Journal of Bacteriology, Mar. 1988, p. 1325–1332
0021-9193/88/031325-08$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 3

Molecular Cloning, Characterization, and Nucleotide Sequence of an Extracellular Amylase Gene from *Aeromonas hydrophila*

KARI S. GOBIUS and JOHN M. PEMBERTON*

Department of Microbiology, University of Queensland, St. Lucia, Queensland 4067, Australia

Received 6 June 1987/Accepted 30 November 1987

The structural gene for excreted amylase from *Aeromonas hydrophila* JMP636 has been cloned within a 2.1-kilobase *Sma*I fragment of DNA. The amylase gene is transcribed from its own promoter in *Escherichia coli*, producing a gene product of M_r 49,000. The amylase gene product is secreted to the periplasm of *E. coli*; however, it is not excreted. Nucleotide sequencing revealed an open reading frame of 1,392 base pairs corresponding to a protein of 464 amino acid residues. A potential signal peptide of 21 amino acid residues is present at the NH_2 terminal of the predicted protein. Three regions of homology with other procaryotic and eucaryotic α-amylases were detected within the predicted amino acid sequence.

Previous studies have revealed that most proteins that are secreted across the cytoplasmic membrane of both gram-positive and gram-negative bacteria are synthesized as precursor molecules (36). The precursor protein is then processed by a signal peptidase which acts to remove 15 to 30 amino acid residues (signal sequences) from the NH_2-terminal end of the precursor, yielding a shorter protein which is then correctly localized in the periplasm, outer membrane, or extracellular environment (30). Such signal sequences appear to be essential for transition through the cytoplasmic membrane and contain discrete cleavage sites which are recognized by the signal peptidase (35, 36). Despite the level of understanding of signal peptidase cleavage and secretion through the cytoplasmic membrane, the molecular mechanisms for the subsequent correct localization of proteins in the periplasm, in the outer membrane, or outside the cell remain largely to be elucidated.

Aeromonas hydrophila excretes a wide range of extracellular enzymes (18), several of which are soluble extracellular toxins implicated as virulence factors in the pathogenesis of this organism (4, 14, 19). Because of the significance of these virulence factors and the number of additional exoenzymes excreted by *A. hydrophila*, this organism has been chosen for the study of the transport of its soluble proteins across the cytoplasmic and outer membranes. Recently, two studies have examined the expression of cloned *A. hydrophila* aerolysin in *Escherichia coli* (3, 19). These studies have shown that although aerolysin is extracellularly excreted by *A. hydrophila* it is only secreted to the periplasm by *E. coli*, indicating that a barrier to excretion of this protein exists in *E. coli* or that the gene encoding aerolysin is not alone sufficient to direct excretion. Aerolysin has a signal peptide which is removed in *A. hydrophila* and probably in *E. coli*.

In this paper we describe the molecular cloning, mapping, expression in *E. coli*, and nucleotide sequence determination of an extracellular amylase gene from *A. hydrophila*.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophages. The bacterial strains, plasmids, and bacteriophages used in this study are listed in Table 1.

Media and culture conditions. Bacterial strains were grown and maintained in Z broth (8), supplemented with 1.5% agar when appropriate. Z starch medium was 0.5% maize starch dissolved in Z broth. Zones of clearing after flooding with dilute iodine solution indicated starch hydrolysis. *E. coli* K38 maxicells were grown in M9 medium (24) supplemented with 0.4% glucose, 20 μg of thiamine per ml, 1 mM $CaCl_2$, 1 mM $MgSO_4$, and 0.01% of 18 amino acids (minus cysteine and methionine). 2YT medium (24) was used for growth of *E. coli* JM109 when this strain was used to propagate recombinant M13 mp18 and mp19 bacteriophages. 2YT medium was supplemented with the chromogenic substrate 5-bromo-4-chloro-3-indoyl-β-D-galactoside at a concentration of 40 μg/ml and isopropyl-β-D-thiogalacto-pyranoside (10 μg/ml) to detect recombinant mp18 and mp19 bacteriophages.

DNA manipulations. Extraction of *A. hydrophila* JMP636 whole cell DNA was performed by the method of Priefer et al. (26). The DNA was partially digested with serial dilutions of *Sau*3A1 for 20 min at 37°C (26), and samples were electrophoresed in 0.5% agarose to select fragments suitable for cloning in pHC79. Undigested λ DNA was used as a molecular weight marker to select appropriately sized fragments of insert DNA. DNA from *E. coli* cosmid clones was extracted by the method of Hansen and Olsen (10). Plasmid DNA was extracted by the method of Naumovski and Friedberg (25). Single-stranded DNA from recombinant M13 mp18 and mp19, to be used for nucleotide sequence determination, was isolated by the method of Messing (23). DNA sequencing was performed essentially as described by Sanger et al. (28). Restriction endonuclease digestion and in vitro ligation were performed according to the suppliers' specifications; however, 100 μg of bovine serum albumin per ml was added to digestions of DNA originating from rapid plasmid extractions. Packaging of recombinant cosmids was performed with a packaging kit from Boehringer Mannheim Biochemicals. Transformation of DNA was by the method of Cohen et al. (5).

Exclusive labeling of plasmid-encoded proteins with the T7 RNA polymerase-promoter system. The labeling procedure used was essentially that of Tabor and Richardson (32) and Tabor (personal communication). Overnight cultures of *E. coli* K38 containing both pGP1-2 and a pT7 recombinant plasmid were used to inoculate 5 ml of Z broth supplemented with 40 μg of ampicillin and kanamycin per ml. Cultures were grown for 2 to 3 h (approximately); when the cultures reached an A_{590} of 0.5, 2 ml of each culture was centrifuged in a bench-top centrifuge. Cell pellets were washed with 5 ml

* Corresponding author.

