Cloning and Characterization of the DNA Region Responsible for
Megacin A-216 Production in Bacillus megaterium 216

Antal Kiss,1* Gabriella Balikó,1 Attila Csorba,2 Tungalag Chuluunbaatar,1
Katalin F. Medzihradszky,2,4 and Lajos Alföldi3

Institute of Biochemistry1, Proteomics Research Group2, Institute of Genetics3, Biological
Research Center of the Hungarian Academy of Sciences, Szeged, Hungary;1 Department
of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA
94143-2240, U.S.A.

Running title: Bacillus megaterium megacin A-216 region

*Corresponding author. Mailing address: Institute of Biochemistry, Biological
Research Center of the Hungarian Academy of Sciences, 6726 Szeged, Temesvári krt.
62., Hungary. Phone: +36 62 599 630, Fax: +36 62 433 506, E-mail: kissa@brc.hu
A. C. present address: Inst. of Medical Chemistry, Szeged University, 6720
Szeged, Dóm tér 8., Hungary
ABSTRACT

*Bacillus megaterium* 216 produces, upon induction, megacin A-216, a bacteriocin, which leads to lysis of the producer cell, kills *B. megaterium* and a few other bacterial species. The DNA region responsible for megacinenogeny was cloned in *B. megaterium*. The nucleotide sequence of a 5494 bp long subfragment was determined and function of the genes on this fragment was studied by generating deletions and analyzing their effects on MegA phenotypes. An ORF encoding a 293 amino acid protein was identified as the gene (*megA*) of megacin A-216. BLAST searches detected sequence similarity between megacin A-216 and proteins with phospholipase A2 activity. Purified biologically active megacin A-216 preparations contained three proteins. Mass spectrometry analysis showed that the largest protein is the full-length translation product of the *megA* gene, whereas the two shorter proteins are fragments of the long protein created by cleavage between Glu-185 and Val-186. The molecular masses of the three polypeptides are 32,855, 21,018 and 11,855 Da, respectively. Comparison of different megacin preparations suggests that the intact chain as well as the two combined fragments can form biologically active megacin. An ORF located next to the *megA* gene and encoding a 91 amino acid protein was shown to be responsible for the relative immunity displayed by the producer strain against megacin A-216. Besides the *megA* gene, at least two other genes, including a gene encoding a 188 amino acid protein sharing high sequence similarity with RNA-polymerase sigma factors were shown to be required for induction of megacin A-216 expression.
INTRODUCTION

Megacins are bacteriocins produced by certain strains of the Gram-positive spore-forming bacterium Bacillus megaterium (13, 17, 36, 39). Megacinogeny was first described in 1954 (17). It was observed that some cells in a growing culture of B. megaterium 216 spontaneously produced a substance, which lysed cells of other B. megaterium strains but did not affect most other bacterial species. Production of this substance, later termed megacin A-216, is inducible by UV light (17), N-methyl-N'-nitro-N-nitrosoguanidine (28), or mitomycin C (23). Induction of megacin A-216 synthesis leads to lysis of the culture in 2-3 hours (17). Cell-free supernatants of an induced culture of B. megaterium 216 kill sensitive cells even in 10^4-fold dilutions, whereas the producer strain 216 exhibits relative immunity to the bacteriocin (17). As bacteriocins in general, megacin A-216 has a rather narrow antibacterial spectrum, besides B. megaterium, it is active against some strains of B. subtilis, B. anthracis, Micrococcus auranticum, M. cinnabareus (17, 18, 28).

Many observations indicate megacin A-216 acts by impairing cell membrane integrity of sensitive bacteria. Microscopic pictures show that the cell content is partially released from the killed bacteria while the cell wall appears to stay intact and the shape of the cells is retained (19). Megacin A-216 is active against protoplasts of megacin-sensitive strains (13, 28). In iso-osmotic medium, the cell material of the protoplasts becomes less dense after megacin treatment, but the cell contours remain visible for a long time. These observations suggest that megacin A-216 somehow damages the membrane permeability barrier but does not destroy the membrane completely. The
bacteriocidic effect of megacin A-216 is inhibited by low temperature, suggesting involvement of enzymatic activity in megacin A-216 action (13).

Megacin A-216, first purified by Holland, was found to be an acidic protein precipitating around pH 4.0 (14). Later studies determined that it has a native molecular weight of ~66,000, contains two different subunits with molecular weights of 30 and 15 kDa, respectively, and displays phospholipase A2 activity (27, 28, 39).

A specific inhibitor of megacin A-216, purified from cultures of \textit{B. megaterium} 216, was suggested to confer immunity to the self-produced bacteriocin. The partially purified substance, whose chemical nature was not identified, inhibited the toxic effect as well as the phospholipase activity of megacin A-216 (26).

After recognition of bacteriocin production in other \textit{B. megaterium} strains, a classification of megacins was proposed (types A, B and C). Based on its inducibility by low level of UV irradiation or mitomycin C treatment, megacin A-216 was classified as type A (15). In a broader context, its physical properties and enzymatic activity place megacin A-216 in Class III of bacteriocins produced by Gram-positives (12, 20). There are only a few bacteriocins (megacin A-216, megacin A-19213 and thuricin), which were shown to have phospholipase activity (9, 28, 39). Interestingly, these bacteriocins, although clearly different proteins, have similar producer hosts (closely related Gram-positives) and similar genetic localization (encoded by plasmids) (9, 39).

A study using protoplast fusion and protoplast transformation demonstrated that the genetic determinants of megacin A-216 production, immunity to megacin A-216 and inducibility of megacin A-216 production were transferable into non-megacinogenic \textit{B. megaterium} strains, and acquisition of the MegA phenotypes was associated with the
transfer of a ~47 kb plasmid (pBM309), one of the ten indigenous plasmids present in *B. megaterium* 216 (32). Fragments of pBM309 were cloned in *B. megaterium*, and the genes responsible for megacin production were localized to an approximately 10 kb region of pBM309, however this DNA region was not analyzed beyond establishing a restriction map (30, 41).

