Stark, G. R., Stein, W. H., and Moore, S. (1960), *J. Biol. Chem.*, *235*, 3177.

Steinrauf, L. K., Peterson, J., and Jensen, L. H. (1958), *J. Am. Chem. Soc. 80*, 3835.

Taborsky, G. (1959), *J. Biol. Chem. 234*, 2915.

Tanford, C., and Hauenstein, J. D. (1956), *J. Am. Chem. Soc. 78*, 5287.

Tanford, C., Hauenstein, J. D., and Rands, D. G. (1955), *J. Am. Chem. Soc. 77*, 6409.

Positions of Amino Acids in Mixed Peptides Produced from Collagen by the Action of Collagenase*

JACOB GREENBERG†, LOUIS FISHMAN, AND MILTON LEVY

From the Department of Biochemistry, New York University College of Dentistry, New York City

Received July 27, 1964

Pure collagen (ichthyocol) was nearly completely digested with purified collagenase. The enzymatic hydrolysate contained 220 peptides per 1000 amino acid residues. The enzymatic digest was degraded by stepwise application of the Edman procedure. It appears that glycine is most abundant in positions 1 and 4, proline in position 2, and hydroxyproline in position 3. There is no significant amount of proline in positions 1, 3, or 4 and no hydroxyproline in positions 1, 2, or 4 of the peptides. An equation correcting for lagging and masking effects during stepwise degradation is developed.

This paper describes attempts to find the positions of the various amino acids in the mixed peptides derived from collagen (ichthyocol) by the action of collagenase. The collagenase of *Clostridium histolyticum* was purified to the extent that it would not attack proteins other than collagen or gelatin derived from it. Its specificity, as determined by its action on peptides, involves a sequence of four amino acids given by Mandl (1961) as -X-proline-R_1-R_2-proline-Y-. Hydroxyproline as well as other amino acids may be at R_1. Hydrolysis is most rapid when R_2 is glycine, and it becomes slower as more complex amino acids occupy this position. The bond between R_1 and R_2 is hydrolyzed. R_1 becomes a carboxyl terminus and R_2 becomes an amino end of the resulting peptides. In conformity with this specificity, glycine is the main N-terminus of these peptides, with much smaller amounts of other amino acids (alanine, glutamic and aspartic acids) in this position (Michaels *et al.*, 1958). The preponderant second amino acid is proline (Grassmann *et al.*, 1961). Disagreement appears about the third position, particularly in the role of glycine and hydroxyproline in filling it. Schrohenloher *et al.* (1959) isolated large amounts of glycylprolylhydroxyproline whereas Grassmann *et al.* (1961) found only small amounts of hydroxyproline at the third position by the Edman method but reported that 26% of the total glycine was in this position. The cyclization step of the Edman method is more critical for glycine (Fraenkel-Conrat and Harris, 1954; Margoliash, 1962; Sjöquist, 1957) than for other amino acids and we have paid special attention to this step in our work. It would be expected from the current understanding of the X-ray diffraction pattern of collagen and the specificity of collagenase that the mixture of peptides should contain preponderantly three, six, nine, and so on, residue peptides.

Studies of the present type will not solve the specific sequence of collagen but will provide a kind of pattern to which known sequences should fit and also serve as a guide to the sequences to be looked for in collagen.

* This investigation was supported by research grants from the National Institutes of Health (GM 4902) and the National Science Foundation (GB 1460).

† On leave from Hebrew University, Jerusalem.

Experimental Procedures

Collagen.—Ichthyocol was prepared from carp swim bladders by extraction with citrate buffers and dialysis against dilute disodium phosphate (Gallop, 1955).

Collagenase.—Crude enzyme (Worthington Biochemical Corp.) was purified by chromatography on hydroxylapatite. The purified material contained 150 units (kinetic suspension assay of Gallop *et al.*, 1957) per mg of protein. It did not digest casein under conditions used for collagenase action.

Digestion.—The reaction mixture (806 mg ichthyocol and 3 mg collagenase in 110 ml of water) was brought from *p*H 5.70 to 6.95 by the addition of 0.13 ml of 0.5 N potassium hydroxide. Three drops of toluene were added and the entire solution was kept stirred at 40° and maintained at *p*H 6.95 by the operation of a *p*H-stat. After 18 hours calcium chloride was added to 0.005 M, and after 48 hours to 0.0075 M. Each addition resulted in an increase in the rate of hydrolysis at the time.

