel. NMR studies had shown that certain proteins were probably partially unstructured in their functional state, but this was an example of a protein that had to be unfolded to carry out a crucial part of its biological function.<sup>1</sup>

One might still argue, says Julie Forman-Kay from the University of Toronto, Canada, who has conducted NMR studies of disordered proteins since the mid-1990s, that 'being exported', although being a necessary biological function, doesn't involve much 'activity' from the protein itself. Examples of more proactive disordered proteins, says Forman-Kay, include 'elastin, which

#### In short

- The classical view of molecular biology is that a protein needs a defined 3D structure to function
- Disorder was ignored by crystallographers, but newer techniques show that dynamic, flexible proteins perform crucial roles in biology
- Disorder lets proteins squeeze into tight spots and be more adaptive in binding multiple targets. This has implications from signalling to cancer regulation

**James Watson's 'central dogma' of biological information flow from DNA to proteins, and the idea that structure determines protein function is being eroded**

requires disorder to impart elasticity in many tissues; clusterin, which acts as a detergent; and portions of the nuclear pore complex that form a regulated gate'.

#### New dynamics

Progress in the methods used to characterise dynamic macromolecules, including NMR spectroscopy and small angle x-ray diffraction, brought in a large number of new examples of proteins disordered under physiological conditions. Gradually, the rigid thinking of structural biologists trained in crystallography gave way to a new view that valued the dynamics of a polypeptide chain.

In addition to the Protein Data Bank with its over 74 000 well-ordered structures of folded proteins, there is now also a database of the more chaotic side of the world, known as disprot.org. Database founder Keith Dunker from Indiana University at Indianapolis, US, says: 'We found that the proteomes of eukaryotes contain a huge fraction of disordered residues. Our current estimates are that about 45 per cent of the amino acids [encoded] in the human genome are predicted to be intrinsically disordered.'

The large number of cellular proteins found to be disordered initially puzzled cell biologists. How can these be present in the cell, seeing that disordered states tend to aggregate into insoluble lumps and thus become unavailable to any biological function? And wouldn't these disordered protein chains be prone to degradation by proteolytic enzymes? The solution to this paradox emerged as their biological functions became better understood.

'What we have found by comprehensive bioinformatics of structured and disordered proteins,' says Dunker, 'is that structured proteins have basically four classes of functions: binding to small molecules, catalysis, membrane transport through structured pores, or structural stabilisation via fibrous assembly systems. Disordered proteins on the other hand are involved mainly in signalling, regulation and control – using their flexibility to carry out these functions.'

Such signalling proteins are crucial, but the cell only needs a very small number of each, and only at very specific times. During their short missions, they are unlikely to encounter other disordered proteins to form aggregates with. And





**Tail-like bacterial flagella rely on a disordered protein to control their construction**

degradation of signal proteins after the signal has expired is a necessary part of their life cycle.

### Induced fit

In some cases, proteins may end their period of disorder when they encounter their target molecule. Some of them literally fold up around their target protein, providing a particularly strong bond to it.

However, Forman-Kay emphasises that it would be wrong to assume that disordered proteins become static upon binding to partners. 'Many examples are showing up of dynamic complexes with multiple ways that the disordered protein can interact with the folded partner,' she says. Examples include observations in her own group<sup>2</sup> that multiple Sic1 motifs bind and release a single site on the partner protein Cdc4, but other groups have found such cases as well. Forman-Kay also cautions that NMR studies of isolated disordered sequences, excised from larger proteins, may not reveal the whole story, and have led people to think that all disordered proteins become static upon binding.

There is a shift, says Forman-Kay, towards the realisation 'that even bound complexes can be highly dynamic'. Researchers call such combinations of molecules 'fuzzy complexes' or 'dynamic complexes'.

Many disordered proteins are not completely disordered – they do not

form what protein researchers call a 'random coil'. Rather, they include a mixture of ordered and disordered functional parts. Traditionally, protein scientists speak of 'domains'. However, the use of this word may lead to philosophical arguments, as the original definition of a domain (as an independently folding unit) depends on the presence of a folded structure.

### Cancer chaos

Among the 'mixed' proteins are many of the human transcription factors, which regulate gene expression and are enormously important for cancer research. Many crystal structures of such proteins in the Protein Data Bank are missing large chunks of sequence, raising the suspicion that these parts of the molecules are intrinsically disordered.

One classic example is the transcription factor p53, notoriously found mutated in more than half of all human tumours. Of its four functional domains, three are highly disordered. Recently, the groups of Jane Dyson and Peter Wright at the Scripps Institute in La Jolla, California, solved the structure of one of these domains in a complex with a natural binding partner.<sup>3</sup> Forman-Kay comments: '[p53] is an interesting example of dynamic complexes, as the same small stretch of disordered protein can bind four

different targets, with the stretch stabilising a different conformation in each case. In the cell, there will be a dynamic equilibrium between these, depending on the presence of the other partners.'

