

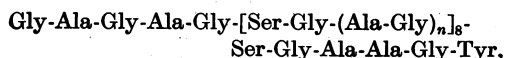
## Some Amino Acid Sequences in the Amorphous Fraction of the Fibroin of *Bombyx mori*

By F. LUCAS, J. T. B. SHAW AND S. G. SMITH  
Shirley Institute, Manchester 20

(Received 9 August 1961)

When an aqueous solution of the fibroin of *Bombyx mori* is treated with chymotrypsin, a white precipitate is formed that contains approximately 60% of the nitrogen of the original fibroin, and the remaining 40% of the nitrogen remains in solution in the form of a mixture of water-soluble peptides. These two fractions, the precipitate and the mixture of soluble peptides, we propose to call fraction Cp and fraction Cs respectively, and these terms replace our earlier nomenclature CTP and CTL, which is no longer acceptable because the former has become the standard abbreviation for cytidine triphosphate. The new system has the advantage that it is applicable by analogy to the fractionation of fibroin by other enzymes: tryptic hydrolysis thus gives rise to fractions Tp and Ts, peptic hydrolysis to fractions Pp and Ps.

The precipitate, fraction Cp, is composed almost entirely of glycine, alanine, and serine, has a molecular weight of about 4000, and has a sequence that can be represented by



where  $n$  is usually 2 and has a mean value of 2 (Lucas, Shaw & Smith, 1957).

The X-ray powder diagram of fraction Cp shows it to be highly crystalline and its reflexions correspond largely with those of powdered undegraded fibroin (Warwicker, 1954). It is likely that it consists of those sections of the polypeptide chains that form the crystalline regions in fibrous fibroin.

Fraction Cs contains all the amino acids with large side chains that occur in fibroin, apart from the tyrosine remaining at the C-terminal position of fraction Cp, and also substantial amounts of glycine, alanine and serine.

Assuming that fraction Cp derives from the crystalline part of fibroin, it follows that fraction Cs comes from the amorphous part. The complete correspondence of these two fractions with the crystalline and amorphous regions cannot, however, at present be proved. This paper deals with the sequence of the three most abundant peptides in fraction Cs. The identity of these three peptides has already been reported briefly (Smith, 1960).

### EXPERIMENTAL

#### Materials

Crystalline chymotrypsin was obtained from The Armour Laboratories and crystalline carboxypeptidase from Worthington Biochemical Corp., N.J., U.S.A. Dry hydrazine was prepared by distilling hydrazine hydrate four times over KOH in an atmosphere of nitrogen. 1-Fluoro-2,4-dinitrobenzene was prepared by the method of Cook & Saunders (1947). Ninhydrin (British Drug Houses Ltd., commercial grade) was used without further purification; hydrindantin was prepared by reduction of ninhydrin with ascorbic acid according to Moore & Stein (1954b). Alloprene, grade D, was a product of Imperial Chemical Industries Ltd.; the fraction that would not pass a 80-mesh sieve was discarded. Brij (a polyoxyethylene lauryl alcohol) was obtained from Honeywell and Stein Ltd. Celite 535 was a product of Johns-Manville and Co. Ltd.

Commercial butan-1-ol was redistilled. Ethyl acetate (B.S.S.) was treated with  $\text{CaCl}_2$  and redistilled. Methyl Cellosolve (2-methoxyethanol) used in the ninhydrin reagent was purified by distillation as described by Eastoe (1955); this procedure considerably reduced the blank values in the ninhydrin method.

**Buffers.** Table 1 gives the composition of the buffers, all those employed on columns for eluting peptides being 0.2M in  $\text{Na}^+$  ions.

**Resins.** Sulphonated polystyrene resin, Zeo-Karb 225, from The Permutit Co. Ltd., was used for the chromatography of peptides, desalting and amino acid analysis. For

Table 1. Composition of buffer solutions

pH		Concn. (g./l.)		Concn. (g./l.)
3.1	NaOH	8.0	Formic acid (90%)	38.2
5.1	Sodium acetate trihydrate	27.22	Acetic acid	4.45
6.0	NaOH	1.81	Potassium hydrogen phthalate	10.2
6.0	$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$	179.0	$\text{Na}_2\text{H}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$	156.0
6.7	$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$	53.72	Citric acid monohydrate	5.25
7.3	$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$	71.6	$\text{Na}_2\text{H}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$	7.8

the last-named the 8% cross-linked resin was used. For chromatography of peptides, a sample of 4½% cross-linked resin, described as having a particle diameter of approx. 50  $\mu$ , was fractionated by wet-sieving. The fraction that did not pass through a 120-mesh sieve was discarded; of the rest, all of which was used, about one-half passed through a 200-mesh sieve. The resin was prepared for use as described by Moore & Stein (1954a) for the similar Dowex 50 resin. The 4½% cross-linked resin, 14–52 mesh, was used for desalting without further fractionation.

The carboxylic resin, Amberlite CG-50 II, passing through a 200-mesh sieve (British Drug Houses Ltd.), was prepared for use by washing with 2N-HCl, followed by 2N-NaOH. It was then washed with water and equilibrated with buffer, pH 3.1 (Table 1).

**Fibroin.** Removal of sericin from Japanese raw silk (obtained from H. T. Gaddum and Co. Ltd., Manchester), and preparation of an aqueous solution of fibroin by means of copper ethylenediamine, normally followed the method given by Drucker, Hainsworth & Smith (1953). In some experiments the aqueous solution of fibroin was prepared by dissolving fibroin (1 g.) in aqueous 5% LiSCN (20 ml.), diluting to 100 ml. with water and dialysing against distilled water until the solution was free from SCN<sup>-</sup> ions.

