Updates to the Symbol Nomenclature For Glycans (SNFG) Guidelines

Running title: Symbol Nomenclature For Glycans

Sriram Neelamegham1,*, Kiyoko Aoki-Kinoshita2, Evan Bolton3, Martin Frank4, Frederique Lisacek5, Thomas Lütteke6, Noel O’Boyle7, Nicolle Packer8, Pamela Stanley9, Philip Toukach10, Ajit Varki11, Robert J. Woods12, and the SNFG Discussion group13

1State University of New York at Buffalo, USA
2Soka University, Japan
3National Library of Medicine (NLM), USA
4Biognos AB Göteborg, Sweden
5Swiss Institute of Bioinformatics (SIB), Switzerland
6GIP GmbH, Offenbach, Germany
7NextMove Software, USA
8Macquarie University, Australia
9Albert Einstein College of Medicine, USA
10Zelinsky Institute of Organic Chemistry, Russia
11University of California, San Diego, USA
12University of Georgia, USA
13Other members of the SNFG Discussion group: Alan Darvill, University of Georgia, USA; Anne Dell, Imperial College London, UK; Bernard Henrissat, Centre National de la Recherche Scientifique (CNRS), France; Carolyn Bertozzi, Stanford University, USA; Gerald Hart, University of Georgia, USA; Hisashi Narimatsu, Research Center of Medical Glycoscience, Japan; Hudson Freeze, Sanford-Burnham-Prebys Research Institute, USA; Issaku Yamada, The Noguchi Institute, Japan; James Paulson, Scripps Research Institute, USA; James Prestegard, University of Georgia, USA; Jamey Marth, University of California Santa Barbara, USA; JFG Vliegenthart, Bijvoet Center, Netherlands; Marilynn Etzler, UC Davis, USA; Markus Aebi, ETH Zürich, Switzerland; Minoru Kamehisa, Kyoto University, Japan; Naoyuki Taniguchi, Riken Global Research Cluster, Japan; Nathan Edwards, Georgetown University, USA; Pauline Rudd, National Institute for Bioprocessing Research & Training, UK; Peter Seeberger, Max-Planck-
Institute of Colloids and Interfaces, Germany; Raja Mazumder, The George Washington University, USA; Rene Ranzinger, University of Georgia, USA; Richard Cummings, Harvard Medical School, USA; Ronald Schnaar, Johns Hopkins University School of Medicine, USA; Serge Perez, French National Centre for Scientific Research, France; Stuart Kornfeld, Washington University in St. Louis, USA; Taroh Kinoshita, Osaka University, Japan; William York, University of Georgia, USA; Yuriy Knirel, Russian Academy of Sciences, Russia

* Correspondence: Sriram Neelamegham, neel@buffalo.edu, 906 Furnas Hall, Buffalo, NY 14260; Ph: 716-645-1200; FAX: 716-645-3822.

Keywords: database/ glycoscience/ glycobiology/ software/ symbol nomenclature

Abstract

The Symbol Nomenclature For Glycans (SNFG) is a community-curated standard for the depiction of monosaccharides and complex glycans using various colored-coded, geometric shapes, along with defined text additions. It is hosted by the National Center for Biotechnology Information at the NCBI-Glycans Page (www.ncbi.nlm.nih.gov/glycans/snfg.html). Several changes have been made to the SNFG page in the last year to update the rules for depicting glycans using the SNFG, to include more examples of use, particularly for non-mammalian organisms, and to provide guidelines for the depiction of ambiguous glycan structures. This Glycoforum article summarizes these recent changes.

Symbol Nomenclature For Glycans

The Symbol Nomenclature For Glycans (SNFG) is a community-curated, broadly utilized standard for the depiction of glycans using various colored, geometric shapes (Varki, A., et al. 2015). It has been adopted by various publications including the journal Glycobiology (Haltiwanger, R.S. 2016) and others listed at https://www.ncbi.nlm.nih.gov/glycans/snfgorg.html, and by major Glycoscience databases (Table I). The historical background of the SNFG, which now spans nearly 40 years, has been previously reviewed (Varki, A., et al. 2015). Currently, the SNFG is hosted by the National Center for Biotechnology Information (NCBI) at the NCBI Glycans Page (www.ncbi.nlm.nih.gov/glycans/snfg.html), and it is curated by an international group of researchers in the field (see SNFG discussion group list). Additionally, various software programs for
sketching glycans have been developed to support the implementation of this nomenclature (Table II).

