The Primary Structure of Aspartate Aminotransferase from Pig Heart Muscle

PARTIAL SEQUENCES DETERMINED BY DIGESTION WITH PEPSIN AND TRYPSIN

By S. DOONAN, HILARY J. DOONAN, FRANCESCA RIVA, C. A. VERNON
and J. M. WALKER

*Department of Chemistry, University College London,
20 Gordon Street, London WC1H 0AJ, U.K.*

and FRANCESCO BOSSA, DONATELLA BARRA, MASSIMO CARLONI
and PAOLO FASELLA

*Istituto di Chimica Biologica e Centro di Biologia Molecolare del Consiglio Nazionale delle Ricerche,
Università di Roma, Città Universitaria-00185 Roma, Italy*

(Received 25 May 1972)

Peptides obtained by tryptic digestion of carboxymethylated and maleylated aspartate aminotransferase and of the aminoethylated enzyme were isolated and the complete amino acid sequences of most of them were determined. Digestion of the carboxymethylated protein with pepsin produced a complex mixture of peptides that allowed some overlapping of the tryptic peptides (Fig. 4); in addition, peptides were obtained that had not been found in either of the tryptic digests. From these studies about 400 amino acid residues were identified. Experimental details and confirmatory data for the results presented here are given in a supplementary paper that has been deposited as Supplementary Publication 50011 at the National Lending Library for Science and Technology, Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1972) 126, 5.

The soluble (cytoplasmic) form of aspartate aminotransferase (L-aspartate–2-oxoglutarate aminotransferase, EC 2.6.1.1) from pig heart muscle has been the subject of intensive study in recent years (for reviews see Guirard & Snell, 1964; Ivanov & Karpeisky, 1969; Braunstein, 1970; Fasella & Turano, 1970). It is generally agreed that both the enzymic reaction and also the reaction catalysed by the cofactor alone proceed by the mechanism proposed independently by Braunstein & Shemyakin (1953) and by Metzler *et al.* (1954). The values of the rate constants describing the rate-limiting steps in enzymic transamination have been determined both by steady-state (Banks *et al.*, 1968a) and by fast-relaxation kinetics (Fasella & Hammes, 1967), and have been compared with the values of the same parameters describing the model system (Banks *et al.*, 1968a). These studies have shown that the presence of the protein part of the holoenzyme produces rate enhancements in the range 10^7 – 10^9 -fold. An understanding of the origin of such rate enhancements must be sought from a knowledge of the three-dimensional structure of the enzyme. As a necessary preliminary to crystallographic studies of aspartate aminotransferase, we have undertaken a collaborative project to determine the primary structure of the enzyme; the

present paper describes the structures of peptides derived from hydrolysis of the enzyme with trypsin and pepsin and the partial sequences obtainable from these data.

During the course of this work preliminary communications describing tryptic and chymotryptic fragments of cytoplasmic aspartate aminotransferase have been published (Ovchinnikov *et al.*, 1971a,b). The results presented below confirm and extend those of the Russian authors.

Experimental

Materials

Aspartate aminotransferase used for digestion with trypsin was prepared by the method of Jenkins *et al.* (1959) as modified by Martinez-Carrion *et al.* (1967). Removal of pyridoxal 5'-phosphate was carried out by dialysis against 1 M-potassium phosphate buffer, pH 5.3, containing glutamic acid (0.1 M) at 2–4°C for 3 days (Turano *et al.*, 1967). The enzyme used for peptic digestion was purchased from Whatman Biochemicals Ltd., Maidstone, Kent, U.K.; this material is prepared essentially as described by Banks *et al.* (1968b).

General experimental methods

The experimental methods employed were based on those generally used for protein structure determination; details are given in Supplementary Publication SUP 50011.

Modification and proteolytic digestion of aspartate aminotransferase

Carboxymethylation (Crestfield *et al.*, 1963) and aminoethylation (Cole, 1967) were carried out on aspartate aminotransferase that had been exposed to urea (8M) for 38 h. Amino acid analysis showed that, in both cases, four out of five cysteine residues had been modified. Cm-Enzyme* was maleylated according to the method of Butler *et al.* (1969). Samples of

* Abbreviations: Cm-Enzyme, H₂NEt-Enzyme and Cm-Mal-Enzyme, carboxymethylated, aminoethylated and maleylated aspartate aminotransferase respectively; CmCys and H₂NEtCys (in sequences), carboxymethylcysteine and aminoethylcysteine respectively.

Cm-Mal-Enzyme and H₂NEt-Enzyme (500mg each) were subjected to digestion with trypsin; peptides from Cm-Mal-Enzyme were demaleylated by incubation at pH 3.5 at 37°C for 24 h. Cm-Enzyme (1250mg) was digested with pepsin.

Peptide separation

Peptide mixtures from the tryptic digests were subjected to preliminary fractionation by gel filtration through Sephadex G-25 followed by Sephadex G-50. Individual fractions from gel filtration were further purified by ion-exchange chromatography on the sulphonated polystyrene resin Bio-Rad AG-50W-X2 and, where necessary, by paper chromatography with the solvent system BAWP (butan-1-ol-acetic acid-water-pyridine, 15:3:12:10, by vol.) (Waley & Watson, 1953). The peptide mixture from the peptic digest was fractionated by passage through Sephadex G-25, and individual fractions were further purified by paper chromatography and by high-voltage paper electrophoresis at pH 3.5 and 6.5.

