Glycoconjugate Journal (1994) **11**:141–151

Gene structure of the 'large' sialidase isoenzyme from *Clostridium perfringens* A99 and its relationship with other clostridial *nanH* proteins

CHRISTINA TRAVING, ROLAND SCHAUER AND PETER ROGGENTIN*

Biochemisches Institut der Christian-Albrechts-Universität, Olshausenstr. 40, D-24098 Kiel

Received 28 December 1993, revised 22 February 1994

Clostridium perfringens possesses two sialidase isoenzymes of different molecular weight. Almost 90% of the gene encoding the 'large' form was found on a 3.1 kb chromosomal fragment (Sau3AI) of strain A99 by hybridization with probes developed from the N-terminal protein sequence and from commonly conserved sialidase motifs ('Asp-boxes'), whereas the remaining 3'-terminal part was detected on a 2.1 kb fragment (*Hind* III) of chromosomal DNA. After combination of both fragments, the resulting *E. coli* clones expressed sialidase activity, the properties of the recombinant sialidase corresponding with those of the wild type enzyme. The entire chromosomal fragment of 3665 bp encompasses the complete sialidase gene of 2082 bp corresponding to 694 amino acids, from which a molecular weight of 72956 for the mature protein can be deduced. The first 41 amino acids are mostly hydrophobic and probably represent a signal peptide. The sialidase structural gene follows a non-coding region with an inverted repeat and a ribosome-binding site. Upstream from the regulatory region, another open reading frame (ORF) was detected. The 3'-terminus of the sialidase structural gene is directly followed by a further ORF of unknown function, which possibly encodes a putative permease or the acylneuraminate pyruvate-lyase involved in sialic acid catabolism. The primary structure of the 'large' isoenzyme is very similar to the sialidase of *Clostridium septicum* (55% identical amino acids), whereas the homology with the 'small' form of the same species is comparatively low (26%).

Keywords: clostridial sialidases (*exo-α-sialidases*, neuraminidases, *nanH*, recommended: *siaH*, EC 3.2.1.18); gene structure; isoenzymes; relationship; *Clostridium perfringens*

Introduction

Sialidases (exo-α-sialidases, neuraminidases, *nanH*, recommended: *siaH*, EC 3.2.1.18) are key enzymes in sialic acid catabolism. They hydrolyse sialic acids terminally bound to the sugar chains of glycoproteins, glycolipids or oligosaccharide molecules. They are involved in the turnover of sialoglycoconjugates in higher animals of the deuterostomate lineage, which possess the corresponding carbohydrate structures. However, sialidases are also widely, but irregularly distributed among microorganisms, which generally lack sialic acid-containing substrates. As they predominantly occur in species that live in close contact with higher animals, they may enhance the pathogenicity of these microorganisms [1–5].

The anaerobic Gram-positive rod *Clostridium perfringens*, a frequent causative agent of gas gangrene [6], is known to possess two sialidase isoenzymes with strikingly different properties [7]. The 'small' isoenzyme that was first isolated by Roggentin *et al.* [8], seems to lack any advantage for the bacterium because, due to the absence of a signal peptide, it cannot be secreted and come in contact with its extracellular substrate. The 'large' form, first isolated by Nees *et al.* [9] is, apart from the enzymes of *Vibrio cholerae* and *Arthrobacter ureafaciens*, one of the most extensively studied and longest known bacterial sialidases. It is purified from *Clostridium perfringens* on an industrial scale for many scientific applications. This enzyme is secreted and has a mainly nutritional function, which is confirmed by the fact that this species also possesses an intracellular acylneuraminate pyruvate-lyase ('aldolase', EC 4.1.3.3) for the degradation of sialic acids [10]. The activity of both enzymes is induced upon addition of free or glycopeptide-bound sialic acid to the culture medium [11], suggesting that both genes are part of an operon. Additionally, the existence of a sialic acid permease was postulated [11], which permits the transport of free sialic acids derived from the host into the bacterial cell.

* To whom correspondence should be addressed.

However, nothing is as yet known about the primary structure and the genes encoding the 'large' sialidase, the acylneuraminate pyruvate-lyase or a putative sialic acid permease. Since the primary structures of several microbial sialidases have already been examined [12], including the gene encoding the 'small' isoenzyme of *Clostridium perfringens* A99 [13], it was one aim of this study to determine the primary structure of this important 'large' sialidase, in order to obtain further insight into the relationship and origin of sialidases. Furthermore, it was of interest to examine whether the gene for this 'large' sialidase is actually organized in an operon together with an 'aldolase' and/or a putative permease gene.

In this report we present the complete nucleotide and predicted amino acid sequence of the 'large' sialidase from *Clostridium perfringens* strain A99 and of ORF2 which is possibly part of a common operon. We also demonstrate the relationship of the 'large' isoenzyme to the 'small' form of the same species and to other clostridial sialidases.

Materials and methods

Bacterial strains, media and vectors

Clostridium perfringens strain A99, isolated from human gas gangrene infection, was a gift of Dr R. Hobrecht (Untersuchungsinstitut I der Bundeswehr, Kronshagen, Germany). *Escherichia coli* JM101 [14] was purchased from Pharmacia (Freiburg, Germany). Vectors for cloning (pUC18, pUC19) and sequencing (M13mp18 and mp19) [15] were from GibcoBRL (Eggenstein, Germany), Boehringer (Mannheim, Germany) or New England Biolabs (Schwalbach, Germany).

For the isolation of total DNA, *Clostridium perfringens* A99 was grown after 1:8 inoculation of 2 l Todd-Hewitt broth (Difco, Detroit, USA) for 8 h at 37 °C under anaerobic conditions.

E. coli was propagated at 37 °C in Luria Bertani broth or on 1.8% agar media. Alternatively, 2 × YT broth (16 g l^{-1} Bacto Tryptone (Difco), 10 g l^{-1} yeast extract (Serva, Heidelberg, Germany), 5 g l^{-1} NaCl (Merck, Darmstadt, Germany)) was used. Media were supplemented with 100 μg ml^{-1} ampicillin for vector selection.

