Determination of the Nucleotide Sequence for the Exonuclease I Structural Gene (sbcB) of Escherichia coli K12

(Received for publication, August 22, 1986)

Gregory J. Phillips and Sidney R. Kushner†

From the Department of Genetics, University of Georgia, Athens, Georgia 30602

The complete nucleotide sequence of the structural gene for Escherichia coli exonuclease I has been determined. The coding region corresponds to a 465-amino acid protein with molecular weight of 53,174. The partial amino acid sequence of purified exonuclease I agrees with that predicted by the DNA sequence. Two putative weak promoters have been localized by S1 nuclease analysis. The sbcB coding sequence contains many non-optimal codons, characteristic of many poorly expressed E. coli genes.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Plasmids—SK4258 is C600 transformed with the runaway replication plasmid pDPK15 (5). pDPK20 is a pBR322 derivative plasmid containing the wild type sbcB gene (6). The M13 vectors mp18 and mp19 (7) were used for construction of all templates, and JM103 (8) was used as the host strain.

Materials—All enzymes were purchased from Boehringer Mannheim or New England Biolabs and used as specified by the manufacturers. Radiochemicals ([α-32P]ATP and [γ-32P]ATP) were obtained from New England Nuclear. Agarose was obtained from FMC Corp.; acrylamide, bisacrylamide, ultra pure urea, and ammonium persulfate from Bio-Rad; TEMED from Eastman-Kodak; antibiotics from Sigma; deoxynucleotides and dideoxynucleotides, 17-mer synthetic primer, and probe-primer from P-L Biochemicals.

DNA Sequencing—The Sanger dideoxy chain termination method (9) was used to determine the DNA sequence, using the modifications recommended by Biggin et al. (10) for the use of [α-32P]ATP. Templates were generated by a combination of forced subcloning into the M13 vectors mp18 and mp19 and the exonuclease III deletion procedure of Henikoff (11).

S1 Nuclease Mapping—A modification of the Berk and Sharp (12) procedure for S1 nuclelease protection was used to determine the site of transcription initiation. RNA was extracted from SK4258 by the phenol method, employing modifications described by Markham et al. (13). End labeling of the 563-bp ClaI-EcoRI DNA fragment by T4 polynucleotide kinase and isolation of the radiolabeled fragment was as described by Maniatis et al. (14). 200 μg of RNA were hybridized with 80,000 cpm of labeled DNA fragment for 16 h at 52 °C in a solution of 80% formamide, 100 mM Pipes, pH 6.8, 400 mM NaCl, and 10 mM NaEDTA. The hybridization mixture was diluted into a 200-μl volume of S1 digestion buffer (100 mM sodium acetate, 1 mM ZnSO4, and 1 mM EDTA) and the RNA-DNA hybrids were digested with 500 units/ml of S1 nuclease (BM) for 30 min at 37 °C. Control experiments were done by incubating the identical 5'-end labeled DNA fragment with 200 μg of tRNA under equivalent conditions prior to S1 digestion. No nonspecific protection was detected. Following S1 digestion, the RNA-DNA hybrids were phenol-extracted, precipitated with 2 volumes of absolute ethanol, and suspended in 20 μl of loading buffer (15) before electrophoresis on an 8% polyacrylamide/urea gel. Maxam and Gilbert (15) sequencing reactions were performed on the same 563-bp DNA fragment and run parallel to the S1 protection experiment. A correction factor of 1 to 1.3 bases (16) was used in determining the base at which transcription initiated.

Nucleotide Sequence Analysis—The computer program DNASEQ (17), developed in the Department of Genetics, University of Georgia, was used for editing, translation, and homology searches. The Intellegenics system was used for the hydroathy plot and for additional homology searches.

Amino Acid Analysis—Exonuclease I protein was purified as previously described (6). 100 picomoles of the protein was analyzed on an Applied Biosystems Amino Acid Sequencer.

Other Methods—Strand-specific hybridization probes were made from M13 mp18 and mp19 derivative templates containing the 1.2-kb EcoRI-PstI fragment (Fig. 1). Primer extension was carried out
as outlined by Hu and Messing (18). RNA dot blots were done as previously described (19).

