SUMMARY

1. Methyl-*O*- β -D-fructofuranoside has been prepared by the action of yeast invertase on sucrose in the presence of methanol.

2. The fructoside has been used as a substrate for a mould-enzyme preparation (Taka-diaxase). An examination of the products showed that both hydrolysis and transfructosylation occurred.

3. Sucrose could not be detected among the products of action upon mixtures of methyl β -fructoside and glucose, either by chromatographic fractionation or by the use of ^{14}C -labelled glucose.

4. Taka-diaxase acting on mixtures of raffinose and glucose formed sucrose. The incorporation of radioactive glucose into the molecule demonstrated that sucrose formation was the result of fructose transfer and not of melibiase (α -galactosidase) action.

5. The implications of these findings are discussed in relation to the hypothesis that invertases form fructosyl-enzyme compounds as common intermediates in hydrolysis and transfructosylation.

Part of this work was done with the technical assistance of Miss B. Dickinson. We are indebted to Dr J. Edelman, Research Institute of Plant Physiology, Imperial College, for letting us see the manuscript of his review (Edelman, 1956) before its publication. One of us (H.J.B.) wishes to thank the British Council for the award of a Scholarship. During part of the work technical assistance was provided

by the Agricultural Research Council, whose help we gratefully acknowledge.

REFERENCES

- Allen, P. J. & Bacon, J. S. D. (1956). *Biochem. J.* **63**, 200.
 Bacon, J. S. D. (1952). *Biochem. J.* **50**, xviii.
 Bacon, J. S. D. (1954). *Biochem. J.* **57**, 320.
 Bealing, F. J. (1953). *Biochem. J.* **55**, 93.
 Bealing, F. J. & Bacon, J. S. D. (1953). *Biochem. J.* **53**, 277.
 Bell, D. J. & Bacon, J. S. D. (1957). *J. chem Soc.* (in the Press.)
 Doudoroff, M. & O'Neal, R. (1945). *J. biol. Chem.* **159**, 585.
 Edelman, J. (1954). *Biochem. J.* **57**, 22.
 Edelman, J. (1956). *Advanc. Enzymol.* **17**, 189.
 Edelman, J. & Bealing, F. J. (1953). *Biochem. J.* **53**, ii.
 Feingold, D. S., Avigad, G. & Hestrin, S. (1956). *Biochem. J.* **64**, 351.
 Fischer, E. H., Kohtès, L. & Fellig, J. (1951). *Helv. chim. acta*, **34**, 1132.
 Hestrin, S., Feingold, D. S. & Avigad, G. (1955). *J. Amer. chem. Soc.* **77**, 6710.
 Hestrin, S., Feingold, D. S. & Avigad, G. (1956). *Biochem. J.* **64**, 340.
 Pazur, J. H. (1952). *Fed. Proc.* **11**, 267.
 Pazur, J. H. (1953). *J. Amer. chem. Soc.* **75**, 6323.
 Pazur, J. H. & French, D. (1952). *J. biol. Chem.* **196**, 265.
 Péaud-Lenoël, C. & Dedonder, R. (1955). *C.R. Acad. Sci., Paris*, **241**, 1418.
 Raybin, H. W. (1933). *J. Amer. chem. Soc.* **55**, 2603.
 Trevelyan, W. E., Procter, D. P. & Harrison, J. S. (1950). *Nature, Lond.*, **166**, 444.
 Wallenfels, K. & Bernt, E. (1952). *Angew. Chem.* **64**, 28.

The Amino Acid Sequence in a Fraction of the Fibroin of *Bombyx mori*

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

(Received 12 November 1956)

The silk fibroin of *Bombyx mori* was the first protein to which the methods of partial hydrolysis were applied in an attempt to elucidate its structure. In 1902, Fischer first isolated from fibroin a dipeptide thought to be glycylalanine, and since that date much work has been done to identify the products of the partial hydrolysis of fibroin, and thus to gain information about its molecular architecture. The considerable literature of this work up to 1943 has been summarized by Synge (1943), and from that date to the end of 1951 by Sanger (1952).

Since 1951, Levy & Slobodian (1952), Slobodian & Levy (1953), Kay & Schroeder (1954) and Joffe (1954) have extended our knowledge of the structure of fibroin by studying the products of its partial hydrolysis, but deductions from such studies of a

minimum repeating sequence throughout the fibroin molecule make assumptions about its regularity that have been shown to be incorrect. The work of Abderhalden & Bahn (1933), Meyer, Fuld & Klemm (1940) and of Zahn and co-workers (1952, 1954*a, b*, 1956) has clearly demonstrated that sections of the fibroin molecule differ in composition from other sections. The latter two authors interpreted these differences by suggesting that fibroin contained crystalline and amorphous phases of differing compositions.

The isolation and identification of the crystalline fraction of fibroin was achieved in these laboratories by the action of chymotrypsin on an aqueous solution of fibroin (Drucker, Hainsworth & Smith, 1953). This fraction, which appeared as a pre-

cipitate, was composed only of the residues of glycine, alanine and serine, with a little tyrosine. It had a molecular weight of approximately 4000, and glycine, alanine and serine were present in the ratio 3:2:1.

The present paper describes further analyses of this fraction of fibroin from a view to establishing the sequence of its amino acids. To do this, two main lines of attack were followed. In the first the fraction was investigated by the methods of partial hydrolysis with quantitative examination of the products so produced. In the second, use was made of a method of specific rearrangement and fission of the bonds involving serine nitrogen atoms similar to that described by Elliott (1952, 1955) and by Desnuelle & Bonjour (1951). The aim of this specific fission was the isolation of a repeating hexapeptide that should be present if the serine residues were evenly spaced throughout the molecule.

Throughout this paper the abbreviations of Brand & Edsall (1947) have been adopted for the amino acid residues, and the convention of Sanger (1952) has been used to indicate the composition, and the sequence of residues, of the peptides.

MATERIALS AND METHODS

Materials

Fibroin of Bombyx mori. Raw silk in the form of reeled 13/15 denier Japanese silk was treated in a boiling 1% solution of soap for 1 hr. to remove the sericin. It was then washed in many changes of hot water, until the final wash water was not alkaline to phenolphthalein. After drying, the silk was extracted in a Soxhlet apparatus first with light petroleum (b.p. 60–80°) and then with ethanol, and dried and washed well with warm water. It was then equilibrated with 0.02N-acetic acid overnight, centrifuged and again washed with warm water until the washings gave a neutral reaction. The silk was then dried in the air.

Aqueous solution of fibroin. This was prepared according to the method of Drucker *et al.* (1953).

Enzyme preparations. Crystalline chymotrypsin was obtained from Armour Laboratories; pure crystalline carboxypeptidase from the Worthington Biochemical Corp. of New Jersey.

Reagents and buffers. Dry hydrazine was prepared by distilling hydrazine hydrate four times over KOH in an atmosphere of nitrogen. 1-Fluoro-2:4-dinitrobenzene (FDNB) was prepared according to the method of Cook & Saunders (1947). Celite 535 was purchased from Johns-Manville and Co. Ltd. Chromatographic papers were

Whatman no. 1, no. 3 and no. 7. Glycylglycine was purchased from Roche Products Ltd., and glycyl-DL-alanine and DL-alanylglycine from L. Light and Co. Table 1 gives the constitution of the buffers used for chromatographic work.

Solvents. *tert.*-Amyl alcohol (technical quality from L. Light and Co.), B.S.S. ethyl acetate, A.R. diethyl ether and A.R. phenol were employed.

