The EMBO Journal vol.5 no.10 pp.2495–2502, 1986

Tetanus toxin: primary structure, expression in *E. coli*, and homology with botulinum toxins

Ulrich Eisel[1,2], Wolfgang Jarausch[1,2], Karin Goretzki[2], Agnes Henschen[3], Joachim Engels[4], Ulrich Weller[2], Martina Hudel[2], Ernst Habermann[2] and Heiner Niemann[1]

[1]Institut für Medizinische Virologie and [2]Buchheim Institut für Pharmakologie der Justus-Liebig-Universität, Frankfurter Str.107, D-6300 Giessen, [3]Max Planck Institut für Biochemie, Martinsried and [4]Institut für Organische Chemie der Johann-Wolfgang-Goethe-Universität, Frankfurt, FRG

Communicated by R.Rott

A pool of synthetic oligonucleotides was used to identify the gene encoding tetanus toxin on a 75-kbp plasmid from a toxigenic non-sporulating strain of *Clostridium tetani*. The nucleotide sequence contained a single open reading frame coding for 1315 amino acids corresponding to a polypeptide with a mol. wt of 150 700. In the mature toxin molecule, proline (2) and serine (458) formed the N termini of the 52 288 mol. wt light chain and the 98 300 mol. wt heavy chain, respectively. Cysteine (467) was involved in the disulfide linkage between the two subchains. The amino acid sequences of the tetanus toxin revealed striking homologies with the partial amino acid sequences of botulinum toxins A, B, and E, indicating that the neurotoxins from *C. tetani* and *C.botulinum* are derived from a common ancestral gene. Overlapping peptides together covering the entire tetanus toxin molecule were synthesized in *Escherichia coli* and identified by monoclonal antibodies. The promoter of the toxin gene was localized in a region extending 322 bp upstream from the ATG codon and was shown to be functional in *E. coli*.

Key words: tetanus toxin/botulinum toxin/promoters/expression in *E. coli*/monoclonal antibodies

Introduction

Tetanus and botulinum toxins are the most potent neurotoxins known to man. They are produced by the anaerobic spore-forming bacteria *Clostridium tetani* and *Clostridium botulinum* as single-chain polypeptides that are subsequently cleaved by endogenous protease(s) to yield the two subchains L (mol. wt 50 000) and H (mol. wt 100 000). The L- and H-chain are covalently linked by a disulfide bridge (for review, see Sakaguchi, 1983; Habermann and Dreyer, 1986). The individual subunits of tetanus toxin are virtually non-toxic after separation (Matsuda and Yoneda, 1975). Tetanus toxin binds via the C-terminal part of the H-chain to ganglioside receptors such as GD_{1b} and GT_{1b}. A fragment with similar binding specificity, termed fragment C, has been prepared from native toxin by papain cleavage (Helting and Zwisler, 1977). Similar to botulinum toxins, tetanus toxin inhibits neurotransmitter release by a still unknown mode of action. Whereas in botulism only peripheral cholinergic synapses are affected (Sakaguchi, 1983), the main pathogenic pathway of tetanus toxin involves binding to nerve terminals, internalization, retrograde axonal transport into the spinal cord and finally transition from post-synaptic to pre-synaptic neurones (Habermann and Dreyer, 1986). Therefore, tetanus toxin may serve as an ideal model compound to study these complex processes at the molecular level. Previous attempts to identify the gene encoding tetanus toxin have focused on plasmid DNA (Finn *et al.*, 1984) as well as on the total clostridial DNA (Fairweather *et al.*, 1986). Based on the analysis of plasmid DNA, we shall present for the first time the complete structure of a clostridial neurotoxin. The amino acid sequence of tetanus toxin, reported here, reveals a close homology to available partial amino acid sequences of botulinum toxins A, B and E. In addition, we shall present results of the expression of individual peptides in *E. coli* and show that the authentic toxin promoter is functional in Gram-negative cells.

Results

N-terminal amino acid sequences of the L- and H-subunits and of fragments C and BII_b of tetanus toxin

To identify the toxin gene with a minimum number of appropriate oligonucleotides, partial amino acid sequences were determined by Edman degradation of the L- and H-subunits as well as of fragments C and BII_b (Figure 1). Separation of the L- and the H-chain was achieved under denaturing conditions by h.p.l.c. as detailed in Materials and methods. The amino acid sequences determined for the light chain (subsequently verified by DNA sequencing) were only partly consistent with those previously published (Robinson and Hash, 1982). In addition, we isolated fragment BII_b from crude toxin preparations according to Bizzini *et al.* (1977). Edman sequencing revealed the identity of this fragment with fragment C.

Cloning strategy and sequencing of the tetanus toxin gene

A pool of 32 heptadecamernucleotides was synthesized which represented all possible DNA sequences predicted for the N-terminal amino acids 8–13 of fragment C excluding the third nucleotide residue of the Ile codon (Figure 1). This pool served as a probe to screen plasmid preparations from various toxigenic and non-toxigenic variants of the Massachusetts strain of *C. tetani*. A plasmid (pE88) of ~75 kbp was isolated from the toxigenic non-sporulating variant E88 and characterized by restriction with endonucleases (Figure 2). Southern blot analyses revealed that a 2.0-kbp *Eco*RI fragment of pE88, designated tet7, was specifically recognized by the 5′- ^{32}P-labelled oligonucleotide mixture. Fragment tet7 was cloned under L3-B1 containment in *E. coli* (registered by ZKBS 755/1) using pUC19 (Yanisch-Perron *et al.*, 1985) as a vector. Toxicity tests of hybridization-positive transformants (including those described below) were performed by intramuscular injections of bacterial lysates into the gastrocnemius muscle of mice. None of the transformants induced signs of tetanus in the test animals.

DNA sequencing (Sanger *et al.*, 1977) of tet7 DNA subcloned into M13mp18/19 proved that this fragment contained the coding sequence of the N terminus of fragment C. Therefore, tet7 was used to screen further libraries obtained from pE88 by digestion with restriction endonucleases *Hpa*II, *Bgl*II and *Eco*RI plus *Hin*dIII, respectively.

© IRL Press Limited, Oxford, England

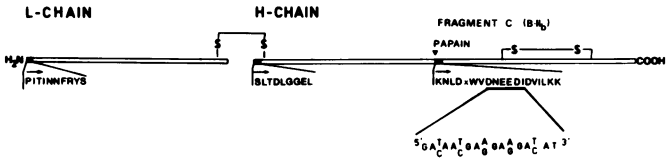

Fig. 1. Structural model of tetanus toxin and partial amino terminal sequences of individual subfragments. Partial amino acid sequences of the L- and H-chains as well as of fragments C and BII_b were determined by Edman degradation. A pool of synthetic oligonucleotides corresponding to amino acids 8–13 of fragment C was used to identify the gene encoding tetanus toxin.

