Isolation and Sequence Analysis of the Gene (*cpdB*) Encoding Periplasmic 2',3'-Cyclic Phosphodiesterase

JU LIU, DENNIS M. BURNS, AND IFOR R. BEACHAM*

School of Science, Griffith University, Nathan, Brisbane, Queensland 4111, Australia

The *cpdB* gene encodes a periplasmic 2',3'-cyclic phosphodiesterase (3'-nucleotidase). This enzyme has been purified previously and the gene is located at 96 min on the *Escherichia coli* chromosome. In this study the *cpdB* gene was cloned from *Cla*I-cleaved DNA, and the gene product was identified. DNA blotting experiments showed that the recombinant plasmid contains a deletion with respect to the expected genomic fragment of approximately 4 kilobases, which extends into the vector. Furthermore, the gene was absent from three other recombinant libraries. Together, these findings suggest the presence in the genome of an adjacent gene whose product is lethal when it is present on a multicopy plasmid. The nucleotide sequence of the *cpdB* gene was also determined. The 5' and 3' untranslated sequences contain characteristic sequences that are involved in the initiation and termination of transcription, including two possible promoters, one of which may contain two overlapping −10 sequences. A strong Shine-Dalgarno sequence is followed by an open reading frame which corresponds to a protein having a molecular weight of 70,954. The first 19 amino acid residues have the characteristics of a signal peptide. The 3' untranslated sequence contains two putative rho-independent transcription terminators having low thermodynamic stability.

The *cpdB* gene of *Escherichia coli* (5) encodes a 2',3'-cyclic phosphodiesterase, which is a bifunctional enzyme that also possesses 3'-nucleotidase activity (1–3), as follows: 2',3'-cyclic UMP→ 3'-UMP→ uridine + PO_4. Interest in the *cpdB* gene derives from the fact that it encodes a secreted (i.e., periplasmic [4]) enzyme and from the fact that there is evidence that the enzyme has two kinetically distinguishable active sites (2). This and other aspects of the enzymology of 2',3'-cyclic phosphodiesterase can potentially be explored by genetic manipulation of the isolated gene (15).

MATERIALS AND METHODS

Bacterial strains and media. All of the media used and selection for CpdB$^+$ by using strain 4K-39-3/9 (CpdB$^-$) as the recipient have been described previously (5, 13). Strain 4K is the parent of strain 4K-39-3/9 (13).

Restriction endonuclease mapping, DNA isolation, DNA manipulation and cloning, and maxi cell analysis. Restriction endonuclease mapping, DNA isolation, DNA manipulation and cloning, and maxicell analysis were also carried out as previously described (8–10, 13).

DNA blotting. For DNA blotting (38) DNA (2 μg) was electrophoresed on a 0.7% agarose gel and transferred to a membrane (Gene Screen; New England Nuclear Corp.) essentially as described previously (34). The membrane was then prehybridized by incubating it in a sealed plastic bag at 65°C for 2 to 5 h in 4× saline-sodium citrate (SSC; 1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin, 0.2% Ficoll (M_r, 400,000 [400K]), and 100 μg of denatured calf-thymus DNA per ml. This solution was then replaced with fresh prehybridization solution containing the probe; the probe was labeled with [^{32}P]dCTP by nick translation (34). Following hybridization for 16 h at 65°C, the membrane was washed once with 1× SSC containing 0.1% sodium dodecyl sulfate and twice with 1× SSC at 65°C for 30 min, dried, and autoradiographed. The fragments

produced by cleavage of λ DNA with *Hin*dIII served as molecular weight markers. The cleaved DNA was electrophoresed on the same agarose gel and blotted onto the membrane. This portion of the membrane was then removed, hybridized with labeled λ DNA, and subsequently aligned with the rest of the membrane.

Enzyme assays. Cyclic phosphodiesterase and 3'-nucleotidase were assayed as described previously (5).

DNA sequencing. DNA fragments were cloned into the replicative form of M13 mp8 or M13 mp9 (26) and were sequenced by the chain termination method (33). These DNA fragments were generated by using the following strategies: (i) by using known restriction sites (see Fig. 1); (ii) by digesting large defined fragments of the insertion with enzymes which cut frequently (*Sau*3A, *Taq*I, and *Hpa*II), followed by random cloning into M13; (iii) shortening large fragments by digestion (from both ends) with BAL 31 nuclease (23), followed by cloning into the *Sma*I site of M13 mp8; and (iv) shortening a given defined fragment essentially in only one direction by a cut-digest-recut procedure. As an example of this latter approach, pDB5 was cut with *Xma*III, digested with BAL 31 nuclease for various times, ethanol precipitated, and then recut with *Eco*RI. The resulting fragments were then cloned directly into M13 mp8 that was cut with *Sma*I and *Eco*RI. Recombinants obtained in this way contained only the region of interest; we assumed that fragments which contained the origin of replication from plasmid pDB5 did not result in viable M13 recombinants. When necessary, the orientation of insertions in M13 vectors was deduced by hybridization, using clones of known orientation and the C-test procedure of Messing (26); however, template DNA rather than phage was used.

RESULTS

Isolation of the *cpdB* gene. The *cpdB* gene has been mapped previously on the *E. coli* chromosome (5) and has been shown to be located on the F-prime plasmid F117 (24; data not shown). Therefore, DNA preparations enriched in F117 DNA were used in the construction of recombinant libraries

* Corresponding author.