The overall goal of the SNFG is to: 1) Facilitate communications and presentations of monosaccharides and glycans for researchers in the Glycosciences, and for scientists and students less familiar with the field; 2) Ensure uniform usage of the nomenclature in the literature, thus helping to ensure scientific accuracy in journal and online publications; and 3) Continue to develop the SNFG and its applications to aid wider use by the scientific community. With these goals in mind, updates have been made this year to the NCBI Glycans page that hosts the SNFG. Specifically, the footnotes to Table 1 at the SNFG page that include the rules for depicting glycans have been modified. No changes were made to the table. The footnotes are now organized into 10 themes in order to help streamline their use. These footnotes provide guidelines for: i. general usage of the SNFG; ii. CMYK/RGB color codes; iii. symbol colors and shapes; iv. ring configurations; v. bond linkage presentation; vi. sialic acids; vii. glycan modifications; viii. amino substitutions; ix. handling ambiguous or partially defined glycans; and x. depicting non-glycan entities using SNFG renderings.

More examples have been included for non-mammalian species (Figure 1), with the realization that there is much greater diversity in monosaccharide composition and modifications in these organisms compared to vertebrates. This necessitates the need to engender greater flexibility with respect to SNFG use. To address this issue, several changes have been made.

First, white symbols of various shapes were previously used to define monosaccharides of unknown/undefined stereochemistry (white circle for hexose, type undefined, white diamond for deoxynonulosonic acids etc.). This usage of white symbols remains in the current revision. For example, Fig. 1A depicts a Drosophila O-glycan that contains generic hexose and N-acetylhexosamine. However, while the white, flat hexagon was previously used to denote both heptoses and unknown monosaccharides, its usage is now restricted to depicting only unknown or partially-defined monosaccharides.
Second, a single, non-italicized letter (A..Z) is now permitted inside white symbols to provide additional information about a monosaccharide that is not defined by a colored symbol. Footnotes/figure legends associated with these single letters convey structural details regarding this annotated symbol. When using the white symbol with a single letter, it is necessary to ensure that the chemical composition (and thus mass) of the entity corresponds to the composition of the selected white generic type. For example, the bacterial glycan in Fig. 1B contains a deoxyhexose variant with known composition. A white triangle with letter ‘A’ is used since this deoxyhexose is not defined by any colored SNFG symbol. A footnote is provided to explain the exact usage. In another case, if a monosaccharide has no equivalent representation in the SNFG table as in Fig. 1C, a pentagon is used. In this case, heparinase action on a glycosaminoglycan results in a chemically-defined product, but the generic type of the entity does not correspond to any of the standard white symbols in the SNFG table. Thus, a white pentagon is used with the letter ‘A’ and an accompanying footnote.

Third, the rules for describing substituents have been reviewed and more examples are provided. For example, Fig. 1D shows a commonly used fluoro-glucose analog, with a footnote providing a full explanation of the modification. Fig. 1E illustrates a multiply-modified sialic acid, where the substituents are presented as a sequential(concatenated) string. To the extent possible, abbreviations used to describe substituents should follow guidelines described in footnote 7 of the updated SNFG page. The presence of variable amounts of substituents may be indicated using the ± symbol or by indicating % presence if known, e.g. "70% 6Ac" to indicate presence of 6Ac on 70% of a residue or repeating unit (Fig. 1F). Less common substituents like ‘4Fo’, which are not among the standard monosaccharide modifications listed on the SNFG page, may also be indicated using footnote as illustrated. The final example illustrates the use of italicized ‘o’ within a pink star to depict ribitol in an O-glycan that is found on mammalian dystroglycan (Fig. 1G).

Please note that the meaning of single letter annotations in colored SNFG symbols are different from those in the white symbols. In the colored symbols, such annotations can only be used to denote either ‘D’/’L’ configuration or ring closure information (‘p’, ‘f’ or ‘o’ for pyranose, furanose and alditol, respectively). In white symbols, they are used to annotate details regarding
generic monosaccharide types as discussed in the above examples. More exhaustive examples of SNFG usage can be found at the NCBI-SNFG page, Essentials of Glycobiology textbook (https://www.ncbi.nlm.nih.gov/books/NBK310274/) and the database resources listed in Table I.

