# The Virulence Gene Activator ToxT from *Vibrio cholerae* Is a Member of the AraC Family of Transcriptional Activators

DARREN E. HIGGINS,<sup>1</sup> EDWARD NAZARENO,<sup>1</sup> AND VICTOR J. DIRITA<sup>1,2\*</sup>

Department of Microbiology and Immunology<sup>1</sup> and Unit for Laboratory Animal Medicine,<sup>2</sup> University of Michigan Medical School, Ann Arbor, Michigan 48109

Received 22 June 1992/Accepted 31 August 1992

Virulence gene expression in Vibrio cholerae is postulated to involve ToxR-dependent activation of the toxT gene followed by ToxT activation of virulence genes, including several of those involved in biogenesis of the toxin-coregulated pilus. ToxR is a transmembrane, DNA-binding protein which is a member of the OmpR subclass of two-component activator systems in bacteria. Data presented in this report demonstrate that ToxT is similar to the AraC family of transcriptional activators identified in a variety of gram-negative bacteria. The toxT open reading frame begins approximately 200 nucleotides from the end of the tcpF gene, which is part of a cluster of genes responsible for production of the toxT is a potential stem-loop structure of an unusual nature which may play a role in regulating expression of toxT mRNA. Experiments with tcpF and toxT cloned behind a strong, constitutive promoter suggest that the two genes can be cotranscribed, but Northern (RNA) blot analysis of V. cholerae suggests that if they are, steady-state levels of their messages may be controlled by a posttranscriptional mechanism. Possible mechanisms for ToxR-dependent expression of toxT are discussed.

Regulation of several genes required for virulence of the human diarrheal pathogen *Vibrio cholerae* is under general control of two proteins, ToxR and ToxS (12, 13, 15, 18, 23). ToxR is an atypical member of the OmpR subclass of two-component regulators in bacteria, in that it does not have the phosphoacceptor domain homology characteristic of this family and thus must become activated in some way other than by being phosphorylated by a sensor kinase (15, 21). It has been proposed that ToxR may function as a homodimer and that ToxS is required for the stability of such dimers (5). Another aspect of ToxR that makes it different from the other members of the family is that although it is a transcriptional activator, it is also a membrane protein that binds the promoter for the operon encoding cholera toxin (ctxAB) while associated with the membrane (15).

Other ToxR-regulated genes are controlled indirectly by it, in that ToxR is required for them to be expressed but it does not directly activate their transcription. These include several genes required for assembly of the toxin-coregulated pilus (TCP) (23), a major colonization factor, and several accessory colonization factor (ACF) genes (18). The observation that ToxR is not sufficient for expression of virulence genes other than ctxAB led to identification of the toxT gene from V. cholerae. toxT encodes a product that can directly activate many ToxR-regulated genes, including tcpA, tcpC, tcpI, and ctxAB, and is itself under ToxR control (6). In this report, we demonstrate that the toxT product is related to the AraC family of transcriptional regulator proteins and demonstrate that toxT is located within the tcp gene cluster downstream of the *tcpF* gene and upstream of the *tcpJ* gene. Preliminary transcription analysis of toxT is also presented.

### MATERIALS AND METHODS

Bacterial strains and plasmids. Escherichia coli VM2 was used to assess ToxT activity. This strain is lysogenic for  $\lambda$ 

carrying a *ctx-lacZ* gene fusion and has been described previously (13). Plasmid pGJ40 is a pBR327 derivative isolated from a plasmid library of *V. cholerae* 569B and was previously shown to encode an activator of several ToxRregulated genes (6). Various derivatives of pGJ40 described herein were generated to determine the approximate location of the *toxT* gene within the cloned *V. cholerae* DNA. Bacterial strains were maintained at  $-70^{\circ}$ C in Luria-Bertani medium plus 20% glycerol.

DNA manipulations. Plasmid DNA was purified by using QIAGEN columns (QIAGEN, Inc.). Cloning and mutagenesis procedures were done by using standard protocols (11). A double-stranded-DNA sequencing protocol (kindly supplied by Yan Su) with synthesized sequencing primers was used. Plasmid DNA (5 µg) was dried under a vacuum, resuspended in 2 µl of 0.2 N NaOH, and incubated at 25°C for 2 to 3 min. Primer (2 µl; approximately 500 to 1,000 ng) was added, and then 2  $\mu$ l of double-distilled water and 2  $\mu$ l of annealing buffer (833 mM Tris-HCl [pH 7.5], 83 mM MgCl<sub>2</sub>) were added and the annealing was allowed to proceed for 2 min at 25°C, at which time 2 µl of 0.2 N HCl was added. Primer template hybrids prepared in this way were subjected to DNA sequence analysis with the Sequenase version 2.0 kit (U.S. Biochemical Co., Cleveland, Ohio). Sequencing products were resolved on 6.0 and 8.0% gels which were fixed, dried, and exposed to Kodak X-Omat film overnight at room temperature.

**RNA analysis.** RNA was isolated from overnight cultures of *V. cholerae* by the hot-phenol method previously described (6). RNA (10  $\mu$ g per lane) was resolved on 1.3% agarose-6% formaldehyde gels in MOPS (morpholine propanesulfonic acid) buffer (11). A separate set of lanes was always loaded and cut off from the remainder of the gel prior to blotting. This was stained with 0.5  $\mu$ g of ethidium bromide per ml in 0.1 M ammonium acetate (11) in order to ensure that equivalent amounts of RNA were loaded on the gels. The unstained portions of the gels were blotted overnight to nitrocellulose (Gelman Scientific, Ann Arbor, Mich.) with-

<sup>\*</sup> Corresponding author.



FIG. 1. Map of pGJ40. The plasmid harbors approximately 7 kbp from V. cholerae 569B cloned as a Sau3A fragment into the BamHI site of pBR327. ToxT activity is defined as the ability to activate a ctx-lacZ fusion gene in E. coli. Mutant derivatives are depicted below pGJ40. Deletions are represented by gaps; asterisks represent the location of frameshift mutations generated by restriction enzyme digestion (SpeI in pVJ60 and XhoI in pVJ44). Restriction enzyme sites are as follows: C, ClaI; G, BglII; H, HindIII; M, MluI; N, NcoI; R, EcoRV; S, SaII; Sp, SpeI; X, XhoI.

out prior treatment. Blots were baked in the morning in a vacuum oven at 80°C. Northern (RNA) analysis was done with restriction fragments removed from agarose gels (GTG agarose; FMC, Rockland, Maine) and labelled with <sup>32</sup>P by using a nick translation kit (Amersham, Arlington Heights, Ill.). For tcpEF, the probe used was an XhoI-SpeI fragment from plasmid pDH8 which contains the 3' end of the tcpEgene and the 5' end of tcpF. The toxT probe used was the *HindIII-XhoI* fragment fully within toxT shown in Fig. 1. Blots were prehybridized for 4 h at 42°C in 50% formamide-5× Denhardt's solution-5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaPO<sub>4</sub>, and 1 mM EDTA [pH 8.0])-0.1% sodium dodecyl sulfate (SDS)-100 µg of denatured salmon sperm DNA per ml. Blots were hybridized overnight at 42°C in the same solution except for the replacement of  $5 \times$ Denhardt's solution with 2× Denhardt's solution. Blots were given two 10-min washes both at 25°C in 2× SSPE-0.1% SDS and 0.2× SSPE-0.1% SDS. After being washed, the blots were dried and exposed to Kodak X-Omat film at -70°C with intensifying screens (Cronex Lightning; Dupont).

Isolation of the O395 toxT locus. Plaques from a  $\lambda$ -GEM11 library of V. cholerae O395 DNA (generously supplied by Kenneth Peterson, LSU Medical Center, Shreveport, La.) were probed with restriction fragments generated from the strain 569B clone pGJ40 (6). Three positive clones were identified, and DNA from one of these was subcloned as an 8.0-kbp EcoRV fragment into the EcoRV site of pBluescript. The resulting plasmid, pDH8, was used as a source of DNA to produce subclones for DNA sequence analysis of the O395 toxT locus.

Nucleotide sequence accession number. The GenBank accession number for the sequences presented in this paper is L01623.

#### RESULTS

Characterization of the V. cholerae 569B toxT locus. The plasmid pGJ40 expressing toxT (Fig. 1) was isolated from a library of Sau3A partial fragments generated from V. cholerae 569B and cloned into the BamHI site of pBR327 (13).

pGJ40 was identified by its ability to activate a ctx-lacZ gene fusion in E. coli and was subsequently shown to activate a number of other virulence-related genes from V. cholerae which also require ToxR for their expression, such as tagA (ToxR-activated gene A), tcpA, and tcpC (6). Mutational analysis of pGJ40 was done in order to identify the region of cloned V. cholerae DNA responsible for activation of ctxlacZ (Fig. 1). This analysis showed that deletion of Vibrio DNA downstream of an XhoI site within pGJ40 abolished ToxT activity, whereas deletion of a nearby ClaI site did not (Fig. 1, cf. pVJ44 and pVJ61). Filling in the XhoI site destroyed toxT, while filling in the upstream SpeI site did not. Thus, toxT resides between the SpeI site and the first ClaI site downstream of it. Consistent with this, analysis of 50 TnphoA insertions in pGJ40 that abolished ToxT activity showed that all but two of these mapped upstream of the MluI site shown in Fig. 1 (data not shown), indicating that toxT is primarily upstream of the *MluI* site and probably ends just downstream of it.