DNA methods and manipulation

Chromosomal DNA of *Clostridium perfringens* strain A99 was isolated according to Marmur [16], modified by Gebers *et al.* [17]. Cloning, preparation of plasmid DNA, ligation, agarose gel electrophoresis and electroelution of DNA-fragments from the gel were performed as described by Sambrook *et al.* [18]. Cells were electroporated in a Gene Pulser™ (Bio-RAD, München, Germany) according to the users' manual. Enzymes for molecular biology were purchased from GibcoBRL, Boehringer or New England Biolabs.

Synthesis and labelling of oligonucleotide probes

The N-terminal amino acid sequence of the purified sialidase protein was determined using a 470A gas-phase sequencer (Applied Biosystems, Pfungstadt, Germany) as described by Eckerskorn *et al.* [19]. From the resulting sequence: V-N-N-S-E-N-L-S-S-L-G-E-Y-K-D-I-N-L-E, two oligonucleotides were created as probes:

Probe A: 5′GTT/A-AAT-AAT-A/TC/GT-GAA-AAT-TT-3′

Probe B: 5′GAA-TAT-AAA-GAT-ATA/T-AAT3′

A further probe (C) is represented by a mixture of two oligonucleotides, each 24 bp in length, which are directed against conserved and repeated sequences ('Asp-boxes') in bacterial sialidases [20]. The probe used for the detection of the fragment encoding the 3′-region of the 'large' sialidase gene (D) was directed against bases 2815–2839 of the known sequence. Another probe (E) was constructed against bases 3641–3660 to screen the sequence downstream from the sialidase gene for another ORF. The probes were synthesized by using an automated DNA-synthesizer (380B, Applied Biosystems) or purchased from R.D. Horstmann (Tropeninstitut, Hamburg, Germany). The oligonucleotides were labelled according to Sambrook *et al.* [18] with [γ-^{32}P]ATP (Amersham, Braunschweig, Germany) and T_4 polynucleotide kinase from New England Biolabs provided with the appropriate buffer.

Blotting and hybridization

DNA was transferred from agarose gels to nylon sheets [21] using a Hybaid™ Vacuum Blotter (Biometra, Göttingen, Germany). *E. coli*-colonies were transferred to nylon filters as described [22]. DNA was denatured and immobilized [23] and hybridization was performed according to Wallace *et al.* [24]. Nylon sheets with immobilized DNA were prehybridized for at least 1 h at the appropriate temperature in 6 × SSC buffer containing 5 × Denhardt's solution [18], 0.5% SDS and 0.1 mg ml^{-1} yeast RNA. The same buffer and temperature were used for overnight hybridization. The filters were washed three times in 6 × SSC buffer at the same temperature for 20 min and then exposed to X-ray films.

DNA sequencing

Sequence analysis was carried out with [α-^{35}S]dATP (Amersham) and the dideoxy chain-termination procedure of Sanger *et al.* [25]. Single-stranded M13mp18 and mp19 templates were prepared [18] and sequenced with the Sequenase DNA sequencing kit (United States Biochemical Corp., Bad Homburg). Electrophoresis was performed at 56 °C in wedge-shaped polyacrylamide gels (6 or 8%) containing 7 M urea by using a LKB Makrophor Sequencing System as described by the supplier (LKB, Bromma, Sweden). Overlapping sequences were obtained by subse-

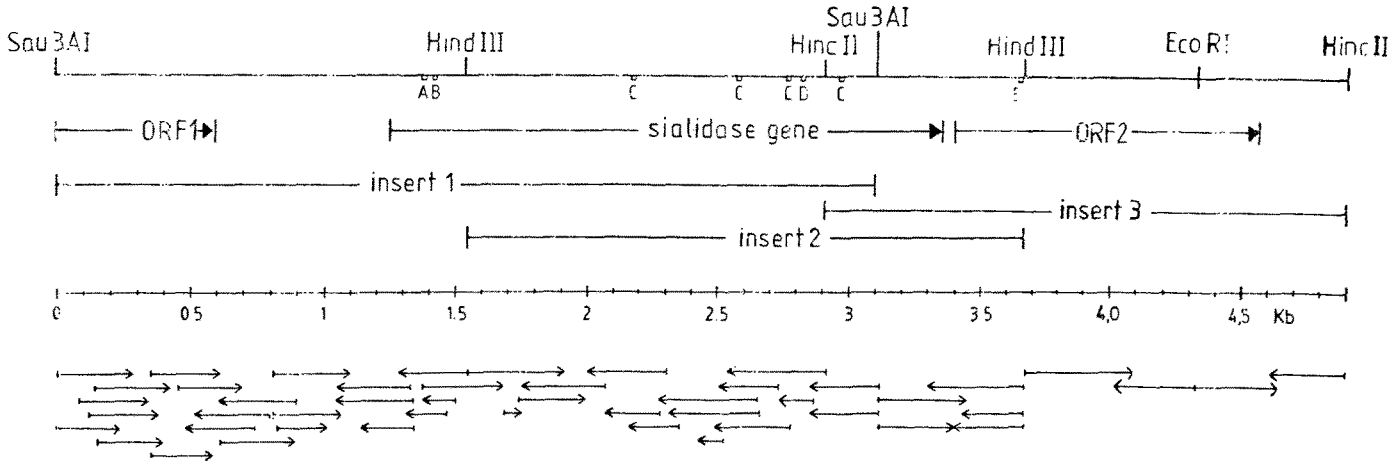

Figure 1. Restriction map of inserts 1, 2 and 3, including the complete gene for the 'large' sialidase isoenzyme of *Clostridium perfringens* A99 and two further open reading frames (ORF1 and 2). Arrows indicate the direction and length of subfragments sequenced. A, B, C, D and E demonstrate the target regions of corresponding oligonucleotide probes (see text).

quent DNA-digestion with exonuclease BAL 31 before subcloning in M13.

Sequence data were analysed with the GENMON program version 4.3 (GBF, Braunschweig, Germany). Sequences were compared with the data of the EMBL 29.0-SWISSPROT 20.0-database of this program.

The amino acid sequence of the 'large' sialidase of *Clostridium perfringens* was aligned by eye with three further clostridial sialidases (Fig. 3). The percentage of identical amino acids between each pair of sialidases was calculated (Table 2), and the similarity values were used to construct a dendrogram by the average linkage method [26].

