



Introduction of a human- and keyboard-friendly N-glycan nomenclature

Friedrich Altmann^{*1}, Johannes Helm¹, Martin Pabst² and Johannes Stadlmann¹

Perspective

Open Access

Address:

¹Department of Chemistry, BOKU University, Vienna, Austria and

²Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

Email:

Friedrich Altmann^{*} - friedrich.altmann@boku.ac.at

^{*} Corresponding author

Keywords:

N-glycans; nomenclature; structural features

Beilstein J. Org. Chem. **2024**, *20*, 607–620.

<https://doi.org/10.3762/bjoc.20.53>

Received: 08 November 2023

Accepted: 27 February 2024

Published: 15 March 2024

Associate Editor: U. Westerlind



© 2024 Altmann et al.; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract

In the beginning was the word. But there were no words for N-glycans, at least, no simple words. Next to chemical formulas, the IUPAC code can be regarded as the best, most reliable and yet immediately comprehensible annotation of oligosaccharide structures of any type from any source. When it comes to N-glycans, the venerable IUPAC code has, however, been widely supplanted by highly simplified terms for N-glycans that count the number of antennae or certain components such as galactoses, sialic acids and fucoses and give only limited room for exact structure description. The highly illustrative – and fortunately now standardized – cartoon depictions gained much ground during the last years. By their very nature, cartoons can neither be written nor spoken. The underlying machine codes (e.g., GlycoCT, WURCS) are definitely not intended for direct use in human communication. So, one might feel the need for a simple, yet intelligible and precise system for alphanumeric descriptions of the hundreds and thousands of N-glycan structures. Here, we present a system that describes N-glycans by defining their terminal elements. To minimize redundancy and length of terms, the common elements of N-glycans are taken as granted. The preset reading order facilitates definition of positional isomers. The combination with elements of the condensed IUPAC code allows to describe even rather complex structural elements. Thus, this “proglycan” coding could be the missing link between drawn structures and software-oriented representations of N-glycan structures. On top, it may greatly facilitate keyboard-based mining for glycan substructures in glycan repositories.

Introduction

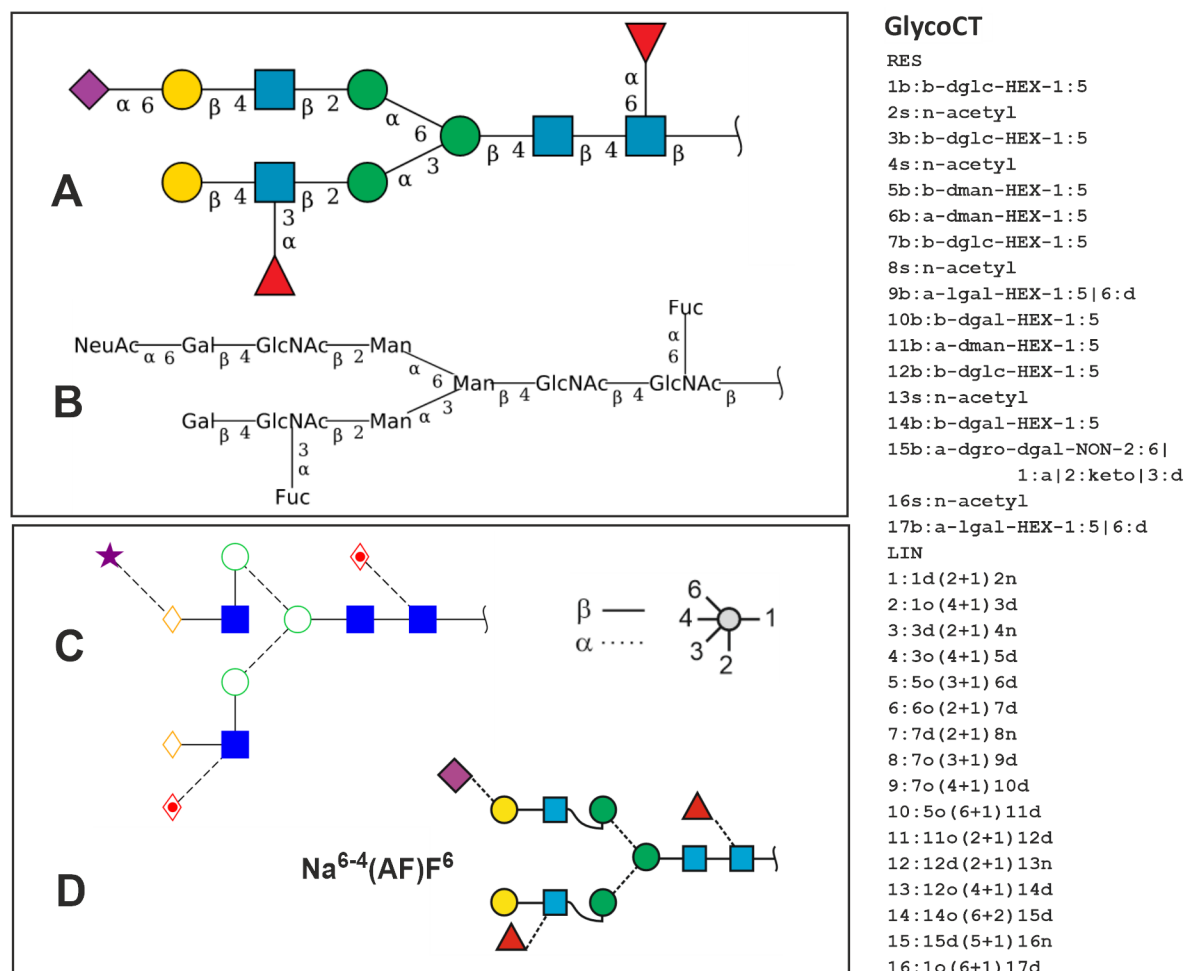
Virtually any article on protein glycosylation starts with imposing assurances about the biological significance of the various structures. This contrasts with the strange awkwardness when structures of interest, e.g., a possible biomarker, shall be

named. Uncompromising carbohydrate chemists will use chemical formulas drawn with ChemDraw, ChemSketch or alike. While this generates rather bulky pictures, chemists rely on the possibility to specify the chemical peculiarities deliberately

introduced by their sorcery. Biochemists and medical chemists, however, rather focus on the native glycan structure as grafted on the protein substrate by the biosynthetic machinery. Thus, biochemists could do with the much simpler IUPAC code [1] (Figure 1), which has the invaluable advantage of being an

alphanumeric code and which nowadays mostly dispenses use of Greek letters for annotating anomericity.

Look – as an example – at entry G75903TQ in the relevant database GlyTouCan (glytoucan.org) [7] for the structure that



IUPAC

NeuAc(a2-6)Gal(b1-4)GlcNAc(b1-2)Man(a1-6)[Fuc(a1-3)[Gal(b1-4)]GlcNAc(b1-2)Man(a1-3)]Man(b1-4)GlcNAc(b1-4)[Fuc(a1-6)]GlcNAc

or
 NeuAc α -6Gal β -4GlcNAc β -2Man β -6(Fuc α -3[Gal β -4]GlcNAc β -2Man α 1-3)Man β -4GlcNAc β -4(Fuc α 1-6)GlcNAc

WURCS

WURCS=2.0/6,12,11/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a1221m-1a_1-5][a2112h-1b_1-5]
 [Aad21122h-2a_2-6_5*NCC/3=O]/1-1-2-3-1-4-5-3-1-5-6-4/a4-b1_a6-l1_b4-c1_c3-d1_c6-h1_d2-e1_e3-f1_e4-g1_h2-i1_j4-j1_j6-k2

GWS

freeEnd--?b1D-GlcNAc,p(--4b1D-GlcNAc,p--4b1Man(--3a1Man--2b1D-GlcNAc,p(--3a1L-Fuc,p)--4b1D-Gal,p)--6a1Man--2b1D-GlcNAc,p--4b1D-Gal,p--6a2D-NeuAc,p)--6a1L-Fuc,p\$MONO,Und,0,0,freeEnd

Figure 1: Descriptions and depictions of a particular complex-type N-glycan structure found in human leukocytes [2]. A: SNFG depiction; B: IUPAC depictions; C: Oxford style; D: smart SNFG. Except for 1D, cartoons and codes were adapted from [3] (© GlyConnect v1.2.0, distributed under the terms of the Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0>). Cross-references in other data repositories are: GlyTouCan [4], GlyGen [5], ChEBI [6].

we will herein baptize **Na⁶⁻⁴(AF)F⁶** and ask yourself, how one could label an Eppendorf vial with the most condensed approved version of the IUPAC code (Figure 1). The graphical depictions are far more illustrative. On top, the “Symbol Nomenclature for Glycans” (SNFG) format, which was pioneered by the Consortium of Functional Glycomics (CFG), with its logical and appealing monosaccharide symbols has fortunately now become standard. Cartoons like this can be generated by GlycanBuilder [8,9], Glyconnect [10], GlyTouCan [7], GlycoWorkbench [9], SugarSketcher [11,12] or GlycoDraw [13] or by a universal graphics program such as, e.g., CorelDraw. The latter allows to merge the SNFG format with Oxford style thereby offering the particular advantage of encoding linkage position and anomericity in the cartoon itself obviating the need for letters, which require a minimal font size to remain readable. The resulting “smart”-SNFG thus transports structural information on a smaller space while essentially retaining the familiar SNFG appearance (Figure 1D). This is, however, just a shy suggestion with no ties to the actual topic of this treatise, even though others have already applied the same logic [14]. The cumbersome struggle for glycan depictions and related software has lately been portrayed in a quite enjoyable review [15].

