

19.7. Nuclear magnetic resonance (NMR) spectroscopy

BY K. WÜTHRICH

19.7.1. Complementary roles of NMR in solution and X-ray crystallography in structural biology

X-ray diffraction in crystals and NMR in solution can both be used to determine the complete three-dimensional structure of biological macromolecules, and to date a significant number of globular protein structures have been independently determined in crystals and in solution (Billeter, 1992). Particularly detailed comparisons of the two states have been made for the α -amylase inhibitor Tendamistat, which also included solving the crystal structure by molecular replacement with the NMR structure (Braun *et al.*, 1989). The dominant impression is one of near-identity of the molecular architecture in solution and in single crystals, which holds for the polypeptide backbone as well as the core side chains.

Although the presently available results show that there is usually close coincidence between both the global molecular architecture and the detailed arrangement of the molecular core in corresponding X-ray and NMR structures of globular proteins, there is also extensive complementarity in the information that is accessible with the two methods: X-ray diffraction can provide the desired information for big molecules and multimolecular assemblies, whereas NMR structure determination is limited to smaller systems [recently introduced new experiments enable solution NMR measurements for molecular weights of 100 000 and beyond (Pervushin *et al.*, 1997; Riek *et al.*, 1999)]. NMR measurements in turn provide quantitative information on both very rapid motions on the subnanosecond timescale (Otting *et al.*, 1991; Peng & Wagner, 1992) and slower dynamic processes (Wüthrich, 1986) which are not manifested in the X-ray data. Examples of low-frequency intramolecular mobility are the ring flips of phenylalanine and tyrosine (Wüthrich & Wagner, 1975), exchange of interior hydration water molecules with the bulk solvent (Otting *et al.*, 1991), and interconversion of disulfide bonds between the *R* and *S* chiral forms (Otting *et al.*, 1993). NMR studies of amide proton exchange and *cis-trans* isomerization of Xxx—Pro peptide bonds (Wüthrich, 1976, 1986) afford additional insight into conformational equilibria in the protein core. Finally, in all instances where a biological macromolecule cannot be crystallized, NMR is currently the only method capable of providing a three-dimensional structure.

Overall, X-ray crystal structures and NMR solution structures provide qualitatively different information on the molecular surface. In the crystals, a sizeable proportion of surface amino-acid side chains are subject to similar packing constraints in protein–protein interfaces as the interior side chains in the protein core, and therefore they are rather precisely defined by the X-ray diffraction data. In NMR solution structures determined according to a standard protocol (Wüthrich, 1995), the surface is usually largely disordered. Surface disorder in NMR structures may in part arise from scarcity of nuclear Overhauser effect (NOE) distance constraints and packing constraints near the protein surface, but, in turn, scarcity of NOE constraints is often a direct consequence of dynamic disorder. Additional NMR experiments that are not part of a standard structure determination protocol can provide information needed for more detailed characterization of the molecular surface, but care must be exercised in the data analysis because of the presence of a multitude of equilibria between two or multiple transient local conformational states, of which the relative populations are usually not independently known.

19.7.2. A standard protocol for NMR structure determination of proteins and nucleic acids

An NMR structure determination involves sample preparation, NMR measurements, assignment of the NMR lines to individual atoms in the polymer chain, collection of conformational constraints, and structure calculation and refinement, where in present practice the sequence of steps usually corresponds to the flow diagram of Fig. 19.7.2.1. As is also indicated in Fig. 19.7.2.1, it is a special feature of protein structure determination by NMR that the secondary polypeptide structure, including the connections between individual segments of regular secondary structure, may be known early on from the data used for obtaining the resonance assignments, *i.e.* before the structure calculation is even started.

For the sample preparation, homogeneous macromolecular material is dissolved at about 1 mM concentration in 0.5 ml of water. The ionic strength, pH and temperature, and possibly the concentration of additives, may then be adjusted, for example, to ensure near-physiological conditions, or denaturing conditions *etc.* The NMR study will often include the preparation of compounds enriched with ^{15}N and/or ^{13}C , and possibly with ^2H (Kay & Gardner, 1997). Uniformly isotope-labelled recombinant proteins are routinely obtained by expression in *Escherichia coli* bacteria grown on minimal media. For RNA and DNA, isotope-labelling techniques are more involved, but labelled nucleic acids will also be commonly available in the future. Being able to work with solutions is generally considered to be a great asset of the NMR method, but there are also potential inherent difficulties. For example, in the course of an investigation it may be nontrivial to achieve identical solution conditions in different NMR samples of the same compound, the absence of which typically results in small chemical-shift differences that slow down the combined analysis of different NMR spectra.

