OXFORD

CarbArrayART: a new software tool for carbohydrate microarray data storage, processing, presentation, and reporting

Yukie Akune[1],* (iD), Sena Arpinar[2], Lisete M Silva[1,3] (iD), Angelina S Palma[4,5] (iD),
Virginia Tajadura-Ortega[1], Kiyoko F Aoki-Kinoshita[6] (iD), René Ranzinger[2], Yan Liu[1] (iD), Ten Feizi[1] (iD)

[1]Glycosciences Laboratory, Department of Metabolism, Digestion and Reproduction, Imperial College, Du Cane Road, London W12 0NN, United Kingdom, [2]Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd, Athens, GA 30602, United States, [3]LAQV-REQUIMTE, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal, [4]UCIBIO, Applied Molecular Biosciences Unit, Department of Chemistry, School of Science and Technology, NOVA University Lisbon, 2819-516 Caparica, Portugal, [5]Associate Laboratory i4HB-Institute for Health and Bioeconomy, School of Science and Technology, NOVA University Lisbon, Lisbon, 2819-516 Caparica, Portugal, [6]Glycan and Life Systems Integration Center (GaLSIC), Soka University, 1-236 Tangi-machi, Hachioji, Tokyo 192-8577, Japan

*Corresponding author: The Glycosciences Laboratory, Burlington Danes Building Room 509, Du Cane Road, London W12 0NN, United Kingdom.
Email: y.akune@imperial.ac.uk

Glycan microarrays are essential tools in glycobiology and are being widely used for assignment of glycan ligands in diverse glycan recognition systems. We have developed a new software, called Carbohydrate microArray Analysis and Reporting Tool (CarbArrayART), to address the need for a distributable application for glycan microarray data management. The main features of CarbArrayART include: (i) Storage of quantified array data from different array layouts with scan data and array-specific metadata, such as lists of arrayed glycans, array geometry, information on glycan-binding samples, and experimental protocols. (ii) Presentation of microarray data as charts, tables, and heatmaps derived from the average fluorescence intensity values that are calculated based on the imaging scan data and array geometry, as well as filtering and sorting functions according to monosaccharide content and glycan sequences. (iii) Data export for reporting in Word, PDF, and Excel formats, together with metadata that are compliant with the guidelines of MIRAGE (Minimum Information Required for A Glycomics Experiment). CarbArrayART is designed for routine use in recording, storage, and management of any slide-based glycan microarray experiment. In conjunction with the MIRAGE guidelines, CarbArrayART addresses issues that are critical for glycobiology, namely, clarity of data for evaluation of reproducibility and validity.

Key words: glycan microarray data management; glycan microarray data presentation; glycan microarray data storage; glycan microarray software; glycoinformatics.

Introduction

Since their inception in 2002, microarrays of sequence-defined glycans (Fukui et al. 2002) have become essential tools in biology and medicine, particularly in elucidating glycan-binding specificities of antibodies, other proteins of the immune system, adhesins of microbial agents, and diverse other carbohydrate-recognizing systems (Rillahan and Paulson 2011; Li and Feizi 2018; Gao et al. 2019; Geissner et al. 2019; Xia and Gildersleeve 2019; Silva et al. 2021). The number and structural diversity of sequence-defined glycans that can be spotted on microarrays are increasing. Currently, they are approaching 1,000 glycans in the largest glycan microarray facilities (Glycosciences Laboratory website https://glycosciences.med.ic.ac.uk; Consortium for Functional Glycomics (CFG) website http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp; National Center for Functional Glycomics (NCFG) https://ncfg.hms.harvard.edu/microarrays).

Beyond data recording, it is desirable to have dedicated software for storage, processing, and display, which includes sorting and filtering functions for the glycan probes according to structural features of those bound or not bound by particular recognition systems. Unique software tools with such functionalities were developed as prototypes (Stoll and Feizi 2009). Indeed, these have been the mainstay of data storage,

presentation, and reporting in the Glycosciences Laboratory at Imperial College London (160 published and over 8,000 internal data sets). However, these software tools were developed over the years, in stages, using Microsoft Office, in house Microsoft Access databases and Visual Studio; as such, they were not readily distributable.

