JOURNAL OF BACTERIOLOGY, Dec. 1996, p. 7260–7264 Vol. 178, No. 24
0021-9193/96/$04.00+0

N-Acetyl-Heparosan Lyase of *Escherichia coli* K5: Gene Cloning and Expression

RICHARD LEGOUX,[1]* PHILIPPE LELONG,[1] CHRISTIANE JOURDE,[1] CLAUDE FEUILLERAT,[1] JOËL CAPDEVIELLE,[2] VIVIANE SURE,[3] EDGARDO FERRAN,[4] MOURAD KAGHAD,[5] BRUNO DELPECH,[5] DAVID SHIRE,[3] PASCUAL FERRARA,[2] GÉRARD LOISON,[1] AND MARC SALOMÉ[1]

Department of Microbiology,[1] Department of Protein Biochemistry,[2] Department of Chemistry,[3] Computer Department,[4] and Department of Gene Molecular Biology,[5] Sanofi-Recherche, 31676-Labège Cedex, France

Received 3 April 1996/Accepted 7 October 1996

The structure of the capsular polysaccharide of *Escherichia coli* K5 is identical to that of *N*-acetyl-heparosan, a nonsulfated precursor of heparin, which makes this *E. coli* antigen an attractive starting point for the chemical synthesis of analogs of low-molecular-weight heparin. This polysaccharide is synthesized as a high-molecular-weight molecule that can be depolymerized by an enzyme displaying endo-β-eliminase activity. The eliminase-encoding gene, designated *elmA*, has been cloned from *E. coli* K5 by expression in *E. coli* K-12. The K-12 genome is devoid of the *elmA* sequence. The *elmA* gene product is 820 amino acids long. Active recombinant eliminase is produced by K-12 cells in both cell-bound and secreted forms. Deletion analyses have shown that the C terminus and the N terminus are required for activity and secretion, respectively.

The K5-specific capsular polysaccharide of *Escherichia coli* (K5 antigen) is composed of regular repeats of 4-β-glucuronyl-(1-4)-α-*N*-acetylglucosaminyl-1 (8). This structure is identical to that of *N*-acetyl-heparosan, a nonsulfated precursor of heparin. This similarity makes the K5 polysaccharide an attractive molecule to be used as the starting point for the chemical synthesis of pharmacologically active analogs of low-molecular-weight (LMW) heparin. *N*-Acetyl-heparosan is released into the medium by *E. coli* K5 as long polymers of LMW, ranging in size from 100 to 200 kDa (unpublished data). Under certain culture conditions, e.g., growth in glucose-containing complex medium (our results to be published elsewhere), *E. coli* K5 produces an enzyme which degrades extracellular *N*-acetyl-heparosan into LMW species by an endo-β-elimination reaction. Like the *N*-acetyl-heparosan lyase of the K5-specific phage (2), this bacterial lyase can be used to fragment high-molecular-weight (HMW) K5 antigen in vitro. However, the bacterial lyase yields oligosaccharides of about 5 kDa as major end products, compared to up to 1,000-kDa fragments obtained with the phage enzyme (2). Since the desirable size of the starting material for synthesis of the heparin analog precisely corresponds to a 5-kDa oligosaccharide, the bacterial enzyme is of special interest for the preparation of LMW *N*-acetyl-heparosan. However, it is produced by *E. coli* K5 only at a low level. We have therefore isolated the corresponding gene by screening *E. coli* K-12 colonies capable of producing *N*-acetyl-heparosan eliminase after transformation with an *E. coli* K5 genomic DNA bank, and we have thus constructed K-12 strains that produced extracellular eliminase at high levels.

