Glycobiology vol. 11 no. 7 pp. 523–531, 2001

Glycosylation analysis of two cysteine proteinase inhibitors from Atlantic salmon skin: di-O-acetylated sialic acids are the major sialic acid species on N-glycans

Anne Ylönen[2], Nisse Kalkkinen[2], Juhani Saarinen[2], Jarl Bøgwald[3], and Jari Helin[1,2]

[2]Institute of Biotechnology, Protein Chemistry Laboratory, P.O. Box 56 (Viikinkaari 9), University of Helsinki, FIN-00014 University of Helsinki, Finland and [3]Norwegian College of Fishery Science, University of Tromsø, N-9037 Tromsø, Norway

Received on September 6, 2000; revised on January 5, 2001; accepted on January 29, 2001

We have recently identified two novel cysteine proteinase inhibitors from the skin of Atlantic salmon (*Salmo salar* L.), named salmon kininogen and salarin. In preliminary experiments, the proteins were found to be both N- as well as O-glycosylated. In the present study we show that both proteins carry biantennary α2,3-sialylated N-glycans. A very high amount of O-acetylated Neu5Ac units are present in the N-glycans, comprising about 60% di-O-acetylated species. Non-O-acetylated Neu5Ac make up less than 5% of the sialic acids in the N-glycans. A small number of Neu5Acα2-8Neu5Ac structures were observed in the N-glycans as well. O-glycans from both proteins were recovered by reductive beta-elimination and were identified by mass spectrometric methods as mono- and disialylated core type 1 tri- and tetrasaccharides. The method used for O-glycan isolation prevented the identification of possible O-acetylation in the O-glycan-bound sialic acids, but O-acetylation was observed in one O-glycosylated peptide isolated from trypsin digest of salarin. The chemical nature of the sialic acid modifications was further studied by liquid chromatography tandem mass spectrometry of 1,2-diamino-4,5-methylenedioxybenzene–derivatized sialic acids, revealing 7-, 8-, and 9- but no 4-O-acetylation. To our knowledge, these are the first observations of sialic acid O-acetylation in N-glycans on fish species and represent clearly the most extensive N-glycan O-acetylation described on any species.

Key words: cysteine proteinase inhibitor/glycoprotein/mass spectrometry/O-acetylation/sialic acid

Introduction

We are studying cysteine proteinase inhibitors present in the skin of various northern Atlantic fish species. During the characterization of two novel inhibitors from Atlantic salmon (*Salmo salar* L.), we found these to be glycoproteins carrying both N- and O-glycosidic glycans (Ylönen *et al.*, 1999). The present study was initiated to characterize the detailed structures of both types of glycans in these glycoproteins, as studies on Atlantic salmon glycoproteins are very scarce. Antithrombin from Atlantic salmon has been shown by lectin blotting to carry complex-type, α2,3-sialylated N-glycans, and no O-glycans were observed (Andersen *et al.*, 2000). The closest relative of Atlantic salmon, brown trout (*Salmo trutta*), has been reported to express glycosylated albumin with sialylated (but otherwise undefined) N-glycans (Metcalf *et al.*, 1998).

The glycosylation in other salmonids shows some intriguing features. Very large N-glycans carrying Lewis X and α2,8-disialyl groups have been described from a rainbow trout (*Oncorhynchus mykiss*) ovarian fluid glycoprotein (Funakoshi *et al.*, 1997). Inoue and Iwasaki described a polysialoglycoprotein from the unfertilized eggs of rainbow trout carrying O-glycosidic oligo-N-glycolylneuraminyl (poly-Neu5Gc) chains (Inoue and Iwasaki, 1978). Later studies demonstrated similar structures in other salmonids as well (Iwasaki and Inoue, 1985). Of special interest are the poly-Neu5Gc chains of kokanee salmon (*Oncorhynchus nerka adonis*), in which the Neu5Gc units are O-acetylated at various positions (Iwasaki *et al.*, 1990).

O-acetylation is perhaps the most common modification found on sialic acids. These substitutions are widespread in higher animals and are also found in several fungi and many, usually pathogenic, bacteria (Varki, 1992; Klein and Roussel, 1998; Alviano *et al.*, 1999). Some intriguing aspects of the biological significance of sialic acid O-acetylation in vertebrates have been established. The lifetimes of serum glycoproteins and blood cells are extended by O-acetylation of their sialic acid residues, because this substitution confers resistance to the action of most sialidases (Corfield *et al.*, 1986), as well as to acid-catalyzed hydrolysis (Varki and Diaz, 1984; Vandamme-Feldhaus and Schauer, 1998). These substitutions have also been implicated in protection against microorganisms, as, for example, the binding of influenza A and B viruses as well as malarial parasites to target cells is abrogated by sialic acid O-acetylation (for review, see Varki, 1992).

In the present study we describe the structures of both N- and O-glycosidic glycans from two cysteine proteinase inhibitors, salmon kininogen and salarin, isolated from Atlantic salmon skin. We present data showing that the N-glycans of these proteins are very highly O-acetylated at their nonreducing terminal sialic acid units, with practically no unsubstituted Neu5Ac present. The major N-glycoproteins in salmon skin, transferrin and GP55, were found to carry similar modifications.

[1]To whom correspondence should be addressed

Results

N-glycosylation of salmon kininogen and salarin

The presence of glycosylation in the two novel cysteine proteinase inhibitors of salmon skin was implied by the difference of the apparent molecular weight (MW) on SDS–polyacrylamide gel electrophoresis (PAGE) and the MW obtained by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). Moreover, by treatment with peptide:N-glycosidase F (PNGase F) the apparent MW of both salmon kininogen and salarin was reduced, thus confirming the presence of N-glycosylation (Ylönen *et al.*, 1999).

To obtain N-glycans for more detailed structural characterization, a sample of salmon kininogen (30 µg, about 0.6 nmol) was subjected to N-glycan liberation by PNGase F. Isolation of the liberated N-glycans by gel filtration yielded about 1 nmol of oligosaccharides (as determined by UV absorbance), thus implying two N-glycosylation sites. The N-glycan pool was subjected to MALDI-TOF MS (Figure 1A). The spectrum shows three clusters of ions, and within each cluster the ions are observed at 42-Da intervals. This implies variable degree of acetylation in the glycan structures, O-acetylation of sialic acid residues being the most probable alternative (Schauer, 1991). The signals were tentatively identified as sialylated biantennary complex-type glycans by their MW: *m/z* 1932.8, 1974.2, and 2016.4, monosialylated glycans with 0, 1, and 2 acetyl (Ac) groups, respectively; *m/z* 2222.4, 2264.9 (+ 1 Ac), 2307.1 (+ 2 Ac), 2349.1 (+ 3 Ac), and 2391.1 (+ 4 Ac), disialylated glycans; and *m/z* 2598.7 (+ 2 Ac), 2640.7 (+ 3 Ac), and 2682.6 (+ 4 Ac), trisialylated glycans (all $[M–H]^-$ ions). The relative intensities of the signals can be used to estimate the molar ratio of the components in the glycan mixture (Papac *et al.*, 1996): the data thus implies that about 60% of sialic acids are di-O-acetylated.

Short alkaline hydrolysis (saponification) of the N-glycans in 0.1 M NaOH yielded glycans devoid of additional acetyl groups, an indication that these indeed were O-acetyl substituents (Varki and Diaz, 1984) (Figure 1B). The MALDI-TOF spectrum of the de-O-acetylated glycans reveals a small signal of monosialylated biantennary glycan (*m/z* 1931.6 $[M–H]^-$), disialylated counterparts (*m/z* 2223.1 $[M–H]^-$, *m/z* 2245.2 $[M+Na–2H]^-$), and trisialylated counterparts (*m/z* 2514.3 $[M–H]^-$, 2536.5 $[M+Na–2H]^-$, 2558.7 $[M+2Na–3H]^-$).

