Volume 13 Number 13 1985 Nucleic Acids Research

# The DNA sequence of the gene (*rnc*) encoding ribonuclease III of *Escherichia coli*

Paul E.March, Joohong Ahnn and Masayori Inouye

Department of Biochemistry, State University of New York, Stony Brook, NY 11794, USA

Received 10 May 1985; Accepted 5 June 1985

ABSTRACT

The DNA sequence of a 1,076 base pair BglI-BamHI fragment containing the entire rnc gene for ribonuclease III (RNase III) was determined. An open reading frame of 681 base pairs was found in this region which encodes a protein of 227 amino acid residues (calculated molecular weight = 25,218). When this open reading frame was cloned into a high expression vector, pIN-III, a protein of apparent molecular weight of 26,000 was produced upon induction of the cloned gene. This product accounted for up to 5% of the total cellular protein, and comigrated with purified RNase III. RNase III enzyme activity was induced in parallel with the production of the 26,000 molecular weight protein. A putative promoter was found 170 base pairs upstream from the initiation codon. In the long leader region a very stable stem-bulge-stem structure was found which closely resembles typical RNase III cleavage sites. This structure may be cleaved by RNase III to auto-regulate the expression of the rnc gene.

## INTRODUCTION

The rnc gene of Escherichia coli has been shown to be responsible for production of ribonuclease III (RNase III[1]) (1), an important enzyme involved in processing rRNA and mRNA precursors(2). This gene is located at 55 min on the E. coli chromosome (3) and was found to be closely linked to the lep operon(4). The DNA sequence of the lep operon has recently been completed in our laboratory (5).

In this report we determined the DNA sequence of the rnc gene and flanking regions, which were found to be immediately downstream of the lep operon. The rnc gene encodes a very basic protein of 227 amino acid residues. Interestingly in the putative leader sequence of the mRNA, upstream from the initiation codon, a stable potential secondary structure can be formed which is analogous to a RNase III cleavage site. A possible role of this secondary structure is discussed in terms of auto-regulation of the rnc gene by RNase III.

MATERIALS AND METHODS

Reagents and materials: Enzymes employed for DNA manipulations (restriction enzymes, T4 DNA ligase and the large fragment of DNA polymerase I) were supplied by Bethesda Research Laboratories. IPTG[1] was obtained from Sigma Chemicals. Deoxy- and dideoxynucleotides were obtained from P-L Biochemicals.

Strains and plasmids: E. coli K12 strain SB221 (lpp lacY hsdR ΔtrpE5 leuB6 recA / F' lacI$^q$ lac pro) (6) was utilized in expression experiments.

Sequencing vectors M13mp8, M13mp18, and M13mp19 and their host strain JM103 were obtained from Bethesda Research Laboratories.

The construction of plasmid pTD101 was described by Date and Wickner(7). Sequencing data indicated that the open reading frame encoding RNase III was contained on a 912 bp[1] PstI-BamHI fragment (see Fig. 2). Therefore pTD101 was digested with PstI and BamHI followed by treatment with the large fragment of DNA polymerase I in the presence of the four dXTPs. The resulting blunt ended 912 bp fragment was purified by gel electrophoresis on a 5% polyacrylamide gel. This fragment was then cloned into the unique XbaI site of the expression vector pIN-III (8). The 5' overhang structures of the XbaI site were filled in by the large fragment of DNA polymerase I to produce blunt ends before ligation to the 912 bp fragment. Transformants were selected in which the rnc fragment was oriented with the BamHI site distal to the lpp- lac promoters; one such transformant was designated pJHA002. The cloned gene in pJHA002 is thus under the control of the strong lpp promoter and the lac promoter-operator such that it is expressed only in the presence of a lac inducer such as IPTG (8).

Expression of RNase III from pJHA002: E. coli SB221 harboring pJHA002 was grown at 37°C in M9-glucose medium (9) supplemented with 20 μg/ml tryptophan, 50 μg/ml ampicillin, and 0.5% Casamino acids. At a density of approximately 4 x $10^8$ cells/ml, IPTG was added to a final concentration of 2 mM. The incubation was continued for another 6h at which point cells were collected by centrifugation and resuspended in a buffer containing 1% SDS[1], 10% glycerol, 10 mM sodium phosphate (pH 7.1), 2% β-mercaptoethanol, and 0.05% Bromphenol blue. Cells were lysed by incubating in a boiling water bath for 10 min and their protein content was analyzed by SDS polyacrylamide gel electrophoresis.

