GENOVIS PRESENTS

A Novel O-glycoprotease with Applications in O-glycan Analysis using Mass Spectrometry

AUTHORS

Rolf Lood, Maria Nordgren, Fredrik Leo, Stephan Björk, Malin Mejàre, Helén Nyhlén, Fredrik Olsson | Genovis AB, Lund, Sweden

INTRODUCTION

The study of O-linked glycosylation within the field of glycomics is known to be a challenging task due to complicated sample preparations and analytical procedures. The field has suffered from lack of O-glycan-specific enzymes and efficient enrichment methods. A novel O-glycan-specific endoprotease was discovered in *Akkermansia muciniphila* and expressed in *E. coli*. This O-glycoprotease (OgpA) digests the peptide backbone N-terminally of serine or threonine residues carrying an O-glycan. A catalytically inactive form of OgpA with retained O-glycan binding was also developed. This inactive OgpA was evaluated for selective affinity purification of O-glycoproteins from various protein samples. Characterization of O-glycoproteins utilizing both OgpA and the inactive OgpA are demonstrated on both simple and complex protein mixtures.

SUMMARY

- Aim: To demonstrate applications of the OgpA and the inactive OgpA in O-glycan analysis using MS
- OgpA is a new, specific endo-O-glycoprotease
- Immobilized and inactive OgpA selectively binds O-glycosylated proteins and peptides at native conditions
- Selective separation of O-glycan peptides from non- and only N-glycosylated peptides
- Selective enrichment of O-glycoproteins was achieved from human serum
- Site-specific O-glycan characterization of peptides from human protein was achieved after enrichment on the O-glycan binding resin
- The OgpA and the O-glycan-binding resin are useful tools for O-glycan analysis using MS

RESULTS

New O-glycan-specific Protease

A new endo-O-protease was discovered in *Akkermansia muciniphila*, a commensal bacterium in the human microbiota. The bacteria colonize distal ileum to rectum and degrade and metabolize highly O-glycosylated mucin. The O-protease (OgpA) was recombinantly expressed in *E. coli* and purified to homogeneity (OpeRATOR™, Genovis). The activity and specificity of the protease was evaluated by digestion of several protein substrates at native conditions. The O-protease displayed activity on an asialylated protein, but the activity was markedly increased on O-glycoproteins pretreated with sialidases (*A. muciniphila*). It was found that the protease only digests proteins decorated with O-glycans. No activity was observed on non-glycosylated proteins nor on proteins with only N-glycan structures (*Fig. 1*).

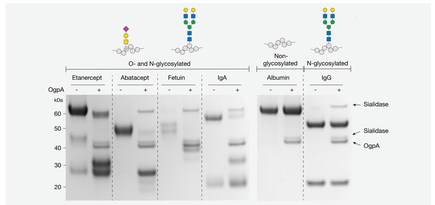


Figure 1. Specificity for O-glycosylated proteins. The OgpA digests only O-glycosylated proteins. No activity was observed on a non-glycosylated protein, or on a protein with only N-glycans. The proteins were pre-treated with sialidases and incubated O/N with the OgpA. The samples were analyzed by reduced SDS-PAGE.

To further verify the specificity of the O-protease, the O-glycosylated proteins erythropoietin (EPO) and human chorionic gonadotropin beta chain (hCGβ) were used as substrates and digested by OgpA at native conditions. EPO is a protein with a single core 1 O-glycan and a specific digestion site N-terminally of the O-glycosylated serine[126] was defined (*Fig. 2*).

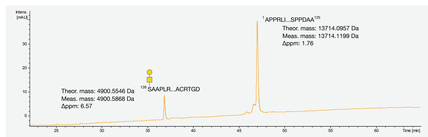


Figure 2. Specific digestion N-terminally of the O-glycosylation site. N-glycans were removed by PNGaseF and sialic acids were removed by a sialidase mix (*A. muciniphila*) in parallel to the digestion with OgpA. The EPO protein carrying one core 1 O-glycan was hydrolyzed at a single specific site N-terminally of the O-glycosylated Serine[126]. After digestion the fragments were reduced by DTT and separated on a RP BEH C4 column (Waters) followed by ESI-Q-TOF MS detection.

hCGβ is a glycoprotein with four O- and two N-glycans, with a constant core 2 at Serine[121,1]. Digestion of this protein with OgpA was done at native conditions with and without sialidase, and the RP-LC-MS/MS result confirms digestion also at core 2 (*Fig. 3*).

