Volume 12 Number 12 1984 Nucleic Acids Research

Molecular cloning and nucleotide sequencing of the *nusB* gene of *E. coli*

Shunsuke Ishii, Eriko Hatada[1], Toshio Maekawa[2] and Fumio Imamoto*

Laboratory of Molecular Genetics, (Riken), The Institute of Physical and Chemical Research, Hirosawa, Wako-shi, Saitama 351, [1]Faculty of Science, Kobe University, Nada, Kobe, Hyogo 657, and [2]The Institute of Medical Science, University of Tokyo, PO Takanawa, Tokyo 108, Japan

Received 20 February 1984; Accepted 18 April 1984

ABSTRACT

The nusB gene of E.coli has been cloned in plasmid pBR322. Using genetic complementation as an assay for the gene, its location in subclones was analyzed, and the nucleotide sequence of this gene and its flanking regions was determined. The coding region consists of 417 base pairs (bp), which specify a protein of 139 amino acids, and the calculated molecular weight of the nusB protein is 15,702. The nusB protein contains 20 acidic and 21 basic amino acids. A significant promoter sequence was not found to be located in the region upstream from the translational initiation codon. The possibility that the nusB gene consists of an operon is discussed. After the coding region, there is a G-C rich inverted repeat sequence followed by a run of Ts, which could be a transcriptional terminator of the nusB gene.

INTRODUCTION

The N gene product of coliphage λ regulates phage gene expression by permitting transcription of the phage delayed early and late genes by overcoming transcriptional termination barriers (1). Many E.coli mutants that fail to support the action of N protein have been isolated, and named nus, an abbreviation for N utilization substance (2,3). Recent observations have indicated that the protein products of the nusA and nusB genes are necessary for efficient expression of some bacterial genes as well, including the trp, lac and rpoB-rpoC genes; the proteins are thought to function by modifying the termination of transcription at intragenic transcriptional barriers (4,5).

The nusA gene product, a 69 K dalton protein with an isoelectric point of 4.6, can itself stimulate the synthesis of ß-galactosidase in a coupled in vitro transcription-translation system (6), and can stimulate the "pausing" of RNA polymerase at some transcription termination signals (7-9). The nusB gene product has been reported to be a protein of about 14 K daltons (10,11), but it has not been purified so far. To obtain information on the function of termination and antitermination factors, we cloned and sequenced the nusB gene.

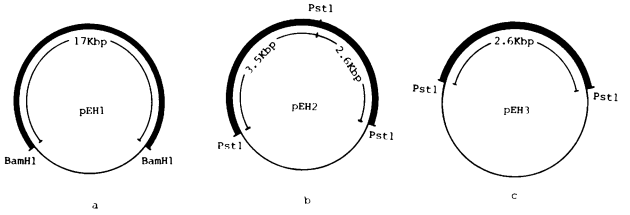

Figure 1. Structures of pEH1, pEH2 and pEH3 plasmid DNA's.

The entire structure of pEH1, pEH2 and pEH3 DNA's are shown. The line indicates the pBR322 vector and the black bar indicates the cloned segment.

MATERIALS AND METHODS

Bacteria, Bacteriophage, Plasmid and Growth Conditions

Strains bearing the *nus*$^+$(K37) and *nusB*(nusB5) alleles (12) and the specialized transducing bacteriophage λ*nusB* (12), used as a source to clone *nusB*, were kindly donated by Dr. D.I.Friedman. The pBR322 plasmid was used as a cloning vehicle (13).

Cells were grown in L-broth (14). Concentrations of 50μg/ml ampicillin and 20μg/ml tetracycline were used.

Enzymes and Chemicals

The restriction enzymes, T4 DNA ligase and T4 polynucleotide kinase were purchased from Takara Shuzo Co., Kyoto, Japan. Bacterial alkaline phosphatase was purchased from Boehringer-Mannheim Co.. (^{32}P)Orthophosphate was purchased from New England Nuclear and (γ-^{32}P)ATP was synthesized as described (15). (α-^{32}P)UTP (410Ci/mmol) was purchased from Amersham Co.. RNA polymerase was purified by the method of Burgess and Jendrisak (16).