TABLE 1. Bacterial strains, plasmids, and bacteriophages

Strain, plasmid, or phage	Genotype or description	Source or reference
Bacteria		
A. hydrophila JMP636	Wild type	8
E. coli K-12		
DH1	F⁻ *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1* λ⁻	9
MM294	*endA1 hsdR17 thi-1*	33
K38	HfrC (λ)	27
JM109	Δ(*pro-lac*) *thi rpsL hsdR4 endA* F' *traD36 proA⁺B⁺ lacI^q lacZ*ΔM15	P. Suntinanalerts
JMP3085	DH1(pJP3101)	This study
Plasmids		
pBR322	Ap^r Tc^r	1
pHC79	Ap^r Tc^r *cos*	13
pUC12	Ap^r	34
pGP1-2	Km^r T7 RNA polymerase	32
pT7-3	Ap^r T7 φ10	S. Tabor
pT7-4	Ap^r T7 φ10	S. Tabor
pT7-5	Ap^r T7 φ10	S. Tabor
pT7-6	Ap^r T7 φ10	S. Tabor
pJP5037	Ap^r Pec⁺[a]	P. Suntinanalerts
pJP3101	Ap^r Amy⁺[b]	This study
pJP3103	Ap^r Amy⁺	This study
pJP3104	Ap^r Amy⁺	This study
pJP3105	Ap^r T7 φ10 Amy⁺	This study
pJP3106	Ap^r T7 φ10 Amy⁺	This study
pJP3107	Ap^r T7 φ10 Amy⁺	This study
pJP3108	Ap^r T7 φ10 Amy⁺	This study
pJP3109	Ap^r Amy⁻	This study
pJP3111	Ap^r Amy⁻	This study
pJP3112	Ap^r Amy⁻	This study
Bacteriophages		
M13 mp18	M13 sequencing vector	Pharmacia
M13 mp19	M13 sequencing vector	Pharmacia

[a] Pec, Pectinase.
[b] Amy, Amylase.

of M9 medium (24) and centrifuged. Cell pellets were then suspended in 1 ml of supplemented M9 medium and grown with aeration at 30°C for 60 min. The temperature of incubation was raised to 42°C for 15 min, and then rifampin was added to a final concentration of 200 μg/ml. Cells were further incubated at 42°C for 10 min. The incubation temperature was lowered to 32°C for 20 min, and then cultures were pulsed with 40 μCi of [^{35}S]methionine for 5 min. Cultures were transferred to Eppendorf tubes and centrifuged for 20 s. Supernatants were removed, and cell pellets were suspended in 120 μl of sample buffer for gel electrophoresis (60 mM Tris hydrochloride [pH 6.8], 1% sodium dodecyl sulfate, 1% 2-mercaptoethanol, 10% glycerol, and 0.01% bromophenol blue). Samples were heated to 95°C for 3 min and loaded onto a 12.5% sodium dodecyl sulfate–polyacrylamide gel (20) or stored at −20°C.

Maxicell fractionation. *E. coli* K38(pGP1-2)(pJP3106) maxicells were labeled with [^{35}S]methionine as described above, except that cells were suspended in 0.5 ml of supplemented M9 medium and labeled with [^{35}S]methionine for 15 min. Maxicells were then fractionated as described by Heppel (12). The labeled bacterial cells were centrifuged, and the supernatant was retained as the extracellular fraction. Cell pellets were gently washed twice with 0.25 ml of 10 mM Tris hydrochloride (pH 8.0), and the supernatants were discarded. Washed cells were suspended in 0.5 ml of 20% sucrose–0.1 mM EDTA–30 mM Tris hydrochloride (pH 7.3) and gently shaken at 180 rpm on a rotary shaker for 10 min. Cells were then pelleted by centrifugation at 10,000 × *g* for 5 min in an Eppendorf centrifuge, the supernatant was removed, and the cells were immediately suspended in 0.5 ml of ice-cold 0.5 mM $MgCl_2$. The cells were shaken for 10 min at 3°C, allowing osmotic shock, and centrifuged as before. The supernatant was retained as the periplasmic fraction. After a second cycle of osmotic shock the cells were centrifuged, suspended in 0.5 ml of 10 mM Tris hydrochloride (pH 7.5), and retained as the cytoplasm-plus-membrane fraction. After fractionation 50 μl of 10× electrophoresis sample buffer and 55 μl of glycerol were added to the extracellular, periplasmic, and cytoplasm-plus-membrane fractions before each fraction was heated to 95°C for 3 min. Portions of 50 μl of each sample were electrophoresed in 12.5% sodium dodecyl sulfate–polyacrylamide gels (20), which were dried down and autoradiographed with Fuji RX X-ray film.

RESULTS

Cloning and subcloning of *A. hydrophila* amylase gene in *E. coli*. A genomic library of *A. hydrophila* JMP636 was constructed in *E. coli* DH1 by using *Sau*3A1 partially digested *A. hydrophila* DNA ligated into the *Bam*HI site of the cosmid pHC79. Among 700 carbenicillin-resistant, tetracycline-susceptible cosmid clones, 3 showed strong amylolytic activity on Z starch agar after 18 h of incubation at 37°C. One clone, JMP3085, harboring the cosmid pJP3101, which contained a 24 kilobase (kb) insert of *A. hydrophila* DNA, was chosen for further study. The cosmid pJP3101 was successively subcloned into the vectors pBR322 and pUC12; finally a 2.1-kb *Sma*I fragment was inserted into the *Sma*I site of pUC12. This plasmid, pJP3103, was used to transform *E. coli* MM294 to amylolytic activity. Expression of pJP3103-encoded amylase by MM294 produced zones of clearing in Z starch agar plates which were equal in clarity and diameter to those of *E. coli* JMP3085; however, pJP3103-encoded amylase zones were produced after only 12 h of incubation at 37°C. Restriction endonuclease analysis was used to construct a restriction map of pJP3103 (Fig. 1a), allowing further delineation of the amylase gene.

Deletion analysis of the amylase gene on pJP3103. Several in vitro deletion mutants of pJP3103 were constructed to determine the approximate extent of the amylase gene. The deletion plasmids pJP3109, pJP3111, and pJP3112 (Fig. 1b) were constructed by digestion of pJP3103 with the restriction endonucleases *Bam*HI, *Sal*I, and *Sph*I, respectively, and religation or the DNA of each sample at low concentrations to ensure that the majority of recombinant molecules would not contain the fragment intended for deletion. pJP3109, pJP3111, and pJP3112, when used to transform MM294, all gave rise to nonamylolytic colonies on Z starch agar. We found that the *Bam*HI and *Sal*I sites of the insert DNA interrupted the amylase gene. Additionally, at least one *Sph*I site (presumably that which lies between the *Bam*HI and *Sal*I sites), and possibly both, also interrupts the amylase gene. After nucleotide sequence determination (see below) a unique *Nco*I site 92 base pairs (bp) upstream of the potential amylase promoter and the *Hin*dIII site present in the pUC12 polylinker were used to delete the nucleotide sequence preceding the potential amylase-coding region. This deletion plasmid was designated pJP3104; when used to transform

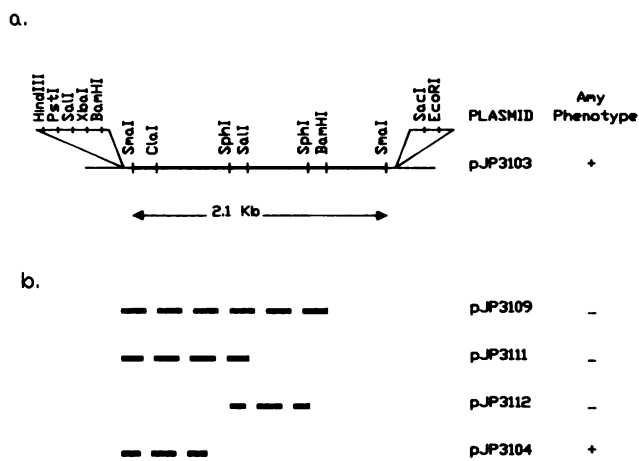

FIG. 1. (a) Restriction map of pJP3103, which encodes the *A. hydrophila* amylase. Thick lines represent the amylase gene insert, and thin lines represent the pUC12 vector moiety. (b) Deletion derivatives of pJP3103. Thick broken lines indicate fragments of DNA which were deleted. The Amy phenotype indicates production of amylase by different deletion mutants.