Other procedures: DNA sequencing was carried out according to the method of Sanger et al. (10). SDS polyacrylamide gel electrophoresis was performed using the gel system described by Anderson et al. (11). RNase III enzyme

activity was measured in a crude lysate as described by Robertson et al. (12). The pI of RNase III was estimated from the predicted amino acid sequence using a computer program purchased from International Biotechnologies Inc. (New Haven, CT).

## RESULTS

DNA Sequence Determination: It was recently reported that the rnc gene is present on an 1.3 kb[1] EcoRI- BamHI fragment which is contained on the plasmid pTD101 (13). This plasmid also contains the lep operon, the entire DNA sequence of which has been recently completed in our laboratory (5). Therefore, using pTD101, the sequence determination was extended to the downstream region of the lep operon to determine the DNA sequence of the rnc gene according to the strategy shown in Figure 1. The entire sequence between the upstream BglI site and the downstream BamHI site was obtained from both strands. The resulting DNA sequence is presented in Figure 2.

Analysis of the DNA sequence reveals that there is a large open reading frame which is initiated by an ATG codon at residue 246 and ends at residue 927 with a TGA stop codon. This initiation codon is preceded by a ribosome binding sequence (-GGT-) which is considered to be rather inefficient (14,15). The spacing between the ribosome binding site and the initiation codon is five bases which is the shortest among the known ribosome binding sites (15). The amino acid sequence from the open reading frame is deduced from the DNA sequence and, as shown in Figure 2, a protein of 227 amino acid residues would

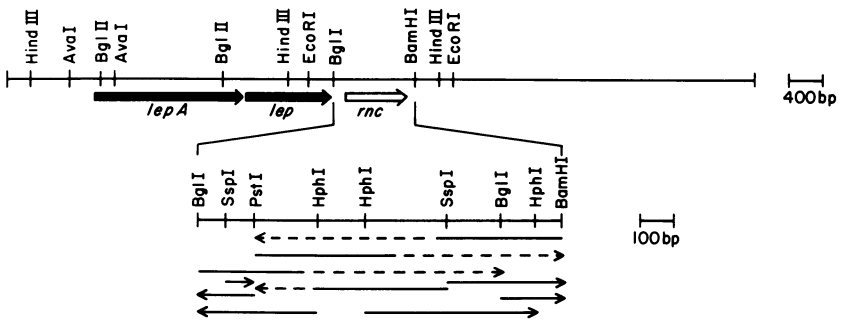
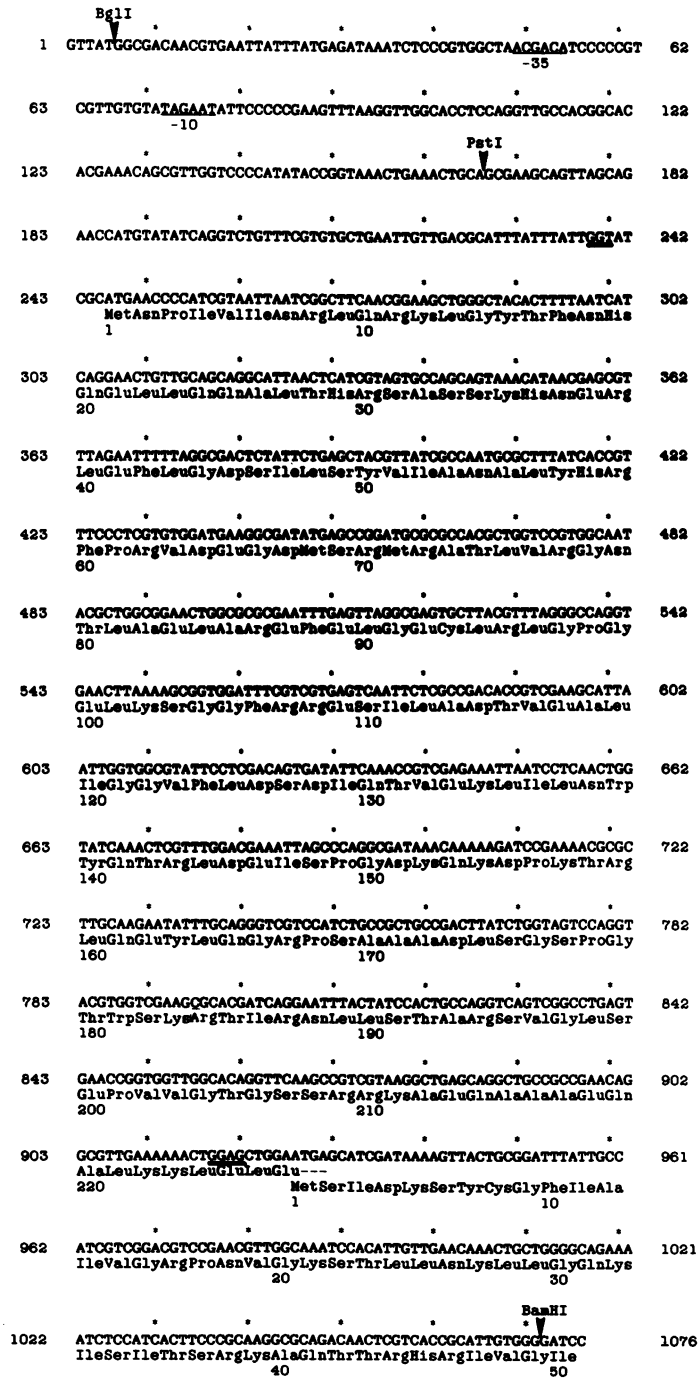