Exoglycosidase digestions followed by MALDI-TOF MS analysis were carried out to obtain structural information of the N-glycans. The intact glycan pool was first treated with Newcastle disease virus (NDV) neuraminidase, which partly removed monoacetylated sialic acids from the glycans (not shown). This data agree with earlier studies on the specificity of neuraminidases against O-acetylated sialoglycoconjugates (Corfield *et al.*, 1986). When the glycan pool was de-O-acetylated and treated with NDV neuraminidase (known to remove only α2,3- and α2,8-linked sialic acids), a complete removal of sialic acid units was seen (Figure 1C): The MALDI-TOF spectrum now shows one glycan species only, *m/z* 1663.5 $[M+Na]^+$,1679.5 $[M+K]^+$, with MW appropriate for nonsialylated biantennary N-glycan.

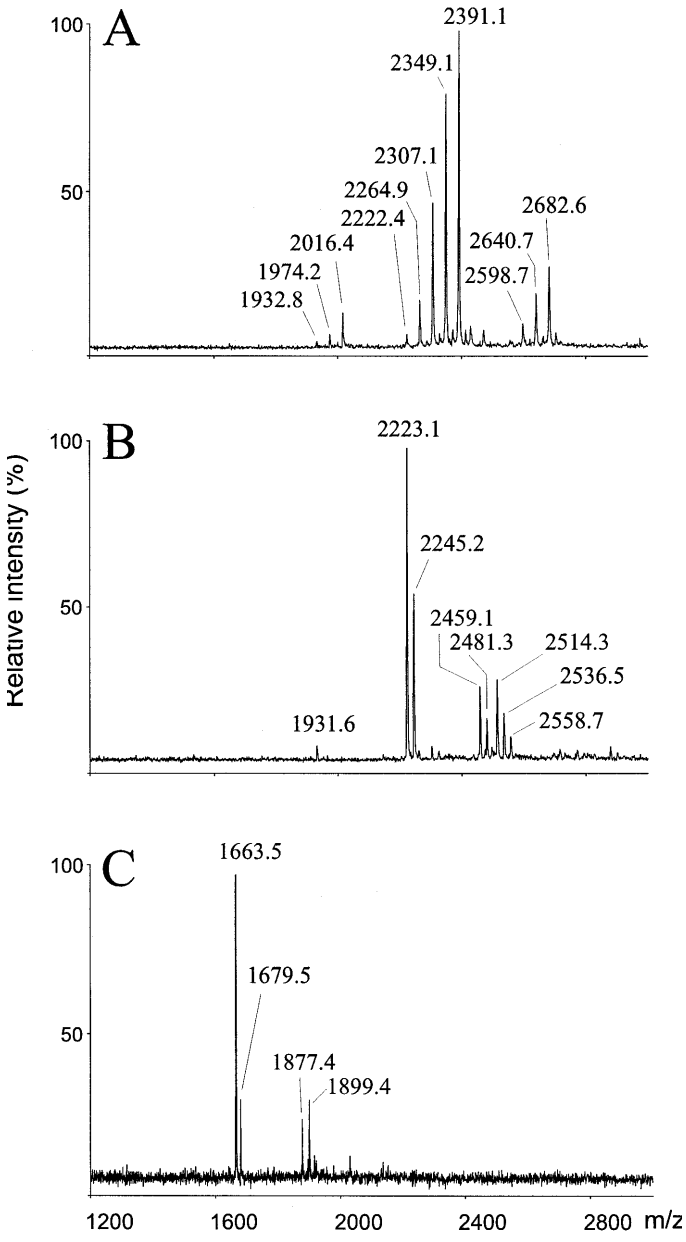

Fig. 1. MALDI-TOF spectra of salmon kininogen N-glycans. (**A**) Intact glycans, (**B**) saponified N-glycans, and (**C**) glycans after treatment with NDV sialidase. Spectra in (**A**) and (**B**) were measured in negative ion mode and (**C**) in positive ion mode. The signals at *m/z* 2459.1 and 2481.3 in (**B**) as well as at *m/z* 1877.4 and 1899.4 in (**C**) are matrix adducts typical to trihydroxyacetophenone matrix when samples contain Na-salts.

Finally, the desialylated glycans were treated with *Streptococcus pneumoniae* β-galactosidase. Under the experimental conditions used, this enzyme is specific for the β1,4-linkage and no Galβ1,3-linkages will be cleaved. Two galactose units were removed as analyzed by MALDI-TOF MS (not shown), a result compatible with β1,4-galactosylated biantennary structure. To sum up, these results indicate that the N-glycans of salmon kininogen were biantennary glycans terminated by O-acetylated α2,3-linked sialic acids.

The trisialylated biantennary species carries a Neu5Ac-Neu5Ac disialosyl group at the nonreducing termini of the N-glycan. This was verified by electrospray ionization (ESI) tandem mass spectrometry (MS/MS) analysis of the trisialylated species in permethylated form (not shown). The disialosyl group was α2,8-linked as established by periodate oxidation of the saponified N-glycans: MALDI-TOF MS analysis of the periodate-oxidized –$NaBH_4$ reduced N-glycans revealed three signals, *m/z* 1873.6, *m/z* 2105.2, and *m/z* 2396.6 ([M–H]$^-$ ions) (not shown). The *m/z* 1873.6 and *m/z* 2105.2 signals are generated from mono- and disialylated glycans, respectively, by the oxidative cleavage of the sialic acid side chain: Each sialic acid is truncated to C7-analogue, and –60 Da loss per each sialic acid is evident in the spectrum. The trisialylated glycan produced *m/z* 2396.6, indicating that only terminal sialic acids were oxidatively cleaved. This is only possible if the sialic acids are α2,8-linked, which renders the side chain of the inner sialic acid unit stable against periodate cleavage. This data is supported by the disappearance of the trisialylated glycan signal after treatment with NDV sialidase (Figure 1B,C). It is noteworthy that there was no more than four O-acetyl groups in the trisialylated biantennary glycan (see Figure 1A). This may suggest that only the terminal Neu5Ac unit is an acceptor for the O-acetyltransferase(s).

A sample of 30 μg (approximately 0.7 nmol) of salarin was treated with PNGase F, and 0.5 nmol of N-glycans were isolated by gel filtration, thus implying one N-glycosylation site in this protein. All structural studies described above for salmon kininogen N-glycans were also performed for salarin N-glycans. Practically identical results were obtained (not shown).

