The Primary Structure of Aspartate Aminotransferase from Pig Heart Muscle DIGESTION WITH A PROTEINASE HAVING SPECIFICITY FOR LYSINE RESIDUES

By SHAWN DOONAN,* HILARY J. DOONAN, ROBIN HANFORD, CHARLES A. VERNON,
JOHN M. WALKER and LUCIA P. DA S. AIROLDI†

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

and

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

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

(Received 2 January 1975)

Carboxymethylated aspartate aminotransferase was digested with a proteinase claimed to be specific for lysine residues. Complete cleavage occurred at 12 of the 19 lysine residues in the protein, but at the remaining seven residues cleavage was either restricted or absent. In addition, cleavage was observed at three of the 26 arginine residues. These results are discussed with reference to the amino acid residues adjacent to points of complete or restricted cleavage. The complete primary structure of aspartate aminotransferase, based on these and other studies, is given. Evidence for the assignment of some acid and amide side chains has been deposited as Supplementary Publication SUP 50050 (11 pp.) at the British Library (Lending Division), Boston Spa, Wetherby, W. Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1975) **145**, 5. The evidence for the assignment of residue 366 was less conclusive than for the other acid and amide side chains and is, therefore, given in the main paper.

We have previously reported results obtained by digestion of the soluble (cytoplasmic) form of aspartate aminotransferase (L-aspartate-2-oxoglutarate aminotransferase, EC 2.6.1.1) from pig heart muscle with trypsin, pepsin, thermolysin, elastase, chymotrypsin, CNBr and dilute acid (Doonan *et al.*, 1972; Bossa *et al.*, 1973; Barra *et al.*, 1974; Hanford *et al.*, 1974; Riva *et al.*, 1974). The combined information from these digests left some overlaps unestablished and the structure has now been completed by digestion of the protein with a newly characterized proteinase from the fruiting body of the basidiomycete *Armillaria mellea* (Walton *et al.*, 1972). This enzyme has been claimed to be an amino endopeptidase with specificity for lysine residues on the basis of its action on casein, fibrinogen, the B chain of insulin and other oligopeptides containing lysine, and its potential usefulness in amino acid-sequence analysis has been recognized (P. L. Walton, C. Jones & S. J. Jackson, unpublished work). The results presented here extend available knowledge about the specificity

of the enzyme; a preliminary account of the work has already been published (Doonan *et al.*, 1974).

In our earlier report (Doonan *et al.*, 1974) some of the acid and amide side chains of aspartate aminotransferase was unassigned. The assignments have now been completed, partly by using a new method involving sequence analysis of peptides after reduction with diborane (Airoldi & Doonan, 1975) and partly by the use of mobility measurements (Offord, 1966); data on the mobility measurements are given in Supplementary Publication SUP 50050. The final structure presented here differs from that reported by Ovchinnikov *et al.* (1973) in the assignment of four acid side chains/amide side chain.

Experimental

Sources of most materials and details of experimental methods have been described previously (Doonan *et al.*, 1972; Bossa *et al.*, 1973). Particular procedures used in the present work are outlined below.

Materials

Aspartate aminotransferase was purchased from Whatman Biochemicals, Springfield Mill, Maidstone,

* To whom reprint requests should be addressed at Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

† Present address: Departamento de Bioquímica, Instituto de Biologia, Universidade Estadual de Campinas, 13100 Campinas, São Paulo, Brazil.

Kent, U.K. *A. mellea* proteinase was a generous gift from Dr. P. L. Walton, Pharmaceutical Division, ICI, Alderley Park, Macclesfield, Cheshire, U.K.

Digestion of modified aspartate aminotransferase

The enzyme was carboxymethylated as previously described (Doonan *et al.*, 1972). Cm-Enzyme* (2 g) was suspended in *N*-ethylmorpholine-acetic acid buffer (0.2 M, pH 7.5, 200 ml). *A. mellea* proteinase (15 mg) was added and the mixture incubated under N₂ at 37°C for 24 h. Acetic acid was added to a final concentration of 30% (v/v), and after 30 min insoluble material (about 700 mg) was removed by centrifugation.

