REFERENCES

Brand, E. & Edsall, J. T. (1947). *Ann. Rev. Biochem.* **16**, 224.
Consden, R., Gordon, A. H. & Martin, A. J. P. (1944). *Biochem. J.* **38**, 224.
Consden, R., Gordon, A. H. & Martin, A. J. P. (1947). *Biochem. J.* **41**, 590.
Consden, R., Gordon, A. H. & Martin, A. J. P. (1948). *Biochem. J.* **42**, 443.
Consden, R., Gordon, A. H. & Martin, A. J. P. (1949). *Biochem. J.* **44**, 548.
Consden, R., Gordon, A. H., Martin, A. J. P. & Synge, R. L. M. (1947). *Biochem. J.* **41**, 596.
Dent, C. E. (1947). *Biochem. J.* **41**, 240.
Dent, C. E. (1948). *Biochem. J.* **43**, 169.
Desnuelle, P. & Casal, A. (1948). *Biochim. Biophys. Acta*, **2**, 64.
Jones, T. S. G. (1949). *Discuss. Faraday Soc.* **7**, 285.
Macpherson, H. T. (1946). *Biochem. J.* **40**, 470.
Partridge, S. M. (1948). *Biochem. J.* **42**, 238.
Partridge, S. M. & Davis, H. F. (1950). *Nature, Lond.*, **165**, 62.
Phillips, D. M. P. (1948). *Nature, Lond.*, **161**, 53.
Sanger, F. (1949*a*). *Biochem. J.* **44**, 126.
Sanger, F. (1949*b*). *Biochem. J.* **45**, 563.
Sanger, F. (1949*c*). *Cold Spr. Harb. Sym. quant. Biol.* **14**, 153.
Sanger, F. & Tuppy, H. (1951). *Biochem. J.* **49**, 481.
Synge, R. L. M. (1943). *Chem. Rev.* **32**, 135.
Synge, R. L. M. (1951). *Biochem. J.* (in the Press).
Theorell, H. & Åkeson, Å. (1942). *Ark. Kemi Min. Geol.* **16A**, no. 8.
Tiselius, A., Drake, B. & Hagdahl, L. (1947). *Experientia*, **3**, 651.
Tristram, G. R. (1949). *Advanc. prot. Chem.* **5**, 83.
Wieland, T. & Wirth, L. (1949). *Ber. dtsch. chem. Ges.* **82**, 468.

The Amino-acid Sequence in the Phenylalanyl Chain of Insulin

2. THE INVESTIGATION OF PEPTIDES FROM ENZYMIC HYDROLYSATES

BY F. SANGER (Beit Memorial Fellow) AND H. TUPPY*
Biochemical Laboratory, University of Cambridge

(*Received* 17 *January* 1951)

In the previous paper (Sanger & Tuppy, 1951) a large number of simple peptides in partial hydrolysates of fraction *B* of oxidized insulin were identified and several amino-acid sequences were deduced from them. It was not possible to determine the complete structure of the polypeptide chain, partly on account of the difficulty of fractionating the less polar peptides containing the aromatic amino-acids and partly because of the relatively great lability of certain bonds such as those involving the amino groups of the hydroxyamino-acids.

In the present paper are described investigations of the peptides produced on hydrolysis of fraction *B* with the proteolytic enzymes, pepsin, chymotrypsin and trypsin. These peptides, which are mostly larger than those present in the acid and alkaline hydrolysates, could be successfully fractionated by paper chromatography, and with the knowledge presented in the previous paper it was possible to determine the structure of most of them and to work out the complete sequence of fraction *B* and hence of the phenylalanyl chain of insulin.

The abbreviations, etc., defined in the previous paper (Sanger & Tuppy, 1951) will be used throughout and the peptides with reference numbers starting *B* 1, *B* 2, *B* 3, *B* 4, *B* 5 refer to those described in that paper.

* Present address: II. Chemisches Universitätslaboratorium, Vienna, Austria.

METHODS

Enzymes

Crystalline pig pepsin was obtained from Armour Laboratories. The chymotrypsin was prepared and several times recrystallized by Mr J. Lightbown of the National Institute for Medical Research, Mill Hill, London. The trypsin was a crystalline preparation from Armour Laboratories. It was recrystallized three times according to the method of Kunitz & Northrop (1936).

Hydrolyses

Experiment Bp (*pepsin*). Fraction *B* (30 mg.) was dissolved in 3 ml. 0·01 M-HCl. A solution of 0·3 mg. pepsin in 0·3 ml. 0·01 M-HCl was added, and the mixture incubated at 37° for 24 hr. After boiling to inactivate the pepsin, the solution was taken to dryness *in vacuo* to remove excess HCl and subjected to ionophoresis in the four-compartment cell of Synge described in the previous paper. This separated the hydrolysate into acid, neutral and basic fractions.

Experiment Bc (*chymotrypsin*). Fraction *B* (50 mg.) was dissolved in 5 ml. water and brought to pH 7·5–8·0 by the addition of $NaHCO_3$. To this was added 2·3 mg. chymotrypsin, the pH again adjusted and the solution incubated at 25° for 24 hr. It was then brought to about pH 3 by addition of HCl and taken to dryness in a desiccator. The hydrolysate was again separated into acidic, neutral and basic fractions by ionophoresis.

Experiment Bt (*trypsin*). Fraction *B* (25 mg.) was dissolved in 2·5 ml. water and the pH adjusted to 7·5 with $NaHCO_3$. Trypsin (1·2 mg.) dissolved in 1·2 ml. water was added and the mixture incubated at 37° for 24 hr. On acidification a

precipitate formed which was removed by centrifugation. It yielded on acid hydrolysis all the amino-acids that are present in fraction *B* and was not investigated further. The solution was then taken to dryness *in vacuo*.

Investigation of the peptides

The techniques used to investigate the peptides were the same as those described in the previous paper. Each fraction was studied on two-dimensional chromatograms using Whatman no. 4 or 3 filter papers. The latter is thicker and is useful when larger amounts of material are required. Butanol-acetic acid was found to be the most generally useful solvent for fractionating these large peptides, and this was usually preceded by cresol. Although in certain cases (e.g. Fig. 2) there were no apparent separations with the first solvent, it was found that much better results were obtained on two-dimensional than on one-dimensional chromatograms. Probably the first solvent removes traces of salt and other impurities and distributes each spot over a wider area.

