Proc. Natl. Acad. Sci. USA
Vol. 78, No. 1, pp. 50–54, January 1981
Biochemistry

Sequence homologies between A subunits of *Escherichia coli* and *Vibrio cholerae* enterotoxins

(DNA sequence/NH_2- and COOH-terminal amino acid sequences/signal peptide)

ELEANOR K. SPICER*, W. MICHAEL KAVANAUGH*, WALTER S. DALLAS†, STANLEY FALKOW‡, WILLIAM H. KONIGSBERG*, AND DAVID E. SCHAFER§¶

*Department of Molecular Biophysics & Biochemistry and §Department of Physiology, Yale University School of Medicine, New Haven, Connecticut 06510; †Cetus Corporation, 600 Bancroft Way, Berkeley, California 94710; ‡Department of Microbiology and Immunology SC42, University of Washington, Seattle, Washington 98195; and §Veterans Administration Medical Center, West Haven, Connecticut 06516

Communicated by Julian M. Sturtevant, August 29, 1980

ABSTRACT **The genes coding for the heat-labile enterotoxin LT produced by *Escherichia coli* have been cloned into the plasmid pBR313. Using DNA derived from the resulting chimeric plasmid, we determined the nucleotide sequence of two regions of the gene coding for the enzymatically active A subunit of LT. Translation of the nucleotide sequence gives the primary structure of the NH_2-terminal and COOH-terminal regions of the LT A subunit. This permits direct comparison of the LT A subunit with the A subunit of cholera toxin. Our results show that the two toxins possess homologous sequences, of varying degrees, in both regions of their primary structure. The order of the component A_1 and A_2 polypeptides is A_1-A_2. The nucleotide sequence predicts the existence of a signal sequence of 18 amino acids at the NH_2-terminus of the A subunit.**

Certain strains of *Escherichia coli* produce enterotoxins that cause severe diarrheal diseases in man and domestic animals. Considerable evidence suggests that one of these enterotoxins, a heat-labile toxin termed LT, functions by a mechanism similar to that of cholera toxin (1), the enterotoxin of *Vibrio cholerae*, and that LT and cholera toxin (CT) are structurally related, having common antigenic determinants (2, 3) and similar subunit structures (4). Epidemiologically, however, the disease caused by LT-producing strains of *E. coli* is much more prevalent than that caused by *V. cholerae* and, typically, it is somewhat less severe in its effects. Efforts to evaluate the extent of structural similarity between the two toxins have been and continue to be hampered by difficulties in obtaining sufficiently pure LT in significant amounts. Fortunately, the DNA encoding LT is plasmid borne, making it possible to isolate the LT genes and to investigate the genes and their products directly (5). Using recombinant DNA technology, Dallas *et al.* (5) have isolated from the large transmissible plasmid P307 (M_r 60,000,000) a DNA fragment (M_r 1,200,000) coding for LT. They have cloned this fragment into the plasmid pBR313, forming the chimeric plasmid EWD299. By using an *E. coli* minicell system containing plasmid EWD299, they demonstrated that LT is composed of two subunits, of estimated M_rs 25,500 and 11,500 (4), called A and B, in analogy to subunits A (M_r 28,000) and B (M_r 11,600) of CT. It now appears that, in LT as in CT, a portion of the A subunit is enzymatically active, causing stimulation of adenylate cyclase in intestinal and other cells (6) and that the B subunits of both toxins function in toxin binding to cells.

The fact that the A subunits of both CT and LT exert their effects by promoting the ADP ribosylation of a M_r 42,000 membrane protein, which functions as a component of the adenylate cyclase system, suggests that they may have similar or perhaps identical active sites. The delineation of similar features of the primary and secondary structures of the two proteins is an obvious step toward understanding how both toxins function. We report here the nucleotide sequence of two regions of the gene coding for the A subunit of LT (LT A) and compare the amino acid sequence of these portions of LT A with the partial amino acid sequence currently known for the A subunit of cholera toxin (CT A). The comparison shows that these regions of the LT A and CT A subunits exhibit direct homology but in strikingly different degrees. Nucleotide sequence data at the distal end of the A cistron shows that the order of the component A_1 and A_2 polypeptides on the gene is A_1-A_2. The DNA sequence further indicates that LT A is synthesized as a precursor containing a signal sequence of 18 additional amino acids at its NH_2-terminus.