in vector pBR322 (6) (see Materials and Methods). Following transformation with plasmid DNA, CpdB$^+$ recombinants were directly selected in a CpdB$^-$ recipient. Successive experiments with recombinant libraries prepared by cloning DNA cleaved with *Hind*III, *Sal*I, *Bam*HI, or *Pst*I did not result in CpdB$^+$ clones, despite the fact that calculation (11) showed that the entire *E. coli* genome was represented with a probability of more than 95%. Selection with a library constructed from *Cla*I-digested DNA yielded a single clone. Recombinant plasmid pDB4 clearly encoded 2′,3′-cyclic phosphodiesterase since (i) it transformed *cpdB*$^-$ strains to CpdB$^+$ and (ii) CpdB$^+$ transformants expressed activity; furthermore the enzyme was somewhat overproduced (four-fold more enzyme than in a wild-type *cpdB*$^+$ strain) (data not shown).
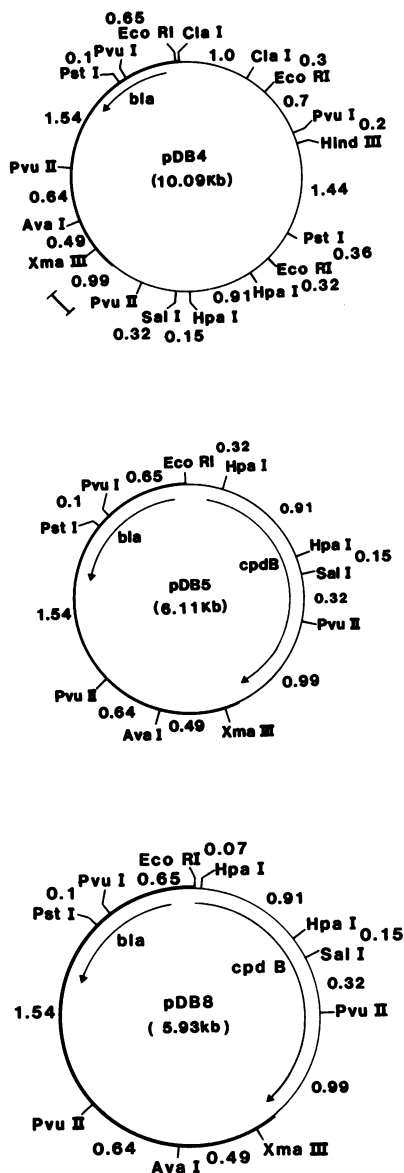

FIG. 1. Restriction maps of plasmids. The bar on the outside of pDB4 indicates the region containing a deletion (see text). The arrows indicate the direction of transcription. The thick lines represent the vector moieties.
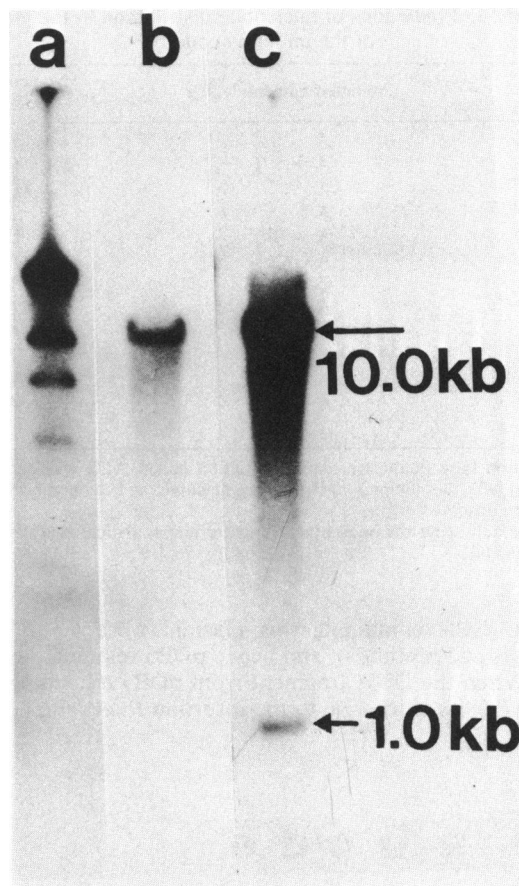
FIG. 2. Southern blotting of *Cla*I-digested *E. coli* DNA. Lane a, λ DNA digested with *Hind*III, probed with labeled λ DNA; lanes b and c, *E. coli* DNA digested with *Cla*I and probed with pDB4 (lane c was exposed longer than lane b).

A restriction map of pDB4 is shown in Fig. 1. The plasmid (10.09 kilobases [kb]) had to contain an insertion with a minimum size of 5.7 kb, based on the size of the vector (4.36 kb). However, this map is not compatible with the simple cloning of a 5.7-kb *Cla*I fragment since a *Cla*I site was found only at one junction with the vector, suggesting that some rearrangement of the DNA had occurred. Also, 1 kb from this *Cla*I site there was a *Cla*I site within the inserted DNA. To establish the genomic arrangement of this DNA, pDB4 was used to probe Southern blots of *Cla*I-restricted chromosomal DNA. Two chromosomal fragments (10 and 1 kb) were found (Fig. 2). Therefore, the 1-kb *Cla*I fragment in pDB4 had to be a result of partial *Cla*I digestion of the DNA used in the construction of pDB4 and not a result of a rearrangement in the plasmid. However, the presence of a genomic 10-kb fragment and the absence of vector *Sal*I, *Hind*III, and *Bam*HI sites clearly indicated that a large deletion (approximately 4 kb) had occurred during the isolation of pDB4, which could account for the absence of a *Cla*I site at one vector-insertion junction, as indicated in Fig. 1. Assuming that the deletion included this original *Cla*I site and extended into the vector, then the location of the vector-insertion junction had to lie between the vector *Xma*III site and the counterclockwise *Pvu*II site (Fig. 1). This was confirmed by DNA sequence analysis (see below).

Identification of the *cpdB* gene. pDB5 (Fig. 1) was derived from pDB4 by digestion of the latter with *Eco*RI, followed by

TABLE 1. Predictions of nucleotide distribution in the vicinity of the initiation codon

Position[a]	Predicted nucleotide(s)[b]	Nucleotide in cpdB gene
−20	A	A
−4	A >> T	C
−3	A	C
−2	A/T	T
−1	A > T	G
3	A/G	A
4	C > A	T
5	A/T	T
6	A >> C	A
10	T >> C	T
11	A > T	T
12	A	A
13	A > T	G

[a] The first base of the translation initiation codon ATG was designated position 0 (40); the numbers upstream are negative, and the numbers downstream are positive.

