

Macromolecular Structure Determination: Comparison of Crystallography and NMR

J Mitchell Guss, *University of Sydney, New South Wales, Australia*

Glenn F King, *University of Connecticut Health Center, Farmington, Connecticut, USA*

The two techniques used to define the three-dimensional structures of biological macromolecules at or near to atomic resolution are X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR). These techniques are based on different physical principles, and may be applicable to different systems, but they often yield complementary results.

Introduction

Most biological macromolecules function correctly only if they are folded into their proper or 'native' shape. To fully understand the details of how a molecule functions, one must first know its structure in as much detail as possible.

The discovery of X-rays by Röntgen at the end of the nineteenth century was rapidly followed by their use for exploring the structures of crystals by von Laue and the Braggs. Experiments on diffraction from biological fibres in the 1930s were accompanied by early experiments on crystals of macromolecules by Dorothy Crowfoot (Hodgkin) and J. D. Bernal. It then took more than twenty years for the technology to be developed sufficiently for John Kendrew to solve the first crystal structure of a protein, namely myoglobin from sperm whale muscle, in 1959. This was followed closely by elucidation of the structure of haemoglobin by Max Perutz.

The assertion by Wolfgang Pauli in 1924 that certain nuclei spin and therefore possess a magnetic moment ultimately led to the first experimental observation of nuclear magnetic resonance in 1945. However, it was not until the advent of multidimensional NMR spectroscopy in the 1970s that the technique matured sufficiently to be applicable to macromolecular systems. The first complete protein structure solved using NMR methods, namely that of proteinase inhibitor IIA from bull seminal plasma, was determined by Kurt Wüthrich's group in 1984.

The Protein Data Bank (see Further Reading), which is the repository for coordinates of macromolecular structures, now contains more than 10 000 entries representing the structures of hundreds of different proteins that have been elucidated using X-ray crystallography and NMR spectroscopy.

Introductory article

Article Contents

- Introduction
- Underlying Physics: What Sort of Radiation Interacts with What in Molecules?
- Transforming Data into a Model (Calculated from Distance Geometry; Built into Electron Density)
- What in the Model Has Been Inferred from Other Information Versus What Comes Directly From the Data?
- Judging Model Quality
- Dynamic Information
- Practical Considerations: Size, Solubility and Stability Versus the Randomness of Crystallizability, Exchange Rates for Complexes, etc.
- Most Common Errors and Pitfalls of Each Technique

Underlying Physics: What Sort of Radiation Interacts with What in Molecules?

The methods of X-ray diffraction and NMR spectroscopy rely on fundamentally different physical processes to generate the information needed to determine the three-dimensional structure of a macromolecule. X-ray diffraction is essentially an imaging technique in which X-rays are scattered by electrons in the atoms of crystal without loss of energy (elastic scattering). The scattered X-rays generate an interference pattern that can be recorded and subsequently transformed to yield an image of the original scattering object, in this case the molecules in the crystal. The choice of X-rays, with wavelengths in the range 0.05–0.25 nm, as the incident radiation is dictated by the need for a radiation with a wavelength comparable to or shorter than the spacing between the atoms. The use of long-wavelength radiation such as visible light would not allow the visualization of small objects such as atoms or small groups of atoms.

Structure determination by NMR is based on the absorption of radiofrequency radiation incident on molecules held in a strong magnetic field. Nuclei with nonzero spin possess a magnetic moment that precesses around the direction of the external magnetic field. Resonance can be detected from the absorption of incident radiation oscillating at the nuclear precession frequency. The resonant frequency varies with the type of nucleus and the strength of the magnetic field. More importantly, this frequency is dependent on the chemical environment of the nucleus and can therefore be used to distinguish nuclei of atoms that are chemically identical but that have different surroundings. The most commonly observed nuclei are

protons (^1H) and the NMR-active nuclei of carbon and nitrogen (^{13}C and ^{15}N , respectively). The linear or one-dimensional NMR spectrum from a biological macromolecule is too crowded to permit identification of individual resonances. Multidimensional NMR techniques have been developed that enable the resonances to be spread out in two or more dimensions. These multidimensional spectra can be interpreted to yield estimates of the distances between hydrogen atoms that lie relatively close together in the folded molecule.