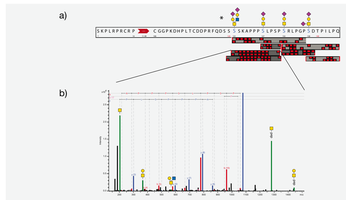


Figure 3. Digestion at core 1 and core 2. OgpA digestion of human chorionic gonadotropin beta chain (hCGβ). **a)** The generated glycopeptides from hCGβ confirm glycosylation at S121, S127, S132, S138, and indicate glycosylation at S120 and T140. RP separation was performed on an AdvanceBio Peptide Map C18 column from Agilent and samples were desalted in-line prior to ESI Q-TOF on a Bruker Impact II MS. *Glycan structures according to Valmu et al. (2006)[1]*. **b)** MS/MS spectra of fragment S121-P131 confirms digestion at the core 2 structure by the presence of HexNAc/HexNAc/Hex as an oxonium ion.

Applications of the O-protease in LC-MS

The unique specificity for O-glycosylated residues on glycoproteins opens for a variety of applications using this new O-protease (OgpA) in the glycoproteomic field. The OgpA was used to map glycosylation sites of the biopharmaceutical etanercept, an Fc-fusion protein with a highly core 1 O-glycosylated region. By using the workflow illustrated in *Fig. 4*, peptides with intact O-glycans, glycopeptides without sialic acids and peptides lacking sugar residues were generated. The intact mass fragments and MS/MS peptides completed the amino acid sequence of etanercept and defined the O-glycosylated serine and threonine residues. A summary of the obtained data is presented in *Fig. 5.*

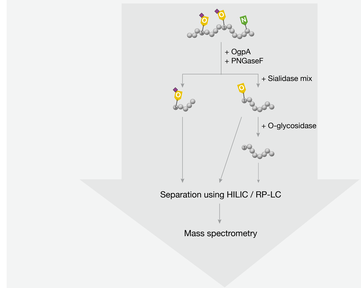
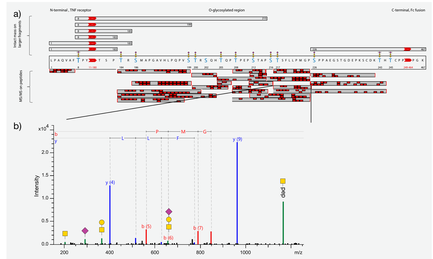


Figure 4. Workflow from intact glycoprotein to LC-MS and LC-MS/MS. In a native one pot O/N digestion reaction, the glycoprotein is hydrolyzed N-terminally of O-glycosylated sites using the OgpA. Addition of Sialidase mix is optional but improves digestion. In a second step, the remaining GalNAc/Gal can be removed from the asialylated glycopeptides using an O-glycosidase (*Streptococcus Oralis*). After reduction, the samples were analyzed on RP or HILIC-LC ESI-Q-TOF MS and MS/MS. PNGaseF is added in the O/N digestion step to reduce heterogeneity.