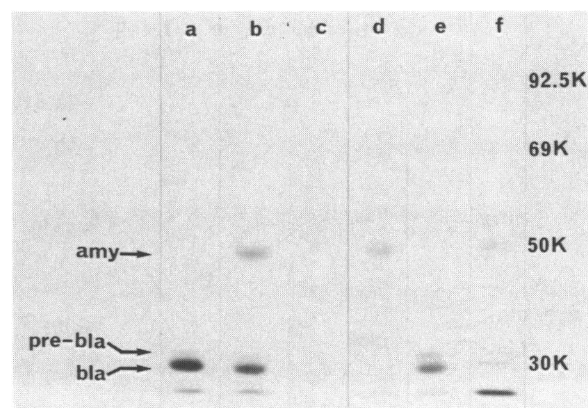

FIG. 3. Polyacrylamide gel electrophoresis of proteins synthesized in *E. coli* K38(pGP1-2) maxicells containing pJP3105 (lane a), pJP3106 (lane b), pJP3107 (lane c), pJP3108 (lane d), and pT7-4 with no insert DNA (lane e). The ^{14}C-labeled molecular weight standards (lane f) are phosphorylase b (92,500), bovine serum albumin (69,000), ovalbumin (50,000), and carbonic anhydrase (30,000). The positions of amylase (*amy*) and β-lactamase (*bla*) genes are indicated at the left.

MM294, pJP3104 gave rise to amylolytic colonies, confirming that the *Nco*I site does not interrupt the amylase-coding region or its promoter.

In vivo transcription-translation of cloned amylase. The cloned *Aeromonas* amylase gene was subcloned into the expression vectors pT7-3, pT7-4, pT7-5, and pT7-6 (S. Tabor personal communication; also see use of pT7-1 [32]) with an *Eco*RI-*Pst*I double digest of pJP3103 and each of the vectors in turn. Double digests ensured the correct orientation of the amylase fragment in each vector. Recombinant pT7-3, pT7-4, pT7-5, pT7-6 clones carrying the *Eco*RI-*Pst*I amylase fragment were designated pJP3105, pJP3106, pJP3107, and pJP3108, respectively (Fig. 2). pJP3105, pJP3106, pJP3107, and pJP3108 were used to transform *E. coli* K38(pGP1-2), enabling the exclusive [^{35}S]methionine labeling of proteins encoded by these plasmids.

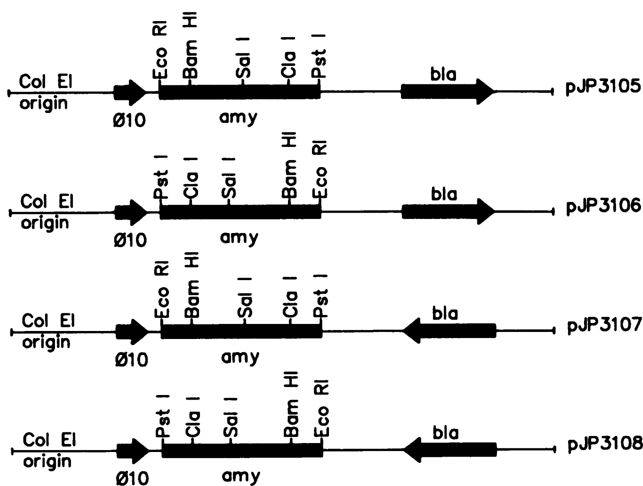

FIG. 2. Schematic representation of recombinant pT7 transcription-translation plasmids carrying the amylase gene fragment from pJP3103. The position and direction of transcription of the bacteriophage T7 ϕ10 promoter and the *bla* gene are indicated. The orientation of the amylase gene fragment (*amy*) in each vector is shown.

When *E. coli* K38(pGP1-2) harboring a pT7 recombinant plasmid is grown in the presence of rifampin at 42°C, mRNA synthesis is initiated solely at the 23-bp phage T7 promoter present on these plasmids (6, 32), since rifampin shuts off *E. coli* K-12 RNA polymerase. Therefore only mRNA synthesized by T7 RNA polymerase is available for translation in these cells, causing the exclusive labeling of pJP3105-, pJP3106-, pJP3107-, and pJP3108-encoded proteins in the presence of [^{35}S]methionine. Figure 3 indicates that transcription of pJP3106 and pJP3108 by T7 RNA polymerase results in the translation of amylase, a protein of M_r 49,000; however transcription of pJP3105 and pJP3107 does not. Translation of the cloned amylase gene on pJP3106 and pJP3108, associated with the absence of translation of amylase from pJP3105 and pJP3107 (Fig. 3), provides evidence that the 5′ end of the amylase gene is closest to the *Pst*I site of pJP3103 (Fig. 1a). Similarly [^{35}S]methionine-labeled pre-β-lactamase and mature β-lactamase are only translated from pJP3105 and pJP3106 transcripts; this result is expected, since the β-lactamase gene is under the control of the ϕ10 promoter in pJP3105 and pJP3106; however in pJP3107 and pJP3108 the β-lactamase gene is placed in the opposite orientation and may not be correctly transcribed by T7 RNA polymerase. These data also indicate that transcription of the cloned amylase gene in *E. coli* is normally initiated at an *A. hydrophila* promoter, since the direction of amylase gene transcription determined above is convergent with transcription of *lacZ* in pJP3103 (pUC12 recombinant). Secondly, *E. coli* K38 cells containing pJP3105, pJP3106, pJP3107, and pJP3108 all show phenotypic expression of the amylase gene in the absence of *E. coli* RNA polymerase-inhibiting rifampin.

