Biochem. J. (1974) **139**, 11–25
Printed in Great Britain

Studies on the Subunit Structure and Amino Acid Sequence of Triose Phosphate Isomerase from Chicken Breast Muscle

By ANNA J. FURTH*, J. D. MILMAN, J. D. PRIDDLE and R. E. OFFORD
Laboratory of Molecular Biophysics, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

(*Received* 24 *July* 1973)

1. Triose phosphate isomerase was prepared by chromatography on DEAE-cellulose of an $(NH_4)_2SO_4$ fraction of an extract of homogenized chicken breast muscle. The product is homogeneous on gel electrophoresis and is suitable for growing crystals for X-ray work. The specific activity is 10000 units/mg and the value for $E^{0.1\%}_{280}$ is 1.20. 2. Comparison between the sum of the amino acid compositions of the tryptic peptides of the protein and the amino acid composition obtained on total hydrolysis of the protein indicates that the relative subunit mass is about 27000. 3. These data, together with the results of the examination of the amino acid compositions of a number of minor peptides, the number of peptides in the tryptic digest and the complete amino acid sequences of the tryptic peptides (the determination of which is described here), give no indication that the subunits are dissimilar. 4. A tentative amino acid sequence is presented for the protein, in which the ordering of the tryptic peptides is derived by homology with the sequence of the rabbit muscle enzyme (Corran & Waley, 1973). 5. An appendix describes the use that was made of mass spectrometry in the determination of some of the sequences. Mass-spectrometric data have been obtained for 35 residues, that is about 15% of the total sequence of the protein. 6. An extended version of the present paper has been deposited as Supplementary Publication SUP 50025 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 may be obtained on the terms given in *Biochem. J.* (1973) **131**, 5.

Triose phosphate isomerase (EC 5.3.1.1) was until recently little studied. However, the superficial simplicity of the elementary reactions of catalysis and its small size among glycolytic enzymes has now led to efforts in many laboratories to elucidate the catalytic mechanism and covalent structure (for reviews of this work see Banner *et al.*, 1971; Knowles *et al.*, 1971).

The present paper describes experiments that indicate that the chicken enzyme is a dimer of subunits having about 250 amino acid residues each. Nothing has been found to suggest that the subunits are dissimilar in amino acid sequence. The principal line of evidence has been obtained by isolating the individual tryptic peptides and determining whether they represent unique sequences. This is a prerequisite to the determination of the complete amino acid sequence of the enzyme, but the work has been planned in the hope that high-resolution electron-density data will become available in this laboratory to obviate conventional overlapping by isolation of peptides from digests with proteolytic enzymes of other specificities.

* Present address: Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE, U.K.

Low-resolution X-ray-crystallographic studies on the enzyme from rabbit muscle (Johnson & Waley, 1967) revealed a twofold molecular symmetry with a subunit weight of 26000 (±3000), whereas those on the chicken enzyme (Banner *et al.*, 1971) indicate a dimeric structure of molecular weight about 49000. The chemical studies in the present paper are entirely consistent with these values.

There is considerable homology between the sequences of the tryptic peptides derived from the chicken enzyme and those of the rabbit enzyme (see Corran & Waley, 1974 for a discussion of the homology). Fig. 4 shows a tentative primary structure of the chicken enzyme on the basis of the CNBr cleavage and the overlapped sequence of the rabbit enzyme (Corran & Waley, 1973).

Materials and Methods

Materials

Iodoacetic acid (BDH Chemicals Ltd., Poole, Dorset, U.K.; 'specially purified for biochemical work') was recrystallized from light petroleum (b.p. 60–80°C). Guanidine hydrochloride was

obtained from BDH Chemicals Ltd. as 'specially purified for biochemical work'. L-α-Amino-β-guanidopropionic acid, DL-β-thienylalanine and DL-norleucine were obtained from Calbiochem. Ltd., London W1H 1AS, U.K. Glycerol 1-phosphate dehydrogenase (EC 1.1.1.8) and NADH (grade I) were obtained from Boehringer Corp. (London) Ltd., London W5 2TZ, U.K. and glyceraldehyde 3-phosphate from Sigma (London) Chemical Co., Kingston-upon-Thames KT2 7BH, U.K. Trypsin [treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one (TPCK) to minimize chymotryptic activity] was obtained from Worthington Biochemicals, Freehold, N.J., U.S.A. CNBr was obtained from Koch–Light Laboratories Ltd., Colnbrook, Bucks., SL3 0BZ, U.K.

Methods

Purification of triose phosphate isomerase from chicken breast muscle. The breast muscle was carved from deep-frozen chickens (MacFisheries Ltd., Oxford, U.K.) and blended in an MSE Ato-Mix with 1 ml of extraction solution [1.3mM-EDTA (disodium salt)–1.0mM-2-mercaptoethanol] /g wet wt. The mixture was stirred for 30min, then centrifuged at 1350*g* for 30min in an MSE Mistral 6L (6×1 litre head). The supernatant was filtered through cheesecloth to remove fat, while the precipitate was re-extracted with one-half the previous volume of extraction mixture and the centrifugation and decantation repeated.

Finely powdered $(NH_4)_2SO_4$ (BDH Chemicals Ltd.; A.R. grade) was added slowly at 4°C over 30min to the combined extracts until 65% saturation (430g added/l) was reached. The mixture was kept at 5°C for 120min and then centrifuged at 23000*g* for 40min in an MSE 'High Speed 18' (6×250ml head). The supernatant was concentrated by ultrafiltration in a 'Diaflo' 2 litre cell with a UM-20E membrane (Amicon Corp., Lexington, Mass., U.S.A.) and then equilibrated against the chromatography buffer (see below) by diafiltration of 10 litres of buffer. Complete removal of $(NH_4)_2SO_4$ was checked by conductivity measurements.

Chromatography on a column (6.8cm×21cm) of DEAE-cellulose (Whatman DE-52) was carried out with starting buffer elution at 4°C at a flow rate of 60ml/h. The buffer was 0.02M-triethanolamine hydrochloride [Sigma (London) Chemical Co.] prepared by diluting a 0.2M solution of the hydrochloride that had been adjusted with NaOH to pH7.8 at 20°C.

