Identification of a New Operon Involved In Listeria monocytogenes Virulence: Its First Gene Encodes a Protein Homologous to Bacterial Metalloproteases

JEROME MENGAUD,¹ CHRISTIANE GEOFFROY,² and PASCALE COSSART¹*

Unité de Génie Microbiologique¹ and Unité des Antigènes Bactériens,² Institut Pasteur, 25 Rue du Docteur Roux, Paris 75015, France

Received 20 August 1990/Accepted 6 December 1990

The region flanking the transposon in a Tn1545-induced lecithinase-negative mutant of Listeria monocytogenes EGD was cloned and sequenced. The transposon had inserted in ORF D, the open reading frame previously identified downstream from hlyA, the gene encoding listeriolysin O. The complete sequence of ORF D from strain EGD has been determined as well as that of two other strains: LO28, a clinical isolate; and LM8, an epidemic strain. ORF D is 1,533 bp long and encodes a protein highly homologous to metalloproteases of bacilli, Serratia sp., Legionella pneumophila, and Pseudomonas aeruginosa. It was renamed prtA. Northern RNA blot analysis indicated that prtA is the first gene of a 6-kb operon, suggesting that the lecithinase-negative phenotype of the mutant might be due to a polar effect of the transposon insertion.

The gram-positive bacterium *Listeria monocytogenes* is an intracellular pathogen discovered in 1926 during an epidemic among laboratory rabbits and guinea pigs (18) and later shown to be a dangerous pathogen for humans, in particular pregnant women, neonates, and immunocompromised people. It also infects healthy people and has been in recent years the cause of several outbreaks traced to contaminated food and resulting in septicemias, meningitis and meningoencephalitis, and death.

After the studies of Mackaness in 1960 (13), virulence of L. monocytogenes was essentially attributed to its capacity to survive and even replicate in professional phagocytes. L. monocytogenes, in addition to its capacity to survive intracellularly, is able to cross the intestinal barrier, to spread from cell to cell, to multiply in the liver and spleen (where it creates foci of infection and abscesses), and to cross the endothelial barriers and gain access to the brain and the placenta. Genetic studies to identify the various factors responsible for these different steps recently converged to establish the crucial role of the thiol-activated toxin produced by L. monocytogenes. This protein, named listeriolysin O (LLO), allows the escape of the bacterium from the phagosome into the cytosol. The crucial demonstration that LLO is a virulence factor was obtained through a gene complementation experiment: transformation of a transposon-induced nonhemolytic mutant with a plasmid carrying hlyA, the gene encoding LLO, led to the concomitant recovery of the hemolytic phenotype and of virulence (8). LLO is, to date, the only bacterial factor identified as an intracellular growth-promoting factor and has even been shown to promote growth of Bacillus subtilis in J774 macrophages (3).

LLO was the first factor to be examined not only because clinical strains are hemolytic and all nonhemolytic strains are avirulent in the mouse model but also because the hemolytic phenotype is easily identified on blood agar plates. Another phenotype easily identified on agar plates is the lecithinase phenotype. Secretion of a lecithinase by L. *monocytogenes* was reported for the first time by Seeliger, who detected halos around colonies on egg yolk plates (22). We recently questioned the role of this virulence factor. Purification of the protein was achieved, and four lecithinase-negative mutants were isolated by independent Tn1545, Tn917, and Tn917-lac mutagenesis. These mutants carried a single copy of the transposon inserted at different loci on the chromosome and their virulence, evaluated by their 50% lethal dose, was affected to different extents (20a and our unpublished results).

In this paper, we report the cloning and sequencing of the DNA flanking Tn1545 in one of the mutants. Unexpectedly, the locus of insertion in the *Listeria* chromosome mapped in ORF D, an open reading frame previously identified downstream from hlyA, the gene coding for LLO (15). The wild-type ORF D was cloned and sequenced from three different strains. It encodes a protein homologous to neutral metalloproteases of bacilli, *Legionella pneumophila*, *Pseudomonas aeruginosa*, and *Serratia* sp. and was renamed *prtA*. Northern RNA analysis indicated that this gene is the first gene of a 6-kb operon, suggesting that a polar effect might explain the lecithinase-negative phenotype of the Tn1545-induced mutant.

MATERIALS AND METHODS

Strains, plasmids, and culture media. The L. monocytogenes strains used in this study were LO28, a wild-type isolate (29); LM8, an epidemic strain (20); and Bug 13, a Tn1545-induced lecithinase-negative mutant of strain EGD (20a). The EGD strain we used is the spontaneous streptomycin-resistant derivative described previously (9). Plasmid pUC18 (32) was used to clone DNA fragments in Escherichia coli MC1061 [F⁻ araD139 Δ (ara leu)7696 Δ lacY74 galU galK hsr hsm strA] (6). E. coli strains were grown at 37°C in LB medium (14). For strains containing pUC derivatives, ampicillin was added at a final concentration of 25 µg/ml in liquid medium and 100 µg/ml in solid medium. L. monocytogenes strains were grown in brain heart infusion (BHI) broth or agar (Difco Laboratories, Detroit, Mich.) at 37°C with aeration.

^{*} Corresponding author.

Chemicals and enzymes. Restriction enzymes and ligase were purchased from Amersham (Buckinghampshire, United Kingdom), Boehringer (Mannheim, Federal Republic of Germany), BioLabs (Beverly, Mass.), Gibco BRL (Cergy Pontoise, France), or Genofit (Geneva, Switzerland) and were used as recommended by the manufacturer. [³⁵S]dATP (3,000 Ci/mmol) was purchased from Amersham.

DNA techniques. DNA techniques and plasmid DNA sequencing were performed as previously described (14, 16). Polymerase chain reaction amplification was performed on chromosomal DNA from strain EGD or LM8 by using two pairs of oligonucleotides chosen from the sequence obtained from strain LO28. The amplified fragments were purified by use of a GeneClean kit (Bio 101, La Jolla, Calif.), after electrophoresis on agarose gels, and sequenced as previously described (16). Listeria chromosomal DNA was purified by a rapid protocol adapted from the method of Ausubel et al. (1). Briefly, cells from 1.5 ml of overnight culture were washed in $0.1 \times$ SSC (1 \times SSC is 0.15 M sodium chloride-0.015 M sodium citrate), suspended in 60 μ l of lysis solution (0.01 M Na₂HPO₄, 20% sucrose, 2.5 mg of lysosyme per ml [freshly added]) and incubated for 1 h at 37°C. Then 507 µl of TE (10 mM Tris HCl, 1 mM EDTA), 30 µl of 10% sodium dodecyl sulfate, and 3 µl of proteinase K (20 mg/ml) were added before a second hour of incubation at 37°C. Cell wall debris, denatured proteins, and polysaccharides were precipitated by adding 100 µl of 5 M sodium chloride and 80 µl of 10% hexadecyltrimethyl ammonium bromide (Sigma, St. Louis, Mo.) in 0.7 M NaCl and incubating for 10 min at 65°C. The mixture was then phenol extracted twice, and DNA was precipitated by adding 0.6 volume of isopropanol and gently inverting the tube. The DNA pellet recovered by brief centrifugation was washed in 70% ethanol, dried, and suspended in 100 µl of TE containing 50 µg of RNase per ml.