Sialidase assays

E. coli clones were sprayed with the synthetic sialidase substrate 4-methylumbelliferyl-α-D-*N*-acetylneuraminic acid (MU-Neu5Ac). Clones expressing sialidase activity were identified by their blue-white fluorescence under UV light (360 nm). Alternatively, *E. coli* clones expressing sialidase activity were grown overnight at 37 °C in 2 ml of 2 × YT broth, harvested by centrifugation, and the sediment was resuspended in 1 ml buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris, pH 8.0) containing 10 mg ml^{-1} lysozyme. After incubation for 30 min at 37 °C, cells were sedimented by centrifugation. The sialidase activity in the supernatant was determined by incubating 10 µl enzyme solution with 80 µl 0.1 M sodium acetate buffer, pH 5.1 or 6.1 and 10 µl of 1 mM MU-Neu5Ac for 10 min at 50 or 37 °C, respectively. Free MU was measured in a M1000 fluorimeter (Perkin-Elmer, Überlingen, Germany). The instrument was calibrated with MU-solutions.

Antibodies inhibiting sialidase activity

Polyclonal antibodies against the natural 'large' sialidase from *Clostridium perfringens* A99 were prepared as described previously [27]. Antibodies against the cloned 'small' isoenzyme were raised in rabbits as described [6, 28]. Sialidase inhibition tests were performed by incubating 20 µl enzyme solution with 20 µl antibody solution for 30 min at room temperature. 50 µl 0.1 M sodium acetate buffer and 10 µl 1 mM MU-Neu5Ac were then added and the remaining enzyme activity was determined as described above.

Results

Cloning of the sialidase gene

After cleavage of chromosomal DNA from *Clostridium perfringens* strain A99 with various restriction enzymes, the resulting fragments were hybridized with three oligonucleotide probes (A, B, C). The fragments obtained with any single enzyme, gave rise to more than one band on the autoradiogram. However, after digestion with *Sau* 3AI, hybridization with probes A, B and C could be assigned to one range of fragments 2.8–3.4 kb in size. After ligation of these fragments into pUC and electroporation of *E. coli* JM101, two of the resulting clones contained an insert of 3.1 kb that still hybridized with all of the three probes. Both inserts exhibited an identical pattern of fragments, when cleaved with various restriction enzymes.

The insert of one clone (insert 1, Fig. 1) was digested with *Hind* III and *Hinc* II (Fig. 1), and the subfragments were sequenced from both ends. This sequence was completed by digestion of the whole insert with exonuclease BAL 31 and sequencing of the resulting subfragments, which overlapped each other. The entire sequence of 3106 bp contained an open reading frame (ORF1, Fig. 1) from position 1–600. Comparison with the data of the EMBL 29.0-SWISSPROT 20.0-database revealed no similarity with other sequences. This ORF is followed by an non-coding region. A further ORF, encoding the sialidase gene, extended from position 1264 to the end of the insert without a stop-codon. This clone did not express sialidase activity.

```
 D  P  I  P  V  E  P  Y  T  N  S  L  K  I  A  I  E  F  F  A  N  N  D  F    24
GATCCTATACCAGTTGAGCCATATACAAATTCCTTAAAGATAGCAATAGAATTTTTTGCAAATAATGATTTT   72

 A  R  F  R  F  V  T  K  Y  T  D  V  D  S  L  L  G  L  N  H  N  G  K  T    48
GCAAGATTTAGGTTCGTAACAAAATATACAGATGTTGATTCATTATTAGGTTTAAATCATAATGGCAAAACA  144

 E  V  R  F  T  I  N  T  D  F  V  I  N  N  Y  E  R  R  T  A  S  L  C  E    72
GAAGTAAGATTTACTATAAATACTGATTTTGTAATAAATAATTATGAGAGAAGAACAGCATCTCTTTGTGAA  216

 R  I  N  A  S  V  K  I  A  K  A  N  Y  P  L  G  F  I  I  A  P  V  F  I    96
AGAATAAATGCCAGTGTGAAAATAGCTAAGGCTAATTATCCATTAGGATTTATAATTGCTCCGGTATTTATA  288

 Y  E  G  W  K  E  D  Y  E  N  L  L  K  D  L  K  E  K  L  P  R  D  L  Q   120
TATGAGGGATGGAAAGAGGATTATGAGAACTTATTAAAGGATTTAAAAGAAAAATTACCAAGGGATTTACAA  360

 H  K  L  T  F  E  V  I  S  H  R  Y  T  T  R  A  K  N  I  I  M  D  I  F   144
CACAAATTAACCTTTGAGGTTATTTCCCATAGGTATACTACAAGAGCTAAAAATATAATAATGGATATATTT  432

 P  D  N  K  L  P  M  D  D  E  K  R  T  F  K  Y  G  Q  F  G  Y  G  K  Y   168
CCAGATAATAAATTGCCTATGGATGATGAAAAGAGGACTTTTAAATATGGTCAATTTGGTTATGGAAAATAT  504

 V  Y  K  K  E  D  I  H  E  I  K  E  F  F  M  E  I  I  D  K  Y  F  P  M   192
GTTTATAAAAAAGAAGATATACATGAAATAAAAGAGTTTTTTATGGAAATAATAGATAAATATTTTCCAATG  576

 S  E  I  K  Y  I  I  *                                                    199
AGTGAAATAAAATATATTATTTAAAAAATTAGTACACTTAGAGTTTTTAAAATGAGTTTGTTTAGTATAATT  648

TTTTTAAGTATCTAAAATTAAAAGAAATTAAAAAAACACCTTTAAGTTTTAAGAAAATTATTTTCAATAAAT  720

TTTATTTGCGAAAATAATTTTCTTTTTATTTAGAGTTAAGAAATATTTTCTTTCTGTAACTCCAAGAGTAAT  792

TGTTTGTATAGCAGTAAAGGTTAGATAATAAGTATTTTGAAATAAGGTTTATGTATCAATAATTAAAGAATA  864

ATAAGAAAGAATAAGAAAATTTTATAGAAAAACTTACCATAAAAATATCTTTATTATATTGAAATTTCTTTA  936

GAAAAGGATTACAATGTTAAATATAAGAAAAATATTTTCAAATTATTCTTATTTTGAAAATACTTTTCATAA 1008
                              <---- ---------- --    -- ---------- ----->
AAATGAAATATTTTAGTATTCAAATTTTAATCGTTGCATTTCAAGAAGTATATTCATAAAAGATTTAAGGAG 1080

AACTAAGAGTAAAAATCTTTTTTTCAAATTAAGGCAAAATTCTACTAAAAGAAGCATTTAAGATATTCTATGT 1152

TATAAATCCAAAGTTACAAATACAATAATAAAAATTTTACTTCGTATGTCATAGTATTTTAAATTAATACTA 1224

                                          M  N  Y  K  G  I  T  L  I  L  T   11
TTTCAAAATAATAAAAAATTATTTATGGGAGGATGTAAAATGAATTACAAAGGGATAACTTTAATTTTAACA 1296

 A  A  M  V  I  S  G  G  N  Y  V  L  V  K  G  S  T  L  D  S  G  K  N  N   35
GCTGCTATGGTTATAAGTGGTGGTAATTATGTATTGGTTAAAGGAAGTACTTTAGACTCAGGGAAAAATAAT 1368

 S  G  Y  E  I  K  V  N  N  S  E  N  L  S  S  L  G  E  Y  K  D  I  N  L   59
AGTGGGTATGAGATTAAAGTAAATAATAGTGAAAATTTAAGTTCACTAGGAGAATATAAGGATATAAATTTA 1440

 E  S  S  N  A  S  N  I  T  Y  D  L  E  K  Y  K  N  L  D  E  G  T  I  V   83
GAAAGTTCAAATGCTTCTAATATTACTTATGATTTAGAAAAATATAAAAATTTAGATGAAGGTACTATTGTT 1512

 V  R  F  N  S  K  D  S  K  I  Q  S  L  L  G  I  S  N  S  K  T  K  N  G  107
GTAAGATTTAATTCCAAAGATTCTAAAATACAAAGCTTATTAGGAATAAGTAATAGCAAGACTAAAAATGGA 1584
```

