Analysis of the *Escherichia coli* Gene Encoding L-Asparaginase II, *ansB*, and Its Regulation by Cyclic AMP Receptor and FNR Proteins

MICHAEL P. JENNINGS AND IFOR R. BEACHAM*

Division of Science and Technology, Griffith University, Nathan, Brisbane, Queensland 4111, Australia

Escherichia coli contains two L-asparaginase isozymes: L-asparaginase I, a low-affinity enzyme located in the cytoplasm, and L-asparaginase II, a high-affinity secreted enzyme. A molecular genetic analysis of the gene (*ansA*) encoding the former enzyme has previously been reported. We now present a molecular study of the gene, *ansB*, encoding L-asparaginase II. This gene was isolated by using oligonucleotide probes, whose sequences were based on the previously determined amino acid sequence. The nucleotide sequence of *ansB*, including 5'- and 3'-untranslated regions, was determined. The amino acid sequence of L-asparaginase II, deduced from this nucleotide sequence, contains differences at 11 positions when compared with the previously determined amino acid sequence. The deduced amino acid sequence also reveals a typical secretory signal peptide of 22 residues. A single region of sequence similarity is observed when *ansA* and *ansB* are compared. The transcriptional start site in *ansB* was determined, allowing the identification of the promoter region. The regulation of *ansB* was studied by using *ansB'-'lacZ* fusions, together with a deletion analysis of the 5' region upstream of the promoter. Regulation by cyclic AMP receptor protein and anerobiosis (FNR protein) was confirmed, and the presence of nucleotide sequence motifs, with homology to cyclic AMP receptor protein and FNR protein-binding sites, investigated.

L-Asparaginase has been found to be effective in the therapy of certain lymphomas and leukemias in both experimental animals and humans and has been used, in combination with other agents, in the treatment of acute lymphoblastic leukemia for some time (2, 10, 37). Only enzymes with a K_m of the order of 10^{-5} M are effective (48).

There has consequently been a great deal of interest in microorganisms, and particularly bacteria, as a source of high-affinity asparaginases and in the detailed structures of these enzymes. *Escherichia coli* produces two L-asparaginases with markedly different properties. One enzyme, L-asparaginase I, has a low affinity for L-asparagine (K_m, 3.5 × 10^{-3} M), is cytoplasmic, and is thought to be constitutively produced (19, 50). L-Asparaginase II, by contrast, is a high-affinity enzyme (K_m, 1.15 × 10^{-5} M) and is secreted to the periplasm, and its expression is positively regulated (6, 8, 13, 20, 35). It is one of a number of bacterial enzymes used in the treatment of acute lymphoblastic leukemia.

Molecular genetic studies have recently resulted in the isolation of the gene (*ansA*) encoding L-asparaginase I and hence the primary structure of this enzyme (19). The amino acid sequence of L-asparaginase II has been previously determined (24). A comparison of these two sequences confirms indirect evidence (40) that they are highly dissimilar; a single region of significant homology, however, is consistent with their likely origin via an ancient gene duplication (19).

It has previously been established that the synthesis of L-asparaginase II is subject to regulation by cyclic AMP receptor protein (CRP) and is also induced by anaerobiosis (6, 8, 35). The latter form of regulation involves the *fnr* gene product (FNR protein) (20), which also activates a number of other anaerobically regulated genes (18, 39).

In this report, we present the isolation and molecular analysis of the gene (*ansB*) encoding L-asparaginase II and hence the primary sequence of this enzyme, which differs from that previously determined. We also present a study of the transcription and regulation of *ansB*.

MATERIALS AND METHODS

Materials. The following radioactively labeled nucleotides were from Amersham Corp.: [γ-³²P]ATP, 3,000 Ci/mmol; [α-³²P]dCTP, 800 Ci/mmol; [α-³⁵S]thiodeoxyATP, 1000 Ci/mmol. Restriction enzymes were obtained from Amersham Corp., New England BioLabs, and Toyobo Chemicals. T7 DNA polymerase (Sequenase) was from the U.S. Biochemical Corp., reverse transcriptase was from Promega Biotec; and polynucleotide kinase and T4 DNA ligase were from New England BioLabs.

Bacteria, bacteriophages, and plasmids. The following strains of *E. coli* K-12 were used: JM103 [Δ(*lac pro*) *thi strA supE endA sbcB hsdR*(F' *traD36 proAB lacIᵍZΔM15*); 25]; MC4100 [F⁻ *araD139* Δ(*argF-lac*)*U169 rpsL150 relA1 flb5301 deoC1 ptsF25 rbsR*]. λRZ5 (λ '*bla 'lacZ lacY⁺*) and MC4100 were kindly supplied by R. P. Gunsalus (21). The plasmid pUC118 (45) was used as a cloning vector. The *lacZ* fusion vector pNM481 was used for *ansB'-'lacZ* fusion construction (28).

Media and growth conditions. Generally, cells were grown at 37°C in LB medium (25). When used, glucose was supplemented to 0.4%. In general, ampicillin was used at a final concentration of 40 μg/ml. Aerobic growth was achieved by gyratory shaking of 5-ml cultures at 300 rpm. Anaerobic cultures were grown in test tubes in an anaerobic jar (BBL Microbiology Systems). Due to an observed instability of the fusion plasmid pMJF1 and subsequent strains (lysogens), assay cultures were inoculated from starter cultures and grown through at least five generations in a higher concentration of ampicillin (500 μg/ml), to maintain selection. The proportion of Lac⁺ Apʳ cells in cultures used for

---

* Corresponding author.

the assays of β-galactosidase was monitored by plating onto MacConkey-lactose agar with and without ampicillin. Medium containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside was used for the selection of recombinant plasmids. In later experiments MacConkey-lactose agar (Oxoid CM7 containing ampicillin [40 μg/ml]) was used instead. Cells were mixed with 10 μl of 100 mM isopropylthiogalactoside before plating. This medium was also used to select λ lysogens.

Southern hybridization. Chromosomal DNA (1 to 2 μg) was digested to completion with suitable restriction endonucleases, electrophoresed on a 0.7% agarose gel, and then transferred to nitrocellulose as previously described (25, 40). The membranes were prehybridized for 4 h at 37°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% sodium dodecyl sulfate, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 0.1% Ficoll (M_r, 400,000), and 100 μg of tRNA per ml. The ^{32}P-labeled DNA probe was added to the prehybridization mix, and hybridization was allowed to proceed overnight. The filters were washed three times in 0.1× SSC, wrapped in plastic film, and autoradiographed. The temperatures for prehybridization and hybridization were 37°C for oligonucleotide probes and 65°C when a restriction fragment, labeled with ^{32}P by nick translation, was used as a probe.

