Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4364–4368, July 1987
Biochemistry

Nucleotide sequence of the Shiga-like toxin genes of *Escherichia coli*

(*Shigella dysenteriae*/ricin)

STEPHEN B. CALDERWOOD*, FRANCOIS AUCLAIR†‡, ARTHUR DONOHUE-ROLFE†, GERALD T. KEUSCH†, AND JOHN J. MEKALANOS*§

*Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115; and †Division of Geographic Medicine and Infectious Disease, New England Medical Center/Tufts University School of Medicine, Boston, MA 02111

Communicated by A. M. Pappenheimer, Jr., March 5, 1987

ABSTRACT **We have determined the nucleotide sequence of the *sltA* and *sltB* genes that encode the Shiga-like toxin (SLT) produced by *Escherichia coli* phage H19B. The amino acid composition of the A and B subunits of SLT is very similar to that previously established for Shiga toxin from *Shigella dysenteriae* 1, and the deduced amino acid sequence of the B subunit of SLT is identical with that reported for the B subunit of Shiga toxin. The genes for the A and B subunits of SLT apparently constitute an operon, with only 12 nucleotides separating the coding regions. There is a 21-base-pair region of dyad symmetry overlapping the proposed promoter of the *slt* operon that may be involved in regulation of SLT production by iron. The peptide sequence of the A subunit of SLT is homologous to the A subunit of the plant toxin ricin, providing evidence for the hypothesis that certain prokaryotic toxins may be evolutionarily related to eukaryotic enzymes.**

Strains of *Escherichia coli* that produce a cytotoxin for Vero cells were first described in 1977 (1). Subsequently this toxin was shown to have immunological cross-reactivity, identical biological activity, and similar subunit structure to Shiga toxin (2, 3), an iron-regulated cytotoxin produced by *Shigella dysenteriae* 1 (4). Because of these similarities, the name Shiga-like toxin (SLT) has been proposed for the Vero cell cytotoxin (5). These toxins contain a single A subunit responsible for inhibiting protein synthesis through the catalytic inactivation of 60S ribosomal subunits (6) and multiple copies of a B subunit that bind the toxin to receptors on the cell surface (4, 7). Since their original description, SLT-producing strains of *E. coli* have been implicated in a wide variety of human disease states including sporadic diarrhea in adults, outbreaks of neonatal diarrhea (8), epidemics of hemorrhagic colitis (9), and the hemolytic-uremic syndrome (10). SLT production has also been detected in strains of *Vibrio cholerae* and *Vibrio parahemolyticus* (11), but what role this toxin plays in infections due to these bacteria is not yet clear.

Production of SLT by *E. coli* has been shown to be a transferable property associated with lysogeny by certain bacteriophages that contain the structural genes for SLT (12, 13). Newland *et al.* (13) have recently established that the SLT structural genes are located on a 2.9-kilobase (kb)-pair *Nco* I–*Eco*RV restriction fragment within a larger 8.5-kb *Eco*RI fragment of bacteriophage DNA; a single *Hin*dIII restriction site within the 2.9-kb fragment is located within the coding sequence of the A subunit of SLT. Similar data have been presented by two other groups for SLT-converting phage H19B (14, 15).

Recently, Seidah *et al.* (16) have determined the complete amino acid sequence of the B subunit of Shiga toxin from *S. dysenteriae* 1. To compare the B subunit sequence of SLT to that of Shiga toxin, facilitate the construction of hybridization probes that might detect the SLT genes of *V. cholerae*, and explore the mechanism of iron regulation of SLT expression, we have determined the nucleotide sequence of the SLT genes from phage H19B of *E. coli*.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. *E. coli* K-12 strains C600 (*F*$^-$λ$^-$ *thi-1 thr-1 leuB6 lacY1 tonA21 supE44*) and C600(H19B) were provided by H. W. Smith, Houghton Poultry Research Station, Huntingdon, Cambridgeshire U.K. (12); strain C600(H19B) is lysogenized by the SLT-converting phage H19B from one of the original strains of *E. coli* shown to produce Vero cell cytotoxin by Konowalchuk *et al.* (1). *E. coli* strain SY327 [*F*$^-$ *araD* Δ(*lac-pro*) *argEam rif nalA recA56*] was used as the host for transformation of recombinant plasmids (17). *E. coli* strain JF626 [*thi* Δ(*lac-pro*) *rpsL supE endA sbcB15 hsdR4/F'traD36 proA*$^+$*B*$^+$ *lacI*^{q} *lacZ*Δ *M15*] was used as the host for propagation of bacteriophage M13 (18). Plasmid pBR327, a deletion derivative of pBR322, was used as the vector in cloning experiments (19).

