Primary Structure of CTP: CMP-3-deoxy-d-manno-octulosonate Cytidyltransferase (CMP-KDO Synthetase) from Escherichia coli*

(Received for publication, May 28, 1986)

Robert C. Goldman‡, Timothy J. Bolling§, William E. Kohlbrenner‡, Yangkil Kim§, and J. Lawrence Fox§

From the §Antimicrobial Research Division and the ‡Department of Molecular Biology, Abbott Laboratories, Abbott Park, Illinois 60064

The gene coding for CTP: CMP-3-deoxy-d-manno-octulosonate cytidyltransferase (CMP-KDO synthetase), kds B, was previously cloned on a 9-kilobase Pst insert of Escherichia coli DNA into pBR 322 (Goldman, R. C., and Kohlbrenner, W. E. (1985) J. Bacteriol. 163, 256–261). Using a transposon mutagenesis approach we have now located kds B on this insert, which facilitated the isolation and sequencing of a 1.3-kilobase segment of DNA containing kds B and putative RNA polymerase and ribosome binding sites. The primary structure of CMP-KDO synthetase predicted by this nucleotide sequence was verified by amino acid composition and sequence analysis of purified CMP-KDO synthetase and cleavage fragments. Our results show that kds B consists of a 744-base open reading frame coding for a 248-amino acid peptide. The molecular weight of CMP-KDO synthetase calculated from the translated sequence is 27,486, taking into account the presence of several key enzymes involved in the assembly of important macromolecular structure which not only serves as part of the barrier provided by the outer membrane of Gram-negative bacteria, but can contribute to the virulence of a particular bacterial species and affect host-pathogen interactions.

CMP-KDO synthetase is also of considerable interest because of the reaction it catalyzes.

\[ \text{CTP} + \text{KDO} + \text{Mg}^{2+} \rightarrow \text{CMP-KDO} + \text{PP} \]

CMP-KDO, which is the activated precursor used for KDO incorporation into lipopolysaccharide (2), is an unusual monophosphoryl sugar nucleotide. The more common diphosphoryl sugar nucleotides are generally formed from phosphorylated sugars (3). However, CMP-KDO synthetase shows an absolute requirement for KDO (4, 5).

The characterization of CMP-KDO synthetase has been limited by the low levels of this enzyme produced in wild-type Gram-negative bacteria such as Escherichia coli (5). We recently reported the cloning of the gene coding for this enzyme, kds B, as part of an 9-kb Pst insert into pBR322 (6). CMP-KDO synthetase expression was increased 40-fold over normal levels when this new plasmid, designated pRG1, was introduced into E. coli (6). Using the enhanced levels of protein available from this overproducer it was possible to demonstrate that the \(\beta\)-pyranose form of KDO, a minor form in solution, is the actual substrate used in CMP-KDO formation (7).

The isolation of the kds B gene was also an important first step in determining the primary structure of CMP-KDO synthetase by nucleic acid sequencing. We report here the strategy used in locating kds B on pRG1, the isolation and sequencing of a 1.3-kb nucleic acid fragment containing kds B, and the verification of the translated sequence by analysis of purified CMP-KDO synthetase and peptide fragments.

EXPERIMENTAL PROCEDURES AND RESULTS2

DISCUSSION

We have subcloned and sequenced a 1.3-kb fragment of nucleic acid from pRG1 that contains kdsB, the structural gene for CMP-KDO synthetase. The molecular weight of CMP-KDO synthetase calculated from the translated gene sequence (27,486) is somewhat lower than previously predicted by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (32,000–40,000) (5, 6). We do not feel this discrepancy is significant considering the uncertainties often encountered in determining molecular weights by this method. The close correspondence of the amino acid content of the native
enzyme to that predicted by the nucleic acid sequence, the
unambiguous overlap of the segments of DNA which were
sequenced, and the large stretches of confirmed amino acid
sequence, including the N and C termini, indicate that the
total structural gene has been correctly sequenced and trans-
lated.