Localization of cloned amylase in *E. coli*. To determine the localization of the cloned amylase gene product, we examined various fractions of osmotically shocked cells for the presence of [^{35}S]methionine-labeled gene product by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The exclusive labeling system of Tabor and Richardson (outlined above) was used to label the gene products of pJP3106 and pJP5037 with [^{35}S]methionine. pJP3106 encodes the cloned *Aeromonas* amylase and β-lactamase, whereas pJP5037 en-

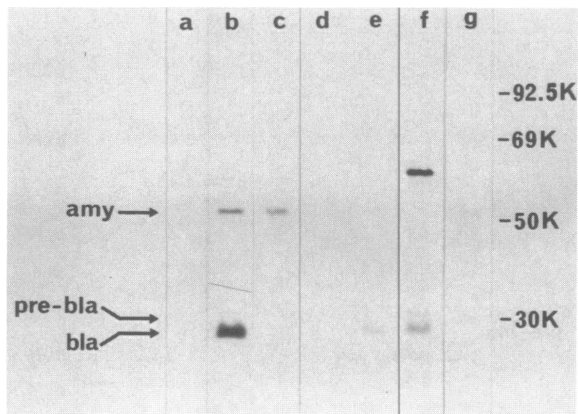

FIG. 4. Cellular fractions of *E. coli* K38(pGP1-2)(pJP3106) and K38(pGP1-2)(pJP5037). K38(pGP1-2)(pJP3106) was fractionated into extracellular (lane a), periplasmic (lane b), and cytoplasm-plus-membrane (lane c) fractions. Lanes d, e, and f are the equivalent fractions of the cytoplasmic control K38(pGP1-2)(pJP5037). Molecular weight controls (lane g) are the same as those in Fig. 3.

codes a cytoplasmic product from *Cellvibrio* sp. (P. Suntinanalerts, unpublished results). The β-lactamase and cytoplasmic gene product served as useful internal controls to monitor the efficiency of osmotic shock and ensure that the cytoplasmic membranes of cells were not disrupted. Cultures (1 ml) of *E. coli* K38(pGP1-2)(pJP3106) and K38(pGP1-2)(pJP5037) were labeled with [^{35}S]methionine for 15 min and immediately fractionated. The amylase gene product was concentrated in the periplasmic and cytoplasmic fractions (Fig. 4, lanes b and c). No amylase gene product was detected in the extracellular fraction (lane a), indicating that the amylase gene product is secreted to the periplasm but is not extracellularly excreted. The apparent accumulation of amylase gene product in the cytoplasmic fraction (Fig. 4, lane c), in contrast to the absence of a similar proportion of mature β-lactamase in this fraction, may indicate that a significant amount of the amylase is trapped in the cytoplasm or cytoplasmic membrane.

Nucleotide sequence of the amylase gene. The nucleotide sequence of the *Sma*I-*Sma*I fragment of pJP3103 (Fig. 1a) was determined by the dideoxy sequencing method of Sanger et al. (28). Subfragments of this clone were inserted in the M13 vectors mp18 and mp19 and sequenced (Fig. 5). A DNA sequence of 2,159 bp was determined in which an open reading frame of 1,392 bp is present (Fig. 6). This open reading frame codes for a protein of 464 amino acids with an M_r of 51,652 and corresponds to the direction of transcription determined above. No other open reading frames sufficient to encode a protein of M_r 49,000 (experimentally determined; Fig. 3) were found.

The deletion analysis data presented above (Fig. 1b) verify that the identified open reading frame indeed codes for the amylase. In particular, a 0.5-kb *Hin*dIII-*Nco*I DNA fragment deleted from pJP3103 after nucleotide sequencing allowed the identification of an unique *Nco*I site 92 bp upstream of the putative amylase promoter. This deletion derivative plasmid, pJP3104, confers an amylolytic phenotype to *E. coli* MM294, confirming that the amylase open reading frame and promoter are uninterrupted by deletion of this fragment and therefore must be encoded within the 1.6 kb proximal to the 3′ end of the insert of pJP3103 (Fig. 1a). In contrast, deletion of nucleotides 40 through 711 of the identified open reading frame (corresponding to the internal *Sph*I-*Sph*I fragment of

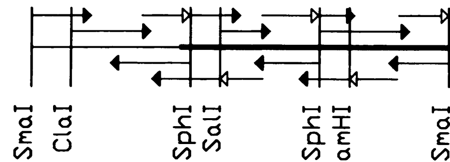

FIG. 5. Strategy used to sequence the insert DNA of pJP3103. Arrows represent the direction and extent of sequencing. Sequences were obtained after priming with the 17-mer universal primer (filled arrowheads) and with 17-mer oligonucleotide primers (open arrowheads) prepared with an Applied Biosystems 380B DNA synthesizer (Applied Biosystems Inc., Foster City, Calif.).

pJP3103) allowed construction of pJP3112, which confers an amylase-negative phenotype to *E. coli* MM294. Similarly, the *Sal*I and *Bam*HI restriction sites that had been determined to interrupt the amylase gene were confirmed to be present within the amylase open reading frame at nucleotide positions 223 through 228 and 876 through 881, respectively. Deletion of the terminal portion of the amylase open reading frame in the region 3′ of an *Xma*III site (nucleotides 1017 through 1022) also results in loss of amylase activity conferred by pJP3103 in *E. coli* MM294 (H. Sakellaris and J. Pemberton, unpublished results).

On the basis of homology to the *E. coli* promoter consensus sequence at positions −35 and −10 (22), a putative promoter was identified in the 5′ region of the sequence, upstream from the amylase initiation codon (Fig. 6). A Shine-Dalgarno sequence (29, 31) AAGGA was located 9 bases upstream of the ATG initiation codon.

The codon usage in the amylase gene is shown in Table 2. To date, no study of codon usage and transfer RNA availability in *A. hydrophila* has been made; thus codon usage in the amylase gene is compared with the data for *E. coli*. The nonrandom use of synonymous codons showed some accord with the rules of Ikemura and Ozeki (17) for *E. coli* nonregulatory genes; however, significant differences were also present. With respect to the use of optimal codons (16), positive bias was shown for these with Phe, Leu, Ile, Thr, Tyr, Gln, Asn, Arg, and Gly; a preference for the use of nonoptimal codons was present with Val, Ala, Pro, Lys, and Glu. Indeed rare codons (2) (Table 2) were predominantly used for Pro (64%). Rare codons were used at a frequency of 13.8% in the entire peptide. Nonoptimal codons (including rare codons) were used at a frequency of 53.2%.

Homology with other amylases. Comparison of the predicted amino acid sequence of the mature *A. hydrophila* amylase with those of amylases from *Aspergillus oryzae*, several *Bacillus* species, and porcine pancreas revealed three regions of homology (Fig. 7). These regions of the *Aspergillus oryzae* amylase have been identified as participants in substrate binding and the active center cleft (21).