MATERIALS AND METHODS

Enzymes and Biochemicals. Restriction enzymes and polynucleotide kinase were purchased from New England Biolabs. *E. coli* DNA polymerase I used for nick translation was obtained from Boehringer Mannheim. Bacterial alkaline phosphatase was purchased from Worthington. [γ-^{32}P]ATP (3000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels) was purchased from New England Nuclear.

Plasmids. Plasmid EWD299 is a chimera of pBR313 and a M_r 1,200,000 fragment of ENT plasmid P307. Construction of EWD299 has been described in detail (5).

Restriction Enzyme Mapping. Restriction fragments were purified by electrophoresis through 5% polyacrylamide gels. DNA was recovered from the gel slices by electroelution through 5% acrylamide gel plugs, followed by precipitation with 0.1 vol of 3 M sodium acetate and 2.5 vol of ethanol. When the DNA was to be sequenced, the pellet was washed twice with 95% ethanol to reduce the sodium acetate concentration and redissolved in distilled water. Restriction site mapping was performed by the method of Smith and Birnstiel (7). There is one large region [400 base pairs (bp)] within the A cistron (between the *Hpa* II and *Taq* I sites; Fig. 1) in which no restriction sites have been found.

DNA Nucleotide Sequence Determination. The nick-translation and chain-terminating method described by Maat and Smith (8) was used for all sequence determinations. Small (100–300 bp) DNA restriction fragments, prepared as described above, were 5′ end labeled by treatment with bacterial alkaline phosphatase, [γ-^{32}P]ATP, and polynucleotide kinase. Labeled fragments were cut with a second restriction enzyme to produce

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.

Abbreviations: LT A, A subunit of heat-labile *E. coli* enterotoxin; CT A, A subunit of cholera enterotoxin; bp, base pair(s).

¶ To whom reprint requests should be addressed.

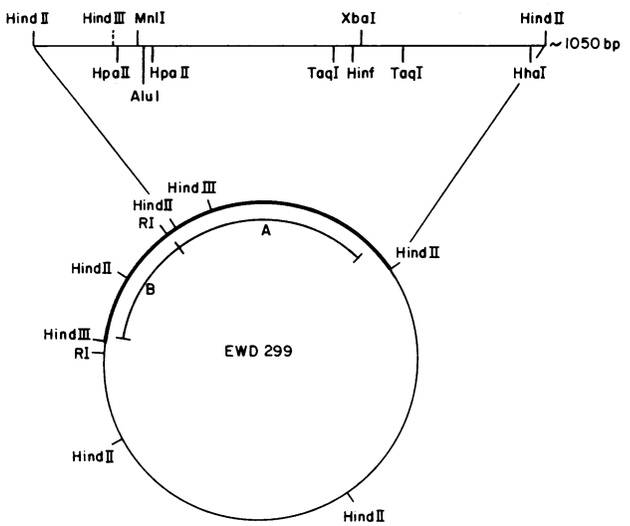

FIG. 1. Known restriction sites and LT gene locations on EWD299. Heavy line represents the DNA taken from P307 and inserted into pBR313. A and B indicate approximate locations of the structural cistrons for LT subunits A and B. The direction of transcription of the LT operon is from right to left (counterclockwise). The *Hin*dIII–*Hin*dIII fragment that contains all of the A structural gene is shown above in greater detail. Restriction sites within the *Hin*dII–*Hin*dII fragment were determined by the Smith–Birnstiel mapping technique. There is one relatively large region (≈400 bp) for which no restriction sites have been found.

singly end-labeled fragments that, after separation on polyacrylamide gels, were sequenced directly. Sequence reaction mixtures were separated by electrophoresis through 8% or 12% polyacrylamide/7M urea gels. After electrophoresis, the gels were dried on Whatman paper, and autoradiography was performed on Dupont Cronex film with a Dupont Lightening-Plus-Intensifying Screen at −70°C.

