Biochem. J. (1974) **139**, 1–10
Printed in Great Britain

The Tryptic Peptides of Rabbit Muscle Triose Phosphate Isomerase

By P. H. CORRAN and S. G. WALEY
Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, U.K.

(*Received* 20 *July* 1973)

1. The peptides obtained by tryptic digestion of *S*-[^{14}C]carboxymethylated rabbit muscle triose phosphate isomerase have been studied. 2. The first step in the fractionation of the tryptic digest was gel filtration on coupled columns of Sephadex G-25 and G-50. Further fractionation was carried out by paper electrophoresis and paper chromatography. 3. The digest contained 26 peptides and three free amino acids. The sizes of the peptides ranged from two to 29 residues. 4. The sequences of the peptides have been determined. 5. The length of the polypeptide chains is about 250 amino acid residues. 6. The variant sequences encountered were due to partial deamidation; this may be one of the reasons for multiple forms of the enzyme. 7. The chicken and rabbit enzymes are compared. 8. Detailed evidence for the sequences of the tryptic peptides has been deposited as Supplementary Publication SUP 50024 at the British Library, Lending Division (formerly the National Lending Library for Science and Technology), Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies can be obtained on the terms given in *Biochem. J.* (1973) **131**, 5.

Triose phosphate isomerase (EC 5.3.1.1.) is not only the most active, but also one of the smallest of the glycolytic enzymes. The molecular weight of the rabbit muscle enzyme has been estimated from ultracentrifuge data to lie in the range 50000–60000 (McVittie *et al.*, 1972; Krietsch *et al.*, 1970; Norton *et al.*, 1970). Crystallographic data suggested a value of about 53000; moreover the hexagonal crystals had a twofold axis of rotation that related two subunits of molecular weight 26200 (Johnson & Waley, 1967). The molecular weight of the polypeptide chains has been estimated as 26500 from polyacrylamide-gel electrophoresis in the presence of sodium dodecyl sulphate (Coulson *et al.*, 1970; Norton *et al.*, 1970). Structural work on the cysteine peptides is consistent with the presence of two identical polypeptide chains (Burton & Waley, 1968; Miller & Waley, 1971*a*). Triose phosphate isomerase from *Bacillus stearothermophilus* is also a dimer and the chains are much the same size as those of the muscle enzyme (Fahey *et al.*, 1971).

The present paper describes the sequences of the tryptic peptides obtained from the rabbit muscle enzyme. Dipeptidylaminopeptidase I (cathepsin C) (McDonald *et al.*, 1969; Lindley, 1972) has been found useful, especially in placing amide groups and in work on small peptides containing tryptophan.

The results now obtained, together with the work on the cysteine peptides (Miller & Waley, 1971*a*), establish the sequences of all the tryptic peptides, and show that the polypeptide chains comprise about 250 amino acid residues. Moreover, these results permit a detailed comparison with the work on the tryptic peptides of the enzyme from chicken muscle (see accompanying paper, Furth *et al.*, 1974). We were greatly helped by knowing their sequences for the shorter tryptic peptides from the chicken enzyme.

Materials and Methods

Enzymes

Rabbit muscle triose phosphate isomerase was obtained from Boehringer Corp. (London) Ltd., London W.5., U.K. Dipeptidylaminopeptidase I was a gift from Dr. J. K. McDonald (Ames Research Centre, National Aeronautics and Space Administration, Calif., U.S.A.). A portion (10–20μl) of dipeptidylaminopeptidase (3.7mg/ml in 1% NaCl–0.1mM-EDTA) was added to about 100nmol of peptide in 100μl of buffer; the buffer was 16mM-HCl–0.8% pyridine, adjusted to pH5 with acetic acid; 1.4mg of dithiothreitol was added to 50ml of the buffer. Digestion was carried out overnight at 37°C. The sources of the other enzymes and the conditions under which they were used are given by Miller & Waley (1971*a*).

Purification of peptides

Chromatography on Sephadex G-25 and G-50 was carried out as described below (Fig. 1). Paper chromatography was carried out in solvent BAWP [butan-1-ol–acetic acid–water–pyridine (15:3:12:10, by vol.) (Waley & Watson, 1953)] and paper electrophoresis in buffers at pH1.8 [20% (v/v) in acetic acid and 2% (v/v) in formic acid], at pH3.5 [pyridine–acetic acid–water (1:10:135, by vol.)], or pH6.5 [pyridine–acetic acid–water (25:1:225,

by vol.)]. The mobilities at pH6.5 are relative to aspartic acid ($m = -1$), and were used to calculate charges as described by Offord (1966). Peptides were detected and peptide 'maps' were prepared as previously described (Miller & Waley, 1971*a*).

Sequence methods

The dansyl–Edman method, with the peptide distributed over several tubes (Gray & Smith, 1970; Gray, 1967*a*,*b*), was used; in some cases, the subtractive method was used concurrently: three-quarters of the material in one tube after each cycle was hydrolysed for amino acid analysis, and one-quarter was used for *N*-terminal analysis by the 'dansyl' technique. The Dns-amino acids were identified by t.l.c. on polyamide sheets (5cm square; Hartley, 1970). The solvent pyridine–acetic acid–water–ethanol (9:16:1000:30, by vol.) (Jörnvall, 1970) was useful for the identification of Dns-histidine.

Amino acid analysis

Hydrolysis was carried out as described previously (Miller & Waley, 1971*a*), and the hydrolysates analysed on a Bio Cal 200 or Bio Cal 100 analyser, equipped with a high-sensitivity colorimeter and linearized output with scale expansion. A one-column system was used.

Qualitative analysis of the hydrolysates of dipeptides was carried out by electrophoresis at pH1.8.