The renewed interest in bacteriocins as antimicrobial agents (5, 8, 10) and in *B. megaterium* as industrial microorganism and cloning host (38) prompted us to revisit megacinogeny of *B. megaterium*. The interest in *B. megaterium* is highlighted by the ongoing whole genome sequencing project (http://www.bios.niu.edu/b_megaterium). Our goal is to study the molecular mechanisms underlying the megacin A-216 action and the relative immunity displayed by the producer strain. The phospholipase A2 activity and its associated natural inhibitor (26) make the megacin A-216 system particularly interesting because phospholipase A2 enzymes play important roles in a wide range of physiological and pathophysiological processes (3), as well as in bacterial virulence (35). Because the recombinant plasmids constructed previously (30, 41) were not available, as a first step toward characterization of the megacin A-216 system, we re-cloned the megacin A-216 region, determined its DNA sequence, and used deletion mapping to identify the genes controlling megacin A-216 production and immunity. Assignment of the *megA* gene was also confirmed by mass spectrometry analysis of purified megacin A-216. We report data showing that the two subunits previously detected in megacin A-216 preparations (39) are proteolytic cleavage products of the protein encoded by the *megA* gene.
MATERIALS AND METHODS

Strains, plasmids and growth conditions. The *B. megaterium* strains 216 (MegA⁺, MegC⁺), KM (Spo⁻, megacin-sensitive), AL (MegA⁺, Arg⁻, Leu⁻, Str⁻, pBM309) and THT (KM, Thr⁻, His⁻, Trp⁻, Str⁻) were described previously (17, 32). The kanamycin-resistant mutant of the THT strain was obtained from K. Fodor. *E. coli* ER1821 F⁻ glnV44, e14 (McrA⁻) endA1 thi-1 Δ(mcrC-mrr)114::IS10 obtained from New England Biolabs was used as *E. coli* cloning host.

To obtain an *E. coli* – *B. megaterium* shuttle vector with a single PstI site, part of the polycloning site of the 4.7 kb plasmid vector pHY300PLK (16) was deleted by digestion with EcoRI and XbaI, filling-in the ends by Klenow polymerase and ligation of the blunted ends to yield pHY301. All recombinant plasmids reported here are based on pHY301. Bacteria were routinely grown in LB medium at 30°C (*B. megaterium*), or at 37°C (*E. coli*). Ampicillin (Ap) was used at 100 µg/ml, tetracycline (Tc) at 10 µg/ml (*B. megaterium*), or at 12.5 µg/ml (*E. coli*), and kanamycin (Kn) at 50 µg/ml concentration.

DNA techniques. Plasmid DNA was introduced into *B. megaterium* by protoplast transformation (42). Plasmid preparation, restriction digestion, agarose gel electrophoresis and cloning of DNA fragments were carried out using standard procedures (33). For plasmid preparation from *B. megaterium*, cells were first converted to protoplasts by incubation with lysozyme to facilitate lysis. DNA sequence was determined by an automated sequencer.
**Megacin assay.** Five µl aliquots of overnight cultures were pipetted and allowed to dry onto the surface of LB agar plates. After incubation at 30°C for 6 hr, the plates were overlaided with LB/Kn soft agar (0.4-0.5%) containing 1/100 vol of a dense culture of the Kn\(^R\) mutant of *B. megaterium* strain THT, and incubated at 30°C overnight. Kanamycin was added to the soft agar to inhibit growth of the colonies tested. A clear zone surrounding the colony identified clones producing megacin.

To study UV light induction of megacin A-216 production, the agar plates with the dried test cultures were incubated at 30°C for 2 hr, then irradiated with UV light (germicid lamp) for 15 sec. After the UV-treatment the plates were returned to 30°C for 4 hr, then megacin production was assayed as described above.

Immunity to megacin A-216 was estimated by titrating a cell-free lysate of a mitomycin C-induced *B. megaterium* 216 culture on the clone tested, and, as a control, on the sensitive strain *B. megaterium* KM(pHY301). Five µl aliquots of serial dilutions of the lysate were pipetted onto the surface of a 3 ml soft agar layer containing 10 µl of a dense *B. megaterium* culture. After an overnight incubation at 30°C, the level of resistance was assessed by determining the highest dilution that caused a clear spot in the bacterial lawn. Under the conditions of the assay, the 10\(^5\)-fold dilution still caused lysis of the sensitive KM strain, whereas clones carrying the intact megA region were not affected even by the undiluted lysate.

**Protein purification.** Megacin A-216 was purified by a modification of methods described previously (14, 39). An overnight culture of *B. megaterium* 216 was diluted 1:100 into fresh LB medium. After shaking at 30°C for two hours, mitomycin C was
added to 0.5 µg/ml concentration, and shaking was continued for 3-4 hours until lysis occurred.

All steps of purification were performed at 4°C. Cell debris was removed by centrifugation. Proteins were precipitated from the supernatant by adding ammonium sulfate to 80% saturation. The precipitated material was dissolved in PC buffer (20 mM sodium phosphate pH 6.0, 10 mM 2-mercaptoethanol, 0.1 mM EDTA and 5% glycerol) and dialyzed against the same buffer. Proteins precipitated during dialysis were removed by centrifugation. The solution was treated with 2% streptomycin sulfate to remove nucleic acids. After centrifugation, the supernatant was dialyzed against PC buffer and loaded onto a DE52 (Whatman) anion exchange column equilibrated with the same buffer. Proteins were eluted with a 0 – 0.5 M linear NaCl gradient in PC buffer. Fractions were assayed for megacin A-216 activity by titration on *B. megaterium* KM strain. Pooled active fractions were dialyzed against PC buffer, and loaded onto a hydroxyapatite column equilibrated with PC buffer. Proteins were eluted with a 20 to 400 mM sodium phosphate pH 6.0 gradient containing 10 mM 2-mercaptoethanol, 0.1 mM EDTA and 5% glycerol. Peak megacin-containing fractions were pooled, and the protein was precipitated with ammonium sulfate (80% saturation). The precipitate was dissolved in a minimal volume of buffer S (20 mM sodium phosphate pH 7.4, 150 mM NaCl, 10 mM 2-mercaptoethanol, 0.1 mM EDTA), and further purified by gel filtration on a Sephacryl S-200 column in buffer S.

