

SOFTWARE TOOLS FOR STORING, PROCESSING AND DISPLAYING CARBOHYDRATE MICROARRAY DATA

MARK STOLL* AND TEN FEIZI#

The Glycosciences Laboratory, Imperial College London, Northwick Park Campus, Watford Road, Harrow, Middlesex HA1 3UJ, U.K.

E-Mail: * m.stoll@imperial.ac.uk, # t.feizi@imperial.ac.uk

Received: 1st March 2010 / Published: 10th December 2010

ABSTRACT

We describe a suite of software modules to store, retrieve and display carbohydrate microarray data. Storage is in a relational database that holds all the microarray data and associated glycan, protein and experimental information. The retrieval and display software has a comprehensive system of sorters, filters and arrangers to allow highly customized presentation of data as charts, tables, 2D matrices and array graphics. Matrices allow arrangement of proteins in one axis and glycans in the other, so that comparisons can be made between the binding patterns of proteins. Sorting and filtering includes a large assortment of built-in parameters that range from glycan features to data grouping in slides and experiments but may also be completely customized to suit individual needs. Charts, tables and matrices are customizable to maximize presentation clarity. There are customizable automatic chart titles, chart axis annotation and scaling, table layouts, matrix arrangements and colour schemes for all graphics. All display output from the software can be saved or printed for permanent record.

BACKGROUND

Over the past 25 years we have developed and have been using the neoglycolipid (NGL) technology for studying the interactions of glycan probes and carbohydrate binding proteins. The lipid linked probes generated have been invaluable for the discovery of novel glycan

ligands, many derived from natural sources and available in only minute amounts. Our laboratory has developed chromatographic and mass spectrometric methods for determining the sequences of the glycan ligands in their NGL forms [1–4].

Emergence of carbohydrate microarrays as powerful tools in biology and biochemistry

After gene and protein microarrays, carbohydrate microarrays have emerged and they are revolutionizing studies of carbohydrate-protein interactions, which are of fundamental biological importance in endogenous recognition systems and pathogen host interactions [5–8]. The advantages of microarrays over the conventional approach are parallel measurements of interactions involving thousands of samples using femtomole amounts of glycan probe. The NGL-based carbohydrate microarray system established in our laboratory at Imperial College is currently one of the two most comprehensive internationally and it includes a diverse repertoire of natural and synthetic saccharide probes (Figure 1), the other system being that of The Consortium for Functional Glycomics (CFG).

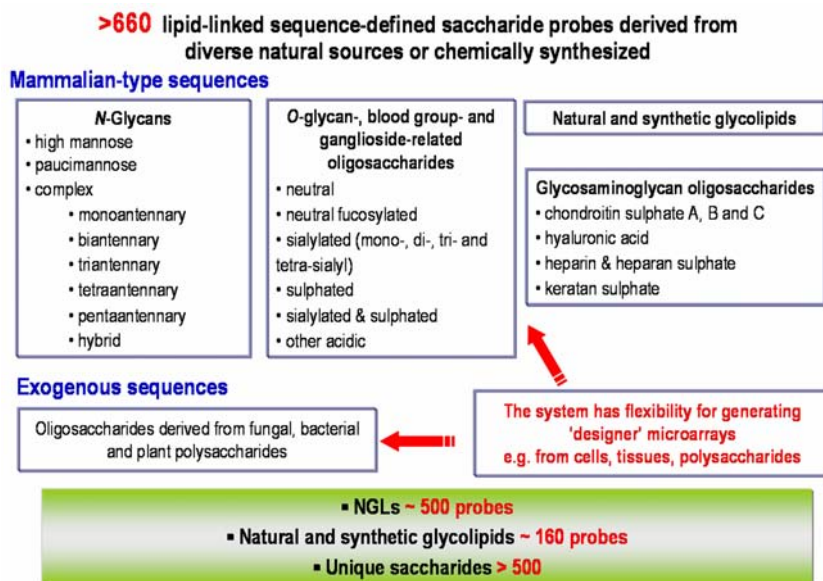


Figure 1. The Glycosciences Laboratory's repertoire of lipid-linked probes. The figure shows the range of different glycans: N-glycans, O-glycans, blood group- and ganglioside-related glycans, natural and synthetic glycolipids, glycosaminoglycans and glycans derived from fungi and bacteria.

Software for microarray data handling and storage

Software is required for carbohydrate microarray data analysis but the commercially available software packages for microarray analyses are designed for gene microarrays. The requirements for carbohydrate microarrays are very different because of the wide range of structural elements present in glycans. These include not only the different monosaccharide units and their substituents, *e.g.* sulphation, phosphorylation and acetylation, but also the linkages and anomeric configurations of the monosaccharides, branching patterns, structural motifs *e.g.* N-glycans, O-glycans, glycosaminoglycans, and presence of antigenic groupings *e.g.* blood group markers. Data formats for glycan storage are being developed by EuroCarbDB, www.eurocarbdb.org (*e.g.* GlycoCT and LINUCS), CFG, www.functionalglycomics.org (*e.g.* GLYDE-II, IUPAC format) and Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.jp/kegg *e.g.* KCF) to address the problem of glycan structural diversity. There is software available for assignments of structure from mass spectrometric (MS) analyses and for automatic annotation of MS spectra [9, 10]. There is also software for manipulation of glycan structure and for conversion to and from cartoon representations of glycans to storage formats such as GlycoCT and LINUCS (EuroCarbDB website), Glyde and IUPAC (CFG website, Links,) and KEGG. A good summary of the present state of glyco-bioinformatics is available [11]. Currently, there is no generally available software dedicated to the analysis, storage and presentation of microarray data at laboratory level of sufficient flexibility to be of general use to any scientist wishing to work in the field.

EuroCarbDB, KEGG and CFG have large on-line database resources in which glycans, proteins and microarray screening data are linked through several databases. A customizable stand-alone software suite that can be tailored to the exact needs of individual research laboratories is desirable. Data stored in such a stand-alone system could potentially be formatted for, linked to and included in international resources.