Peptide nomenclature

The tryptic peptides are given the prefix T; the cysteine peptides obtained previously from the tryptic digest were numbered 1–5 and would now be called T1, T2, etc. Hence the numbering in this paper starts at peptide T6. Further digestion of the tryptic peptides with chymotrypsin, pepsin, subtilisin, thermolysin or dipeptidylaminopeptidase I gave peptides denoted by C, P, S, Th or D respectively. Thus peptide T6 P1 refers to a peptic peptide obtained from the tryptic peptide T6.

Results and Discussion

Fractionation of tryptic digest

The peptide 'maps' (Fig. 6 of Miller & Waley, 1971*a*) suggested that there were about 30 products in the tryptic digest of the carboxymethylated protein. 'Maps' were stained with dilute ninhydrin and the peptides eluted and hydrolysed to gain familiarity with the nature of the peptides and to allocate them to the three CNBr fragments. A few of the simpler peptides were prepared from 'maps'.

On a larger scale, the first step in the fractionation of the tryptic digest of *S*-[^{14}C]carboxymethylated protein (Miller & Waley, 1971*a*) was either ion-exchange chromatography or gel filtration on Sephadex columns. The next steps were paper electrophoresis at pH6.5, 3.5 or 1.8, or paper chromatography in solvent BAWP; one to three 'paper steps' were generally required to obtain the peptides pure. The ion-exchange chromatography on Dowex 50 was carried out as described by Miller & Waley (1971*a*), and the gel filtration was on coupled columns; a column of Sephadex G-50 and one of Sephadex G-25 were used in series. The Sephadex columns were eluted with NH_4HCO_3 (5g/l) in 10% (v/v) propan-2-ol. This medium was chosen because peptides can be detected by measuring E_{220} and the salt can be removed by freeze-drying; the propan-2-ol prevents bacterial contamination and may help to decrease adsorption of aromatic peptides. The resolution was satisfactory (Fig. 1); in particular, the largest tryptic peptide (peptide T11, containing 29 residues) could be obtained pure directly. Normally, four fractions (of 10ml) were pooled for further work.

Ion-exchange chromatography has some advantages in that a greater degree of purification can usually be obtained in one step, and it was especially convenient in the previous work (Miller & Waley, 1971*a*) for the isolation of the readily detectable radioactive *S*-carboxymethyl peptides. On the other hand, in the present work the goal was the isolation of all the tryptic peptides, and here the use of the Sephadex columns as a first step was preferred because the danger of missing irreversibly adsorbed peptides is minimized.

The steps used in the purification of each peptide are given in Table 1; the five peptides containing cysteine (Miller & Waley, 1971*a*) are omitted from Table 1. The dipeptide, lysyl-lysine and the free amino acids, lysine, arginine and glutamine, are discussed below. Table 1 shows the composition of 20 tryptic peptides and these are now discussed individually in the order in which they are believed to be in the chain (Corran & Waley, 1973). The total number of amino acids in the tryptic peptides agrees with the values expected from the composition of the protein (Table 2); further discussion of the composition is deferred until the work on the overlaps of the tryptic peptides is described.

Sequences of peptides

Peptide T17. The sequence of this basic tetrapeptide was established by the dansyl–Edman procedure (Fig. 2).

Peptide T22. This was a basic peptide, and hence asparagine rather than aspartic acid was present, and it contained tryptophan. It is assumed that there

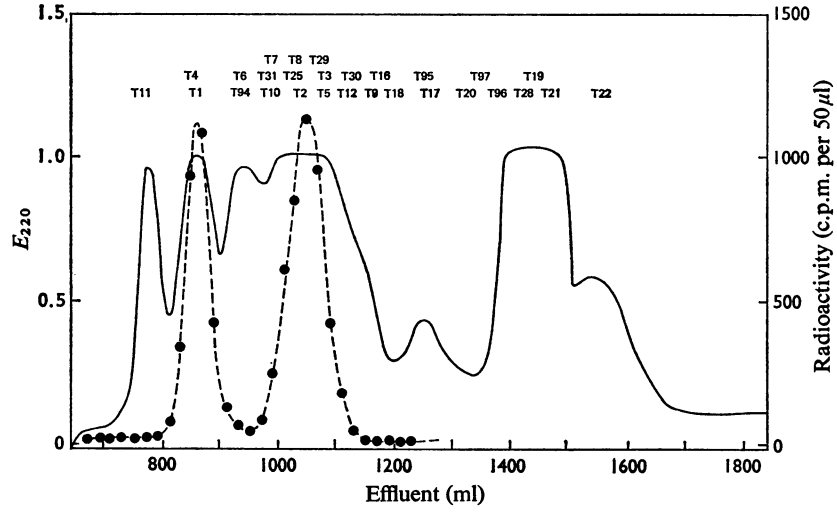


Fig. 1. *Chromatography of tryptic digest of carboxymethylated triose phosphate isomerase*

A column (145cm long × 2.5cm diam.) of Sephadex G-50 was coupled to a column (150cm long × 2.5cm diam.) of Sephadex G-25. Elution with NH_4HCO_3 (5g/l) in 10% (v/v) propan-2-ol was carried out at 15ml/h, and fractions of 10ml were collected. The excluded volume was 530ml; Phenol Red was eluted at 1940ml. Measurements were made of E_{220} (——) and of radioactivity (●) (c.p.m. of 50µl portion); the protein had been alkylated with sodium iodo[2-^{14}C]acetate (Miller & Waley, 1971*a*).

is only one residue of tryptophan. The sequence was determined by the dansyl–Edman method and by carboxypeptidase A (Fig. 2); after the reaction with carboxypeptidase A, electrophoresis at pH3.5 showed that lysine was released; the region where neutral amino acids run to was sewn in and run at pH1.8: tryptophan was present.