### Methods

**Preparation of fraction Cs.** Chymotrypsin (25 mg.), dissolved in a few millilitres of water, was added to an aqueous solution of about 10 g. of fibroin buffered at pH 7.8 and the solution was made up to 250 ml. The solution then contained (per litre) 10.82 g. of Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O and 0.442 g. of NaH<sub>2</sub>PO<sub>4</sub>·2H<sub>2</sub>O. A portion of the solution was taken for determination of nitrogen by the Kjeldahl method. The rest of the solution was incubated at 40° for 24 hr., and the precipitate that formed (fraction Cp) was separated by centrifuging; nitrogen determinations on the supernatant showed that 40% of the original nitrogen was present. This supernatant is a solution of fraction Cs.

Fraction Cs was also prepared in a similar way but without the use of buffers. The solution was then maintained at about pH 8 by the occasional addition of approx. 0.05N-NH<sub>3</sub> during the hydrolysis; the amount of NH<sub>3</sub> nitrogen added was so small relative to the fibroin nitrogen that it did not affect the subsequent determinations. After the precipitate had been removed by centrifuging, the liquor was freeze-dried for storage. Nitrogen determinations showed that the same proportion of the fibroin was precipitated in this method as in the preparation in which phosphate buffer was used.

**Large-scale fractionation of fraction Cs.** The primary fractionation was carried out on a column of Zeo-Karb 225, 2.8 cm. diam. and 150 cm. long. The resin was equilibrated with formate buffer, pH 3.1, and the column was prepared from a suspension of the resin in this buffer. A portion (90 ml.) of a solution of fraction Cs containing about 1.4 g. of peptides was brought to below pH 3.1 with a few drops of conc. HCl and applied to the column. Gradient elution from formate buffer, pH 3.1, to acetate buffer, pH 5.1, was started immediately, a mixing chamber (1 l.) being used. At appropriate times the buffer flowing into the mixing chamber was changed to 0.2N-sodium acetate and finally to 0.2N-NaOH, and the initial temperature of 30° was raised to 50°. All the eluting solutions were 0.2N in Na<sup>+</sup>

ions. No wetting agent was used as this interfered with any subsequent dinitrophenylation experiments by causing formation of emulsions during ether extractions. The eluate was collected in fractions (5 ml.) and portions (0.5 ml.) were taken for reaction with ninhydrin (Moore & Stein, 1954b). After development of the ninhydrin colour, the extinctions of the reaction mixtures were measured at 570 m $\mu$ .

Secondary fractionation of fraction C<sub>8A</sub> was carried out on a column of Amberlite CG-50 II, 1.7 cm. in diameter and 110 cm. long, with the formate buffer, pH 3.1, as eluent at 30°. The eluate was collected in fractions (8 ml.) and portions (0.5 ml.) were assayed by the ninhydrin method.

**Small-scale quantitative fractionation.** In order to obtain further quantitative estimates of the peptides in fraction Cs, fractionations were carried out on a smaller scale so that all the effluent from the column could be treated with ninhydrin. A portion of a solution of fraction Cs containing about 1 mg. of nitrogen was applied to a column of Zeo-Karb 225, 0.9 cm. diam. and 150 cm. long. Gradient elution was performed through a mixing chamber (100 ml.), the sequence of inflowing solutions being formate buffer, pH 3.1, acetate buffer, pH 5.1, and finally 0.2N-NaOH. The temperature was kept at 30° throughout. Brij (0.25%, w/v) was included in the solutions. Fractions (2 ml.) were collected and treated with 1 ml. of ninhydrin reagent. The molar yields of the peptides were calculated on the assumption that they had the same colour yield as standard alanine solutions in citrate buffer, pH 3.25 (Moore, Spackman & Stein, 1958) treated in the same way.

**Amino acid analysis.** Peptide fractions were desalted and hydrolysed by boiling with 6N-HCl for 24 hr. The hydrolysates were evaporated to dryness under reduced pressure and the amino acids present established by two-dimensional paper chromatography. Quantitative determinations of the amino acids in the hydrolysates were carried out by a method essentially that of Moore *et al.* (1958), except that the resin used was Zeo-Karb 225 (8% cross-linked), passing 240 mesh, or by a DNP method which was basically that described by Lucas, Shaw & Smith (1955). The mixtures of DNP-amino acids were fractionated on columns of Celite, buffered at pH 7.2 or at pH 6.7, with buffered ether as the moving phase. The DNP-amino acids were extracted into aqueous 1% NaHCO<sub>3</sub> and estimated spectrophotometrically at 360 m $\mu$ . DNP-valine was further purified by reverse-phase chromatography on columns of Alloprene treated with benzyl alcohol and eluted with phosphate-citrate buffer, pH 6.7.