Library construction and colony hybridization. Enriched DNA libraries were prepared as follows: DNA (approximately 20 μg) was digested with an appropriate restriction enzyme and electrophoresed through a low-melting-temperature agarose gel, and fragments in the desired size range were excised and recovered (4). This DNA was cloned into pUC118 and transformed into JM103. Approximately 500 colonies containing recombinant plasmids were identified on medium containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside and screened by colony hybridization (25), with the same conditions as those used for Southern hybridization.

Recombinant DNA techniques. Oligonucleotides were synthesized with an Applied Biosystems synthesiser (model 380B). The DNA, in NH₄OH, was deprotected by incubation at 55°C overnight, the NH₄OH was removed under vacuum, and the DNA was taken up in distilled water. The DNA was then ethanol precipitated, washed twice with 80% ethanol, dried, and suspended in distilled water. For use as a probe 25 pmol was labeled with 10 U of polynucleotide kinase and 90 μCi of [γ-^{32}P]ATP, and unincorporated precursors removed by using a G25 Sephadex spun column (25). Techniques other than those described below were as described previously (3, 19, 25, 40).

Nucleotide sequence analysis. Suitable DNA fragments were isolated, after agarose (0.7%) or polyacrylamide (4.5%) gel electrophoresis, and subcloned into M13mp18 or M13mp19 (26). Single-stranded DNA was sequenced by the dideoxy-chain termination method (36) with T7 DNA polymerase (42), [α-^{35}S]thiodeoxyATP, and a universal primer; the oligonucleotide probes LP1 and LP2 (Fig. 1) were also used as primers (1.5 pmol per annealing). Sequencing reactions were electrophoresed on 6% denaturing polyacrylamide gels, which were cast as wedge gels with a thickness gradient of 0.4 mm (top) to 1.2 mm (bottom) (7).

RNA transcript mapping. The transcription start site was determined by using primer extension in conjunction with dideoxy sequencing (15). Total RNA was prepared by the method of Aiba et al. (1). The ^{32}P-labeled probe was prepared by using template DNA obtained from the cloned EcoRI-HindIII fragment (from pMJF1) in M13mp18; ^{32}P was incorporated from [α-^{32}P]dATP by using a sequencing primer and T7 DNA polymerase. The product was digested

with DdeI (which cleaves between nucleotide positions −41 and −42; see Fig. 3), yielding a labeled 121-base-pair (bp) single-stranded probe, which was recovered from a denaturing polyacrylamide gel. This probe was hybridized with total RNA, extended with reverse transcriptase, and electrophoresed, alongside dideoxy sequencing reactions, on a 6% denaturing polyacrylamide gel. The template for the sequencing reactions was the same as that used to generate the probe. In initial experiments, bands were observed at unreasonably high molecular weight positions in the gel; these bands were shown to be an artifact of contamination of the RNA with pMJF1 DNA (data not shown). Such bands did not occur when the RNA was treated with RNase-free DNase (RQ1 from Promega Biotec) before use in these experiments.

Construction of ansB'-'lacZ fusions. A 226-bp BstNI-Fnu4HI fragment, encompassing 224 bp of the region upstream of the ansB initiation codon, including 19 bp of coding sequence from the putative open reading frame ORF1, plus 9 codons of the ansB-coding sequence (nucleotide positions −195 to +29; see Fig. 6), was isolated from a polyacrylamide gel and made blunt ended by treatment with mung bean nuclease. The resulting DNA was cloned into the SmaI site of pNM481, transformed into MC4100, and plated on MacConkey-lactose agar to detect Lac⁺ recombinant plasmids. The resulting fusion plasmid, pMJF1 (see Fig. 6), was checked by cloning the EcoRI-HindIII fragment into M13mp18 and sequencing through the ansB'-'lacZ junction. Deletions into the ansB regulatory region were obtained by cleavage of pMJF1 at the upstream vector EcoRI site, followed by treatment with BAL 31 nuclease for various times and then ligation. The deleted plasmids were then transformed into MC4100. The extent of each deletion was first determined approximately by restriction analysis and then by cloning an ScaI-HindIII fragment (ScaI and HindIII are sites in the fusion vector [28]) into M13mp18 and sequencing through the region containing the deletion.

Construction of fusion strains. The ansB'-'lacZ region from these deleted plasmids (see preceding paragraph) and from pMJF1 were each transferred to λRZ5 by homologous recombination and then subsequently to the chromosome of MC4100 by lysogenization (21). In brief, MC4100 isolates containing pMJF1, pMJF1.1, and pMJF1.2 were infected with λRZ5, and plate lysates were prepared and used to transduce MC4100 to Lac⁺ Ap^r by selection on MacConkey-lactose medium containing ampicillin (38). Individual lysogens were isolated and then characterized by restriction analysis as follows. A phage DNA preparation was made, after UV-induced lysis, by immunoprecipitation with Lambdasorb (Promega Biotec). DNA was isolated, and the presence and size of the appropriate fragment within the recombinant phage were ascertained by Southern hybridization.

Enzyme assays. β-Galactosidase was measured on toluenized cells as described by Miller (27). L-Asparaginase II activity was determined on intact cells by measuring the production of ammonia with Nessler reagent; the substrate concentration was 0.1 mM, at which L-asparaginase I activity is not significant (6).