A number of computer-readable formats has been developed of which GlycoCT [16] and WURCS [17] are currently the options of choice. None of these software friendly options is even remotely suitable for labeling of spectra or vials. Therefore, scientists in the huge field of N-glycan research have soon begun to use acronyms. One early adopted system was developed for the antibody field, where basically the number of galactose residues was annotated. In the most fully developed format, the presence of fucose and sialic acid is deliberately mentioned [18] (see also Table 1).

A more versatile set of annotation systems was built around the number of antennae. The term 3A2S denotes a triantennary N-glycan with two sialic acid residues. Obviously, this lean annotation does at first not specify linkage types and branch allocation. Therefore, add-ons were developed to allow a more precise description [19–24]. A structure probably representing the one shown in Figure 1 was described as FcA2G2FS [24]. While concise and informative, several details still remain undefined and we cannot see how the basic architecture of this system could be further developed.

To summarize, a human-, vial- and keyboard-friendly abbreviation system that nevertheless transports unequivocal structural information is still missing. The following chapters will pursue the question if the “proglycan” abbreviation system could be a useful tool, when N-glycan structures shall be talked about,

written on vials or in texts or put in tables. This 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 [25].

Our recent working with 40 isomeric N-glycans all composed of 5 hexoses, 4 HexNAcs and 1 fucose [26,27] reinforced our convictions that a simple and logical naming system is required and that the “proglycan” system could meet the demands. In the following chapters, aspects of how to name structures with different types of fucosylation, sialylation or antennary branching, bisecting GlcNAc, GalNAc, oligomannosidic structures and more will be discussed.

Discussion

The beginnings and the basics

In the 90ties during a guest visit in our lab, Prof. Harry Schachter from Toronto University taught us the term GnGn for the acceptor substrate of fucosyl transferases [28,29]. This “*reductio ad essentialia*” proved very helpful in our work with glycosyltransferases, e.g., insect hexosaminidase [30] and plant fucosyltransferase [31], which dealt with terminal modifications of always the same core. Years later, colleagues from in- and outside my institution engaged in “humanization” of plant N-glycans, which at first entailed removal of the plant-typical residues α 1,3-fucose and xylose [32,33]. Later, the glyco-engineering of plants with the introduction of human features such as β 1,4-galactose, branched and bisected glycans and sialylation required numerous experiments and a handy annotation of dozens of unusual structures for vials and quick notes on paper as well as papers [34–37]. This made us increasingly confident of the underlying principle of the “proglycan” nomenclature, an acronym derived from our then *nom de guerre* “protein-glycosylation analysis” group. By the way, glycan analysis also funneled in activities in the area of allergy diagnosis, where the term MUXF³ enjoys widespread use [38–40]. Finally, our work on the isomer-specific analysis of glycoproteins solidified our faith in a rather universal applicability of this system [26,27].

But now, back to the meaning of “GnGn”. Gn stands for GlcNAc and the two Gn-s symbolize the two terminal residues of a biantennary N-glycan (Figure 2). The GlcNAcs are attached to the common core pentasaccharide. No further definitions are required as the residues preceding the GlcNAcs are unambiguously determined by the biosynthetic pathway of N-glycans. Action of hexosaminidases would generate two isomers with just one GlcNAc and a terminal mannose residue. By defining the reading direction, we can easily distinguish these two isomers (Figure 2).

IUPAC	smart SNFG	proglycan										
$\text{GlcNAc}\beta\text{-2Man}\alpha\text{-6Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc}$ $\text{GlcNAc}\beta\text{-2Man}\alpha\text{-3}$		GnGn										
$\text{Man}\alpha\text{-6Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc}$ $\text{GlcNAc}\beta\text{-2Man}\alpha\text{-3}$		MGn (MGn)										
$\text{GlcNAc}\beta\text{-2Man}\alpha\text{-6Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc}$ $\text{Man}\alpha\text{-3}$		GnM (GnM)										
$\text{Man}\alpha\text{-6Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc}$ $\text{Man}\alpha\text{-3}$		MM										
<table><tr><td>Man M ●</td><td>Gal A ●</td><td>GalNAc An ■</td><td>Neu5Ac Na ◆</td><td>Xyl X ★</td></tr><tr><td>Fuc F ▲</td><td>Glc G ●</td><td>GlcNAc Gn ■</td><td>Neu5Gc Ng ◆</td><td>GlcA Ga ◆</td></tr></table>			Man M ●	Gal A ●	GalNAc An ■	Neu5Ac Na ◆	Xyl X ★	Fuc F ▲	Glc G ●	GlcNAc Gn ■	Neu5Gc Ng ◆	GlcA Ga ◆
Man M ●	Gal A ●	GalNAc An ■	Neu5Ac Na ◆	Xyl X ★								
Fuc F ▲	Glc G ●	GlcNAc Gn ■	Neu5Gc Ng ◆	GlcA Ga ◆								

Figure 2: The initial idea of describing an N-glycan just by its terminal residues. Reading the termini from top to bottom = counter-clockwise introduces arm specificity. The box at the bottom connects abbreviated monosaccharide names with proglycan shorthand names and SNFG symbols.

Galactosylation (including the alpha-Gal epitope) and sialylation

Often, the antennae are elongated by galactose whose one-letter code [41] is A. Added to the acceptor structure GnGn (Figure 2) this leads to three possible products if we only consider β 1,4-linkages. Rarely, however, the linkage is β 1,3. To account for this ambiguity, we use a number defining to define the point in question, in this case 3 or 4 (Figure 3). Writing these numbers as superscripts provides a more comprehensible structure to the term, but it is in no way crucial for the information content.

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 Gal(β 1-4)GlcNAc(β 1-2)Man(α 1-, respectively. Whenever we further elongate the sequence, we need to consider the Gal(β 1-...) 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 [26], whereby the two figures connected by a hyphen point at a disaccharide unit. As alpha-Gal is only known as linked to a beta-Gal residue, this lean annotation is unambiguous unless – very unexpectedly – a novel structural element turns up that makes this spelling invalid.

Sialic acids – usually – are likewise linked to the β -galactose. Sialic acids may occur in the 3- or in the 6-position of this Gal residues. Following the example of alpha-Gal residues, we can write a Neu5Ac(α 2-6)Gal(β 1-4)GlcNAc(β 1-2) antenna as Na⁶⁻⁴ while retaining the complete structural information (Figure 3). Rarely, a sialic acid residue is linked directly to GlcNAc as in bovine fetuin [42]. As this element was found on triantennary N-glycans only, we will deal with it in the relevant chapter below.