Crystals were obtained by concentrating the enzyme by ultrafiltration in buffer to 15mg/ml, adding 5g of finely powdered $(NH_4)_2SO_4$/2ml of solution and leaving the faintly opalescent solution to stand at 4°C. For X-ray work, the solution was filtered through a Millipore filter (1.2μm) before being left at 4°C.

Enzyme activity. Triose phosphate isomerase activity was assayed at 30°C by the method of Beisenherz (1955).

Extinction coefficient. The enzyme was dialysed against degassed distilled water. A measured sample was diluted with triethanolamine buffer (0.2M) by weighing into a cuvette and the absorbance of the solution at 280nm measured on a Zeiss PMQ II spectrophotometer. A measured weight of the remaining solution was dried in an oven at 110°C to constant weight within 0.1%.

Preparation of reduced carboxymethylated triose phosphate isomerase. The protein was dialysed against 0.2M-sodium borate buffer, pH8.5, containing 1mM-EDTA and 0.025mM-dithiothreitol. It was denatured by adding 2g of guanidine hydrochloride/ml of protein solution and incubating at 37°C for 30min. The pH was readjusted if necessary with NaOH to 8.5 and sufficient iodoacetic acid added to give a 5.4-fold molar excess over thiol groups (calculated at 4 thiol groups/subunit plus the dithiothreitol). The solution was incubated at 0°C for 4h. Excess of reagent was then removed by addition of 30μl of 2-mercaptoethanol. The solution was dialysed exhaustively against 0.5% (w/v) NH_4HCO_3 and then freeze-dried.

Amino acid analyses. Samples were hydrolysed at 108°C with 6M-HCl [containing 1% (w/v) phenol to avoid destruction of tyrosine (Sanger & Thompson, 1963)] in tubes which were evacuated and sealed after repeated flushing with N_2. Samples were analysed after hydrolysis with a Beckman 120C automatic amino acid analyser by the method of Benson & Patterson (1965). Cysteine was estimated as carboxymethylcysteine or as cysteic acid (Hirs, 1956). Values for serine, threonine, isoleucine and valine were obtained by extrapolation from the results of 24, 48 and 72h hydrolysis periods.

N-Terminal analysis of peptides. This was carried out by the dansyl chloride procedure. For the whole protein the dansylation was performed in 8M-urea (Gray, 1967*a*). After hydrolysis, the dansyl amino acids were identified by two-dimensional chromatography. The method of Woods & Wang (1967) was followed with, however, a modification that we believe to be desirable. This was the replacement of benzene in the second dimension with the less toxic solvent toluene. This change in solvent caused no significant changes in R_F of the dansyl amino acids.

Edman sequential degradation with dansylation. This was performed by the method of Gray (1967*b*).

C-terminal analysis. This was carried out quantitatively by carboxypeptidase A digestion of performic acid-oxidized enzyme. The oxidized enzyme was dissolved in 0.2M-*N*-ethylmorpholinium acetate

buffer, pH 8.5, by the method of Norton *et al.* (1970). The solution was diluted with a solution of α-amino-β-guanidopropionic acid (final concentration 0.2 μmol/ml) to give a protein concentration of 3 mg/ml. Carboxypeptidase A [1.25 mg as an $(NH_4)_2SO_4$ suspension] was washed free of amino acids by stirring in 0.25 ml of water. The enzyme was centrifuged down, resuspended in 0.25 ml of *N*-ethylmorpholinium acetate buffer, and used without further treatment. Carboxypeptidase A (10 μg in 2 μl) was incubated with 1 mg (in 0.6 ml) of triose phosphate isomerase at 37°C for 1, 2 and 6h. The suspensions were deproteinized by addition of 0.12ml of 10% trichloroacetic acid, and after centrifugation the supernatant solution was applied to the amino acid analyser. Corrections were made for traces of free amino acids that were found in an analysis of undigested performic acid-oxidized triose phosphate isomerase. A carboxypeptidase blank was negative.

The absolute concentration of protein in each sample was determined by removing a sample before digestion and hydrolysing in 6M-HCl. The α-amino-β-guanidopropionic acid (which is eluted after NH_3) then made it possible to calculate absolute amounts of protein in each digest without the need to correct for errors owing to losses on centrifugation.

Cleavage with CNBr. Reduced carboxymethylated triose phosphate isomerase (150mg) in 6ml of 70% (v/v) formic acid, was reacted with CNBr in 30-fold molar excess over methionine residues. After 24h another equal amount of CNBr was added to the mixture and after 30h 25ml of water was added and the reaction mixture freeze-dried. The residue was redissolved and again freeze-dried to remove the last traces of the reagent.

The products were taken up in 7.5 ml of pyridinium acetate buffer, pH3.2 (acetic acid–pyridine–water, 75:4:171, by vol.) and insoluble material separated by centrifugation. The supernatant was freeze-dried, redissolved in the buffer and passed down a column (135 cm × 1.6 cm) of Sephadex G-75 in the same buffer and the eluate was monitored at 280 nm.

Tryptic digestion. Tryptic digestion of a suspension of reduced carboxymethylated enzyme in 0.5% (w/v) NH_4HCO_3 was performed at 37°C for 4h with 1% (w/w) trypsin that had been treated with 1-chloro-4-phenyl-3-L-tosylamidobutan-2-one (TPCK).

Location of peptides on paper. Location of peptides on paper after preparative chromatography or electrophoresis was accomplished by staining a guide-strip cut from one edge of the paper. Cadmium–ninhydrin, Ehrlich's tryptophan stain and chlorination were used, in that order, on the same guide-strip (Offord, 1969).

Paper chromatography. Peptides were separated

Table 1. *Amino acid analysis of chicken triose phosphate isomerase*

The values are on the basis of Leu = 17.