Expression of toxT cloned in pGJ40 was proposed earlier to initiate from the *tet* gene promoter of the cloning vector pBR327 (6). This was supported by the observation that insertion of TnphoA into vector restriction fragments containing the *tet* gene promoter abolished ToxT activity (data not shown). Thus, these data indicate that toxT in pGJ40 is expressed from the *tet* promoter and extends just between the XhoI and ClaI sites shown in Fig. 1.

toxT encodes an AraC-like protein and is part of the tcp gene cluster. The nucleotide sequence of approximately 2.5 kbp of DNA beginning just upstream of the BamHI site at which V. cholerae DNA is inserted in pGJ40 and ending past the ClaI site (Fig. 1) was determined. Analysis of this DNA sequence (Fig. 2) revealed two complete open reading frames (ORFs) and the start of a third on the same strand of DNA as the tet gene promoter. The downstream complete ORF spans the XhoI site and is predicted to encode a protein 32 kDa in size. This ORF was designated toxT, and comparison of this ORF with the National Biomedical Research Foundation and SwissProtein data bases showed it to be similar to several transcriptional regulators from different bacteria which collectively have been designated AraC-like proteins because of

J. BACTERIOL.

Met Ng Tyr Dy Lys The Lee Met Ser Ser The Met Ser The Met Ser Ser The Ser The Yell Tyr Lie Not The Met Ng Tyr Ang Ng Tyr Ang Off Tyr Lie Not The Yell Ng Tyr Ang Off Tyr Lie Not The Yell Ng Tyr Ang Ng Tyr Ang Off Tyr Lie Not The Yell Ng Tyr Ang Ng Tyr Ang Ng Tyr Ang Ng Tyr Lie Not The Yell Ng Tyr Ang Ng Tyr Lie Not The Yell Ng Tyr Ang Ng Tyr Lie Not The Yell Ng Tyr Ang Ng Tyr Lie Not The Yell Ng Tyr Ang Ng Tyr Lie Ng Tyr Ang Ng Tyr Ang Ng Tyr Lie Ng Tyr Ang Ng Tyr Ang Ng Tyr Lie Ng Tyr Ang Ng Tyr Ang Ng Tyr Lie Ng Tyr Ang	1	TCPP GGATCTTC0GTGCGATCTTTTCGCTCGTTGATAAACTAAGTTCATCTTTATAGTGAGGATAT ATG AAA AAA AAC TTA ATG TTA TCA ATC A	TG
9 ATT AGE TGA TT AGE TGA TT AGE GTA THA AGE AGE ATT AGE AGE AGE AGE AGE AGE ATT AGE TGA AGE GTA AGE TA AGE AGE ATT AGE		Net Arg Tyr Lys Lys Thr Leu Net Leu Ser Ile P	et 12
180 Out TCA AND GOD ANT ONG CHC CTH MA THA CT ATT THY ONA DTA THA MA ATT COT ANT AND ATH AND THY CHA MA   261 TOT GTE AND ATH CTH THY CAA ART THA THY AND THY ATH AND ATH AND ATH AND ATH AND ATH THY CHA AND THY PY DE SET THY CHA SHA AND THY PY AND ATH AND THY PY OS SET THY CHA SHA OFT AND THY AND ATH AND THY PY OS SET THY CHA SHA OFT AND THY AND ATH AND THY PY OS SET THY CHA SHA THY THY CHA AND THY AND ATH AND THY AND AND THY AND ATH AND THY AND AND THY AND AND ANT THY CHA AND THY AND AND ANT THY AND ATH AND THY AND ANT THY AND AND ANT ANT AND ANT THY AND AND ANT ANT AND AND ANT ANT AND AND ANT ANT AND AND ANT ANT ANT AND AND ANT ANT ANT AND ANT ANT AND AND ANT ANT AND AND ANT ANT AND AND ANT ANT ANT AND ANT ANT AND AND ANT ANT ANT AND ANT	99	ATT ACA TCA TIT AAC TCA TIT GCA TIT AAC GAT AAT TAT AGT TCA ACC AGT AGG GIT TAT GCT AGG TCT AAT GAA GCT AGG Ile Thr Ser Phe Aen Ser Phe Ala Phe Aen Aep Aen Tyr Ser Ser Thr Ser Thr Val Tyr Ala Thr Ser Aen Glu Ala Thr	39
161 TOT OTE AM ANT TOT TOT OCA AND TAT AND TAT NATE AND TAT AND COL INC OCA TOE ACA AND CAT YEAR DILLY INC.   142 AGA ACA ANA GA ANT TOT TOT OCA COL CAT TATA OCT OUT ANT TAT AND TAT AND TAT TAT AND TAT AND AND TAT TAT AND ANT TAT TAT AND TAT TAT AND TAT AND ANT TAT TAT AND TAT TAT AND CAT AND TAT TAT AND TAT AND ANT TAT TAT AND TAT AND ANT TAT TAT AND ANT TAT TAT AND TAT AND ANT TAT TAT AND CAT AND TAT AND CAT AND	180	GAT TCA AGG GGG AGT GAG CAC CTT AGA TAC CCT TAT TTA GAG TGT ATC AAA ATC GGT ATG AGT AGA GAT TAT CTT GAA AAC Asp Ser Arg Gly Ser Glu His Leu Arg Tyr Pro Tyr Leu Glu Cys Ile Lys Ile Gly Met Ser Arg Aep Tyr Leu Glu Asn	66
142 AGA MAA BA GAT THE DET GET GOD GAT THE GAT OUT OAT THE TAT GAT AND THE TARE GAT THE TARE GAT THE TARE GAT THE TOP OF 100 100 PM BE SER AND AND THE PM BE SER AND THE TARE GAT THE GAT AND GAT AND COD ANT AND GAT ANT GAT AND GAT THE GAT AND GAT ANT COD GAT AND CAN AND COD ANT AND GAT ANT CAN ANT COD ANT THE AND GAT ANT CAN ANT COD ANT THE AND GAT ANT CAN ANT COD ANT AND COD ANT THE AND GAT CAN THE AND GAT ANT CAN ANT CAN ATT CAN AND CAN ATT CAN AND CAN ATT CAN AT CAN ANT COD ANT THE TARE AND COD ANT THE AND CAN ATT CAN AT	261	TOT GTG AAA GTA TCT TTT CCA ACC TCT CAA GAT ATG TTT TAT GAC GCA TAC CCA TOG ACA GAA TCT GAC GGT GCT AAG ACA Cys Val Lys Val Ser Phe Pro Thr Ser Gin Asp Net Phe Tyr Asp Ala Tyr Pro Ser Thr Glu Ser Asp Gly Ala Lys Thr	93
423 OCT GAN AGG ACT THE GAT TOG GAN ATA COL AGG ANT GAN ATT GAG AGT GTT MT TAN TAN GTT ANT GAN GOT ANT GAG AGT GAN ATA COL GAN AGT AGT TAN CAN AGT AGT TOG TAN TAN GAN AGT TAN TAN GAT AGT TOG 11   564 TCT AGG GAN ATT TAN CAN AGE TAT ATT AGG AGT GAN TTT GT AGA AMA GAM AGT GAT GAT TATT AGT ANT GAN AGT AGT VILL 11   565 TCT AGG GAN ATA ATT CAN AGE TAT ATT AGG AGT GAN TTT GT AGA AMA GAM AGT AGT GAT TATT AGT AMT VILL 11   566 GCT AGT AMT ACT CAN GT AMA TTA AGG AGT CAN TTT GTT AGA AMA GAM AGT AGT GAT AGT TATT AGT AMA VILL 11   567 GTT GCA AGT AMT ACT CAN GTT AMA CTA AGT AGT CAN TTT GTT AGA AGA AGT GTT AGT ATA AGT AGT	342	AGA ACA AMA GAA GAT TTC TCT GCT COG CTT TTA GCT GGT GAT TAT GAT AGT TTA CAG AMA TTG TAT ATA GAT TTT TAT CTT Arg Thr Lys Glu Aep Phe Ser Ala Arg Leu Leu Ala Gly Aep Tyr Aep Ser Leu Gln Lys Leu Tyr Ile Aep Phe Tyr Leu	120
504 FOT ACC GOA CTT ANT CAG BOT TA ATT CAG BOT GOA TTT CTT ACA ANA GAA ATT GOT COAT ATT GTT ACT ATT TO TO ANT CTT GOT ATT ATT CAG BOT COA TTT GTT TACA ANA GAA ATT GTT ACT ATT GTT ATT ATT CTT ACA ANT CAG ATT GTT CAR ANT GAT ATT GTT ATT ATT CTT ATT ATT ATT ATT ATT ATT A	423	GCT CAA ACG ACT TTC GAT TGG GAA ATA CCG ACA AGG GAT CAA ATT GAG ACA CTA GTT AAT TAT GCT AAT GAA GGT AAA TTG Ale Gin Thr Thr Phe Asp Trp Giu Iie Pro Thr Arg Asp Gin Iie Giu Thr Leu Vel Asm Tyr Ale Asm Giu Giy Lys Leu	147