RESULTS

Nucleotide Sequence of sbcB—We have previously reported the localization of the sbcB coding region (5), as shown in Fig. 1. The direction of transcription of sbcB was determined by hybridizing two strand-specific probes to total cellular RNA (data not shown). The probes were generated by subcloning the 1.2-kb EcoRI-Smal segment into the M13 vectors mp18 and mp19 (7). The probes were radiolabeled by the primer

![Fig. 1. Restriction map of the sbcB gene region and sequencing strategy. The protein coding region of the sbcB gene is shown by the hatched line. The direction of transcription is shown by the upper arrow. Each lower arrow represents the direction and length of readable DNA sequences determined by the Sanger dideoxy method (9).](image)

Fig. 2. Nucleotide sequence of sbcB coding region and flanking sequences. The deduced amino acid sequence corresponds to a protein of M, 53,174. The first 12 amino acids of the coding region which are underlined correspond to the amino-terminal amino acids determined by analysis of purified enzyme I. Two sites of transcription initiation, as identified by S1 nuclease protection analysis, are indicated by a dot at the +1 position. Two putative -10 promoter regions for these transcription start sites are boxed and are indicated by "A" and "B." A -35 region corresponding to these two -10 regions is also indicated. A putative ribosome binding sequence (20) is underlined starting at the +9 position. A region of dyad symmetry, capable of forming a stem-loop structure, is underlined in the 3'-flanking sequences. This structure is reminiscent of a rho-independent termination site (21). Restriction enzyme recognition sites from Fig. 1 are also indicated.

| Amino acid | Total number |
|------------|--------------|
| Ala        | 50           |
| Leu        | 46           |
| Asp        | 37           |
| Glu        | 31           |
| Arg        | 30           |
| Pro        | 28           |
| Val        | 27           |
| Aan        | 26           |
| Phe        | 24           |
| Thr        | 22           |
| Ile        | 21           |
| Gin        | 19           |
| Gly        | 19           |
| Lys        | 19           |
| Tyr        | 19           |
| His        | 17           |
| Ser        | 13           |
| Met        | 9            |
| Trp        | 9            |
| Cys        | 5            |

Table I

Amino acid composition of sbcB deduced from the nucleotide sequence

Total number of Codons: 465; total molecular weight: 53,174.
Exonuclease I of *E. coli*

457

2

z

I

c

h

'31

III

11

30 60 90 120 150 180 210 240 270 300 330 360 390 420 450

AMINO

ACID

NUMBER

FIG. 3. Hydropathy plot (22) for the *sbcB* coding region. The hydropathy plot was determined by the method of Kyte and Doolittle (22).

A A

C

G G

A

G G+A T+C

C

SI

"FIG. 4. Determination of the in vitro transcription start sites by S1 nuclease protection analysis. The 563-bp *EcoRI-Clal* fragment, 5'-end-labeled at the *EcoRI* site, was hybridized to 200 μg of total cellular RNA form SK4258, an exonuclease I overproducing strain (6). The hybridization conditions, S1 nuclease treatment, and electrophoresis were performed as described under "Experimental Procedures." The S1-protected fragments were run in parallel with Maxam and Gilbert sequencing reactions (15) of the same *Clal-EcoRI* fragment. The arrow indicates two potential transcription initiation sites.

extension procedure described under "Experimental Procedures."

The DNA sequence of 1968 nucleotides was determined from the M13 clones shown in Fig. 1. Sets of nested deletions were produced by exonuclease III digestion of the templates employing the method of Henikoff (11). Using this procedure, greater than 95% of the gene was sequenced in both directions. Fig. 2 shows a potential ribosome binding site (20) located 9 nucleotides preceding an ATG codon which begins an open reading frame of 465 codons. In addition, a potential promoter sequence was also identified by visual inspection. No other putative regulatory sequences, e.g. regions of dyad symmetry, were found upstream of the promoter region, consistent with the notion that *sbcB* is constitutively expressed. An additional feature of the sequence is a potential stem-loop structure which is found between positions 1428 and 1449 following the open reading frame. This structure is potentially a rho-independent termination signal (21).

Amino Acid Sequence of *sbcB*—Fig. 2 also shows the predicted amino acid sequence for a translational open reading frame of 1398 nucleotides. This open reading frame, which begins with an ATG initiation codon at position 25 and terminates with a TAA at position 1421, corresponds to a 465-amino acid protein with a molecular weight of 53,174. The predicted molecular weight is in close agreement with the experimentally determined value of 55,000 for the native protein (6) and 53,700 for the denatured polypeptide (5). The predicted amino acid composition of exonuclease I is shown in Table I.

To determine if the predicted amino acid sequence corresponded to that of the purified exonuclease I protein, the first 12 amino acid residues were determined as described under "Experimental Procedures." These residues, as underlined in Fig. 2, corresponded to the predicted amino acid sequence.

It is interesting to note that there are two ATG codons which initiate the open reading frame. Amino acid sequence analysis of the protein revealed that methionine is the amino-terminal amino acid. Presumably the first methionine is removed from the mature protein.

A hydropathy plot (22) of the exonuclease I sequence is shown in Fig. 3 and reveals several hydrophilic regions, consistent with its role as a nucleic acid specific enzyme.