The amino acid sequence of CMP-KDO synthetase was
compared to the May, 1985 release of the DNA and protein
data bases, and no extensive homologies of statistical signifi-
cance were found, although it is clear structural homologies
may exist in the absence of sequence homologies (21). How-
ever, there is some homology between residues 115–131 of
CMP-KDO synthetase and residues 120–135 of elongation
factor-Tu (EF-Tu). Residues 115, 118–121, and 131
(Leu...Arg, Gln, Val, Gly...Asn) of CMP-KDO synthetase
correspond to residues 120, 123–126, and 135 of EF-Tu (22).
In EF-Tu, residues 120–126 are located at the end of an α-
helix which connects β-strands 4 and 5, which are adjacent to
the nucleotide binding site (23). Residue 135 (asparagine) is
near the end of β-strand 5 and positioned over the guanine
ring (23). This homology is interesting because this same
region of EF-Tu is also homologous to regions of other
nucleotide binding proteins (24). Further studies are necessary
to determine whether this homology reflects common struc-
tural features shared by CMP-KDO synthetase and other
nucleotide binding proteins such as EF-Tu. One aspect of the
structure of CMP-KDO synthetase predicted by the sequence
which is common in nucleotide binding proteins is an alter-
ating α-β structure (25). The secondary structure predictions
for CMP-KDO synthetase (Fig. 4, Miniprint) suggest an
alternating α-β structure between residues 100 and 135.

Proteins expressed at high levels, such as ribosomal pro-
teins and EF-Tu, are coded predominantly by codons which
interact with abundant species of aminocyl-tRNA. Codon
usage in CMP-KDO synthetase follows this pattern in some
cases (arginine, leucine, isoleucine, glutamine, lysine, and
glycine) but not in others (proline, threonine, alanine). CMP-
KDO synthetase represents less than 0.01% of total cell
protein when coded by the single chromosomal gene (5); thus
use of the most abundant aminocyl-tRNA species may not
be necessary for an adequate rate of synthesis.

Downstream from the TAA stop codon of CMP-KDO syn-
thetase are two potential sites for reinitiation of protein
synthesis. ATG is present at base pairs 1166 and 1286, both
of which are preceded by possible ribosome binding sites.
Open reading frames proceed from these ATG codons to the
end of sequenced DNA, and, if used, they could represent the
start of synthesis of up to a 40,000-dalton polypeptide (based
on the insert size in pRG-1). We have isolated these two cold-
sensitive mutations by localized mutagenesis which map ad-
jaent to kds B on the Salmonella typhimurium chromosome,
make a lipid A precursor of an as yet undefined structure, and
are cured by plasmids pRG (but not by the Tn 5-containing
derivatives of pRG1 mentioned above).5 We are currently

4 R. C. Goldman and E. Devine, unpublished data.

investigating the possibility that kds B is the promoter-prox-
imal gene in a polycistronic operon, followed by a second gene
defined by the cold-sensitive chromosomal mutants.

The biosynthesis of KDO and its activation by CMP-KDO
synthetase for transfer into lipid A are essential for the proper
assembly of lipopolysaccharide and are thus critical for growth
and contribute to the virulence of Gram-negative bacteria.5
Information regarding the sequence and expression of kds B,
as well as the catalytic events involved in CMP-KDO forma-
tion, will lead to a better understanding of how Gram-negative
bacteria control the synthesis and assembly of lipopolysac-
charide, an important surface antigen, toxin, and virulence
factor.

Acknowledgments—We would like to thank Jackie Klimala for
excellent technical assistance and Joe Meuth for many helpful dis-
cussions.

REFERENCES

388
Res. Commun. 10, 346–351
3216–3221
145, 1273–1280
259–261
14695–14700
Res. 9, 309–321
453
Acad. Sci. U. S. A. 74, 5463–5467
103, 3185–3191
21, 719–722
Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor,
NY
Mol. Gen. Genet. 177, 65–72
475–507
22. Jones, M. D., Peterson, T. E., Nielsen, K. M., Magnusson, S.,
386
5 R. C. Goldman, submitted for publication.
Sequence of CMP-KDO Synthetase

APPENDIX

DNA sequence of nucleic acid fragment containing ada B and corresponding amino acid sequence of CMP-KDO synthetase. The terminal 12 restriction fragment (357-298) is shown. The translated reading frame is indicated by a solid line. The boxed sequences are the 12 and 22 nucleotide restriction fragments containing the ada B gene product. CMP-KDO synthetase. The unboxed section indicates the -35 consensus sequence (237-258) and a potential ribosome binding site (116 and 130). Potential ribosome binding sites are indicated by potential ribosome binding sites.
The sites of insertion which inactivated the bg b gene were determined by digestion with restriction endonucleases Eco RI and Hind III and analysis of digestion products by gel electrophoresis. Protein products coded by plasmids were analyzed by polyacrylamide gel electrophoresis in 3% SDS after in vitro cotranscription-translation in the presence of $^35$S-methionine.

DNA Sequencing

The 4.2 kb EcoR I/Hind III fragment, containing bg b, was purified and subcloned into an M13 mp8 (Hind III) cut plasmid (16). The insert was further reduced in size by removing a 2.1 kb ECO R I fragment of DNA. The new clone was designated pBG10.