Figure 2.

```
 Y  F  N  F  Y  Y  V  T  N  S  R  V  G  F  E  L  R  N  Q  K  N  E  G  N  T  131
TATTTTAATTTTTATGTTACTAATTCAAGAGTTGGTTTTGAGCTTAGAAATCAAAAAAATGAGGGGAATACA 1656

 Q  S  G  T  E  N  L  V  H  M  Y  K  D  V  A  L  N  D  G  D  N  T  V  A  155
CAAAGTGGAACTGAAAATCTAGTACATATGTATAAAGATGTTGCTTTAAATGATGGGGATAATACTGTTGCC 1728

 L  K  I  E  K  N  K  G  Y  K  L  F  L  N  G  K  I  I  K  E  V  K  D  T  179
TTAAAGATTGAAAAAAATAAAGGTTATAAGCTCTTCCTTAATGGAAAAATCATAAAAGAAGTTAAAGATACA 1800

 N  T  K  F  L  N  N  I  E  N  L  D  S  A  F  I  G  K  T  N  R  Y  G  Q  203
AATACAAAGTTTTTAAATAATATAGAAAATTTAGATAGTGCTTTTATTGGAAAAACTAATAGATATGGACAA 1872

 S  N  E  Y  N  F  K  G  N  I  G  F  M  N  I  Y  N  E  P  L  G  D  D  Y  227
TCTAATGAATATAATTTTAAGGGAAATATAGGTTTTATGAATATATACAATGAACCTTTAGGTGATGATTAT 1944

 L  L  S  K  T  G  E  T  K  A  K  E  E  V  L  V  E  G  A  V  K  T  E  P  251
TTACTTAGTAAGACTGGAGAGACAAAAGCTAAAGAAGAAGTTCTTGTAGAGGGAGCTGTAAAGACAGAGCCA 2016

 V  D  L  F  H  P  G  F  L  N  S  S  N  Y  R  I  P  A  L  F  K  T  K  E  275
GTTGATTTATTTCATCCAGGATTTTTAAATTCTAGTAATTACAGAATACCAGCCTTATTTAAAACAAAGGAA 2088

 G  T  L  I  A  S  I  D  A  R  R  H  G  G  A  D  A  P  N  N  D  I  D  T  299
GGAACTTTAATAGCTTCAATAGATGCAAGAAGACATGGTGGAGCTGATGCTCCTAACAATGATATAGATACA 2160

 A  V  R  R  S  E  D  G  G  K  T  W  D  E  G  Q  I  I  M  D  Y  P  D  K  323
GCAGTTAGAAGAAGCGAAGATGGAGGGAAAACTTGGGATGAAGGACAAATAATAATGGACTATCCTGACAAA 2232
          ------------------------
 S  S  V  I  D  T  T  L  I  Q  D  D  E  T  G  R  I  F  L  L  V  T  H  F  347
TCCTCTGTTATAGATACAACTTTAATTCAAGATGATGAAACAGGAAGAATATTCTTGTTAGTAACTCATTTC 2304

 P  S  K  Y  G  F  W  N  A  G  L  G  S  G  F  K  N  I  D  G  K  E  Y  L  371
CCTTCAAAATATGGTTTTTGGAATGCTGGATTAGGAAGTGGATTTAAAAATATTGATGGAAAAGAATATTTA 2376

 C  L  Y  D  S  S  G  K  E  F  T  V  R  E  N  V  V  Y  D  K  D  G  N  K  395
TGTCTTTACGATTCATCAGGTAAAGAATTTACTGTAAGAGAAAATGTAGTATATGACAAAGATGGCAATAAA 2448

 T  E  Y  T  T  N  A  L  G  D  L  F  K  N  G  T  K  I  D  N  I  N  S  S  419
ACAGAATATACAACAAATGCTTTAGGAGATTTATTTAAAAATGGAACTAAGATAGATAATATAAATTCTAGT 2520

 T  A  P  L  K  A  K  G  T  S  Y  I  N  L  V  Y  S  D  D  D  G  K  T  W  443
ACAGCACCTTTAAAAGCAAAAGGGACAAGCTATATAAATCTTGTATATAGTGATGATGATGGAAAGACTTGG 2592
                                                ------------------------
 S  E  P  Q  N  I  N  F  Q  V  K  K  D  W  M  K  F  L  G  I  A  P  G  R  467
AGTGAGCCACAAAATATTAATTTCCAAGTTAAAAAGGATTGGATGAAGTTTTTAGGAATAGCTCCAGGTAGA 2664

 G  I  Q  I  K  N  G  E  H  K  G  R  I  V  V  P  V  Y  Y  T  N  E  K  G  491
GGAATACAAATAAAAAATGGAGAACATAAAGGAAGAATAGTAGTTCCTGTTTACTATACAAATGAAAAAGGT 2736

 K  Q  S  S  A  V  I  Y  S  D  D  S  G  K  N  W  T  I  G  E  S  P  N  D  515
AAACAATCTAGTGCTGTAATATATAGTGATGATAGTGGTAAGAATTGGACAATAGGAGAATCTCCAAATGAT 2808
                         ------------------------
 N  R  K  L  E  N  G  K  I  I  N  S  K  T  L  S  D  D  A  P  Q  L  T  E  539
AATAGAAAATTAGAAAATGGAAAAATAATAAATTCTAAAACCTTATCAGATGATGCACCTCAATTAACTGAA 2880

 C  Q  V  V  E  M  P  N  G  Q  L  K  L  F  M  R  N  L  S  G  Y  L  N  I  563
TGTCAAGTAGTAGAAATGCCAAACGGTCAACTAAAATTATTTATGAGAAATTTAAGTGGATATTTAAATATT 2952
```