Time of hydrolysis ...	24h	45.5h	72h	Mean values	Integral values	Sum of peptide composition*
Lysine	23.3	24.0	21.8	23.1	23	23
Histidine	7.7	7.5	7.6	7.6	8	8
Arginine	7.8	8.2	6.6	7.5	8	8
Aspartic acid	19.7	19.9	20.5	20.0	20	19
Threonine	9.5	9.5	8.8	10.2†	10	10
Serine	12.6	12.0	10.4	13.5†	13	12
Glutamic acid	24.3	25.9	27.2	25.8	26	26
Proline	8.6	9.7	8.4	8.9	9	7
Glycine	27.0	26.0	28.0	27.0	27	27
Alanine	27.9	28.9	27.6	28.2	28	28
Valine	19.0	22.0	22.6	24.4‡	24	23
Methionine	2.1	2.0	1.9	2.0	2	2
Isoleucine	13.6	15.0	15.4	16.6‡	17	17
Leucine	17	17	17	17	17	17
Tyrosine	3.8	4.0	3.8	3.9	4	4
Phenylalanine	7.7	7.8	7.7	7.8	8	8
Cysteic acid				4.0§	4	5
Tryptophan ||					5	4
Total					253	248

* From Table 3 of Supplementary Publication SUP 50025.

† Extrapolated value at time = 0 obtained from a plot of quantity liberated against time.

‡ Extrapolated values at 1/time = 0 obtained from a plot of quantity liberated against 1/time.

§ Two 24h hydrolyses of performic acid-oxidized enzyme.

|| See Table 5 of Supplementary Publication SUP 50025.

by descending chromatography in butan-1-ol–acetic acid–water–pyridine (15:3:12:10, by vol.) (Waley & Watson, 1953) by using the general techniques described under paper electrophoresis.

Paper electrophoresis. This was carried out at 60–100V/cm in an apparatus similar to that described by Michl (1951) on Whatman 3MC or no. 1 paper at pH1.9, 3.5, 6.5 or 8.9. Buffers and equipment have been described by Ambler (1963). The peptide mixture was dissolved in 0.05M-NH_3 and applied along an origin wide enough to give a loading of 2mg/cm (3MC or 3MM paper) or 0.2μmol/cm (no. 1 paper). After electrophoresis (or chromatography) the peptide bands were located on a stained guide-strip (see above) and cut out. For further purification the strip was sewn to a fresh paper or (where loading was too low) the peptide was eluted with 0.05M-NH_3 (Ingram, 1963) and re-applied to a shorter origin.

Assignment of amide groups. All purification steps on paper were followed as a routine by progress of dyes in a marker mixture run parallel to the peptide mixture. In particular, the marker mixture contained ε-Dnp-lysine (neutral except at pH1.9) and Xylene Cyanol FF (mobility 0.41 relative to aspartic acid at pH6.5). In many cases this permitted immediate discrimination between free carboxyl and amide groups on peptides (Offord, 1966).

Scheme for purification of peptides. The first step in the purification of peptides from a tryptic digest was gel filtration on Sephadex G-50 in 0.5% (w/v) NH_4HCO_3. Up to 160 mg of digest was dissolved in 20ml of buffer, applied to a column (190cm×

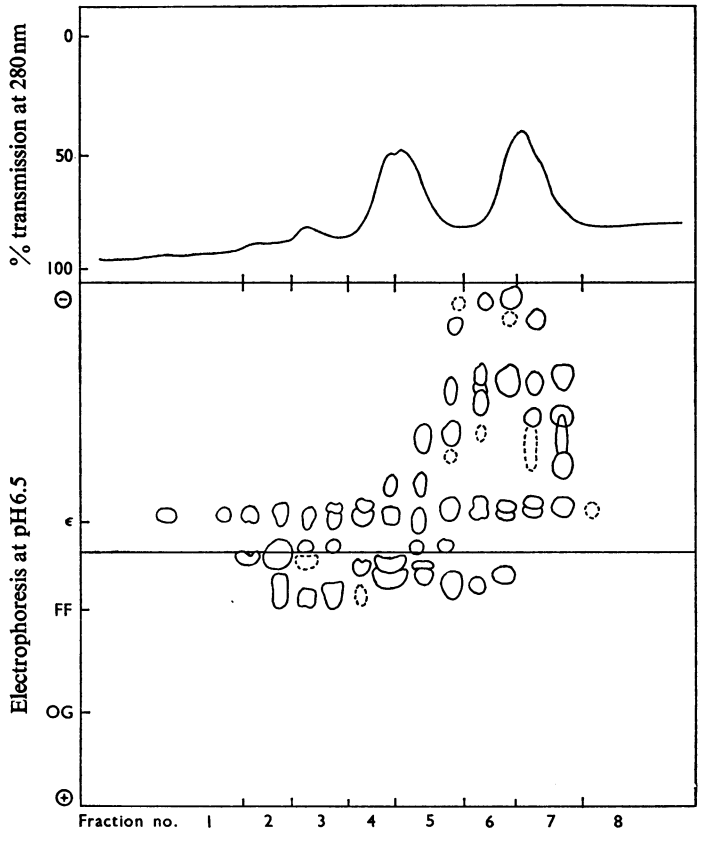


Fig. 1. *Gel filtration of a tryptic digest of carboxymethylated chicken triose phosphate isomerase*

Samples of the effluent were submitted to electrophoresis at pH6.5 and the pattern of peptides is shown below the elution curve. The eluates were combined to give eight separate fractions as indicated. ε represents the position reached by ε-*N*-dinitrophenyl-lysine under endosomotic flow, and FF and OG represent the positions of the cationic dyes Xylene Cyanol FF and Orange G. Other experimental details are given in the text.

4.2cm) and eluted at a flow rate of 130ml/h. Fractions of volume 13ml were collected. The eluate was monitored by absorption at 280nm, and by electrophoretic separation on paper at pH6.5 of 0.1ml portions of every fifth tube (see 'map' in Fig. 1). Tubes were pooled on the basis of this 'map' to give six main fractions.

Each fraction was freeze-dried, and in the second purification step the peptides contained in it were separated by pH6.5 electrophoresis. Subsequent purification steps varied (Table 2 of Supplementary Publication SUP 50025), but, because of the danger of inaccurate amino acid analyses, care was taken where possible to avoid using as a final step either pH6.5 electrophoresis or chromatography in butan-1-ol–acetic acid–water–pyridine (Gonzalez & Offord, 1971).

Results and Discussion

Purification of the enzyme

With the two-step purification procedure described it is possible to obtain 800mg of enzyme/kg wet wt. of breast muscle in a minimum of 2 working days. In the single chromatographic step the enzyme is slightly retarded and is fully resolved from other material provided white (breast) muscle only is used. The specific activity of the enzyme is constant (at approx. 10000 units/mg) across the peak and the protein is homogeneous on polyacrylamide-gel electrophoresis (Davis, 1964).