We have been developing a suite of specialized software tools to support our carbohydrate microarray screening program. This software is in constant use in our laboratory and is being continually modified and tailored to requirements as they arise. We find it an essential adjunct to the biochemical work to help with interpretations of binding specificities.

MICROARRAY SOFTWARE

In the early stages of development, processing of our microarray data was carried out manually using the (Microsoft) Excel spreadsheet. Obtaining graphical and tabular output from a single slide of data would often take days of work. Moreover the lack of flexibility meant that changes in the way information was presented involved much further work. There was clearly a need for software that could radically improve both throughput and flexibility in data handling. Custom designed software would provide both speed and flexibility and would allow analysis and presentation of data to our exact requirements.

Furthermore, by programming in-house, we are able to adjust the software to be in line with developments in the microarray technology and meet the ever-changing requirements for processing and presentation.

Rationale for software development

For rapid and flexible development, an integrated development environment was used, namely Microsoft Office. The Office suite is very familiar to most scientists both specialists and non-specialists so that the learning curve in using the software is relatively shallow. The use of the built-in Visual Basic for Applications (VBA) using the Office Object model and the use of VBA-wrapped Windows API calls for graphics routines has proved wholly adequate in terms of flexibility and performance. All data, including user-defined preferences, are stored in an ACCESS database (DB). EXCEL-based software both transfers data to the DB and retrieves data from it for analysis and display and both EXCEL and WORD are used to store tabular and graphical output. The flow of information is indicated in Figure 2.

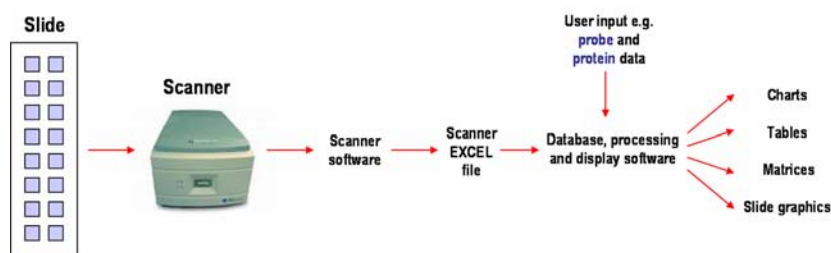


Figure 2. The flow of data from the microarray slide, scanner in EXCEL format, integration with probe and protein data via the input software, storage in an ACCESS database, retrieval and processing with the output software to give charts, tables, matrices and slide graphics.

Microarray format

The NGL microarray technology currently in use in our laboratory uses a screening format in which each glycan probe is used at two levels (2 and 7 (or 5) fmol) each in duplicate. Two fluorescent dyes are used, one (Cy3) to indicate positions of arrayed spots and the other (Alexa 647) to measure binding signals. Each microarray slide typically has the same 64 probes on the 16 nitrocellulose coated pads or two different sets of 64 probes on the two columns of pads. Thus 16 or 8 proteins, respectively, are overlaid for analysis of binding. Sixteen pads each with 64 glycan probes yields 8192 fluorescence signals (Figure 3). Over the past few years our laboratory has accumulated data from over 750 slides (> 9000 binding experiments). To cross-relate and interpret this data has required the development of specialized software for storage, analysis and display.

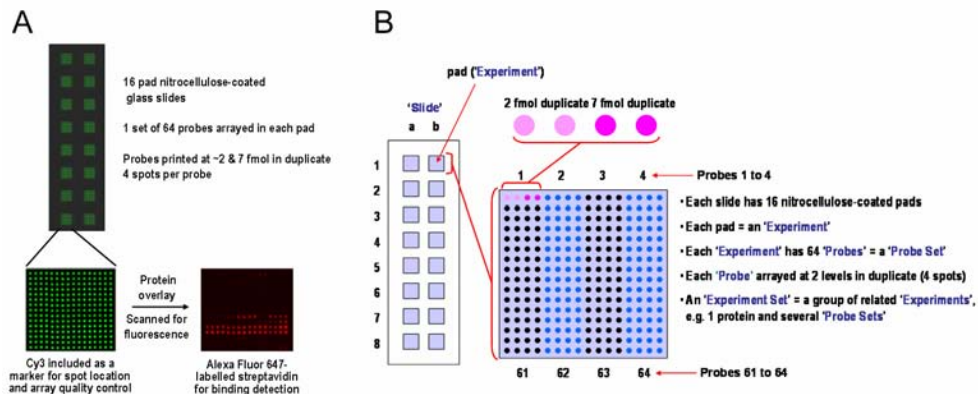


Figure 3. (A) Scanner images of a whole slide showing the Cy3 fluorescence of arrayed spots and a single experiment (nitrocellulose coated pad) showing both Cy3 fluorescence and the Alexa Fluor 647 binding signals. (B) A diagrammatic representation of a microarray slide showing the layout of probes in each experiment and the way 16 experiments are laid out in a slide. The arrangement of probes in 'Probe Sets' and 'Experiments' in 'Experiment Sets' is explained. The scanner reads the microarray slide left to right and top to bottom, one experiment at a time.

DATA INPUT SOFTWARE

Raw microarray data are returned by the scanner in EXCEL spreadsheet format. Each scanner file contains one slide of data but at between one and three laser power levels. Different laser powers are used to find optimum scan conditions to minimise noise for low level signals while avoiding saturation for high levels.

Data input

There is a graphical user interface (GUI) with an interactive slide graphic representing each experiment (16 per slide), to allow transfer of data to the DB from the laser scanner output software, together with associated data on the carbohydrate probes, carbohydrate binding proteins and experimental conditions, which may already be present in the DB associated with other microarray data or which may be added through the input interface as required (Figure 4).

