Proton Magnetic Resonance Spectra of Proteins in Random-Coil Configurations

C. C. McDonald and W. D. Phillips

Contribution No. 1483 from the Central Research Department, Experimental Station, E. I. du Pont de Nemours and Company, Wilmington, Delaware 19898. *Received August 21, 1968*

Abstract: Characteristics of proton magnetic resonance spectra at 220 MHz of random-coil proteins and peptides in neutral aqueous solution are discussed. A general procedure for computing such spectra is presented. Computed spectra are compared with actual spectra of lysozyme, ribonuclease, pepsin, trypsin, apoferredoxin, apoflavodoxin, oxytocin, and cytochrome *c*.

The most general and potentially the most powerful way to use proton magnetic resonance (pmr) spectroscopy to elucidate structures and interactions of native proteins in aqueous solution is direct examination of the proton resonances of hydrogen atoms of the proteins. This approach has recently become much more feasible because of advances in pmr instrumentation. Previously, the principal limitation had been inability to resolve individual protein proton resonances. Pmr spectra of proteins comprise a large number of individual resonances at many different field positions arising from hydrogen atoms in the many different shielding environments of a protein molecule. Since the resonances are confined to a relatively narrow field region, overlap is frequent and is enhanced by the rather broad resonance widths normally exhibited by proteins. Broadening of the resonances arises from anisotropic magnetic dipolar interactions between magnetic nuclei. Such broadening interactions are largely averaged out by molecular tumbling of small molecules in solution so that resonance half-widths of 1 Hz or less are typical. For larger molecules, rotation is slower and anisotropic dipolar broadening becomes evident. Local motions of segments of a large molecule, such as the flexing of the polymer chain of a protein in a random-coil configuration, may reduce dipolar broadening relative to that for a rigid conformation of the same molecule where the resonance line width is determined by over-all molecular rotation. Our studies of pmr spectra of native proteins indicate, however, that these molecules, typically, are sufficiently rigid so that their proton resonance line widths are controlled by over-all molecular rotation. Consequently, resonance half-widths increase with protein molecular weight (as does the multiplicity of resonances), and overlap of individual resonances becomes increasingly serious in limiting resolution. We have found that resonance half-widths for native proteins in the 10,000 to 20,000 molecular weight range are typically 10 to 20 Hz. Overlap of resonances of even such relatively small proteins in their native conformations in aqueous solution is so extensive that in earlier studies at spectrometer resonance frequencies below 100 Hz only an envelope of the individual resonances was usually observed[1] although Bradbury and Scheraga[2] were able to resolve individual C-2 proton resonances of the histidine residues of ribonuclease in studies at 60 MHz.

The separation between chemically shifted proton resonances increases linearly with the magnitude of the polarizing magnetic field. Consequently, resolution of overlapped resonances may be increased by increasing the applied magnetic field and concomitantly the frequency of the radiofrequency field of a pmr spectrometer. Recently, spectrometers operating at 100 and 220 MHz have become available and have been used to obtain pmr spectra of native proteins.[3-12] These instruments provide sufficient resolution of some resonance regions of the pmr spectra of small proteins (molecular weights below about 25,000) in their native conformations to give access to new information on structures and interactions of these proteins. In a preliminary report on the characteristics of pmr spectra of native proteins obtained with a 220 MHz spectrometer,[8] we showed that positions of the observable resolved resonances are partially determined by tertiary conformation since they are shifted considerably when protein configuration is changed from extended random coil where the residue side chains are in an aqueous environment to the native conformation having many residues in an internal protein environment. Furthermore, the characteristics of spectra of native proteins have been found to be very sensitive to perturbations of the native conformations induced by modest changes of environment (pH, temperature, ionic strength) or interactions with other molecules (substrates, inhibitors, etc.). Therefore, it appears that these changes in the pmr spectra of a protein as it is converted from a ran-

(1) M. Saunders, A. Wishnia, and J. G. Kirkwood, *J. Am. Chem. Soc.*, **79**, 3289 (1957); F. A. Bovey, G. V. D. Tiers, and G. Filipovich, *J. Polymer Sci.*, **38**, 73 (1959); A. Wishnia and M. Saunders, *J. Am. Chem. Soc.*, **84**, 4235 (1962); A. Kowalsky, *J. Biol. Chem.*, **237**, 1807 (1962); A. Kowalsky, *ibid.*, **239**, 3299 (1964).

(2) J. H. Bradbury and H. A. Scheraga, *J. Am. Chem. Soc.*, **88**, 4240 (1966).

(3) M. Mandel, *Proc. Natl. Acad. Sci. U. S.*, **52**, 736 (1964); *J. Biol. Chem.*, **240**, 1586 (1965).

(4) D. P. Hollis, G. McDonald, and R. L. Biltonen, *Proc. Natl. Acad. Sci. U. S.*, **58**, 758 (1967).

(5) H. Sternlicht and D. Wilson, *Biochemistry*, **6**, 2881 (1967).

(6) D. H. Meadows, J. L. Markley, J. S. Cohen, and O. Jardetzky, *Proc. Natl. Acad. Sci. U. S.*, **58**, 1307 (1967).

(7) C. C. McDonald and W. D. Phillips, "Proceedings of the Second International Conference on Magnetic Resonance in Biology," Pergammon Press, Inc., New York, N. Y., 1967.

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

(9) J. S. Cohen and O. Jardetzky, *Proc. Natl. Acad. Sci. U. S.*, **60**, 92 (1968).

(10) D. H. Meadows, O. Jardetzky, R. M. Epand, H. H. Ruterjans, and H. A. Scheraga, *ibid.*, **60**, 766 (1968).

(11) R. J. Kurland, D. G. Davis, and C. Ho, *J. Am. Chem. Soc.*, **90**, 2700 (1968).