MATERIALS AND METHODS

Bacterial strains and plasmids. The *E. coli* strains were K5 SEBR 3282 (O10:K5:H4) and K-12 RR1 (Boehringer GmbH, Mannheim, Germany). The *N*-acetyl-heparosan lyase-encoding gene was cloned from a library consisting of fragments obtained from partial *Sau*3A digestion of SEBR 3282 genomic DNA cloned into the *Bam*HI site of pUC18. The ligation mix was used to transform RR1 cells to ampicillin resistance. About 7,000 colonies were obtained. The mean size of the inserts was estimated at 3 to 4 kb, and the percentage of insertion was estimated at 100%. RR1 transformed colonies were divided into 10 pools of 700. Each pool was tested individually by the K5 depolymerization test. One pool was found to induce K5 polysaccharide fragmentation; 1,200 colonies were then isolated from this pool, and cells from each colony were separately resuspended in 3 ml of A buffer (100 mM bis-Tris-propane, 150 mM NaCl, HCl, pH 6.6) and used for the K5 depolymerization test. The *elmA*-bearing pUC18 plasmid isolated with this protocol was designated pEMR838. All the expression plasmids used in this study derived from pEMR466, a urate oxidase expression vector for *E. coli* which has already been described elsewhere (4). *N*-Acetyl-heparosan expression plasmids were obtained by exchanging the urate oxidase-encoding sequence with that encoding *N*-acetyl-heparosan lyase. Plasmid pEMR466 was fully digested by *Nde*I and *Hin*dIII to liberate the urate oxidase-encoding sequence, and the vector large fragment was religated with an *Nde*I-*Hin*dIII fragment corresponding to a PCR-amplified region of the eliminase-encoding sequence. Three different amplified fragments were cloned. The sequences of the primers used in PCR experiments (with the initiation codon shown in bold characters and the *Nde*I site underlined) were as follows: 5′-GATCCAT**ATG**ACGGTCTCAACCGAAGTTG-3′ (primer a), 5′-GATCCA T**ATG**CTGATCCAGCGATGTTTTGGGTGG-3′ (primer b), and 5′-GATC CAT**ATG**CGTGATGGTGTCAGCATTAAGGATTTTGG-3′ (primer c). The sequence of the antisense primer d (with the *Hin*dIII site underlined) was as follows: 5′-GATCAAGCTTATCAATTCCCTGTTAATTGCAAAAC-3′. PCR conditions were as follows: 200 ng of pEMR838 DNA, 100 ng of primer a, b, or c, 100 ng of primer d, and 200 μmol of each of the four deoxynucleoside triphosphates in 50 μl of a buffer composed of 67 mM Tris-HCl (pH 8.8), 16.6 mM $(NH_4)_2SO_4$, 1 mM β-mercaptoethanol, 6.7 mM EDTA, 0.15% Triton X-100, and 200 mg of gelatin. Primers a, b, and c were used for the amplification of the inserts of plasmids pEMR868, pEMR869, and pEMR886, respectively. The open reading frame in each insert starts with a different initiator codon; these codons are numbered I to III. Each amplified fragment was obtained after an 18-cycle reaction, with each cycle corresponding to the following series: 92°C for 1 min, 55°C for 1 min, and 72°C for 3 min.

elmA expression vectors pEMR868, pEMR869, and pEMR886 are therefore essentially similar to the urate oxidase expression vector pEMR466. They contain a pBR327-derived sequence including the origin of replication, the ampicillin resistance gene (*bla*), and the *E. coli lacI* gene. They also contain an expression cassette that includes an IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible artificial promoter, an optimized ribosome binding site, and an fd-derived transcription terminator.

Growth conditions. Cells were grown in standard Luria-Bertani (LB) medium supplemented with ampicillin (100 mg/liter) and 1 mM IPTG, as described below.

DNA protocols. Isolation of plasmids, plasmid transformation, and standard DNA protocols were performed as described by Sambrook et al. (6). DNA sequencing was done on double-stranded DNA with Sequenase (7).