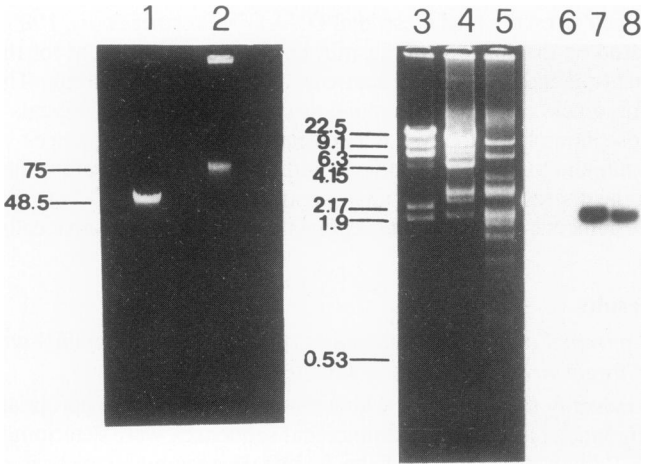

Fig. 2. Characterization of plasmid DNA from a toxigenic non-sporulating Massachusetts strain of *C. tetani*. Plasmid DNA was isolated from strain E88 and analysed on a 0.8% agarose gel. **Lane 1**: lambda DNA as a length marker. **Lane 2**: pE88 DNA digested with *Bam*HI. Restriction fragments obtained by simultaneous digestion with *Sac*I/*Eco*RI (**lane 4**) or *Hin*dIII/*Eco*RI (**lane 5**) were separated together with *Hin*dIII-digested lambda markers (**lane 3**) on a 1% agarose gel and analyzed by Southern hybridization (**lanes 6–8**) using the 5′-^{32}P-labelled oligonucleotide mixture as a probe. Hybridization conditions are described in Materials and methods.

Eight overlapping clones, together spanning >6000 nucleotides were sequenced on both strands and used to establish the toxin gene sequence (Figure 3). They contained a single open reading frame coding for 1315 amino acids, corresponding to a polypeptide with a mol. wt of 150 700. The DNA sequence and the deduced amino acid sequence of tetanus toxin are presented in Figure 4. The coding sequence begins with a codon for methionine which is absent from the mature protein. The second codon specifies proline, the N-terminal residue of the L-chain of tetanus toxin as determined in this and previous studies (Figure 1; Robinson and Hash, 1982).

The toxin gene had an [A+T] content of 72.1% within the coding region and an average of ~81% in the 5′- and 3′-non-coding regions. These findings are in agreement with previous determinations of the [A+T] content of clostridial DNA (Marmur *et al.*, 1963).

Inverted repeat structures in the 5′- and 3′-untranslated regions

The 5′-untranslated region was analyzed for the presence of regulatory sequences. A Shine–Dalgarno consensus sequence [AGGAGA] (Shine and Dalgarno, 1975) was found 14 nucleotides upstream from the initiator ATG codon (Figure 4). A −10 region common to *E. coli* and σ^{55}-dependent *Bacillus subtilis* promoters (Rosenberg and Court, 1979) was present 83 nucleotides upstream from the translation initiation start codon. However,

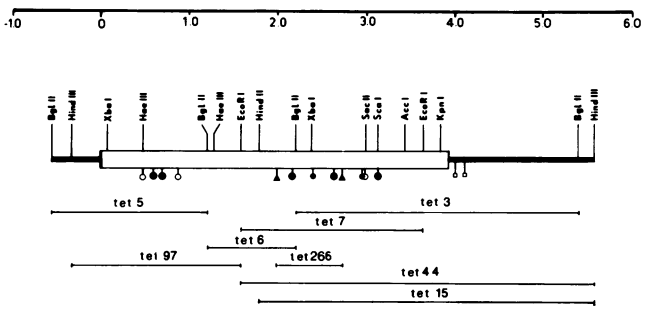

Fig. 3. Restriction map and cloning strategy of the tetanus toxin gene. A 2.0-kb *Eco*RI fragment (tet7) and a 0.75-kb *Hpa*II fragment (tet266) were first recognized by hybridization with the oligonucleotide probe. These fragments were cloned under L3B1 containment in *E. coli* and used to screen pUC12/13 libraries of pE88. Clones tet3, tet5 and tet6 were obtained from a *Bgl*II library and clones tet44 and tet 97 were obtained as *Eco*RI/*Hin*dIII clones. In addition to singular restriction sites shown above the gene map, the following multiple restriction sites were used for subcloning in M13 vectors: *Dra*I (○), *Rsa*I (●), *Hpa*II (▲), *Mae*I(□) and *Sau*3AI (*). The dideoxy chain termination method was used to determine overlapping sequences from both strands. The position of the start (ATG) and stop (TAA) codons corresponding to positions +1 and +3946, respectively, on the scale bar, flank a single open reading frame of 1315 amino acids (open box).

no corresponding −35 region [TTGACA] was detected. In addition, at positions −97 and −63, respectively, sequences similar to the −35 [T-AGGAGA--A] and the −10 [TTT-TTT] σ^{gp28} regions used in virus-infected *B. subtilis* cells (Losick and Pero, 1981) were found. However, without additional experimental data, no precise predictions can be made concerning the transcription initiation site.

Furthermore, it is not clear whether the three predicted inverted repeat strucures in the 5′-flanking region play a regulatory role in the transcription of the toxin gene. It is likely that these stem–loop structures serve as transcription termination signals for a gene positioned further upstream. An open reading frame for a 21 000 mol. wt polypeptide of unknown function was found to be terminated by the TAG codon, 233 nucleotides upstream from the toxin ATG codon (Figure 4).

Sequences with similar dyad symmetry were not detected within the coding sequence of the toxin gene. However, three inverted repeat structures were located immediately downstream from the toxin termination codon (Figure 4). The free energy values calculated for the base-paired regions by the method of Tinoco *et al.* (1973) suggest that these stem–loop structures constitute transcription termination signals of the toxin gene.

Expression of individual domains of the toxin gene in E. coli

Synthesis of a light-chain-specific peptide was first achieved from the pUC recombinant pEJ5 containing 548 bp of the authentic 5′-non-coding region of the toxin gene. In this construct the *lac* promoter of pUC12 was located distally from the toxin ATG codon. Expression from this construct should result in the formation of a 47 399 mol. wt polypeptide. As shown in Figure 5 (lane c), a species of this size was detected in pEJ5-transformants.