Figure 2. *continued*

```
 A T S F D G G A T W D E T V E K D T N V L E P Y  587
GCTACAAGTTTTGATGGGGGAGCTACTTGGGATGAAACAGTAGAGAAAGATACTAATGTTTTAGAGCCATAT 3024
      ----------------------

 C Q L S V I N Y S Q K V D G K D A V I F S N P N  611
TGCCAACTAAGTGTTATAAATTATAGTCAAAAGGTAGATGGAAAGGATGCTGTTATTTTCTCAAATCCAAAT 3096

 A R S R S N G T V R I G L I N Q V G T Y E N G E  635
GCAAGAAGTAGATCAAATGGTACTGTGAGAATTGGATTGATTAATCAAGTTGGAACATACGAAAATGGTGAA 3168

 P K Y E F D W K Y N K L V K P G Y Y A Y S C L T  659
CCTAAATATGAATTTGATTGGAAATATAATAAATTAGTTAAACCAGGTTATTATGCTTATTCTTGTTTAACA 3240

 E L S N G N I G L L Y E G T P S E E M S Y I E M  683
GAATTAAGCAATGGAAATATTGGATTATTATATGAGGGTACACCAAGTGAAGAAATGTCTTATATTGAAATG 3312

 N L K Y L E S G A N K *                         694
AATTTAAAATATTTAGAGAGTGGAGCTAATAAATAAGAAATTATTAAAAATATTATTTTATGGTTTGAAAGG 3384

           M K K I W K Y G V L L A L P L V L A G C   20
GGTAAAAGATTATGAAAAAGATATGGAAGTATGGGGTATTACTTGCATTACCATTAGTTTTAGCGGGATGTG 3456

G N K P S E G S N K N K E I T S I K W T V D G A   44
GGAATAAGCCTTCAGAGGGTAGTAATAAAAATAAAGAAATTACTTCAATTAAGTGGACTGTTGATGGAGCGT 3528

L E A Q K G Q S K N I G T A G L L Y G K T K D Y   68
TAGAGGCTCAAAAAGGTCAAAGCAAAAATATAGGAACAGCAGGGTTACTTTATGGTAAAACAAAAGATTATA 3600

I V V G G G A N F P E E P V A K G G A K K L Y P   92
TAGTTGTTGGAGGTGGAGCTAACTTTCCAGAGGAGCCAGTAGCAAAAGGAGGAGCTAAAAAGCTTTATCCTG 3672

D I Y V L K N E N G D L K Q V D H T T L D Y E I  116
ATATATATGTATTAAAAAATGAAAATGGAGATTTAAAGCAAGTAGATCATACAACATTAGATTATGAAATAG 3744

G Y G S S I T T E E G I Y Y V G G S P N K E E G  140
GTTATGGAAGTTCCATAACTACTGAGGAAGGAATTTATTATGTAGGTGGAAGCCCTAATAAAGAAGAAGGAA 3816

N N V S L I T T D E K G K I G L K H I G D L P F  164
ATAACGTTTCTCTTATTACAACTGATGAAAAAGGTAAAATTGGTTTAAAGCATATTGGAGATTTACCTTTTA 3888

T F S D G F A V K K D K D I Y L G L G K Q D G K  188
CATTTAGTGATGGCTTCGCAGTTAAAAAAGATAAAGATATTTATTTAGGACTTGGAAAACAAGATGGAAAAG 3960

A S N K V Y K F N V E T G K T E E L K E I P G E  212
CTAGTAATAAGGTTTATAAATTTAATGTTGAAACAGGAAAAACAGAAGAATTAAAGGAAATACCAGGAGAAC 4032

P T R N Q A V A Q L L G D D I Y V F S G G D K I  236
CTACTAGAAATCAAGCCGTTGCTCAATTATTAGGTGATGATATATATGTATTTAGTGGTGGAGATAAAATTG 4104

A Y T D G Y K Y S I K D N S W T Q V A D V K I G  260
CTTACACAGATGGTTATAAATATAGCATTAAGGATAATTCATGGACTCAAGTAGCTGACGTTAAAATAGGAA 4176

N E E I S L L G A N S V K L N E D E M L V I G G  284
ATGAAGAAATTTCACTTTTAGGTGCTAATTCAGTTAAATTAAATGAAGATGAAATGCTTGTTATAGGTGGAT 4248

F N K A V Y D D A V K N L N T L K D E S L A E F  308
TTAATAAAGCGGTTTATGATGATGCTGTAAAAAATCTAAATACATTAAAAGATGAAAGTCTAGCTGAATTCA 4320

K K E Y F G R D P Q D F A W N T K V L I Y N S K  332
AGAAAGAGTATTTTGGAAGAGATCCTCAAGATTTTGCTTGGAATACAAAAGTTCTAATTTATAATTCAAAAA 4392
```

