PRIMARY STRUCTURE OF α-SUBUNIT OF DNA-DEPENDENT RNA POLYMERASE FROM *ESCHERICHIA COLI*

Yu. A. OVCHINNIKOV, V. M. LIPKIN, N. N. MODYANOV, O. Yu. CHERTOV, and Yu. V. SMIRNOV

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

Received 14 February 1977

1. Introduction

Transcription of genetic information in bacterial cells is mediated by DNA-dependent RNA polymerase (ribonucleoside-triphosphate:RNA nucleotidyltransferase, EC 2.7.7.6) [1]. The enzyme from *E. coli* (mol. wt 500 000) has been shown to possess a complex structure [2] consisting of two large subunits, β and β′ (mol. wt 155 000 and 165 000, respectively), two α-subunits (mol. wt 40 000) and initiation factor, σ (mol. wt 90 000). Its structural complexity is paralleled by the multistep nature of the transcription process.

Only limited information on the role of the individual subunits in the functioning of RNA polymerase is available, owing in part to our lack of knowledge of its primary and spatial structure. We have, therefore, undertaken an investigation into the primary structure of DNA-dependent RNA polymerase from *E. coli* B.

In the study reported here we have determined the complete amino acid sequence of the RNA polymerase α-subunit (see [3]) using a variety of degradation methods and ^{14}C-enriched amino acid residues for facilitating detection and isolation of the fragments. The polypeptide chain of the α-subunit has mol. wt 36 512 and consists of 329 amino acid residues. In a comparative study of the peptide compositions of various bacterial polymerase α-subunits it has been shown by peptide mapping that they are evolutionarily conservative proteins [4].

2. Materials and methods

Isolation of the *E. coli* B DNA-dependent RNA polymerase α-subunit is described in [4]. ^{14}C-Labeling was achieved by growing *E. coli* in a culture medium enriched with the respective ^{14}C-labeled amino acids.

In the present study use was made of the following reagents: trypsin (TPCK-treated), α-chymotrypsin (Worthington), thermolysin (Calbiochem), protease from *Staphylococcus aureus* (Miles Laboratories), BNPS-skatole (Pierce), ion-exchange resins AG 50 × 4, AG 50 × 2 (Bio-Rad), PA-35 (Beckman), Sephadexes (Pharmacia).

The α-subunit samples were subjected to the following treatments:

(1) Tryptic digestion of the carboxymethylated and all-amino acid-^{14}C-labeled α-subunit (5 μmol, enzyme/substrate ratio 1:30), pH 8.3, 37°C, 5.5 h. The hydrolysate was first separated on the AG 50 × 4 cation exchanger in pyridine/acetate buffer with an exponential pH (3.1–5.0) and a pyridine concentration-gradient (0.2–2.0 M). Peptides in the eluate were detected by means of ninhydrin and radioactivity measurements. Their further separation and purification were carried out by gel-filtration, paper chromatography and paper electrophoresis.

(2) *Staphylococcus* protease digestion of the carboxymethylated α-subunit (5 μmol, enzyme/substrate ratio 1:30), 0.1 M ammonium bicarbonate, pH 8.0, 37°C, 16 h. Isolation was similar to that for the tryptic peptides but with the use of AG 50 × 2 resin.

North-Holland Publishing Company – Amsterdam

(3) Thermolysin hydrolysis of [^{14}C]lysine- and [^{14}C]arginine-labeled α-subunit (0.5 μmol, enzyme/substrate ratio 1:50), 0.1 M ammonium bicarbonate, pH 8.0; 37°C, 4 h. The peptides were fractionated by chromatography using the PA-35 cation-exchanger and Whatman 3 MM paper with peptide detection by autoradiography.

(4) Cyanogen bromide (500-fold excess) cleavage of the α-subunit (3 μmol), 70% formic acid, 20 h. The resultant mixture was chromatographed on Sephadex G-75 in 6 M guanidine–HCl/0.05 M Tris–HCl (pH 8.0)/0.1% 2-mercaptoethanol.

