THE PRIMARY STRUCTURE OF ELONGATION FACTOR G FROM *ESCHERICHIA COLI*

A complete amino acid sequence

Yu. A. OVCHINNIKOV, Yu. B. ALAKHOV, Yu. P. BUNDULIS, M. A. BUNDULE, N. V. DOVGAS, V. P. KOZLOV, L. P. MOTUZ and L. M. VINOKUROV

Institute of Protein Research, USSR Academy of Sciences, 142292 Poustchino, Moscow Region, USSR

Received 26 January 1982

1. Introduction

Elongation factor G (EF-G) is a large protein consisting of one polypeptide chain with $M_r \sim 80\,000$. Studies of this protein by traditional methods used to determine the primary structure did not seem the most rational. We chose the method of limited proteolysis as the first stage of studies permitting, in many cases, cleavage of the protein molecule into a small number of large fragments relatively stable against further effects of protease. On mild tryptic hydrolysis the G-factor is split into 5 fragments [1, cf. 2,3], 4 of which (T_4–T_7) form the complete protein polypeptide chain (for fragment nomenclature and arrangement in the protein chain see fig.1). Thus, the studies of the primary structure of the elongation factor G led to determination of the structure of these 4 fragments and a search for the overlapping peptides for reconstitution of the entire polypeptide chain.

Studies of the structure of fragments T_5–T_7 were relatively simple due to their comparatively low molecular mass. The molecular mass of the N-terminal fragment T_6 is 6500, that of the following fragment T_7, 7500, and that of the C-terminal fragment T_5, 25 000. Traditional chemical and enzymatic methods of polypeptide chain cleavage permitted us earlier to determine the complete primary structures of fragments T_6 [4], T_7 [5,6] and T_5 [7] involving a total of >350 amino acid residues.

To elucidate the complete structure of elongation factor G it was necessary to determine the amino acid sequence of the largest tryptic fragment T_4 representing the middle part of the protein molecule and to obtain the peptides joining all the fragments of limited trypsinolysis into a 1 polypeptide chain. To this end, the products of protein cleavage by cyanogen bromide and those of limited acidic hydrolysis at the Asp–Pro bond were studied.

This paper is devoted to the results of these studies which permitted us to join all the fragments of limited trypsinolysis into one polypeptide chain and to present the complete amino acid sequence of the elongation factor G.

2. Materials and methods

Elongation factor G and the mixture of fragments T_3 and T_4 were isolated using the technique in [2]. The mixture of fragments T_3 and T_4 was used here without separation as fragment T_4 differs from fragment T_3 by the presence at the N-terminus of the latter of the T_7 polypeptide chain. Homogeneity of the protein as well as the absence of any contaminations of other fragments in the mixture of fragments T_3 and T_4 was checked by SDS–polyacrylamide gel electrophoresis and by determination of the N-terminal amino acid sequences by automated Edman degradation.

To establish the structure of fragment T_4 and to obtain overlappings between fragments of limited proteolysis, the products of cyanogen bromide cleavage of a whole molecule of the G-factor were studied. For overlapping of cyanogen bromide peptides, the mixture of fragments T_3 and T_4 was digested by trypsin (TPCK-treated, Worthington) after modification of the fragments with maleic anhydride, by *Staphylococcus aureus* protease (Miles-Lab.) and by cleavage of the G-factor at the Asp–Pro bond [8].

After initial separation on Sephadex G-25 the pep-

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$02.75

tides of cyanogen bromide cleavage were separated on QAE-Sephadex A-25. Large cyanogen bromide and tryptic peptides were subjected to additional digestions with trypsin, chymotrypsin and *Staph. aureus* protease to determine their structure. The digests were separated by peptide mapping on cellulose thin-layer plates (Merck) or by high-performance liquid chromatography on sulphopolystyrene cation-exchange resin (Chromobeads type P, Technicon) with fluorescent detection of peptides in the eluate by *o*-phthalic dialdehyde (Sigma) [9]. To isolate peptides of cyanogen bromide cleavage containing cysteine residues, we used covalent chromatography on thiol-activated Sepharose 4B (Pharmacia) [10].

C-Terminal amino acid residues in fragment T_4 and in the isolated peptides were determined by carboxypeptidases A, B (Worthington) [11] and Y [12] (isolated as in [13]).

The amino acid composition was assessed on the D-500 analyzer (Durrrum). The amount of valine and isoleucine was estimated after 72 h hydrolysis.