(12) K. Wüthrich, R. G. Shulman, and J. Peisach, *Proc. Natl. Acad. Sci. U. S.*, **60**, 273 (1968).

Table I. Positions of Proton Resonances of Amino Acids and Peptides in Neutral D_2O at 40°

Proton type	Equiv protons per residue	Compound	Resonance position,[a] Hz	Est apparent half-width,[b] Hz
Leucine CH_3	6	L-Leucine	208	15
		Glycyl-L-leucine	197	13
		L-Histidyl-L-leucine	193	20
β-CH_2 + γ-CH	3	L-Leucine	374	20
		Glycyl-L-leucine	350	16
		L-Histidyl-L-leucine	345	25
α-CH	1	L-Leucine	813	14
		Glycyl-L-leucine	920	15
		L-Histidyl-L-leucine	≈905	
Isoleucine CH_3	3	L-Isoleucine	202	16
CH_3	3		217	17
CH_2	1		273	29
CH_2	1		319	29
β-CH	1		429	26
α-CH	1		800	14
Valine CH_3	3	L-Valine	214	17
CH_3	3		225	17
β-CH	1		494	24
α-CH	1		789	15
Alanine CH_3	3	L-Alanine	322	18
		Glycyl-L-alanine	294	18
		L-Tyrosyl-L-alanine	292	
		Glycyl-L-phenylalanyl-L-alanine	288	
		L-Tryptophanyl-L-alanine	300	
α-CH	1	L-Alanine	827	18
		Glycyl-L-alanine	916	18
Threonine CH_3	3	L-Threonine	288	16
β-CH	1		928	20
α-CH	1		772	15
Glycine α-CH_2	2	Glycine	776	10
		Glycylglycine	833, 836	
		Glycylglycylglycine	827, 848, 883	
		Glycyl-L-tyrosine	830	22
		Glycyl-L-leucine	835	10
		Glycyl-L-alanine	833	10
		Glycyl-L-asparagine	837	
Lysine γ-CH_2	2	L-Lysine	321	30
	2	Poly-L-lysine	315	35[c]
δ-CH_2	2	L-Lysine	375	25
β-CH_2	2	L-Lysine	412	25
δ-CH_2 + β-CH_2	4	Poly-L-lysine	376	38[c]
ϵ-CH_2	2	L-Lysine	664	22
	2	Poly-L-lysine	660	22[c]
α-CH	1	L-Lysine	821	12
	1	Poly-L-lysine	947	22[c]
Arginine γ-CH_2	2	L-Arginine	368	28
β-CH_2	2		412	24
δ-CH_2	2		709	14
α-CH	1		822	14
Serine β-CH_2	2	L-Serine	865	25
α-CH	1		840	17
Proline γ-CH_2	2	L-Proline	443	21
β-CH_2	1		456	25
β-CH_2	1		510	25
δ-CH_2	2		736	32
α-CH	1		905	23
Glutamic acid β-CH_2	2	L-Glutamic acid	454	20
γ-CH_2	2		512	20
α-CH	1		818	14
Glutamine β-CH_2	2	L-Glutamine	463	20
γ-CH_2	2		534	20
α-CH	1		820	12
Aspartic acid β-CH_3	2	L-Aspartic acid	595	55
α-CH	1		848	22
Asparagine β-CH_2	1	L-Asparagine	627	29
		Glycyl-L-asparagine	580	30
β-CH_2	1	L-Asparagine	641	29
		Glycyl-L-asparagine	615	30
α-CH	1	L-Asparagine	873	20
		Glycyl-L-asparagine	995	
Methionine CH_3	3	L-Methionine	467	10
β-CH_2	2		467	22
γ-CH_2	2		580	16
α-CH	1		842	12

Table I (*Continued*)

Proton type	Equiv protons per residue	Compound	Resonance position,[a] Hz	Est apparent half-width,[b] Hz
Cysteine β-CH_2	2	L-Cysteine	671	12
α-CH	1		870	12
Histidine β-CH_2	2	L-Histidine	700	28
		L-Histidyl-L-leucine	702	17
		Glycyl-L-histidylglycine	690	
α-CH	1	L-Histidine	875	15
		L-Histidyl-L-leucine	≈895	
Imidazole C-4	1	L-Histidine	1558	10
		L-Histidyl-L-leucine	1571	
		Glycyl-L-histidylglycine	1558	
Imidazole C-2	1	L-Histidine	1725	10
		L-Histidyl-L-leucine	1773	
		Glycyl-L-histidylglycine	1758	
Tyrosine β-CH_2	1	Glycyl-L-tyrosine	628	27
		L-Tyrosyl-L-alanine	678	
β-CH_2	1	Glycyl-L-tyrosine	683	27
		L-Tyrosyl-L-alanine	694	
α-CH	1	Glycyl-L-tyrosine	972	
		L-Tyrosyl-L-alanine	906	
Aromatic *ortho* to OH	2	L-Tyrosine (80°)	1514	17
		Glycyl-L-tyrosine	1504	
		L-Tyrosyl-L-alanine	1513	
Aromatic *meta* to OH	2	L-Tyrosine (80°)	1583	17
		Glycyl-L-tyrosine	1572	
		L-Tyrosyl-L-alanine	1578	
Phenylalanine β-CH_2	1	L-Phenylalanine	684	30
		Glycyl-L-phenylalanyl-L-alanine	≈651	
β-CH_2	1	L-Phenylalanine	720	30
		Glycylphenylalanine	≈702	
α-CH	1	L-Phenylalanine	876	
Aromatic	3	L-Phenylalanine	1627	
	2		1616	
	5	Glycylphenylalanine	1607	30
Tryptophan β-CH_2	2	L-Tryptophanyl-L-alanine	743	27
α-CH	1		932	15
Indole C-2	1		1610	10
Indole C-5, C-6	1, 1		1583, 1602	15, 15
Indole C-4, C-7	1, 1		1658, 1690	18, 18

[a] From internal DSS at 220 MHz. [b] For half-width of all resonance components of 10 Hz. [c] Actual.

dom-coil configuration to a folded conformation (native or otherwise) can not only be used to follow the change but can provide information about the latter conformation.