Putative signal peptide. Since *A. hydrophila* amylase is an extracellular enzyme (18), it is expected that such a protein would possess an amino-terminal signal to enable its translocation across the cytoplasmic membrane (36). Analysis of the predicted amino acid sequence of the amylase gene nucleotide sequence allowed the identification of a putative signal peptide (Fig. 8). The amino acid sequence deduced from the first 21 codons of the amylase-coding region shows a high level of conformity to other known signal sequences (35). A positively charged amino acid, Lys, is present adjacent to the initiation codon, followed by an interior core of hydrophobic amino acids. A proline residue is present in

```
                      -35                          -10
CAAGTCCGTA CGAGCCAGCT TGAGGATGCT CACAAAATCG CCGTGCCCTG ATTGCCGTCA

      -81                                          SD        -1
GCAAAGGCCA ATGTGAGGCT CTGGGTGCCC ATTCGGGCAC CCAAACAAGG ACAGACAATA

            10          20          30          40          50          60
ATG AAA AAC ACA GCG GGA ATA CTG GCG ATC GCA GGC ATG CTG ATC GCC CCA CTG GCG CAC
Met Lys Asn Thr Ala Gly Ile Leu Ala Ile Ala Gly Met Leu Ile Ala Pro Leu Ala His

            70          80          90         100         110         120
GCC GAC GTC ATA CTG CAC GCC TTC AAC TGG AAA TAC AGC GAG GTC ACC GCC AAG GCC GAC
Ala Asp Val Ile Leu His Ala Phe Asn Trp Lys Tyr Ser Glu Val Thr Ala Lys Ala Asp
     1
           130         140         150         160         170         180
CTC ATC AAG GGA GCG GGA TAC AAG CAG GTA CTC ATC TCG CCG CCC CTC AAG TCT TCG GGC
Leu Ile Lys Gly Ala Gly Tyr Lys Gln Val Leu Ile Ser Pro Pro Leu Lys Ser Ser Gly

           190         200         210         220         230         240
AAC GAG TGG TGG GCC CGT TAT CAG CCC CAG GAT CTG CGA CTC GTC GAC AGT CCG CTG GGC
Asn Glu Trp Trp Ala Arg Tyr Gln Pro Gln Asp Leu Arg Leu Val Asp Ser Pro Leu Gly
                                         50
           250         260         270         280         290         300
AAC AAG CAG GAT CTT GAG CAG CTG ATC GCC GCC ATG CAG GCC CGC GGC ATC GCC GTC TAT
Asn Lys Gln Asp Leu Glu Gln Leu Ile Ala Ala Met Gln Ala Arg Gly Ile Ala Val Tyr

           310         320         330         340         350         360
GCG GAT GTG GTG CTC AAC CAC ATG GCC AAC GAG AGC TGG AAG CGC AAC GAC CTC AAC TAT
Ala Asp Val Val Leu Asn His Met Ala Asn Glu Ser Trp Lys Arg Asn Asp Leu Asn Tyr

           370         380         390         400         410         420
CCT GGC ACT GAA CTG CTA GGA CAA TAC GCT GCC AAC CCT GAT TAC TAC AGC AGA CAA CGA
Pro Gly Thr Glu Leu Leu Gly Gln Tyr Ala Ala Asn Pro Asp Tyr Tyr Ser Arg Gln Arg
100
           430         440         450         460         470         480
CTG TTT GGT GAC CTG GGG CAG AAT TTA CTG TCG GCG TCC GAT TTC CAC CCC GAG GGC TGC
Leu Phe Gly Asp Leu Gly Gln Asn Leu Leu Ser Ala Ser Asp Phe His Pro Glu Gly Cys

           490         500         510         520         530         540
ATC ACC GAT TGG AGC GAT CCC GGC CAT GTC CAG TAC TGG CGC CTG TGT GGC GGC GCC GGC
Ile Thr Asp Trp Ser Asp Pro Gly His Val Gln Tyr Trp Arg Leu Cys Gly Gly Ala Gly
                                         150
           550         560         570         580         590         600
GAC AAG GGG CTG CCG GAT CTT GAC CCC AAC AAC TGG GTG GTG AGC CAG CAA CAA GCC TAT
Asp Lys Gly Leu Pro Asp Leu Asp Pro Asn Asn Trp Val Val Ser Gln Gln Gln Ala Tyr

           610         620         630         640         650         660
CTC AAG GCG CTC AAG GGG ATG GGG ATC AAG GGC TTT CGG GTC GAT GCG GTC AAG CAC ATG
Leu Lys Ala Leu Lys Gly Met Gly Ile Lys Gly Phe Arg Val Asp Ala Val Lys His Met

           670         680         690         700         710         720
AGC GAT TAC CAG ATC AAC GCC GTG TTC ACC CCC GAG ATC AAG CAG GGC ATG CAC GTG TTT
Ser Asp Tyr Gln Ile Asn Ala Val Phe Thr Pro Glu Ile Lys Gln Gly Met His Val Phe
200
           730         740         750         760         770         780
GGG GAG GTC ATC ACC ACG GGG GGC GCC GGC AGC ACG GAT TAC GAG CGC TTC CTC AAG CCC
Gly Glu Val Ile Thr Thr Gly Gly Ala Gly Ser Thr Asp Tyr Glu Arg Phe Leu Lys Pro

           790         800         810         820         830         840
TAC CTC GAC AAC AGC GGC CAG GGT GCC TAT GAC TTC CCG CTG TTC GCC TCC CTG CGC GGG
Tyr Leu Asp Asn Ser Gly Gln Gly Ala Tyr Asp Phe Pro Leu Phe Ala Ser Leu Arg Gly
                                         250
           850         860         870         880         890         900
GCG CTT GGC TAT GGC GGC AGC ATG AAC CAG CTG GCG GAT CCC GGC GCC TAC GGC CAG GCC
Ala Leu Gly Tyr Gly Gly Ser Met Asn Gln Leu Ala Asp Pro Gly Ala Tyr Gly Gln Ala

           910         920         930         940         950         960
CTG CCG GGC AAC CGG GCG GTC ACC TTC GCC ATC ACC CAC GAC ATC CCC ACC AAC GAC GGC
Leu Pro Gly Asn Arg Ala Val Thr Phe Ala Ile Thr His Asp Ile Pro Thr Asn Asp Gly

           970         980         990        1000        1010        1020
TTC CGC TAC CAG ATC CTG AAC CAG ACC GAC GAG AAG CTG GCC TAT GCC TAT CTG CTC GGC
Phe Arg Tyr Gln Ile Leu Asn Gln Thr Asp Glu Lys Leu Ala Tyr Ala Tyr Leu Leu Gly
300
          1030        1040        1050        1060        1070        1080
CGG GAT GGC GGC TCC CCC CTG GTC TAT TCG GAT CAT GGC GAG ACC CAG GAC AAG GAC GGC
Arg Asp Gly Gly Ser Pro Leu Val Tyr Ser Asp His Gly Glu Thr Gln Asp Lys Asp Gly

          1090        1100        1110        1120        1130        1140
CTG CGC TGG CAG GAC TAC TAC CTG AGA AGC GAT CTC AAG GGG ATG ATC CGC TTC CAC AAC
Leu Arg Trp Gln Asp Tyr Tyr Leu Arg Ser Asp Leu Lys Gly Met Ile Arg Phe His Asn
                                         350
          1150        1160        1170        1180        1190        1200
GCC GTG CAG GGT CAG CCC ATG CAG CTC ATC GGC AGC GGC GAC TGC TTC GTG CTG TTC AAA
Ala Val Gln Gly Gln Pro Met Gln Leu Ile Gly Ser Gly Asp Cys Phe Val Leu Phe Lys

          1210        1220        1230        1240        1250        1260
CGC GGC AAG CAG GGG CTG GTC GGC GTC AAC AAG TGT GAC TAC GAG CAG GAG TAC TGG CTC
Arg Gly Lys Gln Gly Leu Val Gly Val Asn Lys Cys Asp Tyr Glu Gln Glu Tyr Trp Leu

          1270        1280        1290        1300        1310        1320
GAC ACC GCC AAG TTC GAG CTG AAC TGG TAT CGC AAC TAC AAG GAC GTG CTG GAT CAA AGT
Asp Thr Ala Lys Phe Glu Leu Asn Trp Tyr Arg Asn Tyr Lys Asp Val Leu Asp Gln Ser
400
          1330        1340        1350        1360        1370        1380
GCC GTC ATC AAT GTG CAG AGC CAG TGG GTG CGG GTC GCC ATG CCG GCG CGC AGG CCG CCT
Ala Val Ile Asn Val Gln Ser Gln Trp Val Arg Val Ala Met Pro Ala Arg Arg Pro Pro

          1390
                        1400       1410
CTG GCT GCT GAA TAA
Leu Ala Ala Glu ---    GCAGGCAAGT ACTTGCCCGG G
            443
```