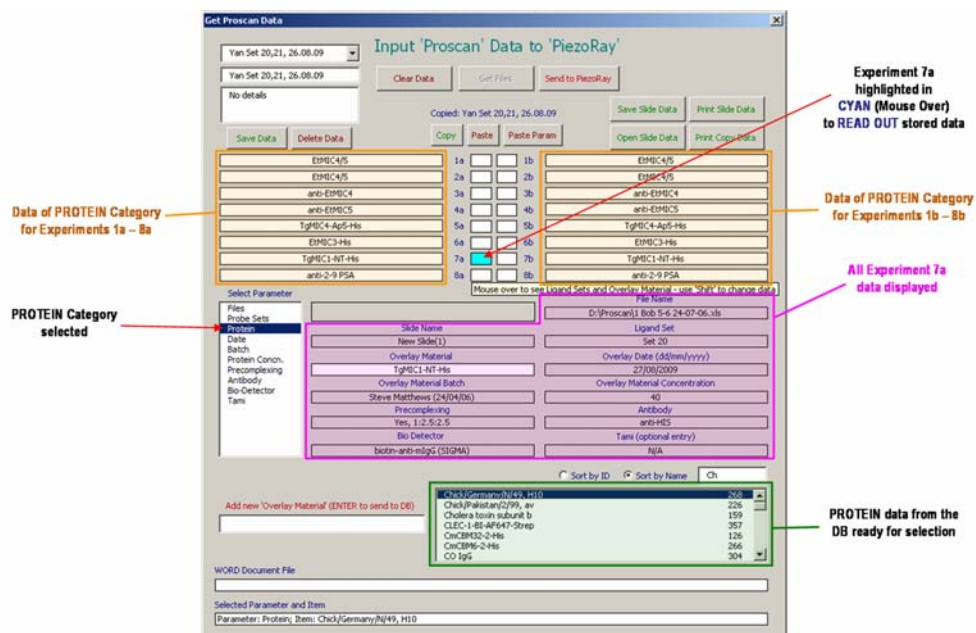


Figure 4. The user-interface of the data input software. The slide graphic is interactive and the mouse can be used to select each experiment either for data input or readout. Each parameter, when selected from the list of categories, reveals the items in the database for that parameter. Items can be selected from the list to associate with each experiment. When all the associations have been made the data are saved in the database.

Data storage

One DB holds all the data on the microarray experiments, carbohydrate probes, carbohydrate binding proteins and experimental conditions. In the DB all microarray data are stored in one table with references to all associated data stored in other tables. Each laser power level is treated as a separate slide of data for storage purposes. To maintain referential integrity the storage process is complex, involving two-way communication between the input software and the DB, which must generate unique reference identifiers for each new 'Slide' and 'Experiment'. Within the main microarray data table each record holds the four Cy3 and four Alexa Fluor 647 fluorescence values for one Glycan Probe/Protein interaction (four microarray spots) together with a slide position (1 – 1024; one for each glycan probe) and a reference each to an 'Experiment' and a Set of 64 Glycan Probes ('Probe Set'). Thus each slide is stored as 1024 records (16 experiments) with 64 records to each 'Experiment'. Glycan Probe data are stored as a text-based structure representation with both structural and immunological data stored in fields that can be used for sorting and filtering. At present we have not adopted *e.g.* the GlycoCT or GLYDE format for storing glycan information but intend to do this in future when standards are more established. In the DB, Glycan Probes

are associated in sets of 64 per ‘Experiment’ in a separate table and ‘Experiments’ are grouped as ‘Experiment Sets’ in yet another table. These associations are made by the user. Other tables hold protein and experimental conditions data all referenced to each ‘Experiment’. User-defined preferences are compressed and stored in special tables as custom character delimited strings. These handle custom sorters and filters, colour schemes, configuration settings etc. When required they are unpacked and used by the software. We use a central comprehensive DB but also smaller individual DBs for trial experiments that may or may not be transferred to the central DB.

Data retrieval

Our data analysis and presentation software (Figure 5) allows retrieval of microarray data from the DB in one of 5 formats; ‘Experiment’, ‘Experiment Set’, ‘Slide’, ‘Glycan Probe’ or ‘Protein’. The full content of the DB under each category is presented to the user as a list at program start-up. Any number of items from any one category can be requested at one time to become the current data to work with. There is also an interactive ‘Find’ feature that allows very specific data to be retrieved in ‘Experiment’ format.

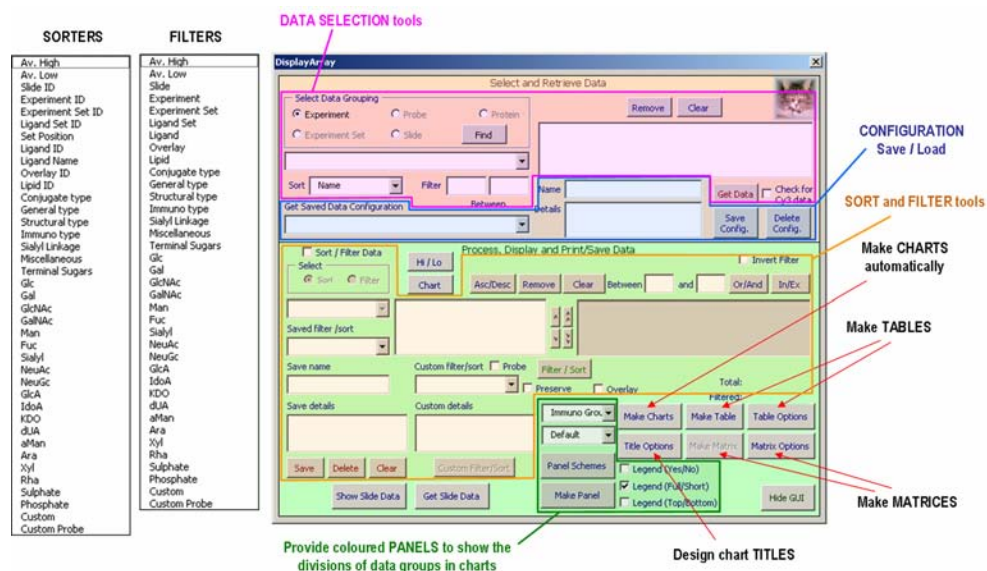


Figure 5. The user-interface of the data retrieval, processing, and presentation software. There are tools for data selection in one of 5 formats, for sorting and filtering data with a wide range of built-in as well as customizable parameters, for making tables, charts and matrices, for saving the state of the software to return to and for viewing data in slide graphic format. (The term ‘ligand’ refers to ‘glycan probe’).

