JOURNAL OF BACTERIOLOGY, Dec. 2007, p. 8643–8650 Vol. 189, No. 23
0021-9193/07/$08.00+0 doi:10.1128/JB.00784-07

Sequencing the Botulinum Neurotoxin Gene and Related Genes in *Clostridium botulinum* Type E Strains Reveals *orfx3* and a Novel Type E Neurotoxin Subtype▽

Ying Chen,[1,2] Hannu Korkeala,[1] Johannes Aarnikunnas,[3] and Miia Lindström[1]*

Department of Food and Environmental Hygiene, Faculty of Veterinary Medicine, University of Helsinki, Finland;[1] *College of Chemistry and Biology, Yantai University, Yantai, People's Republic of China;*[2] *and Department of Basic Veterinary Sciences, Faculty of Veterinary Medicine, University of Helsinki, Finland*[3]

Received 20 May 2007/Accepted 6 September 2007

Three *Clostridium botulinum* type E strains were sequenced for the botulinum neurotoxin (BoNT) gene cluster, and 11 type E strains, representing a wide biodiversity, were sequenced for the *bont*/E gene. The total length of the BoNT/E gene cluster was 12,908 bp, and a novel gene (partial) designated *orfx3*, together with the complete *orfx2* gene, was identified in the three type E strains for the first time. Apart from *orfx3*, the structure and organization of the neurotoxin gene cluster of the three strains were identical to those of previously published ones. Only minor differences (≤3%) in the nucleotide sequences of the gene cluster components were observed among the three strains and the published BoNT/E-producing clostridia. The *orfx3*, *orfx2*, *orfx1*, and *p47* gene sequences of the three type E strains shared homologies of 81%, 67 to 76%, 78 to 79%, and 79 to 85%, respectively, with published sequences for type A1 and A2 *C. botulinum*. Analysis of *bont*/E from the 14 type E strains and 19 previously published BoNT/E-producing clostridia revealed six neurotoxin subtypes, with a new distinct subtype consisting of three Finnish isolates alone. The amino acid sequence of the subtype E6 neurotoxin differed 3 to 6% from the other subtypes, suggesting that these subtype E6 neurotoxins may possess specific antigenic or functional properties.

Clostridium botulinum strains produce one or two of the lethal botulinum neurotoxins (BoNTs) designated A to G according to antigenic properties. These toxins are produced as a progenitor complex consisting of BoNT, hemagglutinin (HA), nontoxic non-HA (NTNH), and probably other uncharacterized components, such as P47 and OrfX (14, 15). The genes encoding these proteins are linked as a cluster and vary in composition and structure among different serotypes and strains. In general, the *ntnh* and *bont* genes are located conjointly and arranged in the downstream region of the gene cluster, whereas the upstream regions vary between different types and subtypes of *C. botulinum* (15).

C. botulinum type E predominates in aquatic environments. It differs from the terrestrial toxinotypes of *C. botulinum* in the arrangement of the neurotoxin gene cluster. Kubota et al. showed that in *C. botulinum* strain Iwanai, the *orfx2*, *orfx1*, and *p47* genes are arranged sequentially in the upstream region of the type E neurotoxin gene cluster (19). The partial *orfx2* (676 bp) was found in the distal upstream region in the same orientation as *orfx1*, whereas *p47* was located immediately upstream of *ntnh* and had the opposite orientation to *orfx2* and *orfx1*. The type E neurotoxin-producing *Clostridium butyricum* strain BL6340 was further shown to be identical to *C. botulinum* Iwanai in toxin gene arrangement (19). To date, no more sequences of type E neurotoxin gene clusters are available, and the sequences of the complete *orfx2* gene and the further upstream region of *orfx2* in type E strains are unknown. Type A2 and F strains have *orfx2*, *orfx1*, and *p47* genes similar to those of type E but with a regulator gene, *botR*, located between the *orfx1* and *p47* genes (15). Recently, a new gene, *orfx3*, in addition to the complete *orfx2* gene, was identified in types A1 and A2 (4).

The nucleotide sequence of *bont*/E has been published for 36 strains (11, 22, 28, 30). BoNTs are composed of a light chain (~50 kDa) and a heavy chain (~100 kDa), linked by a disulfide bond. BoNTs produced by different types of *C. botulinum* share homologies in sequence and structure (20, 23). The light chain is a metalloprotease, while the C- and N-terminal halves of the heavy chain are associated with neurospecific binding and translocation into the nerve cell cytosol, respectively (14). Any variation in amino acid sequence in these functional domains is sufficient to influence the structure, function, and antigenic properties of the BoNT complex (18, 20, 25). Based on amino acid sequences of the neurotoxin, three type E *C. botulinum* and 13 neurotoxigenic *C. butyricum* strains were classified into three subtypes (25), and 23 *C. botulinum* and 13 *C. butyricum* strains were later classified into five subtypes based on the nucleotide sequences of the *bont*/E gene (11).

In this study, three *C. botulinum* type E strains isolated in Finland, Germany, and the United States were chosen for sequencing the neurotoxin gene cluster. Furthermore, *bont*/E genes were sequenced for 11 strains of mainly Finnish origin. The neurotoxin gene cluster was composed of the *bont*/E, *ntnh*, *p47*, *orfx1*, *orfx2*, and *orfx3* genes. The complete *orfx2* and partial *orfx3* genes are reported for the first time in *C. botulinum* type E. Based on nucleotide sequences, *bont*/E genes in the currently sequenced and previously published *C. botulinum* strains and other BoNT/E-producing clostridia were classified

* Corresponding author. Mailing address: Department of Food and Environmental Hygiene, P.O. Box 66, FIN-00014, University of Helsinki, Finland. Phone: 358-9-19157107. Fax: 358-9-19157101. E-mail: miia.lindstrom@helsinki.fi.

▽ Published ahead of print on 28 September 2007.