Mapping the Transcriptional Start Site and Promoter Characterization—in order to determine the start of transcription of the *sbcB* gene, S1 nuclease protection experiments were done. Total cellular RNA was prepared from SK4258, a strain of *E. coli* which contains the cloned *sbcB* gene on a runaway replication plasmid (2). Previous measurements had shown that exonuclease I activity was amplified up to 400-fold in this strain. Total RNA was hybridized to a 563-bp *Clai-EcoRI* fragment, 5'-terminally labeled with [γ-32P]ATP at the *EcoRI* site (Fig. 1). Fig. 4 shows the results of a typical S1 protection experiment run parallel to a Maxam and Gilbert sequencing ladder. After consideration of the 1–1.5-bp correction factor necessary when comprising a DNA fragment with a Maxam and Gilbert sequencing reaction (16), two potential sites of transcriptional initiation were identified. The same two protected DNA fragments were also found in experiments using increased concentrations of S1 nuclease. Two putative –10 promoter regions upstream from these transcriptional start sites were found upstream of the promoter region, consistent with the notion that *sbcB* is constitutively expressed. An additional feature of the sequence is a potential stem-loop structure which is found between positions 1428 and 1449 following the open reading frame. This structure is potentially a rho-independent termination signal (21).

Amino Acid Sequence of *sbcB*—Fig. 2 also shows the predicted amino acid sequence for a translational open reading frame of 1398 nucleotides. This open reading frame, which begins with an ATG initiation codon at position 25 and terminates with a TAA at position 1421, corresponds to a 465-amino acid protein with a molecular weight of 53,174. The predicted molecular weight is in close agreement with the experimentally determined value of 55,000 for the native protein (6) and 53,700 for the denatured polypeptide (5). The predicted amino acid composition of exonuclease I is shown in Table I.

To determine if the predicted amino acid sequence corresponded to that of the purified exonuclease I protein, the first 12 amino acid residues were determined as described under "Experimental Procedures." These residues, as underlined in Fig. 2, corresponded to the predicted amino acid sequence.

It is interesting to note that there are two ATG codons which initiate the open reading frame. Amino acid sequence analysis of the protein revealed that methionine is the amino-terminal amino acid. Presumably the first methionine is removed from the mature protein.

A hydropathy plot (22) of the exonuclease I sequence is shown in Fig. 3 and reveals several hydrophilic regions, consistent with its role as a nucleic acid specific enzyme.
sites can be found. These sites, designated “A” and “B”, are shown in Fig. 2. Both −10 regions appear to share a common −35 region (Fig. 2); however, the spacing between −10 region “A” and −35 is 17 bases, a distance that is highly conserved among prokaryotic promoters (23).

It has been estimated from exonuclease I purification data that there are 40 to 60 molecules/cell of this protein in logarithmically growing cultures. One explanation for the poor expression of this gene is provided by an examination of the proposed promoter region. Using the method of Mulligan et al. (24) for comparing promoter strengths, a homology score of 44.4 using −10 for region “A” and 34.9 for −10 region “B” was obtained. Such homology scores are considered indicative of a weak promoter. These values compare with score of 38.2 for lacI and 51.5 for lexA, two genes known to encode regulatory proteins present in low cellular abundance.

Codon Usage—A correlation has been observed between codon usage biases in E. coli and low cellular abundance, particularly for regulatory proteins (25). Table II shows the codon usage for the sbcB structural gene. Konigsberg and Godson (25) have identified 23 synonymous codons which are infrequently utilized in efficiently expressed, nonregulatory

| Codon | Number of codons | % synonymous codons used | % synonymous used codons | Infrequently used codons |
|-------|------------------|--------------------------|--------------------------|--------------------------|
| TTT (Phe) | 15 | 63 | 44 |
| TTC | 9 | 38 | 56 |
| TTA (Leu) | 9 | 20 | 6 |
| TTG | 7 | 15 | 8 |
| CTT | 5 | 11 | 9 |
| CTC | 2 | 4 | 7 |
| CTA | 2 | 4 | 2 |
| CTG | 21 | 46 | 69 |
| ATT (Ile) | 16 | 76 | 37 |
| ATC | 3 | 14 | 62 |
| ATA | 2 | 10 | 1 |
| ATG (Met) | 9 |
| GTT (Val) | 6 | 22 | 38 |
| GTC | 4 | 15 | 13 |
| GTA | 5 | 19 | 23 |
| GTG | 12 | 44 | 27 |
| TCT (Ser) | 1 | 8 | 27 |
| TCC | 1 | 8 | 26 |
| TCA | 1 | 8 | 8 |
| TCG | 3 | 23 | 11 |
| CCT (Pro) | 5 | 18 | 9 |
| CCC | 5 | 18 | 5 |
| CGA | 4 | 14 | 20 |
| CGG | 14 | 50 | 65 |
| ACT (Thr) | 3 | 14 | 24 |
| ACC | 12 | 54 | 51 |
| ACA | 2 | 9 | 6 |
| ACG | 5 | 23 | 20 |
| GCT (Ala) | 4 | 8 | 28 |
| GCC | 18 | 36 | 19 |
| GCA | 15 | 30 | 23 |
| GCG | 13 | 13 | 30 |
| TAT (Tyr) | 10 | 59 | 41 |
| TAC | 7 | 41 | 59 |

* Codon usage observed in a compilation of 25 efficiently expressed, nonregulatory E. coli genes (25).