DNA from pBG10 was linearized by cutting with Hind III and digested with an amount of Bal 31 exonuclease that removed 100 bp from both the 5' and 3' ends of the DNA sequence. The 3' ends were then filled with 0.1% 2-mercaptoethanal and tale to carboxypeptidase ratio to terminate under nitrogen far Chromatography Cyanogen Bromide Cleavage

2 cyanogen bromide fragments except that the gradient Y~T to the size of the insert was approximated by direct polyacrylamide gel electrophoresis of the phage supernatant.

The 2.1 kb Xho 11Xho ends. the rate of digestion was determined by digestion with TPCK-treated trypsin an 4°C. While the rate of digestion with 8 kb fragments was determined by digestion with 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

SDS-gel electrophoresis of the fragments was conducted on a 15% gel with a 25% gel at the top. The gel was run at 4°C. The gel was stained with Coomassie Blue R 250 and dried (Speed vac) for autoradiography.

Amino Acid Compositions

Amino acid compositions were obtained by hydrolyzing the sample in 5 M HCl containing 0.1% 2 mercaptoethanol and 0.3% phenol at 110°C for 24 hr. Digestion was terminated by adding solid guanidine hydrochloride to a final concentration of 5 M. Digests were separated by HPLC on a SynChropak C-8 column (5 μm x 20 cm) at a linear gradient of 0% acetonitrile to 100% acetonitrile (0 ml/min) at 25°C. The rate of digestion was determined by digestion with TPCK-treated trypsin. The greater the size of the digest, the greater the size of the fragment.

Amino Acid Sequence Determination

Sequence was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

Sequencing was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

Sequence was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

The 4.2 kb EcoR I/Hind III fragment, containing bg b, was purified and subcloned into an M13 mp8 (Hind III) cut plasmid (16). The insert was further reduced in size by removing a 2.1 kb ECO R I fragment of DNA. The new clone was designated pBG10.

DNA from pBG10 was linearized by cutting with Hind III and digested with an amount of Bal 31 exonuclease that removed 100 bp from both the 5' and 3' ends of the DNA sequence. The 3' ends were then filled with 0.1% 2-mercaptoethanal and tale to carboxypeptidase ratio to terminate under nitrogen far Chromatography Cyanogen Bromide Cleavage

2 cyanogen bromide fragments except that the gradient Y~T to the size of the insert was approximated by direct polyacrylamide gel electrophoresis of the phage supernatant.

The 2.1 kb Xho 11Xho ends. the rate of digestion was determined by digestion with TPCK-treated trypsin an 4°C. While the rate of digestion with 8 kb fragments was determined by digestion with 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

SDS-gel electrophoresis of the fragments was conducted on a 15% gel with a 25% gel at the top. The gel was run at 4°C. The gel was stained with Coomassie Blue R 250 and dried (Speed vac) for autoradiography.

Amino Acid Compositions

Amino acid compositions were obtained by hydrolyzing the sample in 5 M HCl containing 0.1% 2 mercaptoethanol and 0.3% phenol at 110°C for 24 hr. Digestion was terminated by adding solid guanidine hydrochloride to a final concentration of 5 M. Digests were separated by HPLC on a SynChropak C-8 column (5 μm x 20 cm) at a linear gradient of 0% acetonitrile to 100% acetonitrile (0 ml/min) at 25°C. The rate of digestion was determined by digestion with TPCK-treated trypsin. The greater the size of the digest, the greater the size of the fragment.

Amino Acid Sequence Determination

Sequence was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

Sequence was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

Sequence was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

Sequence was performed on an Applied Biosystem 470A Gas Phase Sequencer using the modified Sanger's Instrumental and program modifications. An HPLC-grade acetonitrile, without added trifluoroacetic acid throughout, was used to reduce background problem. PTH-Ser and PTH-Thr were identified by reductive amination and phenol: 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

The 4.2 kb EcoR I/Hind III fragment, containing bg b, was purified and subcloned into an M13 mp8 (Hind III) cut plasmid (16). The insert was further reduced in size by removing a 2.1 kb ECO R I fragment of DNA. The new clone was designated pBG10.

DNA from pBG10 was linearized by cutting with Hind III and digested with an amount of Bal 31 exonuclease that removed 100 bp from both the 5' and 3' ends of the DNA sequence. The 3' ends were then filled with 0.1% 2-mercaptoethanal and tale to carboxypeptidase ratio to terminate under nitrogen far Chromatography Cyanogen Bromide Cleavage

2 cyanogen bromide fragments except that the gradient Y~T to the size of the insert was approximated by direct polyacrylamide gel electrophoresis of the phage supernatant.

