Biochem. J. (1980) **187**, 545–563
Printed in Great Britain

The Primary Structure of the Calcium-Transporting Adenosine Triphosphatase of Rabbit Skeletal Sarcoplasmic Reticulum

SOLUBLE TRYPTIC PEPTIDES FROM THE SUCCINYLATED CARBOXYMETHYLATED PROTEIN

Geoffrey ALLEN*
National Institute for Medical Research, Mill Hill, London NW7 1AA, U.K.

(*Received 21 August 1979*)

The isolation and the determination of the amino-acid sequences of the soluble tryptic peptides, derived by cleavage at arginine residues, of the succinylated (3-carboxypropionylated) *S*-carboxymethylated adenosine triphosphatase protein of rabbit skeletal sarcoplasmic reticulum are described. Treatment of the protein with succinic anhydride gave a derivative that was readily digested with trypsin, yielding two distinct sets of peptides. One set comprises large, relatively hydrophobic, peptides that are highly aggregated (or insoluble) in aqueous solution and that have been identified, by several criteria, with the portion of the protein embedded in the lipid bilayer in the sarcoplasmic reticulum. The second set, which is described here, comprises peptides that have properties typical of those derived from soluble globular proteins and that constitute that part of the protein external to the lipid bilayer. The sequences of these soluble tryptic peptides contain 586 unique residues. Details of the isolation of the peptides and the determination of the sequences are contained in Supplementary Publication SUP 50102 (88 pages) which has been deposited with the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1978) **169**, 5.

The Ca^{2+}-transporting ATPase from rabbit skeletal sarcoplasmic reticulum is an integral membrane protein that has been extensively studied. The pure protein may be isolated in substantial quantities with ease, and this enzyme has been used as a convenient system for the study of the structure and function of integral membrane enzymes. Since work in this field requires a knowledge of the detailed structure of the system, and since the structures of few integral membrane proteins have been well characterized, we decided to determine its primary structure. The amino acid sequences around the cysteine residues have been reported (Allen & Green, 1978) and the sequence around the active-site aspartic acid residue is known (Allen & Green, 1976).

The single polypeptide chain has a molecular weight in the range 100000–119000, as determined by SDS/polyacrylamide-gel electrophoresis and sedimentation-equilibrium studies (MacLennan *et al.*, 1971; Martonosi & Halpin, 1971; Louis & Shooter, 1972; Meissner *et al.*, 1973; Thorley-Lawson & Green, 1973; Rizzolo *et al.*, 1976; LeMaire *et al.*, 1976). The determination of the primary structure thus presents a considerable problem, since the longest complete sequence of a protein yet determined is that of β-galactosidase, with mol.wt. 116248 (Fowler & Zabin, 1977). The peculiar properties of some peptides from the protein, presumably those derived from the interior of the membrane, which constitute about one-third of the protein, magnify the difficulties of sequence determination. However, it was found in the present work that the peptides derived by tryptic digestion of the succinylated (3-carboxypropionylated) protein can be divided into two classes. Peptides in one class are highly aggregated, or insoluble, in aqueous solutions, including concentrated urea or guanidinium chloride solutions, whereas peptides in the second class are, on the whole, soluble in aqueous solutions and may be isolated by standard techniques. In the present paper the isolation and sequence determination of peptides of the latter class are described. For the determination of the

Abbreviations used: SDS, sodium dodecyl sulphate; dansyl, 5-dimethylaminonaphthalene-1-sulphonyl.

* Present address: Department of Immunochemistry, The Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, U.K.

0306-3275/80/060545-19$01.50/1

sequences of the long tryptic peptides, use has been made of information derived from other digests of the protein. The isolation and characterization of peptides from digests by α-chymotrypsin, (Allen, 1980) thermolysin, staphylococcal proteinase, pepsin, trypsin (with non-succinylated protein) (Allen *et al.*, 1980*a*) and by CNBr (Allen *et al.*, 1980*b*), and the determination of three extended lengths of sequence are described (Allen *et al.*, 1980*b*) in the accompanying papers.

Most of the detailed work on the isolation of peptides is described in Supplementary Publication SUP 50102.

Materials and Methods

Most of the methods used in the present work have been described in a previous paper (Allen & Green, 1978).

Trypsin (twice-crystallized; Worthington) was treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one ('TPCK'), as described by Kostka & Carpenter (1964), to inhibit chymotryptic activity. Succinic anhydride was recrystallized from chloroform. T.l.c. on silica gel (Polygram SIL G; Machery–Nagel) and cellulose (Machery–Nagel Polygram CEL 300 and CEL 400) was performed with the solvents BAWP (butan-1-ol/acetic acid/water/pyridine, 15:3:12:10, by vol.) (Waley & Watson, 1953), and BAW (butan-1-ol/acetic acid/water, 3:1:1, by vol.) (Ingram, 1956). The plates were 10cm high. Thin-layer electrophoresis was performed with a Shandon Southern apparatus at 20°C. Thin layers 20cm in length were subjected to 400V for 1h, or longer for peptides with low mobility. The buffers used were pH6.5 (pyridine/acetic acid/acetone/water, 20:1:40:140, by vol.), and pH2 (formic acid/acetic acid/water, 1:4:45, by vol.). For the detection of peptides, side strips were cut and stained with ninhydrin/Cd reagent (Heilmann *et al.*, 1957), or (cellulose plates only) the whole plate was stained with 0.001% fluorescamine in acetone containing 1% pyridine. When appropriate, thin layers were radioautographed and side strips were stained with Cl_2/*o*-tolidine (Reindel & Hoppe, 1954) or Ehrlich's reagent (Smith, 1953) for the detection of radioactively labelled, blocked or tryptophan-containing peptides respectively.

Preparation of ATPase protein for digestion

The ATPase lipoprotein was prepared as previously described (Allen & Green, 1978), by using the method described by MacLennan (1970, 1974) with the inclusion of 1mM-*N*-acetylcysteine in the buffers. The lipoprotein was dissolved in 6M-guanidinium chloride and taurodeoxycholate, reduced with dithiothreitol, and carboxymethylated with iodo[^{14}C]acetate (1500c.p.m./nmol) as described previously (Allen & Green, 1978). Concentrations of protein were determined using an $A_{280}^{1\%}$ of 12 (Hardwicke & Green, 1974; Thorley-Lawson & Green, 1977).

The reduced carboxymethylated protein in 6M-guanidinium chloride solution containing taurodeoxycholate was treated with an equal weight of succinic anhydride added gradually over 30min with rapid stirring. NaOH was added to maintain the pH between 7.5 and 8.5 during the reaction. At the end of the reaction, the succinylated protein solution was made about 1mM in 2-mercaptoethanol, and was dialysed against 1mM-EDTA (twice, with 100 times the volume of protein solution) and against 50mM-Tris, adjusted to pH7.0 with HCl, containing 1mM-EDTA, 1mM-2-mercaptoethanol and 0.1% SDS. The lipoprotein formed a gelatinous precipitate during this procedure. The lipoprotein was dissolved by the addition of SDS (10g per g of protein), with heating at 100°C for 2min and homogenization in a glass homogenizer.

The carboxymethylated succinylated lipoprotein was delipidated by chromatography on Sepharose 6B in 0.1% (w/v) SDS. The detergent was removed by ion-exchange chromatography in 8M-urea (Weber & Kuter, 1971), and urea was removed by dialysis against 50mM-NH_4HCO_3. These procedures were similar to those previously described for the carboxymethylated protein (Allen & Green, 1978). The succinylated protein remained soluble at pH8, but at low pH a precipitate formed. The recovery of protein through these procedures was about 70%. The protein was freeze-dried from 50mM-NH_4HCO_3.

Digestion with trypsin

The freeze-dried protein was suspended in water and the pH was adjusted to 8.3 with NaOH. The protein initially formed a gel, but this slowly dissolved to give a clear solution. In a typical preparation, 160mg of protein was dissolved at 37°C in 15ml of 0.5mM-$CaCl_2$. The solution was stirred under N_2 in a Radiometer pH-stat equipped for the addition of 0.1M-NaOH at pH8.3. 1-Chloro-4-phenyl-3-L-tosylamidobutan-2-one-treated trypsin (2mg) was added as a 10mg/ml solution at pH8.3 in water, and digestion was monitored by the uptake of NaOH. After 64min, a further 0.5mg of trypsin was added. After 100min the digestion was stopped by freeze-drying.

Separation of peptides

The digest was soluble in 0.1M-NH_3 at 50mg/ml. The initial fractionation of peptides was by gel filtration on a column (1.8cm × 140cm) of Sephadex G-50 (Superfine grade) as described in the legend to Fig. 1. Combined fractions from the same

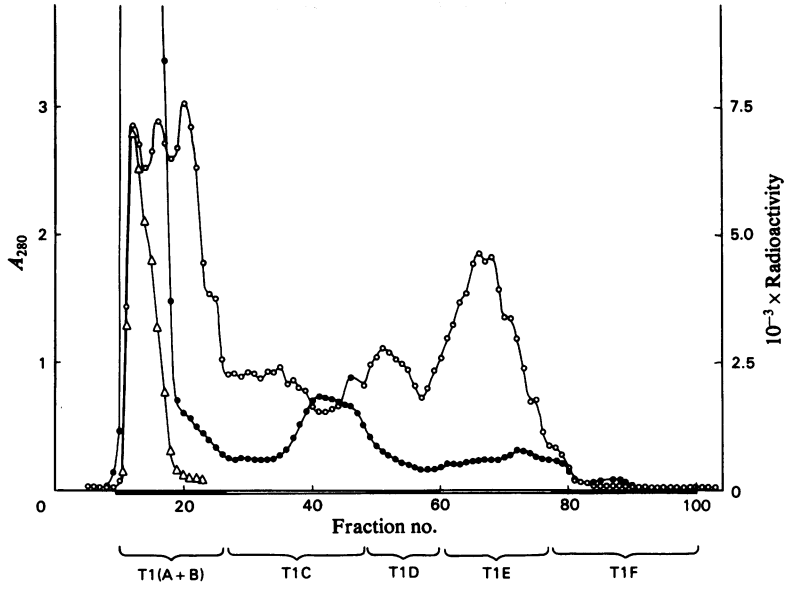

Fig. 1. *Separation of the peptides in the tryptic digest of the ATPase on a column of Sephadex G-50 (Superfine grade)* The column size was 1.8cm × 140cm and the buffer was 50mM-NH_4HCO_3/0.01% thiodiglycol. Fractions of 3.5ml were collected at a flow rate of 15ml/h. Fractions were assayed by absorbance and by liquid-scintillation counting of samples. ●, A_{280}^{1cm}; △, A_{280}^{2mm}; ○, 10^{-3} × ^{14}C radioactivity (c.p.m./0.01ml). Fractions were combined as indicated by the bars and labelled braces at the foot of the Figure, giving the combined fractions T1(A + B), T1C, T1D, T1E and T1F, from which peptides were purified.

parts of the eluates from three chromatographic runs were combined, and further fractionation methods were as described in outline in the Results section below. The details of the separation methods for each peptide are given in Supplementary Publication SUP 50102.

