(1)

AD_____

USE OF RECOMBINANT DNA TECHNIQUES FOR THE PRODUCTION OF A MORE EFFECTIVE ANTHRAX VACCINE

FINAL REPORT

Donald L. Robertson

August 31, 1988

Supported by

U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-85-C-5167

Brigham Young University
Provo, Utah 84602

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Brigham Young University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): Provo, Utah 84602	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U.S. Army Medical Research & Development Comman; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAMD17-85-C-5167
8c. ADDRESS (City, State, and ZIP Code): Ft. Detrick, Frederick, MD 21701-5012	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
63763A	3M2-637630807	AA	041

11. TITLE (Include Security Classification)

(u) Use of Recombinant DNA Techniques for the Production of a More Effective Anthrax Vaccine

12. PERSONAL AUTHOR(S): Donald L. Robertson, Ph.D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 6/1/85 TO 7/31/88	1988 Aug. 31	33

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
06	13		RA 1,
06	01		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

During the course of this contract, we have isolated and characterized each of the Bacillus anthracis toxin genes. Although the PA (pag) gene was cloned and sequenced by researchers in the Bacteriology Division of USAMRIID, the cloning and characterization of the EF (cya) and LF (lef) genes were performed in my laboratory (7,8). In addition, DNA sequence determinations for the cya (9) and lef (unpublished data of author) genes have also been completed in my laboratory.

We have prepared an improved method for the isolation of large quantities of pXO1 and pXO2 from B. anthracis strains. Restriction enzyme cleavage maps for these plasmids have been constructed. We have initiated mutagenesis procedures for the modification of each of the toxin genes and these mutants are being tested for biochemical activity. In addition, wild-type and mutant toxin genes are being inserted into B. subtilis to produce larger quantities

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL: Mary Frances Bostian	22b. TELEPHONE (Include Area Code): (301) 663-7325	22c. OFFICE SYMBOL: SGRD-RMI-S

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

of these proteins for biochemical purposes and for vaccine testing. However, we have not yet placed mutant toxin genes back into B. anthracis, although the wild-type PA and EF genes have been transferred.

In conclusion, it appears that most of the experiments outlined in the original research proposal are completed. (i) The anthrax toxin genes have each been cloned. (ii) Each of the toxin genes have been sequenced. (iii) We have generated some toxin gene mutants to be used in the construction of a safe vaccine. We are also using mutants to help elucidate the biochemical activities of these proteins. (iv) We have expressed the anthrax toxin genes in E. coli and B. subtilis and have constructed expression vectors for B. subtilis and B. anthracis. These recombinant plasmids should allow for high level expression of the toxin proteins for biochemical and immunological purposes. (v) We have identified conserved amino acid homolgy between EF and the Bordetella pertussis calmodulin-dependent adenylate cyclase. This relationship should help to characterize EF better. (vi) Homologies which exist between LF and EF should allow us to examine the interaction between these proteins and PA.

Overall, the research performed under this contract has allowed us to characterize the anthrax toxin genes and to construct important gene mutants. This research is absolutely required for the construction of a safe recombinant DNA derived anthrax vaccine.

SUMMARY OF RESEARCH

Research Goals. The overall goal of the present research is to construct a safe and effective human anthrax vaccine using recombinant DNA techniques. We plan to isolate and characterize the *Bacillus anthracis* toxin genes for protective antigen (PA), lethal factor (LF) and edema factor (EF). The individual toxin genes will be cloned and expressed in *E. coli* and *B. subtilis*. The toxin genes will be modified using site-specific mutagenesis or deletion mutagenesis procedures to generate gene mutants which lack biochemical activity but which are still fully immunologic for use in a recombinant vaccine. These mutant genes can then be inserted back into *B. anthracis* Sterne with the selective removal of wild-type genes. These mutant *B. anthracis* strains will be tested in animals, such as the mouse or guinea pig, for vaccine efficacy.

We will also characterize the *B. anthracis* plasmids pXO1 and pXO2 (1-3). Since we plan to insert the toxin genes back into *B. anthracis* to construct a recombinant vaccine host, we need to know a complete restriction map of pXO1, which contains the toxin genes. In addition, in order to understand the expression of the toxin genes and of the capsule (2-4), we need to physically characterize these plasmids as completely as possible.

Research Achievements. During the course of this contract, we have isolated and characterized each of the *B. anthracis* toxin genes. The PA (*pag*) gene was cloned and initially characterized in the Bacteriology Division of USAMRIID (5). In addition, the DNA sequence for *pag* was also determined by them (6). The cloning and characterization of the EF (*cya*) and LF (*lef*) genes were performed in my laboratory (7,8). The DNA sequences for the *cya* (9) and *lef* (unpublished data of author) genes have also been completed in my laboratory.

An improved method for the isolation of large quantities of pXO1 and pXO2 from *B. anthracis* strains was developed at Brigham Young University (10). Initial restriction enzyme cleavage maps have also been constructed. We have also initiated mutagenesis procedures for the modification of each of the toxin genes. These mutants are currently being tested for biochemical activity. In addition, these gene mutants are being inserted into *B. subtilis* to produce larger quantities of these proteins and for vaccine testing.

FOREWORD

The investigators (Principal Investigator and Graduate Students) have abided by the National Institutes of Health Guidelines for Research Involving Recombinant DNA Molecules (May, 1986). Supplemental guidelines pertaining to the subcloning of the individual *B. anthracis* toxin genes in sporulation competent *B. subtilis* was approved by the NIH committee on toxins March 13, 1986. All recombinant DNA research has also been registered with and approved by the Brigham Young University Institutional Biosafety Committee.

TABLE OF CONTENTS

RESULTS

Isolation and characterization of the edema factor gene (*cya*). The edema factor is a calmodulin-dependent adenylate cyclase (11,12). We have cloned and sequenced the EF gene (*cya*). The DNA sequence and deduced amino acid sequences (9) were reported in previous annual reports and are shown in Appendix I and II. A paper describing the cloning and expression of EF in *E. coli* has been published (8) and a manuscript describing the DNA sequence and its deduced amino acid sequence has been submitted and should soon be accepted by Gene (9).

Several interesting structural features for EF are part of its deduced amino acid sequence. (i) EF apparently contains a 33 amino acid signal peptide which conforms to known *Bacillus* leader sequences in that it starts with charged (mostly positive) and hydrophilic residues (amino acids 1-10), followed by a central core of hydrophobic amino acids (residues 11-23) and then several hydrophilic residues (amino acids 24-33) prior to the start of the mature protein. Proteolytic cleavage apparently occurs at an Ala-Met peptide bond, near the start of a proposed α-helix (see Figure 4A), consistent with signal processing after an Ala or Gly in bacilli (13). PA apparently contains a 29 amino acid leader sequence (6) and LF appears to contain a 33 amino acid leader (see below). Figure 4B shows a comparison between the amino acid sequences near the ends of the EF, PA and LF signal peptides and the apparent position of proteolytic cleavage. Similar amino acids at the ends of these signal peptides may be required for signal peptidase recognition or for secretion. (ii) A very strong *Bacillus* ribosome binding site immediately upstream from the start of the EF protein coding region is present (AAAGGAGGT) which is similar to the identical PA and LF ribosome binding sites (AAAGGAG). (iii) Amino acid residues 347 to 355 of the EF-precursor

protein contains the sequence Gly-x-x-x-x-Gly-Lys-Ser (where x=any amino acid) which is a perfect match to a consensus sequence present in prokaryotic and eukaryotic ATP and GTP binding proteins (14). The Lys residue is part of the ATP binding site of these proteins and appears to be part of the EF ATP binding site as well. That is, using site-specific mutagenesis procedures, we have replaced this Lys within EF with an Asn and cyclase activity was reduced 90-95% (unpublished data of author). (iv) We have also identified a domain in EF which could represent its putative calmodulin-binding site. As described in the EF sequencing paper (9), calmodulin-binding proteins often contain an α-helical region with charged or hydrophilic residues on one side and hydrophobic residues on the other. Such an amphiphilic helical region is present in EF located between amino acid residues 313-323 of the EF-precursor (see Appendix II). (v) No homology between the EF gene or its deduced EF amino acid sequence was observed with either the *E. coli* or yeast adenylate cyclases. However, there is at least three regions of homology in the amino acid sequence between EF and the *B. pertussis* calmodulin-dependent adenylate cyclase. The putative calmodulin-binding site, identified above, is conserved in the *B. pertussis* adenylate cyclase as well (15,16).

As mentioned above, we have also compared the EF amino acid sequence with the calmodulin-dependent adenylate cyclase of *Bordetella pertussis*, the causative agent of whopping cough. The pertussis cyclase appears to function independently of the pertussis toxin, but is a required virulence factor since strains which lack cyclase activity are avirulent (17). Glaser et al. (16) recently showed that the cyclase catalytic domain is about 450 amino acids in length and is part of a larger precursor polypeptide of 1706 amino acids. We performed a homology search between the entire EF (800 amino acids) and pertussis cyclase (1706 amino acids). Three major

regions of homology (labeled #1, #2 and #3 in Appendix III) were observed. These homologous domains are part of the catalytic domain of the pertussis cyclase and are located within the carboxyl terminal 500 amino acids of EF. Domain #1 contains the consensus ATP binding site which is surrounded by highly conserved amino acids. This high degree of amino acid conservation indicates a close evolutionary relatedness for these two proteins. The putative calmodulin-binding site is conserved for these proteins and is shown in Appendix II and III.