The 2.1 kb Xho 11Xho ends. the rate of digestion was determined by digestion with TPCK-treated trypsin an 4°C. While the rate of digestion with 8 kb fragments was determined by digestion with 32000 units of trypsin. The fragments were run on an agarose gel and the greater the size of the digest, the greater the size of the fragment.

SDS-gel electrophoresis of the fragments was conducted on a 15% gel with a 25% gel at the top. The gel was run at 4°C. The gel was stained with Coomassie Blue R 250 and dried (Speed vac) for autoradiography.
Sequence of CMP-KDO Synthetase

Figure 2. Sequencing strategy for KdsB. The arrows show the site of the nucleotide attachment and the direction of sequencing. The solid arrow indicates sequencing performed from the TAA stop codon, whereas the open arrow indicates sequencing performed from the AUG start codon.

The nucleotide sequence of a 1.3 kb segment of DNA containing the KdsB gene is shown in the Appendix. The KdsB amino acid sequence of intact CMP-KDO synthetase was determined by a combination of mass spectrometry and multiple digestion strategies. The recombinant protein is purified by affinity chromatography and gel filtration. The two major species observed by SDS-PAGE are designated 7 and 8. The 7 species contains the aminoterminal amino acid residues 1-208, whereas the 8 species contains residues 1-246. The amino acid sequence is shown using the standard one-letter code. The sequence is consistent with the predicted sequence.

The amino acid sequence of the N-terminal region of the KdsB gene is shown in Table 1. After digestion with cyanogen bromide, the peptide was isolated and sequenced. The amino acid sequence is shown using the standard one-letter code. The sequence is consistent with the predicted sequence.

The amino acid sequence of the C-terminal region of the KdsB gene is shown in Table 2. After digestion with cyanogen bromide, the peptide was isolated and sequenced. The amino acid sequence is shown using the standard one-letter code. The sequence is consistent with the predicted sequence.

Secondary structure predictions for CMP-KDO synthetase using DSSP and tandem algorithns (19) and the hydrophobic index calculations of Kyte and Doolittle (20) are shown in Figure 4 and discussed below.

w.e. kohnbreuer and j. meyer, unpublished results.

<table>
<thead>
<tr>
<th>Alanine</th>
<th>Cys</th>
<th>Asp</th>
<th>Asn</th>
<th>Glu</th>
<th>Gly</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Gly</td>
<td>7</td>
<td>6.7</td>
<td>6.3</td>
<td>5.1</td>
<td>2</td>
</tr>
<tr>
<td>Val</td>
<td>23</td>
<td>23.9</td>
<td>24.4</td>
<td>24.7</td>
<td>2</td>
</tr>
<tr>
<td>Ile</td>
<td>8</td>
<td>9.0</td>
<td>9.0</td>
<td>9.0</td>
<td>2</td>
</tr>
<tr>
<td>Leu</td>
<td>5</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>2</td>
</tr>
<tr>
<td>Thr</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Phe</td>
<td>7</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>2</td>
</tr>
<tr>
<td>Tyr</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Ser</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Lys</td>
<td>1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td>Arg</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>109</td>
<td>109</td>
<td>109</td>
<td>22</td>
</tr>
</tbody>
</table>

The amino acid composition of CMP-KDO synthetase is shown in Table 3. The composition is consistent with the predicted sequence. The predicted sequence is shown using the standard one-letter code. The sequence is consistent with the predicted sequence.

<table>
<thead>
<tr>
<th>Alanine</th>
<th>Cys</th>
<th>Asp</th>
<th>Asn</th>
<th>Glu</th>
<th>Gly</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Gly</td>
<td>7</td>
<td>6.7</td>
<td>6.3</td>
<td>5.1</td>
<td>2</td>
</tr>
<tr>
<td>Val</td>
<td>23</td>
<td>23.9</td>
<td>24.4</td>
<td>24.7</td>
<td>2</td>
</tr>
<tr>
<td>Ile</td>
<td>8</td>
<td>9.0</td>
<td>9.0</td>
<td>9.0</td>
<td>2</td>
</tr>
<tr>
<td>Leu</td>
<td>5</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>2</td>
</tr>
<tr>
<td>Thr</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Phe</td>
<td>7</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>2</td>
</tr>
<tr>
<td>Tyr</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Ser</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Lys</td>
<td>1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>2</td>
</tr>
<tr>
<td>Arg</td>
<td>2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>109</td>
<td>109</td>
<td>109</td>
<td>22</td>
</tr>
</tbody>
</table>