[b] These data were obtained by a computer and statistical analysis of 124 sequences (40).

ligation. Cells containing this plasmid were CpdB[+] (see Materials and Methods), and hence pDB5 retained the cpdB gene. When the DNA fragment from pDB5 extending from the EcoRI site clockwise to the insertion PvuII site (1.7 kb)
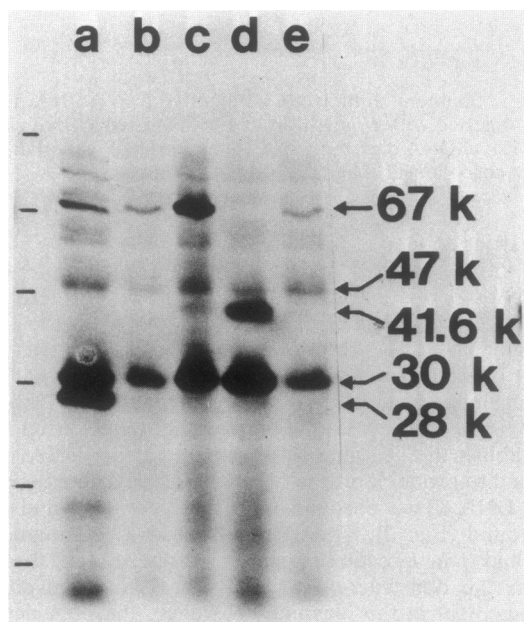

FIG. 3. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of proteins synthesized in maxicells. The maxicells contained pDB4 (lane a), pDB5 (lane b), pDB6 (lane d), or pDB8 (lane e). Lane c contained pAE, which is a plasmid identical to pDB5 except that the vector moiety is derived from pBR327 (12) and not from pBR322; the former vector is present at a higher copy number than pBR322 (12). The positions of the following molecular weight standards, which were labeled with ^{125}I, are indicated on the left (from top to bottom): phosphorylase B (92.5K), bovine serum albumin (66.2K), ovalbumin (45K), carbonic anhydrase (31K), trypsin inhibitor (21.5K), and α-lactalbumin (14.4K). The 30K protein was β-lactamase, which was derived from the vector moiety of all plasmids.

was cloned into pBR322 or the 0.9-kb HpaI fragment was deleted by cleavage and ligation, then the resulting plasmids (pDB6 and pDB7, respectively) were cpdB[−]. Also, when the fragment extending from SalI clockwise to the vector AvaI site (1.80 kb) was cloned into pBR322, the resulting plasmid did not transform cpdB[−] cells to CpdB[+] (data not shown). Hence, the PvuII and SalI sites and one or both of the HpaI sites in pDB5 are within the cpdB gene.

Further localization of the coding sequence was obtained by cleavage of pDB5 at the EcoRI site, followed by BAL 31 nuclease digestion (8, 23) for various times. BAL 31 nuclease-treated DNA was then cleaved with AvaI and ligated to the AvaI-SmaI bla-containing fragment of pHP34 (30). The smallest CpdB[+] plasmid obtained by this procedure was designated pDB8 (Fig. 1).

Expression and direction of transcription of the cpdB gene. The protein products derived from many of the plasmids described above were examined in maxicells (see Materials and Methods). Three major products (28K, 47K, and 67K) were found in pDB4-containing cells (Fig. 3). 2′,3′-cyclic phosphodiesterase has been purified previously, and its M_r is 68,000 (3). Thus, the 67K polypeptide probably represents cyclic phosphodiesterase, and this was confirmed as follows. In maxicells containing pDB5 (cpdB[+]) (and its derivatives), the 28K protein was absent, but the 47K and 67K proteins were retained; this eliminated the 28K gene product as the cpdB gene product (Fig. 3, lanes b, c, and e). When maxicells containing pDB4 were fractionated into periplasmic and cytoplasmic (plus membrane) components (8), the 67K protein, but not the 47K or 28K protein, was clearly found in the periplasm (Fig. 4), as expected if the 67K protein is the product of the cpdB gene. The absence of the 67K protein in maxicells containing pDB6 (which is cpdB[−]) while the 47K protein was present (Fig. 3, lane d) further confirmed the identity of the former protein as the product of
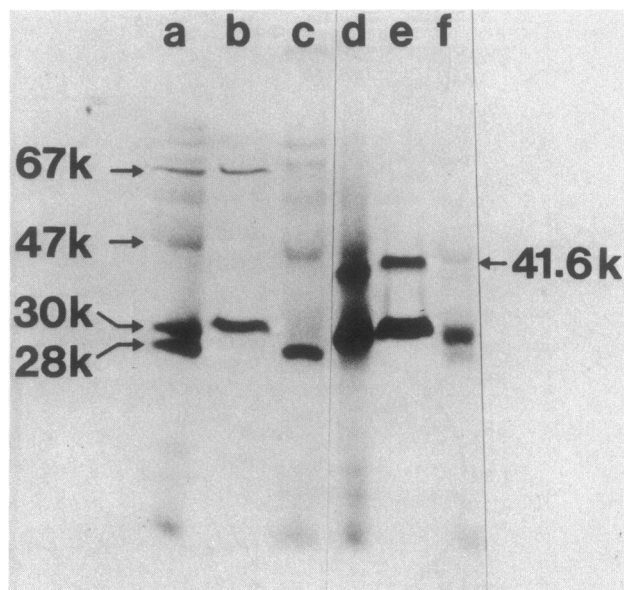

FIG. 4. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of extracts of maxicells containing pDB4 (lanes a through c) or pDB6 (lanes d through f). Maxicells were fractionated into periplasmic fractions (lanes b and e) and fractions containing cytoplasm plus membranes (lanes c and f). Lanes a and d contained unfractionated maxicells.
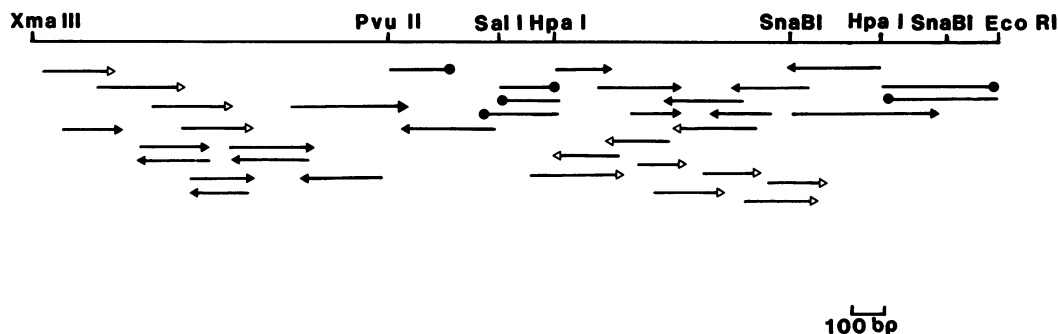
FIG. 5. Sequencing strategy for the *cpdB* gene derived from pDB5 or pAE (see Fig. 1 and the legend to Fig. 3). Both the *Xma*III and *Eco*RI sites were of vector origin. The lines with arrowheads or circles indicate the direction and extent of sequencing. Lines with solid circles indicate restriction fragments cloned directly into M13; lines with open arrowheads indicate fragments shortened at both ends with BAL 31 nuclease; and lines with closed arrowheads indicate fragments that were generated by using the cut-digest-recut strategy (see Materials and Methods). bp, Base pairs.

the *cpdB* gene. Concomitantly with the absence of the 67K protein, a 41.6K polypeptide was apparent; this polypeptide was also secreted (Fig. 4, lanes d through f). These results strongly suggest that the carboxyl end of the protein is lacking as a result of the construction of pDB6 and hence imply the direction of transcription shown in Fig. 1. The direction of transcription was confirmed by sequence analysis.