Fig. 1 The strategy employed to sequence the rnc gene. A linear restriction enzyme map of pTD101 is shown at top. Below this the region sequenced is shown. Within this region only the restriction enzyme sites employed to obtain fragments for sequencing are shown. The direction of sequencing on each fragment is shown at bottom. On long fragments unsequenced regions are indicated by the dashed line.


BglI
-35
-10
PstI
BamHI

be encoded. Its predicted molecular weight is calculated to be 25,218 which is in excellent agreement with the reported apparent molecular weight of RNase III of 25,000 (16).

Identification of the Gene Product of the Open Reading Frame: The open reading frame shown in Figure 2 was cloned into an expression cloning vector, pIN-III, by inserting the 912-bp PstI- BamHI fragment into the unique XbaI site of the vector. Thus, in this construction, the expression of the open reading frame can be controlled by a lac inducer such as IPTG. As shown in Figure 3, in the presence of IPTG a large amount of a protein with an apparent molecular weight of 26,000 is produced (lane 2); in the absence of IPTG, no such product was observed (lane 1). This protein co-migrated with a purified RNase III standard (position indicated by an arrow). In order to determine whether there was any IPTG-inducible RNase III activity, we prepared a lysate by sonicating cells suspended in 10 mM sodium phosphate buffer (pH 7.1). When the lysates prepared from $6 \times 10^6$ cells were used for the RNase III assay, activity was at least 10-fold greater than in the absence of IPTG. These data indicate that the protein of molecular weight 26,000 which is produced from the open reading frame shown in Figure 2 is RNase III. By densitometric analysis, RNase III produced after a 6h induction by IPTG was estimated to be approximately 5% of the total cellular protein.

## DISCUSSION

The DNA sequence of the rnc gene reveals that RNase III possesses 78 polar and 61 charged residues (see Table 1). Of the charged residues, 32 are basic and 29 are acidic, giving the protein an estimated pI of 9.8. The codon usage summarized in Table 1 shows that RNase III employs many minor codons (17), consistent with the fact that RNase III is a minor protein of E. coli. In fact, when the rnc gene was cloned into pBR322, the production of RNase III under its own promoter was very poor (13).

A putative promoter for the rnc gene is found approximately 170 bp upstream from the site of initiation of translation (see Figure 2). At residue 49 the sequence acGACA [upper case letters agree with the consensus

Fig. 2 The DNA sequence of the 1076-bp BglI-BamHI fragment containing the rnc gene. The BglI and BamHI sites are indicated along with the PstI site which was employed to construct the expression vector. The proposed -35 and -10 regions are indicated by an underline and ribosome binding sites discussed in the text are indicated by a heavy underline. The nucleic acid sequence is numbered at the margins and the amino acid sequence of the translated open reading frames is numbered below the amino acid sequence.

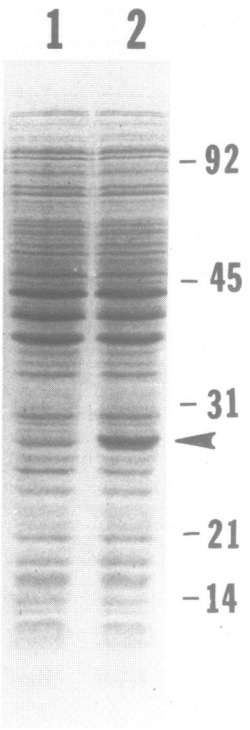



Fig. 3 Expression of the rnc gene product from pJHA002. Lane 1, E. coli strain SB221 harboring pJHA002 grown in the absence of IPTG. Lane 2, as in lane 1, except cells were grown in the presence of IPTG. The position of molecular weight markers and their molecular weight ($x10^{-3}$) is indicated at the right. The position of migration of purified RNase III is indicated by the arrow. Growth conditions were described in the Materials and Methods, and electrophoresis was performed in a 17% SDS polyacrylamide gel.