Encompassing the above-mentioned functionalities, we have been developing a new distributable software package using the Java programming language called Carbohydrate micro-Array Analysis and Reporting Tool (CarbArrayART). This is geared for day-to-day use by experimentalists working on slide-based glycan microarrays. An important feature of CarbArrayART is that it allows users to store centrally glycan microarray data within their laboratory with ease of retrieval, comparison, and mining of data generated. This includes compliance with the glycan microarray guidelines for Minimum Information Required for A Glycomics Experiment [(MIRAGE) (Liu et al. 2016)]. The guidelines were authored by a team of experts for generating interpretable data from a glycan microarray experiment.

CarbArrayART benefits from the GRITS Toolbox, a software package initially developed for storing, processing, and visualizing MS data (Weatherly et al. 2019). GRITS Toolbox allows users to record metadata, such as project and collaborator information, sample metadata, and experimental
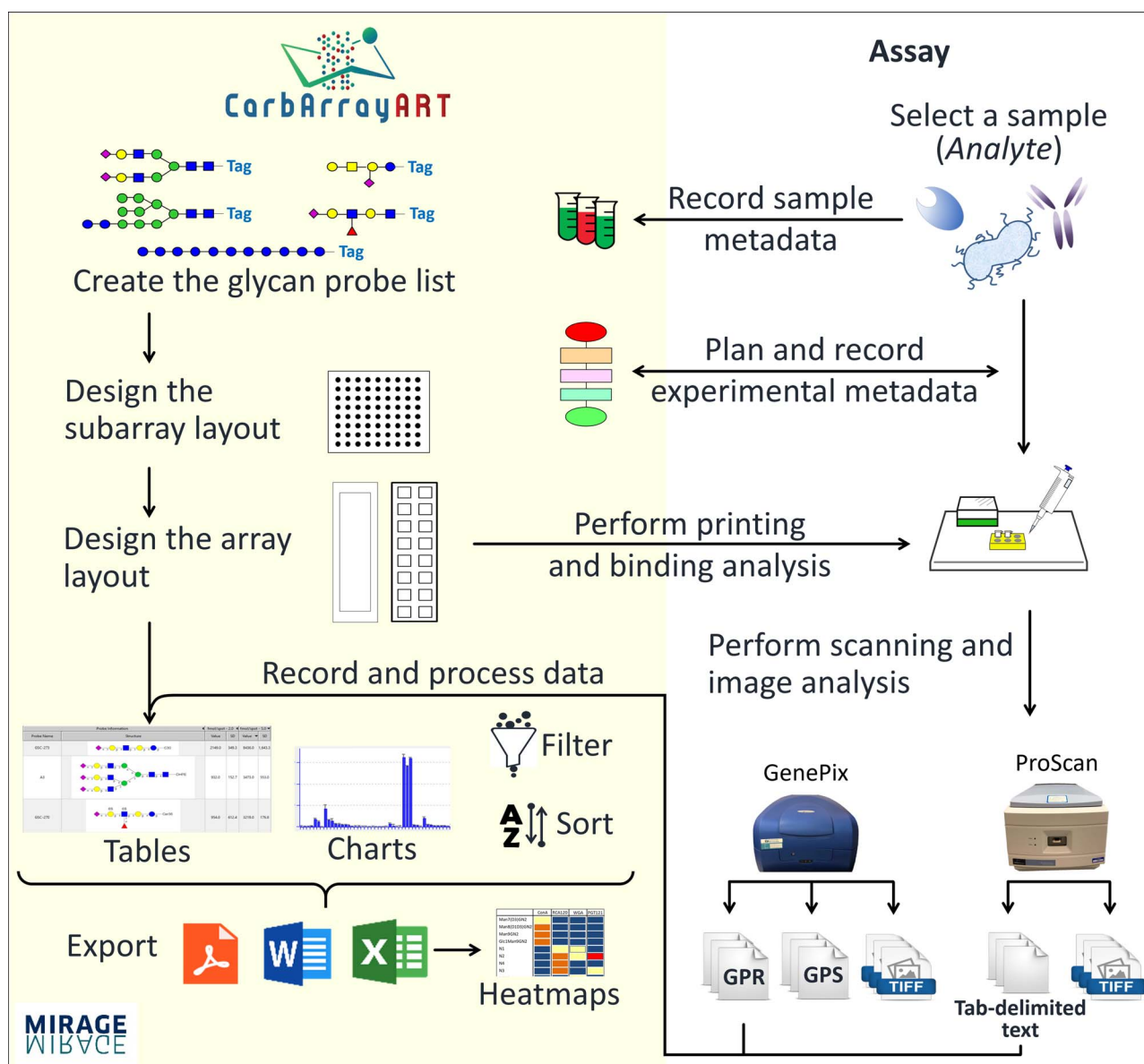
Fig. 1. Schematic representation of the carbohydrate microarray analysis and reporting tool, CarbArrayART.