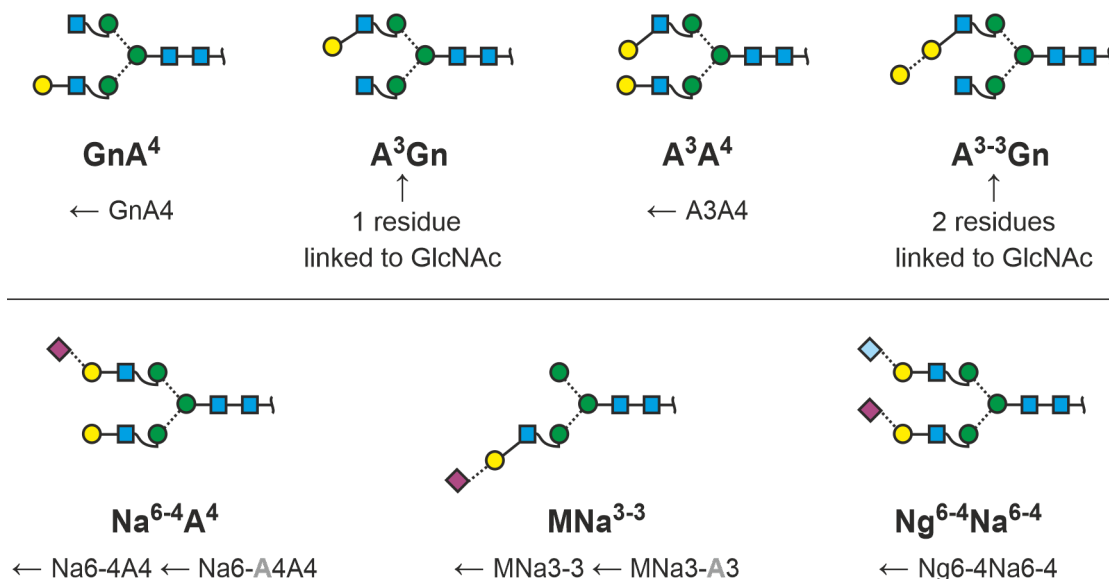


Figure 3: Annotation of antennae with galactose and sialic residues with consideration of linkage options. Note that the use of superscript format for linkages is recommended style but optional as it does not convey information.

Core-fucosylation, bisecting GlcNAc and xylose

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., A⁴A⁴F. In mammals, the core fucose is strictly always in α1,6. If you only work with mammalian samples you may content yourself with this simplification. However, as insect cells and plants are of relevance as biotechnological expression systems for pharmaceutical glycoproteins, we must consider that here fucose may also be found in the α1,3-position. Therefore, superscripts should be used to define the type of core fucose (Figure 4).

A structural element frequently occurring in mammalian glycoproteins, e.g., in immunoglobulins is the bisecting GlcNAc. This shall be added at the end of the abbreviation string by two lower case letters “bi”. In plants and some non-vertebrate animals [43,44], the β-mannosyl residues may instead be decorated with a xylose residue, for which the letter X is added after the term for the 3-arm (Figure 4).

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

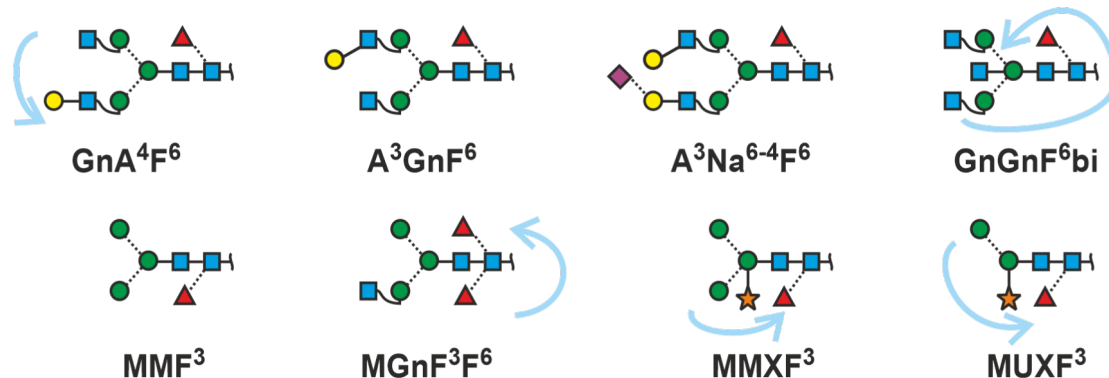


Figure 4: Core substituents are listed in counter-clockwise order.

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 transforms this term to (A⁴F³). A LeA structure would be (F⁴A³) because by default we read the structure counter-clockwise. However, we want to be as elegant and concise as possible. Are we losing any information when omitting the superscripts? Unless novel structural peculiarities of N-glycans emerge, the answer is no. Therefore, we simply write (AF) and (FA) (Figure 5).

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 [41] – we write F²⁻⁴A⁴. We – again – can save one character by omitting the A to arrive at: F²⁻⁴ or F²⁻³ (Figure 5). Why not just F⁴ or F³? Because we must not ignore the galactose. “F⁴” would be a Fuc(α 1-4)GlcNAc substructure, which does not exist.

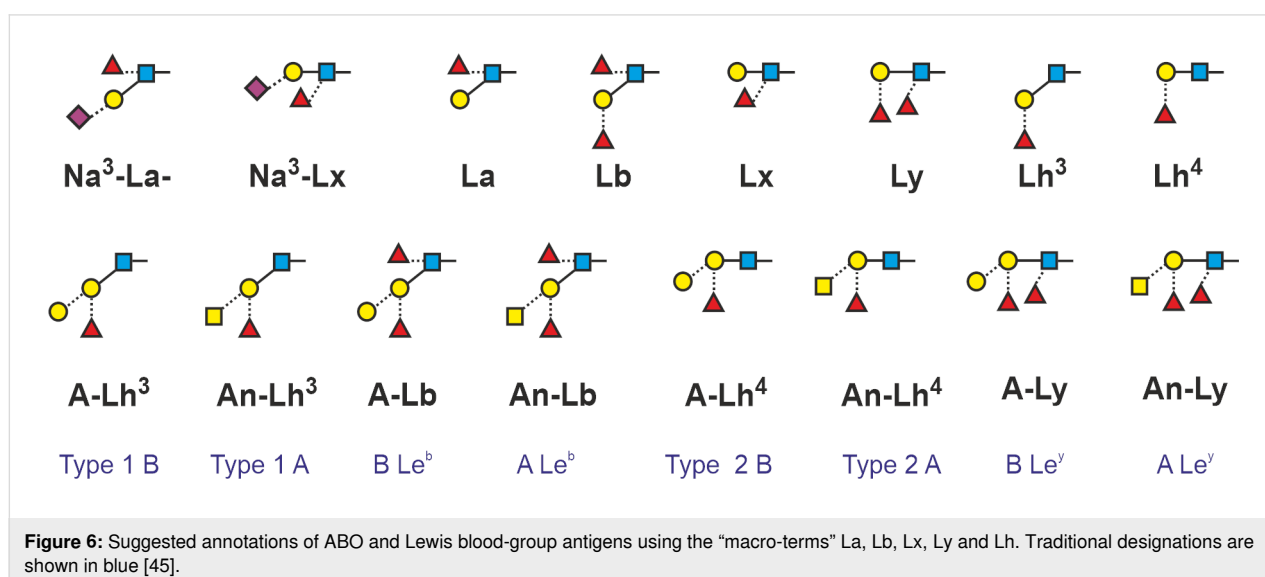
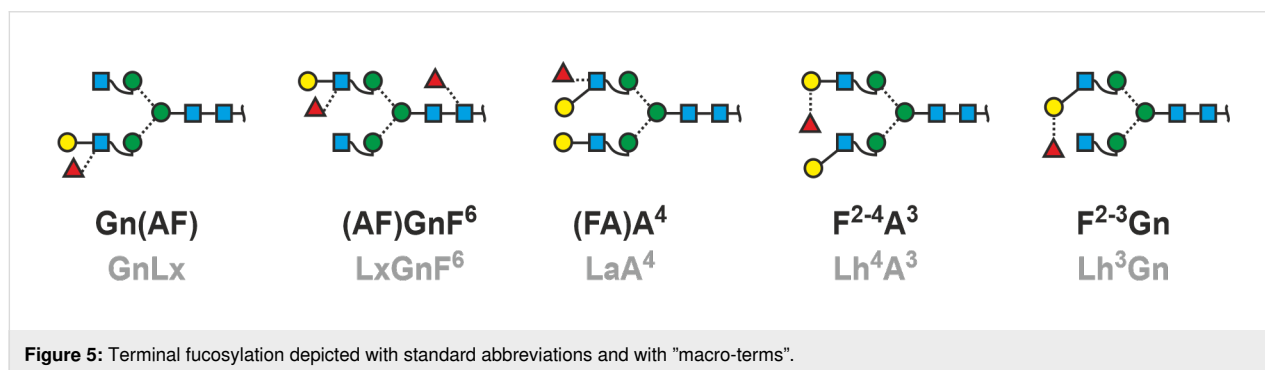
In a rush of boldness, we felt tempted to introduce “macro-terms” for relevant structural elements. The LeX and LeA deter-