TABLE 1. Primers used for sequencing *C. botulinum* type E strains

Gene	Primer	Sequence	Start site	Analysis used for sequence identification for strain[a]:			
				K35	K81	31-2570	Other[b]
orfx3	orfx2E-r6	TTGCAAAAGCAAATCCACCT	265	S	S	S	
	orfx2E-f7	TCCTAAAAGCAAGGGTGATGA	810	S	S	S	
	orfx2E-r7	CGAATTCATCACCCTTGCTT	835	S	S	S	
	orfx3E-r1	TTCCATGTGATTTCAGAAACTG		A, S	A, S	A, S	
orfx2	orfx2E-f1	TGGAAAGAAACAATTTTAAAAGATTCA	34	A, S			
	orfx2E-r1	TGAATGAGTACCGTCTGAATTAGG	330	A, S			
	orfx2E-f2	TTAATCCTTGCTTGGTTTAAAGA	415	A, S	A, S	A, S	
	orfx2E-r2	TCTGCACCCGTAGTCATTTG	674	A, S	A, S	A, S	
	orfx2E-f3	CAATGAGGACATTTATCCTGGAG	921	A, S	A, S	A, S	
	orfx2E-r3	CCATAGCCATAGCTGCAAAAG	1177	S	S	S	
	orfx3E-f1n	TGGCAACAAGTAGTGGGTGT	1498	S	S	S	
	orfx2E-r4	GGTAACGTCACCAGATTCCTC	1614	S	A, S	A, S	
	orfx2E-f5	TGGTTTGATATCCCAAAATGG	1896	S	S	S	
	orfx2E-r5	CCATTTTGGGATATCAAACCA	1916	S	S	S	
orfx1	orfx1E-f1	TGGATCAATTCATTTGGAAAAA	57	A, S			
	orfx1E-f1n	GGGATGGAATTGAAACAAGC	−3		S	S	
	orfx1E-r1	TCCCTTTTTAAAGCAACCAAA	221	A, S			
	orfx1E-f2	AGGGATTATTTAAGTTTATCTTTGTCA	217	A, S	S	S	
	orfx1E-r2	AGAATTCCATTTTTAGTTATCCTTTTT	359	A, S	S	S	
p47	p47E-f1	GAATACCTATGGTTGGGATATCGT	3	A, S	A, S	A, S	
	p47E-r1	TTCATCTTCTTCATATAATTTTCCAC	402	A, S	A, S	A, S	
	p47E-f2	GGTCATAGTTTCAGATCTTAGTGGAA	357	A, S	S		
	p47E-r2	TGCTGCTACTTTAATCCCATACC	825	A, S	S		
	p47E-f3	AAAAATACATTAAATTGGTATGGGATT	787	A, S	S	S	
	p47E-r3	TTTGCTAATTCAAAAGCTAAGGA	1139	A, S	A, S	A, S	
ntnh	ntnhE-f1	TGGTAATTTAAATATTGATTCTCCTG	12	A, S	A, S	A, S	
	ntnhE-r1	AAAAGTTTATAATCAATTCCTCCTGA	728	A, S	A, S	A, S	
	ntnhE-1L	GCAGAAAGTGGAATGGGAAC	472	S	S	S	
	ntnhE-1R	CATGGTTCCCATTCCACTTT	495	S	S	S	
	ntnhE-r2	AAATAATTTACTGGTAATGGTGTGG	1409	A, S	S	S	
	ntnhE-2L	TTCAAATCATGATGCCAGAAA	929	S	S	S	
	ntnhE-2R	TTTCTGGCATCATGATTTGAA	949	S	S	S	
	ntnhE-f3	GCTAAGGAAATTAATACTACCACACCA	1366	A, S	S	S	
	ntnhE-r3	TTCAAAAACACACATAGCAGCA	2169	A, S	A, S	A, S	
	ntnhE-3L	CCATGGATTGGTAGAGCATT	1663	S	S	S	
	ntnhE-3R	TTGTGTCCACCATTGATCTAAA	1920	S	S	S	
	ntnhE-f4	TGCTGCTATGTGTGTTTTTGA	2148	A, S	A, S	A, S	
	ntnhE-r4	TTTCTTTAATATCTTCATTTGCAACAA	2875	A, S	A, S	A, S	
	ntnhE-4L	TTGGAGATACATCCGGTAAAAA	2447	S	S	S	
	ntnhE-4R	AGCCCCAGTTAAATGTATTGC	2544	S	S	S	
	ntnhE-f5	TTGTTGCAAATGAAGATATTAAAGAAA	2849	A, S	A, S	A, S	
	ntnhE-r5	TTTTGGCATATACAGCATCTCC	3525	A, S	S	S	
	ntnhE-5L	TGTTCAAAAATGGGATGAGG	3210	S	S	S	A, S
	ntnhE-5R	CCTCATCCCATTTTTGAACA	3229	S	S	S	
bont/E	bontE-f1	CAGGCGGTTGTCAAGAATTT	68	A, S	A, S	A	S
	bontE-r1	TCCGCTAGCATCTTTATCTAATCC	945	A, S	A, S	A, S	S
	bontE-1L	ATAATGGGAGCAGAGCCTGA	448	S	S	S	S
	bontE-1R	ATCAGGCTCTGCTCCCATTA	468	S	S	S	S
	bontE-r2	TAATGCTGCTTGCACAGGTT	1713	A, S	A, S	A, S	S
	bontE-2L	TTTTTGTGGCTTCCGAGAAT	1304	S	S	S	S
	bontE-2R	ATTCTCGGAAGCCACAAAAA	1323	S	S	S	S
	bontE-f3	AACCTGTGCAAGCAGCATTA	1694	A, S	A, S	A, S	A, S
	bontE-r3	TCTGTATAAGAAGAAAGCTTAAAAGGA	2495	A, S	A, S	A, S	A, S
	bontE-3L	TGAACCCGAGCTTTTAATTCC	1908	S	S	S	S
	bontE-3R	GGAATTAAAAGCTCGGGTTCA	1928	S	S	S	S
	bontE-f4	TCTTGGGAGAGAGTCAGCAAG	2408	A, S	A, S	A, S	S
	bontE-4L	GACATTGCAAGATAATGCAGGA	2883	S	S	S	S
	bontE-4R	TGCATTATCTTGCAATGTCCA	2901	S	S	S	S
	bontE-f5	AAGAATTAGATGAAACAGAAATTCAAA	3155	A, S	S	A, S	S
	bontE-r5	CTTGCCATCCATGTTCTTCA	3751	A, S		A, S	
	bontE-5R	TGGAATTTATGACTTTAGCCGTTTA	3793	A, S	A, S	A, S	A

[a] A, amplification; S, sequencing.
[b] Other strains included all tested type E strains, except K35, K81, and 31-2570.

