Primary Structure of Heat-labile Enterotoxin Produced by
Escherichia coli Pathogenic for Humans*

(Received for publication, October 19, 1983)

Tatsuo Yamamoto‡, Takaaki Tamura†, and Takeshi Yokota‡
From the ‡Department of Bacteriology, School of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo-ku, Tokyo, Japan and the †Department of Microbiology, School of Medicine, Keio University, Shinjuku-ku, Tokyo, Japan

Heat-labile enterotoxin of Escherichia coli pathogenic for humans (LT) or for piglets (LTp) and Vibrio cholerae enterotoxin (CT) are structurally and functionally similar toxins. We have determined the complete nucleotide sequence of the toxA gene which encodes the subunit A of LT (LT A). The deduced amino acid sequence consists of 258 residues including a signal peptide of 18 residues. According to the previously completed LT B sequence (103 residues), the predicted holotoxin (1A5B) of LT comprises 755 residues and has $M_r = 87,866$. With respect to LT A and LT B, secondary structures, local hydrophilicity, and sites for antigenic determinants were predicted. Both codon usage and G + C content of the toxA gene and the LT B gene (toxB) were markedly different from those observed with several E. coli chromosomal genes. Its relatively low G + C content was rather close to that of the V. cholerae chromosome. Although the toxA gene shares a common ancestor with the LTp A gene (eltA), the two genes are apparently distinguishable from each other in their sequences. Like LT B reported previously, the predicted sequence of the catalytic fragment LT A1 also showed more homology to that of CT A1 than did that of LTp A1. In contrast, unique sequences were found in LT A2.

LTp produced by enterotoxigenic Escherichia coli is very similar to Vibrio cholerae enterotoxin (cholera toxin) in terms of molecular structure and mode of action. Like CT, the LT holotoxin consists of one molecule of subunit A (LT A) and five molecules of subunit B (LT B) (1, 2). LT A and LT B are immunologically cross-reactive with CT A and CT B, respectively (3, 4). Also, like CT, LT A undergoes post-translational processing to generate A1 and A2 peptides, which are presumed to be linked to each other by a single disulfide bond (5–7). The A2 portion is presumably important for binding of the A1 portion to subunit B (1). In the case of both LT and CT, subunit B binds to the receptor (G$_{120}$, ganglioside) of the target cells (8–12). Subsequently, subunit A (A1–A2) penetrates across the membrane and the A1 portion, separated from the rest of the molecule, catalyzes ADP ribosylation of the GTP-binding regulatory component of adenylate cyclase in the presence of NAD, resulting in elevation of intracellular cyclic AMP levels (1, 7, 13, 14). In vitro construction of hybrid toxins between LT and CT using the purified subunits A and B has been demonstrated, and it has been confirmed that the corresponding subunit from either toxin has a comparable level of biological activity (15).

Recent accumulation of knowledge has shown that LT from E. coli pathogenic for humans (LT) is distinct from LTp from E. coli pathogenic for piglets (LTp). Nonidentity of the two toxins has been confirmed with respect to immunological properties (4, 16–19), isoelectric points (20), electrophoretic mobility on sodium dodecyl sulfate-polyacrylamide gels (20), and amino acid sequence (21).

Elucidation of the functional mechanism of subunits A and B of those toxins (LT, LTp, and CT) requires detailed knowledge of the structure of the active site. For this, it is necessary to determine the primary structure of the subunits. Especially, the finding of amino acid sequences shared by the three similar toxins (LT, LTp, and CT) would help to gain an understanding of functionally important sequences. Information is totally lacking for sequences at the functional site(s). Comparative analysis of the three sequences (LT, LTp, and CT) would also enable us to predict antigenic determinants which are common to the three toxins as well as those that are unique to each toxin. Knowledge of the antigenic determinants of these toxins is important because of the increasing recognition of the effectiveness of toxoid producing live vaccine candidates against human diarrheal diseases (22). The entire sequences of LTp A and LTp B have already been predicted from their DNA sequences (23, 24), and the entire sequence of CT B has also been determined (25).

In an earlier paper (26), we had predicted the entire sequence of LT B from its DNA sequence, and revealed that LT B more closely resembles CT B than does LTp B. In this paper, we describe the complete nucleotide sequence of the LT A gene (toxA) and the deduced amino acid sequence of LT A. Thus, the entire sequence of the LT subunits has now been completed. Using the predicted amino acid sequences, we discuss secondary structures, local hydrophilicity, and sites for antigenic determinants of LT A and LT B. Homology between LT A, LTp A, and CT A is also discussed.

