Biochem. J. (1980) **187**, 591–616
Printed in Great Britain

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

PEPTIDES DERIVED FROM DIGESTION WITH CYANOGEN BROMIDE, AND THE SEQUENCES OF THREE LONG EXTRAMEMBRANOUS SEGMENTS

Geoffrey ALLEN,* Brian J. TRINNAMAN and N. Michael GREEN
National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

(*Received 21 August 1979*)

The isolation and characterization of the soluble peptides from the CNBr digest of the calcium ion-transporting adenosine triphosphatase protein of rabbit skeletal sarcoplasmic reticulum are described. The 562 unique residues of the protein were placed in sequences. The remaining part of the protein (about 500 residues) yielded long hydrophobic sequences that contained all but one of the tryptophan residues of the protein and that were probably derived largely from the intramembranous parts of the protein. Three long stretches of primary structure, constituting half of the protein, have been reconstructed from the information presented here together with the sequences found in peptides from other digests of the protein. The secondary structures of these sequences have been predicted. A model for the primary structure of the protein is presented and the implications discussed. Details of the isolation of peptides are contained in Supplementary Publication, SUP 50105 (29 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 calcium-ion transporting adenosine triphosphatase protein (Ca^{2+}-ATPase) of rabbit skeletal sarcoplasmic reticulum is an integral membrane protein consisting of a single polypeptide of mol.wt. about 115000, as estimated by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis (Thorley-Lawson & Green, 1973). Electron-microscopic studies show that about two-thirds of the protein extends as a globular projection on the cytoplasmic face of the sarcoplasmic-reticulum vesicles, and the remaining one-third is embedded within the lipid bilayer (Hardwicke & Green, 1974; Stewart & MacLennan, 1974).

This overall structure is reflected in the properties of peptides isolated from several digests of the protein, such that between one-half and two-thirds of the protein is released as soluble, relatively hydrophilic, peptides, whereas the remaining portion is resistant to digestion and gives rise to large, aggregated, relatively hydrophobic peptides (Allen, 1980*a*,*b*; Allen & Green, 1978; Allen *et al.*, 1980).

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

Most of the cysteine residues are reactive in the native protein (Thorley-Lawson & Green, 1977; Murphy, 1976), and are released in the soluble peptides, whereas almost all of the tryptophan residues are close to, or within, the lipid bilayer, and are not released in the soluble peptides.

This clear differentiation between the two types of peptide has prompted us to report the results of structure determination of the parts of the protein external to the lipid bilayer before completion of the total primary structure determination, since the latter must await the determination of the structures of intractable hydrophobic peptides.

The amino acid analysis of the protein (MacLennan *et al.*, 1971; Martonosi & Halpin, 1971; Meissner *et al.*, 1973; Thorley-Lawson & Green, 1975) shows that about 35 methionine residues are present in the polypeptide chain, so about this number of peptides was expected to be present in the CNBr digest. The strategy adopted was to use these CNBr peptides to overlap the tryptic peptides from the succinylated (3-carboxypropionylated) protein, the sequences of most of which have been determined (Allen, 1980*a*). From the evidence provided in

0306-3275/80/060591-26$01.50/1

the present paper, together with information derived from other digests of the protein (Allen, 1980*b*; Allen *et al.*, 1980), the sequences of three large stretches of polypeptide that together constitute more than half of the total sequence of the protein have been reconstructed. These sequences are of interest for the interpretation of the reactivity of the cysteine residues, the properties of the active site and the overall architecture of the protein.

Details of the purification of peptides are contained in Supplementary Publication SUP 50105.

Materials and Methods

Reduced *S*-[^{14}C]carboxymethylated succinylated delipidated Ca^{2+}-ATPase protein was prepared as described previously (Allen, 1980*a*), except that the specific radioactivity of the iodo[^{14}C]acetate used was diluted to 416c.p.m./nmol. Trypsin was a thrice-crystallized product of Worthington Biochemical Co. (Freehold, NJ, U.S.A.); it was treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one ('TPCK') to inhibit chymotryptic activity (Kostka & Carpenter, 1964). Chymotrypsin was also from Worthington. Staphylococcal proteinase was obtained from Miles Laboratories (Stoke Poges, Slough, Berks., U.K.).

CNBr was a product of Aldrich (Gillingham, Dorset, U.K.) and tryptamine hydrochloride was from Eastman (Kodak, Kirkby, Liverpool, U.K.). Sephadex materials were from Pharmacia (G.B.) (London W5, U.K.). DEAE-cellulose was DE-52 from Whatman Biochemicals (Maidstone, Kent, U.K.). Plastic-backed thin-layer cellulose and silica-gel plates were from Macherey–Nagel (Düren, Germany). Other reagents were the purest grades available from BDH (Poole, Dorset, U.K.).

Digestion with CNBr

The protein derivative (250mg, 2.2μmol) was suspended in a mixture of 70% (v/v) formic acid (25ml), tryptamine hydrochloride (50mg) and CNBr (250mg). The mixture was incubated at 20°C under N_2 for 20h, with occasional shaking. Urea (5g) and butan-1-ol (1ml) were added, and the incubation was continued for a further 20h. The viscous solution was evaporated to dryness by rotary evaporation at 30°C. Water (50ml) was added, and the evaporation was repeated. The addition of water and evaporation were repeated three times to remove excess CNBr and formic acid.

Isolation of peptides

The digest was dissolved in water (10ml) with the addition of NH_3 to bring the pH to 9.0. The peptides were separated on a column of Sephadex G-50 as described in Fig. 1 (below). Five sets of fractions, CB-(A + B), CB-C, CB-D, CB-E and CB-F, were combined and freeze-dried. Fraction CB-(A + B) material was further fractionated on a column of Sephadex G-100 (Fig. 2 below). Two sets of fractions were combined and freeze-dried (CB-A and CB-B). Fraction CB-F material was chromatographed on column (2.5cm × 30cm) of Sephadex G-10 in 0.1M-NH_3, but only tryptamine was identified in the eluate.

Peptides were isolated from fractions CB-B, CB-C, CB-D and CB-E by suitable combinations of ion-exchange chromatography, high-resolution gel filtration, paper electrophoresis and t.l.c. Peptides were detected by u.v. absorption (215nm) and liquid-scintillation counting in column fractions, or ninhydrin/Cd staining of side-strips and radioautography on paper or thin layers. The methods have been described (Allen & Green, 1978; Allen, 1980*a*). T.l.c. was generally performed in butan-1-ol/acetic acid/water/pyridine (15:3:12:10, v/v) (solvent system 1). The details of the purification procedures for individual peptides are given in the Results section.

Peptide-structure determination

Amino acid analyses were performed as described by Allen (1980*a*). Peptide sequences were determined by the dansyl–Edman method as described by Allen (1980*a*).

Large CNBr peptides were digested with trypsin, and fragments were isolated by gel filtration followed by t.l.c. The *N*-terminal and amino acid analyses of these fragments were usually sufficient for the identification of them with tryptic peptides of known structure (Allen, 1980*a*).

Reconstruction of extended lengths of sequence

Information from digests with trypsin (Allen & Green, 1978; Allen, 1980*a*; Allen *et al.*, 1980), chymotrypsin (Allen, 1980*b*), pepsin, staphylococcal proteinase and thermolysin (Allen *et al.*, 1980), together with that from the CNBr digest was used to determine the structures of extended lengths of sequence. The basic source of structural information was the sequences of the tryptic peptides derived by cleavage at arginine residues of the succinylated protein (Allen, 1980*a*). These were ordered by the use of overlap peptides from other digests. In order to minimize the possibility of incorrect alignment of peptides, which could arise by chance repetition of sequences around arginine residues, at least three residues on each side of arginine residue were required in the overlap peptides.

Nomenclature

The standard three-letter code for amino acid residues is used, with the addition of the following: Asx and Glx, where discrimination between Asp and

Asn, and Glu and Gln, respectively, has not been made. Chemically modified residues are indicated by asterisks:

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

$\overset{*}{\text{Lys}}$, ε-*N*-3-carboxypropionyl-lysine

$\overset{*}{\text{Met}}$, homoserine or homoserine lactone

This system avoids the use of more than three letters for modified residues, while allowing clearer comparison of peptide sequences. The positions of the peptides within sequences *1–5* (shown in Figs. 3–6) are given in square brackets. Peptides from the CNBr digest are labelled CB-; nomenclature for peptides from other digests is given in Fig. 3 (below).

Results

The freeze-dried reduced carboxymethylated succinylated ATPase protein was not soluble in 70% (v/v) formic acid, but during the digestion with CNBr, with occasional shaking and homogenizing, a viscous, almost clear, solution was formed. The dried digest dissolved completely in dilute NH_3, and the initial fractionation on the Sephadex G-50 column is shown in Fig. 1. The material eluted close to the void volume (fraction A + B) was resolved on a column of Sephadex G-100 into fraction A (at the void volume) and fraction B, as shown in Fig. 2. The peptides in fraction A were aggregated in aqueous solutions. This material was shown by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis to be a complex mixture, including partial degradation products. Some peptides have been isolated from this fraction with phenol-containing solvents (G. Allen & B. J. Trinnaman, unpublished work), but these peptides require much more investigation. Peptides were purified from fractions CB-B–CB-E as described in Supplementary Publication SUP 50105 and as outlined here.

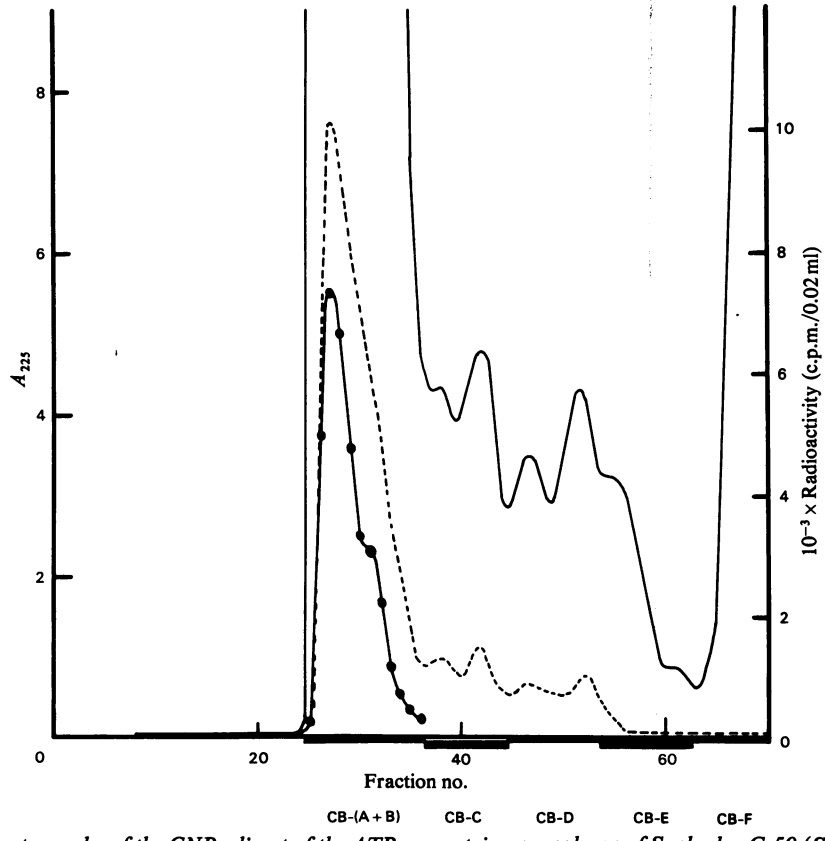

Fig. 1. *Chromatography of the CNBr digest of the ATPase protein on a column of Sephadex G-50 (Superfine grade)*
The column size was 1.7 cm × 141 cm and the buffer was 50 mM-NH_4HCO_3/0.01% thiodiglycol. The flow rate was 20 ml/h and fractions were 5 ml in size. ——, $A_{225}^{1\,cm}$; ●, $A_{225}^{0.5\,mm}$; – – – –, ^{14}C radioactivity (c.p.m./0.02 ml). Fractions were combined as shown by the horizontal bars. The strong absorption in Fraction CB-F was due to tryptamine.

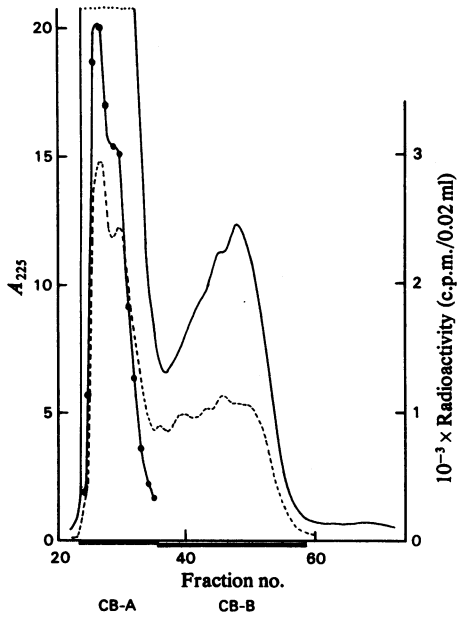

Fig. 2. *Chromatography of material in fraction CB-(A + B) on a column of Sephadex G-100 (Fine grade)*
The column size was 1.5 cm × 140 cm and the buffer was 50 mM-NH_4CO_3/0.01% thiodiglycol. The flow rate was 3 ml/h and fractions were 2.5 ml in size. ——, A_{225}^{10mm}; ●, A_{225}^{2mm}; – – – –, ^{14}C radioactivity (c.p.m./0.02 ml). Fractions were combined as indicated by the horizontal bars.

