SPECTROSCOPY

Protein structure determination… blindfolded

Researchers present a method to generate protein structures using NMR chemical shift data in combination with powerful modeling software, simplifying and shortening the process of NMR-based structure determination.

Protein structure determination by NMR spectroscopy is typically a lengthy process. The slowest step in the overall process is the collection and analysis of nuclear Overhauser enhancement (NOE) spectra to obtain through-space correlation information used to generate a three-dimensional structure.

Many groups have thus been developing methods to minimize the time required for NMR-based structure determination and to simplify the process. The groups of Ad Bax of the US National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health, David Baker of the University of Washington and their colleagues from the Northeast Structural Genomics Consortium (NESG) now present a method that uses chemical shift information alone to generate high-quality protein structures (Shen *et al.*, 2008).

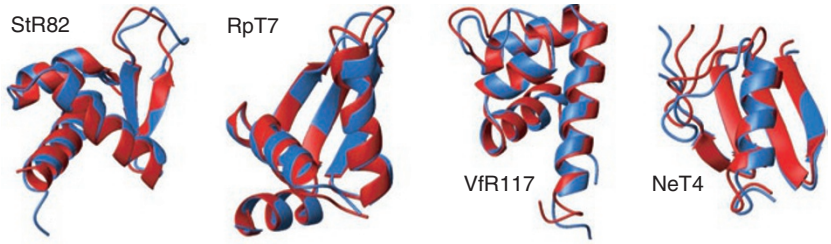

Figure 1 | CS-Rosetta models obtained in a blinded fashion (red) compared to experimental NMR structures (blue). Copyright 2008 National Academy of Sciences, USA

Each nucleus in a molecule experiences a unique chemical environment and thus has a distinct chemical shift in an NMR spectrum. Yang Shen in the Bax group previously developed a program called Sparta, which predicts chemical shifts for proteins of known structure, and used this program to generate a database of hypothetical chemical shifts for more than 5,000 proteins. For a new protein structure, experimental chemical shifts are matched to the database to identify peptide fragments with similar hypothetical chemical shifts. Working with Baker's postdoc Oliver Lange, Shen then used these fragments as inputs for *de novo* protein structure generation with Baker's powerful modeling software, Rosetta, "basically out of the box with very little adaptation," explains Bax. They call the method CS-Rosetta, for chemical shift-Rosetta. Michele Vendruscolo's group from Cambridge University also recently published a very similar concept called Cheshire (Cavalli *et al.*, 2007).

The researchers tested CS-Rosetta on 16 small proteins for which NMR or crystal structures were already available. For all of these test cases, the Rosetta models came within 0.7–1.8 Å of the experimentally derived structures. However, "when you're testing protein structure prediction programs, if you already know the answer

GENE TRANSFER

KNOCKOUTS WITH A TOUCH OF A ZINC FINGER

Capitalizing on the imperfection of error-prone DNA repair, researchers generated knockout mammalian cell lines by targeting a nuclease to a gene of interest via a zinc-finger DNA-binding domain.

Faced with a new gene product, researchers often want to know the phenotype of a cell that is forced to do without it. Knocking out these target genes by homologous recombination is now routine in mouse embryonic stem cells. But as Philip Gregory, vice president of research at Sangamo BioSciences, points out, "most other cell types are quite recalcitrant to homologous recombination–mediated gene knockout, so the efficiency of that process is very low."

But there is another way. In mammalian cells, in addition to homologous recombination, double-strand breaks can be repaired by nonhomologous end joining—an imperfect repair process that helps cells survive in times of stress but results in alteration of genetic information at the site of the break. To take advantage of this imperfection to create a knockout cell line, the researchers needed to introduce a double-strand break within a gene of interest.

Sangamo is a company focused on developing zinc-finger technologies, and for this project they used an engineered zinc-finger nuclease that comprises a zinc-finger DNA-binding domain and an endonuclease domain that cleaves DNA upon dimerization. Thus, two zinc-finger domains are designed to recognize a specific sequence within the target gene; when they bind their target sites, two fused nuclease domains are brought together and create a double-strand break. While repairing this lesion by the endogenous nonhomologous end joining process, the cell effectively 'makes' the knockout.

As proof of concept, Gregory and colleagues targeted the *DHFR* gene in a Chinese hamster ovary (CHO) cell line, which is often used for production of therapeutic recombinant proteins. After transiently transfecting these cells with plasmids encoding different optimized zinc-finger nucleases, they found that 7% of the isolated clones had mutant *DHFR* alleles in one experiment and 3% in another. Sequencing of the PCR-amplified target regions revealed that about one-third of these clones had a disruption of both alleles of the target gene.

One advantage of this approach is that no selection marker is used. Also, in traditional homologous recombination–based knockout approaches, often only one allele is mutated, which requires marker excision and retreatment to isolate a homozygous knockout—a time-consuming process. The surprise in this work was that the frequency of both alleles being mutated was relatively high. It turns out, as Gregory explains, that "cells that have nucleases that are expressed and are sufficiently active to cleave at one allele tend to also cleave at the other, so both alleles are

you never quite know whether you can 100% trust it," notes Bax. "You sort of start saying, 'well this is obvious, this is logical, and clearly I shouldn't take this answer because it's ridiculous'. And in the absence of a blind test, you can never be really sure." Therefore they also tested CS-Rosetta on 9 novel protein targets of the NESG, which were in the process of being solved by traditional NOE-based NMR methods. Once the results were in (**Fig. 1**), "this finally made us believers that this was for real because initially I was worried the results were really too good to be true," says Bax.