To reduce further the size of the putative promoter region, we have studied the expression from pEJ97. In this instance the authentic 5′-non-coding region extended 322 bp upstream from the ATG codon. The coding capacity of the open reading frame corresponded to a 64 444 mol. wt polypeptide. As shown in Figure 5 (lane b), a polypeptide of the predicted size was indeed synthesized. This peptide was recognized by L-chain-specific monoclonal antibodies (not shown) as well as by the H-chain-specific antibody 161D6D9E1. The latter antibody obviously

```
                                                                                                  -211
TAGCATTAAA AAAATTAGAA CCTATAGTAA ATAAATTAAT TAATATATAG TTTTTATAAT TTAATTATGA
                                                          -5.4                          -141
ATAATATTCT TAAGATAAAA AGTAAATTTT TAAAAATTTA AATTTTCAGT TTACAAAAAA TAACCTGATT
                                                                                          -71
ATGTTATATG TAATTGTAAA AAACATATAA AAAATCAGAA AAATTTAGGA GGTATATTAT TAATGGATTA
                  -7.8                                                                      -1
AATAATAATT TTTTAATTTA CTTTTGATTA ATAAATATTA AATGTTTATT TTAATTAGGA GATGATACGT
+1                          -8.0          30                                                60
ATG CCA ATA ACC ATA AAT AAT TTT AGA TAT AGT GAT CCT GTT AAT AAT GAT ACA ATT ATT
Met Pro Ile Thr Ile Asn Asn Phe Arg Tyr Ser Asp Pro Val Asn Asn Asp Thr Ile Ile

                                      90                                               120
ATG ATG GAG CCA CCA TAC TGT AAG GGT CTA GAT ATC TAT TAT AAG GCT TTC AAA ATA ACA
Met Met Glu Pro Pro Tyr Cys Lys Gly Leu Asp Ile Tyr Tyr Lys Ala Phe Lys Ile Thr

                                     150                                               180
GAT CGT ATT TGG ATA GTG CCG GAA AGG TAT GAA TTT GGG ACA AAA CCT GAA GAT TTT AAC
Asp Arg Ile Trp Ile Val Pro Glu Arg Tyr Glu Phe Gly Thr Lys Pro Glu Asp Phe Asn

                                     210                                               240
CCA CCA TCT TCA TTA ATA GAA GGT GCA TCT GAG TAT TAC GAT CCA AAT TAT TTA AGG ACT
Pro Pro Ser Ser Leu Ile Glu Gly Ala Ser Glu Tyr Tyr Asp Pro Asn Tyr Leu Arg Thr

                                     270                                               300
GAT TCT GAT AAA GAT AGA TTT TTA CAA ACC ATG GTA AAA CTG TTT AAC AGA ATT AAA AAC
Asp Ser Asp Lys Asp Arg Phe Leu Gln Thr Met Val Lys Leu Phe Asn Arg Ile Lys Asn

                                     330                                               360
AAT GTA GCA GGT GAA GCC TTA TTA GAT AAG ATA ATA AAT GCC ATA CCT TAC CTT GGA AAT
Asn Val Ala Gly Glu Ala Leu Leu Asp Lys Ile Ile Asn Ala Ile Pro Tyr Leu Gly Asn

                                     390                                               420
TCA TAT TCC TTA CTA GAC AAG TTT GAT ACA AAC TCT AAT TCA GTA TCT TTT AAT TTA TTA
Ser Tyr Ser Leu Leu Asp Lys Phe Asp Thr Asn Ser Asn Ser Val Ser Phe Asn Leu Leu

                                     450                                               480
GAA CAA GAC CCC AGT GGA GCA ACT ACA AAA TCA GCA ATG CTG ACA AAT TTA ATA ATA TTT
Glu Gln Asp Pro Ser Gly Ala Thr Thr Lys Ser Ala Met Leu Thr Asn Leu Ile Ile Phe

                                     510                                               540
GGA CCT GGG CCT GTT TTA AAT AAA AAT GAG GTT AGA GGT ATT GTA TTG AGG GTA GAT AAT
Gly Pro Gly Pro Val Leu Asn Lys Asn Glu Val Arg Gly Ile Val Leu Arg Val Asp Asn

                                     570                                               600
AAA AAT TAC TTC CCA TGT AGA GAT GGT TTT GGC TCA ATA ATG CAA ATG GCA TTT TGC CCA
Lys Asn Tyr Phe Pro Cys Arg Asp Gly Phe Gly Ser Ile Met Gln Met Ala Phe Cys Pro

                                     630                                               660
GAA TAT GTA CCT ACC TTT GAT AAT GTA ATA GAA AAT ATT ACG TCA CTC ACT ATT GGC AAA
Glu Tyr Val Pro Thr Phe Asp Asn Val Ile Glu Asn Ile Thr Ser Leu Thr Ile Gly Lys

                                     690                                               720
AGC AAA TAT TTT CAA GAT CCA GCA TTA CTA TTA ATG CAC GAA CTT ATA CAT GTA CTA CAT
Ser Lys Tyr Phe Gln Asp Pro Ala Leu Leu Leu Met His Glu Leu Ile His Val Leu His

                                     750                                               780
GGT TTA TAC GGA ATG CAG GTA TCA AGC CAT GAA ATT ATT CCA TCC AAA CAA GAA ATT TAT
Gly Leu Tyr Gly Met Gln Val Ser Ser His Glu Ile Ile Pro Ser Lys Gln Glu Ile Tyr

                                     810                                               840
ATG CAG CAT ACA TAT CCA ATA AGT GCT GAA GAA CTA TTC ACT TTT GGC GGA CAG GAT GCT
Met Gln His Thr Tyr Pro Ile Ser Ala Glu Glu Leu Phe Thr Phe Gly Gly Gln Asp Ala

                                     870                                               900
AAT CTT ATA AGT ATT GAT ATA AAA AAC GAT TTA TAT GAA AAA ACT TTA AAT GAT TAT AAA
Asn Leu Ile Ser Ile Asp Ile Lys Asn Asp Leu Tyr Glu Lys Thr Leu Asn Asp Tyr Lys

                                     930                                               960
GCT ATA GCT AAC AAA CTT AGT CAA GTC ACT AGC TGC AAT GAT CCC AAC ATT GAT ATT GAT
Ala Ile Ala Asn Lys Leu Ser Gln Val Thr Ser Cys Asn Asp Pro Asn Ile Asp Ile Asp

                                     990                                              1020
AGC TAC AAA CAA ATA TAT CAA CAA AAA TAT CAA TTC GAT AAA GAT AGC AAT GGA CAA TAT
Ser Tyr Lys Gln Ile Tyr Gln Gln Lys Tyr Gln Phe Asp Lys Asp Ser Asn Gly Gln Tyr

                                    1050                                              1080
ATT GTA AAT GAG GAT AAA TTT CAG ATA CTA TAT AAT AGC ATA ATG TAT GGT TTT ACA GAG
Ile Val Asn Glu Asp Lys Phe Gln Ile Leu Tyr Asn Ser Ile Met Tyr Gly Phe Thr Glu

                                    1110                                              1140
ATT GAA TTG GGA AAA AAA TTT AAT ATA AAA ACT AGA CTT TCT TAT TTT AGT ATG AAT CAT
Ile Glu Leu Gly Lys Lys Phe Asn Ile Lys Thr Arg Leu Ser Tyr Phe Ser Met Asn His

                                    1170                                              1200
GAC CCT GTA AAA ATT CCA AAT TTA TTA GAT GAT ACA ATT TAC AAT GAT ACA GAA GGA TTT
Asp Pro Val Lys Ile Pro Asn Leu Leu Asp Asp Thr Ile Tyr Asn Asp Thr Glu Gly Phe

                                    1230                                              1260
AAT ATA GAA AGC AAA GAT CTG AAA TCT GAA TAT AAA GGA CAA AAT ATG AGG GTA AAT ACA
Asn Ile Glu Ser Lys Asp Leu Lys Ser Glu Tyr Lys Gly Gln Asn Met Arg Val Asn Thr

                                    1290                                              1320
AAT GCT TTT AGA AAT GTT GAT GGA TCA GGC CTA GTT TCA AAA CTT ATT GGC TTA TGT AAA
Asn Ala Phe Arg Asn Val Asp Gly Ser Gly Leu Val Ser Lys Leu Ile Gly Leu Cys Lys

                                    1350                                              1380
AAA ATT ATA CCA CCA ACA AAT ATA AGA GAA AAT TTA TAT AAT AGA ACT GCA TCA TTA ACA
Lys Ile Ile Pro Pro Thr Asn Ile Arg Glu Asn Leu Tyr Asn Arg Thr Ala Ser Leu Thr

                                    1410                                              1440
GAT TTA GGA GGA GAA TTA TGT ATA AAA ATT AAA AAT GAA GAT TTA ACT TTT ATA GCT GAA
Asp Leu Gly Gly Glu Leu Cys Ile Lys Ile Lys Asn Glu Asp Leu Thr Phe Ile Ala Glu
```

```
                                        1470                                    1500
AAA AAT AGC TTT TCA GAA GAA CCA TTT CAA GAT GAA ATA GTT AGT TAT AAT ACA AAA AAT
Lys Asn Ser Phe Ser Glu Glu Pro Phe Gln Asp Glu Ile Val Ser Tyr Asn Thr Lys Asn

                                        1530                                    1560
AAA CCA TTA AAT TTT AAT TAT TCG CTA GAT AAA ATT ATT GTA GAT TAT AAT CTA CAA AGT
Lys Pro Leu Asn Phe Asn Tyr Ser Leu Asp Lys Ile Ile Val Asp Tyr Asn Leu Gln Ser

                                        1590                                    1620
AAA ATT ACA TTA CCT AAT GAT AGG ACA ACC CCA GTT ACA AAA GGA ATT CCA TAT GCT CCA
Lys Ile Thr Leu Pro Asn Asp Arg Thr Thr Pro Val Thr Lys Gly Ile Pro Tyr Ala Pro

                                        1650                                    1680
GAA TAT AAA AGT AAT GCT GCA AGT ACA ATA GAA ATA CAT AAT ATT GAT GAC AAT ACA ATA
Glu Tyr Lys Ser Asn Ala Ala Ser Thr Ile Glu Ile His Asn Ile Asp Asp Asn Thr Ile

                                        1710                                    1740
TAT CAA TAT TTG TAT GCT CAA AAA TCT CCT ACA ACT CTA CAA AGA ATA ACT ATG ACT AAT
Tyr Gln Tyr Leu Tyr Ala Gln Lys Ser Pro Thr Thr Leu Gln Arg Ile Thr Met Thr Asn

                                        1770                                    1800
TCT GTT GAT GAC GCA TTA ATA AAT TCC ACC AAA ATA TAT TCA TAT TTT CCA TCT GTA ATC
Ser Val Asp Asp Ala Leu Ile Asn Ser Thr Lys Ile Tyr Ser Tyr Phe Pro Ser Val Ile

                                        1830                                    1860
AGT AAA GTT AAC CAA GGT GCA CAA GGA ATT TTA TTC TTA CAG TGG GTG AGA GAT ATA ATT
Ser Lys Val Asn Gln Gly Ala Gln Gly Ile Leu Phe Leu Gln Trp Val Arg Asp Ile Ile

                                        1890                                    1920