Identification of salmon kininogen and salarin N-glycosylation sites

Because the complete amino acid sequence of salmon kininogen or salarin is not known, we attempted to establish the N-glycosylation sites by characterizing glycopeptides from both proteins. Samples of salmon kininogen and salarin (about 30 μg each) were alkylated with 4-vinylpyridine and digested with trypsin. The tryptic peptides were separated by narrow-bore reversed phase (RP) chromatography (Figures 2 and 3), and aliquots of the fractions were subjected to MALDI-TOF MS and N-terminal sequencing. Glycopeptides in our experience seldom produce high-quality MALDI-TOF spectra, and identifying them in, for example, spectra of unseparated tryptic peptide mixtures may be impossible. This may partly be ascribed to their ionization properties, but logically also the heterogeneity of the glycan unit (i.e., glycoforms) will spread the signal over a number of mass values. The latter however offers an immediate possibility to identify high-performance liquid chromatography (HPLC) fractions containing glycopeptides, where, for example, differences of 291 Da (N-acetylneuraminic acid) or 146 Da (fucose) may be observed. For salmon kininogen and salarin, the utmost heterogeneity in the N-glycopeptides is sialic acid acetylation, yielding signals 42 Da apart. This is clearly seen in the spectra of salmon kininogen N-glycopeptides (SK-GPs, Figure 2B,C) and salarin N-glycopeptide (Sr-GP2, Figure 3C). N-terminal sequencing of SK-GP1 yielded AQXDSCTQYFVEFNSR, X representing a probable glycosylated asparagine residue in a N-glycosylation consensus sequence. The signals in the MALDI-TOF

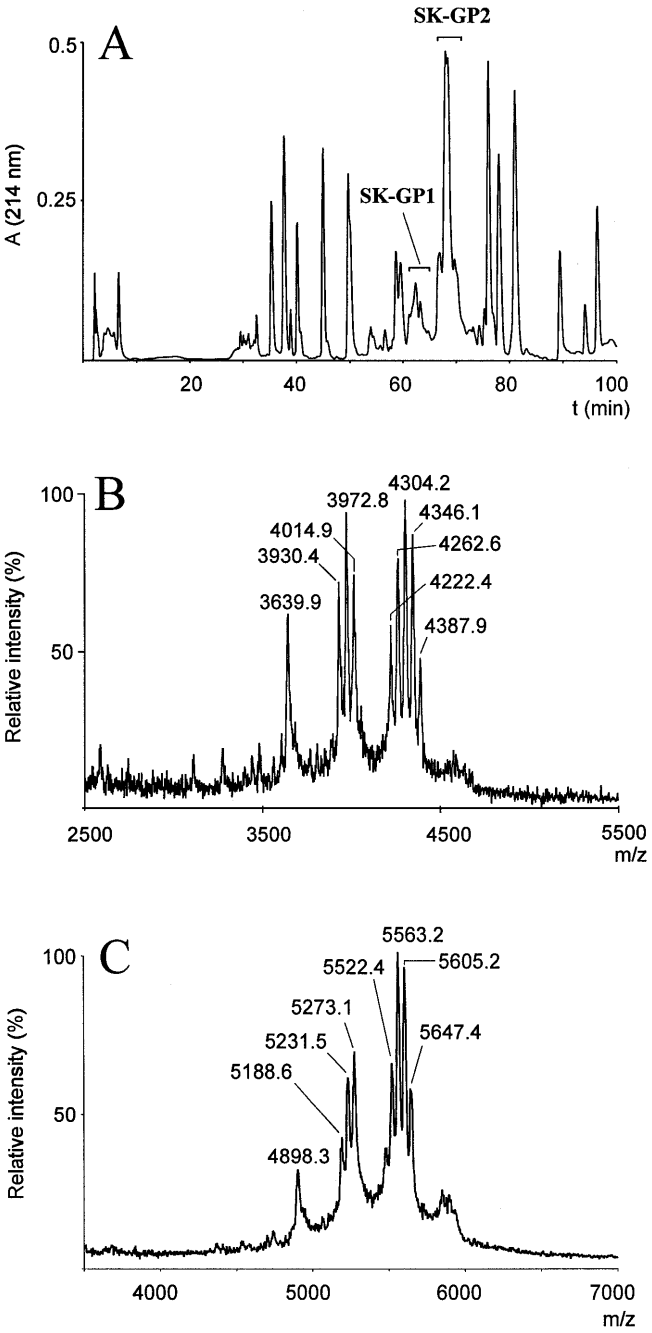

Fig. 2. (**A**) Separation of tryptic peptides from salmon kininogen with narrow-bore RP chromatography. (**B**) MALDI-TOF spectrum of SK-GP1. (**C**) MALDI-TOF spectrum of SK-GP2. Both spectra were recorded in positive ion mode with external calibration. See Table I for assignments.

spectrum of SK-GP1 are assigned as follows (Figure 2B, Table I): *m/z* 3639.9, peptide + nonsialylated biantennary glycan; *m/z* 3930.4, 3972.8 (+ 1 Ac), 4014.9 (+ 2 Ac) monosialylated (O-acetylated) counterpart; *m/z* 4222.4, 4262.6 (+ 1 Ac), 4304.2 (+ 2 Ac), 4346.1 (+ 3 Ac), 4387.9 (+ 4 Ac), disialylated (O-acetylated) counterparts. Compared to the isolated N-glycans, the glycan part in the glycopeptide shows remarkable differences: The appearance of signals corresponding to nonsialylated and

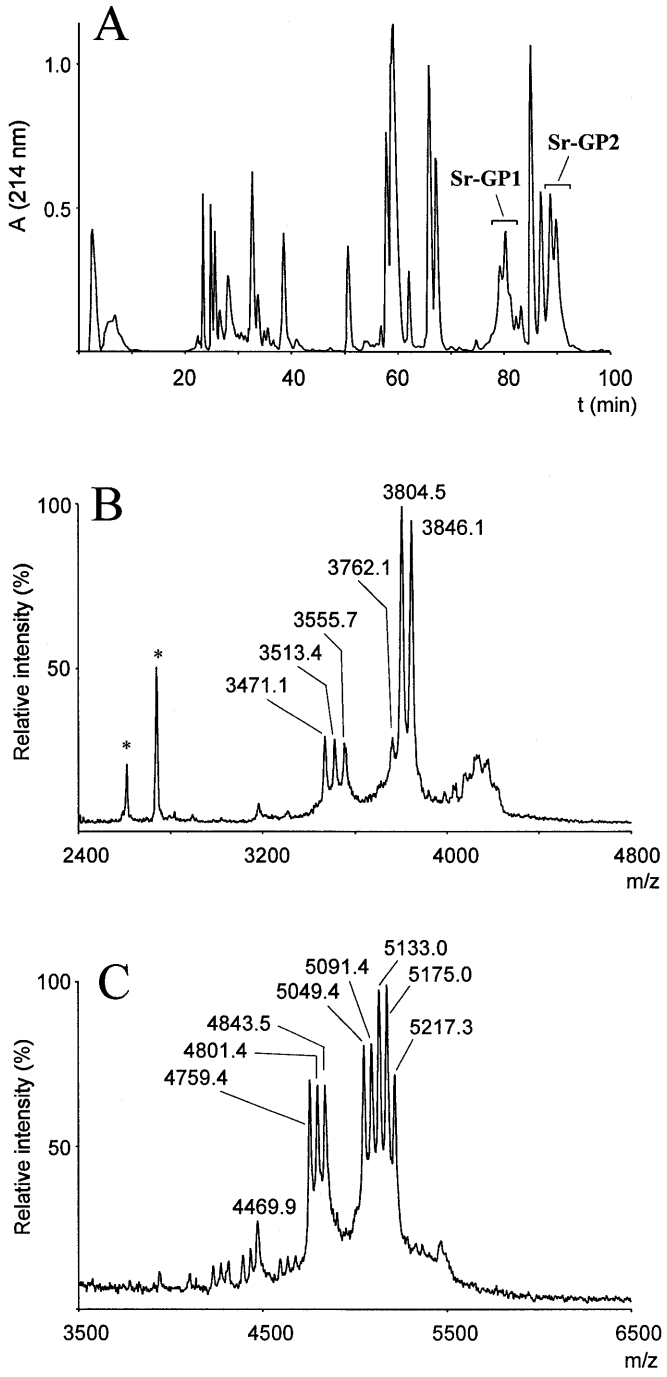

Fig. 3. (A) Separation of tryptic peptides from salarin with narrow-bore RP chromatography. (B) MALDI-TOF spectrum of Sr-GP1, measured in positive ion mode. (C) MALDI-TOF spectrum of Sr-GP2, measured in negative ion mode. See Table I for assignments. Large signals marked with asterisks in (B) are contaminating peptides in the HPLC fraction.