Peptide purification

The solution of soluble peptides was concentrated (to 50 ml) by rotary evaporation and applied to a column (5 cm × 100 cm) of Sephadex G-50 equilibrated in acetic acid (30%, v/v). Four main fractions were obtained which were designated AP-1 to AP-4 in order of elution from the column. Peptides in fractions AP-3 and AP-4 were further purified by paper chromatography in the solvent system butan-1-ol-acetic acid-water-pyridine (15 : 3 : 12 : 10, by vol.) (Waley & Watson, 1953) and by high-voltage paper electrophoresis at pH 6.5, 3.5 or 1.9.

Fraction AP-2 was further purified by ion-exchange chromatography on a column (1.5 cm × 30 cm) of sulphopropyl-Sephadex C-25. The resin was equilibrated in sodium formate buffer (0.1 M, pH 2.8) containing NaCl (0.1 M), and the sample was applied in the same solution. Peptides were eluted by application of a linear gradient of NaCl (from 0.1 M-NaCl to 1.1 M-NaCl in a total elution volume of 300 ml).

Insoluble material, formed on acidification of the original digest, was extracted with acetic acid (30%, v/v), when part of the sample dissolved. The solubilized material was fractionated by passage through a column (2.5 cm × 140 cm) of Sephadex G-150 equilibrated in acetic acid (30%, v/v). The major component was further purified by passage through a column (1.5 cm × 90 cm) of Sephadex G-75 equilibrated in the same solvent.

Purity of peptides

Since most of the peptides had *N*-terminal lysine, purity could not be assessed by *N*-terminal analysis. In general a step of Edman degradation was carried out on peptide fractions which appeared to be pure from electrophoretic or chromatographic analysis, and the appearance of a single new *N*-terminal residue was taken to show that the peptide was pure.

* Abbreviations: Cm-Enzyme, carboxymethylated aspartate aminotransferase; CmCys (in sequences), carboxymethylcysteine.

Amino acid analysis and sequence determination

Samples of peptides (25–50 nmol) were hydrolysed as described by Sanger & Thompson (1963), and quantitative analyses carried out by using a Technicon Auto Analyzer. Amino acid sequences were determined by the dansyl-Edman method described by Hartley (1970).

C-Terminal analysis

The peptide (5–10 nmol) was dissolved in *N*-ethylmorpholine-acetic acid buffer (100 µl, 0.2 M, pH 7.5), and carboxypeptidase A (10–20 µg) was added. The solution was incubated at 37°C, and samples (10 µl) were removed at times 0, 10, 20 and 60 min after addition of the enzyme. The samples were dried under vacuum, dansylated and analysed by t.l.c. in the usual way. The order of liberation of amino acids could usually be deduced by inspection of the intensities of the Dns-amino acid spots as a function of time of hydrolysis.

Peptide nomenclature

Peptides have been given the prefix L followed usually by a number denoting the position of the peptide in the structure of aspartate aminotransferase. The letters a and b (as in L-4a and L-4b) denote a pair of peptides arising from cleavage at a point between two lysine residues. Fragments L-A, L-B and L-C contained internal lysine residues at which cleavage did not occur. The symbol AP-C2 denotes the major fraction obtained from the insoluble part of the digest; peptides produced by digestion of fraction AP-C2 with trypsin have been given the symbol Tr.

Digestion of fraction AP-C2 with trypsin

The sample (20 mg) was dissolved in *N*-ethylmorpholine-acetic acid buffer (2 ml, 0.2 M, pH 7.5), and trypsin was added [1 mg; Sigma (London) Chemical Co., London S.W.6, U.K.; type XI]. The solution was incubated at 37°C for 2 h, after which the digest was dried by rotary evaporation. The components of the digest that were soluble in 10% (v/v) acetic acid were purified by high-voltage paper electrophoresis.

Results

Complete or partial amino acid sequences of the peptides isolated after digestion of Cm-Enzyme with *A. mellea* proteinase were determined by using the

Table 1. Amino acid compositions and purification data for peptides obtained by digestion of *Cm*-Enzyme with *A. mellea* proteinase
The compositions predicted from the proposed structures of the peptides are shown in parentheses. Tryptophan was not determined.

