THE PRIMARY STRUCTURE OF A LEU, ILE AND VAL (LIV)-BINDING PROTEIN FROM *ESCHERICHIA COLI*

Yu. A. OVCHINNIKOV, N. A. ALDANOVA, V. A. GRINKEVICH, N. M. ARZAMAZOVA and I. N. MOROZ

Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Moscow 117312, USSR

Received 11 April 1977

1. Introduction

Lately, proteins selectively binding low molecular weight substrates such as sugars, amino acids, inorganic ions etc. have been isolated from the periplasmic space of various bacterial cells. In conformity with a number of data these proteins seem to be involved in the active transport of these compounds across cellular membranes, at least in the initial (receptor) stage of the process [1,2]. A representative of this class is LIV-protein from *Escherichia coli*, specifically binding the amino acids Leu, Ile and Val [3,4]. It is as yet unclear just how this protein effects the selective binding or how it functions in the cell. A prerequisite for solution of this problem is elucidation of the protein's primary structure so that one could subsequently define its functionally important groups and regions.

This paper reports the complete amino acid sequence of LIV-protein, the first sequencing of a periplasmic binding-protein.

2. Materials and methods

LIV-protein was isolated from a K-12 *E.coli* strain according to [5]. The tryptic peptides of the carboxymethylated (CM) protein and the chymotryptic peptides of the cyanogen bromide fragment M_3 were separated on an AG-50 X 4 cation exchanger in pyridine-acetate buffer with concentration- and pH-gradients. Isolation of the individual peptides from the fractions thus obtained was achieved by means of paper chromatography in a butanol/pyridine/acetic acid/water (15:10:3:2) system and by high-voltage paper electrophoresis at pH 6.5 and pH 1.9. Acidic hydrolysis of the Asp–Pro bonds of the CM-LIV-protein was carried out in 7 M guanidinium-chloride containing pyridine–acetate buffer, pH 2.6, 40°C, 96 h [6]. The fragments obtained by cyanogen bromide cleavage, restricted tryptic hydrolysis of the maleylated CM-LIV-protein (M-CM-LIV) or acidic hydrolysis were subjected to gel-filtration on G-50 Sephadex or P-10 Biogel followed by chromatography on DEAE-cellulose in the presence of 6 M urea. The thermolytic hydrolysate of the cyanogen bromide M_5-fragment was filtered through G-25 Sephadex and the fractions obtained were purified by high-voltage paper electrophoresis. The N-terminal amino acid sequences of the protein and of the large peptides were determined by automatic degradation using a 890 C model Beckman sequencer. The PTH-derivatives were identified by thin-layer chromatography on silica gel, gas–liquid chromatography (GC-65 Beckman chromatograph), mass spectrometry and amino acid analysis of their hydrolysates.

For determination of the amino acid composition of the protein and the peptides use was made of Biocal BC-201 (FRG) and Durrum D-500 (USA) amino acid analyzers.

3. Results and discussion

LIV-Protein consists of a single peptide chain mol. wt 36 000) [4,7] and contains, as will be shown below, one disulfide bridge.

The structural study was begun with an exhaustive tryptic digestion of CM-protein, which yielded all the peptides comprising the protein molecule. The treatment was highly specific, only one of the peptides being due

to trypsin-uncharacteristic scission of the 281–282 peptide bond. The next step–overlap determination of the tryptic peptides–was achieved by splitting the protein chain into large blocks by cyanogen bromide cleavage (5 Met residues), tryptic hydrolysis of the M-CM-protein (7 Arg residues) and limited acid splitting of the three Asp–Pro bonds in the molecule. The resultant blocks were then subjected to automatic degradation on a sequencer.

The reassembly of the LIV-protein structure from the fragments is schematically shown in fig.1. The cyanogen bromide hydrolysate yielded six peptides (M_1–M_6). The N-terminal fragment contained a Met–Ser (11–12) bond very resistant to cyanogen bromide cleavage as a result of which the M_1-peptide was isolated in negligible yield, and attempts to separate fragments M_2 and M_{1+2}, similar in charge and in molecular weight proved unsuccessful. Analysis of the N- and C-terminal structures of fragments M_1–M_6 together with the structural data on the methionine-containing tryptic peptides provided the key to their sequence in the polypeptide chain and thereby permitted establishment of the entire architecture of the molecule.