Nucleotide sequence analysis. The strategy used to sequence the *cpdB* gene is shown in Fig. 5 (see Materials and Methods), and the sequence and predicted amino acid sequence are shown in Fig. 6. The open reading frame encoded a protein having a molecular weight of 70,954; this included a presumed signal sequence having a molecular weight of 1,953.8 (Fig. 7) (see below) and therefore was in very good agreement with the reported value of 68,000 (3) and the value determined by maxicell analysis (see above). The initiation codon shown in Fig. 6 (at position +1) is 6 base pairs downstream from another in-frame ATG codon at position −9. However, we assumed that the protein begins at position +1 since it correlated well with a Shine-Dalgarno sequence 8 base pairs upstream, which is within the usual range of 5 to 9 base pairs (40); a suitable Shine-Dalgarno sequence was not observed in the vicinity of the ATG codon at position −9. Stormo et al. (40) observed a nonrandom distribution of bases around the initiation codon in a statistical analysis of 124 genes. It is notable that in the case of the *cpdB* sequence there was a reasonable correlation with the predictions of this analysis, with agreement at 8 of 13 positions (Table 1). Although this correlation is consistent with the presumed functional significance of these bases (40), it is clearly not sufficient to distinguish translational start sites.

Two putative promoters (pI and pII) were identified on the basis of similarity to consensus −10 and −35 sequences (18, 32) and are indicated on Fig. 6. Interestingly, the −10 region of pI contains two discernible overlapping −10 sequences.

Two overlapping palindromic sequences, one of which may serve as a transcriptional terminator, are indicated in the 3′ untranslated sequence in Fig. 6.

As a secreted protein, cyclic phosphodiesterase would be expected to contain a signal sequence which is cleaved during export (31, 35). The predicted N-terminal amino acid sequence (Fig. 6) does in fact display the main features (21, 35, 48) of a procaryotic signal sequence. A charged residue (lysine) occurs within the first five residues, followed by a

strongly hydrophobic sequence (Fig. 7) (see below). The most likely cleavage site (see below) to yield the mature protein is indicated in Fig. 6.

Codon usage in the *cpdB* gene is summarized in Fig. 8. The fractional use of synonymous codons which are optimal for translation (20) is 0.59. This value is notably low compared with the values for 50 other *E. coli* genes of varying expressivity (20), indicating that the *cpdB* gene is weakly expressed.

Finally, the sequence analysis revealed that the deletion into the vector moiety of pDB4 (and its derivatives), which occurred during its isolation (see above), extends to nucleotide 773 in the pBR322 sequence (41); the sequence in Fig. 6 extends up to this nucleotide.

DISCUSSION

Of the five recombinant DNA libraries used in the attempt to isolate the *cpdB* gene, in only one case did the enzyme used (*Sal*I) cleave the gene (Fig. 1). The absence of the *cpdB* gene in libraries constructed by using the other restriction enzymes, along with the isolation of a clone from which a deletion of adjacent DNA had occurred, strongly suggested that the deleted DNA is lethal when it is cloned into a multicopy plasmid. Thus, it seems possible that an unknown gene product is detrimental to the growth of *E. coli* when it is present in excess, a situation which is not without precedent (39). It may be reasonably predicted that the *cpdB* gene can be simply isolated from a gene library constructed from overlapping DNA fragments, such as those generated by partial *Sau*3A digestion, since some *Sau*3A fragments should contain a intact *cpdB* gene plus a fragmented (lethal) adjacent gene. Using such a library (10), we isolated the *cpdB* gene (Burns and Beacham, unpublished data). However, once again a complex rearrangement in the hybrid plasmid had occurred.

A second explanation may be that an increased level of 2′,3′-cyclic phosphodiesterase is itself detrimental and that the deletion in pDB4 extends into the 3′ end of the *cpdB* gene. However, the nucleotide sequence analysis showed that such a deletion does not extend into the coding region, but it remains possible that the normal transcriptional terminator was deleted, resulting in a more labile mRNA. In any case, neither this activity nor the associated 3′-nucleotidase activity is likely to be inhibitory, particularly since the enzyme is secreted. However, it should be noted that the presence of a cytoplasmic inhibitor (17, 27) of UDP-sugar

A -136   -121   -106   -91   -76
TTTACTTATACCCTA TCGTTAATGAATGCG CCAACTGTGATAGTG TCATCATTTTCAAAG CGTAAAATTGTGGCA
                                                                    -35(pII)

         -61   -46   -31   -16   SD   -1
TTCTTCACTGTTCTA TAAGTAAGACGTTTA TTCTCCCTTTTCTTT CGTATTCCCGATGAT AAAAGGATGTCCCTG
    -35(pI)   -10(pII)   -10(pI)

                    15   30   45   ↓ 60
ATG ATT AAG TTT AGC GCA ACG CTC CTG GCC ACG CTG ATT GCC GCC AGT GTG AAT GCA▼GCG
MET Ile Lys Phe Ser Ala Thr Leu Leu Ala Thr Leu Ile Ala Ala Ser Val Asn Ala Ala

                    75   90   105   120
ACG GTC GAT CTG CGT ATC ATG GAA ACC ACT GAT CTG CAT AGC AAC ATG ATG GAT TTC GAT
Thr Val Asp Leu Arg Ile MET Glu Thr Thr Asp Leu His Ser Asn MET MET Asp Phe Asp

                    135   150   165   180
TAT TAC AAA GAC ACC GCC ACG GAA AAA TTC GGA CTG GTA CGT ACG GCA AGC CTG ATT AAC
Tyr Tyr Lys Asp Thr Ala Thr Glu Lys Phe Gly Leu Val Arg Thr Ala Ser Leu Ile Asn