With these larger peptides the amino-acids present were identified using two-dimensional chromatograms and *N*-terminal residues with the dinitrophenyl (DNP) technique. This was usually done both by identification of the free amino-acids and of the DNP-amino-acid present in a hydrolysate of the DNP-peptide. Due to the small amounts of peptides present it was not usually possible to identify the DNP-amino-acid completely on a number of different columns (Sanger, 1945), but it was usually put on a chromatogram that would distinguish it from other amino-acids likely to be present as terminal residues in the peptide.

Several of the peptides were subjected to further hydrolysis with a different enzyme. This was carried out in the following way: eluates of spots, usually from two chromatograms on Whatman no. 4 filter paper, were transferred to a polythene strip and taken to dryness in a desiccator. A solution (0·1 ml.) containing 0·1 mg. of the enzyme adjusted to a suitable pH was then added. The pepsin was dissolved in 0·01 N-HCl and the trypsin was dissolved in water and adjusted to pH 7·5 with $NaHCO_3$. The drops were then transferred to capillary tubes and incubated at 37° for 24 hr. After putting on to polythene strips again they were acidified, taken to dryness and fractionated on a suitable two-dimensional chromatogram.

RESULTS

Experiment Bp (pepsin)

The results obtained with the neutral and basic fractions of the hydrolysate are recorded in Figs. 1 and 2 and Table 1. The acidic fraction was investigated in three of the experiments. No significant amounts of peptides were obtained in two cases, but in the third peptides 1, 2, 4, 11, 12 and 13 (Table 1) could be detected. The different results were probably due to slight differences of pH in the middle compartment during ionophoresis.

Spot 1. Contains traces of Val.Glu which can be separated on a phenol chromatogram.

Eluates of spot 1 from two phenol/butanol-acetic acid chromatograms were hydrolysed 48 hr. in 10 N-HCl at 37° and the hydrolysate fractionated on a phenol/butanol-acetic acid chromatogram. Besides free amino-acids the peptides Glu.Ala (×), Val.Glu (× ×) and unchanged tripeptide (× × ×) were identified. This demonstrates that its structure must be Val.Glu.Ala. It is the same as spot *B* 1γ10 and has the same R_F values on phenol and cresol. The Val.Glu present has the same R_F value as peptide *B* 1γ9. This indicates that they are probably not present as amides but with the γ-carboxyl group of glutamic acid free.

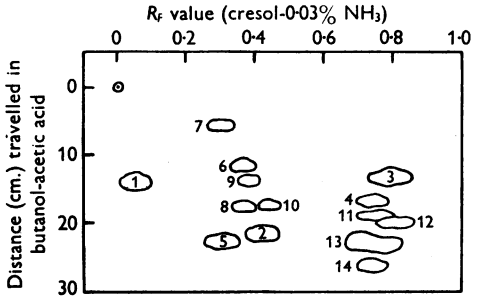


Fig. 1. Chromatogram of the neutral fraction of the peptic hydrolysate (*Bp*). In Figs. 1–6 the fraction was applied in the position marked by the small dotted circle, and the chromatogram run first in the direction represented horizontally, which except in Figs. 2 and 4 was the longer dimension of the filter paper. In each case both solvents were run for one length of the filter paper only (see Table 1).

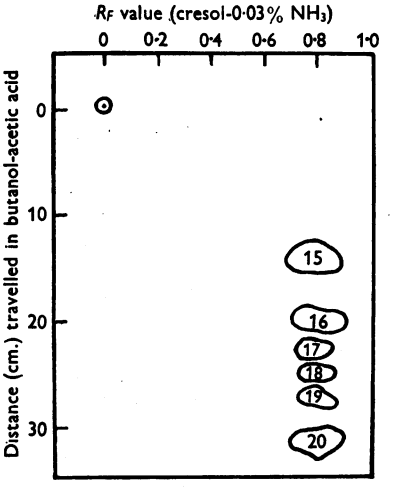


Fig. 2. Chromatogram of the basic fraction of the peptic hydrolysate (*Bp*). The chromatogram was run first down the short dimension of the filter paper with cresol-0·03% NH_3 (see Table 1).

Spot 2. This is not the same as spot *B* 2β15, which contains the same amino-acids, but has leucine as the *N*-terminal residue.

Eluates from two chromatograms were hydrolysed 3 days in 10 N-HCl at 37° and fractionated on a cresol/butanol-acetic acid chromatogram. Besides free amino-acids and small amounts of tripeptides, the dipeptides Val.Glu and Ala.Leu were identified and were present in considerable amounts. From this result it is concluded that the tetrapeptide is Val.Glu.Ala.Leu.

Spot 3. The results in Table 1 indicate that phenylalanine is the *N*-terminal residue. This was confirmed by showing

Table 1. *Peptides from peptic hydrolysate (Bp)*

(Amino-acids formed after hydrolysis of the peptide (*H*) or of the DNP-peptide (*D*).)

Spot no. (Figs. 1 and 2)	*H* or *D*	Amino-acids formed after hydrolysis															
		$CySO_3H$	Asp	Glu	Ser	Gly	Thr	Ala	Lys	Tyr	TyrX	His	Val	Leu	Arg	Phe	Pro
1	*H*	—	—	× × ×	—	—	—	× × ×	—	—	—	—	× × ×	—	—	—	—
	D	—	—	× ×	—	—	—	× ×	—	—	—	—	—	—	—	—	—
2	*H*	—	—	× ×	—	—	—	× ×	—	—	—	—	× ×	× ×	—	—	—
	D	—	—	× ×	—	—	—	× ×	—	—	—	—	?	× ×	—	—	—
3	*H*	× ×	× ×	× ×	× ×	× ×	—	?	—	—	—	× × ×	× ×	× × ×	—	× ×	—
	D	×	×	×	×	× ×	—	?	—	—	—	×*	×	× ×	—	?	—
4	*H*	× ×	—	× ×	—	× × ×	—	—	—	—	—	—	× ×	× ×	× ×	× ×	—
5	*H*	—	—	×	—	—	—	—	—	—	—	—	×	×	—	—	—
6	*H*	?	—	?	—	×	—	—	—	—	—	—	?	?	?	—	—
7	*H*	—	?	?	—	—	—	—	—	—	—	—	?	—	—	—	—
8	*H*	—	×	×	—	—	—	—	—	—	—	—	×	—	—	?	—
9	*H*	—	—	—	—	—	—	—	—	×	—	—	—	—	—	—	—
10	*H*	—	—	—	—	—	—	—	—	—	× ×	—	—	—	—	—	—
11	*H*	× ×	—	× ×	—	× × ×	—	—	—	—	—	—	× ×	× ×	× ×	× × ×	—
12	*H*	× ×	—	× ×	—	× × ×	—	?	—	× ×	—	—	× ×	× ×	× ×	× ×	—
13	*H*	× ×	—	× ×	—	× × ×	—	(× ×)	—	(×)	—	—	× ×	× ×	× ×	× ×	—
14	*H*	—	—	—	—	—	—	×	—	×	—	—	—	× ×	—	× ×	—
15	*H*	—	—	—	—	—	× ×	× ×	× ×	× ×	—	—	—	—	—	—	× ×
	D	—	—	—	—	—	× ×	× ×	×*	—	—	—	—	—	—	—	×
16	*H*	—	—	—	—	—	× × ×	× × ×	× × ×	—	× ×	—	—	—	—	—	× ×
	D	—	—	—	—	—	× ×	× ×	×*	—	—	—	—	—	—	—	×
17	*H*	× ×	—	—	× ×	× ×	—	—	—	—	—	× × ×	—	× × ×	—	—	—
18	*H*	× ×	× ×	× ×	× ×	× × ×	—	—	—	—	—	× × ×	× ×	× × ×	—	×	—
19	*H*	×	×	×	×	×	?	×	×	—	?	×	×	×	—	×	?
20	*H*	×	?	×	×	×	?	×	×	—	?	×	?	×	—	×	?