Purification of DNA

Plasmid DNA was purified essentially by the procedure of Yamamoto *et al*. (17), by Triton lysis of cells followed by Sepharose 4B column chromatography. For small scale preparation of the plasmid, the rapid isolation method of Holmes and Quigley (18) was used. DNA fragments were extracted from agarose gels by the procedure of Yang *et al*. (19) or from acrylamide gels by the method of Maxam and Gilbert (20).

Restriction Enzyme Analysis

The restriction enzymes used were BamHI, PstI, HindIII, SalI, HaeII , TaqI, Sau3A, HinfI, AluI and HincII . All these enzymes were used under the conditions specified by the suppliers. Slab gels (2mm thick X 13cm) of polyacrylamide (acrylamide/bisacrylamide, 19 : 1) or 1% agarose in 90mM Tris-borate (pH8.3), 4mM EDTA and 0.5μg/ml ethidium bromide were used for deter-

mination of the electrophoretic mobilities of the DNA fragments. The sizes of DNA fragments were calculated from these mobilities relative to those of HinfI fragments of pBR322 DNA.

Determination of Nucleotide Sequence

The nucleotide sequence of DNA fragments were determined by the method of Maxam and Gilbert (20). The cleaved products were fractionated on 5%, 8% and 20% polyacrylamide gels in 8M urea.

RESULTS

Construction of Plasmid DNAs Harboring the NusB Gene

A cell with a defective nusB gene can regain sensitivity to λ phage by transformation with a plasmid carrying the functional nusB gene. The E.coli mutant strain nusB5, lacking functional nusB protein, was transformed with a ligation mixture of BamHI fragments from λnusB and pBR322 DNAs. About 300 ampicillin-resistant transformants were screened and 10 colonies with sensitivity to λ phage were found. The plasmid DNAs were extracted from these colonies by the rapid purification method (18). Restriction analysis showed that these plasmids all contain the same 17Kbp insert, though with different orientations (Fig. 1-a). We designated these plasmids as pEH1 and pEH1', respectively, depending on the orientation of the insert.

The pEH1 DNA was partially digested with PstI and then cloned into pBR322. The plasmid DNAs extracted from the tetracycline resistant and λ-sensitive transformants contained an insert of 6.1Kbp. The plasmid was designated as pEH2 (Fig. 1-b). A cell transformed by pEH2 is not only tetracycline-resistant but also ampicillin - resistant, due to reconstruction of the normal bla gene when the insert of 6.1Kbp, containing a part of the bla gene, is joined to the complementary part of the gene carried by the vector.

The 2.6Kbp DNA fragment prepared by complete digestion of pEH2 DNA with PstI was next purified and cloned into pBR322. The transformants thus obtained were tetracycline-resistant and λ-sensitive. This plasmid was designated as pEH3 (Fig. 1-c).

Location of the NusB Gene

To obtain a plasmid containing the shorter insert, we partially digested the 2.6Kbp DNA fragment with Sau3A and then cloned it into the unique BamHI site of pBR322. Eight of 184 ampicillin-resistant (Ap^r) and tetracycline-sensitive (Tc^s) transformants were sensitive to λ phage ($nusB^+$). These 8 Ap^rTc^s nusB colonies were expected to contain plasmid carrying the wild type

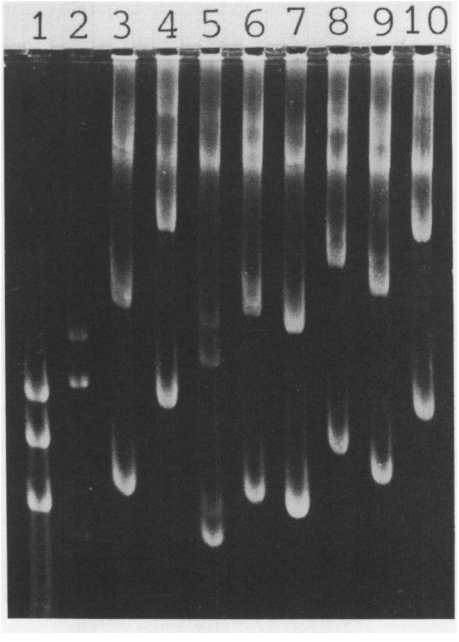

Figure 2. Electrophoretic analysis of nusB$^+$ plasmid DNA's prepared by the rapid boiling method.