Purified megacin preparations were dialyzed against PC buffer and stored at –80°C for several months with only slight loss of biological activity. Protein concentration...
was determined by the Bradford reaction, using BIORAD Protein Assay reagent with a bovine serum albumin calibration curve.

**Estimation of native molecular weight by gel filtration.** A sample of purified megacin A-216 (450 µl, 7.5 mg/ml) was loaded onto a 1 x 50 cm Sephacryl S-200 column in S buffer. Dextran Blue 2000 was used for determination of the exclusion volume, and aldolase, bovine serum albumin, chymotrypsinogen A and RNase A (with molecular weights of 158, 67, 25, and 13.7 kDa, respectively) for calibration.

**Electrophoresis of proteins.** SDS-polyacrylamide gel electrophoresis of proteins was performed in 12 % gels using Coomassie Brilliant Blue R-250 staining.

For electrophoresis under non-denaturing conditions, SDS and 2-mercaptoethanol was omitted. Megacin samples migrated toward the cathode. Proteins were extracted from non-denaturing gels by electro-elution. Gel pieces containing unstained bands were cut out using adjacent gel slices with stained megacin bands and pre-stained molecular weight markers as guide. Electro-elution was performed in Little Blue Tank gel electrophoresis equipment (ISCO) using 1x Tris-glycine (25 mM Tris, 192 mM glycine, pH 8.3) in the tank and 0.2x Tris-glycine in the sample-holding chamber. 200 V (1-2 mA per sample holder) was applied for 7 hours.

**Mass spectrometry.** Samples were prepared for mass spectrometry by in-gel digestion (http://ms-facility.ucsf.edu/ingel.html). Briefly, disulfide bridges were reduced by dithiotreitol then the free Cys residues were alkylated by iodoacetamide. Samples were
incubated with side-chain protected porcine trypsin (Promega) for 4 hours at 37 °C. Unfractionated digests were analyzed on a Bruker Reflex III MALDI-TOF mass spectrometer in reflection mode. One µl of 10 µl digest solution was loaded with 1 µl 2,5-dihydroxybenzoic acid solution (Agilent Technologies) as a matrix for the MS analysis. External calibration was applied. PSD analysis was performed in order to obtain sequence information of selected peptides.

**Bioinformatic methods.** DNA and protein sequence similarity searches were performed using the BLAST program package (1, 2), pairwise comparison of sequences by the BLAST 2 Sequences program (37), conserved domain searches by the CD-Search program (22) and PROSITE signature searches by the ScanProsite program (6). In all cases default settings were used.

**GenBank accession number.** Nucleotide sequence data have been deposited in GenBank (accession number: EU014074).
RESULTS

Cloning of the DNA region responsible for megacinogeny. Plasmid pBM309 carrying the genetic determinants of megacin A-216 production and immunity (32) was purified from *B. megaterium* AL, a strain harboring only this plasmid. We chose to clone the ~10 kb PstI fragment, which, in a previous study (30), was shown to carry the *megA* genes. To obtain a plasmid shuttle vector with a unique PstI site, pHY300PLK (Ap<sup>R</sup>, Tc<sup>R</sup>), originally developed as a *B. subtilis* - *E. coli* shuttle vector (16) was modified to yield pHY301, as described in Materials and Methods. PstI fragments of pBM309 were ligated to PstI-digested pHY301, and the ligated DNA was used to transform protoplasts of the plasmid-less *B. megaterium* THT strain. Transformed cells were regenerated in a tetracycline-containing batch culture. It was assumed that the acquired capacity of megacin A-216 production would endow clones carrying the *megA* genes with a very strong selective advantage. Restriction analysis of the plasmids purified from Tc<sup>R</sup> clones revealed that all contained the same ~10 kb PstI fragment. One plasmid (pMGC1) was selected for further analysis. In the overlay assay, *B. megaterium* THT colonies harboring pMGC1 produced a narrow clear zone around the colony, which became wider, when the colonies were exposed to a short UV irradiation; this indicated that the cloned PstI fragment contained the genetic elements associated with the MegA phenotypes (Fig. 1). Interestingly, pMGC1 conferred MegA<sup>+</sup> phenotype also to *E. coli* ER1821 cells, although the zone of growth inhibition was smaller than around MegA<sup>+</sup> *B. megaterium* colonies (not shown).

A 5.5 kb HindIII fragment of the pMGC1 insert was subcloned in pHY301 to yield pMGC3, and was subjected to further analysis.
Sequence analysis of the DNA region responsible for megacinogeny. The nucleotide sequence (5494 bp) of the cloned HindIII fragment was determined (GenBank accession number: EU014074). Nine open reading frames (ORFs) starting with ATG, encoding a protein longer than 50 amino acids, and non-overlapping with other ORFs were identified (Fig. 2 and Table 1). Translated sequences of all ORFs were used to search the GenBank database for similar amino acid sequences. Start and stop coordinates of the ORFs and results of the BLASTP search are listed in Table 1. Going from left to right on the map of Fig. 2, the translated product of ORF59 showed medium level of similarity to a hypothetical protein of *Bacillus cereus* AH1134, and weaker similarity to a number of other hypothetical bacterial proteins in the database. The predicted protein (P91) encoded by the next ORF showed marginal sequence similarity to a single protein, a hypothetical protein of *Enterococcus faecalis*. Conceptual translation of ORF293A yielded a protein with moderate level of similarity to proteins with phospholipase A2 activity. Moreover, the BLAST search identified conserved phospholipase domains (PLA2_plant, PLA2_bee_venom_like, Phospholip_A2_2, PLA2c, PLA2_like) matching the C-terminal 30% of P293A. In the light of previous results demonstrating phospholipase activity of purified megacin A-216 (28, 39), detection of sequence similarity to proteins with phospholipase activity suggested that ORF293A is the gene encoding megacin A-216. P62, determined by the next ORF, showed very low level of similarity to hypothetical proteins from diverse sources and medium level of similarity to a hypothetical protein of the taxonomically closely related *B. thuringiensis*. The neighboring gene (ORF85) encodes a protein with medium level of similarity to glutaredoxins and related proteins.
The BLAST search also identified an NrdH conserved domain in P85. The predicted product of the next ORF (293B) shares high sequence identity with a large number of serine proteases. The putative ORF188 product is very similar to bacterial RNA polymerase sigma factors containing the Sigma70_r4 conserved domain. The translational start point of P188 is somewhat ambiguous, because there is an in-frame ATG 9 bp upstream of the ATG at 3361. However, there is no appropriately positioned ribosomal binding site in front of the upstream start codon, thus ATG3361 (Table 1) is more likely to be the initiator codon for this protein. Hypothetical translation of the next ORF (ORF73) yields a protein displaying low level of similarity to hypothetical proteins of different Bacillus species. Finally, the predicted ORF185 product is highly similar to several bacterial cell wall hydrolases with N-acetylmuramoyl-L-alanine amidase activity.