* DNP derivative.

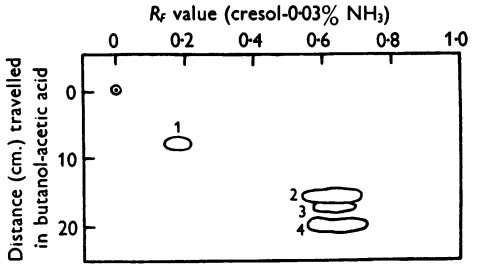


Fig. 3. Chromatogram of the neutral fraction of the chymotryptic hydrolysate (*Bc*) (see Table 2).

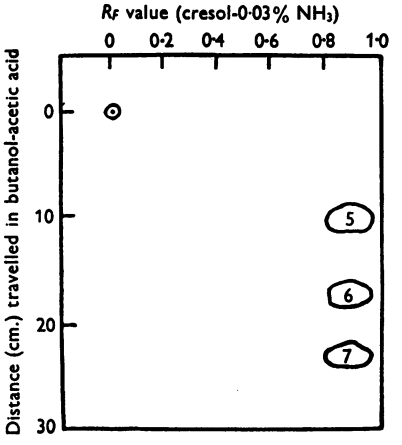


Fig. 4. Chromatogram of the basic fraction of the chymotryptic hydrolysate (*Bc*), run first down the short side of the paper with cresol-0·03 % NH_3 (see Table 2).

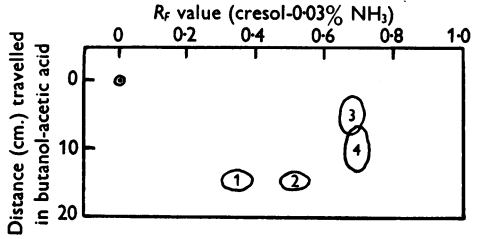


Fig. 5. Chromatogram of peptic hydrolysate of fraction *Bc*2 (see Table 2).

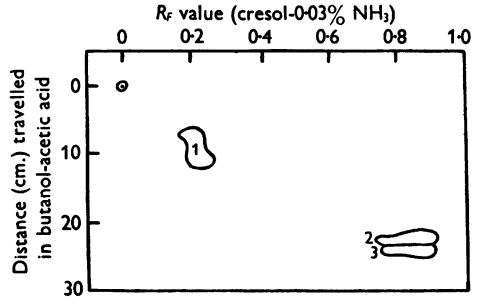


Fig. 6. Chromatogram of the tryptic hydrolysate (*Bt*) (see Table 2).

Table 2. *Peptides from chymotryptic (Bc) and tryptic (Bt) hydrolysates*

Spot no. (Figs. 3, 4, 5 and 6)	Amino-acids formed after hydrolysis																			
	$CySO_3H$	Asp	Glu	Ser	Gly	Thr	Ala	Lys	Tyr	TyrX	His	Val	Leu	Arg	Phe	Pro				
*Bc*1	×				×								×	×						
*Bc*1 (DNP)	×		×		×									×						
*Bc*2	×	×	×	×	×		×		(?)	(×)	×	×	×		×					
*Bc*3	×	×	×	×	×		×		(×)	(×)	×	×	×		×					
*Bc*4	×	×	×	×	×		×				×	×	×	×	×					
*Bc*5					×	×	×	×	×					×	×	×				
*Bc*6						×	×	×	×	×						×				
*Bc*7						×	×	×		×						×				
*Bc*2*p*1			×				×		×			×	×							
*Bc*2*p*2			×				×					×	×							
*Bc*2*p*3			×	×	×		×				×	×	×		×					
*Bc*2*p*4			×	×	×		×				×	×	×		×					
*Bt*1							×									×				
*Bt*2							×		×							×				
*Bt*3										×						×				
*Bt*2*p*1									×							×				
*Bt*2*p*2																				

that the DNP-amino-acid produced on hydrolysis of the DNP-peptide moved at the same R value as DNP-phenylalanine on an ethanol-ligroin column (Sanger, 1945). This distinguishes it from the DNP derivatives of the other amino-acids present in spot 3, except possibly glycine.

Spot 4. On hydrolysis of the DNP derivative, a DNP-amino-acid was produced that was identified as DNP-leucine or DNP-valine. Chromatography of the remaining amino-acids with benzyl alcohol showed strong spots for valine and phenylalanine and only a weak leucine spot, indicating that leucine was the N-terminal residue.

Eluates from two chromatograms were subjected to hydrolysis with trypsin. The hydrolysate was fractionated on a half sheet of Whatman no. 1 filter paper using cresol-0·03 % NH_3/butanol-acetic acid. Only two spots were present, whose R_F values and compositions are shown in Table 3. *Bp*4*t*2 is the same as spot *B* 1γ12*a* (Gly.Phe). Further results with this peptide are given in connexion with peptide *Bc*4.

