1 **Detailed characterization of the *O*-linked glycosylation of the neuropilin-1 c/MAM-domain**

2

3 Markus Windwarder[1], Tamas Yelland[2], Snezana Djordjevic[2], Friedrich Altmann[1]

5

6 Affiliations:
7 [1] Department of Chemistry, University of Natural Resources and Life Sciences, Vienna; Muthgasse 18, 1190
8 Vienna, Austria.
9 Tel.: +43 1 47654 6062; Fax: +43 1 47654 6059.
10 [2] Institute of Structural and Molecular Biology, University College London; Gower Street, Darwin Building,
11 London, WC1E 6BT, United Kingdom

12

13 Abbreviations

14 CID	Collision-induced dissociation
15 ETD	Electron-transfer dissociation
16 FPLC / HPLC	Fast protein / High pressure liquid chromatography
17 Gal	Galactose
18 Gal1	Galectin1
19 GlcNAc	*N*-acetylglucosamin
20 HEK293 cells	Human embryonic kidney cells
21 HexNAc-Hex	*N*-acetylhexosamine linked to a hexose
22 HUVEC	Human umbilical vein endothelial cell
23 Neu5Ac	*N*-acetylneuraminic acid
24 Nrp1	Neuropilin-1
25 (*O*-)GalNAc	(*O*-linked) *N*-acetylgalactosamine
26 (RP)-LC-ESI-MS	(reversed phase)-liquid chromatography-electrospray ionization mass spectrometry
27 PGC	porous graphitic carbon
28 Q-TOF	Quadrupole - Time of flight
29 SPE	Solid phase extraction
30 VEGF	Vascular endothelial growth factor
31 VEGFR2	Vascular endothelial growth factor receptor 2

32

Abstract

Neuropilins are involved in angiogenesis and neuronal development. The membrane proximal domain of neuropilin-1, called **c** or MAM domain based on its sequence conservation, has been implicated in neuropilin oligomerization required for its function. The c/MAM domain of human neuropilin-1 has been recombinantly expressed to allow for investigation of its propensity to engage in molecular interactions with other protein or carbohydrate components on a cell surface. We found that the c/MAM domain was heavily *O*-glycosylated with up to 24 monosaccharide units in the form of disialylated core 1 and core 2 *O*-glycans. Attachment sites were identified on the chymotryptic c/MAM peptide ETGATEKPTVIDSTIQSEFPTY by electron-transfer dissociation mass spectrometry (ETD-MS/MS). For highly glycosylated species consisting of carbohydrate to about 50%, useful results could only be obtained upon partial desialylation. ETD-MS/MS revealed a hierarchical order of the initial *O*-GalNAc addition to the four different glycosylation sites. These findings enable future functional studies about the contribution of the described glycosylations in neuropilin-1 oligomerization and the binding to partner proteins as VEGF or galectin-1.

As a spin-off result the sialidase from *Clostridium perfringens* turned out to discriminate between galactose and *N*-acetylgalactosamine linked sialic acid.

Introduction

Neuropilins (Nrps) are a family of transmembrane proteins involved in a number of signaling pathways. Of these pathways the two best characterized are angiogenesis in combination with the vascular endothelial growth factor (VEGF) receptor (VEGFR) and neuronal development in combination with plexins. Nrp's ability to bind to multiple structurally unrelated ligands has been attributed to its multidomain architecture with a large ectodomain containing four ligand-binding domains (a1a2b1b2) and a membrain proximal c/MAM domain purportedly involved in Nrp oligomerization [1].

Nrp's functionality can also be altered through a post-translational modification on the highly conserved Nrp1 S612, which can be *O*-glycosylated with either heparan sulphate or chondroitin sulphate [2, 3]. Using a S612A mutation a number of functional properties have been attributed to this glycan. With the help of I-125 radiolabelled VEGF Shintani [3] and coworkers showed that the glycosylation increases Nrp affinity for VEGF as well as alters receptor internalization and VEGFR expression levels in smooth muscle cells. Frankel et al [2] investigated S612 glycosylation effects in highly invasive U87MG human glioma cells and suggested that S612 is modified by chondroitin sulphate exclusively and that loss of this modification is associated with increased invasion properties of the cells. Furthermore, they ascribed this behavior to Nrp1-mediated signaling through p130Cas pathway. S612 is not conserved in Nrp2 thus conferring glycosylation-linked and isoform-specific properties to Nrp1 only. Additional *N*-linked glycosylation sites have been identified on N150, N261 and N522 [4, 5] with two more sites predicted based on sequence analysis (N300 and N842, Uniprot NRP1_HUMAN), although no functionality has yet been attributed to these putative modifications.

Other roles for glycans associated with neuropilin-1, in both angiogenesis and neuronal development, have also been postulated. It has been shown that galectin-1 (Gal1), through its carbohydrate-recognition domain, binds Nrp1, although the specific carbohydrate and Nrp1 glycosylation site is undetermined [6]. Gal1 is a homo-dimeric lectin with a conserved carbohydrate-binding site capable of binding the β-galactoside units of *N*- and *O*-linked sugars [7, 8]. The interaction between Nrp1 and Gal1 increased the migration of HUVEC cells in a VEGFR-2-dependent fashion [6]. In addition, it has been shown that binding of Gal1 to Nrp1 increases the phosphorylation levels of VEGFR-2, thus indirectly implicating formation of a complex multiprotein assembly on the cell surface involving Gal1, glycan-modified Nrp1, VEGF and VEGFR-2. Recently, Quinta et al. [9] has shown that Gal1 can play an important role in neuronal regeneration. Gal1, through interactions with Nrp1 prevents semaphorin3A from binding to the Nrp1:plexin complex required to induce neuronal growth cone collapse. Thus, Gal1:Nrp1 interactions have a wide array of cellular effects on different cell types although the molecular basis of this interaction remains undetermined.

The only well-characterized Nrp1 *O*-glycosylation site, so far, is S612, a site of covalent attachment for chondroitin / heparan sulphate. Very recently, 5 *O*-GalNAc glycosylation sites were reported for Nrp1 but without specification of the glycan structures [10]. A stretch of clustered *O*-glycans (4x in a range of 628-646) in close proximity to the known S612 glycosaminoglycan attachment site might affect the (patho)physiological properties of this part of Nrp1 (c/MAM domain). Changes in glycan structure and the type of the specific glycoform of the glycoproteins have been observed in cancer [8, 11-13]. Glycoforms may differ in their affinity to ligands as exemplified by non-fucosylated IgG [14]. While Nrp1 is known to bind Gal1, the exact site of lectin binding has not yet been defined and thus knowledge about sites of modifications as well as structures of bound carbohydrates constitutes an essential basis for further functional studies. Tandem mass spectrometry with

93 electron-transfer dissociation (ETD-MS/MS) on an ion trap instrument has recently been applied to the large and
94 complex multisite *O*-glycopeptide of bovine fetuin [15] and was therefore considered as an appropriate method
95 for the analysis of c/MAM.
96 Here we present a detailed characterization of the *O*-linked glycosylation of the c/MAM-domain of Nrp1 by
97 mass spectrometry. The sites of glycosylation (ETD-MS/MS) as well as the chemical structures of the attached
98 carbohydrates (PGC-LC-MS/MS) were determined revealing a hierarchical order of GalNAc-addition to the
99 different modification sites.

