Mol Gen Genet (1995) 247:670–679

ORIGINAL PAPER

Fred Hofmann · Andrea Herrmann · Ernst Habermann
Christoph von Eichel-Streiber

Sequencing and analysis of the gene encoding the α-toxin of *Clostridium novyi* proves its homology to toxins A and B of *Clostridium difficile*

Received: 21 September 1994 / Accepted: 28 November 1994

Abstract A library of total *Clostridium novyi* DNA was established and screened for the α-toxin gene (*tcnα*) by hybridization with oligonucleotides derived from a partial N-terminal sequence and by using specific antisera. Overlapping subgenic *tcnα* fragments were isolated and subsequently the total sequence of *tcnα* was determined. The 6534 nucleotide open reading frame encodes a polypeptide of M_r 250166 and pI 5.9. The N-terminal α-toxin (Tcnα) sequence MLITREQLMKIASIP determined by Edman degradation confirmed the identity of the reading frame and the assignment of the translation start point. The toxin is not modified posttranslationally at its N-terminus nor does it consist of different subunits. Overall the amino acid sequence shows 48% homology between the Tcnα and both toxins A (TcdA) and B (TcdB) of *Clostridium difficile*. The C-terminal 382 residues of Tcnα constitute a repetitive domain similar to those reported for TcdA and TcdB of *C. difficile*. The individual repeat motifs of these three toxins consist of oligopeptides some 19–52 amino acids in length, arranged in four to five different groups. Genetic, biochemical and pharmacological data thus confirm that the three toxins belong to one subgroup, designated large clostridial cytotoxins (LCT). Further definition of their structure and detailed molecular action should allow the LCTs to be used tools for the analysis of microfilament assembly and function.

Key words Virulence factor · Repetitive domain
Large clostridial cytotoxins · Cytoskeleton · Rho

Communicated by W. Goebel

F. Hofmann · A. Herrmann · C. von Eichel-Streiber (✉)
Verfügungsgebäude für Forschung und Entwicklung,
Institut für Medizinische Mikrobiologie und Hygiene
Obere Zahlbacher Str. 63,
Johannes Gutenberg-Universität Mainz,
D-55101 Mainz, Germany

E. Habermann
Klinische Pharmakologie, Gaffkystr. 11c,
Justus Liebig Universität Gießen, D-35385 Gießen, Germany

Introduction

Toxin production by members of the genus *Clostridium* is positively correlated with pathogenicity. This also holds true for *Clostridium novyi*, of which three serotypes A, B and C, have been differentiated. While serotype C is apathogenic and does not produce toxin, serotypes A and B are responsible for one-third of gas gangrene cases in humans and code for as many as eight different toxins (for review see Hatheway 1990); however, only α-toxin (Tcnα) is produced by both these serotypes. Its pronounced lethal effect has led to the classification of Tcnα as a lethal toxin, alongside toxin A (TcdA) and toxin B (TcdB) of *Clostridium difficile*, and TcsH (haemorrhagic toxin) and TcsL (lethal toxin) of *Clostridium sordellii* (Hatheway 1990). These toxins have been defined as large clostridial cytotoxins (LCTs) (Bette et al. 1991), with reference to the high molecular weight of the single-chain molecules, their in vitro action on the microfilament system of cultured cells and a series of in vivo effects. After intravenous injection, the LD_{50} of LCTs in mice ranges from 50 to 200 ng/kg. Injected into the rat paw they induce a delayed local oedema that lasts for days (Bette et al. 1989, 1991). Tcnα at least exerts this effect by causing dehiscence of the vascular endothelium, reflecting the retraction of endothelial cells induced by the toxin in culture. LCTs added to cells in culture are taken up by receptor-mediated endocytosis into the cytosol and provoke depolymerisation of microfilaments (Thelestam and Gross 1990). Microinjected toxins affect the cytoskeleton similarly; thus activation in the course of the uptake process is not required (Müller et al. 1992; Oksche et al. 1992). Similar microfilament disruption is caused by the small clostridial cytotoxins, C2 and C3, from *Clostridum botulinum* Type C, which ADP-ribosylate G-actin and Rho protein, respectively (Aktories et al. 1986; Chardin et al. 1989). LCT action, however, is not mediated by ADP-ribosylation (Florin and Thelestam 1991; Bette et al. 1991; Just et al. 1994; Selzer et al. 1994).

Tcnα has been purified to homogeneity (Izumi et al. 1983; Bette et al. 1989; Ball et al. 1993). SDS gel electrophoresis gives an apparent molecular weight for the single-chain toxin of between 200 and 250 kDa; the pI is 5.9. Tcnα is immunologically distinct from all the other toxins mentioned, but pharmacologically TcdB and Tcnα closely resemble each other (Bette et al. 1991; Ball et al. 1993). Tcnα production depends on infection of the bacterium with a phage (Eklund 1993). Curing *C. novyi* of the phage leads to loss of Tcnα production, and reinfection to its resumption.

In order to obtain further insights into the structure and possible mode of action of Tcnα, we set out to clone, sequence and analyse the Tcnα gene (*tcnα*). The ensuing comparison of the Tcnα amino acid sequence with those of *C. difficile* TcdA and TcdB helps to elucidate relationships between the three toxins and to establish structural criteria for their classification as members of the LCT family.

Material and methods

Bacterial strains and standard DNA technology

C. novyi type A strain ATCC 19402 was used as the source for isolation of protein and total DNA. It was grown in a medium optimized for toxin production (see below). *Escherichia coli* JM101 or comparable K-12 strains served as hosts for DNA constructs (Yanisch-Perron et al. 1985). Recombinants were grown in LB broth (or agar). Ampicillin (100 μg/ml) was added when selecting for antibiotic resistance.

Cloning of subgenic *C. novyi tcnα* fragments was performed under S2 containment. Standard DNA techniques were applied as outlined by Sambrook et al. (1989). Enzymes were purchased from Boehringer (Mannheim, FRG), Biolabs (Schwalbach, FRG) Pharmacia (Freiburg, FRG) and BRL (Eggersheim, FRG). Chemicals were obtained from Sigma (Deisenhofen, FRG) and Merck (Darmstadt, FRG). Radioactive nucleotides came from NEN DuPont (Bad Homburg, FRG).

Oligonucleotide synthesis

The oligonucleotide probes NOLI-1 (5′-GAACAATTAATGAAAATAGC-3′) and NOLI-2 (5′-GCTATTTTCATTAATTGTTC-3′) corresponding to the N-terminal sequence of Tcnα were synthesized on a fully automated MilliGen/Biosearch Cyclone Plus DNA synthesizer (Millipore, Eschborn, FRG). Other primers for sequencing (not listed here) or polymerase chain reaction (PCR) amplification (given below) were obtained commercially from MWG-Biotech (Ebersberg, FRG). Purity of oligonucleotides was examined by denaturing polyacrylamide gel electrophoresis (Sambrook et al. 1989).

