



## Symbol nomenclature for glycan representation

Ajit Varki<sup>1</sup>, Richard D. Cummings<sup>2</sup>, Jeffrey D. Esko<sup>1</sup>, Hudson H. Freeze<sup>3</sup>, Pamela Stanley<sup>4</sup>, Jamey D. Marth<sup>3,5</sup>, Carolyn R. Bertozzi<sup>6</sup>, Gerald W. Hart<sup>7</sup> and Marilyn E. Etzler<sup>8</sup>

<sup>1</sup> Glycobiology Research and Training Center, University of California, San Diego, CA, USA

<sup>2</sup> Emory University School of Medicine, Atlanta, GA, USA

<sup>3</sup> Burnham Institute for Medical Research, La Jolla, CA, USA

<sup>4</sup> Albert Einstein College of Medicine of Yeshiva University, Bronx, NY, USA

<sup>5</sup> University of California, Santa Barbara, CA, USA

<sup>6</sup> University of California, Berkeley, CA, USA

<sup>7</sup> John Hopkins University School of Medicine, Baltimore, MD, USA

<sup>8</sup> University of California, Davis, CA, USA

The glycan symbol nomenclature proposed by Harvey *et al.* in these pages has relative advantages and disadvantages. The use of symbols to depict glycans originated from Kornfeld in 1978, was systematized in the First Edition of “Essentials of Glycobiology” and updated for the second edition, with input from relevant organizations such as the Consortium for Functional Glycomics. We also note that >200 illustrations in the second edition have already been published using our nomenclature and are available for download at PubMed.

Received: October 16, 2009

Revised: October 16, 2009

Accepted: October 26, 2009



### Keywords:

Glycans / Glycomics / Glycoproteomics / Monosaccharides / Nomenclature / Symbols

It has long been recognized that having a simple way to depict glycan structures is a useful starting point for scientific communication, especially for the many non-expert scientists who are now venturing into the field of glycosciences. The glycan symbol nomenclature that is attributed by Harvey *et al.* [1] to the Consortium for Functional Glycomics (CFG) actually had its origins in 1978, in a paper by Kornfeld *et al.* [2], who realized that standard carbohydrate nomenclature was too complex to convey simple concepts. Variations of the Kornfeld symbol system were subsequently used by numerous authors until 1999, when it was systematized and colorized in the First Edition of “Essentials of Glycobiology” [3]. Given the reasonably wide usage of this codified system, the Editors of “Essentials” decided to update it for the proposed second edition (see

<http://grtc.ucsd.edu/symbol.html> and Supporting Information). Before finalizing the plan, we consulted other relevant organizations, such as the CFG and NLM/NCBI, to ensure compatibility and agreement, realizing of course that there would be no way to please everyone in the field. After further consultations with some others in the community, the CFG adopted the modified Essentials symbol nomenclature (<http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml>) and began to use and publicize it well before the eventual publication of the second edition of Essentials [4], which was recently reviewed by Dwek [5].

With regard to the alternate system proposed by Harvey *et al.*, we can see some relative advantages and disadvantages. For example, while the system gives more information regarding linkages, it might be difficult to use for some vertical representations of glycans. On the other hand, with regard to concerns about use of color in the “Essentials” system, there is actually no problem with discriminating shading in black and white reproductions, if the details of the originally recommended color scheme (<http://grtc.ucsd.edu/symbol.html>, <http://www.functionalglycomics.org/static/consortium/Nomenclature.shtml> and Supporting Information)

**Correspondence:** Professor Ajit Varki, Glycobiology Research and Training Center, University of California, San Diego, 9500 Gilman Drive, MC 0687, La Jolla, CA 92093-0687, USA

**E-mail:** [a1varki@ucsd.edu](mailto:a1varki@ucsd.edu)

**Fax:** +1-858-534-5611

**Abbreviation:** CFG, Consortium for Functional Glycomics

are followed. Both systems do have the limitation that the designated symbols are restricted to monosaccharides found in vertebrate systems, and discussions are under way in various fora to define symbol designations for other monosaccharides found in non-vertebrate taxa. We prefer not to comment further on other specifics regarding the different systems, as many represent matters of opinion and taste. Moreover, >200 illustrations in the second edition of “Essentials” have recently been published using our nomenclature, and they are now freely available for downloading as teaching slides from NCBI/PubMed (<http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=glyco2>).