Characterization of the LF gene (*lef*). We have also cloned the *B. anthracis* LF gene (*lef*) and have determined its entire DNA sequence. We easily identified the start of the LF gene since the first 15 amino acids of the mature LF was previously determined by Dr. J. Schmidt (USAMRIID). The LF DNA sequence and the deduced amino acid sequence are shown in Appendix IV. The LF gene contains a good ribosome binding site (AAAGGAG) which is identical to the proposed PA gene ribosome binding site. The LF-precursor apparently contains a 33 amino acid signal sequence (see Figure 4A) which is removed during secretion. This signal sequence conforms to consensus *Bacillus* leader peptides (and to the EF and PA signal peptides) in that it starts with a polar or charged region followed by 23 non-polar, hydrophobic amino acid residues. After this 33 amino acid leader peptide, the next 16 amino acids correspond exactly to the LF amino acid sequence determined by Dr. Jim Schmidt (USAMRIID), except for one amino acid. Amino acid position +10 of the mature protein (+43 of LF-precursor) is a His (based on the DNA sequence) whereas it was previously reported to be a Lys (based on LF protein sequencing). Interestingly, there is a single Cys in the LF leader, although no Cys residues are in the mature protein. The entire protein sequence of LF is also shown in Appendix V.

There appears to be extensive amino acid homology between LF and EF in the first 300 amino acids of each protein. We have detected 10 closely related domains and three of these highly conserved domains are underlined (labelled #1, #2 and #3) in Appendix II and Appendix V. These homologous regions could represent domains which are required for association with PA prior to cellular uptake. Since these conserved domains in LF and EF are charged, interactions with PA may occur through a series of electrostatic interactions.

Mutagenesis of the anthrax toxin genes. Using site-specific mutagenesis procedures, we have altered the EF gene in order to modify its enzyme activity and to construct EF expression vectors. First, the previously identified ATP binding domain in EF, which conforms to the consensus ATP binding site of other prokaryotic and eukaryotic ATP and GTP binding proteins (14), has a Lys residue which is involved in ATP binding. This amino acid was changed to an Asn in EF. When this mutant EF was isolated from *E. coli*, adenylate cyclase activity was reduced about 90-95% indicating that this Lys is probably involved in ATP binding. However, since total activity was not abolished, other residues are probably also involved. Of particular interest, is the presence of a His two residues prior to this Lys. This His is also conserved in the *B. pertussis* adenylate cyclase (see the ATP binding domain in Appendix III).

We have also removed a *Bgl*II cleavage site within the EF gene and inserted a new *Bgl*II recognition site immediately prior to the start of the protein coding sequence. In another experiment, we inserted a *Bgl*II cleavage site immediately downstream from the PA promoter so that we could fuse the PA promoter to the EF gene. This hybrid toxin gene, when inserted into pBS42 (18) and transformed into *B. subtilis*, expressed EF at a level

at least as great as *B. anthracis* Sterne. We are in the process of determining the precise amount produced using an ELISA or Western blot. EF was secreted from *B. subtilis* and was enzymatically active in an adenylate cyclase assay. Since PA expression is regulated by bicarbonate (19) in *B. anthracis* (Dr. J. Bartkus, USAMRIID, personal communication), we are attempting to transfer this PA promoter-EF gene plasmid into *B. anthracis* by electroporation. Hopefully, this plasmid, when introduced into *B. anthracis*, will produce regulated high levels of EF for purification and analysis. EF gene mutants can also be generated and transferred to *B. anthracis* using this plasmid construction.

Several mutagenesis experiments have also been initiated with the PA gene. Since expression of PA in *B. anthracis* appears to be significantly greater than either LF or EF, we are fusing the PA promoter to both the EF and LF genes for higher levels of expression. In addition, we have started experiments to specifically alter PA. Specifically, we are mutating the Arg-Lys-Lys-Arg sequence (Dr. S. Leppla, USAMRIID, personal communication) in PA which is cleaved with a trypsin-like enzyme when bound to its cellular receptor. After cleavage, the amino terminal 20,000 daltons of PA is removed and PA can now bind either LF or EF. Therefore, by preventing cleavage, LF or EF will not bind and cannot enter the cell. We will alter the amino acids at this location in PA to examine specificity of cleavage and to substitute amino acids which could prevent cleavage. These alterations should also prevent the binding of LF or EF and make these toxin components essentially inactive.

Transcription start sites for the anthrax toxin genes. We have used radiolabeled oligonucleotides, specific for each of the different toxin genes, to determine the start site for toxin gene transcription. Using

mRNA (isolated from *B. anthracis* Sterne) as template, each oligonucleotide was used to prime DNA synthesis (using reverse transcriptase) towards the 5'-end of the respective toxin mRNA. This newly synthesized radioactive DNA was denatured and electrophoresed on a denaturing polyacrylamide gel. Using this approach, we have successfully identified the start sites for PA and LF gene transcription. The PA promoter is apparently located immediately upstream from the start of its coding region with transcription starting about 25 bases before the first start codon for PA translation (6). Likewise, the apparent start for LF gene transcription occurs 25 bases prior to the ATG start codon for LF translation (about nucleotide 456 in Appendix IV). We have not yet been able to localize EF gene transcription. This failure is probably due to the low level of EF mRNA produced in *B. anthracis* which is at least 10-fold lower than either the PA or LF mRNA concentrations (unpublished data of author).

Expression of toxin genes in *B. subtilis* and *B. anthracis*. In an effort to express the anthrax toxin genes in *B. subtilis*, we have cloned each of the toxin genes into *B. subtilis* expression plasmids. Initiatlly, we fused these genes to a regulated promoter and a good ribosome binding site which is present in pSI-1 (20). Using site-specific mutagenesis procedures, we have introduced new *Xba*I recognition sites immediately before the start codons for the PA, EF and LF genes. Following cleavage with *Xba*I, each of the toxin genes was ligated into plasmid pSI-1. When transformed into *B. subtilis*, transcription of the inserted toxin genes is regulated by the *lac* repressor and IPTG (18,20). For example, the amount of PA produced by this fusion was close the expression of PA from PA1 (21).

We have also constructed a plasmid using the T7 promoter cloned upstream from the toxin gene. We cloned the T7 RNA polymerase gene (22) into pSI-1

so that transcription would be controlled by the *lac* promoter, which is inducible with IPTG. Part of this recombinant plasmid which contained the T7 polymerase gene and the erythromycin resistance gene from pE194, was integrated into *B. subtilis* genomic DNA (23,24). *B. subtilis* with this DNA should express T7 RNA polymerase after the addition of IPTG. These cells can then be transformed with a replication competent plasmid containing one of the *B. anthracis* toxin genes (e.g., *cya*, *pag*, or *lef*) cloned downstream from the T7 promoter for gene expression. Although we have not yet tested these recombinants in *B. subtilis*, plasmids containing the toxin genes express toxin in *E. coli* using the T7 polymerase (21). *B. subtilis* containing these plasmids should produce high level, regulated expression of the toxin genes in a safe bacterial host. Toxin is secreted from *B. subtilis* and can be used for purification of individual toxin components.

Isolation and characterization of pXO1 and pXO2. We have developed an efficient plasmid isolation procedure to isolate pure supercoiled pXO1 and pXO2 DNA. This procedure involves chromatography using NACS-37 resins and effectively separates small amounts of genomic DNA from plasmid (10). Our purification protocol does not use CsCl bouyant density gradients since these large plasmids are easily sheared, converting them from supercoiled to relaxed or linear DNA. A typical yield of pXO1 from a one liter culture of *B. anthracis* was about 200 μg, which is close to the maximum amount of DNA expected per liter of culture if these plasmids were present as single copies within *B. anthracis* cells.

Using pure pXO1 and pXO2, we characterized these DNAs using thermal denaturation and bouyant density procedures. Using a T_m analysis, the melting temperatures for pXO1 and pXO2 were 82.5°C ± 0.3°C and 82.2°C ± 0.3°C, respectively. These values correspond to GC contents of 32.2% for

pXO1 and 31.5% for pXO2. Similar experiments using CsCl banding gave GC-contents of 31.1% for pXO1 and 31.4% for pXO2. These values are close to the GC% of *B. anthracis* genomic DNA which is 32.2%.

The restriction maps for pXO1 and pXO2 have been determined for several enzymes which cleave a few times, such as *Pst*I, *Bam*HI, *Cla*I, *Sst*I, *Bgl*II and *Pvu*II (Figures 1 and 2). Experiments to map the more frequent cutting enzymes, such as *Eco*RI and *Hin*dIII, are presently being completed. We have generated recombinant DNA libraries for pXO1 and pXO2 in bacteriophage λ as well as in plasmids in order to generate a complete map for the most common restriction enzymes. A detailed restriction enzyme map of the LF and PA gene regions on pXO1 is also shown in Figure 3.