Peptides in fraction CB-B were resolved on a column of Sephadex G-75 (Superfine grade) into five subfractions, from each of which peptides were purified by ion-exchange chromatography on columns of DEAE-cellulose in gradients of NH_4HCO_3 in 8 M-urea solution. The analyses of the peptides are given in Table 1.

Peptides CB-B2c and CB-B3c [3 (42–154)]

Both of these contained 113 residues. No *N*-terminus was detected. Peptide CB-B3c was digested with trypsin, and the tryptic peptides were resolved on a column of Sephadex G-50 (Superfine grade), followed by t.l.c. (silica gel; solvent system 1) for the smaller peptides. The tryptic peptides were identified from their *N*-terminal and amino acid analyses (Table 1).

Peptide CB-B3a [2 (19–128)]

This was isolated in relatively low yield. The *N*-terminal residue and amino acid analysis showed that this peptide was a partial cleavage product, overlapping peptides CB-B4a and CB-C3b.

Peptides CB-B4a and CB-B5b1 [2 (19–99)]

These were identical, with 81 residues. The partial sequence was:

Gly-(-)-Val-(-)-(-)-Ala-(-)-(-)-(-)-Ser-Val-Glx-

The peptide was digested with trypsin, and peptides were separated on a column of Sephadex G-50, followed by t.l.c. The properties of the purified tryptic peptides are given in Table 1. The partial sequence of peptide CB-B4a. T2 was:

Ala-Val-Asx-Glx-Asx-Lys* (Lys*,Asx,Met*)

Peptides CB-B4c and CB-B5c2 [4 (before 1–82)]

These were almost identical, with about 93 residues. The *N*-terminal sequence gave poor results:

Ile-(-)-Gly-(-)-(-)-(-)-Gly-Thr-Ala-

Peptide CB-B5c2 was digested with trypsin and the tryptic peptides were separated on a column of Sephadex G-50, followed by t.l.c. The properties of the purified peptides are given in Table 1. The partial sequence of peptide CB-B5c2.T2 was:

Val-Glx-Pro-Ser-

which, together with its analysis, identified it with the *N*-terminal part of peptide T1B-1. In addition to the purified peptides, two peptides in fraction CB-B5c2.T3, with *N*-terminal Ile-Gly and Glx, had properties similar to those of peptides T1D-12 and T1D-11 respectively, and a mixture of peptides with *N*-terminal Ala, Arg and Glx in fraction CB-B5c2.T4 could correspond to peptides T1E-22, T1E-21 and T1E-9 (Allen, 1980*a*).

Peptides CB-B5a and CB-B5b2 [3 (188–251)]

These were identical, with 64 residues. Peptide CB-B5a was digested with trypsin, and the tryptic peptides were isolated on a column of Sephadex G-50 followed by t.l.c. The properties of the tryptic peptides are given in Table 1.

Peptides in fraction C

These were resolved by chromatography on a column (1 cm × 142 cm) of Sephadex G-75 (Superfine grade) in 50 mM-NH_4HCO_3/0.01% thiodiglycol into three subfractions: CB-C1, CB-C2 and CB-C3. Only fraction CB-C3 yielded almost pure peptides after further separation. Fraction CB-C3 was resolved into six peptide fractions by ion-exchange chromatography on a column of DEAE-cellulose in a gradient of NH_4HCO_3. The peptides were not

completely pure, as shown by the analyses, and the sequence information obtained was of poor quality:

Peptide CB-C3b [*2* (*100–about 128*)]. This had the partial structure:

Leu-Phe-(-)-Gly-(-)-Asx-Ile-Ala-Ala-Gly-Lys*-Ala-
Leu-Gly-Ile-Val-Ala-Thr-(-)-Gly-

Peptide CB-C3c. This had the partial structure:

Leu-Phe-Ala-Gly-(-)-Asx-Ile-Ala-Ala-Gly-Lys*-Ala-
Leu-Gly-Ile

Peptide CB-C3e. This had the partial structure:

Val-Leu-(-)-Ile-Ser-Leu-Pro-(-)-Ile-Gly-(-)-Asx-Glx-
Ile-Leu-Lys*-Phe-(-)-Ala-(-)-(-)-Tyr-Glx-.

Peptide CB-C3f. This had the partial structure:

Val-Leu-(-)-Ile-Ser-Leu-Pro-(-)-Ile-Gly-Leu-Asx-
Glx-Ile-Leu-Lys*-Phe-(-)-Ala-(-)-(-)-Tyr-Glx

Peptide CB-C3g. This gave sequence results corresponding to a mixture of two peptides:

Val Leu Lys* Ile Ser Leu Pro Val Ile Gly Leu
Glx Cys* Thr Glx Asx His Pro His Phe Glx Gly

Asx Glx Ile Leu Lys* Phe (-) Ala (-) (-) Tyr Glx
Leu Asx (-) (-) (-) Glx Ala Pro Glx Pro (-)

The upper sequence corresponds to that found in peptides CB-C3e and CB-C3f; the lower sequence

Table 1. *Amino acid analyses of CNBr-derived peptides from the Ca^{2+}-ATPase and of proteinase-derived fragments of these peptides*

The results are expressed as the number of residues per molecule of peptide. The results are not corrected for hydrolytic losses. Carboxymethylcysteine was determined from the radioactivity of the peptide. Methionine was converted into a mixture of homoserine, eluted at the position of glutamic acid, and homoserine lactone, eluted between histidine and arginine, and could not be reliably quantified. Detection of approximately one residue per peptide is indicated by +. ε-*N*-(3-Carboxypropionyl)lysine was about 70% hydrolysed and about 30% converted into a compound eluted at the position of valine after 20h hydrolysis. Values in parentheses are the numbers of each residue in the peptide calculated from the deduced sequence. A dash indicates less than 0.2 residue per peptide. Some peptides were analysed only semiquantitatively by the dansyl technique (Allen *et al.*, 1980); these results are expressed thus: +, present; −, absent; ±, faint detection. Tryptophan was determined from the absorption spectra of the peptides, and this amino acid was present only in peptides Cb-B5a and Cb-B5b2, and the tryptic fragment CB-B5a.T1. Identification of the tryptic fragments with known tryptic peptides (Allen, 1980*a*), is indicated below the analyses. Abbreviation used: n.d., not detected.

Peptide	CB-B2c	CB-B3a	CB-B3c	CB-B3c.T3	CB-B3C.T4	CB-B3c.T6b
Yield (nmol)	52	60	75	28	11	6
N-Terminus	n.d.	Gly	n.d.	Ser	Phe	Asx
Amino acid	Content (residues/molecule)					
Cys*	4.4 (4)	—	3.0 (4)	2.0 (2)	1.0 (1)	—
Asp	15.0 (16)	12.6 (12)	16.0 (16)	6.6 (7)	6.0 (6)	1.2 (1)
Thr	9.7 (9)	7.0 (8)	9.3 (9)	6.0 (7)	2.0 (2)	—
Ser	8.9 (8)	7.1 (8)	7.5 (8)	3.1 (3)	2.6 (3)	1.1 (1)
Glu	11.2 (13)	8.4 (8)	12.1 (13)	6.9 (8)	3.0 (3)	0.9 (1)
Pro	2.7 (2)	5.4 (5)	2.0 (2)	—	1.7 (2)	—
Gly	7.7 (7)	8.2 (8)	8.1 (7)	4.3 (4)	3.3 (3)	0.5 (0)
Ala	8.0 (7)	8.4 (9)	8.1 (7)	3.7 (4)	1.4 (1)	0.6 (0)
Val	10.2 (10)	13.1 (14)	9.6 (10)	5.2 (6)	2.3 (2)	1.0 (1)
Met*	+ (2)	+ (2)	+ (2)	+ (1)	—	—
Ile	5.0 (6)	8.6 (11)	5.5 (6)	1.3 (1)	2.7 (4)	—
Leu	10.1 (10)	5.3 (5)	9.9 (10)	6.4 (6)	1.9 (2)	1.1 (1)
Tyr	2.0 (2)	1.2 (1)	2.2 (2)	1.1 (1)	0.8 (1)	—
Phe	5.3 (6)	1.1 (1)	5.8 (6)	2.7 (3)	1.9 (3)	—
His	—	1.1 (1)	—	—	—	—
Lys	9.7 (7)	9.4 (9)	8.7 (7)	2.7 (3)	2.6 (3)	0.9 (1)
Arg	4.9 (4)	7.9 (8)	4.0 (4)	0.5 (1)	0.9 (1)	1.0 (1)
Total	(113)	(110)	(113)	(57)	(37)	(7)
Tryptic peptide					T1C-9	T1D-4

Table 1—(*continued*)

Peptide ...	CB-B4a	CB-B4a.T1a	CB-B4a.T1b	CB-B4a.T2	CB-B4a.T3b	CB-B4a.T3c
Yield (nmol) ...	250	40	50	70	5	25
N-Terminus ...	Gly	Val	Asx	Ala	Ala	Ile
			Content (residues/molecule)			
Amino acid						
*Cys	–	–	–	–	–	–
Asp	10.2 (10)	2.2 (2)	4.0 (4)	2.9 (3)	2.8 (3)	–
Thr	4.2 (4)	2.0 (2)	–	0.3 (0)	–	2.0 (2)
Ser	5.3 (6)	2.7 (3)	–	0.3 (0)	–	1.9 (2)
Glu	6.0 (6)	3.0 (3)	1.1 (1)	1.5 (1)	1.2 (1)	–
Pro	5.1 (5)	3.0 (3)	2.0 (2)	–	–	–
Gly	4.6 (4)	1.2 (1)	2.2 (2)	0.2 (0)	–	–
Ala	5.1 (5)	–	2.1 (2)	1.0 (1)	1.0 (1)	–
Val	11.7 (12)	4.0 (4)	4.7 (5)	1.2 (1)	1.0 (1)	–
*Met	+ (1)	–	–	+ (1)	+ (1)	–
Ile	6.8 (8)	1.8 (2)	2.5 (3)	–	–	2.1 (2)
Leu	3.1 (3)	1.0 (1)	–	–	–	2.2 (2)
Tyr	1.0 (1)	–	–	–	–	–
Phe	0.2 (0)	–	–	–	–	–
His	0.9 (1)	1.1 (1)	–	–	–	–
Lys	7.0 (8)	1.0 (1)	1.0 (1)	1.6 (2)	+ (2)	0.9 (1)
Arg	6.8 (7)	1.0 (1)	1.0 (1)	–	–	1.1 (1)
Total ...	(81)	(24)	(21)	(9)	(9)	(10)
Tryptic peptide ...		T1C-2	T1C-3			T1D-2

Peptide ...	CB-B4a.T4a	CB-B4a.T4b	CB-B4a.T5a	CB-B4a.T5c	CB-B4c	CB-B5a
Yield (nmol) ...	25	30	20	18	80	140
N-Terminus ...	*Lys	Ile	Ala	Gly	Ile	Phe
			Content (residues/molecule)			
Amino acid						
*Cys	–	–	–	–	2.3	1.5 (2)
Asp	–	–	1.1 (1)	–	10.0	4.3 (4)
Thr	–	–	–	–	4.7	7.0 (7)
Ser	1.0 (1)	–	0.3 (0)	0.2 (0)	3.0	0.8 (1)
Glu	1.0 (1)	–	–	–	11.0	5.3 (5)
Pro	–	–	–	–	3.1	4.6 (5)
Gly	–	–	0.3 (0)	1.1 (1)	7.8	5.5 (6)
Ala	–	1.2 (1)	0.9 (1)	–	12.2	3.4 (3)
Val	1.1 (1)	–	–	1.1 (1)	4.3	6.7 (7)
*Met	–	–	–	–	–	+ (2)
Ile	–	1.0 (1)	–	–	7.9	2.1 (3)
Leu	–	–	–	–	4.4	3.9 (4)
Tyr	–	–	–	1.0 (1)	2.6	0.9 (1)
Phe	–	–	–	–	2.9	1.0 (1)
His	–	–	–	–	1.0	–
Lys	0.9 (1)	0.9 (1)	–	1.0 (1)	5.1	3.5 (5)
Arg	1.1 (1)	1.0 (1)	0.9 (1)	1.1 (1)	7.3	6.2 (7)
						Trp + (1)
Total ...	(5)	(4)	(3)	(5)		(64)
Tryptic peptide ...	T1E-12	T1E-14	T1E-10			

Table 1—(*continued*)

Peptide ...	CB-B5a.T1	CB-B5a.T2a	CB-B5a.T2b	CB-B5a.T2c	CB-B5a.T4	CB-B5a.T5n
Yield (nmol) ...	72	13	18	24	47	n.d.
N-Terminus ...	Val	Asx	Asx	Phe	Cys*	Asx
	Content (residues/molecule)					
Amino acid						
Cys*	–	–	–	1.3 (1)	1.1 (1)	–
Asp	0.5 (0)	1.1 (1)	1.1 (1)	1.8 (2)	–	+ (1)
Thr	2.4 (2)	0.8 (1)	0.8 (1)	–	1.3 (1)	+ (1)
Ser	1.0 (1)	0.4 (0)	0.5 (0)	–	–	–
Glu	2.0 (2)	2.1 (2)	2.1 (2)	1.2 (1)	–	–
Pro	1.9 (2)	2.3 (2)	2.2 (2)	1.4 (1)	–	–
Gly	3.0 (3)	0.3 (0)	–	2.4 (2)	–	–
Ala	–	–	–	1.2 (1)	2.1 (2)	–
Val	3.1 (3)	–	–	2.6 (3)	–	–
Met*	+ (1)	+ (1)	+ (1)	–	–	–
Ile	1.5 (2)	–	–	0.8 (1)	–	–
Leu	1.1 (1)	–	–	–	1.8 (2)	+ (1)
Tyr	–	–	–	0.9 (1)	–	–
Phe	–	–	–	0.9 (1)	–	–
His	–	–	–	–	–	–
Lys	2.4 (3)	0.6 (1)	0.8 (1)	1.0 (1)	–	–
Arg	1.4 (1)	0.7 (1)	0.7 (1)	1.7 (2)	0.9 (1)	+ (1)
	Trp + (1)					
Total ...	(22)	(9)	(9)	(17)	(7)	(4)
Tryptic peptide ...	T1C-4				T1E-5	T1E-7