After 88 hours the suspension was centrifuged and the clear supernatant fluid was diluted to 120 ml with water. The solution is A in Table I. The precipitate was suspended in 20 ml of 0.005 M calcium chloride, 0.6 mg collagenase, and 5 drops of toluene, and digestion was continued another 21 hours. A very slight suspension remained which gave a tiny dark brown pellet on centrifuging. The clear supernatant was brought to a final volume of 20 ml (B, Table I).

Amino Acid Analysis.—Samples were hydrolyzed in evacuated sealed tubes with 6 N hydrochloric acid, (triple-distilled from glass) at 110° for 20 hours. The hydrolysates were taken to dryness in evacuated (to 1 or 2 mm Hg) desiccators over sodium hydroxide at room temperature. Amino acid analyses were carried out by the method of Piez and Morris (1960) using Aminex-50 × 12, 25 to 32-μ particles (Bio-Rad Co.), and a Technicon automatic analyzer. The precision in our hands is 5% or better and 0.04 μmole is the smallest amount determined with reasonable accuracy.

We report all of the data in moles of constituent per 1000 moles of residues. No corrections for destruction of individual amino acids during hydrolysis were applied.

Edman Degradation.—The method described by Acher *et al.* (1956) as modified by Light and Smith (1960) was used. The peptides (0.1 of the total digest) in 5 ml of water were treated with 5 ml of 1% phenylisothiocyanate in pyridine. The addition of 1 ml of 25% trimethylamine brought the *p*H to 8.5–9.0 and the mixture was maintained at 40° for 2–3 hours with occasional shaking. A further 0.2–0.3 ml of neat phenylisothiocyanate was added and the temperature was maintained at 40° for an additional hour. The mixture was extracted 5 times with 5-ml portions of cyclohexane and then 5 times with 5-ml portions of benzene. The aqueous phase was dehydrated from the frozen state in an evacuated desiccator containing sodium hydroxide flakes.

The dry residue was sealed in glass with 5 ml of a mixture of 5 volumes of glacial acetic acid and 1 volume of concentrated hydrochloric acid at 40° for 6–8 hours. After this cyclization the mixture was frozen and evaporated at once, as before, to avoid side reactions. The dry residue was dissolved in 5 ml of water and the phenylthiohydantoin derivatives were extracted using five 5-ml portions of ether. The cyclization and extraction was repeated. The absolute scale of the degradation was reduced (decreased reagents and the like) at each step in accord with the amount of peptides present.

The derivatives in the ether extract were subjected to paper chromatography on Whatman No. 1 paper as described by Sjöquist (1953) and Edman and Sjöquist (1956). The developed chromatograms were sprayed with iodine-azide reagent. The use of solvent system F (*n*-heptane–ethylene chloride–70% formic acid, 1:2:2) permitted examination of the developed chromatograms in ultraviolet light.

An aliquot of the aqueous residue was hydrolyzed and its amino acid content was determined. The remainder was subjected to the next step of degradation. The data of Table III were calculated from the amino acid analysis after each step so as to maximize the differences (under (max), Table III) or to minimize the differences (under (min), Table III) using principles described. The differences are the amounts of each amino acid removed at that step.

Glycylprolylglycylglycine was obtained from Mann Research Laboratories. Alanylprolylglycine and glycylprolylhydroxyproline were obtained from Cyclo Chemical Corp. The compositions of these peptides were checked by amino acid analysis.

Results

Amounts of Peptides Produced.—No single method gives with certainty the number of moles of peptides in the digest. Application of the ninhydrin method of Moore and Stein (1948) using glycine as standard (the peptides are preponderantly glycyl-), without correction for hydroxylysine or lysine, gave the figures after ninhydrin in Table I. The formol titration at a final formaldehyde concentration of 2 M was done with a Radiometer Titrigraph with endpoints estimated by the method of slopes as suggested by Dunn and Loshakoff (1936).

The results in Table I show that the total digest contained approximately 220 peptides per 1000 residues with an average size of 5. Since there was no important difference in size of peptides between A and B they were combined for subsequent work.

