

THE AMINO ACID SEQUENCE OF THE LARGE PLAKALBUMIN PEPTIDE AND THE C-TERMINAL SEQUENCE OF OVALBUMIN

By E. O. P. THOMPSON,* R. W. SLEIGH,† and M. B. SMITH†

[Manuscript received November 16, 1970]

Abstract

The amino acid sequence of a 33-residue peptide isolated from plakalbumin by gel filtration in 6M urea at pH 3 and derived from the C-terminal portion of ovalbumin has been determined. Enzyme digestion of the hydrophobic areas by thermolysin, papain, and subtilisin BPN' gave peptides with overlapping sequences. The peptides were fractionated by a combination of paper ionophoresis and chromatography and their sequences determined by the dansyl-Edman technique.

In combination with peptide sequences determined by other workers the sequence of 49 residues at the C-terminus of ovalbumin is proposed.

I. INTRODUCTION

In a previous paper (Sleigh *et al.* 1969) the N-terminal sequence of a large peptide isolated from plakalbumin by dissociation in 6M urea at pH 3 was presented. This peptide was obviously derived from the C-terminal region of ovalbumin since it had the same C-terminal residue as that reported for ovalbumin (Niu and Fraenkel-Conrat 1955). It is characterized by a high proportion of hydrophobic residues. The peptide is fragmented readily by digestion with thermolysin, papain, and subtilisin BPN'. Identification of the peptides obtained has enabled the complete amino acid sequence to be deduced.

II. MATERIALS AND METHODS

(a) Preparation of Plakalbumin Peptide

Plakalbumin was prepared by digestion with subtilisin Carlsberg (subtilopeptidase A) under the conditions described by Smith (1968).

Plakalbumin, 0.8 g, was dissolved in 4 ml water and mixed with 36 ml of a solution 8M in urea and 0.1M in HCl. The solution was applied to a column of Sephadex G75 (400 ml bed volume) equilibrated with 6M urea and 0.05M HCl and the column was eluted with the same solvent at 24 ml/hr. Fractions of 8 ml were collected and analysed for -SH groups (Ellman 1959) with 5,5'-dithiobis(2-nitrobenzoic acid). When stabilization of the thiol groups was desired they were carboxymethylated before gel filtration by the methods previously described (Smith 1968), and the fractions were analysed by the Folin-Lowry method. Urea was removed from the peptide solution by gel filtration on Sephadex G15 (Smith 1968).

(b) Proteolytic Digestion of Plakalbumin Peptide

Digestions with TPCK-trypsin (Worthington) and subtilisin BPN' (Nagarse, Teikoku Chemical Industry Co. Ltd., Osaka, Japan) were carried out in 1% ammonium bicarbonate (pH 8.7) at 37°C for 3-4 hr using 1% enzyme on the weight of substrate.

* School of Biochemistry, University of New South Wales, Kensington, N.S.W. 2033.

† Division of Food Research, CSIRO, Ryde, N.S.W. 2112.

Thermolysin (Chugai Boyaki Co., Osaka, Japan) digestions were done in a 0.2M solution of ammonium acetate at pH 8.5 containing 0.005M CaCl_2 at 37°C for 5–6 hr using 2.5% enzyme on the weight of substrate. The digest was freeze-dried several times after the addition of water to reduce the salt concentration.

The papain (Sigma Chemical Co.) was supplied as a suspension in 0.05M sodium acetate (pH 4.5) containing 27 mg/ml. It was first activated by adding 10 μl to 2.5 ml of a 10% pyridine–water mixture containing 0.005M ethylenediaminetetraacetic acid and 0.005M mercaptoethanol (pH 7.4) and incubating at 37°C for 1 hr. The activated enzyme (0.2 ml) was added to a solution of the peptide (4 mg) in 0.8 ml of 10% pyridine–water containing 0.005M mercaptoethanol, adjusted to pH 7 with 1N acetic acid. After 4 hr digestion the mixture was freeze-dried.