In a final effort to generate a complete gene map of pXO1 and pXO2, we are identifying the number and location of the different RNA transcripts from these plasmids. This project involves the identification of the different promoters and the RNAs made from them. Basically, we are cleaving pXO1 and pXO2 with an enzyme which cleaves these DNAs many times, such as *Mbo*I or *Sau*3A, generating DNA fragments which can ligate to *Bam*HI cleaved plasmids. Using *B. subtilis* plasmids which have been cleaved with *Bam*HI located prior to a promoterless chloramphenicol resistance gene (25), we will insert the pXO1 or pXO2 DNA fragments into these promoter identification plasmids. After transformation of these recombinant plasmids into *B. subtilis*, we will identify bacteria which are now resistant to chloramphenicol. These plasmids will contain a functional promoter (derived from pXO1 or pXO2) driving the transcription of the chloramphenicol resistance gene. The recombinant DNA inserts prepared from these promoter expression plasmids will then be mapped on pXO1 or pXO2. The size and direction of RNA transcription will also be determined. This procedure is very powerful and should allow us to

identify and position most, if not all, of the functional promoters from the *B. anthracis* plasmids, assuming that all these promoters will also function in *B. subtilis*. However, with the recent discovery that we can transform *B. anthracis* using electroporation, we will also be able to transfer these promoter plasmids to *B. anthracis* for promoter identification directly in the parent organism.

CONCLUSIONS

It appears from the data described in this report, that most of the experiments outlined in the original research proposal are essentially completed. (i) The anthrax toxin genes are each cloned. (ii) Each of the toxin genes have also been sequenced. We will be able to study gene expression and to characterize the toxin proteins better. (iii) We can generate toxin gene mutants for the construction of a safe vaccine and to elucidate the biochemical activities of these proteins. (iv) We have expressed the anthrax toxin genes in *E. coli* and *B. subtilis* and have constructed expression vectors, especially for *B. subtilis* and *B. anthracis*, which should allow for high level expression of the toxin proteins for biochemical and immunological purposes. (v) We have determined homology between EF and the pertussis calmodulin-dependent adenylate cyclase which should allow us to better charcterize EF based on conserved domains. In addition homology between LF and EF should allow us to examine the interaction between these proteins and PA. (vi) We have not yet placed mutant toxin genes back into *B. anthracis*, although the wild-type PA and EF genes have been transferred. Overall, our research has allowed us to characterize the anthrax toxin genes and to construct important gene mutants. This research is absolutely

required for the construction of a safe recombinant DNA derived anthrax vaccine.

LITERATURE CITED

1. Mikesell, P., B. E. Ivins, J. D. Ristroph, and T. M. Dreier. 1983. Evidence for plasmid-mediated toxin production in Bacillus anthracis. Infect. Immun. 39:371-376.

2. Green, B.D, L. Battisti, T.M. Koehler, C.B. Thorne, and B.E. Ivins. 1985. Demonstration of a capsule plasmid in *B. anthracis*. Inf. Immun. 49:291-297.

3. Uchida, I., Sekizaki, T., Hashimoto, K., and Terakado, N. 1985. Association of the encapsulation of *Bacillus anthracis* with a 60 megadalton plasmid. J. of Gen. Micro. 131:363-367.

4. Zwartouw, H. T., and H. Smith. 1956. Polyglutamic acid from *Bacillus anthracis* grown in vivo: structure and aggressin activity. Biochem. J. 63:437-442.

5. Vodkin, M. H., and S. H. Leppla. 1983. Cloning of the protective antigen gene of Bacillus anthracis. Cell 34:693-697.

6. Welkos, S.L., Lowe, J.R., Eden-McCuthan, F., Vodkin, M., Leppla, S.H. and J.J. Schmidt, J.J. 1988. Nucleotide sequence of the protective antigen gene of *Bacillus anthracis*. in press, Gene.

7. Robertson, D. L., and S. H. Leppla. 1986. Molecular cloning and expression on *Escherichia coli* of the lethal factor gene of *Bacillus anthracis*. Gene 44:71-78.

8. Tippetts, M.T. and D.L. Robertson. 1988. Molecular cloning and expression of the Bacillus anthracis edema factor toxin gene: a calmodulin-dependent adenylate cyclase. J. Bacteriol. 170:2263-2266.

9. Robertson, D.L., M.T. Tippetts and S.H. Leppla. 1988. Nucleotide sequence of the *Bacillus anthracis* edema factor gene (*cya*): A calmodulin-dependent adenylate cyclase. Submitted to Gene.

10. Kaspar, R.L. and Robertson, D.L. 1987. Purification and physical analysis of *Bacillus anthracis* plasmids pXO1 and pXO2. Biochem. Biophys. Res. Commun. 149:362-368.

11. Leppla, S. H. 1982. Anthrax toxin edema factor: A bacterial adenylate cyclase that increases cyclic AMP concentrations in eukaryotic cells. Proc. Natl. Acad. Sci. USA 79:3162-3166.

12. Leppla, S. H. 1984. *Bacillus anthracis* calmodulin-dependent adenylate cyclase: Chemical and enzymatic properties and interactions with eucaryotic cells. *in* Advances in Cyclic Nucleotide and Protein Phosphorylation Research. ed. P. Greengard et al. Vol. 17, pp. 189-198.

13. Pugsley, A.P. and Schwartz, M. 1985. Export and secretion of proteins by bacteria. FEMS Microbiol. Rev. 32:3-38.

14. Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., Buckel, S.D., Bell, A.W. and Hermodson, M.A. 1986. A family of ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature (London). 323:448-450.

15. Robertson, D.L. 1988. Relationships between the calmodulin-dependent adenylate cyclases produced by *Bacillus anthracis* and *Bordetella pertussis*. Submitted to Infect. Immun.

16. Glaser, P., Ladant, D., Sezer, O., Pichot, F., Ullmann, A. & Danchin, A. 1988. The calmodulin-sensitive adenylate cyclases of *Bordetella pertussis*: cloning and expression in *Escherichia coli*. Mol. Microbiol. 2:19-30.

17. Weiss, A.A., Hewlett, E.L., Myers, G.A. & Falkow, S. 1984. Pertussis toxin and extracytoplasmic adenylate cyclase as virulence factors in *Bordetella pertussis*. J. Infect. Dis. 150:219-222.

18. Band, L. and D.J. Henner. 1984. *Bacillus subtilis* requires a "stringent" Shine-Dalgarno region for gene expression. DNA 3:17-21.

19. Sterne, M. 1939. The effects of different carbon dioxide concentrations on the growth of virulent anthrax strains. Pathogenicity and immunity tests on guinea pigs and sheep with anthrax variants derived from virulent strains. Onderstepoort J. Vet. Sci. Anim. Husb. 13:307-312.

20. Yansura, D.G. and D.J. Henner. 1984. Development of an inducible promoter for controlled gene expression in Bacillus subtilis. Genetics and Biotechnology of Bacilli, pp. 249-263.

21. Ivins, B.E. and S.L. Welkos. 1986. Cloning and expression of the *Bacillus anthracis* protective antigen gene in *Bacillus subtilis*. Infect. Immun. 54:537-542.

22. Studier, F.W. and B.A. Moffatt. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189:113-130.

23. Saunders, C.W., B.J. Schmidt, M.S. Mirot, L.D. Thompson, and M.S. Guyer. 1984. Use of chromosomal integration in the establishment and expression of *blaZ*, a *Staphylococcus aureus* β-lactamase gene, in *Bacillus subtilis*. J. Bacteriol 157:718-726.

24. Saunders, C.W., C. D.B. Banner, S. R. Fahnestock, M. Lindberg, M.S. Mirot, C.S. Rhodes, C.F. Rudolph, B.J. Schmidt, L.D. Thompson, M. Uhlen, and M.S. Guyer. 1984. Use of chromosomal integration for the establishment of heterologous genes in *Bacillus subtilis*. *In* Protein Transport and Secretion, pp. 329-339.

25. Lovett, P.S. and S. Mongkolsuk. 1988. Promoter probe plasmids for Gram-positive Bacteria. *In* Vectors: A survey of molecular cloning vectors and their uses. ed. R.L. Rodriquez and D.T. Denhardt. Butterworths.

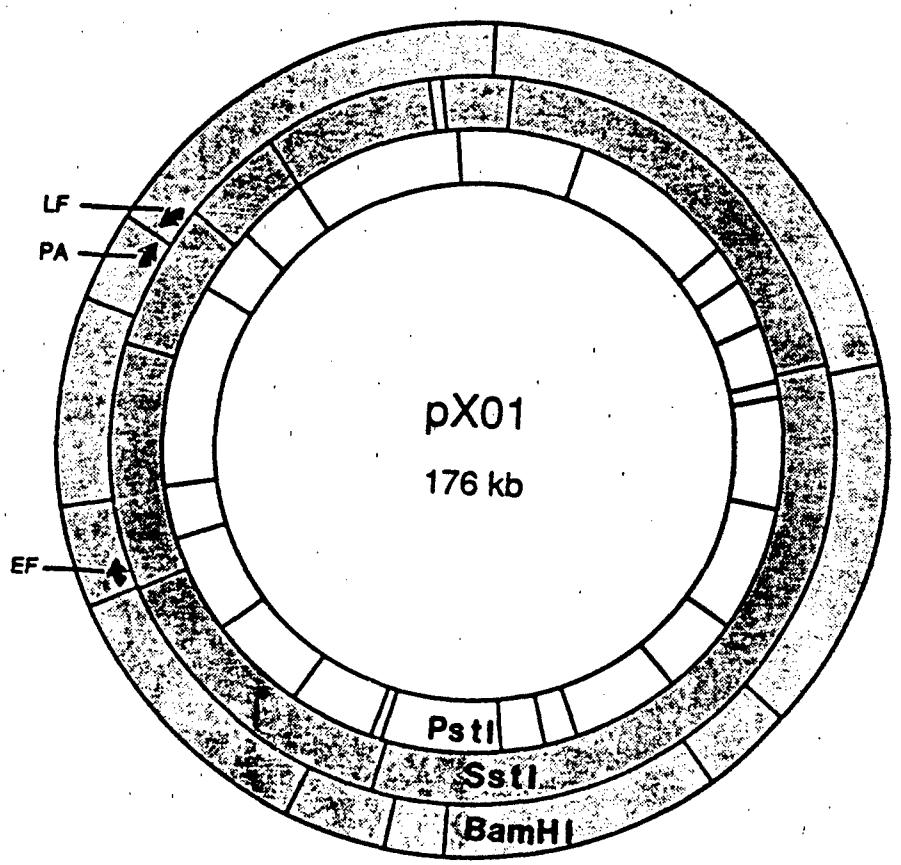


FIGURE 1. Restriction map of pXO1. The positions of the LF, PA and EF genes are depicted. The sizes of DNA fragments for each enzyme are not included due to the lack of space.