100
101

102 **Material and Methods**

103

104 **cDNA cloning, gene expression and protein purification**
105 A portion of the human Nrp1 cDNA, corresponding to protein residues 628-813 was amplified using the
106 following primers: 5'-GCGTAGCTGAAACCGGCGCCACAGAAAAGCCCACGGTC-3', and 5'-
107 GTGGTGGTGGTGTTTTGCTGGTTTTGCACAATCTTCTTGTG-3'. By using the In-Fusion (Clontech Saint-
108 German-en-Laye, France) ligation method, the PCR products were inserted into the pOPINTTGNeo vector
109 (Oxford Protein Production Facility, UK) to generate pOPINTTGNeo_MAM. The resulting expression construct
110 encoded a fusion protein that in addition to human Nrp1 c/MAM domain contained an N-terminal secretion tag
111 and the non-cleavable C-terminal His6-tag.
112 Protein expression was carried out in HEK293 F cells (Invitrogen, Carlsbad, CA, USA) grown in suspension.
113 Typically, 1l flasks seeded with 250ml HEK293 F cells at a density of 1.0-1.4×10^6 cells mL^{-1} were transfected
114 with 312 μg pOPINTTGNeo_MAM plasmid using 470 μg PEI (Sigma Aldrich, Vienna, Austria) mixed in 10
115 mL OptiPRO SFM media (Life technologies, Vienna, Austria) supplemented with 4mM L-Glutamine (Sigma
116 Aldrich). The cells were grown in FreeStyle expression media (Life technologies) while shaken at 125 rpm in a
117 CO_2 incubator. 3-4 days post transfection the medium was collected and cleared by centrifugation at 3000 g for
118 15 min, and filtration through 0.45 μm sterile filter. The clarified media was passed through a 1 mL HisTrap FF
119 column (GE Healthcare, Fairfield, CT, USA) pre-equilibrated with 20 mM Tris-HCl pH 8.0 containing 150 mM
120 NaCl. The column, mounted on an Äkta Purifier FPLC (GE Healthcare), was washed with equilibration buffer
121 supplemented with 20 mM imidazole, before the recombinant protein product was eluted in 20 mM Tris-HCl pH
122 8.0 plus 150 mM NaCl and 500 mM imidazole. The protein was immediately concentrated to about 10 mg/mL
123 by centrifugation with a 10 kDa cutoff concentrator (Millipore, Vienna, Austria) spun at 4000 x g and subjected
124 to size exclusion chromatography on a Superdex S200 10/300 GL column (GE Healthcare) equilibrated in 20
125 mM Tris pH 8.0, 150 mM NaCl. Purified protein eluted from the size exclusion column as a single symmetrical
126 peak (Supplementary Fig. 1). No contaminating protein bands were visible in a Coomassie Blue-stained 15 %
127 SDS-PAGE gel of the eluted fractions corresponding to the main protein peak. About 1 mg of protein per each 1
128 L flask were reproducibly generated Removal of the secretion tag by HEK293 cells and the identity of the N-
129 terminal residue were verified by mass spectrometry analysis as described below. The protein was stored in 20
130 mM Tris pH 8.0, 150 mM NaCl at -80 °C before further analysis.

131

Reductive beta-elimination of *O*-glycans

O- glycans of the recombinant c/MAM domain (50 μg in 10 μl) were released with 1 M sodium borohydride in 100 mM NaOH and purified with Hypercarb SPE cartridges (Hypersep Hypercarb, 25 mg, Thermo Scientific, Vienna). Characterization of these glycans was done via LC-ESI-MS on a porous graphitic carbon (PGC) column (100 x 0.32 mm, Hypercarb, Thermo Scientific) using a Dionex Ultimate 3000 nano-HPLC coupled to an amaZon ion trap (Bruker, Bremen, Germany) as detailed recently [15].

Intact mass measurement

The c/MAM-domain (5 μg) was directly taken for an LC-ESI-MS analysis using a Dionex Ultimate 3000 nano-LC system and a Bruker maXis 4G. The protein was run on a C-5 separation column (Discovery BIO Wide Pore C-5, 50 x 32 mm, 3 μm, Supelco / Sigma Aldrich) at a flow rate of 7 μl /min. After flushing the column for 5 min with 10% acetonitrile (solvent B) in 0.1 % formic acid (solvent A), a gradient up to 75 % acetonitrile within 25 min was applied. Protein species were detected with a modified tune_high method of Bruker with the following settings adapted: mass range 350 – 3000 m/z, spectra rate 0.8 Hz, transfer time and ion cooler RF in basic stepping mode from 91 to 120 μs and from 450 to 700 Vpp, pre pulse storage 14 μs, low mass 400 m/z and in-source CID 30 Ev. Generated protein mass spectra were summed up in a range of 2 min and deconvoluted in Bruker's Data Analysis 4.0 using maximum entropy charge deconvolution.

Fractionation of O-glycopeptides

For initial analysis, the c/MAM-domain (10 μg) was reduced, S-alkylated and digested either with sequencing grade modified trypsin (Promega, Madison, WI, USA) or chymotrypsin (Roche, Vienna, Austria) as described [16]. Subsequent analysis of the generated (glyco)peptides was done via reversed-phase LC-ESI-MS/MS with the said Dionex-HPLC and a maXis 4G Q-TOF MS (Bruker) [15]. For thorough *O*-glycopeptide analysis, 3 mg of S-alkylated c/MAM was digested with chymotrypsin (Roche, Vienna, Austria) overnight 37°C at a substrate to enzyme ration of 1:200. Peptides were fractionated on a BioBasic18 column (250 × 4 mm, 5 μm pore size, Thermo Scientific) with 1 mL / min flow rate with 65 mM ammonium formate of pH 3.0 as solvent A. Glycopeptides were collected in 1 min intervals between 14 and 27 % acetonitrile (solvent B) within 40 min. Fractions were dried in vacuum, taken up in 200 μL water and a 1 μL portion was analyzed via LC-ESI-MS with a very steep gradient to 75 % acetonitrile in 10 min (6 μl / min) on a capillary BioBasic18 column. Glycopeptides were detected with on an Ultima Global Q-TOF MS (Waters-Micromass, Manchester, UK). Fractions of interest were divided and either analyzed as such or after desialylation. To this end, 50 μL aliquots were dried and redissolved in 50 μL neuraminidase buffer (5 mM $CaCl_2$ in 50 mM sodium acetate buffer, pH 5.5) and digested with 1 μL (50 units) α2-3,6,8 neuraminidase (New England Biolabs, Frankfurt/Main, Germany) on 37 °C for 16 h. The glycopeptide species were subsequently purified using C-18 solid phase extraction cartridges (Hypersep C-18, 25 mg, Thermo Scientific). Conditioning was done with 500 μl 80% Acetonitrile (including 0.1 % formic acid) and two flushes with 500 μl 0.1 % formic acid. The samples were applied, washed once with 500 μl 0.1 % formic acid, eluted with 80% acetonitrile plus 0.1 % formic acid and dried in vacuum.

171 **ETD analysis**

172 ETD-MS experiments of the fractions of interest were performed on a Bruker amaZon ion trap equipped with a

173 nanoBooster CaptiveSpray source and prior LC separation. Glycopeptides were flushed onto a C-18 trapping

174 column (C-18 PepMap100, 300 µm x 5 mm, Thermo Scientific, 20 µL / min flow rate), then applied to a

175 separation column (Acclaim PepMap RSLC 75 µm x 250 mm, 2 µm, Thermo Scientific) at a flow rate of 300

176 nL/min and eluted with a gradient from 13 to 40 % acetonitrile in 50 min. The electrospray settings were 1400 V

177 capillary voltage, 500 V end plate offset, 0.2 bar nebulizer pressure, 3 L/min nitrogen gas flow and 150 °C

178 drying gas temperature. The actual ETD reaction was done in Manual MS(n) mode with a 4 Da precursor

179 selection window and ETD reagent times between 40 and 80 ms, depending on the precursor m/z and z values.

180 The ETD data was processed in Data Analysis using the SNAP algorithm for generating average mass spectra of

181 about 1-4 min (about 20-80 single ETD spectra), depending on the peak width. Other ETD settings, procedures

182 and further data evaluation by BioTools 3.2 were done as described [15].

