Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 409–413, January 1984
Biochemistry

$tRNA_{f2}^{Met}$ gene in the leader region of the *nusA* operon in *Escherichia coli*

(antitermination/DNA sequence)

SHUNSUKE ISHII, KAZUYUKI KUROKI, AND FUMIO IMAMOTO

Laboratory of Molecular Genetics, Riken (The Institute of Physical and Chemical Research), Hirosawa, Wako-shi, Saitama, 351, Japan

Communicated by Masayasu Nomura, October 3, 1983

ABSTRACT **The promoter-proximal portion of the operon containing the *Escherichia coli nusA* gene has been cloned. Its nucleotide sequence shows that genes for $tRNA_{f2}^{Met}$ and a 15-kilodalton protein of unknown function precede the nusA protein gene. The sequence suggests that the three genes form a single transcription unit. Consistent with this hypothesis, purified RNA polymerase formed full-length transcripts on the cloned DNA *in vitro*, although transcription was frequently arrested at the intercistronic site(s) between the gene for $tRNA_{f2}^{Met}$ and the 15-kilodalton protein.**

The *N* gene product of coliphage λ permits transcription of the phage-delayed early and late genes by overcoming transcriptional termination barriers (1). Many *Escherichia coli* mutants that fail to support the action of N protein and λ growth have been isolated, and they are named *nus*, standing for N utilization substance (2, 3). The protein products of the genes, nusA and nusB, the former of which has been shown to be the same as L factor (4), are required for efficient transcription of not only λ genes but also *E. coli* genes including the *trp* and *lac* operons (5, 6).

The nusA gene product, a 69-kilodalton (kDa) protein with an isoelectric point of 4.6, binds to N protein (7, 8) or to *E. coli* core RNA polymerase but not to the RNA holopolymerase (9). In a coupled *in vitro* transcription–translation system, it stimulates the synthesis of β-galactosidase (6). It also acts as a factor that causes RNA polymerase to pause at some transcriptional termination signals (10–12).

nus functions may interact with ribosomes in *E. coli*: another mutant, *nusE*, alters the ribosomal protein S10 (13). It seemed possible that the clues to interaction of transcription and translation might also be found in genomic organization of other *nus* genes. We therefore cloned the *nusA* gene region and found two genes, for $tRNA_{f2}^{Met}$ and a 15-kDa protein, that precede the *nusA* gene in a common transcriptional unit.

MATERIALS AND METHODS

Strains and General Methods. The strains of *E. coli*, bacteriophage λ, and plasmids used are listed in Table 1. The methods used for cloning the genes have been described (16). Plasmid DNA was purified as described (17). Slab gels (2 mm thick × 13 cm) of 5% or 3% polyacrylamide or 1% agarose in 90 mM Tris borate, pH 8.3/4 mM EDTA containing ethidium bromide at 0.5 μg/ml were used for determination of the electrophoretic mobilities of the DNA fragment and the RNA transcript. The sizes of the DNA fragments and the RNA transcripts were estimated from their mobilities relative to that of the *Hin*dIII fragments of λ phage DNA or the *Hin*fI fragments of pBR322 DNA. DNA fragments were extracted from agarose or acrylamide gels as described by Yang *et al.* (18) or Maxam and Gilbert (19).

Enzymes and Chemicals. All enzymes were used under the conditions specified by the suppliers. The restriction enzymes, T4 DNA ligase, and T4 polynucleotide kinase were purchased from Takara Shuzo (Kyoto, Japan). Bacterial alkaline phosphatase and RNase T1 were purchased from Boehringer Mannheim. RNase A and RNase U2 were obtained from Sigma. RNase Phy I was supplied by P-L Biochemicals. [^{35}S]Methionine (400 Ci/mmol; 1 Ci = 37 GBq) and $^{32}P_i$ were from New England Nuclear, [α-^{32}P]UTP (410 Ci/mmol) was from Amersham, and [γ-^{32}P]ATP and [γ-^{32}P]GTP (5,000–7,000 Ci/mmol) were synthesized as described by Johnson and Walseth (20). *E. coli* RNA polymerase was prepared by the method of Burgess and Jendrisak (21). NusA protein was purified as reported by Greenblatt *et al.* (11).

