INFECTION AND IMMUNITY, Feb. 1989, p. 645–648
0019-9567/89/020645-04$02.00/0

Vol. 57, No. 2

Rectification of Two *Escherichia coli* Heat-Stable Enterotoxin Allele Sequences and Lack of Biological Effect of Changing the Carboxy-Terminal Tyrosine to Histidine

LUZ-MARIA GUZMAN-VERDUZCO AND YANKEL M. KUPERSZTOCH*

Department of Microbiology, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75235

Received 16 August 1988/Accepted 17 October 1988

Resequencing *estA3*, an allele of the methanol-soluble heat-stable enterotoxin of *Escherichia coli* showed that the proline triplet 19 is in fact an alanine codon; thus, *estA* alleles 3 and 4 were shown to be identical. Resequencing has also shown that the carboxy terminus of another allele, *estA2*, is not the previously inferred histidine triplet but the same tyrosine codon reported for all other *estA* alleles. The improperly inferred histidine codon was used in constructions to fuse *estA2* to the B subunit of the heat-labile enterotoxin gene, and the fused gene products as well as three amino acid insertional mutants containing histidine-72 were not efficiently secreted. We show that the defective secretion is not due to histidine as a carboxy-terminal residue, since site-directed mutagenesis of wild-type tyrosine-72 to histidine did not influence the localization of the activity of the methanol-soluble heat-stable enterotoxin.

The methanol-soluble heat-stable enterotoxins (ST_As) of *Escherichia coli* are a family of extracellular polypeptides that cause diarrhea in animals and humans. Four ST_A genes (*estA* genes) have been cloned and sequenced: *estA1* (11) from a bovine *E. coli* strain, *estA2* (M. DeWilde, M. Ysebeart, and M. Harford, *in* S. B. Levy, R. C. Clowes, and E. L. Koenig (ed.) *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids*, p. 596, 1981), and *estA3* (6) and *estA4* (12) from human Bangladee and human Mexican *E. coli* strains, respectively. The four *estA* genes display a 72-amino-acid open reading frame, and the active extracellular toxins are 18- and 19-amino-acid polypeptides (1, 12).

In an attempt to use the genetic information of *estA2* as an export carrier and to test if it can mobilize the periplasmic heat-labile enterotoxin B subunit (LT_B) (8) to the exterior of the bacterium, the DNA region corresponding to the signal peptide of *eltB*, including its regulatory elements, was substituted by the complete *estA4* devoid of transcription and translation terminators (5); the fused ST_{A2}-LT_B remained cell associated; similarly, ST_{A2} with three additional amino acids at the carboxy terminus ($ST_{His\text{-}72+3}$) was not efficiently secreted into the culture supernatant. These observations suggest that the carboxy-terminal region of the molecule is a determining factor in the extracellular delivery of ST_As.

The carboxy-terminal residues of all ST_As had been determined and/or inferred to be tyrosine except for the histidine inferred in *estA2* (DeWilde et al., 1981). Upon resequencing, we determined that like the other alleles, *estA2* carboxy terminates with a tyrosine codon. Thus, the previously constructed insertional mutant $ST_{His\text{-}72+3}$ and the fused ST_{A2}-LT_B had a His for Tyr mutation preceding the three amino acids and at the junction of the two moieties, respectively (5). To examine the effect of the substitution of His for Tyr-72, wild-type *estA2* was mutagenized, and we show that both ST_{A2} Tyr-72 and ST_{A2} His-72 are extracellular biologically active toxins. Therefore, the gene products of the fused *estA2-eltB* and $estA2_{His\text{-}72+3}$ are not secreted into the culture supernatant by a reason other than the His-72 for Tyr mutation common to both gene products. We also show after resequencing *estA3* that the inferred Pro-19 triplet is an Ala codon; consequently, *estA3* and *estA4* are identical genes.

