Exposition of the N-glycan nomenclature system "proglycan"

Drafted by:Friedrich Altmann, Johannes Helm, Johannes StadlmannUniversity of Natural Resources and Life Sciences, Vienna. AustriaContact:friedrich.altmann@boku.ac.atContact

Teaser

In the beginning was the word but there were no words for N-glycans. Look at the entry in a relevant <u>database</u> for the structure that we will herein baptize **A**⁴**A**⁴**F**⁶ and ask yourself, how you could label your Eppendorf vial with the therein suggested options of which good, old IUPAC code is probably still the most human-friendly one - whereby this an already simplified version:

Gal(b1-4)GlcNAc(b1-2)Man(a1-3)[Gal(b1-4)GlcNAc(b1-2)Man(a1-6)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)] GlcNAc

Another example using a recently published drawing tool [1], the herein used condensed IUPAC code and the respective "proglycan" cartoon and acronym.



N-glycans are just one group in the huge universe of carbohydrates, but in medical biotechnology and bio-pharmaceutics, they supremely reign our attention hierarchy. So, we assume that a comprehensive and logical naming system could be useful and therefore we apply Occam's razor to N-glycan nomenclature.

Content

The beginning Galactosylation (including the alfa-Gal epitope) Gloss about structure cartoons Core-Fucosylation Fucose on antennae Bisecting GlcNAc Sialic acids Multi-antennary glycans LacNAc-repeats High-mannose glycans Plant and insect structures Specialties Comparison with existing systems



Prolog

The number of already existing abbreviation systems for N-glycans cannot rival that of sand grains on sea shores. However, it seems to us that none of the various modes of the currently used naming systems is capable of representing (almost) all N-glycan structures occurring in mammals, insects and plants with a decent number of characters in an unambiguous manner. The big exceptions are the machine codes such as GlycoCT [2], which are by no means apprehensible for the human eye unless somehow translated to visibility [1]. So, quite often researchers give up and just show structure cartoons. These, however, are unsuitable for oral or written transmission or for labeling of vials.

The herein introduced system has proven useful for communication with partners for already many years. The terms MMXF³ and MUXF³ (with or without the superscript number) enjoy widespread use in the allergy diagnosis community. Our recent experience with 40 isomeric N-glycans all composed of 5 hexoses, 4 HexNAcs and 1 fucose [3,4] reinforced our conviction that the "proglycan" system is highly useful. Therefore, we shall not surrender in our fight against the inertia exerted by beaten tracks.

For frequently occurring "default" structures, various naming systems are in use. However, in our eyes, none of these is apt to unambiguously describe a large segment of all possible structures in a systematic, easily understandable manner. A collection of such systems will be shown in the last chapter of the present treatise.

Proglycan Code	IUPAC Code	CFG Code (Vienna style)
GnGn	$GleNAc\beta-2Man\alpha$ $GMan\beta-4GleNAc\beta-4GleNAc$ $GleNAc\beta-2Man\alpha$	
MGn	$\underbrace{\frac{Man}{6}}_{GleNAc\beta} - 2Man\alpha}^{Man} \beta - 4GleNAc\beta - 4GleNAc}$	
GnM	$\underbrace{\frac{\text{GlcNAc}\beta-2\text{Man}\alpha}{6}}_{\text{Man}\beta-4\text{GlcNAc}\beta-4\text{GlcNAc}}$	• •
ММ	$ \frac{Man\alpha}{6} $ $ \frac{6}{3} $ Man\beta - 4GlcNAc\beta - 4GlcNAc Man\alpha	• •

The beginning

Concealing the origin of an idea is widespread practice but we shall ignore this habit. During a brief stay in our lab in Vienna around 1990, Harry Schachter from Toronto taught us the term "GnGn" for the acceptor substrate of fucosyl transferases [5,6]. Gn stands for GlcNAc and the two Gn-s symbolize

the two terminal residues of the biantennary N-glycan. The rest is unambiguously clear. No further definitions are required.

This substrate often fell prey to hexosaminidases leaving two isomers with just one GlcNAc, which can easily be named according to now terminal mannose residue(s) represented by an "M". If two GlcNAc are removed, we end up with "MM". However, if only one mannose gets exposed, we need a rule about the sequence of reading.