**Determination of N-terminal amino acids and sequences.** These were determined by the DNP method. For N-terminal amino acids, a weighed, dry sample of fraction Cs (about 25 mg.), prepared without buffers, was dissolved in 10 ml. of aqueous 2% NaHCO<sub>3</sub>, and was treated for 3–5 hr. with 100 mg. of fluorodinitrobenzene dissolved in 17 ml. of ethanol. The ethanol was then removed by evaporation under reduced pressure and the excess of reagent was extracted with ether. The remaining aqueous phase was acidified with HCl, when much of the DNP-peptide mixture was thrown out of solution. The whole material was evaporated to dryness under reduced pressure, and the resulting mixture of DNP-peptides, dinitrophenol and NaCl was treated with conc. HCl at 40° for 21 days. The dinitrophenol and the ether-soluble DNP-amino acids were extracted into ether, and were separated and

estimated as described above and by Lucas *et al.* (1955). The *N*-terminal amino acids of the separated individual peptides of fraction Cs were estimated in a similar way. The *N*-terminal sequences of these peptides were found by subjecting the DNP-peptides to partial hydrolysis in conc. HCl at 40°, extracting the DNP-amino acids and DNP-peptides with ethyl acetate and separating them by passage down a column of Celite 535 buffered at pH 7.2, with ethyl acetate as eluent. They were identified by comparison of their  $R_f$  values on paper with standard DNP-amino acids or DNP-peptides, and these identifications were confirmed by further hydrolysis of the peptides to amino acids.

*Determination of C-terminal amino acids.* These were determined by the hydrazine method of Akabori, Ohno & Narita (1952), combined with the method of Blackburn & Lee (1954) for separating the basic hydrazides from the free amino acids. A sample of fraction Cs was heated at 105° in a sealed tube with anhydrous hydrazine, the product was evaporated and the residue dissolved in water and passed down an Amberlite IRC-50 column. The hydrazides were retained and the amino acids passed through; these were identified by paper chromatography. *C*-Terminal amino acids of peptides were also determined by means of carboxypeptidase, as described by Lucas *et al.* (1957).

*Determination of chain length.* A sample of fraction Cs or of a particular peptide, dissolved in water, was suitably diluted and portions were treated with ninhydrin (Moore & Stein, 1954b); the extinction of the coloured solution was measured at 570  $\mu$ . Similar portions were refluxed with 6*N*-HCl for 24 hr., evaporated to dryness and made up to standard volume, and portions of the solutions were assayed with ninhydrin as before. The extinctions of the hydrolysed peptides, corrected for dilution, were compared with those of the original solutions, and the ratio of the extinctions gave the mean chain length of the mixture of peptides or the chain length of the individual peptide.

*High-voltage paper electrophoresis.* A laboratory-constructed apparatus in which the paper was held between two water-cooled Perspex plates was used; the maximum potential gradient applied was 25 v/cm. For DNP-amino acids and DNP-peptides, the paper was moistened with aqueous 0.5% NaHCO<sub>3</sub>, and the substance, dissolved in acetone, was applied as a spot to the paper. For untreated peptides, the electrophoresis was done in the formate buffer, pH 3.1, or in 0.02*N*-acetic acid. The peptides were applied to the paper in aqueous solution and were subsequently revealed by spraying with ethanolic ninhydrin.

*Paper chromatography.* For DNP-amino acids and DNP-peptides, Whatman no. 7 paper, buffered with phthalate buffer, pH 6, was used with 2-methylbutan-2-ol as moving phase; or Whatman no. 1 paper was used with concentrated phosphate buffer, pH 6.0, according to the method of Levy (1954). Untreated amino acids and peptides were separated on Whatman no. 1 paper with the solvent systems butan-1-ol-acetic acid-water (4:1:5, by vol.) and phenol-water (7:3, w/w). Potassium cyanide and NH<sub>3</sub>, or 8-hydroxyquinoline, were added to the latter.

*Desalting.* This was achieved by passing the solutions of the peptides through a small column of Zeo-Karb 225, 14-52 mesh, the resin being in the acid form. The acetate and formate ions were washed through with water until the eluate was neutral, and the peptides were then eluted with

*N*-NH<sub>3</sub> solution, the dark front of the NH<sub>3</sub> on the column indicating when collection of the peptides should begin. The Na<sup>+</sup> ions remained behind on the resin, and the ammoniacal solutions of the desalted peptides were evaporated to dryness at 40° under reduced pressure.

## RESULTS

*N-Terminal residues.* The DNP method gave the results shown in Table 2. The separate identities of the other terminal amino acid residues, amounting to 2.7 moles/10<sup>5</sup> g. (2.9%), were not established. The values quoted in the column headed 'moles/10<sup>5</sup> g.' represent actual recoveries of DNP-amino acids from the hydrolysate of the DNP-peptides, and are not corrected for losses caused by decomposition during hydrolysis. Under our conditions of hydrolysis (conc. HCl at 40°) these losses are of the order of 15%, and DNP-glycine decomposes at a rate only slightly greater than that of most of the other DNP-amino acids. The recovered DNP-amino acids, 91.9 moles/10<sup>5</sup> g., therefore correspond to about  $91.9 \times 100/85 = 108.1$  moles/10<sup>5</sup> g. of the original peptide mixture.

The results, taken in conjunction with the analysis of fraction Cp (Lucas *et al.* 1957) which showed that 84% of the *N*-terminal positions in that precipitate are occupied by glycine, show that hydrolysis of fibroin with chymotrypsin produces fission of bonds that are mainly those connecting glycine and tyrosine residues, and that therefore, in fibroin, tyrosine is linked with glycine much more frequently than it would be if the residues were distributed in a random manner.

*C-Terminal residues.* The *C*-terminal residues of fraction Cs were determined by the hydrazine method. The only amino acids recovered were tyrosine and a small amount of phenylalanine.