183

184

185 **Results and Discussion**

186

187 **Initial glycoproteomic analysis of c/MAM**

188 In SDS-PAGE the purified c/MAM domain exhibited an apparent mass of about 25 kDa, which considerable

189 exceeds the expected 22 kDa (Supplementary Fig. 1). Glycosylation was assumed as the reason for this

190 discrepancy. To get an impression of the degree of glycosylation the recombinant c/MAM domain was subjected

191 to intact protein mass determination on a high-resolution Q-TOF instrument (Fig. 1). Intervals of 365 and 291 Da

192 between major peaks indicated the complex pattern to be caused by differently glycosylated protein species. No

193 unglycosylated c/MAM domain could be detected (22.308 kDa average mass, range not shown in Fig. 1). The

194 mass differences between experimental peaks and the calculated protein mass suggested the attachment of

195 several mucin type *O*-GalNAc glycans. To prove this hypothesis, *O*-glycans were released by reductive beta-

196 elimination, purified via SPE cartridges and analyzed by PGC-LC-ESI-MS/MS. The identified *O*-glycans were

197 indeed mucin type core 1 and core 2 glycans (Figure 2) with different modifications and degree of terminal

198 sialylation. The well characterized bovine fetuin *O*-glycans (singly and doubly sialylated core 1 and doubly

199 sialylated core 2 *O*-glycans) were used as reference samples [17]. The c/MAM glycans not only showed the

200 same retention behavior on the hypercarb column but also gave the same fragmentation pattern in the CID-

201 MS/MS spectra as the fetuin *O*-glycans (data not shown). Two peaks were obtained for a singly sialylated core 2

202 glycan. Specific reporter ions in MS/MS spectra helped to resolve the main peak at 19.6 min as sialylated on the

203 6-arm. The other glycan version with sialic acid on the 3-arm eluted some minutes earlier as had also been seen

204 by Pabst et al. [18]. A very early eluting (8.1 min) monosialylated core 1 glycan (m/z 677.2, see Fig. 2) gave a

205 fragment at 515 m/z (Neu5Ac-GalNAc-ol) in CID-MS/MS identifying the peak sialylated on the core-GalNAc

206 residue rather than on the galactose. The fully sialylated core-2 glycan was also found sulfated, however,

207 MS/MS in both positive and negative ion mode could not reveal the exact position of sulfation in the glycan.

208 Further minor glycan structures (with an additional HexNAc attached to the GlcNAc) were proven by CID-

209 MS/MS, whereas the position of a fucose on yet another core-2 glycan could not be resolved.

210 For further investigations of the complex glycosylation pattern and to prove that the modification sites are

211 located in the region described in [10], the c/MAM domain was digested with different proteases and the

6

212 glycopeptides were characterized by RP-LC-ESI-CID-MS/MS. The use of the NetOGlyc 4.0 prediction tool for

213 *O*-GalNac-glycosylation site mapping [10] suggested various amino acid candidates as being potentially

214 modified. All of them were situated in the N-terminal part of the protein. Following protease treatment all

215 peptides but the N-terminal peptide were identified as being non-glycosylated. The *O*-glycopeptide was,

216 however, not found in a tryptic but in a chymotryptic digest, where glycan specific reporter ions (292, 366, etc.)

217 in MS/MS spectra helped to identify the elution positions of the various glycoforms of the chymotryptic peptide

218 ETGATEKPTVIDSTIQSEFPTY (Fig. 3). Other than with the fetuin *O*-glycopeptide, only one peak occurred per

219 composition with the exception of gp336 and gp446. Nevertheless, preparative scale reversed phase fractionation

220 of chymotryptic c/MAM-domain peptides was undertaken for sites specific glycosylation analysis by ETD-MS.

221 Chymotryptic peptides were separated on a "preparative" column using a proportionally scaled up gradient as

222 used for capillary LC-ESI-MS. For correct pooling, all fractions were inspected by LC-ESI-MS on the vintage

223 2002 Micromass Ultima Q-TOF instrument. Interestingly, the amaZon ion trap equipped with the standard ESI

224 source was not capable of detecting all glycopeptide forms. Especially peptides with a high glycan content - in

225 particular sialic acid - could not be found.

226

227 **Determination of glycosylation sites and fine structure with ETD-MS**

228 A comparably poor performance of the ion trap was reported for synthetic sialylated *N*-glycopeptides [19]. In

229 order to perform ETD-MS/MS on the amaZon ion trap, we aimed to find a solution for the said ionization

230 problem with the standard ESI source that prevented ETD analysis of the more intensely glycosylated species in

231 on-line LC and even in direct infusion mode. Nanospray LC-ESI with the nanoBooster CaptiveSpray source was

232 recently reported to promote the ionization of (sialylated) glycopeptides [20]. In fact, the nanoBooster

233 CaptiveSpray allowed ionization of all glyco-variants sufficiently well to obtain useful ETD spectra, however,

234 with the heavily glycosylated forms only when desialylated. Fortunately, in the particular case of Nrp1 *O*-

235 glycosylation, desialylation did not lead to loss of information as all major structures were fully sialylated. With

236 the Nrp1 glycopeptide being considerably smaller than the fetuin *O*-glycopeptide, ETD spectra could be obtained

237 on an LC-ESI (rather than direct infusion) timescale using manual, targeted MS/MS mode, which permits dwell

238 times of well over 1 min.

239 Precursor charge state emerged as a most relevant criterion for ETD of other compounds [21, 22], it made little

240 difference for Nrp1 glycopeptides. In contrast, co-eluting peptides severely affected ionization and hence

241 fragmentation. Therefore, the prior fractionation step proved beneficial. Although nominally the same separation

242 mechanism was applied, subtle differences in selectivity between preparative and analytical column led to more

243 homogenous peaks on the latter.

244 The smallest glycopeptide, gp112, could readily be elucidated as bearing it's doubly sialylated core-1 glycan on

245 T5. Gp224 was modified at T5 and T14 but not S13 as derived from the albeit small c_{13} ion (Fig. 4).

246 With already six sialic acids, the poor ionization on the amaZon ion trap prevented gp336 ETD analysis and so it

247 was the first glycopeptide subjected to desialylation by an α2-3,6,8 neuraminidase. Notably, only three sialic

248 acids were removed this way – enough to provide a strong signal and ETD fragment spectra. At first, we

249 regarded this partial desialylation as the incidental result of insufficient digestion conditions. However, the more

250 highly sialylated structures exhibited an increasingly more complete removal of Neu5Ac with the core 2 type

251 only species being fully desialylated. The overall pattern of incomplete desialylation compellingly implied that

252 the GalNAc bound α2,6-Neu5Ac had resisted the neuraminidase treatment whereas the Gal-linked Neu5Ac

253 residues had been quantitatively removed. A PGC-LC-MS/MS measurement of neuraminidase treated, β-

254 eliminated *O*-glycans finally confirmed this hypothesis (Supplementary Fig. 2).

255 Figure 5 gives an overview of the identified glycopeptide variants with glycans attached to specific amino acids.

256 As mentioned before, glycopeptide forms with one or two glycans are always modified on T5 and T14, with T5

257 being favored over T14. Glycopeptide forms with three modification sites (gp336, gp446 and gp666) showed

258 always occupation at T5 and T14 while usage of T9 or S13 was variable. The differential modification of T9 or

259 S13 in gp336 and gp446 led to the separation of distinct glycoforms in reversed phase chromatography. The

260 early eluting peaks were modified at S13, whereas the later eluting peaks were glycosylated at T9. In contrast,

261 the attachment of os222 on different residues in gp446 did not alter its elution behavior and produced only one

262 peak (Fig.5).

263 To conclude, four sites were used, i.e. T5, T9 S13 and T14 or T629, T633, S637 and T638 when counted in

264 O14786 - NRP1_HUMAN (Fig.3). Interestingly, in gp666, gp778 and gp888, two rather bulky core-2 glycans

265 were found on the neighboring amino acids S13 and T14.