GAT GAT TTT ACC AAT GAA TCT TCA CAA AAA ACT ACT ATT GAT AAA ATT TCA GAT GTA TCC
Asp Asp Phe Thr Asn Glu Ser Ser Gln Lys Thr Thr Ile Asp Lys Ile Ser Asp Val Ser

                                        1950                                    1980
ACT ATT GTT CCT TAT ATA GGA CCC GCA TTA AAC ATT GTA AAA CAA GGC TAT GAG GGA AAC
Thr Ile Val Pro Tyr Ile Gly Pro Ala Leu Asn Ile Val Lys Gln Gly Tyr Glu Gly Asn

                                        2010                                    2040
TTT ATA GGC GCT TTA GAA ACT ACC GGA GTG GTT TTA TTA TTA GAA TAT ATT CCA GAA ATT
Phe Ile Gly Ala Leu Glu Thr Thr Gly Val Val Leu Leu Leu Glu Tyr Ile Pro Glu Ile

                                        2070                                    2100
ACT TTA CCA GTA ATT GCA GCT TTA TCT ATA GCA GAA AGT AGC ACA CAA AAA GAA AAG ATA
Thr Leu Pro Val Ile Ala Ala Leu Ser Ile Ala Glu Ser Ser Thr Gln Lys Glu Lys Ile

                                        2130                                    2160
ATA AAA ACA ATA GAT AAC TTT TTA GAA AAA AGA TAT GAA AAA TGG ATT GAA GTA TAT AAA
Ile Lys Thr Ile Asp Asn Phe Leu Glu Lys Arg Tyr Glu Lys Trp Ile Glu Val Tyr Lys

                                        2190                                    2220
CTA GTA AAA GCA AAA TGG TTA GGC ACA GTT AAT ACG CAA TTC CAA AAA AGA AGT TAT CAA
Leu Val Lys Ala Lys Trp Leu Gly Thr Val Asn Thr Gln Phe Gln Lys Arg Ser Tyr Gln

                                        2250                                    2280
ATG TAT AGA TCT TTA GAA TAT CAA GTA GAT GCA ATA AAA AAA ATA ATA GAC TAT GAA TAT
Met Tyr Arg Ser Leu Glu Tyr Gln Val Asp Ala Ile Lys Lys Ile Ile Asp Tyr Glu Tyr

                                        2310                                    2340
AAA ATA TAT TCA GGA CCT GAT AAG GAA CAA ATT GCC GAC GAA ATT AAT AAT CTG AAA AAC
Lys Ile Tyr Ser Gly Pro Asp Lys Glu Gln Ile Ala Asp Glu Ile Asn Asn Leu Lys Asn

                                        2370                                    2400
AAA CTT GAA GAA AAG GCT AAT AAA GCA ATG ATA AAC ATA AAT ATA TTT ATG AGG GAA AGT
Lys Leu Glu Glu Lys Ala Asn Lys Ala Met Ile Asn Ile Asn Ile Phe Met Arg Glu Ser

                                        2430                                    2460
TCT AGA TCA TTT TTA GTT AAT CAA ATG ATT AAC GAA GCT AAA AAG CAG TTA TTA GAG TTT
Ser Arg Ser Phe Leu Val Asn Gln Met Ile Asn Glu Ala Lys Lys Gln Leu Leu Glu Phe

                                        2490                                    2520
GAT ACT CAA AGC AAA AAT ATT TTA ATG CAG TAT ATA AAA GCA AAT TCT AAA TTT ATA GGT
Asp Thr Gln Ser Lys Asn Ile Leu Met Gln Tyr Ile Lys Ala Asn Ser Lys Phe Ile Gly

                                        2550                                    2580
ATA ACT GAA CTA AAA AAA TTA GAA TCA AAA ATA AAC AAA GTT TTT TCA ACA CCA ATT CCA
Ile Thr Glu Leu Lys Lys Leu Glu Ser Lys Ile Asn Lys Val Phe Ser Thr Pro Ile Pro

                                        2610                                    2640
TTT TCT TAT TCT AAA AAT CTG GAT TGT TGG GTT GAT AAT GAA GAA GAT ATA GAT GTT ATA
Phe Ser Tyr Ser Lys Asn Leu Asp Cys Trp Val Asp Asn Glu Glu Asp Ile Asp Val Ile
                ▲
                                        2670                                    2700
TTA AAA AAG AGT ACA ATT TTA AAT TTA GAT ATT AAT AAT GAT ATT ATA TCA GAT ATA TCT
Leu Lys Lys Ser Thr Ile Leu Asn Leu Asp Ile Asn Asn Asp Ile Ile Ser Asp Ile Ser

                                        2730                                    2760
GGG TTT AAT TCA TCT GTA ATA ACA TAT CCA GAT GCT CAA TTG GTG CCC GGA ATA AAT GGC
Gly Phe Asn Ser Ser Val Ile Thr Tyr Pro Asp Ala Gln Leu Val Pro Gly Ile Asn Gly

                                        2790                                    2820
AAA GCA ATA CAT TTA GTA AAC AAT GAA TCT TCT GAA GTT ATA GTG CAT AAA GCT ATG GAT
Lys Ala Ile His Leu Val Asn Asn Glu Ser Ser Glu Val Ile Val His Lys Ala Met Asp

                                        2850                                    2880
ATT GAA TAT AAT GAT ATG TTT AAT AAT TTT ACC GTT AGC TTT TGG TTG AGG GTT CCT AAA
Ile Glu Tyr Asn Asp Met Phe Asn Asn Phe Thr Val Ser Phe Trp Leu Arg Val Pro Lys

                                        2910                                    2940
GTA TCT GCT AGT CAT TTA GAA CAA TAT GGC ACA AAT GAG TAT TCA ATA ATT AGC TCT ATG
Val Ser Ala Ser His Leu Glu Gln Tyr Gly Thr Asn Glu Tyr Ser Ile Ile Ser Ser Met

                                        2970                                    3000
AAA AAA CAT AGT CTA TCA ATA GGA TCT GGT TGG AGT GTA TCA CTT AAA GGT AAT AAC TTA
Lys Lys His Ser Leu Ser Ile Gly Ser Gly Trp Ser Val Ser Leu Lys Gly Asn Asn Leu

                                        3030                                    3060
ATA TGG ACT TTA AAA GAT TCC GCG GGA GAA GTT AGA CAA ATA ACT TTT AGG GAT TTA CCT
Ile Trp Thr Leu Lys Asp Ser Ala Gly Glu Val Arg Gln Ile Thr Phe Arg Asp Leu Pro
```

```
                                        3090                                    3120
GAT AAA TTT AAT GCT TAT TTA GCA AAT AAA TGG GTT TTT ATA ACT ATT ACT AAT GAT AGA
Asp Lys Phe Asn Ala Tyr Leu Ala Asn Lys Trp Val Phe Ile Thr Ile Thr Asn Asp Arg

                                        3150                                    3180
TTA TCT TCT GCT AAT TTG TAT ATA AAT GGA GTA CTT ATG GGA AGT GCA GAA ATT ACT GGT
Leu Ser Ser Ala Asn Leu Tyr Ile Asn Gly Val Leu Met Gly Ser Ala Glu Ile Thr Gly

                                        3210                                    3240
TTA GGA GCT ATT AGA GAG GAT AAT AAT ATA ACA TTA AAA CTA GAT AGA TGT AAT AAT AAT
Leu Gly Ala Ile Arg Glu Asp Asn Asn Ile Thr Leu Lys Leu Asp Arg Cys Asn Asn Asn

                                        3270                                    3300
AAT CAA TAC GTT TCT ATT GAT AAA TTT AGG ATA TTT TGC AAA GCA TTA AAT CCA AAA GAG
Asn Gln Tyr Val Ser Ile Asp Lys Phe Arg Ile Phe Cys Lys Ala Leu Asn Pro Lys Glu

                                        3330                                    3360
ATT GAA AAA TTA TAC ACA AGT TAT TTA TCT ATA ACC TTT TTA AGA GAC TTC TGG GGA AAC
Ile Glu Lys Leu Tyr Thr Ser Tyr Leu Ser Ile Thr Phe Leu Arg Asp Phe Trp Gly Asn

                                        3390                                    3420
CCT TTA CGA TAT GAT ACA GAA TAT TAT TTA ATA CCA GTA GCT TCT AGT TCT AAA GAT GTT
Pro Leu Arg Tyr Asp Thr Glu Tyr Tyr Leu Ile Pro Val Ala Ser Ser Ser Lys Asp Val

                                        3450                                    3480
CAA TTG AAA AAT ATA ACA GAT TAT ATG TAT TTG ACA AAT GCG CCA TCG TAT ACT AAC GGA
Gln Leu Lys Asn Ile Thr Asp Tyr Met Tyr Leu Thr Asn Ala Pro Ser Tyr Thr Asn Gly

                                        3510                                    3540
AAA TTG AAT ATA TAT TAT AGA AGG TTA TAT AAT GGA CTA AAA TTT ATT ATA AAA AGA TAT
Lys Leu Asn Ile Tyr Tyr Arg Arg Leu Tyr Asn Gly Leu Lys Phe Ile Ile Lys Arg Tyr

                                        3570                                    3600
ACA CCT AAT AAT GAA ATA GAT TCT TTT GTT AAA TCA GGT GAT TTT ATT AAA TTA TAT GTA
Thr Pro Asn Asn Glu Ile Asp Ser Phe Val Lys Ser Gly Asp Phe Ile Lys Leu Tyr Val

                                        3630                                    3660
TCA TAT AAC AAT AAT GAG CAC ATT GTA GGT TAT CCG AAA GAT GGA AAT GCC TTT AAT AAT
Ser Tyr Asn Asn Asn Glu His Ile Val Gly Tyr Pro Lys Asp Gly Asn Ala Phe Asn Asn

                                        3690                                    3720
CTT GAT AGA ATT CTA AGA GTA GGT TAT AAT GCC CCA GGT ATC CCT CTT TAT AAA AAA ATG
Leu Asp Arg Ile Leu Arg Val Gly Tyr Asn Ala Pro Gly Ile Pro Leu Tyr Lys Lys Met

                                        3750                                    3780
GAA GCA GTA AAA TTG CGT GAT TTA AAA ACC TAT TCT GTA CAA CTT AAA TTA TAT GAT GAT
Glu Ala Val Lys Leu Arg Asp Leu Lys Thr Tyr Ser Val Gln Leu Lys Leu Tyr Asp Asp

                                        3810                                    3840
AAA AAT GCA TCT TTA GGA CTA GTA GGT ACC CAT AAT GGT CAA ATA GGC AAC GAT CCA AAT
Lys Asn Ala Ser Leu Gly Leu Val Gly Thr His Asn Gly Gln Ile Gly Asn Asp Pro Asn

                                        3870                                    3900
AGG GAT ATA TTA ATT GCA AGC AAC TGG TAC TTT AAT CAT TTA AAA GAT AAA ATT TTA GGA
Arg Asp Ile Leu Ile Ala Ser Asn Trp Tyr Phe Asn His Leu Lys Asp Lys Ile Leu Gly
                                                                                3960
                                        3930
TGT GAT TGG TAC TTT GTA CCT ACA GAT GAA GGA TGG ACA AAT GAT TAA ACA GAT TGA TAT
Cys Asp Trp Tyr Phe Val Pro Thr Asp Glu Gly Trp Thr Asn Asp End

                                                                                4030
GTTCATGATT ACTCTATATA AAAAATTAAA TAATATAACA ATCTAGCTAT ATTATTTTTG ATTATTTTCT
                                                  -6.4
                                                                                4100
TAATATATAC TAATAAAATA ATCAAAATAG AGCCTATCTT AAATTACTGA AGGGCTGTGT CAAAATAAGA
   -10.2
                                                                          -22.0
TTTTGACACA GCCTCTACTT
```