TABLE 2. Sequences for comparison of BoNT/E gene clusters among *C. botulinum* and *C. butyricum* strains

Accession no.	Species	Strain	Gene(s)	Reference[a]
AM695752	*C. botulinum* type E	K35	*orfx3* through *bont*/E	This study
AM695753	*C. botulinum* type E	K81	*orfx3* through *bont*/E	This study
AM695754	*C. botulinum* type E	31-2570	*orfx3* through *bont*/E	This study
AM695755	*C. botulinum* type E	K3	*bont*/E	This study
AM695756	*C. botulinum* type E	K8	*bont*/E	This study
AM695757	*C. botulinum* type E	K14	*bont*/E	This study
AM695758	*C. botulinum* type E	K15	*bont*/E	This study
AM695759	*C. botulinum* type E	K36	*bont*/E	This study
AM695760	*C. botulinum* type E	K44	*bont*/E	This study
AM695761	*C. botulinum* type E	K51	*bont*/E	This study
AM695762	*C. botulinum* type E	K101	*bont*/E	This study
AM695763	*C. botulinum* type E	K117	*bont*/E	This study
AM695764	*C. botulinum* type E	K119	*bont*/E	This study
AM695765	*C. botulinum* type E	S16	*bont*/E	This study
X62089	*C. botulinum* type E	Beluga	*bont*/E	22
X62683	*C. botulinum* type E	NCTC 11219	*bont*/E	30
AB082519	*C. botulinum* type E	35396	*bont*/E	NA
EF028403	*C. botulinum* type E	E185	*bont*/E	11
EF028404	*C. botulinum* type E	E544	*bont*/E	11
X62088	*C. butyricum*	ATCC 43755	*bont*/E	22
AB037704	*C. butyricum*	LCL155	*bont*/E	28
AB037705	*C. butyricum*	KZ1899	*bont*/E	28
AB037706	*C. butyricum*	KZ1897	*bont*/E	28
AB037707	*C. butyricum*	KZ1898	*bont*/E	28
AB037708	*C. butyricum*	KZ1886	*bont*/E	28
AB037709	*C. butyricum*	KZ1887	*bont*/E	28
AB037710	*C. butyricum*	KZ1889	*bont*/E	28
AB037711	*C. butyricum*	KZ1890	*bont*/E	28
AB037712	*C. butyricum*	KZ1891	*bont*/E	28
AB037713	*C. butyricum*	LCL063	*bont*/E	28
AB037714	*C. butyricum*	LCL095	*bont*/E	28
AB039264	*C. butyricum*	BL6340	*bont*/E	28
AB088207	*C. butyricum*	BL5262	*bont*/E	NA
D12697	*C. botulinum* type E	Mashike	*ntnh*	9
D88418	*C. botulinum* type E	Iwanai	*p47*, *orfx1*, *orfx2*	19
U70780	*C. botulinum* type E	Alaska	*p47* (*p48*)	21
D12739	*C. butyricum*	BL6340	*ntnh*	7
D88419	*C. butyricum*	BL6340	*p47*	19
X96493	*C. botulinum* type A	Kyoto-F	*p47*	5
AB004778	*C. botulinum* type A	Kyoto-F	*orfx1*	19
AY497357	*C. botulinum* type A	NCTC 2916	*p47*, *orfx1*, *orfx2*, *orfx3*	4
AY497358	*C. botulinum* type A	Kyoto-F	*orfx2*, *orfx3*	4
DQ310546	*C. botulinum* type A	Mascarpone	*p47*, *orfx1*, *orfx2*	6

[a] NA, no reference available.

into six subtypes. A previously unidentified subtype consisted of three Finnish *C. botulinum* type E strains.

MATERIALS AND METHODS

Bacterial strains, media, and growth conditions. A total of 14 *C. botulinum* type E strains were analyzed in this study. Strains K3, K8, K14, K15, K35, K36,

TABLE 3. Open reading frames in the neurotoxin gene cluster of *C. botulinum* type E strains K35, K81, and 31-2570

Gene	Positions	Sequence length		Avg GC content (%)		
		Nucleotide	Amino acid	K35	K81	31-2570
orfx3 (partial)	1 to 1355	1,355	451	28.56	28.49	28.49
orfx2	1387 to 3633	2,247	748	27.46	27.55	27.55
orfx1	3647 to 4081	435	144	19.77	19.77	19.77
p47	4368 to 5618	1,251	416	23.34	23.42	23.42
ntnh	5634 to 9125	3,492	1,163	20.90	21.08	21.11
bont/E	9150 to 12908	3,759	1,252	24.55	24.85	24.85

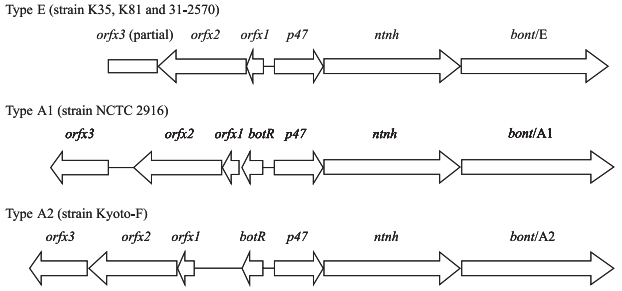

FIG. 1. Schematic representation of the neurotoxin gene cluster in *Clostridium botulinum* type E, A1, and A2 strains. The gene orientation is shown by arrows. Only *orfx3* in the type E gene cluster is a partial gene. Type A1 and A2 neurotoxin gene clusters are derived from *C. botulinum* A1 strain NCTC 2916 (GenBank accession numbers AY497357, Y14238, and X52066) and Kyoto-F (accession numbers AY497358, AB004778, X96493, X87974, and X73423).

