STRUCTURAL BIOLOGY

Protein structure gets exciting

Researchers determined the excited-state structure of a small protein using nuclear magnetic resonance spectroscopy.

A high-resolution protein structure solved by crystallography or nuclear magnetic resonance (NMR) spectroscopy can yield a wealth of information about that protein's biological function. These structures in most cases are static pictures of the native, or ground, state—the most stable state of the protein. However, proteins can also exist in higher-energy states, referred to as excited states, which can be crucial for carrying out biological functions.

Despite their importance, excited-state proteins pose a major experimental challenge. These states can be fleeting, lasting for only milliseconds. Only a very small percentage of proteins in a population exists in an excited state at any given time. Because many excited-state proteins cannot be isolated, solving their structures has been impossible using traditional methods. Lewis Kay of the University of Toronto sums up the conundrum: "How do you go about studying things that are essentially invisible to any of the experimental approaches that you have but that you think are going to be very important biologically?"

By putting more than a decade of methodological advances in NMR spectroscopy together, Kay's group recently solved the structure of an excited state of a small protein called the FF domain, corresponding to a folding intermediate (Korzhnev *et al.*, 2010).

Though transient excited-state structures cannot be directly observed, they do leave their mark on an NMR spectrum. They cause chemical shifts in NMR spectra—which provide information about the local environment of an NMR-active nucleus—to broaden. This is usually frustrating for researchers wanting to obtain a native state structure of a protein. But for this project, broadened chemical shifts were exactly what Kay wanted to see because it meant that the protein population was undergoing exchange between the native and excited state. "We like to get good spectra—but not too good, so that we can actually exploit the broadening," he explains.

To measure chemical shifts for the excited state of the FF domain, Kay's group applied NMR experiments known as relaxation dispersion. These experiments also yielded information about the lifetime and population of the folding intermediate, as well as another parameter called residual dipolar couplings, which provided information about the orientation of the backbone amide bond vectors. The researchers then fed these excited-state data into a powerful tool called CS-Rosetta, a program designed to solve small protein structures using only chemical shift data (Shen *et al.*, 2008).

With this strategy, Kay's group solved the structure of the folding intermediate of the FF domain, which has a lifetime of one millisecond and represents 3% of the protein population. To validate the structure, they made a 'trapped' version of the excited-state protein by truncating a portion of the C terminus to destabilize the protein. The chemical shifts of the trapped excited-state version agreed with those from the 3% excited population, providing strong evidence that the intermediate state structure was correct.

Importantly, the excited-state FF domain structure provides experimental proof that metastable folding intermediates do exist, an issue that has been controversial because of the

GENE EXPRESSION

MANY MINI MIND PROMOTERS

Tools to drive restricted gene expression in the brain.

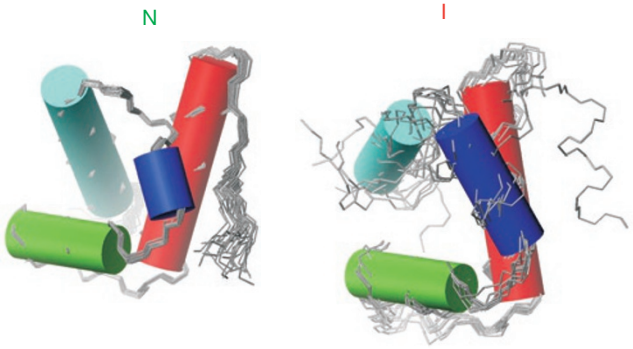
A complex tissue is a mosaic of many cell types and subtypes. For a scientist interested in how a tissue functions or is organized, the ability to target gene expression to restricted sets of cells is a critical experimental tool. And yet, with a few exceptions, such tools are notoriously difficult to come by.

Under the aegis of Genome Canada, a large multidisciplinary group of scientists predominantly at the University of British Columbia, Vancouver, now report the assembly of a pipeline to design mini promoters (MiniPs)—short constructs of human DNA sequence driving restricted gene expression in the mouse brain. The success of the project—named the Pleiades Promoter Project—depended entirely on its scale, explains Elizabeth Simpson, a mouse geneticist and one of the senior scientists directing the project. "I'd always wanted to develop more promoters for expression in the brain, but only by partnering with colleagues could we put together something large enough, and I think that was key," she says. As her collaborator—bioinformatician Wyeth Wasserman—puts it, "I predicted to Beth that somewhere between one in five and one in ten of these constructs would work."

But these odds apparently did not deter the Pleiades team. The biologists first sifted through brain expression data and identified about 200 genes with expression patterns of interest. They handed these over to the bioinformaticians, to attempt to predict regulatory regions driving those patterns. Especially in a mammalian system, this is no easy task. As a first step, Wasserman and his colleagues winnowed the pile of genes to about 60 that they judged to be most tractable for bioinformatic analysis. Simply stated, these were genes with single transcription start sites and relatively short gene length, in which the scientists could define areas of conservation between the mouse and human sequence. "You certainly miss things this way," says Wasserman, "[but] we wanted to start with the genes with the best chance for success."

