Volume 12 Number 12 1984 Nucleic Acids Research

The nucleotide sequence of the *Escherichia coli* K12 *nusB (groNB)* gene

John Swindle*, Janet Ajioka*, Dean Dawson+, Richard Myers, Dana Carroll and Costa Georgopoulos

Department of Cellular, Viral and Molecular Biology, University of Utah School of Medicine, University of Utah, Salt Lake City, UT 84132, USA

Received 12 March 1984; Accepted 18 April 1984

ABSTRACT

The nusB (groNB) gene product of Escherichia coli plays a pivotal role in allowing bacteriophage λ N protein to function as an antiterminator of mRNA transcription and in modulating host gene expression. In addition it is essential for bacterial viability since mutations in it result in a cold-sensitivity phenotype for growth. We have previously cloned the nusB gene and shown it to code for a 14,500-Mr protein. Here we present the primary DNA sequence of the nusB gene. From the sequence we deduce that it codes for a slightly basic protein (21 basic as opposed to 20 acidic amino acids) composed of 139 amino acids with a cumulative 15,689-Mr. The predicted N-terminal amino acid sequence as well as the overall amino acid composition agrees well with that of the purified protein.

INTRODUCTION

Upon infection of E. coli bacteriophage λ-specific RNA transcription is initiated by the host RNA polymerase enzyme at the two early phage promoters, P_L and P_R. Such transcription terminates at various terminator sites, termed t_R in the direction of rightward transcription and t_L in the direction of leftward transcription (reviewed in reference 1). In the presence of the phage-coded N protein, however, the transcription complex becomes refractory to such (as well as other) termination sites resulting in efficient expression of downstream genes. For this to occur functional forms of the host nusA, nusB, nusC and nusE gene products must be present (1,2). Cis-acting regulatory sequences, termed nut and boxA, located between the promoters and terminators have also been shown by mutational analysis to be essential for the formation of such termination-resistant transcriptional complexes (2,3,4,5).

Recent evidence suggests that the nusA protein plays a major role in the recognition of the boxA sequence (5) and that it functionally interacts with the nusB protein (6). The nusB locus has been shown to map at 11 min on the revised E. coli map (7,8,9) and to code for a polypeptide of 14,500-Mr (10,11). Some nusB mutations result in a cold-sensitive phenotype for

bacterial growth (12) suggesting that the nusB protein is an essential E. coli protein. Surprisingly, in the absence of N function both the nusA and nusB proteins have been shown to act as transcriptional termination factors (13,14,15). In this paper we report the DNA sequence of the nusB gene of E. coli K12. The N-terminal amino acid sequence of the purified protein and its overall amino acid composition agree with that predicted by the DNA sequence.

MATERIALS AND METHODS

Bacterial and phage strains. The various E. coli K12 nusB$^+$(groNB$^+$) and nusB$^-$(groNB$^-$) bacterial strains used during the course of this study have been previously described (10,12). The λ cI857nusB$^+$(groNB$^+$) and λ cI857nusB$^+$(groNB$^+$) Δ1 transducing phages have also been previously described (10,12). In this paper we will use the designation nusB (8) instead of groNB, the designation previously used in our laboratory.

Plasmids. The plasmids used were pBR322 (16), pJB8 (17), pLC28 (18) and pMOB45 (19).

Media. L-broth and M9 minimal medium were as previously described (10,12). For selection of a particular drug resistance marker the L-broth or L-agar plates were supplemented with 20 ug/ml of the appropriate antibiotic, tetracycline (tet), chloramphenicol (cam) or ampicillin (amp).

Plasmid DNA extraction and transformations. The procedures followed were essentially those outlined by Maniatis et al.(20). Plasmid and phage DNAs were purified at least once through a CsCl equilibrium gradient before use (20).

Restriction enzyme digestions. The procedures followed were essentially those recommended by the manufacturers. Restriction endonucleases EcoRI, EcoRV, Sau3A, AluI, and BamHI were purchased from Bethesda Research Laboratories (BRL) and FokI, PvuI, BalI, HindIII and AvaI were purchased from New England Biolabs.