Extinction coefficient

When protein concentration was determined by drying to constant weight the value of $E^{0.1\%}_{280}$ was 1.20. This value is higher than that previously reported for the chicken enzyme (Trentham *et al.*, 1969) and is close to the value of 1.21 reported for the rabbit muscle enzyme by Miller & Waley (1971).

Amino acid composition

This is given in Table 1. The results were calculated on the basis of 17 leucine residues/subunit. This gives a total of approximately 250 residues and a molecular weight of 27200. Further, the composition agrees with the number of unique tryptic peptides found to contain phenylalanine, histidine, methionine, cysteine and tyrosine residues (Table 5 of Supplementary Publication SUP 50025).

N- and C-Terminal residues

A single α-*N*-dansyl-amino acid, alanine, was found after hydrolysis of protein labelled by dansylation in urea.

Quantitative analysis of the time-course of digestion with carboxypeptidase showed progressive release of histidine, lysine and alanine in amounts suggesting their sequential release from a unique sequence.

The tryptic digest of reduced carboxymethylated enzyme contained free histidine in high yield, supporting its assignment as the *C*-terminal residue. Burton & Waley (1968) found glutamine to be the corresponding residue in the rabbit muscle enzyme.

Isolation of the tryptic peptides

The native chicken enzyme was largely resistant to trypsin, but could be digested after denaturation by boiling or by carboxymethylation. Enzyme (160mg) was carboxymethylated and digested with trypsin. Initial chromatography of the digest on Sephadex G-50 gave six major fractions (see Fig. 1). Details of the steps used to separate peptides in each fraction are given in Table 2 of Supplementary Publication SUP 50025.

Peptides were numbered as follows. The first numeral indicates the Sephadex fraction from which they came. The second numeral, or pair of numerals, indicates the order of mobility on electrophoresis

T41a	Glu-Ala-Gly-Ile-Thr-Glu-Lys
T58c$_1$	Leu-Asp-Ala-Lys
T512a	Gln-Lys-Leu-Asp-Ala-Lys
T65a	Gln-Lys
T58c$_3$	Ala-Ile-Ala-Asp-Asn-Val-Lys
T58d	Met-Asn-Gly-Lys
T514	Val-Pro-Lys
T61	Leu-Asp-Glu-Arg
T65b	Ala-Pro-Arg
T65d	Leu-Arg
T68d	Arg-Lys
T68ivb	Lys-Lys
T75	His
T76	Arg
T77	Lys

Fig. 2. *Peptides sequenced without further enzymic digestion*

⟶ indicates that residues were identified by the Dns method after Edman degradation. Amide assignments for peptide T58c$_3$ are from electrophoretic-mobility measurements after four cycles of Edman degradation.

at pH6.5, numbering from anode to cathode. The digit 0 is reserved for insoluble peptides. Suffixes a, b, c, P, Q, R were added to indicate the order of migration when other separating systems were used for peptides that were superimposed at pH6.5. The same principles were applied to the naming of fragments that were isolated from digests of the larger tryptic peptides by other enzymes; Th, Ch or P are used to denote cleavage by thermolysin, chymotrypsin or pepsin respectively.

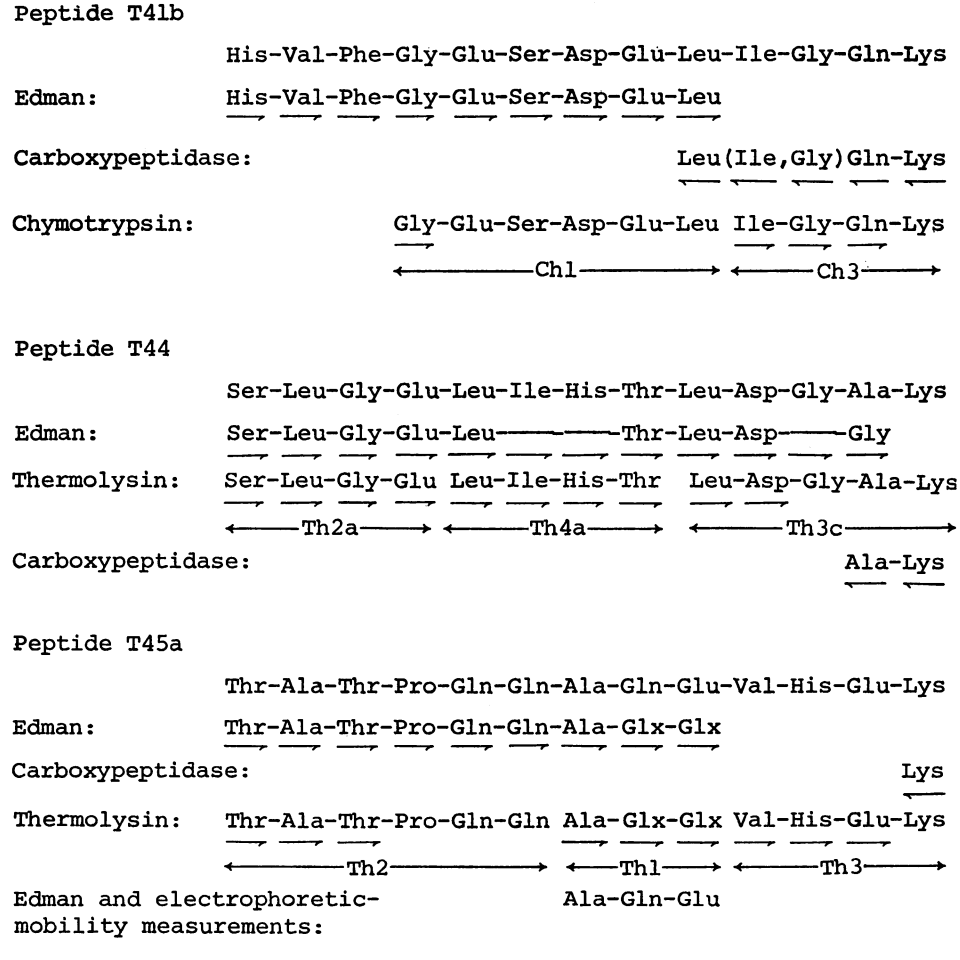
Only two peptides, both of which were large, gave problems because of insolubility at pH6.5. Of these peptide T30c was soluble at pH8.9, whereas peptide T50a, although persistently insoluble, could be purified by removing the soluble contaminants by electrophoresis at pH1.9 and 8.9. The purified peptide T50a could then be recovered in good yield if eluted from paper as soon as possible after electrophoresis.