K5 depolymerization test. Fifty microliters of a 20-g/liter solution of HMW K5

* Corresponding author. Mailing address: Sanofi-Recherche, Labege Innopole BP137, 31676 Labège Cedex, France. Phone: (33) 5 61 00 40 64. Fax: (33) 5 61 00 40 01. Electronic mail address: richard.legoux@tls1.elfsanofi.fr.

antigen was added to 500 μl of each cell suspension. Each reaction mix was incubated at 37°C for 3 to 4 days and then centrifuged at 10,000 × *g* for 3 min. One hundred microliters of the supernatant was precipitated by the addition of 4 volumes of ethanol and then centrifuged again. The pellet was resuspended in 300 μl of 25 mM piperazine-HCl, pH 3.5, and loaded onto a Q-Sepharose microcolumn. The microcolumn was then washed with 2 ml of piperazine buffer, followed by 2 ml of H_2O. The oligosaccharides were eluted with 200 μl of 0.5 M NaCl and analyzed directly by size exclusion chromatography with an SK 2000 SW column (300 by 7.5 mm) containing silica beads (diameter, 1 μm; porosity, 125 Å). Samples were eluted by 0.5 M $NaSO_4$ (1 ml/min) and detected by UV absorption at 205 nm.

***elmA* gene product analyses.** Samples were analyzed by 0.1% sodium dodecyl sulfate (SDS)–10% polyacrylamide gel electrophoresis (PAGE) (3) and stained with Coomassie brilliant blue R-250. A saturated overnight culture of each RR1 transformant was diluted 1:100 in LB medium supplemented with ampicillin and allowed to grow until the culture reached an A_{600} of 0.5 to 0.7. IPTG (final concentration, 1 mM) was added to the cells, and growth continued for another 3 h at 37°C. The bacteria were collected by centrifugation, and the bacterial pellets were either resuspended in loading buffer and incubated in boiling water for 3 min prior to PAGE analysis or incubated in water for sonication (4). In the latter case, the bacterial lysate was centrifuged at 15,000 × *g* for 15 min. Both the supernatant (soluble fraction) and the pellet (insoluble fraction) were incubated at 100°C for 3 min and analyzed. Coomassie brilliant blue R-250-stained bands were cut from the gel. Proteins were submitted to tryptic hydrolysis (5). Peptides were analyzed by mass spectrometry and purified by high-pressure liquid chromatography (5). Some of them were sequenced with an Applied Biosystems 470 A apparatus.

Nucleotide sequence accession number. The *elmA* sequence accession number is X96495 in the EMBL data base.

RESULTS

Isolation of the gene encoding the *N*-acetyl-heparosan lyase (*elmA*) of *E. coli* K5. Seven thousand RR1 colonies transformed by a K5 genomic DNA library were screened for their ability to fragment HMW *N*-acetyl-heparosan. One positive colony was detected. It harbored a plasmid, named pEMR838, which contained a 3-kb insert. The sequence of this insert was entirely determined. A search for open reading frames revealed only one that exceeded 100 codons, which supposedly encoded *N*-acetyl-heparosan lyase (*elmA* gene). The putative translation product, as deduced from the nucleotide sequence (positions 309 to 2768), is 820 amino acids long. The amino acid sequence of the putative *elmA* gene product does not have any significant similarity to any protein in the GenBank database, except for a motif located at C-terminal positions 788 to 812, YGIRYEEALILESAMMRRRVKKLEE, which is similar to a C-terminal sequence of endo-*N*-acetylneuraminidase E of bacteriophage PK1E (YGVRYSEVLILEAAYTRHRLD KLEE) (1). No peptide signal-like sequence was detected at the N terminus of the molecule. The *elmA* open reading frame is preceded by a canonical Shine-Dalgarno sequence. Promoter elements might include a putative −35 box (5′-TTG CAA) at positions 183 to 188, followed by a possible −10 site (5′-TCTATT) at positions 207 to 212.