                    195   210   225   240
GAT GCC CGC AAT GAA GTG AAA AAC AGC GTA CTG GTT GAT AAC GGC GAT TTG ATT CAG GGG
Asp Ala Arg Asn Glu Val Lys Asn Ser Val Leu Val Asp Asn Gly Asp Leu Ile Gln Gly

                    255   270   285   300
AGT CCG CTG GCC GAT TAC ATG TCG GCG AAA GGA TTA AAA GCA GGT GAT ATT CAC CCG GTC
Ser Pro Leu Ala Asp Tyr MET Ser Ala Lys Gly Leu Lys Ala Gly Asp Ile His Pro Val

                    315   330   345   360
TAT AAG GCA TTA AAT ACG CTG GAC TAT ACC GTC GGA ACG CTT GGC AAC CAC GAG TTT AAC
Tyr Lys Ala Leu Asn Thr Leu Asp Tyr Thr Val Gly Thr Leu Gly Asn His Glu Phe Asn

                    375   390   405   420
TAC GGT CTG GAT TAC CTG AAA AAT GCG CTG GCA GGA GCG AAA TTC CCT TAT GTA AAT GCC
Tyr Gly Leu Asp Tyr Leu Lys Asn Ala Leu Ala Gly Ala Lys Phe Pro Tyr Val Asn Ala

                    435   450   465   480
AAC GTC ATT GAC GCC AGA ACC AAA CAG CCA ATG TTT ACA CCG TAT TTA ATT AAA GAC ACC
Asn Val Ile Asp Ala Arg Thr Lys Gln Pro MET Phe Thr Pro Tyr Leu Ile Lys Asp Thr

                    495   510   525   540
GAA GTG GTC GAT AAA GAC GGA AAA AAA CAG ACG CTG AAG ATT GGC TAT ATT GGC GTC GTG
Glu Val Val Asp Lys Asp Gly Lys Lys Gln Thr Leu Lys Ile Gly Tyr Ile Gly Val Val

                    555   570   585   600
CCA CCA CAA ATC ATG GGC TGG GAT AAA GCT AAT TTA TCC GGG AAA GTG ACG GTG AAT GAT
Pro Pro Gln Ile MET Gly Trp Asp Lys Ala Asn Leu Ser Gly Lys Val Thr Val Asn Asp

                    615   630   645   660
ATT ACC GAA ACC GTG CGC AAA TAC GTG CCT GAA ATG CGC GAG AAA GGT GCC GAT GTT GTT
Ile Thr Glu Thr Val Arg Lys Tyr Val Pro Glu MET Arg Glu Lys Gly Ala Asp Val Val

                    675   690   705   720
GTC GTT CTG GCA CAT TCC GGG CTA TCT GCC GAT CCG TAT AAA GTG ATG GCG GAA AAC TCA
Val Val Leu Ala His Ser Gly Leu Ser Ala Asp Pro Tyr Lys Val MET Ala Glu Asn Ser

                    735   750   765   780
GTT TAT TAC CTC AGT GAA ATT CCG GGC GTT AAC GCC ATT ATG TTT GGC CAT GCT CAC GCC
Val Tyr Tyr Leu Ser Glu Ile Pro Gly Val Asn Ala Ile MET Phe Gly His Ala His Ala

                    795   810   825   840
GTT TTC CCA GGT AAA GAT TTT GCT GAT ATC GAA GGG GCT GAT ATC GCC AAA GGC ACG CTG
Val Phe Pro Gly Lys Asp Phe Ala Asp Ile Glu Gly Ala Asp Ile Ala Lys Gly Thr Leu

                    855   870   885   900
AAT GGT GTT CCG GCG GTA ATG CCA GGC ATG TGG GGC GAT CAT CTT GGT GTG GTC GAC TTA
Asn Gly Val Pro Ala Val MET Pro Gly MET Trp Gly Asp His Leu Gly Val Val Asp Leu

                    915   930   945   960
CAA CTC AGT AAT GAC AGC GGT AAA TGG CAG GTG ACG CAG GCG AAA GGC GAA GCA CGA CCG
Gln Leu Ser Asn Asp Ser Gly Lys Trp Gln Val Thr Gln Ala Lys Gly Glu Ala Arg Pro

                    975   990   1005   1020
ATT TAC GAC ATC GCT AAT AAA AAA TCC CTC GCG GCG GAA GAC AGC AAG CTG GTA GAA ACA
Ile Tyr Asp Ile Ala Asn Lys Lys Ser Leu Ala Ala Glu Asp Ser Lys Leu Val Glu Thr

FIG. 6. Nucleotide sequence of the *cpdB* gene and predicted amino acid sequence. The vertical arrow indicates the most likely cleavage site of the signal sequence; this site is designated C_1 in Fig. 7. Two putative promoters are underlined and labeled pI and pII. The Shine-Dalgarno sequence (SD) is also indicated. Inverted repeat sequences, which may function as transcriptional terminators, are indicated in the 3' untranslated region.

