arms of the cross in the B diffraction pattern. Arguments such as those outlined above, based on x-ray fiber diffraction patterns obtained by Maurice Wilkins and Rosalind Franklin, were used by Jim Watson and Francis Crick in the construction of their double-helical model for DNA.

**NMR methods use the magnetic properties of atomic nuclei**

Certain atomic nuclei, such as \(^1\)H, \(^13\)C, \(^15\)N, and \(^31\)P, have a magnetic moment or spin. The chemical environment of such nuclei can be probed by nuclear magnetic resonance (NMR) and this technique can be exploited to give information on the distances between atoms in a molecule. These distances can then be used to derive a three-dimensional model of the molecule. Most structure determinations of protein molecules by NMR have used the spin of \(^1\)H, since hydrogen atoms are abundant in proteins. Small proteins can be analyzed by \(^1\)H (proton) NMR but to study larger proteins and to obtain sufficient data to determine side chain conformations, it is necessary to introduce \(^13\)C and \(^15\)N into the protein. This is usually done by producing the protein in microorganisms grown in media enriched with these isotopes. NMR studies of proteins containing one of the isotopes are called 3-D NMR, and when both \(^13\)C and \(^15\)N are present they are called 4-D NMR.

When protein molecules are placed in a strong magnetic field, the spin of their hydrogen atoms aligns along the field. This equilibrium alignment can be changed to an excited state by applying radio frequency (RF) pulses to the sample. When the nuclei of the protein molecule revert to their equilibrium state, they emit RF radiation that can be measured. The exact frequency of the emitted radiation from each nucleus depends on the molecular environment of the nucleus and is different for each atom, unless they are chemically equivalent and have the same molecular environment (Figure 18.16a). These different frequencies are obtained relative to a reference signal and are called chemical shifts. The nature, duration, and combination of applied RF pulses can be varied enormously, and different molecular properties of the sample can be probed by selecting the appropriate combination of pulses.

**Figure 18.16** One-dimensional NMR spectra. (a) \(^1\)H-NMR spectrum of ethanol. The NMR signals (chemical shifts) for all the hydrogen atoms in this small molecule are clearly separated from each other. In this spectrum, the signal from the CH, protons is split into three peaks and that from the CH, protons into four peaks close to each other, due to the experimental conditions. (b) \(^1\)H-NMR spectrum of a small protein, the C-terminal domain of cellulase, comprising 36 amino acid residues. The NMR signals from many individual hydrogen atoms overlap and peaks are obtained that comprise signals from many hydrogen atoms. (Courtesy of Per Kraulis, Uppsala, from data published in Kraulis et al., Biochemistry 28: 7241–7257, 1989.)
In principle, it is possible to obtain a unique signal (chemical shift) for each hydrogen atom in a protein molecule, except those that are chemically equivalent, for example, the protons on the CH$_3$ side chain of an alanine residue. In practice, however, such one-dimensional NMR spectra of protein molecules (see Figure 18.16b) contain overlapping signals from many hydrogen atoms because the differences in chemical shifts are often smaller than the resolving power of the experiment. In recent years this problem has been bypassed by designing experimental conditions that yield a two-dimensional NMR spectrum, the results of which are usually plotted in a diagram as shown in Figure 18.17.

The diagonal in such a diagram corresponds to a normal one-dimensional NMR spectrum. The peaks off the diagonal result from interactions between hydrogen atoms that are close to each other in space. By varying the nature of the applied RF pulses these off-diagonal peaks can reveal different types of interactions. A COSY (correlation spectroscopy) experiment gives peaks between hydrogen atoms that are covalently connected through one or two other atoms, for example, the hydrogen atoms attached to the nitrogen and C$_\alpha$ atoms within the same amino acid residue (Figure 18.18a). An NOE (nuclear Overhauser effect) spectrum, on the other hand, gives peaks between pairs of hydrogen atoms that are close together in space even if they are from amino acid residues that are quite distant in the primary sequence (see Figure 18.18b).

**Figure 18.18** (a) COSY NMR experiments give signals that correspond to hydrogen atoms that are covalently connected through one or two other atoms. Since hydrogen atoms in two adjacent residues are covalently connected through at least three other atoms (for instance, HCo-C-NH), all COSY signals reveal interactions within the same amino acid residue. These interactions are different for different types of side chains. The NMR signals therefore give a "fingerprint" of each amino acid. The diagram illustrates fingerprints (red) of residues Ala and Ser. (b) NOE NMR experiments give signals that correspond to hydrogen atoms that are close together in space (less than 5 Å), even though they may be far apart in the amino acid sequence. Both secondary and tertiary structures of small protein molecules can be derived from a collection of such signals, which define distance constraints between a number of hydrogen atoms along the polypeptide chain.
Two-dimensional NMR spectra of proteins are interpreted by the method of sequential assignment

Two-dimensional NOE spectra, by specifying which groups are close together in space, contain three-dimensional information about the protein molecule. It is far from trivial, however, to assign the observed peaks in the spectra to hydrogen atoms in specific residues along the polypeptide chain because the order of peaks along the diagonal has no simple relation to the order of amino acids along the polypeptide chain. This problem has in principle been solved in the laboratory of Kurt Wüthrich in the ETH, Zurich, where the method of sequential assignment was developed.

Sequential assignment is based on the differences in the number of hydrogen atoms and their covalent connectivity in the different amino acid residues. Each type of amino acid has a specific set of covalently connected hydrogen atoms that will give a specific combination of cross-peaks, a “fingerprint,” in a COSY spectrum (see Figure 18.18a). From the COSY spectrum it is therefore possible to identify the H atoms that belong to each amino acid residue and, in addition, determine the nature of the side chain of that residue. However, the order of these fingerprints along the diagonal has no relation to the amino acid sequence of the protein. For example, when the fingerprint in one specific region of the COSY spectrum of the lac-repressor segment was assigned to a Ser residue, it was not known whether this fingerprint corresponded to Ser 16, Ser 28, or Ser 31 in the amino acid sequence.

The sequence-specific assignment, however, can be made from NOE spectra (see Figures 18.17 and 18.18b) that record signals from H atoms that are close together in space. In addition to the interactions between H atoms that are far apart in the sequence, these spectra also record interactions between H atoms from sequentially adjacent residues, specifically, interactions from the H atom attached to the main chain N of residue number i + 1 to H atoms bonded to N, Cα, and Cβ of residue number i (Figure 18.19a).

Figure 18.19 (a) Adjacent residues in the amino acid sequence of a protein can be identified from NOE spectra. The H atom attached to residue i + 1 (orange) is close to and interacts with (purple arrows) the H atoms attached to N, Cα, and Cβ of residue i (light green). These interactions give cross-peaks in the NOE spectrum that identify adjacent residues and are used for sequence-specific assignment of the amino acid fingerprints derived from a COSY spectrum. (b) Regions of secondary structure in a protein have specific interactions between hydrogen atoms in sequentially nonadjacent residues that give a characteristic pattern of cross-peaks in an NOE spectrum. In antiparallel β-sheet regions there are interactions between Cα-H atoms of adjacent strands (pink arrows), between N-H and Cα-H atoms (dark purple arrows), and between N-H atoms of adjacent strands (light purple arrows). The corresponding pattern of cross-peaks in an NOE spectrum identifies the residues that form the antiparallel β sheet. Parallel β sheets and α helices are identified in a similar way.
These signals in the NOE spectra therefore make it possible to determine which fingerprint in the COSY spectrum comes from a residue adjacent to the one previously identified. For example, in the case of the lac-repressor fragment the specific Ser residue that was identified from the COSY spectrum was shown in the NOE spectrum to interact with a His residue, which in turn interacted with a Val residue. Comparison with the known amino acid sequence revealed that the tripeptide Ser-His-Val occurred only once, for residues 28-30.

In practice, it is difficult to make unique assignments for longer pieces than di- or tri-peptides, since NOE signals also occur between residues close together in space but far apart in the sequence. Therefore, the peptide segments that have been uniquely identified by NMR are usually matched with corresponding segments in the independently determined amino acid sequence of the protein. Thus knowledge of the amino acid sequence is just as essential for the correct interpretation of NMR spectra as it is for the interpretation of electron-density maps in x-ray crystallography. Whereas x-ray crystallography directly gives an image of a three-dimensional model of the protein molecule, NMR spectroscopy identifies H atoms in the protein that are close together in space, and this information is then used to derive, indirectly, a three-dimensional model of the protein.

**Distance constraints are used to derive possible structures of a protein molecule**

The final result of the sequence-specific assignment of NMR signals, preferably done using interactive computer graphics, is a list of distance constraints from specific hydrogen atoms in one residue to hydrogen atoms in a second residue. The list contains a large number of such distances, which are usually divided into three intervals within the region 1.8 Å to 5 Å, depending on the intensity of the NOE peak. This list immediately identifies the secondary structure elements of the protein molecule because both α-helices and β-sheets have very specific sets of interactions of less than 5 Å between their hydrogen atoms (see Figure 18.19b). It is also possible to derive models of the three-dimensional structure of the protein molecule. However, usually a set of possible structures rather than a unique structure (Figure 18.20) is obtained, each of the possible structures obeying the distance constraints equally well. The sets of possible structures, which are frequently seen in NMR articles, do not, therefore, represent different actual conformations of a protein molecule present in solution. Rather they are simply different structures that are compatible with data obtained by current methods. The primary source of this ambiguity is an insufficient number of measured distance constraints. Because of this ambiguity, the accuracy of an NMR structure is not constant over the whole molecule and is also difficult to quantify.

In addition to the problem of ambiguity, there are other limitations to the use of NMR methods for the determination of protein structures. The most severe concerns the size of the protein molecules whose structures can be determined. Currently, the upper limit is molecules with molecular weights of around 25 kDa, but this limit will be increased in the future by using improved methods and equipment. Furthermore, the method requires highly concentrated protein solutions, on the order of 1–2 mM, with the additional requirement that the protein molecules must not aggregate at these concentrations. In addition, the pH of the solution should be lower than about 6 for proton NMR experiments. The exchanges of the NH protons in the main chain become so fast at higher pH that it is very difficult to observe them with NMR, and the signals from these hydrogen atoms are essential for the sequential assignment procedure.

How well do NMR-derived structures agree with those determined by x-ray methods? The structures of some different globular proteins that have been independently obtained by the two methods—such as bovine pancreatic...
trypsin inhibitor (see Figure 2.14a), plastocyanin (see Figure 2.11c) and thioredoxin from E. coli (see Figure 2.7)—show that NMR and x-ray crystallography give nearly identical results. The minor differences that exist are of the same order of magnitude as usually seen between x-ray structures of unrelated crystal forms of the same protein or determinations made under different experimental conditions. In other words, they are mostly small differences in loop regions of the main chain and different conformations of exposed side chains.

The situation is different for other examples—for example, the peptide hormone glucagon and a small peptide, metallothionein, which binds seven cadmium or zinc atoms. Here large discrepancies were found between the structures determined by x-ray diffraction and NMR methods. The differences in the case of glucagon can be attributed to genuine conformational variability under different experimental conditions, whereas the disagreement in the metallothionein case was later shown to be due to an incorrectly determined x-ray structure. A re-examination of the x-ray data of metallothionein gave a structure very similar to that determined by NMR.

NMR and x-ray crystallography are in many respects complementary. X-ray crystallography deals with the structure of proteins in the crystalline state, while NMR determines the structure in solution. The time scales of the measurements are different: NMR is more suitable for investigation of various dynamic processes such as those during folding, while x-ray crystallography is more suitable for characterization of protein surfaces and the water structure around the protein. X-ray crystallography remains the only method available to determine the structure of large protein molecules, whereas NMR is the method of choice for small protein molecules that might be difficult to crystallize.

**Biochemical studies and molecular structure give complementary functional information**

Our current knowledge of the relation between structure and function of protein molecules is insufficient to deduce the function of a protein from its structure alone, although, as we have seen, structural homology with proteins of known function can sometimes allow this. It is necessary to combine biochemical studies with structural information. Biochemical and cell biological studies can tell us if a protein is a receptor, a transport molecule, or an enzyme and, in addition, which ligands can bind to it, as well as the functional effects of such ligand binding. Studies of the three-dimensional structure of complexes between specific ligands and the protein will then give detailed information on how the active site is constructed and which amino acid residues are involved in ligand binding. Examples that we have described include protein–DNA interaction in Chapters 8, 9, and 10, sugar binding to a sugar transport protein in Chapter 4, and binding of inhibitors to enzymes that cleave peptide bonds in Chapter 11.

The specific role of each amino acid residue for the function of the protein can be tested by making specific mutations of the residue in question and examining the properties of the mutant protein. By combining in this way functional studies in solution, site-directed mutagenesis by recombinant DNA techniques, and three-dimensional structure determination, we are now in a position to gain fresh insights into the way protein molecules work.

**Conclusion**

The three-dimensional structure of protein molecules can be experimentally determined by two different methods, x-ray crystallography and NMR. The interaction of x-rays with electrons in molecules arranged in a crystal is used to obtain an electron-density map of the molecule, which can be interpreted in terms of an atomic model. Recent technical advances, such as powerful computers including graphics work stations, electronic area detectors, and
very strong x-ray sources from synchrotron radiation, have greatly facilitated the use of x-ray crystallography.

Crystallization of proteins can be difficult to achieve and usually requires many different experiments varying a number of parameters, such as pH, temperature, protein concentration, and the nature of solvent and precipitant. Protein crystals contain large channels and holes filled with solvents, which can be used for diffusion of heavy metals into the crystals. The addition of heavy metals is necessary for the phase determination of the diffracted beams.

X-ray structures are determined at different levels of resolution. At low resolution only the shape of the molecule is obtained, whereas at high resolution most atomic positions can be determined to a high degree of accuracy. At medium resolution the fold of the polypeptide chain is usually correctly revealed as well as the approximate positions of the side chains, including those at the active site. The quality of the final three-dimensional model of the protein depends on the resolution of the x-ray data and on the degree of refinement. In a highly refined structure, with an R value less than 0.20 at a resolution around 2.0 Å, the estimated errors in atomic positions are around 0.1 Å to 0.2 Å, provided the amino acid sequence is known.

Biological fibers, such as can be formed by DNA and fibrous proteins, may contain crystallites of highly ordered molecules whose structure can in principle be solved to atomic resolution by x-ray crystallography. In practice, however, these crystallites are rarely as ordered as true crystals, and in order to locate individual atoms it is necessary to introduce stereochemical constraints in the x-ray analysis so that the structure can be refined by molecular modeling.

In NMR the magnetic-spin properties of atomic nuclei within a molecule are used to obtain a list of distance constraints between those atoms in the molecule, from which a three-dimensional structure of the protein molecule can be obtained. The method does not require protein crystals and can be used on protein molecules in concentrated solutions. It is, however, restricted in its use to small protein molecules.

**Selected readings**