266 Three corresponding ETD spectra that were used for modification site determination are given in Figure 4A-C.

267 The gp224 and gp666 (measured as gp660) spectra were chosen as representing one natural and one desialylated

268 glycopeptide. In each case, the yield of c- and z-fragment ions was sufficient for assigning the exact

269 glycosylation sites. In many glycopeptide spectra, a loss of some parts of a glycan during the ETD reaction could

270 be observed. Loss of single sialic acids in natural glycopeptide samples or of HexNAc-Hex (366.1 $[M+H]^+$ see

271 figure 4) was observed. CID-like side reactions during ETD are known from literature [23, 24] and did not

272 compromise the identification of glycan modification sites in any way. Figure 6c shows the ETD spectrum of

273 gp334 that led to the identification of two different glycopeptide forms in the same sample.

274 When comparing the detected glycosylated amino acids with those proposed as modified by the *O*-GalNAc-site

275 prediction tool NetOGlyc 4.0, an over-estimation of glycosylation sites could be seen. The software assessed all

276 actually glycosylated sites as positive, however, also the sites T2, S17 and T21 were predicted to be modified

277 when having a look at recombinant c/MAM domain. With the complete neuropilin-1 sequence, T21 (T645 on

278 NRP-1) was interestingly not suggested to be glycosylated.

279

280 **Substrate specificity of *C. perfringens* neuraminidase**

281

282 The hypothesis that neuraminidase disdains the GalNAc-linked Neu5Ac was tested with disialylated core 1 and

283 core 2 *O*-glycans. In fact, the two Gal-linked sialic acids from core 2 glycans were readily removed and so was

284 one Neu5Ac of the core 1 substrate (Supplementary Fig. 2). The remaining mono-sialylated structure was

285 identified as Neu5Ac (Gal-)GalNAc by PGC-LC-MS/MS. In contrast to glycopeptide linked glycans, GalNAc-

286 linked sialic acid was degraded from free glycans but so much slower that this neuraminidases specificity can be

287 useful for discrimination of differently linked sialic acid.

288

289 **Conclusion**

290

291 Recombinantly expressed neuropilin-1 c/MAM-domain was found to be heavily *O*-glycosylated in the region

292 recently reported as bearing several *O*-GalNAc residues when expressed in Simple Cells, which lack the ability

293 to add the β1,3-Gal to the peptide bound GalNAc [10]. The main focus of this work was the comprehensive

294 definition of the glycosylation sites and the attached glycans expressed by human cells with unquenched *O*-

295 glycosylation capability. The *O*-glycans found were all fully sialylated in contrast to *e.g.* fetuin [15] with one

296 glycan being sulfated as well.

297 ETD-MS proved its reputation for being the method of choice for this kind of analysis [25, 26], whereby the

298 largest difficulty encountered with the highly sialylated neuropilin-1 glycopeptides was their poor ionization in

299 the ESI source on the ion trap used for ETD.

300 A glycopeptide species pyramid with the major glycopeptide gives an overview of the analytical results (Fig. 6).

301 Four amino acids were identified to carry the *O*-glycan modifications (see Fig. 3 B).

302 The isomers found suggest some hierarchy in the order of *O*-glycan addition. T5 was found modified in all cases

303 and thus constituted the first target of a polypeptide GalNAc-transferase, strictly followed by T14. The other two

304 glycosylation sites (T9 and S13) were not subject to such strict hierarchy and can be seen as equal *O*-glycan

305 acceptors. As both core-1 and core-2 *O*-glycans were found on each of the four amino acids, no preference of the

306 processing glycosyltransferases for a certain site - as was seen in other studies [15] - could be observed. The

307 rather equal amounts of core-1 and core-2 glycans observed on c/MAM implies a balance between the sialylation

308 reaction of a T-antigen structure and the buildup of a core-2 structure (addition of a GlcNAc onto the T-antigen)

309 [27].

310 The *O*-glycosylation region of neuropilin-1 is situated between two different outer membrane domains, the

311 c/MAM domain and b2 domain [28]. Regarding the function of this *O*-glycosylation, it can only be speculated

312 that the glycans may help this protein region to adopt the correct, stable conformation by forming a highly

313 hydrophilic sleeve around the connecting peptide stretch. Another interesting fact is that a described

314 glycosaminoglycan modification [3], known to influence VEGF signaling, is only 16 aa upstream of the first

315 detected *O*-GalNAc modification site. Thus, the heavy *O*-GalNAc glycosylation in this part may also play some

316 role in binding VEGFR2 or other binding partners as galectin-1. Certainly, future studies will have to illuminate

317 the probable functions of this novel glycosylation.

318 Finally, two technical aspects shall be emphasized. First, very unlike *N*-glycopeptides of different glycan

319 structure [16, 29], *O*-glycopeptides of differing composition exhibit a large spread of elution times as seen with

320 fetuin and again in this study.

321 Second, the (recombinant) sialidase from *Clostridium perfringens* exhibited an as yet unnoticed substrate

322 specificity in being almost inactive towards the GalNAc-linked α2,6-neuraminic acid in *O*-glycopeptides.

