Determination of Amino Acid Sequence in Di- and Tripeptides by Nuclear Magnetic Resonance Techniques[1]

M. Sheinblatt

Contribution from the Department of Biophysics, Weizmann Institute of Science, Rehovoth, Israel. *Received December 30, 1965*

Abstract: A method is proposed for determination of sequence of amino acid residues in di- and tripeptides using the nmr technique. The method is based on the directions of the shifts of the spectral lines as function of pH and amino acid sequence. Several examples are given which verify the method.

In this paper we propose a general method for the determination of the sequence of amino acids in short-chain peptides using the nuclear magnetic resonance technique.

The method is based on the well-known fact that the position of the spectral line of a group adjacent to a potentially ionizable group depends on the state of ionization of that group.[2-4] Removal of a proton of an acidic group by the addition of base causes the spectral line of its neighboring groups to shift toward higher field, while protonation of the basic group causes the spectral lines of the neighboring groups to shift toward lower field. The magnitude of these shifts generally falls off rapidly with increasing distance between the relevant group and the ionizable center.

A neutral peptide molecule in aqueous solution is a zwitterion; *i.e.*, it has a positively charged group (NH_3^+) with a pK value in the range 7–9 and a negatively charged group (COO^-) with a pK value in the range 2–3. The pH of an aqueous solution of a peptide therefore lies between the two pK's. Thus addition of base to an aqueous solution of a peptide neutralizes the NH_3^+ group. Consequently the spectrum of the adjacent group should shift toward higher field. On the other hand, the addition of acid results in the protonation of the ionized carboxylic group and consequently the spectral line of the adjacent group should shift toward lower field.

The above discussion immediately suggests a method for the determination of the sequence of amino acids in short-chain peptides. Consider a dipeptide XY, where X and Y represent two different amino acid residues. If the sequence is XY (X = NH_3^+RCO, Y = $NHR'COO^-$) addition of base will cause the spectral lines of the group R in X (X(R)) to shift toward higher field as compared to the position of the spectral lines of these groups when the molecule is a zwitterion. Addition of acid, on the other hand, should cause the spectral line of the group R′ in Y (Y(R′)) to shift toward lower field. If the sequence of the dipeptide is reversed, YX($NH_3^+R'CONHRCOO^-$) the spectral line of Y(R′) would shift toward higher field with addition of base, while the spectral line of X(R) would shift toward lower field with the addition of acid.

The above case is illustrated in Figures 1 and 2 where the nmr spectra of acidic, neutral, and basic solutions of glycylalanine and alanylglycine are shown. For Gly-Ala the spectral line of the methylene group of glycine shifts upfield in the basic solution, whereas the spectral lines of the α-CH and to a smaller extent the spectral lines of the CH_3 group of alanine shift downfield in the acidic solution. For the case of Ala-Gly, the spectral lines of α-CH and CH_3 of alanine shift toward higher field with addition of base, whereas the spectral line of the CH_2 group of glycine shift downfield with addition of acid.

The sequence of the amino acids in a tripeptide molecule can be determined similarly on the basis of the directions of the shifts of the spectral lines at different pH's. As an example let us consider a tripeptide XYZ composed of three different amino acid residues, X, Y, and Z. If the sequence of this peptide is XYZ (X = NH_3^+RCO, Y = $NHR'CO$, and Z = $NHR''COO^-$), addition of base should cause the spectral lines of R to shift toward higher field, whereas acidification should cause the spectral line of R″ to shift toward lower field. The spectral line of R′ should be affected only slightly by changes of pH, since the distance between Y and the ionizable centers are relatively larger.

When the peptide molecule contains an amino acid which has three functional groups such as aspartic acid or lysine, care should be taken in the analysis of the shifts of the spectral lines, since some of the shifts are due to the ionization or protonation of the third functional group. However, since the pK of the third functional group differs by at least one unit from that of the functional group of the α-CH, the "extra" shift can be readily distinguished. Such a case is exemplified in Table I-1 for the case of aspartic acid in alanylaspartic acid. At pH 3.9, the ionization of the α-carboxylic group is almost complete (pK_1 = 2.8).[5] Hence, there should be no downfield shift of the spectral line of the α-CH as compared to its position at pH 5.9. On the other hand, the β-carboxylic group is almost completely protonated at pH 3.9, since its pK is 4.5.[5] Consequently the spectral line of the β-CH_2 should be shifted downfield as compared to its position at pH 5.9.