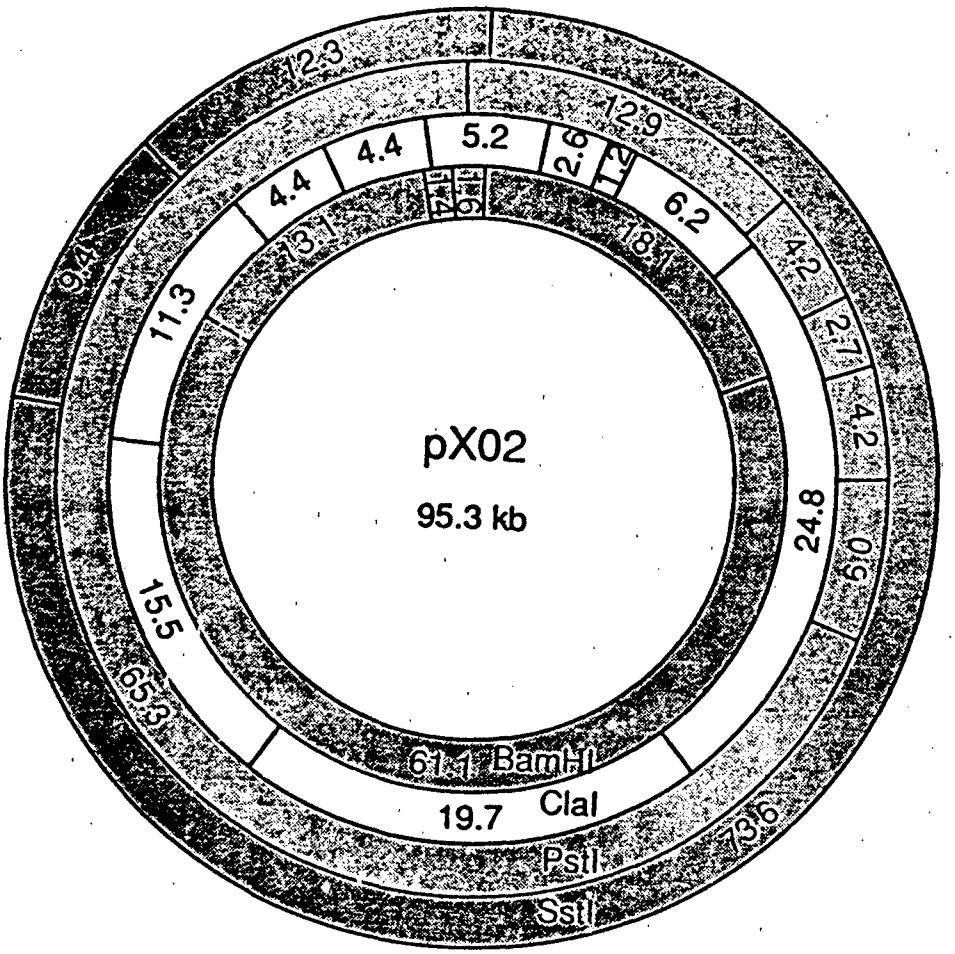
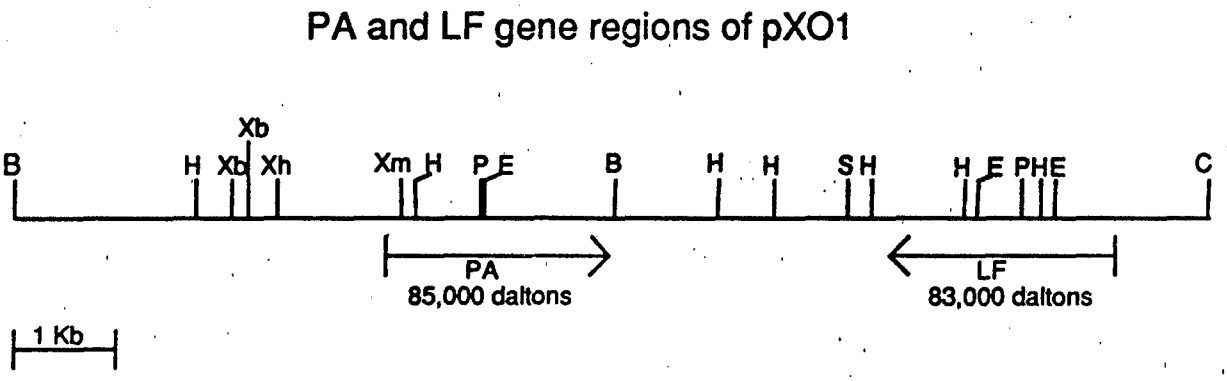


FIGURE 2. Restriction map of pX02.



B -*Bam* HI
H -*Hin* dIII
Xb-*Xba* I
Xh-*Xho* I
Xm-*Xmn* I
P - *Pst* I
E - *Eco* RI
S - *Sst* I
C - *Cla* I

FIGURE 3. Restriction map of the PA and LF gene regions on pXO1.

(A) The signal peptides (in bold) for EF, PA and LF are shown. The proposed secondary structure most likely to be assumed for the first 60 amino acids of each protein is shown (α=α-helix; β=β-sheet; t=β-turn; blank=random coil). The amino terminal amino acid, as determined by Dr. J. Schmidt (USAMRIID), for each mature toxin protein is also shown.

EF signal peptide

```
                                        ↓-start of mature EF
1   MTRNKFIPNKFSIISFSVLL FAISSSQAIEVNAMNEHYTE SDIKRNHKTEKNKTEKEKFK  60
    ααttt   ttβββββββββββ ααα   αααααααααααααα αααααααααααααααααααα
```

PA signal peptide

```
                                   ↓-start of mature PA
1   MKKRKVLIPLMALSTILVSS TGNLEVIQAEVKQENRLLNE SESSSQGLLGYYFSDLNFQA  60
    αααααβββββααβββββ       αααααααααααα     ttttttβββββββt  ααα
```

LF signal peptide

```
                                        ↓-start of mature LF
1   MNIKKEFIKVISMSCLVTAI TLSGPVFIPLVQGAGGHGDV GMHVKEKEKNKDENKRKDEE  60
    ααααααααααααtβββββββ βt   tβββββββ       α αααααααααααααααααααα
```

(B) The amino acid sequence at the end of the anthrax toxin signal peptides is shown. Cleavage occurs after Ala or Gly, consistent with known cleavages after bacilli signal peptides (14). Similar amino acids at the end of the signal peptides (denoted with a vertical bar [|]) probably represents signal peptidase recognition sequences. The numbers (-1 or +1) indicate the last amino acid of the signal peptide and the first amino acid of the mature toxin protein, respectively.

```
                                  -1   +1
EF signal peptide          Glu-Val-Asn-Ala--Met
                               |   |   |
PA signal peptide          Val-Ile-Gln-Ala--Glu
                           |   |   |   |
LF signal peptide          Leu-Val-Gln-Gly--Ala
```

FIGURE 4. Anthrax toxin signal peptides.