Gram-positive bacteria are characterized by strong ribosomal binding sites (24, 25, 31). Sequences located immediately upstream of the ORFs were screened for putative Shine-Dalgarno sequences. All ORFs have preceding sequences with similarity to consensus Bacillus ribosomal binding sites (34) and to a consensus sequence deduced from the 3’-ends of B. megaterium 16S rRNAs (Table 2). The potentially strongest ribosomal binding sites (showing the highest similarity to the consensus sequence), were detected before ORFs 293A, 91 and 293B (Table 2). A consensus Shine-Dalgarno sequence was detected before ORF59, but the distance between the „core“ GGAGG motif and the start codon is probably too big for efficient function (24, 25).

A BLASTN search identified a short, approximately 70 bp region, located between nucleotide postions 4279 – 4345, which shows strong similarity to tRNA-Cys genes of Gram-positive bacteria. The highest sequence similarity (94%) was detected
with the tRNA-Cys-2 gene of *Bacillus thuringiensis serovar konkukian* strain 97-27 (NC_005957, positions 969811 – 969884) of the *B. thuringiensis* genome. An inverted repeat was found between nucleotides 4488-4521 (within ORF185).

The GC content of the sequenced region is 30.85%, lower than the average GC content of *B. megaterium* DNA (37%) (29), and lower than that of two *B. megaterium* plasmids pBM300 (35.2%) and pBM400 (36.5%), whose sequence has become available recently (21, 34).

To our knowledge, transcription signals of *B. megaterium* have not been systematically analyzed, making promoter assignment difficult. A sequence resembling the *B. subtilis* sigma A promoter consensus sequence TTGACA (-35), TATAAT (-10) (11) was found to precede ORF91.

**Functional mapping of the megA region.** Detection of a conserved phospholipase A2 domain strongly suggested, that ORF293A is the structural gene of megacin A-216. To corroborate this assignment, and identify other genes that might play a role in megacin A-216 production and immunity, deletion derivatives of pMGC3 were constructed by removing restriction fragments, or by subcloning fragments of the original HindIII insert in pHY301. The plasmids were constructed in *E. coli* and subsequently introduced into *B. megaterium* THT to test MegA phenotypes. Extension of the deletions and their phenotypic effects are shown on Fig. 1 and 2.

With regard to megacin production, pMGC1 and its deletion derivatives can be classified in four types:
1) The *B. megaterium* THT clone harboring pMGC1 with the ~10 kb PstI insert produced a narrow zone of growth inhibition, which increased substantially upon UV light irradiation. This phenotype was similar to that of the parental *B. megaterium* 216 strain (Fig. 1).

2) pMGC3 carrying the sequenced 5.5 kb HindIII fragment, and its deletion derivatives (pMGC21, pMGC20, pMGC23), in which only ORF293B or ORF185 were inactivated either alone or in combination, produced a wider inhibitory zone around uninduced colonies than pMGC1 (Fig. 1 and 2). The diameter of the inhibitory zone around UV-irradiated colonies was similar to that around induced colonies harboring pMGC1. The effect of UV light was especially evident when samples from young cultures were tested, and the cells, dried onto the surface of the agar plate, were allowed to grow only for a short period of time (2 hours) before the UV-treatment.

3) Clones belonging to the third class in some experiments produced a narrow clear zone around the colony. This was more often detected with pMGC6 and pMGC10, than with pMGC24, but the phenomenon was, in general, not reproducible. UV irradiation did not increase megacin production (Fig. 1). These plasmids (pMGC6, pMGC10, pMGC24) carried shorter or larger deletions affecting several genes located on the 5.5 kb HindIII fragment, but ORF293A was always intact (Fig. 2).

4) Clones categorized in the fourth class did not show megacin production. Some of the plasmids belonging to this class (pMGC26, pMGC17, pMGC30) lacked intact ORF293A, others (pMGC28, pMGC4, pMGC29) had intact ORF293A but carried deletions affecting ORF73 and the tRNA-Cys-2 gene (pMGC28, pMGC4, pMGC29) (Fig. 2).
Megacin immunity was tested by a plate assay using a high-titer cell extract prepared from \emph{B. megaterium} 216. A large deletion removing all ORFs but ORF91 and ORF59 (pMGC30) did not impair megacin immunity, whereas deletions inactivating ORF91 (pMGC24, pMGC17, pMGC40) led to loss of megacin immunity, suggesting that the putative 91 amino acid protein is responsible for the relative resistance of the producer strain to megacin A-216.

Analysis of the phenotypes of the deletion derivatives allowed us to test, whether in addition to ORF293A and ORF91, also other genes have a role in megacinogeny. The lack of phenotypic change associated with the deletion of ORFs 293B and 185 (see above) suggested that neither the putative protease nor the cell wall amidase plays a role in megacin production. A deletion extending into ORF188 led to severe reduction (pMGC6), whereas deletion of ORF73 and the tRNA-Cys-2 gene alone (pMGC28), or in combination with ORF188 (pMGC4, pMGC29) abolished megacin production. All deletions affecting ORF188 or ORF73/tRNA-Cys-2 led to loss of UV-inducibility of megacin production. These observations point to the crucial role of the predicted sigma factor-like protein (P188) in megacin expression and especially in high-level expression after UV-treatment. ORF73 and/or the tRNA-Cys-2 gene seem to be even more important, because their inactivation had more drastic effect, than deletion of ORF188.

