

# The Minimum Information Required for a Glycomics Experiment (MIRAGE) Project: Improving the Standards for Reporting Mass-spectrometry-based Glycoanalytic Data

Daniel Kolarich<sup>‡</sup>, Erdmann Rapp<sup>§</sup>, Weston B. Struwe<sup>¶</sup>, Stuart M. Haslam<sup>||</sup>, Joseph Zaia<sup>\*\*</sup>, Ryan McBride<sup>‡‡</sup>, Sanjay Agravat<sup>§§</sup>, Matthew P. Campbell<sup>¶¶</sup>, Masaki Kato<sup>|||</sup>, Rene Ranzinger<sup>a</sup>, Carsten Kettner<sup>b</sup>, and William S. York<sup>a,c</sup>

The MIRAGE guidelines are being developed in response to a critical need in the glycobiochemistry community to clarify glycoanalytic results so that they are more readily evaluated (in terms of their scope and depth) and to facilitate the reproduction of important results in the laboratory. The molecular and biological complexity of the glycosylation process makes thorough reporting of the results of a glycomics experiment a highly challenging endeavor. The resulting data specify the identity and quantity of complex structures, the precise molecular features of which are sometimes inferred using prior knowledge, such as familiarity with a particular biosynthetic mechanism. Specifying the exact methods and assumptions that were used to assign and quantify reported structures allows the interested scientist to appreciate the scope and depth of the analysis. Mass spectrometry (MS) is the most widely used tool for glycomics experiments. The interpretation and reproducibility of MS-based glycomics data

depend on comprehensive meta-data describing the instrumentation, instrument setup, and data acquisition protocols. The MIRAGE guidelines for MS-based glycomics have been designed to facilitate the collection and sharing of this critical information in order to assist the glycoanalyst in generating data sets with maximum information content and biological relevance. *Molecular & Cellular Proteomics* 12: 10.1074/mcp.O112.026492, 991–995, 2013.

## REPORTING GUIDELINES FOR GLYCOMICS—WHY BOTHER?

The increasing importance of glycoscience in modern biology was recently described in the publication *Transforming Glycoscience: A Roadmap for the Future*, prepared by the U.S. National Academy of Sciences (1). Glycomics, which is one emerging discipline of glycoscience, utilizes diverse analytical and computational techniques aimed at comprehensively identifying and characterizing the repertoire of glycan structures present in an organism, cell, or tissue at a defined time. Recent technical advances have enabled glycan analyses to proceed with increased depth, speed, and efficiency and have led to both the increased publication of glycomics data in carbohydrate-related journals and the accumulation of large data sets on a global scale.

The application of data mining techniques and analytical software tools make it possible to identify relationships among distinct data sets in a way that generates new knowledge. However, the annotation and archiving of information are often carried out in a retrospective way (e.g. by manually extracting it from the literature and importing it into databases). Therefore, database quality is highly dependent on the reliability and depth of literature reports, which can be judged only if the experiments that generate the data are adequately described. Thus, in both publications and databases, the prerequisite for high information quality is comprehensive reporting of the experimental context in which the data were generated.

Unfortunately, a large proportion of published glycomics data do not meet this criterion. Although experimental data are highly dependent on the experimental conditions applied,

From the <sup>‡</sup>Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany; <sup>§</sup>Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, 39106 Magdeburg, Germany; <sup>¶</sup>Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford, OX1 3QZ, UK; <sup>||</sup>Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, UK; <sup>\*\*</sup>Center for Biomedical Mass Spectrometry, Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118; <sup>‡‡</sup>Department of Chemical Physiology, The Scripps Research Institute, 10550, N. Torrey Pines Road, La Jolla, California 92037; <sup>§§</sup>Center for Comprehensive Informatics, Research and Woodruff Health Sciences IT, Emory University, 1784 N. Decatur Rd, Atlanta, Georgia 30322; <sup>¶¶</sup>Biomolecular Frontiers Research Centre, Macquarie University, Sydney, NSW 2109, Australia; <sup>|||</sup>Structural Glycobiology Team, Riken Advanced Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan; <sup>a</sup>Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, Georgia 30602; <sup>b</sup>Beilstein-Institut, Trakehner Str. 7-9, 60487 Frankfurt am Main, Germany