Peptide T16. The sequence of this, the smaller of the two peptides that contain methionine, is given in Fig. 2. The isolated peptide was neutral and thus contains aspartic acid, but 'maps' of the tryptic digest of the CNBr fragment M (Miller & Waley, 1971*a*), after staining with dilute ninhydrin, elution and hydrolysis, gave a basic peptide with the composition: aspartic acid (0.8), glycine (1.1), arginine (1.1). This peptide was not present in a digest of the whole protein. This suggests there is an asparagine residue in the protein and that the isolated neutral peptide has suffered deamidation; this tends to happen with peptides containing the ...Asn-Gly... sequence (see also peptide T94).

Peptide T95. This was a basic peptide with high mobility on electrophoresis at pH3.5 that gave only lysine after hydrolysis (identified by electrophoresis at pH1.8). The peptide had an R_F value of 0.66 relative to lysine on chromatography in solvent BAWP, and this (Waley & Watson, 1953) shows the peptide to be Lys-Lys rather than any higher oligomer.

Peptide T18. This was obtained from the Sephadex fractionation, and was shown by the dansyl–Edman method to be Ala-Ala-Lys; later work has shown that this peptide is not unique, but derives from peptide T31.

Peptide T31. The structure of this peptide was established by cleavage with pepsin, and study of the resulting peptides by the dansyl–Edman method (Fig. 3). Peptide T31 has zero mobility at pH 6.5, and so there are two amide groups. One is in the *C*-terminal peptic peptide (T31 P3), as this is basic. The other amide group is at the *N*-terminus, as the electrophoretic mobility at pH6.5 of the dansyl peptide T31 P1 was only slightly different from that of the dansyl derivative of the peptide remaining after one cycle of Edman degradation. The *C*-terminal moiety of peptide T31 represents peptide T18.

Peptide T12. This is an 'overlap' peptide, comprising peptides T13 and T26. The mobility of peptide T12 indicated the presence of one amide group, which must be in the *N*-terminal position as peptide T13 is basic and peptide T26 is neutral at pH6.5. The structure of these peptides is given in Fig. 4; they are present in CNBr fragment M. Glutamine at the *N*-terminus of peptides tends to cyclize to pyrro-

Table 1. *Amino acid composition of tryptic peptides from carboxymethylated triose phosphate isomerase*

The amino acids compositions are given as molar ratios, derived from hydrolysis for 20h, and usually also for 96h if the peptide contained valine or isoleucine. The methods of purification, which are described in the text, are abbreviated thus: 3.5, 6.5, electrophoresis at pH 3.5 or 6.5 respectively; B, paper chromatography in solvent BAWP; G, chromatography on coupled columns of Sephadex G-25 and G-50; 50, chromatography on Dowex 50; 'Map' refers to peptide 'maps' (see the text). The electrophoretic mobilities (*m*) refer to pH 6.5 and are relative to aspartic acid ($m = -1.00$). Peptides containing tryptophan are assumed to contain one residue of tryptophan. Peptides containing carboxymethylcysteine have been described previously (Miller & Waley, 1971*a*) and are not given here.

Amino acid / Peptide ..	T6	T7	T8	T9	T10	T11	T12	T13	T16	T17	T18	T20	T21	T22	T25	T26	T28	T29	T30	T31	T94
Asp				1.0	1.1	4.0	1.0		1.0					1.0	2.1	1.3	1.0	2.2		1.9	1.1
Thr	1.8	1.1	0.8												1.0				0.9	1.8	1.9
Ser					0.8	1.7				0.9					2.5		0.9				0.8
Glu	4.9	1.0	2.1	1.2	2.9	3.3	1.0	0.9							1.1				2.1	1.1	1.2
Pro	1.1	1.0				2.0	1.1			1.0						1.2					1.0
Gly		2.1	1.0		2.1	3.0			1.0				0.9	2.0						1.2	2.9
Ala	2.1	1.9	1.0			3.1				1.1	1.9				2.1			0.9		2.0	1.1
Val	0.9	3.2			0.9	3.1								1.2	2.1			2.1	1.9		1.1
Met									0.6												0.9
Ile		1.0	1.0		1.2	1.8												0.8		1.0	2.1
Leu		1.0		0.9	1.0	3.1	0.8					1.0	0.9			0.6				3.0	
Tyr		0.8																			
Phe					1.0	2.0								1.9					1.0		1.0
His	1.0				0.8																
Lys	1.1	1.0	1.0		1.0	2.0	2.1	1.1			1.1		1.1	1.0		0.9	1.2	1.0	1.1	1.0	1.0
Arg				1.0					1.0	1.0		1.0			1.1						
Trp		(1)											(1)	(1)			(1)				
Total no. of residues	13	15	7	4	13	29	6	2	4	4	3	2	4	8	12	4	4	7	7	13	16
m	0	0	−0.32	−0.33	−0.29	−0.38	+0.3	+0.54	0	+0.4	+0.56	+0.55	+0.4	+0.3	0	0	0	0	0	0	0
Method of purification	50, B	G, 3.5, B	50; or G, B	'Map'	50, B, 3.5	G	G, B	'Map'	G, 6.5	G, 6.5	'Map'	50	G, B	G, 6.5	G, 3.5	'Map'	50, B, 3.5	50, B	50, B	G, B, 3.5	50, B, 3.5

Table 2. *Amino acid composition of rabbit muscle triose phosphate isomerase*

The values in column (1) are from Norton *et al.* (1970), in column (2) from Krietsch *et al.* (1971) for the main component (α) and in column (3) from Miller & Waley (1971*a*). Each of these has been recalculated on the basis of a total of 248 residues. The values in column (4) are derived from the sum of the tryptic peptides described in the present paper and in Miller & Waley (1971*a*); peptide T5 in Miller & Waley (1971*a*) is now known to have histidine but not lysine. All values are given to the nearest integer.