Amino acid	L-1	L-2	L-3	L-4b	L-4	L-9	L-13	L-14a	L-15a	L-16	L-17	L-18	L-19	AP-C2	
Carboxymethylcysteine	—	—	0.8 (1)	0.8 (1)	0.8 (1)	—	—	—	—	—	—	0.5 (1)	—	2.1 (2)	
Aspartic acid	—	2.6 (3)	3.2 (3)	4.9 (5)	10.8 (13)	—	1.3 (1)	1.0 (1)	1.1 (1)	2.9 (3)	1.1 (1)	1.1 (1)	1.9 (2)	20.0 (23)	
Threonine	—	—	1.4 (1)	1.2 (1)	8.3 (10)	—	—	0.7 (1)	1.0 (1)	3.2 (4)	—	1.7 (2)	1.1 (2)	10.5 (14)	
Serine	0.8 (1)	—	—	3.5 (4)	6.0 (6)	0.9 (1)	1.7 (2)	1.4 (1)	1.9 (2)	1.1 (1)	—	1.0 (1)	0.8 (1)	15.0 (15)	
Glutamic acid	2.8 (3)	1.0 (1)	1.2 (1)	3.9 (4)	9.2 (9)	0.9 (1)	2.9 (3)	1.3 (1)	1.0 (1)	1.0 (1)	2.7 (3)	—	1.2 (1)	20.5 (21)	
Proline	4.0 (4)	1.7 (2)	2.2 (2)	2.2 (2)	6.0 (6)	—	1.2 (1)	1.2 (2)	—	1.9 (2)	—	1.0 (1)	—	12.1 (11)	
Glycine	—	—	2.0 (2)	2.3 (2)	11.0 (12)	—	—	1.0 (1)	—	3.1 (3)	—	2.0 (2)	—	15.6 (18)	
Alanine	3.6 (3)	1.0 (1)	1.1 (1)	4.0 (5)	7.0 (7)	1.1 (1)	—	1.8 (2)	2.0 (2)	—	—	—	1.2 (2)	17.0 (17)	
Valine	2.1 (4)	—	4.0 (5)	—	6.2 (6)	0.9 (1)	0.5 (1)	2.0 (2)	—	—	0.9 (1)	—	1.3 (2)	14.2 (15)	
Methionine	—	—	—	—	—	0.8 (1)	0.9 (1)	0.2 (1)	0.3 (2)	0.2 (1)	—	0.2 (1)	—	0.8 (1)	
Isoleucine	—	1.0 (1)	—	1.8 (2)	4.0 (3)	—	—	1.0 (1)	0.9 (1)	1.4 (2)	0.8 (1)	2.9 (3)	1.3 (1)	7.0 (7)	
Leucine	0.8 (1)	1.0 (1)	2.2 (2)	6.7 (7)	9.1 (8)	—	1.3 (2)	1.0 (1)	2.0 (2)	1.1 (1)	1.0 (1)	2.9 (3)	1.3 (1)	18.8 (17)	
Tyrosine	—	—	0.9 (1)	0.5 (1)	3.3 (4)	—	—	—	—	—	0.8 (1)	0.3 (1)	0.8 (1)	7.1 (7)	
Phenylalanine	1.2 (2)	1.0 (1)	—	1.1 (1)	5.4 (6)	—	—	—	—	1.7 (2)	—	—	—	17.8 (17)	
Lysine	—	1.0 (1)	1.0 (1)	—	4.0 (4)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	10.8 (9)	
Histidine	—	1.1 (1)	—	2.8 (3)	—	—	—	—	—	0.7 (1)	—	0.8 (1)	0.4 (1)	3.4 (4)	
Arginine	—	1.8 (2)	1.9 (2)	2.0 (2)	6.7 (6)	—	0.8 (1)	2.5 (2)	4.0 (4)	—	—	1.2 (1)	—	13.9 (14)	
Purification procedures															
Elution range from Sephadex G-50 (ml)	1000-1370	1000-1370	1000-1370	700-1000*	475-700†	1000-1370	1000-1370	1000-1370	1000-1370	1000-1370	1000-1370	1000-1370	1000-1370	1000-1370	—†
Mobility at pH6.5 relative to arginine	0.10-0.26	0.10-0.26	0.14-0.21	—	—	0.37-0.57	0.10-0.26	0.57-0.73	0.57-0.73	0.14-0.21	0.10-0.26	0.37-0.57	0.10-0.26	—	
R _F value on chromatography	0.91	0.52	0.64	—	—	0.24	0.38	0.56	0.16	0.64	0.38	0.55	0.66	—	
Mobility at pH1.9 relative to arginine	—	1.10	—	—	—	—	0.75	—	—	—	0.62	0.79	—	—	