Peptide ...	CB-B5a.T5b	CB-B5b1	CB-B5b2	CB-B5c2	CB-B5c2.T2
Yield (nmol) ...	n.d.	140	280	165	6
N-Terminus ...	Val	Gly	Phe	Ile	Val
	Content (residues/molecule)				
Amino acid					
Cys*	–	–	1.0 (2)	1.5 (3)	–
Asp	–	10.1 (10)	5.0 (4)	10.3 (10)	4.4 (4)
Thr	+ (2)	4.5 (4)	7.3 (7)	4.9 (5)	2.1 (2)
Ser	–	5.6 (6)	1.0 (1)	3.6 (4)	2.7 (3)
Glu	–	6.2 (6)	4.9 (5)	12.1 (12)	5.1 (5)
Pro	–	5.9 (5)	4.8 (5)	3.0 (3)	2.0 (2)
Gly	+ (1)	4.8 (4)	5.3 (6)	5.8 (6)	3.5 (3)
Ala	–	5.2 (5)	3.1 (3)	9.6 (10)	5.2 (5)
Val	+ (1)	12.1 (12)	6.4 (7)	4.4 (5)	3.0 (3)
Met*	–	+ (1)	+ (2)	+ (1)	+ (2)
Ile	–	7.0 (8)	2.2 (3)	6.5 (7)	3.5 (4)
Leu	–	3.3 (3)	3.9 (4)	4.7 (5)	2.0 (2)
Tyr	–	1.1 (1)	0.9 (1)	2.9 (3)	1.8 (2)
Phe	–	–	1.0 (1)	3.3 (3)	–
His	–	1.0 (1)	–	1.1 (1)	1.1 (1)
Lys	–	7.3 (8)	4.3 (5)	4.1 (4)	3.4 (4)
Arg	+ (1)	6.9 (7)	6.2 (7)	7.8 (8)	–
			Trp (1)		
Total ...	(5)	(81)	(64)	(90)	(42)
Tryptic peptide ...	T1E-16				

Table 1—(*continued*)

Peptide	CB-B5c2.T4a	CB-C3b	CB-C3c	CB-C3c.T1	CB-C3e	CB-C3e.Ch3a	CB-C3f
Yield (nmol)	5	140	290	27	216	13	156
N-Terminus	Glx	Leu	Leu	Leu	Val	Glx	Val
Amino acid	Content (residues/molecule)						
*Cys	+ (1)	–	–	–	1.0	–	0.5
Asp	–	2.3	2.3	2.2	2.2	0.3	3.0
Thr	–	2.8	4.0	3.5	1.7	0.6	1.7
Ser	+ (0)	1.8	2.1	1.8	0.9	0.3	–
Glu	+ (1)	2.1	2.2	2.2	2.2	2.2	5.2
Pro	–	–	–	–	3.4	2.0	2.5
Gly	± (0)	3.9	4.8	4.8	2.2	0.2	1.4
Ala	+ (1)	3.3	4.0	4.3	1.8	1.1	1.2
Val	–	1.8	1.9	2.0	1.3	–	0.5
*Met	–	+	+	+	+	+	+
Ile	–	2.7	3.4	3.1	2.4	–	1.6
Leu	–	2.2	2.4	1.8	4.1	–	2.2
Tyr	–	–	–	–	0.6	–	0.3
Phe	–	0.8	0.9	0.8	2.4	–	2.0
His	–	–	–	–	2.0	–	2.1
Lys	–	1.2	1.5	1.0	1.1	–	–
Arg	+ (1)	1.2	1.0	0.9	0.6	–	0.2
Total	(4)						
Tryptic peptide	T1E-9						

Peptide	CB-C3g	CB-D2a1	CB-D2a4	CB-D2b1	CB-D2c	CB-D2c.T1c	CB-D2c.T4b
Yield (nmol)	156	30	18	60	200	10	24
N-Terminus	Val	Asx	Phe	Asx	Phe	Phe	Val
Amino acid	Content (residues/molecule)						
*Cys	0.4	0.7 (1)	0.7 (1)	0.5 (1)	0.5 (1)	–	–
Asp	6.2	4.0 (5)	1.7 (2)	4.5 (5)	3.0 (2)	1.3 (1)	–
Thr	2.0	0.9 (1)	1.9 (2)	0.9 (1)	1.6 (2)	–	–
Ser	0.6	1.8 (2)	–	1.8 (2)	1.2 (0)	–	–
Glu	8.8	3.1 (3)	1.3 (1)	3.0 (3)	2.1 (1)	1.0 (1)	0.3 (0)
Pro	4.2	–	0.8 (1)	–	2.1 (1)	0.7 (1)	0.7 (1)
Gly	2.2	0.7 (0)	2.8 (3)	0.8 (1)	2.7 (3)	2.4 (2)	0.3 (0)
Ala	1.9	2.1 (2)	1.3 (1)	2.0 (2)	2.0 (1)	1.2 (1)	–
Val	0.8	3.8 (4)	4.1 (5)	3.7 (4)	4.5 (5)	1.9 (2)	1.0 (1)
*Met	++	+ (1)	+ (1)	+ (1)	+ (1)	–	+ (1)
Ile	2.4	1.1 (1)	0.7 (1)	1.1 (1)	1.1 (1)	0.6 (1)	–
Leu	3.3	2.0 (2)	0.4 (0)	2.1 (2)	1.1 (0)	–	–
Tyr	0.4	–	0.7 (1)	–	0.7 (1)	–	–
Phe	3.0	0.8 (1)	0.6 (1)	0.7 (1)	0.8 (1)	0.8 (1)	–
His	2.9	–	–	–	–	–	–
Lys	–	0.7 (1)	1.0 (1)	0.6 (1)	1.0 (1)	0.4 (1)	–
Arg	0.4	3.1 (3)	2.5 (3)	3.0 (3)	3.2 (3)	1.0 (1)	–
Total		(27)	(24)	(27)	(24)	(12)	(3)
Tryptic peptide							

Table 1—(*continued*)

Peptide ...	CB-D2c.T5a	CB-D2d2	CB-D2d2.Ta1	CB-D2e1	CB-D2e1.Ta	CB-D2e1.Tb
Yield (nmol) ...	7	72	25	108	18	18
N-Terminus ...	Cys*	Ser	Ser	Val	Val	Ala
Amino acid	Contents (residues/molecule)					
Cys*	1.0 (1)	0.7 (1)	0.8 (1)	–	–	–
Asp	1.0 (1)	1.0 (1)	–	5.0 (5)	2.9 (3)	2.1 (2)
Thr	–	–	–	1.0 (1)	0.8 (1)	–
Ser	0.4 (0)	3.0 (4)	3.3 (4)	1.1 (1)	1.1 (1)	–
Glu	0.4 (0)	0.8 (1)	–	2.3 (2)	1.9 (2)	–
Pro	–	1.0 (1)	– (1)	–	–	–
Gly	0.4 (0)	1.3 (1)	0.9 (0)	1.2 (1)	1.0 (1)	–
Ala	–	3.1 (3)	1.3 (1)	3.9 (4)	2.9 (3)	0.9 (1)
Val	1.2 (1)	2.0 (2)	1.2 (1)	2.8 (3)	2.8 (3)	–
Met*	–	+ (1)	–	+ (1)	–	+ (1)
Ile	–	–	–	1.7 (2)	0.8 (1)	0.9 (1)
Leu	–	–	–	1.3 (1)	1.1 (1)	–
Tyr	1.0 (1)	0.6 (1)	0.9 (1)	0.9 (1)	–	1.0 (1)
Phe	–	–	–	1.0 (1)	0.9 (1)	–
His	–	–	–	–	–	–
Lys	–	1.0 (2)	1.0 (1)	–	–	–
Arg	1.2 (1)	0.9 (1)	1.0 (1)	1.0 (1)	0.9 (1)	–
Total ...	(5)	(18)	(11)	(24)	(18)	(6)
Tryptic peptide ...	T1E-13					

Peptide ...	CB-D2e2	CB-D3a1	CB-D3b1	CB-D3c1	CB-D3d1	CB-D3d3
Yield (nmol) ...	33	40	30	42	90	160
N-Terminus ...	Val	Gly	Gly	Ala	Gly	Cys*
Amino acid	Content (residues/molecule)					
Cys*	–	0.5 (1)	1.0 (1)	–	–	1.0 (1)
Asp	5.2 (5)	1.0 (1)	1.0 (1)	0.7	0.5 (0)	3.0 (3)
Thr	1.1 (1)	–	–	1.0	1.5 (2)	–
Ser	1.5 (1)	1.1 (1)	1.0 (1)	1.4	1.8 (2)	2.6 (3)
Glu	2.4 (2)	1.2 (1)	1.5 (1)	1.8	1.4 (1)	2.2 (2)
Pro	–	–	–	–	0.6 (0)	–
Gly	1.7 (1)	2.0 (2)	2.3 (2)	1.1	1.6 (2)	0.3 (0)
Ala	4.3 (4)	1.1 (1)	1.3 (1)	1.2	2.8 (3)	1.2 (1)
Val	2.6 (3)	0.8 (1)	0.8 (1)	1.5	1.4 (1)	0.3 (0)
Met*	+ (1)	+ (1)	+ (1)	+	+ (1)	+ (1)
Ile	1.9 (2)	2.6 (3)	2.4 (3)	0.9	–	–
Leu	1.6 (1)	0.9 (1)	1.4 (1)	2.2	0.3 (0)	2.9 (3)
Tyr	1.0 (1)	–	–	–	–	–
Phe	0.9 (1)	–	–	–	–	–
His	–	–	–	–	–	–
Lys	–	–	–	–	1.0 (1)	–
Arg	0.9 (1)	1.8 (2)	1.9 (2)	0.3	0.4 (0)	0.3 (0)
Total ...	(24)	(15)	(15)		(13)	(14)
Tryptic peptide ...						

Table 1—(*continued*)

Peptide ...	CB-D3e5	CB-D3f1	CB-D3f2	CB-D3g1	CB-E1a1a	CB-E2n1
Yield (nmol) ...	120	60	96	252	72	288
N-Terminus ...	Glx	Lys*	Glx	Ile	Leu	Tyr
Amino acid	Content (residues/molecule)					
Cys*	–	–	–	–	–	–
Asp	2.0 (1)	1.2 (1)	1.1 (1)	1.1	1.2 (1)	1.1
Thr	2.1 (2)	0.8 (1)	1.7 (2)	1.0	–	2.0
Ser	–	1.8 (2)	–	–	0.4 (0)	–
Glu	2.5 (2)	2.3 (2)	2.0 (2)	1.3	1.5 (1)	2.2
Pro	–	–	–	–	1.4 (2)	0.8
Gly	2.2 (2)	0.4 (0)	1.8 (2)	–	0.3 (0)	2.0
Ala	1.0 (0)	–	–	1.0	0.3 (0)	1.2
Val	2.5 (3)	–	1.8 (3)	0.3	1.0 (1)	1.0
Met*	+ (1)	+ (1)	+ (1)	+	+ (1)	+
Ile	0.4 (0)	–	–	0.9	–	–
Leu	1.7 (1)	1.0 (1)	1.3 (1)	4.3	1.1 (1)	1.1
Tyr	1.1 (1)	–	0.8 (1)	–	–	1.9
Phe	1.3 (1)	1.6 (2)	1.0 (1)	0.9	–	1.1
His	–	–	–	–	–	2.1
Lys	–	2.4 (3)	–	1.5	0.5 (1)	–
Arg	0.3 (0)	1.7 (2)	–	–	1.0 (1)	–
Total ...	(14)	(15)	(14)		(9)	
Tryptic peptide ...						

Peptide ...	CB-E2n2b	CB-E2b	CB-E2a3a	CB-E2a5
Yield (nmol) ...	144	180	36	30
N-Terminus ...	Val	Tyr	Pro	Ser
Amino acid	Content (residues/molecule)			
Cys*	–	–	–	1.1 (1)
Asp	2.0 (2)	1.0	0.2 (0)	–
Thr	–	1.9	–	–
Ser	1.7 (2)	–	0.2 (0)	0.9 (1)
Glu	0.7 (0)	2.2	1.3 (1)	0.6 (0)
Pro	–	0.8	1.7 (2)	–
Gly	0.3 (0)	2.0	–	–
Ala	–	1.2	–	–
Val	1.0 (1)	1.0	1.0 (1)	1.1 (1)
Met*	+ (1)	+	+ (1)	+ (1)
Ile	–	–	–	–
Leu	1.1 (1)	1.0	–	–
Tyr	–	1.8	–	–
Phe	1.0 (1)	1.0	–	–
His	–	2.2	–	–
Lys	–	–	0.7 (1)	0.9 (1)
Arg	1.0 (1)	–	1.0 (1)	–
Total ...	(9)		(7)	(5)
Tryptic peptide ...				

has been found in peptides from other digests of the protein, by trypsin [peptide B6 of Allen & Green (1978) and peptide T1C-10 of Allen (1980*a*)] and chymotrypsin (peptide Ch.B8a of Allen *et al.*, 1980).