Similarly, the disordered 'domains' of the oncogene product cMyc (which unlike p53 is activated, rather than inactivated in tumours), are mostly disordered in isolation. They can become more ordered when encountering certain binding partners, including a protein called Max. However, experts expect cMyc to remain dynamic overall, and to adjust its structure in different ways for different partners.

### Slippery customers?

So it turns out that transcription factors, much-heralded targets for future drugs, are disordered. Isn't that bad news for drug developers and for medicine more generally? Not necessarily, say Dunker and his colleague Vladimir Uversky. In a recent review, they have argued that the ability of disordered proteins to fold around a target opens new opportunities for small molecule drugs. However, these new drugs remain to be discovered. Failing that, one could still investigate imitating or blocking the well-ordered binding partners of the disordered transcription factor.<sup>4</sup>

A further medically relevant characteristic of disordered proteins is that they present unusual sites for molecular recognition. The function of these sites depends solely on the amino acid sequence (primary structure) of the polypeptide, and not as one would expect in the traditional world view of biochemistry, on 3D structures. As Norman Davey from the European Molecular Biology Laboratory in Heidelberg, Germany, points out, these elements are much more frequent than scientists used to think. By now, more than 150 types have been discovered.<sup>5</sup>

Viruses can mimic at least 50 of these to trick the signalling systems of their host cells. Among these tricksters are human papilloma virus, which causes cervical cancer, and Herpes simplex, which causes cold sores. In this area, too, an improved understanding of how a chain of amino acids can achieve biological function without ordered structure promises medical benefits.

Linear motifs also seem to play a role in the aggregation of  $\tau$  protein in the brains of Alzheimer's patients. They ensure that filaments of

**'Disordered proteins are involved mainly in signalling, regulation and control – using their flexibility to carry out these functions'**



different variants of the  $\tau$  protein, consisting of a  $\beta$ -sheet core surrounded by flexible brushes, can form equally spaced ribbon-like fibrils.<sup>6</sup> The entropy of the disordered chains surrounding the ordered filament core may play a crucial role in keeping the filaments at a distance without specific binding.

#### Rapid evolution


In evolutionary research, disordered proteins have also raised new questions. In folded proteins, most amino acid side chains are involved in interactions with their neighbours in space (which may

not be neighbours in sequence). Therefore, many mutations will severely disrupt the 3D structure and thus the function of a protein. In a disordered polymer chain, by contrast, a single mutation is less likely to cause a noticeable effect. This implies that disordered proteins are under less selection pressure and have therefore more freedom to evolve rapidly.

And yet, even disordered proteins show degrees of family resemblance, highly conserved motifs, and conservation of higher order biological functions.<sup>7</sup> Bioinformatics experts are already trying to use

**Disordered parts of proteins (orange) can mould themselves around targets to bind strongly**

these common features in order to develop methods of predicting disorder from the protein sequence. However, as David Karlin from the University of Oxford, UK's zoology department says, 'a big problem is that the existing algorithms for homology searches and multiple sequence alignments are entirely designed with folded, globular proteins in mind'. These algorithms enable scientists to compare proteins from a wide variety of species to study their evolution. 'In that respect, very little has changed in the last 10 years,' he says. 'People are only starting to tackle the problem.'

As the number of known examples of disordered proteins keeps increasing, the hope is that their relevance will also be recognised among bioinformaticians developing tools for evolutionary and comparative studies. The increasing numbers also promise to give researchers ample material to work out how evolution came up with this paradoxical phenomenon. 

*Michael Gross is a science writer based in Oxford, UK*

**Tumour suppressor p53 is highly disordered, but plays a key role in controlling cancer**

**'p53 is a dynamic complex, the same small stretch of disordered protein can bind four different targets'**

#### References:

- 1 K W Plaxco and M Gross, *Nature*, 1997, **386**, 657 (DOI: 10.1038/386657a0)
- 2 T Mittag *et al*, *Structure*, 2010, **18**, 494 (DOI: 10.1016/j.str.2010.01.020)
- 3 C W Lee *et al*, *Biochemistry*, 2010, **49**, 9964 (DOI: 10.1021/bi1012996)
- 4 A K Dunker and V N Uversky, *Curr. Op. Pharmacol.*, 2010, **10**, 782 (DOI: 10.1016/j.coph.2010.09.005)
- 5 N E Davey, R J Edwards and D C Shields, *Frontiers Biosci.*, 2010, **15**, 801 (DOI: 10.2741/3647)
- 6 S Wegmann *et al*, *J. Biol. Chem.*, 2010, **285**, 27302 (DOI: 10.1074/jbc.m110.145318)
- 7 P Tompa *et al*, *Bioessays*, 2009, **31**, 328 (DOI: 10.1002/bies.200800151)