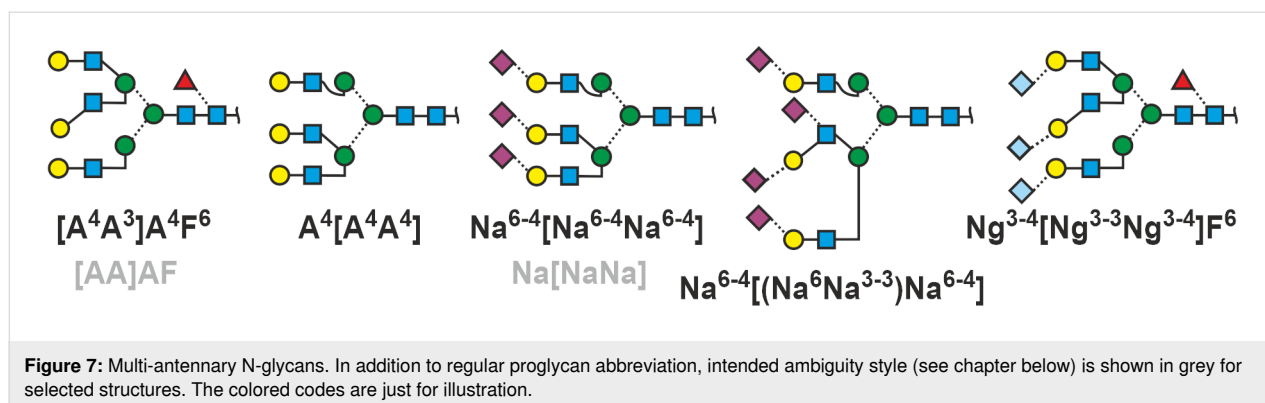
minants (AF) and (FA) could then be pointed up as Lx and La (Figure 5, grey terms). In fact, this logic can be the key for depicting the entire range of Lewis and blood-group antigens as suggested in Figure 6.

Multi-antennary glycans

A branch resulting in a triantennary structure can occur on either arm of an N-glycan. The two antennae ascending from the same mannose are set in square brackets (Figure 7). 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 tetrasialylated triantennary glycan in bovine fetuin can be depicted. Note that the term [(Na³⁻³Na⁶)Na⁶⁻⁴] contains – inside the square brackets – round brackets signifying additional branching. As no other branching point is given, the root residue is GlcNAc.





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- (= LacNAc) unit. The following annotation scheme is a suggestion that may be substituted by a better option. Such LacNAc repeats occur in recombinant erythropoietin [46] where the root galactose was found in 4-linkage. To allow for the occurrence of 3-linked galactose, the superscript “4” is part of the term used here. The large and complex structures **N3.7.2B** in recombinant human erythropoietin [46], aka EPO will thus be written as shown in Figure 8.

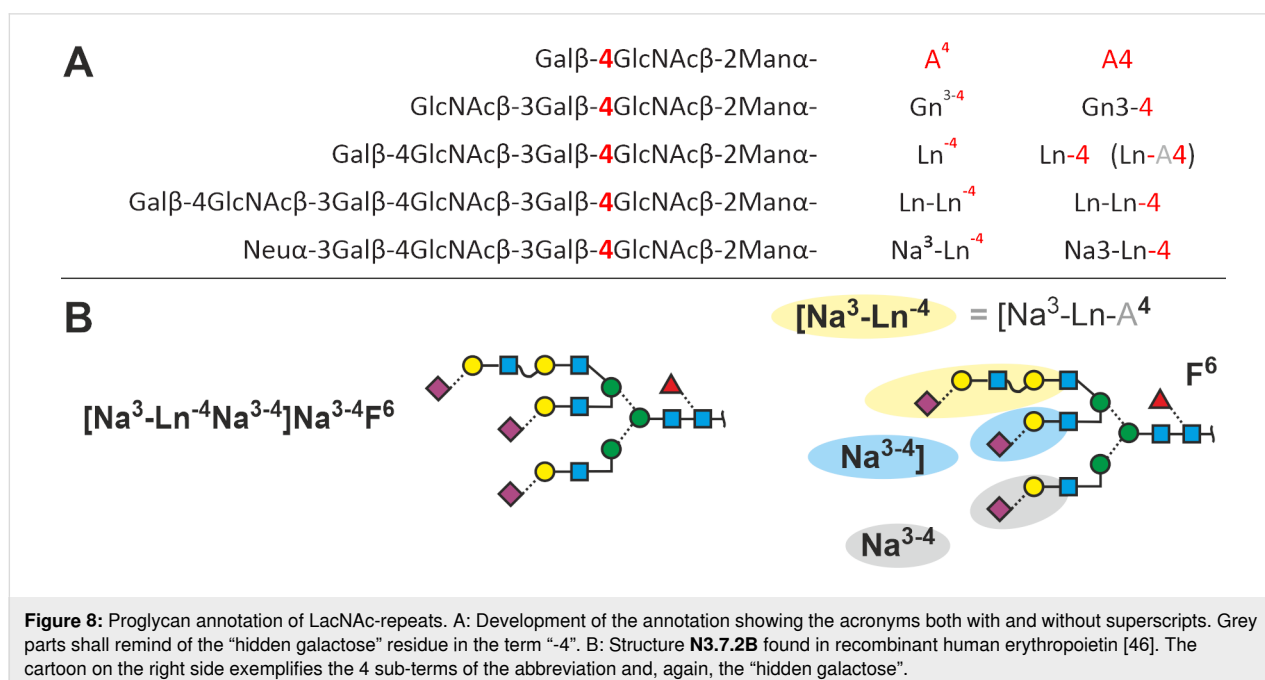
Special structures in brain and bladder

The human brain contains sizable amounts of glycans with the “HNK-1” (from *human natural killer cells*) with sulfated

glucuronic acid [47]. 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 (Figure 9).

Another peculiar structure is that with a Lewis X determinant in the bisecting position [26]. With the rules established so far, even such an exotic item can be named (Figure 9).

The bladder protein uromodulin aka Tamm-Horsfall protein contains glycans with sulfated GalNAc and the Sd^a determinant, which harbors a branch on the galactose residue [48,49]. Based on the rules for Lewis determinants, we use a round bracket embracing the two terminal residues, which do need linkage specifiers as GalNAc has only ever been found in β 1-4 linkage.



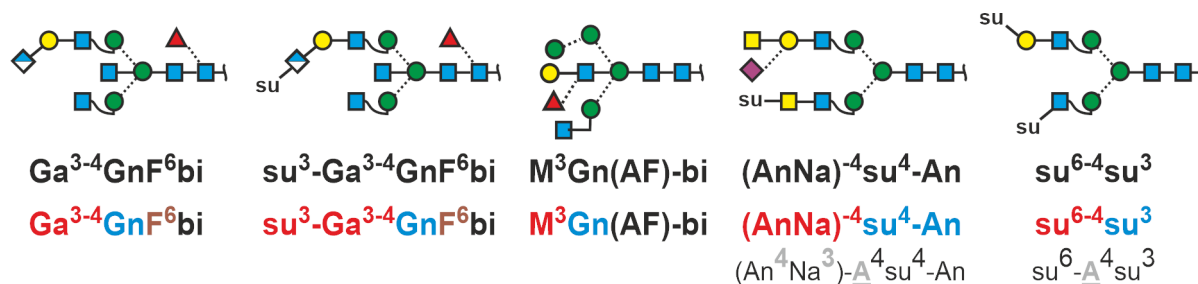


Figure 9: Proglycan codes for HNK-1 structures, a brain glycan with substituted bisecting GlcNAc, a sulfated SDA antigen containing glycan from uromodulin and a hypothetical sulfated N-glycan. The colored codes are just for illustration. The bottom line reveals the hidden, invariable information.

Neu5Ac has no other choice than the 3-position. The two residues are linked to a galactose, hence a hyphen and the pertinent linkage (Figure 9). The letter A itself is hidden as in sialylated antennae (Figure 3).

Finally, a hypothetical N-glycan shall exemplify the annotation of different sequences of sulfated antennae. To facilitate deciphering of these terms, the abbreviations are also given with colors for the 6-arm, 3-arm, and the first extension term (Figure 9).

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 [50]. 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. Noteworthy, the number of numeric figures gives the number of mannose residues in addition to the three of the core. The example $(\text{M}^6\text{M}^3)\text{M}^2$ contains three numbers and hence describes a hexasaccharide (Figure 10).