Clearly there are only two main peptide sequences in these peptides from fraction CB-C, that in peptides CB-C3b and CB-C3c, and that in peptides CB-C3e, CB-C3f and CB-C3g (upper sequence), as well as the known lower sequence in the mixture CB-C3g. These two sequences have not been found among the soluble tryptic peptides from the succinylated protein, and these parts of the polypeptide chain need to be further investigated in conjunction with the aggregated tryptic peptides, which are probably associated with the lipid bilayer in the native protein.

Peptide CB-C3c was digested with trypsin, and tryptic peptides were isolated on a column of Sephadex G-50. The only pure peptide obtained was identical with the total peptide CB-C3c.

Peptide CB-C3e was digested with chymotrypsin and the chymotryptic fragments were isolated by gel filtration on a column of Sephadex G-50 followed by t.l.c. for the smaller peptides. The properties of the only peptide purified, CB-C3e.Ch3a, are given in Table 1.

Peptides from fraction CB-D

These were resolved into three subfractions by chromatography on a column (1 cm × 142 cm) of Sephadex G-50 (Superfine grade) in 50 mM-NH_4HCO_3/0.01% thiodiglycol. Subfraction CB-D1 yielded no pure peptides after further fractionation by t.l.c. Fraction CB-D2 was resolved on a column of DEAE-cellulose into five subfractions, from which peptides were purified by t.l.c. Fraction CB-D3 was similarly resolved. The following peptides were characterized; their properties are listed in Table 1.

Peptide CB-D2a1. This was identical with peptide CB-D2b1.

Peptide CB-D2a4. This was similar to peptide CB-D2c.

Peptide CB-D2b1 [*3* (*128–154*)]. This had the partial sequence:

Asx-(-)-Phe-

The peptide was digested with trypsin and fragments were isolated on a thin-layer cellulose peptide 'map'. The only pure peptide obtained was

Asx-(Leu,Ser,Lys*,Val,Glx,Arg)

with an electrophoretic mobility at pH 6.5 of −0.1. Peptide CB-D2b1 was identified from this data with the *C*-terminal 27 residues of peptide CB-B3c.

Peptide CB-D2c [*3* (*188–212*)] was not completely pure. The partial sequence was:

Phe-Val-Lys*-Gly-Ala-Pro-Glx-Gly-Val-Ile-Asx-(-)-

Cys*-Asx-Tyr-Val-Arg-(-)-Gly-(-)-(-)-(-)-Val-Pro-

The peptide (70 nmol) was digested with trypsin and fragments were separated on a column of Sephadex G-50 (Superfine grade) and purified by t.l.c. The properties of the tryptic peptides are listed on Table 1.

Peptide CB-D2d2 [*3* (*170–187*)]. This had the partial sequence:

Ser-Val-(-)-Cys*-(-)-Pro-Ala-(-)-(-)-(-)-(-)-Ala-Ala-(-)-

Gly-

The peptide was digested with trypsin, and tryptic fragments were resolved by chromatography on a column of Sephadex G-50 (Superfine grade). A single peak was obtained, which contained peptides with *N*-terminal alanine and serine. Only the serine peptide was isolated after t.l.c. on cellulose MN400. The analysis identified this peptide with the *C*-terminal 11 residues of peptide T1D-6, with the sequence:

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

Peptide CB-D2e1 [*4* (*96–119*)]. This had the partial sequence:

Val-Leu-Ala-Asx-Asx-Asx-Phe-Ser-Thr-Ile-Val-Ala-

Ala-Val-Glx-Glx-Gly-Arg-Ala-

The peptide was digested with trypsin and two tryptic peptides were separated by chromatography on a column of Sephadex G-50 (Superfine grade). The properties of the peptides are listed in Table 1. Peptide CB-D2e1.Tb was sequenced as:

Ala-Ile-Tyr-Asx-Asx-Met*

Peptide CB-D2e2. This was identical with peptide CB-D2e1.

Peptide CB-D3a1 and Peptide CB-D3b1 [*3* (*284–298*)]. These were identical, with the partial sequence:

Gly-Ser-Ile-Glx-Leu-Cys*-(-)-Asx-Ala-Gly-Ile-Arg

Peptide CB-D3c1. This had the partial sequence:

Ala-Leu-Ser-Val-Leu-Val-

Peptide CB-D3d1 [*4* (*83–95*)]. This was con-

taminated with peptide CB-E1a1α, but the sequence:

Gly-Ser-Gly-(-)-Ala-Val-Ala-Lys*-Thr-Ala-Ser-Glx-(-)-

could be determined.

Peptide CB-D3d3. This had the sequence:

Cys*-Asx-Ala-Leu-Asx-Ser-Leu-Ser-Glx-Asx-Glx-Ser-Leu-Met*

Peptides CB-D3e5 and CB-D3f2 [*3* (*261–274*)]. These had the partial sequence:

Glx-(-)-Glx-(-)-Asx-Leu-(-)-Phe-Val-Gly-(-)-Val-Gly-

Peptide CB-D3f1 [*3* (*155–169*)]. This was not sequenced.

Peptide CB-D3g1. This had the partial sequence:

Ile-Phe-Ala-Leu-(-)-(-)-Leu-Asx-Leu-Thr-Glx-(-)-Leu

Peptides in fraction CB-E

These were partially resolved on a column (1 cm × 145 cm) of Sephadex G-50 (Superfine grade) in 50 mM-NH_4HCO_3/0.01% thiodiglycol, and were purified by paper electrophoresis at pH 6.5 and by t.l.c.

Peptide CB-E1a1a [*3* (*275–283*)]. This had the sequence:

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

Peptide CB-E2n1. This had the partial sequence:

Tyr-Ala-Glx-Asx-Gly-(-)-Gly-Val-Thr-Tyr-His-Glx-Leu-Thr-Phe-

Peptide CB-E2n2b [*3* (*252–260*)]. This had the sequence:

Val-Leu-Asx-Asx-Ser-Ser-Arg(Phe-Met*)

Peptide CB-E2b. This was identical with peptide CB-E2n1, apart from the electrophoretic mobility (the peptides were basic and neutral respectively). The difference was probably due to the homoserine ⇄ homoserine lactone interconversion.

Peptide CB-E2b. This yielded the following thermolysin fragments:

CB-E2b.Th1, Tyr-Ala-Glx-Asx (Gly) Pro-Gly

CB-E2b.Th2a, His-Glx

CB-E2b.Th2b, Leu-Thr-His

CB-E2b.Th2c1, Phe-Met*

CB-E2b.Th2c2, Val-Thr-Tyr

From these results the sequence of peptides CB-E2n1 and CB-E2b was deduced to be:

Tyr-Ala-Glx-Asx-Gly-Pro-Gly-Val-Thr-Tyr-His-Glx-Leu-Thr-His-Phe-Met*

Peptide Cb-E2a3α [*3* (*277–283*)]. This was isolated in low yield. The sequence was:

Pro-Pro-Arg-Lys*-Glx-Val-Met*

This peptide was a partial-acid-hydrolysis product, derived from peptide CB-E1a1α.

Peptide CB-E2a5 [*3* (*37–41*)]. This had the sequence:

Ser-Val-Cys*(Lys*,Met*)

Reconstruction of long sequences within the Ca^{2+}-ATPase

The *N*-terminal residues, amino acid analyses and partial sequences of the CNBr-derived peptides and of their tryptic fragments, together with the sequences of the soluble tryptic, chymotryptic, peptic, staphylococcal proteinase and thermolytic peptides, have allowed the reconstruction of three long segments of the polypeptide chain that are probably external to the lipid bilayer and that are relatively hydrophilic. These three sequences are presented, together with the positions of peptides from which they were derived, in Figs. 3, 4 and 5. They have been numbered *2*, *3* and *4* in order from the *N*-terminus (see below). Sequence *1*, the *N*-terminal peptide (peptide T1C7 from Allen, 1980*a*), and sequence *5*, the *C*-terminal octapeptide are shown in Fig. 6.

Fig. 3 shows sequence *2*, of 116 residues. An additional 11 residues, not yet placed in sequence, are contained in peptides Ch.B3, CB-C3b and CB-C3a. Most of the sequence was given by peptide CB-B4a (residues *2* (19–99)) and peptide CB-B3a (*2* (19–127)). The sequence of residues *2* (96) to *2* (99) was deduced only from amino acid analyses of residual parts of peptides T3-E10b2b, Ch.C1a, Ch.C1b, CB-B4a, and CB-B4a.T2 together with the specificity of digestion by trypsin and CNBr, and the assignment of these four residues is at present tentative. Residue *2* (5) was determined twice to be glutamic acid (peptides T1C-8, T3-D4c2.Tha1), but once to be glutamine (peptide T3-D4c2). This anomaly is difficult to explain, but the cleavage by staphylococcal proteinase at this residue supports the identification as glutamic acid. Aspartic acid residues *2* (49) and *2* (54) were identified from the mobility of peptide P.E9b only. The acid or amide groups at positions *2* (93, 94, 95 and 98) were not determined, but could be assigned from the results of Klip & MacLennan (1978).

Fig. 4 shows the longest continuous sequence

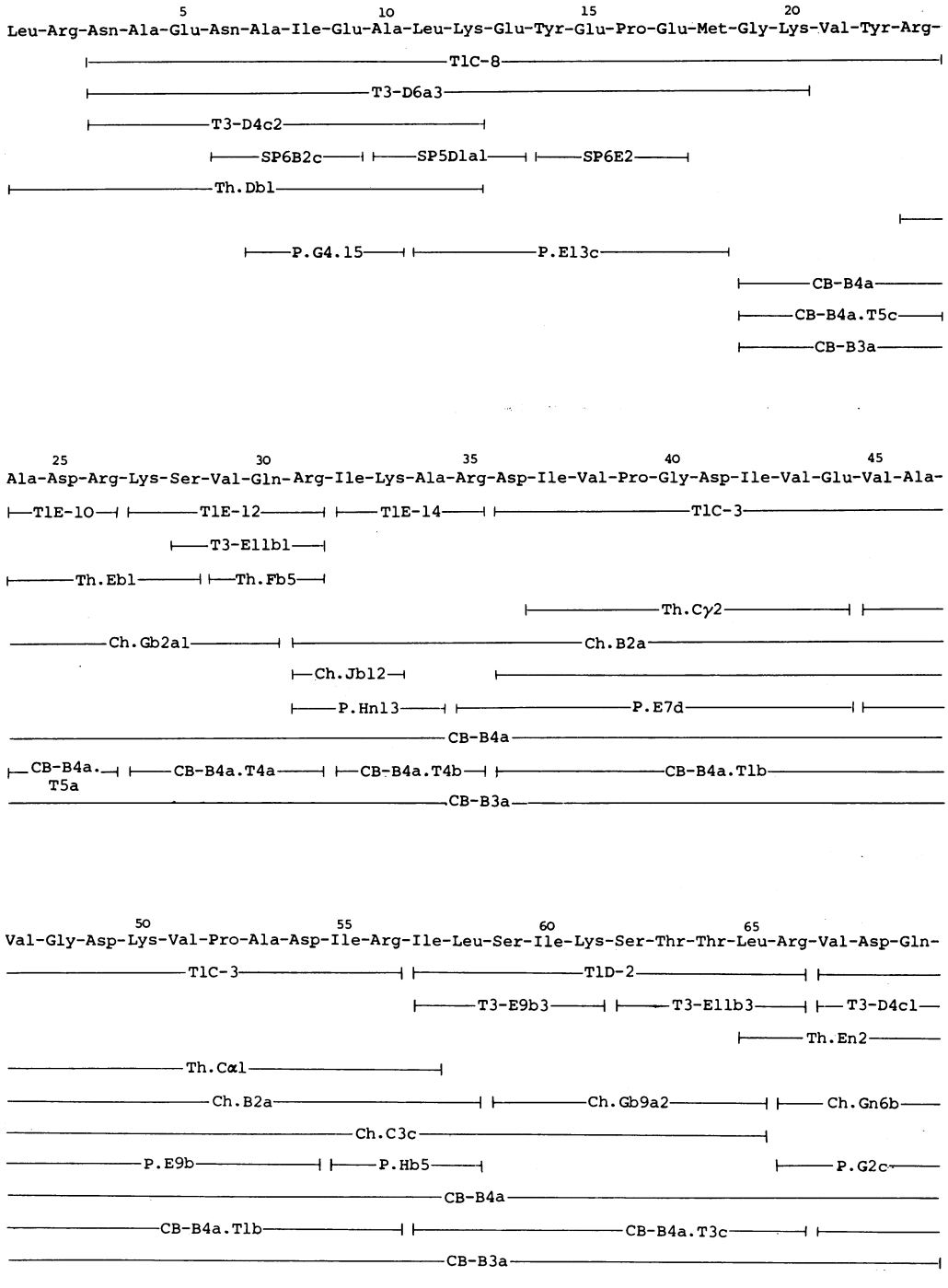

Fig. 3. *Sequence of 116 residues of the ATPase protein (sequence 2)*

The sequence is given on the top line, and below it are the positions of the peptides from which the sequence was deduced. Peptides from the tryptic digest of the succinylated protein (Allen, 1980*a*) are labelled T1. Peptides from the tryptic digests of the protein (Allen *et al.*, 1980) are labelled T2, T3 or T4. Peptides from the thermolytic digest are labelled Th., those from the peptic digest P, and those from the staphylococcal-proteinase digest SP (Allen *et al.*, 1980). Peptides from the chymotryptic digest of the protein are labelled Ch. (Allen, 1980*b*), and those from the CNBr digest, CB (the present work). The acid or amide assignments at positions 93, 94, 95 and 98 were from Klip & MacLennan (1978).