Peptide	Sequence	
TM-1	<u>Ala-Arg</u>	
TM-2	<u>Ser-Tyr-Arg</u>	
TM-3	<u>Ile-Val-Ala-Arg</u>	
TM-4	<u>Ile-Leu-Ser-Met-Arg</u>	
TM-5	<u>Lys-Val-Glu-Gln-Arg</u>	
TM-6	<u>Glx-Asx-Pro-Asx-Pro-Arg</u>	TM-6' <u>Pro-Arg</u>
TM-7	<u>Leu-Ile-Ala-Asx-Phe-Arg</u>	
TM-8	<u>Tyr-Trp-Asx-Thr-Glx-Lys-Arg</u>	
TM-9	<u>Ile-Gly-Ala-Glx-Phe-Leu-Ala-Arg</u>	
TM-10	<u>Lys-Val-Asx-Leu-Gly-Val-Gly-Ala-Tyr-Arg</u>	
TM-11	<u>Val-Thr-Trp-Ser-Asx-Pro-Pro-Ala-Glx-Gly-Ala-Arg</u>	
TM-12	<u>Val-Gly-Gly-Val-Glx-Ser-Leu-Gly-Gly-Thr-Gly-Ala-Leu-Arg</u>	
TM-13	<u>Leu-Ala-Leu-Gly-Asx-Asx-Ser-Pro-Ala-Leu-Glx-Glx-Lys-Arg</u>	
TM-14	<u>Val-Gly-Asx-Leu-Thr-Val-Val-Ala-Lys-Glx-Pro-Asx-Ser-Ile-Leu-Arg</u>	
TM-15	<u>Ile-Ala-Asx-Asx-Ser-Ser-Leu-Asx-His-Glx-Tyr-Leu-Pro-Ile-Leu-Gly-Leu-Ala-Glx-Phe-Arg</u>	

Fig. 1. Amino acid sequences of peptides from tryptic digestion of carboxymethylated and maleylated aspartate aminotransferase

→, Residues from dansyl-Edman method; →, residues from direct Edman method; —, residues from digestion with carboxypeptidases. Residues not underlined were assigned by comparison with other known sequences as described in the text.

The purity of peptides was established by use of the dansylation technique described by Gray (1967). Identification of Dns-amino acids was by chromatography on polyamide t.l.c. plates as described by Woods & Wang (1967). Additional solvent systems were used to distinguish between the dansyl derivatives of acidic and of hydroxy amino acids (Crowshaw *et al.*, 1967) and between the monodansyl derivatives of the basic amino acids (Hartley, 1970). Unequivocal identification of Dns-amino acids was achieved by the use of standards, as recommended by Hartley (1970). A dansylation technique was also used for qualitative amino acid analysis. The residue of amino acids after *N*-terminal analysis was com-

Amino acid sequence determinations

Sequence studies with the dansyl-Edman technique were carried out by the method of Hartley (1970). Phenyl isothiocyanate was purified by distillation at 1.3 N/m² (0.1 mmHg) and stored under N₂; pyridine was purified by distillation at atmospheric pressure.