As a subsidiary demonstration of the nature of these peptides it is noted that the *p*H-stat setting at 6.95, with the production of 1500 μmoles of total peptides requiring 202 μmoles of alkali to maintain this *p*H, fit a $pK = 6.95 + \log 1300/200 = 7.83$, which is appropriate for glycylpeptides (see Greenstein and Winitz, 1961). The formol end point was reasonably sharp as is typical for amino groups, whereas prolyl- and hydroxyprolylpeptides react less avidly with formaldehyde, tending to give an obscure end zone in titration.

Table I
Analysis of Collagenase Digestion of Ichthyocol

	(A)	(B)	Sum
Alkali consumed in *p*H-stat, eq × 10^6	165	37	202
Total nitrogen, eq × 10^6	6970	2240	9210
Formol titer, eq × 10^6	1100	400	1500
Formol peptides/1200 N[a]	189	215	195
Average peptide size; residues per end group	5.35	4.65	5.1
Ninhydrin color, glycine, eq × 10^6	1350	340	1690
Ninhydrin peptides/1200 N	231	185	220
Average peptide size; residues per end group	4.3	5.4	4.5
Edman end groups			1690
Edman "peptides"/1200 N			220

[a] In ichthyocol 1200 N atoms are associated with 1000 residues. The sum of the residue weights is 90,600 g (see complete analysis, Table III).

Degradation of Known Peptides.—The sequences glycylprolyl-, glycylprolylhydroxyprolyl-, and possibly -prolylglycyl- occur in the peptides of collagenase digest of collagen in significant amounts. It was necessary to test the quantitation of the Edman procedure on peptides of this type, since glycine is difficult to cyclize and the behavior of hydroxyproline was not well known.

To test the validity of our procedures the technique described was applied to three peptides of known sequence. One to 3 μmoles of peptides were used. The results are shown in Table II. Each step in the Edman

Table II
Edman Degradation of Known Peptides

Peptide	Amino Acids Remaining after: First Step	Second Step
Glycylprolylhydroxyproline		
Recovery[a]	0.84	0.80
Glycine	<0.04	<0.04
Proline	1.02	<0.04
Hydroxyproline	1.00[b]	1.00[b]
Alanylprolylglycine		
Recovery	0.91	0.98
Alanine	<0.04	<0.04
Proline	1.00[b]	<0.04
Glycine	0.94	1.00[c]
Glycylprolylglycylglycine		
Recovery	0.98	0.65
Glycine	2.00[b]	2.00[b]
Proline	1.03	0.12

[a] Moles of key amino acid recovered per mole of peptide used. [b] Calculation base (key) for this step.

TABLE III
RESIDUAL AMINO ACIDS AFTER SUCCESSIVE STEPS OF EDMAN DEGRADATION[a]

Amino Acids	Ichthyocol	Step 1	Step 2 (max)	Step 2 (min)	Step 3 (max)	Step 3 (min)	Step 4 (max)	Step 4 (min)[b]	Corrected for lag
Hyp	67	67 (0)	68 (+1)	71 (+3)	28 (40)	37 (34)	11 (17)	18 (10)	16 (0)
Asp	46	47 (0)	42 (5)	43 (4)	31 (11)	41 (2)	17 (14)	29 (2)	26 (5)
Thr[c]	18	23 (+5)	15 (8)	16 (7)	7 (8)	9 (6)	2 (5)	4 (3)	3 (4)
Ser[c]	16	20 (+4)	21 (+1)	21 (+1)	13 (8)	17 (5)	4 (9)	6 (7)	6 (7)
Glu	70	70 (0)	65 (5)	67 (3)	49 (16)	63 (5)	30 (19)	49 (0)	44 (5)
Pro	131	131 (0)	44 (87)	45 (86)	26 (18)	35 (10)	12 (14)	20 (6)	18 (8)
Gly	356	176 (180)	150 (26)	155 (21)	120 (30)	155 (0)	51 (69)	85 (35)	77 (43)
Ala	124	109 (15)	90 (19)	93 (16)	58 (32)	76 (17)	24 (34)	40 (18)	36 (22)
Val	19	23 (+4)	9 (14)	10 (13)	6 (3)	8 (2)	4 (2)	8 (+2)	8 (+1)
Met	12	7 (5)	10 (+3)	10 (+3)	5 (5)	6 (4)	2 (3)	4 (1)	3 (2)
Ile	11	11 (0)	6 (5)	6 (5)	3 (3)	4 (2)	2 (1)	4 (0)	3 (0)
Leu	21	22 (+1)	9 (13)	9 (13)	6 (3)	8 (1)	3 (3)	5 (1)	5 (2)
Tyr	4	? (?)	2 (?)	2 (?)	0 (2)	0 (2)			
Phe	14	16 (+2)	7 (9)	7 (9)	5 (2)	6 (1)	2 (3)	4 (1)	3 (2)
Hyl[d]	7	tr	1 (?)	1 (?)	tr	tr			
Lys	29	23 ([d])	10 ([d])	10 ([d])	11 ([d])	11 ([d])	6 ([d])	10 ([d])	9 ([d])
His	5	5 (0)	1 (4)	1 (4)	2 (+1)	3 (+2)	0 (2)	0 (2)	0 (2)
Arg	50	50 (0)	49 (1)	50 (0)	27 (23)	35 (15)	17 (10)	28 (+1)	25 (2)
Total	1000	800 (200)	599 (196)	617 (181)	397 (204)	514 (106)	187 (205)	314 (86)	282 (104)
Sample analyzed (μmoles)	7.6	1.87 and 4.52	15.17		11.7		13.02		
Recovery[e]		80	70		80		65		