Spot 1. When a different preparation of chymotrypsin (obtained by activation of chymotrypsinogen with trypsin) was used this spot was much stronger. It was also present in hydrolysates obtained by the use of unrecrystallized trypsin, and thus appears to be produced by the combined action of chymotrypsin and trypsin on fraction *B*. The small amount present in this experiment may be due to traces of trypsin in the preparation.

Spot 2. This sometimes contains varying amounts of tyrosine and TyrX.

A hydrolysate of the DNP derivative was put on a benzyl alcohol chromatogram. It gave a weak phenylalanine spot but strong spots for leucine and valine, showing phenylalanine to be the N-terminal residue.

The eluates of spot 2 from two chromatograms were treated with pepsin and the hydrolysate fractionated on a half sheet of Whatman no. 1 filter paper using cresol/butanol-acetic acid. The results are shown in Fig. 5 and

Table 3. *Tryptic hydrolysis of fractions Bp* 4 *and Bc* 4

Fraction	R_F value (cresol-0·03 % NH_3)	Distance in butanol-acetic acid (cm.)	Strength	Composition
*Bp*4*t*1	0·19	5·6	× × ×	$CySO_3H$, Glu, Gly, Val, Leu, Arg
*Bp*4*t*2	0·66	14·4	× × ×	Gly, Phe
*Bc*4*t*1	0·21	6·9	× ×	$CySO_3H$, Glu, Gly, Val, Leu, Arg
*Bc*4*t*2	0·87	20·8	× ×	Gly, Phe

Spot 5. A small amount of Leu.Val.Glu identical with peptide *B* 2β14.

Spot 7. Trace of Val.Asp.Glu.

Spot 8. Phe.Val.Asp.Glu. Peptides 7 and 8 probably contain glutamine and asparagine residues (Sanger, 1949*b*) and are therefore not identical with the corresponding peptides from the acid hydrolysates (*B* 1β10, *B* 1β12 respectively).

Spots 9 *and* 10. Free tyrosine and TyrX respectively.

Spots 11 *and* 12. In most experiments these two spots were not clearly separated and their composition varied considerably. The results are from a single experiment. Usually they contain free phenylalanine and leucine and at least two peptides, which contain arginine but no aspartic acid and have variable amounts of tyrosine.

Spot 13. Contains arginine but no aspartic acid. The amounts of tyrosine and alanine varied somewhat from one experiment to another. Alanine varies independently of the other amino-acids so is probably part of a low-molecular peptide (e.g. Ala.Leu). Tyrosine is probably connected with the larger peptide.

Spot 14. Probably a mixture of fast-moving small peptides.

Spot 15. The presence of tyrosine as an N-terminal residue was confirmed by showing that hydrolysis of the DNP derivative gave rise to a DNP-amino-acid that moved on a $CHCl_3$ column at the same rate as *ON*-bis-DNP-tyrosine. This distinguishes it from DNP-threonine.

Spot 16. The same as 15, but containing TyrX instead of tyrosine.

Spot 18. The same as spot 3 (neutral fraction).

Experiment Bc (chymotrypsin)

The results obtained with the chymotryptic hydrolysate are given in Figs. 3 and 4 and Table 2.

Table 2. Spots *Bc*2*p*1 and *Bc*2*p*3 are apparently the same as *Bp*2 and *Bp*3 respectively, which are found in the peptic hydrolysate.

Spot 3. This contains varying amounts of tyrosine and TyrX, showing that it is impure.

A hydrolysate of the DNP derivative was shown on a benzyl alcohol chromatogram to contain a trace of phenylalanine and large amounts of leucine and valine, indicating that phenylalanine was the N-terminal residue.

A peptic hydrolysate gave a chromatogram similar to that obtained with spot *Bc*2 (Fig. 5). Spots corresponding to *Bc*2*p*1 and *Bc*2*p*2 were present but only one other large spot in a position corresponding to that of *Bc*2*p*3 and *Bc*2*p*4. This contained all the amino-acids of the original material, and was thus probably a mixture of *Bc*2*p*3 and *Bc*2*p*4 with unhydrolysed *Bc*3.

Table 4. *Amino-acid composition of peptides Bp* 4*t*2 *and Bc* 4*t*2

(Results expressed as μmol. amino-acid in peptide originating from 10 mg. fraction *B*.)

Peptide	Glycine	Phenylalanine
*Bp*4*t*2	0·3	0·3
*Bc*4*t*2	0·1	0·23

Spot 4. An eluate from one chromatogram was hydrolysed with trypsin and fractionated on cresol-0·03 %NH_3/butanol-acetic acid. Two spots were obtained and the results are shown in Table 3 together with those for the corresponding spots from peptide *Bp*4. Peptides *Bp*4*t*1 and *Bc*4*t*1 are apparently identical. However, peptide *Bc*4*t*2 is different from *Bp*4*t*2 since it moves considerably faster with both solvents. It gives a yellow colour with ninhydrin suggesting that glycine is the N-terminal residue. In Table 4 are shown

the results of an experiment in which the glycine and phenylalanine in the two peptides were estimated by semi-quantitative paper chromatography. Known quantities of authentic amino-acid were run in parallel on a chromatogram and the strength of the ninhydrin spots compared visually. Whereas *Bp* 4*t*2 contains the two amino-acids in equal proportions, *Bc* 4*t*2 has two phenylalanine residues to one glycine residue.

Experiment Bt (trypsin)

The soluble material from the tryptic hydrolysate was fractionated directly without any preliminary group separation. Fig. 6 shows a typical chromatogram. Only three significant spots were present, and their composition is shown in Table 2.

Spot 1. Contained only alanine, and its R_F values and strong ninhydrin reaction indicated that it was the free amino-acid.

Spot 2. The slower moving parts of the spot frequently contained traces of glutamic acid, alanine, valine and leucine, which do not appear to be an integral part of the peptide. It gave a yellow colour with ninhydrin, suggesting that glycine or threonine was the *N*-terminal residue. Hydrolysis of the DNP derivative gave a DNP-amino-acid that moved on a $CHCl_3$ column at the same rate as DNP-glycine (or bis-DNP-lysine or *ON*-bis-DNP-tyrosine) and the spot for free glycine was considerably weaker than in the hydrolysate of the untreated peptide. It was thus concluded that glycine is the *N*-terminal residue.