Preparation of the chymotryptic precipitate of fibroin (CTP). A volume of the aqueous solution of fibroin, containing approx. 10 g. of fibroin, was taken and to it was added between 50 and 150 mg. of crystalline chymotrypsin previously dissolved in a few millilitres of water. The mixed solutions were then buffered at pH 7.8 with phosphate buffer and the whole was made up to 300 ml. The solution then contained 0.433 g. of Na_2HPO_4 and 0.0442 g. of $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}/100$ ml. The whole was incubated at 40° for 4 hr. and the ppt. that formed was separated by centrifuging. It was then washed several times with water, with 0.1N-HCl and then with further water until free from acid. The ppt. was then washed with dry ethanol and with ether and dried over P_2O_5 in a vacuum desiccator.

Preparation of the dinitrophenyl derivative of the CTP. To a solution of 2 g. of NaHCO_3 in 20 ml. of water was added 40 ml. of ethanol in which 0.3 g. of FDNB had been dissolved. A portion (1 g.) of CTP was shaken with this mixture for 4 hr. at 40°, and then two further additions of FDNB, each of 0.3 g., were made, with a 4 hr. period of shaking between the additions and a final shaking of 4 hr., making 12 hr. in all.

The dinitrophenyl derivative (DNP-CTP) was washed successively with 1% NaHCO_3 , ethanol, ether, 1% NaHCO_3 , water and 0.02N-HCl. It was then rewashed with ethanol and with ether, dried, weighed and hydrolysed in 11N-HCl at 40° for various periods of time.

Methods

Chromatography. The absence from the CTP of amino acids with basic or acidic side-chains considerably increases the difficulty of separating the products of its hydrolysis by chromatography. Because the DNP-amino acids and peptides are coloured, it is simpler to separate them than the equivalent colourless untreated compounds. For this reason, and also because a method for the quantitative estimation of DNP-amino acids and peptides has been developed (Lucas, Shaw & Smith, 1955), the work has been carried out mainly with dinitrophenyl derivatives.

For paper chromatography of DNP derivatives the descending moving phase was *tert.*-amyl alcohol saturated with buffer of pH 6, and the stationary phase was Whatman no. 7 paper previously soaked in the same buffer and dried. If the separation of the DNP-peptides or DNP-amino acids aimed merely at identifying them, then the phthalate buffer of pH 6 was used, but if material was to be isolated and estimated quantitatively, then the phosphate buffer of pH 6 was used. This change to phosphate was necessary because, during the elution of DNP derivatives from the paper, the phthalate buffer was removed with them and on acidification gave phthalic acid. When the DNP derivative containing this phthalic acid was chromatographed either on paper or on a column, the pH of the system was locally altered and this gave irregular results. When the amino acids that formed the C-terminal residues were identified, phenol saturated with water was used as the descending

Table 1. *Composition of buffer solutions*

pH	Weights (g./l.)
7.3	71.6 $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ + 7.8 $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$
6.7	107.5 $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ + 27.2 KH_2PO_4
6.0	1.81 NaOH + 10.2 $o\text{-C}_6\text{H}_4(\text{CO}_2\text{H})(\text{CO}_2\text{K})$
6.0	6.2 $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ + 18.1 $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$
5.0	7.2 $\text{H}_3\text{C} \cdot \text{CO}_2\text{H}$ + 38.1 $\text{H}_3\text{C} \cdot \text{CO}_2\text{Na} \cdot 3\text{H}_2\text{O}$
4.0	9.6 $\text{H}_3\text{C} \cdot \text{CO}_2\text{H}$ + 5.44 $\text{H}_3\text{C} \cdot \text{CO}_2\text{Na} \cdot 3\text{H}_2\text{O}$

phase, the paper (Whatman no. 1) being suspended in a tank containing 0.2N-aq. NH_3 soln. and dilute KCN. Ammonium phthalate runs on such a paper chromatogram at the same rate as alanine, and since this ammonium salt also gives a colour with ninhydrin, it interferes seriously with the detection of alanine during the estimation of C-terminal residues.

Columns of Celite 535 were 9 mm. in diameter and buffered with two-thirds of their weight of the appropriate aqueous buffer. Those with ether as moving phase were 40 cm. long, and those with ethyl acetate were 70 cm. long. The column with ethyl acetate was buffered at pH 7.3 and was used for the initial separation of the DNP-peptides and DNP-amino acids. Further detailed separation was achieved by using ether buffered at pH 6.7 for fast-moving short DNP-peptides and DNP-amino acids, and at pH 4 for the slower long DNP-peptides and DNP-seryl peptides.

Estimation of dinitrophenyl-amino acids and dinitrophenyl-peptides. For the estimation of the DNP-amino acids, the u.v. light absorption of their solutions in 1% NaHCO_3 was determined at a wavelength of 360 μm , this wavelength being that at which maximum absorption occurs with these solutions. Standard solutions of the appropriate DNP-amino acids were measured in this way and their molar extinction coefficients determined. These were closely similar, and a mean value of 1.65×10^4 was used to determine the molar concentration of any unknown solution of a DNP-amino acid.

For the estimation of DNP-peptides, the DNP derivatives of a number of synthetic dipeptides, similar to those probably occurring in fibroin, were prepared and their u.v. light-absorption spectra determined. They were all found to have absorption peaks close to 354 μm , and measurements were made at this wavelength in subsequent determinations. A mean molar extinction coefficient of 1.60×10^4 was found and used.

None of the results given in this paper were corrected in any way for losses that occur during hydrolysis and chromatography.

Partial hydrolysis of the CTP

Hydrolysis and fractionation of the products. In each experiment 100 mg. of CTP, previously dried over P_2O_5 , was hydrolysed with 5 ml. of 11N-HCl at 40°. The period of hydrolysis varied from 5 hr. to 48 hr. The hydrolysate was evaporated to dryness and the HCl removed under reduced pressure at 40°. The residue was dissolved in water and made up to 50 ml.; 2 ml. was removed for estimation of total N by the Kjeldahl semi-micro method. A second portion of 2 ml. was taken and to it were added 3 ml. of water containing 100 mg. of NaHCO_3 and 10 ml. of ethanol containing 100 mg. of FDNB. The reaction mixture was allowed to stand at room temperature for 3 hr., after which period the ethanol was removed by vacuum distillation and the solution diluted to about 25 ml. The solution was extracted twice with ether to remove excess of FDNB and acidified with HCl and extracted five times in a separating funnel with 15 ml. portions of ethyl acetate. The ethyl acetate extracts were combined and evaporated to dryness, and redissolved in a small volume of ethyl acetate. The DNP material was fractionated on a Celite column buffered at pH 7.3, with ethyl acetate as the moving phase. On this column, the dinitrophenol obtained from decomposition of FDNB ran fast, together with bis-DNP-tyrosine and DNP-peptides containing tyrosine. These bands were followed, in

order of decreasing rate of movement, by the DNP derivatives of alanine plus Ala. Ala.; glycine; Ala. Gly plus Gly. Ala.; Ala. Gly. Ala.; serine; Ser. Gly plus Gly. Ala. Gly.; and Ser. Gly. Ala. Slower-moving bands, presumably of longer DNP-peptides, were observed, especially in the shorter hydrolyses, but these bands were not investigated.

The following mixtures were separated by column chromatography, the orders being given in decreasing rates of flow. The mixture of DNP-alanine and DNP-Ala. Ala was separated into its two components on a Celite column buffered at pH 6.7 with ether as moving phase, and DNP-Ser. Gly was separated from the DNP-Gly. Ala. Gly on a similar column buffered at pH 4.

DNP-glycine and the mixture of DNP-Ala. Gly plus DNP-Gly. Ala were further purified on a Celite-ether (pH 6.7) column; DNP-Ala. Gly. Ala was purified on a similar column buffered at pH 5, and DNP-serine on one buffered at pH 4.

