# THE COMPLETE AMINO ACID SEQUENCE OF CYTOPLASMIC ASPARTATE AMINOTRANSFERASE FROM PIG HEART*

Yu.A. OVCHINNIKOV, C.A. EGOROV, N.A. ALDANOVA, M.Yu. FEIGINA, V.M. LIPKIN, N.G. ABDULAEV, E.V. GRISHIN, A.P. KISELEV and N.N. MODYANOV
*Shemyakin Institute for Chemistry of Natural Products, USSR Academy of Sciences, Moscow, USSR*

and

A.E. BRAUNSTEIN, O.L. POLYANOVSKY and V.V. NOSIKOV
*Institute of Molecular Biology, USSR Academy of Sciences, Moscow, USSR*

Received 7 November 1972

## 1. Introduction

Aspartate aminotransferase (L-aspartate:2-oxoglutarate aminotransferase, EC 2.6.1.1) is one of the principal pyridoxal-*P*-containing enzymes that catalyse the transamination reactions [3] representing key steps at the intersection between the metabolic pathways of amino acids and dicarboxylic acids.

Although the catalytic mechanism of aspartate aminotransferase has been investigated at the level of substrate–coenzyme models [4], its elucidation in detail requires knowledge of the enzyme's structure, considering, in particular, that the very high rates of the enzymic process are determined by the structural peculiarities of the specific protein(apoenzyme) of the aspartate aminotransferase. Accordingly, we embarked on the task of elucidating the amino acid sequence of this enzyme. In the present paper the concluding stage of the work is reported*.

The object chosen for study was the aspartate aminotransferase of the cytosol of pig heart; the enzyme, which is different from the mitochondrial isozyme [5, 6] was prepared by a previously reported procedure [7]. The enzyme is a complex dimeric protein of high molecular weight; each of the associated subunits consists of a single polypeptide chain and has no disulfide bridges. Indirect evidence (amino acid composition, analysis of N-terminal residues, and peptide maps) testified to the identity of the two subunits [8].

## 2. Experimental

The complete amino acid sequence of the enzyme, shown in fig. 1, was resolved by the combined use of conventional and newer analytical methods. The following were the main steps of the analysis. The polypeptide chain of the protein was split into peptide fragments by either enzymic or chemical means, namely – tryptic, chymotryptic and thermolytic hydrolyses, partial acid hydrolysis and cyanogen bromide cleavage. In addition to ordinary tryptic hydrolysis [9] use was also made of restricted tryptic hydrolysis of the protein following preliminary acylation of the lysine residues [10]. The tryptic cleavage of both the non modified and modified (acylated) protein permitted the isolation and identification of all the peptides constituting the chain of aspartate aminotransferase. The other cleavage procedures were utilized mainly to obtain overlapping sequences. Thus, from the chymotryptic digests of aspartate aminotransferase 68 peptides were isolated, accounting for 341 amino acid residues. Six of the seven expected peptide fragments were isolated from

