[CONTRIBUTION NO. 1875 FROM THE GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY]

The Constitution of Gelatin. Separation and Estimation of Peptides in Partial Hydrolysates

BY W. A. SCHROEDER, LOIS M. KAY, JOANN LEGETTE, LEWIS HONNEN AND F. CHARLOTTE GREEN

RECEIVED JANUARY 25, 1954

The peptides in partial acidic and basic hydrolysates of gelatin have been separated initially by chromatographing on the ion exchanger Dowex-50, and then further resolved by chromatographing in the form of the DNP-peptides on silicic acid–Celite. In this manner, 34 peptides (mainly dipeptides) have been isolated, identified and quantitatively estimated. These peptides account for a minimum of 42% of the threonine, 27% of the serine, 22% of the proline, 13% of the hydroxyproline and 20% of the glycine and alanine in gelatin. The nature and quantity of the peptides which have been isolated from gelatin in this study as well as those which other investigators have identified do not support the assumption that the structure of collagen and gelatin may be described by the repetitive sequence –P–G–R– in which P represents prolyl or hydroxyprolyl, G glycyl and R other amino acid residues. The analytical data provide some evidence that the sequence -gly-pro-hypro-gly- (or similar sequence) may be of frequent occurrence in gelatin and collagen.

In preliminary reports,[1,2] we have presented the initial results of an exploratory small-scale study of partial hydrolysates of gelatin. This study showed that chromatography on the ion exchanger Dowex-50 was very effective in separating the complex mixture into several discrete peptide zones. Some of these peptide zones were then converted into the dinitrophenyl (DNP) derivatives and further fractionated by chromatography on silicic acid; the individual peptides so isolated were then identified. The further study of these hydrolysates on a larger scale has now resulted in the isolation of 45 peptides of which 34 have been identified with certainty or with reasonable certainty; in addition, many peptides were present in amounts too small to warrant detailed investigation. Throughout this work, stress has been placed upon the acquisition of quantitative data with respect both to the amount of each peptide in the hydrolysate and to the molar proportions of each amino acid in each peptide; such quantitative information is of greater assistance in the unambiguous interpretation of the data than mere qualitative itemizing would be.

In the present report, we present a detailed description of experimental techniques which have been used in this larger scale isolation and identification of the peptides and a discussion of the data obtained, with special reference to their bearing upon the structure of gelatin and collagen.

Experimental

Source and Partial Hydrolysis of Gelatin.—The gelatin which was used for the hydrolyses was Wilson U-COP-CO Special Non-Pyrogenic Gelatin No. 47033 from pigskin.

For the acidic hydrolysis, a 2-g. sample of gelatin was dissolved in 24 ml. of 3.6 *N* hydrochloric acid and heated at 37° for one week. The originally light tan-colored solution rapidly became a deeper orange-brown but underwent no apparent change in color after about a day of heating.

For the basic hydrolysis, 1 g. of gelatin was dissolved in 10 ml. of 0.5 *N* sodium hydroxide and maintained at 37° for one week. The hydrolysate was pale yellow.

Ion Exchange Chromatography of the Hydrolysates.—The initial separation of the components of the partial hydrolysates was made on the ion exchanger Dowex-50 by a procedure which differed only in minor detail from the method of Moore and Stein[3] for the separation of amino acids on 100-cm. columns. The chromatographic column, 3.5 cm. in diameter and 100 cm. long, was prepared exactly as they described. The buffers were adjusted within the suggested limits of *p*H and differed only in that the detergent BRIJ 35 was omitted. The major variation in procedure lies in the fact that the temperature of the chromatograms was not altered as recommended by Moore and Stein but instead was maintained at 37.5° until the change to 0.2 *N* sodium hydroxide at which time the column was allowed to come to room temperature. The flow rate of developer through the columns was adjusted to about 50 ml. per hour and approximately 5-ml. fractions were collected. The buffers were progressively changed as shown in Fig. 2. The progress of the chromatogram was assessed by removing 0.5-ml. portions of alternate fractions and applying the ninhydrin procedure of Moore and Stein.[4]

One advantage of ion exchange chromatography for the separation of peptides from partial hydrolysates lies in the fact that the hydrolysate need not be evaporated before chromatography. A sample of the acidic hydrolysate was prepared for chromatography as follows: 4 ml. of hydrolysate was diluted to 105 ml. with *p*H 4.25 buffer and 1 ml. of 6 *N* hydrochloric acid was added. The *p*H of the solution was 1.4; 100 ml. of this solution (containing the products from 262 mg. of gelatin corrected for moisture and ash) was then placed on the ion exchange column. Likewise, 3.5 ml. of the basic hydrolysate and 2.5 ml. of 6 *N* hydrochloric acid were diluted to 105 ml. with *p*H 4.25 buffer to give a *p*H of 1.75; 100 ml. of this solution (equivalent to 274 mg. of gelatin) was placed on the column. Although the volume of the sample is large, there is no deleterious effect on the subsequent separations if the *p*H of the sample is between 1 and 2.

After a zone had been detected by the ninhydrin method, the appropriate fractions were combined, neutralized to *p*H 6.5–7.5, and evaporated to dryness over coned. sulfuric acid in a vacuum desiccator. The residues were stored at room temperature.

Dinitrophenylation of the Peptides.—As previously discussed,[1] the separation of the peptides from the accompanying buffer salts was accomplished by dinitrophenylation which was carried out in the following way. All or part of the dried residue from the combined fractions of a zone was dissolved in an amount of water which varied from 3 to 15 ml. depending upon the amount of salt to be dissolved. To this solution were added 0.1 g. of sodium bicarbonate and 0.1 ml. of 2,4-dinitrofluorobenzene (DNFB) which had been dissolved in a volume of absolute alcohol equal to twice the volume of water required to dissolve the salts. The mixture was then mechanically shaken for 5 hr. during which time the salts precipitated. The reaction mixture was next transferred to a separatory funnel with 20 ml. of water and extracted as shown in Fig. 1. The extraction of soln. 1 with ether removes excess DNFB. Solution 7 is yellow partly because of the presence of dinitrophenol and partly because the presence of alcohol increases the miscibility of the ether and aqueous phases and thus causes some DNP-compound to be carried into the ether. Acidification of soln. 2 and extraction with ethyl acetate will remove all but long DNP-peptides and DNP-arginyl peptides. If soln. 4 is

(1) W. A. Schroeder, L. Honnen and F. C. Green, *Proc. Nat. Acad. Sci.*, **39**, 23 (1953).

(2) W. A. Schroeder, "The Chemical Structure of Proteins," Ciba Foundation, 1953, p. 184.

(3) S. Moore and W. H. Stein, *J. Biol. Chem.*, **192**, 663 (1951).