*Mean chain length.* By the ninhydrin method the ratio of the extinction of fraction Cs after hydrolysis to that before hydrolysis was 7.5; this gives a measure of the mean chain length. The mean chain length may also be calculated from the yield

Table 2. *N-Terminal residues of the mixture of soluble peptides obtained by the chymotryptic hydrolysis of fibroin*

	Moles/10 <sup>5</sup> g. of fraction Cs	Percentage of total <i>N</i> -terminal residues
Glycine	59.3	64.6
Alanine	5.6	6.1
Serine	9.9	10.8
Threonine	1.7	1.8
Aspartic acid }	8.9	9.7
Glutamic acid }		
Valine	3.8	4.2
Other residues	2.7	2.9
Total	91.9	100.1

of *N*-terminal residues determined by the DNP method ( $108.1 \text{ moles}/10^6 \text{ g.}$ ). Since fraction Cs contains 15.9 g. of  $\alpha$ -amino nitrogen/100 g. of peptides or  $1135 \text{ moles}/10^6 \text{ g.}$ , the mean chain length is  $1135/108.1 = 10.5$ .

#### Primary separation of the peptides

**Large-scale fractionation.** A volume of the liquor containing about 1.4 g. of the mixed peptides was applied to the column. Although the material separated into about 20 peaks (Fig. 1), four of these (C<sub>SA</sub>, C<sub>SB</sub>, C<sub>SC</sub> and C<sub>SD</sub>) predominate.

**Small-scale quantitative fractionation.** The elution curves in these experiments were very similar to the curve obtained in large-scale fractionation, although some of the minor peaks of the latter were not clearly seen. The narrower columns gave slightly greater resolution than the wide preparative column and there were indications that the last main peak C<sub>SD</sub> was separating into three.

From four experiments the mean yields of fractions C<sub>SA</sub> and C<sub>SB</sub> were 3.56 and 1.34 moles of peptide/100 moles of nitrogen in fraction Cs. Since 5% of the total nitrogen in the latter is in

the side chains of amino acids and the mean chain length can be taken as 8, 100 moles of nitrogen represent  $95/8 = 11.9$  moles of peptides. Hence the molar yields of the two fractions are  $3.56 \times 100/11.9 = 30.0\%$  and  $1.34 \times 100/11.9 = 11.3\%$  respectively.

#### Investigation of fraction C<sub>SA</sub>

Preliminary amino acid analyses, paper chromatograms and high-voltage-electrophoresis experiments indicated that fraction C<sub>SA</sub> was composed of two major components, with traces of two others. A solution of fraction C<sub>SA</sub> was therefore applied to a column of the carboxylic resin CG-50 and eluted with formate buffer, pH 3.1. Three clearly defined peaks C<sub>SA1</sub>, C<sub>SA2</sub>, and C<sub>SA3</sub>, were obtained (Fig. 2), together with a rather poorly defined fourth peak, C<sub>SA4</sub>, whose chromatographic behaviour was not reproducible. On the basis of the ninhydrin colour yield, fraction C<sub>SA1</sub> represented 2.8% of the material recovered, C<sub>SA2</sub> 48%, C<sub>SA3</sub> 47% and C<sub>SA4</sub> 2.2%. The total recovery of material was 95% of that applied to the column.

**Fraction C<sub>SA1</sub>.** The material was desalted, hydrolysed and analysed by the DNP method. The

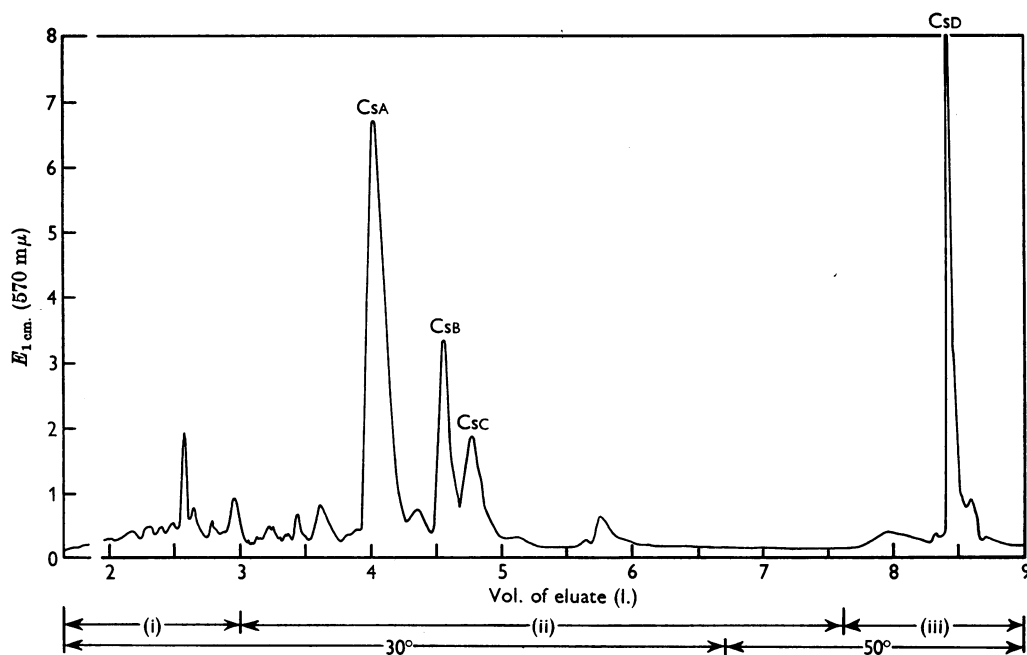


Fig. 1. Primary fractionation of fraction Cs. The liquor (90 ml.; about 1.4 g. of dry peptides) was applied to a column of Zeo-Karb 225, 150 cm. long, 2.8 cm. diam. Gradient elution with a mixing chamber (1 l.): (i) from formate buffer, pH 3.1, to acetate buffer, pH 5.1; (ii) from acetate buffer, pH 5.1, to 0.2M-sodium acetate; (iii) from 0.2M-sodium acetate to 0.2N-NaOH. Portions (0.5 or 0.1 ml.) of the fractions (5 ml.) collected were treated with 1 ml. of ninhydrin reagent and  $E_{1 \text{ cm.}}$  (570  $m\mu$ ) was measured after suitable dilution with ethanol-water (2:1, v/v).  $E_{1 \text{ cm.}}$  was calculated for a standard portion (0.5 ml.) and a standard volume (5 ml.) of diluent.