Active *N*-acetyl-heparosan eliminase is released into the culture medium by *elmA*-expressing *E. coli* K-12 cells. The transformation of RR1 cells by pEMR838 DNA yielded HMW *N*-acetyl-heparosan-fragmenting colonies, suggesting that the large open reading frame encoded *N*-acetyl-heparosan eliminase (Fig. 1). This was ascertained by testing the eliminase production of *E. coli* RR1 cells transformed by various expression plasmids, each containing a different fragment of the pEMR838 insert. In addition to the triplet that opens the *elmA* reading frame, there were two in-frame ATG codons, each preceded by a possible Shine-Dalgarno motif; one was located at nucleotide position +333 and the other was at position +580 of the long open reading frame. Thus, three different constructions, corresponding to three possible starts, were undertaken. These fragments were each placed under the control of an IPTG-induced artificial promoter in pEMR466, an expression vector for *E. coli*, to yield pEMR868, pEMR886, and pEMR869, respectively (Fig. 2). RR1 cells transformed by any of these plasmid DNAs produced active eliminase after the addition of IPTG to the medium. Interestingly, eliminase activity was found both in the culture medium and in the biomass of RR1 cells transformed by either pEMR868 or pEMR886, whereas the activity could be detected only in the biomass of cells expressing the shortest version (Fig. 3). The extracellular eliminase produced by *E. coli* K5 yields short polymers corresponding to 9 to 10 disaccharide residues as major end products of the *N*-acetyl-heparosan fragmentation, whereas the cell-bound form tends to depolymerize the substrate to a smaller size. As 9 to 10 disaccharide units precisely correspond to the size of the material used as the starting point for the synthesis of the LMW heparin substitute, the extracellular

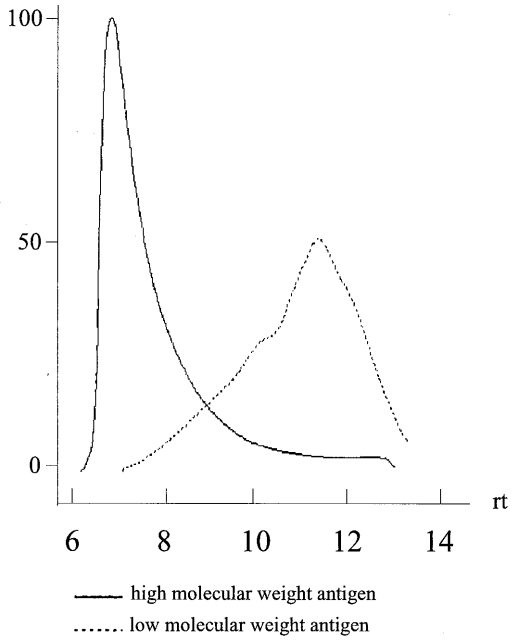


FIG. 1. K5 polysaccharide fragmentation test used to detect RR1 pEMR838 cells. This test was performed as described in Materials and Methods. The oligosaccharides yielded by the fragmentation reaction were analyzed by size exclusion chromatography with an SK 2000 SW column containing silica beads. Samples were eluted by 0.5 M $NaSO_4$ and detected by UV absorption at 205 nm.

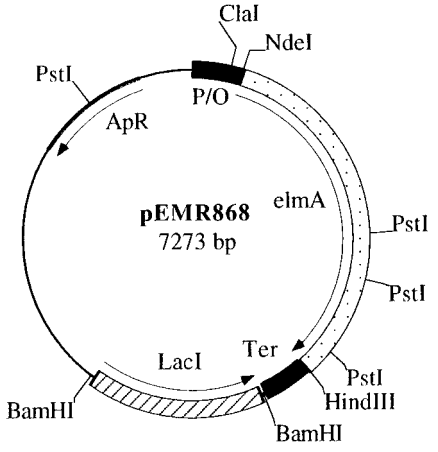


FIG. 2. Construction and structure of plasmid pEMR868. pEMR886 and pEMR869 were constructed in a similar fashion.