The available deletions did not allow separate testing of ORF73 and the tRNA-Cys-2 gene. Puzzlingly, pMGC10 and in some assays pMGC24, which lack both ORF188 and ORF73/tRNA-Cys-2, showed signs of weak megacin production, but inducibility was lost also in these cases.
Purification and analysis of the megacin A-216 protein in vitro. DNA sequence analysis and deletion mapping described above identified a single gene (ORF293A), which encodes megacin A-216. This finding was surprising in the light of previous results, which showed the presence of two different subunits in purified megacin A-216 preparations (39). To address this question, megacin A-216 was purified from lysates of mitomycin C-induced cultures of B. megaterium 216 and B. megaterium THT(pMGC3). Estimations using the biological assay indicated that megacin activity in lysates of the native host was approximately hundredfold higher, than in lysates of the recombinant clone. Megacin preparations were analyzed by SDS-polyacrylamide gel electrophoresis. B. megaterium 216 lysates contained three dominant bands with mobilities corresponding to ~ 40, 30 and 15 kDa (Fig. 3). The 30 and 15 kDa polypeptides appeared to be present in equal molar amounts, whereas the amount of the 40 kDa species varied in different lysates. During the same purification, the relative intensities of the bands did not change. The 30 and 15 kDa polypeptides corresponded to the α and β subunits described previously (39). In the same study the authors also found a 40 kDa polypeptide, which copurified with megacin A-216, but this longer polypeptide was a minor component of their purified megacin preparations, and was not considered to be part of the megacin A-216 system (39). In our hands the 40 kDa protein, designated γ polypeptide, was a major component of the purified preparation, when purification started from B. megaterium 216 extract. Moreover, this protein was the single dominant protein in lysates of B. megaterium THT(pMGC3), and the main component of purified megacin obtained from the recombinant clone (Fig. 3).
To study their relationship, unfractionated tryptic digests of the three proteins purified from the *B. megaterium* 216 lysate were subjected to MALDI-TOF mass spectrometry analysis. Measured masses were compared with calculated masses of predicted tryptic peptides deduced from the sequence of ORFs 293A, 293B, 191(188) and 185. Peptides detected in the 30 kDa sample matched parts of the N-terminal half of the P293A sequence, whereas peptides of the 15 kDa sample matched the C-terminal part of P293A (Fig. 4). The unfractionated tryptic digest of the protein with the apparent molecular mass of ~40 kDa featured tryptic peptides characteristic for both shorter proteins. However, an abundant component with MH⁺ at m/z 2501.7, not matching any predicted peptide was detected only in the 15 kDa sample (Fig. 4). These data indicate that all three proteins of the purified megacin preparation are products of the same gene (ORF293A), and suggest that the α and β polypeptide chains are cleavage products of the full-length γ chain.

In the MALDI-TOF mass spectra of the tryptic digests of the 40 kDa and 30 kDa proteins masses representing peptides [1 – 14] and [2 – 14] of P293A were detected, which indicated that the initiating methionine residue was post-translationally excised from a fraction of the molecules. MALDI-TOF measurements yielded an average molecular mass of 11,832 Da for the intact β-chain (Data not shown). Assuming that the C-terminus of the peptide was intact, this value suggested that the β-chain starts at Val-186. The calculated mass for the 108 amino acid polypeptide beginning with Val-186 and ending with Met-293 is 11,855, which differs only by 0.2% from the experimental value. The predicted MH⁺ (VKLPVPCFNNSTGCCTFSNNGK) of the unique tryptic peptide (m/z: 2501.7, with carbamidomethyl Cys) was indeed detected in the tryptic digest of the
β chain. Its identity was confirmed by post source decay analysis of the peptide (Fig. 5).

These data indicated that the α and β polypeptides were products of proteolytic cleavage between Gln-185 and Val-186 of the full-length γ chain.

The calculated molecular masses (32,855, 21,018 and 11,855 Da for the γ, α, and β chains, respectively) are lower than the values derived from SDS-PAGE (~40, ~30 and ~15 kDa, respectively). There are no signs of post-translational modifications other than the removal of the N-terminal methionine. Analysis of the amino acid sequence shows that megacin A-216 is a hydrophilic molecule, and that the γ chain probably consists of two different domains, with their boundary at the hypothetical cleavage site between the α and β chains. The α chain is rich in acidic residues, resulting in a calculated pI of 3.92, while the β chain contains many Lys and Arg residues and has a theoretical pI of 9.05.

The native molecular weight of megacin A-216 was estimated by gel filtration on a Sephacryl S-200 column. Megacin A-216 eluted in fractions corresponding to 66 – 68 kDa (not shown), which is in accordance with earlier estimates (39). To test the composition of native megacin, purified megacin was electrophoresed in a 12% polyacrylamide gel under non-denaturing and non-reducing conditions. Two major bands, a more intensive slow-moving and a less intensive fast-moving, were detected (Fig. 6A, lanes 1 and 2). Proteins from both bands were electro-eluted, and analyzed by SDS-PAGE (Fig. 6B). The upper (slow) band contained all three chains detected in the purified megacin preparation, whereas the lower band contained only the α chain. In biological activity tests, only the sample extracted from the slow band had megacin activity. If 2-mercaptoethanol was added to megacin samples prior to electrophoresis in non-denaturing gel, the slow band disappeared (Fig. 6A, lanes 3, 4), suggesting that
native megacin contains intra-chain disulfide bonds characteristic for PLA2s, and these
bonds may be important for maintaining its structure. The α chain does not contain Cys
residues, which can explain why reduction did not have a big effect on its mobility. The
β chain, dissociated by 2-mercaptoethanol treatment from the rest of the megacin
molecule, probably did not enter the gel because its pH is higher (9.05) than the pH of the
gel buffer. It is less clear why the γ chain disappeared after treatment with the reducing
agent. A possible explanation is that disruption of the disulfide bonds yielded a
heterogenous population of molecules with different tertiary structures and thus with
different electrophoretic mobilities.

