Glycomics: Building Upon Proteomics to Advance Glycosciences

Lance Wells*‡ and Gerald W. Hart*§

Glycans are among the building blocks of the four major biomolecules of life—carbohydrates, nucleic acids, lipids, and proteins. Intriguingly, the other three major biomolecules of life either contain (*ribo-* or *deoxyribo*nucleic acids) or can be modified with (*glyco*lipids and *glyco*proteins) carbohydrates. To quote one of the leaders in the field of glycobiology, Ajit Varki of the University of California at San Diego, "Despite more than 3 billion years since the origin of life on earth, the powerful forces of biological evolution seem to have failed to generate any living cell that is devoid of a dense and complex array of cell surface glycans" (1). Given the importance of glycans and/or glycoconjugates to all fields of life science and their involvement in virtually every pathophysiological condition affecting mankind, the detection and analysis of these biomolecules is fundamental to understanding life. A recent National Academies of Science report (2) concludes that although a better understanding of glycoscience is required in order to advance improvements in human health and sustainability, efforts and investments in these biomolecules have lagged considerably behind expenditures in other areas of bioscience. This report further suggests that robust glycomic data analysis platforms and rigorously annotated glycan/glycoconjugate databases are needed to move the field forward.

In several aspects, glycomic analysis has followed in the footsteps of proteomics. This provides an opportunity for the field of glycomics to emulate the successes and hopefully avoid some of the pitfalls of proteomics. However, unlike that of nucleic acids and proteins that are template derived, the generation of glycan structures is governed by complex, non-template processes. Furthermore, unlike DNA/RNA and proteins that are composed of building blocks that are predominantly added in one particular linkage in a linear fashion, carbohydrates are made up of monosaccharides that can be added to one another via a variety of linkages and are more often branched than linear. These two facts generate a considerable challenge to the analytical and bioinformatics community. In order to confront this challenge, a meeting was held in conjunction with the Warren Workshop for Glyconjugate Analysis in 2012 in Athens, GA, at the Complex Carbohydrate Research Center with the goal of developing guidelines for the publication of mass-spectrometry-based glycomics data. The guidelines developed as a result of that conference are presented here. The Athens Guidelines for the Publication of Glycomics Data represent the thoughts and efforts of many individuals who are leaders in the fields of glycomics and bioinformatics, including Sanjay Agravat, Parastoo Azadi, Ralph Bradshaw, Alma Burlingame, Matthew Campbell, Robert Chalkley, Cathy Costello, Rick Cummings, Bob Haltiwanger, Jerry Hart, Stuart Haslam, Niclas Karlsson, Carsten Kettner, Daniel Kolarich, Ryan McBride, Al Merrill, Nicki Packer, Vern Reinhold, Rene Renzinger, Pauline Rudd, Doug Sheeley, Weston Struwe, Michael Tiemeyer, Lance Wells, Iain Wilson, Will York, and Joe Zaia. These guidelines, along with a checklist, are also available online at the *Molecular and Cellular Proteomics* website.

In this special issue, we highlight a wide variety of glycomic approaches, through mini-reviews and research articles, that not only advance our understanding of the structural complexity and functional diversity of glycans and glycoconjugates, but also build upon the existing tools and technologies developed by the proteomics community.

Athens Guidelines for the Publication of Glycomic Studies

The identification of free or released glycans, glycopeptides, or glycolipids is commonly accomplished through a combination of approaches. Many studies depend on the acquisition of mass spectra and the conversion of these data into a format suitable for analysis and interpretation. The following information is required by the journal for articles reporting mass spectrometric glycoconjugate analyses.

Clear Definition of the Level of the Glycan Structural Analysis and Its Relationship to the Biological Questions Addressed in the Study—Some glycomic studies need only provide profiling of possible structural classes based upon measured mass values and known biosynthetic pathways, whereas others might require detailed structural analyses to address a biological question. It is essential that authors clearly define the level of the structural analyses being presented and how they are supported by appropriate structural arguments. This is more explicitly outlined for MS analyses in the section "Glycan or Glycoconjugate Identification." It is recommended

that authors use the glycan symbolic representations outlined in *Essentials for Glycobiology* when possible (3).

Search Parameters and Acceptance Criteria—This section applies to free glycans and those released from glycoconjugates. As the glycosylation of proteins is a post(co-)-translational modification, this topic is further covered by the Protein, Peptide, and PTM Guidelines of this journal (available at the *Molecular and Cellular Proteomics* web site).

The following supporting information should be included in the Experimental section of a manuscript for MS-based analyses.

Peak Lists—This refers to the method and/or program (including version number and/or date) used to create the "peak lists" from the original data and the parameters used in the creation of these peak lists, particularly any processing, which might affect the quality of the subsequent database/manual search. Examples include smoothing, signal-to-noise thresholding, charge state assignment, de-isotoping, etc. In cases when additional customized processing of the collections of peak lists has been performed (*e.g.* clustering and/or filtering), the method and/or program (including version number and/or date) should be referenced.