Purification procedures

Elution range from Sephadex G-50 (ml) 1000-1370 1000-1370 1000-1370 700-1000* 475-700† 1000-1370 1000-1370 1000-1370 1000-1370 1000-1370 1000-1370 1000-1370 1000-1370 —†

Mobility at pH 6.5 relative to arginine 0.10-0.26 0.10-0.26 0.14-0.21 — — — 0.37-0.57 0.10-0.26 0.57-0.73 0.14-0.21 0.10-0.26 0.37-0.57 0.10-0.26 —

R_F value on chromatography 0.91 0.52 0.64 — — — 0.24 0.38 0.56 0.16 0.64 0.55 0.66 —

Mobility at pH 1.9 relative to aspartic acid — 1.10 — — — — 0.75 0.75 — — — 0.62 0.79 —

* Purified further by ion-exchange chromatography. Eluted with 0.43M-NaCl (see the Experimental section).

† Not purified further.

‡ Purified by passage through Sephadex G-150 (peak maximum = 450ml) followed by Sephadex G-75 (peak maximum = 115ml) (see the Experimental section).

Peptide	Position in sequence	Sequence data
L-1	1-18	Ala-Pro-Pro-Ser-Val-Phe-Ala <u>Leu-Val-Phe</u>
L-2	19-31	Lys-Leu-Ile-Ala-Asx-Phe-Arg-Glx-Asx....
L-3	32-54	Lys-Val-Asx-Leu-Gly-Val-Gly....
L-4a	55-59	Lys-Val-Glx-Glx-Arg
L-4b	60-97	Ile-Ala-Asx-Asx-Ser-Ser-Leu-Asx-His-Glx-Tyr- Leu-Pro-Ile-Leu-Gly-Leu-Ala-Glx-Phe....
L-A	98-205	Lys-Arg....
L-8	165-(205)	Lys-Arg-Gly-Leu-Asx-Leu-Glx-Gly-Phe-Leu- Ser....
L-9	206-212	Lys-Glx-Ile-Ala-Ser-Val-Met
L-B	213-(323)	Lys-Arg-Arg-Phe-Leu-Phe-Pro-Phe-Phe-Asx- Ser....
L-C	213-274	Lys-Arg-Arg-Phe..... <u>Val-Ala</u>
L-13	275-288	Lys-Glx-Pro-Asx-Ser-Ile-Leu-Arg-Val- Leu-Ser-Glx-Met-Glx
L-14a	289-304	Lys-Ile-Val-Arg-Val-Thr-(Trp)-Ser-Asx- Pro-Pro-Ala-Glx....
L-15a	324-340	Lys-Thr-Met-Ala-Asx-Arg-Ile....
L-16	345-367	Lys-Thr-Pro-Gly-Thr-(Trp)-Asx-His-Ile- Thr-Asx-Glx-Ile-Gly-Met-Phe-Ser-Phe- Thr-Gly-Leu-Asx-Pro
L-17	368-376	Lys-Glx-Val-Glx-Tyr-Leu-Ile-Asx-Glx
L-18	377-394	Lys-His-Ile-Tyr-Leu-Leu-Pro-Ser-Gly- Arg-Ile....
L-19	395-409	Lys-Asx-Leu-Asx-Tyr-Val-Ala....
L-20	410-412	Lys-Ile-Glx

Fig. 1

Peptide	Position in sequence	Sequence
Tr-1	100-113	Val-Gly-Gly-Val-Glx-Ser-Leu-Gly- Gly-Thr-Gly-Ala-Leu-Arg
Tr-2	114-121	Ile-Gly-Ala-Glx-Phe-Leu-Ala-Arg
Tr-3	154-156	Asx-Ile-Arg
Tr-4	157-159	Ser-Tyr-Arg
Tr-5	216-235	Phe-Leu-Phe-Pro-Phe-Phe-Asx-Ser-Ala- Tyr-Glx-Gly-Phe-Ala-Ser-Gly-Asx-Leu- Glx-Lys
Tr-6	242-258	Tyr-Phe-Val-Ser-Glx-Gly-Phe-Glx- Leu-Phe-CmCys-Ala-Glx-Ser-Phe-Ser-Lys
Tr-7	259-266	Asx-Phe-Gly-Leu-Tyr-Asx-Glx-Arg
Tr-8	267-275	Val-Gly-Asx-Leu-Thr-Val-Val-Ala-Lys
Tr-9	276-282	Glx-Pro-Asx-Ser-Ile-Leu-Arg
Tr-10	290-292	Ile-Val-Arg
Tr-11	293-304	Val-Thr-(Trp)-Ser-Asx-Pro-Pro-Ala- Glx-Gly-Ala-Arg
Tr-12	305-308	Ile-Val-Ala-Arg
Tr-13		Free Lys and Arg