Automated determination of the amino acid sequences was done on a sequencer (model 890C, Beckman) using the dimethylallylamine or quadrol programs and on a solid-phase sequencer APS-240 (Rank-Hilger). Automated Edman degradation used for analysis of the peptides obtained at cleavage of the G-factor at the Asp–Pro bond included treatment with fluram [14] to exclude from the Edman reaction peptides with no proline residues at the N-terminus.

The PTH-derivatives of amino acids were identified by thin-layer chromatography on silica gel and HPLC (type HP 1080 A chromatograph) on a 5 μm Ultrasphere RP-18 column in a sodium phosphate buffer using a methanol gradient.

Amino acid sequences of short peptides were determined by the dansyl Edman method [15]. Dansyl derivatives of amino acids were identified by thin-layer chromatography on silica gel. Asparagine and glutamine were determined according to [6].

3. Results and discussion

Studies of the kinetics of tryptic hydrolysis of the G-factor and determination of the N-terminal sequences in the fragments [2] have shown that fragment T_6 is the N-terminal one. Following it along the chain is fragment T_7 containing a functionally important cysteine residue. The GTP-binding site is in the region of

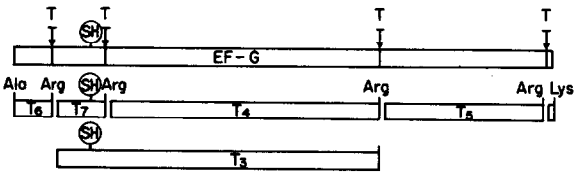


Fig.1. Arrangement of fragments of limited tryptic hydrolysis along the polypeptide chain of the G-factor.

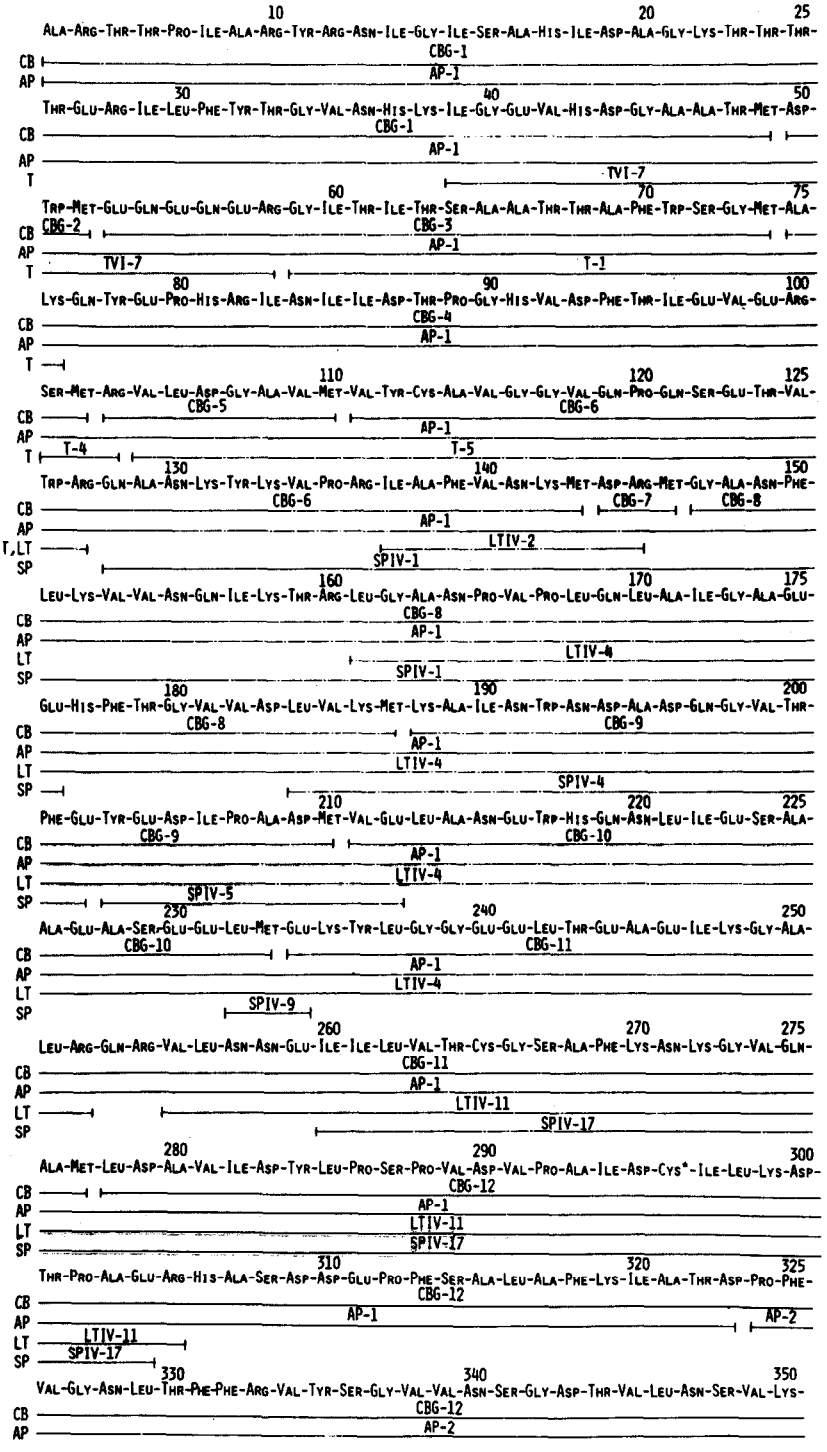
this fragment [17]. Fragment T_7 is followed by fragment T_4 (M_r 41 000) and then by fragment T_5 which is the C-terminal part of the protein molecule (fig.1).