<u>RULE 1: Terminal residues are read from top to bottom = counterclockwise</u>

The first term describes the 6-arm antenna just annotating the non-reducing terminal sugar. The second term describes the 3-arm.

<u>RULE 2: Monosaccharides are depicted by one capital letter. Modifications are specified by a</u> <u>subsequent small letter.</u>

G glucose	Gn N-acetylglucosamine
M mannose	Na N-acetylneuraminic acid
A galactose	Ng N-glycolylneuraminic acid
F Fucose	X Xylose

Galactosylation (including the alfa-Gal epitope)

Often, the antennae are elongated by galactose whose one-letter code [7] is A. Together with above mentioned three acceptors this leads to a total of 4 products if we only consider β 1,4-linkages. Rarely, however, the linkage may be β 1,3. To account for this ambiguity, we use superscripts.

We now know the three basic terminal elements M, Gn and A (either A³ or A⁴) of an antenna. It is clear by definition that "Gn" and "A⁴" annotate the sequences GlcNAc(β 1-2)Man α 1- and "A⁴" Gal(β 1-4)GlcNAc(β 1-2)Man(α 1-, respectively. Whenever we further elongate the sequence, we need to include the terminal Gal residue. The first example for this necessity is the Gal α 1,3-Gal β 1- element. We might resort to something like A α 1-3A⁴ or A α 1-3A³. The suggested annotation, however, is A³⁻⁴ or A³⁻³ [3], whereby the two figures point at the two residues involved. Unless – very unexpectedly – a novel structural element turns up that could be described by this combination, the terms are unambiguous.

GnA ⁴		A ⁴ A ⁴	0-B.9 0-B-9
A ³ Gn		A ⁴ A ³	о-в.е. о-в-е- о ^{-в.е.}
MA ⁴		A ³ A ⁴	
GnA ³⁻⁴	₽ 0 0 0 0	A ³⁻⁴ A ⁴	0 ^{.0-11} 0 0-110

Gloss about structure cartoons

While essentially using the cartoon format of the Consortium of Functional Glycomics (CFG), we mix in a bit of the "Oxford glycobiology" system [8], in which anomericity and linkage position are described solely by the connecting lines. In order to maintain the familiar overall appearance of structures, the 1-2 linkage of antennary GlcNAcs is shown as a bended line.

Strict adherence to the linkage angle rule allows omission of the linkage specifications by characters and this again allows to unambiguous depiction of structures even at very small size.



Core-Fucosylation

The core-fucose constitutes a third terminal residue and hence we introduce a third structure term, "F" for fucose. We could simply write e.g. A^4A^4F . In mammals, the core fucose is strictly always in α 1,6. If you only work with mammalian samples you may content yourself with this simplification. However, as insect cells and plants have some relevance - certainly in biotechnology – we must consider, that here the fucose can sit or sits in the α 1,3-position. Therefore, for the sake of clarity, the superscripts should be used to define the type of core-fucose.

GnGnF ⁶	A ⁴ A ⁴ F ⁶	
A ³ GnF ⁶	MA ⁴ F ⁶	
MMF ³	MMF ³ F ⁶	

Fucose on antennae

Lewis fucoses introduce branching of the antenna. IUPAC nomenclature uses square brackets to identify a branch. We do something similar. However, the two residues, or chains, that are linked to the root of the branch are both put in round brackets. The substitution points are defined by superscripts. So, LeX fucosylation of an A⁴ antenna turns this term into (A⁴F³). This A LeA structure would be (F⁴A³) because we read the structure counter-clockwise. Nothing wrong. However, we want to be elegant and as concise as possible. Are we losing any information when omitting the superscripts? Obviously, not.

Advanced coding:

The terms (AF) and (FA) perfectly describe the terminal structures. In the near future, we may substitute the terms by "macros", i.e. Lx for the Lewis X terminus and La for the Lewis A determinant.