Range	Sequence	O- Composition	Measured Mass	Expected Mass	Δppm
1-7	-.LPAQVAFT	-	744.412	744.417	-7.39
1-185	-.LPA...PVS.R	-	19842.873	19842.895	-1.07
184 - 198	P.TRSMAPGAVHLPQPV.S	-	1559.813	1559.824	-7.12
186 - 199	R.SMAPGAVHLPQPVS.T	-	1405.699	1405.702	-2.13
186 - 204	R.SMAPGAVHLPQPVSTRSQHLT	2 x	2729.256	2729.27	-5.06
200 - 207	L.TRSQHTGP.T	-	953.47	953.468	+1.99
205 - 207	H.TQP.T	-	344.172	344.17	+5.81
208 - 216	P.TPEPSTAPS.T	3 x	1980.796	1980.805	+4.24
212 - 216	P.STAPS.T	-	461.213	461.212	+1.95
217 - 225	S.TSFLLPMGPS.S		1618.7012	1617.693	+17.62
217 - 244	S.TSFLLPMGPSPPAEGSTGDEPKSCDKTHT	-	2885.309	2885.321	-4.16
226 - 242	P.SPPAEGSTGDEPKSCDK.T	-	1703.726	1703.731	-2.93
226-467	P.SPP..PGK-		27210.356	27210.461	-3.87

Figure 5. OgpA digestion in LC-MS analysis of etanercept. a) Both larger C- and N-terminal fragments as well as smaller O-glycosylated fragments were generated using the OgpA. The heterogeneity in O-glycosylation led to overlapping peptides. **b)** Selected mass data of peptides covering the entire amino acid sequence. **c)** Typical MS/MS spectrum with the HexNAc still attached to the peptide (in ESI-Q-TOF-MS, the glycan is often lost from the peptide before fragmentation and found as oxonium ions). RP separation was performed on an AdvanceBio Peptide Map C18 column from Agilent and samples were desalted in-line prior to ESI Q-TOF on a Bruker Impact II MS. HILIC separation was performed on a Waters Acquity BEH Amide column after buffer exchange on Graphite carbon.

Inactive O-protease

The OgpA protein was engineered for high affinity binding to O-glycosylated proteins and peptides without hydrolyzing the peptide bonds at O-glycosylated serines and threonines.

The inactive OgpA was recombinantly expressed in *E. coli*, purified to homogeneity and immobilized on agarose beads to create an affinity resin for O-glycoproteins and -peptides (GlycOCATCH™, Genovis). The specificity and selectivity of the resin were tested by incubation of protein solutions with the resin in presence of the sialidase mix, with end-over-end mixing in microspin columns. After washing the columns, the bound proteins were eluted with 8 M urea. The resin specifically bound proteins carrying O-glycans, separately but also in a mix of proteins. Negligible binding of non- or N-glycosylated proteins were observed (*Fig. 6*). To study the O-glycan selectivity of the resin at the peptide level, a selection of non-glycosylated peptides and an O-glycosylated peptide were used, and the eluted fraction contained solely the core 1 O-glycosylated peptide (*Fig. 7*).

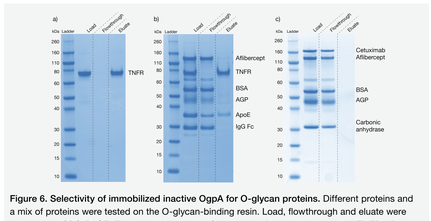


Figure 6. Selectivity of immobilized inactive OgpA for O-glycan proteins. Different proteins and a mix of proteins were tested on the O-glycan-binding resin. Load, flowthrough and eluate were analyzed on SDS-PAGE. The resin with inactive OgpA binds **a)** heavily glycosylated TNFR, **b)** specifically O-glycosylated proteins TNFR and Apolipoprotein E (ApoE) in presence of N- and non- glycosylated proteins. **c)** No binding to proteins lacking O-glycans. TNFR: TNF alpha binding domain, BSA: bovine serum albumin, AGP: alpha-1 acid glycoprotein.

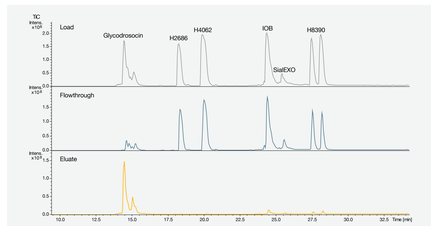


Figure 7. Selective purification of an O-glycosylated peptide. A mix of five synthesized peptides were tested on the O-glycan-binding resin. The data show selective purification of the O-glycan containing peptide. The fractions were analyzed by LC-MS. RP separation was performed on an AdvanceBio Peptide Map C18 column (Agilent) and analyzed by ESI Q-TOF on a Bruker Impact II MS.