While pmr spectra of a random-coil protein provide no direct information on the nature of the folded native protein, they do provide a necessary reference for measurements of spectral changes that occur when the protein converts to a folded conformation. Some proteins can be prepared in a random-coil configuration in neutral aqueous solution for pmr studies by thermal denaturation or a combination of thermal and chemical denaturation (guanidine or urea); other proteins for various reasons cannot. For example, some proteins may aggregate or precipitate as they unfold in such solutions. Sometimes they can be adequately denatured by large changes in pH or use of another solvent (*e.g.*, trifluoroacetic acid) to provide good random-coil pmr spectra. However, resonance positions are sufficiently shifted in these new environments so that the spectra are not useful for comparison with spectra of native proteins in neutral aqueous solution. In many instances, denaturation is not reversible, and if only small quantities of a protein are available it is not desirable to expend it obtaining the pmr spectrum of the random-coil form.

In this paper, some general characteristics of pmr spectra at 220 MHz of proteins in random-coil configurations are discussed. A procedure is introduced for computing some of the characteristics of the pmr spectrum at 220 MHz of any protein in an extended random-coil configuration in neutral aqueous solution. These computed spectra aid in analysis of actual spectra of random-coil proteins and provide random-coil spectra for proteins that are not conveniently observable in this configuration. Interpretation of changes that occur in such pmr spectra when a protein converts from a random coil to a folded conformation will be provided in subsequent reports.

Characteristics of Pmr Spectra of Random-Coil Proteins

In this paper, a random-coil protein is defined as one that is sufficiently extended in solution so that the side chains of the residues are in a solvent environment. Further, all regions of the protein have degrees of motional freedom approaching that of a true extended random-coil polypeptide. This definition thus may in-

clude proteins that retain a few elements of structure such as disulfide bridges or covalent and/or coordinate linkages to prosthetic groups.

Proton magnetic resonance investigations of proteins are usually mainly concerned with examination of solutions of proteins in D_2O so that a major part of the protein spectrum is not obscured by the strong solvent resonance. Protein hydrogen atoms bonded to N or O exchange rapidly with deuterons in neutral D_2O even at room temperature if they are accessible to the solvent as they are in a random-coil protein. Consequently the pmr spectrum of a random-coil protein comprises only resonances of hydrogen atoms bonded to carbon atoms (resonances of prosthetic groups such as heme groups will not be treated here).

Since the side chains of the residues of a random-coil protein are in an aqueous environment, we may expect all protons of a given chemical type, for example, alanine methyl-group protons, to be equivalent and to exhibit resonances at the same field position unless the resonance positions are influenced by the nature of neighbor residues. Experience here and in other investigations[13] has shown that for random-coil proteins the only residue resonance positions that are strongly dependent on the nature of neighbor residues are those arising from protons bonded to α-carbon atoms. Resonance positions of other hydrogen atoms of a given chemical type are, for the purposes of this paper, essentially coincident[14] for all residues of a given type in a protein and, since they are determined by the solvent environment, they are the same for different random-coil proteins. Consequently, it should be possible to select a set of resonance positions for all chemical types of hydrogen atoms in a random-coil protein (except α-hydrogens) and to use them with appropriate values for resonance intensities and line shapes to compute the pmr spectrum of any random-coil protein. This assumption is the basis of the computational procedure that we present here.

Procedure for Computing Schematic Representations of Pmr Spectra of Random-Coil Proteins

Typically, pmr spectra of native proteins are obtained by examination of the proteins in D_2O at pD 7 and 40°. Thus, a set of resonance positions selected for computing random-coil protein spectra should be representative of such an environment. First, the approximate resonance position of each proton type was found by examination of the pmr spectra at 40° of L-amino acids commonly occurring in proteins and a number of short peptides of these amino acids dissolved in D_2O at pD 7. These data are presented in Table I. Throughout this paper, resonance positions are referred to the proton resonance position of the methyl groups of sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) used as an internal reference. Chemical shifts are quoted in hertz (cycles per second) or parts per million and positive numbers indicate shifts to low field from the reference resonance. All spectra were obtained with a Varian Associates 220-MHz pmr spectrometer using experimental techniques reported previously.[8] Assignments of the resonances in Table I are based on previous pmr studies of these or similar compounds at lower spectrometer frequencies[3,13,15] and on the usual considerations of resonance positions, intensities, and spin–spin splittings. The resonance positions of the protons on the side chains of amino acids are influenced by the charged amine and carboxyl groups. These we might term chain-end effects. Thus, use of these resonance positions does not provide a satisfactory set for computing random-coil protein spectra. Some improvement is obtained by using resonance positions of protons of dipeptides at pD 7 and 40°. However, these also are somewhat influenced by end effects (not only the charged groups but hydrogen bonding between end groups and side chains in some instances). The final set of resonance positions was selected by modifying the values in Table I for amino acids and peptides to agree with the actual random-coil spectra of lysozyme and ribonuclease. These proteins can be readily converted to an extended configuration in neutral D_2O by heating above their thermal denaturation temperatures. The thermally denatured proteins retain several disulfide bridges but appear to be sufficiently extended and flexible to conform to the definition of "random coil" used here. In this final selection of resonance positions, some consideration was also given to resonance positions of other actual random-coil pmr spectra of polypeptides and proteins which are discussed below. The selected set of resonance positions is presented in column 3 of Table II. The ultimate test of the validity of the selected set of resonance positions is agreement with the corresponding resonance positions of a wide variety of actual random-coil proteins at 40° and pD 7. Such tests have been made and are discussed below.