Isolation of total *C. novyi* DNA

Strain ATCC 19402 was cultured anaerobically for 72 h at 37°C in 400 ml of a medium containing 2% proteose peptone, 1% yeast extract, 1% glucose, 0.5% $CaCO_3$, and 0.05% sodium thioglycolate. After centrifugation at 5000 rpm for 20 min in a Sorvall GS3 rotor the pellet was suspended in 3 ml 25 mM TRIS, 12.5% sucrose, pH 8.0. Lysozyme (2 g per 100 ml) in 10 mM TRIS-HCl, pH 8.0, 1 mM EDTA was added and the suspension was incubated for 1 h at 37°C. Lysis of the cells was completed by adding 100 μl of 20% SDS and incubating further at 37°C for 20 min. Proteinase K (1 g/100 ml) and 100 μl RNase (10 mg/ml) were added to the cleared lysate, which was further incubated for 30 min at 37°C. The lysates were then treated successively with 5 ml each of phenol and phenol-chloroform. Finally, the DNA was precipitated with 2.5 vol. ethanol (96%), pelleted, dried and dissolved in 3 ml 10 mM TRIS-HCl, 1 mM EDTA at a final DNA concentration of 2–3 mg/ml.

Construction of a *C. novyi* genomic library

The *C. novyi* genomic library was established according to the protocol used for construction of the *C. difficile* library (Eichel-Streiber et al. 1989), with the following modifications. Partial digestion was done with *Sau*3AI. DNA fragments were sized to 2–5 kb and integrated into the *Bam*HI site of pUC18. DNA constructs were transformed into *E. coli* JM101. The number of independent clones was 8000.

Isolation of recombinant α-toxin clones

For screening the library was plated at a density of 6000 colonies per plate on two 23×23 cm plates, grown overnight, transferred to membrane filters and then subjected to colony blotting and colony hybridization as described earlier (Eichel-Streiber et al. 1989, 1992a). Clones producing toxin fragments were identified with a rabbit anti-Tcnα antiserum (Bette et al. 1991), diluted 1:5000. Peroxidase-labelled anti-rabbit second antibody (used at a 1:1000 dilution) was from Dianova (Hamburg, FRG). To reduce background reactivity with proteins from the *E. coli* host, adsorption protein (10%, v/v; prepared according to Eichel-Streiber et al. 1989) was added to the first and second antibody. Positive clones were visualised by diaminobenzidine (0.5 mg/ml) staining (Eichel-Streiber et al. 1989). Colony hybridization was done with DNA probes specific for the 5′ end of *tcnα* (NOLI-1 or NOLI-2). Hybridisation experiments were performed with labelled NOLI-1 and NOLI-2, using the 5′ end labelling kit of Boehringer (Mannheim, FRG) and γ-[^{32}P]ATP (60 mCi/10 pmol, NEN DuPont, Bad Homburg, FRG). Of the 20 pmol of primer labelled, 1 pmol was used in each colony hybridization experiment. Prehybridization was done for 4 h at 42°C, followed by hybridization for 12 h at 42°C and washing at 35°C with 0.1×SSC (15 mM NaCl, 1.5 mM sodium citrate, pH 7.0).

PCR amplification

The two primers 5′-GGTGCGATTCAAGAGGCCACA-3′ (coding strand) and 5′-CGCTCCTAGCAGTCCCGAAAT-3′ (noncoding strand) were used to amplify nucleotides 3013-3273 (numbering follows Genebank accession number Z48636) from chromosomal DNA by PCR. Cloning the product produced pCn312. Amplification was done with 10 pmol primers, 100 ng chromosomal DNA for 30 cycles (denaturation, 95°C, 1 min; annealing, 48°C, 1 min; extension, 72°C, 1 min). The PCR-product was extracted from the agarose gel using Qiaex (Diagen, Düsseldorf, FRG) and subjected to direct sequencing with a Boehringer PCR sequencing kit. For cloning, the PCR fragment was filled in by the Klenow reaction and the DNA was then blunt-end ligated into pUC18 linearized with *Sma*I.

DNA sequencing and computer analysis

Dideoxy sequencing was done as modified by Tabor and Richardson (1987) with double-stranded plasmid DNA and α-[35S]ATP (NEN DuPont, Bad Homburg, FRG) as the radiolabelled nucleotide. Some sequences were determined on a Model 373 DNA sequencer (Applied Biosystems, Weiterstadt, FRG) by Dr. T. Hankeln (Institute for Genetics, Mainz University). The total *tcnα* sequence has been deposited in the EMBL-Database (EMBL, Heidelberg, FRG) under accession number Z48636. For computer

analysis the program packages PC/GENE Amos Bairoch (Genofit, Heidelberg, FRG), HUSAR and DOTPLOT (EMBL, Heidelberg, FRG) were used.

Protein chemistry

Tcnα was prepared according to Bette et al. (1989). Its N-terminus was subjected to microscale sequencing by Dr. D. Linder (Institute of Biochemistry, Gießen University). Protein derived from recombinant constructs was obtained by preparing total lysates (500 µl) from 5-ml overnight cultures of individual clones (Sambrook et al. 1989). Aliquots of 25 µl were subjected to SDS polyacrylamide gel electrophoresis and Western blotting (Eichel-Streiber et al. 1987) using Tcnα-specific antiserum (1:10000) and goat anti-rabbit peroxidase conjugate (1:5000) in the presence of *E. coli* JM101-pUC18 adsorption protein (10%, v/v).

Results

Construction of a *C. novyi* genomic library and specific probes

A genomic library was established in pUC18 by cloning sized fragments (2–5 kb) generated by partial *Sau*3AI digestion of total *C. novyi* DNA (see Materials and methods). This approach of constructing the genomic library offers the possibility to screen for chromosomally encoded virulence factors in addition to the phage encoded *tcnα*. After transformation we obtained 8000 independent clones. Due to the size of the DNA fragments used (2–5 kb) all recombinant clones carried only subgenic fragments of *tcnα*.

Tcnα was purified according to the procedure of Bette et al. (1989); the final fraction was homogeneous as judged by SDS polyacralamide gel electrophoresis; its N-terminal sequence, MLITrEQLMKIASIP, was determined by Edman degradation. Specific antisera raised in rabbits reacted in Western blots with a toxin band of approximately 200 kDa (see below, Fig. 4).

Isolation of overlapping *tcnα* clones

To identify recombinant clones that expressed parts of the Tcnα protein the library was screened by colony blotting with the polyclonal anti-Tcnα serum. Of the clones isolated, pCn47 (Fig. 1A) contained the largest DNA insert. It turned out to carry the 3′ end of the *tcnα* gene. Clone pCn121 was one of several constructs obtained after colony hybridization with a 5′ fragment of pCn47 as DNA probe. Again all recombinants mapped to the 3′ end of *tcnα* only. As attempts to complete cloning by chromosome walking failed, the library was screened for a third time with oligonucleotides NOLI-1 and -2 (see Material and methods) derived from the N-terminal sequence of the toxin. Clones pCn201 and pCn214 were representative of the series of clones obtained by this approach. Partial restriction mapping and sequencing with the complementary primers NOLI 1/2 eventually located the 5′ end of *tcnα* close to an *Acc*I site at position 4.5 kb on the map presented in Fig. 1A. Hybridization experiments between clones pCn214 and pCn121 showed no overlap. Both were sequenced and oligonucleotides (see Materials and methods) were synthesized that bound approximately 100 bp from the gap between them. These primers were used to amplify the missing 260 bp fragment from total *C. novyi* DNA by PCR. Cloning of this fragment resulted in pCn312. The position of this construct was verified by sequencing. The three clones, pCn214, pCn312 and pCn121, encode the entire *tcnα* nucleotide sequence (Fig. 1A).