EXPERIMENTAL PROCEDURES

Enzymes and Chemicals—Restriction endonucleases Hph I and HgiAI were purchased from New England Biolabs. Restriction endonucleases BclI and DdeI and bacterial alkaline phosphatase were purchased from New England Biolabs. Restriction endonucleases and T4 polynucleotide kinase were obtained from Takara Shuzo (Kyoto, Japan). [γ-32P]ATP (5,000 Ci/mmol) was purchased from Amersham Corp.

Plasmids—Plasmid pJY27 is a recombinant plasmid consisting of

* This investigation was supported by Grants-in-Aid for scientific research for the Ministry of Education, Science, and Culture of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom all correspondence should be addressed.

‡ The abbreviations used are: LT, heat-labile enterotoxin; LTp, heat-labile enterotoxin from E. coli pathogenic for piglets; CT, cholera toxin.
cloning vector pACYC177 and an LTh coding, 2.4-kilobase HindIII fragment, which originates from a plasmid of enterotoxigenic E. coli H10407 (serotype 078:H11) isolated from a patient with diarrhea. Construction and characterization of pJY27 have previously been described in detail (27).

**Plasmid DNA Isolation**—For BclI digestion and 32P-labeling experiments, pJY27 DNA was prepared from a DNA adenine methylation-negative (dam) host, E. coli GM3 (kindly provided by Dr. Hisao Uchida, Tokyo University), carrying the plasmid. This technique avoided modification of the BclI site. Plasmid purification procedures have been described previously (28).

**DNA Sequencing**—pJY27 DNA was digested with several restriction endonucleases and the fragments purified in 5% polyacrylamide gels. This was followed by extraction overnight at 37 °C with 10 mM Tris-HCl buffer, pH 8.0, containing 1 mM EDTA. 5' termini of each fragment were dephosphorylated with bacterial alkaline phosphatase. The two 5' termini were then phosphorylated with [γ-32P]ATP and T4 polynucleotide kinase. The fragments were subjected to secondary restriction cleavage with a different restriction endonuclease. The sequence of the singly end labeled fragments was determined by the method of Maxam and Gilbert (29). For sequence analysis, 15 and 20% polyacrylamide, 7 m urea gels were used. After electrophoresis, the gels were overlaid directly on Kodak X-ray film (X-Omat AR6).

**RESULTS AND DISCUSSION**

The LTh operon of E. coli pathogenic for humans is summarized in Fig. 1. It consists of at least two genes, toxA encoding the subunit A of LTh (LTh A) and toxB encoding the subunit B of LTh (LTh B), and is presumably transcribed as a single mRNA (27). The DNA region necessary for LTh production can be cloned as a 2.4-kilobase HindIII fragment (27) as shown in Fig. 1. This fragment was used for DNA sequencing in this study.

**DNA Sequence and Primary Structure of LTh A**—Using the chemical method of Maxam and Gilbert (29), we determined the nucleotide sequence of the DNA region (927 base pairs) in which the entire toxA gene and a further upstream region lie. With respect to the toxA region, 100% of the 5'-3' strand and 99% of the 3'-5' strand were sequenced (Fig. 2). The entire nucleotide sequence of the toxA gene and the deduced amino acid sequence are presented in Fig. 3. The positions of the signal peptide (−18 to −1), the NH2 terminus (1) of LTh A and proteolytic cleavage sites to generate A1 (1–192) and A2 (195–240) were assigned based on the findings described in the assignment of the corresponding regions of LTp A (23) and CT A (5, 30, 32, 33) and using previously determined sequence data of the toxA region at the COOH terminus (36). Comparison of LTh A, LTp A, and CT A sequences are discussed later. Thus, the precursor to LTh A (pre-LTh A) was predicted to consist of 258 amino acids and have Mw = 28,873. The molecular sizes of LTh A, LTh A1, and LTh A2 were calculated to be 27,788, 22,031, and 5,561, respectively, from the predicted amino acid compositions (Fig. 3, Table I). The estimates of LTh A and LTh A1 sizes were in excellent agreement with the LTh A Mw = 27,000 and the LTh A1 Mw = 22,000 determined by sodium dodecyl sulfate-polyacrylamide gel analysis (37). Based on the fact that the LTh holotoxin is composed of one molecule of subunit A and five molecules of subunit B (2), and using the LTh B sequence data which was previously described (26), the holotoxin of LTh was predicted to consist of 755 amino acids and have Mw = 87,866. The predicted molecular size is slightly larger than LTh Mw = 73,000 estimated by gel electrophoresis (16). There was close agreement between the amino acid composition of the LTh holotoxin previously determined (29) and that calculated from the predicted sequences (Table I). The LTh A thus predicted was larger than the previously predicted LTp A (23) by 4 amino acids.