(4) S. Moore and W. H. Stein, *ibid.*, **176**, 367 (1948).

very yellow, it probably contains the latter and requires treatment by a procedure to be described below; it is often pale yellow as a consequence of the presence of unidentified materials. Solution 3 was washed to remove any trace of salt; acidification of the water was necessary to prevent extensive return of DNP-peptides to the water phase. The isolation of any DNP-peptides present in soln. 7 was accomplished by re-extracting into water by the use of the washings (soln. 6). Acidification, extraction, washing and combination with soln. 5 then followed. Solution 9 will usually be yellow because of the presence of dinitrophenol which, of course, will turn colorless if acid is added. Solutions 11 and 13 were colorless.

Chromatography of DNP-peptides on Silicic Acid–Celite.—The DNP-peptides which had been prepared and isolated as described in the preceding section were finally chromatographed on silicic acid–Celite for further separation of mixtures. The chromatographic apparatus, the adsorbent, the packing and prewashing of the column, the types of developer and the method of elution are identical with those used by Green and Kay[5] for the separation of DNP-amino acids on silicic acid–Celite. The extension of the methods of Green and Kay to the separation of DNP-peptides has been described previously.[1,6,7] The sequence of chromatograms and developers which is suggested in references 1 and 7 is often sufficient to separate a simple mixture of DNP-peptides into its components but it is sometimes helpful to chromatograph once more and to develop with 12AA6AL,[8] 16AA8AL or a similar developer. The separation of DNP-peptides in this way is generally satisfactory. Even if the interzones between the DNP-peptides are not wide, the mutual contamination is negligible. Sometimes, of course, complete separation cannot be achieved, but even then quantitative data often permit a conclusion as to the nature of the peptides.

When a DNP-peptide was not extracted by ethyl acetate but remained in soln. 4 (Fig. 1), arginyl peptides were usually present and required a different procedure from that described above. The aqueous solution was evaporated to dryness over concd. sulfuric acid and then chromatographed on 1:1 talc–Celite with *N* hydrochloric acid as the developer. In this procedure, which is similar to one of Sanger,[9] the DNP-arginyl peptides remain fixed while the contaminating salts wash into the filtrate. Elution of the DNP-peptides required 4:1 alcohol–6 *N* hydrochloric acid (by volume) and tended to be somewhat incomplete. The separation of the DNP-arginyl peptides was achieved by the use of the M_{66} system of Sanger.[9]

Identification of the DNP-peptides.—After a DNP-peptide had been separated, an aliquot portion was removed, the ultraviolet spectrum was taken, and the quantity was determined. A known amount of peptide was then hydrolyzed in 10 ml. of refluxing doubly glass-distilled 6 *N* hydrochloric acid. The time of hydrolysis was 2 hr. for a DNP-prolyl or -hydroxyprolyl peptide and 4 hr. for all other peptides. As previously pointed out,[1] the spectrum of a DNP-prolyl or -hydroxyprolyl peptide is characteristic and permits ready distinction from other DNP-peptides. A 4-hr. period of hydrolysis has been found to hydrolyze completely all peptides which have been encountered in this work with the exception of DNP-val-gly.[10] Experiments on the hydrolysis of known DNP-glycyl, -prolyl and -hydroxyprolyl peptides[11] have shown that about two-thirds of the DNP-glycine survives a 4-hr. hydrolysis in refluxing 6 *N* hydrochloric acid, so that no difficulty is encountered in detecting peptides with N-terminal glycine. On the other hand, no DNP-proline survives the 2-hr. hydrolysis of a DNP-prolyl peptide under these conditions. However, about 20% of N-terminal DNP-proline is converted to free proline and unless care is taken in the interpretation of the results it may be concluded erroneously that proline occupied a position in the peptide other than N-terminal. When a DNP-hydroxyprolyl peptide is hydrolyzed for two hours, about 40% of the DNP-hydroxyproline escapes destruction and another 25% is converted to free hydroxyproline. Accordingly, DNP-prolyl and -hydroxyprolyl peptides may be detected readily because of their characteristic spectra and may be distinguished easily on the basis of their hydrolytic behavior.

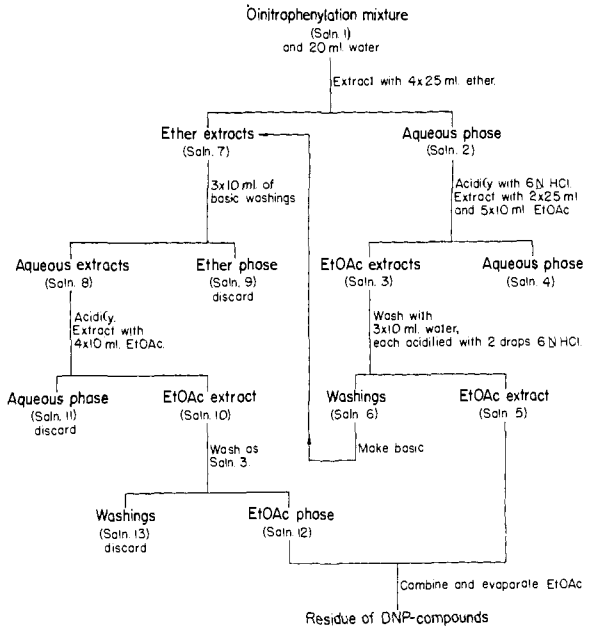

Fig. 1.—Procedure for the extraction of DNP-peptides from the dinitrophenylation mixture.

After the DNP-peptide had been hydrolyzed, the extraction of the N-terminal DNP-amino acid and the subsequent dinitrophenylation of the amino acids which constituted the remainder of the peptide followed the described procedure[6] with the exception that during the latter half of the work the conditions of dinitrophenylation were altered to conform with the best procedure which had been found in studies of the dinitrophenylation of amino acids.[12] The DNP-amino acids so obtained from the peptide were identified by the method of Green and Kay.[5] After identification, the quantity of each amino acid was determined spectrophotometrically as described below.

When DNP-arginyl peptides were being studied a different procedure was necessary. If arginine was N-terminal, the DNP-arginine did not extract from the hydrolysate and had to be separated from the free amino acids by means of talc–Celite as described above. During this procedure the free amino acids passed into the filtrate and could subsequently be dinitrophenylated and studied as required while the arginine was eluted and identified on a Sanger M_{66} column.[9] If arginine was not N-terminal, it was present in the extracted hydrolysate and could then be dinitrophenylated as usual. After extraction of the dinitrophenylation mixture first to remove ether-soluble DNP-amino acids, the aqueous solution of DNP-arginine was passed through talc–Celite to remove salt and the DNP-arginine was then identified on the M_{66} column.