APPENDIX I: Nucleotide Sequence of the EF gene.

```
      10        20        30        40        50        60        70        80        90
TTACTTTTTTATATACTGAATTAAAAAGTCCAAGCACTTATATCGTAATAGATGCTTTCTATTGACCTTATAGTCCTTGAAGTTACGACT

     100       110       120       130       140       150       160       170       180
GACCAATTATGAGACGTTTGCGCTAACCTGCTGAATTCAAAATCCGACTTAGAAATACACATATAGAAATAAACAACGTAATCCATGTCA

     190       200       210       220       230       240       250       260       270
CTGTACCGTTTTTTTACTAAATAAACGAAATCAGTGTAAAAATGAACAGCTGAACTTTATCAACTTAGAATCTCTTTTTTTACTTTAAAT

     280       290       300       310       320       330       340       350       360
GCCTAGCTGTTTTTTCTAATGTTTGTATTTCTAAATATATTTAAATATGAATTGTAGCTGTGTGCCAAGAGTTATAATTAATTTAAATAA
                                              -35 (putative promoter site) -10

     370       380       390       400       410       420       430       440       450
GATTATATTTGTAAATAAAATTGTAATTTAACATGTAGAATAAAGAGATTTTTAGTTTTATTAACAGGATGAAAATCCATAAAACCGTAA

     460       470       480       490       500       510       520       530       540
ATGTGATTTCTAAATTAGTTTAAAATAAAAAACAAGGATTTGCTCAGACTTGAGATGAATATCTAAATATCAAGAACCAAAGGAGGTTTA
                                                                    ribosome binding site

  +1 550       560       570       580       590       600       610       620       630
AGAATGACTAGAAATAAATTTATACCTAATAAGTTTAGTATTATATCCTTTTCAGTATTACTATTTGCTATATCCTCCTCACAGGCTATA
  MetThrArgAsnLysPheIleProAsnLysPheSerIleIleSerPheSerValLeuLeuPheAlaIleSerSerSerGlnAlaIle
                    33 amino acid leader sequence

     640       650       660       670       680       690       700       710       720
GAAGTAAATGCTATGAATGAACATTACACTGAGAGTGATATTAAAAGAAACCATAAAACTGAAAAAAATAAAACTGAAAAAGAAAAATTT
GluValAsnAlaMETAsnGluHisTyrThrGluSerAspIleLysArgAsnHisLysThrGluLysAsnLysThrGluLysGluLysPhe
          1st amino acid of EF

     730       740       750       760       770       780       790       800       810
AAAGACAGTATTAATAACTTAGTTAAAACAGAATTTACCAATGAAACTTTAGATAAAATACAGCAGACACAAGACTTATTAAAAAAGATA
LysAspSerIleAsnAsnLeuValLysThrGluPheThrAsnGluThrLeuAspLysIleGlnGlnThrGlnAspLeuLeuLysLysIle

     820       830       840       850       860       870       880       890       900
CCTAAGGATGTACTTGAAATTTATAGTGAATTAGGAGGAGAAATCTATTTTACAGATATAGATTTAGTAGAACATAAGGAGTTACAAGAT
ProLysAspValLeuGluIleTyrSerGluLeuGlyGlyGluIleTyrPheThrAspIleAspLeuValGluHisLysGluLeuGlnAsp

     910       920       930       940       950       960       970       980       990
TTAAGTGAAGAAGAGAAAAATAGTATGAATAGTAGAGGTGAAAAAGTTCCGTTTGCATCCGTTTTGTATTTGAAAAGAAAAGGGAAACA
LeuSerGluGluGluLysAsnSerMetAsnSerArgGlyGluLysValProPheAlaSerArgPheValPheGluLysLysArgGluThr

    1000      1010      1020      1030      1040      1050      1060      1070      1080
CCTAAATTAATTATAAATATCAAAGATTATGCAATTAATAGTGAACAAAGTAAAGAAGTATATTATGAAATTGGAAAGCGCATTTCTCTT
ProLysLeuIleIleAsnIleLysAspTyrAlaIleAsnSerGluGlnSerLysGluValTyrTyrGluIleGlyLysGlyIleSerLeu

    1090      1100      1110      1120      1130      1140      1150      1160      1170
GATATTATAAGTAAGGATAAATCTCTAGATCCAGAGTTTTTAAATTTAATTAAGAGTTTAAGCGATGATAGTGATAGTAGCGACCTTTTA
AspIleIleSerLysAspLysSerLeuAspProGluPheLeuAsnLeuIleLysSerLeuSerAspAspSerAspSerSerAspLeuLeu
```

1180 1190 1200 1210 1220 1230 1240 1250 1260
TTTAGTCAAAAATTTAAAGAGAAGCTAGAATTGAATAATAAAAGTATAGATATAAATTTTATAAAAGAAAATTTAACTGAATTTCAGCAT
PheSerGlnLysPheLysGluLysLeuGluLeuAsnAsnLysSerIleAspIleAsnPheIleLysGluAsnLeuThrGluPheGlnHis

1270 1280 1290 1300 1310 1320 1330 1340 1350
GCGTTTTCTTTAGCGTTTTCTTATTATTTTGCACCTGACCATAGAACGGTATTAGAGTTATATGCCCCCGACATGTTTGAGTATATGAAT
AlaPheSerLeuAlaPheSerTyrTyrPheAlaProAspHisArgThrValLeuGluLeuTyrAlaProAspMetPheGluTyrMetAsn

1360 1370 1380 1390 1400 1410 1420 1430 1440
AAGTTAGAAAAAGGGGGATTTGAGAAAATAAGTGAAAGTTTGAAGAAAGAAGGTGTGGAAAAAGATAGGATTGATGTGCTGAAAGGAGAA
LysLeuGluLysGlyGlyPheGluLysIleSerGluSerLeuLysLysGluGlyValGluLysAspArgIleAspValLeuLysGlyGlu

1450 1460 1470 1480 1490 1500 1510 1520 1530
AAAGCACTTAAAGCTTCAGGTTTAGTACCAGAACATGCAGATGCTTTTAAAAAAATTGCTAGAGAATTAAATACATATATTCTTTTTAGG
LysAlaLeuLysAlaSerGlyLeuValProGluHisAlaAspAlaPheLysLysIleAlaArgGluLeuAsnThrTyrIleLeuPheArg

1540 1550 1560 1570 1580 1590 1600 1610 1620
CCTGTTAATAAGTTAGCTACAAACCTTATTAAAAGTGGTGTGGCTACAAAGGGATTGAATGAACATGGAAAGAGTTCGGATTGGGGCCCT
ProValAsnLysLeuAlaThrAsnLeuIleLysSerGlyValAlaThrLysGlyLeuAsnGluHisGlyLysSerSerAspTrpGlyPro

1630 1640 1650 1660 1670 1680 1690 1700 1710
GTAGCTGGATACATACCATTTGATCAAGATTTATCTAAGAAGCATGGTCAACAATTAGCTGTCGAGAAAGGAAATTTAGAAAATAAAAAA
ValAlaGlyTyrIleProPheAspGlnAspLeuSerLysLysHisGlyGlnGlnLeuAlaValGluLysGlyAsnLeuGluAsnLysLys

1720 1730 1740 1750 1760 1770 1780 1790 1800
TCAATTACAGAGCATGAAGGTGAAATAGGTAAAATACCATTAAAGTTAGACCATTTAAGAATAGAAGAGTTAAAGGAAAATGGGATAATT
SerIleThrGluHisGluGlyGluIleGlyLysIleProLeuLysLeuAspHisLeuArgIleGluGluLeuLysGluAsnGlyIleIle

1810 1820 1830 1840 1850 1860 1870 1880 1890
TTGAAGGGTAAAAAAGAAATTGATAATGGTAAAAAATATTATTTGTTAGAATCGAATAATCAGGTATATGAATTTAGAATTAGCGATGAA
LeuLysGlyLysLysGluIleAspAsnGlyLysLysTyrTyrLeuLeuGluSerAsnAsnGlnValTyrGluPheArgIleSerAspGlu

1900 1910 1920 1930 1940 1950 1960 1970 1980
AACAACGAAGTACAATACAAGACAAAAGAAGGTAAAATTACTGTTTTAGGGGAAAAATTCAATTGGAGAAATATAGAAGTGATGGCTAAA
AsnAsnGluValGlnTyrLysThrLysGluGlyLysIleThrValLeuGlyGluLysPheAsnTrpArgAsnIleGluValMetAlaLys

1990 2000 2010 2020 2030 2040 2050 2060 2070
AATGTAGAAGGGGTCTTGAAGCCGTTAACAGCTGACTATGATTTATTTGCACTTGCCCCAAGTTTAACAGAAATAAAAAAACAAATACCC
AsnValGluGlyValLeuLysProLeuThrAlaAspTyrAspLeuPheAlaLeuAlaProSerLeuThrGluIleLysLysGlnIlePro

2080 2090 2100 2110 2120 2130 2140 2150 2160
ACAAAAAGAATGGATAAAGTAGTTAACACCCCAAATTCATTAGAAAAGCAAAAAGGTGTTACTAATTTATTGATTAAATATGGAATTGAG
ThrLysArgMetAspLysValValAsnThrProAsnSerLeuGluLysGlnLysGlyValThrAsnLeuLeuIleLysTyrGlyIleGlu

2170 2180 2190 2200 2210 2220 2230 2240 2250
AGGAAACCGGATTCAACTAAGGGAACTTTATCAAATTGGCAAAAACAAATGCTTGATCGTTTGAATGAAGCAGTCAAATATACAGGATAT
ArgLysProAspSerThrLysGlyThrLeuSerAsnTrpGlnLysGlnMetLeuAspArgLeuAsnGluAlaValLysTyrThrGlyTyr