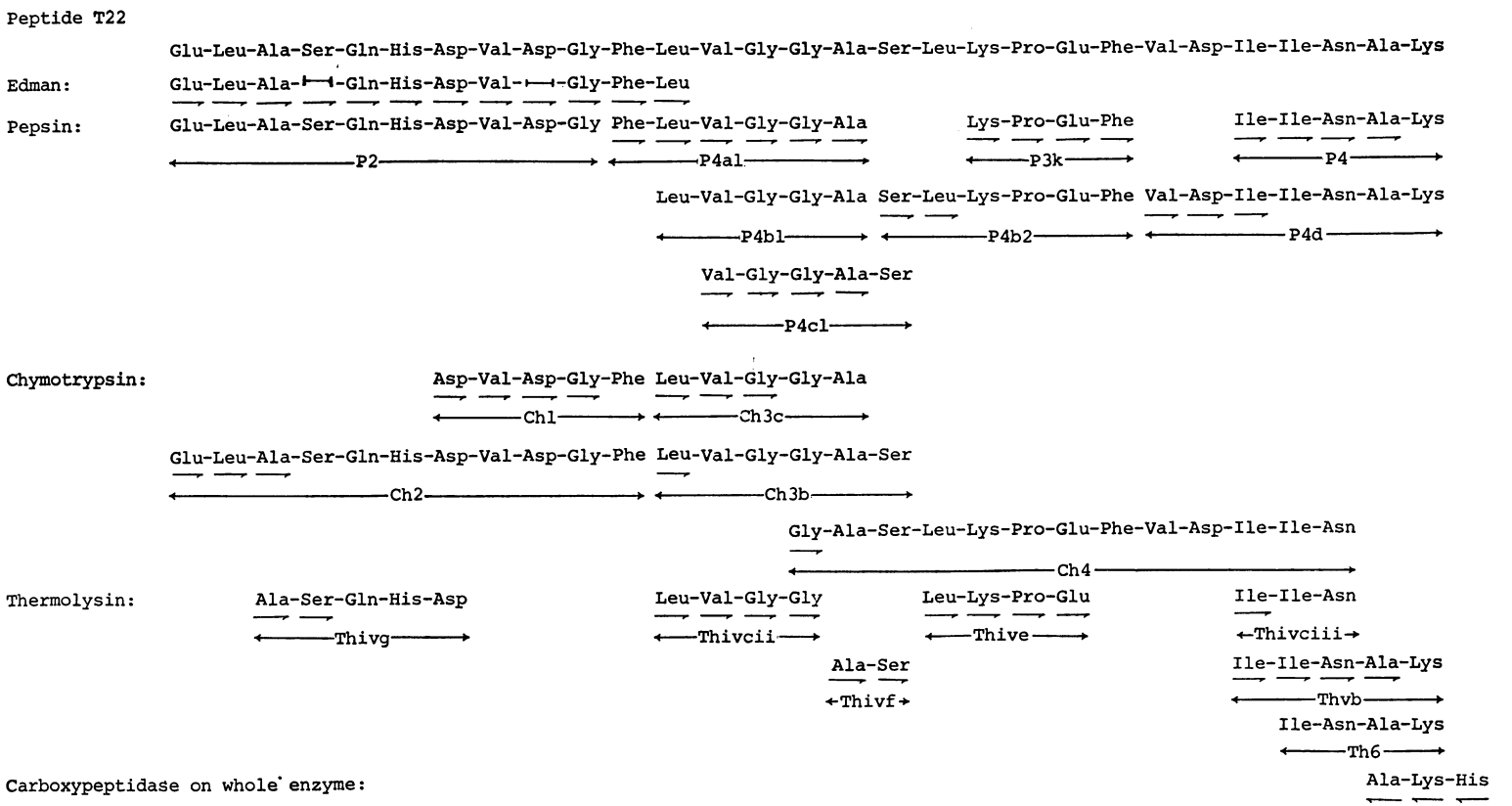
Nature of tryptic peptides

In isolating the fragments from tryptic digestion we were concerned to purify not only the major peptides but also the minor ones. This is essential to make sure that no unique peptide, present in however poor a yield, has been overlooked.