TABLE 4. Identity of BoNT/E gene clusters and member genes among *C. botulinum* K35, K81, and 31-2570

Strains compared	% Identity						
	Gene cluster	*bont*/E	*ntnh*	*p47*	*orfx1*	*orfx2*	*orfx3*
K35 and K81	99.1	98.0	99.5	99.9	100	99.1	99.9
K35 and 31-2570	99.1	98.0	99.6	99.9	100	99.1	99.9
K81 and 31-2570	99.98	99.97	99.97	100	100	100	100

K44, K51, K117, and K119 were Finnish fish isolates, while strain S16 was isolated from sea sediment in Finland. Strains K81 and K101 were German fish isolates. Strain 31-2570 of an unknown source was isolated in the United States. Strains K35, K81, and 31-2570 were sequenced for the entire type E neurotoxin gene cluster. The other 11 strains were sequenced for the *bont*/E gene. All strains, apart from 31-2570, have been isolated in our laboratory and were selected for this study based on the wide geographical and genetic diversity revealed by pulsed-field gel electrophoresis (10), randomly amplified polymorphic DNA (13), and amplified fragment length polymorphism (16) analyses. Egg yolk agar plates were used to subculture these strains and confirm their purity. A single typical colony was picked and inoculated into 10 ml of anaerobic tryptone-peptone-glucose-yeast extract (Difco, Detroit, MI) broth. All cultures were incubated overnight at 30°C under anaerobic conditions.

PCR amplification. Genomic DNA was extracted from *C. botulinum* cultures as described by Keto-Timonen et al. (16). All primers for amplification and sequencing are listed in Table 1.

When the genes corresponding to previously published ones were being sequenced, PCR was applied to amplify suitable DNA fragments using the primers targeted to the published genes (9, 19, 30). To consecutively sequence the whole gene cluster, all adjacent fragments were amplified in an overlapping fashion.

For sequencing the region upstream of the known genes, inverse PCR was used to amplify the unknown fragment with primers orfx1E-r1 and orfx2E-f2 in the K35 strain (24). The genomic DNA (~27 μg) was cut with restriction enzyme HindIII (New England Biolabs, Inc., Ipswich, MA) at 37°C overnight and then ligated with T4 DNA ligase (New England Biolabs, Inc.) to form circular DNA at a DNA concentration of ≤3 ng/μl. The circular DNA was used as a template at a concentration of 2.5 ng/μl (26). After the upstream region of the *orfx2* gene in K35 was sequenced, the primer orfx2E-r4 was designed and used in combination with orfx2E-f2 to amplify the upstream region of the neurotoxin gene cluster in strains K81 and 31-2570.

In sequencing the previously unidentified upstream region of the gene cluster, primers targeted to the consensus sequence of *orfx3* from NCTC 2916 and Kyoto-F (4) were successfully used. This method was considered justified since high similarities were observed in the nucleotide sequences and arrangements of the *orfx2*, *orfx1*, and *p47* genes between the type E strains sequenced here and the previously published type A strains NCTC 2916 and Kyoto-F (4).

The PCRs were conducted using standard protocols (24). Amplification conditions consisted of 30 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and polymerization at 72°C for 80 s (for 4 min if the fragment was 2 kb or more), followed by a final extension at 72°C for 3 min (for 10 min if the fragment was 2 kb or more). Titanium *Taq* DNA polymerase (Clontech Laboratories, Inc., Mountain View, CA) was used in all PCRs. Amplification was performed with a PTC-200 thermal cycler (MJ Research, Watertown, MA).

DNA sequencing. Template DNA was purified from PCR products using a QIAquick PCR purification kit (Qiagen GmbH, Hilden, Germany) or a Montage PCR96 cleanup kit (Millipore Corp., Billerica, MA). Sequencing PCR was conducted using an ABI BigDye Terminator 3.1 cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA) according to the manufacturer's instructions. After precipitation with ethanol or purification with an Agencourt CleanSEQ kit (Agencourt Bioscience Corp., Beverly, MA), the sequencing samples were loaded onto an ABI 310 genetic analyzer or an ABI 3730 genetic analyzer, respectively. The former method was used to sequence the entire neurotoxin gene clusters, and the latter method was used to sequence the 11 neurotoxin genes for more efficient throughput.

Sequence analysis. The sequences of the entire neurotoxin gene clusters and *bont*/E were assembled and aligned by Sequencher software (Gene Codes, Ann Arbor, MI). Bidirectional sequence information from two independent PCRs was used, and the sequences were considered final when threefold coverage for each gene was obtained. Sequences were edited by BioEdit software (v. 7.0.5) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html), and open reading frames were predicted manually and via ORF Finder software (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Alignment and comparison of two or more sequences were performed using Align (http://www.ebi.ac.uk/emboss/align/index.html) or ClustalW software (3).

Type E neurotoxin genes from 33 *C. botulinum* type E strains, including the 14 genes sequenced in this study and 19 previously published ones, were compared using ClustalW software (3). The GenBank accession numbers of the previously published sequences are shown in Table 2.

Nucleotide sequence accession numbers. The sequence data reported in this paper have been submitted to the EMBL Nucleotide Sequence Database under accession numbers AM695752 through AM695765 and are listed in Table 2.

RESULTS

Structure and organization of BoNT/E gene cluster in three type E strains. The neurotoxin gene clusters of *C. botulinum* K35, K81, and 31-2570 consisted of six genes and were each 12,908 bp long (Table 3). For the first time, the entire nucleotide sequences of *orfx2* (2,247 bp) and a partial *orfx3* gene (1,355 bp), located immediately upstream of *orfx2* and having the same orientation as *orfx2*, were determined for the type E BoNT gene cluster (Fig. 1). The structures of the neurotoxin gene clusters in the three *C. botulinum* strains were identical to each other and, apart from the novel elements, to published ones.