955 Got GTT CCA GAT ANA TAT CCA GCT ATT GTT TAT ANA CCA GCT ATT GTT ATT ANA CCA GGA CTA ATG GTT. ACA ATT TC TAT GTT ATA ATA   966 GCA ATT CCC ANT GAA ATT TAT CCA GCT ATT GAA GTT TAT GAA GGA GTT CTA GTG GTT ATA ACA AT TTC TAT GAA ATA CTA GTT GTT ATA ANA CCT GGT TAT AAA ATA CAT GTT GTT ATA ANA CT GTT TA GAA ATA ATA 21   977 GCG GTT TA GAA ATT TAT GCA GAA ATA GAA TAT ATA GAA GTT TAT GAA GGA ATT GTT G	504	TCT AGG GCA CTT AAT CAA GAG TAT ATT ACA GGT CGA TTT CTT ACA AAA GAA AAT GGT CGA TAT GAT ATT GTG AAT GTT GGT Ser Thr Ala Leu Asn Gin Giu Tyr Ile Thr-Gly Arg Phe Leu Thr Lys Giu Asn Giy Arg Tyr Asp Ile Val Asn Val Giy	174
666 ALE THE OCH ANT GUA ATT THY COULD AT THY GUA AGE ATT CTT AGE CAT ATT COT AT ATT ANA OCT OUT GUA ATT ATT CALL ONE ALL LIE SET AND LIE LYE FOO GUY GUA ALL AND COT GUA ATT ATT CALL AND ALL ALL ALL ALL ALL ALL ALL ALL ALL AL	585	GOT GTT CCA GAT AAT ACT CCA GTT AAA CTA CCA GCT ATT GTT TCT AAA CGC GGA CTA ATG GGT ACA ACT TCT GTA GTT AAT Gly Val Pro Asp Asn Thr Pro Val Lys Leu Pro Ale Ile Val Ser Lys Arg Gly Leu Net Gly Thr Thr Ser Val Val Asn	201
747 ACA OTA THA GAA THA GATA AND GAA ATTA AGT AND CAC GAC THA CCA AT CTC TOG CAC TTO CAC THT AND GTT CTT GTT 221   828 GGA GTT CCA CH GCA GAA ACA GGA GTT ATC TAT GAC CCT GTT TAT GAA GAA AT GTT CAA CAC TAT CAC GCA GTT ATTA GAA GAA AT GTT AAA GAC ATA TO AAA ACA TAT CAA CAC TAT AND GAT ATA CAAC 21   909 TTA AGG GOT TAT GTA GAA ACA GGA GTT ATC TAT GAC CT GTT TAT GAA GAA AT GTT AAA CAC TAT CAA CAC TAT ANA ACA ATA CT ATG GTT CAA ATA CT ATG GTT CAA ATA CAA ATG GTT TAT ATA CAA AGG GAA ATTA GTA TAT ATA CAA AGG CAA ATA GTA TAT ATA CAA AGG ATT GTA ATA ATA ATA AGG GAA ATA TAT ATA ATA	666	GCA ATT CCC AAT GAA ATT TAT CCT CAT ATC AAA GTT TAT GAA GGG ACT CTT AGT CGT TTA AAA CCT GGT GGA ATG ATC Ala ile Pro Asn Glu ile Tyr Pro His ile Lys Val Tyr Glu Gly Thr Leu Ser Arg Leu Lys Pro Gly Gly Ala Het Ile	228
828 GOR OTT COC NAT GOR GAR ACA GOR OTT ATC THE GAC COC OTT TAT CAR GAR ACT GAR CAR TO TANA CARC TAR CARC ON UNIT POINT FOR AN OTA ATC AND CAR ATC GAR CARC ATT AND AND CAR ATT ATT AND AND CAR ATT ATT ATT AND AND CAR ATT GAR ATT ATT ATT AND AND CAR ATT GAR ATT ATT ATT ATT ATT ATT ATT ATT ATT A	747	GCA GTA TTA GAA TAT GAC GTA AAT GAA TTA AGT AAG CAC GGG TAT ACC AAT CTG TGG GAT GTG CAG TTT AAG GTT CTT GTT Ala Val Leu Glu Tyr Asp Val Asn Glu Leu Ser Lys His Gly Tyr Thr Asn Leu Trp Asp Val Gln Phe Lys Val Leu Val	255
909 TTA ACO GOT ANG ANG TCT TAT ANT GTA TCT ACA ANT GAC ATO CAT ANT ANT GGT TAC AND TGO TCA ANT ACT ATO TTC TCA ANT Lew Thr Gly Lys Lys Lys Lys Lys Lys Lys Thr Ann Cat Art TAT ANA CA ANT AND AND THE AND GUY TY Lys TTP Ser Asn Thr Met Phe Ser Ann 31   900 TCT ANT TAT ANA ANG CA ANT TA TTA TTA TATA ATA AND ANG ANT GGA ANT GTT TATA ACC ANA GCA TAT TCA GA ANA 33   1071 TTT ANA TANANATAGATAMACHAGCAGCACATTOUTGOCONTTATCTTATCTCAMAMACATAMAMATAMACHAGGATACTTATCTTATCTTATCTATCTATATTATCTATCT	828	GGA GTT CCA CAT GCA GAA ACA GGA GTT ATC TAT GAC CCT GTT TAT GAA GAA ACT GTC AAA CCA TAT CAG CCT AGT AAC AAC Gly Val Pro His Ala Glu Thr Gly Val Ile Tyr Asp Pro Val Tyr Glu Glu Thr Val Lys Pro Tyr Gln Pro Ser Asn Asn	282
990 TCT MAT TAT ANA ACD CAA ATA TTA TTA ACT ANA OGA GAT OGA ACT GOT GTO ANA CTG TAT AGC ANA GCA TAT TCA GAG ANC Ser Amn Tyr Lyr Thr Gin lie Leu Leu Thr Lyr Gly App Gly Ser Gly Val Lyr Leu Tyr Ser Lyr Ala Tyr Ser Glu Ann 33   1071 TIT ANA TANATAGATAACAGCCACATTOGTOGCCUTTATCTTATCTATACTAAAATAAAATAACATAAAACATAAAATAACATAAAATAACATAAAATAACATAAAACATAAAATAACATAAAATAACATAAAATAACATAAAATAAATAAT	909	TTA AGG GGT ANG ANG TTG TAT AAT GTA TGT ACA AAT GAC ATG CAT AAT GGT TAC ANG TGG TGA AAT ACT ATG TTC TCA AAT Leu Thr Gly Lys Lus Leu Tyr Asn Val Ser Thr Asn Asp Met His Asn Gly Tyr Lys Trp Ser Asn Thr Met Phe Ser Asn	309
1071 TTT AAA TAAAATAGATAAGAGGGGGGATTAACTAGGGGGGTTAACTAAAGAAGAAAAAAAA	990	TCT MAT TAT MAN ACG CAN ATA TTA TAT ATA AGA GAT GGA AGT GGA AGT GGT G	336
1179 AACTTOTICTIANGTCHTTANGTAGTTIANATAACTTOGATTIANGTAAACTTIAGTGGGATGGCTCTCTGGGTTIANTTANATATCTTCAAAACGAAGGATGGTATGGGATGGCTCTCTGGGTTIANTTANATATCTTCAAAACGAAGGATGGGATGGCTATGGGGAAAAACTTAAGTAGGATGGGGAAAAACTTAAGTAGTTAGAAACTTAATAACTTAATAATGTTGGAAAGCAAATGAAGGATGGGGAAAAAAATGAGTTTTGAAAAATGAGTAGAAGA	1071	ТТТ АЛА ТАЛАЛТИБАТАЛБАТАЛБАБСКАСТТООТОБСТОТТАТСТТАТСТСАЛАЛАЛСАТАЛАТАЛСАТОЛОГТАСТТАТОТТАТСТТАТОТАЛТАСОТСТ Pho Lys ***	GT 338
1179 ANCTROTICTION OTCOUNT AUGUMENTITY ATTANCTICGA FITTINAMA MAACTITING OTGA ROGCICTC COGGETITATTITATATATCTICGA METAACTITING ANALOCCIC COGGETITATTITATATATCTICGA METAACCICA   1286 ATG ATT GOG ANA ANAL TOT TIT CAA ACT ANT GTA TAT MGA ATG MGT ANALTIT GAT ACT TAC ATA TIC ANT ANC TIA TAT ATA ATA MAT HE CILL CILL CILL CILL CILL CILL CILL CIL			
Text   1286 ART GOG AMA AMA TOT TIT CAA ACT AMT GTA TAT AGA ATO AGT AMA TIT GAT ACT TAC ATA TIC AMT AGC TATA TAT ATA Net Ile Gly Lys Lys Ser Phe Gin The Aen Val Tyr Arg Het Ser Lys Phe Aep Thr Tyr Ile Phe Aen Aen Lew Tyr Ile 27   1367 AGC GAC TAC AMA AND TIC TOG ATA GAT AGT GGA ATT GCA AMG CIT ATA GAT AMA ATT CT TG GTT AGT TAT GAA ATA AMA TA APP Tyr Lys Het Phe Typ Ile Aep Ser Gly Ile Alle ye Lew Ile Aep Lys Aen Cys Lew Val Ser Typ Col Ile Ama Yal Ser Ser Ser Ser Ile Ile Lew Lys Lys Aen Alle Ile Gin Arg Phe Ser Lew The Ser Lew Val Ser Typ Col Ile Ama Yal Ser Ser Ser Ile Ile Lew Lys Lys Aen Alle Ile Gin Arg Phe Ser Lew Thr Ser Lew Val Ser Typ Col Ile Ama Yal Ser Tota ATT ACT ATA CT TA TA CT TA GA AMA ACC GCT ATT CHO GGA TIT CT TTG ACT TAT AGA AMA ATA ATA AMT ATT Ser Val Ile Thr Ile Ser Aep Ser Phe Ile Arg Ser Lew Lys Ser Tyr Ile Lew Gly Aep Lew Het Ile Arg Asn Law Tyr 10 81   1529 TCT GTA ATT ACA ATA TAT CT TOG AAT TOT GAA ATA GAT ATA GAT ATA GT GT CT CT AGT ATA AMG AMA TAT ATA Ser Val Ile Thr Ile Ser Aep Ser Phe Ile Arg Ser Lew Lys Ser Tyr Ile Lew Gly Aep Lew Het Ile Arg Asn Law Tyr 10 81   1510 AAT AAA AAT CT TA TA CT TOG AAT TOT GAA ATA GAT ATA GAT ATA GC GT CT TO TAG AAA AMA ATA ATA ACT TAT ATA Ser Val Ile Thr TAT ATA CT AAT TAT CT TAG AAA AGT THT TIT TA CA GAG ATA AAA ATA ATA AAT TAT CT ATA Ser Val Ile Ana TAT ACT CA ATA ATA CT TAT CT AAAA GT THT TIT TA CA GAG ATA AAA AAA ATA ATA AAA CT TAT TATA Ser Val Ile Ana TYP SER ABG Glu Phe Lew Lys Val Phe Phe Ser Gly Phe Phe Ser Glu Aan AAA AAA ATA TATA ACT ATA Ser Yan Ama TAT TAT CA ATA CAT ATT CO ATA TOG AAA TOG ATA TT TAT AAA AAA ATA TAAA CAT ATT ATA ATA TAT ACT CAT ATA AGT CAT ATA COA TA COA TAT COA ATA TAT CA ATA AAA AAAA A	1179	аасттотусттатотототатотасттатататаасттокаттаталасттакотокатокостостокосттататататататататаластас <u>акотак</u> алокоса	