The selected data are retrieved by using a complex SQL query so that all the fields (including calculated fields) that may be required by the software are available all at once. The returned Recordset Object (RO) is cloned into a Stream Object so that the data is held separately from the DB. A Collection Object also holds a number of smaller Recordsets used to populate lists used extensively by different elements of the software.

Data analysis and presentation

Charts: Charts are a major item in our data analysis and presentation and so much effort has gone into making these as versatile as possible. The fields used for building a histogram chart are copied from the RO to a hidden EXCEL worksheet linked to the software's interactive Chart Object and thus all the data is initially presented in a chart. Any element in the chart can be selected with the mouse to show all the associated data in a separate information window. The data can be scrolled through and/or expanded in, the chart window.

Automatic title generation: The title generator allows construction of saveable title templates from any combination of constant strings and variable parameters whose values are determined by the data present in the chart window. The title generator algorithm checks the data for each variable parameter in the template and looks for a unique title. If none is found a list of possible titles is presented and the user then has options for how to use the information *e.g.* concatenation can be used. There is an 'aliasing' system to allow substitution of variable data values with an alternative if the DB value is unsuitable in a title. There is also an option to condense Probe Set information for readability *e.g.* Set 1, Set 2, Set 4, Set 5, Set 6 would become Sets 1,2,4–6. The title generator operates automatically whenever the contents of the chart window changes, using the current template.

Sorting and filtering: The RO has efficient methods for sorting, filtering and bookmarking which the software uses through the user-interface to allow complex nested sorting and comprehensive filtering with any combination of AND and OR on built-in parameters. There are also special fields returned by the RO used to construct fully customized sorters and filters. The lists of built-in sorters and filters are shown in Figure 5.

Panels: To highlight divisions of data, coloured panels can be added behind the histogram elements of charts and legends may be added. There is a drop-down list of parameters to choose from and colour schemes can be constructed and saved. The panel generator operates by building a bitmap behind the chart elements, using the current colour scheme, after analyzing the chart data with respect to the parameter to be panelled. For example, if a fucose panel were used the panel colour for a particular histogram element would be set by the chosen colour scheme based on the number of fucose monosaccharides present on the

probe. In Figure 7 sialyl-linkage panels are used to show divisions between groups of glycan probes containing, 2-3, 2-6, 2-3/2-6, 2-8 and 2-9 sialyl linkages with the data sorted by a sialyl-linkage sorter.

Other options: There are choices to present averages of either or both the 7 (or 5) or 2 fmol data and the presentation of both probe (x) and fluorescence (y) axes. For example, the amplitude of the fluorescence axis can be chosen to have an absolute value or one related to the peak maximum in the chart data and the numbering of the probe axis can be selected to reflect position in the chart, position in the total retrieved data, position before applying sorters and filters or the relation to a master arrangement defined by a custom sorter. Filtered or unfiltered data can be excluded or shown highlighted for differentiation.

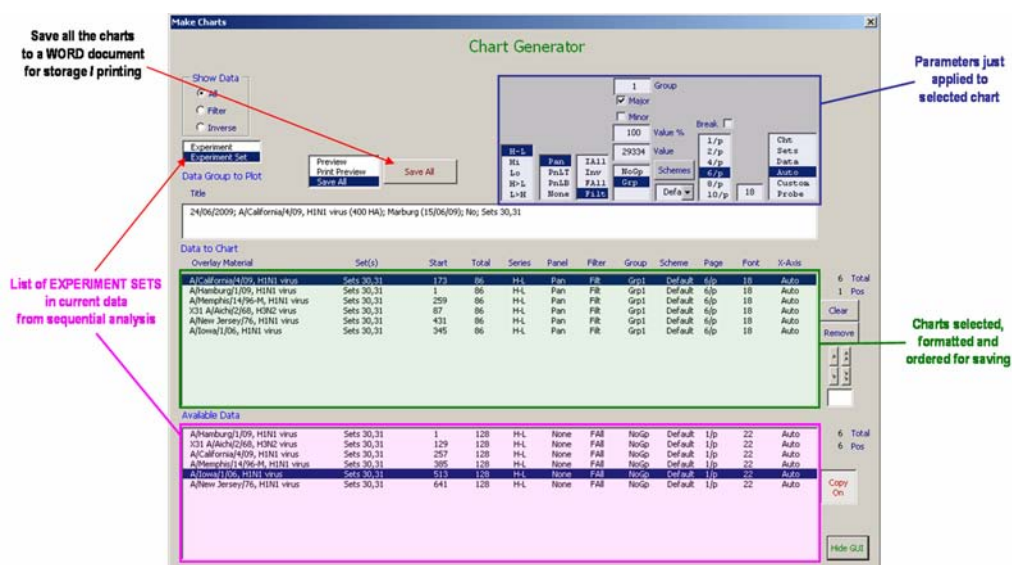


Figure 6. The user-interface of the chart generator. Data are selected from ‘Experiment’ or ‘Experiment Set’. The items under those headings show in the ‘Available Data’ window. Items of each type can be transferred in the required order to the ‘Data to Chart’ window where each item is formatted using the parameters shown in the top right of the interface. There is control over binding value presentation, data panels, filter presentation, y-axis values and grouping, chart colour schemes, number of charts per page, title font size and x-axis format. The chosen charts can be previewed in the chart window, printed directly or saved to a WORD document (Figure 7).