The N-terminal amino acid sequence of the peptides was determined by Edman degradation, the amino acids being identified as dansyl-derivatives and phenylthiohydantoins. The C-terminal sequence was determined by means of carboxypeptidases A, B and C. For larger peptides recourse was had additionally to chymotryptic, thermolytic and *Staphylococcus* protease hydrolysis.

3. Results and discussion

In the sequential study of the RNA polymerase α-subunit structure, use was made of tryptic, *Staphylococcus* protease and thermolytic hydrolysis and of cyanogen bromide, hydroxylamine and BNPS-skatole cleavage.

Success in determining the complete protein sequence was due largely to tryptic hydrolysis that permitted isolation and further study of the entire set of resultant fragments. As indicated, labeling of the protein considerably facilitated detection of the peptides in the course of their isolation.

In the early stages of the sequence study it was found that the tryptic peptides of the α-subunit contained over 30 glutamic acid residues of which only six were adjacent to the residues of basic amino acids. Consequently it was deemed feasible in a second type of protein hydrolysis, to make use of *Staphylococcus aureus* protease, that attacks mainly bonds formed by the α-carboxyl groups of the glutamic acid residues [5].

It is interesting to note that in the process of hydrolysis this enzyme also split the α-subunit at four aspartic acid residues (174, 197, 233 and 280) three of which were connected by an Asp–Leu bond, and also unexpectedly cleaved four serine (21, 49, 266 and 313) bonds.

Staphylococcus protease hydrolysis of the α-subunit led to the isolation of 41 peptides that, after their structural determination, permitted the reassembling of 28 tryptic peptides into larger units. Two highly basic peptides (33–49 and 137–163) could not be isolated from the hydrolysate apparently owing to their irreversible sorption on the resin during the initial stages of separation of the hydrolysate. The structure of these portions of the α-subunit chain was established by the amino acid sequencing of the respective cyanogen bromide fragments.

In order to find additional overlapping fragments, the α-subunit (0.5 μmol) was subjected to hydrolysis by thermolysin. As the most interesting peptides from this hydrolysis were lysine- and arginine-containing ones, a protein sample was obtained from bacteria grown on a medium with ^{14}C-enriched lysine and arginine. The peptide fractions monitored by autoradiographic assay in the course of separation of the thermolytic hydrolysate yielded the peptides quite selectively and with minimum loss.

Valuable information on the structure of the N-terminal and middle parts of the α-subunit molecule was obtained after cyanogen bromide cleavage of the parent protein. Separation by means of Sephadex G-75, at pH 8.0, in 6 M guanidine–HCl and 0.1% 2-mercaptoethanol yielded all the fragments expected and in a homogeneous state. In the absence of mercaptoethanol considerable aggregation of the peptides was observed, evidently due to formation of disulphide bonds. In order to elucidate the structure of the cyanogen bromide fragments containing residues 1–51 and 143–205, they were subjected to additional hydrolysis by chymotrypsin and *Staphylococcus* protease followed by isolation and sequence determination of the resultant peptides. Substantiation of overlap between the tryptic peptides (244–265) and (266–271) was obtained by the directed isolation of the cysteinyl-containing chymotryptic peptide (263–270). To this end the cyanogen bromide fragment (206–316) was carboxymethylated by iodo-[^{14}C]acetic acid and digested by chymotrypsin. The resultant labeled peptide was isolated by means of peptide mapping.

In the sequencing of tryptic peptide (201–218) the sequence asparaginyl–glycyl was discovered; it proved to be the only such one in the protein molecule. For

