Glycoconjugate J (1989) 6:349-353

Conserved Sequences in Bacterial and Viral Sialidases

PETER ROGGENTIN[1], BERND ROTHE[1], JAMES B KAPER[2], JAMES GALEN[2], LOIS LAWRISUK[3], ERIC R VIMR[3] and ROLAND SCHAUER[1]*

[1] *Biochemisches Institut, Christian-Albrechts-Universität, Olshausenstr. 40, D-2300 Kiel, W. Germany*
[2] *Center for Vaccine Development, Division of Geographic Medicine, University of Maryland, Baltimore, Maryland 21201, USA*
[3] *Department of Veterinary Pathobiology, Microbiology Division, College of Veterinary Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

Received April 20, 1989.

Key words: sialidase, sequence homologies, Clostridium perfringens, C. sordellii, Salmonella typhimurium, Vibrio cholerae, Influenza A virus

The genes of the bacterial sialidases from *Clostridium sordellii* G12, *C. perfringens* A99, *Salmonella typhimurium* LT-2 and *Vibrio cholerae* 395 sequenced so far were examined for homologies and were compared with sequences of viral sialidases.

Each of the bacterial sialidases contains a short sequence of twelve amino-acids, which is repeated at four positions in the protein. All these sequences exhibit significant similarities. Comparing the repeated sequences of the four sialidases, five amino-acids were found to be highly conserved at defined positions: Ser-X-Asp-X-Gly-X-Thr-Trp. Additionally, most of the distances between the four repeated regions are also conserved among the different sialidases. The conserved bacterial sequences show similarity with sialidases of influenza A H7N1 and H13N9.

Sialidases (= neuraminidases, EC 3.2.1.18) are known to be essential in metazoan animals of the deuterostomic lineage for the turnover of sialoglycoconjugates. Some microorganisms like protozoa, bacteria and viruses also produce sialidases, which are involved in adhesion, nutrition or invasive destruction [1]. However, substrates for sialidase (e.g. colominic acid) are rarely synthesized by those bacteria expressing sialidase, in contrast to higher animals, which generally produce the enzyme as well as the substrate (oligosaccharides and glycoconjugates).

This situation raises questions about the origin of sialidase. Was the enzyme invented by progenotes, which may also have been able to synthesize a sialidase substrate of unknown function, or was the sialidase first synthesized by ancestors of the Echinodermata in parallel with its substrate?. A comparison of sequence data may give information on the relatedness of sialidases. No homology indicates a polyphyletic origin. If consensus sequences are exhibited, homology may be caused by function or phylogeny. As a first step, sequences of four bacterial sialidases have been investigated and compared with known viral sialidases.

* Author for correspondence

G12	+89	Ala	Lys	**Ser**	Thr	**Asp**	Asn	**Gly**	Gln	**Thr**	**Trp**	Asp	Tyr
A99	+71	Ala	Arg	**Ser**	Thr	**Asp**	Phe	**Gly**	Lys	**Thr**	**Trp**	Ser	Tyr
LT-2	+71	Ala	Arg	**Ser**	Thr	**Asp**	Gly	**Gly**	Lys	**Thr**	**Trp**	Asn	Lys
395	+263	Arg	Thr	**Ser**	Arg	**Asp**	Gly	**Gly**	Ile	**Thr**	**Trp**	Asp	Thr
G12	+158	Val	Tyr	**Ser**	Asp	**Asp**	Asn	**Gly**	Glu	**Thr**	**Trp**	Ser	Asp
A99	+140	Ile	Tyr	**Ser**	Asp	**Asp**	Asn	**Gly**	Leu	**Thr**	**Trp**	Ser	Asn
LT-2	+145	Tyr	Lys	**Ser**	Thr	**Asp**	Asp	**Gly**	Val	**Thr**	Phe	Ser	Lys
395	+585	Ile	Tyr	**Ser**	Asp	**Asp**	Gly	**Gly**	Ser	Asn	**Trp**	Gln	Thr
G12	+226	Ile	Tyr	**Ser**	Lys	**Asp**	Asn	**Gly**	Glu	**Thr**	**Trp**	Thr	Met
A99	+208	Ile	Tyr	**Ser**	Lys	**Asp**	Asn	**Gly**	Glu	**Thr**	**Trp**	Thr	Met
LT-2	+210	Ile	Tyr	**Ser**	Thr	**Asp**	---	**Gly**	Ile	**Thr**	**Trp**	Ser	Leu
395	+653	Phe	Leu	**Ser**	Lys	**Asp**	Gly	**Gly**	Ile	**Thr**	**Trp**	Ser	Leu
G12	+273	Tyr	Ile	**Ser**	Tyr	**Asp**	Met	**Gly**	Ser	**Thr**	**Trp**	Glu	Val
A99	+255	Tyr	Ile	**Ser**	His	**Asp**	Leu	**Gly**	Thr	**Thr**	**Trp**	Glu	Ile
LT-2	+254	Phe	Glu	Thr	Lys	**Asp**	Phe	**Gly**	Lys	**Thr**	**Trp**	Thr	Glu
395	+718	Trp	Phe	**Ser**	Phe	**Asp**	Glu	**Gly**	Val	**Thr**	**Trp**	Lys	Gly
H7N1	+99	Ile	Tyr	**Ser**	Lys	**Asp**	Asn	**Gly**	Ile	Arg	Ile	Gly	Ser
H13N9	+100	Ile	Tyr	Gly	Lys	**Asp**	Asn	Ala	Val	Arg	Ile	Gly	Glu
	+353	Phe	Ser	Tyr	Leu	**Asp**	Gly	**Gly**	Asn	**Thr**	**Trp**	Leu	Gly