* For preceding communications see [1, 2].

```
                                        10                                  20
Ala-Pro-Pro-Ser-Val-Phe-Ala-Glu-Val-Pro-Gln-Ala-Gln-Pro-Val-Leu-Val-Phe-Lys-Leu-Ile-Ala-Asp-Phe-Arg-
                30                                  40                                  50
Glu-Asp-Pro-Asp-Pro-Arg-Lys-Val-Asn-Leu-Gly-Val-Gly-Ala-Tyr-Arg-Thr-Asp-Asp-Cys-Gln-Pro-Trp-Val-Leu-
                                        60                                  70
Pro-Val-Val-Arg-Lys-Val-Glu-Gln-Arg-Ile-Ala-Asn-Asp-Ser-Ser-Leu-Asn-His-Glu-Tyr-Leu-Pro-Ile-Leu-Gly-
                80                                  90                                  100
Leu-Ala-Glu-Phe-Arg-Thr-Cys-Ala-Ser-Arg-Leu-Ala-Leu-Gly-Asp-Asp-Ser-Pro-Ala-Leu-Gln-Glu-Lys-Arg-Val-
                                        110                                 120
Gly-Gly-Val-Gln-Ser-Leu-Gly-Gly-Thr-Gly-Ala-Leu-Arg-Ile-Gly-Ala-Glu-Phe-Leu-Ala-Arg-Trp-Tyr-Asn-Gly-
                130                                 140                                 150
Thr-Asn-Asn-Lys-Asp-Thr-Pro-Val-Tyr-Val-Ser-Ser-Pro-Thr-Trp-Glu-Asn-His-Asp-Gly-Val-Phe-Thr-Thr-Ala-
                                        160                                 170
Gly-Phe-Lys-Asp-Ile-Arg-Ser-Tyr-Arg-Tyr-Trp-Asp-Thr-Glu-Lys-Arg-Gly-Leu-Asp-Leu-Gln-Gly-Phe-Leu-Ser-
                180                                 190                                 200
Asp-Leu-Glu-Asn-Ala-Pro-Glu-Phe-Ser-Ile-Phe-Val-Leu-His-Ala-Cys-Ala-His-Asn-Pro-Thr-Gly-Thr-Asp-Pro-
                                        210                                 220
Thr-Pro-Glu-Gln-Trp-Lys-Gln-Ile-Ala-Ser-Val-Met-Lys-Arg-Arg-Phe-Leu-Phe-Pro-Phe-Phe-Asp-Ser-Ala-Tyr-
                230                                 240                                 250
Gln-Gly-Phe-Ala-Ser-Gly-Asn-Leu-Glu-Lys-Asp-Ala-Trp-Ala-Ile-Arg-Tyr-Phe-Val-Ser-Glu-Gly-Phe-Glu-Leu-
                                        260                                 270
Phe-Cys-Ala-Gln-Ser-Phe-Ser-Lys-Asn-Phe-Gly-Leu-Tyr-Asn-Glu-Arg-Val-Gly-Asn-Leu-Thr-Val-Val-Ala-Lys-
                280                                 290                                 300
Glu-Pro-Asp-Ser-Ile-Leu-Arg-Val-Leu-Ser-Gln-Met-Glu-Lys-Ile-Val-Arg-Val-Thr-Trp-Ser-Asn-Pro-Pro-Ala-
                                        310                                 320
Gln-Gly-Ala-Arg-Ile-Val-Ala-Arg-Thr-Leu-Ser-Asp-Pro-Glu-Leu-Phe-His-Glu-Trp-Thr-Gly-Asn-Val-Lys-Thr-
                330                                 340                                 350
Met-Ala-Asp-Arg-Ile-Leu-Ser-Met-Arg-Ser-Glu-Leu-Arg-Ala-Arg-Leu-Glu-Ala-Leu-Lys-Thr-Pro-Gly-Thr-Trp-
                                        360                                 370
Asn-His-Ile-Thr-Asp-Gln-Ile-Gly-Met-Phe-Ser-Phe-Thr-Gly-Leu-Asn-Pro-Lys-Gln-Val-Glu-Tyr-Leu-Ile-Asn-
                380                                 390                                 400
Glu-Lys-His-Ile-Tyr-Leu-Leu-Pro-Ser-Gly-Arg-Ile-Asn-Met-Cys-Gly-Leu-Thr-Thr-Lys-Asn-Leu-Asp-Tyr-Val-
                                        410
Ala-Thr-Ser-Ile-His-Glu-Ala-Val-Thr-Lys-Ile-Gln
```

Fig. 1. The complete amino acid sequence of cytoplasmic aspartate aminotransferase from pig heart.

the products of cyanogen bromide degradation. The N-terminal sequence of the protein up to the ninth residue was determined on a carboxymethylated specimen of the enzyme with the aid of an AP-01 Sequenator constructed at the Analytical Instruments Designing Bureau of the USSR Academy of Sciences. Carboxypeptidases A and B were employed for determining the C-terminal sequence, the results being subsequently confirmed by isolation of isoleucyl-glutamine in high yield from the tryptic hydrolysate of the enzyme.