The demands on NMR experiments for macromolecular structure determination are currently met by multidimensional NMR at high polarizing magnetic fields (Wüthrich, 1986; Ernst *et al.*, 1987; Cavanagh *et al.*, 1996). With increasing molecular size and concomitant increase of the number of NMR peaks, it becomes more and more difficult to resolve and assign the individual resonances. In heteronuclear three- or four-dimensional (3D or 4D) spectra recorded with compounds that are uniformly labelled with ^{15}N and/or ^{13}C , the peaks are spread out in a third and possibly

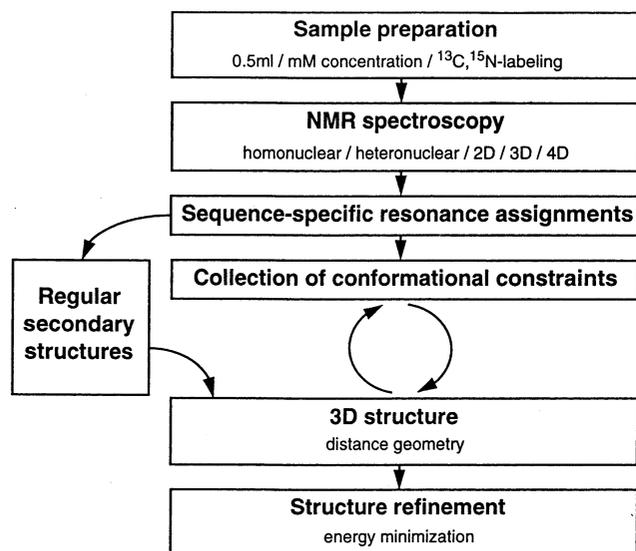


Fig. 19.7.2.1. Diagram outlining the course of a macromolecular structure determination by NMR in solution.

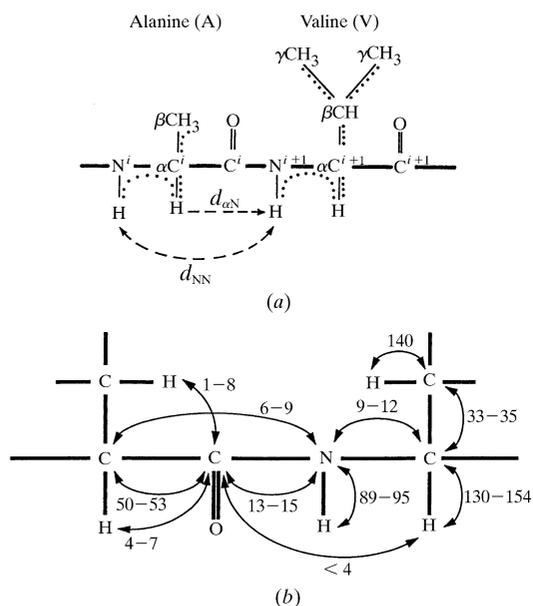


Fig. 19.7.2.2. (a) Sequential resonance assignment based on sequential ^1H - ^1H NOEs. In the dipeptide segment -Ala-Val- the dotted lines indicate ^1H - ^1H relations which can be established by scalar through-bond spin-spin couplings. The broken arrows connect pairs of protons in sequentially neighbouring residues, i and $i+1$, which are related by ^1H - ^1H NOEs that manifest short sequential distances $d_{\alpha\text{N}}$ (between αCH and the amide proton of the following residue) and d_{NN} (between the amide protons of neighbouring residues). (b) Segment of a polypeptide chain with indication of the scalar spin-spin couplings that provide the basis for obtaining sequential assignments by triple-resonance experiments with uniformly $^{13}\text{C}/^{15}\text{N}$ labelled proteins.

fourth dimension along the ^{15}N and ^{13}C chemical-shift axes (for a review, see Bax & Grzesiek, 1993). Additional labelling with ^2H results in slower spin-relaxation rates and hence improved sensitivity and spectral resolution. Although this approach provides ^{13}C and ^{15}N NMR information, which may be used to support the structure determination (*e.g.* Wishart *et al.*, 1991) and to provide supplementary information on dynamic features of the molecule studied, the key purpose of heteronuclear NMR experiments is to enhance the spectral resolution for studies of protons and thus to obtain the maximum possible number of ^1H NMR-based conformational constraints.