Automatic chart generation: There is a comprehensive interface to set-up automatic chart generation from the current data and sorter/filter configuration (Figure 6). There is a choice to divide data into ‘Experiments’ and/or ‘Experiment Sets’ and to apply individual presentational parameters to each chart chosen for saving or printing. All titles are generated using the current template but are editable before outputting. Charts are saved in WORD (Figure 7) as in-line enhanced metafiles. The size of each chart is determined by the charts per page

parameter applied to each chart before saving. The chart generator algorithm gathers all the parameters to be applied to make each chart in sequence and after hiding the chart window from view generates each chart in turn, converts a copy of it to a metafile and adds that into a new, formatted WORD document, which has been created by the software for reception of these objects. The SAVE AS dialogue is then presented to the user.

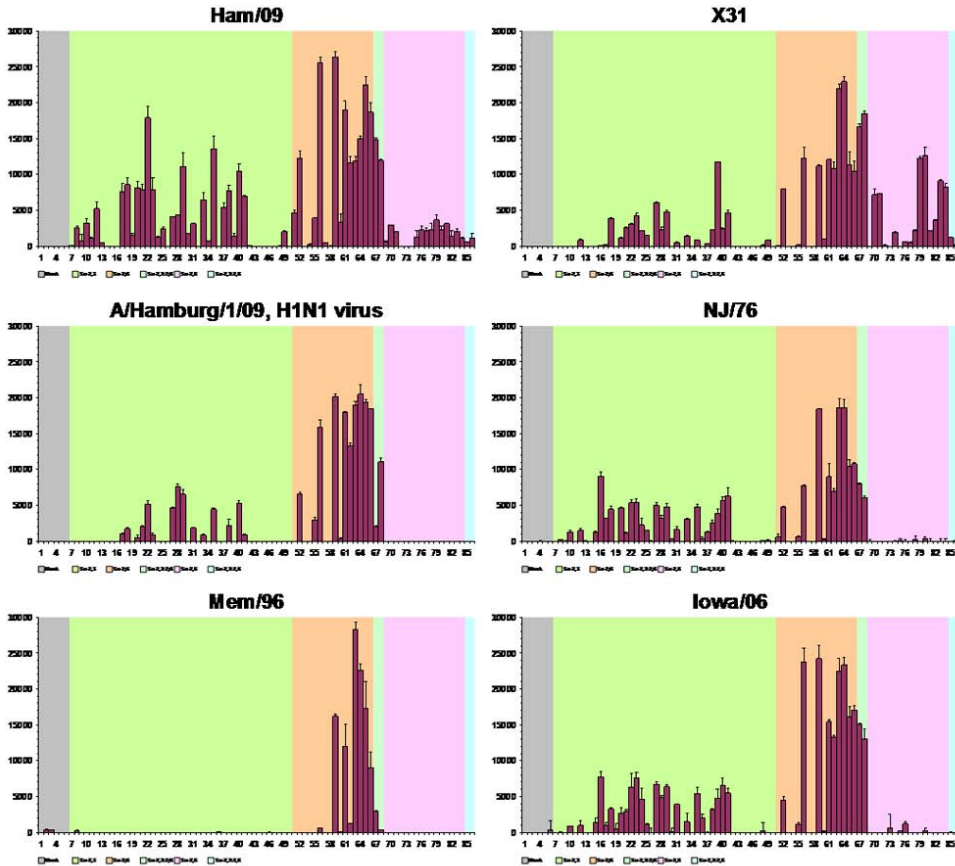


Figure 7. Six charts on a page produced by the chart generator with the settings shown in Figure 6. A sialyl-linkage sorter was used to group the glycan probes. The y-axis shows the fluorescence intensity of the binding signals from 6 influenza viruses. The legend shows the panel colours for non-sialylated and $\alpha 2-3$, $\alpha 2-6$, $\alpha 2-3/\alpha 2-6$, $\alpha 2-8$ and $\alpha 2-3/\alpha 2-8$ sialyl-linked probes. The titles are generated using a protein or virus name template with aliasing to give the wording shown. (taken from [12])

Tables: Tables can be generated from any concurrent region of the data using any selection and order of parameters from a list provided. Repetitions of parameters are allowed. Titles for columns are automatically supplied and the tables are available for immediate printing or copying and pasting to suitable containers. A useful feature is that a master table can be

constructed which can be referenced numerically from other data presentations so the table can be used to determine information for any data point in a chart, matrix or other table which may be a rearranged subset of the master table (Figures 8 and 9).

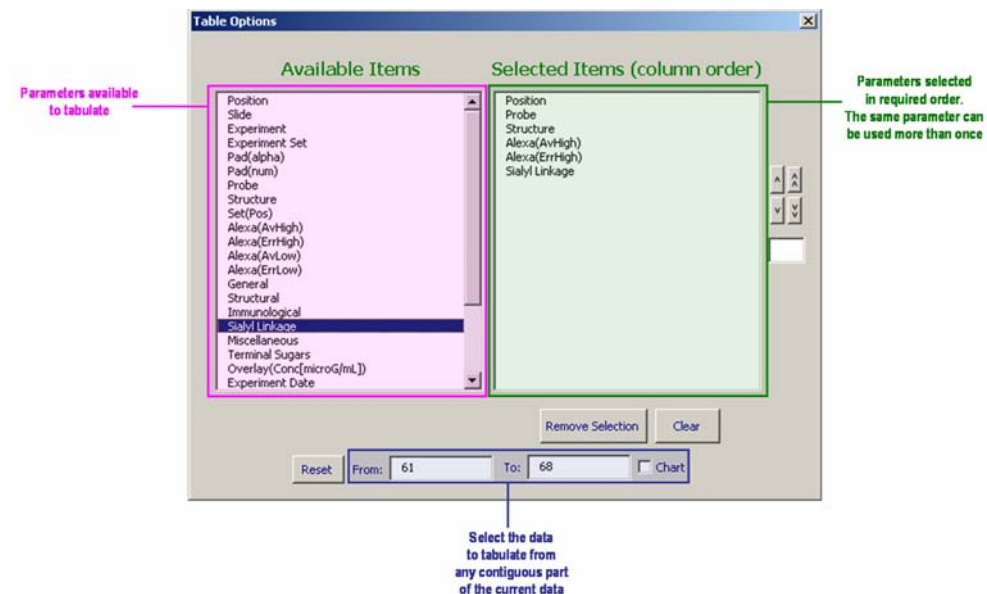


Figure 8. The user-interface of the table generator. ‘Available Items’ are selected from the list in any order. The data to tabulate can be that present in the chart window or any contiguous region of the current data. Column headings, as in Figure 9, are automatically generated.