The *E. coli* strains used were HB101 (5), TG1 (3), and CA8000 (2). The origin of the *estA* genes were plasmids pRIT10250 (*estA2*) (5) and pSLM004 (*estA3*) (6). *estA2* was cloned into pUC8 (5) from pRIT10250 as a 560-base-pair (bp) *Eco*RI fragment to yield pGK22. Plasmid pGK26 (5) was described earlier, and *estA2* in this plasmid has an in-phase addition of the triplets GGA GCT CTC TAG (Gly-Ala-Leu-amber) following the inferred 3′ terminal His codon (CAT). In strain CA8000(pGK26), the *estA* gene product (ST_A) is expected to have Gly-13, Ala-74, and Leu-75 following His-72, while in HB101(pGK26), the presence of the *supE44* gene should yield a fraction of the toxin molecules with the insertion of Gln-76 in place of the amber codon followed by 24 amino acids that precede the UAA termination codon (5). L and T media (5) were used where indicated and supplemented when necessary with ampicillin (100 μg/ml) and tetracycline (12.5 μg/ml).

The two *estA* alleles were sequenced as follows. *estA2* was excised from pGK22 with *Eco*RI and cloned into M13mp18 cut with the same enzyme. *estA3* was cut from pSLM004 with *Hpa*II and cloned into M13mp18 digested with *Acc*I; additionally, two *Eco*RI-*Taq*I fragments derived from *estA2* (254 and 295 bp) were cloned separately into M13mp18 and M13mp19 digested with both *Eco*RI and *Acc*I. The sequence of *estA2* was determined in both orientations of the two *Eco*RI-*Taq*I fragments, passing through each region at least two times, and was done by the dideoxy-chain termination method (9) by using the universal M13 sequencing primer or by using as a primer the 20-mer TTACAACA CAATTCACAGCA complementary to positions 323 to 342 (Fig. 1). Polymerization reactions were done by incubating with Sequenase (United States Biochemical Corp., Cleveland, Ohio) as described by the manufacturer and [α-^{35}S] dATP or with the Klenow fragment of DNA polymerase and [α-^{32}P]dATP (9). The sequencing gels were decompressed at the CG regions by using dITP and Sequenase as suggested by the manufacturer or by running 8% acrylamide-urea (8 M) gels with 25% formamide.

The ST_A activity was determined by the suckling mouse

* Corresponding author.