Amino acid analyses

Peptides were hydrolysed in 6M-HCl/0.1% phenol at 110°C for 20h in sealed glass tubes *in vacuo*. Analyses were performed with a BioCal analyser. *S*-[^{14}C]Carboxymethylcysteine was determined from the radioactivity of the peptide by liquid-scintillation counting. Tryptophan was determined from the u.v. spectra of the peptides (Edelhoch, 1967). Some small peptides isolated in low yield were semi-quantitatively analysed by a dansyl method. The hydrolysate (about 1nmol), dried in a glass tube, 3mm internal diameter, was treated with 0.2M-$NaHCO_3$ (5μl), and the solution was dried *in vacuo*. Water (5μl) and a solution of dansyl chloride in acetone (5μl of 2.5mg/ml) were added. The contents of the tube were mixed and the tube was sealed with Parafilm. After 45min at 20°C, the solution was dried *in vacuo*. A 5μl portion of 6M-HCl was added and the solution was dried *in vacuo* over NaOH pellets. The contents of the tube were dissolved in 2μl of 50% (v/v) pyridine and chromatographed on 5cm × 5cm polyamide layers as described by Bruton & Hartley (1970). The relative intensities of the identified dansyl-amino acids were estimated by eye.

Sequence determinations

Amino acid sequences were determined by the 'dansyl–Edman' method, essentially as described by Hartley (1970). Dansyl-amino acids were identified by chromatography on double-sided polyamide layers. *S*-[^{14}C]Carboxymethylcysteine residues were identified by liquid-scintillation counting of the butyl acetate extracts. ε-*N*-3-Carboxypropionyl-lysine residues were identified as two dansyl derivatives, one close to dansylalanine on the polyamide layers (probably α-*N*-dansyl-ε-*N*-succinimidolysine) and the other running with ε-*N*-dansyl-lysine (probably α-*N*-dansyl-lysine) (G. Allen, unpublished work).

In the later stages of the present work, smaller reaction volumes (as described above for the analyses) and micro polyamide layers (5cm × 5cm) were used, as described by Bruton & Hartley (1970). About 10–20nmol of each peptide was subjected to sequence analysis and up to 20 residues were identified.

For the determination of the sequences of longer peptides, and for distinguishing between aspartic acid and asparagine and glutamic acid and glutamine, further digestion with other proteinases and isolation and characterization of the fragments were performed. Thermolysin was most commonly used, and the conditions of digestion were generally 15 min at 55°C in 50mM-NH_4HCO_3/1mM-$CaCl_2$, with a thermolysin/peptide weight ratio of 1:50. Subtilisin or α-chymotrypsin digestions were at 35°C in 50mM-NH_4HCO_3 for 1h with a proteinase/peptide weight ratio of 1:50. The detailed conditions for each digestion are given in the Supplementary Publication SUP 50102.

Identification of acid or amide residues

Small peptides (1–2nmol) were subjected to electrophoresis at pH6.5 on cellulose MN300 thin-layer plates (20cm × 20cm). The dried plates were stained with ninhydrin/Cd reagent. Standard amino acid mixtures were used as markers and the mobilities of peptides relative to that of aspartic acid were measured. An Offord (1966) plot was used to determine the charge on the peptides. When both acid and amide residues were present in a peptide, the electrophoretic mobility was determined at each stage during the Edman degradation.

Nomenclature of peptides

Peptides from the tryptic digest of the succinylated, carboxymethylated protein are labelled T1. Further symbols are added according to the elution positions from chromatographic columns or positions of bands on paper or thin layers after electrophoresis or chromatography. Peptides from other digests of the protein are labelled as follows: α-chymotryptic digest of the carboxymethylated protein, Ch.; thermolysin digest of the carboxymethylated protein, Th; peptic digest of the carboxymethylated succinylated protein, P; CNBr digest of the carboxymethylated succinylated protein, CB; tryptic digest of the carboxymethylated protein, T2, T4; tryptic digest of the protein, T3; staphylococcal protease digest of the carboxymethylated protein, SP. The positions of the peptides within the five sequences discussed in an accompanying paper (Allen *et al.*, 1980*b*) are shown in square brackets.

Abbreviations

The standard three-letter code for amino acid residues is used. Chemically modified residues are marked thus:

$\overset{*}{\text{Cys}}$, *S*-carboxymethylcysteine

$\overset{*}{\text{Lys}}$, ε-(3-carboxypropionyl)lysine (ε-succinyl-lysine)

Sequences of peptides that were clearly determined by the dansyl–Edman method are marked by single-headed arrows beneath the residues, thus:

Val-Glx-Pro-
$\underrightarrow{\quad}\ \underrightarrow{\quad}\ \underrightarrow{\quad}$

Where the results of the dansyl–Edman method were slightly doubtful, the arrow is broken, thus:

Val-Glx-
$\dashrightarrow\ \dashrightarrow$

Groups of residues that have been determined by amino acid analysis but which have not been placed unambiguously in sequence are enclosed in parentheses and separated by commas. A dash in parentheses indicates that the residue was not identified.

Results

In three preparations of reduced *S*-[^{14}C]carboxymethylated, succinylated ATPase protein, the incorporation of [^{14}C]carboxymethyl groups was 23.7, 24.0 and 20.7groups/molecule, assuming a mol.wt. of 115000. The number of cysteine residues in the reduced protein, as estimated with 5,5′-dithiobis-(2-nitrobenzoate), was 25–27 per molecule (Allen & Green, 1978; Thorley-Lawson & Green, 1977), so the extent of labelling of the cysteine residues was about 90%. Upon addition of trypsin, the hydrolysis was initially rapid, and after 1h only a slow uptake of NaOH remained. For one batch (of 160mg) the calculated uptake of NaOH, assuming cleavage at all arginyl peptide bonds, and pK_a 7.8 for the liberated α-amino groups of peptides (Steinhardt & Beychok, 1964) was 0.56ml, and this was achieved after 30min.

The chromatography of one batch of tryptic digest (230mg) on Sephadex G-50 is shown in Fig. 1. Two other batches gave a similar elution profile. The peptides in fraction (A + B) from the three preparations (totalling 560mg; 4.85μmol, assuming a mol.wt. of 115000) were separated on Sephadex G-100, as shown in Fig. 2.

The solution of fraction A peptides was optically almost clear, but further work showed that they were highly aggregated. Attempts to resolve them in aqueous solution, including guanidinium chloride and urea, were unsuccessful, but they have been resolved by chromatography in phenol-based solution and SDS/polyacrylamide-gel electrophoresis (Allen, 1978). The characterization of these fraction A peptides is at present incomplete. The present paper is concerned with the non-aggregated peptides of fractions B–F (Figs. 1 and 2).

Fraction T1B peptides

The estimated molecular weights of these peptides, from their elution positions on gel-filtration columns, was about 5000–8000. The mixture of

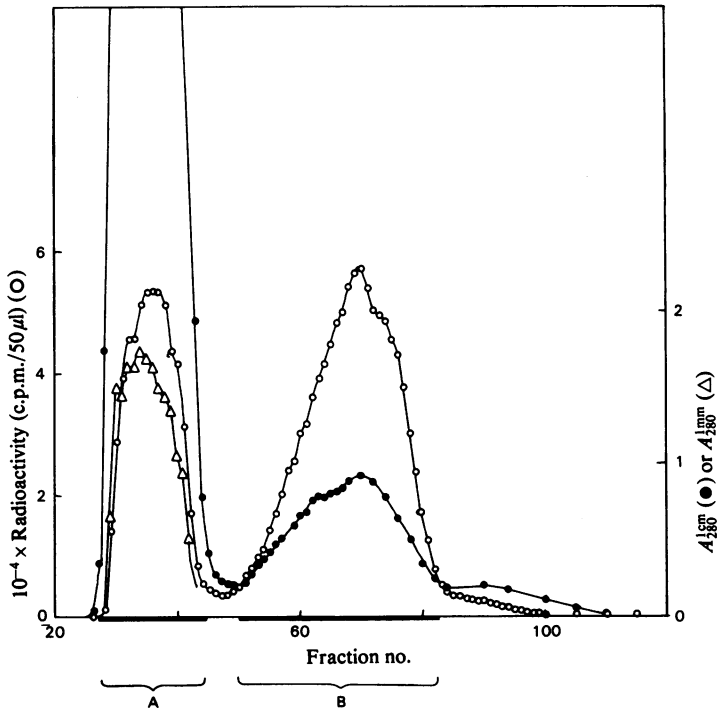

Fig. 2. *Separation of peptides from fraction T1 (A + B) on a column of Sephadex G-100 (Fine grade)*
The column size was 1.4 cm × 131 cm, and the buffer was 50 mM-NH_4HCO_3/0.01% thiodiglycol. Fractions of 2.5 ml were collected at a flow rate of 13 ml/h. ●, A_{280}^{1cm}; △, A_{280}^{1mm}; O, 10^{-4} × ^{14}C radioactivity (c.p.m./0.05 ml). Fractions were combined as indicated by the horizontal bars, to give fractions T1A and T1B.

peptides (combined fraction B of Fig. 2) was fractionated on a column of DEAE-cellulose DE-52, as described in the legend to Fig. 3.

Urea was removed from the combined fractions by gel filtration on a column (2.5 cm × 50 cm) of Sephadex G-10 in 0.1 M-NH_3, 0.01% thiodiglycol (2,2'-thiodiethanol), and the peptide solutions were freeze-dried. As shown in Fig. 3, the resolution of the DEAE-cellulose DE-52 column was poor. However, further resolution of the combined fractions T1B4 to T1B9 by repeated ion-exchange and Sephadex G-75 chromatography, as described in Supplementary Publication SUP 50102, gave three major constituent peptides, and two other peptides in low yield that were shown to be derived by partial non-specific digestion of the major peptides. Each major component was isolated from several of the combined fractions of the DEAE-cellulose DE-52 column eluate (Fig. 3) and subsequent chromatographic steps. The analyses for individual peptides are given in Supplementary Publication SUP 50102. The average analyses of the peptides are given in Table 1.

Peptide T1B-1 [*4(41–113)*]. This peptide was isolated from fractions B5, B6 and B7 of the DEAE-cellulose DE-52 column eluate (Fig. 3), in total yield of 910 nmol (19%). The sequence of this 73-residue peptide was deduced from the sequence analysis of peptides derived from a thermolysin digest of 100 nmol of the peptide, combined with known peptide sequences from other digests of the protein, notably the tryptic and chymotryptic digests, and confirmed by information from digests of the protein by pepsin and CNBr, and digestion of the peptide by α-chymotrypsin. The isolation of thermolysin and chymotrypsin fragments of peptide T1B-1 is described in Supplementary Publication SUP 50102. The thermolysin digest of peptide T1B-4 was also useful, as it was clear that this peptide was the *C*-terminal portion of peptide T1B-1. The total sequence information for peptide T1B-1 is given in Fig. 4. Residue 14 in this sequence was identified as glutamine in a peptide from the peptic digest (P.J2n2), but glutamic acid in a peptide from the tryptic digest (T4-C7a), and its identity needs to be confirmed.