Search Engine—The name and version (or release date) of all programs used for database searching must be provided, as well as the internal energy deposition and dissociation methods used and the appropriate fragment types allowed.

Database/Spectral Library—The name and version (or release date) of all databases or libraries used must be provided. If a database or library was compiled in-house, a complete description of the source of the sequences or spectra is required, as well as the software used for library generation. The number of entries actually searched from each database/library should be included.

Fixed/Variable Modifications—A list of all modifications (reducing end, permethylation, acetylation, metal ions, etc.) considered must be provided, and the author must state whether they are fixed or variable.

Exclusion of Known Contaminants—All omitted peaks from pre-designated contaminants (or whether any of these peaks are used for calibration) must be identified. It must also be stated whether degradation products from the isolation method or from in-source fragmentation were considered.

Specificity—A description of all methods used to generate glycans or glycoconjugates (enzymatic or chemical), including any assumed specificity, must be provided, along with the nature of any efforts employed to quantify the efficiency of release/capture.

Threshold—The criteria used for accepting individual spectra should be stated, along with a justification.

Isobaric/Isomeric Assignments—The criteria (and/or assumptions) used for assigning a particular individual structure should be stated, along with a justification.

Glycan or Glycoconjugate Identification—The information for each glycan identified should be specified in the Results (or Supplemental) section.

All Assignments—A list (in one or more tables) noting any deviation from expected release specificity must be provided.

Precursor Charge and Mass/Charge (m/z)—These values should be listed for each assignment in the same table (including deviation between experimental and theoretical), and significant digits should be consistent with the actual performance of the instrumentation.

Modifications Observed—These alterations (reducing end, adducts, etc.) should be listed for each assignment in the same table.

Number of Assigned Masses—For identifications based on measured mass only, the total number of peaks, both matched and unmatched (at the criteria set above), should be listed in the identification table.

Score(s) (If Used)—The relevant score (depending on the software used) and any associated statistical information obtained for searches conducted must be provided for each glycan in the table. For isobaric/isomeric species, a rationale for how one or multiple structures were selected for inclusion must be included in the text.

For all identifications, it must be possible to view spectra. Submission (with the manuscript) of annotated spectra is required, or the spectra have to be placed in a public database. Web sites maintained by the authors (or their associates) are not acceptable. In addition, the submission of supplementary data in a file format that allows visualization of the spectra (*m/z* and intensity lists) for each glycan assigned is strongly encouraged.

Structural Assignments—The rationale, including literature-based biological assumptions, for assigning structure(s) including monosaccharide composition and linkage must be clearly outlined in both Experimental and Results sections.

Quantification—Manuscripts presenting quantitative results must provide the following information:

(a) All relevant quantification data (as part of identification tables), along with a description of how the raw data were processed to produce these measurements.

(b) A description of how the biological reliability of measurements was validated (using biological replicates, statistical methods, independent experiments, etc.). Studies based on a single biological experiment that lack orthogonal methods of validation are generally not acceptable (except as a dataset for testing bioinformatic systems). If a biological replicate from the same source cannot be obtained (*e.g.* patient sample), a large enough number of similar biological samples, appropriately justified, must be obtained in order to ensure that the conclusions deduced are sound.

(c) A description of the treatment of relevant systematic error effects (incomplete labeling, interference from overlapping precursor ions, etc.).

(d) A description of the treatment of random error issues (rejection of outliers, categorical exclusion of data by means of threshold selection, etc.).

(e) Proper estimates of uncertainty in quantification using replicates and statistical methods. The number of samples (technical and biological) and methods used for determining error analysis must be provided. Standard methods (standard deviation, S.E., *t* test, etc.) or specialized software may be used and must be cited as appropriate.

(f) A description of how isobaric/isomeric species were quantified.

* To whom correspondence should be addressed: Lance Wells, Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, Tel.: 706-542-7806, Fax: 706-542-4412, E-mail: lwells@ccrc.uga.edu; Gerald W. Hart, Department of Biological Chemistry, Johns Hopkins School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205-2185, Tel.: 410-614-5993, Fax: 401-614-8804, E-mail: gwhart@jhmi.edu.

REFERENCES

1. Varki, A. (2011) Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harb. Perspect. Biol.* **3,** 1–14
2. Walt, D., Aoki-Kinoshita, K. F., Bendiak, B., Bertozzi, C. R., Boons, G. J., Darvill, A., Hart, G. W., Kiessling, L. L., Lowe, J., Moon, R. J., Paulson, J. C., Sasisekharan, R., Varki, A. P., Wong, C. H. (2012) *Transforming Glycoscience: A Roadmap for the Future*, National Academies Press, Washington, D.C.
3. Varki, A., Cummins, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E. (2009) Essentials of Glycobiology, Second Edition. Cold Spring Laboratory Press, Cold Spring Harbor, New York