Proc. Natl. Acad. Sci. USA
Vol. 76, No. 4, pp. 1697–1701, April 1979
Biochemistry

Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit β in *Escherichia coli*

(protein sequence/transcription initiation/transcription termination/codon usage/isoaccepting tRNAs)

LEONARD E. POST*, GENEVA D. STRYCHARZ*, MASAYASU NOMURA*†, H. LEWIS‡, AND PATRICK P. DENNIS‡

*Institute for Enzyme Research and Departments of Genetics and Biochemistry, University of Wisconsin, Madison, Wisconsin 53706; and ‡Department of Biochemistry, Faculty of Medicine, University of British Columbia, 2075 Wesbrook Place, Vancouver, British Columbia V6T 1W5, Canada

Contributed by Masayasu Nomura, January 29, 1979

ABSTRACT **The 3072-nucleotide-long sequence of a segment from the 88-min region of the *Escherichia coli* chromosome has been determined. The sequence covers the genes for ribosomal proteins L11 (*rplK*), L1 (*rplA*), L10 (*rplJ*), and L7/L12 (*rplL*), and the 5′ end of the gene for the β subunit of RNA polymerase (*rpoB*), along with the presumed regulatory regions for these genes. The probable locations of the promoter for the first two genes (the L11 operon) and the promoter for the latter three genes (the proximal part of the β operon) have been identified. We have also found that the four ribosomal protein genes preferentially use codons that are recognized efficiently by the most abundant tRNA species. These and other features of the sequence results are discussed in relation to available information obtained from both *in vitro* and *in vivo* experiments on the expression of these ribosomal and RNA polymerase subunit genes.**

Genes specifying ribosomal proteins (r-proteins) and subunits of RNA polymerase are clustered together on the bacterial chromosome (for a review, see ref. 1). At the 88-min region of the *Escherichia coli* chromosome, four genes specifying r-proteins L11, L1, L10, and L7/L12 are located adjacent to the two genes specifying the β and β′ subunits of RNA polymerase (ref. 2; see Fig. 1). Recent experiments have demonstrated that the RNA polymerase subunit genes, *rpoB* (β) and *rpoC* (β′), are cotranscribed with r-protein genes *rplJ* (L10) and *rplL* (L7/L12) (3, 4). However, it has been shown that RNA polymerase subunits and r-proteins are not always coordinately regulated. For example, the synthesis of r-proteins, including L10 and L7/L12, is under stringent control (5, 6), but the synthesis of β and β′ is not (6–8). Rifampicin stimulates production of β and β′, but not that of L7/L12 (9). Thus, the regulation of these genes cannot be explained solely at the level of transcription initiation.

Because of these interesting regulatory features and their obvious importance in regulation of ribosome and RNA polymerase synthesis, we have determined the DNA sequence of the r-protein gene cluster and the 5′ end of the *rpoB* gene, along with the presumed regulatory sequences.

MATERIALS AND METHODS

The sources of all materials as well as the methods used to prepare DNA have been described (10). DNA fragments used in sequencing were from several different sources: λ*rif*^{d}18 (11), λ*rif*^{d}18Δ274 (12), Ch3R11i$^{\lambda}$ (13), Ch3R8i$^{\lambda}$ (like Ch3R11i$^{\lambda}$ except that it contains the *Eco*RI 2.6% fragment instead of the *Eco*RI 4.4%-L fragment; cf. ref. 13 and Fig. 1; constructed by C. S. Jinks in this laboratory), pJC701 (14), and pNO2016 and pNO2017. The latter two are hybrid plasmids derived from pVH51 (15) and carry the *Eco*RI/*Sma* fragment to the left and the *Eco*RI/*Hin*dII fragment to the right of the *Eco*RI 18.6%/4.4%-L site (unpublished work; cf. Fig. 1). The DNA was sequenced by the method of Maxam and Gilbert (16). Dyad symmetry and codon frequency were analyzed using the computer program of Korn *et al.* (17), as modified by R. Littlewood (Biophysics Laboratory, University of Wisconsin, Madison) for the use on a Univac 1110 computer.