Peptide T1B-2 [*3(10–78)*]. This peptide was isolated from fractions B7 and B9 (Fig. 4). Fraction B8, which probably contained this peptide also, was accidentally lost during the isolation procedure. The

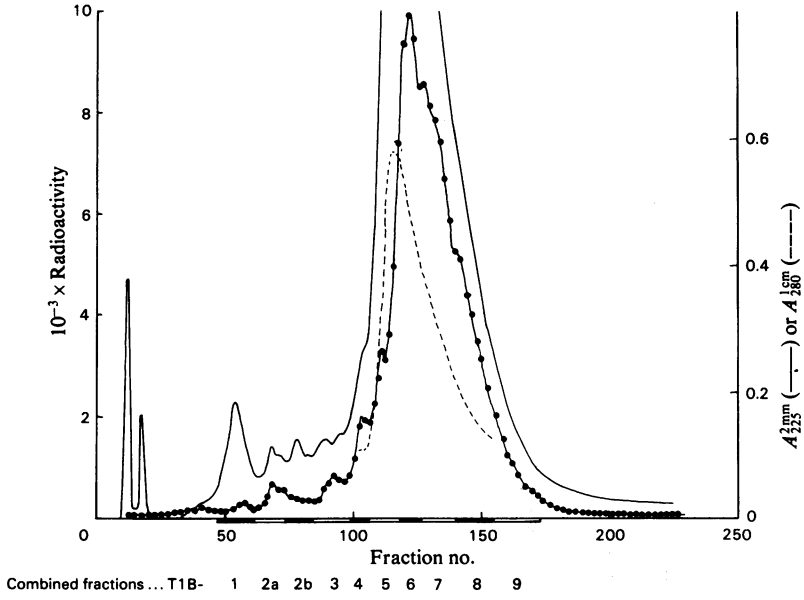

Fig. 3. *Separation of peptides in fraction T1B by ion-exchange chromatography*

The column of DEAE-cellulose DE-52 (1.8 cm × 17 cm) in 8 M-urea/0.01% thiodiglycol was eluted with a gradient prepared from 280 ml of 6 mM-NH_4HCO_3 and 280 ml of 0.6 M-NH_4HCO_3. Fractions of 2.5 ml were collected at a flow rate of 30 ml/h. ——, A^{2mm}_{225}; – – – –, A^{1cm}_{280}; ●, 10^{-3} × ^{14}C radioactivity (c.p.m./0.02 ml). Fractions were combined as indicated by the horizontal bars, and peptides were isolated from the fractions T1B4 to T1B9 by further steps of column chromatography.

Table 1. *Average amino acid analyses for peptides from fraction T1B*

The results are given as mol of amino acid/mol of peptide. All values are derived after hydrolysis for 20 h at 110°C in 6 M-HCl/0.1% phenol *in vacuo*, and are not corrected for hydrolytic losses. [^{14}C]Carboxymethylcysteine was determined from the radioactivity of the peptide (measured by liquid-scintillation counting). Tryptophan was determined from the absorption spectrum (Edelhoch, 1967) of the peptide. The numbers of residues determined from the sequences of the peptides are given in parentheses.

		Content (mol/mol of peptide)				
	Peptide ...	T1B-1	T1B-2	T1B-3	T1B-4	T1B-5
Amino acid	No of analyses of individual peptides ...	6	4	6	2	2
*Cys (from ^{14}C)		0.0 (0)	3.0 (4)	2.0 (2)	0.0 (0)	3.0 (3)
Asp		7.8 (7)	9.0 (9)	7.0 (7)	6.0 (6)	2.0 (2)
Thr		5.3 (5)	8.0 (8)	6.5 (7)	4.5 (4)	5.1 (6)
Ser		5.5 (6)	7.1 (8)	2.8 (3)	2.8 (3)	3.7 (5)
Glu		8.0 (8)	5.3 (5)	7.3 (8)	4.2 (4)	2.2 (2)
Pro		2.0 (2)	2.9 (3)	0.0 (0)	1.2 (1)	1.0 (1)
Gly		6.4 (6)	5.2 (5)	4.1 (4)	5.4 (6)	2.3 (2)
Ala		11.0 (11)	1.5 (1)	3.9 (4)	8.5 (10)	0.9 (0)
Val		6.5 (7)	5.0 (5)	5.6 (6)	4.3 (5)	2.5 (3)
Met		2.1 (3)	1.3 (2)	0.6 (1)	1.6 (2)	1.1 (2)
Ile		4.8 (5)	4.0 (5)	1.1 (1)	2.8 (3)	0.9 (1)
Leu		3.4 (3)	5.0 (5)	5.6 (6)	2.2 (2)	2.9 (3)
Tyr		2.1 (2)	1.0 (1)	1.0 (1)	0.1 (0)	0.0 (0)
Phe		1.1 (1)	3.0 (3)	2.9 (3)	1.0 (1)	0.1 (0)
His		1.2 (1)	0.2 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Lys		5.1 (5)	6.0 (4)	3.0 (3)	3.0 (3)	2.0 (2)
Arg		0.9 (1)	1.0 (1)	0.9 (1)	0.8 (1)	0.1 (0)
Trp (from spectrum)		0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)

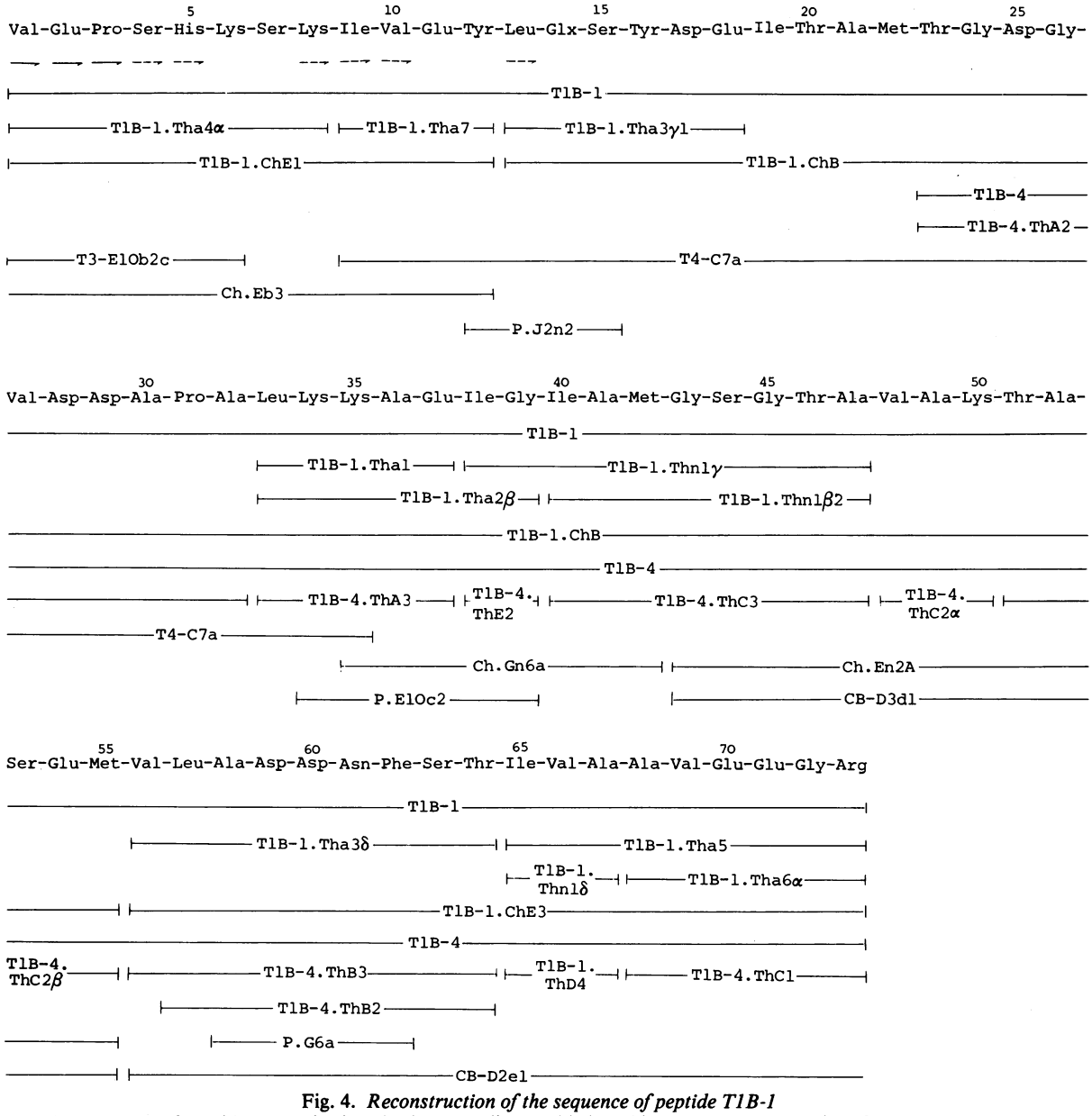

Fig. 4. *Reconstruction of the sequence of peptide T1B-1*

The sequence of peptide T1B-1 is given in the upper line, and below this are shown the peptides from the thermolytic and chymotryptic digests of peptides T1B-1 and T1B-4. The positions of peptides from other digests with trypsin (T3, T4) (Allen *et al.*, 1980*a*), chymotrypsin (Ch.) (Allen, 1980), pepsin (P) (Allen *et al.*, 1980*a*) and CNBr (Allen *et al.*, 1980*b*) (CB), which were also used in the deduction of the sequence of peptide T1B-1, are also included.

yield of T1B-2 was 310nmol (6.4%). The sequence of this 69-residue peptide was determined from the thermolytic fragments, together with known peptide sequences from other digests of the protein. The *N*-terminal 32 residues formed peptide T1B-5, and the *C*-terminal half formed peptide T1C-9, after a non-specific cleavage of a Met–Phe bond. Peptide T1B-2 contains the active-site aspartic acid residue (Allen & Green, 1976). The evidence for the sequence is presented in Fig. 5. A small portion close to the *C*-terminus has not yet been completely sequenced.

Peptide T1B-3 [*3(79–135)*]. This 57-residue peptide was isolated from fractions B5, B6 and B7 (Fig.