The next consideration in selecting a set of standard resonances for computing random-coil protein spectra is the matter of resonance line shape. The resonance shape typically provided by the 220-MHz pmr spectrometer is shown in Figure 1 for the proton resonance of H_2O. This line shape cannot be quantitatively described by a simple mathematical function because of an anomalous intensity distribution in the wings but it can be simulated adequately for the qualitative computation proposed here by a triangle of altitude equal to the resonance height and base equal to twice the resonance width at half-height (half-width). We have found from pmr spectra of a number of polypeptides and proteins in random-coil configurations in neutral D_2O for molecular weights 10,000 to 25,000 that resonance half-widths are about 10 Hz for these flexible chains. These resonance half-widths become somewhat narrower for short peptides and a little broader for longer chains but 10 Hz appears to be a useful standard half-width for an individual resonance that is not split by spin–spin interactions. Actually most of the amino acid proton resonances are split into several closely spaced resonances of different intensities by spin–spin interactions. This fine structure is clearly resolved in pmr spectra of the amino acids where component

(13) A. Nakamura and O. Jardetzky, *Proc. Natl. Acad. Sci. U. S.*, 58, 2212 (1967), and references therein.

(14) Experience so far indicates this statement is generally true, but some exceptions may arise. For example, perturbation of resonance positions by ring-current magnetic fields from a nearest neighbor aromatic residue has not been assessed for a random-coil relationship between various residue types.

(15) M. Takeda and O. Jardetzky, *J. Chem. Phys.*, **26**, 1346 (1957); O. Jardetzky and C. D. Jardetzky, *J. Biol. Chem.*, **233**, 383 (1958).

Table II. Basic Data for Computing Pmr Spectra of Random-Coil Proteins

Proton type	Equiv protons per residue	Resonance triangle position,[a] Hz	Triangle base, Hz	Triangle altitude per residue
Leucine CH_3	6	195	30	40.0
β-CH_2 + γ-CH	3	360	40	15.0
Isoleucine CH_3	6	183	40	30.0
CH_2	1	250	60	3.3
CH_2	1	310	60	3.3
β-CH*	1	425	50	4.0
Valine CH_3	6	205	34	35.2
β-CH	1	495	50	4.0
Alanine CH_3	3	310	36	16.7
Threonine CH_3	3	270	32	18.7
Lysine γ-CH_2	2	315	60	6.7
δ-CH_2 + β-CH_2	4	370	60	13.3
ϵ-CH_2	2	665	44	9.1
Arginine γ-CH_3	2	365	56	7.2
β-CH_2	2	405	48	8.4
δ-CH_2	2	704	28	14.3
Proline γ-CH_2	2	445	42	9.5
β-CH_2	2	465	50	8.0
δ-CH_2*	2	725	60	6.7
Glutamic acid β-CH_2	2	435	40	10.0
γ-CH_2	2	500	40	10.0
Glutamine β-CH_2	2	455	40	10.0
γ-CH_2	2	510	40	10.0
Aspartic acid β-CH_2	2	590	110	3.6
Asparagine β-CH_2	1	613	60	3.3
β-CH_2	1	637	60	3.3
Methionine CH_3	3	454	20	30.0
β-CH_2	2	454	44	9.1
γ-CH_2*	2	565	32	12.5
Cysteine β-CH_3	2	665	24	16.6
Histidine β-CH_2*	2	700	56	7.15
Imidazole C-4	1	1555	20	10.0
Imidazole C-2	1	1740	20	10.0
Tyrosine β-CH_2*	2	655	60	6.7
Aromatic *ortho* to OH	2	1500	34	11.8
Aromatic *meta* to OH	2	1560	34	11.8
Phenylalanine β-CH_2*	1	650	60	3.3
β-CH_2*	1	700	60	3.3
Aromatic	5	1598	60	16.7
Tryptophan β-CH_2*	2	745	54	7.4
Indole C-2	1	1584	20	10.0
Indole C-5, C-6	1, 1	1549, 1566	30, 30	6.7, 6.7
Indole C-4, C-7	1, 1	1638, 1658	36, 36	5.6, 5.6

[a] From internal DSS at 220 MHz.

resonance widths are typically 1 Hz or less, but it is usually not resolved for resonances of amino acid residues of random-coil proteins where the individual resonance half-widths ($\approx$10 Hz) are comparable to or larger than the spin–spin splitting. Consequently, for such resonances, one observes an envelope of the spin–spin fine structure but again the over-all resonance line shape can be approximately simulated by a triangular representation. Thus, we have chosen to represent the set of standard resonances for computation of random-coil protein spectra by a set of triangles centered at the resonance positions shown in Table II. For a resonance arising from a proton that exhibits no spin–spin splitting (in the amino acid spectrum) the resonance triangle has a base of 20 Hz (twice the half-width) and is arbitrarily assigned an altitude of 10 units. Where a resonance does show spin–spin splitting we have estimated from the amino acid spectra what the apparent half-width of the resonance envelope would be if the resonance half-widths of the individual spin–spin components were 10 Hz. The resonance is then represented by a triangle with a base that is twice the value of this half-width. The altitude of the triangle is decreased in proportion to the amount that the base has been increased from 20 Hz to maintain a constant area representing the intensity of a single proton. Finally, the altitude of a resonance triangle is multiplied by the number of chemically equivalent protons in a residue. To illustrate by an example, the three methyl protons of alanine are equivalent and exhibit resonance at the same field position. The resonance is split into a doublet by the α-hydrogen. The selected resonance position for the center of these resonances is 310 Hz, and the selected resonance half-width is 18 Hz. Consequently, the triangle representing one of the methyl protons of an alanine methyl group is a triangle centered at 310 Hz with base 36 Hz and altitude 5.57 units, but since there are three equivalent methyl protons the triangle representing an entire alanine residue has an altitude of 16.7 units. The selected values for the bases and altitudes per residue for triangles representative of other residue proton resonances are tabulated in Table II. Proton

types for which we do not yet have satisfactory information on resonance positions or line widths are noted with asterisks. For these protons, the data are largely based on the appropriate amino acid spectra and they may be improved as they are observed as predominate resonances in polypeptide spectra. To compute the resonance triangles for an actual protein, the triangle altitude per residue is multiplied by the number of residues of that type in the protein. The computed pmr spectrum is obtained by point-by-point summation of "intensities" of the resonance triangles where they overlap.