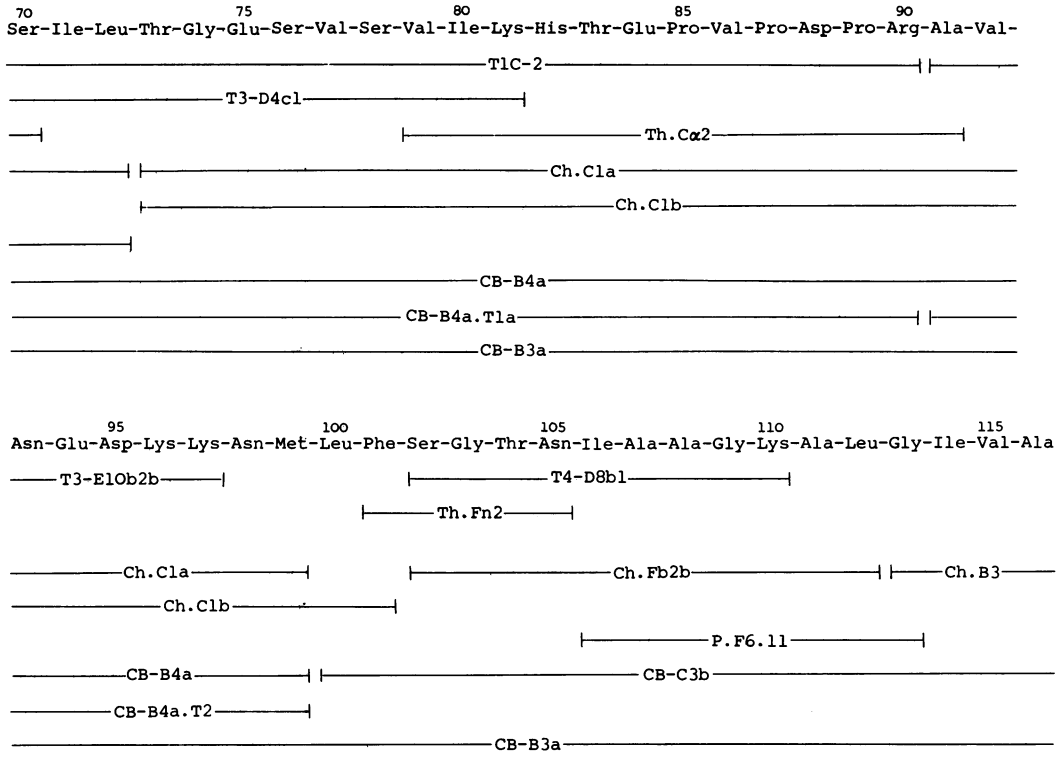

Fig. 3—(*continued*)

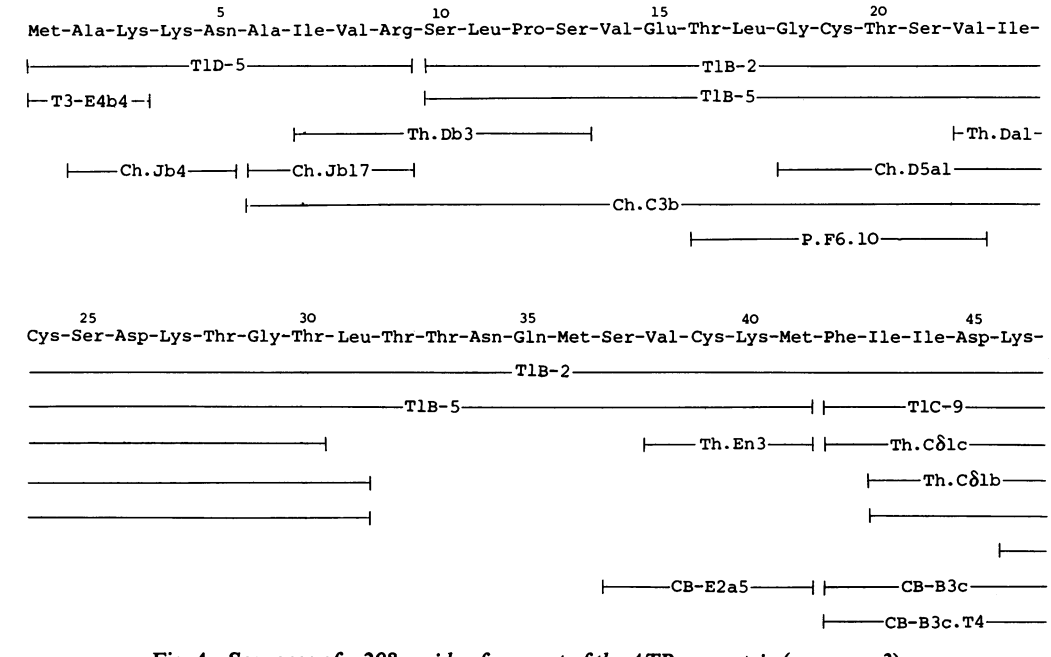

Fig. 4. *Sequence of a 298-residue fragment of the ATPase protein (sequence 3)*
Symbols are as for Fig. 3.

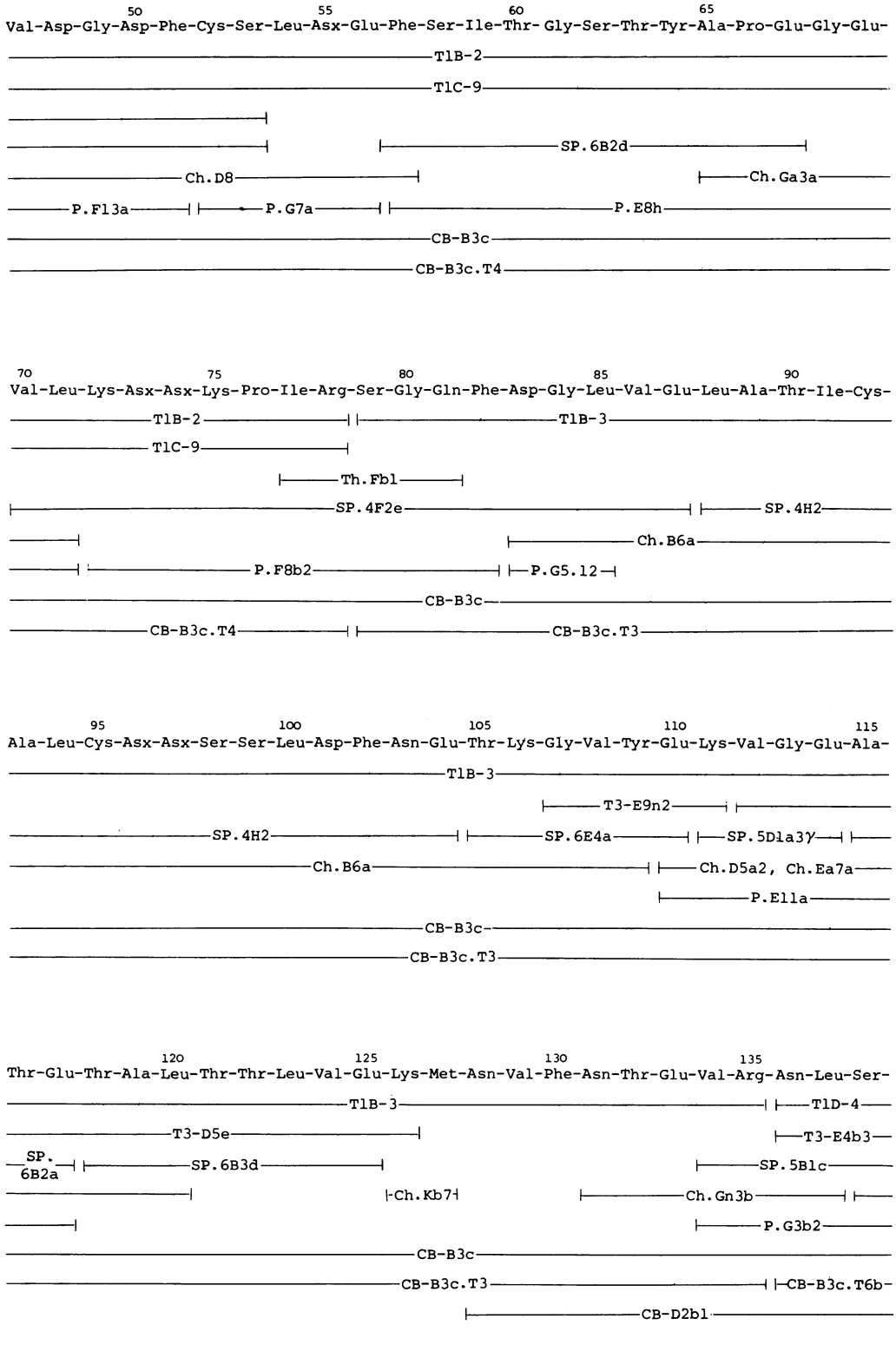

Fig. 4—(*continued*)

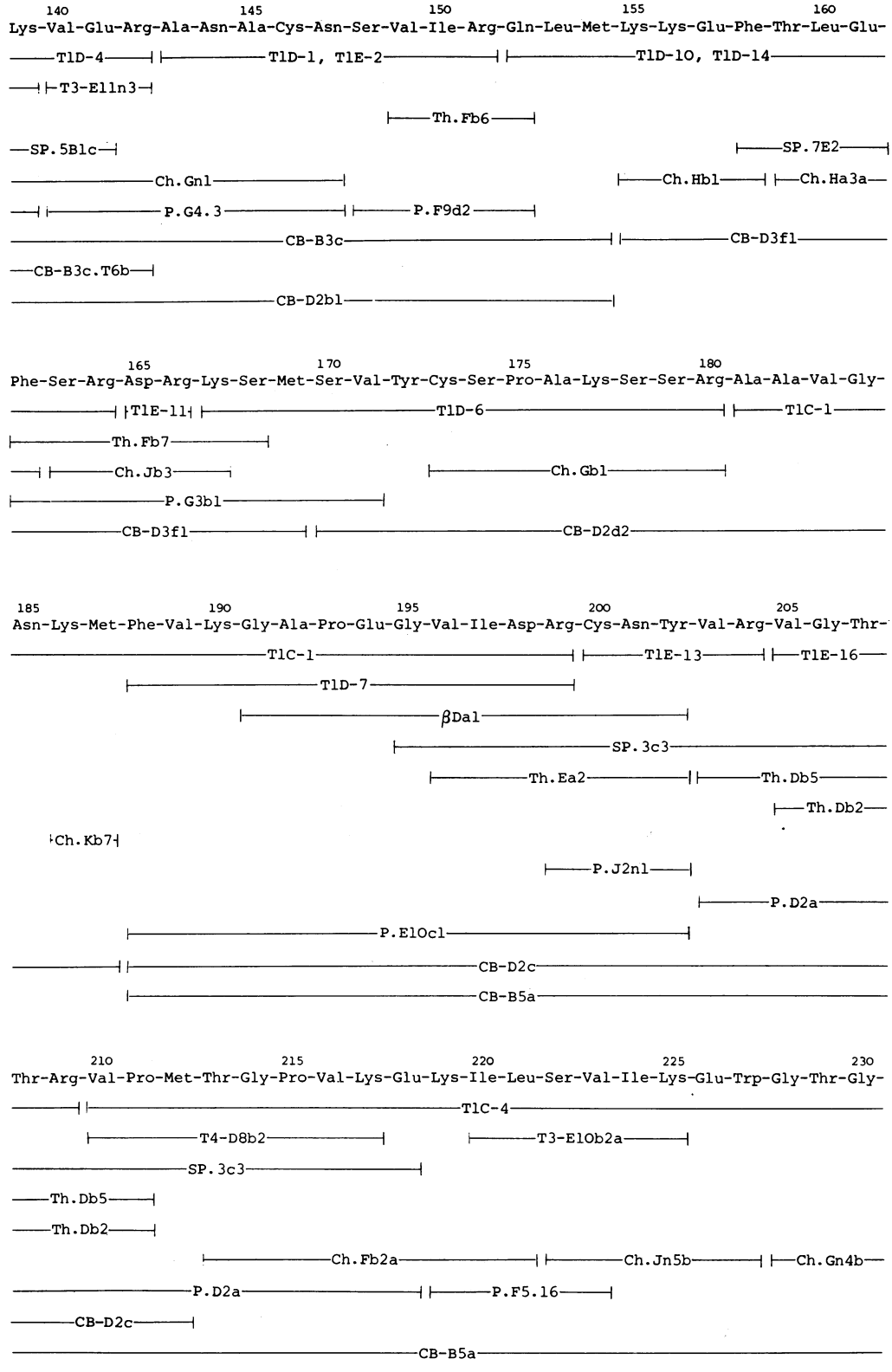

Fig. 4—(*continued*)