Fig. 4. Nucleotide sequence of the tetanus toxin gene and the deduced amino acid sequence. The arrow ↓ between Ala(457) and Ser(458) indicates the start of the heavy chain in agreement with the data obtained by Edman degradation. ▲ indicates the papain cleavage site. Non-coding sequences exhibiting dyad symmetry and representing potential transcription termination signals are underlined by divergent arrows. Numbers represent free energy values in kcal/mol. The putative Shine–Dalgarno sequence is indicated by the boxed region. The TAG termination codon of the preceding 21 000 mol. wt protein is underlined by a dotted line.

recognizes an epitope close to the N terminus of the H-chain. A hydrophilicity analysis of this region according to Hopp and Woods (1981) suggested that this epitope could be created through an exposure of the hydrophilic domain extending from Lys(469) to Thr(476) at the periphery of the toxin molecule.

No expression was obtained when the 5′-non-coding region was deleted from the *Hin*dIII to the *Sna*BI site immediately upstream from the ATG codon (not shown).

Expression from pEJ6 should result in the synthesis of a fused polypeptide consisting of 11 N-terminal amino acids from the *lacZ′*-specified α-peptide, 339 toxin-specific amino acids and 18 C-terminal amino acids derived from the polylinker sequence. The size of the largest toxin-specific peptide detected in lysates of pEJ6 transformants (lane d) is consistent with the 41 583 mol. wt species predicted for this construct.