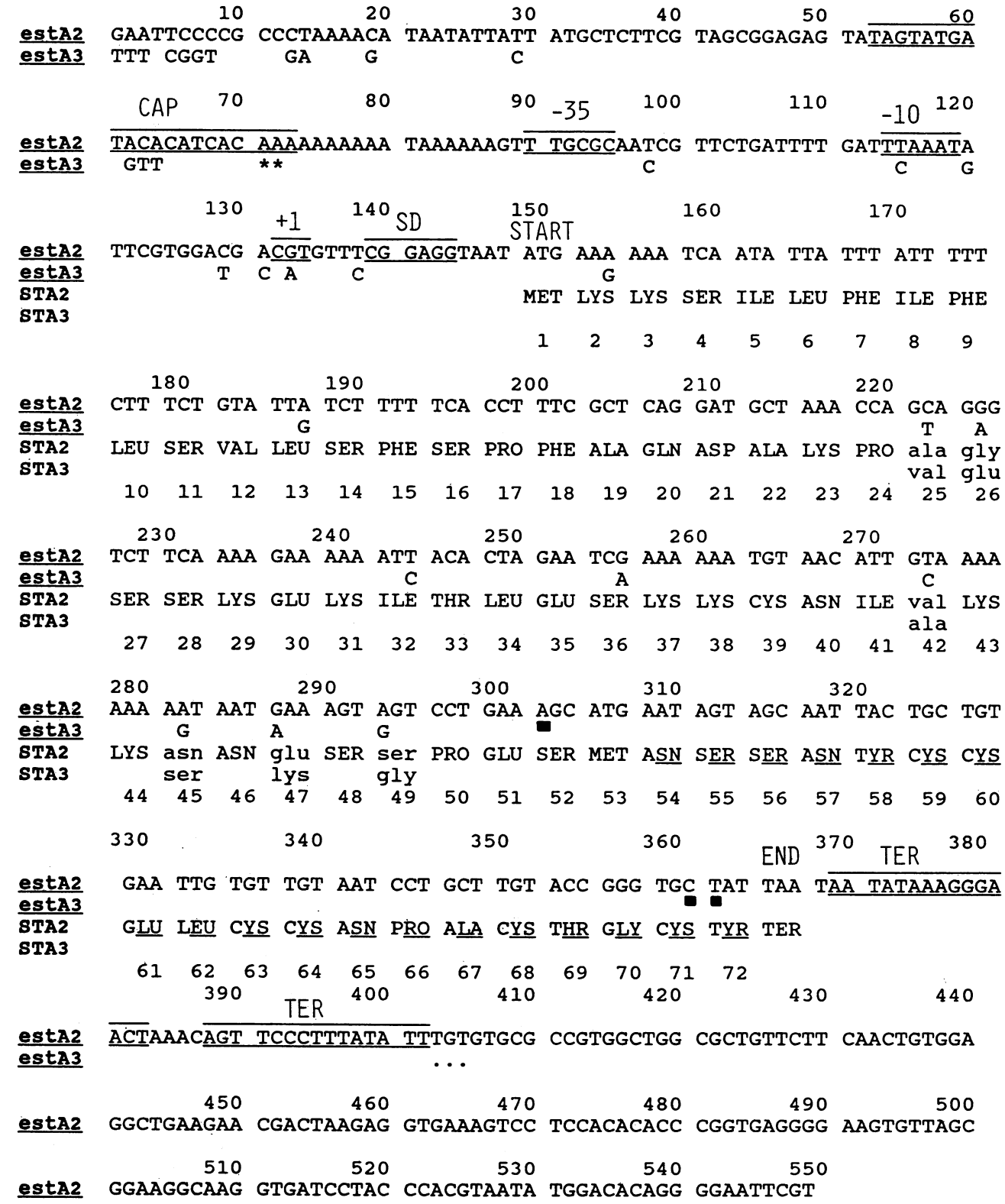

FIG. 1. DNA sequence (*estA2*) and inferred amino acid sequence (ST_{A2}) of the 549-bp *Eco*RI fragment that codes for ST_{A2} and comparison with the sequences of *estA3* and ST_{A3} (6) rectified in nucleotide position 203. The rectified bases of *estA2* are marked by closed squares (■). Nucleotides are numbered above the *estA2* sequence, and amino acids are numbered below the ST_{A3} sequence. For the *estA3* sequence, only the bases that differ from the *estA2* sequence are shown. The hypothetical cyclic AMP receptor protein (CRP)-cyclic AMP recognition site (CAP), the proposed −35, −10, and +1 positions, and the Shine-Delgarno sequence (SD) are shown above the *estA2* sequence. Also indicated are the proposed translation start (START) and end (END) sites. A potential transcriptional terminator (TER) is indicated. The open reading frame is shown as triplets from nucleotide positions 149 to 370. For ST_{A3}, only the amino acids that differ from ST_{A2} are shown. The asterisks in the *estA3* sequence represent bases not found in the sequence, and the dots indicate the last base of *estA3* included in the comparison. The extracellular toxin is underlined.

TABLE 1. Extracellular and cell-associated ST activity in wild-type and mutant *estA2*

Strain	Relevant property	ST[a] activity		
		Supernatant (%)	Sonicate (%)	Supernatant/ sonicate
HB101(pUC8)	Cloning vector	0.0	0.0	
HB101(pGK22)	Tyr-72	128.0 (87.6)	18.0 (12.3)	7.1
HB101(pGK28-2)	Tyr-72 → His	64.0 (87.6)	9.0 (12.3)	7.1
HB101(pGK28-9)	Tyr-72 → His	128.0 (93.4)	9.0 (6.6)	14.2
HB101(pGK26)	Tyr-72 → His + Gly-Ala-Leu	16.0 (55.6)	12.8 (44.4)	1.25

[a] Reciprocal of the dilution that gave a ratio, intestine/(body weight−intestine weight), greater than 0.08 in the suckling mouse model (12).

model (5). The samples to be tested were injected intragastrically, and after 4 h of incubation, the ratio of intestine to total body weight was determined. To localize the ST_A activity, bacteria were harvested by centrifugation from the culture, and the cells were resuspended in fresh media; cell-free supernatants and sonicated cells were assayed for ST_A activity.