1367 ANC GAC TAC ANA NTO TCC TOG ATA GATA NGT DGA ATT GCA MAG CTA TATA GAT ANA TOC TTO GTT GAT TACT TAT GAT ATA ATA AGA AGP TYT LYS NHT PHO TTT IIS ASP SET GIV IIS ALL LYS LEW LIE AGE LYS LEW VAL SET TYT GIU IIS ASP SET 1448 TCA AGT TCA ATA TACT TA CTC AGA GAA ANA TOC GCA ATT TCT TOT ACT TAT GAT GAA ANA TATA ANT GTT SET SET SET SET IIS IIS ES LEW LYS LYS ASM ALS IIS GIN ATT CT TOT ACT TAT TCA GAT GAA ANA TATA ANT GTT SET SET SET SET IIS IIS ES ASP SET PHO IN LYS SET YFT IIS LEW UIT SET ASP GIU ASM TITA TATA TATA TACT ATA CTA ATTA AGT CAT TCA TTT ATA CGT TCA CTA ANA TCT TAC ATT CTA GGT ATA AGG AAT TATA TAT SET VALI IIS TAT IIS SET ASP SET PHO IIS ANT GOT TCA CTA ANA ATC TAC ATT CTA GGT ATA AGG AAT TATA TAT SET VALI IIS TAT TITS SET ASP GIU TTA ATA CGT TCA GTA ANA TCT TAC ATT CTT GGT ATA AGG AAT TATA TAT GUN ANA ANA CAT CTA TTA CTT TOG ANA TOT GGAA CAT AAT GAT ATA GOT GTA CTT CT GAA ATA AGT ATA AGT TTA TATA GUN ANA ANA CAT CTA TTA CTT TOG AAT TOT GGAA CAT AAT GAT ATA GOT GTO CTT CT GAA ANA ANA ATA TAT ACT TTA TATA GUN ANA ANA CAT CTA TTA CAT TAG AT CAG ATT TTT TTT TTA CTA GOG TTO TTT CTT CT GAA ATA AGA TAT AGC TAT ATA GUN ANA ANA CAT CTA TATA CAT TCA CTA TATA GAT TTA CTO CTA CAT ANA CTA TATA COT COT ATA GUN ANA ANA TAT ANT CAT CTA CAT ATA CAT ATT CAT CAT TTA GTA ANA AMA ANA ATA TAT CAG COT ATA TOG COT TTT ATT ATT CAG CAT CAT CAT CAG ATA COA ATA CAT CAT CAT TTT GTA ANA AMA GAA TAT AGC COT ATA TOG CAT ATT TOT GAT GAT CTT CAAT COA ATA COA ATA TAT COA CATA ATA AMA ATA AATA TAA ANA TAT AGA CATT ATA ASP IIS CYS GIV GIU ESU LEW LYS IIS SET CYS LEW YAL LYS SET ASP IIS TTA ANA TT AGA CATA ATA ASP IIS CYS GIV GIU ESU ATA CATA TT CAA ATA AAA AAA AAA ATA TAA CATA TAG CAAT TAT ASP IIS CYS GIV GIU ESU ANA TTA AGA CATA TAT CAA CTT AAA ATA AAAA AAA	1286	Tour And All oge and and it. Til can act ant gir and and and age and all til gir act it. Air it. And the and and the Met lie gir Lys Lys Ser Phe gin the and val tye arg Met Ser Lys Phe asp the tyr lie Phe asm asm Lew Tye lie	27
1448 TCA AGT TCA ATT ATC TTA CTC AMO AMA AMC GCT ATT CHG GGA TTT TCT TTG ACT TA TCA GAT GAA AMT ATA AMT GTT 61   1529 TCT TCT TTA ATC ATT ATC CTA TTA TCA GTT ATCA GTA GTA GAA AMA AMT ATA AMT GTT 61   1529 TCT TCT TA ATC CAT ATT ATC CTT TTA TCA GTT ATCA GTA GTA AMA AMA AGT GTT AGT 61   1529 TCT TCT ATA ATC CAT TTA TCA TTTA TCA TTG GTA CTT ATA AMA TCT TTA CTT GT GTA GTA AMA AMA AMT ATA AMT GTT 10   1510 MAT GAA AMT AAA CAT CTA TTA ATC TTG GAA CAT AMA ATT GTA ATA GCT GTC GTA ATA AGG AMT TA ATA GT 10   1610 AGT GAA AMT AAA CAT CTA TTA CTT TGO AAT TGT GAA CAT AMT GAT ATA GCT GTC TTC TGG AAA GTA AMA GTT TTA AMA GTT TTT ATA GAA TAT AGC TTA TTA AGA GTT ATT ATA GTA TAT TCA GAA TAT AGC TTA ATT AMT GTT TCA AMA GTT TTT TTA CTA GAA AMA AAA ATA AAA ATA AAA CTT ATA AGA TTA AGT TTA GTA AMA GTT TTT TTA AMA ATA GTT ATA AMA GTT GTA AMA GTT GTA AMA GTT GTA AMA GTT TTT ATA AMA ATA TA AC TCT ATA TA GT AGA TTA AGT AGA TTA AGT AGA TTA AGA AG	1367	ARC GRC TRC ARA ATG TTC TOG ATA GAT AGT GGA ATT GCA ANG CTT ATA GAT ARA AAT TGC TTG GTT AGT TAT GAG ATA AAT Asn Asp Tyr Lys Met Phe Trp Ile Asp Ser Gly Ile Ala Lys Leu Ile Asp Lys Asn Cys Leu Val Ser Tyr Glu Ile Asn	54
1529 TCT GTA ATT ACA ATA AGT GAT TCA TT ATA COT TCA CTA AAA TCT TAC ATT CTT GOT GAT CTC ATG ATA AGG AAT TTA TAT 10   1529 TCT GTA ATT ACA ATA AGT GAT TCA TT ATA COT TCA CTA AAA TCT TAC ATT CTT GOT GAT CTC ATG ATA AGG AAT TTA TAT 10   1610 AGT GAA ATT AGT CTA TTA CTT TGG AAT TGT GAA CAT AAT GAT ATA GOT GT TCT GTG AAG TG ATA AAT GAT GT TG AAA 10   1610 AGT GAA ATT AAGT CTA TTA CTT TGG AAT TGT GAA CAT AAT GAT ATA GCT GT TCT CTG AAA GTG GTA ATA GGT TCT AGA 13   1651 GAA ATT AAT TAT TCA GAT GTC CTA AAA GTT TT TTT TCA GGG TCT CT CTG CAAA GTA GAA AAA AAA TAT AAC GTT ATA 13   1691 GAA ATT AAT TAT TCA GAT GTC GTA TG GAA GAT ATT GTG GT TCT TCT TCG AAA GTA AAA AAA TAT AAC GTT ATA 16   1772 TTT ATT ACT GAT GT GT GT ATG GAA AAA ATT TAT TTT TT	1448	TCA AGT TCA ATT ATC TTA CTC ANG ANA ANC GCT ATT CHG OGA TTT TCT TTG ACT TCA TTA TCA GAT GAA AAT ATA AAT GTT Ser Ser Ser Iie Iie Leu Leu Lys Lys Asn Ala Iie Gin Arg Phe Ser Leu Thr Ser Leu Ser Asp Giu Ann Iie Ann Val	81
1610 AGT GAA ANT ANA GAT CTA TTA CTT TOG ANT TOT GAA CAT ANT GAT ATA GCT GTC CTT TCT GAA GTG GTA ANT GGT TTC MARA 13   1651 BET Clu Men Lyw AND Leu Leu Leu Leu TEP ANN CYG Glu His Ann AND JIE Alla Val Leu Ser Clu Val Val Ann Cly THM ANG 13   1651 GAA ATT ANT ATT CA GAT GAT CTT CTT TA CA GOG TTT CTT TT CT GAA GTA GAT ANT GGT TTA ANT GGT TA ATT GAT GLU Val Val Ann Cly TATA 13   1651 GAA TTA ATT ATT TCA GAT GAT CTT GAT GAT GAT ATA GGT TTT TT TT CA GOG TTT CTT TC TCA GAA GTA GAA ANA ANT ATA ACT CTA TAT 14   1610 HA TT ANT ATT CA GAT GAT GAT GAT GAG ANA ATT TCA TOT TTA GTA ANA ANT GAA TAT ACG CGT ATA 16   1772 TTT ATT ACT GAT GAT GAT GAT GAT GAT GAG ANA ATT TCA TOT TTA GTA ANA ANT GAT TAT GAG GAT ANT TOG CGT TGG GCA 18   1853 GAT ATT TGT GGT GAA GTA AGC ANT CGG ATT ATT TT TAT ANA MAG GAT CTA GG GGT ANA GTT AGA ATA ACG GGT GGA GAA TAT CAA TTA AGT TAT CAA TT AGA ATA ACG GGT GGA GAA TAT CAA TT AGA TAT CAA TT AGA TAT AAA CCG AGT GAA TAT AAA CGA ATT GCA TAT CAA TT CAA TT CAA TT CAA TT ATA ATA	1529	TOT GTA ATT ACA ATA AGT GAT TCA TTT ATA OGT TCA CTA AAA TCT TAC ATT CTT GGT GAT CTC ATG ATA AGG AAT TTA TAT Ber Val 11e Thr 11e Ser Amp Ser Phe 11e Arg Ser Leu Lys Ser Tyr 11e Leu Gly Amp Leu Met 11e Arg Amm Leu Tyr	108