The purified megacin preparation was active on protoplasts of the sensitive strain
\textit{B. megaterium} KM, whereas \textit{B. megaterium} 216 was immune against our preparation
(data not shown). Biological activity of the preparation did not decrease after one hour of
incubation at 65°C, in agreement with previous observations demonstrating heat stability
of megacin A-216 (14, 28).
DISCUSSION

Previous studies described several aspects of megacinogeny in the \emph{B. megaterium} 216 strain, including the microbiological phenomenon of megacin A-216 production and location of the \textit{megA} genes on one of the large plasmids present in the host strain (14, 17, 19, 28, 30, 32, 39-41). The main goal of the present work was the cloning and characterization of the genes involved in megacin A-216 production and immunity. Phenotypes (megacin production, effect of UV irradiation) of the primary clone carrying the recombinant plasmid pMGC1 were similar to those observed with the parental \emph{B. megaterium} 216 strain (Fig. 1). The \emph{B. megaterium} clone (pMGC3) carrying the 5494 bp HindIII subfragment of the original 10 kb insert exhibited a wider inhibitory zone around non-induced colonies. It remains to be determined whether this difference is due to a gene present on the PstI fragment but missing from the shorter HindIII fragment, or to the higher copy number of pMGC3. The observation, that both pMGC1 and pMGC3 expressed megacin also in \emph{E. coli}, supports the interpretation that all genes playing a role in megacinogeny are present on the HindIII fragment. An ORF encoding a 293 amino acid protein (ORF293A) was identified as the gene encoding megacin A-216. This assignment was based on three lines of evidence:

\begin{enumerate}
\item Results of deletion mapping. Deletion of ORF293A invariably resulted in MegA- phenotype.
\item Amino acid sequence similarity between the protein encoded by ORF293A and several known proteins with phospholipase A2 activity.
\item Results of mass spectrometry analysis of tryptic peptides obtained from purified megacin A-216.
\end{enumerate}
Megacin preparations purified from *B. megaterium* 216 contained two main proteins with apparent molecular weights of 30 and 15 kDa and a co-purifying 40 kDa protein. These results were in accordance with previous observations (39). Mass spectrometry analysis proved that all three polypeptides are encoded by ORF293A and there is a precursor-product relationship between them. The largest protein (termed \( \gamma \) chain) is the full-length product of ORF293A, whereas the shorter \( \alpha \) and \( \beta \) chains appear to be products of proteolytic cleavage between Gln-185 and Val-186 of the \( \gamma \) chain. The calculated molecular masses of the \( \gamma \), \( \alpha \) and \( \beta \) chains (32,855, 21,018 and 11,855 Da, respectively) are substantially lower than the values deduced from electrophoretic mobility in denaturing gels. The likely reason of the anomalous migration, at least in the case of the \( \alpha \) and \( \gamma \) chains, is the highly acidic composition (\( pI \): 3.92 and 4.51, respectively).

Our purified megacin preparations contained all three chains. The \( \alpha \) and \( \beta \) subunits were always present in approximately equal amounts, whereas the amount of the \( \gamma \) chain was variable. In preparations purified from the parental strain *B. megaterium* 216 the two shorter chains were more abundant, whereas preparations purified from *B. megaterium* THT(pMGC3) were dominated by the \( \gamma \) chain (Fig. 3). Biologically active megacin ran in one band in native polyacrylamide gel and contained all three chains. The most likely interpretation of these results is that the full-length protein, as well as the cleavage products can form active megacin A-216. The native molecular weight of megacin A-216 estimated by gel filtration (~66 kDa) is consistent with an \( \alpha_2\beta_2 \), \( \alpha\beta\gamma \) or \( \gamma_2 \) composition. The hypothetical protease cleaving the primary product has not been identified. It also remains to be determined what role the protease cleavage has in
megacin A-216 activity. Preliminary observations suggest that the specific activity of
megacin purified from \textit{B. megaterium} 216 is higher than that of megacin purified from \textit{B. megaterium} THT(pMGC3), suggesting that the specific protease cleavage leads to an
increase in specific activity.

In the C-terminal third of P293A, approximately corresponding to the \(\beta\) chain,
PLA2 domains were predicted by conserved domain search. A PLA2 catalytic motif
\(\text{DXCCXXHDXCY}\) was recognized by PROSITE search. The predicted phospholipase
A2 domain, with the presence of all conserved catalytic residues, is in agreement with the
PLA2 activity measured in megacin A-216 preparations (28, 39), and with observations
showing that megacin A-216 interferes with membrane integrity (13, 19, 28). A calcium-
binding motif characteristic for Ca-dependent phospholipases or a signal sequence
characteristic for secreted phospholipases have not been detected in the megacin A-216
sequence.

BLAST searches of non-redundant databases of all organisms yielded proteins
with low level similarities to megacin A-216. Most of them were phospholipases or
hypothetical proteins, and the similarities were mostly restricted to the C-terminal region
of P293A, where the PLA2 domain is located.

It is the first case, to our knowledge, that the sequence of a bacterial PLA2 with
antibacterial activity became known. It is known that some eukaryotic PLA2 enzymes
display bactericidal activity. Also, there are bacterial phospholipase A2 enzymes, which
have a role in virulence (prophage-associated Sl protein in \textit{Streptococcus pyogenes}
strains, ExoU in \textit{Pseudomonas} strains (refs. in (35)). In bacteria, only a few proteins with
PLA2 activity have been found so far. Recently, many hypothetical protein sequences in
bacterial genome projects have been predicted to contain PLA2 domain. P293A gave the highest BLAST scores with such hypothetical proteins and with the Group A *Streptococcus* prophage-associated PLA2s.

The hypothetical bacterial proteins with the best BLAST scores share the general organization of megacin A-216: the C-terminal part containing the predicted PLA2 domain is preceded by a longer section rich in acidic residues. One of them is the 248 residue long hypothetical protein BT9727_0864 of the *Bacillus thuringiensis* sv konkukian 97-27 strain. Interestingly, in this case, the similarity extends to proteins encoded by flanking genomic regions of the megacin gene (Fig. 7). Six other predicted proteins encoded by the same, approximately 5 kb region of the *B. thuringiensis* genome share sequence similarity with respective proteins encoded by the *B. megaterium* megacin region. In addition, the region contains a tRNA-Cys gene which has 93% identity with the predicted tRNA-Cys gene in the megacin region. However, the arrangement of these genes is different in *B. megaterium* and in *B. thuringiensis*. Comparison of the two genomic regions suggest, that this group of genes probably moved via lateral transfer between the two species and later evolved by recombinational events.