We do not include in this treatment protons bonded to N or O since they exchange in D_2O, or resonances of protons on prosthetic groups. Furthermore, we have included only proton resonances of the amino acid side chains that fall in the spectral ranges from 0 to 750 Hz and from 1400 to 1800 Hz since these are the most useful spectral regions for comparing random-coil and native protein spectra. Thus, we have excluded all α-hydrogens, the β-hydrogen of threonine, and the two β-hydrogens of serine. These resonances all fall in the spectral region between 750 and 1100 Hz which is typically partially obscured by the resonance of HDO in the D_2O solvent and by spinning side bands of this strong resonance.

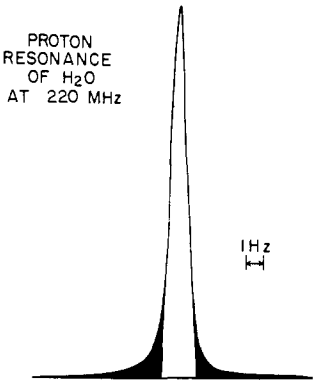

Figure 1. Pmr of H_2O as detected by 220-MHz spectrometer.

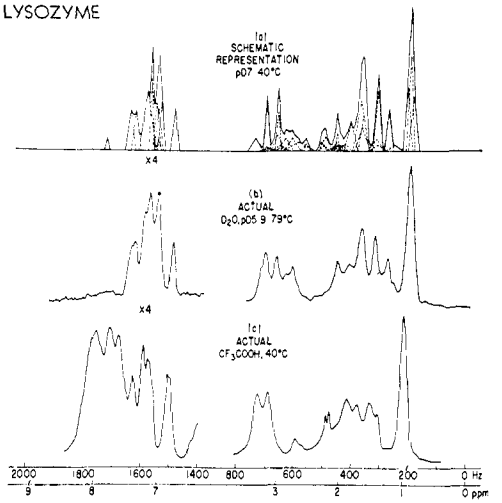

Figure 2. Pmr spectra of HEW lysozyme at 220 MHz.

Comparison of Computed and Actual Pmr Spectra

Confidence in the utility of the computational procedure introduced above has been gained by comparison of computed spectra with actual spectra of a number of polypeptides and proteins that were examined under conditions in which we believe that they have configurations approaching that defined here as "random coil." Some illustrations and discussion of these comparisons are now given.

Lysozyme. Regions of the pmr spectrum exhibited by hen egg white (HEW) lysozyme at 79° (Worthington, salt free, 10% in D_2O, pD 5.9) are shown in Figure 2b. As for the other spectra presented here, computer averaging of random noise was used to improve the signal-to-noise ratio of spectra provided by the spectrometer. This protein undergoes reversible thermal denaturation from 68 to 74° and at 79° is believed to be in an extended random-coil configuration except for the four disulfide bridges which remain intact. The spectrum was obtained at pD 5.9 rather than pD 7 as this protein aggregates considerably above about pD 6. It is our general experience that neither resonance positions nor resonance line widths are appreciably affected by temperature (20–90°) *as long as the protein configuration remains random coil.* Small departures of pD from neutrality do not affect most resonances but may markedly affect resonances of a residue that undergoes a change in ionization in this pD region (*e.g.*, histidine). The effects of pD on random-coil protein spectra are discussed further below.

The pmr spectrum of lysozyme computed using the data of Table II and the amino acid composition of this protein[16] is shown by the solid lines in Figure 2a. The individual resonance triangles for each proton type are shown by dashed lines. Both the actual and schematic spectra have been amplified by the factor indicated (×4) in the spectral region from 1400 to 1800 Hz relative to the intensity of the 0- to 750-Hz spectral region. There is no basis for quantitative intensity comparison of the computed and actual spectra since they are in arbitrary units, but they have been approximately normalized by making the methyl resonances at 200 Hz the same magnitude. The correspondence between the schematic representation and the experimental spectrum is good in most spectral regions. While the experimental lysozyme spectrum was extensively used in arriving at the resonance positions to be used for computed spectra, the agreement in the intensity distributions of the two spectra reinforces the accuracy of the assignments. As expected, resonances appear better resolved in the computed spectrum than in the actual spectrum because in simulating the resonances by triangles we have neglected intensity in the wings.

In the actual lysozyme spectrum the histidine C-2 proton resonance at 1740 Hz is not evident. Meadows, *et al.*[10] have shown that this proton exchanges slowly with deuterium of D_2O in neutral solution at room temperature. It exchanges quite rapidly at elevated temperatures and therefore is typically reduced in intensity or absent in pmr spectra of proteins in neutral D_2O above about 65°.

(16) R. E. Canfield, *J. Biol. Chem.*, 238, 2698 (1963).

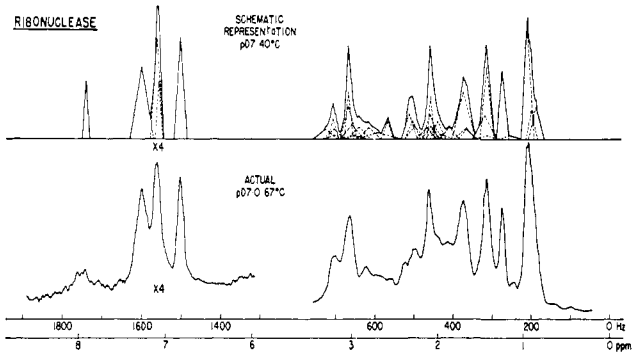

Figure 3. Pmr spectra of bovine pancreatic ribonuclease at 220 MHz.