The N-terminal sequencing of the protein (52 amino acid residues) made it possible to pinpoint 6 tryptic peptide overlappings. Automatic degradation of fragments M_3 and M_5 led to identification of 30 amino acids from each fragment and, thereby, to reassembly of 5 tryptic peptides. The structure of the M_4-fragment (17 amino acid residues) was determined by manual Edman degradation; to determine the structure of fragment M_6 resort was made to hydrolysis by *Staphylococcus aureus* protease.

All the seven arginine residues of the protein molecule are located in its central region (fragments M_3 and M_4) so that tryptic hydrolysis of the M-CM-protein should have given rise to two large and six

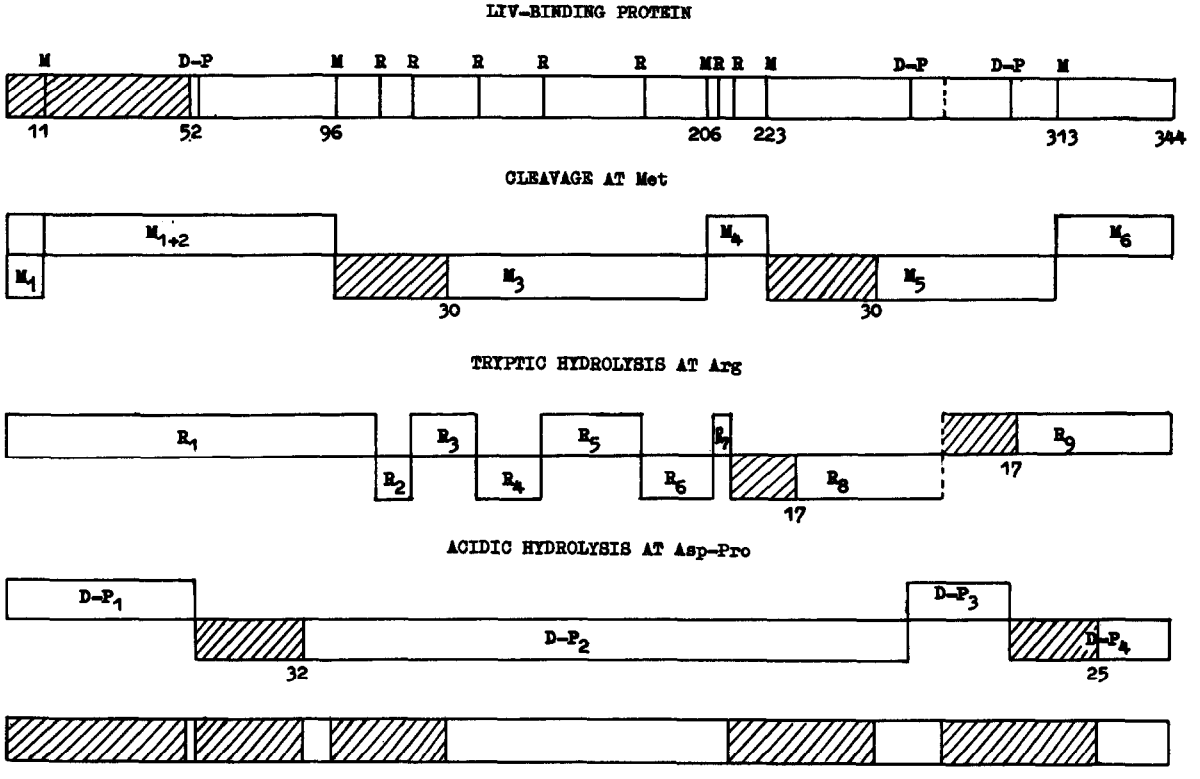

Fig.1. Schematic representation of reassembly of LIV-protein. (M) Cyanogen bromide peptides. (R) Peptides of limited tryptic hydrolysis. (D-P) Peptides of acidic hydrolysis at Asp–Pro. Regions determined by automatic degradation are shaded.