Deletion mapping of the cloned fragment also identified the gene responsible for megacin immunity. This gene (ORF91, *mega*\(^{imm}\)), located adjacently to the *mega* gene, encodes a short, 91 amino acid protein. Viability of cells harboring pMGC24, which has the megacin gene (ORF293A), but lacks the immunity gene (ORF91), is somewhat surprising, and probably indicates the very low level of megacin expression in the absence of ORF188 and ORF73 (see below). Previous studies found in *B. megaterium* 216 a specific inhibitor of megacin A-216, which inhibited both the bacteriocidic action
and the phospholipase activity, and which was suggested to mediate immunity to megacin A-216 (26). Although Ochi et al. did not determine whether the inhibitory substance was a protein, we assume that it was identical with P91. The sequence of the ribosomal binding site preceding ORF91 is more similar to the consensus than that of ORF293A (Table 2), suggesting that the immunity protein is expressed at higher level than megacin A-216.

Although the original goal of this work was cloning and characterization of the megA and megAimm genes, deletion mapping revealed that at least two of three other genes located on the cloned 5.5 kb fragment are also essential for megacin production (Fig. 1 and 2). Of the three adjacent genes, the role of ORF188 was unambiguously demonstrated.

It remains to be determined which of the two other genes (ORF73 or tRNA-Cys-2), or possibly both is(are) required for efficient expression and induction of megacin A-216. ORF188 and ORF73 are preceded by good ribosomal binding sites (especially ORF73, Table 2), suggesting that both encode proteins. The predicted P188 protein shares very high sequence similarity with some sigma factors of Gram-positive bacteria, thus its role might be the modification of RNA polymerase initiation specificity, leading to transcription of the megacin gene. The weak similarity detected between P73 and a few, mostly hypothetical bacterial proteins does not allow a similar prediction for its function. Although ORF73 is more likely to be required for megacin expression, than the tRNA-Cys-2 gene, the role of the latter gene cannot be excluded, given the high Cys content (14/293) of the megacin A-216 protein. In this context it is worth mentioning that the subtle phenotypic difference observed between pMGC10, which usually produced some growth inhibition around the colony, and pMGC29, which always tested negative,
suggests that the glutaredoxin-like P85 might also have some role in megacin A-216
expression. It is tempting to speculate that this effect relates to the putative disulfide
reductase activity of P85 and the large number of cysteines in the megacin protein.

A sequence (AAAAAAGTTGCCATTTGGTCAACTTTTTTTTTTTCTTTTT) resembling rho-independent transcriptional terminators (7) was found downstream of
ORF73, between positions 4191-4226. The lack of a similar sequence between ORF188
and ORF73 suggests that the two genes might form an operon. Studies focusing on the
genetic background of megacinogeny and on the mechanism of megacin action and
immunity are underway in our laboratories.

ACKNOWLEDGEMENTS

This work was supported by the Hungarian Scientific Research Fund (OTKA) grants
T038343 and T049096. We thank Tomoyuki Sako for the plasmid pHY300PLK, Katalin
Fodor for *B. megaterium* THT Kan^R^, and Ibolya Anton and Lilla Lippai Csanády for
technical assistance. We also thank Pál Venetianer for supporting G.B. during the course
of this project and Éva Hunyadi-Gulyás and Éva Klement for help in the early phase of
this work.
REFERENCES


FIGURE LEGENDS

FIG. 1.
Megacin A-216 production by *B. megaterium* THT cells carrying different sections of the megacin A-216 region and by the parental *B. megaterium* 216 strain. Megacin A-216 production was tested by the overlay assay as described in Materials and Methods. Colonies in the bottom row were irradiated by UV light. Numbers refer to recombinant pMGC plasmids (see text and Fig. 2.). B.m., *B. megaterium* 216

FIG. 2.
Functional mapping of the *megA* region. Horizontal bars indicate DNA segments present in the plasmids. Restriction sites used to delete specific DNA segments are shown. ORFs present in the pMGC3 insert and left intact in the deletion derivatives are indicated by filled horizontal arrows and the putative tRNA-Cys gene by empty arrow. Numbers below the arrows in the top row indicate the length of the ORF in amino acids and names above the arrows the putative function derived from BLAST search and functional analysis. *Gltrdx*, ORF encoding a glutaredoxin-like protein; *reg*, ORF encoding a putative regulatory protein playing a role in megacin expression. The capacity to produce megacin A-216 (*MegA* column), immunity to megacin A-216 (*MegA*$_{imm}$ column) and inducibility of megacin production by UV light (UVInd column) is indicated by +, whereas the lack of these features by – signs. Low level of megacin production is indicated by ± sign. *pMGC24*: insert in opposite orientation.
FIG. 3.

Purification of megacin A-216 from *B. megaterium* THT (pMGC3) and *B. megaterium* 216. SDS gel electrophoresis of protein samples obtained in different stages of purification. M, Molecular mass markers (Fermentas); Lane 1, lysate; lane 2, after DEAE-cellulose and Sephacryl-S 200 chromatography; lane 3, lysate; lane 4, after ammonium-sulfate precipitation, streptomycine-sulfate precipitation of nucleic acids and dialysis; lane 5, after DEAE-cellulose chromatography; lane 6, after hydroxyapatite chromatography.

FIG. 4.

MALDI-TOF MS spectra of the unfractionated tryptic digests of a) 40 kDa, b) 30 kDa, c) 15 kDa proteins. Peptides confirmed by PSD analysis are labeled with asterisks. Mass of the predicted N-terminal peptide is underlined.

FIG. 5.