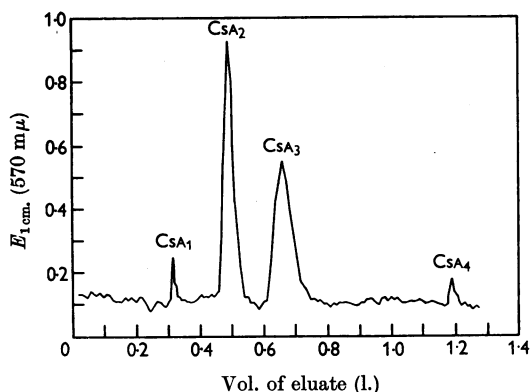


Fig. 2. Chromatography of fraction  $Cs_A$ . Fraction  $Cs_A$  was applied to a column of Amberlite CG-50, 110 cm. long and 1.7 cm. diam. The eluent was formate buffer, pH 3.1, and the temperature  $30^\circ$  throughout. Portions (0.5 ml.) of the fractions (8 ml.) were treated with ninhydrin reagent (1 ml.) and  $E_{1\text{cm.}}$  (570  $m\mu$ ) was measured after dilution with ethanol-water (2:1, v/v; 5 ml.).

main component amino acids were glycine and alanine, but there were small amounts of several other amino acids. The composition was  $Gly_6, Ala_4, X_3$ , where X includes Ser, Asp, Glu, Arg, Tyr, Val, Leu, Ileu and Lys. This peptide is clearly either a complex mixture or a very long peptide.

*Fraction  $Cs_{A_2}$ .* This was desalted and a portion converted into its DNP-derivative. The free peptide and the DNP-peptide were examined by paper chromatography and by paper electrophoresis. A single spot was obtained in each system. The peptide therefore appeared to be pure. It gave a yellow colour, initially, when treated with ninhydrin, suggesting a peptide with a *N*-terminal glycine residue.

Complete hydrolysis of the peptide, by heating at  $100^\circ$  for 24 hr. in 6*N*-HCl, followed by treatment with fluorodinitrobenzene, and separation and estimation of the DNP-amino acids, indicated a composition  $Gly_4, Ala_3, Tyr$ . Hydrolysis of the DNP-peptide showed that glycine was the only *N*-terminal residue. Hydrolysis of the peptide with carboxypeptidase was rapid: after 20 min. only tyrosine, with a faint trace of glycine, was obtained, but after 2 hr., substantial amounts of tyrosine, glycine and alanine had been liberated. The peptide is therefore  $Gly(Ala_3, Gly_3)Tyr$ . The chain length of 8 was confirmed by the ninhydrin method.

When subjected to partial hydrolysis in 12*N*-HCl at  $40^\circ$  for 6 hr. and for 16 hr. the DNP-peptide yielded DNP-glycine, DNP-Gly-Ala, DNP-Gly-Ala-Gly and DNP-Gly( $Ala_2, Gly$ ); the sequence of

the last-named will be DNP-Gly-Ala-Gly-Ala if the DNP-tripeptide derives from the same sequence. Non-terminal amino acids and peptides were glycine, alanine, Gly-Ala, Ala-Gly, Gly-Ala-Gly and Ala-Gly-Ala. No evidence was found for the existence of Gly-Gly in the sequence.

The structure of peptide  $Cs_{A_2}$  is therefore almost certainly Gly-Ala-Gly-Ala-Gly-Ala-Gly-Tyr, although the lability of the Gly-Gly linkage means that the presence of small amounts of such isomeric sequences as Gly-Ala-Gly-Gly-Ala-Gly-Ala-Tyr, although unlikely, cannot be completely ruled out.

*Fraction  $Cs_{A_3}$ .* This was desalted and treated with fluorodinitrobenzene. Both the original peptide and the DNP-derivative ran as single spots on examination by paper chromatography and electrophoresis, and the peptide was therefore presumed to be pure.

After complete hydrolysis with HCl, analysis by the DNP method showed its composition to be  $Gly_4, Ala_2, Val, Tyr$ . Hydrolysis of the DNP-peptide showed that the *N*-terminal residue was glycine. Treatment of the peptide with carboxypeptidase for 20 min. liberated tyrosine only: longer treatment yielded glycine and alanine, but no valine was discerned during treatments lasting up to 2 hr. The chain length of peptide  $Cs_{A_3}$  was found to be 8 by the ninhydrin method.

Partial hydrolysis of the DNP-peptide (6 hr. and 16 hr. in 12*N*-HCl at  $40^\circ$ ) yielded DNP-glycine, DNP-Gly-Ala and DNP-Gly-Ala-Gly. Non-terminal peptides identified included Gly-Ala-Gly, Gly-Ala and Ala-Gly. A peptide containing valine, probably Val-Gly-Ala, was found in small yield, but no Gly-Gly or Ala-Gly-Ala was detected.