Lanes e and f of Figure 5 show the expression of peptides corresponding to the carboxy-terminal domain of tetanus toxin. IPTG induction of *E. coli* transformants harbouring pHKtet15 gave rise to the synthesis of a 82 600 mol. wt protein carrying eight foreign N-terminal amino acids (lane e′). This peptide covers 84% of the heavy chain. Expression from pExtet3 carrying the *Bgl*II tet 3 fragment (Figure 3) downstream from the temperature-sensitive lambda P_L-promoter (Remaut *et al.*, 1981) was exclusively observed at permissive temperature (Figure 5, lane f′). In this instance, a 77 254 mol. wt fusion protein was formed which contained 104 foreign N-terminal amino acids (Strebel *et al.*, 1986). In all situations considerable amounts of smaller size peptides were detected, even if protease-negative *E. coli* strains were used for the expression.

Discussion

We have presented here the structure of tetanus toxin as deduced from the complete nucleotide sequence and from partial amino

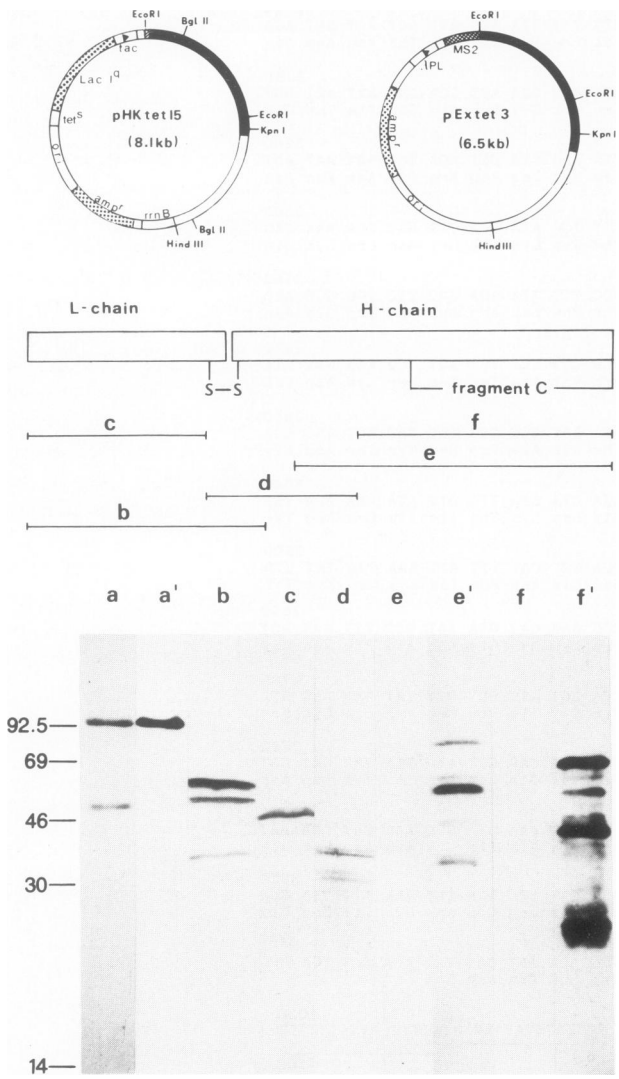

Fig. 5. Expression of various domains of tetanus toxin in *E. coli*. Three different plasmid vectors were used to express toxin-specific fragments as shown in the middle panel. Fragments b, c, and d were expressed from pUC vectors as described in detail in Materials and methods. Note that expression of peptides b and c proceeded from the authentic toxin promoter, while peptide d was synthesized as a *lac* fusion protein. Expression of fragments e and f was obtained from pHKtet15 and pExtet3 as shown in the upper panel (for further details see Materials and methods). *E. coli* strains harbouring recombinant plasmids were grown in liquid cultures, induced where required and lysed. Aliquots were analysed under reducing conditions together with toxin standards on 10% SDS–polyacrylamide slab gels. After transfer to nitrocellulose sheets, polypeptides were identified in the individual lanes with monoclonal antibodies as follows: **tracks a′**, **b** and **d**: H-chain-specific 161D6D9E1 antibody: **track c**: L-chain-specific antibody 99305II2A1; **tracks e**, **e′**, **f** and **f′**: C-fragment-specific antibody 164B4F8C4. **Track a** shows reduced toxin, stained wth Coomassie blue. **Lanes e** and **f** show non-induced bacterial lysates, **lanes e′** and **f′** show the corresponding lysates obtained 90 min after induction with IPTG (**lane e′**) or temperature shift to 42°C (**lane f′**). The positions of mol. wt markers are indicated.

acid sequences. To our knowledge this is the first report of the complete sequence of a clostridial neurotoxin.

Synthetic oligonucleotides corresponding to the N-terminal amino acid sequence of fragment C served as a hybridization probe to identify the structural gene for tetanus toxin on a 75-kbp plasmid from a toxigenic strain of *C. tetani*. Similar conclusions regarding the extrachromosomal association of the toxin gene

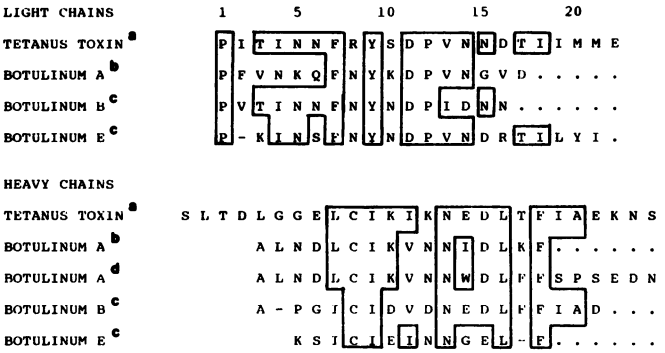

Fig. 6. Structural homologies between tetanus toxin and botulinum toxins A, B and E found in the N-terminal amino acid sequences of the respective L- and H-chains. Amino acids common to tetanus and botulinum neurotoxins are framed with solid lines. (**a**) Sequence presented in this paper, (**b**) Schmidt *et al.*, (1984); (**c**) Schmidt *et al.*, 1985; (**d**) Shone *et al.*, (1985).

have been drawn previously by others (Hara *et al.*, 1977; Laird *et al.*, 1980; Finn *et al.*, 1984). While this work was in progress, Fairweather *et al.* (1986) reported an approach similar to that reported here to obtain partial sequences of the heavy chain. In these studies, however, total clostridial DNA was used.

The extrachromosomal nature of a toxin gene is not a unique feature of tetanus toxin. Diphtheria toxin, for example, is the product of a gene present in the genome of lysogenic corynephages infecting the Gram-positive bacterium *Corynebacterium diphtheriae*. In addition, the genes of the type C and D botulinum toxins have been reported to reside extrachromosomally on the genomes of transducing bacteriophages (Eklund and Poysky, 1981; Iida and Oguma, 1981). It remains to be shown whether the tetanus toxin gene is exclusively located on the 75-kbp plasmid, as shown here, or whether it can exist as part of the bacterial chromosome.

For safety reasons, in order to avoid cloning the entire toxin gene into pUC vectors, only agarose gel-purified fragments of 3.0 kbp and smaller were cloned under L3B1 containment until a complete restriction map of the gene was established. In addition, toxicity tests were performed in mice to exclude the expression of potentially toxic subfragments.

The deduced amino acid sequence of tetanus toxin revealed several features which deserve discussion in context with published data.

No significant modifications of the N terminus of the L-chain take place when the toxin is discharged from *C. tetani*. Only the N-terminal methionine residue is removed from the polypeptide. There is no indication of a signal peptide for secretion as commonly found in secretory proteins from Gram-positive bacteria (Löfdahl *et al.*, 1983). The sequence is in agreement with reports indicating that the toxin contains six free SH groups and two disulfide bridges (Bizzini *et al.*, 1977). Since Cys(467) is the only cysteine residue within the N-terminal domain of the H-chain, it is now clear that a single disulfide bond involving this amino acid connects the L- and the H-chain in the nicked molecule.