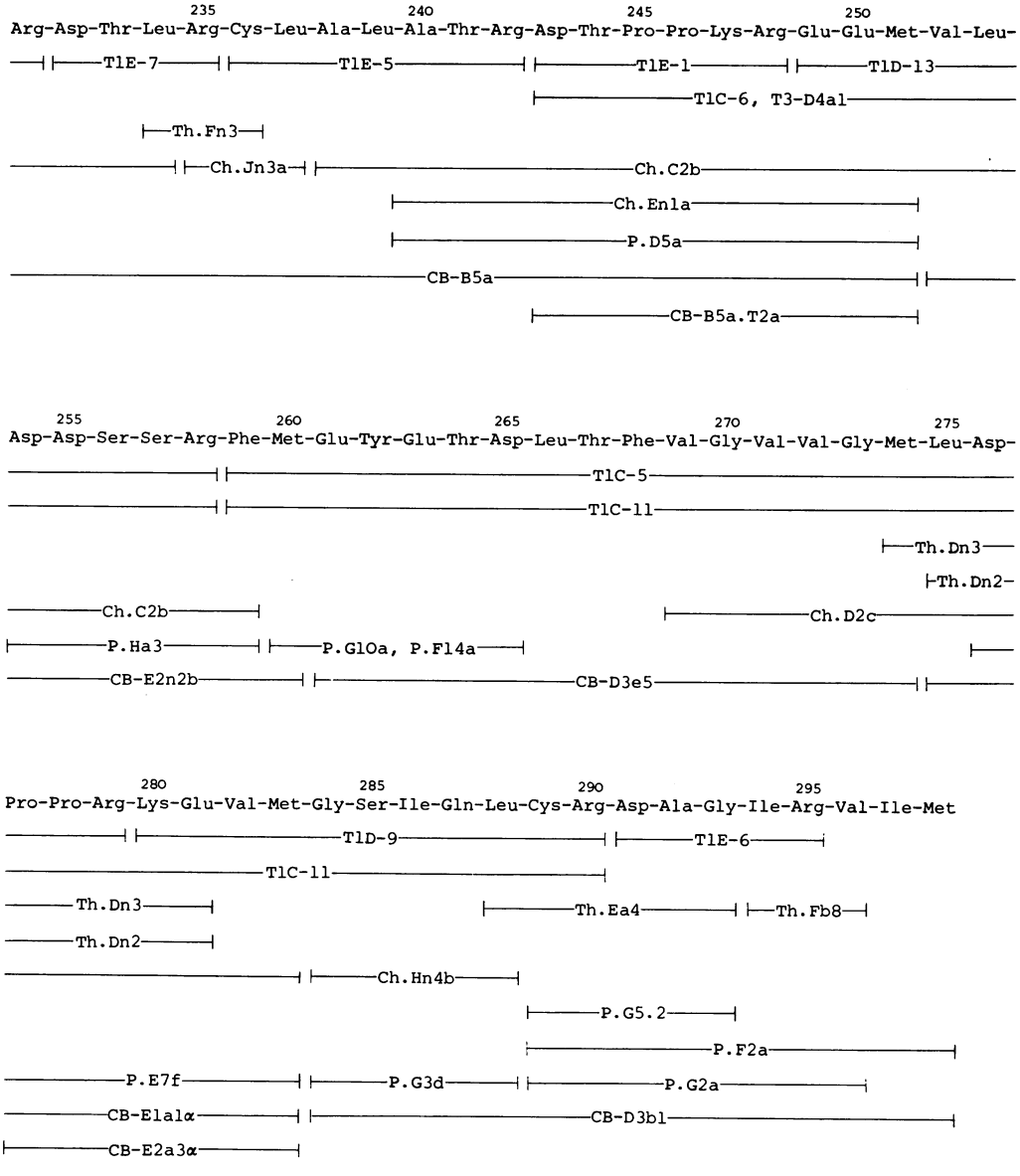

Fig. 4—(*continued*)

(sequence *3*) yet determined in the ATPase protein, of 298 residues. The bulk of this sequence was determined from tryptic peptides T1B-2 (residues *3* (10–78)) and T1B-3 [*3* (79–135)], and CNBr peptides CB-B3c [*3* (42–154)] and CB-B5a [*3* (188–251)]. The sequence from residue *3* (36) to residue *3* (42) was determined only from short peptides that were placed here from the amino acid analyses and the cleavage specificity of CNBr. Acid or amide residues were not determined at positions 55, 73, 74, 96 or 97, owing to the lack of isolation of suitable small peptides in sufficient quantity for the determination of the electrophoretic mobilities at each step in the Edman degradation. Glutamic acid residues were distinguished from the amide at positions 56 and 104 by the specificity of cleavage by *Staphylococcus aureus* V8 proteinase. The active-site aspartic acid residue is at position 26 in this sequence (Allen & Green, 1976).

Fig. 5 shows a sequence of 122 residues (sequence

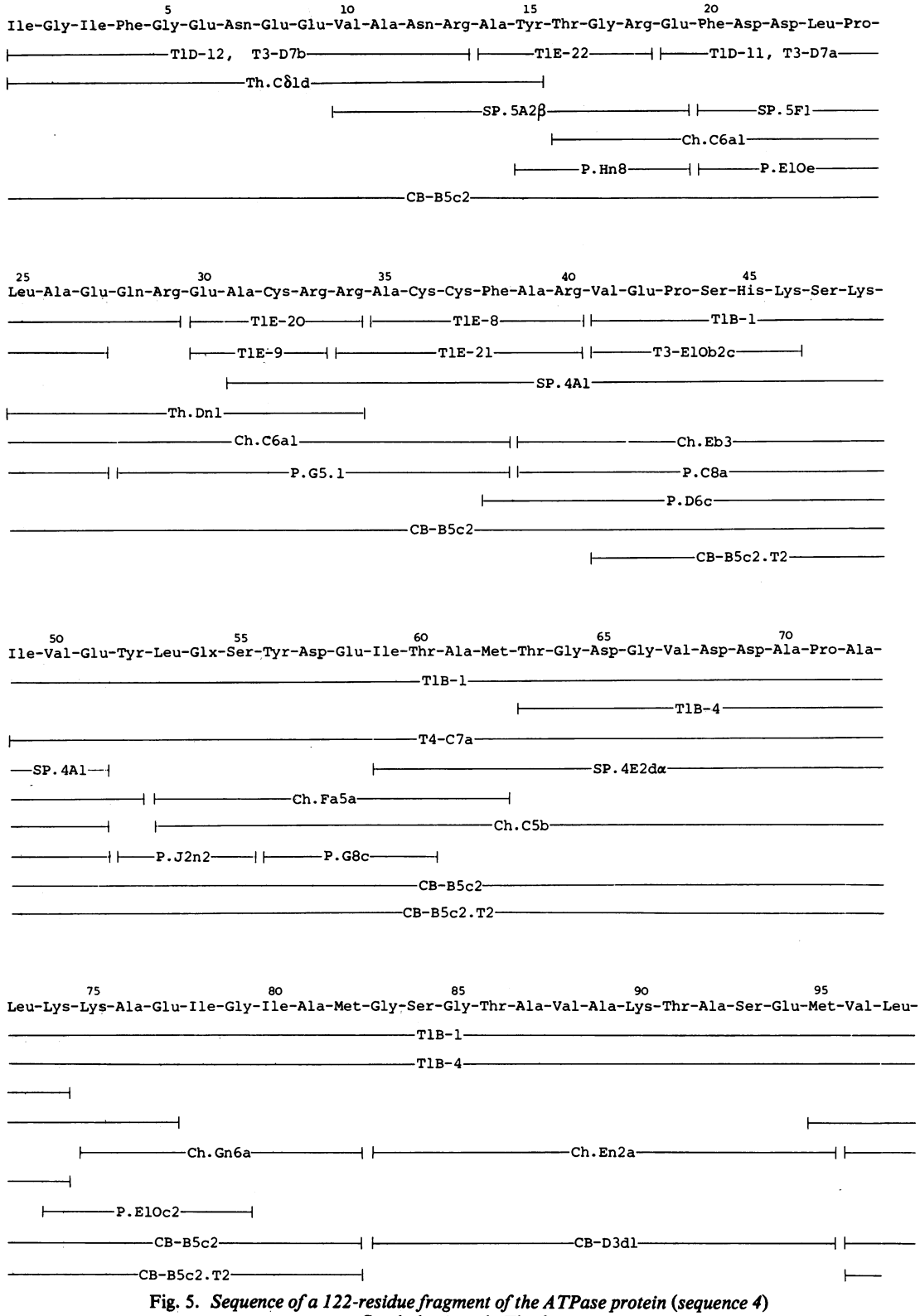

Fig. 5. *Sequence of a 122-residue fragment of the ATPase protein (sequence 4)*
Symbols are as in Fig. 3.

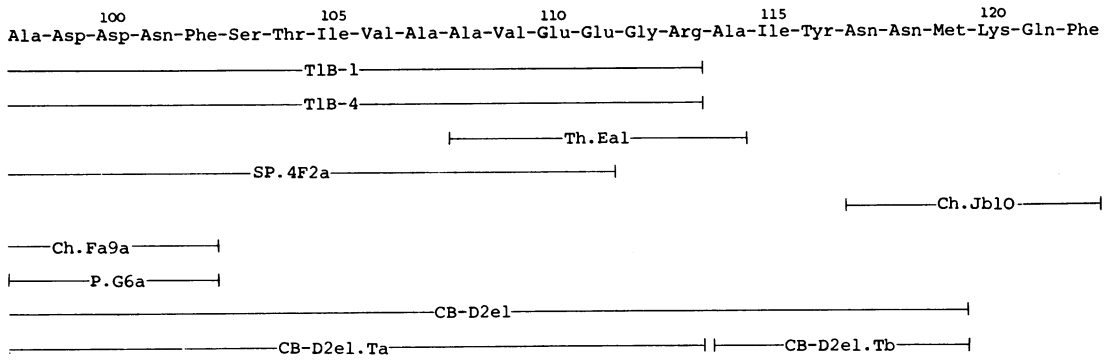

Fig. 5—(*continued*)

4). The *N*-terminal 82 residues of sequence *4* are included within peptide CB-B5c. Most of the sequence was found from the tryptic peptide (from the succinylated protein) T1B-1 (residues 41–113) and was determined at least twice in different peptides. However, there are some relatively-less-well-substantiated parts of the sequence. The sequence from residue *4* (88) to *4* (96) was determined from two peptides (Ch.En2a and CB-D3d1) which were impure by amino acid analysis, but the proposed sequence was also supported by the analyses of thermolysin fragments T1B-4.ThC2α and T1B-4.ThC2β. Most of the acidic and amide residues were clearly determined from the electrophoretic mobilities of small peptides. However, residue *4* (54) was determined to be glutamic acid in peptide T4-C7a.Ch3b, but glutamine in peptide P.Hn18.

Fig. 6 shows the 31-residue *N*-terminal peptide (sequence *1*) and the C-terminal octapeptide (sequence *5*) of the ATPase protein. The evidence that the latter is the *C*-terminal sequence is the known specificity of chymotryptic and tryptic cleavage, which is never after glycine residues, and the high yield of the *C*-terminal tryptic and chymotryptic peptides.

Before considering how the five sequences are aligned in the overall structure of the molecule, it is necessary to consider the extent and the nature of those parts of the molecule that remain unsequenced. The total number of residues in the molecule lies between 900 and 1100, depending on the value chosen for the molecular weight. Taking an approximate molecular weight of 110000 (1017 residues), about 180 residues remain to be placed in sequence. The difference between the total amino acid composition and that of the sequenced peptides is shown in Table 2, along with the composition of those parts that have been sequenced. It can be seen

Sequence *1* (*N*-terminus of the protein), peptide T1C-7:

X-Met-Glu-Ala-Ala-His-Ser-Lys-Ser-Thr-Glu-Glu-Cys-Leu-Ala-Tyr-Phe-Gly-Val-Ser-Glu-Thr-Thr-Gly-Leu-Thr-Pro-Asp-Gln-Val-Lys-Arg(His, Lys)

Sequence *5* (*C*-terminus of the protein):

Ile-Ala-Arg-Asn-Tyr-Leu-Glu-Gly

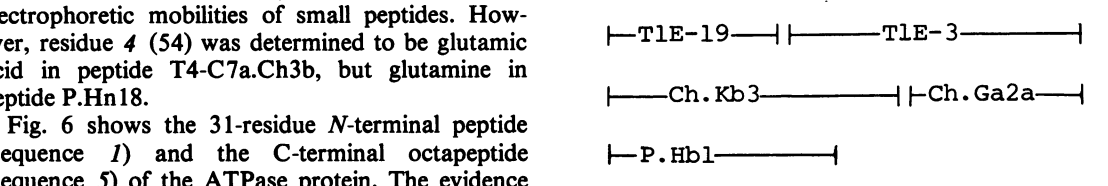

Fig. 6. *Sequence of the N- and C-terminal peptides of the ATPase protein (sequence 1 and 5)*
Symbols are as in Fig. 3. X is an acyl group, determined by Tong (1977) to be acetyl.

that, as expected, the unsequenced region is markedly hydrophobic (only 31% of the residues are polar) and that it contains, in addition to much leucine, phenylalanine and tryptophan, a disproportionately large number of glutamic acid and of proline residues. The long peptides *2*, *3* and *4* and the *N*-terminal peptide (*1*), are much less hydrophobic, resembling water-soluble globular proteins. The peptides making up the 260 residues that have been sequenced but not overlapped with sequences

Table 2. *Amino acid composition of sequenced and non-sequenced regions of the ATPase protein*

	No. of residues			
	Total in sequences *1–5*	Total in other known sequences	Total in ATPase* (mol/110kg)	Non-sequenced residues (by difference)
Lys	40	10	51	1
His	3	8	13	2
Arg	37	8	51	6
Asp	57	21	87	9
Thr	40	14	60	6
Ser	40	13	64	11
Glu	60	25	107	22
Pro	21	14	49	14
Gly	39	20	73	14
Ala	51	21	85	13
Cys	15	6	24	3
Val	52	17	79	10
Met	20	13	34	1
Ile	33	18	59	8
Leu	35	36	98	27
Tyr	14	6	22	2
Phe	17	10	42	15
Trp	1	0	19	18
Total residues ...	575	260	1017	182
Polar residues (Lys,His, Arg,Asx,Thr,Ser,Glx) ...	277	99	433	57
Polar residues (%)	48	38	43	31

* The total number of residues per molecule (mol.wt. 110000) was calculated from the mean of three amino acid compositions (MacLennan *et al.*, 1971; Meissner *et al.*, 1973; Thorley-Lawson & Green, 1975). Where the result for a particular amino acid from one analysis differed markedly from those from the other two, it was disregarded (e.g. serine, histidine, arginine, isoleucine). The number of non-sequenced residues was determined by difference. The numbers of methionine, leucine, tyrosine and phenylalanine residues in the last column are probably underestimates, since these amino acids frequently occur as *N*- and *C*-terminal residues in the shorter peptides and some may therefore have been counted twice in different sequenced peptides. The criteria of Capaldi & Vanderkooi (1976) were used to classify residues as polar or non-polar.