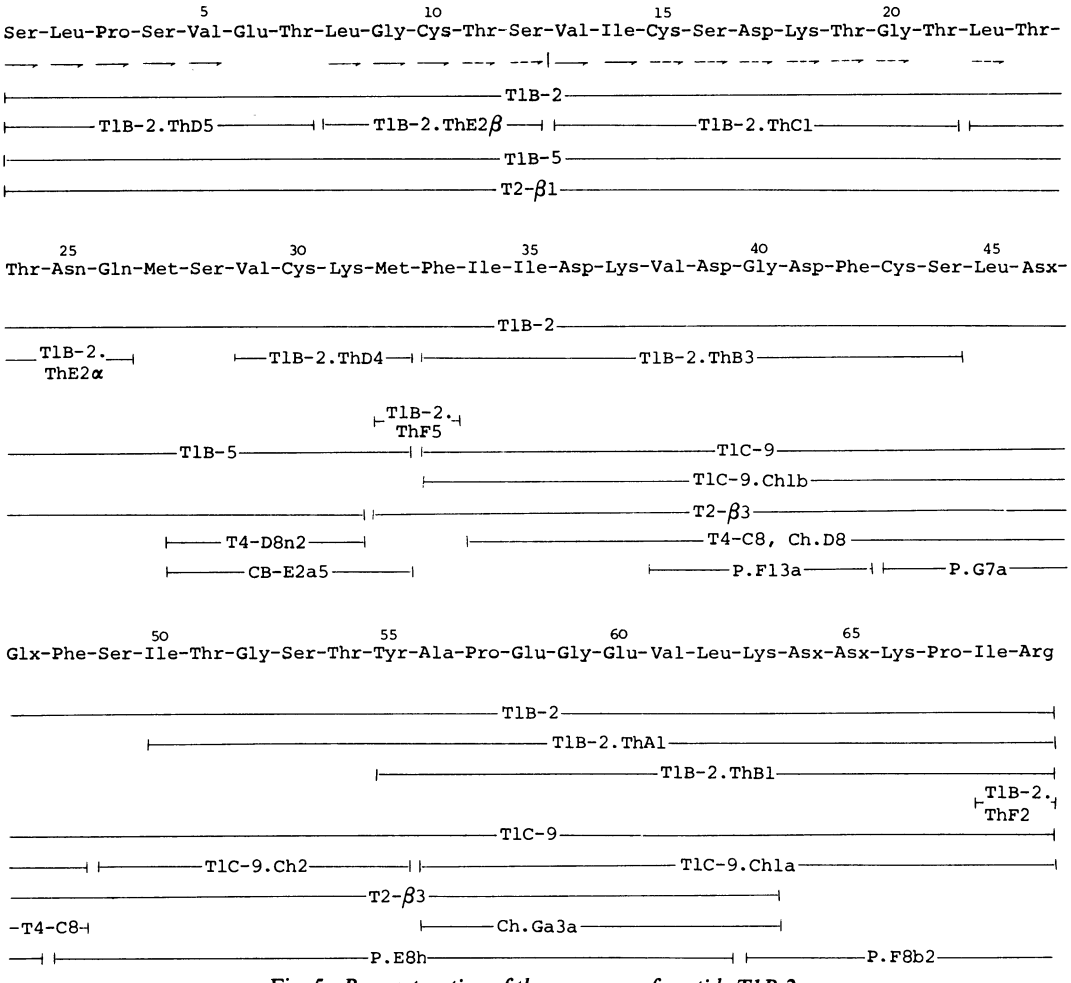

Fig. 5. *Reconstruction of the sequence of peptide T1B-2*

The sequence of peptide T1B-2 is given on the upper line. Below this are shown the positions of the thermolytic fragments of peptide T1B-2, peptide T1B-5, peptide T1C-9, and the chymotryptic fragments of peptide T1C-9. Peptides from other digests of the protein, by trypsin (T2 and T4; Allen & Green, 1978), chymotrypsin (Ch.; Allen, 1980), pepsin (P; Allen *et al.*, 1980*a*) and CNBr (CB; Allen *et al.*, 1980*b*), which were also used in the deduction of the sequence of peptide T1B-2, are also included.

3) in a total yield of 548 nmol (11%). The sequence was determined from the thermolysin peptides, together with peptides from other digests of the protein, as described in Fig. 6.

Peptide T1B-4 [*4(63–113)*]. This peptide was isolated from fractions B4 and B5 in a yield of 182 nmol (3.8%), and the thermolysin peptides were clearly identified with a set of thermolysin peptides from peptide T1B-1. The sequence information is included in Fig. 4. Peptide T1B-4 resulted from cleavage of a Met–Thr bond.

Peptide T1B-5 [*3(10–41)*]. This peptide was isolated from fractions B4 and B5 (Fig. 3) in a yield of 118 nmol (2.4%). The analysis and *N*-terminus identified it as the active-site peptide (Allen & Green, 1976), which is included in peptide T1B-2. Its sequence is given in Fig. 5. The presence of an alanine residue in the analysis is unaccounted for; presumably a contaminating peptide, rich in alanine, is present. The analysis for methionine, which is normally between 60 and 80% in peptides that have been similarly isolated and hydrolysed, indicates that two residues are present. Thus the peptide is a partial cleavage product, with cleavage of the

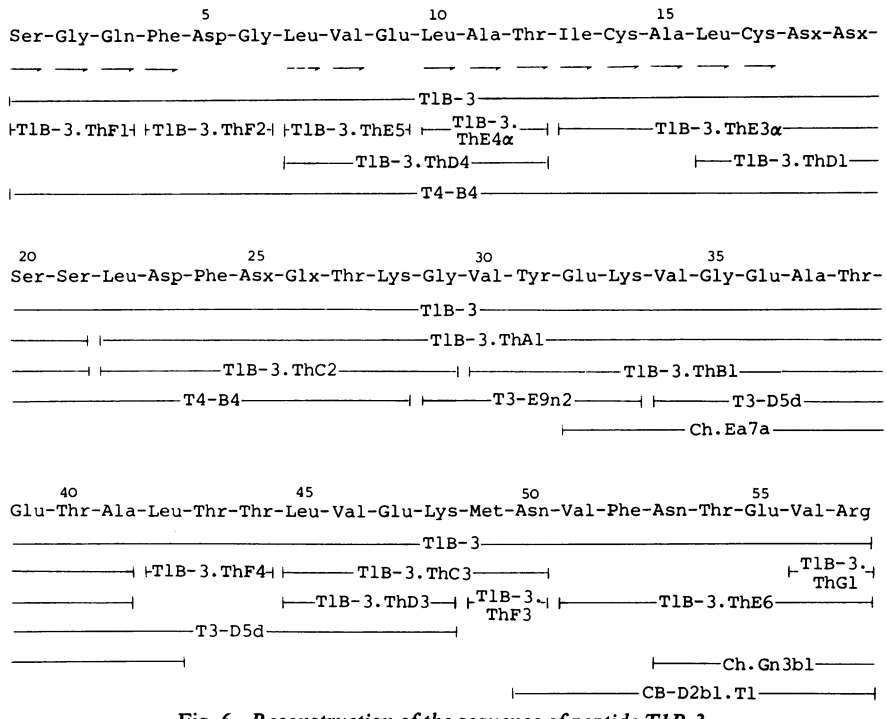

Fig. 6. *Reconstruction of the sequence of peptide T1B-3*

The sequence of peptide T1B-3 is shown on the top line. Below this are shown the positions of the thermolytic fragments of peptide T1B-3, and peptides from other digests of the protein, with trypsin (T4; Allen & Green, 1978), (T3; Allen *et al.*, 1980*a*), chymotrypsin (Ch.; Allen, 1980) and CNBr (CB; Allen *et al.*, 1980*b*), which ere used in the deduction of the sequence of peptide T1B-3.

Met–Phe bond in peptide T1B-2. The complementary peptide was also isolated in low yield (peptide T1C-9).

Fraction T1C peptides (see Fig. 1)

These peptides, with mol.wts. about 2000–5000, as estimated from their elution positions on the Sephadex G-50 column, were fractionated on a column of DEAE-cellulose DE-52 at pH 8, as described in Fig. 7. Further separation and purification steps are described in Supplementary Publication SUP 50102.

Peptide T1C-1 [*3(181–199)*]. This peptide, from fraction C1, had properties (*N*-terminal residue and analysis) identical with those of a peptide isolated from fraction D (peptide T1D-8). The total yield of the peptide was 400nmol (8.2%). The sequence of the *N*-terminal ten residues was obtained directly. The remaining nine residues were deduced from thermolysin fragments, and the sequence was confirmed by the structures of peptides isolated from other digests of the protein. The total sequence information is given in Fig. 8.

Peptide T1C-2 [*2(67–90)*]. This 24-residue peptide was isolated in total yield of 600nmol (12.3%) from fractions C3 and C4 (Fig. 7) after chromatography on DE-52 DEAE-cellulose at pH 5 and Sephadex G-50 chromatography. Of the first 20 residues 18 were identified by direct sequence determination. This sequence was confirmed by the determination of sequences of peptides from tryptic and chymotryptic digests of the carboxymethylated protein. Thermolysin peptides were isolated from peptide T1C-2, but these were not suitable for the completion of the sequence, since the yields were too low. A portion of peptide T1C-2 was digested with subtilisin and several fragments were isolated. The sequence determination of these peptides allowed the completion of the sequence of peptide T1C-2, and the total sequence information is given in Fig. 9.

Peptide T1C-3 [*2(36–56)*]. This 21-residue peptide was isolated from fraction C4 (Fig. 8) after chromatography on Sephadex G-50 and ion-exchange chromatography on DEAE-cellulose DE-52 at pH 5 in urea solution. The yield was 900nmol (18.6%). The sequence was completely determined by the direct dansyl–Edman procedure. Thermolysin fragments were isolated and characterized and some of the acid or amide residues were identified from the electrophoretic mobilities of these frag-

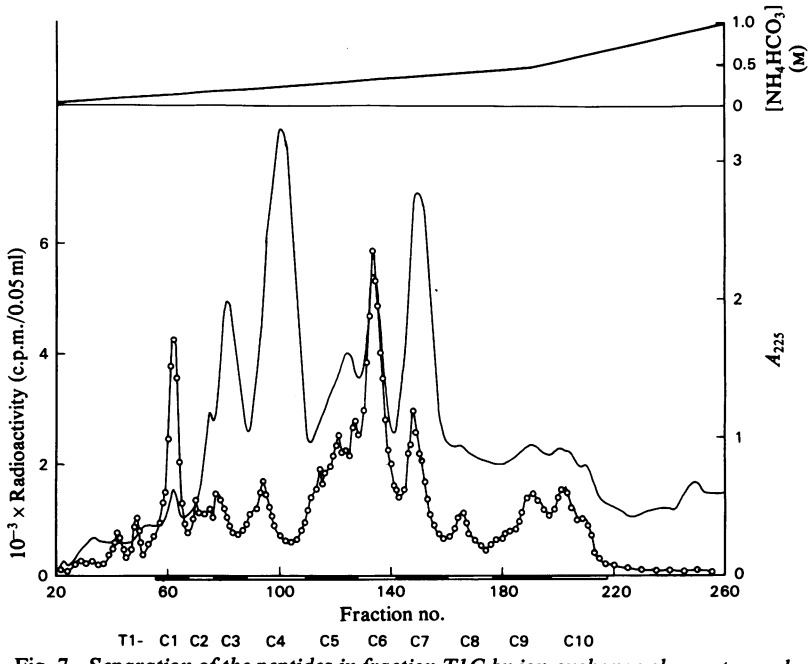

Fig. 7. *Separation of the peptides in fraction T1C by ion-exchange chromatography*

A column (1.8 cm × 23 cm) of DEAE-cellulose DE-52 was used. The sample was applied in 5 ml of 8 M-urea/10 mM-NH_4HCO_3/0.01% thiodiglycol and eluted with 30 ml of 10 mM-NH_4HCO_3/0.01% thiodiglycol, followed by a gradient of NH_4HCO_3 as shown in the upper section of the Figure. Fractions of 5 ml were collected at a flow rate of 40 ml/h. ——, A_{225}^{1cm}; O, 10^{-3} × radioactivity (c.p.m./0.05 ml). Fractions were combined as indicated by the horizontal bars, and peptides were isolated from the combined fractions T1C1 to T1C10 by further steps of column chromatography.