(c) Peptide Fractionation and Amino Acid Analyses

The digests were taken up in 50% pyridine (20–50 $\mu\text{l}/\text{mg}$) and centrifuged to remove insoluble material. The supernatants, 2 mg dry weight for each two-dimensional fractionation, were first subjected to ionophoresis (40 min, 50 V/cm) at pH 6.4 and the peptides in a 1-in. strip were then sewn on a second sheet of Whatman 3MM paper and chromatographed in a mixture of n-butanol–pyridine–acetic acid–water (15:10:3:12 v/v). The neutral peptides were further fractionated by sewing the strip containing them to another sheet of paper, and subjecting them to ionophoresis at pH 1.75. The methods for detection, eluting with 6N HCl, and amino acid analysis were as previously described (Air and Thompson 1969).

(d) Sequence Determination

Peptides for determination of amino acid sequence were stained with 0.02% ninhydrin and eluted from papers after washing thoroughly with acetone. Aqueous pyridine (10–50%) was used for elution by the technique of Sanger and Tuppy (1951). Amino acid sequences were determined by the dansyl–Edman procedure as described by Gray (1967) with the modifications previously given (Air and Thompson 1969, 1971). If the peptide was eluted with 50% pyridine the coupling with phenylisothiocyanate could be performed directly, without drying. This avoids loss of α -amino groups during drying when residual ninhydrin is present.

(e) Cysteic Acid Peptides from Oxidized Plakalbumin Peptide

The peptide in its sulphhydryl form (10 mg) was dissolved in 0.5 ml formic acid and cooled to 0°C. Performic acid reagent (0.5 ml), formed by leaving formic acid (9 vol.) and 30% H_2O_2 (1 vol.) at 25°C for 1 hr, was cooled to 0°C and added. After reaction for 2.5 hr at 0°C the mixture was freeze-dried. Water was added and the solution freeze-dried several times.

Partial acid hydrolysis was carried out with 11N HCl (1 ml) at 37°C for 2.5 days. The mixture was diluted and freeze-dried.

Cysteic acid peptides not containing a basic amino acid residue were separated by passing an aqueous solution through a sulphonated polystyrene ion-exchange resin (Dowex 50-X2, 200 mesh, 10 by 1.0 cm diam.) in the hydrogen form (Kimmel, Thompson, and Smith 1955). The non-absorbed fraction was freeze-dried and fractionated by paper ionophoresis at pH 3.5 followed by chromatography in the solvent described above.

III. RESULTS

(a) Separation of the Large Peptide from Plakalbumin

Figure 1 shows the elution curve of dissociated plakalbumin on G75 Sephadex. The first peak contains the residual protein, which was recovered by diluting the pooled fractions with an equal volume of water and adjusting the pH to 4.9–5.0, when the protein precipitated. The second peak represents the peptide, and is approximately equal in area to the first peak since the peptide and the residual protein each contain two cysteine residues per mole. The peptide was routinely recovered in yields of more than 90% of theoretical yield. When the cysteines were

not alkylated some oxidation occurred on drying but the peptide was readily soluble at pH 8 in the presence of a little dithiothreitol.

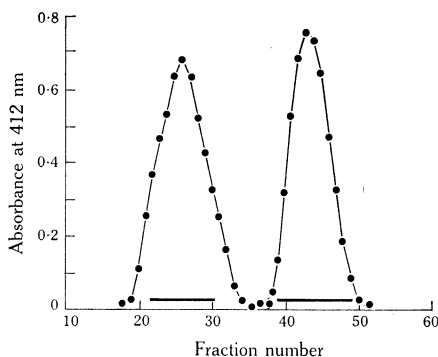


Fig. 1.—Separation of plakalbumin protein and peptide on Sephadex G75 in 6M urea, 0.05M HCl. 0.5 ml of each fraction was mixed with 5 ml of a solution containing 0.5 ionic strength sodium phosphate (pH 7.5), 0.5% sodium dodecyl sulphate, and 0.01% 5,5'-dithiobis(2-nitrobenzoic acid); absorbance measured at 412 nm. Fractions pooled are indicated by bars.