[a] The figures are residues per 1000 residues in the original peptides. The columns headed (max) and (min) are calculated to give maximal and minimal removal of residues as explained in the text. The parenthetical figures show groups removed in the step. Plus signs indicate apparent gains. [b] The differences entered below are max of step 3 − min of step 4. See text. [c] Not corrected for losses during hydrolysis. [d] Not calculated because of ambiguities in reactions of ε-amino groups. [e] Assumes 200 residues lost in steps 1, 2, and 3 and 86 in step 4.

method removed at least 90% the appropriate end group.

Edman Degradation of Ichthyocol Digest.—Table III is a record of the results of the Edman degradation of the digest. Because of unavoidable losses by adsorption, mechanical inefficiencies of recovery, and unknown losses, we did not attempt calculations of absolute quantities of amino acids in the acid hydrolysates, but correlated them on the basis of ratios of residues in each hydrolysate. For this purpose one amino acid in the hydrolysate became the "key" for further calculation. It was generally the one recovered most completely. The ratios of numbers of residues in the hydrolysate to the key were then calculated. If any of these ratios showed an analytically significant rise the selected key was the wrong one, because the degradation could only remove an amino acid or leave it unchanged. If the selected key happened to have been N-terminal all others would show a rise. If it was not N-terminal only analytical errors could affect the ratios. To minimize the effect of such errors on the conclusions we made calculations with several "keys" and selected the best fit, according to the above criteria, for entry in the tables.

An example of the calculation concerning the third step follows. The relative amounts of amino acids left after the third step are shown in Table III. The total amount of amino acids subject to analysis was 11.7 μmoles. For the minimal calculation we could assume that glycine was not removed in this step. The amount of glycine after the second step (min) was 155 residues. The amount in the analyzed sample after the third step was 3.525 μmoles. Using glycine as the key we should multiply each figure by 155/3.525. The analysis after the third step indicated 0.836 μmole of hydroxyproline. The amount of hydroxyproline left is therefore $0.836 \times (155/3.52) = 37$ residues. The amount removed is then $71 - 37 = 34$ residues. Using glutamic acid as key we should multiply by 67.5/1.44. Applied to glycine this indicated 164 residues left after the third step, but this is an increase over the 155 with which we began the step. Therefore glutamic acid is not a valid key. Choice of hydroxyproline indicated the factor to be 70.5/0.836. Applying this indicates 250 μmoles of glycine, too large an amount according to the analysis after step 2. We therefore prefer glycine (Table III).

The same logic lead us to the choice, for minimal calculation, of arginine as key in the second step and glutamic acid in the fourth.

A maximal calculation starts on the premise that the total number of end groups removed at each step is no greater than the initial number of peptides. Maximally one residue is removed per peptide. Thus originally 1000 residues were in 200 peptides. Removal of one residue from each leaves 800 residues at the end of step 1. Losses are thus assumed to be of the entire mixture.