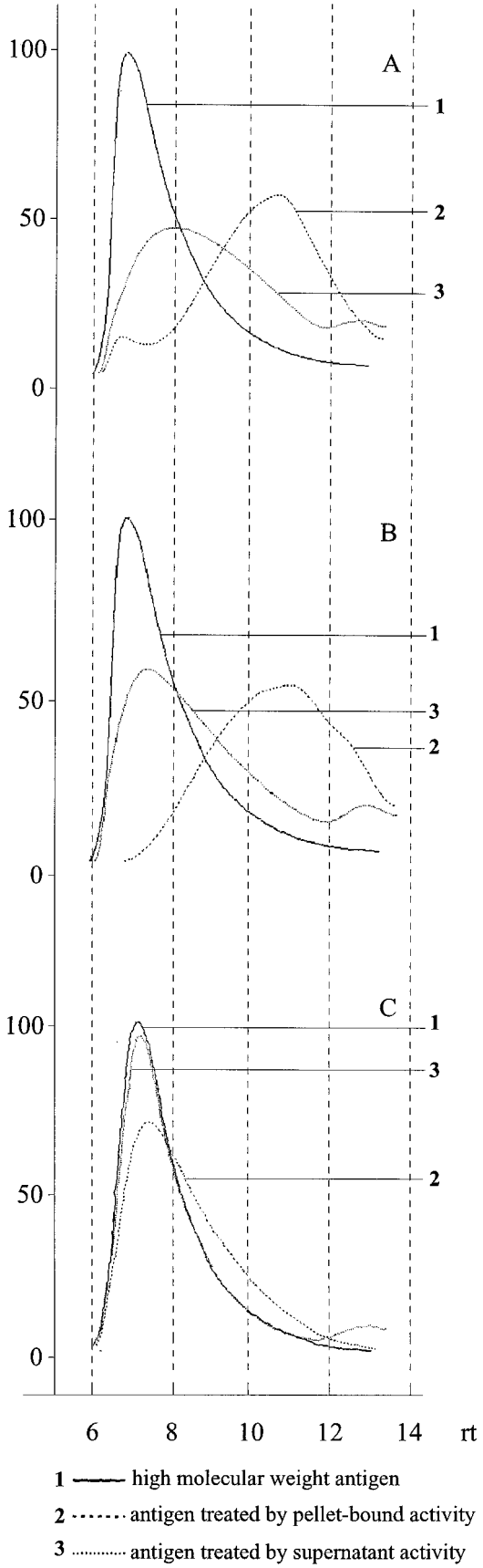


FIG. 3. K5 polysaccharide fragmentation patterns obtained with RR1 cells transformed by pEMR868 (A), pEMR886 (B), and pEMR869 (C). Cells were grown in L broth containing 100 mg of ampicillin per liter to a final optical density at 600 nm of 0.5. Then they were incubated in the presence of 1 mM IPTG for 3 h and finally harvested. Both the pellet and the supernatant, corresponding to about 10^9 cells, were used in the fragmentation test; the method described in the legend to Fig. 1 was used.

enzyme is of particular biotechnological interest. Consequently, pEMR868 was kept for further studies.

Analysis of the enzyme produced by pEMR868-transformed RR1 cells. Crude cell extracts were prepared from RR1 cells transformed by either pEMR868 or pEMR373, a negative control vector, and analyzed in 10% PAGE–0.1% SDS experiments (Fig. 4). Three polypeptides of apparent molecular masses of 89, 70, and 19 kDa were detected in the insoluble fraction from pEMR868-transformed cells and were absent in the negative control. Mass spectrometry data on high-pressure liquid chromatography-purified polypeptides obtained by tryptic hydrolysis of eliminase-derived materials confirmed that the 89- and 70-kDa species were a full-sized and a C terminus-truncated version of eliminase, respectively (Fig. 5). Both forms appeared to be devoid of the N-terminal methionine, as further confirmed by N terminus sequencing on the intact protein (data not shown). The 19-kDa species was in fact identified as hsp16, a 16-kDa heat shock protein whose accumulation was likely provoked by overproduction of the recombinant eliminase. By in vitro-directed mutagenesis we constructed a deleted version of *elmA* lacking the final 15 codons and found that the corresponding gene product was inactive. Thus, the 70-kDa product was probably inactive. It is worth noting that the motif shared with endoneuraminidase has been partially deleted in this inactive construct. On the other hand, Gerardy-Schahn et al. (1) showed that deleting the 38 last C-proximal amino acids from endo-*N*-acetyl-neuraminidase E,