Received December 5, 2012, and in revised form, February 1, 2013

Published, MCP Papers in Press, February 1, 2013, DOI 10.1074/mcp.O112.026492

the descriptions of experimental conditions in the Materials and Methods sections of many publications are often inadvertently or deliberately incomplete.

This issue has been recognized previously by diverse biological and biomedical initiatives that promote reporting standards for analytical data. These include MIAME (2), MIAPE (3), and STRENDA (4). To make it easier for authors to identify appropriate guidelines, a platform project called Minimum Information for Biological and Biomedical Investigations has been developed to provide descriptions for each guideline, including the type of information that is required in order to thoroughly report each particular experiment (5). The need for and success of these initiatives are clearly indicated by the fact that many of these guidelines are already recommended by journals, and the submission of these vital sets of information is often mandatory in order for a manuscript to be considered for publication.

However, the field of glycomics currently lacks such guidelines. This is likely partly because of the diverse number of preparative and analytical methods applied in characterizing glycans and differences in the intended depths of analyses. For example, protein-bound glycans such as *N*-linked or *O*-linked glycans require sample preparation steps that differ quite significantly from protocols used in the analysis of bacterial or plant polysaccharides. Glycans often have very complex structures that cannot be directly inferred from genomics data, as is frequently done to obtain protein sequences. Therefore, diverse analytical techniques are used for glycomics analyses, including those that exclusively utilize HPLC or MS and those that combine more than one method, such as LC-MS/MS analysis. In some cases, minimal information is obtained, as in some glycan mass profiling experiments (e.g. MALDI compositional analysis). In rare cases, detailed structure characterization is performed using NMR.

The application of these techniques can result in varying levels of structural information that, when combined with additional information, such as knowledge of the underlying biosynthetic pathways, often allows a defined structure to be proposed. However, the degree of structural definition and the assumptions that have been made in order to assign each structure are not always well reported.

In summary, the exact experimental conditions for sample preparation and analysis, in combination with the techniques and equipment used, have profound influences on the qualitative and quantitative results generated by a glycomics analysis. Therefore comprehensive description of conditions, techniques and results is required to enable researchers to evaluate and unambiguously interpret the results of these analyses and to reproduce them when necessary.

*The MIRAGE Project*—In 2009, at the Workshop on Analytical and Bioinformatic Glycomics, organized by the Consortium for Functional Glycomics, an international group of glycoscientists concluded that there is an urgent need for the standardization of data reporting in this area (6). Standardiza-



FIG. 1. **Process used within the MIRAGE project for the development of guidelines.** The multistep process established for this purpose includes drafting within the subgroups, refinement within the entire working group, and reviews by the advisory board. Finally, the scientific community is invited to comment so as to achieve broad agreement and minimize potential mistakes and misunderstandings.

tion is required in order to integrate glycomics data that are widely spread among diverse databases and thereby facilitate the development and application of bioinformatic tools for the analysis of these data. This initiative gained significant momentum when international leaders in the development of glycomics analysis techniques and software tools for glycoinformatics were joined by the editors of the major journals that publish glycomics and glycoproteomics research in expressing their willingness to support a standardization initiative. This resulted in the creation of the MIRAGE (Minimum Information Required for a Glycomics Experiment) initiative, led by experts in the fields of glycobiology, glycoanalytics, and glycoinformatics with the goal of creating minimum information guidelines for glycomics. The organization of this international group and their recent conclusions are published on the project website (<http://glycomics.cccr.uga.edu/MIRAGE/>). Membership is open for additional scientists who would like to participate in the work, and input from the scientific community is welcome. Additionally, proposals will be presented and discussed at the biennial Beilstein Symposia on Glyco-Bioinformatics (<http://www.beilstein-institut.de/en/symposia/overview/>).