Ligations. DNA ligations were carried out with T4 DNA ligase (New England Biolabs) as described by Maniatis et al. (20).

DNA sequencing procedures. The basic methodology for labelling DNA with T4 DNA polymerase or AMV reverse transcriptase was that described by Maniatis et al. (20). DNA sequencing followed the procedures outlined by Maxam and Gilbert (21). [α-^{32}P]dNTPs were purchased from Amersham. T4 DNA polymerase was purchased from BRL and AMV reverse transcriptase from Life Sciences, Inc.

Selection for nusB$^+$ transformants. We took advantage of our previous observation that certain nusB mutations (such as nusB2a-1) result in cold

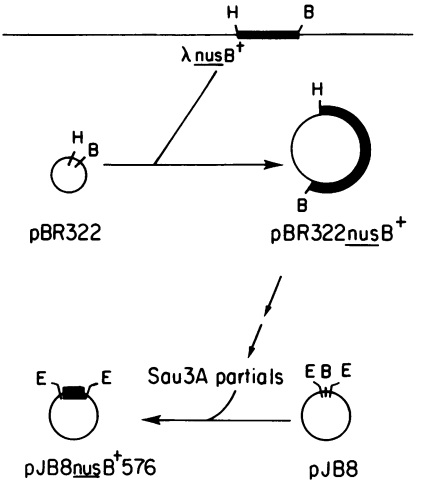

Figure 1. Schematic representation of the cloning of a minimal nusB gene. The nusB-containing HindIII-BamHI fragment of transducing phage λnusB+ (12) was cloned into the corresponding restriction sites of pBR322 resulting in the production of plasmid pBR322nusB+. Partial digestion products of pBR322nusB+ produced with Sau3A restriction endonuclease were cloned into the unique BamHI site of plasmid pMOB45 (19). This resulted in the isolation of plasmid pMOB45nusB+ containing 1,400 bp of bacterial DNA (steps not shown, indicated by arrows). Partial digestion products of plasmid pMOB45nusB+ with Sau3A restriction endonuclease were cloned into the unique BamHI site of plasmid pJB8 (17) resulting in the isolation of plasmid pJB8nusB+576 (see text).

sensitivity for bacterial growth at 30°C (12). NusB2a-1 bacteria transformed with a recombinant plasmid carrying a functional nusB+ gene were distinguished on the basis of their ability to form large colonies after 16 hr at 30°C. Upon further testing they were shown, as expected, to have simultaneously acquired the ability to allow propagation of phage λ.

One- and two-dimensional protein analysis. The procedures for electrophoresis followed those previously described (22).

RESULTS

Cloning of a DNA fragment containing a minimal nusB+ gene. Originally we constructed the λ cI857nusB+ transducing phage by the insertion of phage λ cI857 into the tsx bacterial locus, located near nusB (7,8,9), followed by its excision. This abnormal recombination event resulted in an approximately 7,700 bp bio-type deletion/substitution of λ-specific DNA sequences by the bacterial nusB+ and neighboring genes (Fig. 1; (10)). A DNA fragment produced after digestion with the BamHI and HindIII restriction endonucleases was

cloned into the corresponding restriction sites of plasmid pBR322 (Fig. 1). *E. coli nusB*$^-$ bacteria when transformed with such a recombinant plasmid were shown to become simultaneously Nus$^+$ for growth of phage λ at all temperatures and for bacterial growth at 30°C. Sub-fragments of this bacterial DNA, generated after partial digestion with Sau3A restriction endonuclease were eventually cloned into the unique BamHI site of plasmid pJB8 (17) after the series of steps indicated in Fig. 1. Among the pJB8*nusB*$^+$ recombinant clones (as judged by their ability to transform *nusB*2a-1 bacteria to a Cs$^+$ phenotype at 30°C) we selected the one with the smallest bacterial DNA insert, ∿600 bp, called pJB8*nusB*$^+$576. The advantage of using plasmid pJB8 in this cloning procedure is that its unique BamHl restriction site is immediately flanked by EcoRI sites (17). As a consequence, digestion of pJB8*nusB*$^+$576 with EcoRI restriction endonuclease releases a ∿ 600 bp DNA fragment that contains a functional *nusB*$^+$ gene. After subcloning this EcoRI-derived ∿ 600 bp DNA fragment into pBR322 we showed that all pBR322*nusB*$^+$576 recombinant molecules transformed *nusB*2a-1 bacteria to a Nus$^+$ phenotype regardless of the orientation of the bacterial DNA segment in relation to the vector. This observation suggests, but does not prove, that the ∿ 600 bp DNA fragment may carry an intact promoter for the *nusB*$^+$ gene. Alternatively, read-through transcription from plasmid-specified promoters could be responsible for *nusB* gene expression. By subcloning the EcoRV fragment of plasmid pBR322*nusB*$^+$576 we demonstrated that the *nusB*-specific DNA sequence located approximately 40 bp from one end (see below) is not essential for NusB function. This result suggests that a minimal *nusB* gene is made up of less than 540 bp.