Peptide	Sequence
TAE-1	<u>Ile</u> - <u>Gln</u>
TAE-2	<u>Leu</u> - <u>Lys</u>
TAE-3	<u>Ile</u> - <u>Val</u> - <u>Arg</u>
TAE-4	<u>Asx</u> - <u>Ile</u> - <u>Arg</u>
TAE-5	<u>Ile</u> - <u>Asn</u> - <u>Met</u> - <u>H₂NEtCys</u>
TAE-6	<u>Glx</u> - <u>Val</u> - <u>Glx</u> - <u>Tyr</u>
TAE-7	<u>Thr</u> - <u>Met</u> - <u>Ala</u> - <u>Asx</u> - <u>Arg</u>
TAE-8	<u>Leu</u> - <u>Leu</u> - <u>Pro</u> - <u>Ser</u> - <u>Gly</u> - <u>Arg</u>
TAE-9	<u>Glx</u> - <u>Ile</u> - <u>Ala</u> - <u>Ser</u> - <u>Val</u> - <u>Met</u> - <u>Lys</u>
TAE-10	<u>Val</u> - <u>Leu</u> - <u>Ser</u> - <u>Glx</u> - <u>Met</u> - <u>Gln</u> - <u>Lys</u>
TAE-11	<u>Ala</u> - <u>Arg</u> - <u>Leu</u> - <u>Glx</u> - <u>Ala</u> - <u>Leu</u> - <u>Lys</u>
TAE-11'	<u>Leu</u> - <u>Glx</u> - <u>Ala</u> - <u>Leu</u> - <u>Lys</u>
TAE-12	<u>Tyr</u> - <u>Asx</u> - <u>Gly</u> - <u>Thr</u> - ^{Glx} <u>Asx</u> - <u>Asx</u> - <u>Lys</u>
TAE-12'	<u>Trp</u> - <u>Tyr</u> - <u>Asx</u> - <u>Gly</u> - <u>Thr</u> - ^{Glx} <u>Asx</u> - <u>Asx</u> - <u>Lys</u> - <u>Asx</u> - <u>Thr</u> - <u>Pro</u> - <u>Val</u> - <u>Tyr</u> -(<u>Ala</u> , <u>Asx</u> ₂ , <u>Glx</u> , <u>Gly</u> ₂ , <u>His</u> , <u>Leu</u> , <u>Phe</u> , <u>Pro</u> , <u>Ser</u> ₂ , <u>Thr</u> ₂ , <u>Val</u> ₂)- <u>Phe</u> - <u>Lys</u>
TAE-13	<u>Ser</u> - <u>Phe</u> - <u>Thr</u> - <u>Gly</u> - <u>Leu</u> - <u>Asx</u> - <u>Pro</u> - <u>Lys</u>
TAE-14	<u>Thr</u> - <u>Asx</u> - <u>Asx</u> - <u>H₂NEtCys</u> - <u>Glx</u> - <u>Pro</u> - <u>Trp</u> - <u>Val</u> - <u>Leu</u> - <u>Pro</u> - <u>Val</u> - <u>Val</u> - <u>Arg</u>
TAE-14'	<u>Thr</u> - <u>Asx</u> - <u>Asx</u> - <u>H₂NEtCys</u> - <u>Glx</u> - <u>Pro</u> - <u>Trp</u>
TAE-15	<u>Ala</u> - <u>His</u> - <u>Asx</u> - <u>Pro</u> - <u>Thr</u> - <u>Gly</u> - <u>Thr</u> - <u>Asx</u> - <u>Pro</u> - <u>Thr</u> - <u>Pro</u> - <u>Glx</u> - <u>Glx</u> - <u>Trp</u> - <u>Lys</u>
TAE-16	<u>Thr</u> - <u>Leu</u> - <u>Ser</u> - <u>Asx</u> - <u>Pro</u> - <u>Glx</u> - <u>Leu</u> - <u>Phe</u> - <u>His</u> - <u>Glx</u> - <u>Trp</u> - <u>Thr</u> - <u>Gly</u> - <u>Asx</u> - <u>Val</u> - <u>Lys</u>
TAE-17	<u>Ala</u> - <u>Pro</u> - <u>Pro</u> - <u>Ser</u> - <u>Val</u> - <u>Phe</u> - <u>Ala</u> - <u>Glx</u> - <u>Val</u> - <u>Pro</u> - <u>Glx</u> - <u>Ala</u> - <u>Glx</u> - <u>Val</u> - <u>Leu</u> - <u>Val</u> - <u>Phe</u> - <u>Lys</u>

Fig. 2. Amino acid sequences of peptides from tryptic digestion of aminoethylated aspartate aminotransferase

→, Residues from dansyl-Edman method; ←, residues from digestion with carboxypeptidases. Residues not underlined were assigned by comparison with other known sequences as described in the text.

Tryptophan could not be identified directly by the dansyl method since Dns-tryptophan is destroyed during acid hydrolysis. The presence of tryptophan in a peptide was indicated, however, by the production of an intense purple colour during cyclization and cleavage of the phenylthiocarbamoyl-peptide in trifluoroacetic acid; this colour was produced only weakly during the step in which tryptophan was removed from the peptide. This, combined with failure to detect a Dns-amino acid after the previous degradation cycle, established the position of the tryptophan residue. The C-terminal sequences of some of the tryptic peptides were established by hydrolysis with carboxypeptidases A and B (Ambler, 1967) followed by quantitative amino acid analysis.

Results

Peptides from the tryptic digest of carboxymethylated and maleylated aspartate aminotransferase

Fifteen peptides were isolated in sufficient amount to allow either partial or complete analysis of their amino acid sequences. The sequences are shown in Fig. 1, the peptides being labelled TM-1-TM-15 in order of increasing size. The amino acid compositions of the peptides and data relating to digestion with carboxypeptidases A and B are given in Supplementary Publication SUP 50011.

The sequence of peptide TM-5 was established by the direct Edman method (Edman, 1957, 1960); residues 3 and 4 were shown to be glutamic acid and

