

STRUCTURAL BIOLOGY

Designer labels

Stanley J. Opella

An innovative approach to the isotopic labelling of proteins places the determination of the structures of larger proteins — particularly those in solution — within reach, while improving the accuracy of the method.

With completion of the sequencing of the human and many other genomes, molecular biology is able to return to its roots to focus on, as Francis Crick¹ put it, “the description of [proteins, viruses, bacteria and chromosomes] in terms of their structure, i.e. the spatial distribution of their constituent atoms”. Significantly, this definition includes systems that range from individual globular proteins in aqueous solution or highly regular crystals, to large assemblies of macromolecules, such as virus particles and membranes, that behave like aperiodic solids².

Nuclear magnetic resonance (NMR) spectroscopy can be used on molecules in all physical states and degrees of order, and so is fundamentally capable of determining the structures of proteins in all these environments. Although powerful, the full promise of this technique has yet to be realized. In this issue, Kainosho *et al.* (page 52)³ demonstrate a method to optimize the isotopic make-up of the proteins being studied. This allows far more information to be gleaned by both simplifying and sharpening the spectra.

The potential of NMR spectroscopy for structure determination was obvious from the earliest results, when Pake⁴ showed that the distances and orientations of pairs of nuclei could be determined from the perturbations that the magnetic field of one nucleus causes in a neighbouring nucleus. The initial ¹H NMR spectrum of a protein⁵ showed that a dispersion of chemical shift frequencies results from the local spatial relationships of atoms that are not necessarily bonded to one another; thus, the chemical shift interaction provides a means to distinguish all of the atoms in a protein.

The Achilles' heel⁶ of protein NMR is the ‘size problem’, which has two distinct components. The first is obvious: the spectra of larger polypeptides are more complex, with more signals and consequently overlap within the same limited range of frequencies. This is compounded by the second, more daunting, ‘correlation-time problem’: larger polypeptides (and their complexes) tumble slowly in solution⁷, which results in broad linewidths. The correlation-time problem can be solved under certain circumstances for proteins that are immobile (on NMR timescales) by virtue of their being in the solid state or very large complexes. But the problem has so far been intractable for

larger soluble proteins in their native aqueous environment.

The past half-century of technological developments in NMR for investigating proteins has followed three parallel paths: instrumentation, experimental methods and isotopic labelling. The benefits of advances in all three areas are highly synergistic, despite their seemingly disparate nature. The selective replacement of hydrogen (¹H) with deuterium (²H) in a protein⁸ was the first real step towards the resolution of overlapping NMR spectra, preceding the advent of high-field magnets and the sophisticated magnetic pulse sequences employed today. Deuterium nuclei couple much more weakly with the external magnetic field (and with the nearby ¹H nuclei of interest), so that replacing most ¹H by ²H has a substantial effect on the properties of the remaining ¹H nuclei. In various incarnations, the incorporation of deuterium remains among the most widely used methods in protein NMR, even though, up to now, it required several difficult trade-offs.

When uniform deuteration is performed at very high levels, there are no ¹H nuclei left as sources of magnetization or the couplings that provide structural information. Intermediate levels of deuteration generate a mixture of proteins with the ²H nuclei in different positions, complicating the chemical shift patterns, and the signal intensity decreases because there are fewer ¹H nuclei in the sample.

Kainosho and colleagues³ describe a strategy for optimal isotope labelling for protein NMR that addresses these problems. Their approach goes far beyond just improving sample preparation — it is a major step forward in the attack on NMR's size problem. In addition, it provides a path to structure determination with higher spatial resolution. The authors developed protocols for ‘stereo-array isotope labelling’ (SAIL), which incorporates deuterium into a protein's constituent amino acids such that each carbon or nitrogen nucleus in the final protein will have at most one ¹H nucleus bonded to it, the remaining hydrogen atoms having been replaced by ²H. These amino acids are then used to make the protein in a cell-free expression system that maintains the careful positioning of the isotopes. The patterns of replacement are designed to provide the most information consistent with spectral simplification and isotopic dilution (Fig. 1 on page 52), and it constitutes a general approach to overcoming the limitations of

deuteration in a way that also enhances the contributions of ¹³C and ¹⁵N labelling for both distance and angle measurements throughout the protein. Contemporary alternatives are more limited; their focus is on specific labelling of methyl groups in a highly deuterated background^{9–11}.

The SAIL technology improves NMR spectra because it decreases the number of signals without losing sensitivity for the resonances of interest, and it produces narrower linewidths because the dilution of ¹H nuclei attenuates the line broadening. Moreover, it makes the short-range distance measurements more accurate, and, combined with easier assignment of side-chain resonances, leads to higher-resolution structures. Kainosho *et al.* demonstrate their approach by solving the structures of two proteins — calmodulin (relative molecular mass 17,000) and maltose binding protein (41,000) — at a resolution of about 1 Å. This rivals the highest resolution of structures previously determined for smaller proteins by solution NMR or X-ray crystallography.

The combination of SAIL with other NMR technologies means that protein structure determinations can be expected where accuracy is no longer limited by experimental methodology. Rather, the structures could reflect the fundamental dynamic properties of the backbone and side-chain atoms — portraying the essence of a working protein. Equivalent improvements in solid-state NMR studies can be readily envisaged.

Kainosho and colleagues' study is a significant advance in the field of protein NMR. It demonstrates that by dipping your heel in the river while sailing, the protein size problem will fade away. This will allow the study of proteins selected on the basis of their biological functions or in genuinely unbiased surveys of proteomes, resulting from its ease of automation. ■

Stanley J. Opella is in the Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0307, USA.
e-mail: sopella@ucsd.edu

- Judson, H. F. in *The Eighth Day of Creation* 88 (Cold Spring Harbor Laboratory Press, New York, 1996).
- Schrödinger, E. *What is Life?* (Cambridge Univ. Press, 1944).
- Kainosho, M. *et al.* *Nature* **440**, 52–57 (2006).
- Pake, G. E. *J. Chem. Phys.* **16**, 327–336 (1948).
- Saunders, M., Wishnia, A. & Kirkwood, J. G. *J. Am. Chem. Soc.* **79**, 3289–3290 (1957).
- Homer *The Iliad* (Penguin, London, 1999).
- Einstein, A. *Ann. Phys. (Leipz.)* **17**, 549–560 (1905).
- Markley, J. L., Putter, I. & Jardetzky, O. *Science* **161**, 1249–1251 (1968).
- Chaykovsky, M. M. *et al.* *J. Am. Chem. Soc.* **125**, 15767–15771 (2003).
- Tugarinov, V. & Kay, L. E. *J. Biomol. NMR* **28**, 165–172 (2004).
- Tang, C., Iwahara, J. & Clore, G. M. *J. Biomol. NMR* **33**, 105–121 (2005).