Because glycobiology covers a wide range of different molecules and all the peculiarities of glycan sample preparation and analysis need to be considered, a new set of guidelines is being generated to address diverse information-reporting requirements. The working group has initiated the development of guidelines that take into account the generation, sampling, and storage of glycomics data obtained using MS. These guidelines are derived from the MIAPE-MS guidelines and have been extended to address issues that are unique to glycomics data. The initial version of these MIRAGE-MS guidelines has been reviewed and approved by the MIRAGE advisory board (Fig. 1) and has been made available online so

that the scientific community can offer further comments and refinement. Similar to the MIAPE and MIAMI concepts, MIRAGE identifies specific metadata that significantly increase the value of the associated experimental data. The MIRAGE guidelines will facilitate the collection of this information, for example, by stimulating the development of computational methods to automatically extract this information using software supplied by mass spectrometer manufacturers.

**MIRAGE-MS Guidelines**—The MIAPE committee has had a major effect on proteomics analysis by addressing diverse aspects related to the preparation, analysis, and identification of proteins. Despite the fact that similar methods and instruments are used in glycomics and that many instrument parameters are equally applicable to MS-based proteomics and glycomics, unique experimental requirements differentiate glycan analysis from protein analysis in several respects. These differences are related to the very distinct nature of glycan structures, chemistry, and biosynthesis. One of the major differences is that glycans undergo vibrational dissociation at lower energies than do peptides. For this reason, it is important to include information regarding the mass spectrometer settings used for glycan ion analysis. This information should include the ion source, ion transfer, and ion isolation appropriate for glycan classes. The necessity of the specification of this information is emphasized by the fact that glycans may be analyzed in native or derivatized forms, in positive or negative ionization modes, as cation or anion adducts, or as unattached/unlinked ions. The ability to extract and interpret structural information from the data in a reproducible manner depends on the accurate communication of information regarding sample and instrumental conditions. Therefore, it is particularly important to include minimal information regarding the experimental conditions used for MS-based glycomics.

The MIRAGE guidelines for mass spectral glycoanalysis are relevant to both database deposition and the submission of results to a journal. However, the MIRAGE guidelines are intended to be neither comprehensive nor absolute. In general, the deposition of data to a database requires highly formal parameter descriptions because of the necessity of controlling vocabulary, digital data formats, and other technical characteristics of the information. Nevertheless, restrictive vocabularies and explicit digital data formats are beyond the current scope of the MIRAGE guidelines for mass spectral data, as such requirements are best determined by the database developers. An example of the appropriate application of the MIRAGE guidelines for MS analysis would be the population of a specialized database with the mass spectra of well-characterized standard molecules (*i.e.* “gold standard” spectra). The usefulness of such data collection would most likely depend on compliance with the MIRAGE guidelines (*e.g.* reporting of instrument setup parameters) so that users of the data could design experiments to obtain spectra that were comparable to the standard spectra in the database. The

database developers would undoubtedly impose additional data submission requirements of their own. Conversely, journals are likely to have less formal requirements for data submission than those described in the MIRAGE guidelines. In this context, this journal’s guidelines for the submission of glycomics data serve as a good use case: many explicit details regarding instrumental setup parameters are not required, but the spirit of the MIRAGE guidelines is maintained. That is, the Journal requires information that will allow an expert to judge the quality of the results and to reproduce the overall conclusions of the reported study. Although not fully and formally implemented by the Journal, the MIRAGE guidelines can serve a critical role by prompting the analyst and the reviewer to consider experimental parameters that have a profound effect on the data and their interpretation.