RESULTS AND DISCUSSION

Various DNAs derived from λ*rif*^{d}18 transducing phages (2, 11) were used for the sequence work. Fig. 1 shows the experiments that were done to determine the nucleotide sequence shown in Fig. 2. The sequence begins upstream from the *rplK* (L11) gene, and extends into the *rpoB* (β) gene. Over 90% of the sequence has been determined from both strands of the DNA, increasing its reliability. Furthermore, most of the sequence has been determined from at least two sources of DNA. Because no discrepancy of sequence was found between different DNA sources, one can be quite confident that the DNAs constructed *in vitro* by gene cloning techniques contain exact copies of the sequences on λ*rif*^{d}18.

All four r-protein genes previously assigned to this region (2) were found in the sequence, and the predicted amino acid sequences are given in Fig. 2. The *rplK* (L11), *rplA* (L1), and *rplL* (L7/L12) gene sequences are in perfect agreement with published amino acid sequences (18–20). The amino acid sequence of L10 deduced from the DNA sequence of the *rplJ* gene differs in only two places from the published amino acid sequence (21): the codon corresponding to position 84 (arginine) of the published protein sequence is not in the DNA sequence (between nucleotides 1912 and 1913), and the glutamine at position 116 in the published protein sequence is glutamic acid according to the DNA sequence (nucleotides 2066–2068). Although the sequence of the β subunit of RNA polymerase has not been published, the NH_2-terminal four amino acids have been reported to be Met-Val-Tyr-Arg (22). A sequence coding for such an NH_2 terminus of a protein was not found in the sequence after the end of the *rplL* gene. However, a sequence starting at nucleotide 2972 was found that could code for the first three of the predicted amino acids. Furthermore, the sequence was preceded by a Shine–Dalgarno sequence (ref. 23; see below) and had no downstream termination codons in phase. This sequence is very likely the *rpoB* (β) gene.

The distance between the *rplK* and *rplA* genes is short. Only three nucleotides separate the TAA termination codon of the *rplK* gene from the ATG initiation codon of the *rplA* gene. Thus, the Shine–Dalgarno ribosome-binding sequence (see below) for the *rplA* gene lies within the amino acid coding sequence of the preceding *rplK* gene. Therefore, the synthesis

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.

Abbreviation: r-protein, ribosomal protein.

† To whom reprint requests should be addressed.