1–5, are intermediate in character, with 38% of polar residues, Three of the longer ones (Ch.En5, ChD3f and CB-C3e), accounting for about 50 residues, are sufficiently hydrophobic (75%) to be considered as candidates for membrane-buried peptides, but most of them are not unusually hydrophobic. The *C*-terminal 18 residues of sequence *2* show a similar high hydrophobicity. Since the 'A' peptides in each of the different digests account for somewhat over 40% of the total protein they must include, in addition to the 180 residues of the unsequenced peptides, many of the more polar non-overlapped peptides. Exactly which ones will depend on the cleavage method that has been used to generate the particular set of A peptides. This general conclusion has been confirmed in experiments (G. Allen and B. J. Trinnaman, unpublished work) on one or two 'A' peptides that have been purified. Peptic digestion of these released small soluble peptides found previously in other digests of the whole molecule.

Discussion

CNBr digest

As with other digests of the ATPase protein (Allen & Green, 1978; Allen, 1980*a*,*b*; Allen *et al.*, 1980), a considerable proportion of the protein, including almost all of the tryptophan residues, yielded, after cleavage with CNBr, peptides that although soluble were aggregated in aqueous solution and excluded from Sephadex G-100 beads (fraction CB-A). These peptides are identified with the intramembranous regions of the protein, and special conditions, such as the use of phenol-containing solvents (Allen, 1978) are required for their purification and sequence determination. Some peptides from fraction CB-A have been purified and characterized (G. Allen & B. J. Trinnaman, unpublished work). Analysis by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis of the peptides in fraction CB-A showed that a large number of peptides were present, in widely varying

amounts, showing that cleavage by CNBr was not complete. Similar difficulty of digestion with CNBr of intramembranous parts of a membrane protein have been reported by other workers (Ozols & Gerard, 1977).

Cleavage of the remainder of the protein was, however, satisfactory. As is usually observed, only partial cleavage of methionine–threonine bonds [positions *4* (62–63) and *3* (212–213)] was observed. A methionine–asparagine bond [position *3* (127–128)] was also only partially cleaved. Cleavage of one aspartic acid–proline bond, through acid-catalysed hydrolysis (Piszkiewicz *et al.*, 1970) was observed [position *3* (276–277)].

The isolation of the peptides was slightly hindered by the interconversion of *C*-terminal homoserine and homoserine lactone (Ambler, 1965), which resulted in the distribution of each peptide in two positions on electrophoresis or t.l.c. This interconversion also prevented the use of electrophoretic mobility (Offord, 1966) in the determination of acidic or amide residues in peptides with *C*-terminal homoserine.

The sequences determined in the present paper are very probably from the extramembranous portion of the protein, as discussed in an accompanying paper (Allen, 1980*a*), whereas peptides in fraction CB-A are most likely mainly from the intramembranous portion of the protein (see below). A total of 562 unique residues of the protein were identified in the sequences of the CNBr peptides presented here. This value is similar to the number of residues released in soluble peptides by digestion with trypsin or chymotrypsin (Allen, 1980*a*,*b*), and the sets of peptide sequences contained within these soluble peptides are almost the same.

Properties of the long sequences determined within the ATPase protein

The total number of residues placed within unique sequences of the protein in peptides from the digests so far studied is 835, of which 536 are in the three large fragments reported here. Additional fragments of interest are the 31-residue *N*-terminal peptide, T1C-7 (Allen, 1980*a*) and the *C*-terminal octapeptide, both shown in Fig. 6. The *C*-terminal sequence differs slightly from that proposed by Ohnoki (1978), based on digestion of the protein with carboxypeptidase Y.

The amino acid sequences have been scanned for regions of internal homology, and several pairs of identical tripeptide and tetrapeptide sequences have been found. Many of these could have arisen by chance, since they are not correlated with each other in the sequence. There is, however, one region where four of these homologies show some correlation and this will be considered below in relation to the secondary structure. The sequence Lys-Met-Phe occurs twice [*3* (40–42) and *3* (186–188)] and it is noteworthy that both were found to be cleaved by trypsin at the Met–Phe bond. Three similar sequences containing Pro-Pro are shown in Fig. 7(*a*). A more extended homology (Fig. 7*b*) shows ten identities in 25 residues.

A sequence showing several internal duplications is shown in Fig. 7(*c*), and five sequences related mainly through the pattern of leucine residue in Fig. 7(*d*). It is significant that the latter sequences originate from peptides outside sequences *1–5*, which probably derive from those parts of the chain that interact with the membrane.

From these results it can be seen that the amount of internal homology in the ATPase is very low.

Location of cysteine residues

In a previous paper (Allen & Green, 1978) we identified 18 different sequences containing cysteine residues. Three more peptides can now be added to this list, all of them located outside the five main sequences. These are peptides SP7E1 (forming part of peptide T1C10*a*), CB D3d3 and ThFb1e2. Of those previously identified, one was found in sequence *1*, none in sequence *2*, eleven in sequence *3* and three in sequence *4*. The remaining three were released from the aggregated 'A' peptides by chymotrypsin. It is likely that one or two cysteine residues remain to be found. It is not yet known which cysteine residues form disulphide bonds nor which correspond to reactive and buried thiol groups.

The relation of the long sequences to the major tryptic fragments

The native ATPase protein is rapidly cleaved by trypsin to give two fragments of mol.wts. 60000 and 55000, the former of which is then cleaved into fragments of 33000 and 24000 mol.wt. (Thorley-Lawson & Green, 1973; Migala *et al.*, 1973; Stewart & MacLennan, 1974), which have been labelled A1 (33000), A2 (24000) and B (55000) (Thorley-Lawson & Green, 1977). MacLennan and coworkers (Klip & MacLennan, 1978; Stewart *et al.*, 1976) have investigated the *N*-terminal sequences of these fragments. From the close similarity between the *N*-terminal sequence of one fragment and the sequence following residue 180 of sequence *3*, we conclude that the bond beween residues arginine-180 and alanine-181 is sensitive to trypsin in the native protein. Both of the released *N*-terminal residues had previously been identified as alanine (Thorley-Lawson & Green, 1975). Since the active-site aspartic acid residue is at position 26 of sequence *3*, and is in fragment A1 (Thorley-Lawson & Green, 1973; Stewart & MacLennan, 1974), the *C*-terminal 180-residue sequence of this fragment is determined. The order of the large fragments from the native protein can also be deduced from these results. The remaining 118 residues of sequence *3*

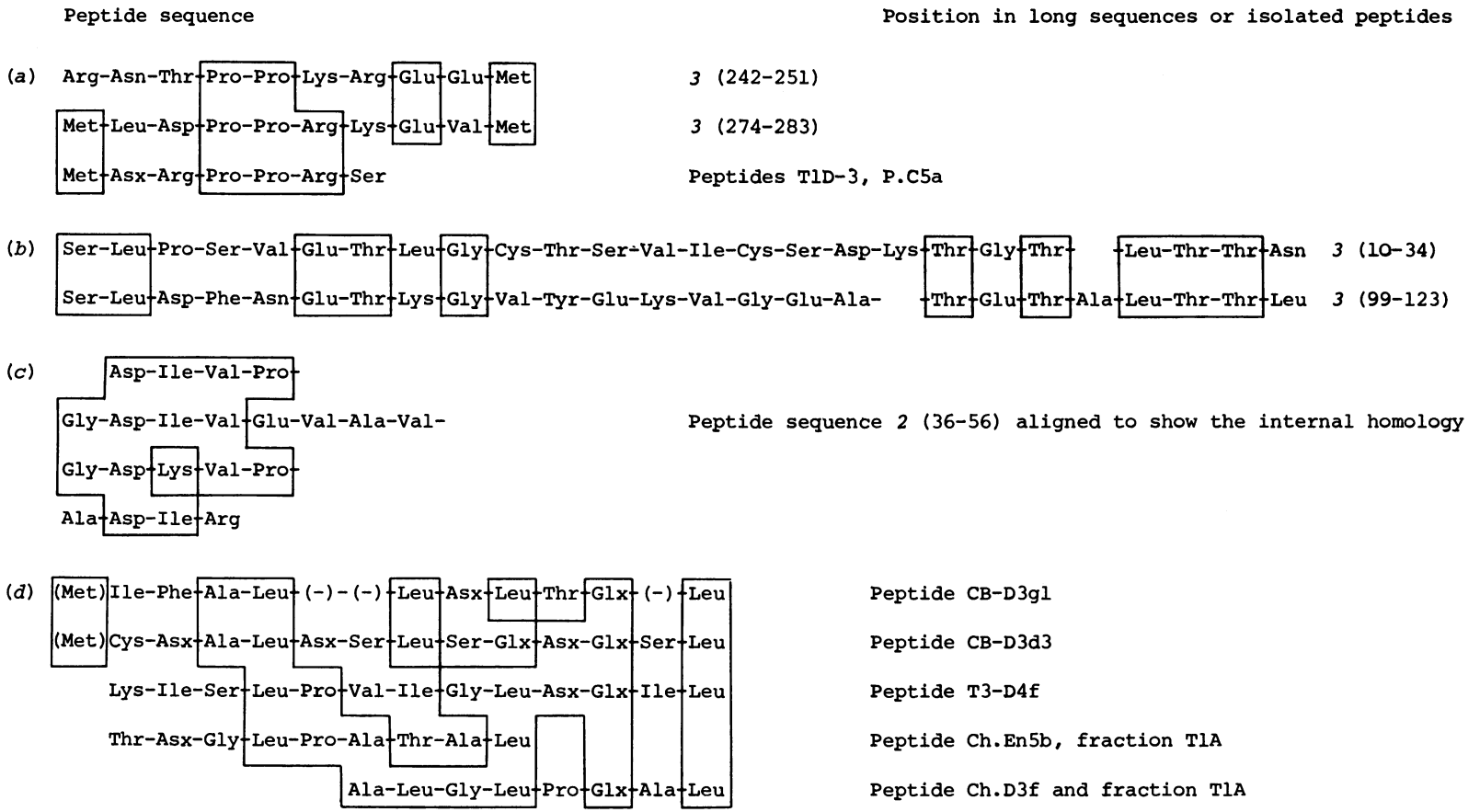

Fig. 7. *Comparison of homologous sequences within the ATPase protein*
Residues that are identical in two or more sequences are enclosed by boxes.

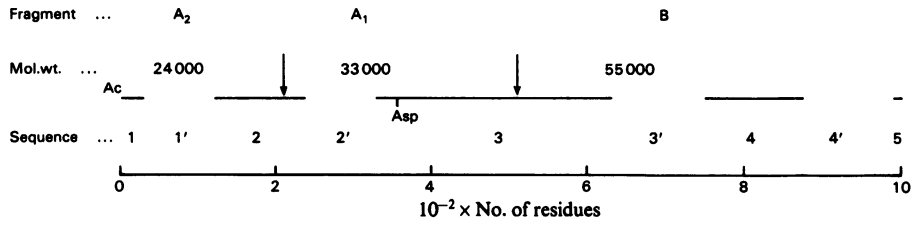

Fig. 8. *Linear map of the ATPase*

The relation between the long sequences described in the present paper and the tryptic fragments A_2, A_1 and B is shown. The four gaps have been numbered 1', 2', 3', and 4'. Their location is approximate to the extent that the molecular weights of the fragments are approximate and the exact position of sequence *4* in the second half of fragment B is unknown.

contain too many (seven) methionine residues for this sequence to be included within fragment A2, which has only three or four methionine residues (Thorley-Lawson & Green, 1975), and these 118 residues thus constitute the *N*-terminal sequence of the fragment B. This conflicts with the results of Stewart *et al.* (1976), but these have recently been reinterpreted (Klip & MacLennan, 1978). Since fragments A1 and A2 are contiguous, the order of these fragments in the protein is thus deduced to be A2-A1-B.

This result has been confirmed by the demonstration that the *N*-terminus of fragment A2 is blocked (Klip & MacLennan, 1978), as is the *N*-terminus of the protein. The second point of cleavage of the native protein by trypsin can also be identified from MacLennan's (Klip & MacLennan, 1978) results as the bond Arg-90–Ala-91 of sequence *2*. These conclusions are summarized in the linear map of the protein shown in Fig. 8.