The number of residues remaining, at the end of step 3, after removing three residues of each peptide $= 1000 - (3 \times 200) = 400$. Since 11.7 μmoles entered the analysis each amount of amino acid found is multiplied by 400/11.7. The analysis indicated 3.52 μmoles of glycine and 0.84 μmole of hydroxyproline. There were therefore 120 residues of glycine and 28 residues of hydroxyproline remaining, and the amount of glycine removed at the third step is $150 - 120 = 30$ and of hydroxyproline $68 - 28 = 40$. At each step these calculations were based on the amounts of peptides actually submitted to the step as determined from the prior step.

Although the phenylthiohydantoin of arginine is not extracted by ether from the aqueous layer (Eriksson and Sjöquist, 1960), the amount converted in the third

step was calculated by subtraction as usual, as justified by our finding that the compound is not significantly hydrolyzed to free arginine at 110° in 24 hours by 6 M hydrochloric acid. Furthermore, no arginine phenylthiohydantoin was observed by paper electrophoresis at *p*H 6.5, after the first and second steps, but was detected by paper electrophoresis after the third step. It is interesting to note that no increase of ornithine, a possible product of the destruction of the arginine derivative (Levy, 1954), occurred in any of the Edman steps.

The disappearances from which we judge the amount of each residue in each position are calculated as differences between both successive minimal and successive maximal estimates. In the first step either calculation gives the same result. The differences are small in the second step and become greater as we go along. We believe the general picture reliable to the fourth step, which needs different treatment to give interpretable numbers; but beyond this the results are not very informative.

To confirm to some degree the nature of the amino acids lost at each step the phenylthiohydantoin derivatives from the step were chromatographed with the results shown in Table IV. The findings aid in the interpretation of the analytical results of Table III. Thus in the third step the maximum calculation indicated loss of 204 residues whereas the minimum indicated 106 residues lost. The amino acids identified as derivatives are mainly hydroxyproline and glycine with smaller spots of proline and alanine. The calculation from minimum levels calls for *no* glycine converted but the maximum indicates 30. We therefore selected the maximum for the third step.

TABLE IV
PHENYLTHIOHYDANTOINS FOUND AFTER EDMAN DEGRADATION OF COLLAGENASE DIGEST

Step	Solvent Systems[a]	Principal PTH[b] Residue	Minor Spots
1	A, F	Gly	
2	A, F	Pro	Gly, Ala
3	A, F	HyPro, Gly	Pro, Ala, (Glu?)
3	Electrophoresis at *p*H 6.5	Arg	
4	A, F	Gly	Ala

[a] Solvent A = *n*-heptane-pyridine, 7:3 (v/v). Solvent F = *n*-heptane–ethylene chloride–75% formic acid, 1:2:2 (v/v), upper phase. [b] PTH = phenylthiohydantoin.

The fourth step also shows large differences between maximum (205) and minimum (86) residues lost. The 200 peptides per 1000 residues in the collagenase digest had an average chain length of 5 residues, and consisted of a mixture of tri- and larger peptides. At the third step the tripeptides in a complete reaction would be completely degraded and their constituents would be completely removed. The fourth step should then show a marked reduction in the number of N-terminal groups. The chief phenylthiohydantoin constituents at step 4 as shown by chromatography are glycine and alanine. No proline, hydroxyproline, aspartic acid, or glutamic acid derivatives were detected in ether extracts using systems A and F or by electrophoresis. Sufficient samples of differing concentration were applied in each case to be detected if either minimum or maximum values were correct. Since none was detected, it appears that the minimum numbers are more valid here The differences entered in Table III under (min) for step 4 are differences between the maxima of the third step and minima of the fourth step. We justify this departure from previous practice by noting that the number of groups removed should decrease drastically between these steps. Step 3 (min) indicates removal of only half the number of groups expected and step 4 (max) indicates removal of much more than was expected. The indicated loss at this point of 35 residues of glycine by the calculation used or 70 by direct differences is larger than any other and confirms the participation of glycine to a large extent in the fourth position.

DISCUSSION

Stepwise degradation is useful in determining the sequence of amino acids in peptides. A limitation on the interpretation of such methods may be serious if the yield at each step is less than 100%. The yield is not likely to be the same for each variety of terminal amino acid and the operator of such a method must give special consideration to the behavior of each kind of amino acid.