FIG. 6. Nucleotide sequence of the *A. hydrophila* amylase gene and predicted amino acid sequence of its gene product. The mRNA identical strand is shown. The sequence is numbered commencing at the initiation codon; bases 5′ to the initiation codon are numbered negatively. The −10 and −35 regions of the putative promoter are indicated, as is a Shine-Dalgarno (SD) sequence. Amino acid residues are numbered under appropriate codons, commencing at the first residue of the putative mature peptide.

position −5, and an alanine residue (position −1) precedes the putative processing site.

DISCUSSION

In this paper we report the cloning into *E. coli* of the gene for extracellular amylase from *A. hydrophila* JMP636. Restriction mapping, transcription-translation analysis and nucleotide sequencing allowed us to determine a 1.6-kb fragment of *A. hydrophila* DNA which encodes a single gene, from which the amylase is well expressed by *E. coli*.

The direction of transcription of the amylase gene inserted into pJP3103 (pUC12 recombinant) is convergent with the transcription of the *lacZ* structural gene, indicating that the amylase gene is under the control of its own promoter. This was verified when active amylase was produced by *E. coli* strains harboring pJP3105 or pJP3106, which carry the amylase gene in opposite orientations. In these plasmids the inserted DNA is also preceded by the bacteriophage T7 $\phi 10$ promoter, which is not recognized by *E. coli* RNA polymerase.

The putative amylase gene promoter (Fig. 6) shows good agreement (four of six nucleotides) with the consensus *E. coli* position −35 sequence, whereas the position −10 sequence shows homology with two of six nucleotides of the consensus *E. coli* position −10 sequence. The two homologous nucleotides in the putative amylase position −10 sequence are in fact two of the three most highly conserved bases of the consensus position −10 sequence observed by Hawley and McClure (11). The spacing between the putative position −35 and −10 hexamers is 18 nucleotides and therefore conforms with the spacing of 17 ± 1 nucleotides in the consensus *E. coli* promoter (22). When compared with the consensus *E. coli* promoter, some functional *E. coli* promoters show similar or weaker homology than the putative amylase gene promoter (11); therefore we propose that this promoter may function in vivo.

The bacteriophage T7 RNA polymerase-promoter system of Tabor and Richardson (32; S. Tabor, personal communication) was used for transcription-translation analysis, allowing the determination of the amylase protein molecular weight by subsequent polyacrylamide gel electrophoresis. The M_r of 49,000 determined by this method shows discrepancy with the M_r of 51,652 predicted from the nucleotide sequence. However, this discrepancy may be accounted for

TABLE 2. Codon usage in the *A. hydrophila* amylase gene[a]

Codon	Amino acid	Usage	Codon	Amino acid	Usage	Codon	Amino acid	Usage	Codon	Amino acid	Usage
UUU	Phe	3	UCU	Ser	1	UAU	Tyr	10	UGU	Cys	2
UUU	Phe	3	UCU	Ser	1	UAU	Tyr	10	UGU	Cys	2
*UUC[a]	Phe	12	UCC	Ser	3	*UAC	Tyr	16	UGC	Cys	2
UUA	Leu	1	**UCA**	Ser	0	UAA		1	UGA		0
UUG	Leu	0		Ser	4	UAG		0	UGG	Trp	11
CUU	Leu	3	**CCU**	Pro	3	CAU	His	2	*CGU	Arg	1
CUC	Leu	14	**CCC**	Pro	11	CAC	His	8	*CGC	Arg	11
CUA	Leu	1	*CCA	Pro	1	CAA	Gln	5	**CGA**	Arg	2
*CUG	Leu	28	*CCG	Pro	7	*CAG	Gln	25	**CGG**	Arg	4
AUU	Ile	0	*ACU	Thr	1	AAU	Asn	2	**AGU**	Ser	2
*AUC	Ile	17	*ACC	Thr	10	*AAC	Asn	21	AGC	Ser	12
AUA	Ile	2	**ACA**	Thr	1	*AAA	Lys	3	**AGA**	Arg	2
AUG	Met	11	ACG	Thr	2	AAG	Lys	20	**AGG**	Arg	1
*GUU	Val	0	*GCU	Ala	3	GAU	Asp	16	*GGU	Gly	3
GUC	Val	14	GCC	Ala	27	GAC	Asp	19	*GGC	Gly	31
*GUA	Val	1	*GCA	Ala	1	*GAA	Glu	2	**GGA**	Gly	4
*GUG	Val	11	*GCG	Ala	12	GAG	Glu	13	**GGG**	Gly	9

[a] *, Codons corresponding to the most abundant tRNA species in *E. coli* (16). Codons in boldface type are rare codons whose corresponding tRNA species in *E. coli* occur with an abundance of 0.3 or less on a scale of 0 to 1.0 (17) and whose percentage use is approximately 10% or less (2).

if the amylase protein observed in polyacrylamide gels (Fig. 3 and 4) is a processed form from which a signal peptide has been removed. The putative mature amylase would have a M_r of 49,593 (from the predicted protein sequence), which is in good agreement with the observed M_r.