These results suggest very strongly that peptide  $Cs_{A_3}$  is similar in composition to  $Cs_{A_2}$  (Gly-Ala-Gly-Ala-Gly-Ala-Gly-Tyr), but that one alanine residue has been replaced by a valine residue. If  $Cs_{A_3}$  is a pure peptide it cannot be Gly-Val-Gly-Ala-Gly-Ala-Gly-Tyr, but must be either Gly-Ala-Gly-Val-Gly-Ala-Gly-Tyr or Gly-Ala-Gly-Ala-Gly-Val-Gly-Tyr. Of these two possibilities the former is slightly to be preferred in view of the results with carboxypeptidase. On the present evidence, however, the possibility cannot be excluded that  $Cs_{A_3}$  is a mixture of these two peptides, or even of all the three isomers. Such mixtures might be exceedingly difficult to separate, and partial hydrolysis of them would be expected to yield just such mixtures of small peptides as were in fact found.

#### *Investigation of fraction $Cs_B$*

Amino acid analysis of the peptide by the DNP method indicated a composition  $Gly_2, Ala, Tyr$ . A more detailed analysis on a resin column

showed the presence of small amounts of other amino acids, the composition being Asp 0.11, Ser 0.11, Glu 0.04, Gly 2.05, Ala 1.07, Ileu 0.04, Leu 0.04, Tyr 1.00. When the fraction was desalted and subjected to paper electrophoresis, one intense spot and one very faint spot were found. The mean chain length by the ninhydrin method was 4.2. This evidence thus indicates that fraction Cs<sub>B</sub> is predominantly a tetrapeptide Gly<sub>2</sub>,Ala,Tyr, containing traces of other peptides. Attempts to separate these components on columns were not successful but the amount of contaminant is so small that it does not interfere with the sequential analysis of the main component.

Treatment of the desalted fraction Cs<sub>B</sub> with carboxypeptidase showed that tyrosine was the first amino acid to be released and is thus in the C-terminal position. Total hydrolysis of the DNP-fraction gave DNP-glycine and no other DNP-amino acid, showing that glycine is N-terminal. Partial hydrolysis of the DNP-fraction in 11N-HCl at 40° for 5 hr. yielded a mixture that was extracted with ethyl acetate, the extract being separated by paper chromatography. Four components were isolated, with the same *R<sub>f</sub>* values as DNP-glycine, DNP-Gly-Ala, DNP-Gly-Ala-Gly and the original DNP-peptide. When eluted from the paper with aqueous 1% NaHCO<sub>3</sub> the first component had an absorption maximum at 360 m $\mu$ , indicating a DNP-amino acid, and the others at 354 m $\mu$ , indicating DNP-peptides. The second substance was hydrolysed completely and only DNP-glycine and free alanine were obtained, in equivalent amounts, confirming that the material was DNP-Gly-Ala. Fraction Cs<sub>B</sub> is therefore predominantly the tetrapeptide Gly-Ala-Gly-Tyr.

#### DISCUSSION

According to the known specificity of chymotrypsin, the peptides that form fraction Cs would all be expected to have C-terminal residues of tyrosine, phenylalanine or tryptophan. Determinations by the hydrazine method confirmed that most of the peptides ended in tyrosine, this being the most abundant of these three amino acids in fibroin, and that a small amount of C-terminal phenylalanine was present. Tryptophan might not be detected by this method, owing to the small amount in fibroin and its extensive decomposition by hydrazine (Locker, 1954).

A calculation of the amino acid composition of fraction Cs (Lucas, Shaw & Smith, 1958) indicated that there were 10.6 residues of tyrosine, 1.8 residues of phenylalanine and 0.5 residue of tryptophan/100 residues. If all these were present in C-terminal positions, the mean chain length of the peptide mixture would be 7.8. The close

agreement between this figure and that determined by the ninhydrin method (7.5) is consistent with the view that most of the tyrosine, phenylalanine and tryptophan of fraction Cs are in C-terminal positions.

The ninhydrin method for chain-length determination is subject to some doubt when a mixture of unknown peptides is being used. Since, however, the majority of the peptides were shown to have N-terminal glycine residues and Harris, Cole & Pon (1956) found that a series of dipeptides with N-terminal glycine had molar-extinction coefficients within 13% of that of alanine, it is likely that no great error is involved in the determination by this method.

For the same reasons the estimation of the yields of the three main peptides, fractions Cs<sub>A2</sub>, Cs<sub>A3</sub> and Cs<sub>B</sub>, which are based on the assumption that these have the same molar extinction coefficients as alanine, are unlikely to be greatly in error.

Two main fractions that contain 39.8 molecules/100 molecules of peptides of fraction Cs consist mainly of three peptides, as shown in Table 3. Thus a large proportion of the amorphous part of fibroin is accounted for by three relatively simple peptide sequences. A striking feature of these sequences is that glycine occupies alternate positions throughout, as it does throughout almost the entire lengths of chain that form the crystalline regions.

Zuber (1958) isolated four neutral peptides containing tyrosine from fraction Cs prepared according to our method, and examined them in a qualitative and semi-quantitative manner. One given by him as (Gly<sub>4</sub>,Ala<sub>3</sub>,Tyr) is probably the same peptide as our Cs<sub>A2</sub>, and another to which he assigns the formula (Gly<sub>5</sub>,Ala<sub>3</sub>,Val,Tyr) may be Cs<sub>A3</sub>.