Amino acid	Column (1)	Column (2)	Column (3)	Column (4)
Asp	21	21	21	21
Thr	15	14	16	15
Ser	12	12	14	12
Glu	29	29	27	27
Pro	9	11	10	10
Gly	25	23	24	24
Ala	29	27	26	27
Val	23	25	24	25
Ile	13	14	14	15
Leu	15	14	15	15
Met	2	2	2	2
Tyr	4	4	4	4
Phe	8	8	8	8
His	4	5	4	4
Lys	21	21	22	21
Arg	8	8	8	8
Cys	5	5	5	5
Trp	5	5	4	5

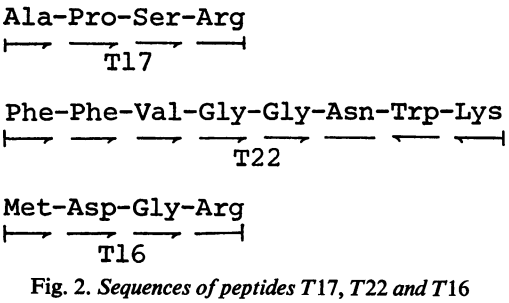


Fig. 2. *Sequences of peptides T17, T22 and T16*

In this and subsequent figures, sequences obtained by the dansyl–Edman method are indicated thus: $|\rightarrow \rightarrow$ etc.; amino acids released by carboxypeptidase A are indicated thus: $\leftarrow$, $\leftarrow|$; residues known to be present from quantitative amino acid analysis are indicated thus :—. T, trypsin.

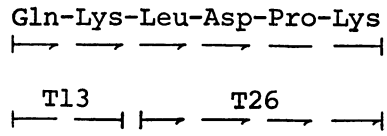


Fig. 4. *Structure of peptide T12*

The symbols are as in Fig. 2. T, trypsin.

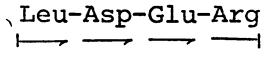
lidone carboxylic acid and small amounts of the cyclized form of peptide T12 were probably present.

Peptide T94. This is the larger of the two peptides containing methionine, and it bridges CNBr fragments M and R (Miller & Waley, 1971*a*). The structure was deduced from the peptides obtained by cleavage with thermolysin and with pepsin (Fig. 5).

Peptide T94 was isolated in both neutral and acidic forms; this was due to the third residue being either asparagine or aspartic acid: the Asn-Gly sequence is particularly prone to hydrolysis of the side-chain amide. The residue is shown as asparagine in Fig. 5, since we take it that the parent form is the amide.

Peptide T10. The structure of this peptide (Fig. 6) was studied by limited digestion with pepsin; two peptides (T10 P2 and T10 P3), which together accounted for all the amino acid residues, were obtained in 70 and 80% yield respectively, together with the tripeptide (T10 P1, 20% yield), and a variant of peptide T10 P3 obtained in 7% yield. Degradation with dipeptidylaminopeptidase gave three dipeptides (Fig. 6) and two other peptides, the larger of which has extra residues of aspartic acid and glutamic acid; apparently these acidic residues brought the digestion by the peptidase to a halt. The variant of peptide T10 P3 apparently has glutamine replaced by glutamic acid; we do not know whether some deamidation accompanied isolation of the peptide or whether there is a variant residue in the protein at this position.

Peptide T9. This peptide was eluted from the 'map' of the tryptic digest of carboxymethylated protein. The peptide is acidic (Table 1), and so amides are absent. The structure was found by the dansyl–Edman method to be:



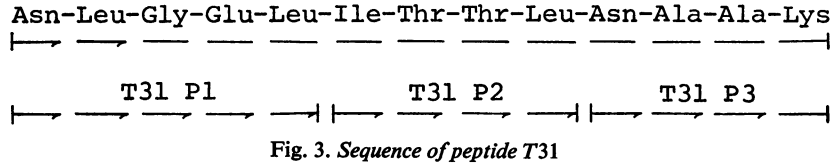


Fig. 3. *Sequence of peptide T31*

The symbols are as in Fig. 2. P, pepsin; T, trypsin.

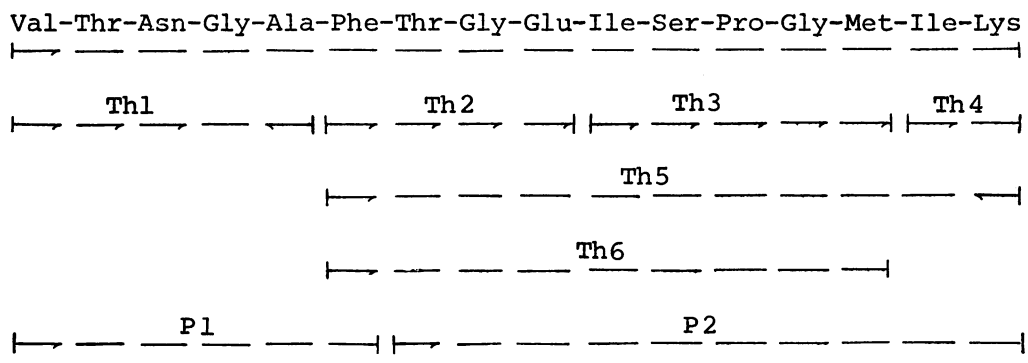


Fig. 5. *Sequence of peptide T*94

The symbols are as in Fig. 2. Th, thermolysin; P, pepsin.