Molecular Biological Techniques. Induction of strain C600 (H19B) by ultraviolet irradiation and preparation of a bacteriophage lysate was done as described (20) except that modified Luria–Bertani broth (21) was used as the growth medium. Standard recombinant DNA techniques were done as described by Maniatis *et al.* (22). Rapid isolation of plasmid DNA was done as described by Birnboim (23), and plasmid DNA was purified by cesium chloride-equilibrium density gradient centrifugation (22).

Southern Hybridization. We derived a mixed oligonucleotide probe for the gene of the B subunit of SLT from the peptide sequence of the Shiga toxin B subunit (16) between residues 14 and 19. This mixed oligonucleotide [composition 5′ GT(AG)TC(AG)TC(AG)TC(AG)TT(AG)TA 3′] was synthesized on an Applied Biosystems synthesizer model 380A (Applied Biosystems, Foster City, CA) and purified by gel electrophoresis at the DNA Synthesis Facility (University of Massachusetts Medical School, Worcester, MA). Dephosphorylated 5′ ends were labeled with [γ-^{32}P]ATP (New England Nuclear) using T4 polynucleotide kinase (International Biotechnologies, New Haven, CT) and separated from unincorporated nucleotide on Sephadex G-50 columns (Sigma) as described (22).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: SLT, Shiga-like toxin.

‡Present address: Division of Infectious Disease, Ottawa Civic Hospital, Ottawa, ON, Canada.
§To whom reprint requests should be addressed.

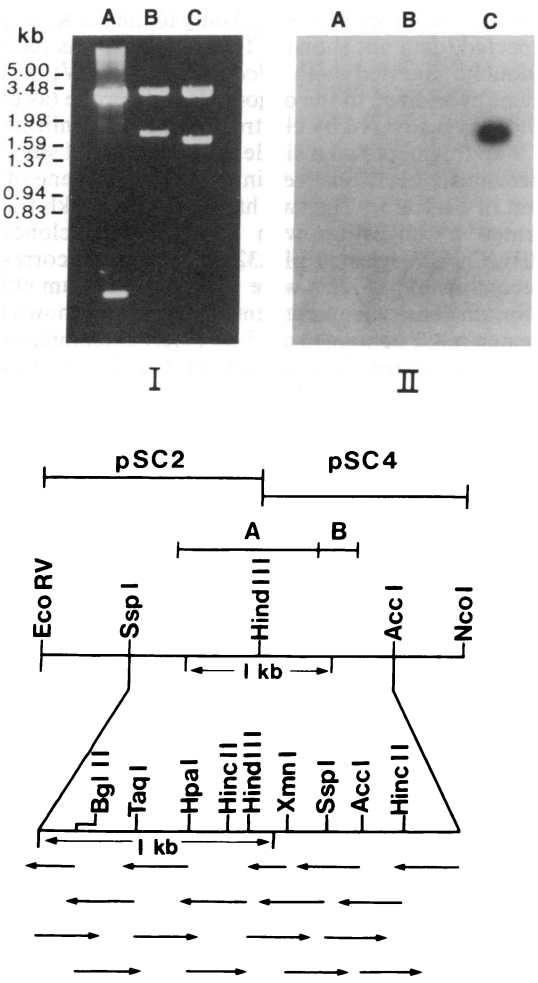

FIG. 1. (*Upper*) Plasmid DNA samples of 1 μg were digested with *Hin*dIII and *Bam*HI, and fragments were separated by electrophoresis through a 1% agarose gel. The gel was stained in ethidium bromide (0.5 μg/ml) and photographed under ultraviolet light (panel I). DNA from the gels was transferred to GenescreenPlus membranes and hybridized to the radioactively labeled SLT B subunit-specific probe (panel II). The positions of the M_r standards are indicated. Lane A, pBR327 (vector plasmid); lane B, pSC2; and lane C, pSC4. The lower bands in lanes B and C are larger than the corresponding inserted fragments because *Bam*HI cuts in the vector plasmid, outside of the *Eco*RV site used for subcloning. Only the insert of pSC4 hybridizes to the synthetic oligonucleotide probe. (*Lower*) The inserts of pSC2 and pSC4 relative to the 2.9-kb *Eco*RV–*Nco* I fragment of phage H19B DNA and the previously reported positions of the A and B subunits of SLT are diagramed. A detailed restriction map of the area between the left-most *Ssp* I and right-most *Acc* I sites is indicated; the positions of subclones and the direction of sequencing are indicated by the arrows at the bottom of the figure.