Isolation and purification of the peptides was achieved by a variety of different techniques including column chromatography with high-resolution ion-exchange resins and celluloses, gel chromatography, electrophoresis and paper chromatography. Structures of the simpler peptides were resolved either directly by Edman degradation, or by the modified technique combining stepwise Edman cleavage with dansylation of the residual peptides and identification of the N-terminal dansyl amino acids [11, 12]. Carboxy- and aminopeptidases were also used for this purpose. Larger peptides were first digested with appropriate proteases, and the resulting fragments were then analysed by the methods indicated above. Mass-spectrometry [13] proved highly convenient for the sequence determination of a large number of smaller peptides (up to octapeptides) obtained by thermolytic digestion of aspartate aminotransferase.

Considerable difficulties were encountered in resolving the structure of this enzyme, mainly on account of the large size of its polypeptide chain, and of peculiarities of the amino acid sequence. For example, the high tendency of the larger peptides for aggregation was a serious obstacle to separation and purification of the aspartate aminotransferase degradation products; this hindrance could be overcome only by the utilization of special procedures (acylation of the lysine residues with maleic or citraconic anhydride; use of concentrated urea solutions, etc.).

We also found that the cleavage by cyanogen bromide is far from complete at methionine residues preceded by a hydroxy amino acid (e.g. at the sequence Ser–Met (residues 332–333) in fig. 1). Furthermore, it was observed in the course of our work that in acid medium (pH ~ 3) there occurred partial hydrolysis of aspartyl–proline bonds. Due to these complications poor yields were obtained of some of the peptidic products, in particular following cyanogen bromide degradation or tryptic cleavage (in the case of the modified-lysine protein), thus causing additional difficulties in purification of the fragments.

It is appropriate to point out that, in general, the methodology for primary structure determination of proteins with long polypeptide chains is a problem still in an early stage of maturity, the experience hitherto gained being confined to elucidation of the structures of glutamate dehydrogenase [14] and γG-immunoglobulin [15]. Therefore the knowledge acquired in the present work may be of value in elaboration of general approaches to sequence studies on polypeptide chains of high molecular weight.

## 3. Results and discussion

Aspartate aminotransferase is apparently the first pyridoxal-*P*-dependent enzyme the complete peptide sequence of which has been resolved. The results of the present work prove that the two subunits of the enzyme are identical, and that the observed electrophoretic heterogeneity of aspartate aminotransferase preparations cannot be due to differences in amino acid sequence of the components isolated. From the structure shown in fig. 1 it is seen that each subunit consists of 412 residues and has the following amino acid composition: $Lys_{19}$, $His_8$, $Arg_{26}$, $(Asp+Asn)_{42}$, $Thr_{25}$, $Ser_{26}$, $(Glu+Gln)_{41}$, $Pro_{24}$, $Gly_{28}$, $Ala_{32}$, $Cys_5$, $Val_{29}$, $Met_6$, $Ile_{19}$, $Leu_{38}$, $Tyr_{12}$, $Phe_{23}$, $Trp_9$. Hence the molecular weight of one polypeptide chain is 46 344 and that of the aspartate aminotransferase holoenzyme (including 2 moles of pyridoxal-*P*) is 93 147.