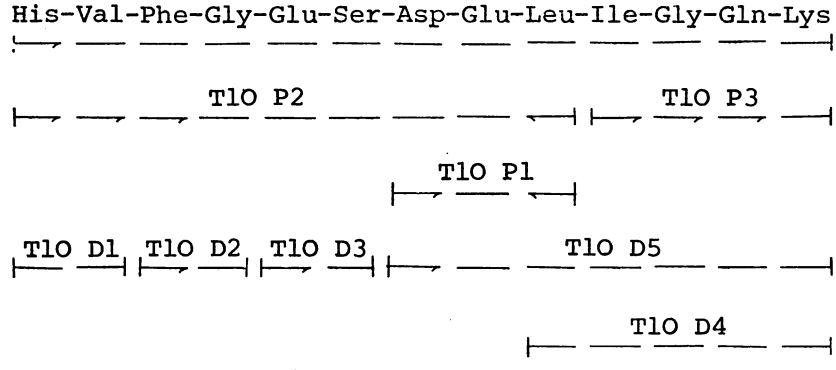


Fig. 6. *Sequence of peptide T*10

The symbols are as in Fig. 2. D denotes peptides obtained with dipeptidylaminopeptidase; T, trypsin; P, pepsin.

*Peptide T*8. This was an acidic heptapeptide; amides were absent. The structure was established by digestion with subtilisin and with thermolysin, supplemented by dansyl–Edman degradation of the parent peptide (Fig. 7). Peptide T8 is derived from the *C*-terminal CNBr fragment R.

*Peptide T*30. The structure of this neutral heptapeptide was studied by the dansyl–Edman method, which led to the sequence: Val-Val-Phe-Glx-Glx-Thr-Lys. This structure was confirmed and the position of the amide established, by cleavage with dipeptidylaminopeptidase. Fractionation of the digest by electrophoresis at pH 6.5 gave an acidic dipeptide containing glutamic acid and phenylalanine, and a basic product ($m = 0.97$) which was free lysine. The fraction neutral at pH 6.5 was fractionated in solvent BAWP and gave a dipeptide with R_F 0.76 giving valine after hydrolysis, and another (R_F 0.27) giving glutamic acid and threonine after hydrolysis. These results lead to the following sequence for peptide T30:

Val-Val-Phe-Glu-Gln-Thr-Lys

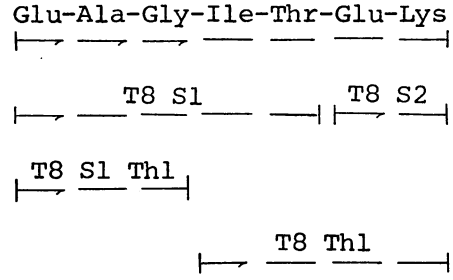


Fig. 7. *Sequence of peptide T*8

The symbols are as in Fig. 2. S, subtilisin; Th, thermolysin; T, trypsin.

*Peptide T*29. The structure of this neutral peptide was established by cleavage with dipeptidylaminopeptidase. The first two products were isolated after electrophoresis at pH 6.5. One product was a basic tripeptide (mobility +0.43, orange colour with the cadmium–ninhydrin reagent), whose composition expressed as molar ratios was: Asp (0.9), Val (0.9), Lys (1.1); this peptide had *N*-terminal Asx, and from

the specificity of trypsin the structure of this tripeptide is: Asn-Val-Lys. Another product was acidic (mobility –0.77), had *N*-terminal alanine and gave comparable amounts of aspartic acid and alanine on hydrolysis; thus this dipeptide is Ala-Asp. The third product (purified by electrophoresis at pH1.9 and chromatography in solvent BAWP by the 'map' procedure) had the composition: Val (0.7), Ile (1.3) and had *N*-terminal valine, as did peptide T29. The sequence deduced from these results is:

Val-Ile-Ala-Asp-Asn-Val-Lys

Peptide T28. This peptide contained tryptophan, and if one residue of tryptophan was present, peptide T28 is a tetrapeptide; it has zero mobility at pH6.5, so amides are absent. Digestion with dipeptidylaminopeptidase gave an acidic peptide that was Ehrlich-positive and contained aspartic acid, and a basic peptide that contained serine and lysine. Peptide T28 had *N*-terminal aspartic acid and hence the structure is deduced to be:

Asp-Trp-Ser-Lys

Peptide T7. This is the tryptic peptide that contains the active site (Hartman, 1971; Miller & Waley, 1971*b*), and it is most easily isolated after gel filtration on Sephadex by chromatography in solvent BAWP, as it has a high R_F value (0.75). Digestion with pepsin gave the three peptides shown in Table 3; the structure of the first, peptide T7 P1, has been described previously (Miller & Waley, 1971*b*). This was the only one of the peptic peptides that gave a positive Ehrlich reaction and it contains one residue of tryptophan (Miller & Waley, 1971*b*), and so there is only one residue of tryptophan in peptide T7. The structure inferred for peptide T7 is shown in Fig. 8.

Table 3. *Peptides obtained by digestion of peptide T7 with pepsin*

Peptide T7 (200nmol) was treated with 5μg of pepsin in 60μl of 5% (v/v) formic acid for 5h at 37°C and the digest fractionated by electrophoresis at pH6.5. The *N*-terminal amino acid is shown in italics.

Peptide	Amino acid composition (molar ratios)	Mobility at pH6.5
T7 P1*	Glu (1.0), Pro (1.2), *Ala* (1.1), Val (0.9), Tyr (0.9)	–0.26
T7 P2	*Val* (1.95), Leu (1.05)	0
T7 P3	Thr (0.9), Gly (2.0), *Ala* (1.0), Ile (1.0), Lys (1.1)	0.38

* This peptide gave a positive Ehrlich reaction.