monosialylated structures is probably mostly due to fragmentation in the MALDI-TOF MS analysis (Juhasz and Costello, 1992). Furthermore, partial de-O-acetylation is very likely under the conditions trypsin digestion was carried out, accounting for the lower overall O-acetylation in the glycopeptide. SK-GP1 was also treated with PNGase F, and MALDI-TOF MS analysis of

Table I. The signal assignments of glycopeptide MALDI-TOF spectra[a]

Glycopeptide/sequence[b]	Signal (*m/z*)	Glycan
SK-GP1		
AQN*DSCTQYFVEFNSR[C]		
	3639.9	$(Hex)_5(HexNAc)_4$
	3930.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_1$
	3972.8	$(Hex)_5(HexNAc)_4(Neu5Ac)_1(OAc)_1$
	4014.9	$(Hex)_5(HexNAc)_4(Neu5Ac)_1(OAc)_2$
	4222.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_2$
	4262.6	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_1$
	4304.2	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_2$
	4346.1	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_3$
	4387.9	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_4$
SK-GP2		
ELNDECHPDPDVELAHCN*STVDVAPWR		
	4898.3	$(Hex)_5(HexNAc)_4$
	5188.6	$(Hex)_5(HexNAc)_4(Neu5Ac)_1$
	5231.5	$(Hex)_5(HexNAc)_4(Neu5Ac)_1(OAc)_1$
	5273.1	$(Hex)_5(HexNAc)_4(Neu5Ac)_1(OAc)_2$
	5522.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_1$
	5563.2	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_2$
	5605.2	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_3$
	5647.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_4$
Sr-GP1		
LMVVFPT*RDGGEEAEVDKEFETWK		
	3471.1	$(Hex)_1(HexNAc)_1(Neu5Ac)_1$
	3513.4	$(Hex)_1(HexNAc)_1(Neu5Ac)_1(OAc)_1$
	3555.7	$(Hex)_1(HexNAc)_1(Neu5Ac)_1(OAc)_2$
	3762.1	$(Hex)_1(HexNAc)_1(Neu5Ac)_2$
	3804.5	$(Hex)_1(HexNAc)_1(Neu5Ac)_2(OAc)_1$
	3846.1	$(Hex)_1(HexNAc)_1(Neu5Ac)_2(OAc)_2$
Sr-GP2		
AEN*GSESFTMGINYFSDMTFEEIPK		
	4469.9	$(Hex)_5(HexNAc)_4$
	4759.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_1$
	4801.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_1(OAc)_1$
	4843.5	$(Hex)_5(HexNAc)_4(Neu5Ac)_1(OAc)_2$
	5049.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_2$
	5091.4	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_1$
	5133.0	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_2$
	5175.0	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_3$
	5217.3	$(Hex)_5(HexNAc)_4(Neu5Ac)_2(OAc)_4$

[a]The MALDI-TOF mass spectra are presented in Figures 2 and 3. All mass values are average and $[M+H]^+$ ions, except those of Sr-GP2, which are $[M-H]^-$ ions.
[b]Peptide sequence analyzed by Edman sequencing.
[c]The glycosylated residue is followed by an asterisk.

the deglycosylated peptide yielded *m/z* 2014.7 (not shown), a value pertinent for the sequence obtained (considering N→D conversion of the glycosylated residue in the PNGase F treatment).

The SK-GP2 gave the sequence ELNDECHPDPDVELAH-CXSTVDVAPWR, X being the glycosylated residue, and its MALDI-TOF spectrum reveals the expected heterogeneity due to the glycan part (Figure 2C, signal assignments in Table I). Deglycosylation of SK-GP2 generated a peptide with *m/z* 3272.3 (not shown), appropriate for the sequence.

Two fractions exhibiting similar O-acetylation and sialic acid heterogeneity were observed among the salarin tryptic peptides (Figure 3). Sr-GP1 was, however, found to be stable against PNGase F treatment, thus implying this fraction to be O-glycopeptide, and its structural characterization will be described below. Sr-GP2 yielded sequence AEXGSESFTM-GINYFSDMTFEEIPK by N-terminal sequencing, and after de-N-glycosylation the peptide exhibited *m/z* 2845.1 in MALDI-TOF MS (not shown), confirming that the third amino acid is a glycosylated asparagine. Assignments of the signals in Sr-GP2 MALDI-TOF spectrum (Figure 3C) are shown in Table I.

Glycopeptides derived from both salmon kininogen and salarin elute in HPLC as partially resolved molecular species explaining their appearance in the chromatogram. The more highly O-acetylated species were found to elute later (not shown), as expected by their increased hydrophobicity.

Characterization of O-glycosylation in salmon kininogen and salarin

O-glycosidically bound glycans in salmon kininogen and salarin were liberated by alkaline reductive β-elimination. The isolated O-glycans were permethylated and analyzed by MALDI-TOF MS. Both proteins were found to carry mostly $(Neu5Ac)_2(Hex)_1(HexNAc\text{-}ol)_1$ species, as deduced from the mass spectrum (not shown). Many salmonids carry oligo–sialic acid structures in polysialoglycoprotein O-glycans (Iwasaki and Inoue, 1985), and therefore it was interesting to subject the salmon skin glycoprotein O-glycans to MS/MS, which is a powerful technique in analyzing isomeric glycans (Reinhold *et al.*, 1995). Figure 4 shows the MS/MS spectrum of salmon kininogen–derived disialylated O-glycan in permethylated form (parent ion $[M+Na]^+$, *m/z* 1256.5). Of particular interest in distinguishing between isomeric structures are the following fragment ions: The Y_{2a}/Y_{1b} ion at *m/z* 506.2 represents a Hex-HexNAc-ol ion, which has lost two substituents, that is, the two sialic acid residues did not form a diNeu5Ac unit. The C_{2a} ion at *m/z* 620.2 is a Neu5Ac-Gal ion, fragmented most probably from the 3-OH of the reduced GalNAc-ol. The Neu5Ac-GalNAc-ol Y_{1a}-H_2O fragment at *m/z* 659.2 also supports the view that Neu5Ac-Gal was linked to 3-OH of GalNAc-ol, because the loss of 3-substituent is often accompanied by water elimination (Egge and Peter-Katalinic, 1987). Taken together, the MS/MS data implies that salmon kininogen tetrasaccharide O-glycan is a core type 1 glycan (Hounsell *et al.*, 1996), carrying sialic acids at both Gal and GalNAc-ol units. There was no signal observed at *m/z* 759, where the B-ion of Neu5Ac-Neu5Ac disialyl group would reside, suggesting that not even a small amount of these structures were present.

Salarin-derived disialylated O-glycan gave practically identical results in MS/MS (not shown). The monosialylated O-glycans observed in the MALDI-TOF MS in low amounts were not studied further.