Fig. 2. Amino acid sequences of peptides isolated after digestion of fraction AP-C2 with trypsin

All the peptides shown were completely sequenced by using the dansyl-Edman method.

Fig. 1. Amino acid-sequence data for peptides isolated after digestion of Cm-Enzyme with *A. mellea* proteinase

The residues shown were identified by the dansyl-Edman method except for those marked $\leftarrow$, which were liberated by carboxypeptidase A and identified as the dansyl derivatives, and the tryptophan residues shown in parentheses, which were positioned by comparison with known regions of the structure. The symbol indicates that the peptide was not completely sequenced. The positions in the sequence of the C-terminal residues of peptides L-8 and L-B are given in parentheses since they were not established with certainty (see the text).

Fig. 3

Fig. 3. Complete amino acid sequence of aspartate aminotransferase showing the positions of peptides produced by cleavage with *A. mellea* proteinase

The numbering system for the peptides is described in the text. Peptides, or parts of peptides, marked with full lines were sequenced by the dansyl-Edman method. Sequences marked with dashed lines were assigned from amino acid compositions and by comparison with known sequences.

dansyl-Edman method, and in some cases by digestion with carboxypeptidase A. The parts of the sequences so determined, and the positions of the peptides in the total structure of the enzyme are shown in Figs. 1 and 3 respectively. The amino acid compositions of most of these peptides are given in Table 1. In general the amino acid compositions were consistent with the sequence data (where these were complete) or with the structures proposed for peptides on the basis of a combination of sequence data and comparison with the known primary structure of the enzyme (Doonan *et al.*, 1974). In some cases (e.g., L-1, L-3, L-13) the yields of isoleucine, leucine, phenylalanine and valine were low, but this could be correlated with the occurrence of sequences in these peptides that were likely to be resistant to acid hydrolysis. Yields of methionine were also low in some cases, presumably because of losses by oxidation. Amino acid compositions were not determined for the small peptides L-4a and L-20, the sequence analyses of which were very easy, or for peptides L-8 and L-C which were obtained in insufficient amount.

Fraction AP-3 from gel filtration contained peptides L-1, L-2, L-3, L-4a, L-9, L-15a, L-16, L-17, L-18 and L-19 in high yields. Fraction AP-4 contained mainly peptide L-20 with smaller amounts of peptides L-4a and L-9. Ion-exchange chromatography of fraction AP-2 gave peptide L-4b as the major component, again in high yield. These peptides between them contained all the amino acid residues from positions 1-97, 206-212 and 324-412 (except for residues 341-344). No other peptides were found which contained any of these residues and hence it appeared that the cleavages which gave rise to the peptides were essentially complete.

The remaining parts of the molecule, i.e. residues 98-205 and 213-323, were digested to a relatively small extent. Fraction AP-1 from gel filtration contained a single component, peptide L-A, with the *N*-terminal sequence Lys-Arg; it proved impossible to sequence the peptide further by the dansyl-Edman procedure owing to lack of solubility. Amino acid analysis (Table 1) suggested that peptide L-A contained residues 98-205 and had three internal lysine residues, although agreement between the observed and predicted analyses was not exact. The only other fragment found originating from this part of the molecule was peptide L-8 which was obtained in very small quantity. The *N*-terminal sequence (Fig. 1) showed that the peptide started at residue 165 and from the specificity of *A. mellea* proteinase, combined with the fact that peptide L-9 (residues 206-212) was obtained in high yield, it was assumed to terminate at residue 205 (Fig. 3).