Ser(458) forms the N terminus of the H-chain. The mol. wts of the resulting L-chain (52 288) and H-chain (98 300) are close to the values determined for both subunits by other methods. We cannot exclude however, that primary cleavage occurs at a different peptide bond and that the N-terminal Ser is the result of the subsequent action of proteases. Lys(865) forms the N terminus of fragment C, in agreement with the results of Fairweather *et al.* (1986). The fragment is identical with fragment BII_b isolated from crude toxin preparations. This indicates that a

cleavage process similar to that catalyzed by papain *in vitro* occurs during storage of the toxin.

In contrast to a previous hypothesis, no duplications of individual sequences are detected (Taylor *et al.*, 1983).

A hydrophilicity analysis according to Hopp and Woods (1981) indicates no domains of striking hydrophobicity that would unambiguously demonstrate an interaction with membranes and thus signal an involvement of this part of the molecule in the internalization process of the toxin. Shorter hydrophobic regions involving up to 10 uncharged amino acids in uninterrupted sequence are randomly distributed over the entire molecule. Only two domains were long enough and sufficiently hydrophobic to potentially span mammalian cell membranes (Wickner and Lodish, 1985). The first, located on the L-chain, extends from Tyr(223) to Ile(253). This sequence contains four of the six His residues present on the L-chain. The second predicted hydrophobic domain involves the sequence from Asn(660) to Ala(691). Each of the regions contains three acidic amino acids. Boquet and Duflot (1982) have shown that fragment B (i.e. tetanus toxin minus fragment C) is capable of forming channels in asolectin vesicles at a pH below 5.0. Neither fragment C nor the isolated L-chain revealed a similar channel-forming capacity (Boquet and Duflot, 1982). Treatment of such vesicles with proteases yielded resistant peptides of 21 000 and 27 000 mol. wt (Roa and Boquet, 1985). It remains to be shown whether the hydrophobic amino acid sequences predicted from the nucleotide sequence are capable of interacting with each other or with still unidentified sequences of the toxin molecule to create channels in a pH-dependent manner.

The search against a protein identification resource library containing 3600 proteins with >800 000 amino acids did not provide significant homologies between the toxin and known proteins. However, an alignment of the N-terminal sequences of the L- and H-chains of tetanus toxin with partial sequences available from the corresponding chains of botulinum toxins A, B, and E revealed a striking homology (Figure 6).

Interestingly, Cys(467), which participates in the disulfide linkage between the two subunits of tetanus toxin, can be aligned with cysteine residues in the three botulinum toxins. Considering the equivalence of certain amino acids in specific positions, and taking into account that most of the mutations depicted in Figure 6 would involve only single base changes, it may be concluded that tetanus and botulinum neurotoxins are not only pharmacologically but also structurally closely related and presumably stem from a common ancestral gene. Future elucidation of the structure and the genetic relationship and the structure of the botulinum neurotoxins might help to explain why the two closely related neurotoxins display clinically different pictures.

There are three lines of evidence suggesting that tetanus toxin is translated from a monocistronic mRNA in Clostridia. Firstly, as shown here, a region exhibiting some of the features of prokaryotic promoters was found upstream from the ATG codon. Secondly, *E. coli* cells that were transformed with recombinants containing this upstream region but lacking an exogeneous promoter, synthesized toxin-specific polypeptides of the expected size. Similar observations regarding the promoter function have been made with the promoter of diphtheria toxin (Kaczorek *et al.*, 1985). Thirdly, the toxin gene is both preceded and followed by palindromic sequences indicating transcription termination. We are currently mapping the precise start and termination sites.

To ultimately assign functional domains involved in binding, ascent and toxicity at a molecular level, we have expressed five overlapping peptides of tetanus toxin in *E. coli*. Four different promoters including the authentic toxin promoter have been tested for this purpose. None of them, however, led to the synthesis of sufficient amounts of toxin-specific protein to allow biological studies. The finding that neither the *trp−lac* fusion promoter nor the P_L promoter of bacteriophage lambda gave rise to the formation of considerable quantities of protein, is paralleled by similar results obtained by *in vitro* transcription/translation studies (B.Anderson, T.Mayer, E.Habermann and H.Niemann, unpublished). Despite normal quantities of full-length RNA transcribed from SP6 vectors, the efficiency of translation in the wheat germ extract was drastically reduced in comparison with mammalian RNA. In addition, prematurely terminated peptides of similar size as observed in the *E. coli* lysates formed the main products. We ascribe these findings to the codon usage related to the low [G+C] content (27.9% overall; 15.5% in the third codon position) of the toxin gene. Although typical for clostridial genes (Marmur *et al.*, 1963), the particular codon usage reflects all the features commonly found in weakly expressed *E. coli* genes (Grosjean and Fiers, 1982).

At present, these considerations argue against a large-scale production of a toxoid based on a genetically engineered non-toxic derivative. However, expression studies at the analytical level may help to define epitopes and functional domains of the toxin molecule. Such experiments are currently in progress.

Materials and methods

Purification of subfragments of tetanus toxin

Crude tetanus toxin was obtained from Behringwerke (Marburg, FRG). The material was reduced with β-mercaptoethanol and subjected to vinylpyridinylation according to Friedman *et al.* (1970). Separation of the L- and H-chains was achieved by h.p.l.c. on an LKB-Ultropak TSK G 3000 SW column (0.75 × 60 cm) and 50 mM Tris/HCl, pH 6.3, containing 6 M guanidine/HCl as an eluent. Peak fractions monitored by u.v. absorption at 280 nm were pooled, dialysed against distilled water and precipitated material was collected by centrifugation. Purified fragment C was a gift from Dr R.O.Thomson (Wellcome Biotechnology Ltd, Beckenham, UK). Fragment BII_b was purified from crude toxin according to Bizzini *et al.* (1977).

Edman degradation

The instrumentation and the microscale separation of phenylthiohydantoin amino acids have been described previously (Edman and Begg, 1967; Lottspeich 1985).

Bacteria and plasmids

The toxigenic non-sporulating E88 variant of the Massachusetts strain of *C.tetani* was obtained from Dr H.Engelhard (Behringwerke, Marburg, FRG) and stock cultures were stored in skimmed milk at −80°C. Plasmid pE88 was isolated from 24 h cultures of E88 in brain heart infusion broth (Merck, Darmstadt, FRG) following a modification of the procedure of Hansen and Olsen (1978). About 20 μg of plasmid DNA was obtained from a 1 litre batch. For subcloning in *E. coli* strain JM101 (Messing, 1979) pUC12,13 and pUC18,19 vectors (Yanisch-Perron *et al.*, 1985) were used.

Enzymes and nucleotides

Restriction endonucleases, T4 DNA ligase and nucleotides were purchased from Boehringer (Mannheim, FRG), T4 polynucleotide kinase was from PL-Pharmacia (Freiburg, FRG).

Hybridization conditions with synthetic oligonucleotides

Oligonucleotides were synthesized by the phosphoamidite method (Beaucage and Caruthers, 1981) using an Applied Biosystems model 380 A automated device. Conditions for hybridization were established by dot blot hybridization. Filters were pre-hybridized in six times standard saline citrate (SSC), 0.5% SDS, five times Denhardt's and 100 μg of sheared denatured herring sperm DNA per ml at 30°C. Hybridization was performed with 10^7 c.p.m. of the oligonucleotide probe in pre-hybridization fluid for 40 h at 30°C. The filter was washed twice for 5 min at room temperature in 6 × SSC and once more for 30 min at 42°C. Southern analyses involving probes radiolabelled by nick-translation were performed as described by Maniatis *et al.*, (1982).