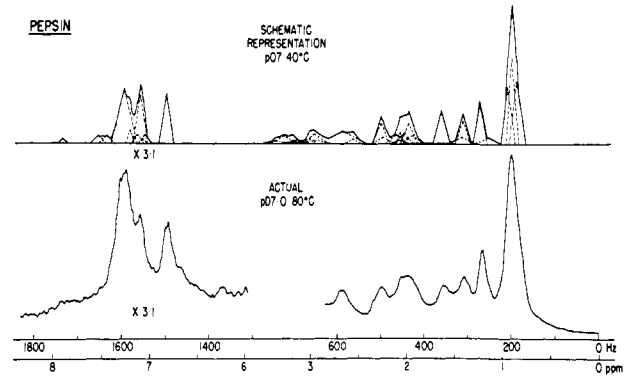

Figure 4. Pmr spectra of porcine pepsin at 220 MHz.

The predicted peak at about 500 Hz is not observed suggesting that the valine, glutamic acid, and glutamine protons expected to contribute in this region may be slightly shifted or broadened relative to the computed resonances. They do provide a prominent resonance at this position if the experiment is done at a lower pD. Similarly, there is little evidence of the expected peak of tryptophan β protons at 745 Hz and this resonance may be shifted into the argine resonance at 704 Hz in the actual spectrum. The peak at 610 Hz is stronger than predicted and it appears that the aspartic acid resonance predicted at 590 Hz is actually at about 610 Hz for lysozyme.

We note that the cysteine β protons predicted as a sharp resonance at 665 Hz appear as a well-defined resonance at this position in the actual lysozyme spectrum, although they occur in residues involved in disulfide bridges rather than in free cysteine side chains. We also note that the sharp resonance of methionine methyl protons predicted at 454 Hz is observable in the actual spectrum although that spectral region is complicated by contributions from many other proton types.

Cohen and Jardetzky[9] have recently reported partial assignments of the low-field region of the pmr spectrum of HEW lysozyme obtained at 100 MHz for several denatured conformations of this protein. Their assignments appear to be generally in agreement with those suggested here. An interesting observation of their studies is the increased motional narrowing of the denatured protein resonances when the disulfide bonds are ruptured.

Ribonuclease. A pmr spectrum of bovine pancreatic ribonuclease at 67° (CalBiochem. A grade 10% in D_2O, pD 7) is shown in Figure 3. In this solvent environment the protein is believed to be in a configuration approaching a random coil except for residual disulfide bridges. The computed spectrum for this protein (amino acid composition from ref 17) also compares very well with the actual spectrum.

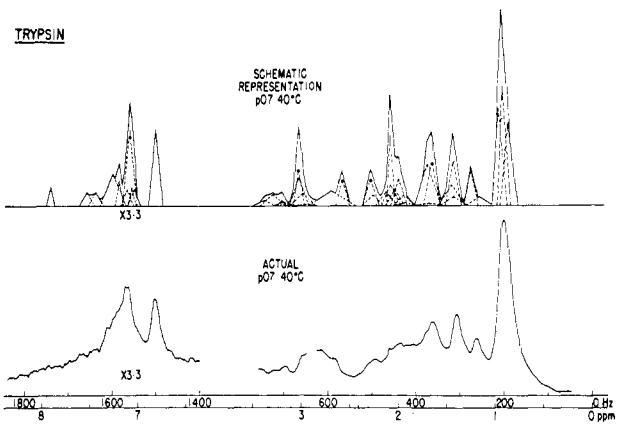

Figure 5. Pmr spectra of trypsin at 220 MHz.

Ribonuclease has no tryptophan residues, so the positions for the low-field "aromatic" phenylalanine and tyrosine resonances are clearly established. The computed and experimental spectra agree rather well throughout the high-field region. We note again a sharp methionine methyl proton resonance in the experimental spectrum near the predicted position (454 Hz), and an apparent contribution of sharp equivalent cystine β protons to the resonance at 665 Hz.

Pepsin. Regions of the pmr spectrum of porcine pepsin at 80° (Worthington, twice crystallized, 15% in D_2O, pD 7) are shown in Figure 4. The schematic representation was computed using the amino acid composition reported for pepsin by Blumenfeld and Perlmann.[18] Pepsin, with a molecular weight of about 35,000, is the largest protein examined in this investigation. The computed and actual spectrum agree well throughout the high-field region. In the low-field region, the resonance of phenylalanine at about 1600 Hz appears somewhat stronger than predicted.

Trypsin. A pmr spectrum of trypsin at 78° (Worthington, twice crystallized, salt-free 10% in D_2O, pD 7) is shown in Figure 5 together with the spectrum computed from the amino acid composition of trypsin reported by Domont, *et al.*[19] Agreement between the spectra is satisfactory in most respects. The strong methionine peaks predicted at 454 and 565 Hz appear to be broadened so that they are not readily detectable in the actual spectrum. This finding indicates that at least some of the 12 methionine residues are in protein regions that do not have a random-coil configuration at pD 7 and 78°.

(17) D. G. Smyth, W. H. Stein, and S. Moore, *J. Biol. Chem.*, 238 227 (1963).
(18) O. O. Blumenfeld and G. E. Perlmann, *J. Gen. Physiol.*, 42, 553 (1959).
(19) G. B. Domont, A. Iachon, L. V. Disitzer, and J. C. Perrone, *Anais Acad. Brasil Cienc.*, 36, 137 (1964); *Chem. Abstr.*, 62, 8074g (1965).