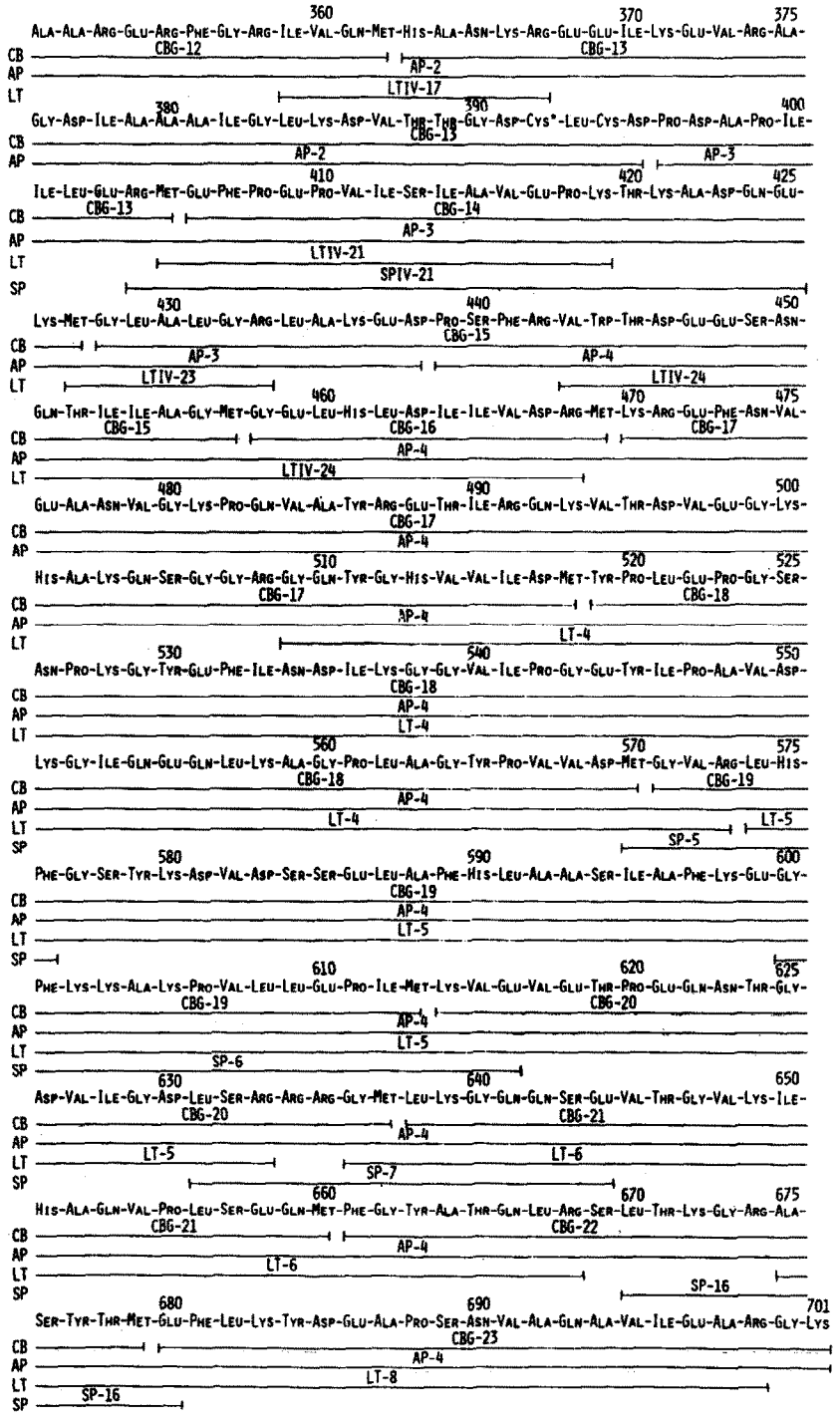
The molecule contains 22 methionine residues and it could be therefore expected that on cyanogen bromide cleavage ⩾23 peptides would be formed. The scheme of cyanogen bromide hydrolyzate separation proposed by us has allowed us to isolate all the 23 peptides. The mixture of peptides was modified by maleic anhydride and then separated on a Sephadex G-25 column. The fractions obtained were subjected to rechromatography on QAE-Sephadex A-25, and the fraction containing low molecular mass peptides was separated on thin-layer cellulose by peptide mapping. After chromatography on QAE-Sephadex A-25 the fractions were subjected once again to rechromatography on QAE-Sephadex columns after a preliminary removal of the protecting maleic group.