protocol as well as storage related data files as archives. All these functionalities are adopted as an integral part of the CarbArrayART data storage and management system for glycan microarrays. To these, we have incorporated microarray-specific features that enable storing, processing, interpretation, and reporting of array data (Fig. 1). These include glycan probe list, array geometry information, assay conditions, and quantified array data files such as GenePix Result files, commonly referred to as GPR. Thus, CarbArrayART is the first distributable software tool that accommodates storage at a laboratory level of glycan microarray data and metadata with easy retrieval, comparison, mining, and sharing of data generated. The information generated with CarbArrayART is eminently usable by bioinformaticians and biologists using glyco-informatics tools that are discussed below.

CarbArrayART architecture
Workspace, project, and analyte

Workspace, project, and analyte are compartments derived from the GRITS Toolbox. When the software is run initially, users are required to create a workspace folder in which the microarray data will be saved. Project is a user-defined folder within the workspace, where microarray data files of a single or multiple analytes and the corresponding metadata are saved. For example, in the entry page for a glycan-binding sample (Analyte) users can record sample metadata compliant with the MIRAGE Glycan Array guidelines. These include the origin of the sample (synthetic, natural, or recombinant), database-associated information (e.g. PDB ID), purity, and quality control information on the sample. In the experiment design tool, users can store the experimental procedures (e.g. conditions used for different incubation and washing steps).

Glycan probes and array geometry

Microarray laboratories have differing approaches for preparing glycan microarrays (Oyelaran and Gildersleeve 2009; Liu et al. 2012; Wang et al. 2014). CarbArrayART has been designed to cater for different slide-based array and file formats.

Glycan probe is the term used to define an arrayed glycan, its sequence, as well as the tag moiety, if applicable. Glycan sequence representations that can be used in CarbArrayART include 2D TEXT (Stoll and Feizi 2009), CFG-IUPAC, GlycoWorkbench Sequence (GWS) (Ceroni et al. 2008; Damerell et al. 2012), Web3 unique representation of carbohydrate structures (WURCS) (Matsubara et al. 2017), and the currently recommended machine-readable format GlycoCT (Herget et al. 2008). Glycans can also be imported using GlyTouCan ID (Fujita et al. 2021) and retrieving their sequence information from that repository.

Tag information is stored separately from the glycan moiety information. For example, users can designate a tag to denote its feature such as Cer32 and Cer42, which are synthetic glycolipids with ceramide having 32 and 42 carbon atoms, respectively. Inclusion the tag information facilitates comparisons of signals elicited by a given glycan sequence with different tags appended; examples are GSC-16 (NeuAcα-3Galß-4Glcß-Cer32) and GSC-18 (NeuAcα-3Galß-4Glcß-Cer42). For arrays of glycans with sequences undefined or arrays of other types of glycoconjugates, the name, source, and other informative features can be entered.