Recombination and sequencing of DNA

Cloning of pE88 fragments was performed under L3B1 containment following the protocols of Maniatis *et al.*, (1982). The chain termination method was used

according to Sanger *et al.* (1977) using [^{35}S]dCTP (New England Nuclear) as the radiolabelled nucleotide.

Expression of toxin-specific gene fragments in E. coli

Five different constructs were made to express overlapping toxin-specific peptides: pEJ5 contained the *Bgl*II tet5 fragment (1773 bp, extending from positions −548 to 1215) cloned into the *Bam*HI site of pUC12. The orientation was verified with *Eco*RI. A translation termination codon 24 nucleotides downstream from the toxin gene/plasmid fusion site terminates an open reading frame corresponding to a 47 399 mol. wt peptide. pEJ97 contained the *Hin*dIII−*Eco*RI tet97 fragment (1926 bp, extending from positions −322 to 1604) cloned into the corresponding sites of pUC12.

pEJ6 contained the *Bgl*II tet6 fragment (positions 1215−2227) in frame with the α-peptide downstream from the *lac* promoter of pUC12. pHKtet15 (Figure 5) contained the *Hin*dII−*Hin*dIII tet15 fragment (positions 1807−5800) cloned into *Sma*I−*Hin*dIII-digested pHK400 (Amann and Kröger, unpublished). This vector provides the *trp*−*lac* fusion promoter, a Shine−Dalgarno sequence and an ATG start codon upstream from the multiple cloning region from pUC12 (Amann and Brosius, 1986). The presence of *lacI*^{q} is designed to suppress constitutive expression from the *lac* promoter. Induction with IPTG leads to a fusion protein containing eight additional N-terminal amino acids. pExtet3 (Figure 5) contained the *Bgl*II tet3 fragment (positions 2227−5600) cloned in-frame into the *Bam*HI site of pEx31 (Strebel *et al.*, 1986). This construct results in the formation of a polypeptide with 99 N-terminal amino acids from the MS2 polymerase and five amino acids from the 5′ region of the polylinker. 537 cells containing the tsCI repressor on a plasmid were grown at 28°C in the presence of kanamycin. Cells were shifted to 42°C for 90 min and lysed (Strebel *et al.*, 1986).

Monoclonal antibodies

The preparation of monoclonal antibodies 99305II2A1 (L-chain), 161D6D9E1 (H-chain) and 164B4F8C4 (fragment C) has been described previously (Goretzki and Habermann, 1985).

Acknowledgements

We thank Drs E.Amann and M.Kröger for supplying the plasmid vector pHK400; Dr E.Beck for the pEx31 vector, Dr H.Engelhard for providing *C. tetani* strains and Dr R.Thomson for a gift of purified fragment C. We further thank Dr Pat Argos for making computer searches, Dr B.Boschek for critically reading the manuscript and R.Schrenk for laboratory assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 47) and by the Bundesministerium des Innern. H.N. was the recipient of a Habilitandenstipendium from the Deutsche Forschungsgemeinschaft. The work is part of the Ph.D. thesis of U.E.

References

Amann,E. and Brosius,J. (1986) *Gene*, **40**, 183−190.
Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Lett.*, **22**, 1859.
Bizzini,B., Stoeckel,K. and Schwab,M. (1977) *J. Neurochem.*, **28**, 529−542.
Boquet,P. and Duflot,E. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7614−7618.
Edman,P. and Begg,G. (1967) *Eur. J. Biochem.*, **1**, 80−91.
Eklund,M.W. and Poysky,F.T. (1981) In Lewis,G.E. (ed.), *Biomedical Aspects of Botulism*. Academic Press, NY, pp. 93−107.
Fairweather,N.F., Lyness,V.A., Pickard,D.J., Allen,G. and Thomson,R.O. (1986) *J. Bacteriol.*, **165**, 21−27.
Finn,C.W., Silver,R.P., Habig,W.H., Hardegree,M.C., Zon,G. and Garon,C.F. (1984) *Science*, **224**, 881−884.
Friedman,M., Krull,L.H. and Cavins,J.F. (1970) *J. Biol. Chem.*, **245**, 3868−3871.
Goretzki,K. and Habermann,E. (1985) *Med. Microbiol. Immunol.*, **174**, 139−150.
Grosjean,H. and Fiers,W. (1982) *Gene*, **18**, 199−209.
Habermann,E. and Dreyer,F. (1986) *Curr. Top. Microbiol. Immunol.*, **129**, 93−179.
Hansen,J.B. and Olsen,R.H. (1978) *J. Bacteriol.*, **135**, 227−238.
Hara,T., Matsuda,M. and Yoneda,M. (1977) *Biken J.*, **20**, 105−115.
Helting,T.B. and Zwisler,O. (1977) *J. Biol. Chem.*, **252**, 187−193.
Hopp,T.P. and Woods,K.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3824−3828.
Iida,H. and Oguma,K. (1981) In Lewis,G.E. (ed.), *Biomedical Aspects of Botulism*. Academic Press, NY, pp. 109−120.
Kaczorek,M., Zettlmeissl,G., Delpeyroux,F. and Streeck,R.W. (1985) *Nucleic Acids Res.*, **13**, 3147−3159.
Laird,W.J., Aaronsen,W., Silver,R.P., Habig,W.H. and Hardegree,M.C. (1980) *J. Infect. Dis.*, **142**, 623.
Löfdahl,S., Guss,B., Uhlén,M., Philipson,L. and Lindberg,M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 697−701.
Lottspeich,F. (1985) *J. Chromatogr.*, **326**, 321−327.
Losick,R., and Pero,J. (1981) *Cell*, **25**, 582−584.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Marmur,J., Falkow,S. and Mandel,M. (1963) *Annu. Rev. Microbiol.*, **17**, 329−372.
Matsuda,M. and Yoneda,M. (1975) *Infect. Immun.*, **12**, 1147−1153.
Messing,J. (1979) *Recombinant DNA Technical Bulletin*, **2**, 43−48.
Remaut,E., Stanssens,P. and Fiers,W. (1981) *Gene*, **15**, 81−93.
Roa,M. and Boquet,P. (1985) *J. Biol. Chem.*, **260**, 6827−6835.
Robinson,J.P. and Hash,J.H. (1982) *Mol. Cell. Biochem.*, **48**, 33−45.
Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, **13**, 319−353.
Sakaguchi,G. (1983) *Pharmacol. Ther.*, **19**, 165−194.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.
Schmidt,J.J., Sathyamoorthy,V. and DasGupta,B.R. (1984) *Biochem. Biophys. Res. Commun.*, **119**, 900−905.
Schmidt,J.J., Sathyamoorthy,V. and DasGupta,B.R. (1985) *Arch. Biochem.*, **238**, 544−548.
Shine,J. and Dalgarno,L. (1975) *Nature*, **254**, 34−38.
Shone,C.C., Hambleton,P. and Melling,J. (1985) *Eur. J. Biochem.*, **151**, 75−82.
Strebel,K., Beck,E., Strohmaier,K. and Schaller,H. (1986) *J. Virol.*, **57**, 983−991.
Taylor,C.F., Britton,P. and van Heyningen,S. (1983) *Biochem. J.*, **209**, 897−899.
Tinoco,I.J., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature*, **248**, 40−41.
Wickner,W.T. and Lodish,H.F. (1985) *Science*, **230**, 400−407.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103−119.

Received on 22 July 1986