Plasmids were prepared by the boiling procedure from 1 ml of bacterial culture (lanes 3-10). Lane 1 - reference linear DNA (PstI digest of pEH2 DNA). The materials in the bands were 4.3, 3.5 and 2.6 kbp, respectively. Lane 2 - the relaxed, linear and super-twisted forms of pBR322 DNA from the top.

nusB gene. The plasmid DNAs were extracted and analyzed by agarose gel electrophoresis as shown in Fig. 2. Three plasmids containing the shorter inserts, designated as pEH4-1, pEH4-2 and pEH4-3 (Fig. 2., lanes 5, 7 and 6), were digested with Sau3A. The resulting cleavage maps are shown in Fig. 3-a. The pEH4-1 clone contains a 572bp insert, pEH4-2 contains a 680bp insert and pEH4-3 contains an 840bp insert in the unique BamHI site of pBR322.

The nusB gene could be transcribed from its own promoter or from the promoter of the tet gene. To determine whether the promoter of the nusB gene is present in the 572bp insert of pEH4-1, we constructed a plasmid containing the 572bp insert of pEH4-1 with the opposite orientation and designated it as pEH4-1'. Transformants containing pEH4-1' were sensitive to λ phage. Thus the nusB gene seems to be expressed by transcription not only from the promoter of the tet gene, but also from either its authentic promoter in the

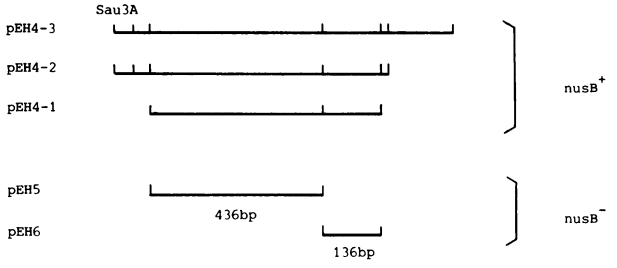

Figure 3. Structures of pEH4-1, pEH4-2, pEH4-3, pEH5 and pEH6 and the Sau3A cleavage sites of the cloned DNA fragments.

Black bars indicate cloned DNA segments. The positions of the Sau3A cleavage sites in the cloned DNA fragments are shown.

572bp insert or another foreign promoter of the vector (see Discussion). Since the molecular weight of the nusB gene product is reported to be about 14K daltons (10, 11), which would be encoded by a region of about 400bp, it was expected that the major portion of the 572bp insert of pEH4-1 must be the coding region of the nusB protein. To verify this, we constructed plasmids carrying 436bp and 136bp Sau3A fragments of the pEH4-1 insert.

The 1,200bp HindIII-SalI fragments of pEH4-1 was digested with Sau3A, and the 436bp and 136bp fragments generated were inserted into the BamHI site of pBR322. The resulting plasmids (pEH5 and pEH6) were analyzed with Sau3A (Fig. 3) and by the complementation test. If the 436bp Sau3A fragment contains the whole coding region of nusB protein, the nusB gene should be transcribed from the tet promoter and its gene product should be synthesized. As nusB5 cells transformed with pEH5 and pEH6 showed the nusB$^-$ phenotype, the 436bp segment is probably not enough to code for the nusB protein, and the coding of the nusB protein requires the other 136bp Sau3A segment. This inference was supported by direct DNA sequencing.

Nucleotide Sequence of the NusB Gene

To determine the nucleotide sequence of the nusB gene, we first digested pEH4-3 with HindIII and SalI to isolate the 1,450bp fragment (see Fig. 3). The cleavage sites of Sau3A, TaqI, HinfI, HaeII , AluI and HincII of this fragment and the strategy for sequence determination are shown in Fig. 4.

The DNA sequence data are given in Fig. 5. The DNA sequence indicates that there are three initiation codons, GTG_1 at position 117, ATG_1 at position 572 and ATG_2 at position 476. The reading frame initiated from GTG_1 is in the opposite direction of those initiated from ATG_1 and ATG_2. No other

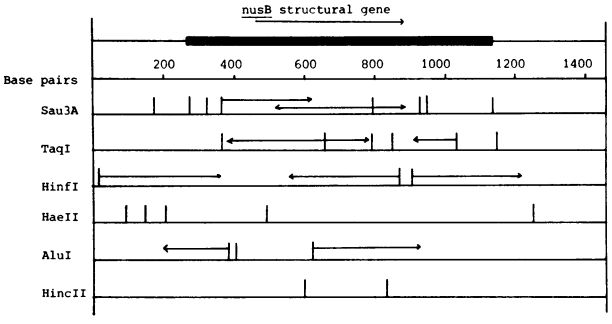

Figure 4. Restriction map of the nusB gene and sequence strategy.