K81 and 31-2570 had identical sequences for *p47*, *orfx1*, *orfx2*, and *orfx3*. These two strains were also highly similar in their *bont*/E and *ntnh* sequences. K35 differed from the other two strains in *bont*/E and *orfx2*, with 98% and 99.1% identities, respectively; however, K35 was highly identical to K81 and 31-2570 in the *p47*, *orfx1*, and *orfx3* gene sequences (Table 4). Marked differences were observed in the GC content of the six genes in the *bont*/E cluster, with *orfx1* showing the lowest and

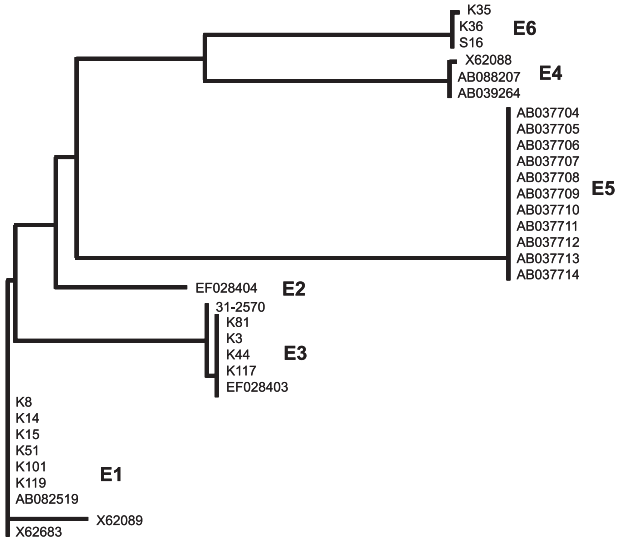

FIG. 2. The *bont*/E genes from 14 *Clostridium botulinum* type E strains sequenced here and the 19 previously published ones were compared using ClustalW software. Six distinct neurotoxin subtypes were described and designated E1 to E6. *C. botulinum* strains fell into subtypes E1, E2, E3, and E6, while subtypes E4 and E5 consisted of *C. butyricum* strains alone. Three Finnish isolates formed a previously unidentified neurotoxin subtype, E6.

TABLE 5. Homology of nucleotide and amino acid sequences for BoNT/E genes from representative strains of each subtype

Strain	% Homology for indicated strain of subtype[a]:																	
	E6				E3				E1				E2		E4		E5	
	K35		K36		K81		31-2570		K8		Beluga		E544		BL5262		LCL155	
	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa
K35			99	99	97	95	98	95	98	96	98	96	98	96	98	96	97	94
K36					98	95	98	96	98	97	98	96	98	96	98	96	97	94
K81							99	99	99	98	99	97	98	97	97	95	97	95
31-2570									99	98	99	97	98	97	97	95	97	95
K8											99	99	99	99	98	97	98	96
Beluga													99	98	98	96	98	96
E544															98	97	97	96
BL5262																	97	95

[a] nt, nucleotide; aa, amino acid.

the partial *orfx3* gene the highest proportions of GC content (Table 3).

Diversity of type E neurotoxin genes. The *bont*/E genes from a total of 33 strains of BoNT/E-producing clostridia were analyzed (Table 2). These strains included the 14 *C. botulinum* type E strains of mainly Finnish origin described in this study and 19 *C. botulinum* and *C. butyricum* strains with *bont*/E sequences available in GenBank. The 33 strains could be classified into six subtypes based on the nucleotide sequences of *bont*/E (Fig. 2). Five Finnish and four other *C. botulinum* strains of diverse origin formed subtype E1. All of these strains were identical in their *bont*/E gene sequences, with the exception of *C. botulinum* Beluga (GenBank accession number X62089), which showed a 1% difference. Subtype E2 included one *C. botulinum* strain E544 (EF028404) alone, while subtype E3 included six *C. botulinum* strains from Finland, Germany, and the United States. Except for strain 31-2570, the strains in subtype E3 had 100% identical *bont*/E genes. All *C. butyricum* strains formed subtypes E4 and E5, with E4 including ATCC 43755 (X62088), BL5262 (AB088207), and BL6340 (AB039264). Subtype E5 was a uniform group containing 11 *C. butyricum* strains. A previously unidentified subtype designated E6 included three *C. botulinum* strains of Finnish origin. Of these, K36 and S16 had 100% sequence identity for *bont*/E.

The homology analysis of the nucleotide and amino acid sequences of BoNT/E within and between the six subtypes is shown in Table 5. Strains within a subtype had only 1% differences in their nucleotide and amino acid sequences, while 1 to 3% and 1 to 5% differences in nucleotide and amino acid sequences, respectively, were observed between the four subtypes of *C. botulinum*. Between the *C. botulinum* and *C. butyricum* subtypes, 2 to 3% and 3 to 6% differences in the nucleotide and amino acid sequences, respectively, were observed. Moreover, the two *C. butyricum* subtypes had corresponding differences of 3% and 5% in their nucleotide and amino acid sequences.

Homology of *ntnh* and *p47*. Among *C. botulinum* type E strains K35, K81, 31-2570, and Mashike (GenBank accession number D12697), the *ntnh* genes were in general 99% homologous in nucleotide and corresponding amino acid sequences; however, strains K35 and Mashike shared 98% amino acid sequence homology for *ntnh*. Furthermore, the differences between the four *C. botulinum* strains and *C. butyricum* BL6340

TABLE 6. Homology of nucleotide and amino acid sequences for the *p47* gene and the *orfx1* gene among types E and A neurotoxin-producing clostridia

Strain	% Homology[a]																	
	C. botulinum type E										*C. butyricum* BL6340		*C. botulinum* type A					
	K35		K81		31-2570		Iwanai		Alaska (*p48*)				NCTC 2916		Kyoto-F		Mascarpone	
	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa
K35			99	99	99	99	99	99	99	99	98	98	85	81	79	72	80	72
K81	100	100			100	100	100	100	100	100	99	98	85	81	79	72	80	72
31-2570	100	100	100	100			100	100	100	100	99	98	85	81	79	72	80	72
Iwanai	100	100	100	100	100	100			100	100	99	98	85	81	79	72	80	72
Alaska											99	98	85	81	79	72	80	72
BL6340													85	80	79	71	80	71
NCTC 2916	78	72	78	72	78	72	78	72							85	79	87	81
Kyoto-F	79	72	79	72	79	72	79	72					93	89			96	95
Mascarpone	79	72	79	72	79	72	79	72					99	99	93	88		

[a] nt, nucleotide; aa, amino acid. Homologies for the *p47* gene sequences are given in the top right portion of the table, whereas those for the *orfx1* gene are given in the lower left portion.