(b) Cysteic Acid Peptides

In the acidic peptide fraction not absorbed on sulphonated polystyrene there were only five peptides, in addition to cysteic acid, in a partial acid hydrolysate of the plakalbumin peptide which had been previously oxidized with performic acid. This fraction was separated by paper ionophoresis at pH 3.5 followed by chromatography in the second (vertical)

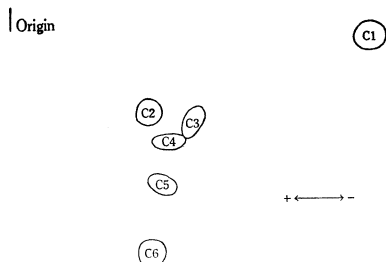


Fig. 2

(C5), (CySO₃H, Val, Ser) (C2), consistent with two sequences Phe-CySO₃H-Ile and CySO₃H-Val-Ser.

The absence of a peptide involving the residue preceding the second cysteic acid probably indicates that the residue is a basic amino acid and the basic peptide was not eluted from the column.

(c) Peptides from Enzyme Digests of the Carboxymethylated Peptide

Peptides obtained by enzymic digestion are designated by a symbol to indicate the enzyme used for the digestion. The symbols used are Tp, trypsin; Th, thermolysin; Pa, papain; S, subtilisin BPN'.

(i) Thermolysin Digestion

The two-dimensional peptide map of a thermolysin digest is shown in Figure 3 and the composition of the peptides in Table 1. The many overlapping sequences indicate that the splitting of a particular bond and generation of a free α -amino group apparently stabilized the adjacent bond, even though it involved a hydrophobic residue as the second residue.

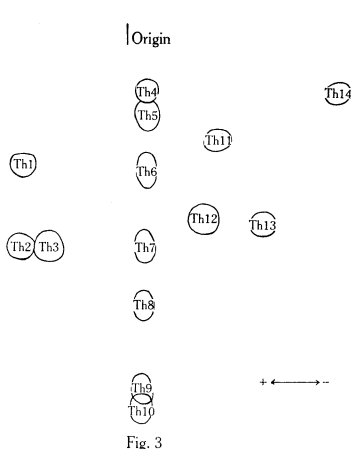


Fig. 3

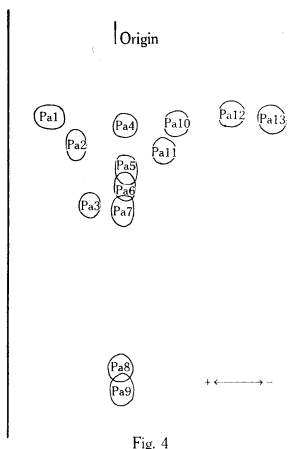


Fig. 4

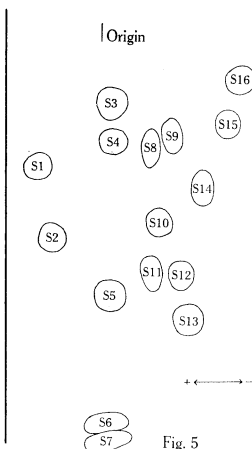


Fig. 5

Figs 3-5.—Peptide map of plakalbumin peptide digested with thermolysin (Fig. 3), papain (Fig. 4), and subtilisin BPN' (Fig. 5). In all cases the fractionation involved paper ionophoresis in pyridine-acetate buffer at pH 6.4 followed by chromatography in the second (vertical) dimension with butanol-pyridine-acetic acid-water (15:10:3:12 v/v). Further resolution of the neutral peptides with similar chromatographic properties was obtained by sewing the neutral strip on another sheet and subjecting them to ionophoresis at pH 1.75.

TABLE 1

AMINO ACID COMPOSITION OF PEPTIDES IN A THERMOLYSIN DIGEST OF PLAKALBUMIN PEPTIDE
Peptides were purified by paper ionophoresis and chromatography as described in the text. Values are given as moles per mole of peptide

Amino Acid	Th1	Th2	Th3	Th4	Th5	Th6	Th7	Th8A*Th8B*	Th9	Th10	Th11	Th12	Th13	Th14
Lys														1.0
His				1.0						0.9	0.9			1.0
Arg						0.9				1.0	1.0	1.0		
SCMC†		0.7	0.7			0.8								
Asp				1.1	0.9						1.0	1.0		
Thr†							0.7							
Ser†	0.8						0.5							
Glu	2.0													
Pro				1.1			1.0				0.7	0.8		
Gly						1.0								
Ala				1.0	1.0			1.1	0.9		1.1	1.1		
Val	1.0						1.1		1.0					
Ile								1.0						1.0
Leu			0.8						0.9	1.0				
Phe		1.0	1.0			1.0		0.5‡	1.0	0.9	0.9	1.8	0.8	

* The A and B designates the neutral peptides resolved by further ionophoretic separation at pH 1.75. A has a lower mobility than B.