Table 1. Strains of *E. coli*, bacteriophage λ, and plasmid used

Strain or plasmid	Relevant genotype	Source or reference
E. coli		
K37	*rpsL,galK2*	2
K95	*nusA1,rpsL,galK2*	2
N17-9 recAam	*trp,rpsL,recA99,uvrA*	H. Ogawa
Bacteriophage		
λ*cI*90		Our laboratory
Plasmid		
pBR322	$Am^R Tc^R$	14
pLC34-4	imm^{colE1}	15
pKU1	*nusA*-expressing fragment of pLC34-4 cloned in pBR322	This work

Determination of Nucleotide Sequences. The nucleotide sequences of DNA fragments were determined as described by Maxam and Gilbert (19). The cleaved products were fractionated on 5% and 20% polyacrylamide gels in 8 M urea. The nucleotide sequence near the 5′ end of RNA was determined by analysis of the products of partial digestion of [γ-^{32}P]GTP-labeled RNA with RNase A, RNase T1, RNase U2, and RNase Phy I essentially as described by Simoncsitz *et al.* (22).

Determination of the NH_2-Terminal Amino Acid Sequence. A manual Edman degradation procedure (23) was used to determine the NH_2-terminal sequence of the nusA protein with slight modification as described by Hase *et al.* (24).

Transcription of DNA Fragment. The 1.9-kilobase-pair (kbp) *Pst* I DNA fragment from pKU1 was transcribed with *E. coli* RNA polymerase. The reaction mixture (200 μl) was 40 mM Tris·HCl, pH 7.8/10 mM $MgCl_2$/0.1 mM dithiothreitol/10% (vol/vol) glycerol/0.04 mM ATP/CTP/GTP/UTP including either [α-^{32}P]UTP (0.6 Ci/mmol), [γ-^{32}P]ATP (31 Ci/mmol), or [γ-^{32}P]GTP (31 Ci/mmol), 1 μg of DNA fragment, and 2 μg of *E. coli* RNA holopolymerase. After incubation of the mixture for 30 min at 37°C, an equal volume of 0.6 M sodium acetate, pH 5.0/1% $NaDodSO_4$ containing 100 μg of *E. coli* tRNA per ml was added. Samples were then extracted with phenol, precipitated with ethanol, and analyzed by gel electrophoresis (in Tris borate/EDTA) on 3%

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kDa, kilodalton(s); kbp, kilobase pair(s).

polyacrylamide gels containing 7 M urea as described (25).

Identification of Plasmid-Encoded Proteins. Plasmid-encoded proteins were located on 12% polyacrylamide/$NaDodSO_4$ gels (26) by the "maxicell" technique (27) with a *uvrA recA phr*$^+$ cell as the plasmid host and [^{35}S]methionine (400 Ci/mmol) at a concentration of 10 μCi/ml. Gels were dried and exposed to XAR-5 x-ray film (Kodak) at −80°C.

RESULTS

Cloning of the *nusA* Gene. The *nusA* gene is near the *argG* gene at 68 min on the *E. coli* chromosome (28). The pLC34-4 plasmid has been reported to carry the *argG* gene (15) and may bear the *nusA* gene. To check this possibility, an *E. coli* mutant strain, nusA1, lacking functional nusA protein was transformed with plasmid pLC34-4. As a transformant was restored to the nusA$^+$ phenotype (the ability to grow λ phage), we concluded that plasmid pLC34-4 carries the functional *nusA* gene. To obtain a plasmid containing the shorter insert, strain nusA1 was transformed with a ligation mixture of *Pst* I fragments from pLC34-4 and pBR322 DNAs. Of about 200 tetracycline-resistant transformants screened, five colonies allowed growth of λ*cI*90 phage at 42°C. The plasmid DNAs extracted from these transformants were analyzed. Analysis of the DNA with restriction enzymes showed that all of these plasmids contain the same DNA insert of about 1.9 kbp, although both orientations are present (Fig. 1). We designated these two types of plasmids pKU1 and pKU1′.

Nucleotide Sequence of the 5′ Upstream Region of the *nusA* Gene. We determined the sequence of the inserted DNA of the 1.9-kbp *Pst* I fragment from pKU1. The sequence analysis strategy is shown in Fig. 2. The partial sequence (Fig. 3) revealed a structural gene for $tRNA_{f2}^{Met}$ (nucleotides 9–85) followed by two open reading frames [nucleotides 320–739 and 770–1801 (the end of the insert of pKU1)].

The second open reading frame, starting at nucleotide 770, permits translation of the nusA protein: the amino acid sequence of five residues from the NH_2 terminus of the nusA protein, which was purified from *E. coli* strain K37, was determined to be NH_2-Met-Asn-Lys-Glu-Ile (data not shown), which is consistent with the sequence predicted from the nucleotide sequence shown in Fig. 3. As more recent sequence data show that pKU1 contains about 70% of the *nusA* coding region (the complete nucleotide sequence of the *nusA* gene will be presented elsewhere), the polypeptide containing the 70% portion from the NH_2 terminus of the nusA protein seemed to complement the nusA1 mutation.