Eluates from two chromatograms were subjected to hydrolysis with pepsin, and the hydrolysate was fractionated on a collidine/benzyl alcohol chromatogram. Only two spots were identified (see Table 2). *Bt* 2*p*1 had $R_F = 0{\cdot}35$ in collidine and had travelled 7 cm. in benzyl alcohol. *Bt* 2*p*2 had $R_F = 0{\cdot}46$ and had travelled 11·5 cm. It is not clear from these results whether peptide *Bt* 2*p*2 is identical with *Bp* 4*t*2 ([Gly, Phe]) or with *Bc* 4*t*2 ([Gly, Phe, Phe]).

Spot 3. The same as spot 2, but containing TyrX instead of tyrosine.

DISCUSSION

With the rather larger peptides present in these enzymic hydrolysates, it was important to know if certain spots contained only one peptide or whether they could be mixtures. Although direct proof of homogeneity is difficult the following factors could be taken into account in assessing the purity of a spot:

(1) Constant amino-acid composition over different parts of the spot and of the same spot from different chromatograms.

(2) Comparison of the R_F value of the spot with that to be expected if all the amino-acids were present in a single peptide and with the R_F values of possible components of a mixture (see Martin, 1949).

(3) The presence of only one *N*-terminal residue is probably the best evidence of purity, since it is unlikely that two different peptides would have the same *N*-terminal residue.

(4) Consideration of the sequences previously shown to be present in fraction *B* would often indicate whether a given spot was likely to be pure or impure.

Structure of the peptides

In the previous paper the sequences Phe.Val.Asp.Glu.His.Leu.$CySO_3H$.Gly, Thr.Pro.Lys.Ala, Gly.Glu.Arg.Gly, Tyr.Leu.Val.$CySO_3H$.Gly and Ser.His.Leu.Val.Glu.Ala were shown to be present in fraction *B*. These sequences account for all the residues present except two phenylalanine, one tyrosine and possibly one leucine and one glutamic acid residue. Other peptides present were Ala.Leu, Gly.Phe and Ala.[Tyr, Leu].

Peptide *Bp* 3 contains aspartic acid and hence at least part of the *N*-terminal sequence Phe.Val.Asp.Glu.His.Leu.$CySO_3H$.Gly, in which the only aspartic acid residue in fraction *B* is present. It has a phenylalanine *N*-terminal residue and contains all the amino-acids present in this sequence and also serine. There is only one serine residue in fraction *B* and this is present in the sequence Ser.His.Leu.Val.Glu.Ala, so that this establishes a connexion between these two sequences. None of the other sequences above could come between the glycine and serine residues, since peptide *Bp* 3 contains no threonine, proline, lysine, arginine or tyrosine. The possibility that a phenylalanine, leucine or glutamic acid residue may come between the glycine and serine should be considered since all of these residues in fraction *B* have not yet been accounted for. Peptide *Bp* 17, which has cysteic acid, glycine, serine, histidine and leucine but no valine, must contain parts of both the above two sequences. It contains no phenylalanine or glutamic acid, so that neither of these two can be between the glycine and serine residue. It is very unlikely that there is a leucine residue here since Gly.Leu has never been observed, whereas the similar Gly.Phe was detected. Also it will be shown below that all the leucine residues of fraction *B* can be accounted for elsewhere. Thus it may be concluded that the *N*-terminal peptide sequence of fraction *B* is Phe.Val.Asp.Glu.His.Leu.$CySO_3H$.Gly.Ser.His.Leu.Val.Glu.Ala. The Gly.Ser bond would be extremely labile and was therefore not found intact in any of the peptides obtained from acid or alkaline hydrolysis. The exact structure of peptide *Bp* 3 cannot be deduced from the present results, but it most probably contains the first eleven residues of the above sequence, since the presence of Val.Glu.Ala.Leu (*Bp* 2) shows that the Leu.Val bond is split by pepsin. It must contain at least the first nine residues as it has serine but not more than thirteen residues as alanine is absent. Peptide *Bp* 17 is most probably His.Leu.$CySO_3H$.Gly.Ser.His.Leu. This is suggested by the presence of Phe.Val.Asp.Glu (*Bp* 8) and Val.Asp.Glu (*Bp* 7) in the peptic hydrolysate, which demonstrates a lability of the Glu.His bond to pepsin.

Peptide *Bp* 2 was shown by partial hydrolysis to be Val.Glu.Ala.Leu. This must be joined to the above

N-terminal sequence, since the other alanine residue of fraction *B* is in the form Lys.Ala. Ala.Leu (*B* 1δ11) was found in the acid hydrolysate. Further evidence for this *N*-terminal sequence is derived from peptides *Bc* 2 and *Bc* 3. These contain the amino-acids of the above sequence and a phenylalanyl *N*-terminal residue. They also have alanine and varying amounts of tyrosine, which shows they are longer than *Bp* 3. On hydrolysis with pepsin *Bc* 2 (Fig. 5, Table 2) gives a peptide (*Bc* 2*p*3) which appears to be identical with *Bp* 3 and also Val.Glu.Ala.Leu (*Bc* 2*p*1) identical with *Bp* 2. It also gives a peptide [Glu, Ala, Tyr, Val, Leu] (*Bc* 2*p*2). This can only be Val.Glu.Ala.Leu.Tyr and establishes the position of tyrosine on the *N*-terminal sequence, which must therefore be Phe.Val.Asp.Glu.His.Leu.$CySO_3H$.-Gly.Ser.His.Leu.Val.Glu.Ala.Leu.Tyr. This will be referred to as sequence 1. Further evidence for this structure comes from peptide *B* 3β9, which must be Ala.Leu.Tyr. Leu.Tyr is probably present in spot *B* 3β10. The indefinite result with the DNP method suggests both Leu.Tyr and Tyr.Leu are present.

The variation in the amounts of tyrosine in spots *Bc* 2 and *Bc* 3 suggests that they are probably mixtures of closely related peptides containing the *N*-terminal sequence and varying amounts of leucine, tyrosine and TyrX. It is probable that pepsin can split the bonds on both sides of this tyrosine residue.

Fraction *B* contains three glycine residues, one of which is present in the above sequence; others are in the two sequences Tyr.Leu.Val.$CySO_3H$.Gly and Gly.Glu.Arg.Gly, which must therefore be parts of the same sequence. Thus the sequence in fraction *B* must be Tyr.Leu.Val.$CySO_3H$.Gly.Glu.Arg.Gly. The only other peptide containing glycine is Gly.Phe (*B* 1γ12*a*) which was obtained from the acid hydrolysate. It can only be combined to the *C*-terminal glycine of the above octapeptide establishing the sequence Tyr.Leu.Val.$CySO_3H$.-Gly.Glu.Arg.Gly.Phe.