† Uncorrected for decomposition. This was considerable in some peptides due to a temperature of hydrolysis of 115-120°C.

‡ The phenylalanine is a contaminating amino acid with similar ionophoretic and chromatographic mobilities. Both DNS-Ile and DNS-Phe were obtained on end-group examination but only one DNS-amino acid at the next two stages of the dansyl-Edman procedure.

Each peptide gave good sequence data and residues assigned in this way are shown in the following tabulation, arrows indicating residues sequenced by the dansyl-Edman procedure:

Th1	Ser-Val-Ser-Glu-Glu	
	→ → →	
Th2	Phe-CMC	
	→ →	
Th3	Leu-Phe-CMC	
	→ → →	
Th4	Ala-Asp-His-Pro	Neutral, therefore aspartyl
	→ → → →	
Th5	Asn-Ala	Neutral, therefore asparaginy
	→ →	
Th6	Phe-Gly-Arg-CMC	
	→ → → →	
Th7	Val-Ser-Pro	
	→ → →	
Th8A	Ile-Ala-Thr	
	→ → →	
Th8B	Ala-Val	
	→ →	
Th9	Leu-Phe	
	→	
Th10	Phe-Leu	
	→	
Th11	Phe-Arg-Ala-Asp-His-Pro	
	→ → → → → →	
Th12	Phe-Arg-Ala-Asp-His-Pro-Phe	
	→ → → → → → →	
Th13	Phe-Arg	
	→ →	
Th14	Ile-Lys-His	
	→ → →	

From the known *N*- and *C*-terminal sequences (Sleigh *et al.* 1969), the amino acid composition, and the sequence data on the peptides the following sequences could be deduced:

Ser-Val-Ser-Glu-Glu-Phe-Arg-Ala-Asp-His-Pro-Phe	Th1 + 12
Leu-Phe-CMC	Th3
Ile-Lys-His	Th14
Asn-Ala-Val	Th5, 8B
Ile-Ala-Thr	Th8A
Phe-Gly-Arg-CMC	Th6
Val-Ser-Pro	Th7

Peptide Th10 indicates that a Phe-Leu sequence must be present and therefore it is assumed that the *C*-terminal phenylalanyl residue shown (Th1 + 12) is not the same residue as in the *N*-terminal phenylalanyl peptide Th6. On this basis the listed sequences account for all the residues in the plakalbumin peptide except for one residue of leucine and one residue of phenylalanine.

A small amount of insoluble peptide material was present. This had phenylalanine as a predominant *N*-terminal residue but other residues were present and it

represented a mixture of partially digested material. The sequences of the cysteic acid peptides suggest that peptide Th3 is linked to Th14 or Th8A, while peptide Th6 must be linked to Th7.

(ii) *Papain Digestion*

The peptide map given by a papain digest is shown in Figure 4. No attempt was made to characterize all peptides as there were some in relatively minor amounts. The amino acid compositions of those which were overlap peptides for the thermolysin peptides (more completely characterized by sequence determination) are shown below:

Pa7	(Lys,CMC,Ile,Phe)	
Pa8	(Gly,Phe ₂)	
Pa9	(Ala,Val,Leu)	
Pa10	(Lys,His,CMC,Asn,Thr,Ala,Ile ₂ ,Phe).	This peptide was slightly basic indicating an asparaginyl residue.
Pa4	(Arg,CMC,Ser,Pro,Val)	
Pa11	(His,Asn,Thr,Ala,Ile).	This peptide was basic indicating an asparaginyl residue.

The points of hydrolysis by the enzyme are shown later (Fig. 6) for the amino acid sequence that was deduced.