Peptide	Method of isolation	Sequence
P1-1	E ₁ , E ₂	<u>Val-Leu-His-Ala-CmCys-Ala-His-Asx-Pro-Thr-Gly-Thr-Asx-Pro-Thr-Pro-Glx-Glx</u>
P1-2	E ₁ , E ₂	<u>Thr-Val-Val-Ala-Lys-Glu-Pro-Asp-Ser-Ile-Leu-Arg-Asp</u>
P1-3	E ₁ , E ₂	<u>Ala-Glu-Val-Pro-Gln-Ala-Gln-Pro-Val-Leu</u>
P1-4	E ₁ , E ₂ , C	<u>Pro-Val-Val-Arg-Lys-Val-Glu-Gln-Arg-Ile-Ala-Asn-Asn-Ser-Ser-Leu</u>
P1-5	E ₁ , C	<u>Ile-Asx-Glx-Lys-His-Ile-Tyr-Leu</u>
P1-6	E ₁ , E ₂ , C	<u>Arg-Val-Leu-Ser-Glx-Met-Glx-Lys-Ile-Val-Arg-Val</u>
P1-7	E ₁ , E ₂	(Arg) <u>Val-Gly-Gly-Val-Glx-Ser-Leu-Gly-Gly-Thr-Gly-Ala</u>
P1-8	E ₁ , E ₂	(Arg) <u>Lys-Val-Asx-Leu</u>
P2-1	E ₁	<u>Ala-Leu-Gly-Asp-Asp-Ser-Pro-Ala-Leu-Glx-Glx</u>
P2-2	E ₁	<u>Pro-Arg-Lys-Val-Asn-Leu-Gly-Val-Gly-Ala-Tyr</u>
P2-3	E ₁ , C, E ₂	<u>Phe-Arg-Glu-Asp-Pro-Asp-Pro-Arg-Lys-Val-Asn-Leu</u>
P2-4	E ₁ , C	<u>Leu-Lys-Thr-Trp-Gly-Thr</u>
P2-5	E ₁ , C, E ₂	<u>Glu-Ala-Leu-Lys-Thr-Pro-Gly-Thr</u>
P2-6	E ₁ , C, E ₂	<u>Asp-Arg-Ile-Leu-Ser-Met</u>
P2-7	E ₁ , C	<u>Glx-Pro-Trp-Val-Leu-Pro-Val</u>
P3-1	E ₁ , E ₂ , C	<u>Lys-Leu-Ile-Ala-Asp-Phe</u>
P3-2	E ₁ , E ₂	<u>Val-Ala-Thr-Ser-Ile-His-Glx</u>
P3-3	E ₁ , C	<u>Ala-Val-Thr-Lys-Ile-Gln</u>
P3-4	E ₁ , E ₂ , C	<u>Thr-Gly-Leu-Asx-Pro-Lys-Glx-Val-Glx-Tyr</u>
P3-5	E ₁ , E ₂ , C	<u>Arg-Thr-Leu-Ser-Asx-Pro-Glx-Leu</u>
P3-6	E ₁	<u>Phe-His-Glx-Trp-Thr-Gly-Asx-Val-Lys-Thr-Met-Ala</u>

Fig. 3.

P5-7	E ₁ , E ₂ , C	<u>Asp-Arg-Ile-Leu-Ser-Met-Arg-Ser-Glu-Leu</u>	P8-4	C, E ₂ , E ₁	<u>Phe-Glu-Leu</u>
P5-1	E ₁ , C	<u>Arg-Ala-Arg-Leu</u>	P8-5	C, E ₂ , E ₁	<u>Ser-Ile-Phe</u>
P5-2	E ₁ , C	<u>Lys-Asp-Ile-Arg-Ser</u>	P8-6	C, E ₂	<u>Asp-Ser-Ala-Tyr-Gln-Gly-Phe</u>
P5-3	E ₁ , C	<u>Ala-Arg-Ile-Val-Ala</u>	P8-7	C, E ₂	<u>Phe-CmCys-Ala-Gln-Ser-Phe</u>
P5-4	E ₁ , C	<u>Arg-Val-Leu-Ser-Glx</u>	P8-8	C, E ₂ , E ₁	<u>Glx-Asx-Tyr-Val-Ala</u>
P5-5	E ₁ , C	<u>Leu-Arg-Ile-Gly-Ala-Glu</u>	P8-9	C, E ₂ , E ₁	<u>Ala-Glu-Phe</u>
P5-6	E ₁ , C	<u>Ala-Pro-Pro-Ser-Val-Phe</u>	P8-10	C, E ₂ , E ₁	<u>Lys-Glx-Ile-Ala</u>
P5-7	E ₁ , C	<u>Val-Ser-Glu-Gly-Gly-Asp</u>	P8-11	C, E ₂	<u>Val-Asx-Phe</u>
P6-1	E ₁ , C	<u>Val-Phe-Lys-Leu</u>	P8-12	C, E ₂	<u>His-Val-Gly-Ala</u>
P6-2	E ₁ , C	<u>Arg-Val-Gly-Asn-Leu</u>	P8-13	C, E ₂	<u>Phe-His-Glx-Trp-Thr-Gly-Asx</u>
P6-3	E ₁ , C	<u>Gly-Val-Gly-Ala</u>	P8-14	C, E ₂	<u>Glx-Asx-Ala-Pro-Glx-Phe</u>
P6-4	E ₁ , C	<u>Ser-Ile-His-Glx</u>	P8-15	C, E ₂	<u>Glx-Lys-Asx-Ala-Trp</u>
P6-5	E ₁ , C	<u>Arg-Ser-Glu-Leu</u>	P8-16	C, E ₂	<u>Ala-Ile-Arg-Tyr</u>
P6-6	E ₁ , C	<u>Leu-Ser-Asp-Leu</u>	P8-17	C, E ₂ , E ₁	<u>Leu-Leu-Glx-Phe-Asx</u>
P6-7	E ₁ , C	<u>Ile-Ala-Asp</u>	P8-18	C, E ₂	<u>Gly-Glx-Phe-Leu</u>
P6-8	E ₁ , C	<u>Leu-Ala-Glu</u>	P9-1	C, E ₁	<u>Tyr-Asx-Glx</u>
P6-9	E ₁ , C	<u>Val-Ser-Glu-Gly</u>	P9-2	C, E ₁	<u>Phe-Glu-Leu-Phe</u>
P6-10	E ₁ , C	<u>Ser-Val-Met</u>	P9-3	C, E ₁	<u>Pro-Glu-Leu-Phe</u>
P8-1	C, E ₂ , E ₁	<u>Phe-Gly-Leu-Tyr-Asx-Glx</u>	P9-4	C, E ₁	<u>Phe-Leu-Phe</u>
P8-2	C, E ₂ , E ₁	<u>Val-Glu-Tyr-Leu</u>	P10-1	C, C	<u>Tyr-Arg-Tyr-Trp-Asx-Thr-Glx</u>
P8-3	C, E ₂	<u>Phe-Val-Leu</u>	P10-2	C, C	<u>Phe-Trp-Gly-Phe</u>
			P10-3	C, C	<u>Phe-Ser-Phe</u>