Transforming Data into a Model (Calculated from Distance Geometry; Built into Electron Density)

NMR is applicable to macromolecules in solution because it can be used to measure scalar quantities that do not vary with molecular orientation. Multidimensional spectra that record scalar or through-bond correlations between NMR-active nuclei are used to assign each NMR resonance to specific nuclei in the macromolecular system. While these spectra constitute the large majority of experiments performed during an NMR-based macromolecular structure determination, they yield relatively little information about macromolecular structure. Some scalar correlation experiments do provide useful structural information in the form of estimates of backbone and side chain torsion angles. However, the most useful structural information comes from multidimensional NMR experiments that record dipolar correlations between hydrogen atoms.

After hydrogen nuclei have absorbed radiofrequency radiation they can return to thermal equilibrium by through-space dipolar interactions with neighbouring nuclei – this process is commonly referred to as longitudinal dipolar relaxation. The strength of the dipolar interaction, which can be measured in nuclear Overhauser enhancement spectroscopy (NOESY) experiments, depends on the reciprocal of the sixth power of the internuclear distance. Thus, NOESY experiments can be analysed to provide semiquantitative measurements of interhydrogen distances for hydrogen pairs separated by less than $\sim 0.6\text{ nm}$; in a 100-residue protein one might expect to obtain several thousand interproton distance measurements. The estimates of dihedral angles and interproton distances are then used to reconstruct a model of the macromolecule. There are several mathematical procedures that can be used for this reconstruction process – many laboratories use distance geometry techniques to provide preliminary structures, which are then refined using restrained molecular dynamics, while some laboratories use only restrained molecular dynamics. In all cases, the aim is to obtain structures that satisfy all of the

experimental data; the reconstruction procedure is repeated many times and an ‘ensemble’ is generated of structures (usually 10–30) that satisfy the experimental data. The precision of the structure determination can be obtained from an overlay of the ensemble of structures. Regions of the structure that do not overlay well indicate either poor-quality NMR data or flexible regions that do not have a single well-defined structure in solution.

In contrast to NMR, the X-ray method ultimately yields an image of the scattering object. Since X-rays are scattered by electrons, the image is of the electron density throughout the crystal. In complete contrast with NMR spectroscopy, where experimentally measured interhydrogen distances are used to reconstruct the macromolecular structure, the small hydrogen atoms generally do not contribute significantly to the electron density map. In the diffraction experiment only the intensities of the diffracted waves can be recorded and all relative phase information is lost. In order to reconstruct the image, both the phases and the amplitudes of the diffracted waves are needed. These missing phases are determined experimentally.

The most widely used technique for structures unrelated to anything previously known is that of multiple isomorphous replacement (MIR) in which heavy atoms, which scatter X-rays more strongly, are added to the crystal. Differences between the diffraction intensities from native crystals and crystals with added heavy atoms may be used to define the heavy-atom positions and then to calculate the phases of the diffracted waves. A relatively new method, multiple wavelength anomalous scattering, uses the fact that X-rays from synchrotron radiation sources can be tuned to the absorption edge of a specified atom. This may be any atom that is present as only a few copies in the molecule, such as a metal ion (which may be intrinsic to the macromolecule) or selenium atoms that have been substituted for sulfur in the methionine residues of proteins using recombinant DNA technology. At wavelengths near to the absorption edge, the X-rays are scattered anomalously and large differences in the diffraction intensities can result from small changes in the energy of the incident X-rays. These ‘anomalous’ atoms can then be treated in a similar fashion to the heavy atoms in the MIR method. Another method of determining the missing phases relies on the fact that the unknown structure is similar to one that has already been determined. This method, termed molecular replacement, locates the known structure in the crystal of the unknown and then enables an initial set of phases to be calculated. Molecular replacement is assuming greater importance as the database of known structures continues to grow.

What in the Model Has Been Inferred from Other Information Versus What Comes Directly From the Data?