When the transcription-translation system was coupled with an osmotic shock procedure (12), the amylase was found localized in the periplasmic shock fluid, indicating that although amylase is able to transit the cytoplasmic membrane it is unable to cross the outer membrane of *E. coli*.

When genes encoding extracellular proteins in their native hosts are cloned in *E. coli*, the gene products are not correctly localized but are accumulated in the periplasm (3, 14). We have shown that the *A. hydrophila* amylase is no exception; clearly the amylase protein does not contain sufficient information in either the signal peptide or the mature protein to ensure translocation across the outer membrane of *E. coli*.

Recent studies have shown that internal deletions of 40 amino acids within the first 84 amino acids of the mature OmpA protein of *E. coli* do not interfere with its correct localization; however deletions of a greater size and further into the protein do (7). A similar situation may exist with respect to the localization of other *E. coli* proteins and the exported proteins of other gram-negative bacteria. In another study, the signal sequences of 43 exported *E. coli* proteins (including OmpA) with different final localizations were subjected to multivariate analysis by measuring the physicochemical profile of each. Signal sequences of proteins with common final localizations were grouped closely together yet were distinct from the physicochemical profiles of signal sequences of proteins destined for alternative final localizations (30). Considering these lines of evidence it is plausible that both the signal sequence and the mature protein interact, in a manner as yet undetermined, to direct and facilitate the translocation and correct localization of exported proteins.

The putative signal peptide of amylase (Fig. 8), deduced from the nucleotide sequence of the cloned gene, shows a number of features that are common to other procaryotic signal sequences (30, 36). A positively charged amino acid (Lys) is present adjacent to the methionine residue at the NH_2 terminal, followed by a hydrophobic core in which 6 of 10 amino acids are hydrophobic and a secondary structure-disrupting proline residue in position −5 (Fig. 8).

On the basis of an analysis of 65 eucaryotic and 20 procaryotic signal sequences, von Heinje (36) proposed that cleavage of signal sequences occurs at specific sites that are defined by the presence of preferred amino acids. Cleavage sites must have small neutral amino acids in positions −3 and −1 (where the signal peptidase cleaves between positions −1 and +1), and an aromatic, charged, or large polar residue must be present in position −2. The putative signal peptide of *A. hydrophila* amylase conforms to each of these

```
REGION 1                 116                 120
Taka amylase A           Val - Asp - Val - Val - Ala - Asn - His - Met - Gly - Tyr - Asp - Gly - Ala
                          95                       100
Porcine pancreas I       Val - Asp - Ala - Val - Ile - Asn - His - Met - Cys - Gly - Ser - Gly - Ala
                          96                 100
B. subtilis              Val - Asp - Ala - Val - Ile - Asn - His - Thr - Thr - Phe - Asp - Tyr - Ala
                         100                                                   110
B. stearothermophilus    Ala - Asp - Val - Val - Phe - Asp - His - Lys - Gly - Gly - Ala - Asp - Gly
                          97             100
B. amyloliquefaciens     Gly - Asp - Val - Val - Leu - Asn - His - Lys - Ala - Gly - Ala - Asp - Ala
                          80                                                   90
A. hydrophila            Ala - Asp - Val - Val - Leu - Asn - His - Met - Ala - Asn - Glu - Ser - Trp

REGION 2                 200                                                   210
Taka amylase A           Ile - Asp - Gly - Leu - Arg - Ile - Asp - Thr - Val - Lys - His
                         191                                             200
Porcine pancreas I       Val - Ala - Gly - Phe - Arg - Ile - Asp - Ala - Ser - Lys - His
                         170                                                   180
B. subtilis              Ala - Asp - Gly - Phe - Arg - Phe - Asp - Ala - Ala - Lys - His
                         228         230
B. stearothermophilus    Ile - Asp - Gly - Phe - Arg - Leu - Asp - Gly - Leu - Lys - His
                         225                     230
B. amyloliquefaciens     Leu - Asp - Gly - Phe - Arg - Ile - Asp - Ala - Ala - Lys - His
                         188                                                   198
A. hydrophila            Ile - Lys - Gly - Phe - Arg - Val - Asp - Ala - Val - Lys - His

REGION 3                 291                                                               300
Taka amylase A           Thr - Phe - Val - Glu - Asn - His - Asp - xxx - Asn - Pro - xxx - xxx - Arg
                         294                           300
Porcine pancreas         Val - Phe - Val - Asp - Asn - His - Asp - xxx - Asn - Gln - xxx - xxx - Arg
                         325                     330
B. stearothermophilus    Thr - Phe - Val - Asp - Asn - His - Asp - Thr - Asn - Pro - Ala - Lys - Arg
                         322                                 330
B. amyloliquefaciens     Thr - Phe - Val - Glu - Asn - His - Asp - Thr - Gln - Pro - Gly - Glu - Ser
                         287             290
A. hydrophila            Thr - Phe - Ala - Ile - Thr - His - Asp - Ile - xxx - Pro - Thr - Asn - Asp
```

FIG. 7. Comparison of three regions of the *A. hydrophila* amylase amino acid sequence with homologous regions in α-amylases from *Aspergillus oryzae* (Taka amylase), Porcine pancreas, *Bacillus subtilis*, *B. stearothermophilus*, and *B. amyloliquefaciens* (15). Homologous sequences are underlined. Numbers indicate the amino acid residue position in each enzyme.

															−5	−4	−3	−2	−1	+1	+2	+3	+4	+5
MET	Lys	Asn	Thr	Ala	Gly	Ile	Leu	Ala	Ile	Ala	Gly	Met	Leu	Ile	Ala	Pro	Leu	Ala	His	Ala	Asp	Val	Ile	Leu	His

← HYDROPHOBIC CORE → ↑

FIG. 8. Putative signal peptide of the *A. hydrophila* amylase precursor. Numbers above the amino acid residues indicate their positions relative to the processing site (36), which is represented by a vertical arrow.

criteria; indeed positions −3 and −1 are filled by Ala residues, which may fit the signal peptidase active site better than any other residue (36).