With the tractable set of genes in hand, the researchers then used several sources of information to identify regulatory regions. They combined prior annotations, chromatin immunoprecipitation data for transcription factors and epigenetic marks, and conservation between mouse and human, and then assembled short (<4 kilobase) MiniPs of human sequence predicted to yield expression in restricted regions of the brain. This information went back into the hands of the biologists, who put these MiniPs to the test in the mouse.

Using single-copy insertion of transgenes at a defined locus,



FF domain structures. The secondary and tertiary structures of the native state (N) differ from the structure of the folding intermediate (I), which is also more disordered. Reprinted with permission from the American Association for the Advancement of Science.

lack of experimental methods to simply detect intermediates, let alone solve their structures. Kay notes that the method could also be used to gain insights into many other interesting questions in biology, ranging from understanding protein misfolding, to molecular recognition, to enzyme catalysis and to how macromolecular machines function.

The future applications of this approach are no doubt 'exciting', but Kay stresses that these are still early days. "There will have to be development of NMR methodologies to keep pace with the biochemical questions that we want to answer," he notes.

Allison Doerr

RESEARCH PAPERS

Korzhnev, D.M. *et al.* A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).

Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA* **105**, 4685–4690 (2008).

Simpson and her colleagues systematically tested about 100 MiniPs for their ability to drive gene expression in the adult mouse brain. A substantial fraction of these constructs, about one-third, yielded interesting patterns. Some constructs apparently recapitulated endogenous expression of the gene from which they were derived, others could drive reporters in only a subset of the endogenously expressing cell types, and still others showed entirely new patterns, but all of them yielded reporter expression in restricted regions of the brain.

Undoubtedly, there is still much to be learned from these tools. "Even with a team like this, and it's obviously not small, we simply couldn't characterize these constructs in the kind of detail that a real biologist wants," says Simpson. But they are making all of the plasmids, embryonic stem cell lines and mice available to other researchers, she emphasizes, and hope that this will allow interested scientists to use them in their own work.

Together with the information available from the bioinformatics analysis, these resources should provide a finer set of tweezers both for probing gene regulation and for a multitude of functional studies in the brain and potentially other tissues.

Natalie de Souza

RESEARCH PAPERS

Portales-Casamar, E. *et al.* A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proc. Natl. Acad. Sci. USA* **107**, 16589–16594 (2010).

NEWS IN BRIEF

MOLECULAR ENGINEERING

Monoclonal antibodies without screening

In developing a pipeline for rapid monoclonal antibody production, Reddy *et al.* describe a strategy that bypasses the need for extensive screening of immortalized B cells or recombinant antibody libraries. Instead, they used high-throughput DNA sequencing to identify the variable region gene repertoires produced by bone marrow plasma cells of immunized mice. They paired the most abundant variable region heavy and light chain genes, reconstructed them using automated gene synthesis and recombinantly expressed the monoclonal antibodies.

Reddy, S.T. *et al. Nat. Biotechnol.* **28**, 965–969 (2010).

IMAGING

Imaging mRNA transport

It is challenging to clearly observe mRNA export from nuclear pores. Grünwald and Singer now describe a super-registration approach for fluorescence microscopy that facilitates measuring distances between different fluorophores with very high time and spatial precision. This technique allowed them to observe the kinetics of β-actin mRNA transport in mammalian cells with an order of magnitude greater precision than previously possible. The method should be useful for imaging other transient cellular processes.

Grünwald, D. & Singer, R.H. *Nature* **467**, 604–607 (2010).

EPIGENETICS

Mapping the histone methyl lysine interactome

Vermeulen *et al.* report an integrated approach to understand how histone methyl lysine modifications regulate gene expression. First, they used quantitative mass spectrometry to identify proteins that bound to trimethyl lysine marks on histones H3 and H4. Then they fluoroscently tagged the identified proteins and performed pulldowns to identify physical interactions. Finally, using chromatin immunoprecipitation and DNA sequencing, they identified the genomic binding sites of the proteins. The resulting dataset is a rich resource for chromatin researchers.

Vermeulen, M. *et al. Cell* **142**, 967–980 (2010).

BIOINFORMATICS

Enzyme discovery without evolution

Before embarking on extensive direct evolution experiments to create enzymes with new activities, it is worth checking first whether such tools already exist in nature. Höhne *et al.* show that enzymes can be discovered by using rational design principles to predict key amino acids that perform the desired activity and then searching protein sequence databases to find proteins that already have these sequences. Applying this approach, they identified 17 (*R*)-enantioselective amine transaminases.

Höhne, M. *et al. Nat. Chem. Biol.* **6**, 807–813 (2010).

BIOPHYSICS

Monitoring membrane protein interactions

Casuso *et al.* show that high-speed atomic force microscopy (HS-AFM) can be used to experimentally monitor bacteriorhodopsin dynamics in native purple membranes from *Halobacterium salinarum*. Whereas AFM has a typical image scanning speed of about one minute, HS-AFM can record an image in about 187 milliseconds, allowing them to observe the movement of ATP-synthase c-rings, undetectable by other methods.

Casuso, I. *et al. Biophys. J.* **99**, L47–L49 (2010).