**Initiation Region for Pre-LTh A Synthesis**—The TAAG sequence (UAAG in mRNA) located 9 nucleotides upstream from the toxA initiation ATG sequence (AUG in mRNA) was predicted to be a ribosome binding site (Fig. 3). Its surrounding region was very A/T rich. The tod gene is predicted to be a ribosome binding site (Fig. 3). Its surrounding region was very A/T rich.

**Base Composition and Codon Usage**—In an earlier paper (26), we have revealed that the toxB sequence encoding pre-LTh B has a relatively low G + C content of 37%. The toxA sequence encoding pre-LTh A shown in Fig. 3 also had a low G + C content of 38%. The entire sequence for pre-LTh A and pre-LTh B had a G + C content of 38%, which is closer to that of the V. cholerae chromosome (43-47%) than to that of the E. coli chromosome (50-51%) (38). A part of the CT A- and CT B-coding sequences which corresponds to one-third of the entire sequence for the LTh genes has been described (33); the G + C content of those regions was found to be 31%.
The codon usage of the toxA and toxB genes is shown in Table II. The codon usage in the two genes contributed to the low G + C content. There was a strong preference for using the A- or U-terminated codons over the G- and C-terminated codons. This preference was significant with Arg, Leu, Ser, Thr, Pro, Ala, Gly, Val, Lys, Asn, Glu, Tyr, Cys, Phe, and Ile. A preference was also apparent for A- or U-initiated codons (see Arg and Leu). Furthermore, of six Cys codons present in the two tox genes, none of them was UGC. And the AGG codon (Arg) that was not used in many E. coli genes (Table II) was frequently used in the LTh genes. This pattern of codon usage was quite different from that seen in a number of E. coli genes, and especially the ribosomal protein genes. Interestingly, some of the unique preferential codon usage was also found in the CT A- and CT B-coding sequences (Table II).

From these observations, together with the fact that the LTh genes are carried by a plasmid (41), one would speculate that the LTh genes may originate from species of bacteria whose DNA has a low G + C content. We have found that the LTh genes are flanked by inverted repeats (37) which is one of the important common features of well known transposons (42). This may support the above hypothesis. V. cholerae carrying CT genes on its chromosome (43) would be one representative of the possible bacterial species which can be expected to be original hosts for the LTh genes.

Local Hydrophilicity Profiles and Predicted Secondary Structures of LTh A and LTh B—The deduced amino acid sequences of pre-LTh A (Fig. 3) and pre-LTh B (26) were subjected to the local hydrophilicity analysis of Hopp and Woods (44), and the predicted amino acid sequences of LTh A (Fig. 3) and LTh B (26) were analyzed by computer for secondary structure according to the procedures of Chou and Fasman (45) (Fig. 4).

In the case of LTh A, the percentages of α-helix and β-sheet structures were 14.6 and 22.5%, respectively; in LTh B, 15.1 and 18.8%, respectively, and in A2, 30.4 and 39.1%, respectively. A large random coil structure (corresponding to 21% of the entire A sequence) stranded A1 and A2. β-Turns were predicted in two positions in A1 and in one position in A2. In the case of LTh B, the α-helix and β-sheet structures comprised 19.4 and 41.7%, respectively. No β-turns were predicted. Thus, LTh A1 contained less ordered structures in the molecule compared to LTh A2 and LTh B. To form a rigid structure, LTh A1 may require LTh B2 and/or LTh B. The predicted secondary structure of LTh A was significantly different from that of LTh A2 (23), probably due to distinct