1691 GAA ATT ANT TAT TCA GAT GAO TTC CTA ANA GTT TTT TTT TCA GOG TTC TCC TCG ANA GTA GAA ANA ANA TAT ANC TCT ATA GLU ILE MEN TYP SER AND GLU PHE LEU LYS VAI PHE PHE SER GLY PHE PHE SER GLY VAI GLU LYS LYS TAR SER ILE 1772 TTT ATT ATT ACT ACT TCA GT GC TAG GAO ANA ATT TAT TTT TTT TCA ANA ATA ANA ATA TAA ACT TCT ATT PHE ILE THA CAP AT CAT CTC ATA GCA TAG GAO ANA ATT CAT TTT GTA ANA ANA ATA TAA ACT CAT ATT APP ILE THA CAP ATA CAT CTC ATA GCA TAG GAO ANA ATT CAN ATT TTG ANA ANA ATA ANA ATA TAA ACT TAT ATA GO CAT TTT GT ANA TTT TCT GOT GAA TTA AGA ACC AAT COG ATO ATT TTG ANA ANA GAA CTA GAD TCT CAO GOA GTA ANG TTT AGA ANA ATT ANA CTA ANT TAA ACT AGA TTA APP ILE CYS GLY GLU LEW APP ALS HET ILE GEU LYS LYS GLU UNG US SER AND GLY VAL LYS PHE ART GLU LEW 11934 ATT ATA MCC ATT CGA ATA TCA ATT TCA ATA CAO ATT ATA ANA ACC GOT GAA TTC ANA ATA ANA CAG ATT GCA TAT CAO TCT 118 ANN SER ILE AFT ILE SER TYS SER ILE SER LEW HET LYS THE GLY GLU PHE LYS GLU ILE ANA CAG ATT GCA TAT CAO TCT 242   2015 OGG TTT GCT TCA TAT TTT TTT TAC ANA TTTA AGA CAC ATA TAT ANA CGA ATT GAA ATT ATT TAC TAT TAT TAT TAT TAT TCD CLY PHE SER THY VAL TYN LAG TCA ACC ATA ATT ANA CTA GAA CAT ATT ANA TTT AGA TAT 274   2056 ACA <u>GGA GTT</u> GCA GAA ANA TA ANA CAG ATC GTT TAC TTA ATT AGA TAT CAA TT TAC GAT TAT TAC GAT TAT TAC GAT TAT TAC GAT TAC CTT TAC TAG ATA TCT ATT TO GAT ATT TAG GAT ATT TAGA ATA CAC ATT TAG ANA TAT ANA CAC ATT TTA GAT ATA TO GAN TAC TAC THE GLU YEL ALS TTA CAA ATA ANA TAC GAT TAC TAG ATA CTT ATT GAT TAC CTT ATT TO ATT GAT ACA CTT ATT GAT ATA TO GAT ATT TAG GAN TAT TAG THE GLU YEL ALA TAT GAA TAC GAT TAC TTAC TTA ATT TAC TAT TAC TAT TAC TAT TAC TAC	1610	AGT GAA AAT AAA GAT CTA TTA CTT TOG AAT TOT GAA CAT AAT GAT ATA OCT GTC CTT TCT GAA GTG GTA AAT GOT TTC AGA Set Glu Aen Lye Aep Leu Leu Leu Trp Aen Cye Glu His Aen Aep Ile Ale Val Leu Set Glu Val Val Aen Gly Phe Arg	135
1772 TTT ATT ACT GAT GAT CTT GAT GCT ATG GAG ANA ATT TCA TGT TA GTA ANA AGT GAT ATT ACG CGT AAT TCG CGT TGG GCA 18   1873 GAT ATT TGT GGT GAA TTA AGA ACG ANT CGG ATG ATT TTG ANA ANG GAT ATT ACG CGT AAT TCG CGT TGG GCA 18   1853 GAT ATT TGT GGT GAA TTA AGA ACG ANT CGG ATG ATT TTG ANA ANA GAA CTA GAG TCT CGA GGA GTA AGT TTAG AAA TTA 18   1934 ATT TGT GGA ATTA AGA ACG ANT CGG ATG ATT TTG ANA ANA GAA CTA GAG TCT CGA GGA GTA AGT TGG ATT ACG TCT 21   1934 ATT ATT ACG ATT CGA TAT TCA ATT TCA CTA ATG ANA ACG GOT GAA TTA AAA CAA ATT ACA TATA CAG TTA TCA CTA ATG ANA ACG GOT GAA TTG AAA ATA ANA CAG ATT GCA TAT CAG TCT 24   2015 GGG TTI GCA TAT CTA TCA TTA TCA ATT TCA CTA ATG ANA ACG GOT GAA TGTA ACA CAG ACT AGA TAT TTA TTA TGT TTG 27   Gug TTT GCT AGT CAT TAT TTT TT CA ACA TAT ACA ACA TGT AGT AGA CAA CAT AGT TAT TTAT T	1691	GAA ATT AAT TAT TCA GAT GAG TTC CTA AAA GTT TTT TTT TCA GGG TTC TCC TGG AAA GTA GAA AAA AAA TAT AAC TCT ATA Glu Ile Amn Tyr Ser Amp Glu Phe Leu Lys Wal Phe Phe Ser Gly Phe Phe Ser Lys Wal Glu Lys Lys Tyr Amn Ser Ile	162
1853 GAT ATT TOT GOT GAA TTA AGA ACG AAT COG ATG ATT TTG AAA AAA GAA CTA GAG TCT CGA GGA GTA AAG TTA AGA ATA AP 211   1853 GAT ATA TOT GOT GAA TTA AGA ACG AAT COG ATG ATT TTG AAA AAA GAA CTA GAG TCT CGA GGA GTA AAG TTA AGA TTA AP 211   1934 ATT AAT AGC ATT CGA ATA TCA ATA TCA CTA ATT CAA TA ATG AAA ACG GOT GAA TTC AAA ATA AAA CGA ATT GCA TAT CAA TCA TTC TTG TTG GIA FDE AFG GIL Lew 211   1934 ATT AAT AGC ATT CGA ATA TCA ATA TCA CTA ATT CAA TA ATG AAA ACG GOT GAA TTC AAA ATA AAA CGA ATT GCA TAT CAA TCA TTC TTG GIA TTG ATG TCA TCA TTT TTT ATT TTT TCT ACA GTA TTT AAG CAA ACC ATG GAA TCA AAT CAA ATA TTA ATT ATA TTT ATG TTG GIY Phe Ala Ser Val Ser Tyr Phe Ser Thr Val Phe Lye Ser Thr Net Aen Val Ala Pro Ser Glu Tyr Lew Phe Met Lew 271   2096 ACA GGA GTT CCA GAA AAA TCA ATC GTT TAC TTG ATC TATA TTT TCG ATT GTA TCA CTG ATT TTG GOT AGT TTT AGT TTGPJ   2096 ACA GGA GTT GCA GAA AAA TA ATC GAT AC GTT TAC TTG ATC TAT TTT CATA CTG ATT TTG GOT AGT TTT AGT TAC TTG GUY Val Ala Glu Lye ** * TTT TCG ATC GAT TAG GUA TAC GTT TAC TTG ATC CTA TTT TCG ATT GTA TCA CTG ATT TTG GOT AGT TTT AGT TAC TTG GUY Val Ala Glu Lye ** * 271   2096 ACA GGA GTT GCA GAA AAA TA ATC GAT TCC TTT TAT TTG ATT TAG TTT TCG ATT GTA TCG CTG TTT TTG GOT AGT TTT AGT TTG GUY Val ALA GLU TYY Val TYY Lew LIE Lew Phe Ser TLE Val Ser Lew UIE Lew UI Ser Lew UI Lew GLU Ser Phe Ser 24	1772	TIT AIT ACT GAT GAT CIT GAT GCT ATG GAG AAA ATT TCA TGT TIA GTA AAA MGT GAT ATT ACG CGT AAT TGG CGT TGG GGA Phe Ile Thr Aep Aep Leu Aep Ale Met Glu Lys Ile Ser Cys Leu Val Lys Ser Aep Ile Thr Arg Aen Trp Arg Trp Ale	189
1934 ATT RAT NGC ATT CGA ATA TCA TAT TCA CTA ATG AMA ACC GOT GAA TTC AMA ATA AMA CAG ATT GCA TAT CAG TCT Ile Asen Ser Ile Arg Ile Ser Tyr Ser Ile Ser Lee Met Lye Thr Gly Glu Phe Lye Sin Ile Ale Tyr Cin Ser 24.   2015 GOG TTT GCT AGC GTT TCA TAT TTT TCT ACA GTA TTT AMG TCA ACC ATG AMA GTA GCA GCA GTA GTA TAT TAT TAT TTA Gly Phe Ale Ser Val Ser Tyr Phe Ser Thr Val Phe Lye Ser Thr Wat Asen Val Ale Pro Ser Glu Tyr Lee Phe Het Leeu 700 271   2056 ACA GGA GTT GCA GAA ATA ATC GTT TAC TTG ATC CTA TTT TCG ATT GTA TCT TTG GOT AGT TTT AGT Thr Gly Val Ala Glu Lye ** 271   2056 MCA GGA GTT GCA GAA ATA ATC GTT TAC TTG ATC CTA TTT TCG ATT GTA TCA TTG GOT AGT TTT AGT Thr Gly Val Ala Glu Lye ** 271   2057 MCA GGA GTT GCA GAA ATA TCA TCG TT TAC TTG ATC CTA TTT TGG ATT GTA TCA TTG GOT AGT TTT AGT Thr Gly Val Ala Glu Lye ** 271	1853	GAT ATT TOT GOT GAA TTA AGA ACG AAT COG ATG ATT TTG AAA AAA GAA CTA GAG TCT CGA GGA GTA AAG TTT AGA GAA TTA Amp ile cys gly glu Leu Arg Thr Amn Arg Met ile Leu Lys Lys Glu Leu glu Ser Arg gly Val Lys Phe Arg Glu Leu	216
2015 00G TTT GCT AGC GTT TCA TAT TTT ACA GTA TTT AMG TCA ACC AG AAT GTA GCA CCA AGT GAA TAT TTA TTT ATG TTG Gly Phe Ala Ser Val Ser Tyr Phe Ser Thr Val Phe Lye Ser Thr Met Aen Val Ala Pro Ser Glu Tyr Leu Phe Met Leu TCpJ 2096 ACA <u>GGA G</u> TT GCA GAA AAA TA ATG GAA TAC GTT TAC TTG ATT TTG GAT TTT GG ATT TTG GGT AGT TTT AGT Thr Gly Val Ala Glu Lye ** * Met Glu Tyr Val Tyr Leu Ile Leu Phe Ser Ile Val Ser Leu Ile Leu Gly Ser Phe Ser 2007	1934	ATT ANT MGC ATT CGA ATA TCA TAT TCA ATT TCA CTA ATG ANA ACC GGT GAN TTC ANA ATA ANA CAG ATT GCA TAT CAG TCT Ile Ann Ber Ile Arg Ile Ser Tyr Ser Ile Ser Law Met Lys Thr Gly Glu Phe Lys Ile Lys Gin 11e Als Tyr Gin Ser	243
TCpJ 2096 ACA <u>GOA O</u> TT GCA GAA AAA TA ATG GAA TAC GTT TAC TTG ATC CTA TTT TCG ATT GTA TCA CTG ATT TTG GOT AGT TTT AGT Thr Gly Vel Ale Glu Lys ** * Met Glu Tyr Vel Tyr Leu Ile Leu Phe Ser Ile Vel Ser Leu Ile Leu Gly Ser Phe Ser 21			
	2015	due fir der met ern tan far fir fer and era fir and fea ace are ant era den den en ant ena far fin fir fie are	270