Using X-ray crystallographic analysis, Matsuura et al. (21) were able to show that the amino acids His-210 and Asp-297 are located in the active center cleft of Taka-amylase A from *Aspergillus oryzae*. The amino acids His-122, His-296, and Asp-297 participate in substrate binding by the same enzyme. Ihara et al. (15) recently used these data to show that α-amylases from *Aspergillus oryzae*, *Bacillus subtilis*, *B. stearothermophilus*, *B. amyloliquefaciens*, *B. licheniformis*, and porcine pancreas possess substantial homology with amino acids that comprise and flank the enzymatically functional regions described above. Similarly, we have found that the predicted amino acid sequence of *A. hydrophila* amylase possesses substantial homology with the same three functional regions. Our data are consistent with the suggestion that active centers of procaryotic amylases may be the same as those of eucaryotic amylases (15).

ACKNOWLEDGMENTS

We gratefully acknowledge research support from the Australian Research Grants Committee (grant D2 85 1822) and the Mayne Bequest Fund. K.S.G. holds the Frank Nicklin Postgraduate Scholarship from the University of Queensland.

We thank S. Tabor for kindly supplying bacterial strains and plasmids, V. Cochrane for typing the manuscript, and A. Al-Zaag and I. Beacham for valuable advice about nucleotide sequencing.

LITERATURE CITED

1. **Bolivar, F., R. Rodriguez, P. J. Greene, M. C. Betlack, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction of new cloning vehicles. II. A multiple cloning system. Gene **2**:95–113.
2. **Burns, D. M., and I. R. Beacham.** 1986. Nucleotide sequence and transcriptional analysis of the *E. coli ushA* gene, encoding periplasmic UDP-sugar hydrolase (5′-nucleotidase): regulation of the *ushA* gene, and the signal sequence of its encoded protein product. Nucleic Acids Res. **14**:4325–4342.
3. **Chakraborty, T., B. Huhle, H. Bergbauer, and W. Goebel.** 1986. Cloning, expression and mapping of the *Aeromonas hydrophila* aerolysin gene determinant in *Escherichia coli* K-12. J. Bacteriol. **167**:368–374.
4. **Chakraborty, T., M. A. Montenegro, S. C. Sanyal, R. Helmuth, E. Bulling, and K. N. Timmis.** 1984. Cloning of enterotoxin gene from *Aeromonas hydrophila* provides conclusive evidence of production of a cytotoxic enterotoxin. Infect. Immun. **46**:435–441.
5. **Cohen, S. N., A. C. Y. Chang, and L. Hsu.** 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *E. coli* by R-factor DNA. Proc. Natl. Acad. Sci. USA **69**:2110–2114.
6. **Dunn, J. J., and F. W. Studier.** 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the location of T7 genetic elements. J. Mol. Biol. **166**:477–535.
7. **Freudl, R., H. Schwarz, M. Klose, N. R. Movva, and U. Henning.** 1985. The nature of information, required for export and sorting, present within the outer membrane protein OmpA of *Escherichia coli* K-12. EMBO J. **4**:3593–3598.
8. **Gobius, K. S., and J. M. Pemberton.** 1986. Use of plasmid pULB113 (RP4::Mini-Mu) to construct a genomic map of *Aeromonas hydrophila*. Curr. Microbiol. **13**:111–115.
9. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
10. **Hansen, J. B., and R. H. Olsen.** 1978. Isolation of large plasmids and characterization of the P2 incompatibility group plasmids pMG1 and pMG5. J. Bacteriol. **153**:227–238.
11. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
12. **Heppel, L. A.** 1971. Osmotic shock, p. 226–227. *In* L. I. Rothfield (ed.), Structure and function of biological membranes. Academic Pres, Inc., New York.
13. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11**:291–298.
14. **Howard, S. P., and J. T. Buckley.** 1986. Molecular cloning and expression in *Escherichia coli* of the structural gene for the hemolytic toxin aerolysin from *Aeromonas hydrophila*. Mol. Gen. Genet. **204**:289–295.
15. **Ihara, H., T. Sasaki, A. Tsuboi, H. Yamagata, N. Tsukagoshi, and S. Udaka.** 1985. Complete nucleotide sequence of a thermophilic amylase gene: homology between prokaryotic and eukaryotic α-amylases at the active sites. J. Biochem. **98**:95–103.
16. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNA's and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151**:389–409.
17. **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.
18. **Janda, J. M.** 1985. Biochemical and exoenzymatic properties of *Aeromonas* species. Diagn. Microbiol. Infect. Dis. **3**:223–232.
19. **Kaper, J. B., H. Lockman, R. R. Colwell, and S. W. Joseph.** 1981. *Aeromonas hydrophila*: ecology and toxigenicity of isolates from an estuary. J. Appl. Bacteriol. **50**:359–377.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
21. **Matsuura, Y., M. Kusunoki, W. Harada, and M. Kakudo.** 1984. Structure and possible catalytic residues of Taka-Amylase A. J. Biochem. **95**:697–702.
22. **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54**:171–204.
23. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
24. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 431–433. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Naumovski, L., and E. C. Friedberg.** 1982. Molecular cloning of eukaryotic genes required for excision repair of UV-irradiated DNA: isolation and partial characterization of the *RAD3* gene of *Saccharomyces cerevisiae*. J. Bacteriol. **152**:323–331.
26. **Priefer, U., R. Simon, and A. Puhler.** 1984. Cloning with cosmids, p. 190–201. *In* A. Puhler and K. N. Timmis (ed.), Advanced molecular genetics. Springer-Verlag, Berlin.
27. **Russel, M., and P. Model.** 1984. Replacement of the *fip* gene of *Escherichia coli* by an inactive gene cloned on a plasmid. J. Bacteriol. **159**:1034–1039.
28. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci.

USA **71**:1342–1346.

30. **Sjöström, M., S. Wold, A. Wieslander, and L. Rilfors.** 1987. Signal peptide amino acid sequences in *Escherichia coli* contain information related to final protein localization. A multivariate data analysis. EMBO. J. **6**:823–831.
31. **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.
32. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
33. **Talmadge, K., and W. Gilbert.** 1980. Construction of plasmid vectors with unique *Pst*I cloning sites in a signal sequence coding region. Gene **12**:235–241.
34. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
35. **von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133**:17–21.
36. **von Heijne, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173**:243–251.