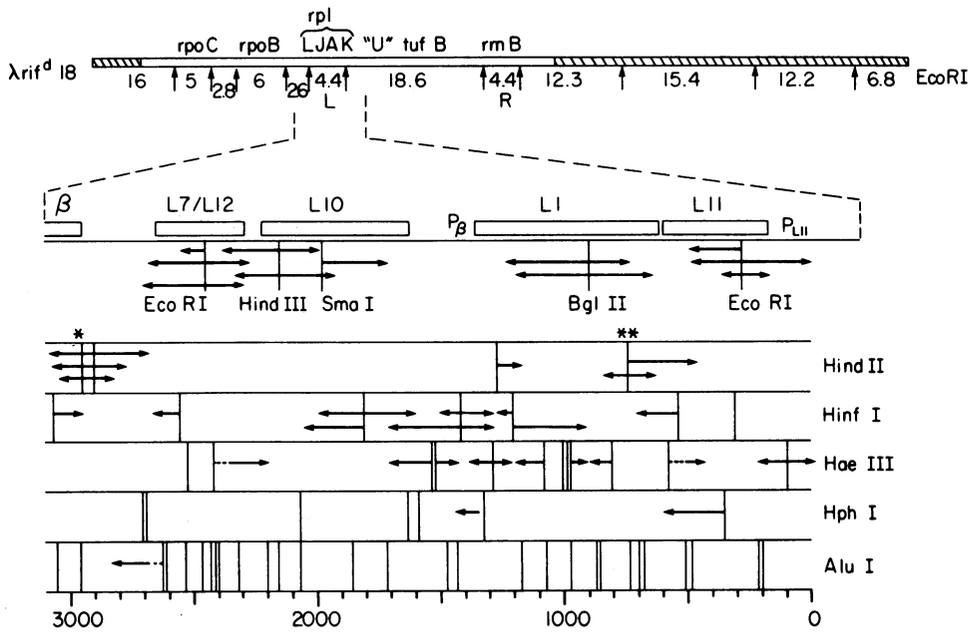

FIG. 1. Restriction enzyme maps and sequencing experiments in the *rpl–rpoB* region. The top of the figure shows the location of genes and the *Eco*RI cleavage sites on λ*rif*[d]18 (cf. refs. 2, 13). Approximate sizes of *Eco*RI fragments are indicated in %-λ units (1 %-λ unit is 465 base pairs). Two 4.4% fragments are called 4.4%-L and 4.4%-R, as indicated. The hatched area represents λ DNA and the open area represents the bacterial DNA carried by the phage. The expanded part of the picture includes the sequenced regions. The single asterisk (*) indicates a *Hin*dII site that is also a *Sal* I site, and the double asterisk (**) indicates a *Hin*dII site that is also a *Hpa* I site. The positions of the *Eco*RI (2), *Hin*dIII (2), *Sma* I (2), and *Sal* I (3) cleavage sites have been reported. The positions of the *Bgl* II and *Hin*dII (including *Sal* I and *Hpa* I) cleavage sites were determined by single or double digestion of purified *Eco*RI fragments. *Hin*fI, *Hae* III, and *Hph* I maps were determined partially by double digestion of purified *Eco*RI fragments and completed by DNA sequencing. The *Alu* I map was determined by DNA sequencing. Each horizontal arrow represents a single Maxam–Gilbert sequencing experiment, and one to five sequencing gels. Broken regions of the arrows indicate bases that were run off the bottom of all of the sequencing gels of that experiment. No attempt has been made to indicate which source of DNA was used for each experiment, although most regions of the DNA have been sequenced from at least two different DNA sources.

of both L11 and L1 takes place almost certainly from a singly polycistronic mRNA transcribed from the L11 operon. In contrast to the above situation, the corresponding intercistronic distances are 413, 67, and 321 nucleotides between *rplA* and *rplJ*, between *rplJ* and *rplL*, and between *rplL* and *rpoB*, respectively. As will be discussed below, these intercistronic regions are probably used to provide regulatory signals, such as promoters or attenuators, or to provide suitable sites for (hypothetical) mRNA processing that might be essential for proper translation.

Promoters. A major reason for determining this DNA sequence was to locate the promoters and examine possible regulatory sequences of these transcription units. *In vitro* transcription experiments (unpublished) similar to those described previously (10) have identified a transcription start site around nucleotide 1348 (±1). This transcript begins with the sequence pppCpC. Because no other specific transcription initiation was observed *in vitro* between this position and *rplJ*, it is likely that this corresponds to the β promoter identified by genetic experiments (3, 4). Consistent with this conclusion is the observation of W. Taylor and R. R. Burgess, (personal communication) that the *Hae* III fragment (1287–1523) from this region specifically binds RNA polymerase in the presence of CTP.

Previous experiments have also indicated that the *rplK* and *rplA* genes are cotranscribed (3, 4) and their promoter (called the "L11 promoter") is located between *rplK* and the gene for "U" (ref. 13; see Fig. 1). Preliminary experiments using small DNA fragments as template indicate that an *in vitro* transcript is initiated at nucleotide 79, and this is likely to be the L11 promoter (unpublished). It should be remembered, however, that these *in vitro* experiments only assign possible start sites, and do not rigorously correlate these promoters to the known *in vivo* transcription units.

The *in vitro* transcripts are both initiated a few bases downstream from sequences which differ from the canonical "Pribnow box," T-A-T-R-A-T-R (R = purine), (24) at two positions. These Pribnow boxes are indicated in Fig. 2. Variation from the perfect Pribnow box sequence by up to two bases has been observed in previously sequenced promoters (see refs. 25 and 26 for reviews). It should be noted that there is a perfect Pribnow box at position 1354–1360 that is only a few nucleotides downstream from the *in vitro* transcription start site; apparently, this Pribnow box is not recognized under the conditions used in our *in vitro* experiments. Therefore, a perfect Pribnow box sequence does not necessarily specify the position of a strong start site.

Another common feature of many promoters is the sequence at approximately position −35 with respect to the start site (see refs. 25 and 26 for reviews). In the case of the L11 promoter, the sequence G-T-T-G-C-A-C-A is similar to other promoters. However, the β promoter has no such common sequence. Thus the sequence features that make the β promoter a promoter are not obvious.