The sequence proposed by Hartman (1971) differs from that given in Fig. 8 in having the first residue as tryptophan instead of valine. The analysis given in Table 1 is based on hydrolysis for 24, 48 and 72h; the value for valine after 24h is only 2.4 residues, essentially the same value as that found by Hartman (1971) (2.3), but the longer times of hydrolysis show that there are three residues of valine in the peptide. It was the absence of alteration in the results of amino acid analysis after the first cycle of Edman degradation that led Hartman (1971) to place tryptophan at the *N*-terminus. The low values from the slowly hydrolysed valylvaline... sequence would thus account for part of Hartman's (1971) findings. We believe that Hartman's (1971) value of two residues of tryptophan (which was based on the E_{280}) is incorrect.

Peptide T6. This peptide was obtained from ion-exchange chromatography; the analysis in Table 1 is based on three 20h hydrolysates, and one 96h one. Digestion with subtilisin gave six peptides whose composition and mobilities were helpful in placing the amide groups; digestion with pepsin gave two peptides. The sequence of peptide T6 is shown in Fig. 9.

Peptide T20. This is a basic dipeptide leucylarginine (Fig. 10).

Peptide T21. This peptide contained tryptophan, and if one residue of tryptophan was present there were four residues altogether. The results from digestion with dipeptidylaminopeptidase lead to the structure shown in Fig. 10.

Peptide T25. This peptide was obtained from ion-exchange chromatography, and also from the Sephadex fractionation; the analysis (Table 1) left it uncertain whether there were two or three residues of serine. The sequence demands three residues of serine. Degradation of the original peptide, and study of the peptides obtained by digestion with thermolysin and with pepsin led to the sequence shown in Fig 11. The peptide is present in CNBr fragment R.

Peptide T11. This was the largest tryptic peptide, and was obtained pure directly from the fractionation on Sephadex; peptide T11 has 29 residues, and the next largest tryptic peptide has 20 residues (peptide 1 of Miller & Waley, 1971*a*). Hydrolysis for 96h

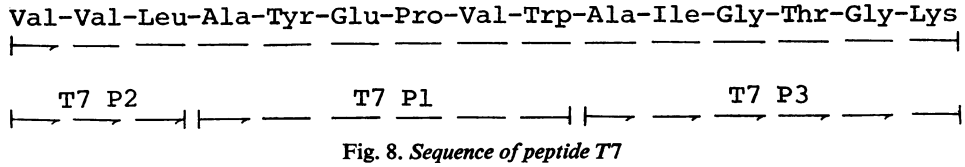


Fig. 8. *Sequence of peptide T7*

The symbols are as in Fig. 2. T, trypsin; P, pepsin.

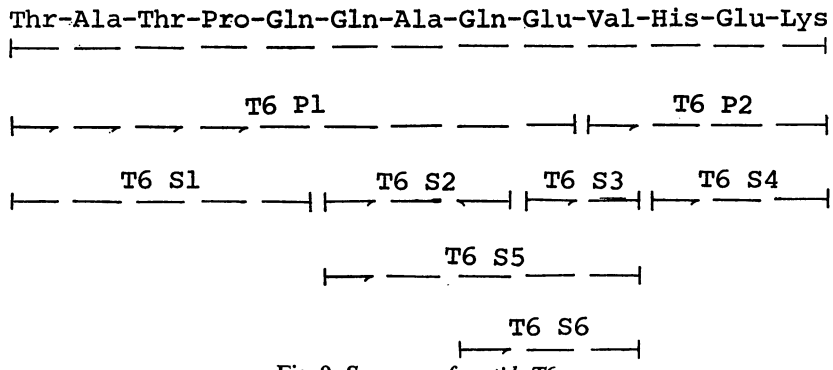


Fig. 9. *Sequence of peptide T6*

The symbols are as in Fig. 2. T, trypsin; P, pepsin; S, subtilisin.

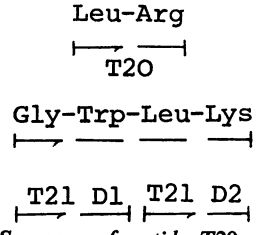


Fig. 10. *Sequences of peptides T20 and T21*

The symbols are as in Fig. 2. T, trypsin; D, dipeptidyl-aminopeptidase.

confirmed that there were three residues of valine and two residues of isoleucine, and the work described below on the sequence confirmed that there were two residues of serine. The mobility of peptide T11 at pH 6.5 suggested the presence of one or two amide groups, and six or five carboxyl groups.

Digestions with pepsin, thermolysin and subtilisin were used to obtain information about the sequence (Fig. 12); the evidence that this is the *C*-terminal tryptic peptide is now outlined. Limited digestion with carboxypeptidase A (various amounts) and amino acid analysis led to the sequence: (Asp, Phe, Val) (Ile, 'Ser') (Ile, Ala, Lys); the last trio of amino acids was fully released even with the smallest amount of enzyme, when the first trio was barely present. Here 'Ser' refers to a peak in the position of serine, but which is asparagine. Peptides T11 P4, T11 Th6 and T11 Th7 (Fig. 12) establish the sequence of the last seven residues of peptide T11 and explain the results with carboxypeptidase. Timed digestions with carboxypeptidase A led Norton *et al.* (1970) to put forward as the *C*-terminal sequence of the whole protein: (Val, Asp, Phe, Ile)Asn-Ile-Ala-Lys-Gln. Since there are two residues of isoleucine in this region, it is difficult to deduce a reliable sequence from such experiments; nevertheless, the results suggest that peptide T11 is the *C*-terminal tryptic peptide.