PSD spectrum of the predicted N-terminal peptide (MH<sup>+</sup> = 2501.7) of the β subunit. PSD segments were smoothed and baseline fitted before merging. For simplicity only the y and b ions are labeled (Nomenclature according to (4)). Cys residues are carbamidomethylated.
FIG. 6.
(A) Electrophoresis of megacin A-216 in 12% non-denaturing polyacrylamide gel at pH 8.8. Lanes 1 and 2, two fractions of purified megacin A-216 after hydroxyapatite chromatography; lanes 3 and 4, samples of the same fractions after reduction with 2-mercaptoethanol; S, slow-migrating band; F, fast-migrating band. (B) SDS-PAGE analysis of proteins electro-eluted from the slow and fast bands in a non-denaturing gel. Lane M, molecular weight marker; lanes P, purified megacin A-216 preparation after Sephacryl-S 200 chromatography; lane S, proteins extracted from the slow band; lane F, proteins extracted from the fast band. (C) Scheme illustrating some important features of the megacin A-216 protein. Filled horizontal bar, segment 1-185 corresponding to the α chain; empty bar, segment 186-293 corresponding to the β chain with the conserved PLA2-like superfamily domain. The protease cleavage site between residues Glu185 and Val186 is indicated by an arrow and positions of cysteine residues by lollipops.

FIG. 7.
Comparison of the megacin region of B. megaterium 216 and a region of the B. thuringiensis sv konkukian 97-27 genome (NC_005957). ORFs with predicted protein products, which, in a pairwise comparison, share amino acid sequence similarity between the two strains, are shown by filled arrows. Corresponding ORF pairs are indicated by the same number below the arrows. Numbers in parentheses refer to the length of the protein in amino acids. RefSeq GIs for the B. thuringiensis hypothetical proteins are 49477011
through 49477019 and apply to the ORFs on the map in the order of left to right. The genomic location of the *B. thuringiensis* tRNA-Cys-2 gene is 969811 – 969884.
<table>
<thead>
<tr>
<th>ORF</th>
<th>Start</th>
<th>Stop</th>
<th>Conserved domain with highest similarity</th>
<th>Protein with highest sequence similarity</th>
<th>Protein Database ID</th>
<th>Organism</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>259</td>
<td>80</td>
<td>none</td>
<td>hypothetical protein BcerAH1_14544</td>
<td>167937449</td>
<td>Bacillus cereus AH1134</td>
<td>9e-13</td>
</tr>
<tr>
<td>91</td>
<td>396</td>
<td>671</td>
<td>none</td>
<td>hypothetical prot. EF0543</td>
<td>29375161</td>
<td>Enterococcus faecalis</td>
<td>1.2</td>
</tr>
<tr>
<td>293A</td>
<td>1593</td>
<td>712</td>
<td>PLA2_plant</td>
<td>hypothetical protein BcerW_26065</td>
<td>167956469</td>
<td>Bacillus cereus W</td>
<td>1e-05</td>
</tr>
<tr>
<td>62</td>
<td>1885</td>
<td>1697</td>
<td>none</td>
<td>hypothetical prot. BT9727_0866</td>
<td>49477018</td>
<td>Bacillus thuringiensis sv konkukian 97-27</td>
<td>5e-06</td>
</tr>
<tr>
<td>85</td>
<td>2145</td>
<td>1888</td>
<td>NrdH</td>
<td>glutaredoxin</td>
<td>169824245</td>
<td>Finegoldia magna ATCC 29328</td>
<td>4e-11</td>
</tr>
<tr>
<td>293B</td>
<td>3158</td>
<td>2277</td>
<td>Peptidase_S8</td>
<td>probable intracellular serine protease</td>
<td>167937446</td>
<td>Bacillus cereus AH1134</td>
<td>3e-68</td>
</tr>
<tr>
<td>188</td>
<td>3361</td>
<td>3927</td>
<td>Sigma70_r4</td>
<td>possible DNA-directed RNA polymerase sigma subunit FliA</td>
<td>167937451</td>
<td>Bacillus cereus AH1134</td>
<td>6e-36</td>
</tr>
<tr>
<td>73</td>
<td>3971</td>
<td>4192</td>
<td>none</td>
<td>hypothetical protein BSG1_14318</td>
<td>149183037</td>
<td>Bacillus sp. SG-1</td>
<td>9e-06</td>
</tr>
<tr>
<td>185</td>
<td>4381</td>
<td>4938</td>
<td>Amidase_3</td>
<td>Putative N-acetylmoramoyl-L-alanine amidase</td>
<td>148747751</td>
<td>Geobacillus virus E2</td>
<td>9e-42</td>
</tr>
</tbody>
</table>
Footnote to Table 1.

Open reading frames starting with ATG, and longer than 50 amino acids and non-overlapping with other ORFs are shown. ORFs are identified by the number of the encoded amino acids and are listed in their order of location from left to right as shown on Figure 2. Start and stop coordinates define nucleotide positions within the sequenced 5494 bp fragment. Proteins encoded by ORFs 91 through 185 have GenBank identifiers ABS44966 through ABS44973.

Identified by BLAST search of the non-redundant protein database (March, 2008).
TABLE 2. Putative ribosomal binding sites

<table>
<thead>
<tr>
<th>ORF</th>
<th>Shine-Dalgarno sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>AAGGAGGTGATcgtactgATG</td>
</tr>
<tr>
<td>91</td>
<td>GGAGGaGAttttttaATG</td>
</tr>
<tr>
<td>293A</td>
<td>atAGGAGGaGAAatATG</td>
</tr>
<tr>
<td>62</td>
<td>ttGGAGtaataagtATG</td>
</tr>
<tr>
<td>85</td>
<td>AagGGtGaaGtctaATG</td>
</tr>
<tr>
<td>293B</td>
<td>gaGAGGTGAaaaagATG</td>
</tr>
<tr>
<td>188</td>
<td>AAGGAatgtATgtaATG</td>
</tr>
<tr>
<td>73</td>
<td>tgGGAGGaaAagttATG</td>
</tr>
<tr>
<td>185</td>
<td>agAAGGtGaaagatATG</td>
</tr>
</tbody>
</table>

The ATG start codons are shown in bold. Nucleotides matching the consensus *B. megaterium* sequence (5’-AAGGAGGTGAT) derived from two GenBank entries (DQ660362 and AY180964) for 16S ribosomal RNA are printed with capital letters.