DNA samples of ≈1 μg were digested with the appropriate restriction endonucleases, and fragments were separated by electrophoresis through a 1% agarose gel. Transfer of DNA to a GeneScreen*Plus* membrane (New England Nuclear) and hybridization of labeled probe were done according to a modification of the method of Southern (24) by Chomczynski and Qasba (25). Probe DNA was added at a concentration of 10 ng/ml (1–5 × 10^5 dpm/ml). After hybridization at 37°C for 24 hr, the membrane was washed at 37°C with 2× standard saline citrate (SSC) (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7) until no further counts eluted off. The membrane was dried at room temperature and exposed to Kodak XRP-1 film (Eastman Kodak) for 4–24 hr.

DNA Sequencing. Fragments of DNA were subcloned into phage M13mp18 and M13mp19 as described by Messing (26), and DNA sequence was determined by the dideoxy chain

```
        10          20          30          40          50          60          70
AATATTGTGT ATTTTTTGTA TTGCAGGATG ACCCTGTAAC GAAGTTTGCG TAACAGCATT TTGCTCTACG

        80          90         100         110         120         130         140
AGTTTGCCAG CCTCCCCCAG TGGCTGGCTT TTTTATGTCC GTAACATCCT GTGTATCAAT AAATGTTGTT

       150         160         170         180         190         200         210
ATCTACGTAC GTCAAGTAGT CGCATGAGAT CTGACCAGAT ATGTTAAGGT TGCAGCTCTC TTTGAATATG
                                                    -35
       220         230         240         250         260         270         280
ATTATCATTT TCATTACGTT ATTGTTACGT TTATCCGGTG CGCCGTAAAA CGCCGTCCTT CAGGGCGTGG
 -10
       290         300         310         320         330
AGGATGTCAA GAATATAGTT ATCGTATGGT GCTCAAGGAG TATTGTGTAA T ATG AAA ATA ATT
                                        SD                  MET Lys Ile Ile  -19

346                 361                 376                 391
ATT TTT AGA GTG CTA ACT TTT TTC TTT GTT ATC TTT TCA GTT AAT GTG GTG GCG
Ile Phe Arg Val Leu Thr Phe Phe Phe Val Ile Phe Ser Val Asn Val Val Ala  - 1

        406                 421                 436                 451
AAG GAA TTT ACC TTA GAC TTC TCG ACT GCA AAG ACG TAT GTA GAT TCG CTG AAT
Lys Glu Phe Thr Leu Asp Phe Ser Thr Ala Lys Thr Tyr Val Asp Ser Leu Asn   18

                466                 481                 496
GTC ATT CGC TCT GCA ATA GGT ACT CCA TTA CAG ACT ATT TCA TCA GGA GGT ACG
Val Ile Arg Ser Ala Ile Gly Thr Pro Leu Gln Thr Ile Ser Ser Gly Gly Thr   36

    511                 526                 541                 556
TCT TTA CTG ATG ATT GAT AGT GGC TCA GGG GAT AAT TTG TTT GCA GTT GAT GTC
Ser Leu Leu MET Ile Asp Ser Gly Ser Gly Asp Asn Leu Phe Ala Val Asp Val   54

            571                 586                 601
AGA GGG ATA GAT CCA GAG GAA GGG CGG TTT AAT AAT CTA CGG CTT ATT GTT GAA
Arg Gly Ile Asp Pro Glu Glu Gly Arg Phe Asn Asn Leu Arg Leu Ile Val Glu   72

616                 631                 646                 661
CGA AAT AAT TTA TAT GTG ACA GGA TTT GTT AAC AGG ACA AAT AAT GTT TTT TAT
Arg Asn Asn Leu Tyr Val Thr Gly Phe Val Asn Arg Thr Asn Asn Val Phe Tyr   90

        676                 691                 706                 721
CGC TTT GCT GAT TTT TCA CAT GTT ACC TTT CCA GGT ACA ACA GCG GTT ACA TTG
Arg Phe Ala Asp Phe Ser His Val Thr Phe Pro Gly Thr Thr Ala Val Thr Leu  108