There are entry tools for: (i) generating the glycan probe list (Glycan Glyco-probe); (ii) designing the layout of spots on a block ("Subarray Layout"); and (iii) arranging positions of subarrays on a slide ("Array Layout"). There is an additional entry tool for array geometry using an Excel file, which contains the block number, spot numbers, and printing conditions for each glycan probe such as the concentrations in arrayed solution or dose arrayed per spot. This is an extended format of GenePix Array List (often referred to as a GAL file, https://www.moleculardevices.com/en/assets/app-note/br/genepix-array-list-gal-files#gref). A template and an example of this is included in the CarbArrayART software package.

Data entry, processing, and storage

Binding signals are acquired with microarray scanners, such as ProScanArray microarray scanner (PerkinElmer) that generates a tab-delimited text file, or GenePix Microarray Scanner (Molecular Devices) that generates GenePix Result (GPR) file. In addition, GenePix Settings (GPS) are acquired, which include parameters such as the PMT voltage, scan area, identification of blocks, brightness, and contrast settings. (https://mdc.custhelp.com/app/answers/detail/a_id/18883/~/genepix%C2%AE-file-formats). At the same time, the TIFF images of the slides are acquired.

In CarbArrayART, the scan result files such as tab-delimited file and GPR are linked to the array geometry information to generate the processed data for presentation. Other files such as GPS and TIFF image files are stored as archived data.

Data presentation and reporting

During the image scanning, the scanner software acquires the fluorescence intensity value for each spot including mean, median, mean minus background, and median minus background. The acquired intensity values for multiple spots of each glycan probe, which are recorded in the scan result file, are averaged. Users can select the averaging method: either the mean of the replicated spots or the mean after eliminating outliers.

In the results page of CarbArrayART, the processed data, which are generated based on the scan result file and array geometry using the selected averaging method, are displayed as tables and charts.

Users can filter and sort the binding signals based on monosaccharides, or other features, e.g. sialyl linkage and oligosaccharide motifs stored in CarbArrayART. Currently, 69 types of monosaccharides and 60 types of substructures and motifs have been adapted from the Symbol Nomenclature for Glycans (Neelamegham et al. 2019) and our previous software tools.

The processed data and metadata are exportable as Excel, PDF, and Word, together with sample and experimental metadata. The exported tabulations can be used for generating heatmaps using the template file included in the CarbArrayART software package.

Discussion and perspectives

Several web resources and software tools for glycan microarray data have been reported. Glycan Array Dashboard (GLAD) (Mehta and Cummings 2019) provides online tools to display glycan microarray results including glycans with relative fluorescence units in bar charts and heatmaps to allow users to look for glycan-binding motifs after sorting. Glycan microarray databases such as DAGR (Sterner et al. 2016), MCAW-DB (Hosoda et al. 2018), GlyMDB (Cao et al. 2020), and CarboGrove (Klamer et al. preprint posted. doi: 10.1101/2021.11.12.468378) store glycan determinants for diverse recognition systems. MotifFinder (Klamer et al. 2021) is a software tool for predicting glycan-binding motifs by introducing a new text-based glycan presentation method and algorithm for data mining. Collectively, these web resources and software tools are important for the glycan microarray community and are highly complementary with CarbArrayART.

CarbArrayART differs in being the first distributable software tool, which accommodates storage at a laboratory level of glycan microarray data and metadata with easy retrieval, comparison, mining, and sharing of data generated.

Clarity, reproducibility, and validity of glycan microarray analysis data are critical topics in glycobiology. CarbArrayART is designed to address these in conjunction with MIRAGE guidelines. The glycan-binding specificities thereby assigned pave the way to detailed studies to establish specific glycan sequences as players in molecular recognition events, for example, in many aspects of cell signaling and cell behavior, and in the initiation of infections and triggering of immunity.