Another difficulty is posed by the blood group H α 1,2-fucose, which is linked to galactose, which in turn can be linked β 1,3- or β 1,4 to GlcNAc. So just putting "F" as the terminal sugar would leave uncertainty. Therefore – using linear code [7] – we write F²-A⁴. We – again - can save one character by omitting the A to arrive at: F²⁻⁴ or F²⁻³. Why not just F⁴ or F³ ? Because, we must not ignore the galactose. "F⁴" would be a Fuc(α 1-4)GlcNAc sub-structure, which does not exist.

RULE 3: If more than one terminal residue occurs on one antenna, these residues are put in brackets. Round brackets are used for all branching except that arising from GlcNAc-transferase IV and V (see below "Multiantennary glycans".

<u>RULE 4:</u> Substituents to the β-galactose are linked to this residue by a hyphen. The terms</u> -A⁴ or -A³ are abbreviated to the superscripts -4 or -3.

(FA)Gn LaGn	A ⁴ (AF) A ⁴ Lx	
A ³ (FA)F ⁶ A ³ LxF ⁶	A ⁴ F ²⁻³ A ⁴ Lh ³	

Bisecting GlcNAc

Bisected N-glycans are a peculiar feature of IgG and especially brain glycans. To annotate this residue, we almost complete our circumnavigation of the N-glycan. Hence, the term "bi" is found at the very end of the abbreviation.

GnGnbi		A ⁴ GnF ⁶ bi (AGnFbi)	
--------	--	--	--

RULE 5: Bisected GlcNAc is indicated by "bi" and is always listed as the last extension term

Sialic acids

Sialic acids are (usually) not linked directly to GlcNAc and therefore the same approach as for alfa-Gal and blood-group H fucose is chosen. Thus, Neu5Ac(α 2-6)Gal(β 1-4)GlcNAc(β 1,2)Man(α 1- shrinks to Na⁶⁻⁴.

Na³⁻³, Na⁶⁻⁴ and the probably not existing Na⁶⁻³ denote the other options for a sialylated antenna. *N*-glycolylneuraminic acid – extremely rare in humans, common to most animals – is abbreviated as "Ng".

The green code is a suggested alleviation for cases where exact linkages are not known, or do not matter.

A ⁴ Na ⁶⁻⁴	Na ⁶⁻⁴ Na ⁶⁻⁴	 ♦ ● ●
Na ³⁻⁴ Ng ³⁻³ F ⁶	Na ⁶⁻⁴ Na ⁶⁻⁴ F ⁶ < NaNaF >	
Na ⁶⁻⁴ (Na ⁶ -A ⁴ F ³)F ⁶ Na ⁶⁻⁴ Na ⁶ -LxF ⁶		

Multi-antennary glycans

A branch leading to three antennae can occur on either arm of an N-glycan. The two antennae ascending from the same mannose are set in square brackets. The square exclusively and immediately tells us that this branch is further branched. By that the two basic types of triantennary glycans are readily told apart.

The proglycan nomenclature reaches its limits here because the terms become lengthy and difficult to read, but then, what is the alternative? Even the strange tetra-sialylated triantennary glycan in bovine fetuin can be depicted. Note that the term $[(Na^{3-3}Na^6)Na^{6-4}]$ contains – inside the squrare brackets - round brackets signifying additional branching. As no other branching point is given, the root residue is GlcNAc.

[A ⁴ A ³]A ⁴ F ⁶ [AA]AF	A ⁴ [A ⁴ A ⁴] A[AA]	
Na ⁶⁻⁴ [Na ⁶⁻⁴ Na ⁶⁻⁴] Na[NaNa]	[Ng ³⁻⁴ Ng ³⁻³]Ng ³⁻⁴ F ⁶ [NgNg]Ng	
Na ³⁻⁴ [Na ³⁻⁴ Na ⁶⁻⁴] Na[NaNa] (major structure of fetuin)	Na ³⁻⁴ [(Na ⁶ Na ³⁻³)Na ⁶⁻⁴] (specialty of fetuin)	

<u>RULE 6: Two mannose-rooted antennae are put in square brackets. The order within brackets</u> <u>follows the "counter-clockwise" RULE 1".</u>

LacNAc repeats

The primary LacNAc disaccharide Gal β 1-4GlcNAc β 1- that is linked to a mannose can be further elongated by the addition of GlcNAc (in β 1-3 linkage to Gal), which usually is followed by the quick addition of Gal to arrive at another Gal β 1-4GlcNAc β 1-, or LacNAc unit.