RESULTS AND DISCUSSION

To determine the primary structure of the LT A subunit, we used the plasmid EWD299 to purify sufficient quantities of the A cistron DNA to permit nucleotide sequence analysis. Fig. 1 is a map of restriction sites and LT cistron locations on EWD299 (5). Studies of *in vitro*-constructed deletion mutants have shown that the A and B subunits are transcribed from a single promoter lying at the right-hand end of the inserted LT DNA (5). The strategy for determining the sequence of the A cistron involved isolation of small restriction fragments (100–200 bp) labeled with ^{32}P at a single 5′ end. These fragments were then subjected to enzymatic sequence analysis by the method of Maat and Smith (8). The regions of the A subunit cistron for which nucleotide sequence data have been obtained are shown in Fig. 2. The horizontal lines below the DNA map indicate the sites of labeling and the direction of sequence obtained from these sites.

The sequence of approximately 35% of the 265 amino acids of CT A has been reported (9–12) (see Fig. 2). In contrast, little is known about the protein chemistry of LT A aside from its amino acid composition (13, 14). There is strong evidence that both CT A and LT A are cleaved by bacterial proteases (6) to produce two polypeptides, designated as A_1 and A_2. In the case of CT A, these two polypeptides have ≈193 and ≈70 amino acids, respectively (15); each contains a single half-cystine residue, and the two are held together by a disulfide bond. Reduction of this disulfide bond results in activation of A_1, giving rise to the suggestion that the half-cystine residue might be near the active site (11).

We have translated the available LT A cistron nucleotide sequence data in all three reading frames into the predicted amino acid sequences and compared them with the known CT A amino acid sequences. Near the COOH-terminal region of CT A_1, the sequence of a 16-amino acid peptide containing the single half-cystine residue is known (9). Translation of the nucleotide sequence in the putative COOH-terminal region of LT A_1 (in the smaller *Hin*dII–*Hin*dIII fragment shown in Fig. 2) predicts a half-cystine-containing peptide in which 7 of 11 neighboring residues are the same as those in CT A. This represents a 64% homology of residues in the immediate vicinity of the half-cystine residues and a 44% homology for the total sequence of 16 amino acids. The LT nucleotide sequence, the predicted LT amino acid sequence, and the known CT A_1 amino acid sequence of the homologous region are shown in Fig. 3.