323

324   **Conflict of interest**

325

326   The authors declare that they have no conflict of interest.

327

328

329

330   **Acknowledgements**

331

332

333

334 **References**

335 1. Giger, R.J., et al., *Semaphorin III: role in neuronal development and structural plasticity.* Prog
336   Brain Res, 1998. **117**: p. 133-49.
337 2. Frankel, P., et al., *Chondroitin sulphate-modified neuropilin 1 is expressed in human tumour*
338   *cells and modulates 3D invasion in the U87MG human glioblastoma cell line through a*
339   *p130Cas-mediated pathway.* EMBO Rep, 2008. **9**(10): p. 983-9.
340 3. Shintani, Y., et al., *Glycosaminoglycan modification of neuropilin-1 modulates VEGFR2*
341   *signaling.* EMBO J, 2006. **25**(13): p. 3045-55.
342 4. Appleton, B.A., et al., *Structural studies of neuropilin/antibody complexes provide insights*
343   *into semaphorin and VEGF binding.* EMBO J, 2007. **26**(23): p. 4902-12.
344 5. Chen, R., et al., *Glycoproteomics analysis of human liver tissue by combination of multiple*
345   *enzyme digestion and hydrazide chemistry.* J Proteome Res, 2009. **8**(2): p. 651-61.
346 6. Hsieh, S.H., et al., *Galectin-1, a novel ligand of neuropilin-1, activates VEGFR-2 signaling and*
347   *modulates the migration of vascular endothelial cells.* Oncogene, 2008. **27**(26): p. 3746-3753.
348 7. Meynier, C., F. Guerlesquin, and P. Roche, *Computational Studies of Human Galectin-1: Role*
349   *of Conserved Tryptophan Residue in Stacking Interaction with Carbohydrate Ligands.* Journal
350   of Biomolecular Structure & Dynamics, 2009. **27**(1): p. 49-57.
351 8. Carlsson, M.C., et al., *Galectin-1-binding glycoforms of haptoglobin with altered intracellular*
352   *trafficking, and increase in metastatic breast cancer patients.* PLoS One, 2011. **6**(10): p.
353   e26560.
354 9. Quinta, H.R., et al., *Glycan-dependent binding of galectin-1 to neuropilin-1 promotes axonal*
355   *regeneration after spinal cord injury.* Cell Death and Differentiation, 2014. **21**(6): p. 941-955.
356 10. Steentoft, C., et al., *Precision mapping of the human O-GalNAc glycoproteome through*
357   *SimpleCell technology.* EMBO J, 2013. **32**(10): p. 1478-88.
358 11. Hakomori, S. and K. Handa, *Glycosphingolipid-dependent cross-talk between glycosynapses*
359   *interfacing tumor cells with their host cells: essential basis to define tumor malignancy.* Febs
360   Letters, 2002. **531**(1): p. 88-92.
361 12. Lau, K.S., et al., *Complex N-glycan number and degree of branching cooperate to regulate cell*
362   *proliferation and differentiation.* Cell, 2007. **129**(1): p. 123-134.
363 13. Theocharis, A.D., et al., *Chondroitin sulfate as a key molecule in the development of*
364   *atherosclerosis and cancer progression.* Adv Pharmacol, 2006. **53**: p. 281-95.
365 14. Shields, R.L., et al., *Lack of fucose on human IgG1 N-linked oligosaccharide improves binding*
366   *to human Fc gamma RIII and antibody-dependent cellular toxicity.* Journal of Biological
367   Chemistry, 2002. **277**(30): p. 26733-26740.
368 15. Windwarder, M. and F. Altmann, *Site-specific analysis of the O-glycosylation of bovine fetuin*
369   *by electron-transfer dissociation mass spectrometry.* J Proteomics, 2014. **108**: p. 258-68.
370 16. Pabst, M., et al., *Glycan profiles of the 27 N-glycosylation sites of the HIV envelope protein*
371   *CN54gp140.* Biol Chem, 2012. **393**(8): p. 719-30.
372 17. Edge, A.S. and R.G. Spiro, *Presence of an O-glycosidically linked hexasaccharide in fetuin.* J
373   Biol Chem, 1987. **262**(33): p. 16135-41.
374 18. Pabst, M., et al., *IL-1beta and TNF-alpha alter the glycophenotype of primary human*
375   *chondrocytes in vitro.* Carbohydr Res, 2010. **345**(10): p. 1389-93.
376 19. Stavenhagen, K., et al., *Quantitative mapping of glycoprotein micro-heterogeneity and*
377   *macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic*
378   *peptides and glycopeptides.* Journal of Mass Spectrometry, 2013. **48**(6): p. 627-639.
379 20. Marx, K., A. Kiehne, and M. Meyer *amaZon speed ETD: Exploring glycopeptides in protein*
380   *mixtures using Fragment Triggered ETD and CaptiveSpray nanoBooster*. 2014, Bruker
381   application note LC-MS-93, 04/2014, Bruker Daltonics GmbH, http://bdal.de, accessed 23
382   October 2014.
383 21. Liu, J. and S.A. McLuckey, *Electron Transfer Dissociation: Effects of Cation Charge State on*
384   *Product Partitioning in Ion/Ion Electron Transfer to Multiply Protonated Polypeptides.* Int J
385   Mass Spectrom, 2012. **330-332**: p. 174-181.

386 22. Good, D.M., et al., *Performance characteristics of electron transfer dissociation mass spectrometry.* Mol Cell Proteomics, 2007. **6**(11): p. 1942-51.

388 23. Catalina, M.I., et al., *Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain.* Rapid Commun Mass Spectrom, 2007. **21**(6): p. 1053-61.

390 24. Mormann, M., H. Paulsen, and J. Peter-Katalinic, *Electron capture dissociation of O-glycosylated peptides: radical site-induced fragmentation of glycosidic bonds.* European Journal of Mass Spectrometry, 2005. **11**(5): p. 497-511.

393 25. Thaysen-Andersen, M., et al., *Site-specific characterisation of densely O-glycosylated mucin-type peptides using electron transfer dissociation ESI-MS/MS.* Electrophoresis, 2011. **32**(24): p. 3536-45.

396 26. Takahashi, K., et al., *Clustered O-glycans of IgA1: defining macro- and microheterogeneity by use of electron capture/transfer dissociation.* Mol Cell Proteomics, 2010. **9**(11): p. 2545-57.

398 27. Brockhausen, I., *Pathways of O-glycan biosynthesis in cancer cells.* Biochim Biophys Acta, 1999. **1473**(1): p. 67-95.

400 28. Djordjevic, S. and P.C. Driscoll, *Targeting VEGF signalling via the neuropilin co-receptor.* Drug Discov Today, 2013. **18**(9-10): p. 447-55.

402 29. Kolarich, D., et al., *Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms.* Proteomics, 2006. **6**(11): p. 3369-80.

404

Fig. 1: **Total mass determination of recombinant neuropilin-1 c/MAM domain.** Deconvoluted average mass spectrum of c/MAM glycoprotein species. The most intense protein forms were named according to the number of attached *N*-acetylhexosamine, hexose and sialic acid sugars. Peaks marked with the same color were found to carry different *O*-glycans on the same amino acid side chains (see also figure 4 and 5). c/MAM glycoprotein was also found truncated of the first N-terminal four amino acids (ETGA - see peak gp224*).

Fig. 2: **Analysis of released *O*-glycans**. Reduced *O*-glycans as obtained by reductive β-elimination were separated by PGC-LC and detected by mass spectrometry.

Fig. 3: **Variety of neuropilin-1 *O*-glycopeptides.** The chymotryptic *O*-glycopeptide of neuropilin-1 c/MAM domain was subjected to RP-LC-ESI-MS with detection by a Q-TOF Ultima MS (Waters Micromass). Panel **A** depicts the extracted ion chromatograms for the major compositions illustrating the largely different retention as caused by different glycosylation. Only two compositions eluted in more than one peak. Panel **B** shows the amino acid sequences of chymotryptic N-terminal glycopeptide of the recombinant neuropilin-1 c/MAM-domain and the corresponding neuropilin-1 Uniprot entry (O14786 - NRP1_HUMAN). Residues marked in red were found to be *O*-glycosylated.

Fig. 4: **Summed ETD spectra of gp224 (a), the desialylated gp660 (b) and gp334 (c).** Identified c- and z-fragment ions are marked in red and blue. Confirmed glycopeptide forms are shown in the upper right corner of the graphs with all corresponding fragment ions depicted in red. In part c, two different glycopeptide variants could be identified (marked with * and °). Fragment ions that discriminate between these forms in the graph are as well marked with * and °.

Fig. 5: **Glycopeptide species of the neuropilin-1 c/MAM domain of recombinant illuminated with ETD-MS**. Glycopeptide variants with exclusively core 1 *O*-glycans are listed in the left column. Forms with core 2 glycans only are displayed at the right side, whereas the variants with both types of core structures are shown in the middle. Forms marked in blue (e.g. gp448) were analyzed after desialylation. As all the major forms considered for determination of exact site occupancy were doubly sialylated (see Fig. 1), the full sialylation regalia of these glycopeptides could be re-constructed as shown. Glycan modifications on S13 and T14 could not be resolved in all glycopeptide species, ambiguities are marked with a line (e.g. gp558). Gp446 marked with a * could not be definitely confirmed or excluded.

Fig. 6: **Glycopeptide pyramide.** The top field shows the relevant part of c/MAM in its unglycosylated state that was, however, not observed. The numbers in red represent c/MAM-peptides with the respective number of sugar

446   residues attached, whereby 224 indicates the presence of 2 HexNAc (1 GalNAc, 1 GlcNAc), 2 galactoses and 4

447   sialic acid residues. Going down one level one *O*-glycan is attached at its step. If the direction points to the left, a

448   112 core 1 glycan is added, if to the left, a 222 core 2 glycan is attached. The colored fields contain peptides with

449   the same number of *O*-glycan moieties. The glycan attachment sites in the peptides are marked in red.

450

451   Supplementary Fig. 1: **Quality control of purified recombinant c/MAM domain.** Size exclusion

452   chromatography and SDS-PAGE demonstrate high protein purity and an apparent size of ~ 25 kDa.