Analysis of the fractions. The amino acid composition of all the bands was determined by completely hydrolysing each eluted band with 11N-HCl for 20 days at 40°. These conditions were chosen because it was found that they gave complete hydrolysis of the DNP-peptides without preferential decomposition of any DNP-glycine that might be present. The hydrolysates were diluted with water and extracted continuously with ether for about 2 hr. to remove the N-terminal DNP-amino acids, which were then separated on a Celite column buffered at pH 6.7 with ether as the moving phase. They were identified and estimated. The amino acids in the aqueous phase were converted into DNP derivatives and separated, identified and estimated by using the modifications of the method of Mills (1952) described by Lucas *et al.* (1955).

Identification of the C-terminal groups of the dinitrophenyl-peptides. The C-terminal groups of the DNP-peptides, especially of the tripeptides, were determined by means of carboxypeptidase. The DNP-peptide was dissolved in 0.1 ml. of 0.05M-ammonium acetate and the pH adjusted to 8–8.5 by addition of aq. NH_3 soln. A volume (10 μl .) of a suspension containing 100 mg. of carboxypeptidase/ml. of water was added and the mixture incubated at 40°. Portions were removed after various periods of time and applied directly to Whatman no. 1 paper, the applied spots being dried by a jet of hot air. The chromatograms were developed with phenol-water, in which the DNP derivatives ran faster than the freed amino acids, which were detected by spraying with ninhydrin solution.

Rearrangement and specific fission

Measurement of the production of amino groups after treatment of the CTP with anhydrous mineral acids. Two series of experiments were carried out, one at 27° with conc. H_2SO_4 and the other at 40° with H_3PO_4 to which had been added P_2O_5 to ensure that the acid remained anhydrous.

A quantity (600 mg.) of CTP was dried and dissolved in 10 g. of H_2SO_4 or H_3PO_4 , and the solutions were incubated for various periods of time. Portions of 0.4 ml. were removed and to each of these was added 5 ml. of ice-cold water, the vessel and contents being cooled in ice-water until required. Free amino N was estimated by the Van Slyke technique. The total N of the treated polypeptide was estimated by means of the Kjeldahl technique on a portion of the solution and the ratio of the total N to the amino N gave a measure of the mean chain length of the material.

It is shown in the Results section of this paper that the production of free amino N by both anhydrous acids at these temperatures proceeded at the same rate, to reach approximately the same value. The yield of peptide, however, was considerably greater when H_3PO_4 was used than when H_2SO_4 was used, and consequently H_3PO_4 was employed exclusively for the rearrangement.

Separation and estimation of the products of the specific fission of the CTP. A quantity (50 mg.) of CTP was dried over P_2O_5 *in vacuo*, transferred to a conical flask having a well-fitting stopper, and 2 g. of H_3PO_4 was added together with about 200 mg. of P_2O_5 . The flask and its contents were incubated at 40° for various periods of time. A mixture of crushed ice and ice-cold water was then added to the flask to bring the total volume to approximately 50 ml., and the mixture treated immediately with a small excess of CaCO_3 to precipitate the phosphate. The contents of the flask were allowed to come to room temperature, and then shaken continuously for 1 hr. with 100 ml. of ethanol containing about 250 mg. of FDNB. Sufficient of a saturated solution of NaHCO_3 was added (usually about 2 ml.) to bring the pH to 8–8.5, and shaking was continued for about 24 hr. with one further addition of 250 mg. of FDNB dissolved in a little ethanol.

The contents of the flask were then filtered through a Büchner funnel, the retained ppt. was washed with ethanol and ice-cold water, and the ethanol removed under reduced pressure. The aqueous solution was treated with about 50 mg. of Na_2CO_3 , transferred to a continuous-extraction apparatus (Mills, 1952), and extracted with ether for 2 hr. to remove excess of FDNB. The aqueous residue was transferred to a separating funnel, acidified with HCl and extracted five times with ethyl acetate to separate the short DNP-peptides and DNP-amino acids. This divided the material, which had been subjected to specific fission and then dinitrophenylated, into two fractions which will be referred to subsequently as the acetate-soluble fraction and the water-soluble fraction.

The water-soluble fraction was made up to 200 ml. and a portion diluted 10 times with water and its optical density determined at $354\text{ m}\mu$ on a u.v. light spectrophotometer. From this value the total molarity of the DNP-peptides in this fraction was calculated. The residue was evaporated to dryness, dissolved in a minimum of water and a sample applied as a strip to Whatman no. 3 paper buffered at pH 6 with phosphate buffer. The chromatogram was developed for 24–48 hr., according to the ambient temperature, with *tert.*-amyl alcohol buffered with the same phosphate buffer of pH 6. The paper was dried in the air, and the four clearly defined bands (I–IV) that had separated were cut from the paper and eluted with water. The optical densities of these fractions were measured at $354\text{ m}\mu$, and the molar yield of each was expressed as a proportion of the weight of the original material. The fastest-moving band (I), of short DNP-peptides, was present in small amount, and was not investigated in any detail. Its chromatographic behaviour appeared to be identical with part of the ethyl acetate-soluble fraction. The amino acid composition of the remaining bands (II, III, IV) was determined after complete hydrolysis by the methods already described.

Sequential analysis of the main product of specific fission. The DNP-peptide III obtained by the treatment of the CTP in anhydrous H_3PO_4 was subjected to partial hydrolysis in 11N-HCl at 40° for various periods of time. The hydrolysates

were extracted with ethyl acetate, and the ethyl acetate-soluble material was investigated in the manner already described for the partial hydrolysates of the CTP. A small amount of dinitroaniline, presumably arising from decomposition of DNP-serine and DNP-seryl peptides, was present in this material, and ran as a fast band with dinitrophenol on the ethyl acetate (pH 7.3) column. The water-soluble fraction was divided into two parts. One half was evaporated to dryness to remove HCl, and the residue was dissolved in water and passed down an acid Amberlite IR-120H resin column. The longer DNP-peptides emerged and the free peptides and amino acids were retained strongly by the column. The longer DNP-peptides were separated into two fractions on Whatman no. 3 paper buffered at pH 6 with *tert.*-amyl alcohol as moving phase. The bands were eluted with water and estimated.

The second half of the water-soluble fraction containing the free peptides and amino acids was treated with FDNB, and the resulting DNP-peptides and DNP-amino acids were extracted with ethyl acetate and separated on Celite columns as already described. After estimation of the separated constituents, each eluted band was hydrolysed completely with HCl and the resulting amino acids were converted into their DNP derivatives and identified and estimated after separation on Celite columns buffered at pH 6.7 with ether as the moving phase.

RESULTS

N-Terminal sequence of the CTP molecule

N-Terminal amino acids. The *N*-terminal amino acid residues of the CTP were determined by treating it with FDNB, hydrolysing with 11N-HCl for 20 days at 40° and separating the DNP-amino acids, which were then estimated by the u.v. light-absorption technique. The results showed that glycine occupies 84% of the terminal amino positions, serine 10% and alanine 4%. Glycine thus occupies the amino end position in the chain far more frequently than it would do if the amino acid sequence were governed only by random arrangement of the amino acids.

N-Terminal peptides. The dinitrophenylated CTP was partially hydrolysed with 11N-HCl at 40° for various periods of time and the products bearing DNP groups were extracted with ethyl acetate. These were separated on a column and their compositions found after complete hydrolysis. The results obtained are given in Table 2.

Table 2. *Products of partial hydrolysis of DNP-CTP*

Hydrolysis was with 11N-HCl at 40° .