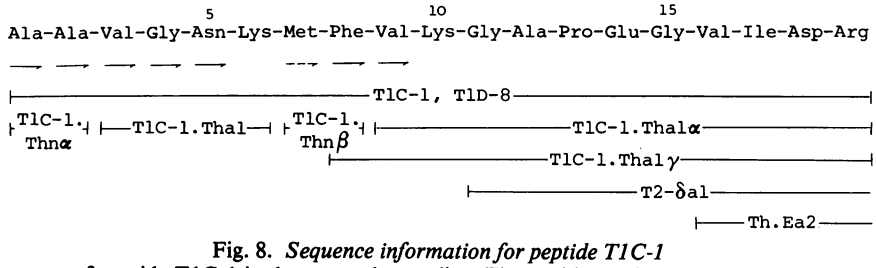

Fig. 8. *Sequence information for peptide T1C-1*

The total sequence of peptide T1C-1 is shown on the top line. The positions of other peptides (thermolysin peptides from peptide T1C-1, tryptic peptide T2-δa1 (Allen & Green, 1978), thermolysin peptide Th.Ea2 (Allen *et al.*, 1980*a*) and CNBr peptide CB-D2c (Allen *et al.*, 1980*b*) whose structures were used to confirm the sequence of peptide T1C-1 are shown beneath.

ments and their degradation products during sequencing. The total sequence information for peptide T1C-3 is given in Fig. 10.

Peptide T1C-4 [*3(210–231)*]. This 22-residue peptide was isolated in a yield of 1080 nmol (22%) from fraction C4 (Fig. 7) by successive chromatography on Sephadex G-50 and on DEAE-cellulose DE-52 at pH 5 in urea. All but three residues were identified by the dansyl–Edman method, and the sequence was completed from thermolytic fragments, as shown in Fig. 11. This was the only soluble tryptic peptide from the succinylated protein found to contain a tryptophan residue. Peptides from other digests of the protein that confirmed the sequence of peptide T1C-4 are also shown in Fig. 11.

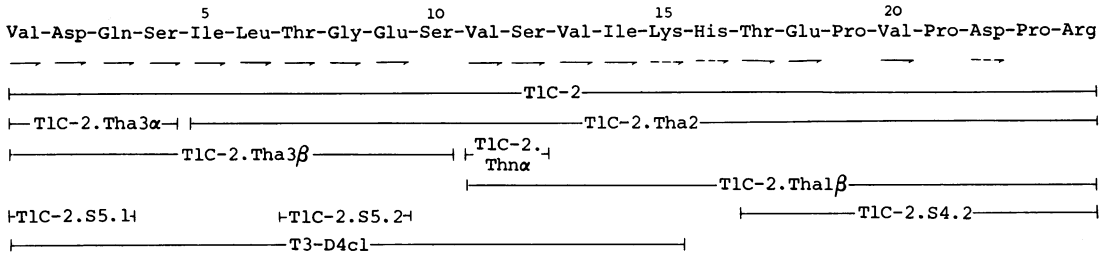

Fig. 9. *Sequence information for peptide T1C-2*
The sequence of peptide T1C-2 is shown on the top line. The positions of thermolytic and subtilisin fragments of the peptide, and of tryptic peptide T3-D4cl (Allen *et al.*, 1980*a*) are shown beneath.

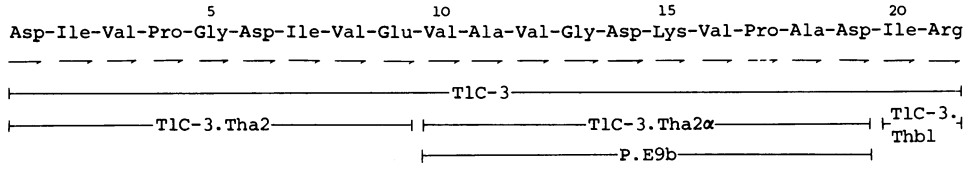

Fig. 10. *Sequence information for peptide T1C-3*
The sequence is shown on the top line. The positions of the thermolysin fragments of peptide T1C-3 are shown beneath.

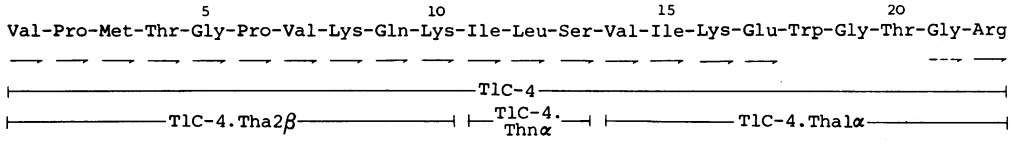

Fig. 11. *Sequence information for peptide T1C-4*
The sequence is shown on the top line. The positions of thermolytic fragments of peptide T1C-4 are shown beneath. The ninth residue was determined as glutamine from the mobility of the peptide T1C-4. Tha2β at pH 6.5; however, this determination is unreliable, as distinction between two and three charges, which is required with this peptide, is inaccurate. This residue has been shown to be glutamic acid in chymotryptic peptide Ch.Fb2a (Allen, 1980) and peptic peptide P.D2a (Allen *et al.*, 1980*a*) by more accurate methods.

Peptide T1C-5 [*3(259–279)*]. This 21-residue peptide was isolated in a yield of 300 nmol (6.2%) from fraction C4 (Fig. 7) by successive chromatography on Sephadex G-50 and on DEAE-cellulose DE-50 at pH 5 in urea solution. The sequence was determined directly to be:

Phe-Met-Glx-Tyr-Glx-Thr-Asx-Leu-Thr-Phe-Val-
Gly-Val-Val-Gly-Met-Leu-Asx-Pro-Pro-Arg

Most of the acid and amide assignments were made from chymotryptic fragments of peptide T1C-11.

Peptide T1C-6 [*3(243–258)*]. This 16-residue peptide was isolated in a yield of 460 nmol (9.5%) from fraction C4 (Fig. 7) by successive chromatography on Sephadex G-50 and on DEAE-cellulose DE-52 in urea solution at pH 5. The sequence was determined by the dansyl–Edman method to be:

Asx-Thr-Pro-Pro-Lys*-Arg-Glx-Glx-Met-Val-Leu-
Asx-Asx-Ser-Ser-Arg

The peptide was a product of incomplete tryptic digestion. The internal arginine residue is between two acidic residues. Identification of the acid or amide residues has been made from other digests of the protein.

Peptide T1C-7 [*1(1–31)*]. This 31-residue peptide was isolated in a yield of 2400nmol (49.5%) from fractions C6, C7 and C8 (Fig. 7). The *N*-terminus was blocked, which suggests that this peptide is the *N*-terminal peptide of the protein, unless the peptide has an *N*-terminal pyrrolidonecarboxylic acid residue, which may have been formed by cyclization of *N*-terminal glutamine during the isolation procedures. The sequence of the peptide was determined from thermolysin-digested fragments, and information from digests of the protein with other proteinases, which was required to complete the ordering of the thermolysin-digested fragments. The *N*-terminal dipeptide, T1C-7.Th5.3, which was blocked and was detected on thin-layer plates by the Cl_2/*o*-tolidine staining method (Reindel & Hoppe, 1954) had the analysis (Met, Glu). This dipeptide (10nmol) was treated with CNBr (2mg) in 50μl of 70% (v/v) formic acid under N_2 for 20h at 45°C. The dried reaction product was dansylated, by using the standard procedure but without hydrolysis, and the products were chromatographed on polyamide layers for the identification of dansyl-amino acids or peptides. The only dansyl derivative observed was dansylglutamic acid. Thus the *N*-terminal sequence is acyl-Met-Glu. A sample of peptide T1C-7.Th2.2, which was also blocked, was similarly treated with CNBr, and the reaction product was dansylated. After hydrolysis, the released *N*-terminal amino acid was shown to be glutamine/glutamic acid. A sample of the peptide T1C-7.Th5.3 was subjected to pH6.5 thin-layer electrophoresis. After staining with Cl_2/*o*-tolidine reagent, a weak spot was visible, in the same position as glutamic acid, which was consistent with the structure acyl-Met-Glu, where the acyl group is small and uncharged. The acyl group has not been identified. Tong (1977) has determined the *N*-terminal sequence of the protein to be acetyl-Met-Glu-Ala. The total sequence information for peptide T1C-7 is given in Fig. 12.

Peptide T1C-8 [*2(3–23)*]. This 21-residue peptide was isolated from fraction C7 (Fig. 7) in a yield of 1860nmol (38%), after chromatography on Sephadex G-50 (Superfine grade). The first 13 residues were sequenced directly. The remaining sequence was determined from thermolysin fragments. Amide assignments were made after determination of the mobilities of thermolysin fragments at pH6.5. The sequence information is presented in Fig. 13.

Peptide T1C-9 [*3(42–78)*]. This peptide was isolated in a yield of 120nmol (2.5%) from fraction C9 (Fig. 7) after chromatography on Sephadex G-50 (Superfine grade). The first ten residues were sequenced, and the results, together with the amino acid analysis, showed that this peptide was the *C*-terminal part of peptide T1B-2. The partial sequence of peptide T1C-9 is therefore included in Fig. 5.