Plant and insect glycans

This chapter arises from the authors' long involvement with glycoproteins from plants and insects, where core fucose occurs in $\alpha 1,3$ -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 (Figure 11). Mosses contain structures with methyl groups [51], non-vertebrates contain numerous “unusual” and remarkable structural features such as methylation, sulfation, and zwitterionic non-

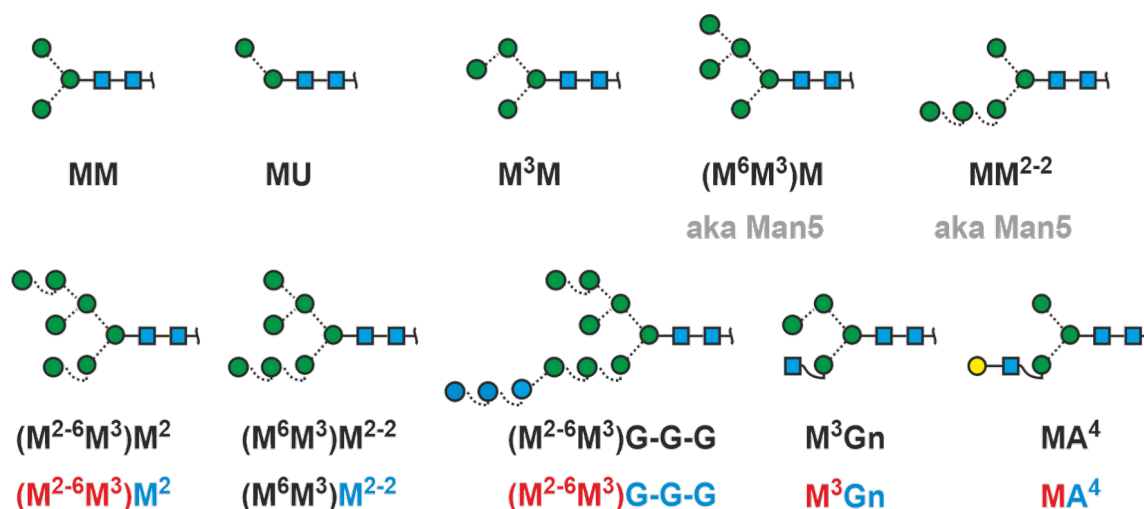


Figure 10: Proglycan annotation of oligomannosidic and some hybrid-type N-glycans.

sugar substituents, again glucuronic acid and often unusual architectures such as substituted core-fucose just as an example [44,52]. While the methylated moss glycans are accessible to the proglycan systems (Figure 11) the extremely diverse and unusual structures found in invertebrates or algae demand IUPAC code or graphical representations [44,53–55].

Moss N-glycans are included here to demonstrate the ability of the proglycan system to annotate nonsugar substituents.

Handling of ambiguities

A disadvantage of the herein described annotation scheme is its unambiguity. A given string stands for one particular structure. All too often, groups of isomers are to be addressed or the exact structure from a group of isomers cannot be defined. This chapter refrains from defining an ultimate solution to this dilemma as we think it is not yet time to impose another layer of complexity onto the herein proposed model. On top, automatic converters will encounter difficulties with such side rules. Nevertheless, for “domestic needs” shall four options be presented (Scheme 1).

Attempt at an application to O-glycans

The short, but varied core-region of O-glycans do not lend themselves to the proglycan logic as much as N-glycans. However, its application is possible as demonstrated in an article on chondrocyte glycans [56]. The branches to the core structures can be dealt with in much the same way as N-glycan antennae. It appears, however, that the particular core type has to be specified as shown in Figure 12. The very simple and frequently

occurring O-glycans T-, Tn- and Sialyl-Tn-antigen may be exempted from such attempts. Likewise, will the rarer core types be left out for the time being.

Comparison with existing abbreviation systems

Finally, we shall compare the herein portrayed naming system with existing ones. There are several systems of which a few occur more often. The most stringent and rigorous way to annotated mass spectrometry data is to give the sum formula without even trying to interpret it in terms of types of hexoses, N-acetylhexosamines (HexNAcs) etc.

Example for $m/z = 2369.84$:

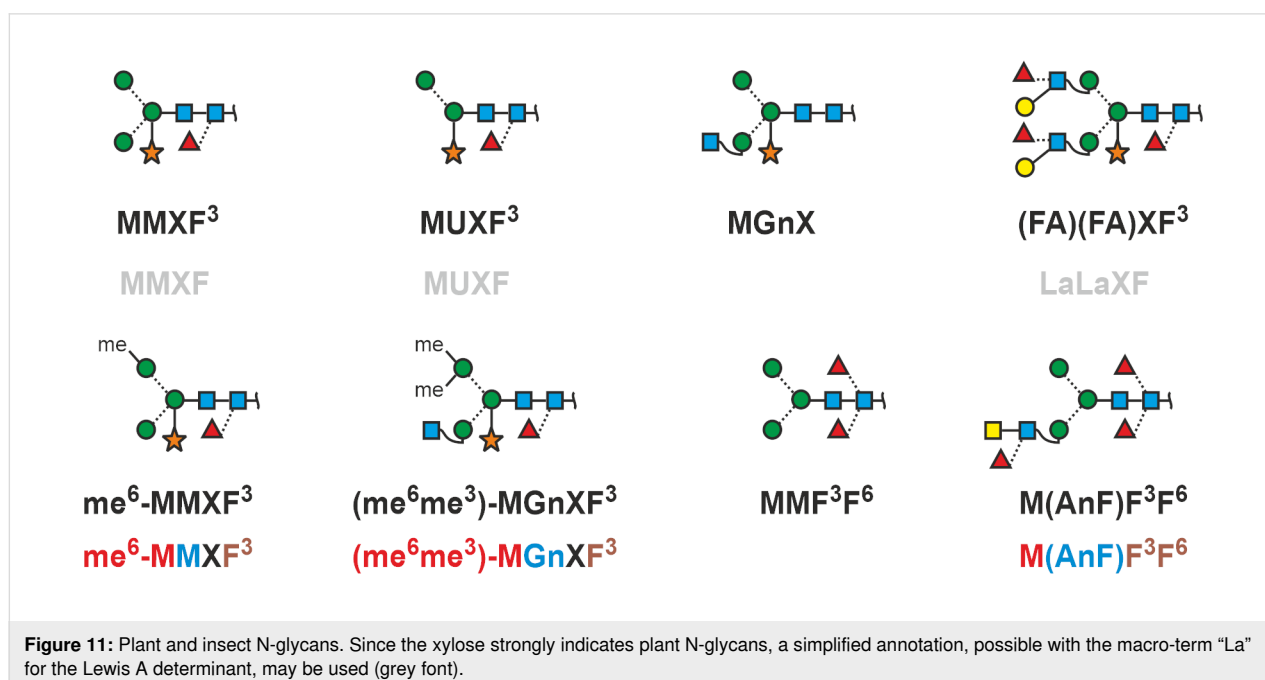
Hex5HexNAc4Neu5Ac2dHex1,

or H5N4S2F1 as, e.g., in [57–59],

or 5_4_2_1 or 5421 as, e.g., in [60] and [61], respectively,

or as in web.expasy.org/glycomod/ (Hex)₂ (HexNAc)₂ (Deoxyhexose)₁ (NeuAc)₂ + (Man)₃(GlcNAc)₂.

These annotation types strictly avoid any overinterpretation of data and should be the explicit starting point of any other naming or drawing exercise. Unfortunately, they neither transport structural information nor do they spoil the human eye. MS spectra evaluation software thus often essentially neglects the overinterpretation problem and suggests a particular structure,



Option 1 - omission of linkage numbers

AGnF could be A⁴GnF⁶ and/or A³GnF⁶ and even A⁴GnF³ and/or A³GnF³ --- 3 terminal residues
 NaGn could be Na⁶⁻⁴Gn, Na³⁻⁴Gn and/or Na³⁻³Gn --- 2 terminal residues
 NaNaA could be any triantennary structure with 2 sialic acids — 3 terminal residues
 MUXF will be MUXF³ or – very theoretically – could be MUXF⁶. --- 3 terminal residues

Option 2 - question marks

Na6-?A?F6 (equals: Na^{6-?}A^{?F}⁶)

A⁴(A^{?F}[?]) will cover glycans with both LeA and LeX structures.

An alternative could be the inclusive term "Le" for both LeA and LeX structures ("La" and "Lx").

This would be a space-saving solution for the notorious problem of ill-defined antennary fucose (Lewis) determinants

Option 3 - listing of alternative structures

Topological alternatives could be listed in the form of GnA⁴ / A⁴Gn.

This solution has the advantage that – as with all other proglycan codes – de-formatting will not cause a loss of information for the fully defined structures.

Scheme 1: Suggestions for the depiction of incompletely defined N-glycan structures.