RESULTS

Synthetic oligonucleotide design and use in Southern hybridization. Oligonucleotide probes were based on the amino acid sequence of asparaginase II, determined by Maita and Matsuda (24). Initially, four degenerate probes (14- and 17-mers) corresponding to regions of low codon degeneracy

```
                    59                              67
Amino Acids         Gln-Asp-Met-Asn-Asp-Asp-Val-Trp-Leu


                    5'                                3'
Possible codons     CAG GAU AUG AAU GAU GAU GUG UGG UU
                    A   C           C   C   C   A       C
                                                U
                                                C

                    3'                              5'
Probe sequence      GTC CTI TAC TTI CTI CTI CAI ACC IA
                    T                       C
```

```
                    117                                       187
Amino Acids         Gly-Tyr-Ile-His-Asp-Gly-Lys-Ile-Asp-Tyr-Gln


                    5'                                            3'
Possible codons     GGG UAU AUU CAU GAU GGG AAA AUU GAU UAU CA
                    A   C   C   C   A   C   G   C   C   C
                    U       A           A           U
                    C                   C

                    3'                                            5'
Probe sequence      CCI ATI TAI GTI CTI CCI TTT TAI CTI ATI GT
                    C                       C   C
```
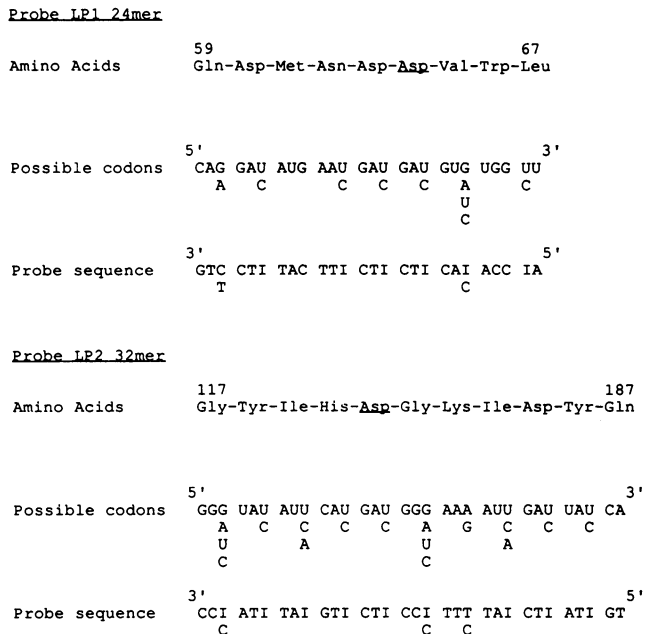
FIG. 1. Design of oligonucleotide probes used in the isolation of *ansB*. Amino acid sequences and their numbering are from the published sequence of L-asparaginase II (24). Underlined aspartate residues are ones that were subsequently found to be asparagine residues, based on the nucleotide sequence (see Fig. 3). All possible codons derived from the amino acid sequences and the sequence of the corresponding probes are shown.

were tested by Southern hybridization with restriction enzyme-digested *E. coli* DNA (see Materials and Methods). However, such probes were not specific, since not only did they hybridize to several fragments in each digest, but none

of these fragments hybridized to more than one probe. Two longer probes containing deoxyinosine, or deoxyinosine and deoxycytidine, residues at ambiguous codon positions were then used (29) (Fig. 1); much longer probes are possible due to the reduced complexity of the oligonucleotide mixture when only one or two residues are incorporated at degenerate codon position, rather than up to four residues. When used in Southern hybridization with restriction endonuclease-digested *E. coli* DNA, a single fragment was revealed; in each case the same fragment was revealed by both probes (data not shown).

Isolation and sequencing of the *ansB* gene. A 2.3-kilobase *Pst*I fragment was chosen for cloning. An enriched library of DNA fragments, encompassing this size range, was constructed and screened with γ-^{32}P-labeled oligonucleotide probe LP2, as described in Materials and Methods. One in 200 recombinants was positive; 1 of these, pMJ11, was analyzed by restriction analysis, and the nucleotide sequence was derived. This analysis revealed that the 3' end of the *ansB* gene (see below) was absent from the cloned fragment in pMJ11. Further Southern hybridization experiments, with the 2.3-kilobase *Pst*I fragment as a probe, revealed a *Hpa*I-*Hind*III fragment, which was cloned, as described above, to yield pMJ12 (Fig. 2); the nucleotide sequence of pMJ12 was determined, and the complete *ansB* gene was reconstructed by ligating the insert *Pst*I fragment from pMJ11 into the *Pst*I site of pMJ12, to give pMJ13 (Fig. 2).

Nucleotide sequence of *ansB* and derived amino acid sequence of L-asparaginase II. The nucleotide sequence reveals two open reading frames. The first of these, ORF1, exists upstream of the 5' end of the sequence shown in Fig. 3. It extends to nucleotide sequence position −179 and terminates with a UAA codon. A possible stem-and-loop terminator structure, with a free energy (ΔG) of −14.6 kcal (ca.
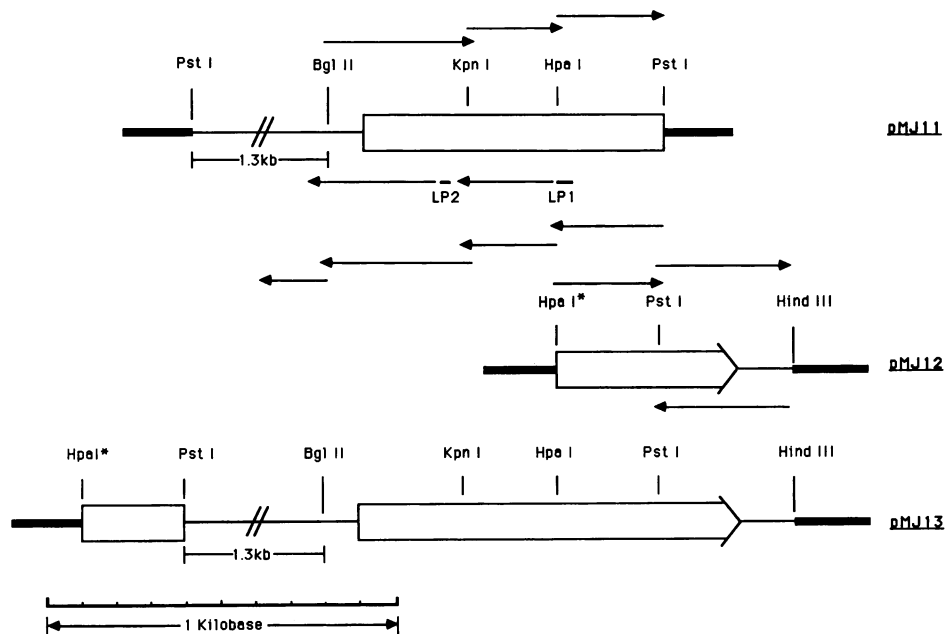