The original purpose of the “Essentials” symbol nomenclature was to describe glycan composition and biosynthesis, whereas the emphasis of the alternate system proposed by Harvey *et al.* [1], is to provide further information about primary structure, *i.e.* linkages. Of course, no symbol system will ever convey a full appreciation of the three-dimensional structure of glycans, which is needed to truly understand how glycans and proteins interact. Moreover, such issues regarding scientific nomenclature generally tend to be more controversial than the underlying science, as there is never one final correct answer, and some aspects are indeed matters of opinion and taste. Only time will tell if one or both systems mentioned here eventually find wide usage, or if some other system or a hybrid version replaces them. In any event, what matters much more is the fact that we now know so much about glycan biology that

there is a clear need for such representations for the larger scientific community.

*Conflict of Interest:* The co-authors are Editors of *Essentials of Glycobiology*, and we herein promote open access to the book figures at PubMed. However, we do not stand to gain financially from this.

## References

- [1] Harvey, D. J., Merry, A. H., Royle, L., P Campbell, M. *et al.*, Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics* 2009, 9, 3796–3801.
- [2] Kornfeld, S., Li, E., Tabas, I., The synthesis of complex-type oligosaccharides II characterization of the processing intermediates in the synthesis of the complex oligosaccharide units of the vesicular stomatitis virus G protein. *J Biol Chem.* 1978, 253, 7771–7778.
- [3] Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H. *et al.* (Eds.), *Essentials of Glycobiology, 1st Edn.*, Cold Spring Harbor Laboratory Press, Plainview, NY 1999.
- [4] Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H. *et al.* (Eds.), *Essentials of Glycobiology, 2nd Edn.*, Cold Spring Harbor Laboratory Press, Plainview, NY 2009.
- [5] Dwek, R. A., Glycobiology: Moving into the mainstream. *Cell* 2009, 137, 1175–1176.

# **“Essentials of Glycobiology”**

**Final Plan for**

**Usage of Sugar Symbols**

**for the Second Edition**

# General Principles

- **Modify monosaccharide symbol set from First Edition of “Essentials” to achieve common ground with as many other interested groups as possible (NCBI, Consortium for Functional Glycomics, KEGG pathways etc.)**
- **Avoid same shape/color but different orientation to represent different sugars since annotation of mass spectra are hard to show as horizontal cartoons.**
- **Choice of symbols should be logical and simple to remember.**
- **Each monosaccharide type (e.g. hexose) should have the same shape, and isomers are to be differentiated by color/black/white/shading.**
- **Use same shading/color for different monosaccharides of same stereochemical designation e.g., Gal, GalNAc, GalA should all be the same shading/color.**
- **To minimise variations, sialic acids and uronic acids are same shape. Only major uronic and sialic acid types are represented.**
- **While color is useful, the system should also function with black and white, and colored representations should survive black-and white printing or xeroxing.**
- **Positions of linkage origin are assumed to be common ones unless indicated. Anomeric notation and destination linkage indicated without spacing/dashes.**
- **Modifications of monosaccharides (e.g., sulfation, O-acetylation) indicated by attached small letters, with numbers indicating linkage positions, if known.**
- **Only common monosaccharides in vertebrate systems are assigned a specific symbol. All other monosaccharides are represented by an open Hexagon, and defined in the figure legend. If there is more than one type of undesignated monosaccharide in a figure, an letter designation internal to the Hexagon can be included to differentiate between them - again, specified in the figure legend.**
- **Note: To reproduce precise colors, use the color sliders. Click on the paint pot icon and select “more colors”. Then, click on sliders in the menu bar and fill in numbers. Or, you can copy an existing fill color by using the eye dropper and touching the color you want.**