Resonance assignments in biopolymers are obtained using the sequential assignment strategy (Dubs *et al.*, 1979; Billeter *et al.*, 1982; Wagner & Wüthrich, 1982). Unique segments of two or several sequentially adjoining amino-acid residues are identified by NMR experiments and then attributed to discrete positions in the polypeptide chain by comparison with the chemically determined amino-acid sequence. The desired relations between protons in sequentially neighbouring amino-acid residues i and $i+1$ can be established by nuclear Overhauser effects (NOEs), manifesting a close approach between αCH_i and NH_{i+1} ($d_{\alpha\text{N}}$), NH_i and NH_{i+1} (d_{NN}) (Fig. 19.7.2.2a), and possibly βCH_i and NH_{i+1} ($d_{\beta\text{N}}$) (Wüthrich, 1986). For Xxx-Pro dipeptide segments, corresponding connectivities are observed with δCH_2 of Pro in the place of the amide proton. For small proteins with up to about 100 amino-acid residues, NOE-based sequential assignments can rely entirely on homonuclear ^1H NMR experiments, and for somewhat bigger proteins they can be established based on the improved resolution of 3D NMR experiments with ^{15}N -labelled proteins. No prior knowledge of the polypeptide conformation is needed, since at least one of the two distances $d_{\alpha\text{N}}$ or d_{NN} (Fig. 19.7.2.2a) is always sufficiently short to be observed by NOEs (Billeter *et al.*, 1982). A further attractive feature of this approach is that the identification of

the sequential NOEs forms an integral part of the data collection for the protein structure determination (see below). Sequential assignments can alternatively be obtained entirely *via* heteronuclear scalar couplings, using recombinant isotope-labelled proteins (Fig. 19.7.2.2b). Using 3D and 4D heteronuclear triple-resonance experiments, the resonance lines of sufficiently large mutually overlapping fragments of the polypeptide chain are grouped together to enable sequence-specific resonance assignments (for a review, see Bax & Grzesiek, 1993). With the implementation of transverse relaxation-optimized spectroscopy (TROSY) elements (Pervushin *et al.*, 1997) into triple-resonance experiments (Salzmann *et al.*, 1998), backbone resonance assignments *via* the spin-spin couplings of Fig. 19.7.2.2(b) can be performed with molecular weights of 100000 and beyond. For nucleic acids, assignment procedures were largely patterned after those used for proteins and have been used successfully for fragments with 40 nucleotides and beyond.