Table III

| Reading frame | sbC | Regulatory genes | Nonregulatory genes |
|--------------|-----|------------------|---------------------|
| 1 (coding)   | 22.2 | 24.1 | 12.1 |
| 2             | 36.3 | 27.3 | 36.5 |
| 3             | 34.6 | 31.2 | 30.8 |

E. coli genes. The percent synonymous codon usage for these genes is also shown in Table II. For 18 of the 23 codons designated as infrequently used, the sbC gene has a higher than expected frequency of occurrence. In addition, it is known that these 23 rare codons occur at a higher frequency in the 2 out-of-frame sites than in the reading frame for highly expressed proteins, while for poorly expressed genes their occurrence is nearly equal across the three reading frames (25).
sbcB is very similar to that of the poorly expressed regulatory proteins.

Homology Comparisons—A comparison of the amino acid sequence of exonuclease I with that of the lambda exonuclease and the T7 gene 6 exonuclease revealed no significant homology. In addition, no significant homology was seen with E. coli DNA polymerase I, which possesses both 3' to 5' and 5' to 3' exonucleolytic activities (26).

DISCUSSION

We have reported the complete nucleotide sequence of the E. coli sbcB gene, encoding the enzyme exonuclease I. Characterization of the 5'-flanking sequences indicate that sbcB is constitutively expressed from two possible inefficient promoters. In addition, similar codon usage biases with those determined for low abundance regulatory proteins have been observed. Since the ribosome binding site appears more than adequate, it is not clear at this time if the low abundance of the exonuclease I protein is the result of either poor transcription or translation or some combination of both factors. However, placing the sbcB structural gene under the control of a T7 promoter only resulted in a 2-fold increase in activity over that previously reported (6).

Acknowledgment—We wish to thank Suzette Lay for assistance in preparation of the manuscript.

REFERENCES

1. Lehman, I. R., and Nussbaum, A. L. (1964) J. Biol. Chem. 239, 2628–2636
2. Kushner, S. R., Nagaishi, H., Templin, A., and Clark, A. J. (1971) Proc. Natl. Acad. Sci. U. S. A. 68, 824–827
3. Kushner, S. R., Nagaishi, H., and Clark, A. J. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1366–1370
4. Vapnek, D., Alton, N. K., Bassett, C. L., and Kushner, S. R. (1976) Proc. Natl. Acad. Sci. U. S. A. 73, 3492–3496
5. Prasher, D., Kasunic, D. A., and Kushner, S. R. (1983) J. Bacteriol. 153, 903–906
6. Prasher, D. C., Conarro, L., and Kushner, S. R. (1983) J. Biol. Chem. 258, 6340–6343
7. Ornander, J., Kempe, T., and Messing, J. (1983) Gene (Amst.) 26, 101–106
8. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309–321
9. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
10. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3963–3965
11. Henikoff, S. (1984) Gene (Amst.) 28, 351–359
12. Berk, A. J., and Sharp, P. A. (1977) Cell 12, 721–732
13. Markham, B. E., Harper, J. E., Mount, D. W., Sancar, G. B., Sancar, A., Rupp, W. D., Kenyon, C. J., and Walker, G. C. (1984) J. Mol. Biol. 178, 237–248
14. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
15. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
16. Sollner-Webb, B., and Reeder, R. H. (1979) Cell 18, 485–499
17. Arnold, J., Eckenrode, V. K., Lemke, K., Phillips, G. J., and Schaeffer, S. W. (1986) Nucleic Acids Res. 14, 239–254
18. Hu, N., and Messing, J. (1982) Gene (Amst.) 17, 271–277
19. Thomas, P. S. (1983) Methods Enzymol. 101, 255–266
20. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 1342–1346
21. Rosenberg, M., and Court, D. (1979) Annu. Rev. Genet. 13, 319–353
22. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132
23. Hawley, D. K., and McClure, W. R. (1983) Nucleic Acids Res. 11, 2237–2255
24. Mulligan, M. E., Hawley, D. K., Entriken, R., and McClure, W. R. (1984) Nucleic Acids Res. 12, 789–800
25. Konigsberg, W., and Godson, G. N. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 637–640
26. Kornberg, A. (1980) DNA Replication, W. H. Freeman and Co., San Francisco

\(^2\) G. J. Phillips and S. R. Kushner, unpublished results.