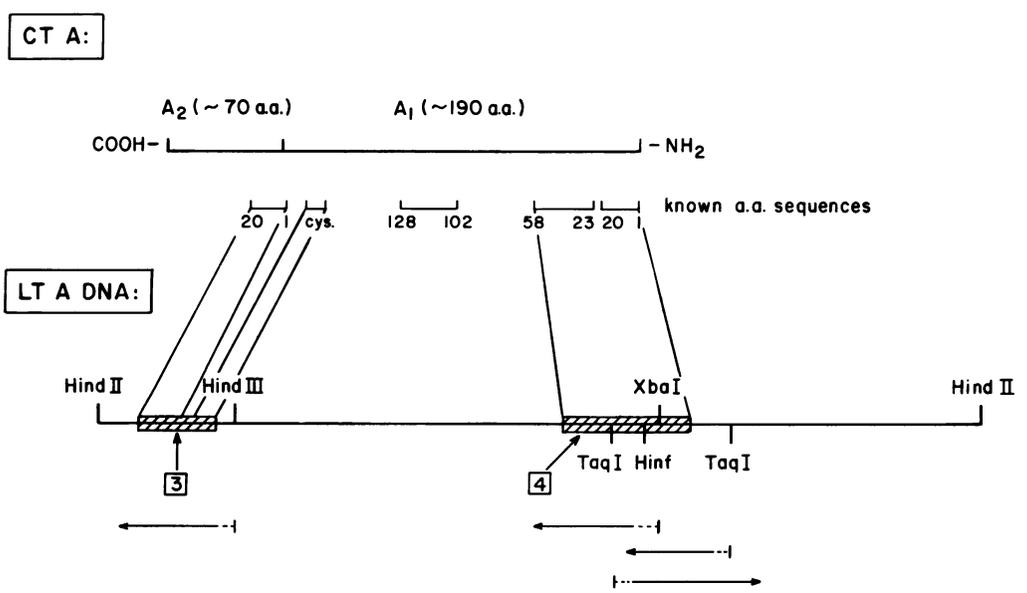

FIG. 2. Correlation of nucleotide sequences determined for the LT A gene (below) and regions of known amino acid sequence of CT A subunit (see Fig. 1). Connecting lines indicate regions of homology between the CT A subunit and the predicted LT A subunit. Horizontal lines below the DNA map indicate the direction and length of sequence determined. [3] and [4] indicate regions whose sequence is given in Figs. 3 and 4, respectively. a.a., amino acid.

```
LT DNA      5' - AGA ACC TGG ATT CAT CAT GCA CCA CAA GGT TGT GGA GAT TCA TCA
            3' - TCT TGG ACC TAA GTA GTA CGT GGT GTT CCA ACA CCT CTA AGT AGT

LT mRNA     5' - AGA ACC UGG AUU CAU CAU GCA CCA CAA GGU UGU GGA GAU UCA UCA

LT a.a.    NH2 - arg-thr-trp-ile-his-his-ala-pro-glu-gly-cys-gly-asp-ser-ser-

CT A1            glx-glx-pro-his-ile-his-ala-pro-gly-gly-cys-pro-asx-ala-pro-



LT DNA      AGA ACA TTC ACA GGT GAT ACT TGT AAT GAG GAG ACC CAG AAT CTG AGC
            TCT TGT AAG TGT CCA CTA TGA ACA TTA CTC CTC TGG GTC TTA GAC TCG

LT mRNA     AGA ACA UUC ACA GGU GAU ACU UGU AAU GAG GAG ACC CAG AAU CUG AGC

LT a.a.     arg-thr-phe-thr-gly-asp-thr-cys-asn-glu-glu-thr-gln-asn-leu-ser-

CT A1       arg-(*)    ⬆
CT A2 (a)              (met)ser-asn-thr-cys-asp-glu-lys-thr-gln-ser-leu-gly-
CT A2 (b)              (met)ser-asp-thr-   -asn-glu-lys-thr-glx-   -leu-gly-
CT A2 (c)              (met)ser-asn-thr-glu-cys-asp-lys-

                         N-terminus of CT A2



LT DNA      ACA ATA TAT CTC AGG GAA TAT CAA TCA AAA GTT ATA GAG GCA CAT ATT
            TGT TAT ATA GAG TCC CTT ATA GTT AGT TTT CAA TAT CTC CGT GTA TAA

LT mRNA     ACA AUA UAU CUC AGG GAA UAU CAA UCA AAA GUU AUA GAG GCA CAU AUU

LT a.a.     thr-ile-tyr-leu-arg-glu-tyr-glu-ser-lys-val-ile-glu-ala-his-ile-

CT A2 (a)   val-lys-phe-leu-asp-glu-tyr-
CT A2 (b)   val-lys-phe-leu-glx-glu-tyr-



LT DNA      TTC AGA CTA TCA GTC AGA GGT TGA
            AAG TCT GAT AGT CAG TCT CCA ACT

LT mRNA     UUC AGA CUA UCA GUC AGA GGU UGA

LT a.a.     phe-arg-leu-ser-val-arg-gly-COOH
```

FIG. 3. Partial homology of LT and CT amino acid sequences in the regions of the A_1 COOH-terminus and the A_2 NH_2-terminus, showing continuity of the two LT chains in this region. The LT amino acid sequence is predicted from the nucleotide sequence data. Homologous amino acids are underlined in the CT sequence. Undetermined amide assignments are assumed to match. Asterisk denotes a sequence of uncertain length at the COOH-terminus of CT A_1. The amino acid sequence of the cysteine-containing peptide of A_1 is from Mendez *et al.* (9). At the NH_2-terminus of A_2, two proposed sequences of 20 amino acid residues according to (a) Klapper *et al.* (10) and (b) Kurosky *et al.* (12) are shown, together with (c) the sequence of the first eight residues as proposed by Mendez *et al.* (9).