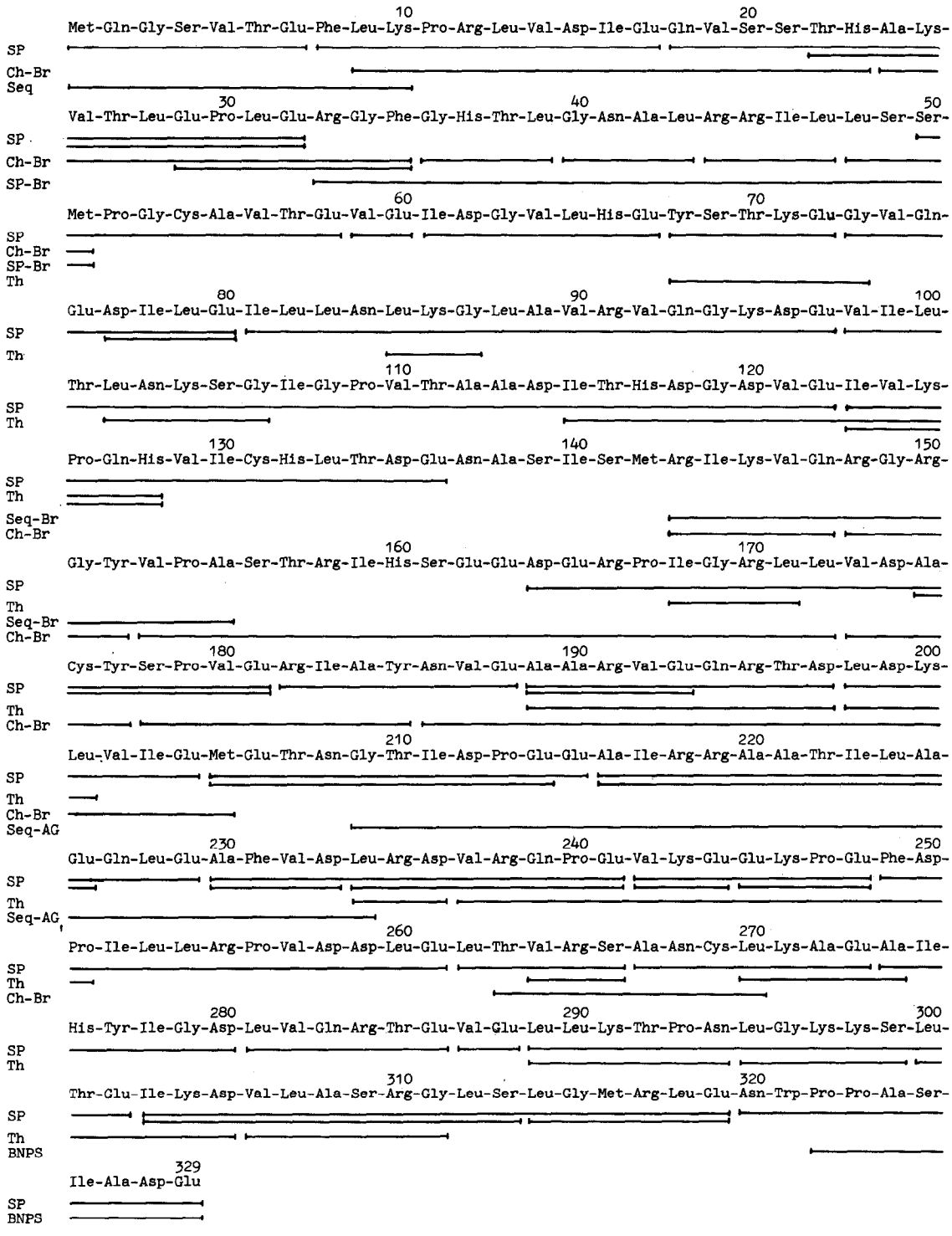
10 20
Met-Gln-Gly-Ser-Val-Thr-Glu-Phe-Leu-Lys-Pro-Arg-Leu-Val-Asp-Ile-Glu-Gln-Val-Ser-Ser-Thr-His-Ala-Lys-
SP
Ch-Br
Seq
30 40 50
Val-Thr-Leu-Glu-Pro-Leu-Glu-Arg-Gly-Phe-Gly-His-Thr-Leu-Gly-Asn-Ala-Leu-Arg-Arg-Ile-Leu-Leu-Ser-Ser-
SP
Ch-Br
SP-Br
60 70
Met-Pro-Gly-Cys-Ala-Val-Thr-Glu-Val-Glu-Ile-Asp-Gly-Val-Leu-His-Glu-Tyr-Ser-Thr-Lys-Glu-Gly-Val-Gln-
SP
Ch-Br
SP-Br
Th
80 90 100
Glu-Asp-Ile-Leu-Glu-Ile-Leu-Leu-Asn-Leu-Lys-Gly-Leu-Ala-Val-Arg-Val-Gln-Gly-Lys-Asp-Glu-Val-Ile-Leu-
SP
Th
110 120
Thr-Leu-Asn-Lys-Ser-Gly-Ile-Gly-Pro-Val-Thr-Ala-Ala-Asp-Ile-Thr-His-Asp-Gly-Asp-Val-Glu-Ile-Val-Lys-
SP
Th
130 140 150
Pro-Gln-His-Val-Ile-Cys-His-Leu-Thr-Asp-Glu-Asn-Ala-Ser-Ile-Ser-Met-Arg-Ile-Lys-Val-Gln-Arg-Gly-Arg-
SP
Th
Seq-Br
Ch-Br
160 170
Gly-Tyr-Val-Pro-Ala-Ser-Thr-Arg-Ile-His-Ser-Glu-Glu-Asp-Glu-Arg-Pro-Ile-Gly-Arg-Leu-Leu-Val-Asp-Ala-
SP
Th
Seq-Br
Ch-Br
180 190 200
Cys-Tyr-Ser-Pro-Val-Glu-Arg-Ile-Ala-Tyr-Asn-Val-Glu-Ala-Ala-Arg-Val-Glu-Gln-Arg-Thr-Asp-Leu-Asp-Lys-
SP
Th
Ch-Br
210 220
Leu-Val-Ile-Glu-Met-Glu-Thr-Asn-Gly-Thr-Ile-Asp-Pro-Glu-Glu-Ala-Ile-Arg-Arg-Ala-Ala-Thr-Ile-Leu-Ala-
SP
Th
Ch-Br
Seq-AG
230 240 250
Glu-Gln-Leu-Glu-Ala-Phe-Val-Asp-Leu-Arg-Asp-Val-Arg-Gln-Pro-Glu-Val-Lys-Glu-Glu-Lys-Pro-Glu-Phe-Asp-
SP
Th
Seq-AG
260 270
Pro-Ile-Leu-Leu-Arg-Pro-Val-Asp-Asp-Leu-Glu-Leu-Thr-Val-Arg-Ser-Ala-Asn-Cys-Leu-Lys-Ala-Glu-Ala-Ile-
SP
Th
Ch-Br
280 290 300
His-Tyr-Ile-Gly-Asp-Leu-Val-Gln-Arg-Thr-Glu-Val-Glu-Leu-Leu-Lys-Thr-Pro-Asn-Leu-Gly-Lys-Lys-Ser-Leu-
SP
Th
310 320
Thr-Glu-Ile-Lys-Asp-Val-Leu-Ala-Ser-Arg-Gly-Leu-Ser-Leu-Gly-Met-Arg-Leu-Glu-Asn-Trp-Pro-Pro-Ala-Ser-
SP
Th
BNPS
329
Ile-Ala-Asp-Glu
SP
BNPS