Figure 1 shows the sequence of the 549-bp *Eco*RI fragment isolated from pRIT10250 that includes *estA2* and the inferred amino acid sequence of this allele. The previously reported *estA2* sequence (DeWilde et al., 1981) spans from nucleotides 10 to 370; His-72 was inferred as the carboxy-terminal residue of *estA2* encoded by the triplet (C_{362}AT); this triplet was now found to be the tyrosine codon (T_{362}AT) followed by two termination codons (TAA and TAA). Therefore, it is implied that ST_{A2}, like all the other ST_As, carboxy terminates with tyrosine. Also differing with the previously communicated *estA2* sequence was the Arg-52 codon (CGC) now determined to be the serine codon AGC (Fig. 1). Outside the *estA2* open reading frame, we identified a hypothetical transcriptional terminator (Fig. 1, nucleotides 369 to 403) that has a stem and loop structure; this structure is conserved in other *estA* genes (6, 12). Figure 1 also shows the complete *estA3* sequence and its inferred 72-amino-acid open reading frame (ST_{A3}). This single nucleotide difference between the previous *estA3* sequence and the current sequence data results in the substitution of the reported CCT (Pro-19) triplet for G_{203}CT (Ala); after this rectification, *estA3* and *estA4* are identical in nucleotide sequence. A comparison of the corrected DNA sequences of *estA* alleles 2 and 3 and the inferred amino acid sequence within the 222-bp coding region shows that 10 bases are different (Fig. 1); when translated, six amino acids differ between the two toxins. These differences are located in the intervening Pro region between the signal peptide (Pre) and the active extracellular toxin (ST).

The *estA2-eltB* fusion that resulted in the intracellular accumulation of pro-ST_{A2}-LT_B was designed with the assumption that ST_{A2} carboxy terminated with histidine (5). We now infer that its amino terminus is tyrosine and examined the effect of the change of tyrosine for histidine on the secretion and activity of ST_{A2}. Using two oligomer primers (15), the synthetic 22-mer mutant oligo-TATTATTA ATG ACA CCC GGTAC and the universal M13 primer, the TAT codon (Tyr) was substituted by the CAT codon (His); one additional base of the preceding triplet, TGC (Cys), was also changed (TGT [Cys]) to facilitate the identification of the mutant DNA during its isolation. Wild-type ST_{A2} and ST_{A2} His-72 were assayed to determine the cellular localization and biological activity of the toxins. The results of these experiments with two independently isolated ST_{A2} His-72 mutants and with wild-type ST_{A2} Tyr-72 are shown in Table 1. It is clear that in an isogenic background, the substitution of Tyr for His did not affect ST_A activity or its localization. In the three cases, the ratio of extracellular to intracellular activity was between 7 and 14; these values are within the experimental error inherent in the animal model used to assay the toxin. In contrast, the toxic activity detected from strain HB101(pGK26) ($ST_{His\text{-}72+3}$) was nearly 10 times lower than the wild-type level; furthermore, 44% of the activity was found associated with the bacteria. The lower ST activity and its cellular association seen with HB101 (pGK26), a *supEA4* strain, was also observed with CA8000 (pGK26), a *supE*$^+$ strain (data not shown). Thus, the addition of these three amino acids at the carboxy termini of STs alters the localization and diminishes the toxigenicity of the strains encoding such a molecule.

ST_As are extracellular *E. coli* polypeptides. Three structural elements have been defined in the 72-amino-acid precursor: a 19-amino-acid peptide (Pre) followed by 34 or 35 amino acids (Pro) which precede the 18- or 19-amino-acid extracellular toxin (ST). Pre-Pro ST_As are first cleaved by an activity susceptible to the energy metabolism inhibitor carbonyl cyanide *m*-chlorophenylhydrazone (J. K. Rasheed, L. M. Guzman-Verduzco, and Y. M. Kupersztoch, manuscript in preparation). The cleavage site in the fused ST_A-LT_B is between Ala-19 and Gln-20 (L. M. Guzman-Verduzco and Y. M. Kupersztoch, manuscript in preparation). The 19 amino acids that constitute Pre have the properties of a signal peptide (13), and the hydropathy of Pre is compatible with a membrane-spanning domain (12). The initially reported sequence of *estA3* (6) showed a single-base difference with *estA4* (12) within the Pre region and a two-base difference with *estA2* in the same region (Fig. 1). The single-base difference between the original *estA3* sequence (6) and *estA4* (12) caused triplet 19 to be proline and alanine codons, respectively. This position is the carboxy-terminal alanine of the *estA2* signal peptide, and it was thought to be important in defining the specificity of cleavage of Pre-Pro-ST to yield Pro-ST. Since we have shown that *estA3* and *estA4* have the alanine codon CCT as triplet 19 in extrapolation of the cleavage of Pre-Pro-ST-LT_B, in these toxins, the processing enzyme should cleave between alanine and glutamine. The rectification of the *estA3* sequence reported here makes *estA4* and *estA3* identical in the open reading frame and the regulatory elements (12). Therefore, this allele should be referred to as *estA3*. When the gene is translated, it is identical in its Pre and ST domains to that of Pre-Pro-ST_{A2} (Fig. 1). There is a high degree of homology (93.4%) between *estA2* and *estA3*. The largest sequence differences are in the intervening Pro region (8 of 102 bp, 92.2%) and in the regulatory region (12 of 100 bp, 88%), while the Pre and ST regions are very similar (55 of 57 bp, 96.5%, and 57 of 57 bp, 100%, respectively). However, following the putative tran-