In the case of a peptide molecule containing a lysine residue, the spectral line of the ϵ-CH_2 group shifts toward higher field only at pH >9 independently of the

(1) (a) A preliminary report of this work was presented before the International Symposium on Nmr, Tokyo, Japan, Sept. 1–3, 1965. (b) This research was supported in part by the U. S. Public Health Service, National Institutes of Health; Agreement No. 235103.

(2) E. Grunwald, A. Loewenstein, and S. Meiboom, *J. Chem. Phys.*, 27, 641 (1957).

(3) F. Taddei and L. Pratt, *J. Chem. Soc.*, 308, 1553 (1964).

(4) M. Sheinblatt, *J. Am. Chem. Soc.*, 87, 572 (1965).

(5) The above values of pK's are those of glycylaspartic acid and are probably the same for alanylaspartic acid.

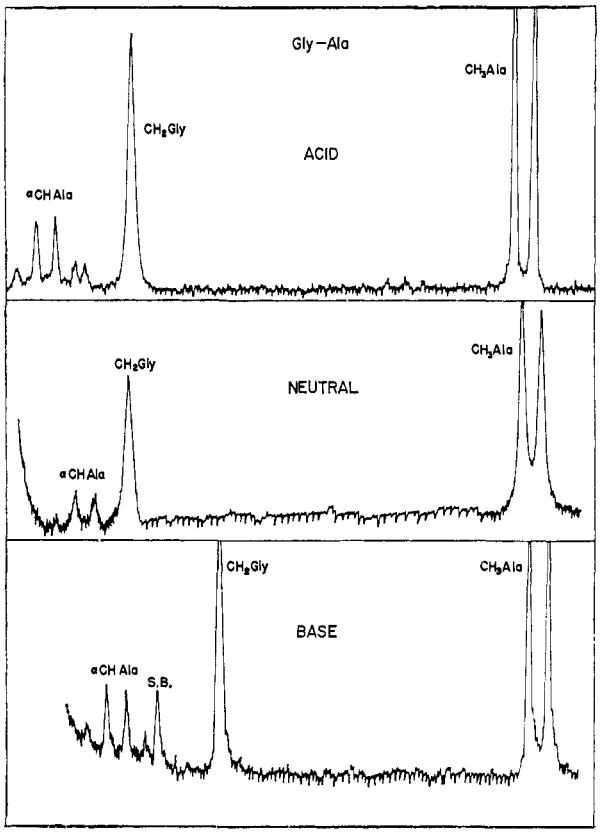

Figure 1. Nmr spectra of D_2O solutions of Gly-Ala at different pH's (the exact pH values are given in Table II-1). The magnetic field increases from left to right. The three spectra are drawn on a common abscissa; the external HDO peak (not seen in the figure) was taken as a common reference.