A number of peptides from the pepsin and chymotrypsin hydrolysates offer further evidence for this sequence. Peptide *Bp* 4 contains arginine as well as valine and cysteic acid and has leucine as the *N*-terminal residue, thus demonstrating the connexion between the two sequences Tyr.Leu.Val.$CySO_3H$.-Gly and Gly.Glu.Arg.Gly. On hydrolysis with trypsin two peptides are produced. One (*Bp* 4*t*2) is Gly.Phe, the other (*Bp* 4*t*1) contains all the amino-acids of *Bp* 4 except phenylalanine. It is probably identical with peptide *Bc* 1, which was very strong when what appears to be a mixture of trypsin and chymotrypsin was used for hydrolysis. It contains a leucine *N*-terminal residue and can only be Leu.Val.$CySO_3H$.Gly.Glu.Arg; *Bp* 4 must be Leu.Val.$CySO_3H$.Gly.Glu.Arg.Gly.Phe, the trypsin splitting after the arginine residue.

Peptide *Bc* 4 is also derived from this sequence and contains the same amino-acids as *Bp* 4. On hydrolysis with trypsin it gives the same hexapeptide Leu.Val.$CySO_3H$.Gly.Glu.Arg (*Bc* 4*t*1), but instead of Gly.Phe it gives a spot (*Bc* 4*t*2) moving considerably faster, but containing only these two amino-acids. A semi-quantitative analysis (Table 4) indicated that the ratio of glycine to phenylalanine in this peptide is 1:2, and the only possible conclusion is that it is Gly.Phe.Phe and that peptide *Bc* 4 is Leu.Val.$CySO_3H$.Gly.Glu.Arg.Gly.Phe.-Phe. The sequence in fraction *B* is thus Tyr.Leu.-Val.$CySO_3H$.Gly.Glu.Arg.Gly.Phe.Phe, which will be referred to as sequence 2.

Spots *Bp* 11, *Bp* 12, *Bp* 13 are probably mixtures of various peptides containing this sequence, with varying amounts of tyrosine, alanine, leucine and phenylalanine. On both sides of this sequence there are several bonds that may be split by pepsin, so that one would expect to find a large number of closely related peptides. Peptide *Bp* 6 is probably a trace of Leu.Val.$CySO_3H$.Gly.Glu.Arg.Gly. It is unlikely that pepsin would split the bond next to arginine, whereas it would be expected to split the bond before the phenylalanine residue.

The only other sequence deduced from the acid hydrolysis experiments that we have not so far considered is Thr.Pro.Lys.Ala, which is present as the tetrapeptide in spot *Bc* 5. Peptides *Bc* 6, *Bc* 7, *Bp* 15, *Bp* 16 all contain the same four residues together with tyrosine or TyrX which is their *N*-terminal residue. Their structure must therefore be Tyr.Thr.Pro.-Lys.Ala. Peptide *Bt* 2 contains threonine, proline and lysine which are unique to this sequence and also phenylalanine, which must be derived from the *C*-terminal end of sequence 2, thus establishing a connexion between these two sequences. Peptide *Bt* 2 has a glycyl *N*-terminal residue and must therefore be Gly.Phe.Phe.Tyr.Thr.Pro.Lys. On hydrolysis with pepsin it gives Tyr.Thr.Pro.Lys (*Bt* 2*p*1) and a peptide (*Bt* 2*p*2) containing glycine and phenylalanine, probably Gly.Phe.Phe. Sequence 2 must now be Tyr.Leu.Val.$CySO_3H$.Gly.Glu.Arg.Gly.-Phe.Phe.Tyr.Thr.Pro.Lys.Ala.

Structure of fraction B

Fraction *B* contains only two tyrosine residues. One is accounted for within sequence 2, the other is present both as the *C*-terminal residue of sequence 1 and as the *N*-terminal residue in sequence 2. The two sequences must therefore have this tyrosine residue in common, and the structure of fraction *B* must be Phe.Val.Asp.Glu.His.Leu.$CySO_3H$.Gly.Ser.-His.Leu.Val.Glu.Ala.Leu.Tyr.Leu.Val.$CySO_3H$.-Gly.Glu.Arg.Gly.Phe.Phe.Tyr.Thr.Pro.Lys.Ala. This structure has alanine as a *C*-terminal residue. This confirms the results of Lens (1949), who studied the action of carboxypeptidases on insulin and

showed that alanine was present as a *C*-terminal residue. Fromageot, Jutisz, Meyer & Pénasse (1950) and Chibnall & Rees (1951) have also independently identified alanine as the *C*-terminal residue in insulin using their respective methods of reduction to β-amino alcohols. The liberation of free alanine (*Bt* 1) by the action of trypsin on fraction *B* also supports this conclusion. Studies with synthetic peptides have shown that trypsin splits bonds involving the carboxyl groups of arginine or lysine. It will therefore be expected to split the bond before the alanine residue, but not that after it. Free alanine could only be produced if it is already present as a *C*-terminal residue.