A peptide usually could be identified unequivocally by comparing the molar proportions of the constituent amino acids with the amount of DNP-peptide which had been hydrolyzed. The recovery of the constituent amino acids usually ranged between 50 and 90% of the amount of DNP-peptide hydrolyzed. If the chromatography of the DNP-peptides on silicic acid did not result in complete separation of a mixture, the nature of the DNP-peptides in the mixture often could be deduced from the quantitative data. The more disparate the amounts of the peptides in the mixture, the simpler such an interpretation naturally was.

As another criterion of identity, the chromatographic behavior of each DNP-peptide was compared either with that of the synthetic compound if it was available or with that of

(5) F. C. Green and L. M. Kay, *Anal. Chem.*, **24**, 726 (1952).
(6) W. A. Schroeder, THIS JOURNAL, **74**, 5118 (1952).
(7) W. A. Schroeder and L. R. Honnen, *ibid.*, **75**, 4615 (1953).
(8) Abbreviations as in Green and Kay.[5]
(9) F. Sanger, *Biochem. J.*, **45**, 563 (1949).
(10) The abbreviations and representation of amino acid sequences follow E. Brand (*Ann. N. Y. Acad. Sci.*, **47**, 187 (1946)) and F. Sanger and H. Tuppy (*Biochem. J.*, **49**, 463 and 481 (1951)).
(11) F. C. Green, unpublished results.
(12) W. A. Schroeder and J. LeGette, THIS JOURNAL, **75**, 4612 (1953).

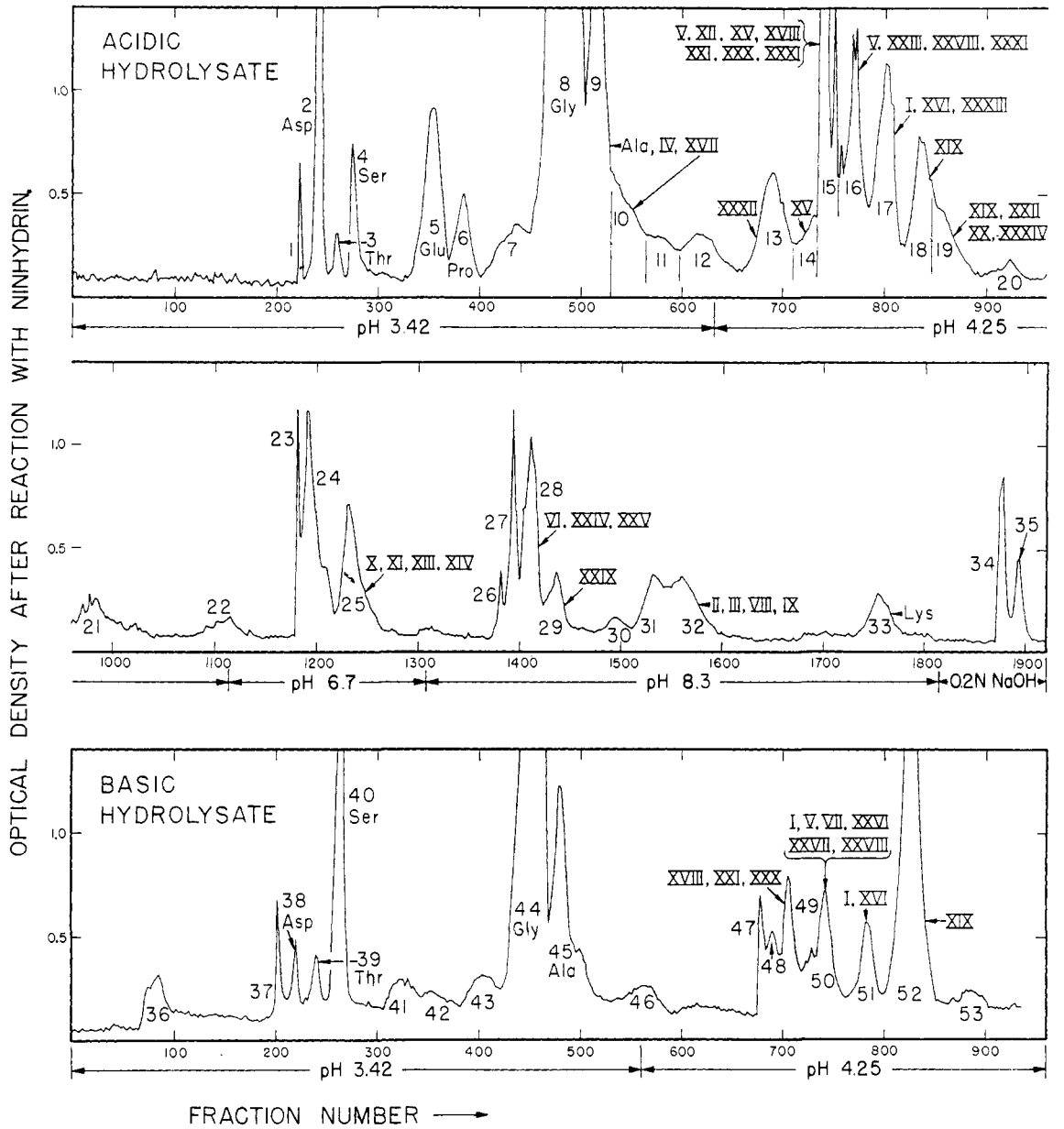

Fig. 2.—Separation of amino acids and peptides in partial acidic and basic hydrolysates of gelatin by means of chromatography on Dowex-50: column dimensions, 3.5 × 100 cm.; temperature of chromatogram, 37.5°; developers, buffers of *p*H shown; fraction size, about 5 ml.; sample, equivalent to about 250 mg. of gelatin.

similar compounds; the comparison was interpreted in the light of the generalizations between the structure of DNP-peptides and their chromatographic behavior as they have been deduced from synthetic peptides.[7] It has been found that in all cases the chromatographic behavior corresponded with that which would be expected of the peptide. This criterion was also valuable in drawing conclusions as to the nature of the peptides in an inseparable mixture: for example, if serine and alanine were found as N-terminal amino acids and threonine and proline as the other amino acids of the peptides, the peptides in such a mixture could only be ser-pro and ala-thr because ser-thr and ala-pro would separate easily.

Quantitative Determinations by Spectrophotometry.—The quantitative determination of all DNP-amino acids and DNP-peptides has been carried out spectrophotometrically in glacial acetic acid. The spectra of all DNP-amino acids and DNP-peptides with the exception of those of DNP-proline, DNP-hydroxyproline and their peptides exhibit a main maximum at 338–343 mμ in this solvent. Hence, the molecular extinction coefficient of a compound with this maximum has been assumed to be 1.61×10^4 liters per mole cm.[6,12] at the maximum except in the instance of lysyl peptides in which the value is assumed to be 3.22×10^4 because of the presence of the two DNP groups. The value 1.75×10^4 was used for DNP-proline, -hydroxyproline and derivatives. The error introduced by the assumption of such average values probably does not exceed ±10%.