Percentage of total *N*-terminal residues found as

Time of hydrolysis (hr.)	DNP-glycine	DNP-Gly. Ala	Longer DNP-peptides
12	23.3	29.7	27.5
24	32.2	26.7	—
48	55.7	30.0	—
5×4	16.4	25.0	39.3

Liberation of DNP-glycine during partial hydrolysis was rapid, in keeping with the known lability of the bond between DNP-glycine and other amino acids (Sanger, 1949). The only DNP-dipeptide produced during the partial hydrolysis was DNP-Gly.Al_a. It is desirable to increase the yield of longer DNP-peptides and to try to prevent formation of free DNP-glycine, and to this end a series of five short hydrolytic treatments each of 4 hr., and each followed by removal of the DNP-peptides formed, was tried. This reduced the production of DNP-glycine, and did not significantly alter the production of DNP-Gly.Al_a, but increased the yield of higher DNP-peptides. Nearly 40% of the residues remained as longer DNP-peptides, and these were separated into two main bands on a Celite-ethyl acetate (pH 6.7) column. One band was a DNP-tripeptide containing two glycine and one alanine residues [DNP-Gly.(Al_a, Gly)] and the other a DNP-tetrapeptide containing two glycine and two alanine residues [DNP-Gly.(Al_a₂, Gly)], glycine occupying the *N*-terminal position in each peptide. The yields of these DNP-peptides after the five short hydrolyses are shown in Table 3.

As is shown below, the complete absence of Gly.Gly from all hydrolysates of the CTP, and the very small amount of Al_a.Al_a, mean that in the tri- and tetra-peptides shown in Table 3 the sequences are alternations of glycine and alanine. Hence it is

shown, by actual isolation of the products, that 59.3% of the total *N*-terminal glycine residues are followed by alanine, and if it is assumed that the free DNP-glycine is wholly derived from DNP-Gly.Al_a, the experimental value becomes 78.5%. It seems reasonable to conclude therefore that practically all the positions next to the *N*-terminal glycine are occupied by alanine.

Partial hydrolysis

Identification and estimation of the products. The products of the partial hydrolysis of the CTP were examined by the methods already given in detail in the Experimental section. These methods proved satisfactory except for DNP-Gly.Al_a and DNP-Al_a.Gly, which could not be adequately separated by any of the systems used. However, by using synthetic mixtures it was shown that DNP-Gly.Al_a and DNP-Al_a.Gly, whilst running together at the same rate, separated from DNP-alanine, DNP-glycine, DNP-Gly.Gly and DNP-Al_a.Al_a, so that the discrete band obtained on the column was likely to be a mixture of DNP-Gly.Al_a and DNP-Al_a.Gly uncontaminated by the other closely related bodies.

On complete hydrolysis of the mixture of the two dipeptides, only DNP-alanine, DNP-glycine, alanine and glycine were obtained, the amount of DNP-alanine being approximately equal to the free glycine, and the amount of DNP-glycine approximately equal to the free alanine. It is clear from these facts that the mixture can only be one of DNP-Al_a.Gly and DNP-Gly.Al_a, and the proportion and amount of each present in the mixture can be derived from the analysis of the complete hydrolysate.

The results obtained from the analyses of the partial hydrolysates of the CTP are given in Table 4. They refer only to the fraction of the hydrolysates that is extractable with ethyl acetate.

Table 3. *Products of five successive hydrolyses of DNP-CTP*

Hydrolysis was with 11*N*-HCl at 40° for 4 hr.

Peptide	Yield (moles/10 ⁶ g.)	Percentage of total Gly terminal groups
DNP-glycine	2.97	19.2
DNP-Gly.Al _a	4.48	28.9
DNP-Gly.(Al _a , Gly)	3.33	21.5
DNP-Gly.(Al _a ₂ , Gly)	1.37	8.9

Table 4. *Products of partial hydrolysis of CTP after various times*

Hydrolysis was with 11*N*-HCl at 40°. N_a/N_p % = g. of amino acid or peptide N/100 g. of CTP N. n.d. = Not determined.

	Time of hydrolysis (hr.)									
	5		10		24		36		48	
	Moles 10 ⁵ g.	N_a/N_p %	Moles 10 ⁵ g.	N_a/N_p %	Moles 10 ⁵ g.	N_a/N_p %	Moles 10 ⁵ g.	N_a/N_p %	Moles 10 ⁵ g.	N_a/N_p %
Glycine	20	1.46	43	3.1	141	10.2	200	14.3	248	17.8
Alanine	14.7	1.06	26	1.89	102	7.3	136	9.8	191	13.7
Serine	9.0	0.64	17.0	1.22	30	2.2	39	2.8	47	3.4
Al _a .Gly	54	7.8	106	15.2	148	21.2	160	22.9	131	18.7
Gly.Al _a	21	3.0	50	7.2	74	10.6	79	11.3	68	9.8
Ser.Gly	46	6.7	73	10.4	95	13.6	114	16.4	101	14.5
Al _a .Al _a	n.d.	—	7.1	1.02	10.7	1.54	10.0	1.44	8.7	1.26
Gly.Al _a .Gly	31	6.7	42	9.1	30	6.4	19.2	4.1	8.5	1.83
Ser.Gly.Al _a	22	4.7	34	7.3	23	5.0	20	4.3	9.2	1.98
Al _a .Gly.Al _a	n.d.	—	13.2	2.9	10.7	2.3	5.8	1.25	3.5	0.75
Total	—	32.1	—	59.3	—	80.3	—	88.6	—	83.7

Identification of the C-terminal amino acid of the DNP-seryl and DNP-alanyl tripeptides by means of carboxypeptidase established their sequence as Ser.Gly.Ala and Ala.Gly.Ala. Control experiments showed that no amino acids were produced by the enzyme alone under the experimental conditions, and DNP-Ser.Gly, DNP-Gly.Ala and DNP-Ala.Gly derived from the synthetic dipeptides gave no free amino acids with carboxypeptidase up to 7 hr. of action by the enzyme. With the glycyl tripeptide the carboxypeptidase released mainly glycine, but in addition a small amount of alanine. The possibility that this tripeptide might contain some DNP-Gly.Gly.Ala was investigated by subjecting it to prolonged action by carboxypeptidase for 24 hr., followed by extraction of the DNP derivatives with ethyl acetate. Examination of this extract by paper chromatography indicated the complete absence of DNP-Gly.Gly, so that the freed alanine could not be derived from DNP-Gly.Gly.Ala. Its origin is therefore uncertain, and it is possible, for the glycine tripeptide, to say only that in the main it consists of Gly.Ala.Gly. There is strong indirect evidence too that the tripeptides are as stated because only those dipeptides that could arise from these sequences are found in their hydrolysates.

The results in Table 4 may be summarized as follows:

(i) The only dipeptides isolated were Ala.Gly, Gly.Ala and Ser.Gly, with a little Ala.Al. Altogether, on a nitrogen basis, these dipeptides accounted for 52% of the CTP in the 36 hr. hydrolysate, which gave maximum recovery of dipeptides.

(ii) The dipeptide Gly.Gly was not isolated even from the shortest hydrolysate, and Ala.Al. only in very small amount.

(iii) The ratio of Ala.Gly to Gly.Al. in the hydrolysates of longer periods than 5 hr. was consistently close to 2:1.

(iv) The only tripeptides obtained were Gly.Al. Gly, Ser.Gly.Al. and Ala.Gly.Al.

Evidence for the absence of glycylglycine from partial hydrolysates. No evidence of the presence of the linkage Gly.Gly was obtained from any of the hydrolysates, but as its absence is of great importance in elucidating the structure of the CTP, a special series of experiments was conducted to establish this fact as unequivocally as possible.

A synthetic mixture was prepared of the DNP derivatives of serine and Gly.Gly, and it was shown that, on a Celite-ethyl acetate (pH 7.3) column they ran together as a single band. However, it was found that they could be separated on a Celite-ether (pH 4) column, the DNP-serine running the faster. The volume of ether necessary completely to elute each component was determined.