Free amino acids in the tryptic digest. Lysine, arginine and glutamine were present in the tryptic digest; they were obtained from the later fractions from the Sephadex fractionation. Glutamine (which had the low mobility on electrophoresis at pH 3.5 characteristic of free amino acids, and was neutral at pH 6.5) is the *C*-terminal amino acid of the protein (Burton & Waley, 1968; Norton *et al.*, 1970; Miller & Waley, 1971*a*).

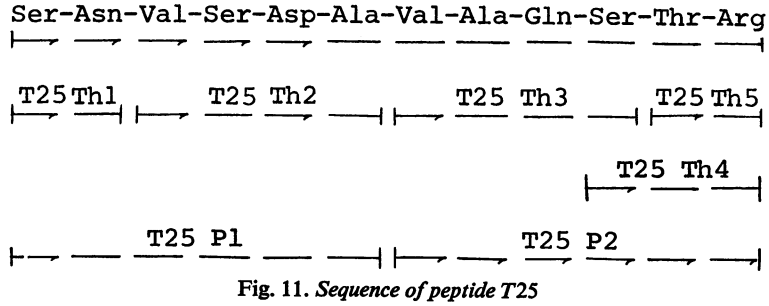


Fig. 11. *Sequence of peptide T25*

The symbols are as in Fig. 2. T, trypsin; Th, thermolysin; P, pepsin.

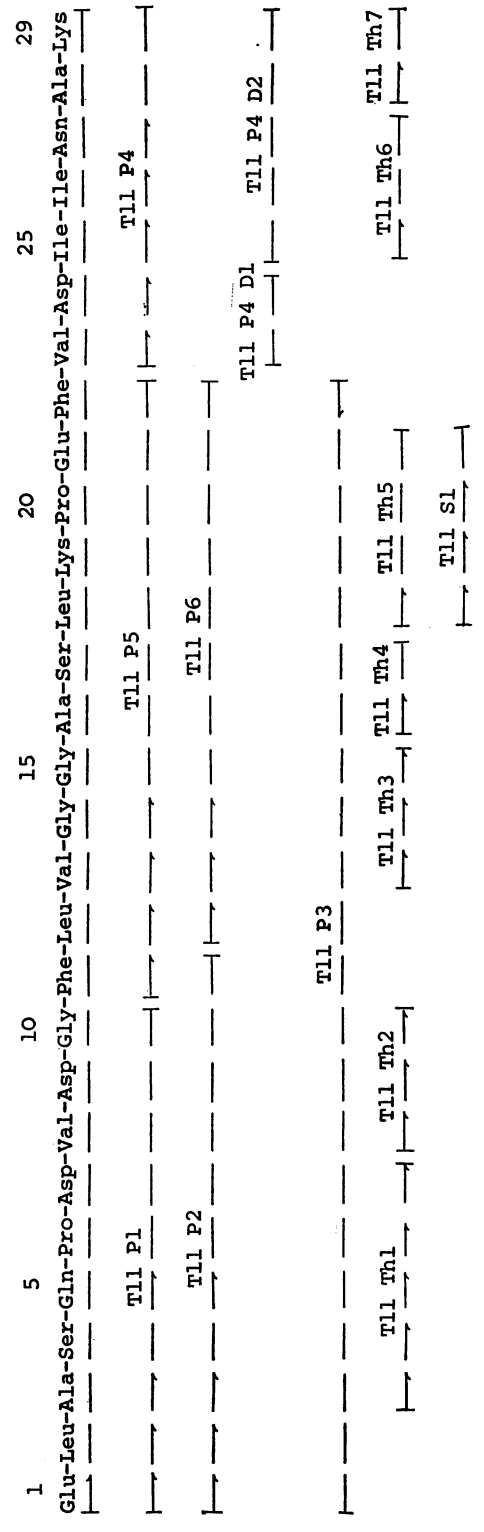


Fig. 12. *Sequence of peptide T*11

The symbols are as in Fig. 2. T, trypsin; P, pepsin; S, subtilisin; Th, thermolysin; D, dipeptidylaminopeptidase.

Variant sequences

Variant forms of three of the tryptic peptides were encountered; these were due to partial deamidation of asparaginyl or glutaminyl residues. Peptides T16 and T94 contain an ...Asn–Gly... sequence and deamidation occurs particularly readily here (Shotton & Hartley, 1970), but in peptide T10 it was a glutaminyl residue that was deamidated.

Partial deamidation may occur during isolation of the peptides; on the other hand, there may be variant forms of the parent protein corresponding to the amidated and deamidated forms of the peptides. The variant forms (isoenzymes) of rabbit muscle triose phosphate isomerase that have been isolated seem to differ in their amino acid compositions (Krietsch *et al.*, 1971), although they do not differ in their catalytic properties (Norton *et al.*, 1970; Krietsch *et al.*, 1971; Lee *et al.*, 1971; Rozacky *et al.*, 1971). These isoenzymes are dimers of the types AA, AB and BB, and it seems possible that part of the differences between the chains may be attributed to the state of amidation of some of the carboxyl groups. Rabbit muscle aldolase contains slightly different chains, owing to the state of amidation of an aspartic acid residue (Lai *et al.*, 1970). Bovine lens α-crystallin contains chains that differ in the replacement of glutamine by glutamic acid (Bloemendal *et al.*, 1972); there is evidence for slow deamidation *in vivo* for both these proteins (Midelfort & Mehler, 1972; Delcour & Papaconstantinou, 1972).