Results

Initial Separation of Peptides on Dowex-50.—The results of the large-scale chromatograms of the partial acidic and basic hydrolysates of gelatin are presented in Fig. 2 in which the optical densities, derived from application of the ninhydrin procedure to the fractions, are plotted against the fraction number. All of the zones in Fig. 2 are numbered; the components of each zone are indicated

TABLE I

QUALITATIVE AND QUANTITATIVE DATA ON SOME PEPTIDES IN ACIDIC AND BASIC PARTIAL HYDROLYSATES OF GELATIN

Number	Peptide or amino acid	Isolated from zone no.[a]	μmoles peptide or amino acid per 250 mg. gelatin[b]	Rank in order of decreasing quantity	Percentage[c] of the total amount of amino acid in gelatin accounted for by amount of peptide isolated											
					Ala	Arg	Asp	Glu	Gly	Hypro	Leu (ileu)	Lys	Pro	Ser	Thr	Val
I	Ala-ala	17	4.6	9	3.5											
		50, 51	2.1		1.6											
II	Ala-arg	32	3.0	14	1.1	2.4										
III	Ala-(arg,gly)	32	0.9	27	0.3	0.7			0.1							
IV	Ala-asp	9	1.1	21	0.4		1.0									
V	Ala-gly	15, 16	13.0	5	5.0				1.4							
		50	4.8		1.8				0.5							
VI	Ala-lys	28	1.9	15	0.7							2.4				
VII	Ala-pro	50	3.0	13	1.1								0.9			
VIII	Arg-(ala,gly)	32	0.2	34	0.1	0.2			>0							
IX	Arg-gly-gly	32	0.4	32		0.3			<0.1							
X	Asp-arg	25	1.2	20		1.0	1.1									
XI	Asp-(arg-gly)	25	0.7	30		0.6	0.7		0.1							
XII	Glu-ala	15	6.6	8	2.5			3.5								
XIII	Glu-arg	25	1.8	16		1.5		1.0								
XIV	Glu-(arg,gly)	25	1.0	23		0.8		0.5	0.1							
XV	Glu-gly	14, 15	4.5	10				2.4	0.5							
XVI	Gly-ala	17	9.0	6	3.4				1.0							
		51	6.0		2.3				0.7							
XVII	Gly-asp	9, 10	1.0	24			1.0		.1							
XVIII	Gly-glu	15	7.0	7				3.7	.8							
		49	2.3					1.2	.3							
XIX	Gly-pro	18, 19	10.4						1.2				3.2			
		52	61.8	1					6.9				19.2			
XX	Gly-pro-ala	19	3.5	12	1.3				0.4				1.1			
XXI	Hypro-gly	15	35.6	2					4.0	12.8						
		49	10.5						1.2	3.8						
XXII	Leu (ileu ?)-glu	19	0.4	33				0.2			0.4					
XXIII	Leu (ileu ?)-hypro	16	1.7	18						0.6	1.7					
XXIV	Lys-gly	28	1.8	17					0.2			2.3				
XXV	Pro-(gly,lys)	28	1.0	25					0.1			1.3	0.3			
XXVI	Pro-ser	50	1.0	26									.3	1.3		
XXVII	Pro-thr	50	0.7	29									.2		1.5	
XXVIII	Ser-ala	16	1.5	19	0.6									2.0		
		50	0.5		0.2									0.7		
XXIX	Ser-arg	29	0.9	28		0.7								1.2		
XXX	Ser-gly	15	17.4	4					1.9					23.0		
		49	1.0						0.1					1.3		
XXXI	Thr-ala	15, 16	1.1	22	0.4										2.4	
XXXII	Thr-gly	13	17.5	3					1.9						38.0	
XXXIII	Val-glu	17	0.5	31				0.3								0.7
XXXIV	Val-gly	19	4.1	11					0.5							5.9
Total[d]			211.4		20.4	8.2	3.8	11.6	20.1	13.4	2.1	6.0	22.0	27.5	41.9	6.6
	Ala	9	74[e]		28											
		45	32		12											
	Asp	2	26				25									
		38	6				6									
	Glu	5	35					18								
	Gly	8	240						29							
		44	148						16							
	Lys	33	8									10				
	Pro	6	38										12			
	Ser	4	12											16		
		40	32											42		
	Thr	3	4												9	
		39	7												15	

[a] Zones 1 to 35 inclusive refer to the acidic hydrolysate; zones 36 to 53 to the basic hydrolysate. [b] All values have been recalculated to 250 mg. of gelatin. Quantities of peptide are listed to the nearest 0.1 μmole of peptide and 1 μmole of amino acid. If the same peptide was isolated from adjacent zones, the combined quantity is listed whereas if the same peptide was isolated from both hydrolysates, the quantities are listed separately. [c] The amino acid composition of gelatin as listed by G. Tristram (*Adv. Prot. Chem.*, **5**, 143 (1949)) has been used as the basis of the calculation of these percentages. [d] If the same peptide was isolated from both hydrolysates, the total includes only the larger of the two amounts; thus, the proline of gly-pro from zone 52 was included in the summation whereas that of gly-pro in zones 18 and 19 was omitted. [e] The quantities of the amino acids listed here are based on the data from the colorimetric ninhydrin procedure and are not derived from actual isolation and identification.

by the abbreviation of the name of the amino acid as well as by Roman numerals for the peptides which may be identified by reference to Table I.

These results may be compared with those of the small-scale exploratory chromatograms.[1,2] Such a comparison reveals that the major difference between the large and small-scale work lies in the improved separation of zones on the large chromato-

gram as may be seen by comparing zones 23 to 32 of Fig. 2 with the corresponding zones of ref. 1. (The numbering of zones in the present article does not coincide with that of ref. 1.) Zones 10, 11, 12, 20, 21 and 22 are not even resolved in the small-scale work and zones 6 (proline) and 7 apparently were coincident in the zone labeled "Pro + ?" (ref. 1). The better separations on the large-scale chromatograms may originate in the fact that different lots of Dowex-50 were used in the preparation of the columns.

On the basis of the small-scale work, one would expect an appreciable zone in approximately the position of zone 19 (Fig. 2). The disappearance of this zone apparently is caused by the aging of the hydrolysate. It should be mentioned that when the large-scale column of the acidic hydrolysate was run the hydrolysate had been stored at 5° for 5 months. Except for the disappearance of this zone, however, this aging of the hydrolysate was without apparent effect; a duplicate small-scale chromatogram which was run at the time that the hydrolysate was 2 months old was in excellent agreement with the original except for the loss of this zone. The basic hydrolysate was only 10 days old when the large-scale chromatogram was run.