A4	Galβ- 4 GlcNAcβ-2Manα-
Gn ³⁻⁴	GlcNAcβ-3Galβ- <mark>4</mark> GlcNAcβ-2Manα-
Ln ⁴	Galβ-4GlcNAcβ-3Galβ- <mark>4</mark> GlcNAcβ-2Manα-
Ln-Ln <mark>4</mark>	$Gal\beta - 4GlcNAc\beta - 3Gal\beta - 4GlcNAc\beta - 3Gal\beta - 4GlcNAc\beta - 2Man\alpha -$

The large and complex structures **N3.7.2B** in recombinant human erythropoietin [9], *aka* EPO will thus be written as:

[Na ³ -Ln ⁴ Na ³⁻⁴]Na ³⁻⁴		
	♦ ^{O-■}	

High-mannose and hybrid type N-glycans

The above defined rules can also be applied to high-mannose, *aka* oligomannosidic type glycans, but this only makes sense if the true structure of a glycan is known, e.g., if glycans are analyzed by porous graphitic carbon chromatography [10]. Only the mannoses that occur in addition to those on the common core have to be explicitly defined and this is realized by their linkage. In the upper line in the example below the number of linkage figures always is the number of mannose residues in addition to the three of the core.

ММ	000	MU	•
M ³ M	••••	(M ⁶ M ³)M ² aka Man6	
(M ⁶ M ³)M aka Man5	0	MM ²⁻²⁻² another Man5	
(M ²⁻⁶ M ³)M ²		(M ⁶ M ³)M ²⁻²	
(M ²⁻⁶ M ²⁻³)G	0,0 0,0 0,0 0,0 0,0 0,0	(M ⁶ M ³)G-G-G	
(M ⁶ M ³)Gn Man5Gn		M ³ A ⁴ Man4A*	

* In Man4A neither the linkage of the terminal mannose nor of the galactose are exactly determined. In some instances, this term may nevertheless be justified.

Plant and insect glycans

This arises from the authors' long occupation with glycoproteins from plants and insects, where core fucose occurs in α 1,3-linkage and where – in plants – a xylose is linked to the β -mannosyl residue of the core. These two peculiarities can easily be included as shown in the examples below.

MMXF ³ MMXF	MUXF ³ MUXF	
MGnX	(FA)(FA)XF ³ LaLaXF ³	
MMF ³ F ⁶ MMFF	M(AnF)F ³ F ⁶	

Specialties

The human brain contains sizable amounts of glycans with the "HNK-1" (from *human natural killer cells*) with sulfated glucuronic acid [11]. Annotating a structure like this requires some form of linear code and the addition of abbreviations for non-sugar substituents, in this case sulfate. Note that the hyphen binds the "su" to "Ga", which in turn is hyphenated to the "4" (or "3"), which stands for the regular antennary galactose.

The bladder protein uromodulin *aka* Tamm-Horsefall protein contains glycans with sulfated GalNAc and the Sd^a determinant, which harbors a branch on the galactose residue [12]. Based on the rules for Lewis determinants, we use a round bracket and a superscript hyphen with the linkage of the galactose residue.

Another peculiar structure is that with a Lewis X determinant in the bisecting position. With the rules established so-far, even such an exotic item can be named.

To facilitate deciphering of these terms, the abbreviations are also given with colors for the 6arm, 3-arm and the extension terms.

Ga ³⁻⁴ GnF ⁶ bi <mark>Ga³⁻⁴GnF⁶bi</mark>	su ³ -Ga ³⁻⁴ GnF ⁶ bi <mark>su³-Ga³⁻⁴GnF⁶bi</mark>	su s
M ³ Gn(AF)-bi M ³ Gn(AF)-bi M ³ GnLx-bi a brain glycan	(An ⁴ Na ³) ⁻⁴ su ⁴ -An ⁴ (An ⁴ Na ³) ⁻⁴ su ⁴ -An ⁴ in Uromodulin	□-0-□0 ♦ su- □-□0

Mosses contain structures with methyl groups [13], non-vertebrates contain numerous "unusual" and remarkable structural features such as methylation, sulfation, and zwitterionic non-sugar substituents, again glucuronic acid and often unusual architectures such as substituted core-fucose just as an example [14,15].