The situation was similar for the region of the molecule from residues 213 to 323. A small amount of a peptide, L-C, was obtained which had the *N*-terminal Lys-Arg-Arg-Phe (residues 213-216). In-

sufficient material was available for amino acid analysis, but treatment with carboxypeptidase A gave the *C*-terminal sequence Val-Ala; this indicated that the peptide terminated at residue 274 (Fig. 3). Similarly, peptides L-13 and L-14a were obtained as minor components of fraction AP-3 but in quantities just sufficient for both sequence analysis and amino acid analysis. Available evidence suggests that the bulk of the material representing residues 213-323 was present as a single fragment (L-B). A fraction (AP-C2) was isolated from the insoluble part of the initial digest and was shown to have the *N*-terminal sequence Lys-Arg-Arg-Phe-Leu-Phe-Pro-Phe-Phe-Asx-Ser. Fraction AP-C2 was digested with trypsin and the soluble components of the tryptic digest were isolated and their complete amino acid sequences determined by using the dansyl-Edman method (Fig. 2). The tryptic peptides covered all the residues from positions 213 to 308 except for 236-241 and 283-289; the latter peptides were probably in the insoluble part of the tryptic digest. The *C*-terminal residue of peptide L-B was assumed to be 323 rather than 308 on the basis of the specificity of *A. mellea* proteinase and the fact that the residues from positions 309 to 323 were not found in any other fraction of the digest; the tryptic peptide containing these residues was presumably insoluble. The interpretation of these results was, however, complicated by the fact that the tryptic digest contained peptides clearly originating from peptide L-A in amounts comparable with those of the peptides from L-B (Fig. 2). This suggested that fraction AP-C2 contained roughly equal amounts of peptides L-A and L-B. Amino acid analysis of the fraction (Table 1) gave results consistent with this proposal although this may have been coincidental; the values given in parentheses for fraction AP-C2 in Table 1 are the sums for all residues in peptides L-A and L-B. The presence of peptide L-A in fraction AP-C2 went undetected on *N*-terminal sequence analysis since, as pointed out above, only the first two amino acid residues of peptide L-A (Lys-Arg) were identifiable by the dansyl-Edman method, and peptide L-B has the same *N*-terminal sequence.

Peptide L-16 (residues 345-367) was used to determine the state of amidation of residue 366. The peptide was digested with chymotrypsin and the digest subjected to electrophoresis at pH 6.5. The neutral band (mobility -0.05) contained peptides with *N*-terminal residues aspartic acid/asparagine, serine, threonine and phenylalanine; complete dansylation of the hydrolysate of the mixture of neutral peptides showed the presence of proline. These results suggested that the neutral band contained a *C*-terminal fragment from peptide L-16 which arose by cleavage either between residues 18 and 19 or between residues 21 and 22 to yield peptides Thr-Gly-Leu-Asx-Pro or Asx-Pro respectively;

in either case the penultimate residue of L-16 would be assigned as glutamine. For unambiguous proof of this point it would have been preferable to isolate the C-terminal fragment from the mixture of neutral peptides, but this was not possible because of the small amount of material available.

Discussion

It is widely recognized that sequence studies of proteins, particularly those with long polypeptide chains, are hampered by lack of methods for specific but limited chain cleavage (see, e.g., Kolb *et al.*, 1974). One solution to this problem is provided by the proteinase from *A. mellea* since the enzyme appears to have a major specificity for lysine residues. Protein cleavage with this proteinase will obviously be of particular value when used in conjunction with the established technique of restricting tryptic cleavage to arginine residues by reversibly blocking lysine side chains (Butler *et al.*, 1969). It is noteworthy that a previous report of a proteinase specific for lysine residues (Wingard *et al.*, 1972) does not appear to have led to applications in primary structure analysis.

The results presented here show that residues adjacent to lysine affect the extent of hydrolysis by *A. mellea* proteinase. Essentially complete cleavage was observed N-terminal to residues 19, 32, 55, 98, 206, 213, 324, 345, 368, 377, 395 and 410; no peptides were found to contain any of these lysine residues internally. Cleavage was not affected by the presence of arginine on either side of the lysine residue (positions 32, 55, 98 and 213) or by proline on the N-terminal side (position 368); Shipolini *et al.* (1974) have shown, however, that cleavage does not occur when proline is C-terminal to the lysine residue. The work reported by Shipolini *et al.* (1974) is of particular note in that the substrate protein was S-aminoethylated and it was found that *A. mellea* proteinase also cleaved the modified protein on the N-terminal side of aminoethylcysteine residues.