FIG. 2. Nucleotide sequences of tcpF and toxT. Nucleotide 1 is the first base of the Sau3A-BamHI fused restriction site between V. cholerae and pBR327 DNAs. The ORFs representing tcpF, toxT, and the first several residues of tcpJ are shown. The sequence presented is for V. cholerae 569B. The final nucleotides of the tcpE coding sequence are left untranslated in the beginning part of the sequence. Putative Shine-Dalgarno ribosome binding sites are underlined.

their similarity to the arabinose operon regulator from *E. coli* and *Salmonella typhimurium* (Fig. 3) (2, 3). The greatest similarities in this family were with FapR, a regulator of the 987P pilus operon from *E. coli* (9) and VirF, a regulator controlling plasmid-encoded virulence genes in the yersiniae (2). ToxT is 19.5% identical over a 256-amino-acid overlap with FapR and is 19.2% identical over a 224-amino-acid overlap with VirF. AraC-like proteins are distinguished primarily by a conserved domain at their C termini which harbors a helix-turn-helix motif typical of DNA-binding proteins (boxed in Fig. 3) (24). Their N termini are nonsimilar and may play a role in specific interactions with effector molecules (20).

Analysis of the other ORFs encoded within this region demonstrated that toxT resides 210 nucleotides downstream of the tcpF gene and its stop codon overlaps the start codon for the tcpJ gene. These genes are required for biogenesis of the TCP. TcpJ is a peptidase responsible for processing the TcpA structural subunit (8), and the function of TcpF is not yet characterized, but TcpF is encoded directly downstream of the gene encoding TcpE; the DNA encoding the last 16 codons for tcpE are shown untranslated upstream of tcpF in Fig. 2. TcpE shows similarity to secretion determinants from different genera of bacteria (22).

The less stringent environmental modulation of virulence gene expression observed in strain 569B compared with that of more-typical strains was postulated to be due to differences within the toxT gene or its expression (4). toxTcontaining clones were therefore isolated from a  $\lambda$  library of *V. cholerae* O395. DNA sequence analysis of one of these clones showed that the sequences of tcpF and toxT and their intergenic region were the same as for strain 569B. Therefore, if the phenotypic differences between 569B and O395 are the result of an alteration in toxT expression, the relevant genetic change resides elsewhere. To date, we have no data that would explain these differences.

Northern blot analysis. Previous work suggested that toxT expression is normally ToxR dependent. This conclusion was drawn from the observation that expression of the insert cloned in pGJ40 from the *tet* promoter was sufficient to restore tcpA and ctxAB expression to a toxR mutant (6). Several lines of evidence point to toxT and not tcpF or tcpJ