scriptional terminator, the homology between the alleles diminishes to 46% (82 of 152 bp) (6, 12).

His-72 was considered to be the correct carboxy-terminal ST_{A2} residue in the previously reported design of the *estA2-eltB* fusion (5). As stated earlier, Pre-Pro-ST_{A2}-LT_B is cleaved to Pro-ST_{A2}-LT_B and remains cell associated (5); similarly, 44% of the ST_A activity encoded by pGK26 ($ST_{His\text{-}72+3}$) (Table 1) remains cell associated. This plasmid should yield an ST_{A2} with His-72 and additions of Gly-73, Ala-74, and Leu-75 in CA8000 (*supE*$^+$) and the insertion of Gln in place of the amber codon in HB101 (*supE44*). In the latter case, 24 triplets of the cloning vehicle are in-phase before the UAA termination codon. Thus, *supE44* in the host bacteria should results in an ST molecule that has the sequence His-72-Gly-Ala-Leu-Gln-Ser-Arg-Pro-Ala-Ala-Gln-Ala-Trp-Arg-Asn-His-Gly-His-Ser-Cys-Phe-Leu-Cys-Glu-Ile-Val-Ile-Arg-Ser at its carboxy terminus. Since a large proportion of the gene product of pGK26 in HB101 and CA8000 as well as that of pGSK51 in HB101 remained cell associated and the gene products have His-72 in common, we considered that the substitution of the natural Tyr-72 (Fig. 1) caused the molecule to transverse ineffectively the membranes of the producing strain. However, this possible explanation was eliminated when wild-type *estA2* was mutagenized to ST_{A2} His-72 and the mutation did not effect export or activity (Table 1). Nevertheless, blocking the ST_A carboxy terminus with three or more amino acids did affect the export of these polypeptides. We propose that for successful mobilization through the membranes, the ST domain has to contain the natural 18 or 19 amino acids not followed by other residues. Experiments are in progress to verify this hypothesis.

The lack of effect on the enterotoxic activity when ST_A Tyr-72 is mutated to ST_A His-72 is compatible with results of previous studies in which a synthetic 13-amino-acid peptide (positions 59 to 71) (Fig. 1) was shown to be toxic in the same animal model (14). Thus, the change of Tyr or its deletion does not affect the toxicity of ST_A. Recently, the pairs of cysteines that form disulfide bridges (10) of *estA1* (also called STp) have been mutagenized (17), and cell-free supernatants of these mutants do not have ST activity. However, the report did not indicate if there was inactive toxin in the supernatant and/or in the cell. Thus, the role of the six ST_A cysteines in the export process needs to be reexamined.

The 34-amino-acid Pro domain of *estA* is the least conserved region of Pre-Pro-ST_A. The six amino acids that differ between Pre-Pro-ST_{A2} and Pre-Pro-ST_{A3} (Fig. 1) all occur in this domain. The predicted conformations (4) of the Pro elements of these two toxins from positions 20 to 43 are predominantly helical structures, even when there are substitutions of a polar residue for a negatively charged one (Gly-26 → Glu). The region with the largest inferred conformational discrepancies lies between residues 44 and 50 and is mainly due to the different residues at position 47 (Glu in ST_{A2}, Lys in ST_{A3}). These changes, therefore, define regions of the molecule where the primary structure probably is not the determining factor in the export process. The role of the conserved amino acids in the Pro domain and their interplay during the export of ST_As have not yet been analyzed. We are currently mutagenizing this region to define its role in the extracellular delivery of this family of enterotoxins.