NOE upper-distance constraints contain the crucial information needed for macromolecular structure determination (Wüthrich, 1986, 1989). To obtain a high-quality structure, the maximum possible number of NOE conformational constraints must be collected as input for the structure calculation. This is accomplished by using the chemical-shift lists obtained as a result of the sequence-specific resonance assignments to attribute the cross peaks in 2D [^1H , ^1H]-NOESY spectra, or 3D and 4D heteronuclear-resolved [^1H , ^1H]-NOESY spectra, to distinct pairs of hydrogen atoms. As indicated in Fig. 19.7.2.1, this data collection is achieved in several cycles, where ambiguities in the NOESY cross-peak assignments can usually be resolved by reference to preliminary structures calculated from incomplete input data sets (Güntert *et al.*, 1993). In present practice, each individual NOE constraint has the format of an allowed distance range, which circumvents intrinsic difficulties that might arise from attempts at quantitative distance measurements, and which is also adjusted to account for possible effects from internal mobility. The lower limit is usually taken to correspond to the sum of two hydrogen atomic radii, *i.e.* 2.0 Å, and the NOE intensities are translated into corresponding upper bounds, typically in steps of 2.5, 3.0 and 4.0 Å. Supplementary conformational constraints, for example, from spin-spin coupling constants (Wüthrich, 1986), residual dipole-dipole couplings (Tjandra & Bax, 1997), pseudocontact shifts and relaxation effects near paramagnetic centres (Banci *et al.*, 1998) *etc.*, are represented in the input by similar allowed ranges, which account for the internal mobility of the 3D structures and the limited accuracy of the individual measurements. Initially, NMR structures were calculated using distance-geometry techniques, and subsequently the principles of distance geometry have been introduced into molecular-dynamics programs in Cartesian coordinates (Brünger *et al.*, 1986) or in torsion-angle space (Güntert *et al.*, 1997). Model calculations performed in conjunction with the initial protein structure determinations had shown that NMR structure calculation depends critically on the density of NOE distance constraints, while it is remarkably robust with regard to low precision of the individual distance constraints (Havel & Wüthrich, 1985). For the common presentation of an NMR structure, one considers the result of a single structure calculation as representing one molecular geometry that is compatible with the NMR data. To investigate further whether or not this solution is unique, the calculation is repeated with different boundary conditions, where for each calculation, convergence is judged by the residual constraint violations. All satisfactory solutions, by this criterion, are included in a group of conformers that is used to represent the NMR structure (Fig. 19.7.2.3). The precision of the structure determination is reflected by the dispersion among this group of conformers. In proteins, larger variations are typically observed near the chain ends, in exposed loops and for surface amino-acid side chains, which

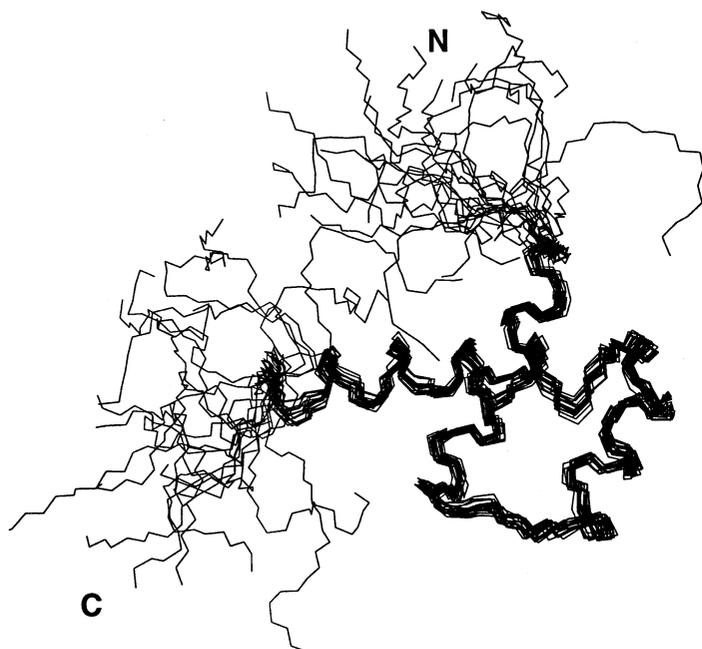


Fig. 19.7.2.3. Polypeptide backbone of 19 energy-refined conformers selected to represent the NMR solution structure of the *Antennapedia* homeodomain. From residues 7 to 59, the structure is well defined by the NMR data. The chain-terminal segments 0–6 and 60–67 are disordered, and additional NMR studies showed that these terminal segments behave like ‘flexible tails’. [Drawing prepared with the atomic coordinates from Qian *et al.* (1989).]

contrasts with the well defined core. For nucleic acids, the ‘global folds’, for example, formation of duplexes, triplexes, quadruplexes, or loops, can be well defined by NMR, but because of the short range of the NOE distance measurements, certain ‘long-range’ features, for example, bending of DNA duplexes, may be more difficult to characterize.