	1				50
ToxT	MIGKKSF	QTNVYRMSKF	DTYIFNNLYI	NDYKMF	WIDSGI
FapR		MKLKNI	HLYNYVVIYT	KNCEIYINKG	NEQVYIPPRM
AraC/Ecoli		MAEAQNDPLL	PGYSFNAHLV	AGLTPIEANG	YLDFFIDRPL
AraC/Salty		MAETQNDPLL	PGYSFNAHLV	AGLTPIEANG	YLDFFIDRPL
VirF			M	ASLEIIKLEW	ATPIFKVVEH
CelD	MMOPVINAPE	IATAREOOLF	NGKNFHVFIY	NKTESISGLH	OHDYYE
	_				
	51				100
TOXT	AKL	IDKNCLVSYE	INSSS	IILLKKNAIO	RFSLTSLSDE
FapR	VAI	FEKNISFNIE	TIRKGD	GVLYESFDMK	HELLTSLRRV
AraC/Ecoli	GMKGYILNLT	IRGOGVVKNO	GREFVCRPGD	ILLFPPGEIH	HYGRHPEARE
AraC/Saltv	GMKGYILNLT	IRGEGVINNN	GEOFVCRPGD	ILLFPPGEIH	HYGRHPDASE
VirF	SODG. LYIL	LOGOISWONS	SOTYDLDEGN	MLFLRRGS	. YAVRCGTKE
CelD	FTLV	LTGRYFOEIN	GKRVLLERGD	FVFIPLGSHH	OSFYEFGATR
					•
	101				150
ToxT	NINVSVITIS	DSFIRS	LKSYILGDLM	IRNLYSENKD	LLLWNCEHND
FapR	. IEPSVKFAA	ESYTNKRS	FK	EKR	IFVKSC
AraC/Ecoli	WYHQWVYFRP	RAYWHEWLNW	PS.IFANTGF	FRPDEAHQPH	FSDLF
AraC/Salty	WYHQWVYFRP	RAYWQEWLTW	PT.IFAQTGF	FRPDEAROPH	FSELF
VirF	PCQLLWIPLP	GSFLSTFLHR	FG.SLLSE	IRRDNATPKP	LLIFNISPIL
CelD	ILNVGISK	RFFEQHYLPL	LPYCFVASQV	YRTNNA	
	151				200
ToxT	IAVLSEVVNG	FREINYSDEF	L.KVF.FSGF	FSKVEKK	YNSIFITDDL
FapR	.SIVIDLFKR	LKD.NGSPEF	T.AIYELAFL	VSKCENPSMF	AISLFSSVAV
AraC/Ecoli	GQIINAG	QGEGRYSELL	AINLLEQLLL	RRMEAI	NESLHPPMDN
AraC/Salty	GQIISAG	QGEGRYSELL	AINLLEQLLL	RRMAVI	NESLHPPMDS
VirF	SQSIQNLCAI	LERSDFPSVL	TQLRIEELLL	LLAFSSQGAL	FLSALRHLGN
CelD	FLTYVETV	ISSLNFRETG	LEEFVEMVTF	YVINRLRHYR	EEQVIDDVPQ
	201				250
ToxT	DAMEKISCLV	KSDITRNWRW	ADICGELRTN	RMILKKELES	R.GVKFRELI
FapR	TFSERIVTLL	FSDLTRKWKL	SDIAEEMHIS	EISVRKRLEQ	E.CLNFNQLI
AraC/Ecoli	RVREACQYIS	DHLADSNFDI	ASVAQHVCLS	PSRLSHLFRQ	QLGISVLSWR
AraC/Salty	RVRDACQYIS	DHLADSHFDI	ASVAQHVCLS	PSRLSHLFRQ	QLGISVLSWR
VirF	RPEERLOKFM	EENYLQGWKL	SKFAREFGMG	LTTFKELFGT	VYGISPRAWI
CelD	WLKSTVEKMH	DKEQFSESAL	ENMVALSAKS	QEYLTRATOR	YYGKTPMQII
	251				300
ToxT	NSIRISYSIS	LMKTGEFK I	K QIAYQSGFA	S VSYFSTVFK	S TMNVAP SEYL
FapR	LDVRMNQAAK	FIIRSDHQ	G MIASLVGYT	S VSYFIKTFK	E YYGVTP KKFE
AraC/Ecoli	EDORISOAKL	LLSTTRMP I	A TVGRNVGFD	D QLYFSRVFK	K CTGASP SEFR
AraC/Salty	EDQRISQAKL	LLSTTRMP I	A TVGRNVGFD	D QLYFSRVFK	K CTGASP SEFR
VirF	SERRILYAHQ	LLLNGKMS I	V DIAMEAGES	S QSYFTQSYRI	R RFGCTP   SQAR
CelD	NEIRINFAKK	QLEMTNYS V	T DIAFEAGYS:	S PSLFIKTFK	K LTSFTP KSYR
	301	315			
ToxT	FMLTGVAEK.				
FapR	IGIKENL	RCNR.			
AraC/Ecoli	AGCEEKVNDV	AVKLS			
AraC/Salty	AGCE	• • • • •			
VirF	LTKIATTG	••••			
CelD	KKLTEFNQ	••••			

#### 'Helix-Turn-Helix' Motif Consensus: [L,I,V]-x2-[L,I,V]-x4-[G]-[I,F,Y]-x5-[F]-x3-[F,Y]-x7-[P]

FIG. 3. Alignment of ToxT and other AraC family members. AraC-like proteins from other bacterial species were aligned by using the PILEUP program in the University of Wisconsin Genetics Computer Group (GCG) package. The helix-turn-helix sequence was generated by using the MOTIFS program in the GCG package. Sequences shown represent AraC proteins from *E. coli* and *S. typhimurium* (2), VirF from *Yersinia enterocolitica* (3), FapR from *E. coli* (9), and CelD (17), a negative regulator of the *cel* operon in *E. coli*.

as being responsible for this phenotype. First, the two *tcp* gene products are predicted to be periplasmically located (8, 22) with no detectable similarity to other transcriptional activators (9), while ToxT shows strong similarity to a family of prokaryotic regulatory gene products. Second, Tn*phoA* insertions within *tcpF* in *V. cholerae* do not abolish TcpA production (22). If *tcpF* were a component of ToxT activity, such mutations would be expected to abolish TcpA synthesis. Third, a frameshift mutation in *toxT* completely abolishes ToxT activity, whereas a frameshift mutation within the *tcpF* gene does not.

The pattern of toxT expression in V. cholerae was determined by Northern blot analysis. RNA was isolated from overnight cultures of wild-type or toxR mutant V. cholerae grown in medium with a pH of 6.5 at 30°C, which favors expression of ToxR-regulated genes (14). In wild-type strain O395, a transcript of approximately 1,400 nucleotides was detected with the tcpEF probe, whereas no detectable message was detected in the toxR mutant JJM43 with the same probe (Fig. 4). When a toxT probe of equivalent specific activity was used to analyze the same batch of RNA, a diffuse and relatively weaker signal beginning at a slightly smaller size than that seen with the *tcpEF* probe was observed in O395 RNA, and only after the *tcpEF* specific signal had been overexposed (Fig. 4). As with *tcpEF*, no RNA was detected with the *toxT* probe in the *toxR* mutant strain (Fig. 4). When wild-type cells were grown at pH 8.5, a condition that down-regulates ToxR-dependent gene expression, no signal was detected with either probe (data not shown). The patterns of hybridization demonstrated in Fig. 4 were consistently observed when other probes encompassing this region were used and when different batches of RNA were analyzed.

### DISCUSSION

Previous work on the ToxR system of virulence gene expression in V. *cholerae* led to a model in which hierarchical expression of activators controls a number of genes (4-6). Data presented in this report support this model by



FIG. 4. Northern blot analysis of V. cholerae. RNA was isolated after overnight growth at 30°C in Luria-Bertani medium with a starting pH of 6.5, and Northern analysis was done as described in Materials and Methods. Restriction fragments corresponding to tcpEF or toxT are indicated. Strain JJM43 is a toxR null mutant. The location of the 23 and 16S rRNA subunits are shown and were identified by ethidium staining lanes with equal amounts of RNA loaded. These RNAs migrate in these gels to the same location as do rRNAs from *E. coli*, which are 2,904 and 1,541 nucleotides long, respectively.

demonstrating that ToxR, itself a member of the OmpR family of activators, is required for expression of toxT, which is demonstrated here to be similar to AraC-like activators identified in other genera of bacteria. AraC-like activators are similar to one another in having a C-terminal helix-turn-helix domain typical of DNA-binding proteins, but N termini within the family are dissimilar. This latter domain may function as a site of effector interaction (20), so perhaps ToxT is a sensor of some intracellular signal generated under conditions in which the ToxR regulon is to be expressed.

The finding that toxT resides within the cluster of genes encoding the biogenesis functions of the TCP is reminiscent of the structure of the 987P pilus gene cluster in *E. coli*, in which the gene encoding the AraC-like regulator FapR is located adjacent to the pilus cluster (9). This finding also adds a layer of complexity to understanding how coordinate control of virulence genes works in *V. cholerae*. While the location of toxT within the *tcp* cluster may suggest that the toxT product might have evolved as a regulator specific to TCP production, ToxT can activate other ToxR-regulated genes not involved in TCP biogenesis (6) and is thus a global regulator, like VirF in yersiniae (3).

Regulated expression of toxT is apparently an important step in the signal transduction pathway controlled by ToxR (6). This conclusion is based on the observations that introJ. BACTERIOL.

duction of toxT under constitutive control of the *tet* promoter into a ToxR mutant of *V. cholerae* restored expression of both TcpA and CtxB and that this expression was independent of pH, which modulates ToxR-regulated gene expression in wild-type cells (6). Northern blots in that study were done with the large *Eco*RV fragment of pGJ40 beginning in pBR327 DNA and proceeding past toxT (and tcpJ; Fig. 1) or an *Eco*RI fragment beginning in pBR327 and ending within the toxT coding sequence (73 nucleotides downstream of the *XhoI* site in Fig. 1) used as a probe. The abundant transcript identified in that tcpEF probe, but, while less abundant, toxT mRNA is subject to the same ToxR-dependent pH modulation. Thus ToxR is responsible for pH-sensitive expression of toxT.