Position	Probe	Structure	Alexa (AvHigh)	Alexa (ErrHigh)	Sialyl Linkage
61	SA(6')-LNFP-VI	$\begin{array}{c} \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-3Gal}\beta\text{-4Glc} \\ \\ \text{Fuc}\alpha\text{-3} \end{array}$	17951	100	2,6-Sialyl termini
62	MSLNH	$\begin{array}{c} \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-6} \\ \\ \text{Gal}\beta\text{-4Glc} \\ \\ \text{Gal}\beta\text{-3GlcNAc}\beta\text{-3} \end{array}$	13303.5	310.5	2,6-Sialyl termini
63	MSLNnH-I	$\begin{array}{c} \text{Gal}\beta\text{-4GlcNAc}\beta\text{-6} \\ \\ \text{Gal}\beta\text{-4Glc} \\ \\ \text{NeuAc}\alpha\text{-6Gal}\beta\text{-3GlcNAc}\beta\text{-3} \end{array}$	18948	568	2,6-Sialyl termini
64	DSLNNH	$\begin{array}{c} \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-6} \\ \\ \text{Gal}\beta\text{-4Glc} \\ \\ \text{NeuAc}\alpha\text{-6Gal}\beta\text{-3GlcNAc}\beta\text{-3} \end{array}$	20490	1339	2,6-Sialyl termini
65	MFMSLNnH	$\begin{array}{c} \text{Gal}\beta\text{-4GlcNAc}\beta\text{-6} \\ \quad \\ \text{Fuc}\alpha\text{-3} \quad \text{Gal}\beta\text{-4Glc} \\ \\ \text{NeuAc}\alpha\text{-6Gal}\beta\text{-3GlcNAc}\beta\text{-3} \end{array}$	19284	510	2,6-Sialyl termini
66	A2(2-6)	$\begin{array}{c} \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-2Man}\alpha\text{-6} \\ \quad \\ \text{Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc} \\ \\ \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-2Man}\alpha\text{3} \end{array}$	18423.5	13.5	2,6-Sialyl termini
67	DST	$\begin{array}{c} \text{NeuAc}\alpha\text{-3Gal}\beta\text{-3GalNAc} \\ \\ \text{NeuAc}\alpha\text{-6} \end{array}$	2076	130	2,3,2,6-Sialyl termini
68	A3	$\begin{array}{c} \text{NeuAc}\alpha\text{-3Gal}\beta\text{-4GlcNAc}\beta\text{-2Man}\alpha\text{-6} \\ \quad \\ \text{Man}\beta\text{-4GlcNAc}\beta\text{-4GlcNAc} \\ \\ \text{NeuAc}\alpha\text{-3Gal}\beta\text{-4GlcNAc}\beta\text{-4Man}\alpha\text{3} \\ \\ \text{NeuAc}\alpha\text{-6Gal}\beta\text{-4GlcNAc}\beta\text{-2} \end{array}$	11033.5	526.5	2,3,2,6-Sialyl termini

Figure 9. A table showing data between data points 61 and 68 inclusive of the current data (Figure 8). The items included are: position in the chart, glycan probe name, probe structure, average binding signal at 5 fmol, error in the 5 fmol duplicates and the 'Sialyl Linkage'.

Matrices: A matrix is defined as a 2D array plot with proteins along one axis and glycan probes along the other (Figures 10, 11). The unique feature is that probes of the same structure from different probe sets are lined up so that binding intensities for a given glycan structure can be easily visualized for each protein in turn. To achieve this, blank spaces are inserted where corresponding data are not available. Matrices are generated in an EXCEL worksheet and are fully interactive so that all data associated with any data point can be viewed in the information window. There are a variety of options including custom colour schemes, matrix- or protein-wide normalization, inclusion of extra parameters such as master table numbers and glycan structures and highlighting of specific data as desired using custom filters.

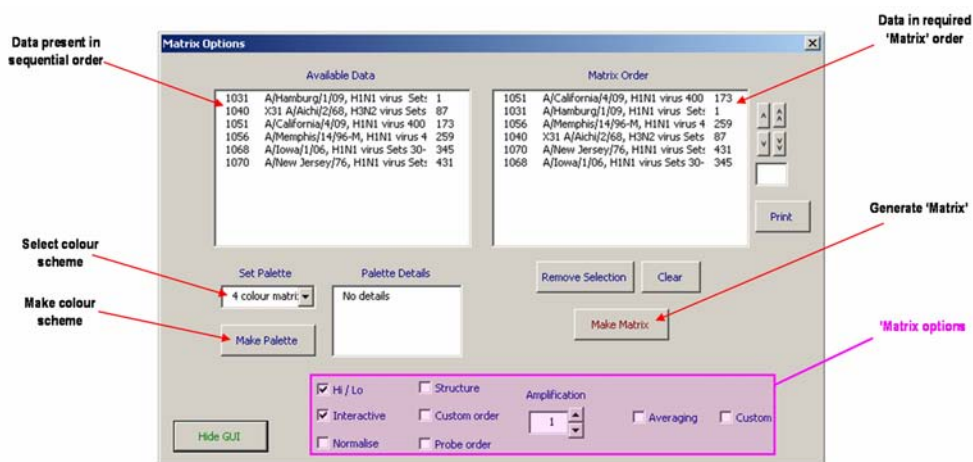


Figure 10. The user-interface of the 'matrix' generator. 'Available Data' are listed and items are transferred to and rearranged as required in 'Matrix Order'. A colour scheme is made or selected from those saved and options are selected. There is control of: data to use (5 or 2 fmo), normalization over all or each protein or using selected data points, highlighting of custom items, inclusion of probe structures and special reference position information. The matrix may be interactive if desired to show all data associated with any selected data point. The data are presented in an EXCEL worksheet as in Figure 11.