B
```
          1035                1050                1065                1080
CTC AAA GCC GAT CAC GAT GCC ACA CGC CAG TTC GTC AGC AAG CCA ATC GGT AAA TCT GCC
Leu Lys Ala Asp His Asp Ala Thr Arg Gln Phe Val Ser Lys Pro Ile Gly Lys Ser Ala

          1095                1110                1125                1140
GAC AAT ATG TAT AGC TAT CTG GCG CTG GTG CAG GAC GAT CCG ACC GTG CAG GTG GTG AAC
Asp Asn MET Tyr Ser Tyr Leu Ala Leu Val Gln Asp Asp Pro Thr Val Gln Val Val Asn

          1155                1170                1185                1200
AAC GCG CAA AAA GCG TAT GTC GAG CAT TAC ATT CAG GGC GAT CCG GAT CTG GCA AAA CTG
Asn Ala Gln Lys Ala Tyr Val Glu His Tyr Ile Gln Gly Asp Pro Asp Leu Ala Lys Leu

          1215                1230                1245                1260
CCG GTG CTT TCA GCT GCC GCA CCG TTT AAA GTC GGT GGT CGC AAA AAT GAC CCG GCA AGC
Pro Val Leu Ser Ala Ala Ala Pro Phe Lys Val Gly Gly Arg Lys Asn Asp Pro Ala Ser

          1275                1290                1305                1320
TAT GTG GAG GTG GAA AAA GGC CAG TTG ACC TTC CGT AAT GCC GCC GAT CTT TAT CTC TAT
Tyr Val Glu Val Glu Lys Gly Gln Leu Thr Phe Arg Asn Ala Ala Asp Leu Tyr Leu Tyr

          1335                1350                1365                1380
CCC AAT ACG CTG ATT GTG GTG AAA GCC AGC GGT AAA GAG GTG AAA GAG TGG CTG GAG TGC
Pro Asn Thr Leu Ile Val Val Lys Ala Ser Gly Lys Glu Val Lys Glu Trp Leu Glu Cys

          1395                1410                1425                1440
TCC GCG GGA CAG TTT AAC CAG ATT GAT CCC AAC AGC ACG AAA CCA CAG TCA CTC ATC AAC
Ser Ala Gly Gln Phe Asn Gln Ile Asp Pro Asn Ser Thr Lys Pro Gln Ser Leu Ile Asn

          1455                1470                1485                1500
TGG GAT GGT TTC CGC ACT TAT AAC TTT GAT GTT ATT GAT GGT GTG AAT TAT CAG ATT GAT
Trp Asp Gly Phe Arg Thr Tyr Asn Phe Asp Val Ile Asp Gly Val Asn Tyr Gln Ile Asp

          1515                1530                1545                1560
GTT ACC CAG CCC GCC CGT TAT GAC GGC GAG TGC CAG ATG ATT AAT GCC AAT GCG GAA AGG
Val Thr Gln Pro Ala Arg Tyr Asp Gly Glu Cys Gln MET Ile Asn Ala Asn Ala Glu Arg

          1575                1590                1605                1620
ATT AAG AAC CTG ACC TTT AAT GCA AGC CGA TTG ATC CGA ACG CCA TGT TCC TGG TTG CCA
Ile Lys Asn Leu Thr Phe Asn Ala Ser Arg Leu Ile Arg Thr Pro Cys Ser Trp Leu Pro

          1635                1650                1665                1680
CCA ATA ACT ATC GCG CTT ACG GGG CAA ATT TGC GGT ACG GGC GAC AGC CAT ATC GCT TTT
Pro Ile Thr Ile Ala Leu Thr Gly Gln Ile Cys Gly Thr Gly Asp Ser His Ile Ala Phe

          1695                1710                1725                1740
GCT TCA CCG GAT GAG AAC CGC TCG GTG CTG GCA GCG TGG ATT GCT GAT GAG TCG AAA CGT
Ala Ser Pro Asp Glu Asn Arg Ser Val Leu Ala Ala Trp Ile Ala Asp Glu Ser Lys Arg

          1755                1770                1785                1800
GCG GGG GAA ATT CAC CCG GCG GCA GAT AAC AAC TGG CGT TTA GCA CCG ATA GCT GGC GAT
Ala Gly Glu Ile His Pro Ala Ala Asp Asn Asn Trp Arg Leu Ala Pro Ile Ala Gly Asp

          1815                1830                1845                1860
AAG AAA CTG GAT ATC CGT TTC GAA ACC TCT CCG TCA GAT AAA GCC GCA GCG TTT ATT AAA
Lys Lys Leu Asp Ile Arg Phe Glu Thr Ser Pro Ser Asp Lys Ala Ala Ala Phe Ile Lys

          1875                1890                1905                1920
GAG AAA GGG CAG TAT CCG ATG AAT AAA GTC GCG ACC GAT GAT ATC GGG TTT GCG ATT TAT
Glu Lys Gly Gln Tyr Pro MET Asn Lys Val Ala Thr Asp Asp Ile Gly Phe Ala Ile Tyr

          1935
CAG GTG GAT TTG AGT AAG TAA
Gln Val Asp Leu Ser Lys
```

```
        1956            1971      L____>  1986    <____I 2001            2016
AACACTTCTTTTTCG GCCTATAAATCATCA ACCGCATCCGGCATT TATTGGATGCATTGG CGCATCTTATCCGCC
                                                   r____>          <____1

      2031            2046
CTACAAGCCATGCAC CGTAGACCAGATAAG CT
```

hydrolase (5'-nucleotidase) may imply that at least small amounts of secreted enzymes can be located in the cytoplasm (42), which is consistent with posttranslational models for secretion in *E. coli* (16, 31).

Finally, a strong promoter at the 3' end of the *cpdB* gene

in the original, nondeleted clones might have allowed transcription into the *ori* region of the plasmid and interfered with replication.

The identification of the 67K protein which is encoded by pDB8 (Fig. 3, lane e) as cyclic phosphodiesterase is based on

the secretion of this protein (Fig. 4) and on a correlation between loss of enzyme activity and loss of the 67K protein following effective deletion of one end of the gene during the construction of pDB6 (Fig. 3, lane d). Furthermore, the molecular weight deduced from the maxicell analysis is close to the reported value (68,000) for the purified enzyme and the calculated value derived from the sequence analysis (see above). pDB8 also encodes a 41.6K protein; however, there is insufficient coding capacity within the inserted DNA of pDB8 to specify both proteins. The origin of the 41.6K protein is not known, but our results are reminiscent of the results of other studies in which more than one protein was seemingly derived from one sequence (8, 14). The reading frame for the 41.6K protein must overlap or lie within the reading frame encoding the 67K protein. Alternatively, the 41.6K protein may be derived posttranslationally from cyclic phosphodiesterase.

The nucleotide sequence reveals putative promoters and terminators which merit further study to determine which one is functional in vivo or whether both are functional. Promotor pII is perhaps less favored since there are only four documented instances of promoters without a thymine residue at the 3' end of the position −10 hexamer (18). If promoter pI is functional, then the overlapping −10 sequences alone may lead to heterogeneity in the 5' ends of the resulting mRNA.

The best terminator, which is centered at nucleotide positions 1,985 and 1,986, forms a potential stem-and-loop structure that is typical of rho-independent terminators (7, 29, 32) with a free energy of −10.2 kcal (42.68 kJ), as calculated by the method of Tinoco et al. (43); this is on the low side of the usual range (7). In many cases the stem-and-loop structure is followed by consecutive thymine residues, but this feature is absent in this instance, in common with several other E. coli genes (7, 25, 28, 37). Both of these features suggest that experimental determination of the transcriptional termination site would be of interest. It may also be noted that a third feature of rho-independent terminators, a downstream TCGT consensus sequence, has been proposed (7); two sequences which closely resemble this sequence are present at nucleotide positions 2,006 through 2,009 and 2,011 through 2,014.