Alkaline β-elimination used in O-glycan isolation destroys O-acetyl substituents, and therefore no evidence was obtained about the presence of O-acetylation in sialic acids of O-glycans. However, as stated above, we found one probable O-glycopeptide among the tryptic peptides of salarin, Sr-GP1 (Figure 3A). This fraction yielded amino acid sequence LMVVFPXRDGGEEAEVDKEFETWK; the nonidentified amino acid is not in an N-glycosylation consensus sequence, and the fraction was resistant to PNGase F, excluding N-glycosylation. The tentative assignment of Sr-GP1 MALDI-TOF

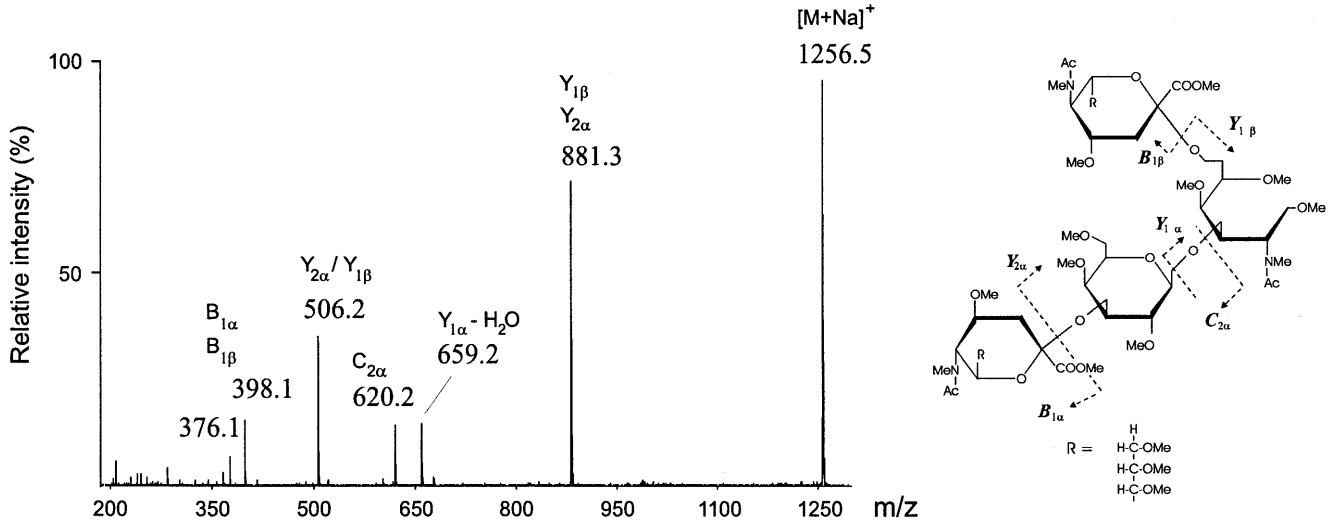

Fig. 4. MS/MS spectrum of permethylated O-glycan from salmon kininogen. The proposed structure with fragment assignments (nomenclature of Damon and Costello, 1988) is also shown.

spectrum (Figure 3B, Table I) suggested the unidentified amino acid to be threonine, which was verified by gas-phase β-elimination of the O-glycopeptide in methylamine (Mirgorodskaya and Roepstorff, 1998). In this procedure, the O-glycosylated residue loses the glycan part and is converted to a stable methylamine derivative. The mass spectrum of methylamine treated O-glycopeptide revealed a major signal at *m/z* 2825.2 (not shown), appropriate for the β-eliminated glycopeptide carrying a methylamine derivative of threonine.

The MALDI-TOF spectrum of Sr-GP1 clearly shows the presence of additional acetylation, implying O-acetylation in the O-glycans of salarin as well. We were not able to detect any O-glycopeptides among the tryptic peptides of salmon kininogen. The O-glycans of this glycoprotein may all reside clustered in one area of the protein backbone that contains no trypsin cleavage sites.

Characterization of sialic acid species

Sialic acids from salmon kininogen and salarin were isolated by hydrolysis in 2 M propionic acid (Mawhinney and Chance, 1994), and purified from proteins by running the hydrolysate through a POROS R1 micro-column. These hydrolysis conditions have been developed to minimize the loss of O-acetyl substituents. The corresponding sialic acid quinoxalinone derivatives were formed by reaction with 1,2-diamino-4,5-methylenedioxybenzene (DMB) reagent (Hara *et al.*, 1989). Preliminary analysis of the sialic acid–DMB derivatives from both salarin and salmon kininogen by LC-MS indicated that Neu5Ac and its mono- and di-O-acetylated species are the only sialic acids present (not shown). Identification of these components by LC-MS/MS is shown in Figure 5. The eluting peaks were identified by MW and characteristic MS/MS fragment pattern (data not shown) as well as by their elution order according to Klein *et al.* (1997), and the proposed structures are indicated in Figure 5. It is probable that the quinoxalinone part largely dictates the ionization properties of the derivatives (all $[M+H]^+$ ions), and thus the peak intensities may be used for estimating the molar ratio of the sialic acid species. The LC-MS/MS analysis showed clearly that 4-O-acetylation is absent in these glycoproteins. The major mono-O-acetylated sialic acid species was identified as Neu5,9Ac_2. Neu5,7Ac_2 and Neu5,8Ac_2 were present as well. The di-O-acetylated species expectedly comprise the three possible structures, carrying O-acetyl groups in 7-, 8-, and 9-OH of Neu5Ac. However, at this stage we are not able to identify the di-O-acetylated isomers in the chromatogram. Attempts to assign the MS/MS spectra of the peaks have so far been unsuccessful. Neither have we found studies in the literature where the relative elution order of di-O-acetylated sialic acid DMB-derivatives would have been reported.

The relatively high amount of non-O-acetylated Neu5Ac observed in LC-MS may indicate that O-glycans are not as highly O-acetylated as N-glycans. Indeed, the disialylated species of salarin O-glycopeptide Sr-GP1 carries no more than two O-acetyl groups, in contrast to the disialylated N-glycopeptide Sr-GP2, where up to four acetyl groups are detected.

Glycosylation of major N-glycoproteins in salmon skin

N-glycosylated proteins in the salmon skin extract were isolated by concanavalin A (ConA) affinity chromatography. The glycoprotein fraction was further separated by RP chromatography, yielding a major peak carrying two components of apparent MW 80 and 55 kDa in SDS–PAGE (not shown). These were subjected to in-gel digestion and identification by mass mapping. The 80-kDa protein was identified as transferrin. The 55-kDa protein (GP55) could not be identified by mass mapping, nor did N-terminal sequencing of five tryptic peptides reveal clear homology to any known protein.

An aliquot of the RP-fraction carrying transferrin and GP55 was subjected to N-glycan liberation, and the isolated N-glycans were analyzed by MALDI-TOF MS. The spectrum was very similar to that obtained from salmon kininogen and salarin N-glycans (not shown), indicating extensive sialic acid O-acetylation also in the N-glycans of these major glycoproteins of salmon skin.

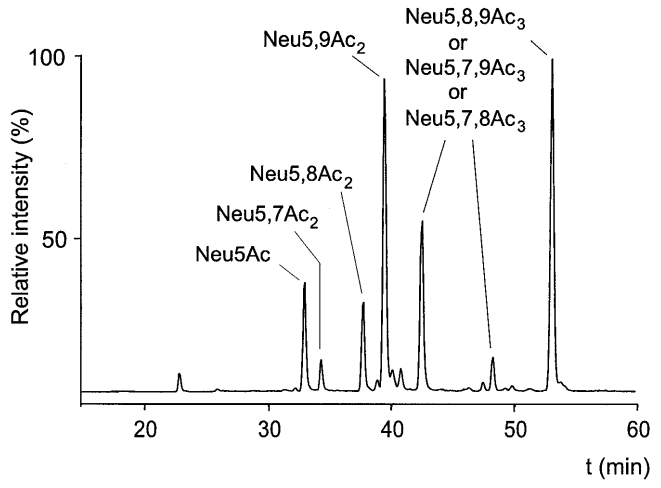

Fig. 5. Characterization of sialic acid species by LC-MS/MS. Sialic acids were liberated from proteins by acid hydrolysis and derivatized with DMB. DMB–sialic acid derivatives were separated on microbore RP-HPLC and detected by Q-TOF mass spectrometer as described in the text.