Analysis of the DNA- and aa-sequence of the α-toxin

Sequencing of the *tcnα* gene revealed an open reading frame (ORF) of 6534 nucleotides (nt) coding for a protein (Tcnα) of M_r 250166. Figure 2 summarizes the sequence of Tcnα deduced from the DNA sequence including 5′- and 3′-untranslated sequences of *tcnα*. Potential promoter sequences (underlined) are analogous to those observed in *tcdA* and *tcdB* of *C. difficile* (Eichel-Streiber 1993). Downstream of *tcnα* an inverted repeat structure is located with a stabilizing free enthalpy (ΔG) of −114 kJ/mol. It probably represents the transcription terminator of *tcnα*.

The Tcnα sequence can be subdivided into three domains (Fig. 1B). Domain II and III were defined according to their special aa-composition. Stretches of hydrophobic amino acids are found at similar positions in Tcnα, TcdA and TcdB. Within this segment, potential membrane-spanning segments were identified by computer analysis (see Table 1 and Fig. 1B). We used these to define domain II, which stretches from the first major hydrophobic segment and the repeat region (for Tcnα see Fig. 2 in between the two arrows). Domain III, starting at amino acid 1797 and running to the C-terminus at position 2178 (Fig. 2, from second arrow to C-terminus) consists of 14 repeating units (each 19–21 or 48–52 amino acids in length). We arranged them in four separate groups of CROPs (*c*lostridial *r*epetitive *o*ligo*p*eptides). The order of *C. novyi* CROPs is given in Fig. 1C. Individual CROPs (NALI1-3, NK1-3, NE1-5, NA1-3) are presented in Fig. 3A, together with their respective consensus sequences. In 10 of the 14 CROP sequences the motif

Table 1 Prediction of potential membrane-spanning regions

Toxin	Amino acid	Sequence[a]
Tcnα	1036–1052	SVASVIIDGINLIAATcsL
	1097–1113	ILGVISVPVAGILVGLP
	1586–1602	FFTVNIQLCPFMIVSGV
TcdA	1031–1047	IPIVSTILDGINLGAAI
	1066–1082	VGVLAINMSLSIAATVA
	1093–1109	IFLLPIAGISAGIPSLV
TcdB	1029–1045	LPIIATIIDGVSLGAAI
	1091–1107	ILLVPLAGISAGIPSLV
	1400–1416	GFVSTcsLFSILEGINAII

[a] The list includes all sequences predicted for the holotoxins

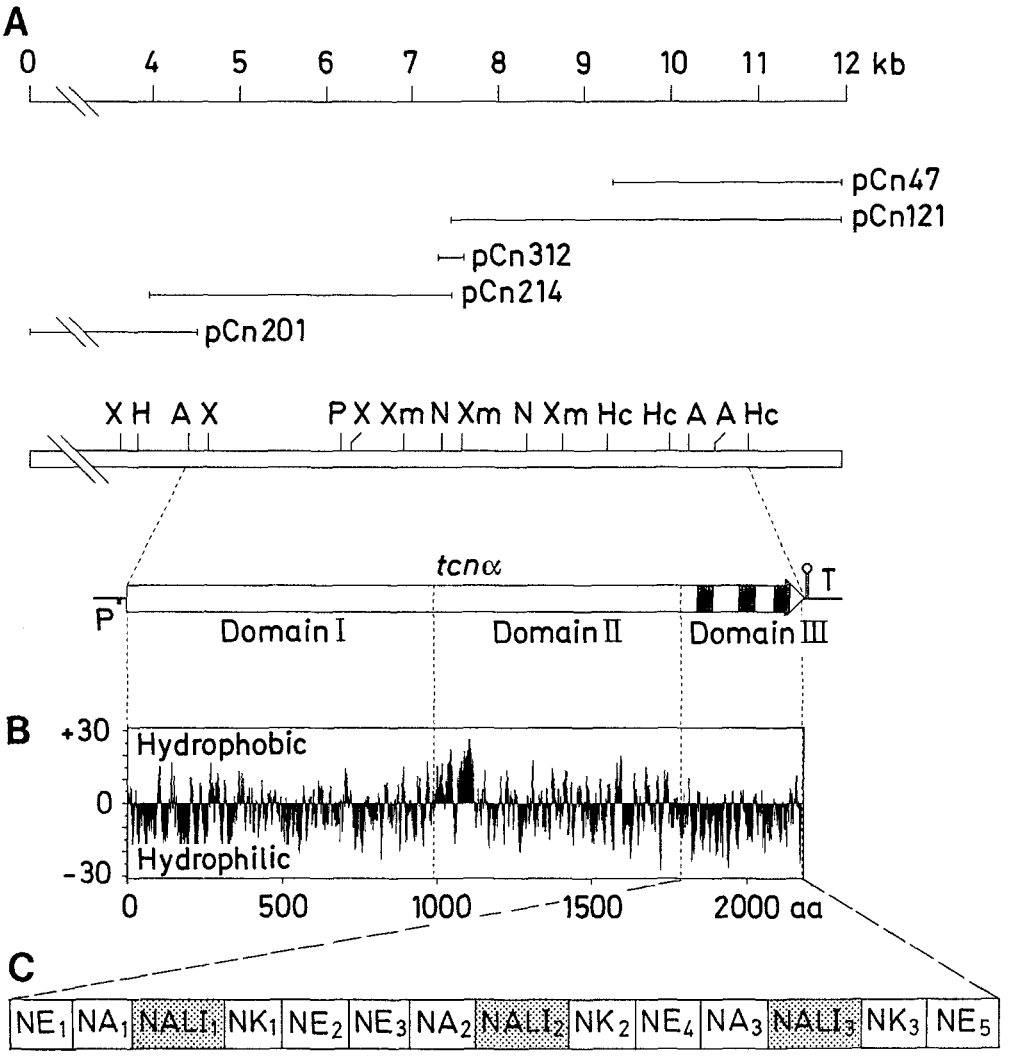

Fig. 1A Partial map of the *Clostridium novyi tcnα* gene. The five clones pCn201, pCn214, pCn312, pCn47 and pCn121 represent approximately 12 kb of *C. novyi* chromosomal DNA, containing the entire *tcnα* gene and sequences upstream and downstream of it. pCn47 was obtained by colony blotting, pCn121 by chromosome walking as outlined in the text, pCn201 and pCn214 by oligonucleotide screening and pCn312 by polymerase chain reaction (PCR) amplification. Cleavage sites for the following enzymes are shown: *A* AccI, *Hc* HincII, *H* HindIII, *N* NdeI, *P* PstI, *X* XbaI, *Xm* XmnI. A potential promoter region (*P*) and a stem-loop (*T*) structure, probably representing the transcription terminator of *tcnα*, are indicated. The Tcnα open reading frame is shown as an *open arrow*; *dotted boxes* close to its tip mark positions of NALI1,2,3 the largest clostridial repetitive oligopeptides (CROPs). **B** Hydropathy plot of Tcnα. The plot was run according to Kyte and Doolittle (1982) with a window of 20 amino acids (aa) and a step size of 5 aa. Tcnα is a hydrophilic molecule with a GRAVY (grand average of hydropathy) index of –3.09. Potential membrane-spanning segments are listed in Table 1. They are predicted to be transmembrane segments. The beginning of the stretch of hydrophobic amino acids defines the N-terminal end of domain II (and thus the C-terminal end of domain I). Domain III comprises the repetitive motifs. **C** Order of *C. novyi* Tcnα CROPs. The Tcnα C-terminal repeats, NALI, NK, NA, and NE, are shown at a larger scale. The three NALI CROPs are numbered sequentially and marked as *dotted boxes*. Their sequences are given in Fig. 3A

G-x_8-YYF is strictly conserved (NK3, NE1, NE2 and NE4 are exceptions; see Fig. 3A). Similar motifs have been found in *C. difficile* TcdA and TcdB (Fig. 3B) and some other proteins of gram-positive bacteria (Eichel-Streiber et al. 1992b).