As there is only a single residue of lysine in plakalbumin peptide, peptide Pa7 indicates that the thermolysin peptide Th3 must be joined to peptide Th14 to give the sequence Leu-Phe-CMC-Ile-Lys-His which agrees with one of the cysteic acid sequences previously deduced.

Also the missing phenylalanyl residue in the thermolysin peptide sequences had been liberated as free phenylalanine (mixed with Th8A) from a Phe-Phe-Gly peptide sequence present in Pa8 which accounts for the single glycine residue known from amino acid analyses to be present in the large plakalbumin peptide.

The cysteic acid peptide C2 as well as the papain peptide Pa4 confirm that the thermolysin peptides Th6 and Th7 are joined together and incorporation of the papain peptide Pa8 gives the *C*-terminal sequence Phe-Phe-Gly-Arg-CMC-Val-Ser-Pro.

The missing leucyl residue is present in Pa9 indicating it follows the Asn-Ala-Val sequence.

The sequences present in the parent molecule are now reduced to:

Ser-Val-Ser-Glu-Glu-Phe-Arg-Ala-Asp-His-Pro-Phe (*N*-terminal)
 Leu-Phe-CMC-Ile-Lys-His
 Asn-Ala-Val-Leu
 Ile-Ala-Thr
 Phe-Phe-Gly-Arg-CMC-Val-Ser-Pro (*C*-terminal)

and further data was required to determine the relative order of the three non-terminal sequences. This is given by the peptide Pa11 which shows that the Ile-Ala-Thr sequence must be joined to the sequence with a *C*-terminal His. This is supported by the asparaginyl residue of the following Asn-Ala-Val-Leu sequence that is present in peptides Pa10 and Pa11.

(iii) *Subtilisin BPN'*

A peptide map of a subtilisin BPN' digest of the large plakalbumin peptide is shown in Figure 5.

The following peptides were identified by quantitative amino acid analysis and provide supplementary evidence for the overlaps of the peptide sequences deduced from the identified thermolysin peptides.

S1	(CMC,Ser,Pro,Val)	
S2	(Ser ₂ ,Glu ₂ ,Val,Phe)	
S3	(Thr,Asn,Ala)	Neutral, indicating an asparaginyl residue
S4	(Lys,CMC,Ile)	
S5	(His,Asp,Pro,Ala,Leu,Phe)	
S6	(Ala,Val,Leu,Phe)	
S7	(Val,Leu,Phe ₂)	
S8	(His,Thr,Asn,Ala,Ile)	Basic, indicating an asparaginyl residue
S9	(His,Arg,Asp,Pro,Ala,Phe)	
S10	(His,Thr,Ala,Ile)	
S11	(His,Arg,Asp,Pro,Ala,Leu,Phe ₂)	
S12	(His,Ala,Ile)	
S13	(Arg,Gly,Phe ₂)	
S14	(Arg,Gly,Phe)	
S15	(Arg,Ala)	
S16	(Arg,Gly)	

Among these subtilisin peptides are several which indicate that the peptide Ile-Ala-Thr precedes the sequence Asn-Ala-Val-Leu. For example, S3 and S8. There are also several peptides, e.g. S12 and S10, which are only consistent with the sequence Ile-Ala-Thr following the Ile-Lys-His sequence.

This then provides sufficient evidence to piece together the sequence of the original peptide, which is shown in Figure 6 together with the indication of the points of cleavage, and the peptides obtained with the different enzymes.