FIG. 2. Restriction maps of plasmids derived during the isolation of *ansB*. pMJ11 and pMJ12 were isolated independently from enriched *E. coli* genomic libraries. pMJ13 is derived from pMJ11 and pMJ12 as described in Results. The *Hpa*I sites with asterisks represent *Hpa*I-*Sma*I junctions. The arrows represent the extent and direction of nucleotide sequencing. The coding sequence of *ansB* is represented by the boxed area, the noncoding region is indicated by a thin line, and the vector is indicated by a thick line. The positions of the two synthetic oligonucleotide probes, LP1 and LP2, are shown. The arrows indicate the position and extent of nucleotide sequencing analysis.

```
         -361            -346            -331            -316            -301
AAATGGGCGCGAAAG CGGTGCTGAAAAGCG GCGGTAACCCATTAC AGAATGTGCTGGGAA GCCTGGGGGGGCTGC

         -286            -271            -256            -241            -226
AATCCTCAATCCAAA CCGAGTGGAAAAAGC AGGAAAAAGATTTCC AGCAGTTTGGCAAAG ATGTTTGTAGCCGCG

         -211            -196            -181            -166            -151
TTGTGACTCTGGAAG ATAGCCGCAAAGCCC TGGTCGGGAATTTAA AATAATCCTCTATTT TAAGACGGCATAATA
                                                  →       ‿‿        →

         -136            -121            -106             -91             -76
CTTTTTTATGCCGTT TAATTCTTCGTTTTG TTACCTGCCTCTAAC TTTGTAGATCTCCAA AATATATTCACGTTG
‿‿

          -61             -46             -31             -16              -1
TAAATTGTTTAACGT CAAATTTCCCATACA GAGCTAAGGGATAAT GCGTAGCGTTCACGT AACTGGAGGAATGAA
-10             *                                                       S.D.

                 15              30              45              60
ATG GAG TTT TTC AAA AAG ACG GCA CTT GCC GCA CTG GTT ATG GGT TTT AGT GGT GCA GCA
MET Glu Phe Phe Lys Lys Thr Ala Leu Ala Ala Leu Val MET Gly Phe Ser Gly Ala Ala
         |
                 75              90             105             120
TTG GCA TTA CCC AAT ATC ACC ATT TTA GCA ACC GGC GGG ACC ATT GCC GGT GGT GGT GAC
Leu Ala Leu Pro Asn Ile Thr Ile Leu Ala Thr Gly Gly Thr Ile Ala Gly Gly Gly Asp
        135
                                150             165             180
TCC GCA ACC AAA TCT AAC TAC ACA GTG GGT AAA GTT GGC GTA GAA AAT CTG GTT AAT GCG
Ser Ala Thr Lys Ser Asn Tyr Thr Val Gly Lys Val Gly Val Glu Asn Leu Val Asn Ala

                195             210             225             240
GTG CCG CAA CTA AAA GAC ATT GCG AAC GTT AAA GGC GAG CAG GTA GTG AAT ATC GGC TCC
Val Pro Gln Leu Lys Asp Ile Ala Asn Val Lys Gly Glu Gln Val Val Asn Ile Gly Ser

                255             270             285             300
CAG GAC ATG AAC GAT AAT GTC TGG CTG ACA CTG GCG AAA AAA ATT AAC ACC GAC TGC GAT
Gln Asp MET Asn Asp Asn Val Trp Leu Thr Leu Ala Lys Lys Ile Asn Thr Asp Cys Asp

                315             330             345             360
AAG ACC GAC GGC TTC GTC ATT ACC CAC GGT ACC GAC ACG ATG GAA GAA ACT GCT TAC TTC
Lys Thr Asp Gly Phe Val Ile Thr His Gly Thr Asp Thr MET Glu Glu Thr Ala Tyr Phe

                375             390             405             420
CTC GAC CTG ACG GTG AAA TGC GAC AAA CCG GTG GTG ATG GTC GGC GCA ATG CGT CCG TCC
Leu Asp Leu Thr Val Lys Cys Asp Lys Pro Val Val MET Val Gly Ala MET Arg Pro Ser

                435             450             465             480
ACG TCT ATG AGC GCA GAC GGT CCA TTC AAC CTG TAT AAC GCG GTA GTG ACC GCA GCT GAT
Thr Ser MET Ser Ala Asp Gly Pro Phe Asn Leu Tyr Asn Ala Val Val Thr Ala Ala Asp

                495             510             525             540
AAA GCC TCC GCC AAC CGT GGC GTG CTG GTA GTG ATG AAT GAC ACC GTG CTT GAT GGC CGT
Lys Ala Ser Ala Asn Arg Gly Val Leu Val Val MET Asn Asp Thr Val Leu Asp Gly Arg

                555             570             585             600
GAC GTC ACC AAA ACC AAC ACC ACC GAC GTA GCG ACC TTC AAG TCT GTT AAC TAC GGT CCT
Asp Val Thr Lys Thr Asn Thr Thr Asp Val Ala Thr Phe Lys Ser Val Asn Tyr Gly Pro

                615             630             645             660
CTG GGT TAC ATT CAC AAC GGT AAG ATT GAC TAC CAG CGT ACC CCG GCA CGT AAG CAT ACC
Leu Gly Tyr Ile His Asn Gly Lys Ile Asp Tyr Gln Arg Thr Pro Ala Arg Lys His Thr

                675             690             705             720
AGC GAC ACG CCA TTC GAT GTC TCT AAG CTG AAT GAA CTG CCG AAA GTC GGC ATT GTT TAT
Ser Asp Thr Pro Phe Asp Val Ser Lys Leu Asn Glu Leu Pro Lys Val Gly Ile Val Tyr

                735             750             765             780
AAC TAC GCT AAC GCA TCC GAT CTT CCG GCT AAA GCA CTG GTA GAT GCG GGC TAT GAT GGC
Asn Tyr Ala Asn Ala Ser Asp Leu Pro Ala Lys Ala Leu Val Asp Ala Gly Tyr Asp Gly

                795             810             825             840
ATC GTT AGC GCT GGT GTG GGT AAC GGC AAC CTG TAT AAA TCT GTG TTC GAC ACG CTG GCG
Ile Val Ser Ala Gly Val Gly Asn Gly Asn Leu Tyr Lys Ser Val Phe Asp Thr Leu Ala

                855             870             885             900
ACC GCC GCG AAA ACC GGT ACT GCA GTC GTG CGT TCT TCC CGC GTA CCG ACG GGC GCT ACC
Thr Ala Ala Lys Thr Gly Thr Ala Val Val Arg Ser Ser Arg Val Pro Thr Gly Ala Thr

                915             930             945             960
ACT CAG GAT GCC GAA GTG GAT GAT GCG AAA TAC GGC TTC GTC GCC TCT GGC ACG CTG AAC
Thr Gln Asp Ala Glu Val Asp Asp Ala Lys Tyr Gly Phe Val Ala Ser Gly Thr Leu Asn

                975             990            1005            1020
CCG CAA AAA GCG CGC GTT CTG CTG CAA CTG GCT CTG ACG CAA ACC AAA GAT CCG CAG CAG
Pro Gln Lys Ala Arg Val Leu Leu Gln Leu Ala Leu Thr Gln Thr Lys Asp Pro Gln Gln

               1035                            1062            1077
ATC CAG CAG ATC TTC AAT CAG TAC TAA TCGCCTCGCCCCGGT ATCGTGCCGGGGCTT TTTCACTTCAG
Ile Gln Gln Ile Phe Asn Gln Tyr                 →                    ‿‿

               1103            1118            1133            1146
ACTCACGTCCATTGC CAATTTTAATTACCC TAATGATAATCACCG GAATAAATTATTT
```