Sequencing strategy. Plasmid pJB8*nusB*$^+$576 was digested to completion with EcoRI restriction endonuclease and the ∿600 bp DNA fragment was separated from the vector and purified after agarose gel electrophoresis (20). The two cohesive ends were labelled by filling in with T4 DNA polymerase in the presence of [α^{32}-P]dATP and excess cold dCTP, dGTP and dTTP (20). The labelled molecules were cut asymmetrically with either AluI or HinfI restriction endonucleases (Fig. 2A). The resulting labelled DNA fragments were separated after agarose gel electrophoresis and sequenced following the procedures outlined by Maxam and Gilbert (21). In order to obtain more and better DNA sequencing data, especially from the middle section of the *nusB*$^+$ gene, pBR322*nusB*$^+$576 was digested with either PvuI, StuI or FokI, labelled with T4 DNA polymerase or AMV reverse transcriptase and cut asymmetrically as indicated in the legend of Fig. 2. Sequencing of the *nusB*-specific DNA fragments gave the results shown in Fig. 2B. The *E. coli* DNA sequence

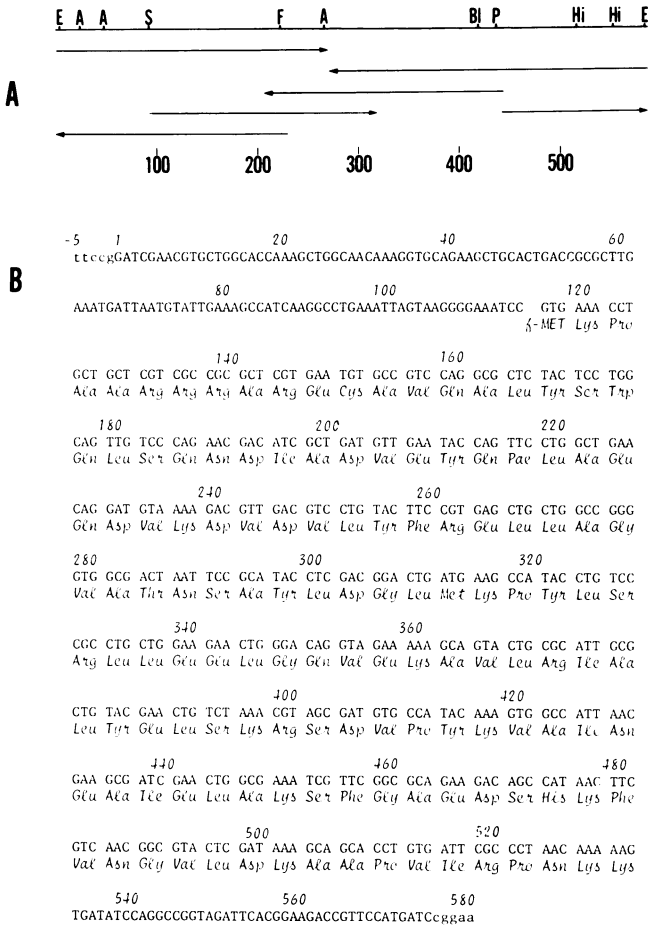

Figure 2. Sequencing strategy and DNA sequence of the nusB$^+$ gene.
A. Partial restriction map and diagram of sequencing analysis. The following restriction endonuclease cleavage sites are shown: E, EcoRI (part of the pJB8-specific DNA sequence); A, AluI; S, StuI; F, FokI; Bl, BalI; P, PvuI; Hi, HinfI. The labelled termini of the restriction fragments are at the beginning of the arrows and the extent of DNA sequencing is indicated by the end of the arrows. Plasmid pJB8nusB$^+$ DNA was cleaved with either PvuI or StuI and the resulting fragments labelled with [α-^{32}P]dGTP after treatment with T4 DNA polymerase in the presence of excess cold dATP, dCTP and dTTP. The labelled fragments were cleaved asymmetrically with EcoRI and BalI respectively and the nusB-specific fragments were sequenced. Plasmid pBR322nusB$^+$576 was cleaved with FokI and the resulting fragments labelled with [α-^{32}P]dTTP using AMV reverse transcriptase. The nusB-specific fragment thusly labelled was sequenced.
B. DNA sequence of the nusB gene. The nusB DNA sequence is shown in capital letters. The neighboring pJB8-specific sequence of plasmid pJB8nusB$^+$576 is shown in small letters. The amino acid composition of the nusB protein is shown starting with the GTG initiation codon at position 117.