RNA techniques. Total cellular RNAs were extracted by a rapid protocol from L. monocytogenes LO28 grown at 37°C with aeration in BHI broth (1). Briefly, cells from a 10-ml culture were suspended in 500 µl of lysis buffer (30 mM Tris HCl [pH 7.4], 100 mM sodium chloride, 5 mM EDTA, 1% sodium dodecyl sulfate, 100 µg of proteinase K per ml [freshly added]), frozen on dry ice, thawed, and sonicated three times for 10 s with a microtip sonicator (power setting, about 30 W). The lysate was then incubated for 1 h at 37°C, phenol extracted twice, and chloroform extracted. The aqueous phase was ethanol precipitated overnight at -20° C. The total RNA and DNA pellet was suspended in 100 µl of DNase digestion buffer (40 mM Tris HCl [pH 7.5], 6 mM MgCl₂) containing 30 U of RNase-free DNase I (Pharmacia) and incubated for 1 h at 37°C. The mixture was phenol extracted, chloroform extracted, and ethanol precipitated overnight at -20° C. The RNA pellet was finally suspended in 100 µl of diethylpyrocarbonate-treated water and stored at -70°C. Yields (about 10 µg of total RNA per ml of culture) are higher than with the technique previously described (15). Northern blotting was performed as described previously (15), except that we used Immobilon N membranes (Millipore) and the rapid hybridization system of Amersham.

RESULTS

To examine the role of lecithinase in virulence of L. monocytogenes, lecithinase-negative mutants were obtained by transposon mutagenesis with Tn1545, Tn917, and Tn917-lac (20a and our unpublished results). In this study, one of these was further analyzed. It contains a single copy of

Tn1545 inserted in the chromosome and has a higher 50% lethal dose than the parental strain, EGD $(10^{7.6} \text{ versus } 10^{6.2})$.

Cloning and sequence analysis of the chromosome-Tn1545 junction of a lecithinase-negative mutant: Tn1545 had inserted in ORF D. Taking advantage of the presence of a kanamycin resistance gene at the left end of Tn1545, we cloned, in pUC18, a 8-kb *Hin*dIII fragment containing the junction between the listerial chromosome and the left part of Tn1545. Using an oligonucleotide located in the left end of Tn1545, we sequenced the transposon-*Listeria* junction. Tn1545 had inserted 530 bp downstream from the beginning of an ORF previously identified immediately downstream from hlyA the structural gene of the LLO and named ORF D. For reasons described below we designated this ORF *prtA*.

Cloning of prtA from L. monocytogenes EGD, LO28, and LM8 and nucleotide sequence comparison. As most of our genetic work on virulence of L. monocytogenes was performed on the clinical strain LO28, we decided to clone and sequence the wild-type prtA gene from strain LO28 as well as that from strain EGD. Strains EGD and LO28, respectively, are of the 1/2a and 1/2c serotype. We had previously noticed a restriction fragment length polymorphism in this region, and therefore we also cloned the region from epidemic strain, LM8, of serotype 4b. In this strain, the second HindIII site located downstream from hlyA is absent (10; unpublished results).

We had previously cloned, in pLis3, a 3,454-bp BamHI chromosomal fragment containing the hlyA gene and the beginning of prtA of strain LO28 (Fig. 1) (15). Using as a probe a 531-bp Sau96-A-BamHI fragment internal to prtA and purified from pLis3, we cloned, in pUC18, a 1,369-bp HindIII fragment (pLis11). We then used as a probe a 455-bp SphI fragment purified from pLis11 to clone a 1,591-bp DraI fragment, giving rise to plasmid pLis22, which contained the 3' end of prtA.

For strain EGD, we also used the 455-bp *SphI* fragment as a probe to clone, in pUC18, a 1,250-bp *SphI* fragment (pLis33). The 1,369-bp *Hin*dIII fragment harbored by plasmid pLis32 was cloned by using as a probe a 1-kb *AccI* fragment internal to *prtA* and purified from the plasmid containing the transposon-*Listeria* junction.

Finally, for strain LM8, with a 412-bp *Hind*III fragment purified from plasmid pLis3 as a probe, *prtA* was nearly entirely cloned in a 1,781-bp *Hind*III fragment (pLis18).

By sequencing on both strands of the DNA the fragments cloned in pLis3, pLis11, and pLis22, we determined the complete nucleotide sequence of *prtA* from strain LO28 (Fig. 2). Inserts of pLis32, pLis33, and pLis18 only partially covered *prtA* of strains EGD and LM8, respectively. To complete the nucleotide sequence of *prtA* of these strains, oligonucleotides were chosen in the sequence of *prtA* region of strain LO28 and used to amplify, by polymerase chain reaction, DNA fragments spanning the 5' and 3' ends of the *prtA* genes from strains EGD and LO28, respectively (Fig. 1). These fragments were directly sequenced.

Sequences of the *prtA* genes of strains LO28, EGD, and LM8 are given in Fig. 2. The *prtA* gene, from the ATG start codon arbitrarily chosen as discussed previously (15) up to the TGA stop codon, is 1,533 bp long. The *prtA* nucleotide content is 38% (G+C). Taking strain LO28 as a reference, only 6 differences were detected in the *prtA* sequence of strain EGD versus 79 differences in the case of strain LM8. Most of the changes affect the third position of codons and either are conservative or change an amino acid for a similar one. The mutation responsible for the restriction fragment length polymorphism observed in the *prtA* region corre-



FIG. 1. Schematic drawing and partial restriction map of the *prtA* gene region from strain LO28. The *hlyA* and *prtA* genes are represented by arrows. Restriction sites: H, *Hin*dIII; B, *Bam*HI, S, *Sph*I, D, *Dra*I. Overlapping restriction DNA fragments from strains LO28, EGD, and LM8, cloned in pUC18 and spanning the *prtA* gene, are represented by thick lines. The names of the pUC derivatives containing these fragments are indicated below the lines. Fragments obtained by polymerase chain reaction amplification for direct sequencing of the 5' and 3' parts of the *prtA* gene from strain EGD and LM8 are reported with no restriction site indication.

sponds to a G-to-A transition in the *Hin*dIII site in position 139 in LM8.

prtA encodes a protein homologous to zinc metalloproteases of the "thermolysin family." The polypeptide encoded by prtA would be 510 amino acids long and have a calculated molecular weight of 57,375. The N-terminal part of the encoded protein has all the characteristics of a signal sequence of a gram-positive bacterium (30). A polypeptide homology search (12) in the NBRF data base revealed substantial similarity between the prtA-encoded protein and the amino acid sequences of metalloproteases from Bacillus stearothermophilus (26), Bacillus thermoproteolyticus (27), Bacillus cereus (23), B. subtilis (31), and Bacillus amyloliquefaciens (28). The prototype of this family of structurally related neutral proteases is thermolysin, the protease produced by B. thermoproteolyticus. Interestingly, two other proteins produced by pathogenic gram-negative bacteria, the major secreted protein (MSP) of L. pneumophila (4) and the elastase of P. aeruginosa (2), also belong to this family.

In all cases where the gene has been cloned and sequenced, a discrepancy between the gene size and the mature protein length was observed, indicating that these proteases are processed from a proprotease into an active form (Fig. 3B). On the basis of protein sequence comparison, the N-terminal amino acid of the Listeria mature protease would be located around position 204 (Fig. 2). The mature form of the protease would have a molecular weight of 34,522. Similarities between the putative Listeria protease and the other proteases are spread along the whole sequence and are stronger when the prtA-encoded protein is compared with proteases from gram-positive bacteria than when it is compared with proteases from gram-negative bacteria. In Fig. 3A, sequences of the putative mature form of the Listeria protease (positions 204 to 510) and mature forms of Bacillus proteases were aligned (21). Percentages of identity calculated on the basis of this alignment are 35% with B.

cereus, 40% with B. thermoproteolyticus and B. stearothermophilus, and 47% with B. subtilis and B. amyloliquefaciens. For L. pneumophila, P. aeruginosa, or Serratia sp., the percentage of identity is around 25 to 30%.