The types of metadata recommended for reporting by the MIRAGE-MS guidelines are divided into five sections (Fig. 2). Overall, Sections 1–3 deal mainly with the instrumental hardware used to generate, fragment, and detect ions, whereas Sections 4 and 5 are focused on data interpretation and handling issues. Section 1, “General Features,” serves as the basis for the required metadata, with global descriptions on the used instrumentation, any particular customizations, and general instrument control parameters such as instrument control software.

Section 2, “Ion Sources,” continues to summarize all crucial parameters for ion generation such as controls of in-source fragmentation or the degree of prompt fragmentation, in addition to other, more common parameters (*e.g.* capillary voltage or laser intensity settings). Glycans contain several types of labile bonds, including bonds to fucose and sialic acid residues and to sulfate and phosphate substituents. It is thus very important to determine whether biologically or chemically significant ions observed in full scan mass spectra arise as a result of prompt fragmentation during the ionization process. The extent of prompt fragmentation can be established by examining data obtained using purified standard glycans, which allows one to demonstrate that the mass spectrometer is tuned properly for analysis of the glycan class in question. For example, if analyzing native *N*-glycans, one can show data obtained using a commercial sialylated *N*-glycan standard to show that sialic acid residues are not lost during ionization under the conditions used. Such data are important in order for readers and database users to evaluate the instrumental conditions under which data were acquired. The MIRAGE-MS guidelines do not require that the experimentalist perform an analysis of prompt fragmentation, but he or she must report whether such an analysis was done and, if so, provide the resulting data.

Section 3, “Ion Transfer and Post-source Components,” asks for instrumental details associated with the transport, gas phase reactions, and detection of ions once they are generated. This critical step is selective for the respective instrument(s) and thus is more difficult to generalize; therefore, a clear separation into major detection and ion transport

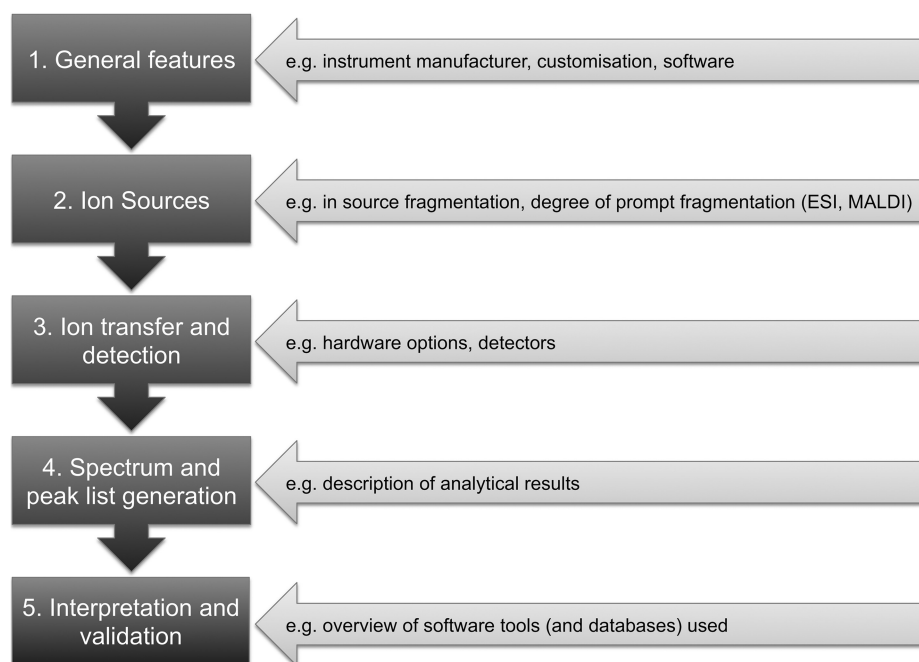


FIG. 2. The five major sections of the MIRAGE-MS guidelines. Sections 1–3 mainly deal with the instrumental hardware used to generate, fragment, and detect ions, and Sections 4 and 5 focus on data interpretation and handling issues.

categories currently found on the market has been introduced.

