We report the complete DNA sequence of the *Escherichia coli* elt A gene, which codes for the A subunit of the heat-labile enterotoxin, LT. The amino acid sequence of the LT A subunit has been deduced from the DNA sequence of elt A. The LT A subunit starts with methionine, ends with leucine, and comprises 254 amino acids. The computed molecular weight of LT A is 29,673.

The A subunit of cholera toxin (CT A) has been shown to be structurally and functionally related to the LT A subunit. Comparison of the primary structure of LT A with the known partial amino acid sequence of CT A indicates that the 2 polypeptides share considerable homology throughout their sequences. The NH₂-terminal regions exhibit the highest degree of homology (91%), while the COOH-terminal region, containing the sole cysteine residue in each toxin, is less conserved (~52%). Alignment of homologous residues in the COOH-terminal regions of LT A and CT A indicates that a likely site for proteolytic cleavage of LT A is after Arg residue 188. The resulting A₂ polypeptide would be 46 amino acids long, would contain a single cysteine residue, and have *Mₘ* = 5261. The elt A nucleotide sequence further predicts that the LT A protein is synthesized in a precursor form, possessing an 18-amino acid signal sequence at its NH₂ terminus.

For most of the past decade it has been known that some enteropathogenic strains of *Escherichia coli* produce a heat-labile enterotoxin known as LT (for review, see (1)). Recent advances in purification of LT from various *E. coli* strains have facilitated studies of its structural and biochemical properties (2, 3). Thus, it is now known that LT is composed of 2 dissimilar subunits, LT A and LT B, which are present in the holotoxin in the ratio of A₁B₅ or A₁B₆ (3). The B subunit functions in binding the toxin to the target intestinal cell, apparently through interaction with gangliosides (4, 5). The A subunit is responsible for the biological effects of the toxin, which it accomplishes by penetrating the cell membrane and activating adenyl cyclase (6, 7). The stimulation of adenyl cyclase is achieved by an NAD-dependent ADP ribosylation of the GTP-regulatory subunit (*Mₘ* = 42,000) of adenyl cyclase (8, 9).

The fact that the disease caused by LT-producing *E. coli* is similar to that produced by *Vibrio cholerae* enterotoxin led investigators to compare the structural and functional properties of the 2 toxins. It was first demonstrated by Gyles and Barnum (10) that the 2 toxins exhibit immunochemical cross-reactivity and hence must have structural similarities. It was subsequently shown that the toxins have similar subunit structures (2, 11-13) and that the A and B subunits of LT are immunochemically related but not identical with the A and B subunits of CT (14, 15). The amino acid sequence of the B subunits of LT (16) and CT (17-19) are now known and have been found to exhibit ~80% homology throughout their primary structure. The LT A and CT A proteins function similarly, stimulating adenyl cyclase by the same mechanism (20, 21). Initially, it was thought that the specific enzymatic activity of LT A was considerably less than that of CT A. Recent studies, however, have demonstrated that the biological activity of LT A and CT A are comparable (2, 3).

These observations raise several interesting questions regarding the similarity of the 2 enterotoxins. 1) Are the amino acid sequences of the A subunits conserved throughout the polypeptide chains or is the conservation limited to certain regions of the protein? 2) If specific regions are strongly conserved, are they associated with special functions, such as NAD binding, ADP ribosylation of adenyl cyclase, the formation of bonds between A and B subunits, and the movement of the A subunit across the cell membrane into the target cell? 3) If amino acid differences are distributed throughout the 2 proteins, are the proteins' overall secondary and tertiary structures still similar?

Answers to these questions depend, in part, upon knowledge of the amino acid sequence of both LT A and CT A. Approximately 35% of the CT A sequence has been determined (22-24). We report here the complete nucleotide sequence of the *E. coli* elt A gene and the amino acid sequence of LT A derived from these data. Comparison of partial amino acid sequences of the 2 toxins shows that they are homologous in varying degrees throughout their primary structures.