                736                 751                 766
TCT GGT GAC AGT AGC TAT ACC ACG TTA CAG CGT GTT GCA GGG ATC AGT CGT ACG
Ser Gly Asp Ser Ser Tyr Thr Thr Leu Gln Arg Val Ala Gly Ile Ser Arg Thr  126

    781                 796                 811                 826
GGG ATG CAG ATA AAT CGC CAT TCG TTG ACT ACT TCT TAT CTG GAT TTA ATG TCG
Gly MET Gln Ile Asn Arg His Ser Leu Thr Thr Ser Tyr Leu Asp Leu MET Ser  144

            841                 856                 871
CAT AGT GGA ACC TCA CTG ACG CAG TCT GTG GCA AGA GCG ATG TTA CGG TTT GTT
His Ser Gly Thr Ser Leu Thr Gln Ser Val Ala Arg Ala MET Leu Arg Phe Val  162

886                 901                 916                 931
ACT GTG ACA GCT GAA GCT TTA CGT TTT CGG CAA ATA CAG AGG GGA TTT CGT ACA
Thr Val Thr Ala Glu Ala Leu Arg Phe Arg Gln Ile Gln Arg Gly Phe Arg Thr  180

        946                 961                 976                 991
ACA CTG GAT GAT CTC AGT GGG CGT TCT TAT GTA ATG ACT GCT GAA GAT GTT GAT
Thr Leu Asp Asp Leu Ser Gly Arg Ser Tyr Val MET Thr Ala Glu Asp Val Asp  198

                1006                1021                1036
CTT ACA TTG AAC TGG GGA AGG TTG AGT AGC GTC CTG CCT GAC TAT CAT GGA CAA
Leu Thr Leu Asn Trp Gly Arg Leu Ser Ser Val Leu Pro Asp Tyr His Gly Gln  216

    1051                1066                1081                1096
GAC TCT GTT CGT GTA GGA AGA ATT TCT TTT GGA AGC ATT AAT GCA ATT CTG GGA
Asp Ser Val Arg Val Gly Arg Ile Ser Phe Gly Ser Ile Asn Ala Ile Leu Gly  234

            1111                1126                1141
AGC GTG GCA TTA ATA CTG AAT TGT CAT CAT CAT GCA TCG CGA GTT GCC AGA ATG
Ser Val Ala Leu Ile Leu Asn Cys His His His Ala Ser Arg Val Ala Arg MET  252

1156                1171                1186                1201
GCA TCT GAT GAG TTT CCT TCT ATG TGT CCG GCA GAT GGA AGA GTC CGT GGG ATT
Ala Ser Asp Glu Phe Pro Ser MET Cys Pro Ala Asp Gly Arg Val Arg Gly Ile  270

        1216                1231                1246                1261
ACG CAC AAT AAA ATA TTG TGG GAT TCA TCC ACT CTG GGG GCA ATT CTG ATG CGC
Thr His Asn Lys Ile Leu Trp Asp Ser Ser Thr Leu Gly Ala Ile Leu MET Arg  288

                1276        1286                1303
AGA ACT ATT AGC AGT TGAGGGGGTA AA ATG AAA AAA ACA TTA TTA ATA GCT GCA
Arg Thr Ile Ser Ser  SD          MET Lys Lys Thr Leu Leu Ile Ala Ala     -12

1318                1333                1348                1363
TCG CTT TCA TTT TTT TCA GCA AGT GCG CTG GCG ACG CCT GAT TGT GTA ACT GGA
Ser Leu Ser Phe Phe Ser Ala Ser Ala Leu Ala Thr Pro Asp Cys Val Thr Gly    7

        1378                1393                1408                1423
AAG GTG GAG TAT ACA AAA TAT AAT GAT GAC GAT ACC TTT ACA GTT AAA GTG GGT
Lys Val Glu Tyr Thr Lys Tyr Asn Asp Asp Asp Thr Phe Thr Val Lys Val Gly   25

                1438                1453                1468
GAT AAA GAA TTA TTT ACC AAC AGA TGG AAT CTT CAG TCT CTT CTT CTC AGT GCG
Asp Lys Glu Leu Phe Thr Asn Arg Trp Asn Leu Gln Ser Leu Leu Leu Ser Ala   43

    1483                1498                1513                1528
CAA ATT ACG GGG ATG ACT GTA ACC ATT AAA ACT AAT GCC TGT CAT AAT GGA GGG
Gln Ile Thr Gly MET Thr Val Thr Ile Lys Thr Asn Ala Cys His Asn Gly Gly   61

            1543                1565        1575        1585
GGA TTC AGC GAA GTT ATT TTT CGT TGACTCAGAA TAGCTCAGTG AAAATAGCAG GCGGAG
Gly Phe Ser Glu Val Ile Phe Arg
```