Bax anticipates that methods like CS-Rosetta and Cheshire will rapidly become accepted in structural genomics because they greatly simplify and shorten the process of structure determination. CS-Rosetta currently is limited, however, to small proteins of about 15 kDa or less. "It appears that the most severe limit is computational; the Rosetta approach really blows up exponentially, and it already takes a humongous amount of computational time," explains Bax. He also notes that as proteins get larger, their folds become more complex. However, Rosetta allows input of any kind of data, so entering additional structural information such as disulfide bond links or just a few long-range NOEs could simplify analyses for larger proteins.

Bax also expects that future developments in NMR spectroscopy will further the application of this technology. He predicts that, "Over the next ten years, we could develop a much more quantitative relationship between chemical shift and structure, at which point one would be able to get atomic-resolution structures better than what we can get from current technology."

Allison Doerr

RESEARCH PAPERS

Cavalli, A. *et al.* Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. USA* **104**, 9615–9620 (2007).

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

simultaneously disrupted at a high frequency." As a result, he estimates that the time required for this approach is less than half of that required for current homologous recombination–based methods, such as those that use adeno-associated virus (AAV)-mediated delivery.

The hardest part of both of these approaches is making the targeting tools, and neither is an off-the-shelf system. "At the moment most labs that find it difficult to make an AAV vector would also find it difficult to make a zinc-finger protein," Gregory notes. "But both are increasingly available through commercial entities and will be available to researchers." Sangamo's zinc-finger proteins and zinc-finger protein nucleases are now being distributed by Sigma-Aldrich.

In the meantime, Sangamo scientists are partnering with Pfizer and others in the industry to use this method to create improved cell lines for manufacturing proteins for human therapeutics. But another direction for this work, according to Gregory, is that it might be a method to create transgenic animals. "Because of the simplicity of the method—the fact that you only have to make a double-stranded break—and the frequency with which you get biallelic targeting, it's an exciting possibility that this will be a method for the creation of transgenic animals."

Irene Kaganman

RESEARCH PAPERS

Santiago, Y. *et al.* Targeted gene knockout in mammalian cells using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA*, published online 21 March 2008.

NEWS IN BRIEF

IMAGING AND VISUALIZATION

Guidelines for reporting gene expression localization

Deutsch *et al.* present guidelines for authors reporting gene expression localization experiments, the minimum information specification for *in situ* hybridization and immunohistochemistry experiments (MISFISHIE). Six types of information—including experimental design, specimens and treatments, reporters, staining, imaging data, and image characterization—should be provided to allow experiments to be reproduced.

Deutsch, E.W. *et al. Nat. Biotechnol.* **26**, 305–312 (2008).

GENE REGULATION

Transgene combinatorics

For *Caenorhabditis elegans*, researchers primarily rely on endogenous enhancers to regulate transgene expression, but this approach restricts them to the spatiotemporal activity of those enhancers during development. Davis *et al.* created a system to turn on transgene activity when a site-specific recombinase removes an intervening marker gene. Different regulatory regions for the transgene and recombinase create a combinatorial code that can better target transgene expression.

Davis, M.W. *et al. PLoS Genet.* **4**, e1000028 (2008).

PROTEIN BIOCHEMISTRY

Monitoring by SPR transcription

Each step of transcription—from initiation to termination—is dictated by the sequence of the nucleic acids involved and by the proteins comprising the RNA polymerase (RNAP) complex. Greive *et al.* applied surface plasmon resonance (SPR) technology to quantitate these steps in solution and determine their rate constants. Using this method they observed processes including RNAP binding DNA as well as transcript release and RNAP dissociation at terminators.

Greive, S.J. *et al. Proc. Natl. Acad. Sci. USA* **105**, 3315–3320 (2008).

MICROSCOPY

Pushing the resolution of electron cryomicroscopy

Though electron cryomicroscopy does not usually yield structures with resolution as high as crystallography, no crystallization is needed, and structures can be observed under near-native conditions. Jiang *et al.* report a 4.5 Å resolution structure for the infectious epsilon15 virus capsid using single-particle electron cryomicroscopy. They obtained this very high resolution for the 22 MDa structure by collecting 20,000 individual particle images and using distributed computing for image processing.

Jiang, W. *et al. Nature* **451**, 1130–1134 (2008).

MICROBIOLOGY

Complete ORF collection for *Vibrio cholerae*

Cholera, caused by the bacterium *V. cholerae*, is still a major public health issue in many parts of the world where it is easily spread through contaminated water. Rolfs *et al.* report a complete FLEXGene clone set containing full-length open reading frames for a pathogenic *V. cholerae* strain, with each clone verified by sequencing to ensure high quality. The authors hope that this resource will stimulate research on cholera, and perhaps lead to new vaccines and therapeutic strategies.

Rolfs, A. *et al. Proc. Natl. Acad. Sci. USA* **105**, 4364–4369 (2008).