Figure 2. *continued*

```
N  N  E  W  K  S  I  G  D  V  P  F  N  A  P  C  G  E  A  L  I  L  D  G    356
ATAATGAGTGGAAGTCAATTGGAGATGTACCATTTAATGCTCCATGTGGAGAAGCATTAATATTAGATGGAG 4464

D  N  I  F  S  I  N  G  E  I  K  P  G  I  R  T  N  K  I  Y  S  G  E  L    380
ATAATATCTTCTCAATAAATGGAGAAATCAAACCAGGAATTAGAACAAATAAAATTTATTCAGGTGAATTAA 4536

I  Y  K  K  *                                                              384
TATATAAAAAGTAAATAAAATGAGATTGAGAAAGTAAAAAAAATAGGCGTGTATCAAAATATAATAAAACTT 4608

ACAGCTTCTAATAGTTGTTAGATCAGCTTTAGAATGTATGTAAGTTTTAAGGAAATTTATTTTGATACATGC 4680

CTTATTAGTTGTTCAATTATAATCTTTGATATTTTAAAGAATAACTTAAGAGAAAATAATATTTTGTAAAAT 4752

TTTCCTAATGATAATATTAAAGCTAAAAAATAGCAAGTATAATAGTGACCATAGATATTATAATTAAGAATA 4824

TTTTTATTAAGGATAATTTATAGCTAAAAGCTTATGAATAAAGGGATGCAACTAATAGTT             4884
```

Figure 2. The complete DNA- and deduced-amino acid sequence of the combined inserts 1, 2 and 3 including the gene for the 'large' sialidase of *Clostridium perfringens* A99 and two surrounding ORFs; arrows indicate inverted repeats; the putative ribosome binding site is boxed; the 'Asp-boxes' are underlined; stop-codons are marked by a star.

Table 1. Properties of the two sialidase isoenzymes of *Clostridium perfringens* A99.

Property	*'Large' isoenzyme*	*'Small' isoenzyme*
Molecular mass (kDa)		
Native	63	33
Denatured	71	43
Deduced from DNA	73	42
pI (pH)	5.2	5.15
Location	broth	cells
T_{opt} (°C)	55	36
pH_{opt}	5.0	6.1
Specific activity (U mg^{-1})	620	736
Substrate specificity	broad	limited

Another oligonucleotide probe (D) was developed from the region of the *Hinc* II-restriction site (Fig. 1) in order to detect the missing 3'-terminal part of the gene. This probe hybridized with a 2.1 kb fragment (*Hind* III) of chromosomal DNA. After cloning, one of the transformants contained the corresponding insert (insert 2, Fig. 1), partly overlapping the insert of the first clone (insert 1, Fig. 1). Within the sequence, the stop-codon of the sialidase gene was found, followed by a non-coding region without termination structures and a further ORF (ORF2, Fig. 1) from nucleotide 3496 to 3665 (Fig. 2), encoding 90 amino acids of an unknown protein with a hydrophobic N-terminus. No sialidase activity was expressed by this clone. From the 3'-terminus of this insert (insert 2 in Fig. 1) a further oligonucleotide probe (E) was deduced to obtain the remaining part of ORF 2. This probe hybridized with a 1.98 kb fragment (*Hinc* II) of chromosomal DNA, which was also subsequently cloned. Four transformants contained an insert still hybridizing with this probe; 0.76 kb of this insert (insert 3 in Fig. 1) overlapped with insert 2. The remaining part of insert 3 was digested with *Eco* RI and the resulting fragments were subcloned for sequencing. These fragments contained the missing part of ORF 2 including a stop-codon. The entire ORF 2 encompasses 1152 bp, corresponding to a putative protein of 384 amino acids, which shows no significant homology with data base sequences (EMBL 29.0-SWISSPROT 20.0-database of GENMON version 4.3).

The complete sequence of all three inserts, encoding the entire sialidase gene and the surrounding open reading frames, is shown in Fig. 2. The mol GC-content of the insert is 26.7% and of the sialidase structural gene is 28.7%.

Analysis of the gene product

Upstream from the start codon of the gene an inverted repeat was found (Fig. 2) that could possibly function as a promoter by forming secondary structures. A predicted ribosome-binding site was present at position −13 to −7. The entire gene encompasses 2082 bp, corresponding to 694 amino acids. The first 41 amino acids following fMet possibly represent a signal peptide, as is indicated by the presence of 28 hydrophobic amino acids. The molecular weight of the mature protein is 72.956 kDa. Adjacent to this signal peptide, the amino acid sequence already determined from N-terminal sequencing of the protein was recognized. A stretch of 12 amino acids was found four times at different positions within the predicted amino acid sequence of the protein (Fig. 2).

The amino acid sequence of the 'large' sialidase from *Clostridium perfringens* was compared with the primary structures of sialidases from related bacterial species and the 'small' isoenzyme of the same species. The alignment of four clostridial sialidase proteins is shown in Fig. 3.

The percentage of identical amino acids between each pair of sialidases was calculated (Table 2), and the similarity values were used to construct a dendrogram that is shown in Fig. 4.

Figure 3. Alignment of four clostridial sialidases. Sialidase primary structures from *Clostridium septicum* (C.se.), *Clostridium perfringens* (C.p.[l], 'large' and C.p.[s], 'small' isoenzyme) and *Clostridium sordellii* (C.so.) were aligned by eye. Residues are numbered from each N-terminus. Capital letters are used for identical amino acids in the respective column. A column with only one type of amino acid present is marked by double bold print. The 'Asp-boxes' are underlined and numbered 1–4. Stops are represented by stars. Gaps are indicated by hyphens.

Table 2. Similarities between clostridial sialidases. Each comparison gives relative percentages of identical amino acids after alignment and after normalizing primary sequences to 100%.