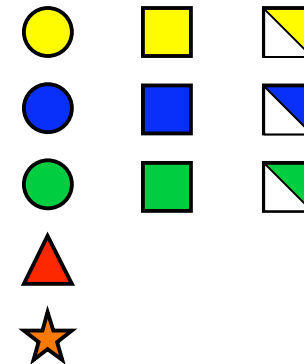
# Essentials Second Edition Symbols : RGB colors

Hexoses: Circles      N-Acetylhexosamines: Squares

Hexosamines: Squares divided diagonally

*Print in color*

- Galactose stereochemistry: **Yellow (255,255,0)** with Black outline
- Glucose stereochemistry: **BLUE (0,0,250)** with Black outline
- Mannose stereochemistry: **GREEN (0,200,50)** with Black outline
- Fucose: **RED (250,0,0)** with Black outline
- Xylose: (5-pointed star) **ORANGE (250,100,0)** with Black outline



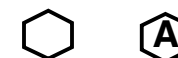
Acidic Sugars (Diamonds)

- Neu5Ac: **PURPLE (125,0,125)** with Black outline
- Neu5Gc: **LIGHT BLUE (200,250,250)** with Black outline
- KDN: **GREEN (0,200,50)** with Pattern & Black outline
- GlcA: **BLUE (0,0,250)**/Upper segment with Black outline
- IdoA: **TAN (150,100,50)**/Lower segment with Black outline
- GalA: **RED (250,0,0)**/Left segment with Black outline
- ManA: **GREEN (0,200,50)**/Right segment with Black outline



Other Monosaccharide

(use letter designation inside symbol to specify if needed)



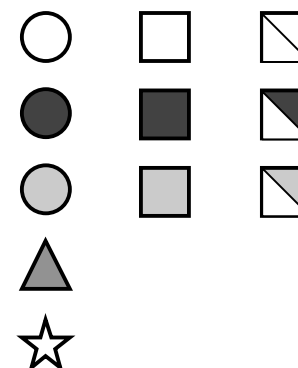
# Essentials Second Edition Symbols : Black & White

Hexoses: Circles      N-Acetylhexosamines: Squares

Hexosamines: Squares divided diagonally

*Print in black & white*

- Galactose stereochemistry: white with Black outline
- Glucose stereochemistry: Black with Black outline
- Mannose stereochemistry: Grey with Black outline
- Fucose: Dark Grey with Black outline
- Xylose: (5-pointed star) with Black outline



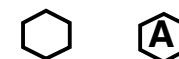
## Acidic Sugars (Diamonds)

- Neu5Ac: Dark Grey with Black outline
- Neu5Gc: White with Black outline
- KDN: Light Grey Pattern & Black outline
- GlcA: Grey upper segment with Black outline
- IdoA: Grey Lower segment with Black outline
- GalA: Grey Left segment with Black outline
- ManA: Grey Right segment with Black outline



## Other Monosaccharide

(use letter designation inside symbol to specify if needed)



# Linkages/Configuration/Modifications

## Unless otherwise indicated:

All monosaccharides are assumed to be in the D-configuration except for Fucose and Iduronic acid, which are in the L- configuration

All glycosidically-linked monosaccharides assumed to be in the pyranose form

All monosaccharide glycosidic linkages are assumed to originate from the 1-position except for the sialic acids, which are linked from the 2-position