FIG. 2. Nucleotide sequence of the *slt* operon from *E. coli* phage H19B. The −35 and −10 regions of the proposed promoter are indicated. Shine–Dalgarno sequences (SD) are located just upstream of the start of the coding regions for SltA (nucleotide 332) and SltB (nucleotide 1289). The proposed signal peptide cleavage sites are indicated by vertical arrows. Numbers to the right of each row refer to amino acid residues in each polypeptide; negative numbers designate residues in the signal peptide, and positive numbers designate those in the mature protein. Horizontal arrows indicate a 21-bp interrupted dyad repeat in the region of the proposed −10 box.

Table 1. Comparison of previously *observed* amino acid compositions of Shiga toxin A and B subunits (4) with those *deduced* here from the nucleotide sequences of Shiga-like toxin A and B subunits

Amino acid	Amino acid composition			
	Subunit A		Subunit B	
	Shiga toxin	SLT	Shiga toxin	SLT
Asp/Asn	33	32	10	10
Thr	28	28	6	10
Ser	30	34	2	3
Glu/Gln	16	14	5	5
Pro	7	6	1	1
Gly	25	23	6	6
Ala	21	18	2	2
Val	22	22	5	6
Met	7	8	1	1
Ile	16	17	3	3
Leu	29	29	5	5
Tyr	7	7	2	2
Phe	16	14	4	4
His	8	8	1	1
Lys	4	3	5	5
Arg	24	26	2	2

termination method of Sanger *et al.* (27). The sequence of both strands of DNA was determined, and each base was confirmed a minimum of four separate times.

Comparison of Protein Sequences with a Computerized Data Base. The deduced peptide sequences of the SLT subunits were analyzed for homology to the protein sequences in the National Biomedical Research Foundation (NBRF) data base,¶ using the IFIND program from Intelligenetics (Intellicorp, Palo Alto, CA), based on the homology search method of Wilbur and Lipman (28). Initial search parameters were as follows: window length of 40; word length of 1, density set to less, fast set to no, and a gap penalty of 1; several additional combinations of search parameters were used to verify the initial results. A comparison was also made between the protein sequences of the SLT subunits and the NBRF data base using the FASTP program of Lipman and Pearson (29); this algorithm scores protein similarities both by identical residues and by amino acid substitutions conserved in evolution. We evaluated the significance of protein similarity scores from the FASTP search using the RDF program of Lipman and Pearson (29).

RESULTS AND DISCUSSION

Nucleotide and Derived Amino Acid Sequence of SLT. Digestion of phage H19B DNA with *Eco*RI revealed a pattern of fragments identical to earlier reports (13, 15); the SLT B subunit-specific probe hybridized only to the 8.5-kb fragment as expected (data not shown). This fragment was recovered and doubly digested with *Nco* I and *Eco*RV; a 2.9-kb fragment hybridized to the oligonucleotide probe (as expected) and was recovered by electroelution. We confirmed that the 2.9-kb fragment had a single internal *Hin*dIII site, which has previously been localized in the structural gene of the A subunit of SLT (13). The two halves of the 2.9-kb fragment generated by digestion with *Hin*dIII were cloned into *Hin*dIII/*Eco*RV-digested pBR327, and the two corresponding recombinant plasmids were purified by cesium chloride-equilibrium density gradient centrifugation. As shown in Fig. 1, plasmid pSC2 contained a 1.5-kb insert extending from the *Eco*RV end to the *Hin*dIII site of the 2.9-kb fragment, whereas pSC4 contained the adjoining 1.4-kb fragment (*Hin*dIII to *Nco* I). The synthetic oligonucleotide probe hybridized specifically to pSC4 (Fig. 1), as expected from previous studies on the localization of the B subunit of SLT (13, 15).

A detailed restriction map of a portion of the 2.9-kb fragment encoding the structural genes for SLT, as well as the position of overlapping subclones used for DNA sequencing, is shown in Fig. 1. The nucleotide sequence from the left-most *Ssp* I site to just beyond the right-most *Hin*cII site is shown in Fig. 2.