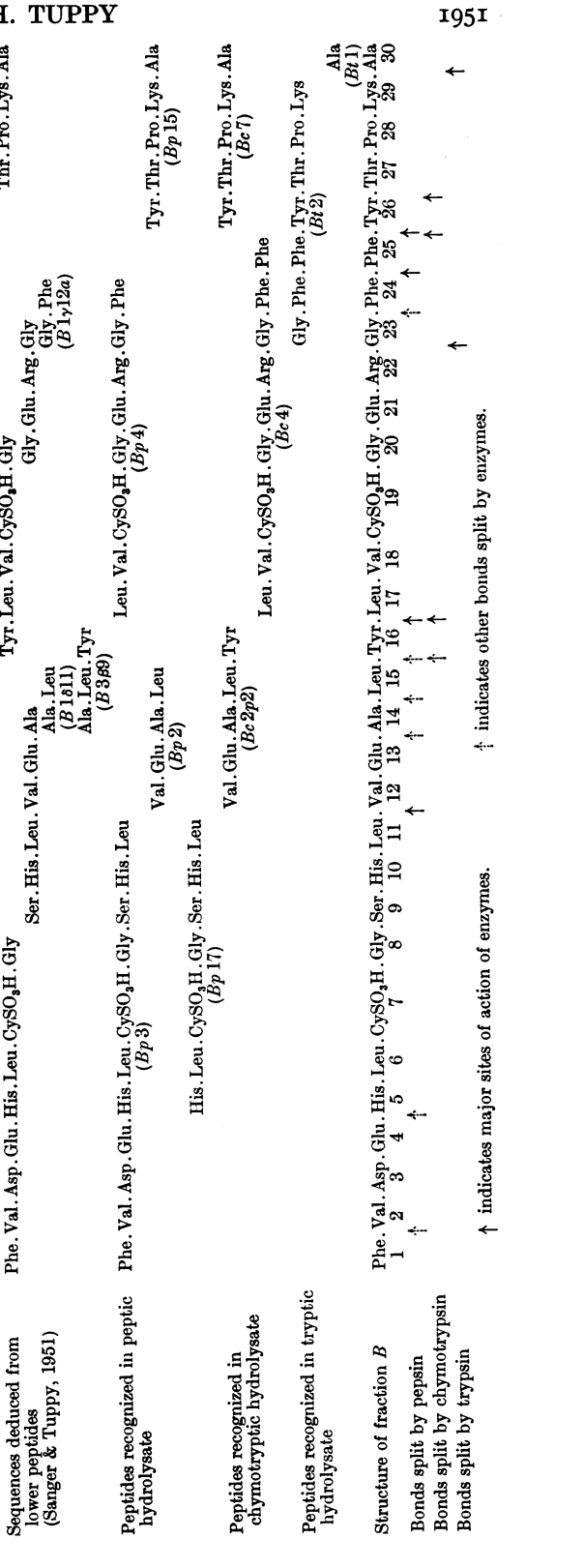
Table 5, which summarizes the foregoing discussion, gives the structure of fraction *B* and of the various peptides that have played a major part in the elucidation of the sequence. The residues have been numbered for convenience in the subsequent discussion. The phenylalanyl chains of insulin have the same structure as fraction *B* except that the $CySO_3H$ residues are replaced by CyS (half-cystine) residues.

The above sequence explains the difficulties experienced in fractionating the fast-moving aromatic peptides (Sanger & Tuppy, 1951). A large proportion of the least polar residues are concentrated together in the two sequences Leu.Tyr.Leu.-Val and Phe.Phe.Tyr, which can be expected to give rise on partial hydrolysis to a large number of closely related peptides having similar R_F values. The difficulty is increased by the presence of TyrX.

Further evidence for the presence of the sequence Phe.Phe was obtained from a chromatogram of an aromatic fraction which was prepared from a 9-day low-temperature acid hydrolysate of insulin by adsorption on charcoal. A peptide spot was identified which had $R_F = 0{\cdot}61$ on collidine and 0·54 on benzyl alcohol and contained only phenylalanine. The free amino-acid had $R_F = 0{\cdot}40$ on collidine, 0·30 on benzyl alcohol on the same chromatogram. The only possible explanation is that the faster moving spot was Phe.Phe.

Since synthetic reactions have been shown to take place in the presence of proteolytic enzymes, the possibility that some of the peptides identified could have been formed by rearrangement of peptide bonds during the course of hydrolysis should be considered. While such a rearrangement would seem rather unlikely, we have attempted as far as possible to obtain evidence for the presence of each bond, either by identifying it in a peptide obtained by acid hydrolysis or else by identifying it in peptides produced by the action of at least two different enzymes. It is extremely unlikely that the same bond would be synthesized in both cases. The only bond to which this does not apply is the Phe.Tyr bond (positions 25, 26). This was intact only in the peptide *Bt* 2, but it

Table 5. *The structure of fraction B of oxidized insulin*



is very improbable that such a bond would be formed by the action of trypsin, since the specificity for synthetic reactions should be similar to that for hydrolytic reactions. Moreover, if any extensive rearrangements had occurred it would almost certainly have been impossible to interpret the results in terms of a single amino-acid sequence for fraction *B*.

The question of location of the amide groups is not completely settled by the present work. The results with the *N*-terminal residues (Sanger, 1949*b*) showed that the aspartic and glutamic acid residues in positions 3 and 4 are in the form of amides, since several different DNP-peptides containing the four residues phenylalanine, valine, aspartic and glutamic acids were detected. According to the amide analysis of fraction *A* (Sanger, 1949*a*) and of insulin (Rees, 1946), there would appear to be only two amide groups in fraction *B*, so that the other two glutamic acid residues (positions 13, 21) would be expected to have free γ-carboxyl groups. If an amide residue is present in a protein it would be expected to be found still intact in most peptides derived from an enzymic hydrolysate, but with the amide group removed in the corresponding peptides from an acid hydrolysate. The peptides Val.Glu.Ala (*Bp*1), Val.Glu (also in *Bp*1) and Leu.Val.Glu (*Bp*5) derived from the peptic hydrolysate all have the same R_F values as the corresponding peptides from the acid hydrolysates (*B*1γ10, *B*1γ9, *B*2β14 respectively), which suggests that they are not present as amides and that the γ-carboxyl group of the glutamic acid residue in position 13 is free in insulin. No such evidence is available for the glutamic acid residue in position 21, but the fact that peptide *Bp*4 (Leu.Val.$CySO_3H$.Gly.Glu.Arg.Gly.Phe) is sometimes found in the acidic fraction from the ionophoresis suggests that the γ-carboxyl group of the glutamic acid residue is also free here.

In recent years several hypotheses of protein structure have been advanced which are based on the assumption of some type of periodic arrangement of amino-acids along the polypeptide chains. It is tempting to assume that this arrangement, which is determined by the unknown mechanisms of protein synthesis, may be rather simple and may follow certain easily discernible principles. An examination of the structure of fraction *B*, however, fails to reveal any simple periodic arrangement of the residues, nor is it possible to formulate any general principles which might govern the order of amino-acids along the protein chains. It may be that when much more extensive data are available such principles may become apparent; meanwhile two interesting features of the structure, which may or may not be significant, seem worth noting.

The presence of three aromatic residues joined together in the sequence Phe.Phe.Tyr is of interest and suggests that there may be a certain tendency for the less polar residues to be grouped together. The sequence Leu.Tyr.Leu.Val is another similar grouping. Such an arrangement might be of importance in connexion with the physical or biological properties of the insulin.