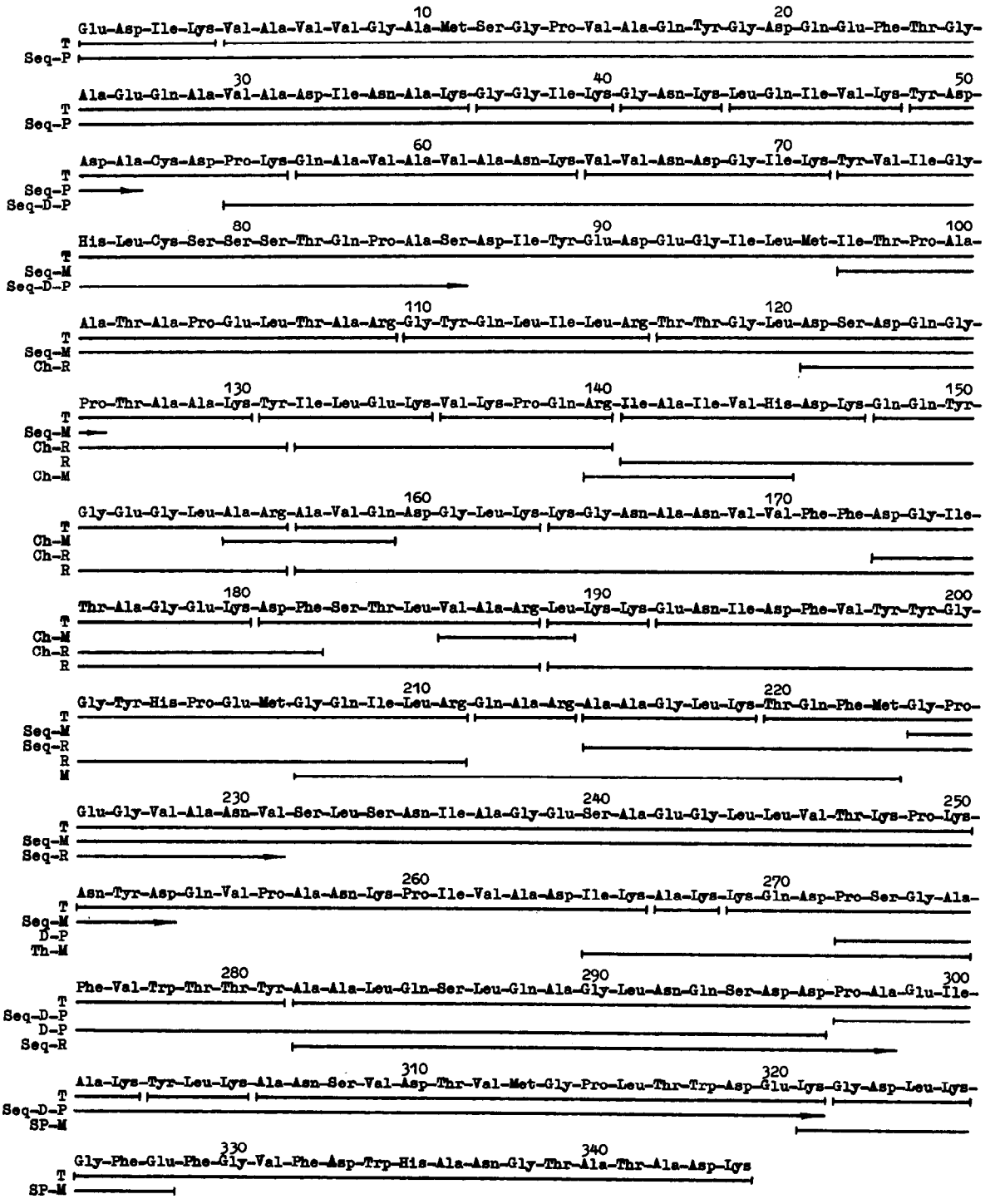

Fig.2. The primary structure of LIV-binding protein. (T) Tryptic peptides. (R) Peptides of limited tryptic hydrolysis. (M) Cyanogen bromide peptides. (D-P) Peptides of acidic hydrolysis at Asp–Pro. Amino acid sequence determined by sequencer: (Seq-P) on the protein itself; (Seq-M, Seq-R, Seq-D-P) on the corresponding fragments. (Ch-M, Th-M, SP-M) Peptides from the chymotryptic, thermolytic and *Staphylococcus* protease digestion of the cyanogen bromide fragments. (Ch-R) Chymotryptic peptides of fragments R.