A portion of a short, 5 hr. hydrolysis of the CTP at 40° was then examined by passing it first down a Celite-ethyl acetate (pH 7.3) column, and eluting the band known to be DNP-serine. This band would also contain any DNP-Gly.Gly, if it were present. This eluate was passed down a Celite-ether (pH 4) column until the DNP-serine was removed, and then a further volume of ether, equal to that previously determined as necessary to remove DNP-Gly.Gly, was passed through the column, the eluate collected and shaken well with a small volume of sodium bicarbonate solution. No colour could be detected in the bicarbonate solution, demonstrating the absence of DNP-Gly.Gly.

The possibility remains, however, that the Gly.Gly linkage is so labile as to be completely destroyed even during so short a hydrolysis as 5 hr. at 40°. The rates of hydrolysis of the dipeptides Gly.Gly, Ala.Gly and Gly.Al. were measured therefore under the same conditions as were used for hydrolysing the CTP (11N-HCl at 40°). The results obtained are given in Table 5.

Thus although Gly.Gly hydrolysed at a rate that is faster than that of either Gly.Al. or Ala.Gly, this higher rate could scarcely account for the non-appearance of Gly.Gly in all the hydrolysates of the CTP were it present in significant amounts.

Rearrangement and specific fission

Rates of production of free amino groups. The values obtained for the ratio of the total nitrogen of the CTP to its amino nitrogen after treatment with anhydrous sulphuric and phosphoric acids for various periods of time are given in Table 6. In their general trend the results are closely similar for the two acids. There is a rapid reduction in the ratio of total nitrogen to amino nitrogen during the first 24 hr., followed by a more gradual diminution.

Because the ratio of glycine to alanine to serine is 3:2:1 in the CTP, the ratio of total nitrogen to amino nitrogen of the CTP during rearrangement in anhydrous acid should reach a steady value of 6. That it does not do so is probably an indication of unwanted fission during the specific rearrangement, and of hydrolytic effects during dilution of the acid before determination of the amino nitrogen. The values derived for the ratio of the total nitrogen to amino nitrogen were used therefore merely as an indication of the order of the time likely to be

Table 5. *Rates of hydrolysis of dipeptides in 11N-HCl at 40°*

	$10^2 \times \text{Velocity constant (hr.}^{-1}\text{)}$	Half-life (hr.)
Gly.Gly	2.35	29.5
Ala.Gly	1.37	50.5
Gly.Al.	1.37	50.5

Table 6. *Liberation of amino groups from CTP*

Anhydrous H ₂ SO ₄ at 27°				Anhydrous H ₃ PO ₄ at 40°			
Time (hr.)	Amino N (mg.)	Total N (mg.)	Total N Amino N	Time (hr.)	Amino N (mg.)	Total N (mg.)	Total N Amino N
5	0.114	1.29	11.1	4	0.079	1.17	15
16	0.205	1.48	7.3	16	0.16	1.10	6.7
24	0.194	1.23	6.3	28	0.28	1.71	6.0
48	0.245	1.23	5.0	48	0.36	1.71	4.7
96	0.268	1.23	4.5	72	0.40	1.71	4.3
				170	0.50	1.71	3.4

Table 7. *Products of specific fission of CTP*

CTP was treated with H₃PO₄ at 40° for various times, then with FDNB, and the water-soluble products were fractionated by paper chromatography. Yields of the products in the four bands obtained are expressed as moles/10⁵ g. of CTP.

Band	Time of treatment (days)				
	2	3	4	5	6
I	5	8	6	10	10
II	15	14	19	24	21
III	46	49	81	77	65
IV	28	31	13	11	7
Total	94	102	119	122	103

necessary for the maximum rearrangement at the serine residues, and a series of experiments was carried out to determine precisely the best conditions for this.

Examination of the water-soluble fraction

Fractionation. The products of the specific fission of the CTP after it had been treated in anhydrous phosphoric acid at 40° for various periods of time were separated, and the yields of the four bands I–IV of the water-soluble fraction were determined. The results obtained are given in Table 7. It is seen that a maximum yield, especially of the band III, is obtained after treatment for between 4 and 5 days.

The total amount of the four bands I–IV recovered after treatment of the CTP with anhydrous sulphuric acid was less than that recovered after treatment with anhydrous phosphoric acid. Thus whereas the best yield from sulphuric acid treatment was 53 moles/10⁵ g. of CTP, it was 122 moles/10⁵ g. with phosphoric acid.

Examination of the main product (band III). The amino acid composition and sequence of band III, which was the principal product of the specific fission, was investigated in detail.

After complete hydrolysis, only glycine, alanine and DNP-serine were found, in the ratios 3:2:0.85. A partial breakdown of DNP-serine occurs on hydrolysis, giving dinitroaniline and, to a less extent, free serine. If the amounts of dinitroaniline and the small amount of free serine recovered are added to the value 0.85, the serine figure is very close to 1. Band III therefore consists of a DNP-hexapeptide of composition DNP-Ser.(Gly₃, Ala₂).

From Table 7 it can be seen that 81 moles of band III/10⁵ g. of CTP were obtained in the aqueous phase. On the basis of the composition of this band, 33.8 % of the original weight of CTP were therefore recovered as hexapeptide after specific fission.

The products of partial hydrolysis of the hexapeptide, band III, were separated and analysed by the methods already described. The longer DNP-peptides separated into two fractions, one of which was identical with unchanged DNP-hexapeptide. The second, faster fraction had a composition corresponding to a mixture of DNP-ser.(Gly₂, Ala) and DNP-Ser.(Gly₂, Ala₂). Attempts to resolve this band by chromatography and electrophoresis failed, and the proportions of each component were calculated from the analytical data. The results obtained after 17 hr. of hydrolysis are given in Table 8, this time being chosen because it gave a high proportion of di- and tri-peptides. Hydrolysis for longer and shorter periods gave no evidence of the presence of short peptides other than those listed, though a careful examination of the mixture was made for the presence of Ala.Ala and Gly.Gly.

Of the total nitrogen of the hexapeptide, 74 % was recovered as the products of its partial hydrolysis.

Examination of Table 8 enables the following conclusions to be drawn: (1) Serine occurs only as the *N*-terminal amino acid. (2) The DNP-serine group is followed by glycine, since only DNP-Ser.Gly and not DNP-Ser.Ala was recovered. (3) The ratio of Ala.Gly to Gly.Ala is of the order of 2:1, as found previously with partial hydrolysates of the CTP. (4) No sequence was found that did not

agree with band III being a DNP-hexapeptide having a structure DNP-Ser.Gly.Ala.Gly.Ala.Gly.

Examination of the DNP-hexapeptide by means of the hydrazine method (Shaw & Smith, 1954) showed that the *C*-terminal group was glycine. A little alanine also appeared on the chromatograms. This was probably released by random hydrolysis caused by the water produced by reduction of the nitro groups in the DNP groups by hydrazine.

When the carboxypeptidase method was used, both glycine and alanine were liberated in approximately equal amounts. This result can probably be explained by the very high rate of liberation of alanine, compared with glycine, from the *C*-terminal position when similar peptides containing these amino acids are treated with carboxypeptidase (Neurath & Schwert, 1950).

Examination of bands II and IV. The bands II and IV were not examined in such detail as III.

Band IV was separated into several fractions by chromatography, and, in one instance, by electrophoresis. None of the fractions was homogeneous, but two (IV₁ and IV₂) were obtained in greater yield than the others. Analyses of bands II, IV₁ and IV₂ are given in Table 9.