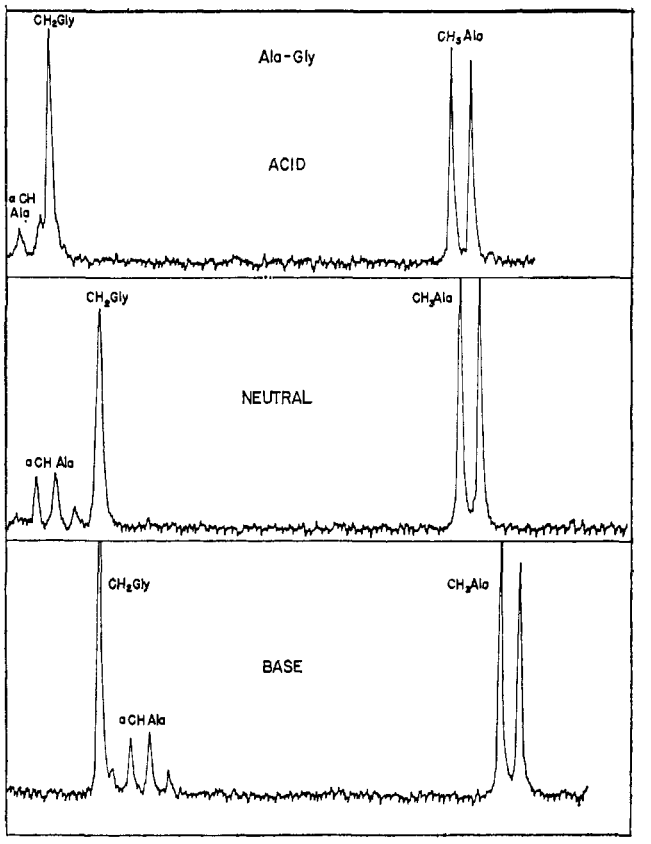

Figure 2. Nmr spectra of D_2O solutions of Ala-Gly at different pH's (the exact pH values are given in Table II-1). The magnetic field increases from left to right. The three spectra are drawn on a common abscissa; the external HDO peak (not seen in the figure) was taken as a common reference.

position of the lysine residue in the molecule. On the other hand the spectral line of the α-CH group at pH's 9 and 0.4 shifts as compared to its position at pH 3.8 up- and downfield, depending on the pH and the position of the lysine residue in the molecule (Table II-4). Thus it is possible to determine the sequence of lysine in the peptide molecule.

Table I. Dipeptides XY[a]

Peptide		α-CH, CH_2X	β-CH_2, CH_3X	α-CH, CH_2Y	β-CH_2, CH_3Y
(1) Alanylaspartic acid	Acid	−1	+1	−20	−18
	Acid*	−2	+2	−2	−16
	Base	+35	+17	+2	+2
(2) β-Alanylglycine	Acid	−5	−2	−18	
	Base	+17	+25	0	
(3) Glycylisoleucine[b]	Acid	−6		−18	−10
	Base	+30		+1	0
(4) Prolylalanine[c]	Acid	−4	−1	−18	−8
	Base	+21	+19	0	0
(5) Glycylthreonine[d]	Acid	−3		−22	−9
	Base	+28		0	0

[a] pH values of the acidic, basic, and the reference solutions for the various peptides were: (alanylaspartic acid) 0.5, 11.1, and 5.3 (acid* −3.9); (β-alanylglycine) 0.5, 12.3, and 6.0; (glycylisoleucine) 0.1, 10.8, and 5.9; (prolylalanine) the exact pH values have not been recorded by a pH meter but only by an indicator; (glycylthreonine) 1.0, 10.6, and 5.8. [b] The shifts of the spectral line of the γ-methylene of the isoleucine residues are acid, −6, and base, 0. [c] The spectral lines of the β- and γ-methylene groups of the proline residue coincide. The shifts of the spectral line of the δ-methylene of the proline residue are acid, −2, and base, +30. [d] The shifts of the spectral line of the γ-methyl group of the threonine residue are acid, −3, and base, 0.

Experimental Section

The high-resolution nmr spectra were obtained with a Varian A-60 spectrometer at room temperature. The pH of a solution was determined with a Radiometer pH meter Type 22r. No correction for pD was made.

Chromatographically pure peptides were obtained from the Yeda Research and Development Co., and Mann Research Laboratories, as well as from private sources.[6]

Results

The shifts of the spectral lines of several dipeptides and tripeptides as functions of pH and amino acid sequence are summarized in the tables given here. The shifts of the spectral lines in these tables are given for acidic and basic solutions, relative to their position in the zwitterionic form.

The degree of acidity and basicity of the solutions were such as to ensure complete protonation of the ionized carboxylic group or the ionization of the NH_3^+ group, respectively. Solutions of sparingly soluble zwitterions were prepared by addition of minimal quantities of HCl. Consequently in these cases the reference state is not exactly the zwitterion. In any case, downfield shifts were always detectable on further addition of HCl. The measured pH of the acid, base, and reference

(6) M. Wilchek, J. Kurtz, and M. Fridkin of the Department of Biophysics, Weizmann Institute of Science, Rehovoth, Israel.