(18,19)] is a possible -35 region, and 17 bases downstream from the -35 region (residue 72) the sequence TAgAAT forms a possible -10 region for the rnc promoter. Other possible promoter sequences could not be found within the 170 bp leader sequence. It is interesting to note that in this long leader sequence an extensive secondary structure can be formed as shown in Figure 4. This stem-bulge-stem motif is typical of RNase III processing sites (20). It is intriguing to speculate that RNase III processes its own mRNA to autoregulate its gene expression. Possibly when too much RNase III is produced it may catalyze the degradation of its own mRNA to reduce the production of RNase III. In this regard it should be pointed out that the PstI site employed for cloning the rnc gene into the pIN-III vector eliminates this secondary structure (see Figure 4). This may be a major reason why such large amounts of RNase III can be produced from our clone.

Downstream of the rnc gene, there was no obvious transcription termination signal. Instead, there is an open reading frame starting from an initiation codon ATG at residue 926 which overlaps the TGA termination codon

TABLE I
Amino Acid Composition and Codon Usage

| Amino Acid | Total Residues | Codon | Number Used | % Codon Usage* | % Codon Usage High Expression** | Amino Acid | Total Residues | Codon | Number Used | % Codon Usage* | % Codon Usage High Expression** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Phe | 6 | TTT | 4 | 67% | 23% | Tyr | 5 | TAT | 3 | 60% | 17% |
| | | TTC | 2 | 33 | 77 | | | TAC | 2 | 40 | 83 |
| Leu | 31 | TTA | 10 | 32 | 3 | His | 4 | CAT | 3 | 75 | 32 |
| | | TTG | 5 | 16 | 4 | | | CAC | 1 | 25 | 68 |
| | | CTT | 3 | 10 | 4 | | | | | | |
| | | CTC | 3 | 10 | 3 | Gln | 11 | CAA | 5 | 45 | 14 |
| | | CTA | 1 | 3 | 0 | | | CAG | 6 | 55 | 86 |
| | | CTG | 9 | 29 | 86 | | | | | | |
| | | | | | | Asn | 8 | AAT | 5 | 62 | 5 |
| Ile | 10 | ATT | 6 | 60 | 21 | | | AAC | 3 | 38 | 95 |
| | | ATC | 4 | 40 | 79 | | | | | | |
| | | ATA | 0 | 0 | 0 | Lys | 11 | AAA | 8 | 73 | 71 |
| | | | | | | | | AAG | 3 | 27 | 29 |
| Met | 3 | ATG | 3 | 100 | 100 | | | | | | |
| | | | | | | Asp | 10 | GAT | 5 | 50 | 31 |
| Val | 10 | GTT | 2 | 20 | 44 | | | GAC | 5 | 50 | 69 |
| | | GTC | 4 | 40 | 7 | | | | | | |
| | | GTA | 2 | 20 | 31 | Glu | 19 | GAA | 12 | 63 | 75 |
| | | GTG | 2 | 20 | 17 | | | GAG | 7 | 37 | 25 |
| Ser | 19 | TCT | 3 | 16 | 44 | Cys | 1 | TGT | 0 | 0 | 20 |
| | | TCC | 1 | 5 | 31 | | | TGC | 1 | 100 | 80 |
| | | TCA | 3 | 16 | 2 | | | | | | |
| | | TCG | 1 | 5 | 0 | Trp | 2 | TGG | 2 | 100 | 100 |
| | | AGT | 5 | 26 | 6 | | | | | | |
| | | AGC | 6 | 32 | 17 | Arg | 21 | CGT | 12 | 57 | 72 |
| | | | | | | | | CGC | 4 | 19 | 26 |
| Pro | 8 | CCT | 1 | 13 | 6 | | | CGA | 0 | 0 | 0 |
| | | CCC | 1 | 13 | 3 | | | CGG | 3 | 14 | 0 |
| | | CCA | 4 | 50 | 18 | | | AGA | 0 | 0 | 0 |
| | | CCG | 2 | 25 | 74 | | | AGG | 2 | 10 | 0 |
| Thr | 12 | ACT | 4 | 33 | 48 | Gly | 18 | GGT | 7 | 39 | 58 |
| | | ACC | 2 | 17 | 41 | | | GGC | 9 | 50 | 40 |
| | | ACA | 1 | 8 | 7 | | | GGA | 1 | 6 | 1 |
| | | ACG | 5 | 42 | 3 | | | GCG | 1 | 6 | 1 |
| Ala | 18 | GCT | 3 | 17 | 50 | | | | | | |
| | | GCC | 9 | 50 | 6 | | | | | | |
| | | GCA | 2 | 11 | 26 | | | | | | |
| | | GCG | 4 | 22 | 18 | | | | | | |