	Clostridium septicum		
Clostridium perfringens		*Clostridium perfringens*	
'large' isoenzyme	55	'large' isoenzyme	
Clostridium perfringens			*Clostridium perfringens*
'small' isoenzyme	30	26	'small' isoenzyme
Clostridium sordellii	32	27	72

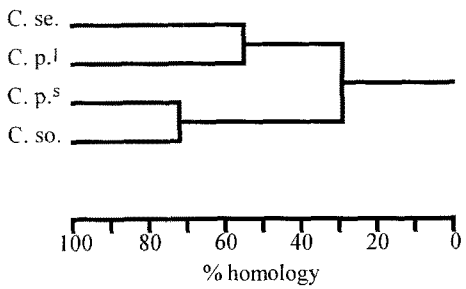

Figure 4. Dendrogram of sialidase relationship calculated from homologous amino acids by the average linkage method. (C. se., *Clostridium septicum*; C.p.^{l}, *Clostridium perfringens* 'large' isoenzyme; C.p.^{s}, *Clostridium perfringens* 'small' isoenzyme; C.so., *Clostridium sordellii*).

Reconstruction of the complete sialidase gene

Overlapping regions of inserts 1 and 2 were removed by restriction digest (*Hinc* II at position 2908 of the sequence, Fig. 1) and the resulting fragments were ligated. After electroporation, clones exhibited sialidase activity after being sprayed with MU-Neu5Ac. However, further cultivation of these clones resulted in strong reduction or complete loss of sialidase activity.

The identity of the cloned sialidase with the wild type 'large' sialidase was confirmed by examining some properties which allow distinction between the 'large' and the 'small' isoenzymes. Antibodies directed against the natural 'large' enzyme inhibited the recombinant enzyme almost completely. Furthermore, the optimum temperature and pH value were similar to the wild type enzyme (55 °C and pH 5.0).

Discussion

The gene encoding the 'large' sialidase isoenzyme of *Clostridium perfringens* A99 was closed after screening the chromosomal DNA of this species with a combination of oligonucleotide probes, which had either been deduced from the N-terminal sequence of the protein or from a conserved motif ('Asp-box'). This strategy has been successfully used to clone various other sialidase genes [13, 23, 29, 30].

With respect to the gene described here, it was difficult to assign the target sequence of every single oligonucleotide probe to a unique band on the autoradiogram as a large number of chromosomal DNA-fragments hybridized with the corresponding probe. This could be due to the presence of wobbles in the oligonucleotide sequences, which made it difficult to choose very stringent conditions for hybridization. On the other hand, the N-terminal probes in particular recognized several quite similar target sequences on the chromosome, which was confirmed by sequencing of such fragments (data not shown). The low GC-content calculated for the sialidase gene (28.7 mol% GC) is in accordance with the value known for the chromosomal DNA (24–27 mol% GC) of *Clostridium perfringens* [31].

Further cultivation of the clones expressing sialidase activity led to rapid decrease or complete loss of this property. Although there is no explanation for this phenomenon so far, it may be assumed that the enzyme protein or even the gene is degraded by endogenous proteases or nucleases. The formation of inclusion bodies containing the gene product, which has already been observed for other recombinant proteins produced in large quantities, may be another explanation.

Certain features of the 'large' sialidase of *Clostridium perfringens* (Table 1) differ markedly from those of the corresponding enzymes from other bacterial species, e.g. the high temperature optimum and the broad range of natural substrates hydrolysed. Nevertheless, its primary structure (Fig. 3) and wide substrate specificity compare well with those of other sialidases, especially with the enzyme from *Clostridium septicum*. In comparison with the 'small' sialidases from *Clostridium perfringens*, *Clostridium sordellii* [23] and *Salmonella typhimurium* [32], which exhibit a limited substrate specificity, the two 'large' enzymes possess additional peptide regions, which may be responsible for the broad range of substrates hydrolysed and may reflect an adaptation to bacterial requirements. It is unknown, whether the additional peptides present in large sialidases originated by insertions or deletions.

The 'Asp-boxes', typical for every microbial sialidase [20], are also present four times in the sequence of the 'large' enzyme of *Clostridium perfringens*. Further sequence motifs or single amino acids were found to be identical in many of the aligned sequences shown in Fig. 3, e.g. the RIP-region formed by amino acids 266–268 of the 'large'

enzyme. The function of the conserved amino acids is as yet unknown. However, the arginine 266 is postulated to be involved in substrate binding [33].

The similarity values calculated from identical amino acids between pairs of aligned sialidase sequences were used to construct a dendrogram (Fig. 4). The high degree of similarity between the 'large' sialidase isoenzyme of *Clostridium perfringens* and the enzyme of *Clostridium septicum* was expected and is due to the phylogenetically close relationship of both species [34]. The homology between the 'large' and the 'small' isoenzymes was surprisingly low. The latter is, however, closely related to the enzymes of *Clostridium sordellii* [23] or even *Salmonella typhimurium* [12, 32, 35].

The presence of promoter-like structures upstream from the sialidase gene and the long distance of 663 bp between its start-codon and the stop-codon of ORF1 make a functional relationship between ORF1 and the sialidase gene unlikely. The short intergenic region between the sialidase gene and ORF2, however, contains no regulatory elements known, thus potentially enabling the coordinate expression of both gene products. In *E. coli*, an inducible catabolic system for sialic acids is known, consisting of an acylneuraminate pyruvate-lyase and a sialic acid permease [36, 37]. A similar catabolic system may also be present in *Clostridium perfringens* [11]. As no similarities could be found between the sequence of ORF2 and the primary structure of the acylneuraminate pyruvate-lyase from *E. coli* [38], the only sequence of an 'aldolase' known so far, ORF2 might encode a sialic acid permease. This is further indicated by a putative signal-sequence and further stretches of hydrophobic amino acids within the amino acid sequence of ORF2, which may correspond to its location within the cytoplasmic membrane. In contrast, the 'aldolase' is known to be a soluble, cytosolic enzyme [10].