A study of 20 procaryotic signal sequences resulted in the formulation of a set of preferred amino acids around the cleavage site (46, 47). The cpdB sequence contains two potential cleavage sites, C_1 and C_2 (Fig. 7), which are consistent with these rules; in particular, both sites have the virtually invariant Ala residue at position −1 relative to the cleavage site, and both have the preferred Ala at position +1. However, C_2 is a less likely cleavage site since (i) the signal sequence would then be unusually short compared with other E. coli signal sequences (48); (ii) there would not be a charged residue between positions +1 and +5 (relative to the cleavage site), which is usual in procaryotes (47), in

Met Ile Lys/Phe Ser Ala Thr Leu Leu Ala Thr Leu

```
    C2                  C1
    ↓                   ↓
        -3  -2  -1↓ +1  +2  +3  +4  +5
    Ile Ala Ala Ser Val Asn Ala Ala Thr Val Asp Leu
```

FIG. 7. Signal sequence of the cpdB gene. The start of the hydrophobic core is indicated by a slash, and two possible cleavage sites are indicated as C_1 and C_2. Residues are numbered with respect to C_1. The molecular weight of the sequence of 19 residues up to C_1 is 1,953.8.

Phe	UUU	12	Ser	UCU	3	Tyr	UAU	19	Cys	UGU	1
Phe	UUC	8	Ser	UCC	5	Tyr	UAC	8	Cys	UGC	3
Leu	UUA	6	Ser	UCA	5	*	UAA	1	*	UGA	0
Leu	UUG	5	Ser	UCG	3	*	UAG	0	Trp	UGG	8
Leu	CUU	5	Pro	CCU	2	His	CAU	6	Arg	CGU	7
Leu	CUC	7	Pro	CCC	3	His	CAC	5	Arg	CGC	7
Leu	CUA	1	Pro	CCA	10	Gln	CAA	4	Arg	CGA	3
Leu	CUG	25	Pro	CCG	17	Gln	CAG	18	Arg	CGG	0
Ile	AUU	25	Thr	ACU	3	Asn	AAU	19	Ser	AGU	5
Ile	AUC	12	Thr	ACC	13	Asn	AAC	19	Ser	AGC	13
Ile	AUA	2	Thr	ACA	3	Lys	AAA	40	Arg	AGA	1
Met	AUG	15	Thr	ACG	16	Lys	AAG	8	Arg	AGG	1
Val	GUU	10	Ala	GCU	10	Asp	GAU	39	Gly	GGU	14
Val	GUC	12	Ala	GCC	23	Asp	GAC	14	Gly	GGC	16
Val	GUA	5	Ala	GCA	16	Glu	GAA	16	Gly	GGA	6
Val	GUG	24	Ala	GCG	21	Glu	GAG	11	Gly	GGG	8

FIG. 8. Codon usage in the E. coli cpdB gene. *, Termination codon.

contrast to C_1, which has an Asp residue at position +4; and (iii) although Leu is tolerated at position −3, Val is highly preferred (47). Therefore we suggest that C_1 is the likely cleavage site. On this basis the mean hydrophobicity per residue was calculated (45) to be −3.25 kJ/mol, which compares favorably with the average of −3.9 kJ/mol for 10 other procaryotic signal sequences (45).

It has been proposed that the charged N termini of signal sequences serve a defined role in the secretory process (21, 22, 44). This hypothesis has been examined in the case of the exported lpp gene product by using site-directed mutagenesis of the two N-terminal lysine residues (22, 44). Since 2',3'-cyclic phosphodiesterase contains only a single N-terminal charged residue, this might prove to be a simple and hence fruitful system with which to pursue similar studies.

2',3'-Cyclic phosphodiesterase is a bifunctional enzyme which catalyzes two consecutive reactions (see above). Kinetic studies on the purified enzyme (2) have indicated that two active sites can be distinguished, which correspond to the two substrates (2',3'-cyclic nucleotides and 3'-nucleotides). Hence, the elucidation of the precise mechanism of action of this enzyme is of considerable interest. A comparison with the extensive studies on microbial RNases (19) would be of additional interest since 2',3'-cyclic nucleotides are intermediates in the breakdown of RNA and also serve as substrates. The characterization and sequencing of the cpdB gene allow a genetic approach to the enzymology of this enzyme by using in vitro and in vivo mutagenesis (15). Presumably, the enzyme can also be excessively produced to a greater extent than reported here, which should facilitate purification and investigation at the protein level. High-level production of several proteins has been achieved by cloning in M13, presumably owing to the high copy number of the replicative form in infected cells (36). Preliminary attempts to improve the yield of 2',3'-cyclic phosphodiesterase obtained with plasmid vectors by cloning the gene in M13 mp8 were not successful (data not shown). This may have been due to regulation of the gene, in which case further genetic manipulation of pDB8 will be necessary.