Discussion

In the present study we describe glycan structures present in two cysteine proteinase inhibitors of Atlantic salmon skin. Both glycoproteins were found to carry biantennary α2,3-sialylated N-glycans, as well as O-glycosidic sialylated core type 1 tri- and tetrasaccharides. The most interesting feature observed was that the N-glycans from these proteins, as well as those from the major N-glycoproteins present in salmon skin, exhibited very highly O-acetylated sialic acids. In fact, O-acetylated sialic acids are practically the sole terminal units in the N-glycans, as insignificant amount of non-O-acetylated N-acetylneuraminic acid is observed even after prolonged exposure to pH 8 during the protein purification (Ylönen *et al.*, 1999). To our knowledge, these structures represent the most extensively O-acetylated N-glycans reported on any species.

Few detailed studies on salmonid N-glycosylation have been presented. Rainbow trout ovarian fluid contains a 54-kDa protein that carries very large tetraantennary N-glycans with I antigenic poly-N-acetyllactosaminoglycans, Lewis X, and α2,8-disialyl groups (Funakoshi *et al.*, 1997). Atlantic salmon antithrombin has been found by lectin blotting studies to carry complex-type, α2,3-sialylated N-glycans (Andersen *et al.*, 2000). No sialic acid O-acetylation was reported in these

studies. However, the presence of O-acetylated sialic acids in salmon antithrombin can not be excluded, as α2,3-sialylation was identified by *Maackia amurensis* lectin, which binds to α2,3-linked sialic acids irrespective of their side chain O-acetylation status (Varki, 1997).

Interestingly, we have recently found high levels of sialic acid O-acetylation in the skin glycoproteins of two other northern Atlantic fish species, spotted wolffish (*Anarhichas minor*) and cod (*Gadus morhua* L.) (Ylönen *et al*., unpublished results). Atlantic salmon as well as other salmonids are considered primitive members of the teleost fishes. The conservation of extensive O-acetylation patterns in more evolved species, like spotted wolffish, may indicate an important function for this sialic acid modification in these species. General protection against microorganisms is one possibility, as O-acetylation of sialic acids is known to confer partial or complete resistance against sialidases and also to mask attachment sites on cells used by several bacteria and viruses (Varki, 1992).

The liver is the major site of kininogen expression in many species (Müller-Esterl *et al.*, 1986). Moreover, salarin seems to be expressed mainly in the liver (Ylönen *et al*., unpublished results). Transferrin, which is mainly expressed in liver as well (Kvingedal *et al.*, 1993), was found to carry similar levels of O-acetylation, and thus it is probable that the very effective sialic acid O-acetylation machinery resides in the salmon liver. This agrees well with earlier studies on sialic acid O-acetylation of N-glycans in rats (Butor *et al.*, 1993). The O-acetyltransferases of salmon liver may however be distinct from those in rat liver, where mainly α2,6-sialylated N-glycans are O-acetylated (Butor *et al.*, 1993).

We noted that two O-acetyl groups at most were present in the disialylated O-glycopeptide (Sr-GP1) isolated from salarin. Substantial loss of O-acetyl groups is unavoidable during the trypsin digestion, but a series of acetylated species carrying up to four O-acetyl groups were still detected in the disialylated N-glycopeptides. The overall acetylation efficiency of O-glycans may thus be lower. Alternatively, it is possible that either one of the sialic acids in the disialylated O-glycan is not a target for the O-acetyltransferase(s). This question however could not be addressed in more detail by the methodology used in this study: There is no practical method to liberate O-glycans from proteins without destroying their O-acetyl substituents. Nuclear magnetic resonance (NMR) is capable of distinguishing substituted sialic acids in branched glycans (Damm *et al.*, 1989): Larger amounts of glycopeptides should therefore be prepared for NMR analysis to resolve whether there is preference for the sialic acid O-acetylation in O-glycans.

Materials and methods

Materials

Salmon kininogen and salarin were purified from the skin of Atlantic salmon as described elsewhere (Ylönen *et al.*, 1999).

Preparation of N-glycans

N-glycans from salmon kininogen and salarin samples were released by enzymatic hydrolysis with PNGase F, and purified as described (Saarinen *et al.*, 1999). The oligosaccharides were recovered as a pool, and no separation of the individual components was attempted prior to MS analyses. The amount of N-glycans was estimated by their UV absorbance in gel filtration chromatography against external N-acetylglucosamine and N-acetylneuraminic acid standards.

The major glycoproteins of the salmon skin extract were isolated by affinity chromatography on a 5 × 45 mm column of ConA sepharose (Amersham Pharmacia Biotech AB, Uppsala, Sweden). Samples of the delipidated skin extract were first run on a Bio-Gel P-6 DG column (10 × 100 mm) in 20 mM Tris–HCl, 250 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$ (ConA buffer) to remove saccharose present in the extract, and then aliquots of 1–3 ml were applied to the ConA column. Nonbinding proteins were removed by washing with five column volumes of the ConA buffer, and the N-glycosylated proteins were then eluted with 5 ml of 0.5 M α-methylmannoside in ConA buffer. Pooled glycoprotein fractions were concentrated with pressure microfiltration (Omegacell 10K, Pall Filtron, Northborough, MA) and separated on a C1 RP chromatography column, yielding a major peak carrying two glycoproteins of 80 kDa and 55 kDa, as evident on SDS–PAGE (not shown). An aliquot of the RP-chromatography fraction (approximately 40 µg total protein) was vacuum-dried and then subjected to PNGase F treatment without attempts to separate the two components.

Preparation of O-glycans

To liberate O-glycosidically linked glycans, vacuum-dried protein samples obtained from RP-chromatography were dissolved in 100 µl of 1 M $NaBH_4$ in 0.1 M NaOH, and incubated for 48 h at 37°C. The liberated oligosaccharide alditols were recovered as described (Nyman *et al.*, 1998).

Permethylation

Vacuum-dried samples were permethylated according to the method of Ciucanu and Kerek (1984). Prior to ESI-MS analyses, the permethylated glycans were purified by adsorption to POROS R3 material (Perseptive Biosystems, Framingham, MA) as follows (Saarinen *et al.*, 1999): about 1 µl of POROS R3 material was packed into an Eppendorf gel-loader tip and washed with water, then the permethylated glycans were applied to the tip in 20% aqueous methanol. The column was washed with 20 µl of 10% aqueous methanol and eluted with 5 µl of 50% aqueous methanol.

Removal of O-acetyl groups

To remove O-acetyl groups from the isolated N-glycans, dry glycan samples were dissolved in 0.1 M NaOH and incubated at 8°C for 80 min. Reaction was terminated by adding one-half volume of 1 M acetic acid. An aliquot of the reaction mixture was drop-dialyzed for 30 min against water on VSWP 02500 membranes (Millipore, Bedford, MA) prior to MALDI-TOF MS analysis.

Periodate oxidation

Saponified N-glycans were incubated in 65 µl of 5 mM $NaIO_4$ on ice in the dark for 15 min, and the excess of periodate was destroyed by addition of 10 µl of 0.3 M ethylene glycol (Van Lenten and Ashwell, 1971). After 30 min at room temperature, 50 µl of 1 M $NaBH_4$ in 0.2 M Na_2CO_3 was added and the mixture was incubated at room temperature for 4 h. The reaction was terminated by adding 2 M acetic acid, and prior to MALDI-TOF MS analysis the N-glycans were desalted by gel filtration.