The position at which a zone would emerge on the larger chromatogram could not be determined exactly from its behavior on the small column. The large column has the same length but 14 times the area of the smaller columns in use in this Laboratory. However, 14 times the volume of developer was not required to effect the emergence of a given zone: aspartic acid needed about 13 times the volume and glycine and alanine only 10 to 11 times the volume. Differences in the packing might well account for the fact that the volume of developer was not in the exact ratio of the sizes of the columns, but this explanation probably does not apply to the observed decrease from zone to zone; the different lots of Dowex-50 which were used in the preparation of the columns may be responsible.

Identity of the Peptides in Zones from the Ion Exchanger.—Table I records the qualitative and quantitative information about those peptides which were identified with certainty or with reasonable certainty. In most instances, these peptides have been obtained in the pure state and it has not often been necessary to interpret the composition of a mixture. Uncertainty sometimes arises not because of the small quantity but because of decomposition during the hydrolysis of the DNP-peptide, a subject which will be discussed below. Some peptides which may be present in the partial hydrolysate but which cannot be identified with much certainty are listed in Table II.

Some of the zones from the ion exchanger have not been investigated further for such reasons as small quantity, poor resolution, etc.; these were zones 11, 20 to 23, 26, 30, 37, 41 to 43, 46 to 49 and 53. Zones 2 to 6, 8, 38, 39, 44 and 45 in all probability contain to a large extent the amino acids indicated in Fig. 2; investigation of zones 9 and 40 showed them to contain largely alanine and serine, respectively, as expected, while zone 33 contained lysine and zone 35 arginine. The compounds in some zones (7, 12, 31 and 34) had undergone extensive or complete decomposition before they could be worked up. Zone 1 contained about 5 compounds in minor amount. In addition to the peptides which were identified in each zone, one or more peptides were present in too small an amount to warrant further study or sometimes complex mixtures were encountered. Zone 24 apparently contained mainly a lysyl peptide but all efforts to ascertain the nature of the remaining amino acids of the peptide proved to be fruitless because of extensive degradation during the hydrolysis of the DNP-derivative. No definite conclusion was reached from the study of zone 36. However, it is probable that longer peptides which contain mainly glycine, proline and hydroxyproline are present.

TABLE II

TENTATIVELY IDENTIFIED PEPTIDES

Peptide	Isolated from zone
Ala-(gly,lys)	28
Ala-hypro	10
Arg-ala	32
Asp-ala	9
Gly-gly	17
Gly-(gly,hypro)	15
Lys-(gly,lys)	27
Pro-ala-gly	19
Pro-glu	17
Pro-hylys	27
Pro-lys	28

Discussion

Quantitative Aspects of the Investigation.—Throughout this investigation stress has been laid on the acquisition of quantitative information. Let us now consider various aspects of the quantitative data which have been accumulated.

After a DNP-peptide had been isolated in as pure a state as possible, a known amount was hydrolyzed and after identification of the constituent amino acids, each was determined quantitatively for comparison with the amount of starting DNP-peptide. In making such a comparison, consideration must be given to losses such as: destruction of the N-terminal DNP-amino acid during hydrolysis of the DNP-peptide,[13] incomplete dinitrophenylation[12] of the free amino acids in the hydrolysate of the DNP-peptide, and losses during manipulation and chromatography.[6] The recovery of the N-terminal amino acid was zero if proline was N-terminal and usually ranged between 50 and 80% uncorrected for chromatographic loss if other amino acids were N-terminal. There is some evidence that the nature of the adjacent amino acid influences the destruction of the N-terminal DNP-amino acid. The recovery of the amino acids other than N-terminal generally was above 75%, especially after the improved procedure of dinitrophenylation[12] was devised. The presence of basic amino acids in the peptide usually resulted in rather low recoveries of all amino acids. Even though the recovery sometimes was low there seldom was reason to believe that any amino acids

(13) R. R. Porter and F. Sanger, *Biochem. J.*, **42**, 287 (1948).

had been overlooked and likewise the molar proportions of the constituent amino acids of the peptide always could be determined.

The chromatographic procedure which was employed to obtain the pure DNP-peptides undoubtedly resulted in losses which may be estimated to approximate 10 to 25%, and in some instances perhaps more.

It is not possible to estimate how completely the peptide zones from the ion exchanger were dinitrophenylated although model experiments[12] lead one to suspect that dinitrophenylation may often be incomplete. The presence of salts in the dinitrophenylation mixture seems to be without adverse effect but the volume of reaction mixture sometimes had to be rather excessive in order to dissolve these salts before the reagents were mixed.

Some conception of the extent to which the materials in the zones from the ion exchanger have been accounted for may be gained by comparing the amounts of DNP-compounds actually isolated with those calculated from the ninhydrin colorimetric data. Table III presents such a comparison.

TABLE III

COMPARISON OF AMOUNTS ISOLATED FROM CERTAIN ZONES WITH THE AMOUNTS CALCULATED BY THE NINHYDRIN PROCEDURE

Zone	Calcd. from ninhydrin[a]	Isolated[a]	$\frac{\text{Isolated}}{\text{Calcd.}} \times 100$
9	75	31	41
10	27	4	15
13	31	18	58
14	12	2	17
15	53	73	138
16	46	15	33
17	38	15	40
18	25	10	40
19	17	9	53
25	25	5	20
28	28	5	18
29	10	1	10
32	15	5	33
40	32	33	103
49	16	14	88
50	16	10	63
51	14	8	57
52	68	62	91

[a] μmoles per 250 mg. of gelatin; values given to the nearest μmole.

The quantities as calculated from the ninhydrin values may well be in error to the extent of ±20% (or more) not only because the color yield has been assumed to be 1.00 relative to leucine (a supposition which may be much in error[14,15]) but also because no effort was made in this work to use the ninhydrin method to obtain very exact results. As much as 10 μmoles should in some instances be added to the quantity isolated in order to take into account many zones which were present in minor amount. The quantities which were isolated vary from a few per cent. to more than 100% of what is calculated to be present. The exceptional recovery from zone 15 is explained by the fact that this zone contains much hypro-gly which produces very little color with ninhydrin. The materials in this zone probably are rather well accounted for and the same is true of zones 40, 49, 50 and 52 but in most of the other zones, the accounting is rather poor. It is difficult to understand why so little was isolated from zone 9 which contains mainly free alanine. Also, the recovery from zone 13 which contained essentially only thr-gly is disappointing. There is no apparent correlation between the percentage isolated and the time elapsed between the ion exchange chromatogram and the study of the zone. That some degradation does occur is shown by the presence of free glycine in zones far removed from zone 8 in which free glycine normally emerges from the ion exchange column. The quantity of glycine in these zones, however, is far too little to account for the "missing" material. Loss of material may have occurred through cyclization of the peptides to form the diketopiperazine. Incomplete dinitrophenylation which may be associated with the structure of a peptide may also be partly responsible for the difficulties. Possibly other methods of preserving the zones until dinitrophenylation would aid in achieving more quantitative results.