Fig. 3 (continued). Amino acid sequences of peptides from peptic digestion of carboxymethylated aspartate aminotransferase

E₁, electrophoresis at pH6.5; E₂, electrophoresis at pH3.5; C, chromatography. —→, Residues from dansyl-Edman method. Residues not underlined were assigned as described in the text. States of amidation were assigned on the basis of electrophoretic mobility at pH6.5 as described in Supplementary Publication SUP 50011.

glutamine respectively. Peptide TM-6' is believed to have arisen from partial acid hydrolysis of the bond between aspartic acid/asparagine-4 and proline-5 of peptide TM-6 during the process of demaleylation (Ovchinnikov *et al.*, 1971a). With peptide TM-14 only the first 11 residues were determined by the dansyl-Edman method; the C-terminal sequence was established by digestion with carboxypeptidases A and B and by comparison with the peptic peptide P1-2 (Fig. 3). The central portion of peptide TM-15 was assigned by comparison with the structures of two peptides isolated from an elastase digest of Cm-Enzyme (H. J. Doonan & J. M. Walker, unpub-

lished work); these peptides have the sequences Ala-Asx-Asx-Ser-Ser-Leu-Asx-His-Glx-Tyr-Leu-Pro-Ile and Leu-Gly-Leu-Ala-Glx-Phe-Arg-Thr-Cys-Ala-Ser. The sequence shown for peptide TM-15 is consistent with its amino acid composition.

Peptides from the tryptic digest of aminoethylated aspartate aminotransferase

Fig. 2 shows the amino acid sequences of peptides isolated from the tryptic digest of H₂NET-Enzyme; confirmatory data are in Supplementary Publication

50011. Other peptides were isolated that were identical with or fragments of those obtained from the digest of Cm-Enzyme.

Peptide TAE-11' was a minor product arising from incomplete cleavage of the bond between arginine-2 and leucine-3 of peptide TAE-11. Similarly peptide TAE-12' arose from incomplete cleavage of the bond between lysine-8 and aspartic acid/asparagine-9 of peptide TAE-12. The C-terminal sequence of peptide TAE-12' was assigned from digestion with carboxypeptidases A and B, and the sequence for residues 3-13 corresponds to that of a peptide isolated from a chymotryptic digest of Cm-Enzyme (F. Riva, unpublished work); the residues so far unassigned, as judged from the amino acid composition, are shown in

parentheses. It is noteworthy that, whereas residue 5 of peptide TAE-12 was identified as glutamic acid/glutamine, that at position 4 of the chymotryptic peptide was aspartic acid/asparagine; the Russian workers (Ovchinnikov *et al.*, 1971*b*) give this residue as aspartic acid/asparagine, and the discrepancy remains to be resolved. The central section of peptide TAE-14 was assigned by comparison with the structures of a chymotryptic peptide from Cm-Enzyme (Arg-Thr-Asx-Asx-CmCys-Glx-Pro-Trp) and the N-terminal sequence (Val-Leu-Pro-Val-Val-Arg . . .) of a peptide obtained by cleavage of Cm-Enzyme with CNBr (R. Hanford, unpublished work). The residue at position 14 of peptide TAE-16 was identified by comparison with peptide P3-6 (Fig. 3).

Fig. 4.

Fig. 4 (continued). Composite peptides from overlapping of fragments from peptic and tryptic digests of modified aspartate aminotransferase (Figs. 1, 2 and 3)

Peptides from the peptic digest of carboxymethylated aspartate aminotransferase

The amino acid sequences of most of the peptides isolated from the peptic digest of Cm-Enzyme are shown in Fig. 3. Owing to the low specificity of pepsin, many fragments of these peptides were also obtained. For example, in addition to peptide P8-1, fragments were isolated with the sequences Phe-Gly-Leu, Phe-Gly-Leu-Tyr, Gly-Leu-Tyr and Tyr-Asx-Glx. This multiple fragmentation greatly increased the difficulty of fractionation of the peptide mixture. Where fragments of major peptides were useful in confirming overlaps or in establishing states of amidation of side chains, the sequences have been included in Fig. 3. Each peptide has been given a number designating the fraction after gel filtration on Sephadex

from which it was obtained and a second number for identification purposes. With a few exceptions the amino acid compositions were determined qualitatively by the method of complete dansylation. Where states of amidation have been assigned these were determined from measurements of mobilities at pH 6.5 (Offord, 1966); details are given in Supplementary Publication 50011.