TABLE 7. Homology of nucleotide and amino acid sequences for the *orfx2* gene and the *orfx3* gene among types E and A neurotoxin-producing clostridia

Strain	% Homology[a]													
	C. botulinum type E								*C. botulinum* type A					
	K35		K81		31-2570		Iwanai (partial)		NCTC 2916		Kyoto-F		Mascarpone (partial)	
	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa
K35			99	98	99	98	99	98	67	49	68	49	76	71
K81	99	99			100	100	99	98	68	50	68	50	76	61
31-2570	99	99	100	100			99	98	68	50	68	50	76	61
Iwanai									75	60	76	60	76	60
NCTC 2916	81	76	81	76	81	76					97	96	98	97
Kyoto-F	81	75	81	75	81	75			94	94			100	100

[a] nt, nucleotide; aa, amino acid. Homologies for the *orfx2* gene sequences are given in the top right portion of the table, whereas those for the *orfx3* gene are given in the lower left portion.

were 1 to 2% in nucleotide sequences and 2% in amino acid sequences.

All the BoNT/E-producing clostridia shared a high homology in the nucleotide and amino acid sequences for *p47*. *C. botulinum* K35 shared 99% homology in the nucleotide and amino acid sequences for *p47* with K81, 31-2570, Iwanai (accession number D88418), and Alaska (U70780, previously named *p48*), the four strains being identical in *p47* sequence. With *C. butyricum* BL6340 (D88419), all *C. botulinum* type E strains had 98 to 99% and 98% homologies in nucleotide and amino acid sequences, respectively (Table 6).

The *p47* genes in the BoNT/E-producing clostridia were markedly different from those in *C. botulinum* type A strains NCTC 2916 (AY497357), Kyoto-F (X96493), and Mascarpone (DQ310546). The differences in nucleotide and amino acid sequences were 15 to 21% and 19 to 29%, respectively (Table 6).

Homology of *orfx1*, *orfx2*, and *orfx3*. No difference was found in *orfx1* among *C. botulinum* type E strains K35, K81, 31-2570, and Iwanai (D88418). However, differences of 21 to 22% and 28% in nucleotide and amino acid sequences, respectively, for *orfx1* were observed among type E, type A1, and type A2 strains (Table 6).

For *orfx2*, there were only differences of 1% and 2% in the nucleotide and amino acid sequences, respectively, among the *C. botulinum* type E strains (Table 7), while again considerable differences were observed among strain types E, A1, and A2. Moreover, for *orfx3*, high homologies (99 to 100%) of nucleotide and amino acid sequences were present among *C. botulinum* type E strains, while homologies of only 81% and 75 to 76% were observed among strain types E, A1, and A2 (Table 7).

Analysis of amino acid sequence of BoNT/E. A total of 149 variable sites were found in the alignment of the BoNT/E amino acid sequences of seven strains representing the different subtypes (Table 8). The variable sites were distributed mainly in the light chain (37.6%) and the C-terminal half of the heavy chain (43.0%), while less variation was observed in the N-terminal half of the heavy chain (19.4%). No variation was seen in the conserved motifs of the functional domains. Compared with the consensus sequence, the sequences of the *C. butyricum* strains forming subtype E5 had maximum variability, with all their variable residues being within the heavy chain. By contrast, the *C. botulinum* K8 sequence had no variation at all. Sequences of strains within the new subtype E6 had the second highest variability distributed along the functional domains of the light chain and the N- and C-terminal halves of the heavy chain. The number and distribution of variable amino acid residues for subtype E6 differed markedly from those of the other subtypes.

DISCUSSION

To our knowledge, we are the first to report partial *orfx3* and complete *orfx2* genes in the BoNT type E gene cluster of three *C. botulinum* strains. Apart from *orfx3*, the other components of the neurotoxin gene cluster were nearly identical to the published sequences of *C. botulinum* Iwanai and *C. butyricum* BL6340 in composition, size, and organization (19).

In addition to BoNT/E-producing clostridia, *orfx1*, *orfx2*, *orfx3*, and *p47* have been found in some type A1 and A2 strains (4, 6). These genes show tremendous similarities among type E, A1, and A2 toxin gene clusters in nucleotide sequence, size, and gap distance (Fig. 1). An interesting feature of our type E strains and the A2 strain Mascarpone (6) is the lack of potential insertion elements detected in the intergenic regions of the neurotoxin gene cluster of type A1 strain NCTC 2916 and type A2 strain Kyoto-F (4). The striking similarities, together with evidence of transposable elements, suggest that gene transfer and recombination have occurred among type A and E neu-

TABLE 8. Numbers of variable amino acid residues in BoNT/E among the representative strains of different subtypes

Location	No. of variable amino acid residues in indicated strain of serotype:						
	E1		E2	E3	E4	E5	E6
	K8	Beluga	E544	31-2570	BL5262	LCL155	K36
Light chain	0	3	1	21	17	0	14
N-terminal half of heavy chain	0	1	0	0	7	9	12
C-terminal half of heavy chain	0	4	10	0	9	30	11

rotoxin-producing clostridia, supporting the theory of gene transfer playing a role in the evolutionary history of the BoNT gene cluster (2). It is tempting to speculate that the lack of insertion elements in our type E strains and strain Mascarpone (6) indicates that these *orfx1*-, *orfx2*-, and *orfx3*-carrying strains represent a conserved ancestor in the evolutionary history of the neurotoxin gene cluster. However, the variable GC content among the neurotoxin genes, particularly *orfx1*, *orfx2*, and *orfx3*, undermines this theory and may suggest the emergence of these genes from a more heterogeneous origin.

We did not detect definite promoters for *orfx1*, *orfx2*, and *orfx3* in the three type E strains. Dineen et al. found that in type A2 strain Kyoto-F, *orfx1*, *orfx2*, and *orfx3* were transcribed as a polycistronic transcript from a conserved promoter 1,179 bases upstream of the *orfx1* start codon and speculated that the corresponding gene products had a coordinated role in the production and expression of the neurotoxin complex (4). Whether the *orfx1*, *orfx2*, and *orfx3* genes play a role in the production of the biologically active BoNT/E complex or whether these genes have lost their meaning through evolution, both possibilities warrant further investigation.