The black bar indicates the cloned segment in pEH4-3. The 1,450 bp HindIII-SalI fragment was isolated from pEH4-3. The positions of restriction endonuclease sites are based on analyses of the sizes of DNA fragments from singly- or doubly- digested 1,450 bp HindIII-SalI DNA fragment. The arrows show the direction and extent of individual sequencing.

reading frame with both strands could code for proteins of more than 100 amino acids. The nusB protein was purified and the amino acid sequence of the eleven residues from its NH_2 terminus was determined to be NH_2-Met-Lys-Pro-Ala-Ala-Arg-Arg-Arg-Ala-Arg-Glu (to be published), which is consistent with the sequence predicted from the nucleotide sequence of the reading frame initiated from GTG_1. So we conclude that the reading frame initiated from GTG_1 can code for the nusB protein. The amino acid composition predicted from the DNA sequence is consistent with the composition determined by amino acid analysis of purified nusB protein (to be published). The structural gene of the nusB protein consists of 417 nucleotides, which corresponds to 139 amino acids (molecular weight, 15,702).

DISCUSSION

Initiation and Termination of Transcription

By restriction analyses of constructed plasmids, we localized the nusB gene in the 572bp region between the Sau3A sites of chromosomal DNA. Judging from the DNA sequence of this region, the structural gene of nusB protein starts about 110bp after the Sau3A site. pEH4-1', in which the nusB gene has the opposite orientation of transcription to that of the tet promoter, gave the $nusB^+$ phenotype. Thus the 572bp insert of pEH4-1' was expected to contain the nusB promoter. However, no RNA species were synthesized by purified RNA polymerase from the Sau3A 436bp DNA fragment in vitro (data not shown). The known promoter regions of E.coli RNA polymerase have certain structural

```
GATCGAACGTGCTGGCACCAAAGCTGGCAACAAAGGTGCAGAAGCTGCAC  1-50
CTAGCTTGCACGACCGTGGTTTCGACCGTTGTTTCCACGTCTTCGACGTG

TGACCGCGCTTGAAATGATTAATGTATTGAAAGCCATCAAGGCCTGAAAT  51-100
ACTGGCGCGAACTTTACTAATTACATAACTTTCGGTAGTTCCGGACTTTA

TAGTAAGGGGAAATCCGTGAAACCTGCTGCTCGTCGCCGCGCTCGTGAGT  101-150
ATCATTCCCCTTTAGGCACTTTGGACGACGAGCAGCGGCGCGAGCACTCA
                fMetLysProAlaAlaArgArgArgAlaArgGluC  1-11

GTGCCGTCCAGGCGCTCTACTCCTGGCAGTTGTCCCAGAACGACATCGCT  151-200
CACGGCAGGTCCGCGAGATGAGGACCGTCAACAGGGTCTTGCTGTAGCGA
ysAlaValGlnAlaLeuTyrSerTrpGlnLeuSerGlnAsnAspIleAla  12-28

GATGTTGAATACCAGTTCCTGGCTGAACAGGATGTAAAAGACGTTGACGT  201-250
CTACAACTTATGGTCAAGGACCGACTTGTCCTACATTTTCTGCAACTGCA
AspValGluTyrGlnPheLeuAlaGluGlnAspValLysAspValAspVa  29-44

CCTGTACTTCCGTGAGCTGCTGGCCGGGGTGGCGACTAATACCGCATACC  251-300
GGACATGAAGGCACTCGACGACCGGCCCCACCGCTGATTATGGCGTATGG
lLeuTyrPheArgGluLeuLeuAlaGlyValAlaThrAsnThrAlaTyrL  45-61

TCGACGGACTGATGAAGCCATACCTGTCCCGCCTGCTGGAAGAACTGGGA  301-350
AGCTGCCTGACTACTTCGGTATGGACAGGGCGGACGACCTTCTTGACCCT
euAspGlyLeuMetLysProTyrLeuSerArgLeuLeuGluGluLeuGly  62-77

CAGGTAGAAAAAGCAGTACTGCGCATTGCGCTGTACGAACTGTCTAAACG  351-400
GTCCATCTTTTTCGTCATGACGCGTAACGCGACATGCTTGACAGATTTGC
GlnValGluLysAlaValLeuArgIleAlaLeuTyrGluLeuSerLysAr  78-95

TAGCGATGTGCCATACAAAGTGGCCATTAACGAAGCGATCGAACTGGCGA  401-450
ATCGCTACACGGTATGTTTCACCGGTAATTGCTTCGCTAGCTTGACCGCT
gSerAspValProTyrLysValAlaIleAsnGluAlaIleGluLeuAlaL  96-111

AATCGTTCGGCGCAGAAGACAGCCATAAGTTCGTCAACGGCGTACTCGAT  451-500
TTAGCAAGCCGCGTCTTCTGTCGGTATTCAAGCAGTTGCCGCATGAGCTA
ysSerPheGlyAlaGluAspSerHisLysPheValAsnGlyValLeuAsp  112-128

AAAGCAGCACCTGTGATTCGCCCTAACAAAAAGTGATATCCAGGCCGGTA  501-550
TTTCGTCGTGGACACTAAGCGGGATTGTTTTTCACTATAGGTCCGGCCAT
LysAlaAlaProValIleArgProAsnLysLys                    129-139

GATTCACGGAAGACCGTTCCATGATCTGGCAGCGATGGGGGCCGATCGCC  551-600
CTAAGTGCCTTCTGGCAAGGTACTAGACCGTCGCTACCCCCGGCTAGCGG

GGCCTTTTCTTTTTACCTGCTGAGGCATAACGTATGGCATGTGGCGAGTT  601-650
CCGGAAAAGAAAAATGGACGACTCCGTATTGCATACCGTACACCGCTCAA

CTCCCTGATTGCCCGTTATTTTGACCGTGTAAG                    651-683
GAGGGACTAACGGGCAATAAAACTGGCACATTC
```