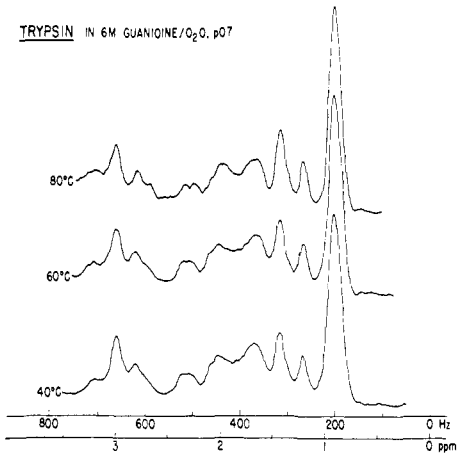

Figure 6. Pmr spectra of trypsin in 6 *M* guanidine at 220 MHz.

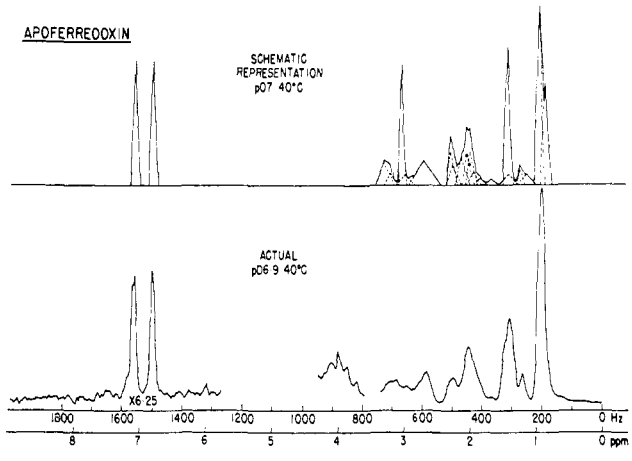

Figure 7. Pmr spectra of apoferredoxin at 220 MHz.

The high-field region of the pmr spectrum of trypsin in 6 *M* guanidine–D_2O at pD 7 is shown in Figure 6 at 40, 60, and 80°. The protein appears to be in a configuration close to random coil even at 40° in 6 *M* guanidine. The resonances are shifted 5 to 15 Hz to lower field by the presence of the high concentration of guanidine in the environment but are not appreciably affected by temperature from 40 to 80°. The methionine resonance regions are somwhat more resolved in 6 *M* guanidine at 80° than in Figure 5 but are still less evident than expected from the computed spectrum.

Apoferredoxin. Portions of the pmr spectrum of apoferredoxin[20] at 40° (a 5.7% solution in D_2O at pD 6.9) are shown in Figure 7. This protein is a single polypeptide chain of 55 residues with 4 disulfide bridges. Agreement between the computed spectrum of apoferredoxin[21] also shown in Figure 7 and the actual spectrum is excellent except at 665 Hz where the expected strong sharp resonance of cystine β-hydrogens is not evident in the actual spectrum. We do not believe that this discrepancy indicates that some regions of rigid structure in the neighborhood of these protons caused their resonances to be shifted or greatly broadened because spectra obtained at 60 and 80° are similar throughout the spectral regions shown in Figure 7. Many of the resonances are quite sharp reflecting the low molecular weight of this protein and that the residues contributing to these resonances are in random-coil regions. For example, indications of spin–spin splitting (7.5 Hz) of the low-field tyrosine resonances are evident. However, the alanine methyl peak in the experimental spectrum shows a shoulder that suggests that some of the alanine residues (of the nine in this protein) remain in structured regions.

Oxytocin. Next we compare computed and actual

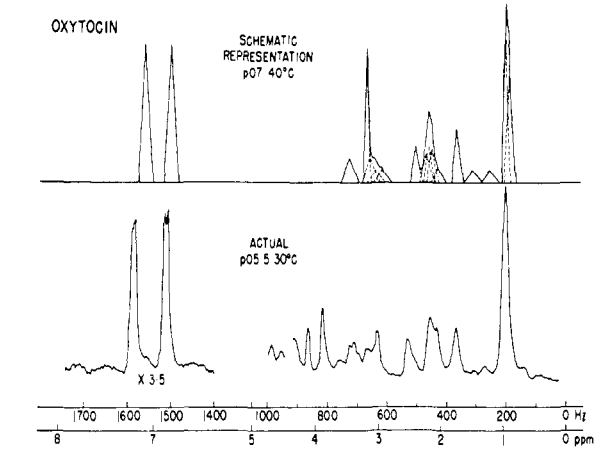

Figure 8. Pmr spectra of oxytocin at 220 MHz.

spectra for the small polypeptide, oxytocin in Figure 8.[22] This peptide has the sequence Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly.[23] The cysteine residues are present as cystine and the C-terminal group is an amide function. Agreement between the computed and actual spectra is satisfactory in most regions, suggesting that most of the residue side chains are in a solvent environment at 30°. The low-field tyrosine resonances are somewhat farther to low field than expected as is the glutamine resonance expected at 510 Hz. The actual spectrum departs significantly from expectation at the cystine β-hydrogen position at 665 Hz. Some if not all of these four proton resonances have shifted to other positions. This discrepancy may be due in part to the fact that one of the cysteine residues is at the N-terminal end of the peptide chain and in part to structural features of oxytocin.

Flavodoxin. Flavodoxin is a flavoprotein of 14,600 molecular weight containing 148 amino acid residues and a flavin mononucleotide prosthetic group.[24] The

(20) This apoferredoxin was prepared by J. C. Rabinowitz, Department of Biochemistry, University of California, Berkeley, Calif., from ferredoxin extracted from the bacterium *Clostridium acidi-urici*. It contains no inorganic iron and no inorganic sulfide, but is constituted from the same amino acids as the native ferredoxin[21] except that the cysteine residues are present as disulfide bonds. The spectrum in Figure 7 was obtained in a collaborative pmr study of *Clostridium acidi-urici* ferredoxin with Professor Rabinowitz.