Peptide T1C-10. This peptide was isolated in low yield (100nmol, 2%) from fraction C10 (Fig. 7) after further chromatography on Sephadex G-50 and DEAE-cellulose DE-52. The analysis revealed the absence of arginine. The content of methionine, carboxymethylcysteine, histidine and phenylalanine suggested that the peptide was related to peptides characterized from the tryptic digest of the carboxymethylated protein (peptide B6 in Allen & Green, 1978) and from the chymotryptic digest (Allen, 1980), peptide Ch.B8a. Peptide T1C-10 was digested with α-chymotrypsin and the isolated

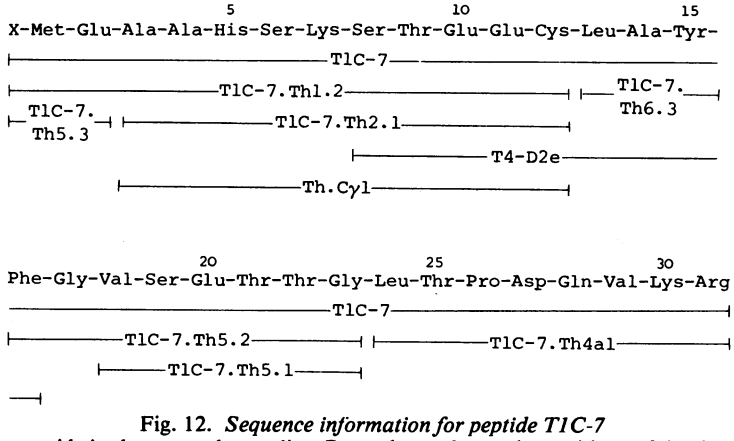

Fig. 12. *Sequence information for peptide T1C-7*

The sequence of the peptide is shown on the top line. Beneath are shown the positions of the thermolytic fragments of peptide T1C-7, peptide T4-D2e (Allen & Green, 1978), and peptide Th.Cγ1 (Allen *et al.*, 1980*a*), which were used to deduce the structure of peptide T1C-7. Residues 10 and 11 were identified as glutamic acid from the mobility of peptide T4-D2e. X is an unidentified acyl group. Tong (1977) determined this to be an acetyl group.

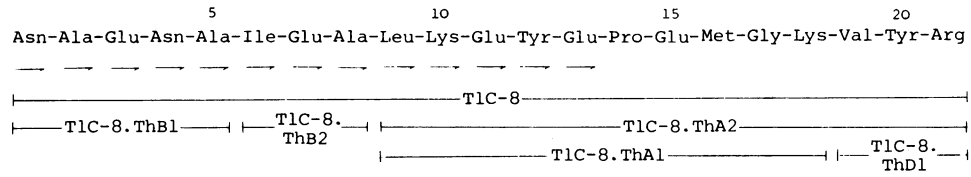

Fig. 13. *Sequence information for peptide T1C-8*
The sequence of the peptide is shown on the top line. The positions of the thermolytic fragments of peptide T1C-8 are shown beneath.

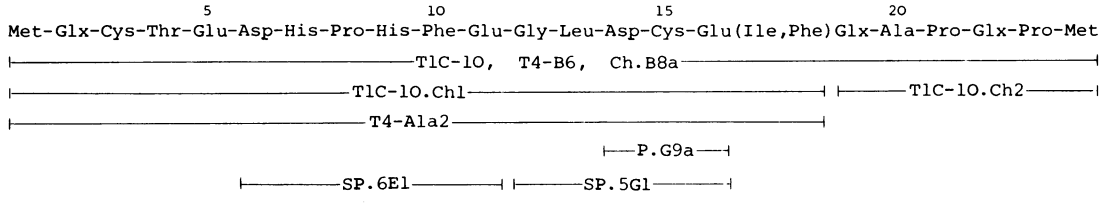

Fig. 14. *Sequence information for peptide T1C-10*
The partial sequence of the peptide is shown on the top line. Beneath is shown the sequence information from chymotryptic fragments of peptide T1C-10 and from other digests of the protein with trypsin (T4-B6 and T4-Ala2; Allen & Green, 1978), chymotrypsin (Ch.B8a; Allen, 1980) and pepsin and staphylococcal proteinase (P.D10a, SP6E1 and SP5G1; Allen *et al.*, 1980*a*). The sequence after residue 14 requires further supporting evidence.

fragments were characterized as described in Supplementary Publication SUP 50102. The sequence information for the peptide is given in Fig. 14. The structure of this part of the primary structure has not been completely determined; peptide T1C-10 was in any case a product of non-specific cleavage, and the yield was low.

Peptide T1C-11 [*3(259–290)*]. This peptide was isolated in low yield (2.4%) from fraction C5. The peptide (50nmol) was digested with α-chymotrypsin (10μg) in 0.1ml of 50mM-NH_4HCO_3 for 12h at 20°C. Peptide fragments were separated as described in Supplementary Publication SUP 50102. These showed that peptide C-11 was an overlap peptide, consisting of peptides T1C-5 and T1D-9. The total sequence of peptide T1C-11 was deduced to be:

Phe-Met-Glu-Tyr-Glu-Thr-Asp-Leu-Thr-Phe-Val-Gly-Val-Val-Gly-Met-Leu-Asx-Pro-Pro-Arg-Lys*-Glx-Val-Met-Gly-Ser-Ile-Gln-Leu-Cys*-Arg

Fraction D peptides

These peptides were estimated by gel filtration to have molecular weights in the range 1000–2000. The mixture was fractionated on a DEAE-cellulose DE-52 column as shown in Fig. 15. Further fractionation steps are described in detail in Supplementary Publication SUP 50102. Column chromatography was generally used, but some paper and thin-layer methods were also employed. The following peptides were characterized.

Peptide T1D-1 [*3(143–151)*]. This peptide was isolated from fraction D2 by paper electrophoresis and chromatography, in a yield of 138nmol (2.8%). The *N*-terminus and amino acid analysis identified this peptide as one previously characterized from the tryptic digest of the carboxymethylated protein (peptide D8n2 in Allen & Green, 1978), with the sequence:

Ala-Asn-Ala-Cys*-Asn-Ser-Val-Ile-Arg

An identical peptide, T1E-2, was isolated from fraction E (below), and the total yield was 548nmol (11.3%).

Peptide T1D-2 [*2(57–66)*]. This peptide was isolated from fraction D2 (Fig. 15) by paper electrophoresis and chromatography in a yield of 360nmol (7.4%). The sequence was determined to be:

Ile-Leu-Ser-Ile-Lys*-Ser-Thr-Thr-Leu-Arg

The two component tryptic peptides, T3-E9b3 and T3-E11b3, from the tryptic digest of the protein have also been characterized (Allen *et al.*, 1980*a*).

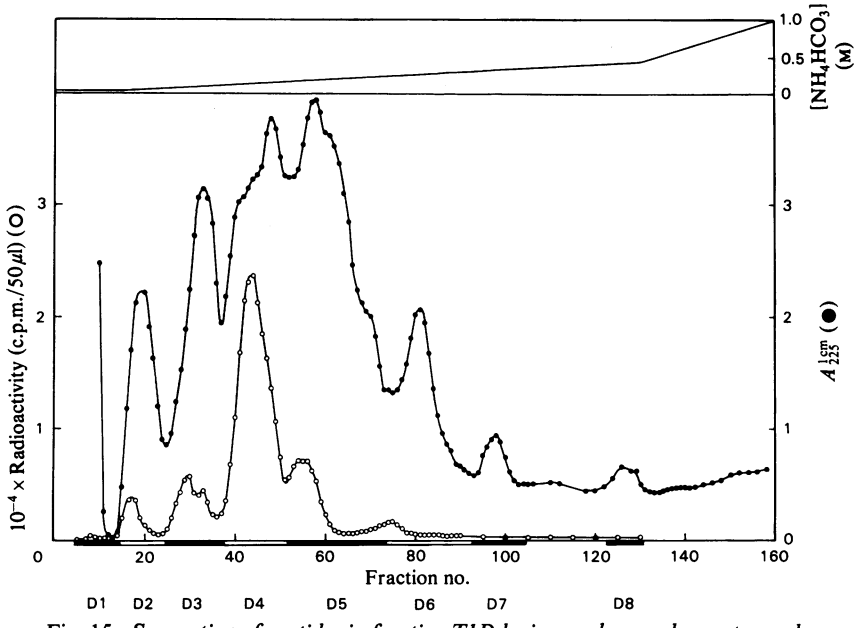

Fig. 15. *Separation of peptides in fraction T1D by ion-exchange chromatography*

The column of DEAE-cellulose DE-52 was 1.2cm × 12cm in size. The buffer was a gradient from 10mM- to 1M-NH_4HCO_3 as shown in the upper section of the Figure. The buffers contained 0.01% thiodiglycol. The sample was applied in 10mM-NH_4HCO_3/8M-urea. Fractions of 2.5ml were collected at a flow rate of 20ml/h. ●, A_{225}^{1cm}; O, 10^{-4} × radioactivity (c.p.m./50μl). Fractions were combined as indicated by the bars at the foot of the figure. Peptides were isolated from the combined fractions (T1D1–T1D8) by a combination of other chromatographic methods.

Peptide T1D-3. This peptide was isolated from fraction D3 (Fig. 15) after paper chromatography, Sephadex G-50 and thin-layer chromatography, in a yield of 90nmol (1.9%). The sequence of the 15-residue peptide was determined to be:

Gly-Phe-Asx-Pro-(?)-Asx-Leu-Asx-Ile-Met-Asx-Arg-Pro-Pro-Arg

The peptide was sequenced three times, on different samples, which were isolated by slightly different procedures. One sample gave aspartic acid/asparagine (Asx) in position 5, but this is inconsistent with the amino acid analysis. A homologous peptide from the peptic digest of the protein gave proline at this position. Further work on a peptide from fraction T1A (Fig. 2) showed that peptide T1D-3 arose from non-specific cleavage at a Leu–Gly bond (G. Allen, unpublished work).

Peptide T1D-4 [*3(136–142)*]. This peptide was isolated from fractions D3 (Fig. 15) and E6 (Fig. 16) in a yield of 1000nmol (20.6%) and its sequence was determined to be:

Asx-Leu-Ser-Lys*-Val-Glx-Arg

The yellow–pink ninhydrin colour showed that the *N*-terminal residue was asparagine. From the electrophoretic mobility, the penultimate residue was therefore glutamic acid.

Peptide T1D-5 [*3(1–9)*]. This peptide was isolated from fraction D3 (Fig. 15) in a yield of 210nmol (4.3%) and its sequence was determined to be:

Met-Ala-Lys*-Lys*-Asx-Ala-Ile-Val-Arg

The fifth residue was shown to be asparagine from the electrophoretic mobility at pH6.5, and from the mobility of the chymotryptic peptide from the carboxymethylated protein, Ch.Jb4, of sequence Ala-Lys-Lys-Asn (Allen, 1980).

Peptide T1D-6 [*3(167–180)*]. This peptide was isolated after several chromatographic procedures from fraction D4 (Fig. 15) in a yield of 750nmol (15.5%). The sequence was determined directly to be:

Lys*-Ser-Met-Ser-Val-Tyr-Cys*-Ser-Pro-Ala-Lys*-Ser-Ser-Arg

Peptide T1D-7 [*3(188–199)*]. This peptide was isolated in low yield (28nmol; 0.6%) from fraction D4 (Fig. 15) after several more fractionation steps. The peptide was not completely pure, but the partial sequence information

Phe-(?)-Lys*-Gly-Ala-Ala-Glx-Gly-Val-Ile-(?)-Arg

showed that the peptide is the same as the *C*-terminal portion of peptide T1C-1, derived from partial non-specific cleavage of a Met–Phe bond.