consists of 576 nucleotides and, as expected, from the cloning strategy shown in Fig. 1, starts and ends with a Sau3A restriction sequence. Only two major open reading frames exist that can code for polypeptides greater than 10,000-Mr. The first open reading frame, shown in Fig. 2B, starts with a GTG initiation codon at nucleotide 117 and terminates with a TGA codon at nucleotide 534. It results in the production of a 139 amino acid long, slightly basic polypeptide of 15,689-Mr. It contains 21 basic amino acids (9 arg, 11 lys and 1 his) and 20 acidic amino acids (9 asp and 11 glu). In addition, it is extremely rich in non-polar amino acids such as ala (18 residues), leu (17 residues) and val (13 residues). The EcoRV restriction site, GATATC, mentioned in the text, begins at position 535 and includes part of the termination codon. This explains why cloning at the EcoRV site does not inactivate the NusB function.

It is not known whether the 576 bp minimal *nusB*$^+$ DNA fragment possesses its own promoter or not. Casual inspection of the sequences upstream of the GTG initiation codon did not reveal an obvious RNA polymerase -35 recognition and -10 binding sites (23). The *nusB*$^+$ DNA sequence was searched using a recently developed computer program, called TARGSEARCH (24), that evaluates potential *E. coli* polymerase promoter sequences. The computer was asked to find potential promoters by demanding 4 out of 6 matches with the -35 position TTGACA concensus sequence and 4 out of 6 matches with the -10 TATAAT concensus sequence (23,25). A potential -35 position sequence starting at nucleotide 50, CTGACC, and a -10 position sequence starting at nucleotide 73, TGTATT, were indicated. Lowering the stringency of the required match (4 out of 6 at position -35 and 3 out of 6 at -10) revealed the existence of two additional potential promoter sequences. One has the -35 sequence starting at nucleotide 77, TTGAAA, and the -10 sequence at position 101, TAGTAA, and the other has the -35 position starting at nucleotide 60, TTGAAA, and the -10 position starting at nucleotide 85, CATCAA. None of these potential promoter sequences appears to be a "good" promoter (23,25) and it is not known whether any of them are used *in vivo* or *in vitro*. A good Shine-Dalgarno ribosomal binding site sequence (26), TAAGGGG, is found at position 104, the expected distance from the GTG initiation codon (27). This is the only potential translational initiation site that the "Preceptron" program (28) found for the entire DNA sequence in the polarity of the *nusB* gene.