453

454   Supplementary Fig. 2: **Proof of novel α2-3,6,8 neuraminidase substrate specificity.** Panel A shows the

455   chromatograms of the two main *O*-glycan species, derived from reductive beta elimination of neuraminidase

456   treated chymotryptic recombinant neuropilin-1. The structure of the core 1 glycan was verified by positive mode

457   CID fragmentation, giving the feature ion 515.1 m/z (panel B). Panel C shows a CID fragmentation spectrum of

458   the main fetuin *O*-glycan being sialylated on the galactose. In contrast to panel C, 515 m/z shows no signal

459   whereas a specific feature ion at 454.1 m/z (Neu5Ac-Gal) appears.

460

461

462 Fig. 1:

463
464

465 Fig. 2:

◆ *N*-acetylneuraminic acid   ● Galactose   ■ *N*-acetylgalactosamine   ■ *N*-acetylglucosamine   ▲ Fucose
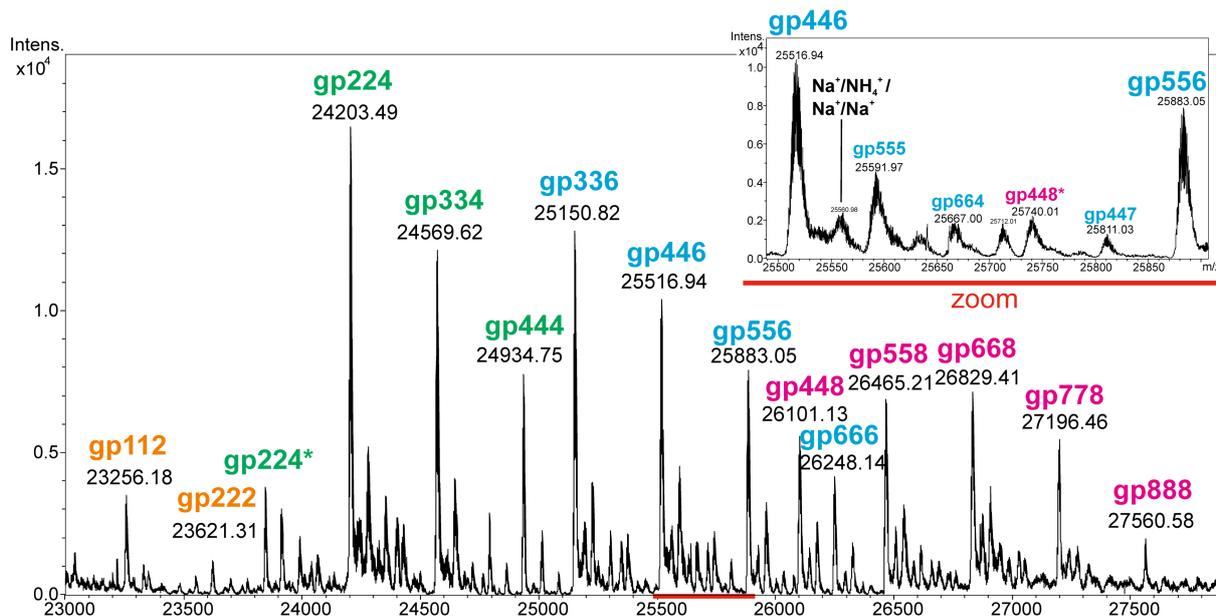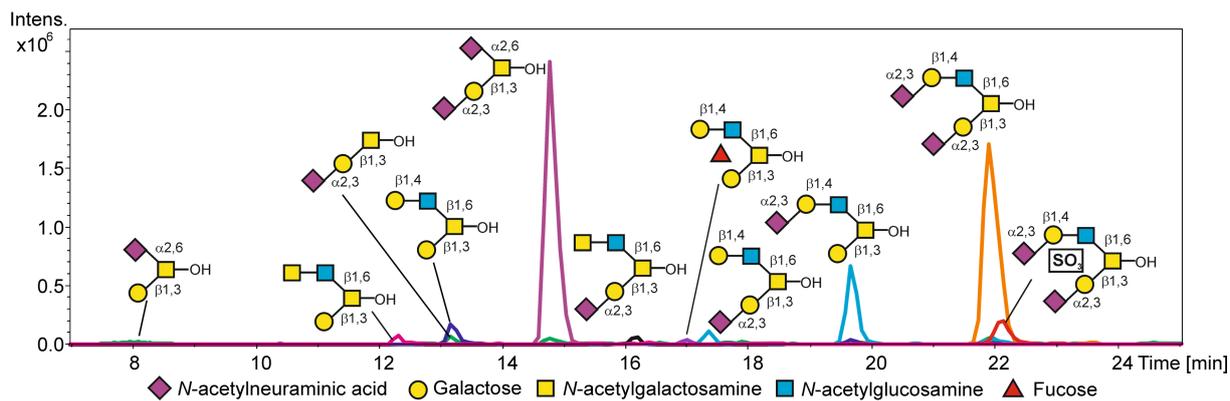
466
467 Fig. 3:

468
469

Fig. 4:

475   Fig. 5:

476
477
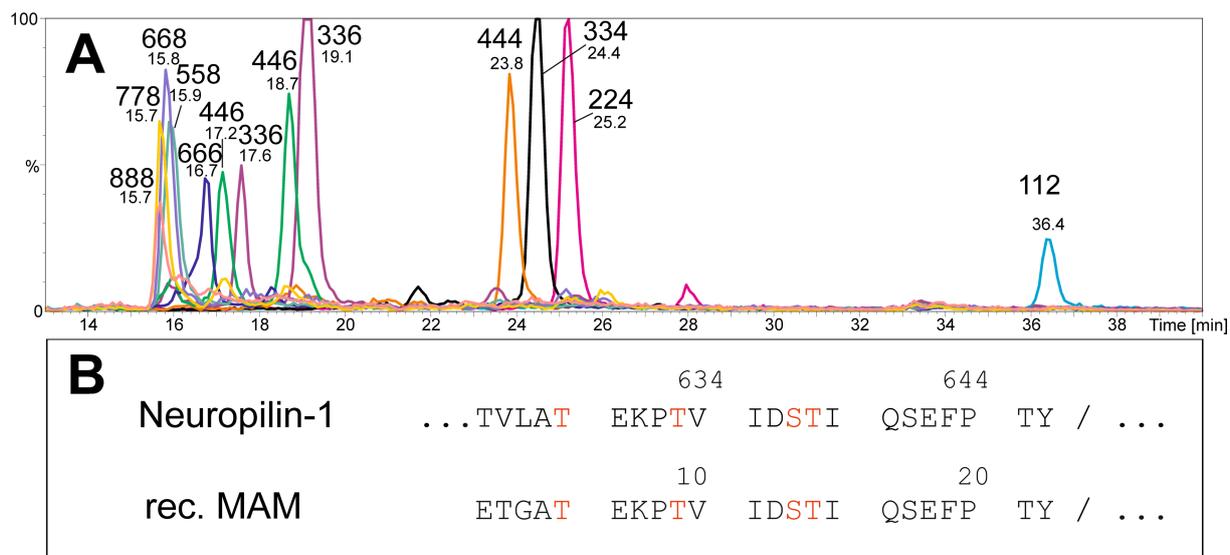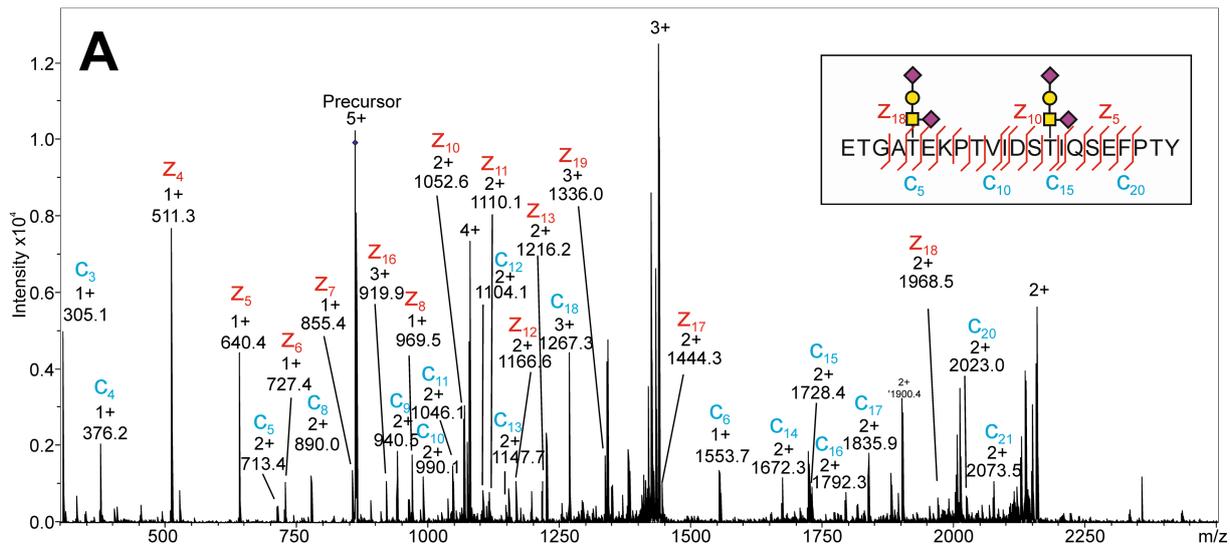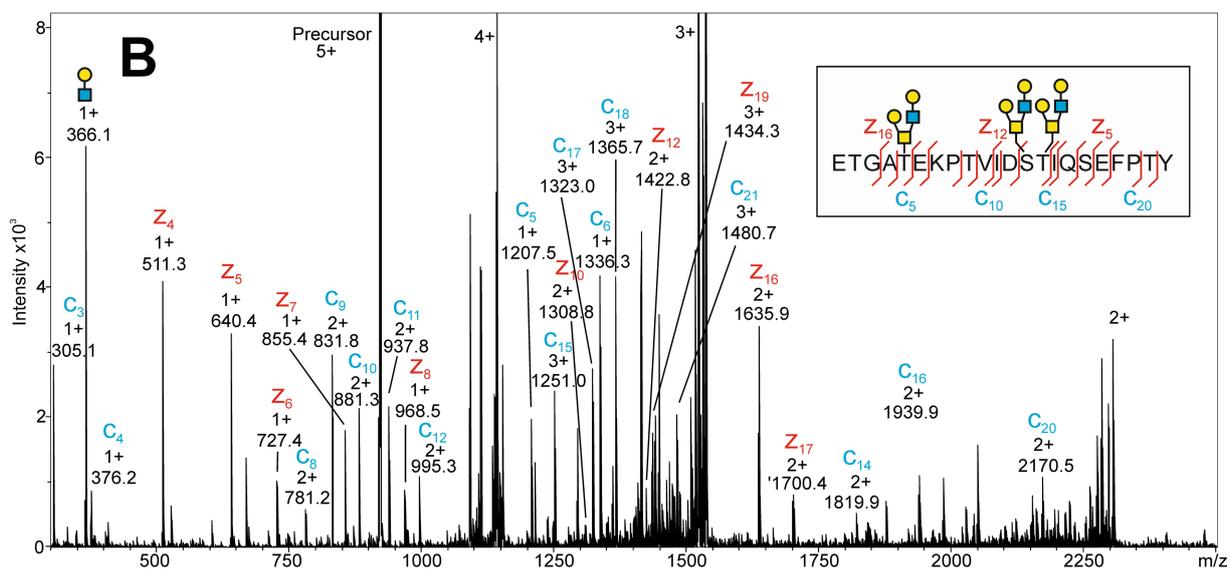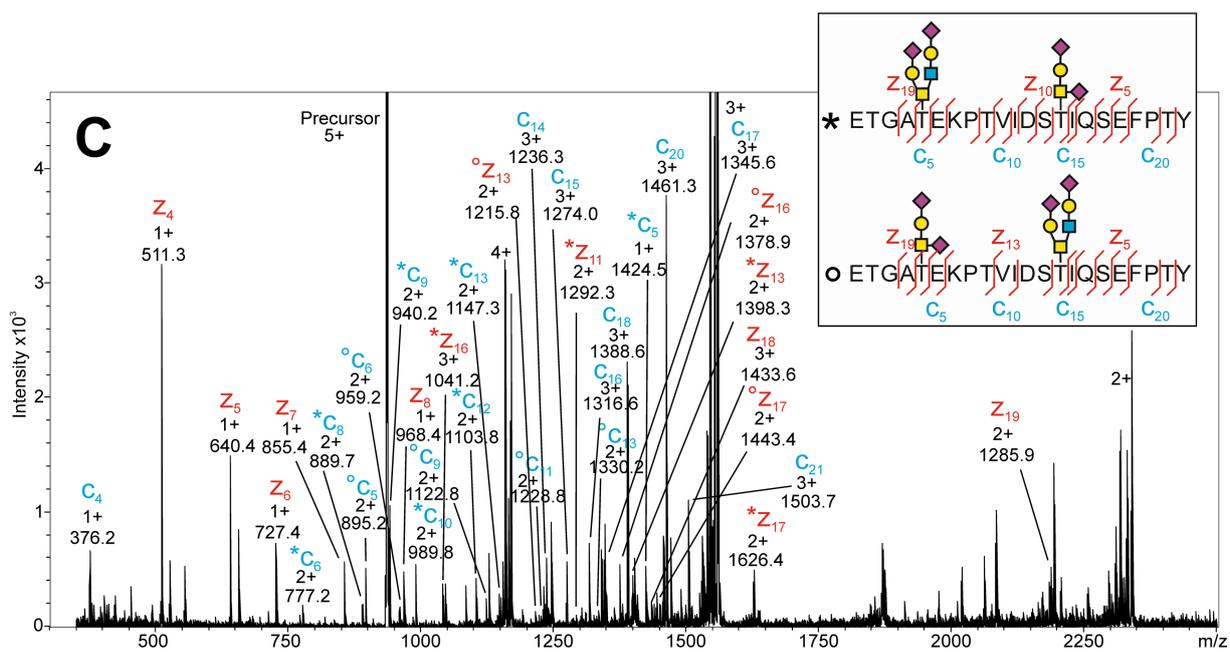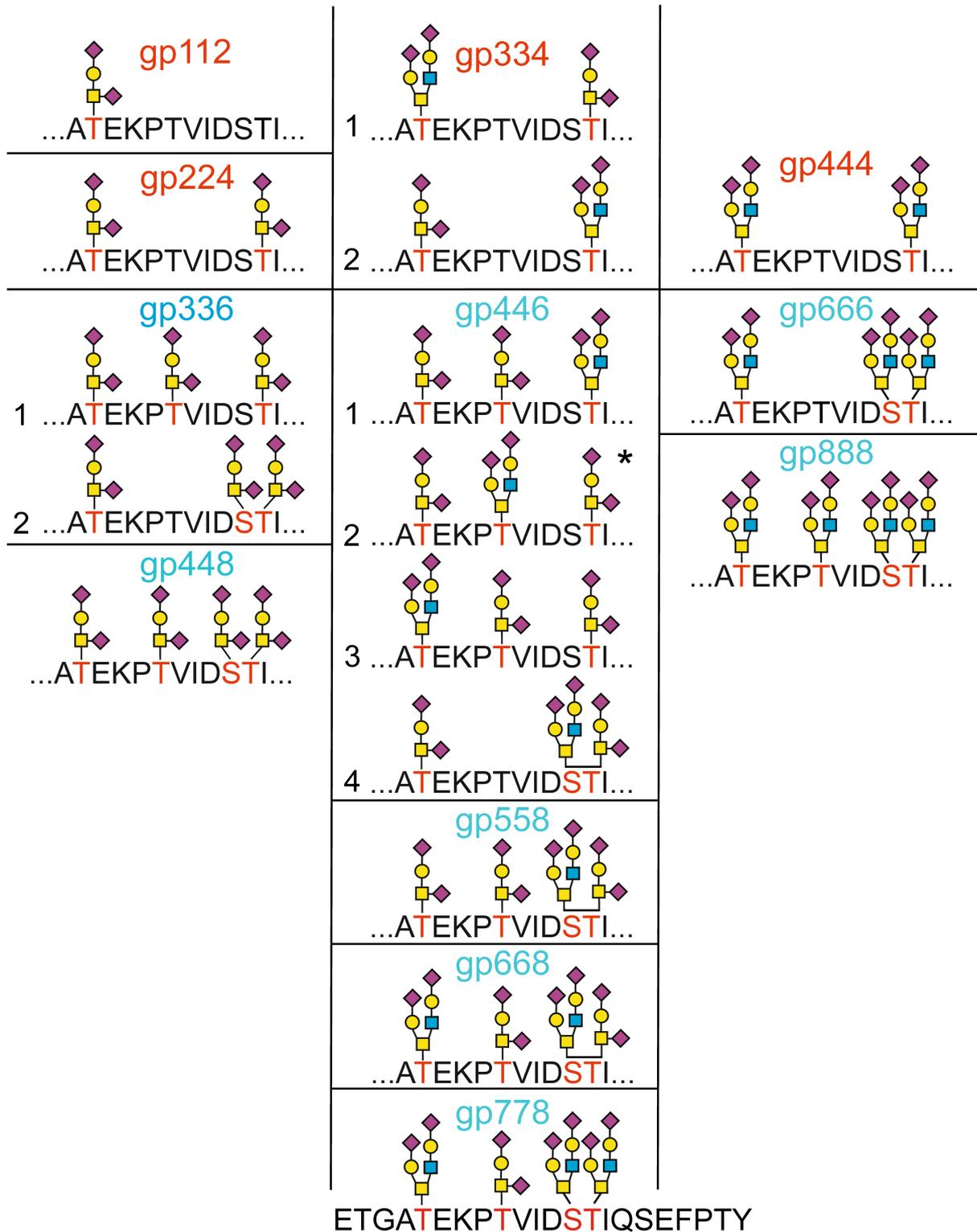
17