The Edman method usually is unequivocal and therefore successful for three or four residues in a peptide. The peptides which escape degradation as well as the less than complete yield from the step under consideration eventually accumulate to a level which defeats the purpose of the method. If one is dealing with an initially pure peptide, isolation of the individual peptides in the degraded mixture (Konigsberg and Hill, 1962) can allow continuing further into the sequence (Smyth *et al.*, 1963). This, of course, is not applicable to the mixture of peptides with which we are dealing. Nevertheless, important insights into the nature of the small peptides present can be had.

Our calculations are based on a knowledge of the total moles of peptides subjected to degradation (see Table I), the reasonable expectation that no increase in total peptides could result from the degradation, that no amino acid after a degradation could increase in amount, and that a phenylthiohydantoin of a particular amino acid would be detectable only if its amino acid occupied a major fraction of the position presumed degraded.

The quantity of each amino acid shown in Table III is in ratio to each of the others as determined by actual analysis, but in absolute amount dependent on the estimate of recovery shown at the bottom of Table III. A further correction comes from the realization that the degradation is not quantitative. It may be assumed for discussion that the yield of removed amino acid remains independent of its nature or its position so that a fraction X of any N-terminal amino acids is removed at a cycle of operations. The fraction not removed is $(1 - X) = Y$ and is also the remaining amount of the original peptide which entered the cycle. Consider a peptide with sequence: ABCDEF.... We may set up a table as follows:

	Remaining Peptide					Removed Amino Acid			
Step	ABCDEF...	BCDEF...	CDEF...	DEF...	EF...	A	B	C	D
1	Y	X				X			
2	Y^2	$2XY$	X^2			XY	X^2		
3	Y^3	$3XY^2$	$3X^2Y$	X^3		XY^2	$2X^2Y$	X^3	
4	Y^4	$4XY^3$	$6X^2Y^2$	$4X^3Y$	X^4	XY^3	$3X^2Y^2$	$3X^3Y$	X^4

TABLE V
POSITION OF VARIOUS AMINO ACIDS IN PEPTIDES

Amino Acids[a]	Residues per 1000 in Digest			
	Position I	Position II	Position III	Position IV
Hyp			55	
Asp		6.2	13.7	3
Thr		8.9	9.6	3
Ser			11	7.6
Glu		6.2	20.6	
Pro		107	1	8
Gly	200	10	36	53
Ala	16.8	21	38	20
Val		16		
Ile		6.2	2.7	
Leu		1.6		3
Phe		11		2
His		3.7	?	(2?)
Arg			31.5	(−8)
Total	217	198	219	100[b]

[a] Amino acids giving zero or questionable numbers omitted. [b] Omits arginine.

The efficiency for step 2 is 0.81. The indicated glycine of position 2 is thus 8/0.81 = 10 residues. In view of the large lag and uncertainties in its estimation this amount may or may not really occupy position 2. It is entered in the table for consistency with experiment and treatment of the data.

The third step in Table III shows 30 glycine residues. These should consist of 200 × 0.009 = 1.8 from the first position plus 10 × 0.162 = 1.6 from the second position; the remainder (26.6) belong in position 3 and represent 26.6/0.729 = 36 residues. Although this number is uncertain, nevertheless it does indicate a considerable amount of glycine (10%) of the total not in the regular pattern usually ascribed to glycine in collagen.

A general formulation of these calculations follows: if A_n is the amount of the amino acid in the position n, ΔA^d is the amount disappearing at step d (as found in Table III), and d, n, F_n^d as have been previously defined. Then:

$$A_n = (\Delta A^d - A_1F_1{}^d - A_2F_2{}^d - \cdots A_{(n-1)}F_{(n-1)})^d/F_n{}^d$$

TABLE VI
AMINO ACID DISTRIBUTION IN COLLAGENASE PEPTIDES

Amino Acids	Position 1 Peptide Base 220		Position 2 Peptide Base 220		Position 3 Peptide Base 220		Position 4 Peptide Base 100		Amino Acids Accounted for (%)
	Occupancy (%)	Specified Amino Acid (%)	Occupancy (%)	Specified Amino Acid (%)	Occupancy (%)	Specified Amino Acid (%)	Occupancy (%)	Specified Amino Acid (%)	
Gly	91	56	4.5	3	16	10	56	16	85
Pro			48	81.5			7	5.5	87
HyPro					25	78			78
Ala	7.5	13.5	9	17	17	31	19	15.5	77
Arg					14	63			63
Peptides (% accounted for)	98.5		61.5		72		82		