Dyad symmetry has not been found to be a general feature of promoters, but striking dyad symmetry has been noted in the tyrosine tRNA promoter (27) and the promoter of the *spc* operon of r-proteins (10). The *str* promoter also has a dyad symmetry (10). The L11 promoter region sequenced in this study contains a region of dyad sequence symmetry, which is indicated in Fig. 2. Similar sequence features are also found around the β promoter; two possible ways to draw symmetry are indicated in Fig. 2. The role of symmetry in these promoters is not known, but they could be recognition sites for unidentified regulatory proteins. In the case of the L11 and β promoters, the symmetry could possibly be related to termination of the upstream transcription unit, as discussed below. Besides the presence of symmetry, there are no obvious features of the L11 and β promoters in common with previously sequenced promoters of r-protein operons (10).

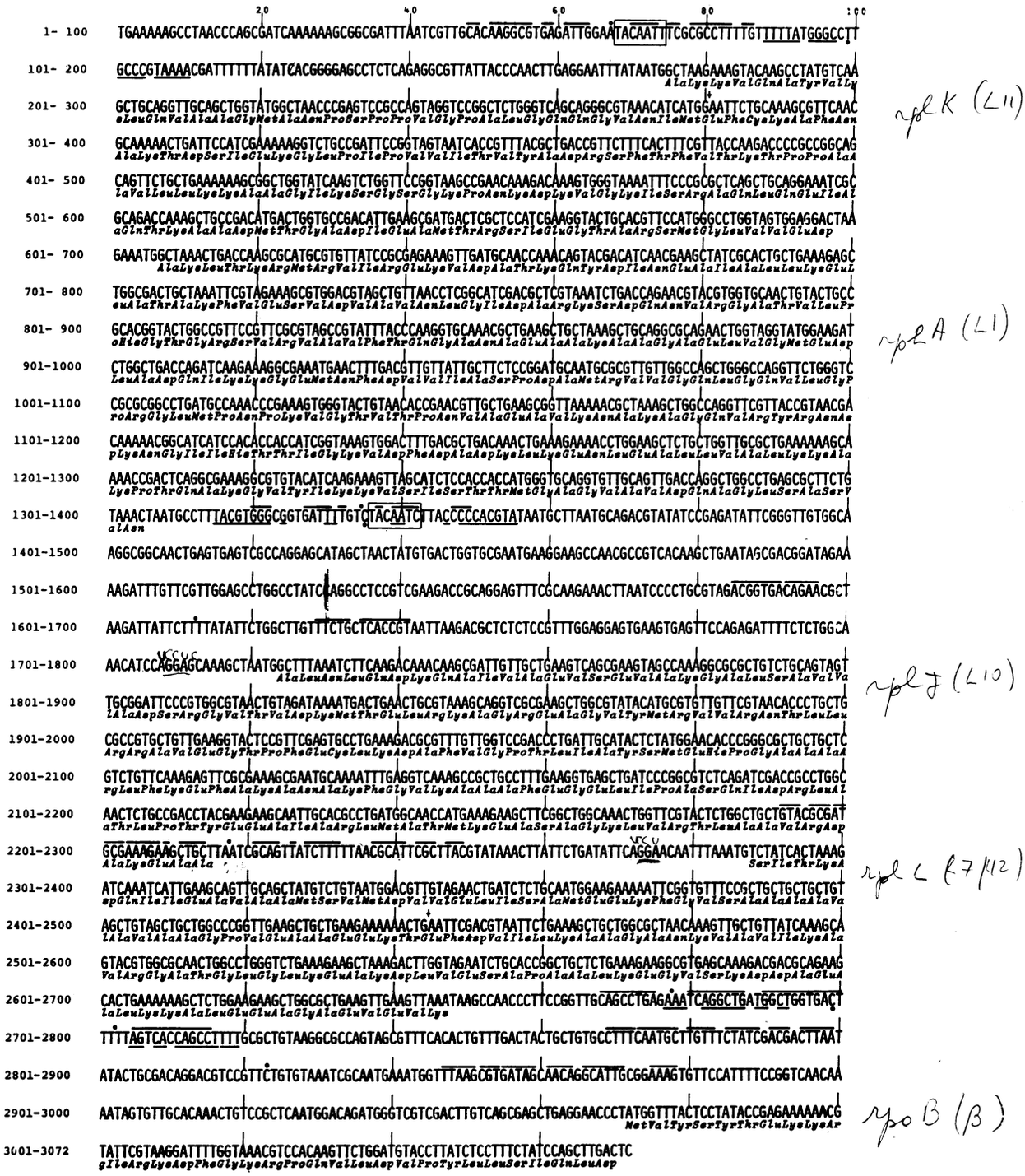

FIG. 2. DNA sequence of the ribosomal protein gene cluster from λ*rif*^{d}18. Only one strand, that with the polarity of the mRNAs, is given. The predicted amino acid sequences of the coded proteins are given below the coding DNA. The *rplK* (L11) gene (including the ATG initiation codon) begins at position 172; the *rplA* (L1) gene begins at position 604; the *rplJ* (L10) gene begins at position 1721; the *rplL* (L7/L12) gene begins at position 2285; and the *rpoB* (β) gene is presumed to begin at position 2972 (see text). The Pribnow sequences of the two promoters are boxed. The box at position 68 is the L11 promoter, and the transcription start site is probably at position 79 (see text). The box at position 1336 is the β promoter, and the start site is at around position 1348 (±1) (see text). Regions of dyad symmetry or possible base pairing are indicated by overlining or underlining with a dot indicating a center of a dyad symmetry. The arrows at positions 290 and 2444 indicate the *Eco*RI sites that define the 4.4%-L fragment.