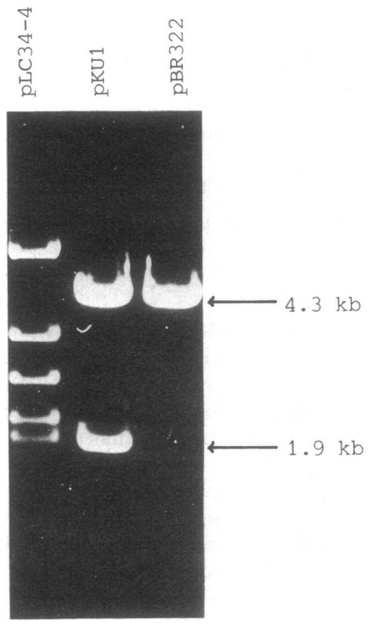

FIG. 1. Electrophoretic analysis of pKU1 and pLC34-4 plasmid DNAs cleaved by *Pst* I. pLC34-4, pKU1, and pBR322 DNAs (1.0 μg each) were digested with *Pst* I and analyzed by electrophoresis on a horizontal 1.0% agarose gel containing ethidium bromide at 0.5 μg/ml.

The first open reading frame (nucleotides 320–739) can encode a protein with a molecular weight calculated as 15,471. The function of this protein is unknown.

Site of Transcriptional Initiation. Among the 840 nucleotides preceding the *nusA* gene, only one nucleotide sequence (nucleotides −36 to −7) fits the canonical −10 region Pribnow sequence (T-A-T-R-A-T-R; R, purine nucleoside) (29) or the −35 region sequence (T-T-G-A-C-A) (30, 31). Thus, it seemed probable that the genes for $tRNA_{f2}^{Met}$, the 15-kDa protein, and the nusA protein form a single transcriptional unit.

To show that only one promoter exists at the nucleotide

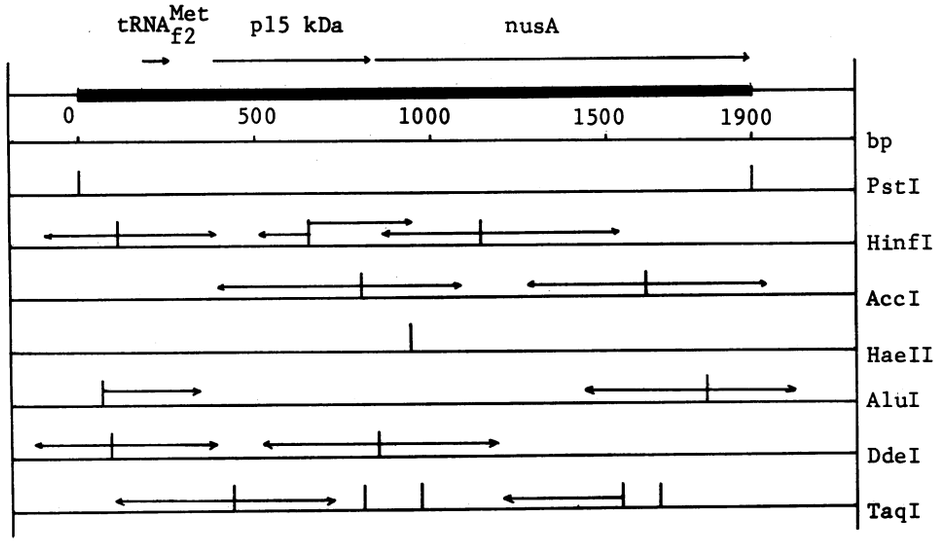

FIG. 2. Restriction map and sequence analysis strategy for the 1.9-kbp insert of pKU1. The black bar indicates the cloned segment in pKU1. The positions of the restriction endonuclease sites are based on analysis of the sizes of DNA fragments from singly or doubly digested 1.9-kbp *Pst* I DNA fragment. The arrows indicate the direction and extent of individual analyses.

FIG. 3. Nucleotide sequence of the 5'-upstream region of the *nusA* gene. The DNA sequence is numbered beginning at the predominant site of initiation of transcription *in vitro*. The Pribnow, −35 recognition, and Shine–Dalgarno sequences are boxed. The $tRNA_{f2}^{Met}$ gene is underlined. The dyad symmetries are indicated by horizontal arrows, and the thymidine-rich sequences preceded by the dyad symmetries are indicated by dotted lines. The predicted amino acid sequence is shown below the DNA sequence.