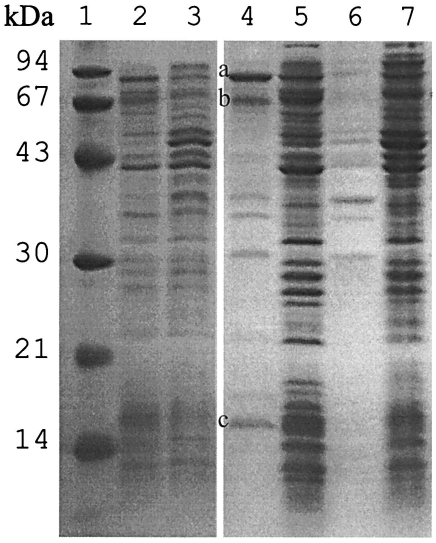
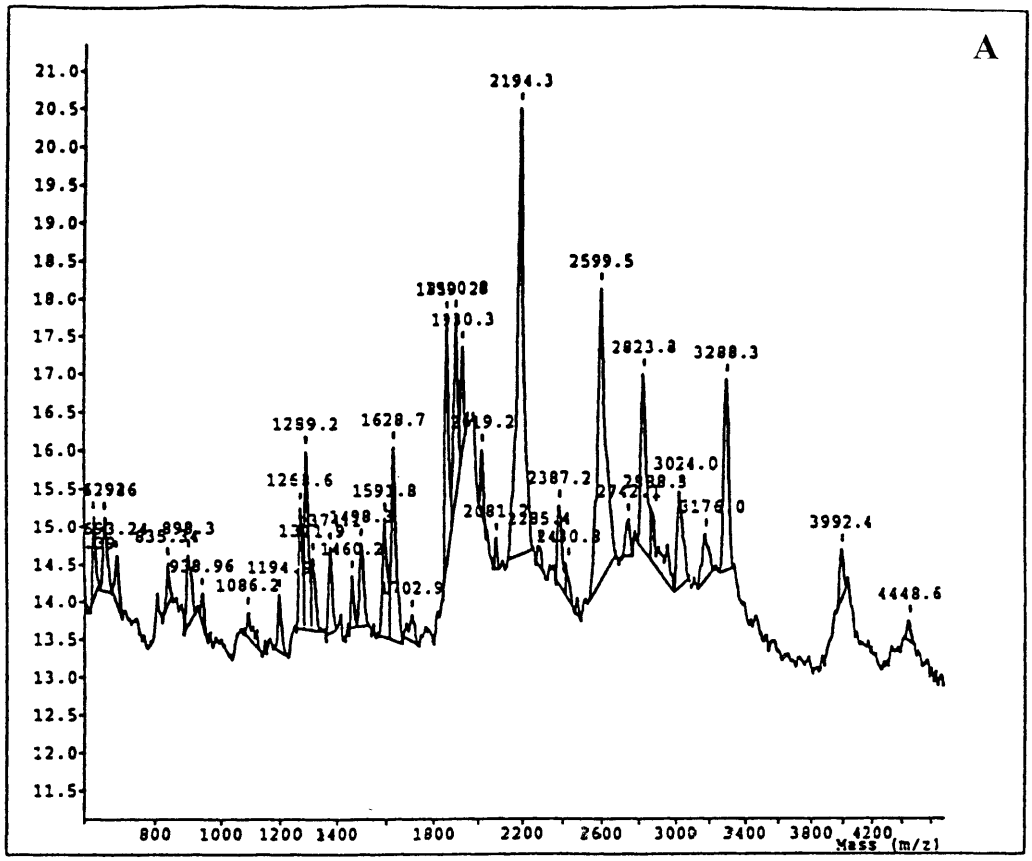