Fig.1

checking the amino acid sequence of this part of the polypeptide chain this (Asn–Gly) bond in the α-subunit was ruptured by hydroxylamine [6] and the C-terminal fragment (209–329) was isolated by chromatography on Sephadex G-100. Analyzed on the 890 C Beckman sequencer this peptide yielded the sequence of residues 209–234.

The sequencer was also used to determine the N-terminal sequence of the RNA polymerase α-subunit (10 residues) whereas the C-terminal sequence was obtained from the peptide resulting from BNPS-skatole cleavage at the only tryptophan (321) linkage of the protein [7].

The combined studies led to the complete sequence of the α-subunit of RNA polymerase shown in fig.1. Since all the fragments of the α-subunit polypeptide chain were isolated from the tryptic hydrolysate, these fragments are not denoted in the figure. Several of the tryptic peptides contained 2–3 residues of basic amino acid because of the resistance of the Lys–Pro (10–11, 246–247) and Arg–Pro (166–167, 255–256) bonds to trypsin attack and also owing to the incomplete hydrolysis of Lys–Asp (95–96), Lys–Val (145–146), Lys–Leu (200–201) and Lys–Lys (297–298) bonds. In the figure all peptides from the *Staphylococcus* protease treatment of the α-subunit are underlined, while only those thermolytic peptides are underlined that yielded valuable overlap information.