The quantitative data which are presented in Table I are minimal values; they are the amounts which were isolated in the final stage of separation immediately before hydrolysis of each peptide. No corrections of any kind have been applied although the preceding discussion indicates that the quantity of some peptides in the hydrolysate actually may have been 50 to 100% greater (perhaps even more). The quantitative data also show that about 20% of glycine, alanine, proline and hydroxyproline (the important amino acids of gelatin) have been found in one peptide form or another.[16] Rather appreciable percentages of these amino acids are also present in the free state although hydroxyproline could not be calculated because its admixture with aspartic acid (zones 2 and 38). Actually only about one-third of the gelatin has been accounted for in the form of free amino acids or peptides. Inability to account for a larger portion lies mainly in deficiencies in technique. It should be noted that no peptides contain histidine, tyrosine, phenylalanine or methionine; the inability to isolate such peptides may in part be due to the small amounts of these amino acids in gelatin.

Conclusions about the Structure of Gelatin.—Proline, hydroxyproline, and glycine comprise so large a fraction of the amino acid residues in gelatin and collagen that it is not surprising to find the suggestion that these proteins are constructed of small repeating units. Thus, Astbury[17] proposed that the repeating unit was the three-amino-acid

(14) Y. P. Dowmont and J. S. Fruton, *J. Biol. Chem.*, **197**, 271 (1952).

(15) M. Ottesen and C. Villee, *Compt. rend. Lab. Carlsberg, Ser. chim.*, **27**, 421 (1951).

(16) As mentioned in footnote *d* of Table I, if the same peptide was isolated from both hydrolysates, the summation of amino acids in peptide form includes the larger of the two amounts. This arbitrary method of summation could, of course, lead to double accounting of an amino acid if, for example, in the sequence -A-B-C-, the peptide A-B were present in one hydrolysate and B-C in the other. Actually, the error would be small except for proline because of the appreciably different amounts of gly-pro in the two hydrolysates.

(17) W. T. Astbury, *J. Intern. Soc. Leather Trades' Chemists*, **24**, 69 (1940).

sequence –P–G–R– where P represents proline or hydroxyproline, G glycine and R one of the remaining residues. Pauling and Corey[18] have used this same sequence for their explicit three-dimensional structure. However, as has been pointed out,[1] the amino acid composition of collagen and gelatin is such that the entire protein could not be composed of such simple units. Collagens and gelatins from a variety of sources have been analyzed by Chibnall,[19] Bowes and Kenten,[20] Neuman[21] and Neuman and Logan[22] and the results of these analyses are gratifyingly concordant. They show that the ratio of proline and hydroxyproline to glycine to other amino acids is very close to 2:3:4. Thus, at most, two-thirds of the proteins could have the sequence –P–G–R–. Let us now consider whether or not the peptides which have been isolated in other investigations and in the present study will permit any definite conclusions as to the structure of these proteins.

Table IV lists the peptides which were isolated by Grassmann and Riederle[23] from an acidic partial hydrolysate of gelatin; by Heyns, Anders and Becker[24] from an alkaline partial hydrolysate of gelatin; and by Kroner, Tabroff and McGarr[25] from an acidic partial hydrolysate of steer hide coliagen. Table I lists the peptides of the present investigation and includes those of a preliminary report.[1] All but six of the peptides in Table IV have also been detected in the present work.

TABLE IV

PEPTIDES FROM GELATIN AND COLLAGEN AS REPORTED IN THE LITERATURE[a]

*Lys-pro-gly[23]		*Ala-gly-ala	
		Gly-ala	
*Ala-ala-gly		*Gly-gly	
Ala-gly		Gly-pro	
*Ala-(gly, glu)	Ref. 24	Hypro-gly	Ref. 25
Glu-gly		*Leu-ala	
Gly-asp		Ser-gly	
		Thr-gly	
		Val-gly	

[a] The asterisk signifies that this peptide was not detected in the present study.

Of the 40 peptides which have been identified, only the following seven agree with the sequence –P–G–R–: ala-pro, gly-ala, gly-asp, gly-glu, hypro-gly, leu(ileu?)-hypro, and lys-pro-gly. Qualitative data permit no definite conclusion about repetitive sequences in gelatin because the small number of agreeing sequences might be quantitatively greater than the disagreeing sequences. Only Grassmann and Riederle[23] and the present paper give quantitative data. Grassmann and Riederle's results indicate that approximately one-third of the lysine is in the sequence lys-pro-gly and hence would account for about 8% of the proline and 3% of the glycine. Our quantitative results show about one agreeing sequence for three which disagree.

It should be mentioned that the most quantitatively important peptide of the present work, gly-pro (which was present in greater amount than all agreeing sequences) was isolated from both the acidic and basic hydrolysates, but in approximately six times the greater amount from the basic hydrolysate. Because of the destruction and interconversion of amino acids which can occur in alkali, it is well to consider whether such reactions might be the source of gly-pro. Wieland and Wirth[26] have shown that when free serine or threonine is heated in hot-saturated barium hydroxide at 110° for 15 hr., glycine is among the products of decomposition but, on the other hand, when cold-saturated barium hydroxide is used, there is no alteration. They state that threonyl peptides form glycine when cold-saturated barium hydroxide is used but give no experimental details. Nicolet and Shinn[27] state that serine is not altered by being boiled for 1 hr. in 0.1 *N* alkali but that, under the same conditions, serine in silk is altered. Sanger and Tuppy[28] found the sequence of gly-pro in alkaline hydrolysates (0.2 *N* sodium hydroxide, 105°, 5 hr.) of the phenylalanyl chain of insulin and concluded that it arose from the sequence thr-pro which was found in acidic hydrolysates. It would appear then that the formation of gly-pro from the sequences -thr-pro- and -ser-pro- must be seriously considered. This possibility can best be discussed in the quantitative terms of a balance sheet of serine and threonine in the hydrolysates as it is given in Table V.