Thirty-six nucleotides downstream (5′→3′ in mRNA) from the codon for cysteine, in the same reading frame, a second cysteine residue is predicted. Comparison of the amino acid sequence in this region with that of the NH_2-terminal portion of CT polypeptide A_2 indicates that seven out of nine amino acids in LT are homologous with the first nine residues of CT A_2. As is shown in Fig. 3, of the total 20 amino acids proposed in CT A_2, 11 are identical in LT A_2 (55% homology). Thus there is considerable, but far from complete, homology between the half-cystine-containing regions of the A_1 and A_2 subunits of LT and CT. Furthermore, the nucleotide sequence data show that the A_2 polypeptide is synthesized at the COOH-terminal end of A_1, contrary to predictions made for CT (9, 11). The most likely sites for the proteolytic cleavage of A are the Arg and Phe residues (see Fig. 3). The M_r of CT A_2 has been estimated as 7500–9700 (9, 13), corresponding to 65–90 amino acid residues. By contrast, our data show that LT A_2 has only 36 residues. In particular, the COOH-terminal sequence, (Phe-Tyr-Val)-Lys-Leu-COOH, proposed (15) for CT A_2 is missing from LT A_2.

In the NH_2-terminal region of CT A_1, the sequences of amino acid residues 1–20 (10) and 23–58 (11) have been reported. Our nucleotide sequence data from the proximal portion of the A subunit structural cistron are shown in detail in Fig. 4. The nucleotide sequence in this region codes for amino acids that are strikingly homologous to those of the NH_2-terminal portion of CT A_1. Specifically, amino acid residues 5–17, 26–30, and 33–58 are completely homologous. Relative to the amino acid numbers assigned by Lai *et al.* (11), it appears that LT may lack 5 amino acids—those at positions 21–24 and 32. Until the final amino acid sequence of CT A_1 in the region of these residues is known, however, it is uncertain whether such deletions in LT actually exist. Altogether, at least 46 of 58 residues are the same in the NH_2-terminal region of the two toxins (≥79% homology).

The LT nucleotide sequence data show that the presumed NH_2-terminal Asn residue is not preceded by a Met residue, as would be required if translation of polypeptide A_1 begins at that site. Upstream from the Asn residue, in the same reading frame, there is a Met residue in position 18. The location of this Met relative to the NH_2-terminal Asn of CT A_1 suggests that, like some other procaryotic proteins (16, 17), LT A is synthesized in