Comparison of rabbit and chicken enzymes

The accompanying paper (Furth *et al.*, 1974) describes the sequences of the tryptic peptides from the chicken muscle enzyme and arranges them in a hypothetical linear order by comparison with the sequence of the rabbit muscle enzyme (Corran & Waley, 1973). We now briefly comment on the extent of the homology and on some of the postulated interchanges. The numbering is from the sequence of Corran & Waley (1973).

Each chain has 248 amino acid residues, but there is one deletion in each chain; presumably an ancestral sequence had 250 residues. Serine-3 in the rabbit enzyme lacks a counterpart in the chicken enzyme, and there is a residue of lysine in the chicken enzyme that intervenes between glycine-16 and

arginine-17 in the rabbit enzyme. There are, apart from these gaps, 32 differences between the chains: the extent of identity is 86%. Fortunately, hardly any of the interchanges involve arginine or lysine, and so there is almost complete correspondence between the tryptic peptides of the two enzymes. Most of the interchanges are of the conservative type commonly found when homologous proteins are compared. Thus 22 of the 32 interchanges belong to the class of interchanges most frequently found in protein families (Dayhoff, 1972), and are also consistent with a change of one nucleotide in the codons. Another four interchanges belong to the next commonest class of interchanges. The remaining six interchanges are listed below, with the residue in the chicken enzyme being given first: histidine-26 → threonine; isoleucine-46 → alanine; proline-70 → threonine; isoleucine-86 → cysteine; valine-202 → glutamine; histidine-224 → proline. None of these six interchanges occurs in the 'middle' stretch of the chain, and indeed there are few interchanges of any kind in this region (see Fig. 4 of Furth *et al.*, 1974), which includes the active-site-residue glutamic acid-165. Whether there is truly a highly conserved region will become clearer when further sequences are compared. The interchanges at positions 26 and 224 are of the 'from histidine' type; the chicken enzyme has eight residues of histidine to the rabbit enzyme's four. Similarly, the interchanges at positions 26 and 70 are of the 'to threonine' type; there are altogether six interchanges of this type. There is one more residue of cysteine in the rabbit enzyme; this is replaced by isoleucine in the chicken enzyme.

The support of the Medical Research Council is gratefully acknowledged, as is the technical assistance of Mrs. C. Moss and Miss R. Sammons. This paper is a contribution from the Oxford Enzyme Group, which is supported by the Science Research Council.

References

Bloemendal, H., Berns, A. J. M., Van der Oudera, F. & de Jong, W. W. W. (1972) *Exp. Eye Res.* **14**, 80–81

Burton, P. M. & Waley, S. G. (1968) *Biochem. J.* **107**, 737–744

Corran, P. H. & Waley, S. G. (1973) *FEBS Lett.* **30**, 97–99

Coulson, A. F. W., Knowles, J. R. & Offord, R. E. (1970) *Chem. Commun.* 7

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, vol. 5, p. 97, National Biomedical Research Foundation, Washington, D. C.

Delcour, J. & Papaconstantinou, J. (1972) *J. Biol. Chem.* **247**, 3289–3295

Fahey, R. C., Kolb, E. & Harris, J. I. (1971) *Biochem. J.* **124**, 77P

Furth, A. J., Millman, J. D., Priddle, J. D. & Offord, R. E. (1974) *Biochem. J.* **139**, 11–25

Gray, W. R. (1967*a*) *Methods Enzymol.* **11**, 139–151

Gray, W. R. (1967*b*) *Methods Enzymol.* **11**, 469–475

Gray, W. R. & Smith, J. F. (1970) *Anal. Biochem.* **33**, 36–42

Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822

Hartman, F. C. (1971) *Biochemistry* **10**, 146–154

Johnson, L. N. & Waley, S. G. (1967) *J. Mol. Biol.* **29**, 321–322

Jörnvall, M. (1970) *Eur. J. Biochem.* **14**, 521–534

Krietsch, W. K. G., Pentchev, P. G., Klingenburg, H., Hofstatter, T. & Bucher, T. (1970) *Eur. J. Biochem.* **14**, 289–300

Krietsch, W. K. G., Pentchev, P. G. & Klingenburg, H. (1971) *Eur. J. Biochem.* **23**, 77–85

Lai, C. Y., Chen, C. & Horecker, B. L. (1970) *Biochem. Biophys. Res. Commun.* **40**, 461–468

Lee, E. W., Barriso, J. A., Pepe, M. & Snyder, R. (1971) *Biochim. Biophys. Acta* **242**, 261–267

Lindley, H. (1972) *Biochem. J.* **126**, 683–688

McDonald, J. K., Zeitman, B. B., Reilly, T. J. & Ellis, S. (1969) *J. Biol. Chem.* **244**, 2693–2709

McVittie, J. D., Esnouf, M. P. & Peacocke, A. R. (1972) *Eur. J. Biochem.* **29**, 67–73

Midelfort, C. F. & Mehler, A. H. (1972) *Proc. Nat. Acad. Sci. U.S.* **69**, 1816–1819

Miller, J. C. & Waley, S. G. (1971*a*) *Biochem. J.* **122**, 209–218

Miller, J. C. & Waley, S. G. (1971*b*) *Biochem. J.* **123**, 163–170

Norton, I. L., Pfuderer, P., Stringer, C. D. & Hartman, F. C. (1970) *Biochemistry* **9**, 4952–4958

Offord, R. E. (1966) *Nature (London)* **211**, 591–593

Rozacky, E. E., Sawyer, T. H., Barton, R. A. & Gracy, R. W. (1971) *Arch. Biochem. Biophys.* **146**, 312–320

Shotton, D. M. & Hartley, B. S. (1970) *Nature (London)* **225**, 802–806

Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337