TABLE V

BALANCE SHEET OF SERINE AND THREONINE IN THE ALKALINE HYDROLYSATE

250 mg. of gelatin contain		76 μmoles of serine
Serine free or in peptides from 250 mg.		
Free serine in zone 40	32.5 μmoles	
Pro-ser in zone 50	1.0	
Total	33.5	
Serine unavailable for conversion		33.5 μmoles
Serine available for conversion		42.5 μmoles
250 mg. of gelatin contain		46 μmoles of threonine
Threonine free or in peptides from 250 mg.		
Free threonine in zone 39	7.0 μmoles	
Pro-thr in zone 50	0.7	
Total	7.7	
Threonine unavailable for conversion		7.7 μmoles
Threonine available for conversion		38.3 μmoles

The table shows that in the alkaline hydrolysate of 250 mg. of gelatin, 42.5 μmoles of serine and 38.3

(18) L. Pauling and R. B. Corey, *Proc. Nat. Acad. Sci.*, **37**, 272 (1951).

(19) A. C. Chibnall, *J. Intern. Soc. Leather Trades' Chemists*, **30**, 1 (1946).

(20) J. H. Bowes and R. H. Kenten, *Biochem. J.*, **43**, 358 (1948).

(21) R. E. Neuman, *Arch. Biochem.*, **24**, 289 (1949).

(22) R. E. Neuman and M. A. Logan, *J. Biol. Chem.*, **184**, 299 (1950).

(23) W. Grassmann and K. Riederle, *Biochem. Z.*, **284**, 177 (1936).

(24) K. Heyns, G. Anders and E. Becker, *Z. physiol. Chem.*, **287**, 120 (1951).

(25) T. D. Kroner, W. Tabroff and J. J. McGarr, THIS JOURNAL, **75**, 4084 (1953).

(26) T. Wieland and L. Wirth, *Chem. Ber.*, **82**, 468 (1949).

(27) B. H. Nicolet and L. O. Shinn, *J. Biol. Chem.*, **140**, 685 (1941).

(28) F. Sanger and H. Tuppy, *Biochem. J.*, **49**, 463 (1951).

μmoles of threonine or a total of about 81 μmoles are not accounted for in some form. The acidic hydrolysate contains 1.1 μmoles of thr-ala and 17.5 μmoles of thr-gly, from neither of which could the sequence gly-pro be formed. Even if we assume that the 7.7 μmoles of free threonine and pro-thr in the alkaline hydrolysate arose from sequences which gave thr-ala and thr-gly in the acidic hydrolysate, about 11 μmoles more of threonine in the form of thr-ala or thr-gly are not available for conversion. Hence, at a maximum, 70 μmoles of the sequences -ser-pro- or -thr-pro- are available for conversion to the 62 μmoles of gly-pro which were isolated from zone 52 or the yield would be 89%. Since a yield of 89% is uncommon in most organic reactions and would be unlikely in as complicated a reaction as this, it is reasonable to conclude that a very considerable portion, if not all, of the gly-pro arose from that sequence in gelatin. Furthermore, it should be noted that in the present experiments the duration of the alkaline hydrolysis was much greater than that of Wieland and Wirth, Nicolet and Shinn and Sanger and Tuppy, but that the temperature was only 37° in contrast to 100–110°. If appreciable amounts of the sequences -ser-pro- or -thr-pro- exist in gelatin one would expect to find such peptides in partial acidic hydrolysates because of the lability of peptide bonds involving the amino groups of serine and threonine as found by Desnuelle and Casal[29] and further substantiated by Sanger and Tuppy.[28] Indeed, Sanger and Tuppy[28] note that the thr-pro bond is resistant to hydrolysis and that large quantities of this peptide were present in partial hydrolysates of insulin. However, not these sequences but the opposite ones, pro-ser and pro-thr, were found in small amounts in gelatin. One might expect to isolate appreciable amounts of either gly-pro or pro-gly from gelatin hydrolysates because glycineproline anhydride was isolated many years ago[30] from tryptic and acidic hydrolysates. Furthermore, Kroner, *et al.*,[25] isolated gly-pro from collagen and we have also found it in the acidic hydrolysate in lesser amount. Hence, we may conclude that the sequence -gly-pro- occurs in gelatin and collagen.

If we return now to further comparison of the peptides with the sequence –P–G–R– it will be seen that the sequence -gly-pro- is the opposite of that required and, because at least 20% of the proline in gelatin has been isolated in this form, it may be concluded that a considerably greater percentage probably is present in this linkage in gelatin. On the other hand, in hypro-gly, which is the next most quantitatively important peptide, the sequence is in accord with the proposed –P–G–R– and probably a rather large percentage of the hydroxyproline is present in this sequence. Of the seven peptides which do agree with the sequence, three (gly-ala, gly-asp and gly-glu) contain no proline or hydroxyproline and their accord with the sequence may be fortuitous. If, then, we note how many sequences disagree and realize further that the isolation of the peptides is indicative of a greater quantity in the protein, it seems reasonable to conclude that the repetitive sequence –P–G–R– is not an essential feature of the structure of collagen and gelatin.

The above evidence is negative. Do the available data give any positive clues to the structure of gelatin? In gelatin and collagen, about one-third of the residues are glycine and about another third is composed of approximately equal amounts of proline, hydroxyproline and alanine. If we consider the distribution of these four amino acids among the 40 peptides we see that 23 contain glycine, 7 proline, 2 hydroxyproline, 17 alanine, 8 neither glycine or alanine, and 5 none of the four. Because glycine is such a predominant component of gelatin, it is not surprising to observe it as a constituent of a majority of the peptides. Proline is fairly wide spread among the peptides, although quantitatively more than 90% is associated with glycine. Hydroxyproline in the main is in the form of hypro-gly. Alanine is distributed among the peptides almost to the same extent as glycine and to a much greater extent than proline or hydroxyproline despite the fact that the contents of alanine, proline and hydroxyproline are about equal. It would appear then that glycine and alanine are widely distributed throughout the protein and that proline and hydroxyproline tend to be in combination with glycine.

Another feature which should be noticed in considering the structure of gelatin is the ease with which it is hydrolyzed to dipeptides. Unfortunately, relatively little is known about the lability of various peptide bonds as a function of the amino acids which form the bond. However, as already mentioned, the work of Desnuelle and Casal[29] shows the lability of the peptide bond associated with the amino group of serine or threonine: the isolation of considerable percentages of the seryl and threonyl peptides in the present work offers further substantiation of this fact. The isolation of much hypro-gly is evidence of the lability of the peptide bond associated with the imino group of hydroxyproline. The experiments of Synge[31,32] indicate a considerable lability of the peptide bond associated with the carboxyl group of glycine. The general distribution of hydroxyamino acids and glycine throughout the molecule thus probably is responsible for the rapid breakdown to many dipeptides.