In conformity with the primary structure of the α-subunit it consists of 329 amino acid residues of the following composition: Asp 21, Asn 9, Thr 19, Ser 17, Glu 36, Gln 10, Pro 16, Gly 20, Ala 23, Cys 4, Val 30, Ile 24, Met 5, Leu 38, Tyr 5, Phe 4, His 8, Lys 16, Arg 23, Trp 1, the mol. wt 36 512. Characteristic features of the protein are the high content of acidic amino acid residues relative to basic amino acid residues and the low number of residues of aromatic amino acids.

The only mutation of the gene of the α-subunit known, hitherto involving the replacement of a leucine residue by histidine, can now be assigned with the accuracy defined in [8] to positions 289 or 290 of the proposed sequence. When *E. coli* is infected with T_4 phage ADP-ribosylation of an arginine residue of the RNA polymerase α-subunit takes place. Goff [9] (cf. [10]) has found that the modified arginine residue is located in the Thr–Val–Arg sequence, i.e., residues 263–265 of the α-polypeptide chain.

No indications have been obtained in the present investigation of any structural differences between the two α-subunit components of the RNA polymerase molecule.

Acknowledgements

The authors wish to thank V. S. Khokhryakov, T. M. Shuvaeva, V. V. Tyurin, N. A. Potapenko, T. M. Rodina, I. G. Akhapkina and S. A. Kocherginskaya for the participation in certain phases of this work and T. I. Muravieva for isolating RNA polymerase. The authors are grateful to Professor E. L. Smith (UCLA, USA) and Professor G. R. Drapeau (Montreal University, Canada) for the samples of *Staphylococcus* protease.

References

[1] Chamberlin, M. (1976) in: RNA Polymerase (Chamberlin, M. and Losick, R. eds) pp. 17–67, Cold Spring Harbor Lab., Cold Spring Harbor.
[2] Burgess, R. R. (1969) J. Biol. Chem. 244, 6168–6176.
[3] Ovchinnikov, Yu. A., Lipkin, V. M., Modyanov, N. N., Chertov, O. Yu., Smirnov, Yu. V., Khokhryakov, V. S. and Shuvaeva, T. M. (1977) Bioorg. Khim. 3, 283–286.
[4] Lipkin, V. M., Modyanov, N. N., Kocherginskaya, S. A., Chertov, O. Yu., Nikiforov, V. G. and Lebedev, A. N. (1976) Bioorg. Khim. 2, 1174–1181.
[5] Drapeau, G. R., Boily, Y. and Houmard, J. (1972) J. Biol. Chem. 247, 6720–6726.
[6] Bornstein, P. (1970) Biochemistry 9, 2408–2421.
[7] Fontana, A. (1972) in: Methods in Enzymology, Vol. XXV, pp. 419–423, Academic Press, NY, London.
[8] Fujiki, H., Palm, P., Zillig, W., Calender, R. and Sunshine, M. (1976) Molec. Gen. Genet. 145, 19–22.
[9] Goff, C. G. (1974) J. Biol. Chem. 249, 6181–6190.
[10] Zillig, W., Fujiki, H. and Mailhammer, R. (1975) J. Biochem. 77, 7p–8p.

Fig.1. The amino acid sequence of DNA-dependent RNA polymerase α-subunit. Designations: SP, peptides from *Staphylococcus* protease hydrolysis; Th, thermolytic peptides; Ch-Br, SP-Br-peptides from chymotryptic or *Staphylococcus* protease digestion of cyanogen bromide fragments; BNPS, peptide obtained by BNPS-skatole splitting. Amino acid sequences determined by means of the sequencer: Seq, on the protein itself; Seq-Br, on the cyanogen bromide fragment; Seq-AG, on the fragment resulting from splitting the protein at the Asn–Gly bond.