Transcription Termination Signals. It was expected initially that the L11 operon mRNA would terminate at the end of *rplA*, before the β promoter. However, there are only approximately 40 nucleotides between *rplA* and the start site of the β promoter. Features common in previously sequenced termination signals include a base-paired stem and loop structure preceding a run of U residues at the 3′ end of the RNA (25) or the sequence C-A-A-T-C-A-A at the 3′ end of the RNA at highly ρ-dependent termination sites (28, 29). None of these features are found in the sequence after *rplA*. It is possible that the symmetry noted above in the β promoter region has some function in the termination of the L11 operon messenger. A relationship between transcription initiation and termination signals has been proposed (28). Alternatively, it remains a possibility that the L11

Table 1. Use of codons in the r-protein genes for L11, L1, L10, and L7/L12

Phe	TTT	6	Ser	TCT	12	Tyr	TAT	1	Cys	TGT	0
Phe	TTC	10	Ser	TCC	9	Tyr	TAC	7	Cys	TGC	2
Leu	TTA	1	Ser	TCA	0		TAA	4		TGA	0
Leu	TTG	1	Ser	TCG	1		TAG	0	Trp	TGG	0
Leu	CTT	1	Pro	CCT	0	His	CAT	0	Arg	CGT	16
Leu	CTC	1	Pro	CCC	0	His	CAC	3	Arg	CGC	12
Leu	CTA	0	Pro	CCA	2	Gln	CAA	5	Arg	CGA	0
Leu	CTG	43	Pro	CCG	21	Gln	CAG	15	Arg	CGG	0
Ile	ATT	9	Thr	ACT	18	Asn	AAT	2	Ser	AGT	1
Ile	ATC	22	Thr	ACC	15	Asn	AAC	18	Ser	AGC	5
Ile	ATA	0	Thr	ACA	1	Lys	AAA	55	Arg	AGA	0
Met	ATG	19	Thr	ACG	0	Lys	AAG	8	Arg	AGG	0
Val	GTT	35	Ala	GCT	60	Asp	GAT	8	Gly	GGT	25
Val	GTC	3	Ala	GCC	7	Asp	GAC	24	Gly	GGC	24
Val	GTA	28	Ala	GCA	31	Glu	GAA	41	Gly	GGA	0
Val	GTG	7	Ala	GCG	16	Glu	GAG	7	Gly	GGG	0

The numbers following the codons for the indicated amino acids are the number of times the codons are used in the sequence shown in Fig. 2. The number of methionine codons does not include the initiation codons.

operon mRNA does not terminate after *rplA*, and transcription continues into the β operon. It should be noted that the presence of a promoter (the β promoter) between the *rplA* and *rplJ* genes was demonstrated only under conditions in which the L11 promoter is not functioning (3, 4). Therefore, an interesting possibility exists that under "normal" conditions all the β operon genes are cotranscribed with the L11 operon genes and that the β promoter discussed above is used only when transcription from an upstream promoter stops.