In the structure given for fraction *B* there are three dipeptide sequences which occur twice in the chain. These are His.Leu (positions 5–6, 10–11), $CySO_3H$.Gly (positions 7–8, 19–20) and Leu.Val (positions 11–12, 17–18). This is more than one would expect from a mere random arrangement and might suggest a certain limitation in the mechanisms by which proteins are synthesized or it might indicate that dipeptide units are the essential building stones of proteins. It is interesting to note that one of these sequences (CyS.Gly) also occurs in glutathione.

It is difficult to say if either of these features will prove to have general significance. They certainly would not be expected if the arrangement were entirely random, but it is probably more pertinent to consider how many of the possible arrangements of the amino-acids found in fraction *B* would not show some interesting or unusual features of this type.

Action of the proteolytic enzymes on fraction B

Since the purpose of this investigation was primarily to determine the structure of fraction *B*, all the precautions necessary for a careful study of the specificity of proteolytic enzymes have not been taken, nor has the exact structure of all the peptides found been determined. Although it is believed that all the enzymes used are essentially homogeneous, no rigorous test of purity was used, nor was it possible to control accurately the pH in the very small volumes of the tryptic and chymotryptic hydrolysis mixtures. However, certain conclusions concerning the action of the enzymes can profitably be drawn, since this is the first time that such a large polypeptide has been studied. In Table 5 are indicated the various linkages that are split by the three enzymes used. It will be seen that the specificities of trypsin and chymotrypsin, which were defined by Bergmann and his associates using synthetic substrates, are almost completely applicable to this large polypeptide and thus probably to proteins also.

The only two bonds that are split to any extent by trypsin are those involving the carboxyl groups of the arginine and lysine residues. An exactly similar specificity was found by Bergmann, Fruton & Pollok (1939) and by Hofmann & Bergmann (1939) using synthetic substrates.

According to Bergmann & Fruton (1937) chymotrypsin splits those bonds involving the carboxyl groups of the aromatic residues phenylalanine and

tyrosine. A typical substrate was *N*-carbobenzyloxytyrosylglycine amide. In fraction *B* three of the aromatic residues are grouped together in the sequence Phe.Phe.Tyr. Since chymotrypsin does not split bonds which are adjacent to a free carboxyl group, the manner in which it would split this sequence would be expected to be rather complex. The present results indicate that the main splitting takes place after the two tyrosine residues (positions 16, 26) and after the third phenylalanine residue (position 25). The small amount of tyrosine in peptide *Bc* 2 suggests that some splitting of the Leu.Tyr bond (positions 15, 16) also occurs.

Pepsin appears to have a considerably wider range of specificity than the other two enzymes. Fruton & Bergmann (1939) showed that bonds involving the amino groups of the aromatic residues were readily split in synthetic substrates. Harington & Pitt-Rivers (1944) demonstrated that Tyr.CySH is also hydrolysed. Here the bond involves the carboxyl group of the tyrosine residue. The present results also suggest that splitting takes place on both sides of the aromatic residues, though here again the situation is rather complex, since in the two sequences Ala.Leu.Tyr.Leu.Val and Phe.Phe.Tyr most of the bonds would be susceptible to hydrolysis. It does not, however, appear that the specificity is limited to bonds adjacent to aromatic residues. Thus the Leu.-Val bond (positions 11, 12) is quite strongly hydrolysed. This is evident since Val.Glu.Ala.Leu (*Bp* 2) is present in considerable amounts. Val.Glu.Ala (*Bp* 1) is also present indicating that the Ala.Leu bond is also broken. Another unexpected finding was that the Glu.His bond (positions 4, 5) is broken. This was evident from the presence of Phe.Val.Asp.Glu (*Bp* 8) and of Val.Asp.Glu (*Bp* 7) in the hydrolysate. As already mentioned, the glutamic acid here is probably in the form of an amide so that the actual link split is glutaminylhistidine. Apart from this, it could be said that pepsin splits those bonds adjacent to tyrosine, phenylalanine and leucine residues. It would appear to have a specificity for the more fat-soluble parts of the peptide chain. The finding of a rather wide specificity is in agreement with the results of Desnuelle, Rovery & Bonjour (1950), who studied the *N*-terminal residues liberated on treatment of ovalbumin and horse globin with pepsin.

SUMMARY

1. Fraction *B* of oxidized insulin was subjected to hydrolysis by pepsin, trypsin and chymotrypsin. The resulting peptides were fractionated by paper chromatography and their structure was investigated.

2. It is concluded that the structure of the phenylalanyl chains of insulin is Phe.Val.Asp-(-NH_2).Glu(-NH_2).His.Leu.CyS.Gly.Ser.His.-Leu.Val.Glu.Ala.Leu.Tyr.Leu.Val.CyS.-Gly.Glu.Arg.Gly.Phe.Phe.Tyr.Thr.Pro.Lys.-Ala.

We wish to express our thanks to Mr J. Lightbown for a gift of chymotrypsin, and to the Medical Research Council for a grant to cover the cost of the insulin. One of us (H.T.) wishes to thank the British Council for a scholarship.

REFERENCES

Bergmann, M. & Fruton, J. S. (1937). *J. biol. Chem.* **118**, 405.
Bergmann, M., Fruton, J. S. & Pollok, H. (1939). *J. biol. Chem.* **127**, 643.
Chibnall, A. C. & Rees, M. W. (1951). *Biochem. J.* **48**, xlvii.
Desnuelle, P., Rovery, M. & Bonjour, G. (1950). *Biochim. Biophys. Acta*, **5**, 116.
Fromageot, C., Jutisz, M., Meyer, D. & Pénasse, L. (1950). *Biochim. Biophys. Acta*, **6**, 283.
Fruton, J. S. & Bergmann, M. (1939). *J. biol. Chem.* **127**, 627.
Harington, C. R. & Pitt-Rivers, R. (1944). *Nature, Lond.*, **154**, 301.
Hofmann, K. & Bergmann, M. (1939). *J. biol. Chem.* **130**, 81.
Kunitz, M. & Northrop, J. H. (1936). *J. gen. Physiol.* **19**, 991.
Lens, J. (1949). *Biochim. Biophys. Acta*, **3**, 367.
Martin, A. J. P. (1949). *Biochem. Soc. Symp.* **3**, 4.
Rees, M. W. (1946). *Biochem. J.* **40**, 632.
Sanger, F. (1945). *Biochem. J.* **39**, 507.
Sanger, F. (1949*a*). *Biochem. J.* **44**, 126.
Sanger, F. (1949*b*). *Biochem. J.* **45**, 563.
Sanger, F. & Tuppy, H. (1951). *Biochem. J.* **49**, 463.