−36 to −7 site of the 1.9-kbp *Pst* I fragment, we transcribed RNA from the purified DNA fragment with purified RNA polymerase and [α-^{32}P]UTP *in vitro*. On electrophoretic analysis of the transcripts, RNA species of about 240 and 1,800 nucleotides were found to be the predominant products (Fig. 4). Because these RNA species could be labeled with [γ-^{32}P]GTP but not with [γ-^{32}P]ATP (data not shown), transcription of each apparently started with GTP. Production of the RNA species of 240 nucleotides might result from termination of transcription at the thymidine clusters located at nucleotides 226–234 or 267–273, which is preceded by a possible palindrome structure of 16 or 13 base pairs of DNA, respectively (nucleotides 210–225, 254–266). In contrast, the RNA species of 1,800 nucleotides is almost as long as the entire template DNA fragment.

To show that the short and long RNA species were transcribed from the same promoter, we analyzed the nucleotide sequences at the 5' ends of each RNA. This sequence was deduced from the products of partial digestion of [γ-^{32}P]GTP-labeled RNAs by various RNases. The results showed that the 5' sequences of two labeled RNA species were both pppG-U-U-U-C-A (data not shown)—i.e., at +1 in the sequence shown in Fig. 3. We also analyzed the RNA products synthesized *in vitro* from the restricted DNA. When the 1.9-kbp *Pst* I DNA fragment from pKU1 was restricted by *Taq* I and used as a template for *in vitro* transcription, only two RNA species—of 240 and 370 nucleotides—were detected. We therefore conclude that genes for $tRNA_{f2}^{Met}$, the 15-kDa protein, and the nusA protein are transcribed as a single transcriptional unit ("operon").

Expression of the *nusA* Operon in "Maxicells." For detection of synthesis of the protein products of this operon, plasmid pKU1 was transformed into maxicell strain N17-9 recAam. After ultraviolet irradiation, the proteins still synthesized in maxicells carrying pKU1 were labeled with [^{35}S]methionine and the bulk protein from labeled cells was analyzed by polyacrylamide gel electrophoresis. Maxicells containing pKU1 directed synthesis of at least four molecular species of protein (Fig. 5). The protein species unique to pKU1 were of 55, 21, and 18 kDa. These three protein spe-

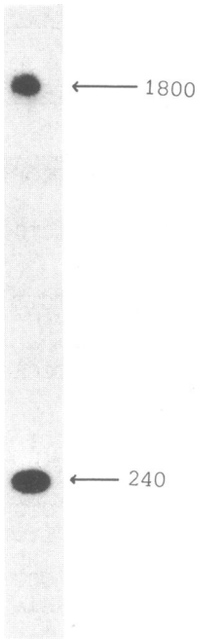

FIG. 4. Autoradiograph of transcripts synthesized from the 1.9-kbp *Pst* I fragment by RNA polymerase. Transcripts were labeled with [α-^{32}P]UTP in the presence of all four NTPs at 0.04 mM each. Electrophoresis in 3% polyacrylamide/urea gels was carried out at 200 V for 15 hr.

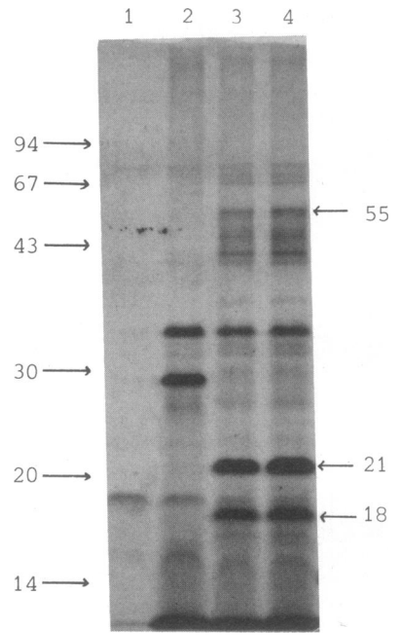

FIG. 5. Radioactive proteins synthesized in a maxicell harboring plasmid pKU1 or plasmid pBR322. Lanes: 1, no plasmid; 2, pBR322; 3 and 4, pKU1. Standards included the following: phosphorylase b (molecular weight, 94,000), bovine serum albumin (67,000), ovalbumin (43,000), carbonic anhydrase (30,000), soybean trypsin inhibitor (20,100), and α-lactalbumin (14,400), shown ($\times 10^{-3}$) on the left. The protein species of 55, 21, and 18 kDa unique to pKU1 are indicated ($\times 10^{-3}$) on the right.

cies probably represent a fragment of the nusA protein, a fragment of β-lactamase, and the 15-kDa protein, respectively. In fact, the 55-kDa protein was shown to be precipitated with a specific rabbit antiserum against nusA protein (A. Ishihama, personal communication).