Figure 1. Comparison of the conserved, repeated amino-acid sequences in the sialidases of *Clostridium sordellii* G12, *C. perfringens* A99, *Salmonella typhimurium* LT-2 and *Vibrio cholerae* 395. The occurrence of comparable sequences is also demonstrated in the sialidases of two types of influenza A virus (H7N1; H13N9). Highly conserved amino-acids are marked by extra bold print.

Materials and Methods

The nucleotide sequences of viral sialidases [2-6] and of a *Clostridium perfringens* A99 sialidase [7] have been published. Meanwhile the genes of sialidases from *C. sordellii* G12 [8], Salmonella typhimurium LT-2 [E.R. Vimr and L. Lawrisuk, unpublished results] and

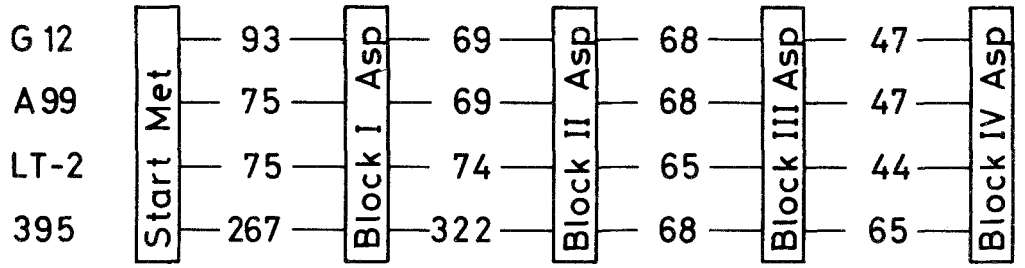

Figure 2. Number of amino-acids separating each of the central aspartic acid residues of the four conserved blocks and the start Met from the first block, respectively, when comparing the sialidases from *C. sordellii* G12, *C. perfringens* A99, *S. typhimurium* LT-2 and *V. cholerae* 395.

Vibrio cholerae 395 [9] have also been cloned and sequenced. The nucleotide and the predicted amino-acid sequences were compared using the Microgenie program (Microsoft, available from Beckman, Munich), which also contains a gene bank.

Results and Discussion

Comparison of the predicted amino-acid sequences reveals that each bacterial sialidase contains a region of twelve amino-acids which is repeated at four positions in the protein (Fig. 1). The repeated regions show homology among the four sialidases. This was not unexpected for the two clostridial sialidases, as the complete sequences have 70% homology [8]. However, a comparison of a clostridial sialidase sequence with that of *V. cholerae* showed that the repeated regions are the only sequences exhibiting significant similarities. A comparable region with up to seven homologous amino-acids was also detected in influenza A virus sialidases [2, 3], but not in sialidases of influenza B [4], Sendai [5] or Newcastle Disease virus [6].

Combining the sequence information of all repeated regions, the value expressed as percentage for the most frequently occurring amino-acid at each of the twelve positions could be determined:
Ile(37)-Tyr(42)-Ser(84)-Lys(32)-Asp(100)-Asn(37)-Gly(94)-Ile(21)-Thr(84)-Trp(84)-Ser(32)-X(11). The values demonstrate that five amino-acids are highly conserved at defined positions: Ser-X-Asp-X-Gly-X-Thr-Trp.

Nothing is known so far about the function of the conserved regions in bacterial sialidases. In influenza A virus (subtype N2) sialidase the Asp at position 103 is part of a turn involved in subunit binding [10]. This position may correspond to the Asp found in the conserved regions at the positions 104 or 105 in the sialidases of N1 and N9 subtypes, respectively, which were used for comparison. This function cannot be applied to bacterial sialidases, since they are reported to have no subunits.