Table II. Dipeptides of the Types XY and YX Where X Stands for Glycine[a]

Y		—CH₂Gly— XY	—CH₂Gly— YX	—α-CHY— XY	—α-CHY— YX	—β-CH₂, CH,CH₃Y— XY	—β-CH₂, CH,CH₃Y— YX	—CH₂,CH₃Y[b]— XY	—CH₂,CH₃Y[b]— YX
(1) Alanine	Acid	0	−18	−14	−5	−3	−3		
	Base	+33	+1	+3	+35	+2	+16		
(2) Hydroxyproline	Acid	0	−18	−12	−5	−12	−1	−4	−3
	Base	+34	−1	+3	+37	+4	+15	+6	+27
(3) Leucine[c]	Acid	0	−18	−8	−3	−4	−3	+1	−2
		(−2)	(−16)			(−4)	(−1)	(+1)	(−1)
	Base	+31	+1	+1	+34	0	+8	0	+1
		(+30)	(0)			(+1)	(+14)	(+1)	(+3)
(4) Lysine[d]	Acid	−13	0	−2	−14	−6	−2	−1	+1
	Base	+7	+23	+27	+3	+2	0	+1	0
	Base*	+10	+33	+45	+3	+15	+5	+27	+27
(5) Phenylalanine	Acid	0	−11	−12	−1	−2	0		
	Base	+33	+4	+1	+36	0	+16		
(6) Proline[e]	Acid	−6	−18	−18	−2	+1	−2	−6	−3
	Base	+26	0	+3	+40	+1	+18	+1	+27
(7) Tryptophan	Acid	0	−13	−5	0	0	0		
	Base	+31	+1	+5	+40	−1	+20		
(8) Tyrosine	Acid	−6	−8	−14	0	−5	+2		
	Base	+28	+3	+1	+38	+2	+20		
(9) Valine	Acid	−5	−16	−19	−4	−6	−2	−5	−1
		(−5)	(−16)			(−5)	(−2)	(−5)	(−1)
	Base	+28	+1	−2	+37	0	+16	−3	+5
		(+29)	(+1)			(0)	(+17)	(−2)	(+7)

[a] pH values of the acidic, basic, and reference solutions of the various peptides were: (Gly-Ala) 0.0, 11.7, and 4.0; (Ala-Gly) 1.0, 12.0, and 6.0; (Gly-Hypro) 0.5, 11.0, and 4.3; (Hypro-Gly) 0.5, 11.5, and 5.0; (Gly-Leu) 2.8 (1.0), 12.0 (10.0), and 4.5 (5.0); (Leu-Gly) 0.7, 12.0, and 5.8; the exact pH value of the H_2O solutions have not been recorded by a pH meter, but only by an indicator; (Gly-Lys) 0.5, 9.0, and 4.1 (base* − 12.3); (Lys-Gly) 0.4, 8.5, and 3.8 (base* −12.8); (Gly-Phe) 0.7, 11.8, and 3.1; (Phe-Gly) 0.2, 12.7, and 4.1; (Gly-Pro) 1.3, 12.2, and 6.3; (Pro-Gly) 1.0, 12.6, and 6.1; (Gly-Try) 0.3, 12.3, and 2.3; (Try-Gly) 0.1, 12.2, and 4.8; (Gly-Tyr) 0.2, 11.9, and 4.2; (Tyr-Gly) 0.3, 11.3, and 6.5; (Gly-Val) 0.7 (0.8), 11.5 (9.5), and 5.5 (5.5); (Val-Gly) 1.0 (0.7), 11.5 (9.3), and 5.8 (5.5). [b] Hydroxyproline-δ-CH_2, leucine-δ-$(CH_3)_2$, lysine-ε-CH_2, proline-δ-CH_2, and valine-γ-$(CH_3)_2$. [c] The spectral lines of the β-CH_2 and γ-CH of the leucine residue coincide. [d] The spectral lines of the β-, γ-, and δ-methylene groups of the lysine residue coincide. [e] The spectral lines of the β- and γ-methylene groups of the proline residue coincide.