Applications of OgpA and O-glycan-binding Resin in LC-MS

The O-glycan affinity columns were used for specific enrichment of O-glycosylated peptides and proteins prior to MS analysis. Using a workflow for tryptic peptides from human IgA, peptides with various degree of glycosylation were specifically enriched on the O-glycan-binding resin and detected in the eluate (*Fig. 8*). The strong binding ability of inactive OgpA to O-glycan structures was also effectively utilized to enrich O-glycosylated proteins from the complex sample mixture of human serum (*Fig. 9*).

Further digestion of the enriched proteins with a combination of trypsin and active OgpA generated peptides with N-terminally glycosylated serines and threonines, which enabled site-specific O-glycan characterization (*Fig. 10*).

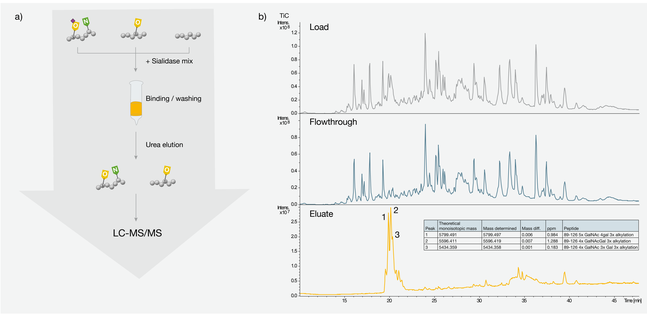


Figure 8. Specific enrichment of O-glycosylated peptides from IgA. a) IgA was trypsin digested, reduced, denatured and alkylated. Trypsin was inhibited, and the sample was incubated on the inactive OgpA resin whilst being desialylated. The resin was washed, and the bound peptides were eluted with 8 M urea. Separation was performed on a RP-LC C18 column (Advance BioPeptide Map 2.1x100 2.7µm from Agilent) and detection on an ESI-Q-TOF Bruker Impact II mass spectrometer. **b)** The generated data confirm that the O-glycosylated tryptic hinge peptide of human IgA was straightforwardly enriched on the resin and eluted free from N-glycan carrying peptides.

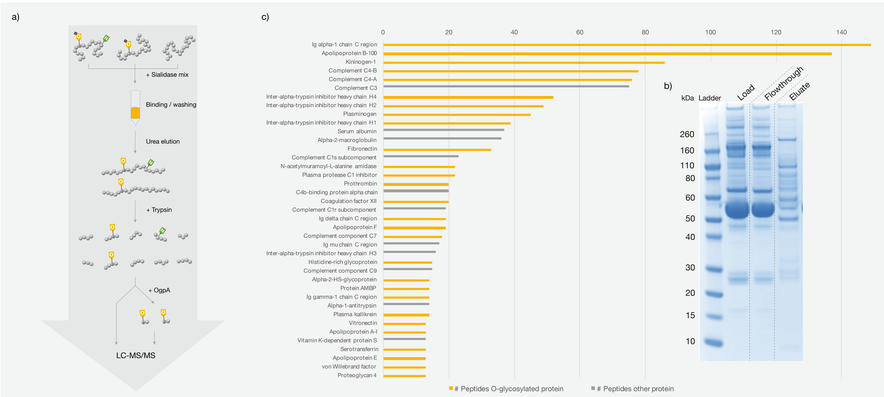
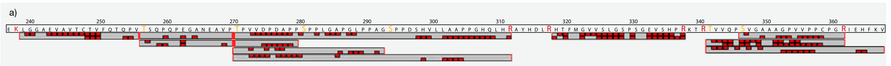