Recombinant Tcnα fragments

The level of expression of recombinant Tcnα fragments in *E. coli* was generally low. Antibody screening with two independently raised polyclonal Tcnα antisera identified only those parts of the gene that matched to the 3′ end. However, the same antisera detected both N- and C-terminal peptides in Western blots (Fig. 4; Table 2). Thus, in Tcnα, unlike TcdA (Eichel-Streiber et al. 1989), the C-terminal repeat seems not to be the immunodominant region of the toxin. The multiplicity of polypeptide bands detected by the antiserum is a consequence of the AT-richness of clostridial DNA. Similar results have been obtained for the expression of *C. difficile* toxins (Eichel-Streiber et al. 1989). Internal initiation and pre-

A

```
aaacaagcca tttctcaaaa aaaagcaaga atactaacag acttgaaaaa taaattttgt aatgattaac aaaaaatatt
aaattattaa aattttatat ttttgtactt ggtatgtatc aataatataa gttaatttat ttaaaggagt gaattttATG
```

B

```
   1 MLITREQLMK IASIPLKRKE PEYNLILDAL ENFNRDIEGT SVKEIYSKLS KLNELVDNYQ TKYPSSGRNL ALENFRDSLY
  81 SELRELIKNS RTSTIASKNL SFIWIGGPIS DQSLEYYNMW KMFNKDYNIR LFYDKNSLLV NTLKTAIIQE SSKVIIEQNQ
 161 SNILDGTYGH NKFYSDRMKL IYRYKRELKM LYENMKQNNS VDDIIINFLS NYFKYDIGKL NNQKENNNNK MIAIGATDIN
 241 TENILTNKLK SYYYQELIQT NNLAAASDIL RIAILKKYGG VYCDLDFLPG VNLSLFNDIS KPNGMDSNYW EAAIFEAIAN
 321 EKKLMNNYPY KYMEQVPSEI KERILSFVRN HDINDLILPL GDIKISQLEI LLSRLKAATG KKTFSNAFII SNNDSLTLNN
 401 LISQLENRYE ILNSIIQEKF KICETYDSYI NSVSELVLET TPKNLSMDGS SFYQQIIGYL SSGFKPEVNS TVFFSGPNIY
 481 SSATCDTYHF IKNTFDMLSS QNQEIFEASN NLYFSKTHDE FKSSWLLRSN IAEKEFQKLI KTYIGRTLNY EDGLNFNKWK
 561 RVTTSELLKV IEEVNSTKIY ENYDLNMILQ IQGDDISYES AVNVFGKNPN KSILIQGVDD FANVFYFENG IVQSDNINNI
 641 LSRFNDIKKI KLTLIGHGEN VFNPKLFGGK TVNDLYTNII KPKLQHLLER EGVILKNKYL KINILGCYMF TPKVDINSTF
 721 VGKLFNKISR DLQPKGFSKN QLEISANKYA IRINREGKRE VLDYFGKWVS NTDLIAEQIS NKYVVYWNEV ENTLSARVEQ
 801 LNKVAEFAKD INSIIQTTNN QELKQSLVNT YADLITTLYS ELLKEDIPFE LDNIQIKERI ILNEISRLHD FSNIILDFYQ
 881 KNNISNNMII LFDSIIKEKD YYNVKLANKI TGETSVIKTY SDSLWNFTNK YKKIVDDIKG IIVKDINGEF IKKADFEIEQ
 961 NPSLLNSAML MQLLIDYKPY TEILTNMNTS LKVQAYAQIF QLSIGAIQEA TEIVTIISDA LNANFNILSK LKVGSSVASV
1041 IIDGINLIAA LTELKNVKTN FERKLIEAKV GMYSIGFILE SSSLISGLLG ATAVSEILGV ISVPVAGILV GLPSLVNNIL
1121 VLGEKYNQIL DYFSKFYPIV GKNPFSIQDN IIIPYDDIAI TELNFKYNKF KYGYAKISGF EGGSGHTYWG NIDHYFSAPS
1201 LDHYIELSIY PALKLNDTNL PKGNVVLLPS GLNKVYKPEI SAIAGANSQE GNGVEVLNLI RNYYVDSNGN TKFPWKYEAP
1281 FEYSFSYMRV EYFDTKVNVI LDNENKTLII PVLTIDEMRN KISYEILGDG GQYNVILPVN QTNINIVSNK NDIWNFDVSY
1361 IVKESKIEDN KFVLDGFINN IFSTLKVSND GFKIGKQFIS IKNTPRAINL SFKINNNIVI VSIYLNHEKS NSITIISSDL
1441 NDIKNNFDNL LDNINYIGLG SISDNTINCI VRNDEVYMEG KIFLNEKKLV FIQNELELHL YDSVNKDSQY LINNPINNVV
1521 KYKDGYIVEG TFLINSTENK YSLYIENNKI MLKGLYLESS VFKTIQDKIY SKEKVNDYIL SLIKKFFTVN IQLCPFMIVS
1601 GVDENNRYLE YMLSTNNKWI INGGYWENDF NNYKIVDFEK CNVIVSGSNK LNSEGDLADT IDVLDKDLEN LYIDSVIIIP
1681 KVYTKKIIIH PIPNNPQINI INTQSIHDKC HLIIDSVLTN NYHWESDGDD LIITNGLDIN IRILQGLSFG FKYKNIYLKF
1761 SNYDELSLND FLLQNYNVKG LYYINGELHY KNIPGDTFEY GWINIDSRWY FFDSINLIAK KGYQEIEGER YYFNPNTGVQ
1841 ESGVFLTPNG LEYFTNKHAS SKRWGRAINY TGWLTLDGNK YYFQSNSKAV TGLQKISDKY YYFNDNGQMQ IKWQIINNNK
1921 YYFDGNTGEA IIGWFNNNKE RYYFDSEGRL LTGYQVIGDK SYYFSDNING NWEEGSGVLK SGIFKTPSGF KLFSSEGDKS
2001 AINYKGWLDL NGNKYYFNSD SIAVTGSYNI KGIQYYFNPK TAVLTNGWYT LDNNNYYVSN GHNVLGYQDI DGKGYYFDPS
2081 TGIQKAGVFP TPNGLRYFTM KPIDGQRWGQ CIDYTGWLHL NGNKYYFGYY NSAVTGWRVL GGKRYFFNIK TGAATTGLLT
2161 LSGKRYYFNE KGEQLTLV
```

C

```
GTAtaatgta aataatagat ttaaaaaagc ctctagttta aagaggcttt tttaaatctt caaaagaatt ttaaaaatta
tcttttaata agatgtactt atagttactt tttaaataaa gccttccagg tatccttacc tacaatacca tcagcactta
```