Little cleavage occurred at three of the lysine residues of aspartate aminotransferase (at positions 165, 275 and 289) and apparently none at the remaining four (at positions 129, 153, 235 and 258). These seven lysine residues are grouped in two regions of the protein and as a result, most of the material constituting residues 98–205 and 213–323 was present as two undigested fragments (peptides L-A and L-B). Of the four lysine residues at which no cleavage occurred, three have aspartic acid on the C-terminal side. Similarly two of the residues at which cleavage was restricted have adjacent glutamic acid at the N- or C-terminal sides. These observations suggest that adjacent acidic side chains restrict or prevent hydrolysis, particularly aspartic acid at the C-terminal side, but the situation is complicated by the fact that cleavage was also restricted in the sequences Gln-

Lys-Ile (288–290) and Ser-Lys-Asn (257–259); on the other hand, complete cleavage occurred at the sequence Glu-Lys-Arg (97–99). It may be that the bonds linking peptides L-A and L-B to the rest of the molecule were hydrolysed very rapidly and that hydrolysis at internal lysine residues in these peptides was restricted by their insolubility. The dependence of rate of cleavage on adjacent acidic side chains obviously requires further study.

Cleavage of the protein was also observed at three arginine residues (positions 59, 304 and 340) and it is noteworthy that in these cases hydrolysis occurred at the C-terminal side of the basic residue. Cleavage at positions 59 and 340 must have been essentially complete since the parent lysine peptides were not found. The degree of cleavage at position 304 is difficult to assess since most of the material from residues 289–323 was present in peptide L-B. The common feature of the arginine residues at which cleavage occurred was a leucine or isoleucine residue at the C-terminal side. It should be noted, however, that aspartate aminotransferase contains four other arginine residues followed either by leucine or isoleucine (positions 85, 113, 229 and 386) at which hydrolysis did not occur. The factors governing cleavage at arginine must, therefore, be more complex than simply the nature of the residue at the C-terminal side.

The results given here were of great value in establishing the final structure of aspartate aminotransferase in conjunction with data reported earlier (Doonan *et al.*, 1972; Bossa *et al.*, 1973; Barra *et al.*, 1974; Hanford *et al.*, 1974; Riva *et al.*, 1974). There is no doubt that, had *A. mellea* proteinase been available at an earlier stage, the sequence analysis would have been greatly simplified. The final structure obtained is shown in Fig. 3; assignment of some of the acid and amide side chains is described in Supplementary Publication SUP 50050. The structure agrees essentially with that reported by Ovchinnikov *et al.* (1973) except that at four positions (63, 144, 288 and 376) we have found amide side chains whereas Ovchinnikov *et al.* and (1973) found the corresponding acids. The four residues in question were identified as the phenylthiohydantoins by the Russian group (Vinogradova *et al.*, 1973); this method should give unambiguous assignments but it is possible that the residues were deaminated during isolation of the respective peptides. It has been suggested, however (John & Jones, 1974), that the so-called sub-forms of cytoplasmic aspartate aminotransferase differ in degree of amidation and this may be the origin of the discrepancies between the two versions of the structure, since in both cases samples of the enzyme containing mixtures of sub-forms were used for sequence studies.

The structure given in Fig. 3 may not be unique in that there is evidence for alternative residues at four

positions in the polypeptide chain; these are at position 127 (asparagine/'Glx'), position 283 (valine/aspartic acid) and position 347 (proline/tryptophan) (Doonan *et al.*, 1972) and at position 157 (serine/alanine) (Bossa *et al.*, 1973; Riva *et al.*, 1974). The residues given at these positions in Fig. 3 are those found in most digests or, when both versions were obtained in the same digest, that occurring in the peptide obtained in greater amount. For example, residue 347 was found as proline in peptides from the digests with elastase (Bossa *et al.*, 1973) and chymotrypsin (Barra *et al.*, 1974) and both as proline and tryptophan in peptides from the digests with *A. mellea* proteinase and with pepsin (Doonan *et al.*, 1972). In addition to these possible substitutions, evidence has been obtained for a three-residue deletion. Two peptides isolated from the chymotryptic digest (Ch-38 and Ch-38'; Barra *et al.*, 1974) had the sequences Val-Ala-Thr-Ser-Ile-His-Glx-Ala-Val-Thr-Lys-Ile-Glx and Val-Ala-Thr-Ser-Ala-Val-Thr-Lys-Ile-Glx. The shorter was obtained in low yield. In the light of these results, population studies of the enzyme would be of interest, although it should be borne in mind that most of the substitutions and the deletion described above would not be detectable by the usual electrophoretic techniques. The only work of this sort so far reported was a study of human placental mitochondrial and cytoplasmic aspartate aminotransferases (Hackel *et al.*, 1972) in which variants of the mitochondrial but not of the cytoplasmic enzyme were detected.