We thank Nigel Harford for supplying plasmid pRIT10250 and the poster in which *estA2* was first reported, Steve Moseley for plasmid pSLM004, Heather Stieglitz and Eric Hansen for critical review of the manuscript, and Cindy Baselski for typing it.

This work was supported in part by Public Health Service grant AI-21698 from the National Institutes of Health.

LITERATURE CITED

1. **Aimoto, S., T. Takao, Y. Shimonishi, S. Hara, T. Takeda, Y. Takeda, and T. Miwatani.** 1982. Amino acid sequence of a heat-stable enterotoxin produced by human enterotoxigenic *E. coli*. Eur. J. Biochem. **129**:257–263.
2. **Brickman, E. L., L. Soll, and J. Beckwith.** 1973. Genetic characterization of mutations which effect catabolite-sensitive operons in *E. coli*, including deletions of the gene for adenyl cyclase. J. Bacteriol. **116**:582–587.
3. **Carter, P., H. Bedouelle, and G. Winter.** 1985. Improved oligonucleotide site-directed mutagenesis using M13 vector. Nucleic Acids Res. **13**:4431–4443.
4. **Garnier, J., D. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
5. **Guzman-Verduzco, L. M., and Y. M. Kupersztoch.** 1987. Fusion of *Escherichia coli* heat-stable enterotoxin and heat-labile enterotoxin B subunit. J. Bacteriol. **169**:5201–5208.
6. **Moseley, S. L., J. M. Hardy, M. I. Huq, P. Echeverria, and S. Falkow.** 1983. Isolation and nucleotide sequence determination of a gene encoding a heat-stable enterotoxin of *E. coli*. Infect. Immun. **39**:1167–1174.
7. **Okamoto, K., K. Okamoto, J. Yukitake, Y. Kawamoto, and A. Miyama.** 1987. Substitutions of cysteine residues of *Escherichia coli* heat-stable enterotoxin by oligonucleotide-directed mutagenesis. Infect. Immun. **55**:2121–2125.
8. **Palva, E. T., T. R. Hirst, S. J. S. Hardy, J. Holmgren, and L. Randal.** 1981. Synthesis of a precursor to the B subunit of heat-labile enterotoxin in *Escherichia coli*. J. Bacteriol. **146**: 325–330.
9. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
10. **Shimonishi, Y., Y. Hidaka, M. Koizumi, M. Hane, T. Aimoto, T. Takeda, T. Miwatani, and Y. Takeda.** 1987. Mode of disulfide bond formation of a heat-stable enterotoxin (ST_h) produced by a human strain of enterotoxigenic *Escherichia coli*. FEBS Lett. **215**:165–170.
11. **So, M., and B. J. McCarthy.** 1980. Nucleotide sequence of the bacterial transposon Tn1681 encoding a heat-stable (ST) toxin and its identification in enterotoxigenic *E. coli* strains. Proc. Natl. Acad. Sci. USA **77**:4011–4015.
12. **Stieglitz, H., L. Cervantes, R. Robledo, R. Fonseca, L. Covarrubias, F. Bolivar, and Y. M. Kupersztoch.** 1988. Cloning, sequencing and expression in Ficoll generated minicells of an *Escherichia coli* heat-stable enterotoxin gene. Plasmid **20**:42–53.
13. **Von Heijne, G.** 1985. Signal sequences. The limits of variation. J. Mol. Biol. **184**:99–105.
14. **Yashimura, S. H., I. Kemua, H. Watanabe, S. Himoto, Y. Shimonishi, S. Hara, T. Takeda, T. Miwatana, and Y. Takeda.** 1985. Essential structure for full enterotoxigenic activity of heat-stable enterotoxin produced by enterotoxigenic *Escherichia coli*. FEBS Lett. **181**:138–142.
15. **Zoller, M. J., and M. Smith.** 1984. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. DNA **3**:479–488.