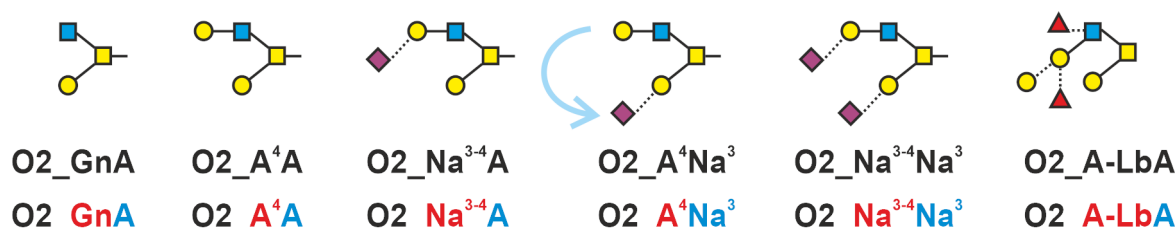


Figure 12: Provisional outline of how O-glycans with the frequent core 2 structure could be named with the proglycan logic.

where actually a range of isomers is possible as recently elaborated for the H5N4F1 composition [26]. A few more prominent alphanumerical codes were developed and shall be described and compared in the following chapter.

The “antibody glycan code” is a perfectly fine convention in the field of – surprise – antibodies, in particular recombinant IgGs, where G2F [62] stands for the therein relevant isomer A⁴A⁴F⁶ and not for any of the 40–50 possible other isobars [26].

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 [57–59].

When particular, exactly known N-glycan structures are to be described, the elaborate “Oxford code” is an option, but the

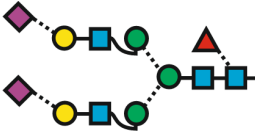
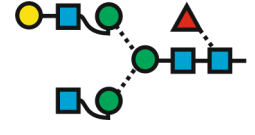
proglycan system appears to be a big step ahead. The following table (Table 1) tries to give an overview of abbreviation systems. By no means does it claim to be comprehensive. The readers' contribution to update and complete this table is highly encouraged.

The proglycan code and the glyco-IT world

Because of its stringent logic, the proglycan codes could be generated from and converted into any other annotation format. Writing appropriate code can be a challenging task for master students. The following deliberation suggests that such an endeavor could well be of practical importance.

Developed to be a human and keyboard-friendly alternative to machine codes, the proglycan code nevertheless qualifies as a valuable component of glycan databases. A simple text search can retrieve all N-glycan structures with a particular terminal motif with a uniquely low technical threshold both for the user

Table 1: Examples of N-glycan abbreviation systems with assessment of the respective abilities.

	reference ^a	number of antennae	sialic acid linkage	type of sialic acid	galactose linkage	fucose linkage	arm location
							
Na ⁶⁻⁴ Na ⁶⁻⁴ F ⁶	proglycan	✓	✓	✓	✓	✓	n.a.
“antibody glycan code” counting the number of galactose residues for biantennary glycans only							
G2FS2	[18,63,64]	(✓)*	x	x	x	x	n.a.
G2FS2α(2,6)	[18,65]	(✓)*	✓	x	x	x	n.a.
“Oxford code” counting number of antennae, galactoses and sialic acids with fucose first							
FA2G2S2	[23,63]	✓	x	(✓)	x	x	n.a.
F(6)A2G2S(6)2	[22]	✓	✓	(✓)	x	(✓)	n.a.
“Oxford code” counting number of antennae, galactoses and sialic acids with fucose last							
A2S2F	[65]	✓	x	(✓)	x	x	n.a.
A2S2G2F ^b	[20,66]	✓	x	✓	(x)	x	n.a.
A2G2FS2 ^b	[21]	✓	x	(✓)	(x)	x	n.a.
+ bisect. GlcNAc A2G2FBS2(6)	[67]	✓	✓	(✓)	(x)	x	n.a.
proglycan							
Na ⁶⁻⁴ Na ⁶⁻⁴ F ⁶	proglycan	✓	✓	✓	✓	✓	n.a.
							
A ⁴ GnF ⁶	proglycan	✓	n.a.	n.a.	✓	✓	✓
“Elaborate Oxford code”^c							
F(6)A2[6]G(4)1	[19]	✓	n.a.	n.a.	✓	✓	✓

^aReferences are examples without claim of completeness. ^bOrder of symbols not fixed. ^cSomehow event-related additions are used to exactly specify structures, e.g., when α-Gal or N-glycolyl-neuraminic acid occur [20] or a large number of isomers are to be named [68].

and the software developer. Proglycan codes inherently would be generated for fully defined structures only. Thereby, they would *en passant* act as a redundancy filter hiding the innumerable incompletely characterized entries.

Perspective

This exposition is not intended to grimly wipe away any other type of annotation. We imagine that it can be of value wherever defined structures shall be named and written down. With increasing uncertainty, any of the systems mentioned in Table 1 may be preferred up to the point where a mere listing of the constituent monosaccharides (e.g., in the form of H5N4F2S1) might be considered as the most suitable annotation. Finally, and for the years to come, a platform for suggestions, updates and related information can be found at <https://www.proglycan.info>.

Supporting Information

Supporting Information File 1

Possible structure terms.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-53-S1.pdf>]

Author Contributions

Friedrich Altmann: conceptualization; writing – original draft; writing – review & editing. Johannes Helm: data curation; validation; writing – original draft; writing – review & editing. Martin Pabst: data curation; writing – original draft. Johannes Stadlmann: conceptualization; formal analysis; writing – original draft.

ORCID® iDs

Friedrich Altmann - <https://orcid.org/0000-0002-0112-7877>Martin Pabst - <https://orcid.org/0000-0001-9897-0723>Johannes Stadlmann - <https://orcid.org/0000-0001-5693-6690>

Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

References

- Sharon, N. *Eur. J. Biochem.* **1986**, *159*, 1–6. doi:10.1111/j.1432-1033.1986.tb09825.x
- Nyman, T. A.; Kalkkinen, N.; Töölö, H.; Helin, J. *Eur. J. Biochem.* **1998**, *253*, 485–493. doi:10.1046/j.1432-1327.1998.2530485.x
- Complex glycan structure #2554; GlyConnect, v1.2.0. <https://glyconnect.expasy.org/all/structures/2554> (accessed Feb 27, 2024).
- International Glycan Structure Repository. GlyTouCan; Accession Number=G75903TQ. <https://glytoucan.org/Structures/Glycans/G75903TQ> (accessed Feb 27, 2024).
- GlyGen: Computational and Informatics Resources for Glycoscience. GlyGen, NIH Glycoscience Common Fund; <https://www.glycogen.org/details-for-glycan/G75903TQ> <https://glygen.org/glycan/G75903TQ>.
- Chemical Entities of Biological Interest (ChEBI); CHEBI glycan Nr. 157583. <https://www.ebi.ac.uk/chebi/displayAutoXrefs.do?chebiid=CHEBI:157583> (accessed Feb 27, 2024).
- Aoki-Kinoshita, K.; Agravat, S.; Aoki, N. P.; Arpinar, S.; Cummings, R. D.; Fujita, A.; Fujita, N.; Hart, G. M.; Haslam, S. M.; Kawasaki, T.; Matsubara, M.; Moreman, K. W.; Okuda, S.; Pierce, M.; Ranzinger, R.; Shikanai, T.; Shinmachi, D.; Solovieva, E.; Suzuki, Y.; Tsuchiya, S.; Yamada, I.; York, W. S.; Zaia, J.; Narimatsu, H. *Nucleic Acids Res.* **2016**, *44*, D1237–D1242. doi:10.1093/nar/gkv1041
- Tsuchiya, S.; Aoki, N. P.; Shinmachi, D.; Matsubara, M.; Yamada, I.; Aoki-Kinoshita, K. F.; Narimatsu, H. *Carbohydr. Res.* **2017**, *445*, 104–116. doi:10.1016/j.carres.2017.04.015
- Damerell, D.; Ceroni, A.; Maass, K.; Ranzinger, R.; Dell, A.; Haslam, S. M. *Biol. Chem.* **2012**, *393*, 1357–1362. doi:10.1515/hsz-2012-0135
- Alloci, D.; Mariethoz, J.; Gastaldello, A.; Gasteiger, E.; Karlsson, N. G.; Kolarich, D.; Packer, N. H.; Lisacek, F. *J. Proteome Res.* **2019**, *18*, 664–677. doi:10.1021/acs.jproteome.8b00766
- Alloci, D.; Suchánková, P.; Costa, R.; Hory, N.; Mariethoz, J.; Vařeková, R.; Toukach, P.; Lisacek, F. *Molecules* **2018**, *23*, 3206. doi:10.3390/molecules23123206
- Yamada, I.; Shiota, M.; Shinmachi, D.; Ono, T.; Tsuchiya, S.; Hosoda, M.; Fujita, A.; Aoki, N. P.; Watanabe, Y.; Fujita, N.; Angata, K.; Kajii, H.; Narimatsu, H.; Okuda, S.; Aoki-Kinoshita, K. F. *Nat. Methods* **2020**, *17*, 649–650. doi:10.1038/s41592-020-0879-8
- Lundström, J.; Urban, J.; Thomès, L.; Bojar, D. *Glycobiology* **2023**, *33*, 927–934. doi:10.1093/glycob/cwad063
- Manz, C.; Grabarics, M.; Hoberg, F.; Pugini, M.; Stuckmann, A.; Struwe, W. B.; Pagel, K. *Analyst* **2019**, *144*, 5292–5298. doi:10.1039/c9an00937j
- Lal, K.; Bermeo, R.; Perez, S. *Beilstein J. Org. Chem.* **2020**, *16*, 2448–2468. doi:10.3762/bjoc.16.199
- Herget, S.; Ranzinger, R.; Maass, K.; Lieth, C.-W. v. d. *Carbohydr. Res.* **2008**, *343*, 2162–2171. doi:10.1016/j.carres.2008.03.011
- Tanaka, K.; Aoki-Kinoshita, K. F.; Kotera, M.; Sawaki, H.; Tsuchiya, S.; Fujita, N.; Shikanai, T.; Kato, M.; Kawano, S.; Yamada, I.; Narimatsu, H. *J. Chem. Inf. Model.* **2014**, *54*, 1558–1566. doi:10.1021/ci400571e
- Echeverria, B.; Etxebarria, J.; Ruiz, N.; Hernandez, Á.; Calvo, J.; Habeger, M.; Reusch, D.; Reichardt, N.-C. *Anal. Chem. (Washington, DC, U. S.)* **2015**, *87*, 11460–11467. doi:10.1021/acs.analchem.5b03135
- Doherty, M.; Bones, J.; McLoughlin, N.; Telford, J. E.; Harmon, B.; DeFelippis, M. R.; Rudd, P. M. *Anal. Biochem.* **2013**, *442*, 10–18. doi:10.1016/j.ab.2013.07.005
- Füssli, F.; Trappe, A.; Carillo, S.; Jakes, C.; Bones, J. *Anal. Chem. (Washington, DC, U. S.)* **2020**, *92*, 5431–5438. doi:10.1021/acs.analchem.0c00185
- Guile, G. R.; Rudd, P. M.; Wing, D. R.; Prime, S. B.; Dwek, R. A. *Anal. Biochem.* **1996**, *240*, 210–226. doi:10.1006/abio.1996.0351
- Mittermayr, S.; Bones, J.; Doherty, M.; Guttman, A.; Rudd, P. M. *J. Proteome Res.* **2011**, *10*, 3820–3829. doi:10.1021/pr200371s
- Planinc, A.; Bones, J.; Dejaegher, B.; Van Antwerpen, P.; Delporte, C. *Anal. Chim. Acta* **2016**, *921*, 13–27. doi:10.1016/j.aca.2016.03.049
- Royle, L.; Roos, A.; Harvey, D. J.; Wormald, M. R.; Van Gijlswijk-Janssen, D.; Redwan, E.-R. M.; Wilson, I. A.; Doha, M. R.; Dwek, R. A.; Rudd, P. M. *J. Biol. Chem.* **2003**, *278*, 20140–20153. doi:10.1074/jbc.m301436200
- Platts-Mills, T. A.; Hilger, C.; Jappe, U.; van Hage, M.; Gadermaier, G.; Spillner, E.; Lidholm, J.; Keshavarz, B.; Aalberse, R. C.; van Ree, R.; Goodman, R. E.; Pomés, A. *Allergy (Oxford, U. K.)* **2021**, *76*, 2383–2394. doi:10.1111/all.14802
- Helm, J.; Grünwald-Gruber, C.; Thader, A.; Urteil, J.; Führer, J.; Stenitzer, D.; Maresch, D.; Neumann, L.; Pabst, M.; Altmann, F. *Anal. Chem. (Washington, DC, U. S.)* **2021**, *93*, 15175–15182. doi:10.1021/acs.analchem.1c03793
- Helm, J.; Hirtler, L.; Altmann, F. *Biomolecules* **2022**, *12*, 85. doi:10.3390/biom12010085
- Gleeson, P. A.; Schachter, H. *J. Biol. Chem.* **1983**, *258*, 6162–6173. doi:10.1016/s0021-9258(18)32387-1
- Staudacher, E.; Altmann, F.; Glössl, J.; März, L.; Schachter, H.; Kamerling, J. P.; Hård, K.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1991**, *199*, 745–751. doi:10.1111/j.1432-1033.1991.tb16179.x
- Altmann, F.; Schwihla, H.; Staudacher, E.; Glössl, J.; März, L. *J. Biol. Chem.* **1995**, *270*, 17344–17349. doi:10.1074/jbc.270.29.17344
- Leiter, H.; Mucha, J.; Staudacher, E.; Grimm, R.; Glössl, J.; Altmann, F. *J. Biol. Chem.* **1999**, *274*, 21830–21839. doi:10.1074/jbc.274.31.21830
- Koprivova, A.; Stemmer, C.; Altmann, F.; Hoffmann, A.; Kopriva, S.; Gorr, G.; Reski, R.; Decker, E. L. *Plant Biotechnol. J.* **2004**, *2*, 517–523. doi:10.1111/j.1467-7652.2004.00100.x
- Strasser, R.; Altmann, F.; Mach, L.; Glössl, J.; Steinkellner, H. *FEBS Lett.* **2004**, *561*, 132–136. doi:10.1016/s0014-5793(04)00150-4
- Strasser, R.; Castilho, A.; Stadlmann, J.; Kunert, R.; Quendler, H.; Gättinger, P.; Jez, J.; Rademacher, T.; Altmann, F.; Mach, L.; Steinkellner, H. *J. Biol. Chem.* **2009**, *284*, 20479–20485. doi:10.1074/jbc.m109.014126
- Castilho, A.; Gättinger, P.; Grass, J.; Jez, J.; Pabst, M.; Altmann, F.; Gorfer, M.; Strasser, R.; Steinkellner, H. *Glycobiology* **2011**, *21*, 813–823. doi:10.1093/glycob/cwr009