FIG. 4. Detection and analysis by SDS-PAGE of *elmA* proteins produced by RR1 cells transformed by pEMR868. Whole-cell lysate (lanes 2 and 3), insoluble fraction (lanes 4 and 6), and soluble fraction (lanes 5 and 7) from RR1 cells were transformed by pEMR868 (lanes 2, 4, and 5) or the negative control vector pEMR373 (lanes 3, 6, and 7). Lane 1, molecular mass standard (Pharmacia, Uppsala, Sweden). Bands corresponding to the material analyzed by mass spectrometry and sequenced are indicated by "a," "b," and "c" in lane 4.



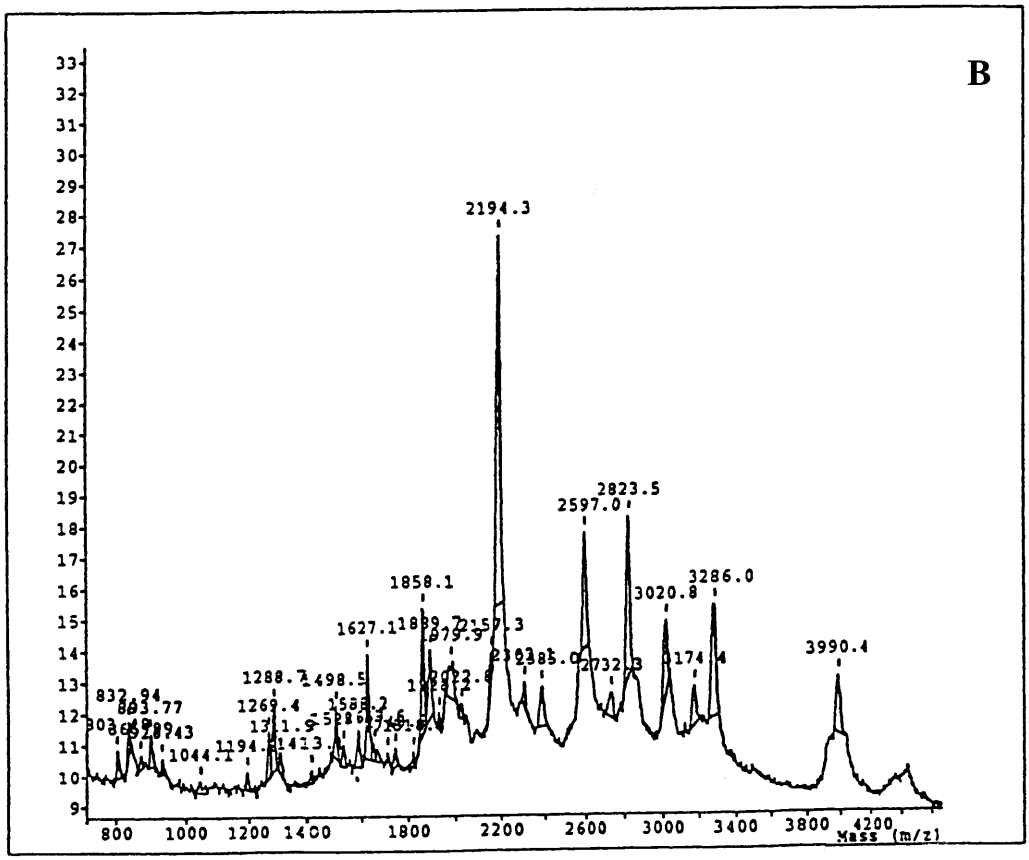


FIG. 5. Mass spectrometry analysis of tryptic digest from the 89-kDa (A) and 70-kDa (B) proteins.

including the whole motif, resulted in enzyme inactivation. It is therefore possible that this motif is required for enzyme activity in both enzymes.