The presence of carboxymethylcysteine and of two glutamic acid/glutamine residues in peptide P1-1 was confirmed by quantitative amino acid analysis. The methionine residues in peptides P1-6 and P3-6 were assigned by comparison with the sequences of peptides TAE-10 and TAE-7 respectively. Peptides P1-7 and P1-8 appeared to be blocked at the *N*-termini and were not degraded by the Edman procedure. The blocked

peptides were cleaved with trypsin and the liberated fragments were degraded in the normal way; cleavage with trypsin probably occurred at arginine residues, since the ϵ -amino groups of any lysine residues would have been blocked by reaction with phenyl isothiocyanate and only unblocked to a small extent by the attempted cyclization. It is noteworthy that whereas tryptophan was found at position 4 of peptide P2-4 the corresponding residue in peptide P2-5 was proline.

Discussion

In spite of extensive studies there is still controversy about some aspects of the molecular structure and properties of the soluble aspartate aminotransferase from pig heart muscle. The molecular weight of the enzyme has not yet been resolved; values of 110000 (Jenkins *et al.*, 1959), 90000–110000 (Polyanovsky & Shpikiter, 1965), 94000 (Martinez-Carrion *et al.*, 1967), 90000 (Feliss & Martinez-Carrion, 1970) and 79000 (Banks *et al.*, 1968b) have been reported. It is, however, agreed that the enzyme dissociates into two subunits at high dilution (Polyanovsky & Ivanov, 1964; Banks *et al.*, 1968a), although there is disagreement about the concentration range over which dissociation occurs and the relative catalytic activities of the monomer and dimer. Dissociation into subunits has also been claimed in response to high and low pH values (Polyanovsky & Shpikiter, 1965), 'succinylation' (i.e. 3-carboxypropionylation) of amino groups (Polyanovsky, 1965), amidation of carboxylate groups (Bossa *et al.*, 1971), treatment with 8M-urea (Banks *et al.*, 1968c) and treatment with guanidine hydrochloride (Feliss & Martinez-Carrion, 1970). It is usually assumed that the two subunits are identical. The evidence for identity of the subunits is that they have the same *N*- and *C*-terminal amino acid residues, that they are of the same molecular weight and that peptide 'maps' of tryptic digests of the dimer show approximately the number of peptides that would be expected on the basis of the molecule consisting of two identical chains (Polyanovsky & Vorotnitskaya, 1965; Martinez-Carrion *et al.*, 1967). Thus, for the initial basis of the present study, it was assumed that the basic unit of the molecule is a polypeptide chain of 350–420 amino acid residues.

A possible complicating factor in the study of the primary structure of soluble aspartate aminotransferase is that the enzyme as isolated is a mixture of so-called 'subforms', i.e. electrophoretically distinct proteins with aspartate aminotransferase activity (Martinez-Carrion *et al.*, 1965). Detailed investigations of the properties of purified samples of the subforms (Martinez-Carrion *et al.*, 1967) have failed to reveal any chemical difference between them, and it has been argued (Banks *et al.*, 1968b) that the subforms differ only in conformation of the polypeptide

chain. Hence it has not been considered necessary to use pure samples of individual subforms for the investigation of the primary structure of the enzyme.

Before the present work and that of Ovchinnikov *et al.* (1971b) little was known about the chemical structure of aspartate aminotransferase. The amino acid composition has been reported by Turano *et al.* (1963), by Martinez-Carrion *et al.* (1967) and by Banks *et al.* (1968b), the results being in close agreement. The large size of the enzyme and the inherent uncertainties in the data, however, preclude a decision between the possible molecular weights on the basis of amino acid composition. The *N*-terminal amino acid sequence has been determined independently by Ovchinnikov *et al.* (1971a) and by Wada *et al.* (1971): in both cases the sequence was found to be Ala-Pro-Pro-Ser-Val-Phe-Ala-Glu-Val. The *C*-terminal sequence has been shown to be Ile-Gln (Wada *et al.*, 1968). Attempts have also been made to determine the amino acid sequence around the lysine residue that is known to be involved in binding the cofactor, pyridoxal 5-phosphate. Hughes *et al.* (1962) showed that reduction of the aldehyde form of the holoenzyme with NaBH₄ resulted in irreversible binding of the cofactor to the enzyme, and they isolated a peptide, after partial digestion of the modified protein, that contained bound cofactor. Subsequently Polyanovsky & Keil (1963) reported the structure Lys(Pxy)-Ser-Asn-Phe for a peptide produced by chymotryptic digestion of the reduced holoenzyme, whereas Morino & Watanabe (1969) claimed the sequence Ser-Lys(Pxy)-Asn-Phe for a peptide produced by the same route; the work by Ovchinnikov *et al.* (1971b) supports the second of these two versions.

Ovchinnikov *et al.* (1971b) have presented preliminary data on peptides from tryptic and chymotryptic digests of aspartate aminotransferase. In all 21 composite peptides were reported, containing 383 amino acid residues. In the present paper we have given the structures of peptides from tryptic and peptic digests of aspartate aminotransferase (Figs. 1, 2 and 3) containing in total 397 distinct amino acid residues. This number may be an over-estimate, since a number of overlaps remain to be established and it is possible that some amino acid residues at the termini of peptides have been counted twice in arriving at the quoted total. Fig. 4 shows the composite peptides (labelled I–XX) that can be derived unambiguously from the fragments listed in Figs. 1, 2 and 3. In some cases (II, VII, XII, XIII and XV) these do not add to the information presented by the Russian workers. In other cases (I, IV, V, VI, VIII, IX, X, XIV, XVII, XVIII and XIX) single amino acid residues or short sequences have been added that extend the previously published sequences and have been of value in obtaining overlaps during subsequent work. Of more interest is peptide XI, which overlaps three tryptic

peptides denoted as TA6-2, TA19-1 and TA18-2 by the Russian workers, and peptide XVI (21 residues), which is missing from the data previously published. The only discrepancies evident between our work and that of the Russian authors are in the state of amidation of side chains. Ovchinnikov *et al.* (1971*b*) give the residues at position 13 of peptide P1-4 as aspartic acid and that at position 6 of peptide TAE-10 as glutamic acid. It will not be possible to determine the origin of these discrepancies until full details of the Russian work are published.