Cebra (1961) has isolated a number of peptides, prepared by chymotryptic hydrolysis of a solution of fibroin made by the use of aqueous lithium bromide. The enzyme was a commercial preparation of  $\alpha$ -chymotrypsin, hydrolysis was carried out at 37° and the pH was maintained at 7.8 by the intermittent addition of sodium hydroxide. Under these conditions hydrolysis was complete after 6 hr. and the mean chain length of the soluble peptides was found to be 15.5 by the DNP method. Cebra isolated two octapeptides and two tetra-

Table 3. *Main components of the mixture of soluble peptides obtained by chymotryptic hydrolysis of fibroin*

Peptide	Molar percentage of fraction Cs	Structure
Cs <sub>A2</sub>	14.4	Gly-Ala-Gly-Ala-Gly-Ala-Gly-Tyr
Cs <sub>A3</sub>	14.1	Gly-(Gly <sub>3</sub> ,Ala <sub>2</sub> ,Val)-Tyr
Cs <sub>B</sub>	11.3	Gly-Ala-Gly-Tyr

peptides which he identified as: (1) Gly-(Gly<sub>3</sub>, Ala<sub>3</sub>)-Tyr; (2) Gly-(Gly<sub>3</sub>, Ala<sub>2</sub>, Val)-Tyr; (3) Gly-(Gly, Ala)-Tyr; (4) Gly-(Gly, Val)-Tyr; but he did not further investigate the sequences of the amino acids in them. The percentage of the total 'soluble peptide nitrogen' represented by these four fractions was 7.22, 6.13, 1.48 and 0.70 respectively. Cebra was also able to demonstrate the presence of dodecapeptides containing, besides glycine, alanine and valine, two tyrosine residues per molecule; their presence clearly showed that, under the conditions he employed, some of the peptide links involving the carboxyl groups of tyrosyl residues had not been hydrolysed. He suggested that the dodecapeptides were of the form: Gly-(Gly<sub>3</sub>, Ala<sub>3</sub>)-Tyr-Gly-(Gly, Val)-Tyr and Gly-(Gly<sub>3</sub>, Ala<sub>2</sub>, Val)-Tyr-Gly-(Gly, Ala)-Tyr; together these dodecapeptides represented approximately 27% of the soluble peptide nitrogen.

Although Cebra's results can be said to corroborate the results presented in this paper in all essential respects, there are various points of difference. Under our conditions hydrolysis with chymotrypsin is more nearly complete, for the mean chain length of the soluble peptides measured by the DNP method is 10.7 compared with Cebra's value of 15.5; and by another method we have obtained a result in substantial agreement with the value of 8 expected from the content of peptide bonds labile to chymotrypsin. It seems probable that with this mixture of peptides, assays of chain length by the DNP method give high results, but the cause cannot at present be determined. Nevertheless we have observed some slight variability in the percentage molar yield of the two octapeptides and of the principal tetrapeptide Gly-Ala-Gly-Tyr, and this variability may well reflect a resistance to chymotryptic cleavage of a proportion of the theoretically vulnerable bonds. If this is so we may expect to encounter small amounts of the dodecapeptides described by Cebra on further investigation of the slower-running peaks of our chromatograms. However, the molar ratios of the octa-

peptides to the tetrapeptide Gly-Ala-Gly-Tyr found by Cebra, i.e. 9.0:7.7:3.6, are not grossly dissimilar from our values of 14.4:14.1:11.3, and as Cebra's dodecapeptides give rise on further chymotryptic attack to equimolar amounts of octa- and tetra-peptide, the discrepancies between the two sets of results are readily explained.

Although glycine alternates throughout the sequences given in Table 3, and throughout fraction Cp (except for one position), it cannot do so in the remaining sequences that await elucidation. It can be seen from Table 4 that glycine accounts for only about one-third of the residues in these remaining sequences. Ziegler & La France (1960) have in fact already isolated from fraction Cs, by paper electrophoresis, a number of peptides containing no glycine, e.g. Ser-Glu-Asp-Tyr, Val-Lys-Phe, Ileu-Thr-Ala-His and Glu-Tyr.

The two octapeptides C<sub>SA2</sub> and C<sub>SA3</sub> found in this investigation are present in equimolar amounts and these amounts bear a ratio of 4:3 with the amount of the tetrapeptide C<sub>SB</sub>. In the absence of a satisfactory value for the molecular weight of fibroin it is not possible to calculate the number of these peptide sequences that occur in each molecule of the original protein. It is interesting, however, to determine the relation between the amounts of these peptides and that of the long peptide of fraction Cp, the major structural element of fibroin.

If 100 nitrogen atoms in fibroin divide so that 60 are in fraction Cp and 40 in fraction Cs, then, since the latter contains all the amino acids contributing side-chain nitrogen, which amounts to 2% of fibroin nitrogen (Lucas *et al.* 1958), we can say that 98 amino acid residues in fibroin divide so that 60 are in fraction Cp and 38 in fraction Cs. Of 100 residues in fibroin therefore 61.2 are in fraction Cp and 38.8 are in fraction Cs. Since the mean chain length of fraction Cp is 59 and of fraction Cs about 8, 61.2 residues in the former represent 1.04 molecules, and 38.8 residues in the latter correspond to 4.9 molecules. Since C<sub>SA2</sub>, C<sub>SA3</sub> and C<sub>SB</sub>

Table 4. *Distribution of amino acid residues among fractions of fibroin*

Amino acid residues (100) in fibroin are fractionated under the action of chymotrypsin into 61.2 residues in the precipitate (fraction Cp) and 38.8 residues in the liquor (fraction Cs) with the distribution shown.