Peptide T1D-9 [*3(280–290)*]. This peptide was isolated from fraction D4 in a yield of 100nmol (2%), and the sequence was determined directly to be:

Lys*-Glx-Val-Met-Gly-Ser-Ile-Glx-Leu-Cys*-Arg

Acid or amide residues 2 and 8 were investigated from other digests of the protein and are glutamic acid and glutamine respectively (Allen *et al.*, 1980*b*).

Peptide T1D-10 [*3(152–164)*]. This peptide was isolated from fraction D5 (Fig. 15) after three more column runs and one thin-layer dimension, in a yield of 38nmol (0.8%). The low yield was probably due to losses during purification and to blocking of the *N*-terminus by cyclization (cf. peptide T1D-14). The sequence was determined to be:

Glx-Leu-Met-Lys*-Lys*-Glx-Phe-Thr-Leu-Glx-Phe-Ser-Arg

The acid and amide residues were not determined from this peptide, but the chymotryptic digest gave the two peptides Lys-Lys-Glu-Phe (Ch.Jb6) and Thr-Leu-Glu-Phe (Ch.H3a) (Allen, 1980).

Peptide T1D-11 [*4(19–29)*]. This peptide was isolated in a yield of 290nmol (6%) from fraction D5 (Fig. 15) after several further steps of purification. The sequence was determined to be:

Glx-Phe-Asx-Asx-Leu-Pro-Leu-Ala-Glx-Glx-Arg

This peptide is identical with peptide T3-D7a from the tryptic digest of the protein (Allen *et al.*, 1980*a*). The assignment of acid and amide residues was

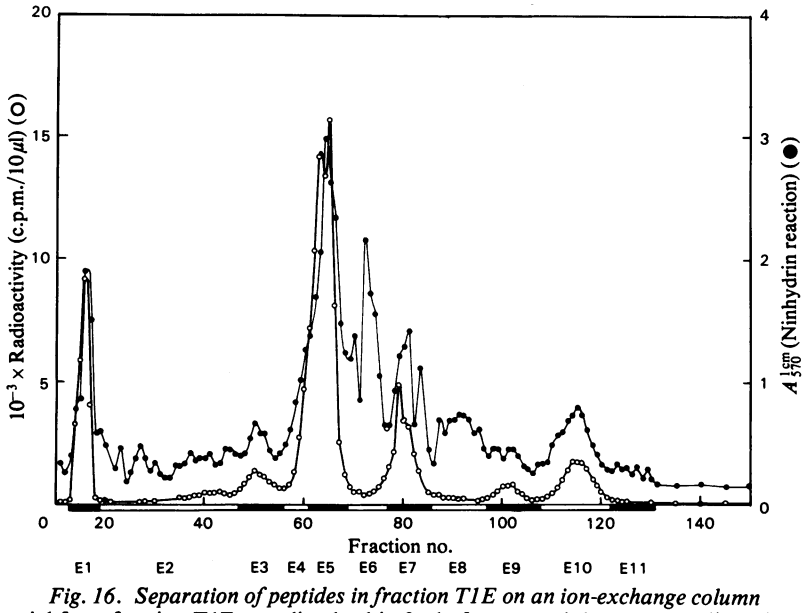

Fig. 16. Separation of peptides in fraction T1E on an ion-exchange column

Freeze-dried material from fraction T1E was dissolved in 2ml of water and the pH was adjusted to 2.0 with HCl. A column (10mm × 400mm) of washed Bio-Rad resin AG50W-X2 (200–400 mesh) was equilibrated in pH 3.1 buffer (prepared from 16.1ml of pyridine, 278ml of acetic acid and 0.1ml of thiodiglycol, made up to 1 litre with water). The sample was applied and the column was eluted with a concave gradient prepared from 100ml of pH 3.1 buffer in the mixing vessel and 200ml of pH 5 buffer (containing 161ml of pyridine, 143ml of acetic acid and 0.1ml of thiodiglycol/litre) in the reservoir vessel of an apparatus consisting of two cylindrical vessels with radii in the ratio 1 : 1.4 connected by a siphon (Schroeder, 1967). Fractions of 1.5ml were collected at a flow rate of 15ml/h at 22°C. A 0.1ml portion of each fraction was subjected to the ninhydrin reaction (Fraenkel-Conrat, 1957) after alkaline hydrolysis (Hirs *et al.*, 1956), and A_{570}^{1cm} (●) was determined. The ^{14}C radioactivity (O) was determined by liquid-scintillation counting of 10μl samples. Fractions were combined for further separation of peptides as indicated by the bars at the foot of the Figure.

made from thermolysin fragments of the latter peptide.

Peptide T1D-12 [*4(1–13)*]. This peptide was isolated in a yield of 145 nmol (3%) from fraction D5 (Fig. 15), and its sequence was determined to be:

Ile-Gly-Ile-Phe-Gly-Glx-Asx-Glx-Glx-Val-Ala-Asx-Arg

This peptide is identical with a peptide, T3-D7b, from the tryptic digest of the protein (Allen *et al.*, 1980*a*), which was determined to have the sequence:

Ile-Gly-Ile-Phe-Gly-Glu-Asn-Glu-Glu-Val-Ala-Asn-Arg

Peptide T1D-13 [*3(249–258)*]. This peptide was isolated from fraction D5 (Fig. 15) in a yield of 48 nmol (1%). Its sequence was determined to be:

Glx-Glx-Met-Val-Leu-Asx-Asx-Ser-Ser-Arg

The low yield precluded determination of the acidic or amide residues. This peptide is the *C*-terminal part of peptide T1C-6.

Peptide T1D-14 [*3(152–164)*]. This peptide was isolated in a yield of 240 nmol (5%) from fraction D5 (Fig. 15). The *N*-terminus was blocked. The amino acid analysis was identical with that of peptide T1D-10. Peptide T1D-14 was digested with thermolysin and the thermolytic fragments were isolated and characterized. These showed that peptide T1D-14 was identical with peptide T1D-10, but for the blocked *N*-terminus, which probably arose by cyclization of *N*-terminal glutamine during isolation of the peptides. The electrophoretic mobilities of the thermolysin fragments also confirmed the assignment of amide and acid residues as given

Table 2. *Sequences of peptides from fraction T1E*
Abbreviation used: n.d., not determined

Peptide	Source (fraction of Fig. 17)	Yield (nmol)	Yield (%)	Sequence
T1E-1 [*3*(243–248)]	E3	156	3.2	Asx-Thr-Pro-Pro-Lys*-Arg
T1E-2 [*3*(143–151)]	E4	410	8.5	Ala(Asx,Ala,Cys*,Asx,Ser,Val,Ile,Arg) (cf. T1D-1)
T1E-3 [*5*(4–8)]	E4	680	14	Asn-Tyr-Leu-Glu-Gly
T1E-4	E4	90	1.9	Val-Val-Ser-Ala-Ala-His
T1E-5 [*3*(236–242)]	E5	131	2.7	Cys*(Leu,Ala,Leu,Ala,Thr,Arg)
T1E-6 [*3*(291–295)]	E5	n.d.	n.d.	Asp-Ala-Gly-Ile-Arg
T1E-7 [*3*(232–235)]	E5	n.d.	n.d.	Asp-Thr-Leu-Arg
T1E-8 [*4*(35–49)]	E5	29	0.6	Ala(Cys*,Cys*,Phe,Ala,Arg)
T1E-9 [*4*(30–33)]	E5	300	6.2	Glx(Ala,Cys*,Arg)
T1E-10 [*2*(24–26)]	E6	670	13.8	Ala-Asp-Arg
T1E-11 [*3*(165–166)]	E6	397	8.2	Asp-Arg
T1E-12 [*2*(27–31)]	E6	n.d.	n.d.	Lys*-Ser-Val-Glx-Arg
T1E-13 [*3*(200–204)]	E7	500	10.3	Cys*(Asx,Tyr,Val,Arg)
T1E-14 [*2*(32–35)]	E7	1200	24.7	Ile-Lys*-Ala-Arg
T1E-15	E7	100	2.1	Ser-Gly-Ile-Gln-Val-Arg
T1E-16 [*3*(205–209)]	E7	415	8.3	Val-Gly-Thr-Thr-Arg
T1E-17	E7	2100	43	Arg
T1E-18	E8	38	0.8	Ala-Leu-Gly-Thr-Arg
T1E-19 [*5*(1–3)]	E9	106	2.2	Ile-Ala-Arg
T1E-20 [*4*(30–34)]	E9	82	1.7	Glu-Ala-Cys*-Arg-Arg
T1E-21 [*4*(34–40)]	E10	100	2.1	Arg-Ala-Cys*(Cys*,Phe,Ala,Arg)
T1E-22 [*4*(14–18)]	E10	480	10	Ala-Tyr-Thr-Gly-Arg

above for peptide T1D-10. The sequence of peptide T1D-14 was therefore (<Glu = pyroglutamic acid):

<Glu-Leu-Met-Lys*-Lys*-Glu-Phe-Thr-Leu-Glu-Phe-Ser-Arg

Although fraction D6 (Fig. 15) contained a considerable amount of peptide material, no pure peptides were isolated from this fraction. Loss of material occurred during paper chromatography.

Fraction T1E peptides

Peptides from fraction E (Fig. 1) were fractionated by cation-exchange chromatography on a column of Bio-Rad AG 50W-X2 resin in a gradient of pyridine/acetic acid buffer, as described in the legend to Fig. 16. Further purification methods were on paper and thin-layer silica gel, and are described in Supplementary Publication SUP 50102. The source of the peptides, the yields, and their sequences are given here in Table 2. Comments on individual peptides are listed below. T1E-1 is the *N*-terminal portion of peptide T1C-6. From the electrophoretic mobility at pH6.5 and the ninhydrin/Cd staining colour, the *N*-terminal residue of this peptide was identified as asparagine. Peptide T1E-2 was identical with peptide T1D-1. Peptide T1E-3 was isolated in a high yield, and the *C*-terminal residue is glycine. This peptide was unlikely to have arisen by non-specific proteinase activity, so it is probably the *C*-terminal peptide of the protein. Confirmatory evidence for this assignment was obtained from the chymotryptic digest, which contained the peptide Leu-Glu-Gly [peptide Ch.Ga2a (Allen, 1980)]. The acid and amide residues were determined from the mobility and the ninhydrin/Cd staining colour of the peptide. Ohnoki (1978) determined the *C*-terminal sequence of the ATPase protein as -Glu-Leu-Gly, from carboxypeptidase Y digests.

Peptide T1E-4 was isolated in low yield, and was probably a product of non-specific digestion of the protein.

Peptide T1E-5 was identical with a known peptide (D8n1a of Allen & Green, 1978) of sequence Cys*-Leu-Ala-Leu-Ala-Thr-Arg.

Peptides T1E-6 and T1E-7 were difficult to resolve, but partial separation was sufficient for the identification of the sequences.

Peptide T1E-8 was identical with a peptide that was previously characterized (peptide E9a1 of Allen & Green, 1978), with the sequence Ala-Cys*-Cys*-Phe-Ala-Arg.