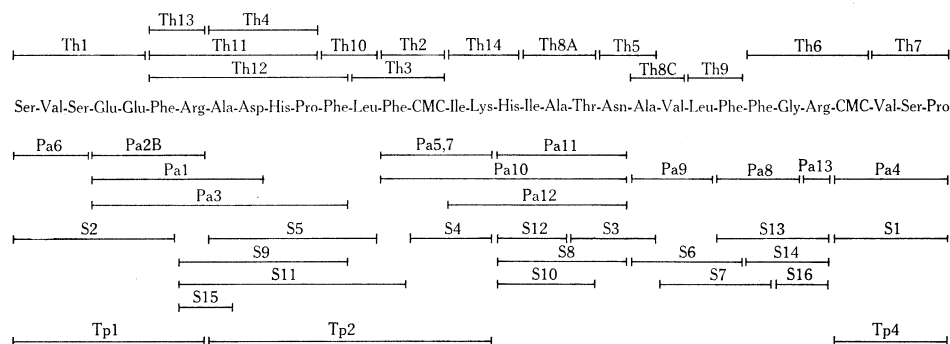


Fig. 6.—Complete amino acid sequence of the large plakalbumin peptide. The points of cleavage by the different proteolytic enzymes are indicated together with the number and designation of each fragment. The symbols used are: Th, thermolysin; Tp, trypsin; Pa, papain; S, subtilisin BPN'. The numbers refer to the position given on peptide maps (Figs. 3-5).

In all the fractionated enzyme digests there were some faint spots which gave only low yields of amino acids on hydrolysis. Some of these were due to free amino acids and some were an oxidized form of the peptides containing *S*-carboxymethyl-cysteine residues.

(iv) *Trypsin Digestion*

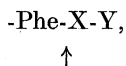
In the early investigation of plakalbumin peptide, digestion with trypsin did not appear very promising, but when the sequence was known it seemed likely that digestion should be possible since there were no Lys-Pro or Arg-Pro sequences. Accordingly a digest was fractionated by paper ionophoresis and chromatography, and among the soluble tryptic peptides which were liberated in reasonable yield were two acidic peptides analysing as (Arg,Ser₂,Glu₂,Val,Phe), peptide Tp1, and (CMC,Ser,Pro,Val), the *C*-terminal peptide Tp4.

Two neutral peptides were obtained, one of which had the composition (Lys,His,CMC,Asp,Pro,Ala,Ile,Leu,Phe₂), corresponding to Tp2. The other peptide contained both lysine and arginine, as well as glutamic acid and serine which are near the *N*-terminus, so that it analysed as the peptide Tp1,2 due to incomplete cleavage of the arginyl bond.

No peptide corresponding to Tp3 was obtained but it is probable that it was not soluble due to its high content of hydrophobic residues (Fig. 6).

IV. DISCUSSION

In investigating the sequence of peptides with hydrophobic regions the use of thermolysin has been recommended (Ambler and Meadway 1968) since it hydrolyses the peptide bonds involving the amino groups of residues with aromatic or hydrophobic side chains. Subtilisin BPN' and papain are useful enzymes to use in conjunction with thermolysin since they tend to split peptide bonds that involve the carboxyl groups of residues with aromatic or hydrophobic side chains, and to a lesser extent those involving lysine and arginine (Smith, Markland, and Glazer 1970). With papain, Berger and Schechter (1970) have defined a specificity site one residue removed from the peptide bond to be split, this site involving hydrophobic side-chains and dominating those involving the basic amino acid residues. This type of specificity, i.e.



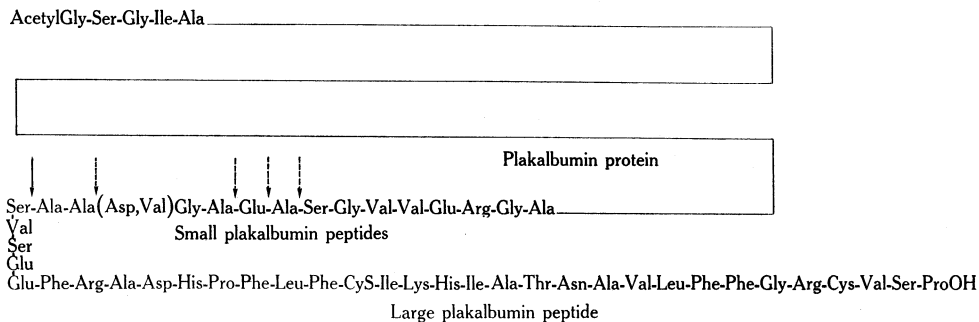
is similar to that described by Gerwin, Stein, and Moore (1966) for the action of the sulphhydryl enzyme streptococcal proteinase. These different specificities are to some extent seen in the bonds hydrolysed in plakalbumin peptide (Fig. 6) but a considerable degree of partial splitting, giving overlapping sequences, is apparent. The extra peptides formed give further information about the sequence but increase the amount of work while decreasing the yields of peptides.