There are two long open reading frames beginning at nucleotides 332 and 1289, respectively, and corresponding to the expected positions of the genes encoding the A and B subunits of SLT. These open reading frames code for polypeptides of M_r 34,804 and 9744 respectively. Both polypeptides have hydrophobic N-terminal regions [identified by the method of Kyte and Doolittle (30)] that conform to bacterial signal sequences (31). Predicted cleavage sites for the signal peptides using the −1, −3 rule of Von Heijne (32) are indicated in Fig. 2. The deduced amino acid compositions of the mature proteins are compared in Table 1 with the previously determined amino acid compositions of the A and B subunits of Shiga toxin from *S. dysenteriae* (4). The excellent correspondence confirms the fidelity of the reading frame and reinforces the close structural similarity between Shiga toxin and the SLT of *E. coli*. In fact, the amino acid sequence determined experimentally for the purified B subunit of Shiga toxin (16) is identical to that of the processed second polypeptide in Fig. 2. The open reading frames in Fig. 2, therefore, correspond to the genetic loci *sltA* and *sltB* coding for the A and B subunits of SLT. No transcription termination signal was identified downstream of *sltB*, and an additional open reading frame may be present (S.B.C. and J.J.M., unpublished observations).

The predicted mature A subunit of SLT has 293 amino acids and a M_r of 32,217, whereas the mature B subunit has 69 amino acids and a M_r of 7692; these are in excellent agreement with previous estimates of the M_r values of SltA and SltB proteins (13). Mild tryptic digestion has been shown to nick the A subunit into A1 and A2 fragments with M_r values of ≈27,000 and 4000, respectively, and joined by a disulfide bond (3). The two cysteine residues at positions 242 and 261

¶Protein Identification Resource (1986) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 10.0.

```
SLT A    (138)  SYLDLMSHSGTSLTQSVARAMLRFVTVTAEALRFRQI  (174)
                ● ●   ● *●● ●* **●●* *  * * * *●● ●●  ●
RICIN A  (149)  SALYYYSTGGTQLP-TLARSFIICIQMISEAARFQYI  (184)

SLT A    (175)  QRGFRTTLDDLSGRSYVMTAEDVDLTLNWGRLSSVL   (210)
                *  ●● *     ●●  **  * ●  ●●●●●* *
RICIN A  (185)  EGEMRTRIRYN-RRSAPDPS-VITLENSWGRLSTAI  (218)
```

FIG. 3. One potential alignment of SltA (SLT A) (residues 138 to 210) with the ricin A subunit (residues 149 to 218), generated by the ALIGN program from IntelliGenetics used as follows: window length of 20, word length of 1, density set to less, and gap penalty of 3. Identical residues are indicated by closed circles and chemically similar residues by asterisks; dashes indicate gaps introduced into the sequence of ricin A. Many additional residues can be aligned with the introduction of further small gaps.

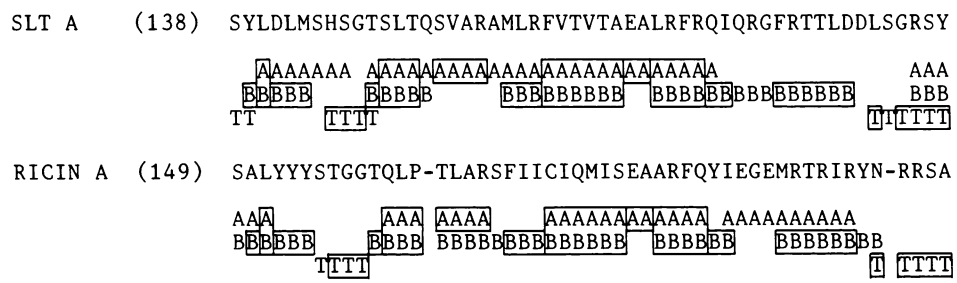

FIG. 4. Predictions of secondary structure by the Chou–Fasman algorithm for SltA (SLT A; residues 138 to 190) and ricin A (residues 149 to 199) aligned as in Fig. 3. A, alpha helix; B, beta sheet; and T, turn. Identical predictions for the corresponding residues are enclosed in boxes.

of SltA span two potential sites of tryptic digestion that could generate A1 and A2 fragments of the observed sizes.