Of the 23 isolated peptides 10 peptides are included in fragment T_4 while 2 peptides also involve some sites of polypeptide chains of fragments T_7 and T_5, and one peptide overlaps fragments T_6 and T_7 (scheme 1). Thus, the study of the structure of fragment T_4 becomes the determination of the amino acid sequences of the 12 peptides of cyanogen bromide cleavage and the search for the overlaps between them. The structure studies of these peptides have required their additional cleavage by chemical or enzymic methods.

To obtain overlapping peptides we have used tryptic digest of fragment T_4 after its modification with maleic anhydride, digest by *Staphylococcus aureus* protease and cleavage of the G-factor at the Asp–Pro bond. There are 3 such bonds in the molecule. Studies of the N-terminal sequences of the formed peptides by the automated Edman method with the use of fluram as a reagent blocking the primary α-amino groups have allowed us to obtain overlaps between the peptides from CBG-13–CBG-17 (scheme 1) and thus to reliably incorporate fragments T_4 and T_5 into one polypeptide chain.

Scheme 1
Complete primary structure of elongation factor G





Solid lines indicate only cyanogen bromide peptides of EF-G and the overlapping peptides obtained from other hydrolyses of T_4–T_7 fragments of EF-G. Fragments of limited trypsinolysis occupy the following positions in the polypeptide chain: T_6, 1–58; T_7, 59–127; T_4, 128–471; T_5, 472–699. CB, peptides of cyanogen bromide cleavage of EF-G. AP, peptides of EF-G hydrolysis at the Asp–Pro bond; T, peptides of tryptic hydrolysis of fragments T_6 and T_7; LT, peptides of tryptic hydrolysis of fragments T_4 and T_5 modified at lysine residues; SP, peptides of hydrolysis of fragments T_4 and T_5 by *S. aureus* protease; Cys* = cysteine residues which form the S–S bond.

From the data of the amino acid analyses and determination of the molecular mass of the fragments of limited trypsinolysis it can be concluded that trypsinolysis does not lead to a nick of any essential regions of the polypeptide chain between these fragments [2]. Separation of the products of cyanogen bromide elongation factor cleavage resulted in the isolation of several peptides, and the determination of their structures allowed us to unite all the fragments of limited trypsinolysis (peptides CBG-3, CBG-6 and CBG-17 in the scheme) into one polypeptide chain. There are no insertions between the fragments of limited trypsinolysis, i.e., trypsin makes point disruptions in the elongation factor G polypeptide chain [4]. We demonstrated [2] that all the fragments of limited trypsinolysis have arginine as the C-terminal amino acid residue, whereas the C-terminal amino acid residue of EF-G is lysine. A structural study of the peptide CBG-23 showed that is elongated by two amino acid residues Gly–Lys as compared with the corresponding peptide of cyanogen bromide cleavage of fragment T_5 [7]. Thus, we have concluded that the cleavage of the C-terminal dipeptide is the result of limited trypsinolysis of the G-factor.