If d is the cycle of reaction, counting the first as one, and n is the position of the residue counting A as one, B as two and so on, a general formulation for the fractional yield F_n^d of the amino acid in the nth position at the dth step is:

$$F_n^d = (d - 1)!\ X^nY^{(d-n)}/(d - n)!(n - 1)!$$

The degradation of glycylprolylpeptides shown in Table II justifies $X = 0.9$ as the efficiency of removal of the end groups from peptides in the Edman degradation. The total of end groups removed at each cycle of reactions has contributions not only from the position corresponding to the cycle number but also from several preceding positions because the efficiency of removal is less than one. For the present purpose only the position corresponding to the step and the two preceding it can contribute significantly; others can be neglected at efficiencies of 0.9.

Application of this principle to the data of Table III suggests that the 180 residues of glycine liberated in step 1 correspond to 180/0.9 = 200 residues in position 1, and that 15 residues of alanine found correspond to 17 residues in position 1 as shown in Table V. The sum of these two is in satisfactory agreement with our estimate of total peptides (Table I).

Step 2 in Table III has a contribution of glycine from position 1 of 200 × 0.09 = 18 glycine residues. Of the 26 glycine residues liberated at step two, 18 must belong in position 1 and the remaining 8 may belong in position 2.

Our calculations of step 4 in Table III were based on "minimum number," i.e., no change in glutamic acid. In order to correct for lag with no change in glutamic acid, we should have in Table III for step 4

$$\Delta Glu_4 - Glu_2F_2{}^4 - Glu_3F_3{}^4 = 0$$

The remaining glutamic acid in Table III after step 4 will then be 49 − 5 = 44. Recalculations of the figures in Table III for step 4 are shown in the last columns assuming 90% yields ($F_4{}^4 = 0.656$). The results obtained showed a disappearance of 100 amino acids in the step, that glycine is the predominant amino acid, and that no hydroxyproline is in this position as is indicated in Table V.

We must, however, keep in mind the possibility that the yield (X) may be different for each amino acid at each step. Only the quantitative estimation of the major peptides such as glycylprolylhydroxyproline and glycylprolylalanine can tell us how far we are correct in our assumptions.

It is questionable whether a digest of a less highly ordered protein can be treated by this method. It is likely that the application of the correction factors to the Edman degradation of a single peptide will enable one to go farther with the degradation, since in a pure peptide the yield (X) can be obtained from the analysis at each step.

Table VI shows the distribution of amino acid residues in collagenase peptides. Our findings confirm the importance of hydroxyproline in position three of the

peptides. Glycine, alanine, and arginine also occurred in significant amounts. Glycine is again the predominant amino acid in the fourth position which indicates that a significant portion of the collagen molecule should have the sequence -glycyl-X-Y-glycyl-Z... This agrees with X-ray diffraction findings.

Schrohenloher *et al.* (1959) isolated glycylprolylhydroxyproline containing 23% of the total hydroxyproline from collagenase digest of collagen. This amount is consistent with, but rather lower than, our findings. Low yields are to be expected from their work since it involved several purification steps with accompanying losses. Our findings from Edman degradation enable us to establish a maximum permissible frequency for a given peptide. However part of the 82% of the total hydroxyproline in the third position (reported here) could be in peptides other than glycylprolylhydroxyproline. The glycine at the third position (10% of the total) indicates the existence of -glycylglycyl- sequences in collagen. Kroner *et al.* (1955) reported finding glycylglycine in acid hydrolysates of collagen. The Edman degradation of a mixture of peptides is not in itself conclusive evidence for the existence of any particular peptide but reflects the distribution of amino acids in the peptide mixtures.

Presently accepted interpretations of the X-ray-diffraction patterns and of the "collagen-fold" requirements postulate sequences similar to those established by the present work, but we emphasize that such sequences are not absolutely fulfilled in reality. In particular proline and hydroxyproline do not occupy equivalent positions: most of the proline found is in position two, a position not occupied at all by hydroxyproline whereas most of the hydroxyproline we find in position 3 with little or none in any other position.