FIG. 3. Nucleotide sequence of the *ansB* region of *E. coli*. The transcriptional start site is indicated with an asterisk (nucleotide position −62), and the presumed −10 hexamer is underlined. The Shine-Dalgarno (S.D.) sequence is also underlined. Sequences underlined with inverted arrows are palindromic sequences. The vertical arrow between the alanine and leucine residues (amino acids 22 and 23, numbering from the initiating methionine residue) indicates the site of signal peptide cleavage (see Results). Boxed regions represent sequences that correspond to oligonucleotide probes (Fig. 1). Amino acids that are single, double, or triple underlined represent single amino acid substitutions, substitutions of two valines for one valine, and an insertion, respectively, when compared with the sequence in reference 24.

```
ansA    GAT TCT TCA GAT ATG ACG CCG GAA GAC TGG CAG CAT ATT GCT GAA GAT ATT AAA GCG CAC    287-346
         •  ••     ••• •        •  •  • •• ••• •    •  ••    ••• •       •  ••
ansB    GGC TCC CAG GAC ATG AAC GAT AAT GTC TGG CTG ACA CTG GCG AAA AAA ATT AAC ACC GAC    235-294
            Ser     Asp Met             Trp         Ala         Ile


ansA    TAT GAC GAC TAT GAT GGT TTT GTC ATT CTG CAC GGC ACC GAC ACG ATG GCG TAT ACC GCC    347-406
         •  ••  •      ••  ••  •• ••• •••     ••• •• ••• ••• ••• ••• •   •  •• ••
ansB    TGC GAT AAG ACC GAC GGC TTC GTC ATT ACC CAC GGT ACC GAC ACG ATG GAA GAA ACT GCT    295-354
            Asp     Asp Gly Phe Val Ile     His Gly Thr Asp Thr Met         Thr Ala


ansA    TCT GCG CTG TCG TTC ATG CTC GAG AAT CTC GGT AAA CCG GTC ATT GTG ACA GGG TCA CAA    407-466
         •      ••      •  ••  • ••     • •    ••• ••• ••  •   ••  •• ••
ansB    TAC TTC CTC     GAC CTG ACG GTG AAA TGC GAC AAA CCG GTG GTG ATG GTC GGC GCA ATG    355-411
            Leu                         Lys Pro Val         Gly


ansA    ATC CCG CTG GCT GAG TTA CGC TCT GAC GGA CAA ATT AAT CTG CTG AAT GCG TTG    467-520
         •••     •      •  ••  • ••• •••  • •  •• ••• • •
ansB    CGT CCG TCC ACG TCT ATG AGC GCA GAC GGT CCA TTC AAC CTG TAT AAC GCG GTA    412-465
            Pro             Asp Gly         Asn Leu     Asn Ala
```

FIG. 4. Nucleotide sequence comparison of *ansA* and *ansB*. The upper sequence is *ansA*, and the lower one is *ansB*. Nucleotide sequence matches are shown by dots between the sequences. Identical amino acids in both sequences are shown below the relevant codon. The nucleotide sequence numbering is shown at the right of each line; for *ansB* the numbering is as in Fig. 3 and 6, and for *ansA* the numbering is as in reference 19.

−61.09 kJ) (43), exists 15 bp downstream from the TAA codon (Fig. 3). The second open reading frame (Fig. 3) is identifiable as the *ansB*-coding sequence (24). Upstream from the codon for the mature N-terminal leucine residue, an initiation codon is present, preceded 6 bp upstream by a strong Shine-Dalgarno sequence. The 22-residue amino acid sequence preceding the experimentally determined N-terminal leucine residue has all the characteristics of a cleavable secretory signal peptide (46, 47): a charged N-terminal region (amino acid residues 1 to 6, where residue 1 is the initiating methionine residue; Fig. 3) and a hydrophobic region (residues 7 through 16) followed by a more polar region (residues 17 through 22). Furthermore, this sequence conforms well with the '−3, −1' rule for the distribution of amino acids N terminal to the cleavage site, derived from a survey of many cleaved signal peptides (46, 47). We conclude that L-asparaginase II, as expected for a periplasmic enzyme, has a 22-residue N-terminal signal peptide which is cleaved between residues 22 and 23 (Fig. 3) to yield a mature protein with an N-terminal leucine residue (24).

The coding sequence of 348 codons terminates with a UAA nonsense codon; 7 bp downstream, a perfect G+C-rich inverted repeat of 8 bp, with a free energy (ΔG) of −22.4 kcal (ca. −93.72 kJ) (43), is followed by a poly(T) sequence. Such a structure, in the mRNA, is typical of one group of rho-independent transcription terminators (31).

The mature asparaginase II sequence consists of 326 amino acids with a calculated M_r of 34594. The primary structure, as deduced from the nucleotide sequence (Fig. 3), deviates from the previously determined sequence (24) at 11 amino acid positions. These changes are of three types: six amino acid substitutions, four positions at which two valine residues are present instead of a single valine, and one leucine residue in the present sequence that is absent in the determined amino acid sequence. These differences are noted in Fig. 3; they may represent real differences between different strains and/or difficulties with the amino acid sequence determination.

Comparison of *ansB* with *ansA*. It is now possible to compare the nucleotide sequences of *ansA* and *ansB*: a matrix comparison (33) reveals a single area of similarity (data not shown), corresponding to that obtained previously in a comparison of the amino acid sequences (Fig. 4) (19). A region of particularly high nucleotide and amino acid similarity is seen between nucleotide positions (in *ansB*) 307 and 342 (inclusive). Within this short sequence, the amino acid sequence identity (91.6%) is greater than the nucleotide sequence identity (80.5%). However, when longer regions encompassing this region are compared, this situation is reversed. For example, the nucleotide sequence and amino acid sequence identity corresponding to the entire region shown in Fig. 4 are 53 and 41%, respectively.

Transcriptional analysis of the *ansB* gene. Total RNA was isolated from MC4100 containing pMJF1 or pNM481 cultured anaerobically in LB or LB containing glucose. The RNA with the probe was annealed, extended, and electrophoresed along with appropriate nucleotide sequencing reactions (see Materials and Methods). The analysis (Fig. 5)