In view of the above facts, the ready isolation of hypro-gly from the hydrolysates is not unexpected. But the even greater quantity of gly-pro implies a still greater lability of the peptide bond associated with the carboxyl group of proline. This conclusion receives substantiation from data in the literature. Consden, Gordon, Martin and Synge[33] concluded that Gramicidin S was a cyclic peptide which had the sequence -val-orn-leu-phe-pro-, but they never were able to isolate pro-val from partial hydrolysates although all other possible dipeptides were identified. Likewise, Sanger and Tuppy[28] isolated all possible dipeptides from acidic hydrolysates of the phenylalanyl chain of insulin with two

(29) P. Desnuelle and A. Casal, *Biochim. Biophys. Acta*, **2**, 64 (1948).
(30) Reviewed by R. L. M. Synge, *Chem. Revs.*, **32**, 157 (1943).
(31) R. L. M. Synge, *Biochem. J.*, **39**, 351 (1945).
(32) R. L. M. Synge, *ibid.*, **44**, 542 (1949).
(33) R. Consden, A. H. Gordon, A. J. P. Martin and R. L. M. Synge *ibid.*, **41**, 596 (1947).

exceptions, pro-lys and gly-ser. On the other hand, small amounts of the rather unusual peptides, pro-ser and pro-thr, have been detected in our partial hydrolysates of gelatin. On the basis of such data, namely that proline and hydroxyproline are present in approximately equal amounts in gelatin, that both seem to be associated largely with glycine, and that a prolyl-hydroxyproline bond probably would be rather labile, one may speculate that sequences of the type -gly-pro-hypro-gly- or -gly-pro-hypro-gly-pro-hypro-gly- may be important in gelatin. Beyond question, a prolyl-hydroxyproline linkage is unusual but collagen and gelatin are unusual proteins because of their high content of proline and because of the mere presence of hydroxyproline. The use of models[34] shows no steric conflicts which would make such a sequence improbable; in fact, models of pro-hypro may be constructed with facility.

If the sequence -gly-pro-hypro-gly were to occur in proteins, it would have interesting structural implications. From molecular models it is clear that the planes of the proline and hydroxyproline rings must be approximately normal to one another because of steric hindrance: thus the sequence -pro-hypro- forces and maintains an approximately 90°-change in the direction of the peptide chain. Although the peptides which have been isolated from gelatin suggest that the sequence -gly-pro-hypro-gly may be present, the models show no reason to exclude the general sequence –R–P–P–R– where R may be any amino acid and P may be either proline or hydroxyproline. If the dark bands which collagen exhibits under the electron microscope represent regions of lesser organization,[35] this lesser organization or more open packing may perhaps be produced by the presence of the sequence –R–P–P–R– which occasions an approximately 90° deviation in the direction of the polypeptide chain.

Acknowledgment.—This investigation was supported in part by a grant from the Lederle Laboratories and in part by a grant-in-aid from E. I. du Pont de Nemours and Company.

(34) R. B. Corey and L. Pauling, *Rev. Sci. Instr.*, **24**, 621 (1953).

(35) An excellent review of the structure of collagen has been given by R. S. Bear, *Adv. Prot. Chem.*, **7**, 69 (1952).

PASADENA 4, CALIFORNIA

[CONTRIBUTION NO. 1882 FROM THE GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY]

The Chromatographic Separation and Identification of Some Peptides in Partial Hydrolysates of Silk Fibroin

BY LOIS M. KAY AND W. A. SCHROEDER

RECEIVED FEBRUARY 26, 1954

By means of chromatography, 12 peptides have been isolated from a partial acidic hydrolysate of silk fibroin, then identified and quantitatively estimated. These peptides account for 60% of the alanine, 50% of the glycine, 47% of the serine, 35% of the tyrosine, 15% of the threonine, 7% of the valine and 5% of the phenylalanine in silk fibroin: these quantities make up almost half of the protein molecule. Few definite conclusions can be drawn from these preliminary experiments about the structure of silk fibroin but they raise some question as to the significance of the minimum repeating sequence -gly-X-ala-gly-ala-gly-X- which has been proposed by Levy and Slobodian.

Synge[1] and Sanger[2] have reviewed the literature pertaining to the amino acid sequence in silk fibroin. There is unequivocal evidence for the presence of ala-gly,[3] gly-ala and gly-tyr in partial hydrolysates of silk fibroin. Other peptides also have been isolated or indirectly estimated[4,5] but their structure has not been demonstrated so convincingly.

In previous investigations in these laboratories, chromatographic methods have been used to separate, identify and estimate 34 peptides in partial hydrolysates of gelatin.[6,7] When these methods were used in small-scale exploratory chromatograms of complete and partial hydrolysates of silk fibroin, the results of Fig. 1 were obtained under the chromatographic conditions given in the legend. The zones which emerge with *p*H 3.42 buffer may with confidence be identified as the free amino acids because the conditions of chromatography are the same as those of Moore and Stein[8] during development with *p*H 3.42 buffer. Zones 2, 3, 4, 7 and 8 have no counterparts in the complete hydrolysate and must be peptides. Experience with gelatin suggested that zone 1 should contain a peptide. Smaller amounts of the amino acids in the corresponding zones of the complete hydrolysate may be present in zones 5 and 6. The partial hydrolysate thus contains a few well-separated zones of peptides in considerable amount. We have chromatographed the partial hydrolysate of silk fibroin on a large scale and shall describe the isolation, identification and estimation of 12 peptides.

Experimental

The silk fibroin was prepared from silk of *Bombyx mori* purchased from a commercial source.

The degumming procedure of Dunn, *et al.*,[9] was followed in detail.

(1) R. L. M. Synge, *Chem. Revs.*, **32**, 135 (1943).

(2) F. Sanger, *Adv. Prot. Chem.*, **7**, 1 (1952).

(3) The abbreviations and representation of amino acid sequences follow E. Brand (*Ann. N. Y. Acad. Sci.*, **47**, 187 (1946)) and F. Sanger (ref. 2).

(4) M. Levy and E. Slobodian, *J. Biol. Chem.*, **199**, 563 (1952).

(5) E. Slobodian and M. Levy, *ibid.*, **201**, 371 (1953).

(6) W. A. Schroeder, L. Honnen and F. C. Green, *Proc. Nat. Acad. Sci.*, **39**, 23 (1953).

(7) W. A. Schroeder, L. M. Kay, J. LeGette, L. Honnen and F. C. Green, THIS JOURNAL, **76**, 3556 (1954).

(8) S. Moore and W. H. Stein, *J. Biol. Chem.*, **192**, 663 (1951).

(9) M. S. Dunn, M. N. Camien, L. B. Rockland, S. Shankman and S. C. Goldberg, *ibid.*, **155**, 591 (1944).