The findings have facilitated the examination of peptides separated from the digest of collagen and may serve as a basis for quantitative estimates of particular peptides. It is confirmed that collagen is a highly ordered protein.

References

Acher, R., Laurila, U. R., and Fromagoet, C. (1956), *Biochim. Biophys. Acta 19*, 97.

Dunn, M. S., and Loshakoff, A. (1936), *J. Biol. Chem. 113*, 359.

Edman, P., and Sjöquist, J. (1956), *Acta Chem. Scand. 10*, 1507.

Eriksson, S., and Sjöquist, J. (1960), *Biochim. Biophys. Acta 45*, 290.

Fraenkel-Conrat, H., and Harris, J. I. (1954), *J. Am. Chem. Soc. 76*, 6058.

Gallop, P. M. (1955), *Arch. Biochem. Biophys. 54*, 486.

Gallop, P. M., Seifter, S., and Meilman, E. (1957), *J. Biol. Chem. 227*, 891.

Grassmann, W., Nordwig, A., and Hörmann, N. (1961), *Z. Physiol. Chem. 323*, 48.

Greenstein, J. P., and Winitz, M. (1961), *in* Chemistry of the Amino Acids, Vol. I, Greenstein, J. P., and Winitz, M., eds., New York, Wiley, p. 489.

Konigsberg, W., and Hill, R. J. (1962), *J. Biol. Chem. 237*, 2547.

Kroner, T. D., Tabroff, W., and McGarr, J. J. (1955), *J. Am. Chem. Soc. 77*, 3356.

Levy, A. L. (1954), *Biochim. Biophys. Acta 15*, 589.

Light, A., and Smith, E. L. (1960), *J. Biol. Chem. 235*, 3151.

Mandl, I. (1961), *Advan. Enzymol. 23*, 163.

Margoliash, E. (1962), *J. Biol. Chem. 237*, 2161.

Michaels, S., Gallop, P. M., Seifter, S., and Meilman, E. (1958), *Biochim. Biophys. Acta 29*, 450.

Moore, S., and Stein, W. H. (1948), *J. Biol. Chem. 176*, 367.

Piez, K. A., and Morris, L. (1960), *Anal. Biochem. 1*, 187.

Schrohenloher, R. E., Ogle, J. D., and Logan, M. A. (1959), *J. Biol. Chem. 234*, 58.

Sjöquist, J. (1953), *Acta Chem. Scand. 7*, 447.

Sjöquist, J. (1957), *Arkiv Kemi 11*, 129.

Smyth, D. G., Stein, W. H., and Moore, S. (1963), *J. Biol. Chem. 238*, 227.

Spectral Characterization of Oligonucleotides by Computational Methods

Arnold W. Pratt, J. Nicolet Toal, George W. Rushizky, and Herbert A. Sober

From the Laboratory of Physiology and the Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Md.

Received July 21, 1964; revised October 2, 1964

Computational methods are presented which permit the quantitative determination of base composition of a hydrolyzed oligonucleotide using only spectrophotometric data. The analysis can be done rapidly with a minimal amount of material (absorbance at 260 mμ of 0.5–1.0 in 3 ml) using any selected catalog of pure component spectra, and the linear programming formulation is applicable in the presence of minimal impurities or blanks. The computational methods are suitable for further spectral characterization. These methods have been used on compounds of known base composition and sequence up to the tetranucleotide level; however it is expected that the spectral analysis and characterization of longer-chain compounds will be equally practicable. The computational methods were used to characterize a homologous series of adenine oligoribonucleotides. The entire absorbance function was found to increase linearly with chain length, indicating that the formation of the phosphodiester linkages completely accounted for the absorbance changes seen and that no secondary structure was present up to a chain length of four.

Improved enzymatic hydrolysis and chromatographic separation procedures make it possible to prepare a large number of oligonucleotide fragments of both ribonucleic acid and deoxyribonucleic acids. For example, the two-dimensional "mapping" procedure (Rushizky and Knight, 1960; Rushizky and Sober, 1962a) permits separation and semiquantitative identification of the mono-, di-, tri-, and tetranucleotides obtained from the enzymatic digestions of RNA and DNA. To characterize these oligonucleotides further usually requires separation and quantitative determination of the constituent base composition and the