The X-ray structure of one of these proteases, thermolysin, has been solved and shown to contain one zinc atom bound to His-142, His-146, and Glu-166. The active site is located around His-231, and two additional sites required for catalysis are located at Glu-143 and Asp-226 (7, 11). The regions from amino acids 346 to 355, 368 to 377, and 428 to 448 of the Listeria protease are very similar to the zincbinding or active site regions of thermolysin (amino acids 138 to 148, 161 to 170, and 222 to 242) (Fig. 3C). On the basis of these similarities, the predicted zinc-binding amino acids of the prtA-encoded protein would be His-349, His-353, and Glu-373, and the predicted active site would be His-437. Glu-350 and Asp-432 might also be essential for activity of the protein, by analogy with Glu-143 and Asp-226 of thermolysin. Three zinc ligands and one active site were also predicted for B. subtilis (31), Serratia sp. (19), P. aeruginosa (2), and L. pneumophila (4) proteases on the basis of their comparison with thermolysin. The predicted zinc-binding regions of prtA-encoded protein are also similar to the zinc-binding regions of these eight other metalloproteases. Recently, sequences of the gp63 metalloproteases produced by the parasitic protozoa Leshmania chagasi and L. major have been reported (17). They contain the conserved zincbinding domain, but the overall similarity with bacterial metalloproteases, including the putative Listeria protease, is very weak (Fig. 3C).

prtA is the first gene of an operon. To evaluate the role of *prtA* in virulence and determine whether the insertion of Tn1545 had a polar effect on adjacent genes, we studied the structural and transcriptional organization of the *prtA* gene region. As previously reported, the upstream region encodes LLO and is transcribed as a monocistronic unit (15). The

EGD

| , ~ 3 | | | G | | | | | | | | | | | | | | | | | | λ | | |
|--------------|----------------|------------|-------|------|------|------|-------------|------------|-------------|------------|-------------|--------|--------|------------|-------------|----------|------------|--------|--------|-------|------|-------|-----|
| LO28 | ATG | ۸A | хX | GT | YYY | CTT | ATT | TGI | XT | C A | тси | ATG | GTA | ата | GCT | TTT | CAG | GCT | CAT | TTC | ACT | ATG | 60 |
| 1 | Het. | ي ل | a . s | er | lya | leu | ile | CV | L11 | <u>e i</u> | ر ها | met. | val | ile. | <u>مام</u> | phe | ملع | هله. | his | phe | thr | met. | |
| | | | | | | | | | | | | | | | | | | | | | asn | | |
| | - | | | | | | | | | | | | | | | | | _ | | | | - | |
| | G | _ | | | G | | _ | _ | | | | | GG | | G | | | T | ~ | | | G | |
| | ACG | GI | × 7 | w | GCA | GAT | TCT | GIG | GG | GG | - | GAA | *** | CTT | CAA | AAT | AAT | ACA | CAA | GCC | *** | AAG | 120 |
| 21 | thr | YA | | V. | إغلم | asp | ser | va. | gl | A d | 10 | gru | туз | Ten | متھ | asn | asn | unr. | g ru | aia | TÀS | TÀR | |
| | ala | | | | - 4 | | | | | | | | arg | | arg | | | 110 | | | | arg | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | . <u>~</u> | ~ | | ~ | ~~~ | - | | . . | | ~ ~ | ~ | ~~~ | - | *** | ~ | ~ | | - | - | - | | 117 | 100 |
| 41 | ACC. | . u | T | SCT. | GAT | TTA | | | 1 11 | | | GAT | TUL | TGC | GAA. | -1- | 1 | GAI | 777 | 141 | 1.00 | | 190 |
| -1 | 482 | PI | | | asb | 160 | - 44 | th | 4 | a p | 20 | eeb | 941 | -74 | 910 | - | -14 | Teb | Pare . | -12 | ~1.4 | 4.011 | |
| | | | | | | | | | • | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | TTT | N | N I | TTA | CTT | GAI | ATC | AC | N N3 | A G | AT . | ANG | CTA | GGC | GTT | ACG | CAT | TAT | ACG | CTC | GCCG | CTA | 240 |
| 61 | phe | 1 | /8 | i le | leu | asp | met | : th | r 13 | 78 a | sp | lys | leu | gly | val | thr | his | tyr | thr | leu | ala | leu | |
| | - | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | λ | | | | | r | | | | | λ | | | | | | | | | |
| | MG | T | т | GGT | GGT | TAC | TT | 3 AO | G | AT J | TA | GAT | GAA | ATC | MG | GTC | CAC | GTC | ACA | œ | GAT | AAT | 300 |
| 81 | 801 | C 84 | er - | gly | gly | ty | r le | u th | r a | sp 4 | ISU | asp | glu | 110 | lys | val | h18 | val | thr | pro | asp | 880 | |
| | | | | asp | | | | | | | | | | | | | | | | | | | |
| | | | | | - | | | | | | | | | c | | | | | | | | | |
| | | | TC | ACT | TT | - | | TGO | a (* | NT 1 | TIG | cac | CAA | GGA | CA-C | CTT | AGG | ATT | ACT | лат | CAA | ATA | 360 |
| 10 | 11/1 | . i | 1. | thr | phe | 11 | | | V A | SD 1 | leu | aln | gln | alv | aln | leu | arg | ile | thr | asn | gln | ile | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | с | | | | | | | т | | | | | | | т | | | | | | |
| | λλi | X X | тт | YCY | GN | | * ** | TGC | TA | TC (| GAA | *** | GCA | TTT | GAA | œc | ATC | GGT | CAG | AGT | GAA | GCT | 420 |
| 12 | 1 1y: | s 1 | 10 | thr | glu | ı ly | | n al | a i | 1. | glu | lys | ala | phe | glu | ala | ile | gly | gln | 80I | glu | ala | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | _ | | - | | | | | | | | | | | | | |
| | ~ | | ~ | | | | . <u>^</u> | T ~ | | T | ~ | ~ | | ~ | | ~ | | | A | | * | 107 | 480 |
| | САС 1 Б-2 | TG | aG | 200 | AG | TA | I GI | COC | | AC I | | GIG | 1 | GAA 0.1 | 1 | . GAA | ATC . | 41- | | AAT | 100 | Aur | -80 |
| - 14 | - 11 | | 41 | TÀI | 3-61 | | ⊥ V38 i1 | - 91 | | an j | pro | Val | TAR | d ra | 1.14 | gru | 116 | 110 | 110 | e a D | | arg | |
| | | | | | | | | - | | | | | | | | | | | | | | | |
| | | | | | | | x | | | • | | | | | | | | | | | | | |
| | AC | λ, | ~ | œ | TT | N GT | T 12 | TA | XT A | тĂ | 7 77 | TTG | ATT | 111 | CT CT | GAG | : 000 | GN | GTT | GCCG | AGT | TGG | 540 |
| 16 | 1 th | ur Ì | lys | arg | ; le | u va | 1 ty | rr al | n i | 10 | lys | leu | ile | phe | | glu | pre | glu | val | ala | 80I | trp | |
| | | | - | | - | | | | | | 2 | | | - | | - | - | | | | | - | |
| | | | | | | | ¢ | : | | | | | | | | | | | | | | | |
| | | . 1 | ۱. | | | | т | : | | | | | | _ | | | ۱ | | | | | | |
| | AT | T (| зтŢ | CN | A GT | ເຜ | c a | CA G | N 1 | œ | ŝ | e con | ATI | TT2 | Ŵ | | ; ch | * *** | ATG | CTI | TCC | GAG | 600 |
| 18 | 11 11 | | al | д'n | n va | 1 84 | ib Ai | LL G. | iu t | hr | gly | - 14 | 114 | 100 | 1 194 | 1 1 1 1 | s gir | | n met | leu | 802 | glu | |
| | | - | 110 | | | | a | Lat | | | | | | | | | | | | | | | |
| | | | c | | | | | | | | | | | | | | | | | | | | |
| | C ¹ | - | ເມັ | ŝ | 6 60 | тс | T M | c c | AC 1 | ~ | GAT | | CN | | s cm | r G(7 | | د م | مم | | : @ | TIG | 660 |
| 20 |)1 va | | alu | AD | a al | | n ti | hr h | in 1 | lvs. | 880 | oh | alt | ala | lei | | 1 1 1 | al | ala | 880 | ATC | leu | |
| | | | | | | 1 | | | | | | | , | | | | | | | | | | |
| | | | | | | T | | | с | | | | | | | | | | | | | | |
| | | | | | | Ť | | | с | | | 1 | ۱. | | | 3 | | | | | | | |
| | C. | rc | CAG | AG | Gα | AC | CA C | AT G | TT I | ATG | XXX | AT | 1 AA: | GA | сти | A TT | T TM | TT | A GTO | GA | AG | ACT | 720 |
| 2 | 21 10 | eu | gln | az | g pi | 0 1 | eu h | is v | al : | net | lys | i i le | | n asj | p len |) ph | a ty: | r les | a val | as | arg | , thr | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | _ | | | _ | _ | | _ | | | |
| | - | | | _ | | | | | | | | _ | | - | 1 | <u> </u> | G | G | N 1 | r | 1 | · | |
| | . ° | AT | ~~~ | | A CI | C A | IA A | GAN | CT ' | TTT | GA1 | TE | | V CV | TAN | I AO | CGN | C AC | r TO | s TT | GGG | ÷ λλλ | 780 |
| 2 | 41 P. | 18 | 174 | g1 | Y 14 | nı i | 10 B | rg t | nr | phe | asp | 5 Tel | a 19 | s hi: | 5 8.81 1 | n th: | r 88 | p th | r 903 | ; pn | gı | , TÀS | |
| | | | | | | | | | | | | | 8.81 | n | 13 | • | | a1 | | | | | |
| | | | | | 6 | | | | | | | | | | 6 | | | | | 2 | | | |
| | 6 | TG | a | ī | א ב | | | ~ • | ат | ATC: | TT | | G (CA) | • m | i ca | | T M- | | ເດ | | GA | r 007 | 840 |
| 2 | 61 v | 1 | val | | | in 1 | vs t | hr | usin : | met. | - ph | • th | | | o gl | uph | | E 80 | r al | va. | Lam | p ala | |
| - | | | | | | | • • | | | | | | | | | | | | | | | | |