Table III. Dipeptides of the Types XY and YX[a]

Amino acid		—α-CHX— XY	—α-CHX— YX	—β-CH, CH₂, CH₃X XY	—β-CH, CH₂, CH₃X YX	—α-CHY— XY	—α-CHY— YX	—β-CH, CH₂, CH₃Y XY	—β-CH, CH₂, CH₃Y YX
(1) X = serine	Acid	−5	−22	−2	−9	−12	−1	−3	−1
Y = alanine	Base	+37	−2	+14	−2	+5	+39	+2	+14
(2) X = phenylalanine	Acid	−6	−16	−1	−6	−15	−6	−4	−3
Y = alanine	Base	+36	0	+16	+1	+4	+36	+2	+21
(3) X = valine	Acid	−6	−15	−4	−9	−16	−2	−7	−1
Y = leucine[b]	Base	+34	−3	+13	−3	+3	+35	−2	+12
(4) X = histidine	Acid	−3	−27	−1	−19	−18	−18	−5	−8
Y = tyrosine[c]	Base	+43	−3	+34	+1	+1	+25	+3	+13
Y = alanine[d]									

[a] pH values of the acidic, basic, and reference solutions of the various peptides were: (Ser-Ala) 1.0, 12.0, and 3.0; (Ala-Ser) 0.5, 11.6, and 5.1; (Phe-Ala) 0.5, 12.1, and 4.0; (Ala-Phe) 0.2, 12.1, and 5.7; (Val-Leu) 0.5, 11.3, and 5.5; (Leu-Val) 1.0, 12.0, and 5.0; (His-Tyr) 0.6, 8.1 and 4.5; and (Ala-His) 0.4, 6.8, and 10.4. [b] The spectral lines of the β-CH_2 and γ-CH of the leucine reisdue coincide. [c] For XY. [d] For YX.

solutions are given in the corresponding tables. The plus and minus signs indicate upfield and downfield shifts. The chemical shifts were measured using either an external reference (HDO) or an internal one, tetramethylammonium chloride (TMA). In some cases both methods were used, and calculated values obtained with the two methods agreed within experimental accuracy. The solvent used was usually D_2O and in some cases H_2O. Unless otherwise specified, the solvent used was D_2O and the chemical shifts were measured from an external HDO reference. When H_2O was used as a solvent and the chemical shifts were measured from an internal TMA reference, the shifts of the spectral lines and the pH's of the solutions are given in parentheses.

In Table I are listed the chemical shifts of the spectral lines of the following dipeptides at different pH's: alanylaspartic acid, β-alanylglycine, glycylisoleucine, and prolylalanine. In Tables II and III the pH dependence of the chemical shifts of 13 pairs of dipep-

Table IV. Tripeptides of Leucine and Two Glycines[a]

Peptide	—α-CH-Leu— Acid	Base	—CH, CH_2-Leu— Acid	Base	—$(CH_3)_2$-Leu— Acid	Base	—CH_2-γ-Gly— Acid	Base	—CH_2-β-Gly— Acid	Base	—CH_2-α-Gly— Acid	Base
Leu-β-Gly-α-Gly	−3	+30	−1	+11	−1	0			0	0	−16	+1
γ-Gly-Leu-α-Gly	−2	+1	−2	0	−2	0	−2	+32			−14	0
γ-Gly-β-Gly-Leu	−12	0	−6	−2	−2	−2	−3	+27	−3	0		

[a] pH values of the acidic, basic, and the reference solutions of the various peptides were: (Leu-Gly-Gly) 0.5, 12.1, and 6.9; (Gly-Leu-Gly) 0.0, 11.8, and 6.8; and (Gly-Gly-Leu) 0.8, 10.6, and 4.0.

Table V. Tripeptides of the Type XY_2 (X-β-Gly-α-Gly)[a]

X	—α-CHX— Acid	Base	—β-CH_2— (CH_3)X Acid	Base	—CH_2— β-Gly Acid	Base	—CH_2— α-Gly Acid	Base
Alanyl	−3	+34	−1	+15	−1	+2	−17	−1
Phenylalanyl	−4	+35	−3	+14	−4	+2	−18	−1
Prolyl[b]	+1	+37	0	+19	0	+4	−10	+5

[a] pH values of the acidic, basic, and reference solutions of the various peptides were: (Ala-Gly-Gly) 0.5, 10.5, and 5.8; (Phe-Gly-Gly) 0.2, 11.8, and 5.1; and (Pro-Gly-Gly) 0.4, 9.6, and 3.6. [b] The spectral lines of the β- and γ-methylene groups of the proline residue coincide. The shifts of the spectral lines of the δ-methylene of the proline residue are acid, 0, and base, +28.