Analysis of the nucleotide sequence of the elt genes suggests that they are co-transcribed from a typical *E. coli* promoter (25, 26), that elt mRNA contains 2 Shine-Dalgarno sequences (27) preceding the AUG translation starts for LT A and LT B, and that LT A, like LT B (16), is subject to post-translational modification.

**MATERIALS AND METHODS**

Enzymes and Biochemicals—Restriction enzymes and polynucleotide kinase were purchased from New England Biolabs and Bethesda Research Labs. *E. coli* DNA polymerase I used for nick translation was obtained from Boehringer Mannheim. Bacterial alkaline phos-
phatase was purchased from Worthington Biochemical Corp. γ-[³²P]ATP (3000 Ci/mmol) was purchased from New England Nuclear and α-[³²P]dATP was purchased from Amersham.

Plasmids—Plasmid EWD299 (Fig. 1) is a chimera of pBR313 and a 1.2 x 10⁸-dalton fragment of ENT plasmid P307. Construction of EWD299 has been described in detail previously (28, 29).

Restriction Enzyme Mapping—Restriction fragments were purified by electrophoresis through 5% polyacrylamide gels. DNA was recovered from gel slices by electroelution through 5% acrylamide gel plugs, followed by precipitation with 0.1 volume of 3 M sodium acetate (bp) DNA restriction fragments, prepared as described above, were 5'-end-labeled by treatment with bacterial alkaline phosphatase, γ-[³²P]ATP and polynucleotide kinase. Labeled fragments were cut with a second restriction enzyme to produce singly end-labeled fragments which, after separation on polyacrylamide gels, were sequenced directly. Sequence reactions were electrophoresed through 8% polyacrylamide-7 M urea gels. DNA was dried down onto Whatman paper and autoradiography was performed with a Dupont Lightening-Plus intensifying screen at −70°C.

In some sequencing experiments, restriction fragments were cloned into M13 mp2, using the strain and technique described by Gronenborn and Messing (32). The inserted DNA was then sequenced in its entirety by the Sanger chain-termination method complementary to the lac region in M13 mp2 (32) and sequenced by the method of Smith and Birnstiel (30). Two strategies for sequencing the A cistron were employed. In one strategy used to analyze elt A. The horizontal lines below the restriction map indicate the start point, direction, and length of the nucleotide sequence that was determined in each experiment.

Primary Structure of LT A—Fig. 3 shows the nucleotide sequence of the elt A structural gene and the derived amino acid sequence of LT A. The sequence data indicate that LT A has 254 amino acids and has a calculated M₀ = 29,673. We have previously suggested that LT A is synthesized as a precursor polypeptide containing an 18-amino acid signal sequence (Ref. 35 and see below). Thus, the mature LT A is predicted to have M₀ = 27,588. Previous estimates of LT A size determined by sodium dodecyl sulfate-polyacrylamide gel analysis range from M₀ = 25,500 (12) to 28,720 (8), and 30,000 (2).

Table I summarizes the amino acid composition of LT A predicted from the DNA sequence. For comparison, the amino method, small restriction fragments (100–200 bp) were labeled with ³²P at a single 5'-end and then subjected to enzymatic sequence analysis by the method of Maat and Smith (31). In cases where overlapping DNA fragments were not easily obtainable, small fragments (~300 bp) were cloned into M13 mp2 (32) and sequenced in their entirety by the Sanger chain-terminating method (33). Fig. 2 summarizes the sequencing strategy used to analyze elt A. The horizontal lines below the restriction map indicate the start point, direction, and length of the nucleotide sequence that was determined in each experiment.