The coding regions of both *sltA* and *sltB* are preceded by sequences homologous to the ribosomal binding sites proposed by Shine and Dalgarno (33). The very short intergenic space between the two coding sequences [12 base pairs (bp)] suggests that they are transcribed *in vivo* as an operon. A region homologous to the consensus sequence for *E. coli* promoters (34) is found upstream of *sltA*; in the vicinity of this proposed promoter, there is a 21-bp region of dyad symmetry (Fig. 2). Most bacterial operator sequences show such 2-fold symmetry (35), and we have evidence that this region of DNA may be involved in iron regulation of the *slt* operon by the *fur* locus of *E. coli* (S.B.C. and J.J.M., unpublished data).

Analysis of Homology Between the SLT Subunits and the NBRF Data Base. The peptide sequences of SltA and SltB were analyzed for homology to the protein sequences in the NBRF data base, using the IFIND program from IntelliGenetics. There was no significant homology of the SLT subunits to diphtheria toxin, *Pseudomonas* exotoxin A, cholera toxin A or B subunits, *E. coli* heat-labile enterotoxin A or B subunits, or *E. coli* heat-stable enterotoxin types 1 or 2. However, the highest score for homology (7.9 standard deviations above the mean) occurred between SltA and the A subunit of ricin, a plant toxin from *Ricinus communis* (castor bean). The ricin A subunit is a polypeptide of 267 amino acids and M_r 29,876, quite similar to SltA (36). Like Shiga toxin, ricin inhibits protein synthesis in eukaryotic cells by catalytic inactivation of the 60S ribosomal subunit. Ricin is also composed of A and B subunits—the A subunit being responsible for catalytic activity and the B subunit being responsible for cell binding. When the FASTP program was used to compare the sequence of SltA with the NBRF data base, comparison with ricin A again gave the highest score for optimized alignment (9.3 standard deviations above the mean of randomly permuted sequences). There was no significant homology between SltB and the B subunit of ricin or other proteins in the data base.

One region of homology between SltA and ricin A subunits is shown in Fig. 3; in this stretch of 73 amino acids, the two proteins have identical residues in 32% and either identical or chemically similar residues in 53% of the positions. Furthermore, the Chou–Fasman algorithm (37) yielded virtually identical predictions of secondary structure in the corresponding regions of these two proteins (Fig. 4). If additional gaps are introduced into the amino acid sequences (by lowering the gap penalty to 1), the entire length of ricin A can be aligned with SltA to give identical residues in 29% of the positions.

The finding of homology between SltA and ricin A is particularly intriguing because it crosses the boundary between prokaryotic and eukaryotic organisms. Several observations suggest this homology is significant: (*i*) Shiga toxin and ricin share similar mechanisms of toxic activity on eukaryotic cells; (*ii*) the similarity score for the comparison between SltA and ricin A was the highest observed in the entire data base, using a variety of search parameters and techniques; this similarity score was well within the range generally considered to have biological significance (29); and (*iii*) the predictions of secondary structure by the algorithm of Chou–Fasman were virtually identical in the area of highest homology between the two peptides. The possibility that prokaryotic toxins are evolutionarily related to eukaryotic enzymes has been suggested before in relation to diphtheria toxin (38, 39) and cholera toxin (40). We now present evidence for this hypothesis by demonstrating amino acid sequence homology between the A subunits of a prokaryotic (SLT) and a eukaryotic toxin (ricin).

This homology may represent conservation of critical amino acid residues forming similar active sites in the A chains of the two toxins. Convergent as well as divergent evolution could explain the degree of homology exhibited, but we favor the hypothesis that both toxin A chains evolved from a common ancestor. This ancestral protein may have had either a prokaryotic or a eukaryotic origin. In the latter case, the precursor may have served an autoregulatory function in the eukaryotic organism, perhaps during development. Molecules highly homologous to the ricin A chain (but lacking a B chain) have been identified in several plant species and have been proposed to serve such a regulatory function (41). Plants such as *R. communis* may have expanded the function of these regulatory molecules to include toxicity for heterologous organisms. Acquisition of the gene for such a regulatory molecule by a prokaryotic organism may have been the first step in the evolution of certain bacterial toxins by mechanisms similar to those used by plant cells in evolving toxic lectins (i.e., the molecular association of an A chain with a target cell-binding B chain).

The authors thank R. K. Taylor, M. J. Betley, and S. F. Carroll for many helpful discussions. This work was supported by Grants AI-18045 to J.J.M., AI-20325 to A.D.-R., and AI-16242 to G.T.K. from the National Institute of Allergy and Infectious Diseases and Grant FRA-302 to J.J.M. from the American Cancer Society. S.B.C. is recipient of an Upjohn Scholar Award and F.A. of a Fellowship Award from the Medical Research Council of Canada.