* % Codon Usage = Number of occurrences of codon ÷ Total codons of a specific residue

** % Codon Usage High expression: calculated as described above, but for a typical highly expressed protein (calculated from data of Gouy and Gautier, 1982).

of the rnc gene (underlined nucleotides denote the overlapping bases). This ATG is preceded by a ribosome binding site (14), GAGG, with a spacing of five nucleotides. The downstream open reading frame continues to the BamHI site without termination codons (see Figure 2); thus, it is possible that RNase III is produced from an operon together with at least one other protein. The data in Figure 2 show that the coding region of the downstream protein contains an amino-terminal domain of 50 amino acid residues which is very rich in basic amino acids. In a previous paper by Watson and Apirion (13), a

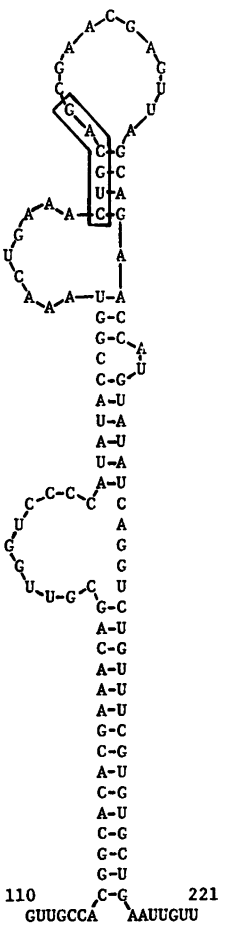



Fig. 4 Possible secondary structure of the rnc gene leader region. A stability of $\Delta G=-39.9$ kcal was calculated, as described by Salser (21), for the structure shown. The PstI site shown in Fig. 2 and employed to construct the expression plasmid pJHA002 is indicated by the boxed residues. The numbers refer to the DNA sequence shown in Fig. 2.

protein of apparent molecular weight 24,700 was detected together with RNase III when a DNA fragment extending towards the 3' end was cloned into pBR322; this protein may be the second protein of the rnc operon.

ACKNOWLEDGEMENTS

We thank Dr. H. Robertson for assaying RNase III activity in a cell lysate and for purified RNase III preparation used as a standard in SDS-polyacrylamide gel electrophoresis. We are grateful to Jack Coleman for critical reading of the manuscript. We thank the National Institute of General Medical Sciences (Grant GM19043) and the American Cancer Society (Grant NP387I) for support of this research. P.E.M. is an American Cancer Society postdoctoral fellow.

Abbreviations used: RNase III, ribonuclease III; IPTG, isopropyl-β-D-galactoside; bp, base pair; SDS, sodium dodecyl sulfate; kb, kilobases.

REFERENCES


1. Kindler, P. Keil, T.U. and Hofschneider, P.H. (1973) Molec. Gen. Genet. 126, 53-69.
2. Dunn, J.J. and Studier, F.W. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 3296-3300.
3. Studier, F.W. (1975) J. Bacteriol. 124, 307-316.
4. Silver, P. and Wickner, W. (1983) J. Bacteriol. 154, 569-572.
5. March, P.E. and Inouye, M. (1985) J. Biol. Chem., in press.
6. Nakamura, K., Masui, Y. and Inouye, M. (1982) J. Mol. Appl. Genet. 1, 289-299.
7. Date, T. and Wickner, W. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 6106-6110.
8. Masui, Y., Coleman, J. and Inouye, M. (1983) in Experimental Manipulation of Gene Expressions (ed. M. Inouye) pp. 15-32, Academic Press, New York.
9. Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
10. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
11. Anderson, C.W., Baum, P.R. and Gesteland, R.F. (1973) J. Virol. 12, 241-252.
12. Robertson, H.D., Webster, R.E. and Zinder, N.D. (1968) J. Biol. Chem. 243, 82-91.
13. Watson, N. and Apirion, D. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 849-853.
14. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346.
15. Kozak, M. (1983) Microbiol. Rev. 47, 1-45.
16. Dunn, J.J. (1976) J. Biol. Chem. 251, 3807-3814.
17. Gouy, M. and Goutier, C. (1982) Nuc. Acid Res. 10, 7055-7074.
18. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
19. Hawley, D. and McClure, W. (1983) Nuc. Acid Res. 11, 2237-2255.
20. Studier, F.W. and Dunn, J.J. (1983) J. Mol. Biol. 166, 477-536.
21. Salser, W. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 985-1002.