478   Fig. 6:
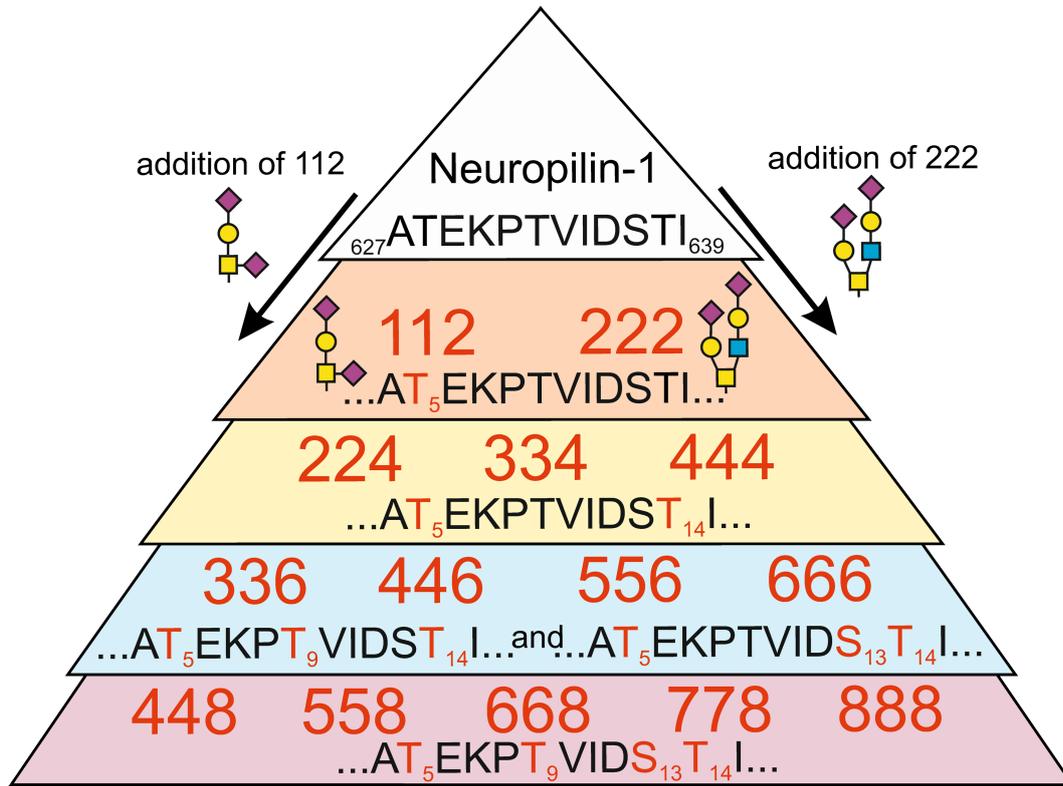
479
480

481   Supplementary Fig. 1:
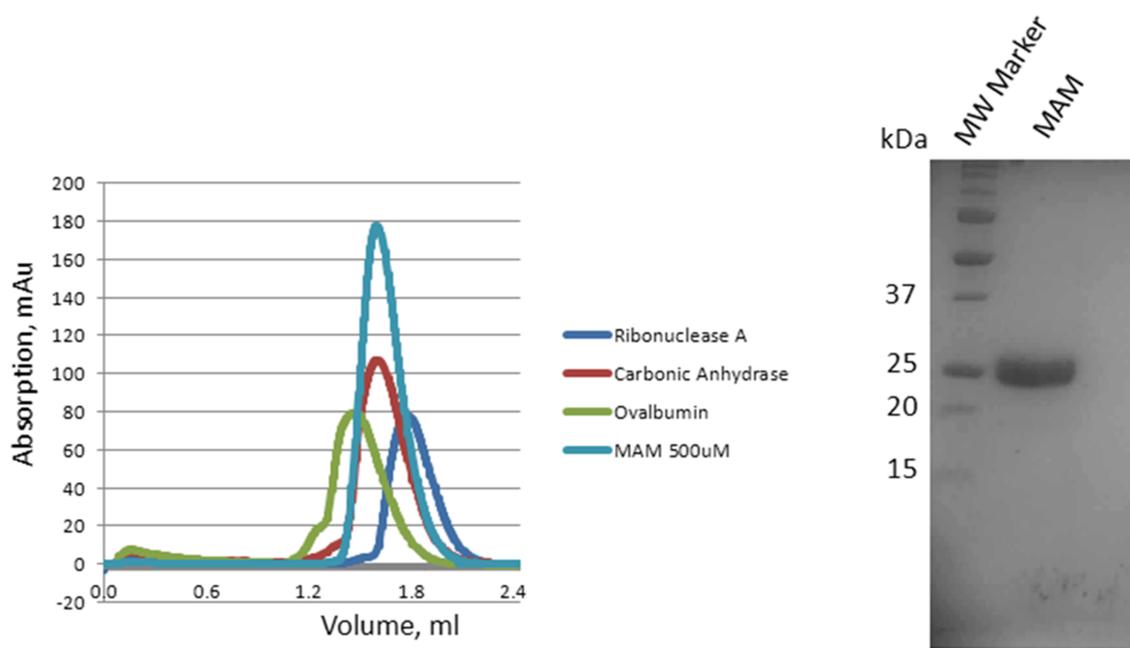
482
483

484   Supplementary Fig. 2:

485