

Bench philosophy (26): Protein structure solution with X-Ray crystallography and NMR

Together for the better

Nuclear magnetic resonance spectroscopy (NMR) and X-Ray crystallography are the main techniques to solve protein structures. Though some crystallographers still hold some reservations against NMR, it is well accepted among structural biologists that both methods may complement each other to get a better picture of a given protein structure.

In the last ten years, the number of annotated genomic sequences of organisms ranging from bacteria, fungi, plants, insects and higher animals, including *Homo sapiens*, has been doubling every 16-18 months in public databases (Jackman et al., *Genome Biol.* 11, 202-5). The Wellcome Trust Collection in London holds a piece of human knowledge (a book!) that contains the complete DNA sequence of the Human Genome Project. The gigantic piece is 121 volumes long, each one of them containing a thousand pages filled with eight-point letter case sequences. If the annotation trend continues at the present speed, probably in a decade from now there will be enough sequence material to print the longest book ever printed, and with only four different letters!

The information extractable from these sequences books may prove to be extremely valuable, not only because it contains the evolutionary molecular history of organisms but also because the codes for the amazing diversity of proteins are truly written there. Bearing in mind that there are roughly 1.75 million species described in literature (an estimation of the total number of organisms on earth is between five and thirty million), the annotation of a large

part of the genomic diversity can be numerically challenging and probably quite redundant. However, this is just the start of the biological pipeline. The astronomical task of translating all the genes into functional and characterised proteins remains to be done.

Typically, a structural biology project begins by focusing on a gene that encodes for a protein, which is believed to be essential for a biological task. The target could be an enzyme involved in a disease or having an important function in the human body. It may also be any essential pathogenic protein or even a complex multifunctional receptor appearing in specific cells. By establishing the three-dimensional arrangement of atoms in the space, it is possible to consider the structural interactions that the macromolecule may make with small molecules (substrates, inhibitors, activators) or with other macromolecules (receptors, other proteins).

Old techniques

For example, insulin is a hormone that decreases blood glucose concentration by accelerating glycolysis. Type II diabetes is an increasingly common disease caused when the CD220 tyrosine kinase recep-

tors become resistant to the protein. The three-dimensional structure of insulin was solved by X-ray crystallography in 1969 by the British Nobel Laureate, Dorothy Crowfoot Hodgkin, after 34 years of struggle! Using pig insulin soaked in different metal solutions (zinc, lead and cadmium), diffraction patterns at 2.8 and 1.9 Å resolutions were acquired. These results showed two different polypeptide chains bound together by a pair of disulphide bridges, arranged in a hexameric fashion in the crystal. As early as her Royal Society Bakerian Lecture in 1972, Dorothy Hodgkin was citing NMR data from other groups that showed that oxitocin (a polypeptide mammalian hormone) had a similar conformation to parts of insulin (Hodgkin, *Proc. R. Soc. Lon. A.* 338, 251-75). Even in the early days of protein structure elucidation, it was recognised that X-ray crystallography and NMR could complement each other.

These days, it is common for the structural biologist to try both techniques for the detailed resolution of their studied proteins, although certain reticence exists, particularly among crystallographers. Not all proteins are amenable to crystallisation or, alternatively, not all proteins may display clear assignable NMR spectra, not to men-

	X-Ray Crystallography	NMR
Requirement	Macromolecule must crystallise and the crystal must diffract	Macromolecule must be soluble and typically isotopically labelled
Physical state	Solid-frozen structure	In buffered solution
Size	Any macromolecule	Small polypeptide (less than 50 kDa)
Time	Long screening and optimisation but short time for processing data	Short for preparation but long time for analysis of the data
Advantages	High atomic resolution usually obtained for well-diffracting crystals	Soluble state closer to real functional environments or foldings
Disadvantages	Necessity of solving the "phase problem", screening and optimisation needed, crystallisation may change the conformation of proteins due to packing interactions	Isotopic labeling and deuterated reagents can be expensive, larger proteins can lead to poor spectra, instability during data collection

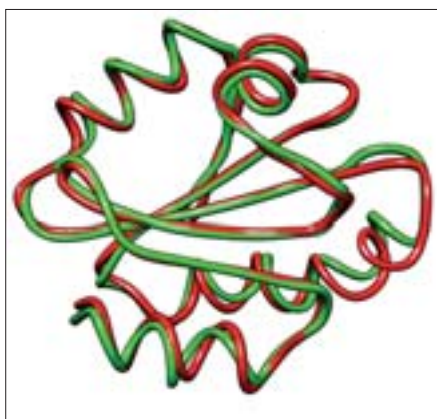
tion the bigger multiprotein complexes or membrane proteins, which are much more difficult to solve, frequently determined using electron microscopy techniques in combination with X-ray or NMR. There is no a priori way of selecting a successful technique for the elucidation of the structure of a protein, only trial and error for each protein usually brings the answer.