| CAT TTT TAC GCA AGT GAA GTG TAC GAA TAC TAT AAA AAT GTC CAC CAA CTA GAG AGT CTA 9 | 00 |
|--|----------------------|
| 281 his phe tyr ala ser glu val tyr glu tyr tyr lys asn val his gln leu glu ser leu | |
| asp | |
| | |
| т т | |
| GAT GOT ANA GOT GOA GAA ATT GAT TOG TTT GTC CAT TAT GOC TTG AAT TGC AAT AAT GOC 9 | 60 |
| and der and der der alt is an tee the vel bis ter de lau an des an an als | |
| Sol asp div ivs div div die ine asp set bie van nie tyr div ieu asn tys asn asn an | |
| | |
| | |
| G | |
| TTT TGG GAT GGC CAA GAA ATT CTT TAT GGA GAT GGG GAC AAA AAG AAT TTC AAA CCA TTT 10 | 20 |
| 321 phe trp asp gly gln glu ile leu tyr gly asp gly asp lys lys asn phe lys pro phe | |
| arg | |
| λ | |
| G λ | |
| TCA TGC GCC ANA ACT ATT GTT GGT CAT GAA CTA ACG CAT GCG GTT ATC CAG TAT TCG GCG 10 | во |
| 341 ser cvs ala lvs thr ile val glv his glu leu thr his ala val ile gln tvr ser ala | |
| | |
| | |
| a b c | |
| OCA THE CAN TAC CAN GOD CAN TOA GOT GOD CTA AND GOD THE GOD GAT GTT THE GOT 11 | 40 |
| all level and the character of all level and the set of the set all and the character of th | •• |
| ber die ver die cit die die die eer die eer eer die eer eek ver bes die | |
| | |
| | |
| | ~~ |
| TAT TIT ATT GC CCA AAT CAT TOG TIG ATT GCT GAG GAT GTC TOT GTG GGT GGG TCG CAA 12 | 00 |
| 381 tyr phe ile ala pro asn his trp leu ile giy glu asp val cys val arg gly ser arg | |
| leu | |
| G | |
| | |
| 6 6 | |
| GAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GCG GCT CAT ATG AAA 12 | 60 |
| GAT GGG CGA ATA AGA ACC ATT AAA GAT CCT GAC AAA TAT AAT CAA GCG GCT CAT ATG AAA 12 401 amp gly arg ile arg mer ile lym amp pro amp lym tyr amn gin alm alm him mmt lym | 60 |
| GAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG GCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys | 60 |
| GAT GGG GGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG GCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys A | 60 |
| GAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG GCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm glm ala ala his met lys A T A T A | 60 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT AAT CAA GGG GCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys λ T A T A T A T A T A T A T A T A T A T T A T T A T T A T T A T T A T T A T T A T T A T T T A T T A T T A T T T T T T T T T T | 50 20 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG GCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys A T GAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT CAT AAT AGT 13 421 amp tur glu gen leu pro leu thr glu glu gly asp trp gly gly val his tyr asm ser | 50 20 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT ANT CAA GGG CT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his set lys 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 | 50 20 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG GCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asn gln ala ala his met lys A GAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GOC GAC TOG GOC GGA GTT CATT AAT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asn ser ile phe | 50 20 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT AAT CAA GCG COT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys amp pro amp lys tyr asm gln ala ala his met lys T A A T A GAT TAC GAA TOG CTT CO. CAC AGA GAA GCC GAC TOG GCC GCA GTT CATT AAT AAT AT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asm ser ile phe | 50 20 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT COT GAC AAA TAT AAT CAA GGG GCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T GAT TAC GAA TOG CTT COA CTC ACA GAG GAA GOC GAC TOG GOC GGA GTT CATT AAT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asm ser ile A GTT ATT GGA AAT AAA GCA GCT TATT ATT ATT AAT CTT GGA AAA GAA AAA ACA GAA 13 | 50 20 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT AAT CAA GCG CCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys $\begin{bmatrix} X \\ A \\ GAT TAC GAA TOG CTT COC ATC ACA GAG GAA GOC GAC TOG GOC GGA GTT \boxed{CATT} TAT AAT AGT 13421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly valided byr asm serile \begin{bmatrix} C \\ G \\ GCT ATC COG AAT AAA GOC GOC TAT AAT ACT ATC ACT GGA AAA GAA AAA AOA GAA 13441 duy ile corp sen us ala ala true ser thr ile thr vas un duy var glu lys thr glu$ | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG CTT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asn gln ala ala his met lys $\begin{array}{c} X \\ A \\ CAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT \begin{array}{c} T \\ T \\ A \\ CAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT \begin{array}{c} T \\ T \\ T \\ A \\ C \\ C$ | 50 20 80 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT AAT CAA GCG CCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln alla alla his met lys A GAT TAC GAA TOG CTT COC ATC ACA GAG GAA GOC GAC TOG GOC GGA GTT $OATT AAT AGT 13$ 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly valided byr asm ser ile A GGT ATC COG AAT ANA GOA GOC TAT AAT ACT ATC ACT ANA CTT GGA ANA GAA ANA AOA GAA 13 441 gly ile pro asm lys alla alla tyr asm thr ile thr lys leu gly lys glu lys thr glu | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT COT GAC AAA TAT AAT CAA GGG CT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asn gln ala ala his met lys T T A C T A C T T A A A C T A C C G C G C G C G C G C G G G A A A A A A A A | 50 20 80 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT AAT CAA GCG COT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T GAT TAC GAA TOG CTT COA CATO AGA GAA GGC GAC TGG GGC GGA GTT \overrightarrow{OXT} TA AAT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asm ser 11e GGT ATC COG AAT ANA GCA GGC CAT TAT AAT ACT ATC ACT ANA CTT GGA ANA GAA ANA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu CCC GTT TAT GGC COG TAT ANT ACT ATC ATT TATA ATG CAT TAGA ANA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT COT GAC AAA TAT AAT CAA GGG CT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asn gln ala ala his met lys T T CAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT \overrightarrow{CMT} TA AAT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asn ser ile C GT ATC CGG AAT AAA GCA GCC TAT ATG AAT CACT AAA CTT GGA AAA GAA AAA ACA GAA 13 441 gly ile pro asn lys ala ala tyr asn thr ile thr lys leu gly lys glu lys thr glu C C C C C C C C C C C C CAT AAG TAC TAT TAT ACG AAA AAA ACA GAT CACT 14 CAG CTT TAT TTT CGC GCC TTA AAG TAC TAT TAT ACG AAA AAA ACA GAT CACT 14 CAG CTT TAT TTT CGC GCC TTA AAG TAC TAT TAT ACG AAA AAA ACC GAT CGC 14 | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GCG CCT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys A GAT TAC GAA TOG CTT COA CTT CACA GAG GAA GCC GAC TGG GCC GGA GTT CAT 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val we tra ser ile GGT ATC COG AAT AAA GCA GCC TAT AAT ACT ATC ACT AAA CTT GGA AAA GAA AAA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu CAG CTT TAT TAT GOC GCC TTA AAG TAC TAT TATA GA AAA GCA GCAA GCG GAC GTT CAT TAT ACC GAT GCG GAA GAA GAA GAA GAA GAA GAA GAA GAA | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT COT GAC AAA TAT AAT CAA GGG CT CAT ATG AA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asn gln ala