Fig. 2A–C The Tcnα sequence. The sequence is based on a file stored at the EMBL database (EMBL Accession No. Z48636). *Letters in lowercase* indicate the nucleotide sequences of the 5′- and 3′-untranslated regions; *uppercase letters* represent in one-letter code the deducted amino acid sequence of the Tcnα ORF. **A** 5′-untranslated DNA sequence including the first codon (ATG). Putative promoter sequences TTAACA (–35 box) and TAAAAT (–10 box) are *underlined*. The *boxed* nucleotides represent the putative Shine-Dalgarno sequence (AGGAGT). **B** Amino acid sequence (1–2178) of Tcnα. Two *vertical arrows* define the C-terminal ends of domains I and II (compare Fig. 1B). **C** 3′-untranslated DNA nucleotide sequence including the C-terminal valine codon (GTA). Between nucleotides (nt) 6548 and 6590 there is an inverted repeat structure (ΔG –114 kJ/mol) that we propose as the transcription terminator of *tcnα* (underlined by convergent arrows)

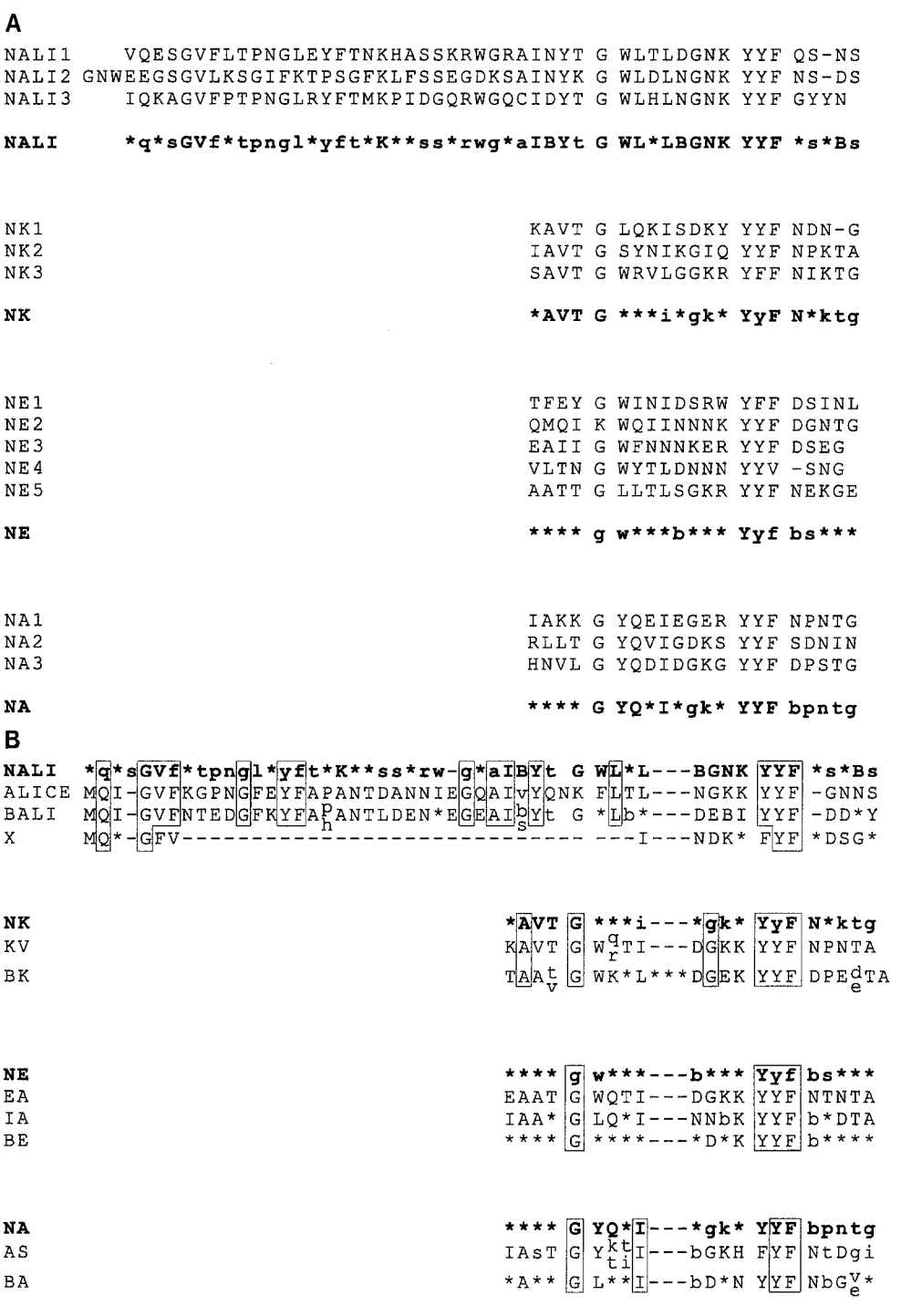

Fig. 3A,B Definition and comparison of the CROPs of Tcnα with those of TcdA and TcdB. **A** The 14 CROP sequences of *C. novyi* are arranged in four groups listed individually together with their consensus sequences (*bold letters*). Three long repeat sequences, NALI 1,2,3, of up to 52 amino acids, were defined. The NALI CROPs are flanked by NA 1,2,3 on the N-terminal and by NK 1,2,3 on the C-terminal side. The stretch NA-NALI-NK is similarly found in *Clostridium difficile* TcdA and TcdB (see Fig. 5B for comparison of consensus CROPs). In Tcnα they are spaced by sequences NE 1–5. All sequences were aligned to a five amino acid stretch C-terminal to the YYF sequence. For the consensus sequences *uppercase letters* indicate positions conserved in all, and *lowercase letters* conservation in the majority of CROPs. An *asterisk* stands for a highly variable amino acid position; shifts introduced in the sequence alignment are represented by *dashes*. The letter B designates Asn or Asp. The G-x_8-YYF motif is conserved in all but four CROPs. **B** Consensus sequences of the Tcnα, TcdA and TcdB CROPs are compared. CROPs of TcdA origin are: ALICE, KV, EA, IA, AS; TcdB CROPs are: BALI, X, BK, BE, BA (see also Fig. 5B). Those positions conserved in the individual CROP groups are *boxed*. CROP IA defined for TcdA is aligned with the NE, BE, EA intervening sequences

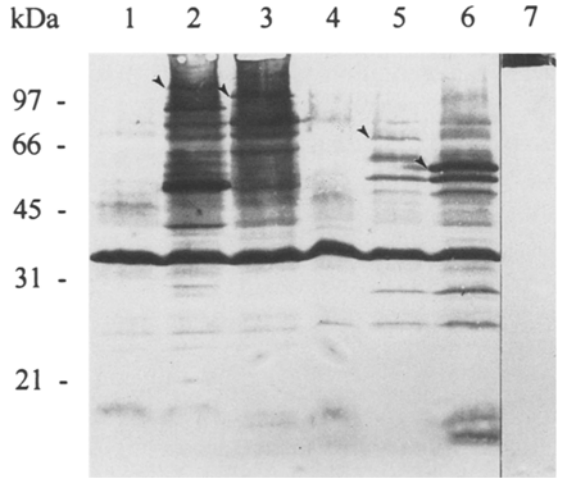

Fig. 4 Western blot of recombinant Tcnα fragments. *C. novyi* Tcnα (0.25 mg, lane 7) and total lysates of recombinant *tcnα* subclones [20 μl each; cells were not induced with isopropyl-1-thio-β-D-galactopyranoside (IPTG)] were separated on 10% SDS-polyacrylamide gels. The proteins were blotted onto nylon membranes, stained using the polyclonal Tcnα antiserum employed for primary screening of the genomic library, and further processed as described in Materials and methods. The major recombinant products are indicated by *arrowheads*. Lane 1, *Escherichia coli* JM101 with pUC18; 2, pCn214; 3, pCns18; 4, pCns11b; 5, pCns12; 6, pCn47; 7, Tcnα. Table 2 summarizes the clones, the fragments they code for and the calculated and determined sizes of Tcnα fragments. The serum detects N- and C-terminal toxin fragments, nevertheless some parts of Tcnα are not immunogenic (see lane 4, pCns11b). The polypeptide at 34 kDa represents an *E. coli* host cell protein stained by the Tcnα antiserum. This cross-reaction is evident only after separation of total lysates of *E. coli* proteins on SDS-polyacrylamide gels and is only marginal in colony blots. Similar results were obtained during analysis of TcdA and TcdB fragments of *C. difficile* (see Eichel-Streiber et al. 1989)