Six subtypes of the type E neurotoxin gene were found among 33 BoNT/E-producing clostridia, based on nucleotide sequences. Five of the subtypes (E1 through E5) are in agreement with a previous subdivision by Hill et al. based on nucleotide sequences (11). While most Finnish *C. botulinum* strains fell into two subtypes, E1 and E3, the three Finnish strains K35, K36, and S16 formed a new subtype, designated E6, based on their nucleotide sequences. The E6 strains clearly differed from the other subtypes (by 3 to 6%) in their amino acid sequences of BoNT/E. Although the overall variation in BoNT/E sequences seems to be lower than that of, for instance, BoNT/A, this level of amino acid variability has been shown to affect the antigenic and biological properties of the neurotoxin (17, 27). Since *C. botulinum* type E strains are frequently found in the Baltic Sea region and in Finnish freshwaters and have been shown to possess a strikingly high biodiversity (10, 13, 16), even more BoNT/E subtypes likely exist among the Finnish *C. botulinum* population. Hence, further investigations of *C. botulinum* strains are warranted.

In the alignment of the BoNT/E amino acid sequences determined in this study or published previously, we found that most of the 149 variable sites were distributed mainly along the catalytic and binding domains. Strains representing different subtypes showed distinct distributions for the variable amino acid residues. For example, *C. botulinum* strain 31-2570 in subtype E3 showed variation only within its catalytic domain. In BoNT/E, this domain residing in the light chain specifically recognizes and cleaves SNAP-25 (synaptosome-associated protein of 25 kDa), and thus blocks neurotransmission in the synaptic vesicle (12, 29). However, the highly conserved zinc-binding motif of HELIHSLH, responsible for protease activity in the catalytic domain (1, 8, 20), was unchanged in the compared BoNT/E sequences.

As opposed to subtype E3, the subtype E5 strain *C. butyricum* LCL155 had all of its variable amino acids within the heavy chain, most of them in the C-terminal binding domain. The high variability in this domain results in a variety of surface structures that can act as neurospecific epitopes, conferring a unique antigenicity and binding to receptors in nerve cells (14, 18, 20). Despite the large number of variable amino acid residues in the binding domain, the conserved motif YL THMRD, responsible for receptor binding in BoNT/E (18, 20), was unaltered in the BoNT/E sequences analyzed.

A relatively low level of variation among the BoNT/E amino acid sequences was observed in the translocation domain at the N terminus of the heavy chain. With the exception of one amino acid shift from Glu to Val (position 637) in *C. butyricum* LCL155 (accession number AB037704), no variation was observed in the hydrophobic domain (positions 624 to 645). The translocation domain is assumed to sense environmental variations and is essential for the channel formation in the endocytosis vesicle membrane for further translocation into the nerve cell cytosol (20). Any changes in the hydrophobic domain will probably affect the channel structure and impair the potential spanning function (18, 30).

As discussed above, the strains of the new subtype E6 were 3 to 6% different from the other subtypes in their BoNT/E amino acid sequences. Although the conserved motifs were not altered, variation was evenly distributed in all three functional domains. Tsukamoto et al. reported the neurotoxin of *C. butyricum* LCL155 to possess different binding activity, toxicity, and antigenicity than those of *C. botulinum* 35396 and *C. butyricum* BL5262 (27), with similar levels of amino acid alterations in the heavy chain alone. Considering that subtype E6 had similar levels of alterations along the entire neurotoxin, it may possess specific antigenic structures and biological properties.

In conclusion, we sequenced the BoNT/E gene cluster in three *C. botulinum* type E strains and the *bont*/E genes in 11 strains of mainly Finnish origin. The partial *orfx3* and complete *orfx2* genes were identified in BoNT/E-producing clostridia for the first time. The structure and organization of the neurotoxin gene cluster in type E strains were identical among strains K35, K81, and 31-2570. Based on the alignment of *bont*/E gene sequences, a total of six neurotoxin subtypes were found among BoNT/E-producing clostridia. A new subtype, E6, consisting of Finnish strains, was observed.

ACKNOWLEDGMENTS

The work was supported by the Academy of Finland (206319).
We thank Hanna Korpunen for technical assistance.

REFERENCES

1. **Binz, T., H. Kurazono, M. Wille, J. Frevert, K. Wernars, and H. Niemann.** 1990. The complete sequence of botulinum neurotoxin type A and comparison with other clostridial neurotoxins. J. Biol. Chem. **265:**9153–9158.
2. **Brüggemann, H.** 2005. Genomics of clostridial pathogens: implication of extrachromosomal elements in pathogenicity. Curr. Opin. Microbiol. **8:**601–605.
3. **Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson.** 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. **31:**3497–3500.
4. **Dineen, S. S., M. Bradshaw, C. E. Karasek, and E. A. Johnson.** 2004. Nucleotide sequence and transcriptional analysis of the type A2 neurotoxin gene cluster in *Clostridium botulinum*. FEMS Microbiol. Lett. **235:**9–16.
5. **East, A. K., M. Bhandari, J. M. Stacey, K. D. Campbell, and M. D. Collins.** 1996. Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic *Clostridium botulinum* types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component. Int. J. Syst. Bacteriol. **46:**1105–1112.
6. **Franciosa, G., A. Maugliani, F. Floridi, and P. Aureli.** 2006. A novel type A2 neurotoxin gene cluster in *Clostridium botulinum* strain Mascarpone. FEMS Microbiol. Lett. **261:**88–94.
7. **Fujii, N., K. Kimura, N. Yokosawa, K. Oguma, T. Yashiki, K. Takeshi, T.**