RESULTS AND DISCUSSION

A recombinant plasmid constructed by Dallas and co-workers (12, 29), EWD299, carrying the elt genes derived from ENT plasmid P307 was used to prepare elt A cistron DNA for nucleotide sequence analysis. A map of restriction sites and LT cistron locations on EWD299 is presented in Fig. 1. LT A and B genes are transcribed from a single promoter lying at the right-hand end of the inserted LT DNA (29). Two strategies for sequencing the A cistron were employed. In one

![Fig. 1. Map of known restriction sites and LT gene locations on EWD299. The heavy line represents the DNA taken from plasmid P307 and inserted into pBR313 (5). A and B indicate the approximate location of the structural cistrons for LT subunits A and B. The direction of transcription of the LT operon is from right to left (counterclockwise). The HindIII-HindII fragment that contains the majority of the A structural gene is shown above in greater detail. Restriction sites within the HindIII-HindII fragment were determined by the Smith-Birnstiel mapping technique (12).](image)

![Fig. 2. Restriction enzyme map of elt A DNA and the location of LT A within the 1.0-kilobase HindII fragment. The strategy used to determine the nucleotide sequence of elt A is indicated by the horizontal lines below the DNA. The lines are grouped to indicate whether the fragments were sequenced by the Maat-Smith (31) or the M13/Sanger techniques (see "Materials and Methods").](image)
Nucleotide Sequence of E. coli elt A Gene

\[ \text{ARG AAA ACT ACT ACT TTC TTT ATT TTA TTA GCA TGC CGA TAT AGT GCC AGA TTA TAC GGT} \]
\[ \text{Met-lys-Asn-Ile-Thr-Phe-Ile-Leu-Leu-Ser-Pro-Leu-Tyr-Ala-Asn-Gly-Asp-Arg-Leu-Tyr-Arg} \]
\[ ^{18} \text{-18} \]
\[ \text{GCT GAC TCT AGA CCC GAC GAT CAA ATA AAA CTT TCC GCT ACT ATT AGC AGA GCT ATT GAG TAC TTC GAT AGA} \]
\[ \text{Ala-Glu-Ser-Arg-Pro-Pro-Asp-Ile-Ile-Arg-Pro-Asp-Leu-Leu-Ser-Pro-Glu-Asn-Gly-Tyr-Tyr-Asp-Arg} \]
\[ ^{19} \text{19} \]
\[ \text{GGA ACT GAA ACT ATT ATT CTG TAT GAT ACC GGC GAG GAA GCA ACC GGC TTT GTC ACA TAT GAT GAG GGA} \]
\[ \text{Gly-Thr-Glu-Met-Asn-Ile-Leu-Tyr-Asp-His-Ala-Glu-Thr-Glu-Thr-Gly-Val-Arg-Tyr-Asp-Asp-Gly} \]
\[ ^{40} \text{40} \]
\[ \text{TAT GGT TGC ACT TCT CTT ACT TIA GAT CTC GAC TTA GCA GAG TAT ATA TTA TCA GGA TAT TCA GTT GCA} \]
\[ \text{Tyr-Val-Ser-Thr-Ser-Leu-Ser-Leu-Cys-Leu-Ala-Gly-Glu-Ser-Ile-Leu-Ser-Tyr-Ser-Leu-Thr} \]
\[ ^{70} \text{70} \]
\[ \text{ATA TAT ATC CTT ACA GAT ATT ATT ATT TAT GAT GGA ATT AGC GCA ACC GCA GAA GAG GAT GGT} \]
\[ \text{Tyr-Val-Ile-Val-Val-Asn-Met-Asp-Asn-Val-Asp-Val-Ser-Ser-Arg-Pro-His-Pro-Tyr-Glu-Gln} \]
\[ ^{90} \text{90} \]
\[ \text{GAC GTT TGC GCA GAT GAA ACA ATA TAT TCT CAT AGA TCG ATT TAT GCT TCT ATT GAT} \]
\[ \text{Tyr-Arg-Leu-Glu-Gly-Val-Glu-Ser-Ser-Glu-Ile-Glu-Tyr-Glu-Trp-Arg-Val-Asp-Tyr-Glu-Val-Ile} \]
\[ ^{110} \text{110} \]
\[ \text{GAA GCA TTA CAT GAT ACC AGG GAA TAT AGA GAC GCG TAT TAC AGA AAT CTG AAT ATA GCT CCG GCA GAG GAT GGT} \]
\[ \text{Glu-Arg-Leu-His-Arg-Asn-Arg-Glu-Tyr-Arg-Arg-Tyr-Arg-Asn-Leu-Asn-Ile-Asp-Pro-Ala-Glu-Asp-Gly} \]
\[ ^{140} \text{140} \]
\[ \text{TAC AGA TTA GCA GCT TTC CCA GCG CAT CAT GCA TCG GAA GCA AGT TGC CCA CTC GAG TGT GCA AGT} \]
\[ \text{Tyr-Arg-Leu-Ala-Glu-Phe-Pro-Pro-Asp-Asp-Asp-Ile-Asp-Trp-Glu-Ala-Glu-Val-Glu-Val-Ile} \]
\[ ^{170} \text{170} \]
\[ \text{TGG GAA GAT GTA TGA AGA AUC AGA GCT GAT ACT TCT AAT GAG GAC ACC CAC AAT CTC AGC AGA ATA TAT CTC} \]
\[ \text{Cys-Glu-Asp-Arg-Arg-Thr-Ile-Thr-Gly-Asp-Thr-Cys-Asn-Glu-Thr-Glu-Asn-Leu-Ser-Thr-Ile-Tyr-Leu} \]
\[ ^{200} \text{200} \]
\[ \text{AGG GAA TAT CCA TGA AAA CTT AGC CAG AAA TAT GAT GCG ATT TCA GGC AGA GGAT GAG TCA GAT AGA ATT} \]
\[ \text{Arg-Glu-Tyr-Glu-Ser-Ile-Val-Ile-Arg-Asn-Ile-Phe-Ser-Asp-Tyr-Glu-Ser-Val-Ile-Arg-Asn-Arg} \]
\[ ^{220} \text{220} \]
\[ \text{GGG GAT GCA TTA TGA} \]
\[ \text{Asp-Asp-Glu-Leu-STOP} \]
\[ ^{236} \text{236} \]