Another box of plethora (sic!) is opened by the highly unusual and diverse N-glycans of microalgae [16-18]. It would be possible to somehow describe also these structures, but for the time being, it does not appear to be a pressing need.

Comparison with existing abbreviation systems

Finally, and hopefully not too late, we shall compare the herein portrayed naming system with existing ones. There are several more or less frequently used systems of which a few occur more often shall be presented and compared.

The most stringent and rigorous way to annotated mass spec data is to give the sum formula without even trying to interpret it in terms of types of hexoses and co

Example for *m*/*z* = 2369.84 = Hex5HexNAc4Neu5Ac2dHex1

= H5N4S2F1 as e.g., in [19,20]

or more condensed

= 5_4_2_1 or 5421 as *e.g.*, in [21] and [22], respectively

or as in web.expasy.org/glycomod/

= (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₂ + (Man)₃(GlcNAc)₂

This annotation type strictly avoids any overinterpretation of data and should be the explicit starting point of any other naming or drawing exercise. Unfortunately, it neither transports structural information nor does it spoil the human eye. MS spectra evaluation software thus often

essentially neglects the overinterpretation problem and suggests a particular structure, where actually a range of isomers is possible as recently emphasized for the H5N4F1 composition [3].

proglycan	antibody	Ox	ford Fuc first		Oxford Fuc last		Elaborate Oxford	
	Ref.	Number of antennae	sialic acid linkage	Typ siali acid	e of c	galactose linkage	fucose linkage	arm location
Na ⁶⁻⁴ Na ⁶⁻⁴ F ⁶	proglycan	\checkmark	\checkmark	V	[\checkmark	\checkmark	n.a.
G2FS2	[25], a, b	(√)*	Х	Х	I	Х	х	n.a.
G2FS2 α(2,6)	[25], c	(√)*	\checkmark	Х	L.	Х	х	n.a.
FA2G2S2	[26], a	\checkmark	Х	Х	I	Х	х	n.a.
F(6)A2G2S(6)2	[27]	\checkmark	\checkmark	(x	:)	Х	(√)	n.a.
A2S2F	С	\checkmark	Х	Х	I	Х	х	n.a.
A2S2G2F ^{&)}	[23], d	\checkmark	Х	(v)	(x)	х	n.a.
A2G2FS2 ^{&)}	[28]	\checkmark	Х	(v	/)	(x)	х	n.a.
A ⁴ GnF ⁶	proglycan	\checkmark	n.a.	n.	a.	\checkmark	\checkmark	\checkmark
F(6)A2[6]G(4)1	[29]	\checkmark	n.a.	n.	a.	\checkmark	\checkmark	\checkmark

^{&)} Examples for the absence of a stringent logic for order of symbols

Commercial sources:

- a) ludger.com/product-catalogue/standards-controls
- b) aspariaglycomics.com/product-category/glycan-standard/
- c) agilent.com/en/product/biopharma-hplc-analysis/glycan-analysis
- d) Thermo Fisher Scientific BioPharma Finder 3.0

For the expression of ideas about structure, two systems are prominent:

- 1) The "**antibody glycan code**" counting the number of galactose residues for biantennary glycans only sufficient for antibodies.
- 2a) The "**Oxford code**" counting number of antennae and galactose residues with fucose first
- 2b) The "Oxford code" counting number of antennae and galactose residues with fucose later

3a/b) The elaborate "Oxford code". Somewhat event-related additions are used to exactly specify structures, *e.g.*, when alfa-Gal or *N*-glycolyl-neuraminic acid occur [23] or a large number of isomers are to be named [24].

The "antibody glycan code" is a perfectly fine convention in the field of – surprise - antibodies, in particular recombinant IgGs, where G2F actually stands for the isomer A⁴A⁴F⁶ and not for any of the 40-50 possible other isobars [3].

The "Oxford code" and its variants likewise have a definite raison d'être especially when rather large and only partially defined structures shall be named. *E.g.*, the term 3A2SF comprises a number of related structures that are most often not told apart by the analytical results [30,31].