In its present form, CarbArrayART is geared for slide-based glycan microarray experiments using sequence defined glycans as well as glycan fractions in glycomic-scale microarray analysis. In the future, we anticipate accommodating other microarray formats such as glycan bead array (Purohit et al. 2018), next-generation glycan microarray (Yan et al. 2019), competitive universal proxy receptor assay (Kitov et al. 2019),

and liquid glycan array (Sojitra et al. 2021). Future versions of CarbArrayART, we will also include text file formats such as .csv and .xml in the export functions to enable the data to be processed further.

The first glycan microarray repository is being developed within GlyGen (York et al. 2020) with support from the NIH Glycoscience Common Fund. GlyGen is the computational and informatics resource for glycoscience to integrate data and knowledge from diverse disciplines relevant to glycobiology. Under development are new features, whereby CarbArrayART will serve as a vehicle for uploading and downloading data to and from the glycan microarray repository.

Funding

Conflict of interest statement: None declared.

Availability and software for download

CarbArrayART operable in Windows 7 (64bit), Windows 10, Mac OS X Lion (10.7.5), Mac OS X Yosemite (10.10.5) and Mac OS Big Sur (11.1).

Software with online user's manual is accessible from http://carbarrayart.org.

References

Cao Y, Park SJ, Mehta AY, Cummings RD, Im W. GlyMDB: glycan microarray database and analysis toolset. *Bioinformatics*. 2020:**36**: 2438–2442.

Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWork-bench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 2008:**7**:1650–1659.

Damerell D, Ceroni A, Maass K, Ranzinger R, Dell A, Haslam SM. The glycan builder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol Chem*. 2012:**393**: 1357–1362.

Fujita A, Aoki NP, Shinmachi D, Matsubara M, Tsuchiya S, Shiota M, Ono T, Yamada I, Aoki-Kinoshita KF. The international glycan repository GlyTouCan version 3.0. *Nucleic Acids Res*. 2021:**49**:D1529–D1533.

Fukui S, Feizi T, Galustian C, Lawson AM, Chai W. Oligosac-charide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nat Biotechnol*. 2002:**20**:1011–1017.

Gao C, Wei M, McKitrick TR, McQuillan AM, Heimburg-Molinaro J, Cummings RD. Glycan microarrays as chemical tools for identify-ing glycan recognition by immune proteins. *Front Chem*. 2019:**7**: 833.

Geissner A, Reinhardt A, Rademacher C, Johannssen T, Monteiro J, Lepenies B, Thépaut M, Fieschi F, Mrázková J, Wimmerova M, et al. Microbe-focused glycan array screening platform. *Proc Natl Acad Sci U S A*. 2019:**116**:1958–1967.

Herget S, Ranzinger R, Maass K, Lieth CW. GlycoCT—a unifying sequence format for carbohydrates. *Carbohydr Res*. 2008:**343**: 2162–2171.

Hosoda M, Takahashi Y, Shiota M, Shinmachi D, Inomoto R, Higashimoto S, Aoki-Kinoshita KF. MCAW-DB: a glycan profile database capturing the ambiguity of glycan recognition patterns. *Carbohydr Res*. 2018:**464**:44–56.

Kitov PI, Kitova EN, Han L, Li Z, Jung J, Rodrigues E, Hunter CD, Cairo CW, Macauley MS, Klassen JS. A quantitative, high-throughput method identifies protein–glycan interactions via mass spectrometry. *Commun Biol*. 2019:**2**:1–7.

Klamer ZL, Haab BB. Combined analysis of multiple glycan-array datasets: new explorations of protein–glycan interactions. *Anal Chem*. 2021:**93**:10925–10933.

Liu Y, Childs RA, Palma AS, Campanero-Rhodes MA, Stoll MS, Chai W, Feizi T. Neoglycolipid-based oligosaccharide microarray system: preparation of NGLs and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. *Methods Mol Biol*. 2012:**808**:117–136.