Thus, the study of fragments of limited trypsinolysis of the G-factor and the products of cyanogen bromide cleavage and hydrolysis at the Asp–Pro bond of the G-factor have allowed us to report the complete amino acid sequence of the protein. Its polypeptide chain consists of 701 amino acid residues, has M_r 77 321.46 and contains 5 cysteine residues, two of which (positions 296 and 392 in the scheme) form the S–S bond. To find the position of the S–S bond, the G-factor was treated with Elman's reagent (to protect the SH-groups from oxidation) then subjected to cyanogen bromide cleavage. After removal of the protecting group the products of cyanogen bromide cleavage were chromatographed on a thiol-activated Sepharose 4B column. After removal of thiol-containing peptides from Sepharose they were carboxymethylated, then reduced by sodium sulphite and again carboxymethylated by ^{14}C-labelled ICH_2COOH. The cysteine residues containing the radioactive label were found in peptides CBG-12 and CBG-13 (scheme 1). This technique provided the isolation of all thiol-containing peptides of the elongation factor G.

These data on the primary structure of the G-factor agree with the data of other laboratories where the structure of some sites of the gene encoding the G-factor has been determined. In particular, studies of *E. coli str*-operon permitted the nucleotide sequence corresponding to 92 amino acid residues from the N-terminus of the G-factor to be established [18], while determination of the nucleotide sequence of the *tufA* gene encoding the elongation factor Tu allowed the structure of the *fus* gene region which encodes the C-terminal part of the G-factor consisting of 20 amino acid residues [19] to be elucidated. In both cases the amino acid sequences established by a direct method and those derived from structure studies of DNA completely coincide.

Elucidation of the primary structure of the G-factor will permit the regularities of the relationship between the structure and biological function to be studied on an entirely new basis. A new possibility appears for the study of the tertiary structure of the G-factor by X-ray structural analysis as large enough crystals have been obtained [20,21] completely including the protein polypeptide chain.

References

[1] Skar, D. C., Rohrbach, M. S. and Bodley, J. W. (1975) Biochemistry 14, 3922–3926.
[2] Alakhov, Yu. B., Motuz, L. P., Stengrevics, O. A., Vinokurov, L. M. and Ovchinnikov, Yu. A. (1977) Bioorg. Khim. 3, 1336–1345.
[3] Alakhov, Yu. B., Stengrevics, O. A., Bundulis, Yu. P., Motuz, L. P. and Vinokurov, L. M. (1979) Bioorg. Khim. 5, 330–339.
[4] Motuz, L. P., Bundulis, Yu. P. and Alakhov, Yu. B. (1979) Bioorg. Khim. 5, 814–827.
[5] Alakhov, Yu. B., Motuz, L. P., Stengrevics, O. A. and Ovchinnikov, Yu. A. (1978) FEBS Lett. 85, 287–290.
[6] Alakhov, Yu. B., Motuz, L. P., Stengrevics, O. A. and Vinokurov, L. M. (1978) Bioorg. Khim. 4, 1301–1313.
[7] Alakhov, Yu. B., Dovgas, N. V., Motuz, L. P., Vinokurov, L. M. and Ovchinnikov, Yu. A. (1981) FEBS Lett. 126, 183–186.
[8] Landon, M. (1977) Methods Enzymol. 47, 145–149.
[9] Creaser, E. H. and Hughes, G. J. (1977) J. Chromatogr. 144, 69–75.
[10] Egorov, C. A., Svenson, A., Ryden, L. and Carlson, J. (1975) Proc. Natl. Acad. Sci. USA 72, 3029–3033.
[11] Amber, R. P. (1967) Methods Enzymol. 11, 155–156.
[12] Hayshi, R. (1977) Methods Enzymol. 47, 84–93.
[13] Jochansen, J. T., Breddam, K. and Ottesen, M. (1976) Carlsberg Res. Commun. 41, 1–14.
[14] Bhown, A. S., Bennet, J. C., Morgan, P. H. and Mole, J. E. (1981) Anal. Biochem. 112, 158–162.
[15] Gray, W. R. (1976) Methods Enzymol. 11, 469–475.
[16] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.

[17] Girshovich, A. S., Bochkareva, E. S., Pozdnyakov, V. A. and Ovchinnikov, Yu. A. (1978) FEBS Lett. 85, 283–286.
[18] Post, L. E. and Nomura, M. (1980) J. Biol. Chem. 225, 4660–4666.
[19] Yokota, T., Sugisaki, H., Takanami, M. and Kaziro, Y. (1980) Gene 12, 25–31.
[20] Reshetnikova, L. S., Savchenko, I. V. and Garber, M. B. (1978) Dokl. Akad. Nauk SSSR 243, 523–525.
[21] Garber, M. B. and Reshetnikova, L. S. (1981) Bioorgan. Khim. 9, 1419–1422.