Ohyama, E. Isogai, and H. Isogai. 1993. Similarity in nucleotide sequence of the gene encoding nontoxic component of botulinum toxin produced by toxigenic *Clostridium butyricum* strain BL6340 and *Clostridium botulinum* type E strain Mashike. Microbiol. Immunol. **37:**395–398.
8. **Fujii, N., K. Kimura, N. Yokosawa, K. Tsuzuki, and K. Oguma.** 1992. A zinc-protease specific domain in botulinum and tetanus neurotoxins. Toxicon **30:**1486–1488.
9. **Fujii, N., K. Kimura, N. Yokosawa, T. Yashiki, K. Tsuzuki, and K. Oguma.** 1993. The complete nucleotide sequence of the gene encoding the nontoxic component of *Clostridium botulinum* type E progenitor toxin. J. Gen. Microbiol. **139:**79–86.
10. **Hielm, S., J. Björkroth, E. Hyytiä, and H. Korkeala.** 1998. Prevalence of *Clostridium botulinum* in Finnish trout farms: pulsed-field gel electrophoresis typing reveals extensive genetic diversity among type E isolates. Appl. Environ. Microbiol. **64:**4161–4167.
11. **Hill, K. K., T. J. Smith, C. H. Helma, L. O. Ticknor, B. T. Foley, R. T. Svensson, J. L. Brown, E. A. Johnson, L. A. Smith, R. T. Okinaka, P. J. Jackson, and J. D. Marks.** 2007. Genetic diversity among botulinum neurotoxin-producing clostridial strains. J. Bacteriol. **189:**818–832.
12. **Humeau, Y., F. Doussau, N. J. Grant, and B. Poulain.** 2000. How botulinum and tetanus neurotoxins block neurotransmitter release. Biochimie **82:**427–446.
13. **Hyytiä, E., S. Hielm, J. Björkroth, and H. Korkeala.** 1999. Biodiversity of *Clostridium botulinum* type E strains isolated from fish and fishery products. Appl. Environ. Microbiol. **65:**2057–2064.
14. **Johnson, E. A.** 2005. Clostridial neurotoxins, p. 491–525. *In* P. Dürre (ed.), Handbook of clostridia. CRC Press, Boca Raton, FL.
15. **Johnson, E. A., and M. Bradshaw.** 2001. *Clostridium botulinum* and its neurotoxins: a metabolic and cellular perspective. Toxicon **39:**1703–1722.
16. **Keto-Timonen, R., M. Nevas, and H. Korkeala.** 2005. Efficient DNA fingerprinting of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis. Appl. Environ. Microbiol. **71:**1148–1154.
17. **Kozaki, S., Y. Kamata, T. Nagai, J. Ogasawara, and G. Sakaguchi.** 1986. The use of monoclonal antibodies to analyze the structure of *Clostridium botulinum* type E derivative toxin. Infect. Immun. **52:**786–791.
18. **Kubota, T., T. Watanabe, N. Yokosawa, K. Tsuzuki, T. Indoh, K. Moriishi, K. Sanda, Y. Maki, K. Inoue, and N. Fujii.** 1997. Epitope regions in the heavy chain of *Clostridium botulinum* type E neurotoxin recognized by monoclonal antibodies. Appl. Environ. Microbiol. **63:**1214–1218.
19. **Kubota, T., N. Yonekura, Y. Hariya, E. Isogai, H. Isogai, K. Amano, and N. Fujii.** 1998. Gene arrangement in the upstream region of *Clostridium botulinum* type E and *Clostridium butyricum* BL6340 progenitor toxin genes is different from that of other types. FEMS Microbiol. Lett. **158:**215–221.
20. **Lacy, D. B., and R. C. Stevens.** 1999. Sequence homology and structural analysis of the clostridial neurotoxins. J. Mol. Biol. **291:**1091–1104.
21. **Li, B., X. Qain, H. K. Sarkar, and B. R. Singh.** 1998. Molecular characterization of type E *Clostridium botulinum* and comparison to other types of *Clostridium botulinum*. Biochim. Biophys. Acta **1395:**21–27.
22. **Poulet, S., D. Hauser, M. Quanz, H. Niemann, and M. R. Popoff.** 1992. Sequences of the botulinal neurotoxin E derived from *Clostridium botulinum* type E (strain Beluga) and *Clostridium butyricum* (strains ATCC 43181 and ATCC 43755). Biochem. Biophys. Res. Commun. **183:**107–113.
23. **Prabakaran, S., W. Tepp, and B. R. DasGupta.** 2001. Botulinum neurotoxin types B and E: purification, limited proteolysis by endoproteinase Glu-C and pepsin, and comparison of their identified cleaved sites relative to the three-dimensional structure of type A neurotoxin. Toxicon **39:**1515–1531.
24. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed., p. 8.1–8.126. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. **Smith, T. J., J. Lou, I. N. Geren, C. M. Forsyth, R. Tsai, S. L. LaPorte, W. H. Tepp, M. Bradshaw, E. A. Johnson, L. A. Smith, and J. D. Marks.** 2005. Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. Infect. Immun. **73:**5450–5457.
26. **Triglia, T., M. G. Peterson, and D. J. Kemp.** 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. **16:**8186.
27. **Tsukamoto, K., M. Mukamoto, T. Kohda, H. Ihara, X. Wang, T. Maegawa, S. Nakamura, T. Karasawa, and S. Kozaki.** 2002. Characterization of *Clostridium butyricum* neurotoxin associated with food-borne botulism. Microb. Pathog. **33:**177–184.
28. **Wang, X., T. Maegawa, T. Karasawa, S. Kozaki, K. Tsukamoto, Y. Gyobu, K. Yamakawa, K. Oguma, Y. Sakaguchi, and S. Nakamura.** 2000. Genetic analysis of type E botulinum toxin-producing *Clostridium butyricum* strains. Appl. Environ. Microbiol. **66:**4992–4997.
29. **Washbourne, P., R. Pellizzari, G. Baldini, M. C. Wilson, and C. Montecucco.** 1997. Botulinum neurotoxin types A and E require the SNARE motif in SNAP-25 for proteolysis. FEBS Lett. **418:**1–5.
30. **Whelan, S. M., M. J. Elmore, N. J. Bodsworth, T. Atkinson, and N. P. Minton.** 1992. The complete amino acid sequence of the *Clostridium botulinum* type-E neurotoxin, derived by nucleotide-sequence analysis of the encoding gene. Eur. J. Biochem. **204:**657–667.