ala his met lys T T CAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT \overrightarrow{CMT} TA AAT AAT TA 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asn ser ile C GT ATC CGG AAT AAA GCA GCC TA AAG AA CAC TAA CTA ACT TAGA AAA AAA ACA GAA 13 441 gly ile pro asn lys ala ala tyr asn thr ile thr lys leu gly lys glu lys thr glu C C C C C C C C C C C C C | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT ANT CAA GCG GCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys A GAT TAC GAA TOG CTT CA CA GAG GAA GCC GAC TOG GGC GGA GTT CAT 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly valles the transfer ile GGT ATC CGG ANT ANA GCA GCC TAT ANT ACT ATC ACT ANA CTT GGA ANA GAA ANA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu CAG CTT TAT TAT COC GCC TTA ANG TAC TAT TATA AGT MAA ANA TOC CAG TTA ACC GAT GGC 14 461 gln leu tyr phe arg ala leu lys tyr tyr leu thr lys lys ser gln phe thr asp ala ala | 50 20 80 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT ANT CAA GCG CCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys $\frac{1}{100}$ asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys $\frac{1}{100}$ asp tyr glue arg ser ile lyr and and and algo ang | 50 20 80 |
| CAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT ANT CAA GGG CT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys A GAT TAC GAA TOG CTT CA CA GAG GAA GGC GAC TOG GGC GGA GTT CAT 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val be pro got ATT AGA GGC GGC TAT ANT ACT ATC ACT AGA CAG GAA AGA ANA AGA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu cAG CTT TAT TAT CGC GCC TAT ANG TAC TAT TAT AGT MAA AGA GAA ANA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu cAG CTT TAT TAT CGC GCC TAT ANG TAC TAT TAT AGT ANA AGA 461 gln leu tyr phe arg ala leu lys tyr tyr leu thr lys lys ser gln phe thr asp ala ala ANA ANA GGG CTT CAA CAA GGA GGG ANA GAT TAT TAT AGT GAA GAT GCT TCT ANA ANA GTT 15 | 50 20 80 40 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT AAT CAA GCG CCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys The GAT TGC GAA TGG CTT CAC AGA GGA AGC GAC TGG GGC GGA GTT CATT ATT AAT AAT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val his tyr asm ser ile GGT ATC CGG AAT ANA GCA GCC TAT AAT ACT ATC ACT AMA CTT GGA AMA GAA AMA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu CAG CTT TAT TAT TGC GGC CCT TA AMA TAT ACT ATC ACT AMA TAT GGA AMA GAA AMA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu CAG CTT TAT TT CGC GCC TTA AMA GTA CTA TAT AGG AMA AMA TCC CAG TTA ACC GAT CGC 14 461 gln leu tyr phe arg ala leu lys tyr tyr leu thr lys lys ser gln phe thr asp ala ala AMA AMA GCG CTT CAA CAA GCA GCG GMA GAT TTA TAT GCT GA GAT GCT TCT AMA AMA GTT 15 61 lys lys ala leu gln gln ala ala lys asp leu tyr gly glu asp ala ser lys lys val | 50 20 80 40 |
| CAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT ANT CAA GGG CCT CAT ATG ANA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T T CAT TAC GAA TOG CTT CA CA GAG GAA GGC GAC TOG GGC GGA GTT TAT ANT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val B GGT ATC GGA ATG AGA GGC GGC TAT ANT ACT ATC ACT AGA CAT GGA ANA GAA ANA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu C C C C C C C C C C C C C | 50 20 80 60 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT AAT CAA GCG CCT CAT ATG AMA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T T T T C C C C C C C C C C C C C | 50 20 80 60 |
| CAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT ANT CAA GGG CT CAT ATG AMA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T T CAT TAC GAA TOG CTT COC CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT TAT AAT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val He by asm ser ile GGT ATC GGA ATA AGA GGC GCC TAT AAT ACT ATC ACT ANA CTT GGA ANA GAA ANA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu C C GGT TAT TAT TTT GGC GCC TTA ANG TAC TAT TAT ACT MAA TOT CAG TAT ACC GAT GGC 14 461 gln leu tyr phe arg ala leu lys tyr tyr leu thr lys lys ser gln phe thr asp ala AAA AMA GCG CT CAA CAA GCA GGG AMA GAT TAT TAT GCT GAA GAT GCT TCT AMA AMA GTT 15 481 lys lys ala leu gln gln ala ala lys asp leu tyr gly glu asp ala ser lys lys val G G G G G G G G G G G G G | 50 20 80 40 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT CCT GAC ANA TAT AAT CAA GCG CCT CAT ATG AMA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T GAT TAC GAA TGG CTT CGA CACA GAG GAA GGC GAC TGG GGC GGA GTT GAT TAT AAT AT 13 421 asp tyr glu ser leu pro leu hr glu glu gly asp trp gly gly valided tyr asm ser ile GGT ATC GGA ATG AA GGA GGC CAT TAT ATT ATT CAT GGA AAA GAA AAA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu C GGG CTT TAT TAT GGC GGC TTA AAG TAC TAT TAT AGG AAA AGT GAC ATAA CGA TAC GAT GAA GAT TAT ATT GGC GAG TTA CGC AT GGG TTA AAGT GAA GAT GGC GAG TTA CGC GAT GAC GAT GAC GAT GAT GAA GAT TAT ATT GGT GAA GAT GCT TCT AAA AAA GTT 15 461 lys lys ala leu gln gln ala ala lys eap leu tyr gly glu asp ala ser lys lys val G GC GAA GCT TGG GAA GCA GGT GGG GTT ACT TAT AAT GGT GAA GAT GCT TCT AAA AAA GTT 15 GC GAA GCT TGG GAA GCA TGG GTT ACT TAT AAT GGT GAA GAT GCT TCT AAA AAA GTT 15 GC GAA GCT TGG GAA GCA TGG GGT TAC TGA 1533 | 50 20 80 40 |
| CAT GGG CGA ATA AGA AGC ATT AAA GAT CCT GAC AAA TAT AAT CAA GGG CT CAT ATG AAA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T T CAT TAC GAA TOG CTT CCA CTC ACA GAG GAA GGC GAC TOG GGC GGA GTT TATT AAT AGT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val He styr asm ser ile GGT ATC GGA ATT AAA GCA GCC TAT AAT ACT ATC ACT AAA CTT GGA AAA GAA AAA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu C C GG TATC CGG AAT AAA GCA GCC TAT AAT ACT ATC ACT AAA CTT GGA AAA GAA AAA ACA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu CAG CTT TAT TTT CGC GCC TTA AAG TAC TAT TAT ACG AAA AAA TCC CAG TTT ACC GAT GGC 14 461 gln leu tyr phe arg ala leu lys tyr tyr leu thr lys lys ser gln phe thr asp ala AAA AAA GCG CT CAA CAA GCA GGG AAA GAT TTA TAT GCT GAA CAT GCT TCT AAA AAA GTT 15 481 lys lys ala leu gln gln ala ala lys asp leu tyr gly glu asp ala ser lys lys val GC GAA GCT TOG GAA CCA GTT GGG GTT AAC TCA 1533 50 ala glu ala trp glu ala val gly val am *** | 50 20 80 90 |
| GAT GGG CGA ATA AGA AGC ATT ANA GAT COT GAC ANA TAT AAT CAA GGG CTA ATG AA 12 401 asp gly arg ile arg ser ile lys asp pro asp lys tyr asm gln ala ala his met lys T GAT TAC GAA TOG CTT COA CTA CAG GGA GGC GAC TOG GGC GGA GTT \overrightarrow{OAT} TAT AAT ATT AAT ATT 13 421 asp tyr glu ser leu pro leu thr glu glu gly asp trp gly gly val as tro GGA ATG AA GGA GGC GAC AGG GAA GGC GAC TOG GGC GAA GAA AA AOA GAA 13 421 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu GGT ATT CGG AAT AAA GOA GGC TAT AAT ACT ATT AAA GTT GGA AAA GAA AAA AOA GAA 13 441 gly ile pro asm lys ala ala tyr asm thr ile thr lys leu gly lys glu lys thr glu GGG GTT TAT TT GGC GGC TTA AAG TAC TAT TAT AGG AAA AAA TOC GAG TTA CCG AT GGG 14 461 gln leu tyr phe arg ala leu lys tyr tyr leu thr lys lys aser gln phe thr asp ala ala AAA AAA GGG CTT CAA CAA GCA GGC GAA GAT TTA TAT GGT GAA GAT GCT TCT AAA AAA GTT 15 481 lys lys ala leu gln gln ala ala lys asp leu tyr gly glu asp ala ser lys lys val G GGA GCT TGG GAA GCA GTG GG GTT ACC TGA 1533 501 ala glu ala trp glu als val gly val asm *** | 50 20 80 90 |