FIG. 5. Transcriptional analysis of the *ansB* gene. Samples were electrophoresed on a 6% denaturing polyacrylamide gel. Lanes 1 through 4 contain primer extensions from ^{32}P-labeled probe with or without RNA as a template (see Materials and Methods): 1, no RNA; 2, RNA from MC4100 containing pNM481; 3, RNA from MC4100 containing pMJF1, grown in LB plus glucose; 4, RNA from MC4100 containing pMJF1, grown in LB. Lanes G, A, T, and C represent nucleotide sequencing reactions. The oligonucleotide in the sequencing reactions, which corresponds to the extension product, is indicated by an arrow.
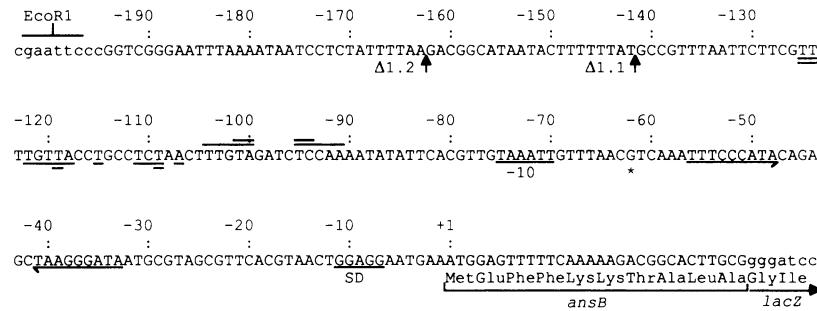
```
EcoR1    -190      -180      -170      -160      -150      -140      -130
 |                 :         :         :         :         :         :
cgaattcccGGTCGGGAATTTAAAATAATCCTCTATTTTAAGACGGCATAATACTTTTTTATGCCGTTTAATTCTTCGTT
                                         Δ1.2↑                Δ1.1↑                ═══

    -120      -110      -100       -90       -80       -70       -60       -50
     :         :         :         :         :         :         :         :
TTGTTACCTGCCTCTAACTTTGTAGATCTCCAAAATATATTCACGTTGTAAATTGTTTAACGTCAAATTTCCCATACAGA
                                                 -10               *

    -40       -30       -20       -10        +1
     :         :         :         :         :
GCTAAGGGATAATGCGTAGCGTTCACGTAACTGGAGGAATGAAATGGAGTTTTTCAAAAAGACGGCACTTGCGggatcc
                                   SD              MetGluPhePheLysLysThrAlaLeuAlaGlyIle
                                                     ansB                          lacZ
```

FIG. 6. The 5' region of *ansB*, showing the extent of deletions, and the site of fusion with *lacZ*. The vertical arrows indicate the extent of deletions in pMJF1.1 and pMJF1.2; the DNA 5' to each arrow is replaced. The −10 and Shine-Dalgarno (SD) sequences are underlined, and the transcriptional start site is indicated with an asterisk. A sequence with homology to the proposed FNR protein-binding site (17) is indicated with lines above the sequence. The underlined sequence shows where a putative CRP-binding site conforms to the CRP consensus sequence (5). Deviations from either consensus sequence are double underlined or overlined. An inverted repeat is underlined with converging arrowed lines. The sequence is numbered as in Fig. 3. The sequences in lowercase letters are derived from the vector (pNM481).

showed one major extension product corresponding to a C residue (in the noncoding strand) at nucleotide position −62 (Fig. 3 and 6), which is hence the transcriptional start point. The relatively lower amount of extended product with RNA from the culture containing glucose (Fig. 5, compare lanes 3 and 4) is consistent with the observed dependence of L-asparaginase II synthesis on CRP (8, 35). At 7 bp upstream from the transcriptional start point is a sequence with good homology with the −10 consensus sequence (Fig. 3 and 6). A sequence is not evident, however, with similarity to the consensus −35 hexamer; lack of a −35 sequence has previously been noted in positively regulated *E. coli* promoters (32).

Regulation of *ansB* by using *ansB'-'lacZ* fusions. Fusions between *ansB* and *lacZ* were constructed to provide a sensitive and facile assay to study the regulation of *ansB*. The initial fusion includes 195 bp of 5'-noncoding sequence and 29 bp of coding sequence from pMJ11. To define the *ansB* regulatory region, deletions were made that extended from the upstream *Eco*RI site into the 5'-untranslated region of *ansB* (Fig. 6). The resulting plasmids were tested for β-galactosidase activity on MacConkey-lactose media, and selected plasmids were sequenced to determine the extent of deletion. Two of these, pMJF1.1 and pMFJ1.2 (Fig. 6), contained the most extensive deletions while remaining Lac⁺. To assay the expression of these *ansB'-'lacZ* deletion derivatives in single copy number, they were transferred to the chromosome of MC4100 via homologous recombination with λRZ5 followed by lysogenization with recombinant phage (see Materials and Methods).

The *ansB* gene is positively regulated by both anaerobiosis, through the FNR protein (20), and by CRP (8, 35). To determine whether the deleted *ansB'-'lacZ* fusions were still regulated by these proteins, the appropriate lysogens were assayed in the presence and absence of oxygen and in the presence and absence of glucose (in the absence of oxygen). The deletions represented by pMJF1.1 and pMJF1.2 did not abolish or reduce anaerobic induction or CRP regulation when compared with the undeleted gene fusion (Fig. 7); this defines the regulatory 5' region of *ansB* to a small 76-bp region between nucleotide position −141 and the transcriptional start point.

We have observed that the anaerobically induced level of L-asparaginase II is no greater in a strain containing pMJ13, and hence containing multiple copies of *ansB*, than in the same strain lacking this plasmid (data not shown). This observation is confirmed by a comparison of strains contain-

ing *ansB'-'lacZ* in multiple copies (pMJF1) and in a single copy (λMJF1), since they contain similar levels of β-galactosidase. The absence of a gene dosage effect is possibly due to limitations imposed by cellular levels of the positive regulatory proteins, FNR protein and CRP.

DISCUSSION

Comparison between homologous genes in related organisms, such as *Salmonella typhimurium* and *E. coli*, reveal that similarity is greater at the amino acid level that at the nucleotide level. In contrast, comparisons of genes presumed to have arisen by duplication within *E. coli* or *S. typhimurium*, or of paired regions of such genes, have revealed that their nucleotide sequence similarity is generally greater than their amino acid similarity (14, 34, 41). This latter situation also holds true for the single region of similarity between *ansA* and *ansB* (see Results). It has been suggested (41) that, as two genes diverge in structure and function, constraints on the use of synonomous codons imposed by bias in codon usage and selection directly at the
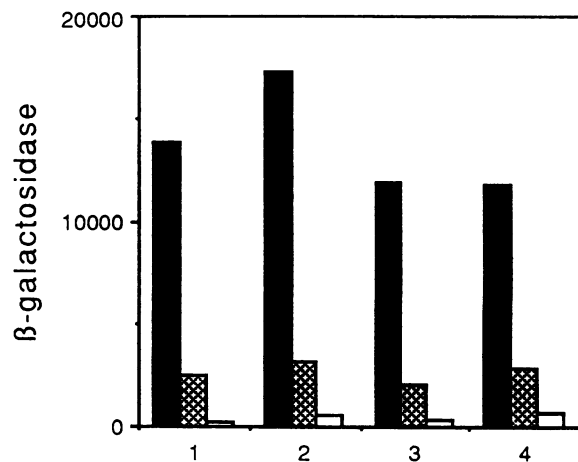

FIG. 7. Effect of anaerobiosis and catabolite repression on *ansB'-'lacZ* expression. Cells were grown anaerobically in LB in the presence (▨) and absence (■) of 0.4% glucose and aerobically in LB alone (□). The following strains were assayed: 1, MC4100(pMJF1); 2, MC4100(λMJF1); 3, MC4100(λMJF1.1); 4, MC4100(λMJF1.2). β-Galactosidase activity is expressed in Miller units (27). The specific activity was corrected in cultures in which the proportion of Lac⁺ Apʳ cells was <100% (see Materials and Methods).

nucleotide sequence level might be factors involved in maintaining greater nucleotide similarity than amino acid similarity. Gene conversion (14) is a less likely explanation for the similarity between *ansA* and *ansB* in view of the relatively low nucleotide sequence identity.