19.7.3. Combined use of single-crystal X-ray diffraction and solution NMR for structure determination

The chemical shifts in proteins or nucleic acids cannot be calculated with sufficiently high precision from the X-ray crystal structures to predict the NMR spectra reliably. This limits the extent to which improved efficiency of an NMR structure determination may be derived from knowledge of the corresponding crystal structure. Nonetheless, for resonance assignments with sequential NOEs, information on the type of connectivity to look for in specified polypeptide segments may be derived. The X-ray structure may also tentatively be used as a starting reference to resolve ambiguities in the assignment of NOE distance constraints (Güntert *et al.*, 1993), which may help to reduce the number of cycles needed to collect the input data for a high-quality structure determination (Fig. 19.7.2.1).

An attractive possibility for combined use of X-ray and NMR data has in the past arisen in situations where both a low-resolution crystallographic electron-density map and a secondary-structure determination by NMR are available. As mentioned before, NMR secondary-structure determination in proteins usually derives directly from the backbone resonance assignments (Wishart *et al.*, 1991; Wüthrich *et al.*, 1984) and does not depend on the availability of a complete three-dimensional structure. Since the covalent connections between regular secondary-structure elements can often also be unambiguously determined, this NMR information may be helpful in tracing the electron-density map for the determination of the corresponding X-ray crystal structure (Kallen

et al., 1991). Furthermore, since identical global architectures are usually found for corresponding crystal and solution structures of globular proteins, NMR structures have been used to solve the corresponding crystal structure by molecular replacement (*e.g.* Braun *et al.*, 1989). Considering the ease with which high-quality X-ray data are obtained nowadays once suitable crystals are available, it remains to be seen whether this kind of combined use of data obtained with the two methods will also play a role in the future.

19.7.4. NMR studies of solvation in solution

In NMR structures, the location of hydration water molecules is determined by the observation of NOEs between water protons and hydrogen atoms of the polypeptide chain. In contrast to the observation of hydration water in X-ray crystal structures, this information is not routinely collected within the scope of a standard protocol for NMR structure determination (Fig. 19.7.2.1), but requires additional experiments (Otting *et al.*, 1991). Because the NOE decreases with the sixth power of the ^1H – ^1H distance, only water molecules in a first hydration layer are typically observed. The NOE intensity is further related to a correlation function describing the stochastic modulation of the dipole–dipole coupling between the interacting protons, which may be governed either by the Brownian rotational tumbling of the hydrated protein molecule or by interruption of the dipolar interaction through translational diffusion of the interacting spins, whichever is faster (Otting *et al.*, 1991). Interior hydration waters are typically observed in identical locations in corresponding crystal and solution structures, but NMR provides additional information on molecular mobility. Completely buried hydration water molecules have thus been found to exchange with the bulk solvent at rates corresponding to millisecond residence times in the protein hydration sites, and measurements with nuclear magnetic relaxation dispersion also revealed exchange rates on the microsecond to nanosecond timescale (Denisov *et al.*, 1996). For surface hydration water, the lifetimes in the hydration sites are typically even shorter, and NMR measures the average duration of these ‘visits’ (Otting *et al.*, 1991; in an alternative analysis of the NMR data, surface hydration has been characterized by reduced diffusion rates of the water, without specifying individual hydration sites; Brüschweiler & Wright, 1994). Diffraction experiments, on the other hand, probe the total fraction of time that a water molecule spends in a particular hydration site, but they are insensitive to the residence time at that site on any particular visit. Furthermore, while hydration water molecules in protein crystals are observed in discrete surface sites that are not blocked by direct protein–protein contacts, the entire surface of a protein in aqueous solution is covered with water molecules. Overall, the NMR view of hydration in solution, which has also been rationalized with long-time molecular-dynamics simulations (*e.g.* Billeter *et al.*, 1996; Brunne *et al.*, 1993), is largely complementary to crystallographic data on hydration.

19.7.5. NMR studies of rate processes and conformational equilibria in three-dimensional macromolecular structures

Similar to the aforementioned hydration studies, information on intramolecular rate processes in macromolecular structures cannot usually be obtained from the standard protocol for NMR structure determination (Fig. 19.7.2.1), but results from additional experiments. The complementarity of such NMR information to crystallographic data is well illustrated by the ‘ring flips’ of phenylalanine and tyrosine (Wüthrich, 1986). The observation of these ring-flipping motions in the basic pancreatic trypsin inhibitor (BPTI)