acid composition of LT A reported by D. Robertson and Clements et al. (3) is given. In general, there is good agreement between the experimental and predicted values. LT A has an unusually high content of charged residues: 24 Arg, 17 Asp, 15 Asn, and 14 Glu. In addition, LT A has an abnormally high content of tyrosine (21 residues) and phenylalanine (11 residues), which may account for its strong affinity for norleucine-Sepharose (2). Sequence Homologies Between LT A and CT A—The amino acid composition of cholera toxin A subunit has been determined in a number of laboratories and is shown in Table I. Notable differences between LT A and CT A are that CT A is richer in Gln and Gln, Asp and Asn, and Gly residues. Lai et al. (25) report that CT A contains 2 lysine and 3 methionine, which is similar to mature LT A which contains 3 lysine and 3 methionine. The sequence of approximately 35% of the estimated 265 amino acids of the CT A subunit has been reported (22-25). We have compared the derived primary structure of LT A with the partial sequence of CT A and have found that considerable homology exists throughout their structures. Fig. 4 presents a comparison of the 2 toxins' primary structures, aligned in such a way as to generate maximum homology. Table II summarizes the percent homology found in the various regions of the 2 toxins. The NH2-terminal regions are highly conserved with residues 5 → 17 and 31 → 54 being completely identical. As shown in Fig. 4, LT A is missing 4 amino acids between residues 20 and 21 which are found in CT A. The middle (residues 90-117) and COOH-terminal regions of the A subunits, although less conserved, also show homology. Using the alignment of the 2 toxin sequences shown in Fig. 4, it appears that the A2 polypeptide of CT lies at the COOH terminus of the A1 polypeptide. There is evidence that LT A, like CT A, is subject to proteolytic cleavage (9) and that LT A activity, although not dependent upon cleavage, is enhanced by it (2, 9, 13). As indicated in Fig. 4, a likely target for proteolytic cleavage of LT A is after Arg 188.