The L11 operon mRNA does contain a structure resembling a transcription termination signal near its 5′ end. The symmetry centered around position 99 (Fig. 2) could form a base-paired stem and loop structure preceding the sequence U-U-U-U-U-U-A in a transcript, which resembles a termination signal. Anti-termination at such a site could regulate transcription of the L11 operon, as has been observed in the case of the *trp* operon (30). Similar structures were observed at the 5′ end of the mRNAs of the *str* and *spc* operons of r-proteins, and the structure in the *str* operon was in fact found to act as a terminator *in vitro* (10).

The symmetry centered within the L11 promoter also precedes the sequence T-T-T-T-T-A. It is possible that this sequence represents the termination of the previous operon, which contains the "U" protein (13).

One position where a termination sequence might have been expected was the region between *rplL* and *rpoB*. Because the syntheses of r-proteins and RNA polymerase subunits are not regulated identically (see introduction), one hypothesis would be that an "attenuator" would regulate termination of transcription between *rplL* and *rpoB*. Although there are 321 nucleotides between these two genes, the function of this long sequence is not obvious. Fig. 2 shows three dyad symmetry patterns at the end of *rplL*, which could be important in regulation of transcription through this region. Such symmetrical sites may represent a site of interaction with some unknown regulatory protein. Fig. 2 also shows a possible region of long-range base pairing, starting at nucleotide 2769. There are several other structures with dyad symmetry that are not shown in the figure.

Translation Initiation Signals. Because most of the r-proteins are present in equimolar amounts, and the synthesis rate of r-proteins in exponentially growing cells appears to be regulated by the amount of mRNA (see ref. 1 for a review), it is expected that r-proteins in the same operon (and possibly all r-proteins except L7/L12; see below) should be translated at equal rates. One feature of a ribosome binding site is the sequence complementary to the 3′ end of 16S rRNA, or the Shine–Dalgarno sequence (23). It was previously noted that the Shine–Dalgarno sequence of several r-protein genes included the sequence G-G-A-G (10). Both the *rplA* and *rplJ* genes have Shine–Dalgarno sequences that include this sequence. However, the ribosome-binding sites of both *rplK* and *rplL* do not have this sequence; the sequences are G-A-G-G or A-G-G-A for *rplK* and A-G-G-A for *rplL*.

Protein L7/L12 is present in multiple copies per ribosome (31, 32), but L11 is not. Thus, our sequence data do not reveal any obvious correlation between Shine–Dalgarno sequences and presumed translational efficiencies. Some unknown structural features of mRNA may play a role in determining translational efficiencies. For example, the large base-paired stem and loop structure (from 2192 through 2249 as indicated in Fig. 2) before the *rplL* gene might be responsible for the higher synthesis rate of L7/L12 relative to that of L10.

Codon Usage. The codon usage in the r-protein genes was found to be highly nonrandom (Table 1). The preferential codon usage was quite different from that observed in single-strand phages (compiled in ref. 33), phage λ repressor (34), and β-lactamase (35), with some similarity to that of the *lacI* gene (36). Both the r-protein genes and the *lacI* gene show a preference for the use of CGY (Y = pyrimidine) for arginine, CTG for leucine, GGY for glycine, ATY for isoleucine, GAA for glutamic acid, and AAA for lysine. For each of these cases (refs. 37, 38, 39, 40, 41, and 42, respectively), the codons preferentially used are those recognized efficiently by the most abundant tRNA species, with the preferences more pronounced for the r-protein genes than for the *lacI* gene. Other examples of preferential codon usage in the r-protein genes, but not in other genes, which correlate with the major tRNA species are alanine (43, 44), valine (45), serine (46), and proline (47). The codons for asparagine and threonine are also nonrandom, but there are not enough data available to assess the significance of this with respect to tRNA species.