When particular, exactly known N-glycan structures are to be described, the elaborate "Oxford code" is an option, but the proglycan system appears to be a big step ahead. The following table tries to give an overview of abbreviation systems. By no means does it claim to be comprehensive. The readers contribution to update and complete this table is highly encouraged.

Finally, we dare to devise a decision triangle that opposes the Oxford and proglycan systems with plain composition as the starting point and save haven.



List of possible "structure terms":

6-arm term

3 arm term

core extension

in both first (6-arm) and second (3-arm) position:

Gn	a GlcNAc linked to the invariant pentasaccharide core			
U	indicates no substituent of the β -mannose.			
А	Galactose linked to GlcNAc, linkage not specified			
A4 / A4	Galactose β1,4-linked to GlcNAc			
A ³ / A3	Galactose β1,3-linked to GlcNAc			
An ⁴ / An4	GalNAc β1,4-linked to GlcNAc			
(AF) / (A ⁴ F ³) /	(A4F3) / Lx Lewis X determinant			
(FA) / (F ⁴ A ³) /	(F4A3) / La Lewis A determinant			
F ²⁻³ / Lh ³	Blood group H determinant on type I chain			
F ²⁻⁴ / Lh ⁴	Blood group H determinant on type II chain			
La	Lewis A determinant			
Lh	Blood group H determinant, probably Lh ³ , maybe Lh ⁴			
Ln	LacNAc extension (Ln-, Ln ³ , Ln ⁴)			
Lx	Lewis X determinant			
Na	Sialic acid linked to Gal-GlcNAc, linkages not specified			
Na ⁶⁻⁴ / Na6-4	Neu5Ac-α2,6-Gal-1,4-GlcNAc-			
Na ³⁻³ / Na3-3	Neu5Ac-α2,3-Gal-1,3-GlcNAc-			
(Na ⁶ Na ³⁻³)	Neu5Ac- α 2,6(Neu5Ac- α 2,3-Gal-1,3)-GlcNAc- (as in fetuin)			

Possible in first (6-arm) position:

М	6-arm mannose of the invariant pentasaccharide core
M ³ / M3	a mannose in 3-linkage to 6-arm mannose of the core
M ⁶ / M6	a mannose in 6-linkage to 6-arm mannose of the core; rare
[GnGn]	GlcNAc-β1,6-(GlcNAc-β1,2-)Man
[A ⁴ A ⁴]	branched 6-arm with two Gal residues (in analogy to [GnGn])

Possible in second (3-arm) position:

Μ	3-arm mannose of the invariant pentasaccharide core
M ² / M2	a mannose in 2-linkage to the invariant pentasaccharide core
M ²⁻² / M2-2	two mannoses in series
[GnGn]	GlcNAc-β1,4-(GlcNAc-β1,2-)Man (in analogy to [GnGn])
$[A^4A^4]$	branched 3-arm with two Gal residues (in analogy to [GnGn])

<u>Possible in the extension positions:</u> – if present; fucosylation before bisection

F ⁶	/ F6	α 1,6-fucosylation of reducing end GlcNAc
F ³	/ F3	α 1,3-fucosylation of reducing end GlcNAc
F ³ F ⁶	/ F3F6	difucosylation of reducing end GlcNAc
Х		β 1,2-xylosylation of the β -mannose (written before fucosylation)
bi		indicates presence of a bisecting GlcNAc

A-bi	bisecting LacNAc
(AF)-bi	bisecting Lewis X

Order of extension terms:

[1] Xyl – [2] substituents of reducing GlcNAc (3 before 6) – [3] bisecting GlcNAc

Non-sugar substituents

su sulfate	
ac acetyl	
me methyl	
po phosphate	
pc phosphocholine	
pe phosphoethanola	amine

References:

1. Mehta AY, Cummings RD (2020) GlycoGlyph: a glycan visualizing, drawing and naming application. Bioinformatics 36 (11):3613-3614. doi:10.1093/bioinformatics/btaa190 2. Herget S, Ranzinger R, Maass K, Lieth CW (2008) GlycoCT-a unifying sequence format for carbohydrates. Carbohydr Res 343 (12):2162-2171. doi:10.1016/j.carres.2008.03.011 3. Helm J, Grunwald-Gruber C, Thader A, Urteil J, Fuhrer J, Stenitzer D, Maresch D, Neumann L, Pabst M, Altmann F (2021) Bisecting Lewis X in Hybrid-Type N-Glycans of Human Brain Revealed by Deep Structural Glycomics. Anal Chem 93 (45):15175-15182. doi:10.1021/acs.analchem.1c03793 4. Helm J, Hirtler L, Altmann F (2022) Towards Mapping of the Human Brain N-Glycome with Standardized Graphitic Carbon Chromatography. Biomolecules 12 (1). doi:10.3390/biom12010085 5. Gleeson PA, Schachter H (1983) Control of glycoprotein synthesis. J Biol Chem 258 (10):6162-6173 6. Staudacher E, Altmann F, Glossl J, Marz L, Schachter H, Kamerling JP, Hard K, Vliegenthart JF (1991) GDP-fucose: beta-N-acetylglucosamine (Fuc to (Fuc alpha 1----6GlcNAc)-Asn-peptide)alpha 1----3fucosyltransferase activity in honeybee (Apis mellifica) venom glands. The difucosylation of asparaginebound N-acetylglucosamine. Eur J Biochem 199 (3):745-751. doi:10.1111/j.1432-1033.1991.tb16179.x 7. Banin E, Neuberger Y, Altshuler Y, Halevi A, Inbar O, Nir DD, A. (2002) A Novel Linear Code® Nomenclature for Complex Carbohydrates. Trends in Glycoscience and Glycotechnology 14:127-137 8. Harvey DJ, Merry AH, Royle L, Campbell MP, Dwek RA, Rudd PM (2009) Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. Proteomics 9 (15):3796-3801. doi:10.1002/pmic.200900096 9. Hokke CH, Bergwerff AA, Van Dedem GW, Kamerling JP, Vliegenthart JF (1995) Structural analysis of

the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric N-acetyllactosamine units. Eur J Biochem 228 (3):981-1008. doi:10.1111/j.1432-1033.1995.tb20350.x

10. Pabst M, Grass J, Toegel S, Liebminger E, Strasser R, Altmann F (2012) Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors. Glycobiology 22 (3):389-399. doi:10.1093/glycob/cwr138

11. Voshol H, van Zuylen CW, Orberger G, Vliegenthart JF, Schachner M (1996) Structure of the HNK-1 carbohydrate epitope on bovine peripheral myelin glycoprotein P0. J Biol Chem 271 (38):22957-22960. doi:10.1074/jbc.271.38.22957

12. van Rooijen JJ, Kamerling JP, Vliegenthart JF (1998) Sulfated di-, tri- and tetraantennary N-glycans in human Tamm-Horsfall glycoprotein. Eur J Biochem 256 (2):471-487. doi:10.1046/j.1432-1327.1998.2560471.x

13. Stenitzer D, Mocsai R, Zechmeister H, Reski R, Decker EL, Altmann F (2022) O-methylated N-glycans Distinguish Mosses from Vascular Plants. Biomolecules 12 (1). doi:10.3390/biom12010136

14. Paschinger K, Wilson IBH (2019) Comparisons of N-glycans across invertebrate phyla. Parasitology 146 (14):1733-1742. doi:10.1017/S0031182019000398

15. Hykollari A, Paschinger K, Wilson IBH (2022) Negative-mode mass spectrometry in the analysis of invertebrate, fungal, and protist N-glycans. Mass Spectrom Rev 41 (6):945-963. doi:10.1002/mas.21693 16. Mocsai R, Blaukopf M, Svehla E, Kosma P, Altmann F (2020) The N-glycans of Chlorella sorokiniana and a related strain contain arabinose but have strikingly different structures. Glycobiology 30 (8):663-676. doi:10.1093/glycob/cwaa012

17. Mocsai R, Figl R, Sutzl L, Fluch S, Altmann F (2020) A first view on the unsuspected intragenus diversity of N-glycans in Chlorella microalgae. Plant J 103 (1):184-196. doi:10.1111/tpj.14718 18. Mocsai R, Kaehlig H, Blaukopf M, Stadlmann J, Kosma P, Altmann F (2021) The Structural Difference of Isobaric N-Glycans of Two Microalgae Samples Reveals Taxonomic Distance. Front Plant Sci 12:643249. doi:10.3389/fpls.2021.643249