We have previously suggested that LT A is synthesized as

\[ ^{2} \text{Personal communication.} \]
Nucleotide Sequence of E. coli elt A Gene

Table II

<table>
<thead>
<tr>
<th>Amino Acid Residues in LT A</th>
<th>Homology with CT A</th>
<th>Conservation of Unusual Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Residue in LT A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 → 20</td>
<td>75</td>
<td>Met 23</td>
</tr>
<tr>
<td>21 → 53</td>
<td>91</td>
<td>Met 36</td>
</tr>
<tr>
<td>90 → 117</td>
<td>78</td>
<td>Met 90</td>
</tr>
<tr>
<td>173 → 188</td>
<td>50</td>
<td>His 103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>His 177</td>
</tr>
<tr>
<td>A2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 → 20</td>
<td>55</td>
<td>Cys 183</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cys 195</td>
</tr>
</tbody>
</table>

FIG. 4. Comparison of the amino acid sequence of LT A with the partial sequences of CT A1 and CT A2. Homologous amino acids are enclosed in boxes. The amino acid sequence of CT A1 is from Lai et al. (25) and the sequence of the cysteine-containing peptide of CT A2 is from Mendez et al. (22). At the NH2 terminus of A1, 2 proposed sequences of 20 amino acid residues according to (a) Klapper et al. (24) and (b) Kurosky et al. (23) are shown. The numbering system refers to proposed positions of amino acids in LT A.

a precursor polypeptide (35). Maturation of LT A would involve cleavage after an alanine residue (indicated in Fig. 4), resulting in loss of a very hydrophobic 18-residue signal sequence (36–38), producing a protein with Asp at its NH2-terminal sequence (36–38). The NH2-terminal aspartic acid residue in LT preparations supports the conclusion that pre-LT A is cleaved to give the mature form. The signal region presumably facilitates movement of LT A into the periplasmic space (39) or to the outer membrane (2). The fact that LT B subunit has been found to contain a signal sequence (16) strengthens the prediction made from the DNA sequence of elt A that LT A is post-translationally modified. Confirmation of this awaits purification of secreted and intracellular forms of LT A.

Secondary Structure of LT A—The secondary structure of LT A as predicted by the rules of Chou and Fasman (40) is shown in Fig. 5. Based on this analysis, LT A contains 26% a-helix, 24% B-sheet, and 50% random coil. As illustrated in Fig. 5, LT A appears to contain 2 folded domains in its NH2- and COOH-terminal regions, separated by a long sequence of residues devoid of B-turns. The arrows in Fig. 5 indicate the proteolytic cleavage sites for the release of the signal peptide and the A1 polypeptide.

Expression of LT Genes—The genes coding for LT A and B subunits were predicted by Dallas et al. (29) to be transcribed into a polycistron mRNA from a single promoter lying proximal to the A subunit gene. We have determined the sequence of 200 nucleotides proximal to elt A and have found sequences characteristic of E. coli RNA polymerase recognition (35) and binding (10) sites (25, 26). The sequence in this region of the operon is shown in Fig. 6, along with the idealized (35) and (10) sequences thought to be recognized by E. coli RNA polymerase. Based on the similarities of the elt A proposed promoter (Fig. 6) and the consensus sequence for E. coli promoters, it is reasonable to predict that elt A mRNA starts with an A nucleotide (indicated at +1 in Fig. 6) and contains a 34-nucleotide long leader region. Within this leader region is a 4-nucleotide potential ribosome binding site 5'-UAAG-3' which could direct ribosomes to start translation at the AUG codon 9 nucleotides downstream (27). There is another AUG codon 21 nucleotides upstream which could serve as a translation initiation site. However, this AUG (at +14) is not preceded by a potential ribosome binding site. Thus, we propose that translation of elt A mRNA starts at the +1 AUG codon and that the NH2-terminal sequence of LT A is Met-Lys-Asn-Ile-etc. Confirmation of this prediction awaits protein chemistry studies of purified, unprocessed LT A.

The fact that the holotoxin form of LT is composed of IA
FIG. 5. Secondary structure of LT A predicted by Chou and Fasman rules. Arrows indicate the proposed proteolytic cleavage sites which release (a) the 18-amino acid signal sequence and (b) the A2 polypeptide.  