### Table I

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Predicted</th>
<th>Direct analysis LTh A, B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>5</td>
<td>6.62 (6.61)</td>
</tr>
<tr>
<td>His</td>
<td>8</td>
<td>1.72 (1.71)</td>
</tr>
<tr>
<td>Arg</td>
<td>22</td>
<td>5.56 (5.47)</td>
</tr>
<tr>
<td>Asp</td>
<td>16</td>
<td>3.36 (3.36)</td>
</tr>
<tr>
<td>Glu</td>
<td>11</td>
<td>8.58 (8.49)</td>
</tr>
<tr>
<td>Glu</td>
<td>27</td>
<td>13.66 (13.66)</td>
</tr>
<tr>
<td>Thr</td>
<td>15</td>
<td>11.11 (11.11)</td>
</tr>
<tr>
<td>Ser</td>
<td>18</td>
<td>8.09 (8.09)</td>
</tr>
<tr>
<td>Thr</td>
<td>15</td>
<td>11.11 (11.11)</td>
</tr>
<tr>
<td>Ser</td>
<td>18</td>
<td>8.09 (8.09)</td>
</tr>
<tr>
<td>Glu</td>
<td>13</td>
<td>12.45 (12.45)</td>
</tr>
</tbody>
</table>

The predicted amino acid composition of the LTh subunits and the holotoxin. Numbers in parentheses represent percentage.

---

Taken from previous data (26).

LT holotoxin is 1A5B (2).

Taken from Ref. 20.

The COOH terminus of LTh A1 was assumed to be Arg 192 (see Figs. 3 and 5).
amino acid sequences.

Signal peptide regions of both pre-LTh A (positions −18 to −1) and pre-LTh B (positions −21 to −1) were extremely hydrophobic as shown in Fig. 4, and preceded by positively charged amino acid residues:

\[
\begin{align*}
-18 & -17 & -16 & -15 \\
\text{Met-Lys-Asn-Ile} & \text{ in pre-LTh A (Fig. 3)}
\end{align*}
\]

\[
\begin{align*}
-21 & -20 & -19 & -18 & -17 \\
\text{Met-Asn-Lys-Val-Lys} & \text{ in pre-LTh B (26)}
\end{align*}
\]

(positively charged residues are underlined). The role of these sequences may be explained by the loop model of the signal peptide (47) in which the positively charged NH2-terminal region allows the attachment of the signal peptide to the negatively charged inner surface of the cytoplasmic membrane by ionic interaction, and subsequently, the next hydrophobic section of the peptide is inserted into the cytoplasmic membrane. Based on this model, we speculate that the signal peptide of pre-LTh A may play a less efficient role at the initial step of the membrane secretion than that of pre-LTh B, because the signal peptide of pre-LTh A has only one positively charged residue whereas that of pre-LTh B has two positively charged residues at the NH2 terminus. This would account, in part, for the different manner of penetration across the bacterial membrane seen with LTh A and LTh B (48), and also for the formation of the holotoxin (1A5B) (1, 2). Interestingly, a similar observation was also made with the signal peptide regions of pre-CT A and pre-CT B (see Ref. 33 for sequence):

\[
\begin{align*}
-18 & -17 & -16 & -15 \\
\text{Met-Val-Lys-Ile} & \text{ in pre-CT A}
\end{align*}
\]

\[
\begin{align*}
-21 & -20 & -19 & -18 & -17 \\
\text{Met-Ile-Lys-Leu-Lys} & \text{ in pre-CT B}
\end{align*}
\]

(positively charged residues are underlined).

LTh A1 contained large strongly hydrophobic regions in the middle of the molecule, corresponding to one-third area of the entire A1 sequence (positions 69–107 and 113–134 in Fig. 4). No such regions were found in LTh A2 and LTh B.
Rather, hydrophilic regions were significant in both LTh A2 and LTh B. This observation is important because the A1 molecule penetrates across the membrane of the target cells (1, 49). Therefore, this penetration may be due to the large hydrophobic regions (internal membrane penetration site?) present in the middle of the A1 molecule, together with the less rigid structure of the A1 molecule (Fig. 4). In this catalytic component A1, most of charged amino acids were present in the first one-third area of the NH2 terminus and the last one-third area near the COOH terminus (Fig. 4). These terminal segments might be responsible for the catalytic activity.

LTh A1 and LTh A2 each contained only 1 Cys in the molecule (Fig. 3, Table I), and the 2 Cys residues were both located in a hydrophilic region on a large random coil structure (Fig. 4). These Cys residues are presumed to link LTh A1 and LTh A2 by a disulfide bond (7), and under hydrophilic conditions. It has been described that the Trp in subunit B plays an important role in the G, binding reaction (11, 12). This Trp 88 was present in a sequence common to LTh A1 and LTh A2 (Fig. 4). Similarly, Glu 66-Arg-Met-Lys-Asp-Thr 71 predicted with LTp B by Hopp and Woods (44). The latter sequence (positions 79-84) was shared only by LTh B and LTp B (Fig. 4, Ref. 26). Interestingly, a sequence, Glu 79-Ala-Lys-Val-Glu-Lys, that is located at the corresponding region of CT B has been predicted by Hopp and Woods (44) to be an antigenic determinant of CT B.