This map agrees with the ordering of the fragments proposed by Klip & MacLennan (1978), although there are some differences of detail between our results and their *N*-terminal sequences. These differences are all at residues that we identified in several peptides as lysine, cysteine, serine or threonine [*2* (97), *2* (102), *2* (104), *3* (186), *3* (190), *3* (199) and *3* (200)], but that they found to be hydrophobic residues (leucine or valine). Since the remainder of the 20 residues' sequences are identical, these differences can probably be attributed to the difficulty of identifying the phenylthiohydantoin derivatives of lysine, cysteine, serine and threonine by the gas-chromatographic method that was used by these authors. In view of the generally low degree of internal homology in the ATPase, the existence of a second almost identical sequence in which these residues are replaced by leucine or valine is very unlikely. [Since the original submission of the present paper, the *N*-terminal sequences of tryptic fragments A1 and B have been reinvestigated by MacLennan and co-workers (Klip *et al.*, 1980), and their results are now in agreement with those reported in the present paper.]

The only difference that this makes to the general conclusions about structure is that our results show that the *N*-terminus of the B-fragment is not especially hydrophobic and is unlikely to be buried in the membrane. This is in accordance with the accessibility of the site to tryptic cleavage.

It can be concluded from the molecular weights of fragments A_1 and A_2 that each has a gap of about 90 residues between the regions of known sequence. The B-fragment has two gaps of undefined relative size totalling about 250 residues. It is known from the labelling of each fragment with a hydrophobic photoactivated azide (Green *et al.*, 1980), from their similar contents of tryptophan and from their firm anchorage in the membrane (Thorley-Lawson & Green, 1975) that each has an extensive membrane-associated segment. It is not yet possible to say whether the B fragment has one or two of such segments, since the exact location of sequence *4* is uncertain.

Conclusions concerning the number of traverses across the membrane in each buried segment require a knowledge of the location of the sequences *1–5* in relation to the membrane surface. The only certain evidence we have is the location of the site of phosphorylation and hence of sequence *3* on the cytoplasmic face. A variety of other evidence suggests that sequences *2* and *4* are also located on this face: (1) the size of the negatively stained cytoplasmic projections show that they account for at least half the molecule, considerably more than the 300 residues of sequence *3*; (2) the intramembranous particles observed in freeze-fractured preparations are located exclusively on the cytoplasmic leaflet of the bilayer, suggesting that the protein is anchored only weakly to the cisternal

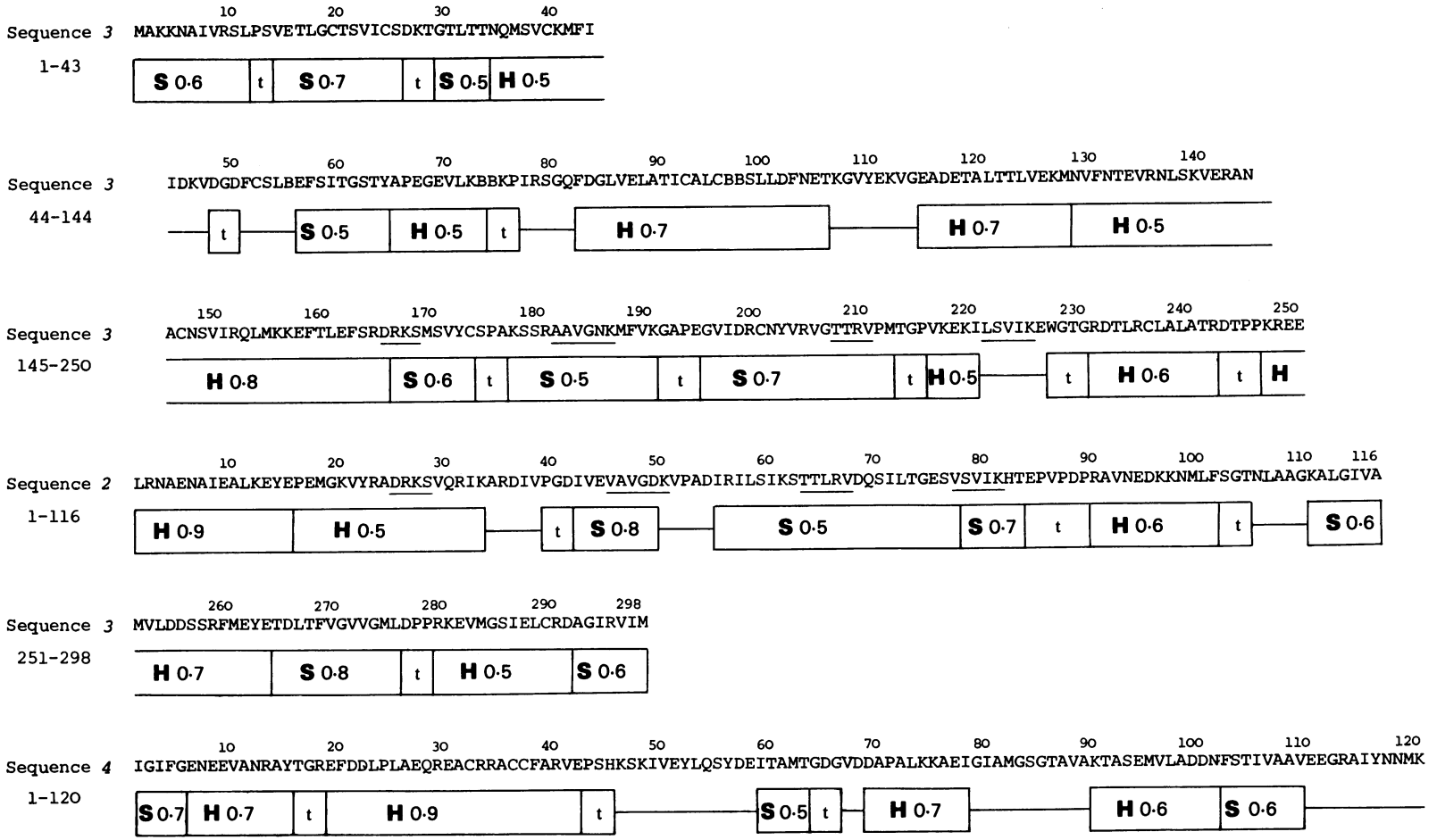

Fig. 9. *Secondary structure of sequences 2, 3 and 4 (McLachlan, 1978)*

Predicted α-helix is indicated by H, predicted β-sheet by S and β-turns by t. The numbers indicate the maximum probability given by the prediction for the segment enclosed. Where the probability is <0.5 for helix or sheet, or <0.4 for turn, no structure is indicated, even though the method predicted helix, sheet or turn in most of these regions, rather than a random structure. Sequence *2* has been inserted between two sections of sequence *3* to facilitate comparison of possible homologous regions. Details of the one-letter notation for amino acids are given in *Biochem J.* (1969) **113**, 1–4.

leaflet; (3) the cleavage of sequence *2* after residue 90 by the action of trypsin on the native membrane suggests that it is located on the cytoplasmic face; (4) all three fragments are iodinated by lactoperoxidase in whole sarcoplasmic reticulum (Thorley-Lawson & Green, 1973, G. M. Hebdon, unpublished work); (5) antibodies raised against the separated fragments bind to inact sarcoplasmic reticulum, which can then be precipitated with anti-immunoglobin serum. The anti-A_1 serum gave most precipitation, but both anti-A_2 and anti-B sera were quite effective precipitants (Stewart *et al.*, 1976); (6) experiments on inside-out vesicles of sarcoplasmic reticulum (Sarzała & Michalak, 1978) showed no binding of these vesicles to thiolated Sepharose. Since sequence *4* has three thiol groups clustered in a very polar region it is likely that this segment is located, together with the remainder of the water-soluble segment of fragment B, on the cytoplasmic surface.

It is not yet possible to say whether the *N*- and *C*-terminal sequences are cytoplasmic or intracisternal. The maximum number of traverses would occur if both the *N*- and the *C*-termini were on the cytoplasmic face. If there were three buried segments, this would require six traverses, whereas four buried segments (i.e. if there were two associated with the B-fragment) would require eight. This argument is based on the assumption that each segment penetrates to the cisternal face of the membrane, for which there is as yet no direct evidence.

Secondary structure of the ATPase

The multiple insertions of the peptide chain into the membrane accounts for the difficulty experienced in releasing the large proteolytic fragments of the ATPase and makes it unlikely that crystallizable fragments of the protein can be obtained. At present the secondary and tertiary structure of the protein are accessible only by indirect methods. The program of McLachlan (1978) was used to predict the secondary structure of sequences *2*, *3* and *4* (Fig. 9) in the hope that some clue to the location of Ca^{2+} or ATP-binding sites might be obtained. In the same Figure we have shown the location of several short internal homologies between sequences *2* and *3* which may have some collective significance. Extensive regions of α-helical and β-structure are evident. The fraction of residues in helical regions is 0.4, somewhat higher than the value of 0.36 for the whole molecule, deduced from the c.d. spectrum (Hardwicke & Green, 1974). Since the membrane-buried regions are also likely to have a high helical content (about 50%) [N. M. Green, unpublished work (c.d. measurements)], it is likely that helix is overpredicted and that a higher probability than 0.5 is required to determine a helical segment. More detailed analysis at this stage would be premature. The most significant general conclusions from Fig. 9 are that the regions of β-structure are segregated from the helical regions, suggesting that any β-sheets are anti-parallel. There is little alternation of helical and β-sheet segments, which might give rise to the parallel β-sheet characteristic of the nucleotide-binding sites of many dehydrogenases, so there is no basis for ascribing an ATP-binding site to such a region. It is noteworthy that several kinases also have nucleotide-binding sites that differ from those of the dehydrogenases (Levine *et al.*, 1978). The phosphorylation site [asparagine-*3* (26)] is on a β-bend situated in an anti-parallel-β-sheet region, followed in the sequence by some weakly predicted helix.

The other noteworthy feature of Fig. 9 is the resemblance between two predominantly β-structural regions in sequences *2* and *3*. Comparison of these regions shows that, in addition to similar location of the predicted secondary structural features, there are four sequence homologies that can be aligned with the help of two short deletions. These possibly homologous regions commence approx. 130 and 500 residues from the *N*-terminus of the peptide chain. This location is not, therefore, consistent with any simple explanation in terms of gene duplication. A firm evaluation of their significance must await the determination of the complete sequence.

Kretsinger (1976) has identified a Ca^{2+}-binding loop in a number of Ca^{2+}-regulatory proteins, which contains several aspartic acid or glutamic acid residues alternating with other residues in a characteristic manner. We have so far found no evidence for sequence of this type in the peptides derived from the ATPase, although there are a number of regions rich in carboxy groups. The absence of homology would not be unexpected, since the ATPase has two interacting sites for Ca^{2+} that function in a totally different manner from those in the regulatory proteins. It is also possible that the appropriate peptides, although likely to be acidic, are associated with the membrane and have not yet been isolated.

We thank Miss S. Lathwell for operating the amino acid analyser, Dr. D. H. MacLennan for making available to us the results of his sequence studies before publication and Dr. A. D. McLachlan for carrying out the analysis of secondary structure of the long peptides.

References

Allen, G. (1978) *FEBS Symp.* **45**, 159–168
Allen, G. (1980*a*) *Biochem. J.* **187**, 545–563
Allen, G. (1980*b*) *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) *Biochem. J.* **187**, 577–589
Ambler, R. P. (1965) *Biochem. J.* **96**, 32P
Capaldi, R. A. & Vanderkooi, G. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 930–932
Green, N. M., Allen, G., Hebdon, G. M. & Thorley-Lawson, D. A. (1977) in *Calcium Binding Proteins and Calcium Function* (Wasserman, R. H., Corradino, R. A., Carafoli, E., Kretsinger, R. H., MacLennan, D. H. & Siegel, F. L., eds.), pp. 164–172, Elsevier/North-Holland, Amsterdam
Green, N. M., Allen, G. & Hebdon, G. M. (1980) *Ann. N.Y. Acad. Sci.* in the press
Hardwicke, P. M. D. & Green, N. M. (1974) *Eur. J. Biochem.* **42**, 183–193
Klip, A. & MacLennan, D. H. (1978) in *Frontiers of Biological Energetics*, vol. 2. (Dutton, L., Leigh, J. & Scarpa, A., eds.), pp. 1137–1147, Academic Press, New York
Klip, A., Reithmeier, R. A. F. & MacLennan, D. H. (1980) *J. Biol. Chem.* in the press
Kostka, V. & Carpenter, F. H. (1964) *J. Biol. Chem.* **239**, 1799–1803
Kretsinger, R. H. (1976) *Annu. Rev. Biochem.* **45**, 239–266
Levine, M., Muirhead, H., Stammers, D. K. & Stuart, D. I. (1978) *Nature (London)* **271**, 621–630
McLachlan, A. D. (1978) *Int. J. Quantum Chem.* **12**, Suppl. **1**, 371–385
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
Migala, A., Agostini, B. & Hasselbach, W. (1973) *Z. Naturforsch.* **28**, 178–182
Murphy, A. J. (1976) *Biochemistry* **15**, 4492–4496
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) *J. Biol. Chem.* **252**, 8549–8553
Piszkiewicz, D., Landon, M. & Smith, E. L. (1970) *Biochem. Biophys. Res. Commun.* **40**, 1173–1178
Sarzała, M. G. & Michalak, M. (1978) *Biochem. Biophys. Acta* **513**, 221–235
Stewart, P. S. & MacLennan, D. H. (1974) *J. Biol. Chem.* **249**, 985–993
Stewart, P. S., MacLennan, D. H. & Shamoo, A. E. (1976) *J. Biol. Chem.* **251**, 712–719
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