Gas-phase β-elimination

Glycopeptide sample was dried in a glass tube, and the tube was placed in a larger glass tube containing 300 µl of 40% aqueous methylamine (Mirgorodskaya and Roepstorff, 1998). The tube was evacuated until cool to the touch, and the sample was allowed to react with methylamine gas for 2.5 h at 55°C. The deglycosylated peptide was dissolved in 50% aqueous acetonitrile, and an aliquot was analyzed by MALDI-TOF MS.

Glycosidase digestions

Liberation of N-glycans from N-glycopeptides was performed by enzymatic hydrolysis with PNGase F: Glycopeptide samples were dissolved in 30 µl of 20 mM sodium phosphate, pH 7.2, containing 0.2% β-octylglucoside, and 1.5 U of PNGase F was added. After 16 h incubation at 37°C, a sample of the mixture was purified on a 1 µl column of POROS R3 as above, and the deglycosylated peptide was analyzed by MALDI-TOF MS.

Digestions with NDV neuraminidase and *S. pneumoniae* β1,4-galactosidase (Oxford Glycosciences, Abingdon, UK) were carried out in 10 µl of 50 mM sodium acetate, pH 5.5, at 37°C; aliquots of 1 µl were removed after 16 h of digestion, drop-dialyzed against water, and analyzed by MALDI-TOF MS. Pure oligosaccharide standards were used as controls for the glycosidase reactions: Neu5Acα2,3Galβ1,4(Fucα1,3)GlcNAcβOMe for NDV neuraminidase, and Galβ1,4GlcNAcβ1,6[Galβ1,3GlcNAcβ1,3]Galβ1,4Glc for β1,4-galactosidase.

Trypsin digestion

Protein samples (20–40 µg) were alkylated with 4-vinylpyridine and desalted by RP-HPLC as described elsewhere (Ylönen *et al.*, 1999). The vacuum-dried samples were dissolved in 30 µl of 0.1 M ammonium bicarbonate, and 0.5 µg of modified trypsin (Promega, Madison, WI) was added. Digestion was performed overnight at 37°C. The tryptic peptides were separated on a narrow-bore RP chromatography as described (Ylönen *et al.*, 1999), and samples from the fractions were analyzed by MALDI-TOF MS.

MS

MALDI-TOF MS was performed on a Biflex™ time-of-flight instrument (Bruker Daltonik, Bremen, Germany) equipped with a nitrogen laser operating at 337 nm.

Underivatized N-glycans were analyzed in either positive ion delayed extraction reflector mode or negative ion delayed extraction linear mode, using 2,4,6-trihydroxyacetophenone (Fluka Chemie AG, Buchs, Switzerland) (3 mg/ml in acetonitrile/20 mM aqueous diammonium citrate, 1:1, by volume) as the matrix as described (Saarinen *et al.*, 1999). Permethylated samples were analyzed in the positive ion delayed extraction reflector mode using 10 mg/ml DHB (Aldrich, Steinheim, Germany) in H_2O as the matrix.

Peptide samples were analyzed using α-cyano-4-hydroxycinnamic acid matrix, as described elsewhere (Saarinen *et al.*, 1999). Mass values for glycopeptides are average; for deglycosylated peptides monoisotopic values were obtained.

ESI mass spectra were collected using a Q-TOF hybrid quadrupole-orthogonal acceleration time-of-flight mass spectrometer (Micromass Ltd., Manchester, UK). Permethylated glycans were dissolved in 50% aqueous methanol containing either 0.5 mM NaOH (O-glycans) or 5 mM ammonium acetate (N-glycans) and injected into the mass spectrometer with a nanoelectrospray ion source at a flow rate of about 30 nl/min. For MS/MS studies the parent ions were selectively transmitted and directed into the collision cell containing argon collision gas. Collision energies were about 60 V for the sodiated singly charged permethylated O-glycans, and about 35 V for the ammoniated doubly charged permethylated N-glycans.

Characterization of O-acetylated sialic acids by LC-MS/MS

Sialic acids were liberated from salmon kininogen and salarin by chemical hydrolysis in 2 M propionic acid (Mawhinney and Chance, 1994). The released sialic acids were purified from proteins by running the hydrolysate through a POROS R1 microcolumn as above, and the eluate (25 µl) was mixed with 25 µl of DMB reagent to produce the corresponding sialic acid quinoxalinone derivatives (Hara *et al.*, 1989; Klein *et al.*, 1997). DMB-derivatized sialic acids were separated by microbore reversed-phase HPLC on an 0.075 × 150 mm PepMap column (NAN75-15-03-C18-PM; LC Packings, Amsterdam) by elution with a linear gradient of acetonitrile (4–40% in 30 min) in 0.1% formic acid. Chromatography was performed at a flow rate of 0.25 µl/min and the eluent was directly injected to Q-TOF mass spectrometer via a distally coated silica capillary needle (20 µm inner diameter, 10 µm tip opening; New Objective Inc., Cambridge, MA). The mass spectrometer was set to scan the HPLC eluent using the following settings: The MS was set to operate in constant CID mode for the entire HPLC run, with the MS set to acquire CID spectra on ions 426.2, 468.2, and 510.2 Da, corresponding to $[M+H]^+$ of non-O-acetylated, mono-O-acetylated, and di-O-acetylated DMB-N-acetylneuraminic acid, respectively. Each spectra were acquired for 1.5 s/parent ion, ramping between the selected ions for the entire HPLC run. The collision energy for all ions was 18 V.

Protein identification by mass mapping

Protein samples were run on SDS–PAGE; bands of interest were cut out of the gel and subsequently hydrolyzed in-gel with trypsin. An aliquot of the peptide mixture obtained was desalted on a POROS R3 microcolumn and analyzed by MALDI-TOF MS as described (Nyman *et al.*, 2000).

Peptide sequencing by Edman degradation

Peptides were sequenced using a Procise 494A sequencer (Perkin-Elmer Applied Biosystems Division, Foster City, CA).

Acknowledgments

The authors wish to thank Ms. Anu Jaakola, Mrs. Gunilla Rönnholm, and Mr. Keijo Virta for excellent technical assistance. This study was partly supported by grants from the Academy of Finland (no. 41413 and no. 46692 to J.H.). This study was also partly supported by the Commission of the European Communities, Agriculture and Fisheries specific RTD program CT97-3508, "Fish cysteine proteinase inhibitors and infectious diseases"; it does not necessarily reflect the commision's views and in no way anticipates its future policy in this area.

Abbreviations

Ac, acetyl; ConA, concanavalin A; DMB, 1,2-diamino-4,5-methylenedioxybenzene; ESI, electrospray ionization; HPLC, high-performance liquid chromatography; LC, liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization-time of flight; MS, mass spectrometry; MS/MS, tandem mass spectrometry; MW, molecular weight; NDV, Newcastle disease virus; Neu5Ac, N-acetylneuraminic acid; Neu5Gc, N-glycolylneuraminic acid; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; PNGase F, peptide:N-glycosidase F; RP, reversed-phase; SK-GP, glycopeptide from salmon kininogen; Sr-GP, glycopeptide from salarin.

References

Alviano, C.S., Travassos, L.R., and Schauer, R. (1999) Sialic acids in fungi: a minireview. *Glycoconj. J.*, **16**, 545–554.

Andersen, Ø., Flengsrud, R., Norberg, K., and Salte, R. (2000) Salmon antithrombin has only three carbohydrate side chains, and shows functional similarities to human β-antithrombin. *Eur. J. Biochem.*, **267**, 1651–1657.