Figure 5. Nucleotide sequence of the nusB gene and its flanking region and the primary structure of the nusB protein.

The nucleotide sequence of the coding strand of DNA from the 5' to 3' end is shown in the upper line and numbered at the right, number one being the Sau3A site of the insert of pEH4-3. Amino acids are numbered at the right starting from N-terminal methionine.

homologies, particularly in two regions about 35 and 10 nucleotides 5'-proximal to the site where transcription begins. Probably, the former provides a site recognized by RNA polymerase, while the latter is the RNA polymerase binding site (Pribnow sequence)(21). In the 110bp region between the Sau3A site and the initiation codon of the nusB protein, there is only one Pribnow-like sequence, T-A-A-T-G-T-A, about 40bp upstream from the initiation codon. This sequence is, however, not as close to the prototype

sequence as the other promoters reported before (22, 23). Among the 20 nucleotides further upstream, there is no sequence that includes the highly commom T-T-G and surrounding sequence existing in the recognition region. It is noteworthy that in this region there are one inverted repeat sequence and one direct repeat sequence (shown by horizontal arrows in Fig. 5). Since a transcriptional regulatory protein factor such as CRP, is known to bind to DNA template at the inverted repeat sequence, efficient transcription of the nusB gene in vitro might need a positive regulatory factor such as CRP. However, it is possible that another strong promoter exists further upstream in the DNA and that the nusB gene is contained in an operon. In fact, in the region upstream from the nusB structural gene, there is an open reading frame which leads to the synthesis of a protein consisting of at least 50 amino acids. Detailed analysis of the 5'-noncoding region flanking the nusB structural gene is underway and results will be presented elsewhere.

About 60bp downstream from the translational termination codon there is a G-C rich inverted repeat sequence followed by a run of Ts. This structure is a typical ρ-independent transcriptional termination signal (24). The transcription presumably terminates somewhere in the T cluster.