Both NMR and X-ray structure determination methods for macromolecules generally require prior knowledge or other information to interpret the structural data. This stems from the fact that in both methods the effective 'resolution' or ability to separate individual objects is not sufficient to distinguish individual atoms. There are rare instances when X-ray diffraction of macromolecules extends to better than 0.1 nm resolution and all aspects of the structure can be determined *ab initio*. In most cases, however, the resolution is in the range 0.18–0.30 nm. Resolution is not a simple parameter in NMR structure determinations, but it may be inferred by comparison with X-ray structures and would typically be in the range 0.20–0.30 nm.

At a resolution of 0.2 nm, the outline of chemical groups, such as the side-chains of proteins or the bases of nucleic acids, can be clearly distinguished. However, to build an atomic model, it is necessary to know the sequence of monomers that make up the macromolecule, accurate structures for the monomers, and the structures of the linking groups. It is sometimes possible to assign the sequence of a protein from a moderate-resolution electron density map or from NMR resonance assignments, but this would never be taken as definitive. Thus, prior information about the sequence of monomers in the protein or nucleic acid is generally essential for both NMR and X-ray structure determinations.

Chemically distinct groups such as the side chains of the amino acids aspartic acid and asparagine could never be distinguished on the basis of an X-ray structure alone, since the atoms involved scatter X-rays almost equally. However, these groups can be readily distinguished on the basis of the differing number of NMR-active nuclei in NMR spectroscopy. NMR spectra can also reveal the protonation state of ionizable amino acid side-chains, whereas this information can often only be indirectly inferred from the pattern of hydrogen bonds and salt bridge interactions in X-ray structures.

In addition to the covalent structural information, the refinement or optimization of structures using either X-ray or NMR data requires the input of restraints to prevent convergence to an energetically unreasonable structure. In practice this is accomplished by adding a relatively simple molecular mechanics force field in the final stages of the structure calculation.

Judging Model Quality

Errors associated with a macromolecular structure determination are of quite different kinds. There are those that

yield a completely erroneous result, and the random errors that define the precision of any experiment. The former will be dealt with in the next section. The result of a structure analysis is a set of three-dimensional coordinates and like any numerical result they should be accompanied by estimated standard deviations. This is certainly a routine publication requirement for the crystal structure of a small molecule. The errors are derived from the standard deviations of the experimental intensity measurements, which are adjusted for known systematic effects. Errors in the atomic coordinates are then calculated from the least-squares structure refinement procedure. In the case of a macromolecule, the situation is complicated by the incorporation of restraints in the refinement procedure and by the use of sparse matrix minimization methods, which do not readily yield errors for the variable parameters. The situation is compounded further for NMR where the distances estimated from the NMR spectra can only be assigned to ranges and not given specific values with associated errors.

One thing that is important to note is that, whether the technique is NMR or X-ray analysis, the errors in a macromolecular structure determination are not uniform throughout the structure and are different for each residue and atom. If parts of a crystal structure are disordered, no significant electron density may be present. If this is the case, then the only restraint on the atomic positions of the affected atoms is the requirement for molecular connectivity. Parts of the structure that are disordered or undergoing motion will be defined poorly whether the structure is determined using NMR or X-ray diffraction. In the former case, fewer distances will be defined for the affected region, and the resulting structure will show large divergences for those parts. Similarly, the X-ray structure will have high thermal parameters for the affected regions. In the case of X-ray analyses, these affects have been quantified by Cruickshank, Read, Jones and others. Cruickshank's diffraction precision indicator is now often quoted in publications of protein structures. The root-mean-square deviation values for the family of structures determined by NMR also give some estimate of the precision, although for the case of NMR this has not been quantified and related to standard deviations of the individual atomic coordinates.

The best single indicator of the quality of an X-ray structure determination is the resolution of the data. This property is generally related directly to the quality of the crystals and with modern instrumentation is less likely to be limited by the equipment. The effect of increased resolution is twofold. Firstly, higher resolution increases the detail visible in electron density maps and reduce the possibility of errors of interpretation. Secondly, as the resolution increases so does the number of observations. The increased ratio of experimental observations to parameters effectively decreases the reliance on the nonexperimental restraints in the refinement procedure.