Elucidation of the amino acid sequence allows to obtain precise information on the topography of functionally important residues in the peptide chain of aspartate aminotransferase. The modification of some of the functional residues and isolation of corresponding peptide fragments have been reported previously by several authors [16–18]. Today it is possible to locate certain residues situated in the enzyme's active site or in its proximity – in particular the lysine residue linked by an aldimine bond with pyridoxal phosphate (Lys-258) and the functionally important tyrosine residue (Tyr-40). Knowledge of the primary structure will facilitate location of other amino acid residues of the enzyme, known to be essential or functionally important, and will provide a basis for sequence determination and comparison in homologous or related protein, as well as for elucidation, in the future, of the three-dimensional structure.

## Acknowledgements

The authors wish to thank A.P. Sukhikh, Yu.V. Smirnov and V.L. Chashchin for the participation in certain phases of this work, and L.D. Makarova and V.S. Khokhryakov for the preparation of aspartate aminotransferase.

## References

[1] Yu.A. Ovchinnikov, A.A. Kiryushkin, Ts.A. Egorov, N.G. Abdulaev, A.P. Kiselev, N.N. Modyanov, E.V. Grishin, E.I. Vinogradova, M.Yu. Feigina, N.A. Aldanova, V.M. Lipkin, A.E. Braunstein and O.L. Polyanovsky, FEBS Letters 12 (1971) 194.

[2] Yu.A. Ovchinnikov, A.A. Kiryushkin, Ts.A. Egorov, N.G. Abdulaev, A.P. Kiselev, N.N. Modyanov, E.V. Grishin, A.P. Sukhikh, E.I. Vinogradova, M.Yu. Feigina, N.A. Aldanova, V.M. Lipkin, A.E. Braunstein, O.L. Polyanovsky and V.V. Nosikov, FEBS Letters 17 (1971) 133.

[3] A.E. Braunstein and M.G. Kritsman, Biokhimiya 3 (1938) 590.

[4] V.I. Ivanov and M.Ya. Karpeisky, Advan. Enzymol. 32 (1969) 22.
[5] H. Wada and Y. Morino, Vitamins Hormones 22 (1964) 411.
[6] M. Martinez-Carrion and D. Tiemer, Biochemistry 6 (1967) 1715.
[7] O.L. Polyanovsky and M. Telegdi, Biokhimiya 30 (1965) 174.
[8] M. Martinez-Carrion, C. Turano, E. Chiancone, F. Rossa, A. Giartosio, F. Riva and P. Fasella, J. Biol. Chem. 242 (1967) 2397.
[9] Yu.A. Ovchinnikov, A.A. Kiryushkin, Ts.A. Egorov, N.G. Abdulaev, A.P. Kiselev and N.N. Modyanov, Biokhimiya 37 (1972) 451.
[10] Yu.A. Ovchinnikov, A.A. Kiryushkin, Ts.A. Egorov, N.N. Modyanov, N.G. Abdulaev, A.P. Kiselev and E.V. Grishin, Biokhimiya 37 (1972) 461.
[11] J. Sjöquist, Biochim. Biophys. Acta 41 (1960) 20.
[12] W.R. Gray, in: Methods in Enzymology, Vol. 11, eds. S.P. Colowick and N.O. Kaplan (Academic Press, New York, 1968) p. 469.
[13] M.M. Shemyakin, Yu.A. Ovchinnikov, E.I. Vinogradova, A.A. Kiryushkin, M.Yu. Feigina, N.A. Aldanova, Yu.B. Alakhov, V.M. Lipkin, A.I. Miroshnikov, B.V. Rosinov and S.A. Kazaryan, FEBS Letters 7 (1970) 8.
[14] K. Moon, D. Pizkiewicz and E.L. Smith, Proc. Natl. Acad. Sci. U.S. 69 (1972) 1380.
[15] G.M. Edelman, Biochemistry 9 (1970) 3197.
[16] O.L. Polyanovsky and B.A. Keil, Biokhimiya 28 (1963) 372.
[17] Y. Morino and T. Watanabe, Biochemistry 8 (1969) 3412.
[18] O.L. Polyanovsky, T.V. Demidkina and C.A. Egorov, FEBS Letters 23 (1972) 262.