The 'large' sialidase of *Clostridium perfringens* seems to be functionally well adapted to its environment, as can be seen from its pH and temperature optimum and the broad range of sialic acid substrates hydrolysed. It is therefore suggested that at an early stage of evolution, the 'large' sialidase gene either evolved in this bacterium or was acquired from an exogenous source. This is supported by its high degree of sequence similarity with the sialidase of *Clostrdium septicum.*

The biological function of a second sialidase isoenzyme in *Clostridium perfringens* that lacks a signal peptide and exhibits limited substrate specificity, remains unclear. It may have been acquired only recently, which is indicated by the proximity of its gene to a phage integration site [39], so that the adaptation to bacterial requirements is not yet completed.

Acknowledgements

We are grateful to Professor Dr Friedrich Lottspeich, Max-Planck-Institut für Biochemie, Martinsried, for N-terminal sequencing, as well as to Dr Werner Liesack, Institut für Allgemeine Mikrobiologie, Universität Kiel and Professor Dr Rolf D. Horstmann, Tropeninstitut Hamburg, for synthesizing DNA probes.

This work was supported by Deutsche Forschungsgemeinschaft grant Scha 202/13, the Sialic Acids Society, Kiel, and the Fonds der Chemischen Industrie.

References

1. Schauer R (1983) In *Structural Carbohydrates in the Liver* (Popper H, Reutter W, Köttgen E, Gudat F, eds) pp. 83–97. Boston: MTP Press.
2. Corfield AP, Lambré CR, Michalski J-C, Schauer R (1992) *Conferences Philippe Laudat 1991*, pp. 111–34. Paris: INSERM.
3. Müller HE (1992) *Bioforum* **1/2**:16–21.
4. Godoy VG, Miller Dallas M, Russo TA, Malamy MH (1993) *Infect Immun* **61**:4415–26.
5. Engstler M, Reuter G, Schauer R (1993) *Mol Biochem Parasitol* **61**:1–14.
6. Roggentin T, Kleineidam RG, Majewski DM, Tirpitz D, Roggentin P, Schauer R (1993a) *J Immunol Meth* **157**:125–33.
7. Roggentin P, Schauer R (1993) In *Genetics and Molecular Biology of Anaerobic Bacteria* (Sebald M, ed) pp. 290–300. New York: Brock/Springer.
8. Roggentin P, Kleineidam RG, Schauer R (1992a) *BioEngineering* **2**:25.
9. Nees S, Veh R, Schauer R (1975) *Hoppe-Seylers Z Physiol Chem* **356**:1027–42.
10. Nees S, Schauer R, Mayer F, Ehrlich K (1976) *Hoppe-Seylers Z Physiol Chem* **357**:839–53.
11. Nees S, Schauer R (1974) *Behring Inst Mitt* **55**:68–78.
12. Roggentin P, Schauer R, Hoyer LL, Vimr ER (1993b) *Mol Microbiol* **9**:915–21.
13. Roggentin P, Rothe B, Lottspeich F, Schauer R (1988a) *FEBS Lett* **238**:314.
14. Messing J (1983) *Methods Enzymol* **101**:20–78.
15. Yanisch-Perron C, Vieira J, Messing J (1985) *Gene* **33**:103–19.
16. Marmur J (1961) *J Mol Biol* **3**:208–18.
17. Gebers R, Wehmeyer U, Roggentin T, Schlesner H, Kölbel-Boelke J, Hirsch P (1985) *Int J Syst Bacteriol* **35**:260–69.
18. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning – A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
19. Eckerskorn C, Mewes W, Goretzki H, Lottspeich F (1988) *Europ J Biochem* **176**:509–19.
20. Roggentin P, Rothe B, Kaper JB, Galen J, Lawrisuk L, Vimr ER, Schauer R (1989) *Glycoconjugate J* **6**:349–53.
21. Southern EM (1975) *J Mol Biol* **98**:503–17.
22. Davis LG, Dibner MD, Battey JF (1986) In *Methods in Molecular Biology*, pp 227–29. New York: Elsevier.
23. Rothe B, Roggentin P, Frank R, Blöcker H, Schauer R (1989) *J Gen Microbiol* **135**:3087–96.
24. Wallace RB, Johnson MJ, Hirose T, Miyake T, Kawashima EH, Itakura K (1981) *Nucl Acids Res* **9**:879–94.
25. Sanger F, Nicklen S, Coulson AR (1977) *Proc Natl Acad Sci USA* **74**:5463–67.

26. Anderberg MR (1973) *Cluster Analysis for Applications.* New York: Academic Press.
27. Roggentin P, Gutschker-Gdaniec GHM, Hobrecht R, Schauer R (1988b) *Clin Chim Acta* **173**:251–62.
28. Schauer R, Sander-Wewer M, Gutschker-Gdaniec GHM, Roggentin P, Randow EA, Hobrecht R (1985) *Clin Chim Acta* **146**:119–27.
29. Rothe B, Rothe B, Roggentin P, Schauer R (1991) *Mol Gen Genetics* **226**:190–97.
30. Henningsen M, Roggentin P, Schauer R (1991) *Biol Chem Hoppe-Seyler* **372**:1065–72.
31. Cato EP, George WL, Finegold SM (1986) In *Bergey's Manual of Systematic Bacteriology* (Sneath PHA, Mair NS, Sharpe ME, Holt JG, eds), Vol. 2, pp. 1141–200. Baltimore: Williams & Wilkins.
32. Hoyer LL, Hamilton AC, Steenbergen SM, Vimr ER (1992) *Mol Microbiol* **6**:873–84.
33. Roggentin T, Kleineidam RG, Schauer R, Roggentin P (1992b) *Glycoconjugate J* **9**:235–40.
34. Dorsch M (1990) *Ein Beitrag zur Phylogenie der Gram-positiven Eubakterien.* Kiel: Dissertation.
35. Guo X, Sinnott ML (1993) *Biochem J* **296**:291–92.
36. Vimr ER, Troy FA (1985a) *J Bacteriol* **164**:845–53.
37. Vimr ER, Troy FA (1985b) *J Bacteriol* **164**:854–60.
38. Ohta Y, Watanabe K, Kimura A (1985) *Nucl Acids Res* **13**:8843–52.
39. Canard B, Cole ST (1990) *FEMS Microbiol Lett* **66**: 323–26.