Peptide T1E-9 was identical with a known peptide (E9a2 of Allen & Green, 1978) of sequence Glu-Ala-Cys*-Arg.

Peptide T1E-13 was identical with a known peptide (E11n of Allen & Green, 1978) of sequence Cys*-Asn-Tyr-Val-Arg.

Peptide T1E-20 was derived from the same part of the primary structure as peptide T1E-9, but by cleavage at the second arginine residue rather than between the two arginine residues of an adjacent pair.

Peptide T1E-21 was similar to peptide T1E-8, but originated by cleavage of an arginine–arginine bond, as shown from other digests of the protein (Allen *et al.*, 1980*b*).

Peptides in fraction F

These peptides were resolved by paper electrophoresis and chromatography. Only one peptide was present in significant amounts, and this was identical with peptide T1E-22.

Discussion

The treatment of the *S*-carboxymethylated derivative of the ATPase with succinic anhydride resulted in a product that dissolved in alkaline solution, in contrast with the reduced carboxymethylated delipidated ATPase, which was insoluble after freeze-drying (Allen & Green, 1978). This soluble derivative was completely and rapidly digested by trypsin, as estimated by the uptake of NaOH. This result is in marked contrast with that with the non-succinylated protein, which was only about 50% digested at lysine and arginine residues (Allen & Green, 1978; Allen *et al.*, 1980*a*). The efficiency of digestion was confirmed during analysis of the peptides, since few arginine bonds in the isolated peptides remained uncleaved.

A few points of non-specific cleavage were observed in the present work: after methionine residues in the sequences Thr-Ala-Met-Thr-Gly (peptide T1B-1), Cys-Lys*-Met-Phe-Ile (peptide T1B-2), Asn-Lys*-Met-Phe-Val (peptide T1C-1), and at the *C*-terminus of peptide T1C-10, and also after histidine (peptide T1E-4) and leucine (at the *N*-terminus of peptide T1D-3). The reason for these cleavages, which occurred to only small extents, was not known; contamination of the trypsin by residual chymotryptic activity was negligible, since points of rapid chymotryptic cleavage, observed by using trypsin not treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one (Allen & Green, 1978) were not attacked. The homology around the cleaved peptide bonds in peptides T1B-2 and T1C-1 is quite remarkable, and may indicate the presence of a proteinase of unusual specificity.

No cleavage of peptide bonds *C*-terminal to lysine residues was observed. The succinylation (3-carboxypropionylation) reaction with succinic anhydride was therefore essentially complete.

The treatment of the protein with succinic anhydride had other advantages also. Most of the larger peptides (apart from those probably derived

from the intramembranous part of the protein) were soluble in alkaline solution and could be separated relatively easily on columns of DEAE-cellulose. The peptides could also be sequenced by using the dansyl–Edman procedure, since the protected lysine residues did not react with phenyl isothiocyanate to give relatively hydrophobic peptide derivatives that would be less soluble. The identification of lysine residues was aided by the formation of two dansyl derivatives after acid hydrolysis of α-*N*-dansyl-ε-*N*-3-carboxypropionyl-lysine peptides. One derivative, running close to dansylalanine on the polyamide layers, was probably the cyclic imide derivative ε-*N*-succinimido-α-*N*-dansyl-lysine (G. Allen, unpublished work).

One disadvantage of the use of succinylated protein, however, was that the thermolysin-derived fragments from the longer tryptic peptides were in many cases multiply negatively charged at pH 6.5, and determination of acid and amide groups from the electrophoretic mobilities of these peptides (Offord, 1966) was not always reliable.

The only problem observed during the isolation of the peptides was that no peptides were purified from fraction T1D-6 (Fig. 15). It is probable that peptides in this fraction were lost during paper chromatography. The use of A_{215} to monitor column fractions is sensitive, and it is unlikely that significant amounts of peptides could be lost through lack of detection. The larger peptides were purified in 8 M-urea solution to minimize losses through irreversible adsorption to DEAE-cellulose.

In all, 39 unique arginine residues and a total of 586 amino acid residues were identified in these soluble tryptic peptides. The amino acid analysis of peptide fraction T1A suggests that about eight arginine residues are present in these large aggregated peptides (Allen, 1978). Although there is a small amount of overlap of the same peptide sequences between the two classes of peptide, these results are consistent with the amino acid analysis of the protein, with a total of about 50 arginine residues (Meissner *et al.*, 1973; MacLennan *et al.*, 1971; Martonosi & Halpin, 1971; Thorley-Lawson & Green, 1975, 1977).

The two classes of peptide, those in fraction T1A and those described in the present work, may be identified with the parts of the protein within and outside of the lipid bilayer, respectively, on the basis of several criteria. Theoretical considerations suggest that peptide sequences within a lipid bilayer should be rich in hydrophobic residues, which is the case with peptides in fraction T1A. Comparison with the properties of peptides derived from intramembranous regions of other proteins, such as glycophorin (Segrest *et al.*, 1972), cytochrome b_5 (Ozols & Gerard, 1977) and bacterial rhodopsin (Bridgen & Walker, 1976; Keefer & Bradshaw, 1977; Ovchinnikov *et al.*, 1977) reveals physical-chemical similarities with peptides in fraction T1A. The distribution of cysteine and tryptophan residues between the two classes of peptide is consistent with the observations that most of the cysteine residues are reactive in the native protein, and exposed to the solvent (Thorley-Lawson & Green, 1977; Murphy, 1976), whereas the tryptophan residues are, on average, distant from the cysteine residues but close to the lipid components of the membrane (Hardwicke, 1976). About 17 of the approximately 26 cysteine and half-cysteine residues, but only one of the approximately 18 tryptophan residues, were determined in the sequences of the soluble peptides described here. Digestion of the ATPase protein within the membrane vesicles has confirmed that most of the tryptophan residues are associated with the lipid bilayer (Green *et al.*, 1978) and that the set of soluble peptides described here are derived from the part of the protein external to the lipid bilayer (Green *et al.*, 1980). However, the presence of insoluble 'core' peptides in the extramembranous part of the protein, which could associate with the peptides in fraction A, cannot be excluded.

Discussion of the characteristics of the peptide sequences is deferred to an accompanying paper (Allen *et al.*, 1980*b*), where long stretches of the primary structure of the protein, reconstructed from these tryptic peptides and from peptides from other digests of the protein, are described.

I thank Ms. J. North for technical assistance, Ms. S. Lathwell for operating the amino acid analyser, and Dr. N. M. Green for support.

References

Allen, G. (1978) *FEBS Symp.* **45**, 159–168
Allen, G. (1980) *Biochem. J.* **187**, 565–575
Allen, G. & Green, N. M. (1976) *FEBS Lett.* **63**, 188–192
Allen, G. & Green, N. M. (1978) *Biochem. J.* **173**, 393–402
Allen, G., Bottomley, R. C. & Trinnaman, B. J. (1980*a*) *Biochem. J.* **187**, 577–589
Allen, G., Trinnaman, B. J. & Green, N. M. (1980*b*) *Biochem. J.* **187**, 591–616
Bridgen, J. & Walker, I. D. (1976) *Biochemistry* **15**, 792–798
Bruton, C. J. & Hartley, B. S. (1970) *J. Mol. Biol.* **52**, 165–178
Edelhoch, H. (1967) *Biochemistry* **6**, 1948–1954
Fowler, A. V. & Zabin, I. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1507–1510
Fraenkel-Conrat, H. (1957) *Methods Enzymol.* **4**, 247–269

Green, N. M., Hebdon, G. M. & Thorley-Lawson, D. A. (1978) *FEBS Symp.* **45**, 149–157
Green, N. M., Allen, G. & Hebdon, G. M. (1980) *Ann. N.Y. Acad. Sci.* in the press
Hardwicke, P. M. D. (1976) *Eur. J. Biochem.* **62**, 431–438
Hardwicke, P. M. D. & Green, N. M. (1974) *Eur. J. Biochem.* **42**, 183–193
Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
Heilmann, J., Barollier, J. & Watzke, E. (1957) *Hoppe-Seyler's Z. Physiol. Chem.* **309**, 219–220
Hirs, C. H. W., Moore, S. & Stein, W. H. (1956) *J. Biol. Chem.* **219**, 623–642
Ingram, V. M. (1956) *Nature (London)* **178**, 792–794
Keefer, L. M. & Bradshaw, R. A. (1977) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **36**, 1799–1804
Kostka, V. & Carpenter, F. H. (1964) *J. Biol. Chem.* **239**, 1799–1803
LeMaire, M., Jørgensen, K. E., Røigaard-Petersen, H. & Møller, J. V. (1976) *Biochemistry* **15**, 5805–5812
Louis, C. & Shooter, E. M. (1972) *Arch. Biochem. Biophys.* **153**, 641–655
MacLennan, D. H. (1970) *J. Biol. Chem.* **245**, 4508–4518
MacLennan, D. H. (1974) *Methods Enzymol.* **32**, 291–302
MacLennan, D. H., Seeman, P., Iles, G. H. & Yip, C. C. (1971) *J. Biol. Chem.* **246**, 2702–2710
Martonosi, A. & Halpin, R. A. (1971) *Arch. Biochem. Biophys.* **144**, 66–77
Meissner, G., Conner, G. E. & Fleischer, S. (1973) *Biochim. Biophys. Acta* **298**, 246–269
Murphy, A. J. (1976) *Biochemistry* **15**, 4492–4496
Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V. & Lobanov, N. A. (1977) *FEBS Lett.* **84**, 1–4
Offord, R. E. (1966) *Nature (London)* **211**, 591–593
Ohnoki, S. (1978) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **37**, 1617
Ozols, J. & Gerard, C. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3725–3729
Reindel, F. & Hoppe, W. (1954) *Chem. Ber.* **87**, 1103–1107
Rizzolo, L., LeMaire, M., Reynolds, J. A. & Tanford, C. (1976) *Biochemistry* **15**, 3433–3437
Schroeder, W. A. (1967) *Methods Enzymol.* **11**, 351–361
Segrest, J. P., Jackson, R. L., Marchesi, V. T., Guyer, R. B. & Terry, W. (1972) *Biochem. Biophys. Res. Commun.* **49**, 964–969
Smith, I. (1953) *Nature (London)* **171**, 43
Steinhardt, J. & Beychok, S. (1964) in *The Proteins*, Neurath, H. (ed.), 2nd edn., vol. 2, pp. 139–304, Academic Press, New York
Thorley-Lawson, D. A. & Green, N. M. (1973) *Eur. J. Biochem.* **40**, 403–413
Thorley-Lawson, D. A. & Green, N. M. (1975) *Eur. J. Biochem.* **59**, 193–200
Thorley-Lawson, D. A. & Green, N. M. (1977) *Biochem. J.* **167**, 739–748
Tong, S. W. (1977) *Biochem. Biophys. Res. Commun.* **74**, 1242–1248
Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337
Weber, K. & Kuter, D. J. (1971) *J. Biol. Chem.* **246**, 4504–4509