Figure 9. Enrichment of O-glycoproteins from human serum. a) Workflow for O-glycan characterization in protein samples. **b)** Human serum (Sigma) and sialidase were incubated on the inactive OgpA resin. Load, flow-through and eluted fractions were visualised on SDS-PAGE. **c)** The eluted proteins were denatured, reduced, alkylated and trypsin digested. Peptides were analyzed using RP-LC-MS/MS on a C18 column (Advance BioPeptide Plus 2.1x150mm 2.7 µm from Agilent Technologies) in a 0.1%FA in MQ: 0.1% FA in 95% ACN gradient at 45°C and a flow of 0.2 ml/min. Detection was on an ESI-Q-TOF Bruker Impact II instrument. Data was converted to mgf format files in Data Analysis Software v 4.4 and searched against the Swiss Prot database. Core-1/ core-2 structures, methionine oxidation and N-terminal acetylation was set as variable modifications and carbamidomethylation as fixed. Identified O-glycosylated proteins and other proteins with >12 matching peptides and a MASCOT score >200 are listed.



Range	Sequence	Measured Mass	z	Expected mass	Mass Δppm	Glycosylation site	Indicated Glycosylation site	Glycoform
238 - 255	K.LGGAEAVVTCTVFQTQPV.T	626.3150	3	1875.9400	- 9.08	-	T 256	-
256 - 269	V.TSQPQPEGANEAVP.T	895.3940	2	1788.7901	- 9.35	T 256	T 270	1 x HexNAcHex
256 - 269	V.TSQPQPEGANEAVP.T	597.2683	3	1788.7901	- 3.93	T 256	T 270	1 x HexNAcHex
270 - 279	P.TPVVDPDAPP.S	686.8192	2	1371.6293	- 3.97	T 270	S 280*	1 x HexNAcHex
270 - 292	P.TPVVDPDAPPSPPLGAPGLPPAG.S	828.7474	3	2483.2319	- 4.64	T 270	S 293*	1 x HexNAcHex
270 - 311	P.TPVVDPDAPPSPPLGAPGLPPAGSPPDSHVLLAAPPGHQLHR.A	972.6827	5	4858.4189	- 8.59	T 270	S 280* or S 293*	2 x HexNAcHex
270 - 311	P.TPVVDPDAPPSPPLGAPGLPPAGSPPDSHVLLAAPPGHQLHR.A	899.6561	5	4493.2867	- 9.47	T 270	-	1 x HexNAcHex
318 - 337	R.HTFMGVVSLGSPSGEVSHPR.K	694.3419	3	2080.0160	- 5.83	-	-	-
341 - 361	R.TVVQPSVGAAAGPVVPPCPGR.I	672.6885	3	2015.0622	- 9.22	T 341*	-	-
341 - 367	R.TVVQPSVGAAAGPVVPPCPGRIRHFKV.-	548.7080	5	2738.5166	- 4.74	T 341*	-	-
346 - 361	P.SVGAAAGPVVPPCPGR.I	619.6342	3	1855.8986	- 9.64	S 346	-	1 x HexNAcHex

*not previously reported, according to Uniprot entry P02765

Figure 10. O-glycan sites of Alpha-2-HS-Glycoprotein in serum. After enrichment of O-glycoproteins from serum as described above, the sample was in addition digested with both trypsin and active OgpA. Separation, detection and data handling was done according to Fig 9, with the addition of OgpA enzyme in protein search. The OgpA-specific digestion N-terminally of glycosylated serines and threonines provides additional information on the actual O-glycan sites. The figure and table present a selection of the peptides and glycopeptides attained to demonstrate the data this workflow can generate also from very complex sample mixtures. In this test, peptides 270-279, 270-292 and 270-311 carrying 2x HexNAcHex indicate that S280 and S293 is glycosylated.

REFERENCES

1. Valmu L, Alfthan H, Hotakainen K, Birken S, Stenman UH. Site-specific glycan analysis of human chorionic gonadotropin beta-subunit from malignancies and pregnancy by liquid chromatography–electrospray mass spectrometry. Glycobiology. 2006; 16:1207–18.