The second major open reading frame is found in the opposite orientation, starting at the ATG initiation codon at position 435 and terminating with a TAA codon at position 70. It codes for a neutral polypeptide of 129 amino

acids extremely rich in phe (17 residues). Using the TARGSEARCH program for RNA polymerase promoter sites (demanding 4 out of 6 matches for both the -35 and -10 conserved sequences) and the "Preceptron" program for translational initiation sites, we found no region on this strand which combines adequate sequences of both types. Nevertheless, we tested whether this open reading frame is being used *in vivo* by cloning the minimal *nusB*$^+$ gene from plasmid pJB8*nusB*$^+$576 in both orientations under the P_L promoter of plasmid pLC28 (18) in strain M72(λ *bio*252*c*I857 H1) (12). Induction of the culture at 42°C in the first orientation produced large amounts of *nusB* protein as judged by its comigration in two-dimensional gel electrophoresis (21) with authentic *nusB* protein synthesized by phage λ *c*I857*nusB*$^+$ after infection of UV-irradiated bacteria (10). No new polypeptide of the expected molecular weight and/or isoelectric point could be identified, however, when the *nusB* sequence was cloned under the P_L promoter in the opposite orientation (data not shown). We conclude that the open reading frame in the opposite orientation is not used *in vivo* in *E. coli*.

The *nusB* protein has been purified after cloning the *nusB*$^+$ gene onto the "runaway" plasmid pMOB45 (19), which resulted in its overproduction. The first eight N-terminal amino acids of the purified *nusB* protein were found to be met-lys-pro-ala-ala-arg-arg-arg, an amino acid sequence identical to that predicted by the DNA sequence (Fig. 2B). Furthermore, the overall amino acid composition of the purified protein agreed well with that predicted by the DNA sequence (data not shown).

DISCUSSION

The isolation of *E. coli* bacterial mutants that block phage λ *N* protein function has provided invaluable information on the processes of bacterial transcription termination and antitermination (reviewed in references 1 and 2). The *nusA* and *nusB* proteins play a key role in both processes. In the presence of the phage λ *N* protein, functional forms of both gene products are essential for the formation of the RNA polymerase antitermination complex (1,2). However, during normal bacterial growth both proteins have been shown to act as auxilliary termination factors at certain sites (13,14). It appears that the *nusA* protein requires the presence of at least the *boxA* sequence to function both as a terminator or antiterminator factor (5,13,14). Although the exact mode of action of the two proteins in these processes has not been elucidated yet, it appears that they functionally interact since some intragenic suppressors of the *nusA*1 mutation map in the *nusB* gene (6). Both

gene products are essential for E. coli viability since mutations exist which result in a temperature sensitive phenotype for bacterial growth (12,29,30).

The proof that the DNA sequence of the nusB gene presented here is the correct one is reflected by the facts (a) that the eight N-terminal amino acids of the purified nusB protein match perfectly those predicted by the DNA sequence, (b) that the overall amino acid composition of the purified nusB protein matches that predicted by the DNA sequence and (c) that TAAGGGG, a sequence similar to the concensus Shine-Dalgarno ribosomal binding domain (26) is found preceding the GTG initiation codon at the expected 7 nucleotide spacing (27). It is not known whether the indicated nusB DNA sequence possesses its own RNA polymerase promoter or is being expressed by readthrough transcription from a plasmid promoter(s). The potential promoter sequences uncovered using the TARGSEARCH program (24) and indicated in the RESULTS section are not very good matches to the concensus -35 and -10 recognition and binding sites of E. coli RNA polymerase (23,25). Since it is not known (a) how much nusB protein is made in these plasmid constructs or (b) what is the minimal amount of intracellular nusB wild type protein necessary to restore the $NusB^+$ phenotype, we cannot be certain about the in vivo use of these potential promoter sequences. The fact that NusB function is expressed when the cloned $nusB^+$ gene is inserted in both orientations argues weakly in favor of a promoter sequence for the nusB gene contained in the 576 bp DNA fragment. Although the minimal nusB gene that we sequenced does not contain an obvious RNA transcription termination site, in the DNA sequence of S. Ishii et al. (31) a rho-independent terminator structure can be found approximately 60 bp past the UGA termination codon. Thus nusB-specific RNA transcripts could conceivably terminate at this site. It is likely then that the nusB gene is the last gene in its own operon; the question whether it is part of a larger operon cannot be conclusively addressed at this time.