1. Konowalchuk, J., Speirs, J. I. & Stavric, S. (1977) *Infect. Immun.* **18,** 775–779.
2. O'Brien, A. D., LaVeck, G. D., Thompson, M. R. & Formal, S. B. (1982) *J. Infect. Dis.* **146,** 763–769.
3. O'Brien, A. D. & LaVeck, G. D. (1983) *Infect. Immun.* **40,** 675–683.
4. Donohue-Rolfe, A., Keusch, G. T., Edson, C., Thorley-Lawson, D. & Jacewicz, M. (1984) *J. Exp. Med.* **160,** 1767–1781.
5. O'Brien, A. D., Lively, T. A., Chen, M. E., Rothman, S. W. & Formal, S. B. (1983) *Lancet* **i,** 702.
6. Reisbig, R., Olsnes, S. & Eiklid, K. (1981) *J. Biol. Chem.* **256,** 8739–8744.
7. Jacewicz, M., Clausen, H., Nudelman, E., Donohue-Rolfe, A. & Keusch, G. T. (1986) *J. Exp. Med.* **163,** 1391–1404.
8. Cleary, T. G., Mathewson, J. J., Faris, E. & Pickering, L. K. (1985) *Infect. Immun.* **47,** 335–337.

9. Johnson, W. M., Lior, H. & Bezanson, G. S. (1983) *Lancet* **i,** 76.
10. Karmali, M. A., Petric, M., Lim, C., Fleming, P. C., Arbus, G. S. & Lior, H. (1985) *J. Infect. Dis.* **151,** 775–782.
11. O'Brien, A. D., Chen, M. E., Holmes, R. K., Kaper, J. & Levine, M. M. (1984) *Lancet* **i,** 77–78.
12. Smith, H. W., Green, P. & Parsell, Z. (1983) *J. Gen. Microbiol.* **129,** 3121–3137.
13. Newland, J. W., Strockbine, N. A., Miller, S. F., O'Brien, A. D. & Holmes, R. K. (1985) *Science* **230,** 179–181.
14. Willshaw, G. A., Smith, H. R., Scotland, S. M. & Rowe, B. (1985) *J. Gen. Microbiol.* **131,** 3047–3053.
15. Huang, A., DeGrandis, S., Friesen, J., Karmali, M., Petric, M., Congi, R. & Brunton, J. L. (1986) *J. Bacteriol.* **166,** 375–379.
16. Seidah, N. G., Donohue-Rolfe, A., Lazure, C., Auclair, F., Keusch, G. T. & Chrétien, M. (1986) *J. Biol. Chem.* **261,** 13928–13931.
17. Goldberg, I. & Mekalanos, J. J. (1986) *J. Bacteriol.* **165,** 715–722.
18. Miller, V. L., Taylor, R. K. & Mekalanos, J. J. (1987) *Cell* **48,** 271–279.
19. Covarrubias, L., Cervantes, L., Covarrubias, A., Soberon, X., Vichido, I., Blanco, A., Kupersztoch-Portnoy, Y. M. & Bolivar, F. (1981) *Gene* **13,** 25–35.
20. Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984) *Experiments With Gene Fusions* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
21. O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W. & Formal, S. B. (1984) *Science* **226,** 694–696.
22. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
23. Birnboim, H. C. (1983) *Methods Enzymol.* **100,** 243–255.
24. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
25. Chomczynski, P. & Qasba, P. (1984) *Biochem. Biophys. Res. Commun.* **122,** 340–344.
26. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
29. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.
30. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
31. von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105.
32. von Heijne, G. (1984) *J. Mol. Biol.* **173,** 243–251.
33. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.
34. Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acids Res.* **11,** 2237–2255.
35. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53,** 293–321.
36. Halling, K. C., Halling, A. C., Murray, E. E., Ladin, B. F., Houston, L. L. & Weaver, R. F. (1985) *Nucleic Acids Res.* **13,** 8019–8033.
37. Chou, P. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
38. Pappenheimer, A. M., Jr., & Gill, D. M. (1973) *Science* **182,** 353–358.
39. Pappenheimer, A. M., Jr. (1984) *J. Hyg.* **93,** 397–404.
40. Moss, J. & Vaughan, M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3621–3624.
41. Jimenez, A. & Vazquez, D. (1985) *Annu. Rev. Microbiol.* **39,** 649–672.