2260 2270 2280 2290 2300 2310 2320 2330 2340
ACAGGGGGGGATGTGGTTAACCATGGCACAGAGCAAGATAATGAAGAGTTTCCTGAAAAAGATAACGAAATTTTTATAATTAATCCAGAA
ThrGlyGlyAspValValAsnHisGlyThrGluGlnAspAsnGluGluPheProGluLysAspAsnGluIlePheIleIleAsnProGlu

```
      2350       2360       2370       2380       2390       2400       2410       2420       2430
GGTGAATTTATATTAACTAAAAATTGGGAGATGACAGGTAGATTTATAGAAAAAAACATTACGGGAAAAGATTATTTATATTATTTTAAC
GlyGluPheIleLeuThrLysAsnTrpGluMetThrGlyArgPheIleGluLysAsnIleThrGlyLysAspTyrLeuTyrTyrPheAsn

      2440       2450       2460       2470       2480       2490       2500       2510       2520
CGTTCTTATAATAAAATAGCTCCTGGTAATAAAGCTTATATTGAGTGGACTGATCCGATTACAAAAGCCAAAATAAATACCATCCCTACG
ArgSerTyrAsnLysIleAlaProGlyAsnLysAlaTyrIleGluTrpThrAspProIleThrLysAlaLysIleAsnThrIleProThr

      2530       2540       2550       2560       2570       2580       2590       2600       2610
TCAGCAGAGTTTATAAAAAACTTATCCAGTATCAGAAGATCTTCAAATGTAGGAGTTTATAAAGATAGTGGCGACAAAGACGAATTTGCA
SerAlaGluPheIleLysAsnLeuSerSerIleArgArgSerSerAsnValGlyValTyrLysAspSerGlyAspLysAspGluPheAla

      2620       2630       2640       2650       2660       2670       2680       2690       2700
AAAAAAGAAAGCGTGAAAAAAATTGCAGGATATTTGTCAGACTATTACAATTCAGCAAATCATATTTTTTCTCAGGAAAAAAAGCGTAAA
LysLysGluSerValLysLysIleAlaGlyTyrLeuSerAspTyrTyrAsnSerAlaAsnHisIlePheSerGlnGluLysLysArgLys

      2710       2720       2730       2740       2750       2760       2770       2780       2790
ATATCAATATTTCGTGGAATCCAAGCCTATAATGAAATTGAAAATGTTCTAAAATCTAAACAAATAGCACCAGAATACAAAAATTATTTT
IleSerIlePheArgGlyIleGlnAlaTyrAsnGluIleGluAsnValLeuLysSerLysGlnIleAlaProGluTyrLysAsnTyrPhe

      2800       2810       2820       2830       2840       2850       2860       2870       2880
CAATATTTAAAGGAAAGGATTACCAATCAAGTTCAATTGCTTCTAACACATCAAAAATCTAATATTGAATTTAAATTATTGTATAAACAA
GlnTyrLeuLysGluArgIleThrAsnGlnValGlnLeuLeuLeuThrHisGlnLysSerAsnIleGluPheLysLeuLeuTyrLysGln

      2890       2900       2910       2920       2930       2940       2950       2960       2970
TTAAACTTTACAGAAAATGAAACGGATAATTTTGAGGTCTTCCAAAAAATTATTGATGAAAAATAAATATATATAATTGTTTTTCTGAAA
LeuAsnPheThrGluAsnGluThrAspAsnPheGluValPheGlnLysIleIleAspGluLys

      2980       2990       3000       3010       3020       3030       3040       3050       3060
ATTCATCATTTTAAAGAAGACACTAGGAATTAAATAGATGTATTGAATAGTTATAGTAATGGTCTTGTATGGACATACCGCTTATACTTT
```

APPENDIX II. EF amino acid sequence

(33 aa signal peptide) ↓-Start of mature EF (767 aa)

```
  1 MTRNKFIPNKFSIISFSVLLFAISSSQAIEVNAMNEHYTESDIKRNHKTEKNKTEKEKFKDSINNLVKTE

 71 FTNETLDKIQQTQDLLKKIPKDVLEIYSELGGEIYFTDIDLVEHKELQDLSEEEKNSMNSRGEKVPFASR

141 FVFEKKRETPKLIINIKDYAINSEQSKEVYYEIGKGISLDIISKDKSLDPEFLNLIKSLSDDSDSSDLLF
                                    #1                                  #2

211 SQKFKEKLELNNKSIDINFIKENLTEFQHAFSLAFSYYFAPDHRTVLELYAPDMFEYMNKLEKGGFEKIS
                                                          #3

281 ESLKKEGVEKDRIDVLKGEKALKASGLVPEHADAFKKIARELNTYILFRPVNKLATNLIKSGVATKGLNE
                  (Potential calmodulin binding site)

351 HGKSSDWGPVAGYIPFDQDLSKKHGQQLAVEKGNLENKKSITEHEGEIGKIPLKLDHLRIEELKENGIIL
    (Putative ATP binding site)

421 KGKKEIDNGKKYYLLESNNQVYEFRISDENNEVQYKTKEGKITVLGEKFNWRNIEVMAKNVEGVLKPLTA

491 DYDLFALAPSLTEIKKQIPTKRMDKVVNTPNSLEKQKGVTNLLIKYGIERKPDSTKGTLSNWQKQMLDRL

561 NEAVKYTGYTGGDVVNHGTEQDNEEFPEKDNEIFIINPEGEFILTKNWEMTGRFIEKNITGKDYLYYFNR

631 SYNKIAPGNKAYIEWTDPITKAKINTIPTSAEFIKNLSSIRRSSNVGVYKDSGDKDEFAKKESVKKIAGY

701 LSDYYNSANHIFSQEKKRKISIFRGIQAYNEIENVLKSKQIAPEYKNYFQYLKERITNQVQLLLTHQKSN

771 IEFKLLYKQLNFTENETDNFEVFQKIIDEK
```

****

The sequence contains 800 amino acids (M_r 92,464):

Ala (A)	32	Leu (L)	69
Arg (R)	22	Lys (K)	103
Asn (N)	61	Met (M)	9
Asp (D)	44	Phe (F)	40
Cys (C)	0	Pro (P)	23
Gln (Q)	27	Ser (S)	55
Glu (E)	82	Thr (T)	39
Gly (G)	40	Trp (W)	5
His (H)	13	Tyr (Y)	34
Ile (I)	68	Val (V)	34

Acidic	(Asp + Glu)	126
Basic	(Arg + Lys)	125
Aromatic	(Phe + Trp + Tyr)	79
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	259

APPENDIX III. Homology Comparison between EF and pertussis cyclase.

1. Domains #1, #2 and #3 represent three highly conserved amino acid domains in EF (top line of each pair) and the pertussis cyclase (bottom line in each pair).

2. The numbers to the left of each line indicates the amino acid position for EF-precursor or the pertussis cyclase.

3. The asterisks (*) indicate the consensus sequences for the ATP binding site for EF and the pertussis cyclase.

APPENDIX IV: Nucleotide Sequence of the LF gene.

```
      10        20        30        40        50        60        70        80        90
AAATTAGGATTTCGGTTATGTTTAGTATTTTTTTAAAATAATAGTATTAAATAGTGGAATGCAAATGATAAATGGGCTTTAAACAAAACT

     100       110       120       130       140       150       160       170       180
AATGAAATAATCTACAAATGGAATTTCTCCAGTTTTAGATTAAACCATACCAAAAAAATCACACTGTCAAGAAAAATGATAGAATCCCTA

     190       200       210       220       230       240       250       260       270
CACTAATTAACATAACCAAATTGGTAGTTATAGGTAGAAACTTATTTATTTCTATAATACCATGCAAAAAAGTAAATATTCTGTTCCATA

     280       290       300       310       320       330       340       350       360
CTATTTTAGTAAATTATTTAGCAAGTAAATTTTGGTGTATAAACAAAGTTTATCTTAATATAAAAAATTACTTTACTTTTATACAGATTA

     370       380       390       400       410       420       430       440       450
AAATGAAAAATTTTTTATGACAAGAAATATTGCCTTTAATTTATGAGGAAATAAGTAAAATTTTCTACATACTTTATTTTATTGTTGAAA

     460       470       480       490       500       510       520       530       540
TGTTCACTTATAAAAAAGGAGAGATTAAATATGAATATAAAAAAAGAATTTATAAAAGTAATTAGTATGTCATGTTTAGTAACAGCAATT
          (r.b.s.)          MetAsnIleLysLysGluPheIleLysValIleSerMetSerCysLeuValThrAlaIle
                                (33 amino acid signal peptide)

     550       560       570       580       590       600       610       620       630
ACTTTGAGTGGTCCCGTCTTTATCCCCCTTGTACAGGGGGCGGGCGGTCATGGTGATGTAGGTATGCACGTAAAAGAGAAAGAGAAAAAT
ThrLeuSerGlyProValPheIleProLeuValGlnGlyAlaGlyGlyHisGlyAspValGlyMetHisValLysGluLysGluLysAsn
                                +1 of mature LF

     640       650       660       670       680       690       700       710       720
AAAGATGAGAATAAGAGAAAAGATGAAGAACGAAATAAAACACAGGAAGAGCATTTAAAGGAAATCATGAAACACATTGTAAAAATAGAA
LysAspGluAsnLysArgLysAspGluGluArgAsnLysThrGlnGluGluHisLeuLysGluIleMetLysHisIleValLysIleGlu

     730       740       750       760       770       780       790       800       810
GTAAAAGGGGAGGAAGCTGTTAAAAAAGAGGCAGCAGAAAAGCTACTTGAGAAAGTACCATCTGATGTTTTAGAGATGTATAAAGCAATT
ValLysGlyGluGluAlaValLysLysGluAlaAlaGluLysLeuLeuGluLysValProSerAspValLeuGluMetTyrLysAlaIle

     820       830       840       850       860       870       880       890       900
GGAGGAAAGATATATATTGTGGATGGTGATATTACAAAACATATATCTTTAGAAGCATTATCTGAAGATAAGAAAAAAAATAAAAGACATT
GlyGlyLysIleTyrIleValAspGlyAspIleThrLysHisIleSerLeuGluAlaLeuSerGluAspLysLysLysIleLysAspIle

     910       920       930       940       950       960       970       980       990
TATGGGAAAGATGCTTTATTACATGAACATTATGTATATGCAAAAGAAGGATATGAACCCGTACTTGTAATCCAATCTTCGGAAGATTAT
TyrGlyLysAspAlaLeuLeuHisGluHisTyrValTyrAlaLysGluGlyTyrGluProValLeuValIleGlnSerSerGluAspTyr

    1000      1010      1020      1030      1040      1050      1060      1070      1080
GTAGAAAATACTGAAAAGGCACTGAACGTTTATTATGAAATAGGTAAGATATTATCAAGGGATATTTTAAGTAAAATTAATCAACCATAT
ValGluAsnThrGluLysAlaLeuAsnValTyrTyrGluIleGlyLysIleLeuSerArgAspIleLeuSerLysIleAsnGlnProTyr

    1090      1100      1110      1120      1130      1140      1150      1160      1170
CAGAAATTTTTAGATGTATTAAATACCATTAAAAATGCATCTGATTCAGATGGACAAGATCTTTTATTTACTAATCAGCTTAAGGAACAT
GlnLysPheLeuAspValLeuAsnThrIleLysAsnAlaSerAspSerAspGlyGlnAspLeuLeuPheThrAsnGlnLeuLysGluHis
```