mature termination are the main reasons for this phenomenon. Why we were unable to isolate pCn214 (compare Fig. 4, lane 2) by colony blotting remains an open question. Lysates of the recombinant *E. coli* cells expressing Tcnα fragments did not show cytotoxic activity on CHO cells (data not shown).

Comparison with other protein sequences

Comparison of *C. novyi* Tcnα with either *C. difficile* TcdA or (TcdB) revealed extensive homology [33% (32%) identity and 15% (16%) conservative substitutions]. Separate comparisons of domains I, II, and III of all three toxins is difficult because of the size differences between the molecules. The proportion of aa positions that are identical in Tcnα, TcdA and TcdB is 22% for domain I, 18% for domain II and 27% in domain III. For domain III, the sequence we compared was restricted to the 11 strictly matching CROPs (only 287 positions of the repeating units; see Fig. 5B and Table 3). This domain is unusual, in combining high levels of variation in length with high levels of conservation in sequence.

Table 2 Calculated and determined sizes of recombinant proteins

Lane in Fig. 4	Clone[a]	*tcnα* sequence encoded[b]	Size of Tcnα fragment (kDa)	
			calculated	determined[c]
2	pCn214	−410–3104	114	114
3	pCns18	405–3104	99	99
4	pCns11b	3212–4388	43	Negative
5	pCns12	4379–(7600)[d]	79	70
6	pCn47	4966–(7600)[d]	58	62

[a] The *C. novyi* inserts are all inserted in-frame into the *lacZ* gene of pUC18
[b] Nucleotide (nt) positions as depicted in EMBL data files; Accession No Z48636
[c] Largest of several polypeptides
[d] Numbers in parentheses have been estimated from partial restriction mapping

Table 3 Biochemical properties of large clostridial cytotoxin (LCT) polypeptides

	Tcnα	TcdA	TcdB
Total length region	2178	2710	2366
Repeat region	382	833	512
Total − repeat region	1796	1877	1854
M_r (Da)	250166	308057	269708
pI	5.9	5.3	4.1
Groups of CROPs	4	5	5
Number of CROPs	14	30	19

Within the Tcnα repetitive region we defined four (Table 3), and within TcdA and TcdB five, groups of CROP sequences (Eichel-Streiber et al. 1992a). The arrangements of Tcnα, TcdA and TcdB CROPs are compared in Fig. 5B. The backbone of the repetitive domains is formed by the large repeats of 48–52 residues together with the adjacent CROPs (NA-NALI-NK for Tcnα). Variations in length of the C-terminal domains result from different numbers of CROPs in the various toxins. The total length of the remainder of the toxins is almost invariant (approx. 1800 amino acids, see Table 3). Thirty CROPs have been described for TcdA 19 for TcdB and 14 for Tcnα. The repeat in Tcnα, (382 residues) is the shortest among the three toxins. *C. difficile* TcdA and TcdB repeats comprise 833 and 512 amino acids, respectively. Sequence homology, again based on the repeating motifs, was also found to GTFs of Streptococci (Eichel-Streiber et al. 1992b). A 3-amino acid motif YYF is almost invariant in all repeats characterized. This motif seems to be essential for the interaction of these proteins with carbohydrates (Eichel-Streiber et al. 1992b).