Residue	Fibroin	Fraction Cp	Fraction Cs		
			Total	Peptides C <sub>SA2</sub> + C <sub>SA3</sub> + C <sub>SB</sub>	Remainder
Glycine	44.6	30.1	14.5	6.7	7.8
Alanine	29.3	20.7	8.6	4.1	4.5
Serine	12.1	9.4	2.7	0.0	2.7
Valine	2.2	0.0	2.2	0.7	1.5
Tyrosine	5.2	1.0	4.2	1.9	2.3
Others	6.6	0.0	6.6	0.0	6.6
Total	100.0	61.2	38.8	13.4	25.4

represent 14.4, 14.1 and 11.3% of fraction Cs respectively, the number of molecules/100 residues in fibroin is therefore: fraction Cp, 1.04; C<sub>SA</sub><sub>2</sub>, 0.71; C<sub>SA</sub><sub>3</sub>, 0.69; C<sub>SB</sub>, 0.55. This corresponds approximately to a molar ratio of fractions Cp:C<sub>SA</sub><sub>2</sub>:C<sub>SA</sub><sub>3</sub>:C<sub>SB</sub> of 6:4:4:3. Thus within the limits of the errors involved there is evidence of stoichiometry in the amounts of the principal elements of the fibroin molecule.

At this juncture one can only speculate on the relative positions in the fibroin molecule occupied by the various peptides that make up fraction Cs. However, evidence gained from the action of chymotrypsin on the precipitate produced by trypsin from a fibroin solution may be significant. Thus Zuber (1958) showed that the neutral peptides isolated by him from fraction Cs were similar to those he isolated by the action of chymotrypsin on a tryptic precipitate from fibroin. Similarly, work in progress in our laboratories has indicated that when the precipitate produced by the action of trypsin on fibroin solution has been further divided by subjecting it subsequently to the action of chymotrypsin, that part of the tryptic precipitate that remains after such treatment is identical with fraction Cp, whereas the peptides in solution contain mainly glycine, alanine, valine and tyrosine.

These facts, taken in conjunction with the molar ratios of fraction Cp and the simple neutral peptides described here, suggest that fraction Cp and these peptides are in close juxtaposition in the fibroin molecule and they may form a transitional phase between the regions of high order represented by fraction Cp and those of low order represented by the sequences containing mainly amino acids with bulky side chains. In such a model serine residues are concentrated in the regions of highest and lowest crystallinity.

#### SUMMARY

1. When the precipitate formed by the action of chymotrypsin on an aqueous solution of silk fibroin has been removed, the liquor remaining consists of a solution of peptides having a mean chain length of about 8.

2. Separation of these peptides by ion-exchange chromatography yields about 20 fractions, most of which are small; two of the four main fractions account for 39.8% of the total molar yield of peptides.

3. One of these two fractions consists of two octapeptides, Gly-Ala-Gly-Ala-Gly-Ala-Gly-Tyr and Gly-(Gly<sub>3</sub>,Ala<sub>2</sub>,Val)-Tyr, and the other is the tetrapeptide Gly-Ala-Gly-Tyr.

4. The molar ratios in which these three peptides occur in the liquor is 4:4:3 respectively, and their relevance to the structure of the fibroin molecule is discussed.

The authors thank Miss D. E. Pinion and Mr C. Holliday for help with some of the experimental work.

#### REFERENCES

- Akabori, S., Ohno, K. & Narita, K. (1952). *Bull. chem. Soc. Japan*, **25**, 214.
- Blackburn, S. & Lee, G. R. (1954). *J. Text. Inst., Manchr*, **45**, T487.
- Cebra, J. J. (1961). *J. Immunol.* **86**, 197.
- Cook, H. G. & Saunders, B. C. (1947). *Biochem. J.* **41**, 558.
- Drucker, B., Hainsworth, R. & Smith, S. G. (1953). *J. Text. Inst., Manchr*, **44**, T420.
- Eastoe, J. E. (1955). *Biochem. J.* **61**, 601.
- Harris, J. I., Cole, R. D. & Pon, N. G. (1956). *Biochem. J.* **62**, 154.
- Levy, A. L. (1954). *Nature, Lond.*, **174**, 126.
- Locker, R. H. (1954). *Biochim. biophys. Acta*, **14**, 533.
- Lucas, F., Shaw, J. T. B. & Smith, S. G. (1955). *J. Text. Inst., Manchr*, **46**, T440.
- Lucas, F., Shaw, J. T. B. & Smith, S. G. (1957). *Biochem. J.* **66**, 468.
- Lucas, F., Shaw, J. T. B. & Smith, S. G. (1958). *Advanc. Protein Chem.* **13**, 107.
- Moore, S., Spackman, O. H. & Stein, W. H. (1958). *Analyt. Chem.* **30**, 1185.
- Moore, S. & Stein, W. H. (1954a). *J. biol. Chem.* **211**, 893.
- Moore, S. & Stein, W. H. (1954b). *J. biol. Chem.* **211**, 907.
- Smith, S. G. (1960). *J. Inst. Chem.* **84**, 390.
- Warwicker, J. O. (1954). *Acta Cryst.* **7**, 565.
- Ziegler, K. & La France, N. H. (1960). *Hoppe-Seyl. Z.* **322**, 21.
- Zuber, H. (1958). *Angew. Chem.* **70**, 656.