Similar but not identical

The most important differences between X-ray crystallography and NMR are summarised in the table below and recent researchers in the field claim that the two techniques are complementary, providing a more complete and efficient strategy for protein structural determination (Snyder et al., *J. Am. Chem. Soc.* 127, 16505-11 and Yee et al., *J. Am. Chem. Soc.* 127, 16512-17). It is well known that structural parts with conformational flexibility are not observable in X-ray crystallography; instead NMR can be used to identify these flexible regions. NMR is a blind technique but displays dynamic information. As a matter of fact, the folding of the protein can be easily observed just by looking at the dispersion of the signals in heteronuclear ^1H - ^{15}N HSQC spectra. Recent work, for example, from John Christodolous' group at the Institute for Structural and Molecular Biology (ISMB) in London has started giving new insights into the way folding occurs after proteins are assembled in the ribosome using NMR analysis (Cabrita et al., *Curr. Opin. Struct. Biol.* 20, 33-45).

At 1st July 2010, there were 66,212 structures deposited in the Protein Data Bank (PDB), split into 57,310 macromolecules solved by X-ray crystallography and 8,462 solved by NMR, accounting for 86.55% and 12.78% respectively. The structures that have been solved independently by NMR and X-ray crystallography are similar but not identical. Let's, for example, take a look at thioredoxins as a specific class of proteins to compare NMR and X-ray solved structures. Thioredoxins are small enzymes that reduce disulphide bridges to dithiol moieties and they are ubiquitous in all of nature's kingdoms. Human thioredoxin has been solved by NMR (PDB: 3TRX) and by X-ray crystallography (PDB: 1ERT), and the structures obtained were very close. The protein backbone of each model is represented in the figure above. The difference in the position of the atoms between the two structures was calculated as the root mean square deviation (rmsd) to be around 0.88 Å, which is less than the distance between

the carbon and the oxygen in the peptide carbonyl (1.23 Å). It is, therefore, clear that the differences between the two structures are very small and from the figure, we can infer that the regions of maximal dissimilarity are the α -helices. Other researchers have found that the packing interactions in the crystal may reduce their mobility. The NMR structures of several proteins appear more distorted than the X-ray structures



Human thioredoxin solved by X-ray crystallography (red, PDB: 1ERT) and by NMR (green, PDB: 3TRX). Molecular image modelled using UCSF Chimera software (Pettersen et al., *J. Comput. Chem.* 25, 1605-12).

and this fact makes sense because in solution their mobility is greater and this phenomenon allows ligand recognition. In such cases, the NMR structure obtained may be more interesting as it shows different conformations that may be involved in recognition and binding.

Also, it is very important to remind the readers that a sound solution obtained by either one of the methods, is just a model, an abstract representation of reality that can, of course, be improved but that will never be perfect! This reminds me of "La trahison des images" from René Magritte, when the artist wrote as the legend to a smoking pipe drawing "Ceci n'est pas une pipe" (This is not a pipe), referring to the drawing as a representation of an object but not the object itself. The same happens with molecular objects and their models, hence interpretation of the models should be performed with caution.

Of course, when an interaction is captured in a crystalline structure with a good R-free value (statistical quantity that measures the agreement between the observed and the computed amplitude of reflections) and a high-resolution, you can really assume that there is a decrease in energy by stabilisation of the complex. For NMR,

the chemical shifts of the residues are very sensitive to variations of the local electronic environment and, therefore, a variation in the chemical shift when introducing a ligand enables determination of the important binding residues.

However, there is always the possibility that the interaction captured with a method for structure resolution has no biological significance as it does not happen *in vivo*. This is causing controversies among the scientific community. For example, in 2006, the most popular anti-tubercular agent to-date, isoniazid, was reported to target dihydrofolate reductase (dhfr) from *Mycobacterium tuberculosis*, based on crystal structures (Argyrou, et al. *Nat. Struct. Mol. Bio.* 13, 408-13). However, a paper from 2010 reported that the minimum inhibitory concentration of isoniazid in plasmid-mediated dfrA over-expression was the same as the wild type in both *M. smegmatis* and *M. tuberculosis* (Fang, et al. *Antimicrob Agents Chemother.* ahead of print). They were also unable to find mutations in the dfrA gene on clinical isoniazid resistant isolates, therefore, they concluded that dihydrofolate reductase is not a significant target of isoniazid in *M. tuberculosis*.

Synergistic effect

Though X-ray crystallography is the preferred structural tool for the determination of three-dimensional protein structures, more and more researchers in the biosciences are finding NMR useful and therefore, the number of NMR-solved structures in public databases is increasing considerably. The advantages and limitations of both techniques, in fact, make them complementary and a much higher probability of success can be achieved if both methods are started simultaneously at the beginning of the project. In the best scenario, if both techniques are successful, the information obtained is much more than the additive result. It is possible to establish the degree of flexibility when the crystallised protein is compared with the protein in solution and this offers a perspective of the effects of chaotic conditions when compared with ordered, packed structures.

JUAN DAVID GUZMAN

Fancy composing an installment of "Bench Philosophy"?

Contact Lab Times
E-mail: editors@lab-times.org