Butor, C., Diaz, S., and Varki, A. (1993) High level O-acetylation of sialic acids on N-linked oligosaccharides of rat liver membranes. Differential subcellular distribution of 7- and 9-O-acetyl groups and of enzymes involved in their regulation. *J. Biol. Chem.*, **268**, 10197–10206.

Ciucanu, I., and Kerek, F. (1984) A simple rapid method for the permethylation of carbohydrates. *Carbohydr. Res.*, **131**, 209–217.

Corfield, A.P., Sander-Wewer, M., Veh, R.W., Wember, M., and Schauer, R. (1986) The action of sialidases on substrates containing O-acetylsialic acids. *Biol. Chem. Hoppe-Seyler*, **367**, 433–439.

Damm, J.B.L., Voshol, H., Hård, K., Kamerling, J.P., and Vliegenthart, J.F.G. (1989) Analysis of N-acetyl-4-O-acetylneuraminic-acid-containing N-linked carbohydrate chains released by peptide-N^4-(N-acetyl-β-glucosaminyl)asparagine amidase F. Application to the structure determination of the carbohydrate chains of equine fibrinogen. *Eur. J. Biochem.*, **180**, 101–110.

Domon, B., and Costello, C.E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.*, **5**, 397–409.

Egge, H., and Peter-Katalinic, J. (1987) Fast atom bombardment mass spectrometry for structural elucidation of glycoconjugates. *Mass Spectrom. Rev.*, **6**, 331–393.

Funakoshi, Y., Taguchi, T., Sato, C., Kitajima, K., Inoue, S., Morris, H.R., Dell, A., and Inoue, Y. (1997) Occurence of terminal α2→8-linked disialylated poly-N-acetyllactosamine chains with Le^X and I antigenic glycotopes in tetraantennary arms of an N-linked glycoprotein isolated from rainbow trout ovarian fluid. *Glycobiology*, **7**, 195–205.

Hara, S., Yamaguchi, M., Takemori, Y., Furuhata, K., Ogura, H., and Nakamura, M. (1989) Determination of mono-O-acetylated N-acetylneuraminic acids in rat sera by fluorometric high-performance liquid chromatography. *Anal. Biochem.*, **179**, 162–166.

Hounsell, E.F., Davies, M.J., and Renouf, D.V. (1996) O-linked protein glycosylation structure and function. *Glycoconj. J.*, **13**, 19–26.

Inoue, S., and Iwasaki, M. (1978) Isolation of a novel glycoprotein from the eggs of rainbow trout: occurence of disialosyl groups on all carbohydrate chains. *Biochem. Biophys. Res. Commun.*, **83**, 1018–1023.

Iwasaki, M., and Inoue, S. (1985) Structures of the carbohydrate units of polysialoglycoproteins isolated from the eggs of four species of salmonid fishes. *Glycoconj. J.*, **2**, 209–228.

Iwasaki, M., Inoue, S., and Troy, F.A. (1990) A new sialic acid analoque, 9-O-acetyl-deaminated neuraminic acid, and α-2,8-linked O-acetylated poly(N-glycolylneuraminyl) chains in a novel polysialoglycoprotein from salmon eggs. *J. Biol. Chem.*, **265**, 2596–2602.

Juhasz, P., and Costello, C.E. (1992) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry of underivatized and permethylated gangliosides. *J. Am. Soc. Mass Spectrom.*, **3**, 785–796.

Klein, A., and Roussel, P. (1998) O-acetylation of sialic acids. *Biochimie*, **80**, 49–57.

Klein, A., Diaz, S., Ferreira, I., Lamblin, G., Roussel, P., and Manzi, A.E. (1997) New sialic acids from biological sources identified by a comprehensive and sensitive approach: liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology*, **7**, 421–432.

Kvingedal, A.M., Rorvik, K.A., and Alestrom, P. (1993) Cloning and characterization of Atlantic salmon (*Salmo salar*) serum transferrin cDNA. *Mol. Mar. Biol. Biotechnol.*, **2**, 233–238.

Mawhinney, T.P., and Chance, D.L. (1994) Hydrolysis of sialic acids and O-acetylated sialic acids with propionic acid. *Anal. Biochem.*, **223**, 164–167.

Metcalf, V.J., Brennan, S.O., Chambers, G.K., and George, P.M. (1998) The albumin of the brown trout (*Salmo trutta*) is a glycoprotein. *Biochim. Biophys. Acta*, **1386**, 90–96.

Mirgorodskaya, E., and Roepstorff, P. (1998) Mass spectrometric determination of O-glycosylation sites using beta-elimination and partial acid hydrolysis. Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, 215.

Müller-Esterl, W., Iwanaga, S., and Nakanishi, S. (1986) Kininogens revisited. *Trends Biochem. Sci.*, **11**, 336–339.

Nyman, T.A., Kalkkinen, N., Tölö, H., and Helin, J. (1998) Structural characterization of N-linked and O-linked oligosaccharides derived from interferon-α2b and interferon-α14c produced by Sendai-virus-induced human peripheral blood leukocytes. *Eur. J. Biochem.*, **253**, 485–493.

Nyman, T.A., Matikainen, S., Sareneva, T., Julkunen, I., and Kalkkinen, N. (2000) Proteome analysis reveals ubiquitin-conjugating enzymes to be a new family of interferon-α-regulated genes. *Eur. J. Biochem.*, **267**, 4011–4019.

Papac, D.I., Wong, A., and Jones, A.J.S. (1996) Analysis of acidic oligosaccharides and glycopeptides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.*, **68**, 3215–3223.

Reinhold, V.N., Reinhold, B.B., and Costello, C.E. (1995) Carbohydrate molecular mass profiling, sequence, linkage & branching data: ES-MS and CID. *Anal. Chem.*, **67**, 1772–1784.

Saarinen, J., Welgus, H.G., Flizar, C.A., Kalkkinen, N., and Helin, J. (1999) N-glycan structures of matrix metalloproteinase-1 derived from human fibroblasts and from HT-1080 fibrosarcoma cells. *Eur. J. Biochem.*, **259**, 829–840.

Schauer, R. (1991) Biosynthesis and function of N- and O-substituted sialic acids. *Glycobiology*, **1**, 449–452.

Vandamme-Feldhaus, V., and Schauer, R. (1998) Characterization of the enzymatic 7-O-acetylation of sialic acids and evidence for enzymatic O-acetyl migration from C-7 to C-9 in bovine submandibular gland. *J. Biochem.*, **124**, 111–121.

Van Lenten, L., and Ashwell, G. (1971) Studies on the chemical and enzymatic modification of glycoproteins. A general method for the tritiation of sialic acid-containing glycoproteins. *J. Biol. Chem.*, **246**, 1889–1894.

Varki, A. (1992) Diversity in the sialic acids. *Glycobiology*, **2**, 25–40.

Varki, A. (1997) Sialic acids as ligands in recognition phenomena. *FASEB J.*, **11**, 248–255.

Varki, A., and Diaz, S. (1984) The release and purification of sialic acids from glycoconjugates: Methods to minimize the loss and migration of O-acetyl groups. *Anal. Biochem.*, **137**, 236–247.

Ylönen, A., Rinne, A., Herttuainen, J., Bøgwald, J., Järvinen, M., and Kalkkinen, N. (1999) Atlantic salmon (*Salmo salar* L.) skin contains a novel kininogen and another cysteine proteinase inhibitor. *Eur. J. Biochem.*, **266**, 1066–1072.