Overall, the goal of the SNFG is to make the field of glycobiology more visual and readily accessible, especially for new users and the larger biomedical community. As we make progress in this endeavor, the group anticipates working with the NCBI Glycan Informatics Advisory Group (GlyAg, www.ncbi.nlm.nih.gov/glycans/glyag.html) to populate more glycan related pages at the NCBI, and to establish links between these pages and other database resources (Table 1). Comments from the scientific community are welcome, including examples of SNFG implementation in databases and software resources.

Acknowledgements

This is a global community-driven project. Some of the individuals who volunteer their time as members of the SNFG Discussion Group have received grant support from national funding agencies including the US National Institutes of Health, Russian Science Foundation, Swiss National Science Foundation, Research Data Australia, Japan Science and Technology Agency (JST) and the National Bioscience Database Center (NBDC) of Japan.
References


FIGURE LEGEND

A. Hexose and HexNAc not specified

C. Lyase reaction on part of heparin sulfate

D. A=4-deoxy-D-arabino-hexose

E. $2[^{18}F] = 2$-deoxy-$2-[^{18}F]$fluoro

F. 4Fo=4-formamido

G. Dystroglycan
### Table I SNFG compliant and actively maintained glycan databases*

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Contents</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian Community of Glycoscience and Glycotechnology</td>
<td>acgg.asia/db/</td>
<td>Includes glycan disease, glycogene, lectin frontier and curated protocols developed by the JCGGDB (Japanese Consortium for Glycobiology and Glycotechnology) and other collaborators</td>
<td>(Maeda, M., et al. 2015)</td>
</tr>
<tr>
<td>CSDB (Carbohydrate Structure Database)</td>
<td>csdb.glycoscience.ru/database</td>
<td>Manually curated bacterial, plant and fungal carbohydrate structure database</td>
<td>(Toukach, P.V. and Egorova, K.S. 2016)</td>
</tr>
<tr>
<td>CSDB_GT (CSDB Glycosyltransferases)</td>
<td>csdb.glycoscience.ru/gt.html</td>
<td>Manually curated database of glycosyltrasferases with confirmed activity (in 2019, of <em>E. coli</em> and <em>A. thaliana</em>).</td>
<td>(Egorova, K.S., et al. 2019)</td>
</tr>
<tr>
<td>GLYCOSCIENCES.de</td>
<td>glycosciences.de</td>
<td>Various tools and databases for glycobiology, with focus on 3D structures &amp; references to related PDB entries</td>
<td>(Bohm, M., et al. 2019)</td>
</tr>
<tr>
<td>Glycostore</td>
<td>glycostore.org</td>
<td>Glycan chromatographic retention properties</td>
<td>(Zhao, S., et al. 2018)</td>
</tr>
<tr>
<td>Glyco3D</td>
<td>glyco3d.cermav.cnrs.fr/home.php</td>
<td>Family of databases covering the 3D features of mono-, poly- &amp; oligo-saccharides, glycosyltrasferases and lectins, and glycosaminoglycan-binding proteins.</td>
<td>(Perez, S., et al. 2015)</td>
</tr>
<tr>
<td>Glycomics@ExPASy</td>
<td><a href="http://www.expasy.org/glycomics">www.expasy.org/glycomics</a></td>
<td>Various glycomics tools developed at the Swiss Institute of Bioinformatics and links to other sites</td>
<td>(Mariethoz, J., et al. 2018)</td>
</tr>
<tr>
<td>GlyCosmos</td>
<td>glycosmos.org</td>
<td>Portal for glycoscience data resources and repositories</td>
<td></td>
</tr>
<tr>
<td>GlyGen</td>
<td><a href="http://www.glygen.org">www.glygen.org</a></td>
<td>Computational and Informatics Resources for Glycoscience research</td>
<td></td>
</tr>
<tr>
<td>Database</td>
<td>Website</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>MonosaccharideDB</td>
<td>monosaccharidedb.org</td>
<td>Monosaccharide database</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** * Table only lists databases that follow a majority of the SNFG nomenclature guidelines, not all glycoscience resources

**SNFG adoption in progress**
## Table II SNFG drawing software

<table>
<thead>
<tr>
<th>Name</th>
<th>Purpose</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D-SNFG in LiteMol</td>
<td>Implementation of 3D-SNFG in LiteMol for 3D display of glycans</td>
<td>(Sehnal, D. and Grant, O.C. 2019)</td>
</tr>
<tr>
<td>SugarSketcher</td>
<td>Drag and drop tool to sketch glycans (web based)</td>
<td>(Alocci, D., et al. 2018)</td>
</tr>
</tbody>
</table>