Section 4, “Spectrum and Peak List Generation and Annotation,” and Section 5, “Interpretation and Validation,” summarize the crucial parameters that form the basis of the analytical results that are generated after spectra have been recorded by the instrument. Detailed description of these parameters is vital because, to date, robust, widely distributed (and thus commonly used) tools for this step are not available. Parameters such as software, software customizations, and databases (if used) are overviewed in these sections. Quantitative aspects, which often play a considerable role in glycomics experiments, are also considered within the guidelines, and a set of parameters judged to be crucial is listed in Subsection 4.d.

Important aspects such as assignment validation and “deduced structure(s)” are specifically listed in order to further emphasize that the interpretation of many MS experiments is highly connected to the well-established (mammalian) glycosylation pathways, and therefore particular structural details are often inferred rather than confirmed via orthogonal techniques. This constitutes one of the major differences existing between MS-based glycomics and proteomics and is clearly reflected in the MIRAGE-MS guidelines. It is clear that mass profiling (without the use of tandem MS) can provide valuable information for both peptide and glycan samples. However, mass profiling of complex samples generated during bottom-up proteomics analysis can lead to unacceptably high false discovery rates, given the statistical probability that unrelated peptides might have the same mass. Thus, tandem MS sequence tags are often required in order for one to confidently identify a large number of distinct proteins in a

complex mixture. A similar state of affairs exists for glycomics experiments. Nevertheless, with certain limitations and caveats, glycomics analysis using single-stage MS can provide data that are adequate to answer a specific experimental question. Examples (e.g. modulation of the *N*-glycosylation of a very well-defined plant protein or control of the incorporation/absence of particular sugar residues such as fucose) clearly show that laborious in-depth experiments are not always required in order to answer a well-defined question (7, 8).

Overall, these MIRAGE-MS guidelines summarize a list of instrumental and experimental parameters that are considered critical in describing MS-related conditions for the acquisition and interpretation of glycoanalysis data. Based on these MIRAGE-MS guidelines, the committee is now working on developing guidelines for various other techniques and approaches (including sample preparation methods) that are commonly used in glycomics analysis. The lack of adequate and generally applicable software tools as utilized in proteomic research adds an additional challenge for glycoconjugate structural determination and structure reporting.

#### CONCLUSIONS

The MIRAGE-MS guidelines have been proposed in order to encourage authors, editors, and reviewers to gather and report all essential information describing a glycomics experiment that is being reported. The guidelines can be viewed in their entirety in the supplementary material or at the project web page. It is important to note that these guidelines are intended neither to dictate the use of particular methods (which should be decided by the experimentalist) nor to serve as a substitute for the review process. The goal of the guide-

lines is to provide a summary of the information describing an MS experiment at a level that allows it to be understood, evaluated, and reproduced. Furthermore, the guidelines provide authors with a framework and standard for defining the depth of structural analysis that supports the structural models reported in the manuscript. This is important for enabling both expert glycoscientists and readers who are less familiar in this area to understand the conclusions of the publication based on a rigorous and comprehensive description of the materials and methods used and the results obtained. As the reader depends on the judgment of the reviewers to set high standards for the publication of glycoanalytic data, this information is absolutely required in order for the reviewers to evaluate the results reported for each experiment described in the manuscript.

\* In 2011, the Beilstein-Institut, a non-profit foundation located in Frankfurt am Main, Germany, assumed financial responsibility for MIRAGE and the advancement of this standardization initiative. This work has also been supported by several other funding agencies. These include the Max Planck Society (D. Kolarich and E. Rapp), the European Union's Seventh Framework Programme (FP7-Health-F5-2011) under Grant No. 278535 (E. Rapp), the Biotechnology and Biological Sciences Research Council (Grant Nos. SF19107, B19088, and BBF008309/1 to S. Haslam), and the Australian government through the project NeCTAR, financed by the Education Investment Fund (M. Campbell). The authors also acknowledge NIGMS (National Institutes of Health) Grant No. P41GM104603 (J. Zaia) and funding received from the NIGMS-funded National Center for Glycomics and Glycoproteomics (National Institutes of Health) (Grant No. 8P41GM103490 to M. J. Pierce and W. S. York).