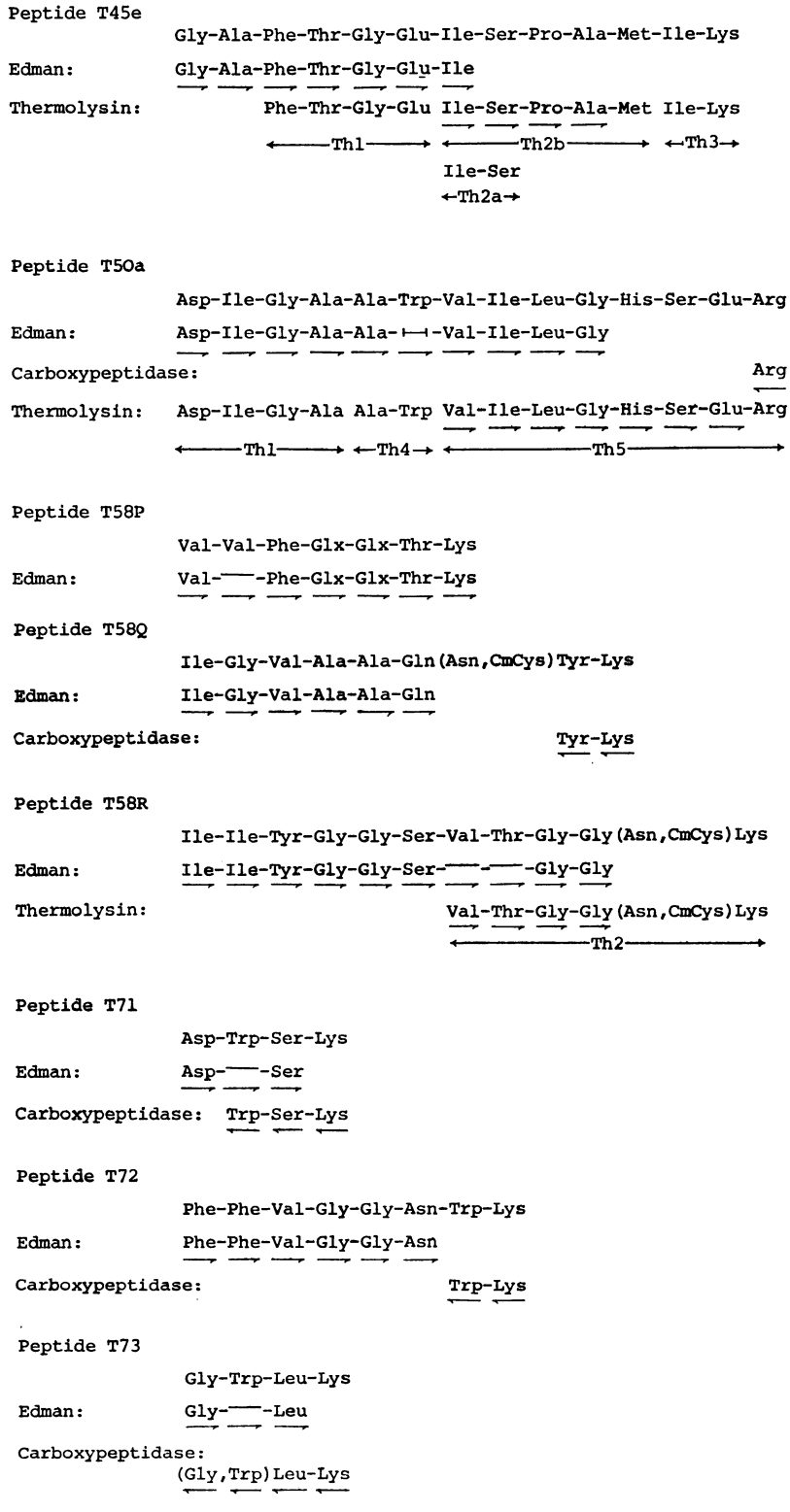
In deciding whether a given peptide was unique or not, two criteria were employed, amino acid composition and the nature of the *N*-terminal amino acid. In cases in which it was still not possible to say whether or not a peptide that had been isolated in poor yield was related to any other, the sequence of the first few residues was determined. Table 3 of Supplementary Publication SUP 50025 shows the composition and *N*-termini of what is thought to be the sum total of the unique peptides in the tryptic digest.

A number of other peptides were obtained in poor




Peptide T22
Glu-Leu-Ala-Ser-Gln-His-Asp-Val-Asp-Gly-Phe-Leu-Val-Gly-Gly-Ala-Ser-Leu-Lys-Pro-Glu-Phe-Val-Asp-Ile-Ile-Asn-Ala-Lys
Edman:
Glu-Leu-Ala-
-Gln-His-Asp-Val-
-Gly-Phe-Leu
Pepsin:
Glu-Leu-Ala-Ser-Gln-His-Asp-Val-Asp-Gly
Phe-Leu-Val-Gly-Gly-Ala
Lys-Pro-Glu-Phe
Ile-Ile-Asn-Ala-Lys
P2
P4al
P3k
P4
Leu-Val-Gly-Gly-Ala
Ser-Leu-Lys-Pro-Glu-Phe
Val-Asp-Ile-Ile-Asn-Ala-Lys
P4b1
P4b2
P4d
Val-Gly-Gly-Ala-Ser
P4c1
Chymotrypsin:
Asp-Val-Asp-Gly-Phe
Leu-Val-Gly-Gly-Ala
Ch1
Ch3c
Glu-Leu-Ala-Ser-Gln-His-Asp-Val-Asp-Gly-Phe
Leu-Val-Gly-Gly-Ala-Ser
Ch2
Ch3b
Gly-Ala-Ser-Leu-Lys-Pro-Glu-Phe-Val-Asp-Ile-Ile-Asn
Ch4
Thermolysin:
Ala-Ser-Gln-His-Asp
Leu-Val-Gly-Gly
Leu-Lys-Pro-Glu
Ile-Ile-Asn
Thivg
Thivcii
Thive
Thivciii
Ala-Ser
Ile-Ile-Asn-Ala-Lys
Thivf
Thvb
Ile-Asn-Ala-Lys
Th6
Carboxypeptidase on whole enzyme:
Ala-Lys-His

Val-Ala-His-Ala-Leu-Ala-Glu-Gly-Leu-Gly-Val-Ile-Ala-CmCys-Ile-Gly-Glu-Lys
Edman:
Val-Ala-
-Ala-Leu-Ala-Glu-Gly-Leu-Gly
Carboxypeptidase:
Lys
Thermolysin:
Val-Ala-His-Ala Leu-Ala-Glu-Gly Leu-Gly
Ile-Gly-Glu-Lys
Th4b
Th1
Th3c
Th3e
Chymotrypsin:
Ala-Glu-Gly-Leu Gly-Val-Ile-Ala-CmCys-Ile-Gly-Glu-Lys
Ch1
Ch3b
Ala-Glu-Gly-Leu-Gly-Val-Ile-Ala-CmCys-Ile-Gly-Glu-Lys
Ch2
Peptide T31a
Leu-Ser-Ala-Asp-Thr-Glu-Val-Val-CmCys-Gly-Ala-Pro-Ser-Ile-Tyr-Leu-Asp-Phe-Ala-Arg
Edman:
-
-Ala-Asp-Thr-Glu-Val-Val-CmCys-Gly-Ala-Pro
Carboxypeptidase:
Ile(Tyr,Leu)Asp-Phe-Ala-Arg
Chymotrypsin:
Leu-Ser-Ala-Asp-Thr-Glu-Val-Val-CmCys-Gly-Ala-Pro-Ser-Ile-Tyr Leu-Asp-Phe
Ch1a
Ch1b
Leu-Asp-Phe-Ala-Arg
Ch3b
Pepsin:
Leu-Ser-Ala-Asp-Thr-Glu
Phe-Ala-Arg
P2a
P6
Asp-Thr-Glu
Leu-Asp-Phe Ala-Arg
P1
P3
P7