The equivalent quantity for resolution for an NMR structure determination would be the number of experimentally derived restraints used in the refinement procedure (interproton distances, dihedral angles and orientational restraints). In the NMR experiment this parameter is directly related to individual residues and atoms and therefore varies throughout the structure, while resolution in the X-ray experiment is a single global parameter.

Dynamic Information

One of the principal differences in the outcomes of structure determinations based on NMR and X-ray diffraction data is the availability of information relating to dynamics or macromolecular motion. In the NMR experiment, measurements are made over a time scale of nanoseconds to seconds, which encompasses many of the important motions of biological macromolecules. It is possible using NMR to directly measure the motions of molecular groups or entire domains.

The scattering event, at the heart of the X-ray diffraction experiment, is fast compared with atomic or molecular motion. However, to record a measurable signal, experiments generally lasting minutes if not hours are required. Furthermore, the diffraction pattern arises from the entire crystal, which may not be uniform throughout. If the molecules exist in randomly different conformers throughout the crystal then the analysis will result in an average. Thus the X-ray diffraction experiment provides both a spatial and a temporal average. One of the parameters used in the refinement of the structure gives some indication of molecular motion. This exponential factor accounts for both real atomic motion that may be occurring in the crystal and the 'disorder' that results from having different conformations in the crystal. This conformational disorder may itself indicate parts of the molecule that would be flexible in solution. In theory, at least, making measurements at different temperatures can separate the two components, but this may be difficult in practice.

Practical Considerations: Size, Solubility and Stability Versus the Randomness of Crystallizability, Exchange Rates for Complexes, etc.

X-ray diffraction is practically limited only by the ability to grow crystals. Structures of large molecular assemblies, such as proteosomes, the ATP synthase complex, the photoreaction centre and viruses, have been solved using X-ray diffraction. Essentially the same methods are

applicable to large structures as to small ones. With present technology, the practical limit for a complete protein structure determination using NMR is around 30–40 kDa; however, this size limit includes numerous small proteins as well as many autonomously folded protein domains. Recent NMR developments that take advantage of anisotropic magnetic interactions promise to significantly increase this size limit in future years by narrowing spectral lines at high magnetic field strengths and by allowing the extraction of orientational restraints in addition to traditional dihedral angle and interproton distance restraints.

The upper size limit for determination of nucleic acid structures using NMR is typically lower than for proteins because of their intrinsically more limited spectral dispersion. However, protein–RNA structures as large as 38 kDa in size have been solved using NMR.

Somewhat surprisingly, given that the most often claimed difference of X-ray and NMR methods is that one is applicable to molecules in the crystal or solid state and the other to molecules in solution, the requirements for obtaining suitable experimental samples are in fact very similar. In both cases the sample should be very pure. This is especially important for X-ray crystallography, where even small amounts of impurities, especially those similar to the target molecule, may prevent crystallization or give crystals that diffract poorly. Recording multidimensional NMR spectra or growing crystals requires considerable time, ranging from hours to weeks or even longer. The samples must therefore be stable for long periods.

For NMR, the sample should be > 95% pure, monodisperse, and concentrated enough to give a measurable signal. NMR is a relatively insensitive spectroscopic technique and therefore the biomolecule or biomolecular complex needs to be in solution at a concentration of about 1 mmol L^{-1} , which corresponds to around 5–10 mg for a 20 kDa protein, depending on the sample size (usually 250–500 μL). The oligomeric state of the protein should be ascertained for proper interpretation of the NMR data; this can be achieved using methods such as analytical ultracentrifugation, dynamic light scattering, gel filtration chromatography, and pulsed-field-gradient NMR. Specialized NMR experiments are required for the extraction of intermolecular interproton distances in oligomers or biomolecular complexes.