° To whom correspondence should be addressed: William S. York, Tel.: +1 706 542 4628; FAX: +1 706 542 4412; E-mail: will@ccrc.uga.edu.

#### REFERENCES

- Committee on Assessing the Importance and Impact of Glycomics and Glycosciences (Walt, D. A., Aoki-Kinoshita, A. F., Bendiak, B., Bertozzi, C. R., Boons, G.-J., Darvill, A., Hart, G., Kiessling, L. L., Lowe, J., Moon, R. J., Paulson, J. C., Sasisekharan, R., Varki, A. P., and Wong, C.-H.) (2012) *Transforming Glycoscience: A Roadmap for the Future*, The National Academies Press Washington, D.C.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371
- Taylor, C. F., Paton, N. W., Lilley, K. S., Binz, P. A., Julian, R. K., Jr., Jones, A. R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E. W., Dunn, M. J., Heck, A. J., Leitner, A., Macht, M., Mann, M., Martens, L., Neubert, T. A., Patterson, S. D., Ping, P., Seymour, S. L., Souda, P., Tsugita, A., Vandeckerkhove, J., Vondriska, T. M., Whitelegge, J. P., Wilkins, M. R., Xenarios, I., Yates, J. R., 3rd, and Hermjakob, H. (2007) The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **25**, 887–893
- Apweiler, R., Cornish-Bowden, A., Hofmeyr, J.-H. S., Kettner, C., Leyh, T. S., Schomburg, D., and Tipton, K. T. (2005) The importance of uniformity in reporting protein-function data. *Trends Biochem. Sci.* **30**, 11–12
- Taylor, C. F., Field, D., Sansone, S. A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C. A., Binz, P. A., Bogue, M., Booth, T., Brazma, A., Brinkman, R. R., Michael Clark, A., Deutsch, E. W., Fiehn, O., Fostel, J., Ghazal, P., Gibson, F., Gray, T., Grimes, G., Hancock, J. M., Hardy, N. W., Hermjakob, H., Julian, R. K., Jr., Kane, M., Kettner, C., Kinsinger, C., Kolker, E., Kuiper, M., Le Novère, N., Leebens-Mack, J., Lewis, S. E., Lord, P., Mallon, A. M., Marthandan, N., Masuya, H., McNally, R., Mehrle, A., Morrison, N., Orchard, S., Quackenbush, J., Reecy, J. M., Robertson, D. G., Rocca-Serra, P., Rodriguez, H., Rosenfelder, H., Santoyo-Lopez, J., Scheuermann, R. H., Schober, D., Smith, B., Snape, J., Stoeckert, C. J., Jr., Tipton, K., Sterk, P., Untergasser, A., Vandesompele, J., and Wiemann, S. (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat. Biotechnol.* **26**, 889–896
- Workshop on Analytic and Bioinformatic Glycomics*, April 16–18, 2009, Bethesda, MD
- Kolarich, D., and Altmann, F. (2000) N-Glycan analysis by matrix-assisted laser desorption/ionization mass spectrometry of electrophoretically separated nonmammalian proteins: application to peanut allergen Ara h 1 and olive pollen allergen Ole e 1. *Anal. Biochem.* **285**, 64–75
- Rillahan, C. D., Antonopoulos, A., Lefort, C. T., Sonon, R., Azadi, P., Ley, K., Dell, A., Haslam, S. M., and Paulson, J. C. (2012) Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat. Chem. Biol.* **8**, 661–668