ACKNOWLEDGEMENTS

We wish to thank Dr. Sandy Parkinson and Dr. Eric Kofoid for help with the TARGSEARCH and "Perceptron" programs, Dr. Bill Gray for N-terminal sequence analysis and amino acid composition, Ms. Jerri Cohenour for preparation of the manuscript and Dr. S. Ishii and Dr. F. Imamoto for exchange of information before publication. This work was supported by NIH Grant #GM23917.

*Present address: Unite d'Immunoparasitologie, Institut Pasteur, 75724 Paris, Cedex 15, France

+Present address: Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115, USA

REFERENCES

1. Friedman, D.I. and Gottesman, M. (1983) In Lambda II, Hendrix, R. et al., eds. pp 21-51, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. Friedman, D.I., Olson, E.R., Georgopoulos, C., Tilly, K., Herskowitz, I. and Banuett, F. (1984) Microbiol. Reviews, in press.
3. Salstrom, J.S. and Szybalski, W. (1978) J. Mol. Biol. 124, 195-221.
4. Olson, E.R., Flamm, E.L. and Friedman, D.I. (1982) Cell 31, 61-70.
5. Friedman, D.I. and Olson, E.R. (1983) Cell 34, 143-149.
6. Ward, D.F., Delong, A. and Gottesman, M.E. (1983) J. Mol. Biol. 168, 73-85.
7. Keppel, F., Georgopoulos, C.P. and Eisen, H. (1974) Biochimie 56, 1503-1509.
8. Friedman, D.I., Baumann, M.F. and Baron, L.S. (1976) Virology 73, 119-127.
9. Bachmann, B.J. (1983) Microbiol. Reviews 47, 180-230.
10. Swindle, J., Ajioka, J. and Georgopoulos, C. (1981) Mol. Gen. Genet. 182, 409-413.
11. Strauch, M. and Friedman, D.I. (1981) Mol. Gen. Genet. 178, 498-501.
12. Georgopoulos, C.P., Swindle, J., Keppel, F., Ballivet, M., Bisig, R. and Eisen, H. (1980) Mol. Gen. Genet. 178, 55-61.
13. Greenblatt, J., McLimont, M. and Hanly, S. (1981) Nature 292, 215-220.
14. Ward, D.F. and Gottesman, M.E. (1981) Nature 292, 212-215.
15. Kingston, R.E. and Chamberlin, M.J. (1981) Cell 27, 523-531.
16. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and Boyer, H.W. (1977) Gene 2, 95-113.
17. Ish-Horwicz, D. and Burke, J.F. (1981) Nucleic Acids Res. 9, 2989-2998.
18. Remaut, E., Straussens, P. and Fiers, W. (1981) Gene 33, 81-93.
19. Bittner, M. and Vapnek, D. (1981) Gene 15, 319-329.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
21. Maxam, A. and Gilbert, W. (1980) In Methods in Enzymology, Grossman, L. and Moldave, K., eds., 65, 499-560.
22. Tilly, K., Murialdo, H. and Georgopoulos, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1629-1633.
23. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
24. Mulligan, M.E., Hawley, D.K., Entriken, R. and McClure, W.R. (1984) Nucleic Acids Res. in press.
25. Hawley, D.K. and McClure, W.R. (1983) Nucleic Acids Res. 11, 2237-2255.
26. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.
27. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981) Ann. Rev. Microbiol. 35, 365-403.
28. Stormo, G.D., Schneider, T.D., Gold, L. and Ehrenfeucht, A. (1982) Nucleic Acids Res. 10, 2997-3011.
29. Nakamura, Y. (1983) Mol. Gen. Genet. 190, 106-203.
30. Friedman, D.I., Schauer, A.T., Mashni, E.J., Olson, E.R. and Baumann, M.F. (1983) in Microbiology 1983, Schlessinger, D. ed, pp 39-42, American Society for Microbiology, Washington, D.C.
31. Ishii, S., Hatada, E., Maekawa, T. and Imamoto, F. (1984) Nucleic Acids Res. this issue.