Band II is shown to be a mixture of tetra- and penta-peptides, and attempts to separate the

mixture chromatographically failed. From the known chromatographic behaviour of *N*-terminal glyceryl peptides, it is probable that the *N*-terminal glyceryl peptide is Gly.(Gly₂, Ala₂). The remaining peptides would then be *N*-terminal seryl peptides of composition *N*-terminal serine 18.3, non-*N*-terminal alanine 27.6 and non-*N*-terminal glycine 40.6, all values being given as molar proportions. This analysis would correspond to a mixture of Ser.(Gly₂, Ala) and Ser.(Gly₂, Ala₂), with a mean chain length of 86.5/18.3 = 4.7.

Examination of the ethyl acetate-soluble fraction

The CTP was treated in anhydrous phosphoric acid for 3, 4 and 5 days, and the products were allowed to react with FDNB. That fraction which was soluble in ethyl acetate was separated and examined. It consisted of short DNP-peptides and DNP-amino acids, together with some longer DNP-peptides and much dinitrophenol. The values obtained for the principal constituents are shown in Table 10.

It was to be expected that the partition of the DNP-amino acids and peptides between water and ethyl acetate would not be sharp, and this is confirmed by the presence of some DNP-peptides longer than three residues in the ethyl acetate fraction. From the longer material recovered from this fraction, a band was separated that had all the

Table 8. *Products of partial hydrolysis of DNP-Ser.(Gly₃, Ala₂)*

Hydrolysis was with 11N-HCl at 40° for 17 hr. DNP N is not included in the total N.

	Serine N (% of total serine N)	Glycine N (% of total glycine N)	Alanine N (% of total alanine N)	N of fraction (% of total N)
DNP-serine	19.6	—	—	3.4
Alanine	—	—	14.4	4.6
Glycine	—	16.1	—	7.9
DNP-Ser. Gly	24	8.3	—	8.2
Gly. Ala	—	7.1	10.9	7.0
Ala. Gly	—	13.3	20.0	13.0
DNP-Ser. (Gly, Ala)	8.7	4.7	3.2	4.5
DNP-Ser. (Gly ₂ , Ala)	2.5	1.7	1.3	1.7
DNP-Ser. (Gly ₂ , Ala ₂)	2.8	2.0	3.0	2.4
DNP-Ser. (Gly ₃ , Ala ₂)	8.4	8.8	9.1	8.7
Gly. (Ala, Gly)	—	10.6	8.1	7.8
Ala. (Gly, Ala)	—	1.3	4.0	1.9
Dinitroaniline	—	—	—	3.0

Table 9. *Amino acid composition of three products of specific fission of CTP*

For the isolation of bands II and IV see Table 7. Band IV was further fractionated by paper chromatography.

	Molar proportions of						Chain length (total N/ <i>N</i> - terminal N)
	<i>N</i> -terminal			Non <i>N</i> -terminal			
	Serine	Glycine	Alanine	Serine	Glycine	Alanine	
Band II	18.3	6.0	0	0	52.6	39.6	4.8
Band IV ₁	7.7	6.2	1.5	26	107	78	15
Band IV ₂	3.1	2.6	2.2	6.6	33.9	26.4	9.5

Table 10. *Products of specific fission of CTP*

CTP was treated with H_3PO_4 at 40° for various times, and then with FDNB. The ethyl acetate-soluble products were fractionated by column chromatography. n.d. = Not determined; +, trace (insufficient for quantitative estimation).

DNP-amino acids and peptides	Moles/ 10^5 g. of CTP after treatment for		
	3 days	4 days	5 days
DNP-alanine	+	+	+
DNP-glycine	5.6	6.4	8.5
DNP-serine	+	+	+
DNP-Ala. Gly plus DNP-Gly. Ala	9.6	13.5	14.0
DNP-Ser. Gly	8.9	11.3	11.0
DNP-Ser. Gly. Ala	7.5	8.3	11.0
DNP-Gly. Ala. Gly	6.1	11.0	12.0
DNP-Ala. Ala	+	+	n.d.
Long DNP-peptides excluding band III	n.d.	33.0	30.0
DNP-hexapeptide, band III	n.d.	5.9	6.9

characteristics of band III of the water-soluble fraction, and this material was almost certainly the DNP-hexapeptide. If the amount recovered from the CTP treated for 4 days in anhydrous phosphoric acid (5.9 moles/ 10^5 g.) is added to the 81 moles/ 10^5 g. isolated in the water-soluble fraction (Table 7), a total of 86.9 moles/ 10^5 g. is obtained. This corresponds to a recovery of 36.2% by weight of the CTP as the hexapeptide Ser. Gly. Ala. Gly. Ala. Gly.

By comparison of Tables 4 and 10, it is seen that similar amino acids and peptide sequences are recovered by both methods, though the yields after treatment of the CTP with anhydrous acid are naturally different.

DISCUSSION

An earlier paper (Drucker *et al.* 1953) described the separation of fractions of silk fibroin by means of pancreatin, and also by means of crystalline chymotrypsin. The fractions appeared as precipitates when a solution of fibroin was treated with the enzymes, and analysis of that produced by the action of pancreatin showed that it was a polypeptide composed almost entirely of residues of glycine, alanine and serine, together with a little tyrosine, the molar ratios of glycine to alanine to serine being 3:2:1. The fraction had a high degree of crystalline order as shown by X-rays, and a molecular weight of about 4000.

We have called the fraction produced by chymotrypsin through specific fission of the bonds involving the carboxyl groups of tyrosine, the chymotryptic precipitate (CTP). It contained 62% of the total nitrogen, and represented 58% of the total weight of fibroin. A detailed analysis of the CTP

(Lucas & Smith, unpublished results) has established its formula as $\text{Gly}_{29}\text{Ala}_{20}\text{Ser}_9\text{Tyr}_1$, and the C-terminal position has been shown to be occupied mainly by tyrosine (Shaw & Smith, 1954). Apart therefore from the presence of a little more tyrosine, the CTP appears to be closely similar in composition to that produced by pancreatin.

Holmes & Smith (1952) have shown by means of the ultracentrifuge that the average molecular weight of regenerated silk fibroin in aqueous solution is approximately 84 000, a value supported by determinations of the lysine, histidine and arginine in fibroin by Corfield, Howitt & Robson (1954). This can be regarded as a minimum value, and the true molecular weight of fibroin may well be considerably higher. Since the average weight of the amino acid residues of fibroin is 78, there are approximately 1100 residues/molecule and, of these, slightly more than 60% will be in the ordered, crystalline portions and the remainder in the amorphous portions. Thus the fibroin molecule can be regarded as consisting of about twelve sections of average molecular weight approximately 4000, these sections being those that, attached to similar sections of neighbouring molecules, constitute the crystalline part of silk. Alternating with these are another twelve sections, of average molecular weight about 3000, which form the amorphous part of silk. This latter value is similar to that suggested by Zahn & Zuber (1954*b*) for the molecular weight of the tyrosine-containing sections of the fibroin molecule.

Consideration of the results given in the present paper enable a detailed concept of the structure of the CTP to be developed. The occurrence of glycine as the *N*-terminal amino acid residue in over 80% of the molecular chains shows that in most cases those sections of the molecules that form the crystalline, orientated part of fibroin are introduced by Tyr. Gly, and this is followed by alanine in at least 59% of the molecules, and probably in a greater proportion than this. The high lability of the Gly. Ser bond and probably also of the Ala. Ser bond means that the possibility of an occasional sequence in which serine occurs near to the *N*-terminal glycine bond cannot be excluded, but it seems likely that the majority of the *N*-terminal sequences that introduce the molecules of the CTP are Gly. Ala. Gly or Gly. Ala. Gly. Ala.