FIG. 2. Complete nucleotide sequence of the *prtA* genes from strains LO28, EGD, and LM8. Only nucleotide changes and amino acid substitutions in the sequences of strains LM8 (serovar 4b) and EGD (serovar 1/2a) compared with the sequence of strain LO28 (serovar 1/2c) are reported. The amino acid sequence deduced from the nucleotide sequence is given under the nucleotide sequence. Nucleotides are numbered on the right side; amino acids are numbered on the left side. The putative signal sequence is underlined. Putative cleavage sites of the signal sequence and of the propeptide (see Results) are indicated by vertical arrows. The insertion point of Tn1545 is indicated (\mathbf{V}). Amino acids involved in zinc atom-binding and catalytic activities are boxed.

presence of putative ORFs downstream from prtA was screened for by use of a program based on codon usage, taking prtA as the reference gene (24). One major ORF with a good probability of being a coding sequence was detected. This ORF, in the same orientation as prtA, started 199 bp downstream from prtA and was not interrupted by a stop codon in the insert of pLis22. In contrast to hlyA and ORF U (15), the ORF identified upstream from hlyA, palindromic sequences corresponding to putative transcriptional termination signals was not found downstream from prtA.

The promoter of prtA was previously mapped and shown to be transcribed independently from a promoter located 150 bp upstream from the putative translation initiation codon of prtA (15). Northern blot analysis determined the size of the prtA transcripts. When total RNAs extracted from bacteria in late-exponential growth were hybridized with a 531-bp Sau96-A-BamHI fragment internal to the prtA gene, one major band was detected that correspond to transcripts of about 5,700 nucleotides (Fig. 4). This is larger than prtA, and the band probably accounts for more than prtA and the ORF located immediately downstream from prtA.

DISCUSSION

In this paper, we have shown (i) that the locus of insertion of a transposon that resulted in a lecithinase-negative phenotype and impairment of virulence mapped in ORF D, which is located downstream from the LLO gene; (ii) that ORF D, designated herein prtA, encodes a protein homologous to bacterial metalloproteases; and (iii) that prtA is the first gene of a 6-kb operon. In addition, we sequenced the gene from three different strains and showed that the gene from an epidemic strain has the most distantly related sequence.

prtA was suspected to play a role in virulence on the basis of two pieces of evidence (i) Southern blotting experiments performed with all species of the genus *Listeria* had shown that this region was specific to *L. monocytogenes* even at a low stringency of hybridization (10). (ii) Promoters of *prtA* and *hlyA* share common structural features, suggesting that those two genes might be similarly regulated (15). Transposon Tn1545, in the less virulent lecithinase-negative mutant, interrupted *prtA*. This indicates that the *prtA* gene region is involved in virulence. Since *prtA* is the first gene of a 6-kb