Fig. 6. DNA sequence of elt operon in the regions proximal to the start of LT A and LT B structural genes. I, the hypothesized LT A mRNA start point is indicated as +1. The proposed promoter region contains a sequence at -10 which is generally homologous to the sequence TATAATG thought to be important for core rRNA polymerase binding (26). Residues in the -35 region also exhibit similarity to sequences found in other E. coli promoters. The nucleotides which are statistically most frequently found in these regions (26) are shown above the elt sequence, with the nucleotides aligned for maximum homology. Nucleotides which are the same in the orthologous sequences are shown above the elt mRNA. The region proximal to the AUG start of translation is shown in Fig. 6. Neither LT A nor LT B mRNAs have sequences which can form stable stem and loop structures in the immediate vicinity of the AUG start codon.

Acknowledgments—We are indebted to Dr. William Konigsberg for considerable advice and encouragement during the course of this work. We are also grateful to Dr. David E. Schafer for useful suggestions and for his continued interest in this work. We thank Dr. W. D. Dains and Dr. Stanley Falkow for gifts of plasmid EWD299 DNA, Dr. Nigel Godson for assistance in the M13 cloning and sequencing technology, and Dr. Nigel Harford for helpful discussions.

REFERENCES


Nucleotide Sequence of E. coli elt A Gene

mRNA having a more efficient ribosome binding site or from the secondary structure of elt B mRNA enhancing translation initiation. We have examined the potential of LT mRNA to form stable stem and loop structures in the translation initiation regions (42) and also compared the complementarity of the two AUG-proximal mRNA regions to the 3'-end of 16 S rRNA (27). The sequence of elt A and elt B mRNA proximal to the AUG start of translation is shown in Fig. 6. Neither LT A nor B mRNAs have sequences which can form stable stem and loop structures in the immediate vicinity of the AUG start codons. Nor is there a significant difference in the extent of complementarity of each mRNA to the 3'-end of 16 S rRNA. We are currently examining in vitro the translation efficiencies of elt A and B cistrons. In addition, we are examining by computer modeling the ability of mRNA sequences distal to the AUG start codons to form stable stem and loop structures to determine if other regions of the mRNAs may influence translation efficiency.

subunit and 5 B subunits (12) leads to the question of whether the 2 cistrons are translated with equal efficiencies. The observation that protein synthesis directed by the LT-plasmid EWD299 in mini-cells results in greater production of B than A subunits (12) indicates that they probably are not, although the actual molar ratio of synthesis is uncertain. A higher rate of B subunit synthesis could result from elt B mRNA having a more efficient ribosome binding site or from the secondary structure of elt B mRNA enhancing translation initiation. We have examined the potential of LT mRNA to form stable stem and loop structures in the translation initiation regions (42) and also compared the complementarity of the two AUG-proximal mRNA regions to the 3'-end of 16 S rRNA (27). The sequence of elt A and elt B mRNA proximal to the AUG start of translation is shown in Fig. 6. Neither LT A nor B mRNAs have sequences which can form stable stem and loop structures in the immediate vicinity of the AUG start codons. Nor is there a significant difference in the extent of complementarity of each mRNA to the 3'-end of 16 S rRNA. We are currently examining in vitro the translation efficiencies of elt A and B cistrons. In addition, we are examining by computer modeling the ability of mRNA sequences distal to the AUG start codons to form stable stem and loop structures to determine if other regions of the mRNAs may influence translation efficiency.

Acknowledgments—We are indebted to Dr. William Konigsberg for considerable advice and encouragement during the course of this work. We are also grateful to Dr. David E. Schafer for useful suggestions and for his continued interest in this work. We thank Dr. Walter Dains and Dr. Stanley Falkow for gifts of plasmid EWD299 DNA, Dr. Nigel Godson for assistance in the M13 cloning and sequencing technology, and Dr. Nigel Harford for helpful discussions.

REFERENCES