DISCUSSION

Two sites for the structural genes of $tRNA_f^{Met}$ in the *E. coli* chromosome have been reported: *metZ* is mapped at 61 min and *metY*, at 69 min (32). $tRNA_{f2}^{Met}$ is a minor fraction of initiator $tRNA_f^{Met}$ molecules and is encoded by *metY* (33). The primary structure of this tRNA differs from that of the major species ($tRNA_{f1}^{Met}$) at one base of the RNA chain (the 47th base from the 5′ end) (34, 35). Data on the nucleotide sequence show that the structural gene of $tRNA_f^{Met}$ existing upstream of the *nusA* operon is *metY*.

The open reading frame (nucleotides 320–739) preceding the *nusA* gene encodes a 15-kDa protein. This protein contains 22 acidic amino acids and 12 basic amino acids and corresponds to no known ribosomal protein (36) or transcription factor (37). It is not clear whether this 15-kDa protein is involved in the antitermination mechanism.

The sequence T-A-G-A-A-T-T located about 10 nucleotides upstream from the site of transcriptional initiation of the *nusA* operon is similar to the consensus sequence of a Pribnow box (30, 31). This sequence resembles the sequence T-A-T-A-A-T-T of the promoter of the *E. coli recA* gene (38) and also of the c17 promoter of phage λ (39). The sequence including T-T-G, found about 25 nucleotides further upstream, may function as the initial recognition signal for RNA polymerase.

The sequence G-A-G-G-T-G-A (nucleotides 753–759) located upstream from the translational initiation site of the *nusA* gene is an excellent Shine–Dalgarno ribosome-binding site (40) and could result in the production of a high level of the nusA protein in *E. coli*. On the other hand, the 15-kDa protein gene does not appear to be preceded by a definite Shine–Dalgarno sequence. The sequence T-A-G (nucleotides 312–314), which precedes the ATG initiation codon by five nucleotides, could serve this function, although its efficiency for ribosome binding may be low.

Preceding the genes for the 15-kDa protein, there are two inverted sequences (nucleotides 210–225 and 254–266) followed in each case by a run of thymidines. This is a typical ρ-independent transcriptional termination signal (41). In fact, about 90% of the transcripts are terminated at these sites *in vitro* (Fig. 4). Since the nusA protein enhances transcription termination at the ρ-independent termination signal (12), the rate of production of nusA protein may be regulated autogenously by decreasing read-through transcription at these two sites.

Plumbridge *et al.* (42) reported that the *infB* gene encoding initiation factor 2 is located at 68 min on the *E. coli* chromosome, which is close to *argG*, *nusA*, *rpsO* (the gene for ribosomal protein S15), and *pnp* (the structural gene for the large subunit of polynucleotide phosphorylase). Two other research groups (ref. 43; T. Kurihara and Y. Nakamura, personal communication) have also reported cloning of the *nusA* region and results of analysis of protein products encoded by the genomic region. From the data obtained by these three groups, the order of the genes is *argG-nusA-infB-rpsO-pnp*. It is possible that the *nusA* operon may include some of these genes in addition to the genes for $tRNA_{f2}^{Met}$ and the 15-kDa protein.

We express appreciation to Drs. H. Matsubara, T. Wada, and S. Wakabayashi (Osaka University) for advice and use of the facilities in determining the NH_2-terminal sequence of the nusA protein.

1. Roberts, J. W. (1969) *Nature (London)* **224**, 1168–1174.
2. Friedman, D. I. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 733–738.
3. Friedman, D. I., Bauman, M. & Baron, L. S. (1976) *Virology* **73**, 119–127.