Table 1 shows the amino acid composition of the monomer of cytoplasmic aspartate aminotransferase, the data being averages of values published by Turano *et al.* (1963) and Banks *et al.* (1968*b*) and calculated for a subunit molecular weight of 46000. Also shown are the total numbers of each amino acid found in the tryptic peptides and the combined totals for the tryptic and peptic peptides (column 3). It is apparent that the totals given in column 3 of Table 1 are in reasonable agreement with the numbers expected if the subunit molecular weight is assumed to be 46000. This strongly suggests that the molecular weight of the dimer is about 92000, although a final decision on this point must await the completion of the amino acid sequence. With the tryptic peptides the situation is somewhat different. Although in some instances (e.g. proline, valine, methionine and arginine) the numbers of residues found were the same or only

slightly less than the numbers expected on the basis of the amino acid composition, in other cases (e.g. aspartic acid/asparagine, isoleucine and cysteine) the numbers were considerably less. This must be taken to show that not all of the molecule was found in the form of tryptic peptides rather than to argue that the molecular weight of the protein is smaller than the assumed value. Hence, even after maximizing tryptic cleavage by aminoethylation of the aspartate aminotransferase, approx. 30% of the amino acid residues were contained in peptides that were not amenable to isolation by the methods used. With the maleylated protein the problem was more severe in that it was possible to isolate only 15 peptides, which accounted for a total of 133 amino acid residues including three of lysine (Fig. 1). Ovchinnikov *et al.* (1971*a*) achieved similar results.

At two positions in the amino acid sequences presented here we have suggested double assignments. At position 10 in peptide X (Fig. 4) we have found both proline and tryptophan; similarly at position 6 in peptide TAE-12' (Fig. 2) we have found both glutamic acid/glutamine and aspartic acid/asparagine. In addition, whereas the C-terminal residue of peptide P1-2 was found to be aspartic acid, a similar peptide has been isolated from a thermolytic digest of Cm-Enzyme in which the residue in this position has been identified as valine. These results must, of course, be viewed with caution until confirmatory work has

Table 1. Amino acid composition of cytoplasmic aspartate aminotransferase and of the peptides isolated in the present work

Amino acid	No. of residues		
	Cytoplasmic enzyme (based on subunit mol.wt. 46000)	Total from tryptic peptides	Total from tryptic and peptic peptides
Asx	44	29	40
Thr	26	16	20
Ser	27	16	25
Glx	41	29	43
Pro	23	20	23
Gly	27	19	27
Ala	31	22	30
Val	29	25	32
Met	5	5	5
Ile	18	12	17
Leu	38	26	37
Tyr	12	7	12
Phe	21	9	26
Trp	11	6	9
Lys	20	14	17
His	8	4	7
Arg	25	20	23
Cys	6	2	4
Total	412	281	397

been carried out. If the observations are correct, however, then a plausible explanation would be that the enzyme exists as two allotypic forms. Such a situation has already been described for bovine carboxypeptidase A (Pétra *et al.*, 1969). If a similar situation exists with aspartate aminotransferase, then tissues from individuals in which both forms are produced would be expected to contain three distinct dimeric enzymes (i.e. A₂, AB and B₂, where A and B represent the two varieties of the monomer). As pointed out above, the aspartate aminotransferase as isolated from pig heart does, in fact, consist of three electrophoretically distinct subforms, but these do not appear to be chemically different (Martinez-Carrion *et al.*, 1967). Moreover, evidence has been presented (Banks *et al.* 1968b) that the individuality of the subforms is a function of differences in three-dimensional structure rather than of differences in chemical composition. The problem of the existence of subforms therefore appears to be separate from that of the possible existence of allotypic forms of the enzyme.

We thank Dr. John Mann, Mr. Domenico Dall'Oco and Mr. Antonio Balestrino for technical assistance. Financial support from the Science Research Council (H. J. D. and J. M. W.), the European Molecular Biology Organization (F. R.) and Roche Products Ltd. is gratefully acknowledged.