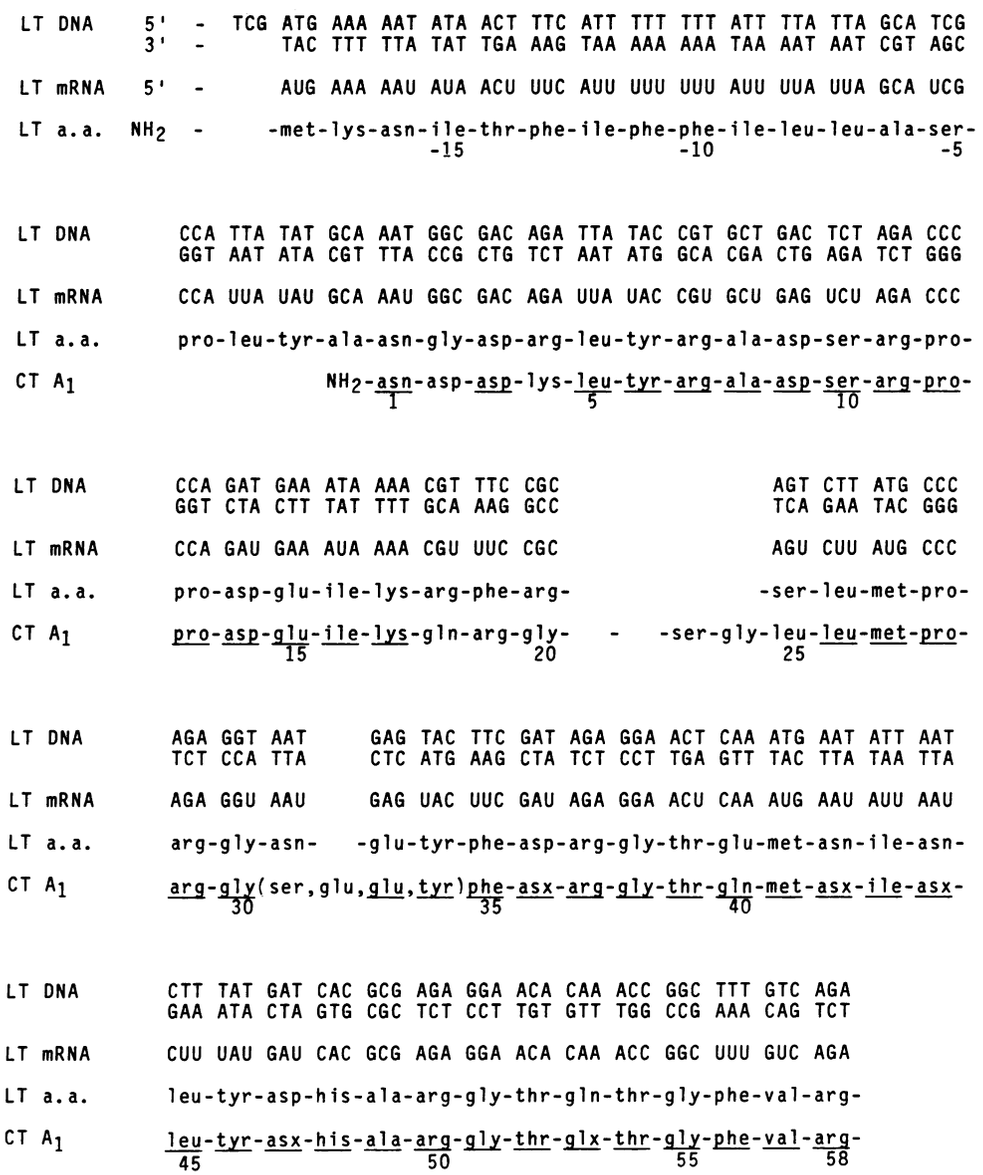

FIG. 4. NH_2-terminal region of the A_1 subunits of LT and CT, showing the proposed leader sequence of LT A_1 and extensive homologies between the two toxins. The numbering system used follows Lai *et al.* (11). The sequence of residues 1–20 is from Klapper *et al.* (10); residues 23–58 are from Lai *et al.* (11). There are no gaps in the LT DNA sequence in regions corresponding to amino acids 21–24 and 32. Homologous amino acids are underlined in the CT sequence. Amide assignments at 42, 44, 47, and 53 are assumed to match. Note that residues 29, 37, 50, and 58 are Arg (see text).

a precursor form containing a leader or signal sequence that is later removed by proteolysis. The role of this signal sequence presumably is to direct the A protein to the cell membrane or facilitate its passage into the periplasmic space (18, 19). Several features of the sequence between the Met and Asn residues make it highly probable that this is a signal sequence: First, its size (18 amino acids) conforms to the range of observed sizes (15–30 amino acids) for procaryotic signal sequences and, second, it complies with three of the four rules proposed in the loop model for the properties of signal peptides (19)—(*i*) Met is followed by Lys, making the NH_2-terminus basic; (*ii*) there is a long run of hydrophobic amino acids, Phe-(-13) through Ala-(-6); and (*iii*) there is an Ala residue proximal to the cleavage site. Other data (20) suggest that the A subunit of CT originally contained a signal sequence.

The fact that the amino acids in the NH_2-terminal portion of the A subunit are more highly conserved than those in the COOH-terminal region suggests that the NH_2-terminal residues may be more important for its function. The conserved region near the NH_2-terminus may play a role, for example, in the binding of the A subunit to the B subunits, in the transport of the A protein into the cell, or in the stimulation of adenylate cyclase. Lai *et al.* (‖) have reported that 35% of the original activity of CT A_1 is displayed by a tryptic fragment of M_r 12,500 that contains all the amino acids from Met-41 to Met-101, plus some proximal and distal to them. Moreover, one or both of the regions preceding Met-41 and following Met-101 are essential for activity. This active fragment overlaps the region of extensive homology (residues 26–58) shown in Fig. 4. It may be significant, therefore, that in Fig. 4, four of the 33 residues are arginine, occupying positions 29, 37, 50, and 58. Reagents that react specifically with the guanidino group of arginine eliminate the toxicity of CT for mammalian cells, but do not interfere with its ability to bind to ganglioside G_{M1} (21). Thus the question arises