The activity assays performed on the extracellular recombinant enzyme confirmed the size of approximately 9 to 10 disaccharide units for the major end product of the depolymerization reaction (not shown). The overproduction factor obtained by using the recombinant system production instead of the K5-based one was estimated 50 to 100. Moreover, the eliminase biosynthesis appears to be tightly regulated in *E. coli* K5, thus limiting the possibilities for improving production by modifying culture parameters in the bioreactor (unpublished results). In contrast, the recombinant strain produces the enzyme at very high levels after induction by IPTG.

The *elmA* gene is specific to *E. coli* K5; it is not found in K-12 strains. The presence of the *elmA* gene in the genome of *E. coli* K5 and of various strains of *E. coli* K-12 was tested by PCR amplification with *elmA*-specific primers. No sequence was amplified from K-12 genomic DNA, whereas fragments of the expected size were obtained with the K5 DNA as template (data not shown).

DISCUSSION

We have cloned and expressed a new gene of *E. coli* which encodes a K5 antigen-specific eliminase. Such an activity is not found in *E. coli* K-12, which is devoid of the corresponding gene. The K5 enzyme is not extensively similar to any proteins found in databases, although it contains a small motif which is also found in a phage neuraminidase and is possibly required for activity. Our enzyme is probably involved in the depolymerization of the K5 capsular polysaccharide. The exact biological role of the eliminase remains to be elucidated. The enzyme is releasable into the medium, the NH_2-proximal sequence appearing to be important for the secretion. No peptide signal-like sequence was detected at the N terminus of the molecule. When grown in LB medium, K5 cells first produce HMW *N*-acetyl-heparosan, which tends to be degraded into LMW species at the end of the culture. This activity is relatively weak and tightly regulated, since it is not found when cells are grown on other media, e.g., glycerol-containing synthetic media. In contrast, recombinant eliminase is produced at a high level and only requires the addition of IPTG. *elmA* gene cloning and expression thus open new possibilities for manufacturing a heparin-like bacterial polymer of LMW.

ACKNOWLEDGMENTS

We thank I. and F. Ørskov (Statens Servminstitut, Copenhagen, Denmark) for providing us with *E. coli* K5 strains and phages.

REFERENCES

1. **Gerardy-Schahn, R., A. Bethe, T. Brennecke, M. Mühlenhoff, M. Eckhardt, S. Ziesing, F. Lottspeich, and M. Frosch.** 1995. Molecular cloning and functional expression of bacteriophage PK1E-encoded endoneuraminidase Endo NE. Mol. Microbiol. **16:**441–450.
2. **Gupta, D. S., B. Jann, and K. Jann.** 1983. Enzymatic degradation of the capsular K5 antigen of *Escherichia coli* by *coli* phage K5. FEMS Microbiol. Lett. **16:**13–18.
3. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.
4. **Legoux, R., B. Delpech, X. Dumont, J.-C. Guillemot, P. Ramond, D. Shire, D. Caput, P. Ferrara, and G. Loison.** 1992. Cloning and expression in *Escherichia coli* of the gene encoding *Aspergillus flavus* urate oxidase. J. Biol. Chem. **267:**8565–8570.
5. **Rosenfeld, J., J. Capdevielle, J. C. Guillemot, and P. Ferrara.** 1992. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. Anal. Biochem. **203:**171–179.
6. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
7. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
8. **Vann, W. F., M. A. Schmidt, B. Jann, and K. Jann.** 1981. The structure of the capsular polysaccharide (K5 antigen) of urinary-tract-infective *Escherichia coli* O10:K5:H4. A polymer similar to desulfo-heparin. Eur. J. Biochem. **116:** 359–364.