ACKNOWLEDGMENTS

LITERATURE CITED

1. **Anraku, Y.** 1964. A new cyclic phosphodiesterase having 3'-nucleotidase activity from *Escherichia coli* B. I. Purification and some properties of the enzyme. J. Biol. Chem. **239:**3412–3419.
2. **Anraku, Y.** 1964. A new cyclic phosphodiesterase having a 3'-nucleotidase activity from *Escherichia coli* B. II. Further studies on substrate specificity and mode of action of the enzyme. J. Biol. Chem. **239:**3420–3424.
3. **Anraku, Y.** 1966. Cyclic phosphodiesterase of *Escherichia coli*. Prog. Nucleic Acid Res. **1:**130–143.
4. **Beacham, I. R.** 1979. Periplasmic enzymes in gram negative bacteria. Int. J. Biochem. **10:**877–883.
5. **Beacham, I. R., and S. Garrett.** 1980. Isolation of *Escherichia coli* mutants (*cpdB*) deficient in periplasmic 2',3'-cyclic phosphodiesterase and genetic mapping of the *cpdB* locus. J. Gen. Microbiol. **119:**31–34.
6. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
7. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
8. **Burns, D. M., L. J. Abraham, and I. R. Beacham.** 1983. Characterisation of the *ush* gene of *Escherichia coli*, and its protein products. Gene **25:**343–353.
9. **Burns, D. M., and I. R. Beacham.** 1983. A method for the ligation of DNA following isolation from low melting temperature agarose. Anal. Biochem. **135:**48–51.
10. **Burns, D. M., and I. R. Beacham.** 1984. Positive selection vectors: a small plasmid vector useful for the direct selection of *Sau*3A-generated overlapping DNA fragments. Gene **27:**323–325.
11. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColEI hybrid plasmids representative of the entire *E. coli* genome. Cell **9:**91–99.
12. **Covarrubias, L., L. Cervantes, A. Covarrubias, X. Soberon, I. Vichido, H. Blanco, Y. M. Kupersztoch-Portnoy, and F. Bolivar.** 1981. Construction and characterisation of new cloning vehicles. V. Mobilisation and coding properties of pBR322 and several deletion derivatives including pBR327 and pBR328. Gene **13:**25–35.
13. **Cowman, A., and I. R. Beacham.** 1980. Molecular cloning of the gene (*ush*) from *Escherichia coli* specifying periplasmic UDP-sugar hydrolase (5'-nucleotidase). Gene **12:**281–286.
14. **Crowlesmith, I., and K. Gamon.** 1982. Rate of translation and kinetics of processing of newly synthesized molecules of two major outer membrane proteins, the OmpA and OmpF proteins of *Escherichia coli* K12. Eur. J. Biochem. **124:**577–583.
15. **Dalbadie-McFarland, G., L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura, and J. H. Richards.** 1982. Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. Proc. Natl. Acad. Sci. USA **79:**6409–6413.
16. **Garwin, J. L., and J. Beckwith.** 1982. Secretion and processing of ribose-binding protein in *Escherichia coli*. J. Bacteriol. **149:**789–792.
17. **Glaser, L., A. Melo, and R. Paul.** 1967. Uridine diphosphate sugar hydrolase. Purification of enzyme and protein inhibitor. J. Biol. Chem. **247:**1944–1954.
18. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2253.
19. **Hill, C., G. Dodson, U. Heinemann, W. Saenger, K. Nakamura, S. Borisov, G. Tischenko, K. Polyakov, and S. Pavlovsky.** 1983. The structure and sequence homology of a family of microbial ribonucleases. Trends Biochem. Sci. **8:**364–369.
20. **Ikemura, T., and H. Ozeki.** 1982. Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. Cold Spring Harbor Symp. Quant. Biol. **47:**1087–1097.
21. **Inouye, M., and S. Halegoua.** 1980. Secretion and membrane localization of proteins in *Escherichia coli*. Crit. Rev. Biochem.

7:339–371.
22. **Inouye, S., X. Soberon, T. Franceschini, K. Nakamura, T. Itakura, and M. Inouye.** 1982. Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. Proc. Natl. Acad. Sci. USA **79:**3438–3441.
23. **Legerski, R. J., J. L. Hodrett, and H. B. Gray.** 1978. Extracellular nucleases of pseudomonas BAL31. III. Use of the double-strand-deoxyriboexonuclease activity as the basis of a convenient method for mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. **5:**1445–1464.
24. **Low, K. B.** 1972. *Escherichia coli* K12 F-prime factors, old and new. Bacteriol. Rev. **36:**587–607.
25. **Magota, K., N. Otsuji, T. Miki, T. Horiuchi, S. Tsunasawa, J. Kondo, F. Sakiyama, M. Amemura, T. Morita, H. Shinagawa, and A. Nakata.** 1984. Nucleotide sequence of the *phoS* gene, the structural gene for the phosphate binding protein of *Escherichia coli*. J. Bacteriol. **157:**909–917.
26. **Messing, J.** 1981. New M13 vectors for cloning. Methods Enzymol. **101:**20–78.
27. **Neu, H. C.** 1967. The 5'-nucleotidase of *Escherichia coli*. II. Surface localisation and purification of the *Escherichia coli* 5'-nucleotidase inhibitor. J. Biol. Chem. **242:**3905–3911.
28. **Ohmori, H., M. Kimura, T. Nagata, and Y. Sakakibara.** 1984. Structural analysis of the dnaA and dnaN genes of *Escherichia coli*. Gene **28:**159–170.
29. **Platt, T.** 1981. Termination of transcription and its regulation in the tryptophan operon of *E. coli*. Cell **24:**10–23.
30. **Prentki, P., and H. M. Krisch.** 1982. A modified pBR322 vector with improved properties for the cloning, recovery and sequencing of blunt-ended DNA fragments. Gene **17:**189–196.
31. **Randall, L. L., and S. J. S. Hardy.** 1984. Export of protein in bacteria. Microbiol. Rev. **48:**290–298.
32. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13:**319–353.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
34. **Sherman, F., G. R. Fink, and C. W. Lawrance (ed.).** 1979. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Silhavy, T. J., S. A. Benson, and S. D. Emr.** 1983. Mechanisms of protein localization. Microbiol. Rev. **47:**313–344.
36. **Slocombe, P., A. Easton, P. Boseley, and D. C. Burke.** 1982. High-level expression of an interferon 22 gene cloned in phage M13 mp7 and subsequent purification with a monoclonal antibody. Proc. Natl. Acad. Sci. USA **79:**5455–5459.
37. **Smith, D. R., and J. M. Calvo.** 1980. Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. Nucleic Acids Res. **8:**2255–2274.
38. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
39. **Stoker, N. G., N. F. Fairweather, and B. G. Spratt.** 1982. Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. Gene **18:**335–341.
40. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10:**2971–2996.
41. **Sutcliffe, J. G.** 1979. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43:**77–90.
42. **Taylor, N., and I. R. Beacham.** 1975. Synthesis and localisation of *Escherichia coli* UDP-glucose hydrolase (5'-nucleotidase), and demonstration of a cytoplasmic inhibitor of this enzyme, in *Salmonella typhimurium*. Biochim. Biophys. Acta **411:**216–221.
43. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1983. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New. Biol. **246:**40–41.

44. **Vlasuk, G. P., S. Inouye, H. Ito, K. Itakura, and M. Inouye.** 1983. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in *Escherichia coli*. J. Biol. Chem. **258:**7141–7148.

45. **von Heijne, G.** 1981. On the hydrophobic nature of signal sequences. Eur. J. Biochem. **116:**419–422.

46. **von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133:**17–21.

47. **von Heijne, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173:**243–251.

48. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12:**5145–5164.