O-esters and ethers are shown attached to the symbol with a number e.g.,

9Ac for 9-O-acetyl group

3S for 3-O-sulfate group

6P for a 6-O-phosphate group

8Me for 8-O-Methyl group

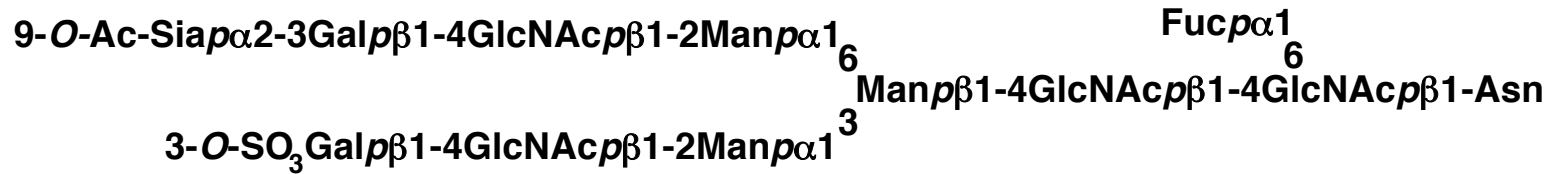
9Acy for 9-O-acyl group

9Lt for 9-O-Lactyl group

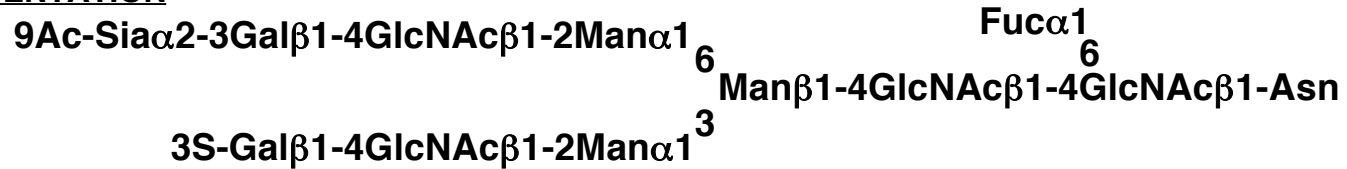
For N-substituted groups assume there is only one Amino group on the monosaccharide with an already known position e.g., use NS for N-sulfate group on Glucosamine, assumed to be at the 2-position