‖ Lai, C.-Y., Cancedda, F. & Duffy, L. K. (1979) *Abstracts of the 11th International Congress of Biochemistry*, Toronto, Canada, p. 207, Abstr. 03-4S173.

whether one of these arginine residues—perhaps Arg 37—might be involved in the active site.

Knowledge of the complete amino acid sequence of LT and of CT will clarify which regions of the protein can tolerate amino acid changes and which cannot.

We are grateful to Dr. G. Nigel Godson for helpful discussions during the course of this work and to Ms. Janelle Noble for excellent technical assistance. This work was supported by National Institutes of Health Research Service Award AG05121-02 (to E.K.S.), National Institutes of Health Grants GM 12607 (to W.H.K.) and GM 24704 (to D.E.S.), and the Veterans Administration (D.E.S.).

1. Evans, D. J., Chin, L. C., Curlin, G. T. & Evans, D. G. (1972) *Nature (London) New Biol.* **236,** 137–138.
2. Gyles, C. L. (1974) *Infect. Immun.* **9,** 564–570.
3. Clements, J. D. & Finkelstein, R. A. (1978) *Infect. Immun.* **21,** 1036–1039.
4. Dallas, W. S. & Falkow, S. (1979) *Nature (London)* **227,** 406–407.
5. Dallas, W. S., Gill, D. M. & Falkow, S. (1979) *J. Bacteriol.* **139,** 850–858.
6. Gill, D. M. & Richardson, S. H. (1980) *J. Infect. Dis.* **141,** 64–70.
7. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398.
8. Maat, J. & Smith, A. J. H. (1978) *Nucleic Acids Res.* **5,** 4537–4545.
9. Mendez, E., Lai, C.-Y. & Wodnar-Filipowicz, A. (1975) *Biochem. Biophys. Res. Commun.* **67,** 1435–1443.
10. Klapper, D. G., Finkelstein, R. A. & Capra, J. D. (1976) *Immunochemistry* **13,** 605–611.
11. Lai, C.-Y., Cancedda, F. & Chang, D. (1979) *Fed. Eur. Biochem. Soc. Lett.* **100,** 85–89.
12. Kurosky, A., Markel, D. E., Touchstone, B. & Peterson, J. W. (1976) *J. Infect. Dis.* **133,** S14–S22.
13. Robertson, D. C., Kunkel, S. L. & Gilligan, P. H. (1980) *Proceedings of the Fifteenth Joint Conference on Cholera,* The U.S.-Japan Cooperative Medical Science Program (NIH, Bethesda, MD), pp. 389–400.
14. Clements, J. D., Yancey, R. J. & Finkelstein, R. A. (1980) *Proceedings of the Fifteenth Joint Conference on Cholera,* The U.S.-Japan Cooperative Medical Science Program (NIH, Bethesda, MD), pp. 27–43.
15. Markel, D. E., Hejtmancik, K. E., Peterson, J. W. & Kurosky, A. (1979) *J. Supramol. Struct.* **10,** 137–149.
16. Chang, C. N., Blobel, G. & Model, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 361–365.
17. Movva, N. R., Nakamura, K. & Inouye, M. (1980) *J. Biol. Chem.* **225,** 27–29.
18. Wickner, W. (1979) *Annu. Rev. Biochem.* **48,** 23–45.
19. Davis, B. D. & Tai, P.-C. (1980) *Nature (London)* **238,** 433–438.
20. Nichols, J. C., Tai, P.-C. & Murphy, J. R. (1980) *The Proceedings of the Fifteenth Joint Conference on Cholera,* U.S.-Japan Cooperative Medical Science Program (NIH, Bethesda, MD), pp. 320–333.
21. Lönnroth, I. & Holmgren, J. (1975) *J. Gen. Microbiol.* **91,** 263–277.