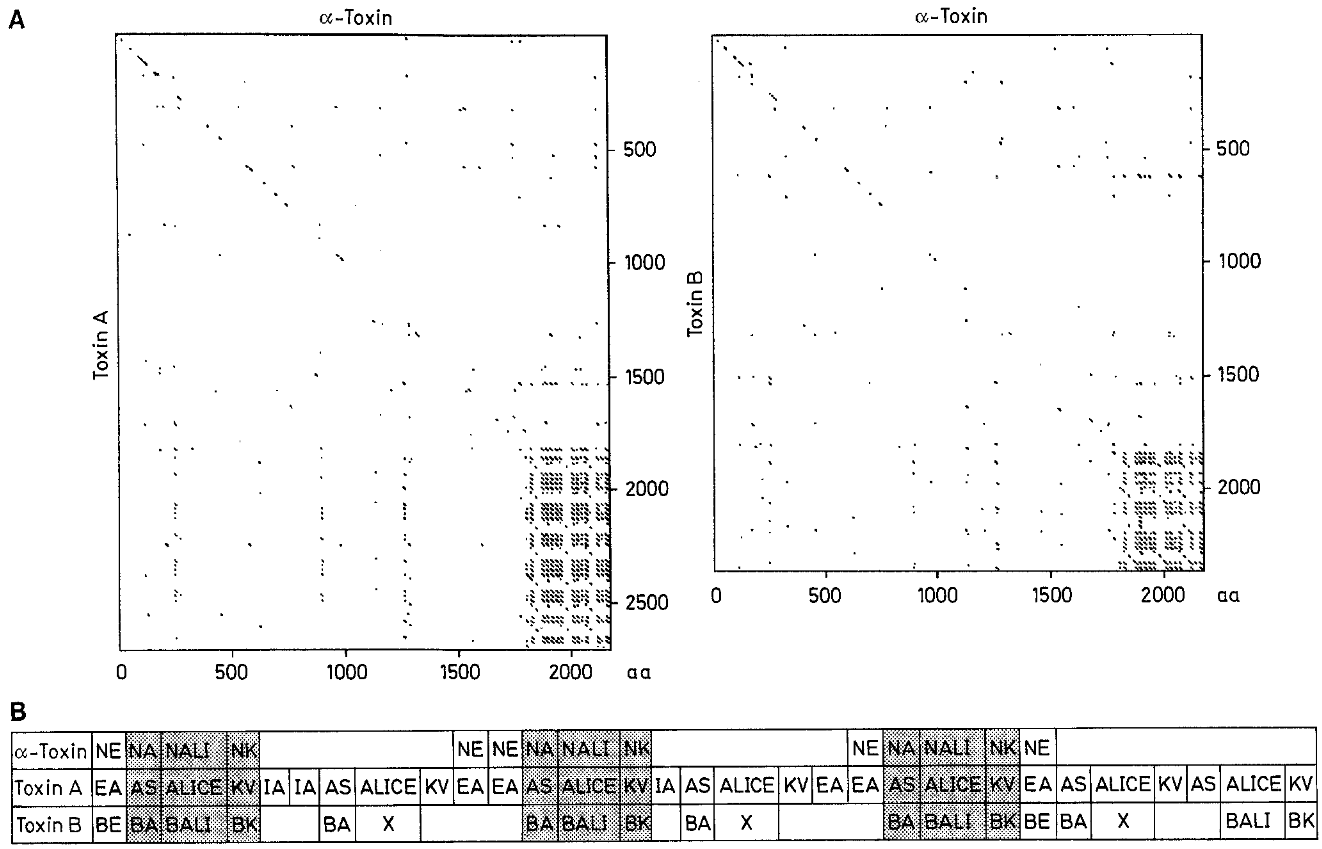

Fig. 5A,B Comparison of Tcnα with TcdA and TcdB of *C. difficile*. **A** Dot matrix plotting was performed with a window of 10 amino acids demanding identity of six positions. Tcnα (*horizontal*) is compared with TcdA and TcdB (*vertical*). There are some signals of identity on the diagonal, and many within the C-terminal repetitive domain. **B** The repeat domains of the three toxins Tcnα, TcdA and TcdB are aligned to highest homology. A general overview, not the individual sequences, is presented. As can be seen, the three CROPs NA-NALI-NK or their homologues AS-ALICE-KV (Eichel-Streiber and Sauerborn 1990) and BA-BALI-BK (Eichel-Streiber et al. 1992a) form the backbone peptides present in all toxins

Discussion

So far classification of Tcnα of *C. novyi*, TcdA and TcdB of *C. difficile* and TcsH and TcsL of *C. sordellii* into the single subgroup designated LCTs has been based on their common biochemical and pharmacological properties (Bette et al. 1989, 1991). This classification is now corroborated by determination of the *tcn*α DNA sequence.

The Tcnα molecule

The *tcnα* gene contains a single ORF encoding a molecule with an M_r of 250 166 and a pI of 5.9. Previous data obtained by SDS gel electrophoresis indicated a size of 200 kDa (Bette et al. 1989; Ball et al. 1993). The differences in size cannot be due to post-translational modification since the N-terminus is preserved in the mature toxin. Processing at its C-terminus would remove the repetitive ligand domain thus leaving a functional inactive toxin (see below). There is also no evidence that Tcnα is composed of distinct subunits. Thus the most probable reason for the discrepancy in M_r is the influence of protein folding on its mobility in SDS gels. Similar differences were observed with tetanus toxin (Kistner et al. 1993). A second discrepancy was observed between the molar fraction of proline of Tcnα determined chemically (Ball et al. 1993) and that deduced from the DNA sequence. The amino acid compositions of Tcnα, TcdA and TcdB are very similar and their proline contents are equally low (2.0–2.3 mol%; Ball et al. determined 14.9 mol%).

Comparison of Tcnα with the toxins of *C. difficile*

Immunologically Tcnα, TcdA and TcdB are unrelated. No antiserum against any of the three toxins crossreacts with the other two toxins. Despite this seeming heterogeneity, TcdA and TcdB of *C. difficile* display considerable homology with 48% amino acid sequence identity, with a further 15% conservative substitutions (Eichel-Streiber et al. 1992a). According to the sequence of Tcnα presented it shares 33%/32% identity and 48% homology with TcdA and TcdB, respectively. Similar degrees of homology have been reported within the family of clostridial neurotoxins (TeTx versus BotxA, 33.8%, and TeTx versus BotxE, 30.4% identity; Binz et al.

1990). Obviously the neurotoxin and the cytotoxin subgroup each evolved from a common ancestor.

The homologies between the three LCTs mentioned are most prominent in their CROPs (see Figs. 3 and 5). Similar repetitive motifs occur in other proteins, for instance glucosyltransferases of Streptococci, where they mediate protein-carbohydrate interactions (Banas et al. 1990; Wren 1991; Eichel-Streiber et al. 1992b). Thus we infer that initial binding of the LCTs to the cell surface is mediated by CROP-carbohydrate interaction. Besides this sequence homology, structural homology between Tcnα, TcdA and TcdB is obvious. The three toxins may be divided into three domains, an N-terminal toxic, an intermediary translocation and a C-terminal region contributing to cellular binding (Eichel-Streiber et al., in press).

Functional aspects of Tcnα

Polyclonal TcdA antiserum preferentially recognizes the repetitive motifs extending throughout the 833 C-terminal amino acids. In TcdA the immunodominant antigenic site of the CROPs (and the entire TcdA) is the motif TIDGKK, which is recognized by monoclonal antibody (mAb) TTC8 (Sauerborn et al. 1994).This sequence was recently identified by fluorescence microscopy as the structure that interacts with carbohydrates expressed on the target cell surface (Sauerborn and Eichel-Streiber, unpublished data). TcdA of *C. difficile* and TcsH of *C. sordellii* both haemagglutinate rabbit erythrocytes. The TIDGKK motif of TcdA is involved in this reaction. Additionally, TcsH reacts with mAb PCG-4, a second TcdA repeat-specific and TTC8 crossreactive mAb (Martinez and Wilkins 1988). Thus we expect to find a TIDGKK-like motif in TcsH; however, its sequence is still unknown.

The pharmacological diversity of LCTs may depend on binding or toxic action. The ligand motif TIDGKK occurs only in TcdA but not in TcdB or in Tcnα. Thus Tcnα uptake should be mediated by a carbohydrate structure different from the Galα1-3Galß1-4GlcNAc sequence, which is part of the TcdA receptor on the cell surface (Tucker and Wilkins 1991; Rolfe and Song 1993). In vitro all LCTs act on a wide variety of cells by disrupting their microfilaments. Cells resistant to TcdB have been obtained by chemical mutagenesis (Florin 1991); the phenotype was shown not to be due to a blockade of toxin uptake. Recently the intracellular target of TcdA, TcdB (Just et al. 1994) and Tcnα (Selzer et al. 1994) was identified as Rho, a small GTP-binding protein. Since Rho is involved in regulating the steady state of polymerisation of F-actin (Ridley and Hall 1992) its inactivation by LCTs may account for the damaged microfilament system. Conversely CNF (cytotoxic necrotising factor) of *E. coli* induces hyperpolymerisation of actin and again the key player is Rho (Oswald et al. 1994). Both effects result from as yet uncharacterized modifications of Rho. The modification of Rho by LCTs interferes with its substrate properties for C3 of *C. botulinum*, an ADP-ribosyltransferase that modifies Rho at Asn42. However, ADP-ribosylation activity has been ruled out for LCTs (Just et al. 1994), as has a protease activity comparable to that of clostridial neurotoxins (Schiavo et al. 1992).

In summary, the structural and functional links between Tcnα, TcdA and B have become much more evident. The LCTs of *C. sordellii* have largely been omitted from the discussion. Their structure is unknown. Moreover the cytopathic action of *C. sordellii* lethal toxin also involves the microfilaments but is morphologically distinct from that of the LCTs mentioned (Bette et al. 1991). Immunologically, antisera against TcsL and TcdB, like those against TcsH and TcdA, are mutually cross-reactive. Hence future work must define the structural and functional features of *C. sordellii* LCTs. The lessons learned from the clostridial neurotoxins may formally apply to the clostridial LCTs too. As with the former group, basic features should be common to all LCTs, whereas details of binding, internalization and substrate specificity obviously differ between the individual toxins.

Acknowledgements The work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB249/A1). We are indebted to M. Weidmann and M. Sauerborn for critically reading the manuscript and to Dr. D. Linder (Gießen) for microsequencing of Tcnα.