1180 1190 1200 1210 1220 1230 1240 1250 1260
CCCACAGACTTTTCTGTAGAATTCTTGGAACAAAATAGCAATGAGGTACAAGAAGTATTTGCGAAAGCTTTTGCATATTATATCGAGCCA
ProThrAspPheSerValGluPheLeuGluGlnAsnSerAsnGluValGlnGluValPheAlaLysAlaPheAlaTyrTyrIleGluPro

1270 1280 1290 1300 1310 1320 1330 1340 1350
CAGCATCGTGATGTTTTACAGCTTTATGCACCGGAAGCTTTTAATTACATGGATAAATTTAACGAACAAGAAATAAATCTATCCTTGGAA
GlnHisArgAspValLeuGlnLeuTyrAlaProGluAlaPheAsnTyrMetAspLysPheAsnGluGlnGluIleAsnLeuSerLeuGlu

1360 1370 1380 1390 1400 1410 1420 1430 1440
GAACTTAAAGATCAACGGATGCTGTCAAGATATGAAAAATGGGAAAAGATAAAACAGCACTATCAACACTGGAGCGATTCTTTATCTGAA
GluLeuLysAspGlnArgMetLeuSerArgTyrGluLysTrpGluLysIleLysGlnHisTyrGlnHisTrpSerAspSerLeuSerGlu

1450 1460 1470 1480 1490 1500 1510 1520 1530
GAAGGAAGAGGACTTTTAAAAAAGCTGCAGATTCCTATTGAGCCAAAGAAAGATGACATAATTCATTCTTTATCTCAAGAAGAAAAAGAG
GluGlyArgGlyLeuLeuLysLysLeuGlnIleProIleGluProLysLysAspAspIleIleHisSerLeuSerGlnGluGluLysGlu

1540 1550 1560 1570 1580 1590 1600 1610 1620
CTTCTAAAAAGAATACAAATTGATAGTAGTGATTTTTTATCTACTGAGGAAAAAGAGTTTTTAAAAAAGCTACAAATTGATATTCGTGAT
LeuLeuLysArgIleGlnIleAspSerSerAspPheLeuSerThrGluGluLysGluPheLeuLysLysLeuGlnIleAspIleArgAsp

1630 1640 1650 1660 1670 1680 1690 1700 1710
TCTTTATCTGAAGAAGAAAAAGAGCTTTTAAATAGAATACAGGTGGATAGTAGTAATCCTTTATCTGAAAAAGAAAAAGAGTTTTTAAAA
SerLeuSerGluGluGluLysGluLeuLeuAsnArgIleGlnValAspSerSerAsnProLeuSerGluLysGluLysGluPheLeuLys

1720 1730 1740 1750 1760 1770 1780 1790 1800
AAGCTGAAACTTGATATTCAACCATATGATATTAATCAAAGGTTGCAAGATACAGGAGGGTTAATTGATAGTCCGTCAATTAATCTTGAT
LysLeuLysLeuAspIleGlnProTyrAspIleAsnGlnArgLeuGlnAspThrGlyGlyLeuIleAspSerProSerIleAsnLeuAsp

1810 1820 1830 1840 1850 1860 1870 1880 1890
GTAAGAAAGCAGTATAAAAGGGATATTCAAAATATTGATGCTTTATTACATCAATCCATTGGAAGTACCTTGTACAATAAAATTTATTTG
ValArgLysGlnTyrLysArgAspIleGlnAsnIleAspAlaLeuLeuHisGlnSerIleGlySerThrLeuTyrAsnLysIleTyrLeu

1900 1910 1920 1930 1940 1950 1960 1970 1980
TATGAAAATATGAATATCAATAACCTTACAGCAACCCTAGGTGCGGATTTAGTTGATTCCACTGATAATACTAAAATTAATAGAGGTATT
TyrGluAsnMetAsnIleAsnAsnLeuThrAlaThrLeuGlyAlaAspLeuValAspSerThrAspAsnThrLysIleAsnArgGlyIle

1990 2000 2010 2020 2030 2040 2050 2060 2070
TTCAATGAATTCAAAAAAAATTTCAAATATAGTATTTCTAGTAACTATATGATTGTTGATATAAATGAAAGGCCTGCATTAGATAATGAG
PheAsnGluPheLysLysAsnPheLysTyrSerIleSerSerAsnTyrMetIleValAspIleAsnGluArgProAlaLeuAspAsnGlu

2080 2090 2100 2110 2120 2130 2140 2150 2160
CGTTTGAAATGGAGAATCCAATTATCACCAGATACTCGAGCAGGATATTTAGAAAATGGAAAGCTTATATTACAAAGAAACATCGGTCTG
ArgLeuLysTrpArgIleGlnLeuSerProAspThrArgAlaGlyTyrLeuGluAsnGlyLysLeuIleLeuGlnArgAsnIleGlyLeu

2170 2180 2190 2200 2210 2220 2230 2240 2250
GAAATAAAGGATGTACAAATAATTAAGCAATCCGAAAAAGAATATATAAGGATTGATGCGAAAGTAGTGCCAAAGAGTAAAATAGATACA
GluIleLysAspValGlnIleIleLysGlnSerGluLysGluTyrIleArgIleAspAlaLysValValProLysSerLysIleAspThr

2260 2270 2280 2290 2300 2310 2320 2330 2340
AAAATTCAAGAAGCACAGTTAAATATAAATCAGGAATGGAATAAAGCATTAGGGTTACCAAAATATACAAAGCTTATTACATTCAACGTG
LysIleGlnGluAlaGlnLeuAsnIleAsnGlnGluTrpAsnLysAlaLeuGlyLeuProLysTyrThrLysLeuIleThrPheAsnVal

```
      2350       2360       2370       2380       2390       2400       2410       2420       2430
CATAATAGATATGCATCCAATATTGTAGAAAGTGCTTATTTAATATTGAATGAATGGAAAAATAATATTCAAAGTGATCTTATAAAAAAG
HisAsnArgTyrAlaSerAsnIleValGluSerAlaTyrLeuIleLeuAsnGluTrpLysAsnAsnIleGlnSerAspLeuIleLysLys

      2440       2450       2460       2470       2480       2490       2500       2510       2520
GTAACAAATTACTTAGTTGATGGTAATGGAAGATTTGTTTTTACCGATATTACTCTCCCTAATATAGCTGAACAATATACACATCAAGAT
ValThrAsnTyrLeuValAspGlyAsnGlyArgPheValPheThrAspIleThrLeuProAsnIleAlaGluGlnTyrThrHisGlnAsp

      2530       2540       2550       2560       2570       2580       2590       2600       2610
GAGATATATGAGCAAGTTCATTCAAAAGGGTTATATGTTCCAGAATCCCGTTCTATATTACTCCATGGACCTTCAAAAGGTGTAGAATTA
GluIleTyrGluGlnValHisSerLysGlyLeuTyrValProGluSerArgSerIleLeuLeuHisGlyProSerLysGlyValGluLeu

      2620       2630       2640       2650       2660       2670       2680       2690       2700
AGGAATGATAGTGAGGGTTTTATACACGAATTTGGACATGCTGTGGATGATTATGCTGGATATCTATTAGATAAGAACCAATCTGATTTA
ArgAsnAspSerGluGlyPheIleHisGluPheGlyHisAlaValAspAspTyrAlaGlyTyrLeuLeuAspLysAsnGlnSerAspLeu

      2710       2720       2730       2740       2750       2760       2770       2780       2790
GTTACAAATTCTAAAAAATTCATTGATATTTTTAAGGAAGAAGGGAGTAATTTAACTTCGTATGGAGAACAAATGAAGCGGAATTTTTT
ValThrAsnSerLysLysPheIleAspIlePheLysGluGluGlySerAsnLeuThrSerTyrGlyArgThrAsnGluAlaGluPhePhe

      2800       2810       2820       2830       2840       2850       2860       2870       2880
GCAGAAGCCTTTAGGTTAATGCATTCTACGGACCATGCTGAACGTTTAAAAGTTCAAAAAAATGCTCCGAAAACTTTCCAATTTATTAAC
AlaGluAlaPheArgLeuMetHisSerThrAspHisAlaGluArgLeuLysValGlnLysAsnAlaProLysThrPheGlnPheIleAsn

      2890       2900       2910       2920       2930       2940       2950       2960       2970
GATCAGATTAAGTTCATTATTAACTCATAAGTAATGTATTAAAAATTTTCAAATGGATTTAATAATAATAATAATAATAATAATAACGGG
AspGlnIleLysPheIleIleAsnSer

      2980       2990       3000       3010       3020       3030       3040       3050       3060
ACCAGCCATTATGAAGCAACTAATTCTAGACTTGATAGTAATTCTTGGGAAGCACCAGATAGTGTAAAAGGTGGCATTGCCAGAATGATA

      3070       3080       3090       3100       3110       3120       3130       3140       3150
TTTTATGTGTTCGTTAGATATGAAGGCAAAAACAATGATCCTGACCTAGAACTTAATGATAATGTTATTAATAATTTAATGCCTTTTATA

      3160       3170       3180       3190       3200       3210       3220       3230       3240
GGAATATTAGTAAAAGTGCCGAAAAGATCCTGTTGCAAAGCTTTTAAAGAACATATTATTCTATCAAGTGGCTGTATATTTTGTGTAATT

      3250       3260       3270       3280       3290
TTCAATAAATTTTGTAATTAAGCATACGTCAAAAAACCGAAATCTGAGCTC
                                                  SstI
```