Preferential codon usage in r-protein genes may reflect the cell's need that the translation of these essential proteins be more efficient than that of a protein such as the *lac* repressor, present in only small amounts. Alternatively, the preferential codon usage could be the consequence of evolution to minimize translational errors in the synthesis of r-proteins. Several studies

(48–50) have shown that when decreased levels of correct amino acyl tRNAs limit the rate of protein synthesis, the rate of occurrence of translational errors increases, presumably due to more effective competition by incorrect amino acyl tRNAs. Thus preferential use of codons corresponding to major tRNA species might be expected to reduce translational errors in r-protein synthesis and thereby prevent "error catastrophe" (51, 52). More sequence analyses of both ribosomal and nonribosomal protein genes should be able to test this idea.

Relation of Sequence to *In Vivo* Experiments. Hybrid plasmid pNO2016 is a derivative of pVH51 and carries the DNA segment extending from the *Eco*RI site at position 281 to the *Sma* I site at position 1986 (unpublished). It was found that strains containing this plasmid produce an RNA species approximately 250 bases long that is not produced in strains without the plasmid (unpublished). Preliminary sequence analysis of this RNA species indicates that the 5′ end of this RNA corresponds to sequences near the *in vitro* β promoter start site. This gives confidence that the β promoter identified *in vitro* is functional as a promoter *in vivo*. The 3′ end of the RNA species is located near position 1600. Why an RNA initiated at the β promoter does not extend into the structural genes is not known. Possibilities include: (*i*) This small RNA may have some function in the cell, and the identified promoter may not really be the promoter for the β operon. (*ii*) There could be an attenuator (with a structure different from known termination structures) at which antitermination could regulate transcription of the β operon. (*iii*) The β operon mRNA could be processed, with cleavage around position 1600. In this regard, it is interesting to note that the sequence around position 1590 can base pair with the sequence around position 1635 (see Fig. 2), which could provide a recognition site for some hypothetical processing enzyme.

Although the DNA sequence reported in this communication does not provide answers to the questions of regulation of these genes, it does provide a basis for designing and interpreting *in vivo* and *in vitro* experiments such as those mentioned above.

We thank Dr. D. Söll for useful discussion on isoaccepting tRNAs and A. Arfsten and T. Ikemura for participating in unpublished experiments discussed in the text. This work was supported in part by the College of Agriculture and Life Sciences, University of Wisconsin, Madison, and by grants from the National Science Foundation (PCM7818490, administered by M.N.), the National Institutes of Health (GM-20427, administered by M.N.; GM-24010, administered by P.P.D.), and the Medical Research Council of Canada (MA-6340, administered by P.P.D.). This is paper no. 2323 from the Laboratory of Genetics, University of Wisconsin, Madison.