Nucleotide and Amino Acid Sequence Homology between Pre-LTh A, Pre-LTp A, and Pre-CT A—Comparison of the sequences is shown in Fig. 5. The previously described pre-LTp A (23) and pre-CT A (5, 30, 32, 33, 50) sequences were aligned to give maximum homology to the pre-LTh A sequence. When the pre-LTh A gene (toxA) sequence and the pre-LTp A gene (eltA) sequence were compared, several large, identical sequences were shared by the two, suggesting that the two genes have a common ancestor. However, there were insertions or deletions of nucleotides in the eltA gene which resulted in nonhomologous amino acid sequences (positions 19-21, 82-86, 191, and 102) or deletion of 4 amino acid residues (positions 27 and 90-92). Table III summarizes the per cent homology found in the regions where the CT A gene sequence has been reported or where amino acid sequences of LTh A, LTp A, and CT A can be compared. In an earlier paper, we have pointed out that sequences unique to LTp B are present, however in no instance are sequences present in both LTp B and CT B but missing in LTh B (26). This
FIG. 5
suggests that LTh B more closely resembles CT B than does LTp B. Similarly, no sequences unique to LTh A were confirmed at either nucleotide or amino acid levels, while sequences unique to LTp A were present (e.g. position 4 in Fig. 5). Unexpectedly, sequences unique to LTp A were found also in A2 regions (positions 213 and 238 in Fig. 5); A2 plays a role only in the binding of A1 to B (1). Amino acid regions which were larger than 6 residues and common to LTh A, LTp A, and CT A were found (Fig. 5) and included in Fig. 4.

In the case of another type of E. coli enterotoxin, heat-stable toxin, proteolytic cleavage takes place between the two Asn residues of so-called "prepro heat-stable toxin" to generate a small mature heat-stable toxin peptide present at the COOH terminus (31, 51). CT A may have also two repeating amino acid residues (Arg-Arg at positions 192 and 193 in Fig. 5) at the proteolytic cleavage site. However, in the case of LTh A (or LTp A), the corresponding sequence was Arg-Thr (Fig. 5).

An undersanding of the amino acid sequence of the toxins would be important especially for construction of toxoids. We have previously constructed genes that encode toxoids of LTh and are developing toxoid producing live vaccines against diarrheal diseases. We are grateful for Dr. Teruko Nakazawa for secondary structure prediction by the method of Chou and Fasman (45), Dr. Toshiya Takano for discussion, and Dr. Yoshifumi Takeda for encouragement during this study.

Acknowledgments—We are grateful to Dr. Teruko Nakazawa for secondary structure prediction by the method of Chou and Fasman (45), Dr. Toshiya Takano for discussion, and Dr. Yoshifumi Takeda for encouragement during this study.

REFERENCES

Fig. 5. Comparison of the coding DNA and amino acid sequences of LTh A, LTp A, and CT A. Top, DNA sequences; bottom, amino acid sequences. The DNA and deduced amino acid sequences of LTp A are from Ref. 23. The DNA and deduced amino acid sequences of CT A are from Ref. 33; residues ambiguously assigned are underlined. Amino acid sequences of CT A which had been previously determined from purified peptides are enclosed in parentheses, and were taken from: Ref. 5 (a); Ref. 32 (b); Ref. 50 (c); and Ref. 30 (d). The DNA and amino acid sequences of LTp A and CT A are aligned to those of LTh A in such a way as to generate maximum homology. In the case of LTp and CT, nucleotide or amino acid residues that are identical with those of LTh are shown as dots. Asterisks (*) indicate residues that differ from those of LTh A. Small boxes with asterisks indicate positions of deletion. Amino acid sequences which are greater than 6 residues and common to LTh A, LTp A, and CT A are marked with a box; dotted boxes indicate possible common regions.
Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans.

T Yamamoto, T Tamura and T Yokota


Access the most updated version of this article at http://www.jbc.org/content/259/8/5037

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/259/8/5037.full.html#ref-list-1