The production of L-asparaginase II in *E. coli* is known to be subject to positive regulation by two pleiotrophic regulatory proteins: the FNR protein (20), which activates a number of genes during anaerobiosis (18, 39, 44); and CRP (8, 35), which controls the initiation of transcription of genes in various catabolite pathways (5, 11, 12). It should be noted that regulation of L-asparaginase II by CRP is not via the FNR protein (i.e., regulation of FNR by CRP). Although synthesis of FNR protein is indeed subject to catabolite repression (39), *fnr'*-*'lacZ* fusions contained in *crp* and *cya* mutant cells can still express 60% of the β-galactosidase activity measured in the wild-type strain (30).

The *ansB'*-*'lacZ* fusion used in this study exhibits regulation by both of the above factors, both in single and multiple copies (Fig. 7). Furthermore, deletion analysis with the fusions λMJF1.1 and λMJF1.2 has defined the amount of 5'-untranslated nucleotide sequence, sufficient to maintain this regulation, to a region between the transcriptional start and a point 76 bp upstream (Fig. 6 and 7). Specific sequences to which the CRP and FNR proteins bind, to stimulate transcription, have been proposed. The CRP recognition sequence has been studied in a large number of genes, and the consensus sequence 5'-AANTGTGANNtNgNTCAaAt at-3' (5) has been confirmed experimentally (5, 11, 16). Consensus sequences to which the FNR protein may bind have been proposed on the basis of sequence homology in the 5'-noncoding regions of FNR-dependent genes (18, 22, 39). Recent experimental evidence has supported the role of these motifs in FNR-mediated regulation. When the inverted repeat 5'-TTGATNNNNATCAA-3' in the *nar* operon (a subset of a region of reported homology [39]) is disrupted by a deletion, a loss of anaerobic induction occurs (23). Furthermore, when a 24-bp sequence from the *nirB* promoter, containing this inverted repeat, is transplanted to the promoter region of *galP1* and placed in a *lacZ* expression vector, the hybrid promoter stimulates the expression of *lacZ* under anaerobic conditions (17). The *ansB* 5'-untranslated sequence required for regulation by FNR protein and CRP, as defined by deletion analysis, was examined for the presence of the regulatory elements described above. A sequence with some homology to the putative FNR protein-binding site was found (Fig. 6). A sequence resembling the CRP consensus sequence was also found (Fig. 6). It must be noted, however, that this sequence departs from the consensus at a position known to be important for CRP binding. The change, a G→T in position 4 of the TGTGA motif, has been shown to significantly reduce CRP binding (16) and is not present in any reported CRP-binding site (5). The presence of a sequence with only partial homology with the FNR protein-binding site and of a nonoptimal CRP site in the *ansB* 5' sequence do not preclude direct interaction of these proteins with this promoter region. The FNR protein-binding sequence is only tentatively defined, and CRP may tolerate the changes noted in the putative binding sequence (Fig. 6).

However, the possibility that these regulatory proteins are acting indirectly on the *ansB* gene should be entertained. This has some precedent in the case of the *fnr*-mediated regulation of formate hydrogenlyase, which is indirect and depends on two other gene products (9). In this regard, and as noted above, the λMJF1.1 deletion retains only 76 bp upstream from the transcriptional start, yet regulation is

retained. This leaves little space for the interaction of multiple regulatory proteins. For example, in the case of *melR*, in which the CRP-binding site is close to the RNA polymerase-binding region, the combined footprint extends up to position −70 (49) relative to the transcriptional start.

If the regulation of the *ansB* gene is indirect with respect to FNR protein, CRP, or both, it may involve an observed stimulation by amino acids (6). The mechanism and the level of regulation by amino acids are as yet unknown. A feature of the sequence that may be involved in such indirect regulation is an inverted repeat present downstream from the transcriptional start, which, if significant, could interact with a regulatory protein to occlude RNA polymerase binding. Future experiments to resolve the regulation of *ansB* will include more extensive deletion analysis and site-directed mutagenesis.

LITERATURE CITED

1. **Aiba, H., S. Adhya, and B. de Crombrugghe.** 1981. Evidence for two functional gal promoters in intact *Escherichia coli* cells. J. Biol. Chem. **256:**11905–11910.
2. **Ammon, H. L., I. T. Weber, A. Wlodawer, R. W. Harrison, G. L. Gilliland, K. C. Murphy, L. Sjölin, and J. Roberts.** 1988. Preliminary crystal structure of *Acinetobacter glutaminasifcans* glutaminase-asparaginase. J. Biol. Chem. **263:**150–156.
3. **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.
4. **Burns, D. M., and I. R. Beacham.** 1983. A method for ligation of DNA following isolation from low melting temperature agarose. Anal. Biochem. **135:**48–51.
5. **Busby, S. J. W.** 1986. Positive regulation in gene expression, p. 51–77. *In* I. R. Booth and C. F. Higgins (ed.), Regulation of gene expression: 25 years on. Cambridge University Press, Cambridge.
6. **Cedar, H., and J. H. Schwartz.** 1967. Localisation of two L-asparaginases in anaerobically grown *Escherichia coli*. J. Biol. Chem. **242:**3753–3755.
7. **Chen, E. Y., and P. H. Seeburg.** 1985. Superoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4:**165–170.
8. **Chesney, R. H.** 1983. *E. coli* L-asparaginase II production in the presence and absence of catabolite activating protein. FEMS Microbiol. Lett. **17:**161–162.
9. **Cole, J. A.** 1988. Reassessment of the role of the FNR transcriptional activator protein in the expression of formate hydrogen-lyase activity of *Escherichia coli* K-12. FEMS. Microbiol. Lett. **55:**169–174.
10. **Crowther, D.** 1971. L-Asparaginase and human malignant disease. Nature (London) **229:**168–171.
11. **De Crombrugghe, B., S. Busby, and H. Buc.** 1984. Cyclic AMP receptor protein: role in transcriptional activation. Science **224:**831–838.
12. **Ebright, R. H., and J. Beckwith.** 1985. The catabolite gene activator protein (CAP) is not required for indole-3-acetic acid to activate transcription of the ara BAD operon of *Escherichia coli* K-12. Mol. Gen. Genet. **201:**51–55.
13. **Ho, P. P. K., E. B. Milikin, J. L. Bobbitt, E. L. Grinnan, P. J. Burck, B. H. Frank, L. D. Boeck, and R. W. Squires.** 1970. Crystalline L-asparaginase from *Escherichia coli* B. I. Purification and chemical characterization. J. Biol. Chem. **245:**3708–3715.
14. **Houmard, J., V. Capuano, T. Coursin, and N. T. deMarsac.**

1988. Genes encoding core components of the phycobilisome in the cyanobacterium *Calothrix* sp. strain PCC 7601: occurrence of a multigene family. J. Bacteriol. **170:**5512–5521.