| A <i>L</i> . <i>B</i> . <i>B</i> . <i>B</i> . <i>B</i> . | monocytogenes subtilis amyloliquefaciens stearothermophilus thermoproteolyticus cereus | adthkdfqalGKGANRLLqrplhvmk aaatGSGTTLKGatvplnis aattGTGTTLKGktvslnis vagastvGVGRGVLGdqkyintt itgtstvGVGRGVLGdqknintt vtgtnkvGTGKGVLGdtkslntt | indlFYLvDRTHKGLIRTFDlkhntdt yeggkYVLrdlSKPTGTQIITYDlqnrqs; sesgkYULrdlSKPTGTQIITYDlqnrey; yssyygYY1qDNTRGSGIFTYDarkyrt1 ystyYYLqDNTRGATIFTYDakrrs1 lsgssYYLqDNTRGATIFTYDakrrs1 | sfgkvv: lpgtlv: lpgtlv: lpgslwto lpgslwao lpgtlwao | snktnmft ssttktft ssttnqft lgdnqfta ladnqffa ladnvfna | dpefssAVD sssqraAVD tssqraAVD sydaa.AVD sydap.AVD aydaSINDA | AHFYASEV) AHYNIGKVJ AHYNIGKVJ AHYYAGVVJ AHYYAGVTJ GAPa . AVD/ | YEYyknvhqleSLDGKGGEIDS YDYfygnfkrnSYDNKGSKIVS YDYfygkfnnSYDNKGGKIVS YDYyknvhgrlSYDGSNAIRS YDYyknvhnrlSYDGNNAAIRS AHYYAGKTYDYykatfnrnLKS |
|--|--|---|---|--|--|--|--|---|
| | | FVHYGLNCNNAFWDGQEILYGDGDKI SVHYGSQYNNAAWTGDQMIYGDGDG SVHYGSRYNNAAWTGDQMIYGDGDG TVHYGRGYNNAFWNGSQMVYGDGDG SVHYSQGYNNAFWNGSQMVYGDGDG TVHYGSNYNNAFWNGSQMVYGDGDG | NFKPFScaktIVg.HELTHAVIQYSAGLEYE FFSPLSgal.DVtaHEMTHGVTQETANLIYE FFSPLSgam.DVtaHEMTHGVTQETANLIYE TFFSPLSggi.DVvgHELTHAVTDYTAGLYQ TFFIPLSggi.DVvgHELTHAVTDYTAGLIYQ TFFSLSggi.DVigHELTHAVTENSSNLIYQ A> | GQSGALNI NQPGALNI NQPGALNI NESGAINI NESGAINI NESGALNI | ESFAD ESFSD ESFSD EAMSDifg EAISDifg BAISDifg B> | VFGYFIA VFGYFND VFGYFND tlvefyanr tlvefyank tlvefydnr | PNHWLIGEI TEDWDIGEI TEDWDIGEI NPDWEIGEI NPDWEIGEI | DVCVRGsrdgrIRSIKDPDK DitVSQPALRSLSNPTK DitVSQPALRSLSNPTK DIYTPGvagdalRSMSDPAK DVYTPGisgdslRSMSDPAK DIYTPGkagdalRSMSDPTK |
| | | YnqaahmkDYESLPLTEEGDWGGVH YnqpdnyaNYRNLPNTDEGDYGGVH Ygqdaf kNYKNLPNTDAGDYGGVH YgdpDHYSKRYTGTQDNGGVH YgdpDHYSKRYTGTQDNGGVH YgdpDHYSKRYTGSSDNGGVH | NSGIPNKAAYnTITKLGKEKT NSGIPNKAAYnTITKLGVEKT NSGIPNKAAYnTITKLGVKAKA NSGIINKAAYllsqggvhygvSVNGIGRDKM INSGIINKAAYllsqggthygvSVVGIGRDKL/ NSGIINKQAYllanggthygvTVTGIGRDKL/ C> | EQLYFRA QQIYYRA EQIYYRA GKIFYRA GKIFYRA GAIYYRA | LKYYLTKK LTTYLTPS LTVYLTPS LVYYLTPT LTQYLTPT NTQYFTQS | SQFTDAKKA STFKDAKAA STFKDAKAA SNFSQLRAA SNFSQLRAA STFSQARAG | LQQAAKDL LIQSARDL LIQSARDL CVQAAADL AVQSATDL SAVQAAADL | YGEDASKKVAEAWEAVGVn YGSTDAAKVEAAWNAVGL. YGSDDAASVEAAWNAVGL. YGstSQEVNSVKQAFNAVGVy YGstSQEVNSVKQAFDAVGVk YGanSAEVAAVKQSFSAVGVn |
| | B | | | | | | | |
| ľ | NH ₂ | | | - | - | | · | COOH |
| | | | | | | | | |
| | Ţ | Ţ | | A | B | Ţ | С | |
| | Signal sequence (~ 25 amino acids) | Propeptide (~ 200 amino acids) | | | Mai (= 32 | ture prote: 0 amino aci | ase ids) | |
| | С | | | | | | | |
| | monocytogenes 8. steanothermophilus 8. thermopoteolyticus 8. subtilis 8. subtilis 8. cereus | A Residu T I V G H E L T H A V 346 - D V V G H E L T H A V 370 - D V V A H E L T H A V 138 - D V T A H E M T H G V 360 - D V T A H E M T H G V 360 - | es B→→→ 155 S G A I X E S F A D 368 - 180 S G A I X E A M S D 393 - 148 S G A I X E A I S D 161 - 170 P G A D X E S F S D 383 - 170 P G A L X E S F S D 383 - | ines 377 1 402 1 170 1 392 1 392 1 | EECC GTQC GTOC DAGC DEGC | W 3 3 7 11 3 1 11 3 1 11 3 1 11 3 1 11 11 11 11 11 11 11 11 11 11 11 11 1 | | Residues Residues Residues 1 P X K A A Y 428 - 448 1 X K A A Y 454 - 474 1 X K A A Y 222 - 242 2 P X K A A Y 440 - 460 2 P X K A A Y 440 - 460 |

FIG. 3. (A) Comparison of the putative mature form of *Listeria* protease and mature forms of *Bacillus* metalloproteases. Blocks of conserved regions are indicated with uppercase letters as described previously (21). Amino acids involved in zinc atom-binding and catalytic activities are indicated (*). (B) Schematic representation of bacterial metalloprotease structure. A, B, and C correspond to highly conserved regions. (C) Amino acid sequence comparison of three highly conserved domains, A, B, and C, among the bacterial metalloproteases. Identical amino acids are indicated by black boxes.

operon, it is highly probable that insertion of Tn1545 had at least a partial polar effect and that a lecithinase gene or a gene involved in the lecithinase activity expression is present at some distance downstream from *prtA*. This hypothesis is corroborated by the recent finding of other lecithinase-negative mutations mapping downstream from *prtA* in a gene presenting no homology with other phospholipase C genes (our unpublished results). Whether the *prtA* gene product itself or adjacent genes play a role in lecithinase expression and in virulence remains to be clarified.

In this study, we cloned and sequenced the prtA gene from

three L. monocytogenes strains, EGD (serovar 1/2a), LO28 (serovar 1/2c), and LM8 (serovar 4b). Those strains had been classified in two groups on the basis of a restriction fragment length polymorphism in the *prtA* gene region (10; unpublished results). By multilocus enzyme electrophoresis analysis, Piffaretti et al. (20) estimated the genetic distances between 175 L. monocytogenes strains and classified them in two distinct groups. In this classification, LO28 and EGD were in the same group; LM8 was in a different group that contained all of the epidemic strains tested (20). Comparison of the three *prtA* sequences is in agreement with a shorter



FIG. 4. (A) Agarose gel electrophoresis of 25 μ g of total RNA from *L. monocytogenes* LO28. Total RNA was extracted as described in Materials and Methods from cultures in late-exponential growth (A_{600} , 1.7). Samples 1, 2, 3, 4, and 5, respectively, correspond to cultures at 10, 11, 12, 13, and 14 h. rRNAs are indicated by arrows and used as size standards (in nucleotides [nts]). (B) Autoradiogram of Northern blot prepared from the gel shown in panel A and hybridized with a ³²P-labeled *prtA* probe. The probe is the 531-bp *Sau*96-A–*Bam*HI fragment internal to *prtA* and purified from pLis3. The approximate size of the major *prtA* transcript is indicated.

genetic distance between EGD and LO28 than between LO28 or EGD and LM8.