19. Blochl C, Wang D, Madunic K, Lageveen-Kammeijer GSM, Huber CG, Wuhrer M, Zhang T (2021) Integrated N- and O-Glycomics of Acute Myeloid Leukemia (AML) Cell Lines. Cells 10 (11). doi:10.3390/cells10113058

20. Gaunitz S, Tjernberg LO, Schedin-Weiss S (2021) The N-glycan profile in cortex and hippocampus is altered in Alzheimer disease. J Neurochem 159 (2):292-304. doi:10.1111/jnc.15202

21. Lee J, Ha S, Kim M, Kim SW, Yun J, Ozcan S, Hwang H, Ji IJ, Yin D, Webster MJ, Shannon Weickert C, Kim JH, Yoo JS, Grimm R, Bahn S, Shin HS, An HJ (2020) Spatial and temporal diversity of glycome expression in mammalian brain. Proc Natl Acad Sci U S A 117 (46):28743-28753. doi:10.1073/pnas.2014207117

22. Barboza M, Solakyildirim K, Knotts TA, Luke J, Gareau MG, Raybould HE, Lebrilla CB (2021) Region-Specific Cell Membrane N-Glycome of Functional Mouse Brain Areas Revealed by nanoLC-MS Analysis. Mol Cell Proteomics 20:100130. doi:10.1016/j.mcpro.2021.100130

23. Fussl F, Trappe A, Carillo S, Jakes C, Bones J (2020) Comparative Elucidation of Cetuximab Heterogeneity on the Intact Protein Level by Cation Exchange Chromatography and Capillary Electrophoresis Coupled to Mass Spectrometry. Anal Chem 92 (7):5431-5438.

doi:10.1021/acs.analchem.0c00185

24. Abrahams JL, Campbell MP, Packer NH (2018) Building a PGC-LC-MS N-glycan retention library and elution mapping resource. Glycoconj J 35 (1):15-29. doi:10.1007/s10719-017-9793-4

25. Echeverria B, Etxebarria J, Ruiz N, Hernandez A, Calvo J, Haberger M, Reusch D, Reichardt NC (2015) Chemo-Enzymatic Synthesis of (13)C Labeled Complex N-Glycans As Internal Standards for the Absolute Glycan Quantification by Mass Spectrometry. Anal Chem 87 (22):11460-11467.

doi:10.1021/acs.analchem.5b03135

26. Planinc A, Bones J, Dejaegher B, Van Antwerpen P, Delporte C (2016) Glycan characterization of biopharmaceuticals: Updates and perspectives. Anal Chim Acta 921:13-27.

doi:10.1016/j.aca.2016.03.049

27. Mittermayr S, Bones J, Doherty M, Guttman A, Rudd PM (2011) Multiplexed analytical glycomics: rapid and confident IgG N-glycan structural elucidation. J Proteome Res 10 (8):3820-3829. doi:10.1021/pr200371s

28. Guile GR, Rudd PM, Wing DR, Prime SB, Dwek RA (1996) A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. Anal Biochem 240 (2):210-226. doi:10.1006/abio.1996.0351

29. Doherty M, Bones J, McLoughlin N, Telford JE, Harmon B, DeFelippis MR, Rudd PM (2013) An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture. Anal Biochem 442 (1):10-18. doi:10.1016/j.ab.2013.07.005

30. Clerc F, Reiding KR, Jansen BC, Kammeijer GS, Bondt A, Wuhrer M (2016) Human plasma protein N-glycosylation. Glycoconj J 33 (3):309-343. doi:10.1007/s10719-015-9626-2

31. Williams SE, Noel M, Lehoux S, Cetinbas M, Xavier RJ, Sadreyev RI, Scolnick EM, Smoller JW, Cummings RD, Mealer RG (2022) Mammalian brain glycoproteins exhibit diminished glycan complexity compared to other tissues. Nat Commun 13 (1):275. doi:10.1038/s41467-021-27781-9