Peptide T45e
Gly-Ala-Phe-Thr-Gly-Glu-Ile-Ser-Pro-Ala-Met-Ile-Lys
Edman:
Gly-Ala-Phe-Thr-Gly-Glu-Ile
Thermolysin:
Phe-Thr-Gly-Glu Ile-Ser-Pro-Ala-Met Ile-Lys
Th1
Th2b
Th3
Ile-Ser
Th2a
Peptide T50a
Asp-Ile-Gly-Ala-Ala-Trp-Val-Ile-Leu-Gly-His-Ser-Glu-Arg
Edman:
Asp-Ile-Gly-Ala-Ala-
-Val-Ile-Leu-Gly
Carboxypeptidase:
Arg
Thermolysin:
Asp-Ile-Gly-Ala Ala-Trp Val-Ile-Leu-Gly-His-Ser-Glu-Arg
Th1
Th4
Th5
Peptide T58P
Val-Val-Phe-Glx-Glx-Thr-Lys
Edman:
Val-
-Phe-Glx-Glx-Thr-Lys
Peptide T58Q
Ile-Gly-Val-Ala-Ala-Gln(Asn,CmCys)Tyr-Lys
Edman:
Ile-Gly-Val-Ala-Ala-Gln
Carboxypeptidase:
Tyr-Lys
Peptide T58R
Ile-Ile-Tyr-Gly-Gly-Ser-Val-Thr-Gly-Gly(Asn,CmCys)Lys
Edman:
Ile-Ile-Tyr-Gly-Gly-Ser-
-
-Gly-Gly
Thermolysin:
Val-Thr-Gly-Gly(Asn,CmCys)Lys
Th2
Peptide T71
Asp-Trp-Ser-Lys
Edman:
Asp-
-Ser
Carboxypeptidase:
Trp-Ser-Lys
Peptide T72
Phe-Phe-Val-Gly-Gly-Asn-Trp-Lys
Edman:
Phe-Phe-Val-Gly-Gly-Asn
Carboxypeptidase:
Trp-Lys
Peptide T73
Gly-Trp-Leu-Lys
Edman:
Gly-
-Leu
Carboxypeptidase:
(Gly,Trp)Leu-Lys

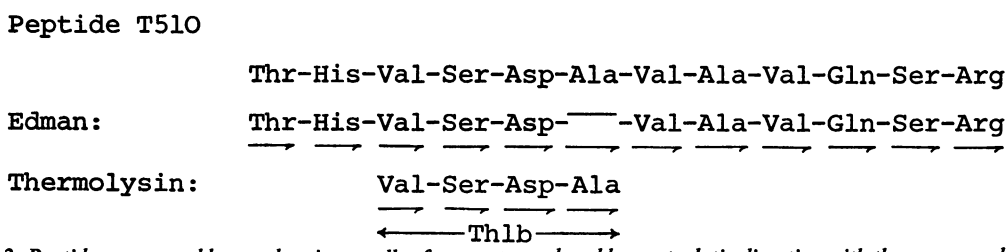


Fig. 3. *Peptides sequenced by overlapping smaller fragments produced by proteolytic digestion with the enzymes shown* —→, indicates residues that were identified by the Dns method after Edman degradation; ←—, indicates residues that were released by carboxypeptidase digestion. CmCys, carboxymethylcysteine. CmCys did not appear in the analyses of the peptides from peptides T30c, T31a, T58Q and T58R, but its negative charge is implied by the observed mobility on paper electrophoresis at pH 6.5 in each case.

yield. Of those that could be isolated in sufficient quantities for analysis and end-group determination, all could be seen to derive from the major peptides (Table 4 of Supplementary Publication SUP 50025).

Since a complete overlapping sequence has not been obtained, it is not possible to state with absolute certainty that every part of the molecule has been recovered in the form of a tryptic peptide. However, the summation in the last column of Table 1 makes it seem unlikely that any great portion can be missing, a conclusion strengthened by homology with the sequence of the rabbit muscle enzyme obtained by Waley and co-workers (see below).

The summation of the composition of the tryptic peptides indicates that the functional unit of the enzyme (mol. wt. 50000) is made up of two subunits. Also, if the subunits differ in sequence, they do so only very slightly. This is confirmed by the amino acid sequences of the peptides (Figs. 2 and 3). If there were slight differences between subunits one might expect to find two peptides with closely similar sequences. Alternatively, a difference might be so slight that it would not be sufficient to permit the separation of the two forms. In such a case there should be ambiguities in the sequences at the site of the difference. We see no evidence for such ambiguities although it is in the nature of sequence work that it is difficult to have complete certainty.

Sequences of the tryptic peptides

Fig. 2 shows the sequences of those tryptic peptides whose sequence can be derived directly by dansyl–Edman degradation, taken with knowledge of the specificity of trypsin itself.

Fig. 3 shows the sequences of those longer peptides that had to be further degraded by proteolytic enzymes of different specificity. The purification schemes of these subfragments shown in the alignments of Fig. 3 are given in Table 5 of Supplementary Publication SUP 50025. The end groups, amino acid compositions and electrophoretic mobility of these subfragments are shown in Table 6 of Supplementary Publication SUP 50025.

The determination of the sequence of the active-site peptide T59 rests on the work of Coulson *et al.* (1970) and Priddle & Offord (1974). In addition, sequences in certain peptides were confirmed, or extended, by mass spectrometry of permethylated fragments (see Appendix to this paper).

Fragments from CNBr cleavage

Chromatography of the soluble products of CNBr digestion on Sephadex G-75 gave two major peptide peaks (Fig. 4 of Supplementary Publication SUP 50025). Each appears to be a pure peptide and they are designated fragments I and II. The dansyl end group of fragment I was alanine (the same as the intact protein), that of fragment II was aspartic acid or its amide. The insoluble fraction of the digest gave a single end group (isoleucine) and was designated fragment III. Only fragments I and II were found to contain homoserine on amino acid analysis (Table 7 of Supplementary Publication SUP 50025). Taken together these data suggest that fragment I represents the *N*-terminal region of the protein, fragment II the central region and fragment III the *C*-terminus.