On the basis of amino acid sequence comparison, prtA was found to encode a protease closely related to metalloproteases produced by Bacillus species and, to a lesser extent, to those produced by gram-negative bacteria such as the elastase of P. aeruginosa, MSP of L. pneumophila, and the protease of Serratia sp. The role of elastase in P. aeruginosa virulence has not been clearly established. For L. pneumophila, it has recently been shown that MSP plays no role in intracellular growth or cell killing (5, 25). Nevertheless, for these two human pathogens, proteolytic activity of the protease is easily detected. This type of activity has not been described in Listeria species. In this work, the gene prtA was shown to be transcribed, but the proteolytic activity, despite several attempts, was not detected. We were unable so far to find the substrate that would allow detection of the proteolytic activity, to detect a crossreacting material with antibodies raised against thermolysin or MSP, or to identify conditions in which proteolytic activity could be expressed. In bacilli metalloproteases are synthesized as proenzymes. Cleavage of the amino-terminal part of the molecule releases the active form of the protease. In Listeria species, this proteolytic activation might occur in environmental conditions that remain to be discovered.

ACKNOWLEDGMENTS

We thank J. Davies for his constant interest in this work, K. Reich and H. Shuman for help in detecting the activity, and M. Horwitz for the gift of anti-MSP monoclonal antibodies.

This work was supported by the World Health Organization as part of its program for "Transdisease Vaccinology," the EEC (action de stimulation, contrat de jumelage no. STJ2-0319-C), the Institut National de la Santé et de la Recherche Scientifique CRE no. 891003, the Centre National de la Recherche Scientifique U.R.A. 209, and the Pasteur Institute.

REFERENCES

- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1990. Current protocols in molecular biology. Greene Publishing and John Wiley & Sons, Inc., New York.
- Bever, R. A., and B. H. Iglewski. 1988. Molecular characterization and nucleotide sequence of the *Pseudomonas aeruginosa* elastase structural gene. J. Bacteriol. 170:4309–4314.
- Bielecki, J., P. Youngman, P. Connelly, and D. A. Portnoy. 1990. Bacillus subtilis expressing a haemolysin gene from *Listeria* monocytogenes can grow in mammalian cells. Nature (London) 345:175-176.
- Black, W. J., F. D. Quinn, and L. S. Tompkins. 1990. Legionella pneumophila zinc metalloprotease is structurally and functionally homologous to Pseudomonas aeruginosa elastase. J. Bacteriol. 172:2608–2613.
- Blander, S. J., L. Szeto, H. A. Shuman, and M. A. Horwitz. 1990. An immunoprotective molecule, the major secretory protein of Legionella pneumophila, is not a virulence factor in a guinea pig model of Legionnaires' disease. J. Clin. Invest. 86:817-824.
- Casadaban, M. J., and S. N. Cohen. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138:179–207.
- 7. Colman, P. M., J. N. Jansonius, and B. W. Matthews. 1972. The structure of thermolysin: an electron density map at 2.3 A resolution. J. Mol. Biol. **70**:701–724.
- Cossart, P., M. F. Vicente, J. Mengaud, F. Baquero, J. C. Perez-Diaz, and P. Berche. 1989. Listeriolysin O is essential for the virulence of *Listeria monocytogenes*: direct evidence obtained by gene complementation. Infect. Immun. 57:3629–3636.
- 9. Gaillard, J. L., P. Berche, and P. Sansonetti. 1986. Transposon mutagenesis as a tool to study the role of hemolysin in the virulence of *Listeria monocytogenes*. Infect. Immun. 52:50–55.
- 10. Gormley, E., J. Mengaud, and P. Cossart. 1989. Sequences homologous to the listeriolysin O gene region of *Listeria monocytogenes* are present in virulent and avirulent haemolytic species of the genus *Listeria*. Res. Microbiol. 140:631–643.
- 11. Holmes, M. A., and B. W. Matthews. 1982. Structure of thermolysin refined at 1.6 angstrom resolution. J. Mol. Biol. 160: 623-639.
- 12. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science 227:1435–1440.
- Mackaness, G. B. 1962. Cellular resistance to infection. J. Exp. Med. 116:381-406.
- 14. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 15. Mengaud, J., M. F. Vicente, and P. Cossart. 1989. Transcriptional mapping and nucleotide sequence of the *Listeria monocytogenes hlyA* region reveal structural features that may be involved in regulation. Infect. Immun. 57:3695–3701.
- Michel, E., K. A. Reich, R. Favier, P. Berche, and P. Cossart. 1990. Attenuated mutants of the intracellular bacterium *Listeria* monocytogenes obtained by single amino-acid substitutions in listeriolysin O. Mol. Microbiol. 4:2167-2178.
- 17. Miller, R. A., S. G. Reed, and M. Parsons. 1990. Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence. Mol. Biochem. Parasitol. 39:267–274.
- Murray, E. G. D., R. E. Webb, and M. B. R. Swann. 1926. A disease of rabbits characterized by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillus *Bacterium* monocytogenes (n. sp.). J. Pathol. Bacteriol. 29:407-439.
- Nakahama, K., K. Yoshimura, R. Marumoto, M. Kikuchi, I. S. Lee, T. Hase, and H. Matsubara. 1986. Cloning and sequencing of *Serratia* protease gene. Nucleic Acids Res. 14:5843–5855.
- Piffaretti, J. C., H. Kressebuch, M. Aeschbacher, J. Bille, E. Bannermann, J. M. Musser, R. K. Selander, and J. Rocourt. 1989. Genetic characterization of clones of the bacterium *Listeria monocytogenes* causing epidemic disease. Proc. Natl. Acad. Sci. USA 86:3818-3822.
- 20a. Ravenau, J., and P. Berche. Personal communication.
- 21. Saurin, W., and P. Marliere. 1986. Simultaneous alignment of

several protein sequences. C. R. Acad. Sci. Paris 303:541-546. 22. Seeliger, H. P. R. 1961. Listeriosis. Karger, Basel.

- 23. Sidler, W., E. Niederer, F. Suter, and H. Zuber. 1986. The
- 23. Suff, W., E. Mederer, F. Suff, and H. Zuber. 1960. The primary structure of *Bacillus cereus* neutral proteinase and comparison with thermolysin and *Bacillus subtilis* neutral proteinase. Biol. Chem. Hoppe-Seyler **367**:643–657.
- 24. Staden, R., and A. D. McLachlan. 1979. Codon preference and its use in identifying protein coding regions in long DNA sequences, p. 141–153. *In* D. Soll and R. J. Roberts (ed.), The applications of computers to research on nucleic acids. IRL Press, Oxford.
- Szeto, L., and H. A. Shuman. 1990. The Legionella pneumophila major secretory protein, a protease, is not required for intracellular growth or cell killing. Infect. Immun. 58:2585–2592.
- Takagi, M., T. Imanaka, and S. Aiba. 1985. Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus*. J. Bacteriol. 163:824-831.
- 27. Titani, K., M. A. Hermodson, L. H. Ericsson, K. A. Walsh, and H. Neurath. 1972. Amino-acid sequence of thermolysin. Nature

(London) New Biol. 238:35-37.

- Vasantha, N., L. D. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula. 1984. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. 159:811-819.
- Vicente, M. F., F. Baquero, and J. C. Perez-Diaz. 1985. Cloning and expression of the *Listeria monocytogenes* haemolysin in *E. coli*. FEMS Microbiol. Lett. 30:77-79.
- Watson, M. E. E. 1984. Compilation of published signal sequences. Nucleic Acids Res. 12:5145-5163.
- 31. Yang, M. Y., E. Ferrari, and D. Henner. 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. J. Bacteriol. 160:15-21.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.