References

Aktories K, Bärmann M, Ohishi I, Tsuyama S, Jakobs KH, Habermann E (1986) Botulinum C2 toxin ADP-ribosylates actin. Nature 322:390–392

Ball DW, van Tassell RL, Roberts MD, Hahn PE, Lyerly DM, Wilkins TD (1993) Purification and characterization of alpha-toxin produced by *Clostridium novyi* type A. Infect Immun 61:2913–2918

Banas JA, Russell RR, Ferretti JJ (1990) Sequence analysis of the gene for the glucan-binding protein of *Streptococcus mutans* Ingbritt. Infect Immun 58:667–673

Bette P, Frevert J, Mauler F, Suttorp N, Habermann E (1989) Pharmacological and biochemical studies of cytotoxicity of *Clostridium novyi* type A alpha-toxin. Infect Immun 57:2507–2513

Bette P, Oksche A, Mauler F, von Eichel-Streiber C, Popoff MR, Habermann E (1991) A comparative biochemical, pharmacological and immunological study of *Clostridium novyi* alpha-toxin, *C. difficile* toxin B and *C. sordellii* lethal toxin. Toxicon 29:877–887

Binz T, Kurazono H, Wille M, Frevert J, Wernars K, Niemann H (1990) The complete sequence of botulinum neurotoxin A and comparison with other clostridial neurotoxins. J Biol Chem 265:9153–9158

Chardin P, Boquet P, Madaule P, Popoff MR, Rubin EJ, Gill DM (1989) The mammalian G protein rhoC is ADP-ribosylated by *Clostridium botulinum* exoenzyme C3 and affects actin microfilaments in Vero cells. EMBO J 8:1087–1092

von Eichel-Streiber C (1993) Molecular biology of the *Clostridium difficile* toxins In: Sebald M (ed) Genetics and molecular biology of the anaerobic bacteria. Springer-Verlag, New York, pp 264–289

von Eichel-Streiber C, Sauerborn M (1990) *Clostridium difficile* toxin A carries a C-terminal repetitive structure homologous to

the carbohydrate binding region of streptococcal glycosyltransferases. Gene 96: 107–113

von Eichel-Streiber C, Harperath U, Bosse D, Hadding U (1987) Purification of two high molecular weight toxins of *Clostridium difficile* which are antigenically related. Microbiol Pathogen 2: 307–318

von Eichel-Streiber C, Suckau D, Wachter M, Hadding U (1989) Cloning and characterization of overlapping DNA fragments of the toxin A gene of *Clostridium difficile*. J Gen Microbiol 135: 55–64

von Eichel-Streiber C, Laufenberg Feldmann R, Sartingen S, Schulze J, Sauerborn M (1992a) Comparative sequence analysis of the *Clostridium difficile* toxins A and B. Mol Gen Genet 233: 260–268

von Eichel-Streiber C, Sauerborn M, Kuramitsu HK (1992b) Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. J Bacteriol 174: 6707–6710

Eichel-Streiber C, Meyer zu Heringdorf D, Habermann E, Sartingen S (1994) Closing in on the toxic domain through analysis of a mutant *Clostridium difficile* cytotoxin. B. Mol Microbiol, in press

Eklund MW (1993) The role of bacteriophages and plasmids in the production of toxins and other biologically active substances by *Clostridium botulinum* and *Clostridium novyi*. In: Sebald M (ed) Genetics and molecular biology of the anaerobic bacteria. Springer-Verlag, New York, pp 179–194

Florin I (1991) Isolation of a fibroblast mutant resistant to *Clostridium difficile* toxins A and B. Microbial Pathogen 11: 337–346

Florin I, Thelestam M (1991) ADP-ribosylation in *Clostridium difficile* toxin-treated cells is not related to cytopathogenicity of toxin B. Biochim Biophys Acta 1091: 51–54

Hatheway CL (1990) Toxigenic clostridia. Clin Microbiol Rev 3: 66–98

Izumi N, Kondo H, Ohishi I, Sakaguchi G (1983) Purification and characterization of alpha-toxin of *Clostridium oedematiens* type A. Jpn J Med Sci Biol 36: 135–146

Just I, Fritz G, Aktories K, Giry M, Popoff MR, Boquet P, Hegenbarth S, von Eichel-Streiber C (1994) *Clostridium difficile* toxin B acts on small GTP-binding protein Rho. J Biol Chem 269: 10706–10712

Kistner A, Sanders D, Habermann E (1993) Disulfide formation in reduced tetanus toxin by thioredoxin: the pharmacological role of the interchain covalent and noncovalent bonds. Toxicon 31: 1423–1434

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132

Martinez RD, Wilkins TD (1988) Purification and characterization of *Clostridium sordellii* hemorrhagic toxin and cross-reactivity with *Clostridium difficile* toxin A (enterotoxin). Infect Immun 56: 1215–1221

Müller H, von Eichel-Streiber C, Habermann E (1992) Morphological changes of cultured endothelial cells after microinjection of toxins that act on the cytoskeleton. Infect Immun 60: 3007–3010

Oksche A, Nakov R, Habermann E (1992) Morphological and biochemical study of cytoskeletal changes in cultured cells afterextracellular application of *Clostridium novyi* alpha-toxin. Infect Immun 60: 3002–3006

Oswald E, Sugai M, Labigne A, Wu HC, Fiorentini C, Boquet P, O'Brien AD (1994) Cytotoxic necrotizing factor type 2 produced by virulent *Escherichia coli* modifies the small GTP-binding proteins Rho involved in assembly of actin stress fibers. Proc Natl Acad Sci USA 91: 3814–3818

Ridley AJ, Hall A (1992) The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. Cell 70: 389–399

Rolfe RD, Song W (1993) Purification of a functional receptor for *Clostridium difficile* toxin A from intestinal brush border membranes of infant hamsters. Clin Infect Dis 16 [Suppl 4]:S219-27

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sauerborn M, Hegenbarth S, Laufenberg-Feldmann R, Leukel P, von Eichel-Streiber C (1994) Monoclonal antibodies discriminating between *Clostridium difficile* toxins A and B. In: Freer J, Aitken R, Alouf JE, Boulnois GJ, Falmagne P, Fehrenbach FJ, Montecucco C, Piemont Y, Rappuoli R, Wadström T and Witholt B (eds) Bacterial protein toxins. Gustav Fischer, Stuttgart Jena New York, Zentralbl Bakteriol Suppl 24: 510–511

Schiavo G, Benfenati F, Poulain B, Rossetto O, Polverino de Laureto P, DasGupta BR, Montecucco C (1992) Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. Nature 359: 832–835

Selzer J, Just I, Habermann E, Aktories K (1994) A cytosolic factor is involved in the action of *Clostridium novyi* α-toxin on the GTP-binding protein Rho. Naunyn-Schmied Arch Pharmacol 349 Suppl:R 15

Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84: 4767–4771

Thelestam M, Gross R (1990) Toxins acting on the cytoskeleton. In: Shier WT, Mebs D (eds) Handbook of toxinology. Marcel Dekker, New York and Basel, pp 423–492

Tucker KD, Wilkins TD (1991) Toxin A of *Clostridium difficile* binds to the human carbohydrate antigens I, X, and Y. Infect Immun 59: 73–78

Wren BW (1991) A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. Mol Microbiol 5: 797–803

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–119