There are no peptides common to the papain, subtilisin BPN', and thermolysin digests, illustrating their suitability, when used in conjunction, in providing overlapping sequences.

The sequence proposed for the large plakalbumin peptide is consistent with the results of other workers. A peptide of similar composition to Tp2 has been described by Fothergill and Fothergill (1970) in a tryptic digest of ovalbumin that had been reduced and carboxymethylated with iodo-[¹⁴C]acetic acid. These authors also obtained the *C*-terminal tryptic peptide Tp4 and presented evidence from double labelling experiments that this carboxymethylcysteine residue was derived from a

disulphide in ovalbumin after reduction. This is not in accord with our results, since the large plakalbumin peptide is obtained without reduction of ovalbumin, and under acidic conditions where disulphide-sulphydryl interchange is unlikely.

Milstein (1968) investigated serine phosphate peptides in digests of ovalbumin and reported a sequence (Milstein 1968, fig. 2) giving a peptide TphcS1 which is clearly the Tp1 sequence of the large plakalbumin peptide. This allows combination of this sequence with that of the present work to give the proposed *C*-terminal 49 residue sequence of ovalbumin (Fig. 7).



V. ACKNOWLEDGMENTS

This work was supported in part by a grant from the Australian Research Grants Committee. The authors are indebted to Miss W. K. Fisher and Mr. R. G. Whittaker for skilled technical assistance.

VI. REFERENCES

- AIR, G. M., and THOMPSON, E. O. P. (1969).—*Aust. J. biol. Sci.* **22**, 1437.
AIR, G. M., and THOMPSON, E. O. P. (1971).—*Aust. J. biol. Sci.* **24**, 75.
AMBLER, R. P., and MEADWAY, R. J. (1968).—*Biochem. J.* **108**, 893.
BERGER, A., and SCHECHTER, I. (1970).—*Proc. R. Soc. Lond. B* **257**, 249.
ELLMAN, G. L. (1959).—*Archs Biochem. Biophys.* **82**, 70.
FOTHERGILL, L. A., and FOTHERGILL, J. E. (1970).—*Biochem. J.* **116**, 555.
GERWIN, B. I., STEIN, W. H., and MOORE, S. (1966).—*J. biol. Chem.* **241**, 3331.
GRAY, W. R. (1967).—In "Methods in Enzymology". (Ed. C. W. H. Hirs.) Vol. XI. p. 469. (Academic Press: New York.)
KIMMEL, J. R., THOMPSON, E. O. P., and SMITH, E. L. (1955).—*J. biol. Chem.* **217**, 151.
MILSTEIN, C. P. (1968).—*Biochem. J.* **110**, 127.
NARITA, K., and ISHII, J. (1962).—*J. Biochem., Tokyo* **52**, 367.
NIU, C.-I., and FRAENKEL-CONRAT, H. (1955).—*J. Am. chem. Soc.* **77**, 5882.
OTTESEN, M. (1958).—*C.r. Trav. Lab. Carlsberg, Sér. Chim.* **30**, 211.
OTTESEN, M., and WOLLENBERGER, A. (1952).—*Nature, Lond.* **170**, 801.
SANGER, F., and TUPPY, H. (1951).—*Biochem. J.* **49**, 463.
SATAKE, K., KURIOKA, S., and NEYASAKI, T. (1961).—*J. Biochem., Tokyo* **50**, 95.
SLEIGH, R., HOSKEN, R., SMITH, M. B., and THOMPSON, E. O. P. (1969).—*Aust. J. biol. Sci.* **22**, 239.
SMITH, E. L., MARKLAND, F. S., and GLAZER, A. N. (1970).—In "Structure-Function Relationships of Proteolytic Enzymes". (Eds. P. Desnuelle, H. Neurath, and M. Ottesen.) p. 160. (Munksgaard: Copenhagen.)
SMITH, M. B. (1968).—*Biochim. biophys. Acta* **154**, 263.