Translation and NusB Protein

Approximately 110 nucleotides downstream from the Sau3A site, a GUG codon begins the structural portion of the nusB gene. About 10 nucleotides upstream from the initiation codon, there is a "Shine-Dalgarno" sequence, T-A-A-G-G-G, which is complementary to a 3'-terminal sequence of E.coli 16S rRNA, and could serve as a ribosome binding site (25). Starting from this GUG, a protein of 139 amino acids (molecular weight, 15,702) would be encoded. This protein would contain 21 basic amino acids (9 Arg and 12 Lys) and 20 acidic amino acids (9 Asp and 11 Glu), and thus would be a neutral protein. Consistent with this inference, maxicells containing pEH4-1 directed the synthesis of a 15K dalton protein with an isoelectric point of about 7.3, shown by two-dimensional gel electrophoresis (to be published).

Ikemura and Ozeki found that the frequency of use of optimal codons is strictly related to the production levels of individual genes in E.coli (26): In the case of highly expressed genes (e.g., tufA and rplA), the optimal codons occupy almost the entire gene, whereas in the case of moderately expressed genes (e.g., trpC and trpA) or weakly expressed genes (e.g., trpR), the occurrence of the optimal codon is less. In collaboration with Dr. Ikemura, we found that the frequency of use of optimal codons of the nusB gene (Fop: 0.780) was very similar to those of the rpoD (Fop: 0.810) and nusA

(Fop: 0.793) genes. It, therefore, seems that the number of nusB protein molecules in a cell is comparable with the numbers of sigma and nusA proteins, which have been estimated to be about 8 X 10^2 molecules per cell (26).

*To whom correspondence should be addressed

REFERENCES

1. Roberts, J.W. (1969) Nature 224, 1168-1174
2. Friedman, D.I. (1971) in The Bacteriophage Lamda, ed., A.A.Hershey, Cold Sprig Harbor, New York, 733-738
3. Friedman, D.I., Bauman, M. and Baron, L.S. (1976) Virology 73, 119-127
4. Kung, H.F., Spears, C. and Weissbach, H. (1975) J. Biol. Chem. 250, 1556-1562
5. Kuroki, K., Ishii, S., Kano, Y., Miyashita, T., Nishi, K. and Imamoto, F. (1982) Molec. Gen. Genet. 185, 369-371
6. Greenblatt, J., Li, J., Adhya, S., Friedman, D.I., Baron, L.S., Redfield, B., Kung, H.F. and Weissbach, H. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1991-1994
7. Kassavetis, G. and Chamberlin, M. (1981) J. Biol. Chem. 256, 2777-2786
8. Greenblatt, J., McLimont, M. and Hanly, S. (1981) Nature 292, 215-220
9. Farnham, P.J., Greenblatt, J. and Platt, T. (1982) Cell 29, 945-951
10. Swindle, J., Ajioka, J. and Georgopoulos, C. (1981) Molec. Gen. Genet. 182, 409-413
11. Strauch, M. and Friedman, D.I. (1981) Molec. Gen. Genet. 182, 498-501
12. Friedman, D.I., Bauman, M. and Baron, L.S. (1976) Virology 73, 119-127
13. Boliver, F., Rosriguenz, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and Boyer, H.W. (1977) Gene 2, 95-113
14. Lennox, E.S. (1955) Virology 1, 190-206
15. Johnson, R.A. and Walseth, T.F. (1979) Advances in Cyclic Nucleotide Research 10, 135-167
16. Burgess, R.R. and Jendrisak, J.J. (1975) Biochemistry 14, 4634-4638
17. Yamamoto, T., Sobel, M.E., Adams. S.L., Avvedimento, V.E., DiLauro, R., Pastan, I., de Crombrugghe, B., Showalter, A., Pesciotta, D., Fietzek, P. and Olsen, B. (1980) J. Biol. Chem. 255, 2612-2615
18. Holmes, D.S. and Quigley, M. (1981) Anal. Biolchem. 114, 193-197
19. Yang, R.C.A., Lis, J. and Wu, R. (1979) in Methods in Enzymology, Wu, R., ed., 68, 176-182
20. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. eds. 65, 499-560
21. Gilbert, W. (1976) in RNA polymerase, Losick, R. and Chamberlin, M., eds., Cold Spring Harbor, N.Y., 193-205
22. Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 784-788
23. Schaller, H., Gray, C. and Herrmann, K. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 737-741
24. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353
25. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346
26. Ikemura, T. and Ozeki, H. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 1087-1097