36. Castilho, A.; Gruber, C.; Thader, A.; Oostenbrink, C.; Pechlaner, M.; Steinkellner, H.; Altmann, F. *mAbs* **2015**, *7*, 863–870. doi:10.1080/19420862.2015.1053683
37. Castilho, A.; Strasser, R.; Stadlmann, J.; Grass, J.; Jez, J.; Gattlinger, P.; Kunert, R.; Quendler, H.; Pabst, M.; Leonard, R.; Altmann, F.; Steinkellner, H. *J. Biol. Chem.* **2010**, *285*, 15923–15930. doi:10.1074/jbc.M109.088401
38. Holzweber, F.; Svehla, E.; Fellner, W.; Dalik, T.; Stubler, S.; Hemmer, W.; Altmann, F. *Allergy (Oxford, U. K.)* **2013**, *68*, 1269–1277. doi:10.1111/all.12229
39. Dramburg, S.; Hilger, C.; Santos, A. F.; de las Vecillas, L.; Aalberse, R. C.; Acevedo, N.; Aglas, L.; Altmann, F.; Arruda, K. L.; Asero, R.; Ballmer-Weber, B.; Barber, D.; Beyer, K.; Biedermann, T.; Bilo, M. B.; Blank, S.; Bosshard, P. P.; Breiteneder, H.; Brough, H. A.; Bublin, M.; Campbell, D.; Caraballo, L.; Caubet, J. C.; Celi, G.; Chapman, M. D.; Chruszcz, M.; Custovic, A.; Czolk, R.; Davies, J.; Douladiris, N.; Eberlein, B.; Ebisawa, M.; Ehlers, A.; Eigenmann, P.; Gadermaier, G.; Giovannini, M.; Gomez, F.; Grohman, R.; Guillet, C.; Hafner, C.; Hamilton, R. G.; Hauser, M.; Hawranek, T.; Hoffmann, H. J.; Holzhauser, T.; Iizuka, T.; Jacquet, A.; Jakob, T.; Janssen-Weets, B.; Jappe, U.; Jutel, M.; Kalic, T.; Kamath, S.; Kespohl, S.; Kleine-Tebbe, J.; Knol, E.; Knulst, A.; Konradsen, J. R.; Korošec, P.; Kuehn, A.; Lack, G.; Le, T.-M.; Lopata, A.; Luengo, O.; Mäkelä, M.; Marra, A. M.; Mills, C.; Morisset, M.; Muraro, A.; Nowak-Węgrzyn, A.; Nugraha, R.; Ollert, M.; Palosuo, K.; Pastorello, E. A.; Patil, S. U.; Platts-Mills, T.; Pomés, A.; Poncet, P.; Potapova, E.; Poulsen, L. K.; Radauer, C.; Radulovic, S.; Raulf, M.; Rougé, P.; Sastre, J.; Sato, S.; Scala, E.; Schmid, J. M.; Schmid-Grendelmeier, P.; Schrama, D.; Sénéchal, H.; Traidl-Hoffmann, C.; Valverde-Monge, M.; van Hage, M.; van Ree, R.; Verhoeckx, K.; Vieths, S.; Wickman, M.; Zakzuk, J.; Matricardi, P. M.; Hoffmann-Sommergruber, K. *Pediatr. Allergy Immunol.* **2023**, *34* (Suppl. 28), e13854. doi:10.1111/pai.13854
40. ImmunoCAP Allergen component MUXF3 CCD, Artikel # 14-5339-01; ThermoFisher Scientific, Inc., Waltham, MA, USA. <https://www.thermofisher.com/phadia/wo/de/product-catalog.html?articleNumber=14-5339-01®ion=AT> (accessed Feb 27, 2024).
41. Banin, E.; Neuberger, Y.; Altshuler, Y.; Halevi, A.; Inbar, O.; Nir, D.; Dukler, A. *Trends Glycosci. Glycotechnol.* **2002**, *14*, 127–137. doi:10.4052/tigg.14.127
42. Green, E. D.; Adelt, G.; Baenziger, J. U.; Wilson, S.; Van Halbeek, H. *J. Biol. Chem.* **1988**, *263*, 18253–18268. doi:10.1016/s0021-9258(19)81354-6
43. Gomord, V.; Fitchette, A.-C.; Menu-Bouaouiche, L.; Saint-Jore-Dupas, C.; Plasson, C.; Michaud, D.; Faye, L. *Plant Biotechnol. J.* **2010**, *8*, 564–587. doi:10.1111/j.1467-7652.2009.00497.x
44. Paschinger, K.; Wilson, I. B. H. *Parasitology* **2019**, *146*, 1733–1742. doi:10.1017/s0031182019000398
45. Marionneau, S.; Cailleau-Thomas, A.; Rocher, J.; Le Moullac-Vaidye, B.; Ruvoën, N.; Clément, M.; Le Pendu, J. *Biochimie* **2001**, *83*, 565–573. doi:10.1016/s0300-9084(01)01321-9
46. Hokke, C. H.; Bergwerff, A. A.; Van Dedem, G. W.; Kamerling, J. P.; Vliegthart, J. F. G. *Eur. J. Biochem.* **1995**, *228*, 981–1008. doi:10.1111/j.1432-1033.1995.0981m.x
47. Voshol, H.; van Zuylen, C. W. E. M.; Orberger, G.; Vliegthart, J. F. G.; Schachner, M. *J. Biol. Chem.* **1996**, *271*, 22957–22960. doi:10.1074/jbc.271.38.22957
48. van Rooijen, J. J. M.; Kamerling, J. P.; Vliegthart, J. F. G. *Eur. J. Biochem.* **1998**, *256*, 471–487. doi:10.1046/j.1432-1327.1998.2560471.x
49. Duca, M.; Malagolini, N.; Dall'Olio, F. *Glycoconjugate J.* **2023**, *40*, 123–133. doi:10.1007/s10719-022-10089-1
50. Pabst, M.; Grass, J.; Toegel, S.; Liebminger, E.; Strasser, R.; Altmann, F. *Glycobiology* **2012**, *22*, 389–399. doi:10.1093/glycob/cwr138
51. Stenitzer, D.; Mócsai, R.; Zechmeister, H.; Reski, R.; Decker, E. L.; Altmann, F. *Biomolecules* **2022**, *12*, 136. doi:10.3390/biom12010136
52. Hykollari, A.; Paschinger, K.; Wilson, I. B. H. *Mass Spectrom. Rev.* **2022**, *41*, 945–963. doi:10.1002/mas.21693
53. Mócsai, R.; Blaukopf, M.; Svehla, E.; Kosma, P.; Altmann, F. *Glycobiology* **2020**, *30*, 663–676. doi:10.1093/glycob/cwaa012
54. Mócsai, R.; Kaehlig, H.; Blaukopf, M.; Stadlmann, J.; Kosma, P.; Altmann, F. *Front. Plant Sci.* **2021**, *12*, 643249. doi:10.3389/fpls.2021.643249
55. Mócsai, R.; Figl, R.; Sützl, L.; Fluch, S.; Altmann, F. *Plant J.* **2020**, *103*, 184–196. doi:10.1111/tpj.14718
56. Pabst, M.; Wu, S. Q.; Grass, J.; Kolb, A.; Chiari, C.; Viernstein, H.; Unger, F. M.; Altmann, F.; Toegel, S. *Carbohydr. Res.* **2010**, *345*, 1389–1393. doi:10.1016/j.carres.2010.02.017
57. Clerc, F.; Reiding, K. R.; Jansen, B. C.; Kammeijer, G. S. M.; Bondt, A.; Wührer, M. *Glycoconjugate J.* **2016**, *33*, 309–343. doi:10.1007/s10719-015-9626-2
58. Shubhakar, A.; Jansen, B. C.; Adams, A. T.; Reiding, K. R.; Venthann, N. T.; Kalla, R.; Bergemalm, D.; Urbanowicz, P. A.; Gardner, R. A.; IBD-BIOM Consortium; Wührer, M.; Halfvarson, J.; Satsangi, J.; Fernandes, D. L.; Spencer, D. I. R. *J. Crohns Colitis* **2023**, *17*, 919–932. doi:10.1093/ecco-jcc/ijad012
59. Williams, S. E.; Noel, M.; Lehoux, S.; Cetinbas, M.; Xavier, R. J.; Sadreyev, R. I.; Scolnick, E. M.; Smoller, J. W.; Cummings, R. D.; Mealer, R. G. *Nat. Commun.* **2022**, *13*, 275. doi:10.1038/s41467-021-27781-9
60. Lee, J.; Ha, S.; Kim, M.; Kim, S.-W.; Yun, J.; Ozcan, S.; Hwang, H.; Ji, I. J.; Yin, D.; Webster, M. J.; Shannon Weickert, C.; Kim, J.-H.; Yoo, J. S.; Grimm, R.; Bahn, S.; Shin, H.-S.; An, H. J. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 28743–28753. doi:10.1073/pnas.2014207117
61. Barboza, M.; Solakyildirim, K.; Knotts, T. A.; Luke, J.; Gareau, M. G.; Raybould, H. E.; Lebrilla, C. B. *Mol. Cell. Proteomics* **2021**, *20*, 100130. doi:10.1016/j.mcpro.2021.100130
62. Wormald, M. R.; Rudd, P. M.; Harvey, D. J.; Chang, S.-C.; Scragg, I. G.; Dwek, R. A. *Biochemistry* **1997**, *36*, 1370–1380. doi:10.1021/bi9621472
63. Standards & Controls, Ludger Ltd., Oxfordshire, UK. <https://www.ludger.com/product-catalogue/standards-controls/> (accessed Feb 27, 2024).
64. Glycan Standard, Asparia Glycomics, San Sebastián, Spain. <https://www.aspariaglycomics.com/product-category/glycan-standard/> (accessed Feb 27, 2024).
65. N-Glycan Sample Preparation & Analysis, Agilent Technologies, Inc., Santa Clara, CA, USA. <https://www.agilent.com/en/product/biopharma-hplc-analysis/glycan-analysis> (accessed Feb 27, 2024).

66. BioPharma Finder Software, ThermoFisher Scientific, Inc., Waltham, MA, USA.
<https://www.thermofisher.com/at/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/biopharma-finder-software.html> (accessed Feb 27, 2024).
67. Mattu, T. S.; Pleass, R. J.; Willis, A. C.; Kilian, M.; Wormald, M. R.; Lellouch, A. C.; Rudd, P. M.; Woof, J. M.; Dwek, R. A. *J. Biol. Chem.* **1998**, *273*, 2260–2272. doi:10.1074/jbc.273.4.2260
68. Abrahams, J. L.; Campbell, M. P.; Packer, N. H. *Glycoconjugate J.* **2018**, *35*, 15–29. doi:10.1007/s10719-017-9793-4

License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (<https://www.beilstein-journals.org/bjoc/terms>), which is identical to the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
<https://doi.org/10.3762/bjoc.20.53>