relatively small peptides. However, the high molecular weight fraction turned out to be a mixture of three components due to non-specific cleavage of the Tyr-Ala (281–282) bond at the C-terminus. Fragments R_2–R_7 are quite small (3–32 amino acid residues) and their structures were determined by manual methods. Fragments R_3 and R_5 were subjected to chymotryptic hydrolysis. In order to determine the overlappings of the arginine containing blocks R_2–R_7 fragment M_3 was subjected to chymotryptic hydrolysis with the purpose of isolating arginine-containing peptides. The amino acid sequence of these peptides combined with the results from the cyanogen bromide degradation of peptide M_3 on the sequencer, yielded the unequivocal sequence of all the R_2–R_7 blocks and consequently, the complete structure of peptide M_3. Automatic degradation of fragment R_9 elucidated the amino acid sequence in the unknown region of the cyanogen bromide peptide M_5.

The structural data on the cyanogen bromide and the limited tryptic peptides made possible the location of the Asp–Pro bonds, the points of attack on the peptide chain in selective acid cleavage under mild conditions. There are three such bonds in the protein and their positions were such that the acid cleavage fragments contained valuable structural information on the initial molecule. Under the chosen hydrolytic conditions partial splitting of the Asp–Gly bonds (3 bonds in the protein), as well as of the Asp–Pro occurred, which complicated the hydrolysate. The automatic and manual degradation of the D-P_4 and D-P_3 fragments revealed the positions of three overlaps thereby considerably narrowing the unknown region in the cyanogen bromide peptide M_5. The missing information on the structure of this peptide was obtained from its thermolytic digestion. The 32 amino acid sequence of the D-P_2 fragment and the overlappings of three tryptic peptides were simultaneously determined on the sequencer, thereby completely establishing the structure of the cyanogen bromide peptide M_{1+2}.

All the Cys-sulfur in the molecule must be in the form of an S–S-bridge. This can be seen from our proof of the presence of such a bridge by means of the procedure described in [8] and also by our isolation from the hydrolysate of the Asp–Pro–split native protein of a dual N-terminal fragment with GLU and Pro at the termini. This fragment treated with mercaptoethanol followed by carboxymethylation yielded the peptides D-P_1 and D-P_2.

The complete amino acid structure of LIV-binding protein is given in fig.2. The protein molecule consists of 344 amino acid residues (mol. wt 36 770) and has the following amino acid composition: Asp-26, Asn-14, Thr-18, Ser-14, Glu-17, Gln-20, Pro-15, Gly-34, Ala-43, Met-5, Val-27, Cys-2, Ile-20, Leu-23, Tyr-13, Phe-10, His-4, Lys-29, Arg-7, Trp-3.

The disulfide bridge involves 26 residues (53–78). An interesting feature of the structure is that the arginine residues are all situated in the central part of the molecule while the tryptophan residues are located in the C-terminus.

References

[1] Oxender, D. L. and Quay, S. (1975) Ann. NY Acad. Sci. 264, 358–372.
[2] Boos, W. (1974) Ann. Rev. Biochem. 43, 123–146.
[3] Piperno, J. R. and Oxender, D. L. (1966) J. Biol. Chem. 242, 5732–5734.
[4] Anraku, Y. (1968) J. Biol. Chem. 243, 3116–3122.
[5] Antonov, V. K., Arsen'eva, E. L., Gavrilova, N. A., Ginodman, L. M. and Krylova, Yu. I. (1973) Biokhimiya 38, 1294–1297.
[6] Fraser, K. J., Poulsen, K. and Haber, R. (1972) Biochemistry 11, 4974–4977.
[7] Penrose, W. R., Nichoalds, G. E., Piperno, J. R., and Oxender, D. L. (1968) J. Biol. Chem. 243, 5921–5928.
[8] Kolesnikova, V. Yu., Skljankina, V. A., Baratova, L. A., Nazarova, T. I. and Avaeva, S. M. (1974) Biokhimiya 39, 287–293.