4. Greenblatt, J., Li, J., Adhya, S., Friedman, D. I., Baron, L. S., Redfield, B., Kung, H. F. & Weissbach, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1991–1994.
5. Kuroki, K., Ishii, S., Kano, Y., Miyashita, T., Nishi, K. & Imamoto, F. (1982) *Mol. Gen. Genet.* **185**, 369–371.
6. Kung, H. F., Spears, C. & Weissbach, H. (1975) *J. Biol. Chem.* **250**, 1556–1562.
7. Greenblatt, J. & Li, J. (1981) *J. Mol. Biol.* **147**, 11–23.
8. Ishii, S., Kuroki, K., Sugino, Y. & Imamoto, F. (1980) *Gene* **10**, 291–300.
9. Greenblatt, J. & Li, J. (1981) *Cell* **24**, 421–428.
10. Kassavetis, G. & Chamberlin, M. (1980) *J. Biol. Chem.* **256**, 2777–2786.
11. Greenblatt, J., Mclimont, M. & Hanly, S. (1981) *Nature (London)* **292**, 215–220.
12. Farnham, P. J., Greenblatt, J. & Platt, T. (1982) *Cell* **29**, 945–951.
13. Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S. & Adhya, S. L. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1115–1119.
14. Boliver, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) *Gene* **2**, 95–113.
15. Takeda, Y., Nishimura, A., Nishimura, Y., Yamada, M., Yasuda, S., Suzuki, H. & Hirota, Y. (1981) *Plasmid* **6**, 86–98.
16. Ishii, S., Hatada, E., Kuroki, K., Kano, Y. & Imamoto, F. (1983) *Microbiology*, 60–62.
17. Yamamoto, T., Sobel, M. E., Adams, S. L., Avvedimonto, V. E., Dilauro, R., Pastan, I., de Crombrugghe, B., Showalter, A., Pesciotta, D., Fietzek, P. & Olsen, B. (1980) *J. Biol. Chem.* **255**, 2612–2615.
18. Yang, R. C. A., Lis, J. & Wu, R. (1979) *Methods Enzymol.* **68**, 176–182.
19. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
20. Johnson, R. A. & Walseth, T. F. (1979) *Adv. Cyclic Nucleotide Res.* **10**, 135–167.
21. Burgess, R. R. & Jendrisak, J. J. (1975) *Biochemistry* **14**, 4634–4642.
22. Simoncsitz, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. & Guilley, H. (1977) *Nature (London)* **269**, 833–836.
23. Blombäck, A., Blombäck, M., Edman, P. & Hessel, B. (1966) *Biochim. Biophys. Acta* **115**, 371–396.
24. Hase, T., Matsubara, H., Hutber, G. N. & Rogers, L. J. (1982) *J. Biochem. (Tokyo)* **92**, 1347–1355.
25. Imamoto, F., Ishii, S., Maekawa, T., Kuroki, K., Kano, Y. & Miyashita, T. (1983) *Microbiology*, 57–59.
26. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
27. Sancar, A., Hack, A. M. & Rupp, W. D. (1979) *J. Bacteriol.* **137**, 692–693.
28. Friedman, D. I. & Baron, L. S. (1974) *Virology* **58**, 141–148.
29. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 784–788.
30. Pribnow, D. (1975) *J. Mol. Biol.* **99**, 419–443.
31. Takanami, M., Sugimoto, K., Sugisaki, H. & Okamoto, T. (1976) *Nature (London)* **260**, 297–302.
32. Ikemura, T. & Ozeki, H. (1977) *J. Mol. Biol.* **117**, 419–446.
33. Ikemura, T., Shimura, Y., Sakano, H. & Ozeki, H. (1975) *J. Mol. Biol.* **96**, 69–86.
34. Dube, S. K., Marcker, K. A., Clark, B. F. C. & Cory, S. (1969) *Eur. J. Biochem.* **8**, 244–255.
35. Dube, S. K. & Marcker, K. A. (1969) *Eur. J. Biochem.* **8**, 256–262.
36. Kurland, C. G. (1977) *Annu. Rev. Biochem.* **46**, 173–200.
37. Ishihama, A., Kajitani, M., Enami, M., Nagasawa, H. & Fukuda, R. (1983) *Microbiology*, 4–6.
38. Horii, T., Ogawa, T. & Ogawa, H. (1979) *Proc. Natl. Acad. Sci. USA* **77**, 313–317.
39. Rosenberg, M., Court, D., Shimatake, H., Brasy, C. & Wulff, D. L. (1978) *Nature (London)* **272**, 414–422.
40. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
41. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
42. Plumbridge, J. A., Howe, J. G., Springer, M., Touati-Schwartz, D., Hershey, J. W. B. & Grunberg-Manago, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5033–5037.
43. Holowachuk, E. W. & Friesen, J. D. (1982) *Mol. Gen. Genet.* **187**, 248–253.