# Examples - Representations of N-Glycans

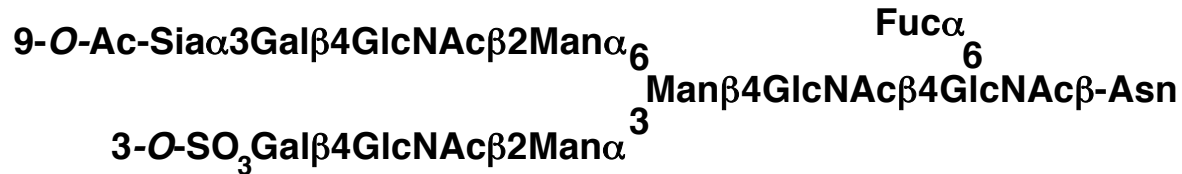
## FULL REPRESENTATION



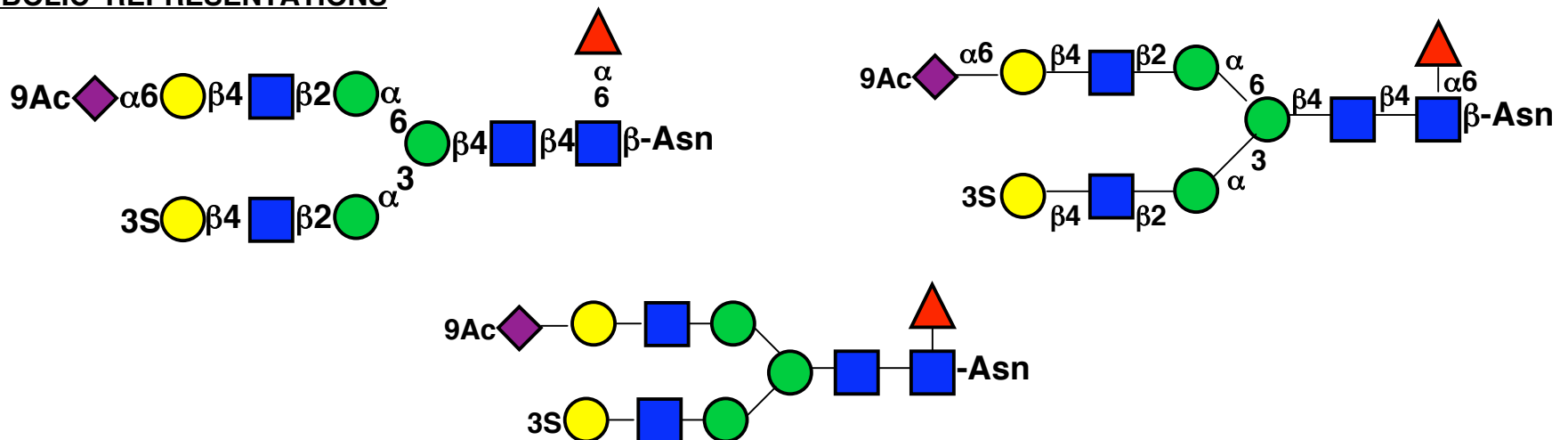
## MODIFIED REPRESENTATION



## SIMPLIFIED REPRESENTATION (IUB-IUPAC JCBN Recommendations, J.Biol.Chem. 262:13-18, 1985)



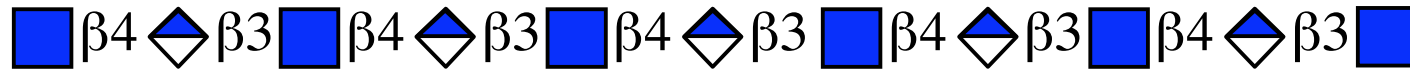
## SYMBOLIC REPRESENTATIONS





## Examples - Simplified Text and Symbolic Representation of Glycosaminoglycans (GAGs)

GlcNAc $\beta$ 4GlcA $\beta$ 3GlcNAc $\beta$ 4GlcA $\beta$ 3GlcNAc $\beta$ 4GlcA $\beta$ 3GlcNAc $\beta$ 4GlcA $\beta$ 3GlcNAc



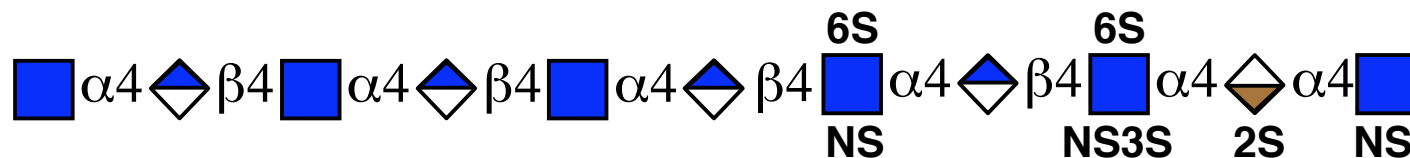
Hyaluronan

GalNAc $4S\beta$ 4GlcA $\beta$ 3GalNAc $4S\beta$ 4IdoA $\alpha$ 3GalNAc $4S\beta$ 4IdoA $2S\alpha$ 3GalNAc $4S\beta$ 4GlcA $\beta$ 3



Chondroitin/Dermatan Sulfate

GlcNAc $\alpha$ 4GlcA $\beta$ 4GlcNAc $\alpha$ 4GlcA $\beta$ 4GlcNAc $\alpha$ 4GlcA $\beta$ 4GlcNS $6S\alpha$ 4GlcA $\beta$ 4GlcNS $3S6S\alpha$ 4IdoA $2S\alpha$ 4GlcNS



Heparin/Heparan Sulfate