The results of the analyses of the products of partial hydrolysis of the CTP, given in Table 4, show that serine is never recovered in peptide form except as the *N*-terminal residue. This is to be expected because of the very high lability, in concentrated mineral acids, of the *N*-peptide bond of serine. Ser. Ala is not recovered but only Ser. Gly, so that in practically every instance, serine must be followed in the molecular chain by glycine. The other dipeptides recovered are Ala. Gly and Gly. Ala,

which are present with remarkable constancy in the ratio 2:1.

Attention must now be directed to the absence of the Gly.Gly linkage. Its non-appearance, even after the shortest periods of hydrolysis, has an important bearing on the elucidation of the amino acid sequence, because this fact, combined with the experimental finding that one-half of the total residues are glycyl, means that every alternate residue in the CTP must be glycine for practically the whole of the molecule. Thus in most of the CTP, glycine must alternate with alanine or with serine.

The regular alternation of glycine with alanine or serine considerably simplifies the elucidation of the amino acid sequence, since otherwise the small variety and low reactivity of the residues that are present, and the absence of any that could be used as 'markers', would make this elucidation almost impossible.

That Gly.Gly can be recovered from partial hydrolysates of proteins that contain it, is shown by the work of Kroner, Tabroff & McGarr (1953), who isolated it from collagen after a period of hydrolysis of 4 days in concentrated hydrochloric acid at 37°, which is a period of hydrolysis considerably longer than any used in our experiments. Schroeder, Kay, Le Gette, Honnen & Green (1954) also tentatively identified Gly.Gly in partial hydrolysates of gelatin after treatment in 3.6*N*-hydrochloric acid at 37° for 7 days, and more recently, by using conditions of hydrolysis identical with those employed during our longest treatment (48 hr.), Schroeder, Kay, Munger & Burt (1956) have isolated large amounts of Gly.Gly from hydrolysates of tussah fibroin.

Because of the high lability of the Gly.Ser bond, it is likely that hydrolysis of the CTP will give rise at an early stage to peptides that have serine as their *N*-terminal residue, and have a general formula Ser.Gly.(Ala.Gly)_{*n*}. If the serine residues are evenly spaced throughout the molecule, *n* will have a value of 2: its value must in any case have an average of 2 because the ratios of glycine:alanine:serine are 3:2:1.

Once the Gly.Ser linkage has been broken, the adjacent Ser.Gly linkage becomes relatively stable, owing to the effect of the charged NH₃⁺ group which tends to repel the approaching H⁺ ions (Sanger, 1952). This accounts for the relatively low yields of serine obtained during the early stages of hydrolysis.

The extent to which the CTP consists of regular repeats of the hexapeptide Ser.Gly.(Ala.Gly)₂ was investigated by specific fission of the chains at the peptide bonds involving the serine amino groups. When the CTP is dissolved in anhydrous phosphoric acid, rearrangement of these bonds occurs, with the production of free amino groups and *O*-acyl (ester) bonds that give an ester-linked chain. When this solution is diluted with ice-water and treated with

FDNB at a pH that increases by stages to about 9, blocking of the free amino groups occurs by their reaction with FDNB, and fission of the corresponding ester bonds takes place. It is not known exactly at what stage this fission occurs, but any free amino groups that have not already reacted with FDNB will tend to revert to their peptide form by a reversal of the *O*-acyl to the *N*-acyl form of linkage. This effect, together with the small amount of random fission that probably occurs in the early stages of the treatment in acid solution, reduces the efficiency of the method and gives yields of the DNP-hexapeptide that are less than they otherwise would be.

The actual isolation during the present work of 36% by weight of the CTP as a hexapeptide is therefore of considerable significance, especially since serine is the only *N*-terminal amino acid found in this hexapeptide. Elliott (1953) has stated from his study of silk fibroin that only 20% of the transformed seryl residues are recoverable as DNP-serine. Hence the amount recovered by us, which represents just over 40% of the total serine, may indicate that the CTP is composed mainly of repeat units of the hexapeptide.

The partial hydrolysis of the recovered hexapeptide gave products that were essentially similar to those given by partial hydrolysis of the CTP. Of the other products of specific fission, bands IV₁ and IV₂ both had mean chain lengths greater than six, but in both instances non-*N*-terminal serine occurred, though this was significantly absent from band II, where the mean chain length was 4.8. This non-*N*-terminal serine may be present either because complete rearrangement had not occurred at all the bonds involving serine nitrogen atoms, or because, if it had occurred, the bond was regenerated in alkaline solution before the amino group had reacted with the FDNB.

It can be concluded therefore that the CTP molecule is composed mainly of repeating units of the form Ser.Gly.(Ala.Gly)_{*n*}. All the peptides found in the partial hydrolysates of the CTP could be derived from this sequence with the exception of Ala.Ala. This was isolated only in a small yield, the amount found being very roughly of the order that would be expected if there were one Ala.Ala bond/molecule of CTP. There is no direct evidence to indicate the position of this bond in the CTP molecule, but it seems more likely that it would occur towards the end of the molecule than in the middle, since in the latter position it would interrupt the regular alternating sequence of glycine residues and probably interfere with the crystallization of these regions. We have therefore shown the Ala.Ala bond in the formula below in the *C*-terminal sequence of the CTP, but this must be taken as a tentative suggestion only.

The CTP is thus composed of an aggregation of those parts of the fibroin molecules that make up the ordered sections of the protein; these parts are probably incorporated in the molecule in the following manner:

where n is usually 2 and always has a mean value of 2.

The results presented in this paper can be compared with those of previous workers in this field. Levy & Slobodian (1952) and Slobodian & Levy (1953) established that the dipeptides Ala. Gly and Gly. Ala occur commonly in silk fibroin, and that the tripeptide Gly. Ala. Gly is an important structural element. On the basis of this work, they proposed a minimum repeating peptide sequence of the type X_1 . Ala. Gly. Ala. Gly. X_2 . Gly. Kay & Schroeder (1954), using the chromatographic technique of Moore & Stein (1951), examined a range of peptides from the partial hydrolysis of silk fibroin, and whilst confirming the high proportions of Ala. Gly and Gly. Ala, which were present in the ratio of approximately 2:1, they found other sequences (for example Ala. Ala and probably Ser. Gly. Ala) that would not fit into the sequence suggested by Levy & Slobodian.

Joffe (1954), as the result of an investigation of tyrosine-containing peptides obtained from fibroin, suggested that the framework of the fibroin structure lies in an alternation of the tetrapeptide Ala. Gly. Ala. Gly with dipeptides such as Ser. Gly, Tyr. Gly and Val. Gly.

The results obtained by these workers have been based on studies of whole fibroin, and deductions about the structure of fibroin from such results have assumed that its molecule possesses a fair degree of regularity in the amino acid sequences throughout its length. As has been shown in this and previous papers, fibroin can be divided into two main fractions that are very different in their amino acid compositions. The fact that workers using unfractionated fibroin obtain results indicating repeat sequences that are not grossly different from our own is merely a reflexion of the fact that the CTP represents 60% of the whole fibroin, and may indicate that, in the remainder, glycine, alanine and serine residues, which account for a considerable part of it, are in the same kind of sequence as they are in the CTP.

Waldschmidt-Leitz & Zeiss (1954, 1955) have worked with fractionated fibroin. They examined the fraction first isolated from fibroin by Drucker *et al.* (1953) by the action of chymotrypsin and suggested for it the formula $[\text{Gly. Ala. Ser. Gly. Ala. Gly}]_n \text{. Tyr.}$ Such a formula is at variance with our results, and also with the authors' own results, because the repeat sequence would give six Gly. Gly

linkages/molecule, yet Gly. Gly was not recovered either by these authors or by us. In addition, the separation of significant amounts of Ala. Gly. Ala during the present work and also by Waldschmidt-Leitz & Zeiss, gives a sequence that could not be obtained from the formula they propose.