How this control is accomplished is not clear. ToxR activation of ctxAB transcription requires the ability of ToxR to bind to a tandemly repeated heptamer, TTTTGAT (15), but this sequence is not present upstream of toxT (Fig. 2), nor is it found within 110 nucleotides upstream of tcpE, which is directly upstream of tcpF (22). A sequence within the intergenic region between tcpF and toxT is identified by computer analysis as a putative prokaryotic promoter element, but whether this actually functions as a promoter is not certain. Since insertions of TnphoA into the tet promoter of pGJ40 abolish toxT activity, it follows that tcpF and toxTare (or at least can be) cotranscribed, although it has not been ruled out that toxT is controlled by its own promoter. Given that the stop codon for tcpE is only 9 nucleotides upstream of the start codon for tcpF, these two genes may be transcribed as a polycistronic mRNA (22). Similarly, toxTand tcpJ overlap one another's coding sequence (Fig. 2), suggesting that they may be cotranscribed. A polycistronic tcpEF transcript would have a predicted size of approximately 2,100 nucleotides, and a polycistronic toxT-tcpJ transcript would have a size of approximately 1,650 nucleotides. The sizes of the mRNAs we observed in this study are inconsistent with these predicted sizes, but since the analysis presented here was done with mRNA isolated from overnight cultures, we cannot rule out the possibility that there is processing of larger ToxR-regulated transcripts within this region.

The DNA between tcpF and toxT contains extensive inverted repeats (Fig. 5A) which would form an unusual structure with a high negative free energy (Fig. 5B). The function of this structure or whether it forms at all is unknown. While it bears certain features typical of type I transcriptional terminators (7), it is different from them in other, perhaps critical, ways. As is often seen in type I terminators, one of the predicted stem-loops ends in a run of thymine residues; however, this stem-loop is A-T rich rather than G-C rich, giving it an overall low free energy. Whether this structure acts as a transcription terminator or plays some other role in RNA metabolism, as an RNase processing site, for instance, is under investigation.

We envision three explanations, which are not exclusive of one another, that may account for some of the observations regarding expression of tcpF (and probably tcpEF) and taxT presented here. First, the taxT expression from pGJ40 that we observe may simply be the result of having tcpFtaxT cloned downstream of a strong, constitutive promoter, and perhaps taxT has its own promoter as well. Second, ToxR-regulated transcription may begin somewhere upstream of tcpF and terminate predominantly after tcpF prior to taxT (perhaps as a function of the stem-loop structure discussed above), but a low percentage of transcripts might



FIG. 5. Nucleotide sequence of the tcpF-toxT intergenic region. (A) Sequence and locations of inverted repeat sequences between the tcpF and toxT genes. Codons for tcpF and toxT are shown in uppercase letters. Asterisks represent bases that are unpaired in the corresponding stem-loop. (B) Predicted stem-loop structure (length, 98 bp; energy, -33.7 kcal/mol) formed from the sequence depicted in A, generated by the FOLD program in the University of Wisconsin GCG package. Numbering of nucleotides is as in Fig. 2. Asterisks indicate stop codon of tcpF.

read through to toxT. The absolute amount of toxT expression would then increase as the expression of the upstream promoter was elevated. This sort of read-through control of a transcriptional regulator has been proposed to control expression of the kdpDE regulatory genes from E. coli (19) and the prfA gene from the gram-positive pathogen Listeria monocytogenes (10). Although there are analogies with the toxT system to be drawn, the RNA species identified by both tcpEF and toxT probes are not the size expected for a read-through transcript encoding them both. Third, portions of a larger ToxR-regulated precursor transcript may exhibit differential stabilities. A stem-loop between lacZ and lacY in E. coli has been proposed to protect the lacZ mRNA from 3' exonucleolytic degradation that removes the downstream portion of the lacZYA transcript (16). Differential stability of portions of a polycistronic mRNA has also been demonstrated for the pap pilus genes in E. coli, in which mRNA for a regulatory gene, papB, is encoded on the same transcript as the papA structural gene but is maintained in a lower concentration by rapid decay (1). In this system, a stem-loop structure downstream of *papA* has been proposed to act as a barrier to decay of the papA transcript (1).

The central issue that arises from all of these possibilities is the precise role played by ToxR, which has not been demonstrated to activate any of the *tcp* genes. Thus, its requirement for regulated expression of toxT in the activator cascade model remains unexplained. Experiments are under way to investigate the various possibilities presented above, and results from them will undoubtedly help elucidate the role of ToxR in toxT expression.

## ACKNOWLEDGMENTS

We thank Ron Taylor and Paul Manning for communicating results prior to publication, Ken Peterson for supplying the  $\lambda$  library of strain O395, and John Mekalanos for helpful discussions and encouragement. We also thank David Friedman and Mike Koomey for reviewing the manuscript.

This work was supported by Public Health Service grants AI-31645 and RR-00200 from the National Institutes of Health. D.E.H. is a recipient of a Rackham Merit Fellowship from the University of Michigan.

## ADDENDUM IN PROOF

The toxT gene is the same as the tcpN gene, whose sequence was recently presented by Ogierman and Manning (Gene 116:93–97, 1992).

#### REFERENCES

- 1. Baga, M., M. Goransson, S. Normark, and B. E. Uhlin. 1988. Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. Cell 52:197-206.
- Burke, K. A., and G. Wilcox. 1987. The araC gene of Citrobacter freundii. Gene 61:243–252.
- Cornelis, G., C. Sluiters, C. Lambert de Rouvroit, and T. Michiels. 1989. Homology between VirF, the transcriptional activator of the Yersinia virulence regulon, and AraC, the Escherichia coli arabinose operon regulator. J. Bacteriol. 171: 254-262.
- DiRita, V. J. 1992. Coordinate control of virulence gene expression by ToxR in Vibrio cholerae. Mol. Microbiol. 6:451–458.
- 5. DiRita, V. J., and J. J. Mekalanos. 1991. Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS, results in signal transduction and transcriptional activation. Cell 64:29–37.
- DiRita, V. J., C. Parsot, G. Jander, and J. J. Mekalanos. 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. Proc. Natl. Acad. Sci. USA 88:5403–5407.
- Friedman, D. I., M. J. Imperiale, and S. L. Adhya. 1987. RNA 3' end formation in the control of gene expression. Annu. Rev. Genet. 21:453–488.
- Kaufman, M. R., J. M. Seyer, and R. K. Taylor. 1991. Processing of TCP pilin by TcpJ typifies a common step intrinsic to a newly recognized pathway of extracellular protein secretion by gram-negative bacteria. Genes Dev. 5:1834–1846.
- Klaasen, P., and F. K. de Graaf. 1990. Characterization of FapR, a positive regulator of expression of the 987P operon in enterotoxigenic Escherichia coli. Mol. Microbiol. 4:1779–1783.
- Leimeister-Wächter, M., E. Domann, and T. Chakraborty. 1992. The expression of virulence genes in *Listeria monocytogenes* is thermoregulated. J. Bacteriol. 174:947–952.
- 11. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Miller, V. L., V. J. DiRita, and J. J. Mekalanos. 1989. Identification of taxS, a regulatory gene whose product enhances ToxR-mediated activation of the cholera toxin promoter. J. Bacteriol. 171:1288-1293.
- Miller, V. L., and J. J. Mekalanos. 1984. Synthesis of cholera toxin is positively regulated at the transcriptional level by toxR. Proc. Natl. Acad. Sci. USA 81:3471-3475.
- Miller, V. L., and J. J. Mekalanos. 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J. Bacteriol. 170:2575–2583.
- Miller, V. L., R. K. Taylor, and J. J. Mekalanos. 1987. Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. Cell 48:271–279.
- Murakawa, G. J., C. Kwan, J. Yamashita, and D. P. Nierlich. 1991. Transcription and decay of the *lac* messenger: role of an intergenic terminator. J. Bacteriol. 173:28–36.
- 17. Parker, L. L., and B. G. Hall. 1990. Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. Genetics 124:455-471.
- Peterson, K. M., and J. J. Mekalanos. 1988. Characterization of the Vibrio cholerae ToxR regulon: identification of novel genes involved in intestinal colonization. Infect. Immun. 56:2822– 2829.

- Polarek, J. W., G. Williams, and W. Epstein. 1992. The products of the kdpDE operon are required for expression of the Kdp ATPase of Escherichia coli. J. Bacteriol. 174:2145–2151.
- Ramos, J. L., F. Rojo, L. Zhou, and K. N. Timmis. 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. Nucleic Acids Res. 18:2149–2152.
- 21. Stock, J. B., A. J. Ninfa, and A. M. Stock. 1989. Protein phosphorylation and regulation of adaptive responses in bacte-

ria. Microbiol. Rev. 53:450-490.

- 22. Taylor, R. K. Personal communication.
- 23. Taylor, R. K., V. L. Miller, D. B. Furlong, and J. J. Mekalanos. 1987. Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with the cholera toxin. Proc. Natl. Acad. Sci. USA 84:2833-2837.
- Wu, J., and B. Weiss. 1991. Two divergently transcribed genes, soxR and soxS, control a superoxide response regulon of *Escherichia coli*. J. Bacteriol. 173:2864–2871.