Ala-Ala, Pro-Gly-Ala, and Pro-Gly-Hypro) of various types as function of pH are summarized in Table VI.

It is worthwhile at this stage to discuss briefly the applicability and usefulness of the method proposed here. The great advantages of this method over those which are usually used are its rapidity, simplicity, and nondestructiveness. While the time required for conventional determination of the sequence of a tripeptide molecule is 2 days or more, it requires only 15–20 min by the present method. The peptide solution does not change as a result of the nmr measurements (except for the formation of NaCl due to changes of pH

Table VI. Tripeptides (XYZ) of Various Types[a]

Peptide		X			Y			Z	
		CH_2			CH_2			α-CH	β-CH
	Acid	−2			−4			−15	−8
		(−1)			(−2)				(−6)
(1) Gly-Gly-Val									
	Base	+29			+1			−2	−3
		(+29)			(+2)				(−1)
		β,γ-CH_2		δ-CH_2	β,γ-CH_2		δ-CH_2	CH_2	
(2) Pro-Pro-Gly	Acid	−1		−1	0		0	−10	
	Base	+12		+32	0		−1	+4	
		β,γ-CH_2		δ-CH_2	β,γ-CH_2		δ-CH_2	α-CH	CH_3
(3) Pro-Pro-Ala	Acid	−1		−1	−1		−1	−10	−2
	Base	+12		+32	0		−2	+5	+2
		CH_2		α-CH			CH_3	α-CH	CH_3
(4) Gly-Ala-Ala	Acid	0		0			0	−12	−4
	Base	+29		0			+1	+1	0
		β,γ-CH_2		δ-CH_2	CH_2			α-CH	CH_3
(5) Pro-Gly-Ala	Acid	0		0	0				−5
	Base	+20		+31	+3				0
		α-CH		δ-CH_2	CH_2			α-CH	δ-CH_2
(6) Pro-Gly-Hypro	Acid	−2		0	0			−10	−2
	Base	+40		+36	+2			+1	+7

[a] pH values of the acidic, basic, and the reference solution for the various peptides were (Gly-Gly-Val) 0.5, 11.5, and 5.8. The exact pH values of the H_2O solution have not been recorded by a pH meter, but only by an indicator: (Pro-Pro-Gly) 0.4, 9.5, and 4.4; (Pro-Pro-Ala) 0.5, 12.0, and 3.9; (Gly-Ala-Ala) the exact pH values have not been recorded by a pH meter, but only by an indicator; (Pro-Gly-Ala) 0.5, 10.0, and 5.0; (Pro-Gly-Hypro) 0.5, 12.2, and 4.3.

tides having the reversed sequence are compared. These pairs are: (1) glycine and alanine, (2) glycine and hydroxyproline, (3) glycine and leucine, (4) glycine and lysine, (5) glycine and phenylalanine, (6) glycine and proline, (7) glycine and tryptophan, (8) glycine and tyrosine, (9) glycine and valine, (10) serine and alanine, (11) phenylalanine and alanine, (12) valine and leucine, and (13) histidine and tyrosine or alanine.

The shifts of the spectral lines as a function of pH and the three possible sequences of amino acid residues of tripeptides of leucine and two glycine are summarized in Table IV. In Table V the shifts of the spectral lines as a function of pH of tripeptide molecules of the type XY_2 (Ala-Gly-Gly, Phe-Gly-Gly, Pro-Gly-Gly) are summarized. Shifts of another six tripeptides (Gly-Gly-Val, Pro-Pro-Gly, Pro-Pro-Ala, Gly-

during the measurements) and therefore can be used for further investigations. The disadvantage of the present method is that it requires relatively large amounts of material (in this work 20–25 mg of peptide were used for each sequence determined). The exact amount depends on the apparatus one uses.[7] For example, a full analysis may be performed on approximately 1–2 mg of material if a Varian HA100 spectrometer is used. Use of a time average device (C.A.T.) may reduce the required amount by a factor of 10 or more.

Acknowledgment. The author wishes to thank Drs. M. Wilchek, J. Kurtz, and M. Fridkin for syntheses and the generous supply of some of the peptides used in this study.

(7) For example, A. Kowalsky and M. Cohn, *Ann. Rev. Biochem.*, 43, 490 (1964).