Reference has already been made to the remarkable consistency with which the molar ratio of Ala. Gly to Gly. Ala approaches 2:1 in hydrolysates of the DNP-hexapeptide (DNP-Ser. Gly. Ala. Gly. Ala. Gly), of the CTP, and also of whole fibroin. In the hexapeptide it is evident that the rates of fission of the bonds joining the residues must differ considerably from each other, since otherwise hydrolysis would yield approximately equimolar proportions of Ala. Gly and Gly. Ala.

In the Appendix an attempt has been made to show that, by making a few reasonable assumptions about rates of fission of the bonds, it is possible to obtain calculated values for the ratio of Ala. Gly to Gly. Ala that are close to the ratio 2:1 found experimentally. If this calculation and the assumptions on which it is based have any significance, it is difficult to see how a ratio of Ala. Gly to Gly. Ala of 2:1 could be derived from a compound of the structure proposed by Waldschmidt-Leitz & Zeiss.

SUMMARY

1. When an aqueous solution of silk fibroin is acted on by chymotrypsin, a precipitate is formed. The sequence of the amino acid residues in this precipitate has been investigated by partial hydrolysis, followed by separation of the products by chromatography and quantitative estimation of their amounts.

2. By use of the specific action of anhydrous phosphoric acid on the seryl linkages, 36% by weight of the precipitate was recovered as a hexapeptide, which was identified as Ser. Gly. Ala. Gly. Ala. Gly.

3. It is suggested, on the basis of the experimental evidence, that the precipitate probably has the formula

in which n is usually 2 and always has a mean value of 2, and the Ala. Ala is either in the introductory, or in the terminal, sequence.

The authors wish to thank Miss D. E. Pinion and Miss J. Robinson for assistance with some of the experimental work.

REFERENCES

- Abderhalden, E. & Bahn, A. (1933). *Hoppe-Seyl. Z.* **219**, 72.
- Brand, E. & Edsall, J. T. (1947). *Annu. Rev. Biochem.* **16**, 223.
- Cook, H. G. & Saunders, B. C. (1947). *Biochem. J.* **41**, 558.

- Corfield, M. C., Howitt, F. O. & Robson, A. (1954). *Nature, Lond.*, **174**, 603.
- Desnuelle, P. & Bonjour, G. (1951). *Biochim. biophys. Acta*, **7**, 451.
- Drucker, B., Hainsworth, R. & Smith, S. G. (1953). *Shirley Inst. Mem.* **26**, 191 (or *J. Text. Inst.* **44**, T420).
- Elliott, D. F. (1952). *Biochem. J.* **50**, 542.
- Elliott, D. F. (1953). *The Chemical Structure of Proteins, Ciba Foundation Symp.* p. 129. London: Churchill.
- Elliott, D. F. (1955). *Biochem. J.* **61**, 264.
- Fischer, E. (1902). *Chemikerztg.* **26**, 939.
- Holmes, F. & Smith, D. I. (1952). *Nature, Lond.*, **169**, 193.
- Joffe, H. G. (1954). *Biokhimiya*, **19**, 495.
- Kay, L. M. & Schroeder, W. A. (1954). *J. Amer. chem. Soc.* **76**, 3564.
- Kroner, T. D., Tabroff, W. & McGarr, J. J. (1953). *J. Amer. chem. Soc.* **75**, 4084.
- Levy, M. & Slobodian, E. (1952). *J. biol. Chem.* **199**, 563.
- Lucas, F., Shaw, J. T. B. & Smith, S. G. (1955). *Shirley Inst. Mem.* **28**, 77 (or *J. Text. Inst.* **46**, T440).
- Meyer, K. H., Fuld, M. & Klemm, O. (1940). *Helv. chim. acta*, **23**, 1441.
- Mills, G. L. (1952). *Biochem. J.* **50**, 707.
- Moore, S. & Stein, W. H. (1951). *J. biol. Chem.* **192**, 663.
- Neurath, H. & Schwert, G. W. (1950). *Chem. Rev.* **46**, 128.
- Sanger, F. (1949). *Biochem. J.* **45**, 563.
- Sanger, F. (1952). *Advanc. Protein Chem.* **7**, 1.
- Schroeder, W. A., Kay, L. M., Le Gette, J., Honnen, L. & Green, F. C. (1954). *J. Amer. chem. Soc.* **76**, 3556.
- Schroeder, W. A., Kay, L. M., Munger, N. & Burt, N. (1956). *J. Amer. chem. Soc.* **78**, 2430.
- Shaw, J. T. B. & Smith, S. G. (1954). *J. Text. Inst.* **45**, T939.
- Slobodian, E. & Levy, M. (1953). *J. biol. Chem.* **201**, 371.
- Synge, R. L. M. (1943). *Chem. Rev.* **32**, 142.
- Waldschmidt-Leitz, E. & Zeiss, O. (1954). *Hoppe-Seyl. Z.* **298**, 239.
- Waldschmidt-Leitz, E. & Zeiss, O. (1955). *Hoppe-Seyl. Z.* **300**, 49.
- Zahn, H. (1952). *Melliand Textilber.* **33**, 1076.
- Zahn, H. & Zuber, H. (1954a). *Mh. Chem.* **85**, 759.
- Zahn, H. & Zuber, H. (1954b). *Textilrdsch.* **9**, 119.
- Zahn, H., Zuber, H., Ditschen, W., Wegerle, D. & Meinenhofer, J. (1956). *Chem. Ber.* **89**, 407.

APPENDIX

Mathematical Derivation of the Yields of the Products of Hydrolysis of the Hexapeptide Ser.Gly.Ala.Gly.Ala.Gly

By F. LUCAS, J. T. B. SHAW, S. G. SMITH AND C. MACK

Shirley Institute, Manchester 20

(Received 12 November 1956)

The problem of relating the composition of a partial hydrolysate of a protein to its amino acid sequence and the hydrolysis constants of its constituent bonds is one of considerable interest to protein chemists, and has been discussed by Sanger (1952) and by Desnuelle (1953). Little progress has been possible in elucidating the problem because of the complex structure of most proteins, and because the rate of hydrolysis of any bond is likely to vary during the course of hydrolysis according to the particular peptide in which the bond occurs.

In the paper to which this is the Appendix, we have reported on a polypeptide (CTP) derived from silk fibroin; CTP is believed to be composed largely of the repeating hexapeptide units Ser.Gly.Ala.Gly.Ala.Gly. Consequently, although CTP has a molecular weight of approximately 4000, its structure is relatively simple, and the variety of its bonds is much less than normally occurs in proteins. Because of the lability of the Gly.Ser bond, this polypeptide would be expected to break down rapidly in concentrated acid to the simple hexapeptide units, and the yields of amino acids and peptides, which we have determined at various stages of the hydrolysis of the polypeptide, would

therefore approximate to those that would be derived from the hexapeptide unit.

As a contribution to the general problem therefore it seemed worth while to calculate the yields of the hydrolysis products by making some simple assumptions about the hydrolysis constants of the bonds. In addition to this it was hoped that the calculations would throw some light on a feature of the experimental results that was not obviously explained by the structure of the hexapeptide, namely the constant occurrence in the partial hydrolysate of Ala.Gly and Gly.Ala in the ratio 2:1.

It is known that polypeptides hydrolyse more rapidly than would be expected from the experimentally determined rates of their bonds in dipeptide form. It is supposed that this comparative stability of dipeptides is due to the effect of the charged NH_3^+ group, which tends to repel the H^+ ions, and, if this is so, it is reasonable to assume that an *N*-terminal peptide bond will also be more stable than the same bond when it occurs elsewhere within the peptide chain.

The rates of hydrolysis of the dipeptides Ser.Gly, Ala.Gly, and Gly.Ala were determined experimentally, the same conditions of hydrolysis being