1. Nomura, M., Morgan, E. A. & Jaskunas, S. R. (1977) *Annu. Rev. Genet.* **11**, 297–347.
2. Lindahl, L., Yamamoto, M., Nomura, M., Kirschbaum, J. B., Allet, B. & Rochaix, J.-D. (1977) *J. Mol. Biol.* **109**, 23–47.
3. Yamamoto, M. & Nomura, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3891–3895.
4. Linn, T. & Scaife, J. (1978) *Nature (London)* **276**, 33–37.
5. Dennis, P. P. & Nomura, M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3819–3824.
6. Reeh, S., Pedersen, S. & Friesen, J. D. (1976) *Mol. Gen. Genet.* **149**, 279–289.
7. Blumenthal, R. M., Lemaux, P. G., Neidhardt, F. C. & Dennis, P. P. (1976) *Mol. Gen. Genet.* **149**, 291–296.
8. Maher, D. L. & Dennis, P. P. (1977) *Mol. Gen. Genet.* **155**, 203–211.
9. Hayward, R. S. & Fyfe, S. K. (1978) *Mol. Gen. Genet.* **160**, 77–88.
10. Post, L. E., Arfsten, A. E., Reusser, F. & Nomura, M. (1978) *Cell* **15**, 215–229.
11. Kirschbaum, J. B. & Konrad, E. B. (1973) *J. Bacteriol.* **116**, 517–526.
12. Yamamoto, M., Lindahl, L. & Nomura, M. (1976) *Cell* **7**, 179–190.
13. Yamamoto, M. & Nomura, M. (1979) *J. Bacteriol.*, **137**, 584–594.
14. Collins, J., Fiil, N. P., Jørgensen, P. & Friesen, J. D. (1976) in *Alfred Benzon Symposium IX: Control of Ribosome Synthesis*, eds. Kjeldgaard, N. O. & Maaløe, O. (Academic, New York), pp. 356–367.
15. Hershfield, V., Boyer, H. W., Chow, L. & Helinski, D. R. (1976) *J. Bacteriol.* **126**, 447–453.
16. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
17. Korn, L. J., Queen, C. L. & Wegman, M. N. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4401–4405.
18. Dognin, M. J. & Wittmann-Liebold, B. (1977) *FEBS Lett.* **84**, 342–346.
19. Brauer, D. & Öchsner, I. (1978) *FEBS Lett.* **96**, 317–321.
20. Terhorst, C., Möller, W., Laursen, R. & Wittmann-Liebold, B. (1973) *Eur. J. Biochem.* **34**, 138–152.
21. Heiland, I., Brauer, D. & Wittmann-Liebold, B. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 1751–1770.
22. Fujiki, H. & Zurek, G. (1975) *FEBS Lett.* **55**, 242–244.
23. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
24. Pribnow, D. (1975) *J. Mol. Biol.* **99**, 419–443.
25. Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–205.
26. Reznikoff, W. S. & Abelson, J. N. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 221–243.
27. Sekiya, T. & Khorana, H. G. (1974) *Proc. Natl. Acad. Sci. USA* **74**, 2978–2982.
28. Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D. L. (1978) *Nature (London)* **272**, 414–423.
29. Kupper, H., Sekiya, T., Rosenberg, M., Egan, J. & Landy, A. (1978) *Nature (London)* **272**, 423–428.
30. Lee, F. & Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4365–4369.
31. Subramanian, A. R. (1975) *J. Mol. Biol.* **95**, 1–8.
32. Hardy, S. J. S. (1975) *Mol. Gen. Genet.* **140**, 253–274.
33. Godson, G. N., Barrell, B. G., Staden, R. & Fiddes, J. C. (1978) *Nature (London)* **276**, 236–247.
34. Sauer, R. T. (1978) *Nature (London)* **276**, 301–302.
35. Sutcliffe, J. G. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3737–3741.
36. Farabaugh, P. J. (1978) *Nature (London)* **274**, 765–769.
37. Celis, T. F. R. & Maas, W. K. (1971) *J. Mol. Biol.* **62**, 179–188.
38. Blank, H. U. & Söll, D. (1971) *J. Biol. Chem.* **246**, 4947–4950.
39. Fleck, E. W. & Carbon, J. (1975) *J. Bacteriol.* **122**, 492–501.
40. Yarus, M. & Barrell, B. G. (1971) *Biochem. Biophys. Res. Commun.* **43**, 729–734.
41. Ohashi, Z., Saneyoshi, M., Harada, F., Hara, H. & Nishimura, S. (1970) *Biochem. Biophys. Res. Commun.* **40**, 866–872.
42. Chakraburtty, K., Steinschneider, A., Case, R. V. & Mehler, A. H. (1975) *Nucleic Acids Res.* **2**, 2069–2075.
43. Williams, R. J., Nagel, W., Roe, B. & Dudock, B. (1974) *Biochem. Biophys. Res. Commun.* **60**, 1215–1221.
44. Lund, E. & Dahlberg, J. E. (1977) *Cell* **11**, 247–262.
45. Kimura, F., Harada, F. & Nishimura, S. (1971) *Biochemistry* **10**, 3277–3283.
46. Roy, K. L. & Söll, D. (1970) *J. Biol. Chem.* **245**, 1394–1400.
47. Söll, D., Cherayil, J. D. & Bock, R. M. (1967) *J. Mol. Biol.* **29**, 97–112.
48. Edelmann, P. & Gallant, J. (1977) *Cell* **10**, 131–137.
49. O'Farrell, P. H. (1978) *Cell* **14**, 545–557.
50. Parker, J., Pollard, J. W., Friesen, J. D. & Stanners, C. P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1091–1095.
51. Orgel, L. E. (1963) *Proc. Natl. Acad. Sci. USA* **49**, 517–521.
52. Edelmann, P. & Gallant, J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3396–3398.