(21) W. Lovenberg, B. B. Buchanan, and J. C. Rabinowitz, *J. Biol. Biol. Chem.*, **238**, 3899 (1963).

(22) From a collaborative pmr investigation of oxytocin with M. Goodman, Department of Chemistry, Brooklyn Polytechnical Institute, New York, N. Y.

(23) D. Jarvis and V. du Vigneaud, *Science*, **143**, 545 (1964).

(24) E. Knight, Jr., and R. W. F. Hardy, *J. Biol. Chem.*, **242**, 1370 (1967).

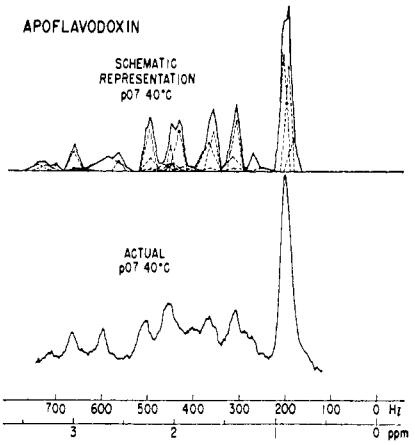

Figure 9. Pmr spectra of apoflavodoxin at 220 MHz.

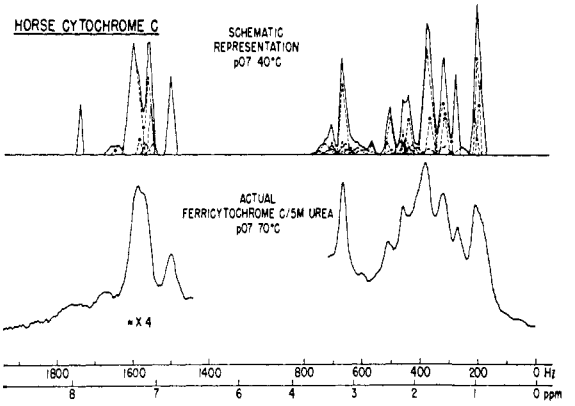

Figure 10. Pmr spectra of horse cytochrome *c* at 220 MHz.

apoprotein (flavodoxin from the bacterium *Clostridium pasteurianum* with the prosthetic group removed) was kindly provided for our examination by Drs. E. Knight and R. W. F. Hardy of this laboratory. The high-field region of the pmr spectra of a solution of apoflavodoxin (D_2O at pD 7) at 40° is compared with a computed spectrum in Figure 9. The rather close agreement suggests that the apoprotein configuration approaches random coil in neutral D_2O at 40°. The pmr spectrum of the native protein is quite different.

Cytochrome *c*. Cytochrome *c* provides a test of the utility of the computational procedure for simulating the random-coil spectrum of a protein with a large prosthetic group (in this case, a heme group). A random-coil pmr spectrum of ferricytochrome *c* in neutral D_2O cannot be obtained simply by heating the solution since the thermally denatured protein aggregates. However, this aggregation is greatly reduced for thermal denaturation in 5 *M* urea. Accordingly, in Figure 10 we compare a computed random-coil spectrum for horse ferricytochrome c[25] with an experimental spectrum for this protein (Sigma Type III) in neutral D_2O, 5 *M* in urea at 70°. Again the two spectra correspond quite well. Shifts of resonance positions because of the high urea concentration seem to be small although the tyrosine resonance predicted at 1560 Hz appears to be shifted somewhat to lower field. The breadth of the methyl resonance regions is greater than predicted and indicates that some regions of the protein retain elements of structure. We do not believe that the proton resonances of the prosthetic group contribute significantly to the spectral regions shown in Figure 10. It is noted that the heme group contains a paramagnetic iron atom (high-spin ferric) which may cause resonances of the protons on the heme group to be greatly broadened and shifted from their usual positions.[11,12,26] The paramagnetic prosthetic group is not expected to affect many of the resonances of the protons of the polypeptide chain when it is in a random-coil configuration.

(25) E. Margoliash and A. Schejter, *Advan. Protein Chem.*, **21**, 174 (1966).
(26) A. Kowalsky, *Biochemistry*, **4**, 2382 (1965).

Summary

We conclude that the correspondence between spectra computed for random-coil proteins by the procedures introduced here and actual spectra of proteins in extended configurations is sufficient to justify use of the computed spectra to compare with pmr spectra of native proteins in cases where, for one reason or another, random-coil spectra cannot be conveniently obtained experimentally. Indeed, we feel that the computational procedure is sufficiently valid so that if a major difference appears between the computed spectrum and the actual spectrum of a protein in an extended configuration (*e.g.*, the missing methionine resonance in trypsin) it may indicate important residual structure in the protein configuration. We anticipate that the resonance positions and widths used in the computational procedure can be improved as additional comparisons are made with experimental spectra, particularly spectra of oligopeptides. Finally, we caution again that the data provided here cannot be used with confidence to compute random-coil spectra of proteins in solution environments that are considerably different from neutral aqueous solution. For example, it may be seen in Figure 2c that they do not provide a good representation of pmr spectra of random-coil spectra of proteins dissolved in trifluoroacetic aid (TFA). For a solution environment such as TFA, an analogous computational procedure could presumably be used if sets of resonance positions and widths were chosen that are appropriate for this environment. In this regard, we note that pmr spectra of insulin in TFA have been successfully analyzed in detail by comparison of the protein spectra with spectra of amino acids in TFA[27]

(27) B. Bak, C. Dambmann, F. Nicolaisen, and E. J. Pedersen, *J. Mol. Spectry.*, **26**, 78 (1968).