The fragments were digested by trypsin and the resulting peptides separated and analysed. Those from fragment I were purified in much the same way as were the tryptic peptides of the whole protein. Those from fragments II and III were eluted from two-dimensional peptide 'maps' after location by staining with 0.02% ninhydrin in acetone (Figs. 5 and 6 of Supplementary Publication SUP 50025) Usually, sufficient material was obtained for the determination of the amino acid composition and *N*-terminal residue (Table 8 of Supplementary Publication SUP 50025).

By these means, nearly all the tryptic peptides isolated from the intact protein could be allocated within the three CNBr fragments of the molecule (Fig. 4). Peptides T50a, T58P, T58Q, T58R and T68d could be identified from the 'maps' obtained after CNBr cleavage. There are reasons why these should not be easy to isolate. Peptide T50a is an

* Two deletions have been introduced in order to maximize the homology.

Fig. 4. *Proposed primary structures of chicken and rabbit triose phosphate isomerases*

insoluble peptide; the T58 series of peptides are neutral peptides, which therefore move to a crowded portion of the 'maps', and which share a tendency to progressive deamidation with consequent dispersion of the material into a number of sites; and peptide T68d is susceptible to variations in the manner and extent of tryptic cleavage during the digestion.

Assembly of the total sequence

Peptide T65b (Ala-Pro-Arg) is at the *N*-terminus of the protein since CNBr fragment I has been established by dansylation as the *N*-terminal fragment and yields only this one tryptic peptide with an alanyl *N*-terminus.

Peptide T22 (followed by histidine, see Fig. 4) is established as the *C*-terminal peptide, since carboxypeptidase treatment of the whole protein indicates that the *C*-terminal sequence is (Phe,Val,Asp,Ile)-(Asn,Ala,Lys)-His. Only two tryptic peptides derived from CNBr fragment III contain phenylalanine (T22 and T41b) and only peptide T22 has the appropriate *C*-terminal sequence. It is also possible, by considering the amino acid sequence of the methionine peptides together with the end group and tryptic peptides of CNBr fragments, to assign peptide T58d to the junction between fragments I and II and peptide T45e to the junction between fragments II and III. The remaining alignments are by homology with the sequence of the rabbit muscle enzyme [Corran & Waley (1973) and see the discussion of the homology in the accompanying paper, Corran & Waley (1974)]. The suggested sequence is shown in Fig. 4.

Conclusion

The results presented above indicate that chicken triose phosphate isomerase is a dimer of identical or closely similar subunits of molecular weight 27000. The amino acid sequence is closely homologous to that of the rabbit muscle enzyme.

An extended version of this paper has been deposited as Supplementary Publication SUP 50025 at the National Lending Library for Science and Technology, Boston Spa, Yorks. LS23 7BQ, U.K. The deposited version has the following additional material: Table 2 (the purification scheme for the tryptic peptides), Table 3 (the compositions, end groups and electrophoretic mobilities of the tryptic peptides), Table 4 (details of the minor peptides that were obtained from the tryptic digest), Table 5 (the purification schemes for the peptides derived from further enzymic digestion of the tryptic peptides), Table 6 (the compositions, end groups and mobilities of the subfragments of the tryptic peptides), Table 7 (the amino acid compositions of the CNBr fragments), Fig. 4 (the separation of the CNBr fragments by gel filtration), Fig. 5 (peptide 'maps' of the tryptic digest of CNBr fragment II) and Fig. 6 (peptide 'maps' of the tryptic digest of CNBr fragment III).

We thank Dr. S. G. Waley, Dr. P. H. Corran and Dr. J. C. Miller for many useful discussions and exchanges of information at all stages. We thank Dr. C. I. Pogson for some of the samples of enzyme that were used in the initial stages of this work. Some of the later samples were prepared for us by Mr. M. Iqbal and colleagues. We thank the following for financial support: Science Research Council (a post-doctoral fellowship for A.J.F. and a studentship for J.D.M.); Medical Research Council (support for R.E.O. and a scholarship for J.D.P.) and Messrs, Merck, Sharp and Dohme (R.E.O.).

References

Ambler, R. P. (1963) *Biochem. J.* **89**, 349–378
Banner, D. W., Bloomer, A. C., Petsko, G. A. & Phillips, D. C. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 151–155
Beisenherz, G. (1955) *Method Enzymol.* **1**, 387–391
Benson, J. V. & Patterson, J. A. (1965) *Anal. Chem.* **37**, 1108–1110
Burton, P. M. & Waley, S. G. (1968) *Biochem. J.* **107**, 737–744
Corran, P. H. & Waley, S. G. (1973) *FEBS Lett.* **30**, 97–99
Corran, P. H. & Waley, S. G. (1974) *Biochem. J.* **139**, 1–10
Coulson, A., Knowles, J. R., Priddle, J. D. & Offord, R. E. (1970) *Nature (London)* **227**, 180–181
Davis, B. J. (1964) *Ann. N.Y. Acad. Sci.* **121**, 404–416
Gonzalez, G. & Offord, R. E. (1971) *Biochem. J.* **125**, 309–317
Gray, W. R. (1967*a*) *Methods Enzymol.* **11**, 139–151
Gray, W. R. (1967*b*) *Methods Enzymol.* **11**, 469–475
Hirs, C. H. W. (1956) *J. Biol. Chem.* **219**, 611–621
Ingram, V. M. (1963) *Methods Enzymol.* **6**, 831–848
Johnson, L. N. & Waley, S. G. (1967) *J. Mol. Biol.* **29**, 321–322
Knowles, J. R., Leadlay, P. F. & Maister, S. G. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 157–164
Michl, H. (1951) *Monatsh. Chem.* **82**, 489–498
Miller, J. C. & Waley, S. G. (1971) *Biochem. J.* **122**, 209–218
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
Offord, R. E. (1969) in *Data for Biochemical Research* (Dawson, R. M. C., Elliott, D. C., Elliott, W. H. & Jones, K. M., eds.), pp. 525–535, Oxford University Press, Oxford
Priddle, J. D. & Offord, R. E. (1974) *FEBS Lett.* in the press
Sanger, F. & Thompson, E. O. P. (1963) *Biochim. Biophys. Acta* **71**, 468–471
Trentham, D. R., McMurray, C. H. & Pogson, C. I. (1969) *Biochem. J.* **114**, 19–24
Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337
Woods, K. R. & Wang, K.-T. (1967) *Biochim. Biophys. Acta* **133**, 369–370