Clostridium sordellii G12

+265	GCT	AAA	AGC	ACG	GAT	AAT	GGA	CAA	ACA	TGG	GAC	TAT
+472	GTT	TAT	TCT	GAT	GAT	AAT	GGA	GAA	ACA	TGG	TCT	GAT
+676	ATA	TAT	TCA	AAG	GAT	AAT	GGA	GAA	ACA	TGG	ACA	ATG
+816	TAT	ATA	TCT	TAT	GAT	ATG	GGA	TCT	ACT	TGG	GAA	GTA

Clostridium perfringens A99

+211	GCA	CGT	AGT	ACT	GAT	TTT	GGA	AAG	ACA	TGG	AGC	TAT
+418	ATT	TAT	TCT	GAT	GAC	AAT	GGA	TTA	ACT	TGG	TCT	AAT
+622	ATC	TAT	TCA	AAG	GAT	AAT	GGT	GAA	ACA	TGG	ACA	ATG
+762	TAC	ATC	TCT	CAT	GAT	TTA	GGA	ACC	ACT	TGG	GAA	ATA

Salmonella typhimurium LT-2

+211	GCC	CGT	AGC	ACA	GAT	GGA	GGG	AAA	ACC	TGG	AAT	AAA
+433	TAT	AAA	TCA	ACC	GAT	GAT	GGC	GTT	ACC	TTT	TCA	AAG
+628	ATA	TAC	TCC	ACT	GAT	GGA	---	ATA	ACA	TGG	TCA	TTG
+760	TTT	GAA	ACA	AAA	GAT	TTT	GGA	AAA	ACA	TGG	ACT	GAG

Vibrio cholerae 395

+786	CGT	ACC	TCA	CGA	GAT	GGC	GGT	ATA	ACT	TGG	GAT	ACC
+1752	ATT	TAC	AGT	GAT	GAT	GGC	GGT	TCA	AAC	TGG	CAA	ACC
+1956	TTT	TTG	AGT	AAA	GAT	GGT	GGA	ATC	ACG	TGG	AGC	CTA
+2151	TGG	TTT	AGC	TTC	GAT	GAA	GGG	GTG	ACA	TGG	AAA	GGA

Influenza virus (A/parrot/Ulster/73; H7N1)

+295	ATA	TAC	AGT	AAA	GAC	AAT	GGT	ATA	AGA	ATT	GGT	TCC

Influenza virus (A/whale/Maine/1/84; H13N9)

+298	ATA	TAT	GGG	AAA	GAC	AAT	GCG	GTA	AGA	ATT	GGA	GAG
+1057	TTT	TCA	TAC	CTA	GAT	GGA	GGT	AAT	ACT	TGG	CTA	GGG

Figure 3. Nucleotide sequences encoding the conserved and repeated amino-acid regions of the sialidases from *C. sordellii* G12, *C. perfringens* A99, *S. typhimurium* LT-2 and *V. cholerae* 395.
The corresponding sequences of sialidases of influenza A virus H7N1 and H13N9 are included for comparison. For orientation, the sequence encoding the central Asp is framed.

Additional information was obtained by comparing the distances between the four blocks of conserved amino-acids in each bacterial sialidase (Fig. 2). In particular, the distance of block II from block III is highly conserved in each of the four sialidases, a finding which is also valid for sialidases from *Clostridium* and *Salmonella* with regard to the distance between the blocks I and IV. The different values obtained for the number of amino-acids between

the starting Met and block I of both clostridial sequences is caused by the presence of a leader peptide in the *C. sordellii* sialidase [8].

Examination of the nucleotides which encode the conserved amino-acids reveals a reduced similarity (Fig. 3). For most of the amino-acids the codon usage varies even in one gene, but some codons are used in the conserved sequences of all sialidases compared. The difference between the members of two very distinct bacterial lineages is not apparent from the codon usage, except for the sixth amino-acid. Only in this codon is the expected preference of A+T by clostridial and G+C by enterobacterial genes realized.

The homologous sequences present in the four bacterial sialidases support the suggestion that the enzymes have one origin. Assuming a monophyletic development of sialidase, this enzyme must have been invented before the major divisions of Gram-negative bacteria split from Gram-positive bacteria in evolution. At present it can only be speculated that the homologous regions may be necessary for sialidase function, a suggestion which is supported by the observation that most of the distances between the four regions are also conserved. More information on this proposal will be obtained by experiments using site-directed mutagenesis.

The sequence information from the conserved regions enables the synthesis of DNA-probes, which may be useful in the finding of homologous genes in other microorganisms or even in animals.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (grant Scha 202/13-1) and Fonds der Chemischen Industrie.

References

1 Schauer R (1982) Sialic Acids, Cell Biol Monogr 10, Springer, Vienna.
2 Steuler H, Rohde W, Scholtissek C (1984) Virology 135:118-24.
3 Air GM, Webster RG, Colman PM, Laver WG (1987) Virology 160:346-54.
4 Shaw MW, Lamb RA, Erickson BW, Briedis DJ, Choppin PW (1982) Proc Natl Acad Sci USA 79:6817-21.
5 Miura N, Nakatani Y, Ishiura M, Uchida T, Okada Y (1985) FEBS Lett 188:112-16.
6 Jorgensen ED, Collins PL, Lomedico PT (1987) Virology 156:12-24.
7 Roggentin P, Rothe B, Lottspeich F, Schauer R (1988) FEBS Lett 238:31-34.
8 Rothe B, Roggentin P, Frank R, Blöcker H, Schauer R (1989) J Gen Microbiol, in press.
9 Vimr ER, Lawrisuk L, Galen J, Kaper JB (1988) J Bacteriol 170:1495-1504.
10 Colman PM, Varghese JN, Laver WG (1983) Nature 303:41-44.