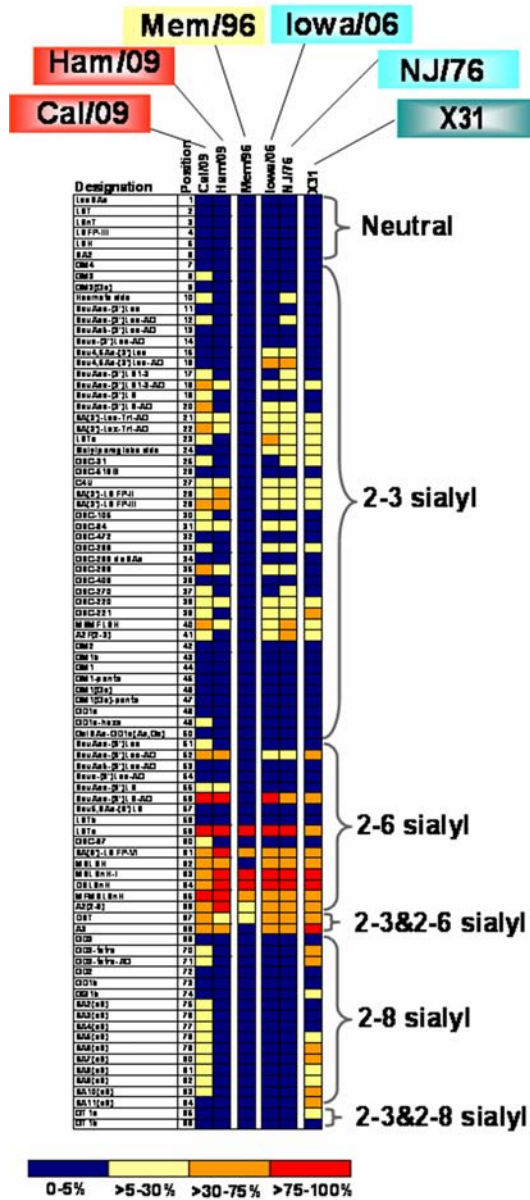


Figure 11. Matrix generated with the settings shown in Figure 10. It shows the same data as in Figure 8 but in a more compact form and more importantly could if necessary apply to data derived from different probe sets which would be more difficult to interpret from charts alone. The relative binding intensities in the matrix were calculated as the percentage of the fluorescence signal intensity at 5 fmol given by the probe most strongly bound by each virus. (taken from [12]).

Slide graphics: There is an option to view all slide data as a slide graphic similar to that produced by the scanner software with a magnified experiment graphic (Figure 12) such that the intensity of each data point is shown as a colour gradient from black to red for the binding signal or black to green for the reference signal. The scale can be selected from linear, quadratic or quartic. The latter two allow enhancement of lower binding levels. There is an option to show negative values in graduated black to blue. Normalization can be slide-wide or experiment-wide. The system is fully interactive showing all the data in the DB for any selected data point. The slide and experiment graphics can be captured as bitmaps for storage or printing. A particular use of this feature is as a way to check new data entry against the original picture of the slide generated by the scanner. The two patterns should correspond. It also offers a useful alternative way to view all the raw data from a slide in a compact interactive format.

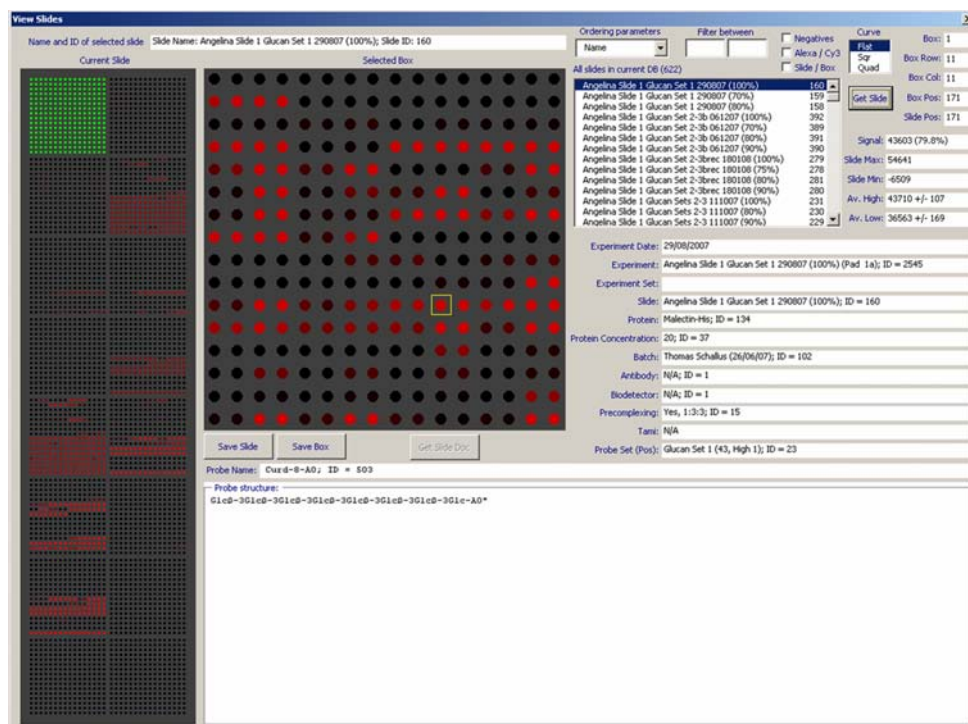


Figure 12. The slide graphic data presentation. The list shows every slide in the database. Selecting a slide will show all its data at once in the same form as that produced by the laser scanner. The enlarged box shows one of the 16 experiments in the slide. Each experiment can be selected in turn. The small yellow box shows the microarray spot selected and all the data associated with that spot are shown on the right hand side of the interface. In the graphic, fluorescence values are shown as black (no signal) to red (maximum binding signal) or green (maximum reference signal). There are options to show the gradient as linear, quadratic or quartic to enhance low values, to show negative values in blue and to normalize over the whole slide or each experiment. The graphics images can be saved as bitmaps for storage.