Predicting whether or not a particular sample will crystallize and thus be amenable to X-ray diffraction is notoriously difficult. Extensive experiments with dynamic light scattering have shown that the best starting point is a monodisperse solution, just the requirement for NMR. The starting concentration for crystallization may range from a few to as much as 100 mg ml^{-1} . While the basic principles of crystallization are well understood, controlling the rate of reaching supersaturation and the formation of nuclei is not predictable. Most macromolecules are therefore crystallized by a random factorial approach in

which selected variables, including precipitant, pH, and added ions, are systematically varied and the most promising combinations are followed up by fine screening.

Complexes, such as an enzyme with bound substrate or inhibitor, need to be sufficiently stable in solution and be in sufficient concentration that crystals can be grown. Most complexes that have been crystallized have micromolar or lower binding affinities. NMR, on the other hand, is able to observe even transient complex formation in solution provided that there is time for the transfer of magnetization. This can provide a very powerful tool for studying the formation of weak complexes. NMR experiments such as saturation and inversion transfer can often be used to measure exchange rates. One particularly convenient feature of NMR spectroscopy is that binding surfaces, such as between a protein and a ligand, can be elucidated without the need to determine the structure of the protein–ligand complex. The biomolecule of interest is simply titrated with the ligand; the resonance frequencies of atoms involved in ligand binding are generally perturbed during the titration, thus revealing the nature of the interaction. This interface mapping technique can be applied to drug screening.

Most Common Errors and Pitfalls of Each Technique

It is possible to make a catastrophic error at an early stage of a crystal structure analysis, for example to assign the wrong space-group symmetry, and then to derive both a molecular and crystal structure that are completely incorrect. Fortunately, this type of error is very rare and can usually be prevented by careful attention to detail and by using the various aids that have been developed for monitoring the course of the progress of a crystal structure analysis. The best known of these tools is the free R factor, which monitors the agreement of the calculated and observed diffraction amplitudes but omits any bias from the model. Careful attention to this parameter prevents over-parametrization of the refinement procedure. In a relatively low-resolution X-ray analysis, the entire chain of the macromolecule may not be visible and breaks in the experimental electron density may lead to an incorrect connectivity in the model. The use of unbiased electron

density maps with combined experimental and calculated phases can help overcome some of these problems.

It is difficult to make a catastrophic error in NMR structure determinations without several critical resonance assignment errors. A single incorrectly assigned interproton distance should not cause serious problems, since it should reveal itself as irreconcilable with the many hundreds to thousands of other experimentally derived restraints. However, several key resonance assignment errors with unfortunate structural consequences have been documented in recent years. A more common and easily avoided problem is incorrectly analysing the spectra of a homooligomeric macromolecule as though it were a monomer; this leads to the designation of what should be intermonomer interproton distance measurements as intramonomer distances, often with disastrous consequences. It should be borne in mind that molecules that are monomeric *in vivo* can often polymerize at the high concentrations necessary for NMR spectroscopy.

Further Reading

- Berman HM, Westbrook J, Feng Z *et al.* (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**: 235–242. [<http://www.rcsb.org/pdb/>]
- Carter WC Jr and Sweet RM (eds) (1997) *Macromolecular Crystallography. Methods in Enzymology*, vols 276 and 277. New York: Academic Press.
- Cavanagh J, Fairbrother WJ, Palmer AG III and Skelton NJ (1996) *Protein NMR Spectroscopy: Principles and Practice*. San Diego: Academic Press.
- Clore GM and Gronenborn AM (1998) New methods of structure refinement for macromolecular structure determination by NMR. *Proceedings of the National Academy of Sciences of the USA* **95**: 5891–5898.
- Drenth J (1999) *Principles of Protein X-ray Crystallography*, 2nd edn. New York: Springer-Verlag.
- Evans JNS (1995) *Biomolecular NMR Spectroscopy*. Oxford: Oxford University Press.
- Glusker JP and Trueblood KN (1985) *Crystal Structure Analysis: A Primer*, 2nd edn. Oxford: Oxford University Press.
- McRee DE (1993) *Practical Protein Crystallography*. San Diego: Academic Press.
- Prestegard JH (1998) New techniques in structural NMR – anisotropic interactions. *Nature Structural Biology* **5**: 517–522.
- Rhodes R (1993) *Crystallography Made Crystal Clear*. San Diego: Academic Press.
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley.