STUDIES OF THE MOLECULAR CONFORMATIONS IN PROTEINS BY ¹H AND ¹³C N.M.R. SPECTROSCOPY

K. WÜTHRICH

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule, Zürich, Switzerland

ABSTRACT

A brief survey of the application of high-resolution proton and carbon-13 n.m.r. spectroscopy for the investigation of the molecular conformations in proteins is presented. The first part of the paper recounts some basic relationships between structure and conformations of proteins and their n.m.r. spectral properties. These are then illustrated with recent experiments including studies of the basic pancreatic trypsin inhibitor, cytochrome b_5 and b_2 and isolated heme groups.

INTRODUCTION

In favourable cases high-resolution n.m.r. spectroscopy can yield data on the molecular conformations in biopolymers, in particular peptides and proteins, which would otherwise be accessible only by single crystal x-ray studies. It is a particular advantage of the n.m.r. method that the application of a magnetic field and irradiation in the radiofrequency range appear to cause only very small perturbations in biological systems, and that proteins can in principle be studied under near-physiological conditions of solvent and temperature. Whereas these potential advantages of n.m.r. for biological research had long been anticipated[1], experimental realization became attractive only during the last few years when superconducting magnets and the Fourier transform technique[2] were introduced into commercial spectrometers. In this paper some basic relationships between molecular structure and n.m.r. spectral properties of proteins are illustrated with recent data obtained from ¹H n.m.r. experiments at 220 MHz and Fourier transform ¹³C n.m.r. studies at 25·14 MHz.

GENERAL ASPECTS

The covalent structure of peptides and proteins consists of a linear array of amino acid residues which are linked by peptide bonds. This is schematically shown in *Figure 1*. There are twenty common amino acids with different side chains R on the α-carbon atom, which are either aliphatic hydrocarbons, or carry various functional groups including different aromatic rings. The amino acid sequences have been determined by chemical methods for a large number of peptides and proteins. For any given amino acid sequence a large number of three-dimensional molecular conformations can in principle be formed by variation of the torsion angles about the single bonds[3]. Since
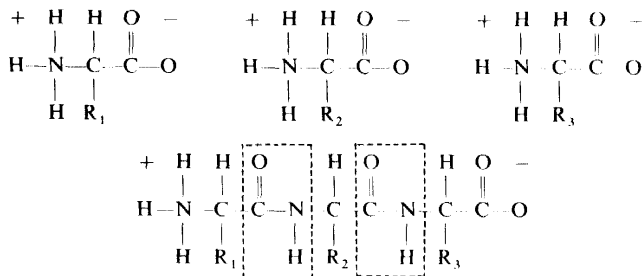
Figure 1. The top row shows three individual amino acid residues which differ in the side chains R_i on the α-carbon atom. Below, the three amino acids have been joined in a tripeptide. For considerations of the peptide conformations the peptide units (dashed squares) are generally assumed to be planar. Different conformations of the peptide backbone can then be uniquely characterized by the torsion angles about the single bonds linking the peptide units with the α-carbon atoms[3]

different molecular conformations are often quite clearly manifested in the n.m.r. spectra, they can in favourable cases be characterized on the basis of the n.m.r. data. The investigation of the molecular conformations by n.m.r. and other methods is of great interest because they are intimately related with the biological roles of the proteins.

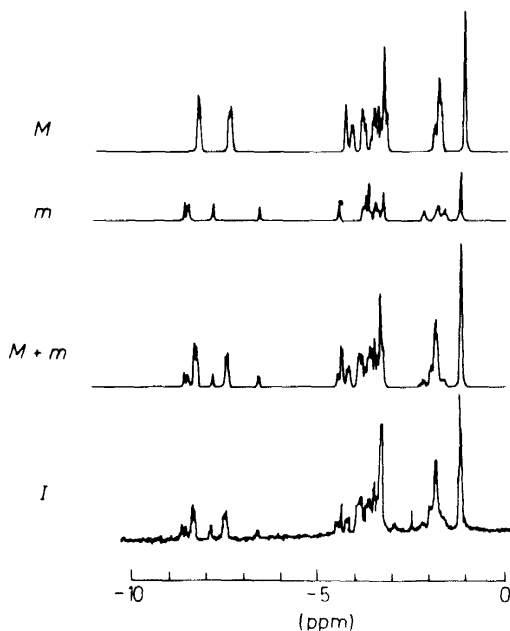To illustrate some n.m.r. spectral properties of small peptides, *Figure 2*

Figure 2. Proton n.m.r. spectrum at 220 MHz of *cyclo*-Glycyl-L-alanyl-glycyl-glycyl-L-prolyl in d_6-DMSO at 22 C. Two different conformations M and m of the peptide are present under these conditions. Chemical shifts are in ppm from internal TMS, where shifts to lower field are indicated by negative numbers: *I*, experimental spectrum, two resonances at -2.5 and -3.3 ppm correspond to the residual protons in d_6-DMSO and to H_2O; $M + m$, simulated spectrum consisting of the sum of the spectra M and m for the individual species

shows the ¹H n.m.r. spectrum of the cyclic pentapeptide *cyclo*-Glycyl-L-alanyl-glycyl-glycyl-L-prolyl[4]. In a solution in dimethyl sulphoxide at ambient temperature there are two different conformations of this peptide, the relative concentrations being approximately 2:1. The n.m.r. spectra of these two species can readily be recognized in the experimental spectrum *Figure 2 I*, and have been individually simulated in *Figure 2, M* and *m*. As is quite generally possible in peptides of this size, the resonances between -1 and -5 ppm, which correspond to the protons bound to carbon atoms, could be individually assigned to the different types of amino acid residues. In this spectral region the relative chemical shifts between corresponding resonances in the species *M* and *m* are mostly of the order 0.1 ppm, and in no case exceed 0.5 ppm[4]. The manifestation of the different molecular conformations is more pronounced in the chemical shift differences of the amide proton resonances between -6 and -9 ppm, and in their dependence upon temperature. These marked differences come from the formation of different transannular hydrogen bonds in the two conformations *M* and *m*, which shield different amide protons from interactions with the solvent. Additional data on the two types of peptide conformations come from the spin–spin coupling constants between the amide and C_α-protons (*Figure 1*), which are related to the torsion angles about the $N-C_\alpha$ bonds by a Karplus-type relationship[5]. *Figure 3* shows that the two
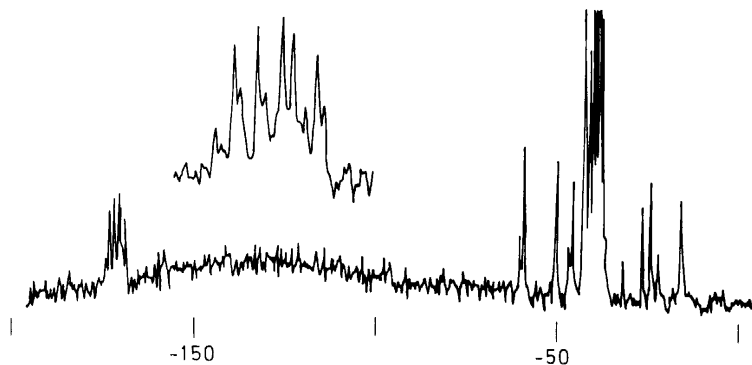

Figure 3. Proton noise-decoupled natural abundance ¹³C Fourier transform n.m.r. spectrum at 25.14 MHz of a 0.3 M solution of *cyclo*-Glycyl-L-alanyl-glycyl-glycyl-L-prolyl in d₆-DMSO, $T = 28°C$. Chemical shifts are in ppm from internal TMS, where shifts to low field are assigned negative values. The septet centred at -39.8 ppm corresponds to d₆-DMSO. The carbonyl region from -165 to -175 ppm is also shown in an expanded scale. Different resonances for the two peptide conformations *M* and *m* (*Figure 2*) are observed for all the carbon atoms except the α-carbons of glycine. Resonance assignments: alanyl C_α -50 ppm, C_β -16 ppm, glycine C_α -43 ppm, proline C_α -60 ppm, C_δ -46 ppm, C_β and C_γ (*M*) -24 and -27 ppm, C_β and C_γ (*m*) -22 and -32 ppm

conformations of this peptide are also manifested in chemical shift differences of the ¹³C resonances[6].

Knowledge of the n.m.r. spectral properties of the individual amino acid residues, obtained from studies of small peptides, is essential for the investigation of proteins, which consist typically of polypeptide chains with one to several hundred amino acid residues. In molecules of this size, the ¹H and

[13]C n.m.r. spectra are rather poorly resolved even at the highest presently available magnetic fields. A spectral analysis based on the observation of the polypeptide backbone resonances, as in the cyclic peptide mentioned above, is then no longer possible. Instead one has to look for particular spectral features which may arise from the close proximity of certain amino acid side chains in the globular native conformations of the proteins. This is illustrated by the [1]H n.m.r. spectra of the basic pancreatic trypsin inhibitor in *Figure 4*[7].
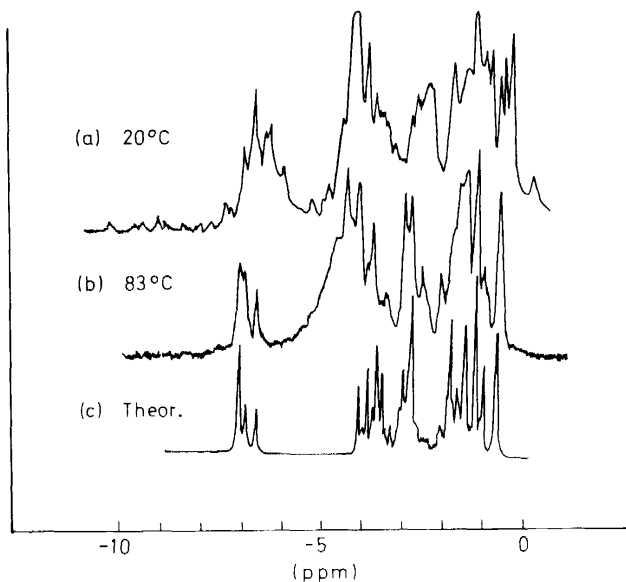

Figure 4. Proton n.m.r. spectra at 220 MHz of the basic pancreatic trypsin inhibitor in a 0.01 M solution in D_2O containing 6 M guanidinium chloride: (*a*) native protein at 20°C: (*b*) denatured protein at 83°C. (*c*) Hypothetical spectrum computed from the amino acid spectra of McDonald and Phillips[8] as described in the text. In the spectra (*a*) and (*b*) the residual protons of the solvent are at −4.5 ppm

The trypsin inhibitor is a small protein which consists of one polypeptide chain with 58 amino acid residues. A hypothetical spectrum (*Figure 4c*) of the molecule in a solution in D_2O, where all the exchangeable protons would be replaced by deuterium, had been computed with the assumption that the resonances of the amino acid side chains in the protein are identical with those observed in the individual amino acids. In the spectral regions from 0 to −3 and −6 to −8 ppm, where the amino acid side chain resonances occur, this hypothetical spectrum is quite similar to that of the denatured protein in *Figure 4b*. On the other hand, the spectrum of the native form of the protein in *Figure 4a* is rather markedly different. In particular, there appear resolved resonances at 0 ppm, between −5 and −6 ppm, and between −7 and −11 ppm. *Figure 5* shows that a corresponding hypothetical [13]C n.m.r. spectrum of the trypsin inhibitor (*Figure 5b*) corresponds again quite closely to that observed for the denatured protein (*Figure 5c*), whereas the resonances of the native molecule are obviously different. In particular, the dispersion of
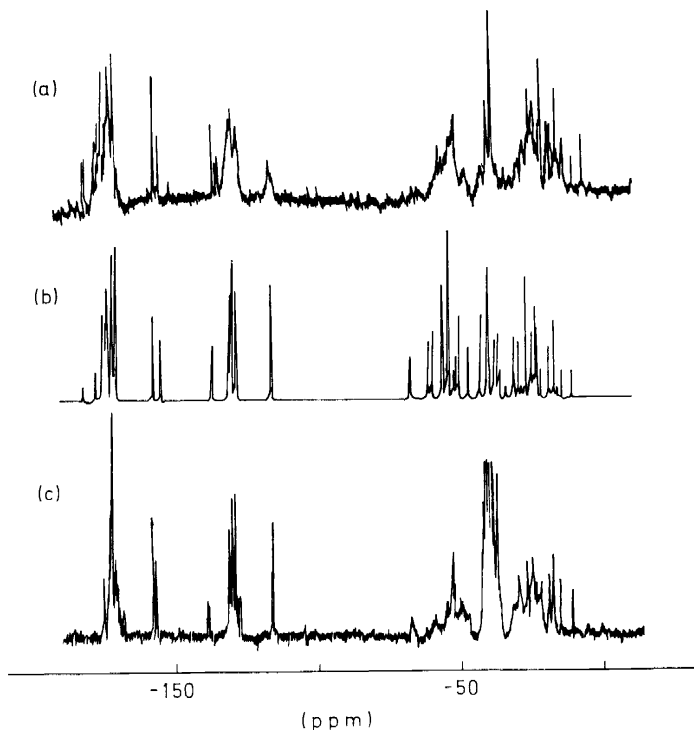
238

Figure 5. Proton noise-decoupled natural abundance ¹³C Fourier transform n.m.r. spectra at 25.14 MHz of the basic pancreatic trypsin inhibitor: (*a*) Native protein in D_2O solution: (*b*) hypothetical spectrum computed from the amino acid spectra as described in the text: (*c*) denatured protein in DMSO solution

chemical shifts among different amino acid residues of a given type is here readily apparent.

The observations in *Figures 4* and *5* can be qualitatively interpreted by the following considerations. The denatured protein is in a random coil confor-mation, i.e. in a generally rather extended form of the polypeptide backbone with the amino acid side chains extending freely into the solvent, as they would in the individual amino acids. In this form of the molecule, all the amino acid side chains of a given type R_i (*Figure 1*) are to a good approximation magneti-cally equivalent[8]. In the biologically functional native conformations of globular proteins the polypeptide chains are uniquely arranged in space, as illustrated in *Figure 6* for myoglobin[9]. Many of the interior parts of these structures are not accessible to the solvent. This is clearly evidenced in the trypsin inhibitor, where approximately 15 amide protons, which can be ob-served between -7.5 and -11 ppm in *Figure 4a*, are only very slowly exchanged in D_2O[7]. On the other hand, there are numerous close contacts even between amino acid side chains which may be far apart in the primary structure. Hence many of the amino acids are in unique environments in the native protein, and this is in turn manifested in unique n.m.r. chemical shifts[10]. From the previously mentioned observations in a cyclic peptide (*Figures 2*
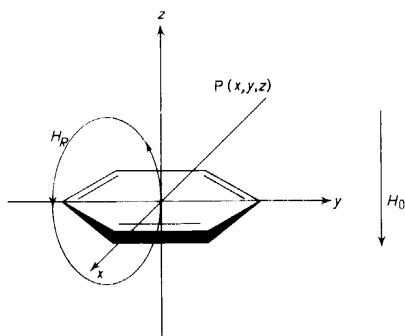
239

Figure 6. x-carbon diagram of the myoglobin molecule obtained from 2 Å x-ray analysis in a single crystal by Kendrew *et al.*[9] The heme group side chains are identified by: M = methyl. V = vinyl, P = propionic acid (see *Figure 11*). (Reproduced with permission from Ref. 9a)

and 3)[4, 6], one would anticipate that through the formation of a globular protein conformation essentially all the nuclear resonances are shifted away from their positions in the random coil form by several tenths of a ppm. That this is actually so can be readily seen in the spectral regions from -1 to -5, and from -6 to -8 ppm of the proton n.m.r. spectra of the trypsin inhibitor (*Figure 4*). However, with the limited spectral resolution obtained for macromolecules, most of the spectral interpretations have so far been concentrated on a relatively small number of resonances which would be particularly strongly affected by the molecular conformations.

In numerous proton n.m.r. studies, the spectral features arising from the anisotropic diamagnetic susceptibility in aromatic rings have been particularly useful[10, 11]. Aromatic rings occur in amino acid side chains and in the prosthetic groups of certain conjugated proteins, where the heme group of myoglobin in *Figure 6* is a typical example. As is schematically shown in *Figure 7*, the resonances of the nuclei located near an aromatic ring, can be shifted upfield or downfield by the local ring current field. For a given ring these ring current shifts depend on the relative positions of the observed protons and the aromatic group, and hence are sensitive to changes of the protein conformation. Ring current shifts due to the aromatic amino acids phenylalanin, tyrosine and tryptophan can be as large as 2 ppm[10], and considerably larger shifts can arise from the proximity to certain prosthetic groups[11]. As will be illustrated below by some experiments with cytochromes,

Figure 7. The local magnetic ring current field H_R of an aromatic molecule. H_0 is the external polarizing field. The field strength at a point P is determined by the position relative to the aromatic ring and by the size of H_R. H_R depends on the total number of π electrons in the aromatic molecule

ring current shifted resonances can thus appear as well-resolved lines in unusual positions even in the spectra of rather large proteins, and can be used as natural probes for studies of the protein conformations[10, 12]. In some proteins individual resonances of hystidyl[13] and tryptophanyl[14] residues could also be identified from their dependence on the solvent, and then used for studies of biological problems. In paramagnetic proteins, the resonances of amino acid residues located near the paramagnetic centre can be considerably shifted by dipole–dipole coupling. These pseudo-contact shifts depend strongly on the molecular conformation, and can thus also serve as probes for the investigation of the protein structure[15]. Overall one can say that the regions in the proteins which have so far been accessible to more detailed ^{1}H n.m.r. studies are often found to be essential for the biological function of the molecule. For example, aromatic amino acids and prosthetic groups are located in the active sites of numerous globular proteins.

^{13}C n.m.r. studies of macromolecules have only more recently become practical, and so far rather few detailed analyses of the chemical shifts in proteins have been presented. On the other hand, studies of the ^{13}C spin-lattice relaxation times have provided very interesting data on the segmental motions of native and denatured proteins[16]. ^{13}C spin relaxation studies appear to be a suitable method for the investigation of molecular dynamics also in more complex biological material, e.g. in biological membranes[17].

^{1}H n.m.r. studies of cytochromes b_5 and b_2

Cytochrome b_5 is a hemoprotein found in the microsomal fraction of liver tissue homogenates. In its biological function it appears to be an electron transport protein[18]. The cytochrome b_5 used for these studies, has a molecular weight of *ca.* 12 000, and consists of one polypeptide chain with 93 amino acid residues and one protoheme IX group (*Figure 11*). The amino acid sequence and the molecular conformation in single crystals of the protein are known[19].

The proton n.m.r. spectrum of the diamagnetic reduced cytochrome b_5 is shown in *Figure 8*. Outside the spectral region from 0 to -10 ppm, which contains the bulk of the resonances, there are five methyl resonances at 1.3, 1.1,
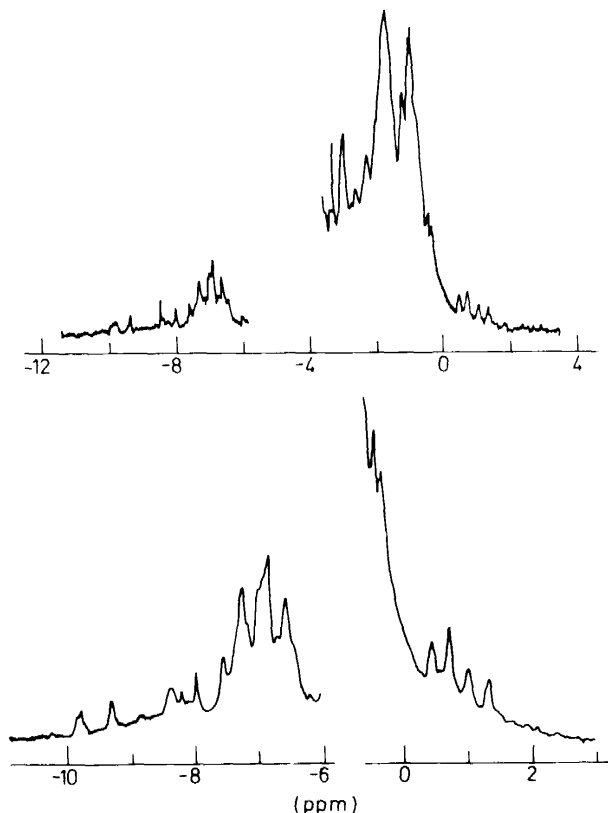
241

Figure 8. Proton n.m.r. spectrum at 220 MHz of a *ca.* 0.008 M solution of ferrocytochrome b_5 in deuterated 0.2 M phosphate buffer, pD = 7.0 at 20 C. The spectral region from −4 to −6 ppm, which contains the resonance of HDO and its spinning sidebands, has been omitted. The spectral regions from −1 to 3 ppm, and from −6 to −10 ppm are also shown on an expanded scale

0.7, 0.7 and 0.5 ppm. From a comparison with a spectrum computed with the single crystal atomic coordinates and with empirical values for the ring current fields in the different aromatic rings, it could be shown that four of these resonances correspond to methyl groups of aliphatic amino acid side chains located near the heme group[20]. The line at 1.1 ppm corresponds to an aliphatic methyl group located near the only tryptophanyl residue in the molecule. The comparison of the experimental spectrum with that computed from the x-ray data led also to the conclusion that the molecular conformations of ferrocytochrome b_5 in single crystals and in solution are quite similar.

The spectrum of the oxidized cytochrome b_5 in *Figure 9* contains more than 25 resolved resonance lines in the spectral regions from −10 to −30 and 0 to 10 ppm. It could be shown that most of these unusual resonance positions come from interactions with the unpaired electron of the low-spin ferric heme iron (Fe^{3+}, $S = 1/2$)[20]. All the lines between −10 and −30 ppm, and several
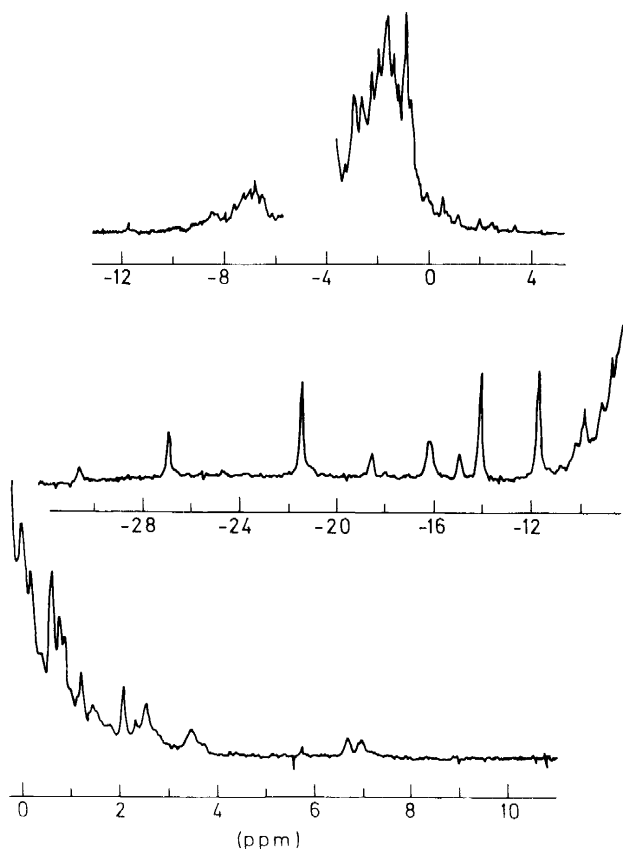
242

Figure 9. 220 MHz proton n.m.r. spectrum of 0.008 M ferricytochrome b_5 in 0.2 M deuterated phosphate buffer, pD 6.3, at 29°C. The three spectral regions are represented with different horizontal and vertical scales. (The resonance intensities of the resolved lines in the regions -10 to -30 and 0 to 10 ppm correspond to 1 to 3 protons, whereas the spectral region from 0 to -10 ppm contains the resonances of approx. 500 protons)

high-field resonances correspond to protons of the heme group (*Figure 11*), which will be further discussed below. A line at 1.2 ppm corresponds to the methyl group near the indole ring of tryptophan which had already been observed in the reduced protein. Eight methyl resonances at high field from 0 ppm were found to come from amino acids near the heme group, which are shifted to high field by dipole–dipole coupling with the electron spin. From the single crystal atomic coordinates[19] and the available data on the electronic *g*-tensor[21], these resonances were assigned to specific amino acid residues. The resonance assignments in the high-field regions of reduced and oxidized cytochrome b_5 are summarized in *Figure 10*[20], which shows a plot of the resonance positions versus reciprocal of temperature. The observed temperature dependences of the resonance positions in the oxidized protein have been extrapolated to $1/T = 0$ (solid lines). The resonance positions in
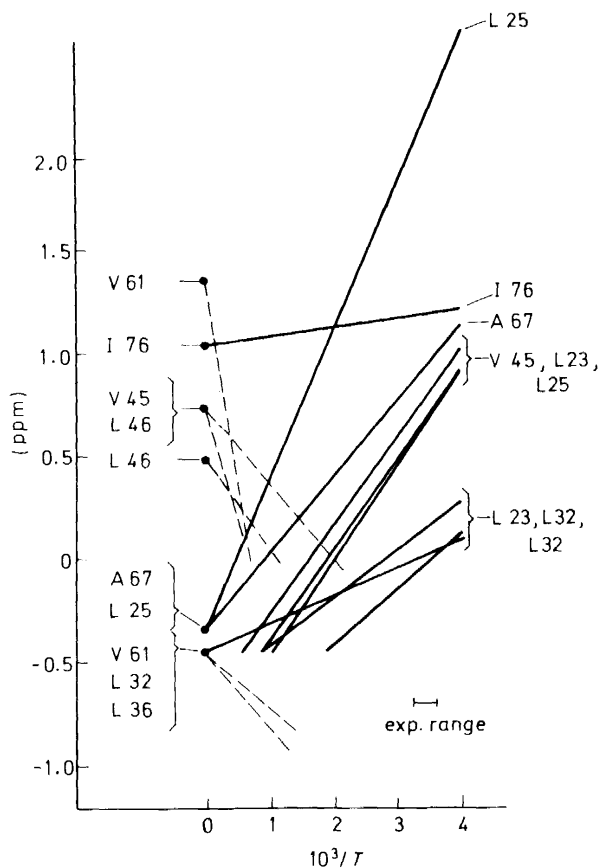
243

Figure 10. Dependence on the reciprocal of temperature and assignments of the high-field methyl resonances in ferric and ferrous cytochrome b_5 (V = valine, I = isoleucine, L = leucine, A = alanine). The points on the left indicate the observed resonance positions in the reduced protein (*Figure 8*). The dashed lines indicate the dependence on the reciprocal of temperature for these resonances in the ferric protein, which were obtained from theoretical considerations. The solid lines show the extrapolation to $1/T = 0$ of the measured temperature dependence of the high-field methyl resonances in the oxidized protein

the diamagnetic reduced protein are plotted at $1/T = 0$, and their approximate dependence on temperature (dashed lines) was obtained from theoretical considerations. It is seen that in this case the ring current field and the pseudo-contact field of the heme have opposite signs for the protons considered, and hence the resonances of different amino acid residues are shifted to high-field positions in the two oxidation states of the protein. In its effect on the proton resonances the heme group in oxidized cytochrome b_5 behaves like a typical 'paramagnetic shift reagent'. The resonance of isoleucine-76, which is located near tryptophan-22 and approximately 17 Å away from the heme iron, is essentially not affected by the paramagnetic centre.

These n.m.r. data on cytochrome b_5 now make it possible to study confor-

mational changes in well-defined regions of this molecule under different conditions. Recently it was also possible on the basis of the cytochrome b_5 data to determine by n.m.r. certain detailed structural features in cytochrome b_2, where neither the complete amino acid sequence nor the single crystal atomic coordinates have as yet been worked out[22].

¹³C n.m.r. studies of the electronic states in paramagnetic heme groups

Spectral interpretations, of the type discussed in the foregoing section, depend to a great extent on the availability of accurate data on the local magnetic fields in the constituent groups of the protein. In this context, additional investigations of the electronic states of the heme groups in hemoproteins are needed[15]. These experiments should further provide additional data for studies of the relationships between the electronic structures of the hemes and the biological roles of the hemoproteins[11].

The proton n.m.r. spectrum of a paramagnetic low-spin ferric complex of protoporphyrin IX, which is the heme group of cytochrome b_5[18, 19] and many other proteins, is shown in *Figure 11*. Since the proton resonances are well
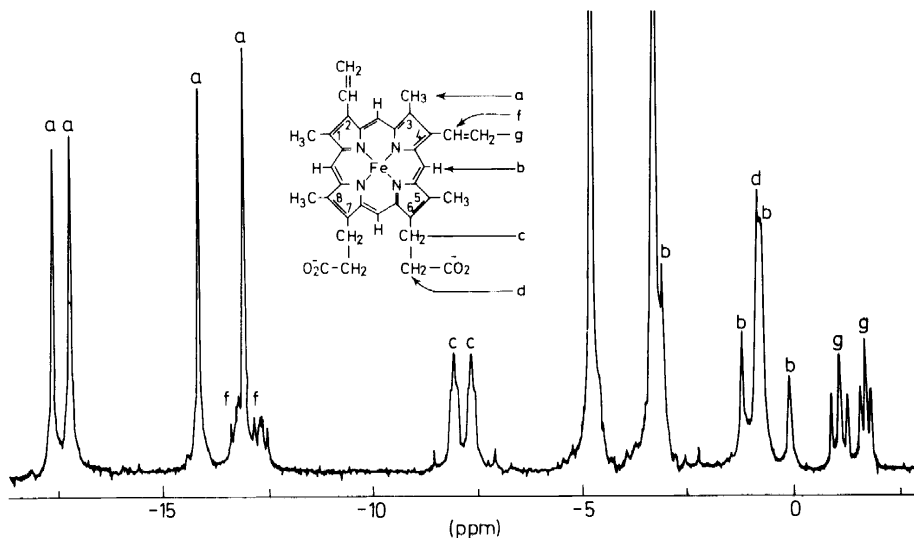

Figure 11. Fourier transform proton n.m.r. spectrum at 100 MHz of the dicyanide complex of iron(III) protoporphyrin IX in CD_3OD, $T = 29°C$. The structure of the complex, where the axial cyanide ligands have been omitted, and the resonance assignments are also indicated

separated by the hyperfine shifts, most of the carbon-13 resonances could be identified in a heteronuclear proton–carbon-13 double resonance experiment with off-resonance proton irradiation at $+5$ ppm[23]. Three carbon-13 spectra are shown in *Figure 12*. In the proton noise-decoupled spectrum (*Figure 12a*) and in the spectrum *Figure 12(c)*, which was recorded without proton irradiation. all the resonances expected from the structure of protoporphyrin IX are observed. In the off-resonance experiment (*Figure 12b*) the different methylene and methine carbons are readily identified from the different residual spin–spin

Figure 12. Fourier transform ^{13}C n.m.r. spectra at 25.14 MHz of the dicyanide complex of iron(III) protoporphyrin IX (*Figure 11*) in CD_3OD, $T = 29$ C. The strong resonance at -49 ppm comes from the solvent: (*a*) proton noise decoupled: (*b*) off-resonance proton irradiation at $+5$ ppm (see *Figure 11*): (*c*) no proton irradiation. The resonance assignments (see *Figure 11* for the nomenclature) which resulted from these and additional experiments[23, 24] are also given

couplings. Complete assignment of the carbon-13 resonances has so far been achieved for several porphyrin complexes with iron(III) and zinc(II)[23,24]. From the combined data on the proton and carbon-13 hyperfine shifts a

rather detailed description of the electronic structures in these complexes can then in principle be obtained through the Karplus–Fraenkél relationships[25].

CONCLUSION

The foregoing discussion has mainly dealt with systems where Nature has provided sufficient spectral resolution for the n.m.r. method to be applicable. In the future it is to be expected that biological systems will generally become more easily accessible for n.m.r. studies through isotopic labelling either by biological or by chemical procedures[26,27]. However, much of the future developments will also depend on further improvements of apparatus and methods. Therefore it is to be hoped that the commendable policy of the organizers of this conference, to bring together researchers working on theoretical and instrumental aspects of n.m.r. and on its applications in biology, will also be adopted at future meetings.

ACKNOWLEDGEMENTS

REFERENCES

[1] O. Jardetzky and C. D. Jardetzky, *Methods of Biochemical Analysis*, Vol. IX. p. 235. Interscience, New York (1962).

[2] R. R. Ernst, *Adv. Magn. Resonance* 2, 1 (1966).

[3] R. E. Dickerson and I. Geis, *The Structure and Action of Proteins*. Harper and Row, New York (1971).

[4] J. P. Meraldi, R. Schwyzer, A. Tun-Kyi and K. Wüthrich, *Helv. Chim. Acta*. 55, 1962 (1972).

[5] V. F. Bystrov, S. L. Portnova, V. L. Tsetlin, V. T. Ivanov and Y. A. Ovchinnikov, *Tetrahedron*. 25, 493 (1969).

[6] K. Wüthrich, A. Tun-Kyi and R. Schwyzer, *FEBS Letters*. 25, 104 (1972).

[7] A. Masson and K. Wüthrich, *FEBS Letters*, 31, 114 (1973).

[8] C. C. McDonald and W. D. Phillips, *J. Am. Chem. Soc.*, 91, 1513 (1969).

[9] (a) R. E. Dickerson. *The Proteins*. 2nd ed., Vol. 2, p. 603. Academic Press, New York (1964). (b) J. C. Kendrew, *Science*, 139, 1259 (1963).

[10] C. C. McDonald and W. D. Phillips, *J. Am. Chem. Soc.*, 89, 6332 (1967)

[11] K. Wüthrich, *Structure and Bonding*, 8, 53 (1970).

[12] K. Wüthrich, *Proc. Nat. Acad. Sci. US*, 63, 1071 (1969).

[13] G. C. K. Roberts and O. Jardetzky, *Adv. Protein Chem.*, 24, 447 (1970).

[14] C. C. McDonald, W. D. Phillips and J. D. Glickson, *J. Am. Chem. Soc.*, 93, 235 (1971).

[15] K. Wüthrich, Proceedings of the 4th International Biophysics Congress, Moscow 1972. in the press.

[16] A. Allerhand, D. Doddrell, V. Glushko, D. W. Cochran, E. Wenkert, P. J. Lawson and F. R. N. Gurd, *J. Am. Chem. Soc.*, 93, 544 (1971).

[17] J. D. Robinson, N. J. M. Birdsall, A. G. Lee and J. C. Metcalfe, *Biochemistry*, 11, 2903 (1972).

[18] J. Ozols and P. Strittmatter, *J. Biol. Chem.*, 244, 6617 (1969).

[19] F. Scott Mathews, M. Levine and P. Argos, *J. Mol. Biol.*, 64, 449 (1972).

[20] R. M. Keller and K. Wüthrich, *Biochim. Biophys. Acta*, 285, 326 (1972).

[21] R. Bois-Poltoratsky and A. Ehrenberg, *Eur. J. Biochem.*, **2**, 361 (1967).

[22] R. M. Keller. O. Groudinsky and K. Wüthrich. *Biochim. Biophys. Acta.* **328**. 233 (1973).

[23] K. Wüthrich and R. Baumann, *Ann. N.Y. Acad. Sci.*, in the press.

[24] K. Wüthrich and R. Baumann, *Helv. Chim. Acta,* **56**, 585 (1973).

[25] M. Karplus and G. K. Fraenkel, *J. Chem. Phys.,* **35**, 1312 (1961).

[26] J. L. Markley, I. Putter and O. Jardetzky, *Science,* **161**, 1249 (1968).

[27] A. M. Nigen, P. Keim. R. C. Marshall, J. S. Morrow and F. R. N. Gurd. *J. Biol. Chem.,* **247**, 4100 (1972).