15. **Hudson, G. S., and B. E. Davidson.** 1984. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K12. J. Mol. Biol. **180:**1023–1051.

16. **Jansen, C., A. M. Gronenborn, and G. M. Clore.** 1987. The binding of the cyclic AMP receptor protein to synthetic DNA sites containing permutations in the consensus sequence TGTGA. Biochem. J. **246:**227–232.

17. **Jayaraman, P. S., K. L. Gaston, J. A. Cole, and S. J. W. Busby.** 1988. The *nirB* promoter of *Escherichia coli*: location of nucleotide sequences essential for regulation by oxygen, the FNR protein and nitrite. Mol. Microbiol. **2:**527–530.

18. **Jayaraman, P. S., T. C. Peakman, S. J. W. Busby, R. V. Quincey, and J. A. Cole.** 1987. Location and sequence of the promoter of the gene for the NADH-dependent nitrite reductase of *Escherichia coli* and its regulation by oxygen, the FNR protein and nitrite. J. Mol. Biol. **196:**781–788.

19. **Jerlström, P. G., D. A. Bezjak, M. P. Jennings, and I. R. Beacham.** 1989. Structure and expression in *Escherichia coli* K-12 of the L-asparaginase I-encoding *ansA* gene, and its flanking regions. Gene **78:**37–46.

20. **Jerlström, P. G., J. Liu, and I. R. Beacham.** 1987. Regulation of *Escherichia coli* L-asparaginase II and L-aspartase by the *fnr* gene product. FEMS Microbiol. Lett. **41:**127–130.

21. **Jones, H. M., and R. P. Gunsalus.** 1987. Regulation of *Escherichia coli* fumarate reductase (*frdABCD*) operon expression by respiratory electron acceptors and the *fnr* gene product. J. Bacteriol. **169:**3340–3349.

22. **Li, S. F., and J. A. DeMoss.** 1987. Promoter region of the *nar* operon of *Escherichia coli*: nucleotide sequence and transcription initiation signals. J. Bacteriol. **169:**4614–4620.

23. **Li, S. F., and J. A. DeMoss.** 1988. Location of sequences in the *nar* promoter of *Escherichia coli* required for regulation by Fnr and NarL. J. Biol. Chem. **263:**13700–13705.

24. **Maita, T., and G. Matsuda.** 1980. The primary structure of L-asparaginase from *Escherichia coli*. Hope-Seyler's Z. Physiol. Chem. **361:**105–117.

25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

26. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101:**20–78.

27. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

28. **Minton, N. P.** 1984. Improved plasmid vectors for the isolation of translational *lac* gene fusions. Gene **31:**269–273.

29. **Ohtsuka, E., S. Matsuki, M. Ikehara, Y. Takakashi, and K. Matsubara.** 1985. An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. J. Biol. Chem. **260:**2605–2608.

30. **Pascal, M. C., V. Bonnefoy, M. Fons, and M. Chippaux.** 1986. Use of gene fusions to study the expression of *fnr*, the regulatory gene of anaerobic electron transfer in *Escherichia coli*. FEMS Microbiol. Lett. **36:**35–39.

31. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55:**339–372.

32. **Raibaud, O., and M. Schwartz.** 1984. Positive control of transcription initiation in bacteria. Annu. Rev. Genet. **18:**173–206.

33. **Reisner, A. H., and C. A. Bucholtz.** 1986. The MTX package of computer programmes for the comparison of sequences of nucleotides and amino acid residues. Nucleic Acids. Res. **14:**233–238.

34. **Riley, M.** 1984. Arrangement and rearrangement of bacterial genomes, p. 285–315. *In* R. P. Mortlock (ed.), Microorganisms as model systems for studying evolution. Plenum Publishing Corp., New York.

35. **Russell, L., and H. Yamazaki.** 1978. The dependence of *Escherichia coli* asparaginase II formation on cyclic AMP and cyclic AMP receptor protein. Can. J. Microbiol. **24:**629–631.

36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

37. **Schwartz, J. H., J. Y. Reeves, and J. D. Broome.** 1966. Two L-asparaginases from *E. coli* and their action against tumours. Proc. Natl. Acad. Sci. USA **56:**1516–1519.

38. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

39. **Spiro, S., and J. R. Guest.** 1987. Regulation and over-expression of the *fnr* gene of *Escherichia coli*. J. Gen. Microbiol. **133:**3279–3288.

40. **Spring, K. J., P. G. Jerlström, D. M. Burns, and I. R. Beacham.** 1986. L-Asparaginase genes in *Escherichia coli*: isolation of mutants and characterization of the *ansA* gene and its protein products. J. Bacteriol. **166:**135–142.

41. **Stokes, H. W., P. W. Betts, and B. G. Hall.** 1985. Sequence of the *ebgA* gene of *Escherichia coli*: comparison with the *lacZ* gene. Mol. Biol. Evol. **2:**469–477.

42. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84:**4767–4771.

43. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246:**40–41.

44. **Unden, G., and J. R. Guest.** 1984. Cyclic AMP and anaerobic gene expression in *E. coli*. FEBS Lett. **170:**321–325.

45. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153:**3–11.

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

47. **Von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.

48. **Wade, H. E.** 1980. Synthesis and functions of microbial asparaginases and glutaminases, p. 563–575. *In* J. W. Payne (ed.), Microorganisms and nitrogen sources. John Wiley & Sons, Inc., New York.

49. **Webster, C., K. Gaston, and S. Busby.** 1988. Transcription from the *Escherichia coli melR* promoter is dependant on the cyclic AMP receptor protein. Gene **68:**297–305.

50. **Willis, R. C., and C. A. Woolfolk.** 1974. Asparagine utilization in *Escherichia coli*. J. Bacteriol. **118:**231–241.