CONCLUSIONS

The software suite described here is an indispensable tool in support of our carbohydrate microarray programme. It allows rapid analysis of new data and comparison with all stored information. The system at present is tailored to our specific needs and is designed around our screening format.

General purpose multi-format software

We are currently developing a new software suite aimed at much more general carbohydrate microarray systems that can handle a wide range of microarray layouts and that will be able to analyse dose-response and inhibition data in any format. We are collaborating internationally with scientists who are experts in bioinformatics in an endeavour to make this new software available to the glyco-community.

ACKNOWLEDGMENTS

We are grateful to our colleagues for continual feedback through usage which has led to many improvements and valuable additions to the software during development. In this regard we acknowledge, Maria A. Campanero-Rhodes, Angelina S. Palma, Robert A. Childs, Yan Liu, Wengang Chai and Alexander M. Lawson. This work was supported by grants from UK Research Councils' Basic Technology initiative 'Glycoarrays' (GR/S79268); UK Engineering and Physical Research Councils Translational Grant EP/G037604/1; UK Biotechnology and Biological Sciences Research Council (BB/E02520X/1 and BB/G000735/1); UK Medical Research Council (G9601454 and G0600512); Wellcome Trust (WT085572MF, NCI Alliance of Glycobiologists for Detection of Cancer and Cancer Risk (U01 CA128416) and Fundação para a Ciência e Tecnologia, Portugal (SFRH/BPD/26515/2006)

REFERENCES

- [1] Tang, P.W., Gooi, H.C., Hardy, M., Lee, Y. C. & Feizi, T. (1985) Novel approach to the study of the antigenicities and receptor functions of carbohydrate chains of glycoproteins. *Biochem. Biophys. Res. Commun.* **132**:474 – 480.
doi: [http://dx.doi.org/10.1016/0006-291X\(85\)91158-1](http://dx.doi.org/10.1016/0006-291X(85)91158-1).
- [2] Feizi, T., Stoll, M.S., Yuen, C.-T., Chai, W. & Lawson, A.M. (1994) Neoglycolipids: probes of oligosaccharide structure, antigenicity and function. *Methods Enzymol.* **230**:484 – 519.
doi: [http://dx.doi.org/10.1016/0076-6879\(94\)30030-5](http://dx.doi.org/10.1016/0076-6879(94)30030-5).
- [3] Stoll, M.S., Feizi, T., Loveless, R.W., Chai, W., Lawson, A.M. & Yuen, C.T. (2000) Fluorescent neoglycolipids. Improved probes for oligosaccharide ligand discovery. *Eur. J. Biochem.* **267**:1795 – 1804.
doi: <http://dx.doi.org/10.1046/j.1432-1327.2000.01178.x>.
- [4] Chai, W., Stoll, M.S., Galustian, C., Lawson, A.M. & Feizi, T. (2003) Neoglycolipid technology – deciphering information content of glycome. *Methods Enzymol.* **362**:160 – 195.
doi: [http://dx.doi.org/10.1016/S0076-6879\(03\)01012-7](http://dx.doi.org/10.1016/S0076-6879(03)01012-7).
- [5] Feizi, T. & Chai, W. (2004) Oligosaccharide microarrays to decipher the glyco code. *Nat. Rev. Mol. Cell Biol.* **5**:582 – 588.
doi: <http://dx.doi.org/10.1038/nrm1428>.
- [6] Paulson, J.C., Blixt, O. & Collins, B.E. (2006) Sweet spots in functional glycomics. *Nat. Chem. Biol.* **2**:238 – 248.
doi: <http://dx.doi.org/10.1038/nchembio785>.
- [7] Horlacher, T. & Seeberger, P.H. (2008) Carbohydrate arrays as tools for research and diagnostics. *Chemical Society Reviews* **37**:1414 – 1422.
doi: <http://dx.doi.org/10.1039/b708016f>.
- [8] Liu, Y., Palma, A.S. & Feizi, T. (2009) Carbohydrate microarrays: key developments in glycobiology. *Biol. Chem.* **390**:647 – 656.
doi: <http://dx.doi.org/10.1515/BC.2009.071>.
- [9] Goldberg, D., Sutton-Smith, M., Paulson, J. & Dell, A. (2005) Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. *Proteomics* **5**:865 – 875.
doi: <http://dx.doi.org/10.1002/pmic.200401071>.
-

- [10] Maass, K., Ranzinger, R., Geyer, H., von der Lieth, C.W. & Geyer, R. (2007) “Glyco-peakfinder” – *de novo* composition analysis of glycoconjugates. *Proteomics* **7**:4435 – 4444.
doi: <http://dx.doi.org/10.1002/pmic.200700253>.
- [11] Aoki-Kinoshita, K.F. (2008) An introduction to bioinformatics for glycomics research. *PLoS. Comput. Biol.* **4**:e1000075.
doi: <http://dx.doi.org/10.1371/journal.pcbi.1000075>.
- [12] Childs, R.A., Palma, A.S., Wharton, S., Matrosovich, T., Liu, Y., Chai, W., Campanero-Rhodes, M.A., Zhang, Y., Eickmann, M., Kiso, M., Hay, A., Matrosovich, M. & Feizi, T. (2009) Receptor-binding specificity of pandemic influenza A (H1N1) 2009 virus determined by carbohydrate microarray. *Nat. Biotechnol.* **27**:797 – 799.
doi: <http://dx.doi.org/10.1038/nbt0909-797>.
-