References

- Ambler, R. P. (1967) *Methods Enzymol.* **11**, 155
- Banks, B. E. C., Bell, M. P., Lawrence, A. J. & Vernon, C. A. (1968a) in *Pyridoxal Catalysis: Enzymes and Model Systems* (Snell, E. E., Braunstein, A. E., Severin, E. S. & Torchinsky, Yu. M., eds.), p. 191, Interscience, New York
- Banks, B. E. C., Doonan, S., Lawrence, A. J. & Vernon, C. A. (1968b) *Eur. J. Biochem.* **5**, 528
- Banks, B. E. C., Doonan, S., Gauldie, J., Lawrence, A. J. & Vernon, C. A. (1968c) *Eur. J. Biochem.* **6**, 507
- Bossa, F., Barra, D., Vecchini, P. & Turano, C. (1971) *Enzymologia* **40**, 369
- Braunstein, A. E. (1970) *FEBS Symp.* **18**, 101
- Braunstein, A. E. & Shemyakin, M. M. (1953) *Biokhimiya* **18**, 393
- Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1969) *Biochem. J.* **112**, 679
- Cole, R. D. (1967) *Methods Enzymol.* **11**, 315
- Crestfield, A. M., Moore, S. & Stein, W. H. (1963) *J. Biol. Chem.* **238**, 622
- Crowshaw, K., Jessup, S. J. & Ramwell, P. W. (1967) *Biochem. J.* **103**, 79
- Edman, P. (1957) *Proc. Roy. Aust. Chem. Inst.* **24**, 434
- Edman, P. (1960) *Ann. N.Y. Acad. Sci.* **88**, 602
- Fasella, P. & Hammes, G. G. (1967) *Biochemistry* **6**, 1798
- Fasella, P. & Turano, C. (1970) *Vitam. Horm. (New York)* **28**, 157
- Feliss, N. & Martinez-Carrion, M. (1970) *Biochem. Biophys. Res. Commun.* **40**, 932
- Gray, W. R. (1967) *Methods Enzymol.* **11**, 469
- Guirard, B. M. & Snell, E. E. (1964) *Compr. Biochem.* **15**, 138
- Hartley, B. S. (1970) *Biochem. J.* **119**, 805
- Hughes, R. C., Jenkins, W. T. & Fischer, E. H. (1962) *Proc. Nat. Acad. Sci. U.S.A.* **48**, 1615
- Ivanov, V. I. & Karpeisky, M. Ya. (1969) *Advan. Enzymol. Relat. Areas Mol. Biol.* **32**, 21
- Jenkins, W. T., Yphantis, D. A. & Sizer, I. W. (1959) *J. Biol. Chem.* **234**, 51
- Martinez-Carrion, M., Riva, F., Turano, C. & Fasella, P. (1965) *Biochem. Biophys. Res. Commun.* **20**, 206
- Martinez-Carrion, M., Turano, C., Chiancone, E., Bossa, F., Giartosio, A., Riva, F. & Fasella, P. (1967) *J. Biol. Chem.* **242**, 2397
- Metzler, D. E., Ikawa, M. & Snell, E. E. (1954) *J. Amer. Chem. Soc.* **76**, 648
- Morino, Y. & Watanabe, T. (1969) *Biochemistry* **8**, 3412
- Offord, R. E. (1966) *Nature (London)* **211**, 591
- Ovchinnikov, Yu. A., Kiryushkin, A. A., Egorov, Ts. A., Abdulaev, N. G., Kiselev, A. P., Modyanov, N. N., Grishin, E. V., Vinogradova, E. I., Feigina, M. Yu., Aldanova, N. A., Lipkin, V. M., Braunstein, A. E. & Polyanovsky, O. L. (1971a) *FEBS Lett.* **12**, 194
- Ovchinnikov, Yu. A., Kiryushkin, A. A., Egorov, Ts. A., Abdulaev, N. G., Kiselev, A. P., Modyanov, N. N., Grishin, E. V., Sukhikh, A. P., Vinogradova, E. I., Feigina, M. Yu., Aldanova, N. A., Lipkin, V. M., Braunstein, A. E., Polyanovsky, O. L. & Nosikov, V. V. (1971b) *FEBS Lett.* **17**, 133
- Pétra, P. H., Bradshaw, R. A., Walsh, K. A. & Neurath, H. (1969) *Biochemistry* **8**, 2762
- Polyanovsky, O. L. (1965) *Biochem. Biophys. Res. Commun.* **19**, 364
- Polyanovsky, O. L. & Ivanov, V. I. (1964) *Biokhimiya* **29**, 728
- Polyanovsky, O. L. & Keil, B. A. (1963) *Biokhimiya* **28**, 372
- Polyanovsky, O. L. & Shpikiter, V. O. (1965) *Dokl. Akad. Nauk SSSR* **163**, 1011
- Polyanovsky, O. L. & Vorotnitskaya, N. E. (1965) *Biokhimiya* **30**, 619
- Sanger, F. & Thompson, E. O. P. (1963) *Biochim. Biophys. Acta* **71**, 468
- Turano, C., Giartosio, A., Riva, F. & Vecchini, P. (1963) in *Chemical and Biological Aspects of Pyridoxal Catalysis* (Snell, E. E., Fasella, P., Braunstein, A. E. & Rossi-Fanelli, A., eds.), p. 149, Pergamon Press, London
- Turano, C., Giartosio, A., Riva, F. & Baroncelli, V. (1967) *Biochem. J.* **104**, 970
- Wada, H., Kagamiyama, H. & Watanabe, T. (1968) in *Symposium on Pyridoxal Enzymes* (Yamada, K., Katunuma, N. & Wada, H., eds.), p. 33, Maruzen, Tokyo
- Wada, H., Watanabe, T. & Miyatake, A. (1971) *Biochem. Biophys. Res. Commun.* **43**, 1318
- Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328
- Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta* **333**, 369