Liu Y, McBride R, Stoll M, Palma AS, Silva L, Agravat S, Aoki-Kinoshita KF, Campbell MP, Costello CE, Dell A, et al. The mini-mum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting glycan microarray-based data. *Glycobiology*. 2016:**27**:280–284.

Matsubara M, Aoki-Kinoshita KF, Aoki NP, Yamada I, Narimatsu H. WURCS 2.0 update to encapsulate ambiguous carbohydrate structures. *J Chem Inf Model*. 2017:**57**:632–637.

Mehta AY, Cummings RD. GLAD: GLycan Array dashboard, a visual analytics tool for glycan microarrays. *Bioinformatics*. 2019:**35**: 3536–3537.

Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lüt-teke T, O'Boyle N, Packer NH, Stanley P, Toukach P, et al. Updates to the symbol nomenclature for glycans guidelines. *Glycobiology*. 2019:**29**:620–624.

Oyelaran O, Gildersleeve JC. Glycan arrays: recent advances and future challenges. *Curr Opin Chem Biol*. 2009:**13**:406–413.

Purohit S, Li T, Guan W, Song X, Song J, Tian Y, Li L, Sharma A, Dun B, Mysona D, et al. Multiplex glycan bead array for high throughput and high content analyses of glycan binding proteins. *Nat Commun*. 2018:**9**:1–12.

Rillahan CD, Paulson JC. Glycan microarrays for decoding the glycome. *Annu Rev Biochem*. 2011:**80**:797–823.

Silva LM, Correia VG, Moreira ASP, Domingues MRM, Ferreira RM, Figueiredo C, Azevedo NF, Marcos-Pinto R, Carneiro F, Magalhaes A, et al. Helicobacter pylori lipopolysaccharide structural domains and their recognition by immune proteins revealed with carbohy-drate microarrays. *Carbohydr Polym*. 2021:**253**:117350.

Sojitra M, Sarkar S, Maghera J, Rodrigues E, Carpenter EJ, Seth S, Ferrer Vinals D, Bennett NJ, Reddy R, Khalil A, et al. Genetically encoded multivalent liquid glycan array displayed on M13 bacteriophage. *Nat Chem Biol*. 2021:**17**:806–816.

Sterner E, Flanagan N, Gildersleeve JC. Perspectives on anti-glycan antibodies gleaned from development of a community resource database. *ACS Chem Biol*. 2016:**11**:1773–1783.

Stoll M, Feizi T. Software tools for storing, processing and displaying carbohydrate microarray data. In: *Proceeding of the Beilstein Sym-posium on Glyco-Bioinformatics*. 2009, p. 123–140.

Wang L, Cummings RD, Smith DF, Huflejt M, Campbell CT, Gilder-sleeve JC, Gerlach JQ, Kilcoyne M, Joshi L, Serna S, et al. Cross-platform comparison of glycan microarray formats. *Glycobiology*. 2014:**24**:507–517.

Weatherly DB, Arpinar FS, Porterfield M, Tiemeyer M, York WS, Ranzinger R. GRITS toolbox-a freely available software for pro-cessing, annotating and archiving glycomics mass spectrometry data. *Glycobiology*. 2019:**29**:452–460.

Xia L, Gildersleeve JC. Anti-glycan IgM repertoires in newborn human cord blood. *PLoS One*. 2019:**14**:e0218575.

Yan M, Zhu Y, Liu X, Lasanajak Y, Xiong J, Lu J, Lin X, Ashline D, Reinhold V, Smith DF, et al. Next-generation glycan microarray enabled by DNA-coded glycan library and next-generation sequenc-ing technology. *Anal Chem*. 2019:**91**:9221–9228.

York WS, Mazumder R, Ranzinger R, Edwards N, Kahsay R, Aoki-Kinoshita KF, Campbell MP, Cummings RD, Feizi T, Martin M, et al. GlyGen: computational and informatics resources for Glycoscience. *Glycobiology*. 2020:**30**:72–73.