Financial support from the Science Research Council (to H. J. D., R. H. and J. M. W.) is gratefully acknowledged.

References

- Aioldi, L. P. da S. & Doonan, S. (1975) *FEBS Lett.* **50**, 155–158
- Barra, D., Bossa, F., Carloni, M., Fasella, P., Martini, F., Petruzzelli, R., Riva, F., Doonan, S., Doonan, H. J., Hanford, R. & Walker, J. M. (1974) *Atti Accad. Naz. Lincei Rend.* **56**, 589–598
- Bossa, F., Barra, D., Carloni, M., Fasella, P., Riva, F., Doonan, S., Doonan, H. J., Hanford, R., Vernon, C. A. & Walker, J. M. (1973) *Biochem. J.* **133**, 805–819
- Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1969) *Biochem. J.* **112**, 679–689
- Doonan, S., Doonan, H. J., Riva, F., Vernon, C. A., Walker, J. M., Bossa, F., Barra, D., Carloni, M. & Fasella, P. (1972) *Biochem. J.* **130**, 443–452
- Doonan, S., Doonan, H. J., Hanford, R., Vernon, C. A., Walker, J. M., Bossa, F., Barra, D., Carloni, M., Fasella, P., Riva, F. & Walton, P. L. (1974) *FEBS Lett.* **38**, 229–233
- Hackel, E., Hopkinson, D. A. & Harris, H. (1972) *Ann. Hum. Genet.* **35**, 491–496
- Hanford, R., Doonan, H. J., Doonan, S., Vernon, C. A., Walker, J. M., Bossa, F., Barra, D., Carloni, M., Fasella, P. & Riva, F. (1974) *Att. Accad. Naz. Lincei Rend.* **56**, 73–83
- Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
- John, R. & Jones, R. (1974) *Biochem. J.* **141**, 401–406
- Kolb, E., Harris, J. I. & Bridgen, J. (1974) *Biochem. J.* **137**, 185–197
- Offord, R. E. (1966) *Nature (London)* **211**, 591–593
- Ovchinnikov, Yu. A., Egorov, Ts. A., Aldanova, N. A., Feigina, M. Yu., Lipkin, V. M., Abdulaev, N. G., Grishin, E. V., Kiselev, A. P., Modyanov, N. N., Braunstein, A. E., Polyakovskiy, O. L. & Nosikov, V. V. (1973) *FEBS Lett.* **29**, 31–34
- Riva, F., Barra, D., Bossa, F., Carloni, M., Fasella, P., Doonan, S., Doonan, H. J., Hanford, R. & Walker, J. M. (1974) *Atti Accad. Naz. Lincei Rend.* **56**, 581–588
- Sanger, F. & Thompson, E. O. P. (1963) *Biochim. Biophys. Acta* **71**, 468–471
- Shipolini, R. A., Callewaert, G. L., Cottrell, R. C. & Vernon, C. A. (1974) *Eur. J. Biochem.* **48**, 465–476
- Vinogradova, E. I., Feigina, M. Yu., Aldanova, N. A., Lipkin, V. M., Smirnov, Yu. V., Potapenko, N. A., Abdulaev, N. G., Kiselev, A. P., Egorov, Ts. A. & Ovchinnikov, Yu. A. (1973) *Biokhimiya* **38**, 3–21
- Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337
- Walton, P. L., Turner, R. W. & Broadbent, D. (1972) *Br. Patent* 1263956
- Wingard, M., Matsueda, G. & Wolfe, R. S. (1972) *J. Bacteriol.* **112**, 940–949