APPENDIX V. LF amino acid sequence

(29 aa signal peptide) ↓-Start of mature LF (780 aa)

```
  1 MNIKKEFIKVISMSCLVTAITLSGPVFIPLVQGAGGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLK
 71 EIMKHIVKIEVKGEEAVKKEAAEKLLEKVPSDVLEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDI
141 YGKDALLHEHYVYAKEGYEPVLVIQSSEDYVENTEKALNVYYEIGKILSRDILSKINQPYQKFLDVLNTI
                                                  #1
211 KNASDSDGQDLLFTNQLKEHPTDFSVEFLEQNSNEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKF
        #2                                                          #3
281 NEQEINLSLEELKDQRMLSRYEKWEKIKQHYQHWSDSLSEEGRGLLKKLQIPIEPKKDDIIHSLSQEEKE
351 LLKRIQIDSSDFLSTEEKEFLKKLQIDIRDSLSEEEKELLNRIQVDSSNPLSEKEKEFLKKLKLDIQPYD
421 INQRLQDTGGLIDSPSINLDVRKQYKRDIQNIDALLHQSIGSTLYNKIYLYENMNINNLTATLGADLVDS
491 TDNTKINRGIFNEFKKNFKYSISSNYMIVDINERPALDNERLKWRIQLSPDTRAGYLENGKLILQRNIGL
561 EIKDVQIIKQSEKEYIRIDAKVVPKSKIDTKIQEAQLNINQEWNKALGLPKYTKLITFNVHNRYASNIVE
631 SAYLILNEWKNNIQSDLIKKVTNYLVDGNGRFVFTDITLPNIAEQYTHQDEIYEQVHSKGLYVPESRSIL
701 LHGPSKGVELRNDSEGFIHEFGHAVDDYAGYLLDKNQSDLVTNSKKFIDIFKEEGSNLTSYGRTNEAEFF
771 AEAFRLMHSTDHAERLKVQKNAPKTFQFINDQIKFIINS
```

****

The sequence contains 809 amino acids (M_r 93,798):

Ala (A)	34	Leu (L)	80
Arg (R)	27	Lys (K)	86
Asn (N)	54	Met (M)	10
Asp (D)	55	Phe (F)	29
Cys (C)	1	Pro (P)	21
Gln (Q)	41	Ser (S)	54
Glu (E)	79	Thr (T)	28
Gly (G)	35	Trp (W)	5
His (H)	21	Tyr (Y)	35
Ile (I)	74	Val (V)	40

Acidic	(Asp + Glu)	134
Basic	(Arg + Lys)	113
Aromatic	(Phe + Trp + Tyr)	69
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	273

PUBLICATIONS

The following articles were published:

Leppla, S.H., D.L. Robertson, S.L. Welkos, L.A. Smith, and M.H. Vodkin. 1986. Cloning and analysis of genes for anthrax toxin components, pp. 275-278. *In* Bacterial protein toxins, Suppl. 15. Zentralblatt für bakteriologie und hygiene. 1. Abteilung. Gustav Fischer, Stuttgart.

Robertson, D. L., and S. H. Leppla. 1986. Molecular cloning and expression in *Escherichia coli* of the lethal factor gene of *Bacillus anthracis*. Gene 44:71-78.

Tippetts, M.T., and D.L. Robertson. 1988. Molecular cloning and expression of the Bacillus anthracis edema factor toxin gene: a calmodulin-dependent adenylate cyclase. J. Bacteriol. 170:2263-2266.

Kaspar, R.L. and Robertson, D.L. 1987. Purification and physical analysis of *Bacillus anthracis* plasmids pXO1 and pXO2. Biochem. Biophys. Res. Commun. 149:362-368.

The following manuscripts are submitted and presently being reviewed for publication:

Robertson, D.L., M.T. Tippetts and S.H. Leppla. 1988. Nucleotide sequence of the *Bacillus anthracis* edema factor gene (*cya*): A calmodulin-dependent adenylate cyclase. submitted to Gene.

Robertson, D.L. 1988. Relationships between the calmodulin-dependent adenylate cyclases produced by *Bacillus anthracis* and *Bordetella pertussis*. submitted to Infection and Immunity.

The following abstracts were published:

Kaspar, R. L. and D. L. Robertson. Purification and analysis of *Bacillus anthracis* plasmids pXO1 and pXO2. Abstr. Annu. Meet. Am. Soc. Microbiol. 1987.

Tippetts, M. T., D. L. Robertson and R. Leavitt. Molecular cloning and characterization of the *Bacillus anthracis* edema factor gene. Abstr. Annu. Meet. Am. Soc. Microbiol. 1987.

Robertson, D.L., T. Tippetts, Y. Luh, T. Bragg and R. Larson. 1988. Biochemical Analysis of the *Bacillus anthracis* Edema Factor Gene: A Calmodulin-Dependent Adenylate Cyclase. 72nd Annual Meeting of the Federation of American Societies for Experimental Biology.

The following invited seminars were given:

Donald L. Robertson. A Biochemical Analysis of the *Bacillus anthracis* Toxin Genes. 8th Annual Rocky Mountain Regional Biochemistry Conference, Pingree Park, Colorado. September, 1987.

Donald L. Robertson. A Biochemical Characterization of the *Bacillus anthracis* Toxin Genes. Brigham Young University Chemistry Department, January, 1988.

PERSONNEL

During the course of this contract, the principal investigator has been Dr. Donald L. Robertson, except for 10 months when Dr. Robertson was on sabbatical leave in the Bacteriology Division at USAMRIID. During this period, Dr. Ronald W. Leavitt served as principal investigator and directed the research of the graduate students.

Graduate students who have done research for this contract have included M. Todd Tippetts (Ph.D. awarded), Kent Hill (Ph.D. awarded), Scott Simpson (M.S. pending), Roger Kaspar (M.S. awarded), Tom Bragg (Ph.D. being completed). Dr. Robert Larson served as a post-doctoral fellow for the last year and performed research on the *B. subtilis* expression plasmids.

The following theses have been accepted:

Tippetts, M. T. 1986. Molecular cloning of the chloroplast genome of *Carthamus tinctorius* L. and of the edema factor gene from *Bacillus anthracis*. Department of Chemistry, Brigham Young University.

Kaspar, R. L. 1986. Purification and characterization of pXO1 and pXO2 plasmids from *Bacillus anthracis*. Department of Chemistry, Brigham Young University.

Luh, Y. 1988. Genetic Modification of the *Bacillus anthracis* Edema Factor Toxin Gene and Construction of Plasmid pBS42-EF which expresses EF in *Escherichia coli* and *Bacillus subtilis*. Department of Chemistry, Brigham Young University.

DISTRIBUTION LIST

5 copies	Commander US Army Medical Research Institute of Infectious Diseases ATTN: SGRD-UIZ-M Fort Detrick, Frederick, MD 21701-5011
1 copy	Commander US Army Medical Research and Development Command ATTN: SGRD-RMI-S Fort Detrick, Frederick, MD 21701-5012
2 copies	Defense Technical Information Center (DTIC) ATTN: DTIC-DDAC Cameron Station Alexandria, VA 22304-6145
1 copy	Dean School of Medicine Uniformed Services University of the Health Sciences 4301 Jones Bridge Road Bethesda, MD 20814-4799
1 copy	Commandant Academy of Health Sciences, US Army ATTN: AHS-CDM Fort Sam Houston, TX 78234-6100